TEMPORAL WEB DYNAMICS AND ITS APPLICATION TO INFORMATION RETRIEVAL

Yi Chang, Fernando Diaz, Anlei Dong, Susan Dumais, Kira Radinsky, Milad Shokouhi

Web content dynamics

WSDM 2013 Tutorial

Schedule

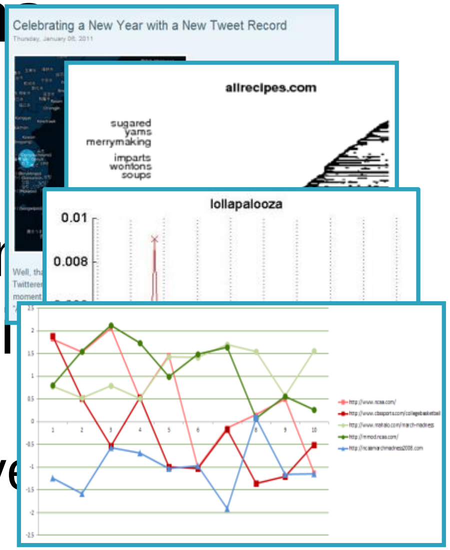
- Introduction (9:00-9:15)
- Modeling Dynamics
  - 9:15-10:15 Web content dynamics [Susan]
  - 10:15-11:15 Web user behavior dynamics [Milad]
  - 11:15-11:30 Break
  - 11:30-13:00 Spatio-temporal analysis [Fernando]
  - Methods for evaluation
- Lunch (13:00-14:30)
- Applications to Information Retrieval
  - 14:30-15:45 Temporal NLP [Kira]
  - News event prediction
  - 15:45-16:00 Break
  - 16:00-17:45 Time-sensitive search [Yi]
  - Time-sensitive recommendations [Anlei]
- Wrap-Up (17:45-18:00)

Schedule

- Introduction (9:00-9:15)
- Modeling Dynamics
  - 9:15-10:15 Web content dynamics [Susan]
  - 10:15-11:15 Web user behavior dynamics [Milad]
  - 11:15-11:30 Break
  - 11:30-13:00 Spatio-temporal analysis [Fernando]
  - Methods for evaluation
- Lunch (13:00-14:30)
- Applications to Information Retrieval
  - 14:30-15:45 Temporal NLP [Kira]
  - News event prediction
  - 15:45-16:00 Break
  - 16:00-17:45 Time-sensitive search [Yi]
  - Time-sensitive recommendations [Anlei]
- Wrap-Up (17:45-18:00)

Time and Time Again …

- Time is pervasive in information system
  - New documents appear all the time
  - Document content changes over time
  - Queries and query volume change over ti
  - What's relevant to a query changes over ti
    - E.g., *U.S. Open 2013* (in June vs. Sept)
    - E.g., *U.S. Open 2013* (before, during, after eve
  - User interaction changes over time
    - E.g., anchor text, "likes", query-click streams, social networks, etc.
  - Relations between entities change over time
    - E.g., President of the U.S. is <> [in 2012 vs. 2004]
- … yet, most information retrieval systems ignore time !



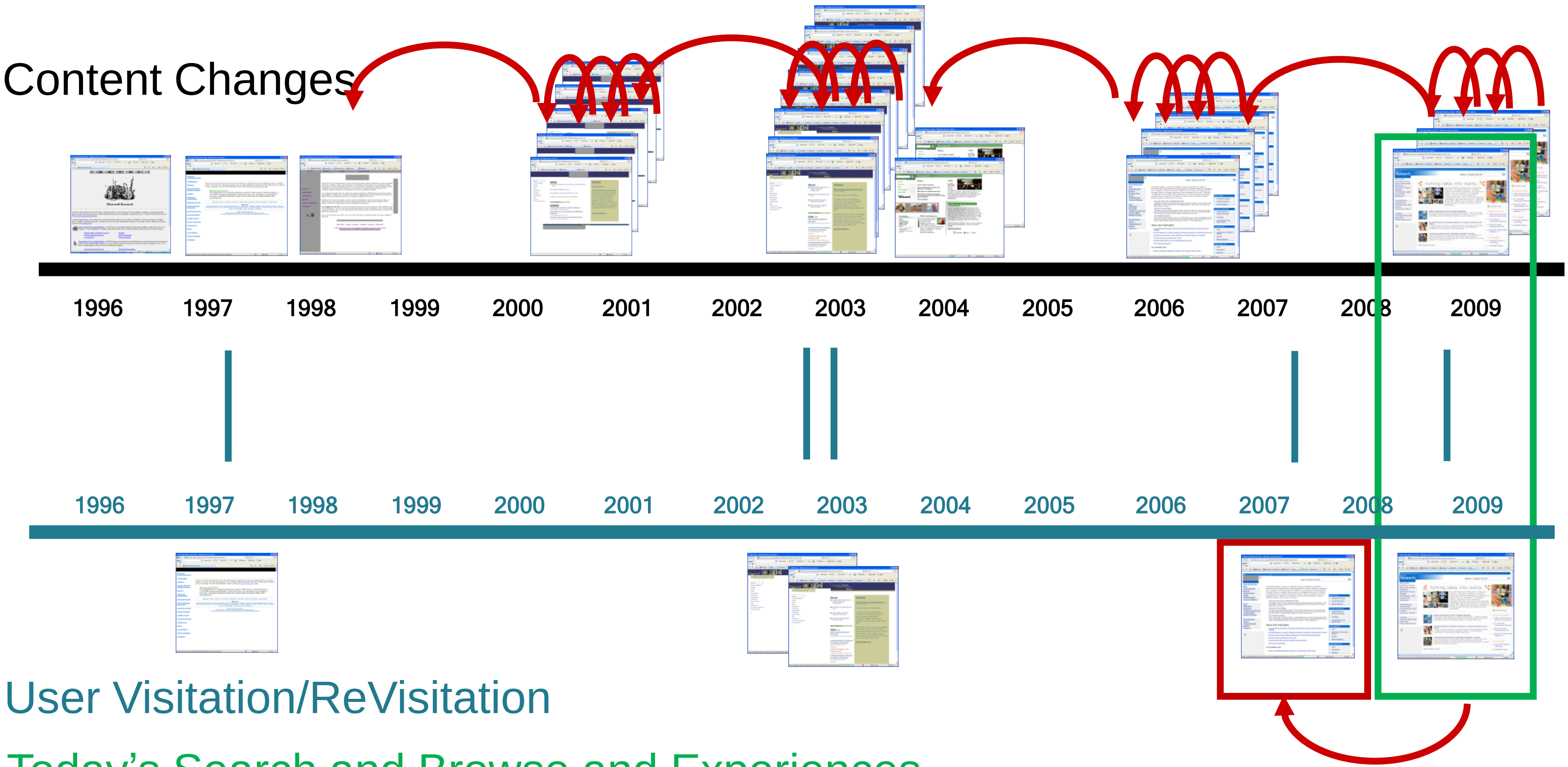
Web Content Dynamics

- Overview
- Change in “persistent” web documents
  - Characterizing content dynamics
  - Systems and applications
- Change in “real-time” content streams
  - Characterizing content dynamics
  - Systems and applications
- Change in Web graphs
  - Web graph evolution
  - Authority and content over time

Content Dynamics

- Easy to capture
- But … few tools or algorithms support dynamics

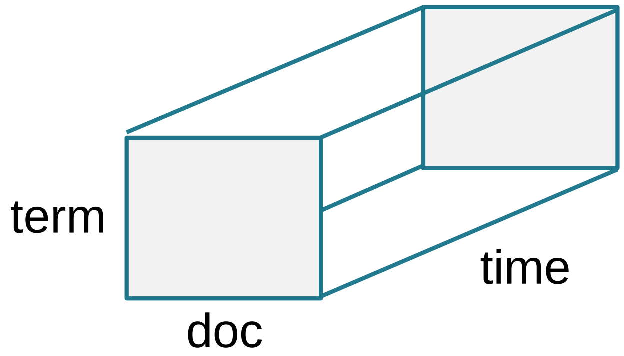
Content Dynamics



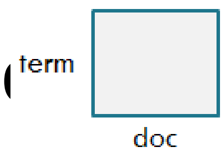
Today’s Search and Browse and Experiences

But, ignores …

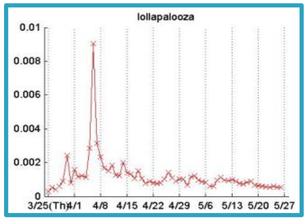
Content Dynamics



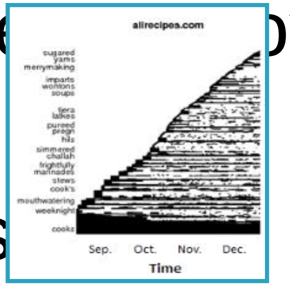
- Traditional IR: single snap sh



- Word/query trends: aggregates over docs



- Document change: aggre over terms
- (Word,Document) trends



Content Dynamics

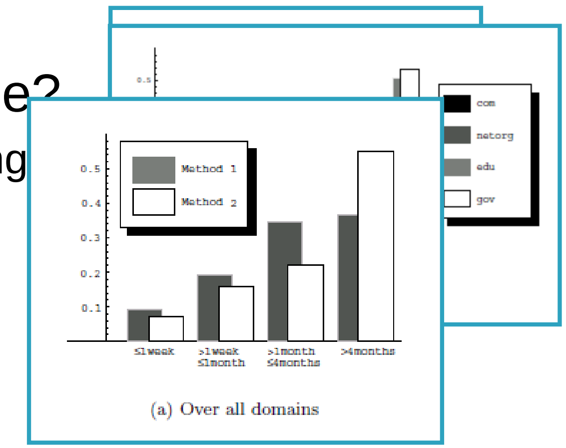
- Types of content
  - Persistent documents (E.g., Web pages that persist over time)
  - Real-time streams (E.g., Twitter, Facebook, blogs)
  - Somewhere in between (E.g., the Web, Wikipedia)
- How content change is discovered
  - Crawling
  - Feeds
  - Wikis

Web Content Dynamics

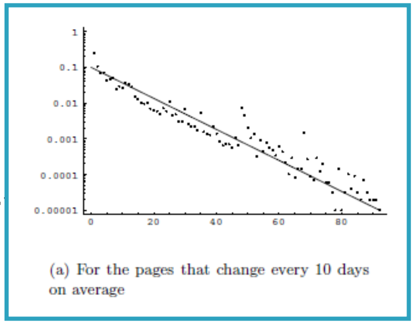
- Overview
- Change in “persistent” web documents
  - Characterizing content dynamics
    - Page-level changes
    - Within-page changes
  - Systems and applications
- Change in “real-time” content streams
  - Characterizing content dynamics
  - Systems and applications
- Change in Web graphs
  - Web graph evolution
  - Authority and content over time

Web Crawling: Cho & Garcia-Molina

- Crawled 720k pages (from 270 popular sites), once per day, 4 months
  - How often does a web page change?
    - 23% change every day; 30% never chang
    - Differs by domain
  - What is the lifespan of a page?
    - ~10% < 1 week; 50% > 4 months
  - Model when a page will change
    - Poisson process - a sequence of random events, occur independently, at a fixed rate λver time ( )
    - PDF : $$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t > 0 \\ 0 & \text{for } t \leq 0 \end{cases}$$
    - Also, Radinsky & Bennett (WSDM 201
  - Use to improved crawling policy



(a) Over all domains

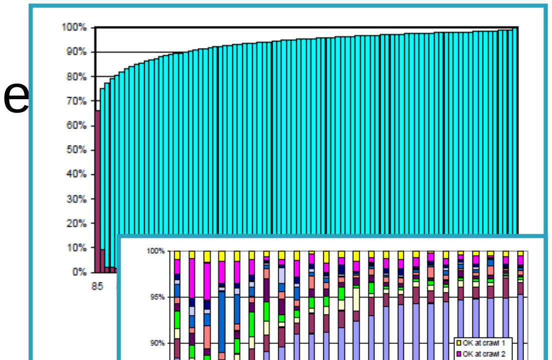


(a) For the pages that change every 10 days on average

J. Cho and H. Garcia-Molina. The evolution of the web and implications for an incremental crawler. VLDB 20

Web Crawling: Fetterly et al.

- Crawled 150m pages (seed Yahoo! home page), once per week, 11 weeks
  - How often does a web page change
    - 67% never changed
  - When was last successful crawl?
    - Avg, 88% on last crawl
    - Varies by domain (.cn 79%, .dk/.gov 95%
  - How much does a web page change?
    - Avg, (~4% >med, 20% small, 10% no text, 67% no char



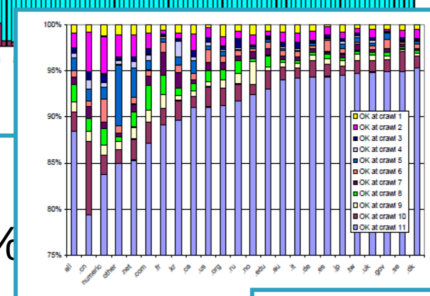


Figure 6: Breakdown showing in last successfully downloaded, brok

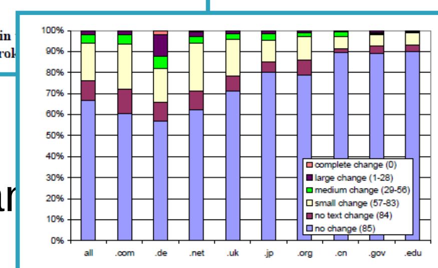
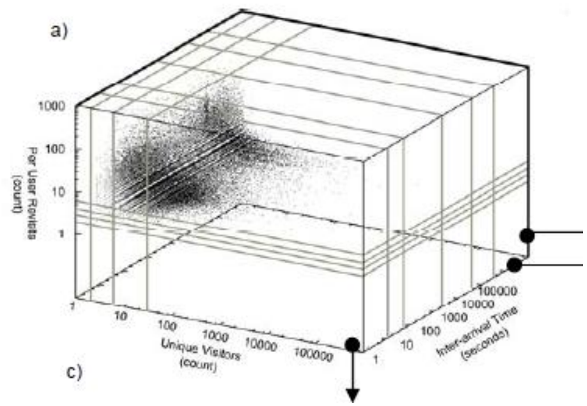


Figure 11: Clustered rates of change, broken down by selected top-level domains, after excluding automatically generated keyword-spam documents.

D. Fetterly, M. Manasse, M. Najork and J. Weiner. A large-scale study of the evolution of web pages. WWW 20

Web Crawling: Adar et al.

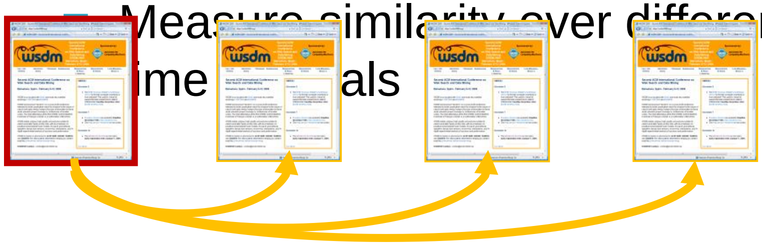
- Crawled 50k pages (usage-sensitive sample), once per hour (at least), 5 weeks
  - Usage-sensitive sample
    - Number of unique users
    - Re-visits per user
    - Inter-visit interval
  - Summary page-level metrics
  - Detailed within-page changes, term longevity
  - Applications to Ranking and UX (Diff-IE)

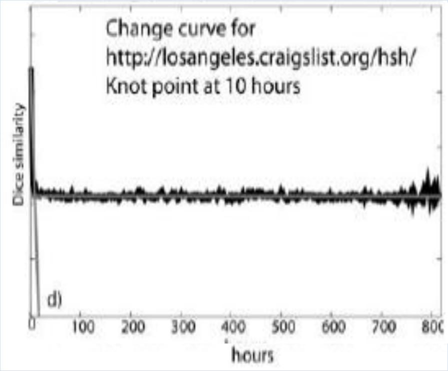


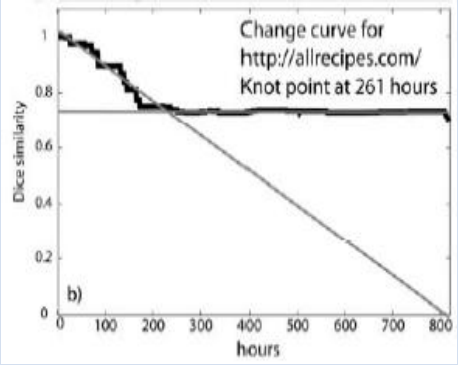
E. Adar, J. Teevan, S. T. Dumais and J. Elsas. The web changes everything: Understanding the dynamics of web content. WSDM 2009.

Adar et al.: Page-level Change

- Summary metrics
  - 67% of visited pages changed
    63% of these changed every hour
  - Popular pages change more frequently, but not by much
  - .com pages change at intermediate frequency, but by more
- Change curves
  - Fixed starting point
  - Measure similarity over different time intervals



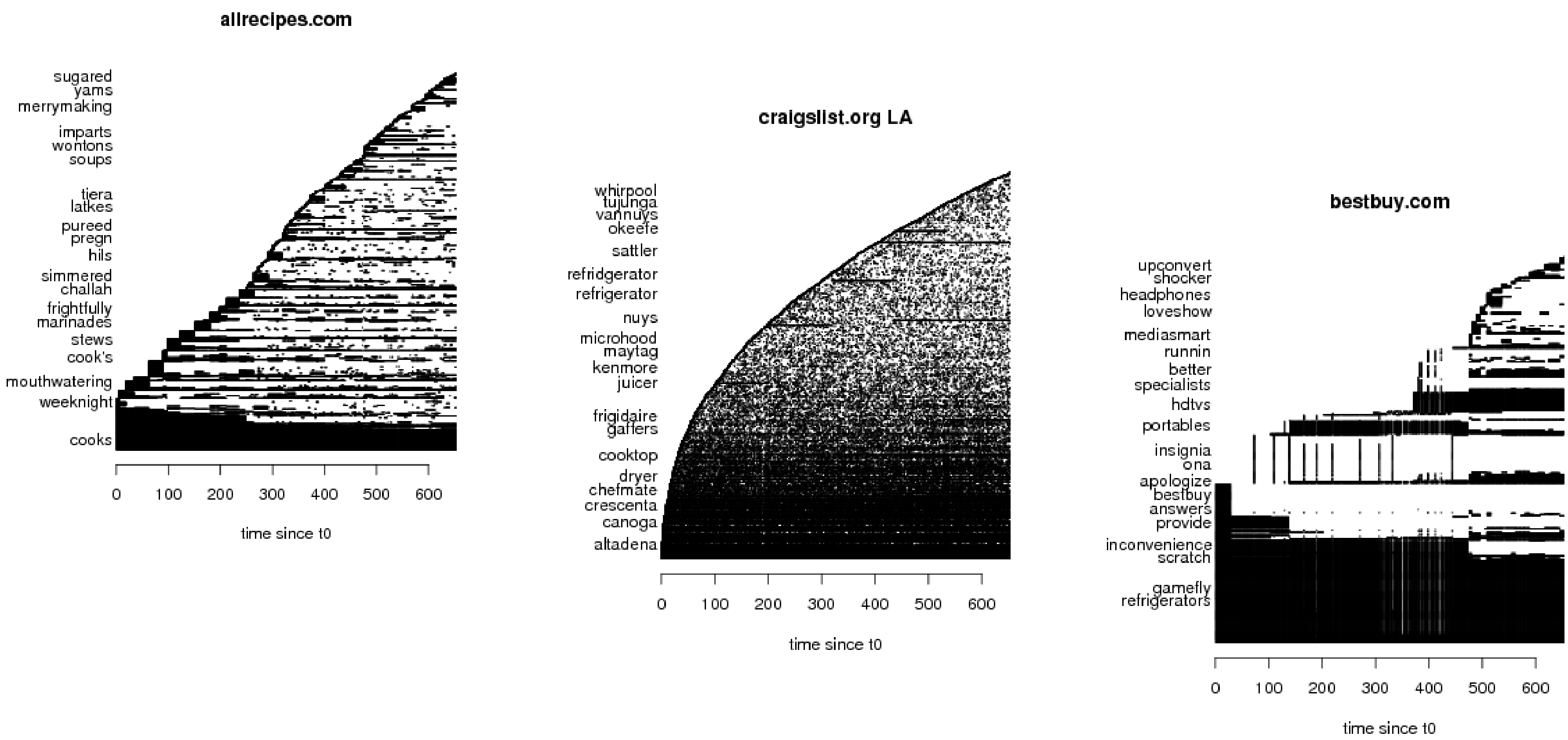




Adar et al: Within-Page Change

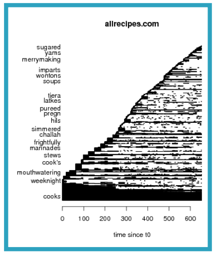
- Term-level changes
  - Divergence from norm
    - cookbooks
    - salads
    - cheese
    - ingredient
    - bbq
    - …
  - "Staying power" in page

Example Term Longevity Graphs



Change and Term Importance

- Traditional IR uses “tf/idf” term weighting
- Time-aware term weighting
  - Elsas & Dumais, WSDM 2010 – language model partitioned by term longevity (+ change prior on

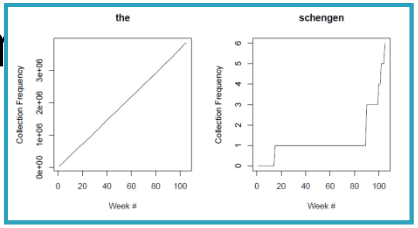


$$P(q|D) = \lambda_L P(q|D_L) + \lambda_M P(q|D_M) + \lambda_S P(q|D_S)$$

  - Aji et al., CIKM 2010 – importance of a term

$$TF_{RHA}(t,d) = \lambda_1 TF_{global}(t,d) + \lambda_2 TF_{burst}(t,d) + \lambda_3 TF(t,d)$$

  - Efron, JASIST 2010 – importance of a term determined by its deviation from linear tin series



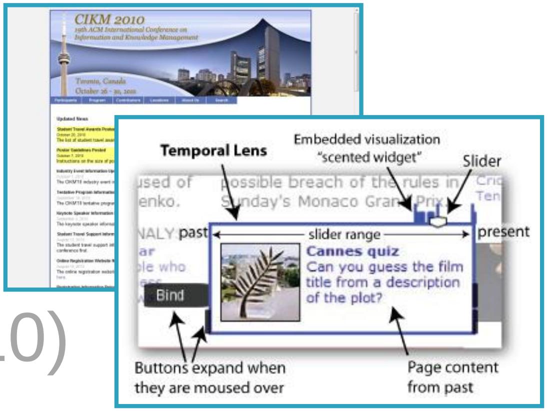
- Used to improve ranking

Systems and Applications

- Systems
  - Internet Archive (e.g., WayBack Machine)
  - Internet Memory Foundation
  - Wikipedia
  - Index structures to support time-travel search
    - Berberich et al. SIGIR 2007, Anand et al. SIGIR 2012.
- Applications
  - Crawling
  - Ranking
  - Query suggestion, burst detection, …
  - User experience

Dynamics and User Experience

- Content changes
  - Diff-IE (Teevan et al., 2008)
  - Zoetrope (Adar et al., 2008)
  - Diffamation (Chevalier et al., 2010)
  - Temporal summaries and snippets …
- Interaction changes
  - Explicit annotations, ratings, “likes”, etc.
  - Implicit interest via interaction patterns
  - Edit wear and read wear (Hill et al., 1992)



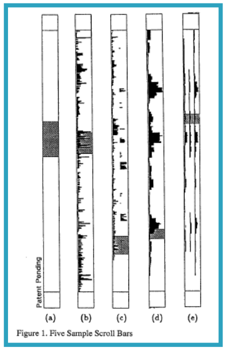
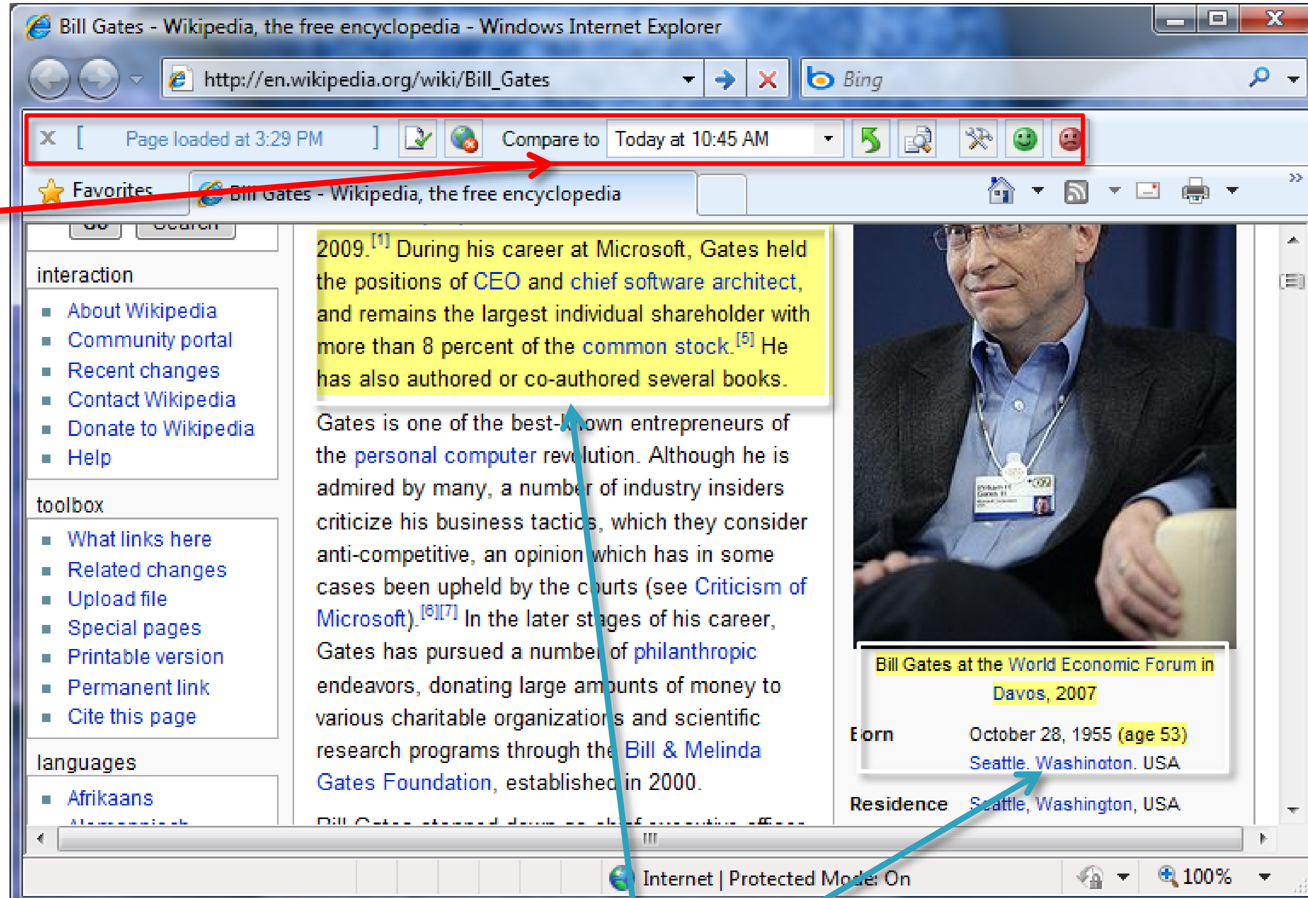


Figure 1. Five Sample Scroll Bars

Diff-IE

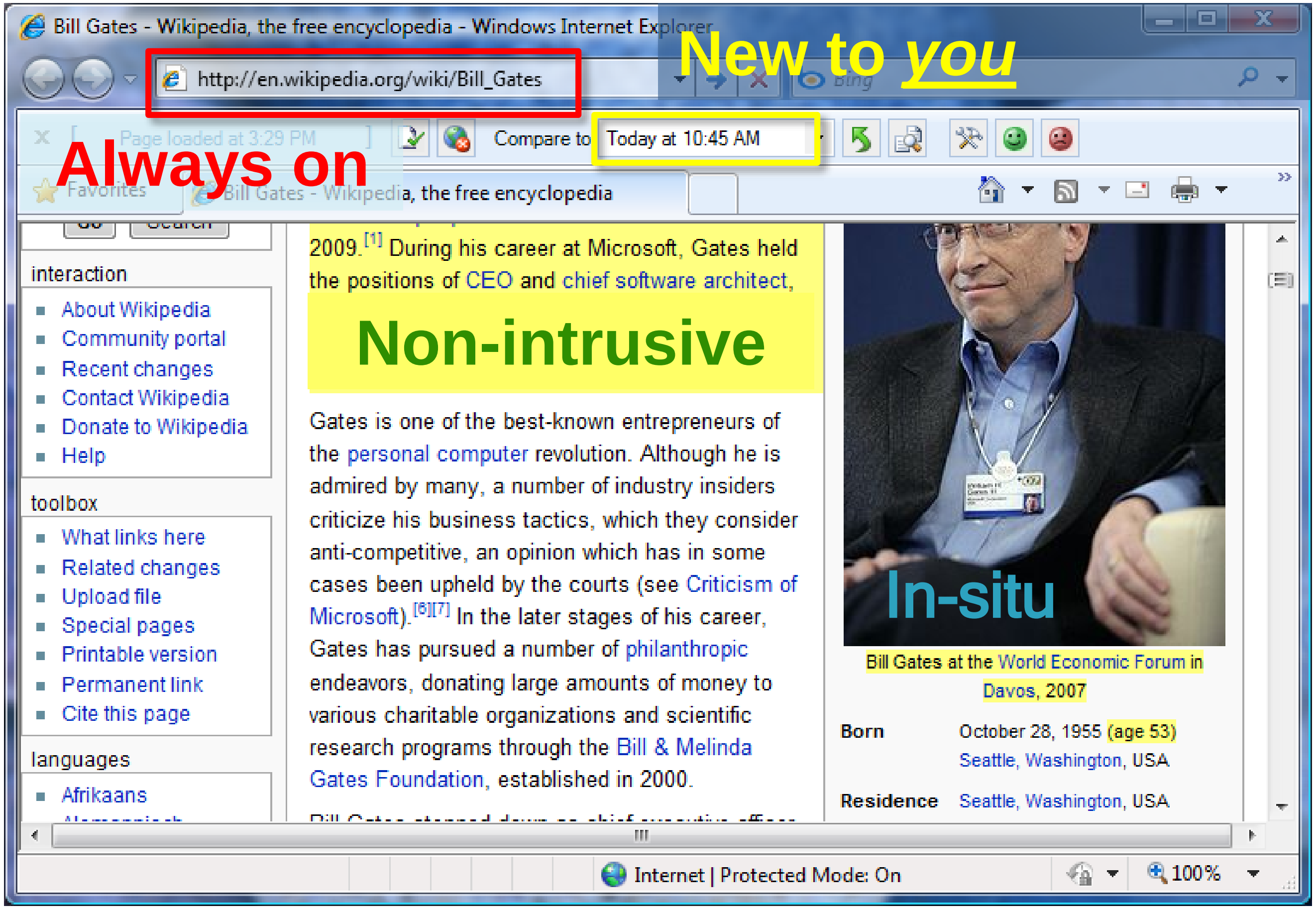
Diff-IE toolbar



Changes to page since your last visit

J. Teevan, S. T. Dumais, D.Liebling and R. Hughes. Changing how people view change on the web. UIST 2009

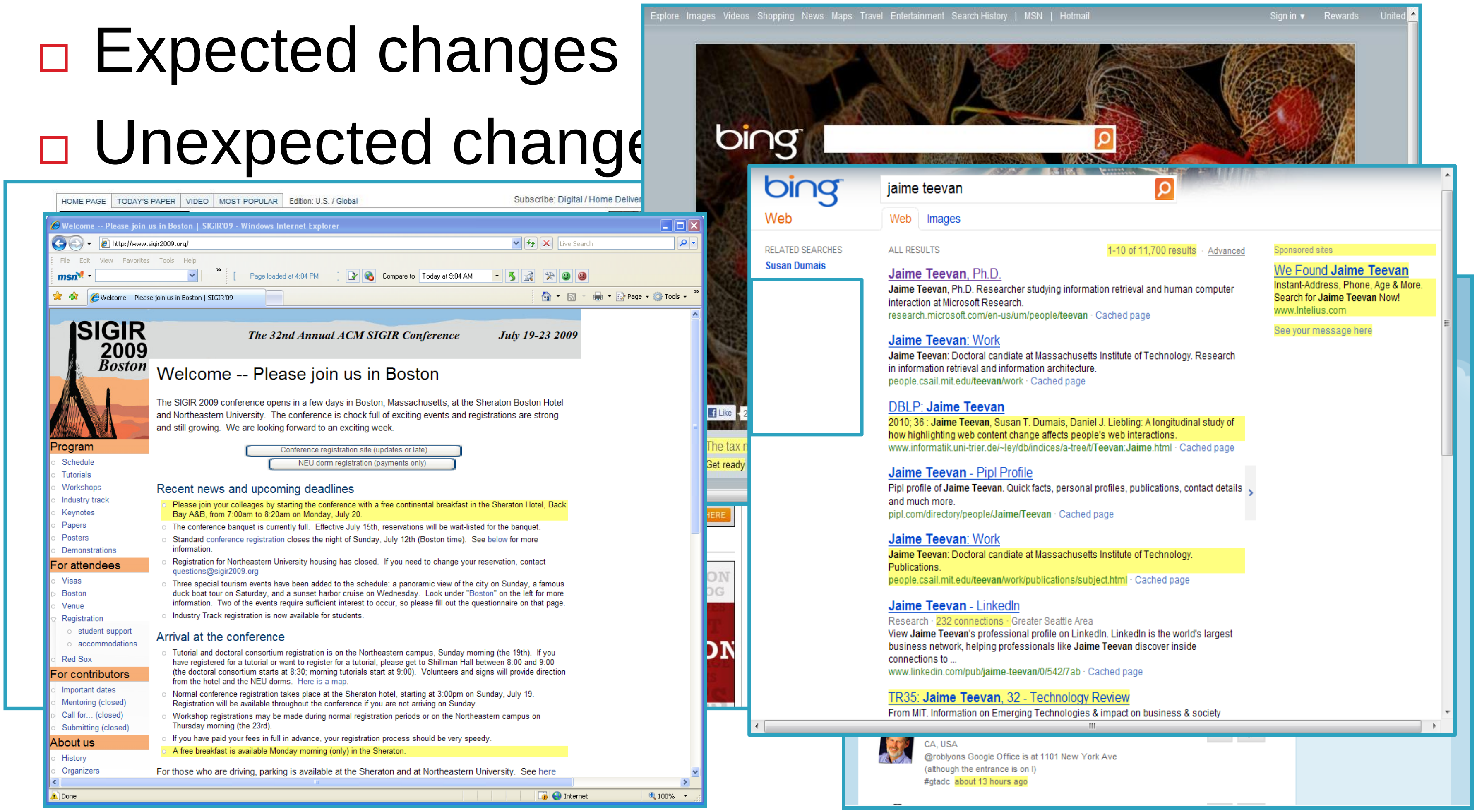
Interesting Features of Diff-IE



Download: http://research.microsoft.com/en-us/projects/diffie/default.aspx

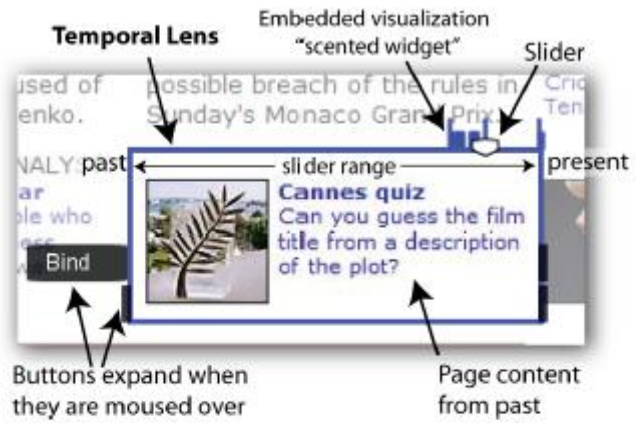
Diff-IE in Action

- Expected changes
- Unexpected changes



Zoetrope

- System that enables interaction with historical Web
- Select regions of interest (x-y location, dom structure, text)
  - E.g., stock price, traffic status, headlines about wsdm, …
- Operators for ma nterest
  - Filter
  - Link
  - Visualize



E. Adar, M. Dontcheva, J. Fogarty and D. Weld. Zoetrope: Interacting with the ephemeral web. UIST 2008

Web Content Dynamics

- Overview
- Change in “persistent” web documents
  - Characterizing content dynamics
  - Systems and applications
- Change in “real-time” content streams
  - Characterizing content dynamics
  - Systems and applications
- Change in Web graphs
  - Web graph evolution
  - Authority and content over time

Change in "Real-Time" Content Streams

- Real-time streams of new content
  - Twitter, Facebook, YouTube, Pinterest, etc.
  - News, Blogs, etc.

- And also …
  - Wikipedia
  - Commerce sites (e.g., EBay, Amazon, etc.)

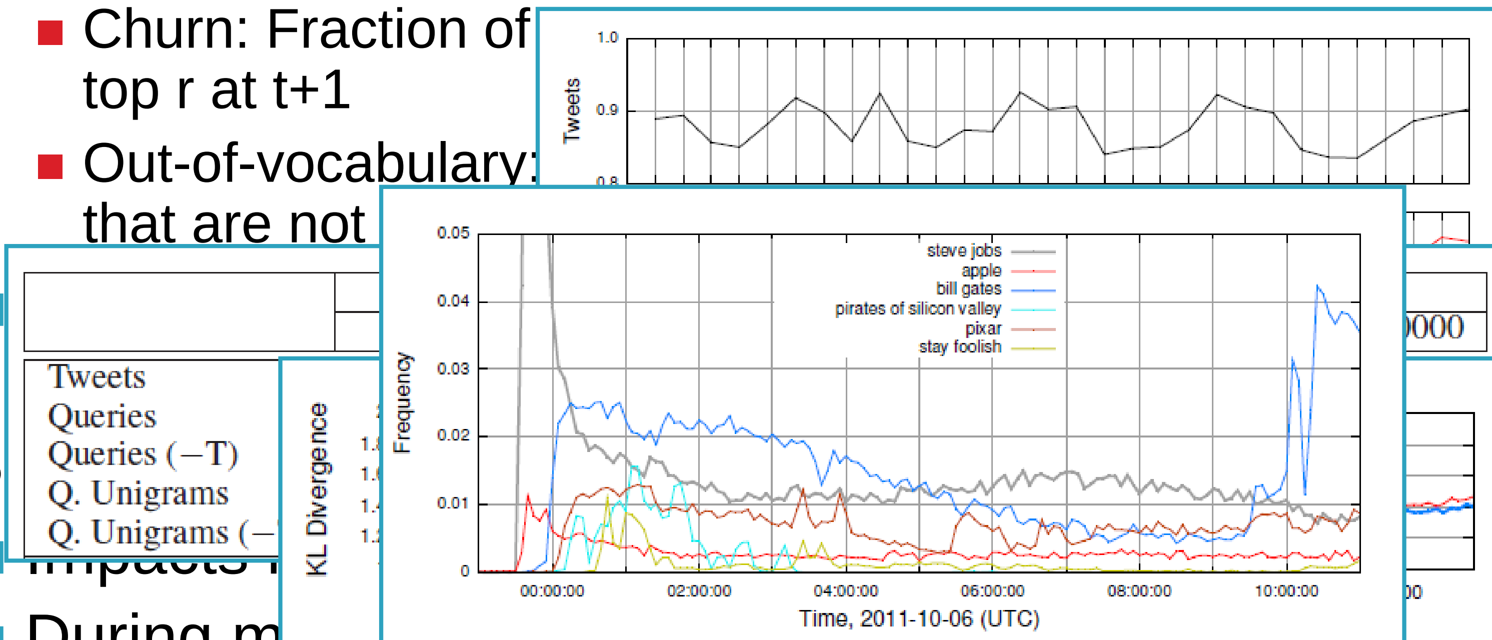
Change in Twitter

- Apr 2010, Twitter and US Library of Congress enter into agreement
- Jan 2013, Status report from [Library of Congress Archive]
  - 171 billion tweets (2006-2012)
  - Tweets/year
    - 21b (2006-2010); 150b (2011-2012)
  - Tweets/day <from Twitter>
    - 200m (6/2011); 400m (6/2012); 500m (10/2012)
  - Max Tweets/second <from Twitter>
    - 7k (Jan 1, 2011); 25k (Dec 11, 2012); 33k (Jan 1, 2013)

- *The Library has not yet provided researchers access to the archive. Currently, executing a single search of just the fixed 2006-2010 archive on the Library's systems could take 24 hours. This is an inadequate situation in which to begin offering access to researchers, as it so severely limits the number of possible searches.*

Temporal Analysis of Twitter

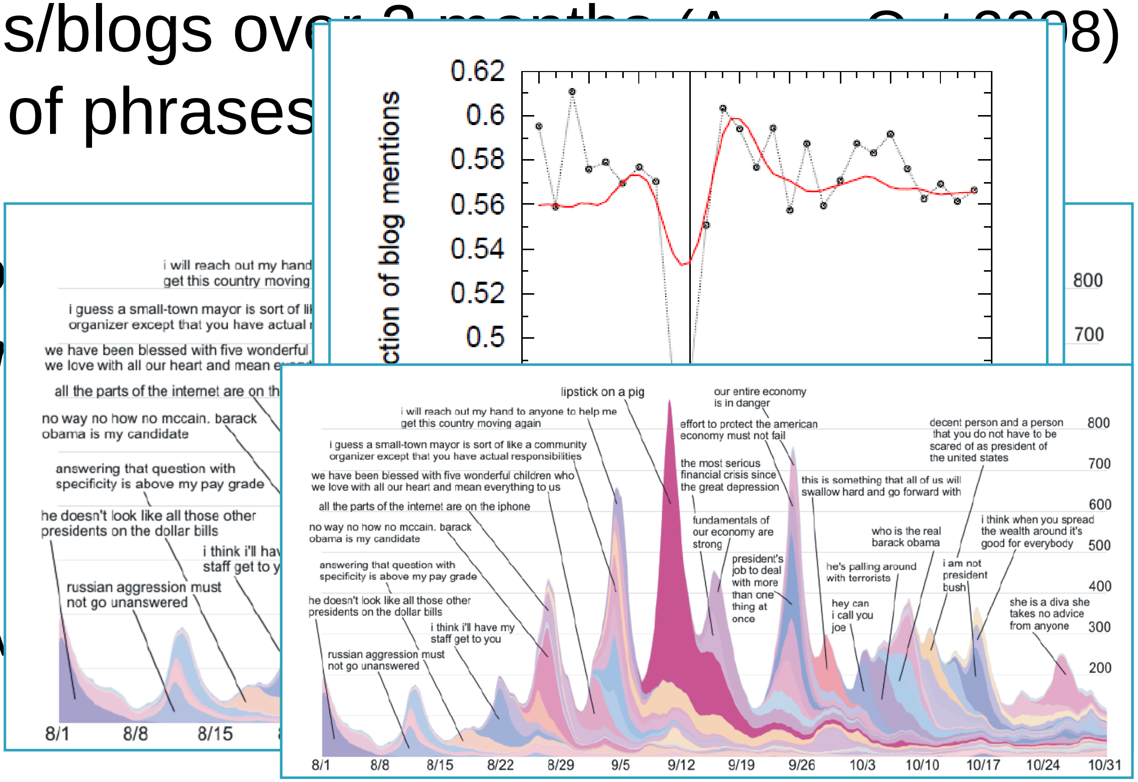
- How different are tweets (and queries) day-over-day?
  - Term (and top-term) distributions
    - KL Divergence (t+1|t)
    - Churn: Fraction of top r at t+1
    - Out-of-vocabulary: that are not
  - nt
- S
  - impacts
  - During major events, sub-hour updates important



J. Lin and G. Mishne. A study of “churn” in tweets and real-time search queries. ICWSM 2012.

Temporal Analysis of “Memes”

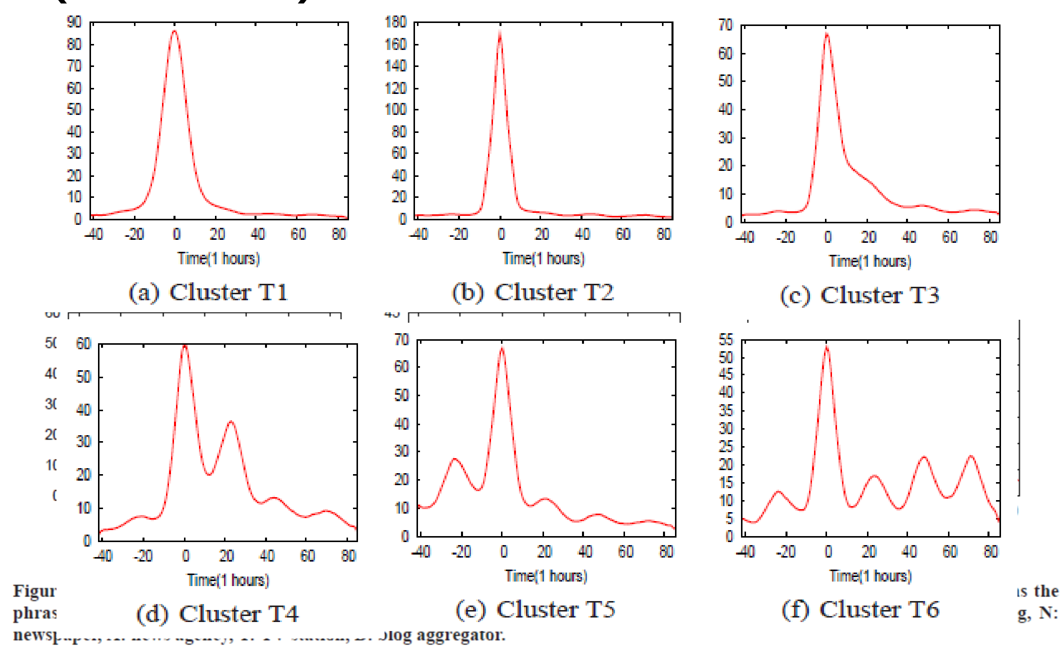
- Tracking short distinctive phrases (“memes”) in news media and blogs
- 90 million articles/blogs ove… …8)
- Cluster variants of phrases
- Global patterns
  - Probabilistic mo…
  - Choose(j) $\propto f($…
- Local patterns
  - Peak of attentio… hours
  - Divergent behav… news and blogs



J. Leskovec, L. Backstrom and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. KDD 2009

Temporal Analysis of Blogs & Twitter

- Patterns of temporal variation
- Short texts over time
  - Short text phrases (memes) <from 170m news articles>
  - Hashtags <from
- Spectral clusterir
  - 6 clusters News/
  - 6 clusters Twitte
- Predict type give

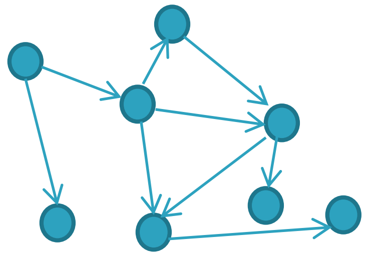


J. Yang and J. Leskovec. Patterns of temporal variation in online media. WSDM 2011.

Web Content Dynamics

- Overview
- Change in “persistent” web documents
  - Characterizing content dynamics
  - Systems and applications
- Change in “real-time” content streams
  - Characterizing content dynamics
  - Systems and applications
- Change in Web graphs
  - Web graph evolution
  - Authority and content over time

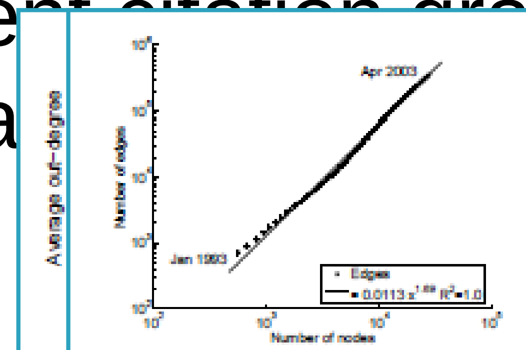
Static Graphs/Networks



- Example graphs: web, tweets, emails, citation networks, etc.
- Properties
  - #nodes, #edges, reciprocity, clustering coefficient, heavy tails for in- and out-degree distributions, size of largest connected component, …
  - Small-world phenomenon
- Models for graph generation
  - Preferential attachment
  - Copying

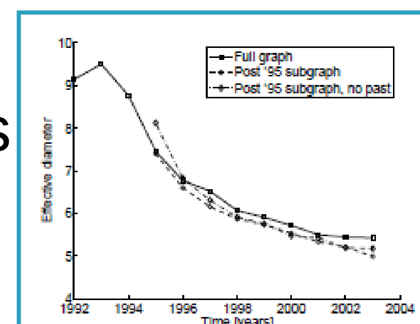
Evolution of Graphs over Time

- ArXiv citation graph, Patent citation graph, Autonomous systems gra graph



(a) arXiv

- Empirical observations
  - Densification
    - Densification: Average out-degree increases over time $e(t) \propto n(t)^a$
    - Densification power law: Nodes r time fit by power law
  - Shrinking effective diameter



(a) arXiv citation graph

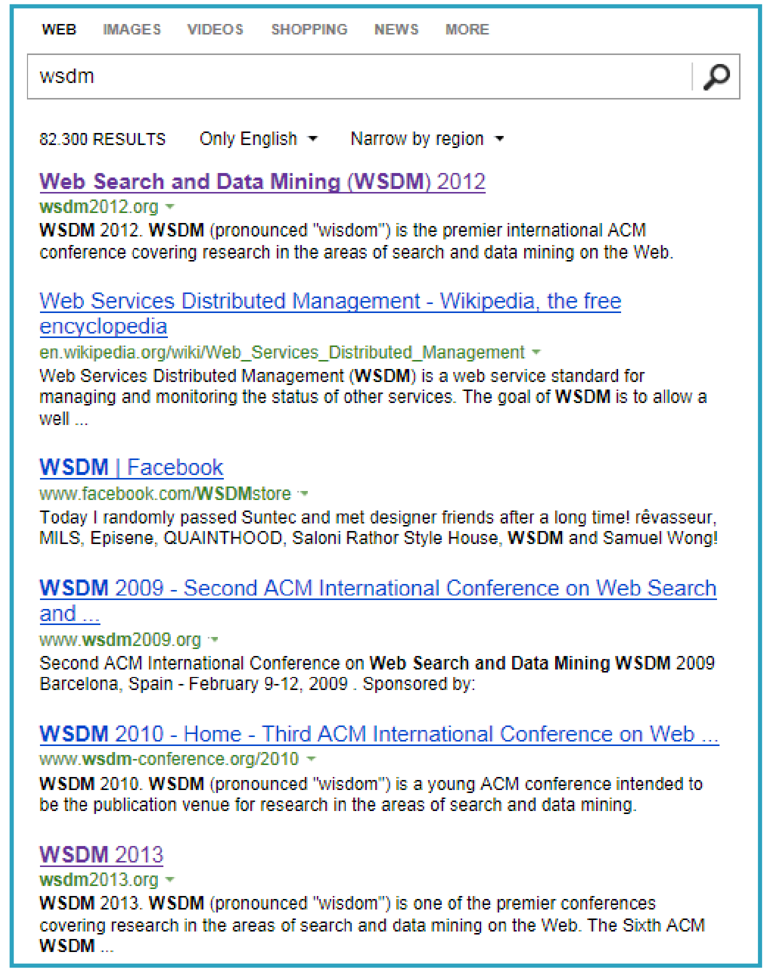
- Generative model

J. Leskovec, J. Kleinberg and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. KDD 2005.

Web Page Authority over Time

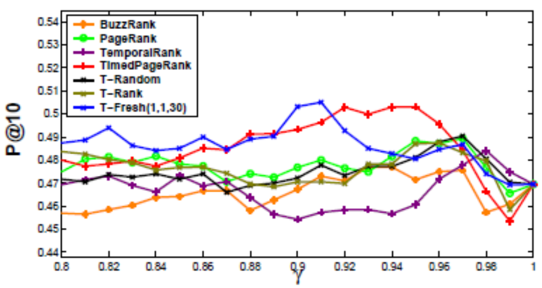
- Query: *wsdm*

- Why is older content ranked higher?
  - Behavioral signals (in-links, clicks) more prevalent for older pages



Web Page Authority over Time

- Modeling page authority over time
  - Multiple web snapshots (.ie domain from IA, 2000-2007)
  - Temporal page profiles (TPP) and temporal in-link profiles (TLP)
  - *Page freshness score,* using exponential decay over time
  - Use freshness score to control authority propagation in a *temporal random surfer model*
    - Web surfer has temporal intent
      (which controls choice of target snapshot)
    - Web surfer prefers fresh content



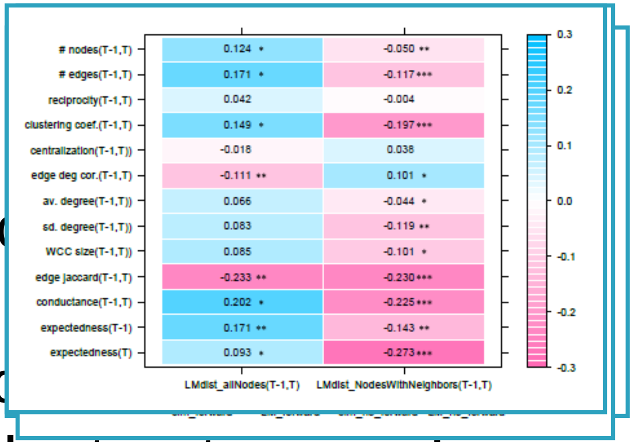
(a) Relevance performance: P@10.

- Rank using combination of content and
  $Score(p) = \gamma BM25 + (1-\gamma)TemporalA$

N. Dai and B. Davison. Freshness matters in flowers, food and web authority. SIGIR 2010

CoEvolution of Structure and Content

- Three networks over time
  - Twitter, Second Life, Enron email
- Characteristics of <u>network structure</u>
  - Standard metrics, Conductance, Expe
- Measures of <u>network content</u>
  - Similarity, Divergence of language mod



- Empirical correspondence of network structure and content diversity and novelty
  - Conductance correl w/ high diversity of content
  - Expectedness correl w/ content novelty
- Simulation model
  - Node policy to forward based on recency, novelty and topicality

C-T Teng et al. Coevolution of network structure and content. ArXiv

Web Content Dynamics

- Overview
- Change in “persistent” web documents
  - Characterizing content dynamics
  - Systems and applications
- Change in “real-time” content streams
  - Characterizing content dynamics
  - Systems and applications
- Change in Web graphs
  - Web graph evolution
  - Authority and content over time

Resources

- Web crawls
  - CLUEWeb’09, CLUEWeb’12 (static snapshots)
  - Common Crawl
  - PageTurner
  - Internet Archive
  - Publication/citation
- Content streams
  - Twitter API, Library of Congress
  - Wikipedia (+ aggregate usage data)
  - Blogs (TREC), Blogs (Spinn3r)
  - Yahoo! update firehose (shutting down Apr 13, 2013)

References

- J. Cho and H. Garcia-Molina. The evolution of the web and implications for an incremental crawler. VLDB 2000.
- A. Ntoulas, J. Cho and C. Olston. What's new on the web? The evolution of the web from a search engine perspective. WWW 2004.
- D. Fetterly, M. Manasse, M. Najork and J. Weiner. A large-scale study of the evolution of web pages. WWW 2003.
- E. Adar, J. Teevan, S. T. Dumais and J. Elsas. The web changes everything: Understanding the dynamics of web content. WSDM 2009.
- J. Teevan, S. T. Dumais, D.Liebling and R. Hughes. Changing how people view change on the web. UIST 2009.
- E. Adar, M. Dontcheva, J. Fogarty and D. Weld. Zoetrope: Interacting with the ephemeral web. UIST 2008.
- J. Lin and G. Mishne. A study of "churn" in tweets and real-time search queries. ICWSM 2012.
- J. Leskovec, L. Backstrom and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. KDD 2009
- J. Yang and J. Leskovec. Patterns of temporal variation in online media. WSDM 2011.
- J. Leskovec, J. Kleinberg and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. KDD 2005.
- N. Dai and B. Davison. Freshness matters in flowers, food and web authority. SIGIR 2010.
- C-T Teng et al. Coevolution of network structure and content. ArXiv.

Temporal Dynamics of Queries & User Behavior

WSDM 2013 Tutorial

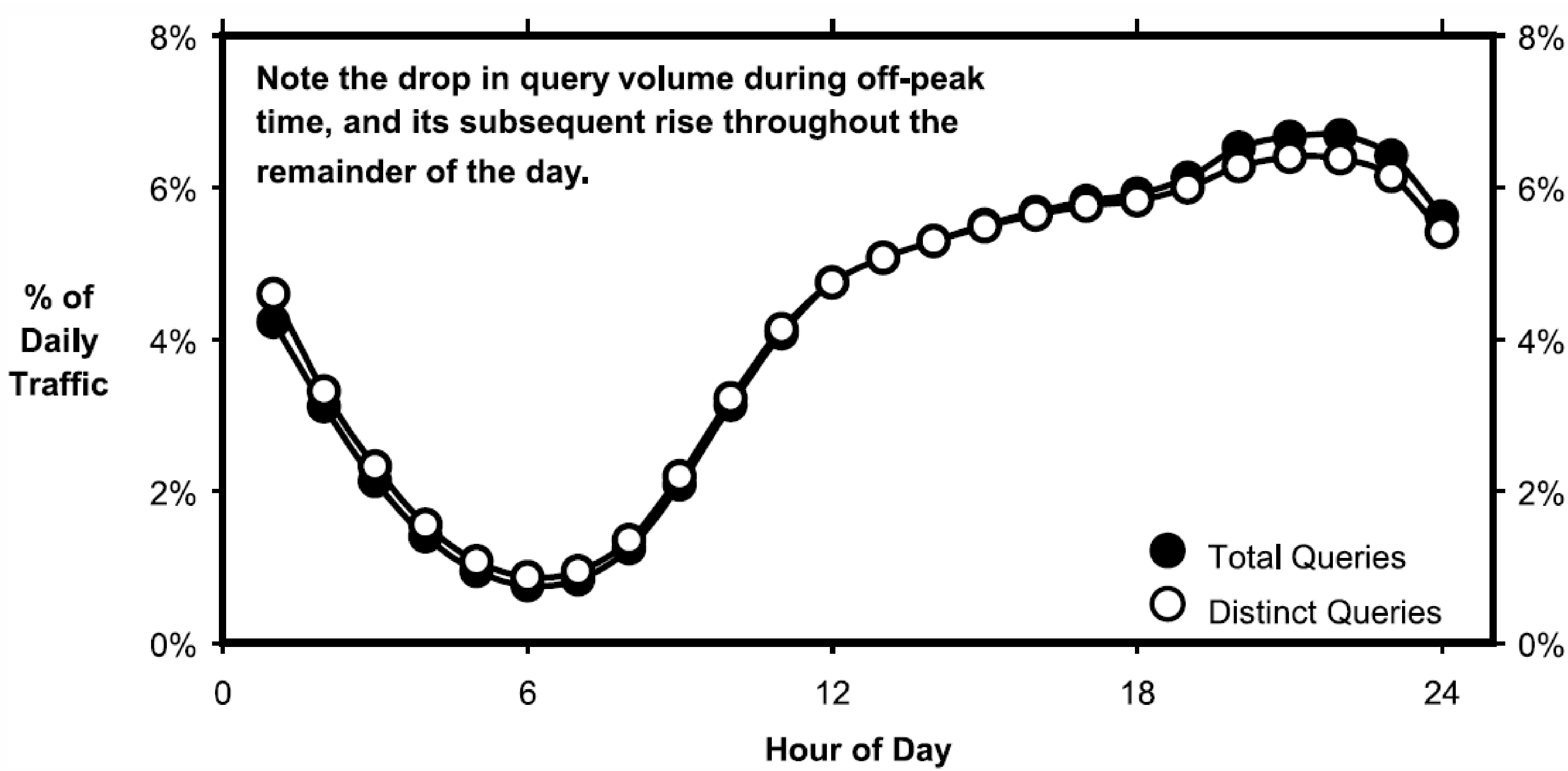
Outline

- Query Dynamics
  - Hourly, Daily & Monthly Trends
- Categorizing Time-Sensitive Queries
  - Spike, Periodicity
- Modeling Query Dynamics
  - Burst detection, Time-Series
- Temporal Patterns in User Behavior
  - Re-finding, Long-term vs. Short-term
- Temporal Patterns & Search Evaluation
  - Predicting Search Satisfaction (SAT)

Query Dynamics

Temporal Analysis of Query Logs

- Hourly analysis of queries [Beitzel et. al, SIGIR2004, JASIST 2007]



Temporal Analysis of Query Logs

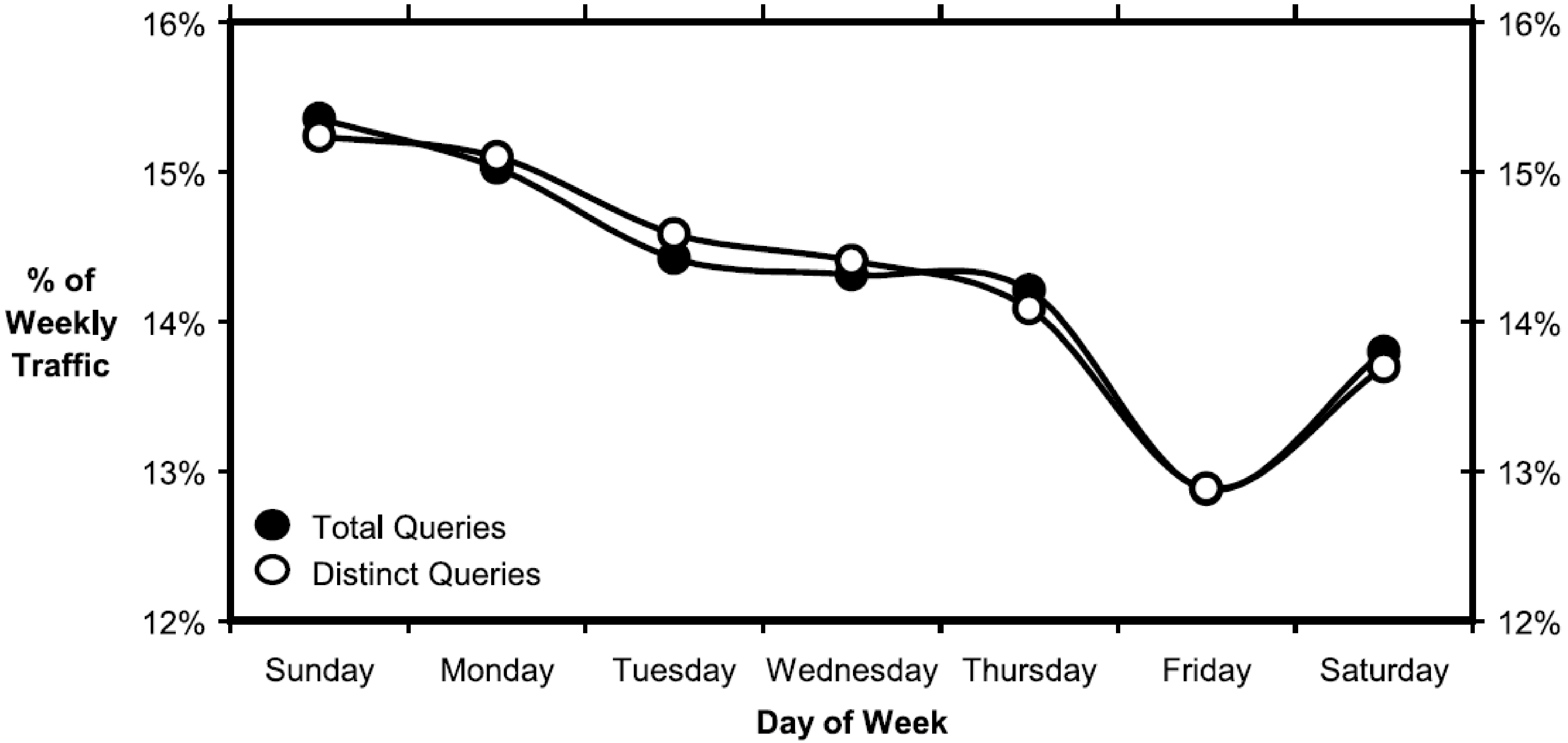


FIG. 3. Average volume of days in the week.

Temporal Analysis of Query Logs

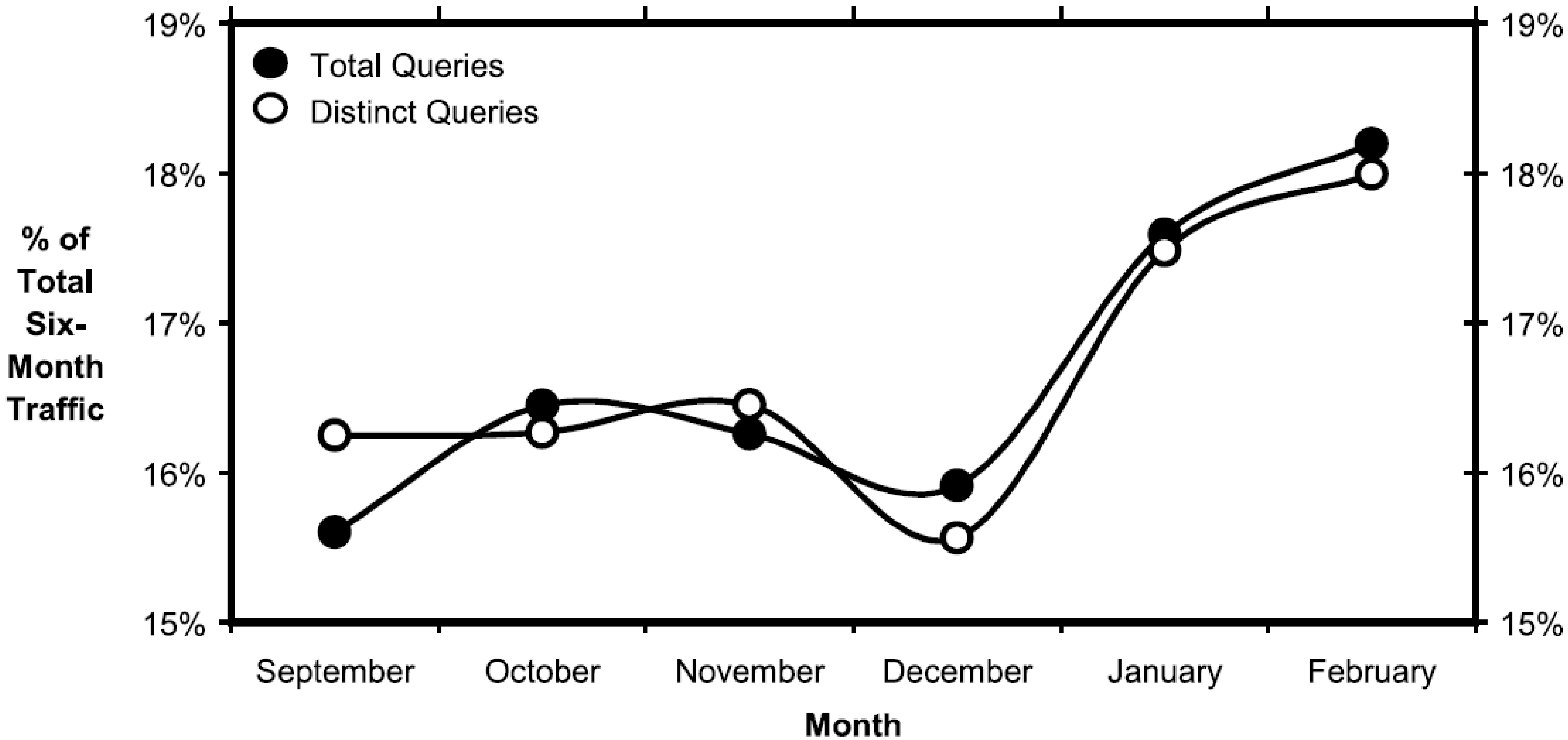


FIG. 5. Query volume by month.

Temporal Analysis of Query Logs

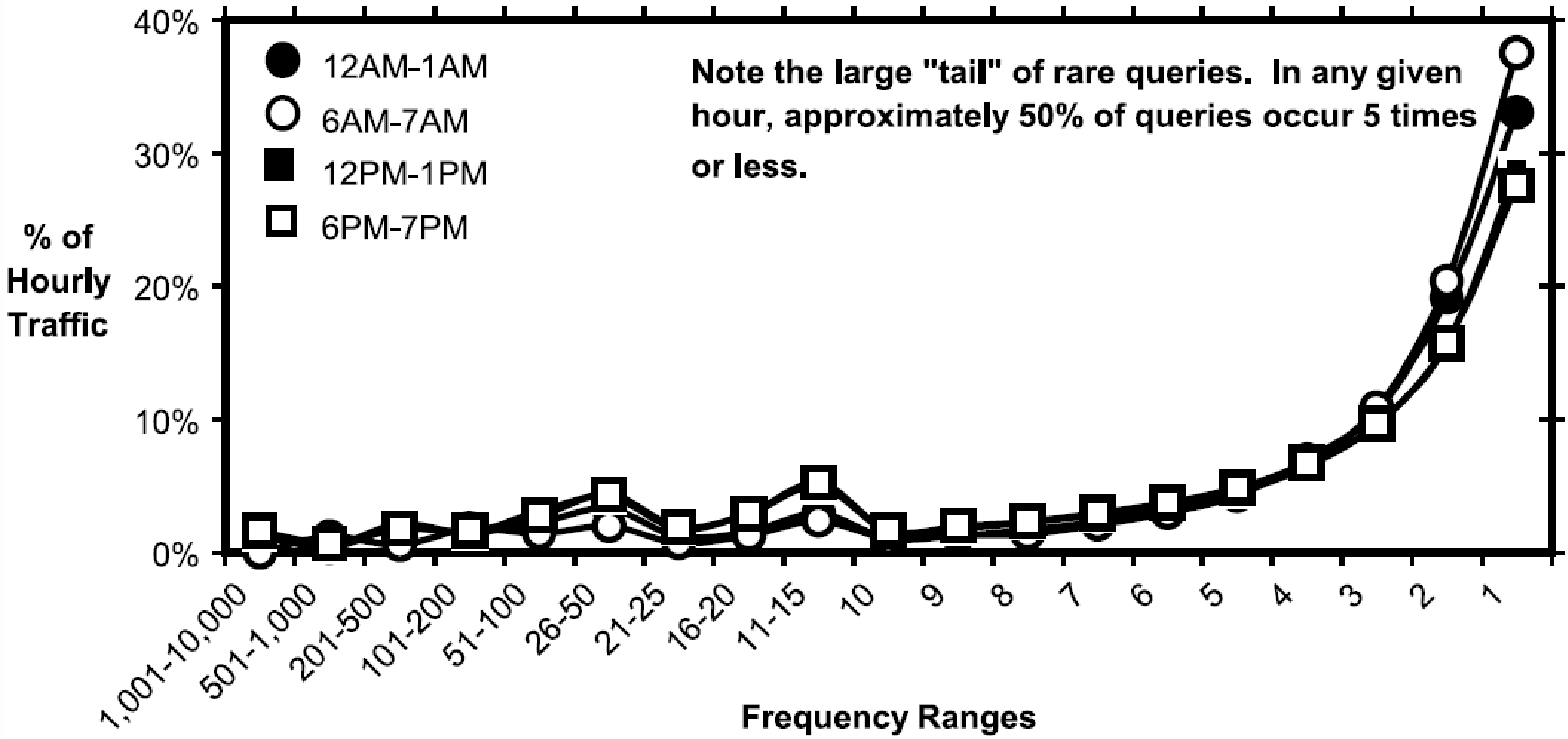


FIG. 2. Frequency distribution for selected hours.

Temporal Analysis of Query Logs

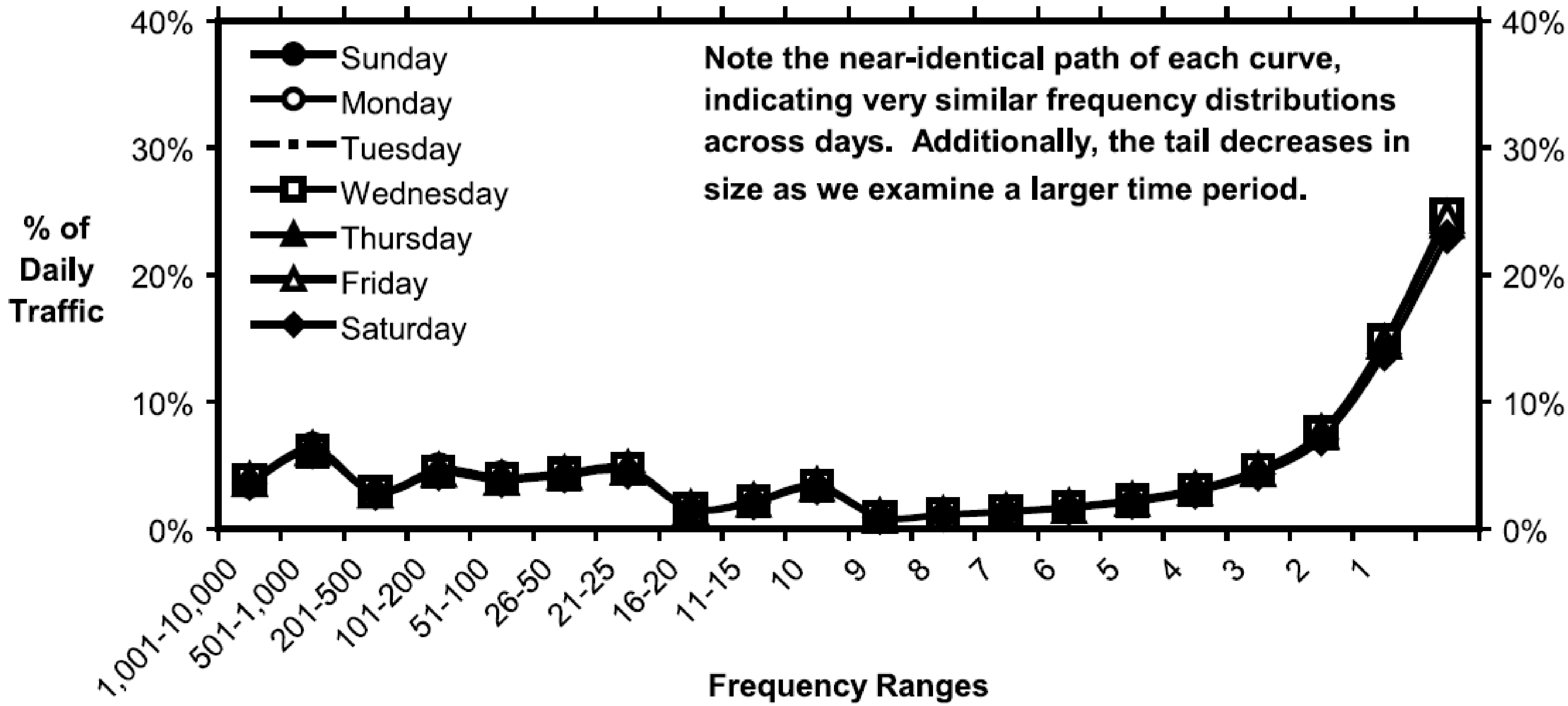


FIG. 4. Average frequency distributions for days in the week.

Temporal Analysis of Query Logs

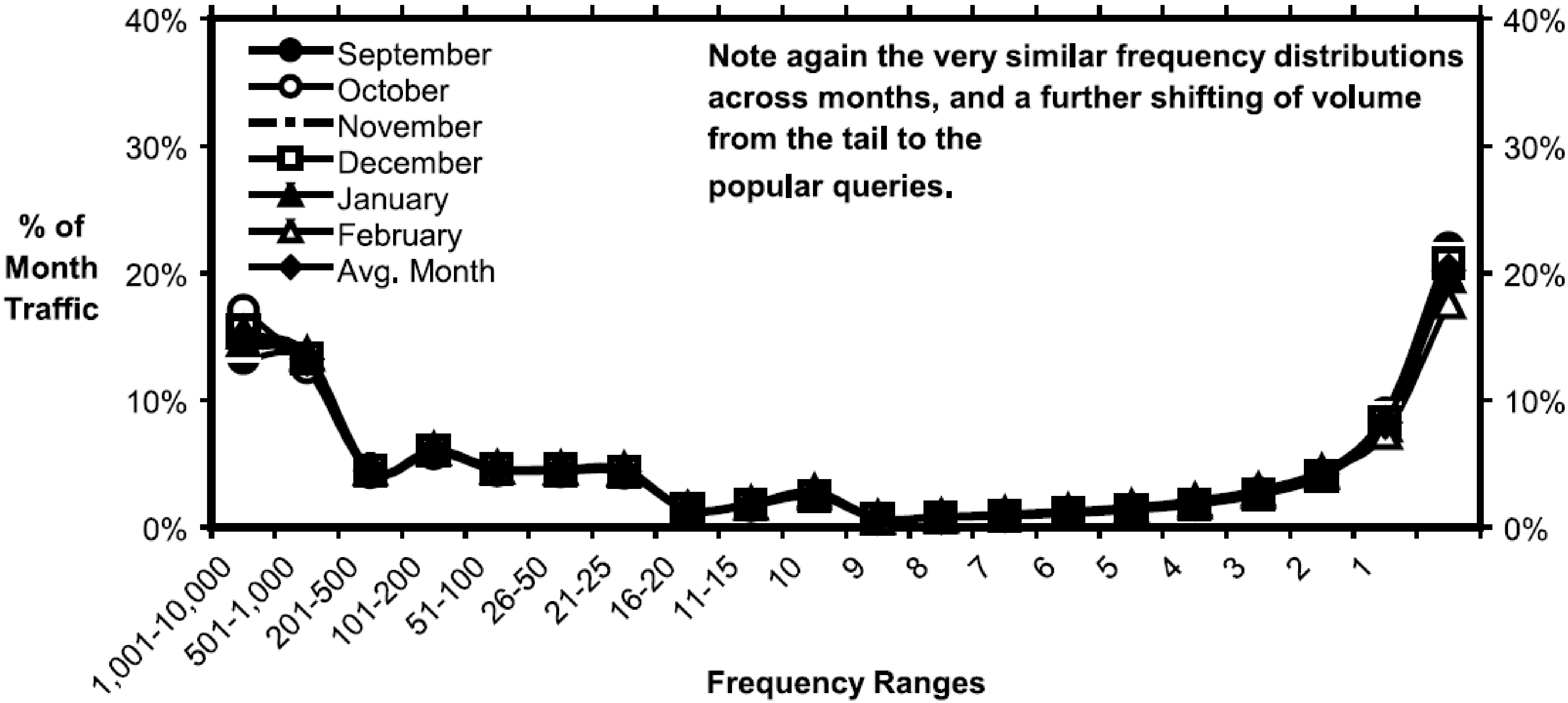


FIG. 6. Frequency distributions by month.

Temporal Analysis of Query Logs

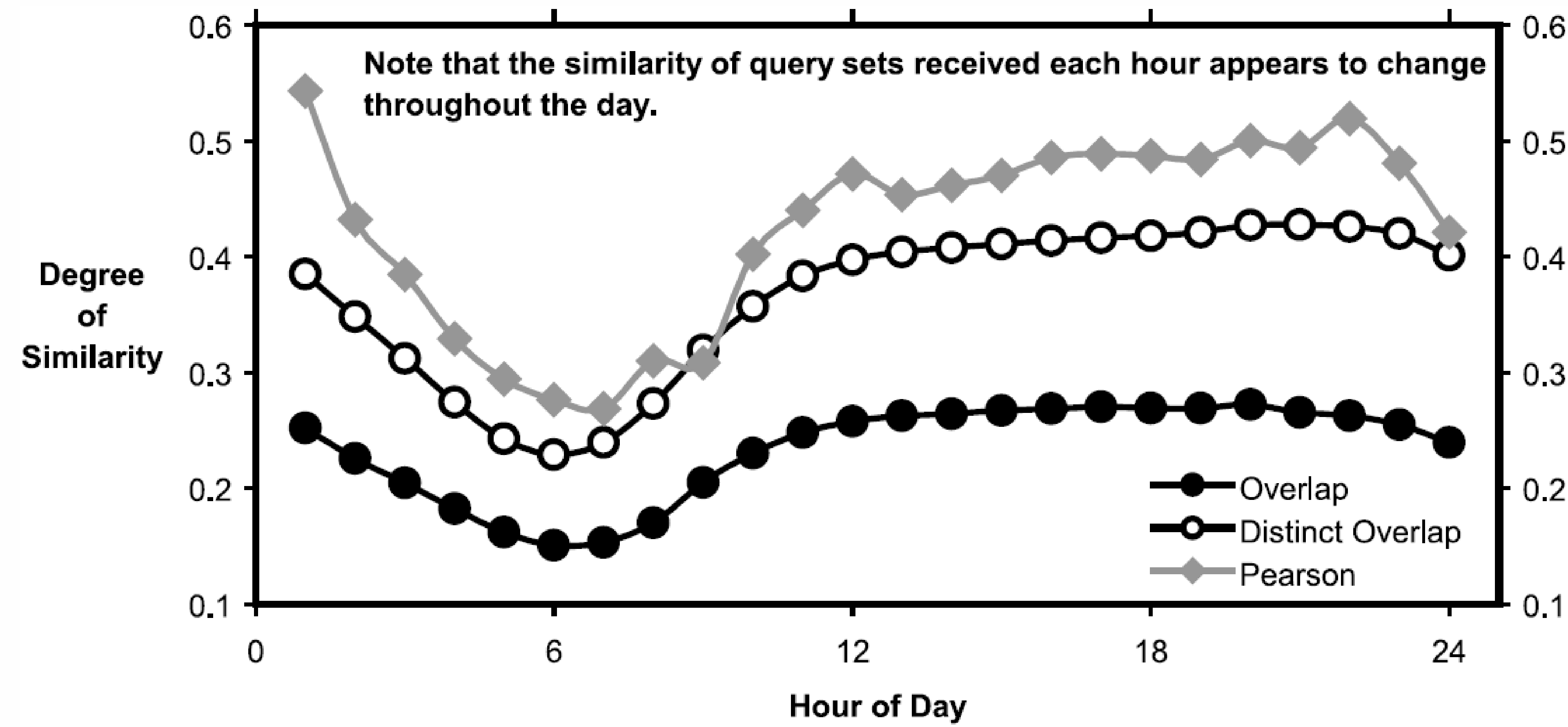
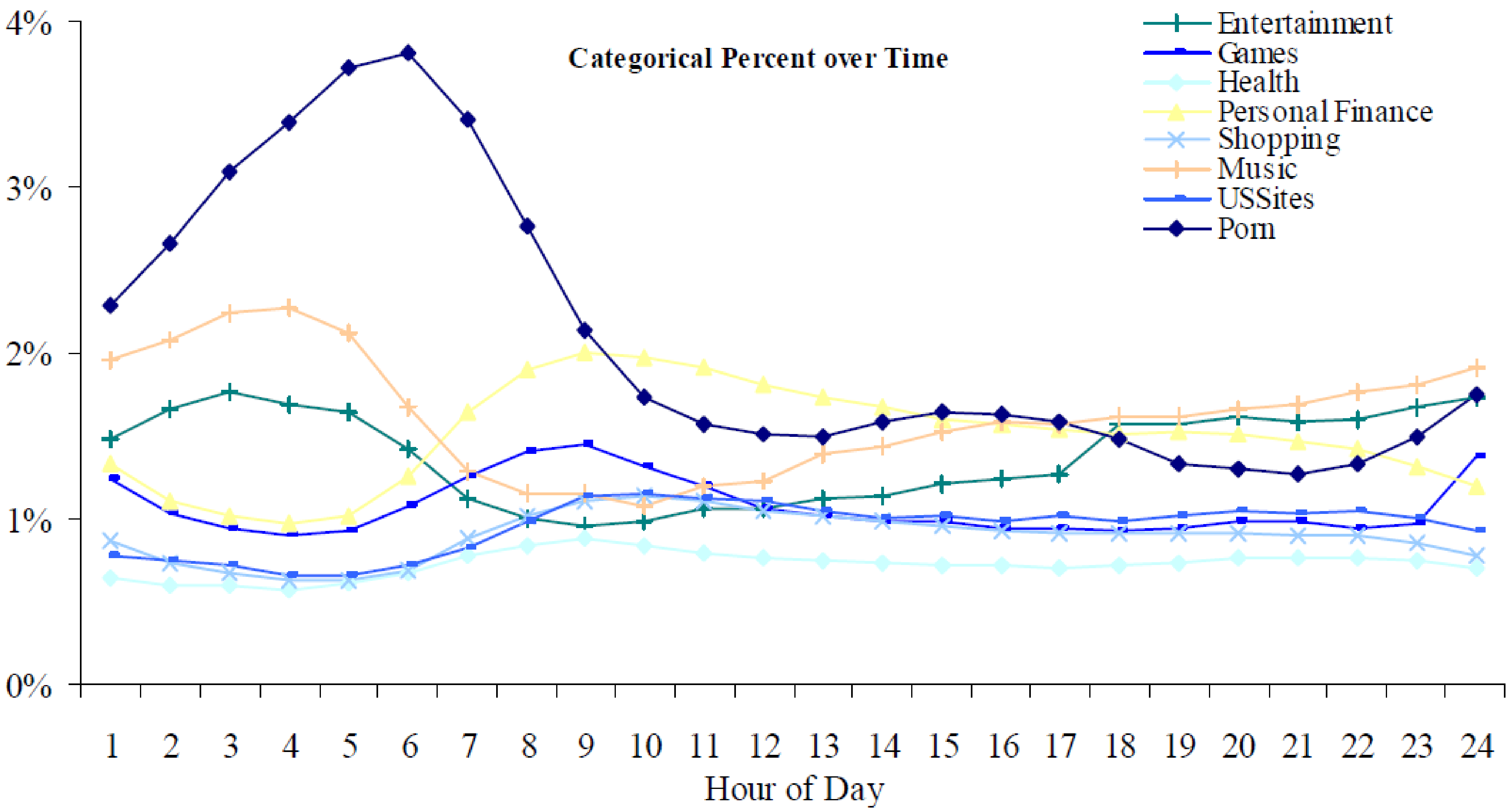
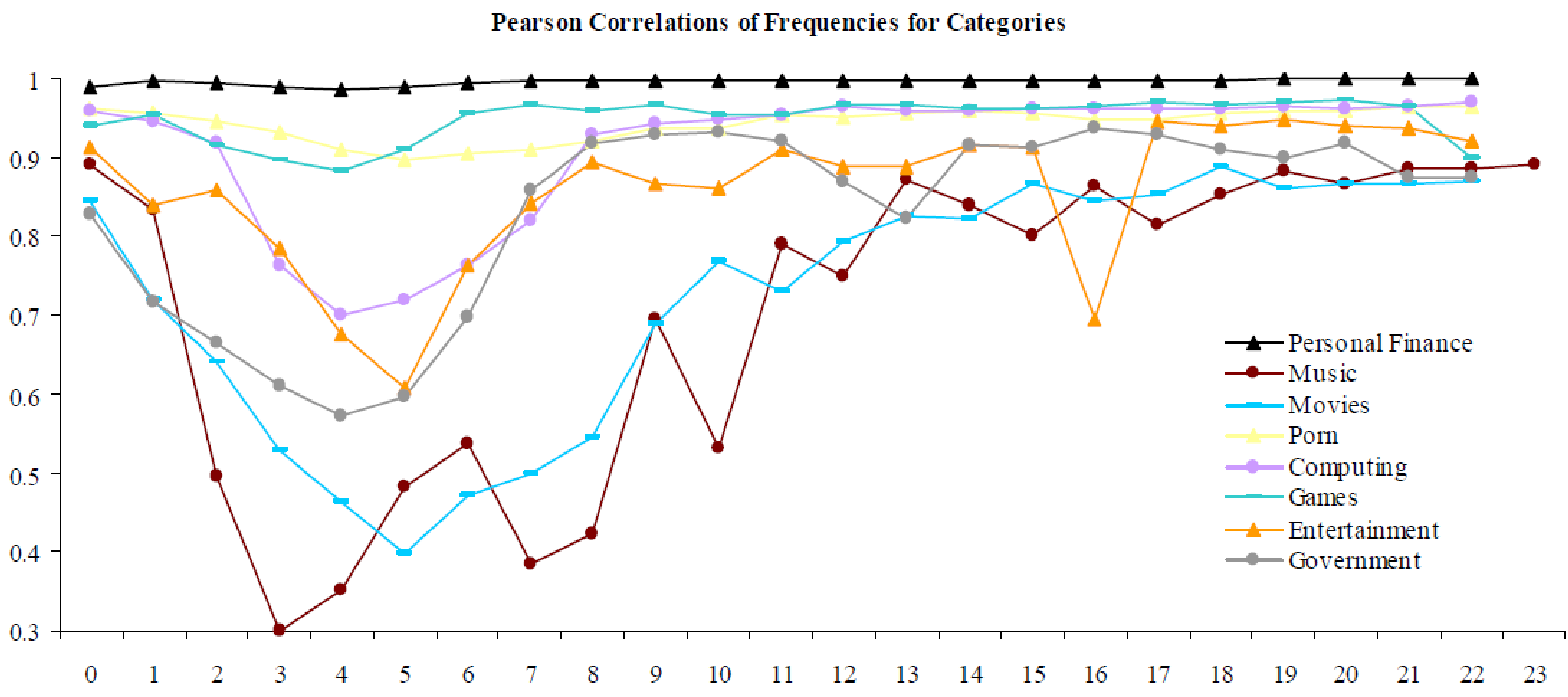


FIG. 7. Average overlap & Pearson correlations of matches from January 2,

Temporal Analysis of Query Logs



Temporal Analysis of Query Logs



Temporal Analysis of Query Logs

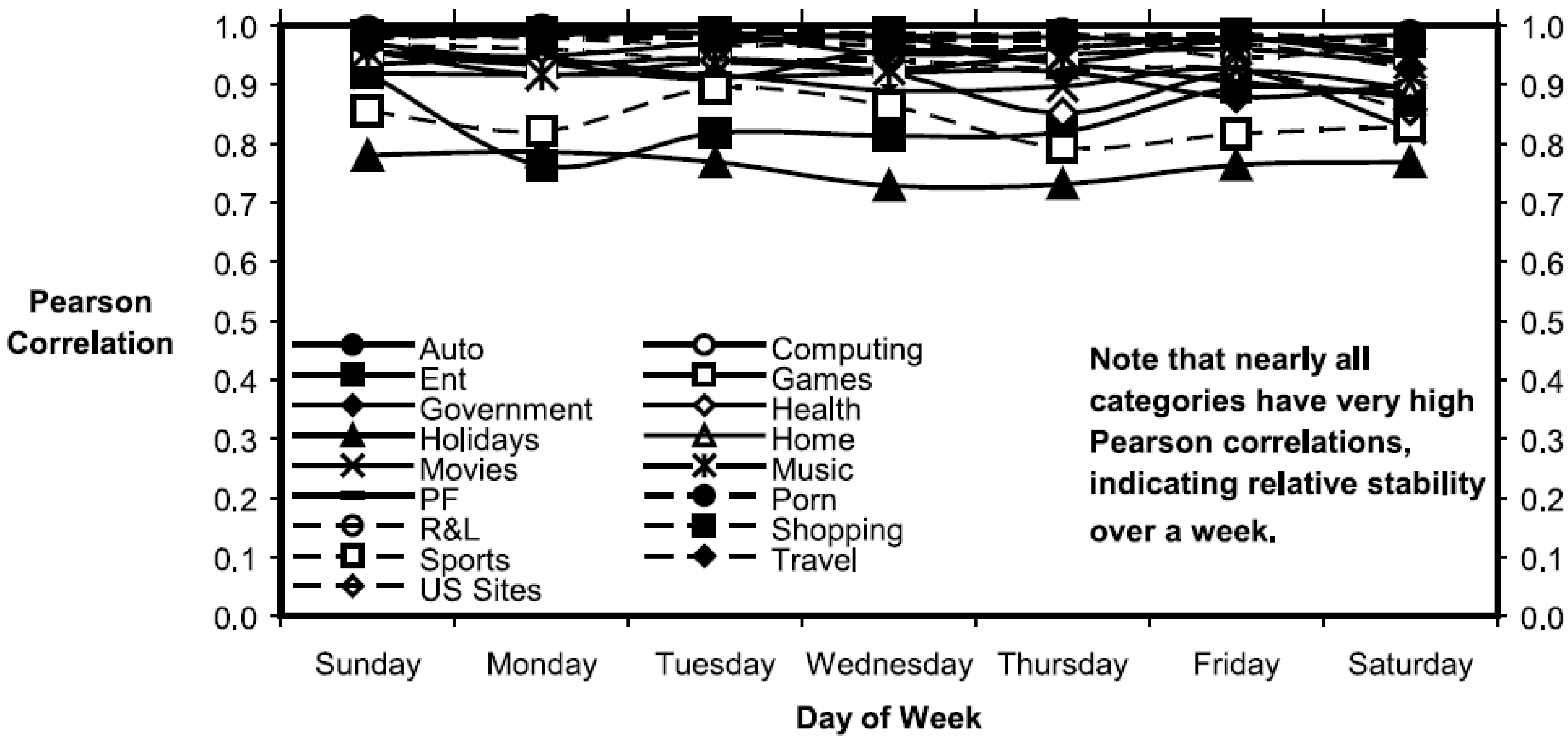


FIG. 22. Pearson correlations of matching query frequencies for each category averaged over days in a week.

Temporal Analysis of Query Logs

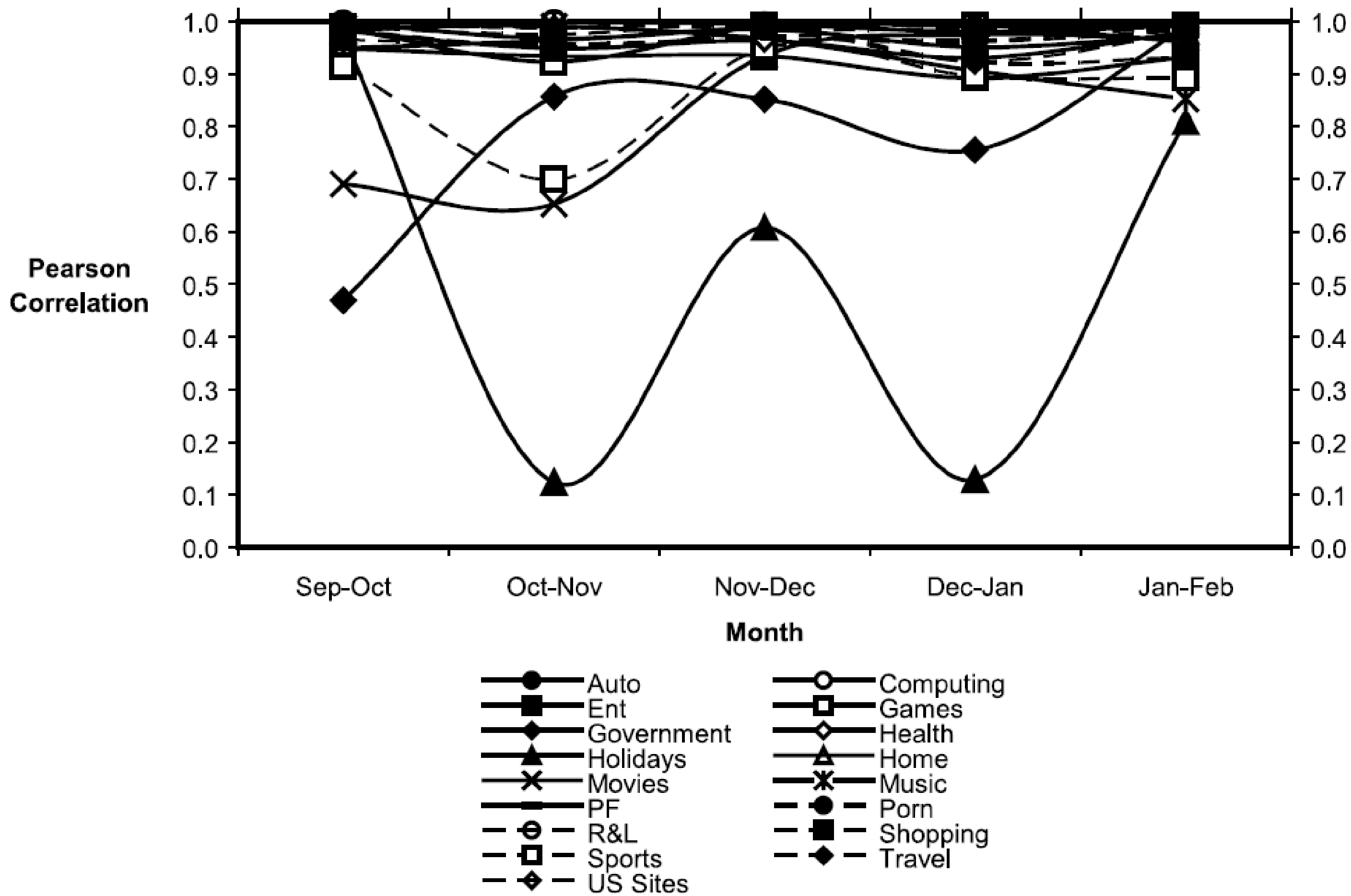


FIG. 23. Pearson correlations of matching query frequencies for each category over 6 months.

Outline

- Query Dynamics
  - Hourly, Daily & Monthly Trends
- Categorizing Time-Sensitive Queries
  - Spike, Periodicity
- Modeling Query Dynamics
  - Burst detection, Time-Series
- Temporal Patterns in User Behavior
  - Re-finding, Long-term vs. Short-term
- Temporal Patterns & Search Evaluation
  - Predicting Search Satisfaction (SAT)

Categorizing Query Dynamics

Burst, Periodicity

Categorizing Temporal Queries

- Temporal query classes [Kulkarni et al., WSDM2011]

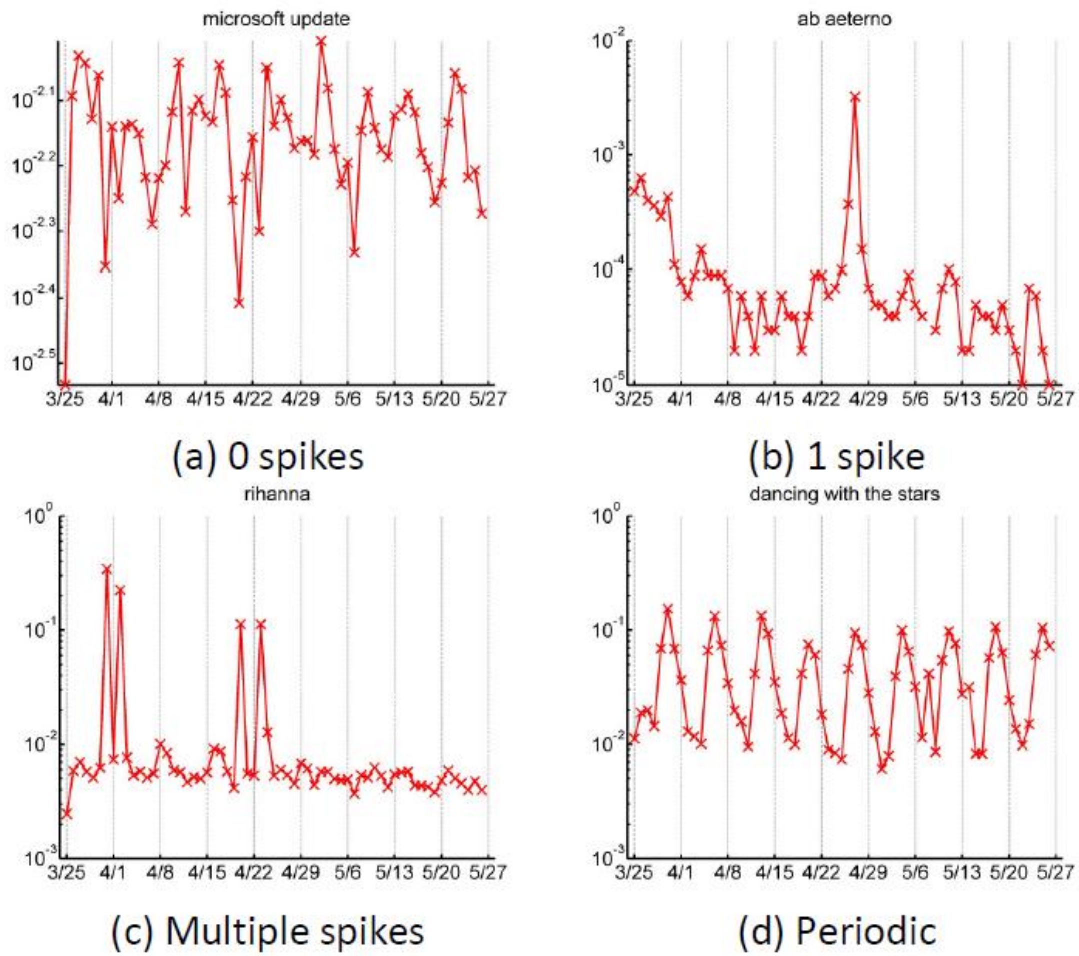


Figure 1. Different queries had different numbers of spikes in query popularity during the study period.

Categorizing Temporal Queries

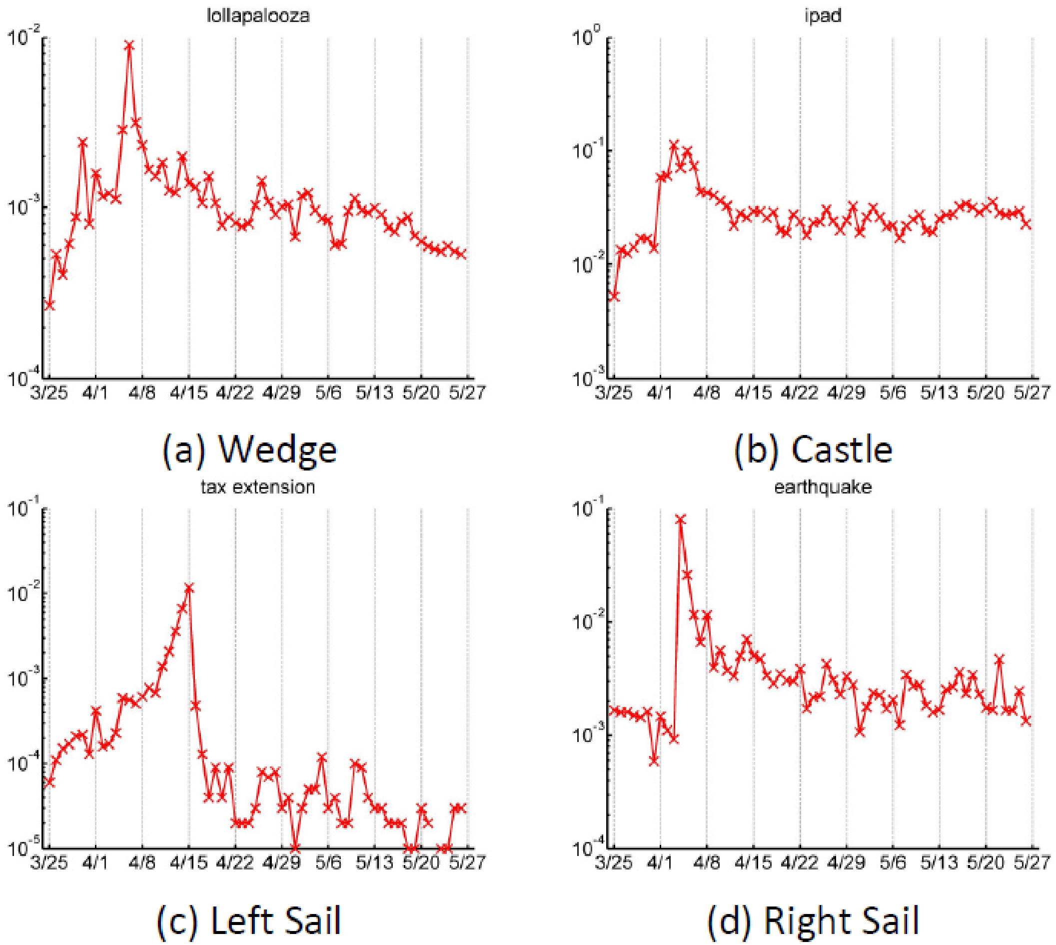
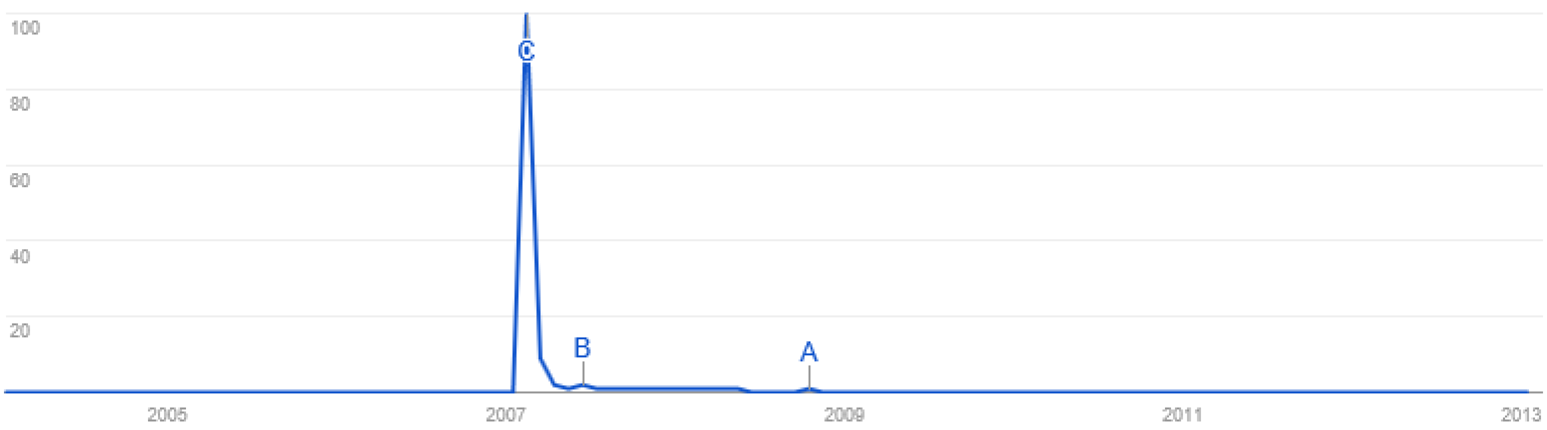


Figure 2. When a query spiked in popularity, the spike could occur in a variety of different shapes.

Categorizing Temporal Queries

- Bald Britney

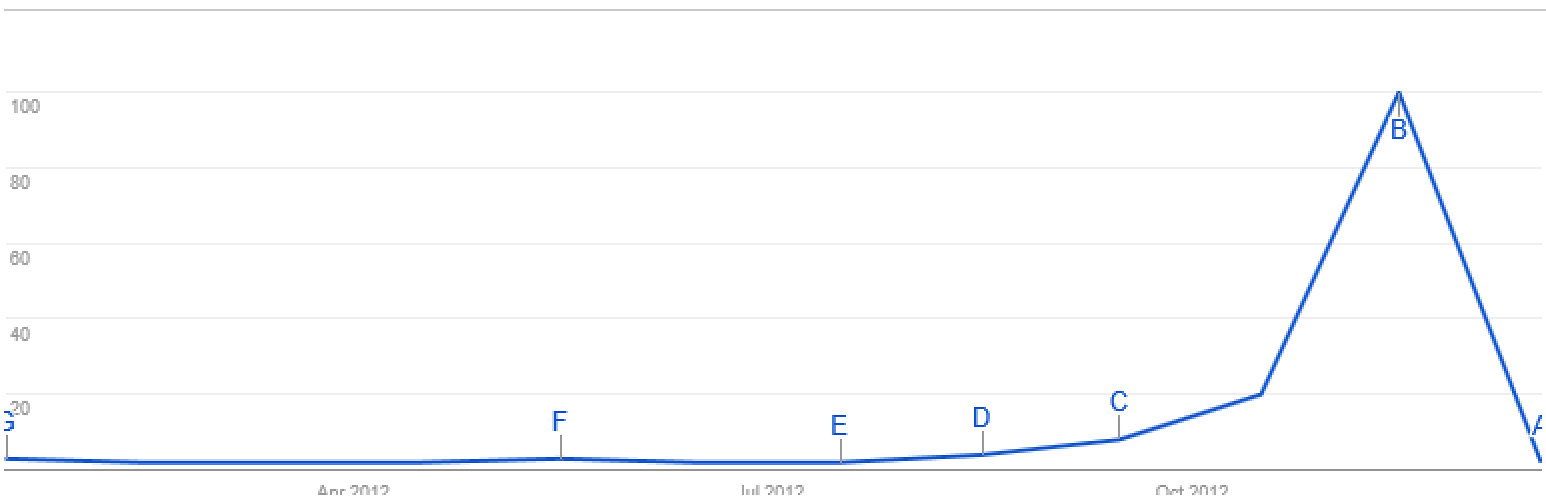


Categorizing Temporal Queries

- Ipad Mini

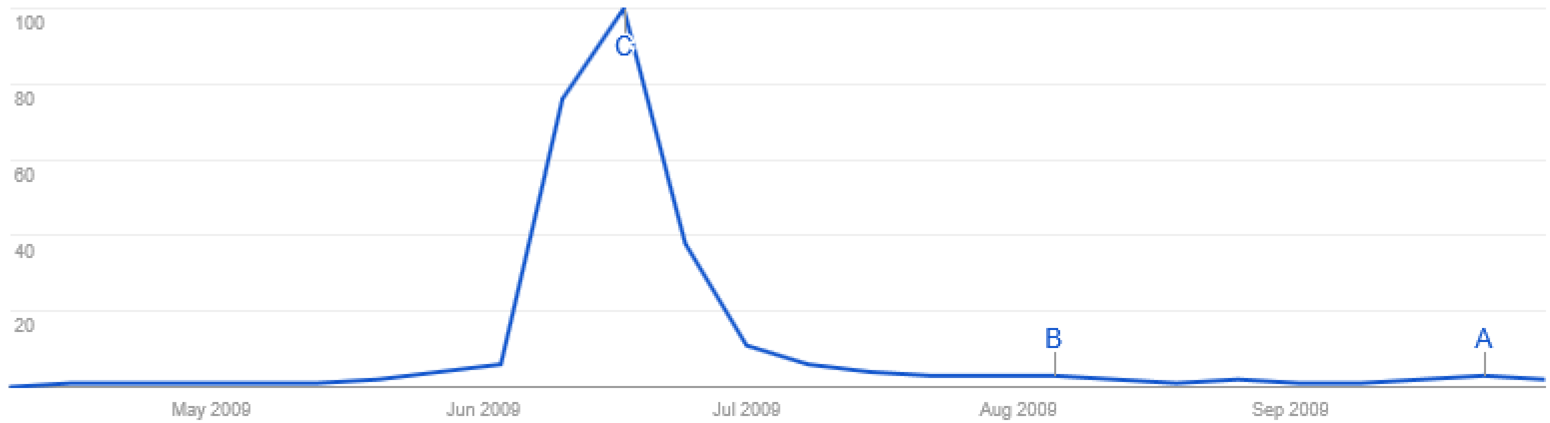
Categorizing Temporal Queries

- US Election



Categorizing Temporal Queries

- Iran Election



Categorizing Temporal Queries

- Query dynamics versus content changes

Table 3. Relationships between query popularity features and measures of result content change. Significant differences ($p < .05$) are shaded.

	Changes in TF		Changes in Dice	
# Spikes	Average	Median	Average	Median
0 (10%)	5.26	[M] 1.70	[M] 2.04	[M] 1.15
1 (47%)	7.52	2.95	[M] 3.01	1.80
M (43%)	8.00	[0] 3.47	[0,1] 4.12	[0] 2.54
Shape				
Castle (15%)	7.90	2.67	3.12	2.06
Sail (11%)	9.88	[W] 1.07	[W] 2.24	1.06
Wedge (54%)	7.58	[S] 3.90	[S] 3.94	2.45
Periodicity				
No (88%)	7.23	2.88	[Y] 3.19	[Y] 1.83
Yes (12%)	9.72	4.44	[N] 4.98	[N] 3.83
Trend				
Down (9%)	11.31	3.08	3.88	2.04
Flat (42%)	7.59	3.78	[UD] 3.83	[UD] 2.60
Up (36%)	7.53	[UD] 3.58	[UD] 3.67	2.28
Up-Down (13%)	7.52	[U] 1.43	[U,F] 2.43	[F] 1.04

Categorizing Temporal Queries

- Click entropy vs. change in intent.

Table 6. Correlation between two measures of change in query intent, including click entropy and the number of top rated results. Significant differences ($p < .05$) are shaded.

	Click Entropy	
	Average	Median
top HR Count	-0.28	-0.35

Table 9. Correlation between measures of change in query intent (*top HR Count*, *Click Entropy*) and change in result content (*Change in TF*, *Change in Dice*). Significant differences ($p < .05$) are shaded.

	Change in TF		Change in Dice	
	Average	Median	Average	Median
top HR Count	-0.16	-0.41	-0.40	-0.35
Click Entropy	0.15	0.48	0.38	0.31

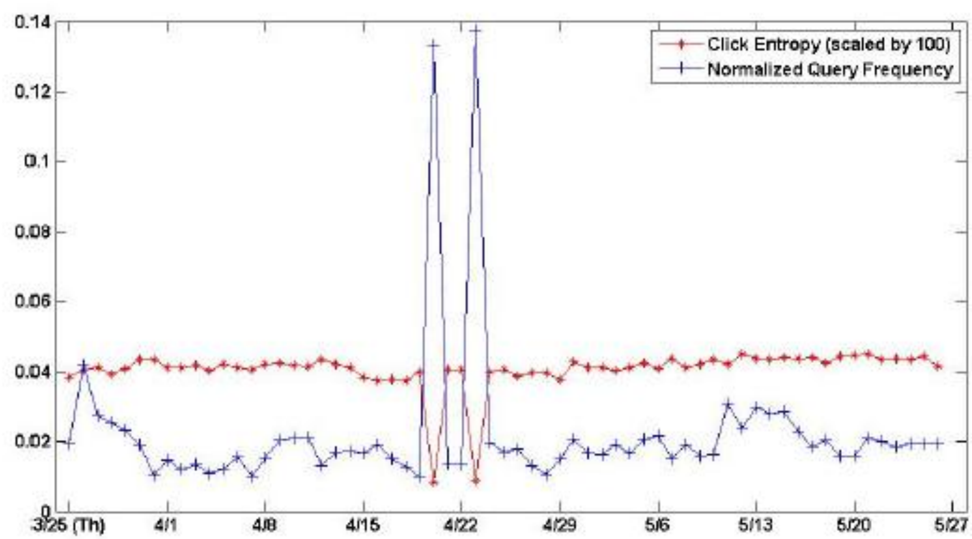


Figure 4. Normalized query frequency and click entropy for the query *lady gaga*.

Outline

- Query Dynamics
  - Hourly, Daily & Monthly Trends
- Categorizing Time-Sensitive Queries
  - Spike, Periodicity
- Modeling Query Dynamics
  - Burst detection, Time-Series
- Temporal Dynamics of User Behavior
  - Re-finding, Long-term vs. Short-term
- Temporal Patterns & Search Evaluation
  - Predicting Search Satisfaction (SAT)

Modeling Query Dynamics

Burst Detection, Time-Series

Burst Detection

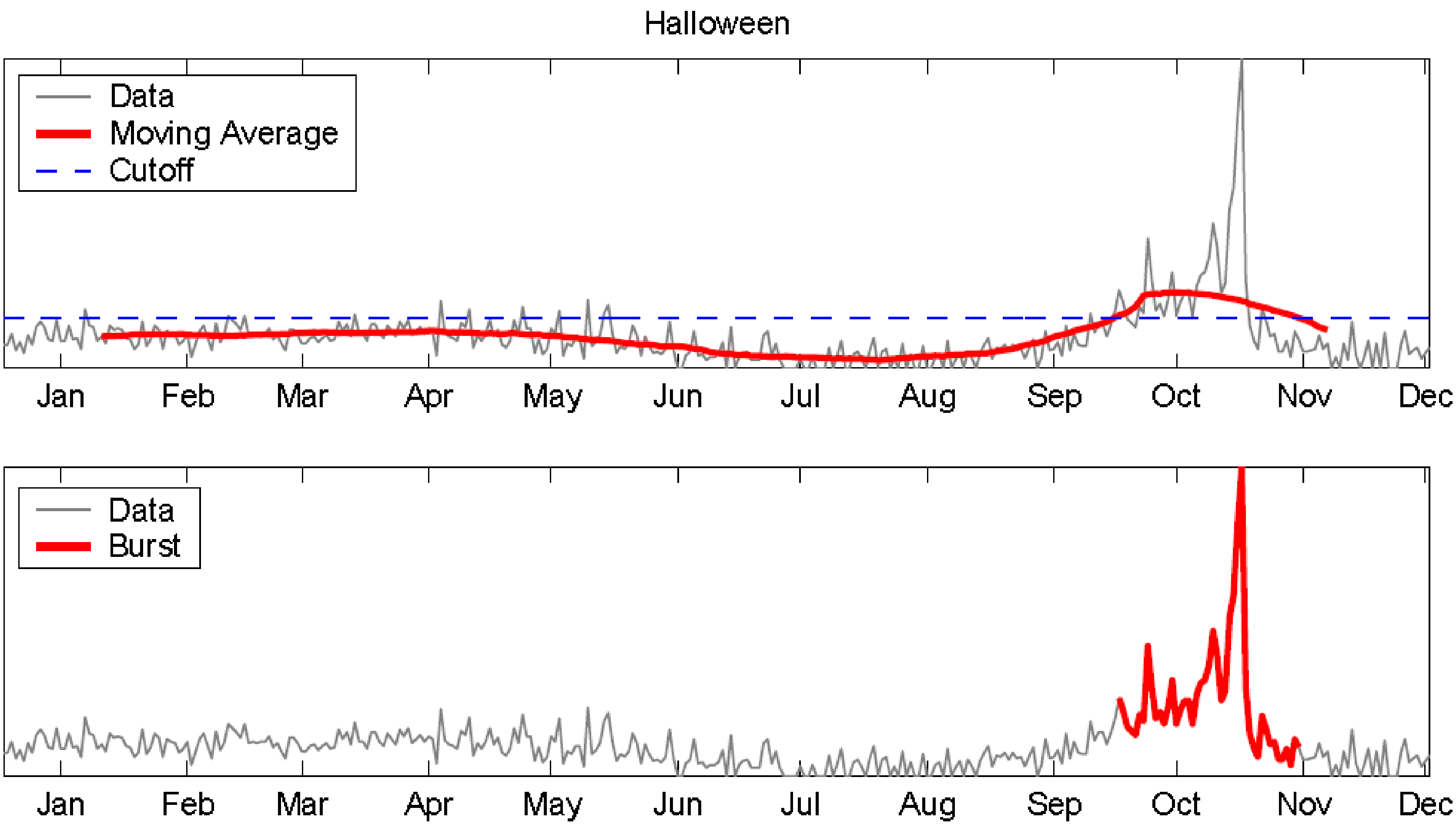


Figure 14: User demand for the query 'Halloween' during 2002 and the bursts discovered

Burst Detection

. . there seems something else in life besides time, something which may conveniently be called “value," something which is measured not by minutes or hours but by intensity, so that when we look at our past it does not stretch back evenly but piles up into a few notable pinnacles, and when we look at the future it seems sometimes a wall, sometimes a cloud, sometimes a sun, but never a chronological chart – ***E.M. Foster***

Burst Detection

- Bursty and hierarchical structure in streams [Kleinberg, KDD2002]
  - Simple randomized model
    - Gap x between messages i and i+1 is distributed according to the "memoryless" exponential density function f(x) = a*exp(-ax)
    - Expected gap = 1/a (rate)
  - A two-state model
    - State q0 (low) with a0 and state q1 (high) with a1
    - The state changes with Pr=p and remains at current state Pr=(1-p)
    - Each state sequence q induces a density function fq over sequences of gap.

Word	Interval of burst
data	1975 SIGMOD — 1979 SIGMOD
base	1975 SIGMOD — 1981 VLDB
application	1975 SIGMOD — 1982 SIGMOD
bases	1975 SIGMOD — 1982 VLDB
design	1975 SIGMOD — 1985 VLDB
relational	1975 SIGMOD — 1989 VLDB
model	1975 SIGMOD — 1992 VLDB
large	1975 VLDB — 1977 VLDB
schema	1975 VLDB — 1980 VLDB
theory	1977 VLDB — 1984 SIGMOD
distributed	1977 VLDB — 1985 SIGMOD
data	1980 VLDB — 1981 VLDB
statistical	1981 VLDB — 1984 VLDB
database	1982 SIGMOD — 1987 VLDB
nested	1984 VLDB — 1991 VLDB
deductive	1985 VLDB — 1994 VLDB
transaction	1987 SIGMOD — 1992 SIGMOD
objects	1987 VLDB — 1992 SIGMOD
object-oriented	1987 SIGMOD — 1994 VLDB
parallel	1989 VLDB — 1996 VLDB
object	1990 SIGMOD — 1996 VLDB
mining	1995 VLDB —
server	1996 SIGMOD — 2000 VLDB
sql	1996 VLDB — 2000 VLDB
warehouse	1996 VLDB —
similarity	1997 SIGMOD —
approximate	1997 VLDB —
web	1998 SIGMOD —
indexing	1999 SIGMOD —
xml	1999 VLDB —

Figure 4: The 30 bursts of highest weight in $\mathcal{B}_2^2$, using titles of all papers from the database conferences SIGMOD and VLDB, 1975-2001.

Burst Clustering

- Burst clustering [Parikh and Sundaresan, KDD2008]
- Matterhorn: new prod
- Cuestas: limited relea

followed by wide-sprea

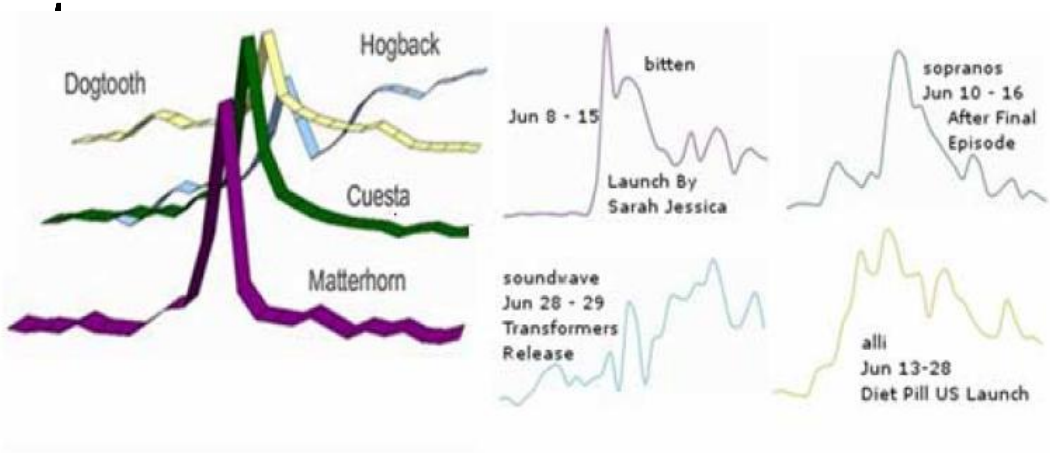
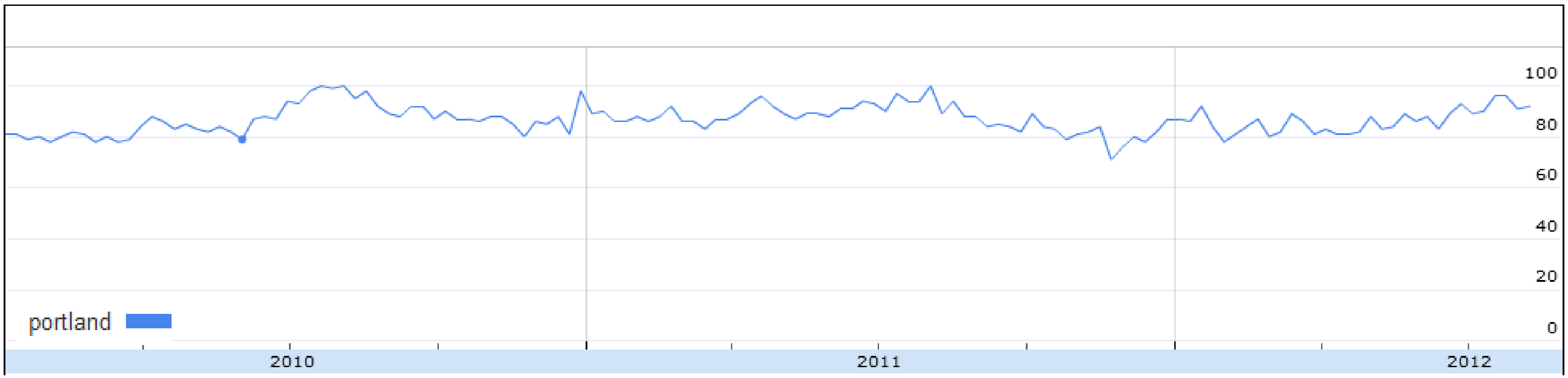


Figure 6 Labeling of Classes. Classes are named based upon the representative shapes of their centroids. X axis represents the time axis (day of the year), with burst period at the center. Y axis shows the relative normalized query frequencies, which gives an indication of the differences in amplitudes between burst and non burst periods for each of the 4 classes. 'bitten' is a Matterhorn, 'sopranos' is a Cuesta, 'alli' is a Dogtooth and 'soundwave' is a Hogback kind of burst.

Time-Series

- A time-series is a set of discrete or continuous observations over time.
  - Applications
    - Data modeling
    - Forecast
  - Examples
    - Sales figures
    - Student enrolment
    - CO_2 rate
    - Query popularity

Time-series (Single Exponential Smoothing)



$$\bar{y}_t = \lambda y_t + (1-\lambda)\bar{y}_{t-1}$$

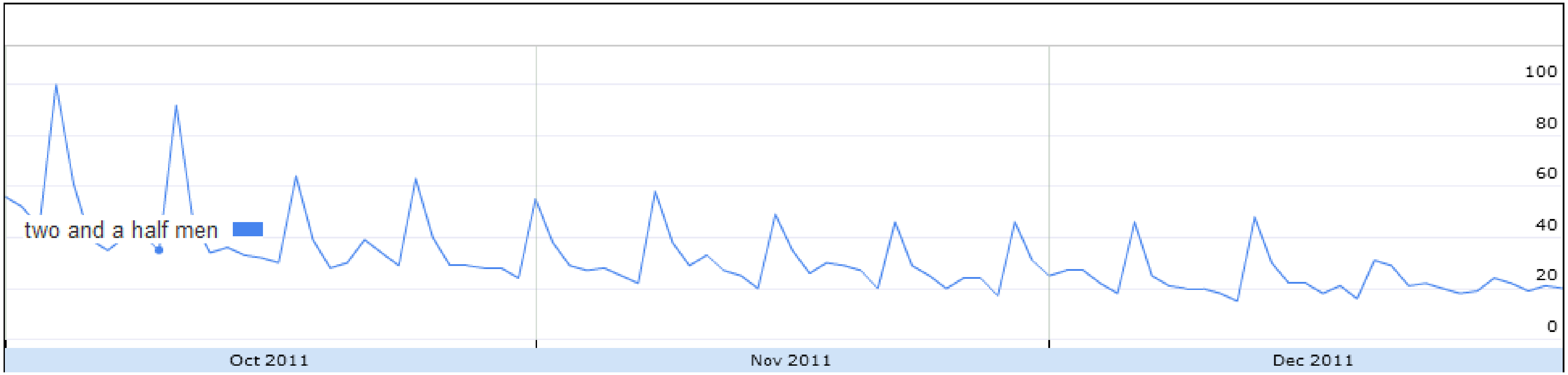
- The data points are modeled with a weighted average.
- $y, \bar{y}, \hat{y}$: Respectively represent actual, smoothed and predicted values at time t.
- λ: Smoot[illegible]ant
- Forecast: $\hat{y}_{t+1} = \bar{y}_t$

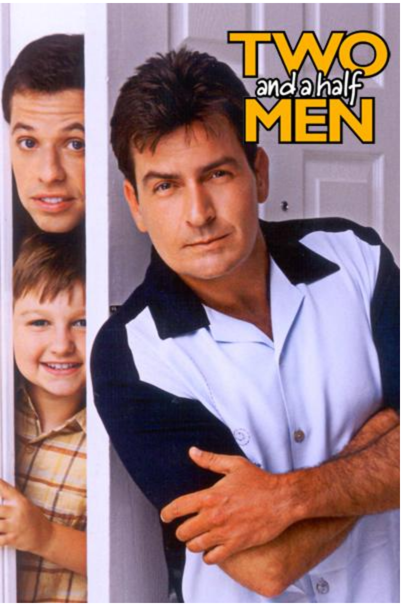
Time-Series (Double Exponential Smoothing)

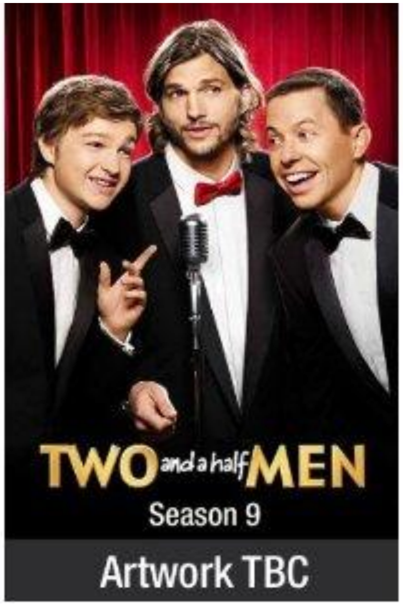
$$
\begin{aligned}
\bar{y}_t &= \lambda_1 y_t + (1-\lambda_1)(\bar{y}_{t-1} + F_{t-1}) \\
F_t &= \lambda_2(\bar{y}_t - \bar{y}_{t-1}) + (1-\lambda_2)F_{t-1}
\end{aligned}
$$

- $y, \bar{y}, \hat{y}$: Respectively represent actual, smoothed and predicted values at time t
- λ_1, λ_2: Smoothing constants
- F_t: Trend $\hat{y}_{t+1} = \bar{y}_t + F_t$
- Forecast:

Time-Series (Trends + Seasonality)







Time-series (Triple Exponential Smoothing)

$$
\begin{aligned}
\bar{y}_t &= \lambda_1(y_t - S_{t-\tau}) + (1-\lambda_1)(\bar{y}_{t-1} + F_{t-1}) \\
F_t &= \lambda_2(\bar{y}_t - \bar{y}_{t-1}) + (1-\lambda_2)F_{t-1} \\
S_t &= \lambda_3(y_t - \bar{y}_t) + (1-\lambda_3)S_{t-\tau} \\
& \lambda_1 + \lambda_2 + \lambda_3 = 1
\end{aligned}
$$

- $y, \bar{y}, \hat{y}$: Respectively represent actual, smoothed and predicted values at time t
- $\lambda_1, \lambda_2, \lambda_3$: Smoothing constants
- F_t: Trend factor at time t
- S_t: Seasonality factor at time t
- τ: Length of seasonal cycle
- Forecast: $\hat{y}_{t+1} = (\bar{y}_t + F_t)S_{t+1-\tau}$

Time-Series

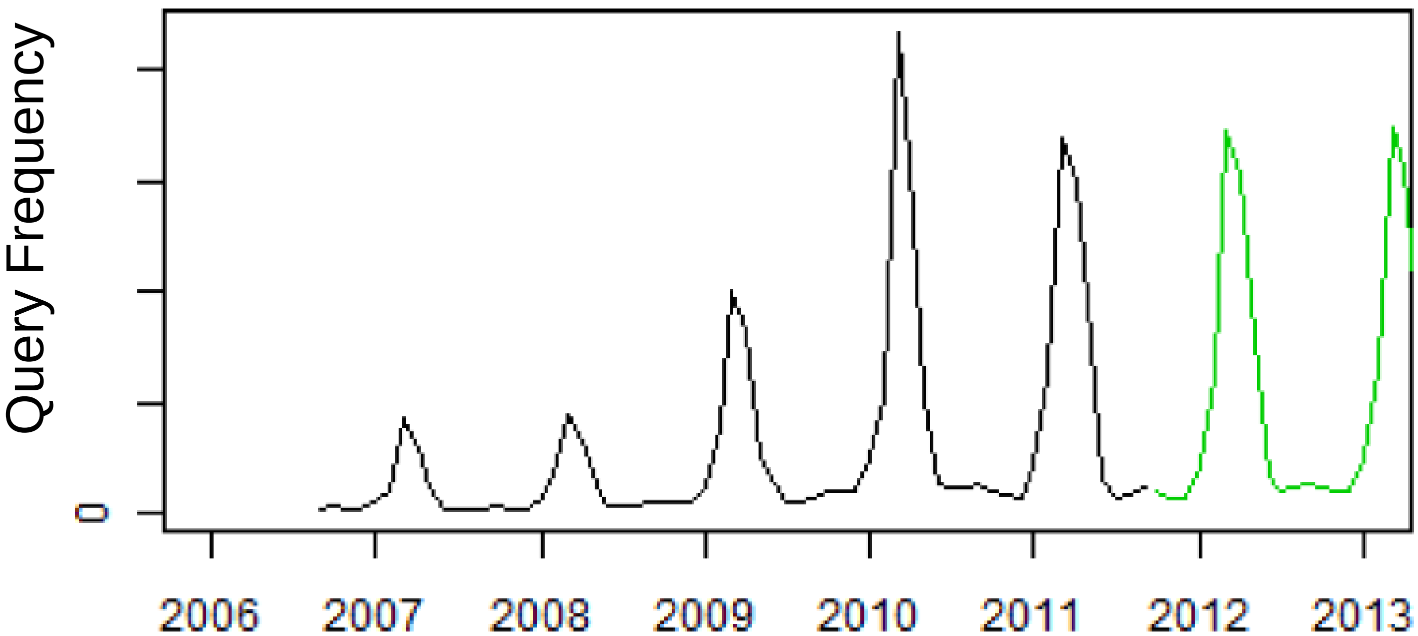


Figure 4: The black line shows monthly frequency values for query *spring flowers* between Sep'06–Jun'11. The green curve depicts the predicted values for the next 24 months (Jul'11–Jun'13) based on triple exponential smoothing. Data from bing.com.

Time-Series Modeling of Queries for Detecting Influenza Epidemics

- Ginsberg et al. [Nature 2009]
- Time-series models for 50 millions of the most popular queries

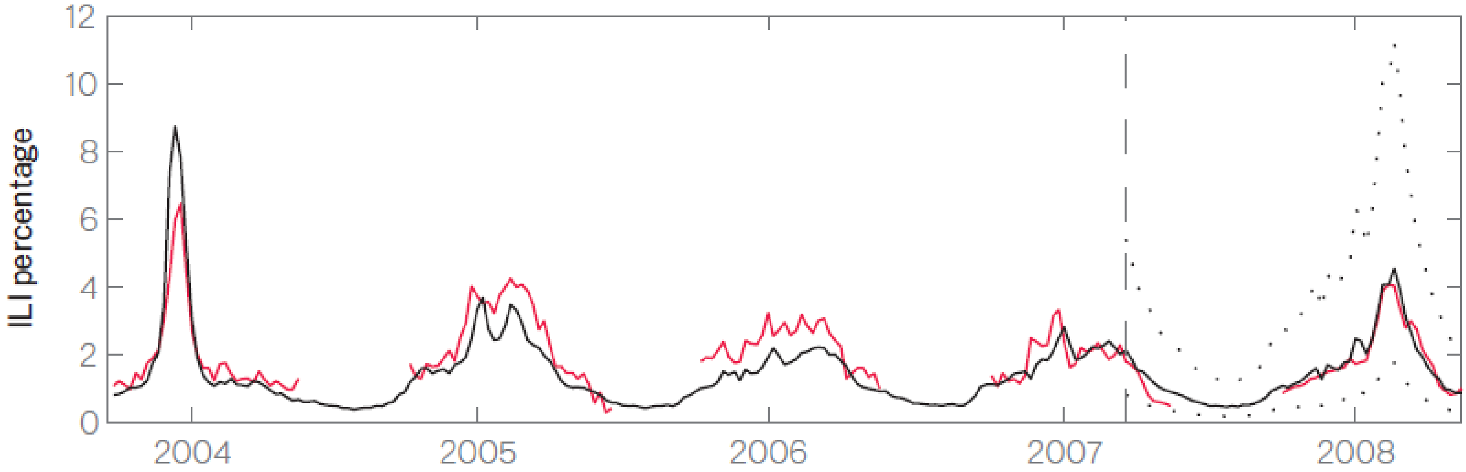


Figure 2: A comparison of model estimates for the Mid-Atlantic Region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated. A correlation of 0.85 was obtained over 128 points from this region to which the model was fit, while a correlation of 0.96 was obtained over 42 validation points. 95% prediction intervals are indicated.

Time-Series Modeling of Queries for Detecting Influenza Epidemics

- Publicly available historical data from the CDC's U.S. was used to train the models.
- The data was matched against the 50 million queries for finding the ones with the highest correlation.

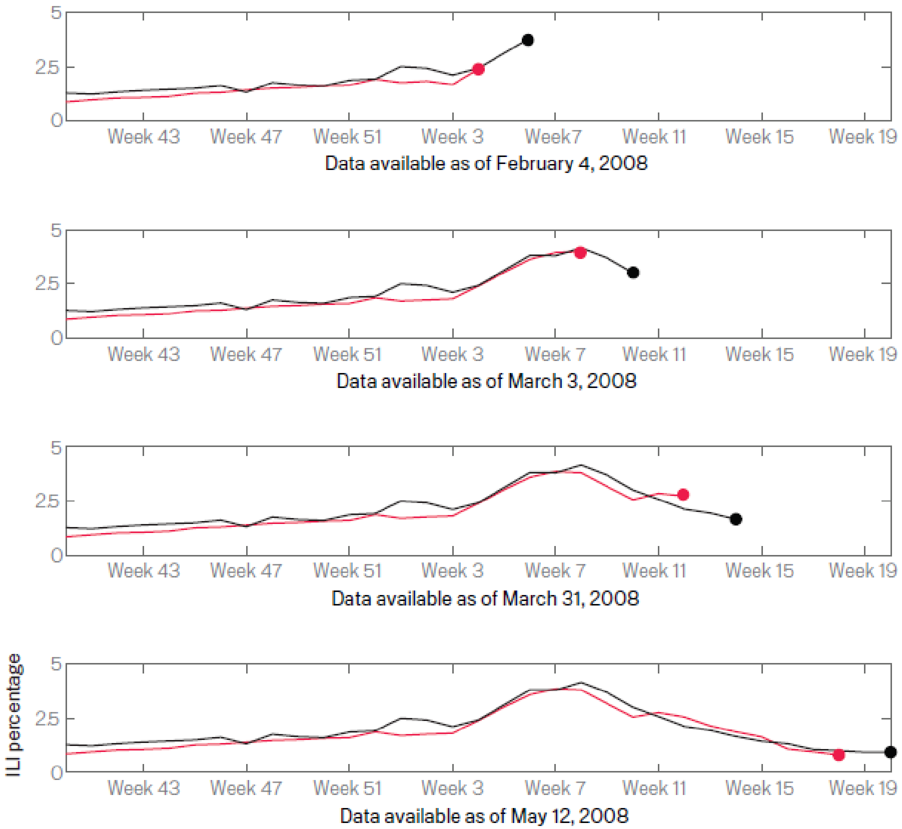


Figure 3: ILI percentages estimated by our model (black) and provided by CDC (red) in the Mid-Atlantic region, showing data available at four points in the 2007-2008 influenza season. During week 5, we detected a sharply increasing ILI percentage in the Mid-Atlantic region; similarly, on March 3, our model indicated that the peak ILI percentage had been reached during week 8, with sharp declines in weeks 9 and 10. Both results were later confirmed by CDC ILI data.

Classifying Seasonal Queries by Time-series

- Classifying seasonal queries [Shokouhi,

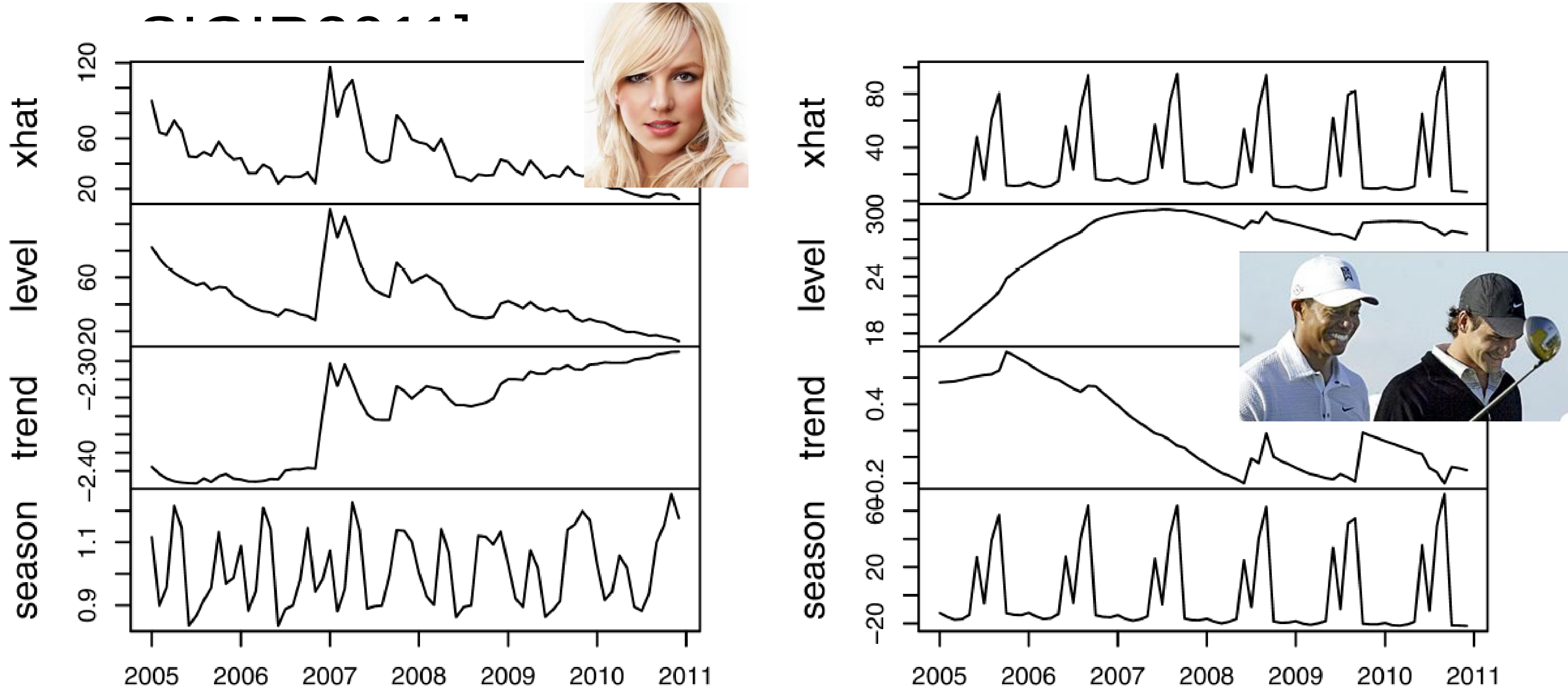


Figure 1: The Holt-Winters decomposition of time-series generated from monthly frequencies of "britney spears" (left), and "us open" (right). The xhat ($\hat{X}$) represents the raw monthly data, and is followed by the decomposed components in each plot. The raw data was collected from Google insight for search.

Periodicity Detection

- Discrete Fourier Transform

2.1 Discrete Fourier Transform. The normalized Discrete Fourier Transform of a sequence $x(n), n = 0, 1 \ldots N-1$ is a sequence of complex numbers $X(f)$:

$$X(f_{k/N}) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi kn}{N}}, \quad k = 0, 1 \ldots N-1$$

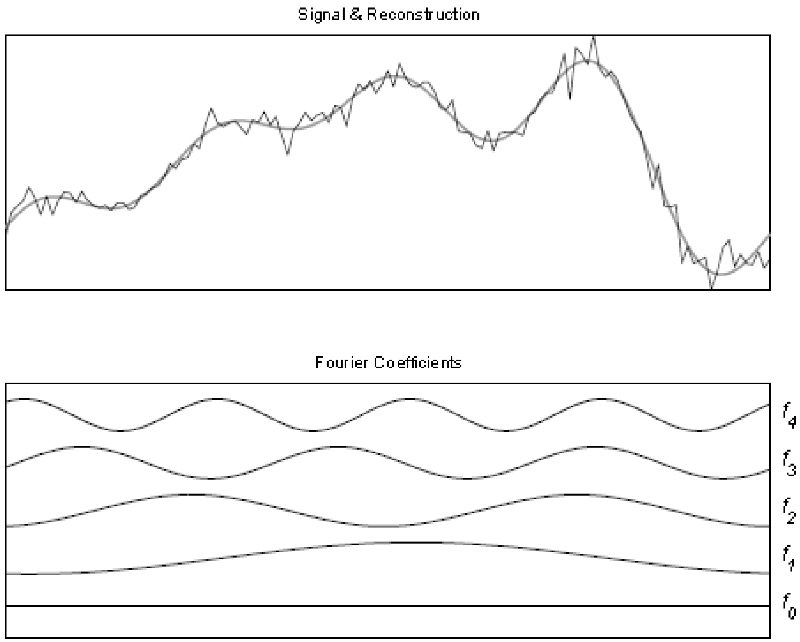


Figure 1: Reconstruction of a signal from its first 5 Fourier coefficients

- Periodogram

2.2.1 Periodogram Suppose that X is the DFT of a sequence x. The *periodogram* $\mathcal{P}$ is provided by the squared length of each Fourier coefficient:

$$\mathcal{P}(f_{k/N}) = \|X(f_{k/N})\|^2 \quad k = 0, 1 \ldots \lceil \frac{N-1}{2} \rceil$$

- Auto-Correlation

ACF), which examines how similar a sequence is to its previous values for different τ lags:

$$ACF(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(\tau) \cdot x(n + \tau)$$

Periodicity Detection

- Periodicity
  - the accuracy deteriorates for large periods
  - Spectral leakage
- Auto-Correlation
  - Automatic discovery of important peaks is more difficult
  - Multiplies of the same basic period also appear as peaks.
  - Low amplitude events of high frequency may look less important.

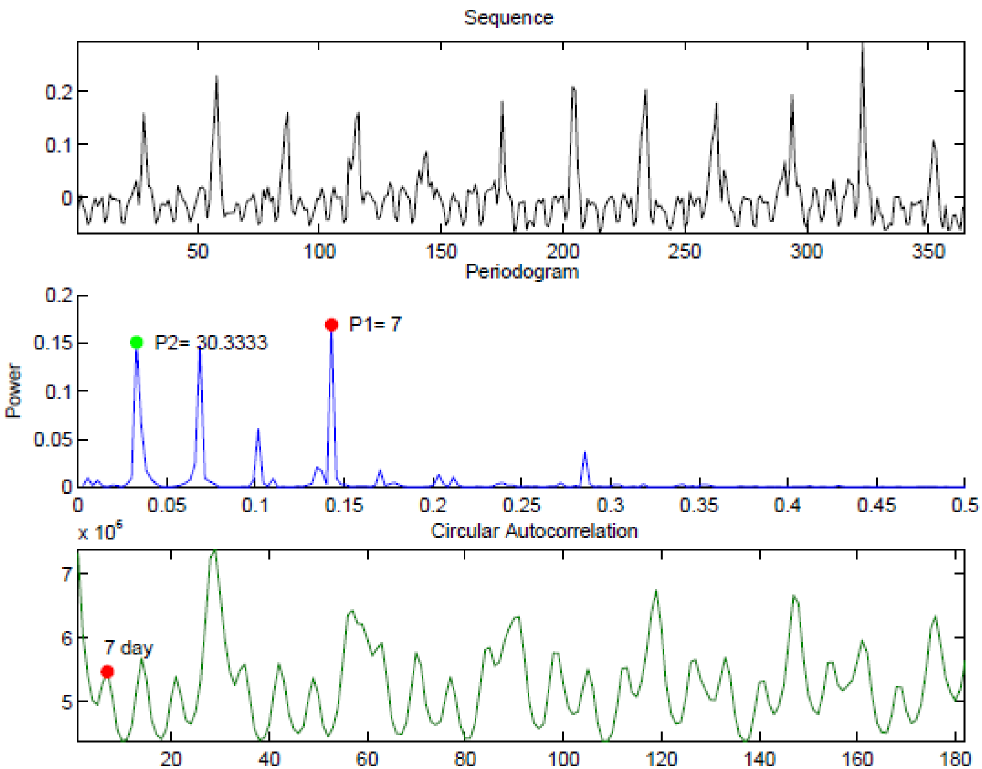


Figure 2: The 7 day period is latent in the autocorrelation graph, because it has lower amplitude (even though it happens with higher frequency). However, the 7 day peak is very obvious in the Periodogram.

Periodicity Detection

- Priodogram + Auto-Correlation[Vlachos et al., SDM2005]

Method	Easy to threshold	Accurate short periods	Accurate large periods	Complexity
Periodogram	yes	yes	no	$O(NlogN)$
Autocorrelation	no	yes	yes	$O(NlogN)$
Combination	yes	yes	yes	$O(NlogN)$

Table 1: Concise comparison of approaches for periodicity detection.

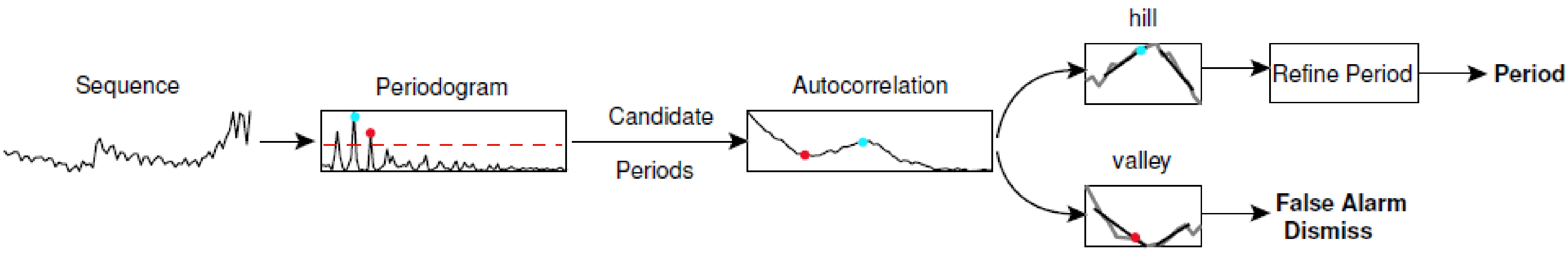


Figure 3: Diagram of our methodology (AUTOPERIOD method)

Periodicity Detection

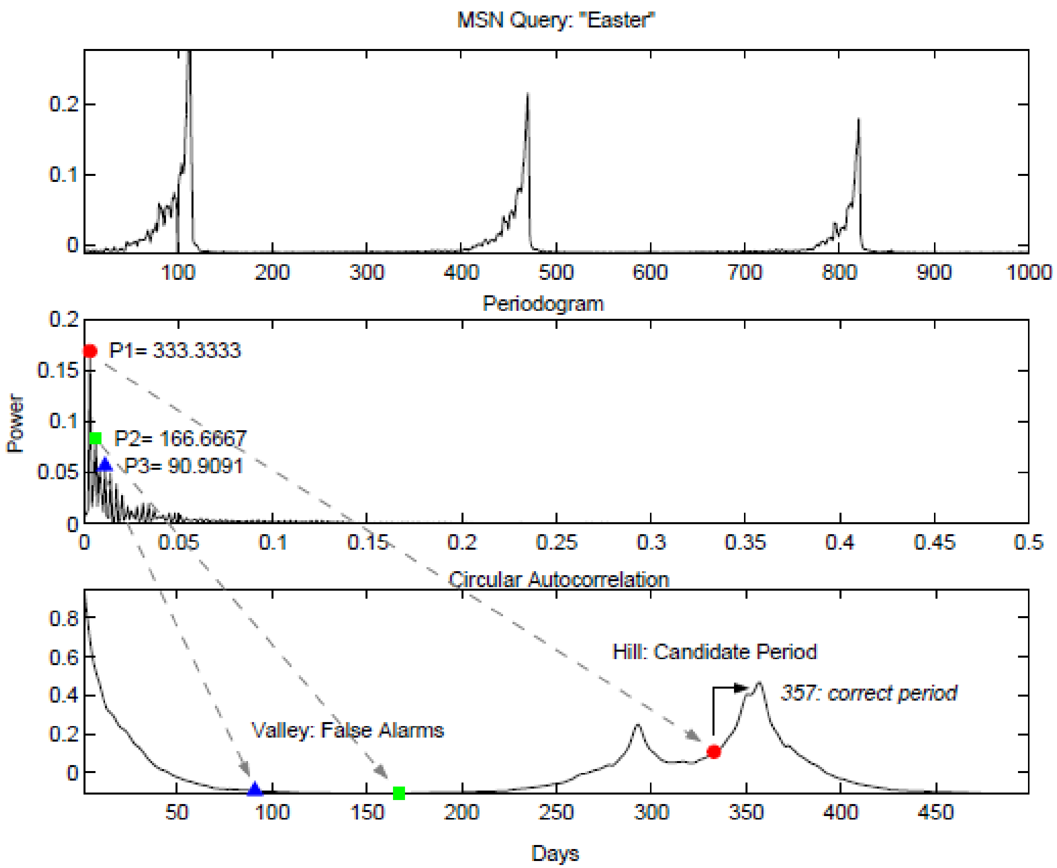


Figure 4: Visual demonstration of our method. Candidate periods from the periodogram are verified against the autocorrelation. Valid periods are further refined utilizing the autocorrelation information.

Learning to Predict Query Frequency

- Learning to predict frequency trends from time-series features [Radinsky et al., WWW2012]

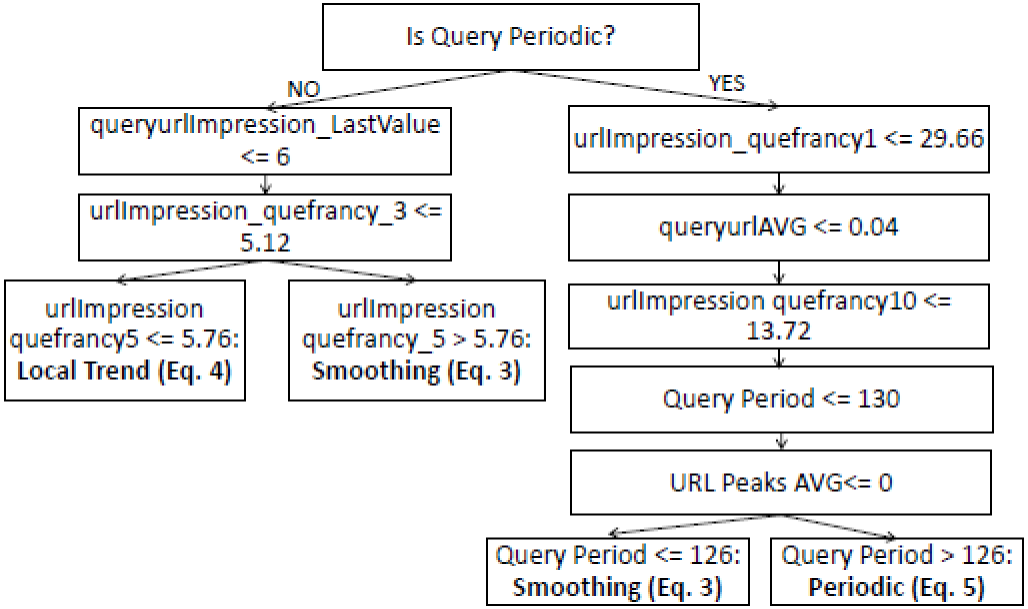


Figure 10: A part of the learned dynamic model.

Outline

- Query Dynamics
  - Hourly, Daily & Monthly Trends
- Categorizing Time-Sensitive Queries
  - Spike, Periodicity
- Modeling Query Dynamics
  - Burst detection, Time-Series
- Temporal Patterns in User Behavior
  - Re-finding, Long-term vs. Short-term
- Temporal Patterns & Search Evaluation
  - Predicting Search Satisfaction (SAT)

Temporal Dynamics of User Behavior

Long/Short History, Re-finding & Re-ranking

Long-term history

- Richardson [TWEB2008]
  - Query effects are long-lasting.
    - Users can be distinguished from their past queries
  - Long-lasting effects are useful for studying
    - Topic hierarchies
    - Temporal evolution of queries.
  - Learning from common similar trends in histories is useful
    - E.g. relationship between medical condition and potential causes.

Long-term history

- Example
  - The medical use of caffeine for migraine is common.
  - Migraine is highly correlated with caffeine in users search histories.

Table I. Dependency Score for Selected Terms vs. "Migraine"

Term	$dep_{migraine}(q)\ (\times 10^{-3})$
coffee	7.4
tea	8.2
coffee maker	10.1
caffeine	22.3
magnesium	24.7
dog	5.5
free	2.3

- Baseball: Beer
- Ski: Wine

Long-term history

- Comparing users by their long history

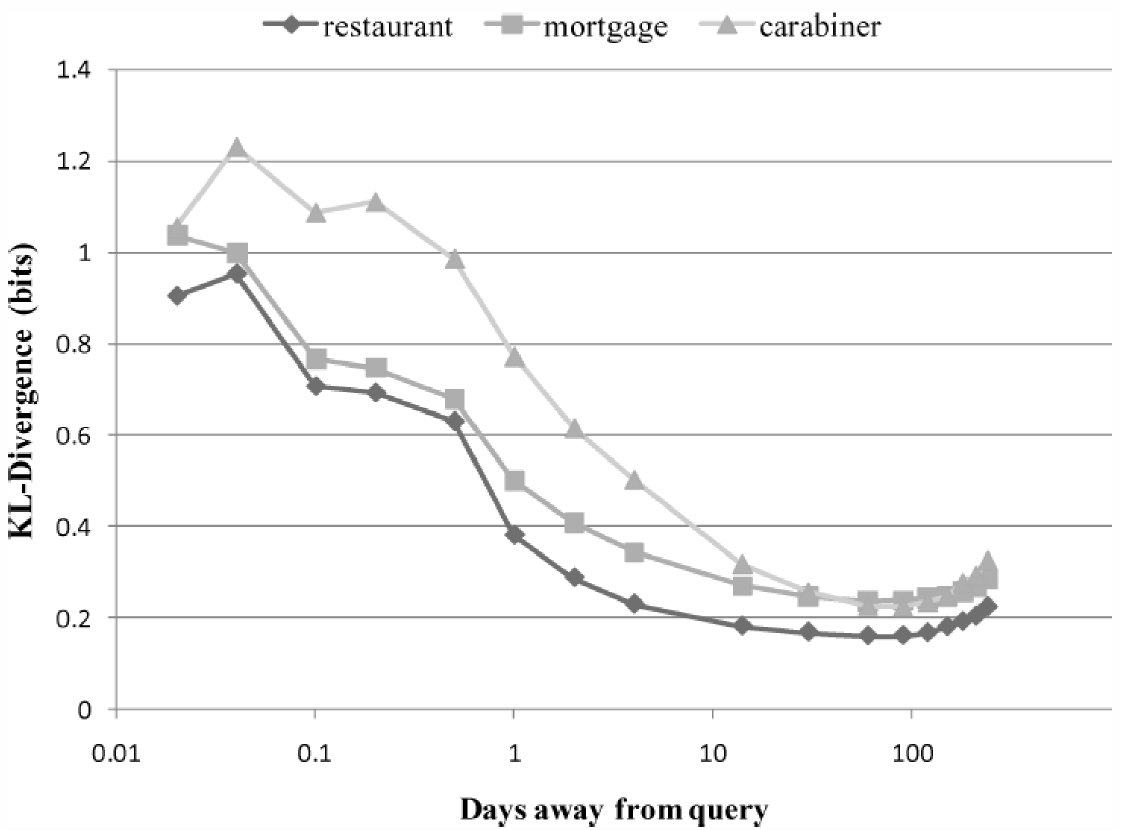


Fig. 2. KL-Divergence between users who issued a given query and the general user population. The divergence decreases over many days, and remains nonzero even months after the query.

Long-term history

- Temporal evolution of information needs

Table II. Queries Correlated with "Mortgage" Over Time (These change dramatically as the query is farther away in time. The users' interests move from mortgage basics to property searching, to insurance and taxes, to furnishings, to pools and patios. Here we give the top 40 terms that did not show up in the previous time period).

Time Period				
0–30 min	1–7 days	7–30d	30–90d	90–365d
mortgage	realtors	llc	kohls	patio
mortage	owner	associates	bath	harbor
mortgage	homes	insurance	overstock	outdoor
calculator	mls	lowes	barn	replacement
mortgages	remax	notary	sears	pools
lenders	property	depot	linens	hampton
calculators	financial	savings	beyond	lawn
countrywide	appraisers	construction	kmart	enterprise
gmac	builders	condo	pottery	ymca
refinance	prudential	business	walmart	vehicle
rates	zillow	secretary	outlet	supply
interest	bankruptcy	furniture	costco	resorts
broker	real	allstate	target	lake
lending	keller	companies	pier	rv
lender	properties	contractors	bed	walgreens
payment	agreement	cost	grill	newport
loan	appraisals	reverse	kitchen	lumber
amro	residential	federal	shield	oak
emc	lease	sale	macys	authority
brokers	county	housing	vacations	concrete
abn	modular	assessors	southwest	vehicles

Long-term history

- Generating topic hierarchies.
  - Long-term history could be more effective than short-term history for generating topics

Table IV. Topic Clusters for "Carabiner", Using Queries More than 90 Days Before or After the Original Query. (Shown are the top terms related to carabiner for various values of the smoothing parameter, m. The terms within a given smoothing value tend to belong to the same topic and level of generality.)

m	Terms
1k	mammut, petzl, botach, kydex, clevis, extrication, trijicon, webbing, eotech, aimpoint, sportiva, utilize, boker, surefire, aiming, coolmax, scarpa, 5d11, nomex, armament

Table V. Topic Clusters for "Carabiner", Using Within-session Queries. (i.e., queries that occurred within half an hour of the original carabiner query. Shown are the top terms related to carabiner for various values of the smoothing parameter, m. Unlike IV, which used queries from a broader time period, we do not see the same topic generalization as the smoothing parameter increases.)

m	Terms
1k	carabiners, caribiner, carbiner, carabeaner, caribener, carabineer, carabener, carabeener, caribeaner, caribiners, karabiner, carabina, biner, screwgate, caribeener, carabine, carrabiner, carabiener, carabeners, carribeaner

Long-term history

- Temporal querying behaviour.
  - Do men buy the ring first or figure out how to propose?

Long-term history

- Temporal querying behaviour.

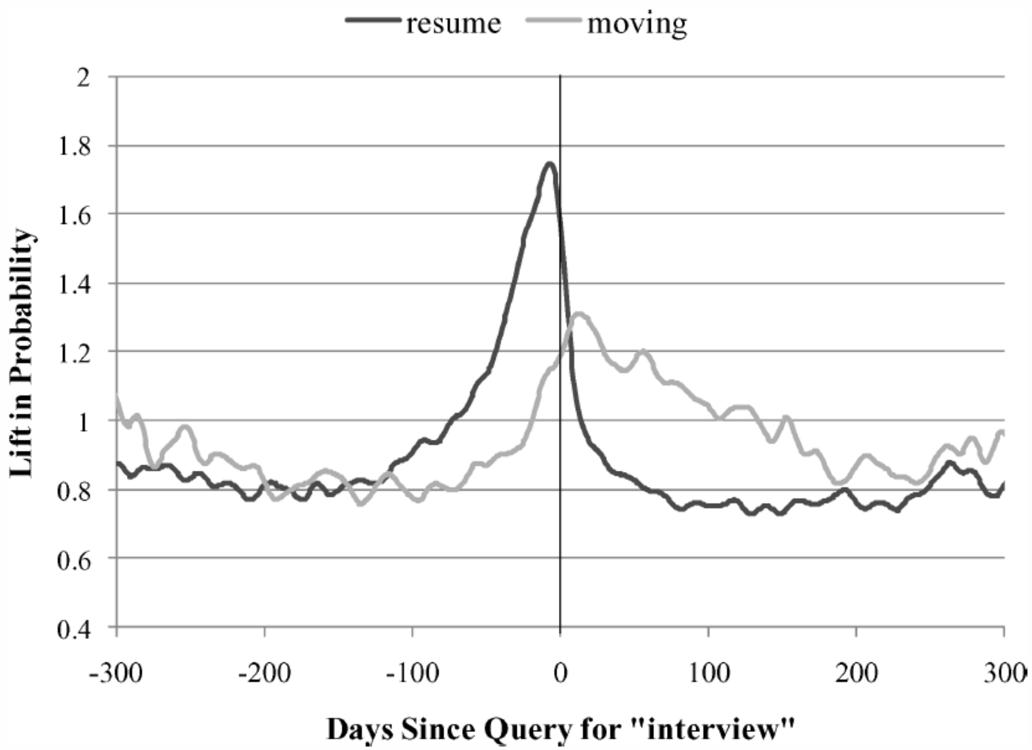
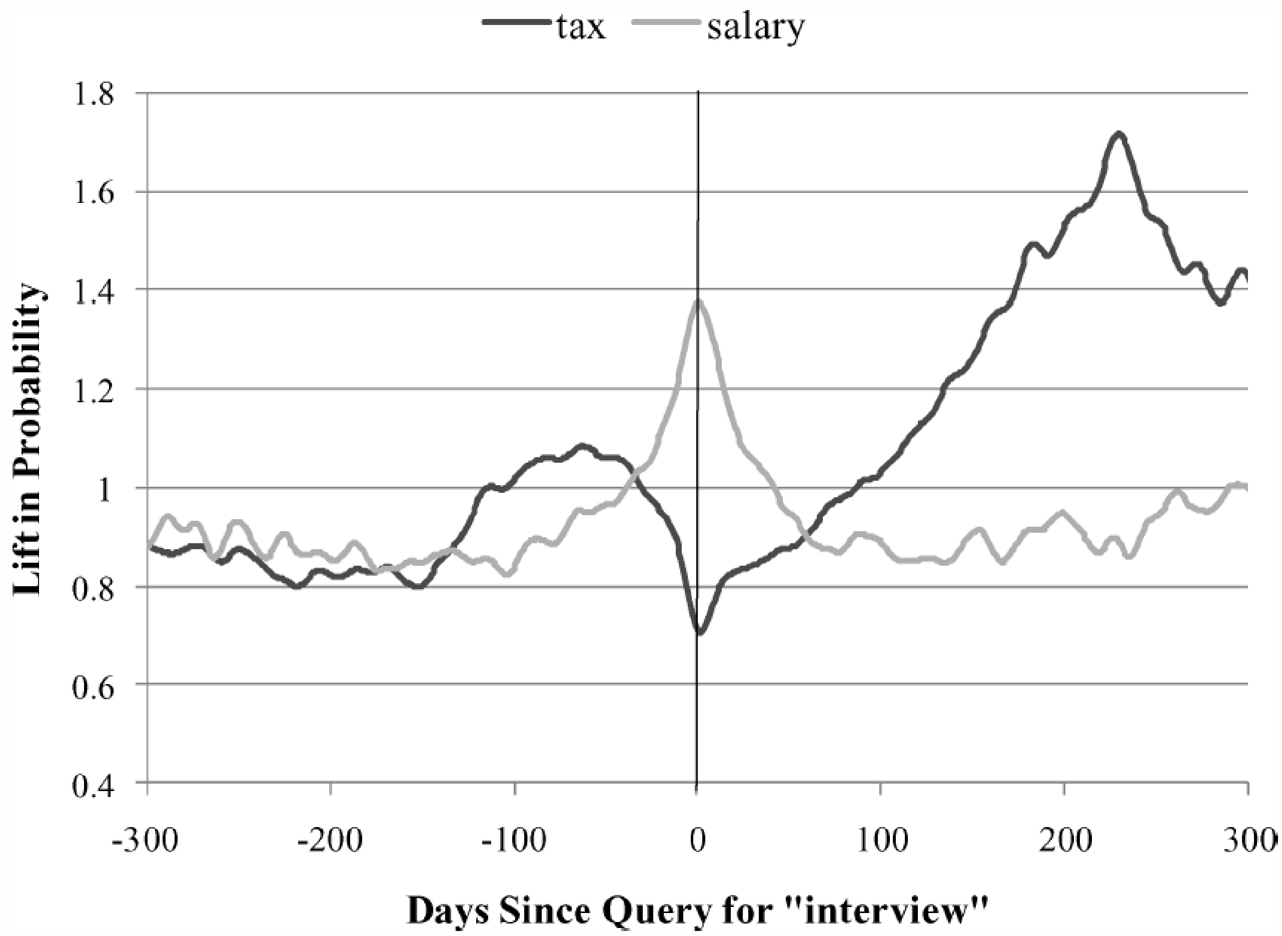


Fig. 4. $S(\delta)$ (lift in probability) for the queries “resume” and “moving” given the reference query “interview”. People begin looking for information on resumes up to 100 days before the interview query; most look immediately before. Users become significantly more interested in moving information after the interview query.

Long-term history

- Temporal querying behaviour.



Long-term history

- Temporal querying behaviour.

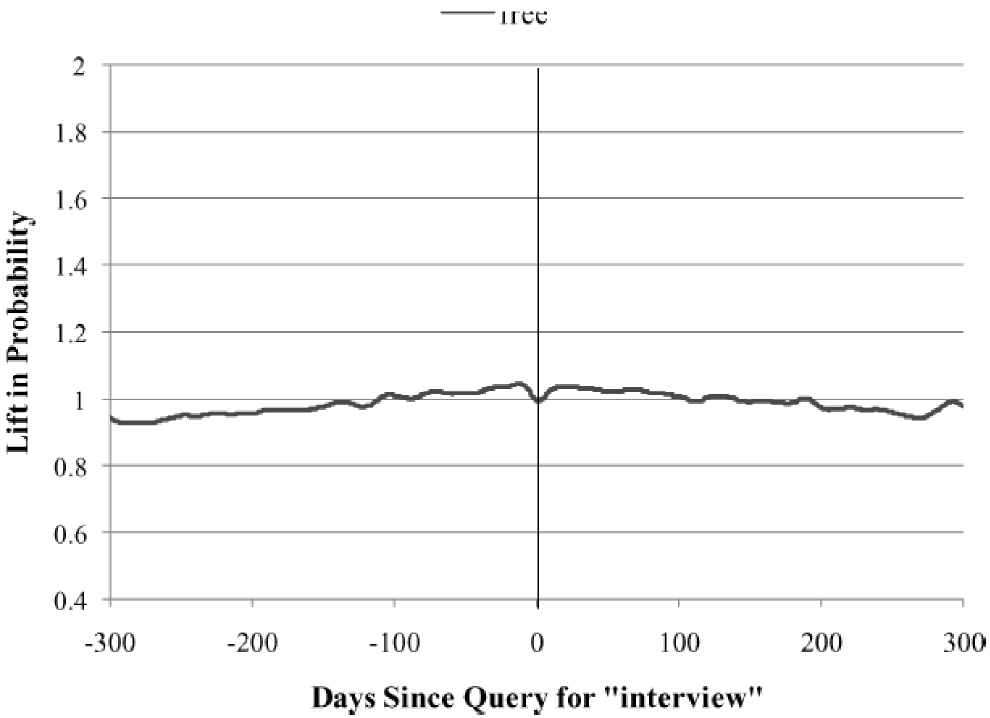


Fig. 6. A temporally unrelated query, ("free"), appears as a flat line.

Table VII. KL-Divergence for Temporal Distribution of Various Terms Relative to the Query "Interview"

Term	KL-Divergence ($\times 10^{-2}$)
resume	5.67
tax	3.12
moving	2.21
salary	2.03
free	0.14

Re-finding

- Traces on query logs of 114 anonymous users [Teevan et al. SIGIR'07]
  - Up to 40% re-finding
- Large-scale log analysis [Tyler & Teevan WSDM2010]
  - 30% of single-click Queries
  - 5% of multi-click queries
  - 66% of re-finding queries are previous queries for later re-findings
  - 48% of re-findings happens within a single session

Re-finding

- Predicting personal navigation [Teevan et al. WSDM11]

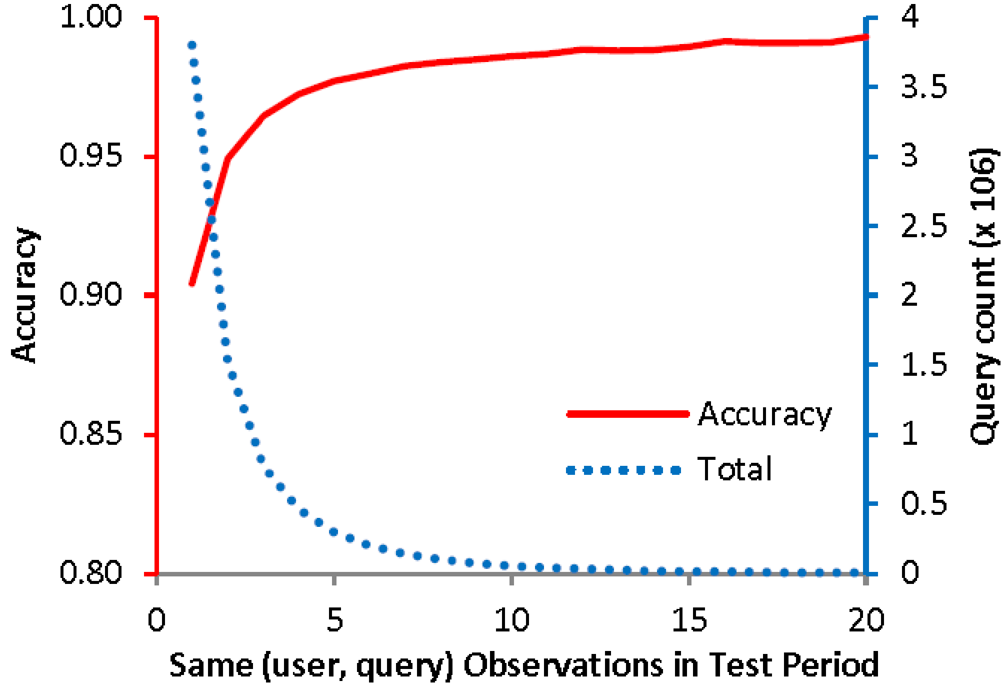


Figure 2. The accuracy of the personal navigation prediction as a function of how often the individual has used the same query for personal navigation.

Re-finding & Re-Ranking

- Predicting personal navigation [Teevan et al. WSDM'11]

Personal Navigation Prediction Algorithm

1. Given a query q_i issued by a user,
2. Select the two most recent queries ($q_{i\text{-}1}$ and $q_{i\text{-}2}$) from the user's history such that:
   - $q_{i\text{-}1} = q_i$ and $q_{i\text{-}2} = q_i$, and
   - $|\text{urls clicked}(q_{i\text{-}1})| > 0$, and
   - $|\text{urls clicked}(q_{i\text{-}2})| > 0$.
3. Predict the user will click $u \in \{\text{urls clicked}(q_{i\text{-}1})\}$ iff:
   - $q_{i\text{-}1} \neq null$ and $q_{i\text{-}2} \neq null$, and
   - $|\text{urls clicked}(q_{i\text{-}1}) \cup \text{urls clicked}(q_{i\text{-}2})| = 1$.

Figure 1. The personal navigation prediction algorithm. A personal navigation query is one that was used to find a particular site the past two times it was issued by the user.

Re-finding & Re-Ranking

- Personal level re-finding [Dou et al., WWW2007]
  - #previous clicks on query-url pairs
  - #previous click on urls from the same topic
  - Re-ranking most effective on comment web search queries with high-entropy click distribution.
  - Using both short-term and long-term contexts is better than using one of them alone.

Long-term vs. Short-term

- Long vs. Short for search personalization [Bennett et al. SIGIR2012]

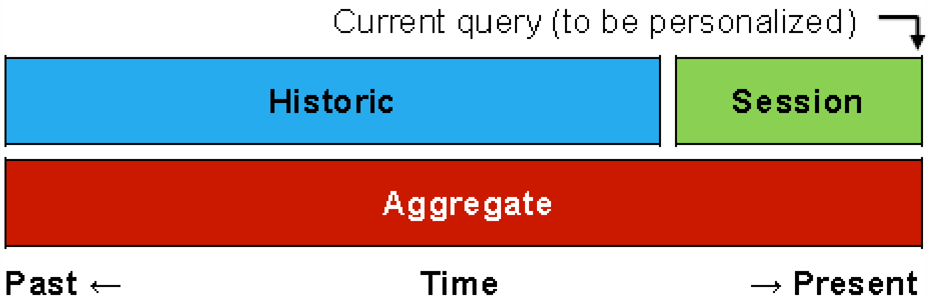


Figure 1. An illustration of how we create profiles from recent (*Session*), past (*Historic*), or a combination (*Aggregate*).

Long-term vs. Short-term

- Long-term gains are generally higher

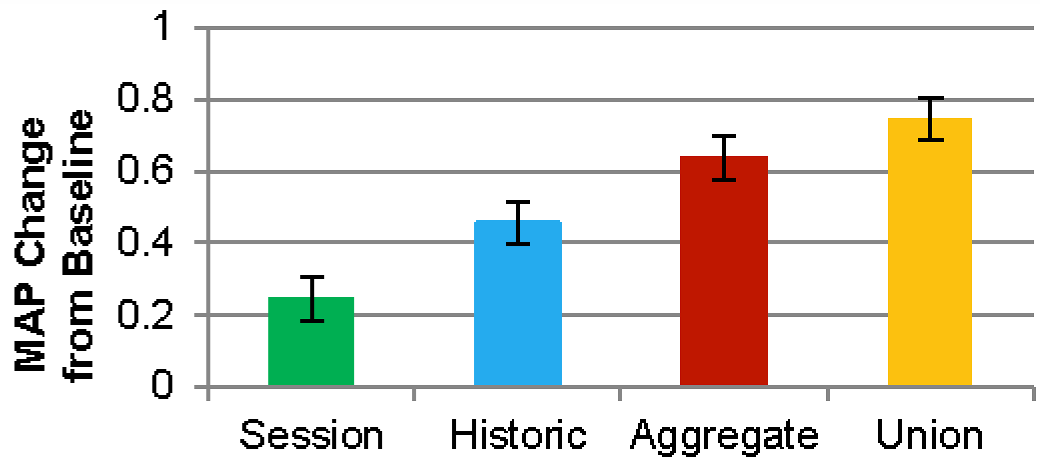


Figure 4. Average change in MAP from baseline ranker MAP

Long-term vs. Short-term

- Long-term features are more effective for personalization early in the session

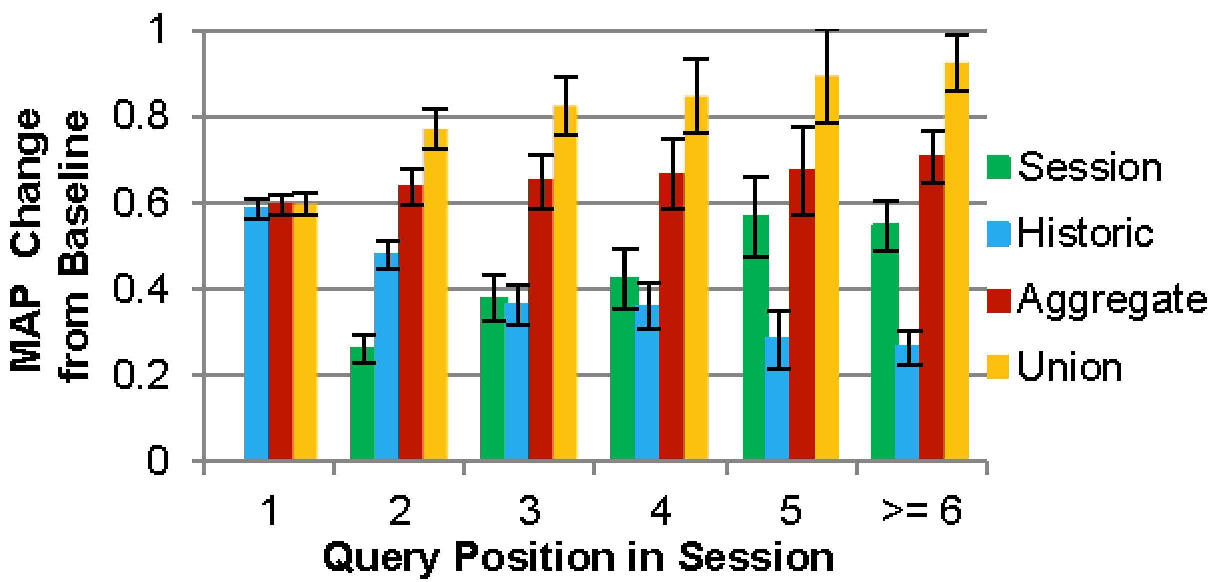


Figure 5. Avg. change in MAP by position of query in session.

Cross-Device Search

- People frequently search cross-device (15% about continuous task) [Wang et al. WSDM2013]

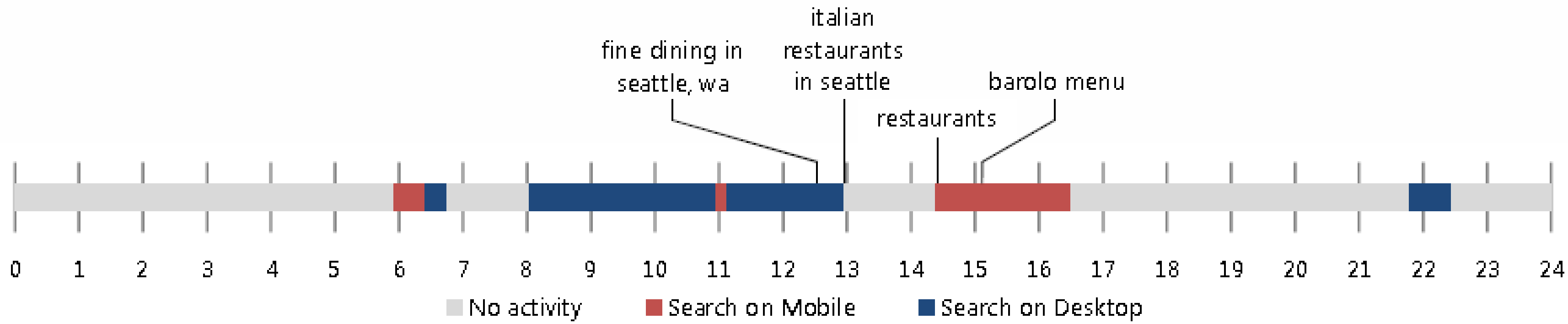


Figure 1. Search activities on mobile and desktop of a fictitious user over the course of a single day. Numbers denote hours from midnight. Queries of interest (relevant to the body of the paper) are included above the figure for reference.

Cross-Device Search

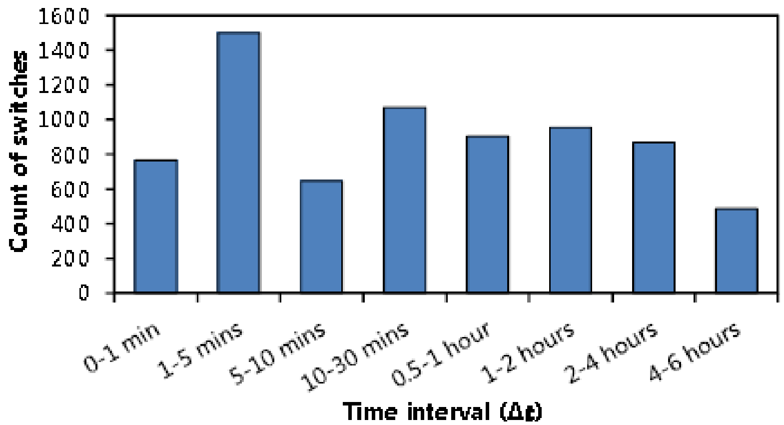


Figure 2. Time interval distribution of same-query switches.

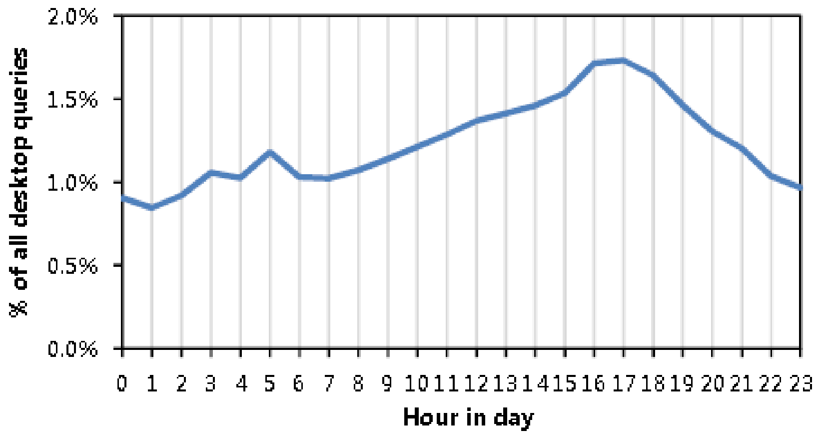


Figure 4. Percentage of pre-switch queries on desktop over time.

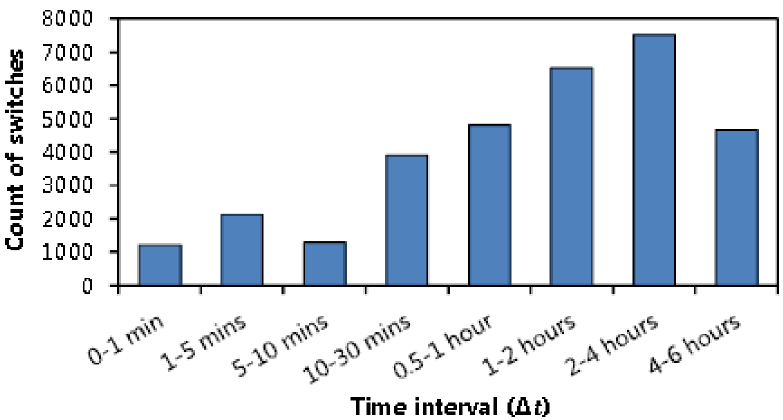


Figure 3. Time interval distribution of different-query switches.

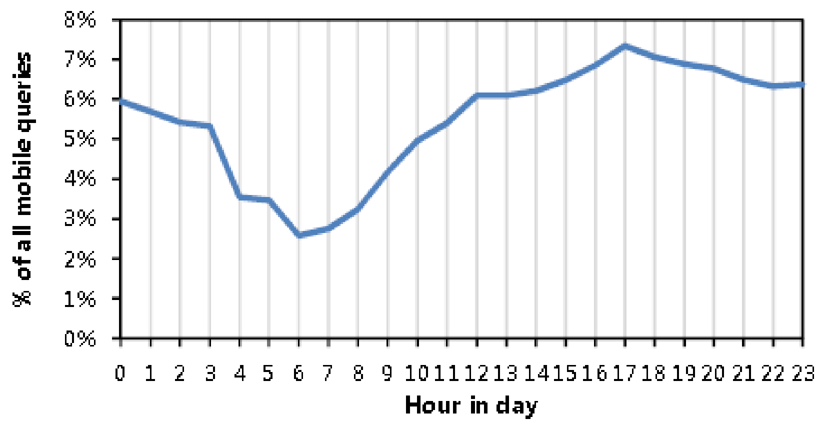


Figure 5. Percentage of post-switch queries on mobile over time.

Outline

- Query Dynamics
  - Hourly, Daily & Monthly Trends
- Categorizing Time-Sensitive Queries
  - Spike, Periodicity
- Modeling Query Dynamics
  - Burst detection, Time-Series
- Temporal Dynamics of User Behavior
  - Re-finding, Long-term vs. Short-term
- Temporal Dynamics of User Behavior for Evaluation
  - Predicting Search Satisfaction (SAT)

Temporal Dynamics of User Behavior for Search Evaluation

Predicting Search Satisfaction & Click Modeling

Search Difficulty vs. Task Time

- 179 participants [Aula et al. CHI2010]
- Difficult tasks take longer

	All tasks	**Successful tasks**	**Unsuccessful tasks**
Average time on task	223.9 (2.36)	176.2 2.24	384.6 (3.52)
Average number of query terms/query	4.77 (0.029)	4.66 (0.030)	5.13 (0.027)
Average number of queries/task	6.71 (0.098)	4.98 (0.070)	12.41 (0.098)
Proportion of queries with advanced operators ('+', '-', 'AND', 'OR', ':')	0.074 (0.0024)	0.056 (0.0046)	0.133 (0.0038)
Proportion of queries with question	0.047 (0.0020)	0.043 (0.0047)	0.060 (0.0025)

Table 1. Descriptive statistics for all tasks and separately for successful and unsuccessful tasks. Values are means (1 std error in brackets).

Search Difficulty vs. Task Time

- More time spent for difficult tasks

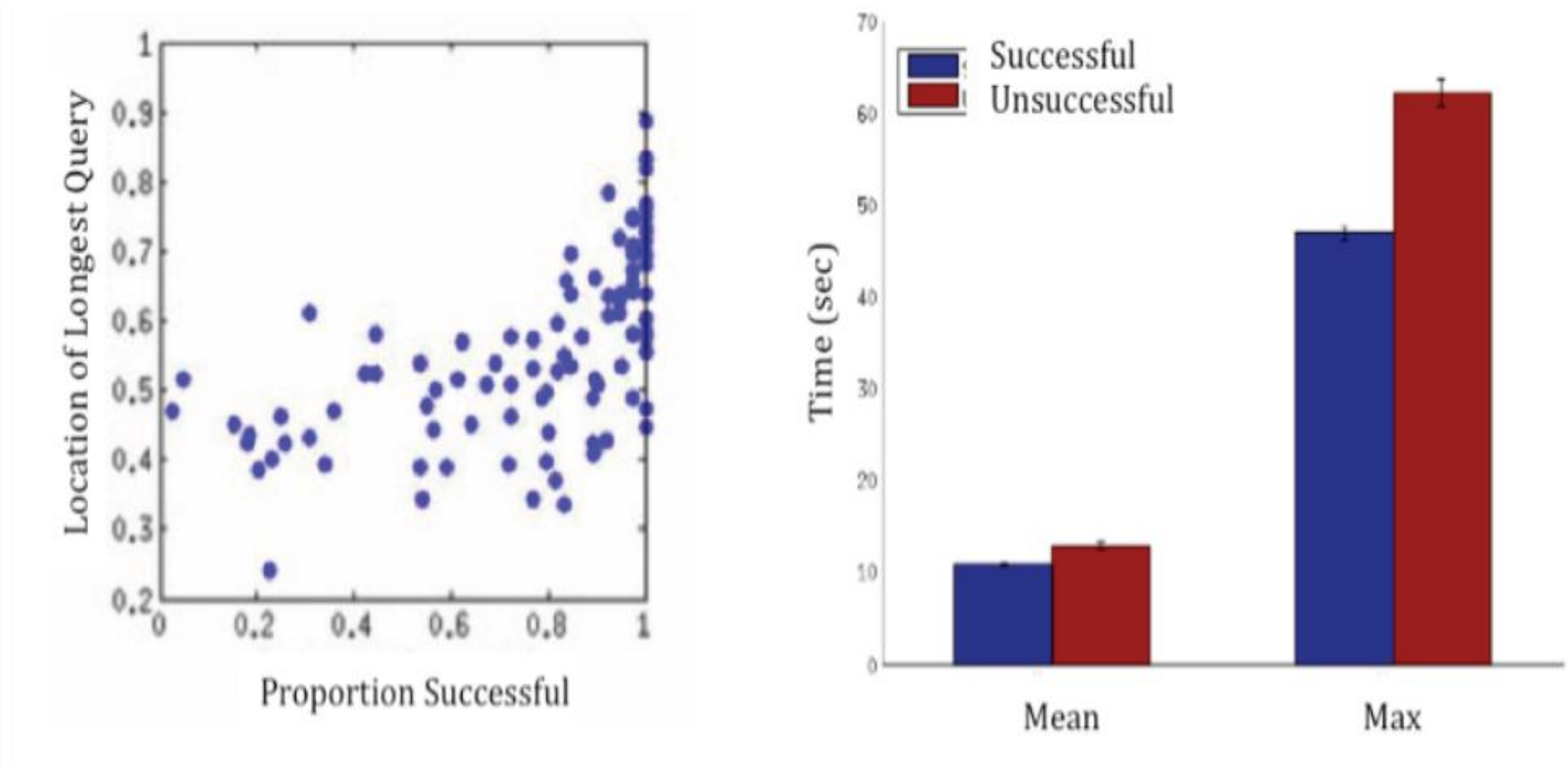


Figure 3. Graph on the left shows the location of the longest query in the search session as a function of the proportion of participants who were successful in the task. Graph on the right shows the mean and maximum time users spent on the search result page in successful and unsuccessful tasks.

Search Difficulty vs. Task Time

- More time spent on SERP for difficult tasks

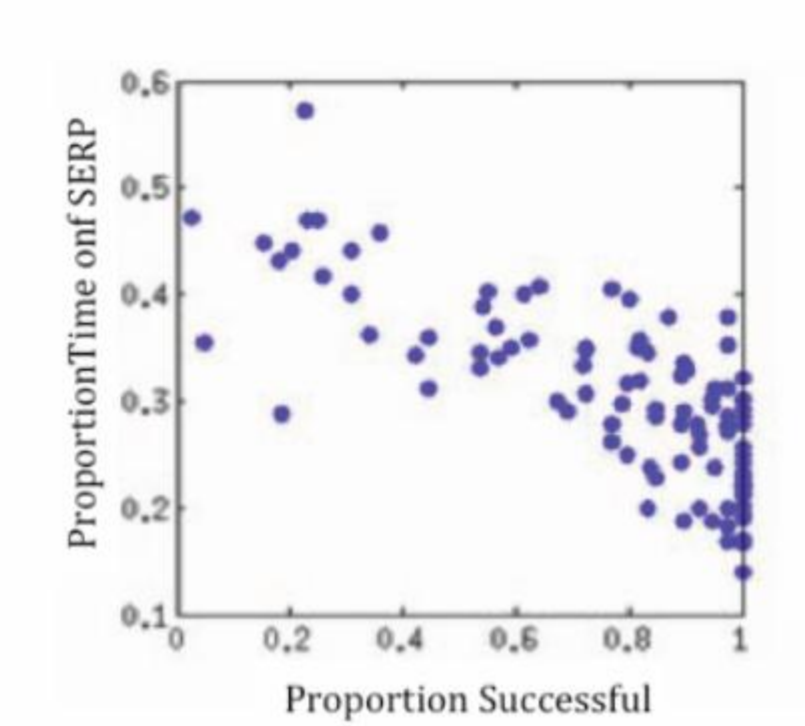


Figure 4. Graph on the left shows the proportion of total task time the users spent on the search result page as a function of task success.

Implicit Measures for evaluation

- Fox et al. [TOIS2005] compared several implicit signals.
- Such signals (e.g. SAT-Clicks) are particularly useful for training personalized rankers

Table 2: Result-Level Implicit Measures

Result Level Measure	Description
Time Difference in Seconds Duration in Seconds	Time spent on a page is represented with two different measures. *Difference in seconds* is time from when the user left the results list to the time they returned. *Duration in seconds* is the subset of the above time during which the page was in focus.
Scrolled, Scrolling Count, Average Seconds Between Scroll, Total Scroll Time, Maximum Scroll	Each time a user scrolled down the page a 'scrolled' event was logged, along with the percentage of the page that the user moved within that scroll and a timestamp.
Time To First Click, Time To First Scroll	Initial activity times. Time to first click and first scroll.

Implicit Measures for evaluation

- SAT-Prediction accuracy based on result-level features.

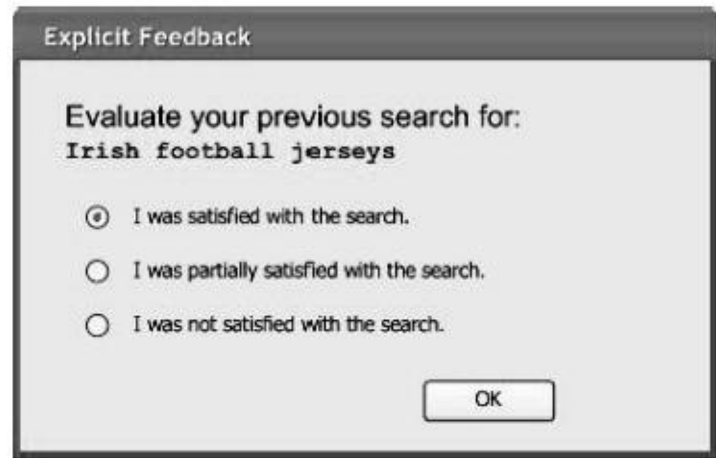
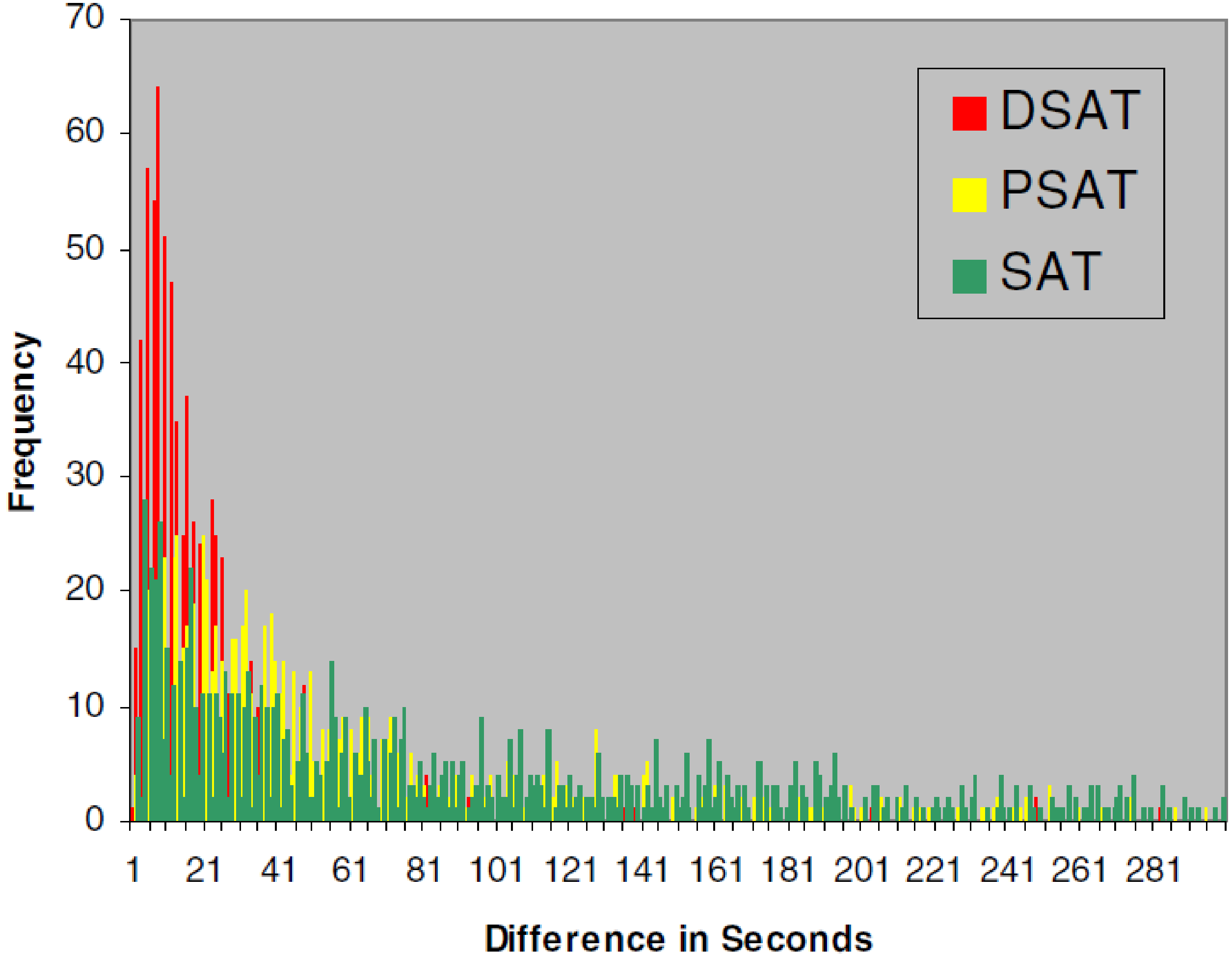


Table 5: Result-Level Predictions using Bayesian model

Levels	SAT	PSAT	DSATs	Accuracy
Predict SAT	172	53	20	70%
Predict PSAT	67	91	36	47%
Predict DSAT	39	86	134	52%

Implicit Measures for evaluation

- Dwell time is positively correlated with SAT.



Implicit Measures for evaluation

- Time to first click for SAT prediction [Hassan et al., CIKM2011]

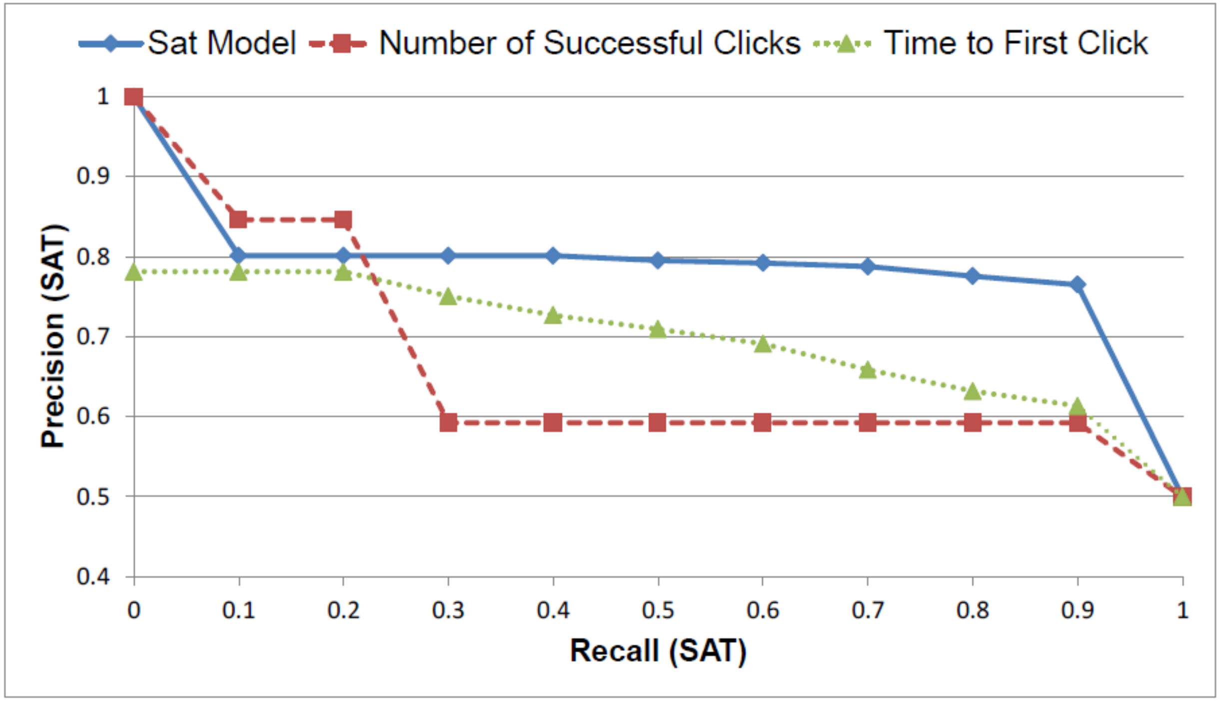


Figure 1: Precision Recall Graph for the SAT Class.

Implicit Measures for evaluation

- Time to first click for DSAT prediction [Hassan et al., CIKM2011]

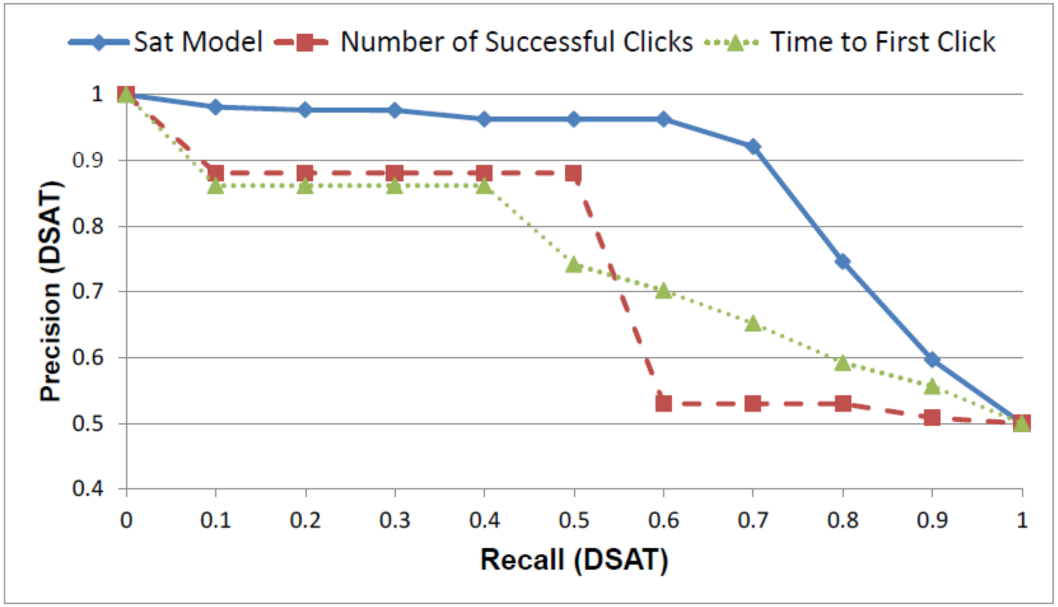


Figure 2: Precision Recall Graph for the DSAT Class.

References

- Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, Jaime Teevan: Modeling and analysis of cross-session search tasks. SIGIR 2011: 5-14
- Sarah K. Tyler, Jaime Teevan: Large scale query log analysis of re-finding. WSDM 2010: 191-200
- Jaime Teevan: How people recall, recognize, and reuse search results. ACM Trans. Inf. Syst. 26(4) (2008)
- Jaime Teevan, Eytan Adar, Rosie Jones, Michael A. S. Potts: Information re-retrieval: repeat queries in Yahoo's logs. SIGIR 2007: 151-158
- Jaime Teevan, Eytan Adar, Rosie Jones, Michael A. S. Potts: History repeats itself: repeat queries in Yahoo's logs. SIGIR 2006: 703-704
- Jaime Teevan: How people recall search result lists. CHI Extended Abstracts 2006: 1415-1420
- Zhicheng Dou, Ruihua Song, Ji-Rong Wen: A large-scale evaluation and analysis of personalized search strategies. WWW 2007: 581-590
- Jaime Teevan, Daniel J. Liebling, Gayathri Ravichandran Geetha: Understanding and predicting personal navigation. WSDM 2011: 85-94
- Matthew Richardson: Learning about the world through long-term query logs. TWEB 2(4) (2008)
- Yu Wang, Xiao Huang, Ryen White, Characterizing and Supporting Cross-Device Search Tasks, WSDM 2013
- Amanda Spink, Minsoo Park, Bernard J. Jansen, Jan O. Pedersen: Multitasking during Web search sessions. Inf. Process. Manage. 42(1): 264-275 (2006)

References

- Anne Aula, Rehan M. Khan, Zhiwei Guan: How does search behavior change as search becomes more difficult? CHI 2010: 35-44
- Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan T. Dumais, Thomas White: Evaluating implicit measures to improve web search. ACM Trans. Inf. Syst. 23(2): 147-168 (2005)
- Ahmed Hassan, Yang Song, Li-wei He: A task level metric for measuring web search satisfaction and its application on improving relevance estimation. CIKM 2011: 125-134
- Zhen Liao, Yang Song, Li-wei He, Yalou Huang: Evaluating the effectiveness of search task trails. WWW 2012: 489-498
- Ryen W. White, Susan T. Dumais: Characterizing and predicting search engine switching behavior. CIKM 2009: 87-96
- Ryen W. White, Jeff Huang: Assessing the scenic route: measuring the value of search trails in web logs. SIGIR 2010: 587-594
- Thorsten Joachims, Laura A. Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, Geri Gay: Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. ACM Trans. Inf. Syst. 25(2) (2007)
- Fan Guo, Chao Liu, Anitha Kannan, Tom Minka, Michael J. Taylor, Yi Min Wang, Christos Faloutsos: Click chain model in web search. WWW 2009: 11-20
- Yuchen Zhang, Weizhu Chen, Dong Wang, Qiang Yang: User-click modeling for understanding and predicting search-behavior. KDD 2011: 1388-1396
- Georges Dupret, Benjamin Piwowarski: A user behavior model for average precision and its generalization to graded judgments. SIGIR 2010: 531-538
- Zeyuan Allen Zhu, Weizhu Chen, Tom Minka, Chenguang Zhu, Zheng Chen: A novel click model and its applications to online advertising. WSDM 2010: 321-330
- Olivier Chapelle, Ya Zhang: A dynamic bayesian network click model for web search ranking. WWW 2009: 1-10

References

- Anagha Kulkarni, Jaime Teevan, Krysta Marie Svore, Susan T. Dumais: Understanding temporal query dynamics. WSDM 2011: 167-176
- Steven M. Beitzel, Eric C. Jensen, Abdur Chowdhury, Ophir Frieder, David A. Grossman: Temporal analysis of a very large topically categorized Web query log. JASIST 58(2): 166-178 (2007)
- Steven M. Beitzel, Eric C. Jensen, Abdur Chowdhury, David A. Grossman, Ophir Frieder: Hourly analysis of a very large topically categorized web query log. SIGIR 2004: 321-328
- Kira Radinsky, Krysta Marie Svore, Susan T. Dumais, Jaime Teevan, Alex Bocharov, Eric Horvitz: Modeling and predicting behavioral dynamics on the web. WWW 2012: 599-608
- Fabrizio Silvestri, Mining Query Logs: Turning Search Usage Data into Knowledge, Foundations and Trends in Information Retrieval, v.4 n.1—2, p.1-174, January 2010
- Michail Vlachos, Philip S. Yu, Vittorio Castelli, Christopher Meek: Structural Periodic Measures for Time-Series Data. Data Min. Knowl. Discov. 12(1): 1-28 (2006)
- Michail Vlachos, Christopher Meek, Zografoula Vagena, Dimitrios Gunopulos: Identifying Similarities, Periodicities and Bursts for Online Search Queries. SIGMOD Conference 2004: 131-142
- Fernando Diaz: Integration of news content into web results. WSDM 2009: 182-191
- Arnd Christian König, Michael Gamon, Qiang Wu: Click-through prediction for news queries. SIGIR 2009: 347-354
- Yoshiyuki Inagaki, Narayanan Sadagopan, Georges Dupret, Anlei Dong, Ciya Liao, Yi Chang, Zhaohui Zheng: Session Based Click Features for Recency Ranking. AAAI 2010
- Milad Shokouhi: Detecting seasonal queries by time-series analysis. SIGIR 2011: 1171-1172
- Jon M. Kleinberg: Bursty and hierarchical structure in streams. KDD 2002: 91-101
- Nish Parikh, Neel Sundaresan: Scalable and near real-time burst detection from eCommerce queries. KDD 2008: 972-980
- Gabriel Pui Cheong Fung, Jeffrey Xu Yu, Philip S. Yu, Hongjun Lu: Parameter Free Bursty Events Detection in Text Streams. VLDB 2005: 181-192
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L, Detecting influenza epidemics using search engine query data, Nature 457, 1012-1014 (19 February 2009)
- Steve Chien, Nicole Immorlica: Semantic similarity between search engine queries using temporal correlation. WWW 2005: 2-11
- Silviu Cucerzan and Eric Brill, Extracting Semantically Related Queries by Exploiting User Session Information Unpublished Draft (submitted to WWW-2006, November 2005)
- Kira Radinsky, Eugene Agichtein, Evgeniy Gabrilovich, Shaul Markovitch: A word at a time: computing word relatedness using temporal semantic analysis. WWW 2011: 337-346
- Qiankun Zhao, Steven C. H. Hoi, Tie-Yan Liu, Sourav S. Bhowmick, Michael R. Lyu, Wei-Ying Ma: Time-dependent semantic similarity measure of queries using historical click-through data. WWW 2006: 543-552

Spatio-temporal and Socio-temporal Trends

WSDM 2013 Tutorial

Schedule

- Introduction (9:00-9:15)
- Modeling Dynamics
  - 9:15-10:15 Web content dynamics [Susan]
  - 10:15-10:45 Web user behavior dynamics [Milad]
  - 10:45-11:00 Break
  - 11:00-11:30 Web user behavior dynamics, cont'd
  - 11:30-13:00 Spatio-temporal analysis [Fernando]
  - Methods for evaluation
- Lunch (13:00-14:30)
- Applications to Information Retrieval
  - 14:30-15:45 Temporal NLP [Kira]
  - News event prediction
  - 15:45-16:00 Break
  - 16:00-17:45 Time-sensitive search [Yi]
  - Time-sensitive recommendations [Anlei]
- Wrap-Up (17:45-18:00)

Multidimensional Dynamics

Information Exist in Context

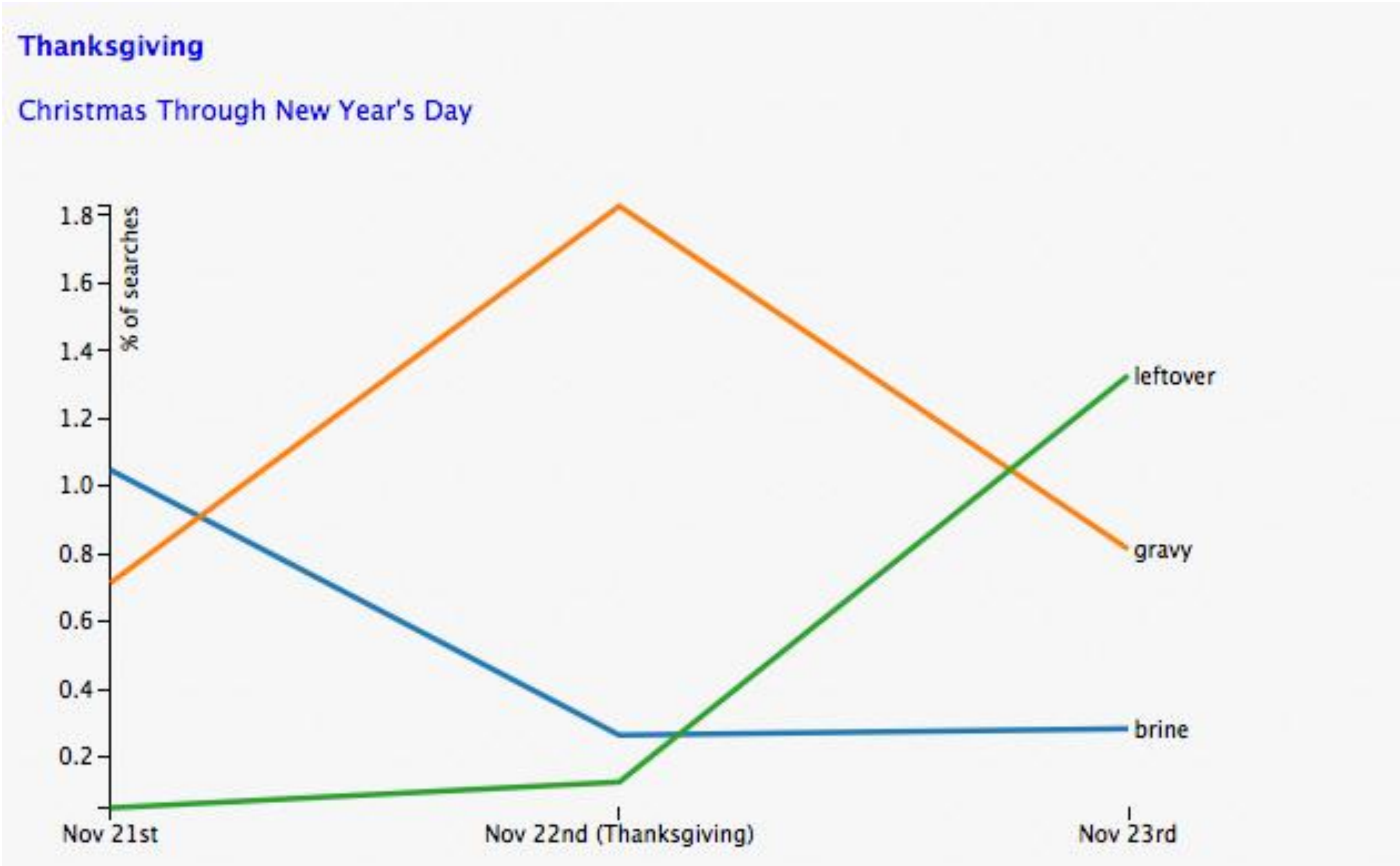
- temporal
  - does the document refer to a specific time?
  - does the information need refer to a specific time?
- geographic
  - does the document refer to a specific location?
  - does the information need refer to a specific location?
- social
  - does the document refer to a specific group of people?
  - does the information need refer to a specific group of people?
- many, many others

Multidimensional Modeling

- Spatiotemporal
  - appropriate when we suspect both temporal and geographic salience.
- Sociotemporal
  - appropriate when we suspect both temporal and social salience.

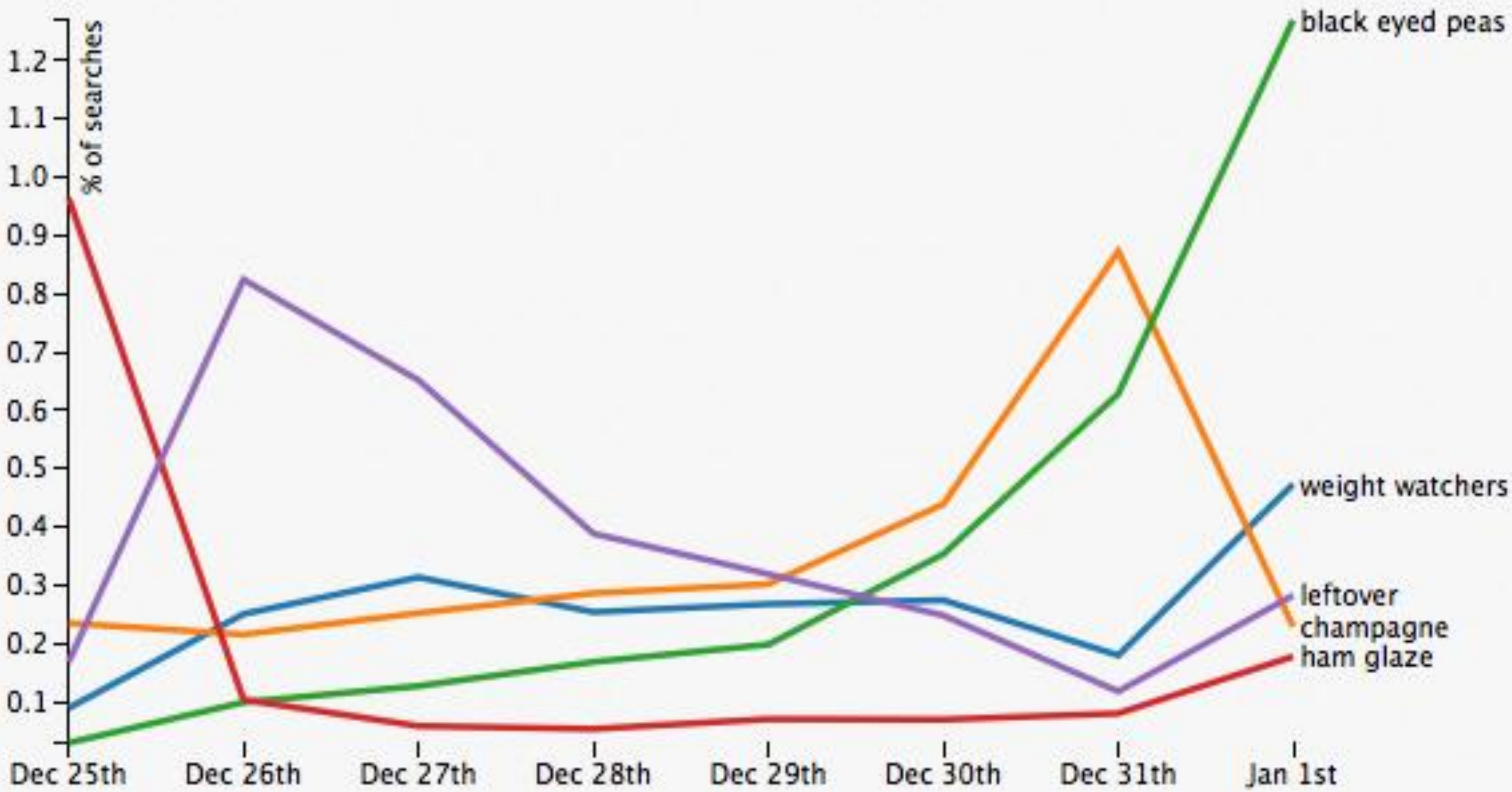
Spatiotemporal Modeling

- Goal: study the ability to capture spatial and temporal aspects for topics.
- Approach: study the ability to capture spatial and temporal aspects for **spatiotemporally acute events**.
  - simplifies the task to topics likely to exhibit capturable behavior.
  - many spatiotemporally acute events receive **a lot** of query and document volume (e.g. natural disasters).

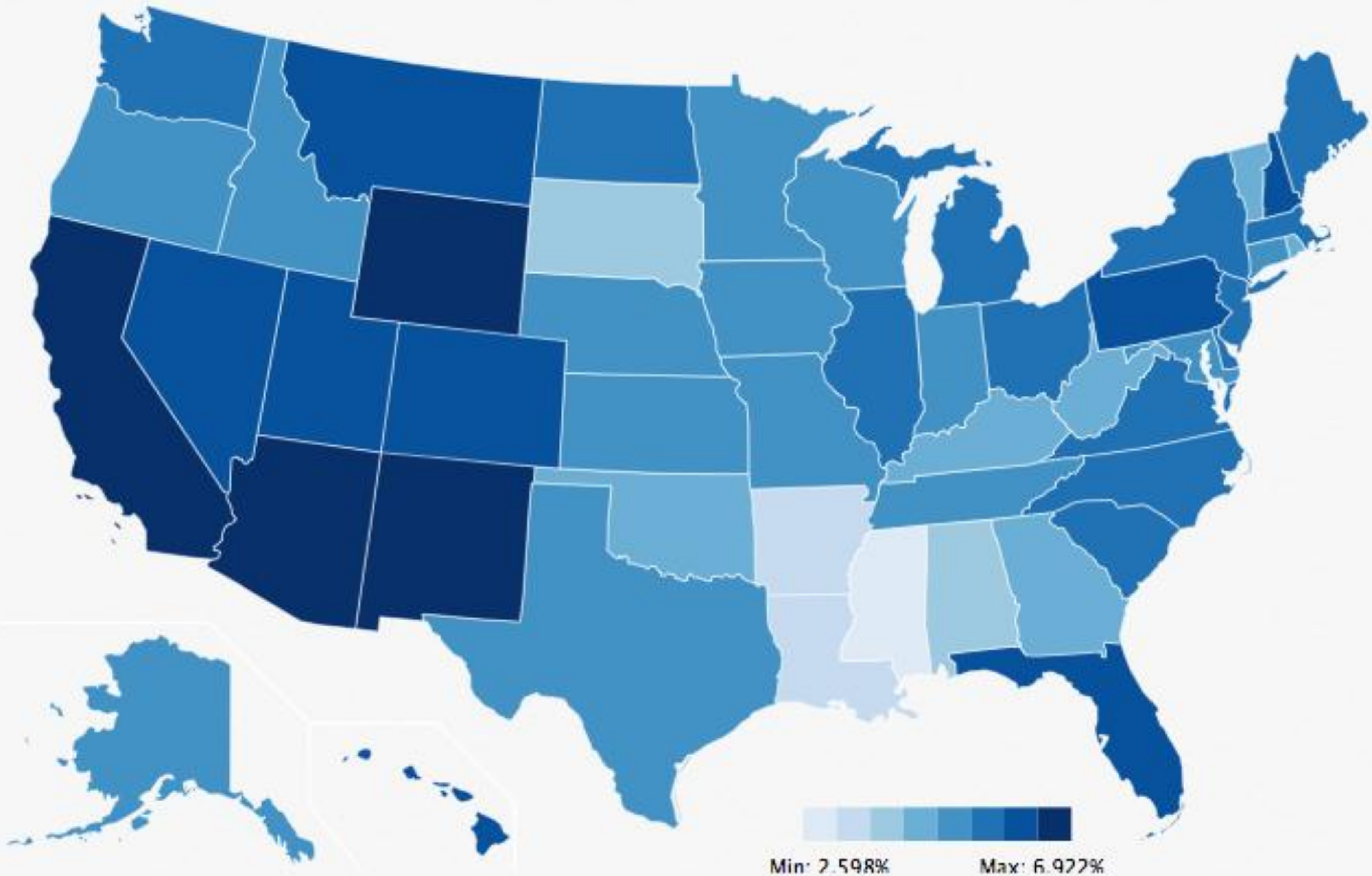

Thanksgiving
Christmas Through New Year's Day
% of searches
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
Nov 21st
Nov 22nd (Thanksgiving)
Nov 23rd
leftover
gravy
brine

Thanksgiving

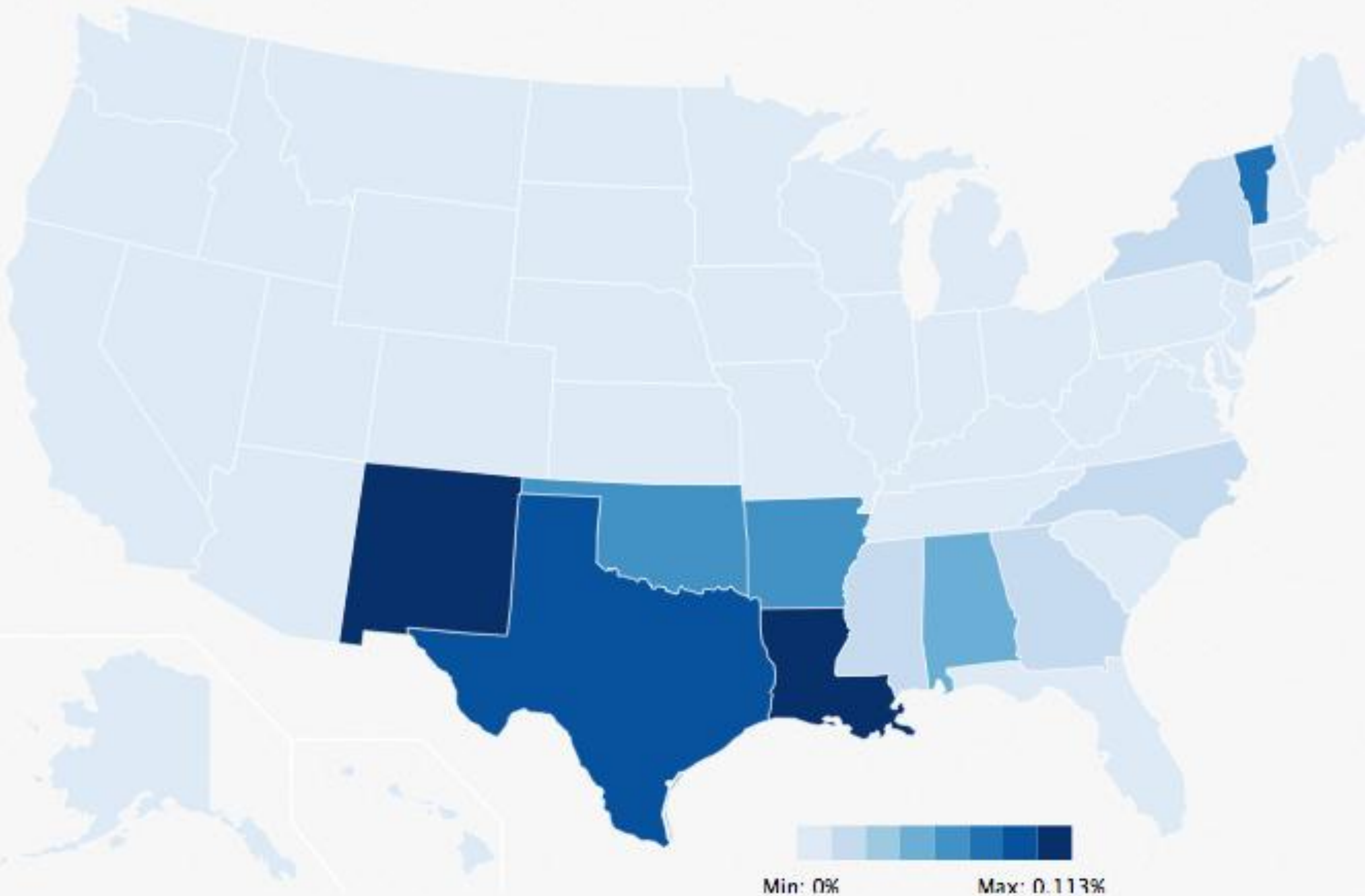
Christmas Through New Year's Day



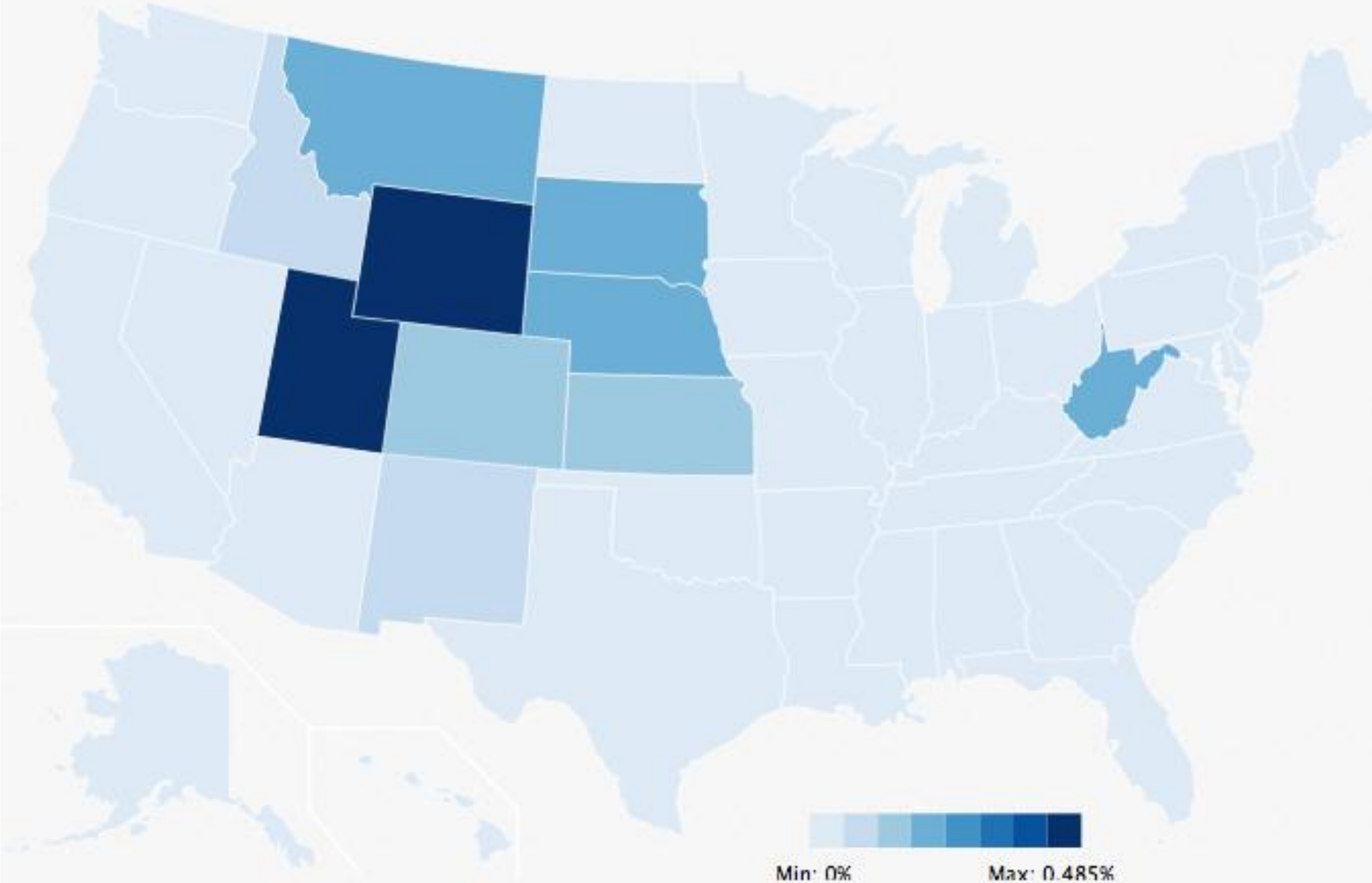
Thanksgiving: Turkey



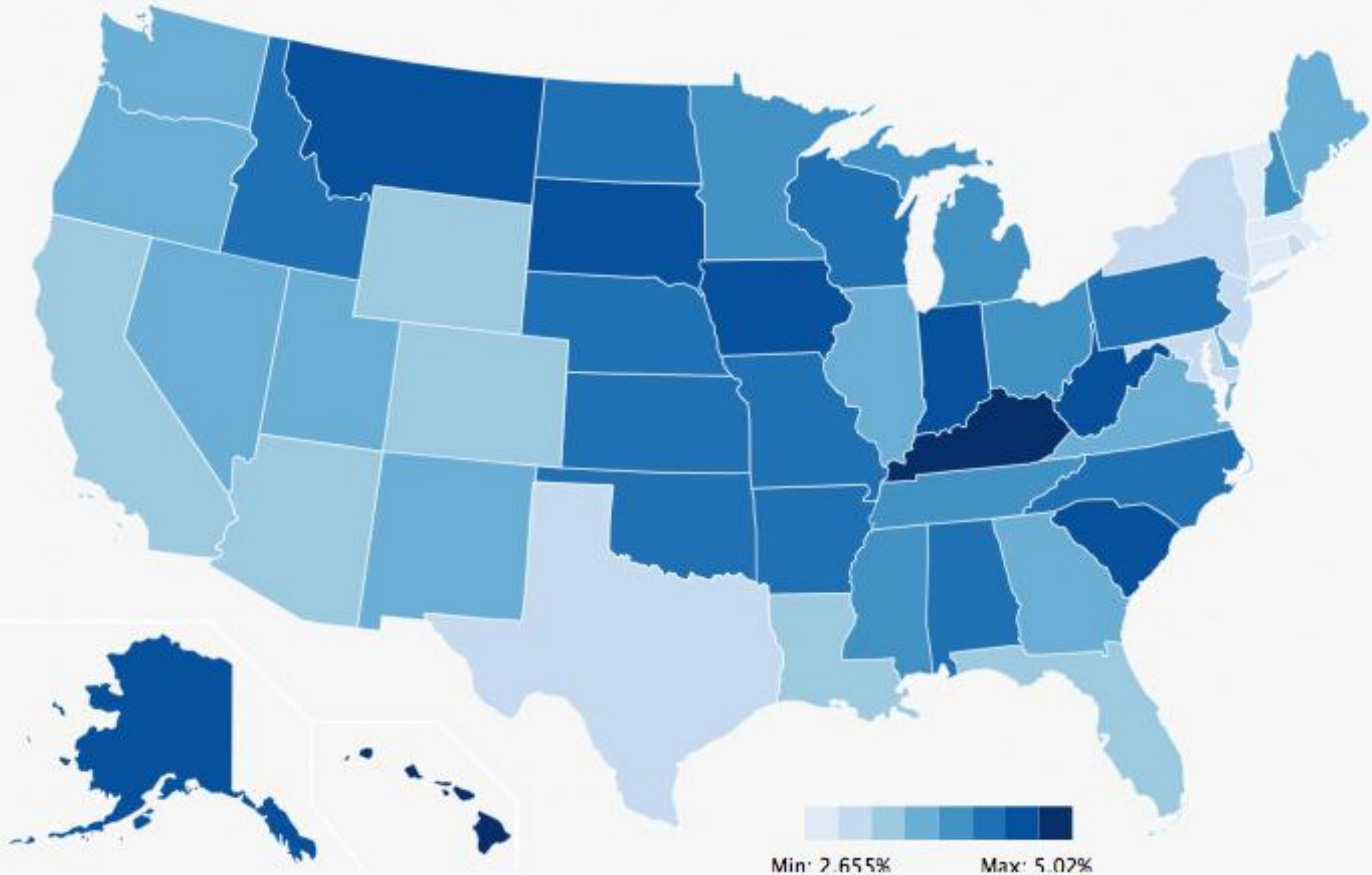
Thanksgiving: Millionaire Pie



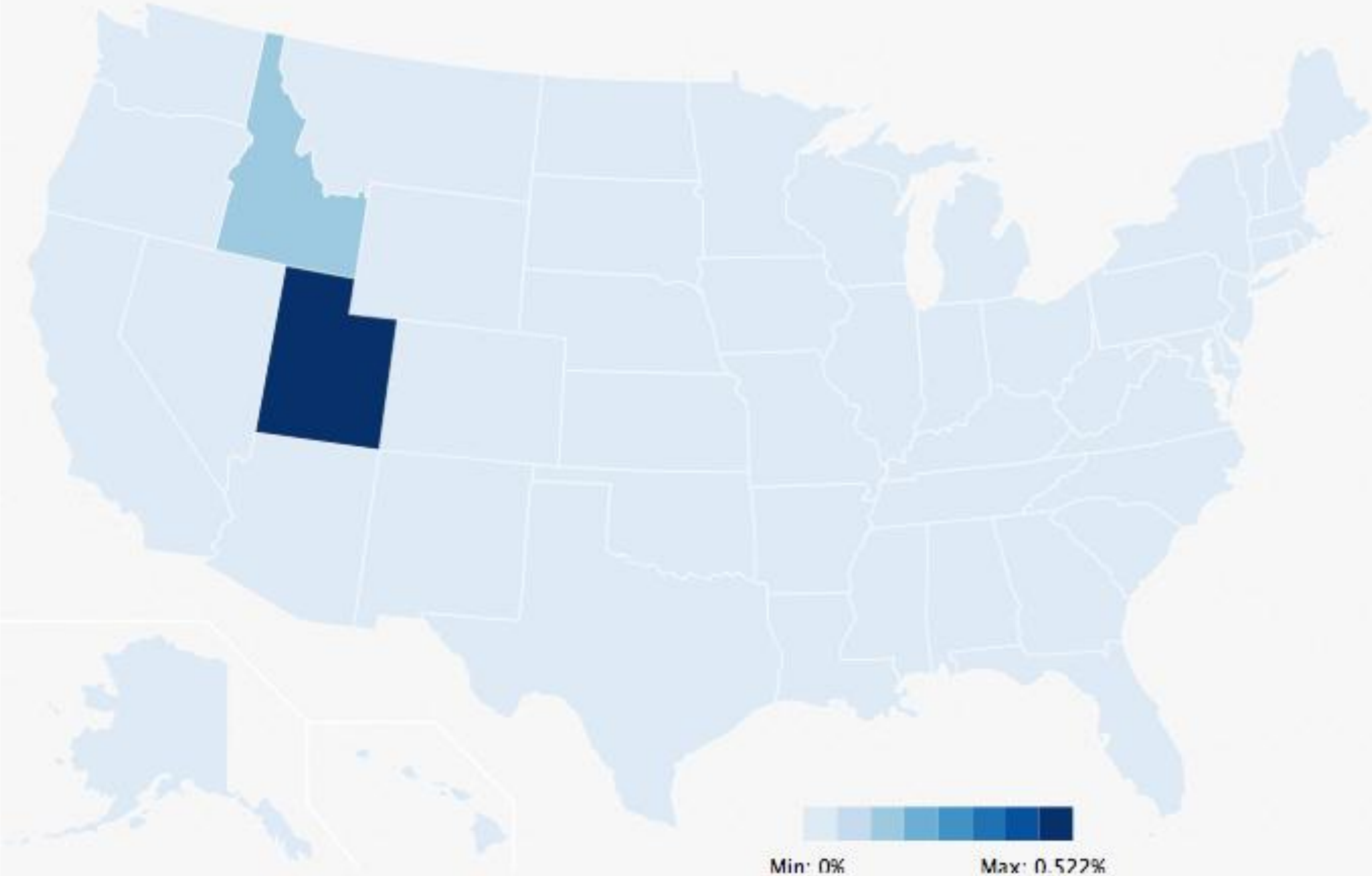
Thanksgiving: Frog Eye Salad



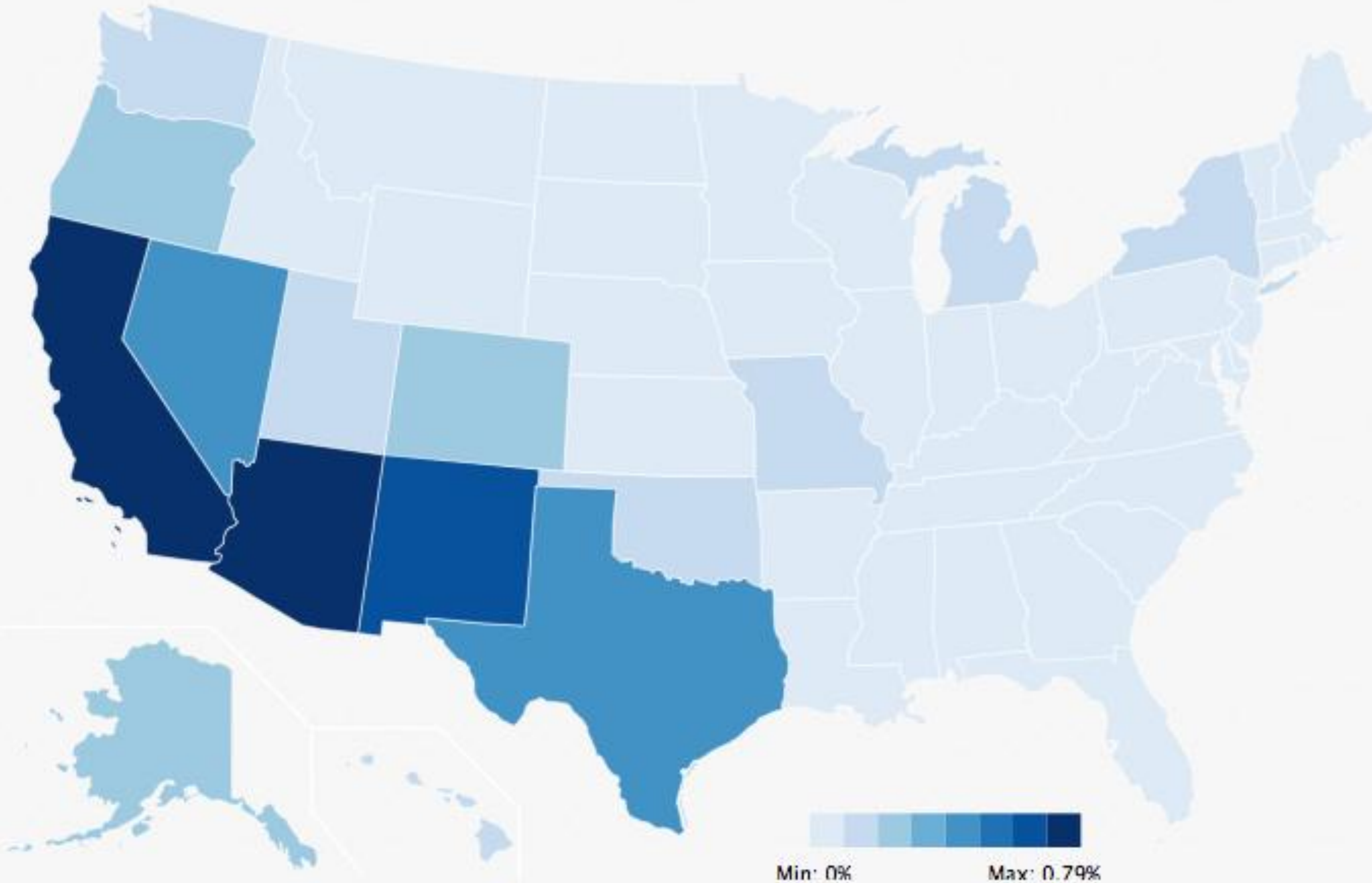
Christmas: Ham



Christmas: Funeral Potatoes



Christmas: Tamale



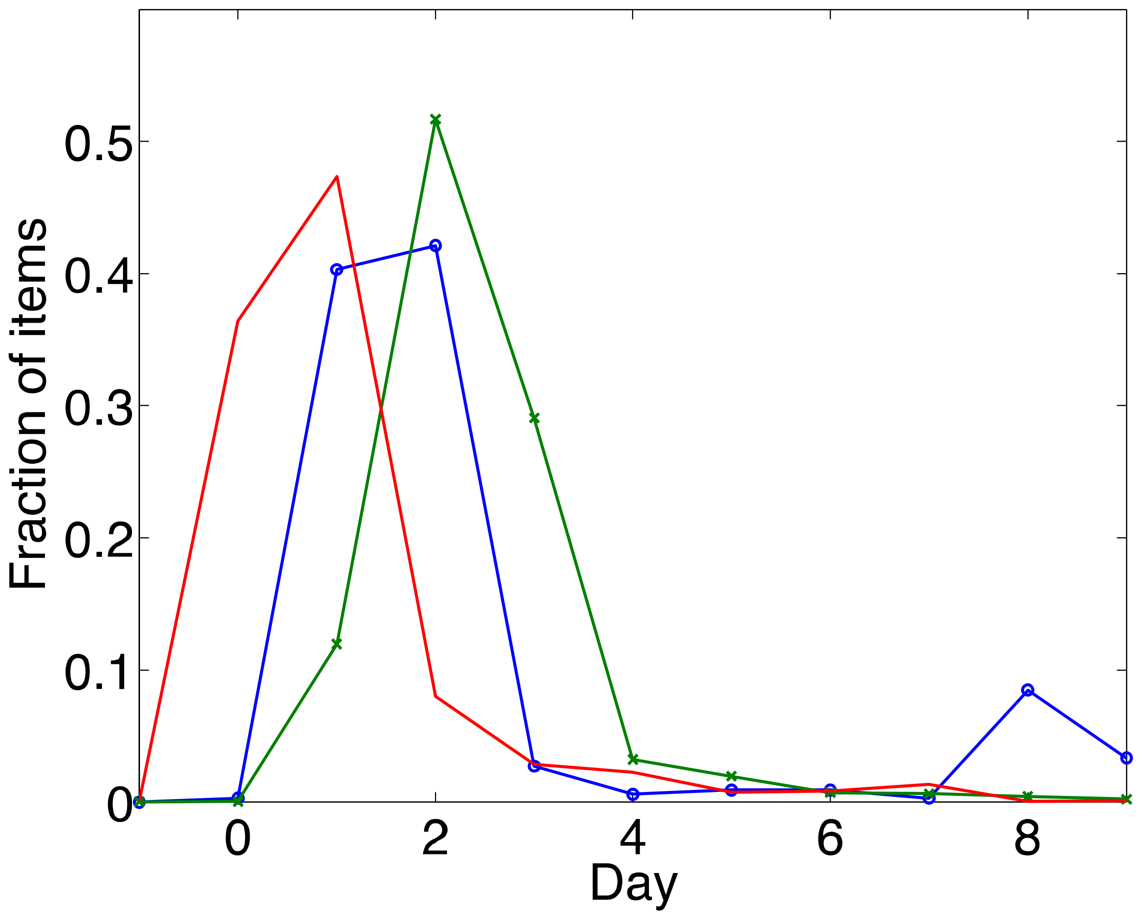
Spatiotemporal Modeling Case Studies

- **News**: exploit text-based *production* to model topics over space and time.
- **Queries**: exploit text-based *demand* to model topics over space and time.
- **Images**: exploit image metadata *production* to model topics over space and time.

Why experiment with news?

- News articles often focus on temporally acute events.
  - natural disaster updates
  - political coverage
- News corpora are easy to deal with
  - availability (e.g. online, LDC)
  - standardized (e.g. LDC corpora, Reuters)
  - clean, journalistic language
  - reliable timestamps

Temporal Sensitivity of News Interest



news (blue), social media (red), and query volume (green) for 2010 New York tornad

[Yom-Tov and Diaz 2011

Geographic Sensitivity of News Interest

Event	Queries	News	Twitter
San Bruno	-0.97	-0.97	-0.97
New York	-0.86	-0.95	-0.90
Alaska	-0.69	-0.91	-0.52

Spearman correlation between physical distance and the fraction of media items and relevant queries, for each of the three events. All correlations are statistically significant at $p < 0.05$.

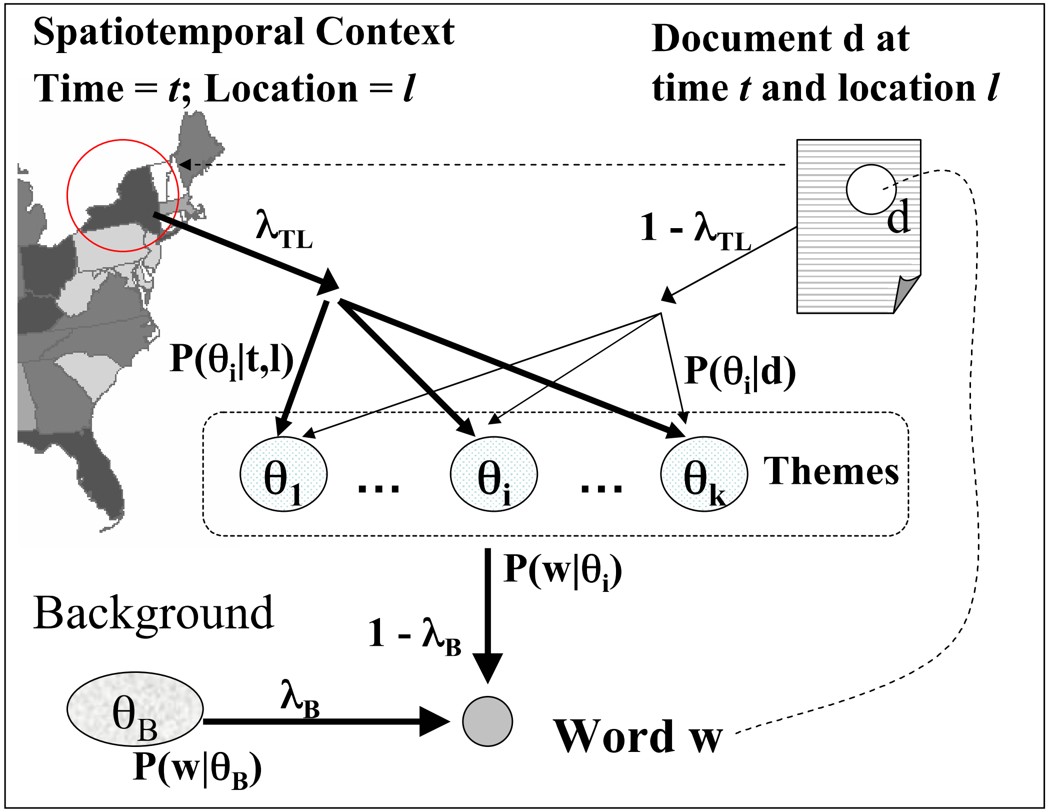
[Yom-Tov and Diaz 2011

Modeling Spatiotemporal News

- Assume that words in an article are sampled from two underlying distributions,
  - a background language model: represents word usage common across time and geography (e.g. determiners, pronouns).
  - a spatiotemporal theme model: represents word usage specific to a time and place (i.e. an event).

[Mei *et al.* 2006]

Modeling Spatiotemporal News



[Mei *et al.* 2006]

Modeling Spatiotemporal News

$$p(w|d,t,l) = \lambda_0 \underbrace{p(w|\theta_B)}_{\text{background}} + \lambda_1 \underbrace{\sum_{j=1}^{k} p(w|\theta_j)p(\theta_j|d)}_{\text{document text}} + \lambda_2 \underbrace{\sum_{j=1}^{k} p(w|\theta_j)p(\theta_j|t,l)}_{\text{document time and location}}$$

$p(w\|\theta_j)$	spatiotemporal language model
$p(\theta_j\|d)$	probability of language model given document text
$p(\theta_j\|t,l)$	probability of language model given document time and location
λ_i	mixing weights

notation modified for clarity.

[Mei *et al.* 2006]

Modeling Spatiotemporal News

- Model parameters
  - background model: maximum likelihood estimate from corpus.

$$p(w|\theta_B) = \frac{\sum_{d \in C} c(w,d)}{\sum_{w \in V} \sum_{d \in C} c(w,d)}$$

  - …similar for document model.
  - theme model: estimated my expectation maximization.

[Mei *et al.* 2006]

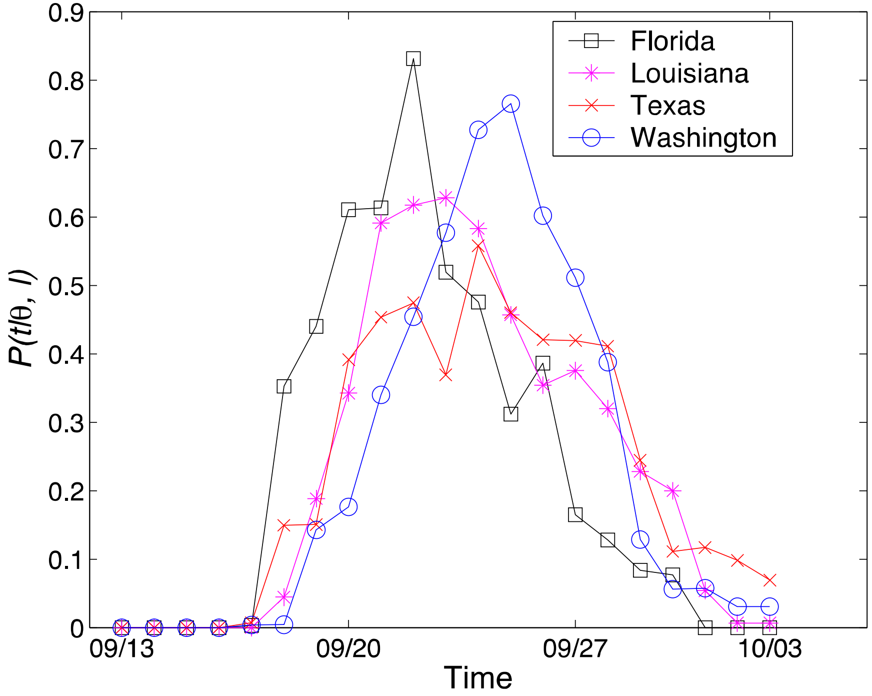
Modeling Spatiotemporal News

Theme 1	Theme 2	Theme 3	Theme 4	Theme 5	Theme 6
Government Response	**New Orleans**	**Oil Price**	**Praying and Blessing**	**Aid and Donation**	**Personal Life**
bush 0.0716374	city 0.0633899	price 0.0772064	god 0.141807	donate 0.120228	i 0.405526
president 0.0610942	orleans 0.0540974	oil 0.0643189	pray 0.047029	relief 0.0769788	my 0.11688
federal 0.0514114	new 0.034188	gas 0.0453731	prayer 0.0417175	red 0.0702266	me 0.0601333
govern 0.0476977	louisiana 0.0234546	increase 0.0209058	love 0.0307544	cross 0.0651472	am 0.0291511
fema 0.0474692	flood 0.0227215	product 0.0202912	life 0.025797	help 0.0507348	think 0.0150206
administrate 0.0233903	evacuate 0.0211225	fuel 0.0188067	bless 0.025475	victim 0.0360877	feel 0.0123928
response 0.0208351	storm 0.01771328	company 0.0181833	lord 0.0177097	organize 0.0220194	know 0.0114889
brown 0.0199573	resident 0.0168828	energy 0.0179985	jesus 0.0162096	effort 0.0207279	something 0.00774544
blame 0.0170033	center 0.0165427	market 0.0167884	will 0.0139161	fund 0.0195033	guess 0.00748368
governor 0.0142153	rescue 0.0128347	gasoline 0.0123526	faith 0.0120621	volunteer 0.0194967	myself 0.00687533

themes extracted from blog posts about Hurricane Katrina.

[Mei *et al.* 2006]

Modeling Spatiotemporal News



"storm" theme broken down by state.

[Mei *et al.* 2006]

News Open Questions

- **Task**: how can this information be used for information access tasks?
- **Granularity:** how can we model small scale/”tail” events underrepresented in the national news?

Why experiment with queries?

- Queries **sometimes** focus on temporally acute events.
  - natural disaster queries
- Temporally acute queries are important
  - information need is **urgent**
  - high-visibility failure

Modeling Spatiotemporal Queries

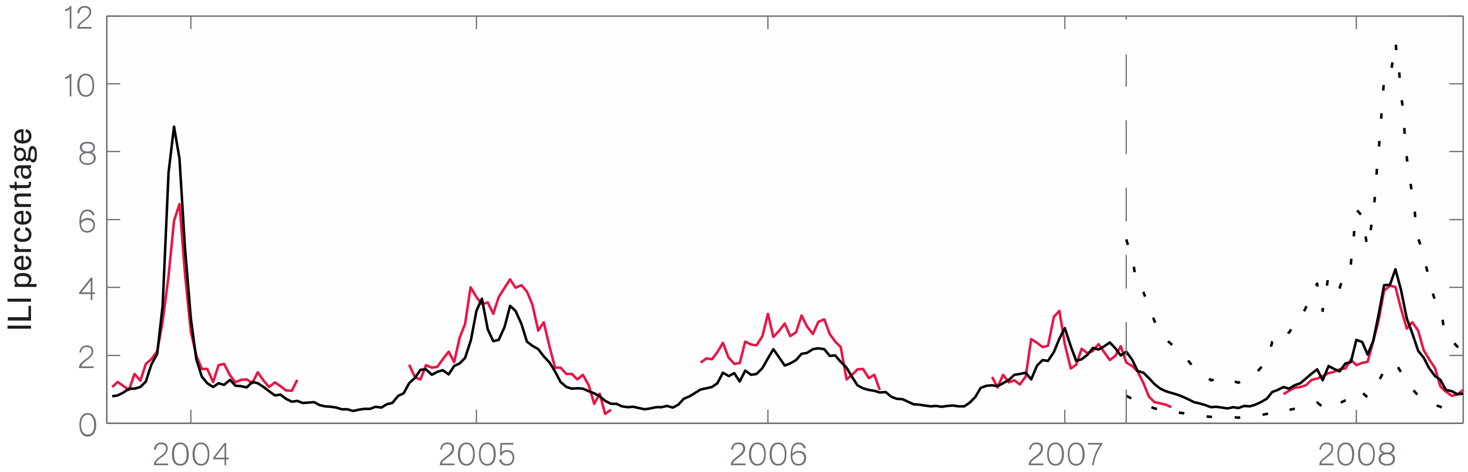
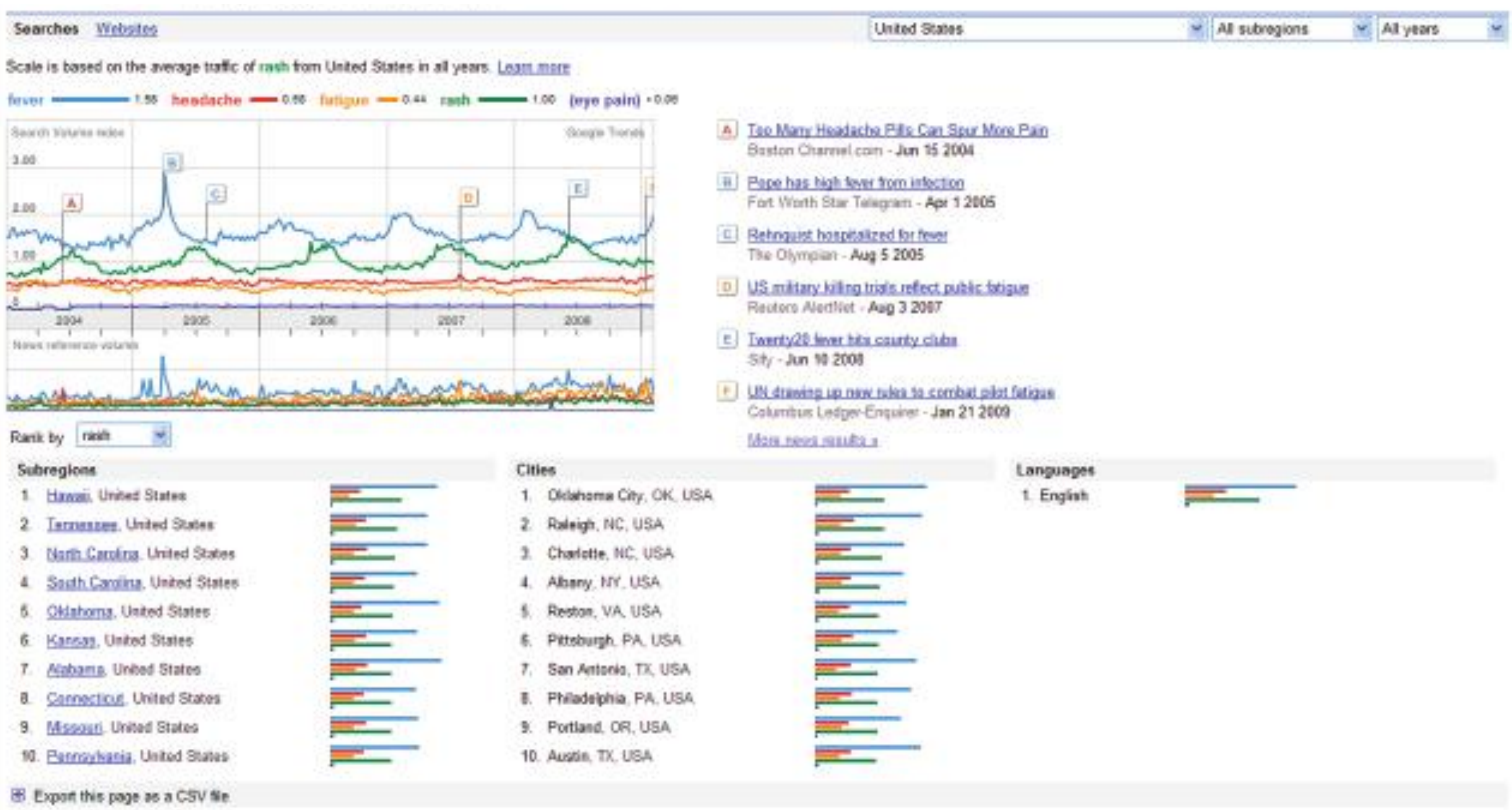


Figure 2: A comparison of model estimates for the Mid-Atlantic Region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated. A correlation of 0.85 was obtained over 128 points from this region to which the model was fit, while a correlation of 0.96 was obtained over 42 validation points. 95% prediction intervals are indicated.

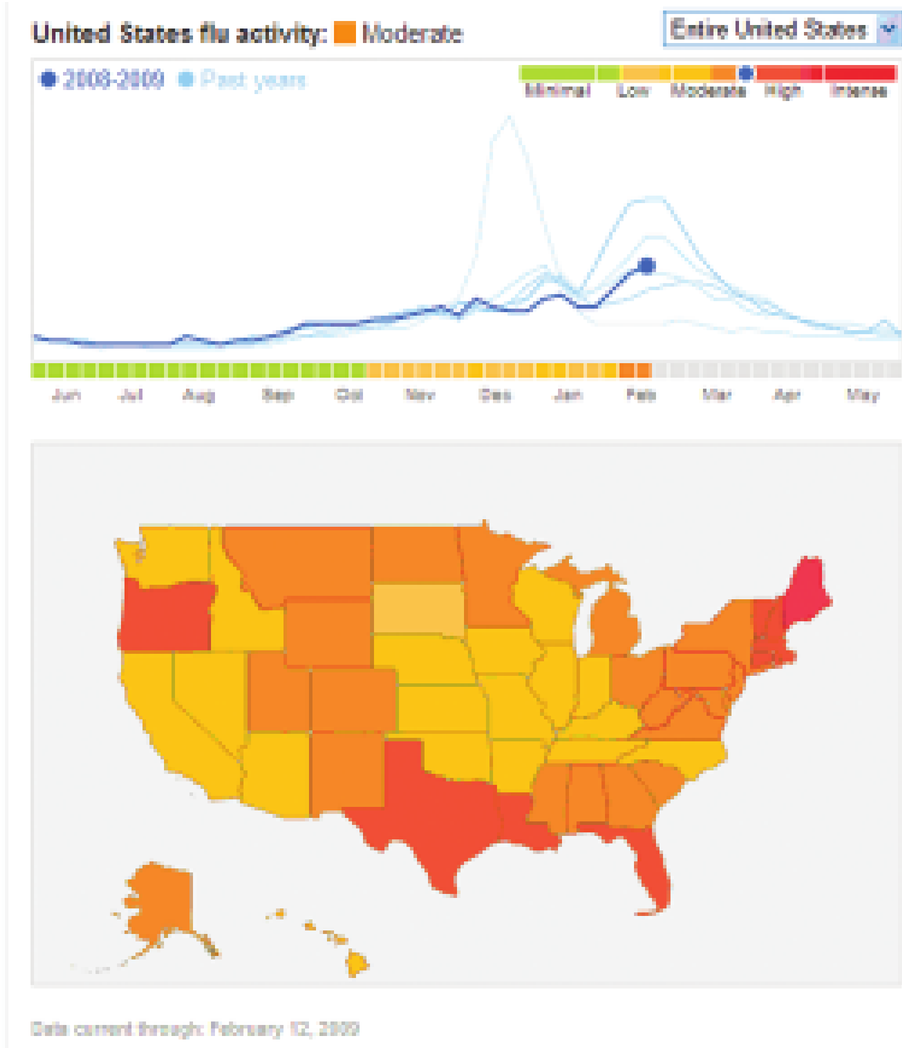
[Ginsberg *et al.* 2009]

Modeling Spatiotemporal Queries



[Carneiro and Mylonakis 2009]

Modeling Spatiotemporal Queries



[Carneiro and Mylonakis 2009]

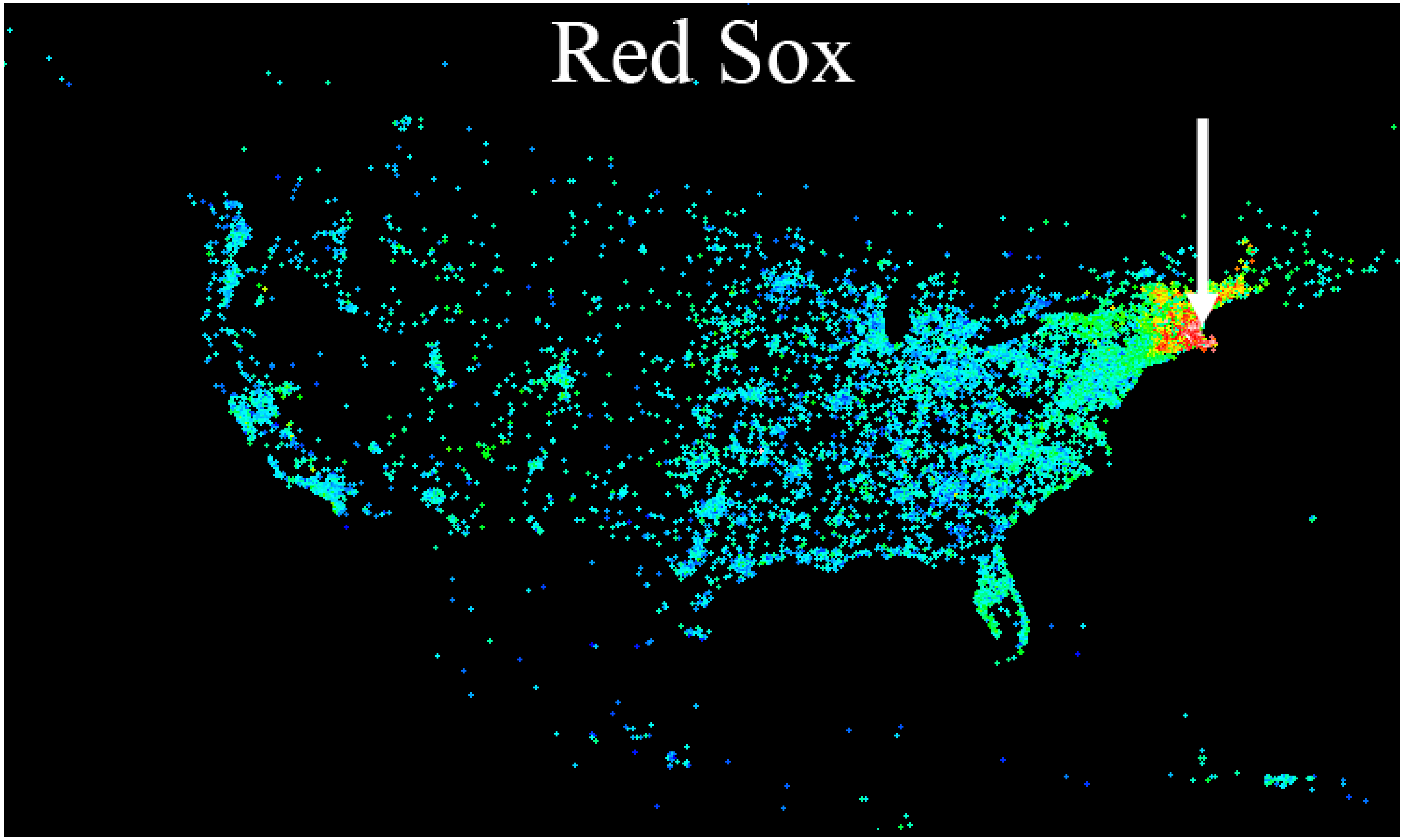
Modeling Spatiotemporal Queries

$$p(q|l) = Cd^{-\alpha}$$

l	user location
d	distance between user location and query center
C	height of peak at query center
α	decay from query center

[Backstrom *et al.* 2009]

Modeling Spatiotemporal Queries

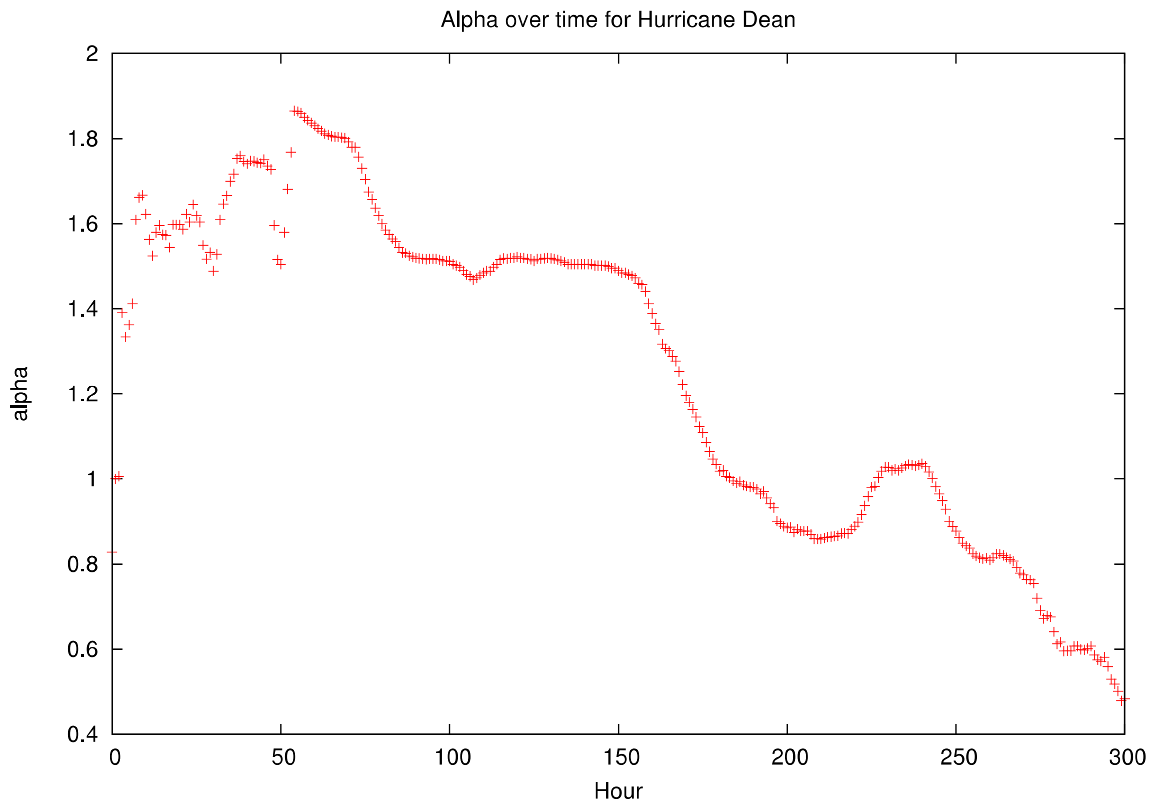


[Backstrom *et al.* 2009]

Modeling Spatiotemporal Queries

[Backstrom *et al.* 2009]

Modeling Spatiotemporal Queries



$$p(q|l) = Cd^{-\alpha}$$

[Backstrom *et al.* 2009]

News Open Questions

- **Task**: how can this information be used for information access tasks?
- **Granularity:** how can we model small scale/”tail” events underrepresented in query logs?
- **More dimensions:** what other dimensions can be incorporated from query logs?

Why experiment with images?

- Photographs are taken at a specific time and place, often with keyword tags.
- Photograph corpora are easy to deal with
  - photographs exist in volume (people like to take pictures)
  - photographs have precise spatiotemporal data
  - photographs are manually tagged (“the food is bad but the portions are large”)

Modeling Spatiotemporal Images

Problem definition:

can time and place semantics for a tag be derived from the tag's location and time usage distribution?

[Rattenbury et al. 2007]

Modeling Spatiotemporal Images

- Short tags can often be attributed to the photo **place** or **event**.
- **place tag**: expected to exhibit significant spatial patterns.
- **event tag**: expected to exhibit significant temporal patterns.
- “significant pattern” refers to a burst of activity in space or time.

#wsdm2013 #rome

[Rattenbury et al. 2007]

Subtasks

1. **scale specification:** at what granularity should we look for patterns?
   - **time:** seconds? minutes? days?
   - **space:** neighborhood? city? state?
2. **segment specification:** how do we partition the dimension for analysis?
   - **time:** uniform segments? volume-weighted? consider diurnal patterns?
   - **space:** uniform grid? political boundaries (e.g. urban, state)?

[Rattenbury et al. 2007]

Subtasks

3. **significance testing:** is the behavior in this segment different from behavior outside of the segment?
   - **time:** compare to before and after? previous day? week? month? year?
   - **space:** compare to all surrounding? similar city?
4. **determine event scale:** how do we aggregate granular results to larger scales?
   - unsmoothed estimate?
   - repeat process for multiple scales?

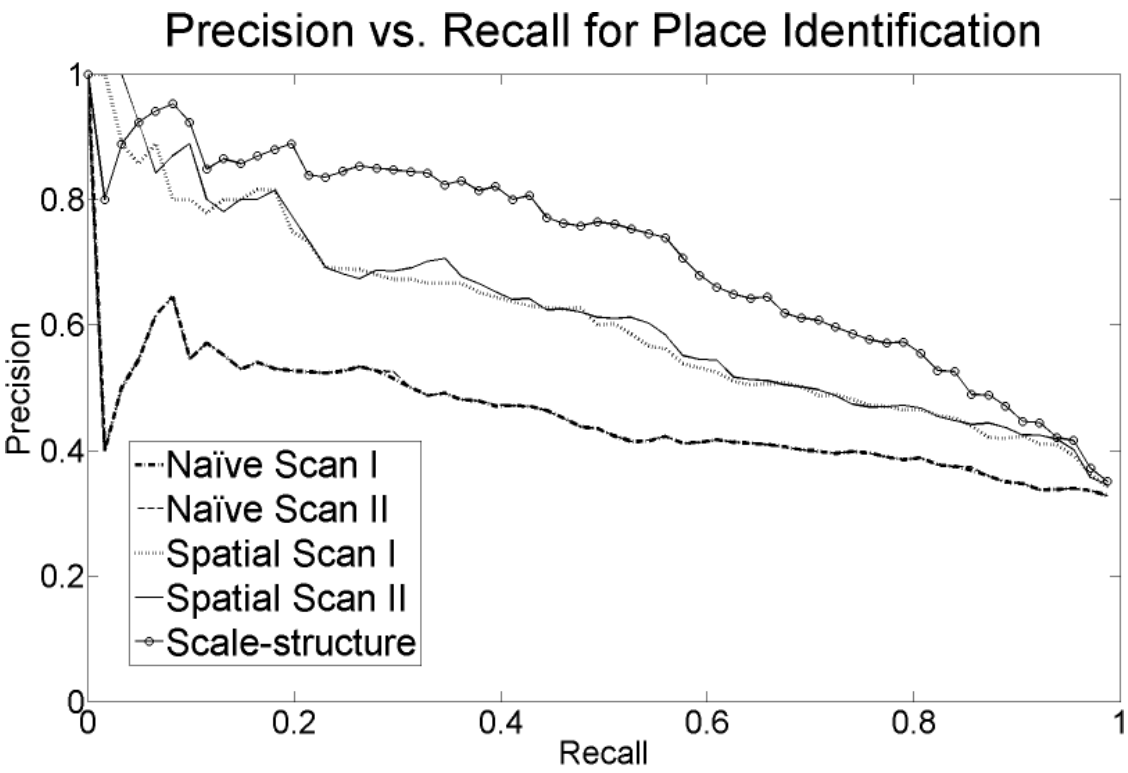
[Rattenbury et al. 2007]

Experiments

- public photograph datasets (e.g. Flickr) often include rich space and time metadata.
- manually judge the events and locations referred to by tags.
- predict whether a tag refers to an event or location, compute precision and recall of labels in ranked list of tags.

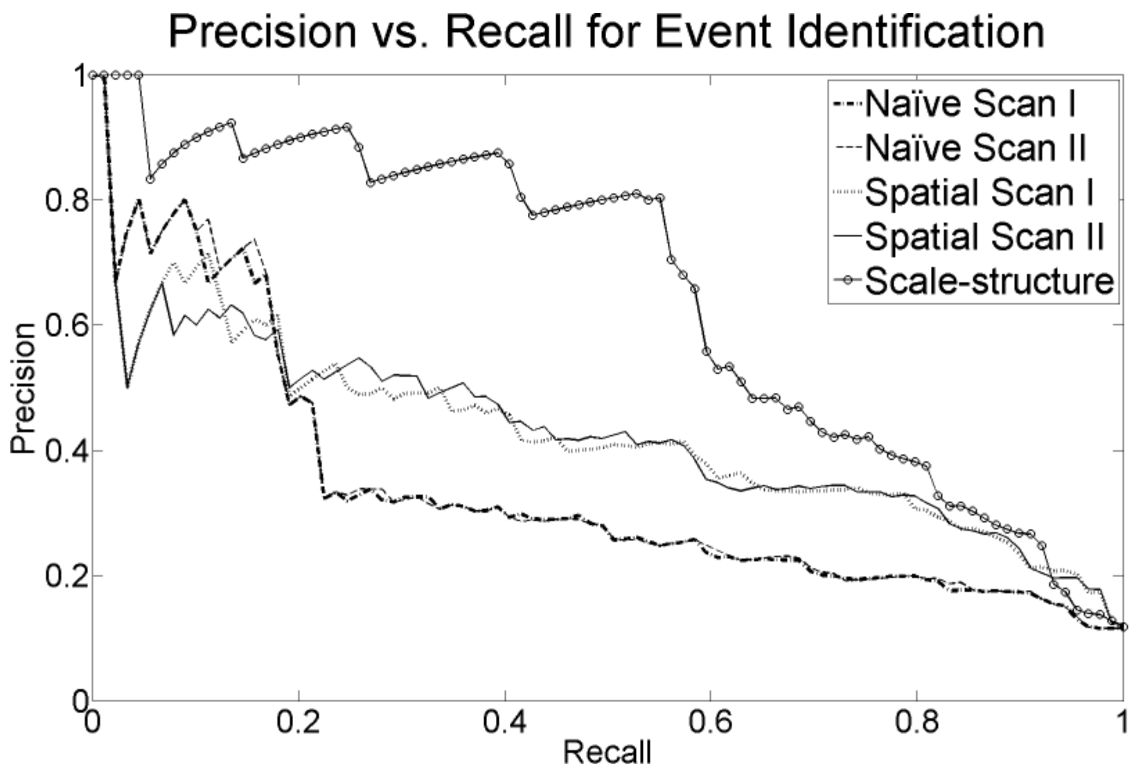
[Rattenbury et al. 2007]

Modeling Spatiotemporal Images



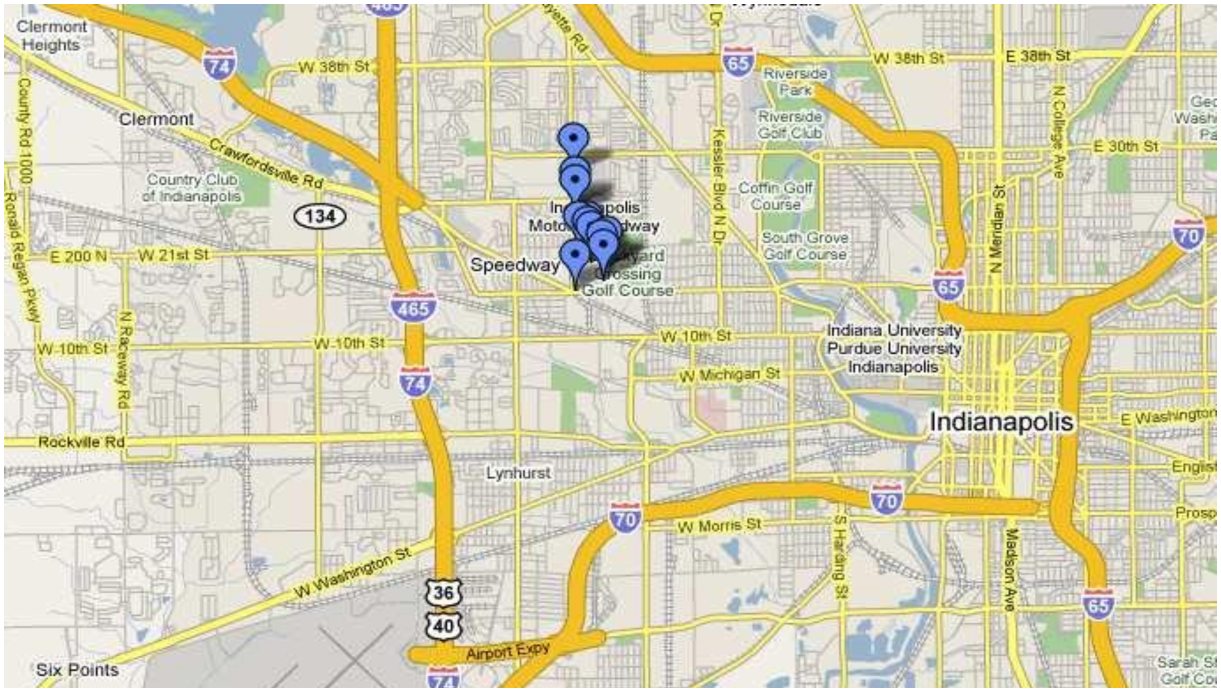
[Rattenbury et al. 2007]

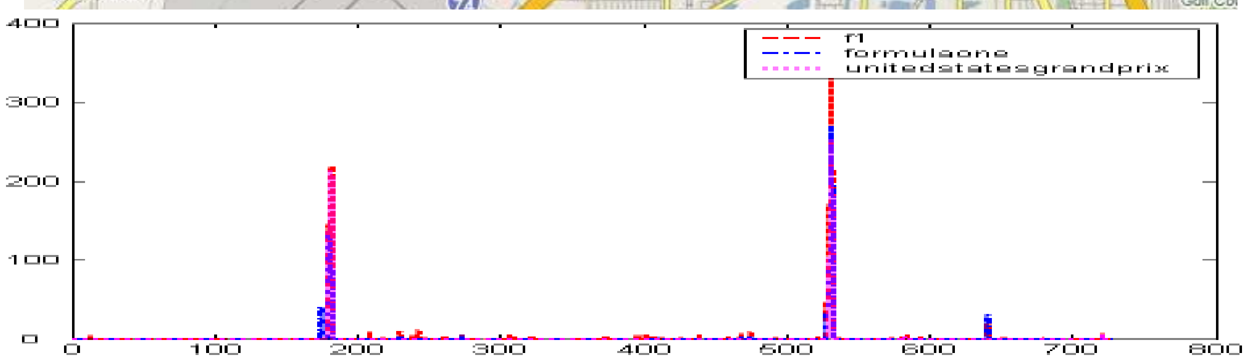
Modeling Spatiotemporal Images



[Rattenbury et al. 2007]

Detecting Periodic Events in Spatiotemporal Images





images with tags,

- f1
- formulaone
- unitedstatesgrandprix

[Chen and Roy 2009]

Detecting Periodic Events in Spatiotemporal Images

Periodic Event Tags	housewarming, bigbear, skysinger, indigogirls, deathvalleynationalpark, legionofdoom, westtexas, californiaadventure, ames, samantha, dealsgap, **grandam**, bymiketravis, detourart, adamhubenig, chincoteague, **nights**, paragliding, leavenworth, thebigapple
Aperiodic Event Tags	bourbonstreet, nueva, theindigogirls, portage, mountdesertisland, tueam, threatdottv, shores, sams, ska, sebastian, boone, dnalounge, greatscott, worldinferno, dawnanddrew, delraybeach, doorcounty, ig, southpadreisland

Table 2: Top 20 event tags detected by SI, where tags in bold are true positives. Tag grandam refers to the car racing event. Tag nights is related to the event of Hollywood nights.

[Chen and Roy 2009]

Image Open Questions

- **Task**: how can this information be used for information access tasks?
- **Granularity:** how can we model small scale/”tail” events underrepresented in images?
- **More dimensions:** what other dimensions can be incorporated from images?

Sociotemporal Modeling

- Goal: study the ability to capture social and temporal aspects for topics.
- Approach: study the ability to capture spatial and temporal aspects for **sociotemporally acute events**.
  - often includes spatiotemporally acute events (news—especially if unexpected—attracts attention)
  - also includes completely virtual events (e.g. `memes’)

Sociotemporal Modeling Case Studies

- **Video Sharing**: users often watch and promote videos over social networks (e.g. email, instant messaging, microblogs).
- **Information Seeking During Disaster:** users often query for information about a disaster if social contacts are affected.

Types of Sociotemporal Topics

- **Exogenous Critical:** topic is propagated throughout the social network by an external stimulus (e.g. earthquake).
- **Endogenous Critical:** topic is propagated throughout the social network without external stimulus (e.g. lolcats).
- **Exogenous Subcritical:** topic does not spread despite external stimulus (e.g. car accident).
- **Endogenous Subcritical:** topic does not spread and is not externally stimulated.

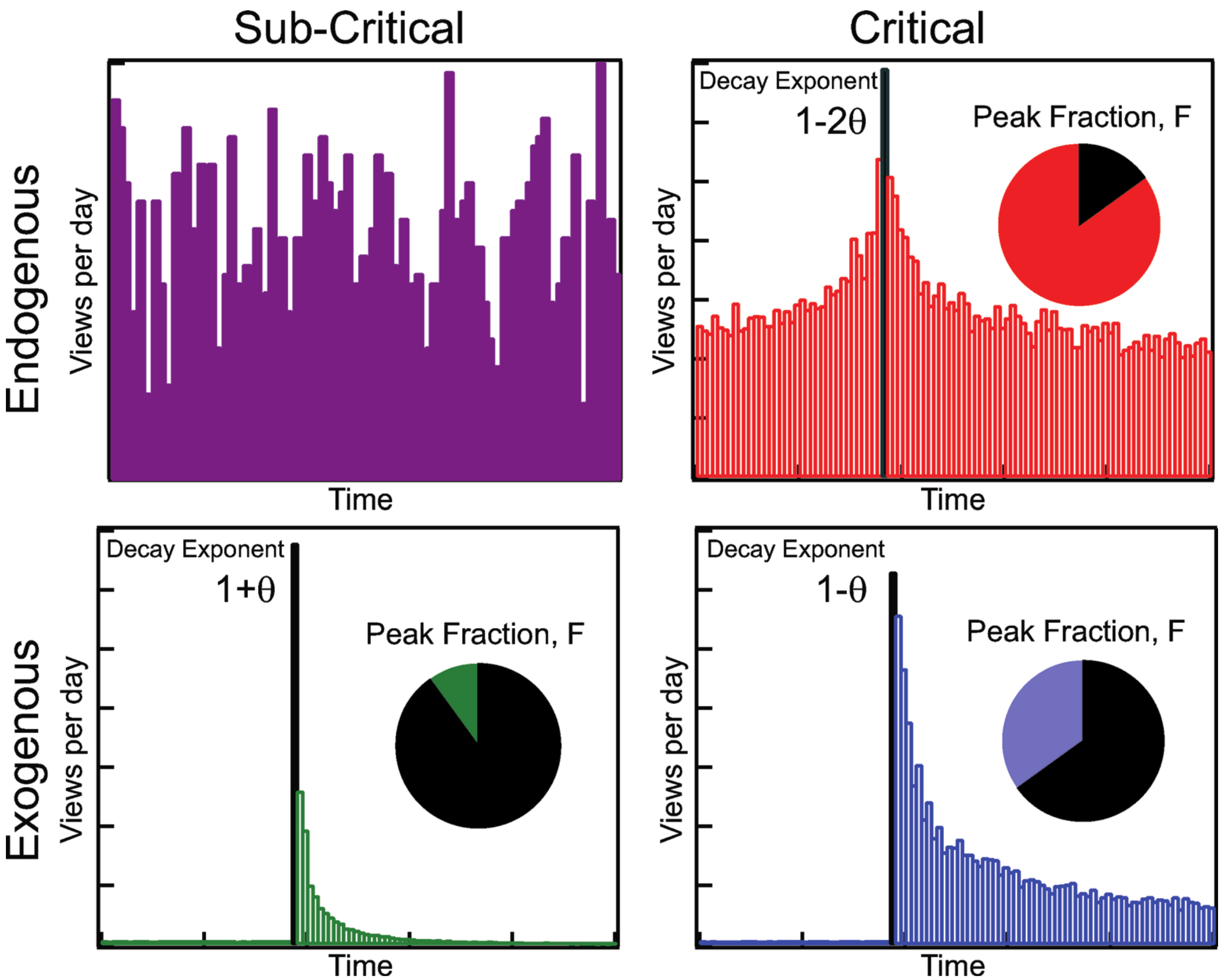
[Crane and Sornette 2008]

Sociotemporal Dynamics of Video Sharing

- **Corpus:** time stamped view information from a video-sharing site.
- **Research Question:** does the viewing information suggest an underlying epidemic model?

[Crane and Sornette 2008]

Types of Sociotemporal Behavior



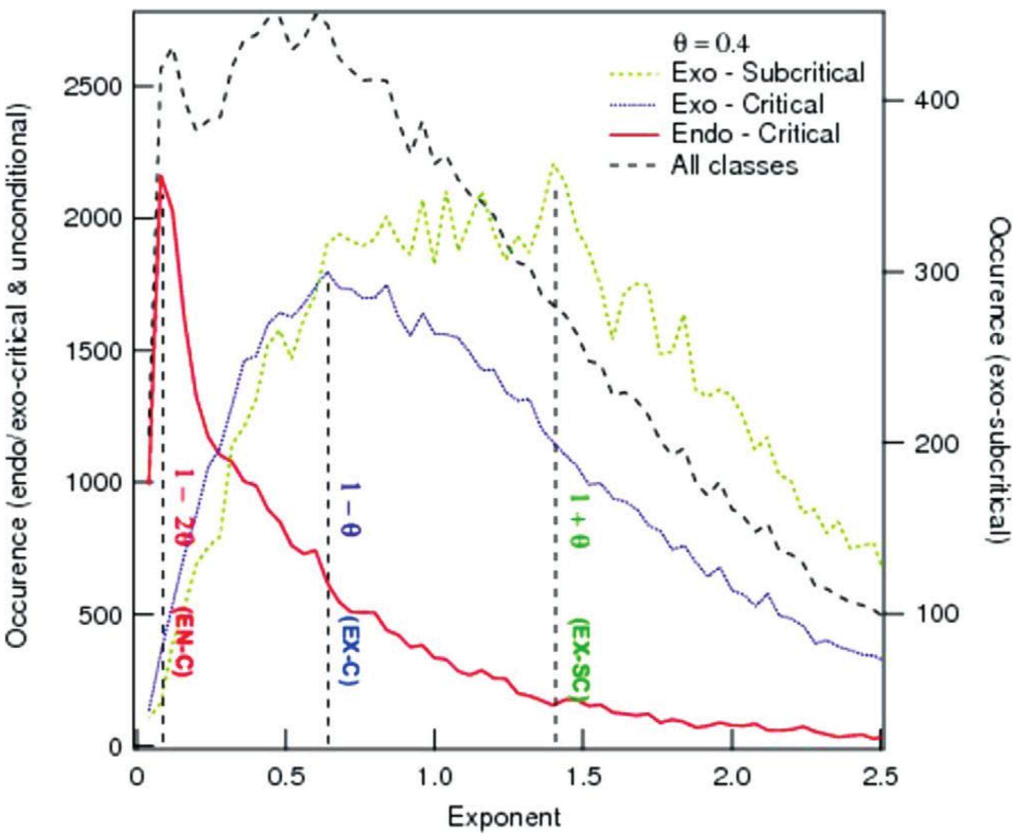
[Crane and Sornette 2008]

Types of Sociotemporal Behavior

	subcritical	crictical
endogenous	$n(t)$	$\|t - t_c\|^{1-2\theta}$
exogenous	$(t - t_c)^{-(1+\theta)}$	$(t - t_c)^{\theta-1}$

[Crane and Sornette 2008]

Types of Sociotemporal Behavior



[Crane and Sornette 2008]

Sociotemporal Dynamics of Video Sharing

- Evidence supports hypothesis of an epidemic process.
- No explicit signals of epidemic processes.

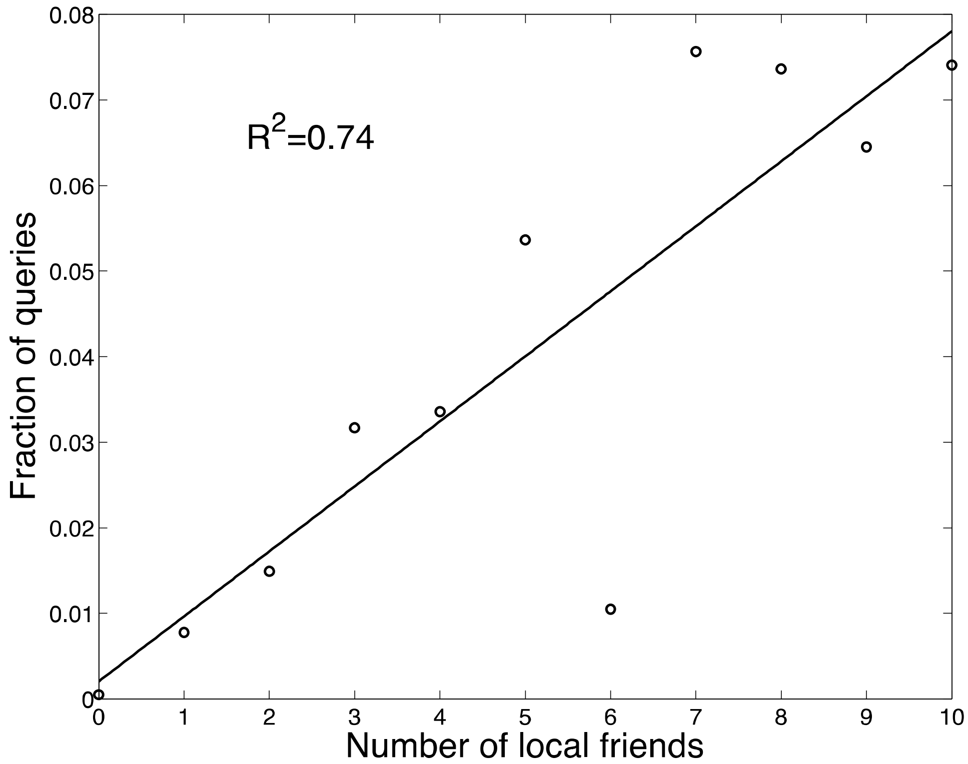
[Crane and Sornette 2008]

Information Seeking During Crisis

- **Hypothesis:** users with friends in areas affected by a crisis event are more likely to seek information about that event than those with no friends in those areas.
- **Test:** Does personalizing ranking by local connections improve retrieval?

[Yom-Tov and Diaz 2010]

Crisis Interest and Social Connections



[Yom-Tov and Diaz 2010]

Social Contacts and Relevance During Crisis

	MAP				P@10			
	Both	Physical	Social	No	Both	Physical	Social	No
San Bruno	0.693	0.692	0.716	0.607	0.234	0.234	0.236	0.228
New York	0.804	0.794	0.788	0.764	0.148	0.148	0.148	0.145
Alaska	0.889	0.886	0.898	0.831	0.167	0.167	0.168	0.167

[Yom-Tov and Diaz 2010]

Multidimensional Modeling Open Questions

- Formal Models
  - no general model capturing spatial, social, and temporal data.
- Tasks
  - need to develop/understand tasks for which multidimensional modeling is important.
- Corpora
  - need to develop standard corpora for sociotemporal modeling.

Methods for Evaluation

Time-Sensitive Tasks

- Web Search
- Topic Detection and Tracking (TDT)
- TREC 2011-2013 Microblog Track
- TREC 2013 Temporal Summarization Track

Web Search

- **Task:** Given a query, provide a ranked list of documents satisfying the user’s information need.
- **Approach:** Collect relevance judgments and evaluate with a judgment-based metric

Normalized Discounted Cumulative Gain (NDCG)

$$\mathrm{NDCG}_n^{\mathbf{y}} = \frac{1}{\mathcal{Z}_n} \sum_{i=1}^{n} \frac{2^{y_i} - 1}{\log_2(i+1)}$$

$$\begin{array}{rl} \mathbf{y} & \text{vector of document gains} \\ \mathbf{n} & \text{rank cutoff} \\ \mathcal{Z}_{\mathbf{n}} & \text{normalizer} \end{array}$$

[Jarvelin and Kekalainen 2002]

Web Search

- **Task:** Given a query, provide a ranked list of documents satisfying the user's information need.
- **Approach:** Collect relevance judgments and evaluate with a judgment-based metric
- **Problem:** For time-sensitive information needs, satisfaction may include more than topical relevance.
- **Solution 1:** Introduce independent, time-sensitive judgments.

Time-Sensitive Gains

$\mathbf{y}^{\mathrm{R}}$	relevance [Jarvelin and Kekalainen 2002]
$\mathbf{y}^{\mathrm{F}}$	freshness [Dong *et al.* 2010; Dai *et al.* 2011]
$\mathbf{y}^{\mathrm{S}}$	staleness [Dong *et al.* 2010]
$\mathbf{y}^{\gamma} = \gamma\mathbf{y}^{\mathrm{R}} + (1-\gamma)\mathbf{y}^{\mathrm{F}}$	combined [Dai *et al.* 2011]
$\mathbf{y}^{d} = \mathbf{y}^{\mathrm{R}} - \mathbf{y}^{\mathrm{S}}$	demotion [Dong *et al.* 2010]

Web Search

- **Task:** Given a query, provide a ranked list of documents satisfying the user’s information need.
- **Approach:** Collect relevance judgments and evaluate with a judgment-based metric
- **Problem:** For time-sensitive information needs, satisfaction may include more than topical relevance.
- **Solution 2:** Rely on implicit behavior (e.g. user clicks) to capture combined target.

[Wang *et al.* 2012]

Open Questions

- **Query sampling:** how to select queries likely to have temporal intent?
- **Judge quality:** how to select topics which are still in the judges "memory"?

Topic Detection and Tracking

- **Topic Tracking:** Keep track of stories similar to a set of example stories.
- **Topic Detection:** Build clusters of stories that discuss the same topic.
- **First Story Detection:** Detect if a story is the first story of a new, unknown topic.

[Allan 2002]

Detection-Error Tradeoff Evaluation

		Reference Annotation	
		Target	Non-Target
System Response	YES (a Target)	Correct	*False Alarm*
	NO (Not a Target)	*Missed Detection*	Correct

[Allan 2002]

Detection Cost

$$C_{\text{det}} = C_{\text{miss}} P_{\text{miss}} P_{\text{target}} + C_{\text{FA}} P_{\text{FA}} (1 - P_{\text{target}})$$

$$P_{\text{miss}} = \frac{\#\text{miss}}{\#\text{target}}$$

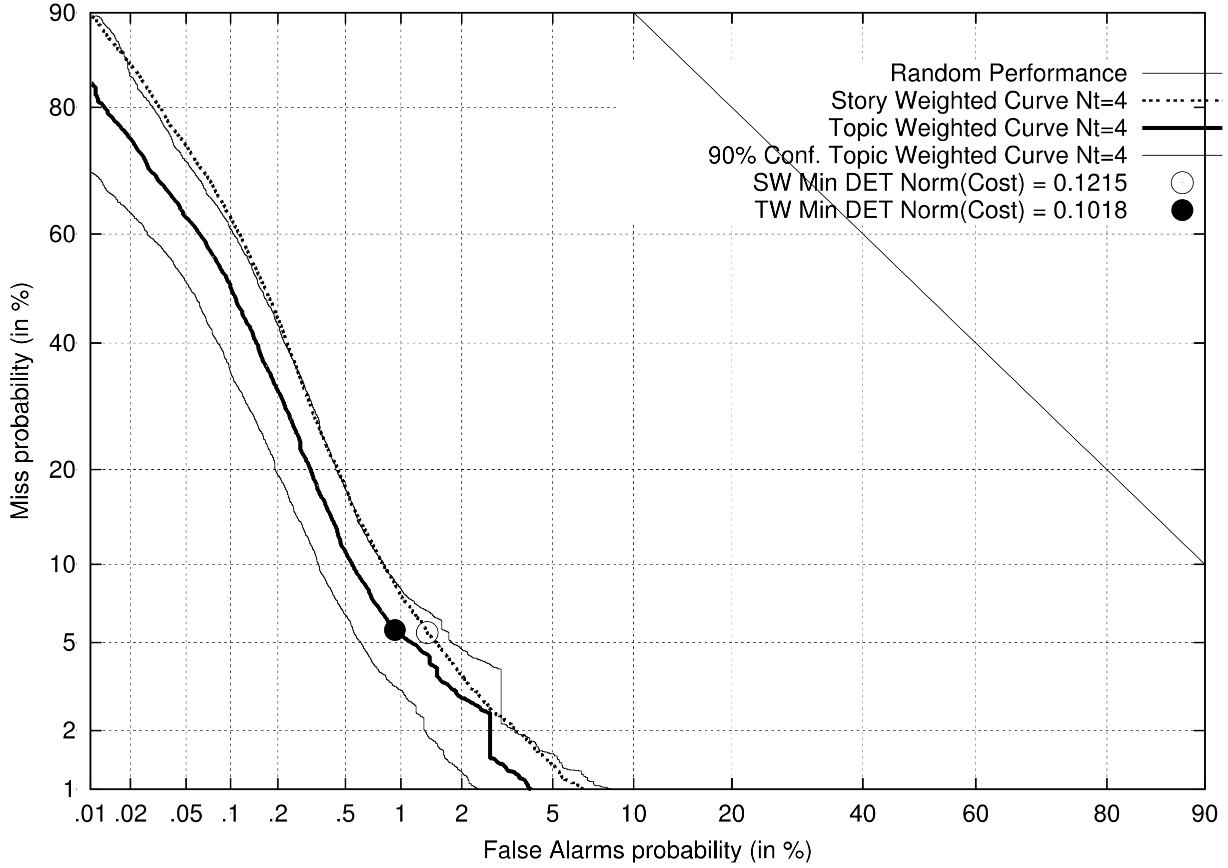
$$P_{\text{FA}} = \frac{\#\text{FA}}{\#\text{nontarget}}$$

C_{miss} cost of a miss

C_{FA} cost of a false alarm

[Allan 2002]

Detection-Error Curve



[Allan 2002]

TREC Microblog

- Retrospective search of a microblog corpus (Twitter).
- Topic definition
  - title: short keyword-style query
  - description: longer explanation of intent
  - time: time at which the query should be issued
- Evaluation
  - topical relevance labels
  - use classic ad hoc metrics with predicted-relevant documents **in reverse chronological order**

[Soboroff *et al.* 2012]

TREC Microblog

- Online filtering of a microblog corpus (Twitter).
- Topic definition
  - title: short keyword-style query
  - description: longer explanation of intent
  - time range: times during which the filtering should occur
- Evaluation
  - topical relevance labels
  - use classic filtering metrics with predicted-relevant documents

[TREC Microblog 2012 Guidelines]

TREC 2013 Temporal Summarization Track

- **Sequential Update Summarization**: broadcast useful, new, and timely sentence-length updates about a developing event.

- **Value Tracking**: can track the value of important event-related attributes (e.g. number of fatalities, financial impact).

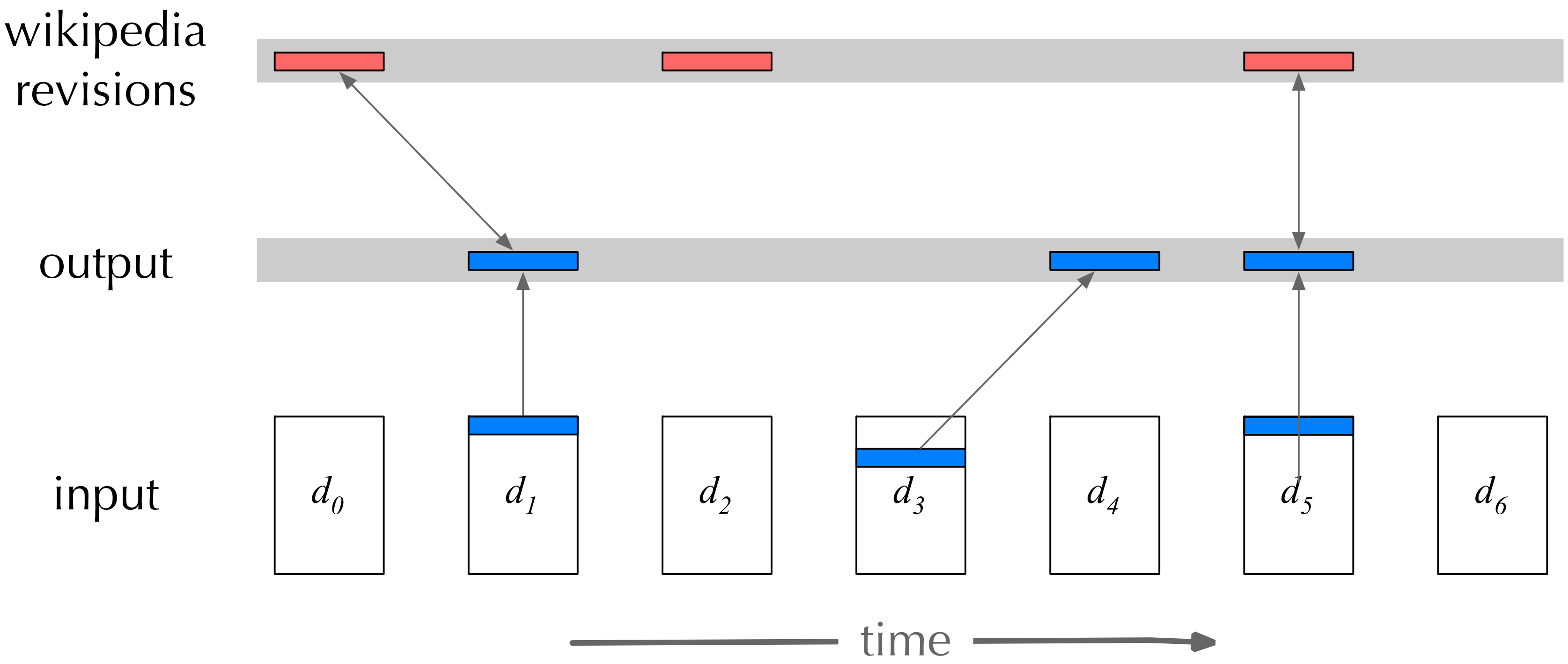
Track Goals

- to develop algorithms which detect sub-events with **low latency**.
- to develop algorithms which **minimize redundant information** in unexpected news events.
- to model information **reliability** in the presence of a dynamic corpus.
- to understand and address the sensitivity of text summarization algorithms in an online, sequential setting.
- to understand and address the sensitivity of information extraction algorithms in dynamic settings.

Sequential Update Summarization

- **corpus:** stream of documents
- **input:** tracking query, event onset time
- **output:** relevant, novel, and timely text updates
- **target:** gold standard, time-stamped updates

Sequential Update Summarization



Corpus

- desired properties
  - timestamped documents
  - topically relevant
  - diverse

Input

- ~10 large events occurring in timespan of corpus
- <event onset time, keyword query>
- <event onset time, first wikipedia revision>

2011 Tōhoku earthquake and tsunami

From Wikipedia, the free encyclopedia

(diff) ← Previous revision | Latest revision (diff) | Newer revision → (diff)

An earthquake occured on 30 km (80 miles) E of Sendai, Honshu, Japan. The earthquake possible to create regional tsunami on the zone.

- USGSEvent ID usc0001xgp http://earthquake.usgs.gov/earthquakes/recenteqsww/Quakes/usc0001xgp.php
- Integrated Tsunami Watcher Service http://www.iibc.in/itws/

Output

- timestamp of the system decision, *not necessarily the the source document*
- id of sentence detected in the annotated corpus
- support
  - id of supporting document(s)

Gold Standard Output

- nuggets semi-automatically derived from wikipedia revision history.

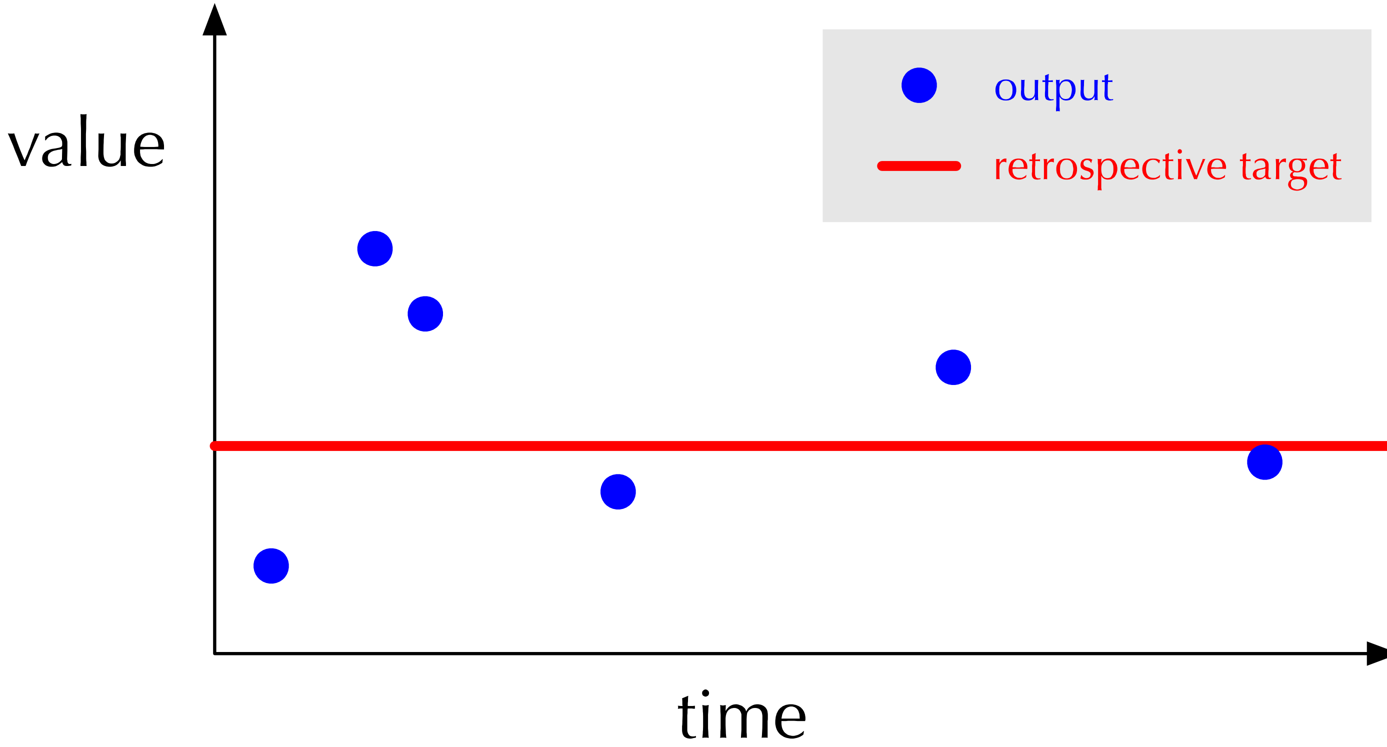
Evaluation

- **precision**: fraction of system updates that match any Gold Standard update.
- **recall**: fraction of Gold Standard updates that are matches by the system.
- **novelty**: fraction of system updates which did not match the same Gold Standard update.
- **timeliness**: difference between the system update time and the matched Gold Standard update time.

Value Tracking

- **corpus:** stream of documents
- **input:** tracking query, event onset time, attribute type
- **output:** running estimate of retrospective attribute value
- **target:** gold standard, retrospective attribute value

Value Tracking



Input

- ~10 large events shared with Task 1
- attributes
  - fatalities
  - financial impact
- <event onset time, keyword query, attribute type>

Output

- estimate
  - extractive
  - generative
- support
  - id of supporting document(s)

Gold Standard Output

- can be extracted from wikipedia infoboxes

Evaluation

- cumulative error rate from event onset to the end of the stream.

Research Problems

- Errors in editorial data
  - older topics are harder to reliable evaluate
- Simulating historic system state
  - need to “rewind the corpus” to the simulate the state of the index at retrieval/decision-making time
  - need to “rewind external information” to prevent “signals from the future”

Schedule

- Introduction (9:00-9:15)
- Modeling Dynamics
  - 9:15-10:15 Web content dynamics [Susan]
  - 10:15-10:45 Web user behavior dynamics [Milad]
  - 10:45-11:00 Break
  - 11:00-11:30 Web user behavior dynamics, cont'd
  - 11:30-13:00 Spatio-temporal analysis [Fernando]
  - Methods for evaluation
- Lunch (13:00-14:30)
- Applications to Information Retrieval
  - 14:30-15:45 Temporal NLP [Kira]
  - News event prediction
  - 15:45-16:00 Break
  - 16:00-17:45 Time-sensitive search [Yi]
  - Time-sensitive recommendations [Anlei]
- Wrap-Up (17:45-18:00)

buon appetito

Temporal NLP & News Prediction

WSDM 2013 Tutorial

Outline

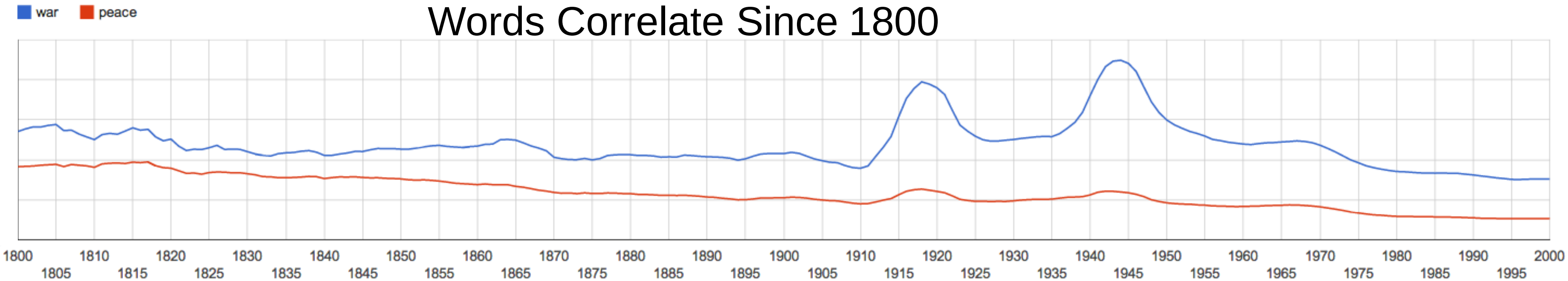
- Temporal Language Models
  - Temporal Word Representation
  - Temporal Document Representation
  - Temporal Topics Representation
- Temporal Information Extraction
- Future Event Prediction from News
  - Future Event Retrieval from text
  - Future Event Retrieval from query stream
  - Future Event Retrieval from social media
- Temporal Summarization
  - Single Timeline
  - Multiple Timeline

Outline

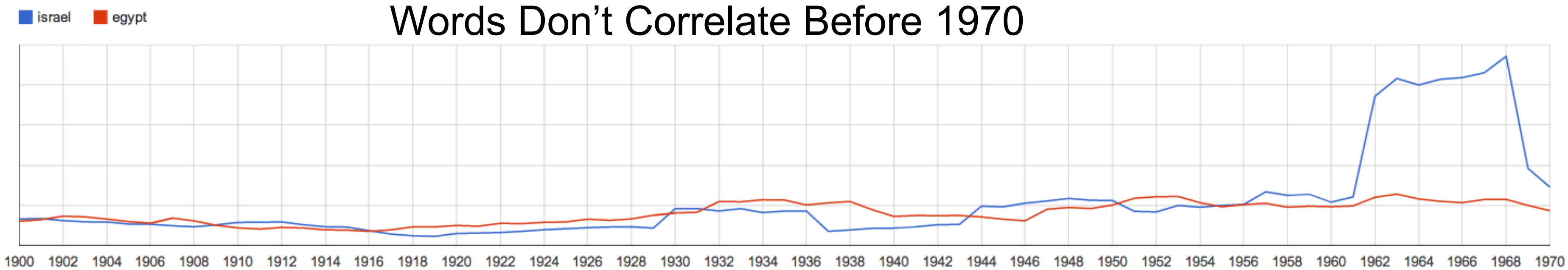
- Temporal Language Models
  - Temporal Word Representation
  - Temporal Document Representation
  - Temporal Topics Representation
- Temporal Information Extraction
- Future Event Prediction from News
  - Future Event Retrieval from text
  - Future Event Retrieval from query stream
  - Future Event Retrieval from social media
- Temporal Summarization
  - Single Timeline
  - Multiple Timeline

Words Over Time

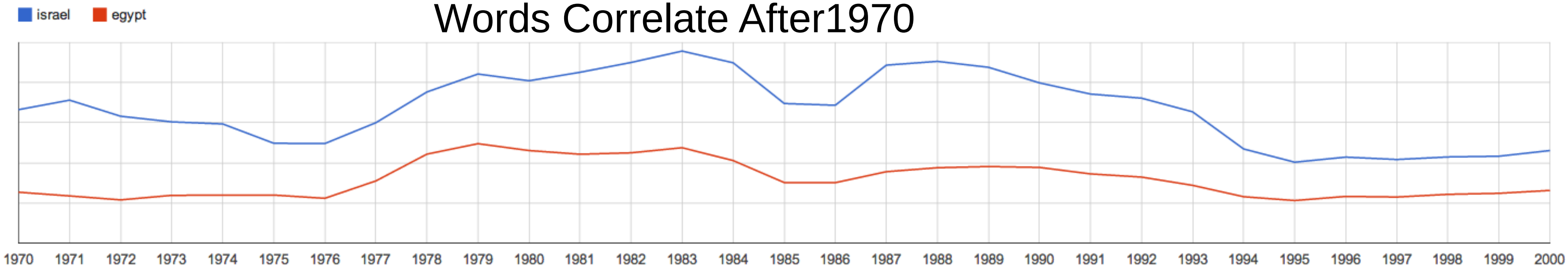
Words Correlate Since 1800



Words Don't Correlate Before 1970



Words Correlate After1970



Words Over Time (Temporal Correlation)

1. Temporal representation of text

Word

Represent
a word using its
query volume

Extend static
representation
with temporal
dynamics

2. Temporal text-similarity

Word1

Cross
Correlation
Or
DTW

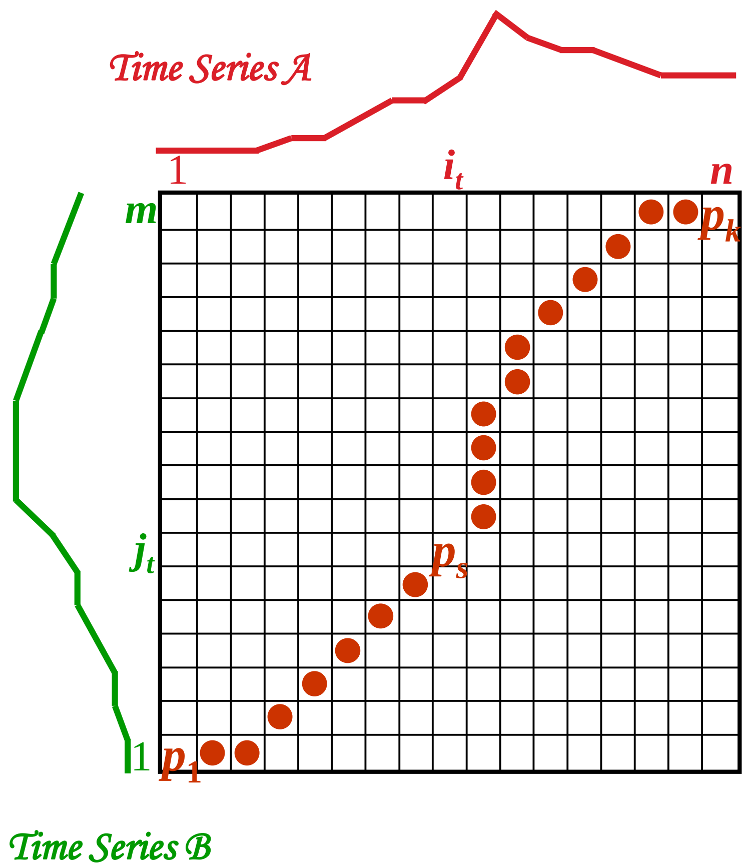
Word 2

Method for computing
semantic relatedness
using the temporal
representation

Steve Chien, Nicole Immorlica: Semantic similarity between search engine queries using temporal correlation. WWW 2005: 2-11

Temporal Correlation Methods (1):

Dynamic time warping (DTW)



Time-weighted distance between $\mathcal{A}$ and $\mathcal{B}$:

$$D(\mathcal{A}, \mathcal{B}) = \sum_{t=1}^{k} d(p_t) \cdot w(t)$$

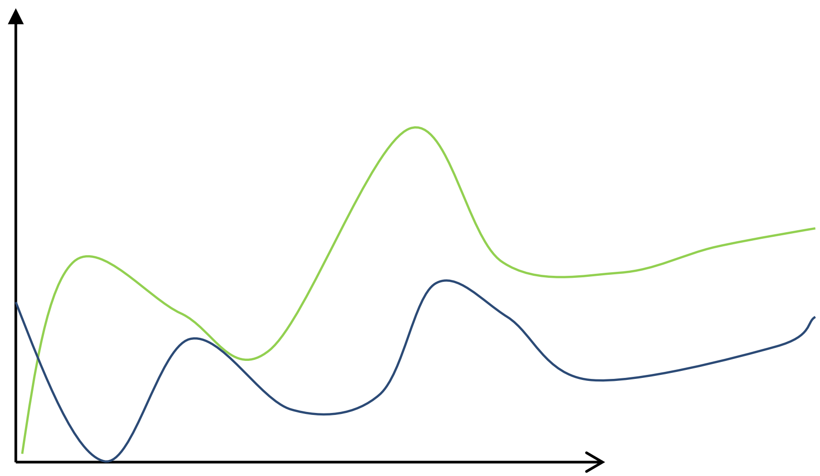
$d(p_s)$: distance between i_t and j_t

$w(t) > 0$: weighting coefficient (with decay over time)

Best alignment path between $\mathcal{A}$ and $\mathcal{B}$:

$$P_0 = \arg\min_{P} (D(\mathcal{A}, \mathcal{B})).$$

Temporal Correlation Methods (2):
Cross correlation



Time-weighted distance between $\mathcal{A}$ and $\mathcal{B}$:

$$D(\mathcal{A}, \mathcal{B}) = \sum_{t=0}^{n} w(t)x(t)y(t-s)$$

$$s = 0, \pm 1, \pm 2, \ldots$$

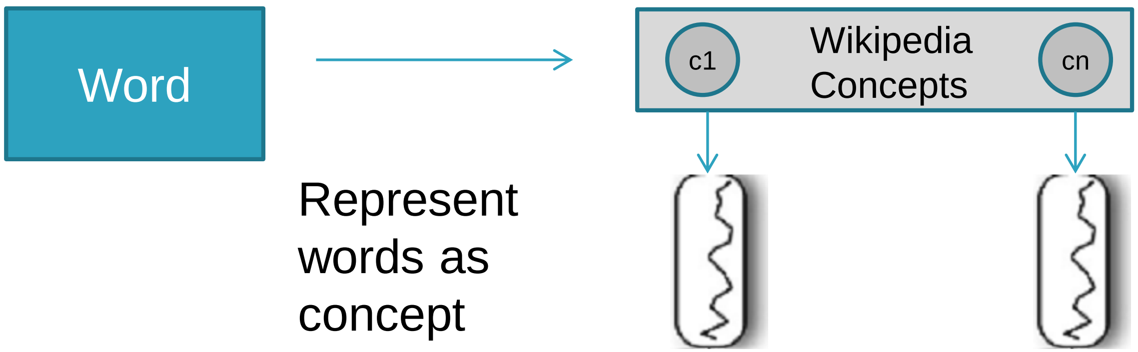
$w(t) > 0$: weighting coefficient (with decay over time)

Best alignment path between $\mathcal{A}$ and $\mathcal{B}$:

$$P_0 = \arg\min_{s}(D(\mathcal{A}, \mathcal{B})).$$

Words Over Time (TSA)

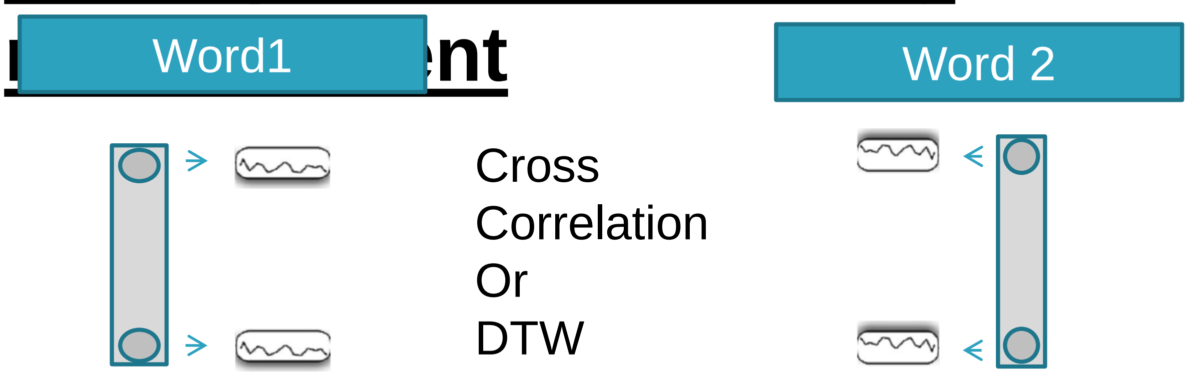
1. Temporal representation of text



Represent words as concept vectors

Extend static representation with temporal dynamics

2. Temporal text-similarity



Cross Correlation Or DTW

Method for computing semantic relatedness using the temporal representation

Kira Radinsky, Eugene Agichtein, Evgeniy Gabrilovich, Shaul Markovitch: A word at a time: computing word relatedness using temporal semantic analysis. WWW 2011: 337-346

Words Over Time (Time Schemas)

1. Temporal representation of text

Word

Clicked Web Pages

P 1

P n

<week, day of the week, hour >, …

Represent words as concept vectors

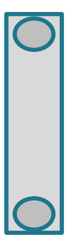
Time Schemas

Extend static representation with temporal dynamics

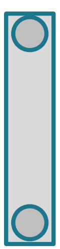
2. Temporal text-similarity
ent

Word1

Word 2



Measure content similarity only during the time schemas

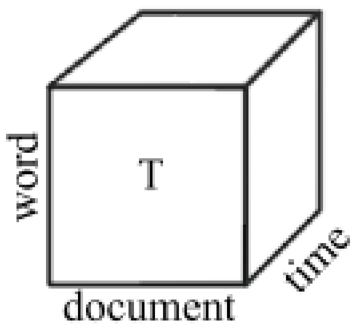


Method for computing semantic relatedness using the temporal representation

Zhao et al. :Time-Dependent Semantic Similarity Measure of Queries Using Historical Click-Through Data. WWW 2006

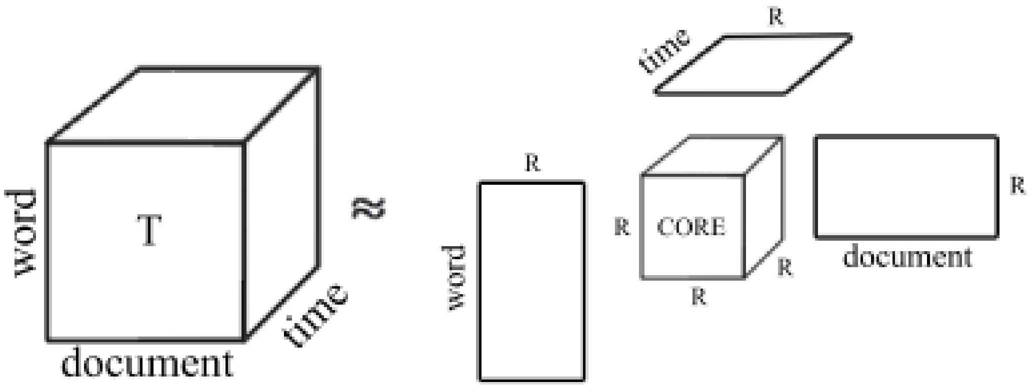
Words Over Time (tLSA)

1. Temporal representation of text



Extend static representation with temporal dynamics

2. Temporal text-similarity me

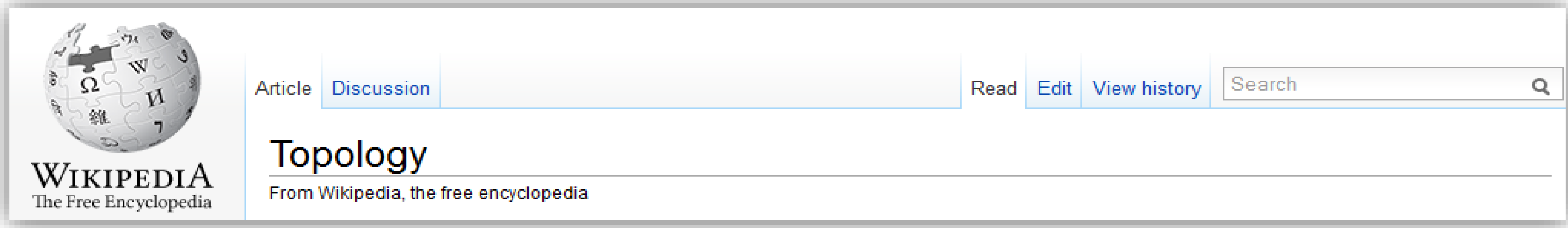


CANDECOMP/ PARAFAC (CP) Decomposition For Tensors

Yu Wang, Eugene Agichtein: Temporal latent semantic analysis for collaboratively generated content: preliminary results. SIGIR 2011: 1145-1146

Documents Over Time (RHA)

Redefine term frequency (TF): a term is relatively important if it appears in the early revisi



First revision

Topology, in mathematics, is both a structure used to capture the notions of continuity, connectedness and convergence, and the name of the branch of mathematics which studies these.

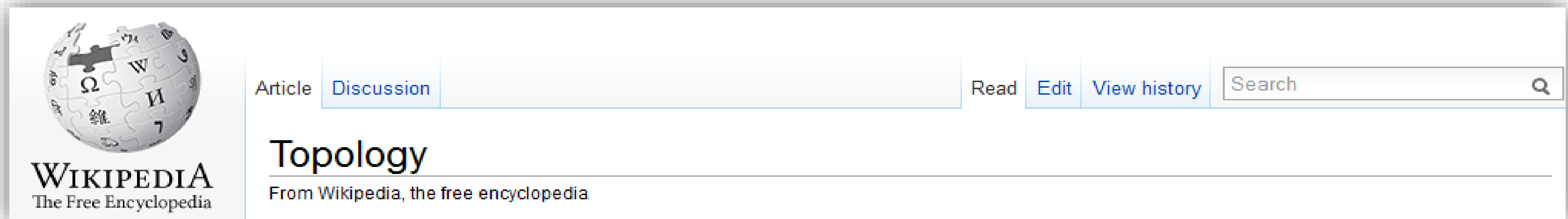
Current version

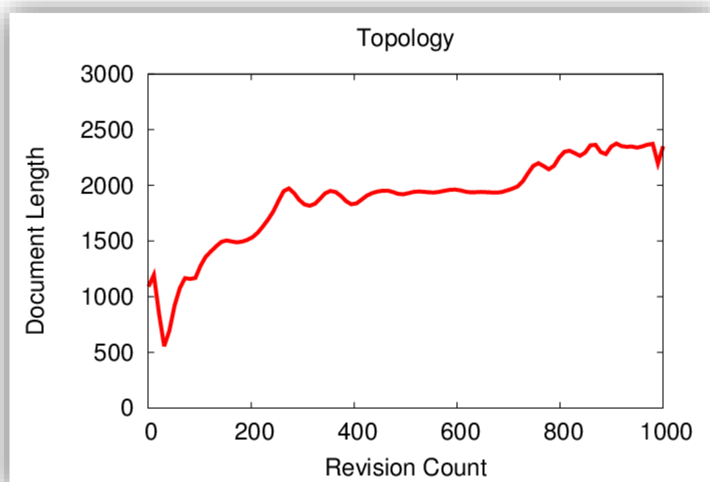
Topology (from the Greek τόπος, “place”, and λόγος, “study”) is a major area of mathematics concerned with spatial properties that are preserved under continuous deformations of objects, for example
…..
basic examples include compactness and connectedness

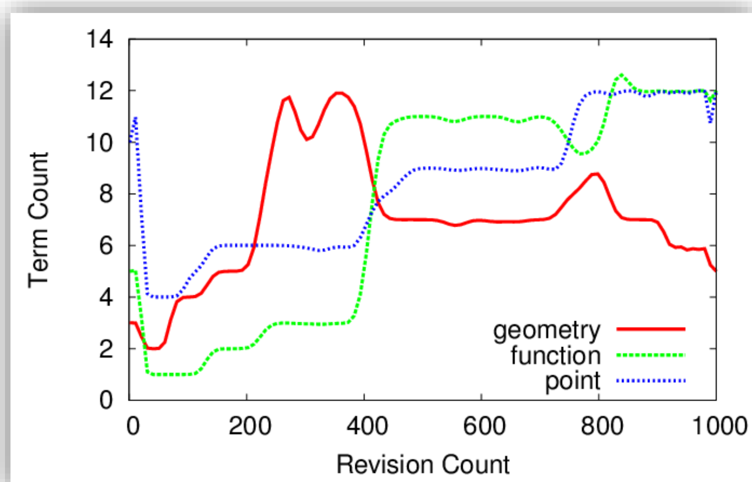
Ablimit Aji, Yu Wang, Eugene Agichtein, Evgeniy Gabrilovich: Using the past to score the present: extending term weighting models through revision history analysis. CIKM 2010: 629-638

Documents Over Time (RHA)

Redefine term frequency (TF): a term is relatively important if it appears in the early revisi





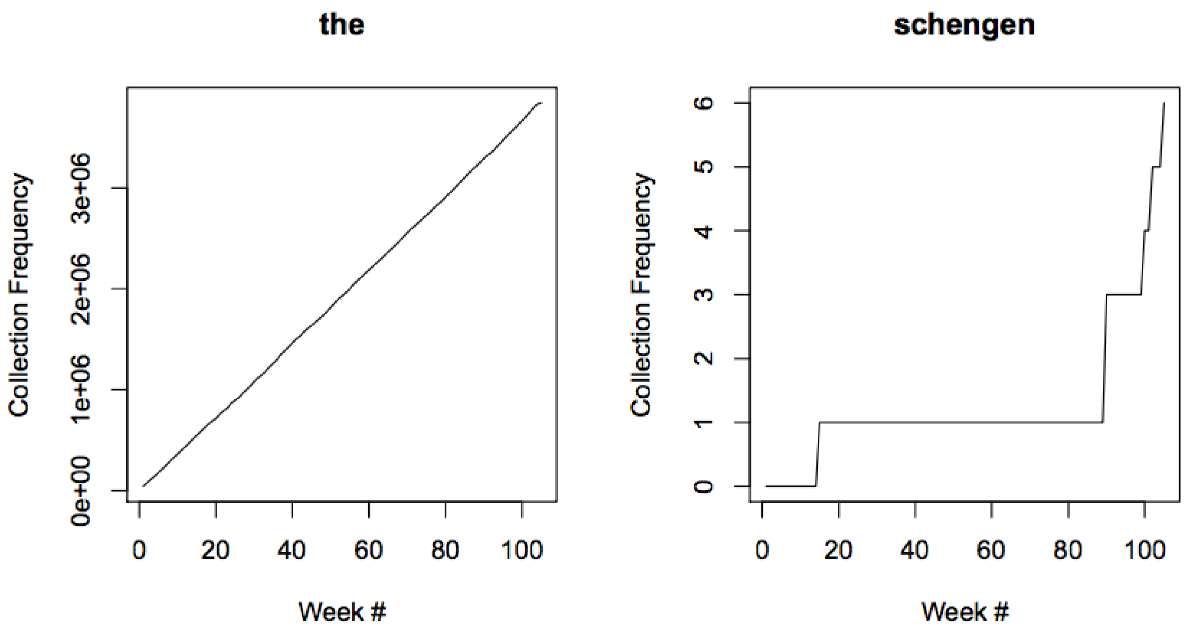


Ablimit Aji, Yu Wang, Eugene Agichtein, Evgeniy Gabrilovich: Using the past to score the present: extending term weighting models through revision history analysis. CIKM 2010: 629-638

Documents Over Time (time series approach)

The temporal behavior of

1. Weak discriminators is easily described by a simple linear time series model,
2. Useful discriminators' distribution over time is too erratic to describe faithfully with a linear model.



Miles Efron: Linear time series models for term weighting in information retrieval. JASIST 61(7): 1299-1312 (2010)

Common Time Series Approaches: The State Space Models

Model

For example, **semi-linear state space** modeling

The prediction for time t

Error at time t

$$Y_t = W(\theta)X_t + \epsilon_t,$$
$$X_{(t+1)} = F(\theta)X_t + G(\theta)\epsilon_t,$$

State vector a time t
(inc. last point, trend, etc.)

Learn Structure and Parameters

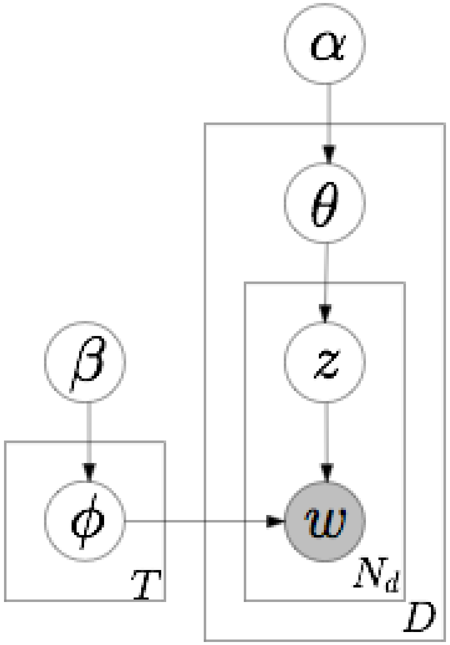
Predict Y

J. Durbin and S. Koopman, Forecasting with Exponential Smoothing (The State Space Approach), 2008

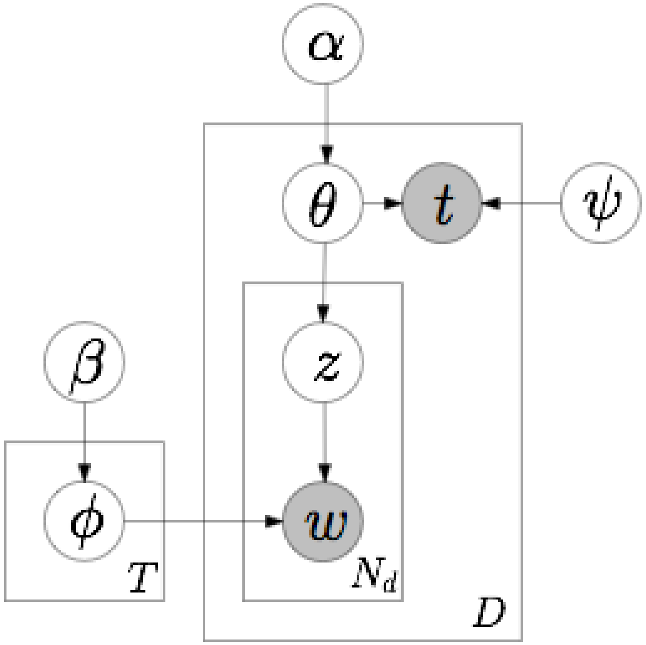
Topics Over Time

- Discretization: Slicing time-ordered data into discrete subsets:
  - Train globally, inspect separately [Griffiths and Steyvers, 2004]
  - Train and inspect separately [Wang, Mohanty and McCallum, 2005]
- Being Markovian: Topic transiting at certain time stamps:
  - The state at time $t + 1$ *or* $t + \Delta t$ is independent of all other history given the state at time t.
  - State-Space model, Hidden Markov model, Kalman filters, etc. [Blei and Lafferty, 2006]
  - Continuous Time Bayesian Network [Nodelman et al., 2002]
- Graphical Models
  - Topics over Time (TOT) [Wang and McCallum, SIGKDD 2006]
  - PAM Over Time (PAMTOT) [Li, Wang and McCallum AAAI Workshops 2006]

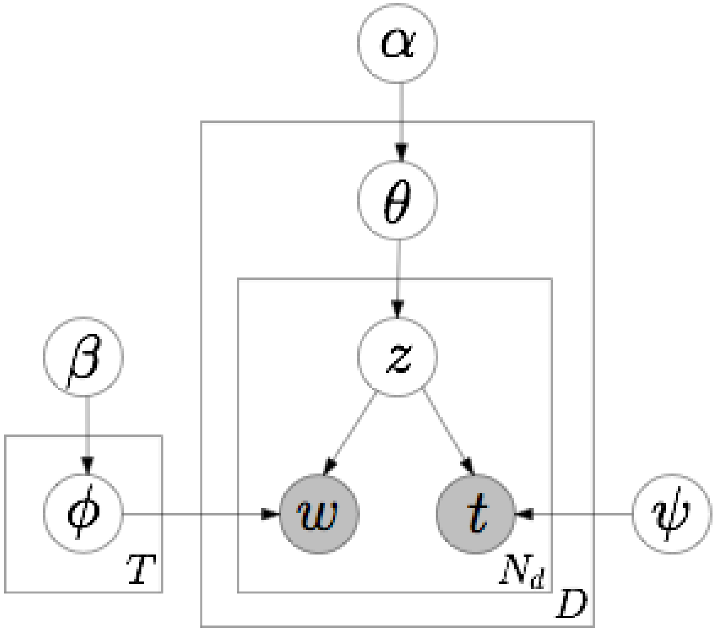
LDA and Topics over Time (ToT)



LDA
(Sampling)



TOT



TOT

[Wang and McCallum, SIGKDD 2006]

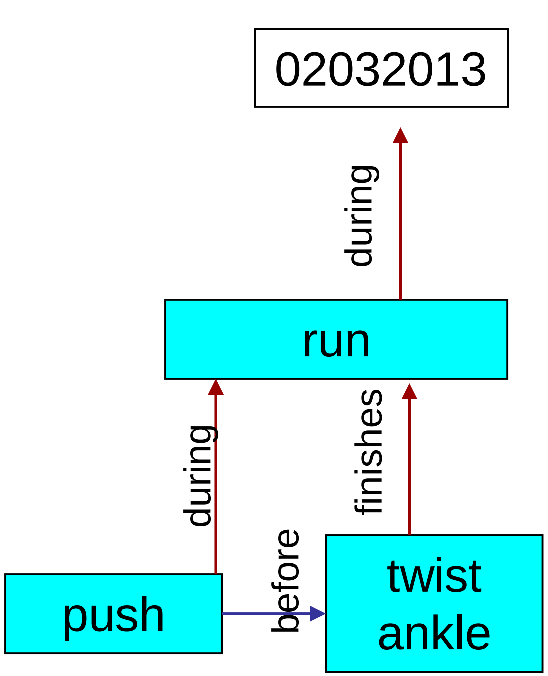
Outline

- Temporal Language Models
  - Temporal Word Representation
  - Temporal Document Representation
  - Temporal Topics Representation
- Temporal Information Extraction
- Future Event Prediction from News
  - Future Event Retrieval from text
  - Future Event Retrieval from query stream
  - Future Event Retrieval from social media
- Temporal Summarization
  - Single Timeline
  - Multiple Timeline

Temporal Information Extraction

Feb. 04, 2013

Yesterday Holly was running a marathon when she twisted her ankle. David had pushed her.



- Input: A natural language discours
- Output: representation of events and their temporal relations
- Applications:
  - Temporal QA
  - Temporal Summarization
  - Temporal Expressions in Query Log

[Mani, IJCAI Tutorial 2007]

Temporal Information Extraction

- Temporal entity (events and attributes) recognition
  - Knowledge-based methods (dictionary and rules)
  - ML based methods (annotated corpus)
    - TimeML
    - Time Expression Recognition and Normalization (TERN)
- Temporal relations discovering
  - Absolute Relations – placing event on timeline
  - Relative Relations – relations between events
- Temporal reasoning
  - Allen's Interval-Based Ontology [Allen, AI'84]

Example: Temporal Web-Mined Rules

- Lexical relations (capturing causal and other relations, etc.)
  - kill => die (always)
  - push => fall (sometimes: *Max fell. John pushed him*.)
- Idea: leverage the distributions found in large corpora
- VerbOcean: database from ISI that contains lexical relations mined from Google searches
  - E.g., X happens before Y, where X and Y are WordNet verbs highly associated in a corpus
- Yields 4199 rules!

Corpora

- News (newswire and broadcast)
  - TimeML: TimeBank, AQUAINT Corpus (all English)
  - TIMEX2: TIDES and TERN English Corpora, Korean Corpus (200 docs), TERN Chinese and Arabic news data (extents only)
- Weblogs
  - TIMEX2 TERN corpus (English, Chinese, Arabic – the latter with extents only)
- Dialogues
  - TIMEX2- 95 Spanish Enthusiast dialogs, and their translations
- Meetings
  - TIMEX2 Spanish portions of UN Parallel corpus (23,000 words)
- Children's Stories
  - Reading Comprehension Exams from MITRE, Remedia: 120 stories, 20K words, CBC: 259 stories, 1/3 tagged, ~50K

Links

- TimeBank:
  - http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T08
- TimeML:
  - www.timeml.org
- TIMEX2/TERN ACE data (English, Chinese, Arabic):
  - timex2.mitre.org
- TIMEX2/3 Tagger:
  - http://complingone.georgetown.edu/~linguist/GU_TIME_DOWNLOAD.HTML

References

1. Berrazega (2012) *Temporal information extraction: A survey*. International Journal on Natural Language Computing (IJNLC)
2. Ling, X., & Weld, D. (2010). *Temporal information extraction.* In Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI).
3. Yoshikawa, K., Riedel, S., Asahara, M., & Matsumoto, Y. (2009). *Jointly identifying temporal relations with markov logic.* In Proceedings of the Third International Joint Conference on Natural Language Processing (ACL IJCNLP).
4. Tatu, M., & Srikanth, M. (2008). *Experiments with reasoning for temporal relations between events*. In Proceedings of the International Conference on Computational Linguistics (COLING).
5. Chambers, N., Wang, S., & Jurafsky, D. (2007). *Classifying temporal relations between events*. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) (Poster).
6. Lapata, M., & Lascarides, A. (2006). Learning sentence-internal temporal relations. Journal of Artificial Intelligence Research (JAIR), 27, 85–117.
7. Mani, I., Pustejovsky, J., and Gaizauskas, R. (eds.). (2005) *The Language of Time: A Reader*. Oxford University Press.
8. Mani, I., and Schiffman, B. (2004). *Temporally Anchoring and Ordering Events in News.* In Pustejovsky, J. and Gaizauskas, R. (eds), Time and Event Recognition in Natural Language. John Benjamins, to appear.
9. Mani, I. (2004). *Recent Developments in Temporal Information Extraction*. In Nicolov, N., and Mitkov, R. Proceedings of RANLP'03, John Benjamins
10. Jang, S., Baldwin, J., and Mani, I. (2004). *Automatic TIMEX2 Tagging of Korean News*. In Mani, I., Pustejovsky, J., and Sundheim, B. (eds.), ACM Transactions on Asian Language Processing: Special issue on Temporal Information Processing.
11. Mani, I., Schiffman, B., and Zhang, J. (2003). *Inferring Temporal Ordering of Events in News.* Short Paper. In Proceedings of the Human Language Technology Conference (HLT-NAACL'03).
12. Ferro, L., Mani, I., Sundheim, B. and Wilson G. (2001). *TIDES Temporal Annotation Guidelines Draft - Version 1.02*. MITRE Technical Report MTR MTR 01W000004. McLean, Virginia: The MITRE Corporation.

Outline

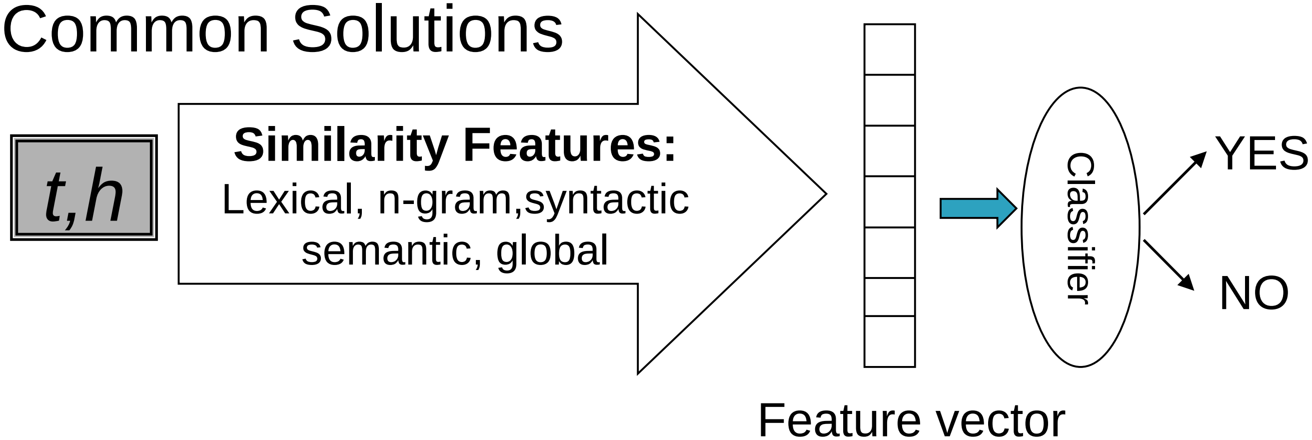
- Temporal Language Models
  - Temporal Word Representation
  - Temporal Document Representation
  - Temporal Topics Representation
- Temporal Information Extraction
- Future Event Prediction from News
  - Future Event Retrieval from text
  - Future Event Retrieval from query stream
  - Future Event Retrieval from social media
- Temporal Summarization
  - Single Timeline
  - Multiple Timeline

Future Event Retrieval from Text (Textual Entailment)

- A directional relation between two text fragments: *Text (t)* and *Hypothesis (h):*

> *t entails h* ($t \Rightarrow h$) if
> humans reading *t* will infer that *h* is most likely true

- Common Solutions



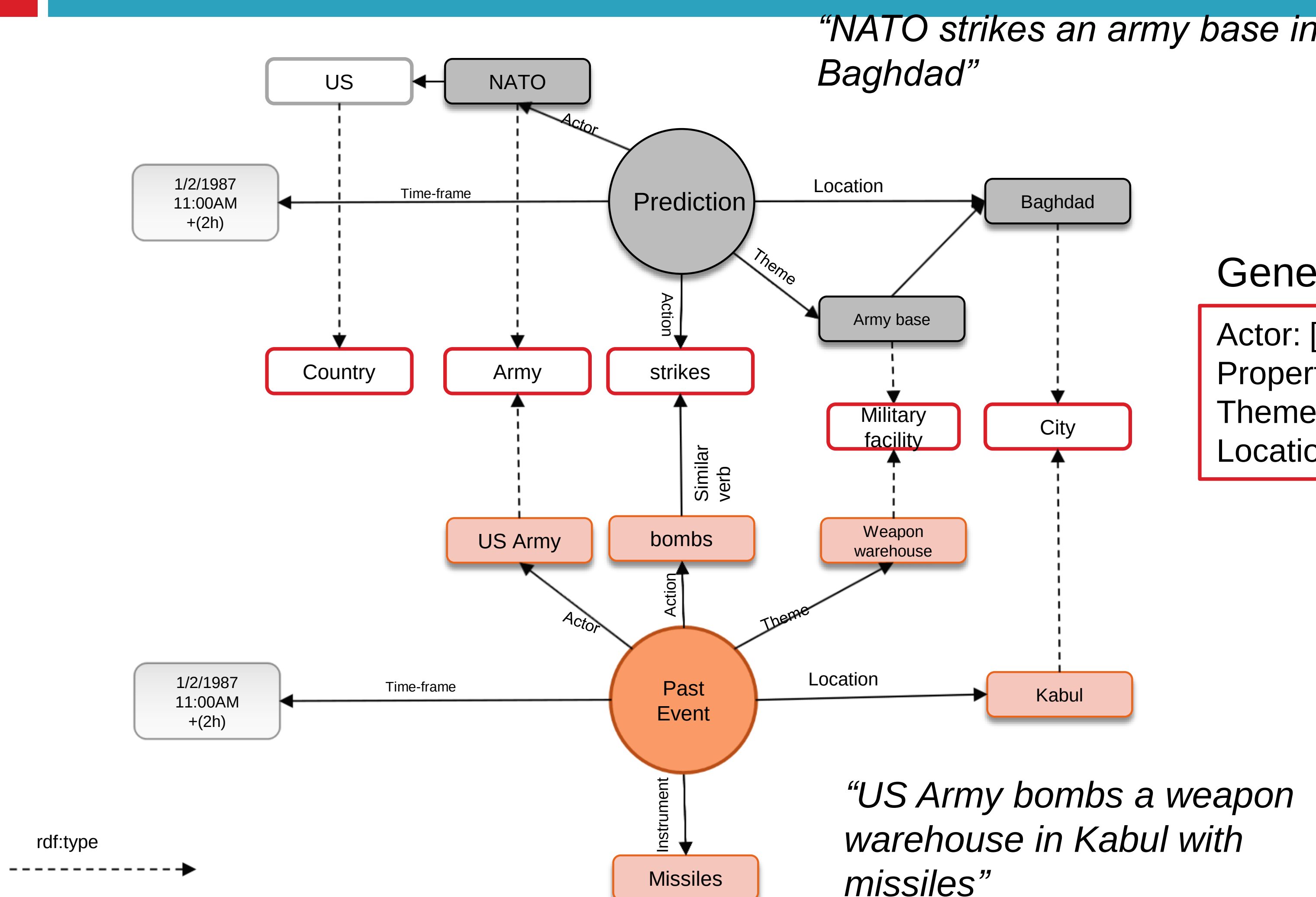
Androutsopoulos and Malakasiotis, JAIR'10
Glickman, Dagan, Koppel, AAAI'05
Dagan, Roth, Zanzotto, ACL'07

http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Portal

Future Event Retrieval from Text (Text Prediction)

- Template-based Approaches [Girju and Moldovan, FLAIRS 2000]
  - Discover lexico-syntactic patterns that can express the causal relation
  - Validate and rank the ambiguous patterns acquired based on semantic constraints on nouns and verbs.
- Co-Occurrences Approaches [Gordon, A. S., Bejan, C. A., & Sagae, K., AAAI 2011]
  - PMI Approaches on words
  - Sentence Proximity in a corpus (e.g., Blogs)
- Human Labeled Corpora
  - Framenet

Future Event Retrieval from Text (Generalized Text Prediction)

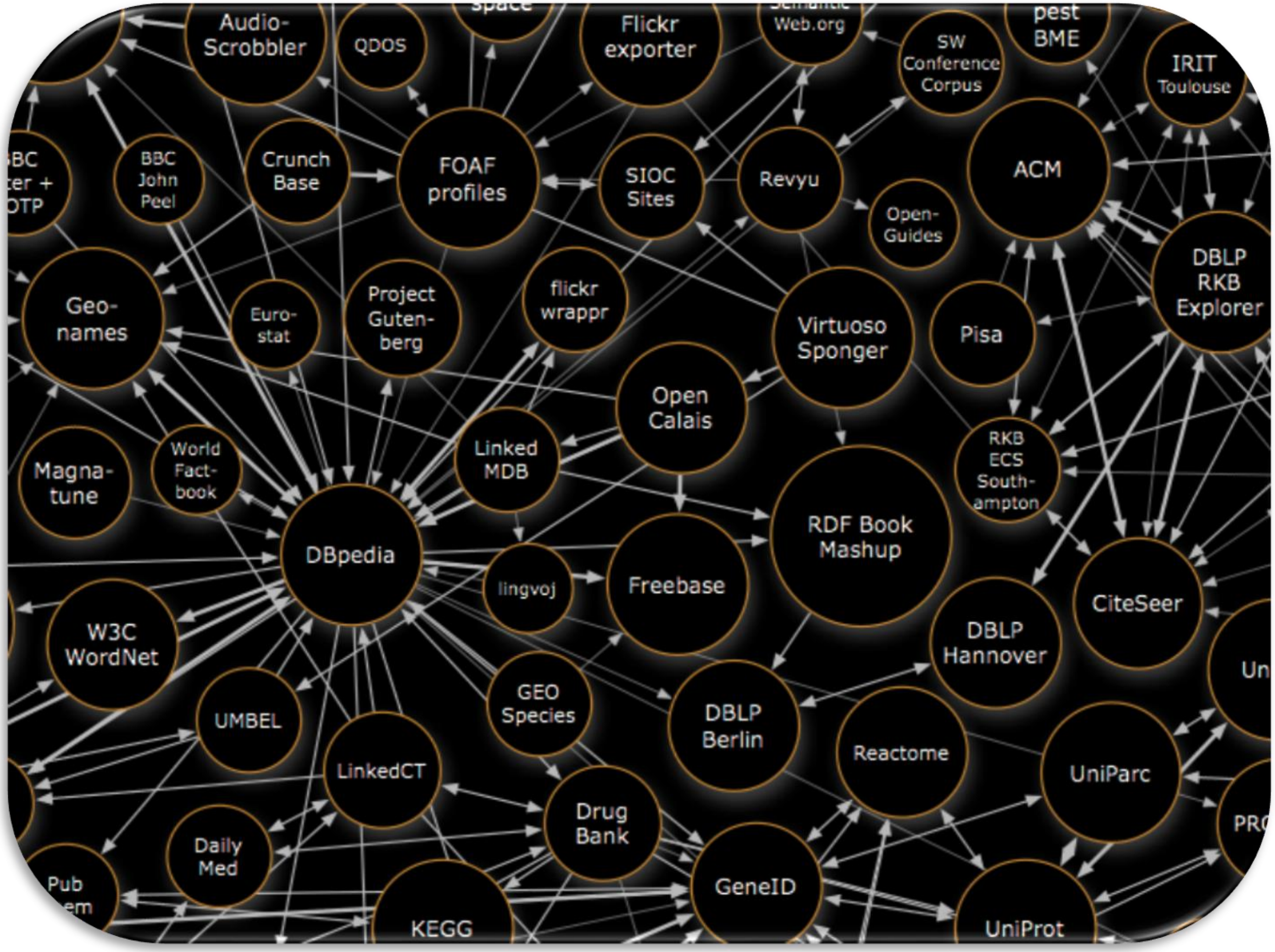


Generalization rule

Actor: [state of Nato]
Property: [Hit1.1]
Theme: [Military facility]
Location: [Arab City]

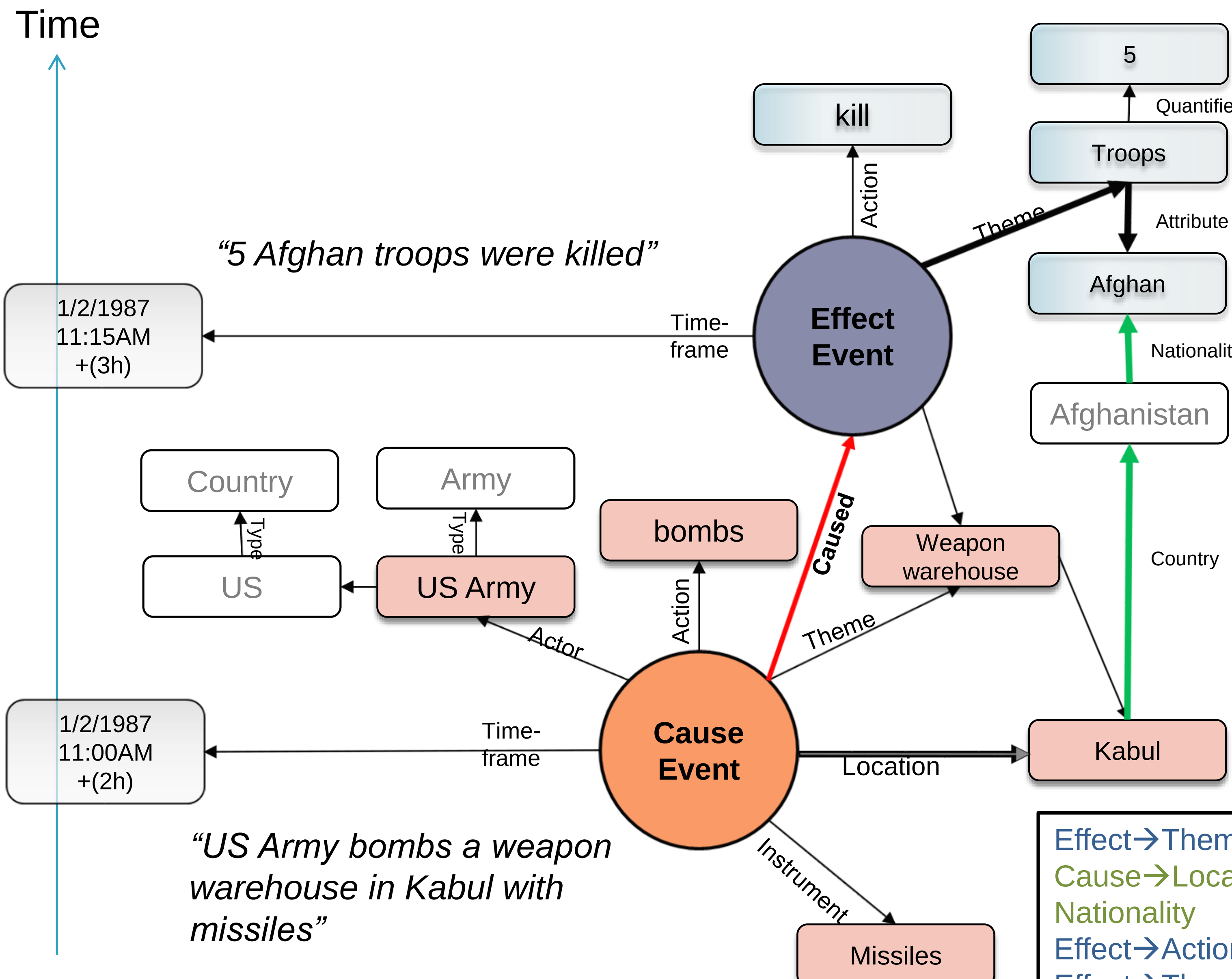
Radinsky, Davidovich, and Markovitch. Learning causality for news events prediction,

Ontology – Linked data



http://www.linkeddata.org

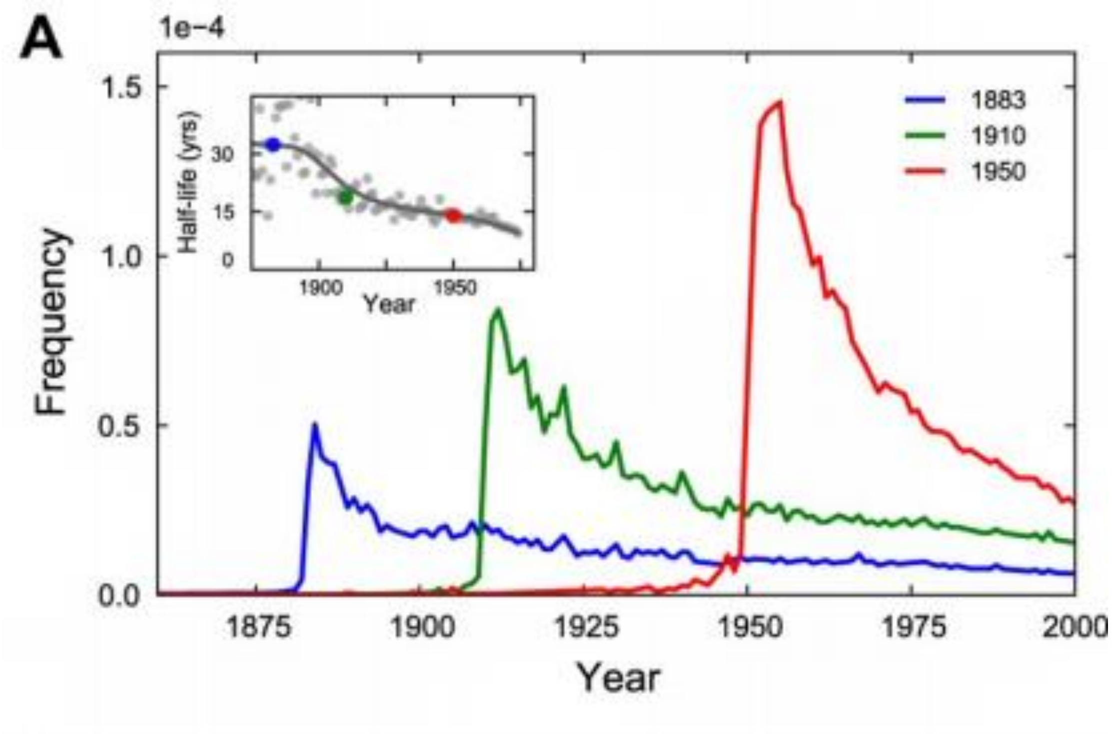
Prediction Rule Generation



Effect→Theme→**Attribute** =
Cause→Location→Country→
Nationality
Effect→Action=kill
Effect→Theme=Troops

Culturomics

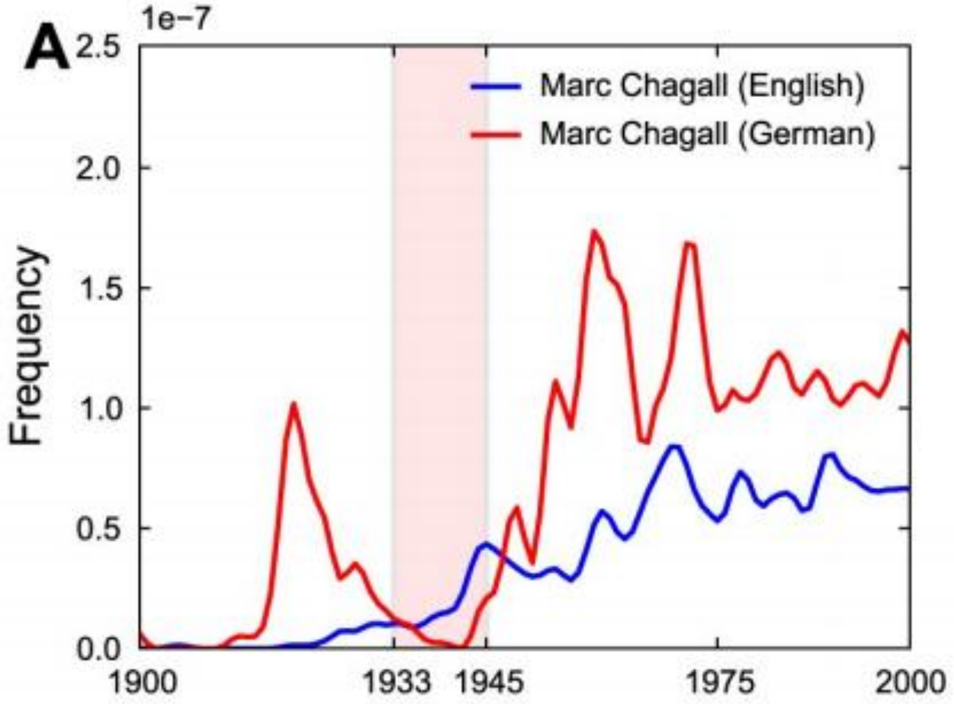
How long is history remembered?



Kalev, . First Monday, 15(9), 2011.
Michel et al. , Science 2011
Yeung and Jatowt CIKM'11

Culturomics

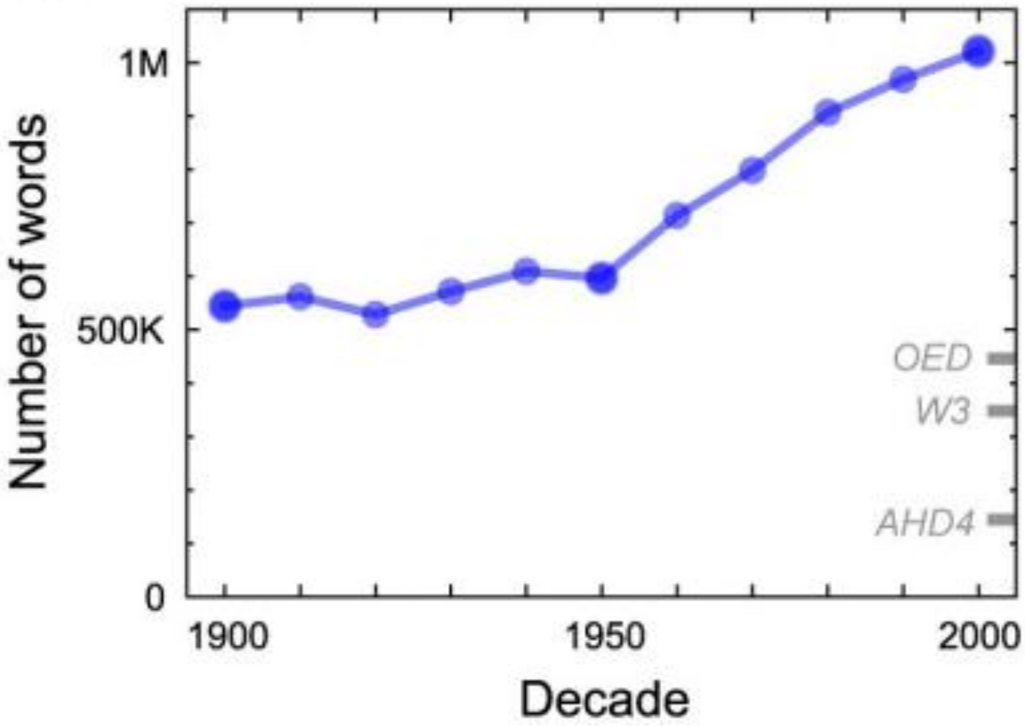
Detecting Censorship and Suppression



Kalev, . First Monday, 15(9), 2011.
Michel et al. , Science 2011
Yeung and Jatowt CIKM’11

Culturomics

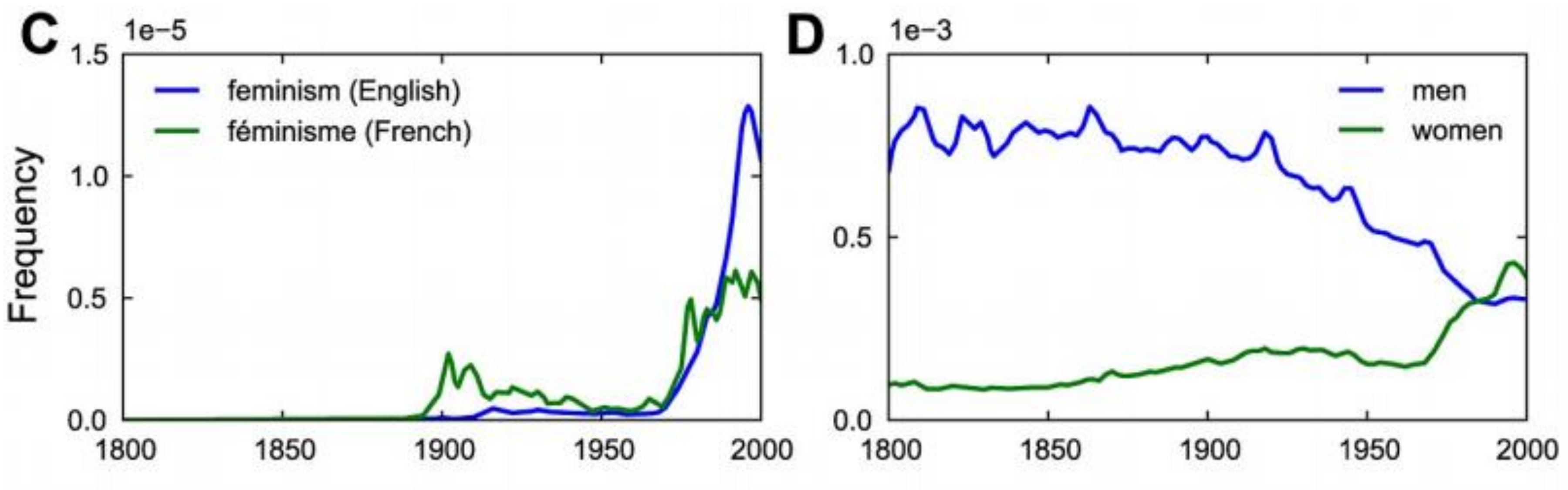
Language Evolution: Size of Lexicon,
Evolution of Grammar



Kalev, . First Monday, 15(9), 2011.
Michel et al. , Science 2011
Yeung and Jatowt CIKM’11

Culturomics

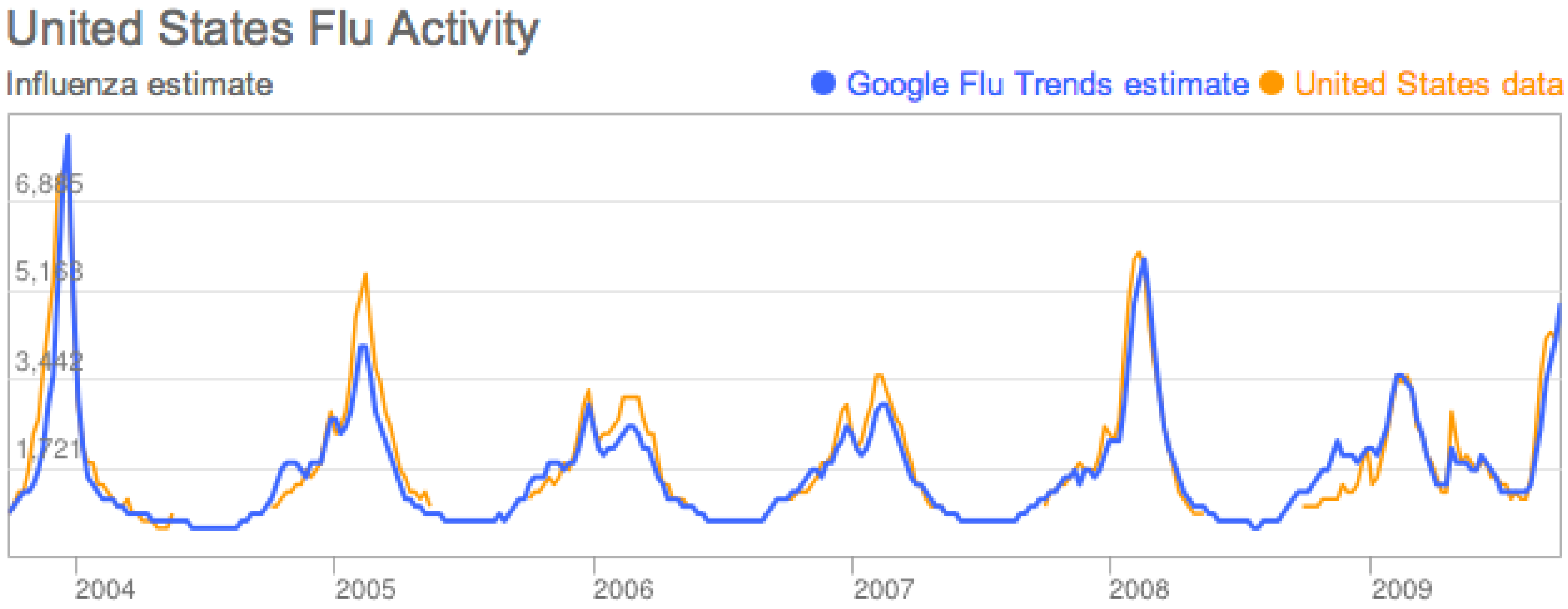
Women Rock!



Kalev, . First Monday, 15(9), 2011.
Michel et al. , Science 2011
Yeung and Jatowt CIKM’11

Future Event Retrieval using query stream

Using query volume [Ginsberg et al., Nature 2009]



Future Event Retrieval using query stream

Using Queries Correlations [Radinsky et al., WI'08]

Indication weight on the candidate

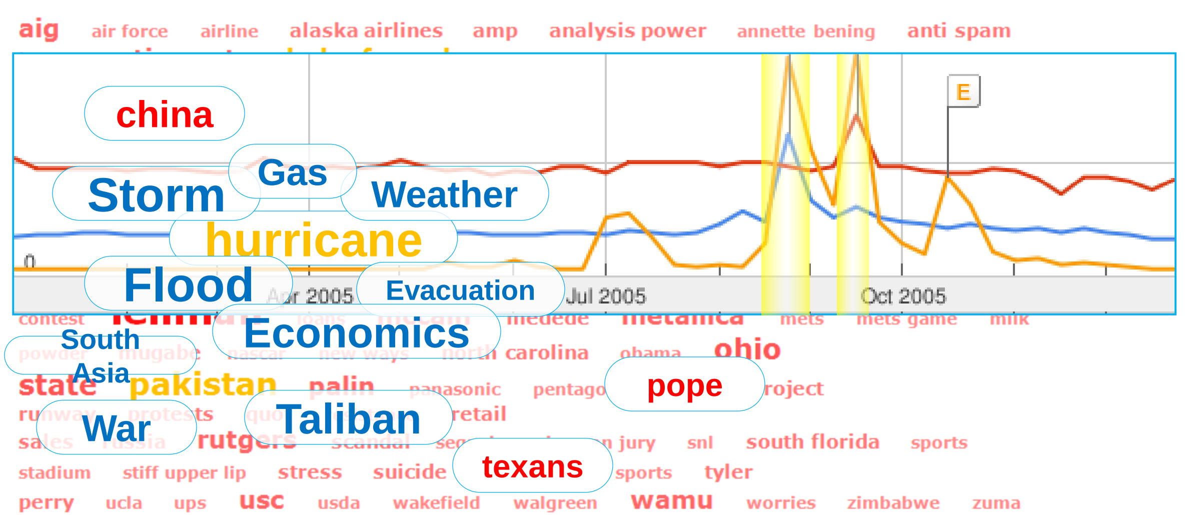
Today's salient terms

Future candidate terms

Likelihood to appear in k days

0.7

0.9

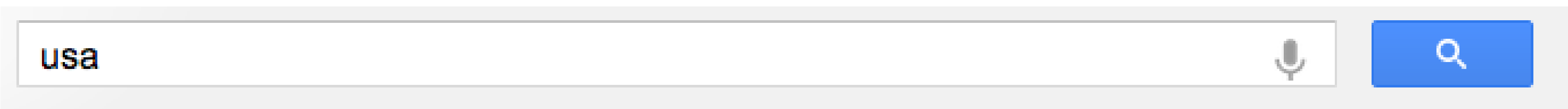


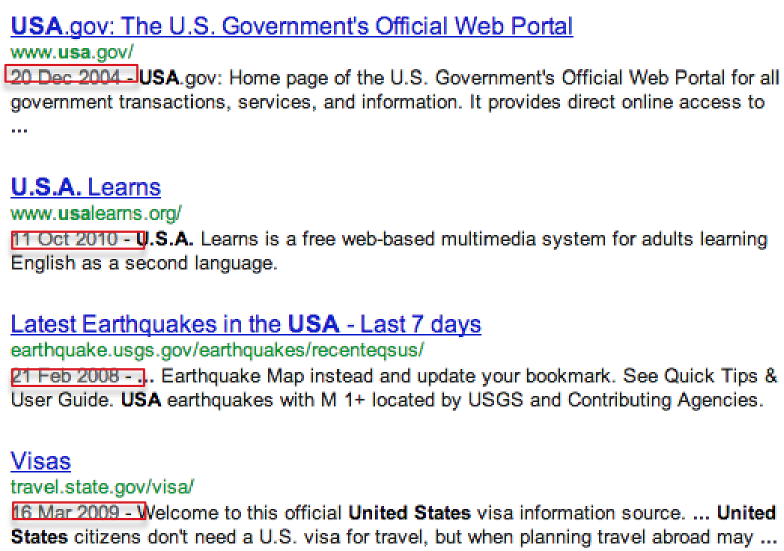
0.12
0.36
0.10
0.40
0.85
0.30
0.05
0.01
0.08

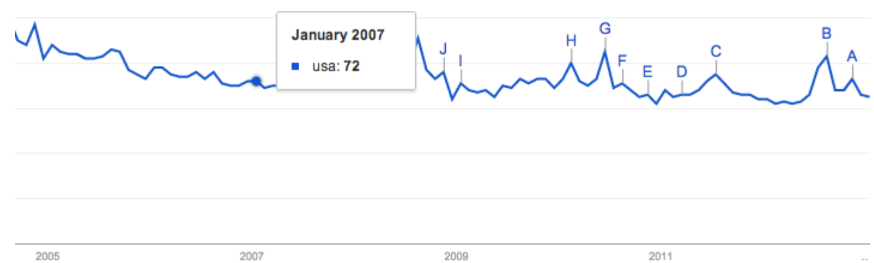
Goal: For each **candidate term** evaluate the probability of it to appear in the future, given **today's terms**.

Future Event Retrieval using query stream

Using relevant documents for future event prediction [Amodeo, Blanco, Brefeld, CIKM’ 11]







Based on publication dates of results bui
a probabilistic model

Future Event Retrieval from social media

- Predicting using Linear Regression on Chatter Rate
  - [S. Asur and B. A. Huberman. Predicting the future with social media, 2010.]
- Predicting Using syntactic and semantic features extracted from text and meta-text
  - [M. Joshi, D. Das, K. Gimpel, and N. A. Smith. Movie reviews and revenues: An experiment in text regression. In In Proc. of NAACL-HLT, 2010.]
- Predicting using Sentiment Analysis
  - [S. Asur and B. A. Huberman. Predicting the future with social media, 2010.]
  - G. Mishne. Predicting movie sales from blogger sentiment. In In AAAI Spring Symposium, 2006.
- Predict future posts
  - Using trending topic modeling and historical data [Wang, Agichtein and Benzi KDD’12]

Outline

- Temporal Language Models
  - Temporal Word Representation
  - Temporal Document Representation
  - Temporal Topics Representation
- Temporal Information Extraction
- Future Event Prediction from News
  - Future Event Retrieval from text
  - Future Event Retrieval from query stream
  - Future Event Retrieval from social media
- Temporal Summarization
  - Single Timeline
  - Multiple Timeline

Temporal Summarization

- Topic detection and tracking (TDT)
  - Lexical similarity, temporal proximity, query relevance, clustering techniques, etc.

    [Allan 02; Allan, Carbonell, Doddington, Yamron, Yang 98;

    Yang, Pierce, Carbonell SIGIR’98 ;J. Zhang, Yang, Ghahramani, NIPS’04.]
  - Named entities, data or place information, domain knowledge

    [Kumaran and Allan SIGIR’04]
- Temporal Summarization/ Storylines
  - Not seek to cluster “topics” like in TDT but to utilize evolutionary correlations of news coherence/diversity for summarization
    [Yan and Zhang SIGIR’11; Shahaf and Guestrin, KDD 2010; Shahaf, Guestrin, Horvitz, WWW 2012; Allan, Gupta, and Khandelwal, SIGIR’01]

Storyline Construction

Given a corpus C and a query q

1. Get set of relevant sentences, i.e. get $SC(q)$= {sentence s from C | s mentions q.}

2. Resolve dates of events in these sentences: $\forall s \in SC(q)$,

   $date(s)$ = {dates of events regarding q mentioned in s}

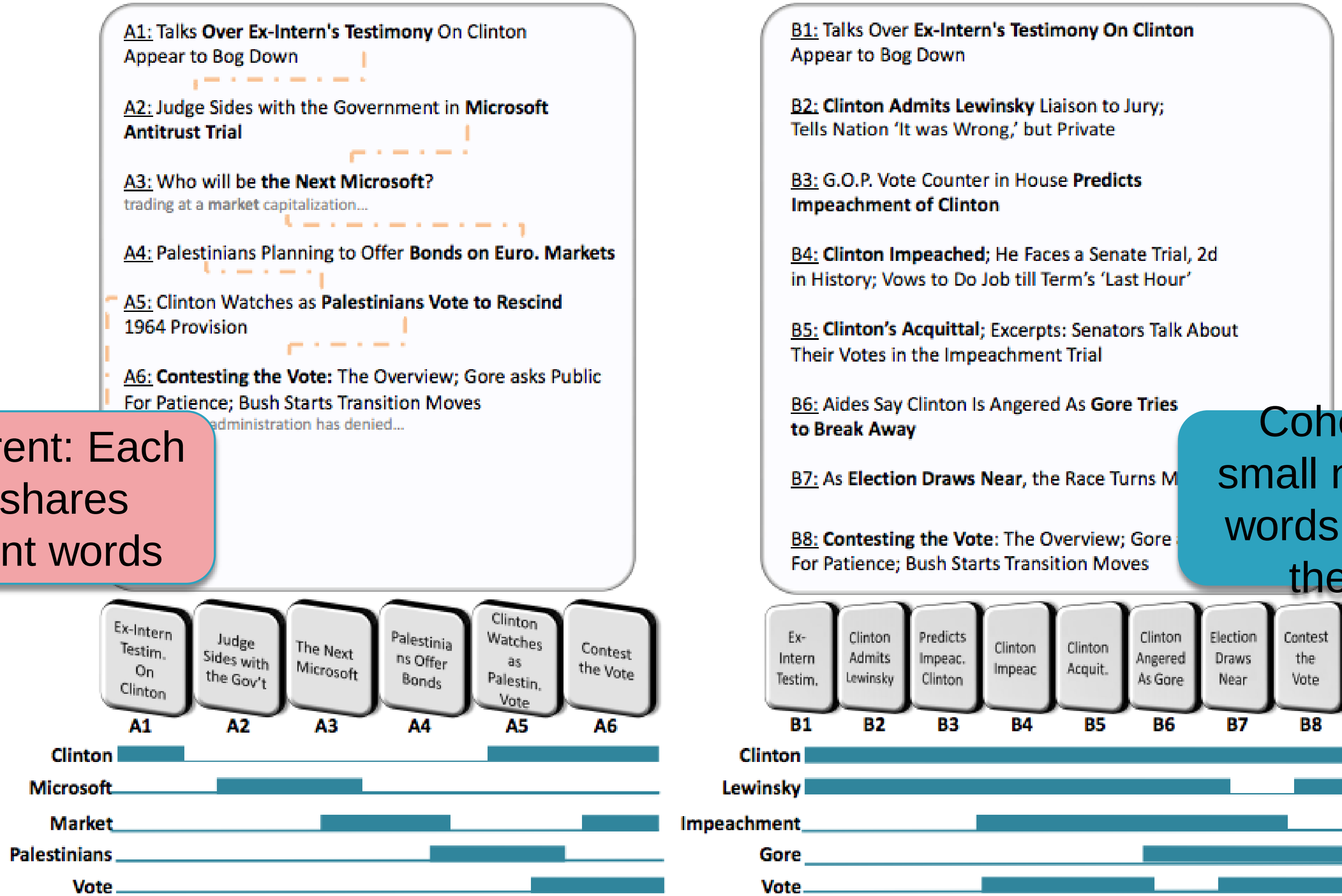
3. Rank the set of sentences

4. Remove duplicate sentences

5. Order top N sentences $\{s_i\}_{1 \le i \le N}$ along a timeline based on $date(s_i)$.

Chieu and Lee SIGIR'04

Good Story Chain (Coherence)



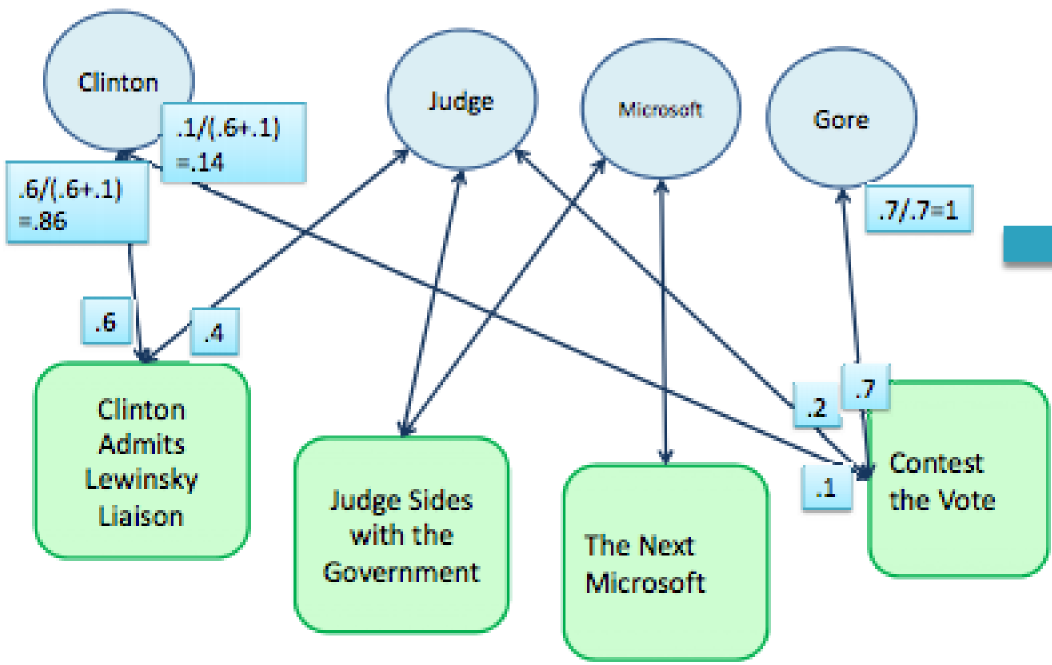
Incoherent: Each pair shares different words

Coherent: a small number of words captures the story

[Shahafand Guestrin, KDD 2010]

Good Story Chain (Word Influence)

Take into consideration the influence of document di to di+1
through the word w. High if:
(1) the two documents are highly connected, and (2) w is important for the connectivit



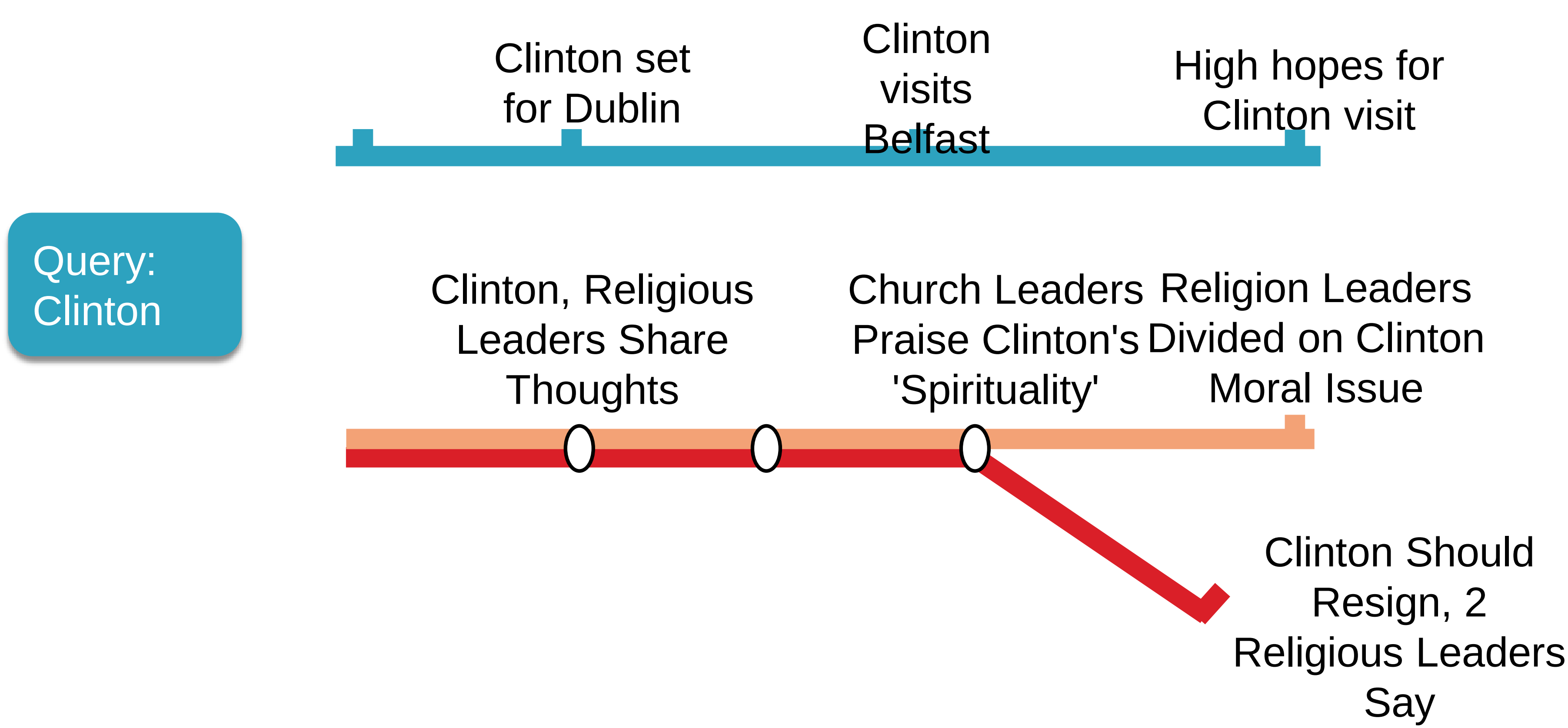
$$Coherence(d_1, ..., d_n) = \max_{activations} \min_{i=1...n-1} \sum_w Influence(d_i, d_{i+1} \mid w) \mathbb{1}(w \ active \ in \ d_i, d_{i+1})$$

Linear Programing Problems

[Shahafand Guestrin, KDD 2010]

Good Multiple Story Chains

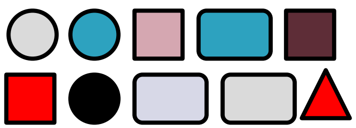
Consider all **coherent** maps with maximum possible **coverage**. Find the most **connect**

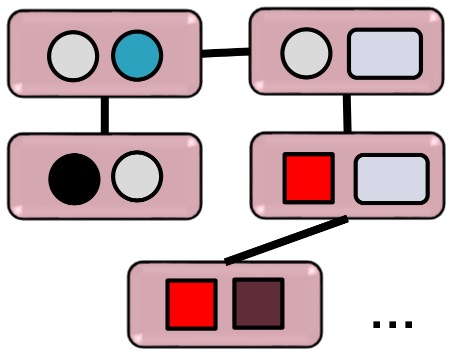


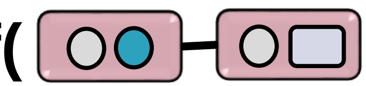
Shahaf, Guestrin, Horvitz: Trains of thought: generating information maps. WWW

Good Multiple Story Chains

Documents D **1. Coherence graph G** **2. Coverage function f**





f() = ?

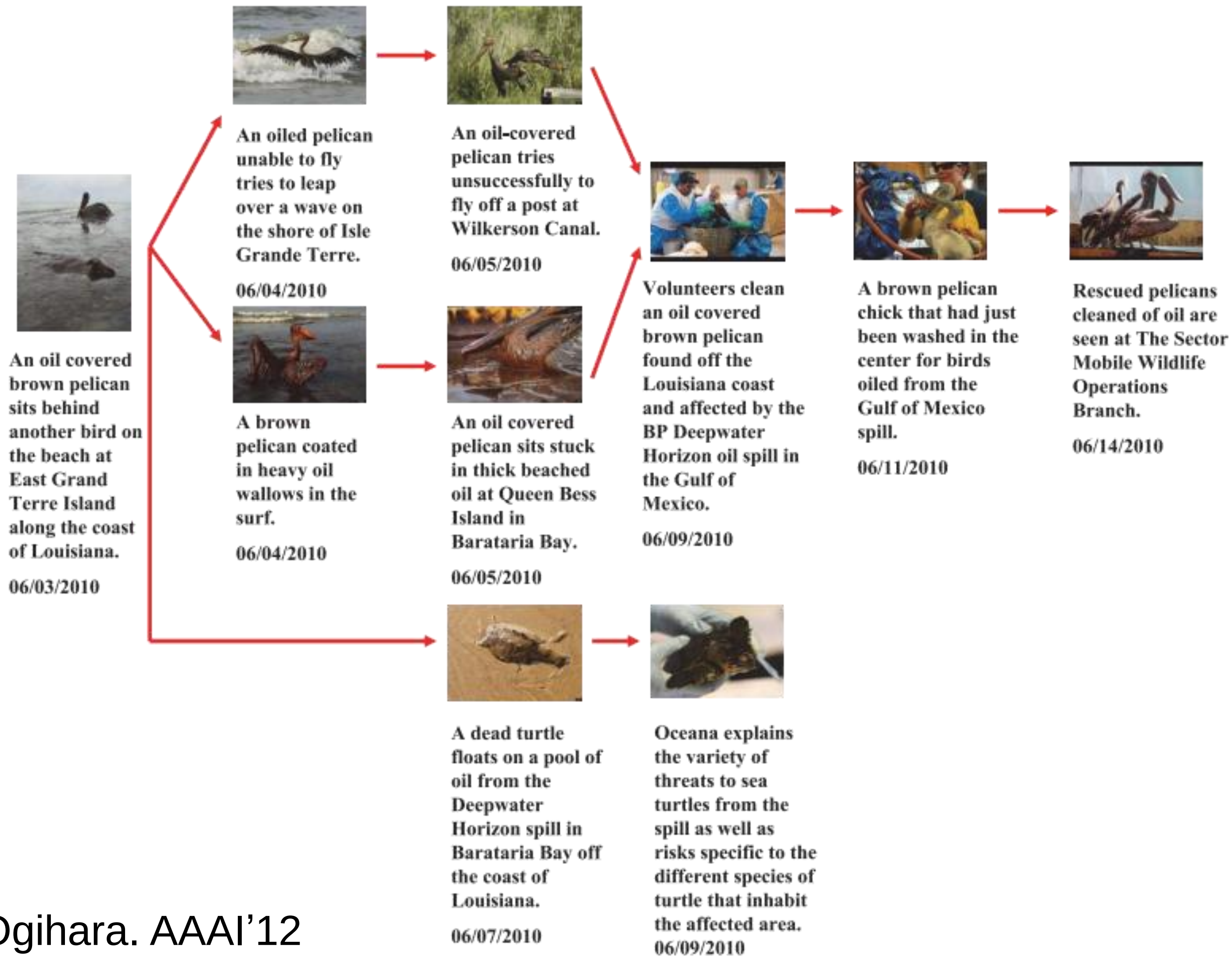
Submodular orienteering [Chekuri & Pal, 2005]
Quasipoly time recursive greedy
O(log OPT) approximation

3. Increase Connectivity

Encodes all m-coherent chains as graph paths

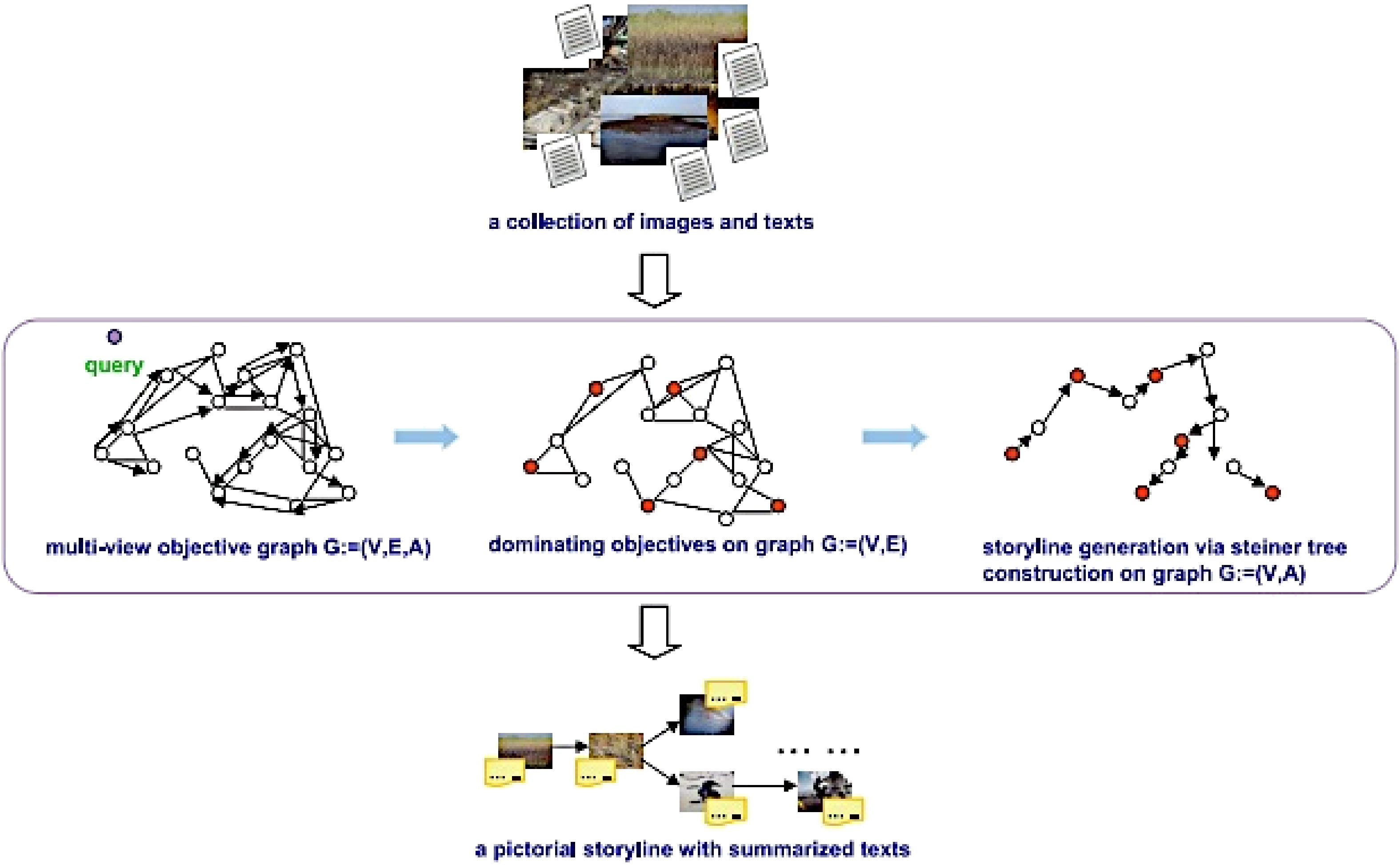
Shahaf, Guestrin, Horvitz: Trains of thought: generating information maps. WWW

Timelines With Images



Wang, Li, Ogihara. AAAI’12

Timelines with Images



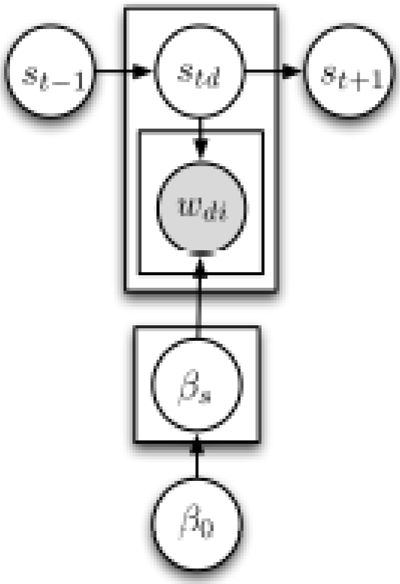
Wang, Li, Ogihara. AAAI'12

Online Timeline creation

- A. Ahmed, Q. Ho, J. Eisenstein, E. Xing, A. J. Smola, and C. H. Teo. Unified analysis of streaming news. In Proc. of WWW, 2011.
- J. Kleinberg. Bursty and hierarchical structure in streams. In KDD, 2002.
- J. Kleinberg. Temporal dynamics of on-line information systems. Data Stream Management: Processing High-Speed Data Streams. Springer, 2006.
- L. Yao, D. Mimno, and A. McCallum. Efficient methods for topic model inference on streaming document collections. In KDD, pages 937–946, 2009.

Online Clustering Model: Recurrent Chinese Restaurant Process

t	time
d	document
(d, i)	position i in document d
s_d	story associated with document d
w_{di}	word i in document d
β_s	word distribution for story s
β_0	prior for word distributions



For each time period $t \in \{1, \ldots, T\}$ do

For each document d in time period t do

i. Draw the storyline indicator: s_{td} via $s_{td} | \mathbf{s_{1:t-1}}, \mathbf{s_{t,1:d-1}}$

ii. If s_{td} is a new storyline draw a distribution over words $\beta_s | \beta_0$

iii. For each i in document draw $w_{di} \sim \beta_{s_{td}}$

$$P(s_{td} | \mathbf{s_{1:t-1}}, \mathbf{s_{t,1:d-1}}) \propto \begin{cases} m'_{ts} + m^{-td}_{ts} & \text{existing story} \\ \gamma & \text{new story} \end{cases}$$

$$m'_{st} = \sum_{\delta=1}^{\Delta} e^{-\frac{\delta}{\lambda}} m_{s,t-\delta}.$$

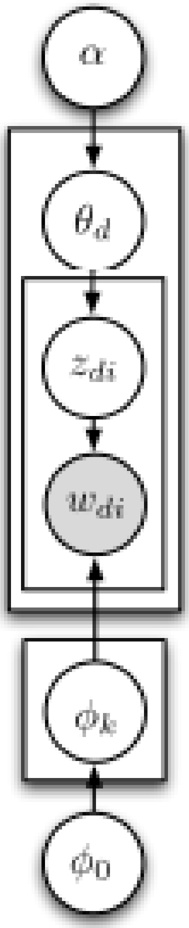
A. Ahmed et al. WWW, 2011.

Topic Model: LDA (reminder)

α	Dirichlet prior over topic distributions
d	document
θ_d	topic distribution for document d
(d,i)	position i in document d
z_{di}	topic associated with word at (d,i)
w_{di}	word at (d,i)
ϕ_0	Dirichlet prior over word distributions for topics
ϕ_k	word distribution for topic k

1. For all topics k do
   (a) Draw word distribution ϕ_k from word prior ϕ_0

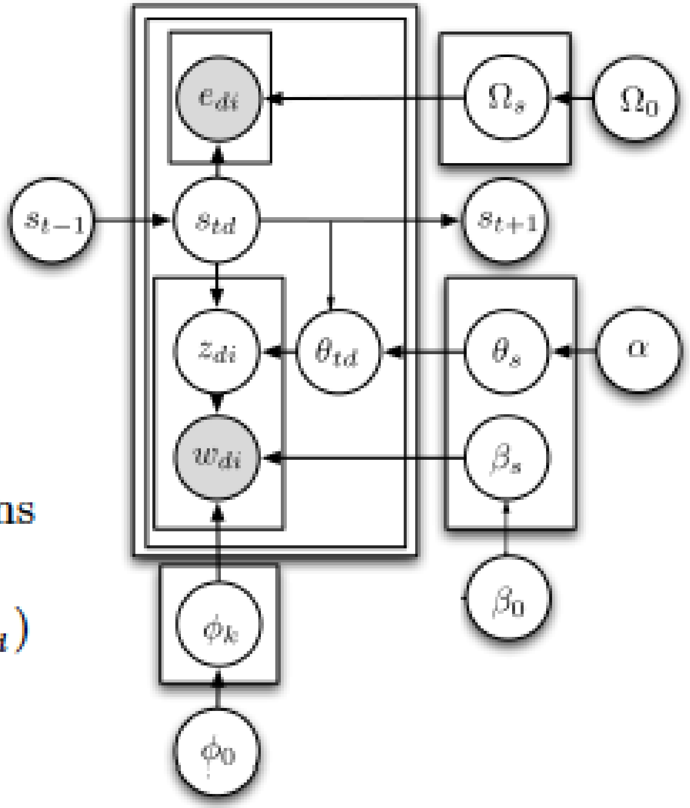
2. For each document d do
   (a) Draw topic distribution θ_d from Dirichlet prior α
   (b) For each position (d,i) in d do
      i. Draw topic z_{di} for position (d,i) from topic distribution θ_d
      ii. Draw word w_{di} for position (d,i) from word distribution $\phi_{z_{di}}$



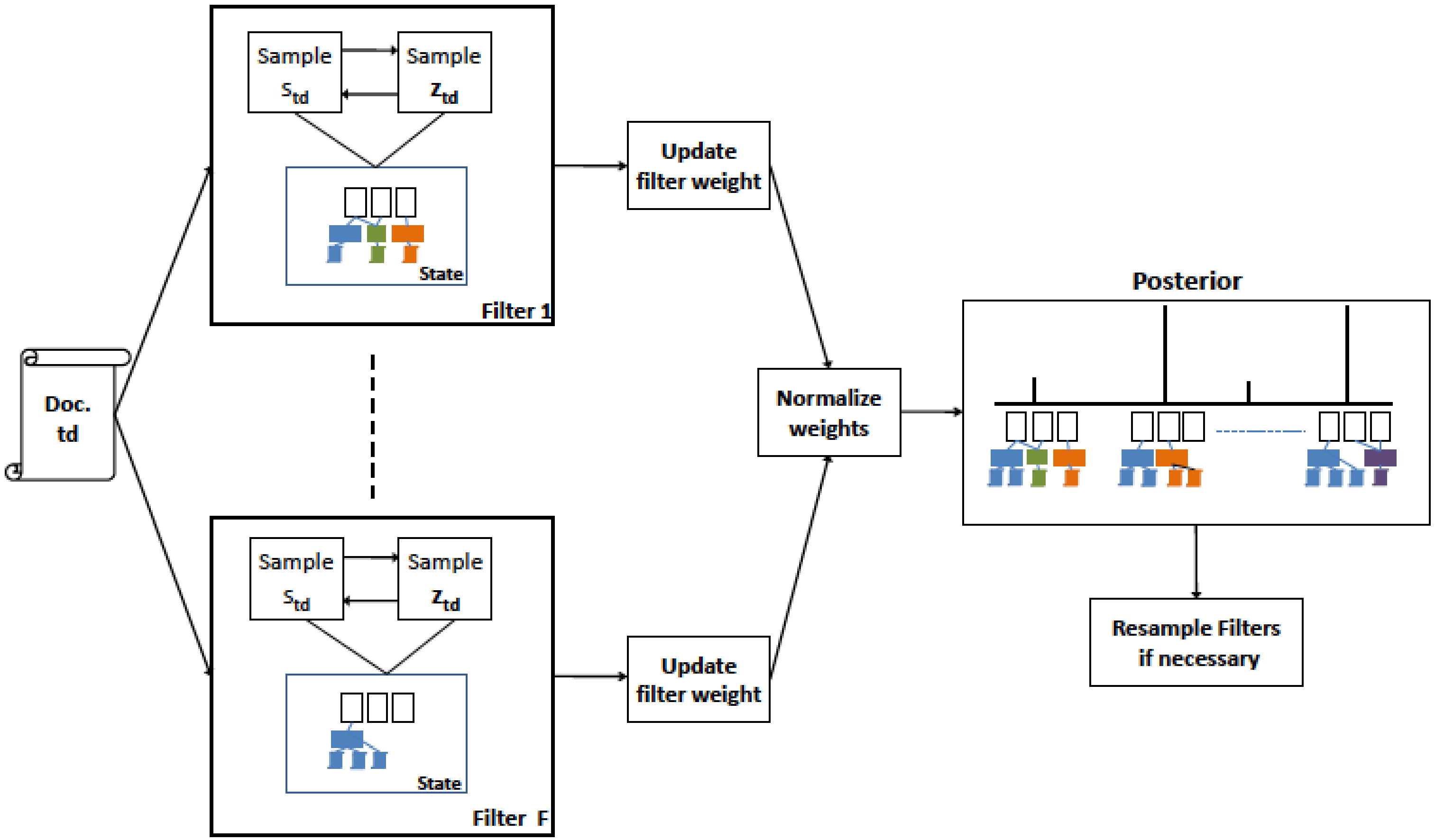
Online Storyline Model

For each time period t from 1 to T do (forward in time)

1. For each document $d \in \{1, \cdots, D_t\}$ in epoch t do
   (a) Draw the storyline indicator:
   $s_{td} | \mathbf{s_{1:t-1}}, \mathbf{s_{t,1:d-1}} \sim RCRP(\gamma, \lambda, \Delta)$
   (b) If s_{td} is a new storyline,
      i. Draw a distribution over words
      $\beta_{s_{\text{new}}} | G_0 \sim \text{Dir}(\beta_0)$
      ii. Draw a distribution over named entities
      $\Omega_{s_{\text{new}}} | G_0 \sim \text{Dir}(\Omega_0)$
      iii. Draw a Distribution over topic proportions
      $\theta_{s_{\text{new}}} \sim \text{Dir}(\alpha)$
   (c) Draw the topic proportions: $\theta_{td} | s_{td} \sim \text{Dir}(\theta_{s_{td}})$
   (d) Draw the words
   $\mathbf{w}_{td} | s_{td} \sim \text{LDA}\left(\theta_{s_{td}}, \{\phi_1, \cdots, \phi_K, \beta_{s_{td}}\}\right)$
   (e) Draw the named entities
   $\mathbf{e}_{td} | s_{td} \sim \text{Mult}(\Omega_{s_{td}})$



Inference: Particle Filtering



Time-sensitive Search & Recommendation

WSDM 2013 Tutorial

Outline

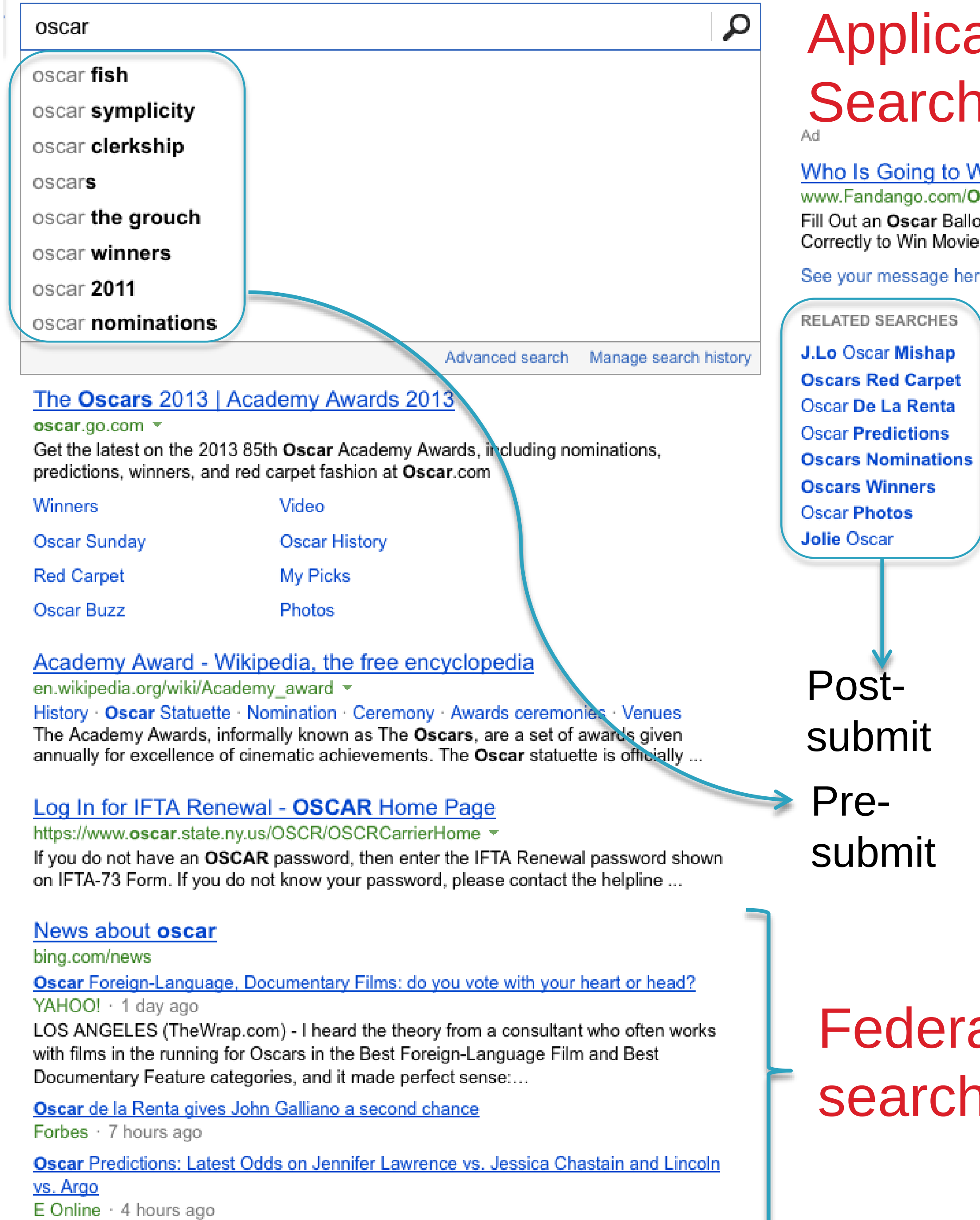
- Modeling Dynamics
  - Web content dynamics [Susan]
  - Web user behavior dynamics [Milad]
  - Spatio-temporal Analysis [Fernando]
  - Methods for evaluation
- Applications to Information Retrieval
  - NLP [Kira]
  - News event prediction [Kira]
  - Time-sensitive search [Dong/Chang]
  - Recommendations [Dong/Chang]

Outline

- Time-sensitive search
  - Time-sensitive ranking relevance
  - Time-sensitive query suggestion
  - Federated search
- Time-sensitive recommendation

SERP

Applications on Search



Ranking (1)

Post-submit

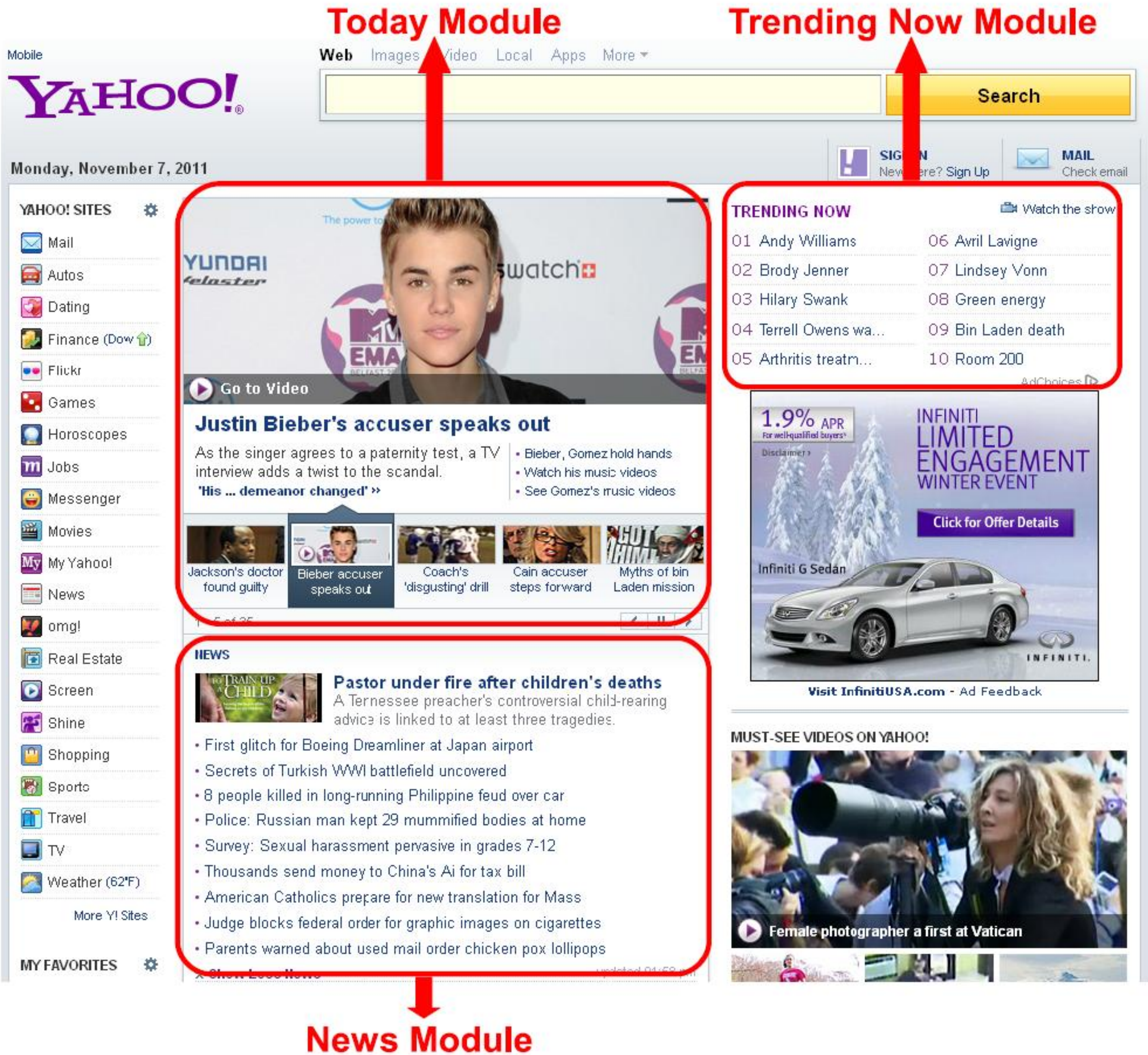
Pre-submit

Query suggestion (2)

Federated search (3)

Portal

Applications on Recommendation



Outline (Anlei Dong and Yi Chang)

- Time-sensitive search
  - Time-sensitive ranking relevance
  - Time-sensitive query suggestion
  - Federated search
- Time-sensitive recommendation

Applications of Time-Sensitive Ranking

- Also called time-aware ranking, recency ranking
- Web search
- Vertical search
  - News search
  - Video search
  - Blog search
  - E-commerce search
  - … …

Problem

- Ranking relevance
  - Topical relevance
  - Authority/popularity/Spam
  - Freshness
  - Local
  - Revenue
  - … …
- How to appropriately combine these factors?
  - Freshness + other relevance

Traditional relevance

Outline for Time-Sensitive Ranking Relevance

- Rule-based approaches
- A learning-to-rank practice
- Leverage Twitter data for improvement
- Joint optimization for relevance and freshness
- Further study: user behavior data

Yearly Recurrent queries

- “WSDM”, “SIGIR”, “Christmas”, “Black Friday”, etc
- Possible solution: query re-writing
  - Solution 1: by query expansion
    - For example, from query “sigir” to “sigir 2009” but
    - Will change query intention, and
    - www.sigir.org better than www.sigir2009.org
  - Solution 2: Double search
    - Use original query, sigir, search first
    - Use query expansion, sigir 2009, search second
    - Then blending two results. BUT
    - Capacity problem and blending algorithm

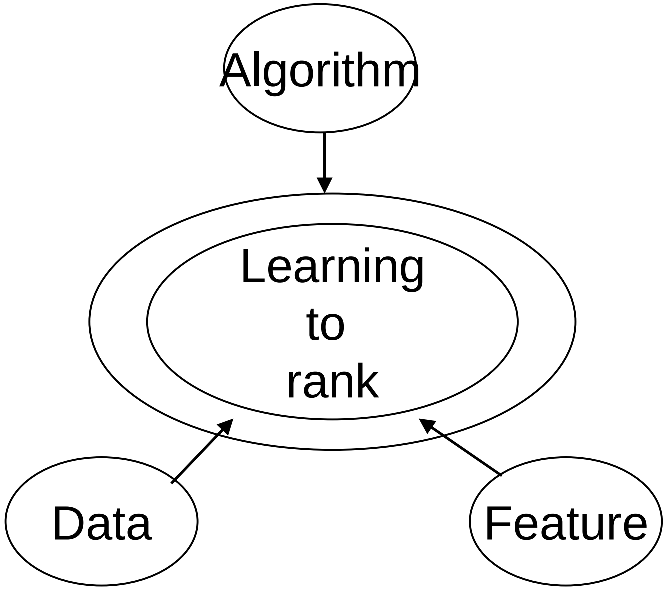
Another Simple Formula

- Combine relevance and freshness by a heuristic rule
  - exponential time-decay rule: $$\textbf{score}_{rank} = \textbf{score}_{relevance} e^{-\beta t}$$

    e.g., [Del Corso, WWW2005]
- Advantage
  - Little training data; fast product delivery;
  - Reasonably good ranking result in practice
- Disadvantage
  - Far from optimal

Learning-to-Rank Solution

- Learning-to-rank: please check the tutorial [Liu WWW09]
- A standard approach



Main Challenges

- Feature Challenges
  - Precise time-stamp for each URL is hard to get
  - Little click information for a fresh URL
  - Few anchor texts for a fresh URL
- Data Challenges
  - Crawling Challenge
  - Labeled data collection challenge
  - Appropriate evaluation metrics
- Ranking Algorithm Challenges
  - Traditional Ranking is poor, since fresh documents lack link or click information
  - Merge different sources of results into 1 ranking

Data: Editorial Label

- Traditional data label:
  - <query, URL> ← ? {perfect, excellent, good, fair, bad}
- Incorporate time:
  - <query, URL, query_time>

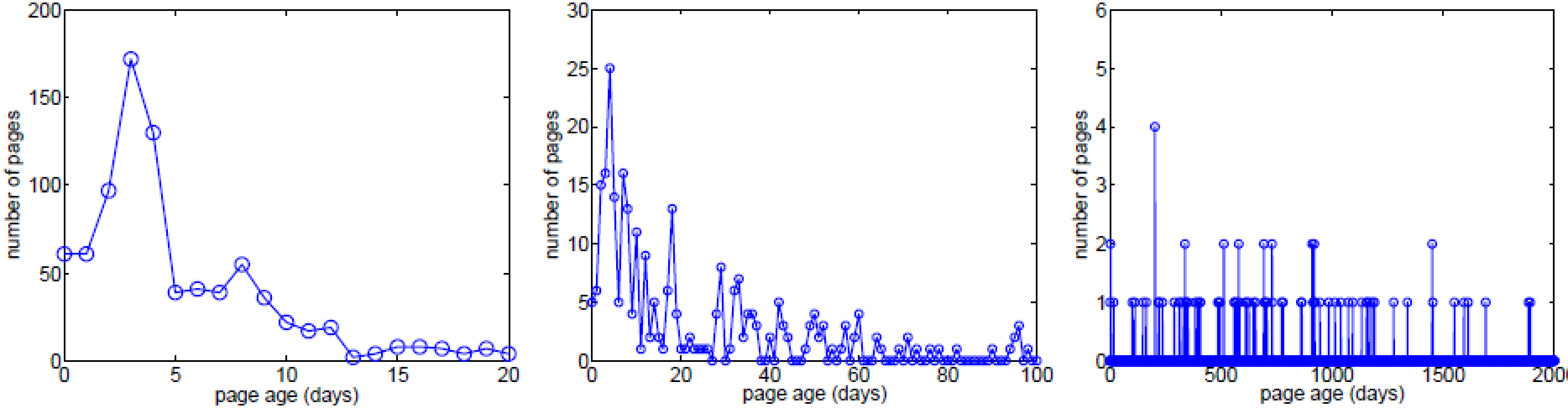
    ← relevance ? {perfect, excellent, good, fair, bad}

    ← freshness ?{latest, ok, a little bit old, totally outdated}
- Learning target:
  - Combine labels by relevance and freshness
  - For example: recency promotion/demotion: {+1, 0, -1, -2}

[Dong, WSDM01]

Freshness: Judge vs. Age

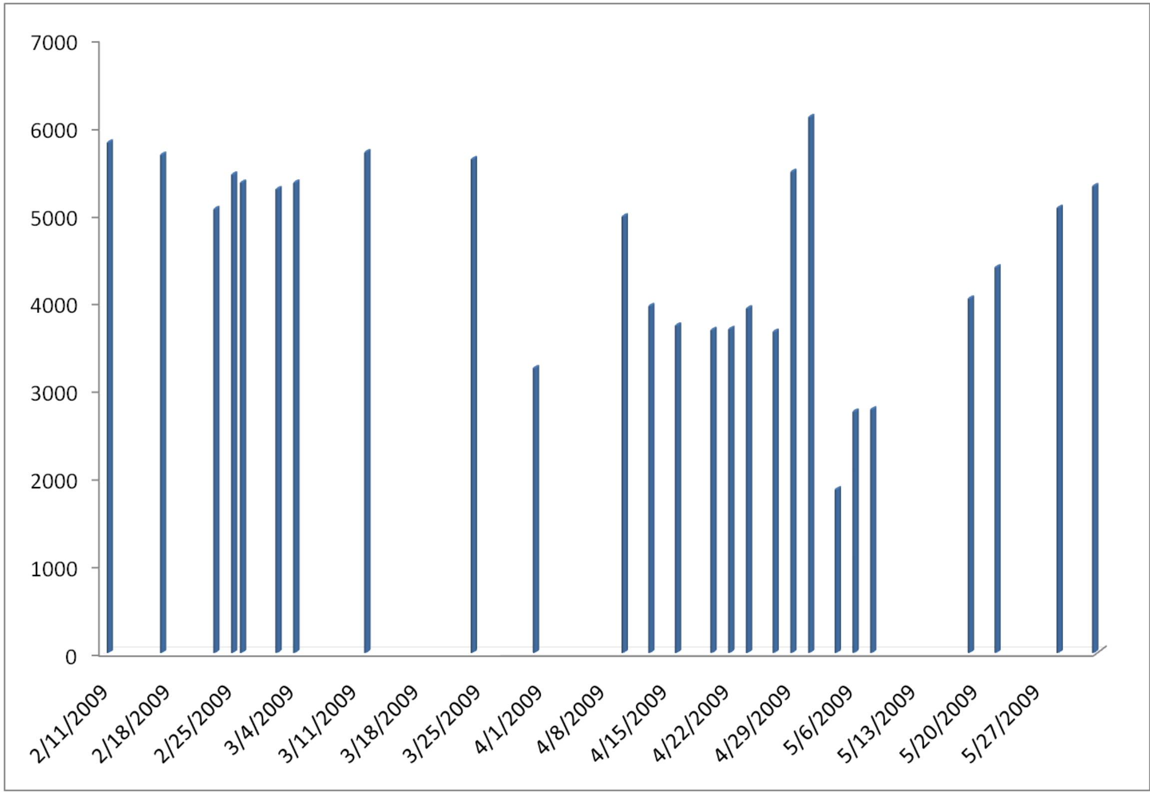
- Subjective vs. objective



(a) no demotion (b) 1-grade demotion (c) 2-grade demotion

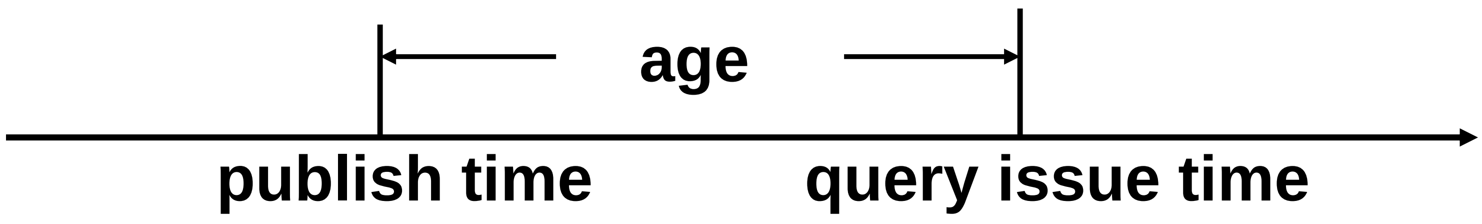
Data: Editorial Data Collection

- Need to collect data periodically
  - Avoid distribution bias
  - Judge immediately



Feature

An ideal case:



But most pages do not have an accurate time!

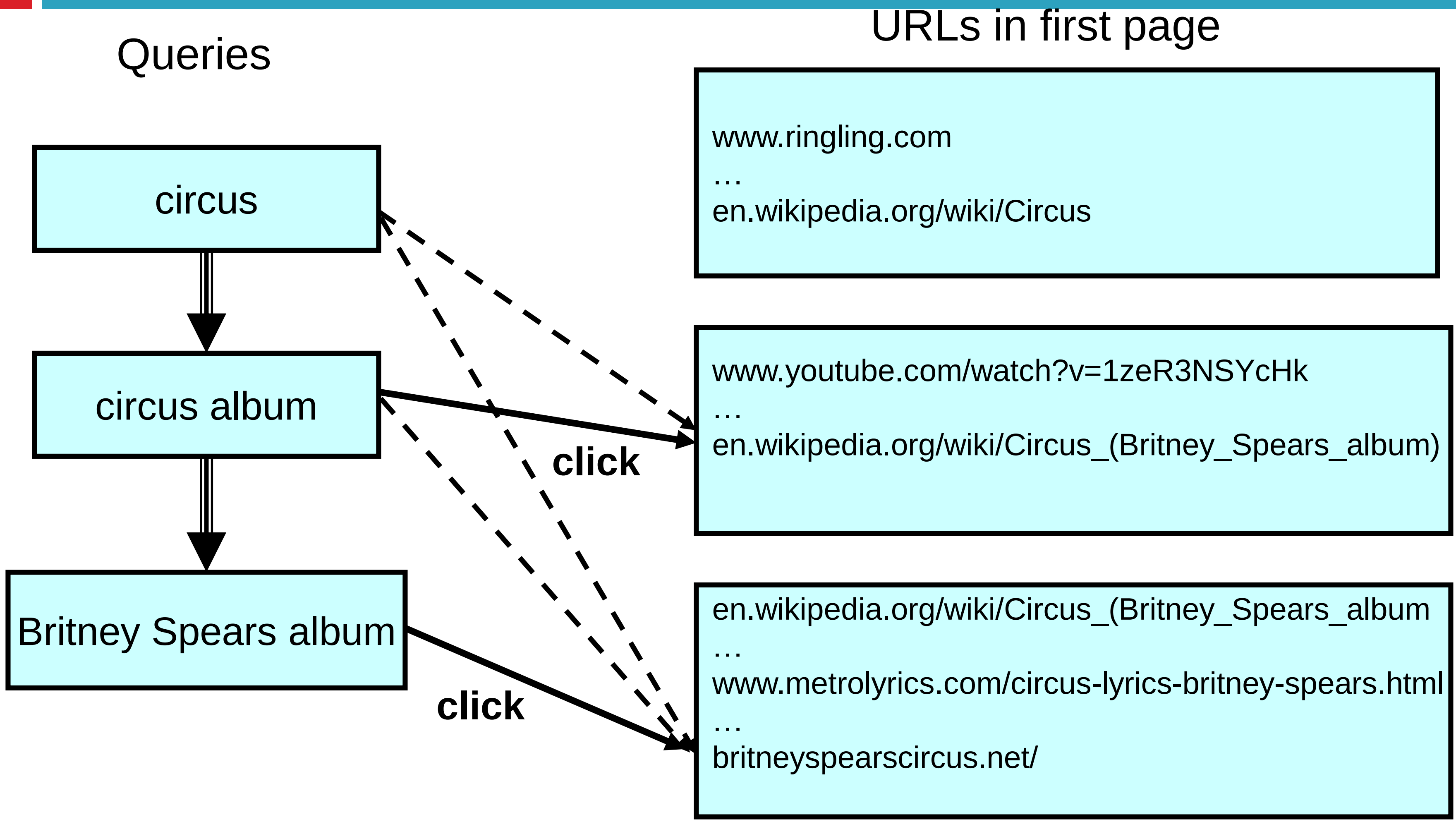
- Some intuitive features
  - Timestamp feature
  - Discovery time feature
  - Query time-sensitivity feature
  - Page classification feature

Click Feature

- Challenge: limited clicks on fresh URLs
- Solution:
  - User may issue a chain of queries for the same information: queries in the chain are strongly related.
  - Use query chains to “smooth” clicks.

[Inagaki AAAI10]

Extend Clicks



Solid arrows : real clicks
Dotted arrows: inferred clicks from query chain

Time-Weighted Click Features

- Recent clicks must be weighted more
  - The shift of user intent must be taken into consideration
  - e.g., should we still rank B. Spears' "Circus" on the top for the query "Circus" after 12 months?
- Time-weighted CTR
  - i refers to day; x is used to control time decay

$$CTR^{w}(q,u,t_q) = \frac{\sum_{i=1,v_i>0}^{t_q} c_i (1+x)^{i-t_q}}{\sum_{i=1,v_i>0}^{t_q} v_i (1+x)^{i-t_q}}$$

Click Buzz Feature

- CTR change over time
  - Compute average CTR_{avg} over a period of time and standard deviation σ
  - BUZZ at a given day is
    - $(CTR_t - CTR_{avg}) / \sigma$
  - Represent how unusual the current CTR is with respect to “normal” CTR for that URL.

Modeling: Leverage Regular Data

- Premise of improving recency
  - Overall relevance should not be hurt!
- Recency training data
  - small amount of query-urls -> Poor relevance
- Regular training data
  - huge amount of query-urls -> Good relevance
- Solution
  - Utilize regular data or model to help recency ranking

Combine Relevance and Recency Data

	Data	Features	Modeling algorithm
Dedicated model	Recency data	Recency features + regular features	GBrank
Over-weighting model	Recency data + Regular data	Recency features + regular features	GBrank
Compositional model	Recency data	Recency features + ranking score	GBrank
Adaptation model	Recency data	Recency features + regular features	1. Regular model as base model 2. Do adaptation

[Dong WSDM10]

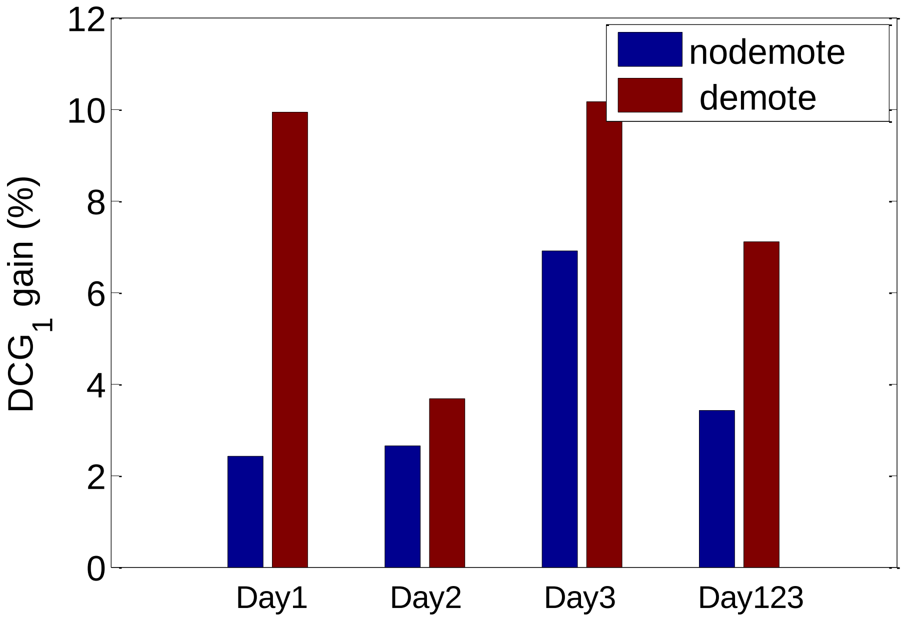
Model Adaptation

Motivation: solve data scalability issues
expensive to have high quality training data for each market/tas

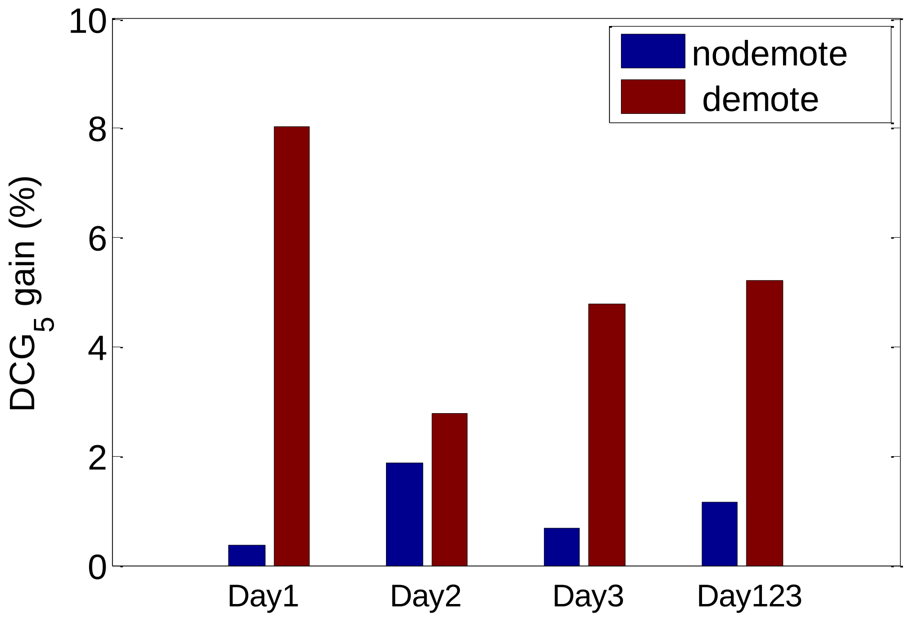
Background:
- Model adaptation is one approach of transfer learning
- Goal: transfer knowledge learned from task (A) → task (B)
- Assumption: there is similarity between A and B

Approach:
- Train a base model A (using Data A)
- Modify model A using Data B → Model A'
- Apply adapted model A' to task B

Online Over-Weighting Results



DCG1



DCG5

Query Classification vs. Query feature

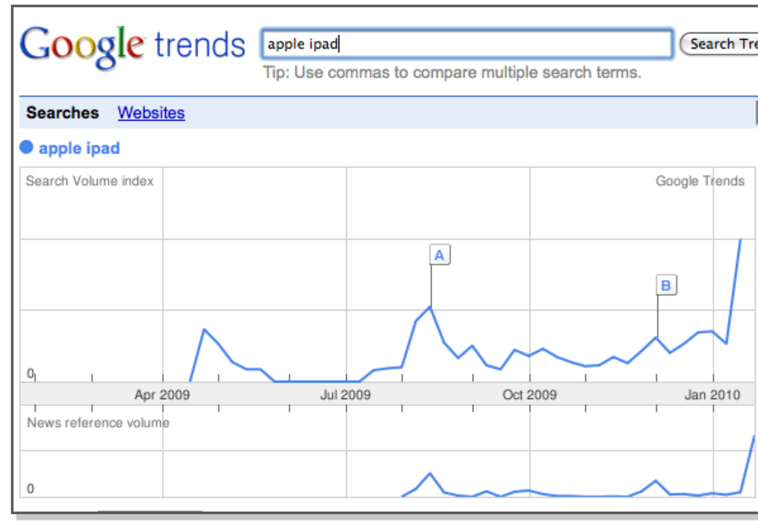
- Approach 1: query classification
  - Step1. determine query type;
    - Breaking-news query? Yearly-recurrent query? … …
  - Step 2. apply corresponding ranking model
  - Divide-and-conquer strategy
  - Effective and straightforward in practice
- Approach 2: query feature
  - A single unified model for all queries
  - E.g. [Dai SIGIR11]

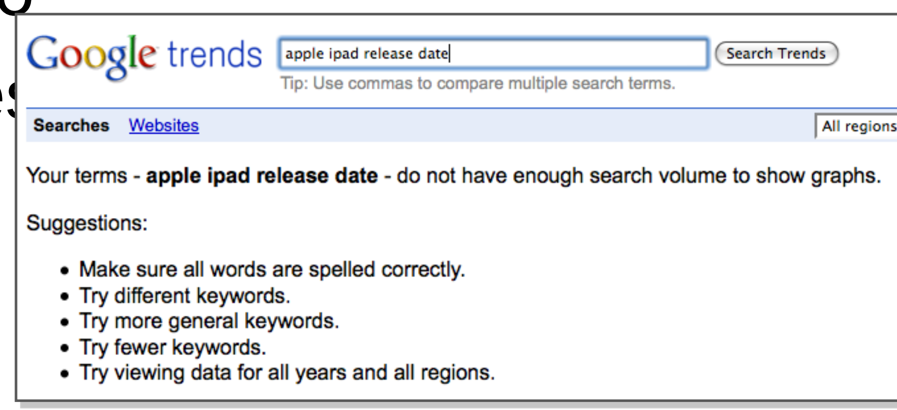
Query Classifier

- Identify, in near real-time, queries about emerging events and news stories
  - E.g., natural disasters; major sport events; latest celebrity gossip; political breaking stories; etc.

Query Classifier

- Standard approach:
  - Maintain temporal model for each query
  - Identify irregularities in model e.g., change in moving average of more than $n\sigma$
  - work well for head queries not so for torso/tail queries





One New Approach

- Rather than maintain a model for each query, maintain a model of each slot of time
- Given a query, determine whether it is predicted by recent models better than by earlier ones
- In practice:
  - Time slot modeling: n-gram language models
  - Model prediction: language model generation likelihood

Compute "Buzziness"

- Approach
  - Reference models, r_i = *{prev_day, prev_week, prev_month}*
  - Language model settings: interpolated bigram model
- Score computation, using Query model

$$\mathrm{buzz}(q, t, Q) = \max_i P(q|M_{Q,t}) - P(q|M_{Q,t-r_i})$$

Content Model

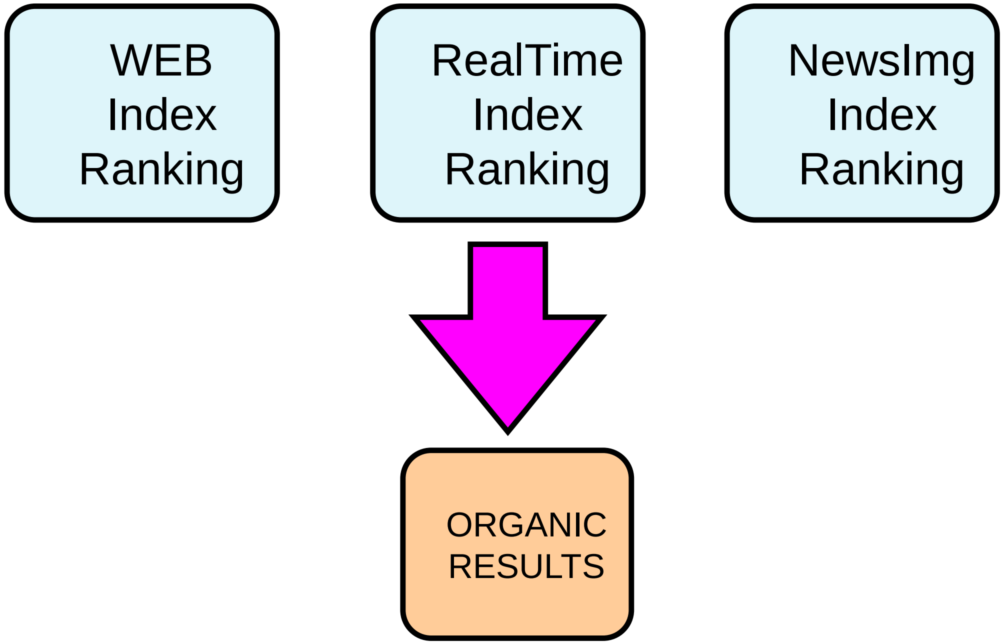
- Not all current events reflected in the query log
- In addition to tracking the query log, we track news headlines from Yahoo! News
  - Top viewed: U.S., Business, World, …
  - RSS feeds updated every 30 minutes
  - Content used for building similar time-slotted LMs

Score, blending Query and Content models:

$$\mathrm{buzz}(q,t) = \lambda_1 \cdot \mathrm{buzz}(q,t,Q) + \lambda_2 \cdot \mathrm{buzz}(q,t,C)$$

Data Blending

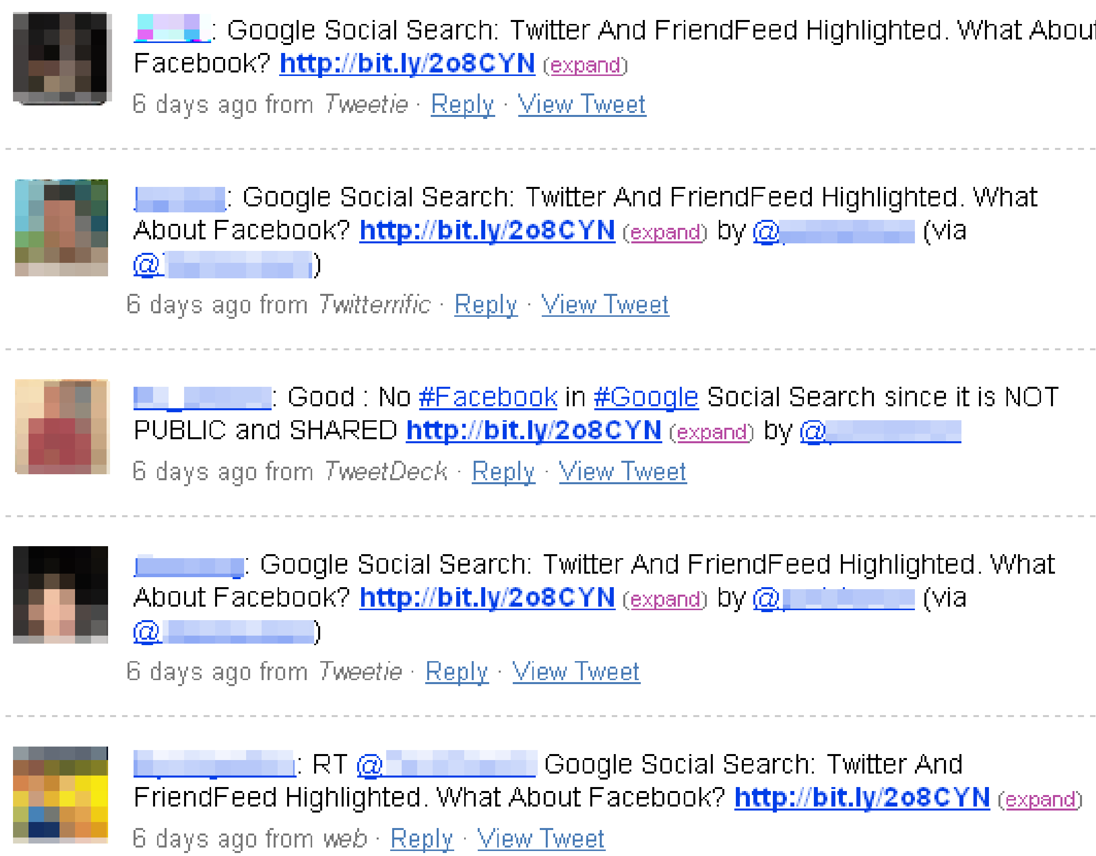
Results from 2+ scoring functions



Single organic result list that maximize relevance

Incorporate Twitter Data to Improve Real-Time Web Search

- To improve Web Search Ranking, not Twitter Search
- Micro-blogging
- Twitter
  - Tweet
  - Twitter User
  - Twitter Tiny URL
    - (Twitter URL)
  - Following Relationshi



[Dong WWW10]

Question

- Can we make use of Twitter to improve real-time crawling?
- Can we utilize Tweets to improve Twitter Tiny URL ranking?
- Can we use social network of Twitter users to improve Twitter Tiny URL ranking?

Motivation

- Twitter Tiny URL contains news/non-news URL, and Twitter Tiny URL could represents diverse and dynamic browsing priority of users;

- The social network among Twitter users data could provide a method to compute popularity of twitter users, and authority of fresh documents;

- Tweets could be leveraged as an extended representation of Twitter Tiny URL;

Crawling Strategy

- Exhaustive crawling strategy for fresh content in real-time is difficult;
- Select high quality Twitter Tiny URL as crawling feeds;
- Twitter Tiny URL could reflect diverse and dynamic browsing priority of users;
- Human intelligence is incorporated into the real-time crawling/indexing system.

Crawl Twitter Tiny URL

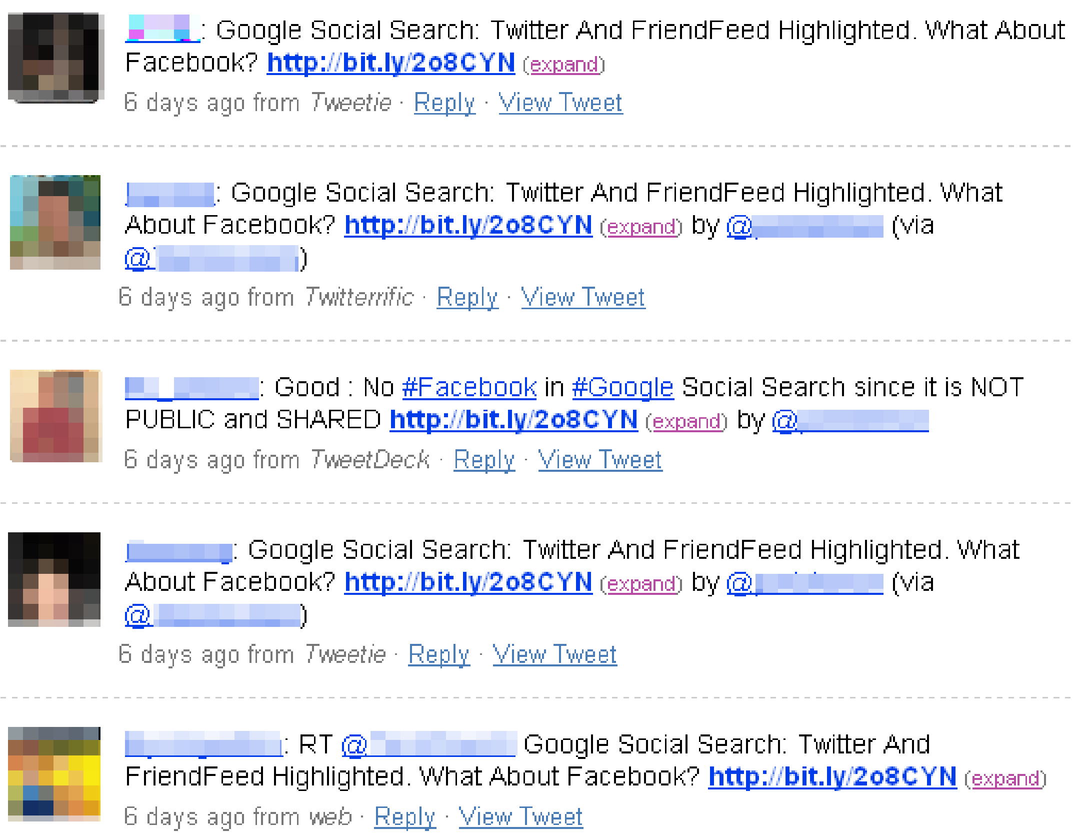
- Majority of Twitter Tiny URL are poor quality
  - Spam, Adult, Self-promotion, etc.
- A set of simple heuristic rules
  - Discard Tiny URL referred by the same Twitter user more than 2 times;
  - Discard Tiny URL only referred by one Twitter user.
- Experiment
  - Based on 5 hour twitter data,
  - about 1 Million Tiny URL,
  - After filtering with the rule, 5.9% high quality Tiny URL remaining.

Twitter Feature

- Text Matching between Query and Tweet
  - Cosine Similarity
  - Exact Matching
  - Proximity Matching
    - Overlapping Terms
    - Extra Terms
    - Missing Terms
  - User Authority Weighted Proximity Matching

Textual Features between Query and Tweet

- Tweets would be a substitute of Anchor Text in real-time.



Social Network Features

- Represent Twitter User as a social network
  - A Vertex represents a Twitter User
  - An Edge represents the follower relationship
  - Apply the PageRank idea
    - The popularity of Twitter Users are generated when it converge.
    - The popularity information is used to update User Authority Weighted Proximity Matching.

Other Features

- Given a Tiny URL, other URL based features include:
  - Average Count Features of the users refer the Tiny URL;
  - Count Features related to the 1st Twitter user refer to the Tiny URL;
  - Count Features related to the most popular Twitter User refer the Tiny URL
  - Count Features
    - # of followers for this user;
    - # of followings for this user;
    - # of posts by this user;
    - # of users retweet the Tiny URL;
    - # of users reply the Tiny URL;

Ranking Strategy

	Data	Features
MLR for Regular URLs	**Regular data**	**Content features** **+ Aggregate Features**
MLR for Twitter URLs	**Twitter** **(Regular) data**	**Content features** **+ Twitter features**

Different Ranking Models

$\mathbf{D}_{\text{regular}}$: training data set from regular data
$\mathbf{D}_{\text{Twitter}}$: training data set from Twitter data

$$\mathcal{M}_{\text{regular}} \leftarrow \textsc{Train-MLR}(\mathbf{D}_{\text{regular}}, \{\mathbf{F}_{\text{content}}, \mathbf{F}_{\text{aggregate}}\})$$
$$\mathcal{M}_{\text{Twitter}} \leftarrow \textsc{Train-MLR}(\mathbf{D}_{\text{Twitter}}, \{\mathbf{F}_{\text{content}}, \mathbf{F}_{\text{Twitter}}\})$$
$$\mathcal{M}_{\text{content}} \leftarrow \textsc{Train-MLR}(\mathbf{D}_{\text{regular}}, \mathbf{F}_{\text{content}})$$
$$\mathbf{y}_{\text{Twitter}} \leftarrow \textsc{Predict}(\mathbf{D}_{\text{Twitter}}, \mathcal{M}_{\text{content}})$$
$$\mathcal{M}_{\text{composite}} \leftarrow \textsc{Train-MLR}(\mathbf{D}_{\text{Twitter}}, \{\mathbf{y}_{\text{Twitter}}, \mathbf{F}_{\text{Twitter}}\})$$

- MLR Model is trained with Gradient Boosted Decision Tree (GBDT) Algorithm.

Rationale of Each Model

MLR + Blending		Advantage & Disadvantage
For Regular URL	**For Twitter URL**	
MRegular	**MRegular**	Favor regular URL, unfavor Twitter URL
MContent	**MContent**	Favor Twitter Tiny URL, unfavor regular URL
MRegular	**MContent**	Twitter Tiny URL will not get promoted
MRegular	**MTwitter**	Tiny URL will be promoted, but relevance of Tiny URL might not be fully leveraged
MRegular	**MComposite**	Tiny URL will be promoted, but relevance of Tiny URL might be

Ranking Result

MLR + Blending		NDCG5	NDCF5	NDCG5 + Recency Demotion
Regular URL	**Twitter URL**			
MRegular	**MRegular**	0.681	0.518	0.666
MContent	**MContent**	0.682 (+0.3%)	0.587 (+11.7%)	0.652 (-2.1%)
MRegular	**MContent**	0.690 (+1.3%)	0.569 (+8.9%)	0.680 (+2.1%)
MRegular	**MTwitter**	0.729 (+6.5%)	0.736 (+29.6%)	0.739 (+9.9%)
MRegular	**MComposite**	0.723 (+5.8%)	0.756 (+31.4%)	0.735 (+9.4%)

Main Findings

- Twitter did contain high quality Tiny URL, which is relevant to some time sensitive queries;
- The text of Tweets can be used to substitute anchor text for those real-time relevant documents;
- The social network of Twitter users can be used to improve ranking.

Simultaneously Optimize Freshness and Relevance

- [Dai SIGIR11]
- Criteria-sensitive divide-and-conquer ranking
  - Multiple rankers corresponding to different query categories
  - Train each ranker by

$$f_i^* = \arg\min_{f_i} \sum_{q \in \mathcal{Q}} \mathcal{I}(q, i) \mathcal{L}_i(\hat{\mathbf{y}}_q, \mathbf{y}_q)$$

Q: training query set;

$I(q, i)$: importance of query q with respect to the ith ranked model

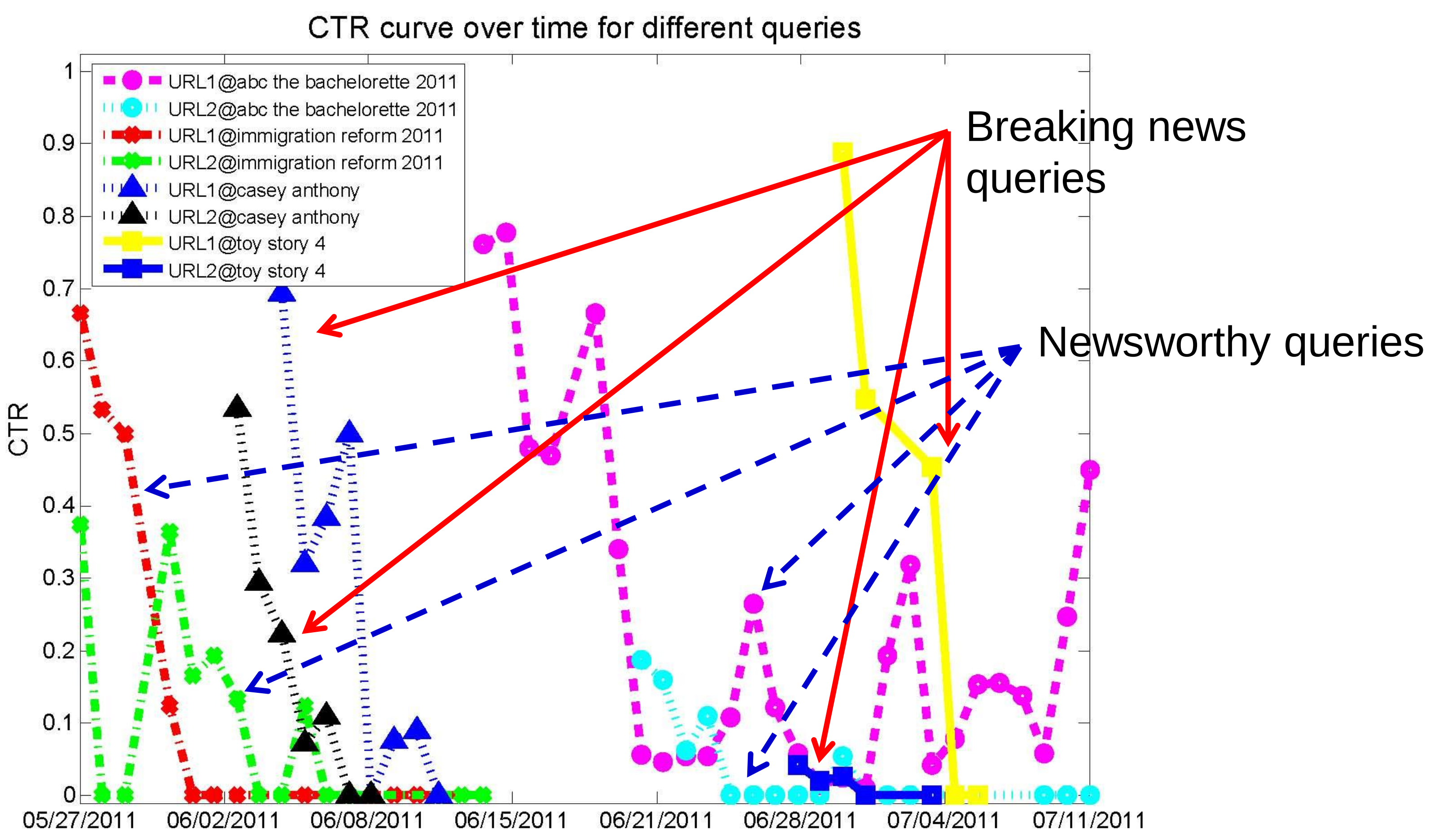
Study User Behavior

- Relevance
  - Topical relatedness
  - Metric: tf*idf, BM25, Language Model
- Freshness
  - Temporal closeness
  - Metric: age, elapsed time
- Trade-off
  - Serve for user's information need

Understand User's Information Need

- User's emphasis on relevance/freshness varies
  - Breaking news queries
    - Prefer latest news reports – freshness driven
    - E.g., "apple company"
  - Newsworthy queries
    - Prefer high coverage and authority news reports – relevance driven
    - E.g., "bin laden death"

Relevance/Freshness Varies



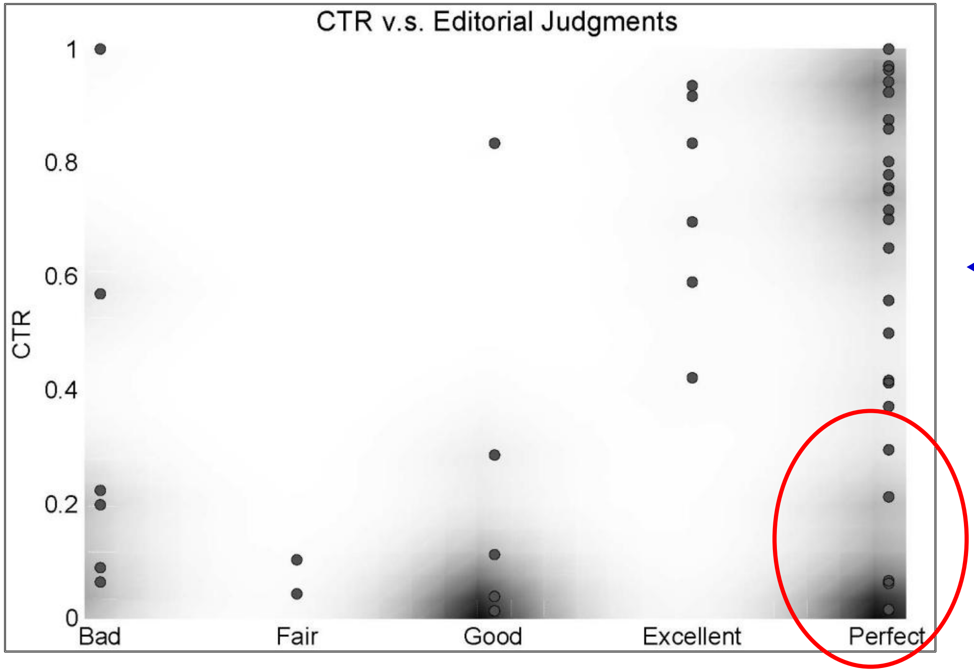
[Wang WWW10]

Access User's Information Need

- Unsupervised integration [Efron SIGIR11, Li CIKM03]
  - Limited on timestamps
- Editor's judgment [Dong WSDM10, Dai SIGIR11]
  - Expensive for timely annotation
  - Inadequate to recover end-user's information need

Editor's Annotation

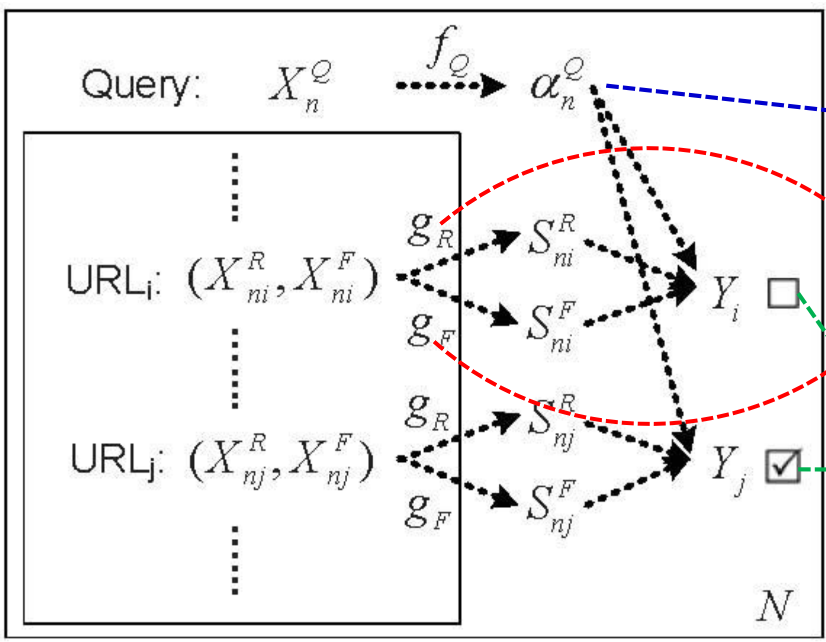
- Freshness-demoted relevance
  - Rule-based hard demotion [Dong WSDM10]
    - E.g., if the result is somewhat outdated, it should be demoted by one grade (e.g., from excellent to good)



Correlation:
0.5764±0.6401

Joint Relevance and Freshness Learning

- JRFL: (Relevance, Freshness) -> Click



Query => trade-off

URL => relevance/freshness

Click => overall impression

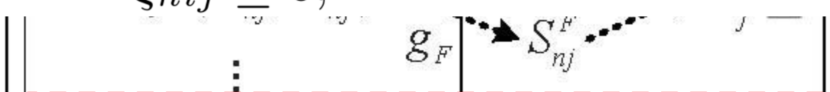
$$Y_{ni} > Y_{nj}$$

Joint Relevance and Freshness Learning

- Model formalization

$$\min_{f_Q,\ g_R,\ g_F,\ \xi} \frac{1}{2}(\|f_Q\| + \|g_R\| + \|g_F\|) + \frac{C}{N}\sum_{n=1}^{N}\sum_{i,j}\xi_{nij}$$

$$\begin{aligned}
\text{s.t.}\quad & \forall(n,i,j), URL_{ni} \succ URL_{nj} \\
& Y_{ni} - Y_{nj} > 1 - \xi_{nij} \\
& 0 \le f_Q(X_n^Q) \le 1 \\
& \xi_{nij} \ge 0,
\end{aligned}$$



$$\boxed{Y_{ni} = \alpha_n^Q \times S_{ni}^F + (1 - \alpha_n^Q) \times S_{ni}^R}$$

Query-specific

Latent

Joint Relevance and Freshness Learning

- Linear instantiation

$$\min_{w_R, w_F, w_Q, \xi} \frac{1}{2}(\|w_Q\|^2 + \|w_R\|^2 + \|w_F\|^2) + \frac{C}{N}\sum_{n=1}^{N}\sum_{i,j}\xi_{nij}$$

$$\begin{aligned}
\text{s.t.} \quad & \forall (n,i,j), U_{ni} \succ U_{nj} \\
& \underline{w_Q^\top X_n^Q} \times \underline{w_F^\top (X_{ni}^F - X_{nj}^F)} \\
& \quad + (1 - w_Q^\top X_n^Q) \times \underline{w_R^\top (X_{ni}^R - X_{nj}^R)} > 1 - \xi_{nij} \\
& 0 \leq w_Q^\top X_n^Q \leq 1 \\
& \xi_{nij} \geq 0.
\end{aligned}$$

- Associative property
  - Relevance/Freshness model learning
  - Query model learning

Temporal Features

- URL freshness features
  - Identify freshness from content analysis

Table 1: Temporal Features for URL freshness

Type	Feature
URL freshness	$\mathbf{age}_{\text{pubdate}}(\text{URL}\mid\text{Query}) = \text{timestamp}(\text{Query}) - \text{pubdate}(\text{URL})$ $\mathbf{age}_{\text{story}}(\text{URL}\mid\text{Query}) = \text{timestamp}(\text{Query}) - \text{pubdate}_{extracted}(\text{URL})$ $\mathbf{LM@1}(\text{URL}\mid\text{Query}, t) = \max_{d \in \text{Corpus(q—t)}[t-1\text{day},t]} \log p(\text{URL}\mid d)$ $\mathbf{LM@5}(\text{URL}\mid\text{Query}, t) = \max_{d \in \text{Corpus(q—t)}[t-5\text{days},t-2\text{days}]} \log p(\text{URL}\mid d)$ $\mathbf{LM@ALL}(\text{URL}\mid\text{Query}, t) = \max_{d \in \text{Corpus(q—t)}[-\infty,t-6\text{days}]} \log p(\text{URL}\mid d)$ $\mathbf{t\text{-}dist}(\text{URL}\mid\text{Query}) = \frac{\mathbf{age}_{\text{pubdate}}(\text{URL}\mid\text{Query}) - \text{mean}[\mathbf{age}_{\text{pubdate}}(\text{URL}\mid\text{Query})]}{\text{dev}[\mathbf{age}_{\text{pubdate}}(\text{URL}\mid\text{Query})]}$

Temporal Features

- Query freshness features
  - Capture latent preference

Table 2: Temporal Features for Query model

Type	Feature
Query Model	$\mathbf{q\_prob}(\text{Query}\vert\text{t}) = \log \frac{Count(Query\vert t)+\delta_q}{\sum_q Count(Query\vert t)+\delta}$
	$\mathbf{u\_prob}(\text{User}\vert\text{t}) = \log \frac{Count(User\vert t)+\lambda_u}{\sum_q Count(User\vert t)+\lambda}$
	$\mathbf{q\_ratio}(\text{Query}\vert\text{t}) = \mathbf{q\_prob}(\text{Query}\vert\text{t}) - \mathbf{q\_prob}(\text{Query}\vert\text{t-1})$
	$\mathbf{u\_ratio}(\text{User}\vert\text{t}) = \mathbf{u\_prob}(\text{User}\vert\text{t}) - \mathbf{u\_prob}(\text{User}\vert\text{t-1})$
	$\mathbf{Ent}(\text{Query}\vert\text{t}) = -p(Query\vert t)\log p(Query\vert t)$
	$\mathbf{CTR}(\text{Query}\vert\text{t}) = \text{mean}\left[\text{CTR(URL}\vert\text{Query, t)}\right]$
	$\mathbf{pub\_mean}(\text{Query}\vert\text{d}) = \text{mean}_{URL\in Corpus(Q\vert t)}\left[\mathbf{age}_{\text{pubdate}}(\text{URL}\vert\text{Query})\right]$
	$\mathbf{pub\_dev}(\text{Query}\vert\text{d}) = \text{dev}_{URL\in Corpus(Q\vert t)}\left[\mathbf{age}_{\text{pubdate}}(\text{URL}\vert\text{Query})\right]$
	$\mathbf{pub\_frq}(\text{Query}\vert\text{t}) = \log \frac{Count(URL\vert d)+\sigma_u}{\sum_{URL} Count(URL\vert t)+\sigma}$

(δ_q, δ), (λ_u, λ) and (σ_u, σ) are the smoothing parameters estimated from the query log.

Experiments

- Data sets
  - Two months' Yahoo! News Search sessions
    - Normal bucket: top 10 positions
    - Random bucket [Li 2011]
      - Randomly shuffled top 4 positions
      - Unbiased evaluation corpus
    - Editor's judgment: 1 day's query log
  - Preference pair selection [Joachims SIGIR05]
    - Click > Skip above
    - Click > Skip next
    - Ordered by Pearson' χ^2 value

Analysis of JRFL

- Relevance and Freshness Learning
  - Baseline: GBRank trained on Dong et al.'s relevance/freshness annotation set
  - Testing corpus: editor's one day annotation set

Table 5: Performance on individual relevance and freshness estimation

	P@1	MAP@3	DCG@5
Relevance GBRank	0.9655	0.3422	14.6026
JRFL Relevance	0.8273	0.2291	14.7962
Freshness GBRank	0.9823	0.4998	18.8597
JRFL Freshness	0.9365	0.3106	19.8228

Query Weight Analysis

Table 6: Query intention analysis by the inferred query weight

Freshness Driven	Relevance Driven
7-Jun-2011, china	5-Jul-2011, casey anthony trial summary
6-Jul-2011, casey anthony trial	9-Jul-2011, nascar qualifying results
24-Jun-2011, nba draft 2011	8-Jul-2011, burbank 100 years parade
28-Jun-2011, libya	10-Jul-2011 gas prices summer 2011
9-Jun-2011, iran	10-Jul-2011, bafta film awards 2011
6-Jun-2011, pakistan	2-Jul-2011, green lantern cast
13-Jun-2011, lebron james	9-Jul-2011, 2011 usga open leaderboard
29-Jun-2011, greece	3-Jul-2011, lake mead water level july 2011
27-May-2011, joplin missing	5-Jul-2011, caylee anthony autopsy report
6-Jun-2011, sarah palin	4-Jul-2011, aurora colorado fireworks 2011

Quantitative Comparison

- Ranking performance
  - Random bucket clicks

Table 8: Comparison On Random Bucket Clicks

Model	FreshDem	RankSVM	GBRank	JRFL
P@1	0.3413	0.3706	0.3882	**0.3969***
P@2	0.3140	0.3372	0.3477	**0.3614***
MAP@3	0.5301	0.5601	0.5751	**0.6012***
MAP@4	0.5859	0.6090	0.6218	**0.6584***
MRR	0.5899	0.6135	0.6261	**0.6335***

* indicates p-value$<$0.05.

Quantitative Comparison

- Ranking performance
  - Normal clicks

Table 9: Comparison On Normal Clicks

Model	FreshDem	RankSVM	GBRank	JRFL
P@1	0.3886	0.5981	0.5896	**0.6164***
P@2	0.2924	0.4166	0.4002	**0.4404***
MAP@3	0.4991	0.7208	0.6849	**0.7502***
MAP@4	0.5245	0.7383	0.7024	**0.7631***
MRR	0.5781	0.7553	0.7355	**0.7702***

* indicates p-value<0.05

Quantitative Comparison

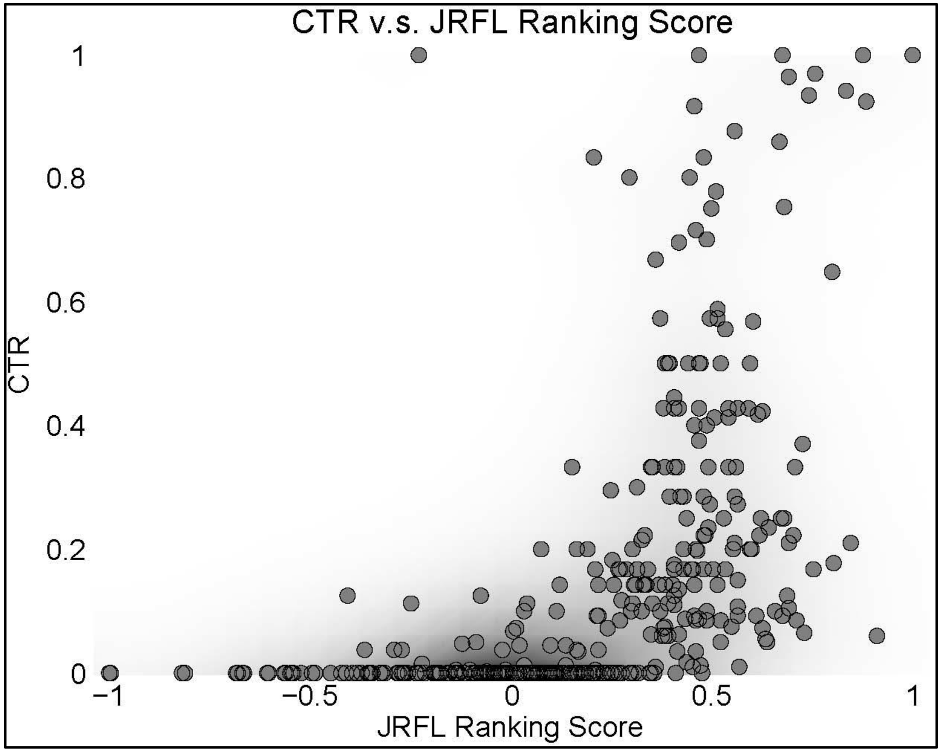
- Ranking performance
  - Editorial annotations

Table 10: Comparison On Editorial Annotations

Model	FreshDem	RankSVM	GBRank	JRFL
P@1	0.9184	0.9626	**0.9870**	0.9508
P@2	0.9043	0.9649	**0.9729**	0.9117
MAP@3	0.3055	0.3628	0.3731	**0.4137**
MAP@4	0.4049	0.4701	**0.4796**	0.4742
MRR	0.9433	0.9783	**0.9920**	0.9745
DCG@1	6.8975	7.9245	**8.1712***	7.2203
DCG@5	15.7175	17.2279	17.7468	**18.9397***

* indicates $p\text{-value}<0.05$.

CTR distribution revisit



Correlation:
0.7163±0.1673

Summary

- Joint Relevance and Freshness Learning
  - Query-specific preference
  - Learning from query logs
  - Temporal features
- Future work
  - Personalized retrieval
    - Broad spectral of user's information need
    - E.g., trustworthiness, opinion

Refs

- [Del Corso WWW05] Gianna M. Del Corso, Antonio Gulli, Francesco Romani: Ranking a stream of news. WWW 2005: 97-106
- [Liu WWW09] Tie-Yan Liu: Tutorial on learning to rank for information retrieval. WWW 2009
- [Dong WSDM10] Anlei Dong, Yi Chang, Zhaohui Zheng, Gilad Mishne, Jing Bai, Ruiqiang Zhang, Karolina Buchner, Ciya Liao, Fernando Diaz: Towards recency ranking in web search. WSDM 2010: 11-20
- [Inagaki AAAI10] Yoshiyuki Inagaki, Narayanan Sadagopan, Georges Dupret, Anlei Dong, Ciya Liao, Yi Chang, Zhaohui Zheng: Session Based Click Features for Recency Ranking. AAAI 2010
- [Dong WWW10] Anlei Dong, Ruiqiang Zhang, Pranam Kolari, Jing Bai, Fernando Diaz, Yi Chang, Zhaohui Zheng, Hongyuan Zha: Time is of the essence: improving recency ranking using Twitter data. WWW 2010: 331-340
- [Zhang EMNLP10] Ruiqiang Zhang, Yuki Konda, Anlei Dong, Pranam Kolari, Yi Chang, Zhaohui Zheng: Learning Recurrent Event Queries for Web Search. EMNLP 2010: 1129-1139
- [Chang SIGIR12] Po-Tzu Chang, Yen-Chieh Huang, Cheng-Lun Yang, Shou-De Lin, Pu-Jen Cheng: Learning-based time-sensitive re-ranking for web search. SIGIR 2012: 1101-1102
- [Kanhabua CIKM12] Nattiya Kanhabua, Kjetil Nørvåg: Learning to rank search results for time-sensitive queries. CIKM 2012: 2463-2466

Refs

- [Wang WWW12] Hongning Wang, Anlei Dong, Lihong Li, Yi Chang, Evgeniy Gabrilovich: Joint relevance and freshness learning from clickthroughs for news search. WWW 2012: 579-588
- [Dai SIGIR11] Na Dai, Milad Shokouhi, Brian D. Davison: Learning to rank for freshness and relevance. SIGIR 2011: 95-104
- [Efron SIGIR11] M. Efron and G. Golovchinsky. Estimation methods for ranking recent information. In SIGIR, pages 495–504, 2011.
- [Li CIKM03] X. Li and W. Croft. Time-based language models. In CIKM, pages 469–475, 2003.
- [Li WSDM11] L. Li, W. Chu, J. Langford, and X. Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In Proceedings of ACM WSDM '11, pages 297–306, 2011.
- [Joachims SIGIR05] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In SIGIR, pages 154–161, 2005.

Outline

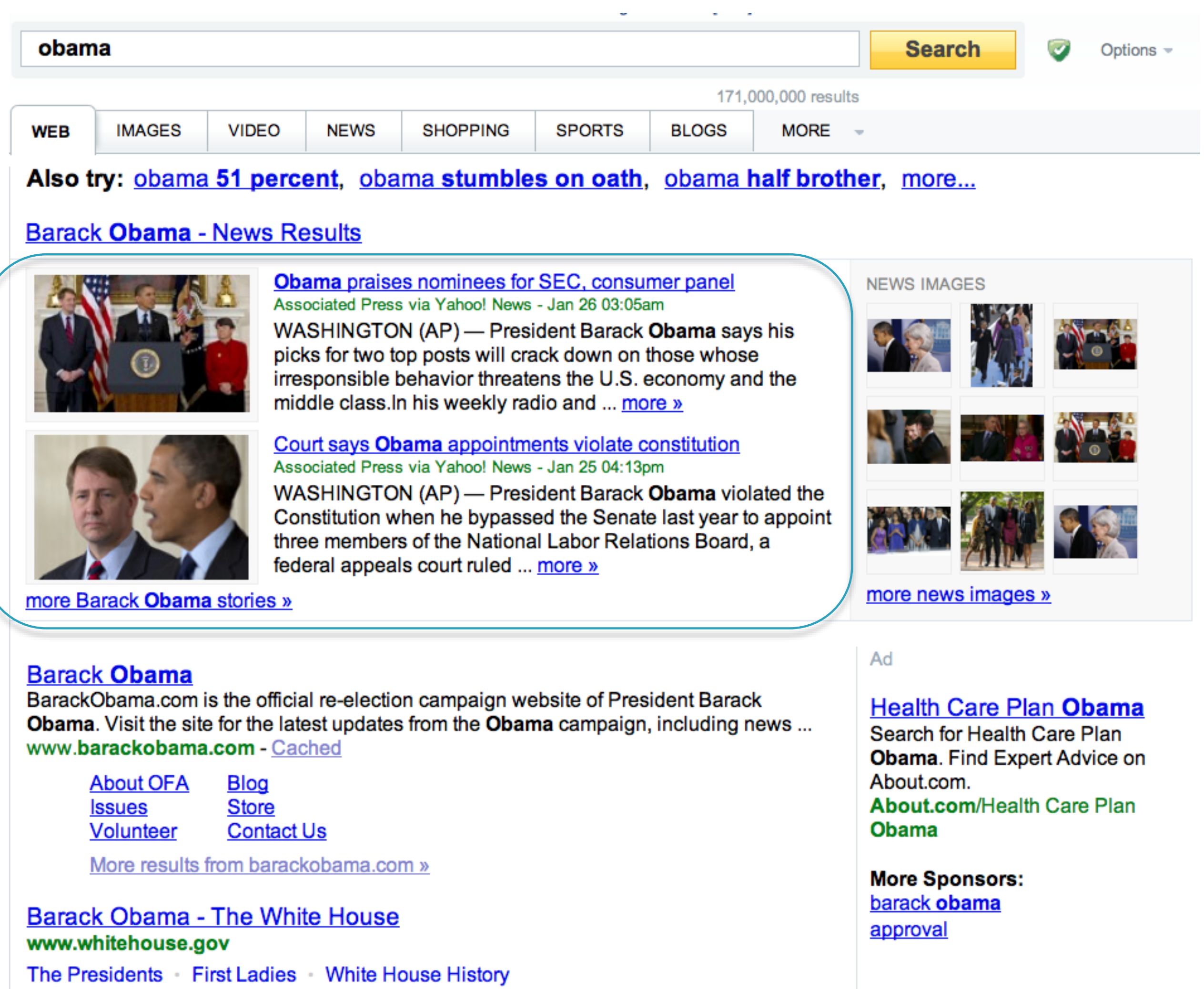
- Time-sensitive search
  - Time-sensitive ranking relevance
  - Federated search
- Time-sensitive recommendation

Federated Search

- In web search engine results
- To integrate vertical search engine results
  - News
  - Local
  - Shopping
  - Finance
  - Movie
  - Travel
  - … …
- Also called DD (direct display)

News DD

News DD



Critical Challenge

- Understand query intent and surface relevant content
  - When to trigger DD?
  - Where to show the DD?
  - Maximize user satisfaction subject to business constrains

Proxy for User Satisfaction

- Strong correlation: CTR & newsworthiness
  - [Diaz WSDM09]
  - Editors label queries for newsworthiness
  - Check the correlation between CTR & labeling
- So user click info can represent query's newsworthiness

Applicability of Existing Approaches

- Web document ranking?
  - CTR is not correlated with query-document relevance
- Query classification?
  - Buzzy words change rapidly
- Online model?
  - No initial CTR data
- Human labeling is very difficult (if not impossible)

Approach by Konig et al. [Konig SIGIR09]

- Data sources for feature computation
  - News corpus
  - Blog corpus
  - Wikipedia corpus
  - … …
- 7-day's data corpus window
  - Small enough for main memory use
- News and Blog complement each other
- Wikipedia is background corpus

Features

- Corpus frequency features
  - frequency of documents matching the query
  - Frequency difference
  - Based on news article title and full text
  - tf-idf method for query term salience
- Context features
  - Breaking news query usually surfaces similar documents
  - On the other hand, “NY Times” return different stories
  - Compute the coherence of returned documents

Features

- Query-only features
  - Ratio of stop words to query length in tokens
  - Ratio of special characters
    - E.g., www.google.com
  - Ratio of capitalization terms
    - Check if query terms are capitalized in news corpus
      - E.g., “Casey Anthony”

Leverage Click Feedback

- [Diaz WSDM09]
- CTR can be estimated simply by

$$\tilde{p}_q^t = \frac{\mathcal{C}_q^t}{\mathcal{V}_q^t}$$

- But
  - Samples are sparse especially at initial stage
  - Click probability is changing over time
- Therefore we need initial guess

Incorporate Prior Estimation into Click Feedback

- Posterior mean:

$$\tilde{p}_q^t = \frac{\mathcal{C}_q^t + \mu \pi_q^t}{\mathcal{V}_q^t + \mu}$$

π_q^t : prior estimation

Small μ: sensitive to early user feedback

Large μ: rely more on prior estimation

- Aggregate clicks/views from similar queries

$$\tilde{\mathcal{C}}_q^t = \mathcal{C}_q^t + \sum_{q'} \mathcal{B}(q, q')\mathcal{C}_{q'}^t$$

$$\tilde{\mathcal{V}}_q^t = \mathcal{V}_q^t + \sum_{q'} \mathcal{B}(q, q')\mathcal{V}_q^t$$

$\mathcal{B}(q_i, q_j)$: query similarity

Features for Prior Estimation

feature	description
`query-last-k`	how many of the last k queries were q
`query-last-k-yesterday`	yesterday, how many of the last k queries were q
`news-last-k`	how many of the last k queries on the news vertical were q
`news-last-k-yesterday`	yesterday, how many of the last k queries on the news vertical were q
`doc-last-k`	how many of the last k documents were retrieved by q
`doc-last-k-yesterday`	yesterday, how many of the last k documents were retrieved by q
`weight-mean-age`	weighting by relevance, how old is the average retrieved document
`weight-stddev-age`	weighting by relevance, what is the standard deviation of retrieved documents

Click Precision and Recall

Baseline: contextual model (prior mean)

Training: use click feedback

Scalability

- Many different verticals
  - News, Shopping, Local, Finance, Movie, Travel, …
  - [Arguello SiGIR09] more features
- Many different markets
  - US, CA, UK, FR, TW, HK, … …
- Need a system that can be applied to all different verticals with minimal effort.
  - Automatic data generation
  - Automatic feature generation
  - Automatic model training/evaluation
    - Not rely on editorial data at all

Exploration

- Uniform Random Exploration over the set of available choices (“actions”)
- Action = Slotting Decision = Slot DD ‘v’ at slot ‘s’ where
  - v in V = set of all legally available DDs.
  - s in S = set of all legally available slots for v, may include NONE.
- Features are logged at the same time.

Generating Data

- Thus each event in the data is a 4-tuple (a, p, x, r)
  - a: Result slotted
  - x: Feature vector
  - r: Observed reward
  - p: Probability of action, Pr(v@s)

Features

- Query features
  - Lexical Features - Bag of words, bigrams, co-occurrence stats, etc.
  - Query attributes - query classification, length, etc.
- Corpus / Vertical level features:
  - Query independent historical CTRs, User preferences etc.
- Post-retrieval features
  - Query-Document match features (ranking scores and features)
  - Global result set features

Summary

- We have introduced
  - Two classical papers on news federation search
  - Scalability issue
- More issues
  - False positive will hurt user experience badly
  - More features

Refs

- [Arguello SIGIR09] Jaime Arguello, Fernando Diaz, Jamie Callan, Jean-Francois Crespo: Sources of evidence for vertical selection. SIGIR 2009: 315-322
- [Diaz WSDM09] Fernando Diaz: Integration of news content into web results. WSDM 2009: 182-191
- [Konig SIGIR09] A. Konig, M. Gamon, and Q. Wu. Click-through prediction for news queries. In Proc. of SIGIR, 2009
- [Kumar WSDM11] Ashok Kumar Ponnuswami, Kumaresh Pattabiraman, Qiang Wu, Ran Gilad-Bachrach, Tapas Kanungo: On composition of a federated web search result page: using online users to provide pairwise preference for heterogeneous verticals. WSDM 2011: 715-724
- [Kumar WWW11] Ashok Kumar Ponnuswami, Kumaresh Pattabiraman, Desmond Brand, Tapas Kanungo: Model characterization curves for federated search using click-logs: predicting user engagement metrics for the span of feasible operating points. WWW 2011: 67-76
- [Arguello CIMK12] Jaime Arguello, Robert Capra: The effect of aggregated search coherence on search behavior. CIKM 2012: 1293-1302
- [Chen WSDM12] Danqi Chen, Weizhu Chen, Haixun Wang, Zheng Chen, Qiang Yang: Beyond ten blue links: enabling user click modeling in federated web search. WSDM 2012: 463-472

Outline

- Time-sensitive search
  - Time-sensitive ranking relevance
  - Time-sensitive query suggestion
  - Federated search
- Time-sensitive recommendation

Web Recommender Systems

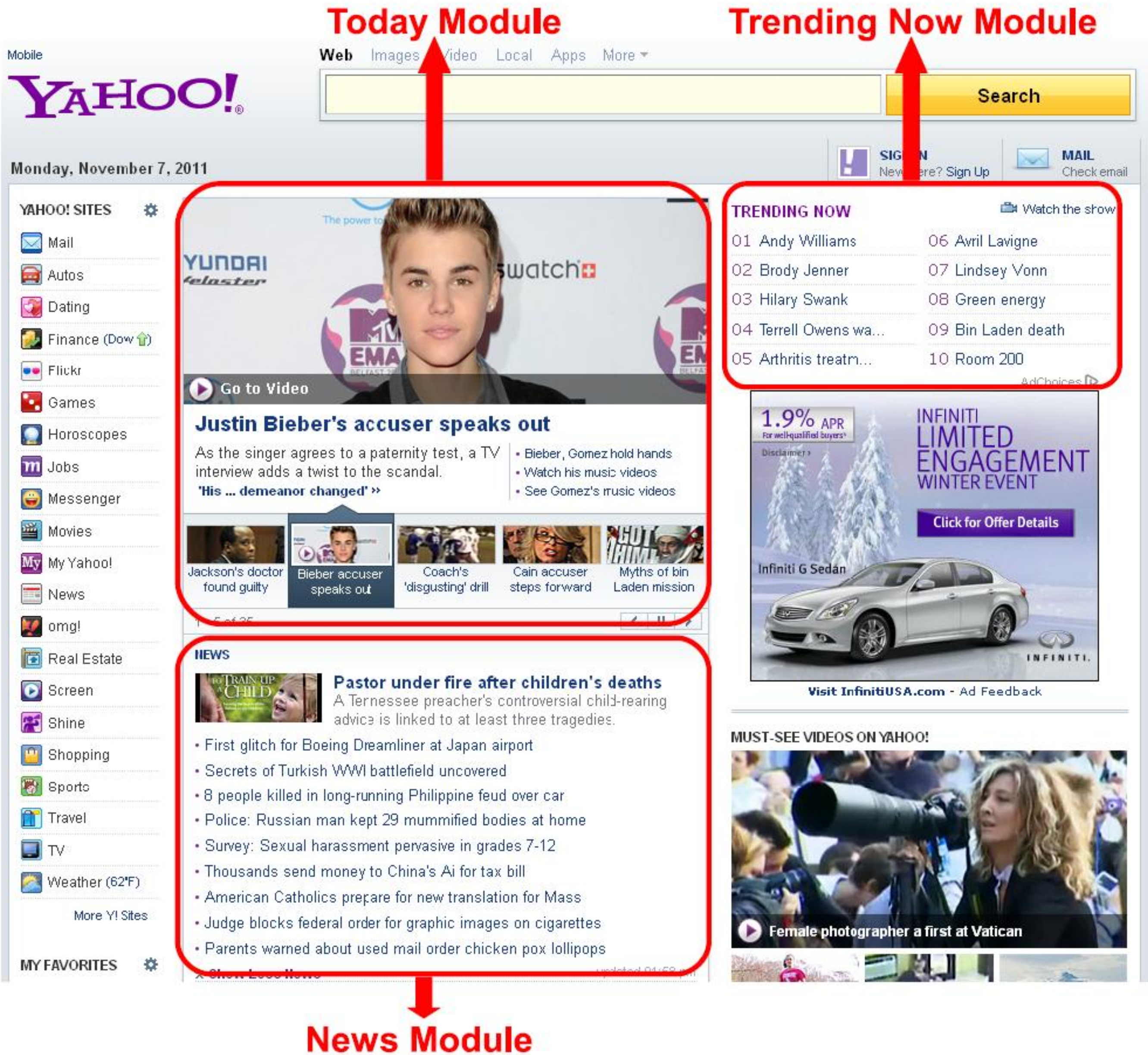
- Recommend **items** to **users** to maximize some **objective(s)**

Outline for Recommendation

- Introduction
- Personalization
- User segmentation
- Action interpretation
- Pairwise preference modeling

Portal

Applications on Recommendation



Scientific Discipline

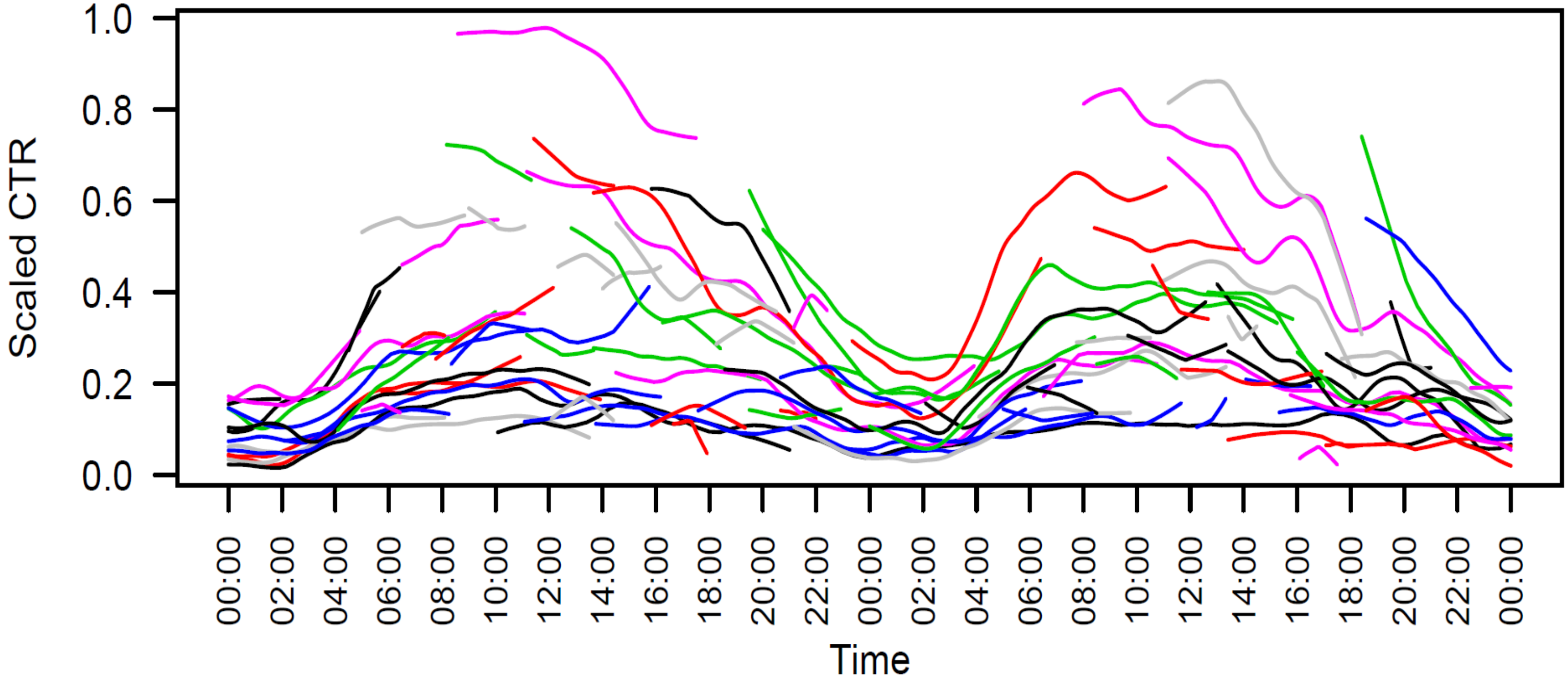
- Machine Learning & Statistics (for learning user-item affinity)
  - Offline Models
  - Online Models
  - Collaborative Filtering
  - Explore/Exploit (bandit problems)
- Multi-Objective Optimization
  - Click-rates (CTR), time-spent, revenue
- User Understanding
  - User profile construction
- Content Understanding
  - Topics, categories, entities, breaking news,…

Some Refs on Previous Research

- Shuang-Hong Yang, Bo Long, Alexander J. Smola, Hongyuan Zha, Zhaohui Zheng: Collaborative competitive filtering: learning recommender using context of user choice. SIGIR 2011: 295-304
- Lihong Li, Wei Chu, John Langford, Xuanhui Wang: Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. WSDM 2011: 297-306
- Wei Chu, Seung-Taek Park: Personalized recommendation on dynamic content using predictive bilinear models. WWW 2009: 691-700
- Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, Xuanhui Wang: Personalized click shaping through lagrangian duality for online recommendation. SIGIR 2012: 485-494
- Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, Xuanhui Wang: Click shaping to optimize multiple objectives. KDD 2011: 132-140
- Deepak Agarwal, Bee-Chung Chen, Bo Long: Localized factor models for multi-context recommendation. KDD 2011: 609-617
- Deepak Agarwal, Bee-Chung Chen: fLDA: matrix factorization through latent dirichlet allocation. WSDM 2010: 91-100

CTR Curves for Dynamic Items

Each curve is the CTR of an item in the Today Module on www.yahoo.com over time



Traffic obtained from a controlled experiment

Things to note:

(a) Short lifetimes, (b) temporal effects, (c) often breaking news stories

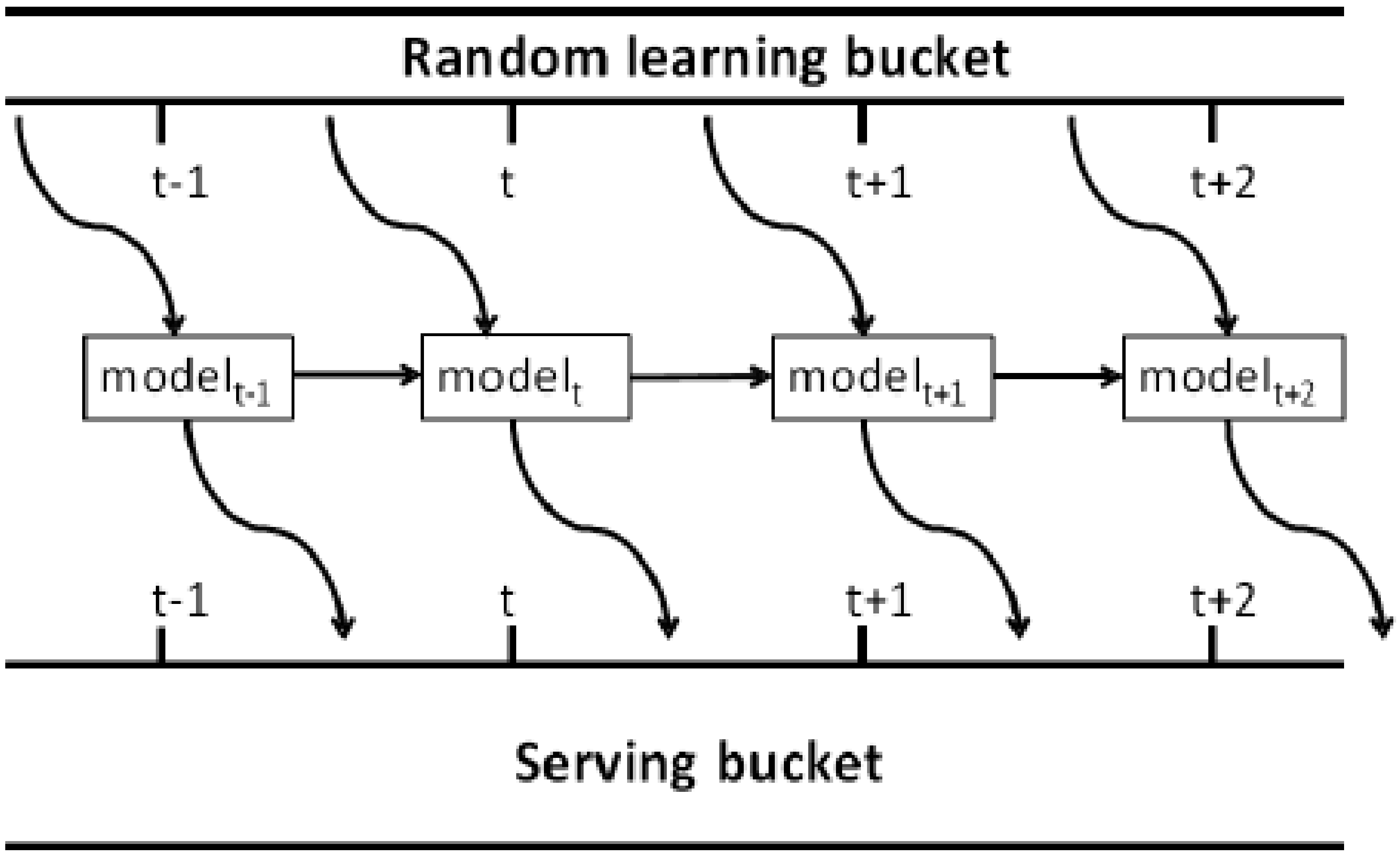
Solutions

- Online learning
  - Content and user interest change fast
  - Offline model cannot capture all of the variations
  - Large amount of user traffic make it possible
- Personalization
  - More relevant to different users

Online Learning

- Ranking model: updated every 5 minutes on users' feedbacks
- Exploration & Exploitation
  - Random bucket (small traffic) for exploration:
    randomly shuffle the ranking of all candidates
  - Serving bucket for exploitation:
    models -> scores -> ranking

Online Learning Flowchart



Per-Item Model

- Each item has a corresponding model.
- For example, **estimated most popular** (**EMP**) model
- Click probability

$$p_{t+1} = \frac{\gamma_t p_t + c_{t,t+1}}{\gamma_t + n_{t,t+1}}$$

where $\gamma_t = w\gamma_{t-1} + n_{t-1,t}$

is sample size.

Outline for Recommendation

- Introduction
- Personalization
- User segmentation
- Action interpretation
- Pairwise preference modeling

Personalization

Gender	CTR
Female	0.24
Male	0.39

Query Category	Gender	CTR
Family	Female	0.34
Family	Male	0.32
Sports	Female	0.16
Sports	Male	0.37
Tech and Gadgets	Female	0.21
Tech and Gadgets	Male	0.44

Query	DMA with highest CTR
SF Giants	San Francisco-Oakland-San Jose
Oregon vs. UCLA	Portland
Texas Rangers	Dallas-Ft. Worth

Age

CTRs are relative values

Personalization Model (I)

- User segmentation
  - Pre-define a few user segments by user features (e.g., age-gender)
  - For each user segment

    apply EMP

Personalization Model (II)

- Online logistic regression (OLR)

$$y = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3 + \cdots$$

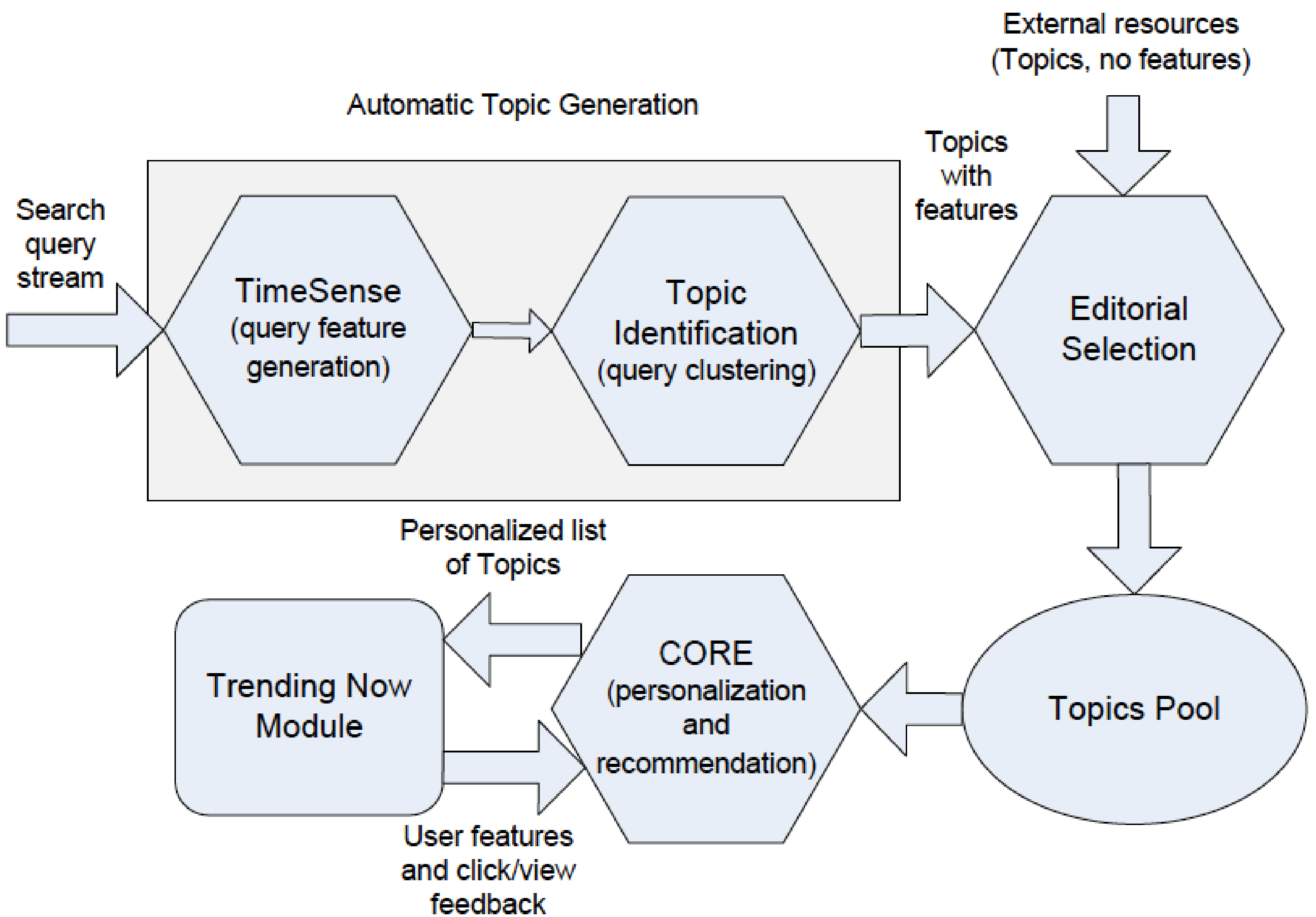
β_0 : intercept term, represent most popular score

$\beta_1, \beta_2, \beta_3, \ldots$: feature weights

$f_1, f_2, f_3, \ldots$: binary user features

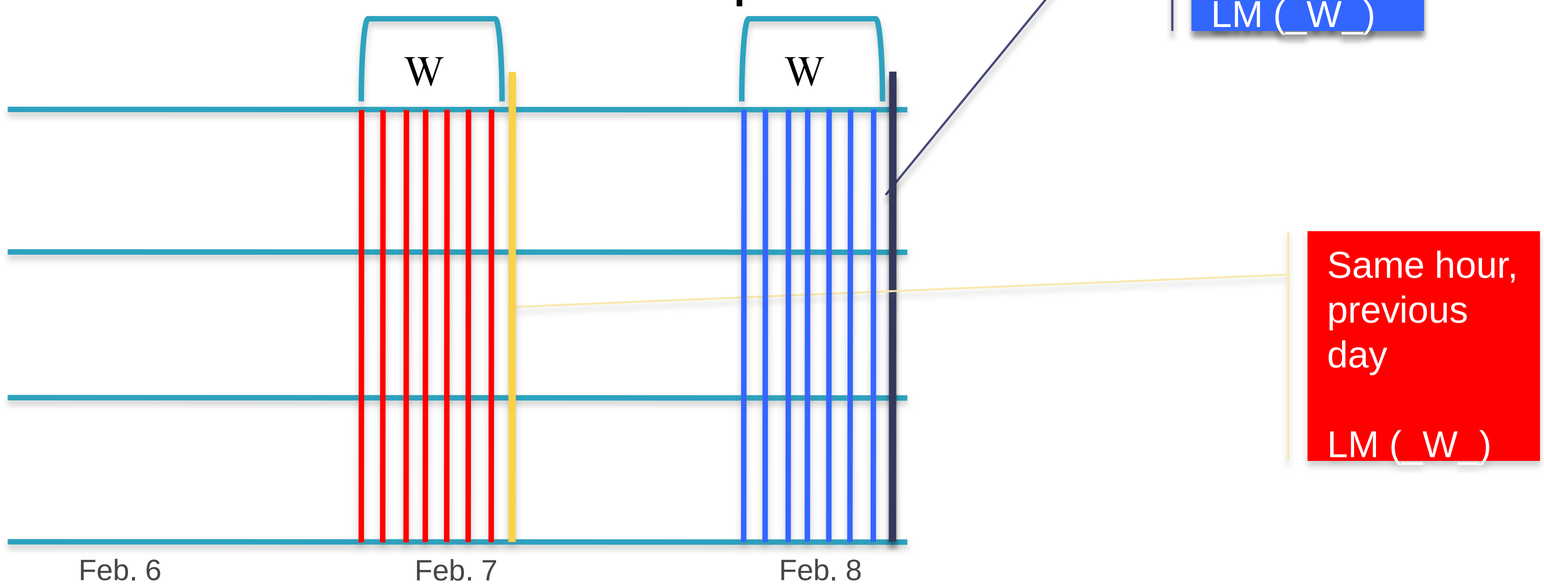
$\beta_i : (\mu_i, \Sigma_i)$

Trending Now Module: Query Recommendation



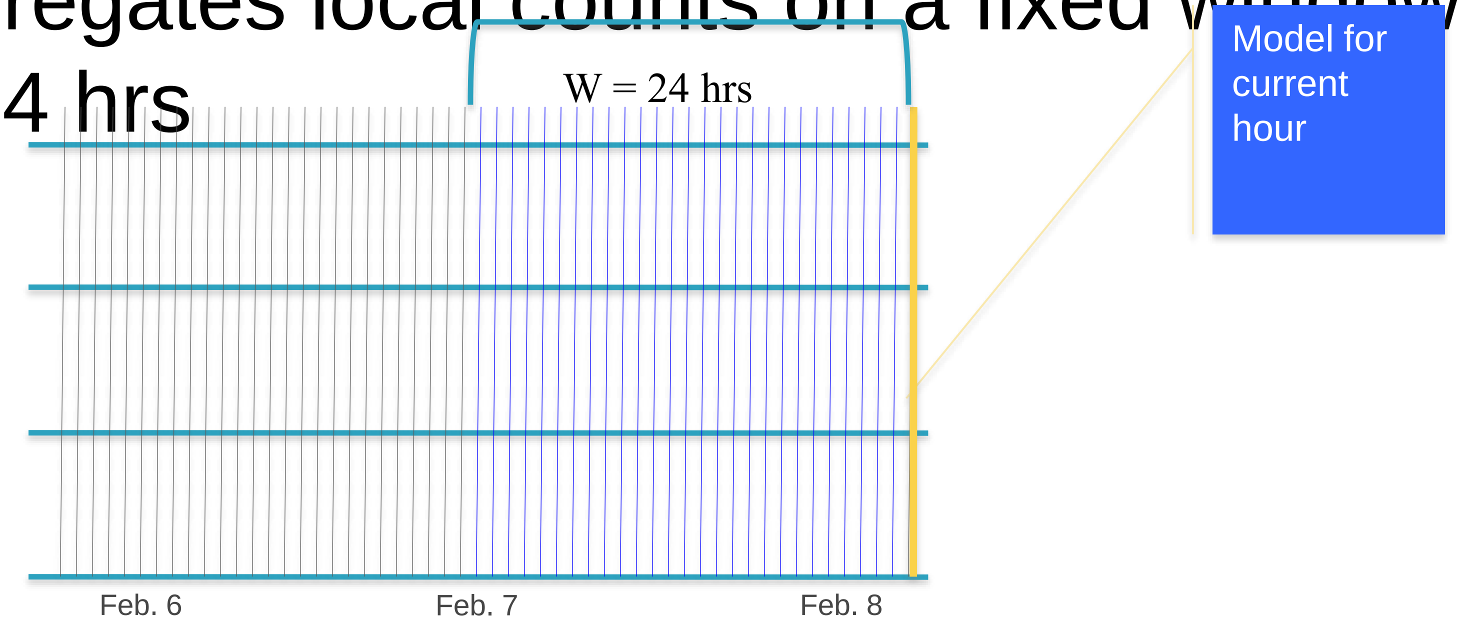
Query Buzz Computation

- ngram based
- uses LM scores based on search queries, queries triggering News DD, and news headlines
- computes the likelihood of the ngrams in a query for:
  - the last hour/window
  - the same hour/window in the previous day
  - the same hour/window in the previous week
  - the same hour/window in the previous month

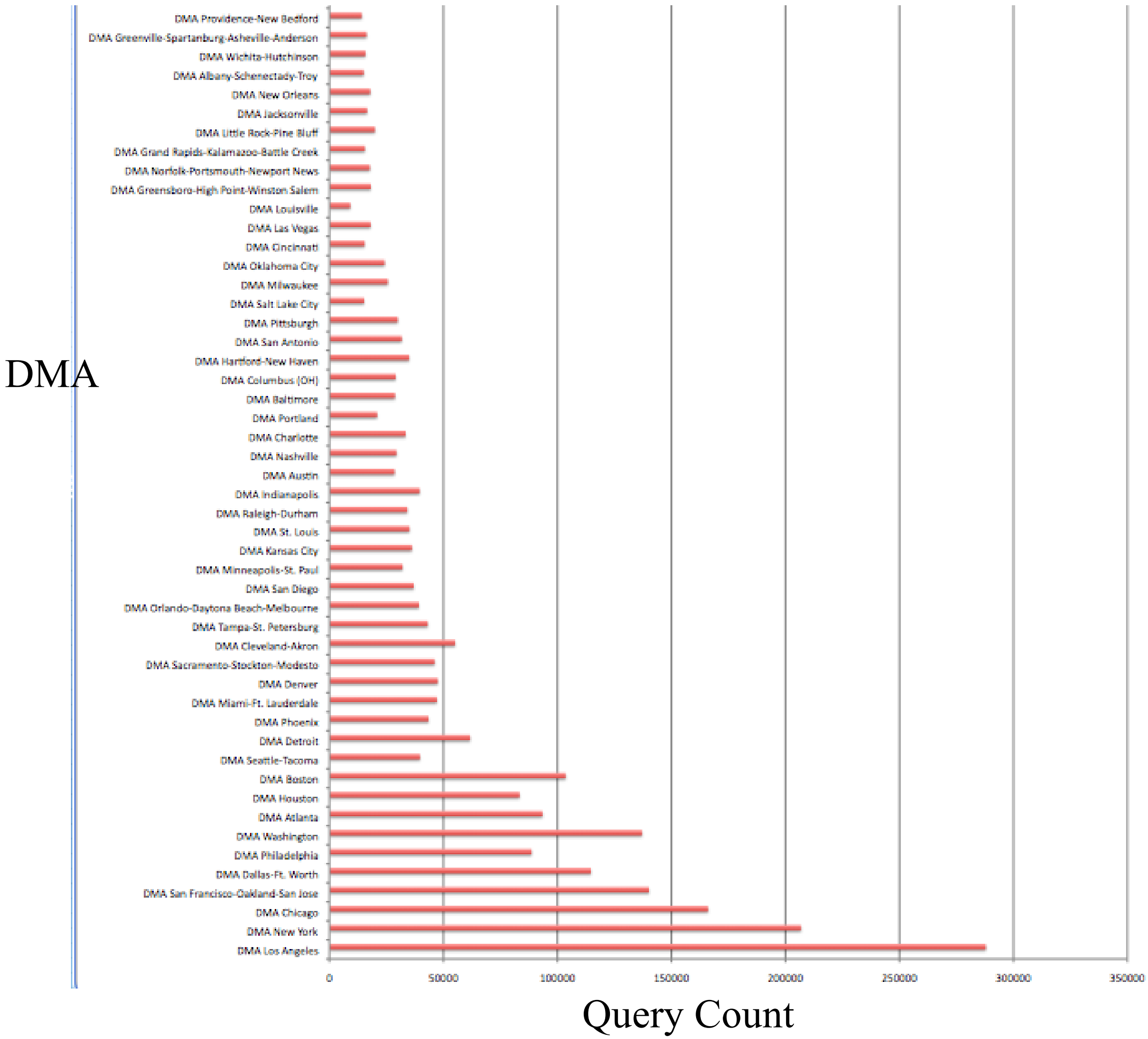


GEO Feature [Bawab KDD12]

- query based
- uses the queries in the TimeSense dictionary
- aggregates local counts on a fixed window of 24 hrs



GEO Capabilities



- DMA: Designated Market Area (Nielsen)
- Top 50 US DMAs
- Log data contains the WOEID/DMA for each query

GEO Model

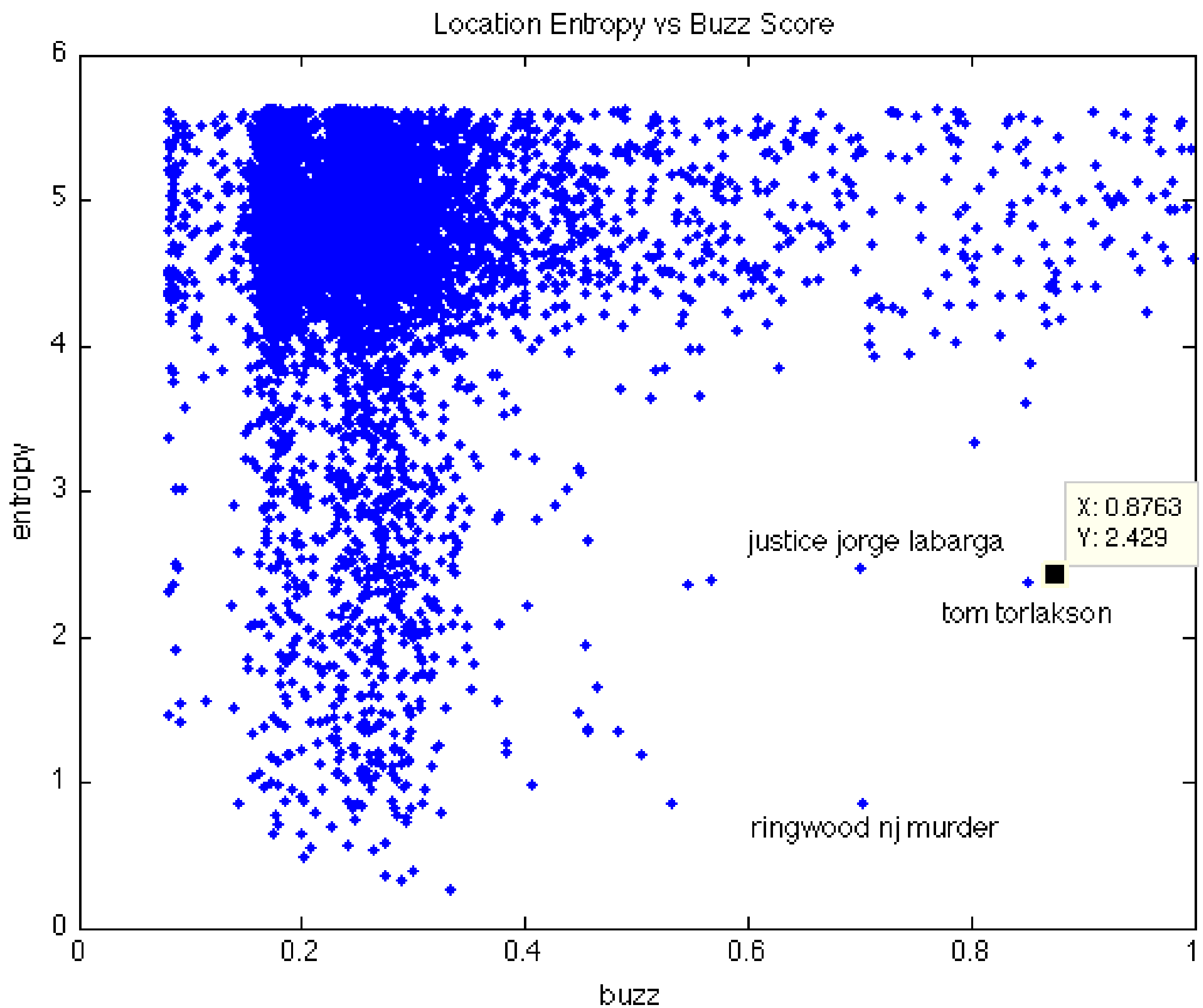
- Entropy of query over DMAs:

$$entropy(dma \mid q) = -\sum_{i=1}^{N} p(dma_i \mid q) * log_2(p(dma_i \mid q))$$

- Posterior probability, normalizes across DMAs, favors larger ones:

$$p(dma_i \mid q) = \frac{v(dma_i, q)}{\sum_{j=1}^{N} v(dma_j, q)}$$

Time-Sensitive vs. Geo-Sensitive



Examples (Buzzy and Local)

Query	Count	Buzz	Entropy	Top DMA nProb
ringwood nj murder	67	0.7024	0.8546	New York = 0.84, Philadelphia = 0.06
tom torlakson	73	0.8506	2.3704	Los Angeles = 0.15, San Fran = 0.16, Sacramento = 0.36, San Diego = 0.21
justice jorge labarga	66	0.7014	2.4733	Miami = 0.19, Tampa = 0.17, Orlando = 0.26, Jacksonville = 0.29
gulf coast claims facility	626	0.5037	1.1892	New Orleans = 0.86
drew brees baby	312	0.4068	0.9781	New Orleans = 0.89

Outline for Recommendation

- Introduction
- Personalization
- User segmentation
- Action interpretation
- Pairwise preference modeling

User Segmentation

- Baseline – heuristic rule
  - E.g., by age-gender

- User behavior information can better reflect users' interests
  - Users with similar behavior patterns are more likely to have similar interests
  - Describing user behaviors:
    - Behavior Targeting (BT) features

Action Interpretation for User Segmentation

- User Segmentation:
  - Use selected features to describe each user
  - Apply clustering methods:
    - K-means
    - Tensor segmentation [Chu KDD09]

[Bian TKDE]

Tensor Segmentation Result

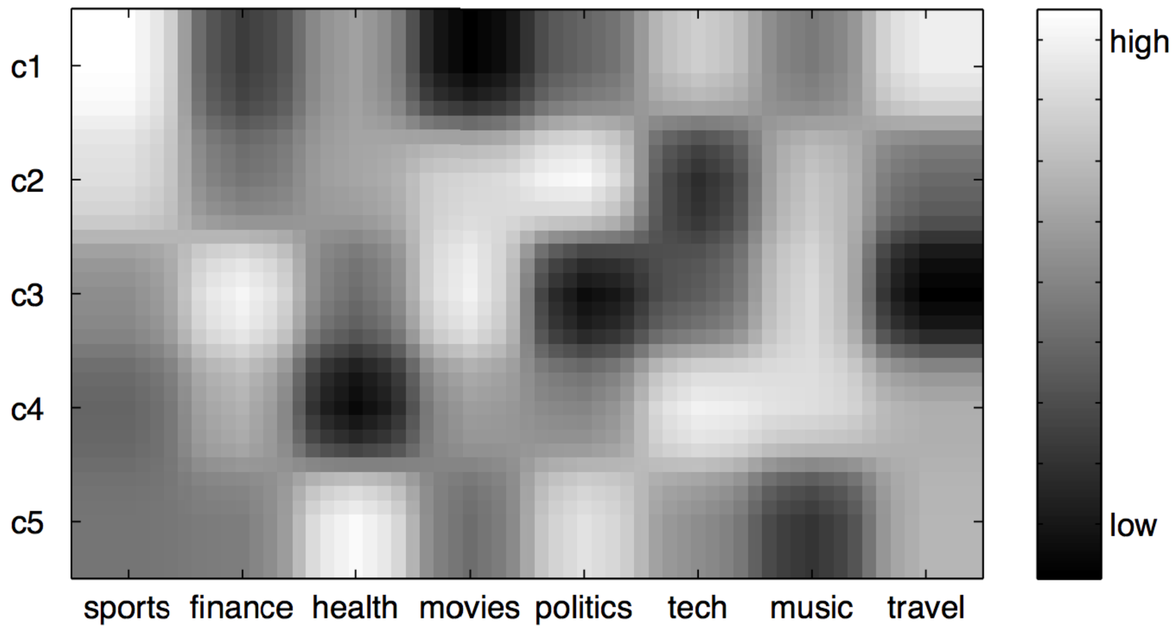


Fig. 6. User segments' preferences on selected item topics in the five example user segments. Each square's gray level indicates the preference of a segment on the corresponding topic, from white (*like*) to black (*dislike*).

Offline Evaluation

- Editorial judge is infeasible
- The correlation between actual clicks and prediction rankings

actual ranking	predicted ranking by Model 1	predicted ranking by Model 2
1 (clicked)	**1**	2
2	5	3
3	4	**1**
4	3	5
5	2	4

Model 1	Model 2
Precision1 = 1	Precision1 = 0
Precision2 = 1	Precision2 = 0
Precision3 = 1	Precision3 = 1
…	…

Compare User Segmentation Approaches

Relative precision gain when training only on *click events* over training on original whole dataset.

Model	prec_1	prec_2	prec_3	prec_4	prec_{10}
EMP	1.81%	-1.57%	-1.79%	-3.65%	-1.72%
EMP-agegender	16.23%	16.07%	15.41%	13.65%	12.58%
EMP-kmeans	20.54%	22.05%	26.39%	26.50%	22.44%
EMP-tensor	22.86%	24.33%	21.02%	22.20%	19.79%

Outline for Recommendation

- Introduction
- Personalization
- User segmentation
- Action interpretation
- Pairwise preference modeling

Action Interpretation for Online Learning

- User is not engaged in every module
- Three event categories
  - Click event
    - user clicked one or more items in the certain module – useful
  - Click-other event
    - contains at least one user action on other modules – not useful
  - Non-click event
    - user has no click action on any module
    - not obvious to determine if the user examine the module
    - we can check user's historic behaviors on this module

User Engagement on Non-Click Events

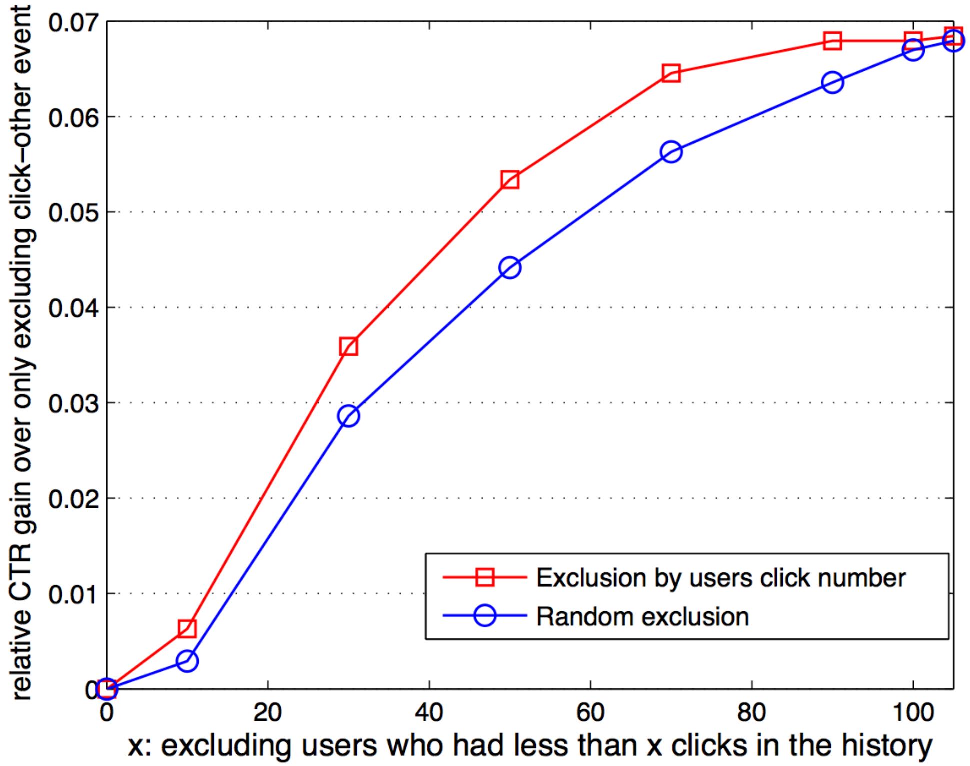


Fig. 8. Relative precision gain when training with data after excluding some *non-click events* over training with data excluding only *click-other events* (using EMP-kmeans).

Remove Click-Other Events

Table 4: Relative precision gain when training without *click-other event* over training on the original whole dataset.

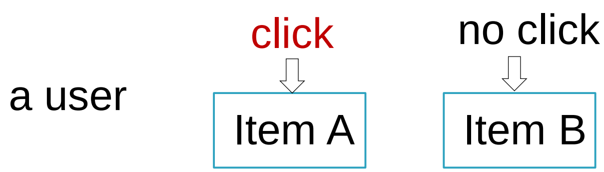
Model	prec_1	prec_2	prec_3	prec_4	prec_{10}
EMP-kmeans	11.11%	7.05%	8.22%	7.70%	5.46%

Outline for Recommendation

- Introduction
- Personalization
- User segmentation
- Action interpretation
- Pairwise preference modeling

Pairwise Preference Learning

- Reality: multiple items displayed at one time
- In one event:



  - Per-item model interpretation:

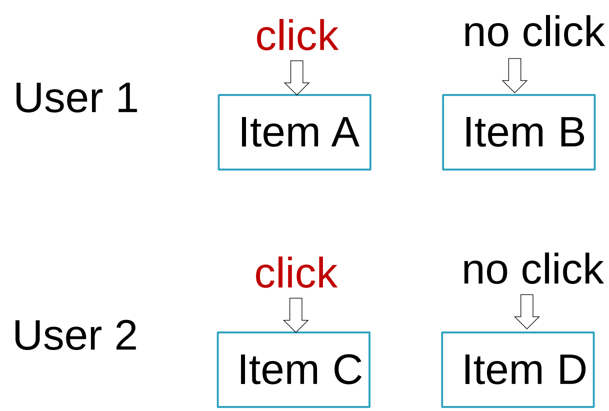
    "Item A was clicked once, Item B was viewed-only once."

  - Preference interpretation:

    "the user liked Item A better than Item B."

[Bian TIST]

Another Example



- By per-item model

  CTR(A) = 1; CTR(B) = 0; CTR(C) = 1; CTR(D) = 0.

  A = C > B = D (wrong due to limited observations)

- Facts are only:

  A > B; C > D; A ? C; A ? D; B ? C; B ?

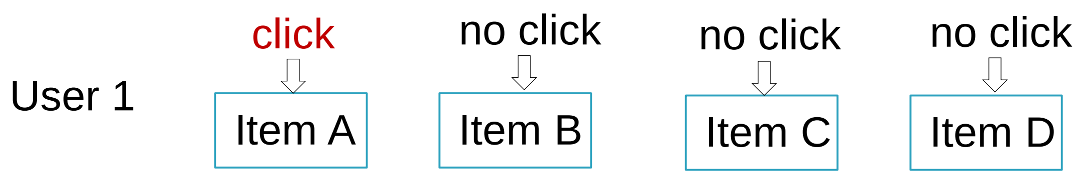
Learning Sample Sparsity

- Many users never really examine the module;
- Candidate pool size >> display number;
- Personalization: makes it even worse

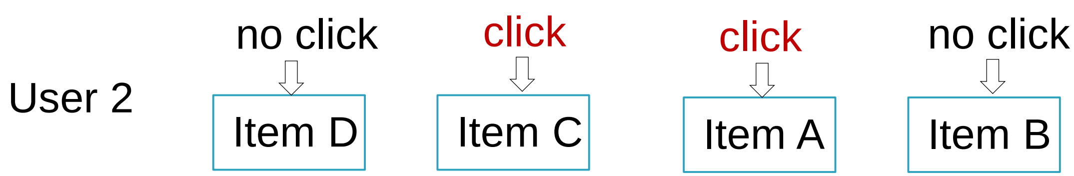
Our Approach for Sample Sparsity

- Use pair-wise preferences for learning
  - Can better deal with sparse problem
  - More straightforward way for final ranking
  - A proven effective approach in search ranking problem.
- Two algorithms
  - Graph-based pairwise learning
  - Bayesian pairwise learning

Preference Extraction



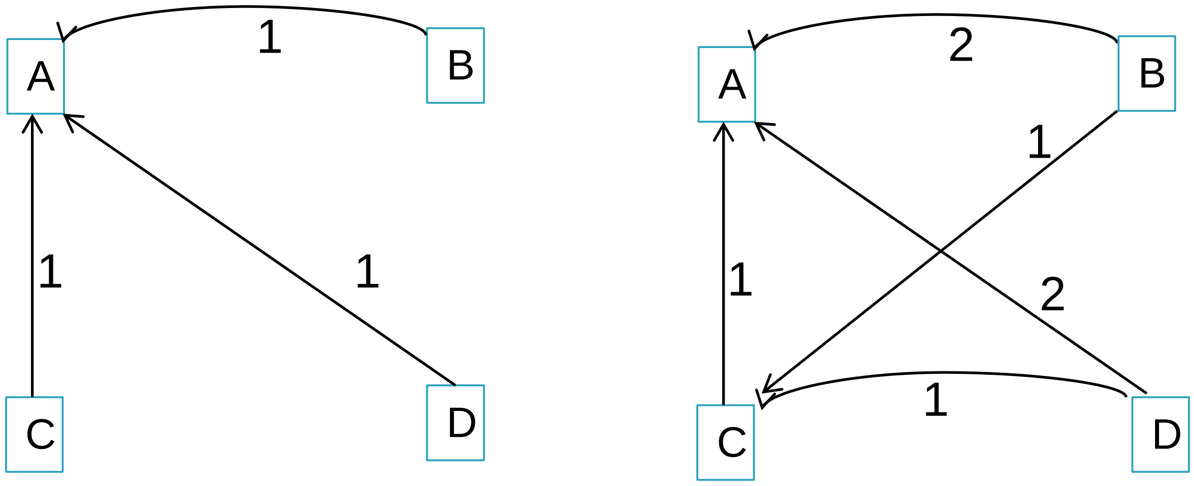
Preferences: A > B; A > C; A > D.



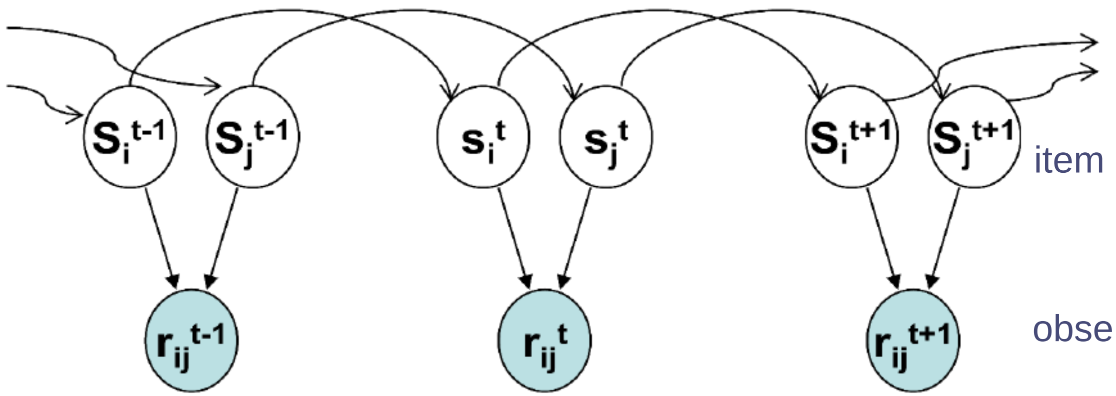
Preferences: C > D; C > B; A > D; A > B.

Graph-Based Pairwise Learning

- Borrow PageRank idea
  - Preferences: A > B;   A > C;   A > D.
  - Preferences: C > D;   C > B;   A > D;   A > B.



Bayesian Pairwise Learning



Bayesian hidden score (**BHS**) model

- Preference distribution:

$$r_{ij}^t \sim p(r_{ij}^t | s_i^t, s_j^t, \alpha), \qquad r_{ij}^t | s_i^t, s_j^t, \alpha \sim N(s_i^t - s_j^t, \alpha)$$

- Attractiveness distribution:

$$s_i^t \sim p(s_i^t | s_i^{t-1}, \lambda), \qquad s_i^t | s_i^{t-1}, \lambda \sim N(s_i^{t-1}, \lambda)$$

Model Optimization

- Likelihood function

$$p(D;\alpha,\lambda)=\prod_{t}\prod_{r_{ij}^{t}\in D}(p(r_{ij}^{t}|s_i^t,s_j^t|,\alpha)p(s_i^t|s_i^{t-1}\lambda)p(s_j^t|s_j^{t-1},\lambda))$$

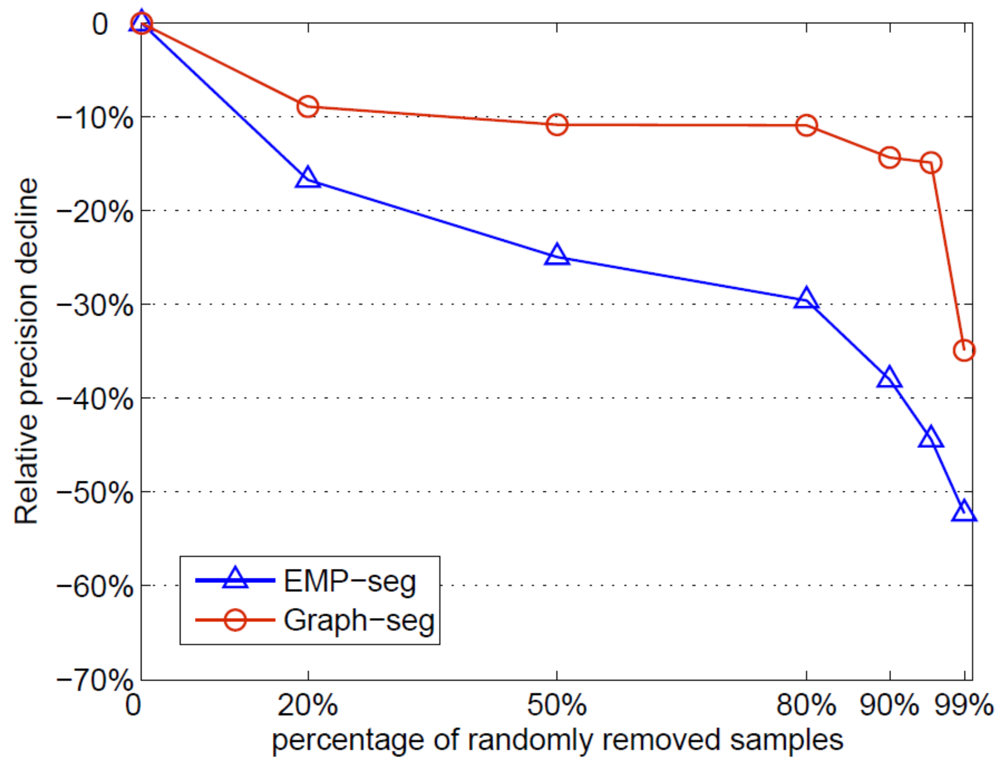
- Final task

$$\min_{s}\sum_{t}\sum_{r_{ij}^{t}\in D}(||r_{ij}^{t}-(s_i^t-s_j^t)||^2+\gamma(||s_i^t-s_j^{t-1}||^2+||s_i^t-s_j^{t-1}||^2),$$

- Optimization:
  - Stochastic gradient descent algorithm

Sample Sparsity Effect

- Trending Now data
- Removing learning samples, compare:
  - Per-item model decline
  - Preference model decline
- Conclusion
  - The fewer samples, the more effective the preference learning approach



Summary

- We have introduced
  - Time-sensitive + geo sensitive
  - User segmentation
  - Action interpretation
  - Pair-wise learning
- We have NOT introduced
  - Many failed efforts
- Many Lessons
  - Appropriate features and sampling are extremely critical in practice

Refs

- [Bian TKDE] Jiang Bian, Anlei Dong, Xiaofeng He, Srihari Reddy, Yi Chang: User action interpretation for personalized content optimization in recommender systems. IEEE Transactions on Knowledge and Data Engineering, to appear.
- [Bawab KDD12] Ziad Al Bawab, George H. Mills, Jean-Francois Crespo: Finding trending local topics in search queries for personalization of a recommendation system. KDD 2012: 397-405
- [Bian TIST] Jiang Bian, Bo Long, Lihong Li, Anlei Dong, Yi Chang, Exploiting User Preferences for Online Learning in Recommender Systems, submitted to ACM Transactions on Intelligent Systems and Technology (TIST)

Summary & Resources

WSDM 2013 Tutorial

Summary and Other Venue

- Wikipedia Page
  - http://en.wikipedia.org/wiki/Temporal_information_retrieval
- Workshops
  - TempWeb WWW'13
  - International Workshop on Big Data Analytics for the Temporal Web (2012)
  - Time-Aware Information Access (TAIA) associated to SIGIR'12
  - Temporal Web Analytics Workshop associated to WWW2011
  - TERQAS (Time and Event Recognition for Question Answering Systems) workshops
  - Workshop on Web Search Result Summarization and Presentation associated to WWW2009
  - Workshop on Temporal Data Mining associated to ICDM2005
  - Workshop on Text Mining associated to KDD2000
- TREC
  - Temporal Summarization Track
  - Microblog Track