,,

Opleiding Informatica

Applying social network analysis to Twitter data
to identify phases in crisis communication

Lonneke Lommers

Supervisors:
Frank Takes & Iris Yocarini

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

30/08/2022

Abstract

This thesis is about applying social network analyses techniques to social media data. Specifically, this thesis will focus on data of the social media platform Twitter, where the objective will be to identify phases during crisis communication. The thesis will focus on data that was generated during hurricane Irma. This hurricane was a very intense natural disaster near the east coast of Florida which resulted in more than 100 deaths in 2017. To understand if it is possible to identify phases in crisis communication for hurricane Irma, we will analyse the underlying social network structure using a number of network measures, which will be plotted over time. Comparing the social network structures and real-life events indicating actual phases during the hurricane, the results indicate that a combination of the size of the network and its density are indicative of possible phase transitions. This indicates that network analysis may be a fruitful way forward to the understanding of phase transitions in crisis communication.

Contents

1 Introduction

In this section the main subject of the thesis is introduced, by explaining the role of social networks during crises and by providing insight in the way that social networks could be analysed. Thereafter, the research question and the thesis overview are shared.

1.1 The role of social networks during crises

When a crisis or disaster unfolds, communication of vital information is key. When such a situation happens, uncertainty increases and people will often seek information to make them feel "in control" of their situation. This occurs, among other things, by turning to social media platforms for the latest information and updates [SLLdG15]. Social media platforms provide access to specific communities formed based on location-bound connections as well as non location-bound connections, made based on similar interests, expertise's and experiences. Additionally, they provide a platform for those communities to connect, share information and create a network of involved organizations, communities and individuals. Social media platforms like Instagram, Facebook and Twitter are frequently used to share and consume time sensitive information and to communicate with other users. Due to the characteristics of Twitter, having a short textual content form and live-posting abilities, this social media platform is especially providing in the need for high accessibility to time-sensitive information during disasters and crises.

During disasters such as an earthquake, support-providing organizations like governmental institutions use social media to communicate with people who are affected by a certain disaster. Users from communities located in disaster stricken areas ([Fel15]; [Kim8a]) share vital information on their social media channels, allowing their followers to share their information and input. The initial sharing of information in a tweet provides for documentation on the content-related information, additionally the resharing by other users provides data on the communication between users. Resharing occurs when another user reshares a tweet, called "retweeting", or leaves a comment under a tweet, mentioning the original user, called "mentioning". The sequential resharing of a tweet, when documented, provides relational data on the communication between the source of the tweet and the target user(s), creating a link between them. For this reason, social media data naturally includes data formatted as relational data, providing the links between their users and thus the elements needed to construct a network.

1.2 Analysing social networks

Platforms like Twitter produce enormous amounts of data. Every post created by a user is stored by Twitter and contains all necessary information about the profiles of users, the user's tweets, the user's followers and the accounts the user is following, among other things like search history. After a disaster has occurred, we can analyse this database, extract specific data and produce a detailed data-set containing the information shared (tweets), by who and with whom during a disaster plotted on a timeline. To analyse this huge amount of social media data, different tools can be used. Some of these technique use natural language processing techniques to analyse the language used in the posts. For example, if analysing or detecting the nature of the words in the posts itself would be priority, techniques like text mining or data mining, which are techniques to process high-quality information from text could provide insight in the value of the information shared by the users. Or in another situation, like when identifying the subjective information of the posts would be the priority, sentiment mining could be used. Using sentiment mining to determine whether data is positive, negative or neutral could be useful in the context of future works regarding this thesis's subject to further understand the content of the information and intent behind the Twitter users. Even though there are many insights those specific techniques could provide, these techniques use the content of the communication and do not provide network analysis needed to test information diffusion though a network. Using social network analysis provides a way to create and analyse a network from a data set, depicting "the sharing of information", regardless of the content of the posts. Quantifying the relationships (links) between these actors through the documentation of their communication can provide insights, like, into the crisis information diffusion through a social media network. To do this, a network can be created from the crisis communication interactions between actors, using the actors as nodes and the action of sharing as the edges. A 'snapshot' of the network structure at a specific point in time could be taken, which can be used as the input when social media analysis is applied on data collected during a crisis situation. In the end, this could gather insights in the crisis communication between actors and potential key actors at a certain time.

Other studies on the subject of crisis communication show that different key actors can take on different roles depending on demand [BB14]. When the roles shift and the relationships between actors change, subsequently the network structure will change. When looking at two different snapshots of the same network at different times, an analysis could be able to identify changes if there are any. Applying network measures, like average degree assortativity coefficient, average density, number of nodes and betweenness centrality to the data and thereafter analysing them over multiple snapshots over time, could be plotted overtime. All exact measures used are described in the "measures" section of this thesis. Since communication over social media is increasing over time, the insights gained could provide valuable knowledge for future disasters. Early recognition of key actors at all stages of a disaster can help with more efficiently spreading or gathering information for governmental organizations to provide aid during disasters. Recognizing the transition between phases in network structure can provide insights into the dynamics of crisis communication via social media and how we can use those platforms more effectively in future disasters.

1.3 Research question

During a disaster, different communication phases can be identified [SLLdG15]. When a crisis shifts into a different phase, the demand for information changes. In some phases this is the location of urgent care provided by emergency or medical organizations, in other phases it is about the guidelines shared by governmental organizations. Different phases provide different key actors, fulfilling different roles in the crisis situation. The structure of the network of communication between these actors could shift over time, depending on the change in the nature or demand of the information needed. Hence, this project will be focusing on the following research question applied to data of hurricane Irma:

How can different phases in crisis communication be identified through analysing changes in network structures underlying Twitter crisis communication?

1.4 Thesis overview

The remainder of this thesis is organised as follows; Section 2 will provide insights in the background information of this thesis; Section 3 will focus on the data used for the research question; the applied method including the results will be described in Section 4 and 5; next, the conclusion and possible future research are described in Section 6.

2 Background

This section will provide insights in related work done on the topic of this thesis. Next to that, additional information about the definitions and terminology used in this thesis will be explained.

2.1 Related Work

In this section, we give a general description of the academic landscape in which this thesis is rooted.

A similar research that focused on the variability in Twitter content across the stages of a natural disaster, discussed the implications for crisis communication [SG16]. Over the years, the topic of using social media platforms to analyse how information is spread has been steadily growing in popularity. Taking inspiration from their time series analyses for the approach of this thesis resulted in the basics of the approach used in this thesis, such as applying density measures. Little is known about the ways in which social media, such as Twitter, function as conduits for information related to crises and emergencies. The current study analyzed the content of over 1,500 tweets that were sent in the days leading up to the landfall of Hurricane Sandy. Time-series analyses reveal that relevant information became less prevalent as the crisis moved from the prodromal to acute phase, and information concerning specific remedial behaviors was absent. Therefore, the research indicates that the social media has introduced a key ingredient, and a potentially unfamiliar variable, into the practice of crisis communication. In this paper, the authors critically assess the social media milestones related to Hurricane Sandy according to situational crisis communication theory (SCCT) (Coombs, 2007). Further, they discuss the crisis lifecycle of Hurricane Sandy with regard to the potential implementation of the STREMII model of social media crisis management, a proposed model originated through this research application. This original model develops from lessons and best practices discovered in historical and contemporary cases of social media crises and crisis management. The researchers acknowledge potential limitations and describe steps for further development of the model through research, all the while recognizing the powerful and paradoxical role of social media in the crisis management process. In reflection of Hurricane Sandy, further qualitative and quantitative examinations of crisis events are encouraged to evaluate the STREMII model continually in the dynamic social media climate and across the vast facets of crisis communication [SLLdG15].

2.2 Definitions

Over the years, data on crisis communication on social media platforms has been collected and documented, creating a data set containing verifiable information of the online communications of all parties affected by specific crisis situations. From this data, the connections between individual data objects in these communities can be structured as a network. This network consist of nodes and edges. The nodes hold information about the node itself and would usually represent a person in the network, while the edges connect the nodes with other nodes, holding information about the relationship between those nodes.

2.2.1 Social network analysis

To analyse the data set, social network analyses will be used. Social network analysis was designed to discover relationships between social entities. In our case, these social entities are Twitter users from the data set and their relationships, quantifying their communication on Twitter. Social network analysis could be applied in several ways of working [CLEZG10] [OG12], but in this thesis the following will be explored:

- Computation of network measures that provide a local (actor level) and global (network level) description of the network, from this point on referred to as ”local measures”and ”global measures”.
- Graphical visualization of the network

2.2.2 Measures

The network structures of social networks can be measured using network measures. Changes of these measures over time indicate a change in the network structure and could possibly be an indication of a phase transition. The relation between the measures used in this thesis and the potential identification of a phase transition will be explained in the Section 2.2.2. The following measures are used in this thesis:

Global Measures

- Global average clustering coefficient
  This measure is the extent to which a node forms triangles with its neighbours.
- Average degree assortativity coefficient Extent to which nodes with similar degrees are connected.
- Average density
  The proportion of actual connections between nodes of a network and the maximum number of nodes possible in that network.
- Number of nodes
  Total number of nodes in the given network structure.
- Number of edges
  Total number of edges between nodes in the given network structure.

Local measures

- Betweenness centrality

  Betweenness centrality is used to measure the number of shortest paths going through a node. Betweenness centrality is measured between 0 and 1. The closer betweenness centrality is to 1, the more paths between two nodes run through the measured node. If the betweenness centrality of node B is 1, all paths from any node A to node C run through node B.

2.2.3 Phases in disasters

In the literature, there is sufficient evidence for the existence of distinguishable phases of communication regarding crisis events. Nevertheless, there is lack of a true consensus about the nature of and number of phases identified in crisis communication. The phase models that are widely used are described in the next sentences. The three-phase model [Coo21], the four phase model from Fink's (1986), Crisis Life Cycle model to mass media content[Fin86], the five phase model [Mit94] and the STREMII model [SG16]. Although these different phase models define different phases, in general each model distinguishes between a pre-disaster phase, an immediate disaster phase, a "solution finding phase" and the learning phase broadly.

From these three previous works, we distil three different phases that can be recognized in crisis communication [SLLdG15]:

- The prodromal time
  This is the period before a disaster. In our case of hurricane Irma, this period could potentially be identified in the data set by tweets which are about an earthquake that could be reaching an area soon, but there is no active danger yet. When uncertainty is high, people will often seek information by turning to social media platforms for the latest information and updates, creating an initial group of users communicating with organizations providing information [SLLdG15].

- The time during disaster
  This is the period during a disaster. In the four phase model from Fink, this is classified as that cute stage. In this thesis, this model is not used, there is note to the nature of this stage. The time during disaster starts with an increase of urgency. In our case of hurricane Irma, this period could potentially be identified in the data set by tweets, which are all about crisis communication.

- The time after disaster
  This is the period after a disaster. In our case of hurricane Irma, this period could potentially be identified in the data set by tweets which are about the recovery of the area after the hurricane and the final numbers of people that were affected by the hurricane.

3 Data

This section will specify the context of the data used while conducting research for this thesis.

3.1 Data description

In this thesis, we will be looking at data from hurricane Irma. Originating from the west coast of Africa on the 27th of August, making her way across the Atlantic Sea and making landfall in the Caribbean and The East of the United States of America. Hurricane Irma was a category 5 hurricane and caused most of her destruction during between the 30th of August to the 12th of September 2017 to Cape Verde, Leeward Islands (especially Barbuda, Saint Barthélemy, Anguilla, Saint Martin and the Virgin Islands), Greater Antilles (Cuba and Puerto Rico), Turks and Caicos Islands, Jamaica, The Bahamas, Eastern United States (especially Florida). At that time, hurricane Irma was considered the most powerful hurricane, later surpassed by hurricane Dorian in 2019.

3.2 Data collection

While this crisis unfolded, people took the conversation to Twitter while using hashtags like #Irma and #HurricaneIrma. The crisis communication exponentially increased and was documented by Twitter all the time. Before the data collection part will be explained in more detail, it's relevant to know that Twitter is a privately held commercial company. Therefore, some specific data sets are not available free of charge when directly requested from Twitter. In this thesis, we focused on the data sets which were available free of charge.

This data set is collected between the 1st of September, starting from 11:44 AM until the 13th of September 13 at 2:23 PM using the following hashtags/topics as key indicators to search for the most relevant data:

- Hurricane Irma
- #Irma
- #HurricaneIrma
- #IrmaStrong

The input for the data set on hurricane Irma is collected from other scholars [Lit20]. The data set was publicly made available for other purposes like research. The data set consists of tweet ID's which refer to tweets that were published during the time and location of hurricane Irma. Twitter's Terms of Service do not allow the full JSON for datasets of tweets to be distributed to third parties. However, they do allow datasets of tweet IDs to be shared. Therefore, we collected the tweet ID's in text form to continue with this specific data set and to make it usable for the intended analysis of this thesis. These tweet ID's consist of a number that is also represented in the URL of the tweets. They are the link to finding the entire tweet and its content and these tweets contain as much information as the source, thus we found this data as extensive enough to be used for this research. To receive the complete contents of these tweets, we can use a Twitter API that releases this data based on the requested ID. In this thesis, we used open source software called "Hydrator"[Hyd20], connected with a Twitter account providing the tweet IDs as an input source. If the contents of a tweet ID is not available, e.g. the tweet is deleted, the Twitter API will filter these tweet IDs out of the output data.

The data used in this thesis needs to consists of:

- The tweet ID.
- The user IDs of the interacting users of the tweets, including:
  - The poster of the tweet.
  - The quoted user, if the tweet is a retweet.
  - The mentioned users, if there are people mentioned (@username).
- The timestamp from when the tweet was posted.
- The content of the message.

4 Methodology

In this section we will discuss which methods were used.

The data set, as mentioned in Section 3 is used to construct a network as described in Section 4.1. The network measures of the tweets will be computed over a timeline and plotted overtime, measuring the differences in those network measures to see how and if the structure of these networks differ over multiple points in time. When comparing those results to the real life timeline of the crisis situation and the expected phase transition points for the disaster, we can compare the actual structural changes in the network with the phase transitions as explained in Section 4.2.

4.1 Network Construction

The data set described in Section 3 provides the elements needed to build a network structure, as described in the network theory elaborated on in Section 2.1.1. To build a network structure based on crisis communication between Twitter users, the nodes and edges of the network must be extracted from the data set. The edges are constructed from the user crisis communication using Twitters tagging system mentioned in Section 1. The nodes are users involved with the interactions captured and additional relevant attributes the profile of a user provides. The edges have associated data concerning the relationship between source and target, as the timestamp of a tweet.

4.2 Approach

Plotting network measures by timestamp in a graph visualizes the changes of the network structure over time. If a measure value changes between two moments with a noticeable difference between them, which could indicate a trend or diversion from a trend, it could tell us that this a significant point in time for the overall network. Using the measures mentioned in Section 2.1.3, two different strategies were chosen for the subsequent measure calculation. Some measures are calculated per node, such as clustering coefficient and the assortativity coefficient. These require an extra processing step to compute them as one value per network and are calculated for every node, summed up and divided by the number of edges per time frame to get a clear input for the analysis and corresponding measures.

4.2.1 Timeframes

While using a movable 'window' functionality, the measures of the nodes in the network structure are calculated within the time frame viewed by this window. When the network measure of nodes in the window are calculated, the window will be moved by an interval value. The window functionality process is decomposed in the appendix Section 7.1 for a step by step reference.

The value determination of the window and interval values is influenced by the complexity of the measure calculated, adjusting the coding and testing the stages of this thesis. Given the scope of the window, a Pandas edge list can be constructed containing all nodes in the given time frame.

The calculated measures together with their respective time frame of the window as reference for the network used in the calculation, will be plotted in a graph. For example, the density of this example Figure 1. At this point, we are able to visualize the measures over time, as the data set

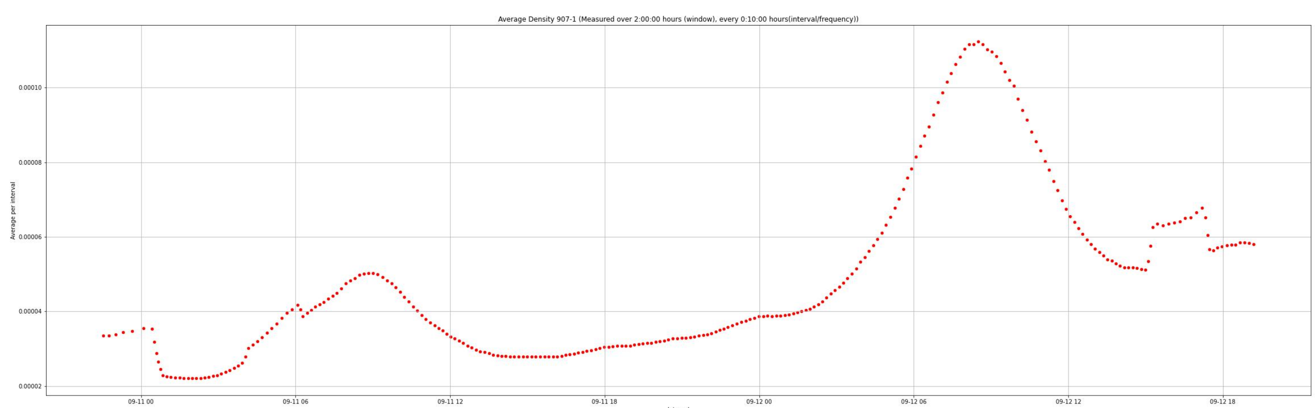

Figure 1: An example of average density plotted over time

can be linked to the timeline of the disaster by adding key timestamps linking to known events from hurricane Irma.

4.2.2 Timeline of the hurricane

One of the key objectives of this thesis is to compare the tested measures against a verifiable timeline of real-world events. The data set we have starts at the 1st of September, starting from 11:44 AM until the 13th of September 13 at 2:23 PM in 2017. The timeline of the data, concerning the communication around hurricane Irma, gives a timeline of circa 13 days and gives us around 17 million tweets to look into. To link the timeline to the data set and to link it to the real life events, a timeline was used with a rough estimation timestamp of the location at some important points of the hurricane's trajectory, as depicted in Figure 2.

Affected Area Name	Timestamp
Barbuda	2017-09-06 7:00:00.000+00:00
St. Martin	2017-09-06 13:00:00.000+00:00
Virgin Islands	2017-09-06 19:00:00.000+00:00
Puerto Rico	2017-09-07 1:00:00.000+00:00
Dominican Republic	2017-09-07 9:00:00.000+00:00
Turks and Caicos	2017-09-08 1:00:00.000+00:00
Great Inagua Island	2017-09-08 7:00:00.000+00:00
Bahamas	2017-09-08 10:00:00.000+00:00

Figure 2: Timeline with approximate timestamps of hurricane Irma's trajectory

4.2.3 Network measures for capturing phases

Average global clustering coefficient
The clustering coefficient is used to indicate the degree to which the nodes in the network tend to cluster together.[BP16] The average clustering coefficient indicates the degree of clustering for the entire network, and the more dense the connection in the network are, the higher the clustering coefficient. Empirical evidence shows that in social networks, as well as real world networks, nodes tens to be clustered together in groups with a relative high density of links. Building on this, at first, the clustering coefficient is expected to be lower When in a crisis situation, uncertainty increases and people will often seek information by turning to social media platforms for the latest information and updates. [SLLdG15]. These updates could start out with more general information, provided by governmental organisations. Later, the crisis progresses and when the information seeking efforts are more locally bound, it could be possible for the clustering to increase and the communication within communities to increase.

Density
Density is the number of edges in the graph compared to the total number of edges possible in that graph, and provides a measure to express how densely the graph is connected with regard to edge

connectivity. This would translate to the number of links between actors, compared to how many interactions between actors there are possible. When density in high, the nodes in the network have more connections with the other nodes in the network and makes the network densely connected. When the density is low, the nodes in the network connect less with other nodes. Thus, when the density measure calculated on our dataset is low, the users are not as densely connected.
People seeking out information when uncertainty is high, makes it more likely for them to gravitate towards information shared by governmental organisations. When the crisis moves to an acute stage of crisis communication, we expect the density to change, due to the changing physicality of the hurricane. When a crisis is in proximity to the user's location, priorities about the nature of the communication might change, resulting in a structural change of the network.

Average degree assortativity coefficient
Assortativity is often used to capture the correlation between nodes. The assorativity coefficient is the Pearson correlation coefficient of degree between pairs of linked nodes. When the degree assortativity coefficient of a node is high, the node mostly connects with nodes with a similar number of connections. For lower values, we expect a large part of the information shared by users to be sourced by a few bigger actors like governmental organisations in a pre- active disaster period. Subsfequently, expecting a turning point when users have been provided and have consumed the 'bigger picture' crisis related information and are communicating more with lower level nodes, and thus expecting the average degree assortativity coefficient to increase.

Number of nodes
The number of nodes measure contains the number of nodes per network within the given timeframe. This calculation is done for every timeframe in the data set. Additionally, these values are depicted in a graph for every timestamp. Since the nodes are representative of the Twitter users involved in the crisis communication about Hurricane Irma, this graph visualizes the course of the quantity of Twitter users involved in crisis communication.

Number of edges
The number of edges contains the links between nodes per network within the given timeframe. This calculation is done for every timeframe in the data set. The number of links betweens the nodes are a representation of the number of relationships between involved users. Hence, the number of edges corresponds to the number of times information is shared, quantifying crisis communication.

Betweenness centrality
The betweenness centrality measure is used to measure the amount of influence the nodes have over the network. Aiming for a single value to computation for comparison to other measures, the betweenness centrality was computed for every node for every network in each timeframe, summed up and divided over the number of nodes in the network, calculating an average betweenness centrality measure. Betweenness centrality is a local measure, measuring on a node level. Computing this measure, could indicate how much influence all nodes have on each other in terms of how densely the nodes are connected.

5 Results

In this section the results will be discussed to understand how these contribute to answering the research question posed in Section 1.

5.1 Experimental Setup

We use Matplotlib [Hun07] to visualize the calculated measures, with “Average per interval” as value on the vertical axis, referring to the average of the measure stated in the title of the graph, and the time on the horizontal axis starting from 11:44 AM until the 13th of September 13 at 2:23 PM. The vertical lines in the plot indicate reference points of the approximate location of hurricane Irma with their abbreviated names, fully stated in the legend in the right upper corner of the graph. The window used while computing the measures contains a time span of 2 hours, and the interval of 0.5 hours. This results in an overlap of nodes of 1.5 hours time span, resulting in a more averaged out course of the graph and making the alignment of the point less sparse, showing the trend of the underlying structure.

The computation and extraction of these lists are executed using a Python script [Fou20] with Pandas data frames [dt20]. The nodes and edges will be used to construct the network using NetworkX [HSS08], and subsequently calculate the measures combining NetworkX, Python and Pandas (see Section 4.2.) Gephi is an open-source network analysis and visualization software package, and is used for both the analytical as the visualization part in this thesis. For example, using Gephi to visualize a network with nodes and edges using the data set provides us with a network visualization similar to Figure 3 as stated below.

Figure 3: An example of how a network between nodes (users) and edges (the interactions) constructed using our dataset, looks like. Made using Gephi [BHJ20].

5.2 Graphs per measure

In this section, we interpret the network measures described in Section 4.2

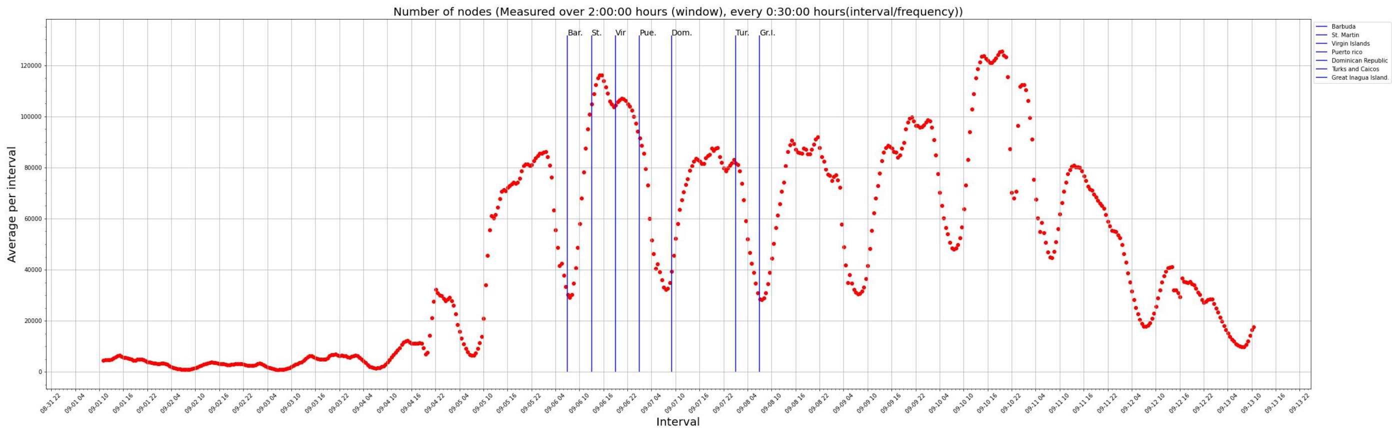

Figure 4: Number of nodes per timeframe.

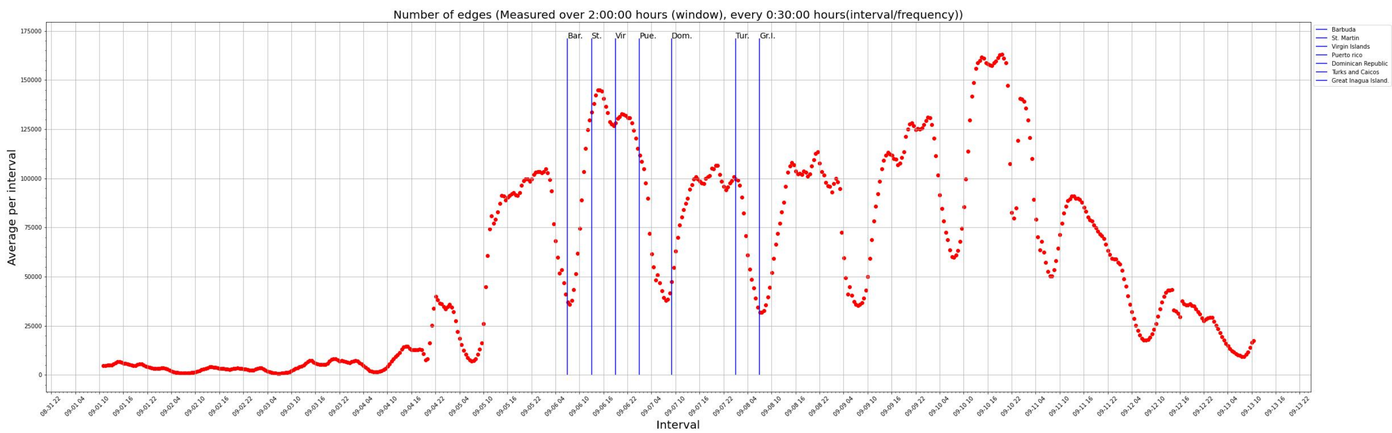

Figure 5: Number of edges per timeframe.

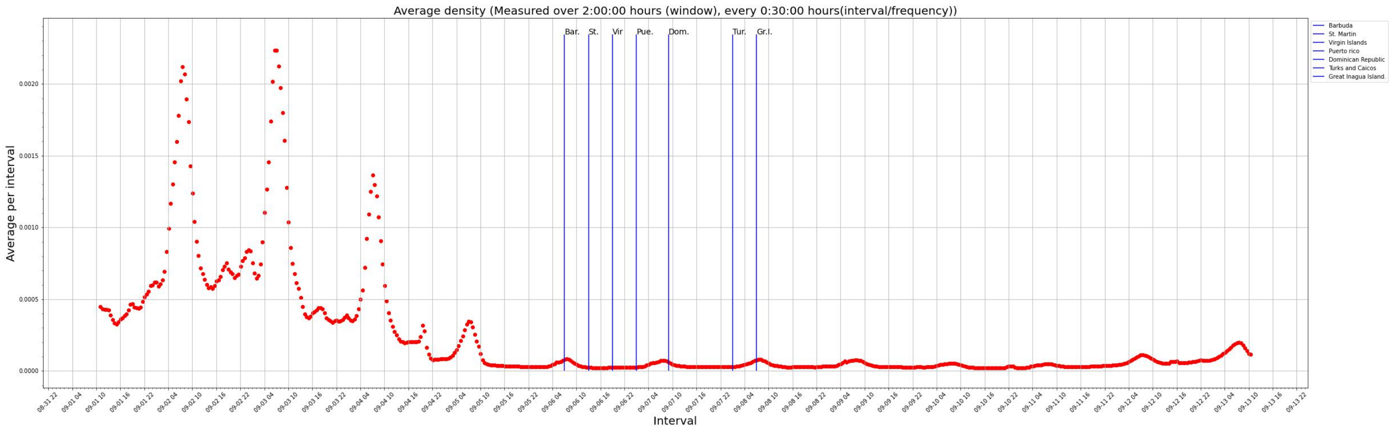

Figure 6: Density per timeframe.

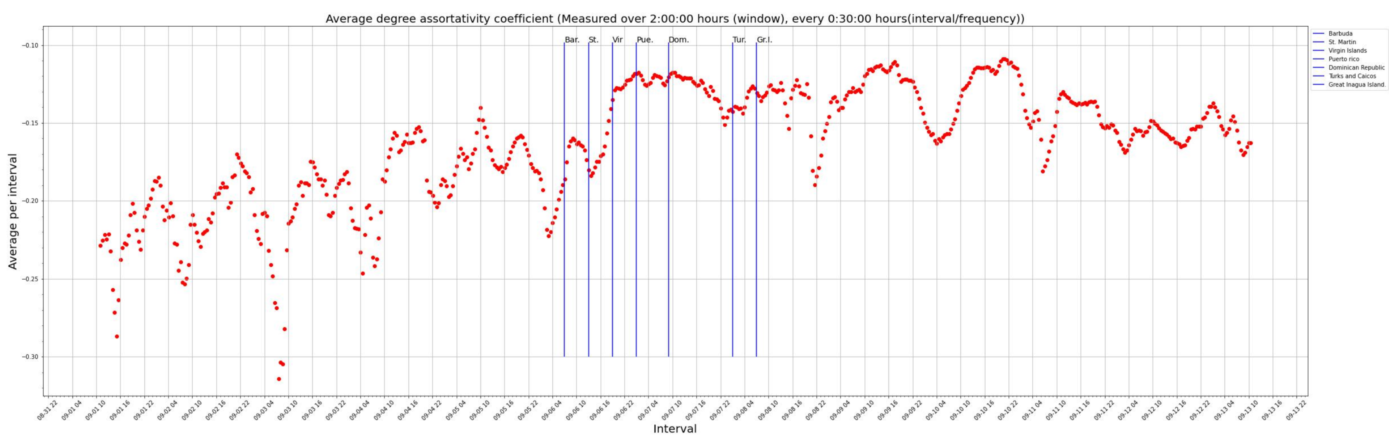

Figure 7: Average degree assortativity coefficient per timeframe.

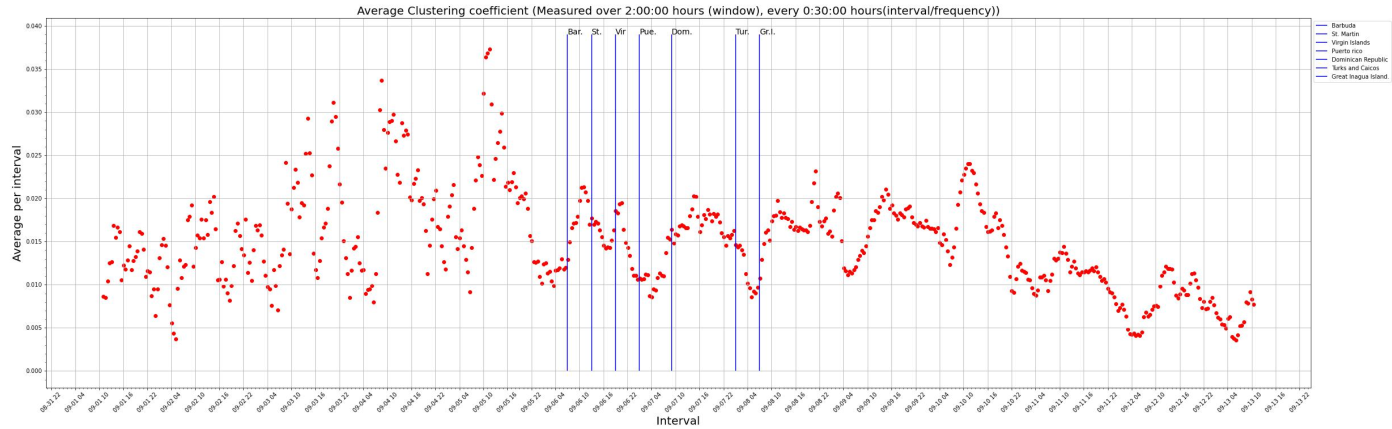

Figure 8: Average clustering coefficient per timeframe.

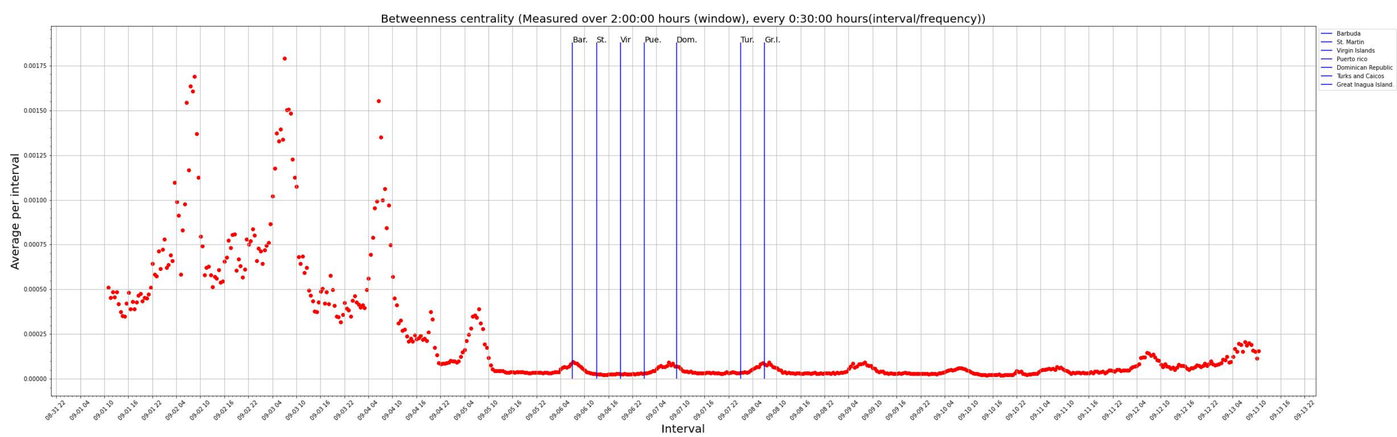

Figure 9: Betweenness centrality per timeframe.

From the above stated figures, the following observations can be made:

- When comparing the shape of Figure 4 and 5, the number of nodes and the number of edges respectively, have a similar shape and development over time, as expected Section 4.2, differing in their overall quantity on the y-axis, both measuring a correlated but not exactly the same measure.
- The nodes and edges both, for the first time, drastically increase within the timespan of 4:00 04-09-2017 and 4:00 05-09-2017, which represents an increase in communication expected in Section 4.2.
- There appears to be a pattern in all measures, corresponding to the day/night rhythm expected from users in the same and neighbouring timezones. During the day, the users are more active on Twitter compared to the user activity at night because the users might be asleep at night.
- The average density depicted in Figure 6 visually decreases within the timestamp of 4:00 04-09-2017 and 4:00 05-09-2017.
- From Figure 7, Figure 8 and Figure 9 no useful conclusions can be drawn.

5.3 Discussion

While comparing the charts, as we can see in the previous subsection, a pattern seems to be visible in every measure graph. We see a clear pattern, in a day-and-night rhythm, explaining the activity and inactivity of users on Twitter measured in the observed measures. However, more importantly, we can generally see a change in the pattern, some figures more clearly than others, around 7 am on the 6th of September, the red line and blue line are crossing each other for the first time. Every chart consists of a red line, which is representing a specific measure, and several blue lines, which are representing the moment in time when the hurricane arrived at residential land for the first time in a specific area. The pattern that we can identify in every chart, can be summarized as follows: when the hurricane starts moving closer to land, the density and corresponding line in the chart start to follow a different trend. This way could either be a situation where the line follows a steep increase, such as Figures 4 and 5 or a steep decrease, such as Figure 6. Before moving to any conclusions, as we will discuss in Section 6, the previous named insight seems to be an indicator of a change in behaviour (indicating a new phase) of Twitter users at the moment that the hurricane is (almost) entering a residential area. Changing the nature of the information the members of the network are searching for.

Observing that the nodes and edges quantity increased, indicating an increase in communication and users. Looking at the density of the network over time, we see a relatively higher density during the earlier stages of the timeline, which late decrease substantially at a given point. The increase of edges and decrease of density confirms expectations made in Section 4.2. The network as a whole appears to be less densely connected, even though both the number of edges and the possible number of edges will increase. The difference in these both these measures, between 4:00 04-09-2017 and 4:00 05-09-2017, indicates a significant change in crisis communication in the network structure.

6 Conclusion and Future work

In this section, the main conclusions of this thesis and potential future research are discussed.

The objective of this thesis was to gather insights on the possibility to identify different phases in crisis communication by analysing the underlying communication and data interactivity on the social media platform Twitter.

By studying the results that were achieved in Section 5 of this thesis, it became clear that there is definitely a change in the network structure constructed from the crisis communication during hurricane Irma, in particular once hurricane Irma moved closer to inhabited areas. Due to the visible shift in number of edges and the density, representing the communication between actors and the connectivity between actors, we see an indication that crisis communication shifts from the prodromal time to the time during disaster.

To fully answer the research question, we believe that two pieces of additional research should be performed. Firstly, we believe that this research should be performed on multiple disasters. Since this research was only focusing on one specific disaster, we can not conclude that a similar pattern would be visible for disasters in other areas. Therefore, in future work, the data set could be extended with multiple variations of other disasters. Secondly, we believe that it should be researched if the pattern that we noticed, as stated in the Section 5, is uniquely applicable for disasters. As an example, it could be the case that other events with a strong impact, like sport events or elections, could follow a similar pattern for the measures used in this thesis. Therefore, in future work the Twitter data that we analysed could be compared to the Twitter data during other events with a similar size or scale, like a sport match or elections, to obtain more generic insights in the relation between real-world events and network structures.

References

[BB14] Axel Bruns and Jean Burgess. Crisis communication in natural disasters: The queensland floods and christchurch earthquakes. *Twitter and society - Digital formations*, 89:373–384, 2014.

[BHJ20] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. February 2020.

[BP16] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, 2016.

[CLEZG10] David Combe, Christine Largeron, Elod Egyed-Zsigmond, and Mathias Géry. A comparative study of social network analysis tools. 2:1–12, 2010.

[Coo21] W Timothy Coombs. *Ongoing crisis communication: Planning, managing, and responding*. Sage Publications, 2021.

[dt20] The Pandas development team. pandas-dev/pandas: Pandas, February 2020.

[Fel15] Contreras S. Karlin B. Basolo V. Matthew R. Sanders B. ... Luke A. Feldman, D. Communicating flood risk: Looking back and forward at traditional and social media outlets. *International Journal of Disaster Risk Reduction*, 15, 12 2015.

[Fin86] S. Fink. *Crisis management: Planning for the inevitable.* New York, N.Y: American Management Association, 1986.

[Fou20] Python Software Foundation. Python, -. *https://www.python.org/*, February 2020.

[HSS08] Aric Hagberg, Pieter Swart, and Daniel Schult. Exploring network structure, dynamics, and function using networkx, 2008.

[Hun07] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[Hyd20] Hydrator - computer software: Documenting the now., February 2020.

[Kim8a] Hastak M. Kim, J. Online human behaviors on social media during disaster responses. the journal of the nps center for homeland defense and security. *https://www.hsaj.org/articles/14135*, page 7–8, 2018a.

[Lit20] Justin Littman. Hurricanes Harvey and Irma tweet IDs, February 2020.

[Mit94] Ian I. Mitroff. Crisis management and environmentalism: A natural fit. *California Management Review*, 36(2):101–113, 1994.

[OG12] Márcia D. B. Oliveira and João Gama. An overview of social network analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2, 2012.

[SG16] Margaret C. Stewart and B. Gail Wilson. The dynamic role of social media during hurricane sandy: An introduction of the STREMII model to weather the storm of the crisis lifecycle. *Computers in Human Behavior*, 54:639–646, 2016.

[SLLdG15] Patric R. Spence, Kenneth A. Lachlan, Xialing Lin, and Maria del Greco. Variability in twitter content across the stages of a natural disaster: Implications for crisis communication. *Communication Quarterly*, 63(2):171–186, 2015.