Temporal point processes in practice

Liangda Li

Yahoo Research

August 4th, 2019

Outline

- Typical real-world applications via TPP
- Dyadic Event in temporal point process
- Marked Event in temporal point process
- Cross-domain Event in temporal point process
- Parametric influence in temporal point process

Influence

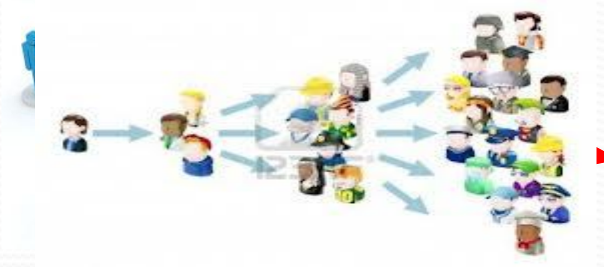
- What?
  - The effect that people have upon the beh
  - Behavior
    - Active: retweet
    - Passive: virus infection
- Why?
  - People interact & learn from the past
- Where?
  - Self-influence
  - Mutual-influence
    - Between individuals
- How?
  - Historical behaviors influence current behaviors


gideonstrumpet @gideonstrumpet 11h
Is this legit? @Popehat @marcorandazza RT @tbuhl: @[me] then I would sue you because I have stated my tweets are not on record comments
Expand

Popehat @Popehat 10h
@gideonstrumpet @marcorandazza @tbuhl no, not legit. Ignorant and preposterous.
Expand

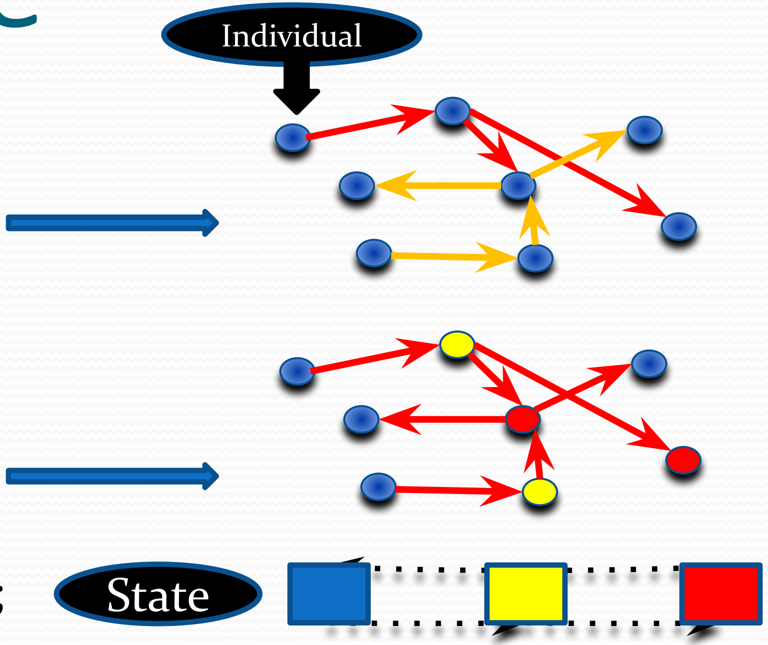
Mark W. Bennett @MarkWBennett 9h
@Popehat @gideonstrumpet @marcorandazza @tbuhl "That's cray cray," as Thing One is wont to say.
Expand

marcorandazza @marcorandazza 9h
@MarkWBennett @Popehat @gideonstrumpet @tbuhl Moronic. You don't get to say something in public and then say "that's off the record"
Expand



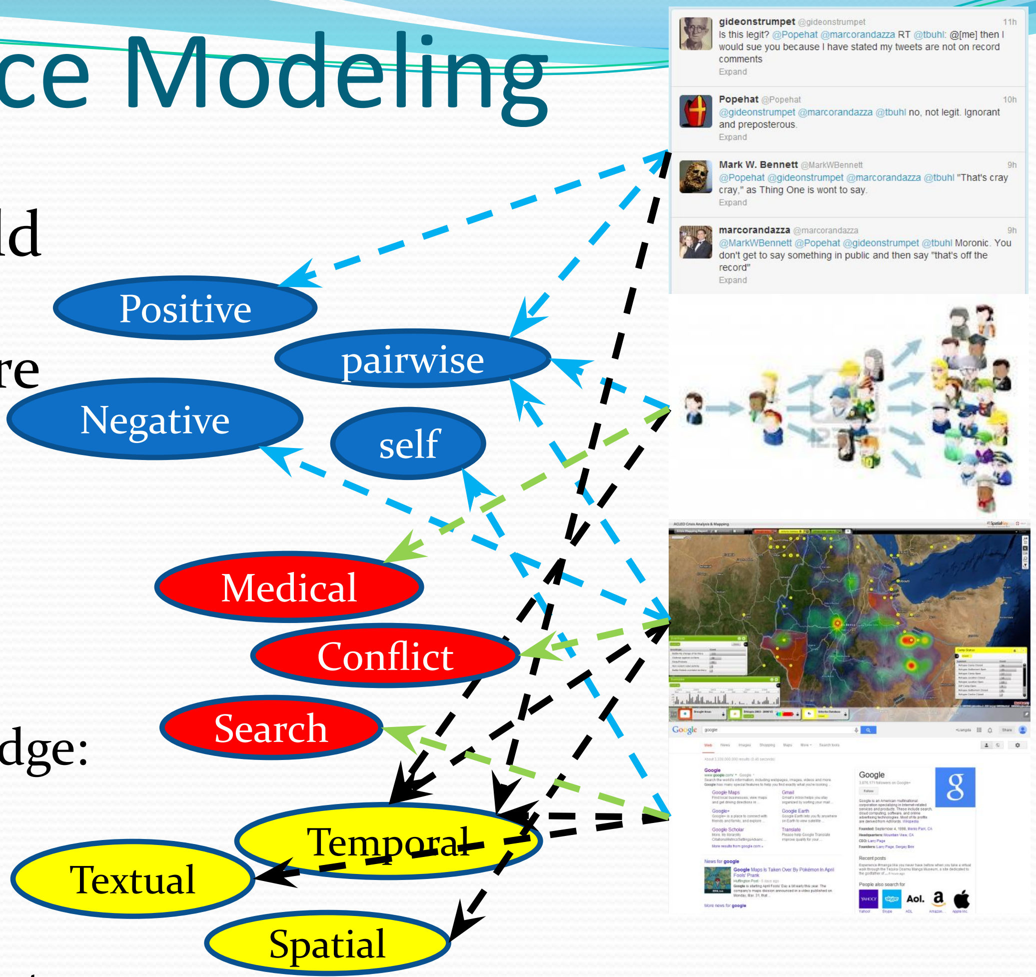
Effect & Importance

- Influenced individual
  - Carry on the same type of behavior
    - Retweet the same post;
    - Infected by the same virus.
  - Respond with some other type of behavior based on certain rules
    - The attack against one country may cause its revenge to the attacker's allies;
    - The results of current search task may trigger a related search task in the next.
- Tracking the diffusion of memes
- Study the chain reactions



Issues in Influence Modeling

- Under different real-world scenarios:
  - The specific influence are diverse:
  - Each demands unique solution using:
    - Domain-specific knowledge:
    - Observed data of specific type:



Outline

- Typical real-world applications via TPP
- Dyadic Event in temporal point process
- Marked Event in temporal point process
- Cross-domain Event in temporal point process
- Parametric influence in temporal point process

Dyadic Event

- Dyadic event: Timestamped interactions involving pairs of actors
  - Email communication, Conflict, Gang rivalry
- More complicated than single actor events
  - Influence between different pairs that shared the same actor
- Actors of events can be unobserved – Dyadic Event Attribution Problem (DEAP)

Data set

- Conflict
  - ACLED: https://www.acleddata.com/
- Gang rivalry
  - LADP: http://www.lapdonline.org/
- Email communication
  - Rnron: http://www.cs.cmu.edu/~enron/

Hawkes Process for DEAP

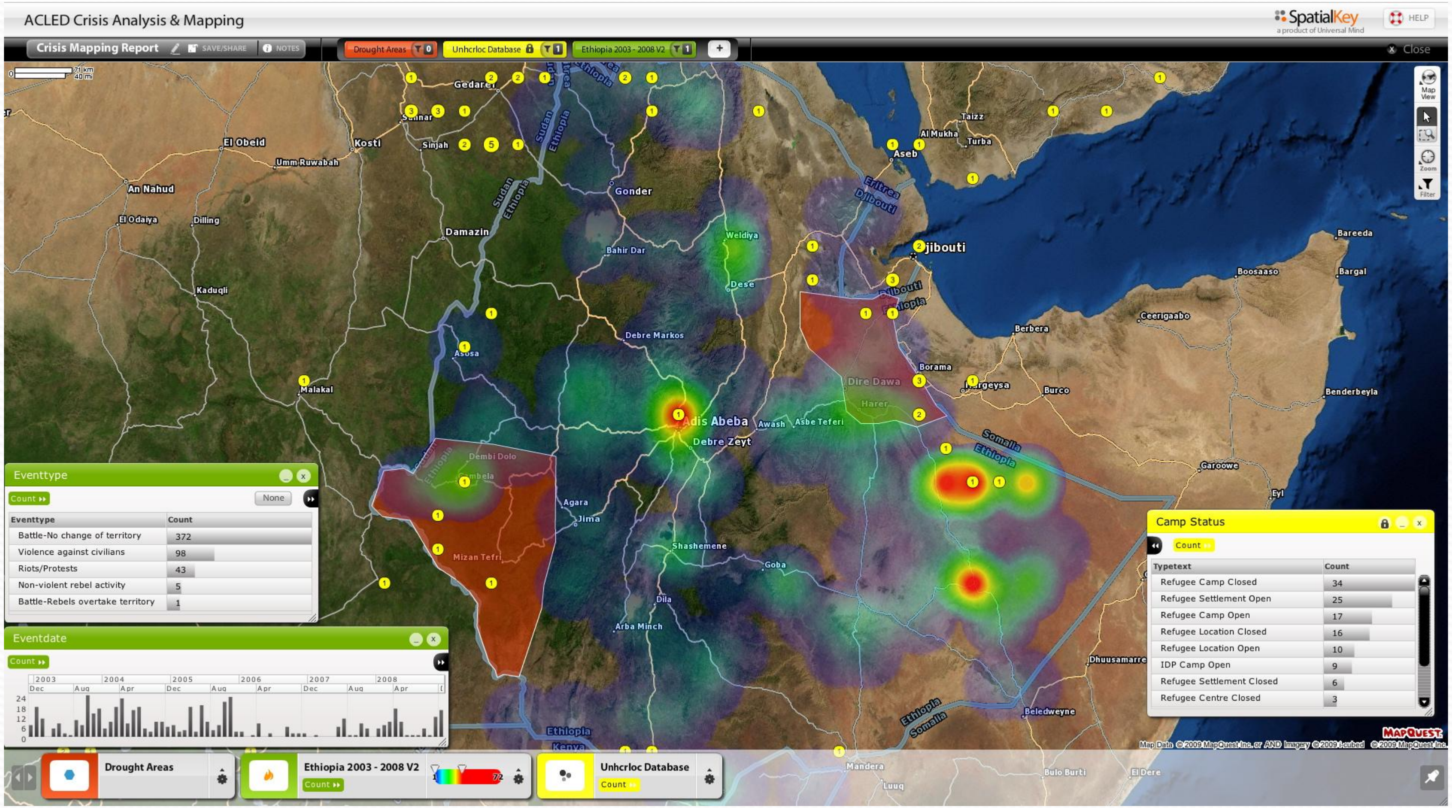
- Introduce binary variable $Z_{n,m}$ to denote whether the n-th event belongs to pair m

$$\lambda_m^*(t) = \mu_m(t) + \sum_{t_l<t} \sum_{m'} \kappa_{m'm}(t - t_l) Z_{l,m'}$$

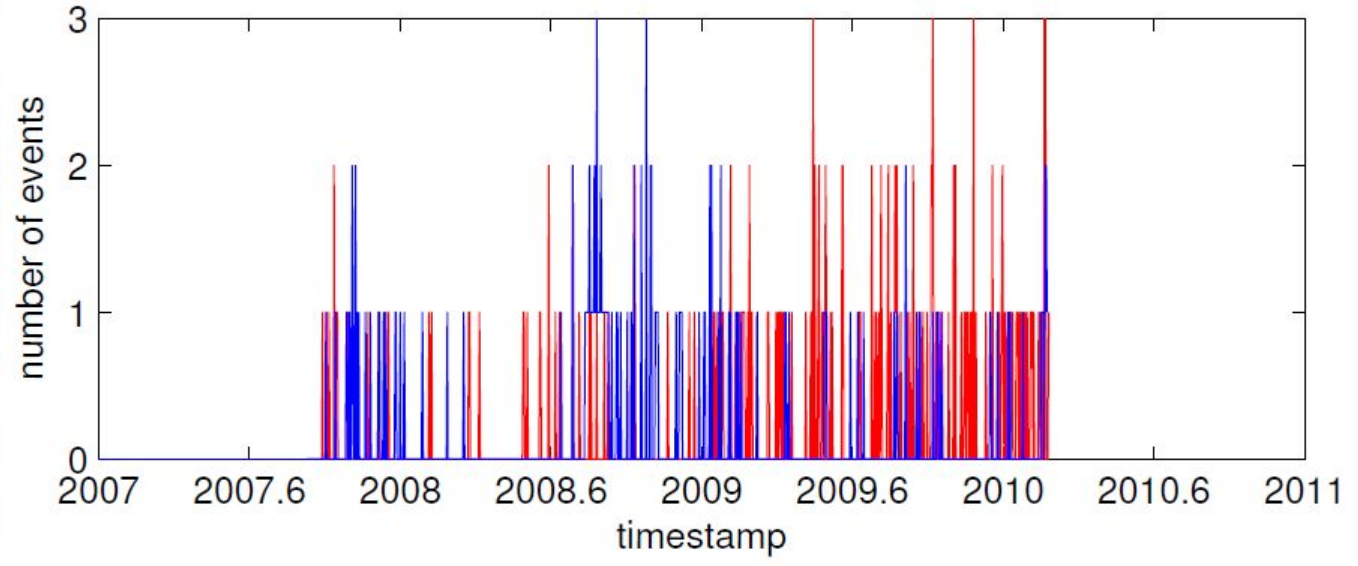
- Expectation–maximization(EM) algorithm [Hegemann, et al., 2013]
  - Variational inference [Li, et al., 2013]
- Additive Model
  - parameterize each actor instead of each actor-pair

$$\mu_m = \mu'_{m1} + \mu'_{m2}, \quad \alpha_m = \alpha'_{m1} + \alpha'_{m2},$$

Scenario - Conflict Data



Self- & mutual-excitation in Conflicts



- One conflict will trigger future conflicts happen between the same actor-pair;
- One conflict will trigger future conflicts that share at least one actor.

Accuracy of Event Attribution

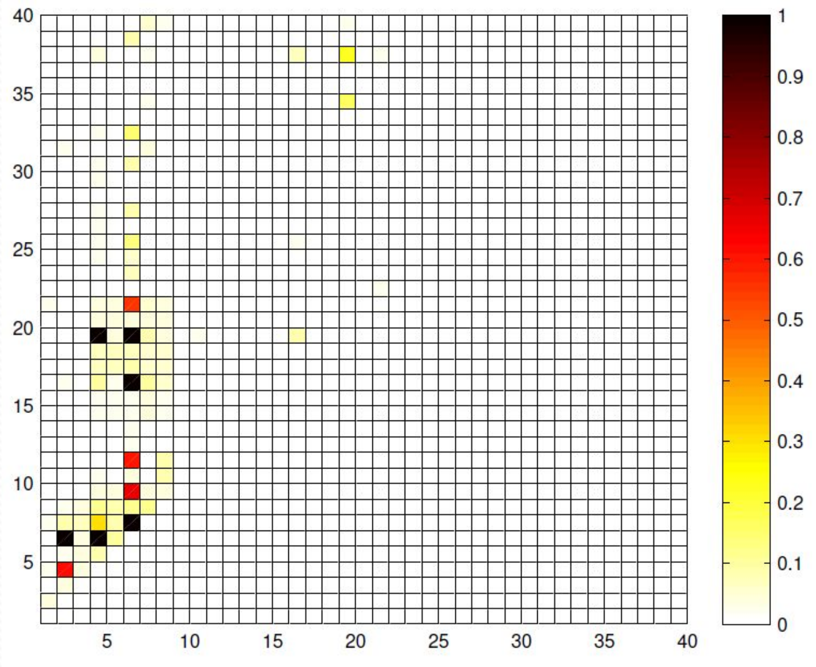
Data set	Method	Top 1	Top 2	Top 3	Top 4	Top 5
Afghanistan	PFHP	11.8%	17.0%	20.2%	21.8%	24.2%
	ESA	12.6%	18.1%	21.3%	23.0%	25.5%
	LPPM	13.4%	20.9%	23.8%	25.4%	26.5%
	MHP	14.6%	23.3%	26.8%	28.6%	30.1%
	AMHP	15.5%	24.0%	27.7%	29.3%	30.8%
Random Guess		0.1%	0.2%	0.3%	0.4%	0.5%
Africa	PFHP	9.0%	14.6%	18.2%	20.0%	22.3%
	ESA	9.9%	15.7%	19.5%	21.3%	23.7%
	LPPM	11.2%	18.7%	21.6%	23.2%	24.4%
	MHP	12.4%	20.9%	24.7%	26.1%	27.5%
	AMHP	13.1%	21.5%	25.4%	26.9%	28.1%
Random Guess		0.1%	0.2%	0.3%	0.4%	0.5%

Afghanistan:
3384 dyadic events
68 actors
1010 actor-pairs

Africa:
52605 dyadic events
3537 actors
1007 actor-pairs

- The underlying dependency network of actor-pairs in real-world data has some special structures.

Relational Graph



Indice of important actors:

4-Civilans
6-Taliban
7-Afghanistan Army
9-Britain Army
11-Afghan Government
16-Police Force
19-ISAF

- Relational graph among actors in Afghanistan Conflict data
- Most sequential conflicts in Afghanistan happened between limited actor-pairs.

Outline

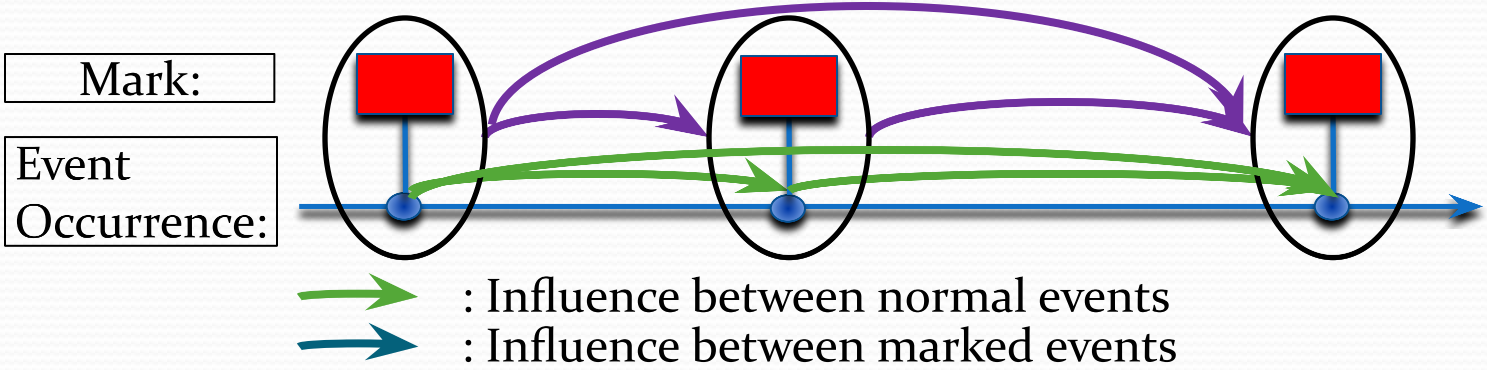
- Typical real-world applications via TPP
- Dyadic Event in temporal point process
- Marked Event in temporal point process
- Cross-domain Event in temporal point process
- Parametric influence in temporal point process

Marked Event

- Mark: detailed information of the corresponding event other than the temporal information.
- Marks can also affect the influence between events.

Event	Mark
Conflict	Casualty
Earthquake	Magnitude
Appliance usage	Consumed energy
Search	Query

Influence Between Marked Events



- How the occurrence and the mark of an event *together* influence the occurrence and the mark of subsequent events in the near future.

Marked Hawkes Processes

- Enables the modeling of the influence between marked events

$$\lambda_t^i = \mu^i + \sum_{m=1}^{M} \phi^{i,k_m}(t - t_m, \xi_m),$$

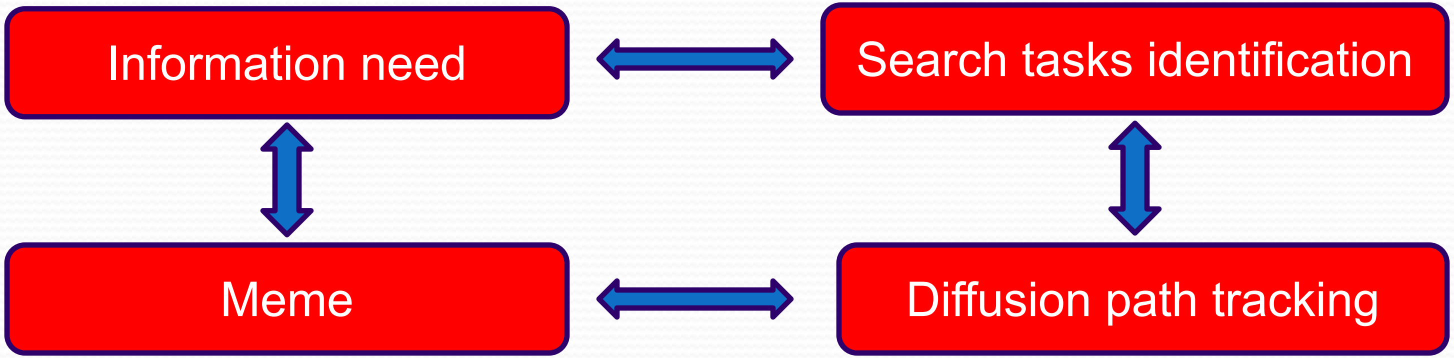
- Directly modeling the relationship between marks and occurrences of different events is difficult
- Assumes a factorized form for the effect of the marks. [Bacry. Et al. 2015]

$$\phi^{ij}(t,\xi) = \phi^{ij}(t)\chi^{ij}(\xi).$$

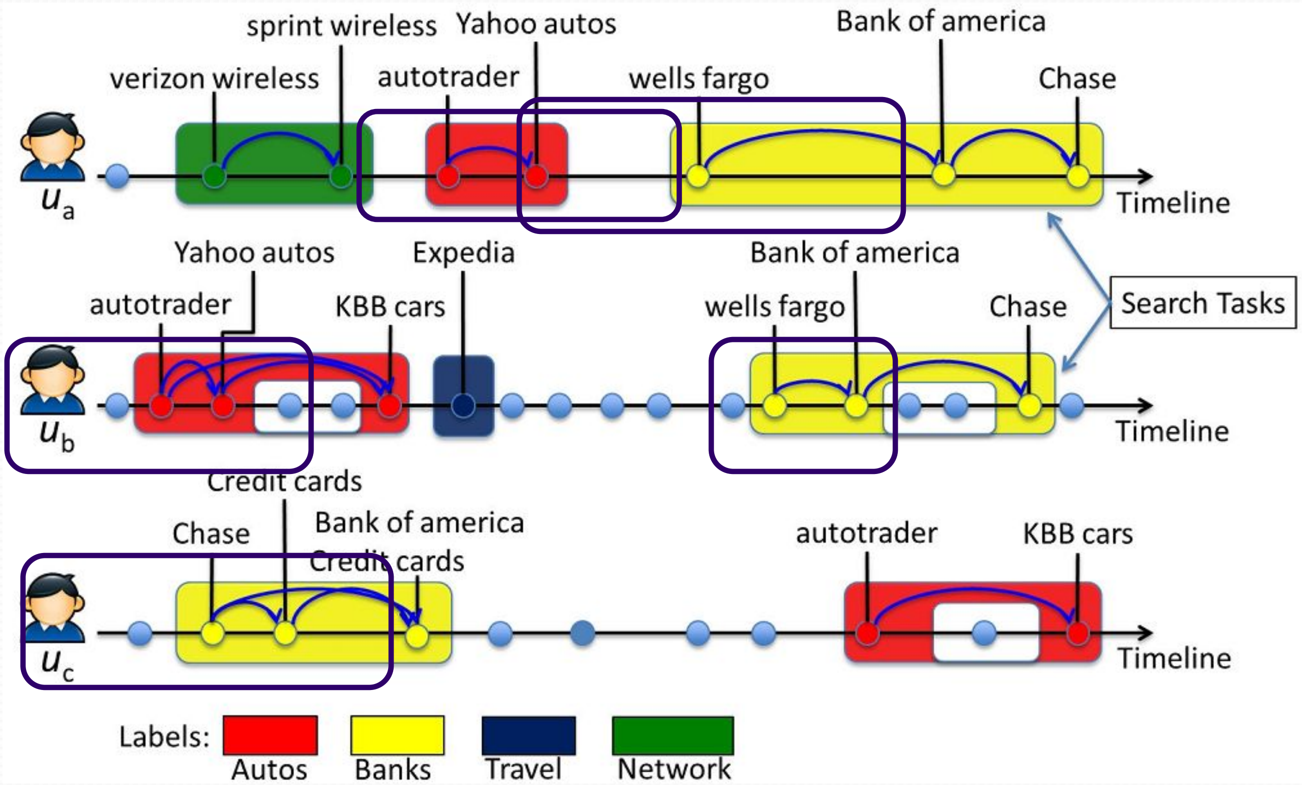
- Utilize mark to better describe the existence and degree of influence

Scenario - Search Task Identification

- Search task
  - A set of queries serving for the same information need.
- Challenge
  - Intertwined multiple intents in a user's query sequence.
- Solution



Consecutive Queries vs Search Tasks



- Consecutive or temporally-close queries issued many times are more like semantically related, i.e., belong to one search task.

Influence in Search Task

- Influence
  - The occurrence of one query raises the probability that the other query will be issued in the near future.

Temporally close query co-occurrence

User's unique query submission frequency

Temporally regular query co-occurrence

$$\lambda_m(t) = \mu_m + \sum_{t_{m,l} < t} R_{m,l,n} \beta_m \kappa(t - t_{m,l})$$

- Issues:
  - Not all temporally-close query-pairs have the actual influence in between.
  - Intractable to obtain an optimal solution of influence existence.

Semantic Influence

- Concentrate on the influence existence between semantically related queries.
- Casting both influence existence and query-topic membership into latent variables.

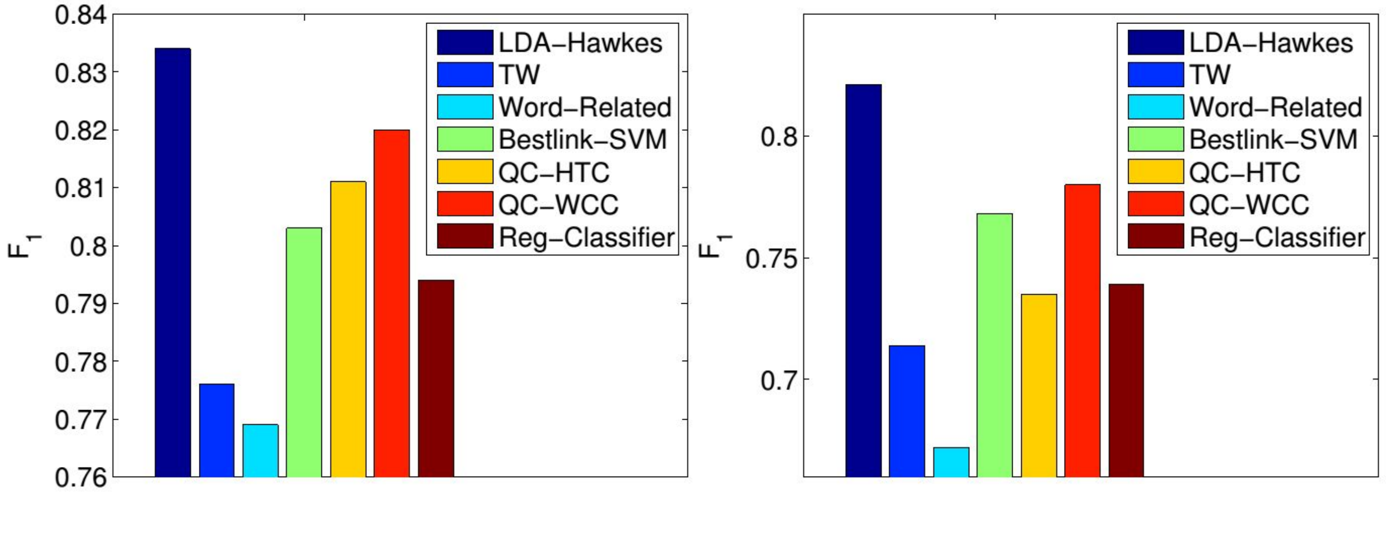
$$R_{m,n,n'} = Y_{m,n}^T * Y_{m,n'}$$

The existence probability of pairwise influence ⟺ The similarity of the memberships of two queries

Hawkes ⟺ LDA

Temporal Information ⟺ Textual Information

Experiments



(a) AOL (b) Yahoo

- Annotated search tasks in AOL & Yahoo.
- LDA-Hawkes > QC > SVM,Reg-Classifier > TW, W-R

Scenario - Energy Disaggregation

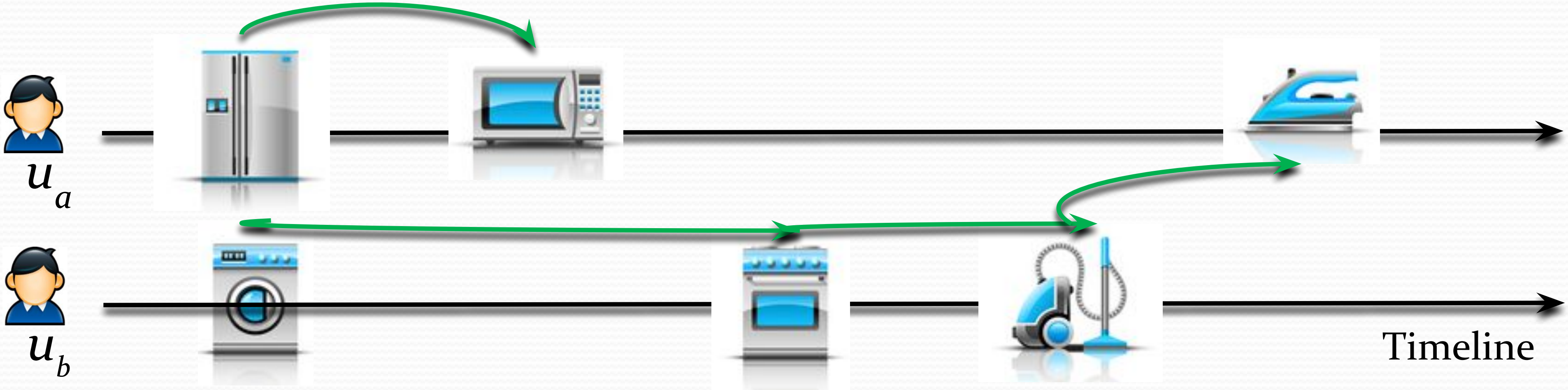
- Energy disaggregation
  - Take a whole home electricity signal and decompose it into its component appliances.
  - Essential for energy conservation
- Fine-grained energy consumption data is not readily available
  - Require numerous additional meters installed on individual appliances

User Energy Usage Behavior

- One powerful cue for breaking down the entire household's energy consumption.
  - how users perform their daily routines.
  - how they share the usage of appliances.
  - users' habits in using certain types of appliances.
- Influence between energy usage behaviors is the key to infer the usage amount

Influence in Energy Usage Behavior

- Why influence modeling is important?



- Influence between energy usage behaviors is hard to model directly.
- Instead, model the influence among various appliances across different time slots.

Marked Hawkes Processes

- Combine multivariate Hawkes processes and topic models

$$\lambda_m(t) = \mu_m + \sum_{t_l < t} \sum_{m'} Y_{m',l} \sum_{k,k'} \boxed{Z_{m,n,k}} Z_{m',l,k'} \boxed{\beta_{m,m',k,k'}} \kappa(t - t_l)$$

The category membership

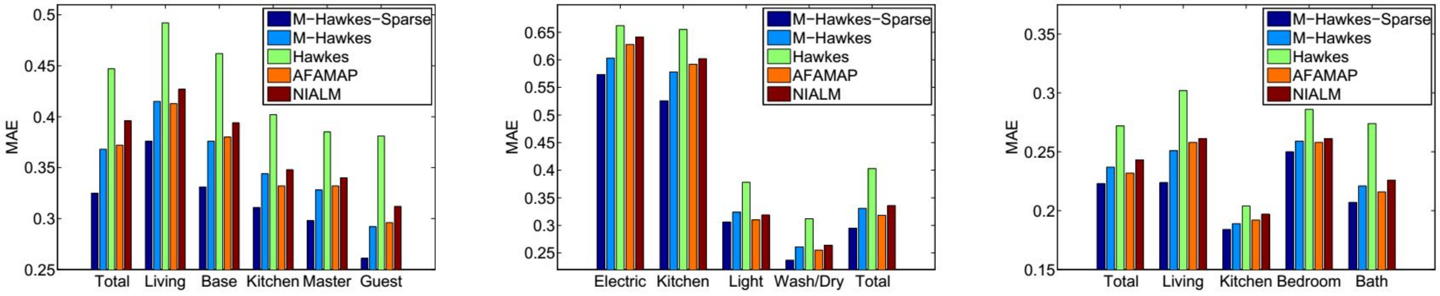
The number of infectivity parameters is O(M²K²)

- Enforce the sparsity of β by imposing lasso type of regularization.

Energy Disaggregation Data Set

- Smart*: 3 homes, 50 appliances
- REDD: 6 homes, 20 appliances
- Pecan: 450+ homes, 20 appliances

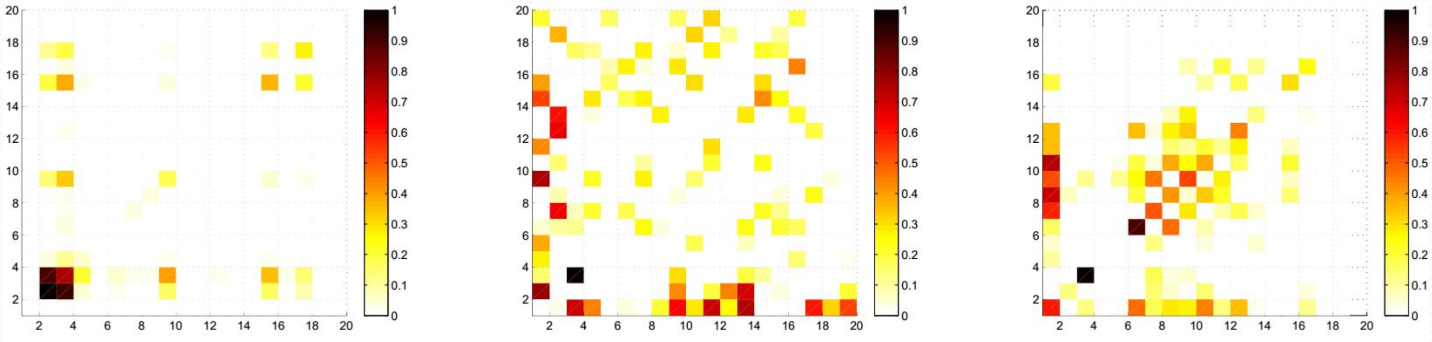
Experiments



(a) Smart*　　(b) REDD　　(c) Pecan

- M-Hawkes-Sparse > M-Hawkes > AFAMP, NIALM > Hawkes
- Only a limited number of dependencies exist between appliances in real world energy consumption.

Energy Usage Pattern



(a) Smart* (b) REDD (c) Pecan

- Smart*: refrigerator-microwave > refrigerator-toaster
- REDD: washer-dryer
- Pecan: refrigerator->microwave > microwave->refrigerator

Outline

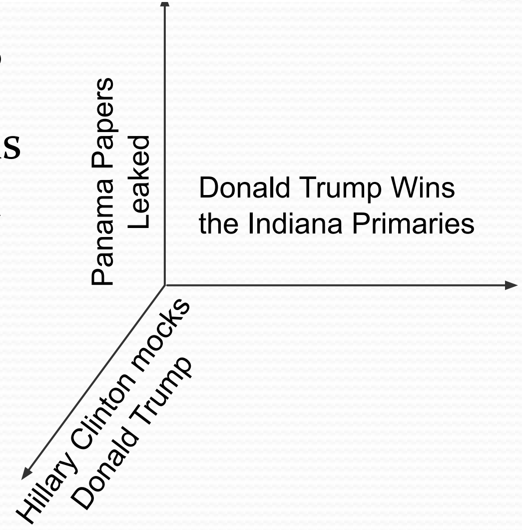
- Typical real-world applications via TPP
- Dyadic Event in temporal point process
- Marked Event in temporal point process
- Cross-domain Event in temporal point process
- Parametric influence in temporal point process

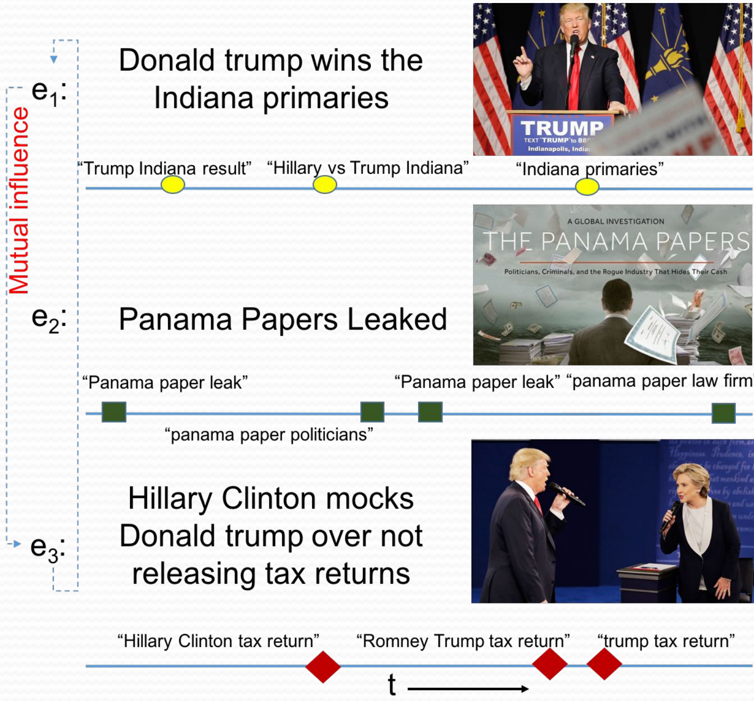
Influence between News and Web Search

$$x_i(W_E, W_q) = \sum_{i=1}^{|W_E|} \frac{\omega(W_{E_i}).IDF(W_{E_i}).TF(W_{E_i}, W_q).(k_1 + 1)}{TF(W_{E_i}, W_q) + k_1.(1 - b + b.\frac{|W_q|}{avgql})}$$

$$\text{subject to} \sum_{i=1}^{|W_E|} \omega(W_{E_i}) = 1$$

- How to influence?
  - Events across domains share the type of data
- Influence type:
  - Single Influence
  - Mutual Influence





Cross-domain Influence TPP Model

- Independent Influence [Santu. et al. 2017]

$$Trend(E,t) = \lambda_0 + \sum_{i=1}^{n} \alpha \cdot TxtSim(W_E, W_i) \cdot e^{-\beta(t-t_i)}$$

- Mutual Influence [Santu. et al. 2018]

Base Influence

Mutual Influence

Decay Function

Impact Function

$$\lambda_j(t) = \eta_j + \sum_{j=1}^{k} v_{ji} \int_{(-\infty,t)\times R} w_j(t-s) g_j(x) e_j(ds \times dx)$$

Data set

Section	Total # of events	Avg. Title Length	Avg. Body Length	Total # of queries	Avg. Textual Sim.
Movies	25	18.88	458.08	193,282	2.49
Sports	15	19.53	508.4	616,449	2.48
US	18	20.38	487.77	204,926	1.99
World	11	18.18	438.81	22,197	1.96

Table 1: Description of Event-Query Joint Dataset

Experiments

Forecast the next most influenced query

Metric	Methods	Movies	Sports	US	World
Accuracy	NF	0.3281	0.4894[2]	0.5717[2]	0.3879
	AR	**0.3879[1]**	0.4794	0.5400	0.4504
	ARD	0.2424	0.1965	0.4410	0.0443
	VAR	0.0023	0.0007	0.0029	0.0001
	IIM	0.3413	0.3660	0.5408	**0.4710[1]**
	JIM	0.3642	0.4688	0.5563	0.3035
	JIM-G	0.3820[2]	**0.5134[1]**	**0.5843[1]**	0.4544[2]

Table 9: Predicting the most frequent query in future

Rank queries based on future influence

Metric	Method	Movies	Sports	US	World
NDCG	NF	0.5914	0.6693	0.8060	0.4465
	AR	0.6713[2]	0.7440[2]	0.7789	0.5200
	ARD	0.2642	0.2977	0.4717	0.0827
	VAR	0.0087	0.0052	0.0136	0.0015
	IIM	0.6355	0.6976	0.8121[2]	**0.6555[1]**
	JIM	0.6484	0.7204	0.8022	0.4809
	JIM-G	**0.6870[1]**	**0.7650[1]**	**0.8430[1]**	0.6062[2]
RBO	NF	0.4349	0.5707	0.6491	0.3665
	AR	0.4947[2]	0.5908[2]	0.6102	0.4130
	ARD	0.1803	0.2191	0.3237	0.0538
	VAR	0.0042	0.0019	0.0045	0.0001
	IIM	0.4562	0.5174	0.6509[2]	**0.4676[1]**
	JIM	0.4782	0.5724	0.6436	0.3048
	JIM-G	**0.5059[1]**	**0.6172[1]**	**0.6764[1]**	0.4332[2]

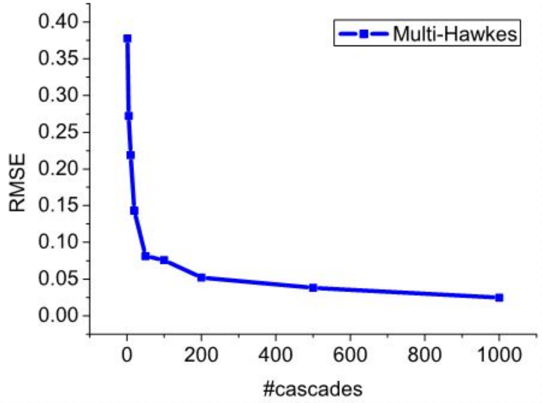
Table 10: Predicting future frequencies for multiple queries. (Wilcoxon's signed rank test at level 0.05)

Outline

- Typical real-world applications via TPP
- Dyadic Event in temporal point process
- Marked Event in temporal point process
- Cross-domain Event in temporal point process
- Parametric influence in temporal point process

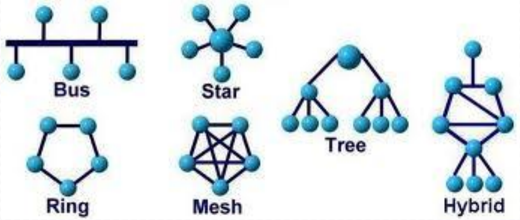
Why Parametric

10 * 10 network



- Problem Complexity
  - $O(M^2)$ α's to learn
    - Hundreds of millions of individuals
  - No sufficient historical events
    - Require multiple cascades
    - The successive event history needs to be segmented into a number of independent cascades in advance.

Why Parametric – cont.



- Dependency in Infectivity Matrix
  - α's are closely related.
  - A priori assumptions on the network topology limit the adaptive social networks of the approaches.
- Time-varying Infectivity
  - Learning separate α for each time interval or with time-dependent function, greatly increase problem complexity.

Parametric Model

- A compact model to parameterize the infectivity between individuals.
- Time-varying features
  - $O(M^2) \longrightarrow$ O(K)
  - Require only one cascades for learning
  - Features incorporate infectivity dependency
    - Simultaneously capture various network topologies
  - Time-varying infectivity

Definition

- For individual-pair (m, m') [Wolfe. et al. 2013, Li. et al. 2014}

$$\alpha_{m,m'} = \beta^T \mathbf{x}_{m,m'}(t)$$

- Optimization problem:

$$\min_{\mu \geq 0, \beta \geq 0} -\mathcal{L}(\mu, \beta) + \boxed{\lambda ||\beta||_1}$$

- Non-differentiable

Select effective features and avoid overfitting

Optimization

- Alternating direction method of multipliers (ADMM)

$$\min_{\mu\geq 0,\beta\geq 0,\mathbf{z}} -\mathcal{L}(\mu,\beta)+\lambda\|\mathbf{z}\|_1, \quad \text{subject to} \quad \beta=\mathbf{z}.$$

ℓ_1

$$\begin{aligned}
\mu^{i+1},\beta^{i+1} &= \operatorname{argmin}_{\mu\geq 0,\beta\geq 0} -\mathcal{L}_\rho(\mu,\beta,\mathbf{z}^i,\mathbf{u}^i),\\
\mathbf{z}^{i+1} &= S_{\lambda/\rho}(\beta^{i+1}+\mathbf{u}^i),\\
\mathbf{u}^{i+1} &= \mathbf{u}^i+\beta^{i+1}-\mathbf{z}^{i+1}.
\end{aligned}$$

ℓ_2

$O(N * K + M)$ Para-Hawkes

- Complexity: $O(N^2 + M^2)$ Multi-dimensional Hawkes

Time-varying Features

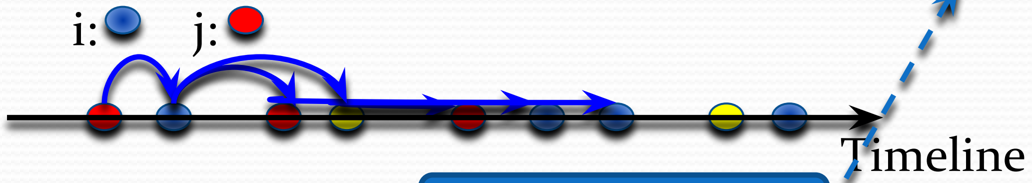
- Individual feature
  - Instant self-property of each individual.
- Dyadic feature
  - Instant relationship between each pair of individuals.

$in(i)$

$i \circlearrowleft^{(v)}$

$i \xrightarrow{(v)} j$

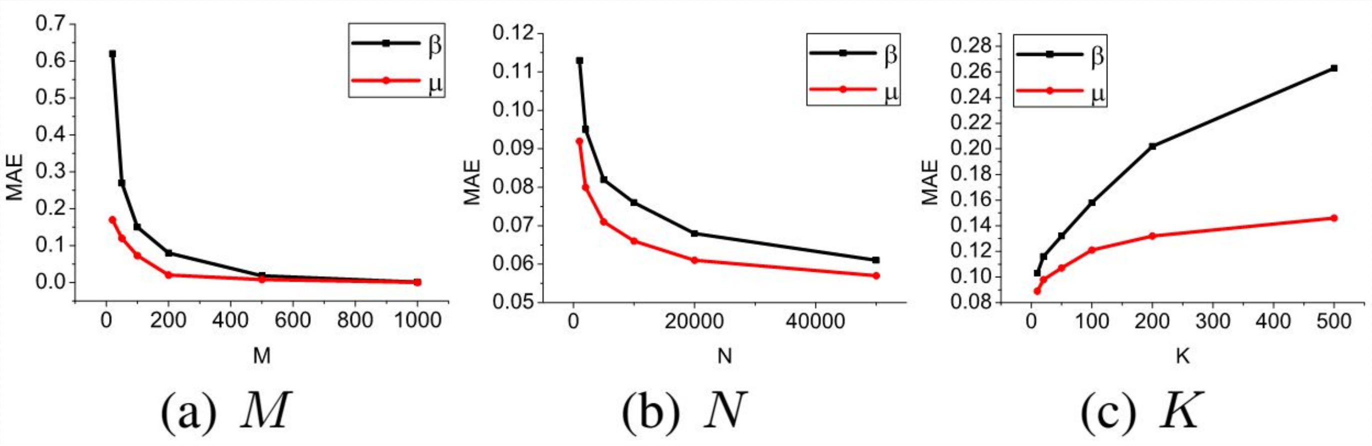
$i \xleftrightarrow{(v,v')} j$



Pattern counting

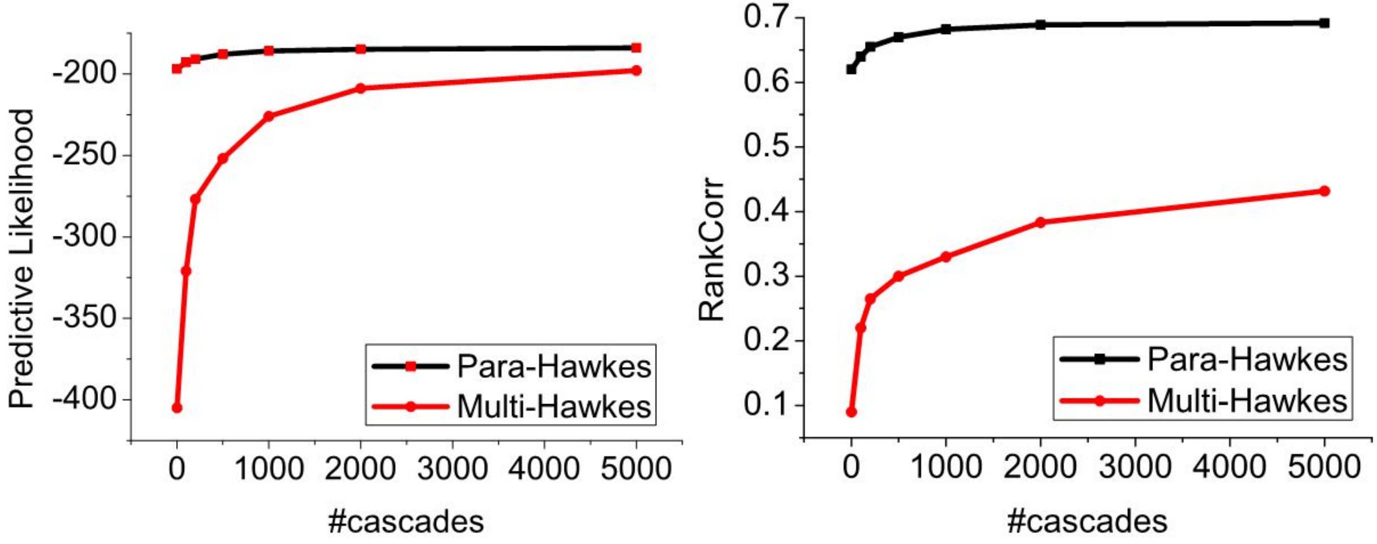
- Formation $\mathbf{x}_{m,m'}(t) = \boxed{\{x(p)(t, \Delta t) | p} \in \mathcal{P}_{m,m'}, \Delta t > 0\}$

Model Dimension Variation



- The impact of model dimension variation on μ is smaller than that on β .

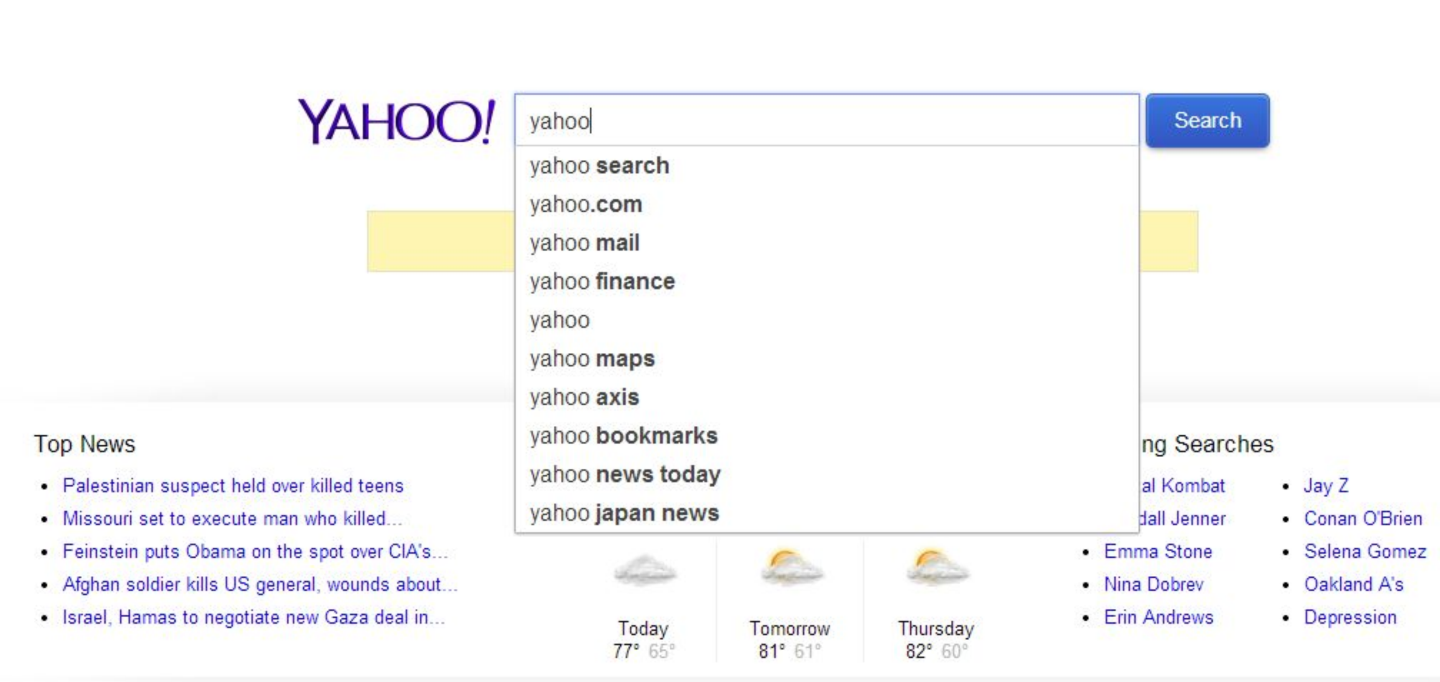
Performance vs #Cascade



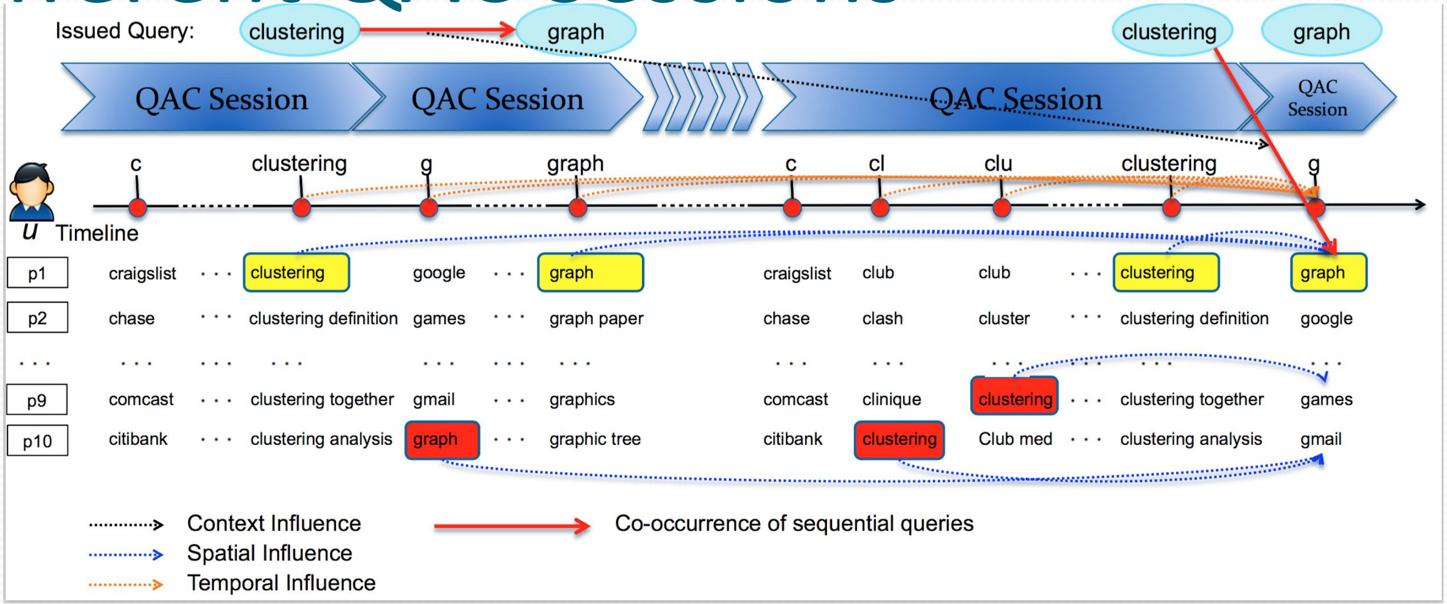
(a) Predictive Likelihood (b) RankCorr

- Works well without multiple cascades

Scenario - Query Auto-Completion



Influence between Click Events in different QAC Sessions



- Influence between users' click choices across different QAC sessions arise from three representative factors:
  - context, position, temporal information.

Factorial Hawkes

- A univariate Hawkes process on each user's issued query sequence.

$$\lambda(t,\mathbf{p}) = \mu + \sum_{t'<t} \beta \underbrace{\mathbf{x}_{q',q}(t)}_{\text{Contextual Factor}}\big(\overbrace{\kappa(t-t')}^{\text{Temporal Factor}} + \alpha\overbrace{\kappa(|\mathbf{p}-\mathbf{p}'|)}^{\text{Spatial Factor}}\big)$$

$$\mathbf{x}_{q',q}(t) = \overbrace{\{x(p)(t,\Delta t)}^{\text{Pattern counting}}|p \in \mathcal{P}_{q',q}, \Delta t > 0\}$$

- A set of contextual features is designed to describe the relationship between the content of a historical query q′ and a current suggested query q.

Query Auto-Completion

Data/Platform	Hawkes	TDCM	RBCM	MPC	UBM	BSS
Measured by MRR@Last						
OldQAC/PC	**0.694**	**0.592**	**0.608**	0.543	0.441	0.545
OldQAC/MB	**0.770**	**0.685**	**0.708**	0.649	0.431	0.650
NewQAC/PC	**0.732**	**0.602**	**0.642**	0.567	0.501	0.552
NewQAC/MB	**0.811**	**0.691**	**0.749**	0.631	0.482	0.654
Measured by MRR@All						
OldQAC/PC	**0.612**	**0.538**	**0.554**	0.464	0.467	**0.531**
OldQAC/MB	**0.671**	**0.611**	**0.629**	0.564	0.471	0.524
NewQAC/PC	**0.664**	**0.578**	**0.602**	0.522	0.508	**0.572**
NewQAC/MB	**0.754**	**0.628**	**0.676**	0.592	0.521	0.554

$$MRR = \frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\text{rank}_q}$$

- RBCM > TDCM > RBCM > MPC, UBM, BBS

Strategy Selection

Data set	T & S & C	T & C	S & C	T & S
Measured by MRR@Last				
OldQAC/PC	0.694	0.658	0.632	0.611
OldQAC/MB	0.770	0.740	0.727	0.720
NewQAC/PC	0.732	0.711	0.691	0.652
NewQAC/MB	0.811	0.798	0.775	0.761
Measured by MRR@All				
OldQAC/PC	0.612	0.588	0.570	0.559
OldQAC/MB	0.671	0.649	0.638	0.634
NewQAC/PC	0.664	0.646	0.625	0.611
NewQAC/MB	0.754	0.719	0.698	0.682

- Factor importance: Context > Temporal > Spatial

Reference

- Rachel A. Hegemann and Erik A. Lewis1 and Andrea L. Bertozzi1An “Estimate & Score Algorithm” for simultaneous parameter estimation and reconstruction of missing data on social networks, 2013
- Liangda Li, and Hongyuan Zha. Dyadic Event Attribution in Social Networks with Mixtures of Hawkes Processes. CIKM 2013.
- Emmanuel Bacry, Iacopo Mastromatteo, Jean-François Muzy, Hawkes processes in finance. 2015.
- Liangda Li, and Hongyuan Zha. Energy Usage Behavior Modeling in Energy Disaggregation via Marked Hawkes Process, AAAI 2015.
- Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, and Hongyuan Zha. Identifying and Labeling Search Tasks via Query-based Hawkes Processes. KDD 2014.
- Patrick O. Perry, and Patrick J. Wolfe. Point process modeling for directed interaction networks. In Journal of the Royal Statistical Society, volume 8, 821–849, 2013
- SK Karmaker Santu, Liangda Li, Dae Hoon Park. Yi Chang, Chenxiang Zha. Modeling the influence of popular trending events on user search behavior. WWW 2017.
- SK Karmaker Santu, Liangda Li, Yi Chang, Chenxiang Zha. Jim: Joint influence modeling for collective search behavior, CIKM 2018.
- Liangda Li, and Hongyuan Zha. Learning Parametric Models for Social Infectivity in Multi-dimensional Hawkes Processes. AAAI 2014.
- Liangda Li, Hongbo Deng, Jianhui Chen, Yi Chang. Learning Parametric Models for Context-Aware Query Auto-Completion via Hawkes Processes. WSDM 2017

Q & A

Thank you!

Appendix