TimeMachine:
Timeline Generation for Knowledge-Base Entities

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

Google

The Problem

- Google is great but not perfect: **Learning about some new topic's history remains challenging**
- Search engines do not provide **representative view** or support **exploration of topic** and its relationships

The Solution: Timeline Visualization

- **Summarizes** most relevant events / relationships
- **Interactive exploration**, e.g. **zoom** and topic-switch
- **Adapts** to available screen estate
- **Our task:** Given entity, generate timeline

Timeline for Robert Downey Jr.

"RDJr was award-nominated for Ally McBeal in 2000."

1990 1995 2000 2005 2010
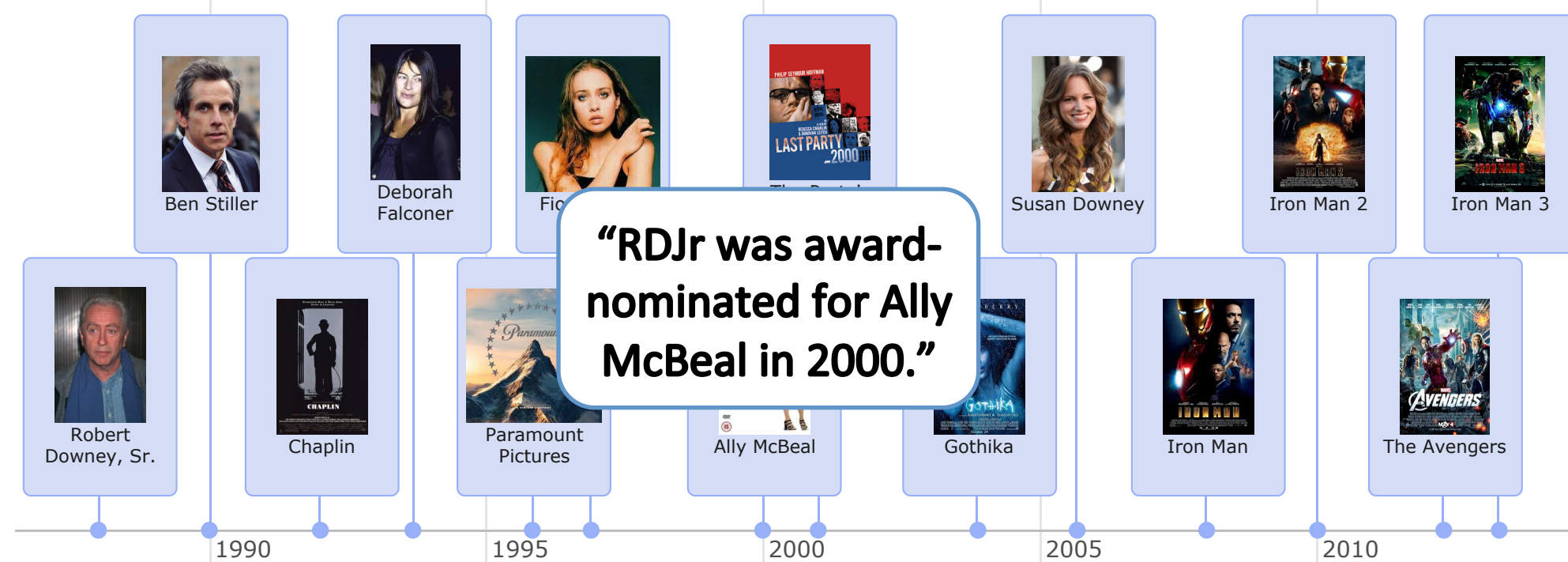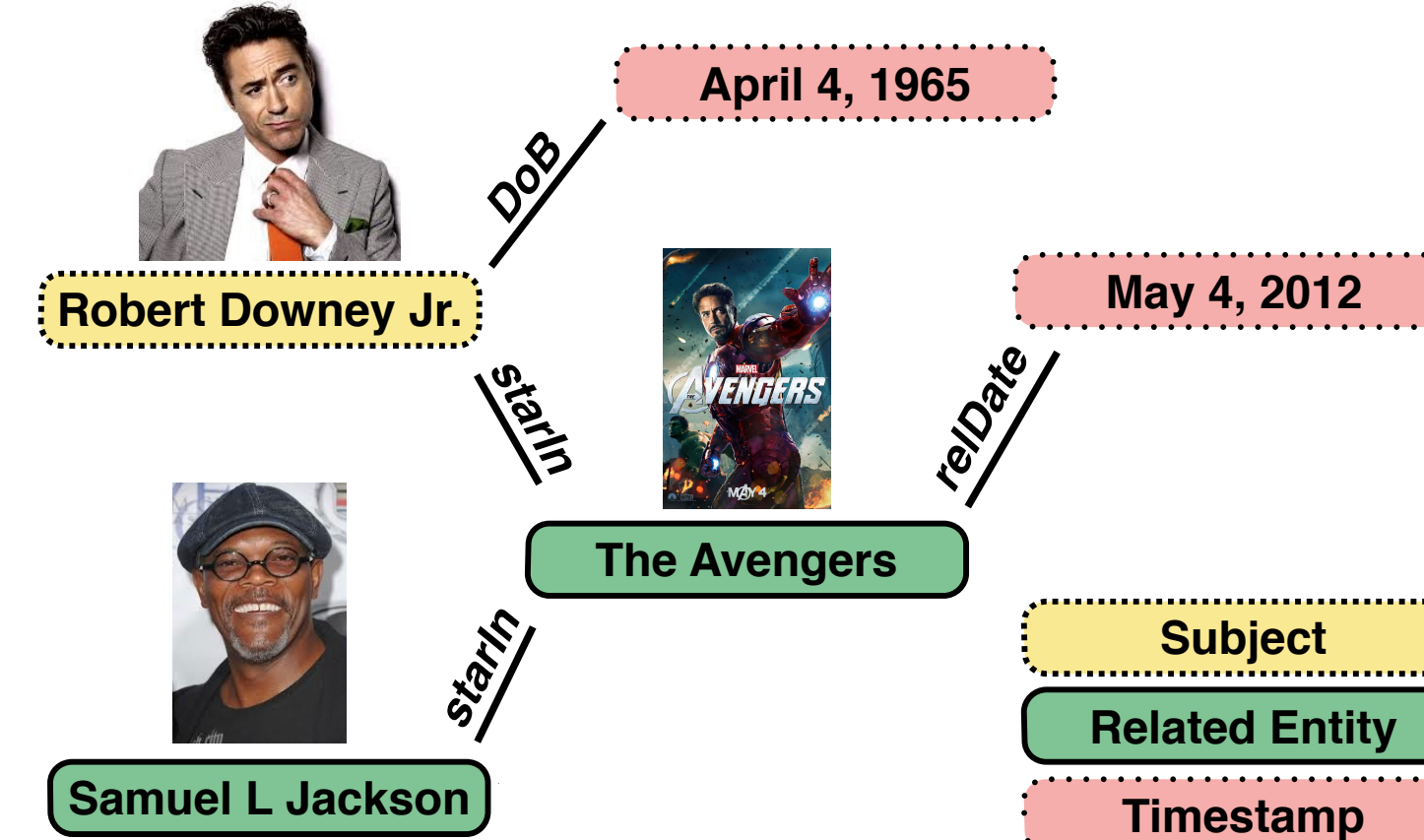
Quality Criteria

1. **Correctness:** Only show actual events/relationships
2. **Relevance:** Only display most "relevant" events
3. **Content Diversity:** Display diverse set of event types
4. **Temporal Diversity:** Produce balanced layout

Our Approach

Entity → Event Generation → List of events → Event Selection → Timeline

Step 1: Event Generation

- Use knowledge base to generate events

Robert Downey Jr. —DoB— April 4, 1965
The Avengers —relDate— May 4, 2012
Samuel L Jackson —starIn—

Subject
Related Entity
Timestamp

- Filtering step to exclude non-informative events

Step 2: Event Selection

Optimization Problem

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

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

1. Correctness (Event Generation)
- Guaranteed by knowledge base & construction

2. Relevance Signals (Objective)
- **Global:** # of search queries for entity
- Entity-Entity & Entity-Date **co-occurrence** from 10B document web corpus (NER + CoRef + NPMI)

3. Content Diversity (Objective)
- Encode **diminishing returns** in objective

4. Temporal Diversity (Constraint)
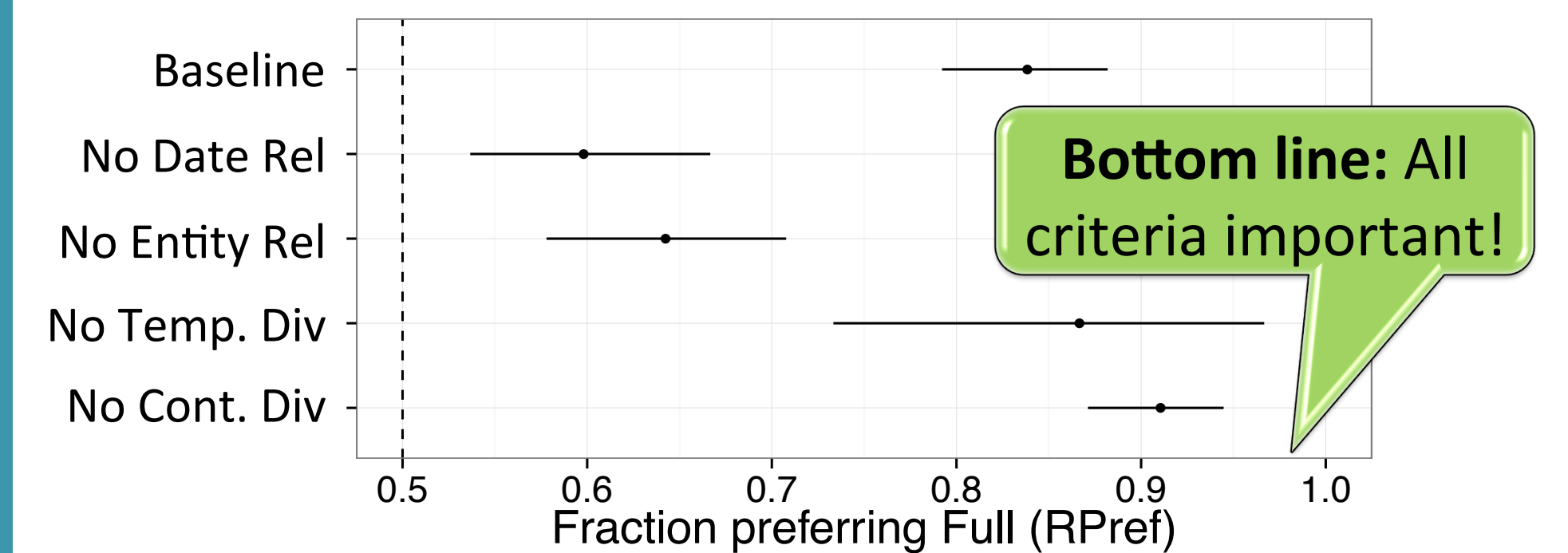- Enforce **balanced layout** during optimization as constraint

Provable approximation guarantee
- Submodular optimization + p-system constraint
- Lazy-greedy algorithm: ≥33% of optimal solution

Experimental Evaluation

- **User studies** on AMT (>1200 raters)
- **Pairwise comparisons** → relative judgments

Baseline
No Date Rel
No Entity Rel
No Temp. Div
No Cont. Div

0.5 0.6 0.7 0.8 0.9 1.0
Fraction preferring Full (RPref)

Bottom line: All criteria important!

- Web-based co-occurrence signals **strongly improve over baseline** (global importance)
- **Temporal and content diversity are crucial** ingredients for good timelines

Another Example: John F Kennedy

"JFK served as President of the United States from 1961."

1958 1959 1960 1961 1962 1963 1964

Conclusions

- **Goal**: Scalable timeline generation
- **Challenge**: Jointly optimize for relevance, content diversity, and temporal diversity
- **Efficient** algorithms with **theoretical guarantees**
- **User studies** show that all criteria are important
- **Demo!** cs.stanford.edu/~althoff/timemachine

Thanks Dafna!