A Network Approach to Detect Heavily Affected Cities and Regions Using Facebooks Movement Data: Final Report

Zhengtao Jin
Stanford University
Computer Science and Slavic Department
zjin2@stanford.edu

Guoqin Ma
Stanford University
Civil Engineering Department
sebsk@stanford.edu

Ende Shen
Stanford University
Computer Science Department
endeshen@stanford.edu

Abstract

Evacuation and returning behaviors and decision making during natural disasters are usually hard to monitor and predict. The paper transforms the data on Facebook's Geoinsight disaster maps for Hurricane Florence into networks and performs network analysis to study evacuation and returning behaviors to identify heavily affected cities and communities within the cities. Methods and metrics we used include degree, betweeness centrality, harmonic centrality, page rank, and clustering coefficient calculation, as well as Louvain algorithm for community detection, principle component analysis, and decoupling methods. The cities identified by our analysis to be severely affected by the disaster are compared with the NOAA flood map that was created post the hurricane, where we find out that the heavily affected cities are correctly detected, and that network analysis can indeed help us gain a nuanced interpretation of disaster data.

1. Introduction

Natural disaster brings great economic loss and threats people's life. The Geoinsight disaster maps developed by Facebook provides valuable data for us to generate meaningful analysis in order to better inform the disaster evacuation process. The motivation for this research stems from the authors' perception that data on these disaster maps are underutilized. A lot of the predictions that can be generated from the maps might help us better estimate the effects of a disaster, and better inform us in the decision making process post a disaster. Network analysis techniques are especially relevant here, as the people's movement across different cities may be modeled as a directed graph, and predictions can be made using network analysis metrics such as degree, betweeness centrality, harmonic centrality, page rank, and clustering coefficient, as well as network analysis algorithms such as louvain algorithm for community detection, principal component analysis, and decoupling methods.

In our paper, we try to harness the Geoinsight disaster maps to monitor people's movement during disasters to study their evacuation and returning behaviors and decision making during disasters, to identify suspicious anomalies which may be associated with unusual accidents, and to estimate regional disaster damage. In order to do so, we will construct networks from Facebook's user movement data and fulfill these tasks with analysis of network metrics mentioned above. One central aim that drives our exploration is to identify heavily affected cities during a natural disaster, and to figure out a way to extract more nuanced information as to what caused the above identified network anomalies, such as power outage. Our estimation will be compared to the NOAA flood map, which identifies heavily flooded areas during the hurricane.

1.1. Information for Event of interest: Hurricane Florence in the United States

Hurricane Florence is our focus in this project.

Hurricane Florence attacked Southeastern and Mid-Atlantic United States (mainly the Carolinas) in September, 2018. It is the wettest hurricanes recorded in the Carolinas and 8th in contiguous United States. It has a detrimental impact on local residents properties and lives, causing heavy rainfall and floods in vast area. The estimate damage in only North Carolina has reached $13 billion.

On September 7, North Carolina declared the state of emergency, followed by South Carolina and Virginia on September 8, Maryland on September 10, Washington D.C. on September 11 and Georgia on September 12. Mandatory evacuation orders were given to some coastal areas in North Carolina, South Carolina, and Virginia on September 10 and 11. On September 14, Hurricane Florence landed United States from Wrightsville Beach, North Carolina, with a strength of Category 1. It dissipated on September

19, with many places still flooded and evacuees unable to return.

2. Related Work

2.1. How Social Ties Influence Hurricane Evacuation Behavior [4]

The paper lays the groundwork for understanding population movement at times of natural disaster. One major contribution that was that it developed many methods and metrics to analyze population movement data post disaster - including ways to identify evacuation behavior, measuring rates of return using Cox proportional hazards models, and finding intuitive results using descriptive statistics.
It shaped our project's focus on community detection by elucidating the fact that social ties amongst population in different regions can in fact shed light on evacuation patterns of population after disaster. We used the process of conducting the analysis - looking at different metrics that might intuitively correlate the above two features, and looking at different metrics and proxies and determine what social, or geographical information might be implied by the population movement data and in turn predict the using these technical and sociological observations.
Although a premise for our project, lacking from the paper is the inclusion of broader features except ones that reveal social ties. The paper served as an inspiration for our methods, but we would look at more sources for the technical details.

2.2. Improving the Robustness of Complex Networks with Preserving Community Structure [10]

Robustness is a decent measure of regional resilience. Therefore, this metric is of great interest to researchers working on disaster management. Robustness could serve as one way for us to determine the significant nodes in the network.

The cited paper particularly focuses on the relationship between robustness and community structure. They make a vital statement that the increase in robustness should not sacrifice the community structure, which is related to the functionality and characteristics of the original network. It is shown that if new edges are added to the network, the robustness of a network certainly betters but the modularity would drastically drop. In order to maintain the modularity of the network, 2 new approaches are provided to ameliorate the robustness: first, let nodes with similar importance in one community connect with each other (which they call it an onion-like structure); second, let highly important nodes only in connection with nodes in the same community.

This paper helped us think of ideas to improve robustness of the people movement network in emergency planning in different scenarios. We considered both weighted and unweighted graphs in our project.

2.3. Ways of Using Facebook's Disaster Data[8][1]

These two websites showcase how Facebook's User Movement data can be used to help with disaster relief. For example, the website [8] shows using these data to re-schedule "hurricane preparedness modules" with detected abnormal patterns, e.g. "more people seem to be congregating around the outer edges of these places versus in the center".

The blog [1] acknowledges that Facebook's data is only representative of people who use Facebook on mobile with Location Services enabled. At the same time, it takes the example of Kaikoura Earthquake in New Zealand, during which Facebook's user movement data more or less matched the evacuation and returning of residents. This blog discusses two other types of map that Facebook constructed, i.e. Facebook Safety Check and Facebook Location Maps. Those two maps focus on specific locations with the interest of learning the extent to which disasters have affected certain areas. However, those maps overlook the fact that evacuations are directional, interconnected, and convoluted; rather, in order to fully understand the influences of disasters, a better model needs to be equipped, i.e. a series of complex networks under time propagation. Therefore, this justifies our need to construct, analyze, and interpret the network model generated with User Movement Data.

3. Data

Facebook launched disaster maps in 2017 to to provide insights in near-real time to help humanitarian organizations coordinate their work and fill crucial gaps in information during disasters [6] [7]. The data are derived from locations of Facebook users enabling location service. They are aggregated and anonymized in order to protect users privacy.

We choose the *Daily Admin Movement Vectors* from Hurricane Florence dataset as our data in this project. The bounding box is shown in figure 1. The covered area mainly includes North Carolina, South Carolina, West Virginia, Virginia, Georgia, and D.C., with some locations falling in New Jersey, Ohio, Kentucky, Tennessee, Delaware, etc. 2 exemplary entries in the dataset are listed in the table below.

Facebook Administrative Movement dataset is similar to edge-lists. For each entry of the dataset, there are 11

Table 1. Properties of normal and anomalous data

	Normal	Anomalous
Date Time	2018-09-10 00:00	2018-09-10 00:00
Ending Location & Ending Region Name	363 Ladson_1	2764 White Marsh_2
Starting Location & Starting Region Name	371 Summerville_1	2756 Parkville_2
Length(km)	9.7035	7.5601
Baseline: People Moving	183.4	0
Crisis: People Moving	185	164
Difference	1.6	164
Percent Change	0.8677	16400
Standard (Z) Score	0.1318	366.7151

features, namely, Date Time, Starting Location, Starting Region Name, Ending Location, Ending Region Name, Length, Baseline: People Moving, Crisis: People Moving, Difference, Percent Change, Standard Z Score. Most of the features are plain and clear as their names. Each region name correspond with one location (integer code). Baseline: People Moving are calculated on a 3-week average before the crisis. A probability distribution, which is used to calculate Standard Z Score, is also drawn from the 3-week data. The Florence dataset cover the dates from Sep 10, when evacuation orders were issued, to Oct 8, recorded every 8 hours. Figure 1) illustrates Facebook user's movement between Sept 9, 20:00 EDT and Sept 10, 4:00 EDT.

4. Methodology

First of all, we perform exploratory data analysis, trying to validate the graph by means of doing degree analysis and community detection to check whether the results conform to the fact.

Then, we use Facebook user movement network to draw the temporal evolution of network metrics such as degree, pagerank, closeness, betweenness, etc. in order to observe potentially different patterns in different evacuation groups (mandatory evacuation, voluntary evacuation, no evacuation) and to identify anomalies. Based on government-released or media news and government official Tweets [5] [2] [9] [3], we classify the cities in the bounding box into 5 categories, namely, mandatory evacuation, voluntary evacuation, mandatory evacuation cancelled, mandatory evacuation for visitors, and no evacuation.

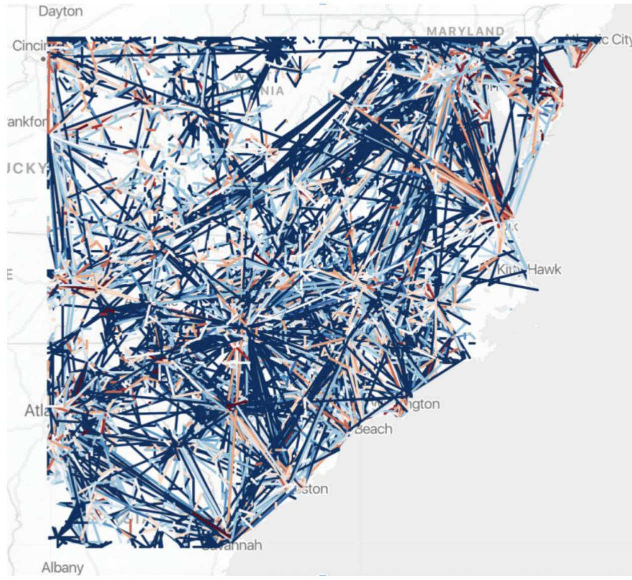

Figure 1. Visualization of Facebook Users Movement Data on September 10 (Screenshot of Facebook Geosight Portal, credit to Facebook). The colors represent standard Z score, with blue being positive and red being negative.

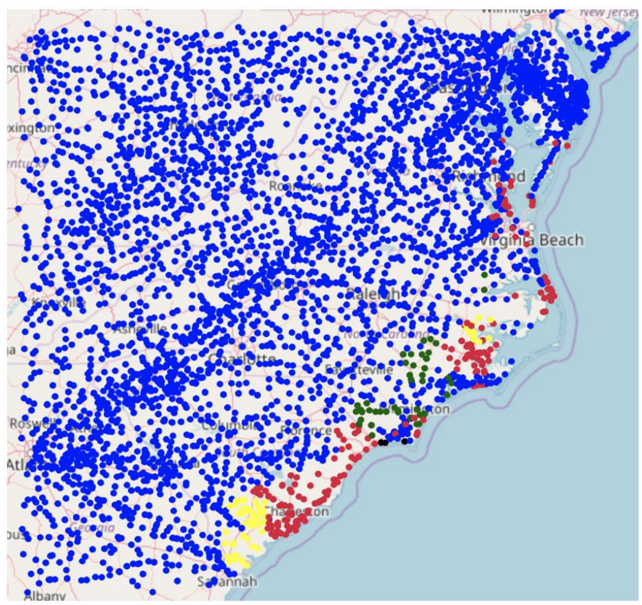

Figure 2. Evacuation map drawn from government-released news and Tweets. Blue: no evacuation; Red: mandatory evacuation; Green: voluntary evacuation; Yellow: mandatory evacuation cancelled; Black: mandatory evacuation (visitors).

Finally, when utilizing Facebook data to monitor people's movement, one critical challenge is: how could we take the effect of power outage and signal lose into consideration? To decouple the effect of people movement and signal loss (The users were not in the network anymore due to power outage or signal tower failure), we derive a balance equation to quantify the effect of power outage and signal loss.

Define $STAY(t)$ as the number of people staying in the node city by the end of time window t.

2 fundamental equations:

$$STAY(t) = LOOP(t) + IN(t)$$

$$STAY(t) = LOOP(t+1) + OUT(t+1)$$

1 key equation - User balance equation:

$$IN(t) + LOOP(t) = OUT(t+1) + LOOP(t+1)$$

1 derived equation:

$$STAY(t+1) - STAY(t) = IN(t+1) - OUT(t+1)$$

Define marginal inconsistency (MI) as:

$$MI = OUT(t+1) + LOOP(t+1) - IN(t) - LOOP(t)$$

MI equals the number of people who **we fail to detect by the end of time window t but resume tracking during time window $t+1$** minus the number of people who **we detect by the end of time window t but lose track of during time window $t+1$** .

By taking the cumulative sum of MI, we get the cumulative inconsistency(CI).

$$CI = OUT(T) - IN(0) + \sum_{t=1}^{T-1} NETOUT(t) + LOOP(T) - LOOP(0)$$

Here we treat the inconsistency before the landing of Hurricane Florence (September 14, 2018) as people turning off their location service during evacuation, and treat the inconsistency after the landing of Hurricane Florence as power outage and/or signal tower failure.

5. Exploratory Data Analysis

In the dataset, one day is divided into 3 time-windows, namely, (EDT) 20:00-4:00, 4:00-12:00, 12:00-20:00. We take the 3 data files on Sept 10, the first day in the dataset to do EDA. The corresponding local time periods are Sept 09 20:00 - Sept 10 4:00, Sept 10 4:00 - Sept 10 12:00, Sept 10 12:00 - Sept 10 20:00, respectively.

By doing EDA, we try to primarily validate the integrity of using Facebook user movement network to study the human behavior in this region.

5.1. Graph Statistics

As a first step, we calculate node and edge number to have a sense of the size of the networks. We find that during between 20:00 and 4:00, the number of Facebook users travelling between cities are evidently less than other 2 time slots. The majority of Facebook users stay in one location during the 8-hour period, which indicates the network have considerable self-edges. During the daytime of Sept 10, people's movement activities are intensified accompanied by mandatory evacuation orders issued in some states.

5.2. Baseline Degree Distribution

We make the following hypotheses:

1. During local time between 20:00 and 4:00, we expect that people do not move much and that they may stay home. Therefore, it is worth exploring if there is a surge in travellers during midnight time, which might be due to disaster Hurricane Florence.
2. During local time between 4:00 and 12:00, people move from home to workplace. The importance of a node city from the graph indicates more on the economic competitiveness of the node city.
3. During local time between 12:00 and 20:00, people move most frequently during daytime. The reason of movement is the most complicated as well.

We construct 3 weighted directed loop-free networks from the 'Baseline: People Moving' feature from each timestamp in the dataset. Figure 3 shows that in the nighttime, many node cities have 0 out-degree. By contrast, Figure 4 shows that in the morning, many node cities have 0 in-degree. Last but not least, Figure 5 shows a more balanced distribution in terms of in-degree and out-degree. These distributions could verify our hypotheses: In the nighttime, people living in small cities or towns tend to return home from their workplace and not go out; in the morning, conversely, people living in small cities or towns tend to go to their workplace in metropolis and there are

Table 2. Basic statistics of movement networks in 3 phases. The number outside the parenthesis is baseline movement and the number inside the parenthesis is the crisis movement. Note that ”crisis movement” does not necessarily indicate movement during crisis. It could also indicate pre-and-post disaster movement. When constructing the graphs, we allow the existence of isolated nodes but drop out the edges with 0 weight (no people moving along these edges), so the numbers of nodes are always the same between ’baseline’ and ’crisis’ while the numbers of edges differ.

time window	Sept 09, 20:00 - Sept 10, 04:00	Sept 10, 04:00 - 12:00	Sept 10, 12:00 - 20:00
number of nodes	2899 (2899)	3036 (3036)	2981 (2981)
number of weighted edges	16479 (20313)	20982 (25115)	20146 (24393)
total FB users	2482878 (2482304)	2505704 (2495538)	2559976 (2604253)
total FB users travelling between cities	429429 (419583)	613458 (699316)	588515 (655375)

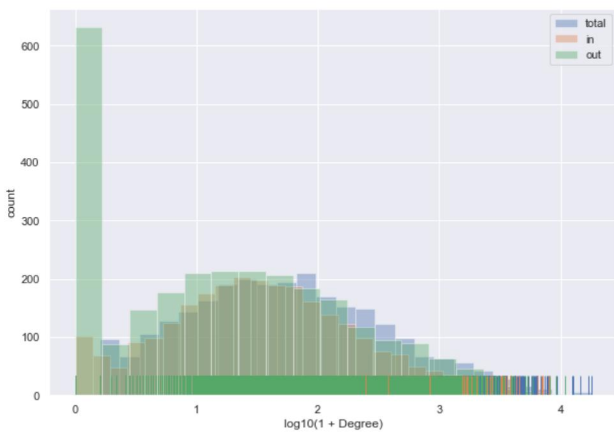

Figure 3. Degree distribution between 20:00 and 4:00

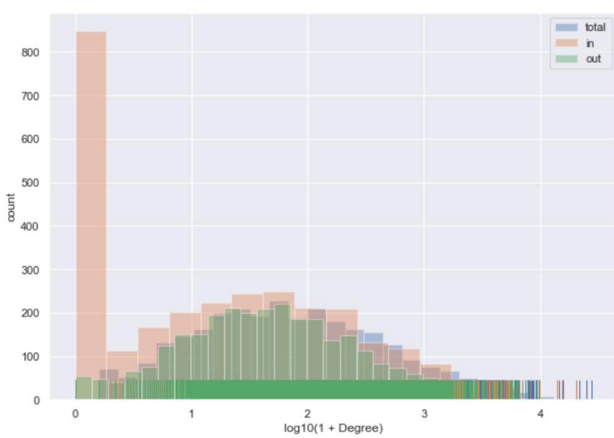

Figure 4. Degree distribution between 4:00 and 12:00

few people heading to small cities or towns; between 11am and 7pm, people may move for different purposes and the in-degree distribution and out-degree distribution are similar.

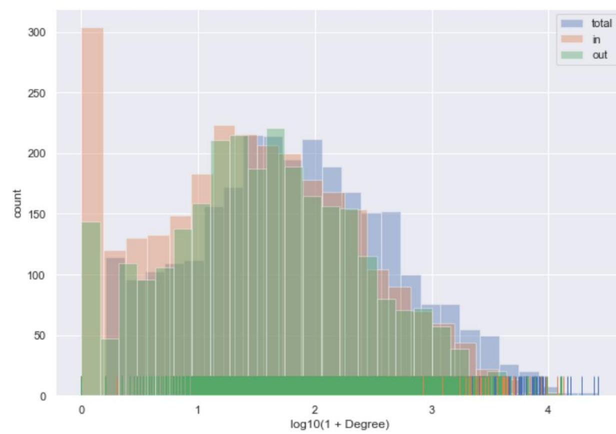

Figure 5. Degree distribution between 12:00 and 20:00

5.3. Community Detection

We use a weighted Louvain algorithm to detect communities in the bounding box. The graph we choose is a weighted directed graph excluding self-edges. The result is shown in Figure 6.

From the graph that is constructed with Facebook user movement data, we could successfully detect the principal metropolitan statistical areas in the targeted region. This gives us more confidence in using Facebook users movement as a representative miniature of the movement of the whole population.

5.4. Robustness

We consider both weighted graphs and unweighted graphs when calculating α of the network.

Since the count vs degree plot for unweighted graph does not have a heavy tail and the degree CCDF of unweighted graph has few points at the beginning, the α for unweighted graph is calculated with the least square method on count vs degree plot. On the contrary, the count vs degree plot for weighted graph have a noisy end, so the

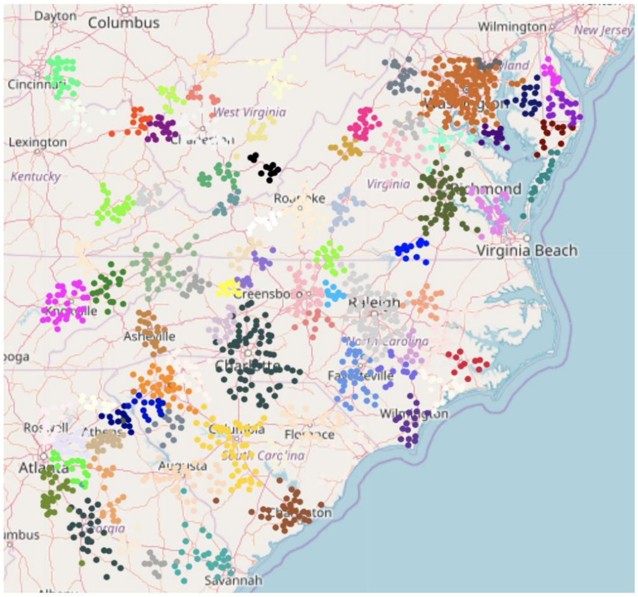

Figure 6. Detected communities with size larger than 10. Major metropolitan statistical areas, such as Charlotte, Raleigh, Charleston, Columbia, Fayetteville, Greensboro, Washington, Atlanta, etc., could be successfully detected.

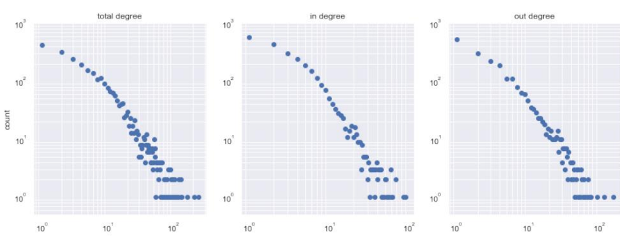

Figure 7. Count-degree plot for unweighted graph Sept 9 20:00-Sept 10 4:00

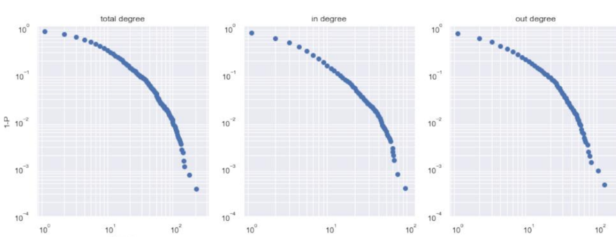

Figure 8. Degree CCDF for unweighted graph Sept 9 20:00-Sept 10 4:00

α for weighted graph is calculated with the least square method on the linear part of the degree CCDF. The results are summarized in Table 3

α values are all below 2. For unweighted graphs, α of in-degree is the largest in nighttime while α of out-degree is the largest in daytime. For weighted graphs, α's for in-degree, out-degree, total-degree do not vary significantly.

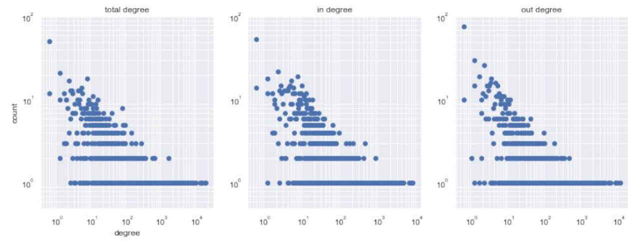

Figure 9. Count-degree plot for weighted graph Sept 9 20:00-Sept 10 4:00

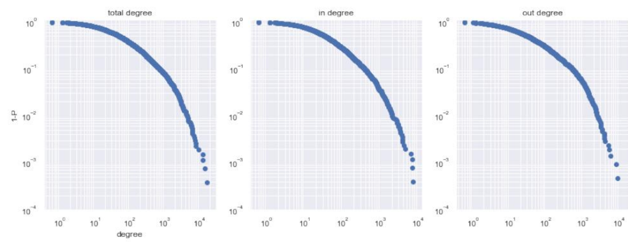

Figure 10. Degree CCDF for weighted graph Sept 9 20:00-Sept 10 4:00

6. Main Results

6.1. metrics

The metrics that we calculate for each time stamp for each node are degree(in, out, total, with and without self-edges), betweenness centrality, harmonic centrality, pagerank (with and without self-edges), and cluster coefficient.

We found that weighted degree, which could be interpreted directly as number of people coming into, moving out, or staying in a nodal city, is the most sensitive network metric among different evacuation groups, as shown in Figure 11 12, 13, 14

The y-axis is standard z score. A clear drop is observed in the groups receiving an evacuation order between Sept 11 and Sept 16, during which mandatory evacuation is ordered and Hurricane Florence landed and its destruction reached maximum. After hitting the bottom, the z score starts to return to 0, indicating people are coming back and power is being restored.

In particular, Figure 11 was generated with the restriction that the standard z-score is confined to $|z| < 5$. We did this in order to eliminate data anomalies. However, if

Table 3. α values. The number outside the parenthesis is baseline movement and the number inside the parenthesis is the crisis movement.

time window	Sept 09, 20:00 - Sept 10, 04:00	Sept 10, 04:00 - 12:00	Sept 10, 12:00 - 20:00
α (in-degree)	1.7434 (1.7216)	1.5026 (1.5186)	1.6559 (1.5684)
α (out-degree)	1.5639 (1.5474)	1.7559 (1.7215)	1.7031 (1.6441)
α (total-degree)	1.4325 (1.4097)	1.4121 (1.3834)	1.4305 (1.3971)
α (weighted in-degree)	1.1075 (1.1628)	1.1079 (1.2071)	1.1083 (1.2382)
α (weighted out-degree)	1.1312 (1.1910)	1.0879 (1.1332)	1.0987 (1.2048)
α (weighted total-degree)	1.0934 (1.1209)	1.0774 (1.1314)	1.0810 (1.1759)

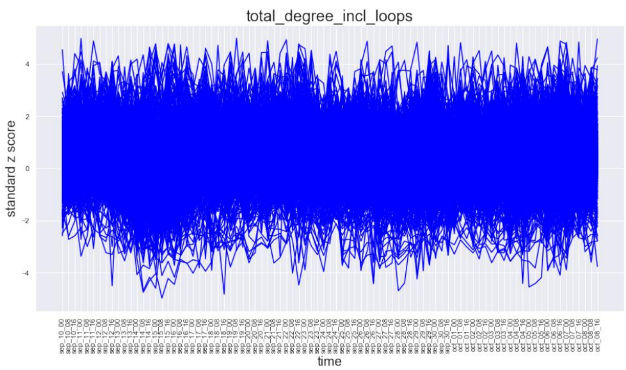

Figure 11. Time evolution of total degree incl. self-edges of nodes in no evacuation group

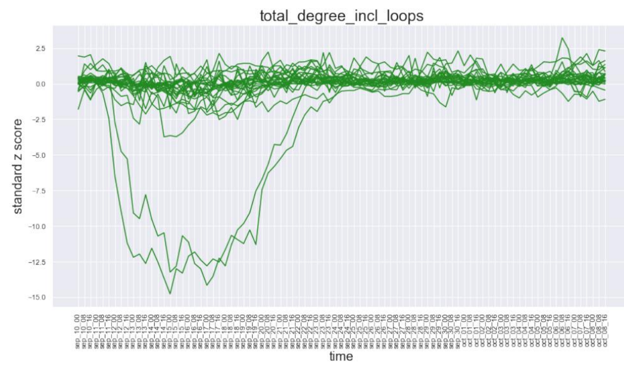

Figure 13. Time evolution of total degree incl. self-edges of nodes in voluntary evacuation group

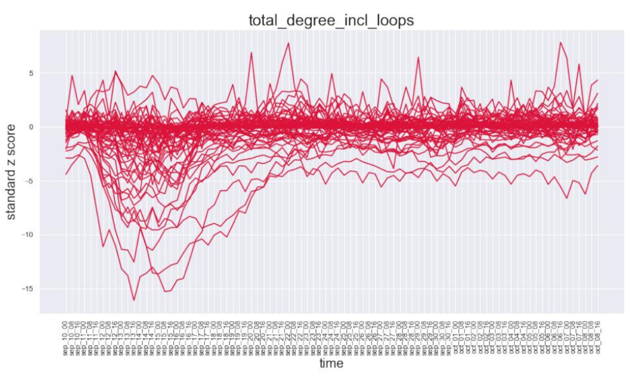

Figure 12. Time evolution of total degree incl. self-edges of nodes in mandatory evacuation group

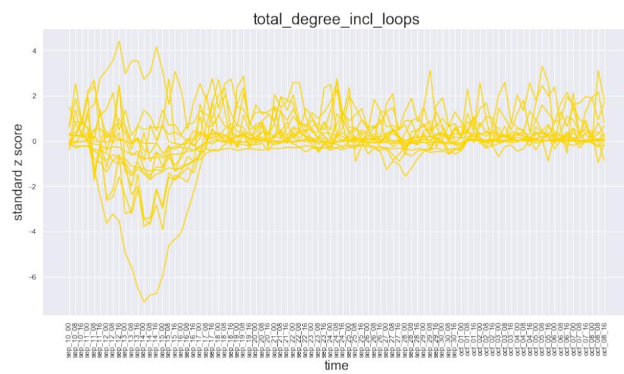

Figure 14. Time evolution of total degree incl. self-edges of nodes in cancelled mandatory evacuation group

we look at the original figure (Figure 15), we do have some findings: those curves with z-score below -10 look like representing evacuation zones. After extracting the specific data (as shown in Figure 16), we found that Jacksonville, Onslow, NC is actually under voluntary evacuation but it is missing in news reports that we referred to. Shallotte, Brunswick, NC is not an unincorporated community but it is really close to the coastline (thus more affected).

6.2. Principal Component Analysis

To better visualize the cities and their difference, we use a dimension reduction technique, PCA to decompose the standard z score time evolution curves of all the cities. The projection is shown in Figure 17 18 19 20 and 21. PC1 explains 42% of variance and PC2 explains 7% of variance.

The main cluster centered at 0 represent the unaffected or lightly-affected cities, the cities falling out of the main cluster are heavily affected. To validate our results, we check the flood map of NOAA (Figure 22) and find Wilmington, Jacksonville, New Bern, Myrtle Beach are all successfully detected in our analysis.

Besides, we find that many cities receiving evacuation

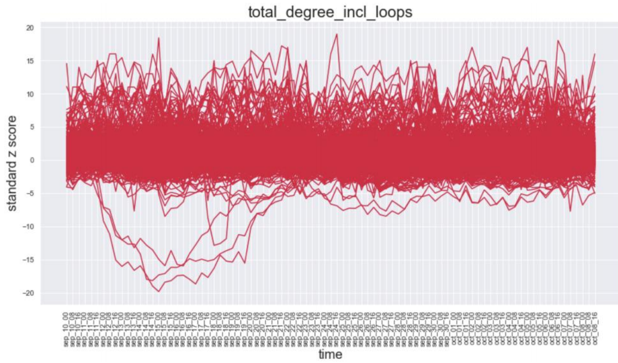

Figure 15. Time evolution of total degree incl. self-edges of nodes in cancelled mandatory evacuation group, without z-score restriction

	sep_10_00	sep_10_08	sep_10_16	sep_11_00	sep_11_08	sep_11_16	sep_12_00	sep_12_08	sep_12_16	sep_
Hubert	0.431641	-1.652304	-1.053576	-1.160326	0.211179	-2.626978	-5.995983	-4.690617	-9.050541	-9.8
Jacksonville_1	-0.420281	-1.538877	-0.232771	-0.832660	-0.654490	-5.503891	-11.879238	-12.175590	-18.405275	-18.9
Jacksonville_2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Pineville_3	4.195734	-0.571981	-0.398215	-0.390974	0.434416	-0.506819	NaN	NaN	NaN	
Sandston_2	0.530283	-1.101672	2.193410	-0.652234	-0.243903	-0.069882	-0.380241	-0.125034	-0.154152	-0.0
Shallotte	-3.478258	-4.082827	-2.656920	-2.722547	-1.974791	-4.224025	-5.038900	-5.379467	-8.514175	-9.3

Figure 16. Cities with z-score below -10

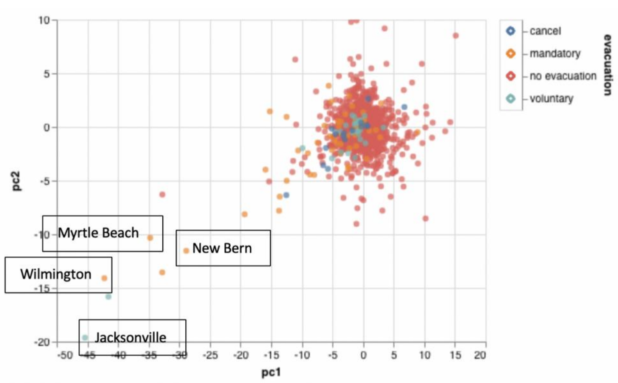

Figure 17. PCA of total degree incl. self-edges

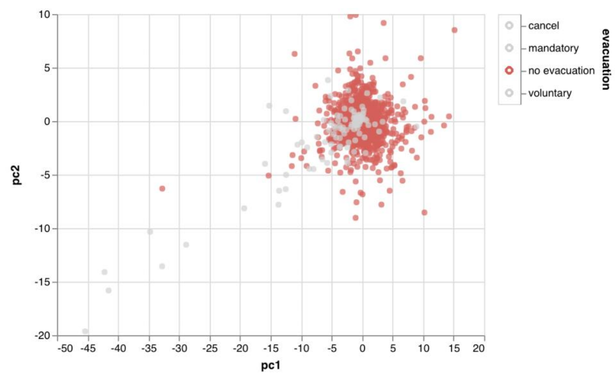

Figure 18. PCA of total degree incl. self-edges (cities without evacuation order)

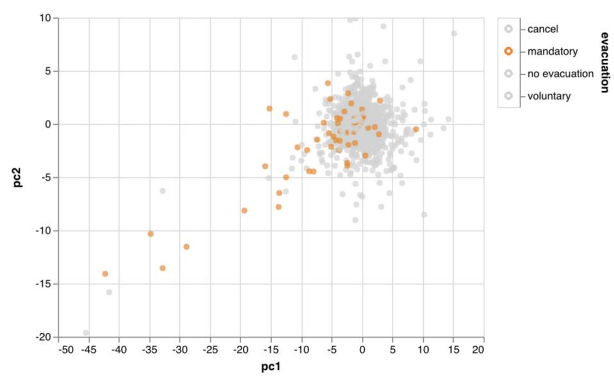

Figure 19. PCA of total degree incl. self-edges (cities with mandatory evacuation order)

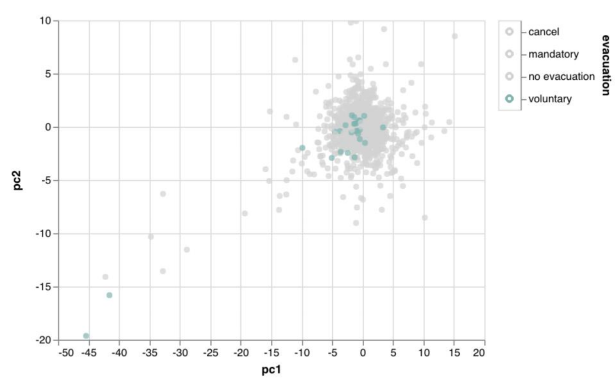

Figure 20. PCA of total degree incl. self-edges (cities with voluntary evacuation order)

orders staying inside the main cluster, which indicates an evacuation order does not necessarily mean heavy local damage. Mandatory evacuation group has a larger proportion of cities falling out of the main cluster than the voluntary evacuation group.

6.3. Power Outage

The anomaly detected in the previous section is a comprehensive one, which is a combination of people leaving, power outage, signal loss, etc. In order to check the effectiveness of the evacuation order (such as residents ignoring the order), we have to decouple the people leaving from the rest factors contributing to the anomaly.

The marginal inconsistency of the temporal data for the mandatory evacuation group is drawn in Figure 23 and 24 in 2 different presentations. The cumulative inconsistency is shown in 25. We attribute the inconsistency before September 14 to evacuation, during which people may turn off their location service to save energy; and attribute the inconsis-

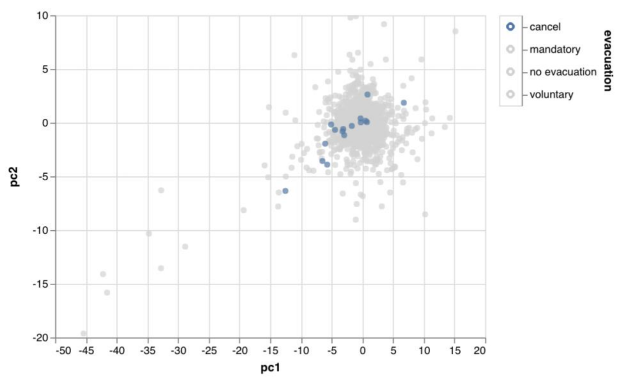

Figure 21. PCA of total degree incl. self-edges (cities with cancelled mandatory evacuation order)

Figure 22. NOAA flood map

tency after to power outage and / or signal tower failure. We use the inconsistency to correct people movement data (Figure 26 and 27). The net out (out degree - in degree) makes more sense for the mandatory evacuation group.

7. Conclusions

Facebook disaster maps have been gaining more attentions from disaster-response groups since it launched. Our study shows that Facebook user movement data could be representative enough to summarize people movement characteristics at the city level and give successful community detection. It is of great help in disaster decision making by detecting heavily affected cities during Hurricane Florence. Last but not least, we define an approach to decouple different factors explaining the anomaly so as to allow for a more detailed analysis and usage of these data in disaster decision making in the future.

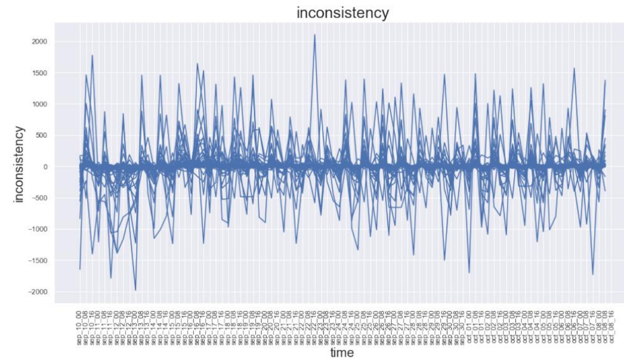

Figure 23. Marginal inconsistency over time

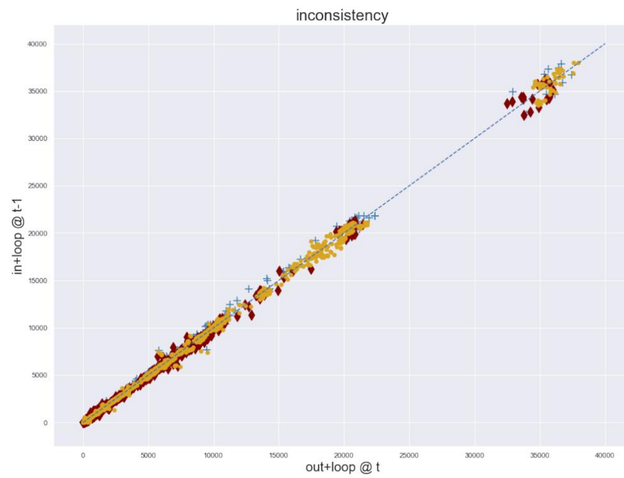

Figure 24. in degree + self edges at time window t-1 vs. out degree + self edges at time window t. (+: before September 14, the landfall of hurricane; diamond: September 14 - September 19; dot: after September 19, the dissipation of hurricane)

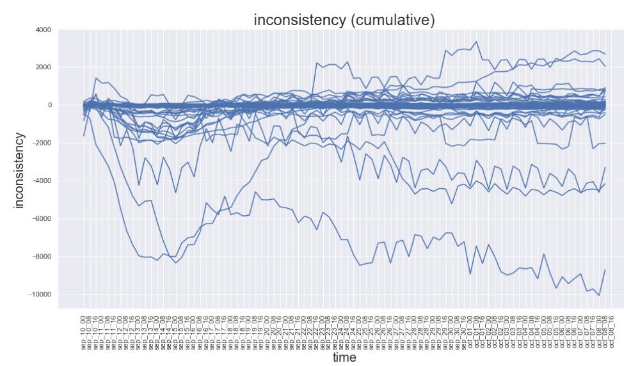

Figure 25. Cumulative inconsistency

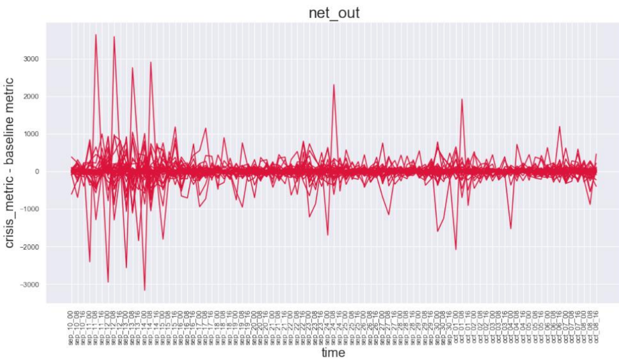

Figure 26. People net-out before correction

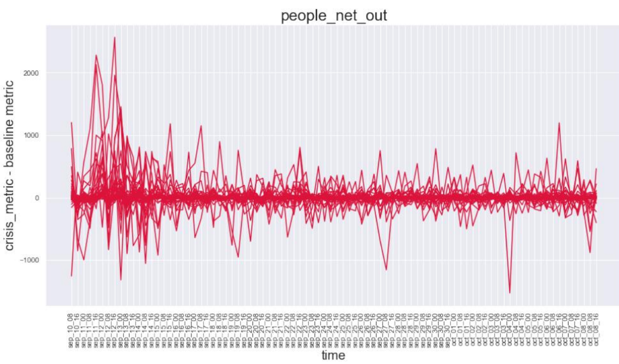

Figure 27. People net-out after correction

8. Personal Contribution

Zhengtao Jin: Literature review; parts of report write-up; problem definition; poster making;

Guoqin Ma: Coding and plotting in data cleansing, EDA, metrics calculation, time evolution, PCA, decoupling of different effects.

Ende Shen: Literature review; write-up of results of time evolution experiments; visualizing changes in total degree by producing .gif images.

References

[1] The future of crisis mapping is finally here, Jun 2017.

[2] abc11. Hurricane Florence evacuation zone: Mandatory orders issued ahead of storm — abc11.com.

[3] C. D. Bill Chappell. Hurricane Florence: Carolinas And Virginia Issue Evacuation Orders : NPR.

[4] D. P. A. Danaë Metaxa-Kakavouli, Paige Maas. How Social Ties Influence Hurricane Evacuation Behavior Facebook Research, 2018.

[5] N. C. Department of Transportation. Ahead of Florence, Evacuations Begin in Coastal Areas.

[6] P. Maas, C. Nayak, A. Dow, A. Gros, W. Mason, I. O. Filiz, C. Diuk, G. Burrows, M. C. Jackman, V. Sharma, C. Lang, W. Malik, and D. Patel. Facebook Disaster Maps: Methodology Facebook Research, 2017.

[7] M. Paige, L. McGorman, C. Nayak, W. Park, and A. Gros. New data tools for relief organizations: network coverage, power, and displacement Facebook Research, 2018.

[8] B. Paynter. How facebook's disaster maps is helping aid organizations serve people affected by florence, Sep 2018.

[9] The Weather Channel. All Hurricane Florence Evacuation Orders State by State — The Weather Channel.

[10] Y. Yang, Z. Li, Y. Chen, X. Zhang, and S. Wang. Improving the Robustness of Complex Networks with Preserving Community Structure. *PLOS ONE*, 10(2):e0116551, 2 2015.