First-passage properties of the Erdős–Renyi random graph

2005 J. Phys. A: Math. Gen. 38 109

(http://iopscience.iop.org/0305-4470/38/1/007)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **38** (2005) 109–123 doi:10.1088/0305-4470/38/1/007

First-passage properties of the Erdős–Renyi random graph

V Sood[1,3], **S Redner**[1,3] **and D ben-Avraham**[2]

[1] Theory Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2] Department of Physics, Clarkson University, Potsdam, NY 13699, USA

E-mail: vsood@bu.edu, redner@bu.edu and qd00@clarkson.edu

Received 14 October 2004, in final form 28 October 2004
Published 8 December 2004
Online at stacks.iop.org/JPhysA/38/109

Abstract

We study the mean time for a random walk to traverse between two arbitrary sites of the Erdős–Renyi random graph. We develop an effective medium approximation that predicts that the mean first-passage time between pairs of nodes, as well as all moments of this first-passage time, are insensitive to the fraction p of occupied links. This prediction qualitatively agrees with numerical simulations away from the percolation threshold. Near the percolation threshold, the statistically meaningful quantity is the mean transit rate, namely, the inverse of the first-passage time. This rate varies non-monotonically with p near the percolation transition. Much of this behaviour can be understood by simple heuristic arguments.

PACS numbers: 02.50.−r, 05.40.Fb, 05.60.−k, 89.75.−k

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In this paper, we study a basic first-passage characteristic of random walks on Erdős–Renyi (ER) random graphs [1], namely, the mean time for a random walk to traverse between two arbitrary sites on the graph. The ER random graph is constructed by taking N sites and introducing a bond between each pair of sites with probability p. When $p = 1$, all possible links exist and this construction gives the complete graph, where each site is connected to all the other $N - 1$ sites in the graph. As p decreases, the random graph undergoes a percolation transition at $p = p_c = 1/N$ [2, 3] that shares many common features with percolation on regular lattices. Another geometrical feature that is relevant for our study of first-passage characteristics is a second connectivity transition at $p_1 = \ln N/N$. For $p > p_1$, all nodes

[3] On leave of absence from Department of Physics, Boston University, USA.

belong to a single component (in the limit $N \to \infty$), while for $p < p_1$ disjoint clusters can exist [2, 3].

Much effort has been devoted to determining basic properties of random walks in disordered or heterogeneous environments [5–8], such as the ER random graph, as well as small-world [9] and scale-free networks [10]. On the random graph, there has been considerable work in determining how basic time scales of the random walk depend on the size of the graph. These include the mixing time—the time scale that determines how the probability distribution approaches its limiting behaviour [9, 11, 12], the cover time—the time for a random walk to visit all sites of the graph [13, 14], and the first-passage time—the time for a random walk to traverse between two specified points for the first time [10] or to return to its starting point for the first time [15].

The goal of this work is to understand by simple physical arguments how a basic first-passage property of random walks on random graphs depends on the concentration of bonds in the graph. We will begin by studying the mean first-passage, or transit, time T_{ab} for a random walk to reach an arbitrary site b on the graph when starting from another arbitrary site a. In general, this transit time T_{ab} does not necessarily equal T_{ba}, and it is sometimes convenient to consider the mean commute time $K_{ab} \equiv T_{ab} + T_{ba}$ to avoid the asymmetry in T_{ab}. However, in averaging over all pairs of sites a and b to obtain statistically meaningful quantities, the asymmetry is eliminated and $\langle K_{ab} \rangle = 2\langle T_{ab} \rangle$.

In practice, mean transit and commute times diverge when the graph consists of more than one component (cluster), because a random walk that starts in one component cannot access sites in different components. We are thus led to study the dependence of the mean commute rate $\mathcal{R}_{ab} \equiv 1/K_{ab}$ as a function of p on the random graph. This rate equals zero for two sites in different components, so that its configurational average is meaningful. Focusing on the commute rate is analogous to considering mean conductance of a random conductor–insulator mixture near the percolation threshold, rather than the mean resistance. The conductance is well behaved near the percolation threshold, while the resistance is divergent for all p in any finite-size system.

In the next section, we construct an effective medium approximation that predicts that the mean commute time is independent of p for $p > p_1$. We also find that the transit time varies weakly with p in the small-dilution limit. In section 3, we present simulation results for the mean commute time and the mean commute rate and find an unexpected non-monotonic behaviour for the latter quantity as a function of p when $p \approx p_c$. In section 4, we outline the relation between the commute time on a network and the two-point conductance on the same network when each link is a unit resistor. We use this electrical network connection to explain the non-monotonicity of the commute rate, first for a tree structure that is a subset of the random graph (section 5), and then for the random graph itself (section 6).

2. Analytic approaches for the transit time

2.1. Effective medium approach

We now develop an effective medium approximation for the commute time of a discrete-time random walk on the random graph [16]. In a single time step, a walk located at a site that is connected to z other sites can hop with probability $1/z$ to any of these neighbours. To compute the mean time for such a random walk to go between two arbitrary sites on any graph by a sequence of nearest-neighbour hops, we use the underlying backward equation [17, 18]. This equation relates the transit time from site a to site b to the transit times from the neighbouring

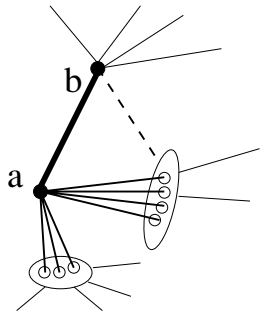

Figure 1. Schematic decomposition of a random graph with starting site a and target site b. The direct link is shown as the thick solid line. After one step via an indirect path (to the sites in of the ovals), either b can be reached directly with probability p (dashed line), while with probability $q = 1 - p$ there is still no direct connection to b.

sites of a to site b as follows:

$$T_{ab} = \sum_{\Pi} P_{\Pi} t_{\Pi} = \sum_{i} p_{a\to i}(\delta t + T_{ib}). \tag{1}$$

The first sum is over all paths Π from a to b, P_{Π} is the probability for the random walk to take the path Π, and t_{Π} is the transit time from a to b along this path. For each path, we then decompose the full transit time into the time to go from a to an intermediate site i after one step plus the time to go from i to b. Thus $p_{a\to i} = 1/z_a$ is the probability of hopping from a to i in a single step, z_a is the degree of a, and δt is the time for each step of the random walk. Without loss of generality, we take $\delta t = 1$.

Let us now construct an effective-medium approximation for the average transit time on the random graph, under the assumption that the graph is connected. This condition implicitly restricts the validity of our approach to the range $p > p_1$, where all nodes belong to a single component. A schematic representation of a random graph, to illustrate our approach, is shown in figure 1. Between two sites a and b on the graph, a direct link to b may exist (thick line) with probability p. If there is no such direct link, then an indirect path must be followed. After a single step on this indirect path (medium lines), there may be a direct link to b with probability p (dashed), or no direct link with probability $q = 1 - p$.

Let us denote by τ the mean transit time to go from a to b under the assumption that a direct link exists, and τ' the transit time from a to b in the absence of a direct link. Then from equation (1) and following an effective-medium assumption, τ obeys the recursion formula

$$\tau = \frac{1}{(N-1)p} + \left[1 - \frac{1}{(N-1)p}\right][p(1+\tau) + q(1+\tau')]. \tag{2}$$

The first term accounts for the walk that goes directly from a to b. This contribution corresponds, in equation (1), to the case where the intermediate site i coincides with b. Since $(N-1)p$ links emanate from a on average, then according to the effective-medium approximation, the probability that a random walk steps along the direct connection is just $1/(N-1)p$. The second set of terms accounts for those walks in which the first step goes to an intermediate site i rather than hitting b directly. In this case, we again apply an effective-medium approximation and posit that after one step of the walk, a direct connection from i to b exists with probability p, or no direct connection exists with probability $1 - p$ (figure 1).

To close this equation, we need an expression for τ', the first-passage time in the absence of a direct connection to b. Applying the same effective-medium approximation as that used

in equation (2), we assume that after the first step of the walk, the terminal site b is directly reachable with probability p, while b is not directly reachable with probability q. Thus τ' obeys

$$\tau' = p(1+\tau) + q(1+\tau'). \tag{3}$$

Solving equations (2) and (3) gives $\tau = N - \frac{1}{p}$ and $\tau' = N$. Finally, we average the transit time over all pairs of terminal points and over all graph configurations. Again in the spirit of an effective medium approximation, this average is simply

$$\langle T\rangle \equiv \langle T_{ab}\rangle = p\tau + (1-p)\tau' = N-1. \tag{4}$$

Surprisingly, $\langle T\rangle$ is independent of p. Thus according to the effective medium approach, the complete graph solution, $\langle T\rangle = N-1$, holds for all p.

The backward equation for the mean transit time can be extended to any positive integer moment of the transit time. Consider, for example, the mean-square transit time. As in the case of the mean time, the governing equation can formally be written as

$$T_{ab}^2 = \sum_{\Pi} P_\Pi t_\Pi^2, \tag{5}$$

For each path, we follow equation (1) and again write the transit time t_Π as $1 + t_{\Pi'}$, namely, the sum of the time for the first step and the time for the remainder of the path. Thus

$$\begin{aligned} T_{ab}^2 &= \sum_{\Pi} P_\Pi (1+t_{\Pi'})^2, \\ &= \sum_{\Pi} P_\Pi \big(1 + 2t_{\Pi'} + t_{\Pi'}^2\big), \\ &= \sum_{i} p_{a\to i}\big(1 + 2\tau_i + \tau_i^2\big). \end{aligned} \tag{6}$$

In going from the second to the last line of this equation, we use the fact that $P_\Pi = \sum_i p_i P_{\Pi'}$, where p_i is the probability of hopping from the starting point to one of its nearest neighbours i, and $P_{\Pi'}$ is the probability for the remainder of the path Π' from i to b. In the last line, the quantities τ_i and τ_i^2 are the mean and mean-square times to reach b when starting from i and the sum is over all neighbours i of the starting point. Strictly speaking, we should write $\langle\tau\rangle$ and $\langle\tau^2\rangle$ for these moments, so that it is obvious that $\langle\tau^2\rangle \neq \langle\tau\rangle^2$. In the following, we drop these angle brackets because the linear and quadratic powers of time always appear separately and there is no ambiguity about where the angle brackets should appear.

The last line of equation (6) is now a backward equation for the second moment of the first-passage time, in which the previously determined first moment is an input to this equation. This construction for the mean-square transit time can be generalized straightforwardly, albeit tediously, to any positive integer moment of the first-passage time. For the random graph, the recursion formula for the mean-square transit time is, in close analogy with equations (2) and (3).

$$\begin{aligned} \tau^2 &= \frac{1}{(N-1)p} + \left[1 - \frac{1}{(N-1)p}\right][p(1+2\tau+\tau^2) + q(1+2\tau'+\tau'^2)], \\ \tau'^2 &= p(1+2\tau+\tau^2) + q(1+2\tau'+\tau'^2). \end{aligned}$$

Using our previously derived results for the first moments, $\tau = N - \frac{1}{p}$ and $\tau' = N$, these recursion formulae are easily solved. We then compute the configuration averaged mean-square transit time, $\langle T^2\rangle \equiv p\tau^2 + q\tau'^2$, and obtain $\langle T^2\rangle = (2N-3)(N-1)$. Thus again, the second moment is independent of p and equals the second moment of the transit time on the complete graph.

More generally, we show that the first-passage probability between any two sites on a random graph, and thus all moments of the first-passage time, are independent of p in the effective-medium approximation. As a preliminary, we first compute the first-passage probability on the complete graph. Let $F(t)$ be the probability that a random walk hits the target site for the first time at time t, and let $F(z) = \sum F(t)z^t$ be the corresponding generating function. For the complete graph, the generating function obeys the recursion formula

$$F(z) = \frac{1}{N-1}z + \frac{N-2}{N-1}zF(z).$$

This equation encodes the fact that after a single step (the factor z) the walk hits the target site with probability $1/(N-1)$, while with probability $(N-2)/(N-1)$ the walk hits another interior site of the graph, at which point the first-passage process is renewed. The solution to this equation is

$$F(z) = \frac{z}{N-1}\left[1 - \left(\frac{N-2}{N-1}\right)\right]^{-1},$$

from which

$$F(t) = \frac{1}{N-1}\left(\frac{N-2}{N-1}\right)^t. \tag{7}$$

Now consider the random graph with bond occupation probability p. Let $\mathcal{F}(t)$ be the first-passage probability from a to b when a bond is present between these two sites, $\mathcal{F}'(t)$ the first-passage probability when this bond is absent, and let $\mathcal{F}(z)$ and $\mathcal{F}'(z)$ be the respective generating functions. In the spirit of our effective medium approximation given in equations (2) and (3), we now have

$$\mathcal{F}(z) = \frac{1}{(N-1)p}z + \left[1 - \frac{1}{(N-1)p}\right][pz\mathcal{F}(z) + qz\mathcal{F}'(z)],$$

$$\mathcal{F}'(z) = pz\mathcal{F}(z) + qz\mathcal{F}'(z).$$

From these two equations, the average first-passage probability $\langle F(z)\rangle = p\mathcal{F}(z) + q\mathcal{F}'(z)$ has the same form as the first-passage probability for the complete graph (equation (7)). Hence all moments of the transit time are independent of p in the effective medium approximation.

2.2. Small dilution limit

We may understand the exact dependence of the mean transit time in the limit $p \to 1$ by considering configurations with a single missing bond. There are four distinct cases to consider: (i) missing link between a and b (1 configuration), (ii) missing link between a and an interior point ($N-2$ configurations), (iii) missing link between an interior point and b ($N-2$ configurations) and (iv) missing link between two interior points (all remaining configurations). Let us denote the mean transit times from a to b for these four configurations by t_1, t_2, t_3 and t_4. By considering each case separately, we obtain the recursion formulae (cf equation (1)):

$$t_1 = 1 + t_3 \qquad t_2 = \frac{1}{N-2} + \frac{N-3}{N-2}(1+t_4)$$

$$t_3 = \frac{1}{N-1} + \frac{1}{N-1}(1+t_1) + \frac{N-3}{N-1}(1+t_3)$$

$$t_4 = \frac{1}{N-1} + \frac{2}{N-1}(1+t_2) + \frac{N-4}{N-1}(1+t_4),$$

with solution

$$t_1 = N + 1 \qquad t_2 = N - 1 - \frac{3}{N} \qquad t_3 = N \qquad t_4 = N - 1 - \frac{2}{N}. \quad (8)$$

Then, by averaging over the appropriate number of configurations for each class, we obtain the mean transit time

$$\langle T \rangle = \frac{N^3 - 2N^2 + N + 4}{N(N-1)} \approx N - 1 + \frac{4}{N^2}. \quad (9)$$

To interpret this result, note that the absence of a single bond corresponds to a bond concentration $p \approx 1 - 2/N^2$. At this value of p, the mean transit time has the asymptotic behaviour $\langle T \rangle / \langle T(p=1) \rangle \sim (1 + 4/N^3) \sim [1 + \sqrt{2}(1-p)^{3/2}]$. The first correction to $\langle T \rangle$ is thus of the order of $(1-p)^{3/2}$, rather than linear in $(1-p)$, as one might naively expect. This small first correction to $\langle T \rangle$ near $p = 1$ makes plausible the effective-medium result that $\langle T \rangle$ is independent of p.

3. Simulation results

To test the effective-medium prediction for the mean first-passage time, we now turn to numerical simulations. For very small systems ($N \leqslant 8$), we have obtained the exact first-passage time by averaging over all configurations of random graphs, over all pairs of endpoints, and over all random walks. For the graph configuration average, each realization is weighted by the factor $p^k q^{E-k}$, where k is the number of occupied links in the graph, $E = N(N-1)/2$ is the total number of possible links, and $q = 1 - p$. We then average over all pairs of endpoints directly. By this averaging, the mean transit time is simply one-half of the mean commute time. Rather than averaging over individual walks directly, we solve exactly the recursion formulae in equation (1) for the transit times between all pairs of points.

For larger systems, the exact enumeration of all graph configurations is impractical. Instead we average over a finite number of graph realizations and endpoint pairs, but still performed the exact average over all random walk trajectories by numerically solving equation (1). For efficiency, we start our simulation with an empty graph, add bonds one at a time and then update the commute times between all pairs of sites in the graph after each bond addition. Each graph is then weighted by $p^k q^{E-k}$ so that we can obtain the commute time as a function of p. We repeat this sequential graph construction over many realizations. The graphs that we obtain by this sequential growth are the same as those obtained by a static construction in which each bond is present with probability $p = 2M/N(N-1)$ when N is large (see [2] for different, but equivalent ways of constructing random graphs).

For the average commute time, we only include connected graphs in the ensemble, while for the average rate, the ensemble consists of all graph configurations. This restriction plays a significant role only for $p < p_1$, where the random graph normally consists of multiple components. Typical results for a graph of 100 sites are shown in figure 2. Above the connectivity threshold $p_1 = \ln N/N$, the average transit time varies slowly with p, in agreement with our effective medium approach. The apparent singularity of the average commute time at a value $p < p_1$ stems from finite size effects.

The behaviour of the mean commute rate is shown in figures 2 and 3. Unexpectedly, this rate is non-monotonic in p for $p \approx p_c = 1/N$, as shown in detail in figure 3. For this plot, we use the average degree, $\mu = p(N-1)$ as the dependent variable, because it has the desirable feature that the percolation transition occurs at the same value $\mu_c = 1$ for all N. The fact that the non-monotonicity in the commute rate occurs near $\mu = 1$ suggests that this anomaly is connected with the percolation transition of the random graph.

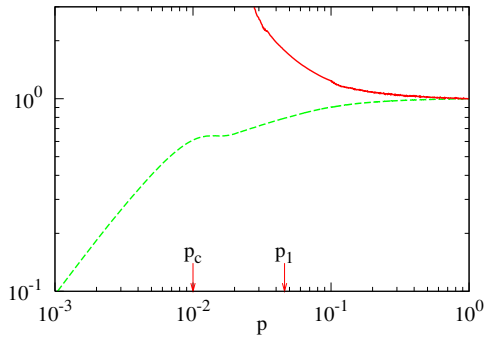

Figure 2. Mean commute time (solid) and mean commute rate (dashed) versus bond occupation probability p for a random graph of $N = 100$ sites. Both quantities are normalized to have the value 1 for the complete graph ($p = 1$). Averages over 10^3 graph realizations were performed for each p. Also shown are the locations of $p_c = 0.01$ and $p_1 \approx 0.046$.

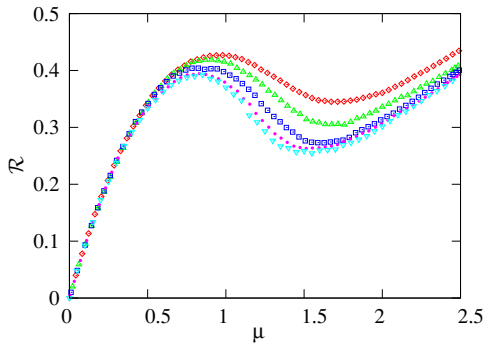

Figure 3. Mean commute rate $\mathcal{R}$ on a random graph for $N = 50$ ($\diamond$), $N = 100$ ($\triangle$), 200 ($\square$), 400 ($\circ$) and 800 ($\triangledown$) sites as a function of the average site degree $\mu = p(N-1)$. These rates are normalized to one for the complete graph limit. Averages over 10^3 graphs were performed for each case.

To understand this non-monotonicity, we first make a connection between the commute rate and the conductance on the same network when each link is a unit resistance, and then analyse the structure of the random graph in the critical regime to evaluate the conductance. This result will then be used to infer the dependence of the commute rate on the mean degree μ.

4. Electrical network connection

In principle, first-passage properties of random walks on a graph can be obtained from the underlying Laplacian matrix of the graph [5]. The eigenvalue spectrum of the Laplacian provides many time-dependent random walk characteristics. This matrix formulation also reveals deep analogies between random walks on a graph and the electrical network problem on the same graph in which each occupied link is a resistor of unit resistance (see [19] for a nice exposition of these connections).

For example, the commute time K_{ab} between a and b and the conductance G_{ab} between these same two sites are simply related by $K_{ab} = 2M/G_{ab}$ [2, 5]. Here M is the number

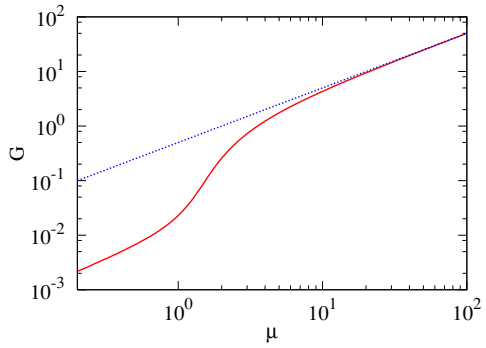

Figure 4. Average two-point conductance (thick solid curve) on a random graph with $N = 100$ sites. The dotted line corresponds to $G = \mu/2$, the asymptotic large-μ form for the conductance.

of bonds in the cluster that contains both a and b. Equivalently, the mean commute rate $\mathcal{R}_{ab} = 1/K_{ab}$ is given by

$$\mathcal{R}_{ab} = \frac{G_{ab}}{2M}. \tag{10}$$

As we shall see, it is much easier to estimate the conductance rather than the commute rate of a random graph by direct means. We will then rely on this connection between G_{ab} and $\mathcal{R}_{ab}$ to determine the latter quantity.

For reasons of numerical convenience, we will often consider the following sum of the rates:

$$\mathcal{R}_a \equiv 2\sum_{b\neq a} \mathcal{R}_{ab} = \frac{1}{M}\sum_b G_{ab}. \tag{11}$$

We include the factor of 2 in the definition because $\mathcal{R}_a$ then equals 1 for the complete graph. We may also sum freely over all sites b in the system equation (11) because $G_{ab} = 0$ for any sites that are not in the same cluster as a. Finally, we obtain the average commute rate for the graph by averaging over all initial sites a:

$$\mathcal{R} \equiv \frac{1}{N}\sum_a \mathcal{R}_a. \tag{12}$$

In the limit of large μ all the sites in the graph belong to the same cluster and $M = \mu N/2$. Thus the average commute rate becomes

$$\mathcal{R} = \frac{1}{N}\frac{2}{\mu N}\sum_{ab} G_{ab} = \frac{2}{\mu}G, \tag{13}$$

where G is the two-point conductance averaged over all pairs of graph endpoints. Thus, as we have discussed, first-passage times and two-point conductances are intimately connected.

For the conductance itself, it is worth noting that this function behaves anomalously near the connectivity transition. Although the conductance must increase monotonically with μ [19], the rate of increase changes for μ in the critical range between 1 and $\ln N$ (figure 4). For large μ, the conductance asymptotically approaches $G = \mu/2$ (dashed line), a result that corresponds to the average commute rate approaching $\mathcal{R} = 1$, in agreement with the result of figure 2.

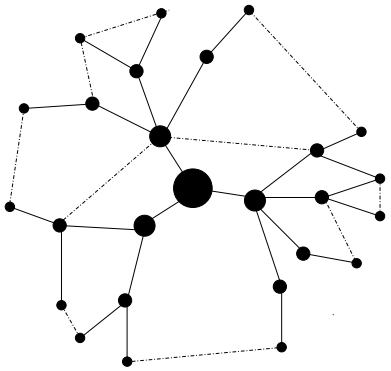

Figure 5. A rooted geodesic tree (RGT). Disks represent sites. The largest black disk is the root. Disks in successive shells have successively smaller radii. The solid lines are links in the RGT. The broken lines are bonds that are added subsequently to generate a random graph cluster.

5. Structure of the random graph

To determine the behaviour of the commute rate for general values of μ, we first need to resolve the structure of random graphs at these values of μ. From this structural information, it is relatively easy to determine the conductance, from which we may then infer the behaviour of the commute rate.

An advantage of formulating the average commute rate as in equation (11) is that only sites that are connected to the starting point a contribute to $\mathcal{R}_a$. Thus we can restrict the endpoint b to lie in the cluster that also contains a. Then averaging over many realizations of these clusters is equivalent to averaging over a in equation (12).

5.1. Rooted geodesic tree

To generate a cluster within the random graph that contains the starting site a of the random walk, we first construct a subset of the cluster that we term the *rooted geodesic tree* (RGT), as shown in figure 5. We can then build the rest of a random graph cluster from the RGT. It is worth noting that the RGT is similar to classical Galton–Watson branching trees [20], and recent work has studied the properties of random walks on such structures [21].

The RGT is a specific subset of a random graph cluster that: (i) spans all the sites in the cluster, and (ii) the distance between a and any site b on the RGT is also the shortest distance between these two sites in the random graph cluster. The notion of the RGT is inspired, in part, by the minimal spanning tree, a construction with useful applications in network flow problems [22][4].

To construct the RGT, we start with N sites and no bonds, and assign one site (denoted by a) as the root; this site is defined to be at level $j = 0$. We generate the RGT as a series of successive shells centred about a. The jth shell, denoted by $\mathbf{S}_j$, contains those sites that can be reached from a after exactly j hops between connected neighbours. We define S_j as the expected number of sites in the jth shell.

Suppose that we have just generated the jth shell. An unassigned site y becomes part of shell $\mathbf{S}_{j+1}$ if a link is created that joins y to an arbitrary site x in $\mathbf{S}_j$. For each unassigned site, at least one such link will be created with probability $1 - (1 - p)^{S_j}$. We then test each

[4] See also section 5.2 of [3] for a construction related to the RGT.

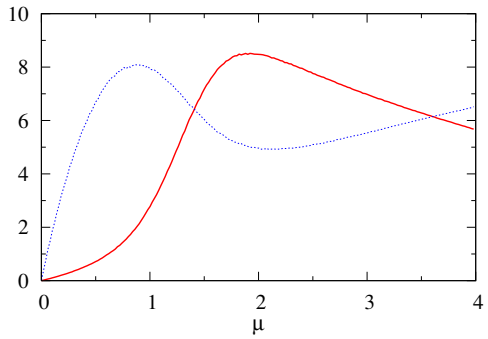

Figure 6. Average RGT radius (solid curve) as a function of the average degree μ. The dotted curve is the average commute rate on the same structure; this rate has been scaled to lie on a similar vertical range. Data are based on 10^5 realizations of RGTs of 100 sites.

possible link between y and the sites in $\mathbf{S}_j$ one by one. Each such link will be created with probability p. When the first such link is created, y becomes an element of $\mathbf{S}_{j+1}$ and we then consider the next unassigned site for potential inclusion in $\mathbf{S}_{j+1}$. If none of the possible links is created, y remains available for inclusion in subsequent shells. After all the unassigned sites have been tested, the shell $\mathbf{S}_{j+1}$ is complete. This growth process continues until either no unassigned sites remain or if all attempts to incorporate the available sites into the current shell fail. In the latter case, the total number of sites in the RGT is less than N.

There are two important subtleties associated with this construction algorithm for an RGT. First, bonds that are not examined in the initial construction of the RGT can only exist between sites in the same or in adjacent shells of the RGT. A second important point is that in building the RGT, each examined bond was tested one time only and is therefore included in the RGT with probability p.

Since the number of sites in successive shells of the RGT grows exponentially in the number of steps away from the root, the radius of the largest cluster (giant component), is given by the criterion $\mu^L \approx \beta N$. Here $\beta = \beta(\mu)$ is the fraction of the initial N sites that belongs to the largest cluster. By the definition of the RGT, the fraction of sites in the largest cluster in the random graph and in the underlying RGT are identical. Thus the radius of the RGT is given by

$$L \sim \frac{\ln N}{\ln \mu} + \frac{\ln \beta}{\ln \mu}. \tag{14}$$

Since β is a rapidly growing function of μ [2, 3], the radius of the giant component of the RGT is an increasing function of μ just above the percolation threshold $\mu_c = 1$. This increase in radius occurs because the RGT acquires progressively more sites with increasing μ. On the other hand for sufficiently large μ, the giant component will contain almost all sites in the graph and the radius of this component will decrease as μ is increased still further (figure 6). This non-monotonic behaviour of the RGT radius on μ is ultimately connected to the non-monotonicity in the commute rate.

5.2. Role of loops on graph structure

Given a rooted geodesic tree, it is possible to augment the tree to generate a realization of a random graph cluster. We merely attempt to add to the RGT each of the bonds between sites on the cluster that were not previously considered in the construction of the RGT itself.

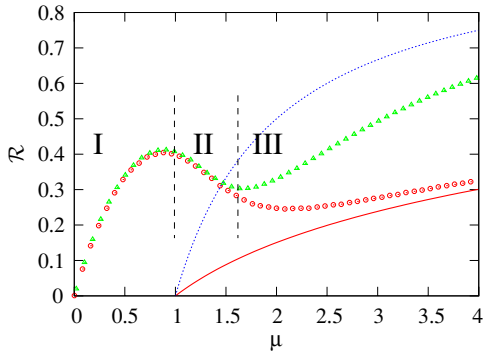

Figure 7. Commute rates on the RGT (○) and on random graphs (△) for $N = 100$ sites based on averages over 10^5 realizations. The lower curve is our RGT prediction $\mathcal{R}_a = \ln\mu/\ln N$ for the large-μ limit. The upper curve in the prediction for the random graph $\mathcal{R} = (\mu - 1)/\mu$ (equation (17)). The approximate locations of regimes I, II and III are indicated.

Each such bond addition attempt is carried out with probability p. As a result of the fact that each of these newly-added bonds and each bond in the RGT is present with probability p, all bonds in the full random graph cluster are present with probability p. Furthermore, both the RGT and the corresponding random graph have the same number of sites and radius for the same values of μ and N. As a result of this equivalence, the RGT undergoes the same percolation transition as the random graph itself when μ passes through 1.

6. The commute rate

We now use the connection between commute rate and conductance to understand the non-monotonicity in the commute rate for a random walk on a random graph. As indicated in figure 7, there are three regimes for the commute rate: (I) an initial increase with μ for small μ; (II) a decrease over an intermediate range; and (III) an ultimate increase for large μ. For regimes I and II, the commute rates on the RGT and the random graph are nearly identical and it is simpler to consider the commute rate on the RGT. We then investigate how adding the links to the RGT to create a random graph affects the commute rate.

6.1. Commute rate on the RGT

For a tree graph, the resistance between two sites is simply the path length between these two sites. Thus the average commute rate in equation (11) has the form

$$\mathcal{R}_a = \frac{1}{V-1}\sum_{b\neq a}\frac{1}{\mathcal{D}_{ab}} = \frac{1}{V-1}\sum_{j=1}^{L}\frac{S_j}{j}. \tag{15}$$

Here the number of links in a tree is one less than the total number of sites V, and $\mathcal{D}_{ab}$ is the distance between a and b. Thus $\mathcal{R}_a$ is the inverse moment of the distance between the root a and all other sites in the tree. The second equality follows from the shell structure of the RGT, where L is the radius of the tree. Thus we need only the statistics of the shell sizes of the RGT to determine the commute rate.

Since each realization of the RGT is distinct, the number of sites V, the radius L, and the shell sizes S_j fluctuate from realization to realization. To calculate the configuration-averaged commute rate $\langle\mathcal{R}_a\rangle$, we first use the algorithm of the previous section to generate RGTs.

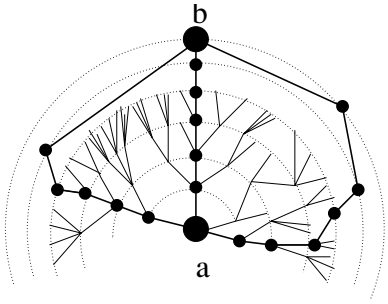

Figure 8. Schematic representation of the random graph. The included RGT is also shown. Loops typically arise at a distance $L = \ln N/\ln \mu$ from the root. A site in this last shell will typically have μ independent paths to the root.

Then we solve the random walk problem on each realization and average equation (15) over realizations to determine the commute rate (figure 7).

To understand the non-monotonicity of the commute rate for the RGT, consider first the small-μ limit. Because isolated sites contribute zero to the average rate, the commute rate must initially increase with μ, as small trees begin to form. Once most sites are no longer isolated, the radii of typical RGTs then increase with μ due to the merging of small trees. This increase in radius causes a decrease in the commute rate, as can be seen by writing the rate in equation (15) as

$$\mathcal{R}_a(L) = \sum_{j=1}^{L} \frac{S_j}{j} \Bigg/ \sum_{j=1}^{L} S_j. \tag{16}$$

In the analogous expression for $\mathcal{R}_a(L+1)$, the numerator increases by $S_{L+1}/(L+1)$ while the denominator increases by S_{L+1}. Thus $\mathcal{R}_a(L)$ is a decreasing function of L, so that a tree with a larger radius will have a smaller commute rate.

As argued in section 5, a further increase in μ will cause the radius of the RGT to eventually decrease with μ. Correspondingly, the commute rate enters regime III and increases with μ. In this regime, we now use the fact that the number of sites in successive shells of the RGT grows exponentially in the distance from the root. Thus the shell at radius L contains almost all of the sites of the RGT. As a naive approximation, we then replace the sum in equation (15) by the last term to give, in the limit of large μ,

$$(\mathcal{R}_a)_{\mathrm{RGT}} \approx \frac{1}{V}\frac{V}{L} \sim \frac{\ln \mu}{\ln N}.$$

This result agrees extremely well with numerical results for the commute rate on RGTs, as shown in figure 7.

6.2. Role of loops on commute rate

We now investigate how adding loops to the RGT to build a random graph affects the behaviour of the mean commute rate. Starting with a realization of an RGT we generate a cluster of the random graph by adding missing bonds, following the procedure discussed in section 5. The addition of these bonds will create loops that provide alternative paths between the root site and the endpoints of a random walk (figure 8). The ostensible effect of these additional paths is to increase the commute rate between the root and any endpoint.

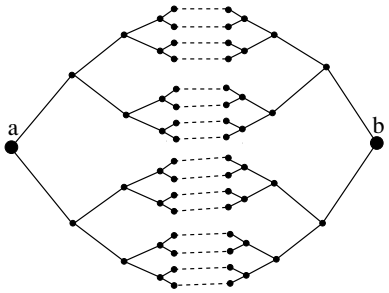

Figure 9. Schematic random graph structure to calculate the conductivity between two sites a and b for mean degree $\mu > 1$. An RGT is grown around both a and b. The two RGTs meet at a distance $O(\ln N/\ln \mu)$ from each of a and b. Broken lines are links between sites in the outermost shells of the two respective RGTs.

To estimate the two-point conductance for a random graph for general $\mu > \ln N$, we start with the picture that the graph consists of two RGTs, one emanating from a and the other from b as depicted in figure 9 (a similar but rigorous argument was given in [23]). For a graph of N sites, the radius of each tree is of order $L \sim \ln(N/2)/\ln \mu$. We argue that these two trees tend to join only at the outermost shell because this is where most of the sites in the trees are located. We further assume that, in the equivalent resistor network, all sites at the same distance from the root are at the same potential. Thus the conductance of the two joining RGTs is simply one-half of the conductance between the root and the last shell of a single RGT.

For this last step, we approximate the RGT by an infinite Cayley tree with branching ratio μ. The resistance between the kth and the $(k+1)$st shell in this tree is $\mu^{-(k+1)}$, since the links between the two shells are in parallel. Because the shells are in series, the resistance from the center to infinity is simply the geometric sum, $\sum_{k=0}^{\infty} \mu^{-(k+1)} = \frac{1}{\mu-1}$. Thus the conductance between a and b is $G_{ab} = (\mu - 1)/2$. Substituting this result in equation (13), then gives the commute rate

$$\mathcal{R} = \frac{\mu - 1}{\mu}. \tag{17}$$

This result converges to 1 as $\mu \to \infty$, in agreement with the effective medium approach in section 2 as well as our simulation results. Closer to the percolation threshold, however, equation (17) and simulation results quantitatively disagree because our naive picture for the structure of the random graph no longer applies.

Thus we observe that the eventual increase in the commute rate (regime III) stems from the combined effect of the decrease in the radius of the underlying RGT embedded within a random graph cluster and the emergence of loops that join two RGTs in the random graph.

7. Conclusion

We studied a basic first-passage characteristic of random walks on random graphs that is related to the time for a walk to travel between two arbitrary points on a graph. We first constructed an effective medium theory and a small dilution approximation for this mean transit time. The former approach predicted that the mean transit time, and also all positive integer moments of the transit time, are independent of the bond concentration p for p greater than the connectivity threshold $p_1 = \ln N/N$. The small dilution approximation also predicts

a slow dependence of the transit time on p near $p = 1$. Our numerical simulation results are in qualitative accord with a transit time that is slowly varying in p for $p > p_1$.

Below the connectivity threshold, the transit time is not well defined because the mean time for a random walk to hop between sites on different components of a disconnected graph is infinite. To avoid this pathology, we studied the inverse of the commute time, namely, the commute rate. We developed a simple heuristic picture for the behaviour of the commute rate that relied on first identifying an embedded rooted geodesic tree (RGT) within an arbitrary random graph cluster. For the RGT, it is simple to compute the commute rate in terms of a geometric picture for the tree and thus argue that this rate is a non-monotonic function of p in the critical regime.

We then presented a simple physical picture for the influence of loops on the behaviour of the commute rate. Qualitatively, the dependence of the radius of the underlying RGT on the bond concentration explains the behaviour of the commute rate close to the percolation threshold. For larger μ loops become an important factor and are ultimately responsible for the non-monotonic dependence of the commute rate on p. While our arguments were heuristic and the approximations made are uncontrolled, they provide an intuitive picture for the structure of random graphs and also provide qualitative and satisfying agreement with simulation results for the commute rate.

Acknowledgments

We thank Paul Krapivsky and Federico Vazquez for helpful suggestions. VS and SR gratefully acknowledge financial support from NSF grants DMR0227670 (at BU) and DOE grant W-7405-ENG-36 (at LANL). DbA similarly acknowledges NSF grant PHY0140094 (DbA) for financial support.

References

[1] Erdős P and Rényi A 1959 *Publ. Math. Debrecen* **6** 290
[2] Bollobás B 1998 *Modern Graph Theory* (New York: Springer)
[3] Janson S, Łuczak T and Ruciński A 2000 *Random Graphs* (New York: Wiley)
[4] Hughes B D 1996 *Random Walks and Random Environments Volume 2: Random Environments* (Oxford: Clarendon Press)
[5] Lovász L 1996 *Combinatorics, Paul Erdős is Eighty* vol 2 ed D Miklós, V T Sós and T Zsőnyi (Budapest: János Bolyai Mathematical Society) pp 353–98
[6] Woess W 2000 *Random Walks on Infinite Graphs and Groups* (Cambridge: Cambridge University Press)
[7] Aldous D and Fill J A *Reversible Markov Chains and Random Walks on Graphs* Online at http://stat-www.berkeley.edu/users/aldous/RWG/book.html
[8] ben-Avraham D and Havlin S 2000 *Diffusion and Reactions in Fractals and Random Systems* (Cambridge: Cambridge University Press)
[9] Almaas E, Kulkarni R V and Stroud D 2003 *Phys. Rev.* E **68** 056105
[10] Noh J D and Rieger H 2004 *Phys. Rev. Lett.* **92** 118701
[11] Bray A J and Rodgers G J 1988 *Phys. Rev.* B **38** 11461
[12] Monasson R 1999 *Eur. Phys. J.* B **12** 555
[13] Jonasson J 1998 *Combin. Probab. Comput.* **7** 265
[14] Cooper C and Frieze A 2003 *Proc. 14th Annual ACM- SIAM Symposium on Discrete Algorithms* vol 140–147 (*SIAM, 2003*)
[15] Masuda N and Konno N 2004 *Phys. Rev.* E **69** 066113
[16] For general presentations of effective medium theory, see e.g., Kirkpatrick S 1973 *Rev. Mod. Phys.* **45** 574 Koplik J 1981 *J. Phys. C: Solid State Phys.* **14** 4821
[17] van Kampen N G 1997 *Stochastic Processes in Physics and Chemistry* 2nd edn (Amsterdam: North-Holland)
[18] Redner S 2001 *A Guide to First-Passage Processes* (New York: Cambridge University Press)

[19] Doyle P G and Snell J L 1984 *Random Walks and Electric Network (Caruso Mathematical Monographs* vol 22*)* (Oberlin, OH: Mathematical Association of America)
Doyle P G and Snell J L 2000 *Preprint* math.PR/0001057
[20] Harris T E 1963 *The Theory of Branching Processes* (Englewood Cliffs, NJ: Prentice-Hall)
[21] Lyons R, Pemantle R and Peres Y 1995 *Ergodic Theory Dynam. Syst.* **15** 593 See also http://mypage.iu.edu/rdlyons/prbtree/prbtree.html for a more comprehensive discussion of diffusion on trees and related structures
[22] See Ahuja R K, Magnanti T L and Orlin J B 1993 *Network Flows: Theory, Algorithms and Applications* (Englewood Cliffs, NJ: Prentice-Hall)
[23] Grimmett G and Kesten H 1984 *J. London Math. Soc.* **30** 171
Grimmett G and Kesten H 2001 *Preprint* math.PR/0107068