LUISS

Department
of Political Science

Ph.D in Politics: Theory, History, Science Programme

Cycle XXXIV

One Size Fits All? An Exploratory Study of Russian and Chinese Information Operations on Twitter

Prof. Dr. Michele Sorice

SUPERVISOR

Prof. Dr. Anja Bechmann

CO-SUPERVISOR

Rolf A. Nijmeijer

CANDIDATE

Academic Year 2023/2024

One Size Fits All? An Exploratory Study of Russian and Chinese Information Operations on Twitter

Doctoral Dissertation

Submitted to the Libera Università Internazionale degli Studi Sociali "Guido Carli"
Department of Political Science

Within the

XXXIV Cycle
of the Politics: Theory, History, Science Programme

By: Rolf A. Nijmeijer

Supervisor:
Prof. Dr. Michele Sorice

Co-supervisor:
Prof. Dr. Anja Bechmann

Many thanks to my friends and family for their support; my supervisors Michele and Anja for their guidance; and all the good people at LUISS for their patience and assistance during the past years.

Contents

Introduction

This dissertation will explore to what extent contemporary state-backed information operations follow similar logics, tactics, and techniques. In other words, if one size fits all. One of the critical issues that has emerged in the fields of communication studies, political science, sociology, and others, has been the rise of disinformation and the various challenges that the phenomenon poses to societies. Simultaneously, an erosion of trust in the institutions and the political processes that shaped (Western) societies in the past decades has been observed (Przeworski, 2019). Contributing to this destructive trend are, purportedly, social media and disinformation. The radical reshaping of the media ecosystem by these 'new media', and the enormous databases of personal information gathered by their owners, make fears about their short- and long-term effects on the health of democracy more than understandable. Abuse of these media by malicious actors, and the convenience and ease with which false information can be produced and disseminated, are indeed cause for concern. This became apparent during the 2016 Brexit referendum and presidential elections in the United States, which brought the disinformation phenomenon to the attention of policy-makers and scholars in the (Western) world.

In response, a wide variety of scholarship has emerged with the aim to explore the many forms of contemporary disinformation, their functioning and effects, and the purveyors involved in the production of this false content. Many of these studies, this dissertation included, have attempted to identify the manner in which propaganda is disseminated through these new media. This rich and ever-growing field of study is multi-disciplinary in the truest sense of the world, which is unsurprising, considering the fact that this issue encompasses nearly the full spectrum of human behaviour, society and technology. However, following after the initial novelty of this threat of hyper technological propaganda, comes the question of *how new* and *how unique* this problem truly is. The field of propaganda studies itself has a long history as well, and involving insights from many different disciplines. The 'best practices' for the manipulation of the subconscious a century ago drew on the latest findings from psychology, in a very similar fashion to the 'neuromarketing' that drives (political) advertising on social media today. The rise of the radio invoked a fear of profound manipulation of the human mind, similar to the public discourse surrounding the role that social media play in our current society. This shall be further explored in the first chapter of this dissertation.

These historical precedents are not just examples that show that *disruptive* new technology will have to be profoundly *destructive* per se, considering how societies adapted to them, but also illustrate that there are certain aspects of propaganda in the mass society that seem to hold true across time and space. One such truism appears to be the notion that propaganda does not radically 'brainwash' its target to adopt beliefs entirely foreign to their previously held convictions. As Jacques Ellul posited sixty years ago, successful propaganda attaches to what is already there, both in the person, and in the society in which they live (Ellul, 1965). In a similar vein, propaganda has to adapt to the medium through which it is propagated and works best if a person that the recipient trusts or knows delivers it (Lazarsfeld & Merton, 1948).

From the aforementioned perspectives, it seems that there is a universal element to propaganda. However, when considered within the specific context in which the manipulative content is spread, this may not be the case. More precisely, there may be psychological and neurological principles that determine how information is processed and retained within a person that are universal for all humans, but there are other factors that will decide how deliberately manipulative content is created and used, and then received by specific target audiences. Strategic doctrine, available technology and means, and the media ecosystem in which the manipulative content is spread, are all variables that will influence a specific instance of deliberate and coordinated attempts at manipulation. Determining to what extent a 'one-size-fits-all' approach to disinformation has merit is what this dissertation aims to achieve.

The framework through which this problem is analysed, is that of 'information operations', which for the purpose of this dissertation is assumed to be any information used by a state actor against another state, state actor, or their population, in order to negatively affect their functioning, decision-making process or opinion forming. However, the term 'information operation' will not be used solely and consistently throughout the entire dissertation, since parts or the entirety of this process have gone by a variety of different names, in various studies throughout the past century.

This dissertation is not a study of the causal relationship between exposure to disinformation that is part of an information operation, and any potential effects thereof on personal beliefs and behaviour; what could be described as the successful fulfilment of the information operation's objectives. Rather, it is an exploratory study of contents and conduct of information operations where, informed by the existing scholarship, the following three hypotheses are leading:

1. Propaganda Logic: successful information operations attach to what is already present in their target.
2. Strategic Logic: successful information operations are adapted to their unique strategic objectives.
3. Medium[1] Logic: successful information operations adapt to the information environment in which they operate.

The first hypothesis is informed by past and contemporary studies on the nature of propaganda operations, which found that introducing completely new ideas to an audience is difficult to accomplish, and that successful propaganda uses pre-existing beliefs and knowledge structures to introduce its information to their audiences. The second hypothesis is based on known military and influence 'doctrines' from the two target countries, which are assumed to shape the objectives and means of the information operations. The third hypothesis is based on the notion that a specific medium operates in a specific manner, in particular social media with their functions and mechanics (hashtags, likes, sharing, repost, etc.). Indeed, adopting certain hashtags in order to 'hijack' discourse on a Twitter topic has been a known strategy for information operations (Darius & Stephany, 2020). As such, an information operation has to adapt to this information environment in order to reach its target and to gain traction within the information ecosystem.

Each step, from the first through the third hypotheses, reduces the level of abstraction, and as such, increases the potential for operations to diverge from one another. Whereas the first hypothesis might assume a potential one-size-fits-all approach to information operations, the third hypothesis assumes that the direct environment of the operation shape its structure and contents, though separate operations conducted on the same platform will likely follow similar logics and apply similar tactics. Together, they will inform the answer to the overarching question that this dissertation tries to answer: to what extent do state-backed information operations on Twitter follow similar logics?

The case studies used in this dissertation to test the hypotheses are information operations on the social media platform Twitter, now called 'X', supposedly conducted by actors from, or acting on behalf of, Russia and China since the year 2010. Both of these states have increased their information operation activities in the past decade, and made it a prominent

[1] Deliberately not identified as 'media logic', which has a broader meaning than what is intended here, which is how the mechanics of a specific medium influence content.

part of their arsenal. Thus, they can be illustrative of the way such operations function in practice. Twitter is the platform of choice due to its prominence as a political communication platform, especially during various wars and crises of the past decade. The platform's use in information operations does fluctuate significantly, due to the turbulent recent history of the platform[2] and social media trends favouring newer platforms. Nevertheless, its role in information operations is still significant. The data used in this dissertation comprises the messages ('tweets') released by the platform itself, as part of the transparency dataset that was published to help counter disinformation[3].

As has been mentioned before, the platform itself has been renamed to 'X', but throughout this dissertation the name 'Twitter' will be used, as well as all terms associated with it, such as 'tweets'. The reason for this is that that name and derived terms are used in the dataset. As for the analysis itself, all the English-language textual data, including metadata, hashtags, tweet text and links to external websites, are analysed though quantitative text analysis in the Python programming environment. The analyses include word frequency, term frequency–inverse document frequency, sentiment analysis and latent dirichlet allocation (LDA) topic modelling.

The hypotheses will be tested through the empirical analysis of the dataset, in order to assess if these assumptions on the nature of propaganda hold true for these information operations on Twitter; to what extent and how the doctrines and strategic interests of these states influence information operations; and to what extent and how the platform shapes the information operations. There are, however, various caveats that should be mentioned here. Due to the use of a pre-made dataset, the opaque nature of information operations and the difficulty to properly identify them, and the fact that such operations often include multiple avenues for information dissemination, it is very unlikely that this assessment provides a full and completely comprehensive overview of these supposed operations. As such, the empirical analyses are illustrative of the principles included within the hypotheses, but will in turn provide the foundation for further research. Thus, the hypotheses guide the empirical assessment, and the empirical assessment puts them to the test and provides insights into how information operations are conducted by these states. If the hypotheses prove incorrect, this

[2] Miah Hammond-Errey, ''Elon Musk's Twitter Is Becoming a Sewer of Disinformation'', *Foreign Policy*, 15 July 2023, https://foreignpolicy.com/2023/07/15/elon-musk-twitter-blue-checks-verification-disinformation-propaganda-russia-china-trust-safety/.
[3] ''Twitter Moderation Research Consortium'', Twitter, https://transparency.twitter.com/en/reports/moderation-research.html.

will provide insights into how these operations may differ from pre-existing trends, and inspire future research to examine thus in more detail.

The data that is used for this dissertation is not unique, since the dataset was harvested and curated by Twitter. It has been examined by analysts from the platform itself, as well as by external researchers.[4] This dissertation's main contributions to the field comprise:

1) assessing the dataset's English-language operations ascribed to Russia and China as a whole, rather than in part;
2) placing these operations in the wider history of information operations and propaganda;
3) comparing these two states in order to assess how they may act similarly or differently in conducting these operations.

Thus it provides a synthesis of several aspects of the study of information operations, that had not been brought together in a similar manner at the time this project was developed. The first chapter of this dissertation will describe the historical and theoretical frameworks of these operations, followed by a chapter which outlines the data and methodology, this is followed by two chapters comprising analyses of the two case studies (Russian and Chinese information operations on Twitter), which in turn are followed by the conclusion.

[4] Twitters own studies of the data have been published on its Blog, see https://blog.twitter.com/. Parts of the dataset have been used by researchers from the Stanford Internet Observatory, among others, see: ''Platform Takedown Reports'', Stanford Internet Observatory Cyber Policy Center, https://cyber.fsi.stanford.edu/io/research/takedowns.

Chapter I
The Theoretical Background of Information Operations

This chapter will discuss some of the past and contemporary scholarship that informs the propaganda, strategic and medium hypotheses, the potentially universal nature of propaganda principles, the adaptation and adoption of strategies that aim to utilise these in an adversarial manner, and what might be done to improve resilience against them.

It can be difficult to discern the true structure, contents and purveyors of information operations. The practical reasons for this obfuscation are clear, since many of these operations are opaque by design. They are intended to take place in the shadows, confusing adversaries as to what the objectives of these operation really are. The confusion that surrounds them, especially in public discourse, can in itself be a form manipulation and misdirection. One example of this is the discourse around the alleged Russian interference in the presidential elections in the United States in 2016, and also the Brexit referendum in the same year. Much has been written about to what extent these interferences took place, and, if they did, how successful they actually were. The prominence of this topic in public discourse may have made this phenomenon appear larger than it actual was. This 'moral panic', as Andreas Jungherr calls it (2021), surrounds prominent manifestations of disinformation and can distract academics and policy-makers from changes and phenomena in society at large, which make disinformation currently salient. Inequality, feelings of insecurity, dissatisfaction with political representation, and reduced trust in institutions, for example, may explain why disinformation gains more traction in contemporary societies in Europe, compared to two decades ago.

Another consequence that ties into this is the potential misjudgement of the actual influence and competence that propagators of influence campaigns actually have. The well-known case of the now defunct consulting firm Cambridge Analytica, and the role it played in effecting the aforementioned events in 2016, has already received much scrutiny. This scrutiny has led to a reappraisal of the firm's significance, towards it being less influential than purported in the firm's own statements (Benkler, Faris, Roberts, 2018). Similar observations have been made by Thomas Rid in his study (2020) of the various Soviet agencies that devoted their time to propaganda, political warfare, and 'active measures'. There is strategic value in presenting oneself as more competent in the art of manipulation than one actually is. In the case of state intelligence agencies, it makes their state's capacity for power projection more impressive. Additionally, the idea that any piece of information that one encounters can be

manufactured with the intent to manipulate will achieve one of the objectives that many of these operations have: the destabilisation of their target and the manufacturing of distrust in existing knowledge structures. For agencies involved in these activities, the added value is that their own influence and prestige within the state apparatus grows. Similar considerations apply to consultancy and public relations firms that engage in the dissemination of propaganda and neuromarketing: it makes them more desirable for potential clients.

The murkiness of influence campaigns extends not just to praxis, but also to the historical, philosophical, sociological, psychological and computer-scientific studies of them. This chapter so far has used many different terms that all cover a (quasi-)similar phenomenon: the use of information to influence (public) opinion and discourse. This variety in terminology characterises the scholarship devoted to this phenomenon, where various studies often study subcategories under the larger umbrella term of 'propaganda', or perhaps study the same phenomenon under a different name. For example, where one scholar may speak of 'political warfare', another may call the same phenomenon a 'disinformation campaign'. One study may focus on the sociology of propaganda and technology, whereas another assesses its strategic utility. This chapter is devoted to bringing together the literature and findings from these various branches of this rich field of study.

A Brief History of Propaganda

Much of the Western scholarship, especially in the Anglophone world, which concerns itself with the creation of the modern theory and practice of propaganda, trace the origins thereof to Edward Bernays. This American public relations agent and nephew of Sigmund Freud wrote several books (1923, 1928) on best practices to influence public opinion, basing his methods on the psychological theories which has uncle had formulated. Here, there is already visible the interconnectedness between the most recent advances in technology and scientific insights, and their use in the 'guidance' of the public. Bernays' contemporary, the influential public intellectual Walter Lippmann succinctly puts into word the need for such practices in his own work on public opinion (1922):

"The amount of attention available is far too small for any scheme in which it was assumed that all the citizens of the nation would, after devoting themselves to the publications of all the intelligence bureaus, become alert, informed, and eager on the multitude of real questions that never do fit well into any broad principle … The purpose, then, is not to burden

every citizen with expert opinions on all questions, but to push that burden away from him towards the responsible administrator" (p. 398-399).

Lippmann's observations on the impossible demands that a high-density information society puts on an individual are salient also, or perhaps especially, in the current information age. However, the notion that the general public should be guided be a few select individuals is easily criticised. It invokes, perhaps unjustly, the image of the totalitarian regimes and dictatorships that use propaganda to bind their subjects to their will. Such 'brainwashing' of the target through the (ab)use of information became a staple of propaganda studies in the 1930s, in the form of the 'hypodermic needle model' or 'one-step-flow' model of communication. This model assumes that information is received directly by the recipient and internalised entirely and unmodified (Hilbert et al., 2017). This model was criticised by Paul Lazarsfeld, based on his own studies on the dissemination and internalisation of information in the 1940s. Lazarsfeld's alternative model suggests that information disseminates through opinion leaders, who influence their followers, whilst they themselves are influenced by other opinion leaders at the same time (1944).

Criticism of these Anglo-Saxon advances in the study of propaganda came from the Frenchman Jacques Ellul, who stressed the difficulty, or rather impossibility, of measuring the exact functioning and effects of propaganda within society at large (1964, 1965). The findings of Lazarsfeld and others were based in small case studies or controlled experiments, which, according to Ellul, fail to represent the complexities of propaganda in society as a whole. Particularly salient for the contemporary debate on disinformation is his notion that propaganda is not an excess of a totalitarian society, but rather a necessity for the modern technological society, which it needs to function. It allows the various segmented parts of society and isolated individuals to be bound together. This particular importance of propaganda for non-totalitarian or pre-totalitarian states was also remarked upon by Hannah Arendt in her writing on this topic in *The Origins of Totalitarianism* (1966). It is this concept, as well as Ellul's notion that effective propaganda attaches itself to that is already present in society that has largely shaped this doctoral project.

The Propaganda Model

The cynicism of Ellul regarding the modern technological society, and the role of propaganda in its functioning, is echoed in Noam Chomsky and Edward Herman's influential 'propaganda model' (1988). Their model attempts to explain the interconnectedness of media, corporations

and government, with a critical appraisal of the criticising function that the former are supposed to adopt towards the latter two. According to the model, media are affected by the concentrated ownership of the media; their reliance on advertisement revenue; their reliance on government information; the negative corporate and government responses to critical media messaging; and anti-communist biases. All of these discipline the media and dissuade them from being overly critical.

Evidently, there are geographical and temporal elements to the model, in the sense that it is based on, and more relevant for, the American media landscape during the Cold War. However, the influence of the model extends beyond these confines, and it may be relevant for contemporary discussions on the role of new/digital media companies, their business models, and their responsibility to facilitate discourse and to combat propaganda. The relationship that is depicted in the model, between media, and the ways in which they may be influenced or beholden to economical and political interests, has elements that can apply to the social media companies of today. Pressure on the latter to enforce strict(er) moderation of content that is deemed harmful by the states and societies in which they operate, and similar pressures exerted by their advertisers, influence the way in which these companies operate.

However, the 'closer' relationship between these companies and their users introduces a different dynamic to the model. Other sources for revenue and information have emerged, through user data and content that allow new media companies to be less dependent on these aforementioned third parties. For example, advertising remains a major source of revenue, but the behavioural models built by tech companies to optimise marketing campaigns, and other utilizations of user data, have added new dimensions to the model. This business model focussed on data gathering rather than news dissemination, which Shoshana Zuboff has called 'surveillance capitalism' (2019), transforms and transcends the traditional relationships between media and audience. The accumulation of personal data, the commodification and sale thereof, its uses in influence campaigns, and the dominance of social media in general, affect the manner in which the propaganda model can applied to such media systems.

These changes do not necessary mean that the propaganda model is no longer applicable to the modern media ecosystem. In many countries, 'traditional media' remain a major source of information for large segments of the population, the elderly in particular. Moreover, after the 2016 US presidential election, American media outlets were very willing to make Russian interference the leading narrative in the news cycle, though the general public did not show an innate interest in this topic, Alan MacLeod argues (2019). Similarly, Chomsky and Herman have both been critical of the Russian interference narrative and overall framing of Russia in

American media, indicating that it serves the interests of the US government and its intelligence agencies. New content restrictions on social media implemented by the major social media platforms, and regulatory hurdles created to hinder foreign media that are critical of the US (such as Sputnik and Russia Today (now RT), are exemplary of the relevance of the five principles for the contemporary US media system (MacLeod, 2019).

These certainly are valid points to raise, in particular considering that free press and free discourse are deemed to be of paramount importance for democracies to function. Questioning dominant, or subtle, narratives in American/Western media systems, and how they may be used to influence public opinion is indeed pertinent and important for a functioning media ecosystem. This has been one of the main contributions of the propaganda model to the study of media, and these contemporary applications of the model apply it effectively to contemporary media. However, this focus on exposing how 'free' Western media are not as free as they appear, seems to generate too charitable an interpretation of opposing narratives. At the time of writing, with RT and Sputnik being banned from the European Union in response to the Russian invasion of Ukraine in 2022, this matter is more pertinent than ever. Preventing media from operating based on political motivations sets dangerous precedents, is potentially harmful to smaller media, and does considerable damage to the credibility of the principle of free press that is professed by Western societies.

However, this presumes that such media were indeed operating as free media, and that any narratives that oppose predominant narratives in Western mainstream media are part of constructive public discourse, produced in adherence to journalistic standards. In the case of RT and Sputnik, presenting an alternative narrative to Western media narratives has indeed always been the objective, but its close ties to the Russian state were a public secret since their inception (Singer & Brookings, 2018), and their media content often included conspiracy theories (Yablokov & Chatterje-Doody, 2022). Therefore, though Western media and governments have generally been antagonistic to Russia and Russian media, this has not been without reason.

The Hybrid Media System

Christian Fuchs argues that the propaganda model is fundamentally about power structures, and in an updated propaganda model for our time should be a critical assessment of social media and the power that they hold. There is an economic, a political and a cultural dimension to this question, which brings us again to a broader, societal approach to this issue (2018). The fundamental disconnect between these past perspectives on propaganda, technology and

society and our modern media landscape, is of course the medium/media in question. Past perspectives discussed media where the message and medium were fundamentally connected, and where the disseminator had full editorial control of the message from production to dissemination. As we have already seen in the previous section, the increasingly dominant 'new', or 'social' media supply the infrastructure for spreading information, but do not necessarily determine each item of content that is disseminated. In other words, it is a *new* media system. When compared to a model like the propaganda model, its complexities become more visible. While the propaganda model assumes a more distinct elites versus population/audience dichotomy, the modern media system is varied, with different actors assuming different roles. Social media are used by politicians to reach their electorates (De Sio & Weber, 2020), but at the same time social media companies increasingly have to comply with regulations[5], driven in part by societal concerns about the threat that they may pose for public discourse. In return, these companies utilize their global reach and the digital infrastructures that they provide, to push back against these efforts, for example by threating to withdraw their services.[6] Lastly, media consumers are data sources, audiences for content, and provide popular pressure for platform regulation.

Andrew Chadwick's influential work on the current hybrid media landscape explores the many ways in which we should "move away from 'either/or' thinking toward 'not only but also' thinking" (p.5, 2017). The rise of 'new media' does not make 'old media' immediately obsolete, rather, there are interconnected groups holding varying degrees of influence, power and prominence. This does not just apply to our current era, but to any era in human history where new technology or social change emerged. Indeed, when we consider the works of the authors and scholars mentioned earlier in this chapter, this concept does seem intuitive. The writings of Bernays and Lippmann show a similar approach to 'managing' the mass society through emerging technologies and insights. Lazarsfeld's descriptions of the fears and worries of his contemporaries regarding the all-powerful radio that will displace the newspaper and manipulate the masses could be published today, if one were to only change the media mentioned. The criticism of Ellul regarding the role of propaganda and technology in modern mass societies does not seem to have lost much of its relevance today.

The resilience of these old patterns and themes is on the one hand a reassurance when one considers current worries about the role of social media in public discourse, the

[5] See for example the European Union's Digital Services Act.
[6] Rory Cellan-Jones, ''Facebook v Australia: Who blinked first?'' *BBC News*, 23 February 2021, https://www.bbc.com/news/technology-56168843.

dissemination of harmful messages and the fragmentation of society in various bubbles. The new is not necessarily politically disruptive and different from the old to the extent that we fear. On the other hand, it is a stark reminder that propaganda, manipulation and the abuse of technology are not new; and it is unlikely that these issues will be solved in a definitive manner.

One example of this process is the steady advance of the information society and the role it has on politics. Philip Howard convincingly demonstrates the growing importance of information technology in the American electoral process, which slowly transitions the realm of politics into a 'hypermediated' campaign that thrives in the interconnected new media ecosystem (2006). Following behind this mediatisation is the role that consultancy firms play in the organisation of modern political campaigns; as well as the importance of grassroots movements in legitimising politics. In the wake of those developments, modern social media politics, built upon the logic of algorithms and the services of public relations firms, seem all but inevitable.

Disinformation and Active Measures

The 1920s were the foundational period for what would become the *disinformation* and *active measures* employed by the Soviet Union and the Russian Republic. Shortly after the Russian Revolution, the new Soviet state realised the importance of intelligence and clandestine measures to safeguard its existence in a world that was hostile to its existence. Through carefully crafted deception campaigns, tsarist loyalists and foreign agents were exposed and arrested (Rid, 2020). While this period in the United States was characterised by the rise of the corporate-driven propaganda crafted by Bernays and successors, the Soviet Union perfected its tradecraft and political warfare capabilities.

This form of political warfare was called 'active measures', and disinformation lies at its core. Though the term invokes the idea of falsehood, it does not mean that a piece of disinformation is false in its entirety. As former KGB-agent and defector Ladislav Bittman describes the term: "Disinformation is a carefully constructed false message leaked into an opponent's communication system to deceive the decision-making elite or the public... To succeed, every information message must at least partially correspond to reality or generally accepted views" (p. 49, 1985).

Though these political warfare tactics are strongly connected to the Soviet Union and Russia, they are part of the arsenal of any state intelligence agency. Much of recent scholarship has been devoted to this Russian use of political warfare, this dissertation included, but that does not mean that their findings, and Bittman's definition, do not apply to disinformation in

the broader, global sense. As Thomas Rid explains, the most fundamental way in which disinformation and active measures have changed over the past decade are the new capabilities that social media offer (2020). These platforms have reduced the costs for creating a disinformation operation and have extended its potential reach beyond that of traditional vectors for disinformation, such as false newspaper articles or leaked documents. The merger of the political, commercial and propagandistic has created a volatile new world where many actors can make their voice heard with a low financial threshold. While foreign information operations were dominant in scholarship following the events of 2016, new scholarship has moved into the domain of domestic manipulation campaigns, which seem to have become dominant in recent years (Somerville, 2020).

In a sense, this new information disorder is a union that ties the American, West-European and Russian branches of propaganda theory together, through the mixing of technology, politics, grassroots movements, corporate influence and societal shifts into one dynamic.

The Spread and Influence of Disinformation

Recent studies on the spread of false information found that such information spread faster and farther than true information, and was difficult to dislodge through fact-checking ex psot the exposure (Vosoughi et al., 2018). Of course, this is very similar to 'clickbait', which is salient and emotionally evocative content to which people gravitate. Such content is the basis for engagement on social media and drives internet traffic to many alternative news websites (Benkler, Faris, Roberts, 2018). It is also another confirmation of the importance of emotionally evocative content for propaganda campaigns to be impactful and the successful (Codevilla, 1992).

The influential study (2017) by Hunt Allcott and Matthew Gentzkow on false news exposure and its effects on the 2016 presidential election in the United States, found that social media were indeed important sources of news to many, but they were not the only news that people consumed. False news that favoured their preferred candidate was predictably more acceptable to the audiences than news that was not. It was difficult to rule out some influence of false news (on social media) on the elections results, and it would indeed be likely, but it is difficult to identify this kind of information as a decisive factor. Another study that was done on the Russian Internet Research Agency's Twitter influence campaign towards the American electorate came to similar conclusions (Bail et al., 2020).

Determining the exact reasons for how and why public opinion is shaped, particularly in the form of political campaigning, is a longstanding issue in the field of political communication. Social media have not improved this predicament, not in the least because of unclarity surrounding the functioning and data collection by these platforms. Some effort has been made by these companies improve access to data, this dissertation is based on such an initiative, but the opacity largely remains. It is therefore unsurprising that the conclusion of many studies on the persuasive significance of disinformation is 'perhaps', and that many studies focus on the structure of disinformation campaigns instead. Once again, this dissertation included. This does not mean that important and illuminating work is not being done on the topic of disinformation's effects, but rather that its clearest effects might be the sowing of doubt and mistrust about institutions, and the manner in which (scientific) knowledge is obtained (Jennings et al., 2021).

Indeed, doubt and mistrust may be the true long-term effects of this phenomenon, and a good argument can be made that this is the main objective of hostile, state-to-state information operations. Democratic societies may find a particularly difficult to counteract these effects, without impeding on their citizen's freedom of expression (Paterson & Hanley). A more pessimistic view on the relationship between popular opinion, (Western) democratic political process, and the effects of false information thereon, as expressed by Cailin O'Connor and James Owen Weathertall (2019), supposes that this issue cannot be resolved within the current systems of political representation. More positive perspectives (Howard, 2020) pose that more regulations for social media platforms may create a healthier online media environment. Nevertheless, the recurring theme in recent literature is that profound revisions are required for the process of aggregating popular opinion to inform political action, and that restrictions of media systems and online platforms are sorely needed to remedy the aforementioned issues.

Beyond the harmful effects on democratic processes and freedom expression, there are other dangers involved in attempts to counter disinformation. One has already been mentioned before: the failure to view disinformation within its larger societal framework, leading to miss-assessments of its potential causes, effects, and solutions (Lewandowsky, Ecker & Cook, 2017). The 'moral panic' framework is useful to place the disinformation phenomenon within its technological and societal developments, and may help to explain these developments themselves (Carlson, 2018). However, adopting such a macro-perspective may create an opposite adverse effect, where the harmful effects of disinformation and its exacerbation of pre-exiting harm are insufficiently acknowledged, which Li Qian Tay et al. mainly interpret as a trend within Western literature on the topic. (Tay et al., 2024). This not only raises the issue

of whether these assessments of disinformation are correct, but in what manner they are shaped by their cultural environment, and in response to which specific disinformation threats.

These different assessments and paradigms regarding the phenomenon of disinformation not only have implications for academia and other researchers on the topic of disinformation and coordinated influence campaigns. As has been mentioned already, Western policies aimed at countering the phenomenon of disinformation, security policy in particular, is struggling to find a response that catches the many legal, moral, societal, technological and practical aspects of disinformation. Informed by the scholarly and other research trends that have emerged in the past years on this topic, they often have to sacrifice one for the other (Hellman & Wagnsson, 2017).

The study of disinformation in Western democracies has grown significantly in the past ten years, largely driven by observed and anticipated antagonistic actions from states hostile to these same democracies. The renewed interest not only applied to the tactics, techniques and protocols (TTPs) of information operations, their structure, contents, effects, and the purveyors involved, but also the strategic cultures that shaped them. The latter certainly is important, as the next few paragraphs will address, but it also raises the issue to what extent the strategic needs and cultures of these Western states themselves influenced the study of disinformation. These may create blind spots regarding the study of information operations, especially if we consider that culture is important for their implementation and success, since the cases studies, cultural frames of reference, specific threat scenarios and strategic needs, and local legal, cultural and moral principles may all (inadvertently) shape how the phenomenon is studied.

Disinformation in International Relations

The section on disinformation and active measures earlier in this chapter already touched upon the discrepancy in appreciation for disinformation tactics between the Soviet intelligence agencies and their Western counterparts. It is important to note that that though the former had and have an established reputation on that front, the latter were not against using political warfare similar to that of the Soviet intelligence services. Nevertheless, academic schools in political science and international relations did not share this enthusiasm for the use of falsehoods as doctrine to advance political interest. La Cour shows that lying in politics and diplomacy has indeed been studied frequently, but that scholars either discarded lying as diplomatically harmful in the long-term, or as a means to obtain domestic gain (La Cour, 2020). This vacancy is rapidly filled by contemporary scholars, as current events increase interest in this topic.

The lack of interest from the academic discipline of international relations for the phenomenon of disinformation likely stems from the predominant belief that its effect within the realm of state-to-state conflict is limited. André Gerrits posits that from a security perspective, the harmful effects of disinformation are limited. He does acknowledge that the effects are more pronounced and worrisome on a domestic level (2019). Since we have already established that pinpointing the precise effects of disinformation is difficult, it is unsurprising that this extends to the state-to-state level which operates farther away from public opinion. Nevertheless, the erosion of trust in politics and institutions, and the potential exploitation of this process by populist politicians, may manifests itself on the world stage in the long-term.

This is acknowledged by Alexander Lanoszka (2019), who mentions this destabilising effect that disinformation targeting a country's population may have. Other reasons for propagating disinformation to international adversaries may be to influence the domestic front of the propagating country instead, through the projecting of lies abouts its foreign enemies. The final reason Lanoszka identifies is one we have identified before: an incorrect assessment by the propagator of its own capacity to influence other states and populations.

These recent advances in scholarship once again raise the issue of what the true objective of foreign information operations is. This might be persuasion, as in convincing the American public to vote for Donald Trump, for example. Obfuscation, as in creating misdirects regarding culpability for the shooting down of Malaysia Airlines Flight 17 by separatist forces in the Donbass Region in 2014. The likely, and rather banal, answer is that there is no singular objective, and that strategic logic, the intended audience, and the intended media vectors used in the operation are decisive in setting the objectives and means. Indeed, information operations are useless if they are not developed in conjunction with other state resources and policies. In addition to these considerations, propagators of such operations also must be aware of the unreliable nature of these operations, the results of which can be hard to predict, and even harder to control once the operation has been launched (Codevilla & Seabury, 2006).

Information Operations

Militarily speaking, what has been called information operations up this point in the dissertation is but one aspect of the military concept of the same name. The other aspects of this kind of operation can be electronic warfare, operations security, psychological operations, military deception and computer network operations (Allen, 2007). This is not a conclusive list: in American military doctrine, there exist slight differences in definitions and subcategories between the different branches of the armed forces, not to mention differences

between various armed forced worldwide. What all of these doctrines share is that they use information in an adversarial manner against the diplomatic, military, economic, information and population resources. The integration of these formerly more disparate parts of national power grown closer and more intertwined through technology, thus increasing their vulnerability to these informational acts of war (Allen, 2007).

As we have already established, adopting social media into this arsenal becomes attractive if there is a strategic interest in targeting the population, but the parameters for success can be hard to determine for outsiders based on observation alone. What makes social media attractive for such an avenue of attack are the cheap and easy means they offer to reach a global audience, and the opinion leaders that may help spread the message to new audiences. Based on the assumption that a successful operation of this nature has to attach to what is already present in a society, it can prey on existing social issues, salient public discourse and trending hashtags (Prier, 2017; Hoogensen Gjørv & Jalonen, 2023).

For the information operation practitioner on social media such as Twitter, high jacking of hashtags and bot spamming, the use of automated accounts to bombard certain conversations on social media with spam messages, can be used to drown out undesirable messages, or instead, boost certain conversation topics. These tactics have been used in Syria, to disrupt the Twitter media discourse by the opposition to Assad's regime. However, as new trends in social media emerge and decline, it is important stay abreast of the new trends in order to determine the optimal method to disseminate information. In other words, emphasise the 'social' of social media, especially in new information operations on platforms like the personality-heavy Instagram and TikTok, which have emerged as the new frontlines for information war (Sidorenko-Bautista, Alonso-López, Paíno-Ambrosio, 2023).

Russian Information Operations

The long history of information war within Russian strategic thought has been mentioned before, but the most recent and current incarnation thereof has a more recent origin: the Russian invasion of Georgia in 2008, where the Russian troops performed poorly in regular warfare. Another catalyst was the wave of revolutions that swept through former Soviet states and the Middle East in the early 2000s and 2010s. Moscow aimed to create a strategy and doctrine that would allow it to exert influence and project power beyond its border, whilst avoiding all-out war. This doctrine of hybrid warfare comprised cyberwarfare, clandestine actions and the information warfare that have discussed in the previous paragraphs (Thornton, 2015). However, hybrid warfare, or *gibridnaya voyna*, has a long tradition in Russian strategic

thinking, and indeed in military history at large. Subversion is the central theme of this kind of warfare, and its exact components and manifestations may differ from Western concepts of hybrid warfare (Fridman, 2017). Russian deception warfare under the umbrella term of *maskirovska* (masking) and the concept of reflexive control are other aspects within Russian strategic thought. Reflexive control is manipulation of an enemy through study of his characteristics, manipulation tactics, and other means, aimed at provoking a specific response (Giannopoulos and Smith, 2021; Thomas, 2004, 2019). The complexities of the various terms, traditions and concepts in Russian strategic thought may be hard to grasp within the terms written here. The aforementioned *gibridnaya voyna*, for example is also what the Russians call the Western concept of hybrid warfare, which is described as being openly challenging an attacking Russia and Russian interests (Galeotti, 2019). Indeed, as reflexive control-expert Vladimir Lefebvre posits, cultural differences that shaped respective Western and Eastern/Russian outlooks and understandings of certain concepts themselves are harmful and disruptive, and the one who best understands and utilises this will be more effective in political confrontations (Lefebvre, 1982).

The Russian incursion in 2014 and subsequent annexation of Crimea in 2014 brought hybrid warfare to forefront and has shaped Western perception of Russian intentions and capabilities. This recent incarnation of Russian hybrid warfare became known under the name 'Gerasimov doctrine', but the term has fallen out of favour since, as the TTPs included in this term may never have been a formal military doctrine (Fridman, 2019,[7]). Indeed, looking at Russian strategy solely through such frameworks may be dangerous, according to Bettina Renz, for it may colour any Russian action as a calculated, militarised move that fits in a master strategy (2016). This could lead to overestimations of Russian capabilities, similar to the potential over appreciation of the role that disinformation can play in shaping public opinion. The progression of the ongoing war in Ukraine, where Russian objectives, tactics, and execution are often not synchronised, lends credence to this assessment. Indeed, there have been instances of Russian interference, economics coercion, military incursion and cyber attacks, aimed at European Union and other Western states. However, the actual success of these operations are not clear, and neither is to what extent they were part of larger strategic operations (Hågen Karlsen, 2019).

[7] Mark Galeotti, ''I'm Sorry for Creating the 'Gerasimov Doctrine''', *Foreign Policy*, 5 March 2018, https://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/.

Studies analysing Russian information operations on Twitter have shown their tendency to adopt and attach to polarising topics and language, with language that is antagonistic towards Western governments and institutions. Often, this means targeting radicals on both ends of the political spectrum (Al-Rawi, 2021). Moreover, the persuasive effect of these campaigns on the individual level remains difficult to determine, but it appears to be less than feared or expected. The reach of these campaigns vis-à-vis the larger Twitter ecosystem is limited. Both of these observations can be explained by the targets and contents of these campaigns, since the radical fringes that they are aimed at are more inward-focused and already strongly entrenched in their views (Bail et al., 2020). Similar observations regarding the amplifying, rather than influencing nature of these campaigns were made by Dawson & Innes (2020) and Eady et al. (2023). The mechanisms of Russian influence abroad comprise many different media, who can operate in conjunction with each other. However, the reach of these various media channels within the larger media ecosystems in which they operate appears to be limited. Much like the Twitter campaigns, RT reaches a limited audience with its narratives critical of the West, compared to the Western media enterprises it competes with (Elswah & Howard, 2021). Thus, following the findings from these various studies, it seems likely that the actual reach of Russian information operations, and their persuasive effects on individuals, may be limited. However, the mere existence of these campaigns, as well as their amplification of polarising topics and content, may have a destabilising effect on societies as a whole, in particular by generating doubt on the veracity of any information and narrative in public discourse. This may indeed be the ultimate objective of these campaigns.

Chinese Information Operations

The shared opposition to Western influence in areas of interests and their internal politics have been cited as reasons for ever closer ties between Beijing and Moscow, including in matters of foreign policy towards the West. However, this relationship is complex, and one should be wary of overstating their closeness (Grajewski, 2022). Where the cooperation and influence between both countries grows more apparent, is that of best practices for information operations. Whereas Chinese doctrine on this matter used to be more domestically oriented, the country's growing global ambition and influence has led to it adapting a more forceful approach to diplomacy and foreign affairs. Russia's hybrid warfare strategies serve as a blueprint for this new power projection (Bachmann, Dowse and Gunneriusson, 2019), though are not its sole inspiration, since American information warfare doctrine has also been influential.

China already adopted the 'Three Warfare Doctrine' at the turn of century, which involves psychological warfare, media warfare and legal warfare. In this doctrine, information operations are integrated within the wider strategic objectives of the Chinese state and its institutions. Legal warfare uses principles of international law to advance Chinese interest abroad, in particular in its efforts to secure its claims over the South China Sea. Psychological warfare aims at demoralizing enemies of the state, foreign and domestic, through coordinated efforts to target and amplify their weaknesses. This is illustrated by China's bellicose rhetoric towards Taiwan and frequent incursions into the latter's territory. Media or public opinion warfare aims to influence foreign audiences and to improve their opinion of China (Charon & Jeangène Vilmer, 2021). Indeed, persuasion appears to be the objective of these operations, whereas amplifying divisions and doubt within target countries falls more into the field of psychological warfare (Cheng 2017). A recent study on an influence campaign on Facebook, which has been attributed to China, did indeed find that the content of this campaign was predominantly positive about the country and its policies (Tambe & Friedman, 2022).

It appears that Chinese information operations so far have been aimed more at improving the country's international standing, rather than sowing divisions like the Russian campaigns appear to do. This does not mean that Chinese information operations are never used for that purpose, but rather that the general trend differs. However, there is some convergence between both countries' on certain topics. The more recent proliferation of Russian-esque disinformation from China followed the Russian information output surrounding the invasion of Ukraine, as well as its media playbook by adopting a Chinese version of Russian foreign influence vector RT (Li, 2022). In general, the information operations strategies of both countries appear to become more similar (Charon & Vilmer, 2021).

As was the case with the study of Russian information operations, recent studies of Chinese operations and doctrine place them within the country's long history and older philosophies on themes of (strategic) deception and manipulation. Writings from ancient philosophers, such as Sun Tzu's *The Art of War*, are used to interpret current Chinese strategy in the realm of information warfare (Fangyi Lu, 2022). These foundational text will undoubtedly inform these strategies in some form, but the question is to what extent. These and other studies do mention the unique domestic situation for China, with its strong censorship of online communications; the specific characteristics of online social media that are used in information operations and affect how they should be implemented; and lastly the large Chinese diaspora, that offers a unique vector for exerting influence abroad. Considering these

many variables that shape strategies, tactics, and specific objectives of information operations, it may be too deterministic to interpret the actions of China, and also Russia, purely from within the confines of their cultural and strategic doctrines, as envisioned by (Western) observes. Doing so may obfuscate and hamper the study of these operations.

This chapter has outlined the theory and concepts that have shaped the hypotheses of this dissertation. We seen some of the persistent similarities between past and present propaganda principles, which have informed the Propaganda Logic Hypothesis. In addition to these, the various political and strategic considerations that may determine the implementation and design of state actor information operations have been explored, which have informed the Strategic Logic Hypothesis. Lastly, some reflections on contemporary social media, their functioning and their potential societal impact have been offered, which in part have shaped the Medium Logic Hypothesis. The latter is also depended on the medium in question, which will be further explored in the next chapter. Overall, this chapter has shown that contextual factors can exert significant influence on the shape that a coordinated information operation takes, but also that our own perceptions, definitions and frameworks to assess these operations are equally influenced by such factors.

Chapter II
Methodology

The previous chapter has described the state of the field of disinformation and propaganda studies, and described some general trends that have been observed and which will guide the empirical analysis. The data for the empirical analysis part of this dissertation is extracted from the social medium Twitter. Though the platform has limited the amount of data that can be extracted, in order to prevent abuse, it is still possible to obtain a significant and relevant dataset through the data that has been released by the platform itself from 2018 onwards. This data is particularly useful for this dissertation, since it has been designated as 'disinformation' by the platform itself, and is assumed to have been created as part of coordinated information operations.

Twitter as a Vector for Political Communication

Twitter was founded in 2006 as a platform for microblogging. Users can share updates on their lives, personal thoughts and any other information they might want to share in the form of small messages called 'tweets'. These tweets allow a maximum of 140 characters per message. Users can designate who can see its contents, within or outside their network of connections, and also add media in the form of pictures and videos. Users can also reply to tweets from others and share them on their own profile through retweeting. In order to help boost exposure of the message through the algorithms used by the platform, users can add keywords to the message in the form of #hashtags.

At the time of the data analysis, Twitter had more than 354 million users[8] and remains one of the premier social media, though its importance has dwindled compared to newer social media such as the video-based medium TikTok. Nevertheless, Twitter remains important in the realm of political communication, where it has been a major tool for political campaigning and for politicians to reach their audiences, as well as for traditional media to obtain insights in rapidly evolving news stories (Jungherr, 2015). Twitter played a particularly influential role in the series of popular uprisings known as the Arab Spring (2010 - 2012), in which the platform was used to coordinate protest and share updates on the developing situation across countries in North-Africa and the Middle-East. It is in the context of these uprisings that some of the

[8]''Number of Twitter Users Worldwide'', Statista, https://www.statista.com/statistics/303681/twitter-users-worldwide/.

forms of disruptive information operations, such as bot spamming, were deployed and refined (Singer and Brooking, 2019).

The platform's prominence as a means for political communication makes it an attractive tool for those engaged in information operations, and those who wish to study them alike. One major caveat for both to keep in mind, however, is the relatively limited audience that the platform has, compared to the population at large. Combined with a relatively older user base compared the image- and video-focused social media, political communication on the platform should be considered more niche and elite, instead of being at the vanguard of online social media trends. Moreover, since the platform was acquired by Elon Musk, who since renamed it 'X', policies regarding moderation and data acquisition have become more fickle and unpredictable. This has made the platform attractive for disinformation again, after years of anti-disinformation measures. Simultaneously, extracting data via the application programming interface (API) to study information operations became more difficult due to the platform's new policy of demanding payment to do so.[9] These developments took place well after the datasets that have been used in this thesis were released, but they are indicative of the platform's continued relevance in the field.

The Twitter Transparency Dataset

Twitter itself is aware of its role as a platform for political communication, partially due to criticism it received regarding its limited efforts to combat propaganda and information operations that were proliferating on the platform over the years. In response to this criticism, the company launched its election integrity dataset initiative in 2018. This initiative released metadata of users that were considered to be bad faith actors by the platform, as well as the tweets and media posted by these accounts. The company decides to designate an account part of an information operation through the platform's own metrics and in cooperation with law enforcement agencies (Gadde and Roth, 2018). Any person who believes they have been falsely identified as a bad faith actor can appeal to the company to have this status revoked.

The current dataset[10] comprises over 200 million tweets that the company alleges originate from intelligence agencies and other coordinated influence campaign organisers, stemming from twenty-two countries, with user activity going as far back as 2009. The most

[9] Jenae Barnes, ''Twitter Ends Its Free API: Here's Who Will Be Affected'', *Forbes*, 3 February 2023, https://www.forbes.com/sites/jenaebarnes/2023/02/03/twitter-ends-its-free-api-heres-who-will-be-affected/.
[10] ''Twitter Moderation Research Consortium'', Twitter, https://transparency.twitter.com/en/reports/moderation-research.html.

prominent states in the dataset are Russia, China and Iran, but countries from Europe, the Middle East, Latin America, Asia and Africa are included as well (see table 1). All the media items – urls of external websites, pictures, and videos – that are tweeted and retweeted by these accounts are included in the datsets. The relevant data for this dissertation comprise accounts and tweets attributed to Russian and Chinese state-sponsored actors, comprising a total of 3780 accounts and 8.892.280 tweets for Russia, and 2159 accounts and 67193 tweets for China (see table 2). The tweets were collected in the period October 2018 to December 2021. The data was downloaded from the Twitter repository on 15 March 2022.

Since the reviewing process of account activity is continuously ongoing, accounts can be removed from the dataset entirely, in case if wrongful identification as a malicious account, or moved from one country subset to another, in case of misidentification. This has indeed been done[11].

Table 1. Countries

Countries Included in the Dataset	
Armenia	Qatar
Bangladesh	Russia
China	Saudi Arabia
Cuba	Serbia
Ecuador	Spain
Egypt	Tanzania
Ghana	Thailand
Honduras	Turkey
Indonesia	United Arab Emirates
Iran	Uganda
Mexico	Venezuela

Table 2. Dataset Metadata

Dataset Metadata						
Country	Accounts	Tweets	Retweets	English Language (Total)	Period Tweets Published	Period Tweets Harvested
Russia	3780	5530978	3361302	1970861	09 May 2009 - 17 March 2021	October 2018 - December 2021
China	2159	59346	7847	22356	24 January 2012 - 05 April 2021	August 2019 - December 2021

[11] Vijaya Gadde and Yoel Roth, "Enabling Further Research of Information Operations on Twitter", Twitter, October 2018, https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.

Advantages of Using Twitter Data

The main advantage of using this data for analysis is that it allows for the use of large amounts of Twitter data, without violating the company's rules on data extraction and data abuse. This has been a major obstacle for researchers interested in analysing any of the social media, ever since abuses like the Cambridge Analytica scandal have made the platforms more restrained with sharing their data. By using data that has been made available by Twitter itself, this issue is circumvented. Extraction of data using Twitter rest and streaming API would have been possible, but this is limited to a certain number of tweets extracted per user account per hour. This is a bottleneck for downloading relevant data on topics that are quickly developing and involve an enormous number of tweets. An additional obstacle has recently emerged, in the form of the platform's new pay-to-use policies regarding APIs.

In addition to the problem of obtaining the data, there is another major complication that is avoided by using this dataset: the identification of disinformation content itself. The qualitative and quantitative methods for identifying fake accounts posting disinformation, and other forms of coordinated manipulation on social media, are rapidly and ever-evolving. These methods involve analysis of (ir)regular positing times for tweets and retweets, where messaging at regular intervals can indicate that automated accounts (bots) being involved. Automated text analysis on content can help identify messages that are intended to be manipulative. The exact manner in which the data that is used is found and extracted (by searching for certain keywords, for example), how the data is analysed, and how the results are classified (genuine activity, manipulated activity, hybrid activity) remains a contentious topic that has to be addressed by each study that uses these methods (Nimmo, 2019). In addition to the above, due to their small size or lack of successful interaction with the platform environment at large, many information operations likely are not noticed before they are taken down by the platform in question (Buntain & Goldbeck, 2017).

By using the data released by Twitter, these issue are avoided. One way to identify manipulated accounts is to scrape the same data twice and identify the tweets and accounts that have been removed for violating Twitter's terms of service. However, these offensive messages may have been posted for any number reasons that do not involve coordinated manipulation attempts, and false misidentification of coordinated influence campaigns is a constant risk. It is highly likely that Twitter has other, additional metrics and insights that it can use to identify coordinated activity on its platform (Beatson et al., 2021). This is particularly plausible when one considers the platform's cooperation with law enforcement and intelligence agencies, since these have their own metrics that can help in identifying bad faith actors. Using the data that

has been curated and released by Twitter itself solves many of the issues that surround disinformation identification, though it can be seen as a 'copout'.

Disadvantages of Using Twitter Data

The clear disadvantage to using this data is that editorial control is ceded to the platform and any standards that it may hold. Information operations by themselves are already opaque, and clarity is nit improved when the exact criteria by which a dataset is created are not known. It seems Twitter generally follows the thread left by accounts that have violated their terms of service, to identify other accounts that frequently interact with these obvious 'bad' accounts. This network of accounts and messages is the foundation for the dataset. However, when we consider the platform's own interests, and the role that power structures play in the managing of media (per Chomsky and Chadwick); and when we consider the black box that is the platform's curation process; some scrutiny of the dataset is required.

What the terms of service are precisely, and what constitutes a 'violation' of these terms, is not something that is decided upon in abstraction. The norms by which the platform operates, how freedom of discourse is defined and when it is deemed proper for the company to intervene are shaped by social and political developments. An example of this is the very existence of this dataset, which was created due to public pressure regarding the role the platform played in the propagation of disinformation during the 2016 presidential election in the United States (Twitter, 2018). The restrictions placed on political advertising on Twitter during the 2020 presidential election in that same country, and the banning of Donald Trump's Twitter account in 2021, are other manifestations of the pressure that society and political events exert on decision-making within the company. Whether those pressures are justified is not relevant to this point.

With those considerations in mind, the contents of the dataset do become somewhat indicative of (geo)political trends. The prominence of Russia, Iran and China can probably be explained by increased fears and interests in Western countries towards those states, to which Twitter responded in this manner. This does not mean that these states are not acting disruptively and maliciously on these platforms, but rather that there is a vacancy in the dataset regarding coordinated manipulative content from Western sources (with the notable exception of Spanish disinformation regarding the Catalan question[12]). For full integrity and academic

[12] ''Disclosing new data to our archive of information operations'', Twitter, 20 September 20219, https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019.

interest, an expansion of the dataset into that direction would be much appreciated. Indeed, one of the few studies that analysed information operations that have been attributed to Western intelligence services found that the tactics, techniques and protocols they employed were very similar to those used by Russia and Iran (Graphika & Stanford Internet Observatory, 2022). Further study of these operations, the tactics and means employed, and the similarities and difference between these various actors is needed.

In spite of these considerations, the general assumption in this dissertation is that the dataset overall is a genuine representation of accounts involved information operations' activity on Twitter. The data, or parts thereof, have already been used in other studies, most notably several conducted by the Stanford Internet Observatory[13], and in Twitters own reports[14]. However, to understand the wider context of these alleged information operations, the actors involved, and their objectives, this partial dataset and methodology is insufficient, and a more in-dept analysis across multiple platforms and languages is needed to analyse such operations in full. Due to constraints in the form of data availability, this dissertation will only analyses these Twitter datasets.

Data Analysis Methodology

The analysis done in this dissertation on the Russia- and China-ascribed Twitter operations involves automated quantitative text analysis in the Python environment, the code for which has been adapted from several sources. In order to process the large amount of data and to identify the major trends that the alleged information operations try to attach to, analysis was done on the main tweet texts and hashtags. The dataset contains tweets in a large variety of languages, but this dissertation is based on English-language tweets only (numbering a total of 1.970.861 tweets for the Russia datasets, and 22.356 tweets for the China datasets), in order to ensure consistency in comprehension of the texts, and to gain insights in information operations that have the goal to reach as wide an audience as possible. Automated translation of the non-English tweets was considered, in order to identify potentially unique characteristics in non-English language operations, but this was abandoned. These automated translations would entail a further loss of linguistic nance and meaning, since the automated text analysis will already lose some of the subtleties and hidden meanings of the tweets.

[13] ''Platform Takedown Reports'', Stanford Internet Observatory Cyber Policy Center, https://cyber.fsi.stanford.edu/io/research/takedowns.
[14] ''Twitter Moderation Research Consortium'', Twitter, https://transparency.twitter.com/en/reports/moderation-research.html.

Visual data was not included in the analysis, since the aim was to analyse the English datasets as a whole, and the total number visual media would be impossible to analyse. Also, due to the enormous scale of the datasets and limitations in available resources, automated text analysis was considered the best option to examine the contents of these datasets.

Metadata Analysis

The first part of the analysis involves simple quantitative analysis of metadata (account creation dates, tweet activity, tweet/retweet ration, tweet languages, see table 3). This analysis is done on the *entire* Russian and Chinese datasets, and not just the English-language tweets. This is done to assess overall trends within the subsets of the dataset, such as account creation dates, which may indicate preferences for the type of accounts and activities used by these operations, such as bot swarms, or sleeper accounts that can be activated when needed.

For the case study Russia, the subsets comprise tweets allegedly propagated by four campaigns (IRA, GRU, REA and RNA). The IRA is the Internet Research Agency, Yevgeny Prigozhin's now defunct 'troll factory'. The GRU dataset comprises campaigns ascribed to the *Glavnoje upravlenije General'nogo shtaba Vooruzhonnykh sil Rossiyskoy Federatsii*, the Russian military intelligence agency. REA and RNA are unknown acronyms. For the case study China, the subsets comprise tweets allegedly propagated on behalf of the Chinese state: the CNHU-subset (government-affiliated) and the CNCC-subset, propagated by private firm Changyu Culture.[15]

[15] ''Analysis of Twitter Takedowns Linked to Venezuela, Mexico, Tanzania, and China'', Stanford Internet Observatory, 2 December 2021, https://cyber.fsi.stanford.edu/io/news/twitter-takedown-december-2021.

Table 3. Metadata Categories

Twitter Account Metadata Categories	
tweetid	in_reply_to_userid
userid	in_reply_to_tweetid
user_display_name	quoted_tweet_tweetid
user_screen_name	is_retweet
user_reported_location	retweet_userid
user_profile_description	retweet_tweetid
user_profile_url	latitude
follower_count	longitude
following_count	quote_count
account_creation_date	reply_count
account_language	like_count
tweet_language	retweet_count
tweet_text	hashtags
tweet_time	urls
tweet_client_name	user_mentions

URL Sharing

In order to identify connections to other potentially disruptive content, such as alternative news sites which may propagate disinformation, the most shared links ('URLs', table 2) to other websites are identified. The assumption is that Tweets used in an information operation will link to other segments of the information operation.

Word Frequency Analysis

After filtering the dataset on English-language tweets only and cleaning the filtered dataset of English-language stop words, emoticons, URLs and links, the second part of the analysis involves word frequency analysis of the main body of the tweets ('tweet_text', table 3) and the hashtags. Wordclouds are generated to identify the key concepts that the tweets discuss and/or attach to. Any hashtags in languages other than English were translated using Google translate. Hashtags are also presented in a heatmap built from a correlation matrix to identify patterns between them, so as to see how hashtags are used to generate traction for the operations on Twitter.[16]

[16] Code adapted from: James, ''Topic Modelling In Python Unsupervised Machine Learning To Find Tweet Topics'', Coding Club, https://ourcodingclub.github.io/tutorials/topic-modelling-python/.

TF-IDF (Term Frequency-Inverse Document Frequency)

To find the most unique terms used in the tweet texts and hashtags, the data was trained and analysed through the methods of Scikit-learn machine learning[17] TF-IDF (term frequency-inverse document frequency), which extracts the most prominent terms relative to each document (tweet_text) it has been extracted from. This will give a rough impression of the overall themes of these operations, which will be refined through topic modelling at a later stage.

Sentiment Analysis

In order to identify the overall sentiment that the tweets and retweets try to evoke, the tweet text for each subset is ran through Flair Sentiment Analysis automated sentiment analysis[18], which attaches emotional values (positive or negative) to the words in the body of the tweets. This is an automated process, which means that subtleties like sarcasm or humour are difficult to identify, which in turn means that this method gives a rough indication of the overall emotional appeal of the tweets. It is therefore not an absolute reflection of the emotional content of the tweets. The assumption here is that Tweets from Russian operations will lean negative considering their preference for divisive content, whereas Tweets from Chinese operations will lean positive, to as part of their effort to present China in a favourable light to foreign audiences.

Latent Dirichlet Allocation (LDA) Topic Modelling

Lastly, the tweet_text data was processed via Latent Dirichlet Allocation (LDA) topic modelling, a probabilistic model that utilises unsupervised machine learning to extract the main topics from a corpus[19]. This widely used method for quantitative text analysis is useful as it allows for grouping documents together, in this case Tweets, based on their features (Chatterjee & Krystyanczuk, 2017). This would help identifying overall themes in the information operations included in the dataset, which will provide more insight into the overall objectives for these operations. The IRA-dataset was by far the largest dataset, and the enormous number of tweets not only required an unsustainable amount of computation power, but also made

[17] Code adapted from: ''Working With Text Data'', Scikit Learn, https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html?highlight=machine+learning.
[18] Code adapted from: nasirsoft93, ''Flair Sentiment Analysis'', GitHub, 12 December 2021, https://github.com/nasirsoft93/NLP/blob/master/Flair_Sentiment_Analysis.txt.
[19] Code adapted from: James, ''Topic Modelling In Python Unsupervised Machine Learning To Find Tweet Topics'', Coding Club, https://ourcodingclub.github.io/tutorials/topic-modelling-python/.

identifying any topics and relevant clusters impossible. This dataset was therefore analysed in smaller batches.

These various tools for quantitative text analysis are expected to show the general themes, some rudimentary structure, and the contents of these information operations. This in turn can help provide insight into the overall objectives of these information operations and the entities that created them, as well as how they adapted to the specific chrematistics of the platform, as per Hypotheses 3 (Medium Logic). The results of these analyses will be presented and analysed in the chapters that follow.

Chapter III

Case Study – Russian Information Operations on Twitter

The alleged Russian Twitter dataset (see figure 3) can be split in subsets of information operations by two entities. One of these is the Russian military intelligence agency Glavnoje upravlenije General'nogo shtaba Vooruzhonnykh sil Rossiyskoy Federatsii (Main Directorate of the General Staff of the Armed Forces of the Russian Federation), generally abbreviated to GRU. The other one is the Agentstvo internet-issledovaniy (Internet Research Agency), generally abbreviated to IRA, which is assumed to have tried to influence the elections in the United States and the Brexit Referendum in 2016 (Singer and Brooking, 2018). The two remaining datasets are subsets of the IRA operations, with particular focus on certain geographical regions. These are called RNA and REA, for reasons unknown to the author.

Table 4. Russia - Metadata

Russia	Agency	From	Till	Accounts	Tweets	Retweets	Total	English Language Tweets
	IRA	09/05/2009	27/12/2020	3644	5484222	3354693	8838915	1937873
	GRU	09/06/2014	18/10/2020	70	22938	3746	26684	14422
	REA	14/08/2012	17/03/2021	16	7663	60	7723	6823
	RNA	20/09/2012	28/12/2020	50	16155	2803	18958	11743
	Total	**09/05/2009**	**17/03/2021**	**3780**	**5530978**	**3361302**	**8892280**	**1970861**

Metadata Analysis

The IRA's activities started shortly after the war in Georgia in 2008 that spurred on the Russian government's shift towards hybrid warfare. These initial actions consisted of creating a large number of accounts to be used in later operations, but not much actual activity (see Appendix B). This holds true for all subsets. The most likely reason for this is to have legitimate accounts available for when the necessity to use them arises, rather than to rely on a mass of throwaway accounts that seem less genuine due to their recent creation data and lack of a distinct user history.

The overall trend for all agencies is to rely on original tweets, rather than spamming retweets, to influence the discourse on the platform. Similarly, since the English language dataset contains very few duplicate messages, spam seems to be rare in all of the subsets. Based on the language divisions in the dataset, the GRU, REA and RNA subsets appear to focus more on English-language foreign operations, whereas the IRA also includes a significant domestic,

Russian-language component. In addition to the English and Russian-language tweets, the REA and RNA datasets also comprise a significant amount of French tweets and retweets. Considering the large number of Africa-related hashtags and keywords within these tweets, it stands to reason that these accounts/tweets are aimed at Francophone countries in Africa.

URL Sharing

Many of the links that are shared within the dataset are now defunct and are untraceable even on archived webpages. However, based on the text in the URLs, a significant portion of them appear to direct to news articles on alternative news websites, with clickbait-heavy, emotionally evocative titles. Other categories include links to YouTube videos (which have been taken down). One piece of salient content that is still retrievable is a Telegram-channel pushed by GRU-accounts. This 'Brussels Snitch' channel promotes content that is critical of the European Union and NATO (table 5).

Table 5. GRU URLs Shared

	GRU URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['http://t.me/brussinf']	146	brussels snitch' Alexander Simonov ▪war correspondent; ▪specialist in the military-political activity of NATO and Western countries.
	['https://www.youtube.com/watch?v=2GuneHJyPSA']	145	Unavailable
	['http://en.farsnews.com/newstext.aspx?nn=13960224000929']	31	Unavailable
	['https://www.youtube.com/watch?v=1S3yIUoU10w']	28	Unavailable
Tweets	['http://t.me/brussinf']	146	See Above
	['https://www.youtube.com/watch?v=2GuneHJyPSA']	145	Unavailable
	['http://en.farsnews.com/newstext.aspx?nn=13960224000929']	31	Unavailable
	['https://www.youtube.com/watch?v=1S3yIUoU10w']	28	Unavailable
Retweets	['http://t.me/brussinf']	146	See Above
	['https://www.youtube.com/watch?v=2GuneHJyPSA']	145	Unavailable
	['http://en.farsnews.com/newstext.aspx?nn=13960224000929']	31	Unavailable
	['https://www.youtube.com/watch?v=1S3yIUoU10w']	28	Unavailable

The RNA dataset is more recent and much of the linked content is still online. These links (table 6) refer to other tweets that are supportive of the military dictatorship that took power in Sudan in 2019 and as such, seems to be part of an operation that aims to promote this regime.

One tweet, in Arabic, that was shared 15 times laments the ongoing strife, the lies in the discourse, and raises the question who might be benefit form this discord, which may allude to foreign interference (from Western powers). Another interesting detail are the shared links to bitcoin websites, which could have been part of scam operations run by the same actors who conduct the information operations.

Table 6. RNA URLs Shared

	RNA URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['https://twitter.com/x_m6y/status/1193182788297646081']	15	Tweet in Arabic: "It saddens me to see my countrymen wrestling and telling lies to each other" Who is the beneficiary?
	['http://okex.com/p2p/btc-to-inr']	9	Bitcoin Website
	['https://worldnews-sy.com/%D9%85%D9%86-%D9%8A%D9%82%D9%81-%D8%B9%D8%A7%	3	Syrian News Website in Arabic
	['https://twitter.com/x_m6y/status/1193182788297646081?s=19']	2	See the Tweet above
Tweets	['https://twitter.com/x_m6y/status/1193182788297646081']	15	See Tweet above
	['http://okex.com/p2p/btc-to-inr']	9	Bitcoin Website
	['https://worldnews-sy.com/%D9%85%D9%86-%D9%8A%D9%82%D9%81-%D8%B9%D8%A7%	3	Syrian News Website in Arabic
	['https://r8qs.adj.st/?adjust_t=ynoul8_q0le9d&adjust_deeplink=mobilelegends://appinvites']	2	Link to Appstore
Retweets	['https://twitter.com/steve_hanke/status/1152581365692428288']	2	Tweet: "Since #Sudan's ruling military council and leaders of the pro-Democracy movement signed a power-sharing agreement on Wednesday, Sudan's #inflation, by my measure, has taken a dramatic fall from 61% to 36% in just 3 days."
	['http://EcoFi.io']	2	Bitcoin Website
	['https://twitter.com/Sudaneseblogs/status/1149015143965237248']	2	Tweet: "Thareed: A document circulating on social networks as a copy of the draft political agreement between the Military Council and the Forces of Freedom and Change, and we note that it may not be the final document, especially for the amendments that appear in it."
	['https://twitter.com/samirasawlani/status/1155559234207191040']	2	Tweet: "Received video of one of the many Protests taking place in Sudan today as people demand civilian rule and demonstrate against Transitional Military Council #SudanUprising"

Word Frequency Analysis and TF-IDF

All subsets share one characterises between them: a strong focus on the Syrian Civil War. Hashtags and tweet texts from the GRU frequently mention and criticise American involvement in the war, attach themselves to the Syria and ISSI hashtags, as well as the Idlib hashtag, which was one of the main rebel strongholds during the war. The tweets call for the end of 'terrorism' in the country (figures 1 and table 7). With keywords such as Militants, Syria, Idlib and terrorists; and hashtags such as #Syria, #StopTerrorism, #Idlib, #WhiteHelmets, there appears to be an attempt to both attach to those trending topics, and

targeting the White Helmets, a volunteer organisation that is active in the conflict and tries to offer humanitarian and medical aid. This organisation has indeed been targeted by hostile disinformation campaign sin the past (Wilson & Starbird, 2020). Similarly, the RNA dataset's keywords and hashtags mainly contain references to the conflict in Syria, while the REA dataset is split mainly between the Syrian conflict and various African countries (see figure 2). The latter involve the Central African Republic and Sudan respectively, countries with regimes that were friendly to Russia. The TF-IDF analysis yielded similar results.

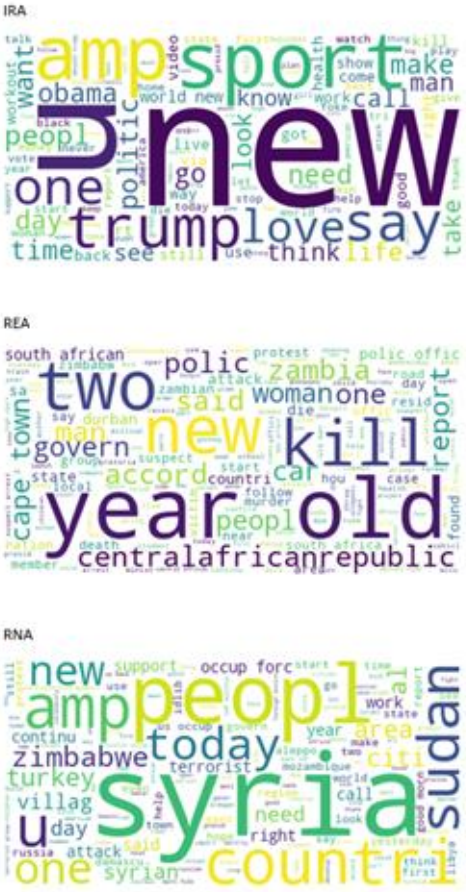
Figure 1. GRU WordCloud

Table 7. GRU Hashtags

	GRU Hashtags			
	Most Shared		Least Shared	
Grand Total	['Syria']	172	['Ahrar_al_Sham', 'Turkey', 'Hazano', 'Idlib']	1
	['Idlib']	105	['October', 'militants', 'Jabhat', 'attack', 'fighters', 'Ahrar', 'occupied', 'territories', 'lost', 'fighters', 'had', 'retreat']	1
	['StopTerrorismInSyria', 'StopHTSinSyria', 'WeAreTheLove']	98	['Nusra', 'Ahrar_al_Sham']	1
	['crime', 'syira']	76	['fire', 'government', 'Syria', 'Idlib']	1
Tweets	['StopTerrorismInSyria', 'StopHTSinSyria', 'WeAreTheLove']	96	['Explosions', 'died', 'coalition']	1
	['Syria']	92	['Americanbase', 'Tanf', 'Syrian']	1
	['Idlib']	85	['Kurdish', 'negotiating', 'Damascus']	1
	['crime', 'syira']	76	['International', 'American']	1
Retweets	['Syria']	80	['Explosions', 'died', 'coalition']	1
	['Idlib']	20	['Americanbase', 'Tanf', 'Syrian']	1
	['Refugees']	20	['Kurdish', 'negotiating', 'Damascus']	1
	['WhiteHelmets']	16	['International', 'American']	1

The IRA's activities (see figure 2) are more diffuse, with the main IRA body being split between American politics (the 2016 elections) and the war in Syria. The hashtags used by the main IRA dataset heavily favour using 'sports', 'news' and 'politics' tags, likely to gain traction with the (American) trending Twitter landscape, while many of the keywords are related to US political figures (Hillary Clinton, Barack Obama, Donald Trump).

Figure 2. IRA/REA/RNA WordClouds



One salient detail in the IRA dataset is the mentioning of German Chancellor Angela Merkel and the German far-right party Alternative für Deutschland (AfD) are the least shared hashtags (table 8), indicating that there was some interest in interfering in German discourse, most likely on the contentious immigration issue that was dominant in public discourse during the time-period covered by the dataset.[20] However, this does not appear to have been a priority for the English-language dataset.

[20] Saim Saeed, ‘’Merkel loses support, far right rises after migration row: poll,’’ Politico, 24 June 2018, https://www.politico.eu/article/angela-merkel-poll-migration-loses-support-far-right-rises-after-migration-row/.

Table 8. IRA/REA/RNA Hashtags

	IRA Hashtags			
	Most Shared		Least Shared	
Grand Total	[news]	152114	[Petry, AfD, CDU, Merkel]	1
	[sports]	97463	[CDU, AfD, SPD, ltwsaar2017, Saarland]	1
	[politics]	73937	[ShowtimeBoxing, Heavyweights]	1
	[СПб] (*St. Petersburg)	50716	['GE2015', 'PeopleNOTpolls']	1
Tweets	[news]	143789	[Госдума, екб] (*State Duma, ekb)	1
	[sports]	92953	[Готово] (*Ready)	1
	[politics]	70848	[Virginia, Cuban]	1
	[СПб] (*St. Petersburg)	37653	['FuerzaAxel']	1
Retweets	[СПб] (*St. Petersburg)	13063	[TravelBanNOW]	1
	[Россия, ОбразРоссии] (*Russia, Image of Russia)	9924	[елг2016, МоскваВыбирает, РоссияВыбирает] (*edg2016, MoscowChoose, RussiaChoose)	1
	[news]	8325	[NoDAPL, WaterIsLife, WaterIsSacred, LifeIsSacred, StandingRock]	1
		8228	['GE2015', 'PeopleNOTpolls']	1

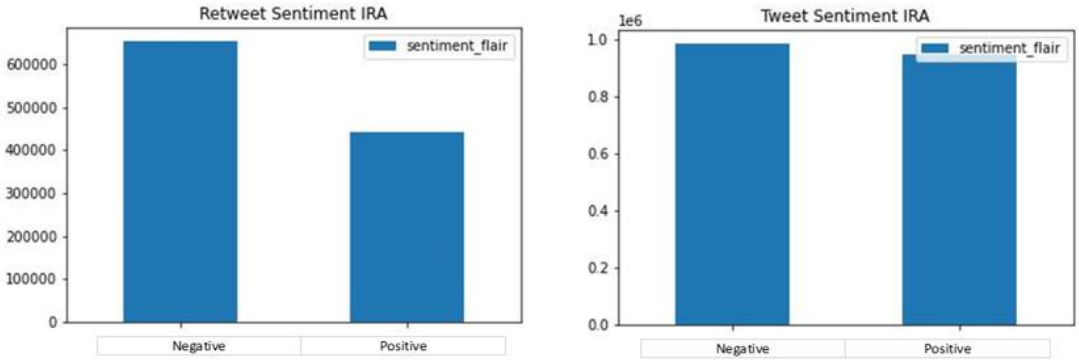
	REA Hashtags			
	Most Shared		Least Shared	
Grand Total	['CentralAfricanRepublic']	263	['Malawi', 'MalawiDecides2019']	1
	['Centrafrique']	231	['Belfast', 'Mpumalanga', 'Corruption']	1
	['Bangui', 'Centrafrique']	40	['NEWS', 'UJ']	1
	['CentralAfricanRepublic', 'COVID19']	26	['financingthefuture']	1
Tweets	['CentralAfricanRepublic']	263	['Cosatu', 'BankStrike']	1
	['Centrafrique']	231	['Khayelitsha']	1
	['Bangui', 'Centrafrique']	40	['Zimbabwe', 'Poachers']	1
	['CentralAfricanRepublic', 'COVID19']	26	['NelsonMandelaDay']	1
Retweets	['WorldCup']	1	['GPS']	1
	['Afrikaans']	1	['SocialMediaBlackout']	1
	['Zlatan']	1	['SocialMediaBlackOut', 'BlackOutWednesdays']	1
	['MandelaDay', 'TheNextChapter', 'ActionAgainstPoverty']	1	['DeepCoach']	1

	RNA Hashtags			
	Most Shared		Least Shared	
Grand Total	['Syria']	516	['Russian', 'ISIS', 'US', 'Euphrates']	1
	['Centrafrique']	198	['Russian', 'Syria', 'Russia']	1
	['Sudan']	139	['Iranian', 'Syrian', 'Damascus']	1
	['Damascus', 'Syria']	128	['AirIndia']	1
Tweets	['Syria']	508	['centrafrique']	1
	['Centrafrique']	140	['Mongoumba', 'Covid19']	1
	['Damascus', 'Syria']	127	['China', 'Sudan']	1
	['Damascus']	107	['healthcare', 'Mozambique']	1
Retweets	['Centrafrique']	58	['MiddleEasternbride', 'Lebanese', 'انفجار بيروت'] (*Explosion Beirut)	1
	['SudanUprising']	34	['Tours']	1
	['Sudan']	33	['covid_19', 'AFP']	1
	['ليبيا'] (*Libya)	30	['باشاغا', 'العثمان'] (*Bashagha', 'Othman (Minister of Health Libya)	1

Sentiment Analysis

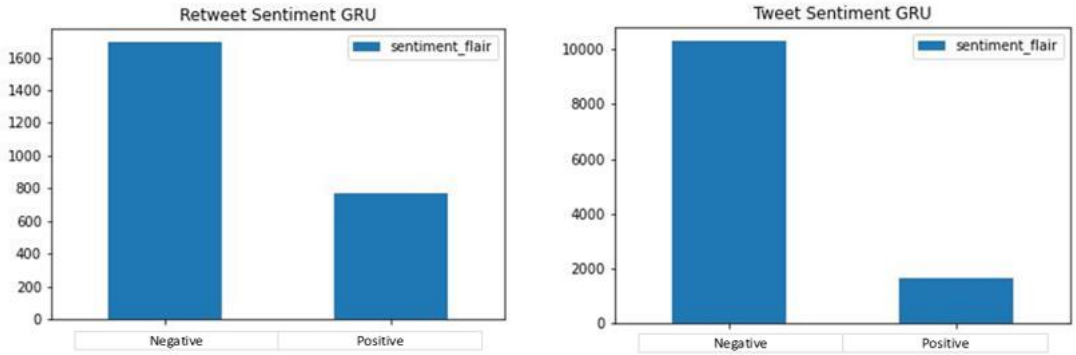
In terms of sentiment, the overall trend for the IRA is a (slight) predominance of negative content over positive content (figure 3). However, the differences are less than what one would expect from coordinated influence operations, since studies have found that that the online ecosystems favours negative messaging. One potential explanation for these unexpected results is the fact that sentiment analysis through machine learning is unsuited to picking up on sarcasm and other textual nuances, which may lead to false classifications.

Figure 3. IRA Sentiments



GRU seems to favour more negative content (figure 4), particularly in its African operations. The RNA and REA tweets and retweets are more negative than positive.

Figure 4. GRU Sentiments



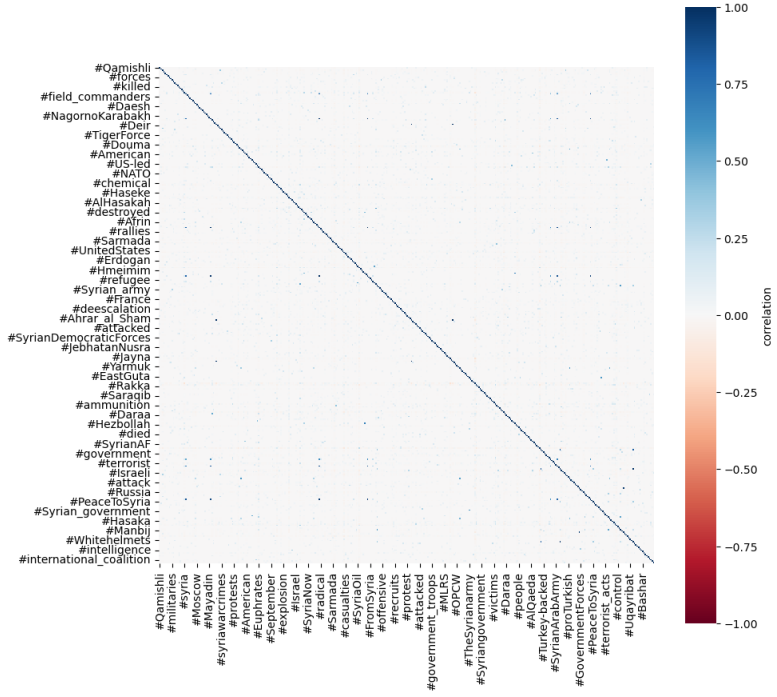
Latent Dirichlet Allocation (LDA) Topic Modelling

Similar to the keyword and hashtag results, the LDA Topic Modelling often yielded diffuse results, without clear connections between topic words, and where these words often were random and mundane, without clear connections to political or salient issues (see table 9 and 6). One strong connection that emerged for the GRU dataset is ‘US’, ‘weapon’, ‘white’, ‘coalition’, ‘kill’, ‘arm’, ‘train’ and ‘terrorists’ (6 and 7) in figure 5. This may indicate attempts to link US/coalition forces to training and arming groups that are involved in the conflict, to terrorists and their involvement in the conflict.

Table 9. GRU Topics

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	return	548.1	camp	865.8	provinc	649.0	al	2277.7	#syria	329.8	€‖	940.2	forc	678.9	â	628.0	syria	638.9	#syria	1103.5
1	#syria	431.8	refuge	467.9	#syria	592.3	milit	717.1	#sdf	321.1	peopl	448.9	zone	629.0	€□	319.1	#syria	467.9	#hts	886.0
2	eastern	371.7	humanitarian	386.3	russian	458.8	sham	696.1	#us	300.2	civilian	342.4	syrian	547.4	#isis	297.7	state	395.3	#terrorists	874.1
3	medic	320.1	#rukban	379.1	militari	435.9	group	644.9	terrorist	289.8	children	326.0	govern	413.3	milit	238.2	us	338.8	#idlib	824.0
4	peopl	293.1	rukban	297.1	syrian	421.4	nusra	526.6	#euphrates	261.9	terrorist	305.5	#idlib	409.4	al	227.8	use	287.2	posit	819.5
5	home	259.1	#syria	271.5	milit	374.8	tahrir	505.1	led	250.5	#syria	264.5	group	391.0	#syria	214.1	#us	286.9	#saa	790.8
6	refuge	241.3	provid	210.7	turkish	355.3	command	503.7	coalit	246.6	kill	230.5	arm	387.1	train	209.1	countri	255.0	#militants	629.1
7	#sdf	239.7	#us	204.1	weapon	342.4	#syria	378.3	#isis	223.3	syrian	210.6	#turkey	386.4	area	192.5	white	237.1	shell	500.5
8	yesterday	229.1	situat	200.6	provoc	314.3	#idlib	355.8	forc	213.2	local	202.1	#turkish	384.8	terrorist	181.8	chemic	226.7	fire	495.9
9	come	215.1	aid	195.1	prepar	314.0	field	315.3	control	204.8	us	184.4	idlib	368.4	ammunit	162.8	helmet	221.1	#hama	461.1

Figure 6. GRU Hashtag Heatmap



For the IRA datasets, there is more data available, as well as some clearer narratives that emerge. In 2014 and early 2015, the accounts mainly mention everyday topics such as sports and breaking news. In January 2015, the dominant narratives refer to the nuclear disaster that took place in Fukushima, Japan in 2011, which in the tweets is linked to the famous nuclear disaster in Chernobyl, Ukraine in 1986, in what was then the Soviet Union. The fact that these two past disasters are part of the Twitter discourse may be due to the release of the documentary *The Russian Woodpecker*, which was released in that same month. This documentary revolves around the Chernobyl disaster.

After mainly discussing sports, the tweets become more political again in the summer of 2015, referencing the riots in Ferguson, Missouri, US following the killing of Michael Brown a year earlier, through #fergusonremembers. Other tweets support Trump in the early stages of the campaign for the US Presidential elections in 2016, with mentions of #trumpbecause. There are no clear narrative connections visible yet, since most of the closely associated words are related to sports and other news items.

From the summer of 2015 onward, the tweets are becoming increasingly political, referencing devise social issues via #blacklivesmatter and #blackisnotacrime and 'Kaepernick', which correlate with sports topics in 2016 (table 10). These reference the kneeling during the singing of the national anthem at US football matches by American football player Collin Kaepernick, to protest racial injustice in the US. Narratives emerge that associate Hillary Clinton with untrustworthiness via #thingsmoretrustedthanhillary, Hillary, #mustbebanned. Other strongly related words are #celebratetrumpwith, #trumpforpresident and maga [Make America Great Again, Donal Trump's campaign slogan]. These continue up to the US presidential elections in November 2016.

Starting in January 2017, with the inauguration of Donal Trump as President of the United States, tweets reference this event. Other strongly connected keywords are 'maga', 'trump' and 'illegal immigrant', likely referencing Trump's strong ant-immigration stance. Support for his presidency is visible in the related hastags: #maga, #makeamericagreatagain, #americafirst. One strong narrative is established through the terms 'Clinton', 'cut', 'kill' and 'sexual', likely referencing the popular narrative that Hillary Clinton and other prominent Democrats were running a child-trafficking and abuse ring.[21]

In 2018, tweets strongly mention 'Mueller', referencing Robert Mueller's investigation of Russian in the 2016 Presidential Elections. In 2019, 'Venezuela' and 'Maduro' feature prominently in the dataset. In 2020, 'pandemic', 'China', 'coronaviru[s]' are unsurprisingly dominant.

[21] Gregor Aisch, Jon Huang and Cecilia Kang, ''Dissecting the #PizzaGate Conspiracy Theories'', *The New York Times*, 10 December 2016, https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html.

Table 10. IRA Topics 2015 & October 2016

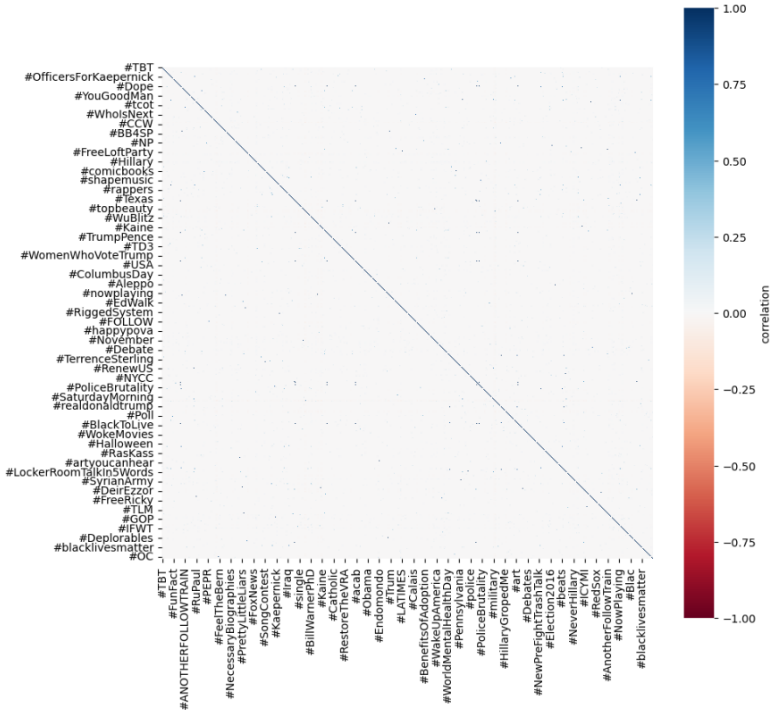
2015

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	–	4689.1	#sports	8285.1	know	3149.5	go	3057.6	foke	3577.7	#news	27641.6	love	7945.2	#politics	4362.5	like	5045.8	–	2477.5
1	one	3842.6	win	1973.5	't	2497.1	#phosphorusdisaster	2681.1	chang	2415.0	#local	5132.1	someon	2827.6	foke	3710.1	feel	2527.5	man	2142.7
2	life	3730.3	game	1459.0	live	1761.2	fail	2675.7	thing	2234.2	#chicago	3673.0	like	2610.8	#foke	2623.1	never	2361.0	say	1227.4
3	want	2652.5	play	1272.3	dream	1620.3	look	2525.2	new	2156.6	polic	2747.1	alway	1813.0	obama	2354.0	make	2246.1	men	1134.0
4	think	2053.8	top	1209.5	don	1254.1	let	2334.1	girl	1965.7	man	2673.4	«	1640.1	#showbiz	1353.4	i'	1427.0	good	1079.4
5	love	1932.0	beat	787.1	well	995.5	make	2202.7	time	1819.1	#loca	2595.1	happi	1620.3	isl	1059.9	peopl	1323.8	would	991.7
6	way	1519.9	team	778.4	#quote	933.1	american	1661.4	#foke	1715.2	'	2486.0	tri	1593.6	ever	961.9	friend	1273.2	reall	979.4
7	got	1309.2	#baseba	651.1	cute	889.8	thing	1654.2	best	1629.6	#breaking	2294.8	peopl	1579.4	...	909.7	good	1244.4	u	913.2
8	live	1265.2	final	637.4	fear	863.6	water	1622.1	year	1507.4	new	2242.4	»	1568.1	#money	874.8	video	1160.1	'	881.6
9	world	1216.5	watch	611.5	amaz	821.6	take	1610.1	cat	1257.1	#business	2048.2	never	1429.4	show	801.0	think	1119.2	money	879.3

October 2016

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	new	2642.8	black	2487.1	...	5270.2	#ruinadinnerinonephrase	1678.5	#news	3111.7	#betteralternativetodebates	2029.3	clinton	3616.4	trump	5496.3	...	3512.7	man	2642.1
1	get	2209.9	amp	1433.4	peopl	2197.9	gt	1628.2	'	2105.5	#rejecteddebatetopics	1394.8	hillari	3210.2	...	2417.8	like	2643.8	polic	2448.2
2	game	895.6	white	1092.3	get	1822.2	play	1438.6	#world	1851.6	#makemehateyouinonephrase	1145.5	...	1685.9	donald	1362.4	look	1467.7	kill	1814.8
3	video	764.8	#nowplaying	1037.1	one	1619.9	hurrican	1047.1	foke	1453.1	listen	1029.4	trump	1567.9	say	1176.3	year	1077.5	year	920.5
4	show	650.0	men	666.6	rt	1617.9	thank	846.6	#foke	1453.1	watch	1011.5	state	1527.0	support	955.3	amp	769.5	offic	916.0
5	amp	607.6	#blacklivesmatter	647.1	amp	1531.5	matthew	835.1	u	1102.3	ft	753.8	email	1126.0	'	909.8	life	739.5	shoot	876.1
6	...	581.6	via	578.8	day	1473.5	go	636.3	follow	834.1	#imtoooldfor	704.2	vote	1000.7	vote	663.9	right	702.3	woman	866.3
7	time	556.6	beat	558.9	think	1282.7	let	596.6	twitter	826.1	album	666.1	elect	902.0	presid	624.0	still	681.9	shot	814.1
8	today	536.8	live	526.5	need	1248.6	world	583.2	say	678.2	track	656.3	campaign	869.4	never	588.0	good	627.6	arrest	807.1
9	via	515.8	women	471.9	make	1137.5	#sports	543.9	readi	650.4	#mybedroominwords	566.1	obama	748.0	hillari	559.0	thing	622.7	charg	806.2

Figure 7. IRA Hashtag Heatmap October 2016



The REA and RNA datasets have fewer strongly connected keywords and narratives. The former mainly include a wide variety of largely unconnected keywords referencing news stories, mainly on crime topics, in various African countries. The latter mainly includes various keywords concerning the Syrian Civil War.

Discussion and Conclusion

The datasets that have been analysed in this chapter were unequal in size and scope. The IRA dataset is much larger and richer than the other three datasets, but this enormous size also means that a lot of 'noise' is included. Many tweets discuss a wide variety of random news items, sports events, or other topics without clear political or societal relevance. This in itself is a relevant finding, especially since this is the case in 2014 and early 2015. This 'mundane normalcy' helps the accounts to avoid attention, risk of banning due to inappropriate content, and to create a believable history on the platform.

These findings are generally in line with other studies on coordinated Russian interference on social media, particularly Twitter. One study observed that supposed IRA-accounts made up a wide, diffuse spectrum of left- and right-leaning content (Golovchenko et al., 2020), likely to be ready to adapt to either a Trump or Clinton victory in the 2016 US elections. Moreover, the impact of these accounts and their discourse may have very limited, though still harmful through their mere presence, and the associated doubt cast on the legitimacy of electoral discourse and outcomes (Eady et al, 2023).

The latter is particularly salient for the attacks on Hillary Clinton, which include likely references to conspiracy theories aimed at the Democratic Party and her, with the supposed objective of demonising and delegitimising them. Other prominent issues in American public discourse and society are also used in messaging, likely to amplify the salience of these issues, and also to gain traction through hijacking prominent and trending topics on Twitter. The GRU, RNA and REA datasets are more limited in geographical and topical scope. The former two are clearly aimed at influencing discourse on the Syrian Civil War. Some of the narratives seem to be aimed at discrediting the White Helmets, to whom the Syrian president and Russian ally Bashar al-Assad was not well-disposed. There are also references to the anti-Assad forces, the Americans and other allies, and their supposed arming and training of terrorists in Syria. The RNA dataset seems to be aimed at supporting Russian allies Sudan and the Central African Republic, but the tweets and discourse appear to be the most diffuse of the four datasets, with only sporadic mention of relevant political events and topics.

To establish whether the topics and narratives mentioned here actually aim to create a dominant narrative on certain issues, further, in-depth analysis of smaller subsets of the data is needed. This can take the form of analysis of the visual media included in the tweets, and close reading of the tweet texts. This is beyond the scope of this current project, but it would be a worthwhile follow-on project.

When we apply the three logics that are part of our analytical framework, we can posit the following. Propaganda Logic: Due to the focus on English-language content in this analysis, some of the nuances that may better expose the use of national and cultural characteristics by these Russian operations are lost. The universal reach of the English language can obfuscate the content that is intended for native-English-speaking countries. The IRA's frequent mention of sports, both in text and in hashtags, may indicate an attempt at gaining traction in the Twitter algorithms by attaching accounts and their messages to trending sports topics. The other accounts are more focused on specific political events and issues, likely in order to gain momentum to influence discourse on those topics. These are tactics shaped by the Logic of the Medium. Hashtags and trending topics that will likely be browsed by those parties interested in them, are heavily targeted by these messages, which may indicate a desire to influence the general global media narrative on these topics. Journalists that followed the 'Idlib' hashtag to abreast of the events there would have to wade through the many messages from these operations, in order to find what they were actually searching for. The general audience interested in these events might even take these messages for the truth. The Strategic Logic here seems to be more to support friendly regimes, rather than to promote uniquely 'Russian' interests, or to improve the international perception of Russia. For the IRA dataset, the strategic logic seems to amplify pre-existing social issues, the discrediting of certain politicians, and to support Donald Trump.

In the established scholarship, the Russian information operations have shaped much of the research and writing on the topic of information operations. In that sense, their tactics, techniques and protocols have become a default model for such operations, and the IRA-dataset in particular follows many of the established norms for such operations.

Chapter IV

Case Study – Chinese Information Operations on Twitter

The Chinese dataset analysed in this paper comprises two subsets, namely the CNCC and the CNHU sets, where the former includes tweets from Changyu Culture, a company in the Xinjiang province of China[22], and the rest of the tweets are generally associated with pro-Chinese Communist Party entities. As we shall see in this chapter, the CNHU subset is heavily focused on the discourse surrounding the Xinjiang province and the internment camps there.

Table 11. China - Metadata

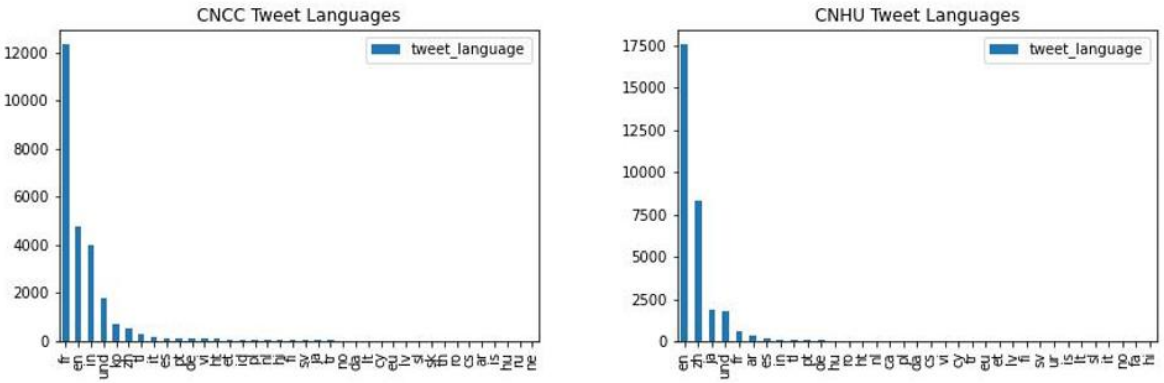
China	Agency	From	Till	Accounts	Tweets	Retweets	Total	English Language Tweets
	CNCC	24/01/2012	29/03/2021	112	30779	5145	35924	4777
	CNHU	20/04/2019	05/04/2021	2047	28567	2702	31269	17579
	Total	**24/01/2012**	**05/04/2021**	**2159**	**59346**	**7847**	**67193**	**22356**

Metadata Analysis

The pattern of account creation dates is similar to that of Russia: front-heavy and tapering off towards the actual main operations activity. Tweet and retweet ratios are similar for both datasets, where the tweets proper are overwhelmingly dominant over retweets. Again, this is similar to the results from the Russian case study. More significant differences emerge in the language distribution of the datasets (figure 8). The 'Xinjiang' dataset has a surprisingly strong presence of French-language tweets, as well as Indonesian tweets. The CNHU subset is predominantly English-language, with Mandarin following as a strong second language category. Duplicate tweets are scarce.

[22] ''Disclosing state-linked information operations we've removed'', Twitter, 2 December 2021, https://blog.twitter.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed.

Figure 8. China – Tweet Languages



URL Sharing

The CNCC subset promotes what appears to be an obscure French-language radio website, which matches the strong presence of French-language tweets (table 12). The reason for the prominence of this website is unclear, for its contents appear not to be relevant to political discourse. This may be an example of an information operation that uses popular media in order to gain traction within the Twitter ecosystem.

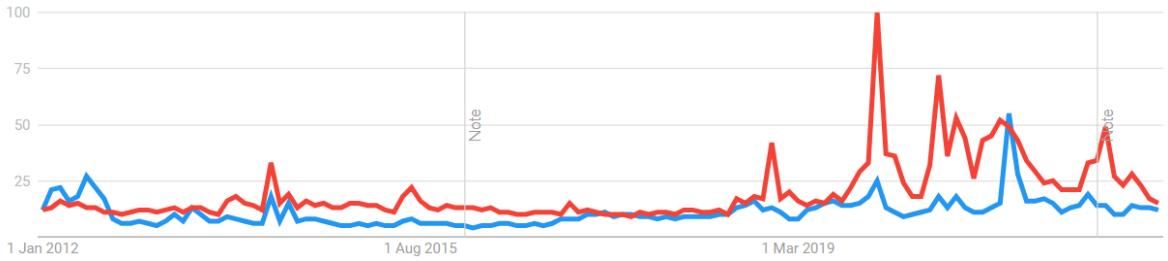
Table 12. URLs Shared

	CNCC URLS Shared		
	URL	Times Shared	Content
Grand Total	['http://mineboxradio.com/']	1159	French Radio Website
	['http://player.mineboxradio.com/']	600	French Radio Website
	['http://www.mineboxradio.com']	318	French Radio Website
	['http://mineboxradio.com']	312	French Radio Website
Tweets	['http://mineboxradio.com/']	1121	French Radio Website
	['http://player.mineboxradio.com/']	586	French Radio Website
	['http://www.mineboxradio.com']	315	French Radio Website
	['http://mineboxradio.com']	285	French Radio Website
Retweets	['http://mineboxradio.com/']	1121	French Radio Website
	['http://player.mineboxradio.com/']	586	French Radio Website
	['http://www.mineboxradio.com']	315	French Radio Website
	['http://mineboxradio.com']	285	French Radio Website

	CNHU URLS Shared		
	URL	Times Shared	Content
Grand Total	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583671.htm']	867	News Article: "Xinjiang's counter-terrorism measures protect human rights"
	['http://t.m.china.org.cn/convert/c_x8BDSHzU.html']	435	News Article: "Vocational education and training is key measure to protect human rights"
	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583673.htm']	281	News Article: "Short videos depict peaceful and happy lives in south Xinjiang"
	['http://t.m.china.org.cn/convert/c_XWWIAWDD.html']	216	News Article: "Real life stories from Xinjiang best refutation of US allegations"
Tweets	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583671.htm']	866	News Article: "Xinjiang's counter-terrorism measures protect human rights"
	['http://t.m.china.org.cn/convert/c_x8BDSHzU.html']	435	News Article: "Vocational education and training is key measure to protect human rights"
	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583673.htm']	273	News Article: "Short videos depict peaceful and happy lives in south Xinjiang"
	['http://t.m.china.org.cn/convert/c_XWWIAWDD.html']	216	News Article: "Real life stories from Xinjiang best refutation of US allegations"
Retweets	['https://www.globaltimes.cn/content/1187107.shtml']	33	News Article: "Xinjiang debunks lies in PBS documentary and biased Western media"
	['https://news.cgtn.com/news/2021-01-07/Xinjiang-s-growth-on-par-with-rest-of-China-new-report-'	29	News Article: "Xinjiang University report proves region's growth on par with rest of China"
	['https://www.pscp.tv/XHNews/1vOGwEAWerqxB?t=57m15s']	25	China Xinhua News Video of Xinjiang Regional Government
	['https://www.pscp.tv/w/cyxYTzFBbWp6Z1Zrb2R6RWV8MUJSSmpCcmFMUExKdxYRJp9C6EesP'	24	Xinjiang is a wonderful land" online promotion event

The CNHU subset shares many news articles in English that promote the Chinese province of Xinjiang as a pleasant place to live, whilst simultaneously calling out Western media that misrepresent the situation there (table 12). This is clearly a response to the international denouncement of the internment of the region's native Uyghur people in 're-education camps'. This issue rose to prominence in international media and reached peak internet search popularity in the first half of 2019 (figure 9).

Figure 9. Google Trends



Google Trends data for the terms 'Uyghur' (red) and 'Xinjiang' (blue). Source: Google Trends, 31 August 2022.

The CNHU information operation began tweeting about the Xinjiang issue in the second half of 2019, and continued promoting positive news about the region in 2020 and 2021, when the accounts were removed by Twitter.

Table 10. Hashtags

	CNCC Hashtags			
	Most Shared		Least Shared	
Grand Total	['MorningBox']	610	['MorningBox', 'serveur', 'illimité', 'OMGSERV']	1
	['MBR']	534	['MBR', 'spoil']	1
	['蓬佩奥', 'StopXinjiangRumors']	438	['JoyeuxNoel']	1
	['LibreAntenneMBR']	258	['Geekopolis', 'Paris', 'MBR']	1
Tweets	['MorningBox']	564	['PlaylistMBR']	1
	['MBR']	507	['MorniingBox']	1
	['蓬佩奥', 'StopXinjiangRumors']	438	['MAILMBR']	1
	['LibreAntenneMBR']	242	['MorningBox', 'MBR', 'Cadeaux', 'RVF', 'OMGSERV']	1
Retweets	['MorningBox']	564	['PlaylistMBR']	1
	['MBR']	507	['MorniingBox']	1
	['蓬佩奥', 'StopXinjiangRumors']	438	['MAILMBR']	1
	['LibreAntenneMBR']	242	['MorningBox', 'MBR', 'Cadeaux', 'RVF', 'OMGSERV']	1

	CNHU Hashtags			
	Most Shared		Least Shared	
Grand Total	['Xinjiang']	3155	['RebiyaKadir']	1
	['xinjiang']	591	['Bitcoin', 'TechnicalAnalysis', 'Forecast', 'PriceAction']	1
	['XinjiangOnline']	497	['BabyBird', 'birds']	1
	['Xinjiang', 'xinjiang']	312	['TLP', 'tlpwelfareactivities', 'tlp_believeislam']	1
Tweets	['Xinjiang']	2711	['穆斯林国家'] (*'Muslim country)	1
	['xinjiang']	591	['新疆', '强迫劳动', 'Uyghur'] (*Xinjiang', 'forced labor)	1
	['XinjiangOnline']	496	['新疆', '维吾尔', 'Xinjiang'] (*Xinjiang', 'Uyghur)	1
	['Xinjiang', 'xinjiang']	312	['新疆', 'America'] (*Xinjiang)	1
Retweets	['Xinjiang']	444	['疫情'] (*'epidemic)	1
	['Pompeo']	85	['coronavirus', 'COVID19']	1
	['XinjiangOnline', 'Xinjiang', 'StopXinjiangRumors', 'Xinjiang', 'Urumqi']	58	['American', 'Trump']	1
	['ZhaoLijian', 'humanrights']	47	['XinjiangOnline']	1

Word Frequency Analysis and TF-IDF

Within the CNCC subset, both in regular term frequency and TF-IDF counts, tweets mentioning ‘Seo Ye-ji’ and ‘Save Me’ are overwhelmingly dominant (figure 10). Both are references to popular culture. Seo Ye-ji is a South-Korean actress and *Save Me* is a British television series. Spread among these are some references to Xinjiang, such as the hashtag '蓬佩奥' ('StopXinjiangRumors').

Figure 10. CNCC Word Frequency (top) and TF-IDF (bottom).

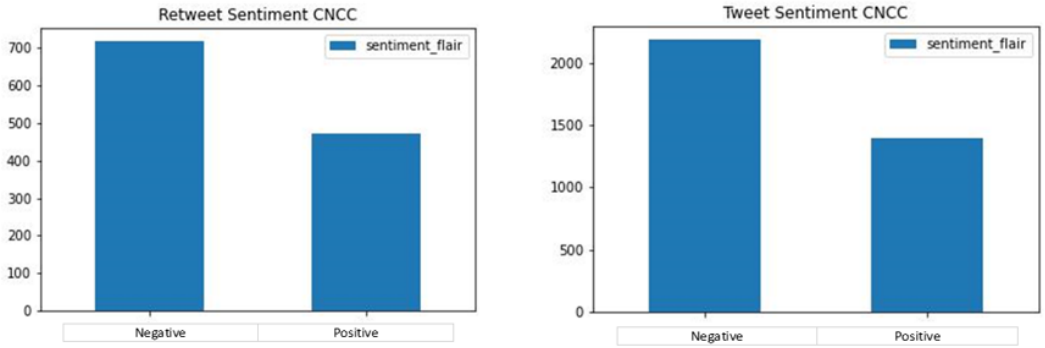
The results for the CNHU subset are almost entirely political and revolve around ‘Xinjiang’ (figure 11). Mentions of ‘terrorism’ and ‘counterterrorism’ are echo the Chinese government’s stance that the actions taken against the Uyghurs are part of counterterrorist measures in the province (Roberts, 2020).

Figure 11. CNHU Word Frequency (top) and TF-IDF (bottom).

Sentiment Analysis

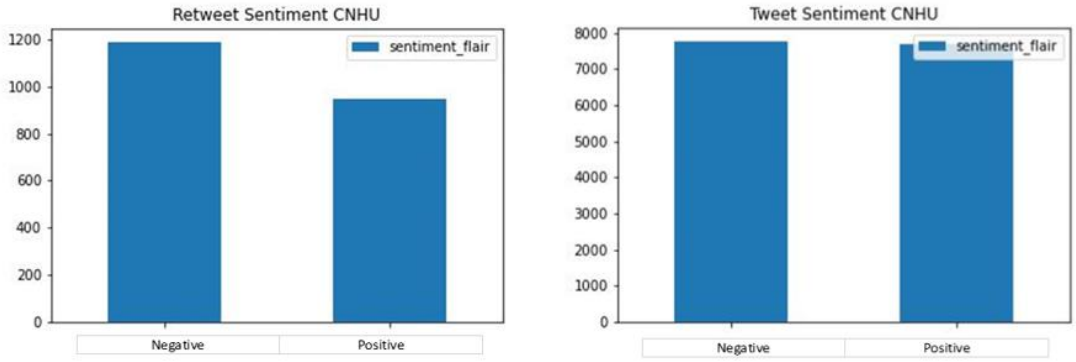
Sentiment analyses of both datasets show another (slight) difference between both subsets, where the CNHU (figure 12) subset has near parity between positive and negative tweets.

Figure 12. CNCC Sentiment



On the other hand, the CNCC surprisingly trends towards positive (re)tweets over negative (re)tweets (figure 13). This can likely be attributed to the strong pop-culture presence within the dataset.

Figure 13. CNHU Sentiment



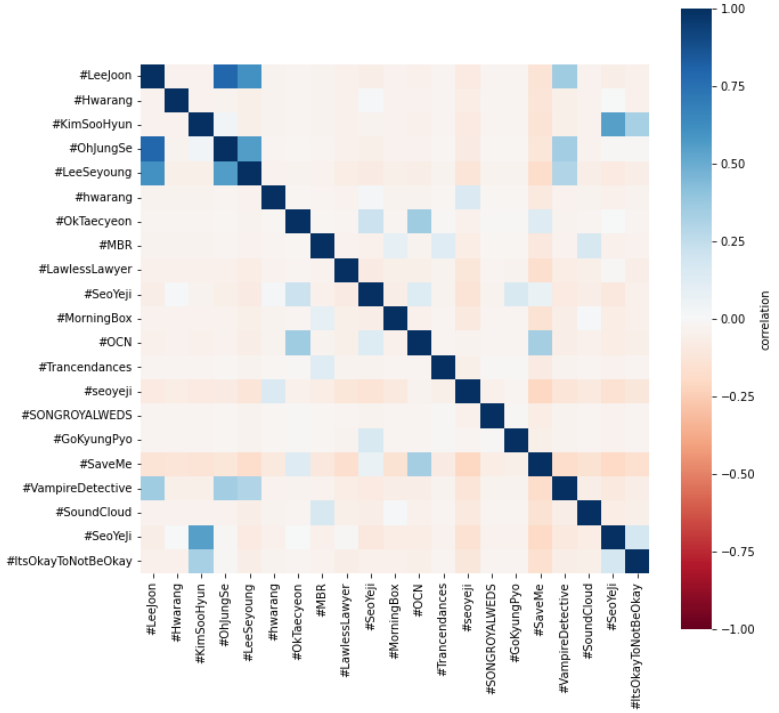
Latent Dirichlet Allocation (LDA) Topic Modelling

The topic modelling for the CNCC dataset mirrored the keyword and hashtag analyses, with the strongest topics being the actress Seo Ye-ji and the television show *Save Me* (table 11, figure 14).

Table 11. CNCC Topics

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	thankyou	94.7	ye	180.1	#seoyeji	259.0	oppa	181.0	...	89.2	#saveme	205.2	😂	184.1	unni	251.7	rt	200.1	look	125.9
1	morn	83.1	drama	148.8	updat	143.1	happi	139.1	get	88.1	#서예지	85.8	watch	158.1	go	130.1	love	155.6	new	72.1
2	mood	76.1	seo	121.3	ig	129.1	want	104.1	yeji	67.1	vampir	84.1	ㅋㅋ	127.0	know	117.4	pleas	110.1	princess	59.1
3	ah	68.6	ji	105.1	good	103.2	birthday	90.1	make	60.1	episod	82.1	like	96.4	let	89.5	think	68.1	le	47.1
4	one	62.1	new	62.1	even	56.2	follow	85.5	hope	58.1	#구해줘	81.1	ㅋㅋㅋ	88.1	pretti	67.1	wait	55.9	sound	44.1
5	#무법변호사	57.1	kim	57.1	sorri	55.1	hello	73.1	smile	49.1	detect	79.1	drama	58.4	gt	64.1	meet	50.1	#hwarang	44.1
6	realli	55.2	beauti	51.1	hard	54.1	much	58.1	peopl	48.1	ep	70.1	join	53.6	great	54.1	like	45.7	face	41.5
7	good	51.0	come	50.8	guy	42.1	thankyou	55.2	take	44.2	still	68.2	help	48.1	girl	53.1	nice	44.4	miss	38.1
8	feel	50.1	film	45.1	see	41.5	wish	54.1	well	43.1	scene	67.9	friend	47.2	mineboxradio	30.1	support	39.1	photo	38.1
9	#lawlesslawyer	48.1	day	44.7	night	39.1	best	42.1	amp	41.1	#leeseyoung	57.1	forget	39.1	alway	29.3	😡	34.1	#soundcloud	36.1

Figure 14. CNCC Hashtag Heatmap

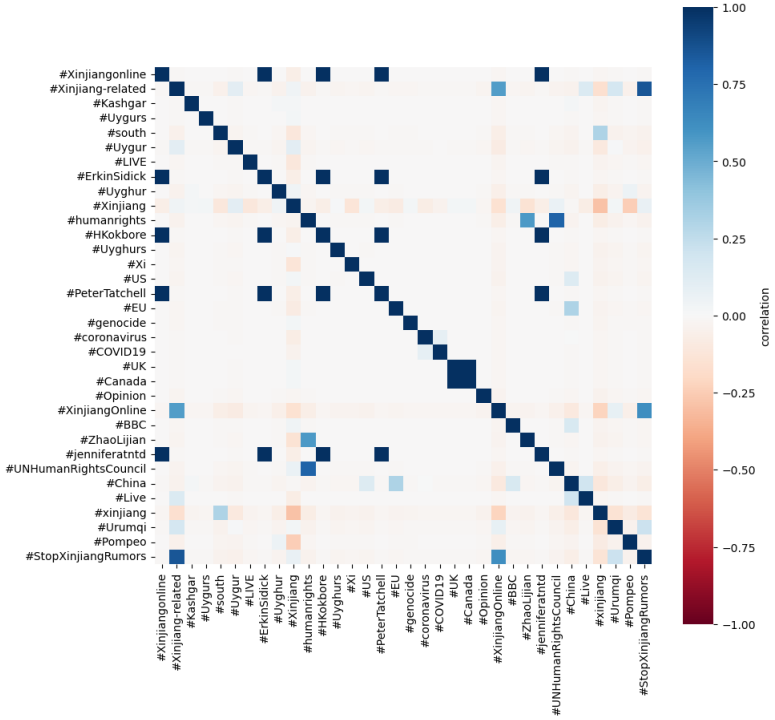


The topics and associated narratives that emerge from the CNHU dataset mainly focus on representing the situation in Xinjiang as positive (see table 12, figure 15), with improvement of human rights for the local population (‘counter’, ‘terror’, ‘protect’, ‘human’, ‘right’, ‘better’, ‘xinjiang’), which we have already seen in the links shared within the tweets.

Table 12. CNHU Topics

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#xinjiang	1785.4	one	336.1	relat	992.8	xinjiang	889.1	live	771.8	xinjiang	717.1	region	488.5	#xinjiang	1442.4	xinjiang	354.1	#xinjiang	939.3
1	right	1640.9	better	314.5	xinjiang	973.7	western	517.8	#xinjiang	700.9	report	435.2	#xinjiang	429.3	issu	725.2	stabil	170.2	lie	516.9
2	human	1588.1	...	202.0	issu	927.2	#xinjiang	365.2	xinjiang	609.6	china	364.3	forc	398.9	relat	625.9	polit	168.0	'	345.0
3	protect	1574.1	man	193.0	foreign	470.3	us	339.3	happi	538.0	say	239.7	xinjiang	370.2	china	550.2	top	153.0	good	287.1
4	measur	1532.0	love	185.1	press	420.1	stori	334.6	peac	521.8	offici	231.1	autonom	355.8	xinjiang	547.6	stress	152.1	...	279.1
5	terror	1182.2	like	167.1	confer	396.1	peopl	332.9	video	485.1	right	178.3	uygur	308.8	wonder	321.0	endur	151.1	pompeo	271.9
6	counter	1172.7	work	160.3	#xinjiang	354.0	real	316.5	short	376.1	refut	167.2	china	303.2	amp	306.4	advisor	150.1	state	267.1
7	xinjiang	1136.6	go	158.0	ministri	308.2	muslim	299.0	south	267.1	im	152.1	never	293.1	#新疆	289.1	two	135.1	china	251.0
8	'	1038.6	make	142.1	china	305.9	call	278.9	depict	254.1	...	151.8	labor	253.5	##xinjiangonline	285.9	know	126.9	fact	248.5
9	covid	742.3	friend	141.0	chines	305.8	life	248.1	futur	181.1	extremist	137.0	us	237.5	##stopxinjiangrumors	270.1	fight	111.7	us	226.5

Figure 15. CNHU Hashtag Heatmap



Discussion and Conclusion

Overall, the 'Chinese' datasets are less rich than those ascribed to Russian information operations. One potential explanation is that it seems the Chinese operations appear to have abandoned the platform at the time when these tweets were posted, in favour of other platforms and influence strategies (Bolsover & Howard, 2019). In general, the use of regular media, diplomatic channels, Chinese cultural institutes, the Chinese diaspora and other social media platforms seem to have become the preferred means to exert influence.

Now let us turn to how these findings correspond with the three hypotheses on the logic that is supposed to guide an information operation. Firstly, the Propaganda Logic. The CNCC dataset appears to attach itself strongly to popular cultural trends in order to embed itself into the Twitter discourse, whereas the CNHU appears to abandon such strategies in favour addressing the topics it tries to influence on directly. This operational directness is also evident in the short time span in which the accounts were created and 'put to work', with a mere few months between their creation and the start of their post activity. In contrast, the CNCC accounts had longer longevity.

These trends can be interpreted as manifestations of different Medium Logics, where one tries to go embed itself in the discourse, and the other tries to directly influence it. The CNCC subset seems to be more regionally focused too, with tweets in regional languages like Indonesian being strongly present within the dataset. The CNHU subset clearly comprises accounts and tweets that were created on short notice to combat mainly Western narratives on the internment camps in Xinjiang. The overwhelming amount of mentions of the province within the tweet texts and hashtags indicate an intend to shape the discourse on this topic on Twitter. One puzzling aspect of this dataset is the significant amount of tweets in Mandarin, since Twitter is blocked in the country. One explanation can be that these tweets are directed at the Chinese diaspora in Western countries. Indeed, this diaspora is assumed to be integrated in the overall strategy of the Chinese state to exert influence abroad (Charon & Vilmer, 2021). Chinese operations in general, appear to be more reliant on participatory propaganda from likeminded citizens themselves, instead of relying on coordinated campaigns conducted by professionals only (Hellman & Wagnsson).

A Stanford Internet Observatory Cyber Policy Study's report on the same dataset, published in 2021, identified similar importance of the Xinjiang- and Covid-issues for these campaigns, the attempt to redirect online discourse into a positive view of China. They also identified the sometimes mundane topics that many of the accounts spoke about, instead of clear political discourse (DiResta, Goldstein, Miller & Wang, 2021). This 'positive' trend is also observed in other studies on Chinese operations on other platforms, such as Facebook, at the time (Tambe & Friedman, 2022). More recent observations of Chinese campaigns, such as those targeting the 2024 Presidential elections in Taiwan, have noticed a more aggressive and negative trend (Köckritz, 2023; Infodemic, 2024). This matches there observed ongoing trend of a convergence in methods and tone between Russian and Chinese information operations.

Overall, those tweets that discuss political issues discuss mainly topics that are important to China's reputation; Uyghurs, Xinjiang and Covid. This seems to indicate that the

Strategic Logic here is to shape discourse on Twitter in a more positive light for the country. This is congruent with other studies (DiResta, Goldstein, Miller & Wang, 2021). The supposed information operations that have been analysed in this chapter seem to differ in the extent to which they follow traditional models for information operations, especially by favouring positive content over negative. However, as has been mentioned before, the more recent trends in Chinese information operations seem to turn towards more aggressive and negative tactics, similar tot those of Russia. As such, more recent operations might indicate a turn to 'one-size-fit- all' for information operations.

Conclusion

This study aimed to identify to what extent information operations on Twitter, supposedly conducted by Russian and Chines state-backed actors, follow similar logics. Three hypotheses, or logics, grounded in extant scholarship on propaganda and information operations, were assumed to guide such activities. A dataset, released by the social media platform Twitter itself, containing accounts and posts (tweets) that were suspected of being part of Russia and China-backed information operations, was analysed using quantitative text analysis methods. This analysis was expected to provide insights in the conduct and content of such operations, and as such support or falsify the hypotheses, as well as answer the overarching question how these operations may converge or diverge. This endeavour has been partially successful, as we shall discuss below, within the framework of the three logics.

1. Propaganda Logic: successful information operations attach to what is already present in their target.

Due to limitations inherent in the research design of the study, this part of the logic was somewhat obscured. This is partially due to the focus on English-language content only, and the universal reach of such content, which eradicates some of the potential cultural characteristics that operations in other languages could reveal. Another reason is the size of the datasets and the broadness of the textual analysis, which created a large amount of 'noise' within the results, obscuring potential narratives within the data.

Some insight has been obtained, however. One example of an information operation trying to embed itself in larger cultural trends or phenomena, was the Chinese CNCC operation's strong integration of popular artists and television series within its messaging. This can be a tactic to attempt to integrate the operation within culturally dominant discourse. Russia IRA's use of sports-related tags can also fit into this category. In the other operations, the direct context in which they operated, such as the Syrian Civil War, the Xinjiang internment camps, and political and social upheaval in Africa, were more decisive in shaping content.

This hypothesis has not been confirmed decisively, and would be well-served by more in-depth follow-on research focussing on information operations on a smaller scale. The Russian preference for negative content fits within the established scholarship, that has found that strong emotional and hateful content is more successful reaching audiences. This differs

from the more positive Chinese operations, that try to improve the global perception of the country, and to counter negative discourse aimed at China on Twitter.

2. Strategic Logic: successful information operations are adapted to their unique strategic objectives.

The major difference between information operations conducted by both states could be identified through reflections on who might 'benefit' from these operations. Most Russian activity appears to be supportive of allied or friendly regimes abroad, or could be adversely affect the internal stability of geopolitical adversaries such as the United States, through the amplification of social discord. The latter is visible in the narratives in the IRA-dataset, which focused on racial relations within the US, and propagated content that referred to conspiracy theories involving Democratic politicians. On the other hand, Chinese information operations are more concerned with issues that align with China's reputation abroad and its right to determine its own internal affairs. This hypothesis is confirmed, but not firmly. The exact goals and objectives of these supposed operations are difficult to discern without further information.

3. Medium Logic: successful information operations adapt to the information environment in which they operate.

All information operations used the logic of Twitter's hashtag system and algorithms to push content, share links and to increase their prominence on topics pertinent to their strategic objectives, confirming this hypothesis.

Overall, the results yielded by the analysis are not as clear and decisive as hoped, and there is still room for further analysis. Some reasons for this are the differences in terms of size and richness of the dataset. The China-data in particular, is rather limited and did not provide clear results within the methodology employed in this project. This is also due to the nature of the datasets, since the exact benchmarks and metrics used in their curation are unknown. This makes it difficult to state with confidence that these are indeed information operations, and that this is indeed their full scope. Indeed, many information operations are unsuccessful in gaining traction, and are removed from the platform before they become part of the discourse. As such, there are biases that are inherent to the dataset, and in order to identify these operations in full, a cross-platform analysis would be needed. This has not been done in this project due to limitations concerning resources and available data, but it would be a worthwhile topic for follow-on research. This is especially relevant since this data is now outdated, considering

changes in the preferred platforms that are employed in information operations, and changes in the platforms of Twitter/X itself.

This study also raises questions on how information operations are defined, identified and studied. The operational difficulty of identifying information operations on a social medium have already been mentioned, but there are other factors to consider as well. To what extent we can accurately identify the actors purveying such an operation; their tactical, strategic and long-term objectives; and their success in achieving those, through only reverse-engineering their operations is an ongoing debate. Ascribing specific intentions to an operation based on their contents alone, or by placing them within the strategic 'doctrine' of the supposed propagator is questionable, and can lead to conclusions and interpretations that are too deterministic. The functioning of the organisation that conducts the operation, their relationships with other actors operating in the same theatre, the cooperation or antagonism between foreign and domestic purveyors of disinformation operations all influence how such an operation might function.

Thus, the three logics that have been used in this dissertation do not necessarily align with each other at all times, since the actors and circumstances involved will not be the same. As to how that influences the answer to the overarching question of this dissertation: there is some similarity in the manner in which the information operations by Russia and China are conducted, particularly regarding the Propaganda and Medium Logics, but the exact Strategic Logic of their operations is likely the most decisive factor. It is also the most difficult one to confirm with certainty. This can, and is encouraged to be explored further in future research.

Bibliography

‘’Analysis of Twitter Takedowns Linked to Venezuela, Mexico, Tanzania, and China’’, Stanford Internet Observatory, 2 December 2021, https://cyber.fsi.stanford.edu/io/news/twitter-takedown-december-2021.

Allcott, Hunt, and Matthew Gentzkow. 2017. “Social Media and Fake News in the 2016 Election.” *Journal of Economic Perspectives* 31 (2): 211–36. https://doi.org/10.1257/jep.31.2.211.

Allen, Patrick. 2007. *Information Operations Planning*. New York: Artech House.

Al-Rawi, Ahmed. 2021. ‘’How Did Russian and Iranian Trolls’ Disinformation Toward Canadian Issues Diverge and Converge?.’’ *Digi War* 2: 21–34. https://doi.org/10.1057/s42984-020-00029-4.

Arendt, Hannah. 1966. *The Origins of Totalitarianism*. London: Penguin Books.

Bail Christopher A, Brian Guay, Emily Maloney, Aiden Combs, D Sunshine Hillygus, Friedlon Merhout, Deen Freelon, Alexander Volfovsky. 2020. ‘’Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017.’’ *Proceedings of the National Academy of Sciences of the United States of America* 117 (1): 243-250. https://doi.org/10.1073/pnas.1906420116.

Barnes, Jenae, ‘’Twitter Ends Its Free API: Here’s Who Will Be Affected’’, Forbes, 3 February 2023, https://www.forbes.com/sites/jenaebarnes/2023/02/03/twitter-ends-its-free-api-heres-who-will-be-affected/.

Beatson, Oliver, Rachel Gibson, Marta Cantijoch Cunill, and Mark Elliot. (2023). ‘’Automation on Twitter: Measuring the Effectiveness of Approaches to Bot Detection.’’ *Social Science Computer Review* 41 (1): 181-200. https://doi.org/10.1177/08944393211034991.

Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Bernays, Edward. 1923. *Crystallizing Public Opinion*. New York: Ig Publishing.

Bernays, Edward. 1928. *Propaganda*. New York: Ig Publishing.

Bittman, Ladislav. 1985. *The KGB and Soviet Disinformation: An Insider’s View*. Oxford: Pergamon-Brassey’s.

Bolsover, Gillian, and Philip N. Howard. 2019. ‘’Chinese Computational Propaganda: Automation, Algorithms and the Manipulation of Information about Chinese Politics on Twitter and Weibo.’’ *Information, Communication & Society* 22 (14): 2063-2080. https://doi.org/10.1080/1369118X.2018.1476576.

Buntain, Cody L. and Jennifer Golbeck. “Automatically Identifying Fake News in Popular Twitter Threads.” *2017 IEEE International Conference on Smart Cloud (SmartCloud) (2017)*: 208-215. https://doi.org/10.1109/SmartCloud.2017.40.

Carlson, Matt. 2018. ‘’ ’Fake News as an Informational Moral Panic: The Symbolic Deviancy of Social Media During the 2016 Us Presidential Election.’’ *Information, Communication & Society* 23 (3): 374-388. https://doi.org/10.1080/1369118X.2018.1505934.

Cellan-Jones, Rory. ‘’Facebook v Australia: Who Blinked First?’’ *BBC News*, 23 February 2021, https://www.bbc.com/news/technology-56168843.

Charon, Paul, and Jean-Baptiste Jeangène Vilmer. 2021. *Chinese Influence Operations: A Machiavellian Moment*. Paris: Ministry for the Armed Forces.

Chatterjee, Siddhartha, and Michal Krystyanczuk. 2017. *Python Social Media Analytics: Analyze and Visualize Data From Twitter, YouTube, GitHub, and More*. Birmingham: Packt Publishing.

Cheng, Dean. 2016. *Cyber Dragon: Inside China's Information Warfare and Cyber Operations*. Santa Barbare: Praeger.

Codevilla, Angelo. 1992. *Informing Statecraft: Intelligence for a New Century*. New York: The Free Press.

Codevilla, Angelo, and Paul Seabury. 2006. *War: Ends and Means*. Dulles: Potomac Books.

‘’Countvectorizer’ Object Has No Attribute ‘get_feature_names_out',’’ StackOverflow, https://stackoverflow.com/questions/70640923/countvectorizer-object-has-no-attribute-get-feature-names-out.

Cour, Christina la. 2020. “Theorising Digital Disinformation in International Relations.” *International Politics* 57 (4): 704–23. https://doi.org/10.1057/s41311-020-00215-x.

Darius, Philipp and Fabian Stephany. 2020. ‘’How the Far-Right Polarises Twitter: 'Highjacking' Hashtags in Times of COVID-19.’’ October 12, 2020. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3709988.

Dawson, Andrew and Martin Innes. 2019. ‘’How Russia's Internet Research Agency Built Its Disinformation Campaign.’’ *The Political Quarterly* 90 (2): 245-256. https://doi.org/10.1111/1467-923X.12690.

De Sio, Lorenzo, and Till Weber. (2020). ‘’Issue Yield, Campaign Communication, and Electoral Performance: A Six-Country Comparative Analysis.’’ *West European Politics* 43 (3): 720-745. https://doi.org/10.1080/01402382.2019.1655968.

DiResta, Renee, Josh A. Goldstein, Carly Miller, and Harvey Wang. (2021). ‘’One Topic, Two Networks: Evaluating Two Chinese Influence Operations on Twitter Related to Xinjiang.’’ Stanford Digital Repository. Available at https://purl.stanford.edu/sn407zm8237.

‘’Disclosing New Data to Our Archive of Information Operations’’, Twitter, 20 September 20219, https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019.

‘’Disclosing State-Linked Information Operations We’ve Removed’’, Twitter, 2 December 2021, https://blog.twitter.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed.

Dov Bachmann, Sascha, Andrew Dowse, and Hakan Gunneriusson. 2019. “Competition Short of War – How Russia’s Hybrid and Grey-Zone Warfare Are a Blueprint for China’s Global Power Ambitions.” *Australian Journal of Defence and Strategic Studies* 1 (1).

Eady, Gregory, Tom Paskhalis, Jan Zilinsky, Richard Bonneau, Jonathan Nagler and Joshua A. Tucker. 2023. ‘’Exposure to the Russian Internet Research Agency Foreign Influence Campaign on Twitter in the 2016 Us Election and Its Relationship to Attitudes and Voting Behavior.’’ *Nature Communications* 14: 1-11. https://doi.org/10.1038/s41467-022-35576-9.

Ellul, Jacques. 1964. *The Technological Society*. New York: Vintage Books.

Ellul, Jacques. 1965. *Propaganda: The Formation of Men’s Attitudes*. New York: Vintage Books.

Elswah, Monah, and Philip N Howard. 2020. ‘’“Anything that Causes Chaos”: The Organizational Behavior of Russia Today (RT).’’ *Journal of Communication* 70 (5): 623–645. https://doi.org/10.1093/joc/jqaa027.

Fangyi Lu, Isabel. 2022. ‘’To Subdue the Enemies Without Fighting: Chinese State-Sponsored Disinformation as Digital Warfare.’’ *Digital War* 3: 96 – 106. https://doi.org/10.1057/s42984-022-00052-7.

Fatih Karabiber, ‘’TF-IDF — Term Frequency-Inverse Document Frequency’’, Learn Datasci, https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/.

Fridman, Fridman. 2017. ‘’Hybrid Warfare or Gibridnaya Voyna?.’’ *The RUSI Journal* 162 (1): 42-49. https://doi.org/10.1080/03071847.2016.1253370.

Fridman, Ofer. 2019. ‘’On ‘Gerasimov Doctrine’: Why the West Fails to Beat Russia to the Punch.’’ *Prism* 8 (2): 101- 112.

Fuchs, Christian. 2018. “Propaganda 2.0: Herman and Chomsky’s Propaganda Model in the Age of the Internet, Big Data and Social Media.” In *The Propaganda Model Today: Filtering Perception and Awareness*, edited by J. Pedro-Carañana, D. Broudy, and J. Klaehn. London: University of Westminster Press.

Gadde, Vijaya, and Yoel Roth, “Enabling Further Research of Information Operations on Twitter,” Twitter, October 2018. https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.

Galeotti, Mark, ‘’I'm Sorry for Creating the ‘Gerasimov Doctrine’’’, *Foreign Policy*, 5 March 2018, https://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/.

Galeotti, Mark. 2019. ‘’The Mythical ‘Gerasimov Doctrine’ and the Language of Threat.’’ *Critical Studies on Security* 7 (2): 157-161. https://doi.org/10.1080/21624887.2018.1441623.

Gerrits, André W.M. 2018. “Disinformation in International Relations: How Important Is It?” *Security and Human Rights* 29 (1–4): 3–23. https://doi.org/10.1163/18750230-02901007.

Golovchenko, Yevgeniy, Cody Buntain, Gregory Eady, Megan A. Brown, and Joshua A. Tucker. 2020. ‘’Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube During the 2016 US Presidential Election.’’ *The International Journal of Press/Politics, 2020, Available at SSRN*: https://ssrn.com/abstract=3552886.

Grajewski, Nicole. 2022. “An Illusory Entente: The Myth of a Russia-China-Iran ‘Axis.’” *Asian Affairs* 53 (1): 164–83. https://doi.org/10.1080/03068374.2022.2029076.

Graphika and Stanford Internet Observatory. 2022. Unheard Voice: Evaluating Five Years of Pro-Western Covert Influence Operations. Stanford Digital Repository. Available at https://purl.stanford.edu/nj914nx9540.

Hammond-Errey, Miah. ‘’Elon Musk’s Twitter Is Becoming a Sewer of Disinformation’’, *Foreign Policy*, 15 July 2023, https://foreignpolicy.com/2023/07/15/elon-musk-twitter-blue-checks-verification-disinformation-propaganda-russia-china-trust-safety/.

Hellman, Maria, and Charlotte Wagnsson. 2017. ‘’How Can European States Respond to Russian Information Warfare? An Analytical Framework.’’ *European Security* 26 (2): 153-170. https://doi.org/10.1080/09662839.2017.1294162.

Herman, Edward, and Noam Chomsky. 1988. *Manufacturing Consent: The Political Economy of the Mass Media*.

Hilbert, Martin, Javier Vásquez, Daniel Halpern, Sebastián Valenzuela, and Eduardo Arriagada. 2017. “One Step, Two Step, Network Step? Complementary Perspectives on Communication Flows in Twittered Citizen Protests.” *Social Science Computer Review* 35 (4): 444–61. https://doi.org/10.1177/0894439316639561.

Hoogensen Gjørv, Gunhild, and Outi Jalonen. 2023. ‘’Identity as a tool for disinformation: Exploiting social divisions in modern societies.’’ Hybrid CoE Strategic Analysis 34, Hybrid Centre of Excellence, Helsinki.

Howard, Philip N., and Samuel Woolley. 2018. *Computational Propaganda. Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford: Oxford University Press.

Howard, Philip N. 2020. *Lie Machines. How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: Yale University Press.

Infodemic. 2024. *2023 Dec W4 – 2024 Taiwan Presidential Election Information Manipulation AI Observation Report*. Available at https://ailabs.tw/wp-content/uploads/2024/01/2023-Dec-W4-2024-Taiwan-Presidential-Election-Information-Manipulation-AI-Observation-Report.pdf.

James, ‘’Topic Modelling In Python Unsupervised Machine Learning To Find Tweet Topics’’, Coding Club, https://ourcodingclub.github.io/tutorials/topic-modelling-python/.

Jennings, Will, Gerry Stoker, Hannah Bunting, Viktor Orri Valgarðsson, Jennifer Gaskell, Daniel Devine, Lawrence McKay, and Melinda C. Mills. 2021. “Lack of Trust, Conspiracy Beliefs, and Social Media Use Predict COVID-19 Vaccine Hesitancy.” *Vaccines* 9 (6): 593. https://doi.org/10.3390/vaccines9060593.

Jungherr, Andreas, and Ralph Schroeder. 2021. “Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy.” *Social Media and Society* 7 (1). https://doi.org/10.1177/2056305121988928.

Karabiber, Fatih. n.d. “TF-IDF — Term Frequency-Inverse Document Frequency.” Accessed August 31, 2022. https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/.

Karlsen, Geir Hågen. 2019. “Divide and Rule: Ten Lessons about Russian Political Influence Activities in Europe.’’ *Palgrave Communications* 5 (19): 1-14. https://doi.org/10.1057/s41599-019-0227-8.

Köckritz, Angela. 2023. ‘’In a Savvy Disinformation Offensive, China Takes Aim at Taiwan Election.’’ *Mercator Institute for China Studies*. Available at https://merics.org/sites/default/files/2023-12/MERICS%20Report%20Disinformation%20Taiwan%20December%202023.pdf.

Lanoszka, Alexander. 2019. “Disinformation in International Politics.” *European Journal of International Security* 4 (2): 227–48. https://doi.org/10.1017/eis.2019.6.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1944. *The People’s Choice. How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lazarsfeld, Paul F., and Robert K. Merton. 1948. ‘’Mass Communication, Popular Taste, and Organized Social Action.’’ In *The Communication of Ideas*, edited by Lyman Bryson, New York: Harper.

Lefebvre, Vladimir. 1982. *Algebra of Conscience. A Comparative Analysis of Western and Soviet Ethical Systems*. Dordrecht: Springer.

Lewandowsky, Stephan, Ullrich K.H. Ecker, and John Cook. 2017. ‘’Beyond Misinformation: Understanding and Coping with the “Post-Truth” Era.’’ *Journal of Applied Research in Memory and Cognition* 6 (4): 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008.

Li, Yuan. 2022. “How China Embraces Russian Propaganda and Its Version of the War.” *The New York Times*, May 4, 2022. https://www.nytimes.com/2022/03/04/business/china-russia-ukraine-disinformation.html.

Lippmann, Walter. 1922. *Public Opinion*. New York: Harcourt, Brace & Co.

MacLeod, Alan. 2019. ‘’Fake News, Russian Bots and Putin’s Puppets.’’ In *Propaganda in the Information Age. Still Manufacturing Consent*, edited by Alan MacLeod. London: Routledge.

nasirsoft93, ‘’WordCloud’’, GitHub, 11 October 2021, https://github.com/nasirsoft93/python/blob/master/WordCloud.txt.

nasirsoft93, ‘’Flair Sentiment Analysis’’, GitHub, 12 December 2021, https://github.com/nasirsoft93/NLP/blob/master/Flair_Sentiment_Analysis.txt.

Nimmo, Ben. 2019. “Measuring Traffic Manipulation on Twitter.” Working Paper 2019.1, Oxford Internet Institute, Oxford.

O'Connor, Cailin, and James Owen Weatherall. 2019. *The Misinformation Age: How False Beliefs Spread*. New Haven: Yale University Press.

Paterson, Thomas, and Lauren Hanley. 2020. ‘’Political Warfare in the Digital Age: Cyber Subversion, Information Operations and ‘Deep Fakes’.’’ *Australian Journal of International Affairs* 74 (4): 439-454. https://doi.org/10.1080/10357718.2020.1734772.

Prier, Jarred. 2020. “Commanding the Trend. Social Media as Information Warfare.” In *Information Warfare in the Age of Cyber Conflict*, edited by Christopher Whyte, Trevor Thrall, and Brian Mazanec. Abingdon-on-Thames: Routledge.

Przeworski, Adam. 2019. *Crises of Democracy*. Cambridge: Cambridge University Press.

“Python for Data Science - Part 1: Visualizing Tweets and Retweets,” Cornell Virtual Workshop, https://cvw.cac.cornell.edu/PyDataSci1/tweets_retweets.

Renz, Bettina. 2016. “Russia and ‘Hybrid Warfare.’” *Contemporary Politics* 22 (3): 283–300. https://doi.org/10.1080/13569775.2016.1201316.

Rid, Thomas. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare*. New York: Farrar, Straus and Giroux.

Singer, P. W., and E. T. Brooking. 2018. *Likewar: The Weaponization of Social Media*. Boston: Eamon Dolan/Houghton Mifflin Harcourt.

Sidorenko-Bautista, Pavel, Nadia Alonso-López, and Adriana Paíno-Ambrosio. 2023. ‘’TikTok as a New Paradigm for Information in the Ukrainian War. A Study from the West of the Initial Coverage of the Conflict through This Platform.’’ *Estudios sobre el Mensaje Periodístico* 29 (2): 737-748. https://dx.doi.org/10.5209/esmp.84815.

Somerville, Alistair, and Jonas Heering. 2020. “The Disinformation Shift: From Foreign to Domestic.” *Georgetown Journal of International Affairs*. November 2020. https://gjia.georgetown.edu/2020/11/28/the-disinformation-shift-from-foreign-to-domestic/.

Starbird, Kate, and Tom Wilson. (2020). Cross-Platform Disinformation Campaigns: Lessons Learned and Next Steps. *Harvard Kennedy School (HKS) Misinformation Review* 1 (1): 1-11. https://doi.org/10.37016/mr-2020-002.

Statista, “Number of Twitter Users Worldwide,” 2021, https://www.statista.com/statistics/303681/twitter-users-worldwide/.

Sudalai Rajkumar, ‘’Getting Started with Text Preprocessing’’, Kaggle, https://www.kaggle.com/code/sudalairajkumar/getting-started-with-text-preprocessing.

Tambe, Arjun M., and Toni Friedman. 2022. ‘’Chinese State Media Facebook Ads Are Linked to Changes in News Coverage of China Worldwide.’’ *Harvard Kennedy School (HKS) Misinformation Review* 3 (1): 1-28. https://doi.org/10.37016/mr-2020-88.

Tay, Li Qian, Stephan Lewandowsky, Mark J. Hurlstone, Tim Kurz, and Ullrich K. H. Ecker. 2024. ‘’Thinking Clearly about Misinformation.’’ *Communications Psychology* 2 (4): 1-5. https://doi.org/10.1038/s44271-023-00054-5.

Thomas, Tomothy L. 2004. ‘’Russia’s Reflexive Control Theory and the Military.’’ *Journal of Slavic Military Studies* 17: 237-256. https://doi.org/10.1080/13518040490450529.

Thomas, Tomothy L. 2019. *Russia’s Reflexive Control Theory: Manipulating an Opponent to One’s Advantage*. Mclean: The MITRE Corporation.

Thornton, Rod. 2015. “The Changing Nature of Modern Warfare.” *The RUSI Journal* 160 (4): 40–48. https://doi.org/10.1080/03071847.2015.1079047.

‘’Tweets_retweets’’, Cornell Virtual Workshop, https://cvw.cac.cornell.edu/PyDataSci1/tweets_retweets.

’’Update on Twitter’s Review of the 2016 US Election,” Twitter, 19 January 2018, https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The Spread of True and False News Online.” *Science* 359 (6380): 1146–51. https://doi.org/10.1126/science.aap9559.

‘’Working With Text Data’’, Scikit Learn, https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html?highlight=machine+learning.

Yablokov, Ilya, and Precious N Chatterje-Doody. 2022. *Russia Today and Conspiracy Theories. People, Power and Politics on RT*. London: Routledge.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Appendix A

Python Code Used

Establishing the Time Frame of the Dataset

```
import pandas as pd

data = pd.read_csv (r'path_to_location/Entity.csv')

df = data['tweet_time']

print(min(df))
print(max(df))
```

Establishing the Tweets versus Retweets Ratio

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Entity.csv')

tweets_counted = data['is_retweet'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Tweet Type")
plt.savefig("Entity Tweet Type.jpg")
```

Establishing the Number of Account Languages Used

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Entity.csv')

tweets_counted = data['account_language'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Account Languages")
plt.savefig("Entity Account Languages.jpg")
```

Establishing the Number of Tweet Languages Used

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Entity.csv')

tweets_counted = data['tweet_language'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Tweet Languages")
plt.savefig("Entity Tweet Languages.jpg")
```

Establishing the Account Creation Dates

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv r'path_to_location/Entity.csv')

tweets_counted = data['account_creation_date'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Account Creation Date")
ax= plt.subplot()
plt.setp(ax.get_xticklabels(), rotation=90, ha='right')
plt.savefig("Entity Account Creation Date.jpg")
```

Main Hashtags Used

```
import pandas as pd

data = pd.read_csv (r'path_to_location/Entity.csv')

tweets_counted = data['hashtags'].value_counts()
print(tweets_counted)
```

Main URLS Used

```
import pandas as pd

data = pd.read_csv (r'path_to_location/Entity.csv'

tweets_counted = data['urls'].value_counts()
print(tweets_counted)
```

Main User Mentions

```
import pandas as pd

data = pd.read_csv (r'path_to_location/Entity.csv')

tweets_counted = data['user_mentions'].value_counts()
print(tweets_counted)
```

Like Counts

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Entity.csv'

tweets_counted = data['like_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Overall Likes")
plt.savefig("Entity Overall Likes.jpg")
tweets_counted = data['like_count'].value_counts()
print(tweets_counted)
```

Overall Follower Count

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Entity.csv'

tweets_counted = data['follower_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Overall Follower Count")
plt.savefig("Entity Overall Follower Count.jpg")
```

Overall Following Count

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Entity.csv')

tweets_counted = data['following_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Overall Following Count")
plt.savefig("Entity Overall Following Count.jpg")
```

Filter Tweet Metadata

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv(r'path_to_location/Entity.csv')
filtered = df[(df['is_retweet'] == False)]

tweets_counted = filtered['account_creation_date'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Tweet Account Creation Date (Tweets)")
plt.savefig("Entity Tweet Account Creation Date (Tweets).jpg")

tweets_counted = filtered['account_language'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Tweet Account Language (Tweets)")
plt.savefig("Entity Tweet Account Languages (Tweets).jpg")

tweets_counted = filtered['tweet_language'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Tweet Language (Tweets)")
plt.savefig("Entity Tweet Language (Tweets).jpg")

tweets_counted = filtered['hashtags'].value_counts()
print(tweets_counted)

tweets_counted = filtered['urls'].value_counts()
print(tweets_counted)
```

```
tweets_counted = filtered['user_mentions'].value_counts()
print(tweets_counted)

tweets_counted = filtered['like_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Tweet Likes (Tweets)")
plt.savefig("Entity Tweet Likes (Tweets).jpg")

tweets_counted = filtered['follower_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Tweet Follower Count (Tweets)")
plt.savefig("Entity Tweet Follower Count (Tweets).jpg")

tweets_counted = filtered['following_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Tweet Following Count (Tweets)")
plt.savefig("Entity Tweet Following Count (Tweets).jpg")
```

Filter Retweet Metadata

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv(r'path_to_location/Entity.csv')
filtered = df[(df['is_retweet'] == True)]

tweets_counted = filtered['account_creation_date'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Account Creation Date (Retweets)")
plt.savefig("Entity Account Creation Date (Retweets).jpg")

tweets_counted = filtered['account_language'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Account Language (Retweets)")
plt.savefig("Entity Account Language (Retweets).jpg")

tweets_counted = filtered['tweet_language'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Entity Tweet Language (Retweets)")
plt.savefig("Entity Tweet Language (Retweets).jpg")

tweets_counted = filtered['hashtags'].value_counts()
print(tweets_counted)

tweets_counted = filtered['urls'].value_counts()
print(tweets_counted)
```

```
tweets_counted = filtered['user_mentions'].value_counts()
print(tweets_counted)

tweets_counted = filtered['like_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Likes (Retweets)")
plt.savefig("Entity Likes (Retweets).jpg")

tweets_counted = filtered['follower_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Follower Count (Retweets)")
plt.savefig("Entity Follower Count (Retweets).jpg")

tweets_counted = filtered['following_count'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.line(title="Entity Following Count (Retweets)")
plt.savefig("Entity Following Count (Retweets).jpg")
```

Creating an English Language Only Twitter Dataset

```
import pandas as pd

fileloc = " r'path_to_location/Entity.csv'"
df = pd.read_csv(fileloc)
corpus = df[df['tweet_language'].str.contains("en", na=False)]

corpus.to_csv("Entity English.csv", index=False, encoding='utf-8-sig')
```

Counting URLS Mentioned in Retweets

```
import pandas as pd

df = pd.read_csv("path_to_location/Entity English.csv")

filtered = df[(df['is_retweet'] == True)]

tweets_counted = filtered['urls'].value_counts()

df2 = tweets_counted
df2.to_csv("URLS Retweet Entity.csv", index=False, encoding='utf-8-sig')
```

Counting URLS Mentioned in Tweets

```
import pandas as pd

df = pd.read_csv("r'path_to_location/Entity English.csv")

filtered = df[(df['is_retweet'] == False)]

tweets_counted = filtered['urls'].value_counts()

df2 = tweets_counted
df2.to_csv("URLS Tweets Entity.csv", index=False, encoding='utf-8-sig')
```

Making a Wordcloud for Hashtags

```
from wordcloud import WordCloud
import pandas as pd
import matplotlib.pylab as plt

data_file = pd.read_csv(r'path_to_location/Entity English.csv')

wordcloud = WordCloud(width=1600 ,
height=800,background_color="White",colormap="Set2").generate("
".join(map(str,data_file['hashtags'])))
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)

wordcloud.to_file ('path_to_location/Entity Hashtag Wordcloud.png')
plt.show()
```

Making a Wordcloud From TF-IDF on Hashtags

```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pylab as plt
from sklearn.feature_extraction.text import TfidfVectorizer
import re
import string
pd.options.mode.chained_assignment = None

full_df = pd.read_csv("r'path_to_location/Entity English.csv")
df = full_df[["hashtags"]]
df["hashtags"] = df["hashtags"].astype(str)
full_df.head()

df["text_lower"] = df["hashtags"].str.lower()

df2 = df["text_lower"]

df.drop(["text_lower"], axis=1, inplace=True)

PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    """custom function to remove the punctuation"""
    return text.translate(str.maketrans('', '', PUNCT_TO_REMOVE))

df["text_wo_punct"] = df2.apply(lambda text: remove_punctuation(text))

df3 = df["text_wo_punct"]

from nltk.corpus import stopwords
'', ''.join(stopwords.words('english'))

STOPWORDS = set(stopwords.words('english'))
```

```
def remove_stopwords(text):
    """custom function to remove the stopwords"""
    return " ".join([word for word in str(text).split() if word not in STOPWORDS])

df["no_stopwords"] = df3.apply(lambda text: remove_stopwords(text))

df4 = df["no_stopwords"]

def remove_urls(text):
    url_pattern = re.compile(r'https?://\S+|www\.\S+')
    return url_pattern.sub(r'', string)

df["no_urls"] = df4.apply(lambda text: remove_punctuation(text))

df5 = df["no_urls"]

def remove_html(text):
    html_pattern = re.compile('<.*?>')
    return html_pattern.sub(r'', string)

df["no_html"] = df5.apply(lambda text: remove_punctuation(text))

df6 = df["no_html"]

corpus = df6

tr_idf_model  = TfidfVectorizer()
tf_idf_vector = tr_idf_model.fit_transform(corpus)

tf_idf_array = tf_idf_vector.toarray()

words_set = tr_idf_model.get_feature_names()

df_tf_idf = pd.DataFrame(tf_idf_array, columns = words_set)
```

```
df2 = df_tf_idf

df2.T.sum(axis=1)
wordcloud = WordCloud(background_color="white",
max_words=50).generate_from_frequencies(df2.T.sum(axis=1))
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)

wordcloud.to_file ('Hashtag TF-IDF Entity.png')
plt.show()
```

Extracting Duplicate Spam Messages

```
import pandas as pd

df = pd.read_csv(r'path_to_location/Entity English.csv')

df2 = df[['account_creation_date', 'tweet_text', 'tweet_time']]

corpus = df2[df2.duplicated()]

corpus.to_csv("Duplicates Entity.csv", index=False, encoding='utf-8-sig')
```

Establishing a Tweet Timeline for Duplicates[23]

```
import matplotlib.pyplot as plt
import pandas as pd

tweet_df = pd.read_csv('path_to_location/Duplicates Entity.csv', \
    parse_dates=['tweet_time'])

tweet_df_5min = tweet_df.groupby(pd.Grouper(key='tweet_time')).size()
tweet_df_5min.plot(figsize=(18,6))
plt.ylabel('Tweets')
plt.title('Entity Duplicates Publication Timeline')
plt.grid(True)
```

Establishing an Account Creation Date Timeline for Duplicates[24]

```
import matplotlib.pyplot as plt
import pandas as pd

tweet_df = pd.read_csv('path_to_locationDuplicates Entity.csv', \
    parse_dates=['account_creation_date'])

tweet_df_5min = tweet_df.groupby(pd.Grouper(key='tweet_time')).size()
tweet_df_5min.plot(figsize=(18,6))
plt.ylabel('Tweets')
plt.title('Entity Account Creation Date Timeline')
plt.grid(True)
```

---

[23] Source: ‘’Tweets_retweets’’, Cornell Virtual Workshop, https://cvw.cac.cornell.edu/PyDataSci1/tweets_retweets.

[24] Ibidem.

TF-IDF on Duplicates

```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pylab as plt
from sklearn.feature_extraction.text import TfidfVectorizer

df =  pd.read_csv("path_to_locationDuplicates Entity.csv")
text = df["tweet_text"]

tr_idf_model  = TfidfVectorizer()
tf_idf_vector = tr_idf_model.fit_transform(text)

tf_idf_array = tf_idf_vector.toarray()

words_set = tr_idf_model.get_feature_names()

df_tf_idf = pd.DataFrame(tf_idf_array, columns = words_set)

df2 = df_tf_idf

df2.T.sum(axis=1)
wordcloud = WordCloud(background_color="white",
max_words=50).generate_from_frequencies(df2.T.sum(axis=1))
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)

wordcloud.to_file ('path_to_locationDuplicate Account TF-IDF Wordcloud.png')
plt.show()
```

Sentiment Analysis on Duplicates[25]

```
import pandas as pd
from flair.models import TextClassifier
from flair.data import Sentence
sia = TextClassifier.load('en-sentiment')

data_file = pd.read_csv('path_to_location/Duplicates Entity.csv')

def sentiment_Flair(x):
    sentence = Sentence(x)
    sia.predict(sentence)
    score = sentence.labels[0]
    if "POSITIVE" in str(score):
        return "positive"
    elif "NEGATIVE" in str(score):
        return "negative"
    else:
        return "neutral"

data_file['sentiment_flair'] = data_file['tweet_text'].apply(lambda x: sentiment_Flair(x))

csv_data = data_file.to_csv('path_to_locationEntity Duplicates Sentiment.csv')
```

[25] Source: nasirsoft93, ‘’Flair_Sentiment_Analysis.txt’’, GitHub, 12 December 2021, https://github.com/nasirsoft93/NLP/blob/master/Flair_Sentiment_Analysis.txt.

Creating the Main Corpus for Analysis

```
import pandas as pd

df = pd.read_csv('path_to_location/Entity English.csv')

corpus = df[['tweet_text', 'tweet_time', 'is_retweet']]

corpus.to_csv("Corpus Sentiment Entity.csv", index=False, encoding='utf-8-sig')
```

Creating a Corpus for Tweet Text Only

```
import pandas as pd

df = pd.read_csv('path_to_location/Entity English.csv')

corpus = df['tweet_text']

corpus.to_csv("Text Only Entity.csv", index=False, encoding='utf-8-sig')
```

Pre-process the Text CSV[26]

```
import pandas as pd
import re
import string
pd.options.mode.chained_assignment = None

full_df = pd.read_csv("path_to_location/Text Only Entity.csv")
df = full_df[["tweet_text"]]
df["tweet_text"] = df["tweet_text"].astype(str)
full_df.head()
```

[26] Source: Sudalai Rajkumar, ‘’Getting started with Text Preprocessing’’, Kaggle, https://www.kaggle.com/code/sudalairajkumar/getting-started-with-text-preprocessing.

```
df["text_lower"] = df["tweet_text"].str.lower()

df2 = df["text_lower"]

df.drop(["text_lower"], axis=1, inplace=True)

PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    """custom function to remove the punctuation"""
    return text.translate(str.maketrans('', '', PUNCT_TO_REMOVE))

df["text_wo_punct"] = df2.apply(lambda text: remove_punctuation(text))

df3 = df["text_wo_punct"]

from nltk.corpus import stopwords
'', ''.join(stopwords.words('english'))

STOPWORDS = set(stopwords.words('english'))
def remove_stopwords(text):
    """custom function to remove the stopwords"""
    return " ".join([word for word in str(text).split() if word not in STOPWORDS])

df["no_stopwords"] = df3.apply(lambda text: remove_stopwords(text))

df4 = df["no_stopwords"]

def remove_urls(text):
    url_pattern = re.compile(r'https?://\S+|www\.\S+')
    return url_pattern.sub(r'', string)

df["no_urls"] = df4.apply(lambda text: remove_punctuation(text))
```

```
df5 = df["no_urls"]

def remove_html(text):
    html_pattern = re.compile('<.*?>')
    return html_pattern.sub(r'', string)

df["no_html"] = df5.apply(lambda text: remove_punctuation(text))

df6 = df["no_html"]

corpus = df6

corpus.to_csv("Processed Entity.csv", index=False, encoding='utf-8-sig')
```

Making a Wordcloud for Term Frequency[27]

```
from wordcloud import WordCloud
import pandas as pd
import matplotlib.pylab as plt

data_file = pd.read_csv('path_to_location/Processed Entity.csv')

wordcloud = WordCloud(width=1600 ,
height=800,background_color="White",colormap="Set2").generate("
".join(map(str,data_file['no_html'])))
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)

wordcloud.to_file ('path_to_location/Entity Term Frequency Wordcloud.png')
```

[27] Source: nasirsoft93, ‘’WordCloud.txt’’, GitHub, 11 October 2021, https://github.com/nasirsoft93/python/blob/master/WordCloud.txt.

```
plt.show()
```

TF-IDF[28]

```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pylab as plt
from sklearn.feature_extraction.text import TfidfVectorizer

text = open("path_to_location/Processed Entity.csv", "r", encoding="utf8")

tr_idf_model  = TfidfVectorizer()
tf_idf_vector = tr_idf_model.fit_transform(text)

tf_idf_array = tf_idf_vector.toarray()

words_set = tr_idf_model.get_feature_names()

df_tf_idf = pd.DataFrame(tf_idf_array, columns = words_set)

df = df_tf_idf

df.T.sum(axis=1)
wordcloud = WordCloud(background_color="white",
max_words=50).generate_from_frequencies(df.T.sum(axis=1))
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)
```

[28] Source: Fatih Karabiber, ‘’TF-IDF — Term Frequency-Inverse Document Frequency’’, Learn Datasci, https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/.

```
wordcloud.to_file ('path_to_locationTF-IDF Entity Wordcloud.png')
plt.show()
```

Creating a Corpus for Sentiment Analysis on Retweets

```
import pandas as pd

df = pd.read_csv("path_to_location/Entity English.csv")

retweets = df[(df['is_retweet'] == True)]

df2 = retweets[['tweet_text', 'tweet_time']]
df2.to_csv("Retweet Text Entity.csv", index=False, encoding='utf-8-sig')
```

Creating a Corpus for Sentiment Analysis on Tweets

```
import pandas as pd

df = pd.read_csv("path_to_location/Entity English.csv")

retweets = df[(df['is_retweet'] == False)]

df2 = retweets[['tweet_text', 'tweet_time']]
df2.to_csv("Tweet Text Entity.csv", index=False, encoding='utf-8-sig')
```

Sentiment Analysis on Retweets[29]

```
import pandas as pd
from flair.models import TextClassifier
from flair.data import Sentence
sia = TextClassifier.load('en-sentiment')

data_file = pd.read_csv('path_to_location/Retweet Text Entity.csv')

def sentiment_Flair(x):
    sentence = Sentence(x)
    sia.predict(sentence)
    score = sentence.labels[0]
    if "POSITIVE" in str(score):
        return "positive"
    elif "NEGATIVE" in str(score):
        return "negative"
    else:
        return "neutral"

data_file['sentiment_flair'] = data_file['tweet_text'].apply(lambda x: sentiment_Flair(x))

csv_data = data_file.to_csv('Retweet Sentiment Entity.csv')
```

[29] Source: nasirsoft93, ‘’Flair_Sentiment_Analysis.txt’’, GitHub, 12 December 2021, https://github.com/nasirsoft93/NLP/blob/master/Flair_Sentiment_Analysis.txt.

Retweet Sentiment Plots

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Retweet Sentiment Entity.csv')

tweets_counted = data['sentiment_flair'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Retweet Sentiment Entity")
plt.savefig("Retweet Sentiment Entity.jpg")
```

Sentiment Analysis on Tweets

```
import pandas as pd
from flair.models import TextClassifier
from flair.data import Sentence
sia = TextClassifier.load('en-sentiment')

data_file = pd.read_csv('path_to_location/Tweet Text Entity.csv')

def sentiment_Flair(x):
    sentence = Sentence(x)
    sia.predict(sentence)
    score = sentence.labels[0]
    if "POSITIVE" in str(score):
        return "positive"
    elif "NEGATIVE" in str(score):
        return "negative"
    else:
        return "neutral"
```

```
data_file['sentiment_flair'] = data_file['tweet_text'].apply(lambda x: sentiment_Flair(x))

csv_data = data_file.to_csv('Tweet Sentiment Entity.csv')
```

Tweet Sentiment Plots

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv (r'path_to_location/Tweet Sentiment Entity.csv')

tweets_counted = data['sentiment_flair'].value_counts()
print(tweets_counted)

df = pd.DataFrame(tweets_counted)
df.plot.bar(title="Tweet Sentiment Entity")
plt.savefig("Tweet Sentiment Entity.jpg")
```

Latent Dirichlet Allocation (LDA) Topic Modelling

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
import re

df = pd.read_csv('path_to_file/Entity.csv')

def my_normal_function(x):
    return x**2 + 10
my_lambda_function = lambda x: x**2 + 10
```

```
df['is_retweet'] = df['tweet'].apply(lambda x: x[:2]=='RT')
df['is_retweet'].sum()  # number of retweets

df.loc[df['is_retweet']].tweet.unique().size

df.groupby(['tweet']).size().reset_index(name='counts')\
  .sort_values('counts', ascending=False).head(10)

counts = df.groupby(['tweet']).size()\
        .reset_index(name='counts')\
        .counts

my_bins = np.arange(0,counts.max()+2, 1)-0.5

plt.figure()
plt.hist(counts, bins = my_bins)
plt.xlabels = np.arange(1,counts.max()+1, 1)
plt.xlabel('copies of each tweet')
plt.ylabel('frequency')
plt.yscale('log')
plt.show()

def find_retweeted(tweet):
    '''This function will extract the twitter handles of retweed people'''
    return re.findall('(?<=RT\s)(@[A-Za-z]+[A-Za-z0-9-_]+)', tweet)

def find_mentioned(tweet):
    '''This function will extract the twitter handles of people mentioned in the tweet'''
    return re.findall('(?<!RT\s)(@[A-Za-z]+[A-Za-z0-9-_]+)', tweet)

def find_hashtags(tweet):
    '''This function will extract hashtags'''
    return re.findall('(#[A-Za-z]+[A-Za-z0-9-_]+)', tweet)
```

```
df['retweeted'] = df.tweet.apply(find_retweeted)
print(df['retweeted'])
df['mentioned'] = df.tweet.apply(find_mentioned)
print(df['mentioned'])
df['hashtags'] = df.tweet.apply(find_hashtags)
print(df['hashtags'])

hashtags_list_df = df.loc[
                    df.hashtags.apply(
                       lambda hashtags_list: hashtags_list !=[]
                    ),['hashtags']]

flattened_hashtags_df = pd.DataFrame(
   [hashtag for hashtags_list in hashtags_list_df.hashtags
   for hashtag in hashtags_list],
   columns=['hashtag'])

flattened_hashtags_df['hashtag'].unique().size

popular_hashtags = flattened_hashtags_df.groupby('hashtag').size()\
                                .reset_index(name='counts')\
                                .sort_values('counts', ascending=False)\
                                .reset_index(drop=True)

counts = flattened_hashtags_df.groupby(['hashtag']).size()\
                        .reset_index(name='counts')\
                        .counts

my_bins = np.arange(0,counts.max()+2, 5)-0.5

plt.figure()
plt.hist(counts, bins = my_bins)
plt.xlabels = np.arange(1,counts.max()+1, 1)
```

```
plt.xlabel('hashtag number of appearances')
plt.ylabel('frequency')
plt.yscale('log')
plt.show()

min_appearance = 10
popular_hashtags_set = set(popular_hashtags[
                    popular_hashtags.counts>=min_appearance
                    ]['hashtag'])

hashtags_list_df['popular_hashtags'] = hashtags_list_df.hashtags.apply(
        lambda hashtag_list: [hashtag for hashtag in hashtag_list
                        if hashtag in popular_hashtags_set])

popular_hashtags_list_df = hashtags_list_df.loc[
        hashtags_list_df.popular_hashtags.apply(lambda hashtag_list: hashtag_list !=[])]

hashtag_vector_df = popular_hashtags_list_df.loc[:, ['popular_hashtags']]

for hashtag in popular_hashtags_set:
    # make columns to encode presence of hashtags
    hashtag_vector_df['{}'.format(hashtag)] = hashtag_vector_df.popular_hashtags.apply(
        lambda hashtag_list: int(hashtag in hashtag_list))

hashtag_matrix = hashtag_vector_df.drop('popular_hashtags', axis=1)

correlations = hashtag_matrix.corr()

plt.figure(figsize=(10,10))
sns.heatmap(correlations,
    cmap='RdBu',
    vmin=-1,
    vmax=1,
    square = True,
```

```
    cbar_kws={'label':'correlation'})
plt.show()

import nltk
from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords

def remove_links(tweet):
    '''Takes a string and removes web links from it'''
    tweet = re.sub(r'http\S+', '', tweet) # remove http links
    tweet = re.sub(r'bit.ly/\S+', '', tweet) # rempve bitly links
    tweet = tweet.strip('[link]') # remove [links]
    return tweet

def remove_users(tweet):
    '''Takes a string and removes retweet and @user information'''
    tweet = re.sub('(RT\s@[A-Za-z]+[A-Za-z0-9-_]+)', '', tweet) # remove retweet
    tweet = re.sub('(@[A-Za-z]+[A-Za-z0-9-_]+)', '', tweet) # remove tweeted at
    return tweet

my_stopwords = nltk.corpus.stopwords.words('english')
word_rooter = nltk.stem.snowball.PorterStemmer(ignore_stopwords=False).stem
my_punctuation = '!"$%&\'()*+,-./:;<=>?[\\]^_`{|}~•@'

def clean_tweet(tweet, bigrams=False):
    tweet = remove_users(tweet)
    tweet = remove_links(tweet)
    tweet = tweet.lower() # lower case
    tweet = re.sub('['+my_punctuation + ']+', ' ', tweet) # strip punctuation
    tweet = re.sub('\s+', ' ', tweet) #remove double spacing
    tweet = re.sub('([0-9]+)', '', tweet) # remove numbers
    tweet_token_list = [word for word in tweet.split(' ')
                        if word not in my_stopwords] # remove stopwords
```

```
    tweet_token_list = [word_rooter(word) if '#' not in word else word
                        for word in tweet_token_list] # apply word rooter
    if bigrams:
        tweet_token_list = tweet_token_list+[tweet_token_list[i]+'_'+tweet_token_list[i+1]
                                             for i in range(len(tweet_token_list)-1)]
    tweet = ' '.join(tweet_token_list)
    return tweet

df['clean_tweet'] = df.tweet.apply(clean_tweet)

from wordcloud import WordCloud
import matplotlib.pylab as plt

long_string = ','.join(list(df['clean_tweet'].values))

wordcloud = WordCloud(background_color="white", max_words=5000, contour_width=3,
contour_color='steelblue')

wordcloud.generate(long_string)

wordcloud.to_image()
wordcloud.to_file ('path_to_file/Wordcloud Entity.png')
plt.show()

from sklearn.feature_extraction.text import CountVectorizer

vectorizer = CountVectorizer(max_df=0.9, min_df=25, token_pattern='\w+|\$[\d\.]+|\S+')

tf = vectorizer.fit_transform(df['clean_tweet']).toarray()

# tf_feature_names tells us what word each column in the matric represents
tf_feature_names = vectorizer.get_feature_names_out()

from sklearn.decomposition import LatentDirichletAllocation
```

```
number_of_topics = 10

model = LatentDirichletAllocation(n_components=number_of_topics, random_state=0)

model.fit(tf)

def display_topics(model, feature_names, no_top_words):
    topic_dict = {}
    for topic_idx, topic in enumerate(model.components_):
        topic_dict["Topic %d words" % (topic_idx)]= ['{}'.format(feature_names[i])
                    for i in topic.argsort()[:-no_top_words - 1:-1]]
        topic_dict["Topic %d weights" % (topic_idx)]= ['{:.1f}'.format(topic[i])
                    for i in topic.argsort()[:-no_top_words - 1:-1]]
    return pd.DataFrame(topic_dict)

no_top_words = 10
display_topics(model, tf_feature_names, no_top_words)
```

Appendix B

Graphs and Tables

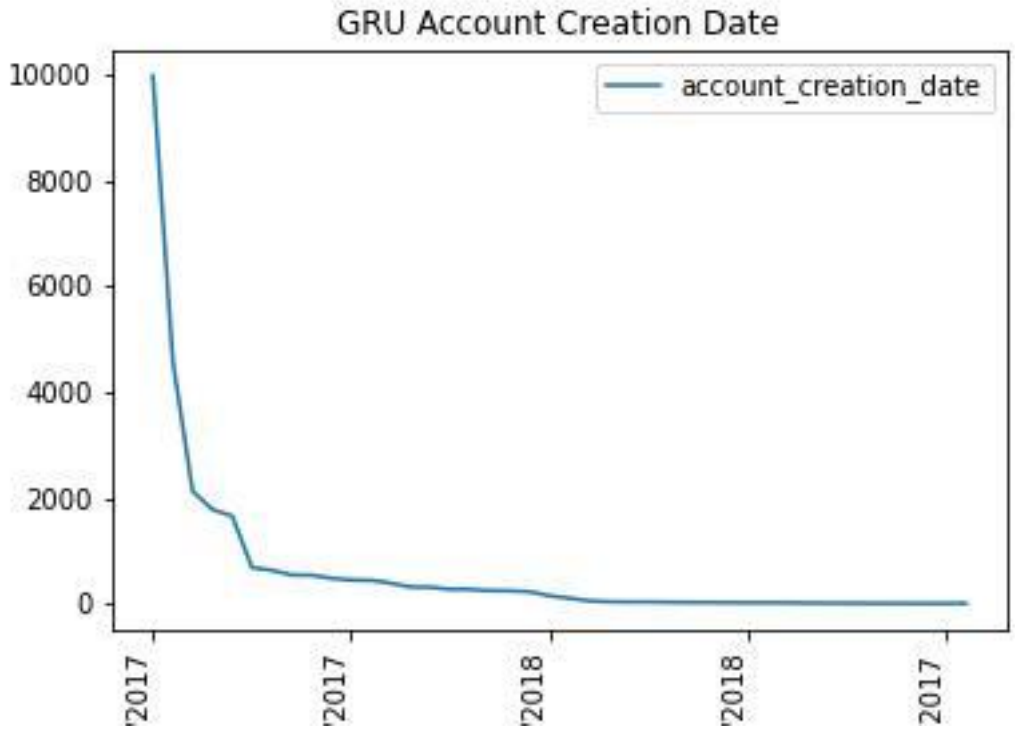
Russia

I. Data Overview

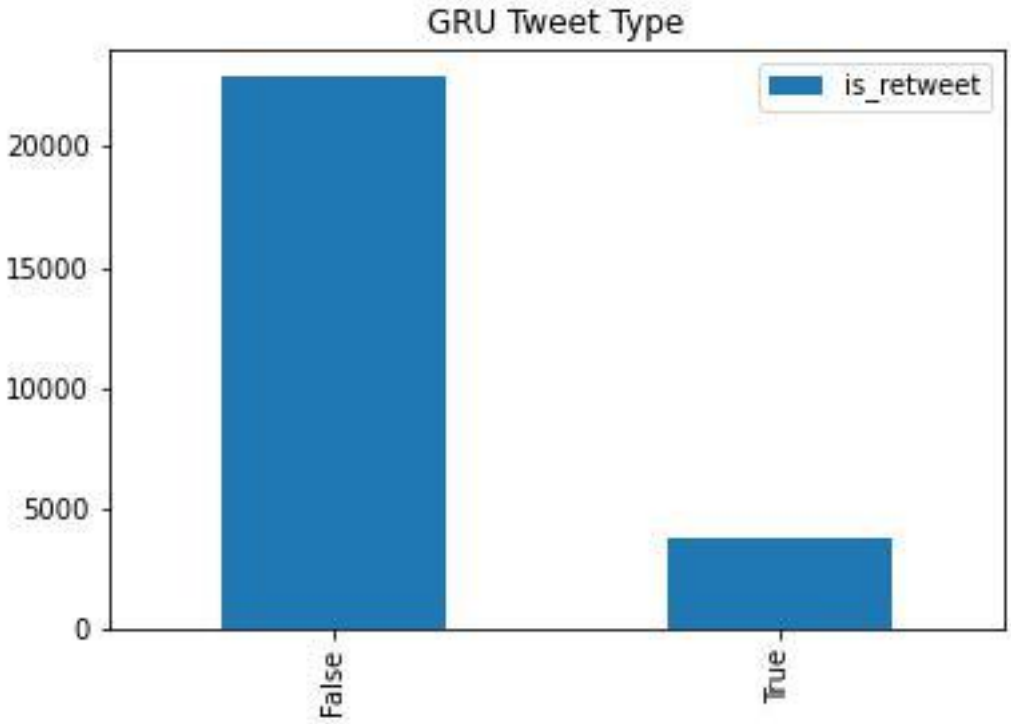
Russia	Agency	From	Till	Accounts	Tweets	Retweets	Total	English Language Tweets
	IRA	09/05/2009	27/12/2020	3644	5484222	3354693	8838915	1937873
	GRU	09/06/2014	18/10/2020	70	22938	3746	26684	14422
	REA	14/08/2012	17/03/2021	16	7663	60	7723	6823
	RNA	20/09/2012	28/12/2020	50	16155	2803	18958	11743
	Total	**09/05/2009**	**17/03/2021**	**3780**	**5530978**	**3361302**	**8892280**	**1970861**

II. GRU

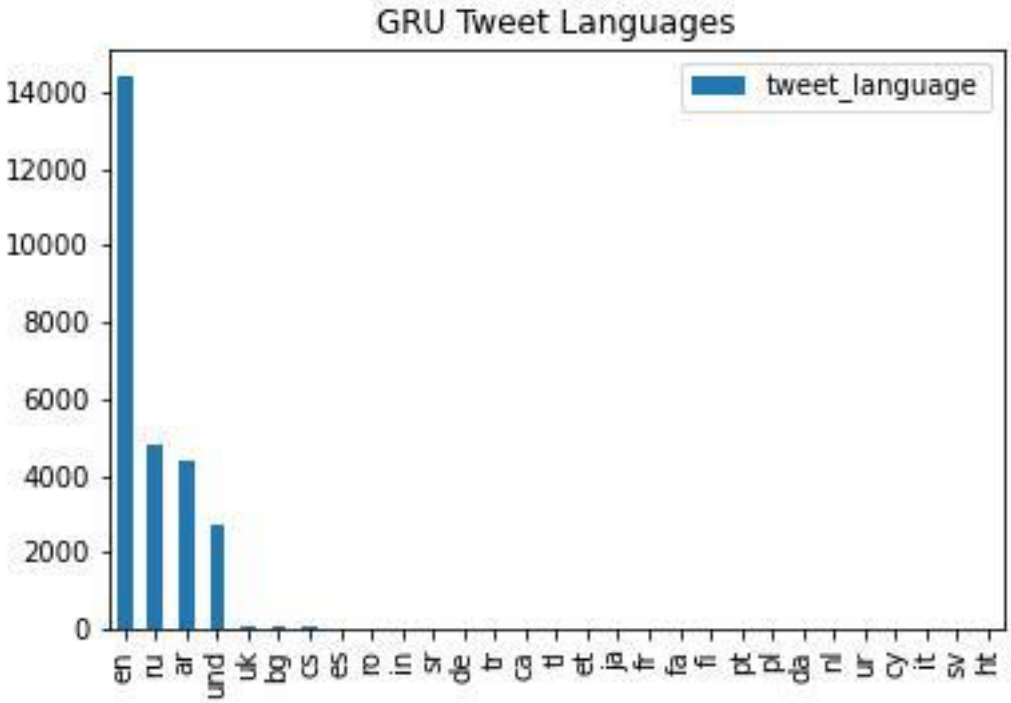
Account Creation Dates



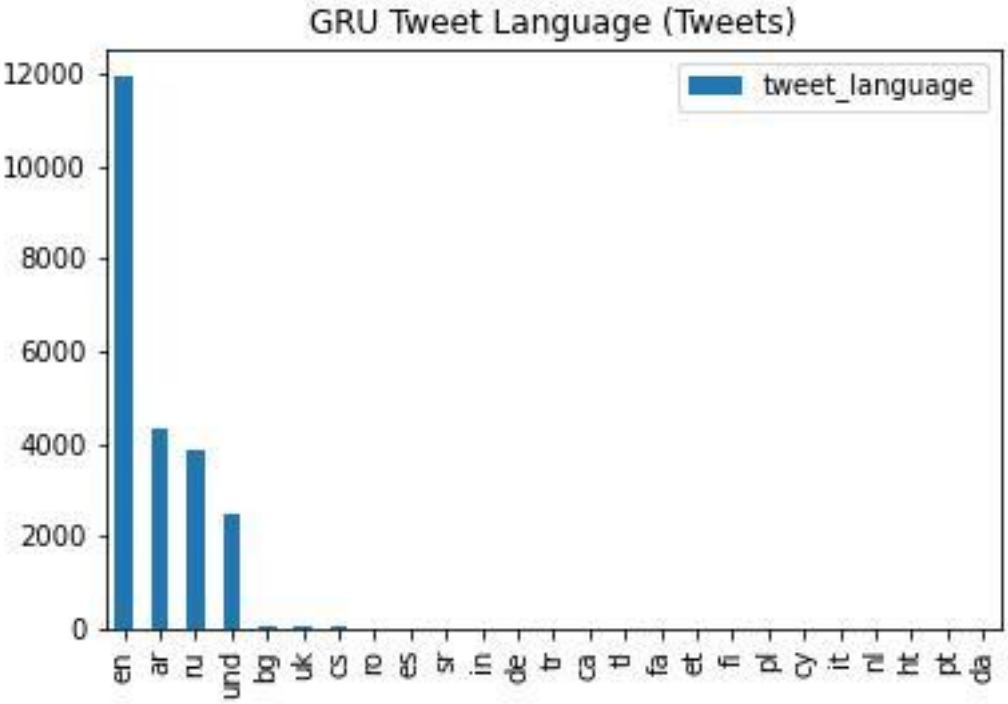
Tweet Types



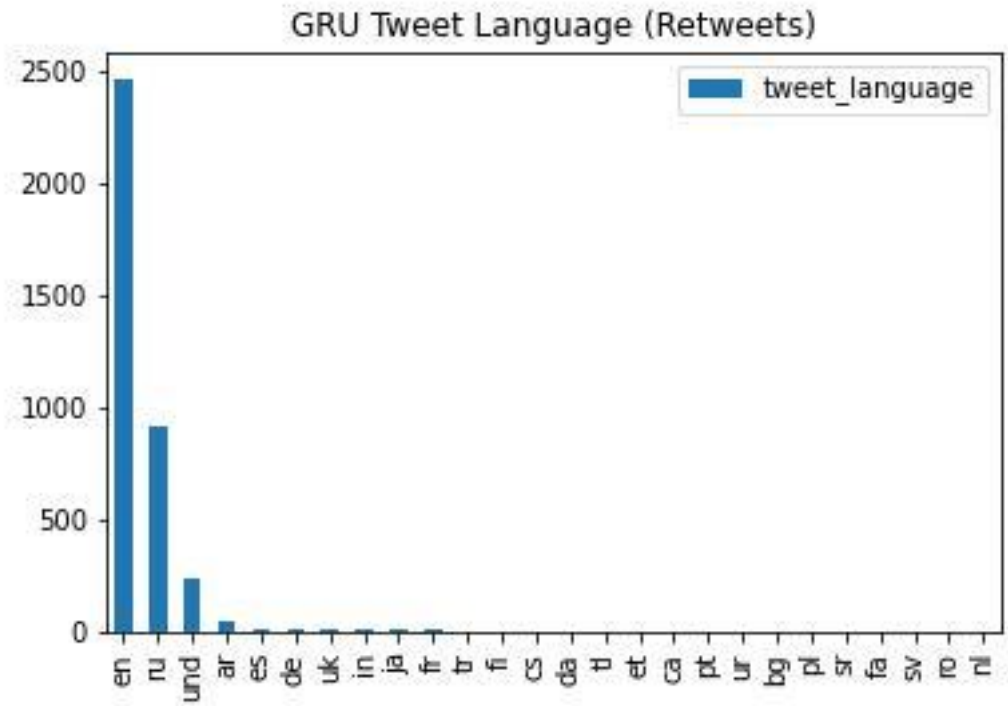
Tweet Languages (Overall)



Tweet Languages (Tweets)



Tweet Languages (Retweets)



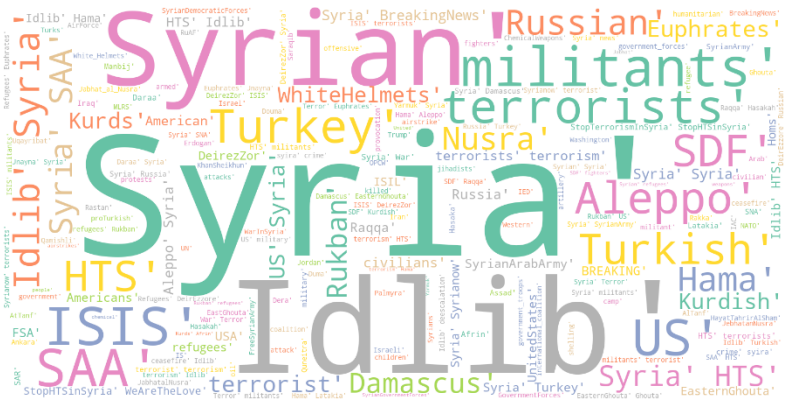
Term Frequency

TF–IDF (Term Frequency–Inverse Document Frequency)

Hashtags Shared

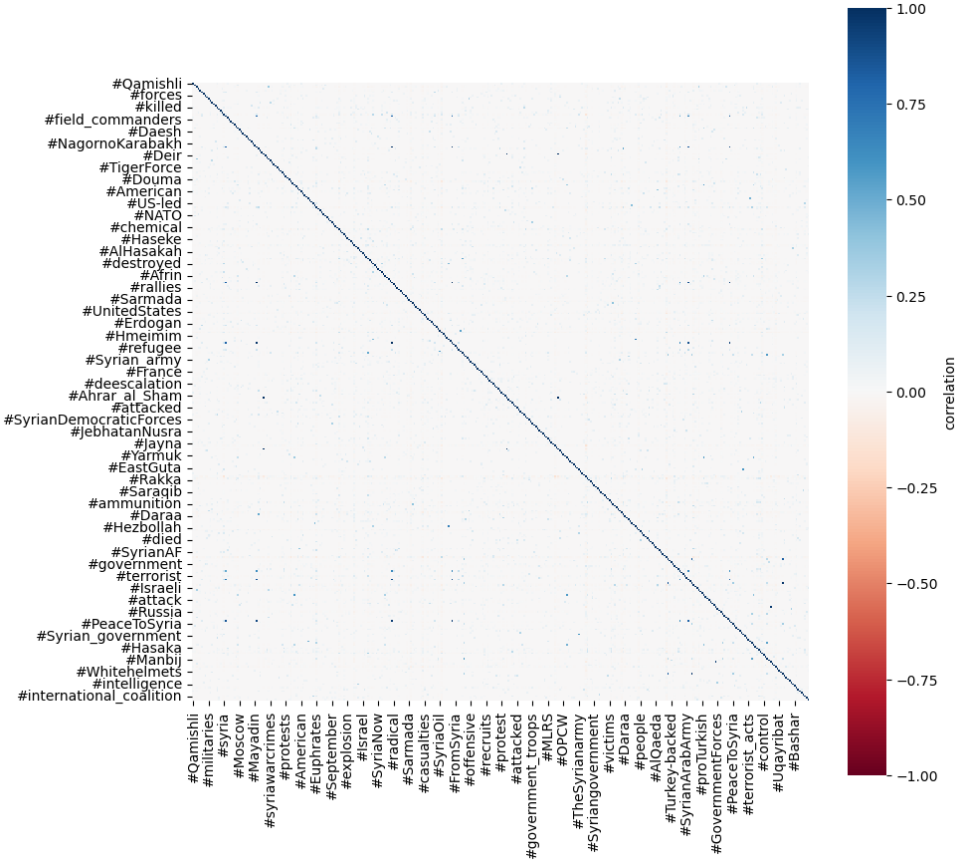
	GRU Hashtags			
	Most Shared		**Least Shared**	
Grand Total	['Syria']	172	['Ahrar_al_Sham', 'Turkey', 'Hazano', 'Idlib']	1
	['Idlib']	105	['October', 'militants', 'Jabhat', 'attack', 'fighters', 'Ahrar', 'occupied', 'territories', 'lost', 'fighters', 'had', 'retreat']	1
	['StopTerrorismInSyria', 'StopHTSinSyria', 'WeAreTheLove']	98	['Nusra', 'Ahrar_al_Sham']	1
	['crime', 'syira']	76	['fire', 'government', 'Syria', 'Idlib']	1
Tweets	['StopTerrorismInSyria', 'StopHTSinSyria', 'WeAreTheLove']	96	['Explosions', 'died', 'coalition']	1
	['Syria']	92	['Americanbase', 'Tanf', 'Syrian']	1
	['Idlib']	85	['Kurdish', 'negotiating', 'Damascus']	1
	['crime', 'syira']	76	['International', 'American']	1
Retweets	['Syria']	80	['Explosions', 'died', 'coalition']	1
	['Idlib']	20	['Americanbase', 'Tanf', 'Syrian']	1
	['Refugees']	20	['Kurdish', 'negotiating', 'Damascus']	1
	['WhiteHelmets']	16	['International', 'American']	1

Hashtags



TF–IDF Hashtags

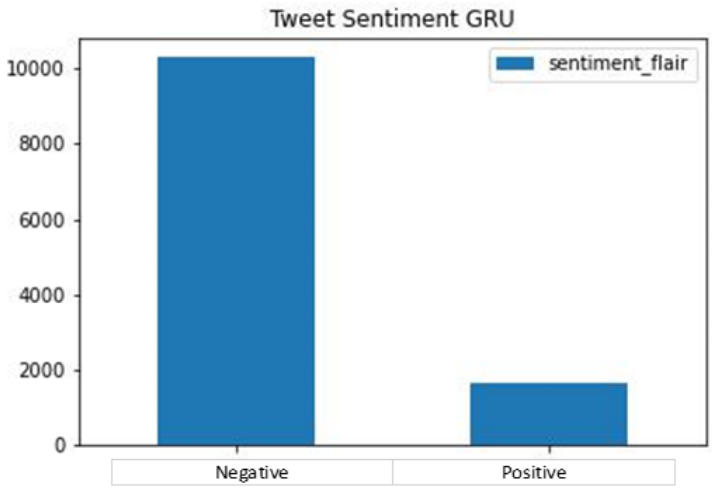
Hashtags Heatmap


#Qamishli
#forces
#killed
#field_commanders
#Daesh
#NagornoKarabakh
#Deir
#TigerForce
#Douma
#American
#US-led
#NATO
#chemical
#Haseke
#AlHasakah
#destroyed
#Afrin
#rallies
#Sarmada
#UnitedStates
#Erdogan
#Hmeimim
#refugee
#Syrian_army
#FFrance
#deescalation
#Ahrar_al_Sham
#attacked
#SyrianDemocraticForces
#JebhatanNusra
#Jayna
#Yarmuk
#EastGuta
#Rakka
#Saraqib
#ammunition
#Daraa
#Hezbollah
#died
#SyrianAF
#government
#terrorist
#Israeli
#attack
#Russia
#PeaceToSyria
#Syrian_government
#Hasaka
#Manbij
#Whitehelmets
#intelligence
#international_coalition
#militaries
#syria
#Moscow
#Mayadin
#syriawarcrimes
#protests
#American
#Euphrates
#September
#explosion
#Israel
#SyriaNow
#radical
#Sarmada
#casualties
#SyriaOil
#FromSyria
#offensive
#recruits
#protest
#attacked
#government_troops
#MLRS
#OPCW
#TheSyrianarmy
#Syriangovernment
#victims
#Daraa
#people
#AlQaeda
#Turkey-backed
#SyrianArabArmy
#proTurkish
#GovernmentForces
#PeaceToSyria
#terrorist_acts
#control
#Uqayribat
#Bashar
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
correlation

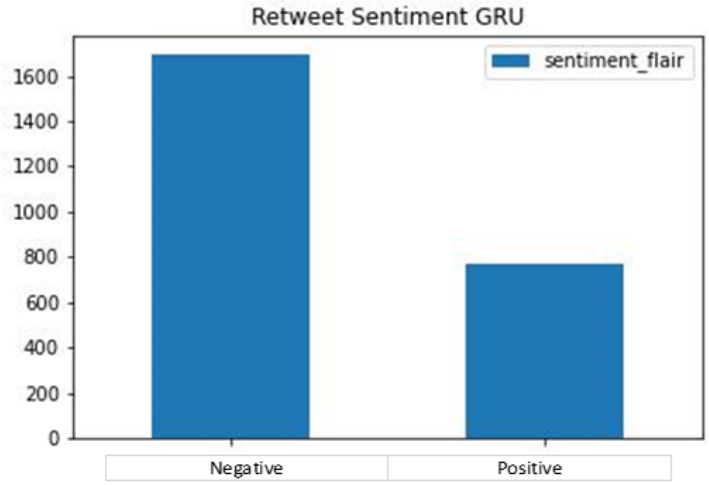
URLS Shared

	GRU URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['http://t.me/brussinf']	146	brussels snitch' Alexander Simonov ▪war correspondent; ▪specialist in the military-political activity of NATO and Western countries.
	['https://www.youtube.com/watch?v=2GuneHJyPSA']	145	Unavailable
	['http://en.farsnews.com/newstext.aspx?nn=13960224000929']	31	Unavailable
	['https://www.youtube.com/watch?v=1S3yIUoU10w']	28	Unavailable
Tweets	['http://t.me/brussinf']	146	See Above
	['https://www.youtube.com/watch?v=2GuneHJyPSA']	145	Unavailable
	['http://en.farsnews.com/newstext.aspx?nn=13960224000929']	31	Unavailable
	['https://www.youtube.com/watch?v=1S3yIUoU10w']	28	Unavailable
Retweets	['http://t.me/brussinf']	146	See Above
	['https://www.youtube.com/watch?v=2GuneHJyPSA']	145	Unavailable
	['http://en.farsnews.com/newstext.aspx?nn=13960224000929']	31	Unavailable
	['https://www.youtube.com/watch?v=1S3yIUoU10w']	28	Unavailable

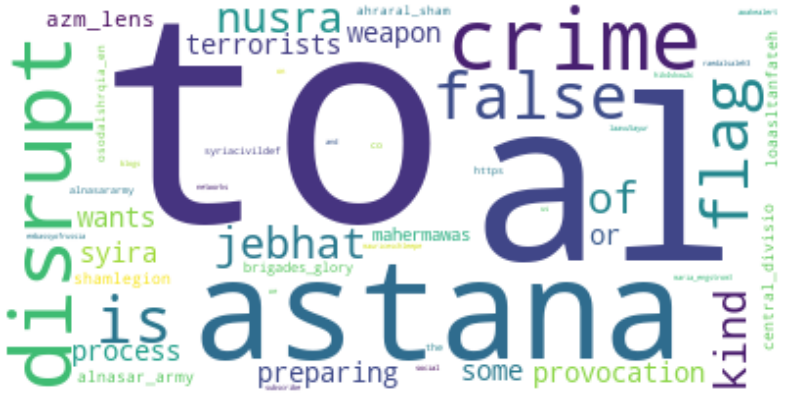
Tweet Sentiment Analysis



Retweet Sentiment Analysis



TF–IDF on Duplicate Tweets

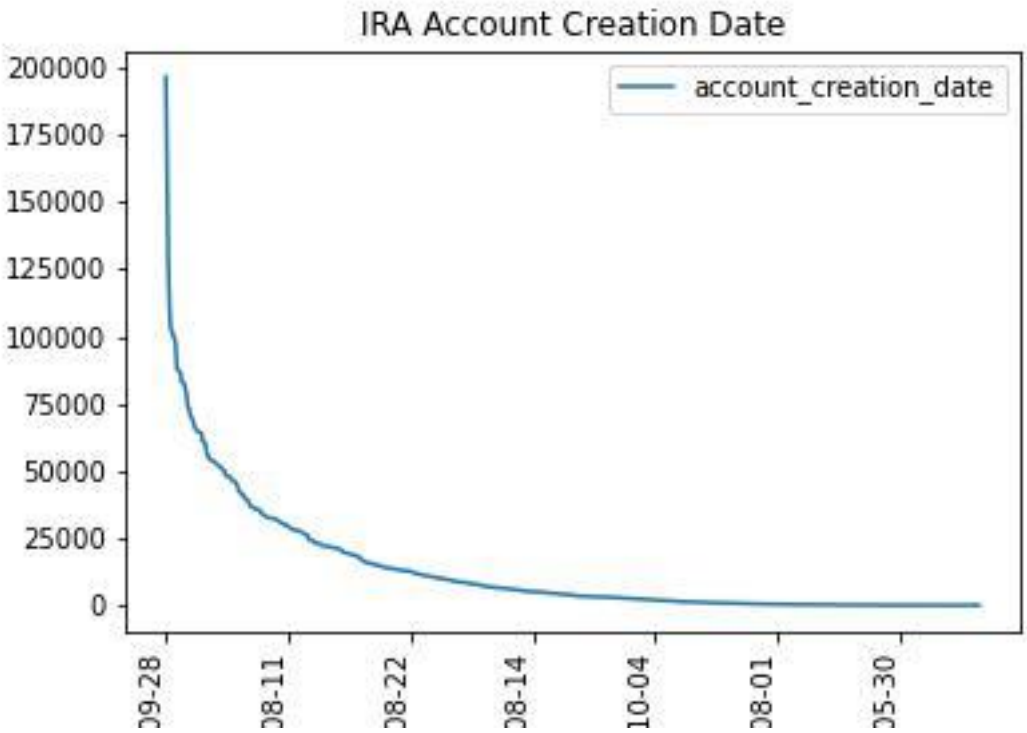


Topics

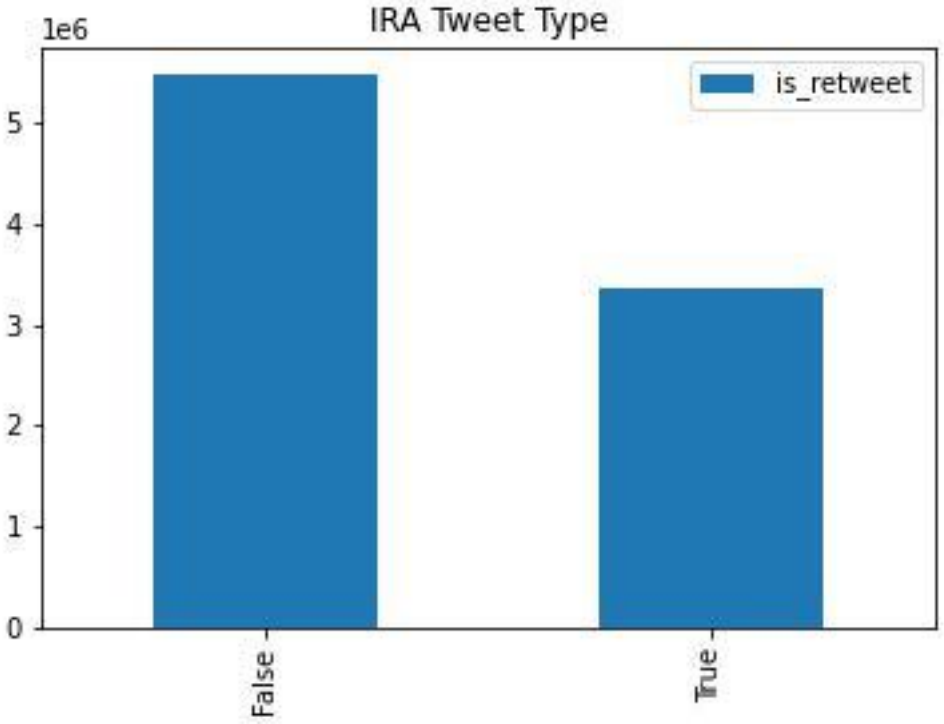
	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	return	548.1	camp	865.8	provinc	649.0	al	2277.7	#syria	329.8	€|	940.2	forc	678.9	â	628.0	syria	638.9	#syria	1103.5
1	#syria	431.8	refuge	467.9	#syria	592.3	milit	717.1	#sdf	321.1	peopl	448.9	zone	629.0	€□	319.1	#syria	467.9	#hts	886.0
2	eastern	371.7	humanitarian	386.3	russian	458.8	sham	696.1	#us	300.2	civilian	342.4	syrian	547.4	#isis	297.7	state	395.3	#terrorists	874.1
3	medic	320.1	#rukban	379.1	militari	435.9	group	644.9	terrorist	289.8	children	326.0	govern	413.3	milit	238.2	us	338.8	#idlib	824.0
4	peopl	293.1	rukban	297.1	syrian	421.4	nusra	526.6	#euphrates	261.9	terrorist	305.5	#idlib	409.4	al	227.8	use	287.2	posit	819.5
5	home	259.1	#syria	271.5	milit	374.8	tahrir	505.1	led	250.5	#syria	264.5	group	391.0	#syria	214.1	#us	286.9	#saa	790.8
6	refuge	241.3	provid	210.7	turkish	355.3	command	503.7	coalit	246.6	kill	230.5	arm	387.1	train	209.1	countri	255.0	#militants	629.1
7	#sdf	239.7	#us	204.1	weapon	342.4	#syria	378.3	#isis	223.3	syrian	210.6	#turkey	386.4	area	192.5	white	237.1	shell	500.5
8	yesterday	229.1	situat	200.6	provoc	314.3	#idlib	355.8	forc	213.2	local	202.1	#turkish	384.8	terrorist	181.8	chemic	226.7	fire	495.9
9	come	215.1	aid	195.1	prepar	314.0	field	315.3	control	204.8	us	184.4	idlib	368.4	ammunit	162.8	helmet	221.1	#hama	461.1

III. IRA

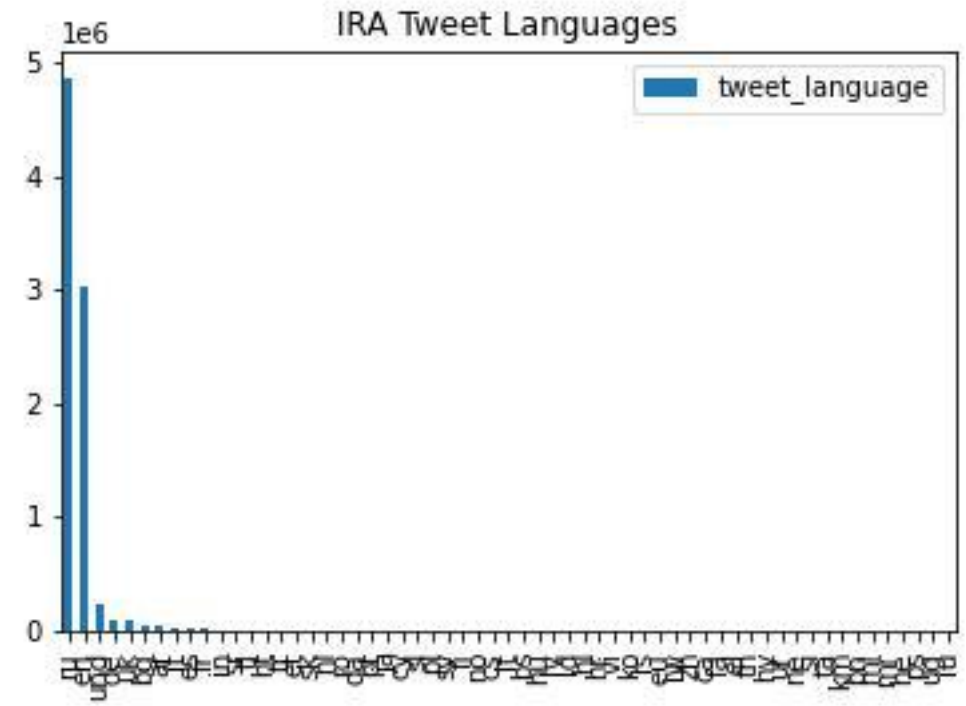
Account Creation Dates (Whole Subset)



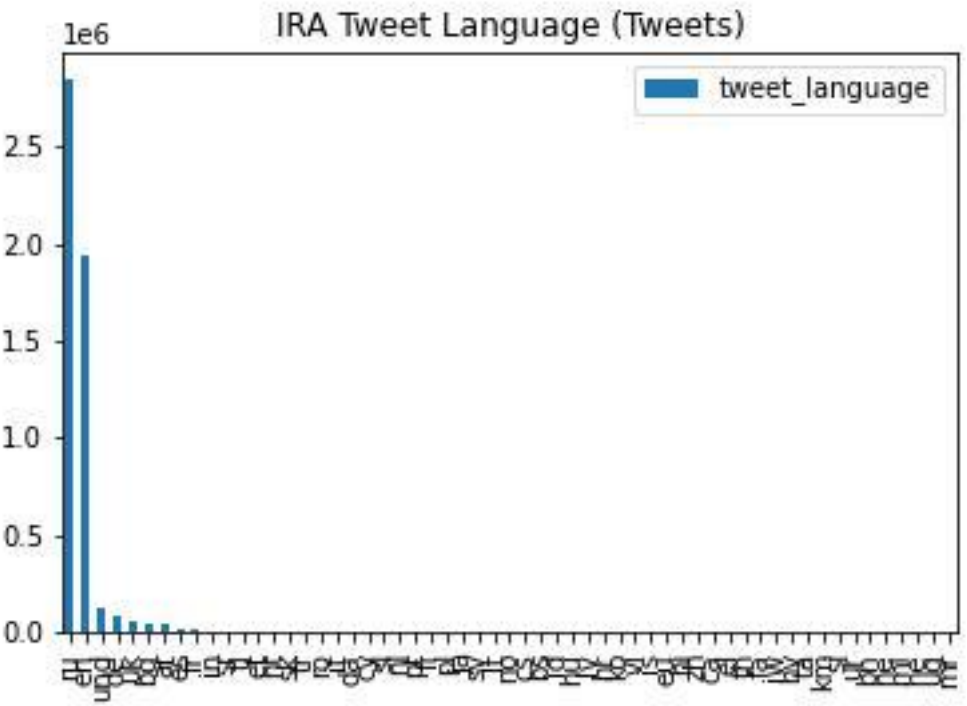
Tweet Types (Whole Subset)



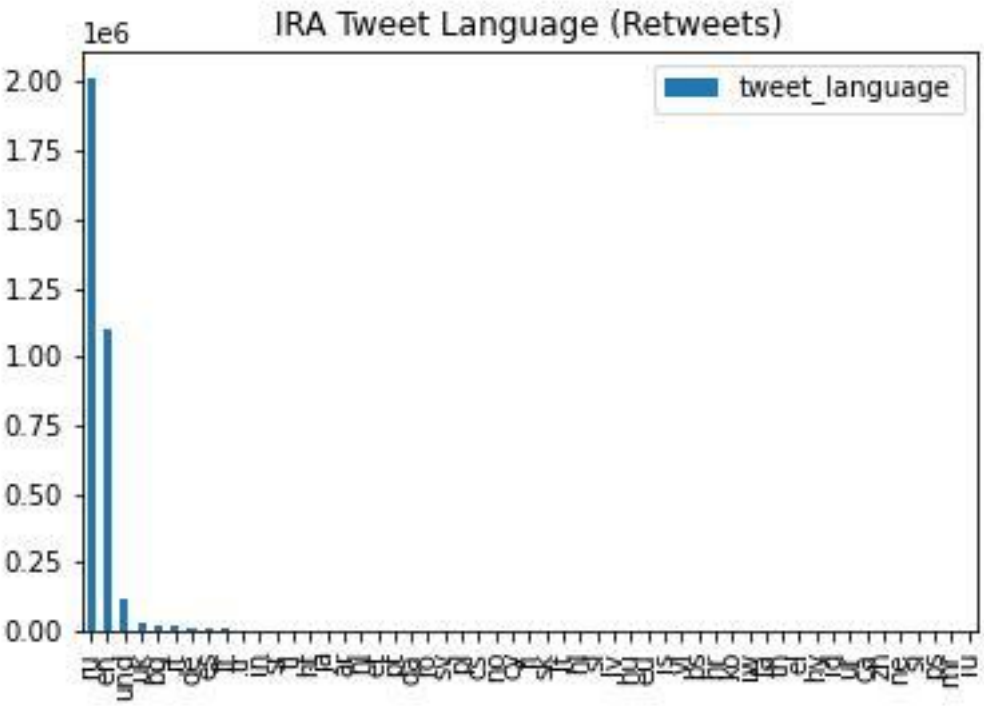
Tweet Languages (Overall, Whole Subset)



Tweet Languages (Tweets Whole Subset)



Tweet Languages (Retweets Whole Subset)

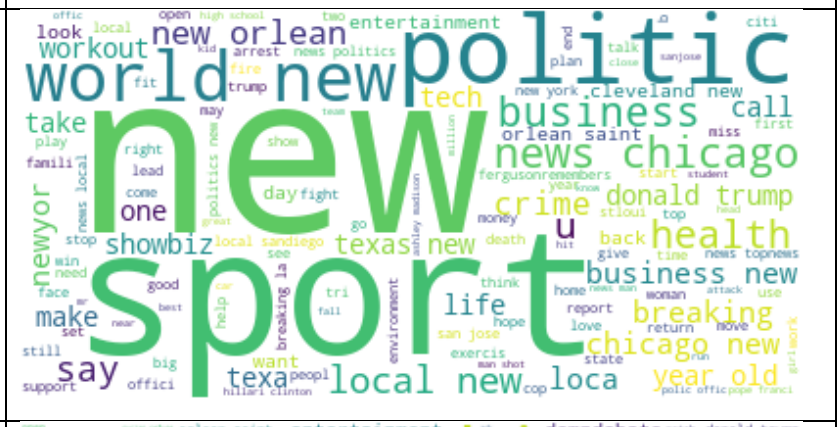


Term Frequency (Whole Subset)

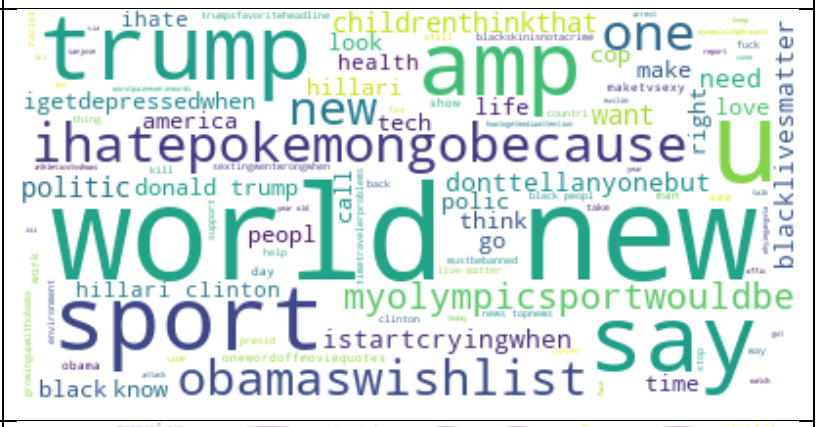
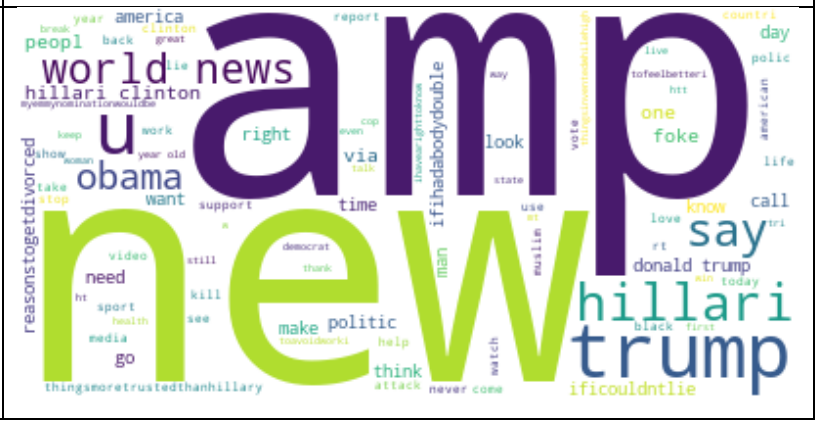
Term Frequency 2014

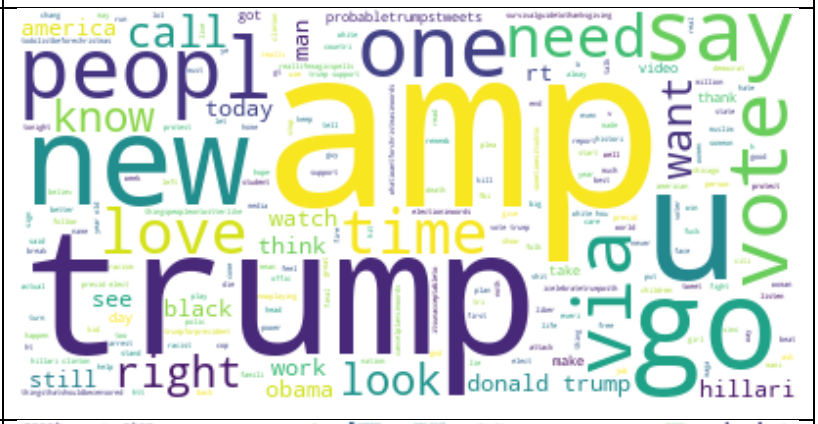
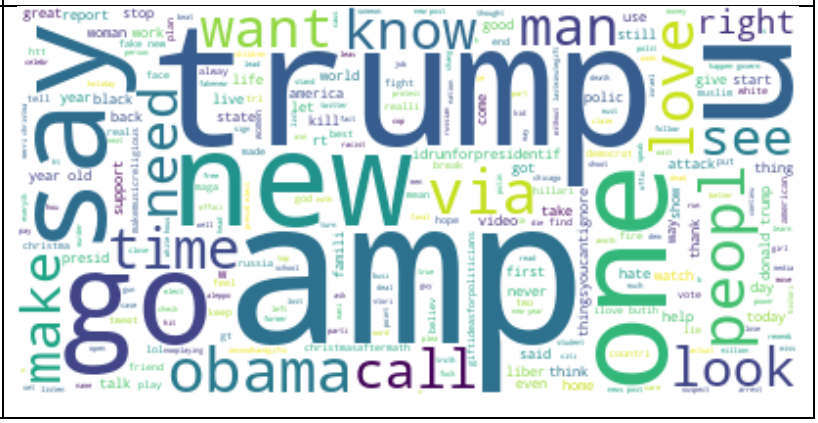
Term Frequency 2015

January	
February - March	
April - May	

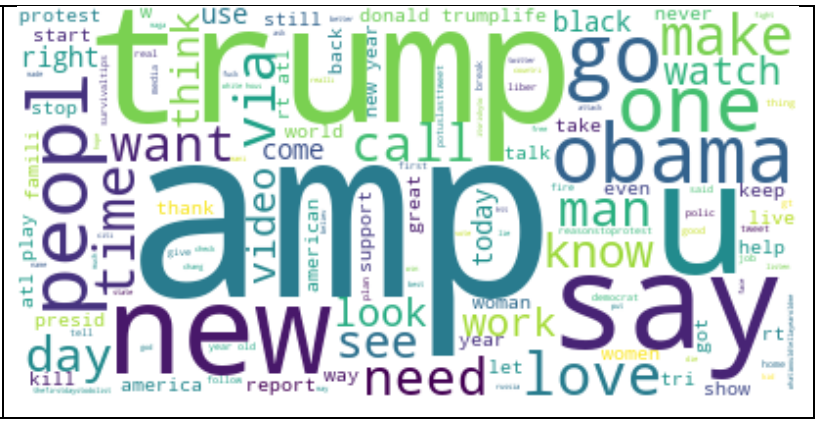
June	
July	
August - September	
October - December	

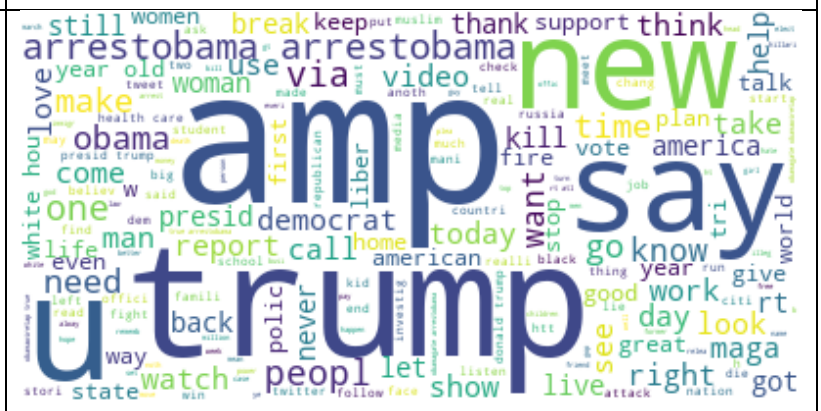
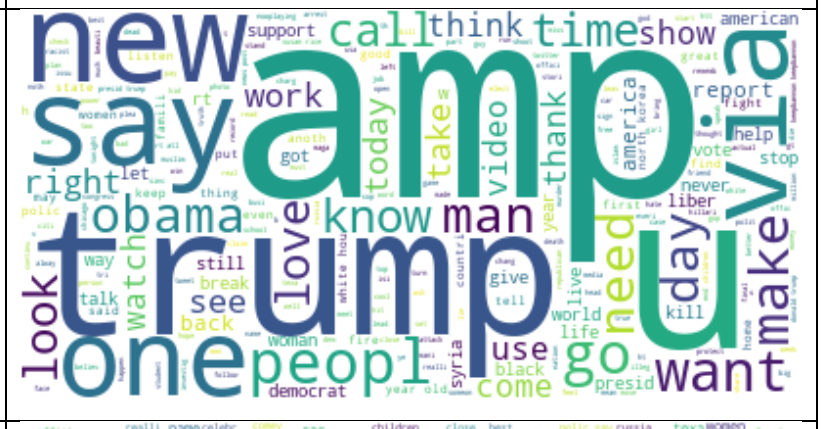
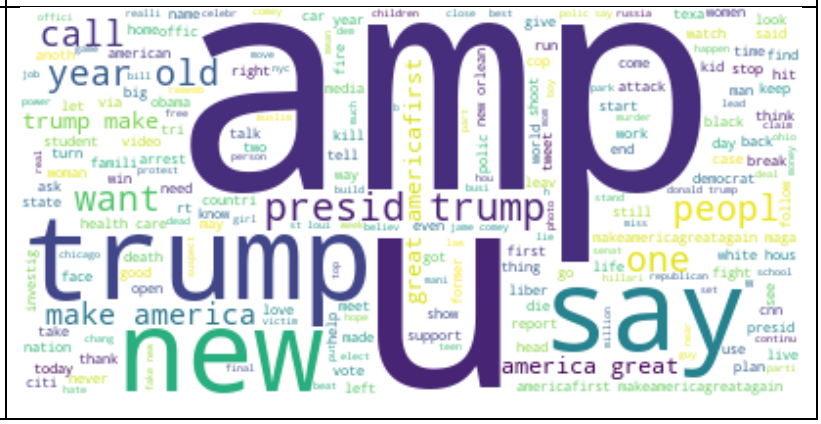
Term Frequency 2016

January - March	
April – June	
July - August	
September	

October	
November	
December	

2017

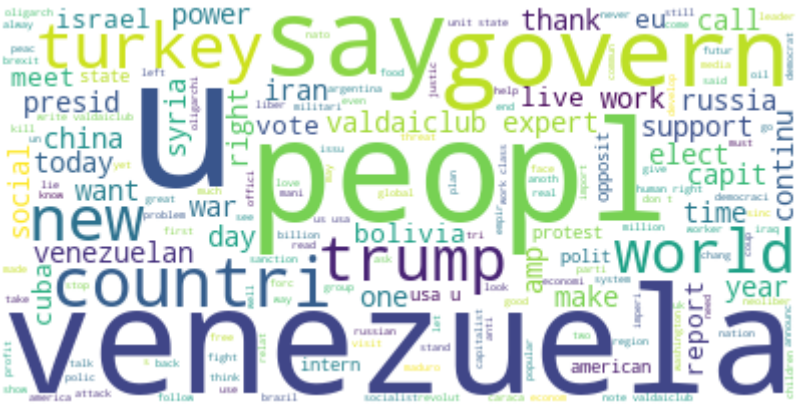
January	

February	
March	
April	
May - July	

August - December	

Term Frequency 2018

Term Frequency 2019



Term Frequency 2020

venezuela
op
ed
u
peopl
new
countri
china
coronaviru
trump
world
turkey
today
elect
american
russia
million
libya
state
social
nation
iran
one
covid
presid
govern
venezuelan
support
worker
time
say
war
day
intern
continu
capit
live
need
year
go
want
amp
first
thank
global
polit

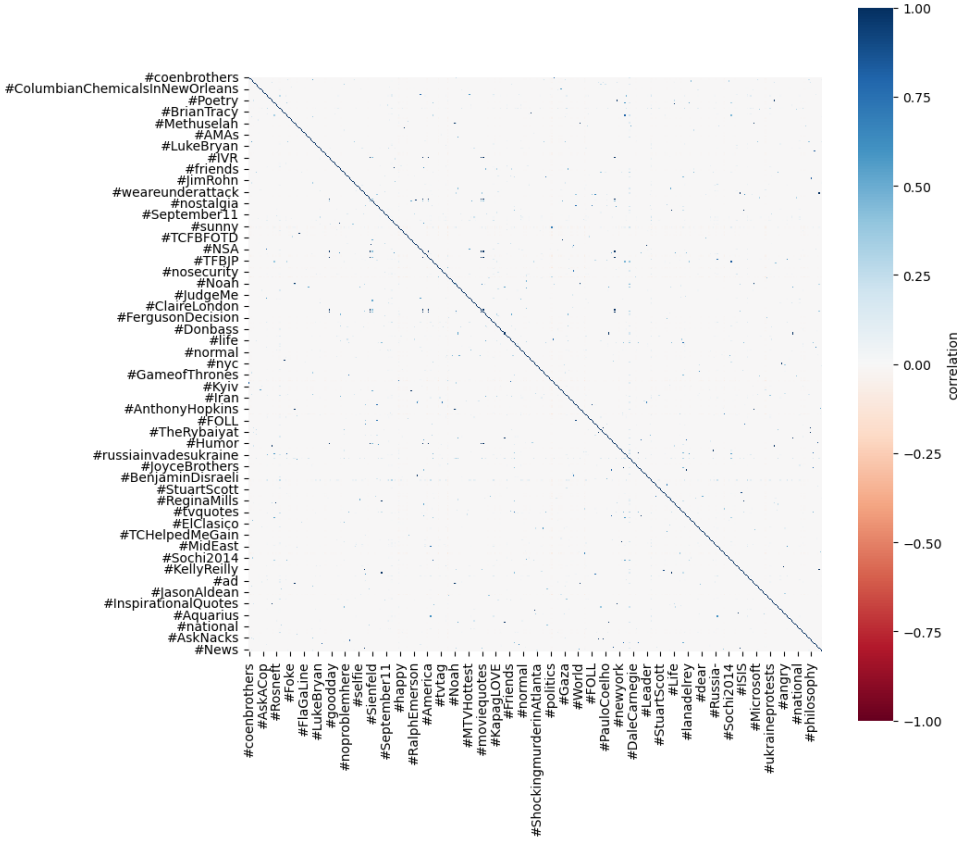
Hashtags Shared

	IRA Hashtags				
	Most Shared		Least Shared		
Grand Total	[news]	152114	[Petry, AfD, CDU, Merkel]	1	
	[sports]	97463	[CDU, AfD, SPD, ltwsaar2017, Saarland]	1	
	[politics]	73937	[ShowtimeBoxing, Heavyweights]	1	
	[СПб] (*St. Petersburg)	50716	['GE2015', 'PeopleNOTpolls']	1	
Tweets	[news]	143789	[Госдума, екб] (*State Duma, ekb)	1	
	[sports]	92953	[Готово] (*Ready)	1	
	[politics]	70848	[Virginia, Cuban]	1	
	[СПб] (*St. Petersburg)	37653	['FuerzaAxel']	1	
Retweets	[СПб] (*St. Petersburg)	13063	[TravelBanNOW]	1	
	[РоссияОбразРоссии] (*Russia, Image of Russia)	9924	[елг2016, МоскваВыбирает, РоссияВыбирает] (*edg2016, MoscowChoose, RussiaChoose)	1	
	[news]	8325	[NoDAPL, WaterIsLife, WaterIsSacred, LifeIsSacred, StandingRock]	1	
		8228	['GE2015', 'PeopleNOTpolls']	1	

Hashtags

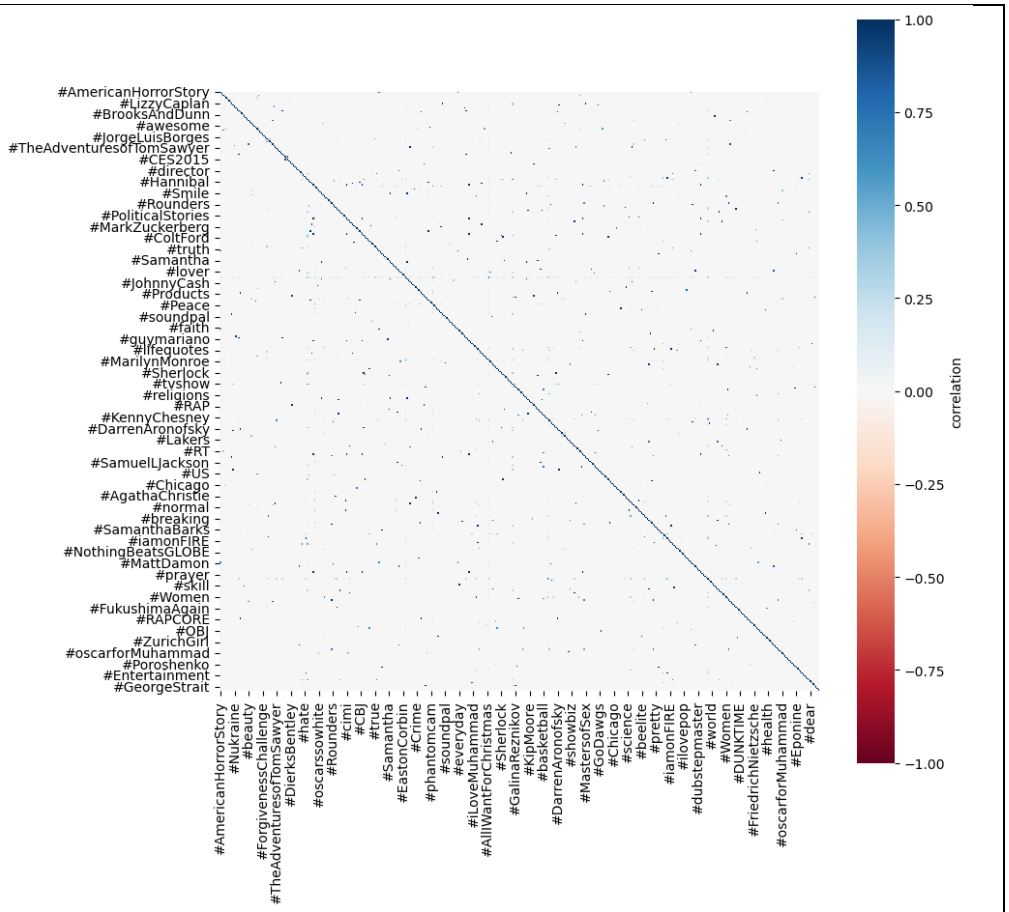
Hashtags Heatmaps

Hashtags Heatmap2014

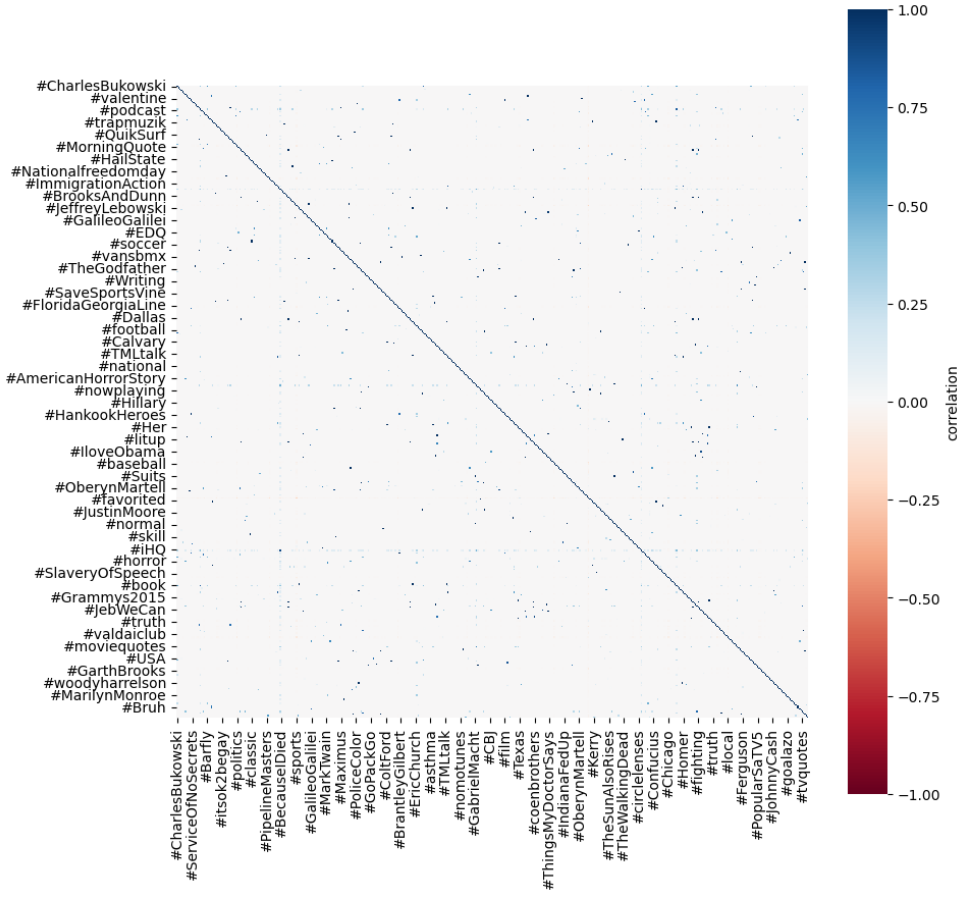


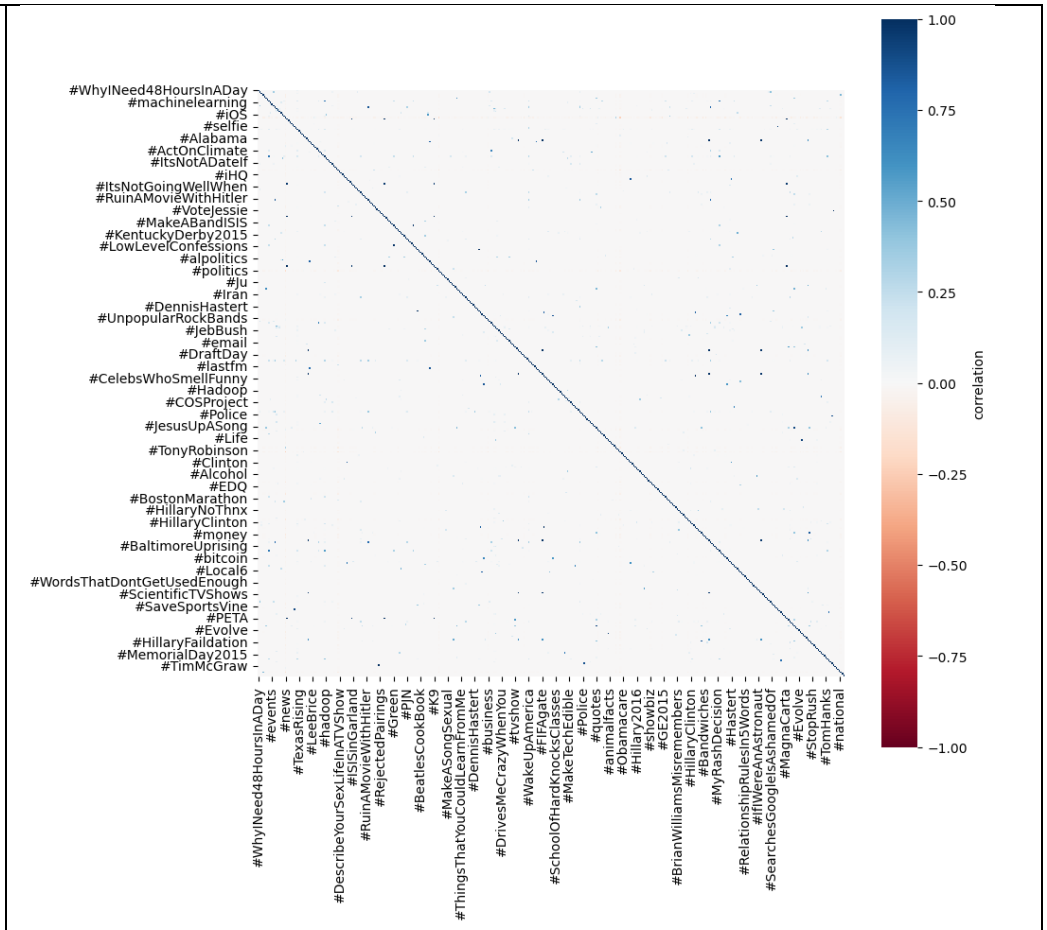
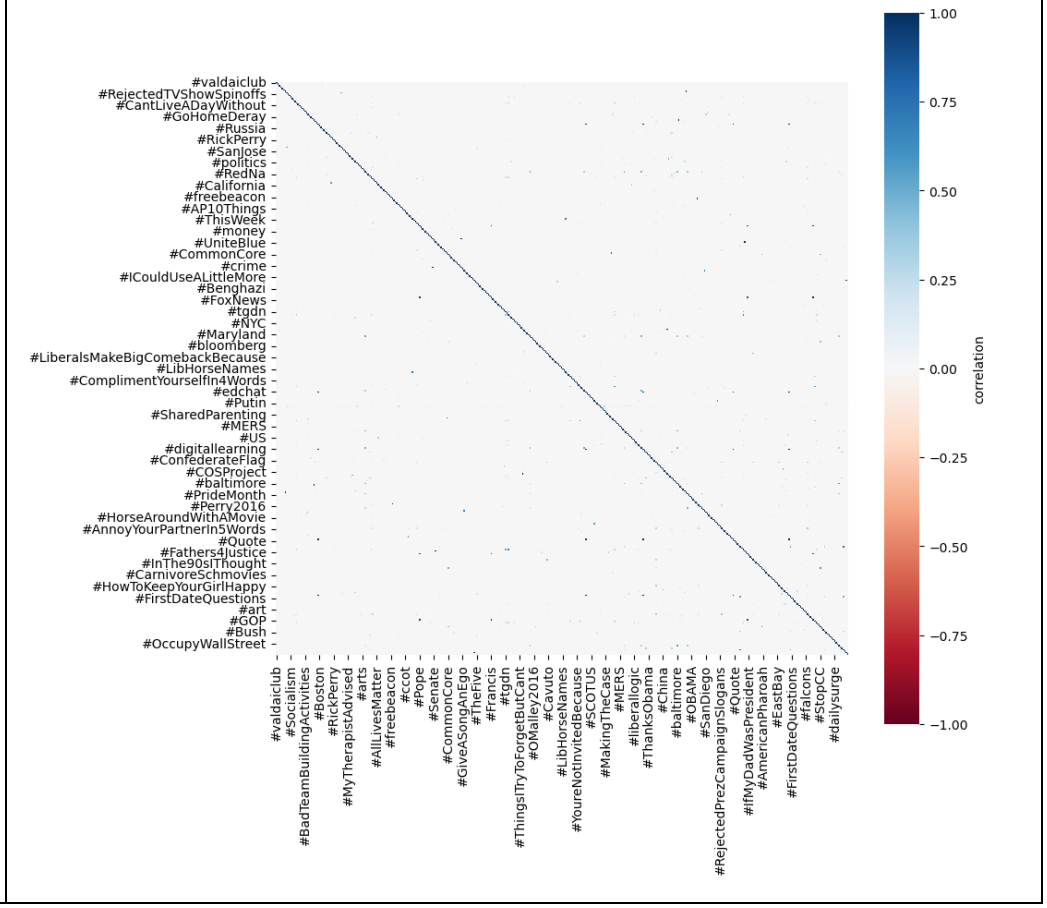
Hashtags Heatmap 2015

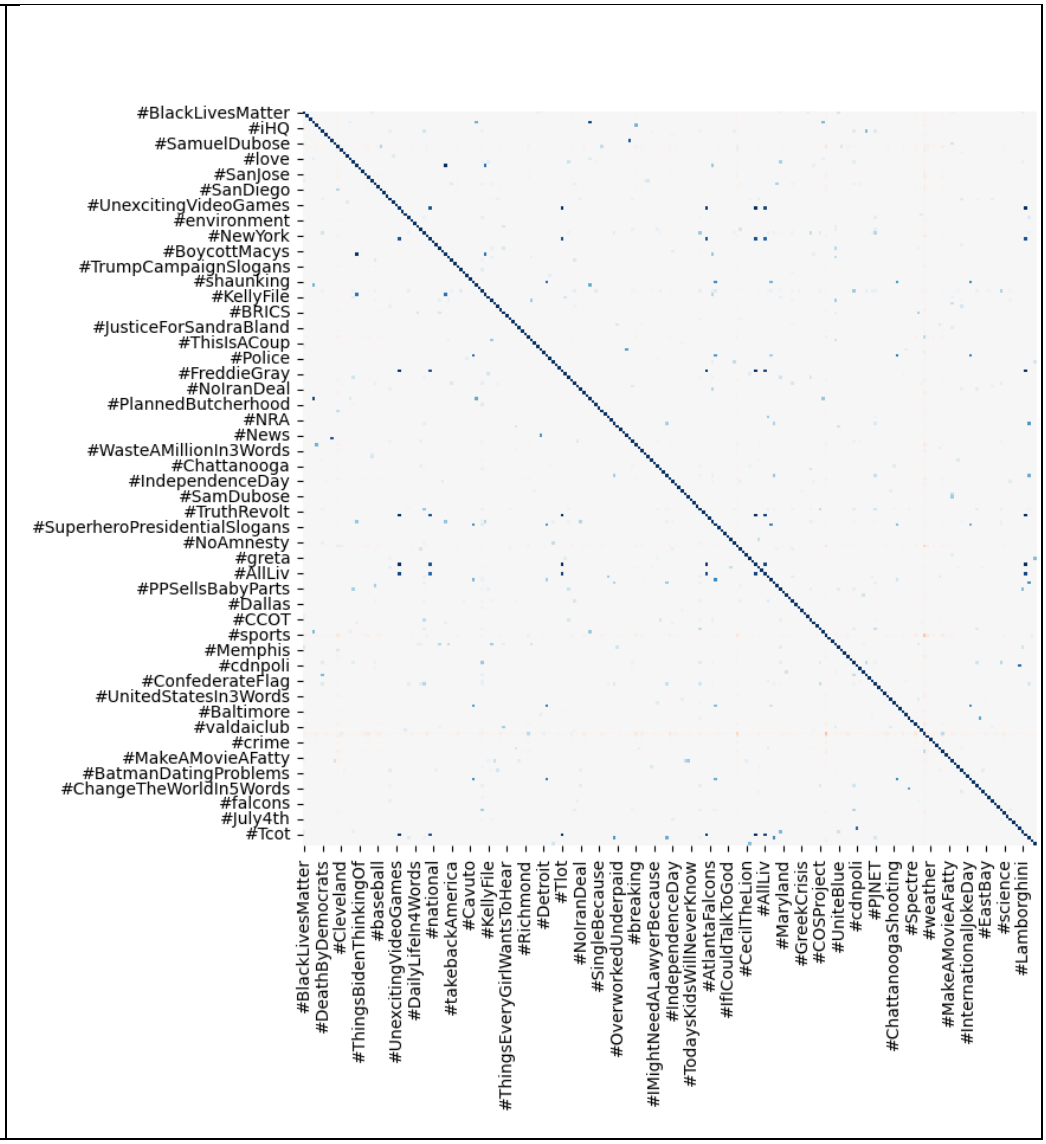
January

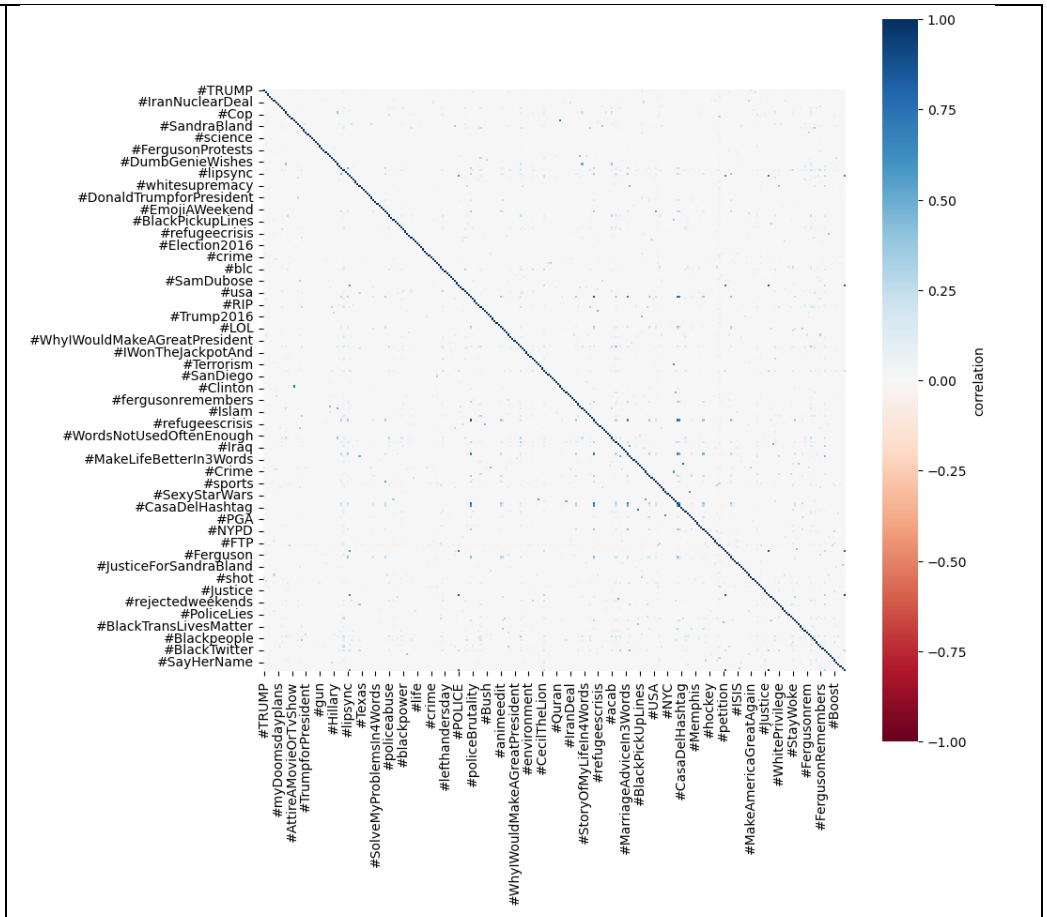
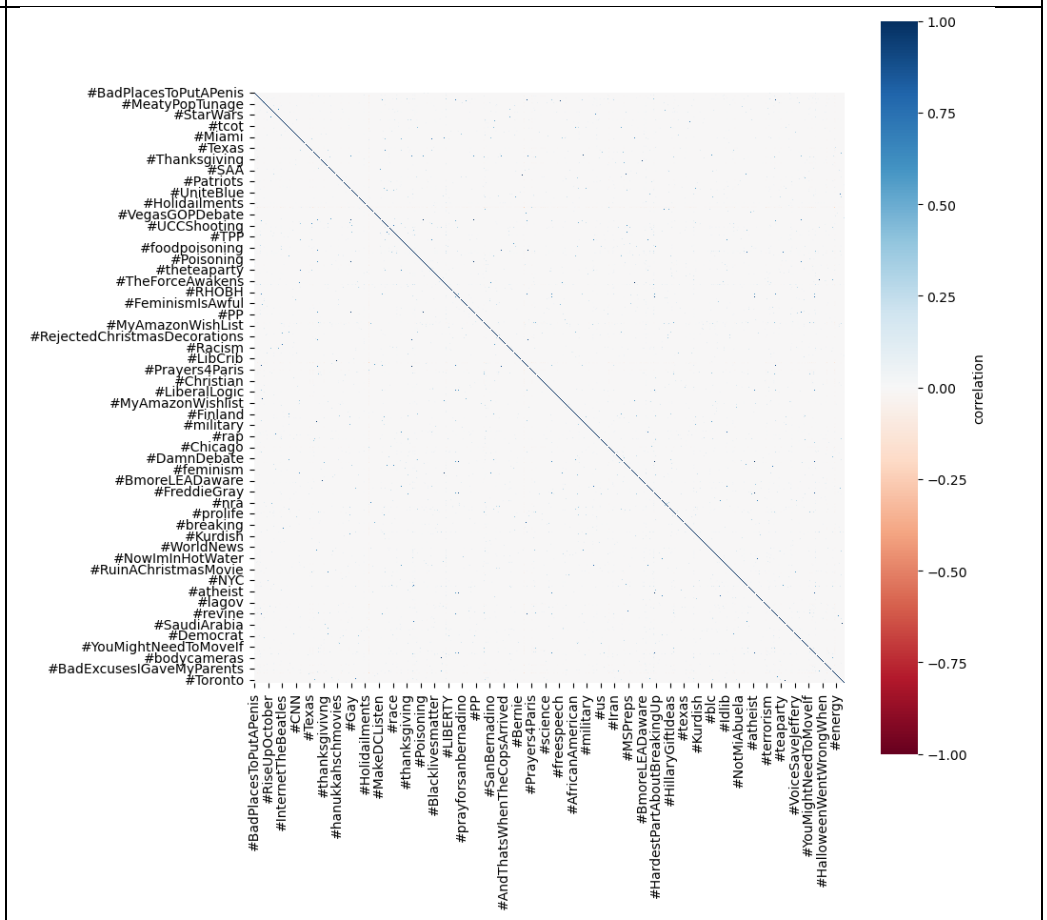


February - March

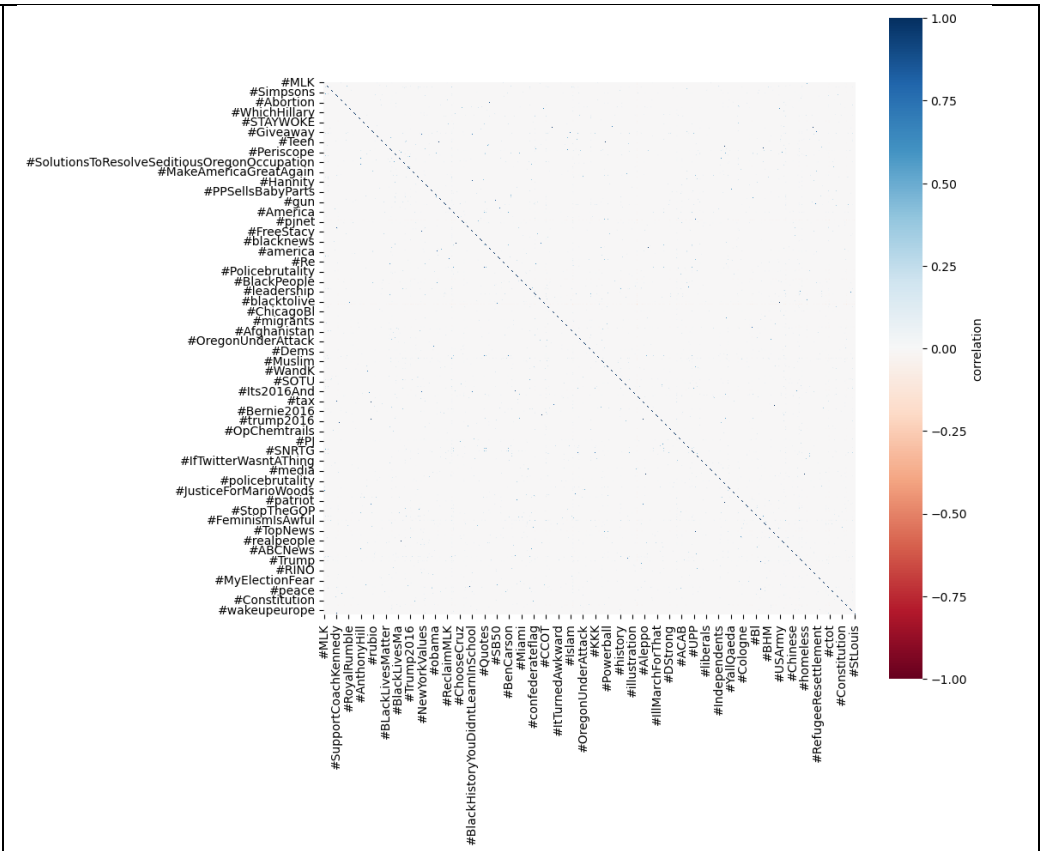


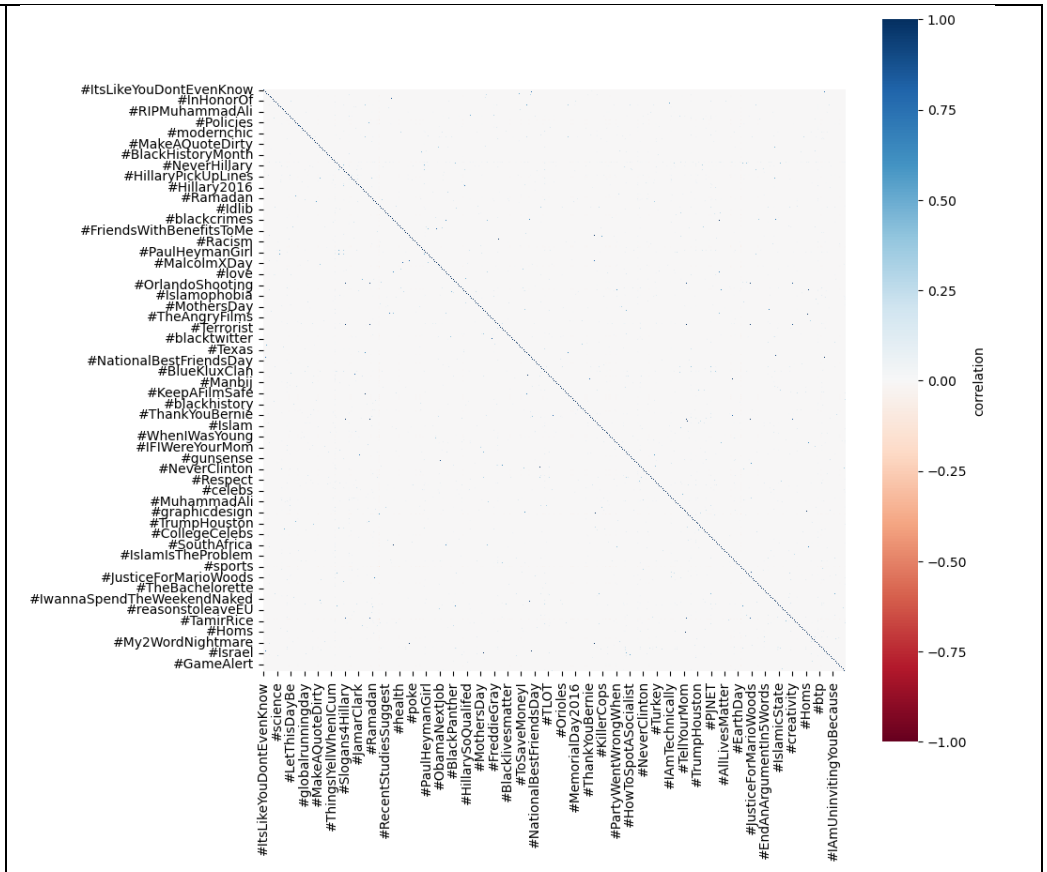
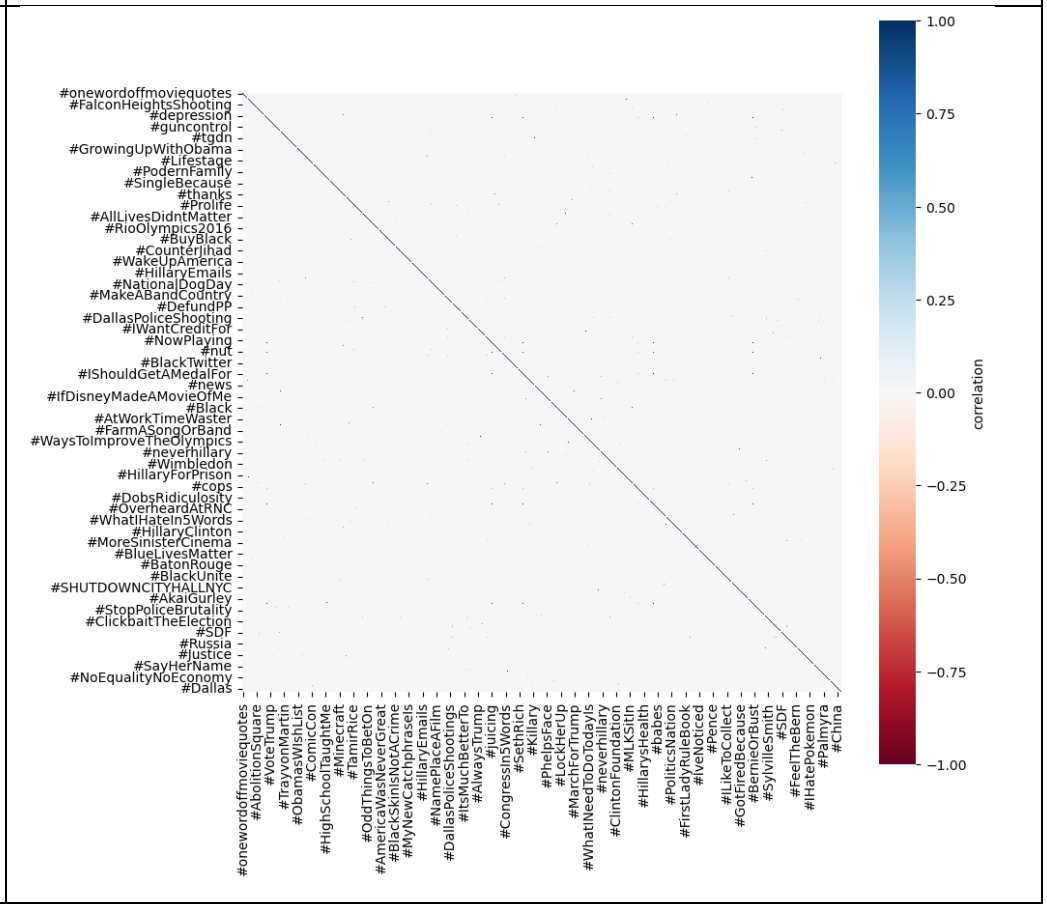
April - May	
June	

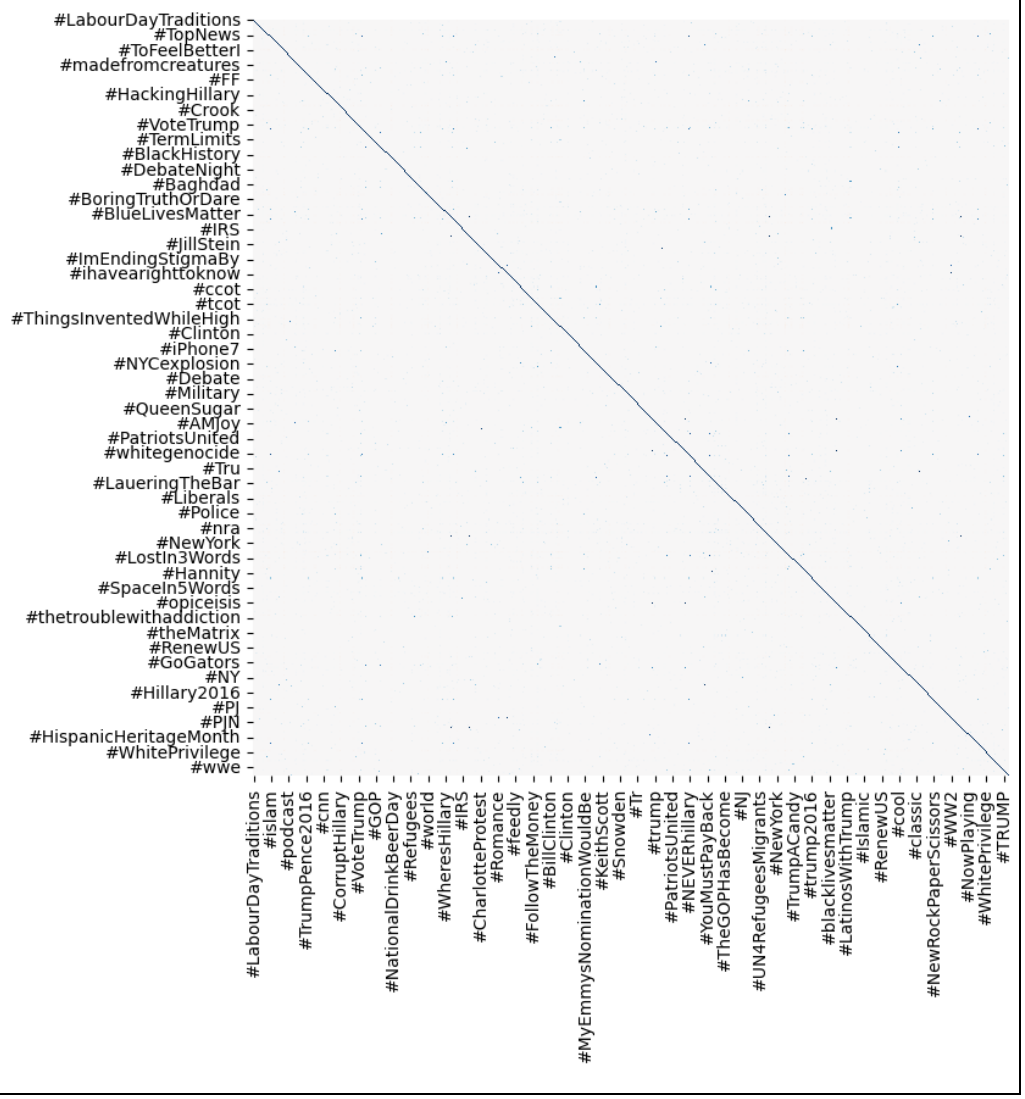
July	 #BlackLivesMatter #iHQ #SamuelDubose #love #SanJose #SanDiego #UnexcitingVideoGames #environment #NewYork #BoycottMacys #TrumpCampaignSlogans #shaunking #KellyFile #BRICS #JusticeForSandraBland #ThisIsACoup #Police #FreddieGray #NoIranDeal #PlannedButcherhood #NRA #News #WasteAMillionIn3Words #Chattanooga #IndependenceDay #SamDubose #TruthRevolt #SuperheroPresidentialSlogans #NoAmnesty #greta #AllLiv #PPSellsBabyParts #Dallas #CCOT #sports #Memphis #cdnpoli #ConfederateFlag #UnitedStatesIn3Words #Baltimore #valdaiclub #crime #MakeAMovieAFatty #BatmanDatingProblems #ChangeTheWorldIn5Words #falcons #July4th #Tcot #BlackLivesMatter #DeathByDemocrats #Cleveland #ThingsBidenThinkingOf #baseball #UnexcitingVideoGames #DailyLifeIn4Words #national #takebackAmerica #Cavuto #KellyFile #ThingsEveryGirlWantsToHear #Richmond #Detroit #Tlot #NoIranDeal #SingleBecause #OverworkedUnderpaid #breaking #IMightNeedALawyerBecause #IndependenceDay #TodaysKidsWillNeverKnow #AtlantaFalcons #IfiCouldTalkToGod #CecilTheLion #AllLiv #Maryland #GreekCrisis #COSProject #UniteBlue #cdnpoli #PJNET #ChattanoogaShooting #Spectre #weather #MakeAMovieAFatty #InternationalJokeDay #EastBay #science #Lamborghini

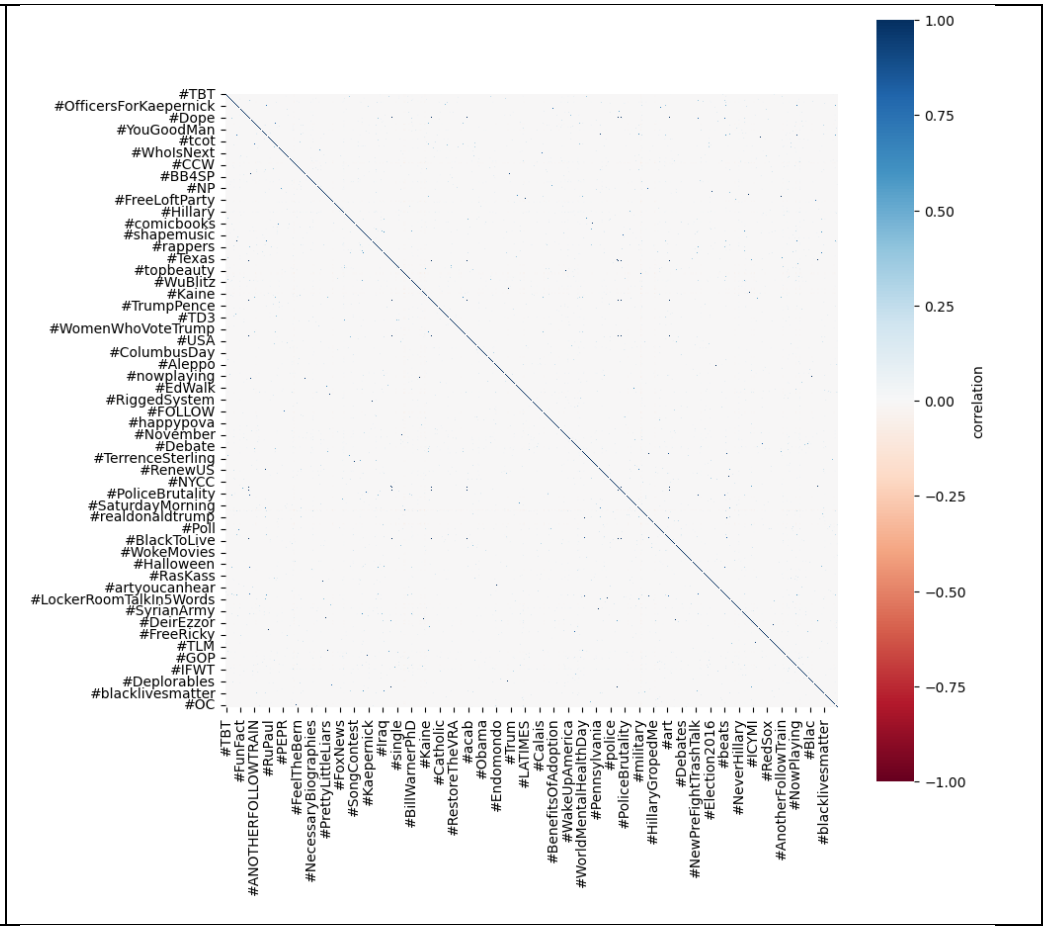
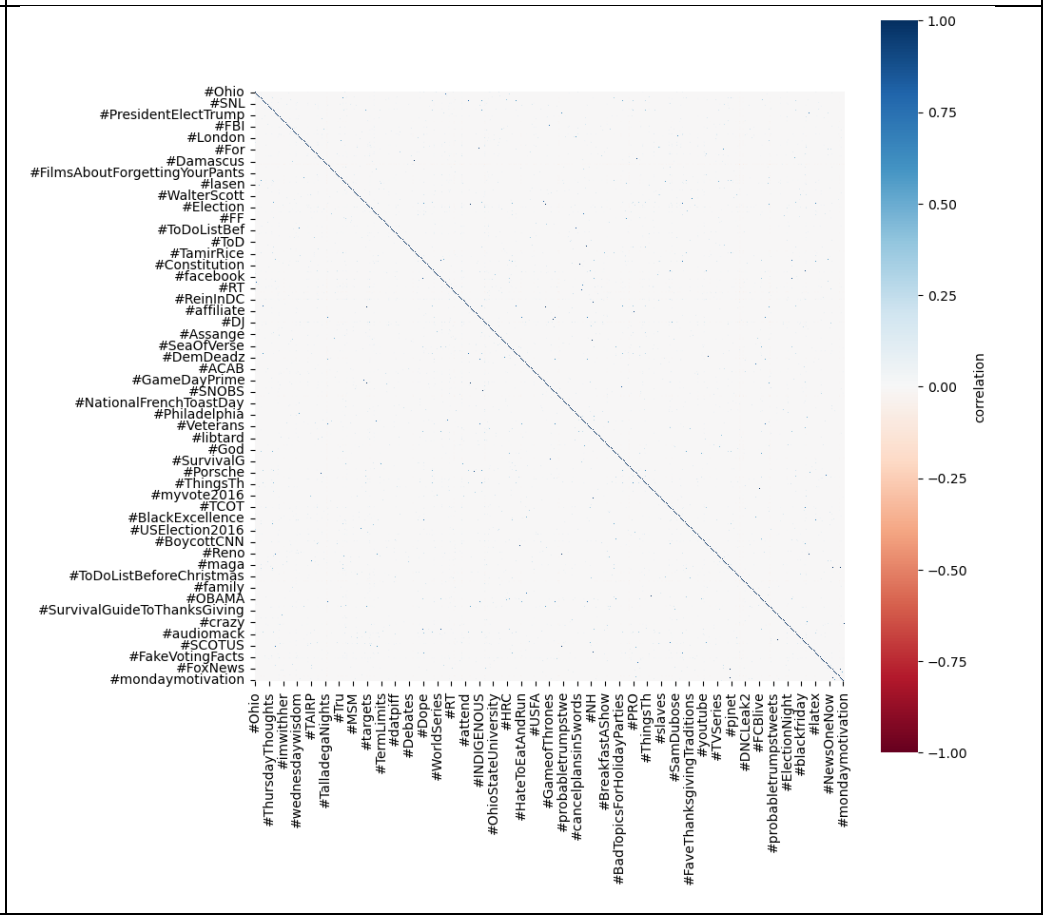
August - September	
October - December	

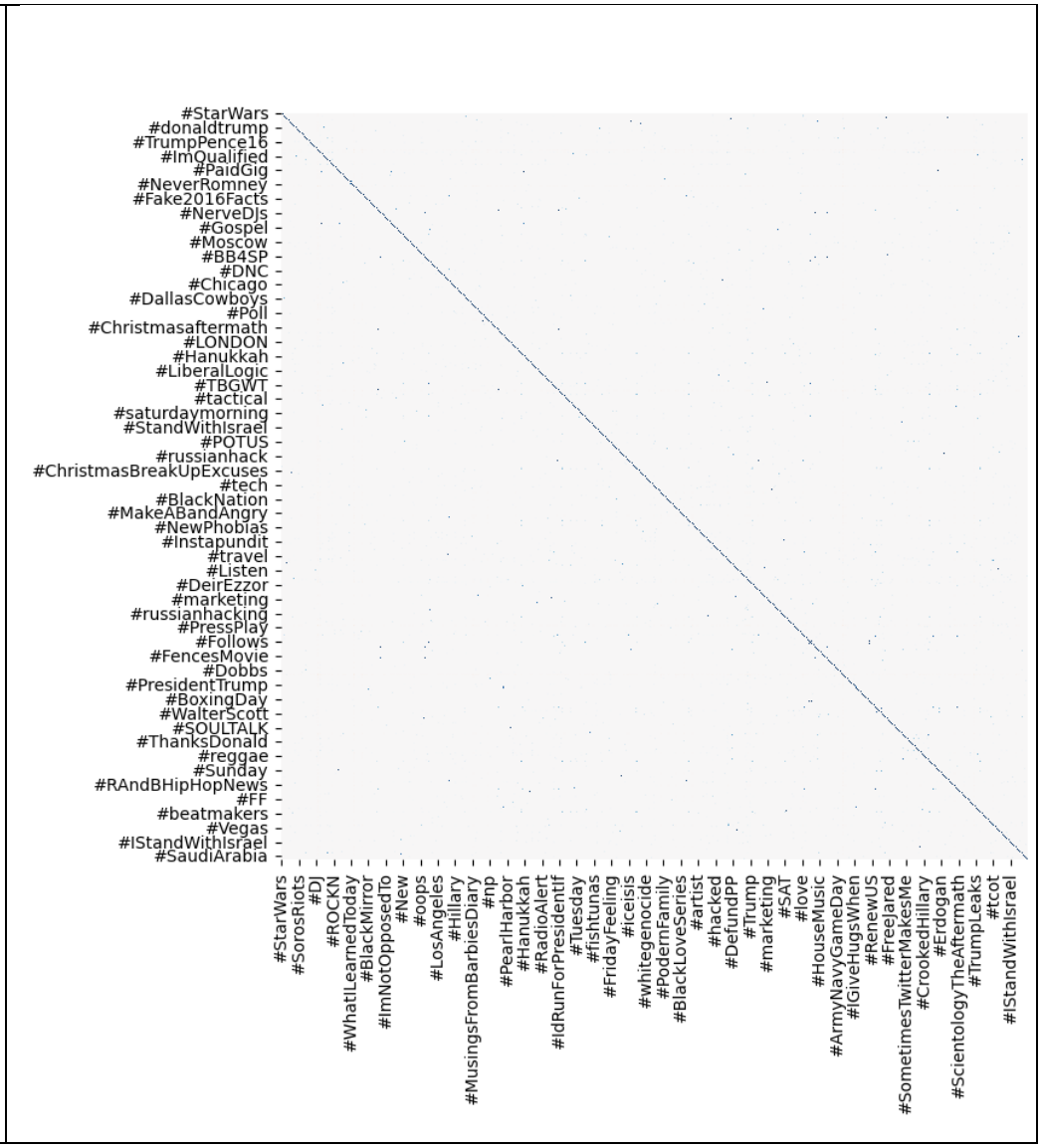
Hashtags Heatmap 2016

January - March	

April – June	
July - August	

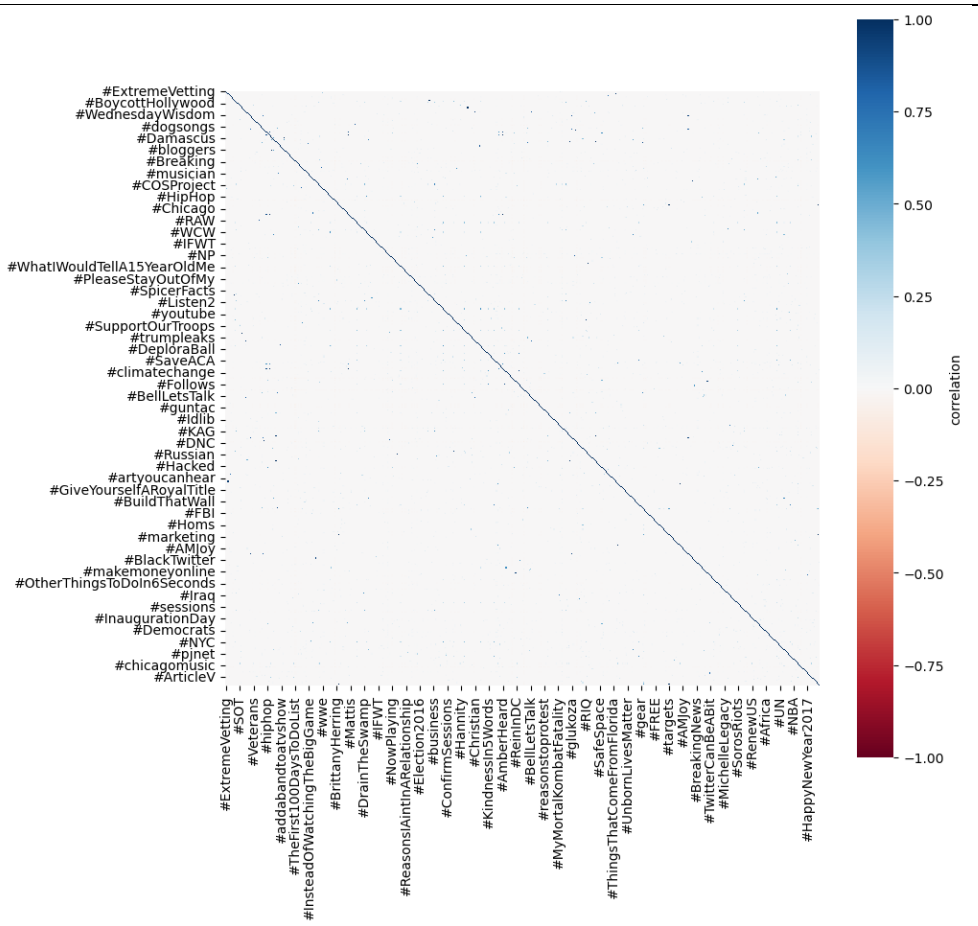
September	 #LabourDayTraditions #TopNews #ToFeelBetterI #madefromcreatures #FF #HackingHillary #Crook #VoteTrump #TermLimits #BlackHistory #DebateNight #Baghdad #BoringTruthOrDare #BlueLivesMatter #IRS #JillStein #ImEndingStigmaBy #ihavearighttoknow #ccot #tcot #ThingsInventedWhileHigh #Clinton #iPhone7 #NYCexplosion #Debate #Military #QueenSugar #AMJoy #PatriotsUnited #whitegenocide #Tru #LaueringTheBar #Liberals #Police #nra #NewYork #LostIn3Words #Hannity #SpaceIn5Words #opiceisis #thetroublewithaddiction #theMatrix #RenewUS #GoGators #NY #Hillary2016 #PJ #PJN #HispanicHeritageMonth #WhitePrivilege #wwe #LabourDayTraditions #islam #podcast #TrumpPence2016 #cnn #CorruptHillary #VoteTrump #GOP #NationalDrinkBeerDay #Refugees #world #WheresHillary #IRS #CharlotteProtest #Romance #feedly #FollowTheMoney #BillClinton #Clinton #MyEmmysNominationWouldBe #KeithScott #Snowden #tr #trump #PatriotsUnited #NEVERhillary #YouMustPayBack #TheGOPHasBecome #NJ #UN4RefugeesMigrants #NewYork #TrumpACandy #trump2016 #blacklivesmatter #LatinosWithTrump #Islamic #RenewUS #cool #classic #NewRockPaperScissors #WW2 #NowPlaying #WhitePrivilege #TRUMP

October	
November	

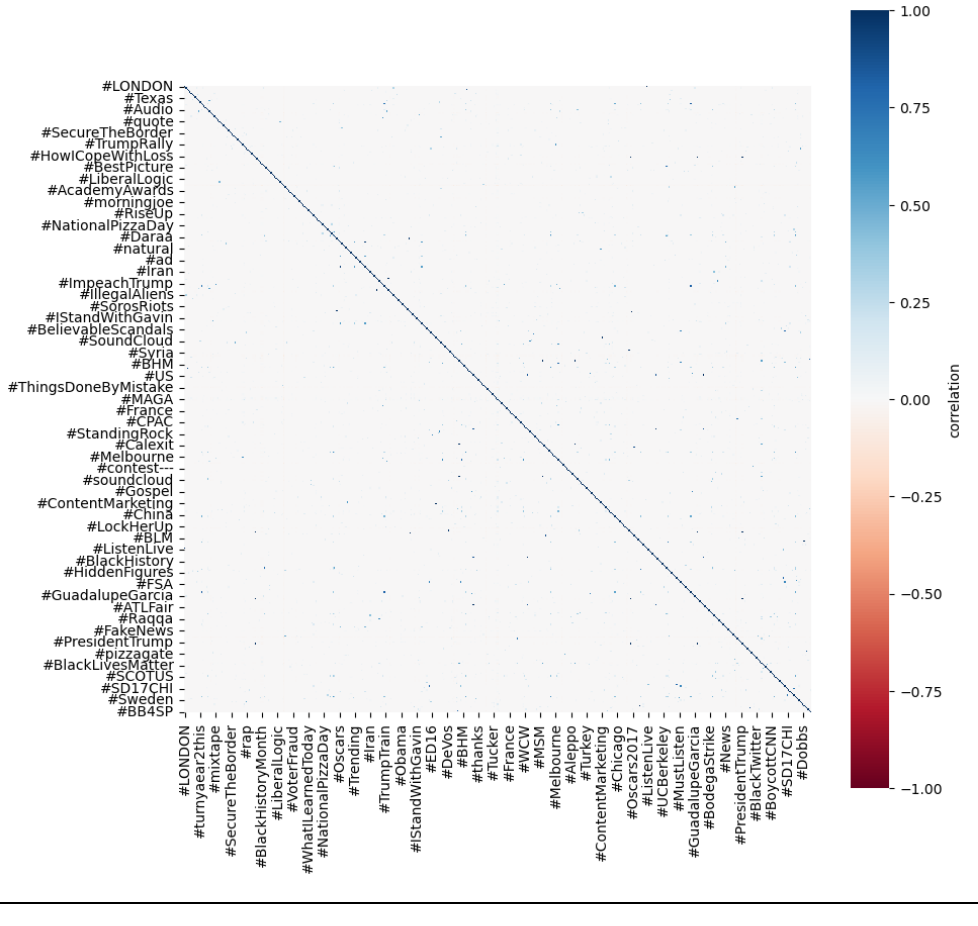
December	

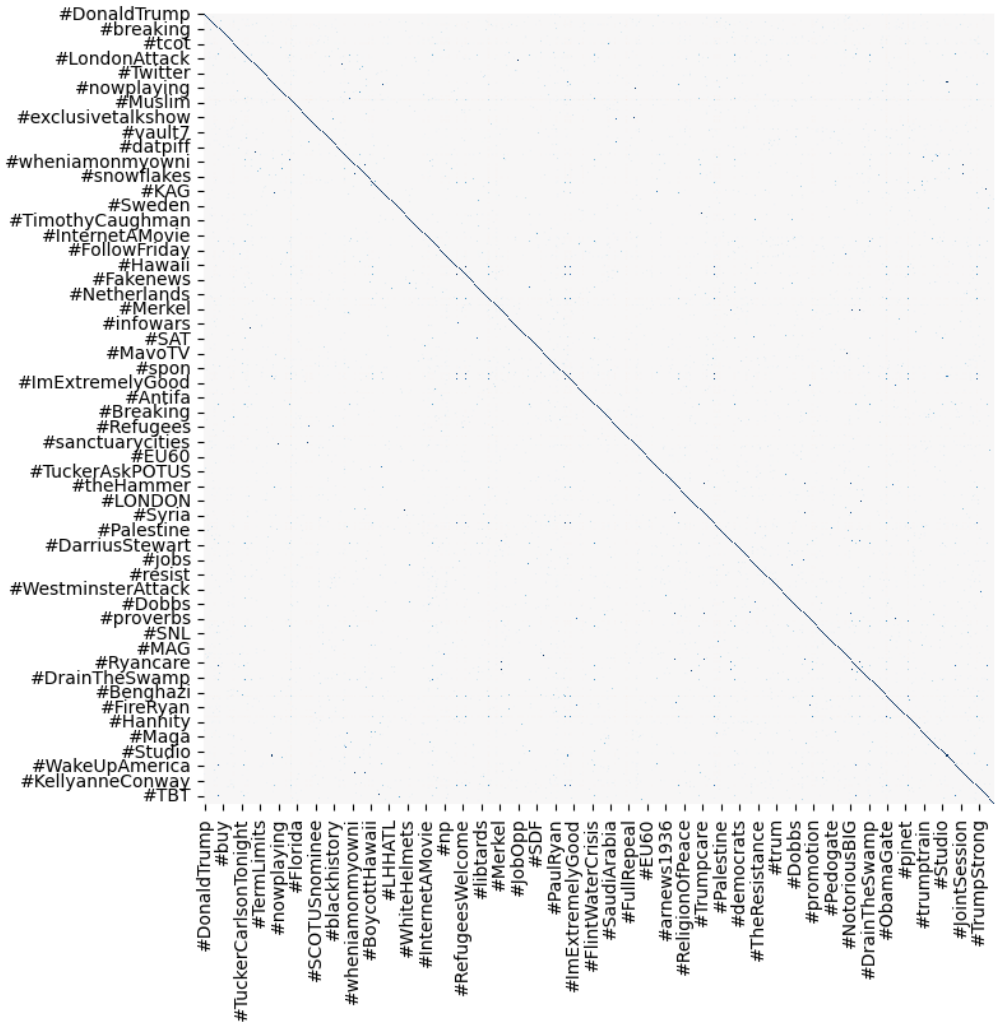
2017

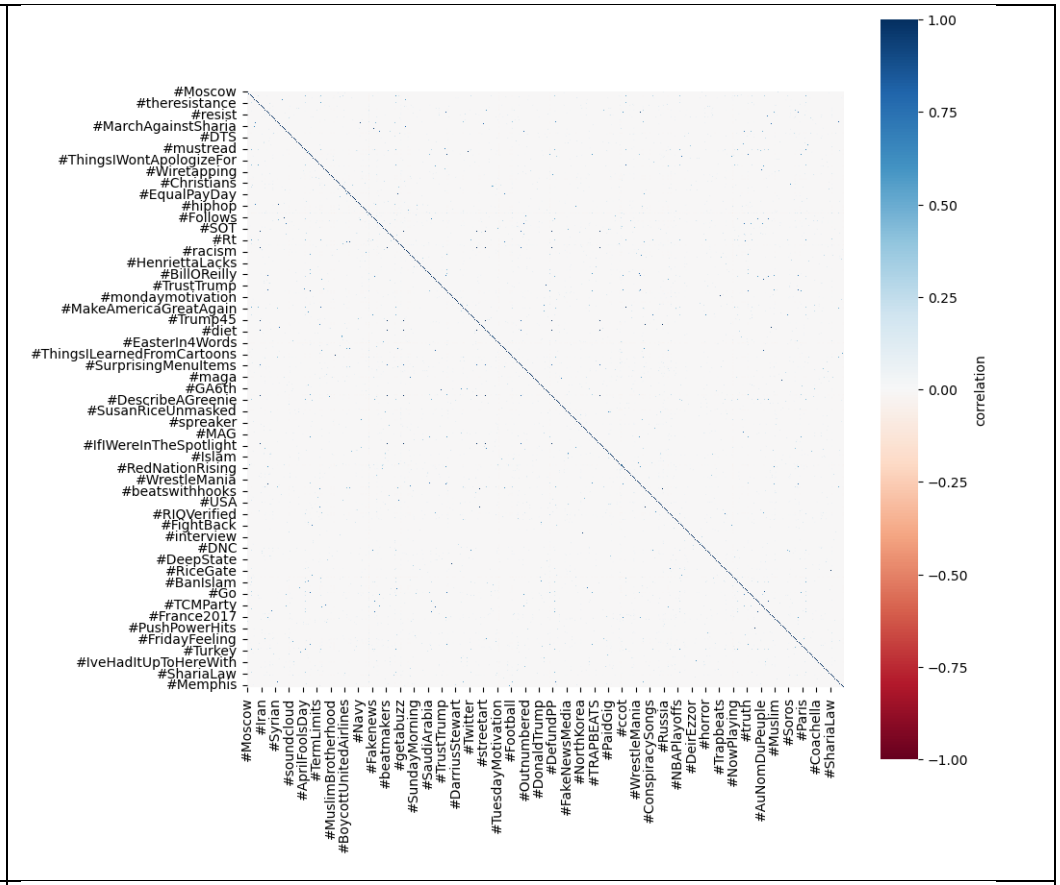
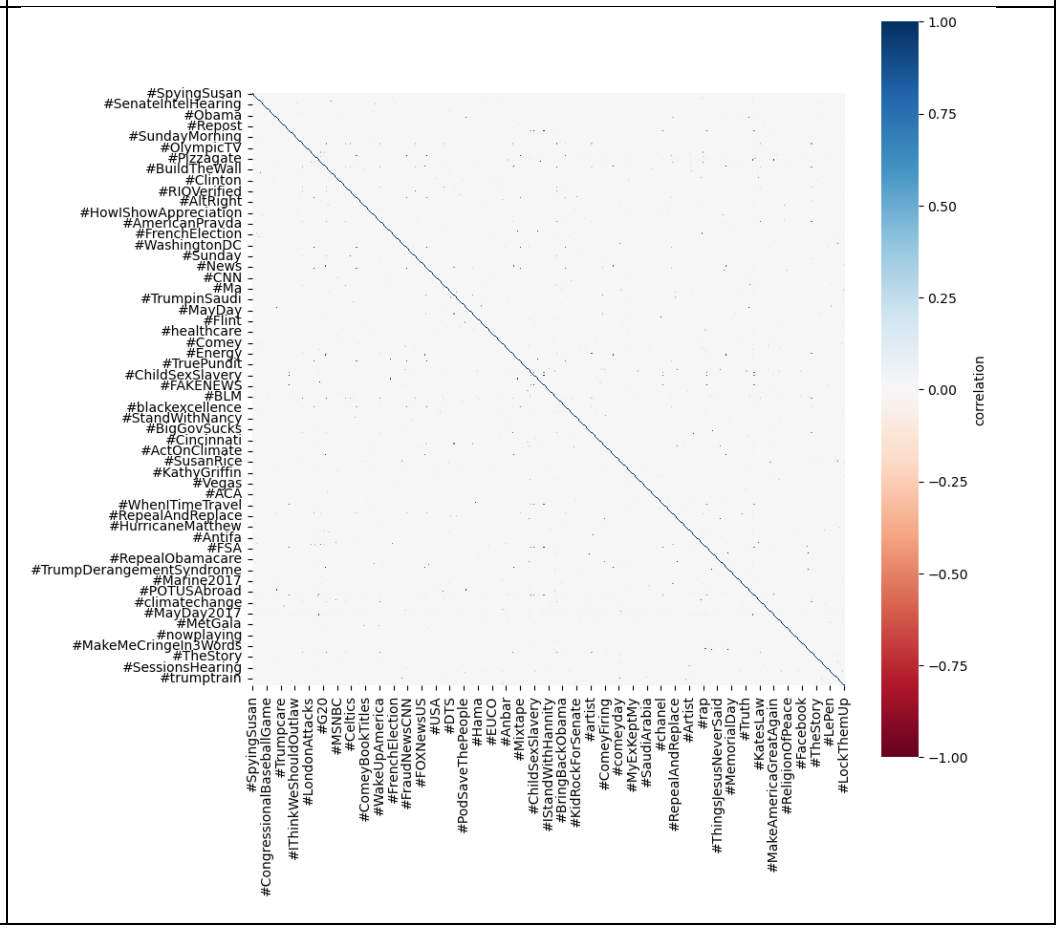
January

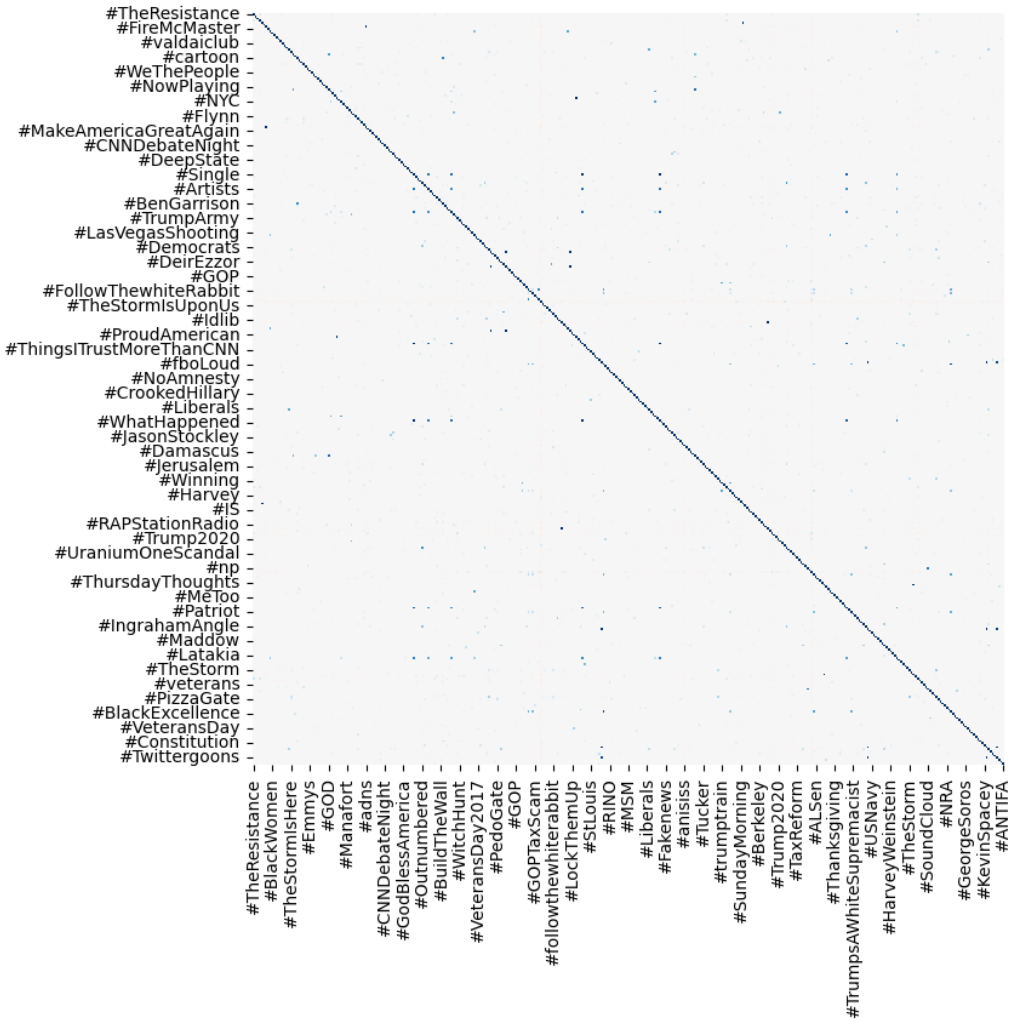


February

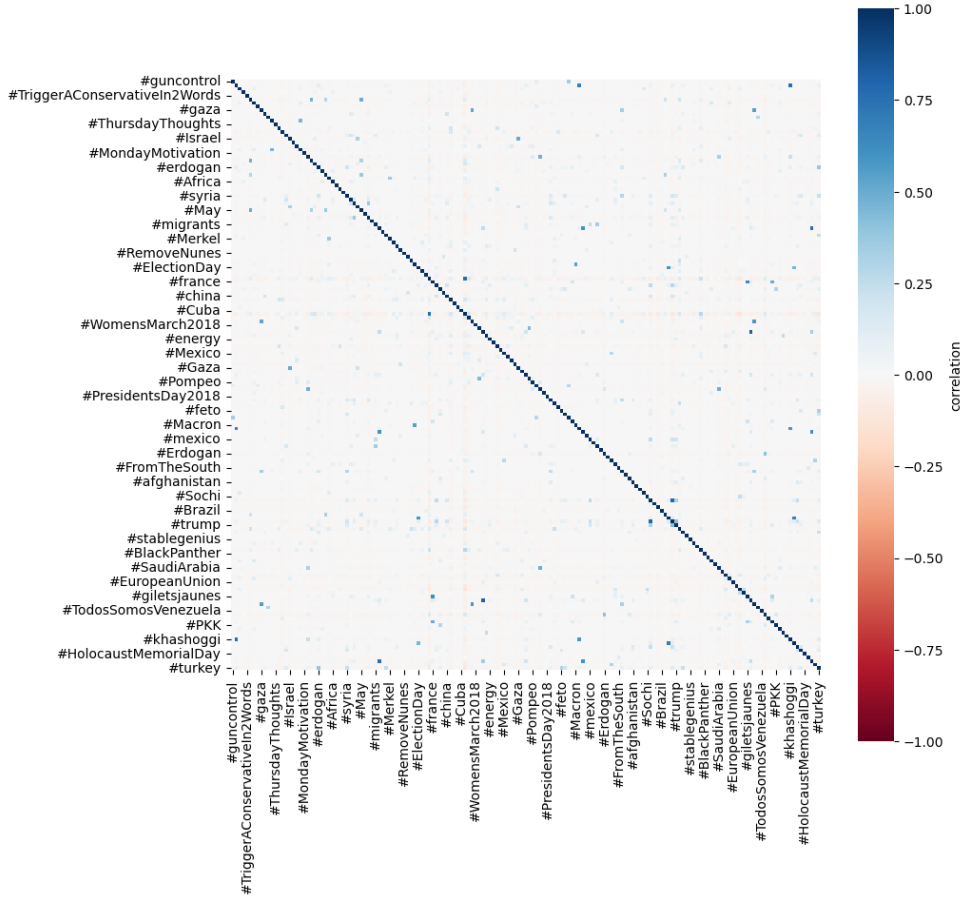


March	 #DonaldTrump #breaking #tcot #LondonAttack #Twitter #nowplaying #Muslim #exclusivetalkshow #vault7 #datpiff #wheniamonmyowni #snowflakes #KAG #Sweden #TimothyCaughman #InternetAMovie #FollowFriday #Hawaii #Fakenews #Netherlands #Merkel #infowars #SAT #MavoTV #spon #ImExtremelyGood #Antifa #Breaking #Refugees #sanctuarycities #EU60 #TuckerAskPOTUS #theHammer #LONDON #Syria #Palestine #DarriusStewart #jobs #resist #WestminsterAttack #Dobbs #proverbs #SNL #MAG #Ryancare #DrainTheSwamp #Benghazi #FireRyan #Hannity #Maga #Studio #WakeUpAmerica #KellyanneConway #TBT #DonaldTrump #buy #TuckerCarlsonTonight #TermLimits #nowplaying #Florida #SCOTUSnominee #blackhistory #wheniamonmyowni #BoycottHawaii #LHHATL #WhiteHelmets #InternetAMovie #np #RefugeesWelcome #libtards #Merkel #JobOpp #SDF #PaulRyan #ImExtremelyGood #FlintWaterCrisis #SaudiArabia #FullRepeal #EU60 #arnews1936 #ReligionOfPeace #Trumpcare #Palestine #democrats #TheResistance #trum #Dobbs #promotion #Pedogate #NotoriousBIG #DrainTheSwamp #ObamaGate #pjnet #trumptrain #Studio #JointSession #TrumpStrong

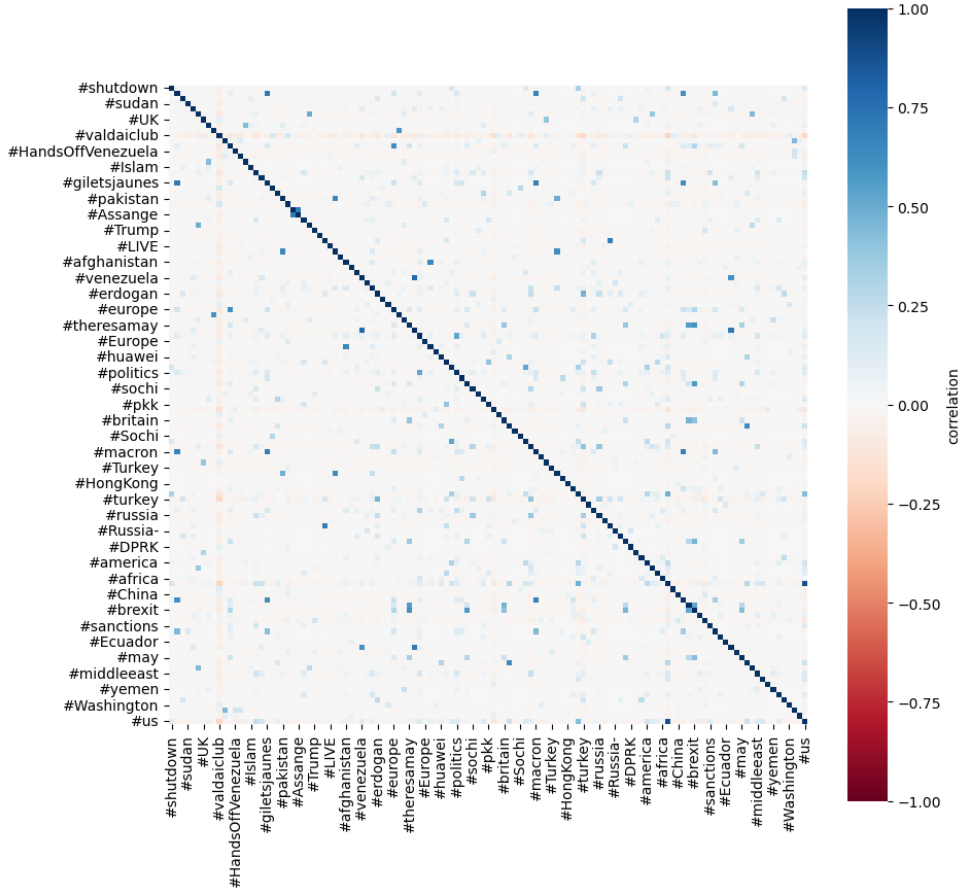
April	
May - July	

August - December	

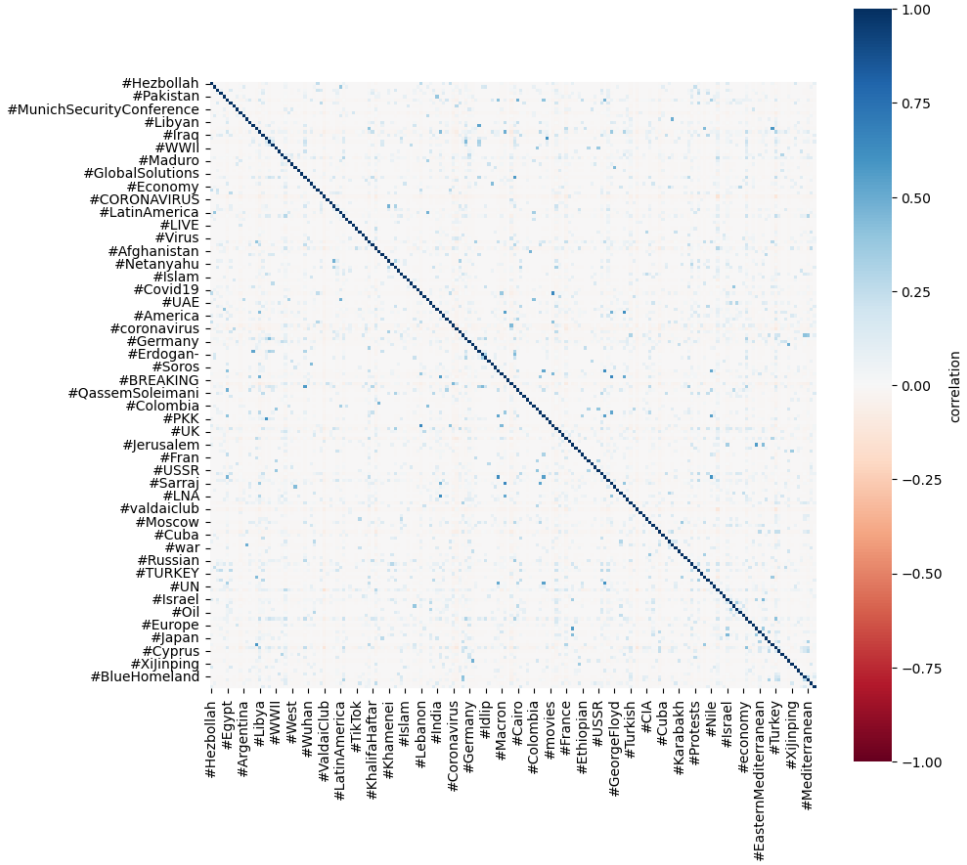
Hashtags Heatmap 2018



Hashtags Heatmap 2019



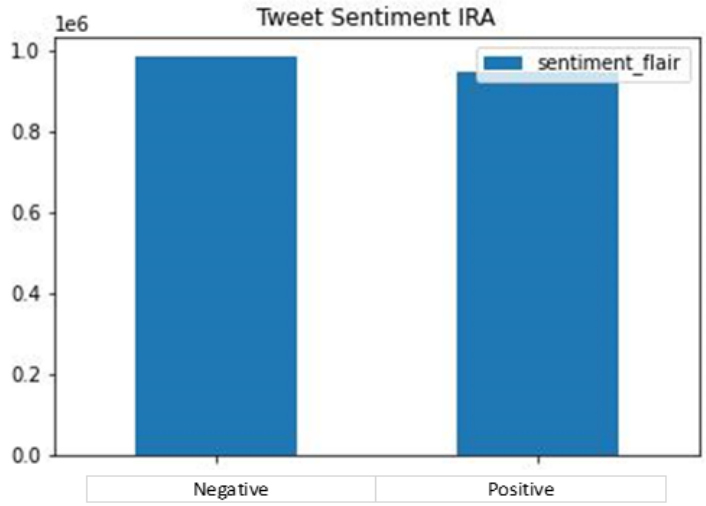
Hashtags Heatmap 2020


#Hezbollah
#Pakistan
#MunichSecurityConference
#Libyan
#Iraq
#WWII
#Maduro
#GlobalSolutions
#Economy
#CORONAVIRUS
#LatinAmerica
#LIVE
#Virus
#Afghanistan
#Netanyahu
#Islam
#Covid19
#UAE
#America
#coronavirus
#Germany
#Erdogan-
#Soros
#BREAKING
#QassemSoleimani
#Colombia
#PKK
#UK
#Jerusalem
#Fran
#USSR
#Sarraj
#LNA
#valdaiclub
#Moscow
#Cuba
#war
#Russian
#TURKEY
#UN
#Israel
#Oil
#Europe
#Japan
#Cyprus
#Xijinping
#BlueHomeland
#Hezbollah
#Egypt
#Argentina
#Libya
#WWII
#West
#Wuhan
#ValdaiClub
#LatinAmerica
#TnkTok
#KhalifaHaftar
#Khamenei
#Islam
#Lebanon
#India
#Coronavirus
#Germany
#Idlip
#Macron
#Cairo
#Colombia
#movies
#France
#Ethiopian
#USSR
#GeorgeFloyd
#Turkish
#CIA
#Cuba
#Karabakh
#Protests
#Nile
#Israel
#economy
#EasternMediterranean
#Turkey
#Xijinping
#Mediterranean
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
correlation

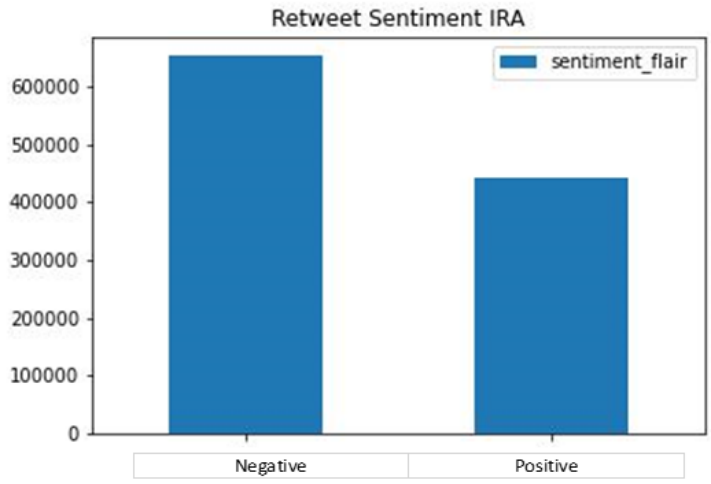
URLS Shared (Whole Subset)

	IRA URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	[http://1063.mobi]	4403	Not Active
	[http://kiev-news.com/13876-detskaya-bolnica-v-donecke-popala-pod-obstrel.html]	2867	Car Selling Website
	[http://kievsmi.net/novosti/politics/47097-klimkin-peregovory-ukrainy-i-rossii-idut-neprosto.html]	2801	Unavailable
Tweets	[http://bit.ly/1D3LvAc]	2235	Unavailable
	[http://kiev-news.com/13876-detskaya-bolnica-v-donecke-popala-pod-obstrel.html]	1976	Car Selling Website
	[http://kievsmi.net/novosti/politics/47097-klimkin-peregovory-ukrainy-i-rossii-idut-neprosto.html]	1941	Unavailable
Retweets	[http://1063.mobi]	4403	Not Active
	[http://USFREEDOMARMY.COM]	1394	Military Recruitment Website
	[http://www.theunder.us]	1235	Unavailable
	[http://payday-loans-24.com/]	1202	Scam Website

Tweet Sentiment Analysis (Whole Subset)



Retweet Sentiment Analysis (Whole Subset)



TF–IDF on Duplicate Tweets (Whole Subset)

Topics

Topics 2014

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#usa	6416.5	good	5679.9	one	2681.5	know	5187.7	wanna	4004.3	give	5600.5	#columbianchemicals	4743.3	one	7613.1	way	2708.8	#love	17051.7
1	'	4914.6	say	4509.6	help	2638.1	man	4598.0	peopl	3833.1	want	5180.0	happen	3291.8	us	4284.9	long	2661.4	love	11017.0
2	#imho	4596.1	believ	3342.8	day	2151.8	feel	4390.7	look	3739.2	life	4496.6	make	3224.5	love	4197.3	world	2401.1	#true	8328.2
3	't	3764.1	realli	3069.3	new	2086.1	got	4262.1	like	3407.7	never	4422.3	fall	3045.1	live	3921.9	someth	2292.4	#rap	7125.8
4	u	2354.4	want	2621.4	fuck	2050.7	come	4216.2	away	3241.1	happi	3952.9	explos	2898.1	life	3736.7	understand	2167.2	like	5366.3
5	#deadhorse	1915.4	foke	2461.0	afraid	2012.3	life	2544.6	heart	3069.3	day	3503.6	never	2818.8	let	3276.2	would	1879.6	need	3277.9
6	it	1886.1	–	2244.9	#russia	1991.1	peopl	2508.9	everi	2702.8	think	3279.2	best	2417.0	go	3118.4	fight	1825.5	still	3221.9
7	need	1827.2	chang	2218.4	ever	1934.5	mind	2439.2	eye	2687.4	#quote	3151.0	thought	2370.0	take	3078.1	#america	1744.8	right	2962.4
8	don	1742.1	stop	2021.8	must	1843.3	like	2387.1	walk	2473.7	made	3071.1	plant	2308.1	noth	2747.4	hell	1649.8	sky	2564.7
9	make	1712.8	go	2005.6	us	1719.0	time	2145.9	hurt	2386.5	#true	2642.2	true	1971.5	much	2185.2	use	1642.3	word	2497.7

Topics 2015

January

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	–	4263.3	#sports	3862.9	live	2049.4	know	2565.6	#fukushima	14575.0	love	5060.6	#news	3752.2	love	2598.3	#news	9794.0	foke	5294.0
1	thing	2367.4	world	2047.3	#quote	1233.3	care	1689.3	ukrain	10380.1	like	4357.9	#politics	3123.7	go	2555.0	man	1503.1	#foke	3034.1
2	chang	2011.5	make	1918.8	–	1180.5	got	1085.2	nuclear	7538.1	want	3451.4	new	1213.5	get	2435.6	#chicago	1503.1	#oscarformuhammad	1539.1
3	someth	1992.1	win	1023.2	#quotes	1100.1	hate	1075.1	#fukushimaagain	6881.8	life	3375.0	obama	1190.7	one	1919.6	polic	1424.6	man	649.8
4	time	1847.3	good	850.8	make	941.8	alway	956.8	#nukraine	6208.1	–	3009.5	hous	874.7	't	1879.9	't	1405.1	#whoismuhammad	547.1
5	give	1699.3	game	788.1	#iamonfire	817.5	lie	916.9	npp	6015.1	look	2735.0	u	762.1	«	1771.1	#local	1339.2	year	520.7
6	peopl	1468.3	love	756.6	life	725.6	offici	835.8	ukrainian	5264.9	someon	2525.9	#health	751.3	»	1734.1	#loca	1139.3	"	464.7
7	never	1359.4	heart	653.5	mistak	635.7	bowl	792.8	chernobyl	4941.1	make	2004.8	#business	646.7	'	1459.4	new	1012.3	#oscarssowhite	453.1
8	believ	1332.4	smile	632.9	man	601.9	super	745.1	new	4352.8	never	1796.1	citi	610.3	come	1418.9	kill	968.4	king	444.2
9	right	1267.2	cri	578.7	never	600.6	anyth	713.9	danger	4202.3	one	1733.0	'	575.2	day	1343.4	'	947.2	get	434.1

February – March

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	–	4689.1	#sports	8285.1	know	3149.5	go	3057.6	foke	3577.7	#news	27641.6	love	7945.2	#politics	4362.5	like	5045.8	–	2477.5
1	one	3842.6	win	1973.5	't	2497.1	#phosphorusdisaster	2681.1	chang	2415.0	#local	5132.1	someon	2827.6	foke	3710.1	feel	2527.5	man	2142.7
2	life	3730.3	game	1459.0	live	1761.2	fall	2675.7	thing	2234.2	#chicago	3673.0	like	2610.8	#foke	2623.1	never	2361.0	say	1227.4
3	want	2652.5	play	1272.3	dream	1620.3	look	2525.2	new	2156.6	polic	2747.1	alway	1813.0	obama	2354.0	make	2246.1	men	1134.0
4	think	2053.8	top	1209.5	don	1254.1	let	2334.1	girl	1965.7	man	2673.4	«	1640.1	#showbiz	1353.4	i	1427.0	good	1079.4
5	love	1932.0	beat	787.1	well	995.5	make	2202.7	time	1819.1	#loca	2595.1	happi	1620.3	isi	1059.9	peopl	1323.8	would	991.7
6	way	1519.9	team	778.4	#quote	933.1	american	1661.4	#foke	1715.2	'	2486.0	tri	1593.6	ever	961.9	friend	1273.2	realli	979.4
7	got	1309.2	#baseba	651.1	cute	889.8	thing	1654.2	best	1629.6	#breaking	2294.8	peopl	1579.4	…	909.7	good	1244.4	u	913.2
8	live	1265.2	final	637.4	fear	863.6	water	1622.1	year	1507.4	new	2242.4	»	1568.1	#money	874.8	video	1180.1	'	881.6
9	world	1216.5	watch	611.5	amaz	821.6	take	1610.1	cat	1257.1	#business	2048.2	never	1429.4	show	801.0	think	1119.2	money	879.3

April – May

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#local	4563.7	#news	13470.2	#news	2674.1	#news	4733.1	love	2842.1	make	1797.3	#politics	6454.7	#sports	9263.9	#news	3911.3	like	1967.7
1	#news	3228.1	man	4277.2	#showbiz	1887.1	#business	2158.7	life	1604.6	new	1519.1	obama	1558.1	game	2199.2	#health	2301.9	peopl	1557.9
2	san	906.5	#chicago	4270.5	#sports	1702.0	#texas	1225.4	–	1424.9	want	1249.4	clinton	1180.1	#news	1909.8	#politics	1562.4	…	1523.9
3	fire	837.9	polic	3827.9	#entertainment	1246.9	texa	1136.7	one	1326.8	orlean	899.1	hous	1042.9	win	1320.7	bill	1251.9	go	1223.7
4	#breaking	688.1	#local	2288.3	top	1237.0	#money	980.1	know	1147.8	look	792.9	#news	1011.8	'	1187.9	new	1186.6	never	1117.7
5	#sanjose	661.3	kill	2197.5	'	1087.8	court	843.6	want	985.6	best	709.8	presid	905.4	lead	931.8	state	880.6	't	995.5
6	day	619.4	shoot	1960.1	#fattenamovie	886.1	law	828.3	day	909.0	#news	682.6	hillari	867.3	play	888.6	#business	794.0	get	943.3
7	year	583.6	#loca	1723.1	star	821.1	school	731.3	alway	874.1	hate	570.1	gop	663.4	live	782.0	tax	715.8	#reasonIamnotasuperhero	941.1
8	miss	529.5	charg	1674.5	red	803.1	high	602.3	great	835.7	bad	567.3	vote	647.1	#breaking	673.0	say	696.1	time	939.6
9	#sandiego	502.1	woman	1662.9	th	787.1	flood	596.9	thing	827.2	day	498.4	run	614.8	coach	634.5	#newyor	653.4	think	922.7

June

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#sports	6077.6	nyc	2649.5	obama	10471.6	via	4674.7	#news	8471.2	#politics	5673.9	stock	5399.1	#news	5529.6	#news	3433.2	#news	3674.4
1	financ	1568.9	#sports	950.3	make	5715.0	selfi	2482.1	#local	4903.2	#news	1520.3	#business	1376.6	#chicago	1871.5	game	2556.5	new	2365.9
2	final	1271.4	home	860.4	money	4663.1	#health	1167.9	man	4200.7	hous	1243.8	greec	945.1	'	1154.9	video	2202.9	court	1452.8
3	warrior	957.9	go	808.4	…	3385.7	new	839.4	polic	3329.4	'	1186.7	us	934.7	–	1094.2	'	2000.8	rule	1170.1
4	open	948.3	back	646.2	like	3103.6	#newyor	821.9	kill	2349.6	vote	1120.6	market	929.1	bbsp	1004.9	#sports	1669.8	#politics	1135.6
5	nyc	940.3	day	642.8	peopl	1966.0	health	718.0	shoot	2073.7	clinton	1100.9	china	698.0	➡	929.8	#showbiz	1047.6	york	890.5
6	】	802.1	red	624.4	get	1869.9	say	698.8	#loca	1507.6	obama	1048.1	buy	680.0	die	913.5	win	1013.8	state	824.7
7	night	767.8	blackhawk	609.4	want	1650.3	#news	655.6	charg	1463.0	tax	932.8	trade	589.8	#breaking	804.8	american	910.7	gay	823.1
8	profile	732.1	san	546.3	go	1624.3	state	556.0	woman	1439.9	bush	856.3	#money	571.6	texa	771.9	flag	707.7	marriag	819.2
9	【see	722.1	big	541.5	one	1462.9	use	468.9	arrest	1416.1	gop	832.1	greek	510.1	#texas	722.8	show	703.1	suprem	718.1

July

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	workout	41201.0	#politics	4815.1	#news	4082.6	#news	6773.4	u	1622.3	#sports	8098.0	exercis	8876.6	#news	4955.1	diet	1804.3	weight	8387.8
1	good	4883.3	obama	4104.9	kill	1600.1	#business	1586.8	exercis	1542.9	#news	1576.8	fat	2185.1	man	3868.1	plan	1511.5	lose	6577.0
2	go	4638.1	trump	2047.9	#local	1332.6	#chicago	1339.7	#news	1308.6	win	1323.0	like	1256.8	polic	2837.0	world	1110.7	loss	2156.7
3	need	3990.3	deal	1239.3	fire	1020.0	new	1205.5	#world	935.7	new	1157.6	fit	1255.1	shoot	1718.3	…	1082.4	one	1486.4
4	get	3579.9	say	1137.5	#world	956.2	'	1079.6	stop	751.0	game	1142.6	much	1087.0	woman	1518.8	black	850.9	tri	1190.6
5	today	3363.0	iran	1080.1	die	843.9	#health	891.1	say	661.6	star	1079.1	eat	961.1	year	1450.4	women	832.6	need	1134.2
6	exercis	2741.7	presid	953.7	look	758.0	school	755.3	fight	643.2	'	927.6	get	910.6	#local	1332.0	video	832.3	…	1009.9
7	day	2178.5	hous	896.4	#unitedstatesinwords	736.8	citi	725.3	right	600.2	lead	636.0	post	906.0	charg	1201.8	us	784.9	want	912.1
8	morn	2134.8	'	829.7	juli	702.1	#tech	703.3	join	547.9	top	614.1	food	799.8	shot	1197.1	#business	716.7	thing	838.9
9	feel	2060.0	donald	791.9	crash	554.2	report	631.3	't	542.6	#showbiz	586.2	look	799.1	old	1163.9	via	656.9	keep	747.1

August – September

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#news	6294.1	workout	2819.1	#business	3989.1	#news	10716.0	#news	3665.7	#sports	5876.7	#sports	7463.8	ashley	893.7	#politics	8688.3	'	5107.7
1	new	3923.3	#fergusonremembers	2105.2	#news	2749.5	man	5349.2	u	2131.7	win	1184.0	lsu	1413.7	#politics	806.7	trump	2780.9	#news	4673.7
2	#world	2348.6	like	1837.2	#health	1787.1	polic	5152.1	pope	886.1	#baseba	781.1	#news	1382.7	debat	701.8	clinton	1440.7	#chicago	2517.9
3	orlean	1567.1	peopl	1374.1	plan	1689.6	#local	4342.4	report	866.5	red	688.0	open	1256.1	#sports	683.3	obama	1433.4	#showbiz	965.4
4	saint	1415.8	need	1318.0	#politics	1350.2	shoot	2782.1	#health	771.0	world	665.6	footbal	1142.9	madison	603.6	say	1131.1	#entertainment	752.5
5	say	740.7	good	1216.7	new	1197.6	kill	2351.2	#world	728.6	beat	658.8	game	915.8	news	574.7	#news	915.3	miss	720.9
6	train	561.4	get	1205.0	million	759.8	#crime	2003.1	state	725.2	lead	648.5	coach	794.5	trump	571.3	donald	871.8	show	697.9
7	york	560.3	go	1191.8	j	758.1	shot	1998.1	#topnews	714.1	jame	594.7	start	792.3	#cleveland	493.1	presid	848.4	bear	678.5
8	problem	458.1	#trumpbecause	1114.0	#tech	711.6	charg	1905.8	air	551.7	sport	587.6	time	784.5	#news	489.1	iran	840.2	chicago	638.0
9	kill	457.4	exercis	1109.1	health	692.1	fire	1867.3	us	516.5	th	509.2	season	706.1	john	462.8	court	788.3	#loca	633.5

October – December

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	'	2094.0	#news	12348.9	#politics	10489.7	#kochfarms	3534.1	#news	2141.3	#blacklivesmatter	1295.1	#sports	15806.6	#news	4746.6	gun	3294.6	#news	10659.6
1	new	1762.1	man	6962.8	#news	6046.8	…	3184.2	#demndebate	2133.1	lead	1001.6	game	3112.4	#business	3441.2	black	2459.3	#chicago	3783.3
2	't	1599.1	polic	5720.1	#world	5067.4	like	1911.3	#demdebate	1734.1	#blacktwitter	965.0	#news	3074.4	#health	3402.6	san	1906.6	new	3762.3
3	#news	1396.4	kill	3404.5	trump	3920.8	#turkey	1871.1	case	1268.8	white	871.7	win	2457.1	#politics	2292.7	…	1617.2	orlean	2112.4
4	#vegasgopdebate	1142.6	shoot	2981.4	say	2101.2	know	1839.4	student	1220.8	world	701.1	'	1908.1	u	1624.2	peopl	1461.1	'	1269.9
5	#gopdebate	1096.6	woman	2579.1	clinton	1907.0	one	1672.1	#tech	1220.2	n	692.0	lsu	1836.8	new	1523.5	#prayerscalifornia	1339.1	night	1134.5
6	#entertainment	1041.8	#local	2503.1	debat	1736.3	peopl	1596.6	protest	1213.1	get	677.8	saint	1469.9	school	1443.2	muslim	1251.1	#sports	1092.6
7	#showbiz	996.1	shot	2485.4	obama	1703.3	thanksgiv	1566.1	court	1164.1	point	672.6	footbal	1394.3	state	1011.8	shoot	1108.5	year	1052.4
8	star	946.4	year	2406.4	syria	1513.1	go	1460.6	#topnews	1135.5	obama	650.0	week	1274.0	#environment	949.3	live	1037.9	#cleveland	1004.0
9	#local	912.3	charg	2405.8	gop	1507.3	good	1455.1	plan	1111.1	#sports	627.1	season	1236.6	#topnews	883.7	polic	1005.0	chicago	914.5

Topics 2016

January – March

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#news	13280.2	#news	6478.2	#tcot	3990.1	#health	3291.1	#politics	9763.3	#sports	10430.5	loan	2563.1	day	1659.7	'	4276.0	...	3846.8
1	#world	4132.7	man	6098.3	#pjnet	3396.8	love	1750.9	trump	5547.2	#news	1922.4	payday	2165.1	right	1590.2	#news	3945.6	like	3092.7
2	#business	2103.2	polic	5988.0	#islamkills	3166.1	...	1151.1	#sports	2878.8	bowl	1436.8	gt	1692.1	good	1180.9	court	1741.1	black	2810.5
3	u	1923.2	kill	2578.6	#ccot	2526.1	amp	1147.2	clinton	2568.5	lsu	1436.2	onlin	1653.1	berni	1165.1	#entertainment	1317.1	peopl	2367.0
4	state	1789.8	shoot	2524.2	refuge	2030.1	go	1028.6	new	2399.5	miss	1390.2	approv	1566.1	need	991.3	water	1177.2	#blacktolive	1797.4
5	new	1772.3	woman	2386.0	obama	1568.6	get	931.7	cruz	2185.7	super	1357.8	x	1491.1	new	918.3	say	1132.3	#gopdebate	1721.9
6	#local	1486.4	arrest	2277.9	mt	1556.1	zika	917.1	win	1988.4	sign	1060.5	#politics	1463.1	know	902.6	#topnews	1033.0	one	1693.3
7	say	1228.0	fire	2207.9	#brussels	1486.1	would	835.3	gop	1525.3	game	1030.6	minut	1389.1	free	822.9	case	973.7	think	1645.3
8	#tech	1086.1	year	2145.4	#wakeupamerica	1484.9	#staywoke	830.3	sander	1493.6	coach	852.1	lt	1341.1	#ididmybestbut	736.7	obama	971.4	get	1615.4
9	citi	1039.5	#local	2105.4	us	1452.8	peopl	808.0	#news	1384.2	nfl	800.7	hous	1110.7	saint	669.1	#politics	930.1	white	1609.3

April – June

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	black	5168.8	#news	10811.0	#business	1638.2	#politics	7262.1	get	1563.5	one	2437.0	man	3408.6	#sports	7657.1	#obamanextjob	1989.1	#ineedalawyerbecause	1833.2
1	peopl	4193.8	#world	4706.1	#health	1491.7	trump	6543.3	#thingseveryboywantstohear	1325.0	love	2112.8	year	2879.0	'	4006.1	...	1615.8	#partywentwrongwhen	1460.8
2	white	2088.9	kill	1850.0	#tech	1437.6	clinton	2615.8	game	1230.8	day	2105.0	#sports	1996.4	#news	2183.9	like	1369.4	#addpotinamovie	1422.1
3	#blacklivesmatter	2018.8	polic	1763.2	#ifiwereyourmom	1132.7	sander	1692.1	#wheniwasyoung	963.4	...	1733.8	polic	1629.6	win	1354.0	look	1232.3	want	1391.4
4	...	1347.6	say	1594.7	...	1061.3	vote	1474.7	n	889.3	go	1442.1	charg	1557.1	san	1154.1	#iamnotthepresidentbecause	1182.5	never	1021.9
5	women	1308.7	u	1583.0	would	934.2	donald	1196.0	play	821.2	see	1285.7	#news	1552.4	#entertainment	1102.0	america	1147.7	like	979.4
6	't	1258.1	#topnews	1157.0	new	931.6	obama	1170.4	start	793.6	#armywomenproblems	1252.1	old	1493.0	game	1002.5	#presentstrumpgot	1081.6	thank	947.1
7	live	1237.0	fire	1125.0	u	904.6	campaign	1049.7	#tosavemoneyi	790.6	good	1166.6	arrest	1382.4	miss	952.4	make	980.5	best	941.7
8	us	1125.3	state	1120.0	use	782.3	hillari	1037.2	take	725.7	get	1141.7	woman	1296.6	#sanjose	832.2	amp	889.4	chang	939.4
9	american	1061.6	#environment	1052.1	tax	765.8	berni	1020.5	water	724.7	give	1030.8	cop	1096.4	warrior	679.9	great	886.7	#thingsnottodoatthezoo	885.7

July – August

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	trump	3571.2	peopl	4724.2	live	1471.2	#news	3959.1	#sports	2363.1	#istartcryingwhen	2062.9	...	1639.9	#childrenthinkthat	2039.9	polic	2883.5	#ihatepokemongobecause	2279.6
1	hillari	2035.7	#mustbebanned	2488.4	#blackskinisnotacrime	1396.1	#world	3353.1	#myolympicsportwouldbe	2109.1	see	1203.9	one	1263.1	#maketvsexy	842.1	man	1680.2	#donttellanyonebut	1814.4
2	clinton	1727.8	american	1358.9	black	1376.4	u	1232.7	win	770.1	#igetdepressedwhen	1122.3	hate	1049.5	show	630.9	black	1333.9	#igetdepressedwhen	1265.3
3	#politics	1094.9	white	1076.4	go	1280.5	state	991.9	#mustbebanned	604.1	like	907.3	need	817.2	time	613.9	kill	1254.7	look	1165.2
4	donald	800.5	america	850.8	#blacklivesmatter	1154.6	attack	899.1	olymp	551.1	watch	616.3	like	782.1	money	610.0	cop	1183.9	like	1163.6
5	vote	794.2	right	793.1	amp	1038.3	say	865.2	#mynewcatchphraseis	483.1	#sextingwentwrongwhen	582.1	#ihate	769.5	#growingupwithobama	576.6	#topnews	1090.6	good	1112.0
6	presid	756.7	...	730.0	...	973.8	#environment	697.1	rio	422.1	#worstpaineverinwords	576.1	amp	717.1	rt	561.5	offic	1027.5	...	1106.3
7	'	722.8	#onewordoffmoviequotes	697.1	get	876.1	kill	649.3	game	418.9	last	525.3	still	688.5	#timetravelerproblems	559.7	shoot	970.6	know	955.1
8	support	678.9	countri	640.0	matter	755.3	islam	538.7	gold	383.7	#ishouldgetamedalfor	523.2	twitter	609.6	#health	517.2	new	873.7	play	852.3
9	obama	663.3	make	608.2	stop	723.7	#sextingwentwrongwhen	500.3	start	377.1	#tech	485.2	get	579.2	get	506.4	shot	721.3	get	724.0

September

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	...	1307.4	#news	3626.6	would	1940.0	#reasonstogetdivorced	1599.8	polic	1877.7	#myemmynominationwouldbe	1925.4	#tofeelbetteri	2280.4	...	2039.0	#tcot	848.8	trump	5641.0
1	state	1085.5	#world	1681.1	#ifihadabodydouble	1326.1	love	1193.0	man	1735.0	like	1882.2	...	2194.5	amp	1298.8	nation	637.7	hillari	3903.7
2	obama	928.5	kill	1067.3	#thingsmoretrustedthanhillary	1026.3	#thingsinventedwhilehigh	643.1	#toavoidworki	1408.8	best	1237.0	make	1409.7	#fishtv	773.1	amp	536.4	clinton	3178.1
3	mt	698.9	#foke	836.0	#ificouldntlie	884.5	come	475.7	black	1296.0	look	877.9	know	1180.4	obama	645.9	thank	536.0	'	1282.8
4	law	614.2	foke	836.0	...	766.6	go	404.4	offic	882.6	game	861.3	bomb	1098.4	start	628.2	...	524.9	...	1244.7
5	#pjnet	552.4	say	773.7	amp	580.9	#ihavearighttoknow	397.0	white	851.3	play	654.7	hillari	999.7	work	600.8	school	468.4	vote	1230.8
6	#politics	505.3	u	700.2	peopl	575.8	#toavoidworki	363.7	year	846.9	#sports	605.9	thing	946.3	get	587.0	god	441.6	support	1153.8
7	us	499.8	n	695.8	#ihavearighttoknow	565.7	#maga	356.9	cop	844.3	ever	572.8	go	798.1	rt	545.7	day	404.6	donald	1011.9
8	via	376.6	new	673.6	well	498.5	one	349.6	shoot	841.2	#ireallylikeyoubut	515.1	amp	749.1	good	542.0	#health	391.1	media	896.4
9	'	375.6	citi	586.2	lie	493.0	#trump	318.0	charg	702.7	get	507.7	terror	704.9	need	541.5	clinton	389.4	#politics	860.3

November

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	...	1118.0	hate	759.2	...	2683.8	trump	5001.4	man	1727.5	play	1552.0	...	6371.3	trump	1969.6	love	1907.1	#todolistbeforechristmas	3133.3
1	#electioninwords	1112.1	...	699.1	get	1333.9	vote	3218.6	polic	1371.0	gt	1414.1	amp	2245.2	'	1419.4	new	959.8	get	1568.5
2	great	918.9	stand	609.2	u	1082.8	elect	1865.8	kill	1170.0	#thingsthatshouldbecensored	1402.7	white	1819.2	...	1070.3	one	918.2	like	1428.4
3	make	717.5	like	562.3	#icelebratetrumpwith	1025.8	#trumpforpresident	1410.1	video	1157.2	#reallifemagicspells	997.1	hillari	1191.4	us	909.7	get	701.3	#sometimesitsokto	1331.0
4	thanksgiv	635.4	peopl	514.9	good	1025.0	win	1356.2	new	982.4	talk	958.9	peopl	1104.0	rt	873.5	...	694.1	...	1119.5
5	america	633.3	amp	511.0	thing	821.3	hillari	1204.7	offic	840.4	make	654.5	clinton	981.8	news	810.6	best	688.3	make	1038.7
6	tweet	623.9	#survivalguidetothanksgiving	492.5	presid	743.9	donald	1010.6	'	724.0	peopl	531.7	right	935.9	thank	787.8	happi	653.6	time	1027.4
7	see	620.2	use	430.8	show	736.8	clinton	979.4	woman	710.8	castro	524.1	go	918.5	follow	781.8	#probabletrumpstweets	646.6	year	903.1
8	much	619.3	attack	408.4	amp	729.1	#maga	968.0	shot	693.8	real	502.6	know	867.5	amp	751.5	thank	591.5	go	886.2
9	come	612.2	state	395.0	feel	635.5	presid	905.3	shoot	682.1	#thingspeopleontwitterlike	498.5	want	835.6	day	704.2	amp	578.5	christma	868.1

October

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	new	2642.8	black	2487.1	...	5270.2	#ruinadinnerinonephrase	1678.5	#news	3111.7	#betteralternativetodebates	2029.3	clinton	3616.4	trump	5496.3	...	3512.7	man	2642.1
1	get	2209.9	amp	1433.4	peopl	2197.9	gt	1628.2	'	2105.5	#rejecteddebatetopics	1394.8	hillari	3210.2	...	2417.8	like	2643.8	polic	2448.2
2	game	895.6	white	1092.3	get	1822.2	play	1438.6	#world	1851.6	#makemehateyouinonephrase	1145.5	...	1685.9	donald	1362.4	look	1467.7	kill	1814.8
3	video	764.8	#nowplaying	1037.1	one	1619.9	hurrican	1047.1	foke	1453.1	listen	1029.4	trump	1567.9	say	1176.3	year	1077.5	year	920.5
4	show	650.0	men	666.6	rt	1617.9	thank	846.6	#foke	1453.1	watch	1011.5	state	1527.0	support	955.3	amp	769.5	offic	916.0
5	amp	607.6	#blacklivesmatter	647.1	amp	1531.5	matthew	835.1	u	1102.3	ft	753.8	email	1126.0	'	909.8	life	739.5	shoot	876.1
6	...	581.6	via	578.8	day	1473.5	go	636.3	follow	834.1	#imtoooldfor	704.2	vote	1000.7	vote	663.9	right	702.3	woman	866.3
7	time	556.8	beat	558.9	think	1282.7	let	596.8	twitter	826.1	album	666.1	elect	902.0	presid	624.0	still	681.9	shot	814.1
8	today	536.8	live	526.5	need	1248.6	world	583.2	say	678.2	track	656.3	campaign	869.4	never	588.0	good	627.6	arrest	807.1
9	via	515.8	women	471.9	make	1137.5	#sports	543.9	readi	650.4	#mybedroominwords	566.1	obama	748.0	hillari	559.0	thing	622.7	charg	806.2

December

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	year	4165.9	man	3015.0	...	2949.3	vote	1526.4	news	2054.1	play	1937.5	#thingsyoucantignore	3131.0	...	4582.5	...	3582.9	...	2680.9
1	'	4007.5	polic	2689.3	amp	1856.1	#giftideasforpoliticians	1401.8	#secondhandgifts	1736.4	amp	1909.0	love	2609.1	#inwords	1642.1	trump	2178.7	right	1743.7
2	black	2500.1	trump	1770.1	peopl	1046.6	#christmasaftermath	1283.9	#idrunforpresidentif	1693.4	gt	1761.7	white	1512.0	amp	1302.9	thing	1409.9	trump	1224.8
3	new	2333.3	kill	1532.0	us	988.2	trump	1002.0	#alternativeacronyminterpretations	1688.2	russian	1271.3	hate	1244.1	time	1001.2	never	1394.2	rt	1095.1
4	...	1365.7	woman	1236.6	god	834.1	elector	731.0	post	1574.1	get	1215.3	christma	983.9	make	966.3	work	1330.2	fuck	1052.1
5	live	1341.2	fire	1229.8	get	700.1	colleg	651.1	could	1155.6	music	1194.1	day	953.2	us	948.1	go	1317.2	know	1031.7
6	old	1230.1	shoot	1168.6	follow	681.0	way	649.0	fake	963.1	video	998.8	one	918.1	via	939.9	read	1029.3	want	884.8
7	peopl	799.0	offic	1131.6	word	669.8	't	580.6	trump	904.4	new	960.5	thank	812.3	get	842.2	realli	1023.6	amp	865.9
8	women	692.9	arrest	963.7	obama	650.2	...	572.1	one	738.3	#nowplaying	955.0	...	754.0	media	723.9	good	982.6	obama	851.2
9	boy	612.1	shot	962.9	think	643.4	state	553.2	new	724.3	russia	866.7	make	708.0	trump	669.7	like	920.2	say	788.4

Topics 2017

January

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	get	2132.1	...	3291.5	trump	4683.3	...	3567.7	man	3113.9	trump	2587.4	play	2029.0	#survivaltips	1289.7	new	1243.3	...	1948.4
1	...	1769.9	peopl	2361.1	...	3647.6	trump	2054.5	polic	2518.1	'	1144.3	rt	1904.5	#myfarewellwordswouldbe	1115.1	fire	855.5	#itsriskyto	1664.1
2	good	1313.0	black	2223.8	presid	1967.4	women	1553.5	year	2006.9	rt	980.8	gt	1820.1	'	630.0	citi	775.2	thank	1464.3
3	know	1230.2	white	1716.8	obama	1458.0	'	1306.4	kill	1718.2	order	927.9	atl	774.1	game	611.9	'	585.4	u	1310.6
4	like	1031.3	amp	1440.6	american	1409.5	march	1166.0	shoot	1299.1	#thefirstdaystodolist	727.5	ft	771.1	first	517.9	close	550.4	love	1181.4
5	go	1005.7	day	1273.3	amp	1323.2	media	951.1	woman	1219.3	obama	572.9	#nowplaying	682.1	#drunkbands	487.1	st	547.6	#potuslasttweet	989.2
6	never	901.3	stop	1173.8	us	1239.2	say	863.4	new	1200.1	state	538.2	fuck	623.1	#sexysports	473.1	show	496.7	leav	729.3
7	make	870.7	one	1113.6	protest	1118.4	news	825.6	old	1009.6	senat	536.1	amp	600.8	issu	463.0	day	457.8	amp	686.1
8	let	847.1	need	1111.3	inaugur	1051.8	elect	743.7	arrest	947.9	hous	530.8	name	503.7	win	462.7	park	436.1	look	682.9
9	one	843.5	get	984.3	muslim	785.8	russia	741.9	shot	870.8	vote	529.2	#dogsongs	476.1	state	391.4	home	432.2	god	639.1

February

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	...	1754.7	polic	1886.2	...	4137.4	...	2968.5	rt	1657.6	trump	6631.5	year	1397.5	time	1454.5	man	1350.9	#valentinesdayinwords	2544.1
1	trump	1084.6	man	994.3	black	1882.1	get	2175.8	gt	1468.1	'	2910.9	best	1013.1	news	1039.2	charg	741.1	#howtoloseyourjob	1435.4
2	#rejectedstatueofliibertysigns	850.8	fire	780.8	amp	1544.6	like	1979.0	new	916.7	...	2078.7	...	1003.9	new	865.9	kill	733.1	#thingsdonebymistake	1189.4
3	come	744.2	home	765.5	white	952.8	peopl	1172.2	play	803.9	presid	1718.1	th	802.3	'	752.8	murder	491.2	day	912.9
4	immigr	688.7	arrest	693.0	peopl	815.2	make	1158.8	video	700.9	obama	1048.9	one	776.5	fake	735.9	death	473.3	school	885.0
5	take	498.1	shot	692.2	u	783.7	go	1099.2	listen	688.1	ban	924.1	last	691.0	citi	664.9	win	471.0	love	579.5
6	would	454.0	found	642.1	women	778.9	thank	1048.5	watch	605.0	hous	822.4	ever	642.5	trump	644.7	child	402.5	tell	543.3
7	turn	449.1	woman	627.2	say	608.0	good	879.3	show	564.8	media	717.4	day	553.1	bill	602.6	state	380.5	boss	529.3
8	countri	431.5	shoot	616.6	histori	597.2	amp	805.5	bowl	548.1	american	705.9	night	551.1	court	552.8	accus	378.4	work	514.9
9	america	428.9	offic	612.3	w	558.4	look	792.2	super	541.1	nation	692.5	old	538.5	senat	470.1	woman	367.7	#howtoescapeadate	449.0

March

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	school	1325.0	#maga	2345.2	...	5644.4	new	3266.9	get	1431.4	make	1665.7	care	1344.1	trump	7074.8	💥 #arrestobama 💥	2403.1	man	2666.1
1	state	1101.4	trump	1536.1	year	2538.0	'	2124.6	amp	1348.4	day	1607.6	health	1129.7	...	6655.8	true	900.2	polic	2437.1
2	illeg	895.8	judg	841.4	like	1664.6	hous	1266.9	...	1288.7	...	918.4	help	1100.7	obama	2830.8	black	793.9	kill	1497.5
3	immigr	796.3	thank	759.2	know	1514.3	trump	918.4	'	1202.9	us	818.8	'	1066.6	presid	1658.7	want	792.8	shoot	1191.1
4	law	732.8	pleas	745.8	amp	1512.8	obamacar	739.8	live	1152.4	women	802.6	gop	827.9	amp	1642.4	men	621.0	arrest	1178.4
5	student	716.1	...	695.2	rt	1276.9	tax	659.4	time	1104.3	great	770.1	bill	804.0	news	1257.9	#obamagate	597.9	woman	1045.0
6	#rejectedlegominifigures	682.1	ban	593.1	peopl	1230.1	releas	558.0	start	1081.0	much	748.0	war	803.2	russian	1177.0	...	528.0	offic	993.7
7	high	671.1	support	581.6	old	1203.9	white	551.6	best	916.5	work	739.0	trump	773.2	say	1164.2	#nowplaying	518.0	charg	885.8
8	citi	612.7	love	580.3	get	1094.1	repeal	469.6	win	672.1	america	723.6	...	725.7	media	1085.6	go	511.0	attack	868.8
9	st	587.0	final	494.0	gt	1030.0	post	459.7	promot	651.4	photo	719.5	't	652.1	russia	915.2	#obamawiretap	464.1	shot	815.1

April

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	...	6501.1	get	2783.7	first	1235.6	new	4020.0	...	1875.9	trump	3903.5	...	3551.4	man	3030.8	...	1614.6	...	2728.8
1	amp	3809.7	...	2559.6	🚨	984.1	post	1097.2	like	1763.3	#maga	1673.6	us	2134.0	polic	2741.1	rt	1526.7	trump	2476.5
2	citi	1027.9	go	1297.5	#nowplaying	894.1	watch	1067.2	peopl	1586.4	obama	1528.1	u	1333.6	kill	1599.6	'	1065.2	need	1423.2
3	america	838.5	use	1033.0	trump	893.1	music	947.2	think	1136.2	news	1378.8	syria	1331.3	year	1356.8	gt	1048.2	tri	1105.3
4	illeg	773.7	'	1032.4	student	768.6	video	705.6	right	1026.2	presid	1223.3	amp	1164.0	shoot	1206.1	democrat	968.7	russia	1084.7
5	w	759.5	tax	1009.3	😐	665.1	'	686.1	black	1017.5	'	1017.0	thank	1155.5	woman	1198.1	play	788.9	vote	1047.8
6	fight	733.7	real	848.4	school	534.6	win	669.2	day	927.3	morn	833.1	love	1064.8	arrest	1147.6	#easterinwords	754.5	support	846.1
7	law	725.7	best	761.8	drop	522.4	shit	598.7	one	926.0	bill	833.1	north	963.0	old	1146.4	court	730.7	elect	801.0
8	back	667.6	pay	700.6	day	498.2	state	575.8	white	924.2	rice	807.1	korea	942.1	suspect	916.1	want	714.5	hillari	764.1
9	world	653.7	one	680.2	cool	460.1	check	542.5	look	924.1	via	770.4	trump	815.4	dead	889.9	senat	600.7	clinton	761.1

May – July

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	kill	1272.8	amp	1919.2	black	2632.3	man	6325.7	'	3321.0	trump	5392.9	fire	2552.1	...	2077.1	...	8128.9	...	3515.7
1	state	1090.8	new	1895.4	white	1857.9	polic	6121.9	school	2004.8	presid	4206.6	counti	976.8	love	1680.9	trump	5864.4	get	2283.9
2	court	1020.3	...	1376.9	hous	1342.7	shoot	2868.7	trump	1237.9	make	4047.8	case	968.8	live	1142.6	amp	2308.0	go	1707.5
3	run	861.9	get	1195.4	american	958.7	year	2757.8	student	1192.3	#maga	3452.1	st	938.0	one	1112.8	comey	1727.2	care	1692.6
4	hit	820.5	video	1032.2	plan	874.5	woman	2647.0	new	1049.2	usususususususususususus	3251.1	new	884.7	day	1103.6	russia	1639.8	health	1658.5
5	attack	699.0	play	942.9	tax	793.0	old	2326.2	law	969.5	america	2937.1	area	769.6	like	1062.0	news	1545.8	bill	1355.7
6	crash	687.4	th	766.2	'	784.2	kill	2313.9	high	894.0	great	2777.1	day	725.0	good	992.6	media	1500.3	senat	1197.8
7	prison	678.7	star	765.0	citi	756.4	arrest	2245.8	hous	876.7	#americafirst	2014.1	us	707.1	peopl	980.2	say	1319.5	back	1123.5
8	win	651.5	left	747.6	new	752.8	offic	1908.7	u	618.4	#makeamericagreatagain	1669.1	'	612.8	thing	908.9	rt	1226.0	time	1116.4
9	islam	528.5	game	700.5	busi	718.5	shot	1871.0	special	593.7	must	963.3	loui	577.7	water	827.1	cnn	1111.9	vote	1033.6

August – December

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	...	1956.2	...	2253.9	white	1542.1	trump	6249.5	...	3760.4	tax	1226.1	...	2373.3	'	1499.3	...	7337.8	'	2089.5
1	vote	1078.6	news	1303.4	black	793.2	presid	2479.6	rt	2587.0	...	1175.7	man	1200.7	via	1000.2	enlist	4954.1	trump	1761.9
2	day	982.6	trump	1136.6	racist	774.2	obama	1567.4	fbi	790.6	new	1051.8	peopl	1072.8	't	981.1	amp	3071.0	north	737.1
3	republican	897.5	year	993.3	hous	692.7	...	1060.9	mueller	767.4	rt	773.4	amp	835.6	...	893.1	patriot	2713.1	korea	723.1
4	like	803.2	support	872.7	american	662.1	'	943.8	barbmuenchen	722.3	polic	659.1	one	731.3	moor	832.1	at	2505.1	nation	524.8
5	senat	715.1	rt	868.6	hey	499.1	hillari	681.1	investig	692.0	illeg	622.4	win	583.5	accus	819.6	us	2219.4	break	510.9
6	see	709.2	fake	699.0	al	473.3	clinton	648.4	clinton	628.8	cut	524.0	kill	525.8	sexual	781.1	armi	2124.1	u	462.6
7	democrat	677.2	cnn	668.0	senat	426.5	"	590.9	know	579.3	citi	495.1	women	508.6	roy	643.6	america	1479.9	say	420.1
8	think	572.8	attack	602.9	fraud	403.0	donald	554.2	amp	503.6	law	488.5	help	485.1	break	561.3	#usfa	1440.1	...	414.3
9	trump	571.5	show	561.4	school	394.7	call	488.0	want	466.8	immigr	431.1	woman	485.0	assault	403.1	stand	1417.4	nfl	414.3

Topics 2018

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	vote	227.8	trump	2659.4	...	322.2	state	240.1	peopl	598.0	"	250.0	black	344.1	peopl	336.9	state	321.1	#uwi	659.9
1	may	134.4	presid	517.6	day	269.5	us	226.9	...	506.3	discuss	188.2	year	305.9	countri	257.2	right	304.4	#unitedworldinternational	528.4
2	first	129.0	...	288.9	women	241.4	militari	162.6	venezuela	420.3	#valdaiclub	181.1	obama	239.6	...	245.9	...	220.9	#turkey	388.6
3	elect	106.2	donald	272.0	today	190.1	#uwi	145.5	social	294.0	russia	177.5	't	237.6	us	236.9	unit	189.3	presid	343.1
4	#brexit	106.1	like	223.9	peopl	150.6	israel	133.4	work	276.5	saudi	171.5	...	228.3	war	188.8	us	162.8	said	299.2
5	union	105.8	'	214.5	school	147.0	forc	131.4	capit	243.4	expert	166.1	one	216.7	'	178.7	presid	112.9	'	285.7
6	parti	97.5	mueller	210.1	shoot	135.0	'	128.6	revolut	158.7	'	141.4	gun	198.7	children	153.5	year	102.7	#us	232.5
7	leav	95.0	say	167.7	march	124.0	#unitedworldinternational	116.8	empir	157.4	russian	138.0	time	184.7	kill	141.6	human	101.6	minist	213.3
8	america	93.4	want	152.7	student	123.2	...	114.8	capitalist	154.1	meet	125.4	old	177.4	white	141.3	govern	98.4	turkey	207.9
9	brexit	88.1	white	148.5	thank	105.2	unit	109.5	world	141.0	#russia	118.4	like	172.8	american	125.0	nation	95.9	u	189.0

Topics 2019

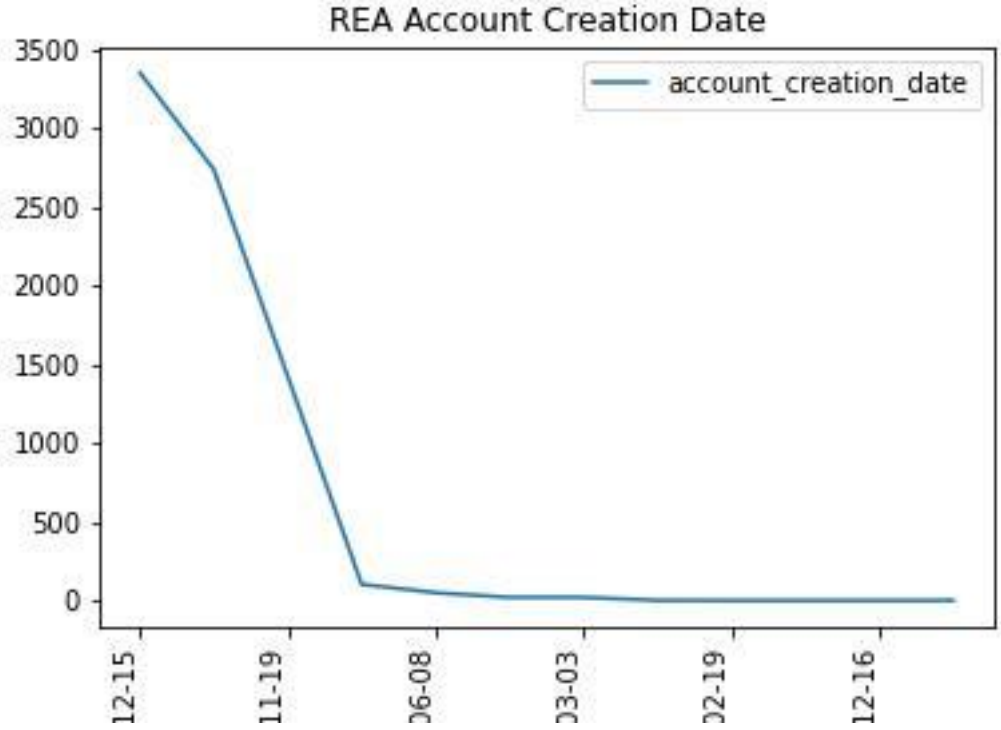
	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	russia	183.3	...	582.9	...	649.3	elect	180.7	peopl	429.8	#us	286.7	...	345.7	peopl	206.3	...	852.4	#valdaiclub	462.9
1	#turkey	181.8	bolivia	125.1	presid	268.6	th	128.3	venezuela	322.0	#usa	253.8	't	185.0	china	180.5	us	301.1	expert	355.1
2	'	144.5	presid	121.8	#venezuela	236.1	parti	122.7	social	276.0	u	234.5	like	175.1	right	176.8	peopl	220.4	write	171.1
3	turkey	130.6	america	120.1	venezuelan	185.7	discuss	97.2	revolut	150.5	trump	229.8	want	156.7	'	166.6	work	216.1	note	106.1
4	syria	110.8	first	111.0	venezuela	170.1	"	96.6	world	148.5	said	180.0	state	153.8	...	158.6	year	171.5	relat	87.0
5	#syria	69.1	coup	108.0	opposit	159.3	european	93.7	thank	140.5	'	155.2	media	136.7	capit	141.6	million	119.5	russian	85.2
6	turkish	64.9	year	97.9	support	112.6	'	90.1	govern	139.3	say	145.5	venezuela	133.5	war	133.1	live	119.1	director	83.8
7	us	63.6	evo	86.1	maduro	105.2	club	78.0	...	138.2	#trump	118.3	peopl	125.3	human	129.0	class	117.0	#us	81.5
8	end	61.0	democrat	84.6	amp	92.6	valdai	77.0	bolivarian	114.1	iran	115.5	would	110.6	it	94.0	day	115.5	#china	81.0
9	peac	57.1	attack	80.1	govern	83.7	confer	70.1	right	109.2	foreign	114.5	power	91.1	us	84.0	venezuela	108.4	question	78.1

Topics 2020

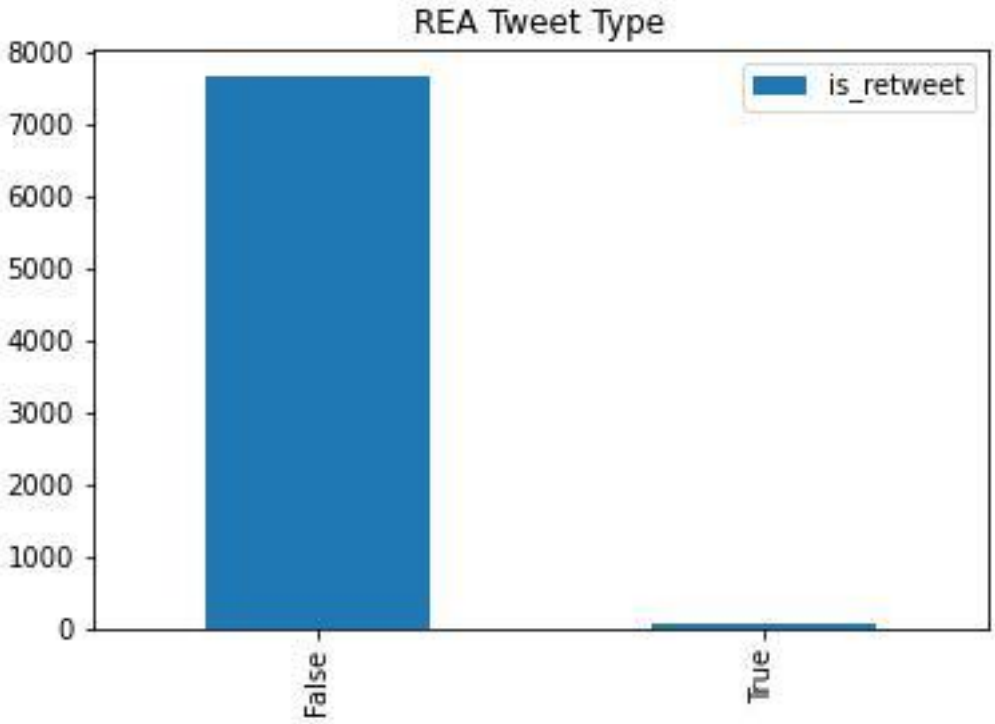
	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	...	239.5	#valdaiclub	263.4	...	386.0	messag	158.1	op	630.1	turkey	216.1	...	326.3	...	685.5	war	283.4	...	623.3
1	'	129.0	econom	187.3	us	245.9	like	152.7	ed	628.1	'	154.8	peopl	228.4	presid	228.7	us	130.6	peopl	454.8
2	war	125.7	'	183.4	world	185.3	would	151.6	#us	313.0	plan	106.6	'	160.4	call	127.2	...	103.2	venezuela	269.4
3	us	123.4	"	169.8	china	169.8	direct	146.1	#turkey	280.8	east	105.0	million	134.6	'	117.9	new	93.3	social	201.9
4	protest	94.9	pandem	167.0	use	139.3	follow	137.2	#coronavirus	263.5	middl	103.3	us	109.6	nation	106.6	billion	74.1	right	179.5
5	crime	91.1	discuss	162.1	coronaviru	125.8	lie	119.6	#covid	259.9	...	83.0	worker	106.4	minist	102.6	#us	68.5	human	162.8
6	articl	87.3	world	158.6	new	105.9	could	118.4	#russia	201.5	look	72.9	health	105.7	say	98.4	iran	68.1	year	148.3
7	't	84.3	global	141.8	sanction	104.5	send	116.1	#eu	199.2	#erdogan	68.2	make	103.0	state	98.1	#middleeast	66.1	work	142.1
8	support	78.8	onlin	125.4	time	102.7	...	114.4	тк#turkey	165.7	latest	64.3	get	101.5	foreign	89.3	trump	65.2	govern	141.1
9	way	72.8	titl	112.1	state	96.4	pleas	106.1	#china	163.1	like	60.6	via	90.4	right	84.2	ethiopia	65.1	american	131.1

IV. REA

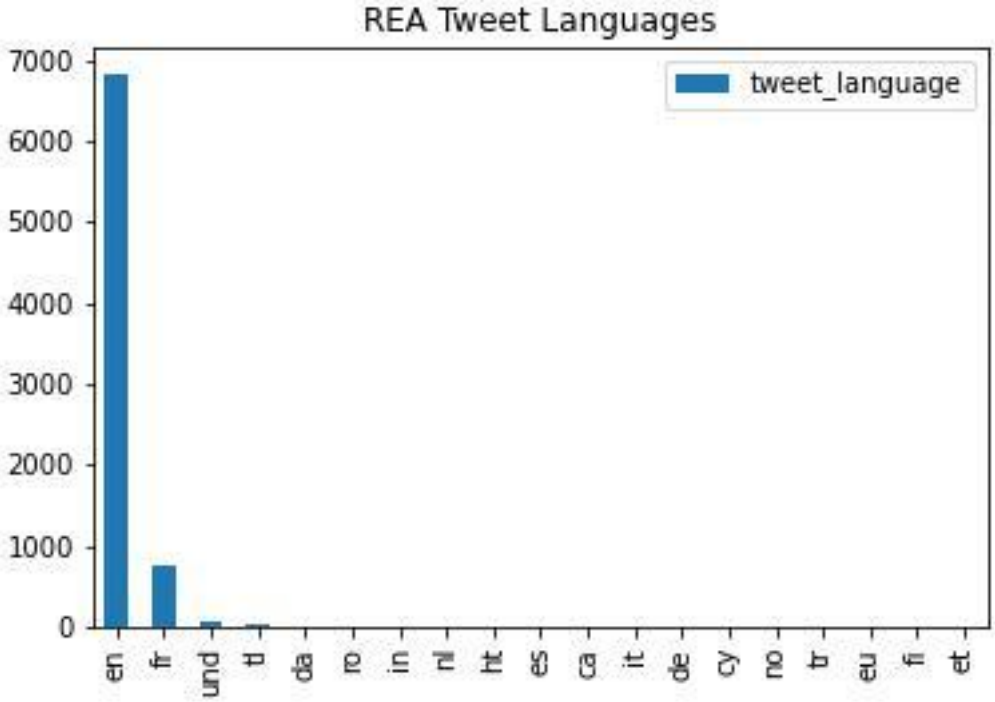
Account Creation Dates


REA Account Creation Date
account_creation_date
3500
3000
2500
2000
1500
1000
500
0
12-15
11-19
06-08
03-03
02-19
12-16

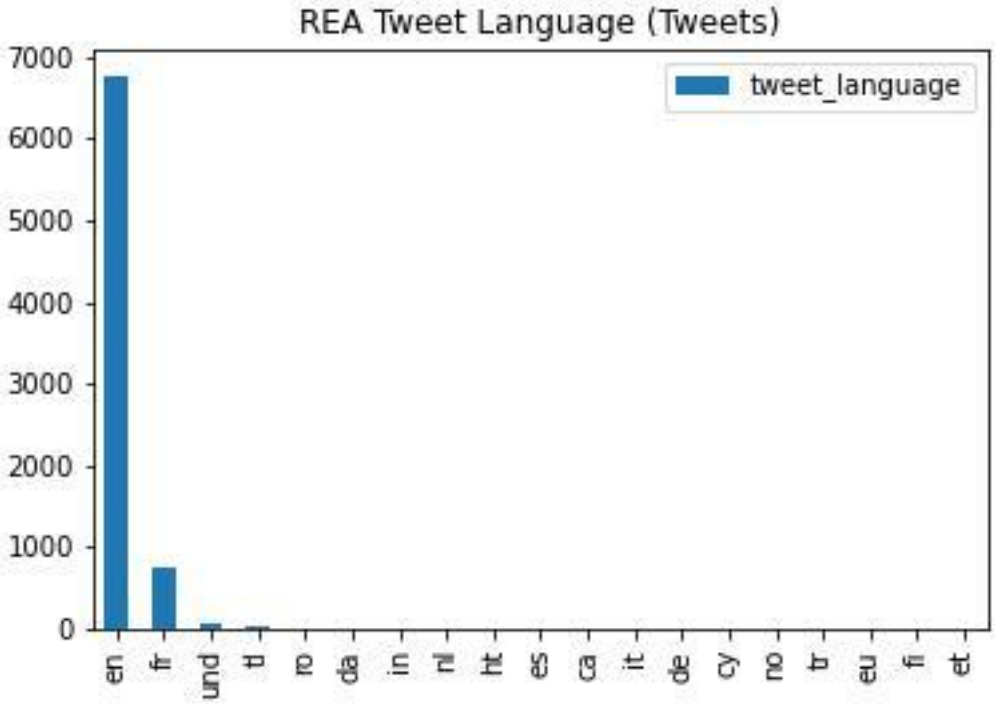
Tweet Types



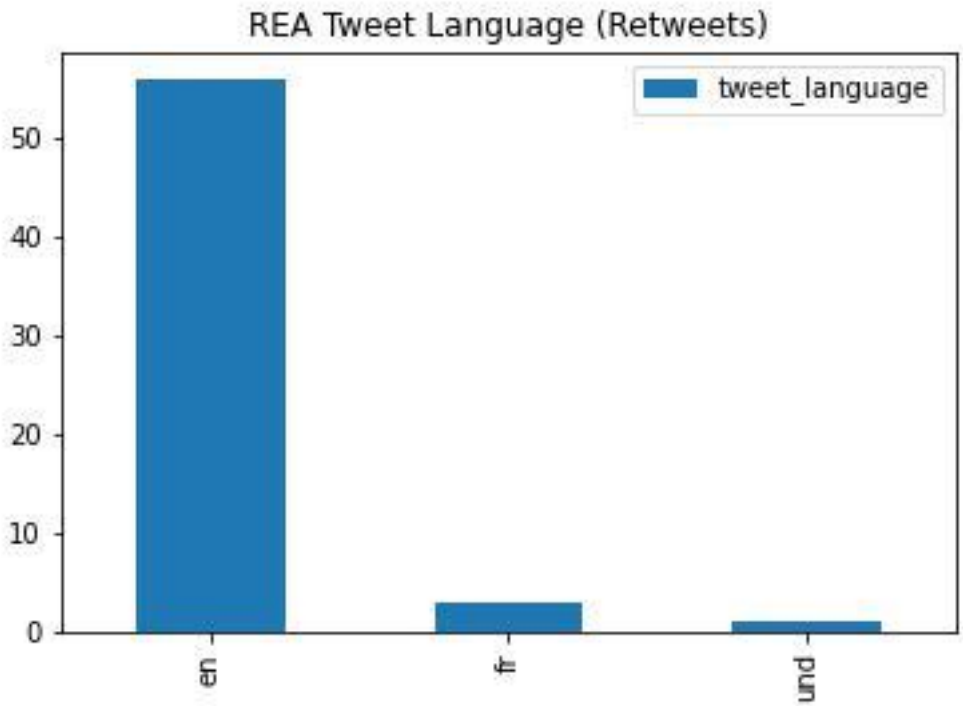
Tweet Languages (Overall)



Tweet Languages (Tweets)



Tweet Languages (Retweets)



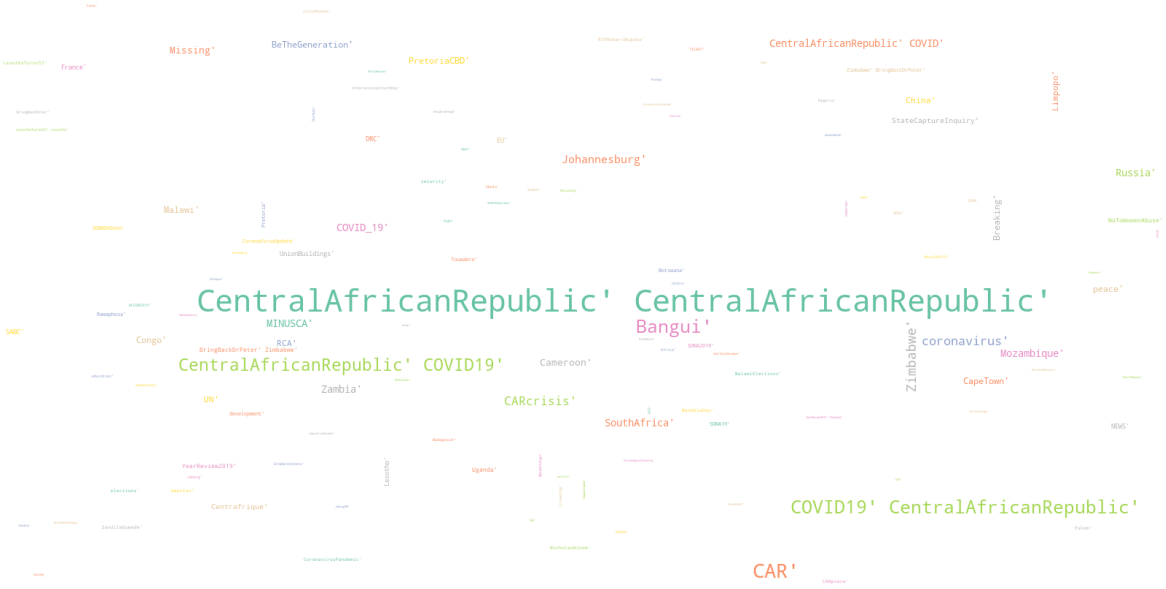
Term Frequency

TF–IDF (Term Frequency–Inverse Document Frequency)

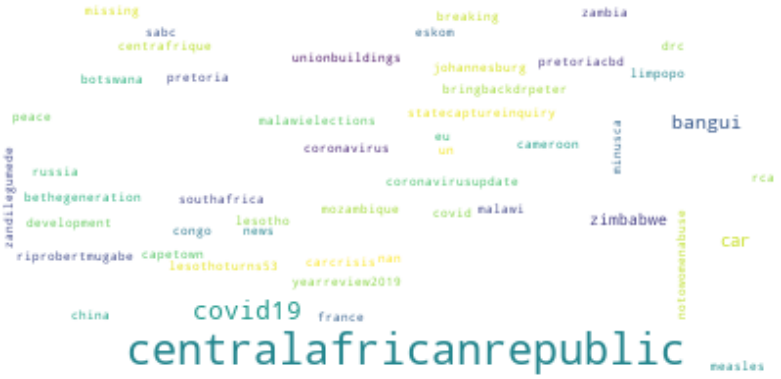
Hashtags Shared

	REA Hashtags			
	Most Shared		Least Shared	
Grand Total	['CentralAfricanRepublic']	263	['Malawi', 'MalawiDecides2019']	1
	['Centrafrique']	231	['Belfast', 'Mpumalanga', 'Corruption']	1
	['Bangui', 'Centrafrique']	40	['NEWS', 'UJ']	1
	['CentralAfricanRepublic', 'COVID19']	26	['financingthefuture']	1
Tweets	['CentralAfricanRepublic']	263	['Cosatu', 'BankStrike']	1
	['Centrafrique']	231	['Khayelitsha']	1
	['Bangui', 'Centrafrique']	40	['Zimbabwe', 'Poachers']	1
	['CentralAfricanRepublic', 'COVID19']	26	['NelsonMandelaDay']	1
Retweets	['WorldCup']	1	['GPS']	1
	['Afrikaans']	1	['SocialMediaBlackout']	1
	['Zlatan']	1	['SocialMediaBlackOut', 'BlackOutWednesdays']	1
	['MandelaDay', 'TheNextChapter', 'ActionAgainstPoverty']	1	['DeepCoach']	1

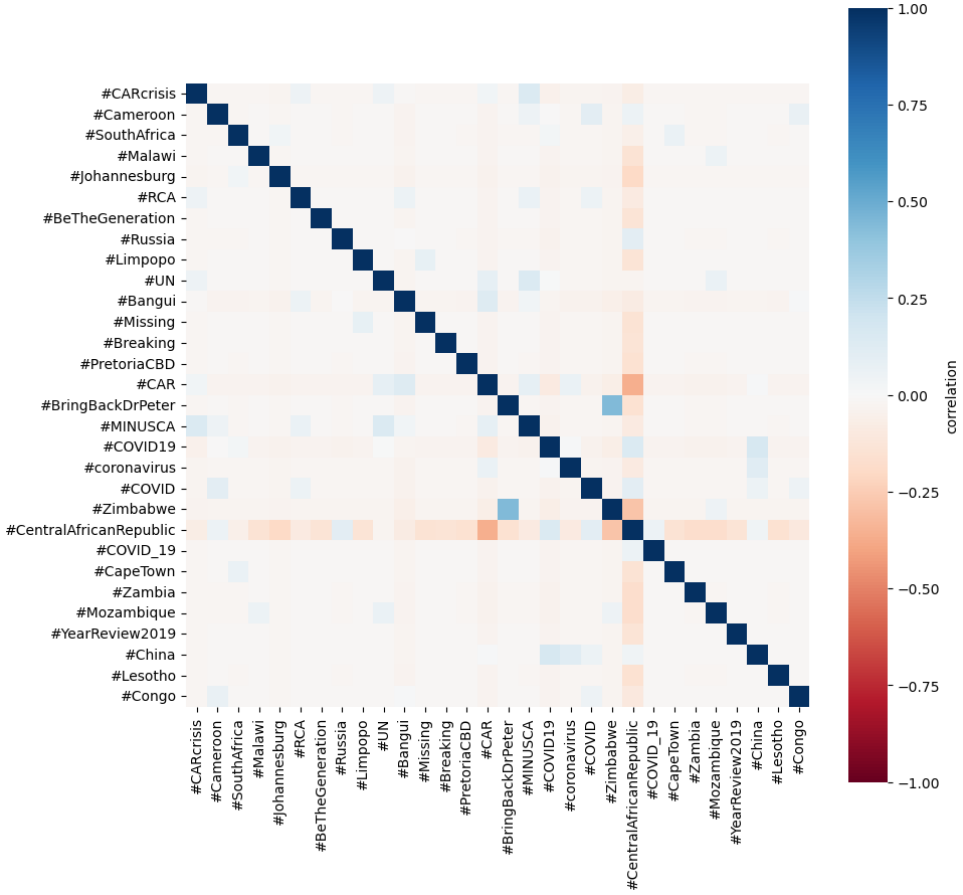
Hashtags



TF–IDF Hashtags



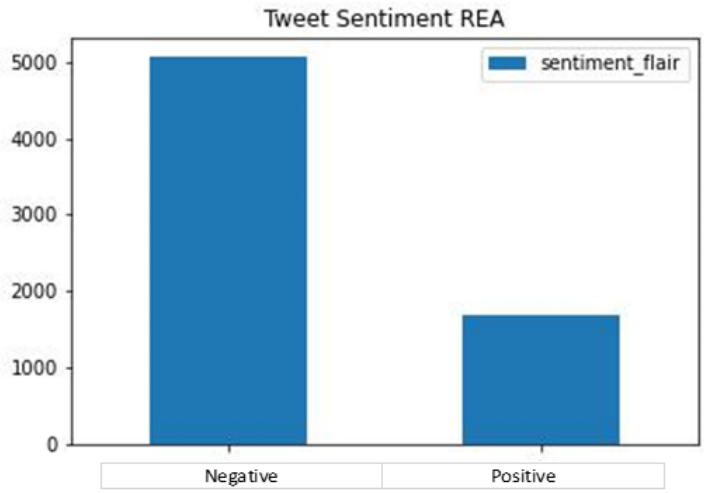
Hashtags Heatmap



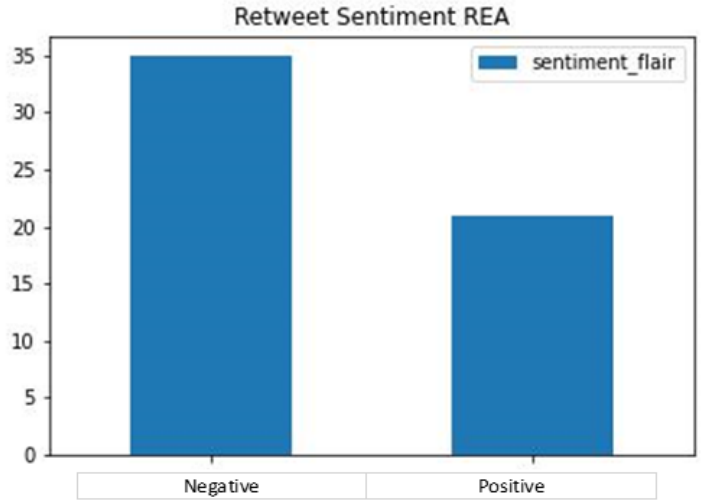
URLS Shared

	REA URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['https://livereport.co.za/2019/06/05/bloemfontein-man-left-with-his-tongue-bitten-off-after-alleged	4	Blocked
	['https://livereport.co.za/2019/05/22/kzn-man-to-spend-life-behind-bars-for-raping-ex-girlfriend/A']	4	Blocked
	['https://livereport.co.za/2019/05/23/ufs-students-on-protest-campus-security-guard-injured/A']	3	Blocked
	['https://livereport.co.za/2019/05/15/attempted-hijacking-in-midrand-family-drive-off-leaving-hijack	3	Blocked
Tweets	['https://livereport.co.za/2019/05/22/kzn-man-to-spend-life-behind-bars-for-raping-ex-girlfriend/A']	4	Blocked
	['https://livereport.co.za/2019/06/05/bloemfontein-man-left-with-his-tongue-bitt['https://livereport.c	4	Blocked
	['https://livereport.co.za/2019/05/15/attempted-hijacking-in-midrand-family-drive-off-leaving-hijack	3	Blocked
	['https://livereport.co.za/2019/05/23/ufs-students-on-protest-campus-security-guard-injured/A']	3	Blocked
Retweets	['http://bit.ly/2tIlR25']	1	African News Website in English
	['http://ow.ly/2fmj30iuFNC']	1	Sport News Website
	['http://rfi.my/4upS.T']	1	African News Website in French
	['https://www.joburgtoday247.tv/?t=v&mid=Opx2ruw8&fid=PoZLrghV']	1	News

Tweet Sentiment Analysis



Retweet Sentiment Analysis

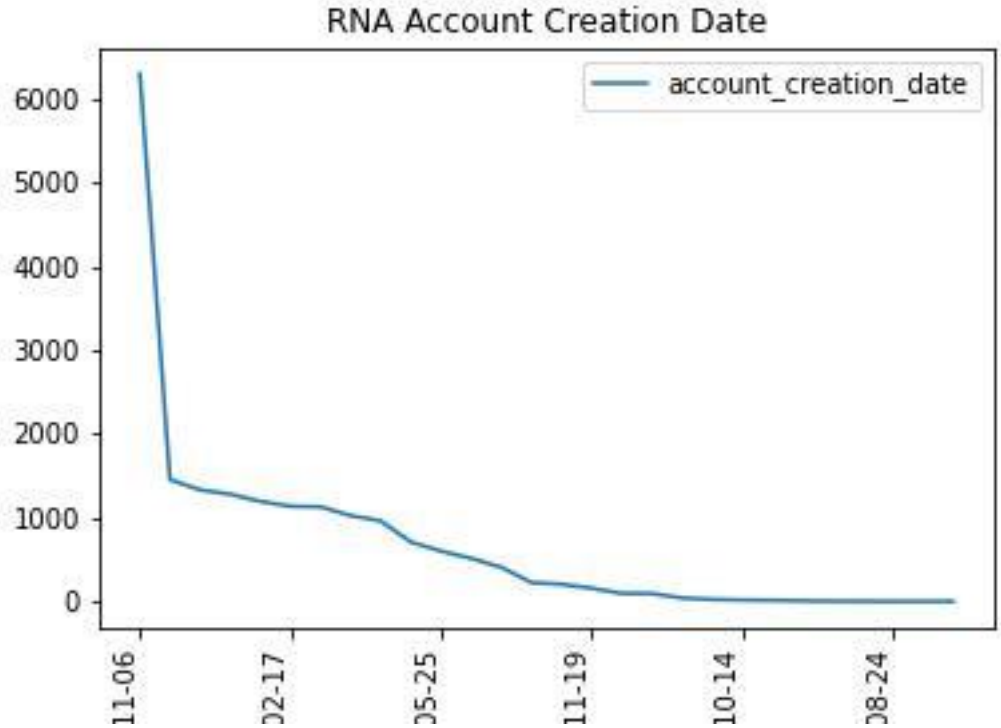


Topics

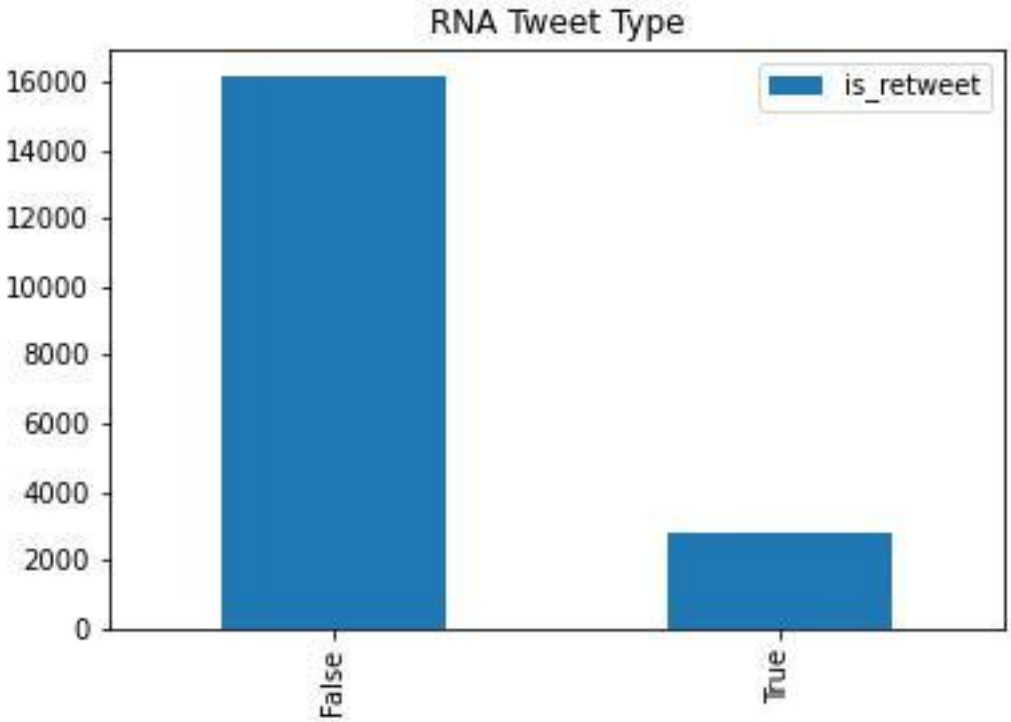
	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	protest	221.4	#centralafricanrepublic	480.0	'	267.7	cape	252.9	polic	513.0	school	255.1	water	118.1	'	481.8	car	258.6	arrest	527.9
1	forc	151.0	presid	357.5	ramaphosa	133.2	fire	212.1	year	502.4	year	250.5	citi	112.3	africa	279.8	driver	204.0	suspect	362.7
2	african	137.9	minist	230.0	anc	131.6	town	197.6	man	469.9	old	239.2	watch	108.9	south	262.1	road	192.1	polic	338.3
3	group	132.3	nation	207.5	court	122.1	shot	195.0	murder	371.9	found	196.3	state	92.5	zambia	261.7	vehicl	188.3	r	245.7
4	attack	124.5	african	196.8	cyril	81.8	man	168.2	old	362.7	bodi	191.4	peopl	88.5	zambian	209.9	crash	174.1	possess	223.3
5	south	111.4	elect	150.6	public	79.7	kill	149.8	court	329.7	woman	153.0	live	85.6	zimbabw	173.1	peopl	153.6	drug	195.1
6	arm	103.0	parti	118.6	vote	76.4	park	105.7	suspect	317.5	hospit	134.8	day	78.6	new	151.5	accid	147.6	offic	140.8
7	polic	97.4	gener	117.5	read	75.6	morn	104.8	offic	259.2	girl	115.9	place	70.8	presid	148.5	die	140.6	found	123.1
8	student	95.8	#covid	114.1	botswana	72.4	dead	97.3	arrest	254.3	polic	112.2	resid	65.4	govern	146.0	taxi	130.6	firearm	122.2
9	violenc	82.2	visit	109.6	full	72.1	die	90.9	appear	227.1	death	110.9	new	64.3	nation	137.0	injur	125.8	oper	105.8

V. RNA

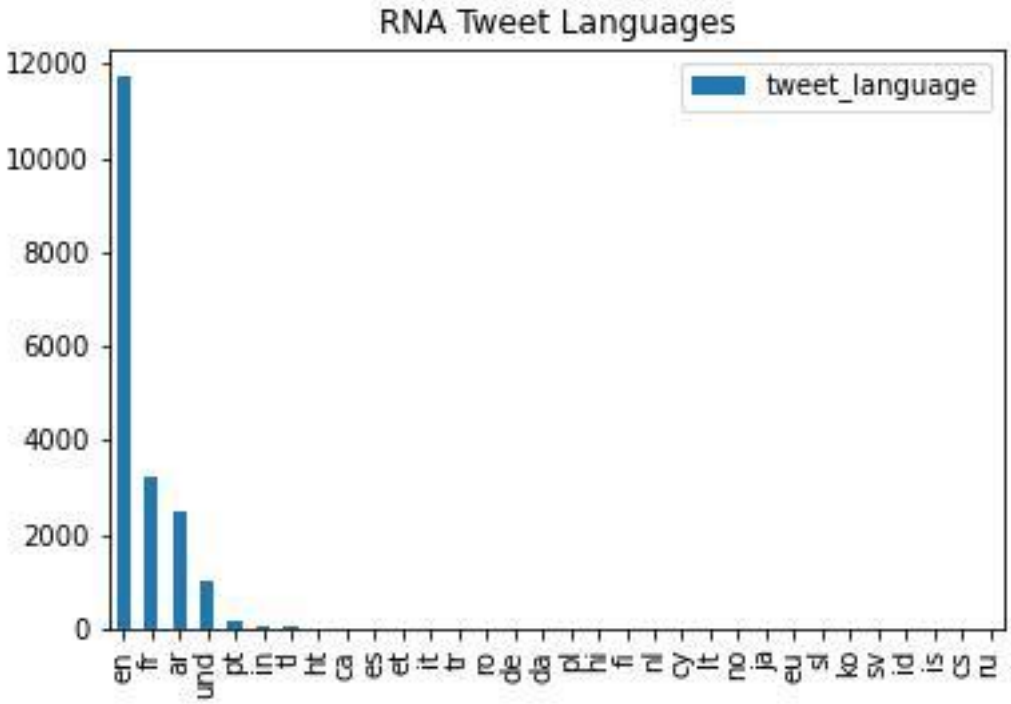
Account Creation Dates



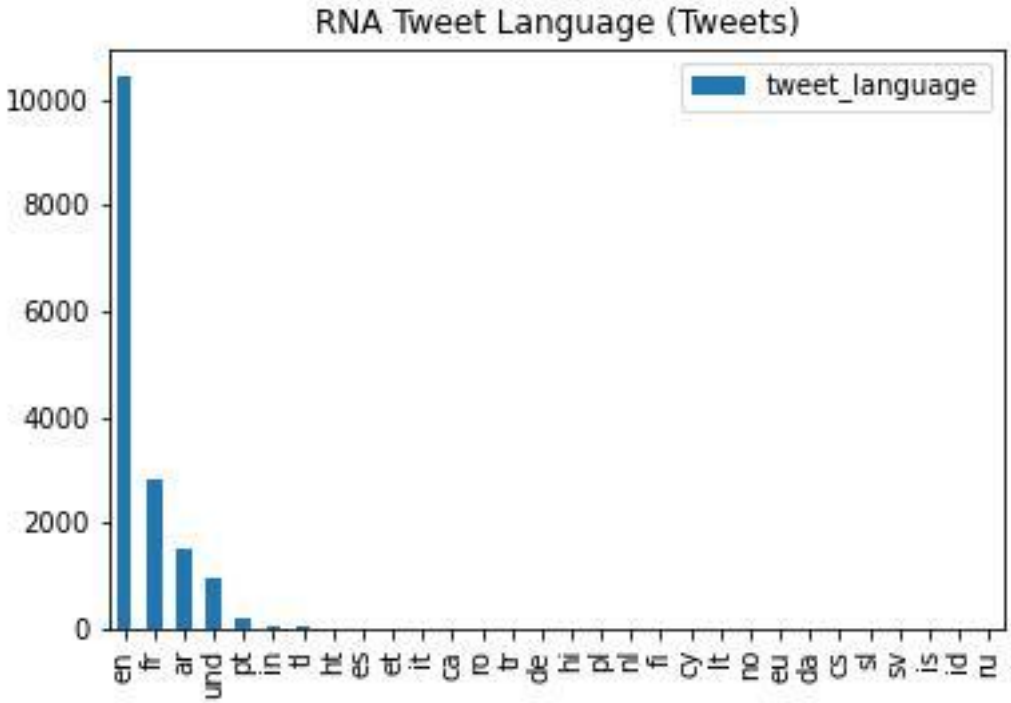
Tweet Types



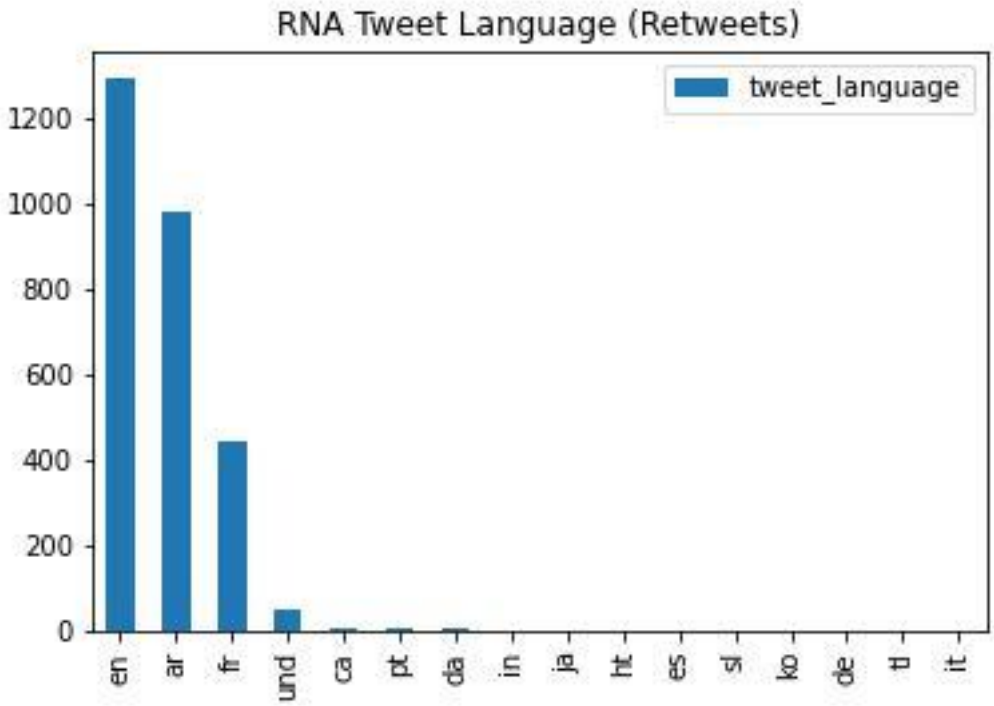
Tweet Languages (Overall)



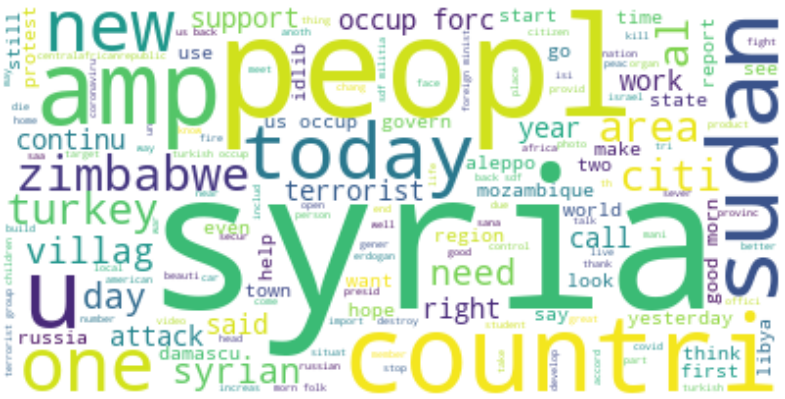
Tweet Languages (Tweets)



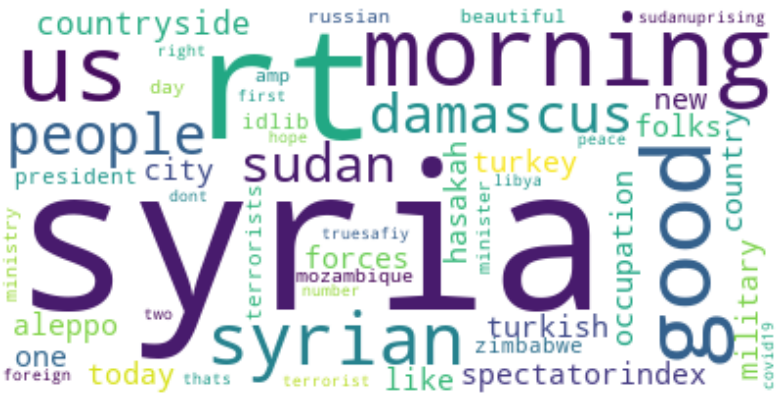
Tweet Languages (Retweets)



Term Frequency



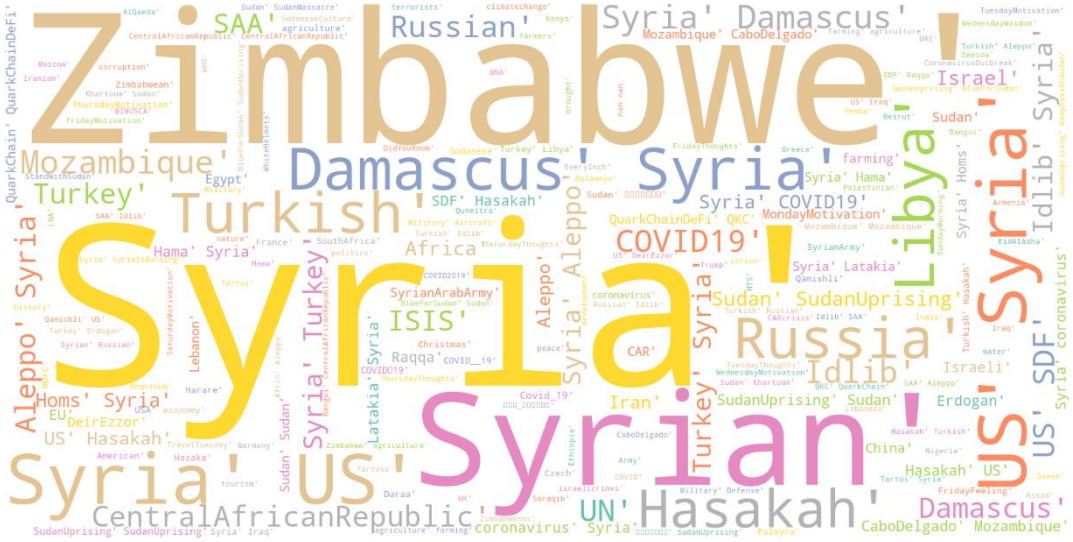
TF–IDF (Term Frequency–Inverse Document Frequency)



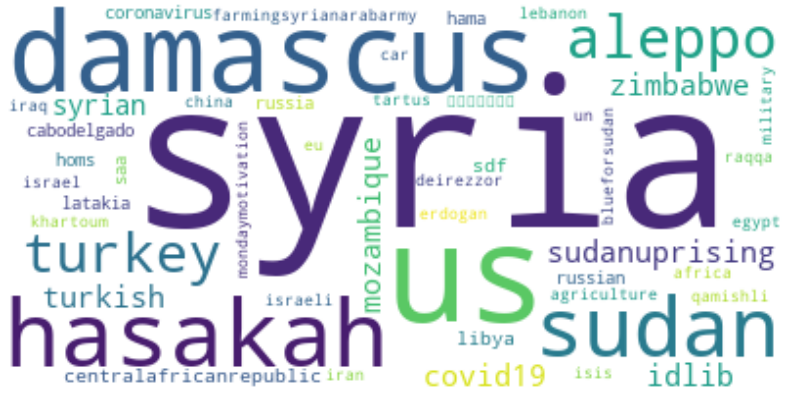
Hashtags Shared

	RNA Hashtags			
	Most Shared		Least Shared	
Grand Total	['Syria']	516	['Russian', 'ISIS', 'US', 'Euphrates']	1
	['Centrafrique']	198	['Russian', 'Syria', 'Russia']	1
	['Sudan']	139	['Iranian', 'Syrian', 'Damascus']	1
	['Damascus', 'Syria']	128	['AirIndia']	1
Tweets	['Syria']	508	['centrafrique']	1
	['Centrafrique']	140	['Mongoumba', 'Covid19']	1
	['Damascus', 'Syria']	127	['China', 'Sudan']	1
	['Damascus']	107	['healthcare', 'Mozambique']	1
Retweets	['Centrafrique']	58	['MiddleEasternbride', 'Lebanese', 'انفجار_بيروت'] (*Explosion Beirut)	1
	['SudanUprising']	34	['Tours']	1
	['Sudan']	33	['covid_19', 'AFP']	1
	['ليبيا'] (*Libya)	30	['باشاغا', 'العثمان'] (*Bashagha', 'Othman (Minister of Health Libya)	1

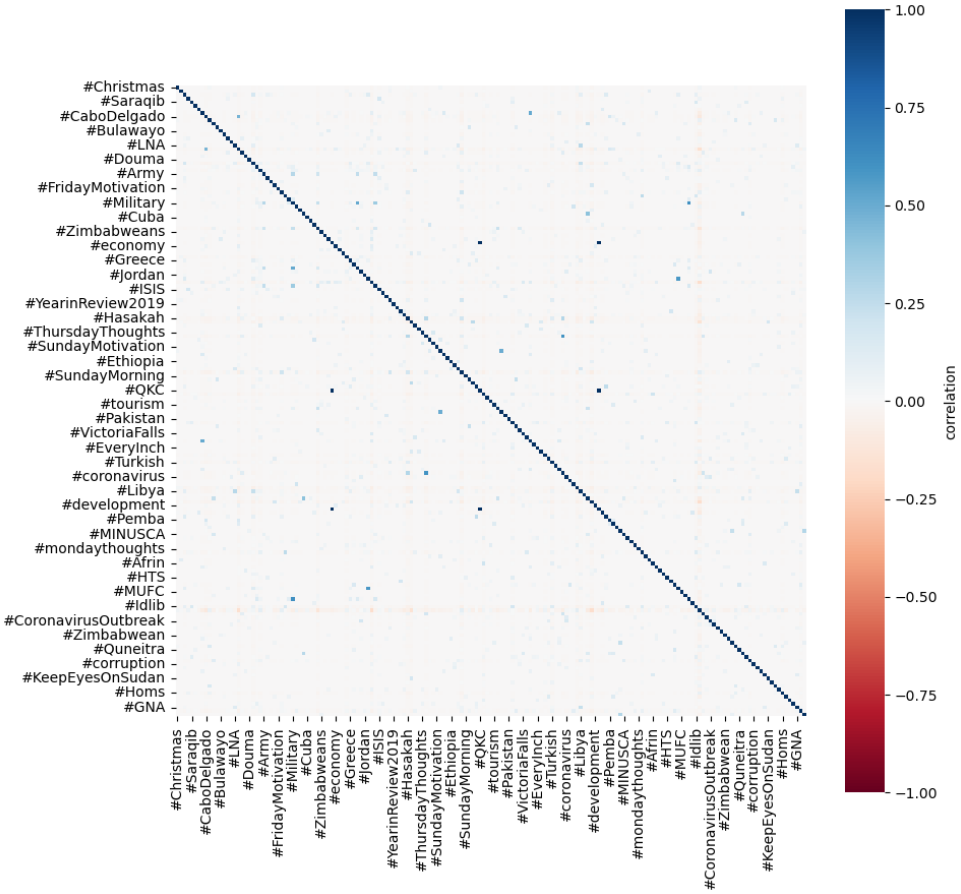
Hashtags



TF–IDF Hashtags



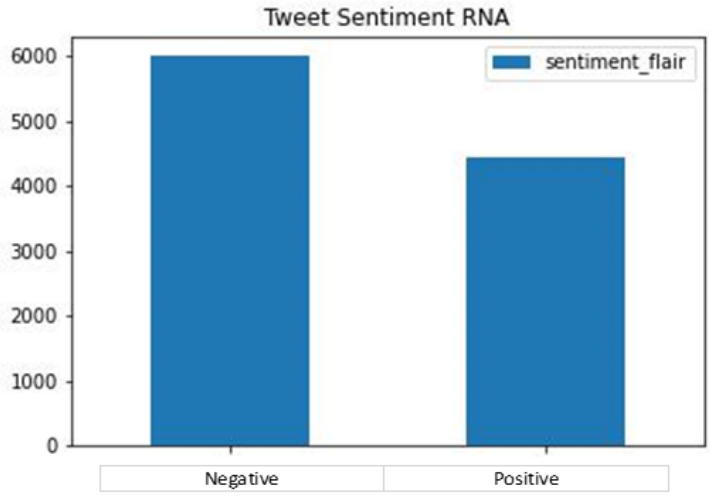
Hashtags Heatmap



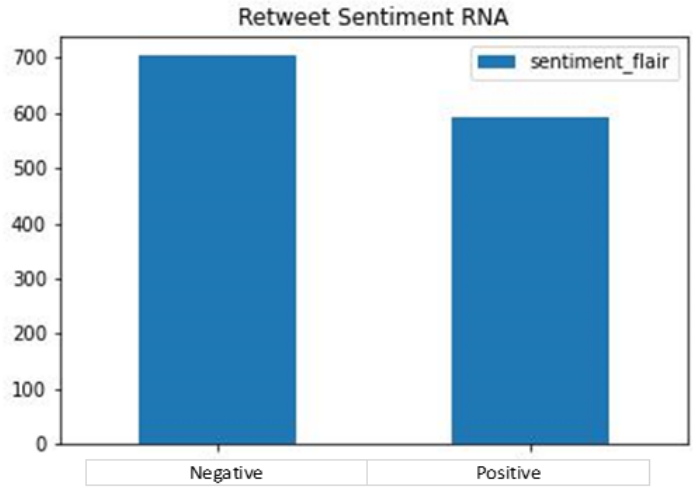
URLS Shared

	RNA URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['https://twitter.com/x_m6y/status/1193182788297646081']	15	Tweet in Arabic: "It saddens me to see my countrymen wrestling and telling lies to each other" Who is the beneficiary?
	['http://okex.com/p2p/btc-to-inr']	9	Bitcoin Website
	['https://worldnews-sy.com/%D9%85%D9%86-%D9%8A%D9%82%D9%81-%D8%B9%D8%A7%	3	Syrian News Website in Arabic
	['https://twitter.com/x_m6y/status/1193182788297646081?s=19']	2	See the Tweet above
Tweets	['https://twitter.com/x_m6y/status/1193182788297646081']	15	See Tweet above
	['http://okex.com/p2p/btc-to-inr']	9	Bitcoin Website
	['https://worldnews-sy.com/%D9%85%D9%86-%D9%8A%D9%82%D9%81-%D8%B9%D8%A7%	3	Syrian News Website in Arabic
	['https://r8qs.adj.st/?adjust_t=ynoul8_q0le9d&adjust_deeplink=mobilelegends://appinvites']	2	Link to Appstore
Retweets	['https://twitter.com/steve_hanke/status/1152581365692428288']	2	Tweet: "Since #Sudan's ruling military council and leaders of the pro-Democracy movement signed a power-sharing agreement on Wednesday, Sudan's #inflation, by my measure, has taken a dramatic fall from 61% to 36% in just 3 days."
	['http://EcoFi.io']	2	Bitcoin Website
	['https://twitter.com/Sudaneseblogs/status/1149015143965237248']	2	Tweet: "Thareed: A document circulating on social networks as a copy of the draft political agreement between the Military Council and the Forces of Freedom and Change, and we note that it may not be the final document, especially for the amendments that appear in it."
	['https://twitter.com/samirasawlani/status/1155559234207191040']	2	Tweet: "Received video of one of the many Protests taking place in Sudan today as people demand civilian rule and demonstrate against Transitional Military Council #SudanUprising"

Tweet Sentiment Analysis



Retweet Sentiment Analysis



TF–IDF on Duplicate Tweets

Topics

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	good	295.5	...	370.3	#us	566.9	#turkey	247.0	#syria	1324.4	peopl	295.5	#centralafricanrepublic	213.7	al	565.0	sudan	278.9	#syria	365.8
1	peopl	251.5	#zimbabwe	287.3	occup	443.7	#libya	212.1	syrian	391.0	die	115.9	african	160.6	terrorist	522.3	#sudan	270.7	#covid	299.5
2	hope	234.0	amp	238.4	forc	418.9	presid	156.0	minist	341.9	children	102.2	use	159.3	countrysid	434.1	#sudanuprising	256.0	ministri	267.9
3	like	207.5	'	233.6	al	302.7	militari	154.8	foreign	315.3	live	99.9	#mozambique	159.1	#syria	402.8	protest	213.2	health	249.4
4	one	179.6	countri	176.0	#hasakah	261.3	council	145.9	#damascus	264.5	kill	94.2	countri	128.1	fire	354.9	water	212.6	case	248.1
5	know	173.8	us	170.7	#syria	257.1	agreement	145.5	al	233.0	man	91.5	forc	127.7	citi	308.5	day	206.3	syrian	240.3
6	see	167.0	world	144.3	countrysid	255.7	countri	126.4	syria	211.7	two	91.2	arm	121.4	#idlib	279.5	...	178.9	announc	202.2
7	🙏	161.1	africa	129.7	back	250.6	meet	122.3	presid	159.0	one	85.6	like	109.3	villag	269.2	gener	126.1	number	163.5
8	get	159.8	new	127.0	militia	180.5	...	117.8	terror	148.6	school	82.9	join	105.1	attack	246.6	student	120.9	#coronavirus	160.1
9	time	153.9	help	126.1	#sdf	176.3	state	103.0	morn	144.0	#farming	81.5	win	105.1	#aleppo	234.7	today	120.2	new	158.4

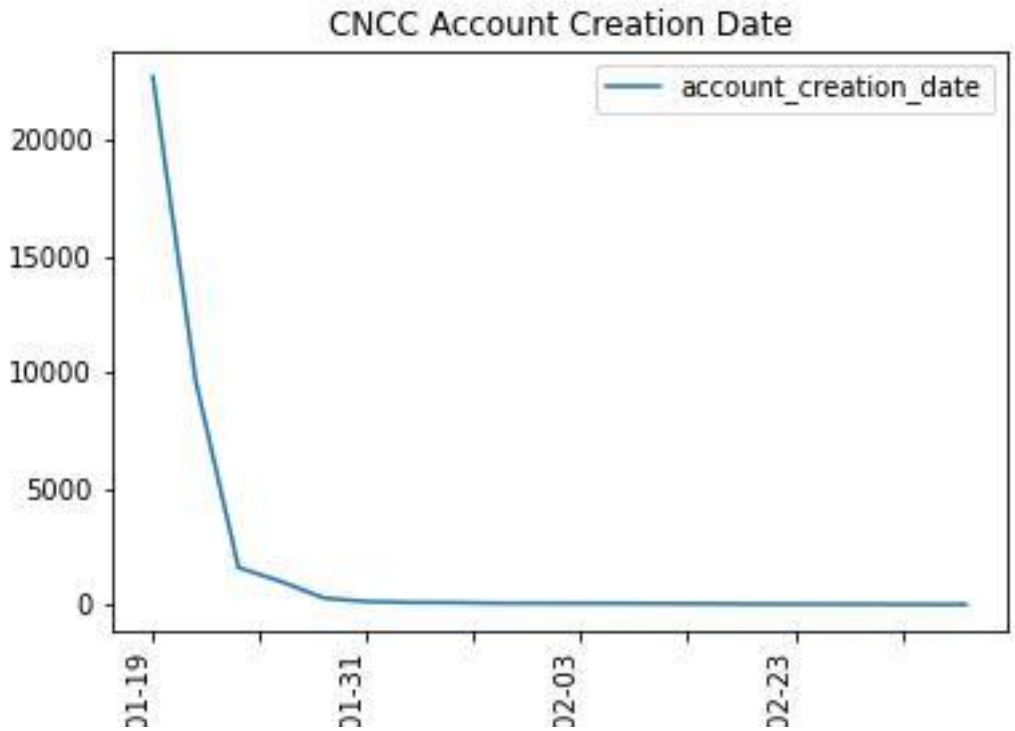
China

I. Data Overview

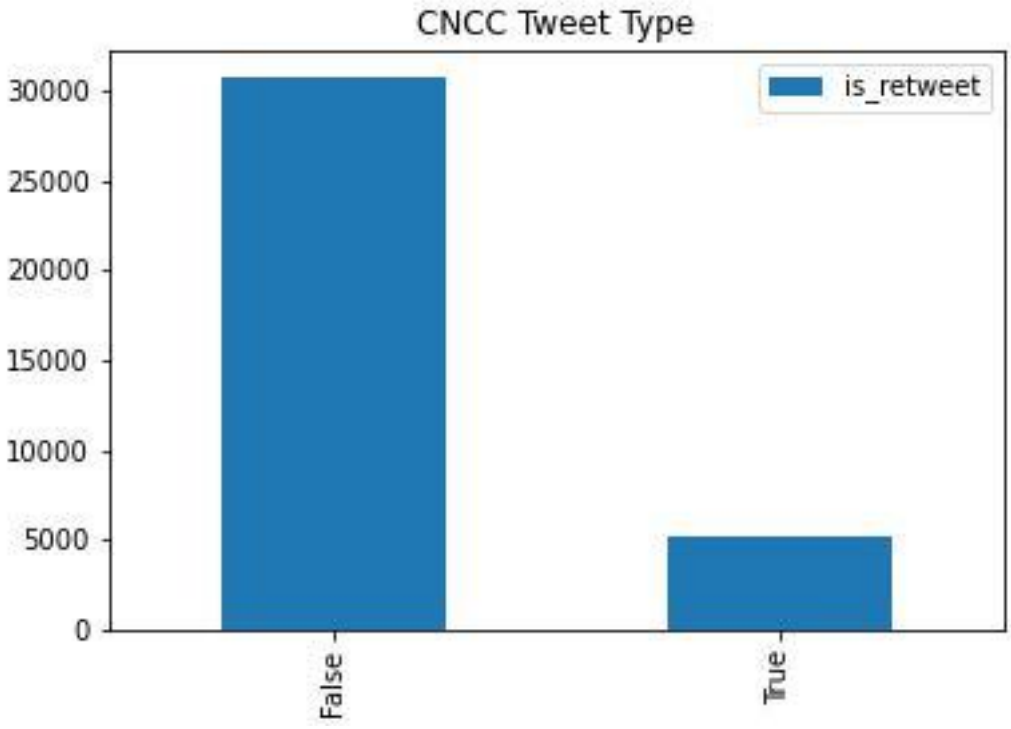
China	Agency	From	Till	Accounts	Tweets	Retweets	Total	English Language Tweets
	CNCC	24/01/2012	29/03/2021	112	30779	5145	35924	4777
	CNHU	20/04/2019	05/04/2021	2047	28567	2702	31269	17579
	Total	**24/01/2012**	**05/04/2021**	**2159**	**59346**	**7847**	**67193**	**22356**

II. CNCC

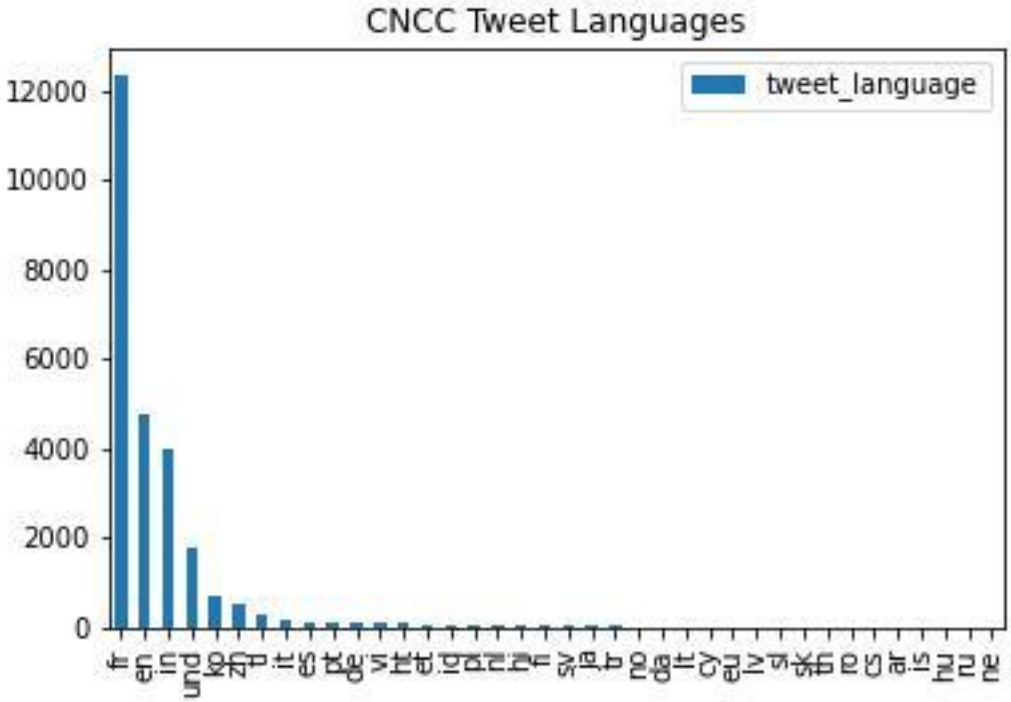
Account Creation Dates



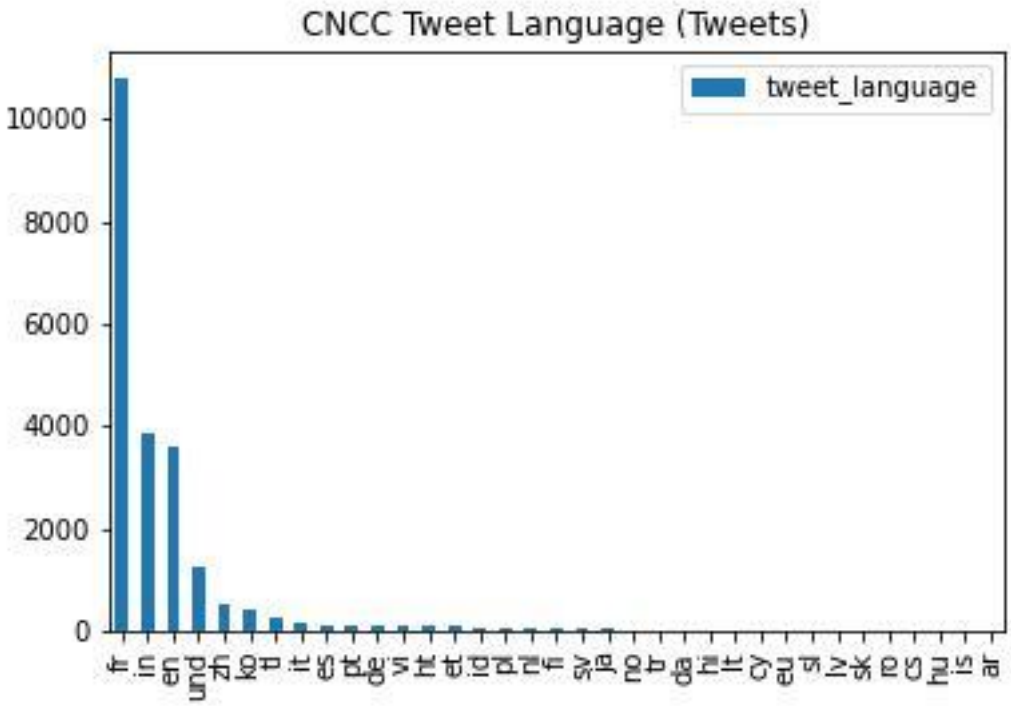
Tweet Types



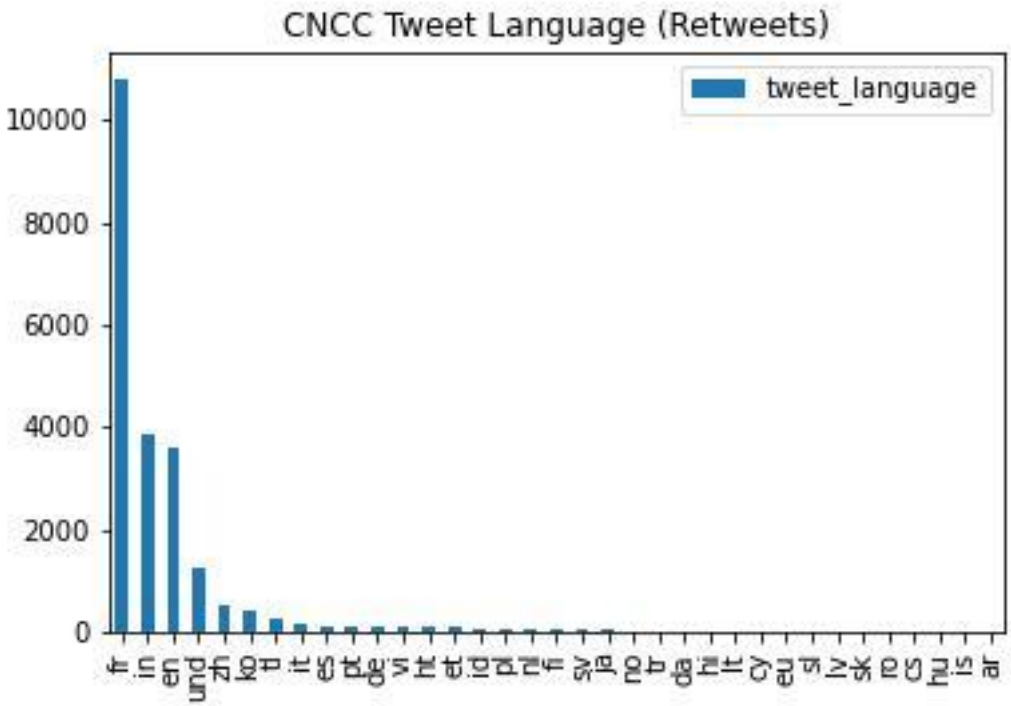
Tweet Languages (Overall)



Tweet Languages (Tweets)

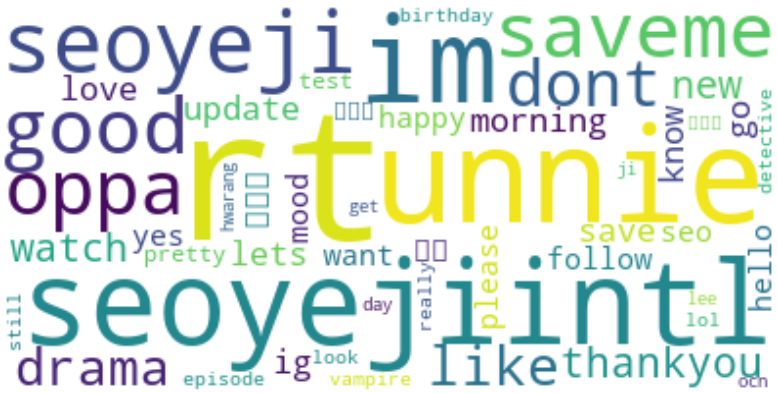


Tweet Languages (Retweets)



Term Frequency

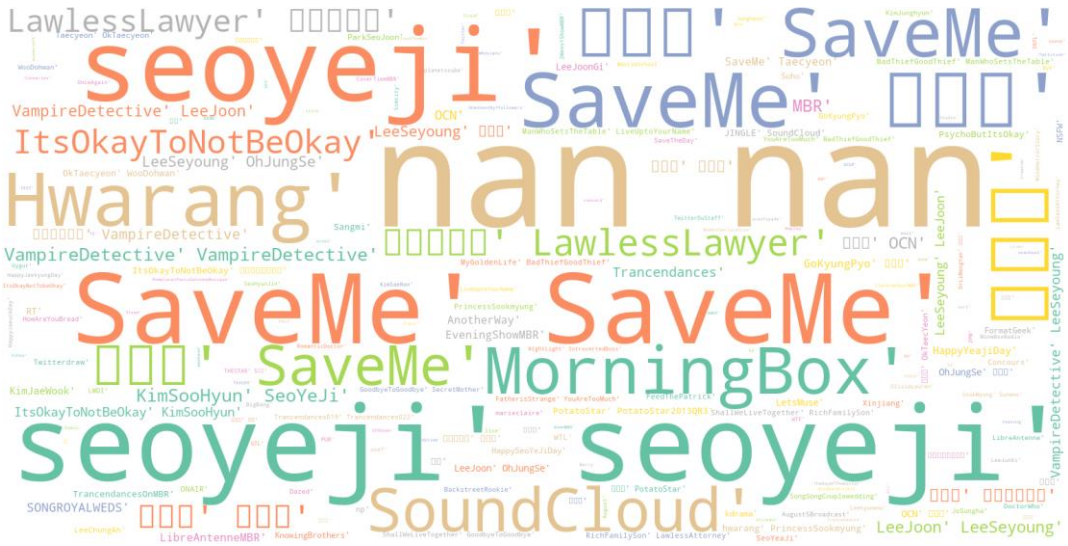
TF–IDF (Term Frequency–Inverse Document Frequency)



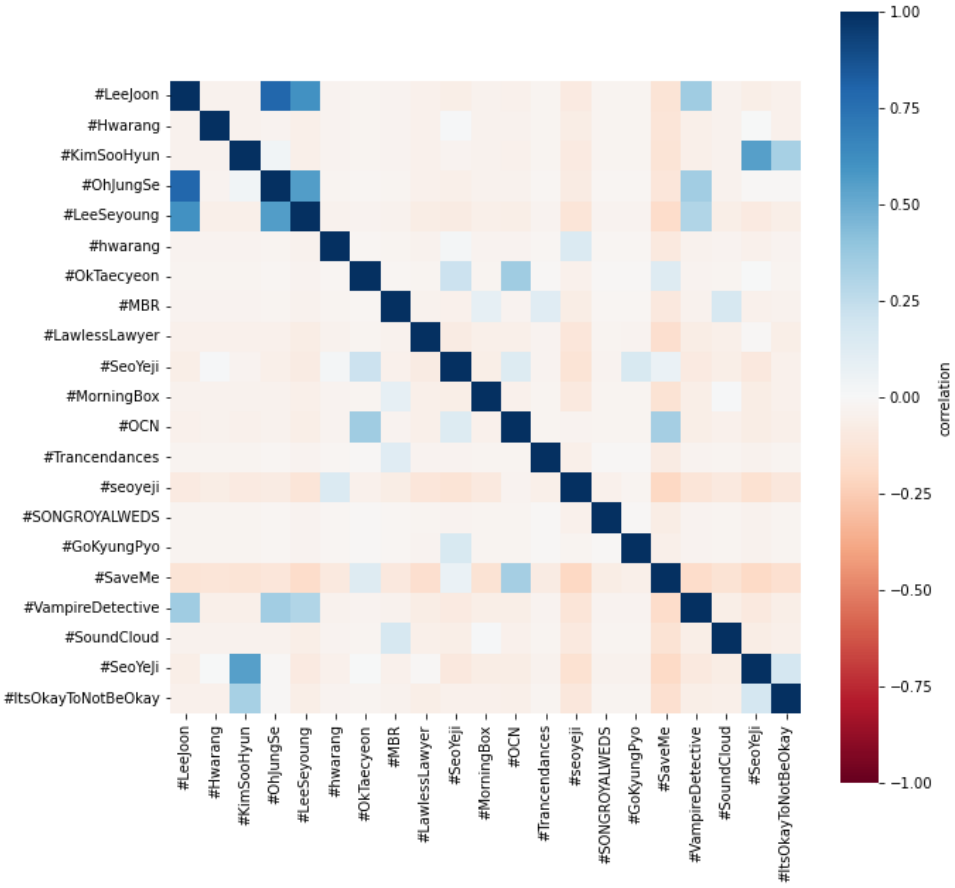
Hashtags Shared

	CNCC Hashtags			
	Most Shared		Least Shared	
Grand Total	['MorningBox']	610	['MorningBox', 'serveur', 'illimité', 'OMGSERV']	1
	['MBR']	534	['MBR', 'spoil']	1
	['蓬佩奥', 'StopXinjiangRumors']	438	['JoyeuxNoel']	1
	['LibreAntenneMBR']	258	['Geekopolis', 'Paris', 'MBR']	1
Tweets	['MorningBox']	564	['PlaylistMBR']	1
	['MBR']	507	['MorniingBox']	1
	['蓬佩奥', 'StopXinjiangRumors']	438	['MAILMBR']	1
	['LibreAntenneMBR']	242	['MorningBox', 'MBR', 'Cadeaux', 'RVF', 'OMGSERV']	1
Retweets	['MorningBox']	564	['PlaylistMBR']	1
	['MBR']	507	['MorniingBox']	1
	['蓬佩奥', 'StopXinjiangRumors']	438	['MAILMBR']	1
	['LibreAntenneMBR']	242	['MorningBox', 'MBR', 'Cadeaux', 'RVF', 'OMGSERV']	1

Hashtags



Hashtags Heatmap

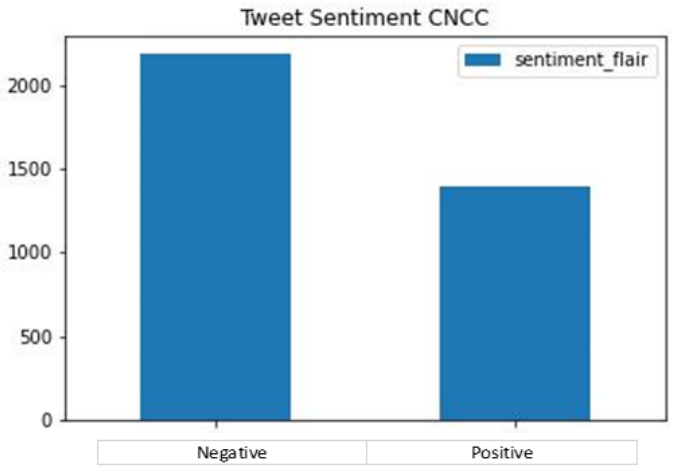


TF–IDF Hashtags

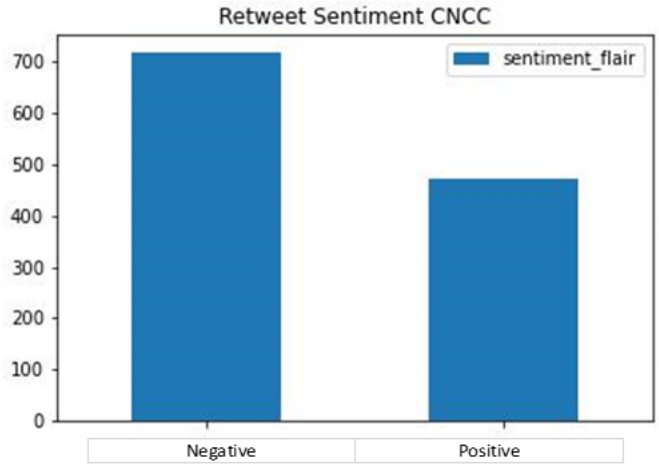
URLS Shared

	CNCC URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['http://mineboxradio.com/']	1159	French Radio Website
	['http://player.mineboxradio.com/']	600	French Radio Website
	['http://www.mineboxradio.com']	318	French Radio Website
	['http://mineboxradio.com']	312	French Radio Website
Tweets	['http://mineboxradio.com/']	1121	French Radio Website
	['http://player.mineboxradio.com/']	586	French Radio Website
	['http://www.mineboxradio.com']	315	French Radio Website
	['http://mineboxradio.com']	285	French Radio Website
Retweets	['http://mineboxradio.com/']	1121	French Radio Website
	['http://player.mineboxradio.com/']	586	French Radio Website
	['http://www.mineboxradio.com']	315	French Radio Website
	['http://mineboxradio.com']	285	French Radio Website

Tweet Sentiment Analysis



Retweet Sentiment Analysis

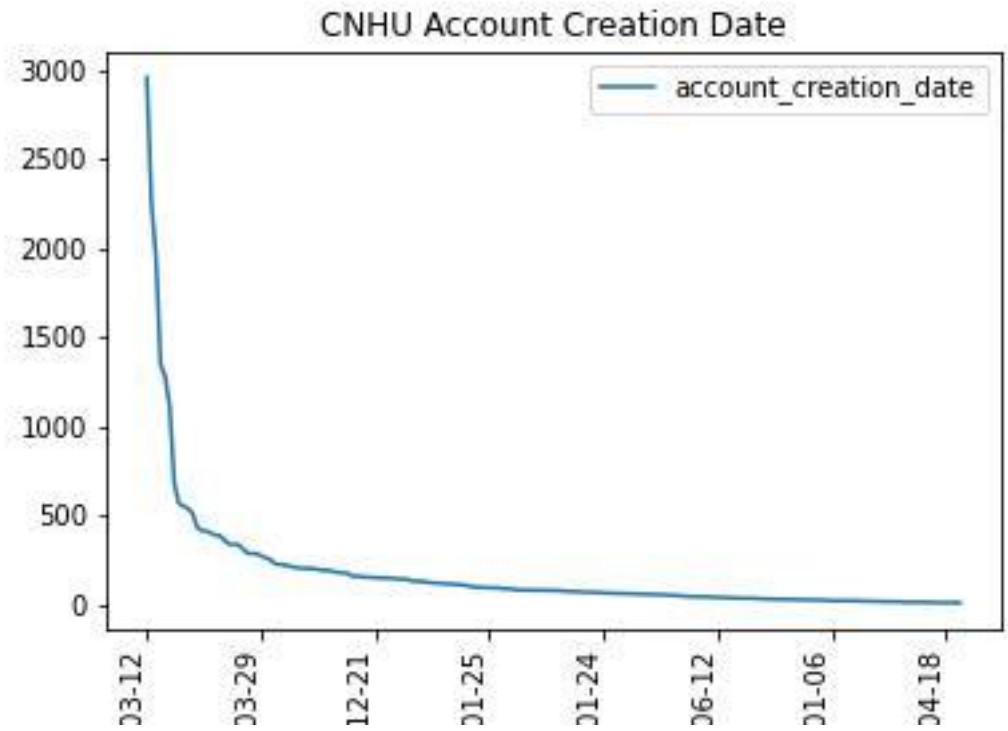


Topics

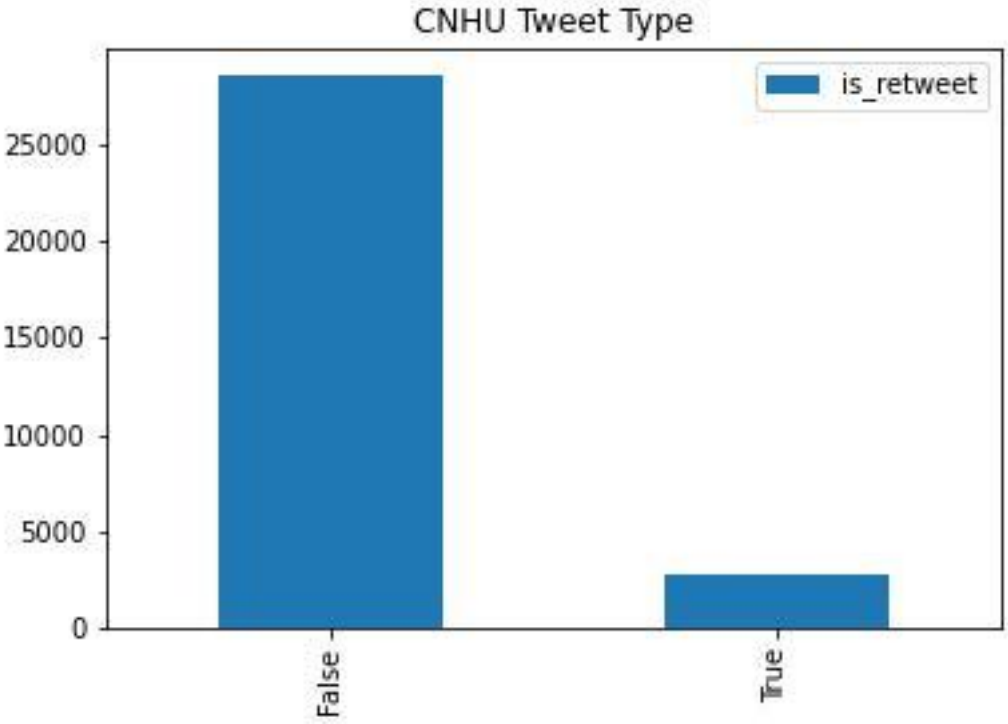
	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	thankyou	94.7	ye	180.1	#seoyeji	259.0	oppa	181.0	...	89.2	#saveme	205.2	😂	184.1	unni	251.7	rt	200.1	look	125.9
1	morn	83.1	drama	148.8	updat	143.1	happi	139.1	get	88.1	#서예지	85.8	watch	158.1	go	130.1	love	155.6	new	72.1
2	mood	76.1	seo	121.3	ig	129.1	want	104.1	yeji	67.1	vampir	84.1	ㅋㅋ	127.0	know	117.4	pleas	110.1	princess	59.1
3	ah	68.6	ji	105.1	good	103.2	birthday	90.1	make	60.1	episod	82.1	like	96.4	let	89.5	think	68.1	le	47.1
4	one	62.1	new	62.1	even	56.2	follow	85.5	hope	58.1	#구해줘	81.1	ㅋㅋㅋ	88.1	pretti	67.1	wait	55.9	sound	44.1
5	#루법변호사	57.1	kim	57.1	sorri	55.1	hello	73.1	smile	49.1	detect	79.1	drama	58.4	gt	64.1	meet	50.1	#hwarang	44.1
6	reall	55.2	beauti	51.1	hard	54.1	much	58.1	peopl	48.1	ep	70.1	join	53.6	great	54.1	like	45.7	face	41.5
7	good	51.0	come	50.8	guy	42.1	thankyou	55.2	take	44.2	still	68.2	help	48.1	girl	53.1	nice	44.4	miss	38.1
8	feel	50.1	film	45.1	see	41.5	wish	54.1	well	43.1	scene	67.9	friend	47.2	mineboxradio	30.1	support	39.1	photo	38.1
9	#lawlessla wyer	48.1	day	44.7	night	39.1	best	42.1	amp	41.1	#leeseyoung	57.1	forget	39.1	alway	29.3	😍	34.1	#soundcloud	36.1

III. CNHU

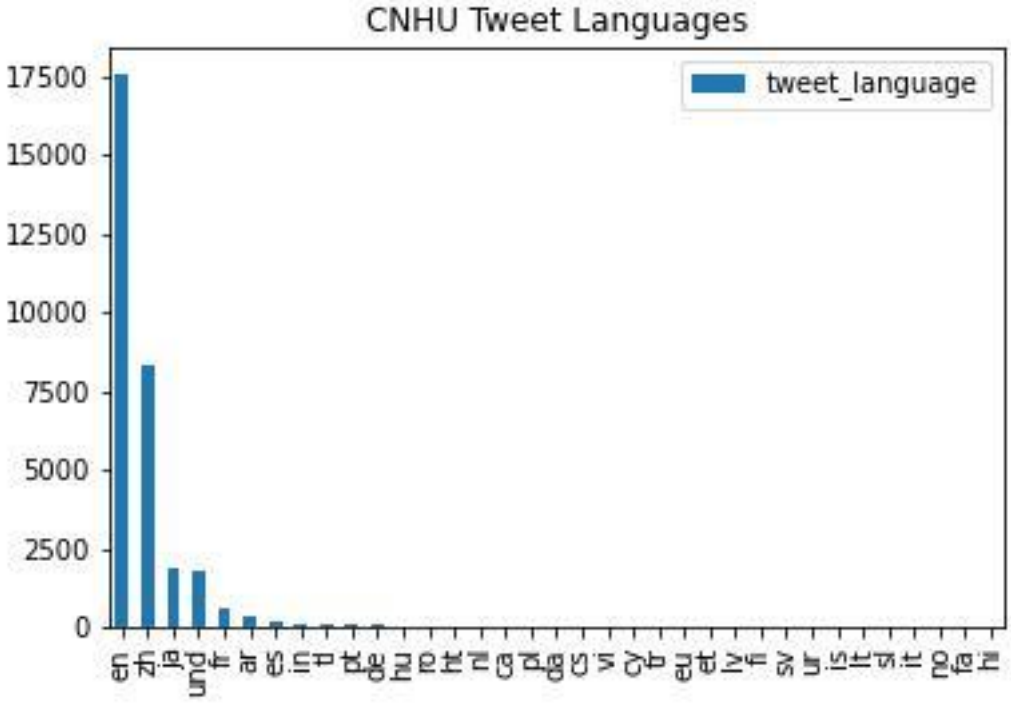
Account Creation Dates



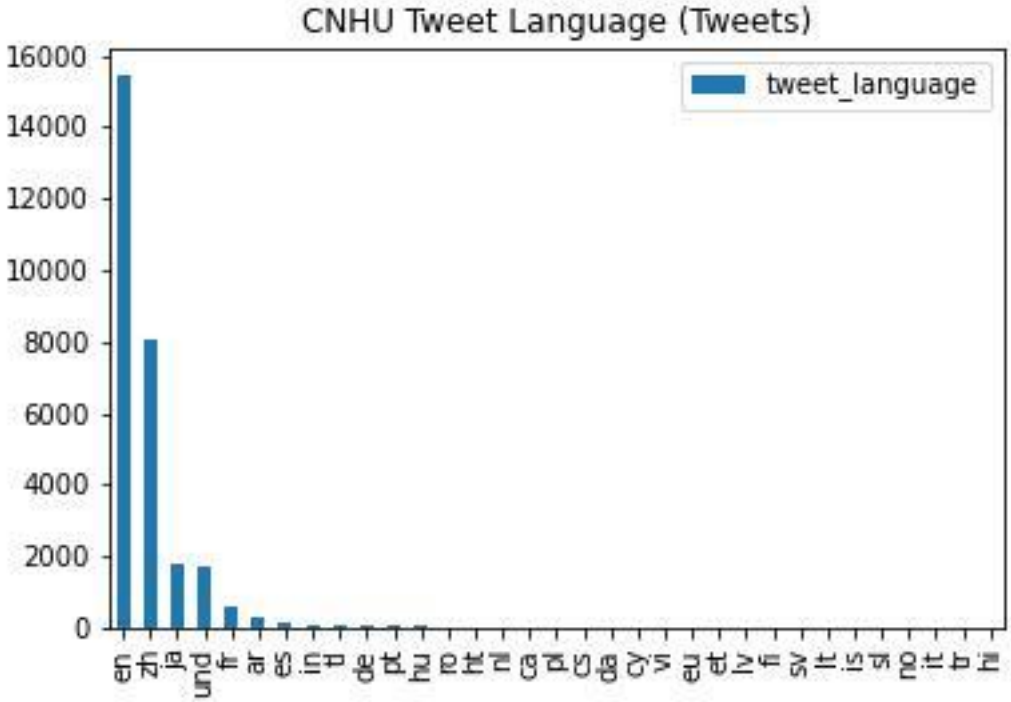
Tweet Types



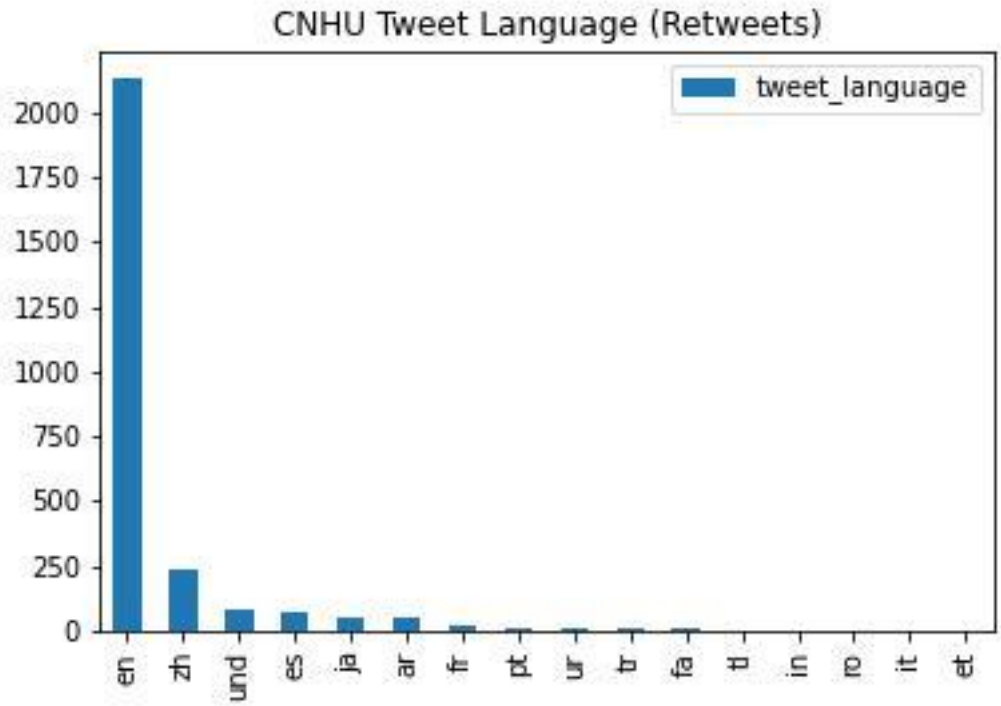
Tweet Languages (Overall)



Tweet Languages (Tweets)

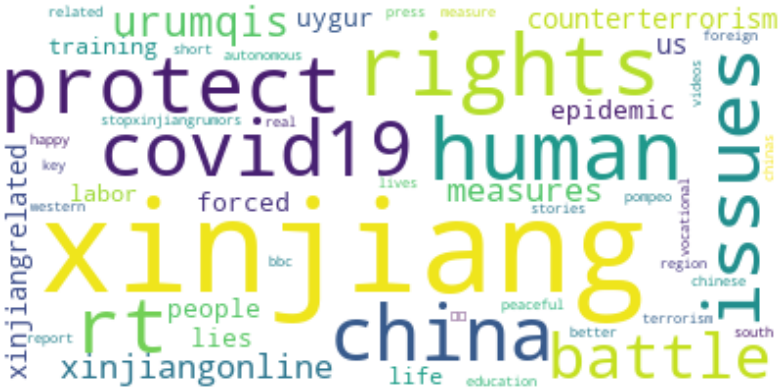


Tweet Languages (Retweets)



Term Frequency

TF–IDF (Term Frequency–Inverse Document Frequency)



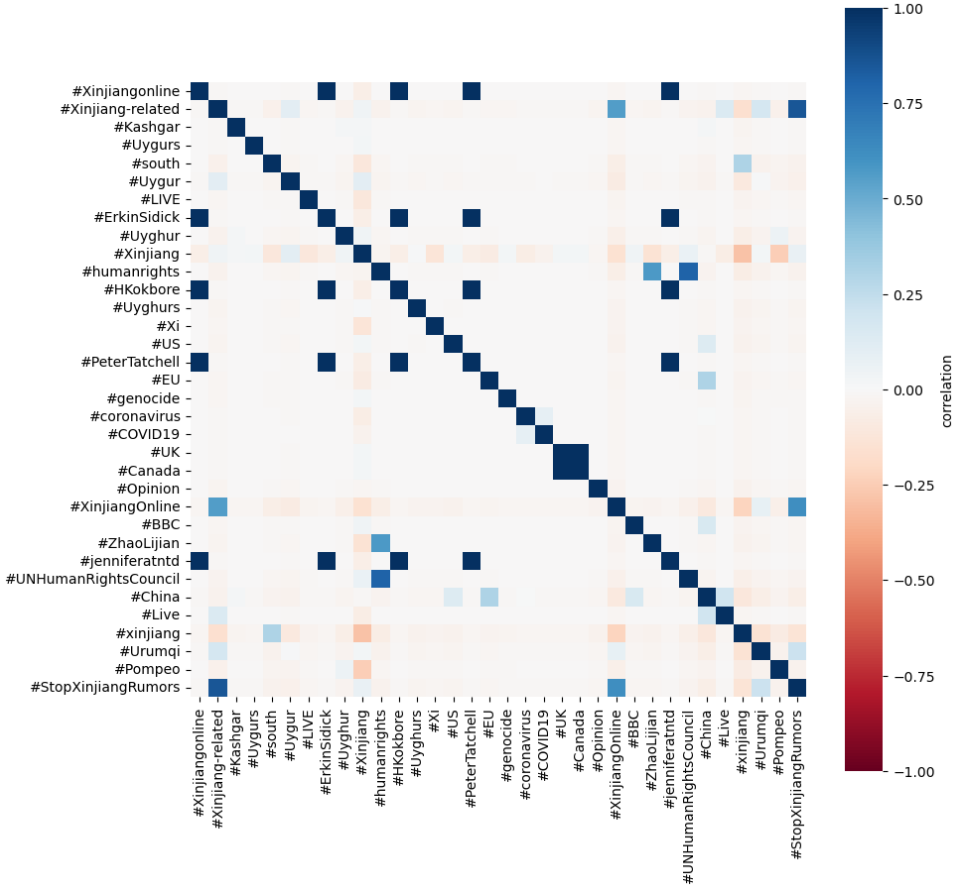
Hashtags Shared

	CNHU Hashtags			
	Most Shared		Least Shared	
Grand Total	['Xinjiang']	3155	['RebiyaKadir']	1
	['xinjiang']	591	['Bitcoin', 'TechnicalAnalysis', 'Forecast', 'PriceAction']	1
	['XinjiangOnline']	497	['BabyBird', 'birds']	1
	['Xinjiang', 'xinjiang']	312	['TLP', 'tlpwelfareactivities', 'tlp_believeislam']	1
Tweets	['Xinjiang']	2711	['穆斯林国家'] (*'Muslim country)	1
	['xinjiang']	591	['新疆', '强迫劳动', 'Uyghur'] (*Xinjiang', 'forced labor)	1
	['XinjiangOnline']	496	['新疆', '维吾尔', 'Xinjiang'] (*Xinjiang', 'Uyghur)	1
	['Xinjiang', 'xinjiang']	312	['新疆', 'America'] (*Xinjiang)	1
Retweets	['Xinjiang']	444	['疫情'] (*'epidemic)	1
	['Pompeo']	85	['coronavirus', 'COVID19']	1
	['XinjiangOnline', 'Xinjiang', 'StopXinjiangRumors', 'Xinjiang', 'Urumqi']	58	['American', 'Trump']	1
	['ZhaoLijian', 'humanrights']	47	['XinjiangOnline']	1

Hashtags

TF–IDF Hashtags

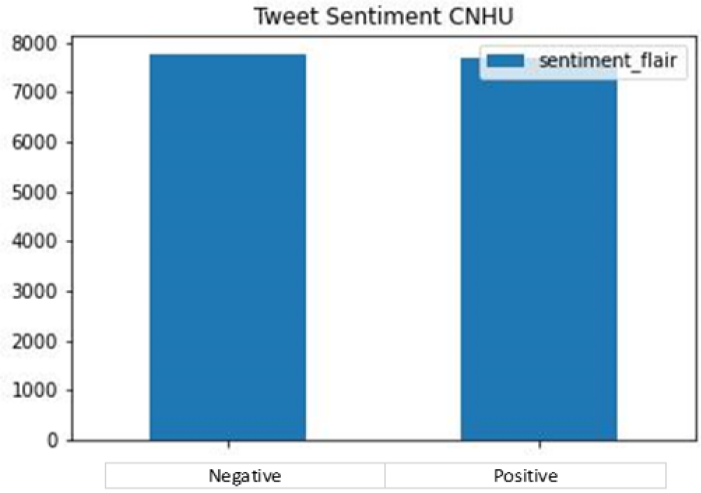
Hashtags Heatmap


#Xinjiangonline
#Xinjiang-related
#Kashgar
#Uygurs
#south
#Uygur
#LIVE
#ErkinSidick
#Uyghur
#Xinjiang
#humanrights
#HKokbore
#Uyghurs
#Xi
#US
#PeterTatchell
#EU
#genocide
#coronavirus
#COVID19
#UK
#Canada
#Opinion
#XinjiangOnline
#BBC
#ZhaoLijian
#jenniferatntd
#UNHumanRightsCouncil
#China
#Live
#xinjiang
#Urumqi
#Pompeo
#StopXinjiangRumors
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
correlation

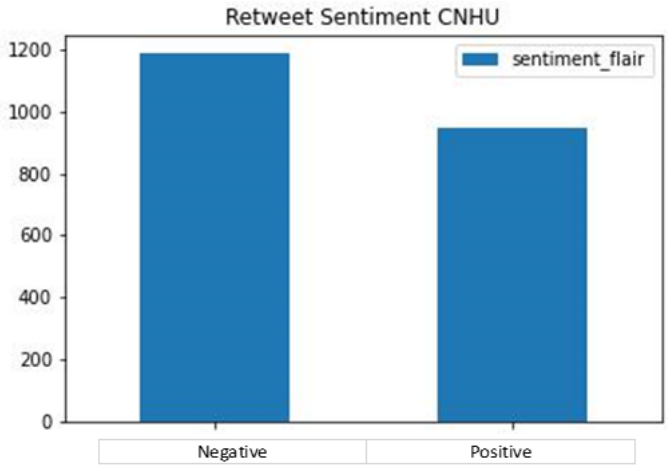
URLS Shared

	CNHU URLS Shared		
	URL	**Times Shared**	**Content**
Grand Total	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583671.htm']	867	News Article: "Xinjiang's counter-terrorism measures protect human rights"
	['http://t.m.china.org.cn/convert/c_x8BDSHzU.html']	435	News Article: "Vocational education and training is key measure to protect human rights"
	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583673.htm']	281	News Article: "Short videos depict peaceful and happy lives in south Xinjiang"
	['http://t.m.china.org.cn/convert/c_XWWIAWDD.html']	216	News Article: "Real life stories from Xinjiang best refutation of US allegations"
Tweets	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583671.htm']	866	News Article: "Xinjiang's counter-terrorism measures protect human rights"
	['http://t.m.china.org.cn/convert/c_x8BDSHzU.html']	435	News Article: "Vocational education and training is key measure to protect human rights"
	['http://regional.chinadaily.com.cn/en/2021-01/19/c_583673.htm']	273	News Article: "Short videos depict peaceful and happy lives in south Xinjiang"
	['http://t.m.china.org.cn/convert/c_XWWIAWDD.html']	216	News Article: "Real life stories from Xinjiang best refutation of US allegations"
Retweets	['https://www.globaltimes.cn/content/1187107.shtml']	33	News Article: "Xinjiang debunks lies in PBS documentary and biased Western media"
	['https://news.cgtn.com/news/2021-01-07/Xinjiang-s-growth-on-par-with-rest-of-China-new-report-'	29	News Article: "Xinjiang University report proves region's growth on par with rest of China"
	['https://www.pscp.tv/XHNews/1vOGwEAWerqxB?t=57m15s']	25	China Xinhua News Video of Xinjiang Regional Government
	['https://www.pscp.tv/w/cyxYTzFBbWp6Z1Zrb2R6RWV8MUJSSmpCcmFMUExKdxYRJp9C6EesPT	24	Xinjiang is a wonderful land" online promotion event

Tweet Sentiment Analysis



Retweet Sentiment Analysis



TF–IDF on Duplicate Tweets

Topics

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topic 9 words	Topic 9 weights
0	#xinjiang	1785.4	one	336.1	relat	992.8	xinjiang	889.1	live	771.8	xinjiang	717.1	region	488.5	#xinjiang	1442.4	xinjiang	354.1	#xinjiang	939.3
1	right	1640.9	better	314.5	xinjiang	973.7	western	517.8	#xinjiang	700.9	report	435.2	#xinjiang	429.3	issu	725.2	stabil	170.2	lie	516.9
2	human	1588.1	...	202.0	issu	927.2	#xinjiang	365.2	xinjiang	609.6	china	364.3	forc	398.9	relat	625.9	polit	168.0	'	345.0
3	protect	1574.1	man	193.0	foreign	470.3	us	339.3	happi	538.0	say	239.7	xinjiang	370.2	china	550.2	top	153.0	good	287.1
4	measur	1532.0	love	185.1	press	420.1	stori	334.6	peac	521.8	offici	231.1	autonom	355.8	xinjiang	547.6	stress	152.1	...	279.1
5	terror	1182.2	like	167.1	confer	396.1	peopl	332.9	video	485.1	right	178.3	uygur	308.8	wonder	321.0	endur	151.1	pompeo	271.9
6	counter	1172.7	work	160.3	#xinjiang	354.0	real	316.5	short	376.1	refut	167.2	china	303.2	amp	306.4	advisor	150.1	state	267.1
7	xinjiang	1136.6	go	158.0	ministri	308.2	muslim	299.0	south	267.1	im	152.1	never	293.1	#新疆	289.1	two	135.1	china	251.0
8	'	1038.6	make	142.1	china	305.9	call	278.9	depict	254.1	...	151.8	labor	253.5	##xinjiangonline	285.9	know	126.9	fact	248.5
9	covid	742.3	friend	141.0	chines	305.8	life	248.1	futur	181.1	extremist	137.0	us	237.5	##stopxinjiangrumors	270.1	fight	111.7	us	226.5