



TweetViz: Twitter Data Visualization

D. Stojanovski, I. Dimitrovski, G. Madjarov
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University in Skopje



A decorative graphic consisting of a grid of colored squares in shades of green, blue, purple, orange, and pink, arranged in a pattern that tapers to the right.

Outline

- Motivation
- Data collection
- User-orientated visualizations
- Keyword-orientated visualizations
- Conclusion



- Social network
- Micro-blogging service
- Share ideas and information
- 271 million monthly active users (July, 2014)
 - 645 million active registered users
- Institutions, companies, politicians, celebrities ...
- 500 million tweets are sent per day



- Tweet – maximum 140 characters long
- Retweet – share a tweet from another user
- Followers – follow another user’s activity
- Mention – including a user in tweet
- Reply – engaging in interaction
- Hashtag – tweet topic






Motivation

- Widely popular
- Generates enormous amount of data every day
 - 143199 Tweets per second (in Japan after watching Castle in the Sky)
- Analyze Twitter activity
- Obtain knowledge from the data



- Web tool for visualizing and analyzing Twitter data
- Modules
 - Data collection
 - User-orientated visualizations
 - Keyword-orientated visualizations
- Google Charts, d3 (Data-Driven Documents)
- Simple user interface

194.149.136.27/Paper/TweetViz/public/

Visualize everything that comes to mind

bitcoin Visualize

Word cloud | Timeline | Scatter | **Tweets**

User	Date	Rts	Fvs	Tweet
ge282	2014-10-05 23:40	0	0	DealBook: Price of Bitcoin Tumbles http://t.co/uV1kibDDwk #Tech #Innovation https://t.co/qPhzqDHSm
BitcoinGood	2014-10-05 23:40	0	0	[FORHIRE] Translations from English to Danish http://t.co/AG93iS01uK #BTC
Bitcoinallday	2014-10-05 23:40	0	0	Bitcoin Looking More and More Like a Real Currency http://t.co/8RCbfvy0n9
528crypto	2014-10-05 23:40	0	0	Bitcoin News in Review: Bitcoin Price, Bill Gates, JPMorgan, and More http://t.co/9Qzg2ZI2tt
HourlyBitcoin	2014-10-05 23:40	0	1	Visit the site above and collect your bitcoins - B http://t.co/chuRwBgmmm B EARN #Bitcoin HERE
cryptoticks	2014-10-05 23:40	0	0	Bitfinex Prices LAST: \$302.76 BID: \$302.71 ASK: \$302.76 VOL: 76916.03 BTC #bitcoin #bitfinex
nicotaku88	2014-10-05 23:40	0	0	Free BitCoin 0.30BTC daily For Free! http://t.co/Jb12yqup51 <-!! #Bitcoin #FreeBTC [777]
CryptoCoinRDT	2014-10-05 23:39	0	0	Can someone explain Bitcoin's price plummet....? http://t.co/eVOvBbjor7
Getrealbitcoins	2014-10-05 23:39	0	0	Bitcoin News in Review: Bitcoin Price, Bill Gates, JPMorgan, and More http://t.co/Ed9859LUwu
KyleDeauna1992	2014-10-05 23:39	0	0	"Price of Bitcoin Tumbles" by SYDNEY EMBER via NYT http://t.co/pvZfjheCg2 http://t.co/3DYyiNGJ0s

The system is available at: <http://194.149.136.27/Paper/TweetViz/public/>



Data collection

- Twitter API
- Twitter user data
- Tweets with keyword or hashtag – Twitter Search



Example Request

GET

https://api.twitter.com/1.1/search/tweets.json?q=%23freebandnames&since_id=24012619984051000&max_id=250126199840518145&result_type=mixed&count=4

Example Result

```
{
  "statuses": [
    {
      "coordinates": null,
      "favorited": false,
      "truncated": false,
      "created_at": "Mon Sep 24 03:35:21 +0000 2012",
      "id_str": "250075927172759552",
      "entities": {
        "urls": [

        ],
        "hashtags": [
          {
            "text": "freebandnames",
            "indices": [
              20,
              34
            ]
          }
        ]
      }
    }
  ]
}
```



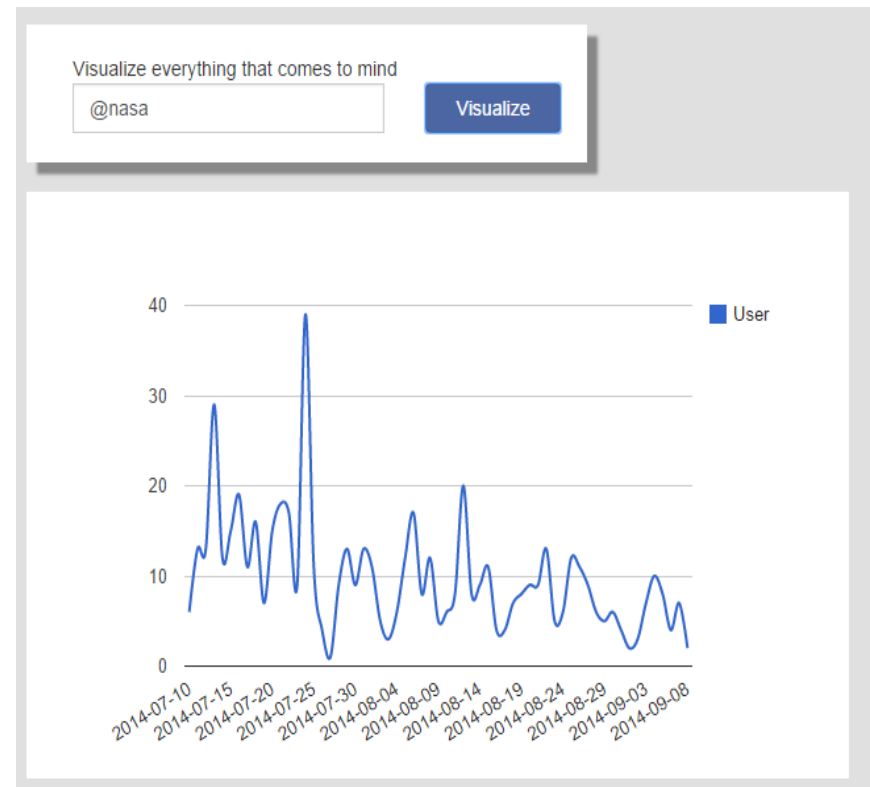
Data collection

- Tweets collected from group of users
- LDA model built on text from tweets
 - 1.4 million tweets



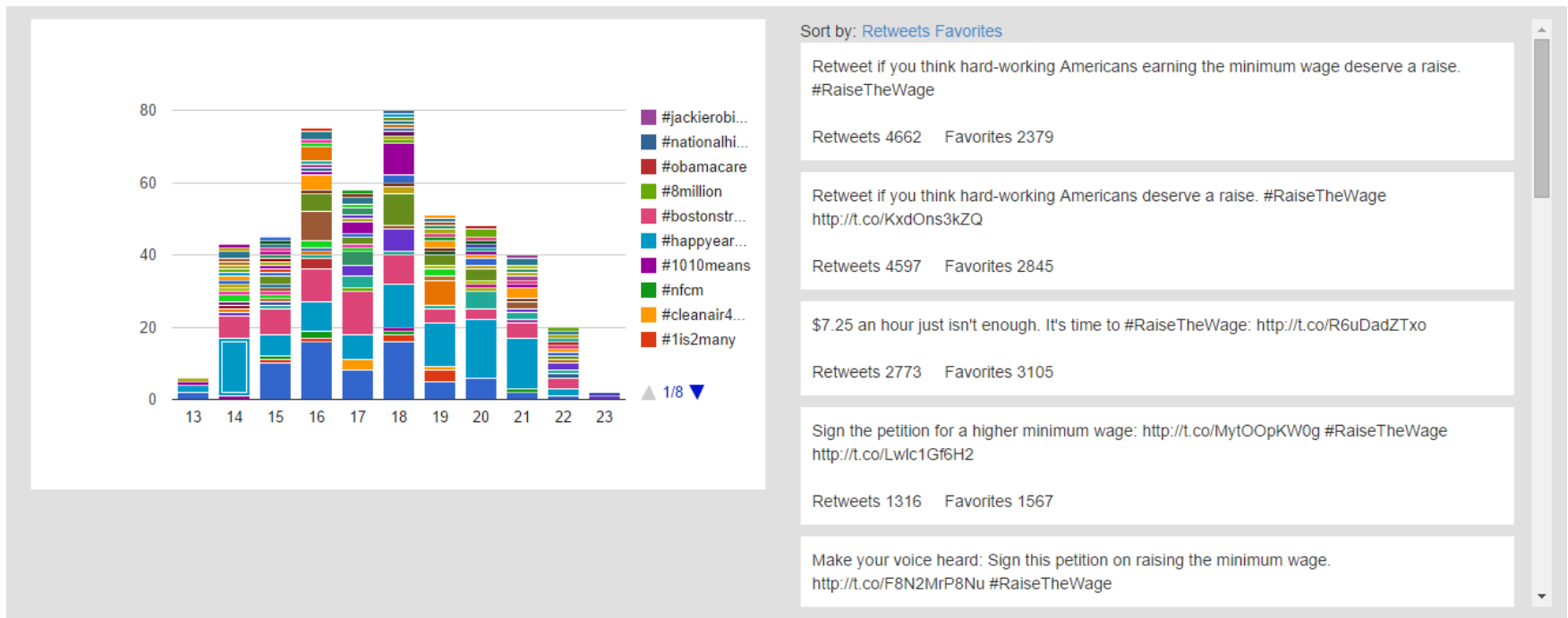
User-orientated visualizations

- Analyze user activity and interests
- General and hashtag related activity visualizations
- Interactive visualizations
- Additional information and related tweets



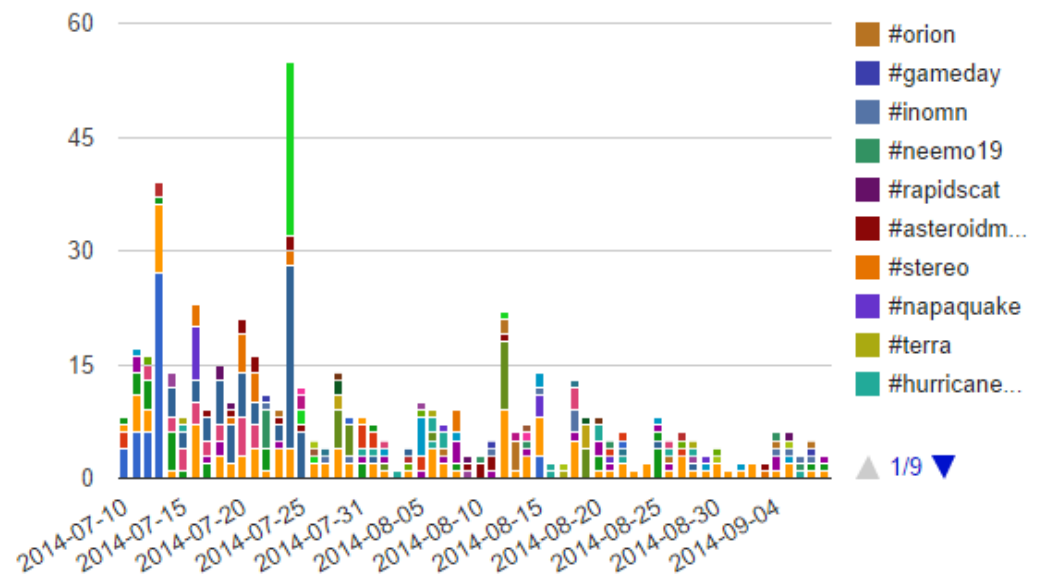
User-orientated visualizations

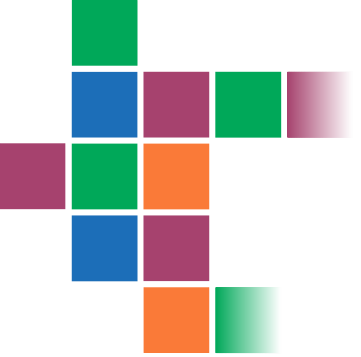
- User-hashtag at different times of day
- Shows tweets on click



User-orientated visualizations

- Stacked column chart showing hashtag distribution in a time interval





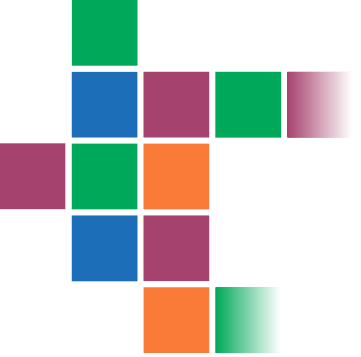
Keyword-orientated visualizations

- Tweets containing some keyword or hashtag
- Twitter Search
- Simple way of analyzing activity surrounding a keyword
- Visualizations similar to the user-orientated ones

Keyword-orientated visualizations

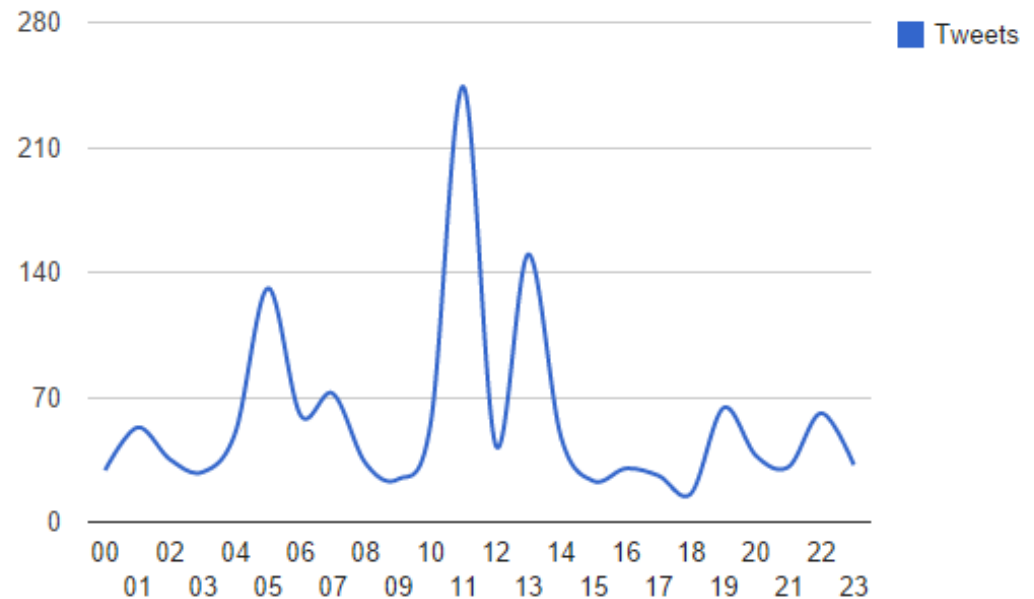
- Word cloud
 - Stopwords removal
 - More frequent – larger dimension
- Shows context surrounding a keyword





Keyword-orientated visualizations

- Activity around a specific keyword in different times of day
- Track changes in interest about the term





Topic distribution visualization

- Topic distribution in a set of tweets
- Track changes in user activity and interests
- Text preprocessing (stop words removal and stemming)
- Latent Dirichlet Allocation
 - Tweet = mix of topics
 - Topic = mix of words
 - Accompanied by probabilities



Latent Dirichlet Allocation

Example:

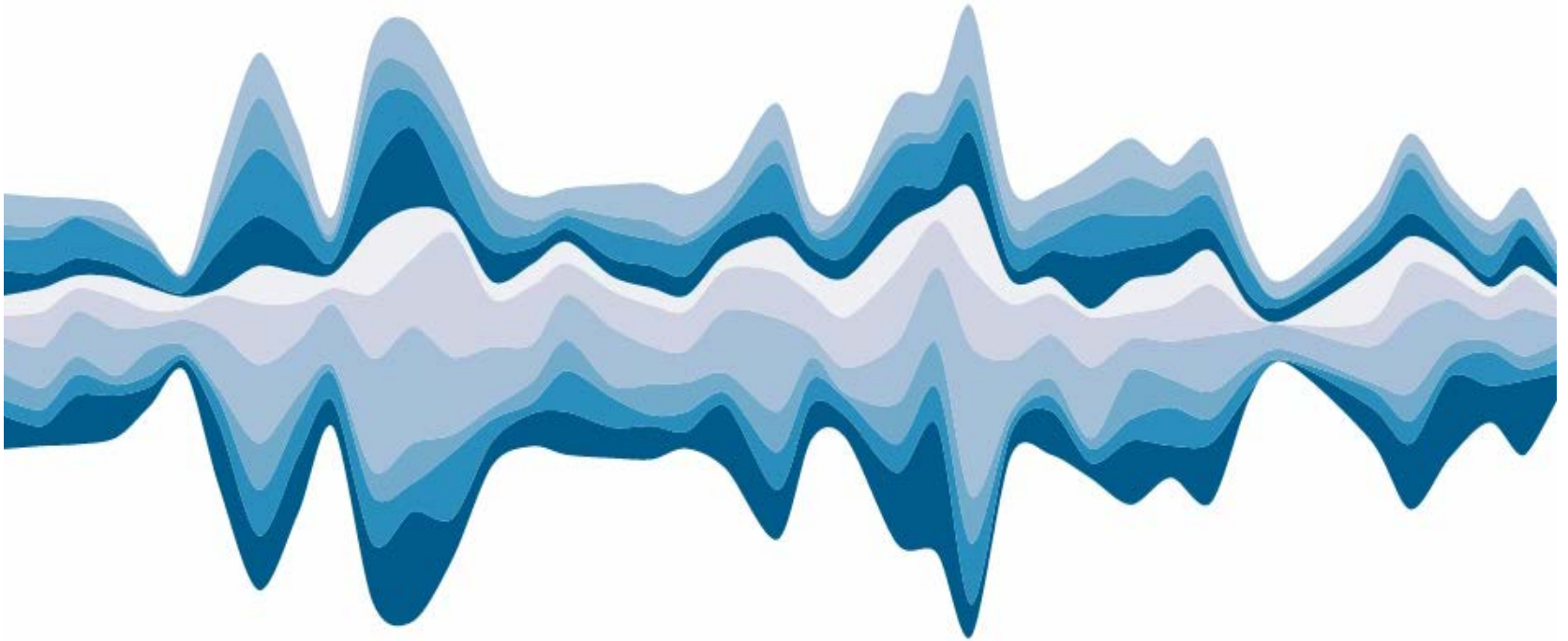
topic #1: $0.026 * \text{relay} + 0.026 * \text{athletics} + 0.025 * \text{metres} + 0.023 * \text{freestyle} \dots$

tweet #1: $0.122 * \text{topic\#1} + 0 * \text{topic\#2} + 0.675 * \text{topic\#3} \dots$

Each topic is represented as a probability distribution over a number of words



Streamgraph





Streamgraph

- The Streamgraph is separated into time slices
- Each time slice is consisted of a set of tweets
- Time slices containing more tweets will have larger y-axis values
 - A layer's height in a certain time interval is dependent on the presence of the related topic in the set of tweets
- Bring topics with greater differences in distribution to the top and bottom of the Streamgraph as oppose to those with lower differences that end up in the middle
 - This adds to a clearer way of presenting the layers and differentiating between them



Streamgraph

- Each layer represents a topic
- 10 layers (topics)
- Broader range of colors
- Layout ordering
- Interactive visualization
- Information about words that make up a topic



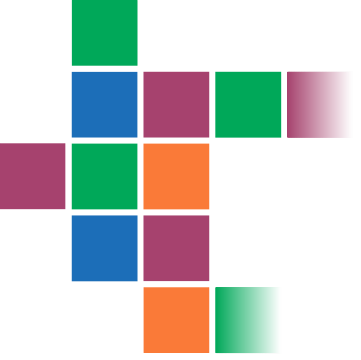
Topic 3

start	0.0381225142864
app	0.0378702298203
mobil	0.0245224027991
futur	0.0221991246815
appl	0.0176901548123
new	0.0158879682935
phone	0.0151669032312



Conclusion

- Improved way of visualizing Twitter activity
- Gain better understanding of Twitter data



Questions?