Contact Networks for Mathematical Models of Infectious Disease

Karen McCulloch
RMIT University

This project focuses on the research question 'What are the percolation properties of contact networks that are spatial but also have a strongly overdispersed degree distribution?' It involved modifying existing code in the programming language R to produce algorithms for networks that combine geographical constraints on connections with individual heterogeneity in how often such connections occur. The percolation properties of the resulting networks are of interest because long range connectivity in such networks corresponds to the possibility of disease outbreaks; where the contacts in the network are of the type that allows transmission of the infectious agent. Part of this project also focused on having a look at how a real wildlife contact network compares to the random networks produced in addressing the research question above.

Knowledge of the structure of human and animal contact networks is of vital importance in terms of studying the spread of infectious diseases. The structure tells us whether the disease is likely to spread, how much of the population will be affected and what effort is required to stop an outbreak. A recent example of why contact networks are important is the Tasmanian devil and the facial tumour disease. Note here that the Tasmanian devil is just a motivating example; the above research question which I will be addressing is more theoretical and general. The Tasmanian devil is the world's largest surviving carnivorous marsupial. In 2008 the Tasmanian devils' status was upgraded to endangered species due to its vulnerability to the facial tumour disease. Devil facial tumour disease is a fatal condition in Tasmanian devils which is characterised by the appearance of facial cancers. The tumours or cancers are first noticed in and around the mouth as small lesions or lumps. These develop into large tumours around the face and neck and sometimes even in other parts of the body. Devil facial tumour disease is very unusual as it is one of only three recorded cancers that can spread like a contagious disease. The cancer is passed from devil to devil through biting. Due to the lack of genetic diversity among the Tasmanian devils the tumour cells aren't rejected by their immune system thus causing the disease to spread.[1]

[1] http://www.dpiw.tas.gov.au/inter.nsf/WebPages/LBUN-5QF86G

In February 2010, the devil facial tumour disease was detected across more than 60% of Tasmania. So, is the devil population well connected? And does it percolate? Before we can answer these questions we need to define a few terms and properties that are important when looking at the topology of a network. The square lattice is an example of a network which is composed of individual nodes or vertices. A bond, or edge, is a connection between any two vertices. A cluster is a group of vertices connected by bonds. The degree of a vertex is the number of connections or edges it has. If a network is well connected then we say that it percolates. Note that contact networks can be of different types. It depends on the type of contact that a disease uses to transmit. It can be direct contact (such as biting) or indirect contact (such as flu which transmits via the environment). With respect to direct contacts, the Tasmanian devil population appears to be well connected and thus we can say that it percolates.

If we randomly generate bonds between vertices to form a random network, then each bond has a probability, p, of existing. The higher the probability, p, the more likely it is that the network has an infinite cluster. An infinite cluster, or large component, is a cluster which spans the network. There is a critical point, p=pc, at which below this point an infinite cluster does not exist. Above this point an infinite cluster always exists, and the chance that an arbitrary node belongs to it increases. This critical point is known as the percolation threshold. If p>pc the network is well connected and thus percolates. If p<pc the network is not well connected and does not percolate. On the square lattice the percolation threshold is 0.5.

What happens when contact networks are spatial *and* have high individual variation? Note here that these are opposing ideas, when a network is spatial the likelihood of a disease spreading is low – spatial networks require lots of "little hops" to get anywhere; when individual variation in degree is high this creates "hubs" in the network, or superspreaders, and the likelihood of a disease spreading is higher. Using the programming language R I produced algorithms for networks that combine geographical constraints on connections with individual heterogeneity in how often such connections occur. This was achieved by randomly assigning each vertex a coordinate in space and then allocating each of these vertices a random 'fitness', xi. The fitness of a vertex represents the socialability (willingness to be social) of each individual, for example in the Tasmanian Devil population two dominant males might be more likely to come into contact with each other than, say, two shy males.

To develop a plausible model for the contact rates of pairs of individuals we started with a purely geographical model

$$k_{ij} = e^{-\lambda s_{ij}}$$

Where sij is the Euclidean distance between node i and node j and lambda is a scaling parameter. However, with this model, the average contact rate varies with lambda. To compare the behaviour of the networks it is convenient to have the average contact rate constant. Therefore we multiplied the above model by lambda squared to form

$$k_{ij} = \lambda^2 e^{-\lambda s_{ij}}$$

Now, it can be shown that the average contact rate of individuals is independent of lambda.

The heterogeneity of individuals is also important when determining which individuals have contact with other individuals. For this reason, we want to form a model that includes the purely geographical model, but also takes into account the heterogeneity of individuals to see if it produces plausible networks

$$k_{ij} = \lambda^2 e^{-\lambda s_{ij}} x_i x_j$$

We chose to multiply the 'fitness' of node i and node j in this model so that the probability of node i been connected to node j is equally weighted. That is both the sociableness of node i and node j are of equal importance when determining whether or not they come into contact with each other.

i. How does changing lambda affect the network?

We took a small range of lambda values to see how these random networks behaved when the scaling parameter lambda was changing and everything else was kept constant. We found that for this particular contact rate, as lambda increases the network became more spatial.

ii. How does varying lambda change the percolation threshold?

Using the same three values of lambda as before, we had a look at how the percolation threshold behaved. As the value of lambda increased, it appeared that so did the percolation threshold for these random networks being generated. See figure 1 in the Appendix.

iii. How does changing the individual variation affect the network?

We took a small range of values for the variance of the individual fitness's to see how these random networks behaved when the individual variation was changing and everything else was kept constant. As the individual variation was increased, it appeared that the networks had an increasing number of isolated vertices and superspreaders.

iv. What effect does variation have on the percolation threshold?

Using the same range of values for the individual variation as before, we had a look at how the percolation threshold behaved. We found that for each of the values of lambda used previously, the variance had different effects. For smaller lambda values of 0.5 and 1, as the variation increased the percolation threshold was pushed closer to zero. However, for the larger lambda value of 2, as the variation increased there was no clear effect on the percolation threshold. Also, for all lambda values used previously as the variation increased the probability that an arbitrary vertex would belong to an infinite cluster decreased. See figures 2 and 3 in the Appendix.

In most networks high individual variation in contact rate translates into networks having a few, highly connected individuals. In epidemiology, such individuals are sometimes referred to as superspreaders. In human contact networks superspreaders

often have a significant influence on the spread of disease. This is the case for sexually transmitted diseases such as HIV where control is made more difficult by superspreaders.

However, from the above results, for our spatial networks and our "fitness" model that generates individual heterogeneity, it is possible that the influence of superspreaders on the spread of disease is less significant the higher the value of lambda. One possibility is this is because for strongly spatial networks the superspreaders may tend to be isolated from the majority of the network, despite being superspreaders. Thus their effect on the spread of disease – which is usually to accelerate spread and increase the chances of an outbreak – would be local rather than felt throughout the entire network.

The second part of this project was to address the question of how do random networks compare to real wildlife networks? For this part of the project we used data previously collected for Voles. A Vole is a small rodent resembling a field mouse. The Vole data was collected by setting one hundred traps up in a 100m by 100m grid on different sites. Each Vole was given a unique ID number; this number was recorded every time a Vole was trapped. The Voles were said to have contact with other Voles if they were caught in the same trap on the same site.

Using the programming language R I created code to extract the relevant information from excel files. This information included Vole ID numbers, the number of times each Vole was caught and the specific traps they were caught in. Each Vole was then given an x and y coordinate to represent their coordinate in space; these coordinates were determined by the traps each Vole was caught in. Edges were then drawn between two Voles, or vertices, if they had a trap in common. Once the Vole contact network had been set up, code was added to compute the lengths of all the edges within the network so that an edge length distribution histogram could be produced. Similarly, we also computed the individual node degrees (or counted the number of connects, edges, that each Vole had in the network) in the network and produced a degree distribution.

Using the same spatial point pattern as for the Vole data, we then modified the way that edges were drawn between two vertices. Instead of using the information about the Voles to draw edges between vertices we randomly assigned each edge to two vertices. We then compared the degree distribution and the edge length distribution of the edges determined using the Vole data and of randomly assigned edges.

It is important to note that we can only visually compare the Vole and the random networks at this stage without performing further statistical inferences. This is merely the preliminary stages of analysing the contact network of voles. By visual comparison the random network and the Vole network have significant differences in structure (figure 4 in Appendix). We can also see this by comparing the degree distribution and the edge length distribution of both networks (figures 5 and 6 in Appendix). As the

variance of the degree of the Vole's model is significantly higher than the mean, this suggests that there is a significant amount of individual heterogeneity between the individual Voles. That is, some Voles are very highly connected or social while others only have a few connections or are perhaps shy individuals. The mean edge lengths also suggest that in the vole data network they form connections with other voles that are closer to them than in the random network. It appears that, based only on the small amount of research we have done, the vole network is both spatial and has an overdispersed degree distribution.

I wish to thank my supervisor Dr Stephen Davis for all his guidance and support as without which this project would not have been possible. I would also like to thank AMSI for their generous support which allowed me to undertake this project and gain invaluable research experience. Lastly I would like to thank the CSIRO for hosting the Big Day In as this was a fantastic event which I have benefited greatly from.

Appendix

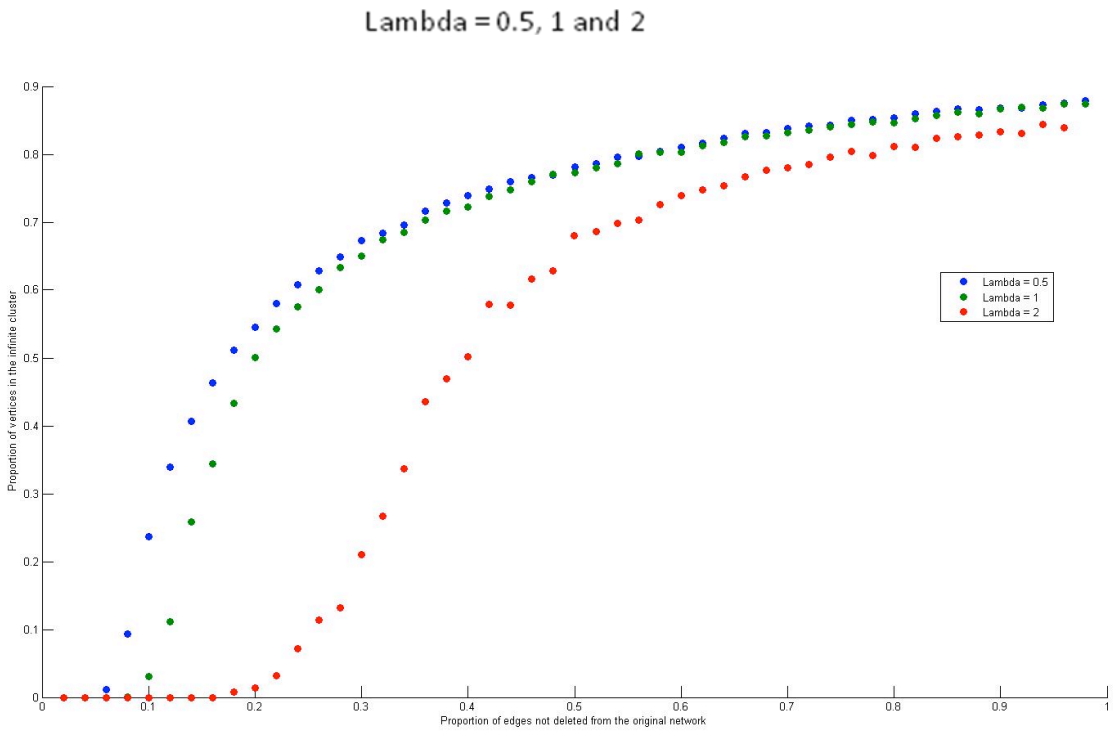

Figure 1. Effect of varying lambda on the percolation threshold. Lambda = 0.5, 1 and 2 are the blue, green and red curves respectively. The y-axis is the proportion of vertices in the infinite cluster and the x-axis is the proportion of edges not deleted from the original network generated.

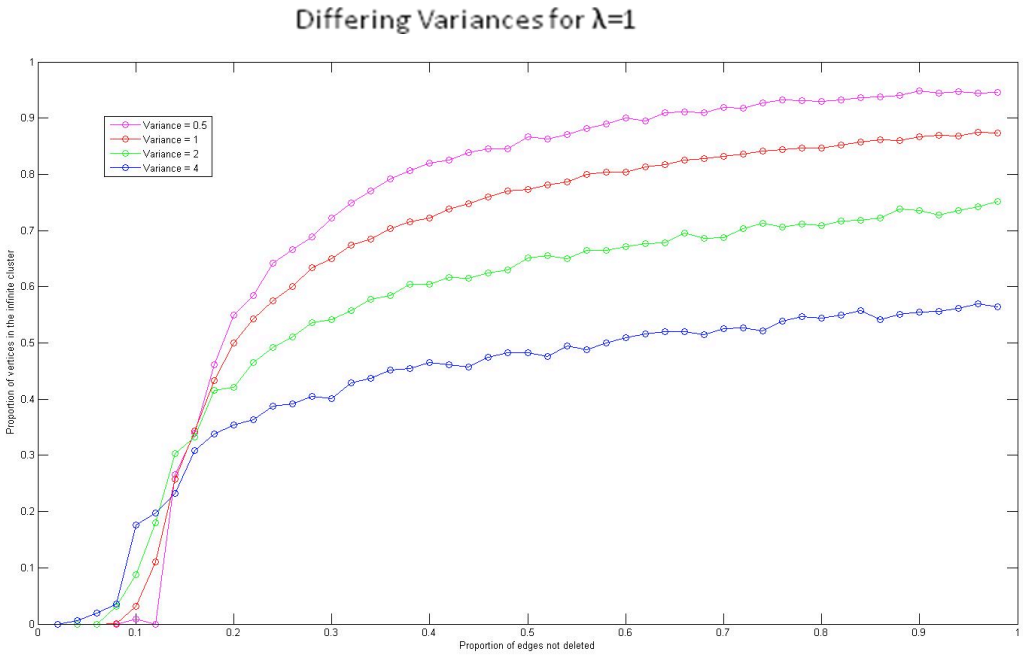

Figure 2. Effect of changing the variance on the percolation threshold. Lambda = 1 for Variances of 0.5, 1, 2 and 4 are the pink, red, green and blue curves respectively. The y-axis is the proportion of vertices in the infinite cluster and the x-axis is the proportion of edges not deleted from the original network generated.

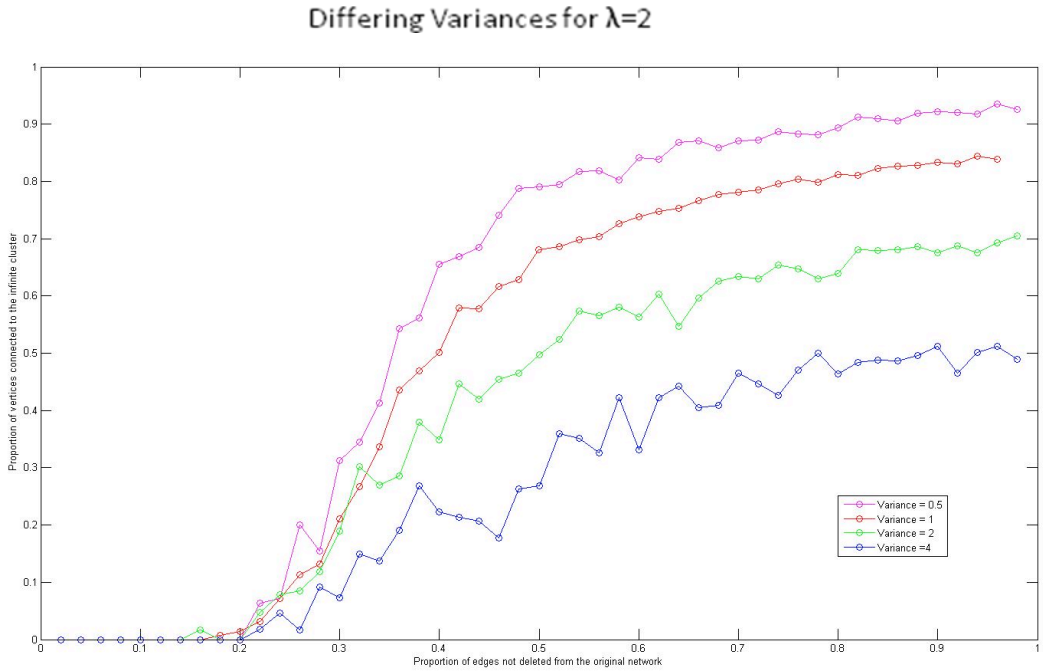

Figure 3. Effect of changing the variance on the percolation threshold. Lambda = 2 for Variances of 0.5, 1, 2 and 4 are the pink, red, green and blue curves respectively. The y-axis is the proportion of vertices in the infinite cluster and the x-axis is the proportion of edges not deleted from the original network generated.

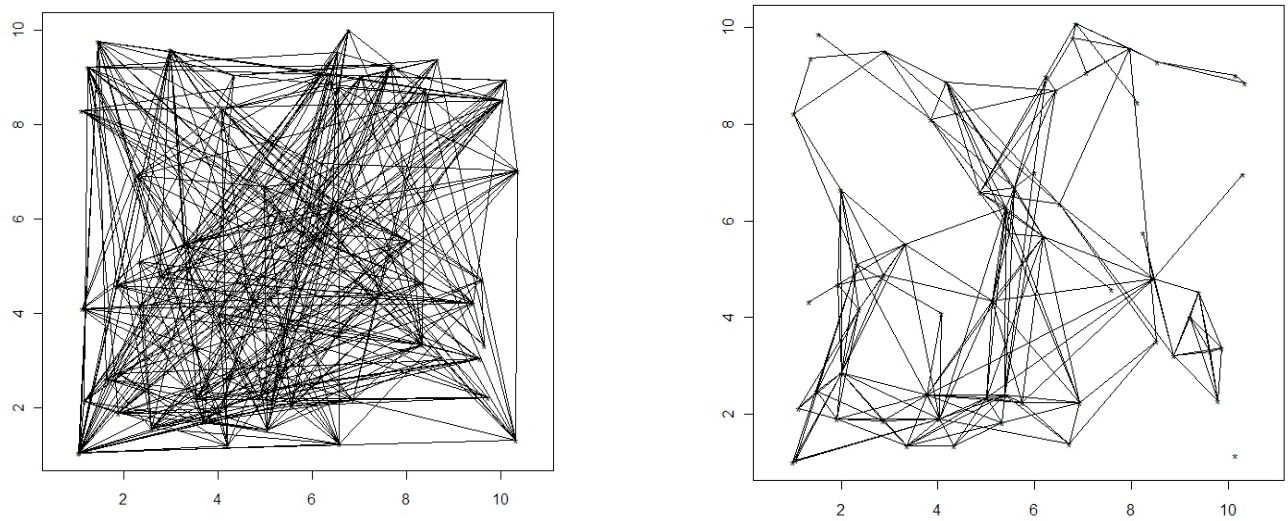

Figure 4. Left: Random network. Right: Voles network

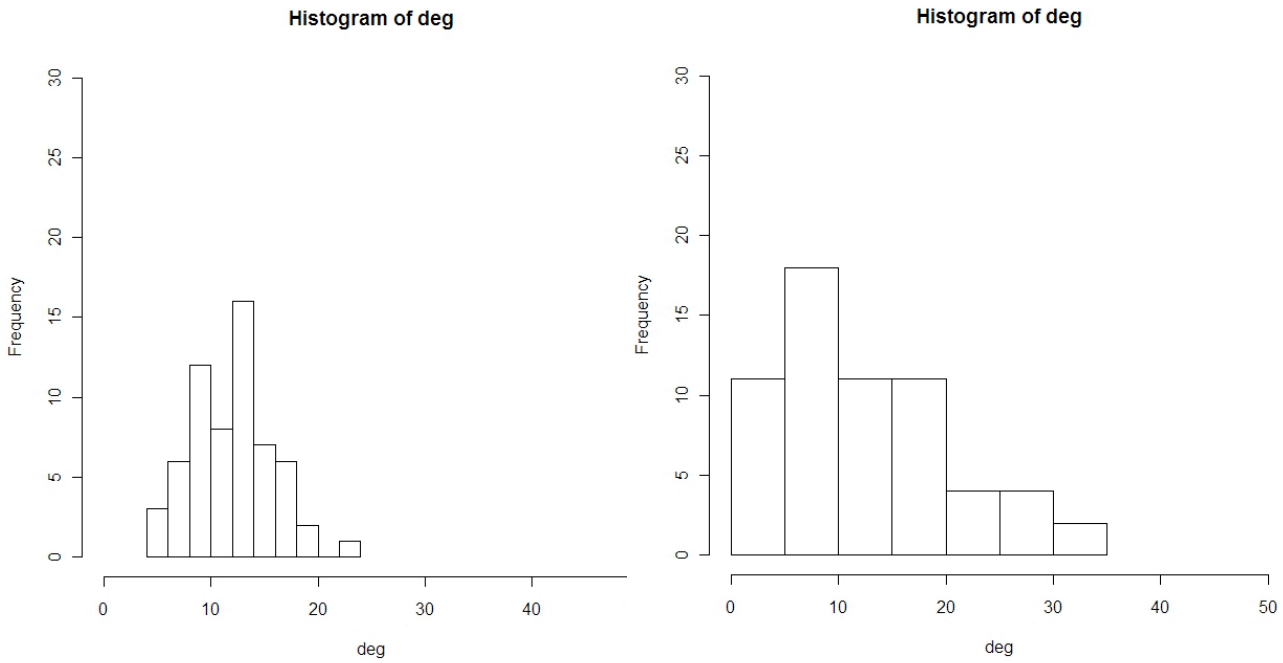

Figure 5. Degree distributions of the Random network and the Voles network respectively.

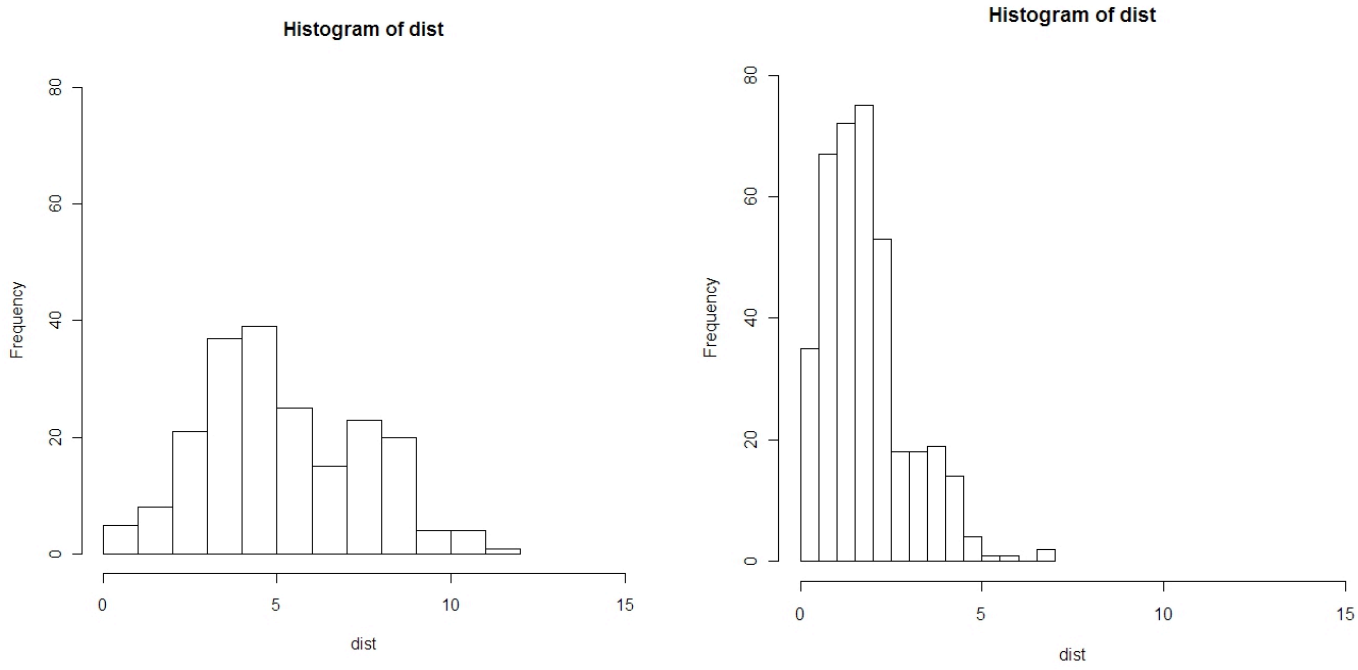

Figure 6: Edge length distributions of the Random network and the Voles network respectively.

Karen McCulloch received a 2010/11 AMSI Vacation Research Scholarship