Neural Models for News Events

Vinay Setty

vinaysetty.net

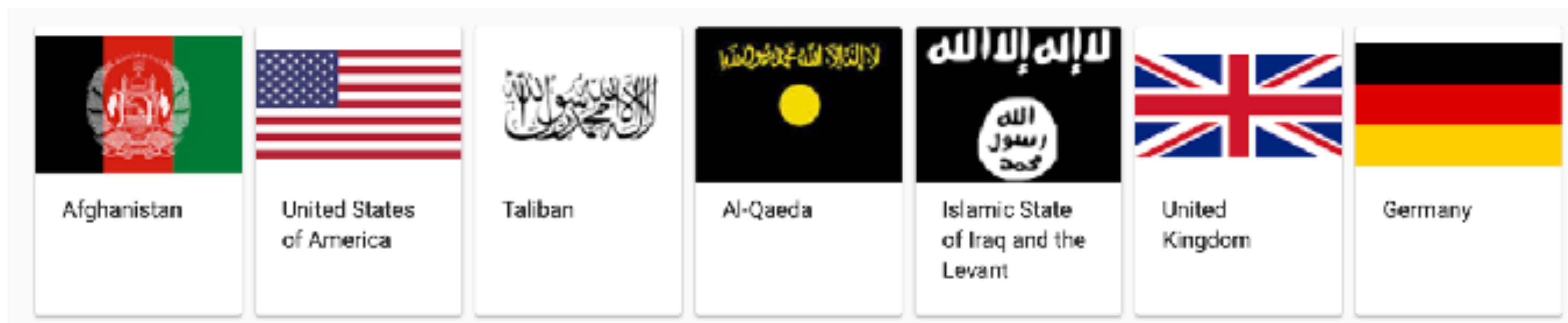
vinay.j.setty@uis.no

Aalborg University, Denmark

Joint work with: Arunav Mishra, Klaus Berberich, Supratim Das, Katja Hose

News Events

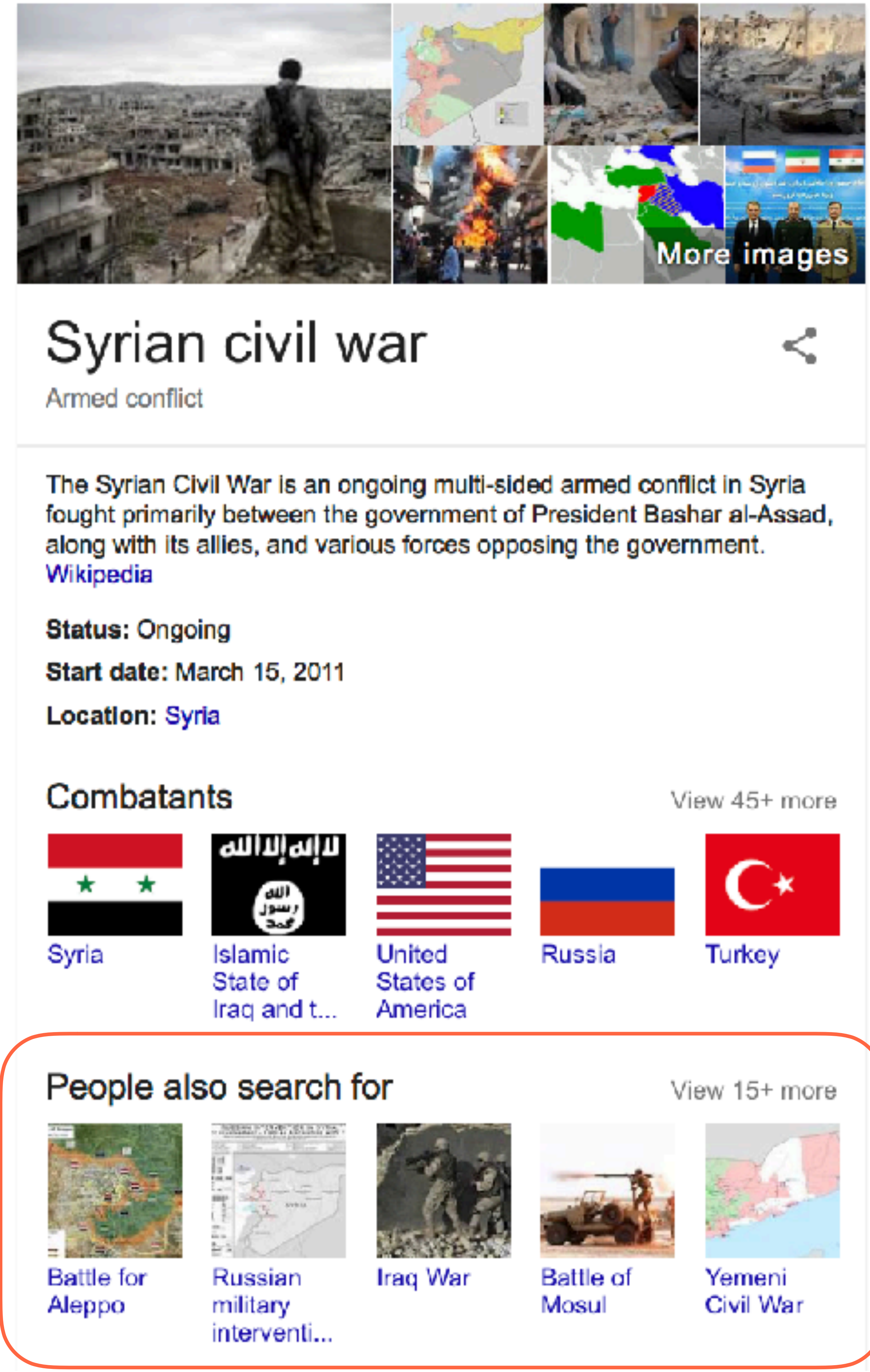
- Wars, natural disasters, elections etc
- Each event involves one or more of
  - People
  - Organizations
  - Locations
  - Time



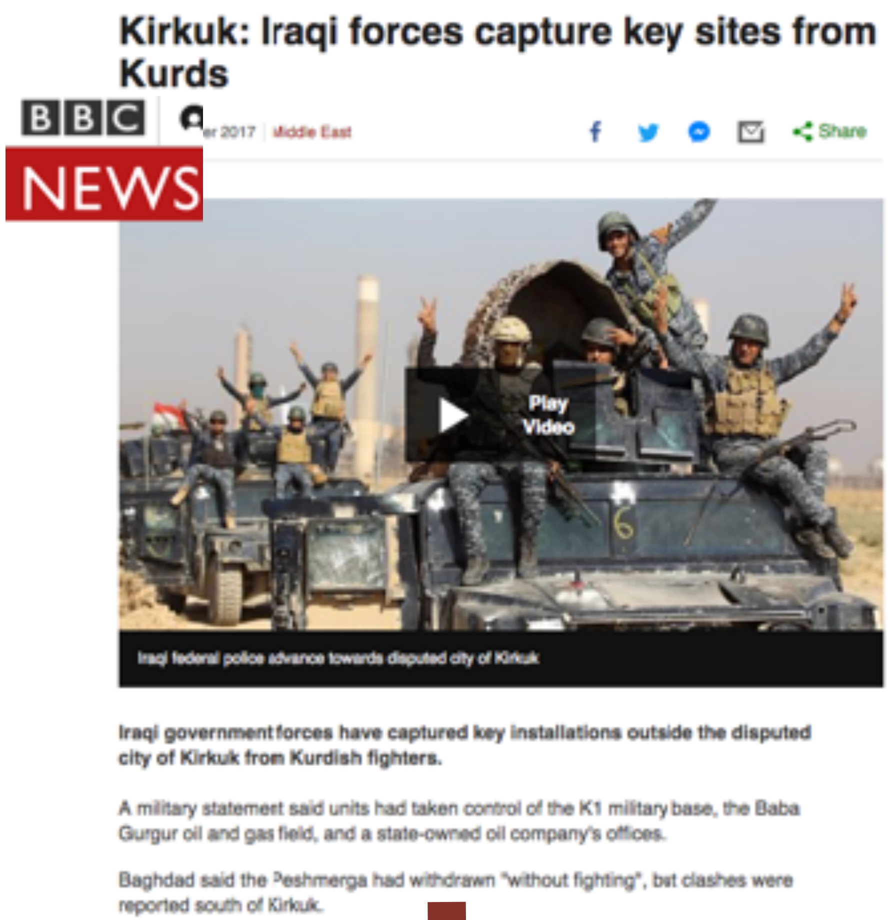
- **Research Question:** How to learn latent feature vectors to represent news events for various tasks?

Motivation

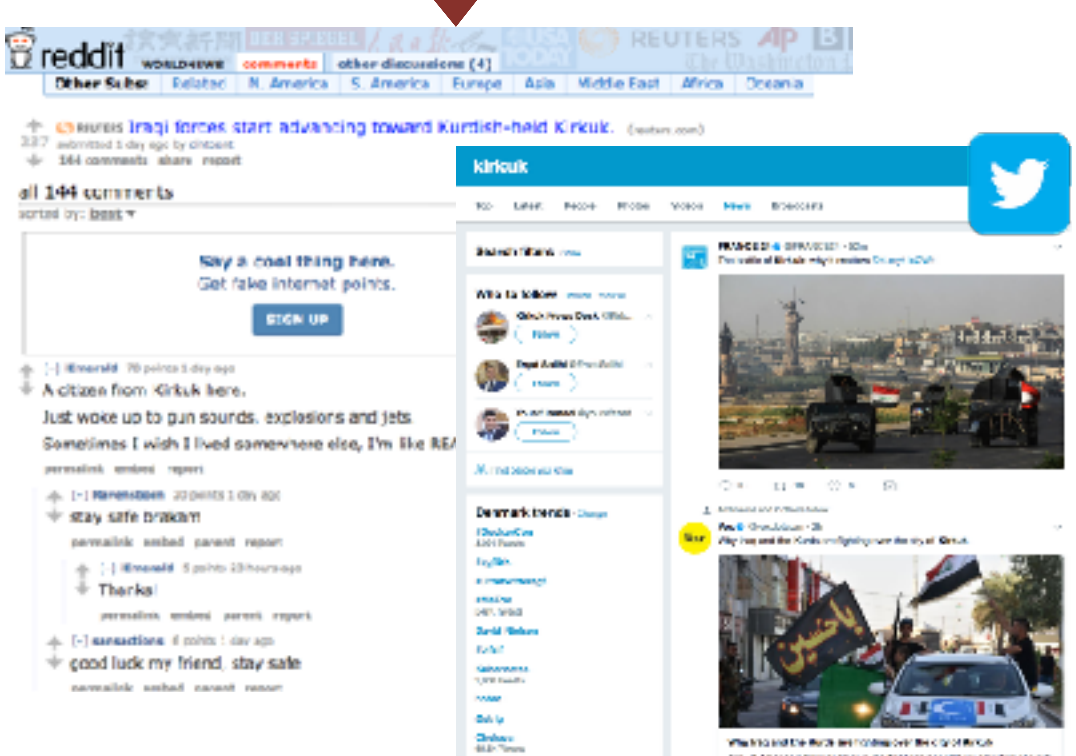
- Retrieving similar events
  - For journalists or historians
  - “Wars similar to Syrian Cvil War” - Spatio-Temporal dimensions
  - For news recommendation
- Similarity between news event latent features vectors



Motivation: Linking news events



WIKIPEDIA
The Free Encyclopedia



Motivation: Summarization, Ranking and Selection

Wikipedia Current Events Portal (WCEP)

October 17, 2017 (Tuesday) edit history watch

Armed conflicts and attacks

- Syrian Civil War
  - Battle of Raqqa
    - The Syrian Democratic Forces announce the end of the battle for ISIL's *de-facto* capital. (Al Jazeera)
- War in Afghanistan
  - A suicide attack on a police compound in Gardez, Paktia Province, kills at least 20 people, while a separate gun attack at a training facility in neighboring Ghazni Province leaves another 15 people dead. (Radio Free Europe/Radio Liberty)
- 2017 Iraqi–Kurdish conflict
  - The Iraqi Army and allied militias continue to seize Kurdish-held territory, taking over several key cities including Khanaqin near the Iranian border, Jalawla, Bashiqa, Sinjar and Rabia, as well as the Mosul Dam. (*The Washington Post*)

Motivation: Determining Event Focus Times

- For temporal IR
- Question Answering
- Knowledge extraction

Emmanuel Macron was sworn in as the President of France

- Event descriptions with no temporal expressions in them
- They may have different Focus Times than that of their host documents

Merkel heads for another term in Germany

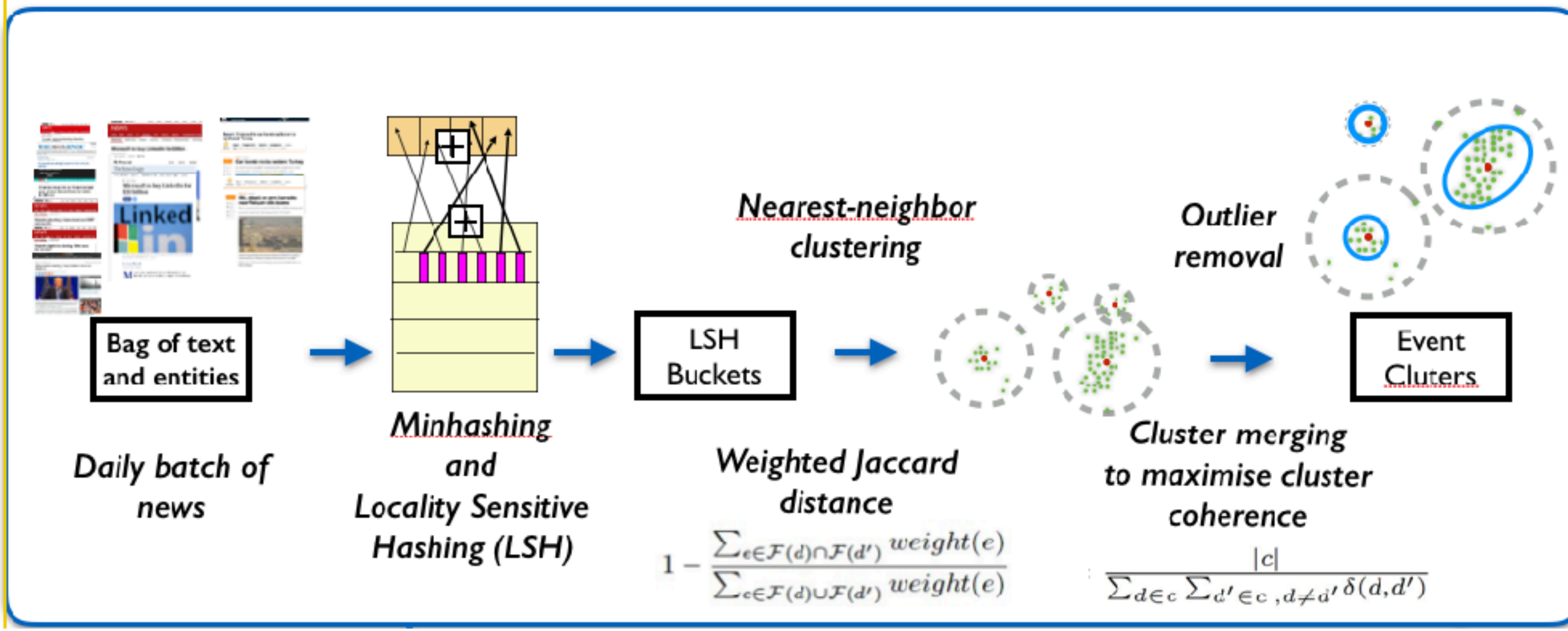
- It is important to disambiguate the queries to right focus time

State-of-the-art

- Use textual, temporal and entity features from news articles to represent events
  - Bag-of-words
  - Tf-idf vectors [Li et al., 2005]
  - Language Models [Lee et al., IR 2014]
- Multimodal distributions in text, space and time [Mishra et al., ECIR 2016]

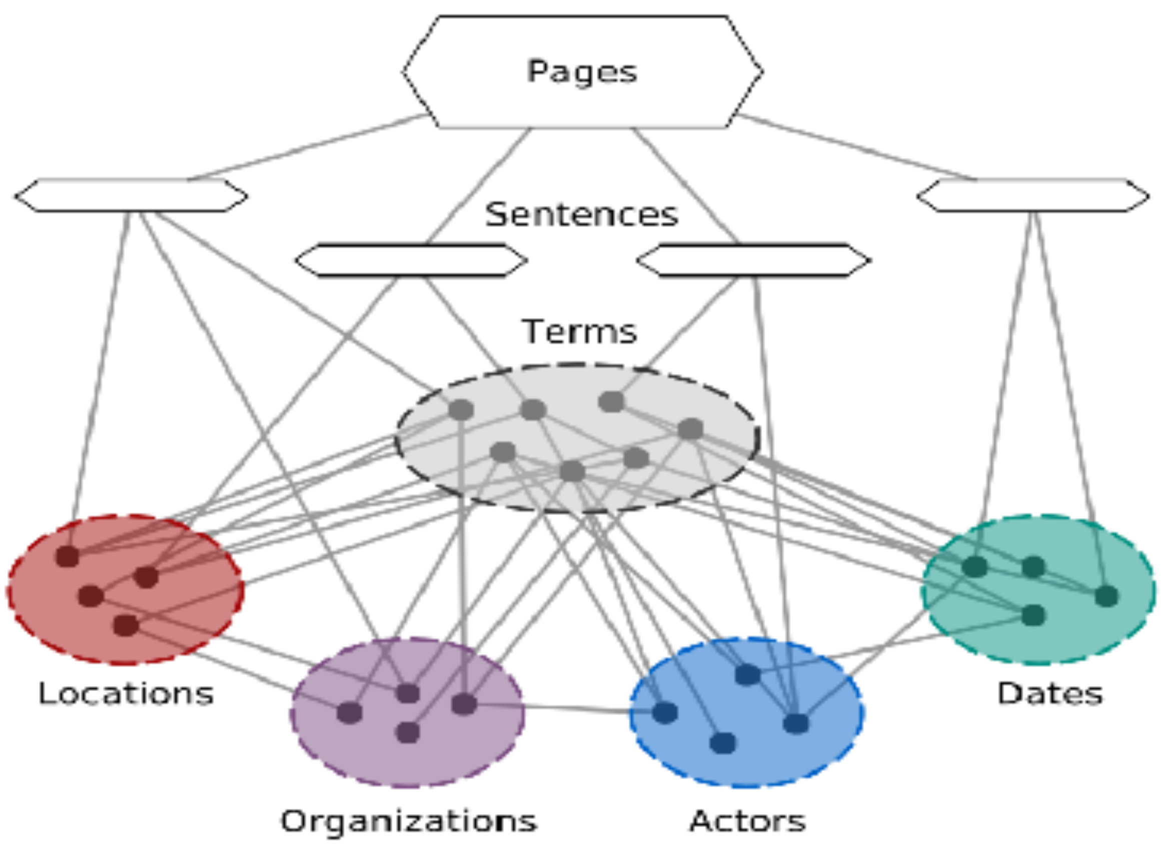
State-of-the-art: Clustering Approach

Events represented by clusters of news articles about same event [Setty et al., WSDM 2017]

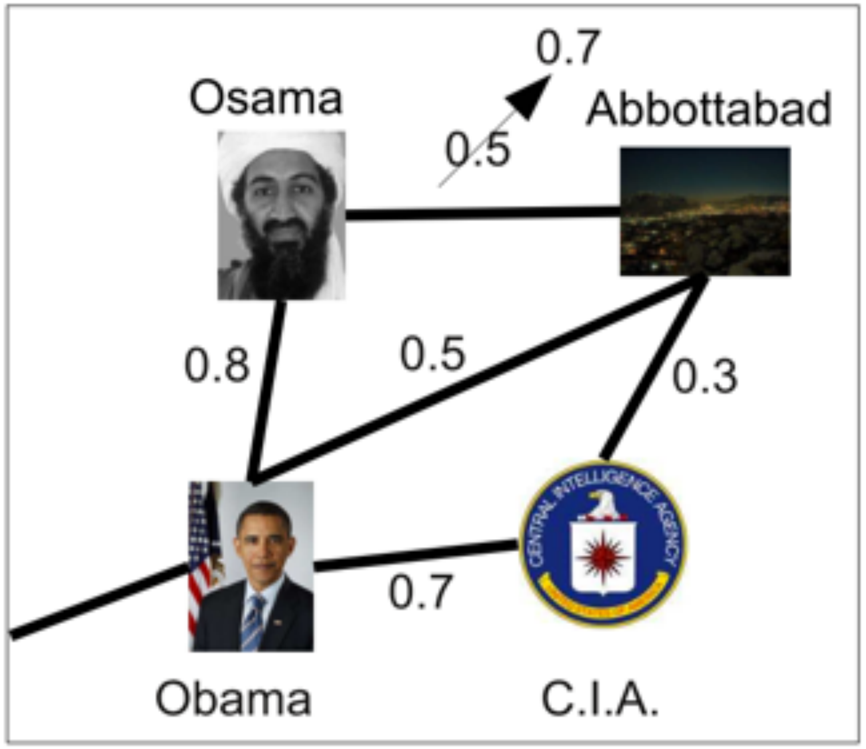


State-of-the-art: Graphs

- LOAD graph [Spitz et al., SIGIR 2016]
  - Links between entities, events and dates



- Co-occurring entities [Angel et al., 2012]
  - Links between entities in same documents
  - Dense-subgraphs are assumed to be events



Evolving Entity Graph

Neural Models for Event Embeddings

- Distributed representation of events considering text, entities, locations, and time
- Context of words, entities and time captures the semantics of events
- Avoid handcrafting features?
- Address the sparsity of event descriptions using distributed representation
- Avoid the ambiguity in event descriptions
  - E.g. "Germany became the world champion of football"

Take 1 : Word Vectors

- Learn word embeddings (Word2Vec)
  - Skip-gram model with negative sampling
- Event vector by combining
    - **Word vectors** in the event description
- Vectors for **temporal expressions**
  - Treat temporal expressions as special tokens and learn word vectors

Global Temporal Vectors

- Corpus-wide word vectors associated with temporal expressions
- News events in a given time period are diverse
- Global vectors of temporal expressions are hence overly generic

Year : 2002

The physical Euro is officially introduced in the Eurozone countries.

Riots and mass killings in the Indian state of Gujarat.

Queen Elizabeth II of the United Kingdom celebrates her Golden Jubilee

Event-Specific Temporal Vectors

Query model improved by Linearly combining with a feedback model estimated from pseudo-relevant documents [Zhai Et al., 2008]

Riots and mass killings in the Indian state of Gujarat.

Applying standard query likelihood model

Retrieve a set of Pseudo-Relevant documents

THE HINDU

Online edition of India's National Newspaper

Front Page

Shoot orders in many Gujarat towns, toll over 200

A Hindu mob waves swords at an opposing Muslim mob during street battles in Ahmedabad on Friday. FACE OF FEAR: A Muslim seeks mercy from rioters. - Reuters

Gujarat riot death toll revealed

India has for the first time published detailed figures on the number of people killed in the religious riots in the western state of Gujarat in 2002.

The government told parliament that 790 Muslims and 254 Hindus were killed, 223 more people reported missing and another 2,500 injured.

Death for 11, life sentence for 20 in Godhra train burning case

Collect words co-occurring with the temporal expressions

Generate Event specific temporal by combining global temporal vector (analogous to a feedback model)

$$\mathbf{t_e} = (1-\alpha)\mathbf{t_g} + \alpha\Big(\frac{1}{N}\sum_{w\in S}\mathbf{w}\Big)$$

Event specific vector

Global vectors (for same time period)

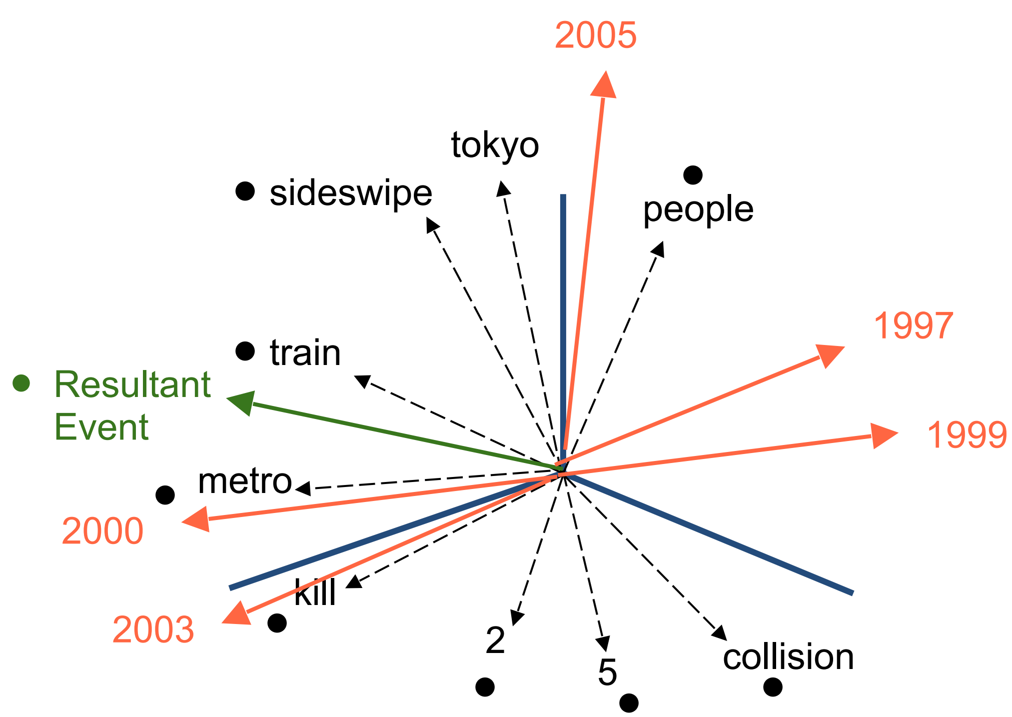
Vectors of co-occurring words

Event Focus Time Ranking

- Early Fusion
  - Compute centroid of word vectors of the event description
    - Simple mean $\mathbf{e} = \frac{1}{n}\left(\mathbf{V}_{w_1} + \mathbf{V}_{w_2} + \cdots + \mathbf{V}_{w_n}\right)$
    - tf-idf weighted $\mathbf{e} = \left(\mathbf{V}_{w_1} \times \alpha_{w_1} + \mathbf{V}_{w_2} \times \alpha_{w_2} + \cdots + \mathbf{V}_{w_n} \times \alpha_{w_n}\right)$
  - Compute a list of temporal vectors from the temporal expressions in the event-related documents (from before)
  - Then the focus time is the temporal vector with maximum similarity (using cosine similarity)

Early Fusion Example

Event: *"A sideswipe collision of 2 Tokyo Metro trains kills 5 people."*
Focus Time: 2000



Late Fusion Motivation

- Some words may be too generic causing noise in vectors
  - For example in "A sideswipe collision of 2 Tokyo Metro trains kills 5 people."
  - Words like trains, kills etc are too generic and make it harder to distinguish different focus times
- Focus on words specific to the event: sideswipe, Tokyo, Metro and collision
- Treat named entities specially (entity vector)

Late Fusion

- Each word in the event description is treated as an independent query
- Ranked list of temporal vectors are selected based on similarity scores
- The temporal vectors are then normalised and aggregated

Event Query: *“A sideswipe collision of 2 Tokyo Metro trains kills 5 people.”*
Focus Time: 2000

sideswipe	collision	2	tokyo	metro	trains	kills	5	people		Final
2002	2000	2002	2001	2000	2000	2000	2000	2001		2000
2003	2002	1994	2002	2005	2005	2005	2005	2008		2008
1997	2007	1996	1999	1997	1992	1992	1992	1990	Σ	1990
2005	1998	2005	2008	2006	2010	2010	2006	2003		2003
1999	1999	1996	1995	1991	1991	1991	1992	1992		1992

Evaluation Setup

- Background Corpus
  - English Gigaword corpus - 9 million news articles published between 1994 and 2010 taken from five different news sources
  - ClueWeb12-B13 corpus (CW12) - about 50 million web pages crawled in 2012
- Queries
  - 100 Random events from Wikipedia year pages to serve as our test queries
  - Note some queries are outside the span of Giga collection (represented as Giga-OC)

Baseline Approaches

Methods	Description
LOAD	Spitz et al., SIGIR 2016
Adj	Estimating focus time for documents Jatowt et al. CIKM 2013
ML	Maximum Likelihood
LDA	Latent Dirichlet Allocation-based method
EF_MEAN	Early Fusion with Mean of vectors
EF_TFIDF	Early Fusion using TF-IDF weights
LF_TFIDF	Late Fusion with TF-IDF weights
LF_MEAN	Late Fusion with Mean of vectors
LF_NER	Late Fusion with Named Entities

Event Focus Time Estimation Results

Methods	Giga-All		CW-All		Giga-OC		CW-OC	
	MRR	Giga-All	MRR	CW-All	MRR	Giga-OC	MRR	CW-OC
LOAD	0.459	0.567	0.216	0.375	0.074	0.234	0.087	0.247
Adj	0.516	0.609	0.431	0.551	0.159	0.301	0.19	0.331
ML	**0.642**	**0.736**	0.471	0.584	0.206	0.367	0.327	0.448
LDA	0.192	0.354	0.114	0.288	0.07	0.229	0.056	0.22
EF_MEAN	0.607	0.689	0.617	0.694	0.571	0.641	**0.613**	**0.674**
EF_TFIDF	0.598	0.683	0.622	0.699	**0.575**	**0.645**	0.606	0.671
LF_TFIDF	0.56	0.654	0.611	0.676	0.493	0.584	0.561	0.618
LF_MEAN	0.57	0.662	**0.636**	**0.71**	0.557	0.63	0.59	0.642
LF_NER	0.623	0.702	0.609	0.689	0.549	0.626	0.598	0.664

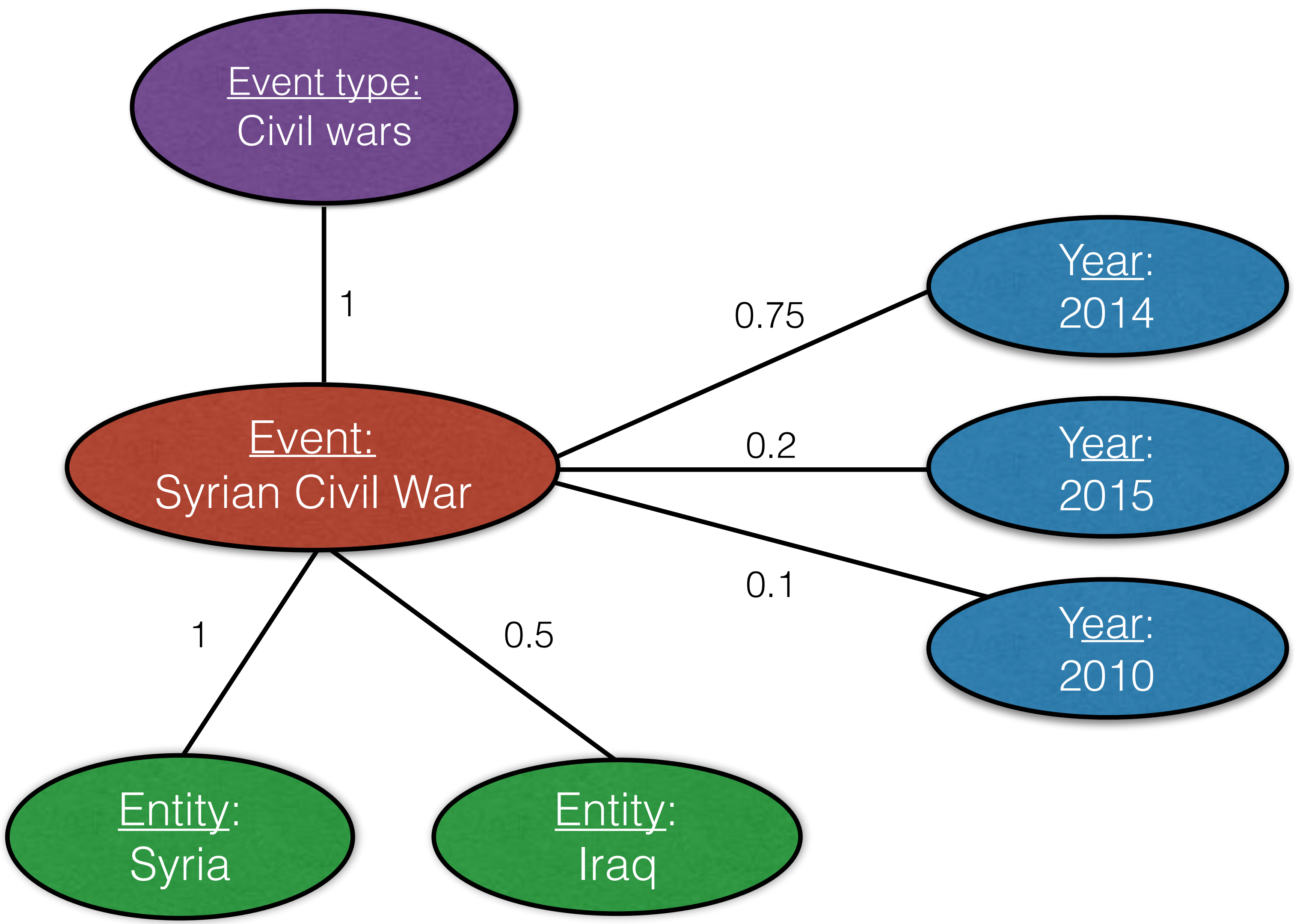
Insights

- **Best performing query:** "Republic of China Army executes 19 unarmed Vietnamese refugees on Donggang beach, Lieyu, Kinmen off Mainland China." Foucs time: 1987
  - Adj method fails due to lack of any temporally focused words
  - LF-NER performs best using time vector based on co-occurring words within sentences
- **Worst performing query:** "Several explosions at a military dump in Lagos, Nigeria kill more than 1,000. Focus time: 2002"
  - Few in-discriminative terms such as explosion, kill and military.
  - LOAD graph finds correct answer due to entities Lagos and Nigeria

Take 2 : Network Embeddings

- Network embeddings are useful for various network analysis tasks and prediction tasks [Deepwalk, Node2Vec]
- Can we use similar techniques for news events?
- Word embeddings are learned form sequential text they can miss cross-document relationships
- News events can be modelled as networks
  - Events and event categories, entities, temporal expressions as nodes
  - Edges between the nodes are added if they co-occur
  - Edges can be weighted as well depending on how often they co-occur (tf-idf)

Network Construction



Network Embeddings

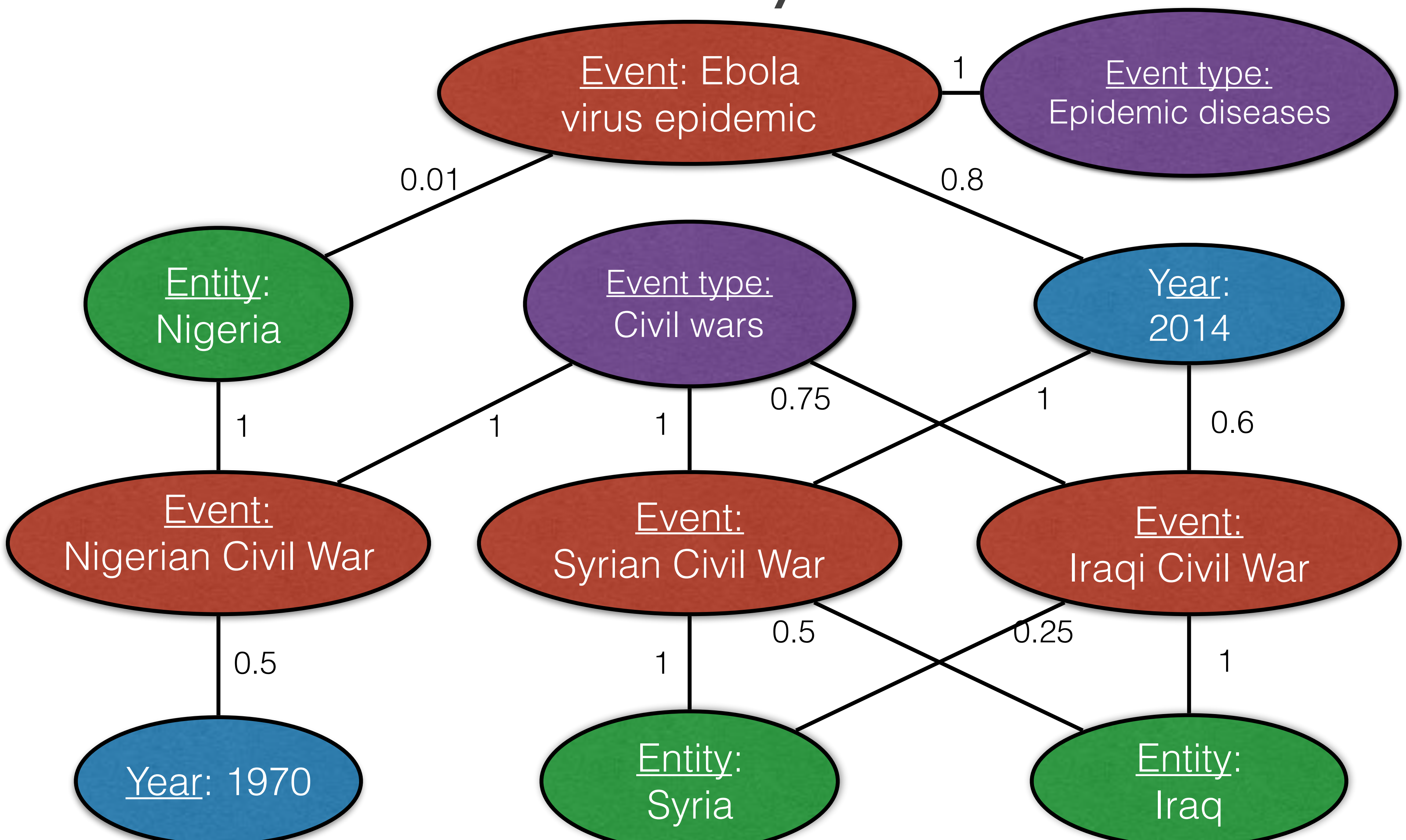
- Our goal is to learn a feature vector ***F(v)*** for each node v
- ***N(v)*** represents neighbourhood of node v
- The goal is to maximise the probability of predicting the neighbourhood of a node v given its *F(v)*

$$\max \sum \log Pr(N(v)|\mathcal{F}(v))$$

- This is solved using softmax function parametrised by the dot product of feature vectors
- This is approximated using random walks

Heterogeneous Network Embeddings

- It is important to treat nodes according to node types
  - For example, for Syrian Civil War other civil wars are very important
  - but other events from same year not so much



Event-Centric Biased Random Walk

- Transition probabilities depend on
  - Distances between source node and the destination node
    - To simulate DFS and BFS

$$\alpha_{vx} = \begin{cases} \frac{1}{p}, & \text{if } d_{tx} = 0 \\ 1, & \text{if } d_{tx} = 1 \\ \frac{1}{q}, & \text{if } d_{tx} = 2 \end{cases}$$

  - Node type
    - Same event types are always visited
    - Temporal and entity nodes are visited based on their neighbourhood overlap

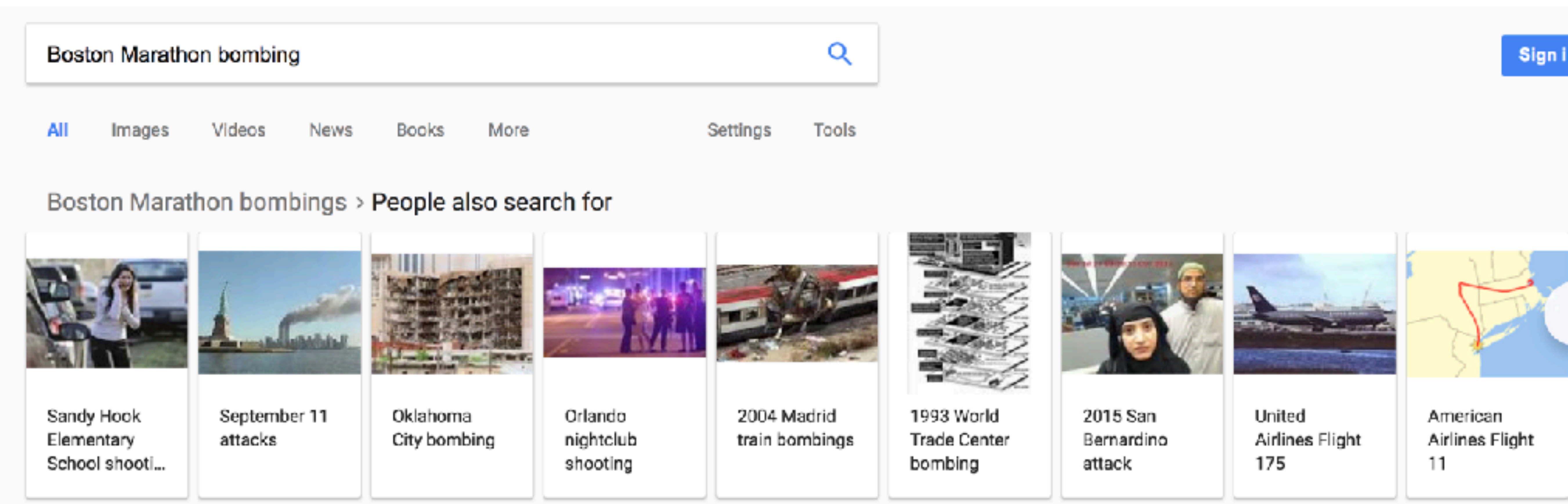
$$\beta_{vx} = \begin{cases} 1, & \text{if v, x have same event type} \\ \frac{|N(v) \cap N(x)|}{|N(v) \cup N(x)|}, & \text{if x is an entity or year} \end{cases}$$

Experimental Setup

- News events from Wikipedia Current Events Portal
  - Over 7000 events from 2007 to 2017
  - External news article links from these events
- Entities annotated using AIDA
- Event categories from Wikipedia
- Temporal Expressions annotated using HeidelTime

Similar Events Task

- Ground-truth from Google's "People also search for" news events
- Crawled around 2000 events



Similar Events Task : Results

	P@10	P@20	P@30	P@40	R@10	R@20	R@30	R@40
Node2vec Event-Centric	0.364	0.262	0.212	0.175	0.357	0.488	0.582	0.638
Node2Vec (Grover and Leskovec, 2016)	0.313	0.228	0.181	0.151	0.315	0.442	0.518	0.570
LOAD Graph (Spitz and Gertz, 2016)	0.131	0.100	0.084	0.072	0.135	0.198	0.246	0.280

- LOAD graph relies on immediate neighbourhood of nodes to rank
- Network embeddings cover broader/deeper neighbourhood
- But in general the precision and recall values are low
- The ground-truth from Google crawl is not really “the ground-truth”
  - We also conducted manual evaluation on CrowdFlower and got a P@10 of 0.784

Event Linking Task

- Task: Given a news article link it to appropriate Wikipedia Event page


BBC NEWS

Kirkuk: Iraqi forces capture key sites from Kurds

16 October 2017 | Middle East

Share

Play Video

Iraqi federal police advance towards disputed city of Kirkuk

Iraqi government forces have captured key installations outside the disputed city of Kirkuk from Kurdish fighters.

A military statement said units had taken control of the K1 military base, the Baba Gurgur oil and gas field, and a state-owned oil company's offices.

Baghdad said the Peshmerga had withdrawn "without fighting", but clashes were reported south of Kirkuk.

- Ground-truth from Wikipedia Current Events Portal (WCEP)

	P@1	P@10	P@20	MRR
Node2vec Event-Centric	0.643	0.845	0.899	0.487
Node2vec (Grover and Leskovec, 2016)	0.425	0.628	0.712	0.364
LOAD Graph (Spitz and Gertz, 2016)	0.112	0.303	0.397	0.212

Summary

- First steps towards learning latent features for news events
- Two approaches
  - Fusing word vectors of event descriptions
  - Network embeddings
- Evaluated for several tasks
  - Event focus time estimation
  - Similar events
  - Event linking

Shameless Advertisement

- We are hiring!
- A PhD position on “Neural models for news events” is available at University of Stavanger!!
- Contact:
  - vinay.j.setty@uis.no
  - http://vinaysetty.net