

What describes an event?

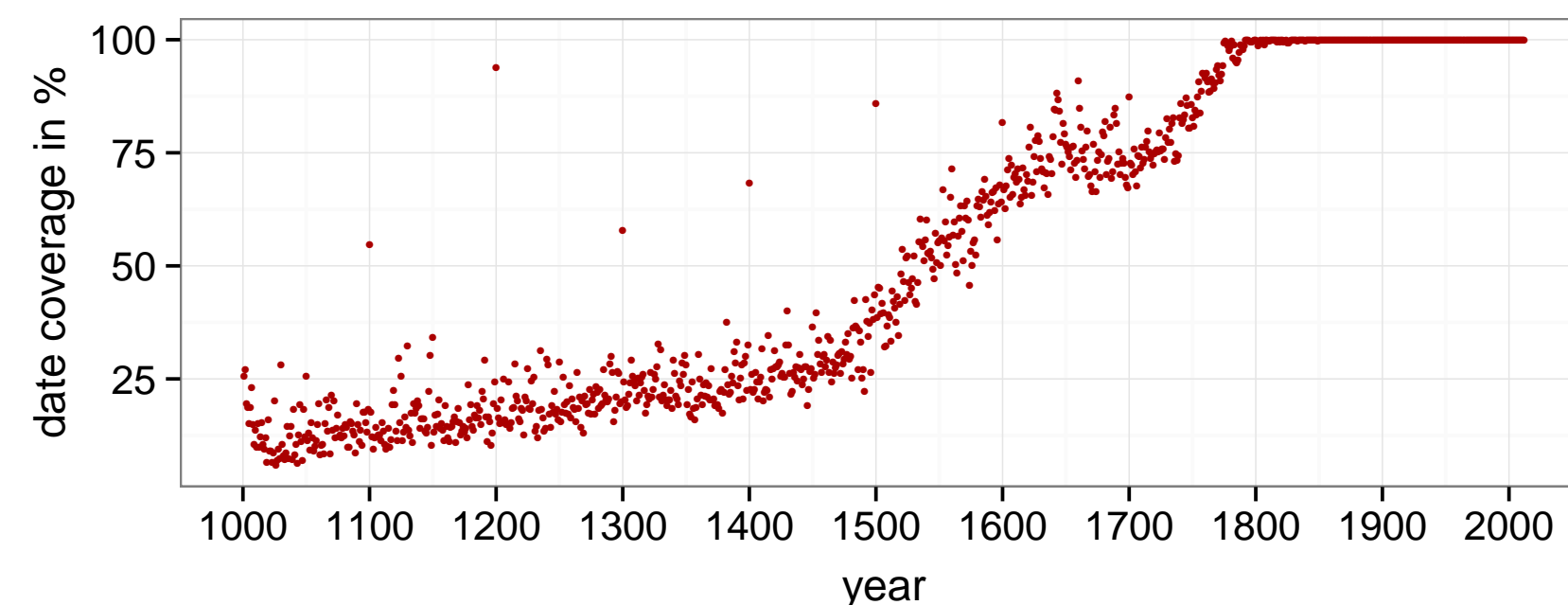


Textual representations and mentions of events have a **spatial** and **temporal** component and involve a set of **actors**.

First Approach: Temporal Profiles

With the English Wikipedia as corpus:

- Extract **terms** (content words)
- Extract **dates** at granularity levels day, month and year

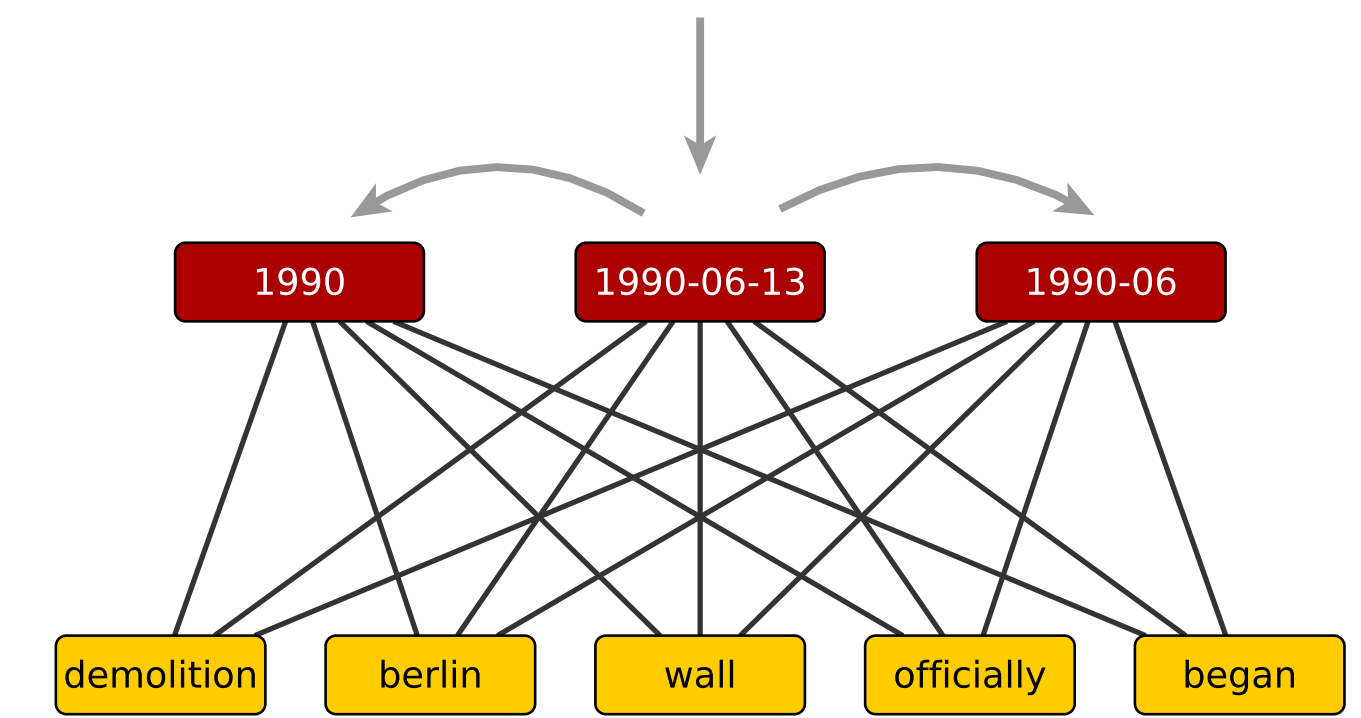


Concept: Use term-time co-occurrences to construct temporal term profiles and contextual date profiles to extract information from the text [1].

Term-Time Co-occurrences

Create bipartite sentence-graphs:

The Demolition of the Berlin Wall officially began on 13 June 1990.



Combine into one graph representation

$G = (T \cup D, E, \omega)$ with $\omega : E \rightarrow \mathbb{N}$:

$|T| = 3,748,730$ terms

$|D| = 210,375$ dates

$|E| = 110,639,525$ edges

Significance- & Similarity-Measures

Information Retrieval

Leverage the graph representation to identify significant co-occurrences of dates and terms in the corpus and structural equivalence. Therefore, introduce ranking functions between sets of nodes:

$$r_{XY} : X \rightarrow \mathbb{R}^{|Y|}$$

where $X, Y \in \{D, T\}$.

Heterogeneous ranking ($X \neq Y$):

- Connection strength between nodes in different sets (partitions)
- Adapt *tf-idf* to the term-date graph:

$$tf(d, t) \equiv \omega(d, t)$$

$$idf(t) \equiv \frac{|D|}{deg(t)}$$

$$tf-idf(d, t) := \omega(d, t) \log \frac{|D|}{deg(t)}$$

Homogeneous ranking ($X = Y$)

- Similarity within the same node set
- No direct links available
- Leverage existing links between sets
- Use for example a cosine similarity of adjacency vectors:

$$\cos(t_a, t_b) := \frac{\sum t_{a_i} \cdot t_{b_i}}{\sqrt{\sum t_{a_i}^2 \cdot \sum t_{b_i}^2}}$$

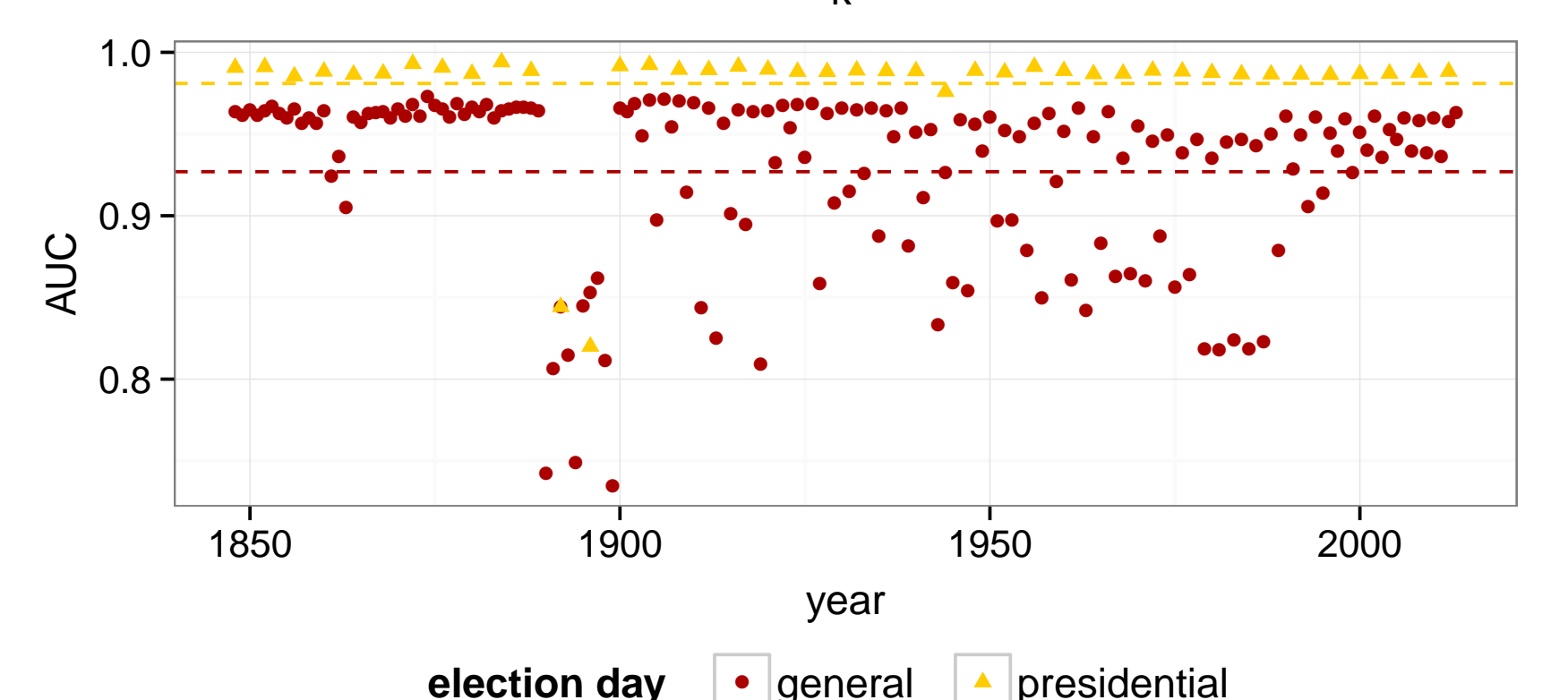
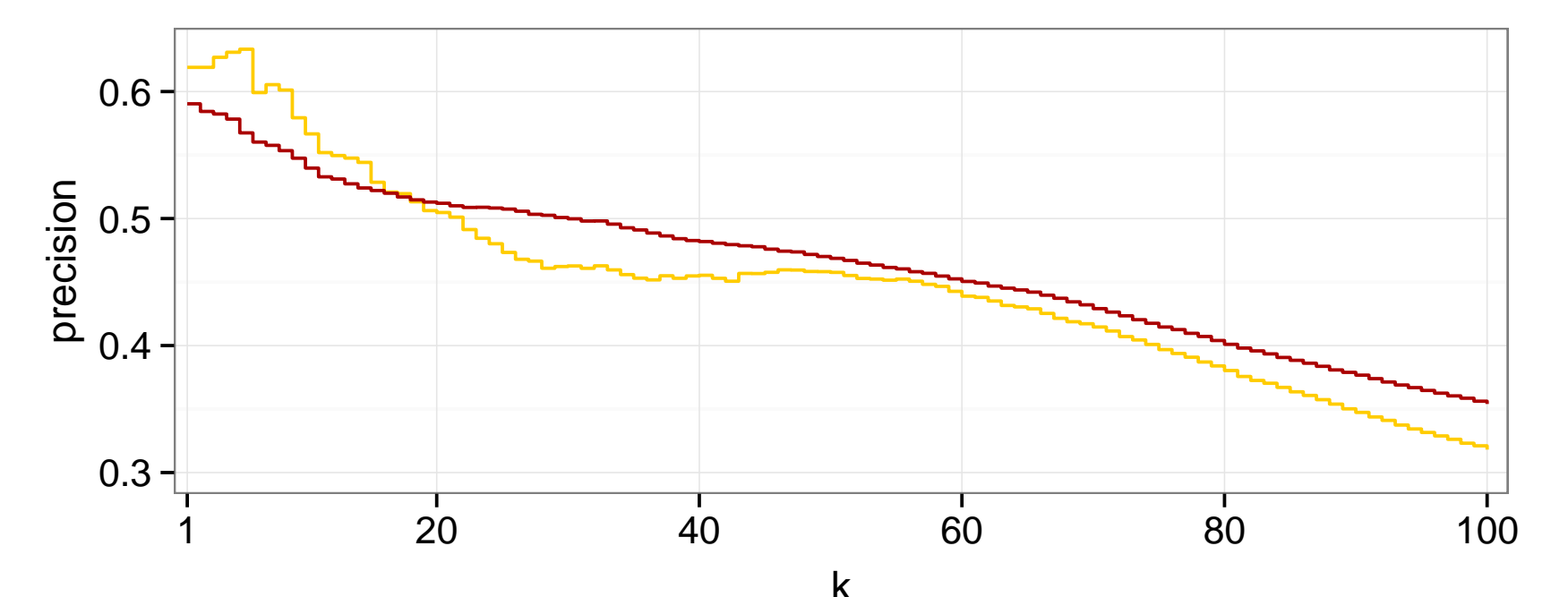
Examples of rankings:

Query: "1215-06-15"				Query: "Tsunami"			
	tf-idf	ω	deg(t)		tf-idf	ω	deg(t)
carta	79.7	14	709	2004	3097.2	1374	393475
magna	71.2	14	1298	2011	2753.9	1313	460264
barons	46.9	10	1928	2011-03	1878.5	464	65407
runnymede	40.5	6	247	2004-12-26	1658.0	238	3536
king	20.4	12	38400	2011-03-11	1474.2	226	5508
oaths	17.1	3	714	2005	1030.6	476	430107
king's	15.1	5	10200	2004-12	734.8	162	40186
repudiation	13.6	2	231	2005-01	465.5	102	39062
fealty	12.4	2	424	2006	301.7	147	481555
john	11.8	11	71893	2010	295.2	148	510254

Evaluation

Ranking for U.S. Election Days

- Annually, varies between Nov 2 - Nov 8
- Presidential election every 4 years
- Idea: similar dates to Election Days are also Election Days (in different years)

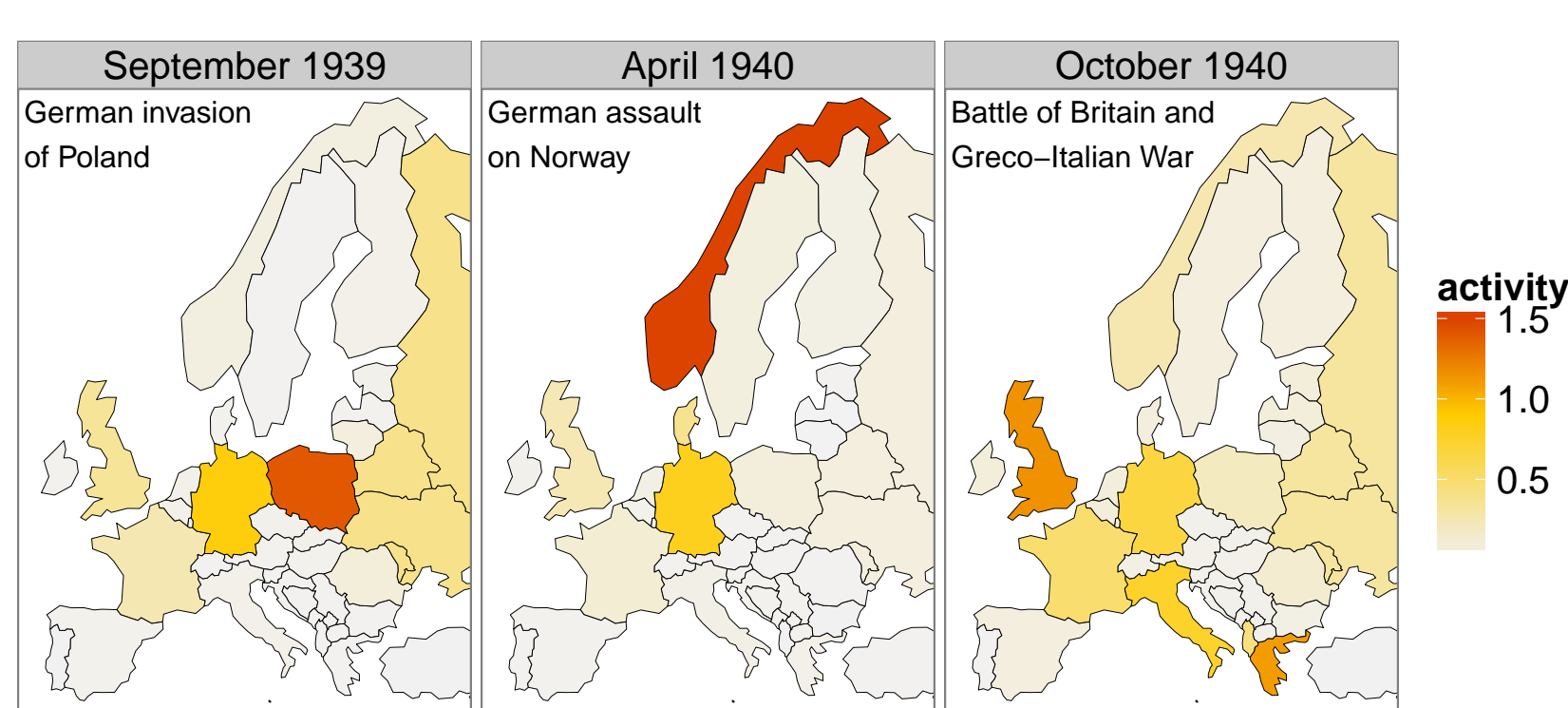


Application: Identifying Hot-Spots

Pinpoint events during World War II:

- Select a set of countries c
- For each country, define the corresponding name $t_n(c)$ and adjective $t_a(c)$ (e.g., *Italy* and *Italian*)
- Compute an activity score from the individual *tf-idf* scores for a given date of interest and highlight accordingly

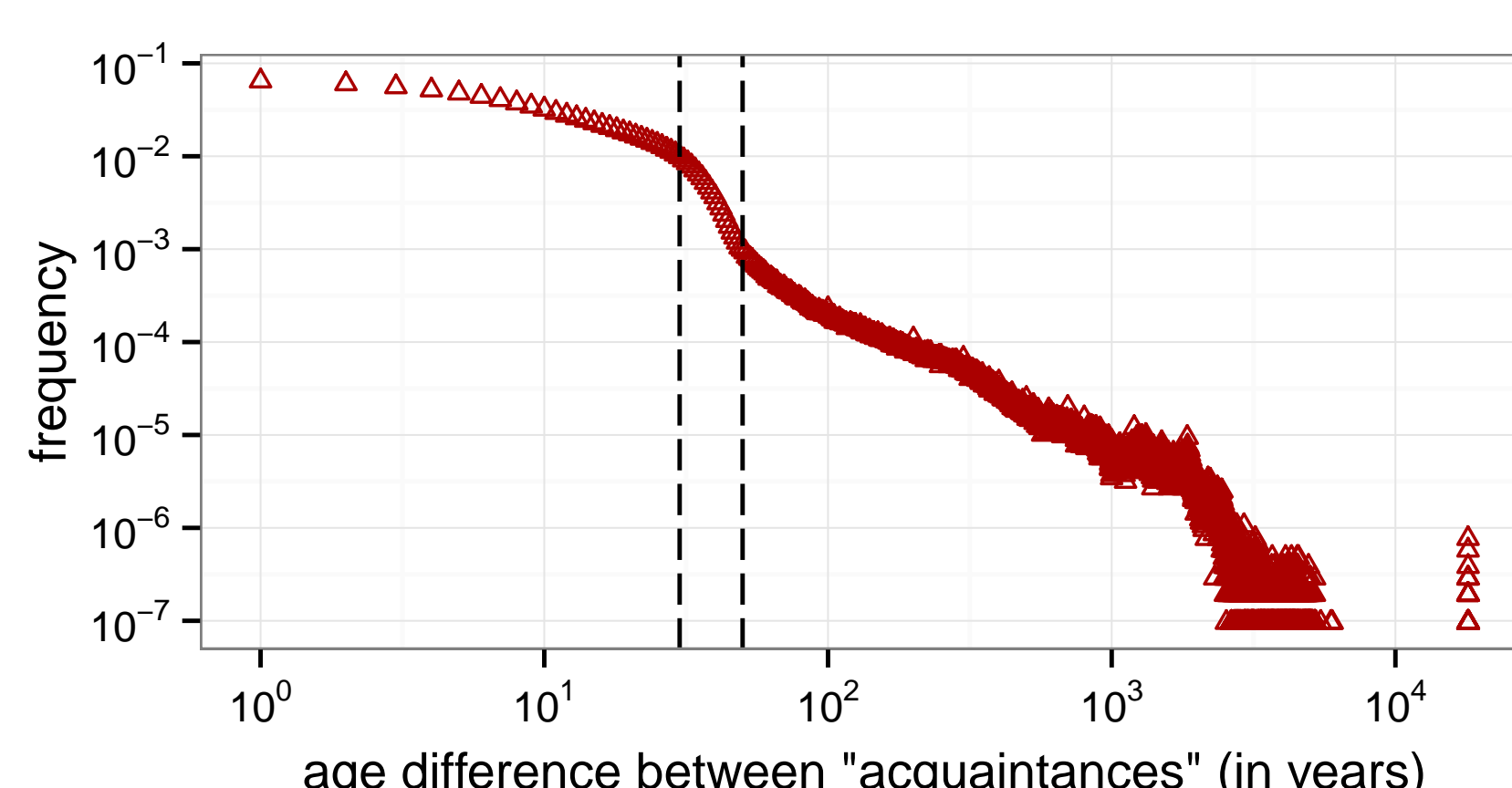
$$act(c, d) := \frac{tf-idf(d, t_n(c)) + tf-idf(d, t_a(c))}{\max[tf-idf(d, \cdot)]}$$



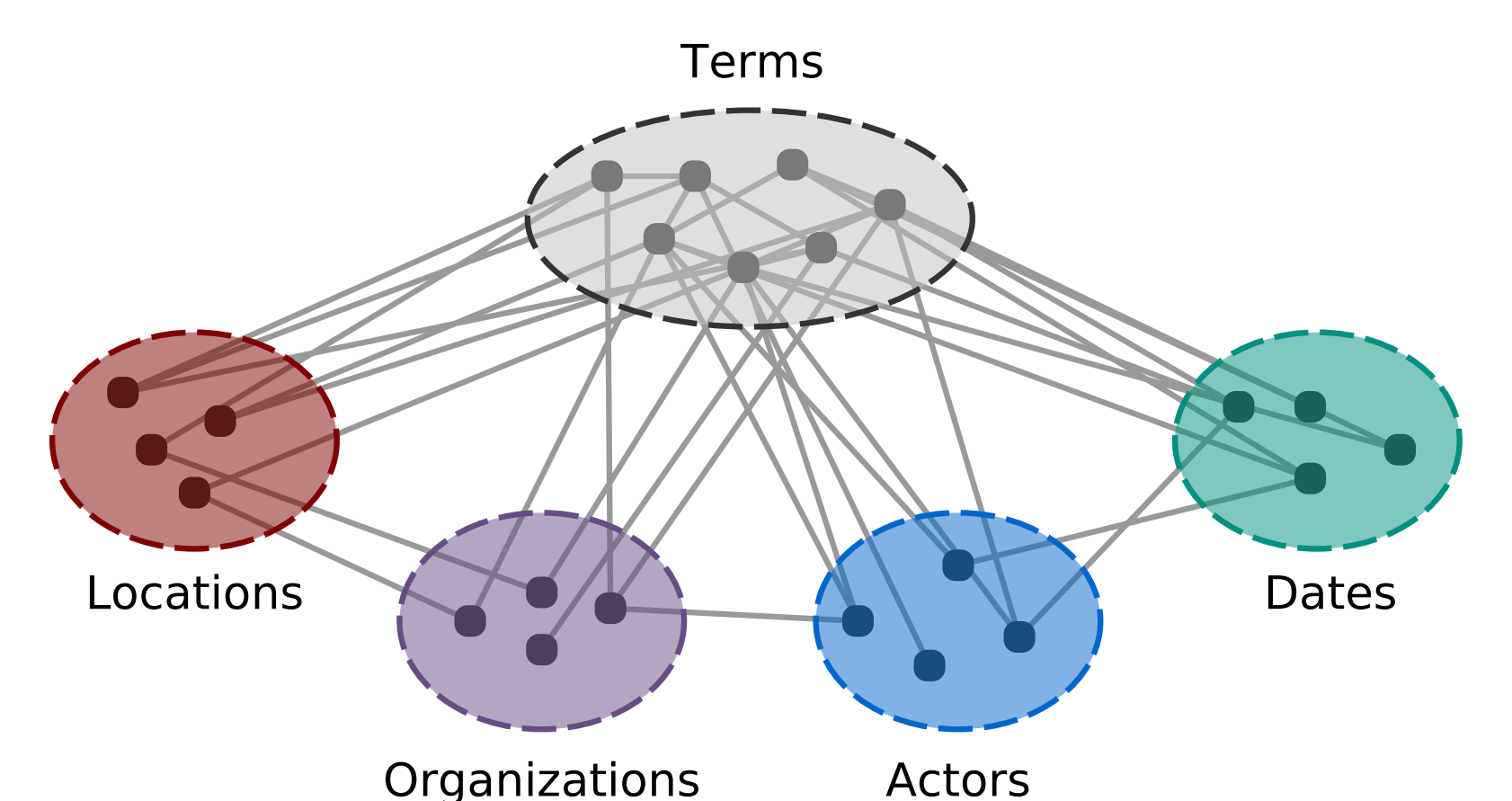
Outlook: Weight decay by distance

Beyond a bag of words:

- Consider co-occurrences in the entire document, not just in sentences
- Use a weight function for edges that decays with distance
- Aggregate individual co-occurrences
- When applied to the construction of a social network from person mentions on Wikipedia: results in relationships with natural age profiles [2].



Outlook: The LOAD Model



Extension to a Multi-Partite Graph

- Include *spatial information*
- Include *persons*
- Include *organizations*
- Include relationships between terms

Identifying Events:

- Based on incomplete information
- Without need for disambiguation

References

- [1] A. Spitz, J. Strötgen, T. Bögel and M. Gertz: **Terms in Time and Times in Context: A Graph-based Term-Time Ranking Model**. *WWW '15 Companion*, 1375-1380, 2015
- [2] J. Geiß, A. Spitz and M. Gertz: **Beyond Friendships and Followers: The Wikipedia Social Network**. *ASONAM '15*, 2015

This work was presented at the 3rd Heidelberg Laureate Forum, 23.08 – 28.08, 2015, Heidelberg, Germany.

Contact Information:

Andreas Spitz

spitz@informatik.uni-heidelberg.de

<http://dbs.ifi.uni-heidelberg.de/>

