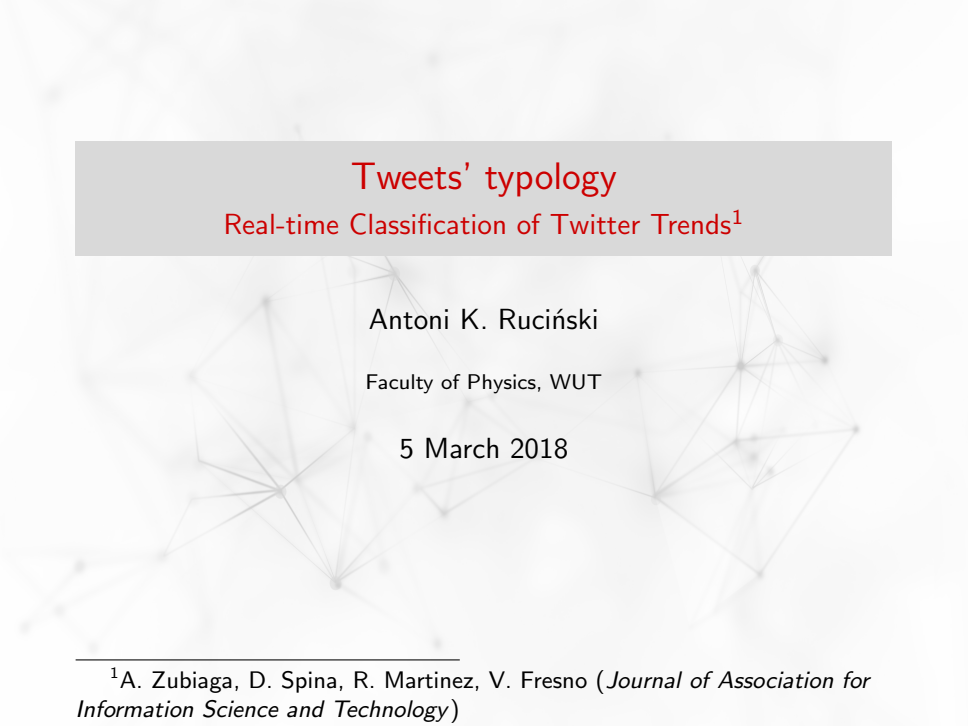


Powered by:

ANRU



The background of the slide features a complex network diagram with numerous nodes and connecting lines, rendered in a light gray color. A prominent gray rectangular banner is positioned at the top of the slide, containing the title and subtitle in red text.

Tweets' typology

Real-time Classification of Twitter Trends¹

Antoni K. Ruciński

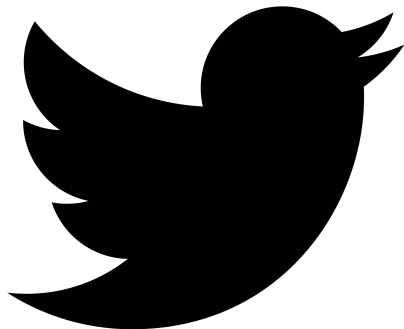
Faculty of Physics, WUT

5 March 2018

¹A. Zubiaga, D. Spina, R. Martinez, V. Fresno (*Journal of Association for Information Science and Technology*)

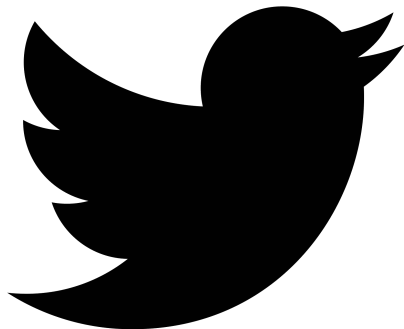
Interaction syntax include:

- **User mentions**
(e.g You should all follow @username, she is always abreast of breaking news and interesting stuff)
- **Replies**
(e.g @username I agree with you)
- **Retweets**
(e.g RT @username: Text of the original tweet.)
- **Hashtags**
(e.g (...) #hollywood #oscars)



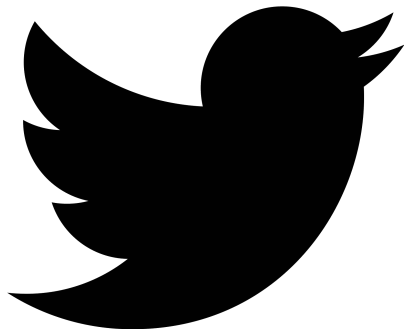
Interaction syntax include:

- **User mentions**
(e.g You should all follow @username, she is always abreast of breaking news and interesting stuff)
- **Replies**
(e.g @username I agree with you)
- **Retweets**
(e.g RT @username: Text of the original tweet.)
- **Hashtags**
(e.g (...) #hollywood #oscars)



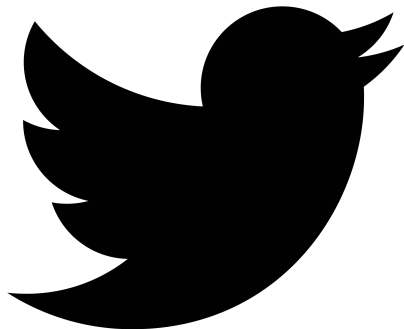
Interaction syntax include:

- **User mentions**
(e.g You should all follow @username, she is always abreast of breaking news and interesting stuff)
- **Replies**
(e.g @username I agree with you)
- **Retweets**
(e.g RT @username: Text of the original tweet.)
- **Hashtags**
(e.g (...) #hollywood #oscars)



Interaction syntax include:

- **User mentions**
(e.g You should all follow @username, she is always abreast of breaking news and interesting stuff)
- **Replies**
(e.g @username I agree with you)
- **Retweets**
(e.g RT @username: Text of the original tweet.)
- **Hashtags**
(e.g (...) #hollywood #oscars)



Trending Topics

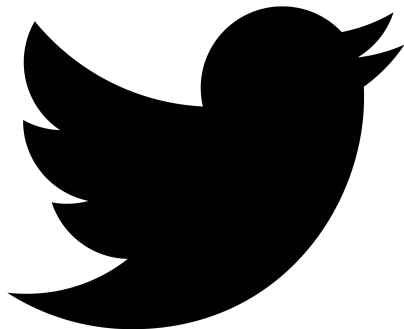
According to Twitter: *topics that are immediately popular, rather than topics that have been popular for a while or on a daily basis*

TABLE 1. Example of a trending topic, and some associated tweets.

Trending topic: Interpol	
User	Content of the tweet
@u1	Interpol issues arrests warrants for Gaddafi & 15 senior Libyan officials. #Libya
@u2	RT @u1: Interpol issues arrests warrants for Gaddafi & 15 senior Libyan officials. #Libya
@u3	@u1—so Interpol cannot act until he & family leave Libya, is that right? Assuming he is toppled?
@u4	RT @u2: RT @u1: Interpol issues arrests warrants for Gaddafi & 15 senior Libyan officials. #Libya
@u5	Interpol has issued international alert for Muammar Gaddafi & 15 other family members & close associates—Telegraph http://bit.ly/h9GwYI

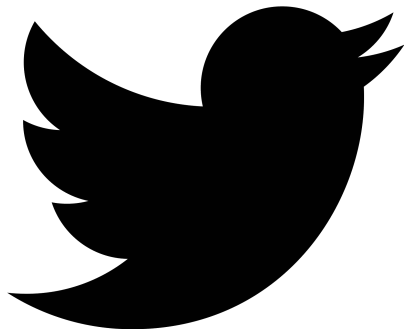
Types of triggers:

- **NEWS**
(produced by newsworthy event that major news outlets had reported by the time the trend popped up or will report soon after its appearance on Twitter)
- **ONGOING EVENTS**
(introduced by a community of users tweeting about an event as it unfolds, e.g. soccer game, musical festival)
- **MEMES**
(viral ideas initiated by an organisation or individual that were popular enough to be spread widely; interesting or amusing)
- **COMMEMORATIVES**
(congratulations on a birthday, celebrations)



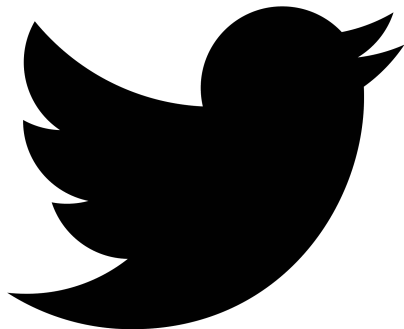
Types of triggers:

- **NEWS**
(produced by newsworthy event that major news outlets had reported by the time the trend popped up or will report soon after its appearance on Twitter)
- **ONGOING EVENTS**
(introduced by a community of users tweeting about an event as it unfolds, e.g. soccer game, musical festival)
- **MEMES**
(viral ideas initiated by an organisation or individual that were popular enough to be spread widely; interesting or amusing)
- **COMMEMORATIVES**
(congratulations on birthday, celebrations)



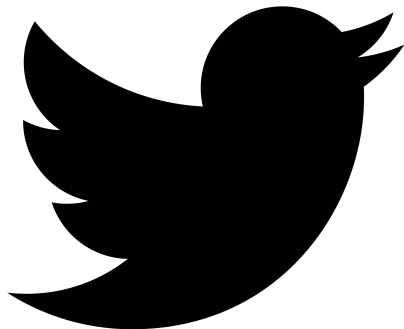
Types of triggers:

- **NEWS**
(produced by newsworthy event that major news outlets had reported by the time the trend popped up or will report soon after its appearance on Twitter)
- **ONGOING EVENTS**
(introduced by a community of users tweeting about an event as it unfolds, e.g. soccer game, musical festival)
- **MEMES**
(viral ideas initiated by an organisation or individual that were popular enough to be spread widely; interesting or amusing)
- **COMMEMORATIVES**
(congratulations on birthday, celebrations)



Types of triggers:

- **NEWS**
(produced by newsworthy event that major news outlets had reported by the time the trend popped up or will report soon after its appearance on Twitter)
- **ONGOING EVENTS**
(introduced by a community of users tweeting about an event as it unfolds, e.g. soccer game, musical festival)
- **MEMES**
(viral ideas initiated by an organisation or individual that were popular enough to be spread widely; interesting or amusing)
- **COMMEMORATIVES**
(congratulations on a birthday, celebrations)



Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hastags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hastags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hashtags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

Averages means

- Retweets (depth)
- Retweets (ratio)
- # Hastags
- Length
- # Exclamations
- # Questions
- # Links
- Topic repetitions
- Replies
- Spread velocity

$$AM(f)_t = \frac{1}{|T|} \sum_{i=1}^{|T|} f_i \quad (1)$$

$AM(f)$ – arithmetic mean of feature f , $|T|$ – # of tweets in trending topic, f_i – value of feature for the tweet i ;

$$AM(sv)_t = \frac{|T|}{\Delta t} \quad (2)$$

Δt – # seconds from the first to the last tweet on trending topic.

Diversity values

- **User diversity**
 - Retweeted user diversity
 - Hashtag diversity
- Language diversity
- Vocabulary diversity

Diversity values

- User diversity
- Retweeted user diversity
- Hashtag diversity
- Language diversity
- Vocabulary diversity

Diversity values

- User diversity
- Retweeted user diversity
- Hashtag diversity
- Language diversity
- Vocabulary diversity

Diversity values

- User diversity
- Retweeted user diversity
- Hashtag diversity
- Language diversity
- Vocabulary diversity

Diversity values

- User diversity
- Retweeted user diversity
- Hashtag diversity
- Language diversity
- Vocabulary diversity

Diversity values

- User diversity
- Retweeted user diversity
- Hashtag diversity
- Language diversity
- Vocabulary diversity

$$H'(f)_t = - \sum_{j=1}^s (p_{jt} \ln p_{jt}) \quad (3)$$

$$p_{jt} = \frac{n_{jt}}{N} \quad (4)$$

$H'(f)_t$ – Shano's index of feature f for the trending topic t , n_{jt} – population of the value j , S – # different values, N – total population, p_{jt} – observed probability if the value j

Features of Trending Topics

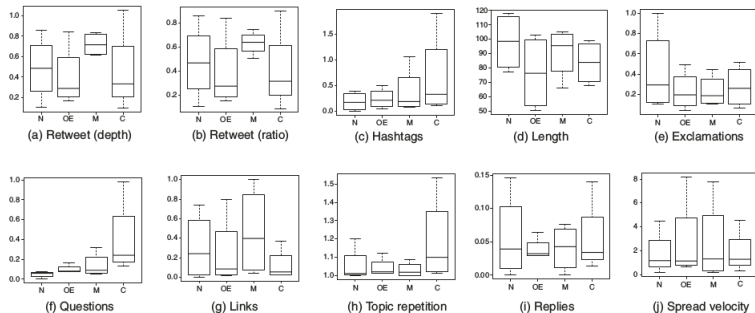


FIG. 1. Box/whisker plots for features relying on average values (N: News; OE: Ongoing Events; M: Memes; C: Commemoratives).

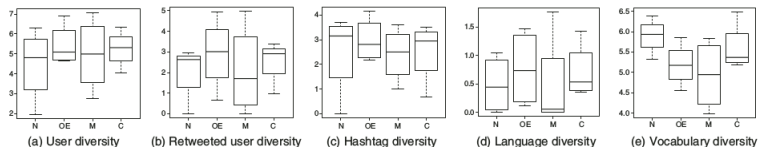


FIG. 2. Box/whisker plots for features relying on diversity values (N: News; OE: Ongoing Events; M: Memes; C: Commemoratives).

Analysis of text

- **COMMEMORATIVE**

congratulation words: *happy, birthday* as well as time related words: *day, anos (=year)*

- **ONGOING EVENTS**

words refer to the present: *live, tonight*; action: *watching, watch*; media: *tv*

- **NEWS**

terminology is not specific

- **MEMES**

terminology is not specific

Analysis of text

■ **COMMEMORATIVE**

congratulation words: *happy, birthday* as well as time related words: *day, anos (=year)*

■ **ONGOING EVENTS**

words refer to the present: *live, tonight*; action: *watching, watch*; media: *tv*

■ **NEWS**

terminology is not specific

■ **MEMES**

terminology is not specific

Analysis of text

- **COMMEMORATIVE**

congratulation words: *happy, birthday* as well as time related words: *day, anos (=year)*

- **ONGOING EVENTS**

words refer to the present: *live, tonight*; action: *watching, watch*; media: *tv*

- **NEWS**

terminology is not specific

- **MEMES**

terminology is not specific

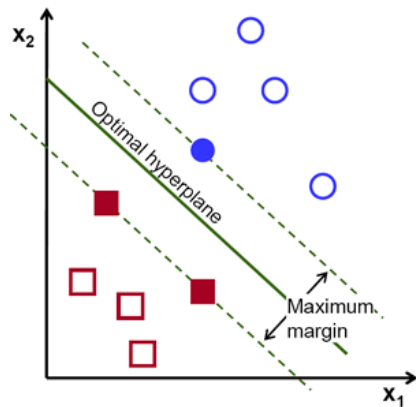
one-against-all

$$C_i = \arg_{\max} m_i \quad (5)$$

m_i – margin outputted by the classifier i

Two different representations:

- Twitter features: 15 dimensions
- bag-of-words: # occurrences of each term in a trend



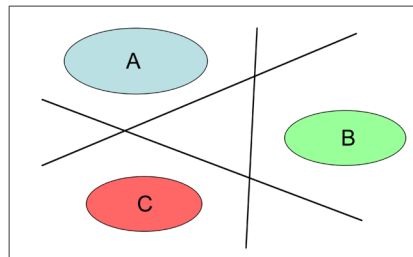
one-against-all

$$C_i = \arg_{\max} m_i \quad (5)$$

m_i – margin outputted by the classifier i

Two different representations:

- Twitter features: 15 dimensions
- bag-of-words: # occurrences of each term in a trend



Examples

$$C_i^* = \mathit{argmax}_{j=1\dots k} S_{tj} \quad (6)$$

k – # classes; t –trending topic; S_{ti} – sum giving by:

$$S_{ti} = \sum_{v=1}^{|V|} m_{vti} \quad (7)$$

$|V|$ – # classifiers; n_{vti} –margin produced by classifier v for the topic t in the class i

TABLE 5. Example of classifier committees on a trending topic. Classifier A labels it as news, and classifier B labels it as meme, while adding both up using classifier committees would label it as an ongoing event.

	N	OE	M	C
Classifier A	1.2	1.1	0.6	0.3
Classifier B	0.5	1.0	1.2	0.3
Classifier committees	1.7	2.1	1.8	0.6

TABLE 6. Accuracy and Cohen's kappa statistic of the trend classification by type of representation.

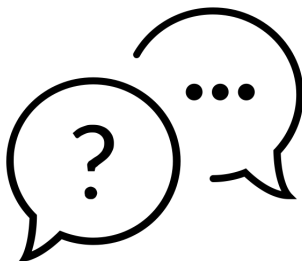
	Accuracy	Kappa
Bag-of-words	0.752	0.530
Twitter features	0.784	0.604
Committees	0.812	0.649

Conclusions

- 4 types of trending topics: *news, ongoing events, memes, commemoratives*;
- 2 different features: *averages* and *diversity values*;
- SVM classifier

Thank you for your attention!

Q&A.



Powered by:

ANRU