Topic 5: Network Constrained Spatial Point Patterns Analysis (NetSPPA)

Dr. Kam Tin Seong
Assoc. Professor of Information Systems (Practice)

School of Computing and Information Systems,
Singapore Management University

2021-5-5 (updated: 2021-05-30)

Content

- Motivation
- Network Constrained Kernel Density Estimation (NetKDE)
- Network Constrained 2nd-Order Point Patterns Analysis
  - Network Constrained K-function
  - Network Constrained G-function
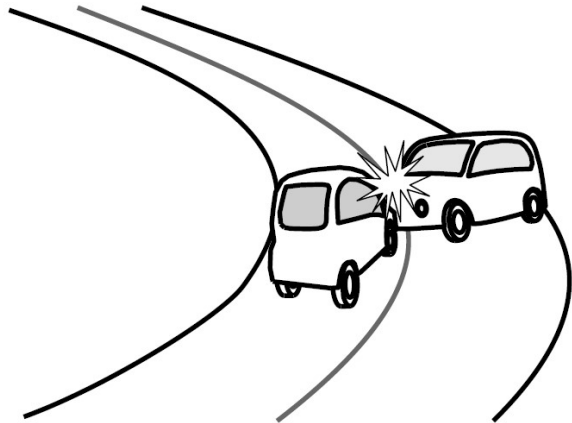
Motivation

- Real world geography are heterogeneous and anisotropy.

What is Network Constrained events

network (constrained) events

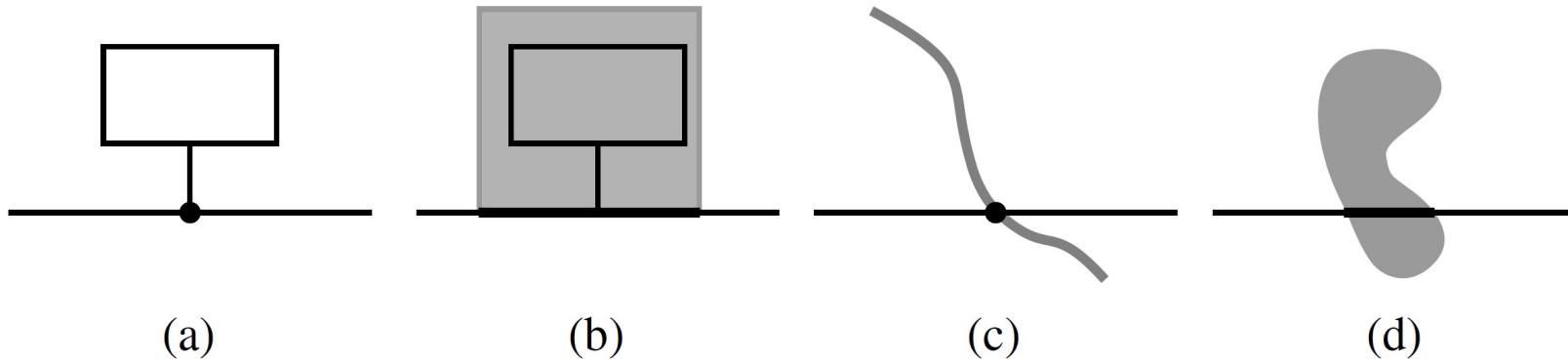on-network events   alongside-network events

Alongside-network relations

- (a) an access point (the black circle) of a polygon to a network (the horizontal line segment),
- (b) a boundary line segment of a polygon shared with a network (the bold line segment)
- (c) an intersection point of two networks (the black circles),
- (d) a network intersecting an area (the bold line segment).

(a)   (b)   (c)   (d)

Real world network constrained events

- A traffic accident events at downtown area.
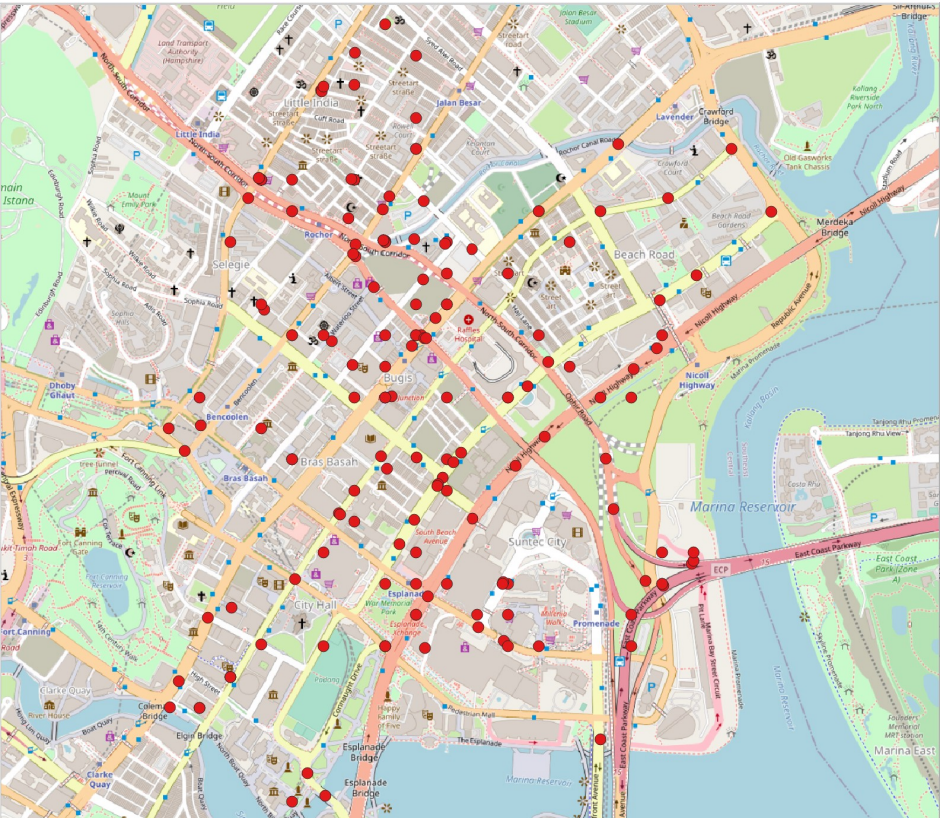

- Distribution of Airbnb listing at Central region.

Examples of real world questions

Typical network constrained events questionsare as follows:

- How can we obtain the catchment areas of parking lots in a downtown area including one-way streets, assuming that drivers access their nearest parking lots?
- Do coffee outlets tend to stand side-by-side alongside streets in a downtown area?
- Do Airbnb listings tend to locate near to MRT stations?
- Is the roadside retail rental price of a street segment similar to those of the adjacent street segments?
- How can we locate clusters of childcare centres within HDB towns?
- How can we estimate the density of traffic accidents and street crimes incidence, and how can we identify locations where the densities of those occurrence are high, referred to as black spots and hot spots?
- How can we spatially interpolate an unknown PM2.5 density at an arbitrary point on a road using known PM2.5 densities at observation points at CBD?

Why is it important to consider the underlying geography?

- With reference to (a), we will conclude that the spatial event is nonrandomly distributed points on a bounded plane,
- By constraining the points on a network, (b) reveals a randomly distributed patterns.
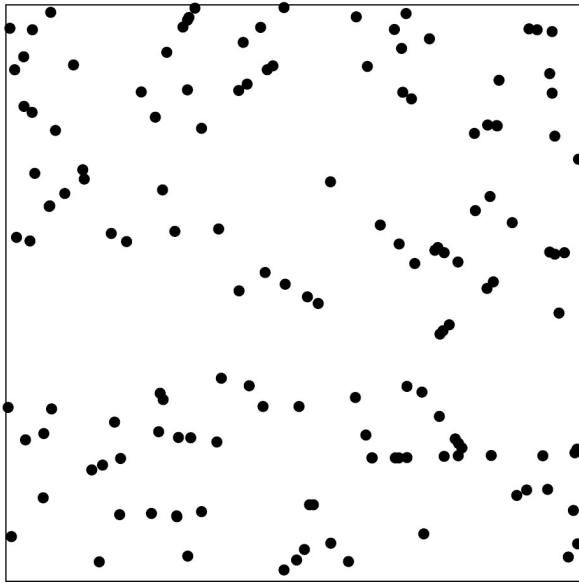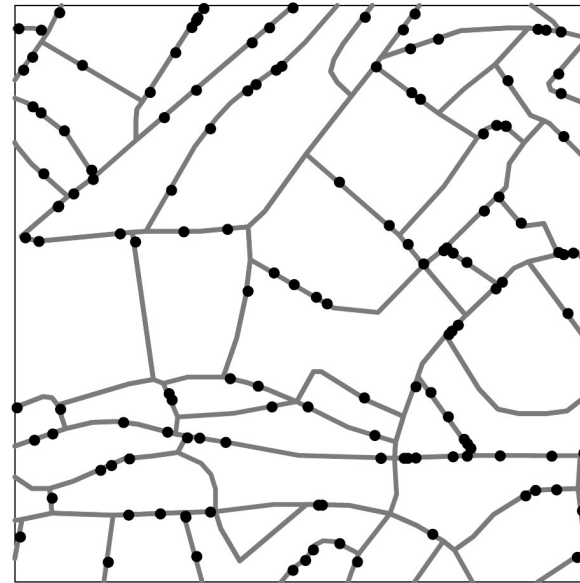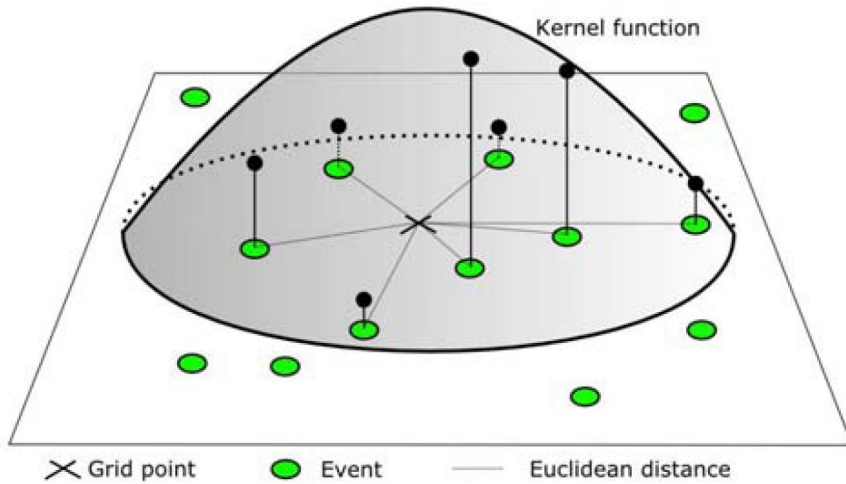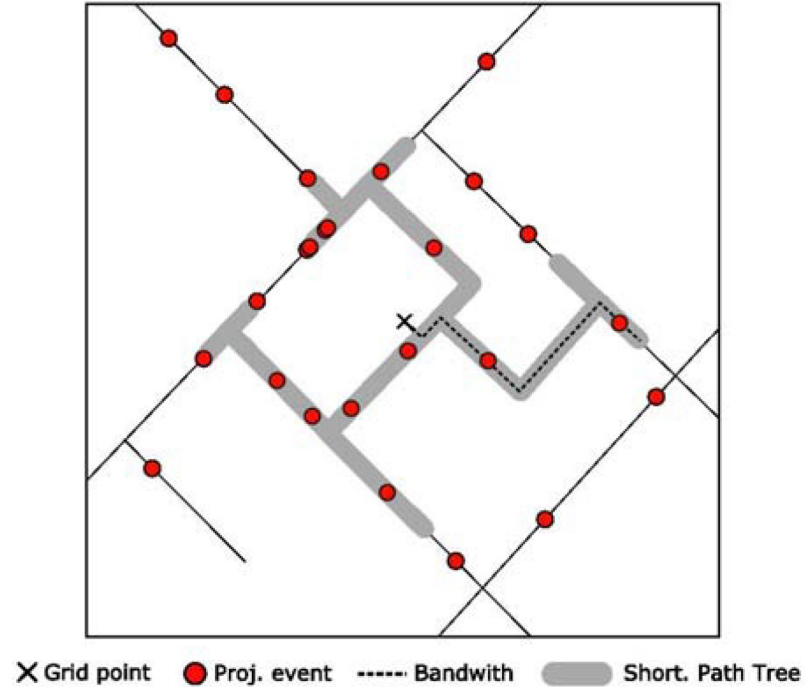

(a)   (b)

Planar KDE versus Network Constraied KDE (NetKDE)

- Kernel function of the Kernel Density estimation

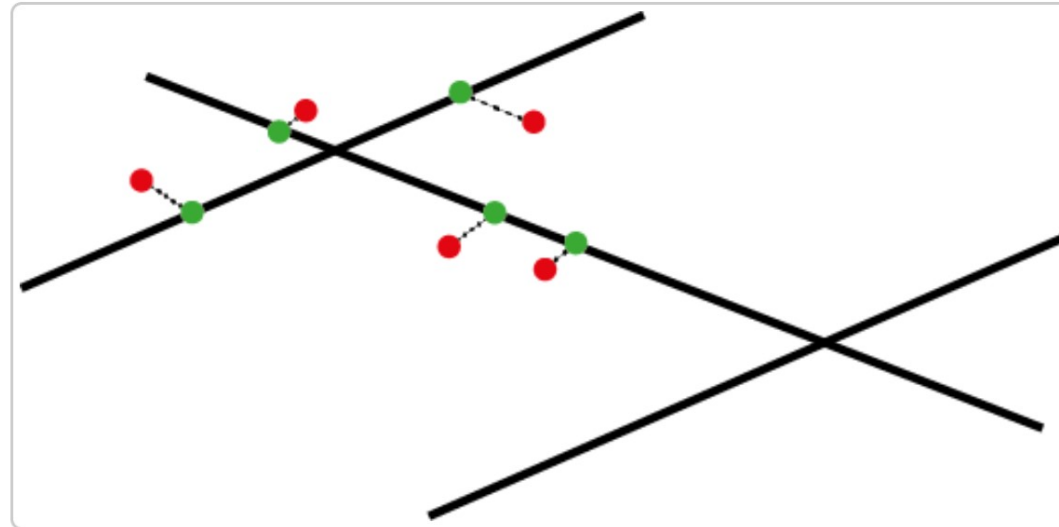- Shortest Path Tree: The grid point and the activities are projected on the nearest edge.

NetKDE algorithm: Basic step

- To perform a NKDE, the events must be snapped on the network. The snapped events are shown here in green.

Reference: Network Kernel Density Estimate

NetKDE: The kernel function

The mass of each event can be seen as a third dimension and is evaluated by a selected kernel function (K) within a specified bandwidth. The kernel function must satisfy the following conditions:

$K(x) >= 0$, if $x < bandwidth$
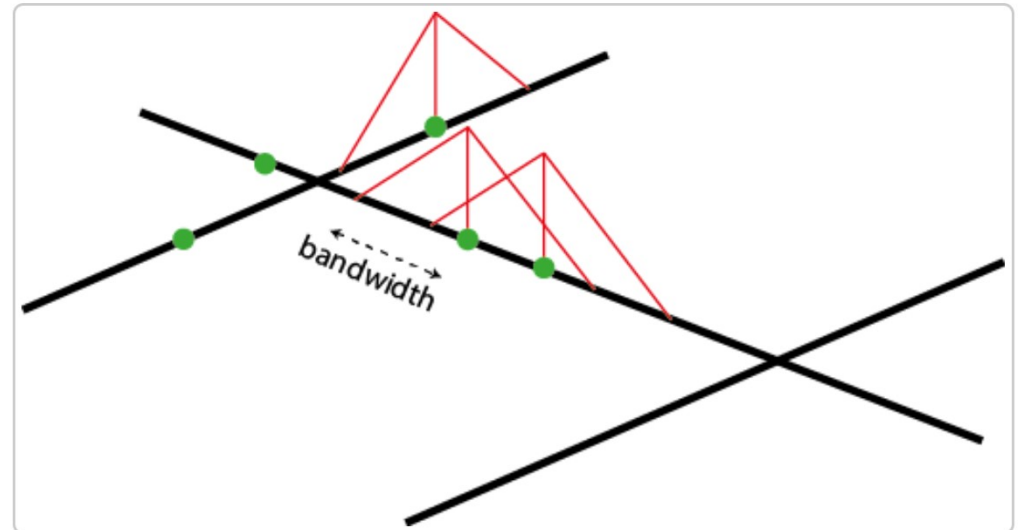
$K(x) = 0$, if $x > bandwidth$

$\int K(x) = 1$

The total mass of an event is 1, and is spread according to the function K within the bandwidth distance.

Reference: Network Kernel Density Estimate

We can see that the "influence" of each point is limited within the bandwidth and decreases when we move away from the event.

More than one point

In the figure below, 3 sampling points (s1, s2 and s3) are added in blue.
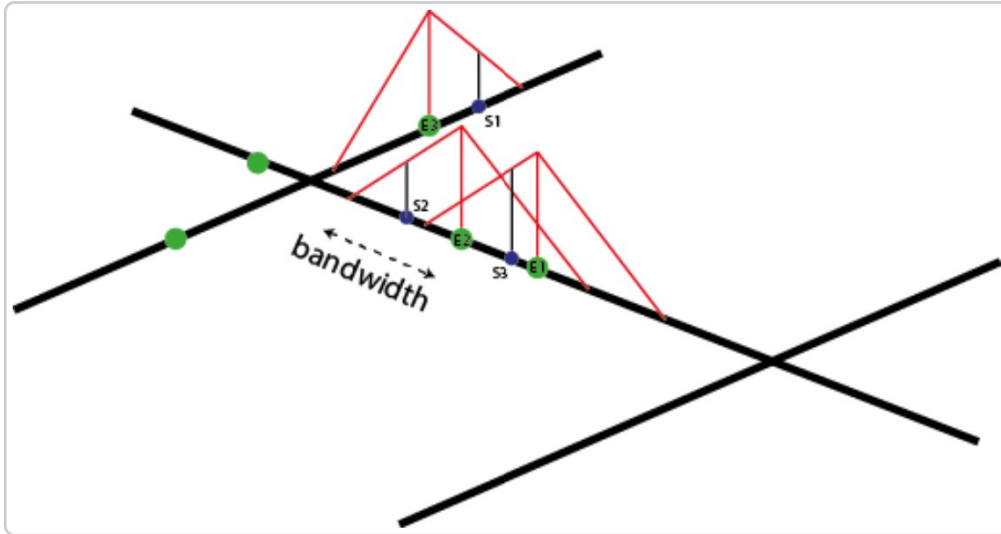
The NetKDE formula can be defined as follow:

$$d_{s1} = \frac{1}{bw^2} K(dist_{s1;e1})$$

$$d_{s2} = \frac{1}{bw^2} K(dist_{s2;e2})$$

$$d_{s3} = \frac{1}{bw^2} (K(dist_{s3;e2}) + K(dist_{s3;e3}))$$

Reference: Network Kernel Density Estimate

The general formula of NetKDE

The general formula of NetKDE is defined as:

$$d_{si} = \frac{1}{bw^2} \sum_{j=1}^{n} K(dist_{si;ej})$$

with dsi the density estimated at the sample point si, bw the bandwidth and ej an event.

The proposed kernel functions in the spNetwork package are:

- gaussian kernel: $k(\frac{d_{il}}{r}) = \frac{1}{\sqrt{2\pi}} * \exp(-\frac{d_{il}^2}{2r^2})$

- quartic kernel: $k(\frac{d_{il}}{r}) = \frac{3}{\pi} * (1 - \frac{d_{il}^2}{r^2})$

- epanechnikov kernel: $k(\frac{d_{il}}{r}) = \frac{3}{4} * (1 - \frac{d_{il}^2}{r^2})$

- triangle kernel: $k(\frac{d_{il}}{r}) = 1 - |\frac{d_{il}}{r}|$

- uniform kernel: $k(\frac{d_{il}}{r}) = |\frac{1}{r}|$

- triweight kernel: $k(\frac{d_{il}}{r}) = \frac{35}{32} * (1 - \frac{d_{il}}{r}^2)^3$

- tricube kernel: $k(\frac{d_{il}}{r}) = \frac{70}{81} * (1 - |\frac{d_{il}}{r}|^3)^3$

- cosine kernel: $k(\frac{d_{il}}{r}) = \frac{\pi}{4} * (cos(\frac{\pi}{2} * \frac{d_{il}}{r}))$

with r the bandwidth, and d_{il} the distance between an event and the sampling point.

Reference: Network Kernel Density Estimate

The Simple Method

The simple method was proposed by Xie and Yan (2008). They defined the NetKDE with the following formula:

$$d_{si} = \frac{1}{bw} \sum_{j=1}^{n} K(dist_{si;ej})$$

with d_{si} the density estimated at sample the point *si*, $dist_{si;ej}$ the network distance between *si* and event *j*, with all distances $dist_{si;ej} < bw$.

Pros	Cons
quick to calculate	biased (not a true kernel)
intuitive	overestimate the densities
continuous	

Reference: Network Kernel Density Estimate

Discontinuous NetKDE

The algorithm is proposed by Sugihara, Satoh, and Okabe (2010). It is easily presented visually below:
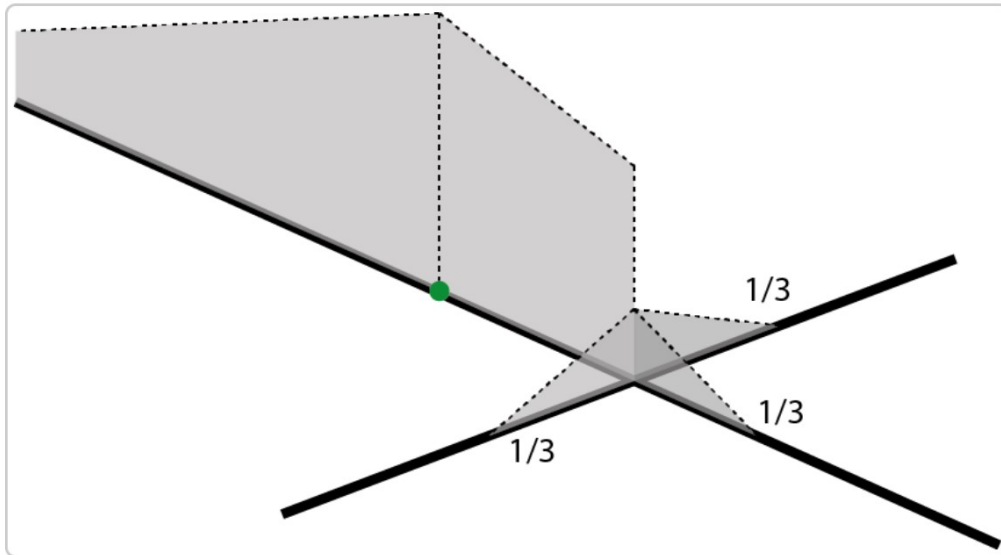
Pros	Cons
quick to calculate	counter-intuitive
unbiased	discontinuous
	contrasted results

Reference: Network Kernel Density Estimate

Continuous NetKDE

The continuous NKDE merges the best of the two worlds:

- it adjusts the values of the NetKDE at intersections to ensure that it integrates to 1 on its domain, and
- applies a backward correction to force the density values to be continuous.

This process is accomplished by a recursive function. It is more time consuming, so it might be necessary to stop it when the recursion is too deep.
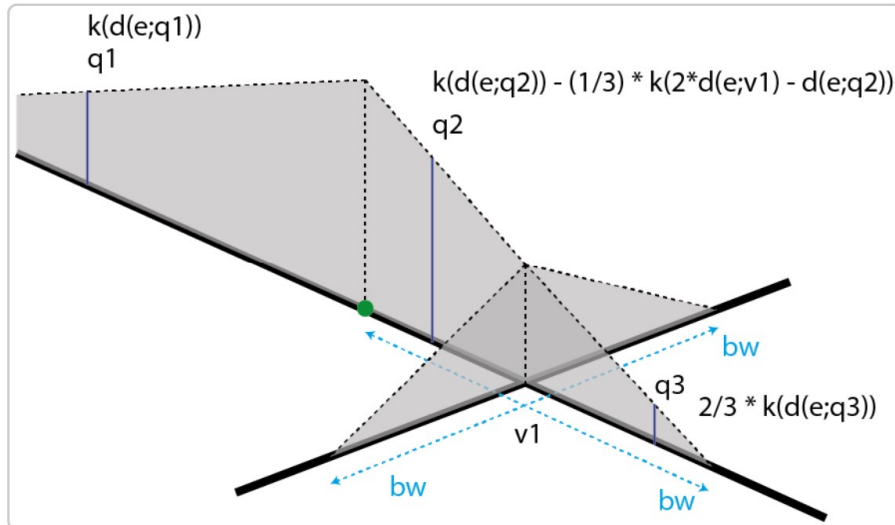
| Pros	Cons
unbiased	long calculation time
continuous	smoother values

Reference: Network Kernel Density Estimate

Adaptive bandwidth NetKDE

1. A fixed reference bandwidth (h_0) is selected. For each event (e_i) on the network, the density of the spatial process is evaluated by using this reference bandwidth ($\tilde{f} h_0(e_i)$).
2. For each event, the adaptive bandwidth is calculated with the following formula. A trimming values can be defined to limit the maximum adaptive bandwidth.
3. The densities at sampling points are then evaluated by using the obtained vector of bandwidth for each event.

$$h(e_i) = h_0 * \frac{1}{\sqrt{\tilde{f} h_0(e_i)}} * \frac{1}{\gamma_f}$$

$$\gamma_f = \exp\left(\frac{\sum_i log\left(\frac{1}{\sqrt{\tilde{f} h_0(e_i)}}\right)}{n}\right)$$

This method is called event oriented, because the new bandwidths are evaluated for the events rather than the sampling points. The sample oriented method is not implemented because not compatible with the algorithms used to calculate the discontinuous and continuous NKDE.

Reference: Network Kernel Density Estimate

Planar K-function versus Network K-function
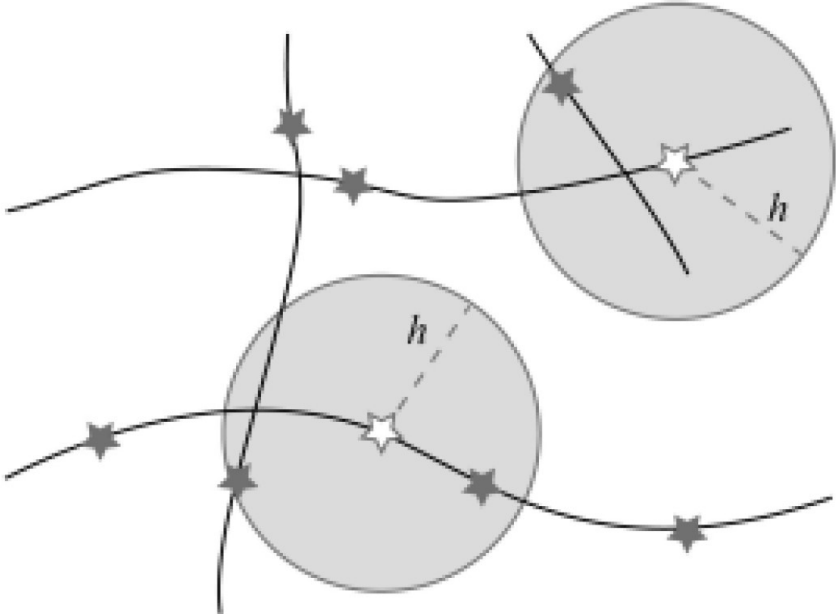
(a) Planar *K*-function
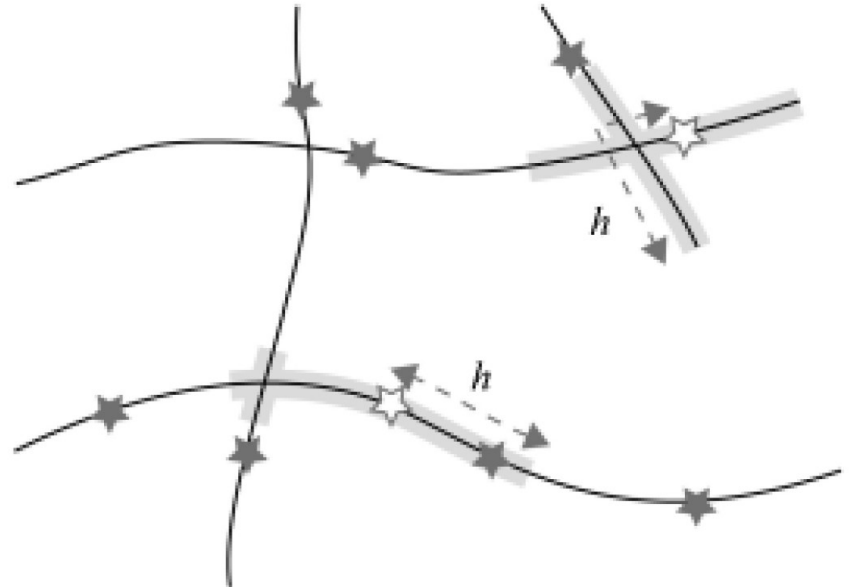
(b) Network *K*-function

Typical Network K-function question

Ho: The observed spatial point events (i.e airbnb listings, coffee outlets, traffic accident locations etc) are uniformly distributed over a street network in a study area.

- The assumption of the binomial point process implies the hypothesis that objects represented by P (say, airbnb listings) are uniformly and independently distributed over the street network Lp.

- If this hypothesis is rejected, we may infer that the spatial point events are spatially interacting and dependent on each other; as a result, they may form nonuniform patterns.

Spatial Point Patterns on network

In general, there are two types of nonuniform spatial point patterns on network, they are:

- clustering whereby the spatial point events tend to be close together (such as in Figure a), and
- repelling (also know as regular) whereby the spatial point event tend to keep away from each other (such as in Figure b).

(a) (b)
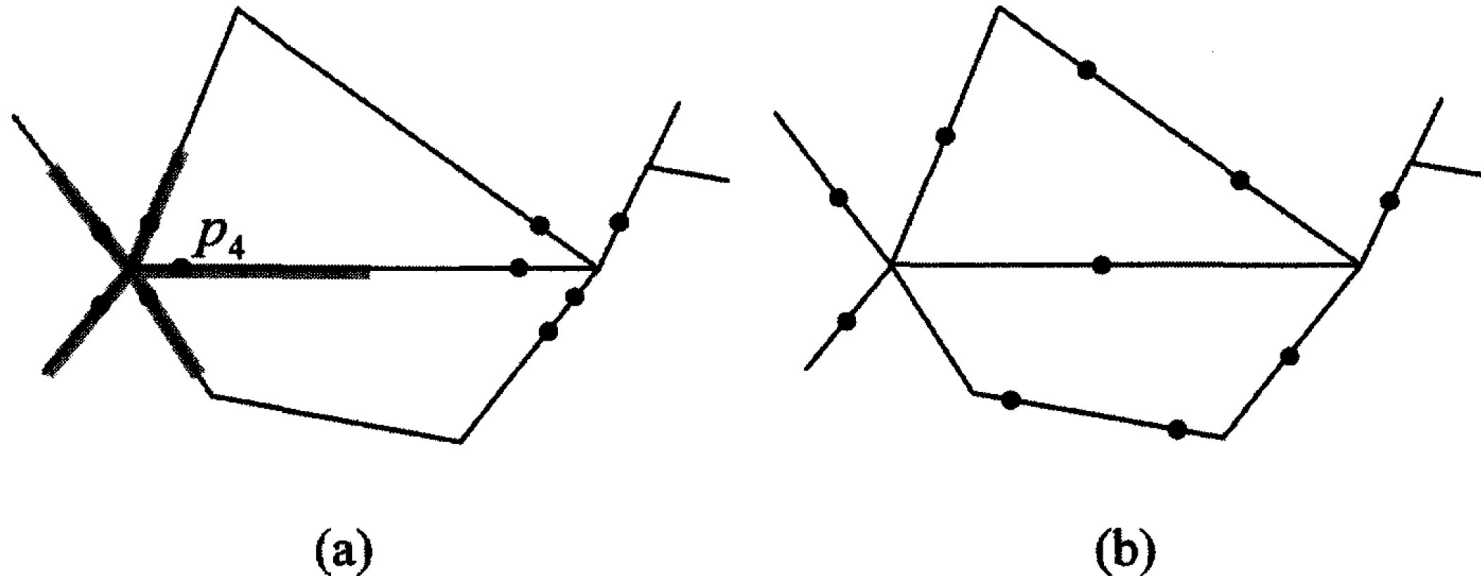
The Network K-function

Under the assumption of the binomial point process, network k-function K(t) (Atsuyuki Okabe and Ikuho Yarnada, 2001) is defined as:

$$K(t) = \frac{1}{\rho} \, \mathrm{E} \left(\begin{array}{c} \text{the number of points } P \text{ within} \\ \text{network distance } t \text{ of a point } p_i \text{ of } P \end{array} \right)$$

where $\mathrm{E}(\cdot)$ is the expected value with respect to $p_1, \ldots, p_n$ $(p_i \in P)$ which follow the binomial point process, and ρ is the density of points of P, that is, $\rho = n/|L_T|$. We call

Reference: Atsuyuki Okabe and Ikuho Yarnada (2001) "The K-Function Method on a Network and Its Computational Implementation", **Geographical Analysis**, Vol. 33, No. 3, pp. 271-290

Typical Network Cross K-function question

For instance, the set A may be the set of Airbnb listings and the set B may be the set of MRT stations. We are concerned with whether the locations of MRT stations affect the distribution of Airbnb listing.

Ho: Airbnb listings are distributed according to the binomial point process.

- This assumption implies that Airbnb listings are uniformly and independently distributed over LT regardless of the locations of MRT stations.

- If the above hypothesis is rejected, we may infer that the locations of MRT stations affect the distribution of Airbnb listing.

It should be noted that no assumption is made with respect to the distribution of points B.

The Network Cross K-function

To examine the above hypothesis, we define a function, $K^{ba}(t)$, for two sets of points A and B by

$$K^{ba}(t) = \frac{1}{\rho_a} \mathrm{E}\left(\begin{array}{l} \text{the number of points } A \text{ within} \\ \text{network distance } t \text{ of a point } b_i \text{ in } B \end{array} \right), \qquad (11)$$

where $\mathrm{E}(\cdot)$ is the expected value with respect to $b_1, ..., b_{n_b}$ $(b_i \in B)$ which follow the binomial point process, and ρ_a is the density of points of A, that is, $\rho_a = n_a/|L_T|$. We call the function $K^{ba}(t)$ the *network cross K-function of A relative to B* for the binomial point process. Note that $K^{ba}(t) \neq K^{ab}(t)$ in general.

Reference: Atsuyuki Okabe and Ikuho Yarnada (2001) "The K-Function Method on a Network and Its Computational Implementation", **Geographical Analysis**, Vol. 33, No. 3, pp. 271-290

References

Okabe, Atsuyuki and Yarnada, Ikuho (2001) "The K-Function Method on a Network and Its Computational Implementation", *Geographical Analysis*, Vol. 33, No. 3, pp. 271-290

Abramson, Ian S. 1982. "On Bandwidth Variation in Kernel Estimates-a Square Root Law." *The Annals of Statistics*, 1217–23.

Okabe, Atsuyuki, and Kokichi Sugihara. (2012) **Spatial Analysis Along Networks: Statistical and Computational Methods**. John Wiley & Sons.

Sugihara, Kokichi, Toshiaki Satoh, and Atsuyuki Okabe (2010) "Simple and Unbiased Kernel Function for Network Analysis." *In 2010 10th International Symposium on Communications and Information Technologies*, 827–32. IEEE.

Xie, Zhixiao, and Jun Yan (2008) "Kernel Density Estimation of Traffic Accidents in a Network Space." *Computers, Environment and Urban Systems* 32 (5): 396–406.