Fast Network K-function-based Spatial Analysis

Tsz Nam Chan
Hong Kong Baptist University
edisonchan@comp.hkbu.edu.hk

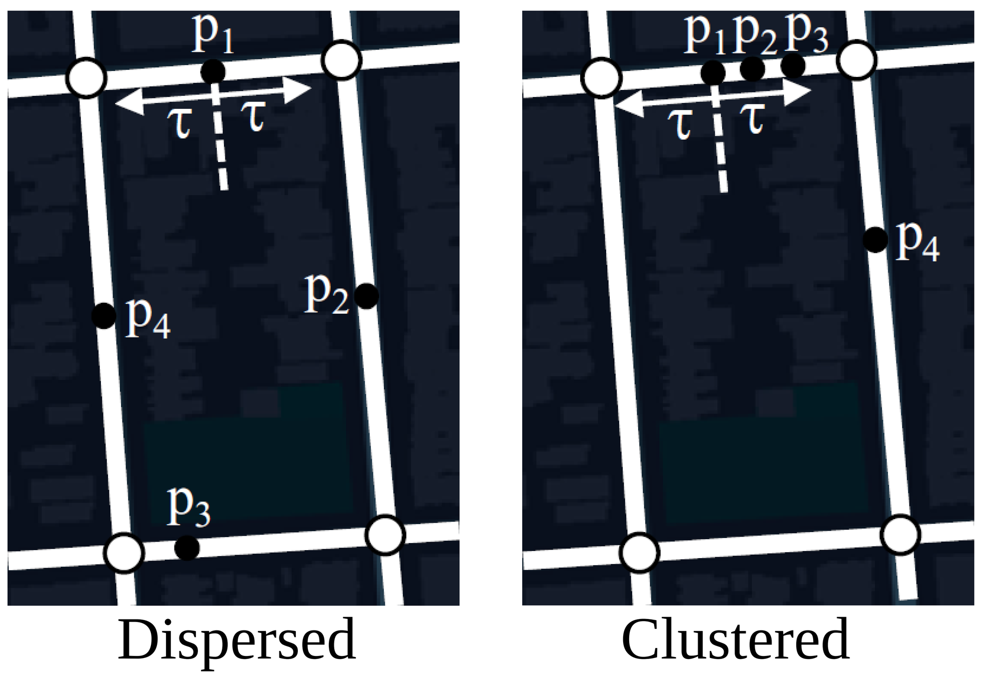
Leong Hou U
University of Macau
SKL of Internet of Things for Smart City
paklonip@um.edu.mo

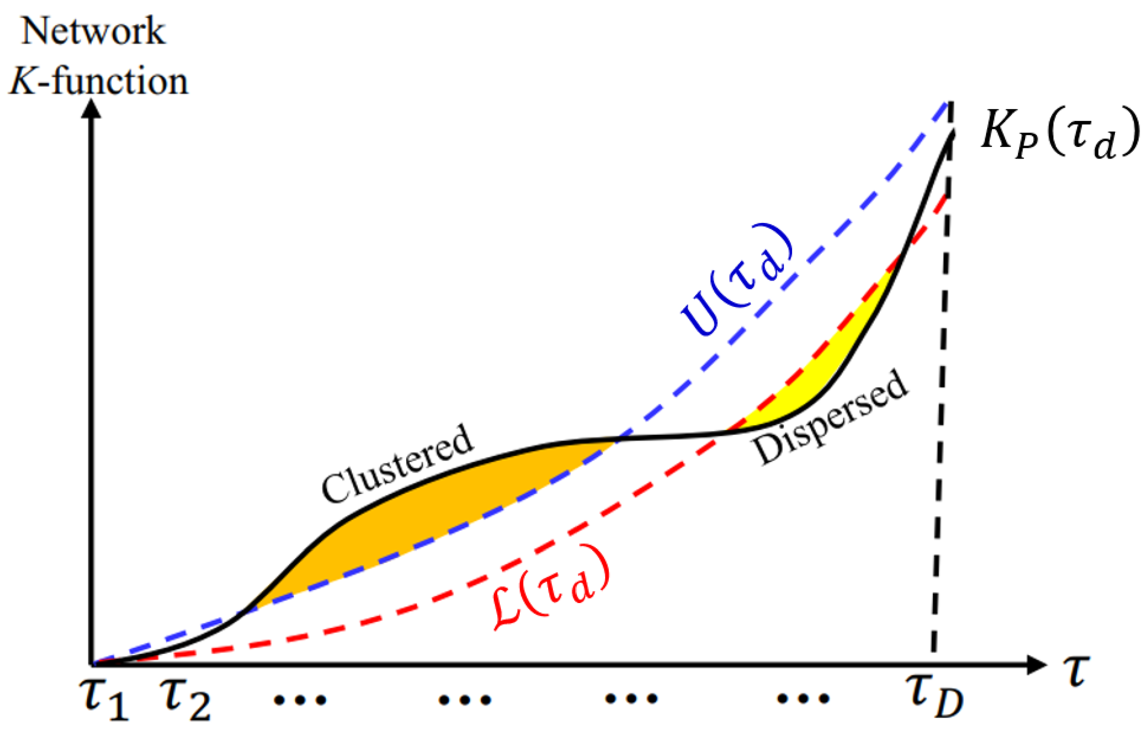
Yun Peng
Institute of Artificial Intelligence and Blockchain, Guangzhou University
yunpeng@gzhu.edu.cn

Byron Choi
Hong Kong Baptist University
bchoi@comp.hkbu.edu.hk

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

Overview of Network K-function





A network K-function plot

- A fundamental GIS operation for checking whether the clusters/hotspots (discovered by the clustering/hotspot detection algorithms) in the datasets are significant.
- Network K-function is defined as:

$$\underbrace{K_P}_{\text{Dataset}}(\tau) = \sum_{p_i \in P} \sum_{\substack{p_j \in P \\ p_j \neq p_i}} \mathbb{I}(\overbrace{dist_{\underbrace{G}_{\text{Road network}}}(p_i, p_j)}^{\text{Shortest path distance}} \leq \underbrace{\tau}_{\text{Threshold}}) \qquad O(n(T_{SP} + n)) \text{ time ☹}$$

where $\mathbb{I}$ denotes an indicator function.

- Domain experts need to:
  1. Provide a road network $G = (V, E)$, a location dataset $P = \{p_1, p_2, \dots, p_n\}$, and D thresholds, which are $\tau_1, \tau_2, \dots, \tau_D$.
  2. Randomly generate L datasets, which are $R_1, R_2, \dots, R_L$, in the road network G.
  3. For each threshold τ_d $(1 \leq d \leq D)$, compute the following three terms.

     (1) $K_P(\tau_d)$

     (2) $\mathcal{L}(\tau_d) = \min\left(K_{R_1}(\tau_d), K_{R_2}(\tau_d), \dots, K_{R_L}(\tau_d)\right)$

     (3) $U(\tau_d) = \max\left(K_{R_1}(\tau_d), K_{R_2}(\tau_d), \dots, K_{R_L}(\tau_d)\right)$
- Generating a network K-function plot takes $O(LDn(T_{SP} + n))$ time ☹

Decompose the Network K-function

Let $P(e)$ be the set of data points in an edge e of a road network.

$$K_P(\tau) = \sum_{p_i \in P} \sum_{\substack{p_j \in P \\ p_j \neq p_i}} \mathbb{I}(dist_G(p_i, p_j) \leq \tau)$$

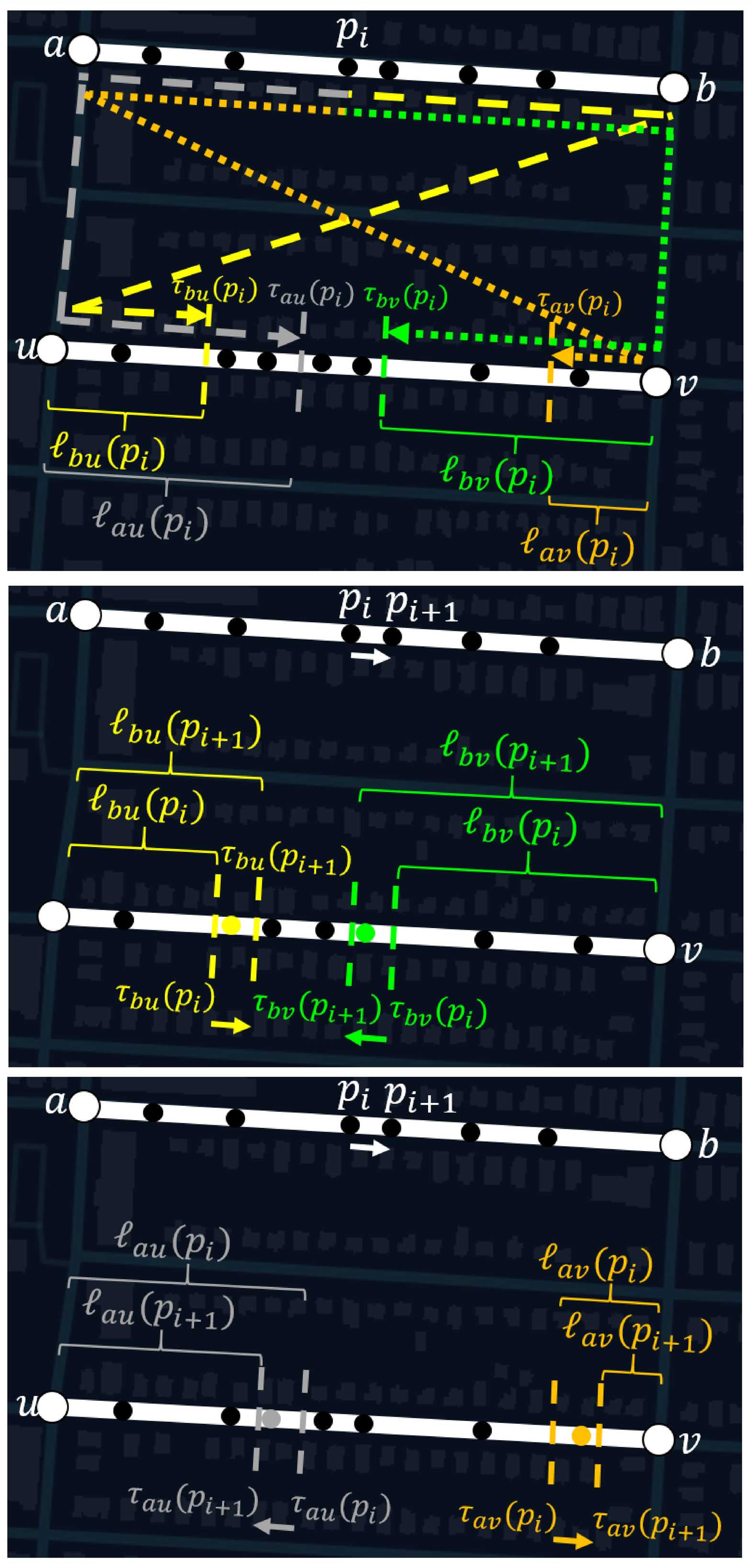
$$= \sum_{\tilde{e} \in E} \sum_{p_i \in P(\tilde{e})} \sum_{e \in E} \sum_{p_j \in P(e)} \mathbb{I}(dist_G(p_i, p_j) \leq \tau)$$

$$= \sum_{\tilde{e} \in E} \sum_{e \in E} C_P^{(\tilde{e},e)}(\tau)$$

where $C_P^{(\tilde{e},e)}(\tau) = \sum_{p_i \in P(\tilde{e})} \sum_{p_j \in P(e)} \mathbb{I}(dist_G(p_i, p_j) \leq \tau)$

Question: Can we reduce the time complexity for computing the $(\tilde{e}, e)$-count function $C_P^{(\tilde{e},e)}(\tau)$ (e.g., from $O(|P(\tilde{e})||P(e)|)$ to $O(|P(\tilde{e})| + |P(e)|)$)?

Neighbor-Sharing (NS) Method



- There are four possible routes with length τ from any data point p_i in the edge $\tilde{e} = (a, b)$ to the edge $e = (u, v)$. We have:

$$\tau_{au}(p_i) = \tau - dist_G(p_i, a) - dist_G(a, u)$$
$$\tau_{bu}(p_i) = \tau - dist_G(p_i, b) - dist_G(b, u)$$
$$\tau_{av}(p_i) = \tau - dist_G(p_i, a) - dist_G(a, v)$$
$$\tau_{bv}(p_i) = \tau - dist_G(p_i, b) - dist_G(b, v)$$

- Maintain four sets of data points in the edge $e = (u, v)$.

$$\ell_{au}(p_i) = \{p_j \in P(e) : dist_G(u, p_j) \leq \tau_{au}(p_i)\}$$
$$\ell_{bu}(p_i) = \{p_j \in P(e) : dist_G(u, p_j) \leq \tau_{bu}(p_i)\}$$
$$\ell_{av}(p_i) = \{p_j \in P(e) : dist_G(v, p_j) \leq \tau_{av}(p_i)\}$$
$$\ell_{bv}(p_i) = \{p_j \in P(e) : dist_G(v, p_j) \leq \tau_{bv}(p_i)\}$$

- These four sets of data points exhibit the monotonicity property.
- Takes $O(|P(\tilde{e})| + |P(e)|)$ time to compute $C_P^{(\tilde{e},e)}(\tau)$ ☺
- Takes $O(LD(|E|T_{SP} + n|E|) + Ln\log n)$ time to generate the network K-function plot (Refer to the paper) ☺

Theoretical Results

Method	Time complexity	Space complexity
RQS [7, 44–46] (cf. Section 2.2)	$O(LDn(T_{SP} + n))$	$O(\lvert V\rvert + \lvert E\rvert + nL + \mathcal{S}_{SP})$
SPS [53] (cf. Section 2.3)	$O(LD(\lvert E\rvert T_{SP} + n^2))$	
CA (cf. Section 3.2)	$O\big(LD(\lvert E\rvert T_{SP} + n\lvert E\rvert \log(\frac{n}{\lvert E\rvert})) + Ln\log n\big)$ (cf. Theorem 1)	$O(\lvert V\rvert + \lvert E\rvert + nL + \mathcal{S}_{SP})$ (cf. Theorem 4)
NS (cf. Section 3.3)	$O(LD(\lvert E\rvert T_{SP} + n\lvert E\rvert) + Ln\log n)$ (cf. Theorem 2)	
$CA^{(ASPS)}$ (cf. Sections 3.2 and 3.4)	$O\big(\lvert E\rvert T_{SP} + nLD\lvert E\rvert \log(\frac{n}{\lvert E\rvert}) + Ln\log n\big)$ (cf. Theorem 3)	
$NS^{(ASPS)}$ (cf. Sections 3.3 and 3.4)	$O(\lvert E\rvert T_{SP} + nLD\lvert E\rvert + Ln\log n)$ (cf. Theorem 3)	

Experimental Results