Spatial Networks

- "Sideways" talk – theorems exist but are peripheral to the most interesting conceptual issues.

- Lots of distantly related work, but not much closely related.

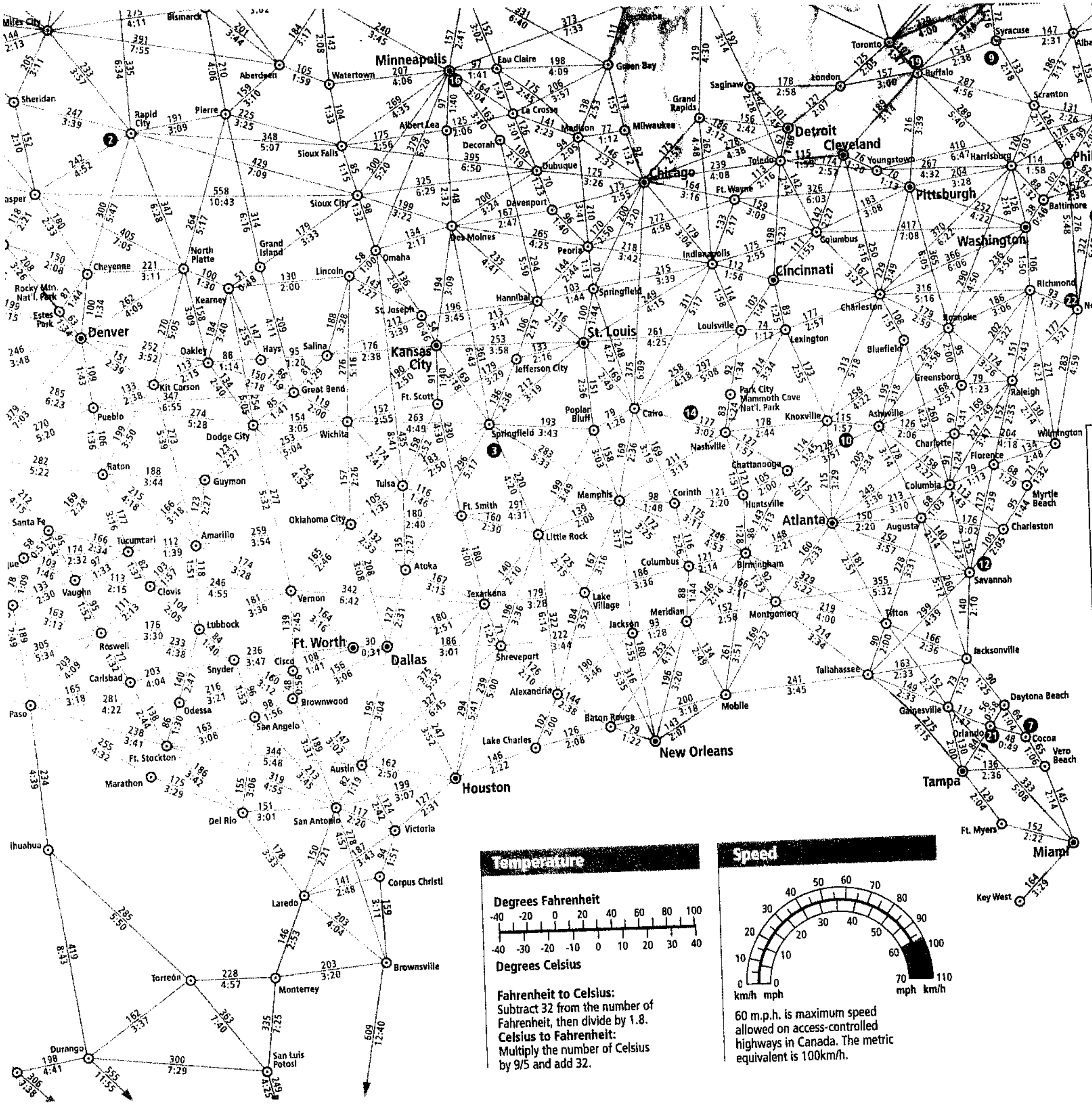
- Scope for undergraduate research projects.

- I'll spotlight one topic, worthy of more study: proximity networks on random points.

(R)OUTES

driving times map

This map shows both the distance and the approximate driving time between hundreds of cities across North America. You can also get mileages and driving times at go.randmcnally.com/MC by typing in any two cities or addresses.



Mileages

277 Black numerals indicate mileage in statute miles.

Drive Times

7:55 Blue numerals indicate driving time.

Driving time shown is approximate under normal conditions. Consideration has been given to topography, number of towns along the route, congested urban areas, and the speed limit imposed by each state.

Allowances should be made for night driving and unusually fast or slow drivers.

Points of Interest

1. Banff Nat'l. Park, AB
2. Black Hills, SD
3. Branson, MO
4. Butchart Gardens, BC
5. Cape Cod, MA
6. Cape Hatteras, NC
7. Kennedy Space Center, FL
8. Disneyland, CA
9. Finger Lakes, NY
10. Great Smoky Mts. Nat'l. Park, TN
11. Hearst San Simeon, CA
12. Hilton Head Island, SC
13. Lake Tahoe, CA/NV
14. Land Between The Lakes, KY/TN
15. Laurentian Mountains, QC
16. Mall of America, MN
17. Monterey Peninsula, CA
18. Mystic Seaport, CT
19. Niagara Falls, NY
20. Rehoboth Beach, DE
21. Walt Disney World, FL
22. Williamsburg, VA

Temperature

Degrees Fahrenheit

-40 -20 0 20 40 60 80 100

-40 -30 -20 -10 0 10 20 30 40

Degrees Celsius

Fahrenheit to Celsius: Subtract 32 from the number of Fahrenheit, then divide by 1.8.
Celsius to Fahrenheit: Multiply the number of Celsius by 9/5 and add 32.

Speed

km/h 0 10 20 30 40 50 60 70 80 90 100 110

mph 0 10 20 30 40 50 60 70

60 m.p.h. is maximum speed allowed on access-controlled highways in Canada. The metric equivalent is 100km/h.

Consider the problem of designing a road network to link n given cities. Assume that some main goal of the network is to provide short routes. Want to set up a toy model of "benefits" and "costs".

- "cost" = total network length.

- "benefit" = some statistic measuring route-lengths between cities.

Conceptual starting point for this research project: To what extent are real-world "evolved" networks approximately optimal (i.e. maximize benefit for given cost)?

At first sight, "just an empirical question", but

- What's a good choice for the "benefit" statistic? (Goldilocks!)

- Algorithm for finding optimal network?

- Theoretical properties of optimal networks on random points?

- There are many possible uses for models for connected networks on random points. What are some mathematically **tractable** models?

For two cities i, j in a given network, consider

$$r(i,j) = \frac{\text{route-length } i \text{ to } j}{\text{straight line distance } i \text{ to } j} - 1$$

so that "$r(i,j) = 0.2$" means that route-length is 20% longer than straight line distance. With n cities we get $\binom{n}{2}$ such numbers $r(i,j)$; what is a reasonable way to combine these into a single number R to be used as a descriptive statistic? Two natural possibilities are

$$\begin{aligned} R_{\text{max}} &:= \max_{j \neq i} \; r(i,j) \\ R_{\text{ave}} &:= \text{ave}_{(i,j)} \; r(i,j) \end{aligned}$$

where $\text{ave}_{(i,j)}$ denotes average over all distinct pairs (i,j).

I will discuss these in a minute.

Need a notion of “normalized total network length” for comparing networks with different spatial scales and different numbers n of cities.

Convention; Choose unit of length so that ave density of cities equals 1 per unit area.

Define $L =$ (total network length)$/n$.

This is the right scaling because nearest n'bor cities will be at distance $O(1)$.

We can get an intuitive feeling for L by looking at familiar networks on regularly-spaced vertices.

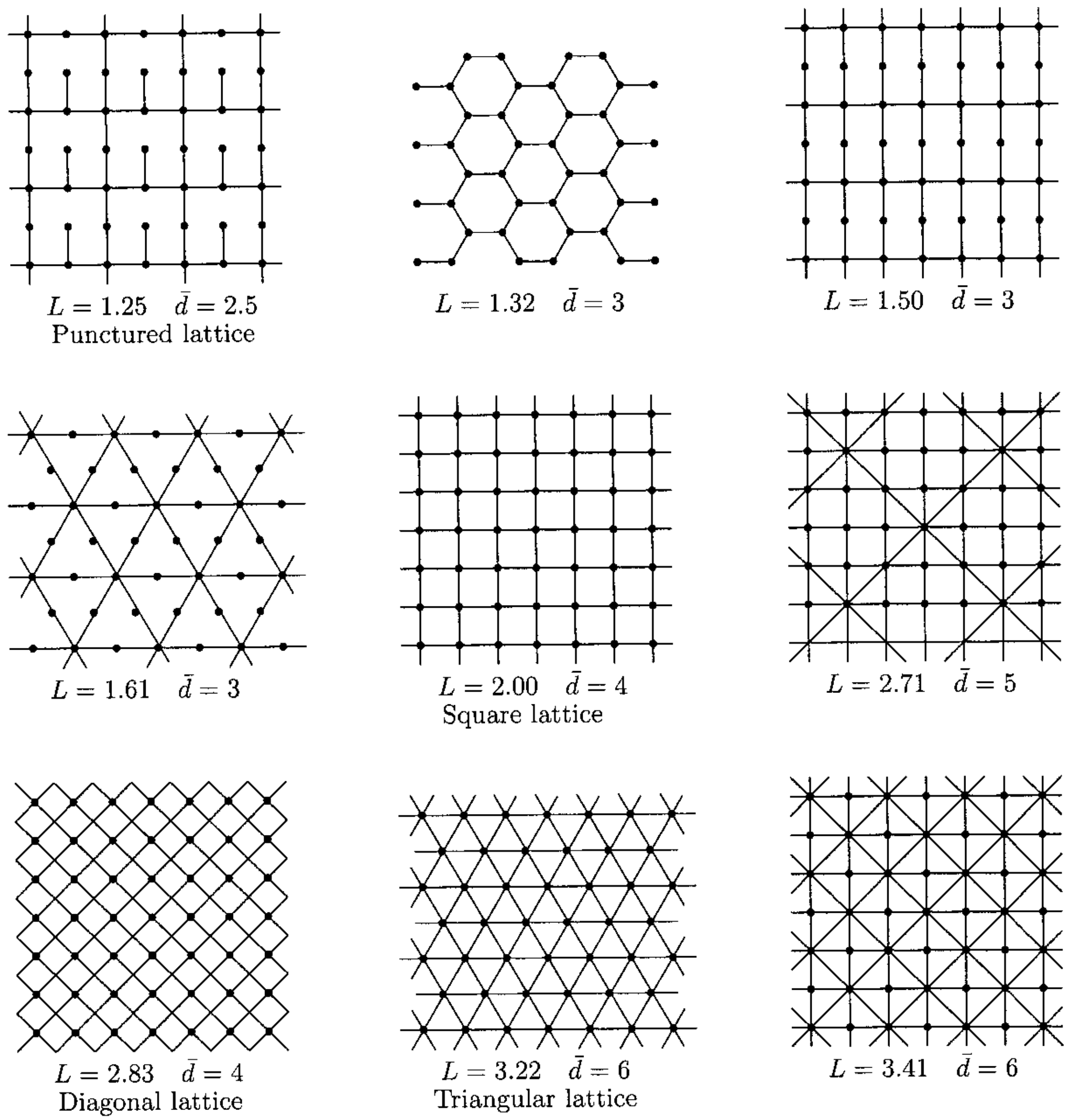


Figure 4. Variant square, triangular and hexagonal lattices. Drawn so that the density of cities is the same in each diagram, and ordered by value of L.

The statistic R_{max} has been studied in the context of the design of **geometric spanner networks** (Narasimhan-Smid 2007) where it is called the *stretch*. However, being an "extremal" statistic R_{max} seems unsatisfactory as a descriptor of real world networks – for instance, it seems unreasonable to characterize the U.K. rail network as inefficient simply because there is no very direct route between Oxford and Cambridge.

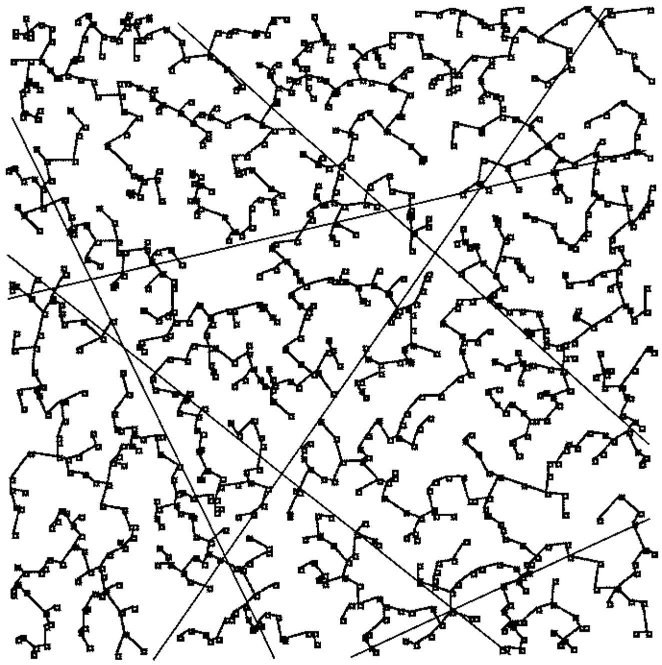


Figure 2. Efficient or inefficient? R_{ave} would judge this network efficient in the $n \to \infty$ limit. [1]

We propose a statistic R which is intermediate between R_{ave} and $R_{\max}$.

$$\rho(d) := \text{ mean value of } r(i,j) \text{ over city-pairs with } d(i,j) = d$$

$$R := \max_{0 \le d < \infty} \rho(d). \tag{3}$$

In words, $R = 0.2$ means that on every scale of distance, route-lengths are on average at most 20% longer than straight line distance.

On an intuitive level, R provides a sensible and intepretable way to compare efficiency of different networks in providing short routes. On a technical level, we see two advantages and one disadvantage of using R instead of R_{ave}.

Advantage 1. Using R to measure efficiency, there is a meaningful $n \to \infty$ limit for the network length/efficiency tradeoff (the function $\Xi(\theta)$ discussed

[1]The tree is actually a MST, taken from the home page of Leland Wilkinson `http://www.spss.com/research/wilkinson`, because we could not find a suitable Steiner tree picture. Similarly, Table 1 quotes the MST rather than the Steiner tree because we could not find data for normalized length for the Steiner tree.

The statistic R_{ave} has a more subtle drawback. Consider a network consisting of
(a) the minimum-length connected network (Steiner tree) on given cities;
and (b) a superimposed sparse collection of randomly oriented lines (*Poisson line process*). By choosing the density of lines to be sufficiently low, one can make the normalized network length be arbitrarily close to the minimum needed for connectivity. But [Aldous - Kendall 2008] for this network $R_{\text{ave}} \to 0$ as $n \to \infty$.

Of course no-one would build a road network looking like this to link cities, because there are many pairs of nearby cities with only very indirect routes between them. The disadvantage of R_{ave} is that (for large n) most city-pairs are far apart, so the fact that a given network has a small value of R_{ave} says nothing about route-lengths between nearby cities.

delete No

A peripheral theorem.

Notation set-up: Configuration $\mathbf{x}^n$ of n cities in square of area n, so side-length $\sqrt{n}$.
len($ST(\mathbf{x}_n)$) = length of Steiner tree.
$d(i,j)$ = distance and $\ell(i,j)$ = route-length in some network $\mathcal{G}(\mathbf{x}_n)$, whose total length is len($\mathcal{G}(\mathbf{x}_n)$).

Assertion last slide: we can construct $\mathcal{G}(\mathbf{x}^n)$ so that

$$(i)\ \text{len}(\mathcal{G}(\mathbf{x}_n)) - \text{len}(ST(\mathbf{x}_n)) = o(n)$$

$$(ii)\ \text{ave}_{i,j}(\ell(i,j) - d(i,j)) = o(n^{1/2}).$$

Theorem 1 (with Wilf Kendall) *True, and can improve (ii) to*

$$(ii)\ \text{ave}_{i,j}(\ell(i,j) - d(i,j)) = o(\omega_n \log n)$$

for $\omega_n \to \infty$ arbitrarily slowly.

Conceptual point: R_{ave} not helpful, at least in $n \to \infty$ setting.

We propose a statistic R which is intermediate between R_{ave} and R_{max}.

$$\rho(d) := \text{ mean value of } r(i,j) \text{ over}$$
$$\text{city-pairs with } d(i,j) = d$$

$$R := \max_{0 \le d < \infty} \rho(d).$$

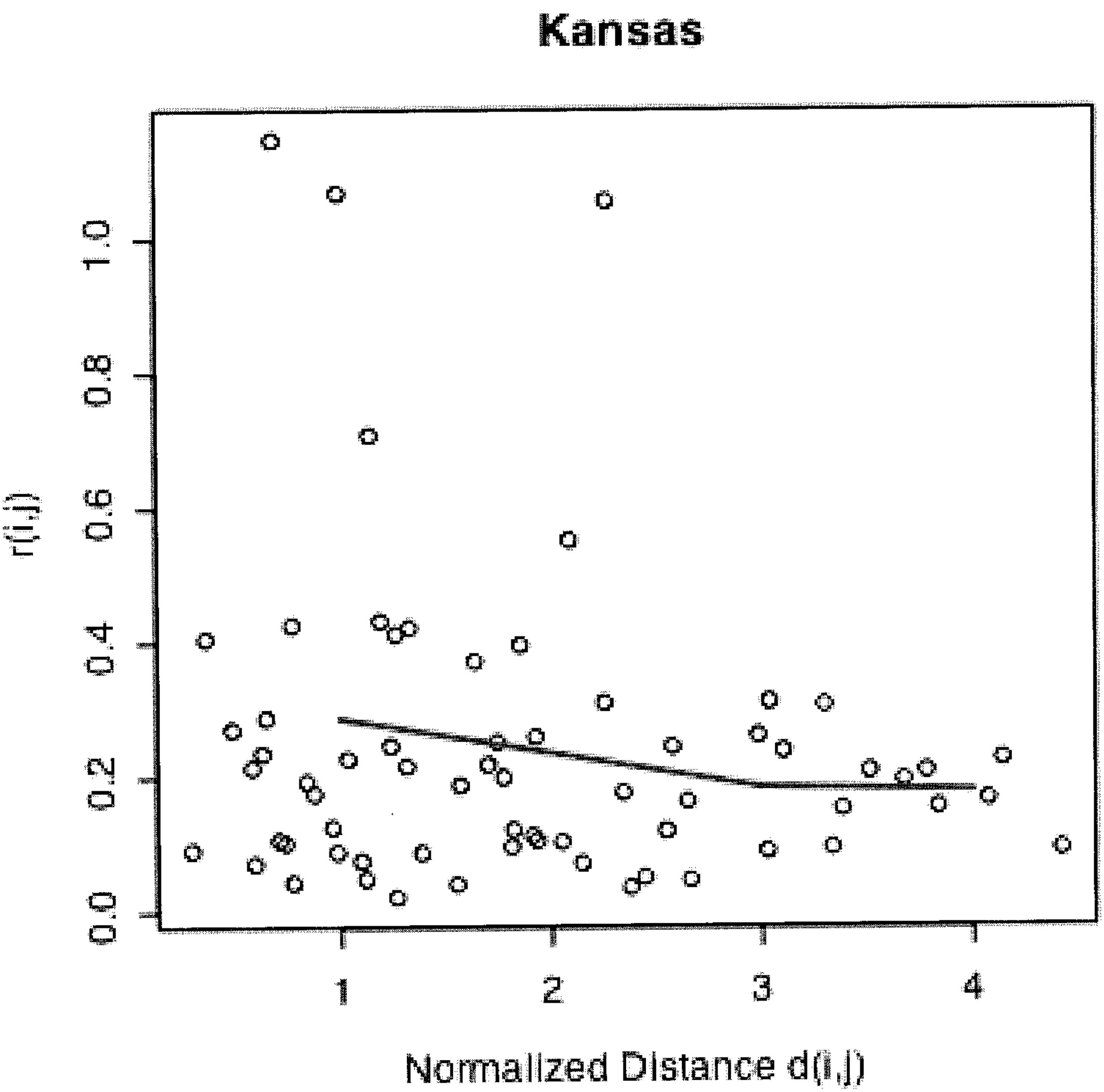
In words, $R = 0.2$ means that on every scale of distance, route-lengths are on average at most 20% longer than straight line distance.

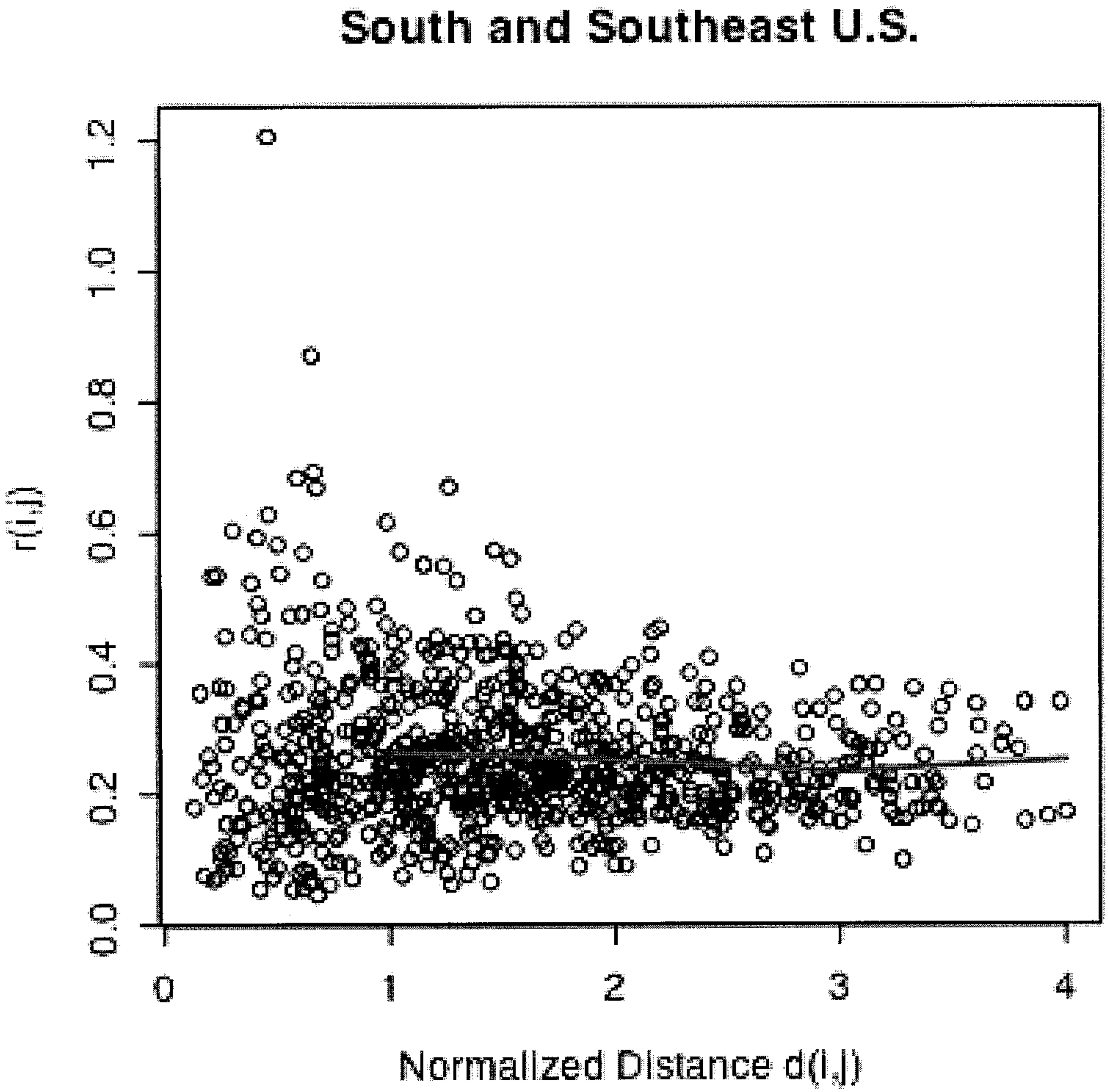
Discussion points:

(a) Permits meaningful $n \to \infty$ limit for the network length/efficiency tradeoff (in either random model for city positions, or worst-case), and so in particular it makes sense to compare the values of R for networks with different n.

(b) Finesses modelling issue that traffic volume depends on distance.

(c) Need to discretize for finite n.


Kansas
r(i,j)
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
Normalized Distance d(i,j)


South and Southeast U.S.
r(i,j)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
2
3
4
Normalized Distance d(i,j)

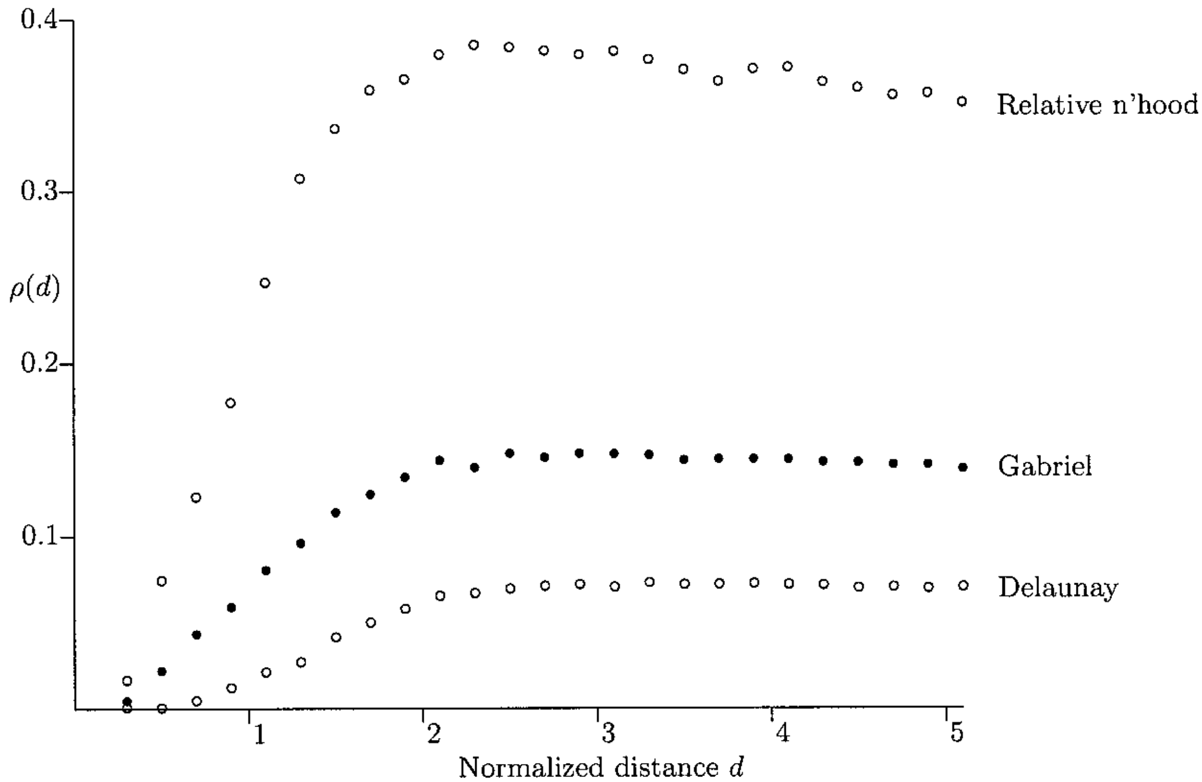


Figure 6. The function $\rho(d)$ for three theoretical networks on random cities. Irregularities are Monte Carlo random variation.

This characteristic shape is also visible in real-world data [7]. It is easy to interpret. Any efficient network will tend to place roads directly between unusually close city-pairs, implying that $\rho(d)$ should be small for $d < 1$. For large d the presence of multiple alternate routes helps prevent $\rho(d)$ from growing.. At distance $2 - 3$ from a typical city x_0 there are about $\pi 3^2 - \pi 2^2 \approx 16$ other cities x_i. For some of these x_i there will be cities x' near the straight line from x_0 to x_i, so the network designer can create roads from x_0 to x' to x_i. The difficulty arises where there is no such intermediate city x': including a direct road (x_0, x_i) will increase L, but not including it will increase $\rho(2.5)$.

Tractible models for connected networks

Of the many ways to construct random networks in two-dimensional space, the following two are perhaps best known.

(i) [*Small world* models]. Start with the usual square grid of vertices and edges; add extra edges (i, j) with probability $p(i, j)$ for prescribed $p(\cdot)$:.

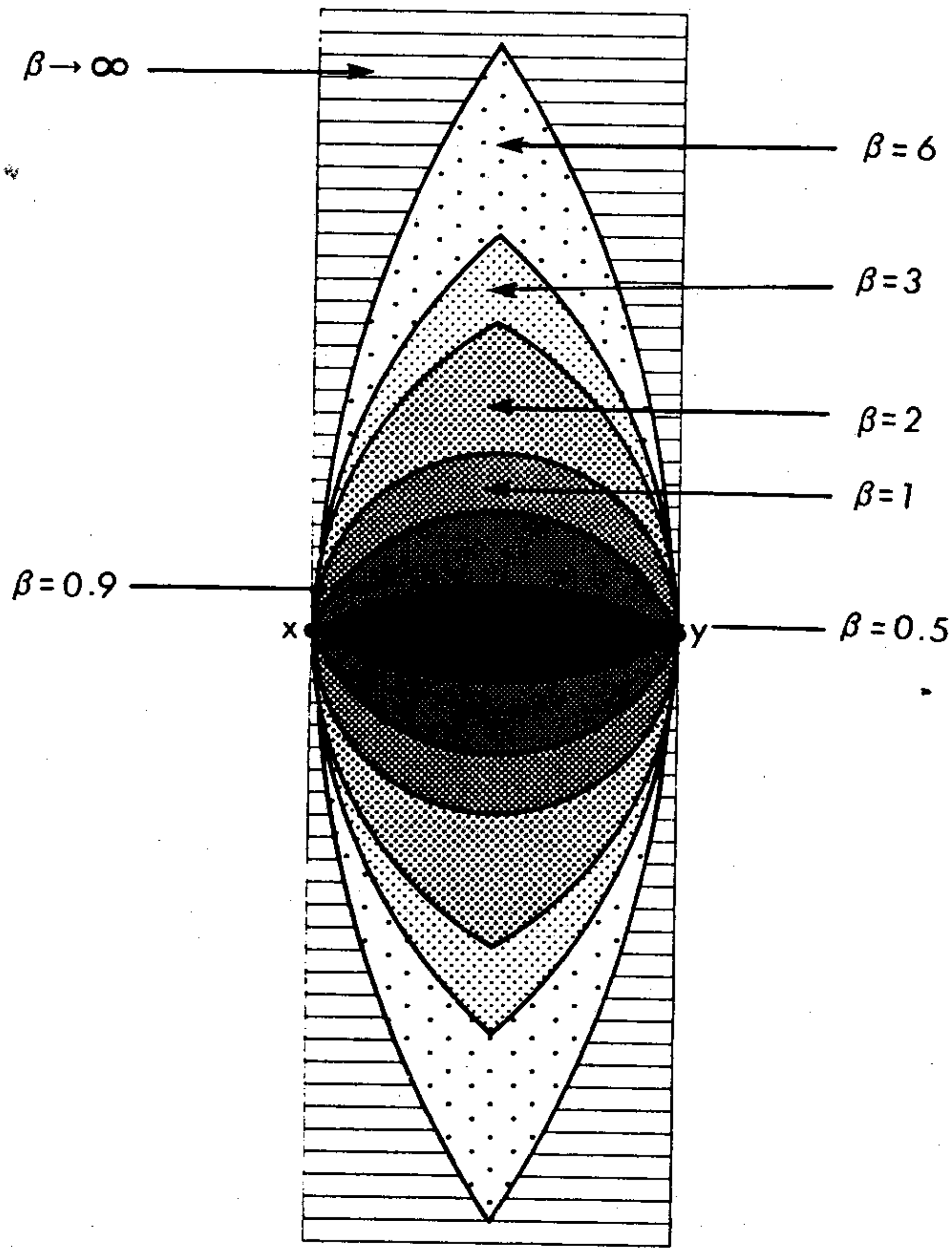
(ii) [*Random geometric graph*]. Start with a Poisson point process of vertices; put an edge between vertices i, j if the Euclidean distance $d(i, j)$ is less than some prescribed constant c.

Note that in model (ii), for any configuration of vertices there is a determinstic rule for defining edges; one then applies this rule to random points. We study a different model with the same "deterministic" feature. The point of the model is that it is both **connected** and has **finite normalized length**.

Spotlight topic : Proximity graphs

Write v_- and v_+ for the points $(-\frac{1}{2}, 0)$ and $(\frac{1}{2}, 0)$. The **lune** is the intersection of the open discs of radii 1 centered at v_- and v_+. So v_- and v_+ are not in the lune but are on its boundary. Define a **template** A to be a subset of $\mathbb{R}^2$ such that
(i) A is a subset of the lune;
(ii) A contains the line segment (v_-, v_+);
(iii) A is invariant under reflection (left - right and top - bottom)
(iv) A is open.

For arbitrary points x, y in $\mathbb{R}^2$, define $A(x, y)$ to be the image of A under the transformation (translation, rotation and scaling) that takes (v_-, v_+) to (x, y).

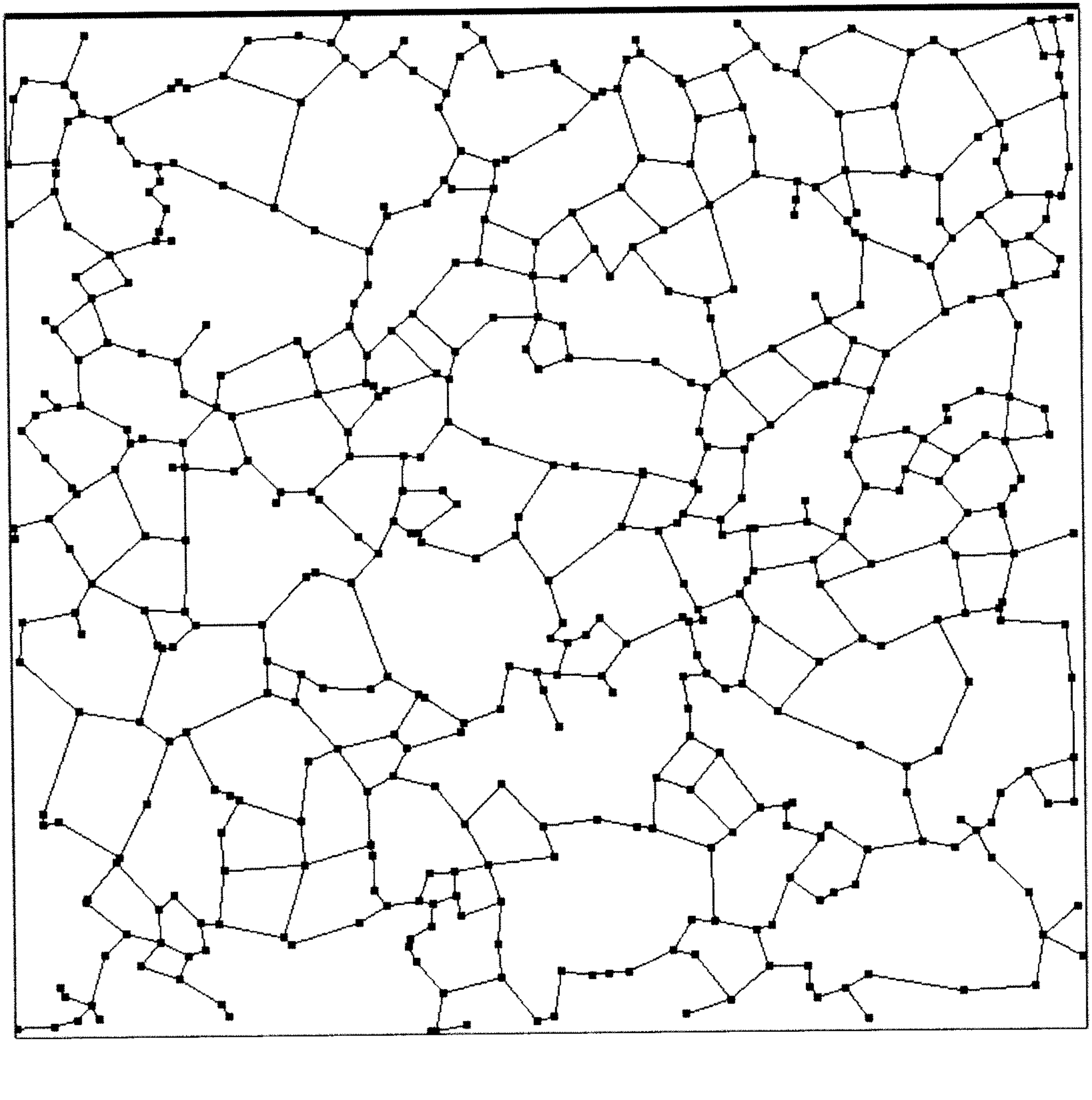

$\beta \to \infty$
$\beta = 6$
$\beta = 3$
$\beta = 2$
$\beta = 1$
$\beta = 0.9$
x
y
$\beta = 0.5$

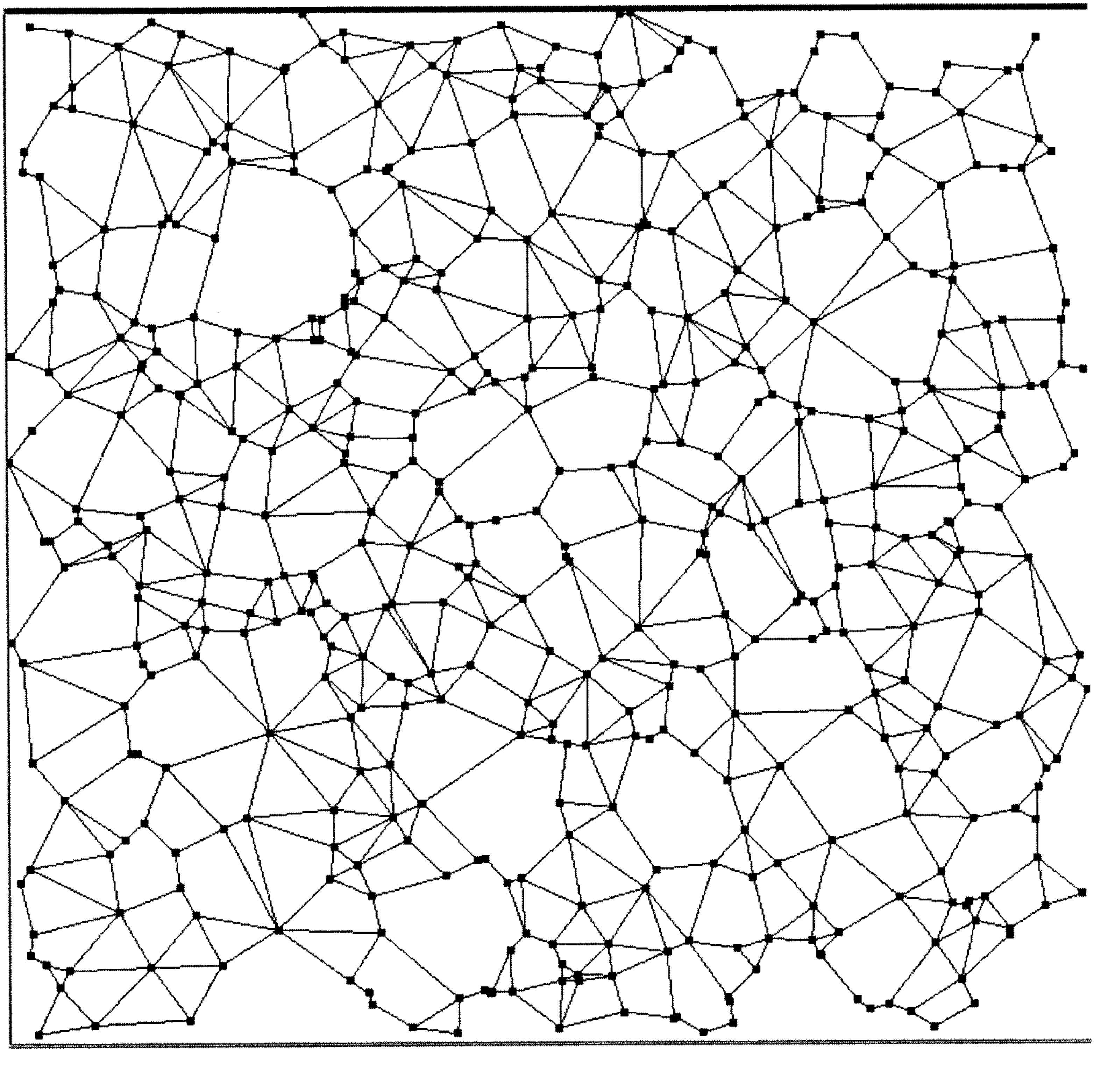
Definition. Given a template A and a locally finite set $\mathbf{x}$ of vertices, the associated **proximity graph** G has edges defined by: for each $x, y \in \mathbf{x}$,

$$(x, y) \text{ is an edge of } G \text{ iff}$$
$$A(x, y) \text{ contains no vertex of } \mathcal{V}.$$

Note that replacing A by a subset A' can only increase the edge-set. It is easy to check that, if A is the lune, then G contains the minimum spanning tree, and it follows that (for finite $\mathcal{V}$) the proximity graph is always connected.

There are two "named" special cases. If A is the lune then G is called the **relative neighborhood graph**. If A is the disc centered at the origin with radius 1/2 then G is called the **Gabriel graph**.





Discussion. This general idea has been studied since the 1980s in computational geometry and pattern recognition, focusing on specific examples such as the relative neighborhood graph, the Gabriel graph and an interpolating family called *beta-skeletons*. We are imagining proximity graphs as toy models for road networks (an idea already noted in e.g. Kirkpatrick and Radke (1985) but not investigated very thoroughly). That is, whether or not there is a direct road from city i to a nearby city j depends (partly) on whether some other city k is between i and j.

Finally, note that the Delaunay triangulation does not fit the "proximity graph" definition (though it has a related description).

Let's consider $\mathcal{G}_A$, the proximity graph associated with a Poisson point process of rate 1 on $\mathbb{R}^2$. To indicate that $\mathcal{G}_A$ is at least somewhat tractable, note

Lemma 2 *Write* $a =$ area(A). *Then for* $\mathcal{G}_A$

$$\begin{aligned} \textit{mean edge-length per unit area} &= \frac{\pi^{3/2}}{4a^{3/2}} \\ \textit{mean vertex degree} &= \frac{\pi}{a}. \end{aligned}$$

One could continue along the lines of Lemma 2 to write down complicated integral expressions for (e.g.) the mean number of triangles per unit area in $\mathcal{G}_A$. In contrast to random geometric graphs [see Penrose monograph] there seems only one known non-elementary result about $\mathcal{G}_A$ – the model deserves more study.

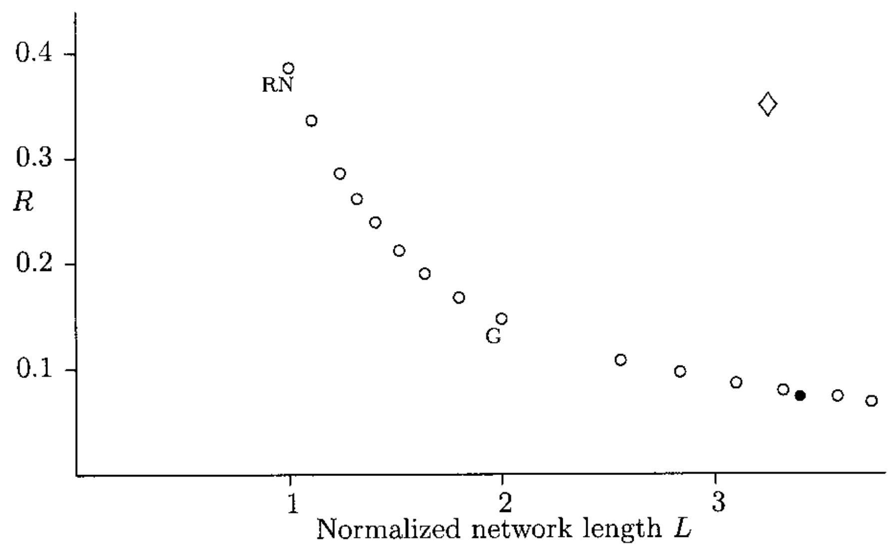


Figure 7. The normalized network length L and the route length efficiency statistic R for certain networks on random points. The $\circ$ show the beta-skeleton family, with RN the relative neighborhood graph and G the Gabriel graph. The $\bullet$ shows the Delaunay triangulation and $\diamond$ shows the Hammersley network.

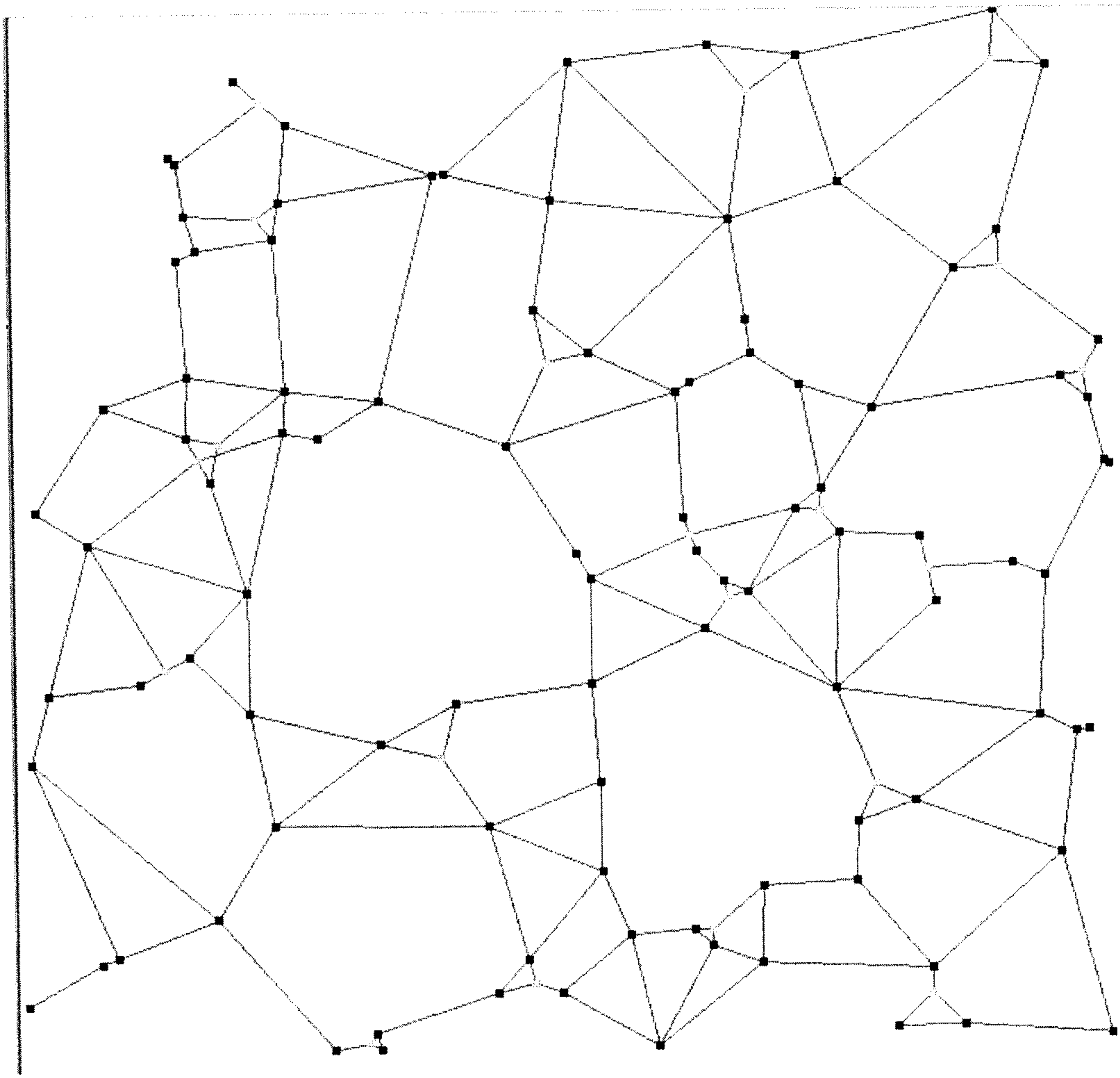
Notes on Table 1. (a) Values of R from our simulations with $n = 2,500$.
(b) Value of L for MST from Monte Carlo [15]. In principle one can calculate arbitrarily close bounds [8] but this has never been carried through. Of course $\bar{d} = 2$ for any tree.
(c) The Gabriel graph and the relative neighborhood graph fit the assumptions of Lemma 1 below with $c = \pi/4$ and $c = \frac{2\pi}{3} - \frac{\sqrt{3}}{4}$ respectively, and their table entries for L and $\bar{d}$ are obtained from Lemma 1, as are the values for β-skeletons in Figure 7.
(d) For the Hammersley network, every degree equals 4, so $L = 2\times$(mean edge-length). It follows from theory [3] that a typical edge, say NE from (x, y), goes to a city at position $(x + \xi_x, y + \xi_y)$, where ξ_x and ξ_y are independent with Exponential(1) distribution. So mean edge-length equals

$$\int_0^\infty \int_0^\infty \sqrt{x^2 + y^2}\; e^{-x-y}\; dxdy \approx 1.62. \tag{7}$$

In the random model of cities, the figure shows the $R - L$ trade-off for the standard family of proximity graphs.

Perhaps the central object in our project is the (unknown) function $R = \Psi(L)$ giving the **optimal** value of R for given network length θ. This means optimal over all possible networks on the random cities.

Intuitively, we expect to be able to construct networks that improve over these proximity graphs. for instance by introducing junctions. But in ongoing simulation projects [by undergraduates] our first three ideas failed to work well . . . . . .



Peripheral theory, in progress.

Heuristics suggest that we can attain the long-network asymptotics

$$R \approx L^{-4/3} \text{ for large } L.$$

Our function $R = \Psi(L)$ is one of four functions, depending on choice of

- random or worst-case city positions
- R or R_{max}.

Seems reasonable to conjecture that each has the same long-network scaling exponent α

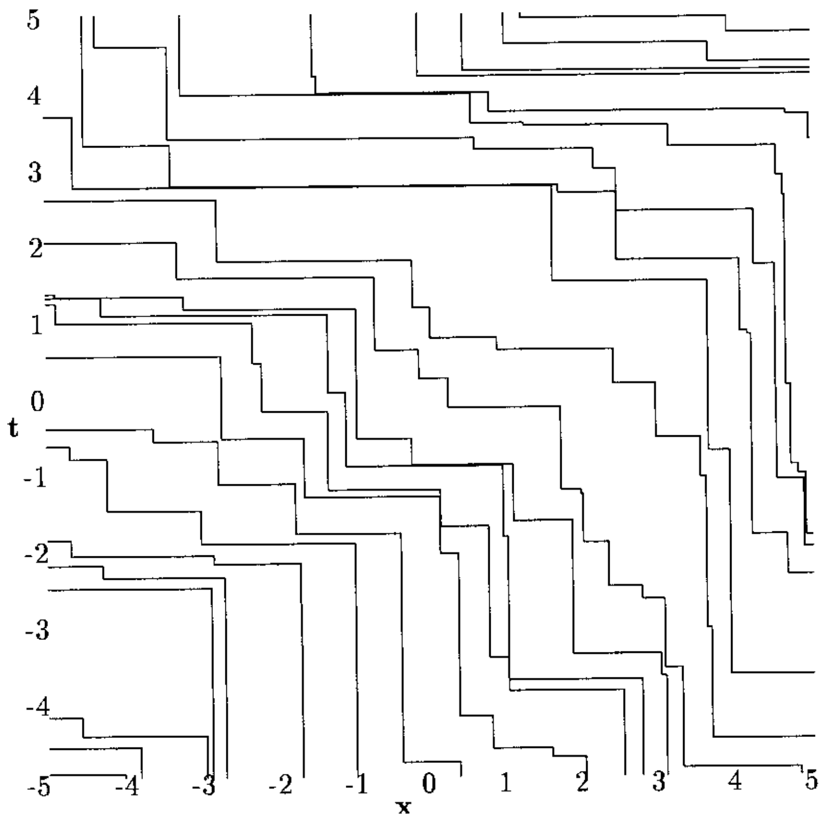
$$\Psi(L) \approx L^{-\alpha} \text{ for large } L.$$

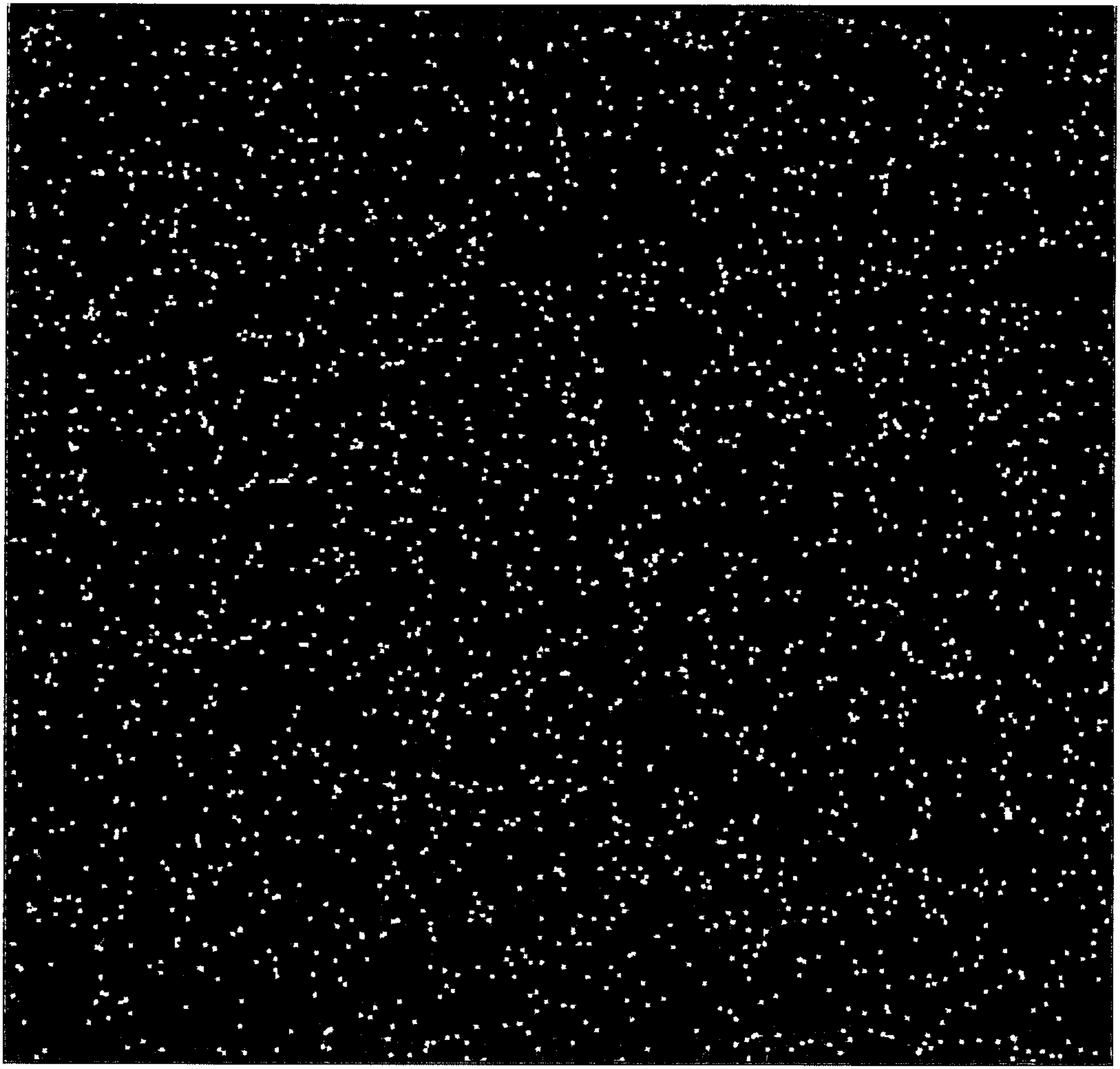
Spotlight topic #2

A (slightly) new idea – networks whose construction depends on the fact that we have a Poisson process of city positions.

For a picturesque description, imagine one-eyed frogs sitting on an infinitely long, thin log, each being able to see only the part of the log to their left before the next frog. At random times and positions (precisely, as a space-time Poisson point process of rate 1) a fly lands on the log, at which instant the (unique) frog which can see it jumps left to the fly's position and eats it. This defines a continuous time Markov process (**Hammersley process**) whose states are the configurations of positions of all the frogs. There is a *stationary* version of the process in which at each time, the positions of the frogs form a Poisson (rate 1) point process.

Now consider the space-time trajectories of all the frogs, drawn with time increasing upwards on the page. For each frog, the part of the trajectory between the ends of two successive jumps consists of an upward edge (the frog remains in place as time increases) followed by a leftward edge (the frog jumps left).


5
4
3
2
1
0
t
-1
-2
-3
-4
-5
-4
-3
-2
-1
0
1
2
3
4
5
x



Reinterpreting the time axis as a second space axis, and introducing compass directions, the part of the trajectory becomes a North edge followed by a West edge. Now replace these two edges by a single North-West straight edge. Doing this procedure for each frog and each pair of successive jumps, we obtain a collection of NW paths; that is, a network in which each city (the reinterpreted space-time random points) has a edge to the NW and an edge to the SE.

Repeat with the same realization of the space-time Poisson point process but with frogs jumping rightwards instead of leftwards. This yields a network on the infinite Poisson point process, which we name the **oriented Hammersley network**. In this network each city has degree 4, with one road going in each direction NW, NE, SE, SW. We can calculate the normalized length of this network as

$$\theta = 3.25 \ldots .$$

Final comments on proximity graphs.

1. Elementary inclusion of edge-sets in the deterministic case:

MST $\subseteq$ relative n'hood $\subseteq$ Gabriel $\subseteq$ Delaunay .

2. In contrast, for certain proximity graphs on random cities, it is known that

$$R_{\text{max}}(n) \sim \sqrt{\frac{\log n}{\log \log n}} . \qquad (1)$$

This contrasts with the Delaunay triangulation for which $R_{\text{max}} \leq 1.42$. Fact (1), from Bose-Devroye-Evans-Kirkpatrick (2006), is perhaps the **only** known non-elementary result about random proximity graphs.

3. Speculative applications of random proximity graphs.

Random proximity graphs seem an interesting object of study from many viewpoints, in particular as an attractive alternative to random geometric graphs for modeling spatial networks that are connected by design. As well as being natural models for road networks, they might be useful in modeling communication networks suffering line of sight interference.

At a more mathematical level, for questions such as spread-out percolation or critical value of contact processes, random proximity graphs with small A are an interesting alternative to the usual lattice- or random graph-based models. For instance, it is natural to conjecture that the critical value p^*_A for a random proximity graph with template A satisfies

$$p^*_A \sim \pi^{-1} \text{area}(A) \text{ as } \text{area}(A) \to 0 \qquad (2)$$

and that the critical value λ^*_A for the contact process has the same asymptotics.