The Newspaper-Microblogging Ecosystem:

Measuring Signal Strength and Connectivity of the American Newspaper

Geography // Information Systems // University of Nevada, Reno
May 2019

ABSTRACT:

The purveyance of information through the traditional paperbound newspaper has been hybridized by online news and the spread of news microblogging platforms. Coexistence between a paperbound newspaper and its affiliated microblogging profile permits a broader geographic range of informational dilation to the public. However, this new ecosystem of traditional and digital media can adversely affect the quality of media and journalism. The shift in medium from traditional to hybrid also advances a new informational user-interface for the news consumer. This research project examines the localities of traditional paperbound newspapers in the United States, identifying areas of high and low newspaper density by state. From this benchmark, signal strength for each newspaper is tested with a Twitter API in terms of tweets on the microblogging platform. The nine states chosen for the signal strength test were Nevada, California, Arizona, New York, Illinois, Alabama, Arkansas, Florida, and Colorado. A hub and spoke network analysis for each state is also conducted, which isolated the tweet activity of key users for each state. A final analysis of user characteristics was established based on identifiable traits of users inside and outside of the tweet network. Inferences suggest a power-law dynamic for the most active hubs and newspapers in the state networks. Results also suggest broad commonalities in user profile types, including many bot and cyborg accounts.

KEYWORDS: microblogging, Twitter bot, newspaper, network analysis, citizen journalism.

INTRODUCTION:

The traditional paperbound newspaper has served its community well, both in America and the world at large, providing a professional and thorough exposition of news events on a small and large scale, and reinforcing the values of place and community within its hometown and across its circulation radius. As the media and business structure of the newspaper evolved, it eventually reached a saturation point in the market and could no longer steadily expand. In fact, well before the advent of the internet, the e-newspaper, and the entire gamut of online news mediums, the viewership of traditional newspapers was declining. According to *The Vanishing Newspaper*, when the newspaper is "measured by household penetration (average daily circulation as a percent of households), this mature industry peaked in the 1920's at 130%.....by 2001, newspaper household penetration was down to 54 percent," (Meyer 4-5). This changing dilation level of paperbound newspapers into the public eye was substantial but didn't threaten the normative social fabric that newspapers had patched together across our democratic framework.

As the internet moved in, the fate of the newspaper quickly moved nearer to its judgement day. The judgement day for the traditional newspapers was summoned upon the same way that technologic shifts and new business paradigms set forth: seamless and without a blink. Just as the buggy had no choice to be usurped by the car, or the telegraph by the telephone, or the brick and mortar store by ecommerce, or the tape and DVD movie rental store by online streaming, the paperbound newspaper and all its piled anchors of traditional journalism practice were upended (Rushkoff 27). The upending of the newspaper didn't erase it completely from the earth. In fact, similar to most historic technologic shifts, new technologies

are layered on top of old ones in order of relative power and importance; the old technologies don't disappear but are reassigned a subservient role in the technologic hierarchy (Brock 89). The tape of time cannot be rewound and the attacks on the integrity of traditional news journalism cannot be vindicated, but a new cerebral programming of what journalism is and how it can be represented in the digital world can help move it forward. As stated in *Out of Print*,

> "the internet is not simply a new publishing system, allowing faster, wider distribution of material assembled and edited as it has always been. The changes wrought by digital technology are transformative and not adaptive: they require journalism to be rethought," (Brock 1).

With the transformative shift in the newspaper and its journalistic methods, new semantics need to be applied to the profession. First and foremost, what is "news", and secondly, what is "journalism"? Traditional newspapers and news mediums follow a top-down structure where journalists represent a professional position at the top of a long-established media triangle from producer to consumer (Tolochko and Boomgaarden 1783).

With the spread of online news mediums and microblogging platforms came the rise to fame of the citizen journalist. The basic structure of the citizen journalist is quite humbling; citizen journalism cuts nearly all the middle men and provides a direct feed into the jugular of event and opinion. For the citizen journalist to have the freedom to post "news" on social media, YouTube, microblogging platforms, personal news websites, or any other online transmission vector feels like a heroic stab at the heart of big media, a cybernetic tossing of tea into the harbors as a symbolic show of freedom in the face of a media tyranny. And these basic defining tenets of citizen journalism are true. Citizen journalism is real and it's grassroots, but is it really "news"? And more bluntly, "if anyone can be a journalist thanks to cheap, simple electronic publishing technology, what is a journalist and can we define what they do?" (Brock 2).

Each day these grassroot forces embed deeper into the informational wiring of America and world media. The counterintuitive and puzzling feature of the freeing up and decentralizing of the media environment is that is hasn't helped the democratic process, nor has it upended the domination of big media holdings. One of the issues with the new digital media environment is that its connectivity doesn't move in the same patterns. Information flows from the outside in and has allowed many more people to openly participate, but without under-empowered people seeking active empowerment through its means, it becomes swallowed up and owned by similarly structured big media holdings of the newspaper era. In simple words, "each new media revolution appears to offer people a new opportunity to wrest that control from an elite few and reestablish the social bonds that media has compromised," and this opportunity must be seized (Rushkoff 59). In hindsight, the previous media age may have been controlled by the few, but the printed newspaper served as a "traditional source of the information citizens rely on for their everyday orientation and the exercise of democratic responsibilities," (Thurman and Fletcher 543). The toppling of this centralized force has awakened new possibilities, and also new forms of unrest.

As the definition and role of the newspaper, the news, and the journalist gets rewritten, other systemic changes are simultaneously happening. The decline of the traditional newspaper, the rise of the e-newspaper, and the domination of news and "pseudo-news" through the new medium channels of social networks, microblogging, news websites, and vlogging have not only forced news and journalism to be redefined but have also changed our relationship with information. The interactivity of information and news has opened up more dialogue to challenge and change. Consequently, the impermanence and changeability of this new

interactive digital medium has taken away the authority that the fixed nature of print once laid forth (Brock 86-87). Figuratively and metaphorically, the paper trail of our news is gone. Bouncing around in its predecessor's shoes, a new and revived vehicle of information transmission has awoken from its pre-cybernetic lull, and we call it the meme vehicle.

The meme vehicle represents all news and information transmitters, including all human beings. The meme, to ensure its survival and propagation, catches a free ride on any meme vehicle; it can be viewed as a replicated sequence, or an information replicator that is a complex message capable of replicating itself (Dawkins 15). In the internet age, the meme is the basic title of a message. While the memetic force of pervasive language and newspaper titles has kept the meme healthy and running before interconnected web platforms existed, the wild world of the web has given the meme access to an entire world network for its proliferation. Interestingly, the message content of the meme doesn't determine its fate; provoking or disgusted, loved or hated, liked or disliked, news or fake-news, the meme moves on regardless of appeal. A meme lives inside of a "meme-pool", which we could loosely bound within the constraints of human language interaction; memes live inside of brains and then travel outwards (Gleick 313).

Much of internet-based news on social networks and microblogging platform depend on the meme for information diffusion. The "shock-and-awe" factor of a single phrase, perhaps accompanied by an image or perhaps not, has become the paradigm for the new journalism standard of the digital age. Shocking titles such as:

"*A woman live-streams' her husband dying",*

or perhaps,

"*Congressman transmits smartphone pictures of his genitals*,"

have allowed the meme to replicate while also demanding society's attention to the "news" (Rushkoff 34). As the meme looks to survive regardless of a like or a dislike, of having taste or being tasteless, the refreshed feeds of our online news sources have succeeded in bringing us closer to this type of news. Political blogs, commonly associated with the microblogging platform of Twitter, are a great nesting ground for citizen journalism, for the proliferation of the meme, and for the transmitting of factual and non-factual news. It turns out that the political blog is

> "mostly user-centric and viewed in opposition to the professional, corporate media. Although there are many manifestations of citizen journalism, political blogs resemble the conventional newspaper columns much more than any other form of citizen journalism," (Tolochko and Boomgaarden 1786).

Understanding the advantageous and detrimental manifestations of the microblogging platform is crucial. This digital cypher, along with other social media platforms that commanded the news in new ways, needs to be understood as an official news platform of the digital age.

THEORETICAL PROBLEM:

The Decline of the Newspaper, The Rise of Online News and Microblogging Platforms, and the Growth of Automated News Feeds-

This research investigation attempts to better understand the Twitter microblogging platform. With the Twitter API, microblogging signal strength can be interpreted as a function of how many traditional newspaper titles are being tweeted. Tweets may be in reference to the paper

or digital version and will help determine which hybridized newspapers are best represented on the Twitter microblogging network. A fortuitous and equivalating gesture by the social media giant Twitter, has been the open accessibility to Twitter data and user data via the API. Novice developers can explore trends in user data and more advanced developers can create automated systems to perform tasks within the platform. The Twitter social media and microblogging network is very robust due to the nature of the platform; users may follow anyone they wish, and this permits the construction of a very complex network. In fact, "the Web, blogosphere, online social networks and human contact networks all belong to a class of "scale-free networks,"" (Java et al. 58).

Understanding the semantic overlay of the term "microblogging" helps to realize the broad uses that a service like Twitter might offer to the public. The term microblogging

> "is a contemporary phenomenon that refers to the broadcasting of brief messages to some or all members of the sender's social network through a specific web-based service," (Hennig-Thurau & Wiertz 375).

Its reach can be from the most mundane recitals of daily activities to the mass-publication of sentiment by a world leader. This new form of communication with one another can be referred to as "microblogging word of mouth", or MWOM, and has proven very influencial at predicting the outcomes of foreseen events (Hennig-Thurau & Wiertz 376). Understanding the underlying sentiment of its users and the ultimate purpose of one's presence on the Twitter microblogging platform has become a very important task. The rise of automated Twitter bots that grow from the API platform's open rule book has allowed for the era of the "automated news feed". News bots can gather links and data from sources of their choice and transmit them to all of their followers. Other more malicious bots designed with "the specific intention to mislead, exploit, and manipulate social media discourse with spam, malware, misinformation, slander, or even just noise," are also granted free access into the realm of Twitter (Ferrara et al. 96).

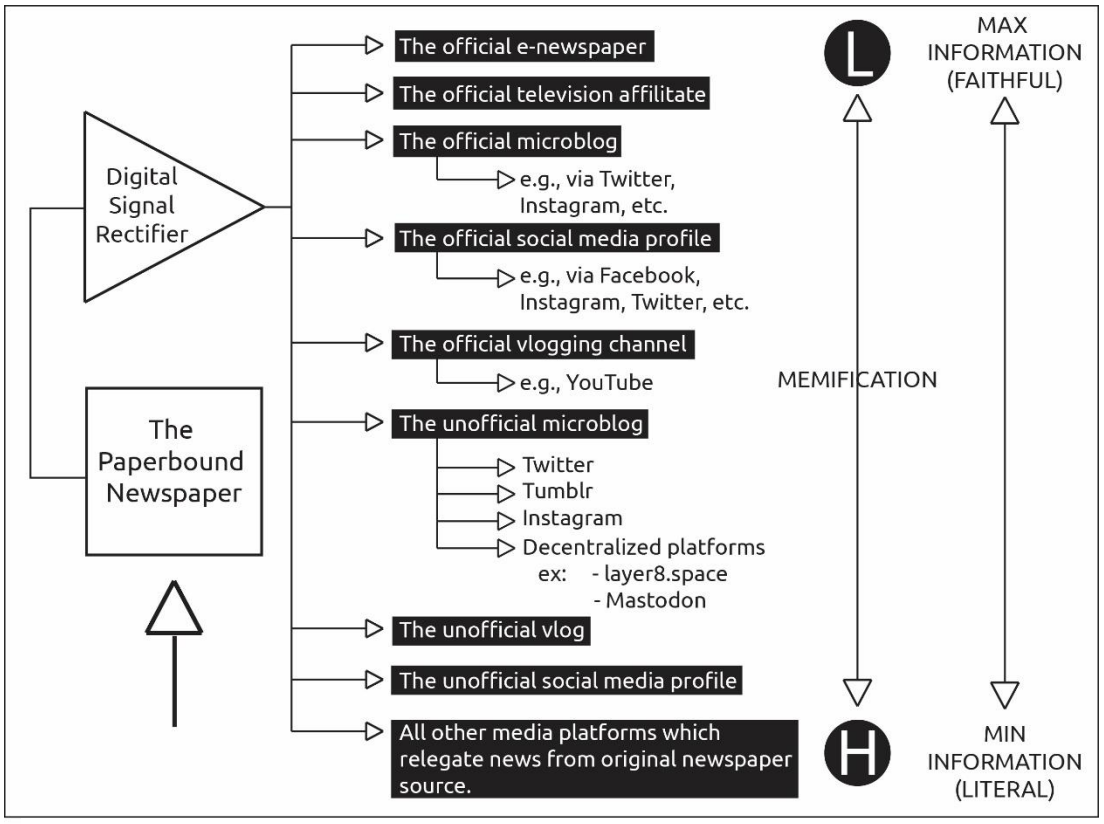


Fig.1: A schematic detailing the transformative process that news information goes through as its medium changes from physical to digital form.

The stakes are very high. The current state of Twitter microblogging combines the very best and very worst of the new media paradigm; citizen journalism has given everyone a voice, and that is positive. Nonetheless, the transmission of unauthentic and deceptive media across these channels has further heightened the state of disarray in a fragile American and world political system.

By understanding the signal strength of microblogging platforms in terms of newspapers will provide a baseline value to be assessed through time. By understanding the network dynamics of how different users tweet more or less frequently within a given state, will help to understand user behavior and diversity of news sources. Lastly, by gaining a better grasp on who's doing the tweeting, we can better assess the validity of the entire newspaper-microblogging ecosystem.

METHODS AND DATA:

A.) Acquisition and organization of newspaper data –

A baseline of total newspapers in the United States was provided by the Wikipedia page "*List of Newspapers in the United States".* This list proved robust, but not complete, and totaled 4,852 newspapers within continental and territorial lands. For this baseline list to be suitable for geocoding with the Google API, city data needed to have correct syntax within Microsoft Excel that appeared as "City, ST". So, Python .split() and Excel columns to rows tool were used.

B.) Geocoding newspaper list with Python script and Google API –

For the geocoding of each newspaper based on its "City, State" format, a Google API account and key were required. This allowed the geocoding process to be facilitated through a code in Python. Of the 5,054 U.S. newspapers on the input list, 4,852 were successfully geocoded. This list output of 4,852 city newspapers with accurate coordinate data served as the foundation for the geospatial analysis.

C.) Use Twitter API in R Studio to search for tweeted newspaper titles–

With the twitteR and plyr packages installed in R Studio, each state's newspapers were searched in tweets across the entire Twitter "tweetscape". Free API accounts enable tweets to be gathered for up to 8 days prior to the search query; the queries for this study were conducted between March 13th – April 23rd of 2019. Each newspaper was searched for as a single string (ex: "\"Alabama Baptist\"") to ensure that the tweet results returned were connected to the newspaper. A visual check of the tweet message for possible title strings that may have picked up noise was conducted. For example, the Alabaman newspaper, *La Lagniappe*, drew considerable noise as the title picked up tweets related to its word usage in the French language. In cases such as these, the search string was given a more specific title and then checked for tweet accuracy. The string "*La Lagniappe"* was lengthened to "*Mobile La Lagniappe",* and in other cases the article "*The"* was added at the beginning of the string, or "*newspaper"* added to the end of the string. The searchTwitter function from twitteR searched for tweets corresponding to each newspaper string. A maximum value of 15,000 tweets to return was selected to avoid hitting the max tweet wall assigned by Twitter.

For each newspaper with tweets, a dataframe or "tweetframe" was created that contained every tweet plus its user information. An additional column to each tweetframe was added to parse the newspaper name to each row of tweet information. A master tweetframe was made

to contain only the pertinent information for the study, such as "*username*", "*contents of tweet*", "*retweet count*", "*is retweet*", "*retweeted*", and "*newspaper name*". To finalize the tweet search, all newspaper tweetframes were concatenated by row (rbind()) to create a masterfile of tweet data for each state. The masterfile was exported as a csv, as was a separate list giving a unique count for each username to provide a total tweet count by each user in the set (data.frame(table(df$column).

D.) Left outer join statements in Python to make master tweet file –

A succession of three left outer joins were performed in Python Spyder to build a complete master tweet file for each state. An outer join was chosen so that all the data was preserved, allowing it to be examined for false or unsuccessful joins. For example, on the first left outer join which combined the masterfile of tweet data with the list of geocoded newspapers, the common column for joining was "newspaper". An example is below:

```
import pandas as pd
import numpy as np
output_filename = '….State_tweet_xy_merge.csv'
State_alltweets = pd.read_csv('……..Rmasterfile.csv', encoding = 'unicode escape')
geocoded_news = pd.read_csv('……State_news_geocoded.csv', encoding = 'unicode)

        mergedState = pd.merge(State_alltweets, geocoded_news, how='outer',
                                        left_on= 'newspaper', right_on= 'newspaper')
```

Analyzing the finished outer join could show where syntax was not equivalent, thus producing no join in tables. For example, many newspaper titles on the right side of the join had spaces before the first word and were therefore inequivalent to the left table. Fixing the syntax errors to create two identical keys for the join was made easier by using the *Notepad ++* software which could detect all symbol types in any string. The final and corrected left outer join #1 was then exported as a csv.

Left outer join #2 and #3 each combined the output result from the previous join with unique tweet count data for each user(#2) and city population data for each newspaper (#3). The final left outer join (#3) contained all tweet data, all geocode xy data, all unique counts, and the city population for each newspaper. Once exported as a csv file, the dataset was reordered in descending fashion from most to least unique tweet count, and then save as the final masterfile for the state.

E.) Making centroids in Python and network constraints –

Reading in the final masterfile saved in descending order, each qualified user can be indexed in a top-down manner. The centroid function consists of defining a function in Python that will average the x and y values of the indexed user's variety of newspapers in their tweets. In this study, the term centroid is modified to refer to "a user's center of geographic interest," as it creates a mean-center value for the sum of newspapers that hail from different cities. Rather than using the centroid value to represent a potential homebase for the user, defining the centroid as a user's center of geographic interest provides a central xy value to serve as the center node of each user's individual hub and spoke network. Each qualifying centroid is printed, copied, and pasted into a csv file centroid list, defining the username with its x and y centroid. The centroid coordinates are also copied and pasted to the final masterfile. From here, each qualifying user is given a unique csv table with the centroid's xy values as the final two columns.

Network Constraint #1:

To qualify for the newspaper-microblogging network, a user must have tweeted three or more unique newspaper titles within the state.

Network Constraint #2:

Users within the network that have identical paths (ratio of tweets to unique papers) will each be counted for record, but only one of these users will be quantified and displayed for the final state network summary. This constraint ensures that the total edge count for each state won't be miscounted by duplicate user networks.

Ex: CannabisBizNews and CannaFunds each tweeted 1x to the New York Newspapers of: "New York Daily News", "The Buffalo News", "Watertown Daily Times", and "Long Island Business News". This identical hub and spoke path was only included once in the network analytics and map for the New York.

Network Constraint #3:

If a user tweets 3 or more unique newspapers all within the same city with the same xy coordinates, this user will qualify for the network and for a centroid value. However, because this type of user has a "zero" edge count, they cannot be included in the network summary or map. Even though these users qualify and have produced a substantial number of tweets for a single node, these tweet values will not be used.

F.) Illustrate centroid list and build users' edges in ArcMap –

The Excel centroid csv file for each state is read into ArcMap and converted to a points shapefile. Each qualified users' centroid csv file is read in and converted to points. Using the xy to line feature, the users' unique centroid value is given lines (edges) that connect its variety of city newspapers. Upon completion, all cartographic design elements are included and the final state network map is exported as a JPEG.

G.) Quantify edge count for each city (degree of node) using R Studio and Python –

The Excel csv file created for each qualified user is read into R Studio (ex: read.csv()) and given a count for unique newspaper titles (ex: data.frame(table(df$column)). The number of rows of the csv file read in will verify the user's total number of tweets and the unique count will verify the number of unique edges for each user. To understand the total unique edge count for each city and state, each user's count results are printed, copied, and pasted to a csv file. The completed csv file is read back into the same R Studio session and given another unique count which determines the total unique edges for each newspaper. This result is also exported from R Studio as a csv file.

Opening the last csv export in Python and reading it in as a Pandas dataframe, the file is given a left inner join with the list of geocoded newspapers for the desired state. The purpose of the inner join is so that the join output will only show data for the distinct newspapers in the list. The completed inner join of all newspaper edges is exported as a csv file.

H.) Final edge count in Excel –

The exported file of all newspaper edges for a given state is opened in Excel and sorted by city name. A quick summation for total edges in each city is performed and then tallied in a new *master state network* csv file.

ex:

newspaper	City	edges
Der Blatt	Bronx, NY	1
Der Yid	Bronx, NY	1
Hamodia	Bronx, NY	6
Jewish Press	Bronx, NY	6

Bronx, NY total edges (total degree of node) = 14

I.) Network summary statistics in Excel: Calculating Beta, Diversity and Power indices –

A master state network file, or *network summary*, was derived in Excel. The file includes columns containing each city's "Total tweets in network", "Total Edges", "Total Vertices", "Global tweet count for users inside and outside of network", "Max tweet value in network", "Beta Index", "Power Score", "Diversity Score", and "Power and Diversity Score"

Purpose of Beta Index:

The beta index was calculated as a simple measure of the degree of connectivity for each city. The measure is defined as the proportion of edges/vertices.

$$\beta = \frac{\text{total edges for city}}{\text{total vertices in network}}$$

Purpose of Power Score:

Due to the potential of certain key nodes (city nodes or user centroid nodes) to provide disproportionately higher amounts of tweets than others, a Power Score was devised to illustrate the relationship between a city hub and its total tweet count. High Power Scores for a city hub show that these cities received a majority of the tweet activity in the entire network distribution.

$$\text{Power Score} = \left(\frac{\text{city's total tweets}}{\text{highest total tweet value in network}} \right) \text{X } 100$$

Purpose of Diversity Score:

The Diversity Score was devised as a countermeasure to the Power Score, giving high scores to city nodes or user centroid nodes that had a high ratio of distinct newspapers to tweets. This score rewards a city that received a small number of tweets by only a few unique users, or a city that was only tweeted one time by one user.

$$\text{Diversity Score} = \left(\frac{\text{city's total edges}}{\text{city's total tweets}} \right) \text{X } 100$$

J.) Map outputs in Python with matplotlib.pyplot and geopandas –

Map outputs of the contiguous United States were produced within a Jupyter notebook session in Python. The required packages for the map making process were matplotlib.pyplot, geopandas, and pandas.

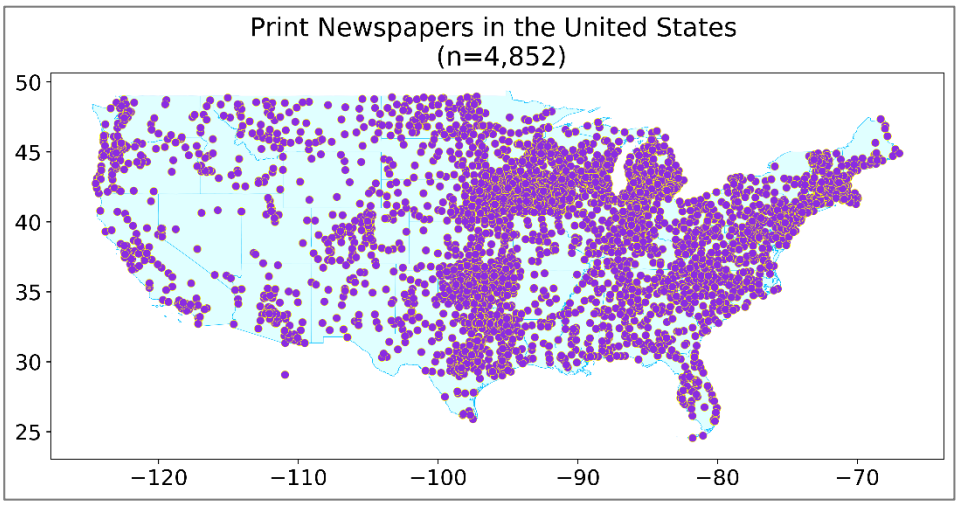
K.) Scatterplots of User Behavior in Python and R –

Scatterplots of users' profile information were plotted in R (ggplot package) with the assistance of tailored datasets from Python. The Python pandas function for joining tables was used again to bridge username information (exported from R) with tweet frequency information. The final scatterplot output in R measures the relationship of friends to followers for different user profiles.

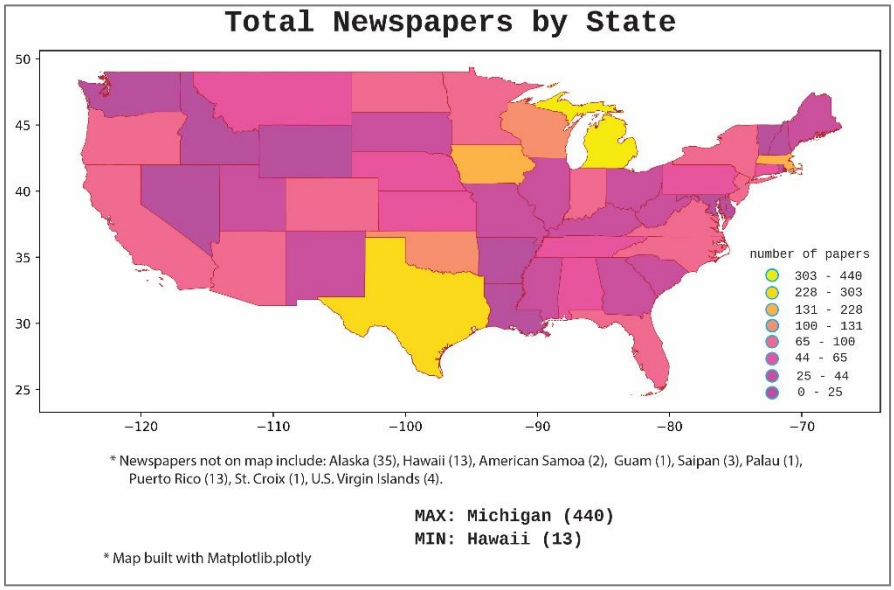
RESULTS:

A. Mapping traditional newspaper locations in the United States-

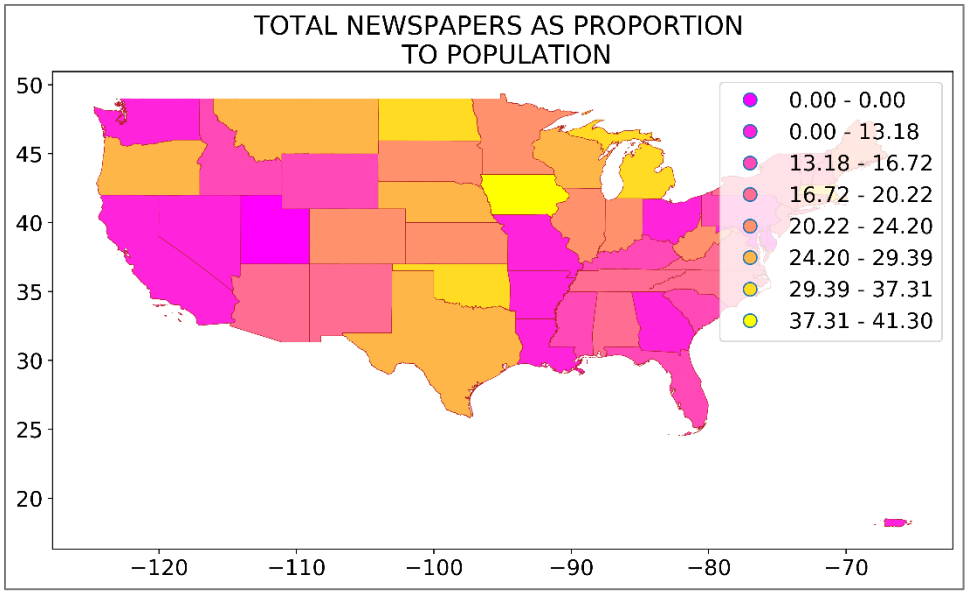
Of the entire list of U.S. newspapers, a total of 4,852 newspapers were geocoded. A higher than predicted number of newspapers were found in the Midwest, which is deserving of further research.



Newspaper frequency was found to be highest in the state of Michigan (440 newspapers) and lowest in the state of Hawaii (13).



Newspaper frequency as a proportion to state population showed that Iowa had the greatest ratio.

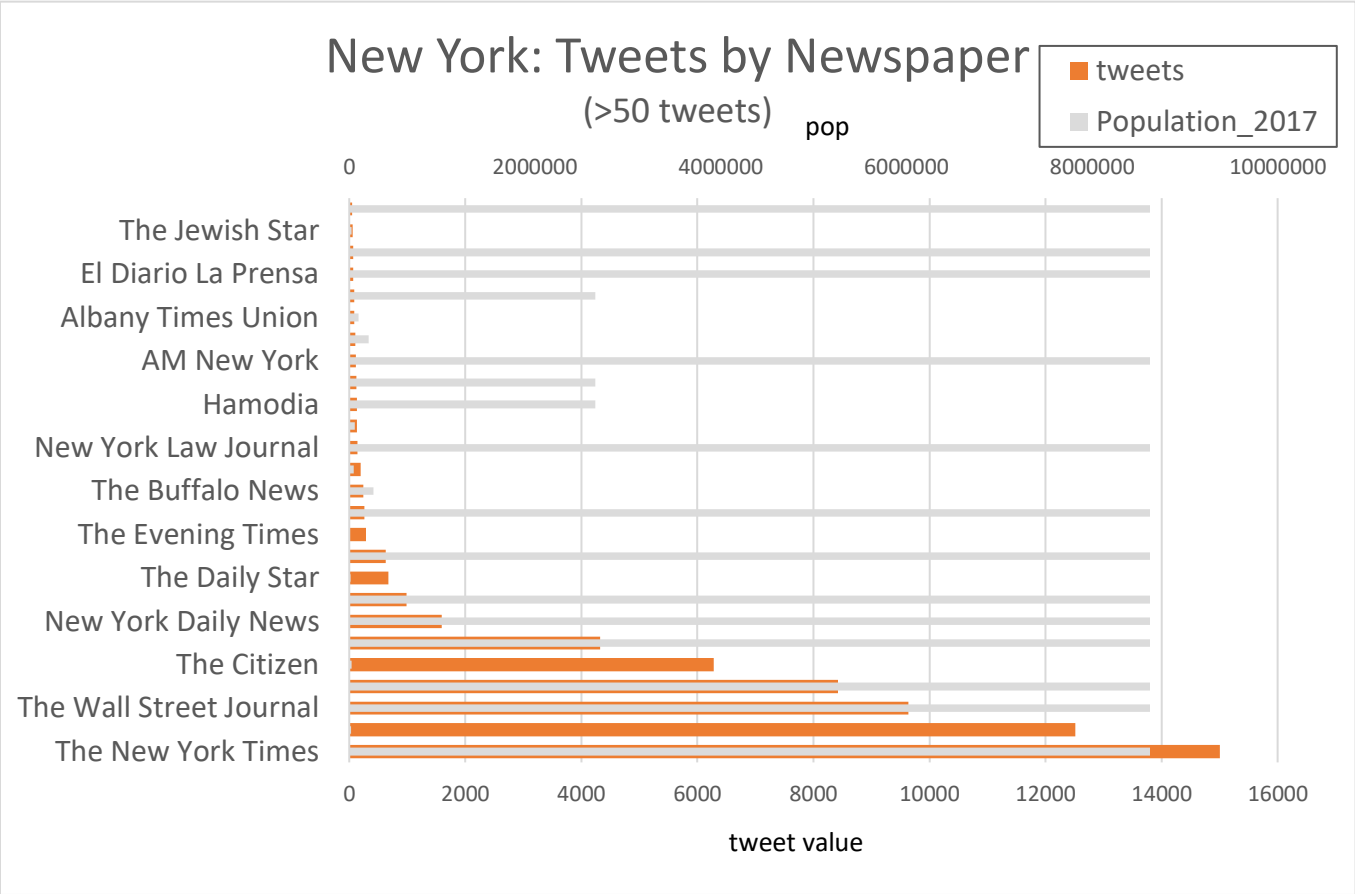


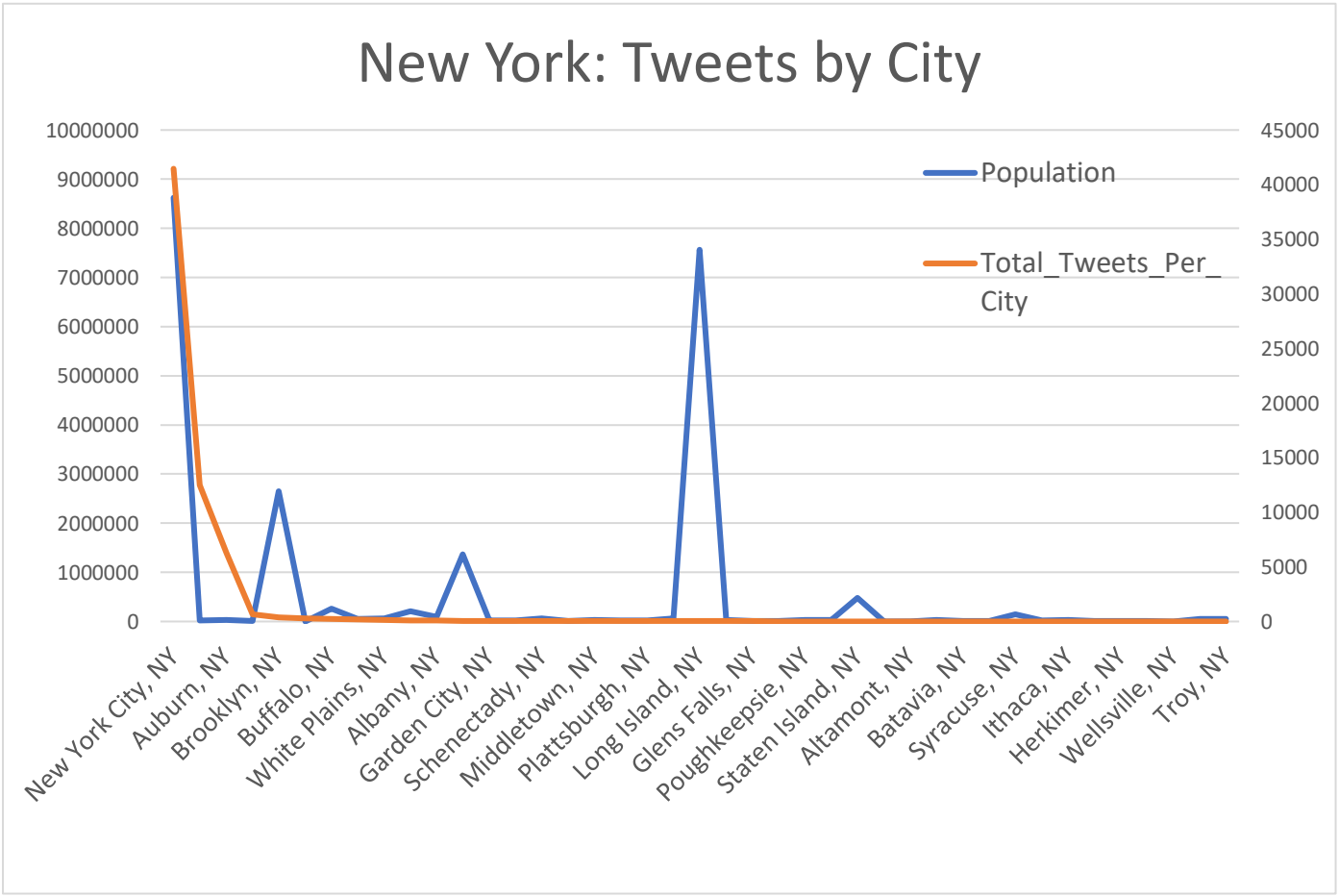
B. Signal Strength of Twitter Microblogging Platform in Terms of Tweeted Newspapers –

The 9 states tested for newspaper-microblogging signal strength were Nevada, California, Arizona, Colorado, Florida, New York, Illinois, Alabama, and Arkansas. Original plans were to test all 50 states, but due to time constraints these 9 strategic states were chosen. These nine states encompass our three largest cities (L.A., Chicago, and NYC) while the remaining 6 states were chosen to form state clusters in the south and southwest United States.

Twitter data was gathered between March 13th – April 23rd of 2019, and the aggregate amount of tweets was 125,235 tweets that pertained to our search criteria. Of these tweets, the highest single-state tweet score was New York with 62,995 tweets, followed by Florida (21,397), California (14,671), Illinois (5,725), Alabama (3,058), Colorado (2,553), Arizona (1,222), Nevada (623), and Arkansas (400).

Inside of the newspaper-user networks for each state, California had the best overall coverage with 89 of 109 newspapers tweeted and also had the highest number of city nodes (53). However, New York reigned king in terms of users who tweeted three or more unique newspapers within the state, and had the highest sheer number of tweets within the network. New York also had the highest city node in the study; New York City had 266 unique edges. This score is over double the next highest degree node of Los Angeles with 105 unique edges.
A scaling relationship between city nodes is clearly evident in many states and needs to be studied further. An interesting number of unique nodes with high diversity scores (# of edges / # of tweets) suggests that the Twitter microblogging platform is reaching smaller newspaper and news venues. Temporal studies need to be done to determine how the news centers flux with time.


New York: Tweets by Newspaper
(>50 tweets)
pop
tweets
Population_2017
0
2000000
4000000
6000000
8000000
10000000
The Jewish Star
El Diario La Prensa
Albany Times Union
AM New York
Hamodia
New York Law Journal
The Buffalo News
The Evening Times
The Daily Star
New York Daily News
The Citizen
The Wall Street Journal
The New York Times
0
2000
4000
6000
8000
10000
12000
14000
16000
tweet value


New York: Tweets by City
10000000
9000000
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
Population
Total_Tweets_Per_City
New York City, NY
Auburn, NY
Brooklyn, NY
Buffalo, NY
White Plains, NY
Albany, NY
Garden City, NY
Schenectady, NY
Middletown, NY
Plattsburgh, NY
Long Island, NY
Glens Falls, NY
Poughkeepsie, NY
Staten Island, NY
Altamont, NY
Batavia, NY
Syracuse, NY
Ithaca, NY
Herkimer, NY
Wellsville, NY
Troy, NY

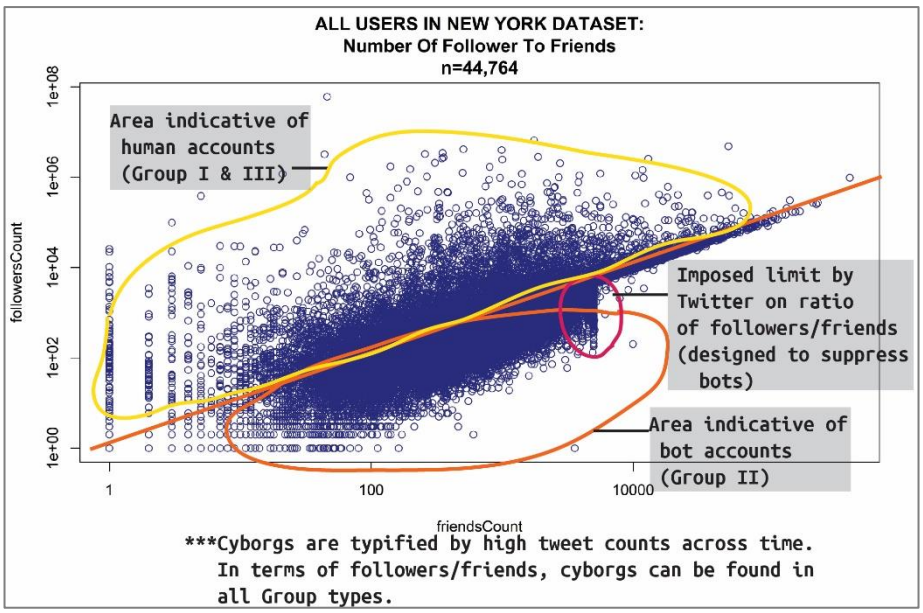
B. Understanding User Profiles in Study

Repeated usernames in multiple states suggested the possibility that many users were automated bots servicing a variety of purposes. Some typical measures to detect a bot include: the unique time stamp of the tweet, the content of the tweet, the sentiment of the tweet, the content of the tweet, whether the tweet is a retweet or not, and the network features of the user's links (Ferrara et al.). Another key metric for detecting a bot is to understand the relationship between followers and friends of a user. This metric reciprocates the network rules of the social media proxy in which Twitter operates; bots will tend to follow a large amount of people and have a disproportionately low amount of followers (Chu et al. 3). The followers to friends scatterplot can be broken down into three groups:

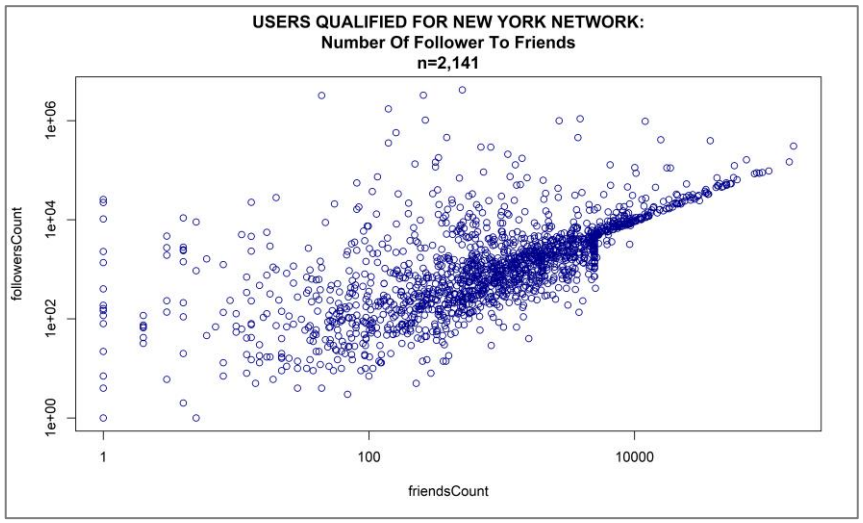
Group I: number of followers is clearly greater than friends
Group II: number of friends is clearly greater than followers
Group III: points stay close to the diagonal trend (followers ≈ friends)

Bot accounts are typified but not limited to a large number of group II users, where they lack equivalence in followers. Human accounts are typified but not limited to group III users, where amount of friends roughly equals amount of users. This represents the reciprocity seen in the follower-friend relationships seen in social media. Human accounts also contain many Group I users (celebrities, high-profile users...). Cyborg accounts are typified by high tweet counts, which isn't measured by the three groups. That being said, cyborg accounts can fit into Group I,II, & III (Chu et al. 3).

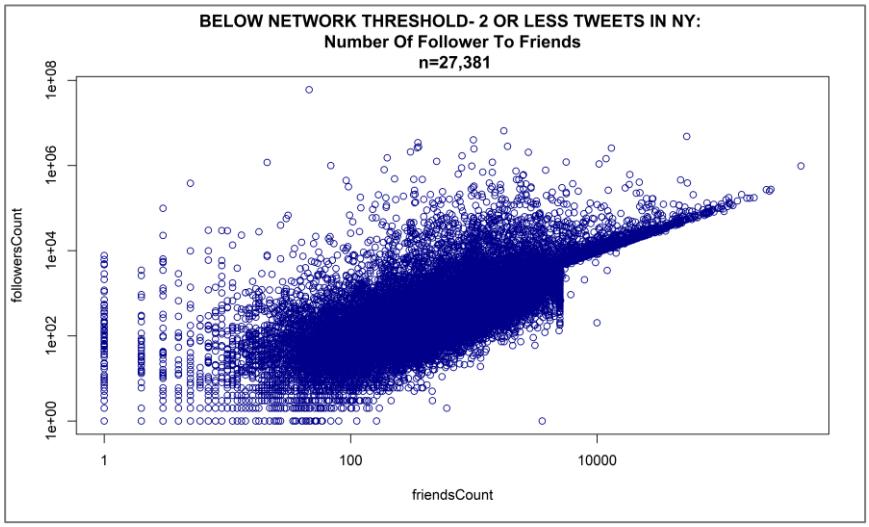
SET 1:
ALL USERS IN NEW YORK DATASET BROKEN DOWN BY GROUP TYPES:



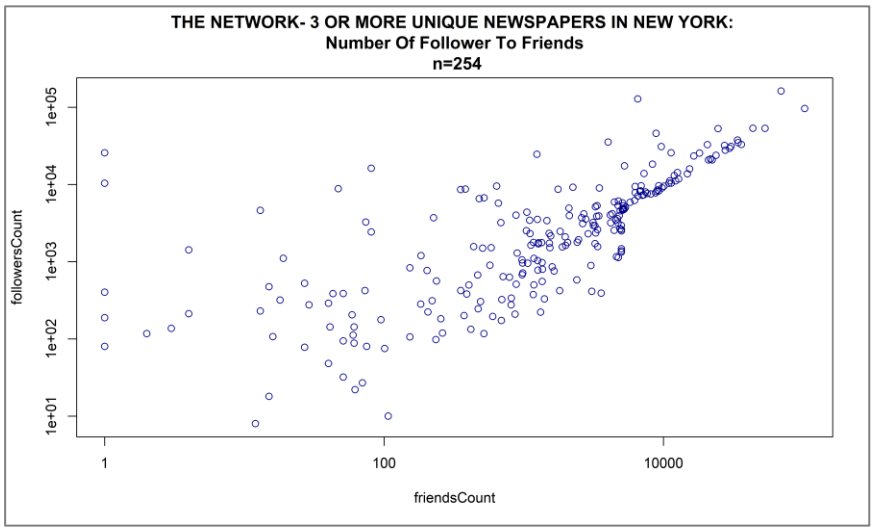
SET 2:
ALL USERS WHO MET INITIAL NETWORK CRITERIA OF THREE OR MORE TWEETS:



SET 3:
ALL USERS WHO WERE BELOW NETWORK MINUMUM



SET 4:
ALL USERS IN FINAL NETWORK



The results from the plots of user profile behavior illustrate that the users who tweeted newspapers in New York are suggestive of all three user types. Improving the confidence in labeling the type of each user account will require more tests on user behavior. A crucial test to be done next is to measure the time intervals between tweets; this is a key method to discern between a random (human) tweeting pattern or a non-random (bot/cyborg) pattern (Chu et al. 3). For the meantime, a pragmatic way to find out more about the type of user is to search their profile page. For example, two prolific tweeters that registered in each of the nine surveyed states were "*EcoInternetDrGB*" and "*LidarMonkey*". Both Twitter profiles smell of a bot or cyborg automated account due to their unhuman-like large amounts of tweets posted over short periods of time. If there is really automation in these users' behaviors, then they must be identified as either bot or cyborg. While bots are 100% automated, a cyborg profile will interchange between human and automated input (Chu et al. 2). Both EcoInternetDrGB and LidarMonkey have profiles which are almost completely comprised of links and nearly devoid of personal sentiment blogging, another suggestion that these profiles are bots or cyborgs.

The user *EcoInternetDrGB* appears to be a peaceful and constructive news bot, spreading information in defense of the climate and environment. *EcoInternetDrGB* has a very high Diversity score, as the user tweets a high variety of newspapers within each state. While *EcoInternetDrGB* doesn't have a followers to friends ratio typical of a bot (23.4 K followers to 16.5 friends), other user behavior traits infer automation in the account.

The user *LidarMonkey* appears to be another peaceful and constructive news bot and spreads a unique message across the Twitter platform. Innovation and geospatial technology appear to be the true interest of *LidarMonkey*. Interestingly, *LidarMonkey* tweets a broad variety of newspapers across each of the 9 test states. The followers to friends ratio for *LidarMonkey* fits the archetype for an automated account (12K followers to 12.7K friends); however, more user behavior metrics will need to be tested to be certain.

UNCERTAINTY:

Understanding the true nature of Twitter accounts will require much deeper research. A more complete map of newspapers in the U.S. will provide a better understanding of the newspaper-microblogging relationship. Understanding other media platforms in conjunction with the newspaper and microblog will provide a better analysis.

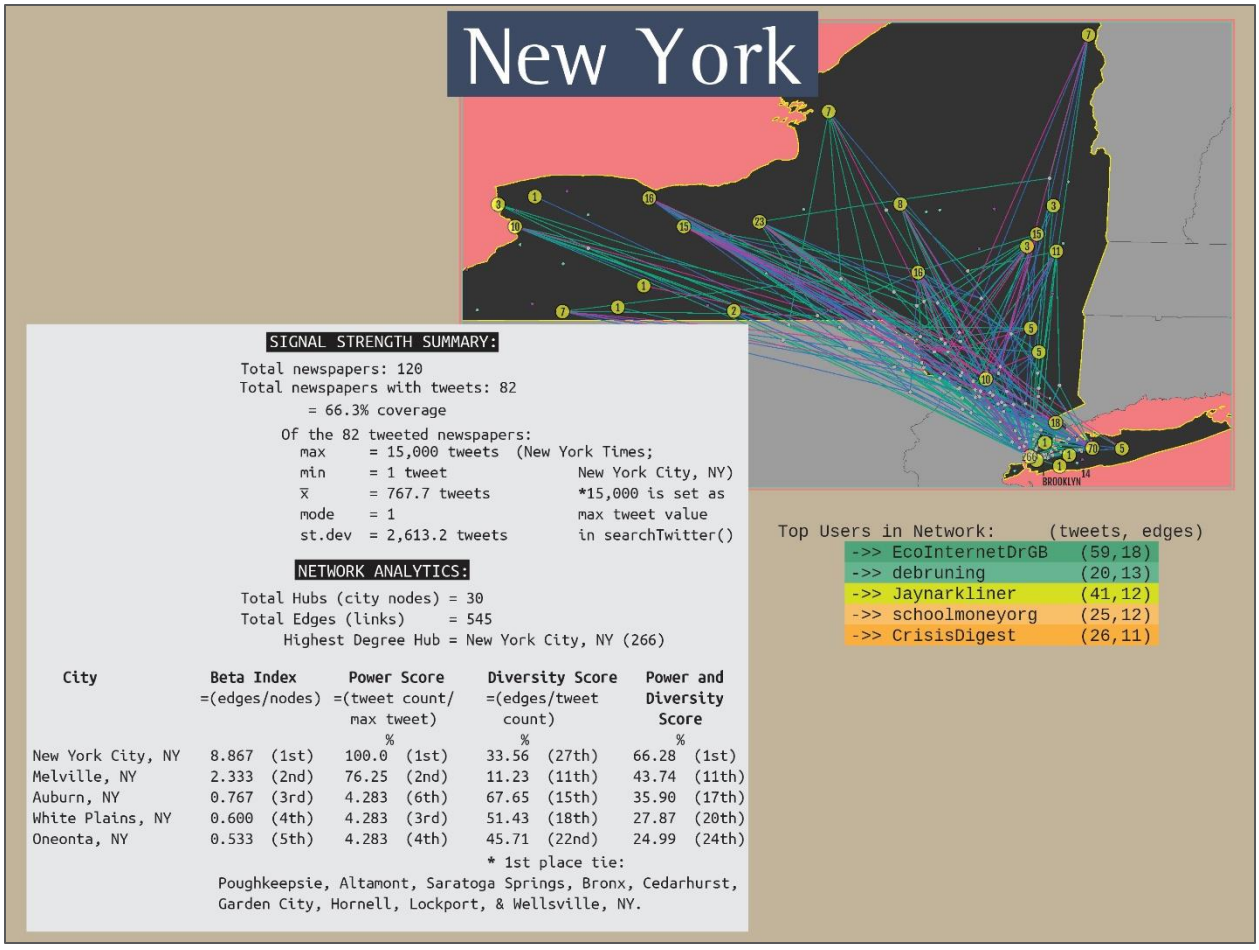


Fig.2: Network map of Twitter users who tweeted 3 or more unique newspapers in New York.

Works Cited

Brock, George. "Out of Print: Newspapers, Journalism And The Business Of News In The Digital Age." KoganPage, 2013. eBook (EBSCOhost).

Chu, Zi,; Gianvecchio, S.; Wang, H.; Jajodia, S. "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?". *IEEE*, vol. 9, no. X, 2012, pp. 1-14.

Dawkins, Richard, "The Selfish Gene." New York, Oxford University Press, 1989.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. "The Rise of Social Bots." *Communications of the ACM*, vol. 59, no. 7, July 2016, pp. 96-104.

Gleick, James. "The Information: A History, A Theory, A Flood." New York, Vintage Books, 2011.

Google API. Geocoded list of 4,852 U.S. cities in March 2019. https://cloud.google.com/maps-platform/.

Hennig-Thurau, Thorsten & Caroline Wiertz. "Does Twitter matter? The impact of microblogging word mouth on consumers' adoption of new movies." *Journal of the Academy of Marketing Science*, vol. 43, 2015, pp. 375-394.

Java, A.; Finin, T.; Song, X.; Tseng, B. "Why We Twitter: Understanding Microblogging Usage and Communities." *Association for Computing Machinery*, vol. 1, no. 1, August 2007, pp. 56-65.

Meyer, Philip. "The Vanishing Newspaper: Saving Journalism in the Information Age." University of Missouri Press, 2004. eBook(ProQuest).

Rushkoff, Douglas. *Team Human.* New York, W.W. Norton & Company, 2019

Thurman, Neil & Richard Fletcher. "Has Digital Distribution Rejunvenated Readership?". *Journalism Studies*, vol. 20, no. 4, 2019, pp. 542-562.

Tolochko, Petro & Hajo G. Boomgaarden. "Analysis of Linguistic Complexity in Professional and Citizen Media." *Journalism Studies*, vol. 19, no. 12, 2018, 1786-1803. eJournal(Taylor&FrancisOnline), https://doi-org.unr.idm.oclc.org/10.1080/1461670X.2017.1305285.

Twitter API. Tweet data collected from March 13 – April 23, 2019. https://developer.twitter.com/en/dashboard.