The idemetric property: when most distances are (almost) the same

George Barmpalias[1] Neng Huang[2] Andrew Lewis-Pye[3] Angsheng Li[4]
Xuechen Li[5] Yicheng Pan[4] Tim Roughgarden[6]

January 15th, 2019

Abstract. We introduce the *idemetric* property, which formalises the idea that most nodes in a graph have similar distances between them, and which turns out to be quite standard amongst small-world network models. Modulo reasonable sparsity assumptions, we are then able to show that a strong form of idemetricity is actually equivalent to a very weak expander condition (PUMP). This provides a direct way of providing short proofs that small-world network models such as the Watts-Strogatz model are strongly idemetric (for a wide range of parameters), and also provides further evidence that being idemetric is a common property.

We then consider how satisfaction of the idemetric property is relevant to algorithm design. For idemetric graphs we observe, for example, that a single breadth-first search provides a solution to the all-pairs shortest paths problem, so long as one is prepared to accept paths which are of stretch close to 2 with high probability. Since we are able to show that Kleinberg's model is idemetric, these results contrast nicely with the well known negative results of Kleinberg concerning efficient decentralised algorithms for finding short paths: for precisely the same model as Kleinberg's negative results hold, we are able to show that very efficient (and decentralised) algorithms exist if one allows for reasonable preprocessing. For deterministic distributed routing algorithms we are also able to obtain results proving that less routing information is required for idemetric graphs than in the worst case in order to achieve stretch less than 3 with high probability: while $\Omega(n^2)$ routing information is required in the worst case for stretch strictly less than 3 on almost all pairs, for idemetric graphs the total routing information required is $O(nlog(n))$.

[1]State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, 100190, P. R. China.

[2]School of Computer Science, University of Chinese Academy of Sciences, Beijing, P. R. China.

[3]Department of Mathematics, London School of Economics, London, UK.

[4]State Key Laboratory of Software Development Environment, School of Computer Science, BeiHang University, 100191, Beijing, P. R. China.

[5]Department of Computer Science, University of Toronto, Canada.

[6]Department of Computer Science, Stanford, USA.

1 Introduction

One of the basic tasks of network science is to identify properties which are common to most real-world networks of interest, from telecommunication and computer networks, to networks arising in biological or social contexts. A well known example of such a property is given by the fact that these networks tend to be *small worlds*, i.e. nodes in real-world networks tend to have short paths between them. The phrase "six degrees of separation", which originated with the experiments of Stanley Milgram [SM67] in the 60s, sums up the idea that two people normally have a short path of contacts connecting them. This idea has subsequently been extensively verified [MK89, WS98, AJB99] and has since permeated popular culture.

As well as the existence of short paths, however, a notable feature of many of these networks is that they have *strongly concentrated distance distributions*. Most distances, in other words, are almost the same. Close to 60% of Facebook user pairs are at distance 5 in the underlying friendship network [RH16], for example, while over 80% of actor pairs on the IMDB[1] are at a distance either 4 or 5.[2] Perhaps to an even greater extent, this is a feature which also carries over to most small-world network models. Figure 1 shows examples of distance distributions for three of the best known small-world models, the Erdős-Rényi [ER60], Preferential Attachment [BA99] and Watts-Strogatz [WS98] models. As we shall see, in these and most other small-world models, such as the Chung-Lu [CL02, CL03] and Norros-Reittu [NR06] models, the distance distributions can be *proved* to become proportionately more concentrated (in a certain formal sense) as the size of the network increases.

In order to study this phenomenon in a formal setting, of course we need a mathematical definition which makes precise the idea that "most distances are almost the same". Definition 1.1 below is new, and seems natural.

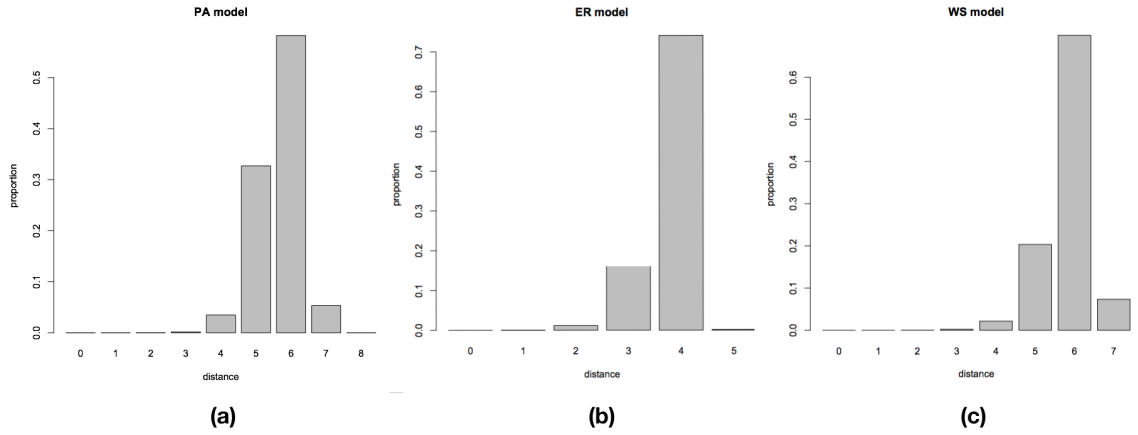

Figure 1: These graphs show examples of distance distributions for three small-world network models. Distance is displayed on the x-axis, while the y-axis shows the proportion of pairs at that distance: (a) is for the Preferential Attachment (PA) model, with 10^6 nodes and mean degree 6; (b) is for the Erdős-Rényi model $G(n,p)$, with $n = 5\times 10^4$ nodes and expected degree 25; (c) is for the Watts-Strogatz model with 10^6 nodes, in which each node is initially connected to 10 neighbours either side, and the rewiring probability is 0.2.

[1] The data for the Internet Movie Database can be downloaded at http://www.imdb.com/interfaces

[2] The graph considered here is the same as that in the so called *Kevin Bacon game*: nodes are actors or actresses, and there is an edge between two nodes if they have appeared in the same film.

Let $\mathbb{G}$ be a random network model for which distances are defined in the standard fashion,[3] which produces for every $n \in \mathbb{N}$ (and perhaps taking other inputs, which we consider for now to be fixed) an ensemble of graphs with n nodes, i.e. a probability distribution over a certain set of graphs with n nodes. For each n we let G_n be a graph sampled from the distribution given by $\mathbb{G}$. Recall that if X is a random variable and $\{X_n\}_{n\geq 0}$ is a sequence of random variables, then we say that X_n tends to X *in probability*, denoted $X_n \xrightarrow{P} X$, if for every $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

Definition 1.1. *We say $\mathbb{G}$ is **idemetric** if there exists a finite valued function f (i.e. $f(n) \neq \infty$), such that if u_n and v_n are nodes chosen uniformly at random from G_n, then $X_n = d(u_n, v_n)/f(n) \xrightarrow{P} 1$.*[4]

For network models in which the graph produced may not be connected but does have a giant component, a weaker notion than idemetric is also useful. We shall say an event occurs *with high probability* if it occurs with probability tending to 1 as $n \to \infty$.

Definition 1.2. *We say that a random network model $\mathbb{G}$ is **weakly idemetric** if both:*

1. *There exists $\kappa \in (0, 1]$ such that, with high probability, the largest component of G_n is of size at least κn.*
2. *There exists a function $f(n)$ such that if u_n and v_n are nodes chosen uniformly at random from the same largest component of G_n, then $X_n = d(u_n, v_n)/f(n) \xrightarrow{P} 1$.*

Let us say that a network model is *partially unbounded* if there exists $\epsilon > 0$ such that for all b the following holds for sufficiently large n: if u_n and v_n are nodes chosen uniformly at random from G_n, then $d(u_n, v_n) > b$ with probability $> \epsilon$. Many of the network models which have been shown to be idemetric (without the term itself being used) actually satisfy a property which is stronger, providing the model is partially unbounded:

Definition 1.3. *We say $\mathbb{G}$ is **strongly idemetric** (SI) if there exists a finite valued function f, and a bound b_ϵ for each $\epsilon > 0$ (independent of n), such that, for all sufficiently large n, if u_n and v_n are chosen uniformly at random from G_n, then $|f(n) - d(u_n, v_n)| < b_\epsilon$ with probability $> (1 - \epsilon)$.*

So we actually have three closely related notions: weakly idemetric, idemetric and strongly idemetric. While the idemetric definition might seem sufficient to correctly capture the intended notion, the strongly idemetric definition will turn out to be at least as significant due to how commonly it is satisfied. Immediate evidence that these definitions are of interest is then provided by the fact that many of the best known small-world models have actually already been shown to be (weakly/strongly) idemetric. Classical results in graph theory (see [RD06] or [RH]) suffice to show, for example, that when the expected degree of each node is greater than 1, the Erdős-Rényi random graph $G(n, p)$ is weakly idemetric (although that term is not used in the literature). Bounded degree expanders are known to be strongly idemetric [JSW09]. Similarly, without the term itself being applied, a range of inhomogeneous random graph models, such as the Preferential Attachment model of Barabási and Albert [BA99], the Chung-Lu [CL02, CL03] model and the Norros-Reittu model [NR06], have been shown to be idemetric for a wide range of parameters. In fact there are a number of ways of making the definition of the Preferential Attachment model precise,

[3] If a and b are nodes in a (possibly directed) graph, then we let $d(a, b)$ denote the length of a shortest path from a to b (where all edges are traversed in the forward direction if the graph is directed, and with $d(a, b) = \infty$ if no such path exists). In the case of a weighted graph, the above definition still applies, if the length of a path is defined to be the sum of all edge weights in the path. Note that in a directed graph it may not be the case that $d(a, b) = d(b, a)$. In an undirected graph, however, d is a metric.

[4] To be clear, X_n is defined by first choosing G_n according to the distribution given by G, and then choosing u_n and v_n uniformly at random from G_n.

and a generalised model has often been considered (see, for example, [DHH10]) which depends on two parameters, an integer $m \geq 1$ and a real $\delta > -m$: roughly speaking, in every step a new node is added to the network and connected to m existing nodes with a probability proportional to their degree plus δ. This produces a degree distribution with power law exponent $\tau = 3 + \delta/m$, meaning that all values of $\tau > 2$ are possible (which is the motivation for including δ in the definition of the model). Let the graph distance H_n be the distance between two randomly chosen nodes in the giant component. When $m = 1$ and $\delta > -1$, we have [BP94, RH]:

$$\frac{H_n}{\log(n)} \xrightarrow{P} \frac{2(1+\delta)}{2+\delta}.$$

When $m \geq 1$ and $\delta > 0$ (so that $\tau > 3$) there exist $0 < a_1 < a_2 < \infty$ such that, as $n \to \infty$, $\mathbb{P}(a_1 \log(n) \leq H_n \leq a_2 \log(n)) = 1 - o(1)$. If $m \geq 2$ and $\delta = 0$ (so that $\tau = 3$) then [BR04, RH]:

$$H_n \frac{\log\log(n)}{\log(n)} \xrightarrow{P} 1.$$

Finally, if $m \geq 2$ and $\delta \in (-m, 0)$ then [DHH10, DMM12, RH]:

$$\frac{H_n}{\log\log(n)} \xrightarrow{P} \frac{4}{|\log(\tau - 2)|}.$$

When $m \geq 2$ the graph produced is connected with high probability, meaning that these results suffice to give idemetricity. When $m = 1$ the connectivity properties depend somewhat on the details of the model (such as whether self-edges are allowed), but the established results still suffice to establish that the corresponding models are at least weakly idemetric. For a detailed summary of these results together with self-contained proofs see [RH].

Of course the Preferential Attachment model was introduced to try and explain the power law degree distributions observed in many real networks. An alternative and related approach which has been extensively studied, is to consider the configuration model applied to an i.i.d. sequence of degrees with a power-law degree distribution. According to this model one starts by sampling the degree sequence from a power law and then connects nodes with the sampled degrees at random (possibly resulting in a multigraph). The results obtained here are broadly similar, although constant multiplicative factors may differ [HHM05, HHZ07]. When $\tau \in (2, 3)$ graph distance centres around $2\log\log(n)/|\log(\tau - 2)|$, while for $\tau > 3$ the graph distance grows logarithmically with n.

In this paper we shall show further that the Watts-Strogatz model [WS98], and the Kleinberg model [JK00] are both idemetric for a wide range of parameters.

Given that it seems most (if not all) well-known small world models are at least weakly idemetric for a wide range of parameters, we then distinguish two clear aims.

1. In the spirit of the programme of research on quasi-random graphs by Fan Chung et al. [CGW89, CG08], which is aimed at showing that many conditions satisfied by random graphs are equivalent to each other, we would like to characterise the (weakly/strongly) idemetric network models.

2. We would like to understand how one can make use of knowledge that the idemetric property is satisfied. What is the significance for algorithm design?

In Section 2 we address the first of these aims. Modulo reasonable sparsity assumptions, we establish the surprising fact that, for unweighted undirected networks, being strongly idemetric is actually *equivalent* to a very weak expander condition, which we call PUMP and which is described in Definition 1.5 below. This provides a direct way of providing short proofs that networks models are strongly idemetric, and we immediately apply it to establish that the Watts-Strogatz model is strongly idemetric for a wide range of parameters.

Definition 1.4. *For a node u, we write $B_u(r)$ to denote the ball around u of radius r, while $\overline{B_u(r)}$ denotes the complement of $B_u(r)$. For sets of nodes X, Y, we write $e(X, Y)$ to denote the number of edges incident to one node in X and one node in Y.*

Definition 1.5. *$\mathbb{G}$ is a weak ball expander (PUMP) if whenever $0 < \epsilon < \frac{1}{2}$ there exists $\alpha_\epsilon > 0$ such that, for all sufficiently large n, if u is chosen uniformly at random then with probability $> 1-\epsilon$ both of the following hold: (i) there exists r with $|B_u(r)| \geq \epsilon n$, (ii) for all r with $\epsilon n \leq |B_u(r)| \leq (1-\epsilon)n$, $e(B_u(r), \overline{B_u(r)}) \geq \alpha_\epsilon |B_u(r)|$.*

In Section 3 we then address the second aim above. For idemetric network models we observe that a single breadth-first search provides a solution to the all-pairs shortest paths problem, so long as one is prepared to accept paths which are of stretch close to 2 with high probability.[5] The significance of this result stems from the broad class of networks to which it applies. One might hope, nevertheless, to be able to achieve smaller stretch, by further restricting the class of network models considered in a reasonable fashion:

Question 1.6. *Does there exist a condition which is satisfied by most/all well-known small world network models, and which allows for an efficient[6] solution to the all-pairs shortest paths problem, with stretch < 2?*

Since we are then able to show that Kleinberg's model is idemetric for a wide range of parameters, our results here contrast nicely with the well known negative results of Kleinberg [JK00] concerning efficient decentralised algorithms for finding short paths: for precisely the same model as Kleinberg's negative results hold, we are able to show that very efficient (and decentralised) algorithms exist if one allows for reasonable preprocessing.

For deterministic distributed routing algorithms we are also able to obtain results proving that less routing information is required for idemetric graphs than the worst case in order to achieve stretch less than 3 with high probability: while $\Omega(n^2)$ routing information is required in the worst case for stretch strictly less than 3 on almost all pairs, for idemetric graphs the total routing information required is $O(nlog(n))$.

2 Characterising the idemetric network models

Throughout this section, we restrict attention to undirected, unweighted graphs. The results of this section thus apply to all small-world models mentioned in the paper other than the Kleinberg model, which is directed and is dealt with later. If one is interested in real-world networks, then it makes sense to restrict

[5] Recall that a path from a to b is of stretch k if it is at most k times as long as the shortest path.

[6] By an efficient algorithm in this context we mean one with query times which are $O(\log(n))$, and with preprocessing time $O(n)$ for sparse graphs.

attention further to graphs which are sparse, and in particular to network models in which individual nodes have finite expected degree in the limit. A natural way to formalise this is as follows.

Definition 2.1. *Let deg(u) denote the degree of the node u. If G_n is sampled from the distribution given by $\mathbb{G}$ and u is chosen uniformly at random from G_n, then $\boldsymbol{p}_n(d)$ is the probability that deg(u) = d. We'll say an undirected network model is of finite expected degree (FED) if there exists a probability distribution $\boldsymbol{p}$ on $\mathbb{N}$ with finite mean such that:*

(F1) For each $d \in \mathbb{N}$, $lim_{n\to\infty}\boldsymbol{p}_n(d) = \boldsymbol{p}(d)$;

(F2) $lim_{n\to\infty} \sum_d d\boldsymbol{p}_n(d) = \sum_d d\boldsymbol{p}(d)$.

So in Definition 2.1, (F1) and (F2) just say that $\boldsymbol{p}_n$ converges nicely to $\boldsymbol{p}$ with finite mean as $n \to \infty$. FED is a natural sparsity condition which we can expect to be satisfied by the small-world network models we study (for 'realistic' parameter values): all of the undirected network models mentioned in this paper satisfy the condition for a wide range of parameter values. The following sparsity condition, however, is the one we shall actually need in our proofs, and is implied by FED. Let's say a set of nodes U is a *node cover* for a set of edges E, if every edge from E is incident with at least one node from U.

Definition 2.2. *$\mathbb{G}$ is **uniformly sparse** (US) if, for each $\epsilon > 0$, there exists a constant C_ϵ such that the following holds with probability at least $1 - \epsilon$ for all sufficiently large n: for any set of $y \geq \epsilon n$ edges from G_n, every node covering is of size at least y/C_ϵ.*

Lemma 2.3. *If $\mathbb{G}$ is FED, then it is US.*

Proof. Let $\boldsymbol{p}$ be as guaranteed by satisfaction of FED. Define $M_n = \sum_d d\boldsymbol{p}_n(d)$ and $M = \sum_d d\boldsymbol{p}(d)$. By a *formal property C* of nodes, we mean a set of pairs (G_n, u). For fixed $n \in \mathbb{N}$, if (G_n is sampled according to the distribution given by $\mathbb{G}$ and) u is chosen uniformly at random from G_n, then we define:

$$M_n(C) = \sum_d d \cdot \mathbb{P}[(G_n, u) \in C \;\&\; \deg(u) = d].$$

So $M_n(C)$ can be thought of as the contribution to the mean M_n given by C. For any formal properties C_1 and C_2, note that:

$$M_n \geq M_n(C_1) + M_n(C_2) - M_n(C_1 \cap C_2). \tag{1}$$

Given $\epsilon > 0$, let D be such $\sum_{d\leq D} d\boldsymbol{p}(d) > M - \epsilon^2/3$. Let C_1 be the set of all pairs (G_n, u) such that $u \in G_n$ is of degree at most D. Satisfaction of FED guarantees that, for all sufficiently large n, $\sum_{d\leq D} d\boldsymbol{p}_n(d) > M_n - \epsilon^2/2$. So, to phrase this another way, for all sufficiently large n we have:

$$M_n(C_1) > M_n - \epsilon^2/2. \tag{2}$$

Now choose $\mu > 0$ to be small, and suppose towards a contradiction that there exist infinitely many n for which the following condition holds:

($\star_n$): With probability at least ϵ, the $\lceil \mu n \rceil$ many nodes of highest degree have total degree summing to at least ϵn.

Let C_2 be the set of all pairs (G_n, u) such that the $\lceil \mu n \rceil$ many nodes of highest degree in G_n have total degree summing to at least ϵn, and such that u is one of one of those $\lceil \mu n \rceil$ nodes of highest degree in G_n. When ($\star_n$) holds, we have that:

$$M_n(C_2) \geq \epsilon^2. \tag{3}$$

On the other hand, since D is a bound on the degree of u whenever $(G_n, u) \in C_1 \cap C_2$, we also have:

$$M_n(C_1 \cap C_2) \leq D\mu. \tag{4}$$

So long as μ is chosen sufficiently small, (1)–(4) then produce the required contradiction. We conclude that, for an appropriate choice of $\mu > 0$ and for all sufficiently large n, it holds with probability at least $1 - \epsilon$ that the $\lceil \mu n \rceil$ many nodes of highest degree have total degree summing to less than ϵn. Since this holds for all $\epsilon > 0$, and since (F2) holds, we can then strengthen this statement slightly. For each $\epsilon > 0$ we can choose $K_\epsilon \in \mathbb{N}$ such that, for an appropriate choice of $\mu_\epsilon > 0$, the following holds with probability at least $1 - \epsilon$ for all sufficiently large n: the $\lceil \mu_\epsilon n \rceil$ many nodes of highest degree have total degree summing to less than ϵn and the total number of edges in G_n is less than $K_\epsilon n$.

Now if the $\lceil \mu_\epsilon n \rceil$ many nodes of highest degree have total degree summing to less than ϵn, it clearly holds that no set of $\lceil \mu_\epsilon n \rceil$ many nodes can act as a node cover for any set of $\geq \epsilon n$ many edges. US is therefore satisfied for $C_\epsilon = K_\epsilon / \mu_\epsilon$. □

As mentioned in the introduction, it turns out that a characterisation of the strongly idemetric (SI) network models can be given in terms of *expander graphs*, which have been studied intensively by mathematicians and computer scientists since the 1970s and have applications in the design and analysis of communication networks, in the theory of error correcting codes and in pseudorandomness. For background on expanders see [HLW06].

Definition 2.4. *We say G with $|G| = n$ is an α-edge expander if the following holds for any set of nodes S with $|S| \leq n/2$: $e(S, \bar{S}) \geq \alpha |S|$.*

Definition 2.4, however, is too strong for our purposes, since it requires nice behaviour with respect to *all* S with $|S| \leq n/2$. Definition 1.5, restated below, weakens Definition 2.4 by requiring that the given condition holds only when S is a large ball $B_u(r)$, and even then only *most of the time*. Since Definition 1.5 only describes conditions on S with $\epsilon n \leq |S| \leq (1 - \epsilon)n$, we can drop the restriction that $|S| \leq n/2$ without making the definition any stronger.

Definition 1.5. *$\mathbb{G}$ is a weak ball expander (PUMP) if whenever $0 < \epsilon < \frac{1}{2}$ there exists $\alpha_\epsilon > 0$ such that, for all sufficiently large n, if u is chosen uniformly at random then with probability $> 1 - \epsilon$ both of the following hold: (i) there exists r with $|B_u(r)| \geq \epsilon n$, (ii) for all r with $\epsilon n \leq |B_u(r)| \leq (1 - \epsilon)n$, $e(B_u(r), \overline{B_u(r)}) \geq \alpha_\epsilon |B_u(r)|$.*

Definition 2.5. *We define $r_u(\epsilon)$ to be the least r for which $|B_u(r)| \geq \epsilon n$, or to be undefined if no such r exists.*

Lemma 2.6. *If $\mathbb{G}$ is a PUMP and US, then it is SI.*

Proof. For each node u, consider the following condition:

> $\star(u, \epsilon, \alpha_\epsilon)$: There exists r with $|B_u(r)| \geq \epsilon n$, and for all r with $\epsilon n \leq |B_u(r)| \leq (1 - \epsilon)n$, $e(B_u(r), \overline{B_u(r)}) \geq \alpha_\epsilon |B_u(r)|$.

To establish that $\mathbb{G}$ is SI, we use a condition which is equivalent to PUMP:

P2 : For every $\epsilon > 0$, there exists $\alpha_\epsilon > 0$ such that, for all sufficiently large n, the following holds with probability $> 1 - \epsilon$: $\star(u, \epsilon, \alpha_\epsilon)$ holds for at least $(1 - \epsilon)n$ many nodes in G_n.

Suppose $\mathbb{G}$ satisfies P2 and US. Given arbitrary $\epsilon > 0$, it suffices to show the existence of b_ϵ, such that for all sufficiently large n, if u, v, u', v' are nodes chosen uniformly at random from G_n, then $|d(u, v) - d(u', v')| < b_\epsilon$

with probability $> (1-\epsilon)$. To this end, let ϵ_1 be sufficiently small compared to ϵ (we shall come back and specify precisely what this means later). Let α_{ϵ_1} be the constant guaranteed by satisfaction of P2 w.r.t. ϵ_1, and write α to denote α_{ϵ_1}. Let $C_{\alpha\epsilon_1}$ be the corresponding constant guaranteed by satisfaction of US, and write C for $C_{\alpha\epsilon_1}$. Finally, to complete the initial round of definitions, put $\beta := \alpha/C$.

Fix a node u and suppose that $\star(u, \epsilon_1, \alpha)$ holds. Very roughly, the idea now is as follows. Let r be such that $\epsilon_1 n \leq |B_u(r)| \leq (1-\epsilon_1)n$. The expander property PUMP means we are guaranteed that $e(B_u(r), \overline{B_u(r)}) \geq \alpha|B_u(r)|$. So long as it holds for any set of $y \geq \alpha\epsilon_1 n$ edges from G_n, that every node covering is of size at least y/C, it then follows that $|B_u(r+1)| \geq (1+\beta)|B_u(r)|$. This exponential growth means that most nodes will be within some constant range of distances from u. Then a simple transitivity argument can be used to show that this constant range is the same for most nodes u.

Now let us describe the details of that argument more precisely. Since $\star(u, \epsilon_1, \alpha)$ holds, any ball $B_u(r_0)$ of size s with $\epsilon_1 n \leq s \leq (1-\epsilon_1)n$ has least αs edges coming out of it (i.e. $e(B_u(r), \overline{B_u(r)}) \geq \alpha s$), which are incident with at least βs distinct nodes outside $B_u(r)$ – so long, that is, as it holds for any set of $y \geq \alpha\epsilon_1 n$ edges from G_n, that every node covering is of size at least y/C. $B_u(r_0+1)$ therefore has at least $(1+\beta)s$ nodes. Iterating this argument, ball $B_u(r_0+r)$ has at least $\min((1+\beta)^r s, (1-\epsilon_1)n)$ nodes, for any $r \in \mathbb{N}$. In particular, letting $r_0 = r_u(\epsilon_1)$, it follows that:

$$r_u(1-\epsilon_1) - r_u(\epsilon_1) \leq \log_{1+\beta}\left(\frac{1-\epsilon_1}{\epsilon_1}\right). \tag{5}$$

Define $b_\epsilon := 3(\log_{1+\beta}(1-\epsilon_1) - \log_{1+\beta}(\epsilon_1))$. To present the transitivity argument referred to above, let E_1 be the set of all unordered node pairs $\{u, v\}$ such that $\star(u, \epsilon_1, \alpha)$ and $\star(v, \epsilon_1, \alpha)$ both hold, and such that $r_u(\epsilon_1) \leq d(u, v) \leq r_u(1-\epsilon)$. Note that, so long as $\star(u, \epsilon_1, \alpha)$ holds for at least $(1-\epsilon_1)n$ many nodes in G_n, there are at most $\epsilon_1 n^2 + 2\epsilon_1 n^2 = 3\epsilon_1 n^2$ many node pairs which do not belong to E_1. If at least $18\epsilon_1 n$ many nodes had degree less than $\frac{2}{3}n$ in E_1, i.e. for at least $18\epsilon_1 n$ many nodes u there were less than $\frac{2}{3}n$ nodes v with $\{u, v\} \in E_1$, this would imply the existence of more than $18\epsilon_1 n \cdot \frac{1}{3}n$ ordered node pairs not belonging to E_1, meaning more than $3\epsilon_1 n^2$ unordered node pairs which don't belong to E_1. We conclude that less than $18\epsilon_1 n$ many nodes have degree less than $\frac{2}{3}n$ in E_1. Define E_2 to be all those node pairs $\{u, v\} \in E_1$, for which both u and v have degree at least $\frac{2}{3}n$ in E_1, so that there are less than $3\epsilon_1 n^2 + 18\epsilon_1 n^2 = 21\epsilon_1 n^2$ many node pairs which do not belong to E_2. The point of defining E_2 this way is:

(†) If $\{u, v\}, \{w, x\} \in E_2$, then both u and w have degree at least $\frac{2}{3}n$ in E_1, meaning that there exists y with $\{u, y\}, \{w, y\} \in E_1$.

Note that for any two node pairs that are adjacent in E_1, i.e. for any two node pairs in E_1 of the form $\{u, v\}, \{v, w\}$, the difference between their distances is at most $\frac{1}{3}b_\epsilon$, by (5). More generally, it then follows from (†), that if $\{u, v\}, \{w, x\} \in E_2$, the difference between their distances is at most b_ϵ.

To finish the proof, let us consider how ϵ_1 should be defined so that b_ϵ suffices for the satisfaction of SI w.r.t. ϵ. For all sufficiently large n, we were given that with probability at most ϵ_1, it fails to be the case that $\star(u, \epsilon, \alpha_\epsilon)$ holds for at least $(1-\epsilon_1)n$ many nodes in G_n. For all sufficiently large n, there is also probability at most $\alpha\epsilon_1$ that it fails to be the case that, for any set of $y \geq \alpha\epsilon_1 n$ edges from G_n, every node covering is of size at least y/C. We can assume that $\alpha < 1$. When we choose the two pairs $\{u, v\}$ and $\{u', v'\}$, in order to ensure that $|d(u, v) - d(u', v')| < b_\epsilon$, we require both of these pairs to be in E_2. Since there are $n(n-1)/2$ unordered pairs, this happens with probability at least $\left(1 - \frac{2 \cdot 21\epsilon_1 n^2}{n(n-1)}\right)^2$. So for all sufficiently large n the probability that at least one of the pairs is not in E_2 is at most $85\epsilon_1$. It thus suffices that $\epsilon_1 < \epsilon/87$. □

Lemma 2.7. *If $\mathbb{G}$ is SI and US, then it is a PUMP.*

Proof. Given US, suppose that the condition PUMP fails to hold. So there exists $\epsilon_0 > 0$, such that for all $\alpha > 0$ there exist infinitely many n for which the following holds with probability at least ϵ_0: for u chosen uniformly at random $\star(u, \epsilon_0, \alpha)$ fails to hold. It suffices to show that if ϵ is chosen small enough compared to ϵ_0, then, given any $b_\epsilon \in \mathbb{N}$ there exist infinitely many n such that if u, v, u', v' are nodes chosen uniformly at random from G_n then, with probability $\geq \epsilon$, either one of the distances $d(u, v), d(u'v')$ is infinite or else $|d(u, v) - d(u', v')| > b_\epsilon$. Choose $\epsilon < (\epsilon_0)^2/4$ and suppose given $b_\epsilon > 0$. For any node u, let N_u be the $n\epsilon_0/2$ nodes closest to u (with ties broken arbitrarily) and let F_u be the $n\epsilon_0/2$ nodes furthest from u. If $\star(u, \epsilon_0, \alpha)$ fails then this happens either for reason (1), that there does not exist r with $|B_u(r)| \geq \epsilon_0 n$, or for reason (2), that there exists r with $\epsilon_0 n \leq |B_u(r)| \leq (1 - \epsilon_0)n$, such that $e(B_u(r), \overline{B_u(r)}) < \alpha|B_u(r)|$. In the case that $\star(u, \epsilon_0, \alpha)$ fails for reason (1), v is then chosen such that $d(u, v)$ is infinite with probability $> 1-\epsilon_0$, which we can assume is greater than $\epsilon_0/4$. It remains to show that if $\alpha > 0$ is chosen small enough, and if $\star(u, \epsilon_0, \alpha)$ fails for reason (2), then with probability at least $1/2$, all nodes in N_u are at a distance greater than $2b_\epsilon$ from all nodes in F_u: In that case there exists $A \in \{N, F\}$ for which we are guaranteed $|d(u, v) - d(u', v')| > b_\epsilon$ so long as v is chosen in A_u.

To establish the claim we make use of US. For a set of nodes U, we let $B_U(r)$ denote $\cup_{u\in U} B_u(r)$. For any $\epsilon_1 > 0$, let $C_{\epsilon_1/2} > 1$ be the constant guaranteed by satisfaction of US with respect to $\epsilon_1/2$, and define $f(1, \epsilon_1) = \epsilon_1/2C_{\epsilon_1/2}$. So long as a node covering of any set of $n\epsilon_1/2$ many edges must be of size at least $nf(1, \epsilon_1)$ ($< n\epsilon_1/2$), we conclude that U must be of size at least $nf(1, \epsilon_1)$ in order for it to be the case that $B_U(1) \geq n\epsilon_1$. Now we can iterate this idea. For any given $\epsilon_1 > 0$ and $k \in \mathbb{N}$, let β_0 be such that $\epsilon_1/f(k, \epsilon_1) = \beta_0^k$. Let $C_{f(k,\epsilon_1)/2}$ be the constant guaranteed by satisfaction of US with respect to $f(k, \epsilon_1)/2$, and let β_1 be the maximum of β_0 and $2C_{f(k,\epsilon_1)/2}$. Then we define $f(k + 1, \epsilon_1) = \epsilon_1/\beta_1^{k+1}$. Now define $\alpha = f(2b_\epsilon, \epsilon_0/4)$, where we can assume $b_\epsilon \in \mathbb{N}$. US ensures that for all sufficiently large n, the following occurs with probability $> 1 - \alpha$:

($\dagger_\alpha$) For all U, if $|U| < \alpha n$ then $|B_U(2b_\epsilon)| < \epsilon_0 n/4$.

If $\star(u, \epsilon_0, \alpha)$ fails for reason (2), then there exists r with $\epsilon_0 n \leq |B_u(r)| \leq (1 - \epsilon_0)n$ for which $|B_u(r + 1) - B_u(r)| < \alpha n$, which, so long as ($\dagger_\alpha$) holds, means that $r_u(1 - \epsilon_0/2) > r + 2b_\epsilon \geq r_u(\epsilon_0/2) + 2b_\epsilon$. Since we can assume $\alpha < 1/2$, this suffices to establish the required claim, that with probability at least $1/2$, all nodes in N_u are at a distance greater than $2b_\epsilon$ from all nodes in F_u. □

Lemmas 2.3, 2.6 and 2.7 give us the promised characterisation.

Theorem 2.8. *If $\mathbb{G}$ is FED, then it is strongly idemetric iff it is a PUMP.*

Along with the Erdős-Rényi random graph and the Preferential Attachment model of Barabási and Albert, the Watts-Strogatz model is one of the best known and most studied small-world models. The definition is given in the Appendix, and depends on two parameters p and m: roughly, p is the rewiring probability, while m is the number of neighbours on each side prior to rewiring.

Using the characterisation given by Theorem 2.8, we get the following result. The proof appears in the Appendix. While the proof is simpler in the case that $pm > 1$, this condition can presumably be significantly weakened.

Theorem 2.9. *The Watts-Strogatz model is strongly idemetric when $pm > 1$.*

3 Path finding

In this section, we consider how satisfaction of the idemetric property may be useful for algorithm design, and in particular for path finding.

The task of finding the shortest path between two nodes in a (possibly weighted and/or directed) graph is a fundamental problem in computer science, which has been extensively studied since the 1950s. Since often one would like to make queries for multiple pairs of nodes in the same graph, a number of variants of the problem also become significant:

The Single Source Shortest Paths (SSSP) Problem. The well known algorithm of Dijkstra [ED59] was originally formulated to give the shortest path between a single pair of nodes, but the version which is now better known solves the SSSP problem: a single node is fixed as the "source" node and the algorithm then finds shortest paths from the source to all other nodes in the graph. For a graph with n nodes and m edges, this algorithm for the SSSP problem terminates in time $O(n^2)$, but has been repeatedly improved upon. Thorup's algorithm[7] [MT99], for example, runs in time $O(m)$ for undirected weighted graphs with non-negative integer weights. For undirected unweighted graphs, a simple breadth-first search suffices, terminating in time $O(n)$ for sparse graphs.

The All Pairs Shortest Paths (APSP) Problem. For the APSP problem one is required to output a data structure encoding all shortest paths between any pair of nodes. When a pair of nodes (a, b) is queried, the data structure should return a shortest path from a to b in time $O(\ell)$, where ℓ is the number of edges on this path. The classic Floyd-Warshall algorithm [RF62, SW62] solves this problem in time $O(n^3)$ for directed graphs. Running Thorup's solution to the SSSP for each node in an undirected graph gives an $O(nm)$ algorithm, which is much more efficient when the graph is sparse. For dense directed graphs, on the other hand, Williams' algorithm [RW14] is the best known, running in time $O(n^3/2^{\Theta(\log n)^{1/2}})$.

As pointed out by Thorup and Zwick [TZ05], however, there are many contexts in which the above solutions to the APSP problem are not satisfactory. Quite simply, the time and space complexity bounds provided by these algorithms are not practical for many real-world graphs of interest. One would like algorithms for which the preprocessing time – the time required to produce the data structure – is close to linear in n, or ideally, which can be run efficiently in an online and decentralised fashion, responding to changes in the graph structure as they occur. It therefore becomes natural to consider ways in which one can reasonably make the problem easier, and thereby obtain more efficient solutions. One option is to accept *approximate* solutions, i.e. algorithms which produce paths which are close to optimal. This can be made precise by requiring paths of small *stretch*, where a path from a to b is of stretch k if it is at most k times as long as the shortest path. Another way in which to make the problem easier is to restrict attention to graphs with convenient properties. This will be a reasonable thing to do, so long as we consider properties which one can expect to be satisfied by graphs arising from real-world networks. Of course, our interest here is in the extent to which the restriction to idemetric networks us useful.

3.1 Finding short paths in idemetric networks

For now let us restrict attention to unweighted graphs, although similar arguments can be made for the weighted case. If $\mathbb{G}$ is idemetric then we can obtain an approximate solution to the APSP problem very

[7] Here and throughout the paper we work under standard RAM model assumptions.

simply as follows. For each n let G_n be a graph generated according to $\mathbb{G}$ with n nodes, and let r_n be a *beacon* which is chosen uniformly at random in G_n. We can then carry out a breadth-first search from r_n, and then again with edge directions reversed if the graph is directed, in order to find, for all $a \in G_n$, a shortest path $p(a, r_n)$ from a to r_n and a shortest path $p(r_n, a)$ from r_n to a. In network models with the small-world property, each node a can then store these short paths $p(a, r_n)$ and $p(r_n, a)$, taking space at most $O(\text{polylog}(n))$ for each node. For a_n and b_n chosen uniformly at random from the nodes in G_n, let ℓ_n be the length of the path from a_n to b_n given by concatenating $p(a_n, r_n)$ and $p(r_n, b_n)$. Since $\mathbb{G}$ is idemetric, we then have that:

$$\frac{\ell_n}{d(a_n, b_n)} \to 2 \text{ in probability.}$$

We therefore have:

Observation 3.1. *If $\mathbb{G}$ is idemetric then the all-pairs shortest path problem can be reduced to the single-source shortest path problem, so long as one is prepared to accept solutions which are of stretch $2 + o(1)$ with high probability.*

Note that if $\mathbb{G}$ is also of small diameter (as is the case for the Watts-Strogatz and PA models, for example), then even in the small proportion of cases where stretch $2 + o(1)$ fails, a short path will still be given by the simple algorithm above. If $\mathbb{G}$ is weakly idemetric, then similar results hold if we restrict to nodes in the giant component. The significance of Observation 3.1 stems from the broad class of networks to which it applies. Greedy routing, for example, can often be very efficient in networks which come with an appropriate geometric embedding, see for example [BK17]. The attraction of Observation 3.1, however, is that it can be applied in scenarios where there is no given geometric embedding (and where it is not even clear how greedy routing would be defined).

It is also worth observing, that if we restrict attention to graphs of small diameter ($O(\text{polylog}(n))$ say), then one can run a decentralised version of the algorithm above, which may be seen as a distributed version of the Bellman-Ford algorithm, and which stills runs quite efficiently in an online fashion with the nodes computing in parallel. For the sake of simplicity, let's consider graphs of bounded degree. Rather than unilaterally choosing a single beacon, it suffices to have each node choose to be a beacon with probability $\log(n)/n$. Then we can consider a stage by stage process, such that during each stage $s + 1$ and for each beacon r, each node u compares the shortest path $u_s(r)$ it has previously seen from u to r, with each of the paths given by taking the edge to a neighbour v and then following $v_s(r)$ (with information as to which nodes are beacons disseminating simultaneously). If the graph is of diameter ℓ, then correct values are obtained after at most ℓ many of these update stages, or ℓ many stages after any changes to the graph. Each stage involves each node passing $O(\text{polylog}(n))$ many bits to each of its neighbours.

The question of finding short paths in small-world graphs was addressed by Kleinberg in [JK00]. Since he considered decentralised algorithms in the absence of any preprocessing, however, his results were largely negative – the central point of that paper was that while short paths may exist, this does not mean that they can be easily found. For a variant of the the Watts-Strogatz model, now referred to as the Kleinberg model, he was able to prove that, while short ($O(\log(n))$) paths will exist between all pairs of nodes with high probability, no efficient decentralised algorithms exist for finding short paths (in the absence of preprocessing, and for a wide range of parameter inputs). To contrast with Kleinberg's negative results, we prove next that the Kleinberg model is idemetric. So for precisely the graphs which Kleinberg is able to conclude efficient decentralised algorithms do not exist in the absence of preprocessing, we establish that very efficient decentralised algorithms do exist when reasonable levels of (even decentralised) preprocessing are permitted.

We are able to conclude this simply because the network model is idemetric. First of all, let us define the model.

The Kleinberg model. The model we consider is exactly the same as that in Kleinberg [JK00]. For any square number n, we begin with a set of nodes identified with the lattice points in a $\sqrt{n} \times \sqrt{n}$ square, $\{(x, y) : x \in \{1, 2, ..., \sqrt{n}\}, y \in \{1, 2, ..., \sqrt{n}\}\}$. The *lattice distance* between two nodes (x, y) and (r, s) is defined to be the number of lattice steps separating them when we fix periodic boundary conditions: $d_\ell((x, y), (s, t))$ is the minimum value of $|s' - x'| + |t' - y'|$, where the minimum is taken over all values $x' = x \pm \sqrt{n}, y' = y \pm \sqrt{n}, s' = s \pm \sqrt{n}, t' = t \pm \sqrt{n}$. For a universal constant $p \geq 1$, each node u has a directed edge to every other node within lattice distance p. For universal constants $q \geq 0$ and $r \geq 0$, we also construct directed edges from u to q other nodes (the long-range contacts) using independent random trials; the ith directed edge from u has endpoint v with probability proportional to $[d_\ell(u, v)]^{-r}$ (to obtain a probability distribution, we divide this quantity by the appropriate normalising constant[8]). This defines the network model $\mathbb{KL}(r, p, q)$. Note that the lattice distance between two nodes $d_\ell(a, b)$ may be quite different than $d(a, b)$.

Since the Kleinberg model gives directed graphs, we cannot apply Theorem 2.8 in order to prove that it is idemetric. We shall give a more direct proof, which is also more informative since it allows us to deduce the precise function f (see Definition 1.1) with respect to which the condition of being idemetric holds. As well as its relationship to Observation 3.1, the fact that the Kleinberg model is idemetric, is of significant interest in its own right.

Theorem 3.2. *For all $p, q \in \mathbb{N}$ with $p, q \geq 1$, and for all $r \in \mathbb{R}$ with $r \in [0, 2)$, the random network model $\mathbb{KL}(r, p, q)$ is idemetric.*

Proof. We consider first the case $p = q = 1$. Once we have dealt with this case, generalising to arbitrary $p, q \geq 1$ will then be straightforward. The proof for the case $p = q = 1$ is split into four sections A-D, and we then consider the general case $p, q \geq 1$ in Section E.

(A) The setup. Let a and b be nodes chosen uniformly at random from amongst the nodes in G, a graph with n nodes generated according to $\mathbb{KL}(r, p, q)$. We can consider the sets of nodes $A_1^*, A_2^*, \ldots$ and $B_1^*, B_2^* \ldots$, where:

$$A_i^* = \{u : d(a, u) = i - 1\}, \qquad B_i^* = \{u : d(u, b) = i - 1\}.$$

We are interested in finding the least i such that $b \in A_i^*$, and so are interested in the values $A_i = |A_i^*|$ (and later will also be interested in $B_i = |B_i^*|$). Clearly $A_1^* = \{a\}$, so $A_1 = 1$, and then A_2^* consists of the four lattice neighbours of a, together with the node which is the outbound long-range contact of a. As we continue to consider A_i^* and A_i for larger values of i, the process is complicated, however, by potential collisions: distinct nodes u and v may have outbound long-range contacts which are near to each other. The outbound long-range contact of a *could* be one of its four lattice neighbours, for example, or two of the lattice neighbours of a could have long-range contacts which are within distance one of each other. The latter possibility could then lead to double counting in A_4, unless one is careful. To give a useful upper bound for A_i, we therefore consider a simplified process in which, roughly speaking, every long-range contact appears on a new two dimensional lattice at an infinite lattice distance from the previous node.

[8]For the case $r = 0$, we fix the convention that $0^0 = 1$, so that the long-range outbound contact of a node u could be itself.

Figure 2: A schematic depiction of the growth of the sequence $\{C_i\}_{i\in\mathbb{N}}$.

Formally, we can simply consider the sequence C_i, where $C_i = 0$ for $i < 0$ and, for $i \geq 0$:

$$C_0 = 1, \;\; C_{i+1} = C_i + \sum_{j\geq 1} C_{i-j} \cdot 4j = C_i + 4C_{i-1} + 8C_{i-2} + 12C_{i-3} \cdots \tag{6}$$

So, as depicted in Figure 2, $C_1 = 1, C_2 = 5, C_3 = 17, C_4 = 57$, and so on, and C_i is the value that A_i would take were it not for the possibility of collisions, as described above.

The enumeration of $\cup_i C_i^*$. It will also be useful to have a counterpart C_i^* to A_i^* and B_i^*, so that C_i^* is a set of nodes with $|C_i^*| = C_i$. To this end, we consider 3-dimensional points (x, y, z), which we shall call *3-nodes*, and specify that the lattice distance between (x, y, z) and (x', y', z') is infinite when $z \neq z'$, and is equal to the lattice distance between (x, y) and (x', y') otherwise. If $a = (x_a, y_a)$, we let $C_1^* = \{(x_a, y_a, 0)\}$. In order to enumerate C_{i+1}^*, we take each element $u = (x, y, z)$ of C_i^* in turn and proceed as follows.

1. Enumerate into C_{i+1}^* all 3-node lattice neighbours of u (i.e. those 3-nodes at lattice distance 1) which have not already been enumerated into $\cup_{j\leq i+1} C_j^*$.
2. Let $v = (x', y')$ be the outbound long-range contact of (x, y). Let z' be such that no nodes with third coordinate z' have been enumerated into $\cup_{j\leq i+1} C_j^*$ before, and enumerate (x', y', z') into C_{i+1}^* as the outbound long-range contact of u.

We say that the 3-node $u = (x, y, z) \in C_i^*$ *corresponds* to the node (x, y). Note that, due to collisions, it may be the case that a node (x, y) has several 3-nodes in C_i^* corresponding to it. The process above specifies an enumeration of $\cup_i C_i^*$, and we also consider it to specify an enumeration of $\cup_i A_i^*$ in the obvious way: nodes in this set are enumerated in the order of their first corresponding 3-nodes in $\cup_i C_i^*$. Distances between 3-nodes are defined in the standard way, in terms of the number of edges on the shortest directed path between two nodes: there are directed edges from each 3-node u to each of its lattice neighbours, i.e. those v such that $d_\ell(u, v) = 1$, and a directed edge from u to its outbound long-range contact. For 3-nodes u and v, we say that v is a descendant of u if u is a 3 node in C_i^*, v is a 3-node in C_j^* for some $j > i$ and $d(u, v) = j - i$. We will normally be interested in the descendants of u, only in the case that u is a long-range outbound contact.

It follows immediately from the definitions that for all i, $A_i \leq C_i$. Since $C_i = C_{i-1} + \sum_{j\geq 1} C_{i-j-1} \cdot 4j$ for $i > 0$, equation (6) can then be rewritten for $i > 0$:

$$C_{i+1} = 2C_i + 3C_{i-1} + 4 \sum_{j\geq 2} C_{i-j}.$$

Given that, for $i > 2$, $C_{i-1} = 2C_{i-2} + 3C_{i-3} + 4 \sum_{j\geq 2} C_{i-j-2}$, this in turn then gives, for $i > 2$:

$$C_{i+1} = 2C_i + 4C_{i-1} + 2C_{i-2} + C_{i-3}.$$

Standard techniques can then be applied in order to solve this linear recurrence relation. Since the characteristic polynomial

$$x^4 - 2x^3 - 4x^2 - 2x - 1 \tag{7}$$

has a largest root

$$\alpha \approx 3.38298, \tag{8}$$

it follows that for some constant $\rho > 0$:

$$\lim_{i\to\infty} \frac{C_i}{\rho \cdot \alpha^i} = 1. \tag{9}$$

Let us define:

$$\ell_n = \log_\alpha n.$$

It is immediate that $C_{i+1} \geq \sum_{j\leq i} C_j$. To within an additive constant, we then have that ℓ_n is the least k such that $\sum_{i=1}^{k} C_i \geq n$. Note that, for $i > 0$, $C_{i+1} \geq 2C_i$. For any $\epsilon > 0$, it follows that:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{(1-\epsilon)\ell_n} C_i}{n} = 0.$$

Since $\sum_{i=1}^{(1-\epsilon)\ell_n} A_i \leq \sum_{i=1}^{(1-\epsilon)\ell_n} C_i$, we conclude that, for any $\epsilon > 0$, with high probability:

$$d(a, b) > (1 - \epsilon)\ell_n.$$

To prove the theorem, it thus suffices to show that, for any $\epsilon > 0$, the following holds with high probability:

$$d(a, b) < (1 + \epsilon)\ell_n. \tag{10}$$

(B) Proving (10). First of all we need to establish a useful bound for the distribution on long-range contacts. Suppose that we are given arbitrary nodes u and v, but that we do not know the value of u' which is the outbound long-range contact of u. Let $\ln(n)$ denote the natural logarithm. Then we shall show that, irrespective of the given relative positions of u and v, the fact that $r \in [0, 2)$ implies:

$$\mathbb{P}(u' = v) \geq \frac{1}{4n\ln(n)}. \tag{11}$$

The proof of (11) is given in the Appendix. Towards proving (10), fix ϵ with $0 < \epsilon < 1/6$, and for the remainder of the proof let $\ell^* = \frac{(1+\epsilon)\ell_n}{2}$ (note that we drop the subscript n for convenience). The basic idea, as depicted in Figure 3, is that so long as $A^*_{\ell^*}$ and $B^*_{\ell^*}$ are reasonably large, it is very likely the case that one of the nodes in $A^*_{\ell^*}$ has some node in $B^*_{\ell^*}$ as an outbound long-range contact. This gives a short path from a to b. More precisely, what we do is to show that for any constant $\mu > 1$, the following both hold with high probability:

$$A_{\ell^*} > \mu n^{1/2}\ln(n) \qquad \text{and} \qquad B_{\ell^*} > \mu n^{1/2}\ln(n). \tag{12}$$

If this holds for $\mu > 4$ then either $\bigcup_{i=1}^{\ell^*} A^*_i$ and $\bigcup_{i=1}^{\ell^*} B^*_i$ already have non-empty intersection, which gives the required short path from a to b, or else we can apply (11) to conclude that the probability every member of $A^*_{\ell^*}$ will fail to have a member of $B^*_{\ell^*}$ as a long-range contact is bounded above by:

$$\left(1 - \frac{4n^{1/2}\ln(n)}{4n\ln(n)}\right)^{\mu n^{1/2}} = \left(1 - \frac{1}{n^{1/2}}\right)^{\mu n^{1/2}} \to e^{-\mu} \text{ as } n \to \infty.$$

Given that μ was arbitrary, (12) therefore suffices to give (10). It remains to establish (12).

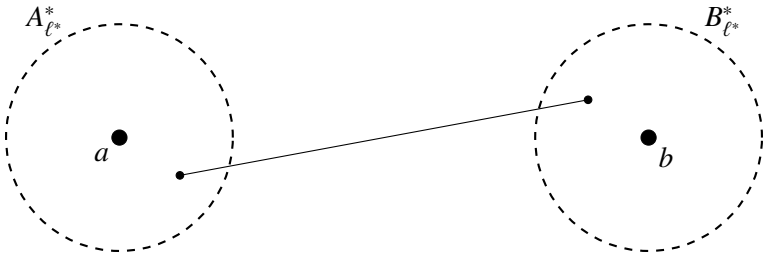

Figure 3: If A_{ℓ^*} and B_{ℓ^*} are much larger than $n^{1/2}\ln(n)$ there is likely to be an edge between them, giving a short path from a to b.

(C) Proving (12) for A_{ℓ^*}. We deal first with A_{ℓ^*} – achieving the lower bound for B_{ℓ^*} then uses most of the same ideas but is complicated by the fact that nodes do not have a fixed number of inbound long-range contacts. We deal with the case for B_{ℓ^*} in Section D of the proof. To achieve the lower bound for A_{ℓ^*}, we make some new definitions. Given the enumeration of $\cup_i C^*_i$ specified previously, we say $u \in C^*_{i+1}$ is *collision causing* if both:

(a) It is a long-range outbound contact of an element of C^*_i, and corresponds to a node in $\bigcup_{j=1}^{i+1} A^*_j$ which is within lattice distance $\log_2(n)$ of another previously enumerated element of $\bigcup_{j=1}^{i+1} A^*_j$, and;

(b) It is not the descendant of a 3-node which is already collision causing.

Since we follow the process for less than $\log_2(n)/2$ many steps (i.e. $\ell^* < \log_2(n)/2$), long-range outbound contacts at a lattice distance $> \log_2(n)$ from all other nodes will not lead to collisions of the kind discussed previously. We then define the *discounted* 3-nodes, to be all those 3-nodes in $\bigcup_{i=1}^{\ell^*} C^*_i$ which are either collision causing, or else are descendants of a collision causing 3-node. The key point of these definitions is that any two 3-nodes in $\bigcup_{i=1}^{\ell^*} C^*_i$ which are not discounted correspond to distinct nodes in $\bigcup_{i=1}^{\ell^*} A^*_i$. For each i, let $\mathtt{nd}(C^*_i)$ be all those 3-nodes in C^*_i which are not discounted, and define $\mathtt{nd}(C_i) = |\mathtt{nd}(C^*_i)|$. The basic idea is now that we want to show, for all sufficiently large n and for all $i \leq \ell^*$, that $\mathtt{nd}(C_i)$ is unlikely to be very much smaller than C_i. Since distinct nodes in $\mathtt{nd}(C^*_i)$ correspond to distinct nodes in A^*_i, and since it follows from the definition of ℓ^* that $C_{\ell^*} = \Omega(n^{(1+\epsilon)/2})$, this will suffice to give the probabilistic lower bound for A_{ℓ^*} in (12).

In order to give the probabilistic lower bound for $\mathtt{nd}(C_i)$ just discussed, we will need to bound the probability that a given long-range contact is collision causing. So suppose that we are given an arbitrary node u and an arbitrary set of nodes C. While we know which node u is, and we know the elements of C, suppose that we do not know the value of u', which is the outbound long-range contact of u. We want to show that if $|C| \leq n^{2/3}$, then:

$$\mathbb{P}(u' \in C) = o(1). \tag{13}$$

The proof of (13) is given in the appendix. Using this bound, we can now provide the probabilistic lower bound for $\mathtt{nd}(C_i)$ discussed previously. Let α be defined as before, in (8). What we want to show is that, for every $\alpha_0 < \alpha$, the following holds with high probability:

$$\mathtt{nd}(C_{\ell^*}) > \alpha_0^{\ell^*}. \tag{14}$$

Given (14), let $\alpha_0 < \alpha$ be such that $\log_\alpha \alpha_0 > (1+\epsilon/2)/(1+\epsilon)$. Then with high probability:

$$A_{\ell^*} \geq \mathtt{nd}(C_{\ell^*}) > \alpha_0^{\ell^*} = (\alpha^{\log_\alpha \alpha_0})^{(1+\epsilon)\ell_n/2} > \alpha^{(1+\epsilon/2)\ell_n/2} = n^{\frac{1}{2}+\frac{\epsilon}{4}}.$$

For any constant $\mu > 0$, the last term $n^{\frac{1}{2}+\frac{\epsilon}{4}}$ is greater than $\mu n^{1/2}\ln(n)$ for all sufficiently large n. So this gives (12) for A_{ℓ^*}, as required.

To prove (12) for A_{ℓ^*}, it therefore remains to use (13) in order to establish (14). To this end, consider the version of the recurrence relation (6) which results when, in each generation, the nodes in C_i^* only have $(1-\beta)C_i$ many long-range outbound contacts (for $\beta < 1$). We'll use E_i to denote the new resulting sequence of values:

$$E_0 = 1/(1-\beta), \quad E_{i+1} = (1-\beta)E_i + \sum_{j\geq 1} E_{i-j} \cdot 4j(1-\beta). \tag{15}$$

Of course $(1-\beta)E_i$ may not be integer valued, but the sequence given by this recurrence relation is meaningful in giving a lower bound to the number of non-discounted 3-nodes there will be in each generation, in a context where we always have some integer number $\geq (1-\beta)\mathtt{nd}(C_i)$ of non-collision causing long-range outbound contacts for elements of $\mathtt{nd}(C_i^*)$. The same algebraic manipulations as before can be applied, in order to reduce (15) to:

$$E_{i+1} = (2-\beta)E_i + (4-3\beta)E_{i-1} + (2-3\beta)E_{i-2} + (1-\beta)E_{i-3}.$$

This is gives a characteristic equation which is a function of β:

$$f(x,\beta) = x^4 - (2-\beta)x^3 - (4-3\beta)x^2 - (2-3\beta)x - (1-\beta). \tag{16}$$

Now in a neighbourhood of $x = \alpha$, $\beta = 0$, the derivates of f with respect to x and β are both positive, meaning that the largest root of f is a continuous decreasing function of β in some neighbourhood of this point. For each $\alpha_0 < \alpha$ sufficiently close to α, it follows that there exists $\beta > 0$ for which α_0 is the largest root of $f(x,\beta)$. Similarly, for each $\alpha_0 > \alpha$ sufficiently close to α, there exists $\beta < 0$ for which α_0 is the largest root of $f(x,\beta)$.

So suppose given $\alpha_0 < \alpha$ and ϵ' with $0 < \epsilon' < 0.5$. To prove (14) it suffices to show, for all sufficiently large n, that $\mathtt{nd}(C_{\ell^*}) > \alpha_0^{\ell^*}$ with probability $> 1-\epsilon'$. Choose α_1 and β, with $\alpha_0 < \alpha_1 < \alpha$, $0 < \beta < 0.5$, and such that α_1 is the largest root of $f(x,\beta)$. Suppose we are given $\mathtt{nd}(C_i^*)$ for some $i < \ell^*$. For all sufficiently large n, the fact that we chose $\epsilon < 1/6$ in the definition of ℓ^* means that:

$$\left|\cup_{j\leq \ell^*} C_j^*\right| < n^{2/3}. \tag{17}$$

Then, according to (13), for sufficiently large n the number of outbound long-range contacts of elements of $\mathtt{nd}(C_i^*)$ which are collision causing, is stochastically dominated by:

$$X \sim \mathrm{Bin}(\mathtt{nd}(C_i), \beta/2).$$

Applying the Chernoff bound to this stochastically dominating binomial, we conclude that for some $I_\beta > 0$, and for all sufficiently large n, the probability that more than a β proportion of the outbound long-range contacts of elements of $\mathtt{nd}(C_i^*)$ are collision causing is bounded above by:

$$e^{-I_\beta \mathtt{nd}(C_i)}.$$

Now $\mathtt{nd}(C_i)$ is guaranteed to grow at a certain rate, since $\mathtt{nd}(C_i) \geq 4(i-1)$ for $i \geq 2$. We can therefore choose i_0 such that $e^{-I_\beta 4(i_0-1)} < \epsilon'/2$, and this means that with probability 1,

$$e^{-I_\beta \mathtt{nd}(C_{i_0})} < \epsilon'/2.$$

We chose $\beta < 0.5$. So if at most a β proportion of the long-range contacts of elements of $\mathtt{nd}(C^*_{i_0})$ are collision causing, we have $\mathtt{nd}(C_{i_0+1}) \geq 1.5\mathtt{nd}(C_{i_0})$. Since $\epsilon' < 0.5$ this gives:

$$e^{-I_\beta \mathtt{nd}(C_{i_0+1})} < (\epsilon'/2)^{1.5} = \epsilon' \sqrt{2\epsilon'}/4 < \epsilon'/4.$$

Then so long as at most a β proportion of the long-range contacts of elements of $\mathtt{nd}(C^*_{i_0+1})$ are collision causing, we have $e^{-I_\beta \mathtt{nd}(C_{i_0+2})} < \epsilon'/8$, and so on. Define $\Pi_{\beta,\epsilon'}$ to be the following event: for all i with $i_0 \leq i < \ell^*$, at most a β proportion of the long-range contacts of elements of $\mathtt{nd}(C^*_i)$ are collision causing. The analysis above then gives, for sufficiently large n:

$$\mathbb{P}(\Pi_{\beta,\epsilon'}) > 1 - \epsilon'.$$

To complete the argument, consider all those 3-nodes at a lattice distance of $i_0 - 1$ from a. These are the 3-nodes we considered above, which are guaranteed to belong to $\mathtt{nd}(C^*_{i_0})$. So long as $\Pi_{\beta,\epsilon'}$ holds, we can give a lower bound for the number of descendants of these nodes which belong to $\mathtt{nd}(C^*_{i_0+i-1})$ for each $i > 0$ through the series:

$$E_0 = 4(i_0 - 1)/(1 - \beta) \quad E_{i+1} = (1 - \beta)E_i + \sum_{j \geq 1} E_{i-j} \cdot 4j(1 - \beta). \tag{18}$$

There exists a constant $\rho > 0$ such that $\lim_{i\to\infty} E_i/(\rho\alpha_1^i) = 1$. Since $\alpha_0 < \alpha_1$, we can choose $i_1 > i_0$ such that $\mathtt{nd}(C^*_{\ell^*}) > \alpha_0^{\ell^*}$ whenever $\Pi_{\beta,\epsilon'}$ holds and $\ell^* > i_1$. This gives (14) as required.

(D) Proving (12) for B_{ℓ^*}. As remarked on previously, establishing (12) for B_{ℓ^*} is complicated by the fact that nodes do not have a fixed number of inbound long-range contacts. The proof is similar to the case for A_{ℓ^*}, however, and so Section D of the proof is given in the Appendix.

(E) The case $p, q \geq 1$. The generalised form of the recurrence relation (6) is given as follows. For $i < 0$, $C_i = 0$. Then $C_0 = 1/q$ and, for $i \geq 0$:

$$C_{i+1} = qC_i + \sum_{j \geq 1} \left(4p^2(i - 1/2) + 2p\right) qC_{i-j}.$$

Using this expression for C_i, we get that, for $i > 0$:

$$C_{i+1} = (q + 1)C_i + (2p^2 + 2p - 1)qC_{i-1} + \sum_{j \geq 2} 4p^2 qC_{i-j}.$$

For $i > 2$, this means that $C_{i-1} = (q + 1)C_{i-2} + (2p^2 + 2p - 1)qC_{i-3} + \sum_{j\geq 2} 4p^2 qC_{i-j-2}$. For $i > 2$ we then have:

$$C_{i+1} = (q + 1)C_i + \left((2p^2 + 2p - 1)q + 1\right) C_{i-1} + (4p^2 q - q - 1)C_{i-2} + (2p^2 - 2p + 1)qC_{i-3}.$$

The largest root α of the corresponding characteristic polynomial is the same as the largest eigenvalue of the matrix:

$$\boldsymbol{M} = \begin{bmatrix} q + 1 & (2p^2 + 2p - 1)q + 1 & 4p^2 q - q - 1 & (2p^2 - 2p + 1)q \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{19}$$

That the largest eigenvalue α of $\boldsymbol{M}$ is positive and real (and occurs with multiplicity 1) follows directly from the Perron-Frobenius Theorem for non-negative matrices. In a manner precisely analogous to the proof for $p = q = 1$, we conclude that, for some constant $\rho > 0$:

$$\lim_{i\to\infty} \frac{C_i}{\rho \cdot \alpha^i} = 1.$$

The remainder of the proof then goes through with only the obvious required modifications. The definitions of C_i^* and D_i^* (where D_i^* is defined in Section D of the proof, in the Appendix) have to be adjusted to incorporate the increased number of neighbours, for example, and the expression $\log_2(n)$ must be replaced throughout with $p\log_2(n)$. The expected number of inbound long-range contacts for each node is now q, and distributions on the number of inbound long-range contacts must be adjusted accordingly. The general form of (16) is:

$$x^4 - (q(1-\beta)+1)x^3 - \Big((2p^2+2p-1)q(1-\beta)+1\Big)x^2 - (4p^2q(1-\beta) - q(1-\beta) - 1)x - (2p^2-2p+1)q(1-\beta).$$

Again we have that if α is the largest root then in a neighbourhood of $x = \alpha, \beta = 0$, the derivates of f with respect to x and β are both positive. □

So the proof establishes that the Kleinberg model is idemetric with respect to $f = \log_\alpha n$, where α is the largest eigenvalue of the matrix $\boldsymbol{M}$ given in (19).

3.2 Memory efficient routing with small stretch

In this subsection we consider deterministic distributed routing algorithms, but focus instead on the total routing information required in order to achieve paths of small stretch for almost all pairs. As before we shall consider methods which give short paths even when stretch $2 + o(1)$ fails, so long as the graph is of small diameter (as is the case for most small-world models which give connected graphs with high probability). The standard set-up we consider is as in [GG01], and we refer the reader to that paper for the precise details of the framework. The nodes in a graph are given labels in $\{1, \ldots, n\}$ and each edge adjacent to a node u is given an *output port* number in $\{1, \ldots, \deg(u)\}$ with respect to u, where $\deg(u)$ is the degree of u. Our results here are strong in the sense that lower bounds on routing information allow arbitrary relabelling of nodes and ports, while upper bounds assume one is given arbitrary and fixed labelings of nodes and ports. A routing function R is a pair (P, H), where P is the *port* function and H is the *header* function. For any two distinct nodes u and v, R produces a path $u = u_0, \ldots, u_k = v$ of nodes, along with a sequence $h_0, \ldots, h_k$ of headers and a sequence $p_0, \ldots, p_k$ of ports. We stipulate that $h_0 = v$ and $p_0 = 0$. In general, a message with header h_i, arriving at node u_i through input port p_i, is forwarded to the output port $P(u_i, p_i, h_i) = p_{i+1}$ with a new header $H(u_i, p_i, h_i) = h_{i+1}$. Routing functions are defined as a set of local routing functions $R_u = (P_u, H_u)$, one for each node u. The memory requirement of node u is the length of the smallest program that computes R_u. The total memory requirement for a routing function is the sum of the memory requirements of the nodes.

The first observation to be made here, is that we can employ a similar trick to that described in Section 3.1 for idemetric network models. To deal with G_n we pick a beacon r_n uniformly at random and perform a breadth-first search from r_n, defining a spanning tree T. At each node u other than r_n we simply store a port number $p(u)$ which takes them the first step on a shortest path to r_n. At r_n, meanwhile, we store $O(n\log n)$ bits of information: for each other node u, r_n stores the predecessor v in T together with v's port number

for the edge to u. For (u, v) as input, u begins by sending a header of the form $(v, 0)$ to the node at its port $p(u)$. The last coordinate 0 indicates that we are in the phase of the routing process where a path is being followed to the beacon. The next node v therefore passes on the same header, and so on, until the beacon is reached. Upon reaching r_n, the beacon then finds the sequence of port numbers $q_0, \ldots, q_d$ given by T and leading to v if followed in turn, before passing on the header $(q_1, \ldots, q_d, 1)$ to the node w at its output port q_0. Since the last coordinate 1 indicates that we are now moving away from the beacon, w then passes the header $(q_2, \ldots, q_d, 1)$ to the next node, and so on, until the header is the singleton (1) and v is reached. This argument gives the following result, where we say that a statement holds for almost all pairs if it holds with high probability for r_n, u, v chosen uniformly at random.

Theorem 3.3. *For idemetric networks total memory requirement $O(nlog(n))$ suffices to achieve stretch $2 + o(1)$ for almost all pairs.*

The next result shows that, in fact, the problem is provably easier in the idemetric case than in the general case.

Theorem 3.4 (Essentially Gavoille and Gengler, [GG01])**.** *Total memory requirement $\Omega(n^2)$ is required for general networks, to achieve stretch < 3 routing for almost all pairs.*

Proof. This can be obtained by a modification of the proof given by Gavoille and Genger, which proves the same statement when stretch < 3 is required for all rather than almost all pairs. In the Appendix we describe the necessary modifications. □

4 Discussion

Many real networks have strongly concentrated distance distributions: Figure 4 gives some examples derived from the Social Computing Data Repository. Rather than real data, however, the focus of this paper has been on *network models*. We have seen here that most (if not all) of the well known small-world network models are at least weakly idemetric for a wide range of parameters. While the terminology itself was not used, this had previously been shown for the Erdős-Rényi, Preferential Attachment, Chung-Lu, and Norros-Reittu models, as well as for bounded degree expanders. In Sections 2 and 3, we gave proofs for the Watts-Strogatz and the Kleinberg models. Modulo the sparsity condition FED, in fact, we were able to show that being strongly idemetric is equivalent to the weak expander condition PUMP for undirected and unweighted network models. As well as providing further evidence that it is a common property, this characterisation gives an easy way of providing short proofs that network models are idemetric. It remains, however, to thoroughly investigate the extent to which the prevalence of (weakly/strongly) idemetric models is reflected in real-world data. It would be of considerable interest to find that it is almost universally the case (as it seems to be for models) that real-world networks expand through mechanisms which cause the distance distributions to become proportionately more concentrated as the network grows. If, on the other hand, there are significant instances in which the distance distribution is seen not to remain concentrated as the size of the network increases, then this raises an obvious question: can one define natural small-world models, which produce realistic graphs (in the sense that they produce graphs with the standard properties associated with real-world networks, such as being scale-free, having high clustering coefficients etc) and which are *not* at least weakly idemetric?

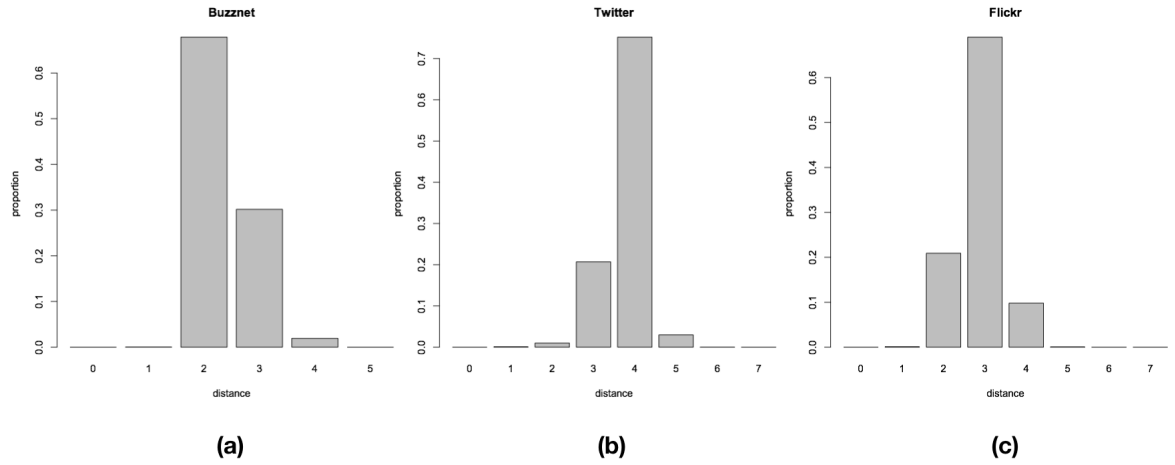

Figure 4: These graphs show distance distributions for three datasets from the Social Computing Data Repository, which can be found at http://socialcomputing.asu.edu/pages/datasets. Distance is displayed on the x-axis, while the y-axis shows the proportion of pairs at that distance: (a) Buzznet (101168 nodes, 4284534 edges); (b) Twitter (11316811 nodes, 85331846 edges); (c) Flickr (80513 nodes, 5899882 edges).

Data accessibility. All experimental data is presented in Figures 1 and 4 in its entirety.

Competing interests. There are no competing interests to declare.

Authors' contributions. All authors contributed to the construction of the proofs in this paper.

Acknowledgements. None.

Funding statement. Lewis-Pye was partially supported by a Royal Society University Research Fellowship. This research was partially supported by NSFC grant Nos 61772503 and 2014CB340302. Barmpalias was supported by the 1000 Talents Program for Young Scholars from the Chinese Government No. D1101130.

References

[AJB99] Réka Albert, Hawoong Jeong & Albert-László Barabási. The diameter of the World Wide Web. *Nature*, 401, 130, 1999.

[BA99] Albert-László Barabási & Réka Albert. Emergence of scaling in random networks. *Science*. 286 (5439): 509–512, 1999.

[BR04] Béla Bollobás & Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1), 5-34, 2004.

[BK17] Karl Bringmann, Ralph Keusch, Johannes Lengler, Yannic Maus & Anisur R. Molla. Greedy routing and the algorithmic small-world phenomenon. *2017 ACM Symposium on Principles of Distributed Computing (PODC '17)*, Washington, DC, USA, July 25-27, 371–380, 2017

[CGH18] Francesco Caravenna, Alessandro Garavaglia & Remco van der Hofstad. Diameter in ultra-small scale-free random graphs. *Random Structures and Algorithms*, to appear.

[CG08] Fan Chun & Ronald Graham. Quasi-random graphs with given degree sequences. *Random Structures and Algorithms*, 32 (1), 1–19, 2008.

[CGW89] Fan Chung, Ronald Graham & Richard Wilson. Quasi-random graphs. *Combinatorica*, 9 (4), 345–362, 1989.

[CL02] Fan Chung & Linyuan Lu. The average distance in a random graph with given expected degrees. *Proceedings of the National Academy of Sciences USA*, 99 (25), 15879–15882, 2002.

[CL03] Fan Chung & Linyuan Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1 (1), 91–113.

[DMM12] Steffen Dereich, Christian Mönch & Peter Mörters. Typical distances in ultrasmall random networks. *Advances in Applied Probability*, 44(2), 583–601, 2012.

[ED59] Edsger Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1: 269–271, 1959.

[DHH10] Sander Dommers, Remco van der Hofstad & Gerard Hooghiemstra. Diameters in preferential attachment graphs. *Journal of Statistical Physics*, 139, 72–107, 2010.

[RD06] Rick Durrett. Random Graph Dynamics. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, New York, USA, 2006.

[ER60] Paul Erdős & Alfréd Rényi. On the evolution of random graphs. *Publication of The Mathematic Institute of The Hungarian Academy of Sciences*, 17–61, 1960.

[RF62] Robert Floyd. Algorithm 97. *Communications of the Association for Computing Machinery*, volume 5, issue 6, page 345, 1962.

[GG01] Cyrill Gavoille & Marc Gengler. Space-Efficiency for Routing Schemes of Stretch Factor Three. *Journal of Parallel and Distributed Computing*. 61, 697–687, 2001.

[GP96] Cyril Gavoille & Stephane Perennes. Memory requirement for routing in distributed networks. *15th Annual ACM symposium on Principles of Distributed Computing (PODC)*, 125-133, 1996.

[RH16] Remco van der Hofstad. Random Graphs and Complex Networks. Volume 1. Cambridge University Press, 2016.

[RH] Remco van der Hofstad. Random Graphs and Complex Networks. Volume 2. Cambridge University Press, to appear.

[HHM05] Remco van der Hofstad, Gerard Hooghiemstra & Piet van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures & Algorithms*, 27(1), 76–123, 2005.

[HHZ07] Remco van der Hofstad, Gerard Hooghiemstra & Dmitri Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electronic Journal of Probability*, 12(25), 703–766, 2007

[HLW06] Shlomo Hoory, Nathan Linal & Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43 (4), 439–561, 2006.

[JK00] Jon Kleinberg. The small world phenomenon: an algorithmic perspective. *Proceedings of the 32nd ACM Symposium on Theory of Computing*, 163–170, 2000.

[JSW09] Jon Kleinberg, Aleksandrs Slivkins & Tom Wexler. Triangulation and embedding using small sets of beacons. *Journal of the ACM (JACM)*, Volume 56 Issue 6, 2009.

[MK89] Manfred Kochen, editor. *The small world*. Norwood, N.J. Ablex Pub., 1989.

[SM67] Stanley Milgram. The Small World Problem. *Psychology Today*. 2: 60–67, 1967.

[NR06] Ilkka Norros & Hannu Reittu. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38 (1), 59–75, 2006.

[BP94] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures and Algorithms*, 5, 337?347, 1994.

[MT99] Mikkel Thorup. Undirected single-source shortest paths with positive integer weights in linear time. *Journal of the Association for Computing Machinery*, 46 (3): 362–394, 1999.

[TZ05] Mikkel Thorup & Uri Zwick. Approximate distance oracles. *Journal of the Association for Computing Machinery*, 52 (1), 1–24, 2005.

[SW62] Stephen Warshall. A theorem on Boolean matrices. *Journal of the Association for Computing Machinery*, 9:11–12, 1962.

[WS98] Duncan Watts & Steven Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393, 440–442, 1998.

[RW14] Ryan Williams. Faster all-pairs shortest paths via circuit complexity. *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, 664–673, 2014.

5 Appendix

5.1 The definition of the Watts-Strogatz model

The model is defined via a two stage process. We begin with a regular ring lattice, i.e. n nodes are arranged in a circle, such that each has an edge to the m nearest nodes on each side. So if the nodes are labelled $0, \ldots, n-1$, then there is an edge between distinct nodes i and j iff $|i-j| \leq m \bmod n$. For every node i, we then take each edge (i, j) such that $0 < (j-i) \bmod n \leq m$ (i.e. such that j is 'to the right' of i) and 'rewire' it with probability $p \in [0, 1]$. Rewiring is done by replacing the edge (i, j) with an edge (i, k), where k is chosen with uniform probability from all possible values that avoid self-loops and multi-edges. Whether or not the edge is rewired, i is referred to as the 'fixed node' of the edge.

5.2 The proof of Theorem 2.9

We use the same notation as in the definition of the Watts-Strogatz model in Section 5.1. So n, m and p are as defined there. To prove the theorem, we must establish that when $pm > 1$, the Watts-Strogatz model is both FED and a PUMP.

Establishing that the model is FED. When rewiring occurs and the edge (i, j) is replaced with an edge (i, k), we refer to k as the 'new node' of the edge (i, k). If u is chosen uniformly at random then $\deg(u)$ can be expressed as the sum of two terms $N(u) + F_n(u)$, where $N(u) \sim m + \text{Bin}(m, 1-p)$ and $F_n(u)$ is the number of edges for which u is the new node.

For reasons that will soon become clear, our first aim is to show that for every $\epsilon > 0$, there exists a constant c_ϵ such that with probability $> 1 - e^{-c_\epsilon n}$:

($\dagger_\epsilon$) All nodes are of degree $< \epsilon n$.

So suppose $\epsilon > 0$. Since the total degree of all nodes is $2mn$, it immediately follows that at most a constant number $2mn/\epsilon n = 2m/\epsilon$ of nodes can have degree $\geq \epsilon n$. Consider the rewiring process as it proceeds around the ring, starting with node 0, then moving on to node 1, and so on. When we come to consider the rewirings for $v \neq u$, suppose that v has degree $\leq \epsilon n$. In that case, no matter what has happened previously during the rewiring process, the probability that v rewires an edge for which u becomes the new node is less than $pm/(1-2\epsilon)n$ for all sufficiently large n. For all sufficiently large n, $F_n(u)$ is thus stochastically dominated by $X \sim 2m/\epsilon + \text{Bin}(n, pm/(1-2\epsilon)n)$. Applying Chernoff bounds, we conclude that for some constant c, the probability u has degree $\geq \epsilon n$ is less than e^{-cn}. Since u was chosen uniformly at random, the probability that at least one node has degree ϵn is therefore bounded above by ne^{-cn}. We can thus choose c_ϵ as required.

Next we want to show that $F_n(u)$ converges in distribution to $Y \sim \text{Poisson}(pm)$.[9] To this end, consider again the rewiring process as it proceeds around the ring. Suppose $\epsilon > 0$ and consider first the case that $\dagger_\epsilon$ holds. When we come to consider the rewirings for $v \neq u$, the probability that v rewires an edge for which u becomes the new node is less than $pm/(1-2\epsilon)n$, so long as n is sufficiently large. For all sufficiently large n, $F_n(u)$ is thus stochastically strictly dominated by $Y_{\epsilon,n} \sim \text{Bin}(n, pm/(1-2\epsilon)n)$ if we condition on satisfaction of $\dagger_\epsilon$. Since the probability that $\dagger_\epsilon$ fails is exponentially small, however, (and since $F_n(u)$ is

[9] Poisson(pm) denotes a Poisson distribution with mean pm.

always less than mn) we conclude that $F_n(u)$ is stochastically dominated by $Y_{\epsilon,n} \sim \text{Bin}(n, pm/(1-2\epsilon)n)$ for all sufficiently large n, even if we don't condition on satisfaction of $\dagger_\epsilon$. Let $\epsilon_n \to 0$ be chosen such that $F_n(u)$ is stochastically dominated by $Y_{\epsilon_n,n}$, and define $Y_n = Y_{\epsilon_n,n}$. As $n \to \infty$, Y_n converges in distribution to $Y \sim \text{Poisson}(mp)$. In general, for any discrete random variables X_n, Y_n and Y, if we have that (a) for all sufficiently large n, Y_n stochastically dominates X_n, (b) every X_n has mean M, and (c) $\mathbb{E}(Y_n) \to M$, while the sequence Y_n converges in distribution to Y, then it follows that the sequence X_n converges in distribution to Y. The expected degree of u is $2m$, so since $\mathbb{E}(N(u)) = 2m - pm$, this immediately implies $\mathbb{E}(F_n(u)) = mp$. It thus follows that $F_n(u)$ converges in distribution to $Y \sim \text{Poisson}(mp)$, as claimed.

Let $Y \sim \text{Poisson}(mp)$, and define $\boldsymbol{p}(d) = \mathbb{P}(N(u) + Y = d)$. We have shown already that (F1) from Definition 2.1 holds with respect to $\boldsymbol{p}$. That (F2) holds is immediate, since $\sum_d d\boldsymbol{p}_n(d) = 2m$, which is the mean of the distribution $\boldsymbol{p}$.

Establishing that the model is a PUMP. It suffices to verify that the definition is satisfied for all sufficiently small $\epsilon > 0$. In order to see this, note that while the condition (i) (that there exists r with $|B_u(r)| \geq \epsilon n$) is stronger for larger ϵ, the fact that $\alpha_\epsilon > 0$ means that when both (i) and (ii) from the definition are satisfied with respect to $\epsilon > 0$, (i) is also satisfied for all $\epsilon' > \epsilon$ with $\epsilon' < \frac{1}{2}$.

For each u and $r \geq 1$, we let $B^*_r(u) = B_r(u) - B_{r-1}(u)$. First of all we look to establish the following lemma:

Lemma 5.1. *For all $\epsilon > 0$, and $N \in \mathbb{N}$, there exists $M \in \mathbb{N}$ such that for all sufficiently large n, if u is chosen uniformly at random then $|B^*_M(u)| \geq N$ with probability $> 1 - \epsilon$.*

Proof. Since $pm > 1$, we have $m \geq 2$. For u which is chosen uniformly at random from G_n, and $r \in \mathbb{N}$, define $C_0(u) = \{u\}$. Given $C_r(u)$, define $C_{r+1}(u)$ to be the union of $C_r(u)$ and the set of all nodes j such that there exists $i \in C_r(u)$ and an edge (i, j) for which i is the fixed node. For each $r \in \mathbb{N}$, define $C^*_{r+1}(u) = C_{r+1}(u) - C_r(u)$. For $r \geq 2$, let q_r be the number of nodes $i \in C_{r-2}$ which are the fixed node of at least one edge which was rewired. For each $r \geq 2$ and each possible value of q_r it holds with high probability that $|C^*_r(u)| > 2q_r$, and that $C^*_r(u) \subseteq B^*_r(u)$. For fixed N, the probability that $q_r < N$ tends to 0 as $r \to \infty$ (since the rewirings are independent events). So the result holds for M large compared to N. □

Now let $\epsilon > 0$ be sufficiently small that $pm(1-\epsilon) > 1$. If $|B_u(r)^*| = N$, then let $v_1, \ldots, v_N$ be an enumeration of $B_u(r)^*$. In order to put a probabilistic lower bound on $|B_u(r+1)^*|$, we can take each v_i in turn, and consider the expected value of ρ_i, which is the number of nodes outside $B_u(r)$ which rewire an edge for which v_i is the new node, and which do not rewire any edges for which some v_j with $j < i$ is the new node. If $|B_u(r+1)| < \epsilon n$, as we come to consider v_i, the expected value of ρ_i is greater than $pm(1-\epsilon)$, no matter the values for ρ_j, $j < i$. We conclude that $|B_u(r+1)^*|$ stochastically dominates X which is a sum of N i.i.d random variables, each with mean $pm(1-\epsilon)$. If we choose $\beta > 1$ with $\beta < pm(1-\epsilon)$, then, applying Chernoff bounds, we get that for some $I_\beta > 0$, either $|B_u(r+1)| \geq \epsilon n$, or else the probability that $|B_u(r+1)^*| < \beta|B_u(r)^*|$ is bounded above by $e^{-I_\beta N}$. Applying the same argument, and assuming that $|B_u(r+1)^*| \geq \beta|B_u(r)^*|$, we then have that either $|B_u(r+2)| \geq \epsilon n$, or else the probability that $|B_u(r+2)^*| < \beta|B_u(r+1)^*|$ is bounded above by $e^{-I_\beta N\beta}$. Iterating this argument, we conclude that if $|B_u(M)^*| = N$ then the probability that there fails to exist any r for which $B_u(r) \geq \epsilon n$, is bounded above by:

$$e^{-I_\beta N} + e^{-I_\beta N\beta} + e^{-I_\beta N\beta^2} + \ldots$$

For any $\epsilon > 0$, the expression above is less than ϵ for all sufficiently large N. So, for sufficiently small $\epsilon > 0$, (i) from the definition of PUMP holds with probability $> 1 - \epsilon$ for all sufficiently large n, by Lemma 5.1. In fact, a more detailed analysis of the argument above allows us to draw a stronger conclusion. If $|B_u(M)^*| = N$, and for all $r \in [M, M']$ we have $|B_u(r+1)^*|/|B_u(r)^*| \geq \beta$, then for each β_1 with $1 < \beta_1 < \beta$ it holds for all sufficiently large M' that $|B_u(M')^*|/|B_u(M')| > 1 - \beta_1^{-1}$. We conclude that for all sufficiently small $\epsilon > 0$, there exists $\alpha_1 > 0$ such that the following holds with probability $> 1 - \epsilon$ for all sufficiently large n: there exists r with $|B_u(r)| \geq \epsilon n$ and such that $|B_u(r)^*| \geq \alpha_1 n$. In order to conclude the proof, suppose $|B_u(r)^*| \geq \alpha_1 n$, and under the assumption that $|B_u(r+1)| \leq (1-\epsilon)n$ consider the number of nodes outside $B_u(r)$ which rewire an edge for which the new node is in $B_u(r)^*$. For some $J_\epsilon > 0$ the probability that this number is not at least $\epsilon\alpha_1 n$ is bounded above by $e^{-J_\epsilon n}$. Condition (ii) from the definition therefore holds for all sufficiently large n, if we set $\alpha_\epsilon = \epsilon$.

5.3 The proof of (11)

In order to establish (11), note that $\mathbb{P}(u' = v)$ is minimised by taking v at a maximum possible lattice distance from u, and by taking r as large as possible. So it suffices to prove the result for $r = 2$, and when v is at a maximum lattice distance from u. In that case:

$$\mathbb{P}(u' = v) = d_\ell(u, v)^{-2} / \sum_{v' \neq u} d_\ell(u, v')^{-2}. \tag{20}$$

Then we have:

$$\sum_{v' \neq u} d_\ell(u, v')^{-2} \leq \sum_{j=1}^{\sqrt{n}} (4j)(j^{-2}) = 4 \sum_{j=1}^{\sqrt{n}} j^{-1} \leq 4(1 + \ln(\sqrt{n})) \leq 4\ln(n), \tag{21}$$

where the second inequality follows from the general bound $\sum_{j=1}^{x} j^{-1} \leq (1 + \ln(x))$. Then (20) and (21) combine to give (11), as required.

5.4 The proof of (13)

Recall that we are given an arbitrary node u and an arbitrary set of nodes C. While we know which node u is, and we know the elements of C, we do not know the value of u', which is the outbound long-range contact of u. We want to show that if $|C| \leq n^{2/3}$, then $\mathbb{P}(u' \in C) = o(1)$.

In order to prove (13), we define the following sets of nodes:

$$H_i = \{x : \ 2^{i-1} \leq d_\ell(u, x) < 2^i\}, \qquad I_k = \{x : \ d_\ell(u, x) = k\}.$$

The number of nodes in $\bigcup_{j=1}^{i} H_j$ is $2^{i+1}(2^i - 1)$, so long as n is sufficiently large. We say that H_i is *complete* for n when the size of $\bigcup_{j=1}^{i} H_j$ takes this maximum value. Similarly, we say that I_k is complete for n when it is of size $4k$. In what follows we shall be concerned with $\mathbb{P}(u' \in H_i)$ and $\mathbb{P}(u' \in I_k)$ for various i and k. When we consider such values, the implicit assumption will always be that n is sufficiently large that H_i and I_k are complete for n.

To establish (13), our aim is to show that, for each $r \in [0,2)$, there exists $\kappa_r > 0$ such that:

$$\frac{\mathbb{P}(u' \in H_{i+1})}{\mathbb{P}(u' \in H_i)} \geq 2^{\kappa_r} \text{ for all sufficiently large } i. \tag{22}$$

If H_i is complete for n then $|\cup_{j=1}^{i} H_j| = 2^{i+1}(2^i - 1)$. Note also that $\lim_{i\to\infty} 2^{i+1}(2^i-1)/(2\cdot 4^i) = 1$. It follows that for $\ell = \log_4(n)$ there exists a constant c such that, for all sufficiently large n:

$$\left|\cup_{j=1}^{\frac{2}{3}\ell+c} H_j\right| > n^{2/3}, \text{ while simultaneously } H_{\ell-c} \text{ is complete for } n.$$

Given (22), this means that (13) holds when C is chosen to be a set of $n^{2/3}$ many nodes which are as near as possible to u, and therefore holds for all C with $|C| \leq n^{2/3}$.

To prove (22), choose β with $0 < \beta < 2 - r$. For each $I_k \subseteq H_i$ consider $I_{2k} \cup I_{2k+1} \subseteq H_{i+1}$. For all sufficiently large k:

$$\frac{\mathbb{P}(u' \in I_{2k+1})}{\mathbb{P}(u' \in I_{2k})} > 2^{-\beta}. \tag{23}$$

Also:

$$\frac{\mathbb{P}(u' \in I_{2k})}{\mathbb{P}(u' \in I_k)} = 2^{1-r}. \tag{24}$$

Combining (23) and (24), we get that for all sufficiently large k:

$$\frac{\mathbb{P}(u' \in I_{2k+1}) + \mathbb{P}(u' \in I_{2k})}{\mathbb{P}(u' \in I_k)} > 2^{2-r-\beta},$$

so $\kappa_r = 2 - r - \beta$ suffices.

5.5 (D) Proving (12) for B_{ℓ^*}

We begin by defining the sets D_i^*, which are to B_i^* as C_i^* is to A_i^*. As we enumerate the sets D_i^*, it is useful to monitor whether or not we have considered the long-range inbound contacts of each node before: all nodes start as *unseen* and will be labelled *seen* during the enumeration once we have considered their long-range inbound contacts. If $b = (x_b, y_b)$, we let $D_1^* = \{(x_b, y_b, 0)\}$. In order to enumerate D_{i+1}^*, we take each element $u = (x, y, z)$ of D_i^* in turn and proceed as follows.

1. Enumerate into D_{i+1}^* all 3-node lattice neighbours of u which have not already been enumerated into $\cup_{j\leq i+1} D_j^*$.
2. We divide into two cases. If (x, y) is not labelled as *seen* then proceed according to (a) below. Otherwise proceed according to (b).
   (a) Let $v_1, \ldots, v_k$ be the inbound long-range contacts of (x, y), where $v_j = (x_j, y_j)$. For each $j \in [1, k]$ in turn, let z_j be such that no 3-nodes with third coordinate z_j have been enumerated into $\cup_{j\leq i+1} D_j^*$ before, and enumerate (x_j, y_j, z_j) into D_{i+1}^* as an inbound long-range contact of u. Label (x, y) as *seen*.

(b) Let k be sampled from a distribution $\mathrm{Bin}(n, 1/n)$ (independent from all other distributions considered). Let $z_1, ..., z_k$ be distinct and such that, for each $i \in [1, k]$, no node with third coordinate z_j has been enumerated into $\cup_{j \le i+1} D^*_j$ before. For each $i \in [1, k]$ enumerate $(0, 0, z_i)$ into D^*_{i+1} as an inbound long-range contact of u.

We let $D_i = |D^*_i|$. The definitions we gave previously for collision causing nodes, discounted nodes and descendants are applied to the 3-nodes in $\bigcup_i D^*_i$ in the obvious way – replacing "outbound" everywhere with "inbound", replacing A^*_i with B^*_i, and replacing $d(u, v)$ in the definition of descendant with $d(v, u)$ (we are now interested in 3-nodes as elements of D^*_i rather than C^*_i, so these definitions *replace* rather than extend the previous ones).

In order to prove (12) for B_{ℓ^*}, it suffices to establish the following analogue of (14). For every $\alpha_0 < \alpha$, the following holds with high probability:

$$\mathtt{nd}(D_{\ell^*}) > \alpha_0^{\ell^*}. \tag{25}$$

The proof would proceed much as it did for (14), except that there are now *two* non-deterministic factors influencing $\mathtt{nd}(D_i)$: as well as the fact that 3-nodes may or may not be discounted, they may also have from 0 to n many inbound long-range contacts (0 to $n-1$ if $r \neq 0$). Whereas previously it followed directly from (9) that (17) held for all sufficiently large n, now we have to provide further proof that with high probability:

$$\left|\cup_{j \le \ell^*} D^*_j\right| < n^{2/3}. \tag{26}$$

Only after (26) is established will we be in a position to conclude that, with high probability, the probabilistic bound (13) can be applied at all stages $i < \ell^*$. To prove (25) we therefore first have to prove that, for every $\alpha_1 > \alpha$, the following holds with high probability:

$$\left|\cup_{j \le \ell^*} D^*_j\right| < \alpha_1^{\ell^*}. \tag{27}$$

In order to see that (27) gives (26), choose $\alpha_1 > \alpha$ such that $\log_\alpha(\alpha_1) < 4/(3(1+\epsilon))$. Then (27) gives that with high probability:

$$\left|\cup_{j \le \ell^*} D^*_j\right| < \alpha_1^{\ell^*} = (\alpha^{\log_\alpha \alpha_1})^{\ell^*} < \alpha^{\frac{4(1+\epsilon)\ell_n}{6(1+\epsilon)}} = n^{2/3}.$$

We define $\ell^\dagger$ to be the minimum of ℓ^* and the first i such that $|\cup_{j \le i+1} D^*_j| \ge n^{2/3}$. In order to establish that with high probability $\ell^\dagger = \ell^*$ and (27) holds, we can argue much as in the proof of (14). So suppose given $\alpha_1 > \alpha$ and ϵ' with $0 < \epsilon' < 0.5$. To prove (27) it suffices to show, for all sufficiently large n, that $\left|\cup_{j \le \ell^*} D^*_j\right| < \alpha_1^{\ell^*}$ with probability $> 1 - \epsilon'$. Choose α_2 and β, with $\alpha < \alpha_2 < \alpha_1$, $0 < \beta < 0.5$, and such that α_2 is the largest root of $f(x, -\beta)$, where f is as defined in (16). Now suppose we are given $\cup_{j \le i} D^*_j$ for some $i < \ell^\dagger$ — so for some $i < \ell^\dagger$ we are told precisely the elements of D^*_j for each $j \le i$ (but we are not told the value of $\ell^\dagger$). Before we are told the long-range inbound contacts of a given node $u \in D^*_i$ during the enumeration of D^*_{i+1}, we do not know the outbound long range contacts of any nodes other than some of those we have enumerated into $\cup_{j \le i+1} D^*_i$ already, and the outbound long-range contacts of all other nodes remain uniformly and independently distributed amongst the unseen nodes. The number of inbound long-range contacts of elements of D^*_i is therefore stochastically dominated by X, which is the sum of D_i i.i.d. random variables, each with distribution $\mathrm{Bin}(n, 1/(n - n^{2/3}))$. At the same time, the number of inbound long-range contacts of elements of D^*_i stochastically dominates X', which is the sum of D_i i.i.d. random variables, each with distribution $\mathrm{Bin}(n - n^{2/3}, 1/n)$. Applying Chernoff bounds, we conclude that for some

$I_\beta > 0$, and for all sufficiently large n, the probability that the number of inbound long-range contacts of elements of D_i^* is outside the interval $[(1-\beta)D_i, (1+\beta)D_i]$ is bounded above by[10]:

$$e^{-I_\beta D_i}.$$

We can then continue to argue much as in the proof of (14). Let i_0 be such that $e^{-I_\beta 4(i_0-1)} < \epsilon'/4$. We chose $\beta < 0.5$. So if the number of inbound long-range contacts of elements of D_i^* is at least $(1-\beta)D_i$, we have $D_{i_0+1} \geq 1.5 D_{i_0}$, which means that $e^{-I_\beta D_{i_0+1}} < \epsilon'/8$, and so on. We can also choose N_0 which depends on ϵ' but is independent of n, and such that $\mathbb{P}(D_{i_0} > N_0) < \epsilon'/2$. Define $\Pi_{\beta,\epsilon'}$ to be the following event: $D_{i_0} \leq N_0$ and, for all i with $i_0 \leq i < \ell^\dagger$, the number of inbound long-range contacts of elements of D_i^* is in the interval $[D_i(1-\beta), D_i(1+\beta)]$. The analysis above then gives, for sufficiently large n:

$$\mathbb{P}(\Pi_{\beta,\epsilon'}) > 1 - \epsilon'.$$

Consider the recurrence relation:

$$E_0 = N_0/(1-\beta) \quad E_{i+1} = (1-\beta)E_i + \sum_{j \geq 1} E_{i-j} \cdot 4j(1-\beta).$$

So long as $\Pi_{\beta,\epsilon'}$ holds, we have:

$$\forall i \in [i_0, \ell^\dagger], \quad D_i \leq E_{i-i_0+1}.$$

Since there exists $\rho > 0$ for which

$$\lim_{i\to\infty} \frac{E_{i-i_0+1}}{\rho \alpha_2^i} = 1,$$

and since $\alpha_2 < \alpha_1$, we can choose i_1 such that:

$$\forall i \in [i_1, \ell^\dagger], \quad \left|\cup_{j \leq i} D_j^*\right| < \alpha_1^i. \tag{28}$$

We also have that $E_{\ell^*} < n^{2/3}$ for all sufficiently large n, which means that so long as $\Pi_{\beta.\epsilon'}$ holds and n is sufficiently large, $\ell^\dagger = \ell^*$. Thus (28) gives (27), as required.

With (26) established, we can then prove (25) in almost exactly the same way that we proved (14). Suppose given $\alpha_0 < \alpha$ and ϵ' with $0 < \epsilon' < 0.5$. To prove (25) it suffices to show, for all sufficiently large n, that $\mathtt{nd}(D_{\ell^*}) > \alpha_0^{\ell^*}$ with probability $> 1 - \epsilon'$. Choose α_1 and β, with $\alpha_0 < \alpha_1 < \alpha$, $0 < \beta < 0.5$, and such that α_1 is the largest root of $f(x,\beta)$, where f is as in (16). Let $\ell^\dagger$ be the minimum of ℓ^* and the first i such that $|\cup_{j \leq i+1} D_j^*| \geq n^{2/3}$. By (27) it holds with high probability that $\ell^\dagger = \ell^*$. Suppose we are given $\mathtt{nd}(D_i^*)$ for some $i < \ell^\dagger$. Then, according to (13), for sufficiently large n the number of inbound long-range contacts of elements of $\mathtt{nd}(D_i^*)$ which are not collision causing, stochastically dominates X which is the sum of $\mathtt{nd}(D_i)$ many i.i.d random variables, each with mean $1 - \beta/2$. Applying the Chernoff bound to X, we conclude that for some $I_\beta > 0$, and for all sufficiently large n, the probability that we fail to have at least $(1-\beta)\mathtt{nd}(D_i)$ many long-range inbound contacts of elements of $\mathtt{nd}(D_i^*)$ which are not collision causing is bounded above by:

$$e^{-I_\beta \mathtt{nd}(D_i)}.$$

[10]Here I_β has a different but similar definition to its previous use in the proof of (14). Similarly $\Pi_{\beta,\epsilon}$ has a similar but different definition here in the proof of (27).

Now $\mathtt{nd}(D_i)$ is guaranteed to grow at a certain rate, since $\mathtt{nd}(D_i) \geq 4(i-1)$ for $i \geq 2$. We can therefore choose i_0 such that $e^{-I_\beta 4(i_0-1)} < \epsilon'/2$, and this means that with probability 1,

$$e^{-I_\beta \mathtt{nd}(D_{i_0})} < \epsilon'/2.$$

We chose $\beta < 0.5$. So if at most a β proportion of the long-range contacts of elements of $\mathtt{nd}(D^*_{i_0})$ are collision causing, we have $\mathtt{nd}(D_{i_0+1}) \geq 1.5\mathtt{nd}(D_{i_0})$. Since $\epsilon' < 0.5$ this gives:

$$e^{-I_\beta \mathtt{nd}(D_{i_0+1})} < (\epsilon'/2)^{1.5} = \epsilon' \sqrt{2\epsilon'}/4 < \epsilon'/4.$$

Then so long as at least $(1-\beta)\mathtt{nd}(D_{i_0+1})$ many inbound long-range contacts of elements of $\mathtt{nd}(D^*_{i_0+1})$ are not collision causing, we have $e^{-I_\beta \mathtt{nd}(D_{i_0+2})} < \epsilon'/8$, and so on. Define $\Pi_{\beta,\epsilon'}$ to be the following event: for all i with $i_0 \leq i < \ell^\dagger$, at least $(1-\beta)\mathtt{nd}(D_i)$ many inbound long-range contacts of elements of $\mathtt{nd}(D^*_i)$ are not collision causing. The analysis above then gives, for sufficiently large n:

$$\mathbb{P}(\Pi_{\beta,\epsilon'}) > 1 - \epsilon'.$$

To complete the argument, consider all those 3-nodes at a lattice distance of $i_0 - 1$ from b. These 3-nodes are guaranteed to belong to $\mathtt{nd}(D^*_{i_0})$. So long as $\Pi_{\beta,\epsilon'}$ holds, we can give a lower bound for the number of descendants of these nodes which belong to $\mathtt{nd}(D^*_{i_0+i-1})$ for each $i \leq \ell^\dagger - i_0 + 1$ through the series:

$$E_0 = 4(i_0-1)/(1-\beta) \quad E_{i+1} = (1-\beta)E_i + \sum_{j\geq 1} E_{i-j} \cdot 4j(1-\beta). \tag{29}$$

There exists a constant $\rho > 0$ such that $\lim_{i\to\infty} E_i/(\rho\alpha_1^i) = 1$. Since $\alpha_0 < \alpha_1$, we can choose $i_1 > i_0$ such that $\mathtt{nd}(D^*_{\ell^*}) > \alpha_0^{\ell^*}$ whenever $\Pi_{\beta,\epsilon'}$ holds, $\ell^\dagger = \ell^*$ and $\ell^* > i_1$. This gives (25) as required.

5.6 Modifications required to change the proof of [GG01] into a proof of Theorem 3.4

We assume full knowledge of the proof of Gavoille and Gengler in [GG01]. In that paper, Lemma 1 gives an inequality concerning the ratio between binomial coefficients, which we now replace with the following, which can also be proved using Stirling's approximation.

Lemma 5.2. *Suppose $\pi(q) : \mathbb{N} \to \mathbb{N}$ satisfies the condition that $\pi(q)/q \to 0$ as $q \to \infty$. Then there exists a real $\gamma > 1$ and a constant $d > 0$ such that for all q which are multiples of 5:*

$$\frac{\binom{q}{4q/5}}{\binom{4q/5}{\pi(q)}\binom{q}{\pi(q)}\binom{2q/5}{q/5}} > d \cdot \gamma^q$$

Gavoille and Gengler specify for each n a set of graphs with n nodes, where the nodes are partitioned into three sets $V = W \cup A \cup B$, where $|W| = q$ and $|A| = |B| = p$. Each possible graph is identified by a matrix M which specifies which nodes in A and W are neighbours (by construction this suffices to give the same information for B and W). The basic form of the argument involves showing that, from a routing function with stretch < 3, M can be recovered using only a small amount of extra information. The construction of M from the routing function consists of three steps. In the first step a matrix M' is produced, in which entries are port numbers given by the routing function as responses to certain appropriately defined queries.

In the next step a matrix, let us call it M_0 (although it is not named in the presentation given in [GG01]), is produced from M' by setting each entry to 1 if its numeral value is unique amongst the values in that row, and setting it to 0 otherwise. A crucial feature of M_0 is that any entries which are set to 1 are 'correct', in the sense that the corresponding entry is 1 in M.[11] Now that we work with a routing function which only gives stretch < 3 for *almost all* pairs, however, the difficulty is that a matrix M_1 will be produced in place of M_0, which may not satisfy this correctness condition for 1 entries. It suffices to show that only a small amount of information is required in order to convert M_1 into M_0. We can carry out this conversion in two stages.

Stage 1. In the first stage, we take each row of M_1 in turn, and change any 1s which are 0s in M_0 into 0s. The number of port labels for nodes in A, means that any row of M_1 has less than $c = 4q/5$ many 1s. Note also that, as observed in [GG01], any row of M_0 has at most $2q/5$ many 0s. In all but $o(p)$ of the rows, specifying which 1s should be flipped requires specifying at most $o(q)$ many of the (less than $4q/5$ many) 1s in that row of M_1 (since stretch < 3 holds for almost almost all pairs). For $o(p)$ of the rows, however, we may have to specify up to $2q/5$ many 1s to change to 0. Overall, for some function $\pi : \mathbb{N} \to \mathbb{N}$ such that $\pi(q)/q \to 0$ as $q \to \infty$, this gives the following upper bound for the information required to carry out the first stage:

$$p\log\binom{4q/5}{\pi(q)} + o(p)\log\binom{q}{2q/5}.$$

Stage 2. In the second stage we take the modified form of M_1 resulting from Stage 1 (in an abuse of notation, we shall still refer to it as M_1), and we change any 0s which are 1s in M_0 into 1s. Arguing similarly to as in Stage 1, we get a (generous) upper bound on the amount of information required in order to specify the appropriate bits to be flipped:

$$p\log\binom{q}{\pi(q)} + o(p)\log\binom{q}{4q/5}.$$

With these modifications in place, we need to consider how they affect the bound on K, which is the number of bits required to store the routing function. In [GG01] the relevant bound is:

$$K \geq p\log\left(\binom{q}{4q/5} \Big/ \binom{2q/5}{q/5}\right) - O(n\log(n)). \tag{30}$$

In place of (30) we now have:

$$K \geq p\log\left(\binom{q}{4q/5} \Big/ \binom{4q/5}{\pi(q)}\binom{q}{\pi(q)}\binom{2q/5}{q/5}\right) - o(p)\log\binom{q}{2q/5} - o(p)\log\binom{q}{4q/5} - O(n\log(n)). \tag{31}$$

By Lemma 5.2, and since $o(p)\log\binom{q}{2q/5}$ and $o(p)\log\binom{q}{4q/5}$ are both $o(n^2)$, it still holds that $K = \Omega(n^2)$, as required.

[11] There is a small oversight in [GG01], that when the routing function for a_i is being queried, actually the port number for b_i could appear as a unique entry in the corresponding row, meaning that a non-neighbour of a_i in W is incorrectly marked as a neighbour. This is easily corrected with information specifying the port label for a_i corresponding to the edge to b_i.