Entropy of Random Geometric Graphs

Justin P. Coon

with thanks to N. Warsi, O. Georgiou and C. P. Dettmann

21 September, 2015

Random Geometric Graphs and Spatially Embedded Networks

The model...

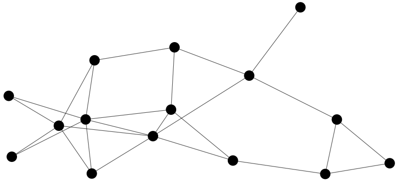
- Node positions are random
- Existence of an edge (i, j) depends on distance between nodes i and j
- Notion of distance implies embedding in some space (e.g., Euclidean)

Applications...

- Human/animal interactions
- Nano processes (e.g., carbon nanotubes on a polymer substrate)
- Wireless communication networks

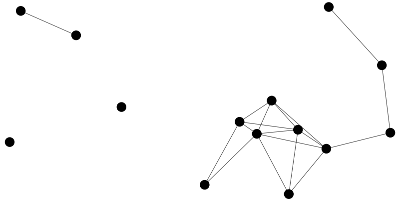
Erdős-Rényi Graph vs Random Geometric Graph

ER Graphs



- Edge probability $\wp$
- Independent of embedding

RG Graphs



- Edge probability $\wp(s_{i,j})$
- Distribution of $s_{i,j}$ important

Graph Entropy

Consider a graph $G = (\mathcal{V}, \mathcal{E})$ to be a set of vertices (nodes) and a set of edges (links).

- $\binom{n}{2} = n(n-1)/2$ possible edge configurations
- $2^{n(n-1)/2}$ possible graphs without considering node locations

The entropy of G is a measure of disorder or the amount of information contained in the graph distribution.

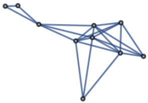
$$H(G) = \mathbb{E}[-\log(\mathbb{P}(G))]$$

** Logarithms are base e.*

ER Graph Entropy

For an ER graph, all edges occur independently with probability $\wp$. To calculate entropy...

- Map G to $n(n-1)/2$ Bernoulli variables $X_1, \ldots, X_{n(n-1)/2}$



$\mapsto$ **110 100 010 . . .**

- Independence of $\{X_i\}$ implies

$$H(G) = H(X_1, \ldots, X_{n(n-1)/2}) = \sum_i H(X_i) = \binom{n}{2} H(\wp)$$

- Entropy of single edge is

$$H(\wp) = -\wp \log \wp - (1-\wp)\log(1-\wp)$$

ER Graph Entropy

Examples

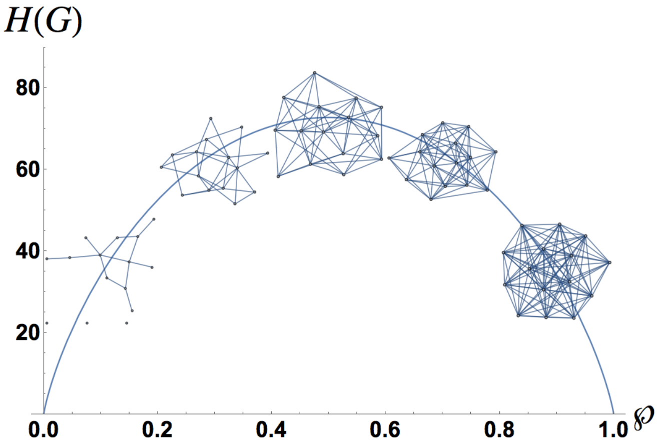


Figure : Entropy of ER graph with $n = 15$ nodes.

RG Graph Entropy

For an RG graph, edge (i, j) occurs with probability $\wp_{i,j}$, which depends on the distance $s_{i,j} = \|\mathbf{r}_i - \mathbf{r}_j\|$. Consider *conditional entropy*...

$$
\begin{aligned}
H(G|S) &= \mathbb{E}[H(G|s_{1,2}, \ldots, s_{n-1,n})] \\
&\leq \sum_{i<j} \mathbb{E}[H(X_{i,j}|s_{i,j})] && \text{(independence bound)} \\
&= \binom{n}{2} \mathbb{E}[H(X|s)] \\
&\leq \binom{n}{2} H(\overline{\wp}) && \text{(Jensen's inequality)}
\end{aligned}
$$

where

$$\overline{\wp} := \mathbb{E}[\wp(s_{i,j})].$$

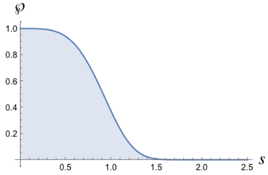
Pair Connection Function

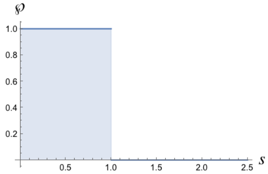
Interesting properties arise from the **distribution of the pair distance** and the **pair connection function**. Let...

$$\wp(s) = \exp(-(s/s_0)^\eta)$$

- Models Rayleigh fading in wireless networks
- Versatility offered by parameter η: probabilistic to deterministic

$$\Rightarrow \overline{\wp} = \int_0^D f(s) \exp(-(s/s_0)^\eta)\,\mathrm{d}s$$





Uniform Node Distributions in a Compact Domain

In d dimensions, the **set covariance** of a convex set $\mathcal{K}$ is given by

$$c_{\mathcal{K}}(\mathbf{s}) = \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{K}}(\mathbf{x}) \mathbf{1}_{\mathcal{K}}(\mathbf{x} - \mathbf{s}) \, \mathrm{d}\mathbf{x}, \qquad \mathbf{s} \in \mathbb{R}^d$$

where $\mathbf{1}_{\mathcal{K}}(\mathbf{x})$ is the indicator function for $\mathbf{x} \in \mathcal{K}$.

Uniform Node Distributions in a Compact Domain

In d dimensions, the **set covariance** of a convex set $\mathcal{K}$ is given by

$$c_{\mathcal{K}}(\mathbf{s}) = \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{K}}(\mathbf{x}) \mathbf{1}_{\mathcal{K}}(\mathbf{x} - \mathbf{s}) \, \mathrm{d}\mathbf{x}, \qquad \mathbf{s} \in \mathbb{R}^d$$

where $\mathbf{1}_{\mathcal{K}}(\mathbf{x})$ is the indicator function for $\mathbf{x} \in \mathcal{K}$.

The **isotropised set covariance** is given by

$$\overline{c}_{\mathcal{K}}(s) = \int_{\mathcal{S}^{d-1}} c_{\mathcal{K}}(s\mathbf{u}) \, \mathrm{d}\mathbf{u}, \qquad s \geq 0$$

where $\mathbf{u}$ is a vector denoting a point on the unit sphere $\mathcal{S}^{d-1}$.

Uniform Node Distributions in a Compact Domain

In d dimensions, the **set covariance** of a convex set $\mathcal{K}$ is given by

$$c_{\mathcal{K}}(\mathbf{s}) = \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{K}}(\mathbf{x})\mathbf{1}_{\mathcal{K}}(\mathbf{x}-\mathbf{s})\,\mathrm{d}\mathbf{x}, \qquad \mathbf{s} \in \mathbb{R}^d$$

where $\mathbf{1}_{\mathcal{K}}(\mathbf{x})$ is the indicator function for $\mathbf{x} \in \mathcal{K}$.

The **isotropised set covariance** is given by

$$\overline{c}_{\mathcal{K}}(s) = \int_{\mathcal{S}^{d-1}} c_{\mathcal{K}}(s\mathbf{u})\,\mathrm{d}\mathbf{u}, \qquad s \geq 0$$

where $\mathbf{u}$ is a vector denoting a point on the unit sphere $\mathcal{S}^{d-1}$.

The **pair distance** probability density function is given by

$$f(s) = \frac{2\pi^{d/2} s^{d-1} \overline{c}_{\mathcal{K}}(s)}{\Gamma(d/2)\,\mathrm{vol}(\mathcal{K})^2}.$$

Entropy vs Connection Range s_0

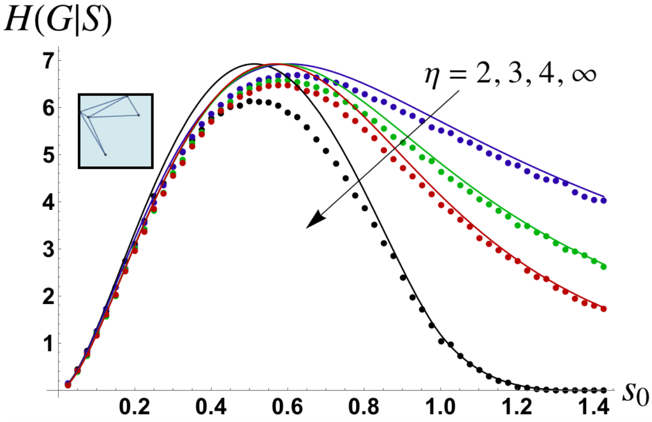
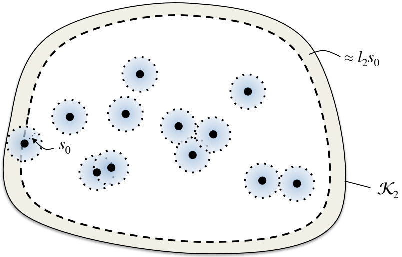


Figure : Entropy of RG graph with $n = 5$ nodes and pair connection functions with typical range s_0 for a unit square. Solid line: upper bound. Markers: numerical simulations.

Uniform Node Distributions in a 2D Compact Domain

Small Typical Connection Range

Consider a compact domain $\mathcal{K}_2$ with area a_2, boundary length l_2 and small typical connection range $s_0 \ll \pi a_2 / l_2 \ll D$.



- Pair distance distribution

$$f(s) = \frac{2\pi s}{a_2}\left(1 - \frac{l_2 s}{\pi a_2}\right)$$

- Soft pair connection probability

$$\overline{\wp} = \frac{2\pi s_0^2 \Gamma(2/\eta)}{\eta a_2}\left(1 - \frac{l_2 s_0}{\pi a_2}\frac{\Gamma(3/\eta)}{\Gamma(2/\eta)}\right)$$

- Hard pair connection probability

$$\overline{\wp} = \frac{\pi s_0^2}{a_2}\left(1 - \frac{2 l_2 s_0}{3\pi a_2}\right)$$

Entropy for Small s_0

Let $u_2 = 2\pi\Gamma(2/\eta)/(\eta a_2)$ denote the *fractional area of a unit soft disc*. For $s_0 \to 0$, we have

$$H(G|S) \leq \binom{n}{2} \left(2u_2 \log\left(\frac{1}{s_0}\right) s_0^2 + u_2(1 - \log u_2)s_0^2 + O(s_0^3 \log s_0) \right)$$

- For fixed s_0, the bound increases with n.
- For fixed n, the bound decreases with s_0 (toward a completely disconnected graph).

Entropy for Small s_0

What about $s_0 = g(n)$?

Fact

For typical connection distances that decay according to the relation

$$s_0^2 \log\left(\frac{1}{s_0}\right) = o\left(\frac{1}{n^2}\right)$$

the entropy $H(G|S) \to 0$ *as* $n \to \infty$.

An Entropy Limit for Small s_0

Fact

The upper bound on the entropy of a graph in $\mathcal{K}_2$ will tend to a limit $\nu > 0$ as $n \to \infty$ if

$$s_0(n) = \exp\left(\frac{1}{2} W_m\left(-\frac{2\nu}{u_2 n^2}\right)\right) = \frac{\sqrt{\nu}}{n\sqrt{u_2 \log n}}\left(1 + O\left(\frac{1}{\log n}\right)\right)$$

where $W_m(x)$ is the lower branch ($-1/e \leq x < 0$ and $W_m \leq -1$) of the solution to $x = W \exp W$.

Entropy vs Number of Nodes n

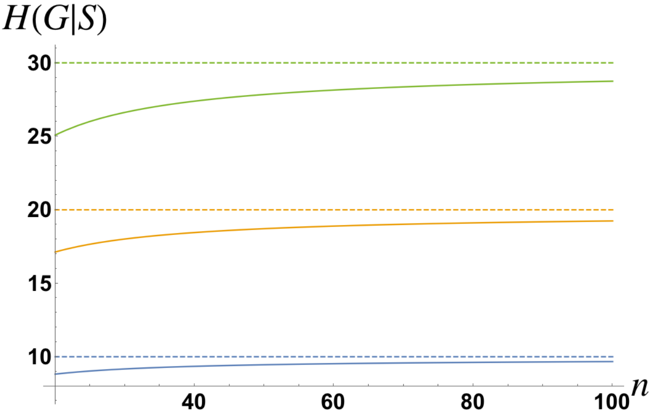


Figure : Upper bound on the entropy of an RG graph in a unit square with $s_0 = \exp\left(\frac{1}{2} W_m \left(-\frac{2\nu}{u_2 n^2}\right)\right)$. Solid line: theory. Dashed line: limit.

Connectivity for Small s_0

Fact

The probability that a graph in $\mathcal{K}_2$ is completely disconnected is well approximated by

$$P = (1 - \overline{\wp})^{n(n-1)/2}.$$

Thus, for $s_0 \to 0$ and $n \to \infty$, a graph is almost surely completely disconnected if

$$s_0 = o\left(\frac{1}{n}\right).$$

The probability of a completely disconnected graph will tend to a limit $\varphi \in (0, 1)$ if

$$s_0(n) = \frac{1}{n}\sqrt{\frac{2}{u_2}\log\left(\frac{1}{\varphi}\right)}$$

Connectivity vs Entropy for Small s_0

Interesting Fact

Letting s_0 tend to zero such that the upper bound on $H(G|S)$ tends to $\nu > 0$ yields

$$P \sim 1 - \frac{\nu}{2 \log n}.$$

I.e., typical connection distances that yield a positive limit on the entropy bound result in almost surely completely disconnected graphs (albeit with a slow convergence).

Arbitrary Node Configurations for $s_0 \gg D$

For the **hard disc model**, the graph G is complete, so $H(G|S) = H(G) = 0$.

For the **soft model** (i.e., $\eta < \infty$)

$$\overline{\wp} = 1 - \frac{\mathbb{E}[s^\eta]}{s_0^\eta} + O\left(\frac{1}{s_0^{2\eta}}\right)$$

provided the ηth, 2ηth, ... moments exist. The conditional entropy is thus bounded by

$$H(G|S) \leq \binom{n}{2} \frac{\eta \mathbb{E}[s^\eta] \log s_0}{s_0^\eta} + O\left(\frac{1}{s_0^\eta}\right), \qquad \eta < \infty$$

Room for Improvement...

- Consider $d > 2$; all small s_0 results generalise
- More to consider for $s_0 \gg D$
- Exact calculations (difficult)
- Strengthen the bound
- Consider other pair connection functions
- Explicit calculations for specific bounding geometries
- Maximum entropy points
- Structural entropy
- Applications (e.g., routing/forwarding tables, protocol design)