# Network Analysis With Twitter Data

Zachary C. Steinert-Threlkeld

# Purpose

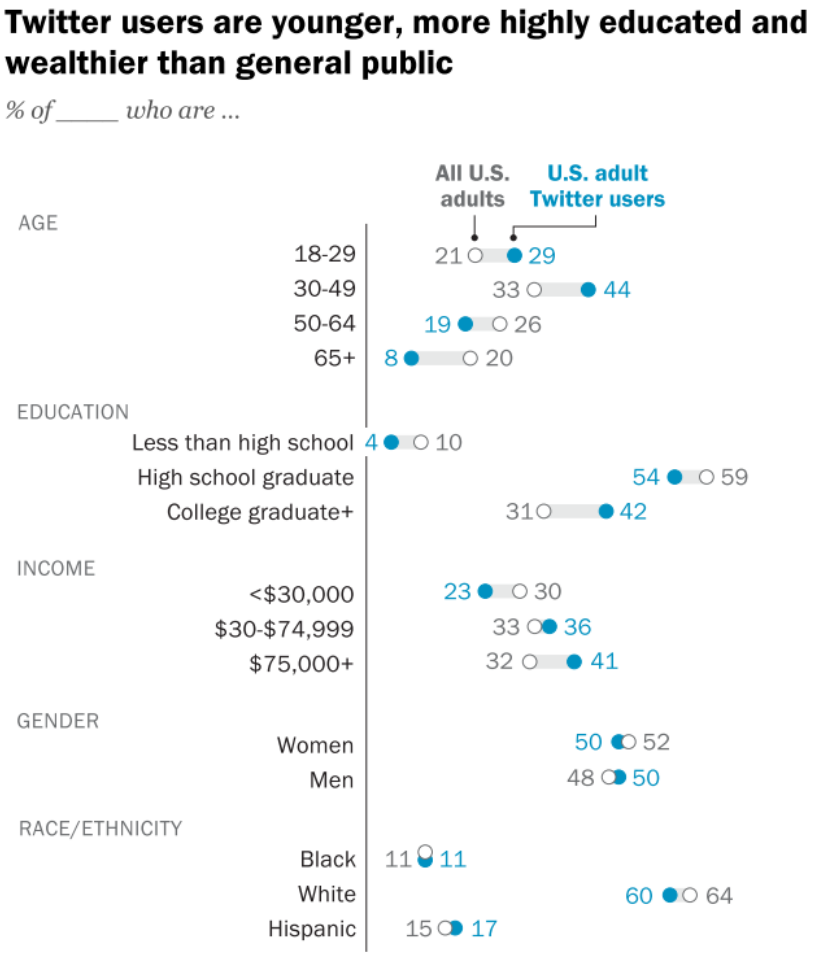
Explain how to get data from Twitter, types of network data available from it, and example studies.

# Who Uses Twitter and Why

The following figures come from two Pew Research Center reports:

1. "Sizing Up Twitter Users" (2019)
2. "The Behaviors and Attitudes of U.S. Adults on Twitter" (2021)

# Demographics

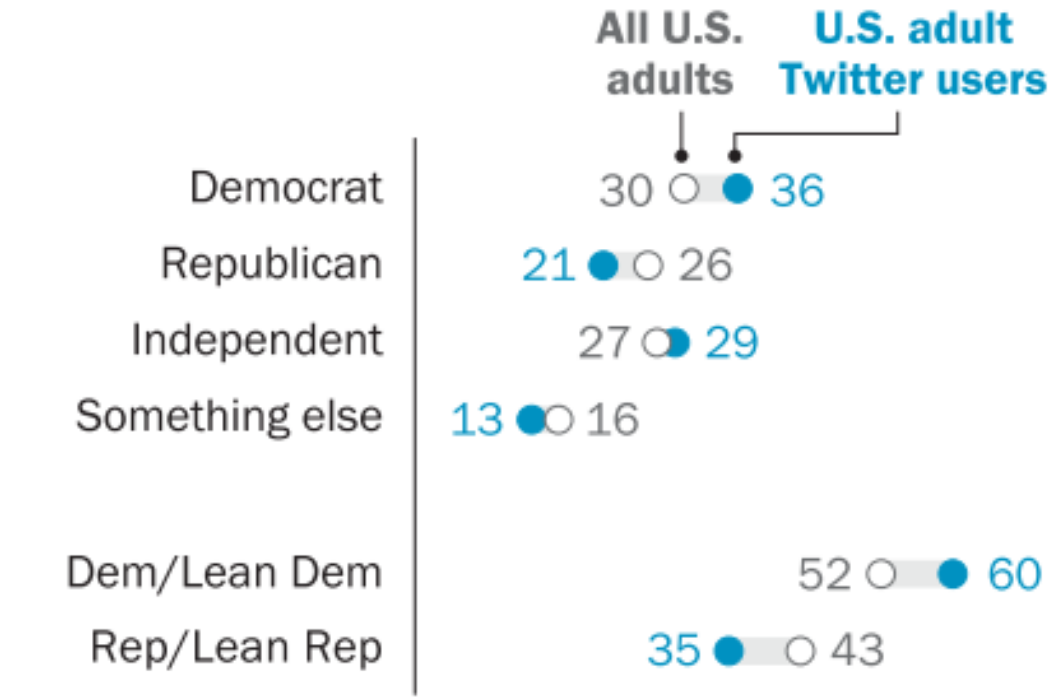



Note: Whites and blacks include only non-Hispanics. Hispanics are of any race.
Source: Survey of U.S. adult Twitter users conducted Nov. 21-Dec. 17, 2018, and survey of U.S. adults conducted Nov. 7-11, 2018.
"Sizing Up Twitter Users"

**PEW RESEARCH CENTER**

## Twitter users more likely to identify as Democrats than Republicans

*% of _____ who identify as...*




Source: Survey of U.S. adult Twitter users conducted Nov. 21-Dec. 17, 2018, and survey of U.S. adults conducted Nov. 7-11, 2018. "Sizing Up Twitter Users"

**PEW RESEARCH CENTER**

# Public

## Majority of Twitter users surveyed have public accounts, even if they say it is set to private or are not sure of their settings

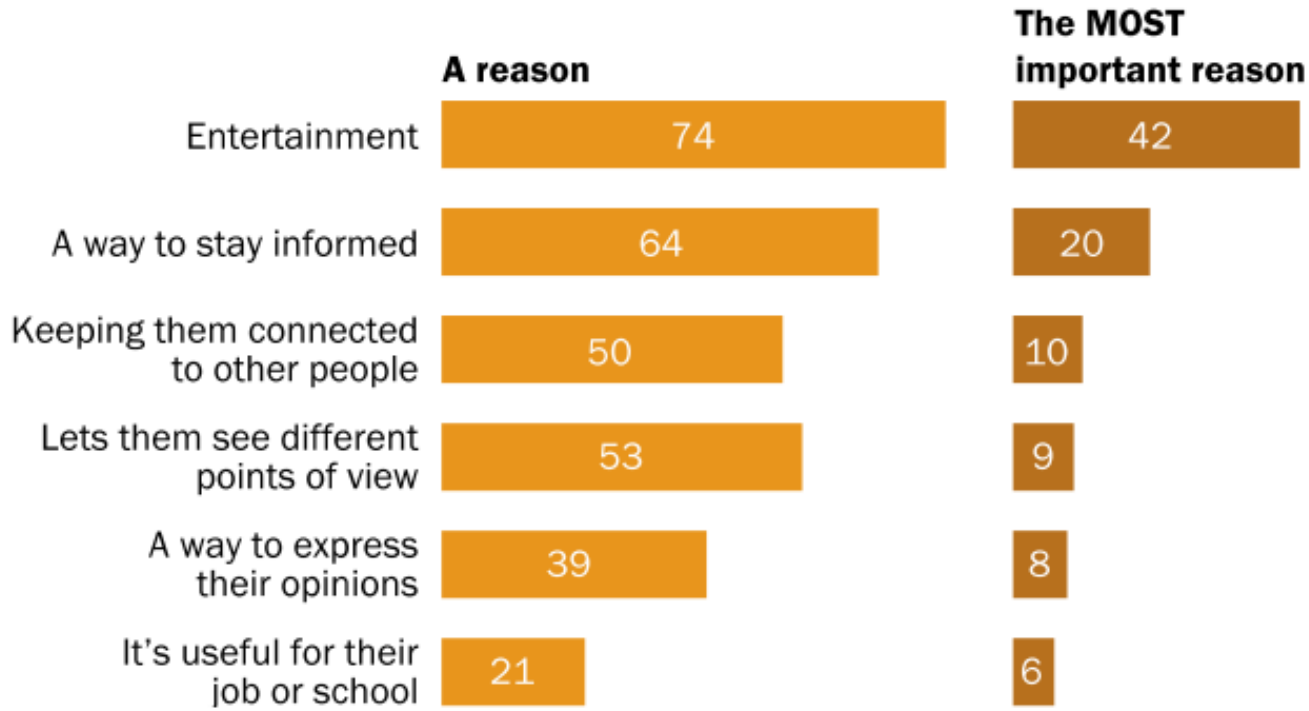

Note: Based on users who provided a valid handle.
Source: Survey of U.S. adult Twitter users conducted May 17-31, 2021. Data on respondents' Twitter accounts collected as of July 28, 2021, using Twitter API.
"The Behaviors and Attitudes of U.S. Adults on Twitter"

PEW RESEARCH CENTER

# Reasons

## About four-in-ten Twitter users say the most important reason they use the site is for entertainment ...

*% of U.S. adult Twitter users who say each of the following is ____ they use Twitter*




## ... but users 50 and older are slightly more likely than younger users to say it is to stay informed, connected to others, see different points of view

# Twitter for Research

- Early platform
- Global *with a norm of public production*.
- Responsive to events
- Easy to get data
  - Text is easy to analyze
  - Free free free free free free free

# Tweet

**ZacharyST**
@ZacharyST

What papers do I need to read about measuring emotion in tweets.  Fear, anger, shame etc., not valence (positive, negative, neutral.).  #rstats #twitter #nlp #python #nltk

1:46 PM · May 1, 2020 · Twitter Web App

View Tweet activity

**3** Retweets **12** Likes

Tweet your reply

Reply

**Jason Jeffrey Jones** @jasonjones_jjj · Aug 10, 2020
Replying to @ZacharyST
Paper from the DeepMoji project at MIT: arxiv.org/abs/1708.00524 "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm"

# From Tweets

**Centrality**

- In-degree (followers) and out-degree (following)
- Delivered with the tweet (v1.1) or user object (v2).

# From Tweets

**Structure**

- Retweet
- User mention
- Bipartite
  - Location
  - Hashtags

# "Actual" Edges

- Follower IDs
- Following IDs

# Tweet

**ZacharyST**
@ZacharyST

What papers do I need to read about measuring emotion in tweets. Fear, anger, shame etc., not valence (positive, negative, neutral.). #rstats #twitter #nlp #python #nltk

1:46 PM · May 1, 2020 · Twitter Web App

View Tweet activity

**3** Retweets **12** Likes

Tweet your reply

Reply

**Jason Jeffrey Jones** @jasonjones_jjj · Aug 10, 2020
Replying to @ZacharyST
Paper from the DeepMoji project at MIT: arxiv.org/abs/1708.00524 "Using millions of emoji occurrences to learn any-domain representations for

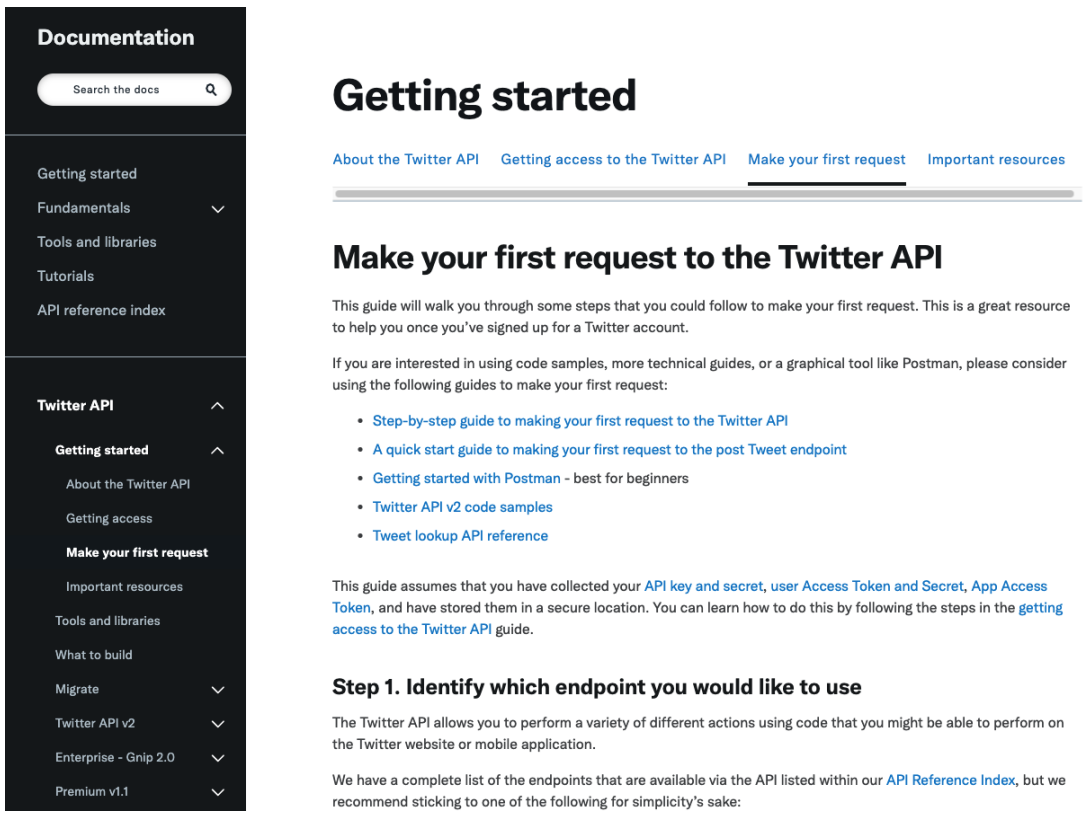
# Get Access

https://developer.twitter.com/en/portal/petition/essential/basic-info

# As Data, API Documentation

To learn about the data available and how to download it, read the Application Programming Interface (API) documentation at https://developer.twitter.com/en/docs/twitter-api.

Example of getting started:



# API in Transition

Twitter is phasing out v1.1 of the API and replacing it with v2.

- v2 improvements: potential to access every tweet ever; more detailed query parameters; conversation ID.
- v2 downgrades: monthly tweet quota (except for sample stream), data delivered requires more processing.

# As Data, JavaScript Object Notation (JSON)

JSON is a formatting standard for delivering data. It is a method of formatting text just like a .csv file is. It is equivalent to R's list and Python's dictionary.

Example:

```
{ "employees":[
    {"firstName":"John", "lastName":"Doe"},
    {"firstName":"Anna", "lastName":"Smith"},
    {"firstName":"Peter", "lastName":"Jones"}
]
}
```

# As Data, API v1.1

{"in_reply_to_user_id": null, "in_reply_to_status_id_str": null, "retweeted": false, "truncated": true, "favorited": false, "text": "What papers do I need to read about measuring emotion in tweets. Fear, anger, shame etc., not valence (positive, nǍ026 https://t.co/Ja3yF9CcuJ", "contributors": null, "place": null, "in_reply_to_screen_name": null, "favorite_count": 12, "created_at": "Fri May 01 20:46:42 +0000 2020", "user": {"follow_request_sent": false, "friends_count": 582, "notifications": false, "withheld_in_countries": [], "is_translator": false, "contributors_enabled": false, "location": "Los Angeles, CA", "profile_image_url": "http://pbs.twimg.com/profile_images/378800000467355220/5ccd66a33f4f8a4e41712f675eb4cd6e_normal.jpeg", "is_translation_enabled": false, "profile_link_color": "FA743E", "time_zone": null, "profile_background_image_url": "http://abs.twimg.com/images/themes/theme12/bg.gif", "screen_name": "ZacharyST", "url": "http://t.co/2mmy0mI5j8", "created_at": "Thu Nov 13 15:52:14 +0000 2008", "profile_use_background_image": false, "profile_image_url_https": "https://pbs.twimg.com/profile_images/378800000467355220/5ccd66a33f4f8a4e41712f675eb4cd6e_normal.jpeg", "id": 17367248, "profile_background_color": "000000", "following": false, "followers_count": 633, "lang": null, "favourites_count": 1081, "id_str": "17367248", "profile_background_tile": false, "profile_text_color": "000000", "translator_type": "none", "name": "ZacharyST", "default_profile_image": false, "profile_sidebar_border_color": "000000", "geo_enabled": true, "utc_offset": null, "statuses_count": 1488, "has_extended_profile": false, "entities": {"url": {"urls": [{"indices": [0, 22], "url": "http://t.co/2mmy0mI5j8", "display_url": "zacharyst.com", "expanded_url": "http://www.zacharyst.com"}]}, "description": {"urls": []}}, "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme12/bg.gif", "description": "Prof. of public policy at UCLA's Luskin School of Public Affairs. Protests, computational methods, social networks. Also enjoy other things.", "profile_sidebar_fill_color": "000000", "listed_count": 10, "protected": false, "default_profile": false, "verified": false}, "in_reply_to_status_id": null, "id": 1256324187917914113, "retweet_count": 3, "lang": "en", "id_str": "1256324187917914113", "coordinates": null, "source": "¡a href=ḧttps://mobile.twitter.comr̈el=n̈ofollow¿Twitter Web App¡/a¿", "in_reply_to_user_id_str": null, "geo": null, "is_quote_status": false, "entities": {"urls": [{"indices": [117, 140], "url": "https://t.co/Ja3yF9CcuJ", "display_url": "twitter.com/i/web/status/1Ǎ026", "expanded_url": "https://twitter.com/i/web/status/1256324187917914113"}], "user_mentions": [], "hashtags": [], "symbols": []}}

# As Data, API v1.1


```
Out[6]:
{'contributors': None,
 'coordinates': None,
 'created_at': 'Fri May 01 20:46:42 +0000 2020',
 'entities': {'hashtags': [],
  'symbols': [],
  'urls': [{'display_url': 'twitter.com/i/web/status/1…',
    'expanded_url': 'https://twitter.com/i/web/status/1256324187917914113',
    'indices': [117, 140],
    'url': 'https://t.co/Ja3yF9CcuJ'}],
  'user_mentions': []},
 'favorite_count': 12,
 'favorited': False,
 'geo': None,
 'id': 1256324187917914113,
 'id_str': '1256324187917914113',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'place': None,
 'retweet_count': 3,
 'retweeted': False,
 'source': '<a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>',
 'text': 'What papers do I need to read about measuring emotion in tweets.  Fear, anger, shame etc., not valence (positive, n… https://t.co/Ja3yF9CcuJ',
 'truncated': True,
 'user': {'contributors_enabled': False,
  'created_at': 'Thu Nov 13 15:52:14 +0000 2008',
  'default_profile': False,
  'default_profile_image': False,
  'description': "Prof. of public policy at UCLA's Luskin School of Public Affairs. Protests, computational methods, social networks.  Also enjoy other things.",
  'entities': {'description': {'urls': []},
   'url': {'urls': [{'display_url': 'zacharyst.com',
      'expanded_url': 'http://www.zacharyst.com',
      'indices': [0, 22],
      'url': 'http://t.co/2mmy0mI5j8'}]}},
  'favourites_count': 1081,
  'follow_request_sent': False,
  'followers_count': 633,
  'following': False,
  'friends_count': 582,
  'geo_enabled': True,
  'has_extended_profile': False,
  'id': 17367248,
  'id_str': '17367248',
  'is_translation_enabled': False,
  'is_translator': False,
  'lang': None,
  'listed_count': 10,
  'location': 'Los Angeles, CA',
  'name': 'ZacharyST',
  'notifications': False,
  'profile_background_color': '000000',
  'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme12/bg.gif',
  'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme12/bg.gif',
  'profile_background_tile': False,
  'profile_image_url': 'http://pbs.twimg.com/profile_images/378800000467355220/5ccd66a33f4f8a4e41712f675eb4cd6e_normal.jpeg',
  'profile_image_url_https': 'https://pbs.twimg.com/profile_images/378800000467355220/5ccd66a33f4f8a4e41712f675eb4cd6e_normal.jpeg',
  'profile_link_color': 'FA743E',
  'profile_sidebar_border_color': '000000',
  'profile_sidebar_fill_color': '000000',
  'profile_text_color': '000000',
  'profile_use_background_image': False,
  'protected': False,
  'screen_name': 'ZacharyST',
  'statuses_count': 1488,
  'time_zone': None,
  'translator_type': 'none',
  'url': 'http://t.co/2mmy0mI5j8',
  'utc_offset': None,
  'verified': False,
  'withheld_in_countries': []}}
```

# As Data, API v2

```
In [109]: result.keys()
Out[109]: dict_keys(['meta', 'data', 'includes'])
```

{"entities": {"hashtags": [{"tag": "rstats", "start": 136, "end": 143}, {"tag": "twitter", "start": 144, "end": 152}, {"tag": "nlp", "start": 153, "end": 157}, {"tag": "python", "start": 158, "end": 165}, {"tag": "nltk", "start": 166, "end": 171}]}, "text": "What papers do I need to read about measuring emotion in tweets. Fear, anger, shame etc., not valence (positive, negative, neutral.). #rstats #twitter #nlp #python #nltk", "created_at": "2020-05-01T20:46:42.000Z", "conversation_id": "1256324187917914113", "source": "Twitter Web App", "id": "1256324187917914113", "possibly_sensitive": false, "author_id": "17367248", "lang": "en"}

{"users": [{"public_metrics": {"following_count": 583, "followers_count": 633, "tweet_count": 1488, "listed_count": 10}, "name": "ZacharyST", "pinned_tweet_id": "1256324187917914113", "location": "Los Angeles, CA", "profile_image_url": "https://pbs.twimg.com/profile_images/378800000467355220/5ccd66a33f4f8a4e41712f675eb4cd6e_normal.jpeg", "created_at": "2008-11-13T15:52:14.000Z", "username": "ZacharyST", "entities": {"url": {"urls": [{"expanded_url": "http://www.zacharyst.com", "display_url": "zacharyst.com", "url": "http://t.co/2mmy0mI5j8", "start": 0, "end": 22}]}}, "url": "http://t.co/2mmy0mI5j8", "description": "Prof. of public policy at UCLA's Luskin School of Public Affairs. Protests, computational methods, social networks. Also enjoy other things.", "id": "17367248", "protected": false, "verified": false}]}

# API - Streaming Endpoints

- Receive tweets in real time
  - Random sample
  - Filtered sample parameters: bounding box, specific place, language, keywords, users.
- Quantity
  - API v1.1: 1% of all tweets, $\approx$ 5 million per day.
  - API v2: $\leq$ 500,000, 2 million, or 10 million per month.

# API - Search Endpoints

- Retrieve tweets created before the time of query.
  - API v1.1: previous seven days only.
  - API v2: previous seven days only *unless Twitter grants Academic Research access*.
- API v2 parameters: location (place name, country code, point + radius), keywords, language, emoji, user, retweets, & many more.

# API - Other Useful Endpoints

1. Users lookup
2. Users' follower or friends
   - Especially powerful with users lookup.
   - Can infer when a follower started following an account.
3. User's tweets
   - Previous 3200 in v1.1 (GET statuses/user_timeline)
   - Use search endpoint in v2, previous week or all (Academic Research)

# Sample Code in Several Languages

Twitter provides sample code for v2 of the API at this Github repository.

# R Packages

- academictwitteR - API v2 only; can convert JSON objects.
- rtweet - API v1.1 only; converts JSON objects.
- streamR - API v1.1 streaming; JSON or .csv; what I use.
- twitteR - do not use (not updated since 2015)

# Other Sources of Data

- DocNow Catalog
- Harvard Dataverse
- Twitter Transparency - Information Operations
- Google "<keyword> tweets dataset"
- Ask people to share tweet IDs.

# Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter (2011)

**Types of Data**

- GET statuses/user_timeline, easy to guess user IDs; 08.2009-01.2010; 3 billion tweets, 60 million users.
- Network: user mentions $t \geq 3$ times.
  - 8,509,140 nodes in giant connected component, 50,814,366 edges
  - Directed

# Main Result

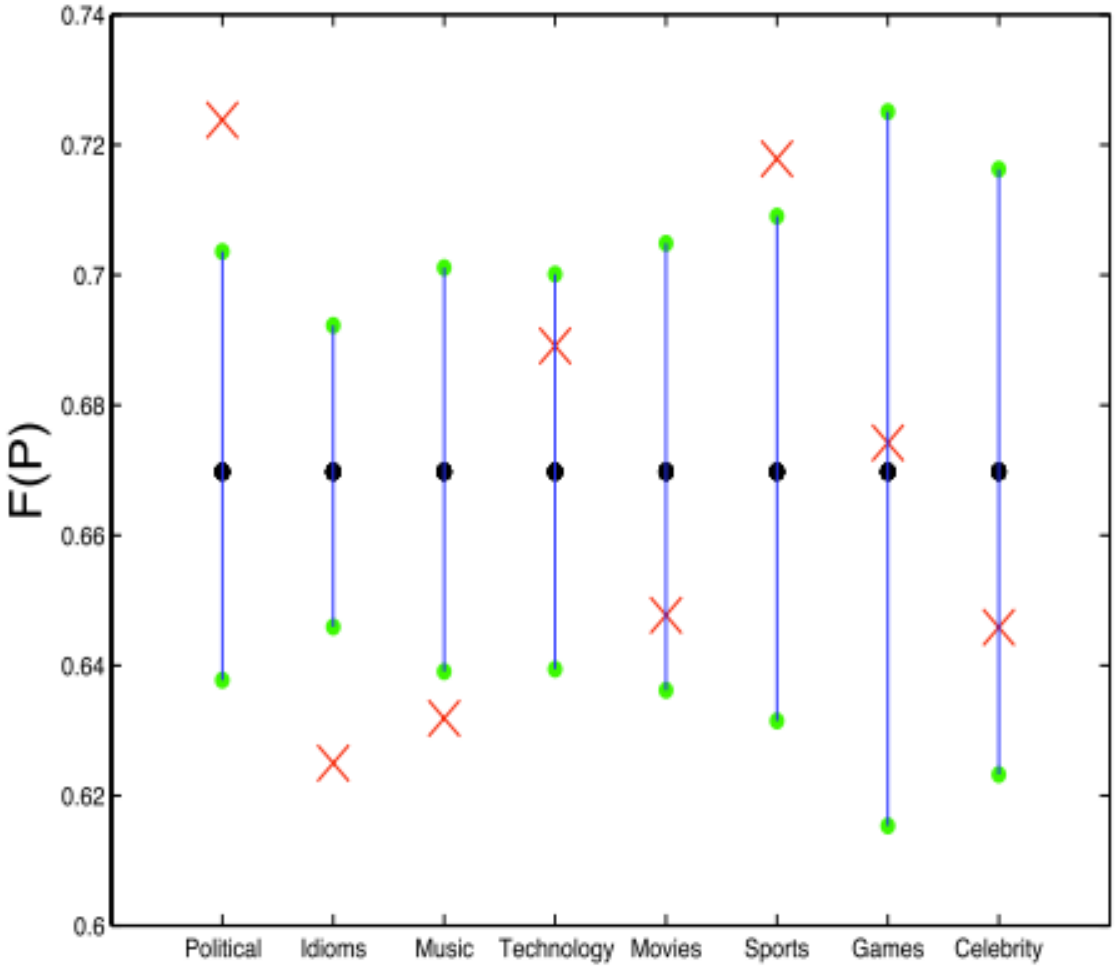



**Figure 2:** $F(P)$ **for the different types of hashtags.The black dots are the average** $F(P)$ **among all hashtags, the red x is the average for the specific category, and the green dots indicate the 90% expected interval where the average for the specific set of**

# Sentiment Cascades in the 15M Movement (2015)

**Types of Data**

- Vague: SMMART platform is proprietary, "no further details are available".
- Keywords to identify tweets, then extract follower and following network. Keep only "active followers".
- K-core centrality.
- Sentiment (positive, negative, neutral) from the SentiStrength dictionary.

# Main Result

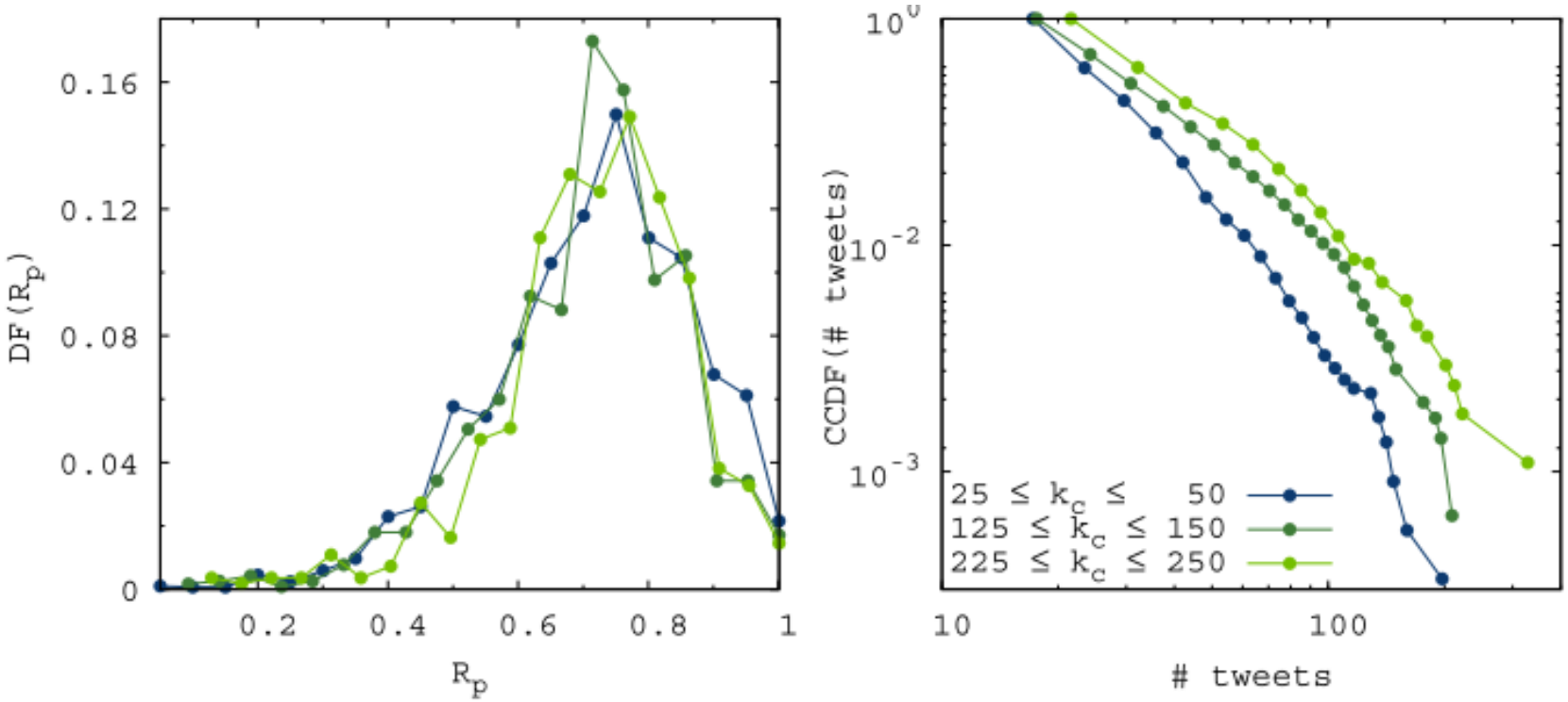



**Figure 3 Positivity and engagement for three ranges of integration in the movement.** Left: probability density function of the ratio of positive tweets of participants, for three ranges of $k_c$. Right: CCDF of the engagement of participants, measured by their amount of tweets about 15M, for three intervals of $k_c$. Participants with higher integration in the movement are more engaged and active in the online medium.

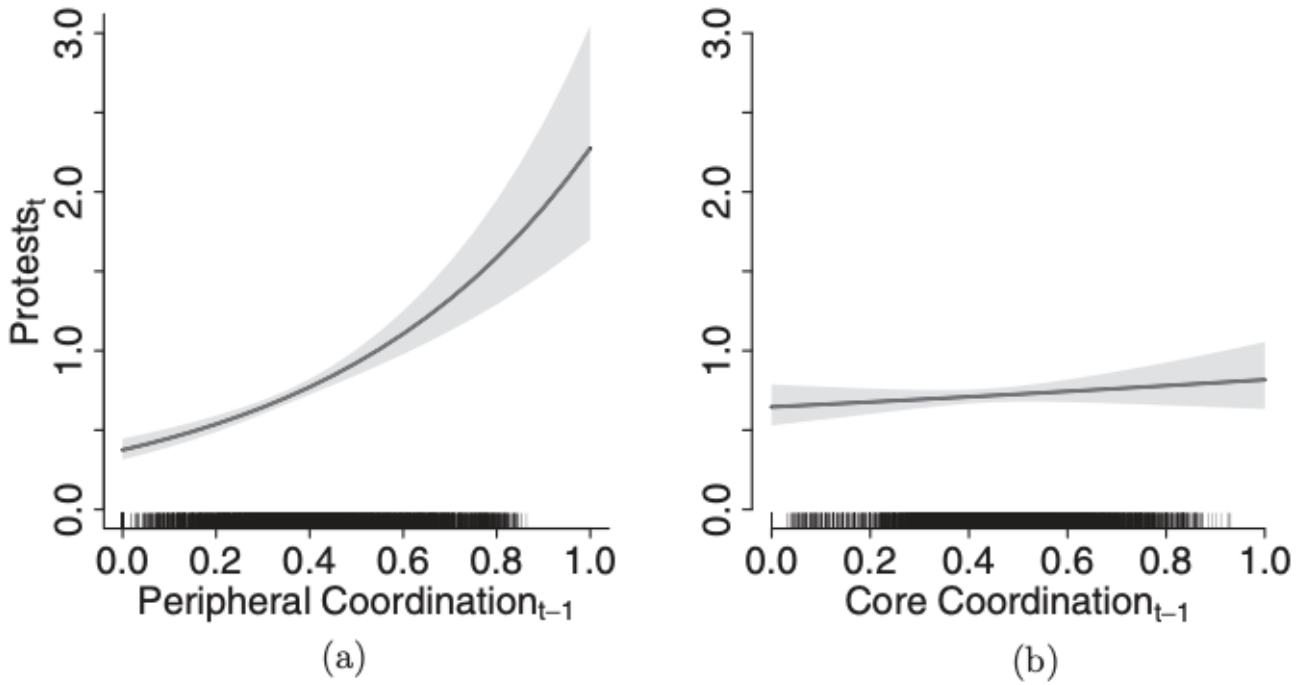
# Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring (2017)

**Types of Data**

- Request streamed tweets from Alessandro Vespignani at Northeastern University.
- Number of followers per user, provided with each tweet.

# Main Result

**FIGURE 2. Marginal Effects of Peripheral and Core Coordination**



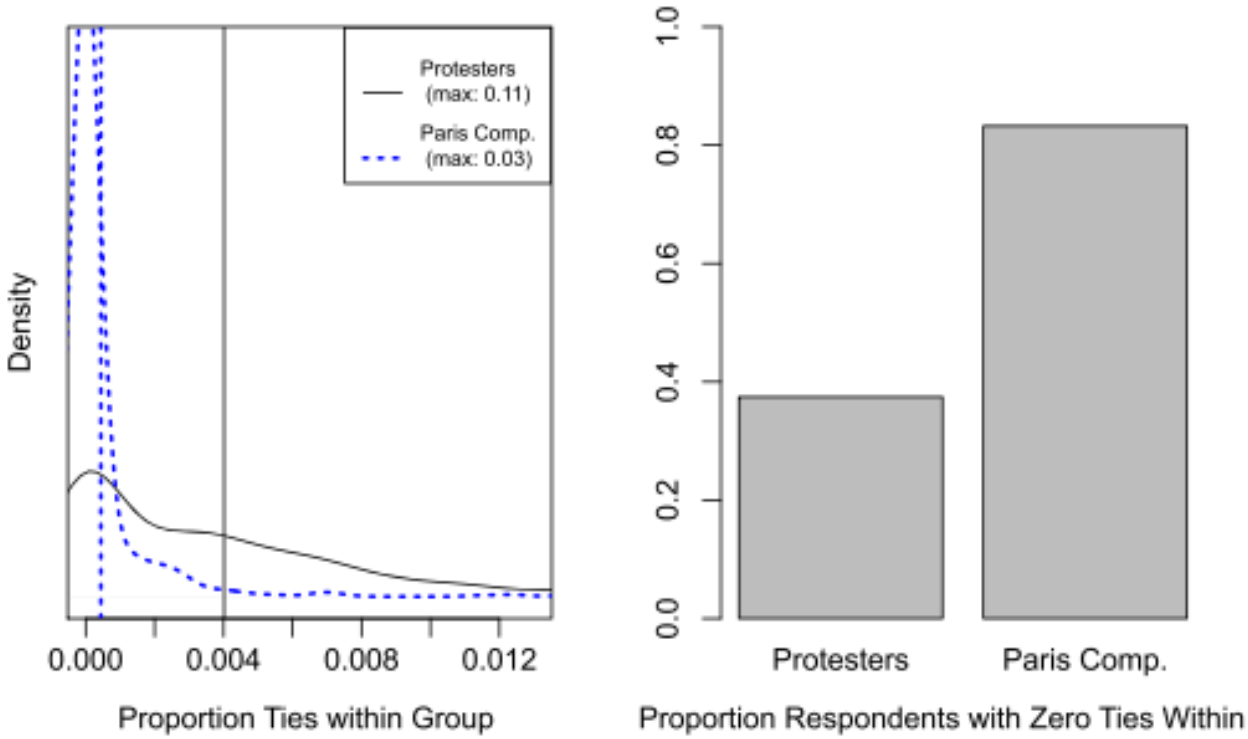
# Social Networks and Protest Participation: Evidence from 130 Million Twitter Users (2019)

**Types of Data**

- Tweets with one of 7 hashtags (API endpoint unclear) to identify users tweeting from protest and from nearby
- Followers and followers' followers for each of these 1,528 accounts.

# Main Result

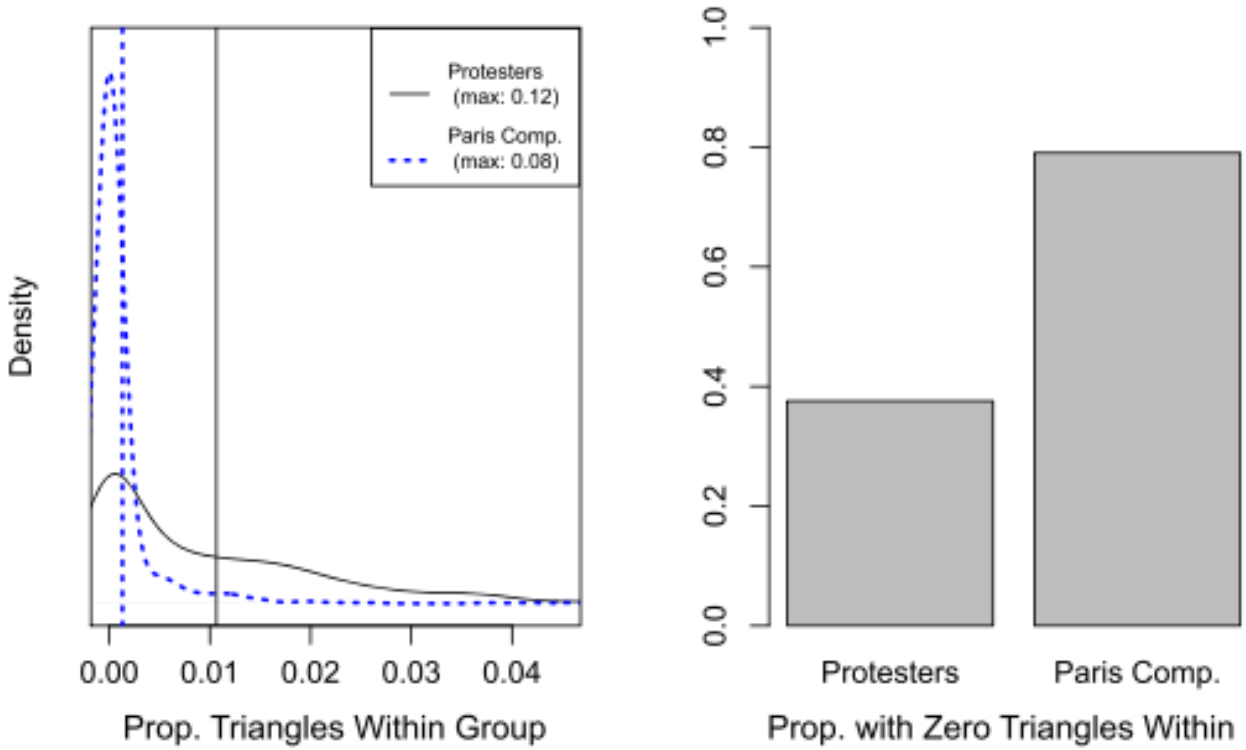
**FIGURE 3 Distribution of the Proportion of Ties within Each Group, Zoomed In**




*Note:* The left panel shows the distribution of the proportion of protesters' ties to other protesters and the distribution of the proportion of users' ties in the Paris comparison set to others in the Paris comparison set. Vertical lines show the distributions' means. The right panel shows the proportion of each sample with zero ties to others within the sample. Strong support is indicated for Hypothesis 1.

# Main Result

**FIGURE 5 Distributions of Individuals' Triangles That Entail Another in the Relevant Set**




*Note*: The left panel shows the distribution of the proportion of protesters' triangles that entail at least one other protester, and the distribution of the proportion of individuals in the Paris comparison set's triads that entail at least one other in that set. Vertical lines show the distributions' means. The right panel shows the proportion of respondents with no triangles that entail anyone else in the relevant set. Strong support is indicated for Hypothesis 3.

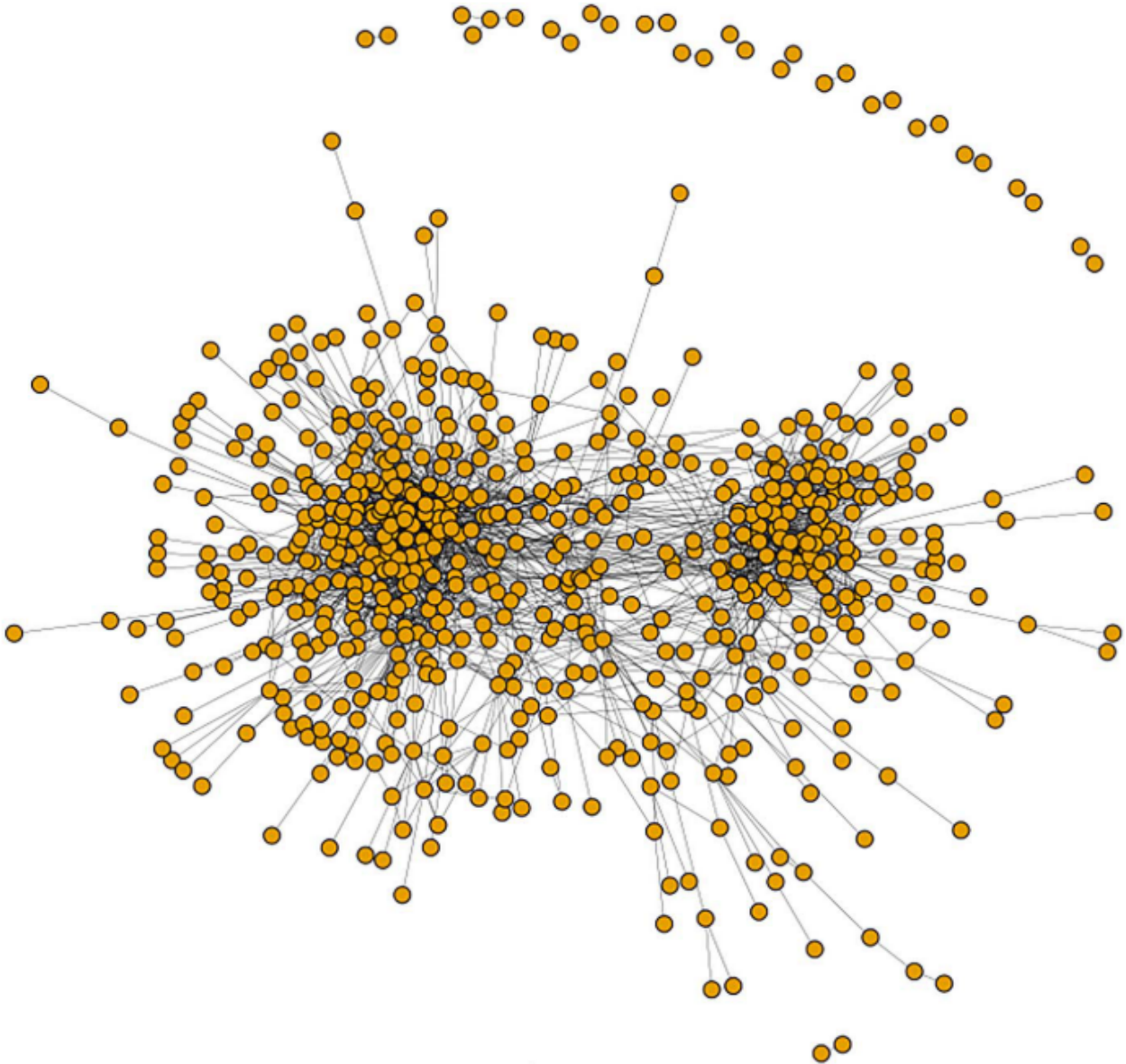
# Evaluating the Mobilization Effect of Online Political Network Structures: A Comparison between the Black Lives Matter Network and Ideal Type Network Configurations (2021)

**Types of Data**

- Follower list of @BLMSacramento: 1215 followers.
- Follower list of these 1215.
- Edge = mutual following.

# Main Result

**Figure 5. The BLM network**



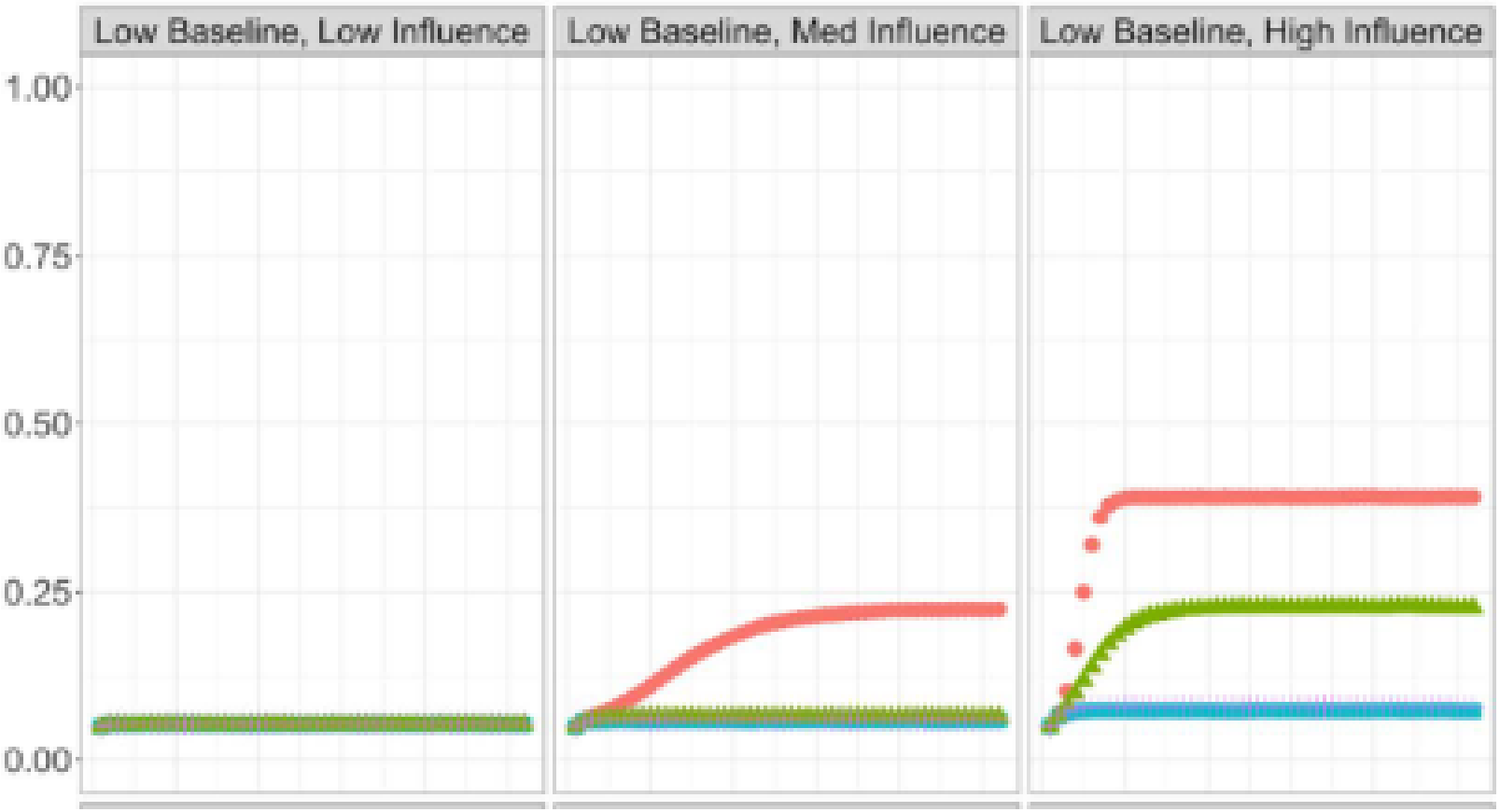
# Main Result

**Table 2. Comparison of Network Statistics**

| | BLM | Small world | Village | Opinion leader |
|---|---|---|---|---|
| Number of nodes | 655 | 655 | 648 | 655 |
| Number of ties | 2335 | 2620 | 2430 | 2610 |
| Density | 0.011 | 0.012 | 0.012 | 0.012 |
| Transitivity | 0.198 | 0.237 | 0.863 | 0.035 |
| SD degree | 9.195 | 1.49 | 0.706 | 7.562 |
| Diameter | 10 | 6.030 | 13.94 | 5.002 |

# Main Result

**Figure 6. Simulation results for hypothetical scenarios**



# Conclusion

- You go to war with the army you have, not the army you want.
- Use Twitter's v1.1 API as long as possible.
- Learn how to download tweets based on their tweet ID.
- Defining a network is an art.