

TimeMachine

TIMELINE GENERATION FOR KNOWLEDGE-BASE ENTITIES

Tim Althoff (Stanford CS),
Xin Luna Dong, Kevin Murphy, Safa Alai,
Van Dang, Wei Zhang (Google)

Really common scenario!

Learning about new topics is hard

**Even learning about a
simple topic like a person!**



Web Search Results

Search results for "robert downey jr" showing 34M documents. The results include a Wikipedia entry, news articles, and an IMDb profile.

robert downey jr

Web News

34M documents

About 34,000,000

Robert Downey, Jr. - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Robert_Downey,_Jr. - Wikipedia -
Robert John Downey, Jr. (born April 4, 1965) is an American actor whose career has included critical and popular success in his youth, followed by a period of ...
Robert Downey, Sr. - Robert Downey Jr. filmography - Susan Downey - Chaplin

In the news

 **How Much Did Robert Downey Jr Make for 'Avengers 2' & 'Captain America: Civil War'?**
/FILM - 10 hours ago
A couple years ago, Robert Downey Jr. hinted that he might be done with superhero movies ...

Robert Downey Jr. Sent Chris Pratt The Greatest 'Jurassic World' Meme
Huffington Post - 2 days ago

Robert Downey Jr. Makes A Startling Amount Of Money From Marvel
Cinema Blend - 3 hours ago

[More news for robert downey jr](#)

Robert Downey Jr. - IMDb
www.imdb.com/name/nm0000375 - Internet Movie Database
Robert Downey Jr., Actor: The Avengers. Robert Downey Jr. has the most respected actors in Hollywood. With an amazing list of cre

Results don't provide representative view

Robert Downey Jr (@RobertDowneyJr) | Twitter

<https://twitter.com/robertdowneyjr>
357 tweets • 178 photos/videos • 4.93M followers. "This might get out of hand, @prattprattpratt... <http://t.co/ZsFCGOdFC7>"

Robert Downey Jr | Facebook

<https://www.facebook.com/robertdowneyjr>
Robert Downey Jr. 25344451 likes · 1916960 talking about this. The Official Page of Robert Downey Jr Twitter: <http://rdj.lk/twitter> Instagram:...

Robert Downey, Jr. (@robertdowneyjr) • Instagram photos ...

<https://instagram.com/robertdowneyjr?hl=en>
Official Instagram of Robert Downey, Jr. Facebook: <https://www.facebook.com/robertdowneyjr> Twitter: <http://rdj.lk/twitter>

Doesn't support exploration of topic & history

Structured Information

Manual curation



Robert Downey Jr.

Actor

Robert John Downey, Jr. is an American actor whose career has included critical and popular success in his youth, followed by a period of substance abuse and legal troubles, and a resurgence of commercial success in middle age. [Wikipedia](#)

Born: April 4, 1965 (age 50), Manhattan, New York City, NY

Height: 5' 9" (1.74 m)

Spouse: Susan Downey (m. 2005), Deborah Falconer (m. 1992–2004)

Parents: Elsie Ford, Robert Downey, Jr.

Children: Indio Falconer Downey, Exton Elias Downey, Avri Roel Downey

Bias towards recent blockbusters

Movies

[View 45+ more](#)



The Avengers
2012



Avengers:
Age of
Ultron
2015



Iron Man
2008



Iron Man 3
2013



The Judge
2014

People also search for

[View 10+ more](#)



Susan
Downey
Spouse



Robert
Downey, Sr.
Father



Scarlett
Johansson



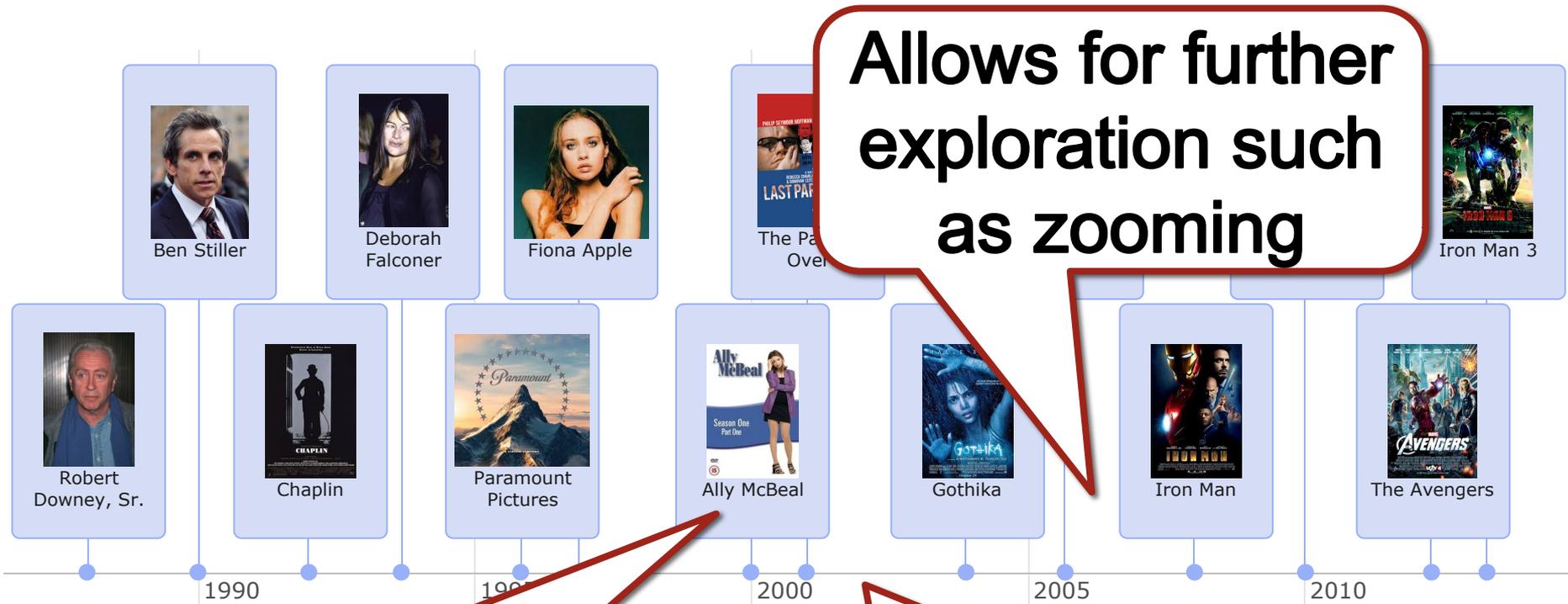
Chris Evans



Chris
Hemsworth

Missing context
of relationship

Our System: TimeMachine



Allows for further exploration such as zooming

“RDJr was award-nominated for Ally McBeal in 2000.”

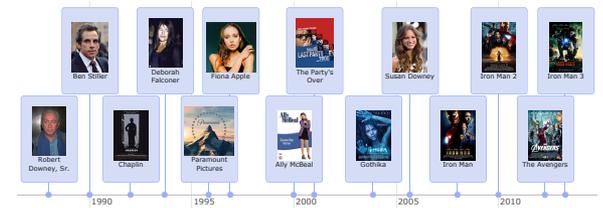
Better overview through timeline visualization

Timeline Generation

Problem Definition



Entity



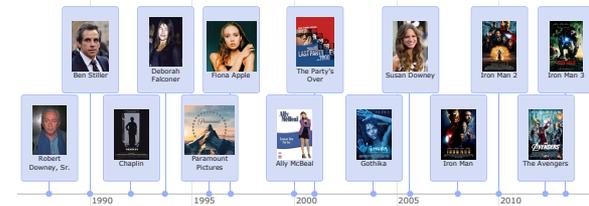
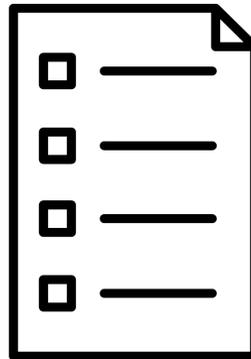
Timeline

- Timeline needs to be rendered on device
- Generation must be fast (zoom + interaction)

Timeline Generation Approach

1. Event Generation

2. Event Selection



Entity

List of events

Timeline

List of Events

Subject	Related Entity	Time	Description
R. D. Jr	Robert D. Sr	1988	In movie directed by
	Jon Bon Jovi	2004	TV show app. with
	Susan Downey	2005	Got married
	Iron Man	2008	Award for movie
	Avengers	2010	Acted in movie

1. Event Generation

Candidate Event Generation



Robert Downey Jr.

DoB

April 4, 1965

Subject

Related Entity

Timestamp

Resulting Candidate Events

Subject	Related Entity	Time	Description
R. D. Jr	-	1965	Was born
	The Avengers	2012	Acted in movie
	Samuel L Jackson	2012	Co-starred in The Avengers



737 events total

Event Filtering

- Some non-informative events:
 - any American
 - nationality: USA
 - founded 1776
- **Frequency Filter:** events commonly associated with a large number of subjects are unlikely to be interesting (like IDF) (no “nationality → founded”)
- **Existence Filter:** filter out events before entity begins to exist (no “parent → DOB”)



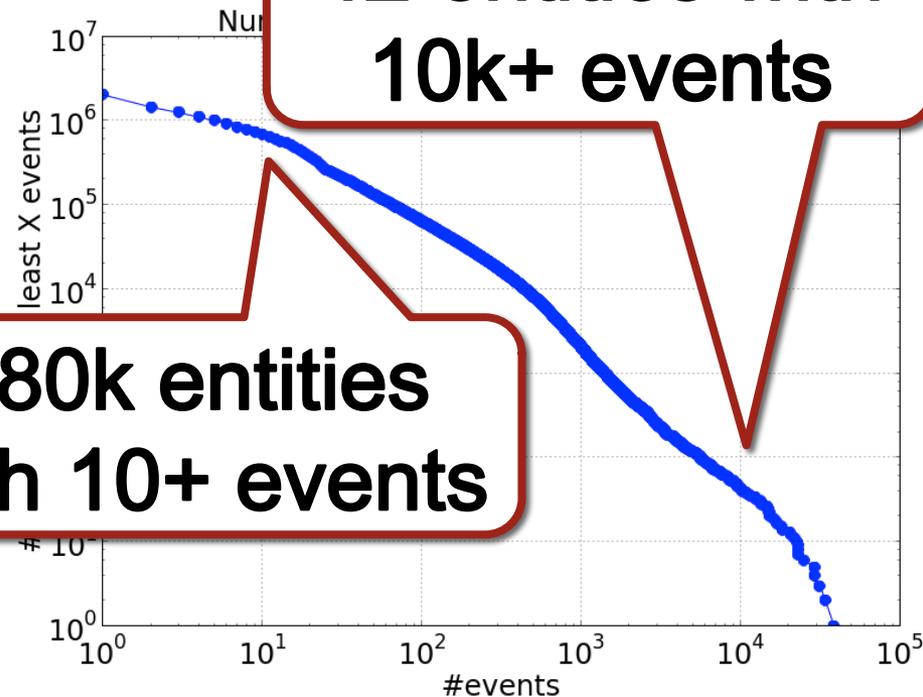
Event Filtering Evaluation

- 87% precision (two indep. raters)
- Generate many candidate events (Freebase)

Event Selection becomes very important but challenging!

680k entities with 10+ events

42 entities with 10k+ events



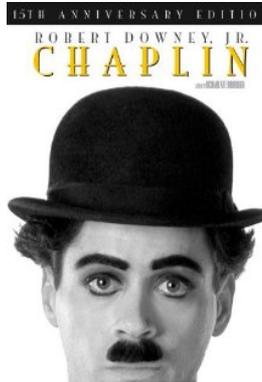
2. Event Selection

Timeline Quality Criteria

1. Correctness



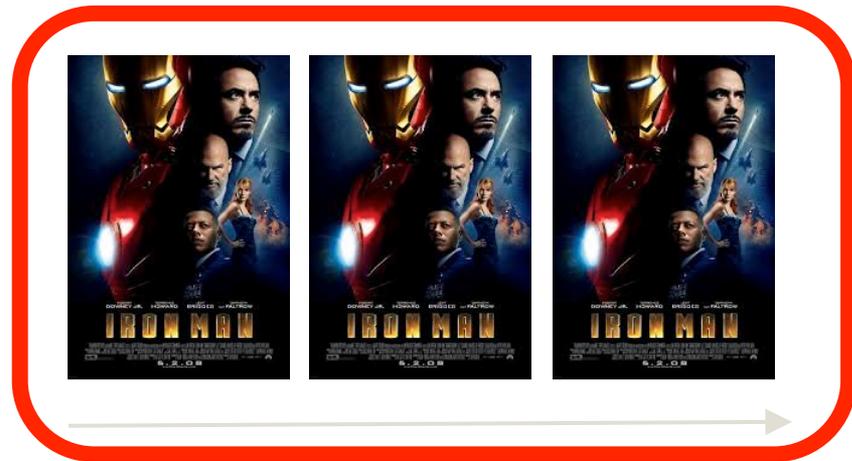
2. Relevance



Timeline Quality Criteria (cont.)

3. Content Diversity

Encourage selection of different entities

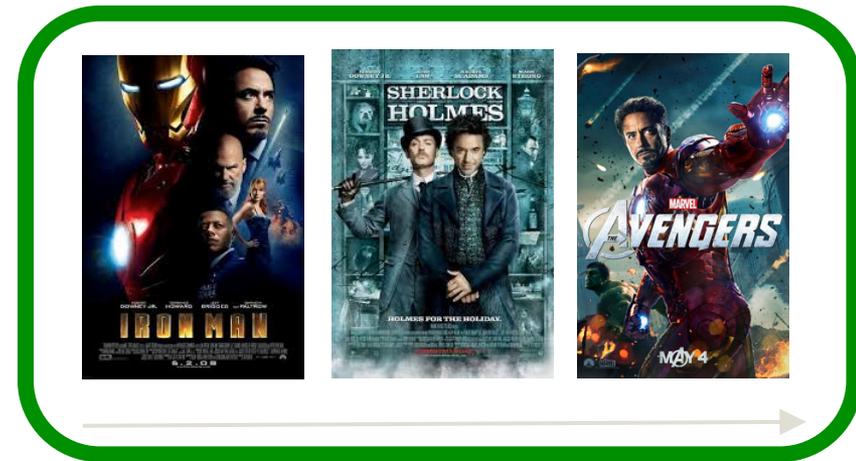


US Release

Award

EU Release

VS



Award

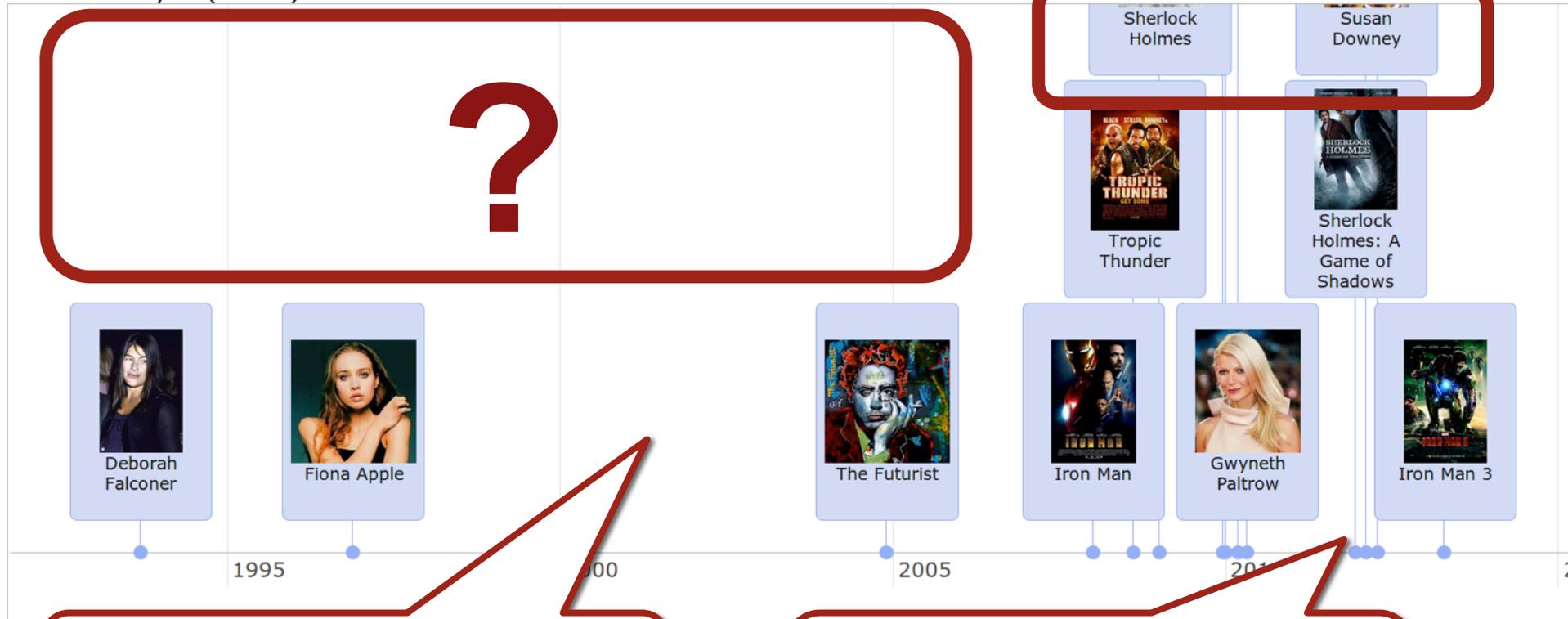
Award

US Release

Timeline Quality Criteria (cont.)

4. Temporal Diversity / Layout

Robert Downey Jr. (1965—)



Looks unaesthetic

Hard to interact with events

Optimization Problem

Candidate Set:
1. Correctness
(by construction)

Objective:
2. Relevance &
3. Content Diversity

$$T^* = \arg \max_{T \subseteq E} \textit{Relevance}(T)$$

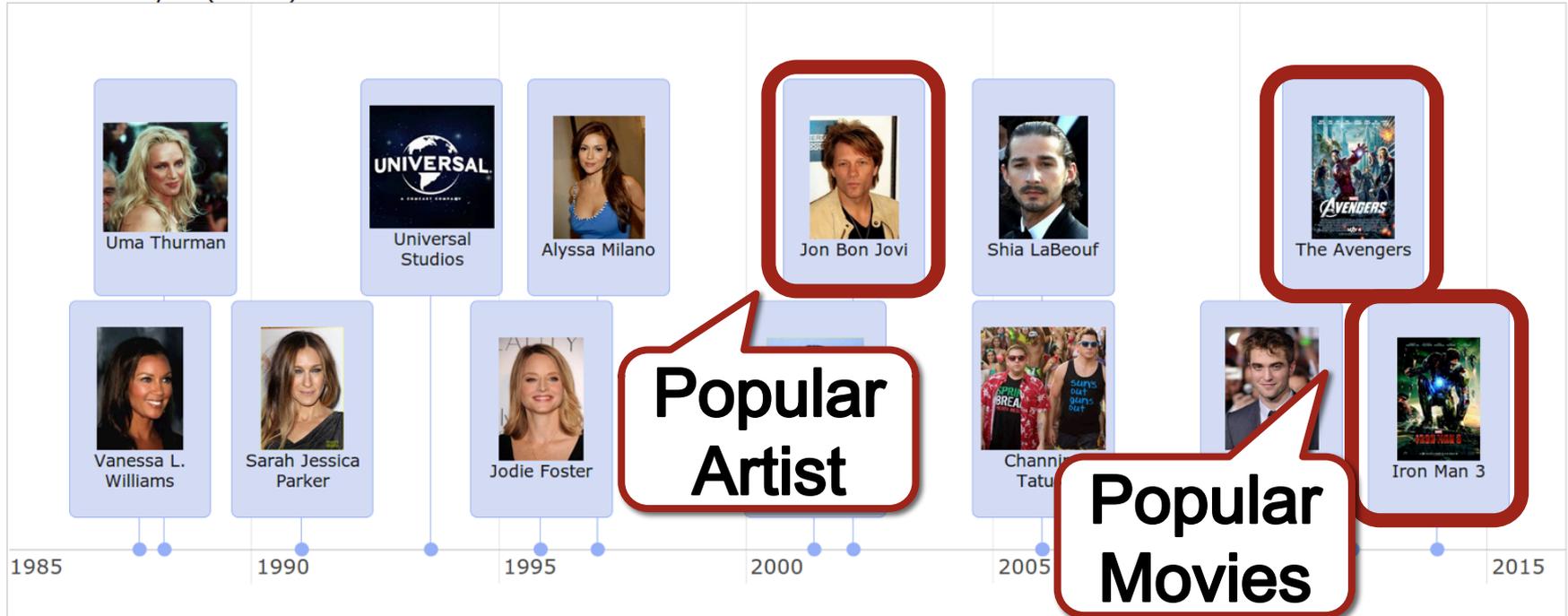
s.t. *Temporal Diversity*(T)

Constraint:
4. Temporal Diversity

(2) Relevance Signals

- Baseline: global relevance signal
- Based on # search queries for entity

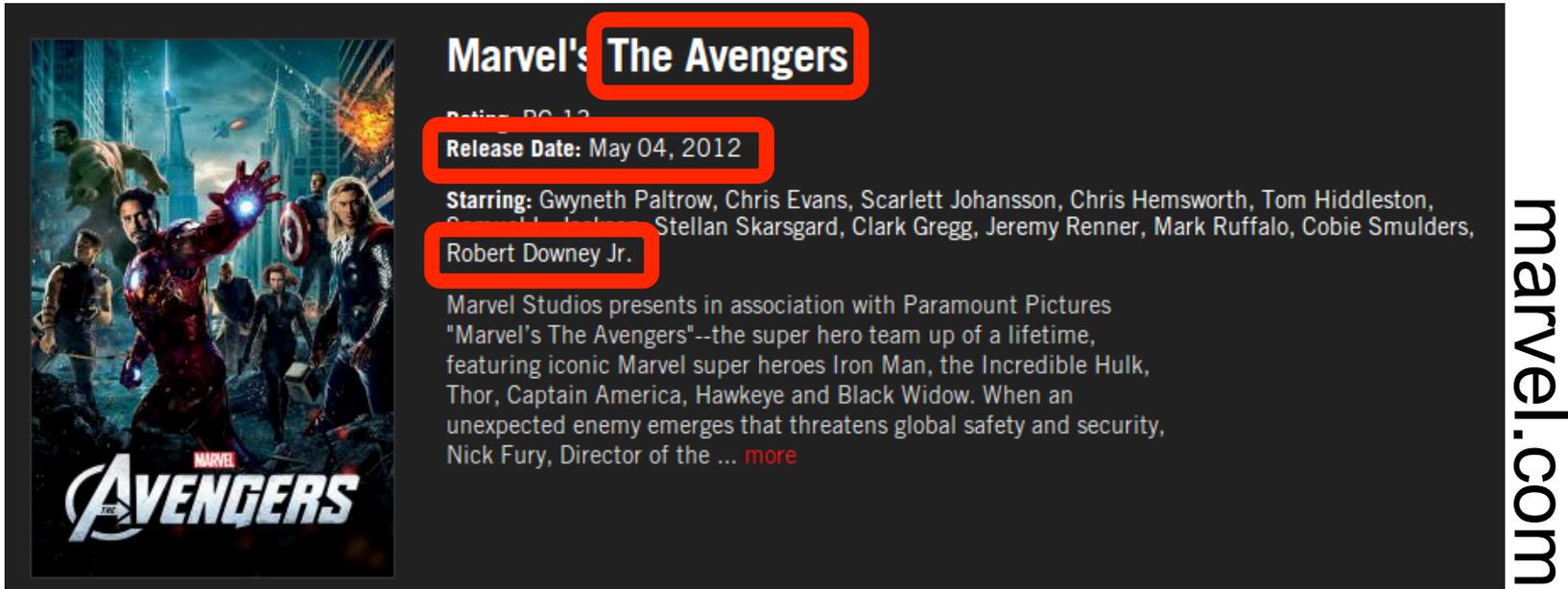
Robert Downey Jr. (1965—)



- Biased towards popular but unspecific events

How to improve relevance signal

- Use **co-occurrences** on web scale



Marvel's **The Avengers**

Rating: PG-13

Release Date: May 04, 2012

Starring: Gwyneth Paltrow, Chris Evans, Scarlett Johansson, Chris Hemsworth, Tom Hiddleston, Samuel L. Jackson, Stellan Skarsgard, Clark Gregg, Jeremy Renner, Mark Ruffalo, Cobie Smulders, **Robert Downey Jr.**

Marvel Studios presents in association with Paramount Pictures "Marvel's The Avengers"--the super hero team up of a lifetime, featuring iconic Marvel super heroes Iron Man, the Incredible Hulk, Thor, Captain America, Hawkeye and Black Widow. When an unexpected enemy emerges that threatens global safety and security, Nick Fury, Director of the ... [more](#)

marvel.com

- “Robert Downey Jr” and “May 4, 2012” occurs 173 times on 71 different webpages
- US Release date of *The Avengers*

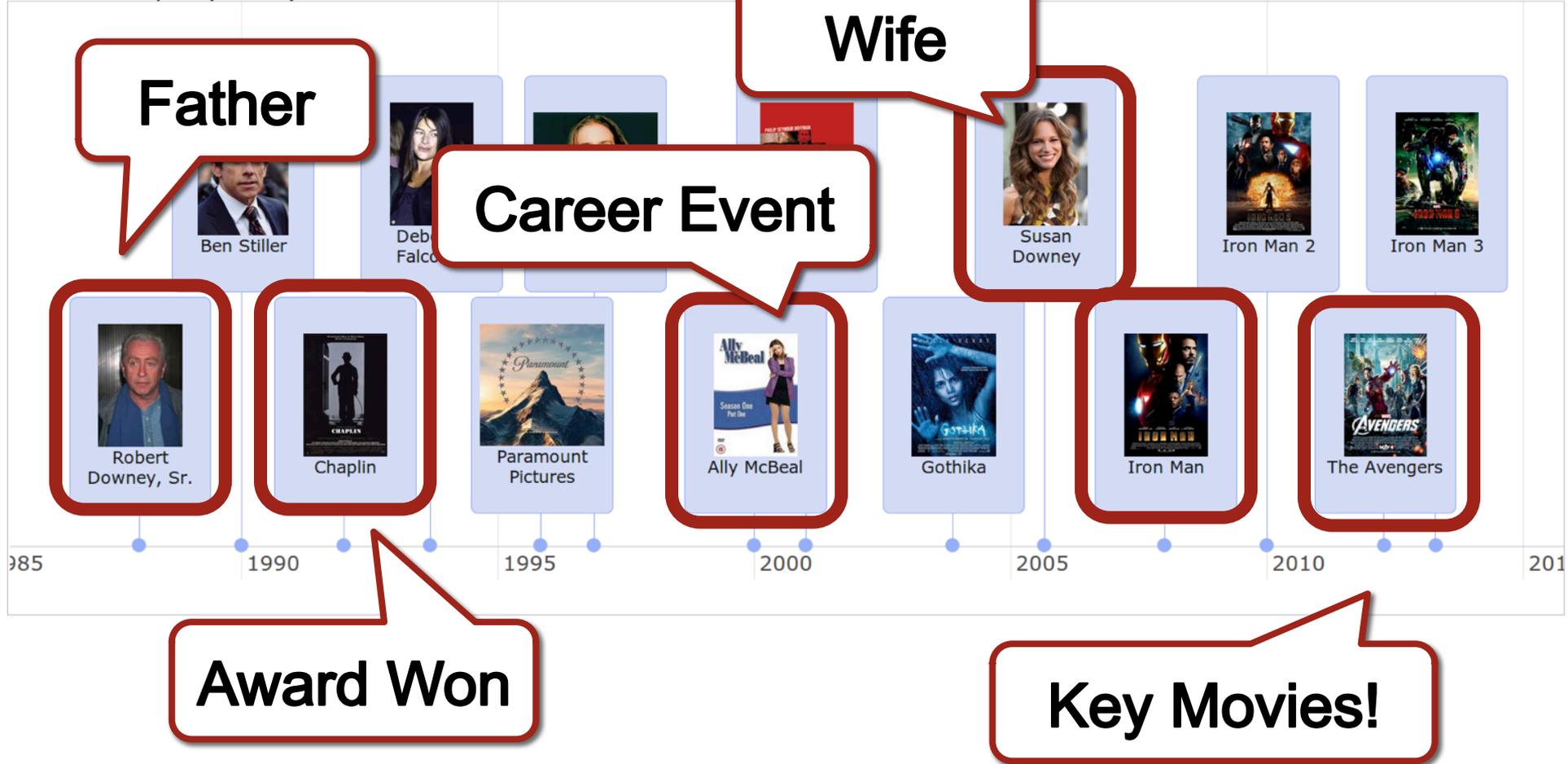
Assigning scores to entities/dates

- Run NLP tools across **large web corpus of 10B documents** (NER + CoRef)
- **Extract entity-entity and entity-date co-occurrences** within small windows
- **Normalize counts** using Normalized Pointwise Mutual Information (NPMI)
 - accounts for popular entities/dates



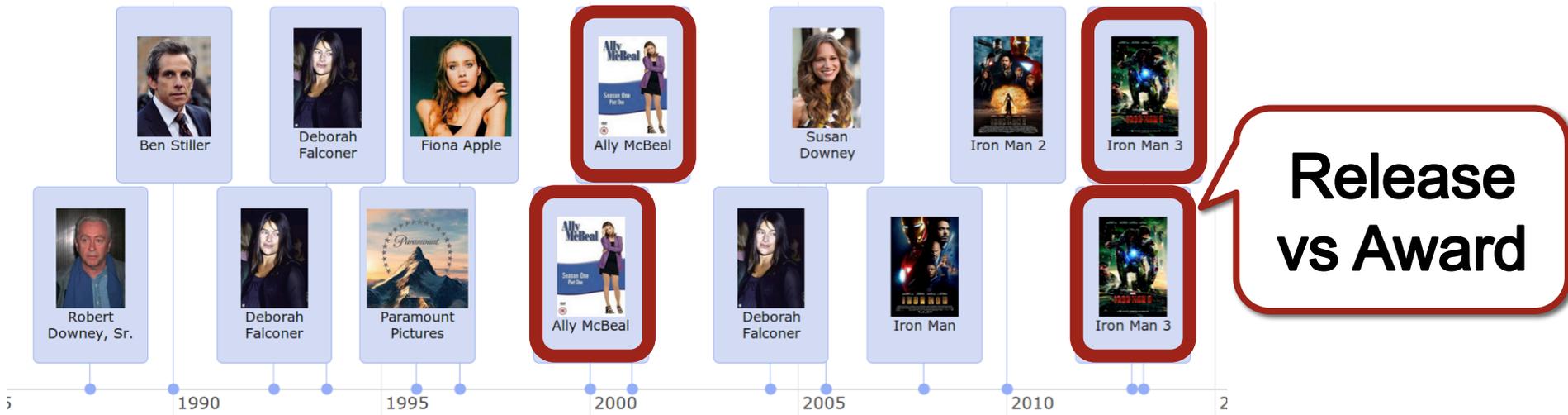
Improvements from New Signal

Robert Downey Jr. (1965—)



- Large improvement over global relevance

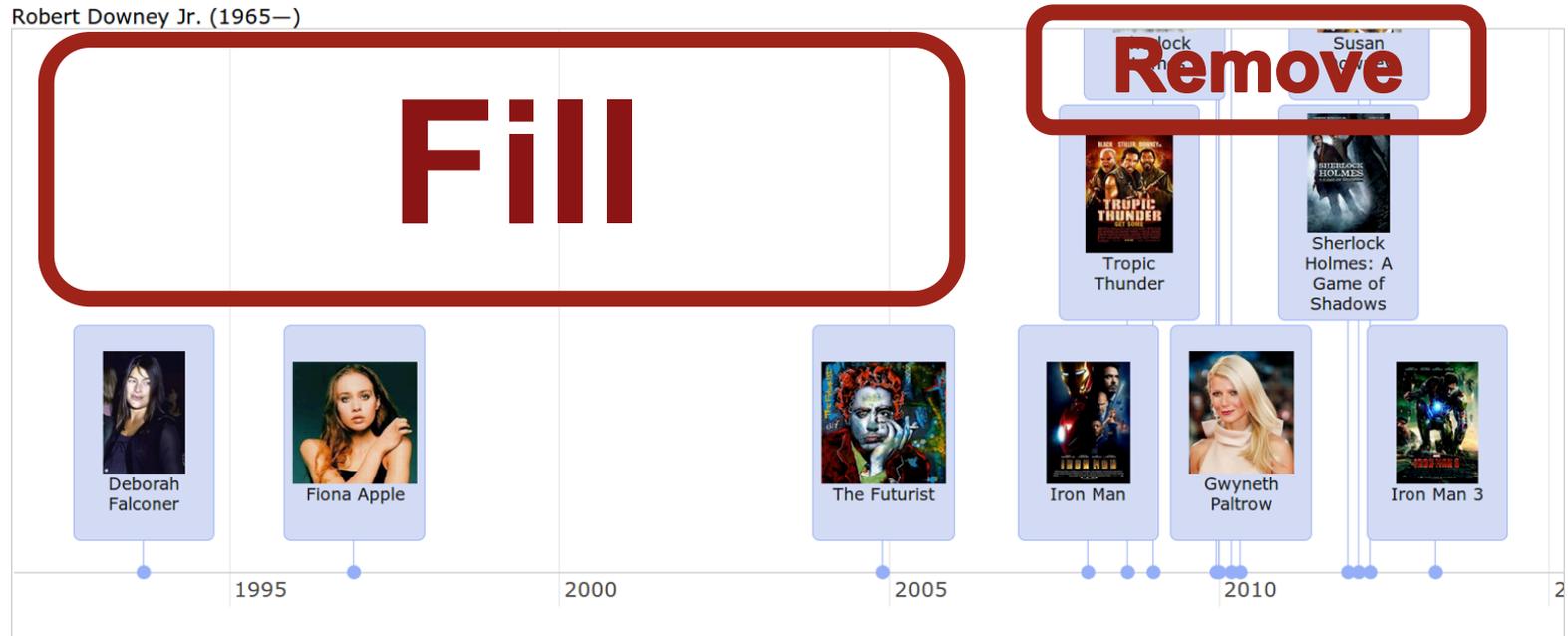
(3) Content Diversity



Encode diminishing returns in objective function:



(4) Temporal Diversity



- **Constraint** in optimization problem: “Event boxes cannot overlap.”
- Enforce **balanced layout** during optimization

Submodular Optimization

$$T^* = \arg \max_{T \subseteq E} \text{Relevance}(T)$$

$$\text{s.t. } \text{TemporalDiversity}(T)$$

- Relevance(T) is **submodular**
 - “Diminishing returns”
 - Less reward for adding to “bigger set”

$X, Y \subseteq \Omega, X \subseteq Y, \text{ and } x \in \Omega \setminus Y:$

$$f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y)$$

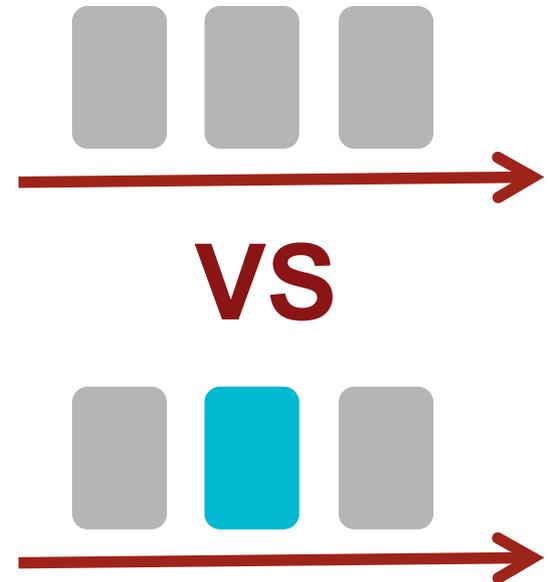
Algorithm & Theoretical Results

- **Fast approximation** through **lazy-greedy**
 - Allows for zoom and interaction
- **Provable approximation guarantee**
 - Worst-case: 33% of optimal solution!
 - **Still holds with complex constraint!**
 - We prove: constraint structure induces independence family that is a **p-system**
(Calinescu et al. 2011)

3. Evaluation

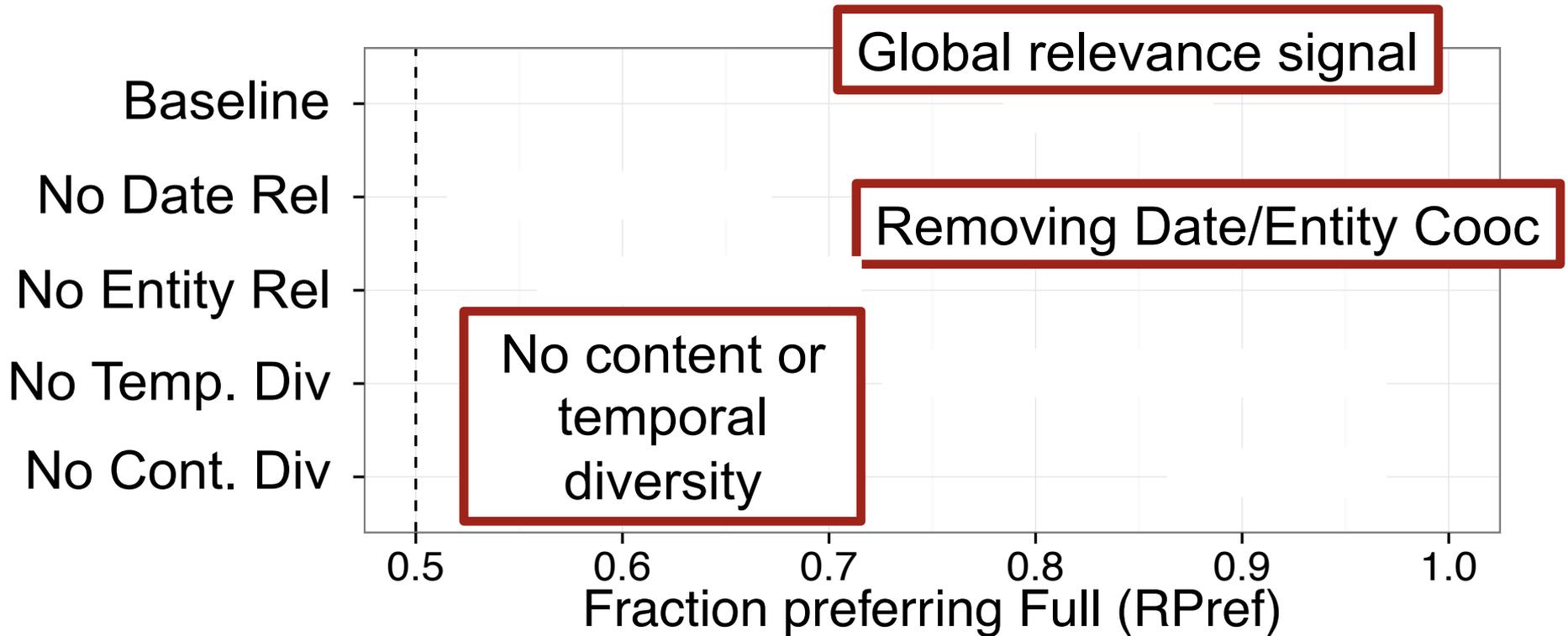
Experimental Evaluation

- **User study** on Amazon Mechanical Turk
 - Rating of timelines challenging, possibly subjective, and no ground truth available
- **Pairwise comparisons** using 250 entities
 - Relative judgements
 - Explanations
- **Large-scale**
 - >1200 raters
 - >6000 tasks



Results

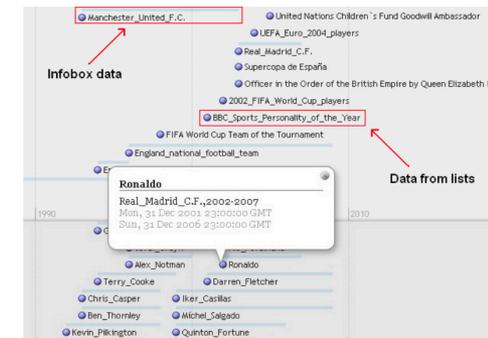
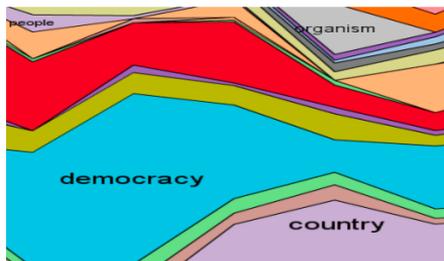
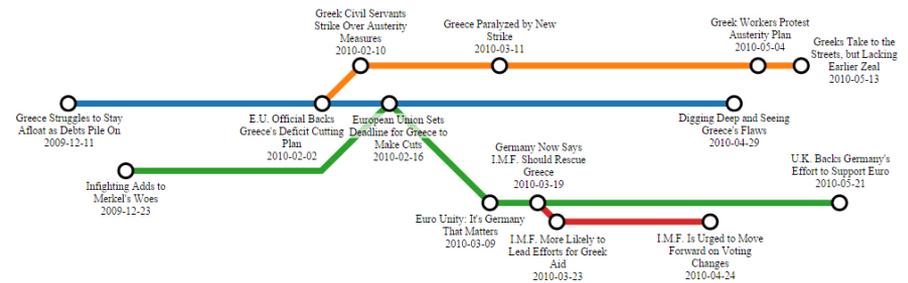
In what fraction of cases do raters prefer our full method over the ablated baselines?



Full = “everything” = Global + Web Cooc + TD + CD

Related Work

- Document summarization: (Allan et al. 2001)
- Submodular optimization: (Krause & Golovin 2014), (Calinescu et al. 2011), (Nemhauser et al. 1978)
- Maps of information: (Shahaf et al. 2013)
- Timelines based on knowledge bases: (Mazeika et al. 2011), (Tuan et al. 2011), (Wang et al. 2010)



Conclusions

- **TimeMachine:** Automatic timeline generation for knowledge base entities
- **Submodular optimization framework**
 - Jointly optimizes for relevance, content diversity, and temporal diversity
 - Proved near-optimal performance guarantees
- **User studies** show
 - Web-based co-occurrence signals improve over baseline model (global importance)
 - Temporal and content diversity are crucial

Thanks!

Contact

@timalthoff althoff@cs.stanford.edu

Check out the demo!

cs.stanford.edu/~althoff/timemachine

Paper / Proofs / Slides

cs.stanford.edu/~althoff

Acknowledgements

Evgeniy Gabrilovich, Arun Chaganty, Stefanie Jegelka, Karthik Raman, Sujith Ravi, Ravi Kumar, Jeff Tamer, Patri Friedman, Danila Sinopalnikov, Alexander Lyashuk, Jure Leskovec, David Hallac, Caroline Suen