EARTH: Accelerating Spatiotemporal Network K-function-based Analytics

Tsz Nam Chan
Shenzhen University
edisonchan@szu.edu.cn

Hongwei Ye
Shenzhen University
yehongwei2023@email.szu.edu.cn

Leong Hou U
University of Macau
ryanlhu@um.edu.mo

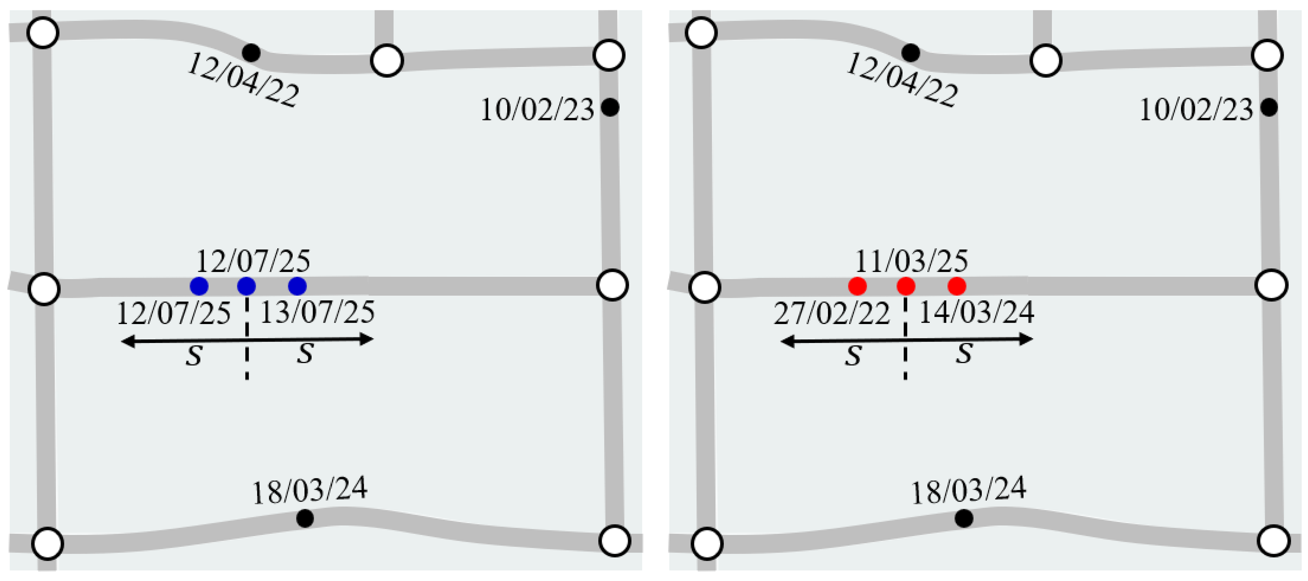
Yun Peng
Guangzhou University
yunpeng@gzhu.edu.cn

Dingming Wu
Shenzhen University
dingming@szu.edu.cn

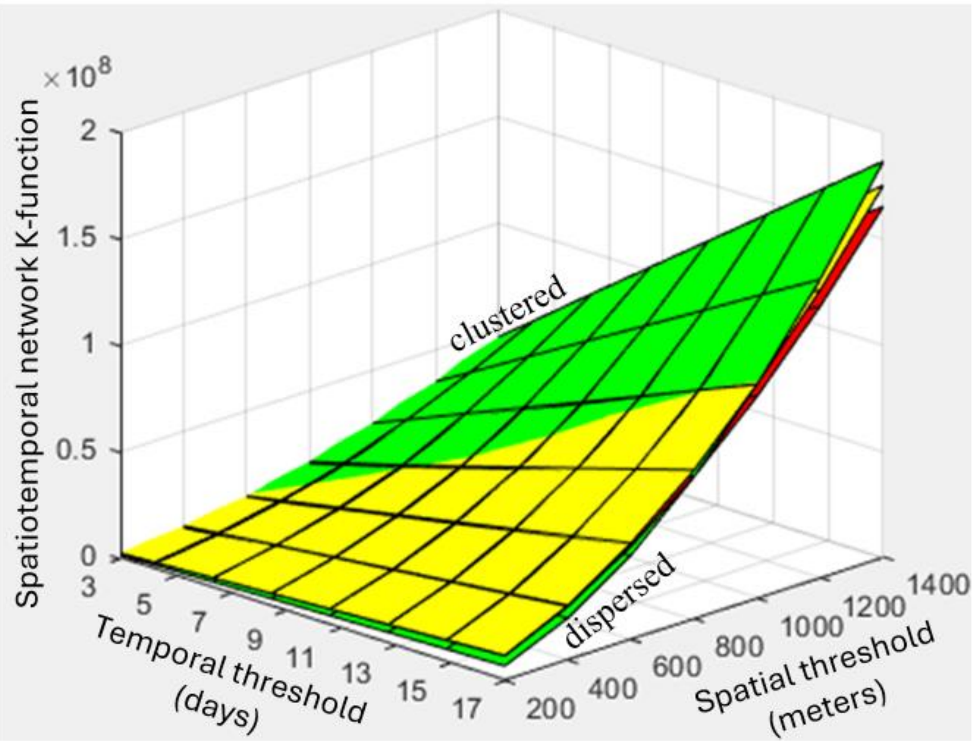
Jianliang Xu
Hong Kong Baptist University
xujl@comp.hkbu.edu.hk

Christian S. Jensen
Aalborg University
csj@cs.aau.dk

What is Spatiotemporal Network K-function-based Analytics?



(a) Dispersed (b) Clustered



(c) Spatiotemporal network K-function plot

- Count the number of data points that are within the spatial threshold s and the temporal threshold t from each data point.

- Spatiotemporal network K-function:
  - A spatiotemporal dataset $P = \{(\mathrm{p}_1, t_{\mathrm{p}_1}), (\mathrm{p}_2, t_{\mathrm{p}_2}), \dots, (\mathrm{p}_n, t_{\mathrm{p}_n})\}$ with size n on a road network $G = (V, E)$
  - Spatial threshold s
  - Temporal threshold t

$$K_P(s,t) = \sum_{(\mathrm{p}_i, t_{\mathrm{p}_i}) \in P} \sum_{\substack{(\mathrm{p}_j, t_{\mathrm{p}_j}) \in P \\ j \neq i}} \mathfrak{I}\left(d_G(\mathrm{p}_i, \mathrm{p}_j) \leq s, d\left(t_{\mathrm{p}_1}, t_{\mathrm{p}_j}\right) \leq t\right)$$

- Spatiotemporal network K-function plot:
  - A spatiotemporal dataset P and L randomly generated datasets $R_1, R_2, \dots, R_L$ (with the same size n)
  - M spatial thresholds $s_1, s_2, \dots, s_M$
  - T temporal thresholds $t_1, t_2, \dots, t_T$
  - Compute $K_P(s_\alpha, t_\beta)$ (green plane), $\mathcal{L}(s_\alpha, t_\beta)$ (red plane), and $\mathcal{U}(s_\alpha, t_\beta)$ (yellow plane)

$$\mathcal{L}(s_\alpha, t_\beta) = \min\left(K_{R_1}(s_\alpha, t_\beta), K_{R_2}(s_\alpha, t_\beta), \dots, K_{R_L}(s_\alpha, t_\beta)\right)$$

$$\mathcal{U}(s_\alpha, t_\beta) = \max\left(K_{R_1}(s_\alpha, t_\beta), K_{R_2}(s_\alpha, t_\beta), \dots, K_{R_L}(s_\alpha, t_\beta)\right)$$

Very Slow!

- Applications
  - Determine whether the spatiotemporal clusters/hotspots are meaningful.
  - Help for choosing the correct parameters for those statistical/clustering models.

Our Solution: EARTH

- Decomposition of $K_P(s,t)$

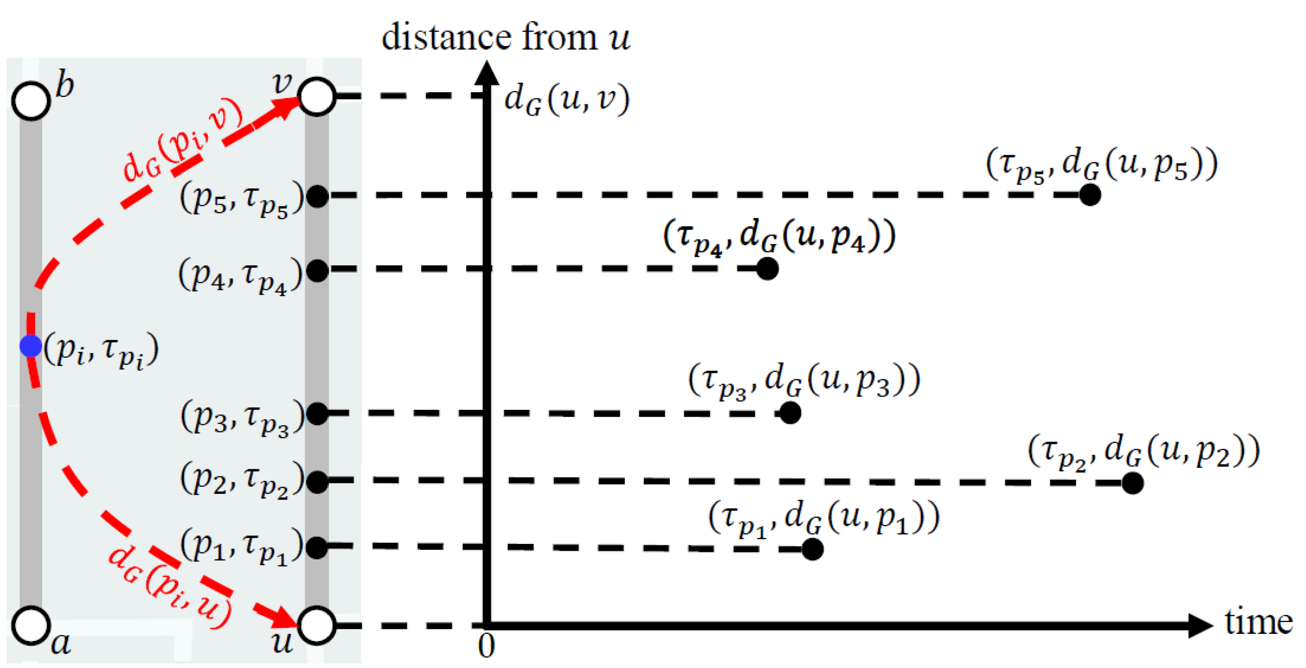
$$K_P(s,t) = \sum_{(\mathrm{p}_i, t_{\mathrm{p}_i}) \in P} \sum_{e \in E} C_{P(e)}^{(\mathrm{p}_i, t_{\mathrm{p}_i})}(s,t)$$

where

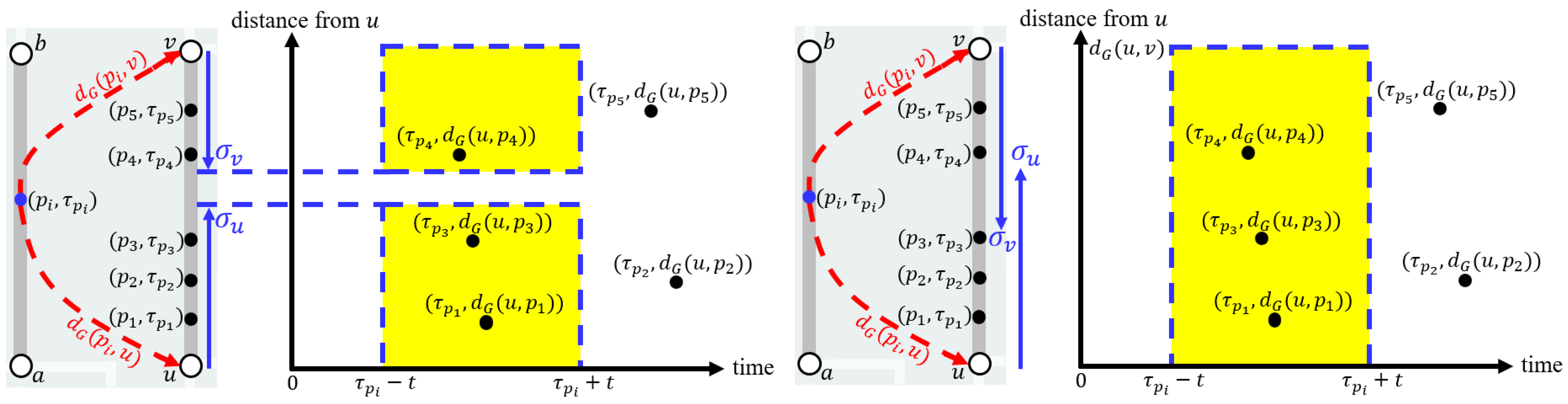
$$C_{P(e)}^{(\mathrm{p}_i, t_{\mathrm{p}_i})}(s,t) = \sum_{\substack{(\mathrm{p}_j, t_{\mathrm{p}_j}) \in P(e) \\ j \neq i}} \mathfrak{I}\left(d_G(\mathrm{p}_i, \mathrm{p}_j) \leq s, d\left(t_{\mathrm{p}_i}, t_{\mathrm{p}_j}\right) \leq t\right)$$

- Computing $C_{P(e)}^{(\mathrm{p}_i, t_{\mathrm{p}_i})}(s,t)$ takes $O(|P(e)|)$ time. ☹

- Core idea: Mapping data points for each edge $e = (u, v)$ onto a two-dimensional (time, distance)-plane.

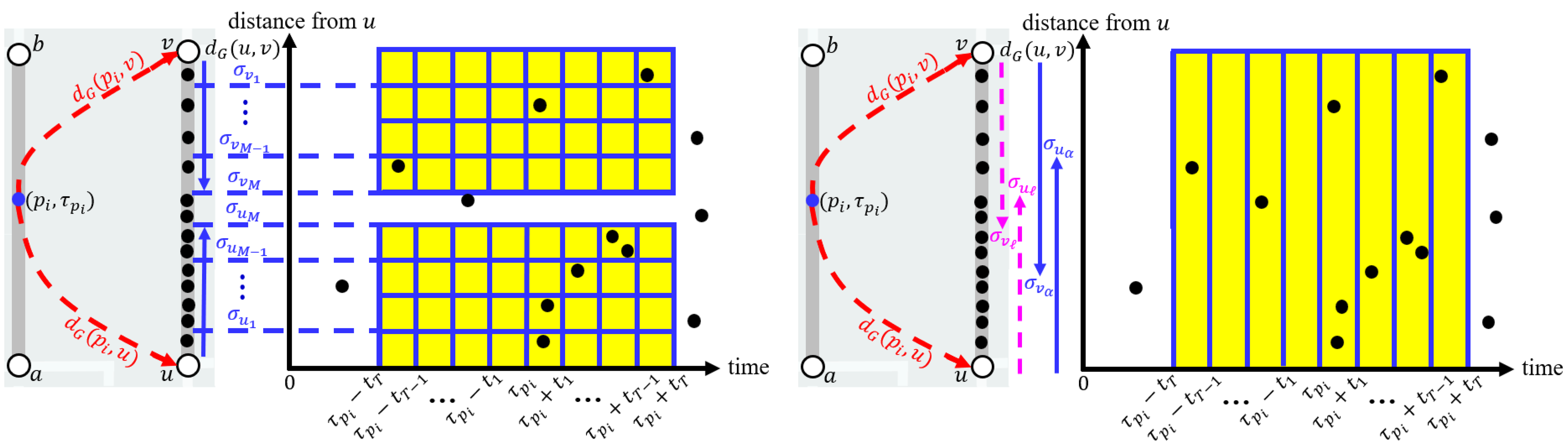


- Edge-Augmentation with Range-tree (EAR) $O(\log |P(e)|)$ time ☺



(a) $\sigma_u + \sigma_v < d_G(u,v)$ (b) $\sigma_u + \sigma_v \geq d_G(u,v)$

- Multi-Threshold Sharing (MTS) $O(MT + |P(e)|)$ time ☺



(a) $\sigma_{u_M} + \sigma_{v_M} < d_G(u,v)$ (b) $\sigma_{u_M} + \sigma_{v_M} \geq d_G(u,v)$

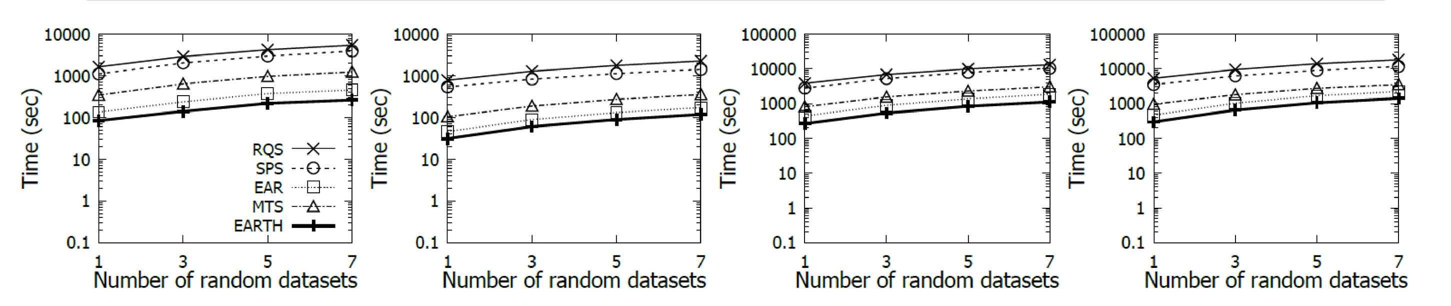
- EARTH = Edge-Augmentation with Range-tree + multi-THreshold sharing

Theoretical Performance

Problem	Method	Time complexity	Space complexity
Spatiotemporal network K-function	RQS [54]	$O(n(\mathcal{T}_{\mathrm{SP}} + n))$	$O(\mathcal{S}_{\mathrm{SP}} + n)$
	SPS [55]	$O(\|E\|\mathcal{T}_{\mathrm{SP}} + n^2)$	
	EAR (Section 4.1)	$O(\|E\|\mathcal{T}_{\mathrm{SP}} + n\|E\| \log\left(\frac{n}{\|E\|}\right) + n \log n)$	$O(\mathcal{S}_{\mathrm{SP}} + n)$ (Section 4.4)
Spatiotemporal network K-function plot	RQS [54]	$O(LMTn(\mathcal{T}_{\mathrm{SP}} + n))$	$O(\mathcal{S}_{\mathrm{SP}} + Ln + LMT)$
	SPS [55]	$O(LMT(\|E\|\mathcal{T}_{\mathrm{SP}} + n^2))$	
	EAR (Section 4.1)	$O(\|E\|\mathcal{T}_{\mathrm{SP}} + LMTn\|E\| \log\left(\frac{n}{\|E\|}\right) + Ln \log n)$	$O(\mathcal{S}_{\mathrm{SP}} + Ln + LMT)$ (Section 4.4)
	MTS (Section 4.2)	$O(\|E\|\mathcal{T}_{\mathrm{SP}} + Ln^2 + LMTn\|E\|)$	
	EARTH (Section 4.3)	$O(\|E\|\mathcal{T}_{\mathrm{SP}} + \min\left(LMTn\|E\| \log\left(\frac{n}{\|E\|}\right) + Ln \log n, Ln^2 + LMTn\|E\|\right))$	

Experimental Results

Name	n	$\|V\|$	$\|E\|$	Category	Ref.
Seattle	543,969	11,383	37,472	Crime events	[8]
London	822,382	19,984	75,609	Traffic accidents	[6]
New York	1,542,738	41,307	115,021	Traffic accidents	[3]
Los Angeles	2,108,154	35,933	109,708	Crime events	[2]



(a) Seattle (b) London (c) New York (d) Los Angeles