

Refining Imprecise Spatio-temporal Events: A Network-based Approach

Andreas Spitz¹, Johanna Geiß¹, Michael Gertz¹,
Stefan Hagedorn² and Kai-Uwe Sattler²

¹ Heidelberg University, Institute of Computer Science
Database Systems Research Group, Heidelberg

²Technical University of Ilmenau
Databases and Information Systems Group, Ilmenau

{spitz, geiss, gertz}@informatik.uni-heidelberg.de
{stefan.hagedorn, kus}@tu-ilmenau.de

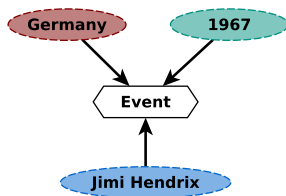
10th GIR Workshop
San Francisco,  2016



What is an Event?

Event

*"The Jimi Hendrix Experience
toured Germany in 1967."*



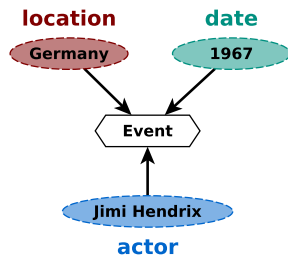
What is an Event?

Event

"The Jimi Hendrix Experience toured Germany in 1967."

Definition: Event

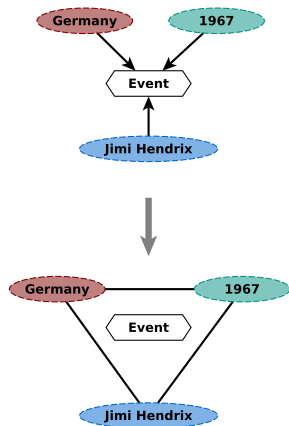
*"Something that happens at a given **place** and **time** between a group of **actors**."* [All02]



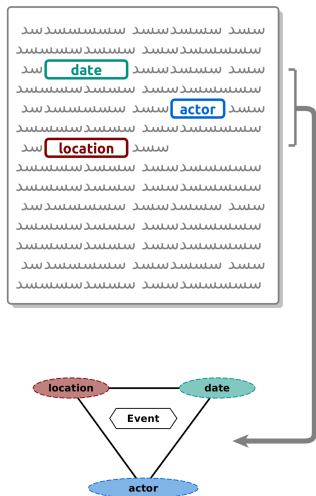
Event Triangles

Intuition:

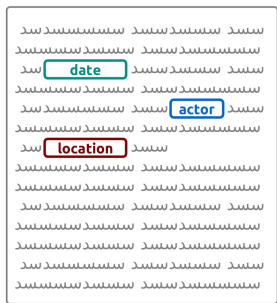
- Events correspond to triangular structures of entities
- Participating entities can be used to extract events
- Linguistic, sentence-based event detection can improve results



Event Extraction



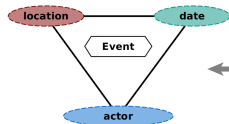
Event Extraction



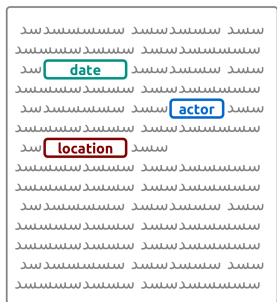
Generating weights for events:

- Similarity of entities [GSG15]:

$$\phi(r, r') := \exp\left(-\frac{\text{dist}(r, r')}{2}\right)$$



Event Extraction



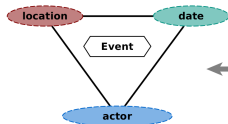
Generating weights for events:

- Similarity of entities [GSG15]:

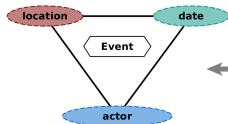
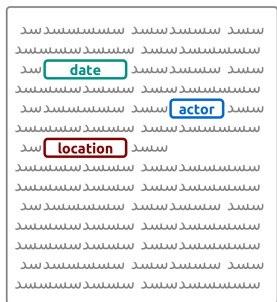
$$\phi(r, r') := \exp\left(-\frac{\text{dist}(r, r')}{2}\right)$$

- Weight of individual events:

$$\omega(e_i) = \min\{\phi(l, t), \phi(l, a), \phi(t, a)\}$$



Event Extraction



Generating weights for events:

- Similarity of entities [GSG15]:

$$\phi(r, r') := \exp\left(-\frac{\text{dist}(r, r')}{2}\right)$$

- Weight of individual events:

$$\omega(e_i) = \min\{\phi(l, t), \phi(l, a), \phi(t, a)\}$$

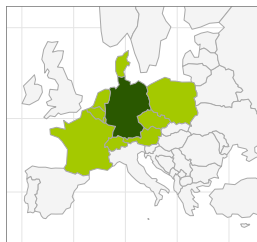
- Weight of aggregated events:

$$\omega(e) := \sum_{i=1}^k \omega(e_i)$$

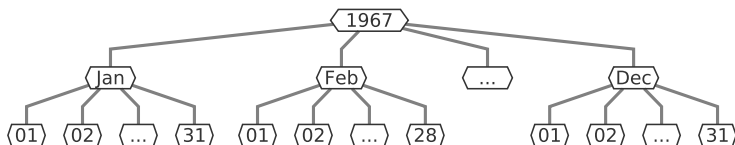
Geographic Background Network

Locations can be connected in a location network G_L based on:

- Geographic containment
- Geographic neighbourhood
- Reachability
- Semantic similarity
- Context-dependent relations



Temporal Background Network



Dates are inherently connected in a temporal network G_T :

- Temporal containment is straightforward
- Heterogeneous granularity levels are possible

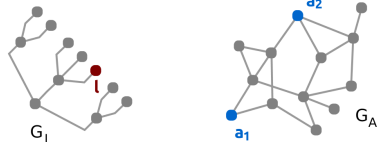
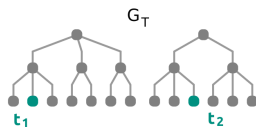
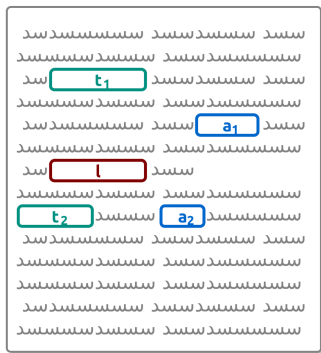
Social Background Network

Actors form a social network G_A and are connected by:

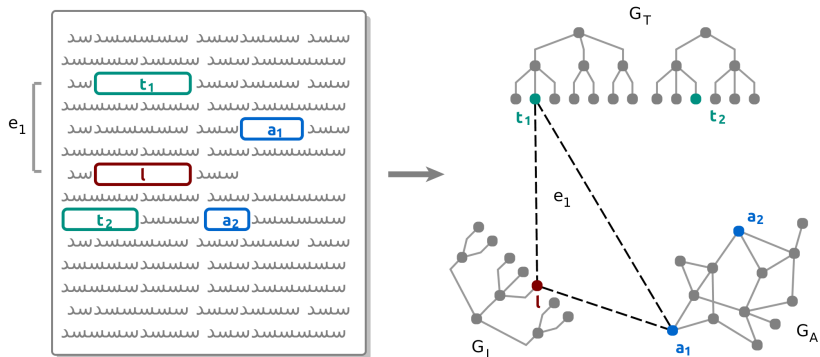
- Acquaintance or Relation
- Collaboration
- Organization membership
- Context-dependent similarities



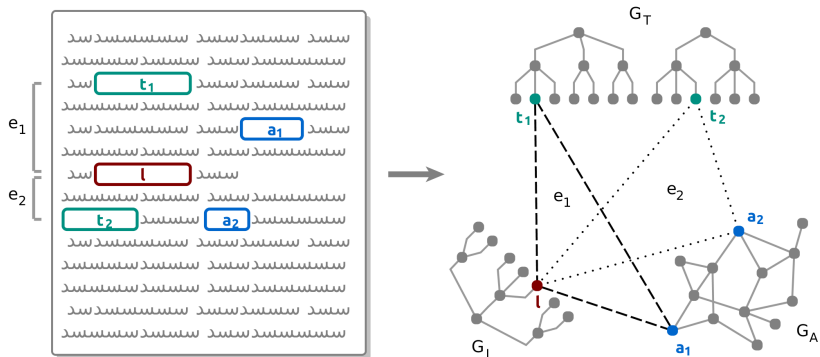
Event Hypergraph



Event Hypergraph

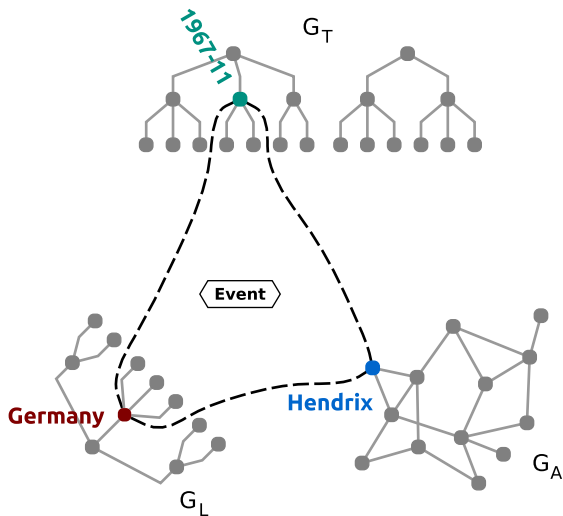


Event Hypergraph

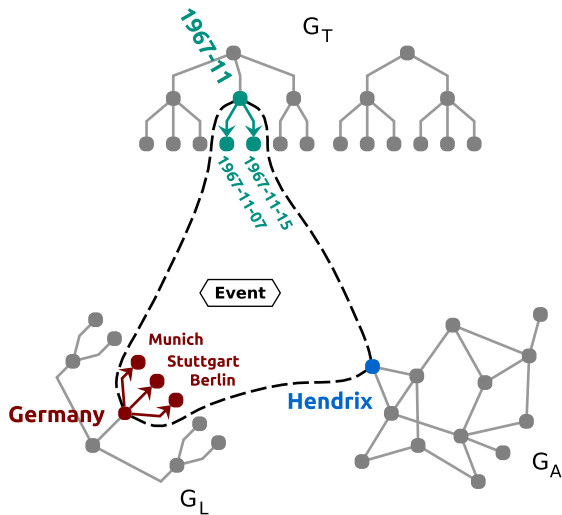


Event Refinement

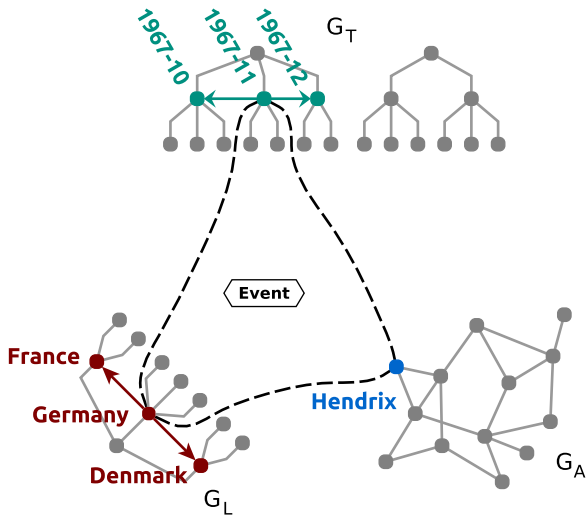
Event Refinement



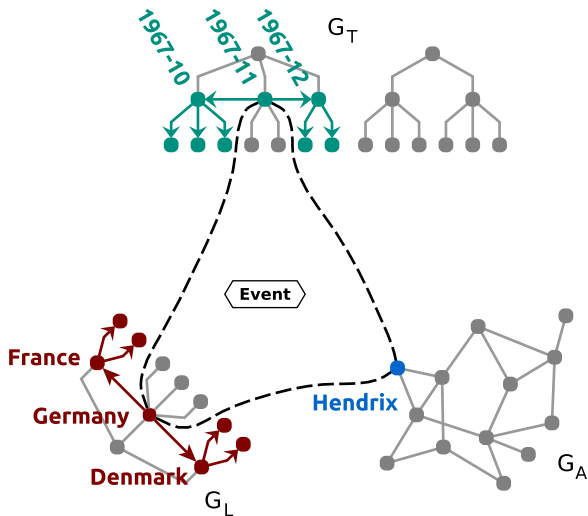
Granularity Refinement



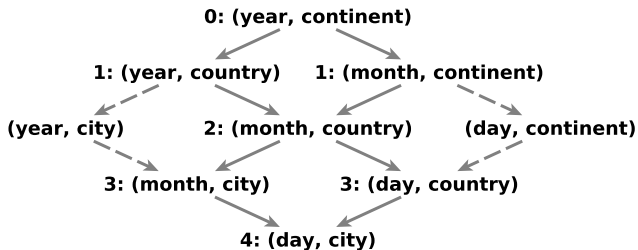
Neighbourhood Refinement



Neighbourhood Refinement



Event Refinement: Stratification



Evaluation

Extraction of Events from Wikipedia



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikipedia store

Interaction

Help

Article [Talk](#)

1984 Summer Olympics

From Wikipedia, the free encyclopedia

The **1984 Summer Olympics**, officially known as the **Games of the XXIII Olympiad**, was an international multi-sport event held in **Los Angeles, California, United States** in 1984. **When Tehran**, the only other interested city on the international level, declined to bid due to the concurrent Iranian political and social changes, the IOC awarded Los Angeles the Games. On a second occasion Los Angeles hosted the event, which was hosted in 1932.

Los Angeles

From Wikipedia, the free encyclopedia

Coordinates: 34°05′N 118°15′W

This article is about the U.S. city. For the county in California, see [Los Angeles County](#). For other uses, see [Los Angeles \(disambiguation\)](#).
 "L.A." and "City of Los Angeles" redirect here. For other uses, see [L.A. \(disambiguation\)](#) and [City of Los Angeles \(disambiguation\)](#).

Los Angeles (/ˈlɒsˈæŋdʒələs/ *LOSS AN-jə-ləs* or *LOSS AN-jə-liss*)^[k] officially the **City of Los Angeles** and often known by its initials **L.A.**, is the second-largest city in the United States, the most populous city in the U.S. state of California, and the county seat of Los Angeles County. Situated in Southern California, Los Angeles is known for its mediterranean climate, ethnic diversity, sprawling metropolis, and as a major center of the American entertainment industry. Los Angeles lies in a large coastal basin surrounded on three sides by mountains reaching up to and over 10,000 feet (3,000 m).

Los Angeles, California

Charter city

City of Los Angeles



Tehran

From Wikipedia, the free encyclopedia

Coordinates: 35°41′46″N 51°23′23″E

This article is about the Iranian capital city. For other uses, see [Tehran \(disambiguation\)](#).

Tehran (/tɛhˈrɑːn/; Persian: تهران‎ - *Tehrān*) is the capital of Iran and Tehran Province. With a population of around 9 million in the city and 16 million in the wider metropolitan area,^[k] Tehran is the largest city and urban area of Iran, the second largest city in Western Asia, and the 3rd largest in the Middle East. It is ranked 29th in the world by the population of its metropolitan area.^[k]

In the Classical era, part of the present-day city of Tehran was occupied by a Median city which in the Avesta occurs as *Ahagā*.^[l] It was destroyed by the Mongols in the early 13th century, and remains now as a city in Tehran Province, located towards the south end of the modern-day city of Tehran.

تهران

Metropolis

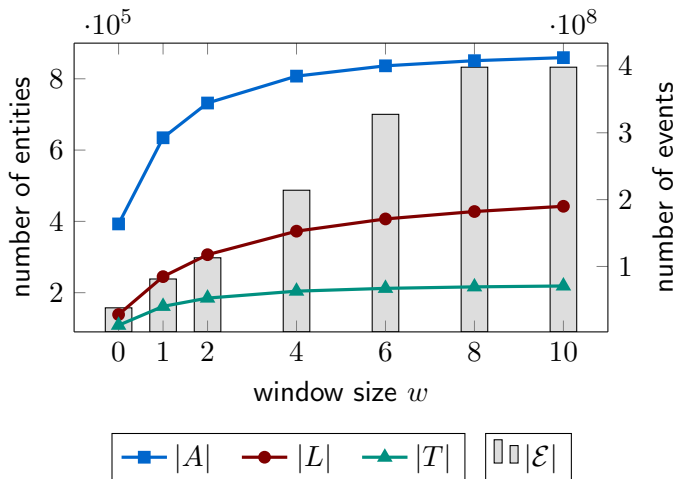
تهران - Tehran Metropolis





WIKIDATA

Entities and Events by Window Size



Background Network Construction

Location Network:

- From Wikidata: continents, countries and cities
- Containment edges between hierarchy levels
- Neighbourhood edges between adjacent countries
- Neighbourhood edges between close cities

Temporal Network:

- Temporal Tagging with HeideTime [SG13]
- Containment edges between years, months and days
- Neighbourhood edges within granularity layers

Ground Truth Event Queries

We obtain events by

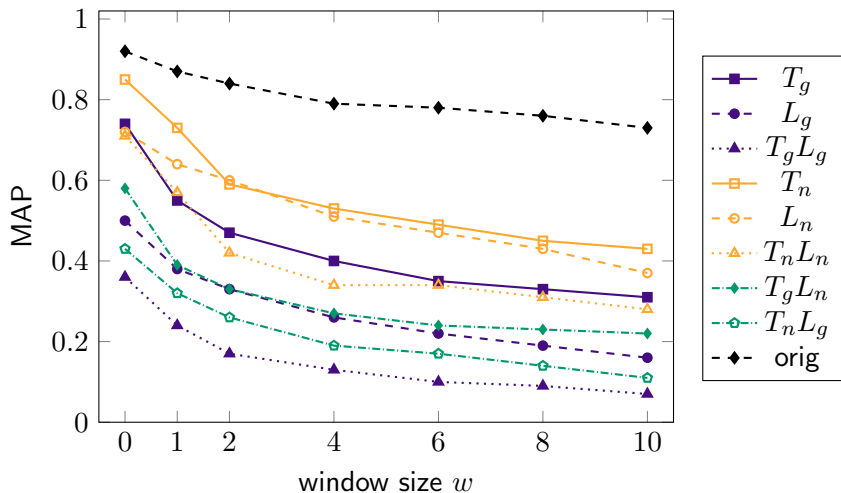
- Named Entity Recognition in online news articles
- Extraction of the entity triples of events by hand

We generate queries by making events less certain in the combination of dimensions

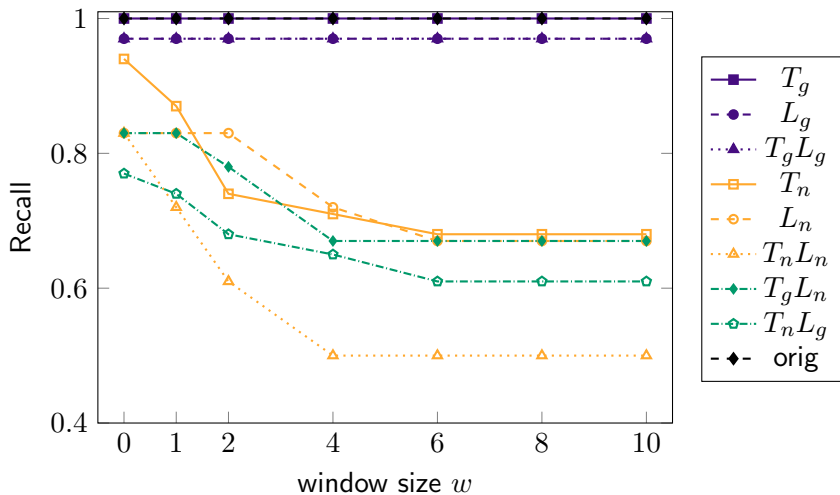
- Location L
- Time T
- Granularity g
- Neighbourhood n



Evaluation Results (Precision)



Evaluation Results (Recall)



Summary

New method for event representation:

- As a hypergraph model
- Backed by underlying entity networks
- Compatible with any entity-based definition of event

Graph-based event refinement offers:

- Spatio-temporal refinement in two dimensions: neighbourhood and granularity
- Efficient computation due to localized queries
- Language independence

Ongoing Work

Directions for ongoing event refinement research:

- Include measures of granularity and neighbourhood in social background networks
- Include hierarchical or organization networks
- Add event structures beyond triangles

The event network and the background networks are available for download.



<http://dbs.ifi.uni-heidelberg.de/index.php?id=data>

The event network and the background networks are available for download.



<http://dbs.ifi.uni-heidelberg.de/index.php?id=data>

Thank you!
Questions?

Bibliography



James Allan.

Topic Detection and Tracking: Event-Based Information Organization,
volume 12.

Springer Science & Business Media, 2002.



Johanna Geiß, Andreas Spitz, and Michael Gertz.

Beyond Friendships and Followers: The Wikipedia Social Network.
In *ASONAM*, 2015.



Jannik Strötgen and Michael Gertz.

Multilingual and cross-domain temporal tagging.
Language Resources and Evaluation, 47(2), 2013.



Denny Vrandečić and Markus Krötzsch.

Wikidata: A Free Collaborative Knowledgebase.
Communications of the ACM, 57(10), 2014.