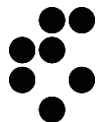


An approach to creating a time-series dataset for news propagation: Ukraine-war case study

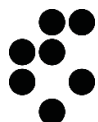
SiKDD October 9th, 2023

Abdul Sittar, Dunja Mladenic



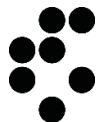
Motivations

- Classification and forecasting of barriers to news spreading
 - LLMs
 - Inaccurate or misleading results
 - Optimized for applications
 - Reporting time: impact, context, public safety
 - Barrier definition: geographical location - context
 - Cascading structure: narratives building and story telling
 - Classical, deep learning, transformers



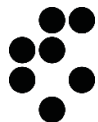
Outline

- Introduction
 - Barriers to news spreading
 - Topic modeling
 - Time-series dataset
- Approach
- Time-series creation
- Annotations of time-series
- Statistical analysis and evaluations
- Conclusions and Future work



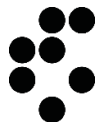
Introduction

- Events – Event-Registry
 - Reasons based on cultural, economic, political, linguistic, or geographical reasons
 - FIFA World cup
 - Sri-Lankan economic crisis and Ukraine-war
- News spreading patterns
 - Geographical barrier
 - Topic modeling

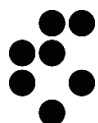
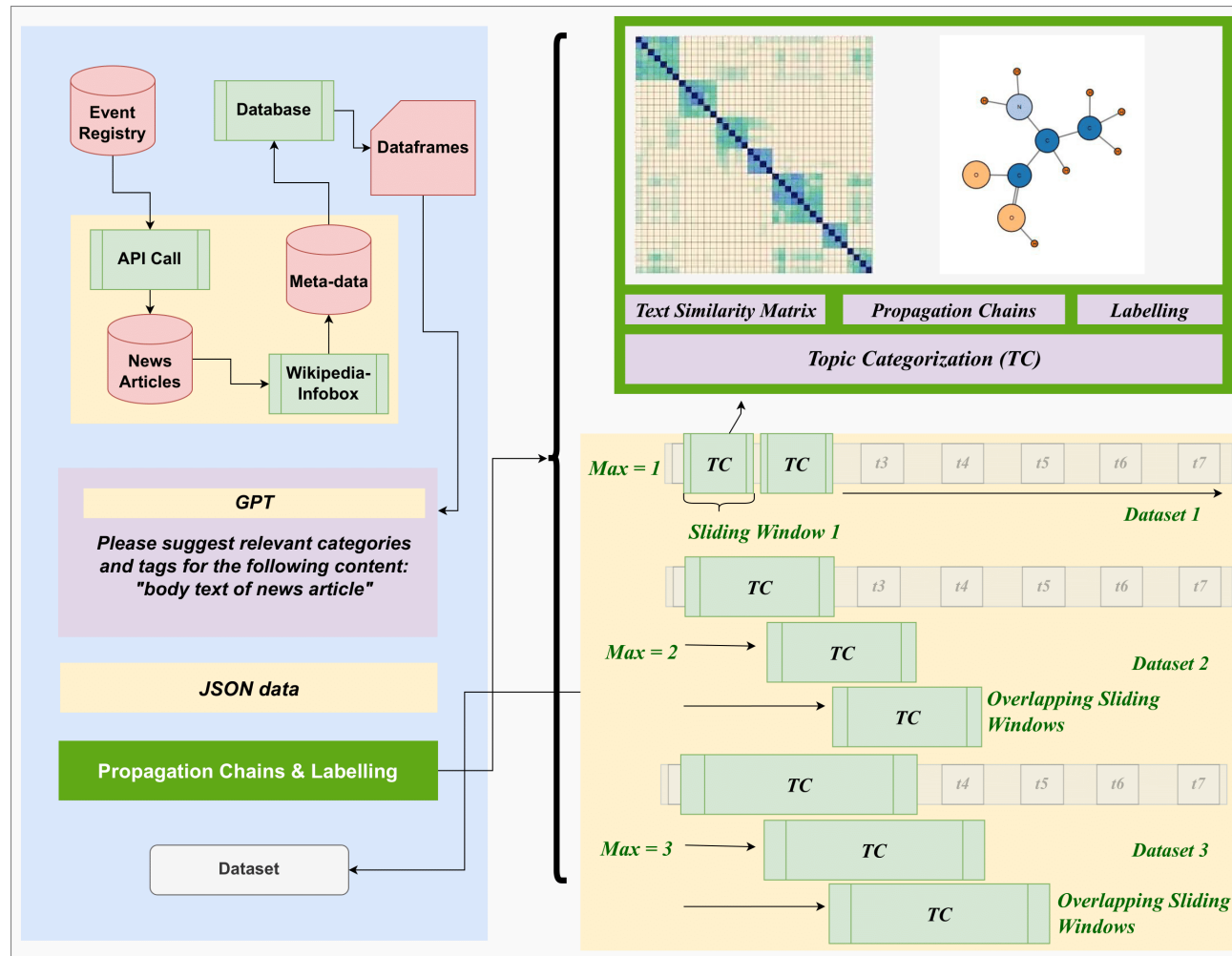


Introduction

- Time-series
 - News propagation
 - Size and time of cascading
- Time-series datasets
 - Prediction stock market values
 - Forecast price trends immediately after press release publication
- Methodological considerations are necessary

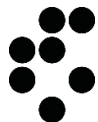


Approach

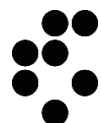
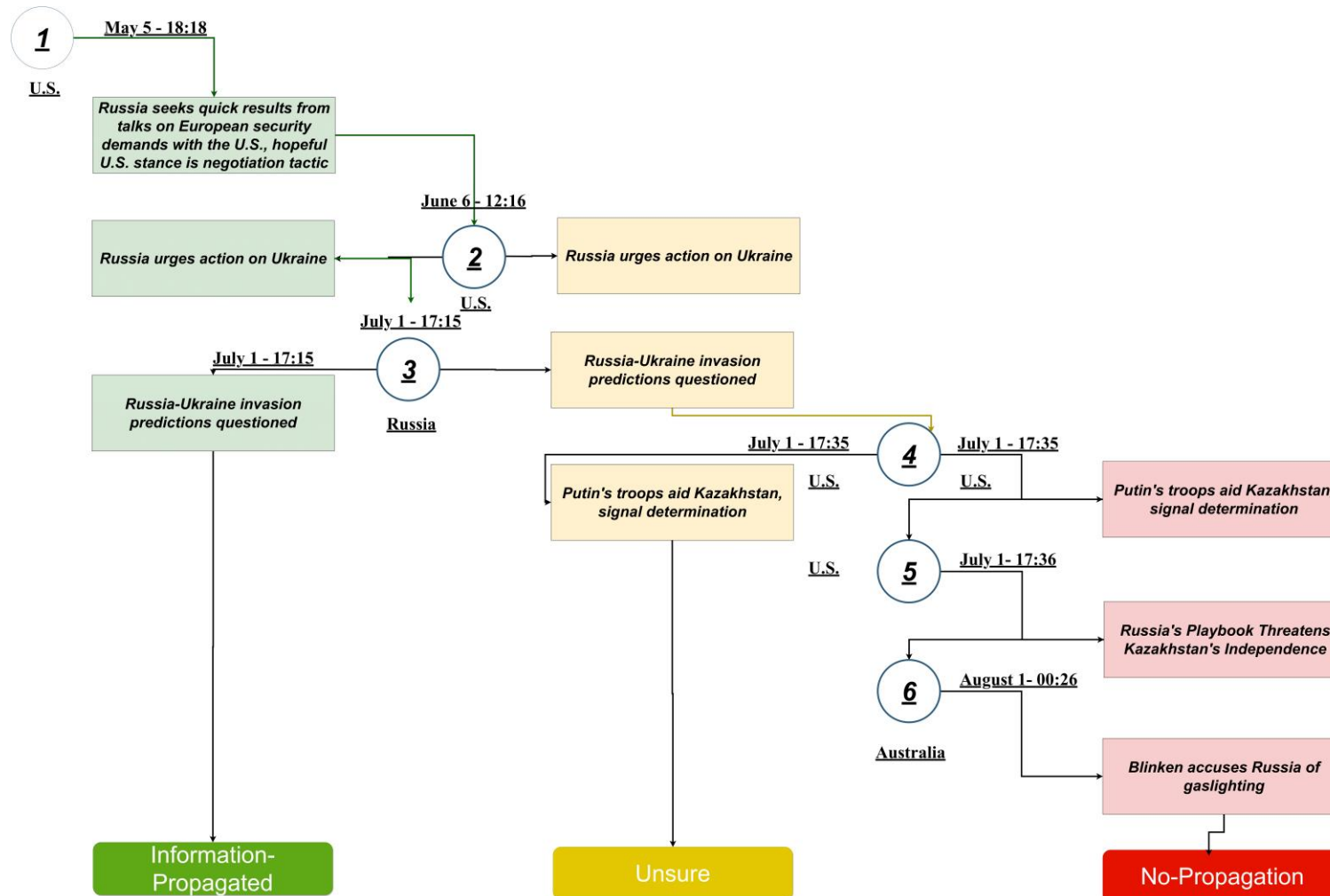


Annotations of time-series

- Semantic similarity
 - Sentence transformers
- Chat-GPT Summarization
 - Tags, categories, and summary of news articles
- Types of time-series
 - Propagating, not-propagating, unsure

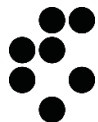


Example

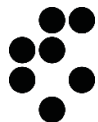
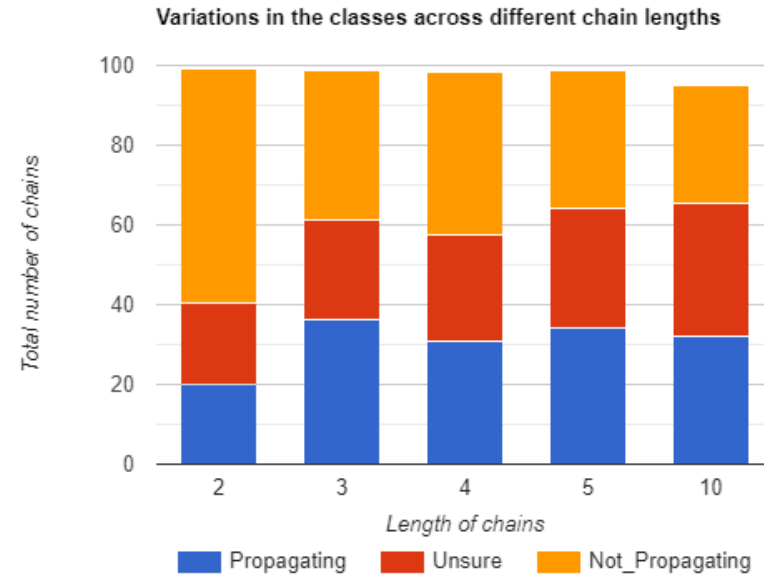
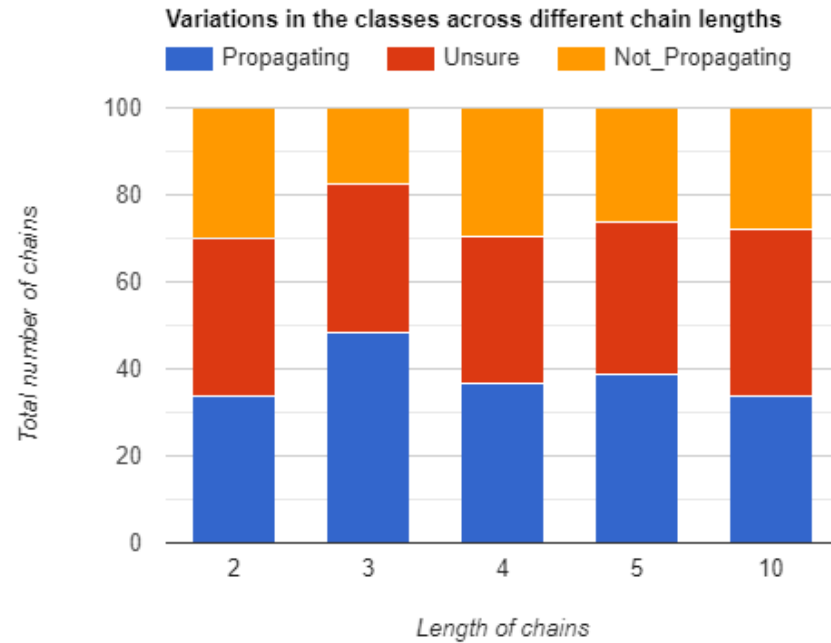


Evaluations and Statistical analysis

- 50 time-series of different lengths for all three labels
 - Propagating – one or two themes of discussion
 - Not-Propagating – Different point of views
 - Unsure – three or four sub-topics
- Effect of window size
 - The noise of overlapping topics also increased



Evaluations and Statistical analysis



Conclusions and Future work

- An approach to creating a time-series
- Same approach for different events
- Extend geographical barrier with political, economic, and cultural
- Prediction and forecasting
 - Classical methods, deep learning, transformer-based methods, and large language models (LLMs)

