

# All-in-one: Multi-task Learning for Rumour Verification

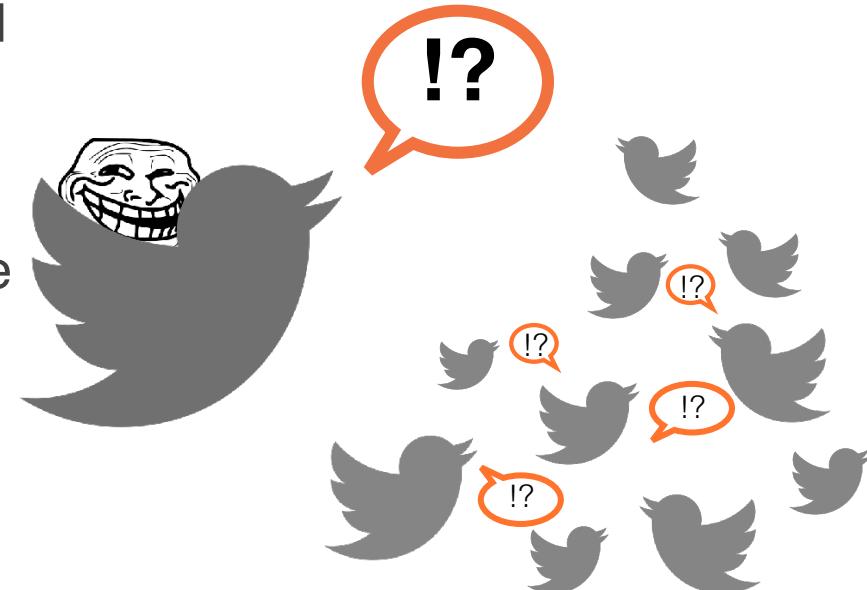
Elena Kochkina, Maria Liakata and Arkaitz Zubiaga



# Rumours on social media

WARWICK

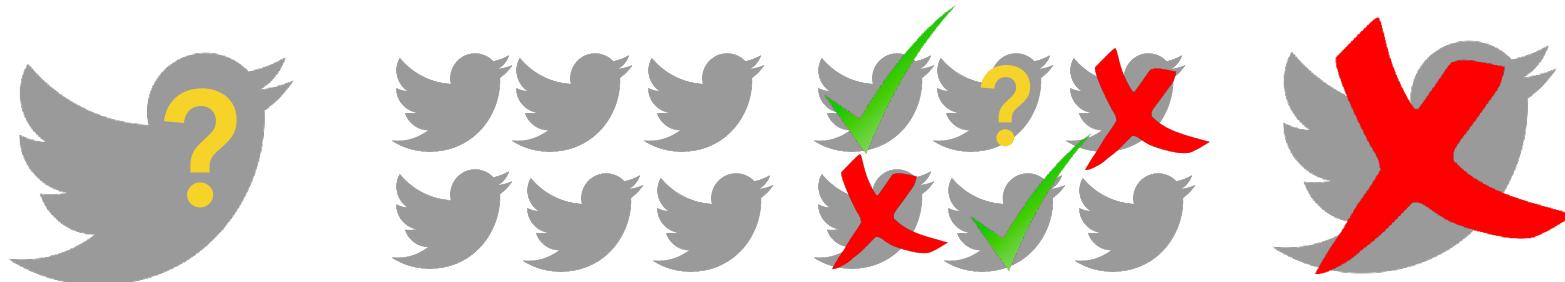
- **Rumour** is an unverified story related to the event and spreading on social media.
- The rapid spread of rumours can have **immediate and far-reaching consequences**.
- **Manual verification is not scalable**



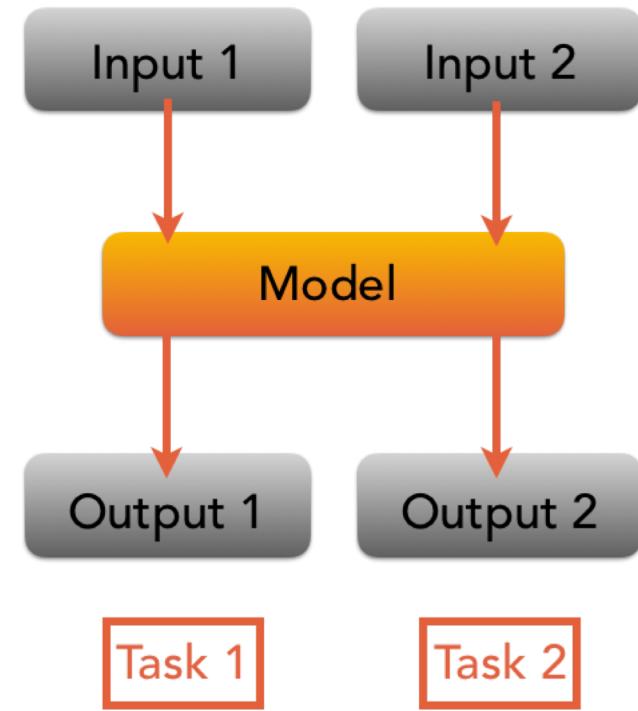
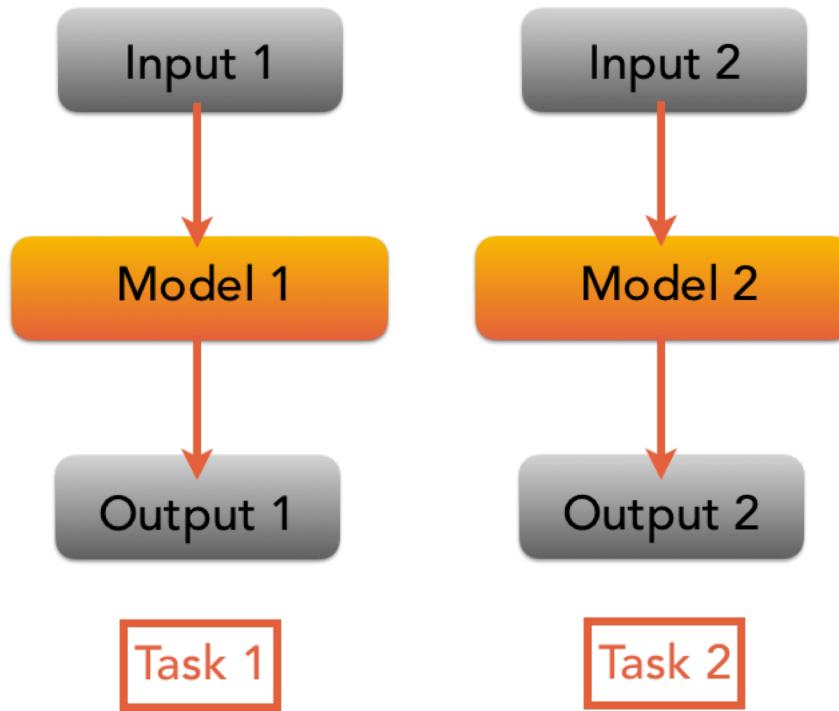
**Need for an automated system that would assist in this process**

# Rumour verification pipeline

WARWICK



# Multi-task learning I



# Multi-task learning II

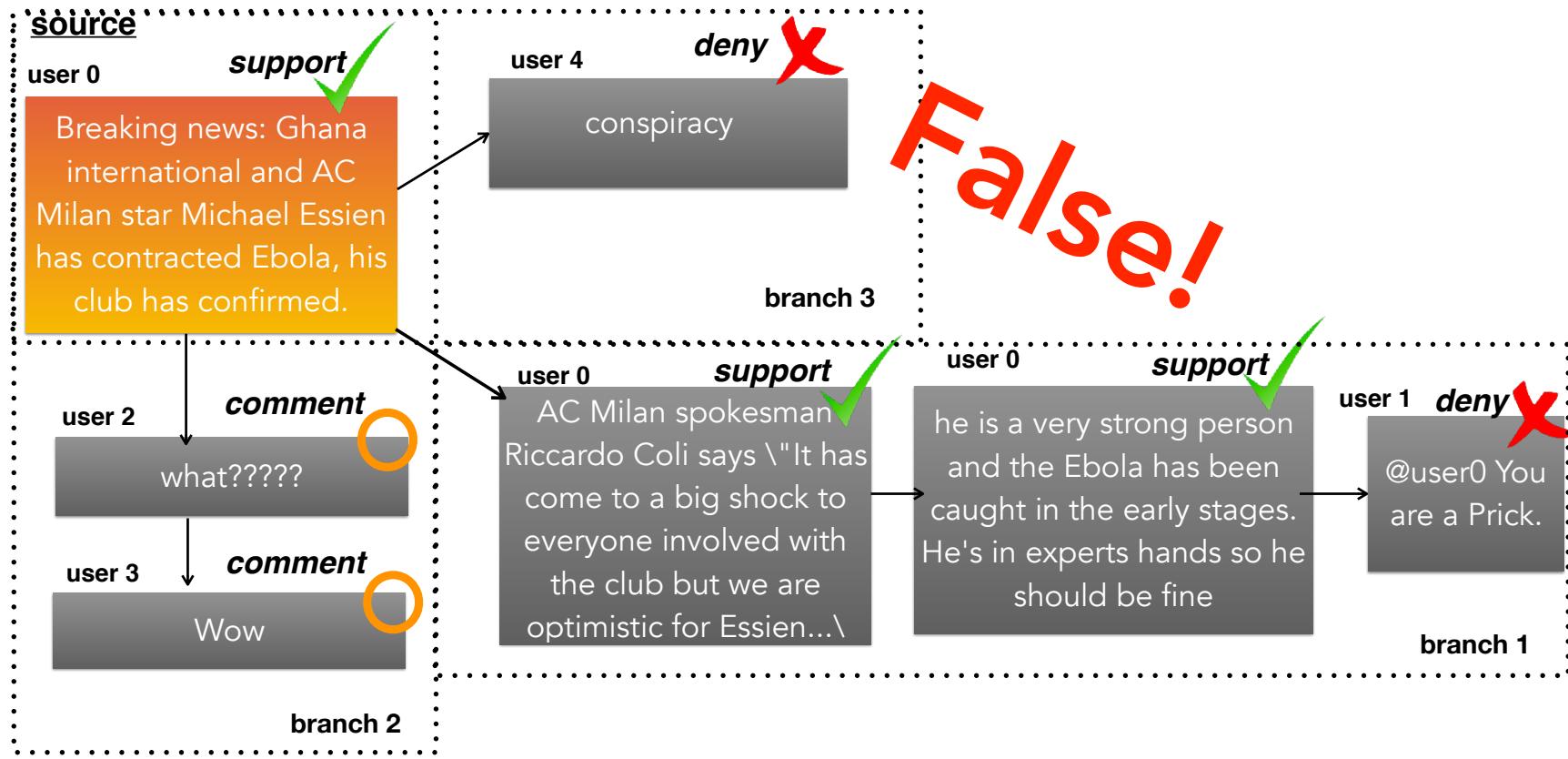


Success is attributed to:

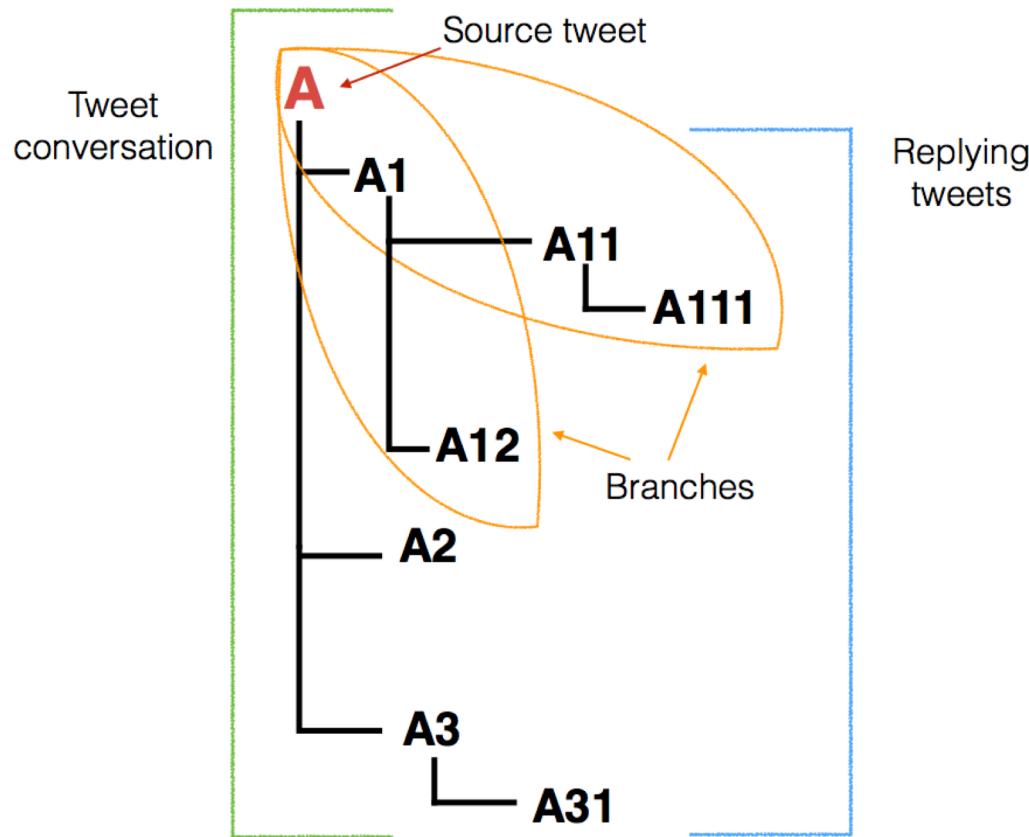
- **'hints' in shared representation**
- **increased training data**
- **regularisation**

Usually **auxiliary tasks** are **low-level tasks** like POS tagging, Chunking, Log frequency of each word.

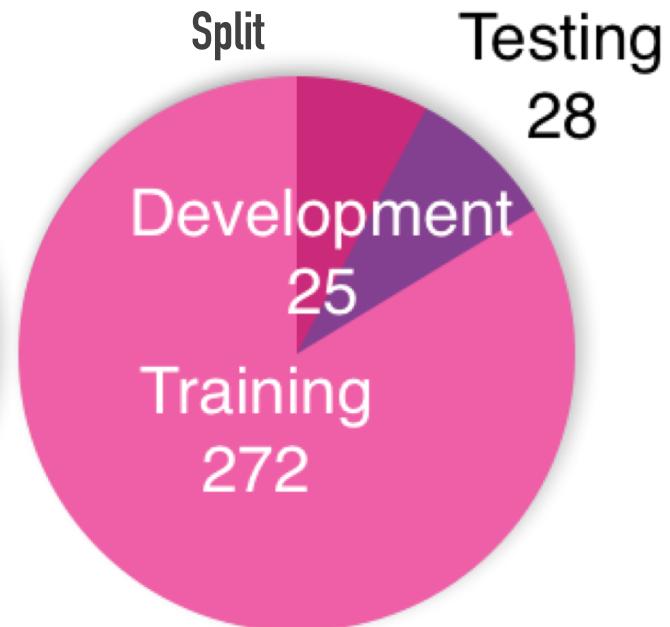
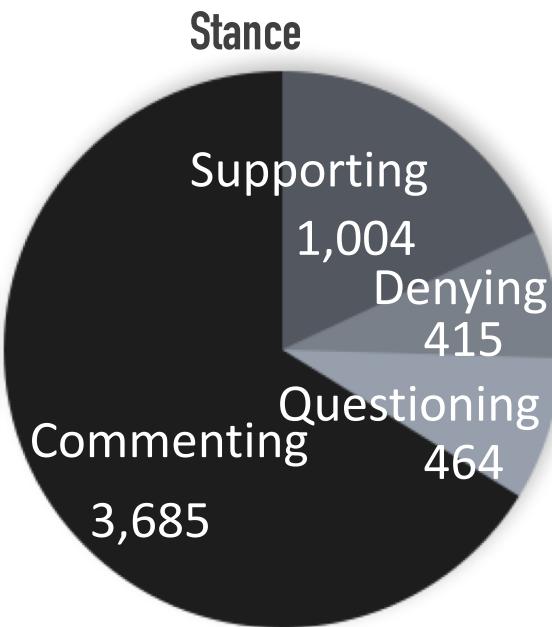
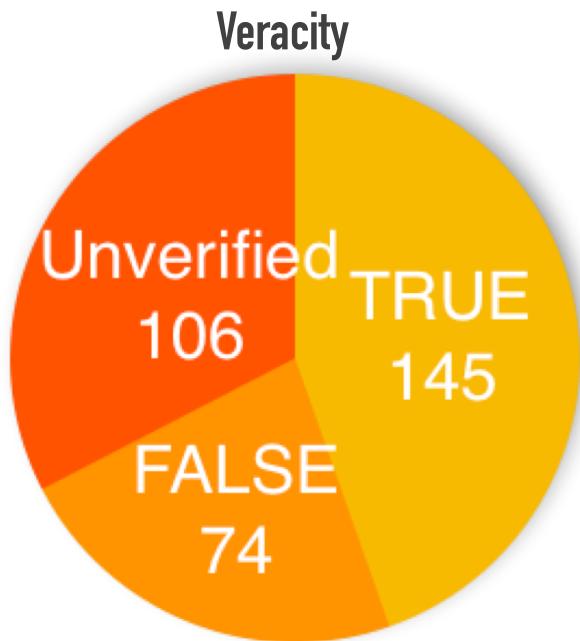
# Conversation example



# Conversational structure



# Datasets. RumourEval

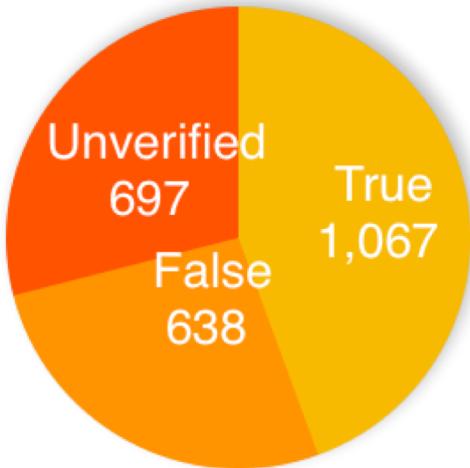


# Datasets. PHEME

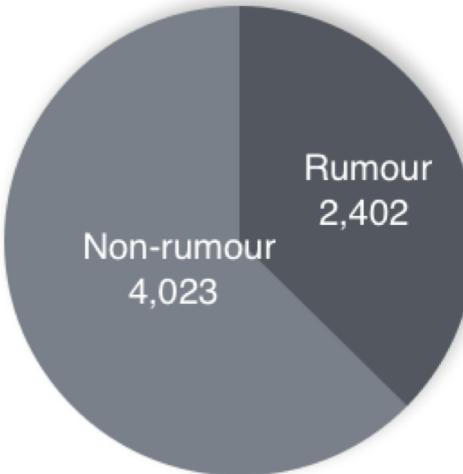


WARWICK

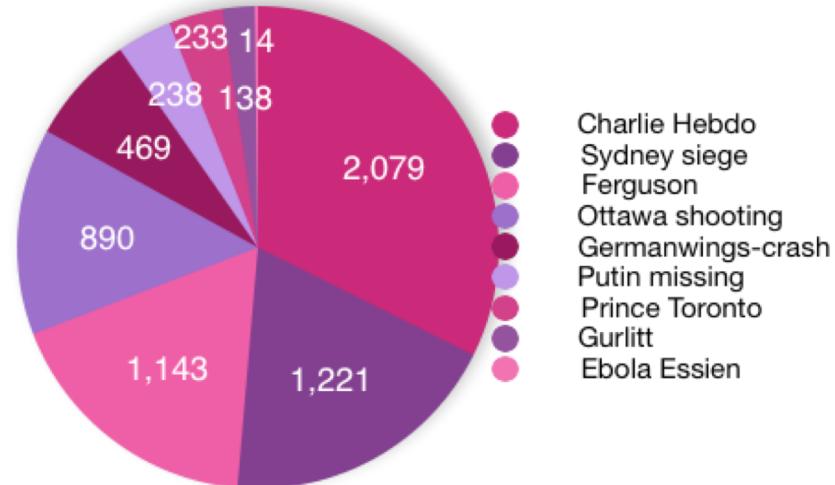
Veracity



Detection



Events



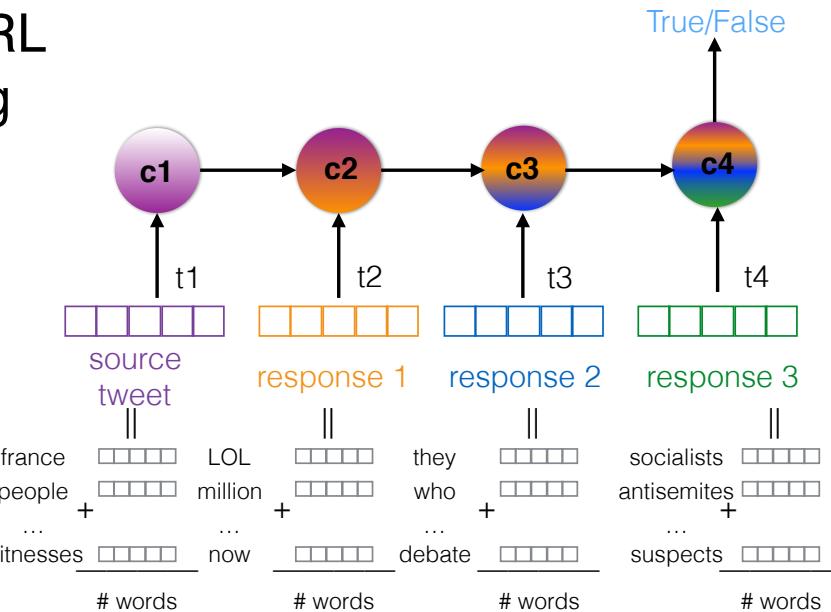
# Baselines

- **Majority** - always predicts majority class (True).
- **NileTMRG** (Enayet and El-Beltagy, 2017)

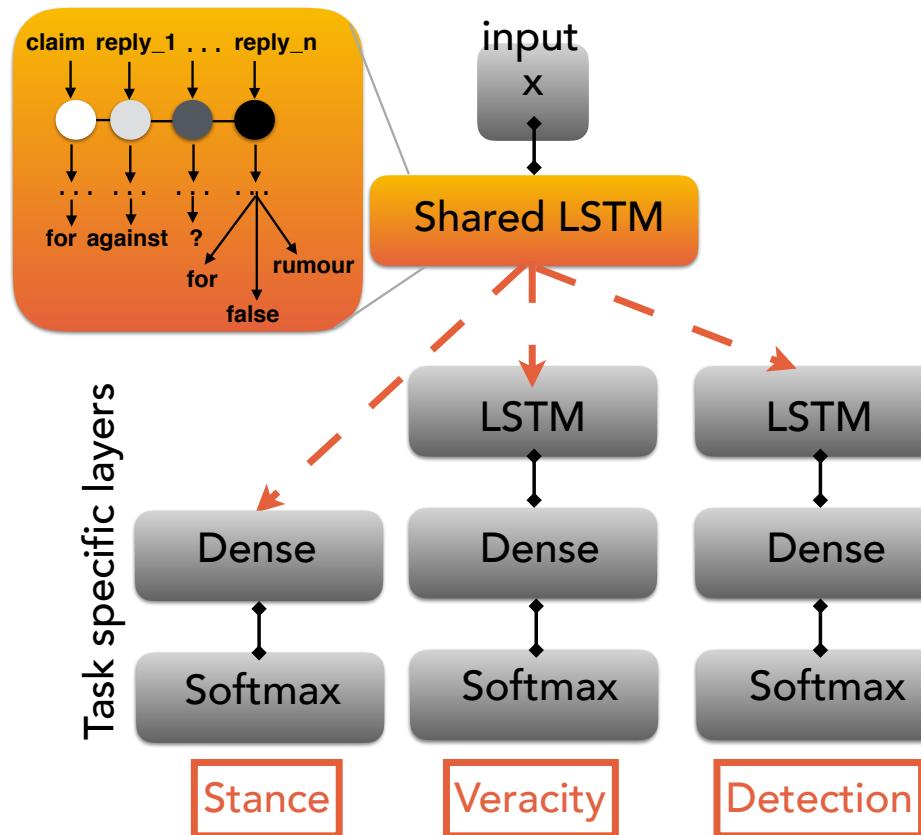
Linear SVM with BoW, presence of #, URL and proportion of supporting, questioning and denying tweets

- **branchLSTM**

This is a single task model that utilises conversation structure. branchLSTM is a basis of MTL models.



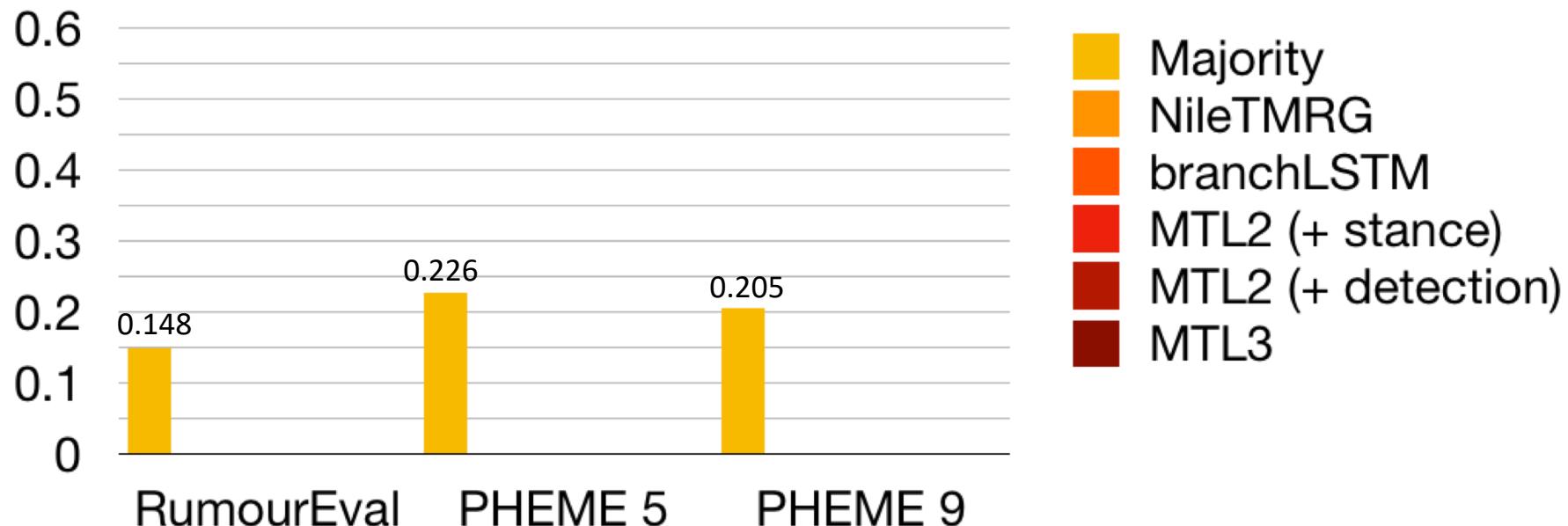
# Multi-task models



# Results



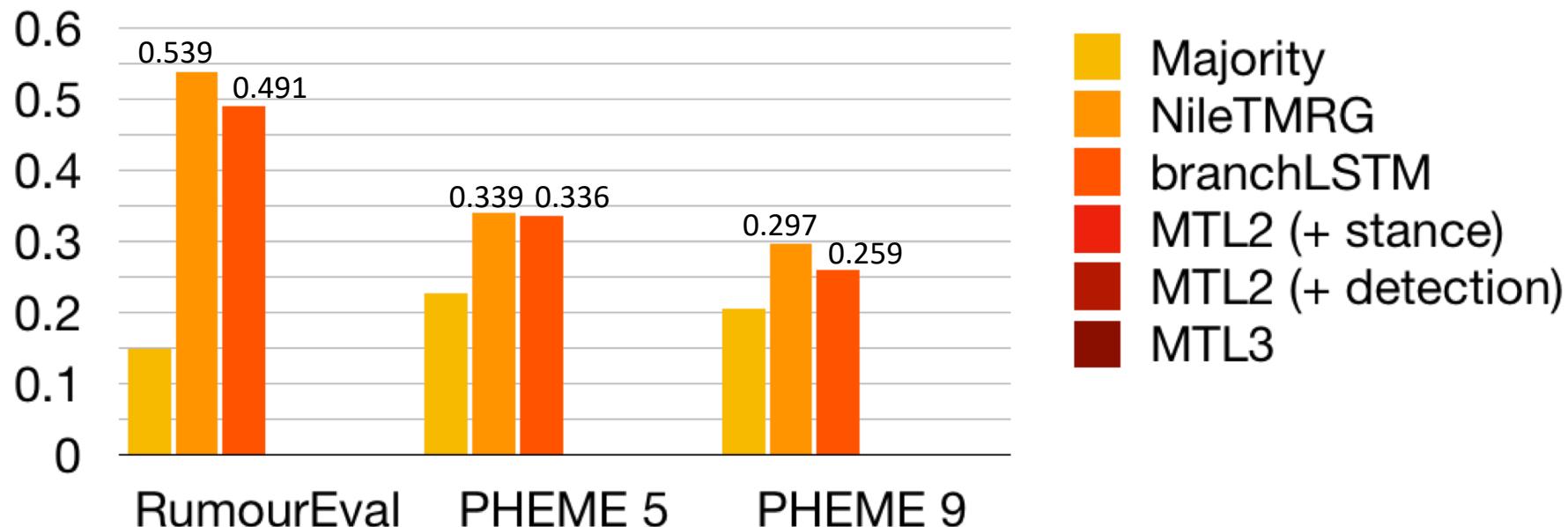
## Macro F1 score



# Results



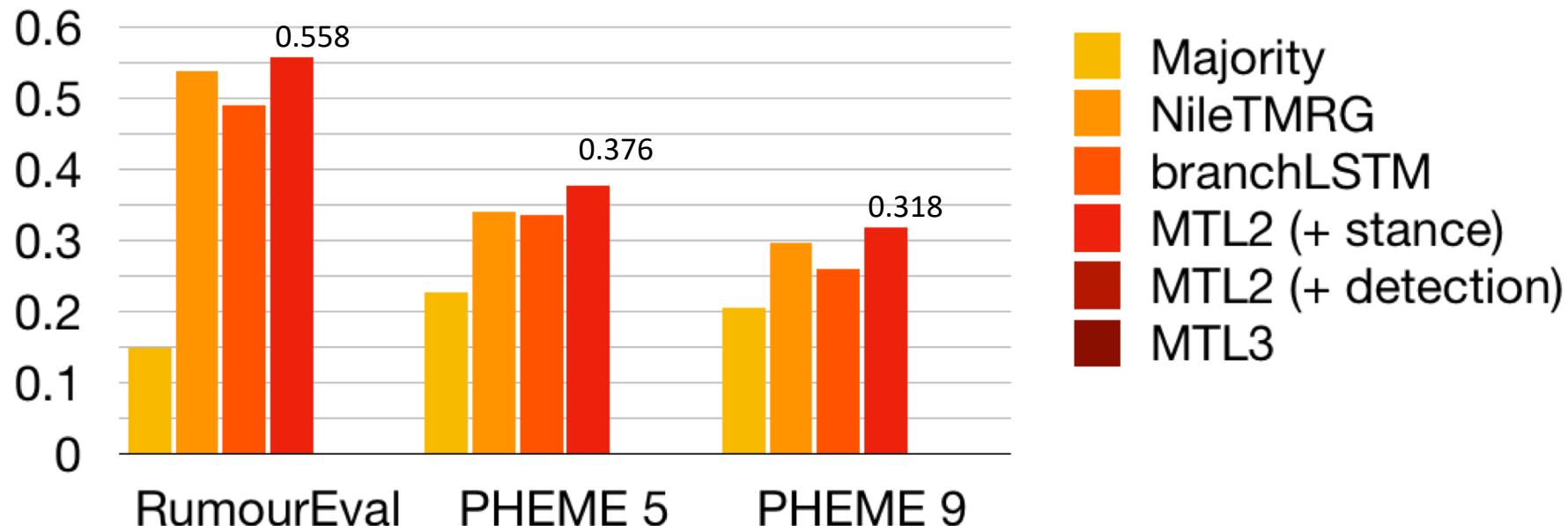
## Macro F1 score



# Results



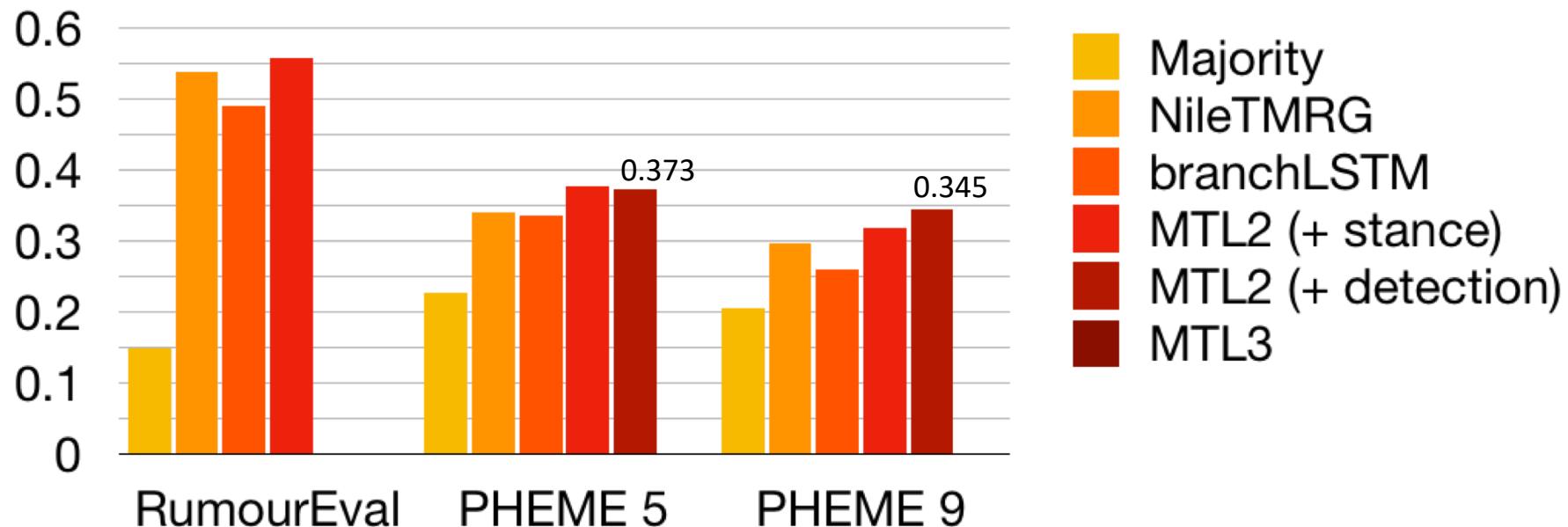
## Macro F1 score



# Results

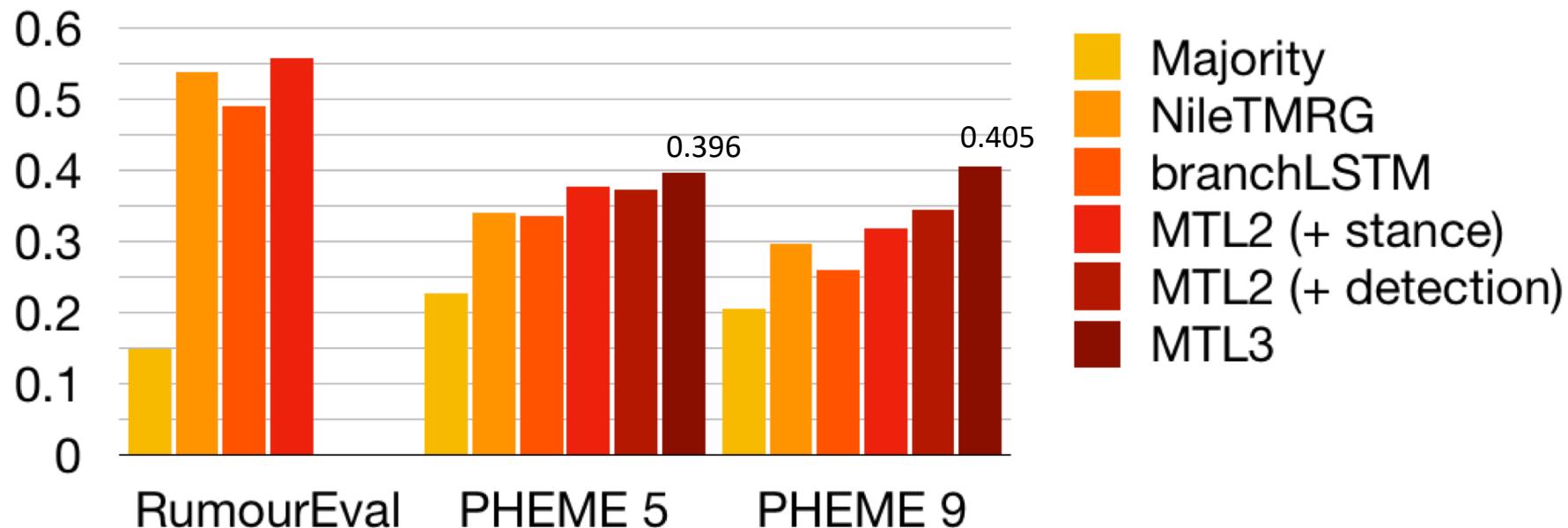


## Macro F1 score



# Results

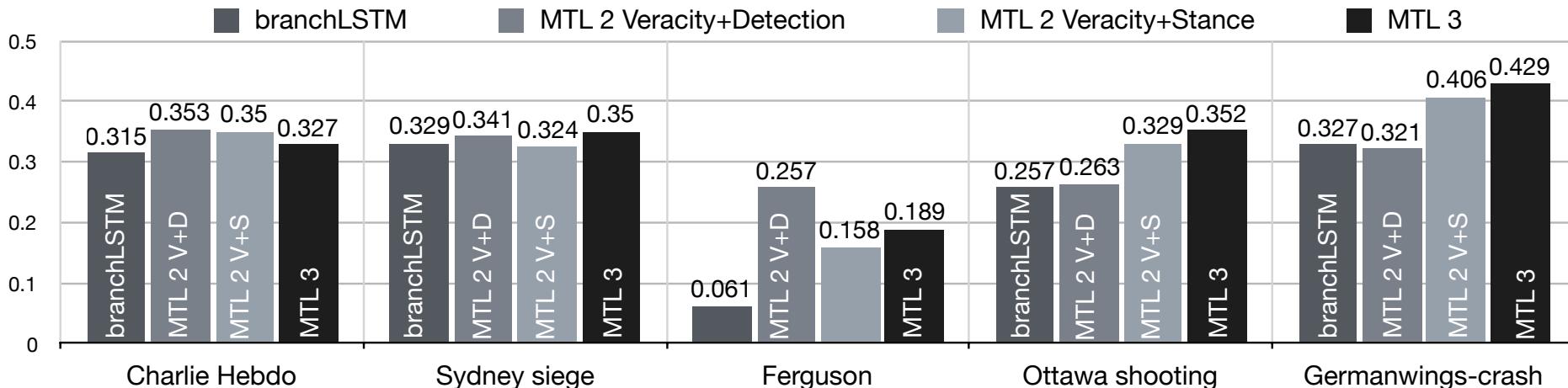
## Macro F1 score



# Results. Per-event and per-class results analysis

WARWICK

	MacroF	Accuracy	TRUE	FALSE	Unverified
<b>Charlie Hebdo</b>	0.327	0.369	0.502	0.227	0.251
<b>Sydney siege</b>	0.35	0.575	0.731	0.153	0.168
<b>Ferguson</b>	0.189	0.338	0.058	0	0.508
<b>Ottawa shooting</b>	0.352	0.645	0.789	0.168	0.1
<b>Germanwings-crash</b>	0.429	0.42	0.538	0.358	0.364



# Conclusions

The joint learning of two tasks from the verification pipeline outperforms a single-learning approach to rumour verification. The combination of all three tasks leads to further performance improvements.

# Future work

- Scheduling
- Task hierarchy
- Extra features





Elena Kochkina



E.Kochkina



@Elena\_Kochkina



Maria Liakata



M.Liakata



@xrysoflhs



Arkaitz Zubiaga



A.Zubiaga



@arkaitz

The  
Alan Turing  
Institute



---

The Leverhulme Trust

[@warwick.ac.uk](https://warwick.ac.uk)

Code available: <https://github.com/kochkinaelena/Multitask4Veracity>

WARWICK

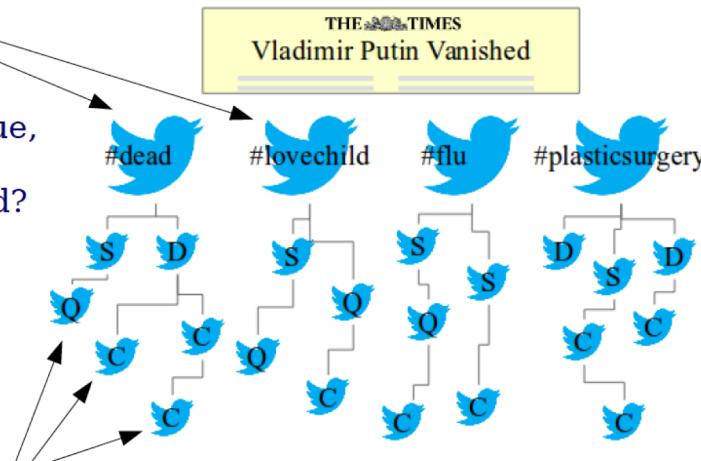
# SemEval-2019

International Workshop on Semantic Evaluation

## Fact vs fiction

### Task 7: RumourEval 2019: Determining Rumour Veracity and Support for Rumours

**Task B:**  
Classify  
rumour  
posts. True,  
false or  
unverified?



**Task A:** Classify discussion posts.  
Support, deny, query or comment?

### What has changed since 2017?

- New rumours added
- Reddit data as well as Twitter
- New languages added
- Baselines Available

### Organizers:

- Genevieve Gorrell
- Ahmet Aker
- Leon Derczynski
- Elena Kochkina
- Maria Liakata
- Arkaitz Zubiaga
- Kalina Bontcheva