

#filterbubble

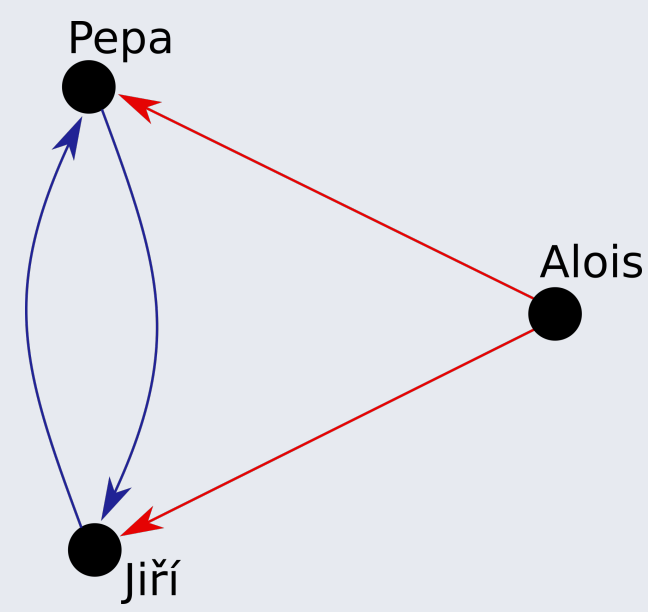
Filter Bubble

Living in one's own information environment.

- ▶ occurs on social networks
- ▶ caused by preferential algorithms
- ▶ first mentioned by Eli Pariser (2011)

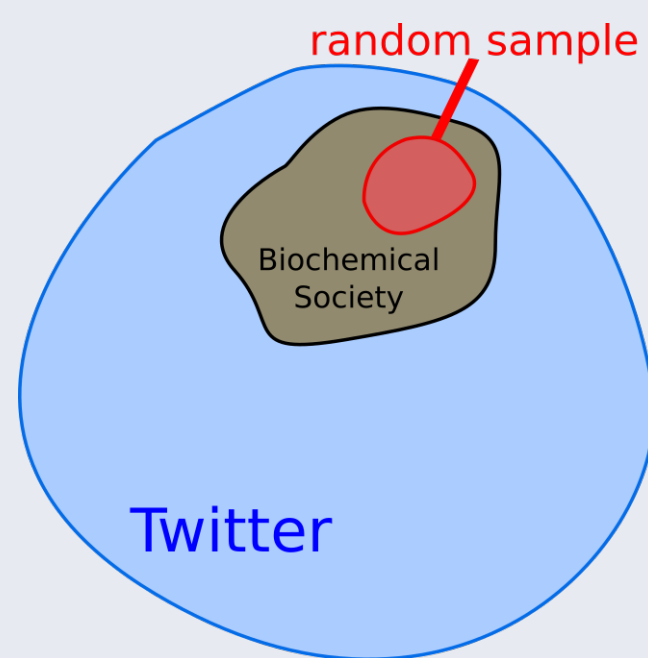
1. Twitter

- ▶ microblogging platform
- ▶ **following, followers** system
- ▶ Twitter API is suitable data source



2. Studied groups selection

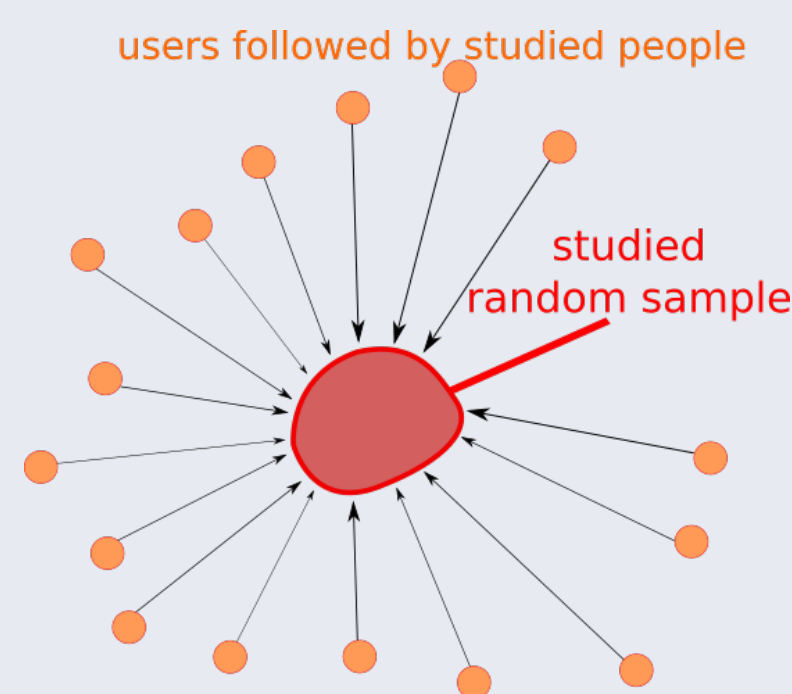
- ▶ random sample from followers of the significant group



Twitter → Biochemical Society → **studied people**

3. Tweets collection

- ▶ analysing content affecting the studied people



- ▶ i. e. content from **followed people**

4. Tweets filtering

- ▶ filter only tweets on given topic
- Keyword "**Trump**":

- ✗ I had fish and chips for lunch.
- ✓ I'm glad Donald **Trump** is the president of the USA.

5. Sentimental analysis

- ▶ measure sentiment of collected tweets
 - ▶ **positive** vs. **negative** tweets
- Donald Trump is a terrible person.*
(0.14)

Donald Trump is a great person.
(0.95)

Motivation

Threats for democracy:

1. content **homogeneity**
2. loss of objectivity
3. **radicalization**

Content homogeneity is a huge **threat for democratic systems**. Our aim is to develop a new method for **detecting and measuring** *Filter Bubble* effects. This would provide us a new way to **study and defend from negatives** of *Filter Bubble*.

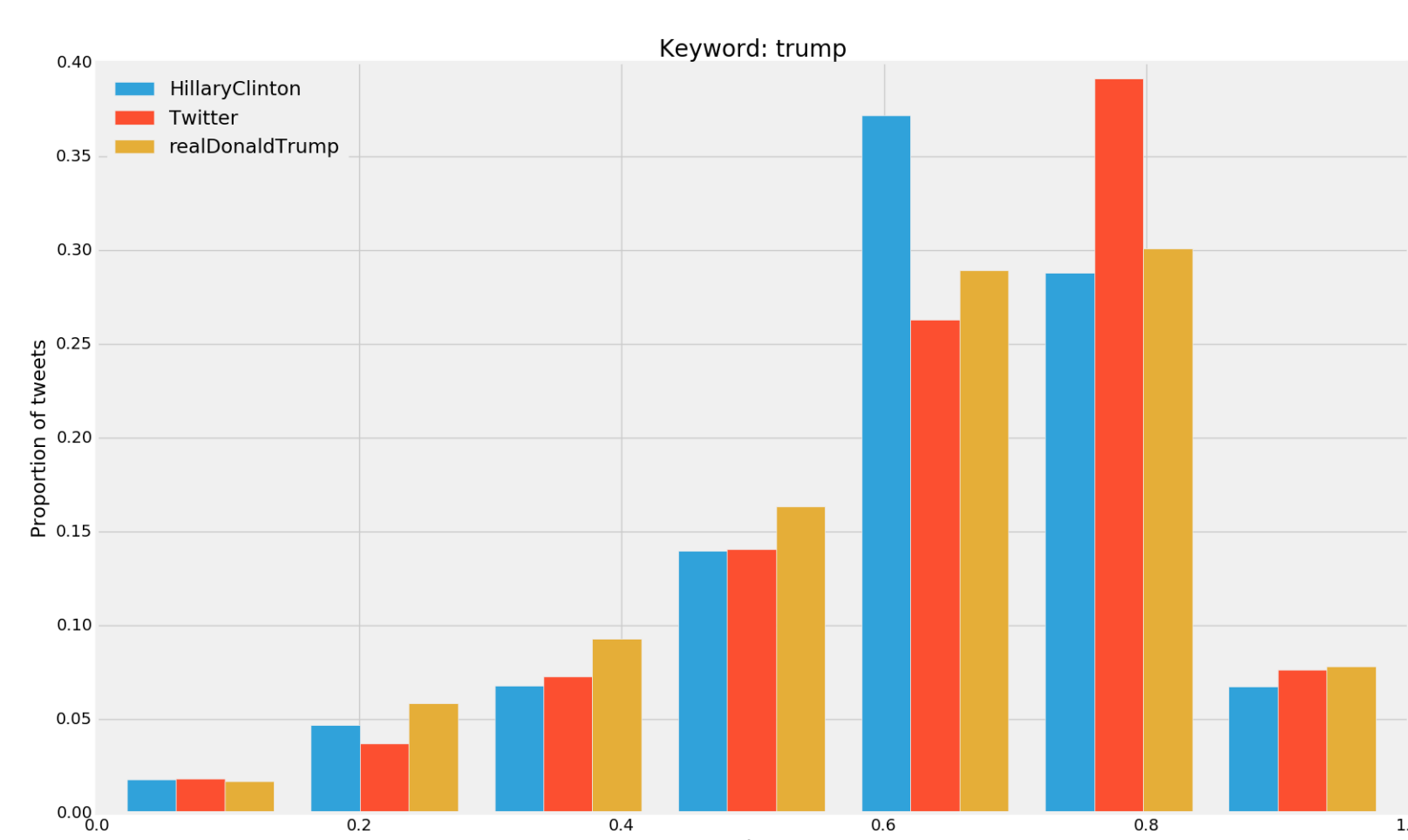
Goals:

- ▶ filter bubble detection
- ▶ filter bubble quantification

Measurements

Studied groups:

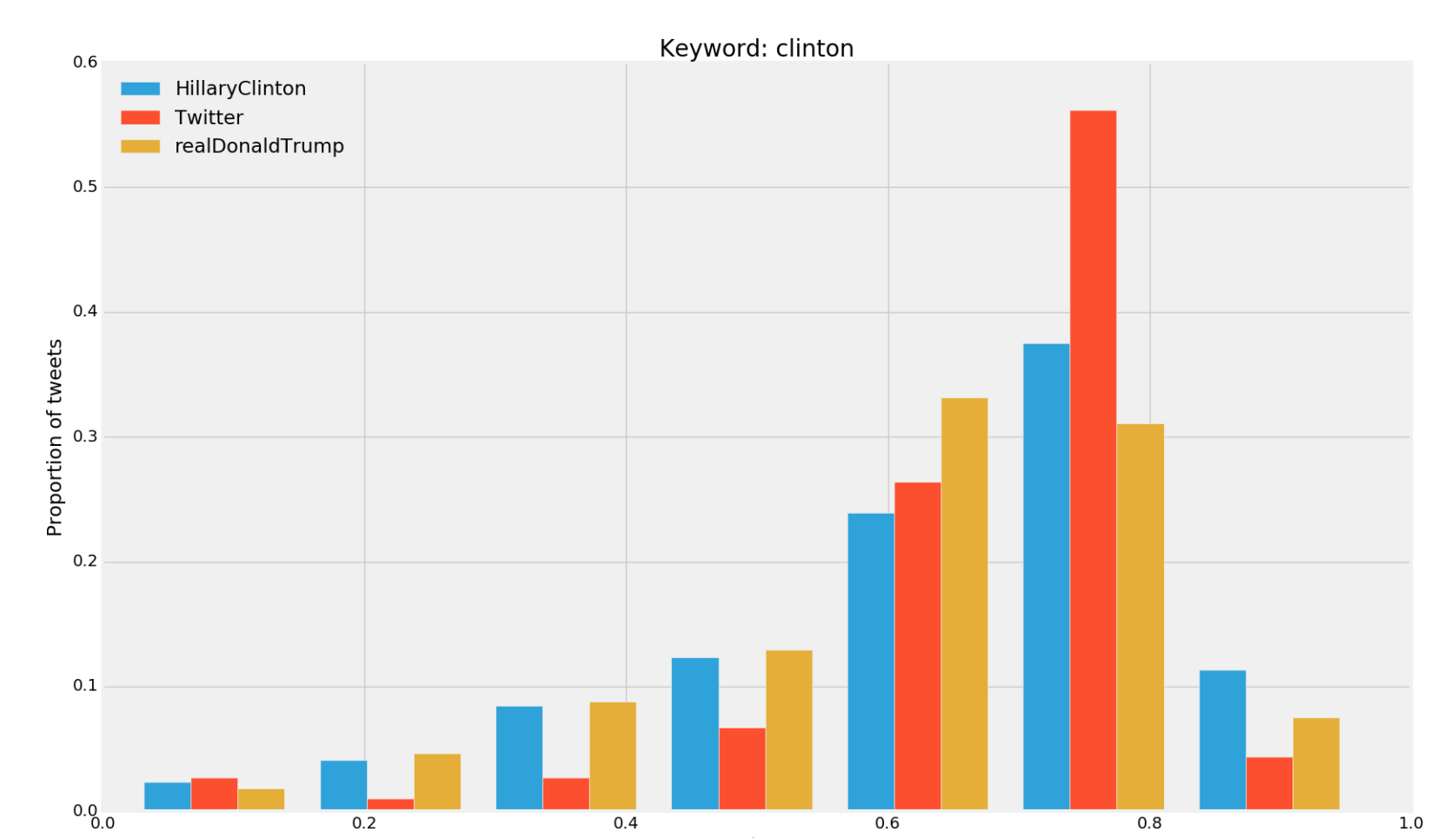
- ▶ Hillary Clinton's supporters
- ▶ Donald Trump's supporters



Normalized sentiment histogram for topic *Donald Trump*.

Topics:

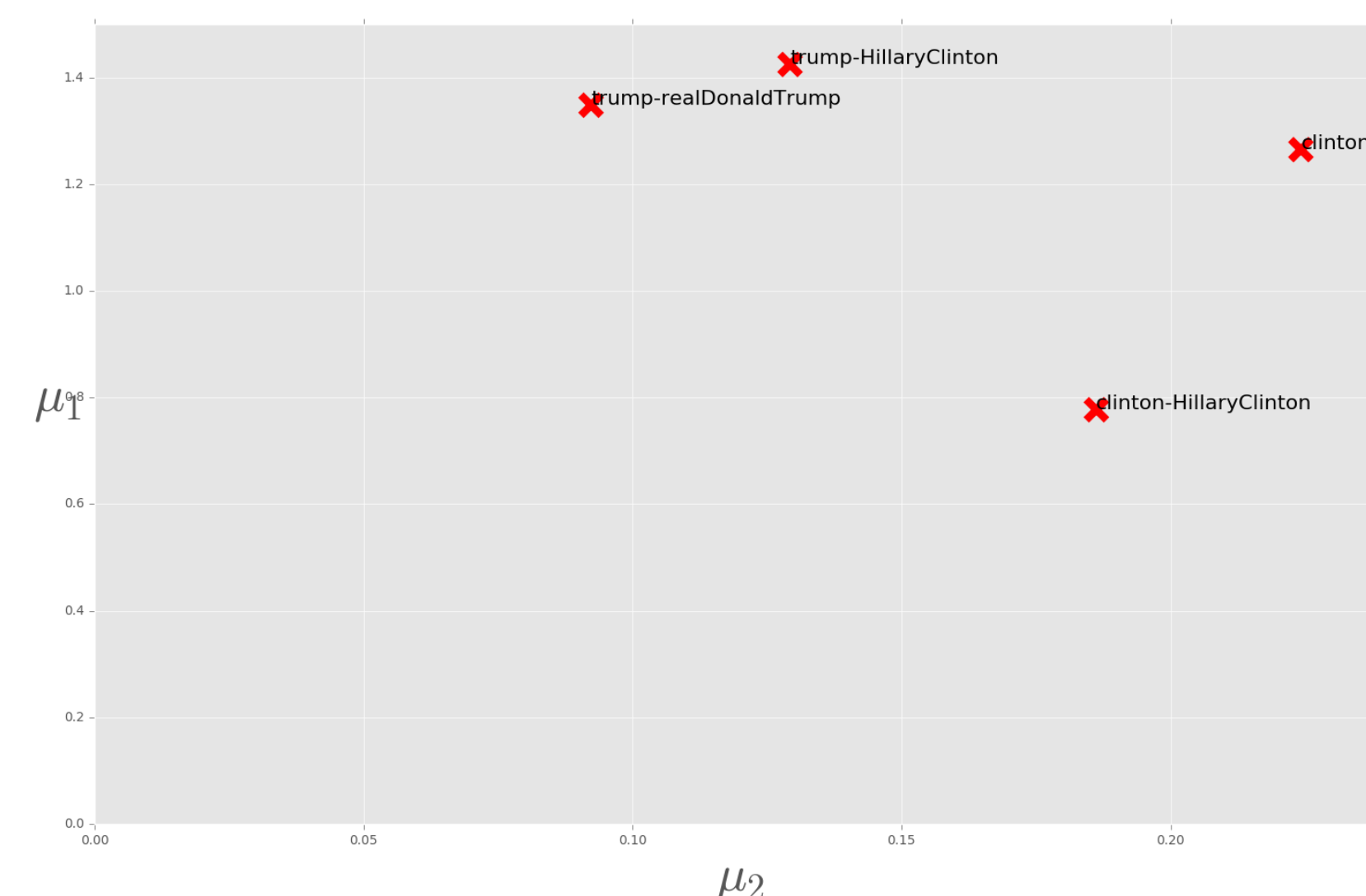
- ▶ Hillary Clinton
- ▶ Donald Trump



Normalized sentiment histogram for topic *Hilary Clinton*.

Proposed measure

We define group affected by *Filter Bubble* as a group that lays in different information environment than randomly sampled group.



Phase diagram of our measurements.

- ▶ μ_1 : distance of proportion of tweets on given topic from random sample
- ▶ $\mu_1(G) = \frac{|p_T - p_G|}{p_t}$

The environment may differ from average in two major ways:

1. the **number of tweets** on given topic (μ_1),
2. **sentiment distribution** (μ_2).

Further from origin means they receive **less balanced** information.

- ▶ μ_2 : distance of sentiment histogram of group G from random sample (T)
- ▶ $\mu_2(G) = \sqrt{\sum_i (S_T^i - S_G^i)^2}$

Conclusion

We have achieved:

1. To the best of our knowledge - **the first measure** of *Filter Bubble* proposed.
2. Large scale measurements → data **noise reduction**.
3. More **straightforward** than traditional research.

Future plans:

1. **Real world** usage.
2. Modify methodology for use **outside the Twitter**.
3. Develop measure that encodes information about **direction of sentiment**