

Analysis of Twitter's Trending Topics in Comparison to Top News Headlines

Surya Keswani, Zion Calvo & Donald Stewart
University of California, Santa Cruz



Introduction

Twitter services millions of people with content that ranges from news to memes. The platform’s explosive growth over the last decade and a half has caught the attention of many researchers. They ask: does Twitter censor it’s trending topics list and should it be allowed to continue this practice? Controversies have emerged when important and newsworthy topics such as the elections or Ebola do not trend but Justin Bieber is frequently trending on the platform [2]. The objective of this project is to analyze what Twitter flags as a trending topic and what factors contribute to a topic receiving the trending label. We compare the trending topics to the top news headlines put out by some of the largest news sources in the world.



Abstract

We explore the types of content that people consume on social media. This presents the question: Are people receiving their daily world/political news from social media or are they simply being consumed by other pop culture news? Is the content that people are consuming worthy of being “trending”? Is pop culture news important news for users to be receiving? This article addresses the trends and the type of content that users are consuming on Twitter.

Models and Algorithms

Getting a deeper look into Twitter’s machine learning algorithm requires the collection and analysis of data that incases Twitter's most prevalent subjects and most popular news headlines. The “Twitter Trending Topics” dataset was modified, a headline dataset was web scrapped, and visualizations were created using Google’s Facets Data Visualization Library.

Dataset Modification: By iterating through the hashed data, the number of tweets, unique users, and average traffic was obtained and added to each trending topic in the data set.

Facets: This library builds interactive visualizations so users can investigate their data for patterns. This library is especially useful when analyzing large and complex amounts of data. By providing this library with a “Pandas” data frame, the Facets library manipulates it into a graph that can instantly alter its axes with any field of the dataset. Our Facet visualization reveals the most common tendencies within the most popular topics.

Dataset

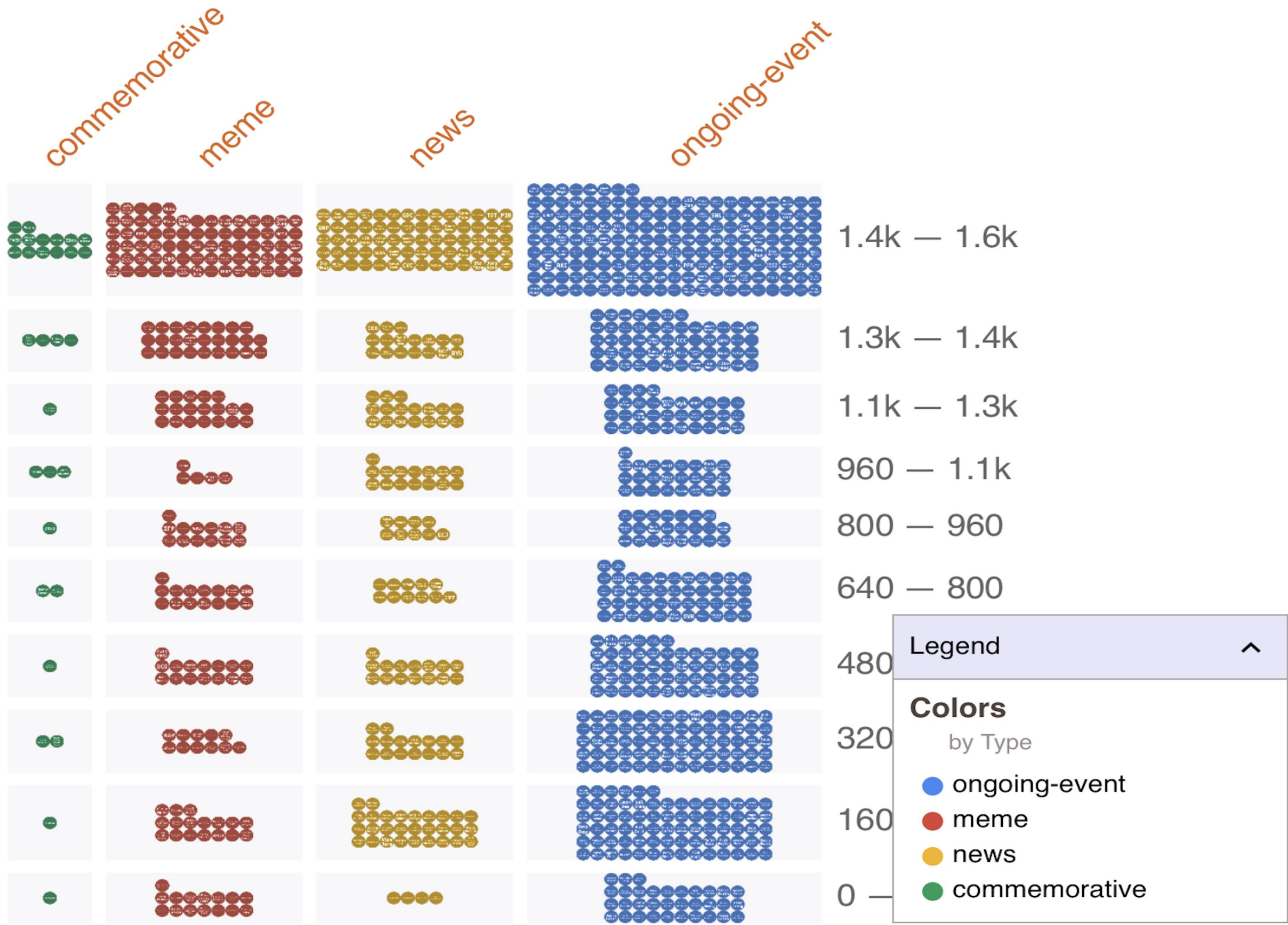
To analyze the most frequently viewed subject matter on Twitter, the dataset “Twitter Trending Topics” was utilized. It contains trending topics from 03/01/2011 - 03/08/2011. This dataset originally consisted of five fields: Md5 Hash, Date, Topic, and Type. We append three new fields to the existing dataset: the number of Tweets, Unique Users, and Average Traffic. These addition fields create more numeric features that allow for a further investigation of the algorithm Twitter used (at the date specified above) to determine a topic trending.

	Md5 Hash	Date	Topic	Type	# Tweets	Unique Users	Average Traffic
0	e86fdb234c7429d9baa70c5f104dbecb	Sunday 3/6/11	#mileysnarl	ongoing-event	1358	536	947.0
1	fe2dca2b2bfca21e6c83567a531d6534e	Sunday 3/6/11	Leap Year	ongoing-event	1496	1335	1415.5
2	cf90d03111f5e65f9bc7b5e079fa30	Saturday 3/5/11	Sober Valley Lodge	meme	1480	1376	1433.0
3	748089790952e9dbfb174b7ef93ad72b	Thursday 3/3/11	Howard Davies	news	557	452	504.5
4	3b7dd11b6654dd5f5745285deddbd9	Thursday 3/3/11	Sky News	news	1483	1203	1343.0
...
1031	05c739a093794343728b6d5d802db10b	Wednesday 3/2/11	Antonio Damasio	ongoing-event	197	166	181.5
1032	8669b32af04ae65b8bc38df5662545ea	Saturday 3/5/11	Shelia Mello	ongoing-event	190	170	180.0
1033	4e49ec9f0f039f91a5b34278d15484f3	Friday 3/4/11	Gareca	ongoing-event	284	222	253.0
1034	d93d84a006f530cb2b33ebf824d48e	Monday 3/7/11	Patricia Poeta	ongoing-event	309	283	296.0
1035	98d79b7652a65b4a4b260ca2b2b5583d	Thursday 3/3/11	PIB	news	1497	1176	1336.5

Figure 1: An example of the Pandas Data Frame

Creating comparison data to have context of pressing events throughout the time period, we web scraped headlines from five new sources: The New York Times, The Wall Street Journal, Fox News, The Washington Post, and CNN. This data set contains the following fields: Source, Date, and Headline.

Top Trending Topics by Number & Type of Tweets



The figure illustrates that approximately 50% of trending topics have 1400 to 1600 tweets associated with them. The volume of tweets is an important factor in determining if a topic becomes a trending label. The dataset also reveals that Twitter follows a pre-set distribution when classifying topics as Trending. On average, approximately 55% of the Trending topics are ongoing events, 20% are memes, 20% are news, and 5% are commemorative.

Results

Top 5 Trending Topics based on Average Traffic Below:

	Md5 Hash	Date	Topic	Type	# Tweets	Unique Users	Average Traffic
607	f6c5c0ae6c636b204045d21267608be5	Saturday 3/5/11	YOURMAN	meme	1492	1479	1485.5
711	6401813085837e9f1f8b27eae98ac4f3	Saturday 3/5/11	Mountain Dew	meme	1493	1454	1473.5
1011	5eb1c412235828a21e3767c12ab9975	Friday 3/4/11	Fast Times	ongoing-event	1498	1441	1469.5
10	e8611c0126305e7a3af7eb6317dc699	Tuesday 3/1/11	RIP Jane Russell	news	1497	1441	1469.0
408	257c941753478852a2436136bc72ab61	Wednesday 3/2/11	TOP10 Profile STALKERS	meme	1496	1439	1467.5

Top 5 Trending Topics based on Unique Users Below:

	Md5 Hash	Date	Topic	Type	# Tweets	Unique Users	Average Traffic
607	f6c5c0ae6c636b204045d21267608be5	Saturday 3/5/11	YOURMAN	meme	1492	1479	1485.5
711	6401813085837e9f1f8b27eae98ac4f3	Saturday 3/5/11	Mountain Dew	meme	1493	1454	1473.5
877	463636c775a2ba72afdf29d240b4a19	Wednesday 3/2/11	Nancy Grace	meme	1464	1445	1454.5
736	c80e68f4191a8150a52cfc543a8bdaf	Wednesday 3/2/11	Still Winning	meme	1462	1442	1452.0
10	e8611c0126305e7a3af7eb6317dc699	Tuesday 3/1/11	RIP Jane Russell	news	1497	1441	1469.0

Top 5 Trending Topics based on # of Tweets Below:

	Md5 Hash	Date	Topic	Type	# Tweets	Unique Users	Average Traffic
70	050feefcd07d85023c61753474848a48	Monday 3/7/11	Rosenmontag	meme	1500	1298	1399.0
114	4a4de7aa4553a1bd4feaf73042d479f6b	Tuesday 3/1/11	Robot Chicken	ongoing-event	1500	1279	1389.5
460	6a5c42c43e58b8e6d7ab5da8d965c51	Wednesday 3/2/11	Wild Thing	ongoing-event	1500	1429	1464.5
602	d6d9a283570262f6978f7dbab2c5bc41	Wednesday 3/2/11	Kalla	news	1500	1203	1351.5
809	9f084e11ff093258187e9a5e19c1a	Thursday 3/3/11	Wellington Paulista	ongoing-event	1500	987	1243.5

The figure above shows the top five trending topics classified by Average Traffic, # of Unique Users, and # of Tweets per topic. During the week this dataset was collected, a civil war broke out in Libya. Based on the 3 popularity metrics measured (Average Traffic, # of Unique Users, and # of Tweets), none of the top 5 charts cover the Libyan Civil War as trending. In fact, the Libyan Civil War is not mentioned in the top 35 Trending topics for any of the popularity measures. Of the top 35 news headlines we have curated from a diverse pool of news sources, 16 of the 35 headlines cover the Civil War. The trending topics seem to be more focused on popular culture as the top trending topics are dictated by memes.



REFERENCES

- [1] Elisa Shearer and Katerina Eva Matsa. 2019. News Use Across Social Media Platforms 2018. (December 2019). Retrieved March 5, 2020
- [2] Tarleton Gillespie. 2011. Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us. (October 2011). Retrieved March 5, 2020
- [3] Lahle Wolfe. 2019. Twitter Statistics: How Many People Use Twitter. (October 2019). Retrieved March 5, 2020
- [4] Anon.
- [5] A. Zubiaga, D. Spina, V. Fresno, R. Martinez. *Real-Time Classification of Twitter Trends* Journal of the American Society for Information Science and Technology (JASIST). In Press.