



Finding the Story in the TweetStack

Mining Spatio-temporal Clusters for Event Correlation and Visualization

Rahul Potharaju, Andrew Newell, Cristina Nita-Rotaru

Department of Computer Science, Purdue University

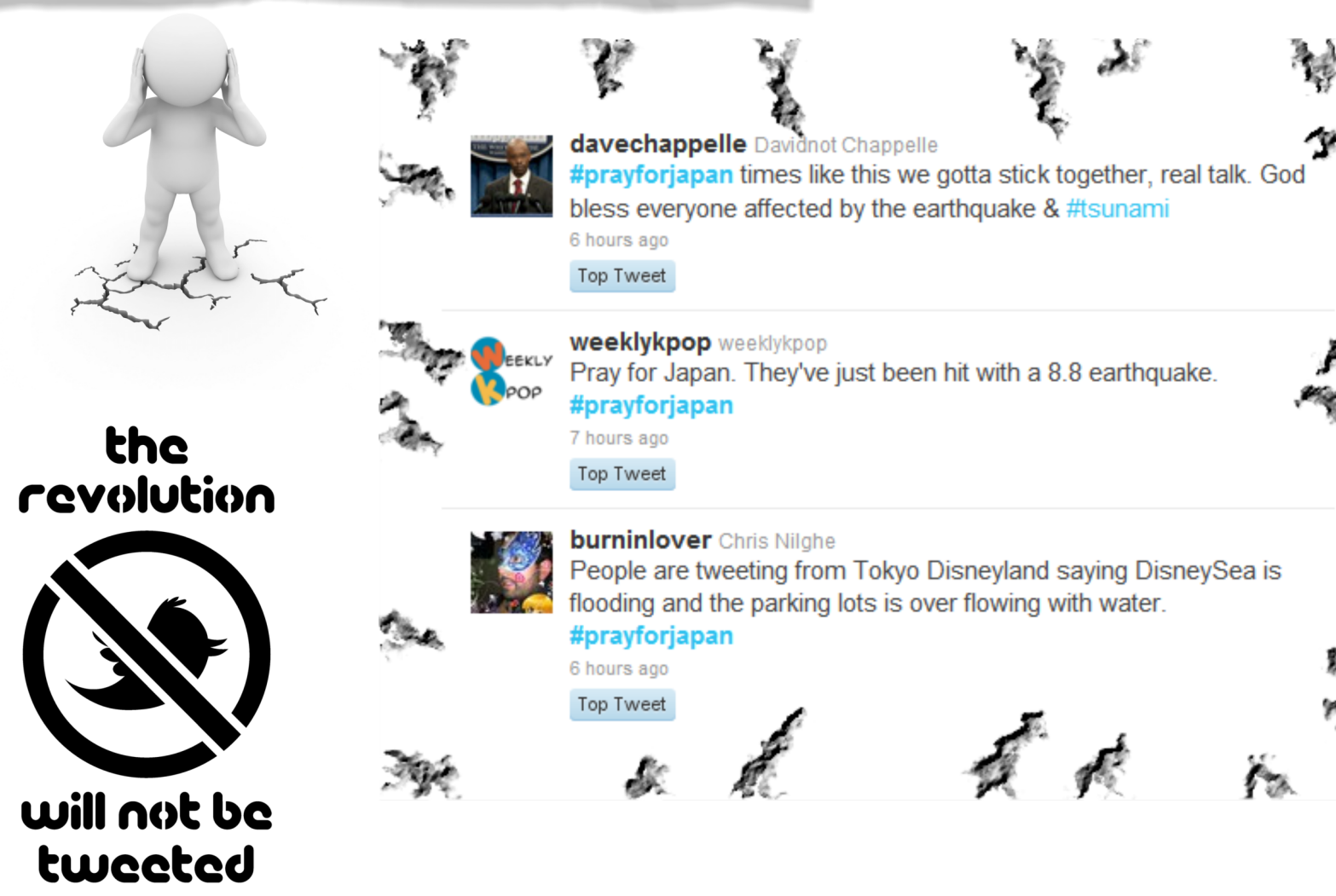


TWEETING AT THE SPEED OF LIGHT?

Early **event detection** carries substantial value in various domains

CAN TINY TWEETS PREDICT FINANCIAL MARKETS?

Twitter may be able to foresee the ebbs and flows of the stock market better than any financial analyst.



Top All-Time

ladygaga	\$10,351,818
justinbieber	\$9,171,231
katyperry	\$8,019,386
rihanna	\$7,431,234
shakira	\$7,318,221
britneyspears	\$7,009,520
KimKardashian	\$6,811,456
barackobama	\$6,432,456
taylorswift13	\$5,814,213
selenagomez	\$5,369,471



Tremendous **real-time** capabilities in social media



TEMPORAL VIEW OF TWITTER

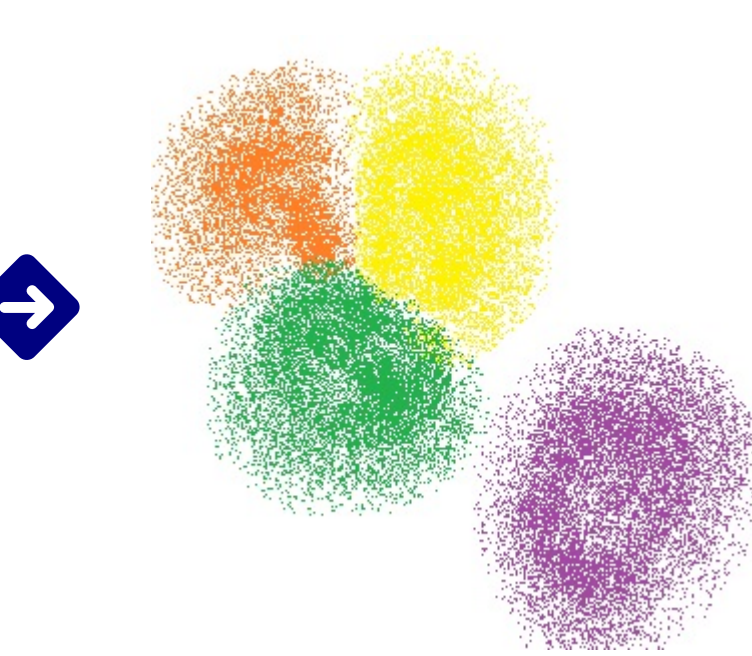
Limitations of Conventional Techniques

Natural Language Processing is not very effective for tweets! Short words, new abbreviations and word disambiguation.

Key Idea

Leverage signal processing to pre-process tweets into clusters. Subsequently improve semantic interpretation using natural language processing.

1: Construct Time Series



2: Cluster similar patterns

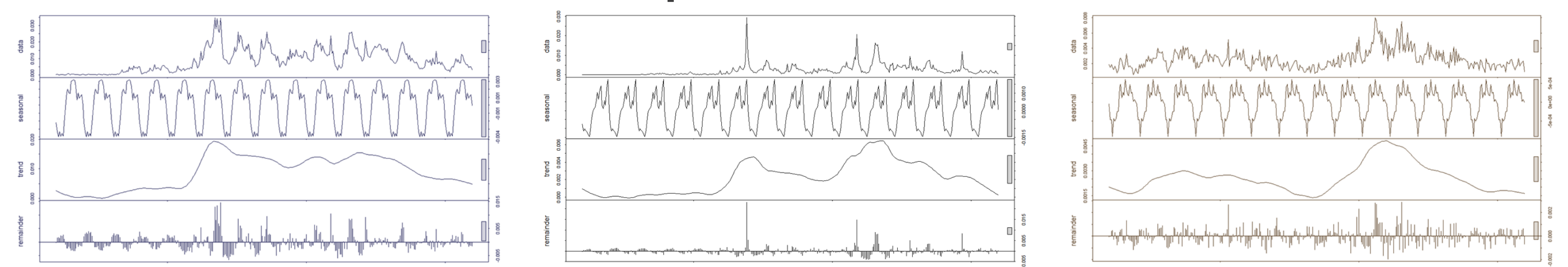
3: Semantic Correlation



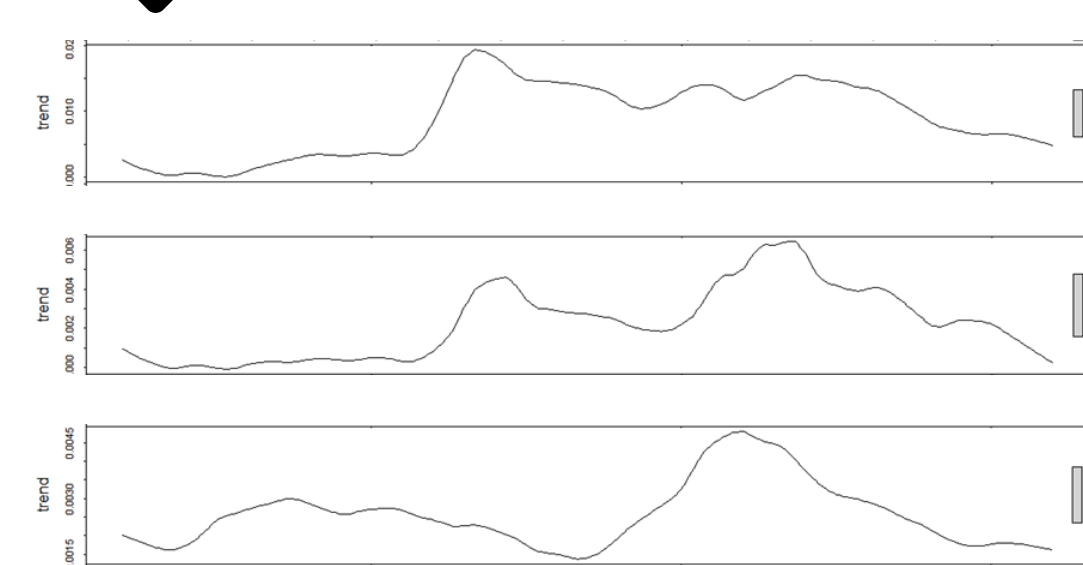
PRELIMINARY INSIGHTS

ARE TWEETS RELATING REAL-WORLD ENTITIES CORRELATED?

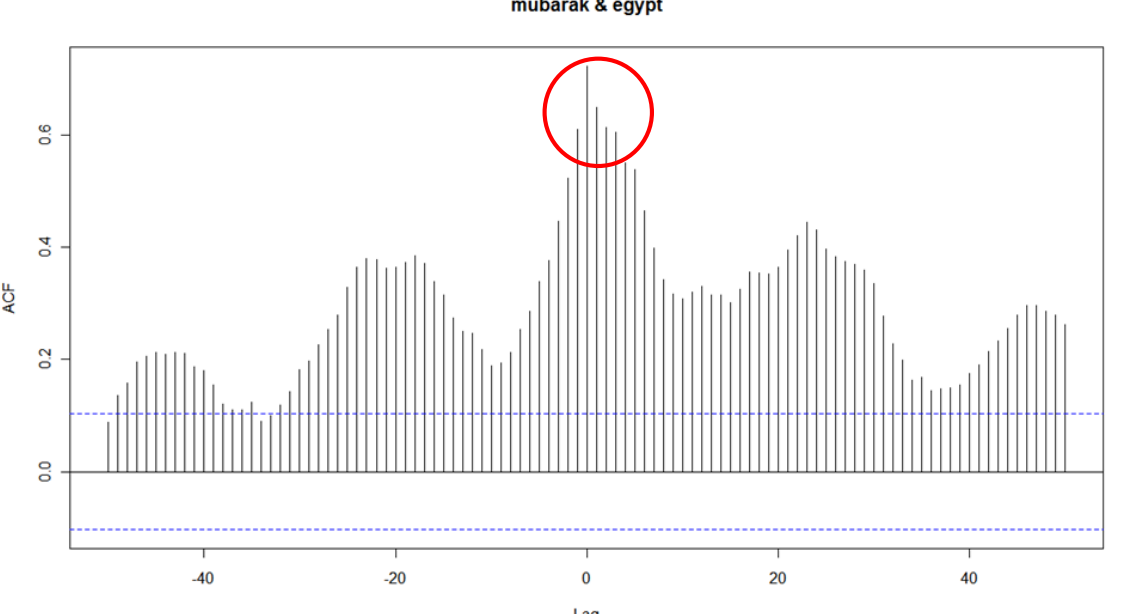
Seasonal Trend Decomposition based on Loess Smoother



Extract Trend Lines



Cross Correlation



CURRENTLY IN THE PIPELINE

Clustering Time Series

- Computing cross correlation is expensive!
- Convert time series into another representation
- Cluster (k-means or hierarchical) this representation
- Verify cluster utility

