

# Percolation on a spatial network with individual heterogeneity as a model for disease spread among animal host populations

S. A. Davis<sup>a</sup>

<sup>a</sup>RMIT University, GPO Box 2476V, Melbourne, Victoria 3001, AUSTRALIA  
Email: [stephen.davis@rmit.edu.au](mailto:stephen.davis@rmit.edu.au)

## Abstract:

Recent efforts to directly or indirectly observe animal contact networks reflect an increasing awareness that the spread of infectious diseases, and their control, can be critically affected by the contact structure of the host population. This has long been realised for sexually-transmitted diseases of humans but has also recently been shown for seasonal influenza, where only casual contact, or indirect contact, is required for transmission. Many animals, though by no means all, are radically less mobile than humans and their social contact networks are consequently spatial networks where each host can be assigned a location in space and the rate of contact between two hosts depends on the distance between them. Such contact networks have received little attention and are poorly understood. Here we present a model for animal contact networks that allows for both spatial constraints and individual heterogeneity: let  $x_i$  be the sociability of host  $i$ , let  $k_{ij}$  be the *rate* of contact between hosts  $i$  and  $j$ , and let  $s_{ij}$  be the Euclidean distance between  $i$  and  $j$ , then

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

where  $\lambda$  determines the scale over which the spatial constraints operate. To study the transmission of an infectious agent on large scale realisations of these network models we use long-range percolation. That is, we relate the probability of an open, *directed* edge between  $i$  and  $j$ , given  $i$  is infected, to the contact rate between the two hosts by

$$p_{ij} = 1 - e^{-v k_{ij} \tau_i},$$

where  $v$  is the probability of actual transmission given contact occurs and  $\tau_i$  is the infectious period of host  $i$ . We define outbreaks in terms of spatial spread: occurring when the distance between a newly infected host and an initially infected host exceeds certain milestones. The preliminary results presented here show that for at least this model of contacts (i) the stochasticity in model behaviour is a combination of the spatial constraints and the variance of the  $x_i$  rather than just the variance of the  $x_i$ , and (ii) stronger spatial constraints on contact rates tends to nullify the effect of individual heterogeneity to increase pathogen spread. We would hence argue that tailored methods of network analysis are needed that can quantify the role of space in determining contact rates between animal hosts. At both a national and global scale, public health and veterinary authorities periodically face challenges from new pathogens that arise in wildlife and livestock populations. These can pose a threat to human health, such as in the case of recent and recurrent outbreaks of Hendra virus, or, in the case of the Devil Facial Tumour Disease, they may threaten native species. Our preparedness and ability to control such pathogens will be markedly improved as we better understand the contact networks of animal hosts, and the implications of their structure for the spread of infectious disease.

**Keywords:** mathematical model, epidemiology, contact network, wildlife.

## 1 INTRODUCTION

The recent flurry of field studies on wildlife contact networks (see Bohm et al. [2009]; Craft et al. [2009]; Hamede et al. [2009]; Perkins et al. [2009]; Porphyre et al. [2008]) have been motivated by a general interest in the epidemiological consequences of such networks, or by an interest in a particular pathogen that is endemic in the population being observed. Underlying such studies is an increasing awareness that the dynamics of infectious diseases, and their control, can be critically affected by the contact structure of the host population. For example, it is now well understood that human sexual networks have degree distributions with relatively fat tails (Liljeros et al. [2001]) which has important implications for the dynamics and control of sexually transmitted diseases (May [2006]). More recently, epidemiologists are understanding that there can be equally important consequences where only casual contact, or indirect contact, is required for transmission. In a fascinating study, Christakis and Fowler [2010] found that even in the case of influenza, where the contacts required for transmission are of a very casual nature, the social network topology can be used to achieve early detection. This is done by monitoring the friends of randomly selected individuals, a tactic that succeeds because highly connected individuals are both likely to be monitored and are typically among the first to become infected.

Less is known or understood for disease systems of wildlife. Two of the recent studies on wildlife contact networks, reported by Porphyre et al. [2008] and Bohm et al. [2009], have been motivated by concern for bovine tuberculosis and both reported high individual heterogeneity in the observed contact rates. This is important because individual heterogeneity creates a contact network that is easier for a pathogen to spread through, compared to one in which there is low or no variation. A feature of animals though is that they are far less mobile in comparison to modern humans, and they are often territorial. That is, two animals may normally have contact only if their home ranges overlap, with exceptions arising from long distance dispersal events. In such cases, the contact networks will reflect strong spatial constraints on who in the population has contact with whom, the extreme case being that contact only occurs between animals having adjacent territories. This means that some pathogens of wildlife will have a spatial barrier to overcome. That is, they must transmit enough to escape local build up of infected and recovered animals which can lead to early fade-out.

An example of a pathogen that faces strong spatial constraints in some of its host populations is sylvatic plague (*Yersinia pestis* infection in wildlife hosts). In particular, Davis et al. [2008] showed that for plague in great gerbils in Central Asia, there is empirical evidence for a sharp difference between the scale at which flea movements occur (these are necessary for transmission between family groups of the rodent host and they are almost always  $<200$  metres) and the scale of the landscapes occupied by the rodent species and monitored for plague by public health authorities (these areas are hundreds of square kilometres). One approach to modelling such systems is long-range bond percolation, where hosts are the nodes of a (spatial) network and bonds (or open edges) between nodes represent transmission of the pathogen (in the case of plague, an infectious flea has been transported between family groups).

Here we use long-range bond percolation to investigate pathogen dynamics on model contact networks that include both spatial constraints and individual heterogeneity. The algorithm used to generate contacts makes use of the good-get-richer mechanism proposed by Caldarelli et al. [2002] where each node is given a “fitness” value generated from a probability distribution  $\rho(x)$  and the rate of contact between two nodes is determined by the sum or product of these values. In this context the values represent sociability, an intrinsic willingness to make or have contact with others. Fitness is a misleading label that we will not use from this point on and replace with sociability. The model proposes that the rate of contact between two animals depends on the sociability of *both*. This is intuitively appealing for contacts of wildlife. Furthermore, Caldarelli et al. [2002] showed that the good-get-richer mechanism can give rise to a wide range of network types, including scale-free networks from non-scale-free distributions for  $\rho(x)$ .

## 2 THE NETWORK MODEL

Consider a population of hosts, each with a (unique) location in an area  $A$ . This position represents a central location for the host, but hosts are not fixed in space and they move freely about such that contacts occur. Let the spatial arrangement of these central locations for the hosts in  $A$  follow complete spatial randomness, defined by Diggle [2003], with each point being an independent random sample from the

uniform distribution on  $A$ . For each pair of hosts,  $i$  and  $j$ , let the *rate* of contact between them be given by  $k_{ij}$ , and let  $s_{ij}$  be the Euclidean distance between their central locations. For contact rates giving rise to purely geographical networks,

$$k_{ij} = e^{-\lambda s_{ij}}, \quad (1)$$

where  $\lambda$  essentially determines the spatial scale at which distances between nodes begin to affect their rate of contact.

Several studies of wildlife contact networks though, suggest substantial variation between individual animals in the rates of contact and the numbers of contacts. One way of incorporating such heterogeneity is to assign a "fitness" to each host that represents the sociability of that animal, its willingness to either initiate contact with other individuals in the population or to respond favourably when approached. Let  $x_i$  then be the sociability of host  $i$ , a random number drawn from a probability distribution  $\rho(x)$ , and consider

$$k_{ij} = f(s_{ij}, x_i, x_j). \quad (2)$$

Such an expression for the  $k_{ij}$  effectively adds geographical constraints to the good-get-richer mechanism proposed by Caldarelli et al. [2002]. In this paper, we consider the particular model

$$k_{ij} = \lambda^2 e^{-\lambda s_{ij}} x_i x_j. \quad (3)$$

We make this choice because it has the useful property that  $E[k_i] = E[\sum_j k_{ij}]$  is independent of  $\lambda$  (see Appendix). This is useful because varying  $\lambda$  does not change the average degree of the contact network (or the total contact rate of the network). It does mean, though, that for sharper spatial constraints (higher values of  $\lambda$ ) hosts will tend to have higher contact rates with their closer neighbours to make up for the loss of contacts with more distant neighbours. And hence, the mean number of *new* contacts does change with  $\lambda$ .

We choose the gamma distribution for  $\rho(x)$  as a flexible, 2-parameter distribution. This allows us to change the variance of the  $x_i$  while keeping the mean value of  $x_i$  constant and (consequently) keeping the average degree of the contact network constant (Appendix).

### 3 LONG-RANGE PATHOGEN PERCOLATION

To study the spread of a pathogen on the family of contact networks just described we consider long-range percolation where we let  $p_{ij}$  be the probability of an open edge between node  $i$  and node  $j$ . This is a directed edge and the probability is a conditional probability – the probability of transmission from host  $i$  to host  $j$  *given* that host  $i$  is infected. Note that an open directed edge from  $i$  to  $j$  does not imply an open directed edge from  $j$  to  $i$ . The relationship between the contact rate of hosts  $i$  and  $j$  and the conditional probability of transmission from  $i$  to  $j$  is,

$$p_{ij} = 1 - e^{-vk_{ij}\tau_i} \quad (4)$$

where  $v$  is the probability of actual transmission of the pathogen given contact occurs and where we have introduced  $\tau_i$  to represent the length of time that host  $i$  is infectious. The  $\tau_i$  are assumed to be identically and independently distributed. The product  $vk_{ij}\tau_i$  is the density of a Poisson process, representing the average number of *infectious* contacts between  $i$  and  $j$  that occur over a time period of length  $\tau_i$ . The equation for  $p_{ij}$  is the probability that there is at least 1 such infectious contact. In this paper we consider only the case of a fixed infectious period ( $\tau_i = \tau$ ) so that the individual heterogeneity in the model arises

only from variation in the local spatial arrangement of hosts and from variation in the sociability values (the  $x_i$ ) of the hosts.

We simulate outbreaks by “seeding” the network with a single infectious node. In the simulations our spatial networks are finite and we therefore choose a central node - chosen randomly but within a relatively small distance from the network centre - to begin the epidemic. For each simulation we then draw open edges, starting from the initially infected node and using the probabilities given by equation (4). We explore the open cluster that the initial node belongs to by following all open edges and labeling the end nodes as infectious. For each infectious node we again apply equation (4). When all open edges are explored and there are no more infectious nodes then the extent of the epidemic is known and the set of nodes that have been infected defines the open cluster to which the initially infected node belongs. The size of the cluster corresponds to both the severity of the epidemic and its spatial extent. Alternatively, simulations come to an end when the conditions (see below) that define an outbreak have been satisfied.

#### 4 OUTBREAK DEFINITION AND PERCOLATION THRESHOLDS

In disease systems where hosts are randomly mixing, predicting whether a pathogen will spread or fade out equates to estimating the quantity  $R_0$ , which is defined as the expected number of secondary cases from a single primary case in a fully susceptible population. If this number is greater than 1, then the number of infected is expected to grow, but if it is less than 1 then an infected, on average, will not replace itself and there will be no outbreak.

In the context of spatial spread the percolation threshold, denoted  $p_c$ , has a similar threshold effect to  $R_0 = 1$ . Given an infinite spatial network and a regular arrangement of nodes where  $p_{ij} = p$  for neighbouring nodes and 0 otherwise, then the percolation threshold has the property that for  $p > p_c$  there is a positive probability that an arbitrary node belongs to an infinite open cluster. For values of  $p$  less than  $p_c$  the probability that an arbitrary node belongs to an infinite open cluster is 0 (for a thorough treatment of the topic see Grimmett [1999]). For long-range percolation – where open edges between nodes that are not nearest neighbours are allowed – then, referring to equation (4) we can, for example, define  $v_c$  as the critical probability of transmission, below which an infinite open cluster does not occur.

In either contexts, when  $R_0 > 1$  or when  $p > p_c$  the appearance of a single infection does not imply there will be an outbreak, or even that an outbreak is the most likely outcome. The only implication is that it is *possible* for a single infection to then lead to an infinite number of cases (if there were an infinite number of hosts).

For spatial spread of a pathogen, it is natural to define successful spread in terms of spatial criteria such as the distance between newly infected hosts and the initially infected host reaching certain milestones, rather than the number of infected hosts per se. We choose this approach here, though it introduces a level of arbitrariness. Without specifying a particular scale, we set  $A$  to be 100,000 square units and the node density to be 1.5 nodes per unit area (implying a spatial network of 150,000 hosts). We also set  $\tau_i = \tau = 2$  and  $E[x_i] = 1/3$  where there is also a level of arbitrariness except in the sense that together with  $\lambda = 0.75-1.75$  and a coefficient of variation for the  $x_i$  between 0 and (approximately) 3, they placed the model into a region of parameter space that gave interesting behaviour.

Figure 1 shows the results of simulations where  $v$  was varied from 0 to 1 in steps of 0.01 for 6 combinations of the scale parameter  $\lambda$  (0.75, 1.25 or 1.75) and the co-efficient of variation for the  $x_i$  (0, 1.7 or 3). The proportions of the simulations which resulted in new infections 37, 74, 111 and 148 units from the centre of the landscape were recorded after 500 simulations ( $v$  is fixed for the 500 simulations). To save computing time, if for a particular value of  $v$  the proportion of simulations for new infections 37, 74, 111 and 148 units away from the centre were the same (this means that every time a new infection occurred more than 37 units away then a new infection also appeared more than 148 units away) then simulations for higher values of  $v$  were not done. Visually, referring to Figures 1b-g, this value of  $v$  is where the four coloured curves converge. The results show that for some parameter combinations the 4 curves converge very quickly, and for others there is considerably more stochasticity in the model behaviour, i.e. the curves do not converge quickly or do not converge at all. Results for higher values of  $\lambda$  (stronger spatial constraints on contact rates) suggest that the effects of increasing individual heterogeneity are considerably dampened.

Note that we have not attempted to determine  $v_c$ . Inspection of the sets of curves in Figure 1b–g would suggest that a  $v_c$  could be estimated from (some of) them. It is worth noting that a simple estimate for  $R_0$  on a network is  $v\tau\bar{k}$ , where  $\bar{k}$  is the mean contact rate. This would imply the same threshold value ( $v_c = 1/(\tau\bar{k})$ ) for *all* of the scenarios in Figure 1. This is because  $\tau$  is fixed and  $\bar{k}$  does not change with either  $\lambda$  or  $\text{Var}[x_i]$ .

## 5 CONCLUSIONS

Emerging infectious disease events are increasing at a global scale (Jones et al. [2008]) with the majority (71.8%) of them originating in wildlife. In Australia, conservation and public health issues have recently arisen from novel pathogens such as Tasmanian Devil Facial Tumour (Hamede et al. [2009]) and Hendra virus in bats of the genus *Pteropus* (Halpin et al. [2000]). When faced with novel pathogens epidemiologists are very often concerned with  $R_0$  as this quantity can determine (i) the effort required to control a pathogen in a particular host population, and (ii) the range of host populations and circumstances for which the pathogen is expected to invade. However, the dynamics of pathogens in animal host populations that are territorial or simply much less mobile than humans are poorly predicted by  $R_0$ . The model presented here is a step towards (i) encouraging field workers to measure the presence of strong spatial constraints on contact rates, and (ii) understanding better how to interpret such spatial information for the management and control of infectious disease.

## REFERENCES

- Bohm, M., M. R. Hutchings, and P. C. L. White (2009). Contact networks in a wildlife-livestock host community: Identifying high-risk individuals in the transmission of bovine tb among badgers and cattle. *PLoS ONE* 4, e5016.
- Caldarelli, G., A. Capocci, P. De Los Rios, and M. A. Muoz (2002). Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters* 89, 258702(4).
- Christakis, N. A. and J. H. Fowler (2010). Social network sensors for early detection of contagious outbreaks. *PLoS ONE* 5, e12948.
- Craft, M. E., E. Volz, C. Packer, and L. A. Meyers (2009). Distinguishing epidemic waves from disease spillover in a wildlife population. *Proceedings of the Royal Society B* 276, 1777–1785.
- Davis, S., P. Trapman, H. Leirs, and J. A. P. Heesterbeek (2008). The abundance threshold for plague as a critical percolation phenomenon. *Nature* 454, 634–637.
- Diggle, P. (2003). *Statistical analysis of spatial point patterns*. New York: Oxford University Press Inc.
- Grimmett, G. (1999). *Percolation*. Berlin: Springer-Verlag.
- Halpin, K., P. L. Young, H. E. Field, and J. S. Mackenzie (2000). Isolation of hendra virus from pteropid bats: a natural reservoir of hendra virus. *Journal of General Virology* 81, 1927–1932.
- Hamede, R. K., J. Bashford, H. McCallum, and M. Jones (2009). Contact networks in a wild tasmanian devil (*Sarcophilus harrisii*) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters* 12, 1147–1157.
- Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–994.
- Liljeros, F., C. R. Edling, L. A. Nunes Amaral, H. E. Stanley, and Y. Åberg (2001). The web of human sexual contacts. *Nature* 411, 907–908.
- May, R. M. (2006). Network structure and the biology of populations. *TRENDS in Ecology and Evolution* 21, 394–399.

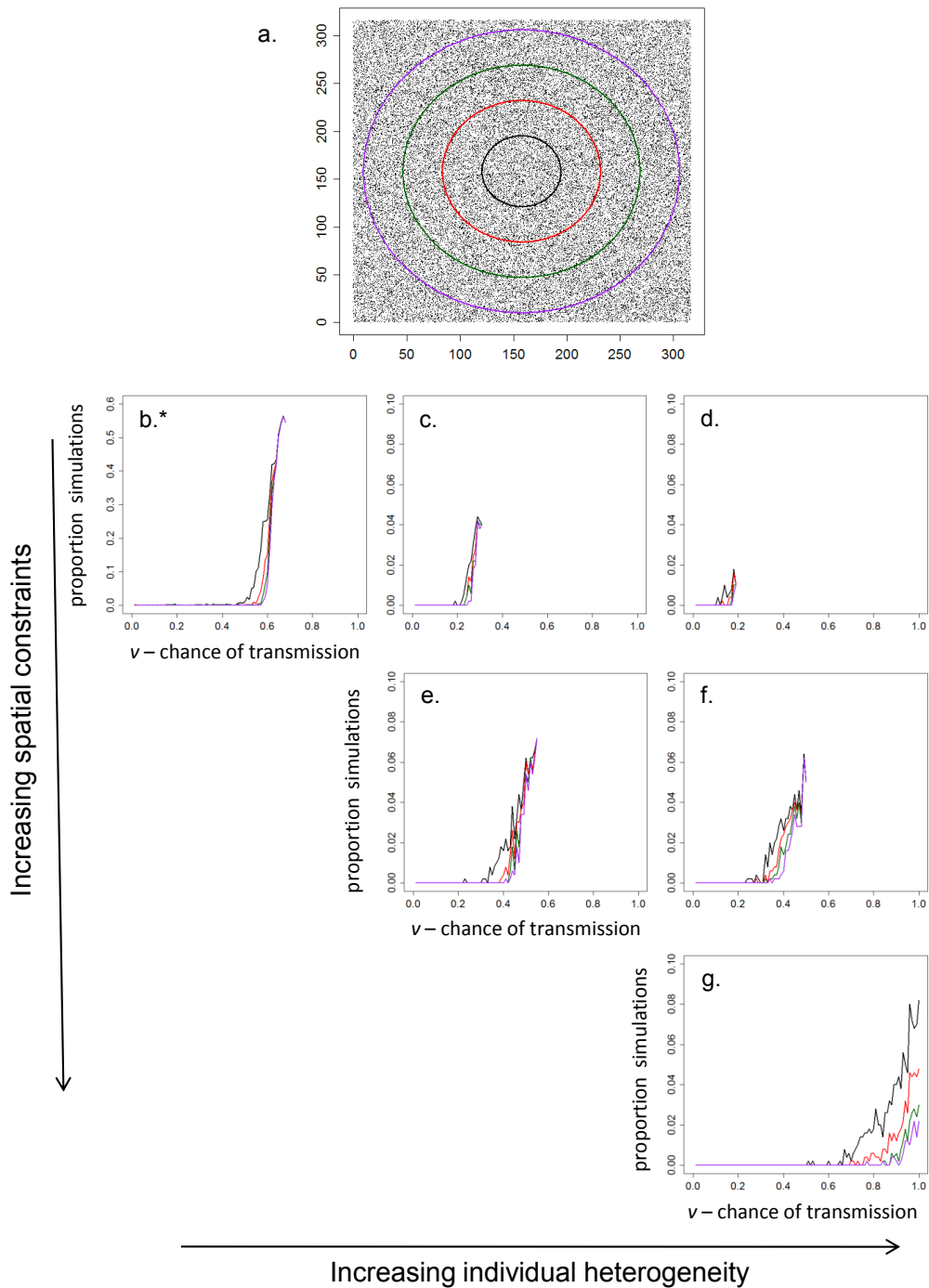


Figure 1: Long-range percolation results for the spatial good-get-richer model for 9 combinations of the scale parameter  $\lambda$  and the co-efficient of variation for the  $x_i$ . The landscape (a) is 100,000 square units and the density of nodes is 1.5. The coloured circles represent the milestones used to track the size of the outbreak (the open cluster to which the initially infected node belongs) and respectively represent the infection traveling 37, 74, 111 and 148 units from the centre of the landscape. The proportions of simulations for which the infection passes each milestone is given by the curve of the corresponding colour in **b-g**. The combinations of  $\lambda$  and the co-efficient of variation for the  $x_i$ , listed in alphabetical order and corresponding to **b** through **g**, were: (0.75, 0), (0.75, 1.7), (0.75, 3.0), (1.25, 1.7), (1.25, 3.0) and (1.75, 3.0). \*Note the different scale for the vertical axis.

Perkins, S. E., F. Cagnacci, A. Stradiotto, D. Arnoldi, and P. J. Hudson (2009). Comparison of social networks derived from ecological data: implications for inferring infectious disease dynamics. *Journal of Animal Ecology* 78, 1015–1022.

Porphyre, T., M. Stevenson, R. Jackson, and J. McKenzie (2008). Influence of contact heterogeneity on tb reproduction ratio  $R_0$  in a free-living brushtail possum *Trichosurus vulpecula* population. *Veterinary Research* 39:31.

## 6 APPENDIX

We show here that  $E[k_i] = E[\sum_j k_{ij}]$  is independent of  $\lambda$  when the  $k_{ij}$  are given by (3). Making this substitution we have,

$$\begin{aligned} E \left[ \sum_j \lambda^2 e^{-\lambda s_{ij}} x_i x_j \right] &= E[x_i] E[x_j] E \left[ \sum_j \lambda^2 e^{-\lambda s_{ij}} \right] \\ &= \bar{x}^2 E \left[ \sum_j \lambda^2 e^{-\lambda s_{ij}} \right] \end{aligned}$$

where we have introduced  $E[x_i] = \bar{x}$  and used the independence of the  $x_i$ , the independence of the positions of the nodes, and the independence of the  $x_i$  from the positions of the nodes. We now consider this quantity when  $A$  is infinite. Let  $H(s)$  be the average number of nodes that are at distance  $s_{ij}$  from an arbitrary node in  $A$  (where the average is taken over all nodes in  $A$ ). Here we may write,

$$\bar{x}^2 E \left[ \sum_j \lambda^2 e^{-\lambda s_{ij}} \right] = \bar{x}^2 \int_0^\infty \lambda^2 e^{-\lambda s} H(s) ds.$$

The positions of the nodes in  $A$  are each an independent random sample from the uniform distribution on  $A$ . This is equivalent to the number of nodes falling in an area  $|A'|$  following a Poisson distribution with intensity  $c|A'|$  with  $c$  a constant (see Diggle [2003]). Consider then the expected number of nodes in the area between a circle of radius  $s$  and a circle of radius  $s + \delta s$ . This is

$$\begin{aligned} E[\text{number of nodes}] &= c (\pi(s + \delta s)^2 - \pi s^2) \\ &= c (2\pi s \delta s + \pi(\delta s)^2) \sim 2\pi c s \text{ as } \delta s \rightarrow 0 \end{aligned}$$

With  $H(s) = 2\pi c s$  then;

$$\begin{aligned} \int_0^\infty \lambda^2 e^{-\lambda s} H(s) ds &= 2\pi c \lambda^2 \int_0^\infty s e^{-\lambda s} ds \\ &= 2\pi c \lambda^2 \left[ \frac{e^{-\lambda s}}{\lambda^2} (-\lambda s - 1) \right]_0^\infty \\ &= 2\pi c \lambda^2 \frac{1}{\lambda^2} = 2\pi c \end{aligned}$$

This leaves  $E[k_i] = E[\sum_j k_{ij}] \approx 2\pi c \bar{x}^2$  (it is an approximate relationship for a finite region  $A$ ), which is independent of  $\lambda$ .