Nowcasting the Stance of Social Media Users in a Sudden Vote:

The Case of the Greek Referendum

A. Tsakalidis, N. Aletras, A.I. Cristea, M. Liakata

The Alan Turing Institute

27th ACM International Conference on Information and Knowledge Management (CIKM 2018)

Introduction

Using Online Social Media (OSM) to predict a political index

Introduction

Using Online Social Media (OSM) to predict a political index

Macro-level (Gayo-Avello, 2012)

- Demographic/OSM bias (Morstatter et al., 2013; Miranda Filho et al., 2015)
- Train on opinion polls? (Lampos et al., 2013; Tsakalidis et al. 2015)
- Evaluation: 5-10 instances

Introduction

Using Online Social Media (OSM) to predict a political index

Macro-level

(Gayo-Avello, 2012)

- Demographic/OSM bias (Morstatter et al., 2013; Miranda Filho et al., 2015)
- Train on opinion polls? (Lampos et al., 2013; Tsakalidis et al. 2015)
- Evaluation: 5-10 instances

Micro-level

- Ideology prediction (Rao et al., 2010; Volkova et al., 2014; Preotiuc-Pietro et al., 2018)
- Evaluation: static – real world setting?

Introduction

Using Online Social Media (OSM) to predict a political index

Macro-level

(Gayo-Avello, 2012)

- Demographic/OSM bias (Morstatter et al., 2013; Miranda Filho et al., 2015)
- Train on opinion polls? (Lampos et al., 2013; Tsakalidis et al. 2015)
- Evaluation: 5-10 instances

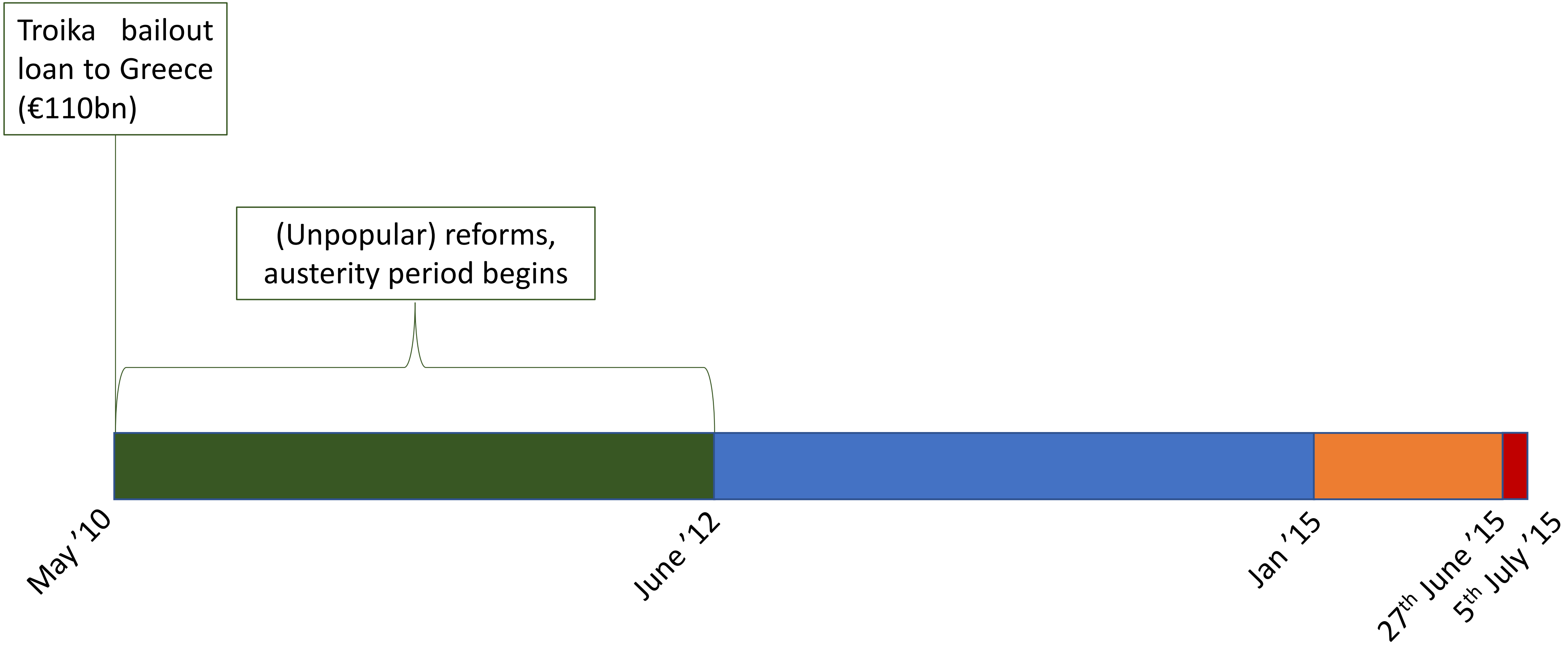
Micro-level

- Ideology prediction (Rao et al., 2010; Volkova et al., 2014; Preotiuc-Pietro et al., 2018)
- Evaluation: static – real world setting?

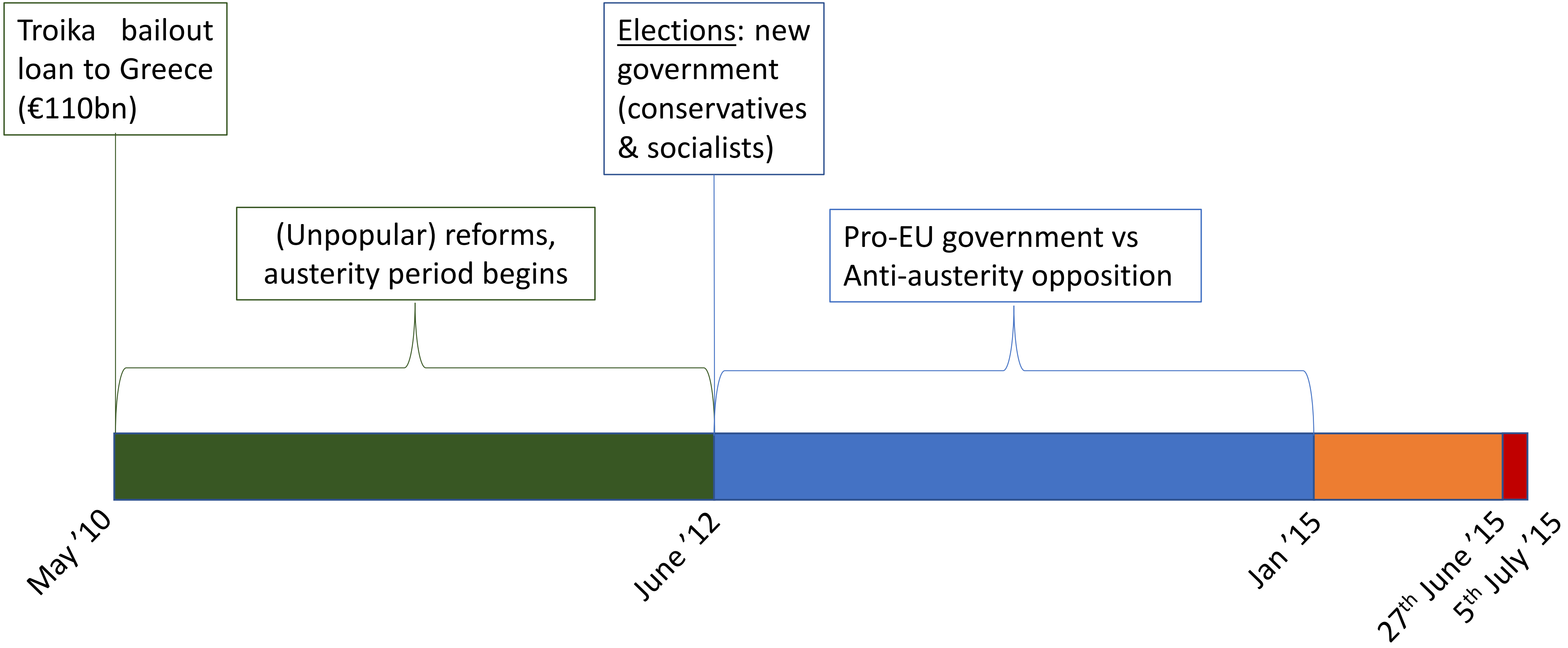
What about major/sudden electoral cases in which we have no historical evidence?

Greek crisis in a nutshell

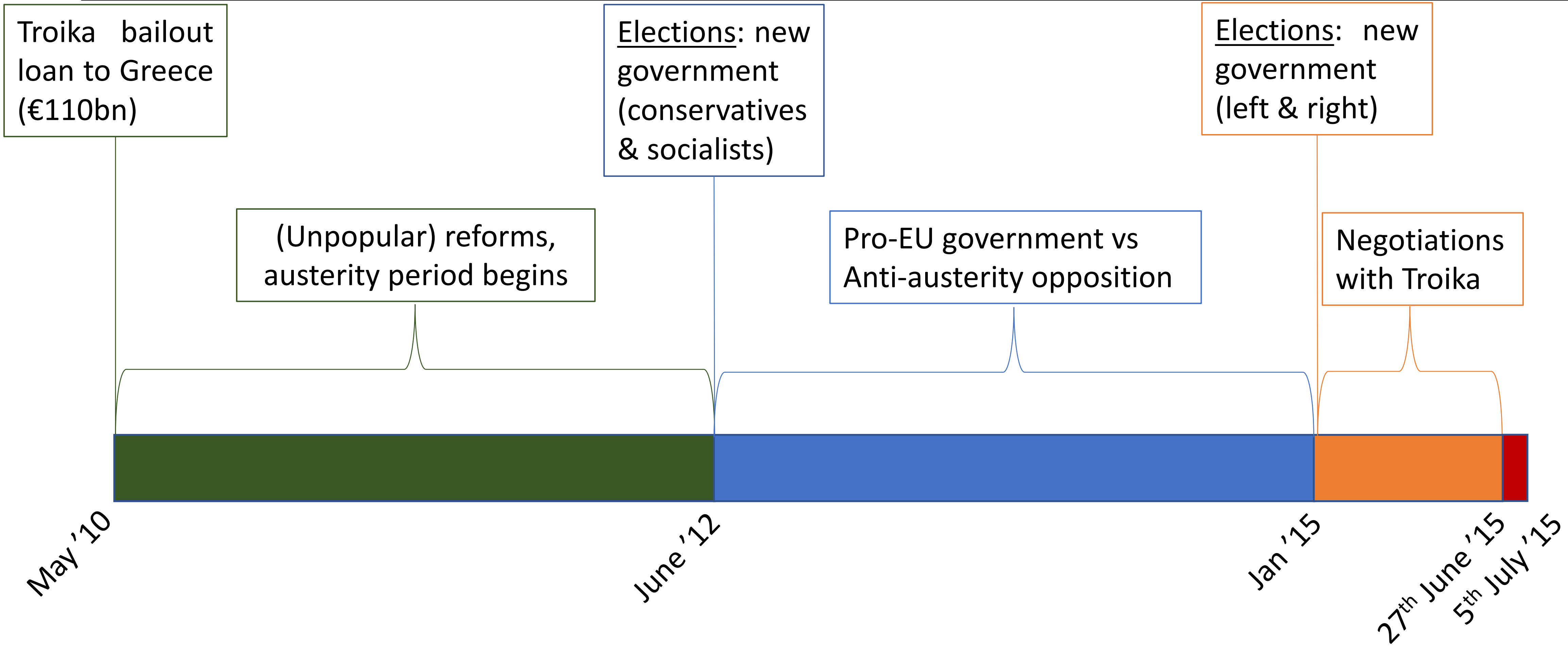
Greek crisis in a nutshell



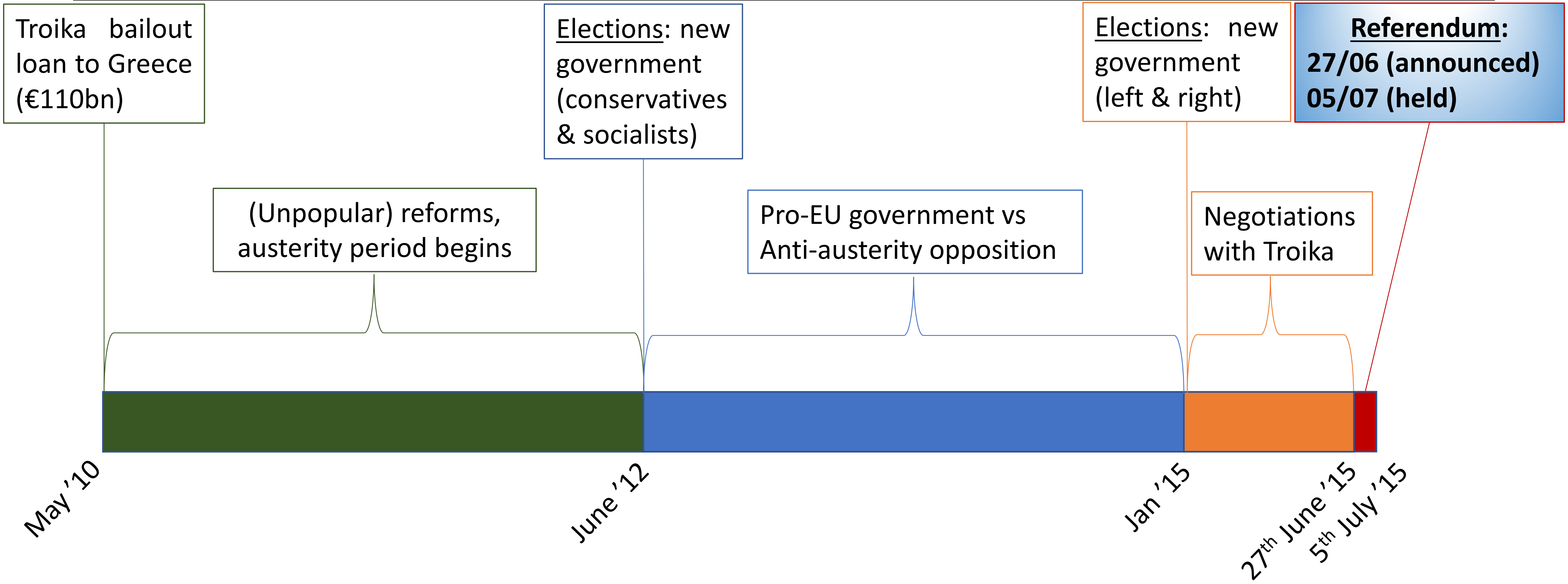
Greek crisis in a nutshell



Greek crisis in a nutshell



Greek crisis in a nutshell



27th June-5th July extreme polarisation, demonstrations, capital controls…

Task Description

Aim: *nowcast the stance (YES/NO) of social media users*
(binary task over time: real-world setting)

Task Description

Aim: *nowcast the stance (YES/NO) of social media users*

(binary task over time: real-world setting)

For *t* in [0, 1, …, 8]: *#[26/06-04/07]*

1. Assume training set of *n* users at *t*:
2. Learn a function at *t*:
3. Evaluate performance on test users at *t*

$$D_t = \{(x_t^{(1)}, y^{(1)}), ..., (x_t^{(n)}, y^{(n)})\}$$

$$\hat{y} = f_t(x_t^{(j)})$$

Task Description

Aim: *nowcast the stance (YES/NO) of social media users*
(binary task over time: real-world setting)

For *t* in [0, 1, …, 8]: *#[26/06-04/07]*

1. Assume training set of *n* users at *t*: $D_t = \{(x_t^{(1)}, y^{(1)}), \ldots, (x_t^{(n)}, y^{(n)})\}$
2. Learn a function at *t*: $\hat{y} = f_t(x_t^{(j)})$
3. Evaluate performance on test users at *t*

Challenges:

- real-world application; no manual effort – training set?
- combine *asynchronous* and *time sensitive* information sources?

Dataset

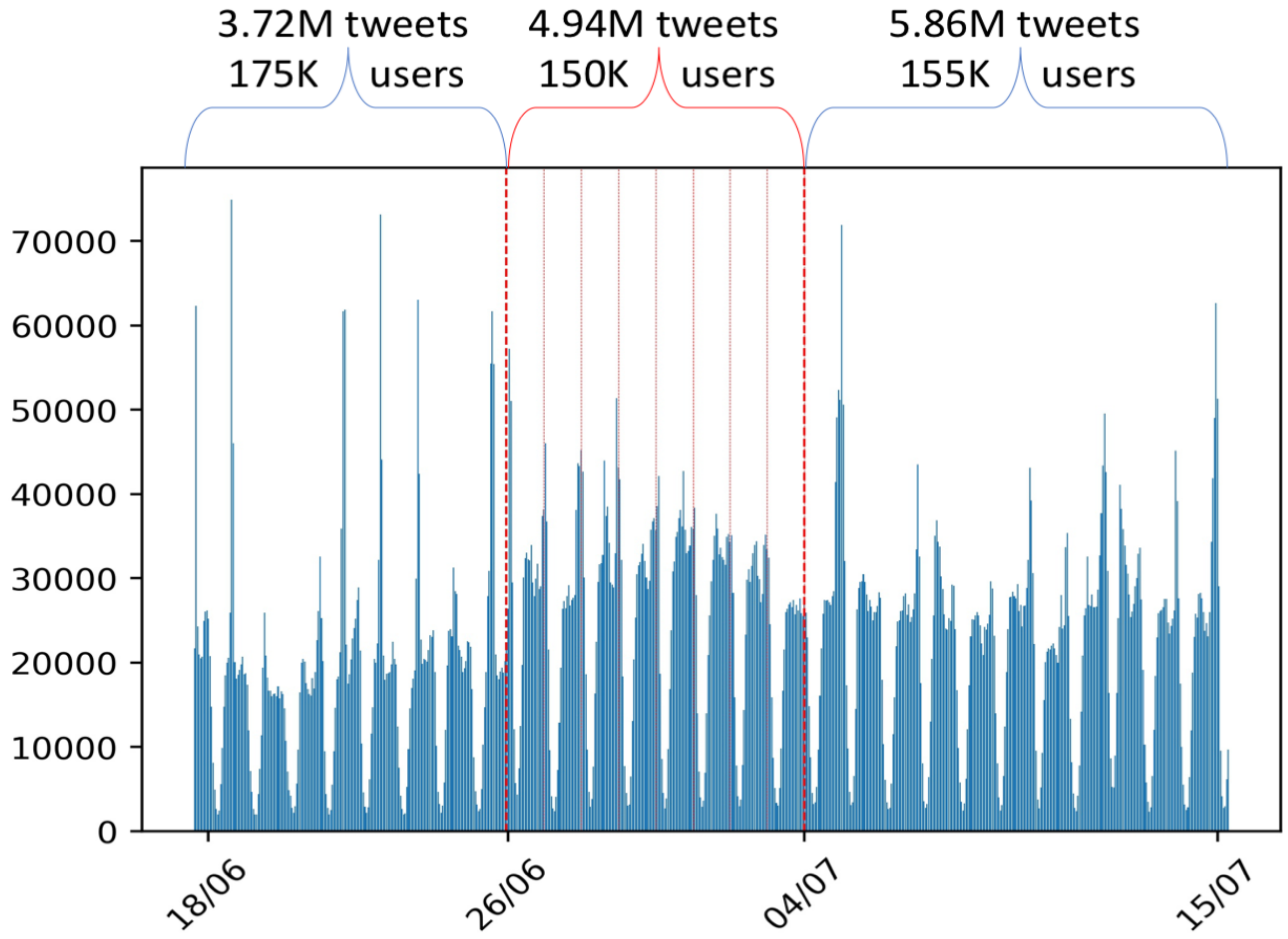
Twitter Streaming API (Greek stop-words)

- 14.6M tweets
- 304K users

Dataset

Twitter Streaming API (Greek stop-words)

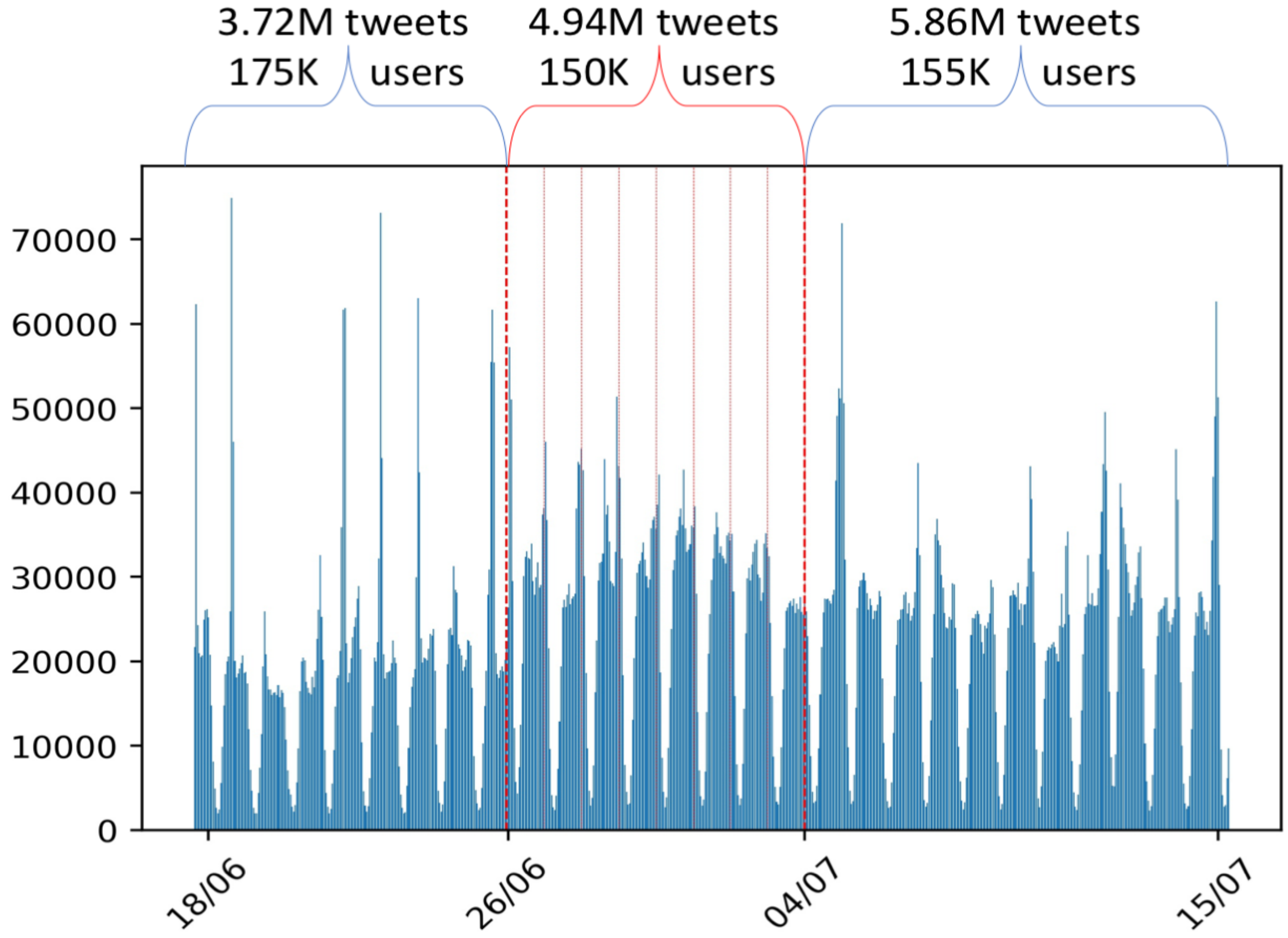
- 14.6M tweets
- 304K users



Dataset

Twitter Streaming API (Greek stop-words)

- 14.6M tweets
- 304K users



Test set: annotate 2.7K users

- >10 tweets; >3 polarised hashtags
- <u>Annotation</u>: YES/NO (or N/A)
- Cohen's κ: .98 (.74)

- Final set: 2,197 users
- 77/23 NO/YES (result: 61/39)

Training Set

Seed accounts with "known" labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Training Set

Seed accounts with “known” labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Assumption: user RTs pro-EU accounts => vote for “YES”

Training Set

Seed accounts with “known” labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Assumption: user RTs pro-EU accounts => vote for “YES”

$$scores = [PMI(u, YES) - PMI(u, NO), \forall u]$$

Threshold t

Training Set

Seed accounts with “known” labels:

- political parties (pro-EU vs anti-austerity)
- MPs
- keyword-based profile search

148/119
YES/NO

Assumption: user RTs pro-EU accounts => vote for “YES”

$$scores = [PMI(u, YES) - PMI(u, NO), \forall u]$$

Threshold t

$$scores[u] > t \Rightarrow YES$$

965 YES

$$scores[u] < -t \Rightarrow NO$$

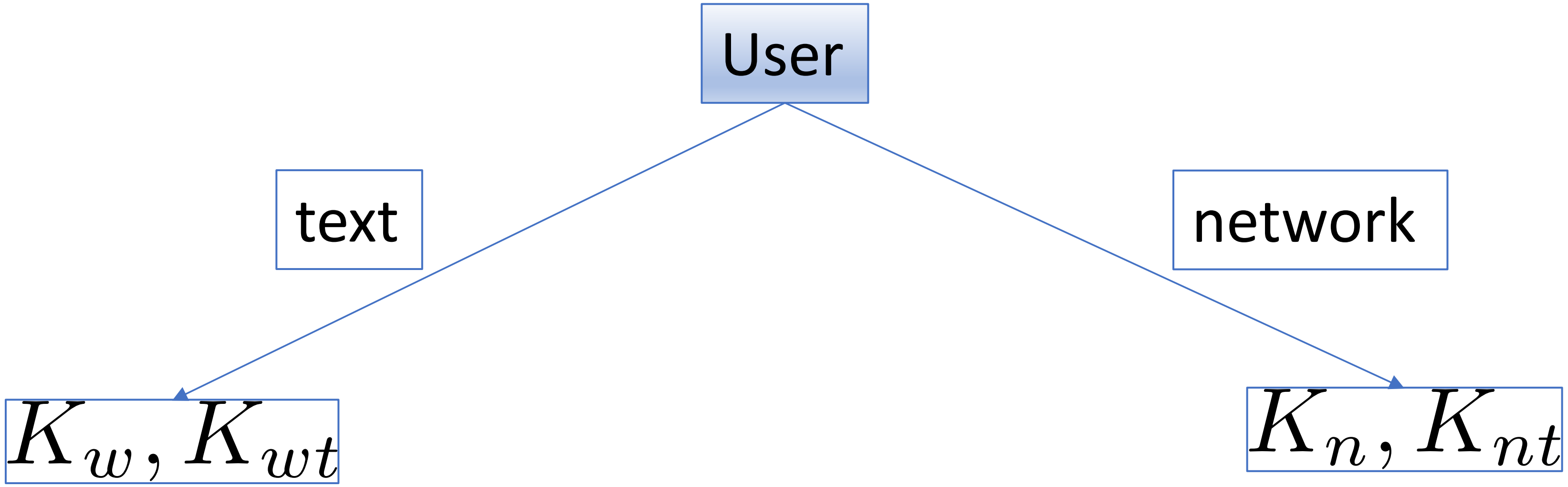
1,165 NO

Data (summary)

	#users	#tweets	YES (%)	NO (%)
Train	2,121	867K	45%	55%
Test	2,197	768K	23%	77%

Methodology

Convolution kernels



Methodology

User

Convolution kernels

text

network

K_w, K_{wt}

K_n, K_{nt}

Kernel combination

$$w_w K_w + w_{wt} K_{wt} + w_n K_n + w_{nt} K_{nt}$$

Convolution Kernels

Let Z_i be user representations (based on text or network structure) over times T_i:

$$Z_a = \{z_a^1, \ldots, z_a^N\} \qquad T_a = \{t_a^1, \ldots, t_a^N\}$$

$$Z_b = \{z_b^1, \ldots, z_b^M\} \qquad T_b = \{t_b^1, \ldots, t_b^M\}$$

Convolution Kernels

Let Z_i be user representations (based on text or network structure) over times T_i:

$$Z_a = \{z_a^1, \dots, z_a^N\} \qquad T_a = \{t_a^1, \dots, t_a^N\}$$

$$Z_b = \{z_b^1, \dots, z_b^M\} \qquad T_b = \{t_b^1, \dots, t_b^M\}$$

TEXT: tweet representation
NET: retweet-based network representation

We can define a convolution kernel as:

$$K_z(a,b) = \frac{1}{|Z_a|\,|Z_b|} \sum_{i,j} k_z(z_a^i, z_b^j)$$

K_w (TEXT)
K_n (NET)

Convolution Kernels

Let Z_i be user representations (based on text or network structure) over times T_i:

$$Z_a = \{z_a^1, \ldots, z_a^N\} \qquad T_a = \{t_a^1, \ldots, t_a^N\}$$

$$Z_b = \{z_b^1, \ldots, z_b^M\} \qquad T_b = \{t_b^1, \ldots, t_b^M\}$$

TEXT: tweet representation
NET: retweet-based network representation

We can define a convolution kernel as:

$$K_z(a,b) = \frac{1}{|Z_a|\,|Z_b|} \sum_{i,j} k_z(z_a^i, z_b^j)$$

K_w (TEXT)
K_n (NET)

Account for time (Lukasik & Cohn, 2016):

$$K_{zt}(a,b) = \frac{1}{|Z_a|\,|Z_b|} \sum_{i,j} k_z(z_a^i, z_b^j) k_t(t_a^i, t_b^j)$$

K_{wt} (TEXT)
K_{nt} (NET)

Kernel Combination

Kernel summation:

$$K(a,b) = \sum_{k=1}^{|K|} K_k(a,b)$$

Kernel Combination

Kernel summation:

$$K(a,b) = \sum_{k=1}^{|K|} K_k(a,b)$$

Multiple (convolution) kernel learning (Sonnenburg et al., 2006):

$$K(a,b) = \sum_{k=1}^{|K|} w_k K_k(a,b)$$

$$f(x) = sign\left(\sum_{i=1}^{N} a_i K(x, x_i) + b\right)$$

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Baselines

- LR, FF, RF, SVM on feature aggregates (avg word2vec; final LINE embeddings)

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Baselines

- LR, FF, RF, SVM on feature aggregates (avg word2vec; final LINE embeddings)

Our models

- Single modality: SVM_w , SVM_{wt} , SVM_n , SVM_{nt}
- Combinations: SVM_+ , MCKL

Experiments

Features

- TEXT: word2vec (Mikolov et al., 2013); 50 dim; trained on 15M tweets
- NETWORK: LINE (Tang et al., 2015); 50 dim; trained every 12 hours

Baselines

- LR, FF, RF, SVM on feature aggregates (avg word2vec; final LINE embeddings)

Our models

- Single modality: SVM_w , SVM_{wt} , SVM_n , SVM_{nt}
- Combinations: SVM_+ , MCKL

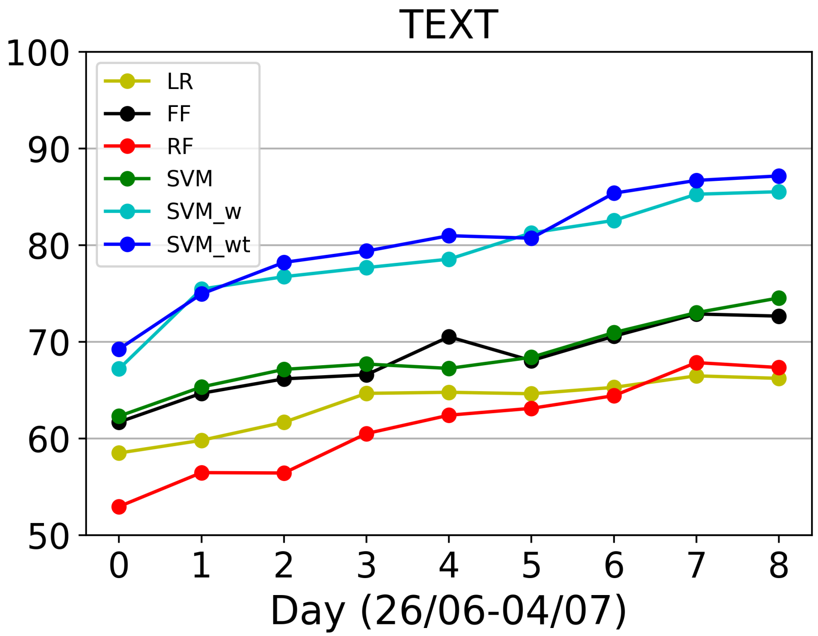
Evaluation

- macro-average F-score
- *nine* evaluation time points (every midnight)

Results

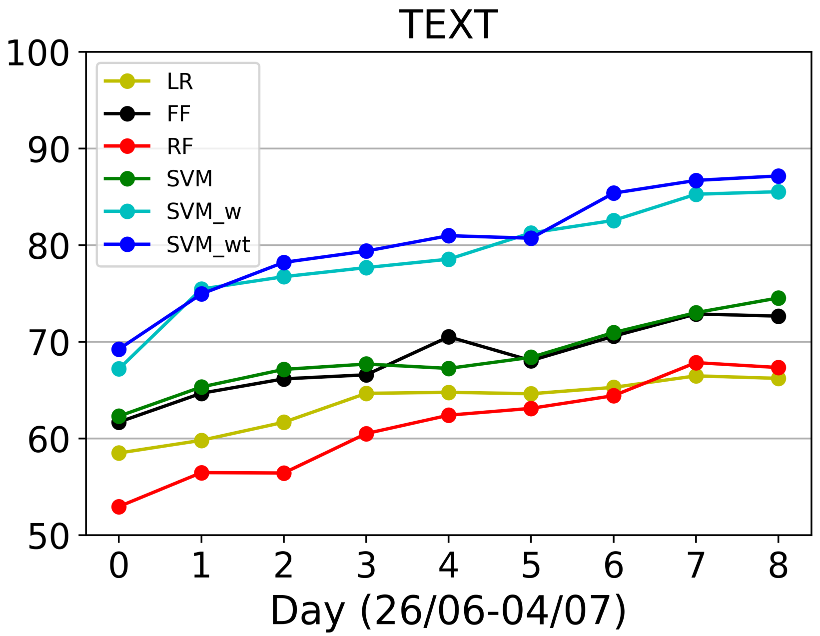
	LR	FF	RF	SVM	SVM_f	SVM_{ft}	SVM_+	MCKL
TEXT								
NET								
BOTH								

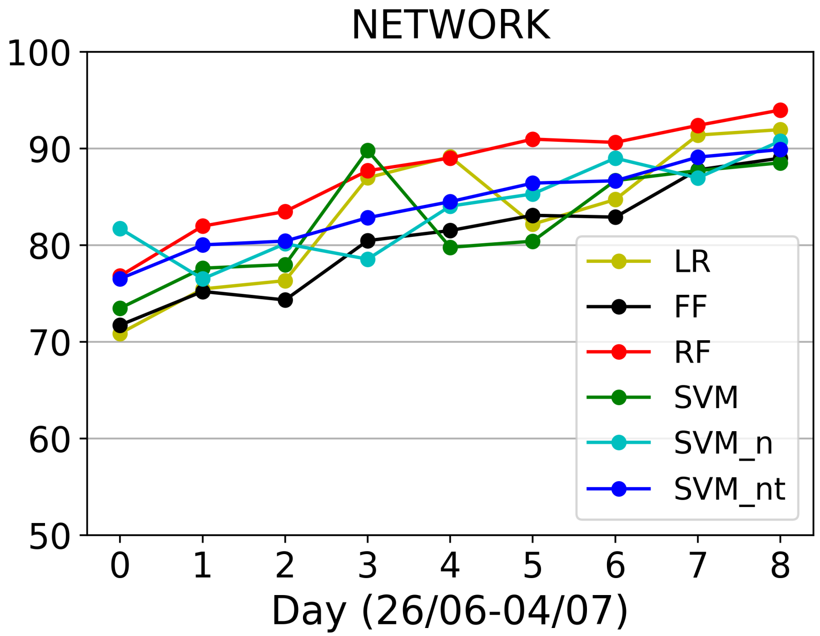
Results



	LR	FF	RF	SVM	SVM_f	SVM_{ft}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET								
BOTH								

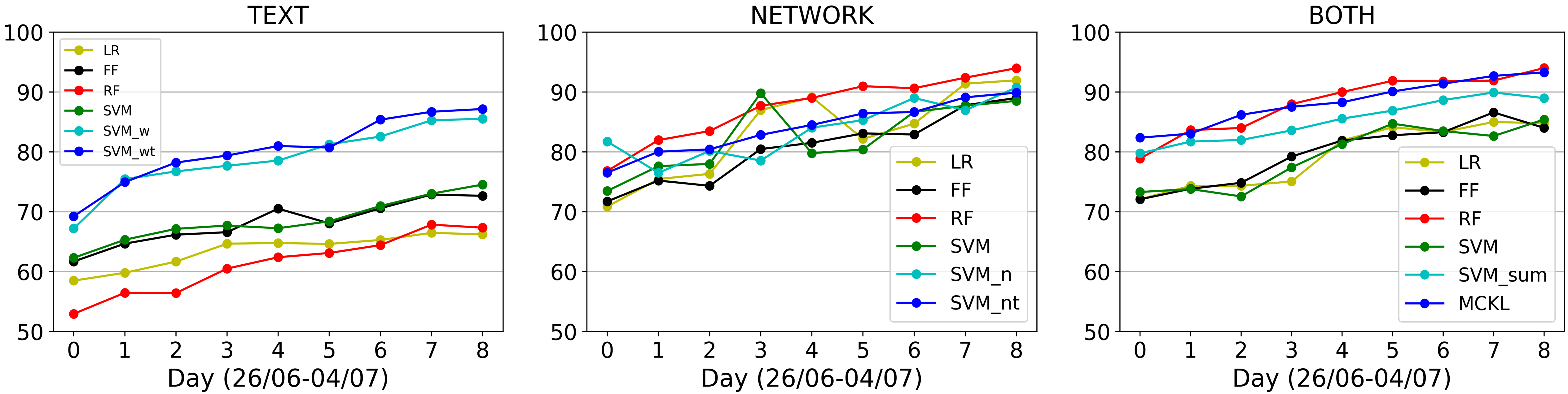
Results





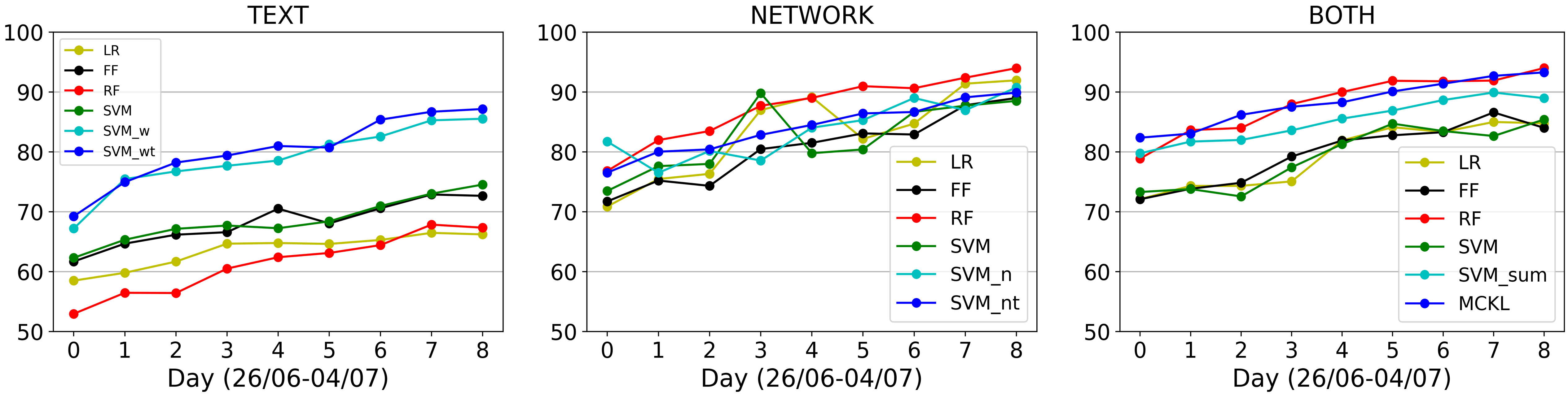
	LR	FF	RF	SVM	SVM_f	SVM_{ft}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET	83.21	80.66	**87.43**	82.43	83.65	84.03	--	--
BOTH								

Results



	LR	FF	RF	SVM	SVM_f	SVM_{ft}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET	83.21	80.66	**87.43**	82.43	83.65	84.03	--	--
BOTH	79.43	79.83	88.22	79.39	--	--	85.22	**88.31**

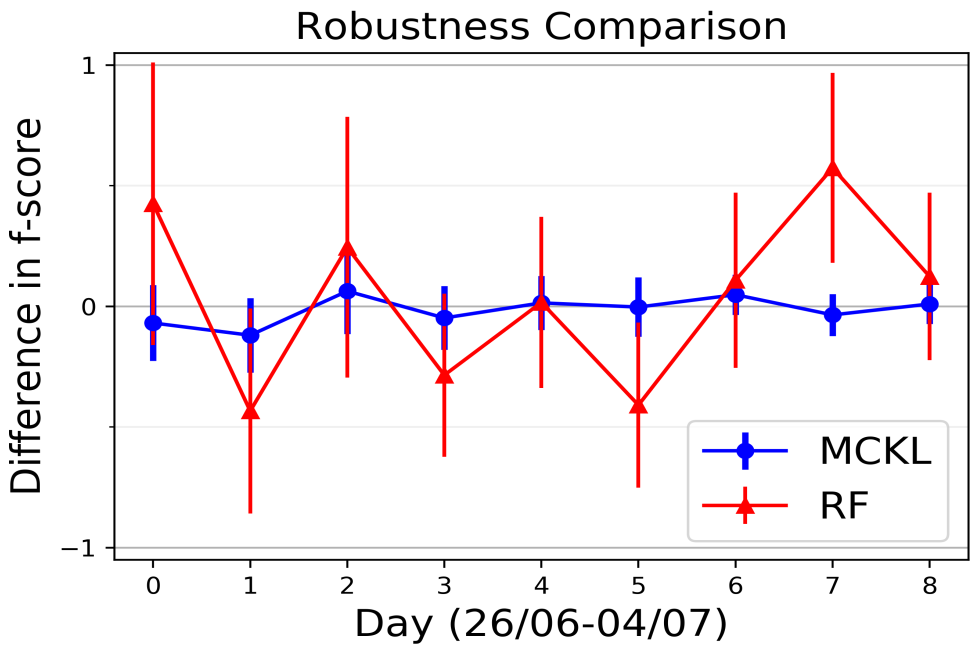
Results



	LR	FF	RF	SVM	SVM_f	SVM_{ft}	SVM_+	MCKL
TEXT	63.55	68.19	61.27	68.51	78.91	**80.30**	--	--
NET	83.21	80.66	**87.43**	82.43	83.65	84.03	--	--
BOTH	79.43	79.83	88.22	79.39	--	--	85.22	**88.31**

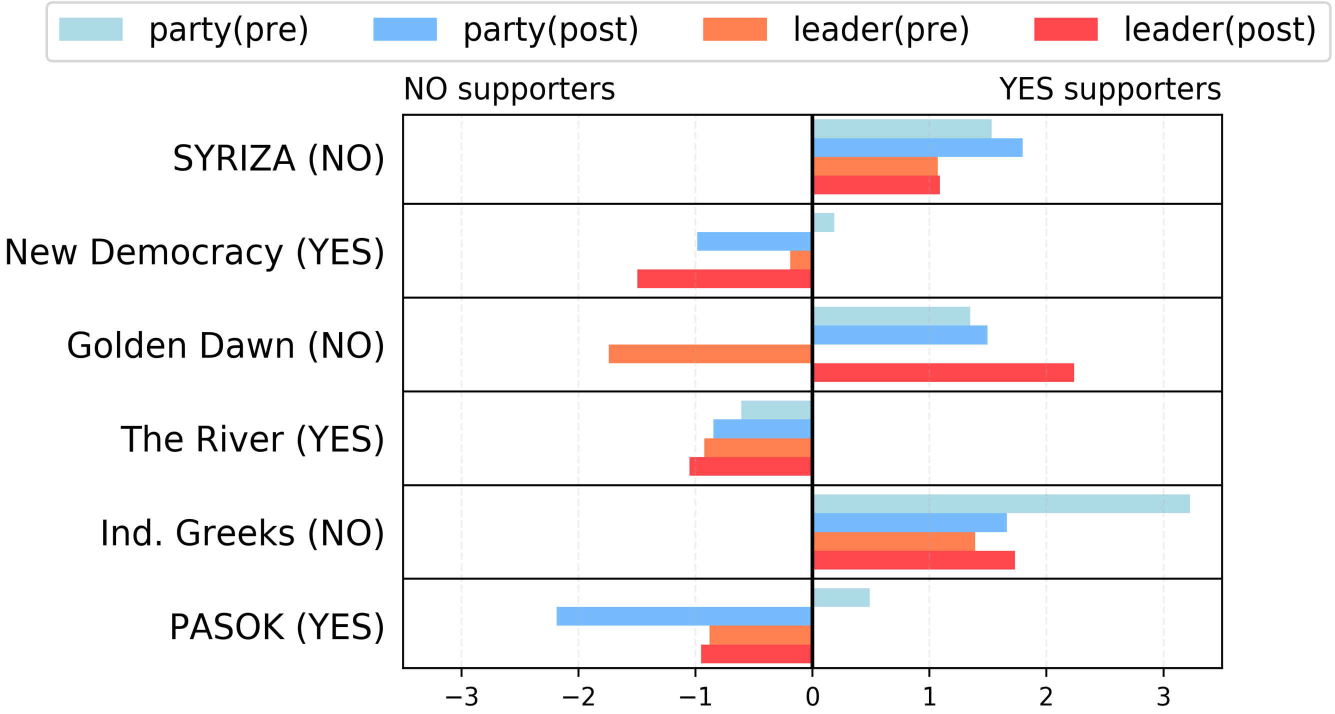
Results

RF vs MCKL: 100 experiments with added noisy features



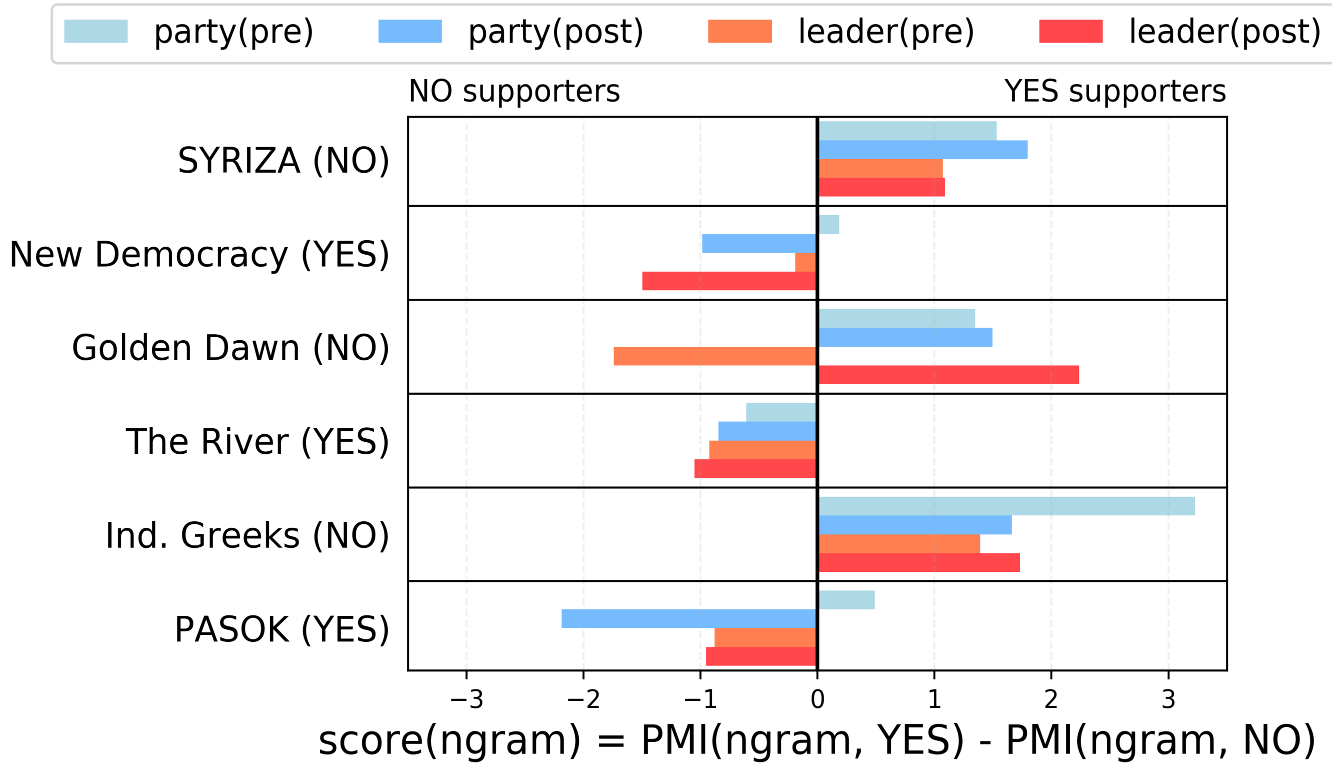
Discussion (text)

Discussion (text)



score(ngram) = PMI(ngram, YES) - PMI(ngram, NO)

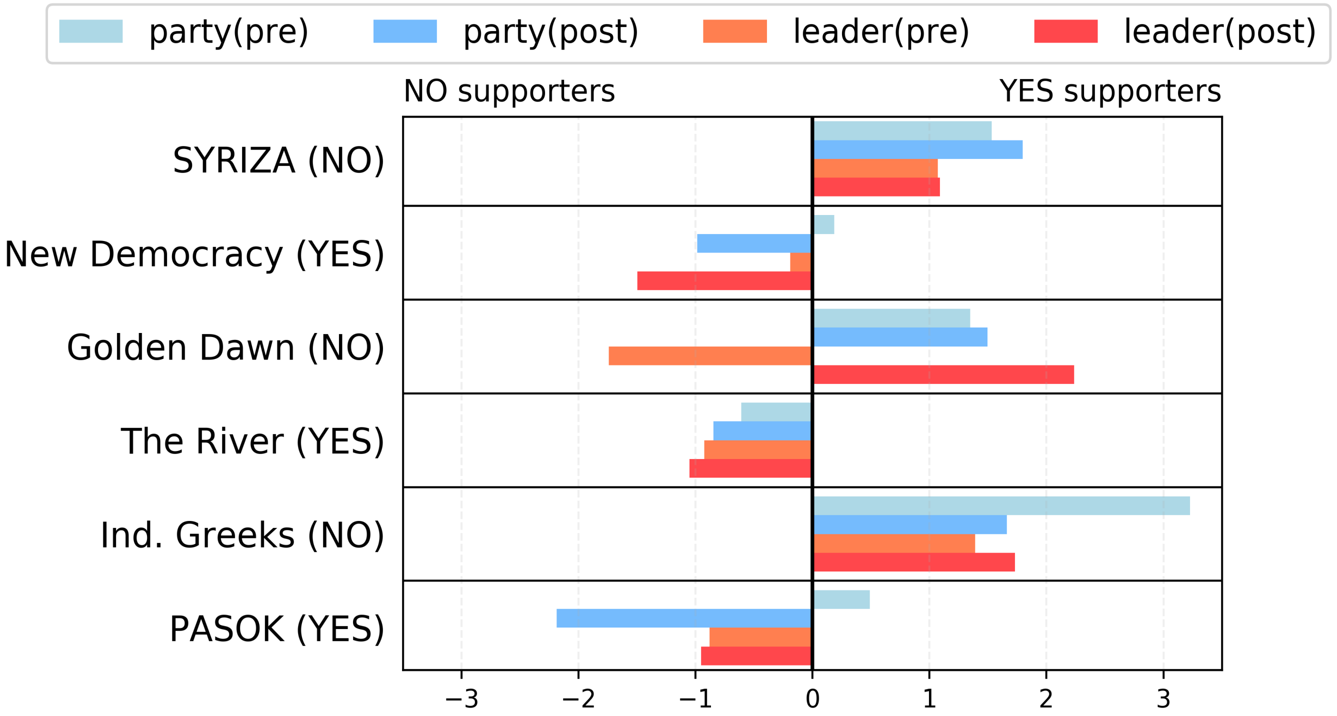
Discussion (text)



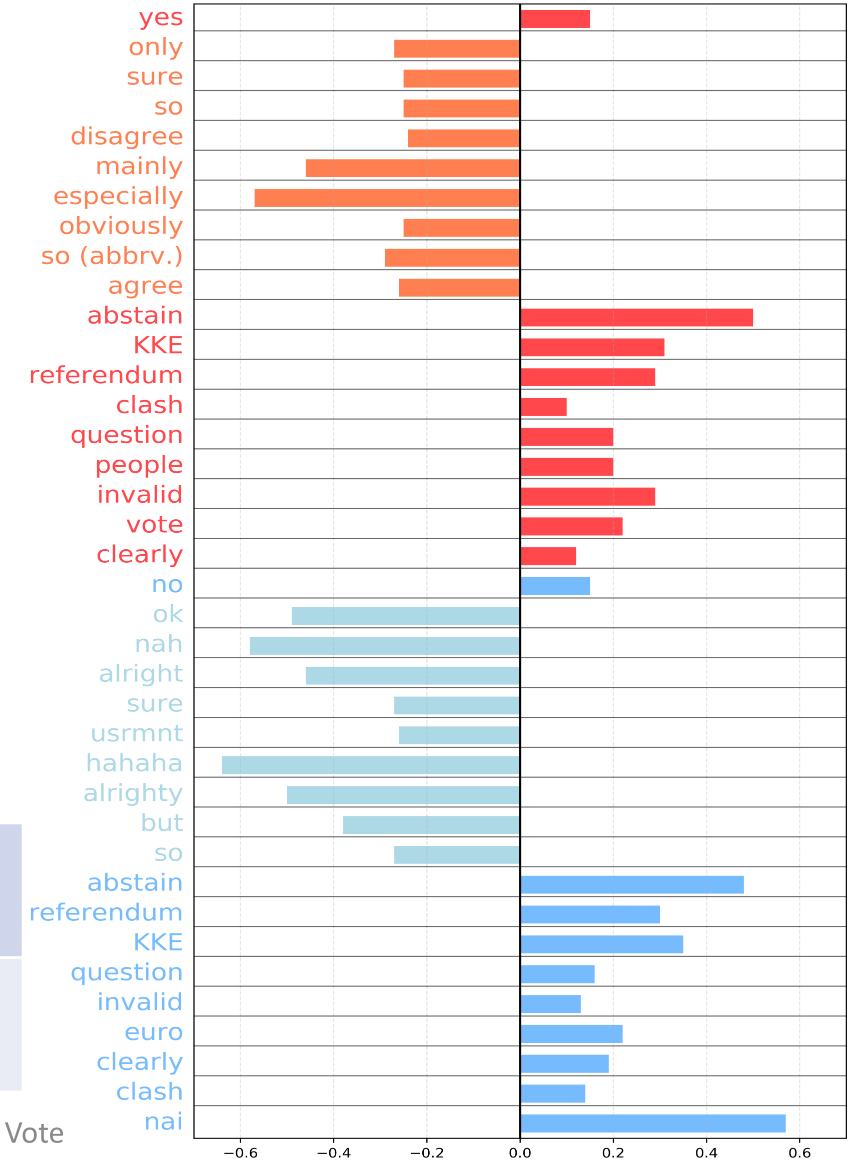
They say that there is a long queue of people in ATMs and they show 6 people waiting; this is not a queue, this is PASOK.

I want to write something funny regarding the statements made by Kammenos, but I cannot find something funnier than the statements made by Kammenos.

Discussion (text)



score(ngram) = PMI(ngram, YES) - PMI(ngram, NO)

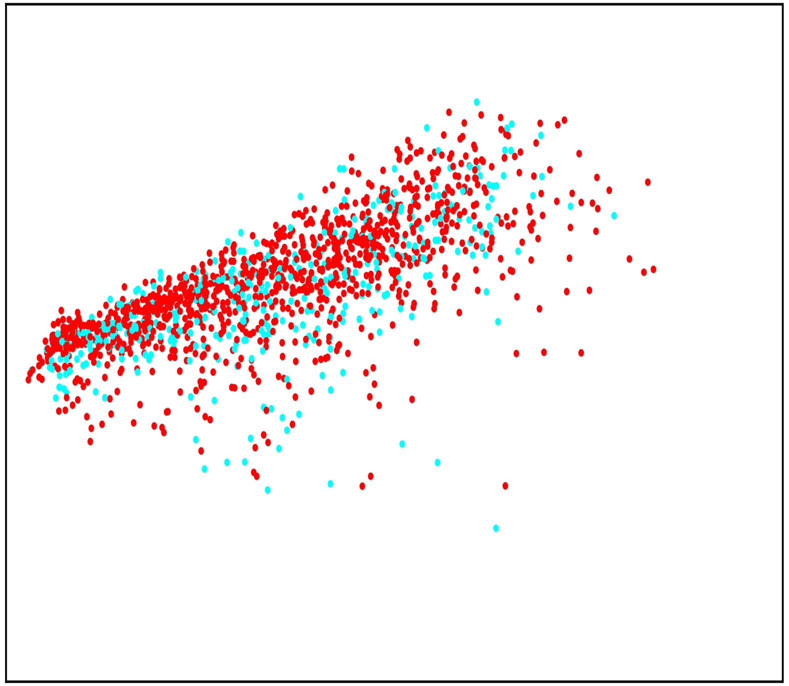


They say that there is a long queue of people in ATMs and they show 6 people waiting; this is not a queue, this is PASOK.

I want to write something funny regarding the statements made by Kammenos, but I cannot find something funnier than the statements made by Kammenos.

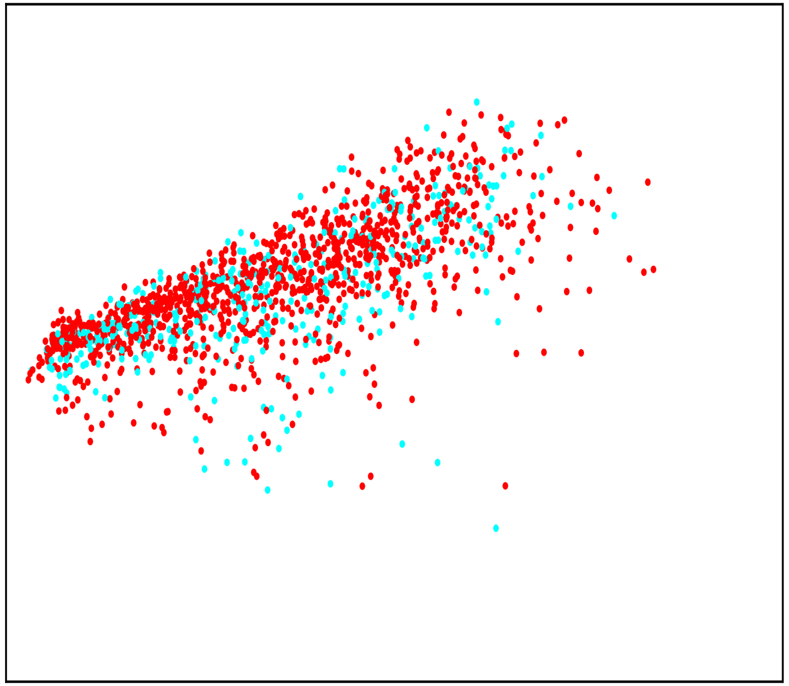
Discussion (network)

Discussion (network)

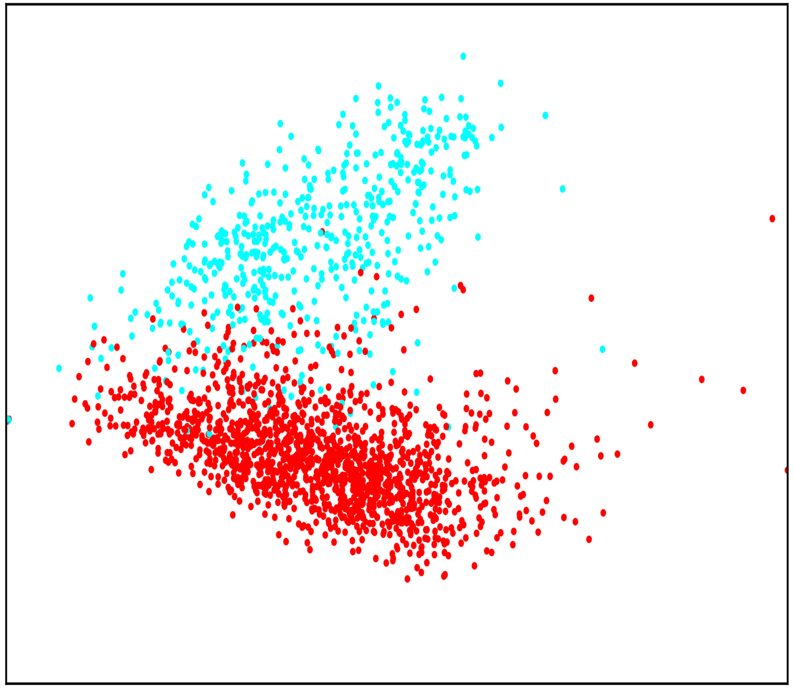


User representations **before** the announcement of the referendum

Discussion (network)

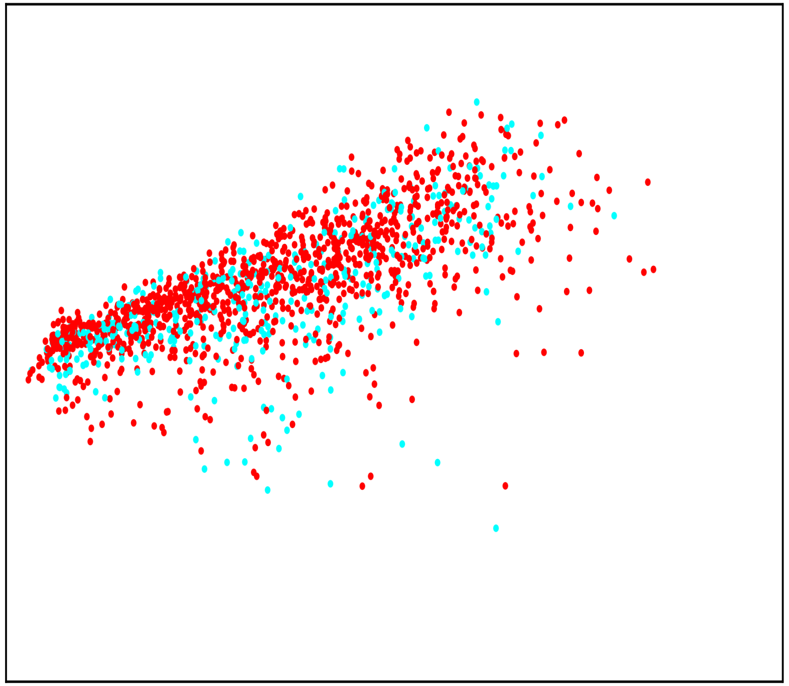


User representations **before** the announcement of the referendum

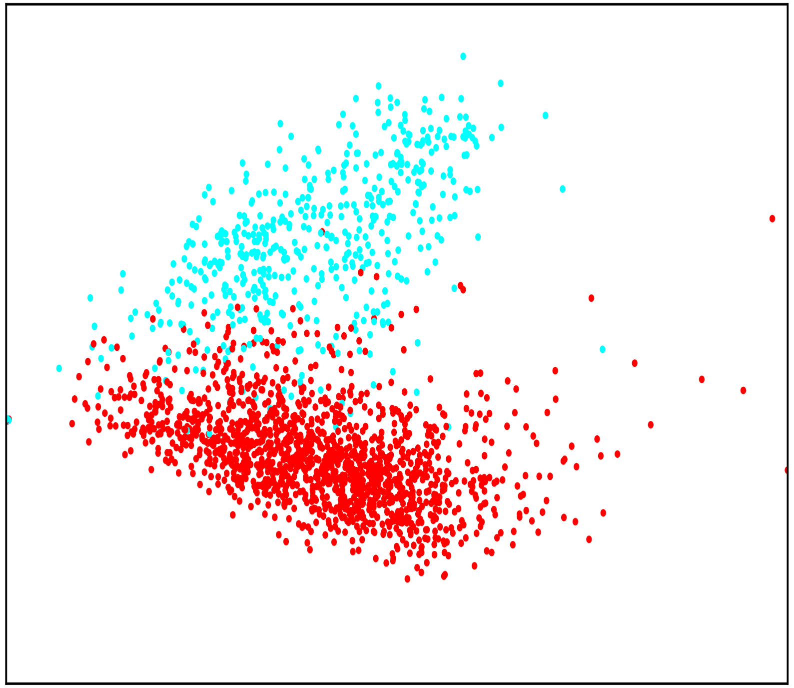


User representations **after** the announcement of the referendum

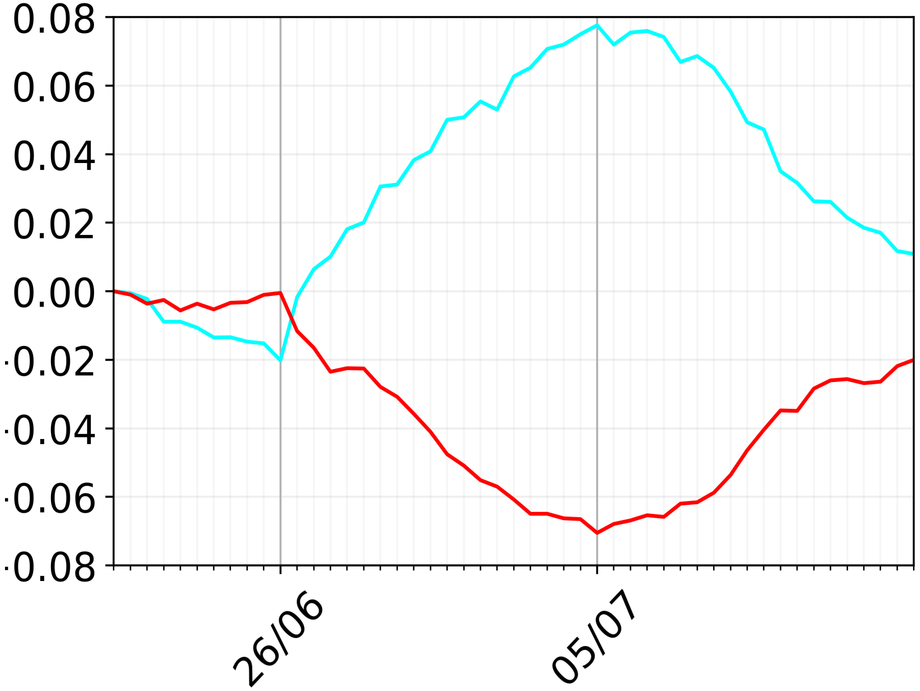
Discussion (network)



User representations **before** the announcement of the referendum



User representations **after** the announcement of the referendum



User representations in a sliding window:
sim(YES, YES*)-sim(YES, NO*)
sim(NO, YES*)-sim(NO, NO*)

Conclusion

Nowcasting the voting intention during the Greek Referendum (2015)

Evaluation: real-time setting
Temporal modelling: highly important (text)
Network structure: most predictive

MCKL boosts performance of weaker kernels/models in a robust way

Conclusion

Nowcasting the voting intention during the Greek Referendum (2015)

Evaluation: real-time setting
Temporal modelling: highly important (text)
Network structure: most predictive

MCKL boosts performance of weaker kernels/models in a robust way

Future work:

- Random sample of test users
- Longer lasting electoral races
- More modalities (e.g., location, user profile)

Thank you!

Any questions?

@adtsakal
@nikaletras
@xrysoflhs

atsakalidis@turing.ac.uk

References

Daniel Gayo-Avello. 2012. I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper – A Balanced Survey on Election Prediction using Twitter Data. arXiv preprint arXiv:1204.6441 .

Vasileios Lampos, Daniel Preoțiuc-Pietro, and Trevor Cohn. 2013. A user-centric model of voting intention from social media. In ACL, volume 1, pages 993–1003.

Michal Lukasik and Trevor Cohn. 2016. Convolution Kernels for Discriminative Learning from Streaming Text. In AAAI, pages 2757–2763.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 .

Renato Miranda Filho, Jussara M. Almeida, and Gisele L. Pappa. 2015. Twitter population sample bias and its impact on predictive outcomes: a case study on elections. In Advances in Social Networks Analysis and Mining, pages 1254–1261.

Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley. "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose." In ICWSM. 2013.

Daniel Preotiuc-Pietro, Ye Liu, Daniel Hopkins, and Lyle Ungar. 2017. Beyond Binary Labels: Political Ideology Prediction of Twitter Users. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vol. 1. 729–740.

Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying Latent User Attributes in Twitter. In Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents. ACM, 37–44.

Sören Sonnenburg, Gunnar Rätsch, Christin Schäfer, and Bernhard Schölkopf. 2006. Large scale multiple kernel learning. Journal of Machine Learning Research, 7(Jul):1531–1565.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. LINE: Large-scale information network embedding. In WWW, pages 1067–1077.

Adam Tsakalidis, Symeon Papadopoulos, Alexandra I. Cristea, and Yiannis Kompatsiaris. 2015. Predicting elections for multiple countries using Twitter and polls. IEEE Intelligent Systems, 30(2):10–17.

Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring User Political Preferences from Streaming Communications. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vol. 1. 186–196.