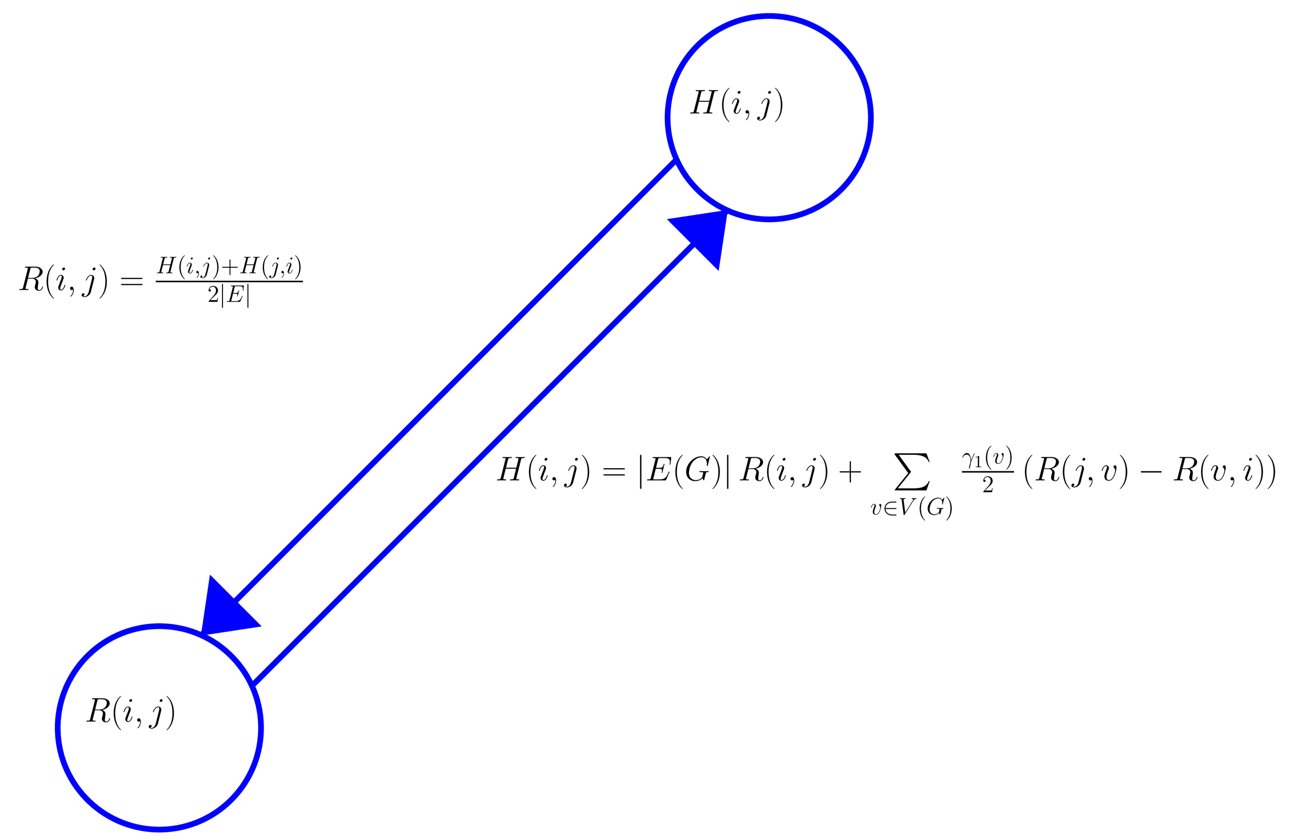
Hitting times and effective resistance in Erdős-Rényi graphs.

John Sylvester, University of Warwick
Mathematics Institute, Department of Statistics

Introduction

Let $G = (V, E)$ be a connected, simple graph on n vertices. The hitting time $H(i, j)$, $i, j \in V(G)$ is the expected time taken for a simple random walk on G to hit j when started from i. Another function on vertices of $i, j \in V(G)$ is the effective resistance $R(i, j)$. In physical terms this is the energy dissipated by a unit current between i and j when we consider the graph as an electrical network where every edge has unit resistance. The hitting times and effective resistance are closely related [1]:



An Erdős-Rényi random graph $\mathcal{G} \overset{d}{\sim} \mathcal{G}(n,p)$ is a simple graph on n vertices sampled according to the law $\mathbb{P}$ which is the product measure over edges where each edge occurs independently with probability $p := p(n)$, see [2]. If we consider hitting times and effective resistances in an Erdős-Rényi random graph $\mathcal{G} \overset{d}{\sim} \mathcal{G}(n,p)$ then $H(i,j)$ and $R(i,j)$, $i, j \in V(\mathcal{G})$, are random variables on the space of n-vertex simple graphs.

Results

We require $\mathcal{G}$ is sparsely connected, this means p must satisfy the following

$$c \log n \le np \le n^{1/10} \quad \text{for any } c > 1. \tag{1}$$

The first theorem concerns the expectation of the hitting times and effective resistances.

Theorem 1

Let $\mathcal{G} \overset{d}{\sim} \mathcal{G}(n,p)$ satisfy (1). Then for all $i, j \in V(\mathcal{G})$,

$$\mathbb{E}\left[H(i,j) \,\middle|\, \mathcal{G} \text{ connected}\right] \sim n$$

and

$$\mathbb{E}\left[R(i,j) \,\middle|\, \mathcal{G} \text{ connected}\right] \sim \frac{2}{np}.$$

We also have concentration around the means for hitting times and effective resistances.

Theorem 2

Let $\mathcal{G} \overset{d}{\sim} \mathcal{G}(n,p)$ satisfy (1) and $i, j \in V(\mathcal{G})$. Then with high probability

$$(1 - o(1))\frac{2}{np} \le R(i,j) \le (1 + o(1))\frac{2}{np}$$

and

$$(1 - o(1))\, n \le H(i,j) \le (1 + o(1))\, n.$$

More precise versions of these two theorems giving explicit errors and probabilities are stated and proven in [3]. Similar results for commute times, Kirchoff index, cover cost, random starting times, mean hitting times and the Kemeny constant are also obtained.

Discussion

Since the size of the neighbourhoods in $\mathcal{G}(n,p)$ are distributed $Bin(n-1,p)$ our results show that the main contribution to $R(i,j)$ comes from the first neighbourhoods of $i, j \in V(\mathcal{G})$. The hitting times in the complete graphs K_n are all $n-1$, with this in mind our results imply you can do Bernoulli percolation on K_n removing all but a vanishing proportion of the edges and on average you won't notice the simple random walk between two points taking any longer.

The lower assumption on p comes from the connectivity threshold and the upper one comes from our construction. The fact the results do not hold for large p is not of so much importance as when $np \ge (\log n)^{C_0}$ for large enough C_0 the results follow from [4]. The novel element to our results is at the lower end of the range of suitable p where it is hard to obtain the spectral estimates needed to apply the results in [4].

The idea of the proof is to control $R(i,j)$ by showing the main contribution comes from the first neighbourhoods of i and j. The lower bound on $R(i,j)$ follows from neglecting contributions to $R(i,j)$ from edges outside the first neighbourhoods of i, j, the upper bound is more complex and is mentioned next. We can then control $H(i,j)$ by the previously mentioned formula relating it to $R(i,j)$.

Upper bound on $R(i, j)$

- Run a modified breadth first search on G starting simultaneously from i and j, removing edges and vertices which create paths from i to j or cycles. From this process we define the various "pruned" neighbourhoods $\Phi_1(x), \Psi_1(x), \Gamma_i^*(x)$ of a vertex $x \in V$.
- Using these neighbourhoods we define $\mathcal{A}_{i,j}^{n,k}$, the strong k path property:

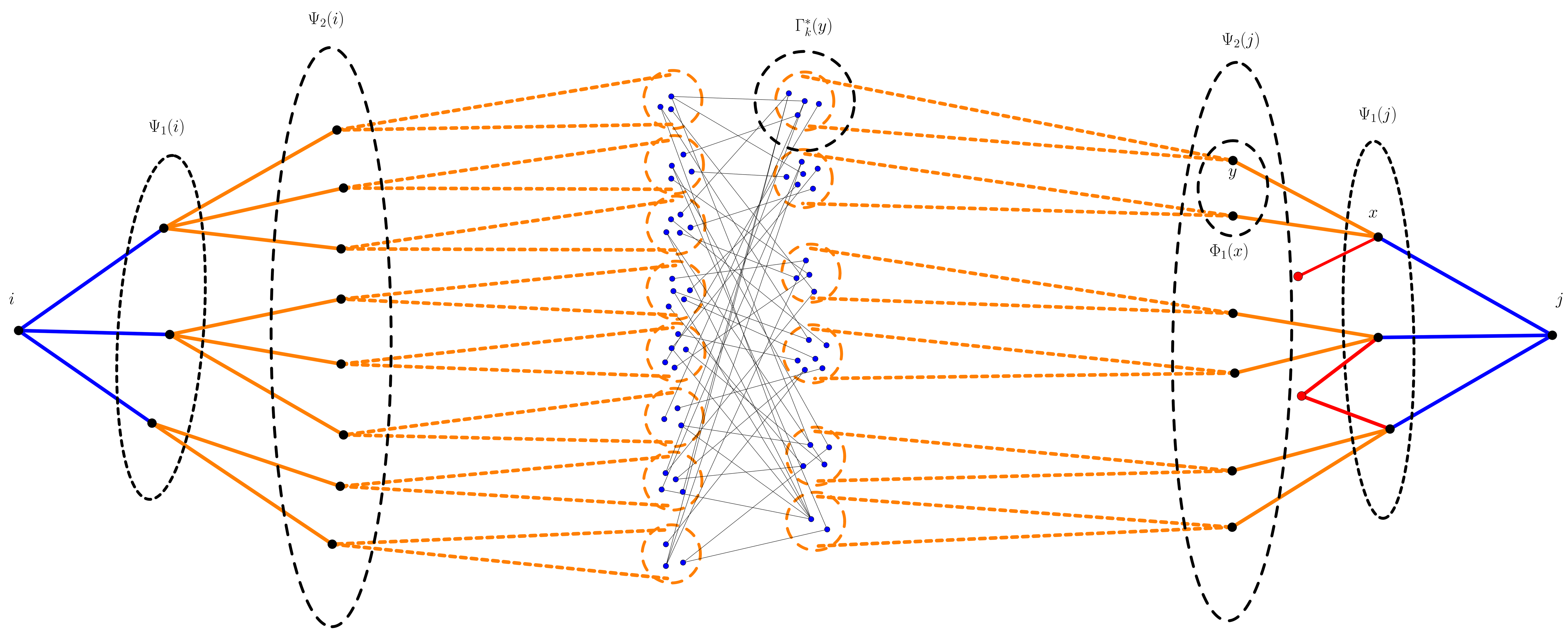
$$\mathcal{A}_{i,j}^{n,k} := \left\{G : V = [n], \forall (x,y) \in \Psi_2(i) \times \Psi_2(j), \exists (u,v) \in \Gamma_k^*(x) \times \Gamma_k^*(y), uv \in E\right\}.$$

- If $G \in \mathcal{A}_{i,j}^{n,k}$ then G contains a "nice" subgraph (shown in the diagram below). Using a flow through this subgraph the following upper bound can be given for $G \in \mathcal{A}_{i,j}^{n,k}$,

$$R(i,j) \le \frac{1}{|\Psi_1(i)|} + \frac{1}{|\Psi_1(j)|} + \sum_{a=1}^{|\Psi_1(i)|} \frac{k+2}{|\Psi_1(i)|^2 |\Phi_1(i_a)|} + \sum_{b=1}^{|\Psi_1(j)|} \frac{k+2}{|\Psi_1(j)|^2 |\Phi_1(j_b)|}.$$

- For $\mathcal{G}(n,p)$ couplings are used to show $|\Phi_1(x)|, |\Psi_1(x)|, |\Gamma_i^*(x)|$ are all close in distribution to binomial rv's with means $np(1-o(1))$ or $(np)^i(1-o(1))$ in the case of $|\Gamma_i^*(x)|$.
- Neighbourhood branching estimates can then be used to show that if $\mathcal{G} \overset{d}{\sim} \mathcal{G}(n,p)$ satisfies (1) then $\mathcal{G} \in \mathcal{A}_{u,v}^{n,k}$ and $k = O\left((\log n)/\log(np)\right)$ with high probability.
- Expectation and high probability bounds can then be obtained for the right hand side of the upper bound on $R(i,j)$ using the couplings for $|\Phi_1(x)|, |\Psi_1(x)|$.

The strong k-path property



Acknowledgements

I would like to thank my supervisors Agelos Georgakopoulos and David Croydon for their help and guidance.

This work is supported by EPSRC as part of the MASDOC DTC at the University of Warwick. Grant No. EP/HO23364/1.

References

[1] P. Tetali, Random walks and the effective resistance of networks. Journal of Theoretical Probability **4**(1), 101-109.

[2] P. Erdős & A. Rényi , On random graphs I. Math. Debrecen, **6** 1959, 290–297.

[3] J. Sylvester, Hitting times and effective resistance in sparsely connected Erdős-Rényi graphs. In preparation.

[4] U. von Luxburg, A. Radl & M. Hein, Hitting and Commute Times in Large Random Neighborhood Graphs. Journal of Machine Learning Research **15** 2014, 1751-1798.