Emma Burger 12/13/16

SI 330 Final Project Written Report

U.S. Twitter Trends and News Sources

***Project Background & Motivation***

Since the election, there’s been a lot of buzz about echo chambers and the stratified nature of the news we consume. Increasingly, it’s become clear that not only is the news that different groups consume completely different, but that online social networks only further stratify the political conversation, as they have a strong tendency to create echo chambers. People tend to follow others with whom they agree, leading to a high degree of homophily in our social networks. This issue is particularly prevalent on Twitter, which is used more widely as a news-sharing source than many other popular social media such as Instagram and Snapchat.

My project uses nationwide Twitter trends as a metric for the news people are reading and talking about, especially on social media. The aim is to find out which news sources have the greatest overlap with trending Twitter topics. Presumably, the news sources with the highest degree of overlap are the most reflective of the interests and views of the American people.

While the political climate provided the impetus for writing this program, I did not limit the scope of this experiment to strictly political news. Rather, I was interested in finding out the topics most compelling to American Twitter users. As a result, I sought to represent a wide range of genres, topics, viewpoints, target demographics, etc. in my selection of news sources. Broadly speaking, these were sports, traditional news/politics, and lifestyle/entertainment websites, with a diversity of biases within each. Within news, for instance, FOX News and MSNBC were each included, with the aim of seeing which better represented the top Twitter trends on their homepages. The specific sites used in this program are listed below.

**Sports: News:  Lifestyle/Entertainment:**

ESPN FiveThirtyEight Refinery29

Sports Blog Nation CNN BuzzFeed

Bleacher Report FOX News People Magazine

FOX Sports MSNBC

NBC Sports Yahoo News

CBS Sports The Huffington Post

Yahoo Sports The New York Times

The Huddle Politico

Sports Illustrated

***Data Sources Used***

The data sources used in this project were the Twitter API and the homepages of all the sites above. Below is the information for each of these sources.

***Name:*** Twitter API

***Size:*** 1 call made

***Location:*** https://dev.twitter.com/overview/documentation

***Format:*** Web API

***Access Method:*** Gets web API, returns JSON

***Name:*** Homepage URLs from various news sources listed above

***Size:*** between 10KB to 1MB

***Locations:***

http://www.espn.com/

http://www.sbnation.com/

http://bleacherreport.com/

http://www.foxsports.com/

http://www.nbcsports.com/

http://www.cbssports.com/

http://sports.yahoo.com/

http://thehuddle.com/

http://www.si.com/

http://fivethirtyeight.com/

http://www.cnn.com/

http://www.foxnews.com/

http://www.msnbc.com/

https://www.yahoo.com/

http://www.huffingtonpost.com/

http://www.nytimes.com/

http://www.politico.com/

http://www.refinery29.com/

https://www.buzzfeed.com/

http://people.com/

***Format:*** HTML page

***Access Method:*** Fetch and parse HTML in order to isolate school ranking

***Project Data Flow***

1. Fetch each news source homepage’s URL and save the HTML in a variable.

*Input:* URL

*Output:* HTML document

2. Parse the HTML data using BeautifulSoup

*Input:* HTML document

*Output:* String of text on each site’s homepage

3. Retrieve a list of the top US trends from the Twitter API.

*Input:* URL for Twitter API trends

*Output:* List of top U.S. Twitter trends

4. For each news source, count the occurrences of each topic on the trend list on the source’s homepage.

*Input:* List of top U.S. Twitter trends

*Output:* Integer representing trend occurrences count for each source

5. Divide the count (from step 4) for each news source by the length of the trend list to calculate percentage overlap.

*Input:* Integer representing trend occurrences count for each source

*Output:* Integer representing percentage overlap for each source

6. Save these figures to a CSV file.

*Input:* Integer representing percentage overlap for each source

*Output:* CSV file containing each source and their respective percentage overlaps

7. Use the data to produce a bar graph in Plot.ly

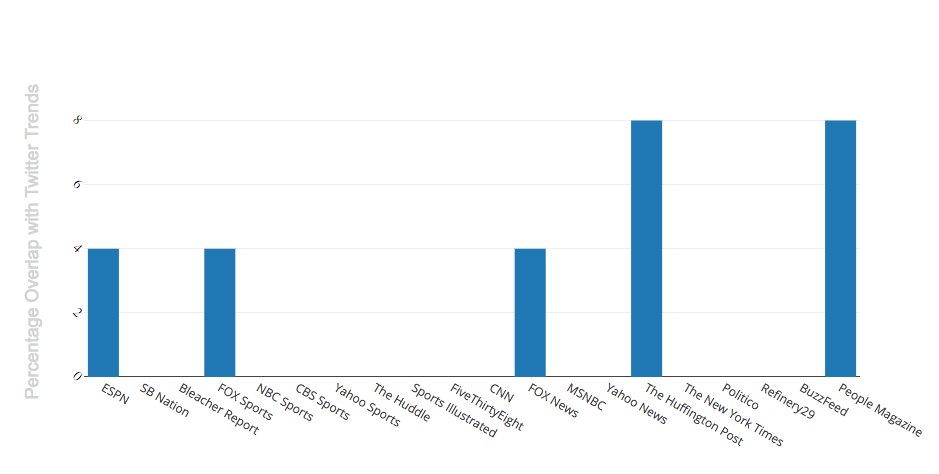
*Input:* CSV file containing each source and their respective percentage overlaps

*Output:* Plot.ly bar graph depicting said percentage overlaps

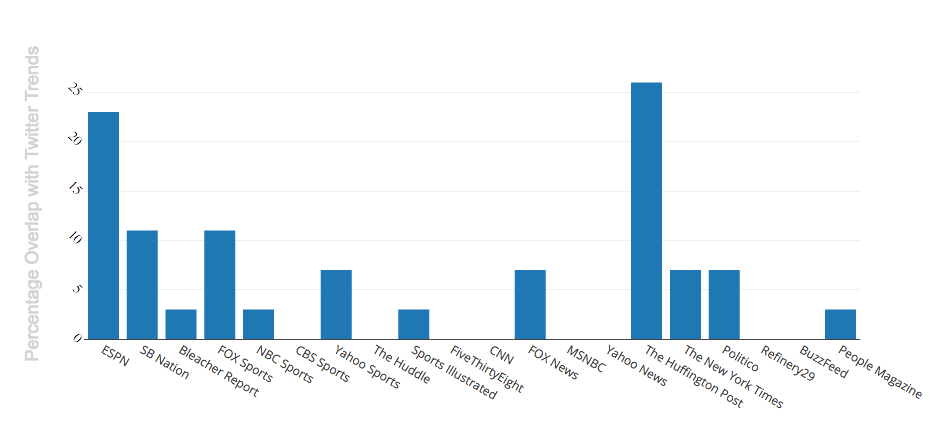
***Findings & Outcomes***

Below are the varied graphical outputs from running the program throughout the day on December 9th.

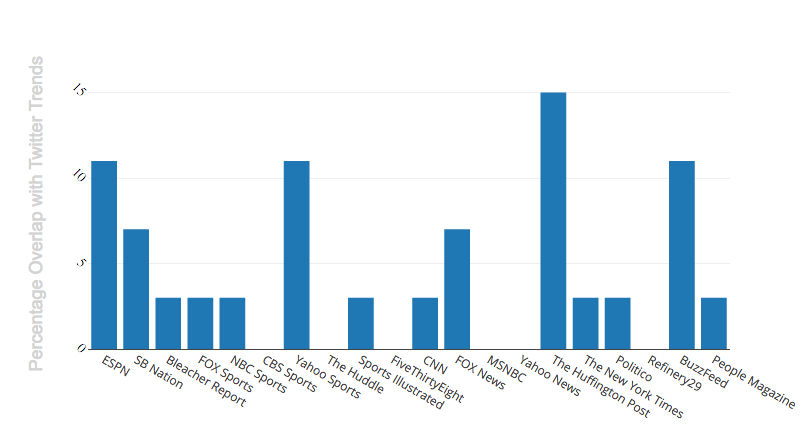
12/9 at 10:00 am



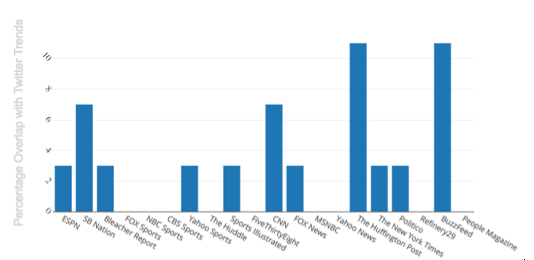
12/9 at 1:00 pm



12/9 at 3:00 pm



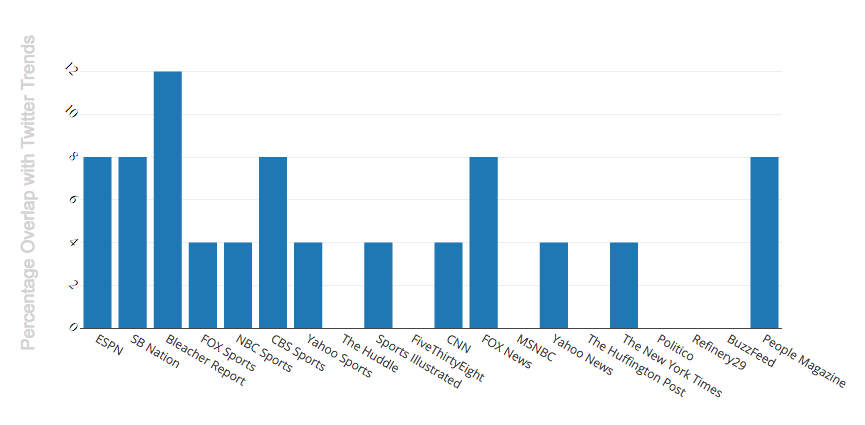
12/9 at 5:00 pm



***Conclusions & Further Study***

Prior to launching the project, my hypothesis was that the sports media would be disproportionately reflected in the trending Twitter topics in comparison with news and lifestyle/entertainment. In particular, I predicted ESPN would have the greatest overlap, being the most prominent of the sports news sources used in the project. Part of the reason this was my prediction was because when I was first researching the possibility of this project, athletes and sports games represented a vast majority of trending Twitter topics. Having run the code now multiple times on different days and at different times during the day, I’ve found that Twitter trends and their representation in various news sources varies significantly and with great frequency depending on the events of any given day, even down to the hour. This immediacy is likely due to the fact that Twitter is a real-time medium, where live tweeting is common and publication is immediate. It is especially easy on Twitter for news to fluctuate significantly throughout the day because messages are limited to 140-characters. This limitation restricts users’ ability to take time to reflect on an event and rather, is conducive to gauging the public’s immediate reaction to a piece of news.

Running the code Monday December 12th at 1:00 pm, Twitter’s trends are 'Lane Kiffin', 'Instagram Live', 'The Tainted Election', 'Carly Fiorina', 'The Russians', 'F-35', 'Flip or Flop', 'ACL and MCL', 'Monty Rice', 'Michael Floyd', 'iOS 10.2', 'AP Top 25', 'AP All-American', 'Black Is Gold', 'DeShone Kizer', 'Tiny Desk Concert', 'Dunkirk', 'Player of the Week', 'Live 360', 'Game Theory', 'AP Poll', 'Lytle Tunnel', "Shooter's Choice", and 'Lionel Richie'. Within the top three, all are related to major news stories still developing Monday morning. Lane Kiffin was appointed head coach of Florida Atlantic’s football team; the live streaming function on Instagram was released; and Paul Krugman wrote a viral op-ed confirming the Russian government’s role in tipping the 2016 election for Donald Trump. This evidence supports the broader tendency shown by my program for the overlap between trends and news sources to be more reflective of the current state of the world (down to the hour) rather than a broader bias inherent to Twitter’s users that favors any one news source or genre over any other. Of all the sources, *Bleacher Report* currently has the greatest overlap with the current trend list.



Compared to the trials run throughout the day on December 9th, *Bleacher Report* holds a significant lead in their coverage of Twitter trends, as it trailed either ESPN or SB Nation all day on the 9th. In the news and politics realm, note the decrease in overlap with *The Huffington Post*, which was a leader in overlap all day on the 9th. These dramatic shifts exemplify the impact of day to day changes in current events on the interaction between news sources and Twitter users.

Similar changes in the data were also reflected when the code was run throughout the day December 8th and December 10th, indicating that there is no one site that consistently exceeds the others. Beyond that, the category of site with the most overlap is not stagnant either. The results from December 9th, for instance, do not support my initial hypothesis that sports news sources in particular would have a much greater overlap with Twitter trends than their news or entertainment/lifestyle counterparts. When there’s a big game on, the sports sites are likely to reign supreme; had I run this same program during the election, it would likely have been tipped heavily in favor of news and politics sites such as *CNN*, *FiveThirtyEight* and *Politico*. Due to the time constraints of this project, it’s impossible to tell whether these past few days of running the code are representative of these trends as a whole. However, this limited sample seems to indicate that there is great variety between Twitter users’ news sources, which appear to be constantly in flux. Continuing to run the code could provide a more holistic picture and allow us to gain insight more broadly into where most users get their news—or at least where their sentiments are echoed most.

***References***

<https://thomassileo.name/blog/2013/01/25/using-twitter-rest-api-v1-dot-1-with-python/>

Used some of his code for providing authorization information to the Twitter API. Citation is noted in-line in my code.