

Predicting and Visualizing Short-Term Change in Word Representation and Usage in VKontakte Social Network



Pacific Northwest
NATIONAL LABORATORY

Ian Stewart¹, Dustin Arendt², Eric Bell², and Svitlana Volkova²

¹Georgia Institute of Technology, ²Pacific Northwest National Laboratory

Proudly Operated by Battelle Since 1965

Motivation

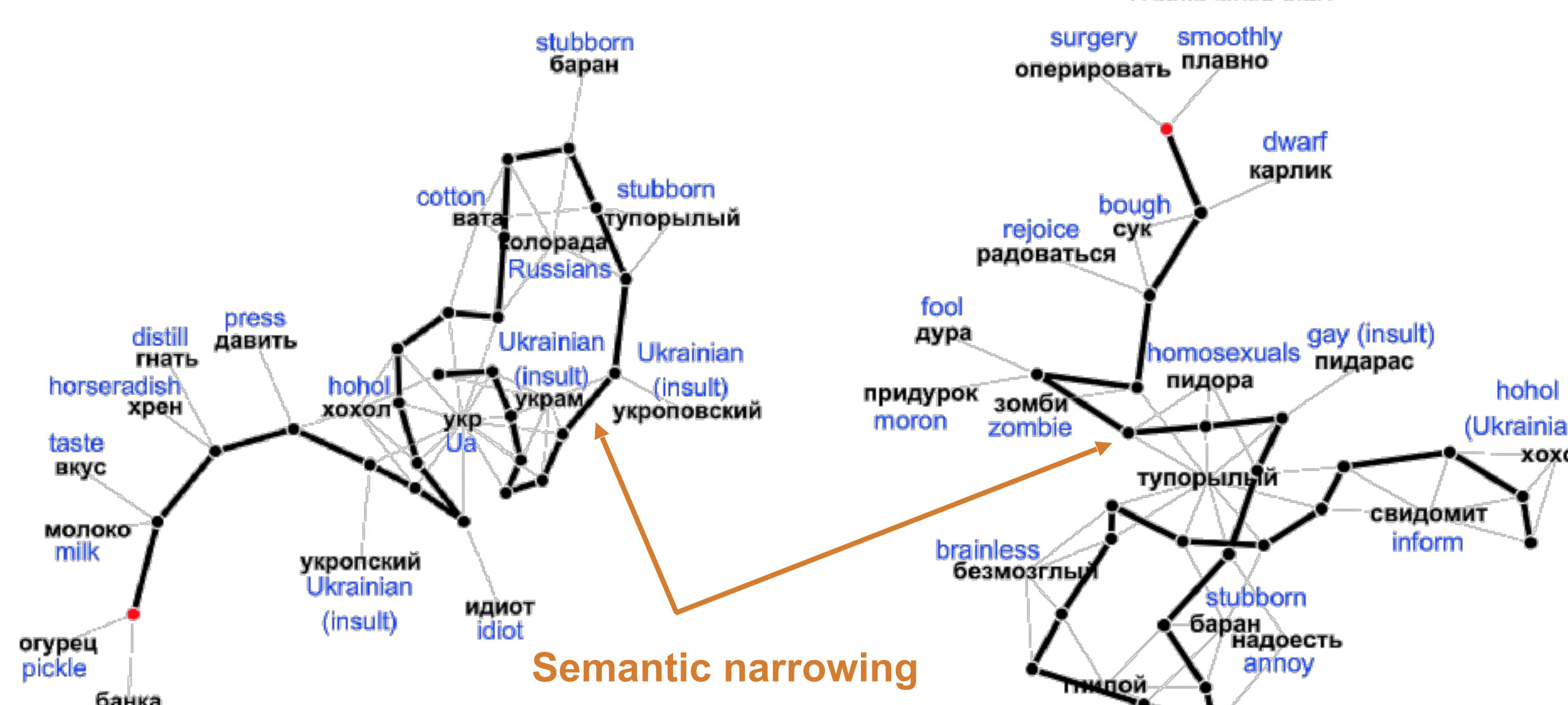
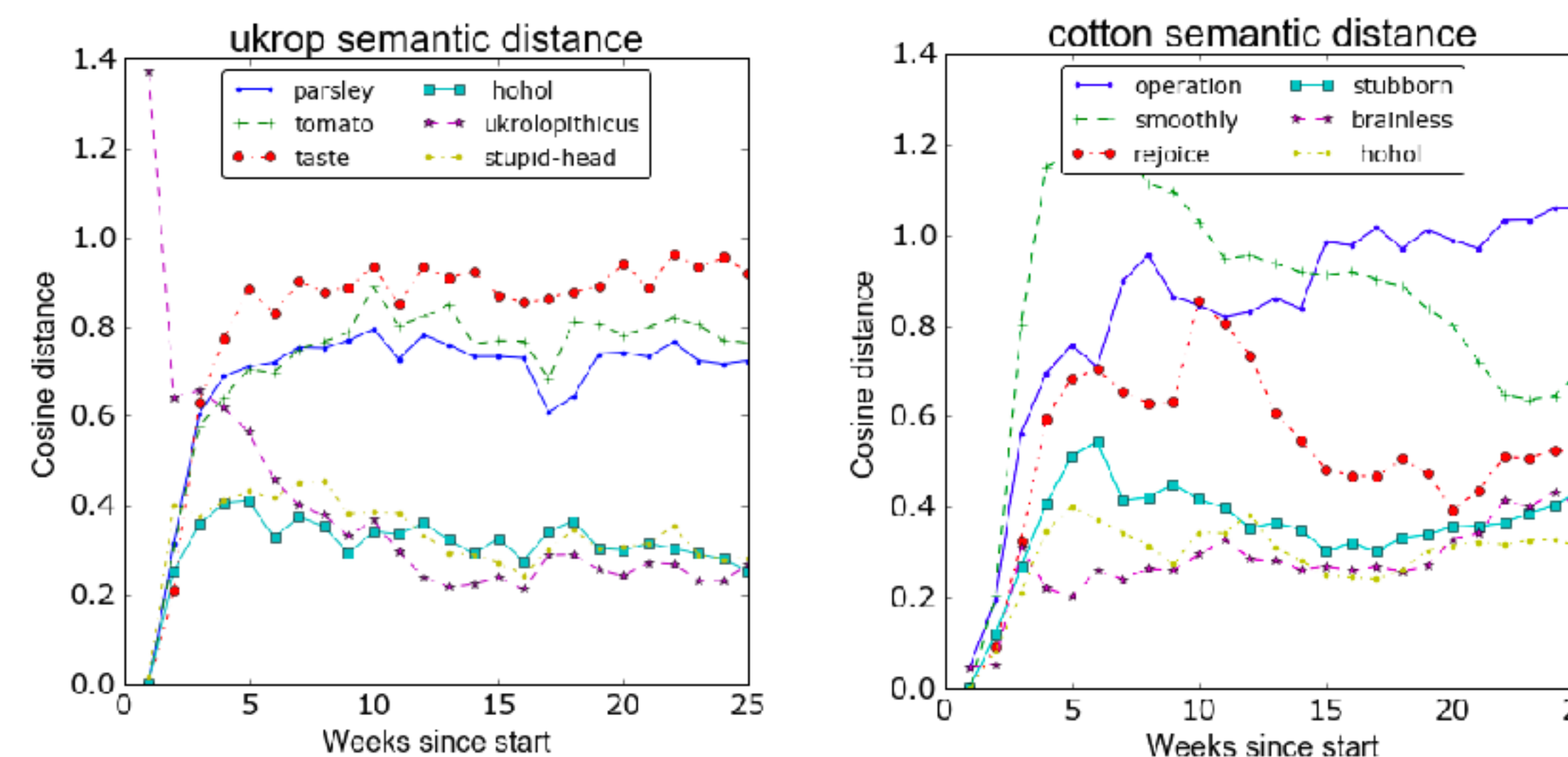
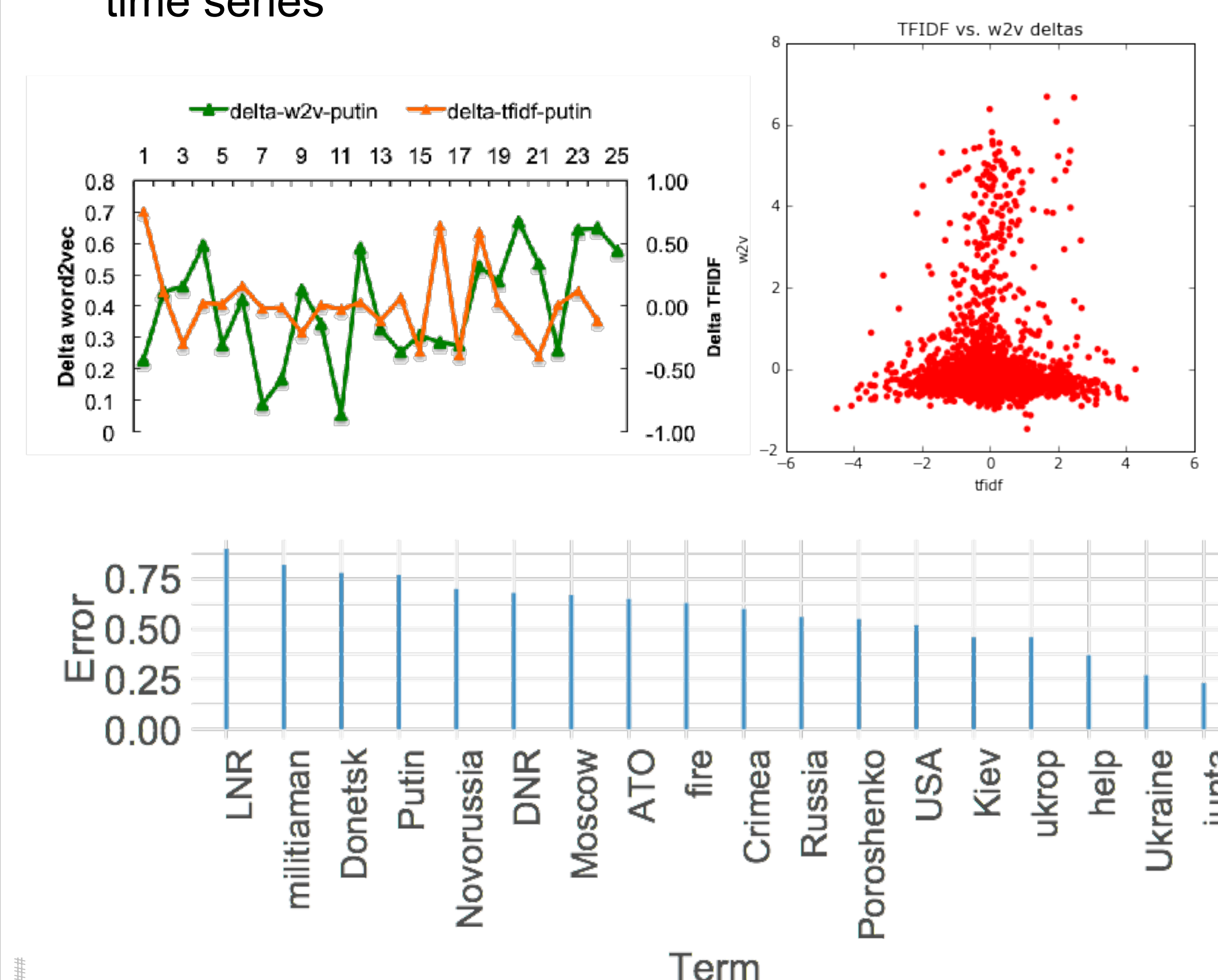
- ❑ Crisis can lead to **language change**, such as when neutral words become politically charged
- ❑ укроп <ukrop> “dill” → “Ukrainian (insult)”
- ❑ Analysts benefit from both **usage** (frequency) and **representation** (semantic) dynamics
- ❑ Beyond **dynamic topic models** and **trending topics**

Contributions

- I. Develop predictive models and show that representation shift can be inferred from prior shift and concept drift
- II. Relate measures of concept drift and representation shift to track language dynamics in social media

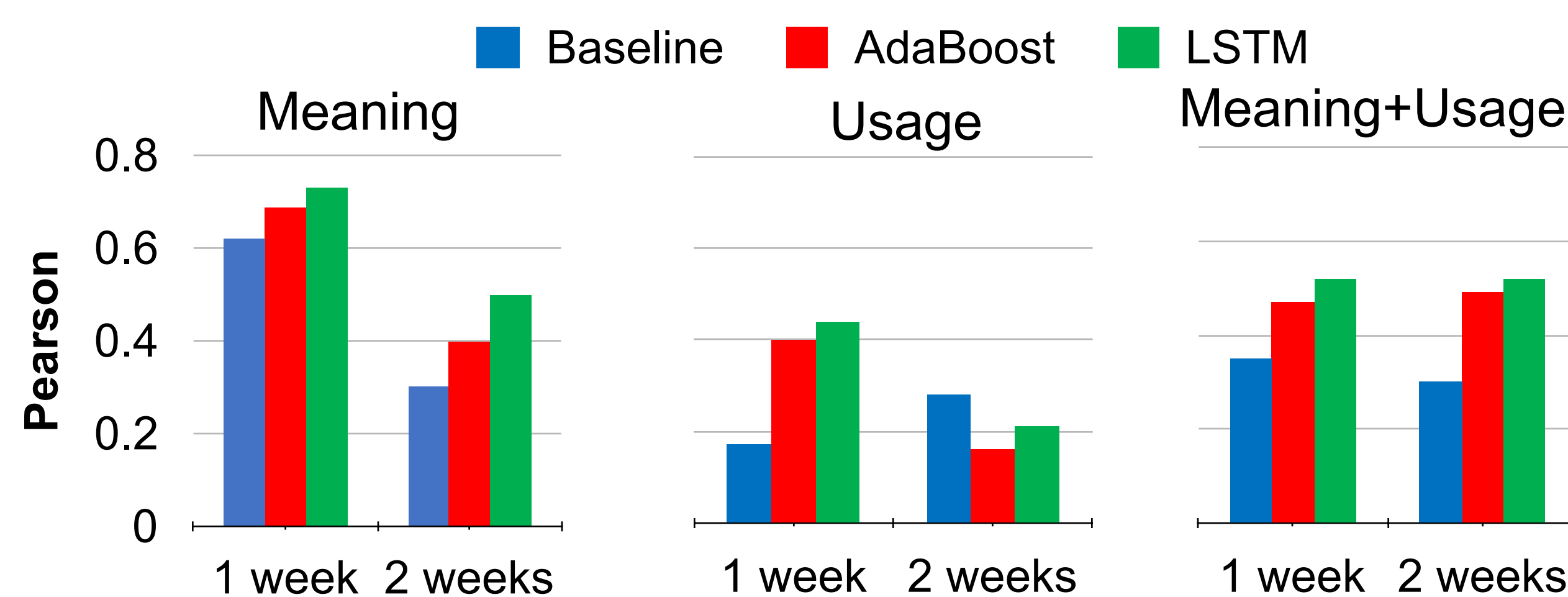
VKontakte Data

- ❑ 600,000 posts, 167 words/post
- ❑ 25 weeks, 09/2014 – 03/2015
- ❑ Usage (TFIDF) and representations (word2vec) weekly time series



Prediction Experiments

- ❑ Task: predict representation shift from (1) meaning shift, (2) concept drift, (3) concept drift + meaning shift
- ❑ Regression with LSTMs, AdaBoost, Baseline
- ❑ Evaluation: Pearson, Root Mean Squared Error



RMSE when predicting representation shift (RS)

	Meaning → RS		Usage → RS		Both → RS	
	1 week	2 weeks	1 week	2 weeks	1 week	2 weeks
Mean	5.0	5.6	47.4	59.5	5.2	5.4
AdaBoost	5.7	6.6	7.3	7.9	5.2	6.4
LSTM	4.2	3.9	5.6	4.4	3.2	4.3

Impact

- ❑ Track and visualize unexpected **changes in language** in social media during crisis events
- ❑ Isolate linguistic units with **dynamic contexts**
- ❑ Predict **evolution of semantic concepts** in a stream of text over time
- ❑ Improve real-time **stream summarization** and **event detection**

ABOUT

Pacific Northwest
National Laboratory

The Pacific Northwest National Laboratory, located in southeastern Washington State, is a U.S. Department of Energy Office of Science laboratory that solves complex problems in energy, national security, and the environment, and advances scientific frontiers in the chemical, biological, materials, environmental, and computational sciences. The Laboratory employs nearly 5,000 staff members, has an annual budget in excess of \$1 billion, and has been managed by Ohio-based Battelle since 1965.

For more information on the science you see here, please contact:

Dr. Svitlana Volkova
Data Science and Analytics
National Security Directorate
Richland, WA 99352
(509) 372-6585
svitlana.volkova@pnnl.gov
<https://www.cs.jhu.edu/~svitlana/>

Extended paper

arXiv.org > cs > arXiv:1703.07012