Learning Dynamic Context Graphs for Predicting Social Events

Songgaojun Deng,
Huzefa Rangwala,
Yue Ning

Department of Computer Science
Songgaojun Deng
sdeng4@stevens.edu

Learning Dynamic Context Graphs for Predicting Social Events

Songgaojun Deng
Stevens Institute of Technology
Hoboken, New Jersey
sdeng4@stevens.edu

Huzefa Rangwala
George Mason University
Fairfax, Virginia
rangwala@cs.gmu.edu

Yue Ning
Stevens Institute of Technology
Hoboken, New Jersey
yue.ning@stevens.edu

Songgaojun Deng
6? August 2019

KDD 2019

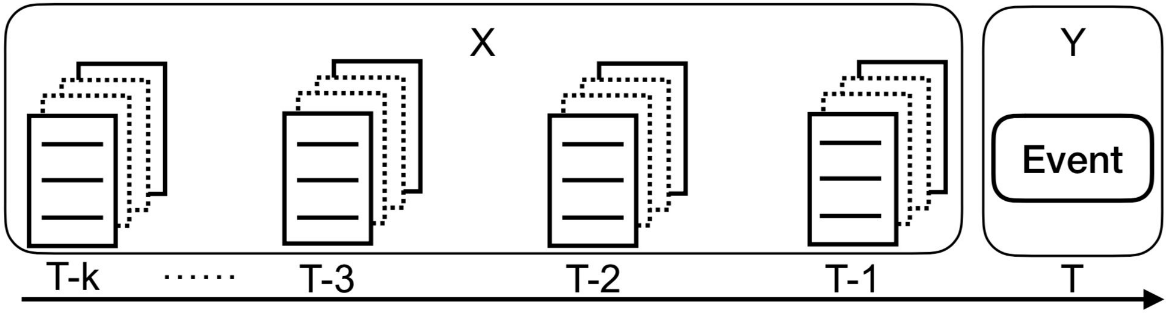
Outline

- Motivation and Background
- Graph Convolutional Network
- *DynamicGCN* for event forecasting
- Experiment Evaluation
- Dynamic Context Graphs
- Conclusion

Motivation

- Forecasting social events such as civil unrest movements is an important and challenging problem

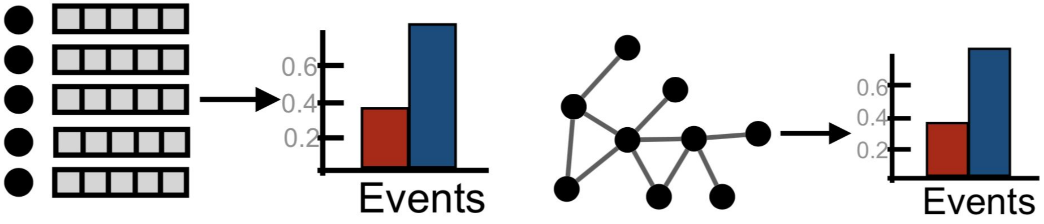
- From the perspective of human analysts and policy makers, forecasting algorithms should
  - not only make accurate predictions
  - but also provide supporting evidence/clue
    - e.g., the causal factors related to the event



Motivation

- **Model contextual information** for event forecasting

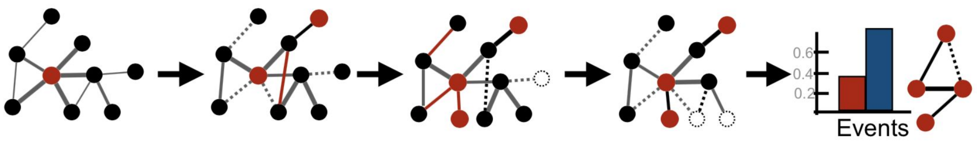
- Challenges
  - uncertainty of context structure and formulation
  - high dimensional features
  - adaptation of features over time



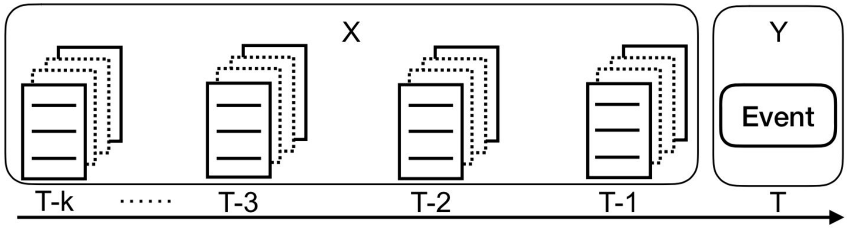
(a) Non-Graph Features (b) Static Graph

Contributions

- Develop a novel graph-based model for predicting events
  - Dynamic graph structures over time are discovered for understanding the events
- Propose a temporal encoded feature module to adapt to the dynamic features over time.
  - semantic embeddings ⊕ graph embeddings



(c) Dynamic Context Graph



(d) Raw Input Data

Graph Convolutional Network

Main idea: Pass messages between pairs of nodes

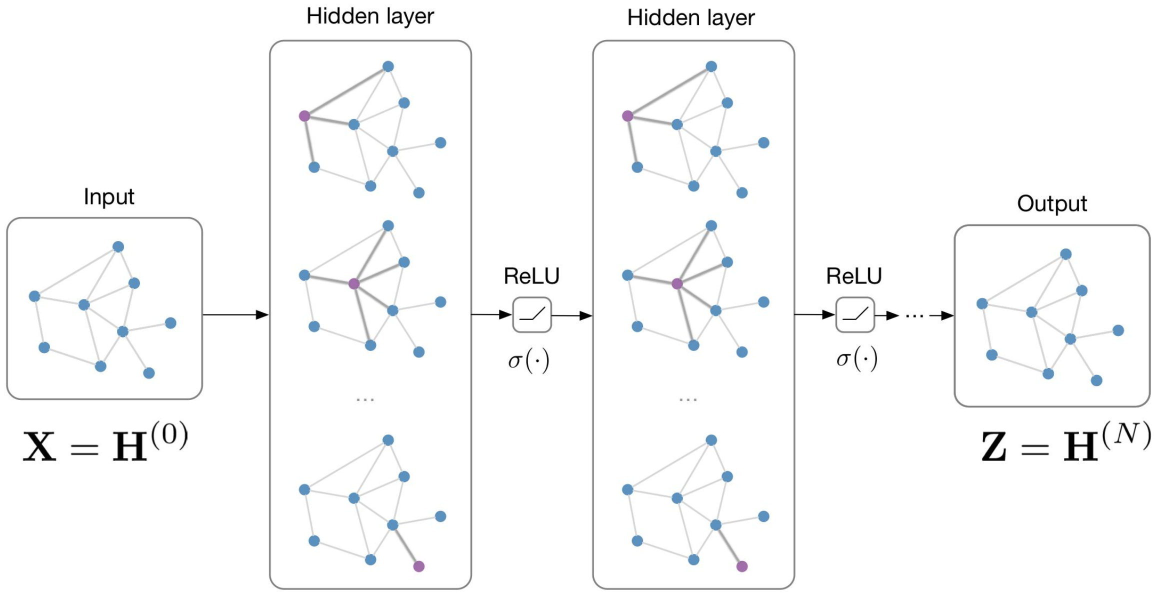
Graph: $G = (\mathcal{V}, \mathcal{E})$

$\mathcal{V}$: Set of nodes $\{v_i\}$, $|\mathcal{V}| = N$

$\mathcal{E}$: Set of edges $\{(v_i, v_j)\}$

Notation: $\mathcal{G} = (\mathbf{A}, \mathbf{X})$

- Adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$
- Feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$



$\mathbf{X} = \mathbf{H}^{(0)}$ $\quad$ $\mathbf{Z} = \mathbf{H}^{(N)}$

$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

Problem formulation

- Data
  - Given a city c, we collect published news articles for k consecutive days prior to a date t as the raw input $x\_(c, t)$ = {docs in days $t-1 \rightarrow t-k$}
  - If a target event occurs on day t, $y\_(c,t)$= 1
  - Otherwise, if there is no event on the day t as well as the previous three days, $y\_(c,t)$= 0

- Key objectives

$$X \xrightarrow{\text{encode}} \text{Graphs} \xrightarrow{\text{model}} y$$

Methodology

- Encoding
  - Dynamic Graph
  - Feature Representation
- Model Framework
  - Input Layer
  - Dynamic GCN Based Network Encoding
  - Temporal Encoded Features (TE)
  - Masked Nonlinear Transformation Layer

Encoding

- Dynamic Graph
  - Define a sequence of adjacency matrix $[A_{t-k}, \ldots, A_{t-1}]$

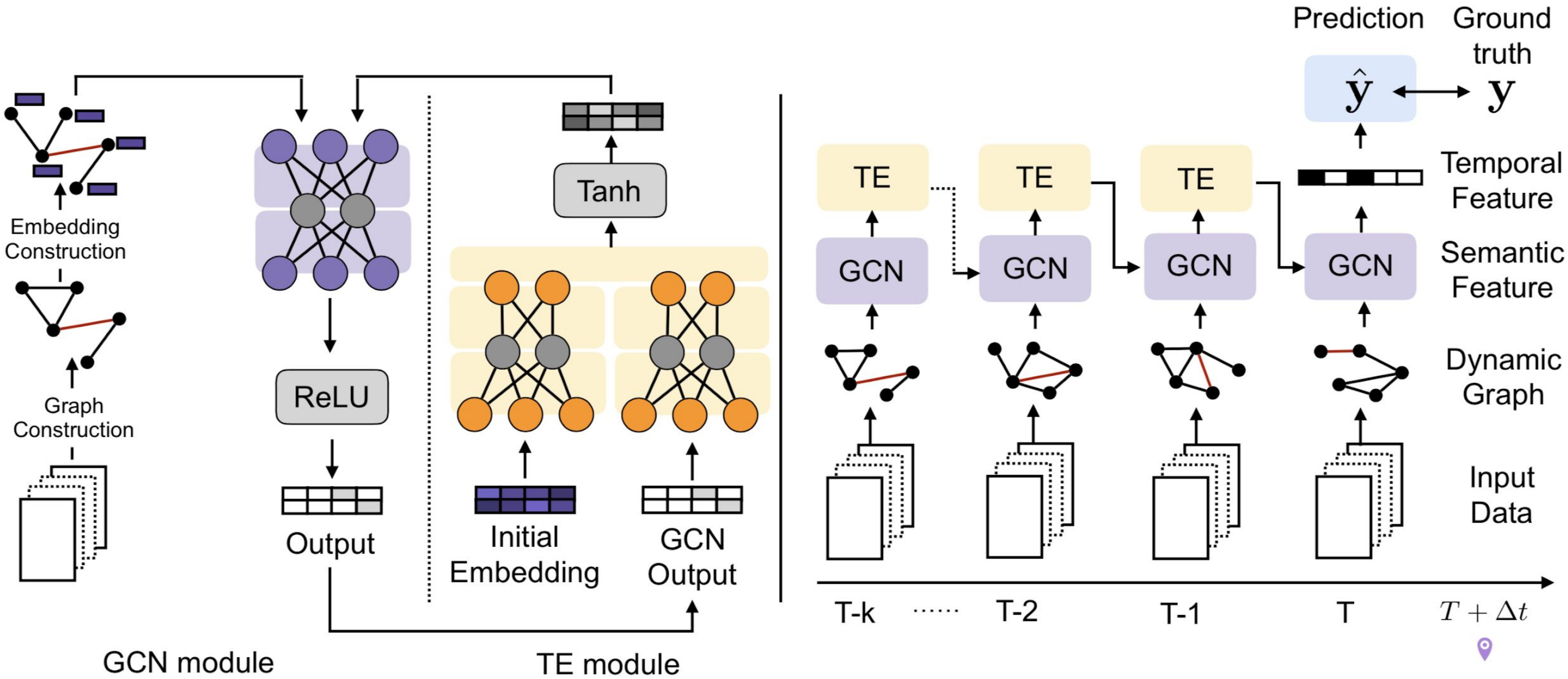
$$A_t[i,j] = \begin{cases} \mathrm{PMI}_t(i,j) & \mathrm{PMI}_t(i,j) > 0 \\ 0 & \text{otherwise} \end{cases}$$

  - Document-based pointwise mutual information (PMI)

$$\mathrm{PMI}_t(i,j) = \log \frac{d(i,j)}{d(i)d(j)/D}$$

- Feature Representation
  - Word embedding (encode syntactic and semantic info)

Model Framework



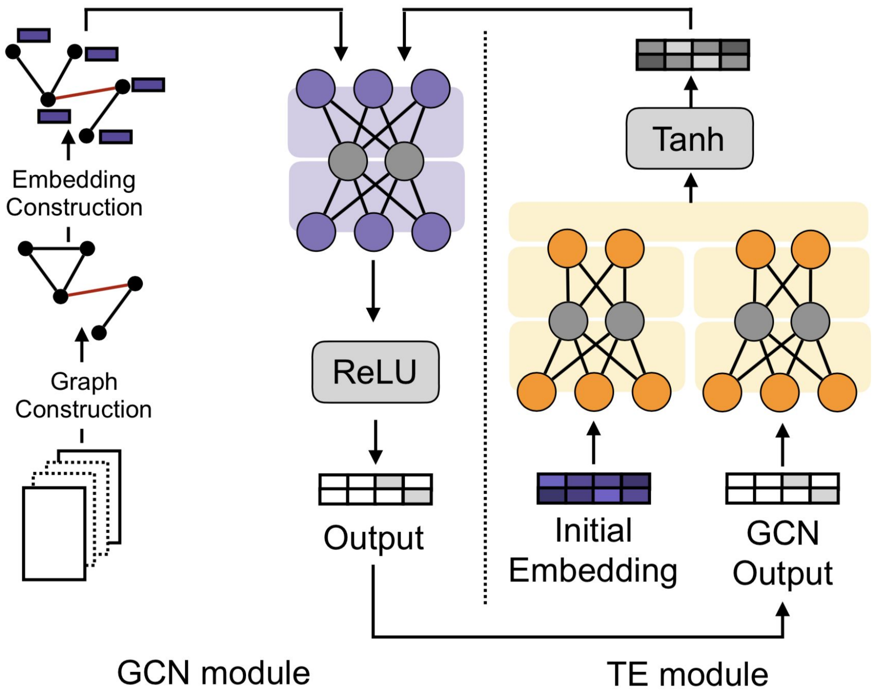
Temporal Encoding Module

- Temporal encoded features

$$H_p^{(t)} = H_{t-1} W_p^{(t-1)} + b_p^{(t-1)}$$
$$H_e^{(t)} = H_0 W_e^{(t-1)} + b_e^{(t-1)}$$
$$\tilde{H}_t = \tanh([H_p^{(t)} \parallel H_e^{(t)}])$$

- Graph convolution based on dynamic graphs and features

$$H_{t+1} = g(\hat{A}_t \tilde{H}_t W^{(t)} + b^{(t)})$$



Temporal dependency is passed through each GCN layer to the graph nodes

Optimization

- Binary cross entropy

$$\mathcal{L} = -\sum y \ln \hat{y}$$

- All model parameters can be trained via back-propagation and optimized using the Adam optimization algorithm

Datasets and Metrics

- Data (English news articles)
  - India
  - Egypt
  - Thailand
  - Russia

Table 1: Dataset Statistics.

data set	#news	#vocabulary	#sample	#pos	#neg
India	111,653	75,994	12,249	4,586	7,663
Egypt	30,867	19,680	3,788	1,469	2,319
Thailand	19,410	27,281	1,883	715	1,168
Russia	85,527	49,776	3,552	1,171	2,381

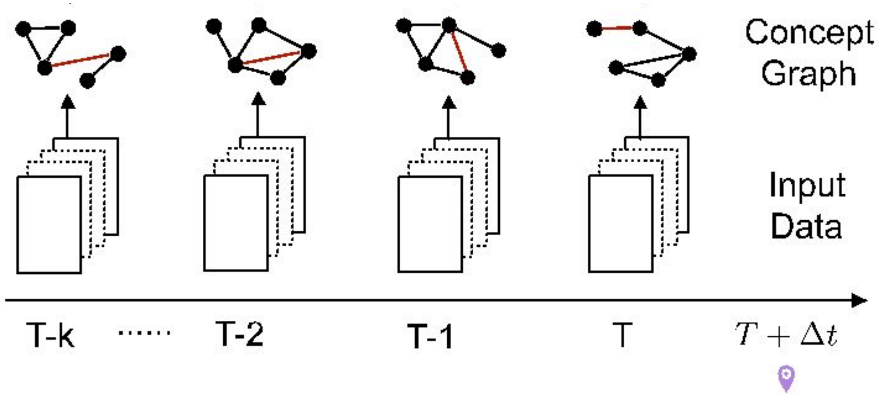
- Evaluation
  - Prediction performance
    - Precision (Prec.)
    - Recall (Rec.)
    - F1-Score (F1)
  - Dynamic Context Graphs

Model Comparison

- Logistic Regression (LR)
- Nested Multi-Instance Learning (nMIL) (Ning, etc. KDD16)
  - A hierarchical multi-instance learning framework for forecasting events and identifying historical documents as event precursors. It uses document embeddings.
- Graph Convolutional Network (GCN) (Kipf, etc. ICML17)
- GCN+GRU (T-GCN) (Zhao, etc. 2018)
  - It combines GCN and GRU to capture spatio-temporal correlations in traffic data.
  - GCN+LSTM
  - GCN+RNN

Experiments - Data Processing

- Data preprocessing (for each country)
  - Preprocess all the news data by cleaning and tokenizing words
  - Remove stop words and keep only stemmed words
  - Train word embeddings
- Construct raw input data
  - Collect news in days $t-1 \rightarrow t-k$, respectively
- Construct training samples (for each raw input data)
  - Extract top keywords based on news in k days
  - Build dynamic graphs over time



Experiments

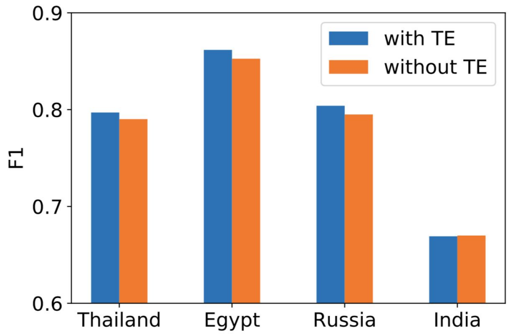
Performance comparison on four data sets

Table 2: Performance comparison on test set. (average over 20 trials)

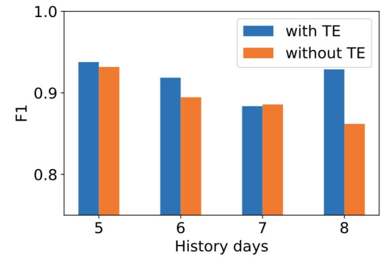
		Thailand			Egypt			India			Russia		
		F1	Rec.	Prec.	F1	Rec.	Prec.	F1	Rec.	Prec.	F1	Rec.	Prec.
Non-Temp.	LR (Count)	0.7701	0.7129	**0.8372**	0.7945	0.7468	0.8488	0.6182	0.5589	0.6916	0.7389	0.7205	0.7582
	LR (word TF-IDF)	0.7151	0.6337	0.8205	0.7795	0.7511	0.8102	0.543	0.4335	0.7266	0.7048	0.6894	0.7208
	LR (N-Gram TF-IDF)	0.7293	0.6535	0.825	0.761	0.7039	0.8283	0.5515	0.4411	**0.7355**	0.7143	0.7143	0.7143
	GCN	0.7613	0.758	0.7663	0.8491	0.8161	0.8787	0.6533	0.6271	0.6853	0.784	**0.8262**	0.7469
Temporal	nMIL	0.7304	0.6614	0.8155	0.7234	0.7969	0.6623	0.6277	**0.7193**	0.5567	0.7595	0.7692	0.750
	GCN+GRU	0.7825	0.7686	0.7999	0.85	0.825	0.8775	0.6547	0.6215	0.6963	0.7866	0.8087	0.7677
	GCN+LSTM	0.7813	0.7702	0.7938	0.8507	0.8271	0.8766	0.6493	0.6137	0.6914	0.7858	0.7914	0.7812
	GCN+RNN	0.7566	0.7553	0.7585	0.851	0.8204	0.8851	0.6416	0.6016	0.6892	0.7868	0.8088	0.7667
	DynamicGCN	**0.797**	**0.7734**	0.8248	**0.8617**	**0.8285**	**0.8984**	**0.6692**	0.6275	0.7196	**0.804**	0.7988	**0.8101**

Experiments

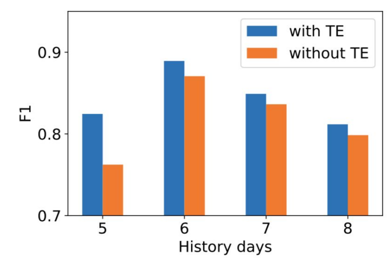
Sensitivity analysis on Temporal Encoding Module



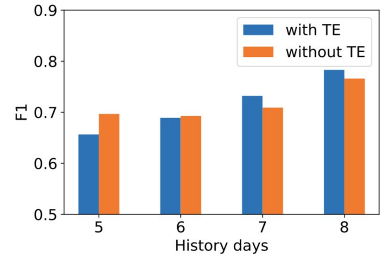
(e) Four countries



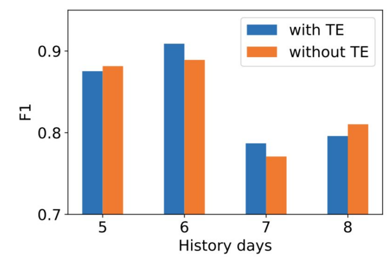
(a) Thailand



(b) Egypt



(c) India

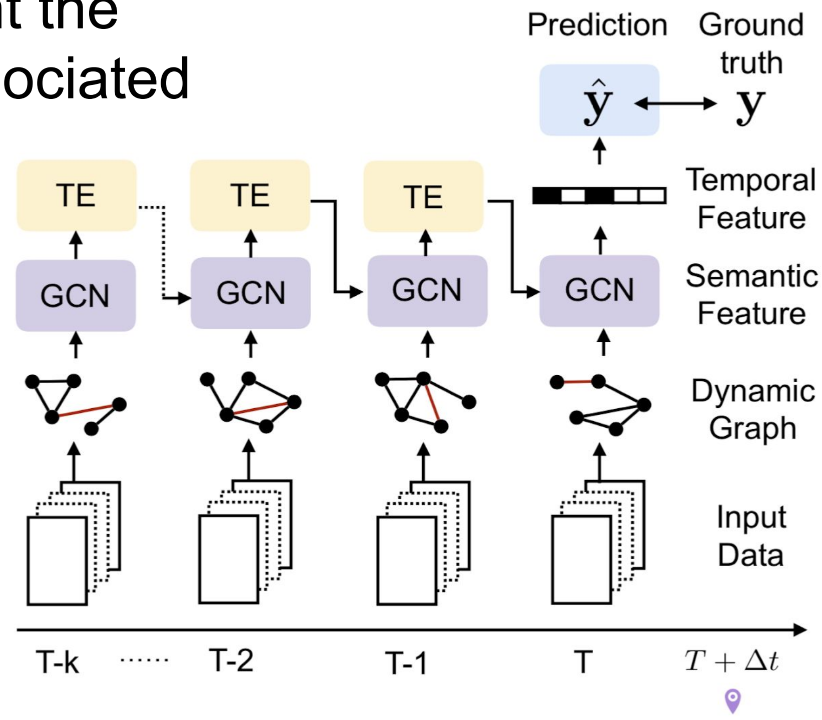


(d) Russia

Dynamic Context Graphs

Context Graph Generation

- Extract Nodes
  - A Gaussian distribution N(μ ≈ 0,σ2).
  - We set the threshold range (μ, μ + 2σ] for sampling nodes to represent the dynamic contextual graphs associated with the target events.

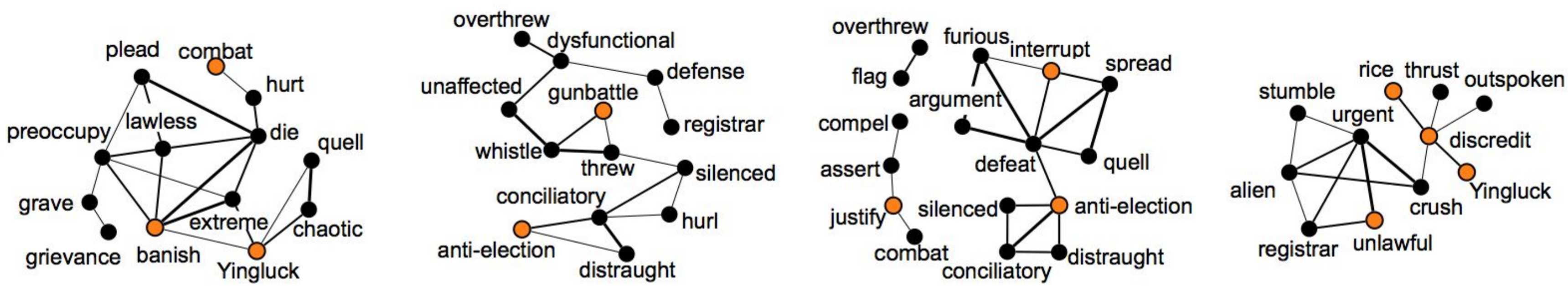


- Subgraph Construction

Case study

An event in Thailand

02/01/2014 02/02/2014 02/03/2014 02/05/2014



Violence grips Thai capital on eve of vote called by Yingluck.

Thailand started voting.Voters blocked by anti-election groups squared off with scuffles and hurled objects.

Election Commission asked the national police chief to maintain law and order. Thai Protests Disrupt Vote.

Yingluck's former commerce ministers were suspected of being involved in improper rice deals.

02/07/2014

Conclusion and future work

- A novel dynamic graph convolutional model with a temporal encoded feature module for event forecasting and for identifying dynamic context graphs.
- We demonstrated the effectiveness of the proposed model on large-scale real-world open source datasets.

    - Automatic relationship extraction for entities in dynamic graphs
    - Long-term event dependency
    - Consider multiple geo locations simultaneously and study the influence from neighbor locations

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

stevens.edu

Thank You

sdeng4@stevens.edu