Exploring Sentiment Dynamics on Spatially Embedded Random Networks

Matt Garrod, Supervisor: Dr Nick Jones,
Department of Mathematics, Imperial College London

EPSRC Centre for Doctoral Training
Mathematics of Planet Earth
Imperial College London
University of Reading

Motivation: Modeling Sentiment Dynamics

- Meeting global climate change goals will require large scale cooperation between individuals in society. In order achieve this it is crucial to be able to understand and quantify the factors which influence **public sentiment**.
- Climate sentiment can be though of as a *stochastic process* occurring on a social network.
- Previous studies have used data from **online social networks** in order to study public sentiment on climate change [1], [2].
- However, in many cases the underlying social network is **unobserved** and we may only have access (or partial access) to characteristics of individuals (e.g GPS coordiantes or sociodemographic information).

Q: How can we explore how the interplay between individuals experience of the climate and their social positions effects their opinions?

Figure : An example of a spatial network. The strength and number of social ties between different districts in London can depend on both their physical distance and demographic attributes. (Image courtesy of Till Hoffmann)

One approach to modeling social networks with unobserved connections is through **spatial network models** such as the Random Geometric Graph (see box below). These models allow us to incorporate the following attributes of individuals or groups:

- Their coordinates in physical space, $\underline{x}$
- A vector of demographic attributes $\underline{b}$ (e.g age, income, education level, ethnicity)
- Their experience of an **external climate field** $f(\underline{x}, t)$. f might represent average temperature, rainfall or some other climatological variable.

Random Geometric Graphs

A well studied spatial network model is the **Random Geometric Graph** (RGG). An RGG with N nodes in some domain D can be constructed as follows:

1) Sample N positions uniformly in D to obtain a vector of random positions $\underline{X} = (X_1, ..., X_N)$.
2) Construct a matrix of distances between nodes $d_{ij} = |X_i - X_j|$.
3) Connect two nodes i and j with a probability given by $\mathbb{P}(A_{ij} = A_{ji} = 1) = \gamma(d_{ij})$, where $\gamma : \mathbb{R} \to [0, 1]$ is typically a decreasing function of distance.

A common choice of connection function is a step function of the form:

$$\gamma(r) = \begin{cases} 1 & \text{if} \quad r \leq r_c \\ 0 & \text{if} \quad \text{otherwise} \end{cases} , \qquad (1)$$

where $r_c \in \mathbb{R}^+$ is a parameter which sets the threshold for connectivity between nodes.

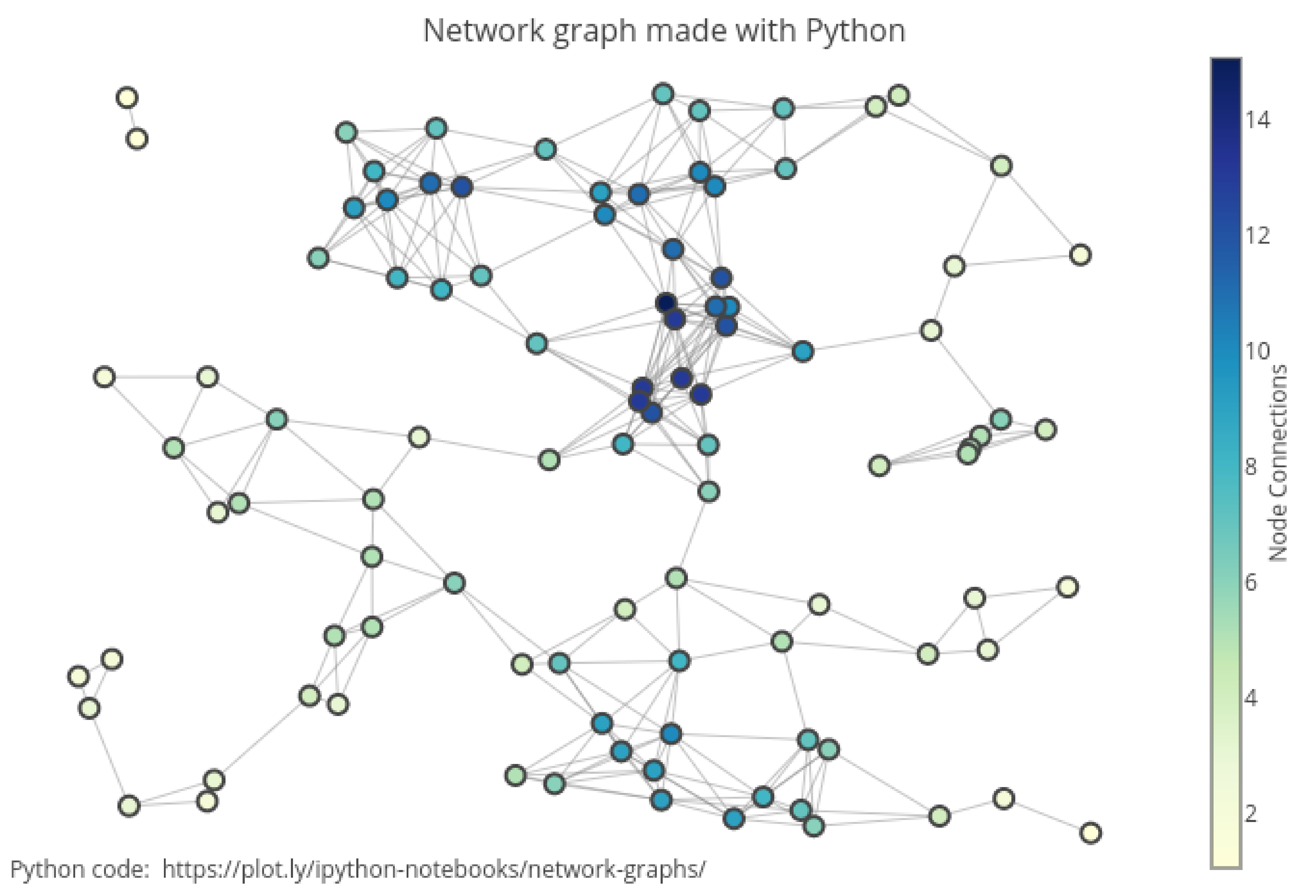


Figure : Typical realisation of an RGG for $N = 100$ with $D = [0, 1]^2$

The adjacency matrix: We can describe networks with an adjacency matrix A. For undirected networks we set $A_{ij} = A_{ji} = 1$ if nodes i and j share a link and $A_{ij} = A_{ji} = 0$ otherwise.

Modeling Social Networks

Information about nodes typically comes in the form of coordinates X_i in some *metric space* M. The principle of **homophily** states that nodes with similar properties are more likely to share a social tie.

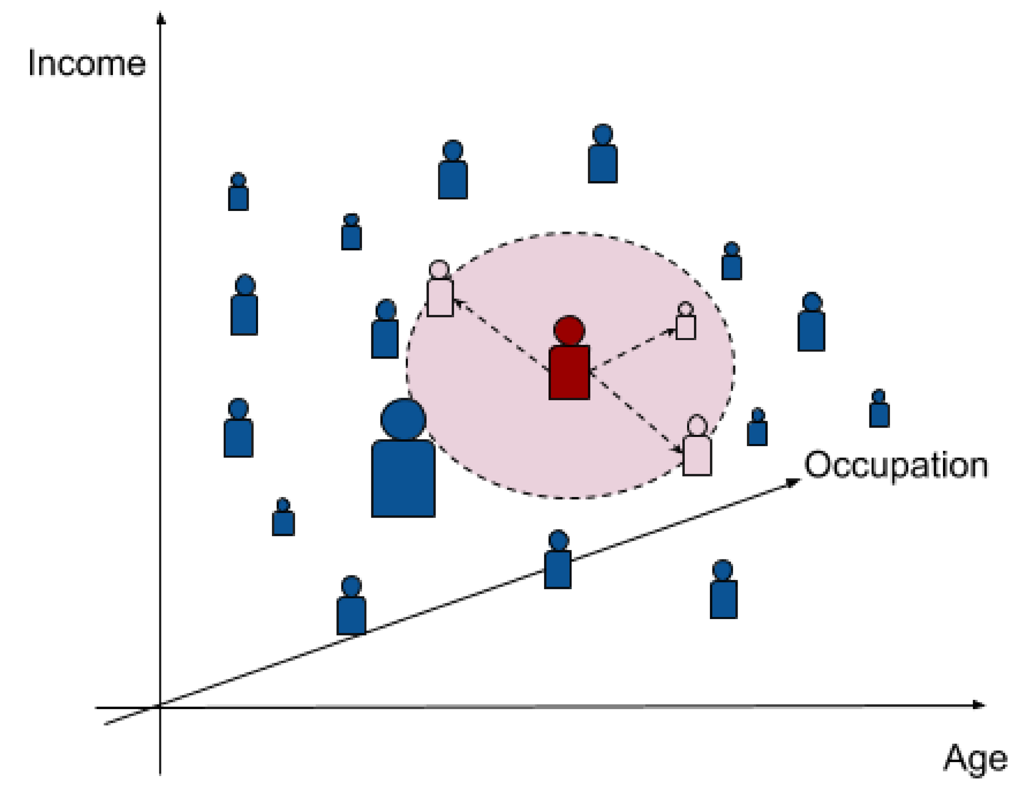


Figure : Individuals in society be though of as coordinates in some multi-dimensional social space. In social networks individuals are more likely to share a connection with those who lie closer to them in the space.

Current Work - Understanding Variability in Network Dynamics

Problem Statement:

1) Most network models involve uncertainty in the positions or connections between nodes. **Q:** How does variability in structure of spatial networks influence variability in important network properties?

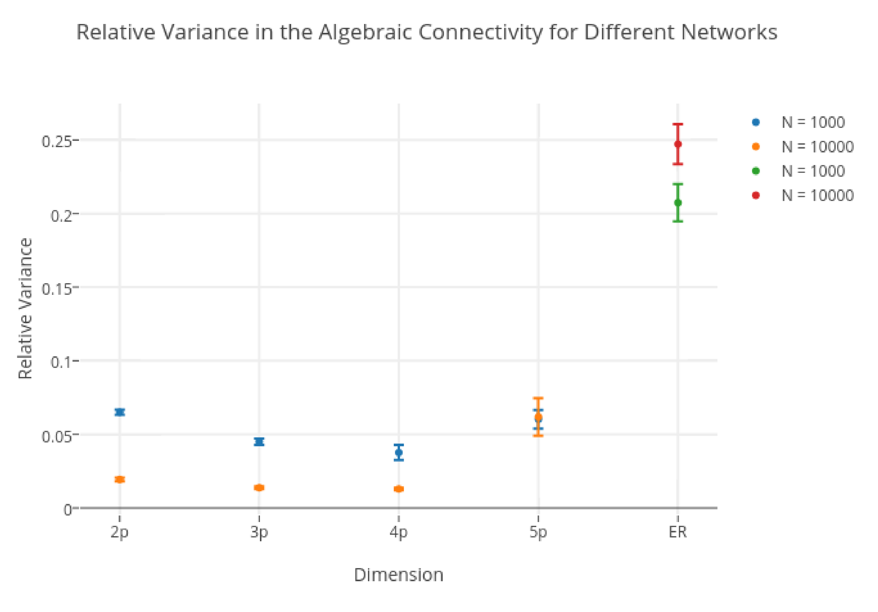
2) **Q:** Does knowledge of node locations and the probability of connection between individuals give us a meaningful predictive advantage?

- There are many metrics which can be used to study the dynamical properties of networks. For instance, the **algebraic connectivity**, μ_2 is related to various dynamical properties of networks. The algebraic connectivity is inversely proportional to the the *characteristic timescale of diffusion* on the network ($\tau \approx \frac{1}{\mu_2}$).
- For random networks the algebraic connectivity will be a **random variable** with distribution $\mathbb{P}(\mu_2)$. We can quantify the variability in μ_2 by considering the **relative variance** (also known as the coefficient of variation) :

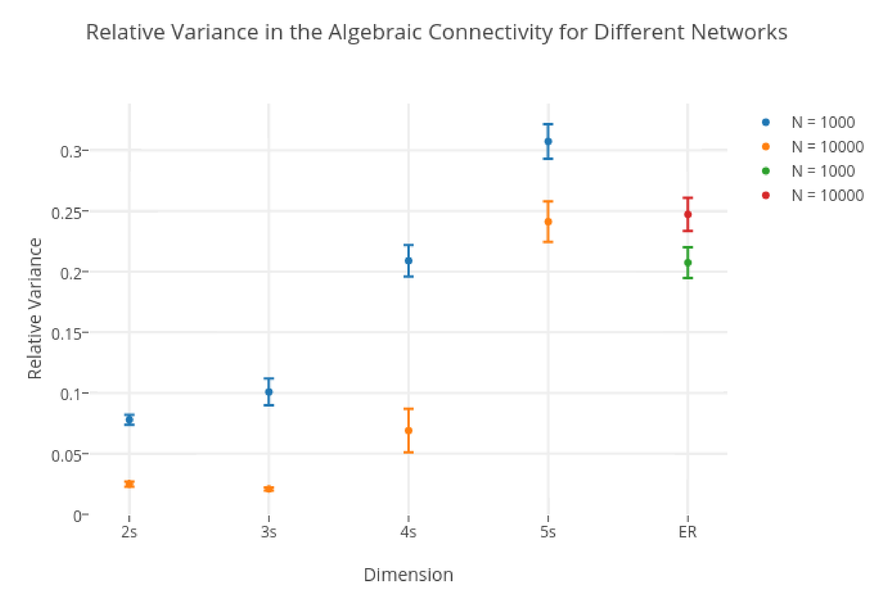
$$CV(\mu_2) = \frac{\sigma_{\mu_2}}{\mathbb{E}(\mu_2)} \qquad (2)$$

Conclusions (so far...)

- The intrinsic variation in the algebraic connectivity is larger for RGGs in higher dimensions. Knowing node locations in an RGG completely specifies the network structure (and therefore the value of μ_2).
- Consequently, in higher dimensions knowing node locations can give you "more of a win" compared with the case in low dimensions where variability between ensemble members is typically small.



(a) **Periodic Boundaries**



(b) **Solid Boundaries**

Figure : Scatter plot showing the relative variance in the algebraic connectivity distribution for RGGs in different dimensions. Shown for mean degree parameter $\kappa = 15.0$ for both periodic and solid boundary conditions. Also shown are the corresponding relative variance values for Erdős-Rényi Random Graphs (ER).

References

[1] Kirilenko, Andrei P., Tatiana Molodtsova, and Svetlana O. Stepchenkova. "People as sensors: Mass media and local temperature influence climate change discussion on Twitter." Global Environmental Change 30 (2015): 92-100.
[2] Williams, Hywel TP, et al. "Network analysis reveals open forums and echo chambers in social media discussions of climate change." Global Environmental Change 32 (2015): 126-138.