Return probabilities and hitting times of random walks on sparse Erdős-Rényi graphs

O. C. Martin and P. Šulc

Univ Paris-Sud, LPTMS ; CNRS, UMR8626, Orsay, F-91405, France.

(Dated: April 23, 2009)

We consider random walks on random graphs, focusing on return probabilities and hitting times for sparse Erdős-Rényi graphs. We show how to solve for the distribution of these quantities in the thermodynamic limit and we find that these distributions exhibit structures on all scales.

PACS numbers: 05.40.-a,05.40.Fb,46.65.+g

I. Introduction

Random walks are some of the simplest stochastic processes \[1\, 2\] and yet they arise in many scientific fields such as pure mathematics, statistical physics or even biology \[3\, 4\, 5\, 6\]. A fundamental quantity for computing probabilities is the return probability \(1/N\) at large \(N\). W e thus map the problem where at time 0 a walker is equi-distributed on all nodes \(n\) to the node \(i\). The matrix \(D\) is diagonal; its \(i\)-th diagonal element \(D_{ii}\) is equal to the degree \(d_i\) of the \(i\)-th node.

To investigate the hitting time of the walker to go from \(s\) to \(t\), it is enough to initialize the vector \(v^{(0)}\) to be zero on all nodes except at \(s\) where it is 1, and to impose absorbing conditions at the target node \(t\), i.e., \(v^{(n)}_{i} = 0\) at all \(n\). Then the probability of having a first passage time equal to \(n\) is given by the flux into node \(t\) at that time step \[7\]. A modified treatment of the walker allows one to obtain the probability of return to \(s\).

Our mathematical solution concerns Erdős-Rényi graphs in the ensemble \(G(N,p)\), where \(N\) is the total number of nodes and each pair of nodes has probability \(p\) to be connected by an edge. For sparse graphs, \(p = c/N\) where \(c = (d)\) is the mean degree of nodes. We shall also consider fixed degree random graphs, also called random \(d\)-regular graphs, where each node has exactly degree \(d\) but connections are otherwise random \[17\].

III. Hitting times on random \(d\)-regular graphs

Let us first compute the hitting time on random regular graphs, exploiting their local tree-like nature. Indeed, loops can arise in random \(d\)-regular graphs \[17\] but their typical length is \(O(\ln(N))\). Thus it is expected that most properties can be obtained by studying what happens locally, as long as boundary conditions at “infinity” are properly handled. Such an approach has been used in many contexts with a high level of success \[18\, 19\].

For a given random regular graph, of fixed degree \(d\), we consider a node \(t\) and ask what is the mean of \(H(s,t)\) when averaged over all \(s\). We need to solve a diffusion problem where at time \(n = 0\) a walker is equi-distributed amongst the \(N-1\) nodes \(s\) (\(t \neq s\)) and if the walker hits node \(t\) it gets absorbed. If one denotes by \(F^{(n)}\) the probability flux into node \(t\) at step \(n\), then the hitting time averaged over all \(s\) is given by the first moment of \(n\) distributed as \(F^{(n)}\).

In the neighborhood of \(t\), the graph is a Cayley tree with probability one at large number of nodes \(N\) and thus does not depend on \(t\) in the large \(N\) limit. Given
the diffusion-absorption process, the vector of probabilities quickly converges to the dominant eigenvector of the master equation (that with the largest eigenvalue, decaying the slowest). In the limit of large $N$, the decay rate goes to zero and all the transient behavior (associated with the other eigenvectors) becomes irrelevant. When $N \to \infty$, it is then enough to determine the dominant eigenvector, imposing zero boundary conditions at the root node $t$ (labeled 0 hereafter) and $1/(N-1)$ boundary conditions for the far away nodes. As $N \to \infty$, the recurrence equation that is satisfied by the eigenvector’s elements leads to $dA_{k+1} = A_{k+2} + (d-1)A_k$ where $A_k$ is the sum of the probabilities on the nodes that are at distance $k$ from the root node. Solving this, subject to the normalization and boundary conditions, leads to the value of $A_1$ and thus the flux flowing into the absorbing node: $F = A_1/d$.

Note that since at large $N$ only the leading eigenvector matters, the first passage time is exponentially distributed with a mean given by the inverse of this flux. This then gives for random $d$-regular graphs a hitting time behaving at large $N$ as

$$
\frac{H}{N} \sim \frac{d-1}{d-2}.
$$

Finally, it is worth noting that for random $d$-regular graphs, with probability 1 in the large $N$ limit, the ratio $H(s,t)/N$ does not depend on the starting node $s$. Also, because of the regularity of the graph, this quantity does not depend on $t$ either.

IV. Probability of return on random $d$-regular graphs

On any finite graph, a walker leaving node $s$ will return with probability one. Nevertheless, if one considers the distribution of return times for increasing values of $N$, one will find that there is a $N \to \infty$ limiting point-wise distribution but which does not integrate to 1. Indeed, in that limit, the return times will be finite with probability $\hat{r}$ and will diverge linearly in $N$ with probability $1 - \hat{r}$. If $\hat{r} \neq 1$, the walk is transient. On the infinite Cayley tree, $\hat{r}$ can be computed simply by using the homogeneity of the graph as follows.

Take $s$ to be the root of an infinite Cayley tree. The walker must make a first step; let it be to one of its neighbors $s'$. Define $r$ as the probability for the walk to return to $s$ given that it has stepped to $s'$. Using the equivalence of all nodes, one can write a series for $r$:

$$
r = \frac{1}{d} + \frac{(d-1)r}{d} \frac{1}{d^2} + \frac{(d-1)r^2}{d^3} + \ldots
$$

where $d \geq 2$ is the degree of the Cayley tree. In this series, the term of $O(r^k)$ corresponds to the probability that the walk returns $k$ times to node $s'$ before going back to the root $s$. Summing this geometric series gives two possible values: $r = 1$ and $r = 1/(d-1)$. Furthermore, it is easy to see that $\hat{r} = r$. If $d = 2$, we have a one dimensional walker and $\hat{r} = 1$. For $d \geq 3$, the walk is transient and $\hat{r} = 1/(d-1)$.

V. Probability of return on Erdős-Rényi random graphs

Here we extend the previous calculation of return probabilities to the case of Erdős-Rényi graphs. Just as for the random $d$-regular graphs, we exploit the fact that with probability 1 in the large $N$ limit the neighborhood of a node belonging to a sparse Erdős-Rényi graph is locally tree-like. We denote by $c = \langle d \rangle$ the mean degree of these graphs; the probability to have a node of degree $d$ is $P(d) = e^{-c}c^d/d!$, i.e., is given by the Poisson distribution.

To find the probability to return in a finite number of steps (at large $N$) for a walker starting on the root node (hereafter referred to as 0), we reconsider the series of Eq. (3). Suppose that at the first step the walker moves to the neighbor $j$ of the root node, and that $d_j$ is the connectivity of that node. If the walker is to return to 0, it can do so immediately, or it can perform $k$ loops from $j$ (avoiding 0), stepping back to 0 only after its $(k+1)$th visit to node $j$. By a loop from $j$, we mean a step to one of the $d_j-1$ neighbors of $j$ other than 0, then a finite number of steps that do not visit $j$, and then finally a return to $j$. The point is that in our system the walker cannot come back to 0 other than through the edge connecting $j$ to 0: any other route requires going to “infinity” and thus an infinite number of steps. (Since we are dealing with a return probability on an infinite graph, the walks returning to 0 must have a finite number of steps.)

For the edges connecting node $j$ to a node other than 0, let the return probabilities be $r_j(1)$, $r_j(2)$, . . . $r_j(d_j-1)$. Given these $r_j$s, the probability $r_0$ to return to the root node if the walk’s first step is to node $j$ is

$$
r_0 = \frac{1}{d_j - \sum_{m=1}^{d_j-1} r_j(m)}.
$$

However, the $r_j(m)$ are i.i.d. random variables belonging to a distribution $\rho(r)$. In the Erdős-Rényi ensemble, 0 connects to a random node ($j$ here) which itself connects to other random nodes. The distribution of $r_0$ is thus the same as that of the $r_j$s, and Eq. (4) determines implicitly a self-consistent functional equation for $\rho(r)$. This can be written formally as:

$$
\rho(r) = \sum_{z=0}^{\infty} P(z) \int dr_1 \ldots \int dr_z \rho(r_1) \ldots \rho(r_z)
$$

$$
\times \frac{1}{1 + \sum_{i=1}^{z}(1-r_i) - r}
$$

where $P(z)$ is the Poisson distribution (of $z = d_j - 1$), and $\delta(x)$ is the Dirac delta function. Also, note that in this formula the $z = 0$ term must be interpreted as $P(0)\delta(1-r)$.

We have solved for $\rho$ by numerical iteration, demanding a stable distribution. Because $\rho$ has both a continu-
VI. Hitting times on Erdős-Rényi random graphs

To compute the hitting time \( H(s,t) \), we take \( s \) and \( t \) to be on the same connected component whose size we denote by \( N_\infty \). For Erdős-Rényi graphs, we work beyond the percolation threshold, \( c > 1 \), on the “infinite” component, so \( N_\infty \approx (1-\Delta)N \). With probability 1, the hitting time \( H(s,t) \) scales with \( N \), has negligible fluctuations with \( s \), and depends only the neighborhood properties of \( t \). We thus focus on \( H(t) \), the mean of \( H(s,t) \) when averaging over all nodes \( s \) distinct from \( t \). This problem has been solved for dense Erdős-Rényi graphs and leads to \( H(t) = N + o(N) \) \[12\]. For the sparse case, no exact treatment has been proposed, but a mean-field like approximation gives rather good results \[14\]. We now provide an exact mathematical approach.

As explained previously, we can follow the probability of finding the walker on any node. The initial condition is that every node except \( t \) is occupied with the same probability \( 1/(N_\infty - 1) \). The absorption at node \( t \), hereafter labeled 0, imposes \( \psi_0^{(n)} = 0 \) at all times. The master equation for this process is therefore

\[
\psi_i^{(n+1)} = \left( TAD^{-1}\psi_i^{(n)} \right)_t,
\]

where \( T_{ij} = \delta_{ij}(1-\delta_{0i}) \). Denote by \( S \) the leading eigenvector of the diffusion operator \( AD^{-1} \) having no absorption, with eigenvalue 1. For a normalisation of the probabilities to 1, one has \( S_i = d_i/(N_\infty \langle d \rangle_\infty) \) where \( d_i \) is the degree of the \( i \)-th node. Furthermore, \( \langle d \rangle_\infty \) is the mean degree on the connected component considered, which in our case is not \( c \) because we have the constraint of belonging to the infinite component, instead it is

\[
\langle d \rangle_\infty = \frac{\sum_{z=1}^{\infty} z(1-\Delta^z)P(z)}{\sum_{z=1}^{\infty}(1-\Delta^z)P(z)}.
\]

It is easy to check that under evolution without absorption \( S \) is unchanged: since the walk is on a connected component, this is the only normalized steady state distribution. We now introduce the vector \( b^{(n)} \) that represents the difference between the vector \( S \) and the vector \( \psi^{(n)} \):

\[
\frac{1}{N_\infty} b_I^{(n)} = \frac{1}{N_\infty} \langle d \rangle_\infty - \psi_I^{(n)}.
\]

The absorption condition at 0 then imposes \( b_0^{(n)} = d_0/(\langle d \rangle_\infty) \) for all \( n \). Far away from the root node, the distribution quickly relaxes to the leading eigenvector of the diffusion equation. In the \( N_\infty \to \infty \) limit, almost all nodes are oblivious to the absorption, so we can compute the hitting time by assuming that \( \psi_m^{(n)} \) is equal to \( S_m \) for
all nodes $m$ at “infinity”, which gives us the boundary condition $b_m^{(n)} = 0$ at all times.

Now we can interpret the evolution equation for $b^{(n)}$ as describing a process of multiple random walkers diffusing on the graph, with in addition a fixed source at the root node. Specifically, at each time step $n$, $b_0^{(n)}$ new walkers are created at the root and step away while any walkers incoming to the root are removed from the system. With increasing number of iterations, the vector $b^{(n)}$ converges to a steady-state $\tilde{b}$ (as $v^{(n)}$ converges to $\tilde{v}$, a leading eigenvector of $TAD^{-1}$) in which for each edge $(0j)$ connected to the root node, there is an outgoing flux of $1/(d)_{\infty}$ and a corresponding incoming flux of $r_j/(d)_{\infty}$ where $r_j$ is the probability of return to 0 of a walker given that it has stepped to $j$. The flux into $b_0$ is then equal to the flux of “returning” random walkers:

$$\sum_{j \neq 0} \frac{1}{d_j} \tilde{b}_j = \frac{1}{\langle d \rangle_{\infty}} \sum_{j \neq 0} r_j.$$  

Coming back to the the formalism based on $\tilde{v}$, i.e., the leading eigenvector of $TAD^{-1}$, the net total flux $F(t)$ into the absorbing node 0 is given by

$$F(0) = \sum_{j \neq 0} \frac{1}{d_j} \tilde{v}_j.$$  

Using Eqs. (7) and (8) one obtains the final expression

$$F(0) = \frac{1}{N_{\infty} \langle d \rangle_{\infty}} \sum_{j \neq 0} (1 - r_j).$$

In the previous section we derived the distribution of $r_j$, from which one easily obtains the distribution for $H(0) = 1/F(0)$. First, for each value $z \geq 1$ of $d_0$ (the degree of the root node), we compute the distribution of $F(0)$. The delta function part of this distribution (at $F(0) = 0$) is removed and the remaining distribution if rescaled to have norm 1. This corresponds to enforcing the constraint that the absorbing node is on the infinite component of the Erdős-Rényi graph: the part of the distribution of $F(0)$ which gives zero flux corresponds to being on a finite component. Second, the distribution of $H_z = 1/F(0)$ is extracted: call it $\mu_z(H_z)$. Finally, given all the distributions $\mu_z$ ($1 \leq z < \infty$), the distribution of $H$ at a random node is obtained by averaging the $\mu_z$ with their respective weights:

$$\mu(H) = \sum_{z=1}^{\infty} \frac{\mu_z(H_z) P(z)(1 - \Delta^z)}{\sum_{j=1}^{\infty} P(j)(1 - \Delta^j)}.$$  

An example of such a distribution is shown in Fig. 2 when $\langle d \rangle = 4$. Furthermore, the distribution of $H$ also gives the distribution of first passage times since at large $N$, for each value of $H$, the first passage time $n$ is distributed as $\exp(-n/H)$. Finally, to obtain the mean hitting time, it is enough to compute the mean of the distribution of $H$. We have done so and show in Fig. 3 the resulting values, normalized by $N_{\infty}$, as a function of the mean degree of the graphs. At large $\langle d \rangle$, the ratio converges to 1 with $O(1/(\langle d \rangle))$ corrections: one recovers the dense graph result. Also, the behavior is very smooth and we find that it differs from the value when the degree does not fluctuate (the case of random $d$-regular graphs) also by $O(1/(\langle d \rangle))$.

VII. Comparison with numerical simulations

Fig. 4 shows the mean hitting times on the largest connected component of an Erdős-Rényi graph with mean degree $\langle d \rangle = 4$. The estimation from Eq. (11) is compared with values obtained from a numerical simulation in which we followed the probability vector $v^{(n)}$ as in Eq. (5). The mean hitting times were then averaged over multiple graphs. The errorbars are shown as well. We found that values determined from the simulations approach their large $N$ limit rather fast and that this limit is compatible with our analytical result, the relative differ-

![FIG. 3: Mean hitting times divided by $N_{\infty}$ for Erdős-Rényi graphs in the limit of large graphs, as a function of the parameter $c = \langle d \rangle$ equal to the graphs’ mean node degree. $N_{\infty}$ is the size of the “infinite” component, $N_{\infty} \approx (1 - \Delta)N$ for graphs of $N$ nodes.](image)

![FIG. 4: Plot comparing numerical simulation with analytical results. The x axis shows the size of the largest connected component of the graph, the y axis shows the mean hitting time for such a component.](image)
ence being compatible with a $O(1/N)$ convergence. The same conclusion also holds in the context of random $d$-regular graphs (cf. Eq. (2)).

VIII. Discussion and conclusion

We considered random walks on random graphs, focusing on two quantities: the distribution of hitting times and the probability that a walker will return to its starting point in a finite time. (The hitting time is the mean of first passage times.) We derived a way to calculate the large $N$ behavior of these quantities on two families of random graphs, finding non-trivial and intricate distributions associated with the discrete nature of possible neighborhoods of a node. Finally, we compared the calculated results with numerical simulation and found excellent agreement, supporting the expectation that the loops in these graphs can be treated by appropriate boundary conditions on infinite trees.

Acknowledgments—We thank Satya Majumdar for enlightening discussions. This work was supported by the EEC’s IST project GENNETEC - 034952. The LPTMS is an Unité de Recherche de l’Université Paris-Sud associées au CNRS.

[1] B. Hughes, Random Walks and Random Environments (Clarendon Press, Oxford, 1996).
[2] J.-P. Bouchaud and A. Georges, Phys. Rep. 195, 127 (1990).
[3] E. Montroll and K. Shuler, PNAS 76 (12), 6030 (1979).
[4] M. Newman, Social Networks 27, 39 (2005).
[5] J. Noh and H. Rieger, Phys. Rev. Lett. 92, 118701 (2004).
[6] O. Benichou, M. Coppey, M. Moreau, P. Suet, and R. Voituriez, Phys. Rev. Lett. 94, 198101 (2005).
[7] W. Feller, An introduction to probability theory and its applications (John Wiley and sons, New York, NY, 1950).
[8] S. Redner, A guide to First-Passage Processes (Cambridge University Press, Cambridge, 2001).
[9] L. Lovász, in Paul Erdős is Eighty (Combinatorics, 1993).
[10] B. Bollobás, Random Graphs (Cambridge University Press, Cambridge, 2001).
[11] P. Erdös and A. Rényi, Publ. Math. Debrecen 6, 290 (1959).
[12] F. Chung, L. Lu, and V. Vu, Proceedings of the National Academy of Sciences 100, 6313 (2003).
[13] V. Sood, S. Redner, and D. Ben-Avraham, J. Phys. A 38, 109 (2005).
[14] A. Baronchelli and V. Loreto, Phys. Rev. E 73, 026103 (2006).
[15] G. Biroli and R. Monasson, J. Phys. A 32, L255 (1999).
[16] G. Semerjian and L. F. Cugliandolo, J. Phys. A 35, 4837 (2002).
[17] N. Wormald, in Surveys in Combinatorics (Cambridge University Press, 1999), pp. 239–298.
[18] D. J. Aldous, Probability Theory and Related Fields 18(4), 381 (2001), math.PR/0010063.
[19] M. Mézard, G. Parisi, and R. Zecchina, Science 297, 812 (2002).