ONE, TWO AND THREE TIMES $\log n/n$

RICHARD MYCROFT (†)

Suppose that n cities are all connected to one another by roads, and the time needed to traverse each road is a random variable with uniform distribution $\mathcal{U}(0,1)$, independent of the time for any other road. For a given pair of cities, we wish to know the typical time taken to travel between these cities by the fastest route (of course, this may not be by the 'direct' route).

To formalise this problem, consider the complete graph K_n on vertex set $[n] := \{1, 2, \dots, n\}$, where each vertex represents a city. Then assign to each edge ij a weight T_{ij}, where the weights T_{ij} are independent random variables with uniform distribution $\mathcal{U}(0,1)$. So T_{ij} represents the travel time along the road connecting i and j. Then the parameter we are interested in is X_{ij}, which for each i and j denotes the total weight of the path of least weight from i to j (i.e. the shortest total travelling time).

Theorem 1. *As $n \to \infty$:*

(i) *For any fixed i and j,*

$$\frac{X_{ij}}{\log n/n} \xrightarrow{p} 1.$$

(ii) *For any fixed i,*

$$\frac{\max_j X_{ij}}{\log n/n} \xrightarrow{p} 2.$$

(iii) *Also,*

$$\frac{\max_{i,j} X_{ij}}{\log n/n} \xrightarrow{p} 3.$$

So typically the fastest route between two cities has duration around $\log n/n$. But for each city there is a city which takes longer (typically $2\log n/n$) to reach, and it takes even longer (typically $3\log n/n$), to travel between the most distant pairs of cities.

The following simple proposition helps to explain these results.

Proposition 2. *As $n \to \infty$,*

$$\frac{\max_i \min_{j\neq i} X_{ij}}{\log n/n} \xrightarrow{p} 1.$$

This suggests that whilst most cities are fairly well-connected (you can get from there to most other cities in time around $\log n/n$), there are a small number of cities which are poorly connected, from which it takes time roughly $\log n/n$ to get anywhere at all. So to get from a given city i to a poorly-connected city j, it may take time $\log n/n$ to get across the network of well-connected cities, followed by a further $\log n/n$ to reach the destination city, giving the total time of $2\log n/n$ as in Theorem 1(ii). Similarly, travelling between

(†) All contents of this abstract are from the paper 'One, Two and Three Times $\log n/n$' by Svante Janson, University of Uppsala.

two poorly-connected cities may take three steps, each of time $\log n/n$, for a total time of $3\log n/n$ as in Theorem 1(iii).

The following standard coupling result both generalises Theorem 1 and is essential to its proof.

Proposition 3. *Let $\mathcal{D}$ be a non-negative probability distribution with $\mathbb{P}(\mathcal{D} \leq t) = t + o(t)$ as $t \to 0$. Then the validity of Theorem 1 is not affected if the independent random variables T_{ij} instead have distribution $\mathcal{D}$ (rather than being uniformly distributed).*

By Proposition 3, we may assume that the random variables T_{ij} have the exponential distribution Exp(1) (so $\mathbb{P}(T_{ij} \leq t) = (1 - e^{-t}) = t + o(t)$ as $t \to 0$). This will be very useful in proving Theorem 1, due to the memoryless property of the exponential distribution.

Fact 4 (Memoryless propery of the exponential distribution)**.** *Let $X \sim Exp(\lambda)$. Then for any s and t:*

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t).$$

So to prove Theorem 1(i) and (ii), we consider the problem as a first-passage percolation viewpoint. A good way to think of this is that a disease breaks out in city i, and spreads along each edge after a Exp(1)-distributed waiting time (from the moment that one end-vertex is infected). Then X_{ij} is the time until city j is infected, and $\max_j X_{ij}$ is the time until all cities are infected. The fact that the waiting times are exponentially distributed means that the typical values of these times can be calculated without too much difficulty.

Furthermore, the observation that $\mathbb{P}(\max_{i,j} X_{ij} \geq a \log n/n) \leq n\mathbb{P}(\max_j X_{ij} \geq a \log n/n)$ (by symmetry) allows us to obtain an upper bound for Theorem 1(iii) fairly simply using the same approach, since the right hand side of this inequality is amenable to a percolation approach (for the left hand side, where would the disease start?). However, the proof of the lower bound of Theorem 1(iii) is longer and more technical.

Janson goes on to consider the asymptotic distributions of X_{ij} and $\max_j X_{ij}$, which are not normal, and also the length of the path of minimal weight, but I do not expect to have time to cover these topics.