Counting k-Hop Paths in the Random Connection Model

Alexander Kartun-Giles[1]

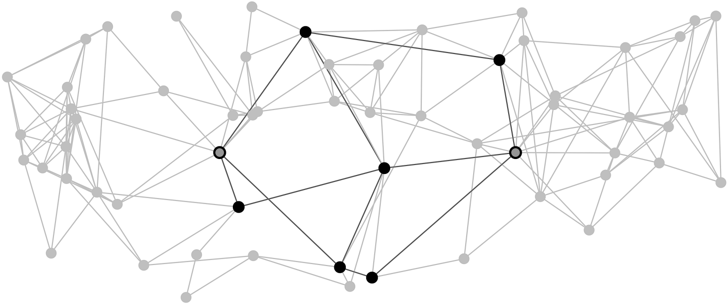
[1]Complex Systems Group
School of Mathematical Sciences
Queen Mary University of London

Spatially Embedded Networks, Oriel College, Oxford
13th of September 2017

Outline

Outline

Introduction: k-Hop Paths in the RCM



- With $\mathcal{X}_{\lambda(n)} \subset [0,1]^d$ a homogeneous Poisson point process of intensity $\lambda(n)\mathrm{d}x$, $\mathrm{d}x$ Lesbegue measure on $\mathbb{R}^d$, we consider the length **in hops** of paths running between two vertices $x, y \in \mathcal{X}_n$ in the random geometric graph

$$G\left(\mathcal{X}_{\lambda(n)}, \{(x,y) \in \mathcal{X}_{\lambda(n)} \times \mathcal{X}_{\lambda(n)} : 1\{u_{||x-y||} < H\left(||x-y||\right)\} > 0\}\right)$$

with independent $u_{||x-y||} \in U[0,1]$, $H : \mathbb{R}^+ \to [0,1]$ a connection function and $||x-y||$ distance according to some norm.

Motivation

Either counting the length k of paths, or the number $\sigma_k(||x-y||)$ of k-hop paths, is interesting for a number of reasons.

- Firstly, understanding the realtion between $d_{Euc}(||x-y||)$ and $d_{Graph}(||x-y||)$ stems from the fact that one can then find upper bounds for the diameter of G, denoted by $\text{diam}(G)$ (length in hops of the longest shortest path in G), that hold a.a.s., and the runtime complexity of many algorithms can often be bounded from above in terms of this diameter. Consider the problem of broadcasting information, see *'Diameter and Broadcast Time of Random Geometric Graphs in Arbitrary Dimensions' (Friedrich, Sauerwald, and Stauffer 2011)*.

Motivation

Either counting the length k of paths, or the number $\sigma_k(||x-y||)$ of k-hop paths, is interesting for a number of reasons.

- ....the main result is that, in the random geometric graph with H (the connection function) the **indicator of a ball of radius** $r(n)$ centered at the points of $\mathcal{X}_{\lambda(n)} \subset [0,1]^d$ of intensity $\lambda(n)\mathrm{d}x$, then there exists $K > 0$ such that

$$d_{Graph}\left(||x-y||\right) \leq \frac{K d_{Euc}\left(||x-y||\right)}{r\left(n\right)}$$

whenever $\lambda(n)\mathrm{d}x \to \infty$ with n, $r = \omega\left(\lambda^{-1/d}(n)\log^{1/d}\left(n\right)\right)$ i.e. the graph is in the connectivity regime $r(n) > r_c(n)$, and the Euclidean distance $||x-y|| = \omega\left(\log^{-1}\left(n\right)\right)$ i.e. for sufficiently large distances. Importantly $K = 1 + o(1)$ in this regime (Díaz et al. 2016).

Motivation

Either counting the length k of paths, or the number $\sigma_k(||x-y||)$ of k-hop paths, is interesting for a number of reasons.

- Secondly, the number of paths $\sigma_k(||x-y||)$ running between two vertices at displacement $||x-y||$ is related to the number of loops in the graph which intersect both x and y. Two three hop paths $\sigma_3(||x-y||) = 2$ makes a loop of length six, for example. These loops matter because the Erdős-Rényi random graph has no small loops a.a.s, and so *looks locally like a tree*. The random c-regular graph is similar. But the random geometric graph, as well as e.g. scale free networks, are on the other hand not locally tree-like.

Motivation

Either counting the length k of paths, or the number $\sigma_k(||x - y||)$ of k-hop paths, is interesting for a number of reasons.

- …This affects the validity of approximate models e.g. Sherrington-Kirkpatrick model of spin glass (which is a weighted subgraph of the complete graph without degree correlations), since they do not account for the existence of small loops. The strong interactions between spins are key features of these physical systems, and follow from the underlying geometric structure. Also, with frustration a result of odd-length loops, the loop counts are related to the amount of glassiness present in the system at low temperature.

In general, *the absence of a locally tree-like strucuture* is characteristic of an underlying spatial point process of arbitrary density.

Expectation of the number of paths

Theorem (Expected number of k-hop paths for the specific case of Rayleigh fading)

Take the so-called Rayleigh fading connection function

$$H(r) = \exp\left(-\beta r^2\right)$$

and define a new Poisson point process $\mathcal{X}_\lambda^\star$ *which is* $\mathcal{X}_\lambda$ *conditioned on containing two specific points* $x, y \in \mathbb{R}^d$ *at Euclidean distance* $||x - y||$. *Consider those two vertices* x, y *in the vertex set of the random geometric graph* $\mathcal{G}_H = (\mathcal{X}_\lambda^\star, E)$, *and set* $x = z_0, y = z_k$. *Then, in* $\mathcal{G}_H$, *the expected number of distinct non-repeating sequences of* k *sequential edges starting at* x *and terminating at* y *is*

$$\mathbb{E}\sigma_k = \frac{1}{k}\left(\frac{\lambda\pi}{\beta}\right)^{k-1} \exp\left(\frac{-\beta||x-y||^2}{k}\right).$$

Variance of the number of paths

Theorem (Variance of the number of three-hop paths for the specific case of Rayleigh fading)

Take the so-called Rayleigh fading connection function $H(r) = \exp\left(-\beta r^2\right)$ and define a new Poisson point process $\mathcal{X}_\lambda^\star$ which is $\mathcal{X}_\lambda$ conditioned on containing two specific points $x, y \in \mathbb{R}^d$ at Euclidean distance $||x-y||$. Consider those two vertices x, y in the vertex set of the random geometric graph $\mathcal{G}_H = (\mathcal{X}_\lambda^\star, E)$, and set $x = z_0, y = z_k$. Then, in $\mathcal{G}_H$, the expected number of distinct non-repeating sequences of k sequential edges starting at x and terminating at y is

$$\mathrm{Var}\left(\sigma_3\right) = \mathbb{E}\sigma_3 + \frac{\pi^3\lambda^3}{\beta^3}\left(\frac{1}{4}\exp\left(\frac{-\beta\|x-y\|^2}{2}\right)\right.$$
$$\left. + \frac{1}{6}\exp\left(\frac{-3\beta\|x-y\|^2}{4}\right)\right) + \frac{\pi^2\lambda^2}{8\beta^2}\exp\left(-\beta\|x-y\|^2\right).$$

Counting small subgraphs with indicator functions

Theorem (Slivnyak-Mecke Formula)

Let $t \in \mathbb{N}$. For any measurable real valued function f defined on the product of $(\mathbb{R}^d)^t \times \mathcal{G}$, where $\mathcal{G}$ is the space of all graphs on finite subsets of $[0,1]^d$, given a connection function H, the following relation holds

$$\mathbb{E} \sum_{X_1,\ldots,X_t \in \mathcal{Y}}^{\neq} f\left(X_1,\ldots,X_t,\mathcal{G}_H\left(\mathcal{Y} \setminus \{X_1,\ldots,X_t\}\right)\right)$$

$$= n^t \int_{[0,1]^d} \mathrm{d}x_1 \cdots \int_{[0,1]^d} \mathrm{d}x_t \mathbb{E} f\left(x_1,\ldots,x_t,\mathcal{G}_H\left(\mathcal{Y}\right)\right) \quad (1)$$

where $\mathcal{Y} \subset [0,1]^d$, $\mathbb{E}\|\mathcal{Y}\| = n$, and $\sum^{\neq}$ means the sum over all ordered t-tuples of distinct points in $\mathcal{Y}$.

This is little more than $\mathbb{E}\left[A + B + \ldots\right] = \mathbb{E}A + \mathbb{E}B + \ldots$.

Outline

Example: Counting two hop paths

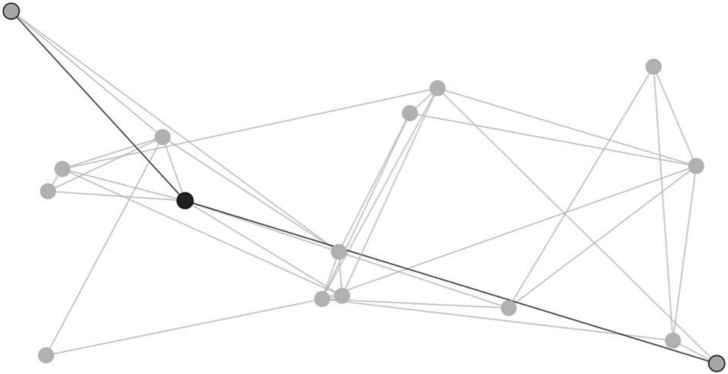
Define the *path-existence function* g to be the following product

$$g\left(z_1, \ldots, z_{k-1}, \mathcal{G}_H\left(\mathcal{Y}^\star\right)\right) = \prod_{i=0}^{k-1} \mathbf{1}\{z_i \leftrightarrow z_{i+1}\} \tag{2}$$

The expected value of this function is then just the product of the connection probabilities H of the inter-point distance along the sequence $z_0, \ldots, z_k$, i.e.

$$\mathbb{E}g(z_1, \ldots, z_{k-1}, \mathcal{G}_H\left(\mathcal{Y}^\star\right)) = \prod_{i=0}^{k-1} H\left(\|z_i - z_{i+1}\|\right) \tag{3}$$

Example: Counting two hop paths



Example: Counting two hop paths

For example, for two-hop paths we have

$$g\left(z_1, \mathcal{G}_H\left(\mathcal{Y}^\star\right)\right) = \prod_{i=0}^{1} \mathbf{1}\{z_i \leftrightarrow z_{i+1}\} = \mathbf{1}\{z_0 \leftrightarrow z_1\}\mathbf{1}\{z_1 \leftrightarrow z_2\}$$

(two edges), and the expected value of this function is then just the product of the connection probabilities H of the inter-point distance along the sequence $z_0 \to z_1 \to z_2$, i.e. $\mathbb{E}g\left(z_1, \mathcal{G}_H\left(\mathcal{Y}^\star\right)\right) = \prod_{i=0}^{1} H\left(\|z_i - z_{i+1}\|\right)$ and so

$$\mathbb{E}\sum_{X_1 \in \mathcal{Y}^\star}^{\neq} \{g\left(X_1, \mathcal{G}_H\left(\mathcal{Y}^\star\right)\right) = \lambda \int_{\mathbb{R}^2} \mathbb{E}g\left(z_1, \mathcal{G}_H\left(\mathcal{Y}^\star\right)\right) \mathrm{d}z_1 = \lambda \int_{\mathbb{R}^2} H\left(\|z_0 - z_1\|\right) H\left(\|z_1 - z_2\|\right) \mathrm{d}z_1$$

Example: Counting two hop paths

And with the Rayleigh fading connection function

$$H(r) = \exp\left(-\beta r^2\right)$$

this integral is

$$\lambda \int_{\mathbb{R}^2} H\left(\|z_0 - z_1\|\right) H\left(\|z_1 - z_2\|\right) \mathrm{d}z_1 = \frac{\lambda\pi}{2\beta} \exp\left(\frac{-\beta||x-y||^2}{2}\right)$$

and for three hop paths

$$\lambda^2 \int_{\mathbb{R}^2} H\left(\|z_0 - z_1\|\right) H\left(\|z_1 - z_2\|\right) H\left(\|z_2 - z_3\|\right) \mathrm{d}z_1 \mathrm{d}z_2 = \frac{\lambda^2\pi^2}{3\beta^2} \exp\left(\frac{-\beta||x-y||^2}{3}\right)$$

and so on for each k. These are expected numbers of k-hop paths.

Outline

t-tuples of paths

The $k = 2$ hop paths are Poisson:

$$\sigma_2(||x-y||) = \text{Poisson}\left(\frac{\lambda\pi}{2\beta}\exp\left(\frac{-\beta||x-y||^2}{2}\right)\right)$$

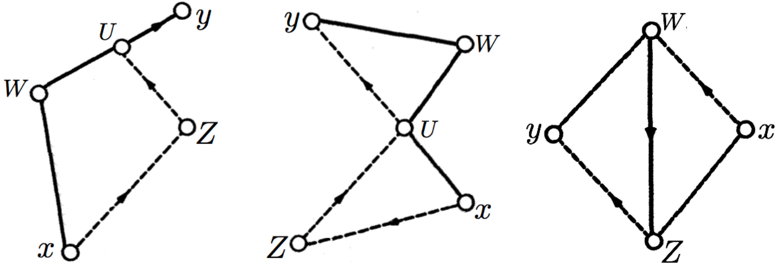
but the $k = 3$ hop paths are not Poisson. So calculate the variance using:

$$\sigma_3^2(||x-y||) = \Sigma_0 + \Sigma_1 + \Sigma_2$$

where for $i = 0, 1, 2$ the integer Σ_i denotes the number of ordered pairs of three hop paths with i vertices in common. This breaks down σ^t into t-tuples of paths, classed into catagories.

Substructures

For example, Σ_0 counts 2-tuples which do not intersect (not shown), while $\Sigma_{1(1)}$ and $\Sigma_{1(2)}$ count paths which intersect at a single vertex:



Σ_2 (right picture) counts 2-tuples which intersect at all vertices.

Outline

Paths which don't intersect

We can quickly evaluate the term Σ_0, which is the following sum over ordered 4-tuples of points in $\mathcal{Y}^\star$,

$$\Sigma_0 = \sum_{V,W,X,Y \in \mathcal{Y}^\star}^{\neq} g(V,W)\, g(X,Y)$$

The Mecke formula implies that

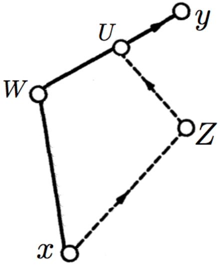
$$\begin{aligned} \mathbb{E}\Sigma_0 &= \rho^4 \int_{\mathbb{R}^8} \mathbb{E}\left(g(z_1,z_2)\, g(z_3,z_4)\right) \mathrm{d}z_1 \mathrm{d}z_2 \mathrm{d}z_3 \mathrm{d}z_4 \\ &= (\mathbb{E}\sigma_3)^2 \end{aligned}$$

which cancels with a term in the definition of the variance

$$\mathrm{Var}(\sigma_3) = \mathbb{E}(\sigma_3^2) - (\mathbb{E}(\sigma_3))^2$$

so now $\mathrm{Var}(\sigma_3) = \mathbb{E}\Sigma_1 + \mathbb{E}\Sigma_2$.

Paths which intersect



We break Σ_1 down into two separate contributions. The first is depicted on the left. This refers to the term:

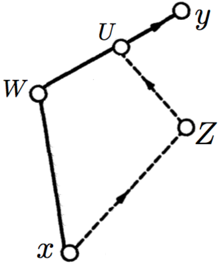
$$\sum_{U \in \mathcal{Y}^\star} \sum_{W,Z \in \mathcal{Y}^\star \setminus \{U\}}^{\neq} g\left(U, W\right) g\left(U, Z\right)$$

so count all 2-tuples of paths from $U \to x$ to get

$$\mathbb{E}\Sigma_{1(1)} = 2\rho \int_{\mathbb{R}^2} H\left(\|y - U\|\right) \mathbb{E}\left[\left(\sum_{X \in \mathcal{Y}^\star \setminus \{U\}} g\left(X\right)\right)_2\right] \mathrm{d}U$$

where the factor of two is due to symmetry.

Paths which do intersect

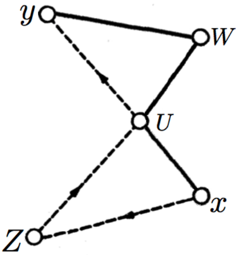


We break Σ_1 down into two separate contributions. The first is depicted on the left. This refers to the term:

$$\sum_{U \in \mathcal{Y}^\star} \sum_{W,Z \in \mathcal{Y}^\star \setminus \{U\}}^{\neq} g(U,W)\, g(U,Z)$$

$$\begin{aligned}
\mathbb{E}\Sigma_{1(1)} &= 2\rho \int_{\mathbb{R}^2} H(\|y-U\|)\, \mathbb{E}\left[\left(\sum_{X \in \mathcal{Y}^\star \setminus \{U\}} g(X)\right)_2\right] \mathrm{d}U \\
&= 2\rho^3 \int_{\mathbb{R}^2} H(\|y-U\|) \times \left(\int_{\mathbb{R}^2} H(\|x-z\|)\, H(\|z-U\|)\, \mathrm{d}z\right)^2 \mathrm{d}U
\end{aligned}$$

Paths which intersect



And now, notice how the next term is a slightly different sum where 2-tuples intersect at a single vertex, but in a different way:

$$\Sigma_{1(2)} = \sum_{U \in \mathcal{Y}^{\star}} \sum_{Z \in \mathcal{Y}^{\star} \setminus \{W\}} g(Z, U) \sum_{W \in \mathcal{Y}^{\star} \setminus \{Z\}} g(U, W)$$

Paths which do intersect

The two inner sums are in fact just counting the number of two hop paths between x and U, and U and y, then pairing them with each other

$$\mathbb{E}\Sigma_{1(2)} = \rho \int_{\mathbb{R}^2} H\left(\|x - U\|\right) H\left(\|U - y\|\right) \times \mathbb{E}\left[\sum_{Z \in \mathcal{Y}^\star \setminus \{W\}} g\left(Z, U\right) \sum_{W \in \mathcal{Y}^\star \setminus \{Z\}} g\left(U, W\right)\right] \mathrm{d}U$$

which simplifies to

$$\mathbb{E}\Sigma_{1(2)} = \rho^3 \int_{\mathbb{R}^2} H\left(\|x - U\|\right) H\left(\|U - y\|\right) \times \left(\int_{\mathbb{R}^2} H\left(\|x - z\|\right) H\left(\|z - U\|\right) \mathrm{d}z\right) \times \left(\int_{\mathbb{R}^2} H\left(\|U - z\|\right) H\left(\|z - y\|\right) \mathrm{d}z\right) \mathrm{d}U$$

Paths which do intersect

Since the terms are just integrals of products of the connection function over the whole plane, they can be evaluated exactly, with the two we've just calculated

$$\mathbb{E}\Sigma_{1(1)} = \frac{\pi^3\lambda^3}{4\beta^3}\exp\left(\frac{-\beta\|x-y\|^2}{2}\right)$$

$$\mathbb{E}\Sigma_{1(2)} = \frac{\pi^3\lambda^3}{6\beta^3}\exp\left(\frac{-3\beta\|x-y\|^2}{4}\right)$$

and the final term

$$\mathbb{E}\Sigma_2 = \frac{\pi^2\lambda^2}{8\beta^3}\exp\left(-\beta\|x-y\|^2\right)$$

calculated in a similar manner. This provides the variance, since

$$\mathrm{Var}(\sigma_3) = \left(\mathbb{E}\Sigma_1(1) + \mathbb{E}\Sigma_1(2)\right) + \mathbb{E}\Sigma_2$$

Higher moments

What we really want is factorial moments, since:

$$P(\sigma_k = t) = \frac{1}{t!}\sum_{i\geq 0}\frac{(-1)^i}{i!}\mathbb{E}\sigma(\sigma-1)\ldots(\sigma-t-i+1)$$

for example, the path existence probability

$$P(\sigma_k > 0) = 1 - \sum_{i\geq 0}\frac{(-1)^i}{i!}\mathbb{E}\left[(\sigma_k)_i\right]$$

where $(\sigma_k)_i$ is the i^{th} factorial moment, with the zeroth moment equal to unity. Also, the partial sums upper and lower bound this path existence probability in turn.

With our results, we have these factorial moments for $i = 0, 1, 2$, but not higher. At least this provides a bound on the path existence probability which does not use a "mean field" model, where spatial dependence is ignored.

Conclusions

Conclusions:

1. We can use the Slyviniak-Mecke formula to count tuples k-hop paths, and deduce some factorial moments.
2. This leads to bounds on the path existence probability, which improve as you calculate more factorial moments.

Outlook: Can someone find a recursion/general theory of these moments?

What do they tell us about the relation between graph distance and Euclidean distance in the random connection model?

Also, **can we count the expected number of loops of length** L?

Thank you.