Complex-Network Modelling and Inference

Lecture 12: Random Graphs: spatially-embedded and small-world networks

Matthew Roughan
<matthew.roughan@adelaide.edu.au>
https://roughan.info/notes/Network_Modelling/

School of Mathematical Sciences,
University of Adelaide

August 18, 2021

Section 1

Spatially-Embedded Random Graphs

Space and networks

Many networks have nodes embedded in space

- Most physical networks
  - *e.g.*, electricity
  - *e.g.*, Internet
  - *e.g.*, air-plane routes
- Many biological networks
  - *e.g.*, animal interaction networks
  - *e.g.*, epidemiological contact graphs
  - *e.g.*, neural networks
- Many social networks
  - most people have a locus around which they spend most time

Space and networks

In many settings a longer link is more expensive in some sense

- *e.g.*, Longer electricity wires or telephone cables are more costly to build
- *e.g.*, Contact between animals requires them to move over larger distances, and hence expend more energy.
- *e.g.*, Neural networks – sending signals over longer distances can be slower, costing speed.
- *e.g.*, Wireless connections require more power over longer distances, and will therefore create more interference.

The natural consequence is that longer links will be less likely in spatially embedded networks.

Waxman graphs

Early examples

- [Wax88] Take $n = |N|$ nodes, distributed uniformly at random on a square (or potentially some other space)
- [ZCB96] connect nodes i and j randomly with probability

$$p = \alpha f(d_{ij}).$$

where d_{ij} is the Euclidean distance between them.

  - example 1

$$p = \alpha\, e^{-\beta d_{ij}}.$$

  - example 2

$$p = \alpha\, u(d_{ij} - r),$$

where $u(x)$ is the unit step function (at zero).

- key point is that links still chosen independently *conditional* on the locations of the nodes

Spatially Embedded Random Graphs (SERNs)

1. Place n nodes *randomly* within a defined region R of a metric space Ω
   1. randomly usually means uniformly, *i.e.*, the probability of a node occurring in any sub-region r is proportional to the area of r.
   2. a metric space means it has a distance metric $d(x, y)$
2. Probability of an edge (i, j) is

$$p_{ij} = q\, f_\theta(s\, d_{i,j}),$$

where

$$\begin{aligned} f(\cdot) &= \text{a distance deterrence function,} \\ q &= \text{a thinning parameter,} \\ s &= \text{a scale parameter,} \\ \theta &= \text{other parameters.} \end{aligned}$$

Don't panic! Most cases use standard Euclidean spaces, and simple deterrence functions.

e.g., Waxman Random Graphs

Points are chosen in the unit square, and

$$p(d_{i,j}) = q\, e^{-s d_{i,j}}.$$

Notes

- I use a different parameterisation from the literature
  - to be consistent with other models
  - because the literature gets it wrong about half the time
- The idea has been reinvented multiple times, but Waxman was the first as far as I know.

e.g., Random Geometric Graphs

Points are chosen in the unit square, and

$$p(d_{i,j}) = q\, H(1 - s d_{i,j}),$$

where $H(\cdot)$ is the Heavyside step function.

Notes

- Also called
  - random plane network [Gil61]
  - random connection models
  - random distance graphs

Other common examples

- GER: $p_{i,j} = q$, where $q \in (0, 1]$ [Gil59, ER60];
- Clipped Waxman: $p_{i,j} = \min\left(q\, e^{-sd_{i,j}}, 1\right)$, where $s \in [0, \infty), q \in (0, \infty)$;
- Mixed Waxman-threshold: $p_{i,j} = q\, e^{-sd_{i,j}} H(r - sd_{i,j})$, where $s \in [0, \infty), q \in (0, 1], r \in [0, \infty)$;
- Power law: $p_{i,j} = q\,(1 + s\, d_{i,j})^{-\theta_2}$, (*e.g., range-dependent random graphs*) [Gri02, Far02, GD07, HM03], and
- Cauchy: $p_{i,j} = q\,(1 + (sd_{i,j})^2)^{-1}$ [Avi08].
- Exponential: $p_{i,j} = \dfrac{q\, e^{-d_{i,j}}}{L - d_{i,j}}$, [ZCD97];

Common features

- Distance deterrence function $f(\cdot)$ is non-increasing
  - it doesn't have to be, but all cases I know have this feature, and it relates to the motivation
- Usually in a finite region, but don't have to be, and often we are interested in asymptotic limits
- Underlying point process of node locations is a n-D Poisson Process
  - doesn't have to be, but not much work where it isn't
  - produces the uniformly at random result

Sidebar on the line picking problem

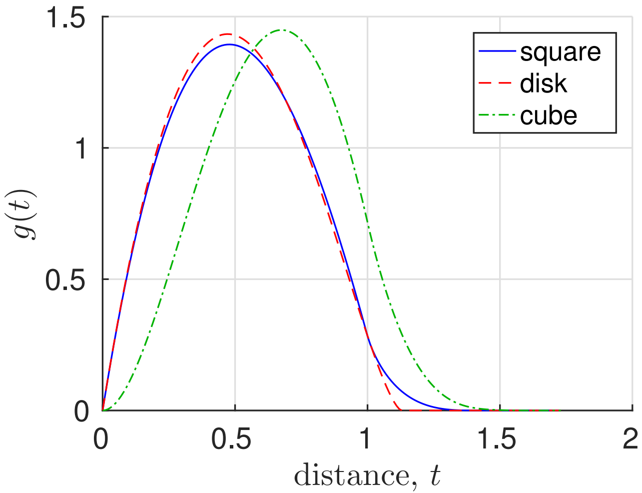
The *line-picking problem* is the problem of finding the probability distribution of the length of a random line in some region. More precisely, choose two points at random in the region, what is the distribution of distances between them.

- Lots of results are known
  - *e.g.*, square line picking, the probability density is

$$g(t) = \begin{cases} 2t(t^2 - 4t + \pi) \ \text{ for } 0 \leq t \leq 1, \\ 2t\left[4\sqrt{t^2-1} - (t^2+2-\pi) - 4\tan^{-1}\left(\sqrt{t^2-1}\right)\right] \\ \qquad \text{for } 1 \leq t \leq \sqrt{2}. \end{cases}$$

- It's not hard to calculate numerically even when it isn't known
- Major determining factors:
  - size of region
  - dimension of space
  - "aspect ratio" of region

Sidebar on the line picking problem



Waxman in detail

Points chosen in $R \subset \Omega$, and

$$p(d_{i,j}) = q\, e^{-sd_{i,j}}.$$

Probability of an arbitrary link (prior to knowing the distances):

$$Prob\{(i,j) \in \mathcal{E} \mid q, s\} = q \int_0^\infty \exp(-st) g(t)\, dt = q\tilde{G}(s),$$

where $\tilde{G}(s)$ is the Laplace transform of $g(t)$ (from line-picking)
When

- $s = 0$, the Laplace trans. becomes the normalisation constraint so $p(d_{i,j}) = q = Prob\{(i,j) \in \mathcal{E} \mid q, s\}$

Waxman in detail

Average number of edges

$$\mathbb{E}[|E|] = \frac{n(n-1)}{2} Prob\{(i,j) \in \mathcal{E} \mid q, s\} = \frac{n(n-1)q\tilde{G}(s)}{2}$$

From the handshake theorem, the average node degree is

$$\bar{k} = (n-1)q\tilde{G}(s)$$

Waxman in detail

Distribution of the lengths of edges $f(d\,|q,s)$ can be derived

$$
\begin{aligned}
f(d \mid q,s) &= \text{Prob}\left\{d_{ij} = d \mid (i,j) \in \mathcal{E}\right\} \\
&= \frac{\text{Prob}\left\{d_{ij} = d \;\&\; (i,j) \in \mathcal{E}\right\}}{\text{Prob}\left\{(i,j) \in \mathcal{E}\right\}} \\
&= \frac{\text{Prob}\left\{(i,j) \in \mathcal{E} \mid d_{(i,j)} = d; q, s\right\} \text{Prob}\left\{d_{ij} = d\right\}}{\text{Prob}\left\{(i,j) \in \mathcal{E} \mid q, s\right\}} \\
&= \frac{q \exp(-sd) g(d)}{\int_{t=0}^{\infty} q \exp(-st) g(t)\, dt} \\
&= \frac{g(d) \exp(-sd)}{\tilde{G}}
\end{aligned}
$$

Waxman in detail

Average distance of edges

$$
\begin{aligned}
\mathbb{E}[d \mid s] &= \frac{1}{\tilde{G}(s)} \int_0^\infty t g(t) \exp(-st) \, dt \\
&= -\frac{\tilde{G}'(s)}{\tilde{G}(s)}.
\end{aligned}
$$

Parameter estimation

The above result suggests a moment-based estimator

- We match the means to get s, *e.g.*, , measure

$$\hat{d} = \frac{1}{|E|} \sum_{e} d_{i,j},$$

  and find $\hat{s}$ such that

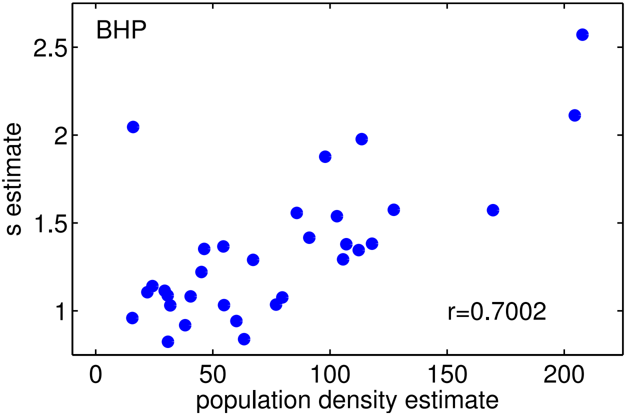
$$\frac{\tilde{G}'(s)}{\tilde{G}(s)} = -\hat{d}$$

- Estimate q exactly how we did for Gilbert-Erdős-Rényi
  - convert it into estimate of Binomial parameter
- It turns out this is also the MLE (Maximum Likelihood Estimator)
- $|N|$, $|E|$ and $\bar{d}$ are sufficient statistics

Example: voles [DAS+14]

- Large marked-capture-recapture experiment with voles
  - English-Scottish border over a 7-year period to study
  - Field voles *(Microtus agrestis)*
- Trap locations in a grid
  - traps emptied multiple times
  - contact between voles presumed if they were caught in same trap (on separate occasions)
  - generate multiple graphs at 4 different sites, and different time periods
- Fitted several models
  - their favoured model was similar to Waxman, but a little more complex
  - I have done a Waxman fit
  - most interesting is relationship between population density and $\hat{s}$

Example: voles [DAS+14]



- As population density increases, voles travel shorter distances
- Has important consequences for disease transmission

Section 2

Small-world graphs

Six degrees of Kevin Bacon

Here's a game

- Pick an actor (any actor will do)
- Determine the shortest path to Kevin Bacon on the graph of movies in actors collaborate
- Famous result is that these paths rarely have more than 6 hops (6 degrees of separation).
  - *e.g.*, Nicole Kidman → Jeff Perry → Kevin Bacon
  - actually there are lots of cases for this particular connection

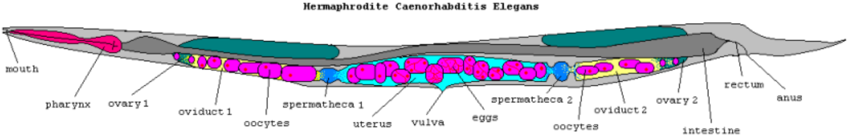
http://oracleofbacon.org/cgi-bin/movielinks

Experiment 1: letter forwarding

Stanley Milgram [Mil67] experiment

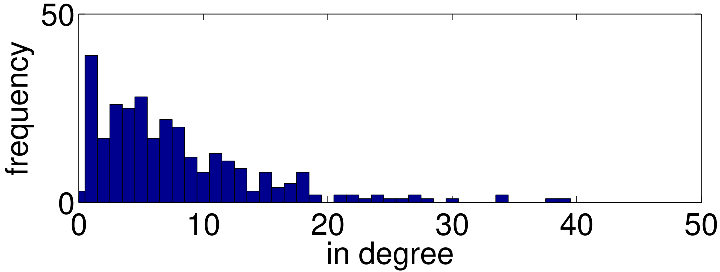
- gave people envelopes and the name of a target stranger
- mission: get envelope to destination
- method forward to someone they know (on 1st name basis)
- interesting result was how short the path-lengths were
  - the average was 6
  - 6-degrees of separation
- lots of problems with the study, but also lots of interest

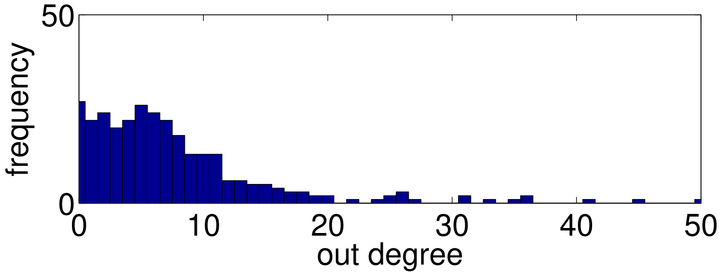
Experiment 2: *Caenorhabditis elegans*



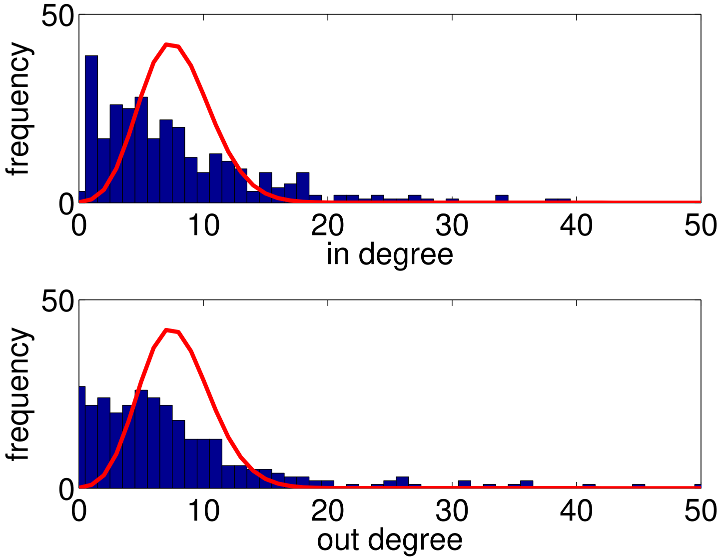
- *C. elegans* [SH77, SSWT83] is a small (∼1mm) soil nematode (worm)
- its very simple (only 959 somatic cells)
- its neural network was mapped in the 70's and 80's [AT76, WSNB86]
  - 302 neurons
  - database available
    http://ims.dse.ibaraki.ac.jp/ccep/

Experiment 2: *C. elegans*

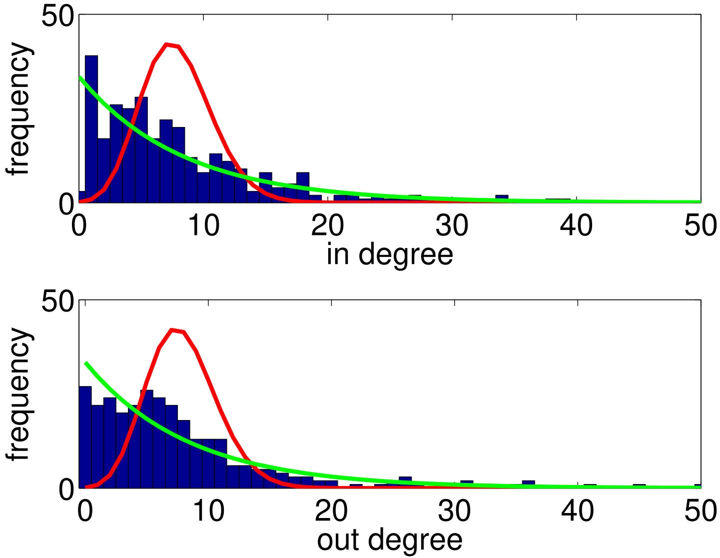




Experiment 2: *C. elegans*



Experiment 2: *C. elegans*



Experiment 2: *C. elegans*

- mean path length 2.45
  - reported as 2.65 in [WS98]
  - much smaller than the size (nodes $= 302$) of the network
  - equivalent ER random network with same n and e mean distance is 2.34 (reported as 2.25 in [WS98])
- mean (local) clustering coefficient 0.31
  - reported as 0.28 in [WS98]
  - equivalent ER random network with same n and e mean distance is 0.055 (reported as 0.05 in [WS98])
- notably:
  - distances are similar
  - clustering is completely different

Small world networks

- Watts and Strogatz [WS98] noted that many networks have two properties
  - short path distances
  - high clustering
- They proposed a model
  - start with a highly regular network with lots of clustering
  - rewire some links at random

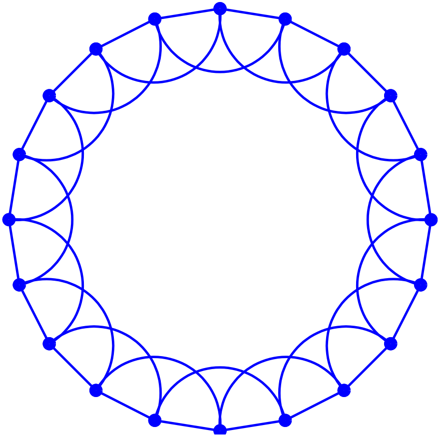
Small world networks

Formally:

- start with a regular ring network
  - n nodes on a circle, linked to k nearest neighbours
  - k even for it to be regular
- rewire each link with probability p
  - take an existing link, and send it to a random node

Regular ring network

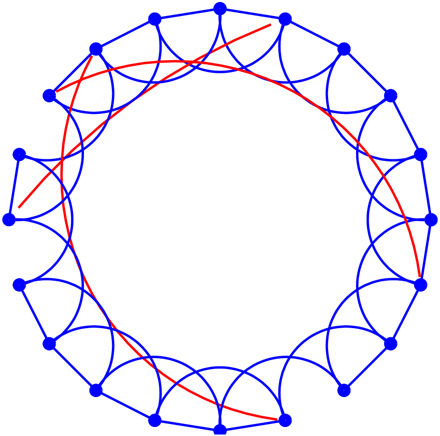
$n = 20$ and $k = 4$



$p = 0$

Regular ring network

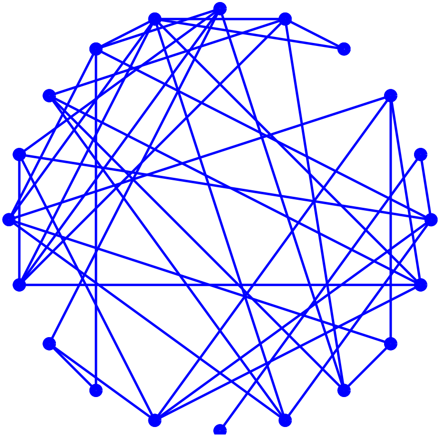
$n = 20$ and $k = 4$



$p = 0 \ \rightarrow$ increasing randomness

Regular ring network

$n = 20$ and $k = 4$



$p = 0 \ \rightarrow$ increasing randomness $\rightarrow p = 1$

Small world network clustering

Theorem

The clustering coefficient for the k-regular ring (and $n > k$) is

$$C = 1 - \frac{e(e-1)}{k(k-1)},$$

where k is the number of neighbours (k even) and

$$e = k/2 + 1.$$

In the limit for large k is $C \to 0.75$

Small world network clustering

Proof.

We can derive the clustering coefficient for the regular ring as follows:

- WLOG start from node 1.
- Each node has k neighbours, so the clustering coefficient will be

$$C = \frac{n_e}{k(k-1)/2}$$

  where n_e is the number of links between the neighbours.
- It is easier to consider the number of missing links, so that

$$C = 1 - \frac{n_e^C}{k(k-1)/2}$$

  where n_e^C is the number of missing links between neighbours.
- List the neighbours in consecutive order around the ring from "left" to "right".

Small world network clustering

continued.

- The first neighbour will connect to $k/2$ nodes (including node 1, which we ignore because it isn't a "neighbour"), *i.e.,* $k/2$ of the nodes. So it misses $k/2$ links.
- The second neighbour will connect to $k/2$ nodes to its right (including node 1), and one to the left, so it misses $k/2 - 1$ links.
- Continue until we get to the $k/2$ node (immediately to the left of node 1), and this will miss 1 link.
- Repeat the same argument from the other side of the original node, but remember to divide the total by 2 because links are undirected.
- The total number of missed links will be

$$
\begin{aligned}
n_e^C &= 2\left[(k/2) + (k/2 - 1) + \cdots + 1\right]/2 \\
&= (k/2 + 1)(k/2)/2 \\
&= e(e-1)/2.
\end{aligned}
$$

Small world network clustering

Make clustering a function of p, the probability of rewiring

- ER random graph $C(1) \simeq 0$
- Regular ring (in limit) has $C(0) = 0.75$
- Small-world network $p \in (0, 1)$ interpolates between these

Small world networks path length

Distances in regular ring (n, k):

- WLOG take an arbitrary start node, say 1
- The furtherest node away on the ring is distance $n/2$
- We can reach this node in steps of $k/2$
- So distance to this node is $\lceil n/k \rceil$ hops
- But we want an average over all nodes. They are equi-spaced around the ring, so where n is divisible by k a crude average can be obtained by dividing by two.

$$E[L] = \frac{n}{2k}$$

- However, we should take into account the fact that the last hop, won't be length $k/2$ for many of the intermediate nodes. So the path lengths will be slightly longer. The extra distance can be seen in the first hop, where dividing by $k/2$ would tell you less than one hop, but there is exactly one. It can be approximated by $(k/2 - 1)/(k/2)$.

Small world networks path length

Make L a function of p, the probability of rewiring

- ER random graph distances are fairly short (assuming connectivity)

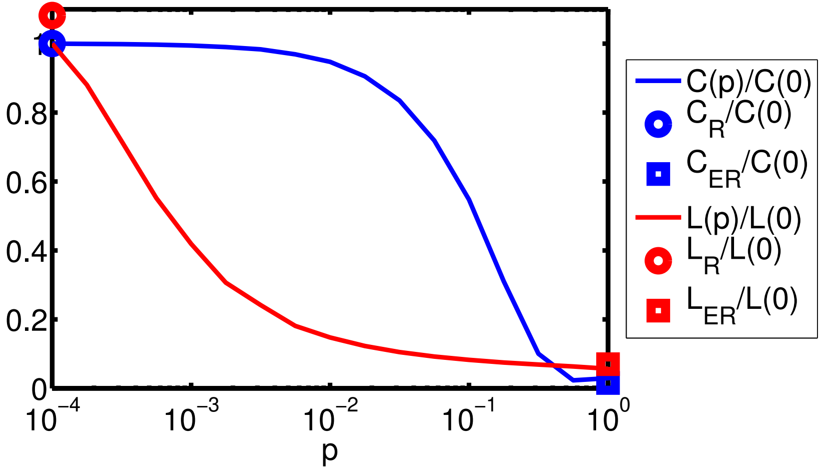
$$L(1) \simeq \log n / \log k$$

- Regular ring

$$L(0) \simeq n/2k + (k/2 - 1)/(k/2)$$

  where k is the number of neighbours (k even)
- Small-world network interpolates between these

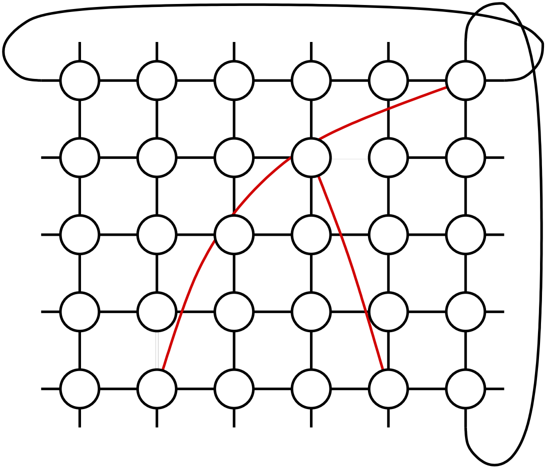
Small world networks features



Generalizations

Small world on a lattice

- a ring is 1D
- what about starting with a lattice on a torus



Parameter estimation

- I haven't seen any formal literature on estimation
- Could hack up something that matches $\bar{L}$ and $\bar{C}$
- Do I think it's worth it?

Further reading I

D. G. Albertson and J. N. Thomson, *The pharynx of caenorhabditis elegans*, Phil. Trans. R. Soc. London B **275** (1976), 299–325.

Chen Avin, *Distance graphs: from random geometric graphs to Bernoulli graphs and between*, Fifth International Workshop on Foundations of Mobile Computing (New York, NY, USA), DIALM-POMC '08, ACM, 2008, pp. 71–78.

Stephen Davis, Babak Abbasi, Shrupa Shah, Sandra Telfer, and Mike Begon, *Spatial analyses of wildlife contact networks*, Journal of The Royal Society Interface **12** (2014), no. 102, http://rsif.royalsocietypublishing.org/content/12/102/20141004.short.

P. Erdős and A. Rényi, *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960), 17–61.

Andras Farago, *Scalable analysis and design of ad hoc networks via random graph theory*, Dial-M'02, 2002.

Further reading II

Cigdem Gunduz-Demir, *Mathematical modeling of the malignancy of cancer using graph evolution*, Mathematical Biosciences **209** (2007), no. 2, 514–527.

E.N. Gilbert, *Random graphs*, Annals of Mathematical Statistics **30** (1959), 1441–1144.

E. N. Gilbert, *Random plane networks*, Journal of the Society for Industrial and Applied Mathematics **9** (1961), no. 4, 533–543.

Peter Grindrod, *Range-dependent random graphs and their application to modeling large small-world proteome datasets*, Phys. Rev. E **66** (2002), no. 066702.

R. Hekmat and P. Van Mieghem, *Degree distribution and hopcount in wireless ad-hoc networks*, Networks, 2003. ICON2003, 2003, pp. 603–609.

Further reading III

- Stanley Milgram, *The small world problem*, Psychology Today **1** (1967), no. 1, 60–67.
- J. E. Sulston and H. R. Horvitz, *Post-embryonic cell lineages of the nematode Caenorhabditis elegans*, Developmental Biology **56** (1977), 110–156.
- J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson, *The embryonic cell lineage of the nematode Caenorhabditis elegans*, Developmental Biology **100** (1983), 64–119.
- B.M. Waxman, *Routing of multipoint connections*, IEEE J. Select. Areas Commun. **6** (1988), no. 9, 1617–1622.
- Duncan J. Watts and Steven H. Strogatz, *Collective dynamics of 'small-world' networks*, Letters to Nature **393** (1998), no. 4, 440–442.

Further reading IV

J. G. White, E. Southgate, Thomson J. N., and S. Brenner, *The structure of the nervous system of the nematode Caenorhabditis elegans*, Phil. Trans. R. Soc. London B **314** (1986), 1–340.

Ellen W. Zegura, Ken Calvert, and S. Bhattacharjee, *How to model an internetwork*, IEEE Infocom (San Francisco, CA), 1996, pp. 594–602.

Ellen W. Zegura, Kenneth L. Calvert, and Michael J. Donahoo, *A quantitative comparison of graph-based models for Internet topology*, IEEE/ACM Transactions on Networking **5** (1997), no. 6, 770–783.