



首届致远学术节 学生科研成果展示

DANCING LINES:

An Analytical Scheme to Depict Cross-Platform Event Popularity

Author: Jinning Li, Tianxiang Gao, et al. Tutor: Xiaofeng Gao

Background:

With the development of Internet technology, the primary media for information propagation have been shifting to online media like social networks, search engines, web portals.

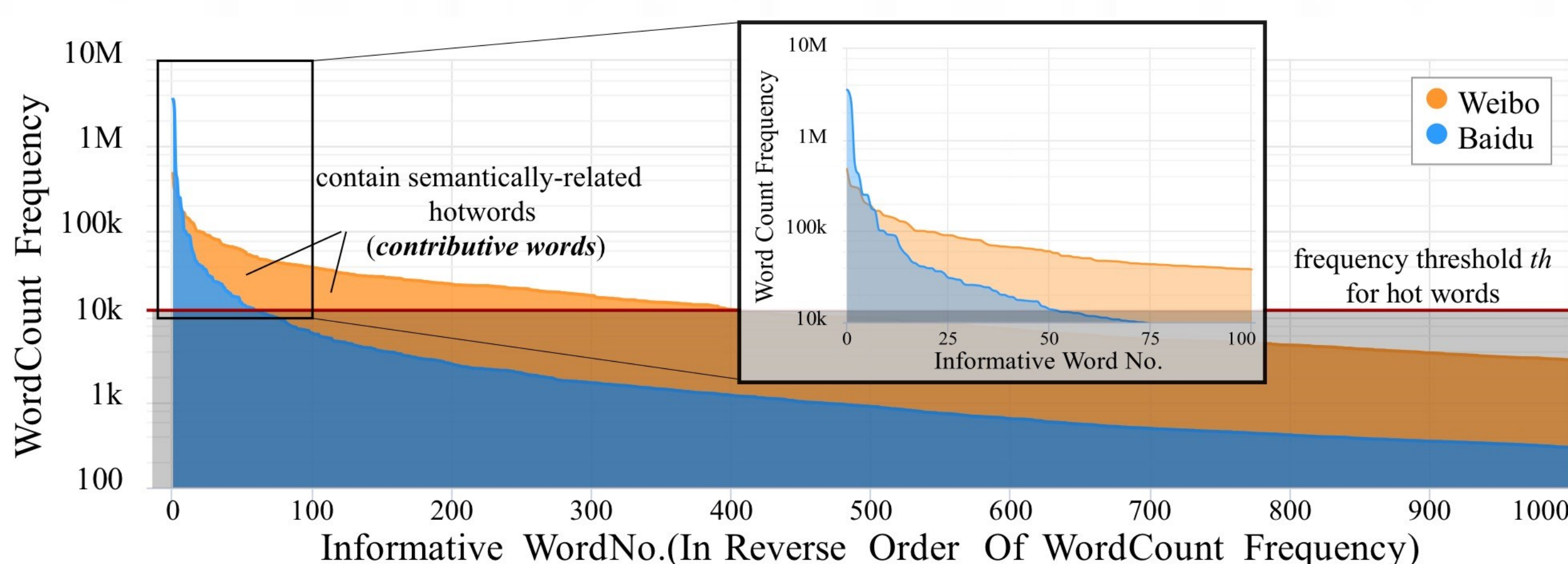
Popular events are usually disseminated on multiple media. Depicting and analyzing event popularity across different platforms plays a vital role in tracking the public concerns and understanding the event disseminations.

Object:

Quantify the *event popularity time series* (EPTS) based on the datasets of Baidu and Weibo. Align and analyze EPTSs of different media and visualize the result.

TF-SW:

TF-SW is a semantic-aware popularity quantification model based on an integrated weight coefficient that leverages Word2Vec and TextRank algorithm.



Discard the words unrelated to certain events

$$sem(w_i, w_j) = \beta \cdot \frac{\mathbf{w}_i^R \cdot \mathbf{w}_j^R}{\|\mathbf{w}_i^R\| \cdot \|\mathbf{w}_j^R\|} + (1 - \beta) \cdot \frac{\mathbf{w}_i^D \cdot \mathbf{w}_j^D}{\|\mathbf{w}_i^D\| \cdot \|\mathbf{w}_j^D\|}$$

$$str(w_i, w_j) = \frac{\sum_{c_l \in w_i \cap w_j} num(c_l, w_i) \cdot num(c_l, w_j)}{\sqrt{\sum_{c_l \in w_i} num(c_l, w_i)^2} \cdot \sqrt{\sum_{c_l \in w_j} num(c_l, w_j)^2}}$$

$$sim(w_i, w_j) = \gamma \cdot sem(w_i, w_j) + (1 - \gamma) \cdot str(w_i, w_j)$$

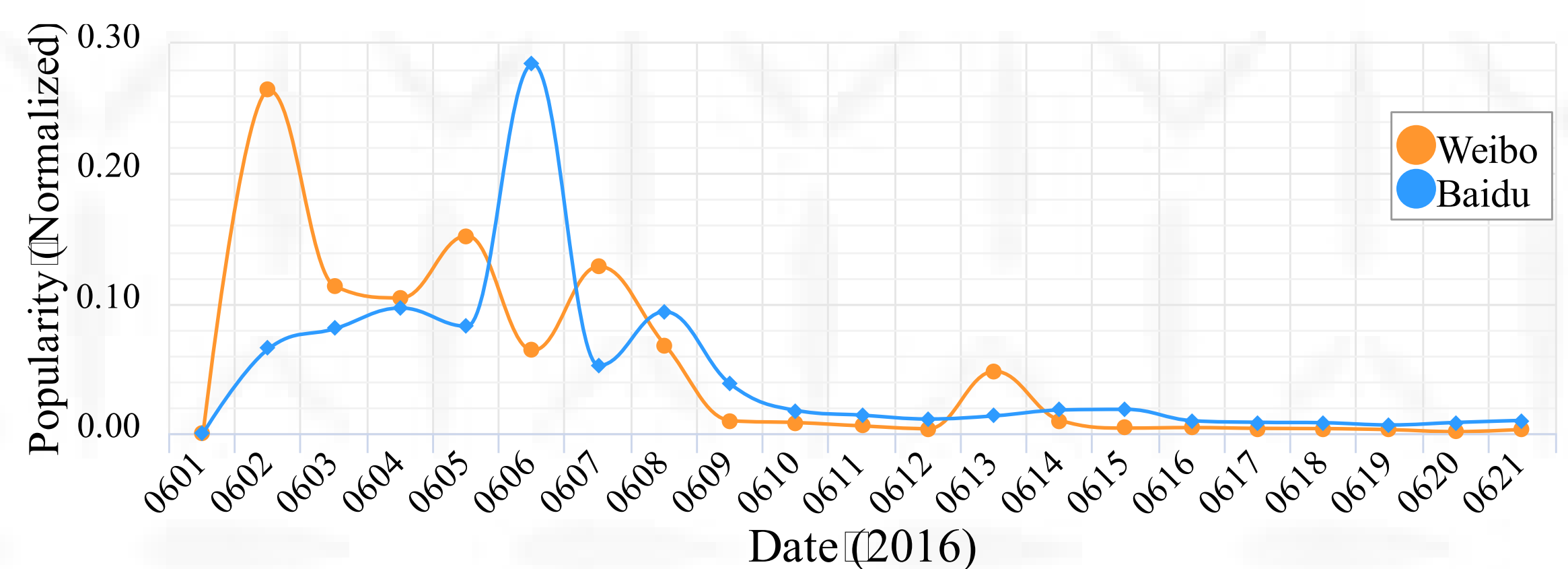
Similarity utilizing semantic and lexical relations

$$TR(w_i) = \frac{1 - \theta}{|\mathcal{E}|} + \theta \cdot \sum_{j \rightarrow i} \frac{sim(w_i, w_j)}{\sum_{k \rightarrow j} sim(w_k, w_j)} \cdot TR(w_j)$$

TextRank gives the importance of each word

$$pop(w_k^i) = fre(w_k^i) \cdot weight(w_k^i)$$

$$weight(w_j^i) = \frac{TR(w_j^i)}{|C_i|} \cdot \sum_{w_k^i \in E_i} fre(w_k^i)$$



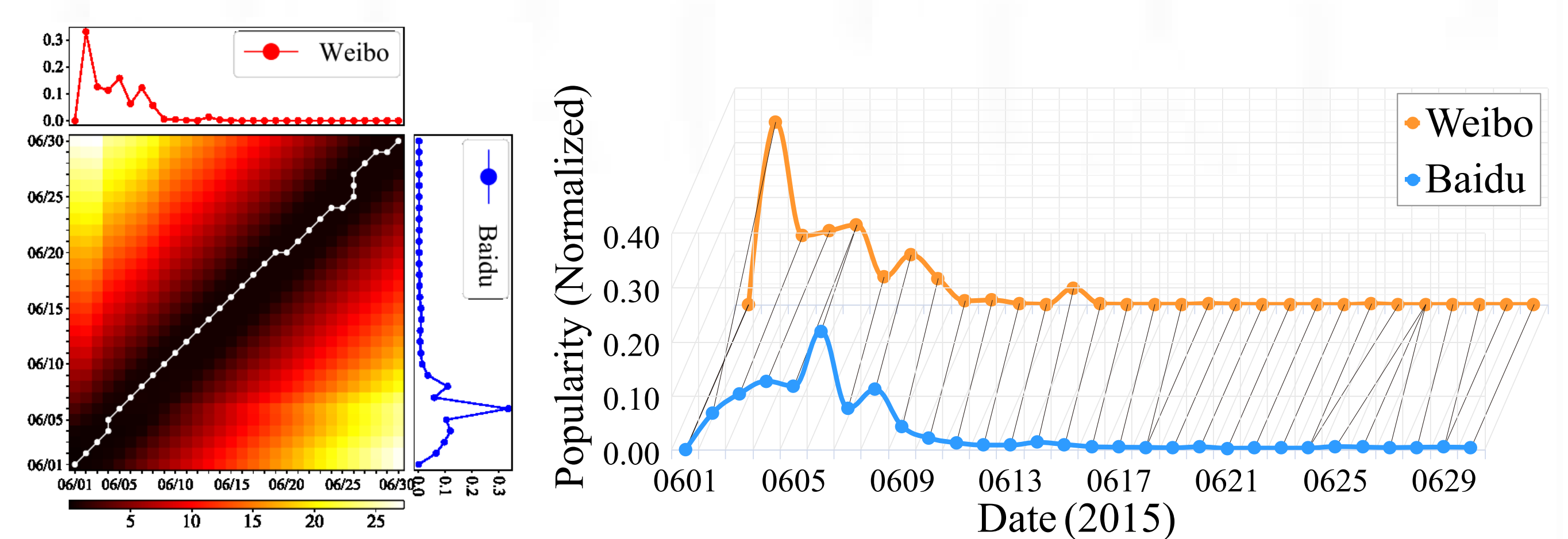
Event popularity time series is generated

ωDTW-CD:

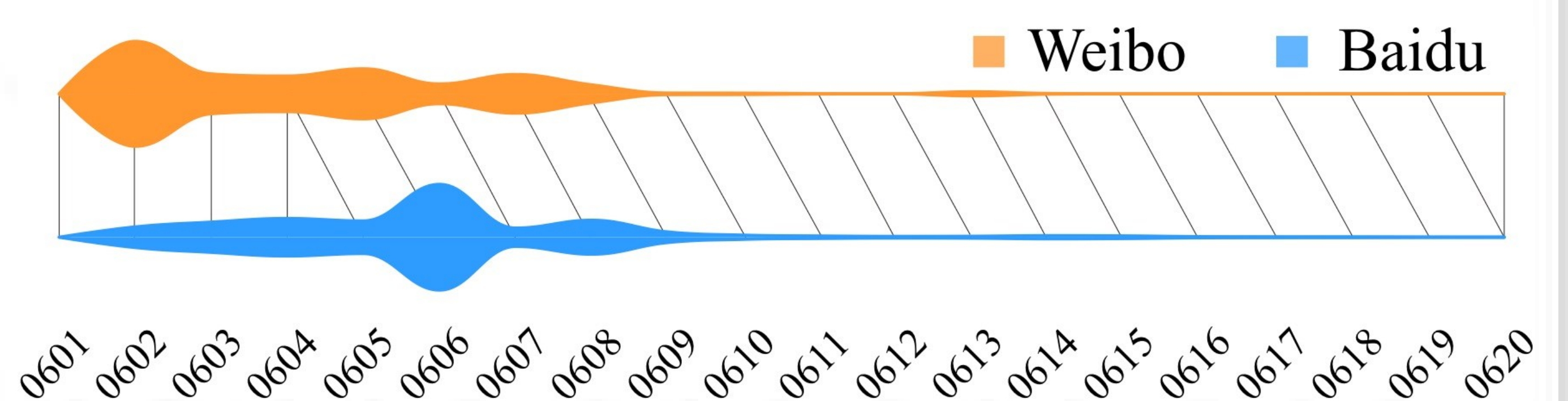
ωDTW-CD is a pairwise EPTSs alignment model using extended Dynamic Time Warping method. It generates matches between the temporally warped EPTSs.

$$dist^C(i, j) = \sqrt[3]{dist^E(i, j) \cdot dist^L(i, j) \cdot dist^D(i, j)}$$

Temporal differences and shapes are considered to avoid the unrealistic far-match and singularity problem.



The DTW heatmap and Aligned EPTSs



Lead-Lag stripes for aligned EPTSs

Experiment:

Our result is proved to be more accuracy and robust than other model on depicting event popularity.

Overall Rank	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6	Event 7	Event 8	Event 9
1	wDTW-CD3	wDTW-CD2	wDDTW	wDTW-CD3	DTW-CD	wDTW-CD3	wDTW-CD3	DDTWbias	DTWbias
2	wDTW-CD1	wDTW	wDTW-CD3	wDTW-CD1	wDTW-CD3	DDTWbias	wDTW	wDTW-CD1	DDTWbias
3	wDTW-CD2	wDTW-CD1	wDTW-CD1	wDTW-CD1	DDTW	DDTWbias	DDTW	wDTW-CD2	wDTW-CD3
4	DDTWbias	DTWbias	wDTW-CD2	wDDTW	DDTWbias	DTW-CD	wDTW-CD1	DTWbias	wDTW
5	wDTW	wDDTW	DTWbias	wDTW-CD2	wDDTW	wDTW-CD2	wDTW-CD1	DDTWbias	wDTW-CD1
6	wDDTW	wDTW-CD3	DDTWbias	wDTWbias	wDTWbias	DTWbias	wDTW-CD1	wDTW	wDDTW
7	DTWbias	DDTWbias	wDTW	wDTW	wDTW-CD2	wDTW-CD2	wDDTW	wDTWbias	wDTW-CD2
8	DDTW	DDTW	DDTW	DTW-CD	DDTW	DTWbias	DTW	DDTW	DTW-CD
9	DTW-CD	DTW	DTW-CD	DTWbias	wDTW	DTW	wDTW-CD2	DTW	DDTW
10	DTW	DTW-CD	DTW	DTW	DTW	DTW-CD	DTW	DTW-CD	DTW

Keywords: Cross Platform, Event Popularity, EPTS, Word2Vec, TextRank, DTW, Visualization

个人信息: 2015级, 计算机科学
邮箱: lijinning@sjtu.edu.cn