Open Data for Open Cities: Re-use and discovery level applied to the spatial point analysis process on linear networks
point pattern session

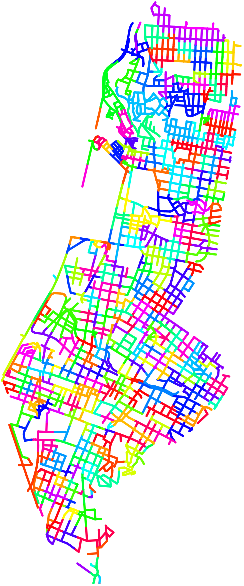
Mohammad Mehdi Moradi[a]

Institute of New Imaging Technologies (INIT), University Jaume I, Castellon, Spain[a]

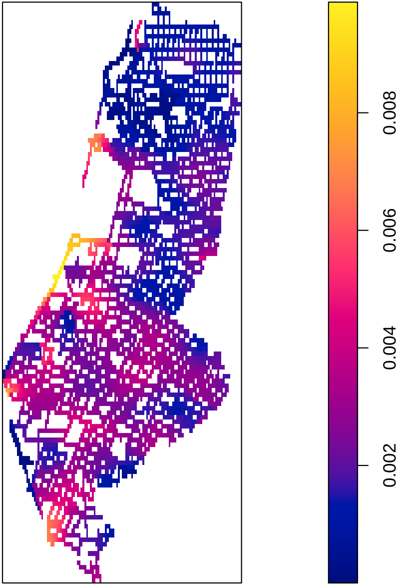
Outline

Traffic accidents in an area of Medellin, Colombia

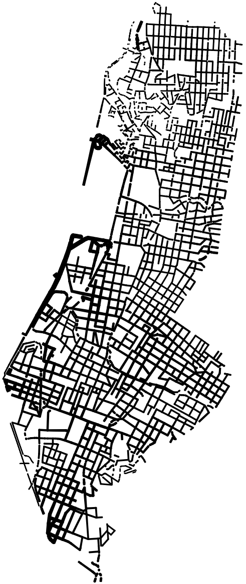
Traffic accidents in an area of Medellin, Colombia



Traffic accidents in an area of Medellin, Colombia



Traffic accidents in an area of Medellin, Colombia



Mathematical structure of linear networks

A line segment with endpoints $u \in \mathbb{R}^2$ and $v \in \mathbb{R}^2 \setminus \{u\}$ is given by $l = [u, v] = \{tu + (1-t)v : 0 \leq t \leq 1\} \subseteq \mathbb{R}^2$, $u \neq v$ and a *linear network* L is defined as the union

$$L = \bigcup_{i=1}^{k} l_i$$

of $k \geq 1$ line segments $l_i = [u_i, v_i]$, the *components* of L, which we assume satisfy $l_i \nsubseteq l_j$, $i \neq j$, and boundedness, i.e. $|l_i| = |u_i - v_i| < \infty$.

Mathematical structure of linear networks

A line segment with endpoints $u \in \mathbb{R}^2$ and $v \in \mathbb{R}^2 \setminus \{u\}$ is given by $l = [u, v] = \{tu + (1-t)v : 0 \leq t \leq 1\} \subseteq \mathbb{R}^2$, $u \neq v$ and a *linear network* L is defined as the union

$$L = \bigcup_{i=1}^{k} l_i$$

of $k \geq 1$ line segments $l_i = [u_i, v_i]$, the *components* of L, which we assume satisfy $l_i \not\subseteq l_j$, $i \neq j$, and boundedness, i.e. $|l_i| = |u_i - v_i| < \infty$.

The *total network length* is given by the sum of the component lengths, i.e.

$$|L| = \sum_{i=1}^{k} |l_i|,$$

whereby $|L| < \infty$ if $k < \infty$. On the other hand, if $k = \infty$, since each component has positive length, we have that $|L| = \infty$.

Mathematical structure of linear networks

A *path* from a starting point $u \in L$ to some end point $v \in L$ is a sequence of points $u_0, \ldots, u_n \in L$, $n > 1$, such that $u_0 = u$, $u_n = v$ and $[u_i, u_{i+1}] \subseteq L$ for all $i = 0, \ldots, n-1$.
Denoting by $p_{u,v}$ the collection of all paths between $u \in L$ and $v \in L$, the shortest-path (geodesic) distance between u and v is given by

$$\begin{aligned} d_L(u,v) &= \min_{(u_0,\ldots,u_n)\in p_{u,v}} \sum_{i=0}^{n-1} |u_i - u_{i+1}| \\ &= \min_{(u_0,\ldots,u_n)\in p_{u,v}} |u - u_1| + \sum_{i=1}^{n-2} |u_{i+1} - u_i| + |v - u_{n-1}|. \end{aligned}$$

Network example

Example

The first very simple example is any $L = \bigcup_{i=1}^{\infty} l_i$, $|L| = \infty$, *where each node has degree 2. Since the 1-dimensional Euclidean space* $\mathbb{R}$ *may be expressed as a countable union of compacts, we note that* L *is isometric to* $\mathbb{R}$ *and may be viewed as a bending of the real line at* $k = \infty$ *places. For the particular case where* $L = \mathbb{R}$ *it additionally follows that* $d_L(\cdot, \cdot)$ *coincides with the Euclidean metric.*

Example

The second example is the graph $L = \mathbb{Z}^2$, *which has its nodes at all integer pairs* $(i, j) \in \mathbb{R}^2$, $i, j \in \mathbb{Z}$. *The edges* $\mathcal{E}_L$ *are formed by the line segments joining the nodes horizontally and vertically. Note that we may scale everything by an arbitrary positive constant* $\alpha > 0$ *so that all edges have length* α.

Linear network point processes

Assume X is a point process on linear network L with intensity measure μ,

$$\mu(A) = E(N(A)) = \int_A \rho(u) \mathrm{d}_1 u, \qquad A \subseteq L,$$

where $N(A)$ is the number of points belong to subnetwork A, we then say X has *intensity function* $\rho(u)$, $u \in L$.

Linear network point processes

Example

A Poisson process X on L with intensity function $\rho(\cdot)$ is a LNPP such that $X(A_1), \ldots, X(A_n)$ are mutually independent for any disjoint line segments $A_1, \ldots, A_n \subseteq L$, $n \geq 1$, and $X(A)$ is Poisson distributed with mean $\mu(A) = \int_A \rho(u) \mathrm{d}_1 u$. Conditioning on $X(L) = n$, we obtain a classical random sample with density $f(u) = \rho(u)/\mu(L) = \rho(u)/n$, a so-called Binomial point process (Møller and Waagepetersen; 2004).

Voronoi tessellations/estimator

In the context of a linear network L, the *Voronoi cell/subnetwork* of $x \in X$ is given by (Okabe and Sugihara; 2012)

$$\mathcal{V}_x = \mathcal{V}_x(X) = \{u \in L : d_L(x,u) \leq d_L(y,u) \text{ for all } y \in X \setminus \{x\}\}.$$

Definition

Given a LNPP X on a linear (sub)network L with intensity function $\rho(u)$, $u \in L$, the *Voronoi intensity estimator* is given by

$$\hat{\rho}^V(u) = \hat{\rho}^V(u; X, L) = \sum_{x \in X \cap L} \frac{\mathbf{1}\{u \in \mathcal{V}_x(X) \cap L\}}{|\mathcal{V}_x(X) \cap L|}, \qquad u \in L,$$

if $X \cap L \neq \emptyset$; set $\hat{\rho}^V(u) = 0$ otherwise.

Kernel smoothing

The Kernel-based intensity estimator (Diggle corrected) of a point pattern on a linear network L is as

$$\hat{\lambda}_{\epsilon}^{(\mathrm{D})}(\mathbf{u}) = \sum_{i=1}^{n} \frac{\kappa_{\epsilon}(d_L(\mathbf{u}, \mathbf{v}_i))}{C_{L,\epsilon}(\mathbf{v}_i)}, \qquad \mathbf{u} \in L, \tag{1}$$

where κ_ϵ is a one-dimensional kernel function with bandwidth ϵ, and

$$C_{L,\epsilon}(\mathbf{v}) = \int_L \kappa_{\epsilon}(d_L(\mathbf{v}, \mathbf{u}))\mathrm{d}_1 u, \qquad \mathbf{v} \in X \cap L, \tag{2}$$

is an edge-correction factor (Moradi et al.; 2016).

Kernel smoothing

The diffusion estimator of intensity function of a point pattern on a linear network L with bandwidth $\epsilon = \sqrt{t}$ is as

$$\hat{\lambda}_\epsilon(\mathbf{u}) = \sum_{i=1}^{n} \kappa_{\epsilon^2}(\mathbf{u}|x_i), \qquad \mathbf{u} \in L \tag{3}$$

where κ_t is the heat kernel (McSwiggan et al.; 2016).

How to choose bandwidth parameter ϵ

Common possibilities:

How to choose bandwidth parameter ϵ

Common possibilities:

- use the smoothing bandwidth parameter ϵ which miximaize (Loader (1999))

$$\sum_{i=1}^{n} log(\hat{\lambda}_{-i,\epsilon}(u_i)) - \int_L \hat{\lambda}_{\epsilon}(u_i)\mathrm{d}_1 u \tag{4}$$

How to choose bandwidth parameter ϵ

Common possibilities:

- use the smoothing bandwidth parameter ϵ which miximaize (Loader (1999))

$$\sum_{i=1}^{n} log(\hat{\lambda}_{-i,\epsilon}(u_i)) - \int_L \hat{\lambda}_{\epsilon}(u_i)\mathrm{d}_1 u \tag{4}$$

- use the smoothing bandwidth parameter ϵ which minimizes (Cronie and van Lieshout (2016))

$$\left| \sum_{i=1}^{n} \frac{1}{\hat{\lambda}_{\epsilon}^{(\mathrm{D})}(u_i)} - |L| \right|.$$

Acknowledgements

Author gratefully acknowledges funding from the European union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, http:// www.geo-c.eu/). M.M. Moradi is also grateful to Adrian Baddelay and Greg Mcswiggan for providing us some codes in the latest version of R package spatstat.

Reference

Cronie, O. and van Lieshout, M. (2016). Bandwidth selection for kernel estimators of the spatial intensity function, *arXiv preprint arXiv:1611.10221* .

Loader, C. (1999). *Local Regression and Likelihood*, Springer, New York.

McSwiggan, G., Baddeley, A. and Nair, G. (2016). Kernel density estimation on a linear network, *Scandinavian Journal of Statistics* . doi: 10.1111/sjos.12255.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*, CRC Press.

Moradi, M. M., Rodriguez-Cortes, F. and Mateu, J. (2016). On the intensity estimator of spatial point patterns on linear networks, *submitted* .

Okabe, A. and Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods*, John Wiley & Sons.