Quantifying Controversy on Social Media

Kiran Garimella*, Gianmarco De Francisci Morales[#], Aristides Gionis*, Michael Mathioudakis*

*Aalto University/HIIT, [#]Qatar Computing Research Institute

{kiran.garimella,aristides.gionis,michael.mathioudakis}@aalto.fi, gdfm@acm.org

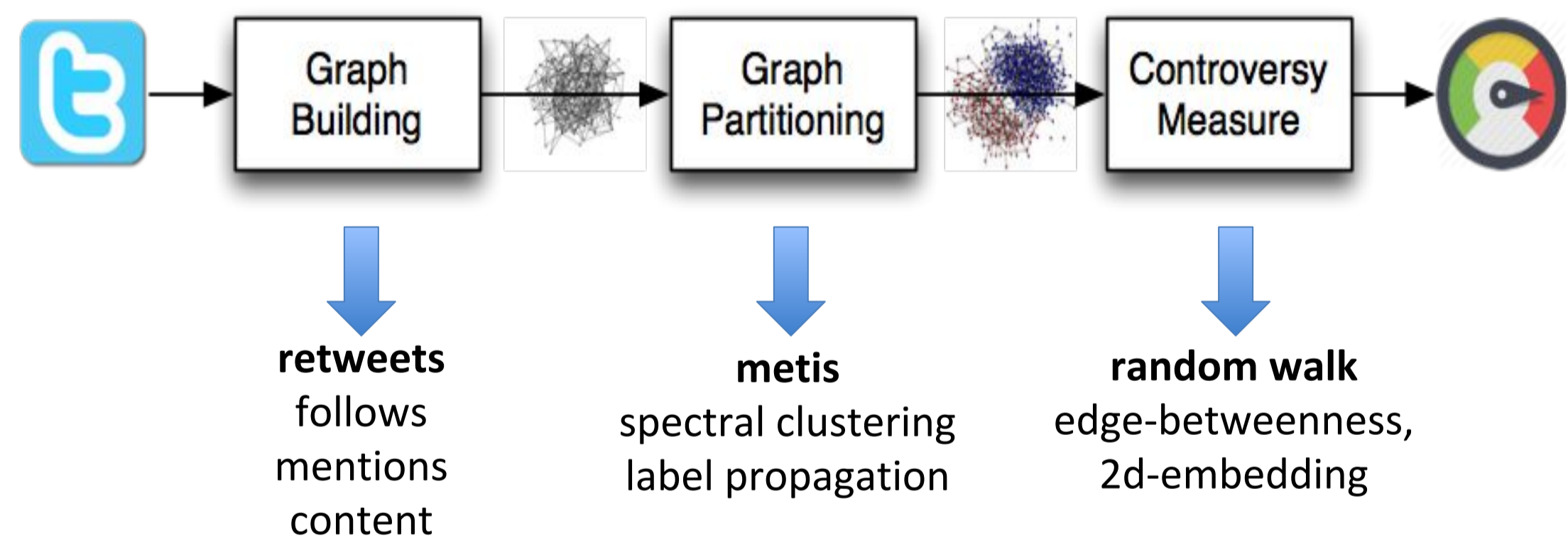A demo of this research will appear at CSCW 2016.
Details available at: https://users.ics.aalto.fi/kiran/controversy

#beefban

#russia_march

#sxsw

#germanwings
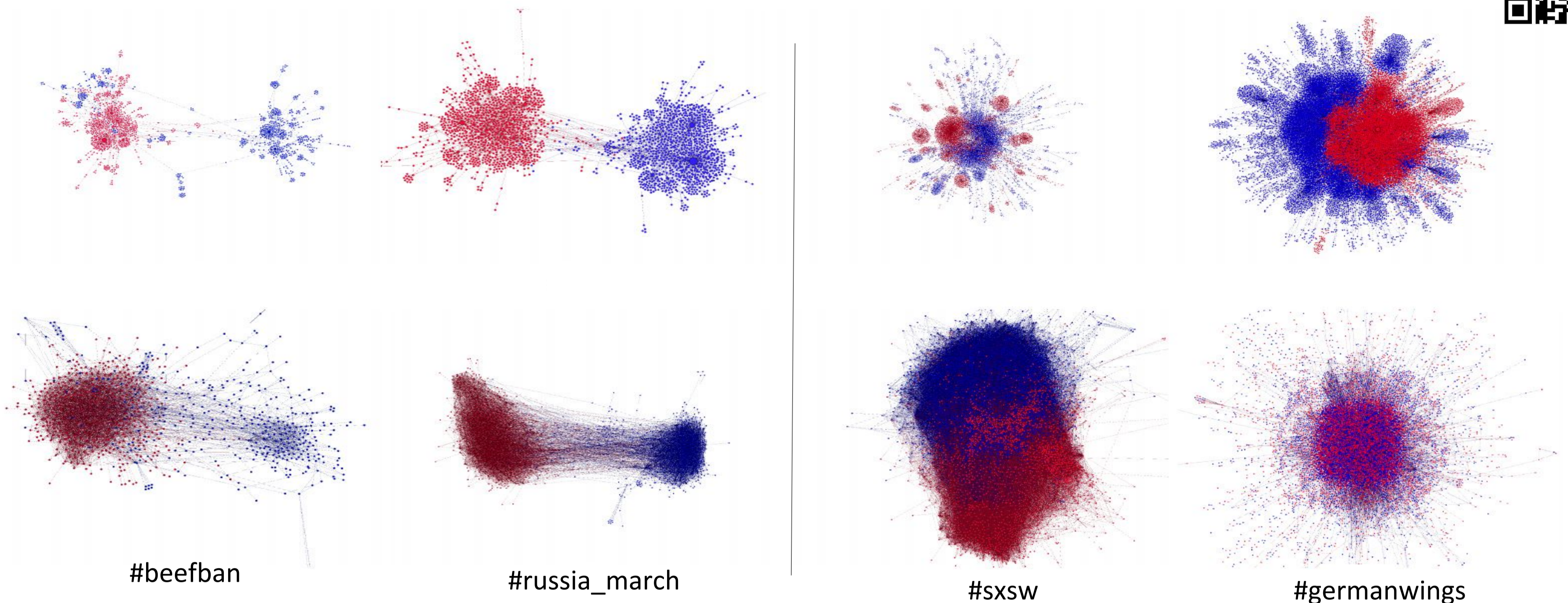
Problem

Detect controversial topics on Social Media

Topic : one hashtag

Controversy : users split into two sides

Motivation

- Discovering controversial topics in the wild, with no prior knowledge, and no domain expertise
- Quantifying the controversy intensity
- Compare the degree of controversy across networks
- Can we generalize on the graph structure?

Results

- Random walk measure works the best
- Sentiment variance works!
- Identifying controversies is not easy
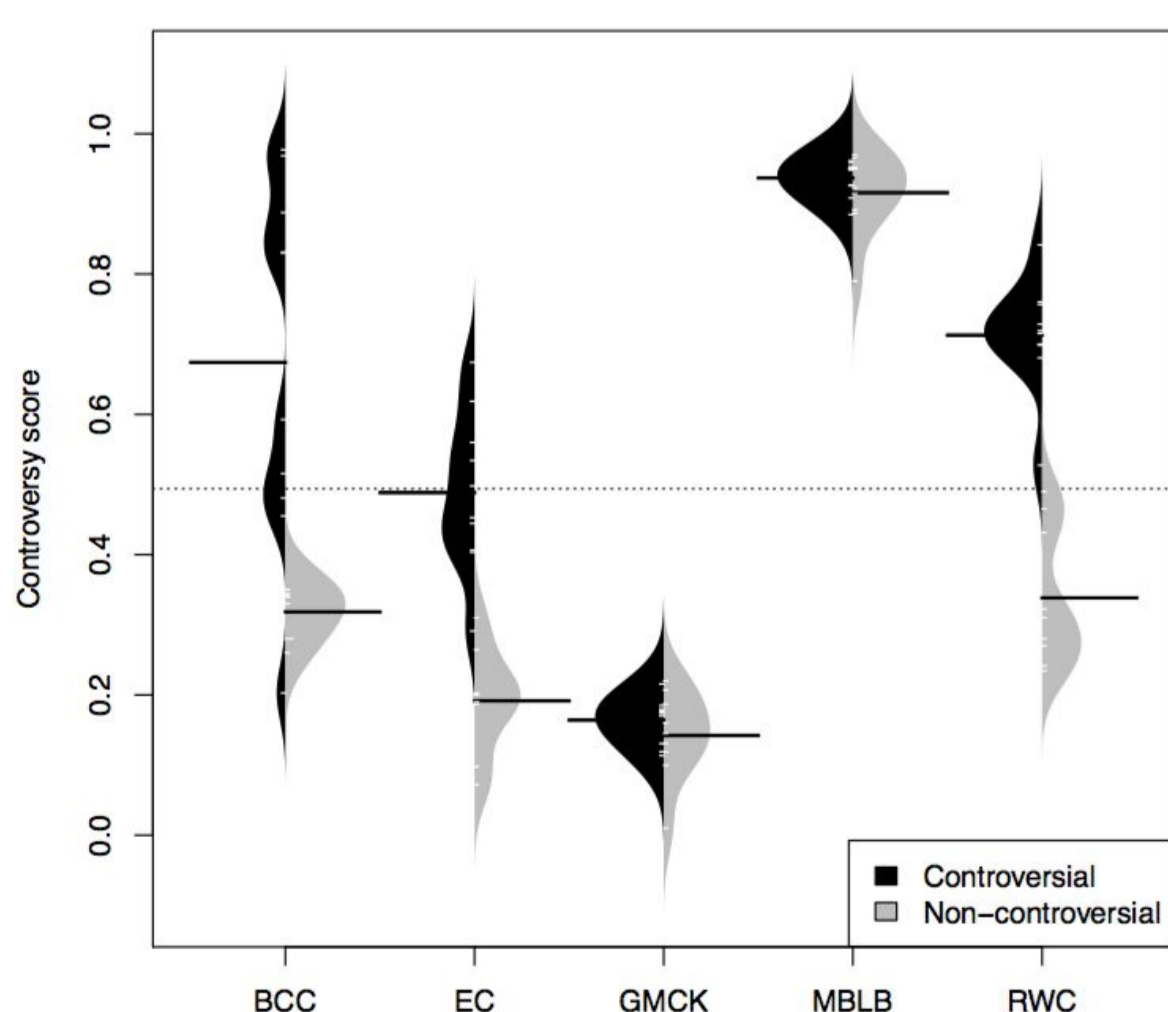
Controversy scores on retweet graphs for 20 controversial and non-controversial datasets

Sentiment variance controversy score

Methodology

retweets
follows
mentions
content

metis
spectral clustering
label propagation
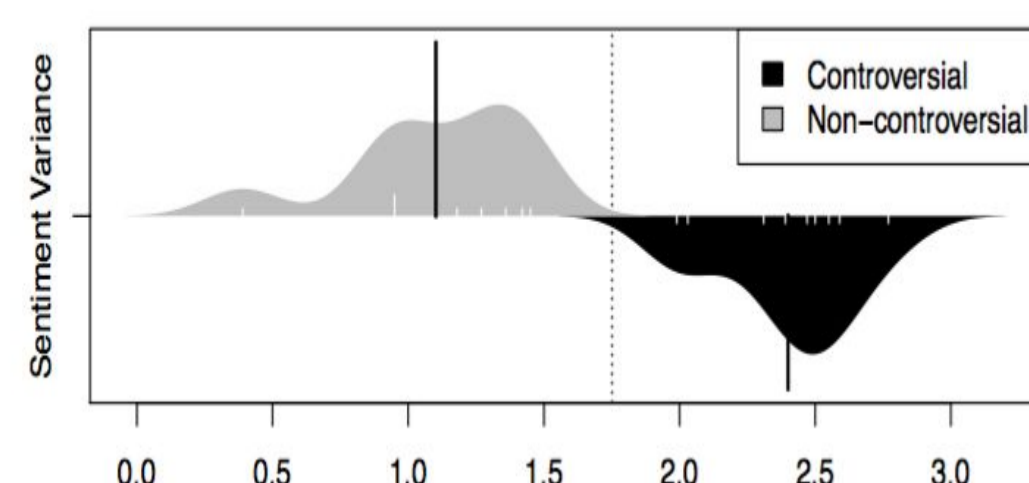
random walk
edge-betweenness,
2d-embedding

Approaches

Random walk : How likely a random user on either side is to be exposed to authoritative content from the opposite side.

Betweenness : How does the distribution of betweenness on the cut differ from other random edges in the graph.

Embedding : How separated are the two sides in a low dimensional embedding.

Sentiment Variance : Variance of content sentiment on both sides.

Limitations

- Not applicable to multi-sided controversies
- Dependance on graph partitioning method
- Controversy measures not to scale

Future Directions

- Track evolution of controversies automatically
- Automatic stance identification
- Balanced news recommendation