

# **IDC 407, Network Science**

## Information Sharing and Epidemic Modelling in Ecological Networks

Dhruva Sambrani (MS18163)

Aalhad Bhatt (MS18118)

Dhananjay Joshi (MS18065)

21<sup>st</sup> March 2022

---

### **Contents**

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Network Parameters</b>	<b>2</b>
<b>3</b>	<b>Models</b>	<b>3</b>
<b>4</b>	<b>Methods</b>	<b>3</b>
4.1	Extracting Data . . . . .	3
4.2	Implementing the Network . . . . .	4
4.3	Output . . . . .	4
4.4	Correlation Measures . . . . .	4
<b>5</b>	<b>Results</b>	<b>5</b>
<b>6</b>	<b>Conclusions</b>	<b>6</b>

---

# 1 Introduction

Social and contact networks are receiving increased attention as powerful tools to investigate disease spread in animal populations. In this term paper we carry out a comparative analysis of social networks of about 400 animal species from different taxa: mammals, birds, reptiles and insects in the context of some common disease spread models and try to draw meaningful conclusions. We also consider some information sharing models and compare the responses generated by different networks depending on their varying network parameters.

## 2 Network Parameters

We used the following set of network parameters to characterize the current database:

- **Number of Nodes and Edges:** For the present analysis, we only considered undirected animal networks having a minimum of 50 nodes and 100 edges. This filtered out the excessively sparse networks.

- **Average Degree:**  $\frac{1}{N} \sum_i^N \sum_j A_{ij}$  : The mean degree of a node in the network

- **Connectance** : Refers to the proportion of the realized links of the total number of possible links:

$$\frac{\sum_{i,j} A_{i,j}}{N(N-1)}$$

- **Assortativity** : Assortativity is a number between -1 and 1 that gives us a measure of mixing in the network. A degree of assortativity of 1 means that the network is perfectly mixed.

$$A = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) \delta(c_i, c_j)}{2m - \sum_{ij} (k_i k_j / 2m) \delta(c_i, c_j)}$$

- **Betweenness Centrality:** The betweenness centrality of a node gives us an idea about how many times that node lies between the geodesic path between two randomly selected vertices of the network. Here we consider the average of betweenness centrality measure of all vertices.

$$X = \frac{\sum_i \sum_{ts} n^i}{N}$$

- **Eigenvector Centrality:** Eigenvector Centrality awards a vertex a centrality score based upon the number of neighbors and how important a neighbour is. The average Eigenvector Centrality here is given by:

$$X = \frac{\sum_i \kappa_1^{-1} \sum_j A_{ij} x_i}{N}$$

- **Clustering Coefficient:** The clustering coefficient gives us a measure of transitivity in the network. It is given by :

$$C = \frac{\text{No. of triangles} \times 6}{\text{Total number of paths of length 2}}$$

- **Mean Path Length:** The average value of shortest paths between all the nodes:

$$M = \frac{\sum_{i,j} \mu_{i,j}}{2N}$$

where  $\mu_{i,j}$  is the shortest path between nodes  $i$  and  $j$ .

### 3 Models

We consider 5 different epidemic models, all of which are stochastic and discrete time. These are types of compartment models where each node belongs to any one compartment. Transitions, or how a node goes from one compartment to another are based on which compartment it and its neighbours belong to. The compartments under consideration in the models are Susceptible (S), Infected (I), Removed (R), Exposed (E).

1. SI
2. SIR
3. SIS
4. SEIS
5. SEIR

Since many of the models have the same transitions, the mechanism for each transition and the associated parameters are mentioned below.

- $S \rightarrow I$ : Each S node becomes I with a probability of transition  $\beta$  at each step. The transition requires that at least one neighbour be I. Only occurs if the model does not have E class.
- $S \rightarrow E$ : Each S node becomes E with a probability of transition  $\beta$  at each step. The transition requires that at least one neighbour be I.
- $E \rightarrow I$ : Each E node becomes I with a probability of transition  $\alpha$  at each step.
- $I \rightarrow R$ : Each I node becomes R with a probability of transition  $\gamma$  at each step.
- $I \rightarrow S$ : Each I node becomes S with a probability of transition  $\lambda$  at each step. Only happens in SIS and SEIS.

Each model was initialised with 0.05 of the total nodes being randomly chosen and changed to I. Rest were S. Parameter values chosen were  $\beta = 0.1$ ,  $\alpha = 0.05$ ,  $\gamma = 0.005$ , and  $\lambda = 0.005$ .

## 4 Methods

### 4.1 Extracting Data

We used the animal network repository compiled by the authors of this [1] paper. They have collected data from various ecology publications and converted into a python ready format using NetworkX.

## 4.2 Implementing the Network

We used the python libraries **Networkx** to filter suitable networks from the database and calculate the various network parameters. The epidemic and information sharing models were implemented using the python **ND.Lib** library.

## 4.3 Output

We found out the normalised total infection duration after running the model for 5000 time steps. This was calculated by dividing summing over the duration each nodes was in the I state and dividing by the size of the network. This gives a measure of the severity of the infection both in terms of proportion of the nodes infected and the duration. This was plotted against the different network parameters for each network and infection model.

## 4.4 Correlation Measures

We calculated two types of correlation measures for each of the Network Parameter vs the total infected plot for each model.

- **Pearson Correlation Coefficient:** This is a measure of linear correlation between two independent variables : X and Y and is given by the relation:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Where  $\text{cov}(X,Y)$  is the covariance between X and Y, and  $\sigma_X \sigma_Y$  are the respective standard deviations. The value of  $\rho$  varies between -1 to 1 where -1 represents negative correlation, 0 means no correlation and 1 refers to complete correlation.

- **Distance Correlation:** This measure checks for both linear and non linear correlation and returns a value between 0 (non correlated and independent) to 1 (maximally correlated and linear). Distance correlation does not compare the distribution of the two random variables around their respective means. Instead, it is based off of the pairwise Euclidean distances between two points.

$$a_{ij} = ||X_i - X_j||, b_{ij} = ||Y_i - Y_j||$$

$$A_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, B_{ij} = b_{ij} - \bar{b}_{i.} - \bar{b}_{.j} + \bar{b}_{..}$$

$$\text{dcov}(X,Y) = \frac{1}{n^2} \sum_{i,j} A_{ij} B_{i,j}$$

$$\text{dcor}(X,Y) = \frac{\text{dcov}(X,Y)}{\sqrt{\text{dcov}(X,X) \text{dcov}(Y,Y)}}$$

## 5 Results

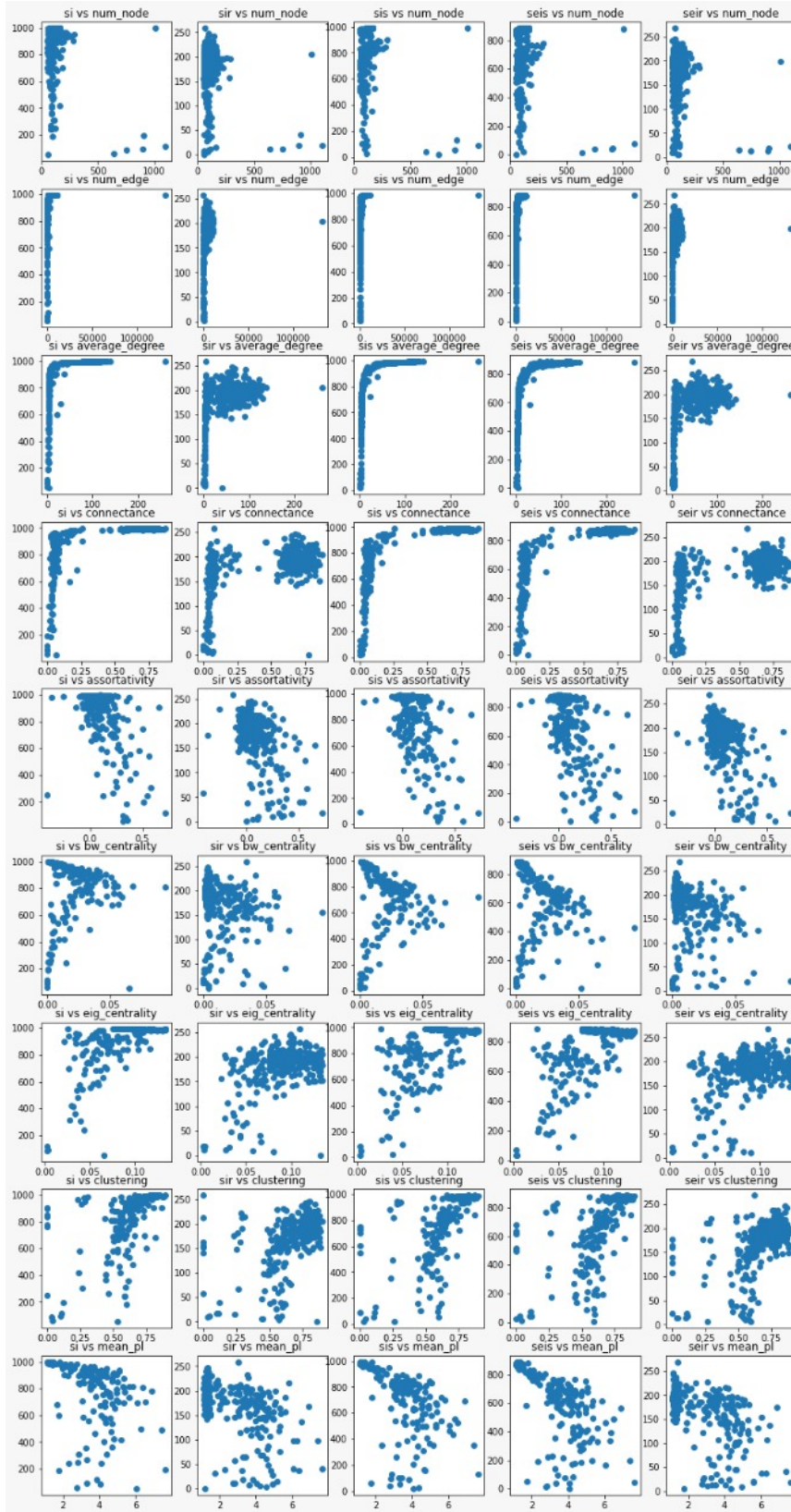
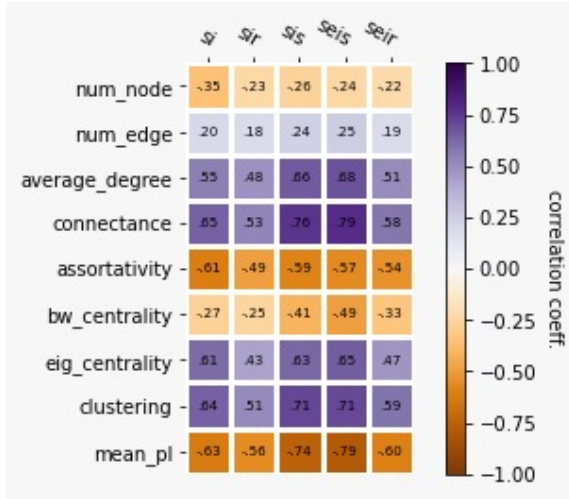
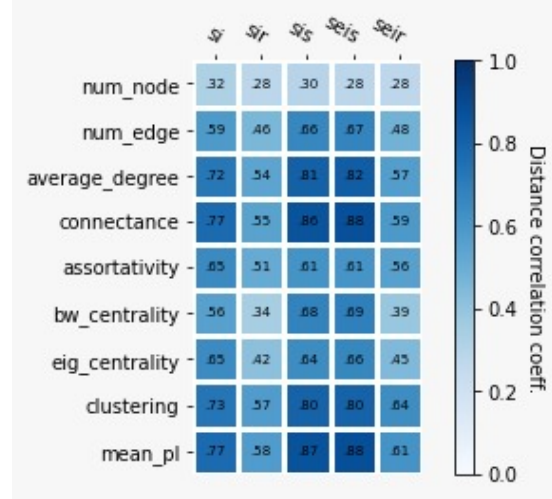


Figure 1: Net infection vs various network parameters



(a) Pearson Coefficient for each network parameter and epidemic model



(b) Distance Correlation for each network parameter and epidemic model

Figure 2: The correlation coefficients reveal which network parameters affect the rate of spread of infection the most

**NOTE** The SI model can be used as a proxy for information sharing as it has 2 states, 0 corresponding to someone who doesn't know the information and 1 corresponding to someone who does. Once you know the information, you stay in that state forever.

## 6 Conclusions

Across all the models, the correlation coefficients for normalised total infection duration with different network measures were mostly uniform. SIR and SEIR showed slightly different values from the other 3 while being fairly similar to each other. In the other 3, SI was slightly different from the others. There was no large scale difference, and if one model had a strongly positive correlation, then all the others did as well. As a result, we can speak of the correlation coefficients without specific reference to any model.

For Pearson's Correlation Coefficient, we see that connectance and clustering have the strongest positive correlations while average shortest path length has the strongest negative one. This makes sense, as longer paths will slow disease spread down. High clustering and connectance mean that there are multiple paths for the disease to spread, and therefore it can spread faster. Betweenness centrality, number of nodes and number of edges seem to play the smallest role.

Pearson's Correlation Coefficient only looks at linear relationships however, and it might not detect certain important relationships. The direction of the correlation matches what we would expect. More nodes and higher degree assortativity imply weaker disease spread, while the number of edges implies faster spread.

To solve this problem, we also compute the distance correlation. Distance correlation does not specify whether the relationship is positive or negative, but gives a way to calculate the strength independent of the form of the data. Here, the main changes are for betweenness centrality and number of edges, which both now have stronger correlations.

## References

- [1] Sah, P., Méndez, J.D. & Bansal, S. *A multi-species repository of social networks* ([link](#))
- [2] Wikipedia *Pearson Correlation Coefficient* ([link](#))
- [3] Peter Gleeson *Finding Correlations in Non-Linear Data* ([link](#))
- [4] Mark Newmann *Networks* ([link](#))