#### Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric Graphs

Matthew Garrod

EPSRC Centre for Doctoral Training in the Mathematics of Planet Earth

Imperial College London

Supervisor: Nick Jones

September 11, 2017

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





#### Motivation

- Random Geometric Graphs
- When is Knowing Node Locations Useful?
- The Algebraic Connectivity
- 2 Large Fluctuations in the Algebraic Connectivity
- Why? Correlations between the Algebraic Connectivity and Other Network Properties





#### Random Geometric Graphs

We construct a **random geometric graph** (RGG) as follows:

- Draw a vector of random positions  $\underline{X} = (X_1, X_2, ..., X_N)$  from some distribution P(X).
- Connect nodes *i* and *j* if their Euclidean distance is less than *R* ie.  $|X_i X_j| < R$ .



Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





#### Random Geometric Graphs



(a) Uniform distribution in  $[0,1]^2$  (b) 2D Gaussian distribution in  $\mathbb{R}^2$ 

Figure: Examples of RGGs with different node distributions.

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





#### Random Geometric Graphs

- Node Locations need not represent positions in physical space. They may represent any continuous (or perhaps discrete) **node attributes**.
- e.g in social networks they might represent the **'social coordinates'** of individuals. Different dimensions might represent characteristics such as age, income, social background etc..
- Spatial networks/Random Geometric Graphs provide a framework for modeling networks which show **assortative mixing** (ie. nodes with similar characteristics more likely to share connections).
- > Motivation to consider spaces with  $d \ge 2!$





#### Randomness in RGGs



Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric EPSRC Engineering and Physical Sciences Research Council



#### When is Knowledge of Node Locations Useful?



 $\mathbb{P}(X_1, X_2, ..., X_N) \xrightarrow{\text{Estin}}_{\substack{\text{Leng}\\ \text{Prop}}}$ 

Estimates of Mean Shortest Path Length, Mean Degree, Dynamical Properties etc..





#### When is Knowledge of Node Locations Useful?



ペロト 《 伊 ト 《 臣 ト 《 臣 ト ○ 臣 · の Q ()





#### The Coefficient of Variation

The Coefficient of Variation can be used to measure the size of fluctuations in a property, q, relative to its mean. It is defined as the standard deviation divided by the mean:

$$CV(q) = rac{\sigma(q)}{\langle q \rangle}$$
 (1)

e.g if CV = 0.1 then we have fluctuations of the order 10% the size of the mean.

• If we want to estimate some q precisely for a large graph then fluctuations greater than 5 - 10% the size of the mean might be considered as large fluctuations.





#### The Algebraic Connectivity

#### **Laplacian Matrix**



Note: we consider  $\mu_2$  for the largest connected component of the network.





### Methodology

We consider three distinct RGG ensembles:

- Periodic Boundaries: Uniform distribution in [0, 1]<sup>d</sup> with toroidal boundary conditions. We refer to these networks as PRGGs
- Solid Boundaries: Uniform distribution in [0, 1]<sup>d</sup> with solid boundaries. We refer to these as SRGGs
- **3** Gaussian Node Positions: Node locations drawn from multivariate gaussian density  $f(x_1, ..., x_d) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{(x_1^2 + ... + x_d^2)}{2}}$  on  $\mathbb{R}^d$ . We will refer to networks of this class as GRGGs.





#### Methodology





(a) Uniform distribution in  $[0,1]^2$ 

(b) 2D Gaussian distribution in  $\mathbb{R}^2$ 

Figure

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





### Methodology



Figure: Graphic illustrating the method used to sample from  $\mathbb{P}(\mu_2)$  for RGGs.

- We also extract the *largest connected component* of the network before computing μ<sub>2</sub> and other network properties.
- We will use  $\kappa$  to refer to the mean degree of ensembles  $\rightarrow (\Xi) \rightarrow (\Xi) \rightarrow (\Xi)$





The value of  $CV(\mu_2)$  can be relatively large (> 0.1) even for networks of size  $N = 10^5$ .



Figure: Plots showing estimates of  $CV(\mu_2)$  as a function of the dimension for the three different RGG ensembles. Shown for the case of mean degree,  $\kappa = 20.0$  for network ensembles of sizes  $N = 10^4$  and  $10^5$ . Also shown is the  $CV(\mu_2)$  for the ER random graph model.

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





#### Motivation

- Random Geometric Graphs
- When is Knowing Node Locations Useful?
- The Algebraic Connectivity

Large Fluctuations in the Algebraic Connectivity

Why? - Correlations between the Algebraic Connectivity and Other Network Properties

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric

#### - < ロ > < 母 > < 臣 > < 臣 > 三 回 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < 巴 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □





Large fluctuations in the  $\mu_2$  of GRGGs are associated with the presence of weakly connected subgraphs.



(a)  $\mu_2 = 0.01, F_P = 0.03$  (b)  $\mu_2 = 0.07, F_P = 0.13$  (c)  $\mu_2 = 0.16, F_P = 0.46$ 

Figure: Plots showing networks drawn from the ensemble of GRGGs with N = 1000, d = 2 and  $\kappa = 20.0$ . Nodes are colored according to partition corresponding to the eigenvector associated with  $\mu_2$ .  $F_P$  is the fraction of nodes in the smaller partition.

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





Large Fluctuations in the  $\mu_2$  of SRGGs are associated with the presence a small number of weakly connected nodes.



(a)  $\mu_2 = 0.23$ ,  $F_P = 0.02$  (b)  $\mu_2 = 0.35$ ,  $F_P = 0.04$  (c)  $\mu_2 = 1.7$ ,  $F_P = 0.47$ 

Figure: Plots showing networks drawn from the ensemble of SRGGs with N = 1000, d = 5 and  $\kappa = 20.0$ . Nodes are colored according to partition corresponding to the eigenvector associated with  $\mu_2$ .  $F_P$  is the fraction of nodes in the smaller partition.

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





# Different factors influence $\mu_2$ for RGGs in different dimensions (Solid BCs)



Figure: Scatter plot showing the joint distribution of  $\mu_2$  and *average shortest path length*, *L*, for RGGs with N = 1000. Shown for  $\kappa = 20.0$  with solid boundaries. Points are colored according to the minimum degree of corresponding network ensemble member.

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric ロマム時マスポットで、





# Different factors influence $\mu_2$ for RGGs in different dimensions (Periodic BCs)



Figure: Scatter plot showing the joint distribution of  $\mu_2$  and *average shortest path length*, *L*, for RGGs with N = 1000. Shown for  $\kappa = 20.0$  with periodic boundaries. Points are colored according to the minimum degree of corresponding network ensemble member.

Training

nac

- ロト - ( 同 ト - ( 三 ト - 4

The relationship between  $\mu_2$  and other network properties within a network ensemble is dependent on both the ensemble type and the spatial dimension



Figure: Plot showing Spearmans rank correlation coefficient,  $\rho$  between  $\mu_2$  and other network properties as a function of dimension. Shown for N = 1000,  $\kappa = 20.0$ . Error bars show 95% confidence intervals on  $\rho$  estimated via bootsrapping.

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





## Is Knowing Node Locations Useful? (Revisited)

- In RGGs with **uniform node distributions**  $\mu_2$  is strongly influenced by **microscopic properties** ie. the minimum degree
- **②** In RGGs with **Gaussian node distributions**  $\mu_2$  is influenced by the presence of **mesoscopic structures**.

In case 1) we might need exact knowledge of node locations in order to make predictions. In case 2) we may be able to make more gains without knowing all node locations precisely.

- Real situations may be modeled more realistically by **Soft Random Geometric Graphs** where the probability of two nodes being connected decays as a function of their separation distance.
- We expect our results to generalize to some choices of connection function.



Mothem

SQR

## Conclusions

- The presence of boundaries and heterogeneity in the distribution of nodes are important factors for driving fluctuations in the algebraic connectivity of RGGs.
- We have observed these large fluctuations in various large but finite sized RGG ensembles.
- The covariation of the algebraic connectivity and network properties is dependent both on the class of RGG ensemble and the dimensionality of the space.
- Knowledge of node locations in RGGs allows us to exactly specify the network structure and compute the algebraic connectivity.
- -> in the cases where  $CV(\mu_2)$  is large, knowledge of node locations in RGGs gives us a predictive advantage over only having knowledge of the distribution of node locations.





#### Future Directions

- Extension to **Soft RGGs** How much do network properties vary when we condition on knowledge of the node locations?
- Analytic approaches can we use techniques from random matrix theory to derive distributional properties of μ<sub>2</sub> or other network properties?





# Thanks for listening!! Any Questions?

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric

EPSRC Engineering and Physical Sciences Research Council



#### Mathematical Properties of RGGs

- Many properties of RGGs are hard to compute analytically
- For example, there is no closed form expression for the probability of a given ensemble member.
- However, we can estimate some of the important properties such as the **mean degree**

$$\kappa = \int_{\mathbb{D}} \gamma(|x - y|) dx dy$$
 (2)

The above equation works best as  $N \to \infty$  with **periodic boundary conditions** 

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric





#### Generation of RGGs with Specified Degree

For the case where the domain of interest is of unit volume the probability of some node j falling within the connection radius of a node i is given by the volume of the ball of radius R in d dimensions. Multiplying by this quantity by the number of remaining nodes in the network gives us an expression for the mean degree of the form:

$$\kappa = (N-1)\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d+2}{2})}R^d.$$
(3)

Inverting this formula allows us to obtain an expression for the connection radius required to generate RGGs with mean degree  $\kappa$  in d dimensions for a network with N nodes:

$$R = \frac{1}{\sqrt{\pi}} \left( \frac{\kappa}{N-1} \Gamma\left(\frac{d+2}{2}\right) \right)^{\frac{1}{d}}.$$
 (4)

This approach works well in practice for RGGs with periodic boundary conditions. However, for RGGs generated in  $[0,1]^d$  with solid boundaries the empirically observed mean degree of the ensemble  $\frac{d}{d} = \frac{1}{2} \frac{1}{$ 

Matthew Garrod Large Fluctuations in the Algebraic Connectivity of Gaussian Random Geometric



