A MATRIX PERTURBATION VIEW OF
THE SMALL WORLD PHENOMENON*

DESMOND J. HIGHAM[†]

Abstract. We use techniques from applied matrix analysis to study small world cutoff in a Markov chain. Our model consists of a periodic random walk plus uniform jumps. This has a direct interpretation as a teleporting random walk, of the type used by search engines to locate web pages, on a simple ring network. More loosely, the model may be regarded as an analogue of the original small world network of Watts and Strogatz [*Nature*, 393 (1998), pp. 440–442]. We measure the small world property by expressing the mean hitting time, averaged over all states, in terms of the expected number of shortcuts per random walk. This average mean hitting time is equivalent to the expected number of steps between a pair of states chosen uniformly at random. The analysis involves nonstandard matrix perturbation theory and the results come with rigorous and sharp asymptotic error estimates. Although developed in a different context, the resulting cutoff diagram agrees closely with that arising from the mean-field network theory of Newman, Moore, and Watts [*Phys. Rev. Lett.*, 84 (2000), pp. 3201–3204].

Key words. Google, Markov chain, matrix perturbation, mean hitting time, optional sampling theorem, partially random graph, random walk, Sherman–Morrison formula, teleporting, web search engine

AMS subject classifications. 65F35, 65C40, 60J27

DOI. 10.1137/S0895479802406142

1. Introduction. We show here that a small world cutoff arises in a simple random walk setting that is amenable to rigorous analysis via matrix perturbation theory. Our model is derived by adding uniform jumps to a periodic, one-dimensional random walk. Increasing the jump probability allows us to interpolate between completely local and completely global behavior. The small world property is then quantified by the average or maximum of the mean hitting times.

Although it is simplistic, we believe that this model is relevant to many physical, sociological, epidemiological, and computational applications, as it combines the traditional notion of diffusion on a lattice [3, 4, 16, 20] with the type of partially random connectivity that has recently been used to describe complex, real-life networks [6, 12, 15, 17, 18, 22, 23, 24]. In particular, we mention that the original work of Watts and Strogatz [25] included a disease simulation that is in a similar spirit to our model.

More specifically, the idea of taking a "random walk plus shortcuts" is used by web search engines. Here, the fundamental task is to locate all web pages by following hyperlinks. A simple random walk—finding all links out of the current page and choosing one of them uniformly—is liable to reach a dead end or to cycle. To avoid this, it is common to jump occasionally to a page chosen uniformly at random. Adding jumps in this way is known as *teleporting* [10]. The search engine Google uses just such an algorithm [14]. Our results apply directly to the case of teleporting on a ring lattice and quantify, in terms of the teleporting parameter, the expected number of links that must be followed to reach a given target.

†Department of Mathematics, University of Strathclyde, Glasgow G1 1XH, UK (djh@maths.strath.ac.uk).

Although the correspondence is not exact, we originally developed this model by analogy with the randomized network approach of Watts and Strogatz [25]. In that work, the authors showed experimentally that by replacing a small number of connections by new connections between randomly chosen nodes, that is, by *randomly rewiring* a few times, the small world property is roused before the clustering property is lost. They coined the phrase *small world phenomenon* to describe the unlikely alliance of the clustering and small world properties. Initially, the small world property was verified through numerical simulation. More recently, Newman, Moore, and Watts [19] gave a semiheuristic analysis of a closely related model in the limit of large network size. Here random links do not replace existing links but instead are added to the network and are thus referred to as shortcuts. The resulting mean-field expression for the path length is shown as a dashed line in Figure 3.3 below. An unsatisfactory feature of the treatment in [19] is that it is designed to be valid for either a large or small number of shortcuts, that is, $x \ll 1$ or $x \gg 1$ on the x-axis in Figure 3.3. This does not cover the interesting cutoff region where the average path length sharply decreases as a function of the average number of shortcuts added. However, simulations reported in [19] showed that the mean-field expression continues to give a reasonable fit in this range. A fully rigorous analysis that applies only for a large number of shortcuts ($x \gg 1$) has been given in [1].

For our random walk model, we measure the small world property as the maximum or average mean hitting time, rather than the expected path length. One of the key advantages of this approach is that it permits a rigorous analysis in the asymptotic limit of a large number of states. Further, the analysis is sharp; we obtain exact expressions for the leading terms in the expansions. Our results include what appears to be the first rigorous analysis of a small world cutoff effect for the interesting $O(1)$ shortcuts regime. Quite remarkably, the analytical cutoff diagram that we derive is in close agreement with the one that has been found experimentally for the Watts–Strogatz network model.

In the next section we set up the random walk as a Markov chain and state results about the mean hitting times. The results are interpreted in section 3. We show that a certain scaling of the interpolation parameter (in terms of the chain length) has a particular physical significance. For this scaling, we obtain a cutoff diagram that illustrates the small world phenomenon and may be compared with that of the Watts–Strogatz network model. Section 4 is the heart of the paper. Here we prove the key results using techniques from numerical analysis to capture the effect of a certain structured perturbation on a linear algebraic system. Because the perturbation depends on the dimension of the system, the usual "$(I + E)^{-1} = I - E + O(\|E\|^2)$" expansion cannot be employed in general. Section 5 points to possible future work.

2. The Markov chain approach.

2.1. The model. We begin by setting up the relevant mathematical concepts. A discrete time, finite state Markov chain is a stochastic process $\{X_n\}_{n \geq 0}$ that can be characterized by a *transition matrix* P. We suppose that there are N states, labeled 1 to N, so $P \in \mathbb{R}^{N \times N}$. The value p_{ij} specifies the probability that $X_{n+1} = j$ given that $X_n = i$, that is,

$$\mathbb{P}(X_{n+1} = j | X_n = i) = p_{ij},$$

with all $p_{ij} \geq 0$ and $\sum_{j=1}^{N} p_{ij} = 1$. We will always make the process start at state 1, so $X_0 = 1$ with probability 1. The *mean hitting time* for state j is the average

number of steps taken by the process before first reaching state j. More precisely, the mean hitting time for state j is the expected value of the random variable $h^j(\omega) := \inf\{n \geq 0 : X_n(\omega) = j\}$. We let $\mathbf{z} \in \mathbb{R}^{N-1}$ denote the vector of mean hitting times for states $2, 3, \ldots, N$, so z_j is the mean hitting time for state $j+1$. A standard result that may be found, for example, in [20, Theorem 1.3.5] shows that $\mathbf{z}$ is the minimal nonnegative solution to the system of linear equations

$$(2.1) \qquad \left(I - \widehat{P}\right)\mathbf{z} = \mathbf{e}.$$

Here, $\widehat{P} \in \mathbb{R}^{(N-1)\times(N-1)}$ is formed by removing the first row and column from P, so $\widehat{p}_{ij} = p_{i+1,j+1}$, and $\mathbf{e} := [1, 1, \ldots, 1]^T \in \mathbb{R}^{N-1}$. We find it natural to use the mean hitting time as an analogue of the path length in order to measure the "small world" size of the Markov chain. We will consider the *maximum mean hitting time*

$$(2.2) \qquad \mathrm{mht}_{\max}(P) := \max_{1 \leq i \leq N-1} z_i$$

and the *average mean hitting time*

$$(2.3) \qquad \mathrm{mht}_{\mathrm{ave}}(P) := \frac{1}{N-1} \sum_{i=1}^{N-1} z_i.$$

We note that $\mathrm{mht}_{\mathrm{ave}}(P)$ has the agreeable interpretation as the expected number of steps between a pair of sates chosen uniformly at random. There are, of course, other hitting time measures, such as the expected value of $\max_{1 \leq j \leq N} h^j(\omega)$, that may be of interest. We focus on (2.2) and (2.3) because we believe them to be natural choices and because they can be studied via matrix analysis.

By analogy with the basic ring network in [25], we consider the Markov chain with transition matrix

$$(2.4) \qquad P_0 = \begin{bmatrix} 0 & \frac{1}{2} & & & & & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & \\ & \frac{1}{2} & 0 & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \frac{1}{2} \\ \frac{1}{2} & & & & \frac{1}{2} & 0 \end{bmatrix} \in \mathbb{R}^{N \times N}.$$

Here, at each step the process moves to either of the two neighboring states with equal probability (with 1 and N regarded as neighbors). This could also be described as a symmetric, one-dimensional, periodic random walk. With this choice of transition matrix, the system (2.1) becomes

$$(2.5) \qquad T\mathbf{z} = \mathbf{e},$$

where $T := \mathrm{tridiag}(-\frac{1}{2}, 1, -\frac{1}{2})$ and $\mathrm{tridiag}(a, b, c)$ denotes a tridiagonal Toeplitz matrix of the form

$$\begin{bmatrix} b & c & & & \\ a & b & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & c \\ & & & a & b \end{bmatrix} \in \mathbb{R}^{(N-1)\times(N-1)}.$$

It is well known, and easily verified, that (2.5) has the unique solution $z_i = i(N - i)$. Hence,

$$(2.6) \qquad \mathrm{mht}_{\max}(P_0) := \frac{N^2}{4} + O(1) \quad \text{and} \quad \mathrm{mht}_{\mathrm{ave}}(P_0) := \frac{N(N + 1)}{6}.$$

Now we perturb the basic transition matrix P_0 in (2.4) by resetting all zero entries to $\epsilon > 0$. This gives the transition matrix

$$(2.7) \qquad P_\epsilon = \begin{bmatrix} \epsilon & \frac{1}{2} - \widehat{\epsilon} & \epsilon & \cdots & \epsilon & \frac{1}{2} - \widehat{\epsilon} \\ \frac{1}{2} - \widehat{\epsilon} & \epsilon & \frac{1}{2} - \widehat{\epsilon} & \epsilon & \cdots & \epsilon \\ \epsilon & \frac{1}{2} - \widehat{\epsilon} & \epsilon & \frac{1}{2} - \widehat{\epsilon} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \epsilon \\ \epsilon & \cdots & \epsilon & \frac{1}{2} - \widehat{\epsilon} & \epsilon & \frac{1}{2} - \widehat{\epsilon} \\ \frac{1}{2} - \widehat{\epsilon} & \epsilon & \cdots & \epsilon & \frac{1}{2} - \widehat{\epsilon} & \epsilon \end{bmatrix} \in \mathbb{R}^{N \times N},$$

where in order to keep the row sums equal to one we require

$$(2.8) \qquad \widehat{\epsilon} = \frac{N - 2}{2}\epsilon,$$

and in order to maintain nonnegative entries we require

$$(2.9) \qquad \epsilon \le \frac{1}{N - 2}.$$

On each step the Markov chain with transition matrix P_ϵ moves to either of the two neighboring states with equal probability $\frac{1}{2} - \frac{N-2}{2}\epsilon$ and to each nonneighboring state with probability ϵ. This is precisely the teleporting idea described in section 1 applied to a ring network and, more loosely, is analogous to the rewiring operation used to generate small world networks. We may regard ϵ as a parameter that allows us to interpolate between a local and a global process.

The main issue that we address in this work is how the mean hitting times are reduced as ϵ is increased from zero. This leads to an interesting problem in matrix perturbation theory. We will compute expressions for the *maximum mean hitting time reduction ratio*

$$(2.10) \qquad \frac{\mathrm{mht}_{\max}(P_\epsilon)}{\mathrm{mht}_{\max}(P_0)}$$

and the *average mean hitting time reduction ratio*

$$(2.11) \qquad \frac{\mathrm{mht}_{\mathrm{ave}}(P_\epsilon)}{\mathrm{mht}_{\mathrm{ave}}(P_0)}$$

for small ϵ and large N. The constraint (2.9) shows that ϵ must scale with N, and hence we consider the power law relationship

$$(2.12) \qquad \epsilon := \frac{K}{N^\alpha} \quad \text{for fixed } K > 0 \text{ and } \alpha > 1 \quad \text{in the limit } N \to \infty.$$

For reference, note that in the case where $\epsilon = 1/N$, all entries of P_ϵ are equal. This is the fully global regime where the process moves to any other state with equal probability. In this case, it follows from (2.1) (or from basic probabilistic arguments) that the mean hitting time vector has all entries $z_j = N$. Hence, from (2.6),

$$\frac{\mathrm{mht}_{\max}(P_{1/N})}{\mathrm{mht}_{\max}(P_0)} = \frac{4}{N} + O(N^{-3}) \quad \text{and} \quad \frac{\mathrm{mht}_{\mathrm{ave}}(P_{1/N})}{\mathrm{mht}_{\mathrm{ave}}(P_0)} = \frac{6}{N + 1}.$$

2.2. Results. Theorems 2.1, 2.2, and 2.3 below completely characterize the reduction ratios for all ϵ in (2.12), to leading order in N. Proofs are given in section 4.

THEOREM 2.1. *For $\alpha > 3$,*

$$\frac{\mathrm{mht_{max}}(P_\epsilon)}{\mathrm{mht_{max}}(P_0)} = 1 + O(N^{3-\alpha}) \quad \text{and} \quad \frac{\mathrm{mht_{ave}}(P_\epsilon)}{\mathrm{mht_{ave}}(P_0)} = 1 + O(N^{3-\alpha}).$$

THEOREM 2.2. *For $\alpha = 3$,*

$$(2.13) \qquad \frac{\mathrm{mht_{max}}(P_\epsilon)}{\mathrm{mht_{max}}(P_0)} = \frac{2\sqrt{2}}{\sqrt{K}} \tanh \frac{\sqrt{K}}{2\sqrt{2}} + O(N^{-1})$$

and

$$(2.14) \qquad \frac{\mathrm{mht_{ave}}(P_\epsilon)}{\mathrm{mht_{ave}}(P_0)} = \frac{6}{K} \left(\frac{\sqrt{2K}}{2 \tanh \frac{\sqrt{2K}}{2}} - 1 \right) + O(N^{-1}).$$

THEOREM 2.3. *For $1 < \alpha < 3$,*

$$(2.15) \qquad \frac{\mathrm{mht_{max}}(P_\epsilon)}{\mathrm{mht_{max}}(P_0)} = \frac{2\sqrt{2}}{\sqrt{K}} N^{\frac{\alpha-3}{2}} + O(N^{-1})$$

and

$$(2.16) \qquad \frac{\mathrm{mht_{ave}}(P_\epsilon)}{\mathrm{mht_{ave}}(P_0)} = \frac{3\sqrt{2}}{\sqrt{K}} N^{\frac{\alpha-3}{2}} - \frac{6}{K} N^{\alpha-3} + O(N^{-1}).$$

(We remark that the second term on the right-hand side of (2.16) can be absorbed into the final $O(N^{-1})$ term for $\alpha \leq 2$.)

3. Interpretation and discussion. The theorems show that there is a threshold at $\alpha = 3$. For larger α values, the ϵ perturbation has no effect on the mean hitting time reduction ratios in the $N \to \infty$ limit. For $\alpha = 3$, the reduction ratio has a fixed, nonzero value for each K. For α below 3, the ϵ perturbation dominates the process, giving a reduction ratio that is asymptotically zero.

In the case of networks, the small world phenomenon has been characterized by expressing some measure of the average path length in terms of the expected number of shortcuts added [19, 25]. An appropriate characterization in our Markov chain setting is to measure the average mean hitting time, $\mathrm{mht_{ave}}(P_\epsilon)$, as a function of the expected number of shortcuts (teleportings) *taken* per random walk. (We say that a shortcut takes place from step n to step $n+1$ if $X_{n+1} \neq (X_n \pm 1) \mod N$.) Now, on each step, the probability of a shortcut is $\epsilon(N-2)$. Define the process M_n by

$$(3.1) \qquad M_n := (\text{number of shortcuts up to step } n) - n\epsilon(N-2).$$

Subtracting the drift in this way produces a martingale, that is, $\mathbb{E}M_n = 0$. Since h^j is a stopping time, the optional sampling theorem [11, Chapter 3, Corollary 3.1] may be applied to give $\mathbb{E}M_{h^j} = \mathbb{E}M_0 = 0$. Using this in (3.1), we find that the expected number of shortcuts up to the hitting time for state j is given by $\epsilon(N-2)\mathbb{E}h^j$. So if we let W_ϵ denote the average over all states of the expected number of shortcuts taken per random walk, then $W_\epsilon = \epsilon(N-2)\mathrm{mht_{ave}}(P_\epsilon)$. Applying Theorems 2.1–2.3 leads immediately to the following corollary.
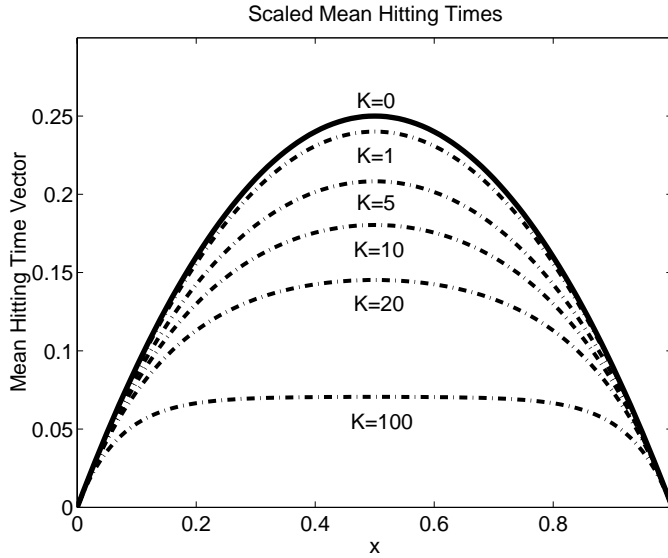
Scaled Mean Hitting Times

FIG. 3.1. *Curves describing the vector of scaled mean hitting times with $\alpha = 3$ for different values of K in (2.12).*

COROLLARY 3.1. *The average over all states of the expected number of shortcuts taken per random walk, W_ϵ, has the following properties:*
  1. *For $\alpha > 3$, $W_\epsilon \to 0$ as $N \to \infty$.*
  2. *For $\alpha = 3$,*

$$W_\epsilon = \frac{\sqrt{2K}}{2 \tanh \frac{\sqrt{2K}}{2}} - 1 + O(N^{-1}).$$

  3. *For $1 < \alpha < 3$, $W_\epsilon \to \infty$ as $N \to \infty$.*

Corollary 3.1 distinguishes $\alpha = 3$ as the appropriate regime in which to search for the small world phenomenon—it is only in this case that the ϵ perturbation introduces a nonzero but bounded number of shortcuts. So henceforth we consider only the case $\alpha = 3$. Note that this scaling is easily arrived at via the following heuristic arguments. Typical excursions on the basic ring take $O(N^2)$ steps. For the P_ϵ model, the probability of a shortcut on each step is $(N-2)\epsilon = O(N^{1-\alpha})$. Hence, if the $O(N^2)$ excursion length is preserved and a finite number of shortcuts are to be taken, then a reasonable guess is to set $O(N^2) \times O(N^{1-\alpha}) = O(1)$, giving $\alpha = 3$. However, since our analysis provides the coefficients associated with the leading order asymptotics, we are able to investigate the model more closely.

Returning our attention to the individual mean hitting times, for $\alpha = 3$ it follows from the analysis in section 4 (more precisely, from (4.5), (4.11), and (4.15)–(4.17)) that z_j is perturbed to $z_{\epsilon j}$, where

$$(3.2) \quad z_{\epsilon j} = \frac{N^2}{\sqrt{2K} \tanh \frac{\sqrt{2K}}{2}} \left[1 - \frac{\cosh \sqrt{2K}(x_j - \frac{1}{2})}{\cosh \frac{\sqrt{2K}}{2}} \right] + O(N), \quad \text{with } x_j := \frac{j}{N}.$$

In Figure 3.1 we plot curves for the mean hitting time vector, as given by the first term on the right-hand side of (3.2), scaled by N^2. We show the cases $K = 1, 5, 10, 20, 100$.
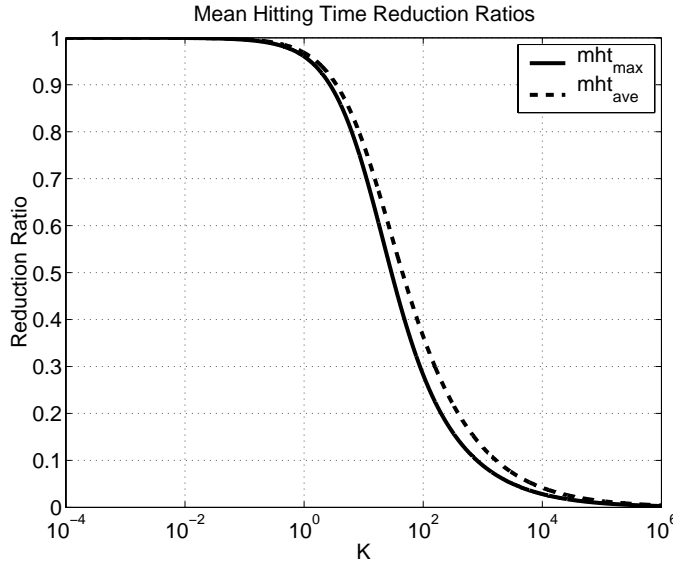
FIG. 3.2. *Leading term in the maximum (solid line) and average (dashed line) mean hitting time reduction ratios as a function of K, from Theorem 2.2.*

(Note that components in the scaled vector $\mathbf{z}_\epsilon/N^2$ are found by evaluating the curves at the equally spaced points $1/N, 2/N, \ldots, (N-1)/N$ along the x-axis.) We have also plotted the $K = 0$ case, that is, $\epsilon = 0$, which corresponds to the parabola $x(1 - x)$. The outcome is intuitively reasonable—the mean hitting times decrease and the profile flattens as K, and hence ϵ, increases.

Separate numerical testing indicated that (3.2) is sharp—the remainder term behaves like a nonzero multiple of N.

Turning now to the mean hitting time reduction ratios, (2.10) and (2.11), Figure 3.2 plots the leading terms in (2.13) and (2.14) as functions of K. (Note that the horizontal axis is logarithmically scaled in order to zoom in on the region of interest.) We see that there is a rapid decay when K is increased beyond ≈ 1.

To look for the small world phenomenon, we now plot the average mean hitting time reduction ratio, $\mathrm{mht}_{\mathrm{ave}}(P_\epsilon)$, as a function of the the average over all states of the expected number of shortcuts taken per random walk, W_ϵ. From Theorem 2.2 and Corollary 3.1, these may be computed via the parametric form

$$(3.3) \qquad W_\epsilon = \frac{\sqrt{2K}}{2\tanh\frac{\sqrt{2K}}{2}} - 1 \quad \text{and} \quad \mathrm{mht}_{\mathrm{ave}}(P_\epsilon) = \frac{6}{K}W_\epsilon,$$

with an error of $O(N^{-1})$. The solid line in Figure 3.3 shows the resulting curve. A sharp cutoff is noticeable as the number of shortcuts increases from around $\frac{1}{2}$ to 50—the small world effect kicks in abruptly when only a small number of shortcuts are taken.

It is possible to compare the behavior of this model with that of the $k = 1$ version of the Newman–Moore–Watts network model [19], which is closely related to the corresponding Watts–Strogatz model [25]. In the network model, we begin with a ring of N nodes, where node i is connected to node j if $|i-j| = 1 \mod N$. This "local" network is interpolated toward the "global" by adding shortcuts between randomly
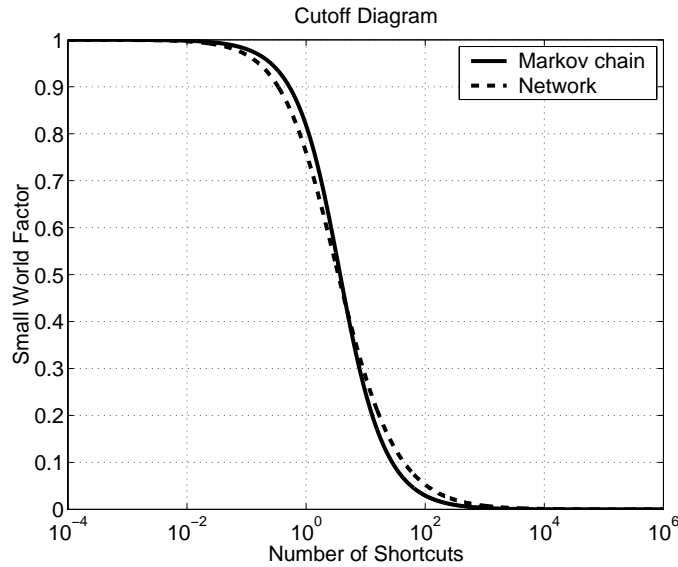
FIG. 3.3. *Solid line is (3.3): x-axis is average number of shortcuts per excursion, W_ϵ, y-axis is reduction ratio for average mean hitting time, $\mathrm{mht}_{\mathrm{ave}}(P_\epsilon)/\mathrm{mht}_{\mathrm{ave}}(P_0)$. Dashed line is, from [19], $y = 4f(x)$ with f defined in (3.4): x-axis is average number of shortcuts per network, y-axis is reduction ratio for average path length.*

selected nodes. The mean-field expression in [19] for the average path length as a function of the expected number of shortcuts is $y = Nf(x)$, where

$$(3.4) \qquad\qquad f(x) = \frac{1}{2\sqrt{x^2 + 2x}} \tanh^{-1} \frac{x}{\sqrt{x^2 + 2x}}.$$

Since the average path length when there are no shortcuts is, to leading order, $N/4$, the curve $y = 4f(x)$ gives the reduction in the average path length in terms of the average number of shortcuts. This is plotted with a dashed line in Figure 3.3. As we mentioned in section 1, the authors note in [19] that their mean-field approximation involves assumptions that are valid only for values that correspond to $x \ll 1$ and $x \gg 1$ on the x-axis of Figure 3.3. However, simulations show that the curve also gives quite an accurate description of the cutoff region around $x = 1$; see [19, Figure 2] or [22, Figure 5].

Overall, although the two measures are fundamentally different, Figure 3.3 shows that there is a remarkable qualitative and quantitative agreement between the small world cutoff behavior in the Markov chain and randomized network models. In particular, the mean-field theory predicts that an average of 3.5 shortcuts per network are needed to give a reduction of $\frac{1}{2}$ in the average path length (consistent with the simulations of [25]). The average number of shortcuts per random walk required to give an average mean hitting time reduction ratio of $\frac{1}{2}$ is 3.7. The corresponding figures for a reduction of $\frac{1}{10}$ are 44 and 28, respectively.

So far, we have focused on measuring mean hitting times by analogy with path length. Is there a corresponding analogue of the clustering property? One possibility is to consider how rapidly the Markov chain converges to its equilibrium distribution. We may regard the chain as *not* being clustered if it tends quickly to equilibrium— that is, transient behavior rapidly gives way to steady state behavior. The rate at

which equilibrium is approached can be bounded above and below in terms of the spectrum of the transition matrix; see, for example, [2, Theorem 10.3]. In our case P_ϵ in (2.7) is circulant, and hence its eigenvalues can be calculated explicitly. For $\alpha = 3$ we find that with N even there is an eigenvalue of modulus $1 - K/N^2$ and with N odd there is a repeated eigenvalue of modulus $1 - (K + \pi^2)/N^2 + O(N^{-4})$. We may conclude that for large N the process is slow to approach equilibrium and hence remains clustered. In this sense, when the cutoff in Figure 3.3 takes place we have captured the small world phenomenon.

4. Proofs.

4.1. Preliminaries. Theorems 2.1–2.3 concern matrix perturbation theory. The vector $\mathbf{z}_\epsilon \in \mathbb{R}^{N-1}$ is the (minimal nonnegative) solution to

$$(4.1) \qquad T_\epsilon \mathbf{z}_\epsilon = \mathbf{e},$$

where $T_\epsilon = T + \mathrm{tridiag}(\frac{N\epsilon}{2}, 0, \frac{N\epsilon}{2}) - \epsilon \mathbf{e}\mathbf{e}^T$. We have $\mathrm{mht}_{\max}(P_\epsilon) = \|\mathbf{z}_\epsilon\|_\infty$ and $\mathrm{mht}_{\mathrm{ave}}(P_\epsilon) = \|\mathbf{z}_\epsilon\|_1/(N-1)$, where $\|\cdot\|_\infty$ and $\|\cdot\|_1$ are used to denote the vector ∞ and 1 norms and their induced matrix norms, respectively. We are thus concerned with the normwise effect on the size of the solution when (2.5) is perturbed to (4.1). For the $\alpha > 3$ case, a standard expansion can be used; see section 4.2. However, for $\alpha \le 3$ this approach is no longer applicable—special care is needed to deal with the dependence of the perturbation on the dimension N; see section 4.3.

We find it useful to let

$$(4.2) \qquad \widehat{T}_\epsilon = T + \Delta T_\epsilon,$$

with

$$(4.3) \qquad \Delta T_\epsilon = \mathrm{tridiag}(\tfrac{N\epsilon}{2}, 0, \tfrac{N\epsilon}{2}).$$

Note that $\widehat{T}_\epsilon$ is diagonally dominant and hence nonsingular. We also let $\mathbf{y}_\epsilon \in \mathbb{R}^{N-1}$ satisfy

$$(4.4) \qquad \widehat{T}_\epsilon \mathbf{y}_\epsilon = \mathbf{e}.$$

Now, we may use the Sherman–Morrison formula [9, p. 490] to deal with the rank one perturbation that converts $\widehat{T}_\epsilon$ to T_ϵ. First note that the inequality $1 - \epsilon \mathbf{e}^T \mathbf{y}_\epsilon \ne 0$ follows from the analysis below. (More precisely, it follows from (4.9) for $\alpha > 3$ and from (4.16) for $1 < \alpha \le 3$.) Hence, by the Sherman–Morrison formula, $\widehat{T}_\epsilon$ in (4.2) is nonsingular and

$$
\begin{aligned}
\mathbf{z}_\epsilon = T_\epsilon^{-1} \mathbf{e} &= \left(\widehat{T}_\epsilon - \epsilon \mathbf{e}\mathbf{e}^T \right)^{-1} \mathbf{e} \\
&= \left(\widehat{T}_\epsilon^{-1} + \frac{\epsilon \widehat{T}_\epsilon^{-1} \mathbf{e}\mathbf{e}^T \widehat{T}_\epsilon^{-1}}{1 - \epsilon \mathbf{e}^T \widehat{T}_\epsilon^{-1} \mathbf{e}} \right) \mathbf{e} \\
(4.5) \qquad\qquad &= \frac{1}{1 - \epsilon \mathbf{e}^T \mathbf{y}_\epsilon} \mathbf{y}_\epsilon.
\end{aligned}
$$

We also note a few more facts. First, recall that A is defined to be a *Stieltjes matrix*, that is, a symmetric M-matrix, if $A^{-1} \ge 0$ and $a_{ij} \le 0$ for $i \ne j$. (Inequalities between vectors or matrices are understood to hold for all components.) Further, any

strictly or irreducibly diagonally dominant symmetric A with $a_{ij} \leq 0$ for $i \neq j$ and $a_{ii} > 0$ for all i is a Stieltjes matrix; see, for example, [21, Theorem 6.2.17]. It follows that T and $\widehat{T}_\epsilon$ are Stieltjes matrices, and hence $T^{-1} \geq 0$ and $\widehat{T}_\epsilon^{-1} \geq 0$. Further,

$$(4.6) \qquad \| T^{-1} \|_1 = \| T^{-1} \|_\infty = \| T^{-1}\mathbf{e} \|_\infty = \frac{N^2}{4} + O(1).$$

4.2. Proof of Theorem 2.1.
Proof. First, note that

$$\| \Delta T_\epsilon \|_\infty = \| \Delta T_\epsilon \|_1 = O(N^{1-\alpha})$$

and hence

$$\| T^{-1}\Delta T_\epsilon \|_\infty = \| T^{-1}\Delta T_\epsilon \|_1 \leq \| T^{-1} \|_1 \| \Delta T_\epsilon \|_1 = O(N^{3-\alpha}).$$

Since $\alpha > 3$ we have $\| T^{-1}\Delta T_\epsilon \|_\infty \to 0$ and $\| T^{-1}\Delta T_\epsilon \|_1 \to 0$. We may thus appeal to standard perturbation theory and expand $(I + T^{-1}\Delta T_\epsilon)^{-1}$ in powers of $T^{-1}\Delta T_\epsilon$; see, for example, [5, Lemma 2.1]. We have

$$
\begin{aligned}
\| \mathbf{y}_\epsilon \|_\infty &= \| (T + \Delta T_\epsilon)^{-1}\mathbf{e} \|_\infty \\
&= \| (I + T^{-1}\Delta T_\epsilon)^{-1}T^{-1}\mathbf{e} \|_\infty \\
&= \| \left[I - T^{-1}\Delta T_\epsilon + O(\| T^{-1}\Delta T_\epsilon \|_\infty^2) \right] T^{-1}\mathbf{e} \|_\infty) \\
&= \| T^{-1}\mathbf{e} \|_\infty + O(\| T^{-1}\Delta T_\epsilon \|_\infty \| T^{-1}\mathbf{e} \|_\infty) \\
&= \| \mathbf{z} \|_\infty \left(1 + O(N^{3-\alpha}) \right).
\end{aligned}
$$

$$(4.7)$$

Similarly,

$$(4.8) \qquad \| \mathbf{y}_\epsilon \|_1 = \| (T + \Delta T_\epsilon)^{-1}\mathbf{e} \|_1 = \| \mathbf{z} \|_1 \left(1 + O(N^{3-\alpha}) \right).$$

Since $\mathbf{y}_\epsilon = (T + \Delta T_\epsilon)^{-1}\mathbf{e} \geq 0$, this also shows that $\mathbf{e}^T\mathbf{y}_\epsilon = O(N^3)$, and hence

$$(4.9) \qquad 1 - \epsilon\mathbf{e}^T\mathbf{y}_\epsilon = 1 + O(N^{3-\alpha}).$$

Using (4.5), this gives

$$\mathbf{z}_\epsilon = \left(1 + O(N^{3-\alpha}) \right) \mathbf{y}_\epsilon.$$

So, from (4.7)

$$\| \mathbf{z}_\epsilon \|_\infty = \| \mathbf{z} \|_\infty \left(1 + O(N^{3-\alpha}) \right)$$

and from (4.8)

$$\| \mathbf{z}_\epsilon \|_1 = \| \mathbf{z} \|_1 \left(1 + O(N^{3-\alpha}) \right),$$

as required.   □

4.3. Proofs of Theorems 2.2 and 2.3.
We begin this subsection by discussing the main ideas in the proofs of Theorems 2.2 and 2.3 and introducing some notation before proving a lemma that formalizes the key steps.

Note that by using the Sherman–Morrison formula to establish (4.5) we have essentially reduced the problem to the study of $\mathbf{y}_\epsilon$ in (4.4). This system may be written as

$$\left(\frac{1}{\Delta x^2}\mathrm{tridiag}(1,-2,1) - 2\epsilon N^3\mathrm{tridiag}(\tfrac{1}{2},0,\tfrac{1}{2})\right)\mathbf{y}_\epsilon = -2N^2\mathbf{e},$$

where $\Delta x = 1/N$. This may be interpreted as a finite difference formula applied to the boundary value problem

$$(4.10)\qquad y''(x) - 2\epsilon N^3 y(x) = -2N^2, \quad 0 \le x \le 1, \quad y(0) = y(1) = 0.$$

The finite difference formula applies standard central differences to the $y''(x)$ term but uses slightly unusual symmetric averaging for the $y(x)$ term. The boundary value problem (4.10) has solution

$$(4.11)\qquad y_{\mathrm{bvp}}(x) = \frac{1}{\epsilon N}\left[1 - \frac{\cosh\gamma(x - \tfrac{1}{2})}{\cosh\frac{\gamma}{2}}\right],$$

where $\gamma := \sqrt{2\epsilon N^3}$. We let $\mathbf{y}_{\mathrm{bvp}} \in \mathbb{R}^{N-1}$ denote the vector whose ith component is given by $y_{\mathrm{bvp}}(x_i)$, where $x_i = i\Delta x$.

If the finite difference method is successful, then we would expect $\mathbf{y}_{\mathrm{bvp}}$ to form a good approximation to $\mathbf{y}_\epsilon$, and this is the basis of our analysis. We note, however, that some care is required since, unlike in the scenario normally studied by numerical analysts, the underlying problem (4.10) depends on the discretization parameter Δx (through N). However, by carefully adapting the traditional M-matrix type analysis (see, for example [21, Chapter 6]) and exploiting the special structure of the problem, it is possible to obtain a useful result. (As an aside, we mention that the original system (4.1) could be analyzed through a finite difference framework by regarding $\mathbf{e}\mathbf{e}^T$ as approximating a scaled integral operator. However, we found it more convenient to invoke Sherman–Morrison.)

To proceed, we therefore define the truncation error vector $\boldsymbol{\tau} \in \mathbb{R}^{N-1}$ by

$$\tau_i := \frac{1}{\Delta x^2}\left[y_{\mathrm{bvp}}(x_i - \Delta x) - 2y_{\mathrm{bvp}}(x_i) + y_{\mathrm{bvp}}(x_i + \Delta x)\right]$$
$$(4.12)\qquad\qquad - \frac{\gamma^2}{2}\left[y_{\mathrm{bvp}}(x_i - \Delta x) + y_{\mathrm{bvp}}(x_i + \Delta x)\right] + 2N^2.$$

Equivalently, we may write

$$(4.13)\qquad \widehat{T}_\epsilon \mathbf{y}_{\mathrm{bvp}} = \mathbf{e} - \frac{\Delta x^2}{2}\boldsymbol{\tau}.$$

LEMMA 4.1. *Suppose $1 < \alpha \le 3$. Then the truncation error $\boldsymbol{\tau}$ satisfies $\tau_i > 0$ for all i (for sufficiently large N) with*

$$(4.14)\qquad \|\boldsymbol{\tau}\|_\infty = O(N^{3-\alpha}) \quad\text{and}\quad \|\boldsymbol{\tau}\|_1 = O(N^{\frac{5-\alpha}{2}}).$$

Further,

$$(4.15)\qquad \|\mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}}\|_\infty = O(1)$$

and

$$(4.16) \qquad 1 - \epsilon \mathbf{e}^T \mathbf{y}_\epsilon = \frac{2 \tanh \frac{\gamma}{2}}{\gamma} + O(N^{-1}).$$

Proof. It follows from (4.11) that

$$(4.17) \qquad \| \mathbf{y}_{\text{bvp}} \|_\infty = O\left(\frac{1}{\epsilon N}\right) = O(N^{\alpha-1}).$$

Also

$$\mathbf{e}^T \mathbf{y}_{\text{bvp}} = \frac{N-1}{\epsilon N} - \frac{2}{\epsilon N} \frac{\sum_{i=1}^{N-1} e^{\gamma(i\Delta x - \frac{1}{2})}}{e^{\gamma/2} + e^{-\gamma/2}}.$$

By summing the geometric series and exploiting the fact that $\gamma = O(N^{(3-\alpha)/2})$ we find

$$(4.18) \qquad \mathbf{e}^T \mathbf{y}_{\text{bvp}} = \frac{1}{\epsilon} - \frac{2 \tanh \frac{\gamma}{2}}{\epsilon \gamma} + O(N^{\alpha-1}).$$

(Note that for $1 < \alpha < 3$ the $\tanh \frac{\gamma}{2}$ factor in (4.18) may be replaced by 1.)

To estimate $\boldsymbol{\tau}$, we note that since $y_{\text{bvp}} \in C^4[0,1]$, Taylor expansions give

$$y_{\text{bvp}}(x_i - \Delta x) - 2y_{\text{bvp}}(x_i) + y_{\text{bvp}}(x_i + \Delta x) = \Delta x^2 y''_{\text{bvp}}(x_i) + \frac{\Delta x^4}{4!} \left[y^{\text{IV}}_{\text{bvp}}(\xi_i^1) + y^{\text{IV}}_{\text{bvp}}(\xi_i^2)\right],$$

$$y_{\text{bvp}}(x_i - \Delta x) + y_{\text{bvp}}(x_i + \Delta x) = 2y_{\text{bvp}}(x_i) + \frac{\Delta x^2}{2} \left[y''_{\text{bvp}}(\zeta_i^1) + y''_{\text{bvp}}(\zeta_i^2)\right],$$

where $\xi_i^1, \zeta_i^1 \in [x_{i-1}, x_i]$ and $\xi_i^2, \zeta_i^2 \in [x_i, x_{i+1}]$. It follows from (4.12) that

$$\tau_i = \frac{\Delta x^2}{4!} \left[y^{\text{IV}}_{\text{bvp}}(\xi_i^1) + y^{\text{IV}}_{\text{bvp}}(\xi_i^2)\right] - \frac{\gamma^2 \Delta x^2}{4} \left[y''_{\text{bvp}}(\zeta_i^1) + y''_{\text{bvp}}(\zeta_i^2)\right].$$

Now, since $y^{\text{IV}}_{\text{bvp}}(x) = \gamma^2 y''_{\text{bvp}}(x)$, we have

$$(4.19) \qquad \tau_i = \frac{\gamma^2 \Delta x^2}{4} \left[\frac{y''_{\text{bvp}}(\xi_i^1) + y''_{\text{bvp}}(\xi_i^2)}{6} - \left(y''_{\text{bvp}}(\zeta_i^1) + y''_{\text{bvp}}(\zeta_i^2)\right)\right].$$

Another Taylor expansion gives

$$|y''_{\text{bvp}}(\xi_i^1) - y''_{\text{bvp}}(x_i)| \le \Delta x |y'''(\xi_i^{1,1})| \le \gamma \Delta x |y''(\xi_i^{1,1})|$$

for some $\xi_i^{1,1} \in [x_{i-1}, x_i]$, and thus

$$|y''_{\text{bvp}}(\xi_i^1) - y''_{\text{bvp}}(x_i)| \le \gamma \Delta x \left(|y''_{\text{bvp}}(x_i)| + \Delta x |y'''(\xi_i^{1,2})|\right)$$

$$\le \gamma \Delta x \left(|y''_{\text{bvp}}(x_i)| + \gamma \Delta x |y''(\xi_i^{1,2})|\right)$$

for some $\xi_i^{1,2} \in [x_{i-1}, x_i]$. Continuing this argument we find

$$|y''_{\text{bvp}}(\xi_i^1) - y''_{\text{bvp}}(x_i)| \le |y''_{\text{bvp}}(x_i)| \sum_{k=1}^{l} (\gamma \Delta x)^k + (\gamma \Delta x)^{l+1} \max_{[x_{i-1}, x_i]} |y''_{\text{bvp}}(x)|$$

$$(4.20) \qquad \le |y''_{\text{bvp}}(x_i)| \frac{\gamma \Delta x}{1 - \gamma \Delta x} + (\gamma \Delta x)^{l+1} \max_{[x_{i-1}, x_i]} |y''_{\text{bvp}}(x)|$$

for any $l \geq 1$. By taking l sufficiently large, we can make the second term in (4.20) negligible, and hence

$$y''_{\text{bvp}}(\xi_i^1) = y''_{\text{bvp}}(x_i)\left(1 + O(\gamma \Delta x)\right).$$

Similarly, this expansion holds for $y''_{\text{bvp}}(\xi_i^2)$, $y''_{\text{bvp}}(\zeta_i^1)$, and $y''_{\text{bvp}}(\zeta_i^2)$, so, in (4.19),

$$\tau_i = \gamma^2 \Delta x^2 y''_{\text{bvp}}(x_i)\left(\frac{-5}{12} + O(\gamma \Delta x)\right).$$

Since $y''_{\text{bvp}}(x_i) < 0$ for all i, the positivity of τ_i follows. Using $\max_{[0,1]}|y''_{\text{bvp}}(x)| = O(\gamma^2/(\epsilon N)) = O(N^2)$ we then find that $\|\boldsymbol{\tau}\|_\infty = O(N^{3-\alpha})$.

To bound $\|\boldsymbol{\tau}\|_1$ we note from (4.11) that

$$\sum_{i=0}^{N} |y''_{\text{bvp}}(x_i)| = \gamma^2 \sum_{i=0}^{N}\left[\frac{1}{\epsilon N} - y_{\text{bvp}}(x_i)\right] \leq \frac{\gamma^2(N+1)}{\epsilon N} + \gamma^2 \|\mathbf{y}_{\text{bvp}}\|_1.$$

From (4.18) we have $\|\mathbf{y}_{\text{bvp}}\|_1 = O(N^{3(\alpha-1)/2})$, so

$$(4.21) \qquad \sum_{i=0}^{N} |y''_{\text{bvp}}(x_i)| = O(N^{(3+\alpha)/2}).$$

Since $|y''_{\text{bvp}}(x)|$ takes its extreme value over $[x_i, x_{i+1}]$ at an endpoint, we have, from (4.19) and (4.21),

$$\|\boldsymbol{\tau}\|_1 \leq \frac{\gamma^2 \Delta x^2}{4}\left(\frac{1}{6} + \frac{1}{6} + 1 + 1\right)\sum_{i=0}^{N} |y''_{\text{bvp}}(x_i)| = O(N^{(5-\alpha)/2}).$$

Now from (2.5) and (4.4) we have $\mathbf{z} - \mathbf{y}_\epsilon = T^{-1}\Delta T_\epsilon \mathbf{y}_\epsilon$. We know that $T^{-1} \geq 0$, $\Delta T_\epsilon \geq 0$, and $\mathbf{y}_\epsilon \geq 0$ (because $T + \Delta T_\epsilon$ is Stieltjes). Hence $\mathbf{z} - \mathbf{y}_\epsilon \geq 0$, that is,

$$(4.22) \qquad T^{-1}\mathbf{e} \geq \widehat{T_\epsilon}^{-1}\mathbf{e}.$$

Then from (4.4) and (4.13) we have

$$(4.23) \qquad \mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}} = \frac{\Delta x^2}{2}\widehat{T_\epsilon}^{-1}\boldsymbol{\tau},$$

so, using (4.22),

$$(4.24) \qquad |\mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}}| \leq \frac{\Delta x^2}{2}\|\boldsymbol{\tau}\|_\infty \widehat{T_\epsilon}^{-1}\mathbf{e}$$

$$\leq \frac{\Delta x^2}{2}\|\boldsymbol{\tau}\|_\infty T^{-1}\mathbf{e}.$$

Hence, using (4.6) and (4.14),

$$(4.25) \qquad \|\mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}}\|_\infty \leq \frac{\Delta x^2}{2}\|\boldsymbol{\tau}\|_\infty \|T^{-1}\mathbf{e}\|_\infty = O(N^{3-\alpha}).$$

We now refine this bound for $\alpha < 3$. From (4.17) and (4.25) we have

$$(4.26) \quad \|\widehat{T_\epsilon}^{-1}\mathbf{e}\|_\infty = \|\mathbf{y}_\epsilon\|_\infty \leq \|\mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}}\|_\infty + \|\mathbf{y}_{\text{bvp}}\|_\infty = O(N^{3-\alpha}) + O(N^{\alpha-1}).$$

For $2 \leq \alpha < 3$ the $O(N^{\alpha-1})$ term dominates, and so after taking norms in (4.24) we have

$$(4.27) \qquad \| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{-2} N^{3-\alpha} N^{\alpha-1}) = O(N^0).$$

For $1 < \alpha < 2$, in (4.26) we have $\| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty = O(N^{3-\alpha})$. Using this in (4.24) gives

$$\| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{-2} N^{3-\alpha} N^{3-\alpha}) = O(N^{4-2\alpha}).$$

Hence,

$$(4.28) \qquad \| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty \leq \| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty + \| \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{4-2\alpha}) + O(N^{\alpha-1}).$$

For $\alpha \geq 5/3$, the $O(N^{\alpha-1})$ term dominates and we may use (4.24) to recover (4.27). For $1 < \alpha < 5/3$, in (4.28) we have $\| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty = O(N^{4-2\alpha})$. Using this in (4.24) gives

$$\| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{-2} N^{3-\alpha} N^{4-2\alpha}) = O(N^{5-3\alpha}).$$

Hence

$$(4.29) \qquad \| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty \leq \| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty + \| \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{5-3\alpha}) + O(N^{\alpha-1}).$$

For $\alpha \geq 6/4$, the $O(N^{\alpha-1})$ term dominates and we may use (4.24) to recover (4.27). For $1 < \alpha < 6/4$, in (4.29) we have $\| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty = O(N^{5-3\alpha})$. Using this in (4.24) gives

$$\| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{-2} N^{3-\alpha} N^{5-3\alpha}) = O(N^{6-4\alpha}).$$

Hence

$$(4.30) \qquad \| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty \leq \| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty + \| \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^{6-4\alpha}) + O(N^{\alpha-1}).$$

For $\alpha \geq 7/5$, the $O(N^{\alpha-1})$ term dominates and we may use (4.24) to recover (4.27).

The pattern is now clear. Given any integer $k \geq 1$ we can establish $\| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty = O(N^0)$ for $(k+2)/k \leq \alpha \leq 3$, which confirms (4.15).

From (4.14), (4.15), (4.17), and (4.23) we have

$$\begin{aligned}
\| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_1 &\leq \frac{\Delta x^2}{2} \| \widehat{T}_\epsilon^{-1} \|_1 \| \boldsymbol{\tau} \|_1 \\
&= \frac{\Delta x^2}{2} \| \widehat{T}_\epsilon^{-1} \mathbf{e} \|_\infty \| \boldsymbol{\tau} \|_1 \\
&= \frac{\Delta x^2}{2} \| \mathbf{y}_\epsilon \|_\infty \| \boldsymbol{\tau} \|_1 \\
&\leq \frac{\Delta x^2}{2} [\| \mathbf{y}_{\mathrm{bvp}} \|_\infty + \| \mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} \|_\infty] \| \boldsymbol{\tau} \|_1 \\
&= O(N^{-2} N^{\alpha-1} N^{(5-\alpha)/2}) \\
&= O(N^{(\alpha-1)/2}).
\end{aligned}$$

Since $\mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}} = \frac{\Delta x^2}{2} \widehat{T}_\epsilon^{-1} \boldsymbol{\tau} \geq 0$, this is equivalent to

$$\mathbf{e}^T (\mathbf{y}_\epsilon - \mathbf{y}_{\mathrm{bvp}}) = O(N^{(\alpha-1)/2}),$$

which gives, using (4.18),

$$1 - \epsilon \mathbf{e}^T \mathbf{y}_\epsilon = 1 - \epsilon \mathbf{e}^T \mathbf{y}_{\text{bvp}} + O(N^{(-\alpha-1)/2}) = \frac{2 \tanh \frac{\gamma}{2}}{\gamma} + O(N^{-1}),$$

completing the proof. □

We are now in a position to prove Theorems 2.2 and 2.3.

Proof of Theorems 2.2 and 2.3. From (4.5), (4.15), (4.16), and (4.17) we obtain

$$\| \mathbf{z}_\epsilon \|_\infty = \frac{\gamma}{2 \tanh \frac{\gamma}{2}} \| \mathbf{y}_{\text{bvp}} \|_\infty + O(N).$$

For $1 < \alpha < 3$, $\| \mathbf{y}_{\text{bvp}} \|_\infty$ equals $y_{\text{bvp}}(\frac{1}{2})$ plus exponentially small terms, and for $\alpha = 3$, $\| \mathbf{y}_{\text{bvp}} \|_\infty = y_{\text{bvp}}(\frac{1}{2}) + O(1)$. So

$$\| \mathbf{z}_\epsilon \|_\infty = \frac{\gamma}{2 \tanh \frac{\gamma}{2}} y_{\text{bvp}}(\tfrac{1}{2}) + O(N)$$

for $1 < \alpha \leq 3$. This simplifies to

$$\| \mathbf{z}_\epsilon \|_\infty = \frac{\gamma \tanh \frac{\gamma}{4}}{2 \epsilon N} + O(N).$$

Using $\| \mathbf{z} \|_\infty = N^2/4 + O(1)$ we thus have

$$(4.31) \qquad \frac{\| \mathbf{z}_\epsilon \|_\infty}{\| \mathbf{z} \|_\infty} = \frac{2 \gamma \tanh \frac{\gamma}{4}}{\epsilon N^3} + O(N^{-1}).$$

For $\alpha = 3$ we have $\epsilon = KN^{-3}$ and $\gamma = \sqrt{2K}$. Inserting this into (4.31) gives (2.13). For $1 < \alpha < 3$ we have $\gamma = \sqrt{2K} N^{(3-\alpha)/2}$ and (2.15) follows.

For the 1-norm result, we first note that $\mathbf{y}_\epsilon \geq 0$, $\mathbf{y}_{\text{bvp}} \geq 0$ and $\mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}} \geq 0$, so that

$$\| \mathbf{y}_\epsilon \|_1 = \| \mathbf{y}_{\text{bvp}} \|_1 + \| \mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}} \|_1 = \| \mathbf{y}_{\text{bvp}} \|_1 + O(N \| \mathbf{y}_\epsilon - \mathbf{y}_{\text{bvp}} \|_\infty).$$

Using (4.5), (4.15), (4.16), and (4.18) we find

$$\| \mathbf{z}_\epsilon \|_1 = \frac{\gamma}{2 \epsilon \tanh \frac{\gamma}{2}} - \frac{1}{\epsilon} + O(N^2).$$

Scaling by $\| \mathbf{z} \|_1 = \frac{(N-1)N(N+1)}{6}$ and inserting $\gamma = \sqrt{2 \epsilon N^3}$ gives the estimates (2.14) and (2.16). □

5. Final remarks. Our aim in this work was to show via matrix perturbation theory that the small world phenomenon arises in the context of Markov chains. The results are fully rigorous, with sharp error estimates that vanish as the system size increases.

There are, of course, many ways in which the Markov chain model may be extended or altered. The two most obvious directions are perhaps moving to higher dimensions and considering more complex underlying lattice topologies. Further, instead of giving equal weight to all nonneighboring states we could, for example, introduce range-dependent perturbations to the transition matrix of the form $f(|i - j|)$ for some suitable function f. Grindrod [7, 8] has recently produced some elegant results for analogous network models. Alternatively, we could perturb only a small, fixed number of zeros in (2.4). We note that Liu, Strang, and Ott [13] have characterized the effect of this type of modification on the spectrum of a structured matrix. In all cases, the techniques developed here form a useful starting point for further analysis.

REFERENCES

[1] A. D. Barbour and G. Reinert, *Small worlds*, Random Structures Algorithms, 19 (2001), pp. 54–74.

[2] E. Behrends, *Introduction to Markov Chains with Special Emphasis on Rapid Mixing*, Vieweg, Braunschweig, 2000.

[3] H. C. Berg, *Random Walks in Biology*, Princeton University Press, Princeton, NJ, 1983.

[4] H. Caswell, *Matrix Population Models*, 2nd ed., Sinauer Associates, Sunderland, MA, 2001.

[5] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[6] P. M. Gleiss, P. F. Stadler, A. Wagner, and D. A. Fell, *Relevant cycles in chemical reaction networks*, Adv. Complex Systems, 4 (2001), pp. 207–226.

[7] P. Grindrod, *Range-dependent random graphs and their application to modeling large small-world proteome datasets*, Phys. Rev. E, 66 (2002), article 066702.

[8] P. Grindrod, *Modeling proteome networks with range-dependent graphs*, Amer. J. Pharmaco Genomics, 3 (2003), pp. 1–4.

[9] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.

[10] S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub, *Extrapolation methods for accelerating pagerank computations*, in Proceedings of the Twelfth International World Wide Web Conference, Budapest, Hungary, 2003.

[11] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, San Diego, 1975.

[12] J. Kleinberg, *Navigation in a small world*, Nature, 406 (2000), p. 845.

[13] X. Liu, G. Strang, and S. Ott, *Localized eigenvectors from widely spaced matrix modifications*, SIAM J. Discrete Math., 16 (2003), pp. 479–498.

[14] C. Moler, *The world's largest matrix computation*, MATLAB News and Notes, October 2002.

[15] J. M. Montoya and R. V. Solé, *Small world patterns in food webs*, J. Theoret. Biology, 214 (2002), pp. 405–412.

[16] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, UK, 1995.

[17] M. E. J. Newman, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA, 98 (2001), pp. 404–409.

[18] M. E. J. Newman, *The structure and function of networks*, Comput. Phys. Comm., 147 (2002), pp. 40–45.

[19] M. E. J. Newman, C. Moore, and D. J. Watts, *Mean-field solution of the small-world network model*, Phys. Rev. Lett., 84 (2000), pp. 3201–3204.

[20] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge, UK, 1997.

[21] J. M. Ortega, *Numerical Analysis: A Second Course*, SIAM, Philadelphia, 1990.

[22] S. H. Strogatz, *Exploring complex networks*, Nature, 410 (2001), pp. 268–276.

[23] A. Wagner and D. A. Fell, *The small world inside large metabolic networks*, Proc. Roy. Soc. London Ser. B., 268 (2001), pp. 1803–1810.

[24] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton, NJ, 1999.

[25] D. J. Watts and S. H. Strogatz, *Collective dynamics of "small-world" networks*, Nature, 393 (1998), pp. 440–442.