GEODESICS AND ALMOST GEODESIC CYCLES IN RANDOM REGULAR GRAPHS

ITAI BENJAMINI, CARLOS HOPPEN, ERAN OFEK, PAWEŁ PRAŁAT, AND NICK WORMALD

ABSTRACT. A geodesic in a graph G is a shortest path between two vertices of G. For a specific function $e(n)$ of n, we define an almost geodesic cycle C in G to be a cycle in which for every two vertices u and v in C, the distance $d_G(u,v)$ is at least $d_C(u,v) - e(n)$. Let $\omega(n)$ be any function tending to infinity with n. We consider a random d-regular graph on n vertices. We show that almost all pairs of vertices belong to an almost geodesic cycle C with $e(n) = \log_{d-1}\log_{d-1} n + \omega(n)$ and $|C| = 2\log_{d-1} n + O(\omega(n))$. Along the way, we obtain results on near-geodesic paths. We also give the limiting distribution of the number of geodesics between two random vertices in this random graph.

1. Introduction

A *geodesic* in a graph G is a shortest path between two vertices of G, and a *geodesic cycle* C in G is a cycle in which, for every two vertices u and v in C, there is a geodesic between u and v contained in C. The term "geodesic" comes from the discrete metric space naturally associated with a graph, whose elements are the vertices of the graph, and where the distance between two vertices is the length of a shortest path joining them in the graph. The study of geodesic cycles in a graph provides information about the "shape" of the graph. For instance, it is easy to see that if a cycle of length t is embedded into a tree, then some vertex of the tree receives two vertices whose distance around the cycle is 'large' (it is not hard to show a lower bound of $t/3$). In this sense, a graph with a long geodesic cycle has a very different shape from a tree. We do not define "shape", but one might take the minimum metric distortion of an embedding of one graph in another to be a measure of the difference of their shapes. In connection with the influence of geodesics upon the shape of a graph, Angel, Holroyd, Romik and Virág [2] study random geodesics in a graph defined from permutations, and conjecture that the geodesics lie close to great circles in a particular Euclidean embedding of the graph.

Many properties of random graphs, and in particular random regular graphs, have been studied in the past. Imposing the regularity constraint has a strong influences on diameter and connectivity, and we were motivated by the considerations above to ask whether these graphs (for degree at least 3) have geodesic paths and cycles between two randomly chosen vertices. Unfortunately, results on geodesic cycles have proved hard to obtain. However, a similar feeling of the "shape" of a graph can be gained by considering the following relaxation of the definition of geodesic cycles. In the terminology of Bonk

The second author acnowledges the support of FAPESP (2007/56496-3).
The fourth author is supported by MITACS and NSERC.
The fifth author acknowledges the support of the Canadian Research Chairs Program and NSERC.

and Schramm [6], a *k-roughly geodesic cycle* is one in which, for every two vertices u and v, the distance $d_G(u,v)$ is at least $d_C(u,v)-k$. Let $\omega(n)$ be any function tending to infinity with n, and consider the error function $e(n)=\log_{d-1}\log_{d-1}n+\omega(n)$. We define an *almost geodesic cycle* C in G to be one that is $e(n)$-roughly geodesic. Almost geodesic cycles are in some sense as difficult as geodesic cycles to embed into trees with low metric distortion. We investigate the existence of almost geodesic cycles through random pairs of vertices in a random regular graph, and related questions on geodesics and paths that are nearly geodesic, in a sense to be made precise.

Our results on almost geodesic cycles cannot be directly translated to results about geodesic cycles. For instance, we do not know whether almost all pairs of vertices lie in a geodesic cycle. However, the proof of our main theorem allows us to draw conclusions about another measure of shape, namely hyperbolicity. Following Gromov [8], a graph G is *δ-hyperbolic* if, for every four vertices u,v,w,z in G, the two largest values in the set $\{d(u,v)+d(w,z), d(u,w)+d(v,z), d(u,z)+d(v,w)\}$ differ by at most 2δ. (So, for instance, a tree is 0-hyperbolic.) This graph invariant has far-reaching consequences in the design of algorithms. As an easy consequence of our work, we establish that a random d-regular graph, for $d\geq 3$, is a.a.s. not δ-hyperbolic for $\delta=(\log_{d-1}n)/2-\omega(n)$. On the other hand, it is easy to see that this graph must be δ-hyperbolic for δ equal to half of its diameter, which is a.a.s. $(\log_{d-1}n)/2+O(\log\log n)$ by the main result of [5]. See Section 4 for more details.

Our results refer to the probability space of random d-regular graphs with uniform probability distribution. This space is denoted $\mathcal{G}_{n,d}$, and asymptotics (such as "asymptotically almost surely", which we abbreviate to a.a.s.) are for $n\to\infty$ with $d\geq 3$ fixed, and n even if d is odd.

Some related previous research focused on finding (edge/internally)-disjoint paths with many sources and targets. Frieze and Zhao [7] showed that for sufficiently large d there exist fixed positive constants α and β such that a graph G taken from $\mathcal{G}_{n,d}$ a.a.s. has the following property: for any choice of k pairs $\{(a_i,b_i)\mid i=1,\dots,k\}$, satisfying

(i) $k\leq\lceil\alpha dn/\log_d n\rceil$, and
(ii) for each vertex v: $|i : a_i=v|+|i : b_i=v|\leq\beta d$,

there exist edge-disjoint paths in G connecting a_i to b_i for all $i=1,2,\dots,k$. This result is optimal up to constant factors. The paths returned by their algorithm are of length of at least $10\log_d n$.

Our focus is different as it comes from different motivation: showing the existence of an abundance of almost geodesic cycles in $\mathcal{G}_{n,d}$. We obtain a result on internally disjoint paths referring to one pair of vertices fixed before the graph is chosen. This is a much weaker model than the model of [7], that dealt with $\Theta(n/\log n)$ pairs given by an adversary after the graph is chosen. However, we show the existence of disjoint paths that approximate the optimal path (whose length is a.a.s. in $[\log_{d-1}n-\omega(n),\log_{d-1}n+\omega(n)]$) by an additive factor of $O(\omega(n))$, for any function $\omega(n)$ such that $\lim_{n\to\infty}\omega(n)=\infty$, whereas the result of [7] gives at best a constant multiplicative factor. Additionally, our result holds for all $d\geq 3$. We show that it is not only true that randomly chosen vertices u and v are a.a.s. connected by d internally disjoint paths, but these paths may all be chosen with length within a small error $\omega(n)$ of their distance $d(u,v)$. Hence, we prove that their distance in the graph can only be increased by $\omega(n)$ if at least d vertices are deleted. As u and v have degree d, the deletion of this many vertices is sufficient to disconnect them from the graph. Note that the existence of the maximum possible number of internally disjoint paths, d, that there can possibly be between any two vertices u and v in a random regular graph G is an immediate consequence of G being a.a.s. d-connected, as was shown independently by Bollobás [4] and Wormald [9].

More precisely, the first main result in this paper shows the existence of paths of the above form with the additional property that every pair of them creates an almost geodesic cycle.

Theorem 1.1. *Take any integer $d \geq 3$ and any function $\omega(n)$ with $\omega(n) \to \infty$. Let $G \in \mathcal{G}_{n,d}$ and choose vertices u and v in $V(G)$ independently with uniform probability. Then a.a.s. the following hold:*

(i) $|d(u,v) - \log_{d-1} n| < \omega(n)$,

(ii) *there are d paths connecting u and v, all of length at most $\log_{d-1} n + \omega(n)$, such that the subgraph induced by each pair of these paths is an almost geodesic cycle.*

Note that the d paths in (ii) are pairwise internally disjoint because each pair of them induces a cycle.

We may obtain the lower bound in part (i) of the theorem from an elementary observation. Note that, given $G \in \mathcal{G}_{n,d}$, the number of vertices at distance at most i from a vertex u is bounded above by

$$1 + d + d(d-1) + \ldots + d(d-1)^{i-1} = O\left((d-1)^i\right).$$

So, there are $O\left(n(d-1)^{-\omega(n)}\right)$ vertices at distance $i = \log_{d-1} n - \omega(n)$ from any given vertex of G, where $\omega(n) \to \infty$. As a consequence, if two vertices of G are chosen independently with uniform probability, then the probability that the second vertex is at distance at most $i = \log_{d-1} n - \omega(n)$ from the first is at most

$$\frac{1}{n} O\left(n(d-1)^{-\omega(n)}\right) = O\left((d-1)^{-\omega(n)}\right),$$

and therefore, a.a.s.

$$d(u,v) \geq \log_{d-1} n - \omega(n) \tag{1}$$

if u, v are vertices chosen independently with uniform probability in $G \in \mathcal{G}_{n,d}$ and $\omega(n)$ is a function satisfying $\omega(n) \to \infty$. The fact that a.a.s. $d(u,v) \leq \log_{d-1} n + \omega(n)$ will follow from our study of the distribution of the number of geodesics in $G \in \mathcal{G}_{n,d}$.

The rest of the proof requires more sophisticated arguments. Instead of working directly in the uniform probability space of random regular graphs on n vertices $\mathcal{G}_{n,d}$, we use the *pairing model* of random regular graphs, first introduced by Bollobás [3], which is described next. Suppose that dn is even, as in the case of random regular graphs, and consider dn points partitioned into n labelled cells $v_1, \ldots, v_n$ of d points each. A *pairing* of these points is a perfect matching of them into $dn/2$ pairs. Given a pairing P, we may construct a multigraph $G(P)$, with loops allowed, as follows: the vertices are the cells $v_1, \ldots, v_n$, and a pair $\{x, y\}$ in P corresponds to an edge $v_i v_j$ in $G(P)$ if x and y are contained in the cells v_i and v_j, respectively. It is an easy fact that the probability of a random pairing corresponding to a given simple graph G is independent of the graph, hence the restriction of the probability space of random pairings to simple graphs is precisely $\mathcal{G}_{n,d}$. Moreover, it is well known that a random pairing generates a simple graph with probability asymptotic to a constant depending on d, so that any event holding a.a.s. over a probability space of random pairings also holds a.a.s. over the corresponding space $\mathcal{G}_{n,d}$. For this reason, asymptotic results over random pairings suffice for our purposes. The advantage of using this model is that the pairs may be chosen sequentially so that the next pair is chosen uniformly at random over the remaining (unchosen) points. For more information on this model, see [10].

In a slightly different direction, we also investigate the limiting distribution of the number of geodesics joining two randomly chosen vertices in a random regular graph. Indeed, we obtain the following precise result.

Theorem 1.2. *Fix an integer $l \geq 1$. The probability that two vertices u, v chosen independently with uniform probability in $G \in \mathcal{G}_{n,d}$ are joined by exactly l distinct geodesics is asymptotic to*

$$\sum_{k=-\infty}^{\infty} \frac{\left(d(d-1)^{2k-2\gamma(n,d)-2}\right)^l}{l!} \exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)-1}}{d-2}\right) \times \\ \times \left(1+(d-1)^l \exp(-d(d-1)^{2k-2\gamma(n,d)-1})\right).$$

Despite its daunting appearance, this formula answers a very natural question, namely the probability, as n tends to infinity, that two typical vertices in a random regular graph on n vertices are joined by exactly l shortest paths.

In Section 2 of this paper, the number of geodesics is investigated and Theorem 1.2 is proved. The study of almost geodesic cycles, including the proof of Theorem 1.1, is in Section 3. Some consequences of our work and other final remarks are in Section 4.

2. Distribution of the number of geodesics

The first portion of our argument is a simplified version of part of the argument of Bollobás and Fernandez de la Vega [5]. We consider the process in which the neighbourhoods of u and v are exposed step by step. First, the neighbours of u and v are revealed, then the vertices at distance two, and so on. This sequential exposure of the random regular graph is analysed using the random pairing model mentioned in the Introduction.

Let $N_i(u)$ denote the set of vertices at distance at most i from u. Note that, in the early stages of this process, the graphs grown from u and v tend to be trees, hence the number n_i of elements in $N_i(u)$ is approximately

$$n_{i-1} + (d-1)(n_{i-1} - n_{i-2}).$$

Let f_i denote the number of vertices in a balanced d-regular tree, that is,

$$f_i = 1 + d\sum_{j=0}^{i-1}(d-1)^j = 1 + \frac{d((d-1)^i - 1)}{d-2},$$

and let

$$i_0 = \left\lfloor \frac{1}{2}\log_{d-1} n \right\rfloor. \tag{2}$$

Lemma 2.1. *Let $\omega(n)$ be any function of n such that $\omega(n) \to \infty$. For $i \leq i_0 - \omega(n)$ a.a.s. the cardinality n_i of $N_i(u)$ equals f_i. Moreover, for $i \leq i_0 + \omega(n)$ a.a.s.*

$$n_i = f_i - O\Big(\omega(n)(d-1)^{3(i-i_0)+\omega(n)}\Big).$$

Proof. First note that it is sufficient to consider the case when $\omega(n) = o(\log n)$.

Since f_i denotes the number of vertices in a balanced tree where every non-leaf vertex has degree d, the first assertion follows if we show that a.a.s. the set of vertices at distance at most $i \le i_0 - \omega(n)$ of a vertex u induces a tree. In other words, if we expose, step by step, the vertices at distance $1, 2, \ldots, i$ from u, we have to avoid, at step j, edges that induce cycles. So, we wish not to find edges between any two vertices at distance j from u or edges that join any two vertices at distance j to a same vertex at distance $j+1$ from u. We shall refer to edges of this form as 'bad edges'. Note that the expected number of 'bad edges' at step $i+1$ is equal to $O(n_i^2/n) = O(f_i^2/n) = O((d-1)^{2i}/n)$.

Consider $i_1 = \lfloor \frac{1}{2}\log_{d-1} n - \omega(n)\rfloor$. The expected number of 'bad edges' found up to step i_1 is equal to

$$\sum_{j=0}^{i_1-1} O\big((d-1)^{2j}/n\big) = O\big((d-1)^{2i_1}/n\big) = O\big((d-1)^{-2\omega(n)}\big) = o(1)\,.$$

Thus, by Markov's inequality, a.a.s. there are no 'bad edges' until step i_1, hence a.a.s. $N_{i_1}(u)$ is a tree and $n_i = f_i$ for $i \le i_1$.

In order to prove the second assertion, notice that the expected number of 'bad edges' added between step i_1+1 and step i, $i \le \lfloor i_0 + \omega(n)\rfloor \le \lfloor \frac{1}{2}\log_{d-1} n + \omega(n)\rfloor$ is equal to

$$\sum_{j=i_1}^{i-1} O\big((d-1)^{2j}/n\big) = O\big((d-1)^{2i}/n\big) = O\big((d-1)^{2(i-i_0)}\big)\,.$$

Thus, again by Markov's inequality, a.a.s. the total number of 'bad edges' at time i is at most $O\big(\omega(n)(d-1)^{2(i-i_0)}\big)$. Notice that one 'bad edge' added in this time interval can destroy at most two tree branches of size $O\big((d-1)^{i-i_0+\omega(n)}\big)$. This occurs because the 'bad edge' creates a cycle instead of exposing a new vertex v. The branch of descendants of v, which would appear had v been exposed and had the process continued as a balanced d-regular tree, is therefore destroyed and has size at most $1 + (d-1) + \cdots + (d-1)^{i-i_0+\omega(n)} = O\big((d-1)^{i-i_0+\omega(n)}\big)$.

Thus, we have a.a.s.

$$n_i = f_i - O\big(\omega(n)(d-1)^{2(i-i_0)}\big)\cdot O\big((d-1)^{i-i_0+\omega(n)}\big).$$

This completes the proof of the lemma. □

Immediately from this lemma, we have

Corollary 2.2. *For $i = i_0 + o(\log n)$ a.a.s.*

$$n_i = f_i - n^{o(1)} = n^{1/2+o(1)}\,.$$

In the remainder of this notes, let u, v be vertices chosen independently with uniform probability in a graph $G \in \mathcal{G}_{n,d}$ and consider the process of exposing the neighbourhoods of u and v introduced in Lemma 2.1. We say that the neighbourhoods of u and v *join at time* i if $N_{i-1}(u) \cap N_{i-1}(v) = \emptyset$ and $N_i(u) \cap N_i(v) \neq \emptyset$. Moreover, given functions $g = g(n)$ and $h = h(n)$, we say that g is *asymptotic* to h, denoted by $f \sim g$ if

$$\lim_{n\to\infty} \frac{g(n)}{h(n)} = 1.$$

Also, whenever a result that holds for any $\omega(n)$ satisfying $\lim_{n\to\infty}\omega(n) = \infty$ is proven, we shall assume without loss of generality that $\omega(n) = o(\log n)$.

Lemma 2.3. *Let k be a fixed integer and define $\gamma(n,d) = \frac{1}{2}\log_{d-1} n - i_0$, the fractional part of $\frac{1}{2}\log_{d-1} n$. Then*

$$\mathbb{P}(N_{i_0+k}(u) \cap N_{i_0+k}(v) = \emptyset) \sim \exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)}}{d-2}\right).$$

Proof. Denote S_i the event that the neighbourhoods of u and v are separate at time i, that is, $N_j(u)$ and $N_j(v)$ did not join up to time i. We claim that

$$\mathbb{P}(S_{i_0+k} \mid S_{i_0+k-1}) \sim \exp\left(-d^2(d-1)^{2k-2\gamma(n,d)-2}\right).$$

This implies the result for the following reasons. If M is a positive integer, $-M < k$,

$$\begin{aligned}\mathbb{P}(S_{i_0+k}) &= \mathbb{P}(S_{i_0-M}) \\ &\quad \times \prod_{l=-M+1}^{k} \mathbb{P}(S_{i_0+l} \mid S_{i_0+l-1}).\end{aligned}$$

Furthermore, equation (1) establishes that a.a.s. $d(u,v) > 2i_0 - \omega(n)$ for any function $\omega(n)$ with $\lim_{n\to\infty}\omega(n) = \infty$. In particular, given $\epsilon > 0$, we can choose $M = M_\epsilon > 0$ sufficiently large so that $\mathbb{P}(S_{i_0-M}) > 1-\epsilon$. Given such an M, we use the previous equation to derive

$$\begin{aligned}\mathbb{P}(S_{i_0+k}) &> (1-\epsilon)\prod_{l=-M+1}^{k} \exp\left(-d^2(d-1)^{2l-2\gamma(n,d)-2}\right)(1-o(1)) \\ &\sim (1-\epsilon)\exp\left(\sum_{l=-M+1}^{k} -\frac{d^2(d-1)^{2l}}{(d-1)^{2+2\gamma(n,d)}}\right) \\ &= (1-\epsilon)\exp\left(-\frac{d\big((d-1)^{2k+2M}-1\big)}{(d-1)^{2M+2\gamma(n,d)}(d-2)}\right).\end{aligned}$$

The same calculations also lead us to

$$\begin{aligned}\mathbb{P}(S_{i_0+k}) &< \prod_{l=-M+1}^{k} \exp\left(-\frac{d^2(d-1)^{2l}}{(d-1)^{2+2\gamma(n,d)}}\right)(1-o(1)) \\ &\sim \exp\left(-\frac{d\big((d-1)^{2k+2M}-1\big)}{(d-1)^{2M+2\gamma(n,d)}(d-2)}\right).\end{aligned}$$

Putting the last two equations together and letting $\epsilon \to 0$, during which we may assume $M_\epsilon \to \infty$, we have

$$\mathbb{P}(S_{i_0+k}) \sim \exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)}}{d-2}\right),$$

as required.

We now focus on proving the claim. First we would like to find the expected number of joins at time $i = i_0+k$ given that $N_{i-1}(u) \cap N_{i-1}(v) = \emptyset$. Let $U_{i-1} = N_{i-1}(u) \setminus N_{i-2}(u)$ and $V_{i-1} = N_{i-1}(v) \setminus N_{i-2}(v)$. These are the sets of vertices at distance $i-1$ from u

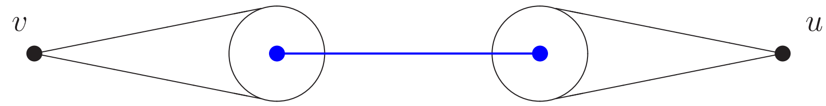

Figure 1. First case – odd length.

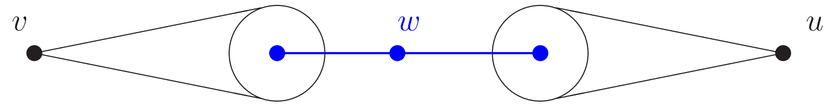

Figure 2. Second case – even length.

and v, respectively. Observe that, by Lemma 2.1

$$\begin{aligned}|U_{i-1}| &= |N_{i-1}(u)| - |N_{i-2}(u)| = f_{i-1} - f_{i-2} + O\left(\omega(d-1)^{3k+\omega(n)}\right) \\ &= d(d-1)^{i-2} + O\left(\omega(d-1)^{3k+\omega(n)}\right) = d(d-1)^{k-\gamma(n,d)-2}\sqrt{n} + O\left(\omega(d-1)^{3k+\omega(n)}\right) \\ &\sim d(d-1)^{k-\gamma(n,d)-2}\sqrt{n}.\end{aligned} \tag{3}$$

Clearly, the same holds for $|V_{i-1}|$.

We have to consider two types of join at time i. The first type (see Figure 1) consists of edges that join one vertex in U_{i-1} to a vertex in V_{i-1} and therefore create uv-paths of odd length. The random variable corresponding to the number of such joins created at time i is denoted by O_i. The second type (see Figure 2) contains joins such that a vertex in $V(G) \setminus (N_{i-1}(u) \cup N_{i-1}(v))$ has neighbours in each of U_{i-1} and V_{i-1}. This generates a path of even length between u and v, and the random variable for the number of these joins is E_i.

Consider joins of the first type. Recall that we are considering the process of exposing the neighbourhoods of u and v step by step. After the first $i-1$ steps, we have exposed the sets $N_{i-1}(u)$ and $N_{i-1}(v)$, which are assumed to be disjoint. Recall that, according to the pairing model (see Introduction), any vertex in U_{i-1} and V_{i-1} can be regarded as a cell of distinct points, where the number of points corresponds to the number of unexposed neighbours of this vertex. The probability that one given point joins another is then asymptotic to $1/(dn)$, since any pair of unmatched points is equally likely to be paired and the whole process has, by Corollary 2.2, matched at most $n^{1/2+o(1)} = o(n)$ pairs of points to this moment. Asymptotically, there are $|U_{i-1}||V_{i-1}|(d-1)^2$ pairs of points such that one is associated with a vertex in U_{i-1} and the other with a vertex in V_{i-1}. This is because the hypothesis $i = i_0 + k$ implies, by Lemma 2.1, that the number of vertices in U_{i-1} or V_{i-1} incident with 'bad edges' created at step $i-2$ is a.a.s. at most $O(\omega(n))$ for any $\omega(n) \to \infty$, and it is clear that each vertex in U_{i-1} or V_{i-1} with degree larger than 1 in $G[N_{i-1}(u)] \cup G[N_{i-1}(v)]$ is incident with a 'bad edge'.

Thus, the expected number O_{i-1} of edges of the first type joining the neighbourhoods of u and v at time $i-1$ satisfies

$$\mathbb{E}\left(O_{i-1} \mid S_{i-1}\right) \sim \frac{(d-1)^2}{dn} |U_{i-1}||V_{i-1}|. \tag{4}$$

A similar argument shows that

$$\mathbb{E}\left(E_{i-1} \mid S_{i-1}\right) \sim \frac{d(d-1)(d-1)^2}{d^2n^2} n|U_{i-1}||V_{i-1}| = \frac{(d-1)^3}{dn}|U_{i-1}||V_{i-1}|\,. \qquad (5)$$

As a consequence,

$$\begin{aligned}\mathbb{E}(\text{number of joins at time } i = i_0+k \mid S_{i-1}) &= \mathbb{E}\left(O_{i-1} \mid S_{i-1}\right) + \mathbb{E}\left(E_{i-1} \mid S_{i-1}\right)\\ = \frac{(d-1)^2}{n}|U_{i-1}||V_{i-1}| &\sim \frac{d^2(d-1)^{2i-2}}{n} = \frac{d^2(d-1)^{2k}}{(d-1)^{2+2\gamma(n,d)}}.\end{aligned}$$

In this equation, the cardinalities of U_{i-1} and V_{i-1} have been approximated by equation (3).

We wish to apply the method of moments to establish

$$\mathbb{P}(S_{i_0+k}|\ S_{i_0+k-1}) \sim \exp\left(-d^2(d-1)^{2k-2\gamma(n,d)-2}\right),$$

so we have to verify that the j-th factorial moment of the random variable Z counting the number of joins at time $i = i_0 + k$ satisfies $\mathbb{E}([Z]_j) = \frac{\mathbb{E}(Z)^j}{j!}$, for all $j \geq 2$.

Let $j \geq 2$ and suppose that the subgraphs induced by $N_{i-1}(u)$ and $N_{i-1}(v)$ are disjoint. As before, let $U_{i-1} = N_{i-1}(u) \setminus N_{i-2}(u)$, $V_{i-1} = N_{i-1}(v) \setminus N_{i-2}(v)$, and, given $r \in U_{i-1}$, $s \in V_{i-1}$, $t \in V(G) \setminus (N_{i-1}(v) \cup N_{i-1}(u))$, introduce indicator random variables $X_{(r,s)}$ for the event that rs is an edge in G and $Y_{(r,s,t)}$ for the event that rt and st are both edges in G. So,

$$Z = \sum_{r \in U_{i-1},\ s \in V_{i-1}} X_{(r,s)} + \sum_{\substack{r \in U_{i-1},\ s \in V_{i-1}\\ t \in V(G) \setminus (N_{i-1}(v) \cup N_{i-1}(u))}} Y_{(r,s,t)}$$

is the random variable counting the number of joins that appear between the neighbourhoods of u and v at step i.

The j-th factorial moment of Z is given by

$$\mathbb{E}([Z]_j) = \sum_{l=0}^{j} \sum_{\star} \mathbb{P}((X_{(r_m,s_m)} = 1,\ 1 \leq m \leq l) \wedge (Y_{(r_m,s_m,t_m)} = 1,\ l+1 \leq m \leq j)), \quad (6)$$

where, for any given l, $\sum_\star$ denotes the sum over all distinct ordered pairs (r_m, s_m), $1 \leq m \leq l$, and (r_m, s_m, t_m), $l+1 \leq m \leq j$.

We shall prove later that the relevant terms in this sum are the ones for which all the ordered pairs are disjoint, that is, there is no repetition of vertices among the j events. Assuming this, we obtain

$$\begin{aligned}\mathbb{E}([Z]_j) = \sum_{l=0}^{j} \binom{|U_{i-1}|}{j}\binom{|V_{i-1}|}{j}\binom{n-o(n)}{j-l}\binom{j}{l}^2 l![(j-l)!]^2 \times\\ \times \left(\frac{(d-1)^2}{dn-o(n)}\right)^l \left(\frac{(d-1)^3}{dn^2-o(n^2)}\right)^{j-l}.\end{aligned}$$

This is because there are $\binom{|U_{i-1}|}{j}\binom{|V_{i-1}|}{j}\binom{n-o(n)}{j-l}$ ways of choosing j vertices in each of U_i and V_i, and of choosing $j-l$ vertices in $V(G) \setminus (N_{i-1}(u) \cup N_{i-1}(v))$. Moreover, pairing l of the chosen vertices in U_i with l of the chosen vertices in V_i can be done in $\binom{j}{l}^2 l!$ ways, whereas there are $[(j-l)!]^2$ ways of creating triples on the remaining chosen vertices in U_i, V_i and the vertices chosen in $V(G) \setminus (N_{i-1}(u) \cup N_{i-1}(v))$. Now that we fixed distinct ordered pairs (r_m, s_m), $1 \le m \le l$, and (r_m, s_m, t_m), $l+1 \le m \le j$, the term $\left(\frac{(d-1)^2}{dn-o(n)}\right)^l \left(\frac{(d-1)^3}{dn^2-o(n^2)}\right)^{j-l}$ corresponds to the probability that all the events $X_{(r_m,s_m)} = 1$ and $Y_{(r_m,s_m,t_m)} = 1$ occur simultaneously, since there is only a finite number of them.

The previous sum is asymptotic to

$$\begin{aligned}
&\sum_{l=0}^{j} \frac{|U_{i-1}|^j}{j!} \frac{|V_{i-1}|^j}{j!} \frac{n^{j-l}}{(j-l)!} \frac{j!^2}{l!} \Big(\frac{(d-1)^2}{dn}\Big)^l \Big(\frac{(d-1)^3}{dn^2}\Big)^{j-l} \\
&= \frac{|U_{i-1}|^j |V_{i-1}|^j (d-1)^{2j}}{n^j d^j j!} \sum_{l=0}^{j} \frac{(d-1)^{j-l} j!}{l!(j-l)!} \\
&= \frac{|U_{i-1}|^j |V_{i-1}|^j (d-1)^{2j}}{n^j d^j j!} \sum_{l=0}^{j} \binom{j}{l} (d-1)^{j-l} \\
&= \frac{|U_{i-1}|^j |V_{i-1}|^j (d-1)^{2j}}{n^j j!} = \frac{1}{j!} (\mathbb{E}Z)^j = \frac{1}{j!} \left(\frac{d^2 (d-1)^{2k}}{(d-1)^{2+2\gamma(n,d)}} \right)^j .
\end{aligned}$$

It remains to show that indeed the sum over all disjoint ordered pairs (r_m, s_m), $1 \le m \le l$, and (r_m, s_m, t_m), $l+1 \le m \le j$, is asymptotic to the sum over all distinct ordered pairs. Suppose that there are $j-a$ distinct elements appearing in the first coordinate, $j-b$ in the second and $j-l-c$ in the third, where $a+b+c \ge 1$. The terms of this form in equation (6) are bounded above by

$$\begin{aligned}
\sum_{l=0}^{j} \sum_{**} \binom{j-1}{a} \binom{|U_{i-1}|}{j-a} \binom{j-1}{b} \binom{|V_{i-1}|}{j-b} \binom{j-l-1}{c} \binom{n-o(n)}{j-l-c} \times \\
\times \binom{j}{l}^2 l! [(j-l)!]^2 \Big(\frac{(d-1)^2}{dn-o(n)}\Big)^l \Big(\frac{(d-1)^3}{dn^2-o(n^2)}\Big)^{j-l},
\end{aligned}$$

where $\sum_{**}$ denotes the sum over all triples $(a,b,c) \in \{0,\dots,j-1\}^2 \times \{0,\dots,j-l-1\}$ satisfying $a+b+c \ge 1$. This is because there are $\binom{|U_{i-1}|}{j-a}$ ways of choosing $j-a$ vertices in U_{i-1} and $\binom{j-1}{a}$ ways of building a multi-set of cardinality j with $j-a$ given elements (and using all of them). The same is true for choosing vertices in V_{i-1} and $V(G) \setminus (N_{i-1}(u) \cup N_{i-1}(v))$. Our last expression is smaller or equal to

$$\begin{aligned}
\sum_{l=0}^{j} \sum_{**} \frac{(j-1)^{a+b}(j-l-1)^c}{a!\,b!\,c!} \frac{|U_{i-1}|^{j-a}}{(j-a)!} \frac{|V_{i-1}|^{j-b}}{(j-b)!} \frac{n^{j-l-c}}{(j-l-c)!} \times \\
\times \frac{j!^2}{l!} \Big(\frac{(d-1)^2}{dn-o(n)}\Big)^l \Big(\frac{(d-1)^3}{dn^2-o(n^2)}\Big)^{j-l} .
\end{aligned}$$

If we divide this by $\frac{|U_{i-1}|^j |V_{i-1}|^j (d-1)^{2j}}{n^j j!}$, this is asymptotic (with respect to n) to

$$\sum_{l=0}^{j} \sum_{**} \frac{\mathcal{K}(a,b,c,j,l,d)}{|U_{i-1}|^a |V_{i-1}|^b n^c},$$

where $\mathcal{K}(a,b,c,j,l,d)$ does not depend on n. Since $|U_{i-1}|^a |V_{i-1}|^b n^c \to \infty$ as $n \to \infty$ for every $a+b+c \ge 1$, we conclude that the above sum tends to zero as n tends to infinity and therefore the terms related to non-disjoint tuples in equation (6) can indeed be ignored to compute $\mathbb{E}([Z]_j)$.

Given this, the method of moments implies that

$$\mathbb{P}(\text{no joins at time } i = i_0 + k \mid N_{i-1}(v) \cap N_{i-1}(u) = \emptyset) \sim \exp\left(-d^2(d-1)^{2k-2\gamma(n,d)-2}\right),$$

which completes the proof of the claim and therefore establishes the lemma. □

We are now prepared to prove the result mentioned at the end of the last section.

Corollary 2.4. *Let u, v be vertices chosen independently with uniform probability in $G \in \mathcal{G}_{n,d}$. For any function $\omega(n)$ such that $\omega(n) \to \infty$, the assertion $d(u,v) < \log_{d-1} n + \omega(n)$ holds a.a.s.*

Proof. Let $\epsilon > 0$. Lemma 2.3 implies that the probability of the event D_k that u and v are at distance greater than $2i_0 + 2k$ is asymptotic to

$$\exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)}}{d-2}\right),$$

where k a fixed integer and $\gamma(n,d)$ is the fractional part of $\frac{1}{2}\log_{d-1} n$. So, $\mathbb{P}(D_k) < \epsilon$ for k sufficiently large, and the result follows. □

Lemma 2.5. *Let k be an integer and let i_0 and $\gamma(n,d)$ be the integer and fractional parts of $\frac{1}{2}\log_{d-1} n$, respectively. Define O_i to be the random variable counting the number of uv-paths of odd length, that is, paths of the first case, created at step $i = i_0 + k$. Let E_i be the equivalent random variable for paths of even length. Then*

(i) *With $\mu_k = d(d-1)^{2k-2\gamma(n,d)-2}$,*

$$\mathbb{P}(O_{i_0+k} = j \mid N_{i_0+k-1}(u) \cap N_{i_0+k-1}(v) = \emptyset) \sim \frac{\mu_k^j}{j!}\exp(-\mu_k).$$

(ii) *With $\nu_k = d(d-1)^{2k-2\gamma(n,d)-1}$,*

$$\mathbb{P}(E_{i_0+k} = j \mid N_{i_0+k-1}(u) \cap N_{i_0+k-1}(v) = \emptyset \wedge O_{i_0+k} = 0) \sim \frac{\nu_k^j}{j!}\exp(-\nu_k).$$

Proof. This can be proven by the method of moments using calculations very similar to the ones in the previous lemma, proceeding separately for joins of the first type and joins of the second type. For the former, we condition on the event that no joins occurred in previous steps of the process, and, for the latter, we further assume that no joins of the first type occurred in the current step. The details are omitted. □

We observe that, alternatively, the proofs of the previous lemma and of Lemma 2.3 could be unified by considering joint factorial moments of random variables for joins of the first type and of the second type.

We are now ready to deduce one of the main results, Theorem 1.2, which is now restated.

Theorem 1.2. *Fix an integer $l \geq 1$. The probability that two vertices u, v chosen independently with uniform probability in $G \in \mathcal{G}_{n,d}$ are joined by exactly l distinct* *geodesics is asymptotic to*

$$\sum_{k=-\infty}^{\infty} \frac{\left(d(d-1)^{2k-2\gamma(n,d)-2}\right)^l}{l!} \exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)-1}}{d-2}\right) \times$$
$$\times \left(1+(d-1)^l \exp(-d(d-1)^{2k-2\gamma(n,d)-1})\right).$$

Proof. Let J_i denote the event that the neighbourhoods of u and v join at time i, and let $\hat{Z}_l$ be the event that u and v are connected by exactly l geodesics. Note that given J_i, all geodesics must be present in the neighbourhood exposure process at time i. Observe that, because a geodesic is a shortest path between u and v, the event $\hat{Z}_l$ occurs if, at the time the neighbourhoods join, there are exactly l geodesics created by joins of the first type, or exactly l created by joins of the second type. So, given a positive integer M,

$$\mathbb{P}(\hat{Z}_l) = \mathbb{P}(\hat{Z}_l \wedge \bigcup_{k=1}^{i_0-M} J_k) + \sum_{k=-M+1}^{M-1} \mathbb{P}(\hat{Z}_l \wedge J_{i_0+k}) + \mathbb{P}(\hat{Z}_l \wedge \bigcup_{k\geq i_0+M} J_k).$$

The first and last element in the right-hand side can be made less than ϵ, for any given $\epsilon > 0$, by choosing $M = M_\epsilon$ sufficiently large, as ensured by Corollary 2.4 and by the fact that equation (1) holds a.a.s.

We now treat the terms of the form $\mathbb{P}(\hat{Z}_l \wedge J_{i_0+k})$ for integers $k \in [-M+1, M-1]$. A significant simplification to our calculations comes from the fact that, when the neighbourhoods first join, a single geodesic is produced by each particular paired that joins. To this end, let B_i be the bad event that some vertex in $U_{i-1} = N_{i-1}(u) \setminus N_{i-2}(u)$ with more than one path back to u, or a similar vertex in $V_{i-1} = N_{i-1}(v) \setminus N_{i-2}(v)$ with more than one path back to v, is involved in a join in the step $i = i_0 + k$. We call such vertices u and v bad vertices. Define Z_l to be the event that exactly l pairs of vertices join. The difference between $\mathbb{P}(Z_l \wedge J_i)$ and $\mathbb{P}(\hat{Z}_l \wedge J_i)$ is at most $\mathbb{P}(B_i)$.

We first show $\mathbb{P}(B_i) = o(1)$. Since our subsequent computation of events involves probabilities that are not $o(1)$, we may then compute with Z_l. Let X_i be the random variable for the number of bad vertices in a join at time i, given that the neighbourhoods of u and v are separate at time $i-1$. Clearly, $\mathbb{P}(B_i) = \mathbb{P}(X_i \geq 1)$, which is bounded above by $\mathbb{E}(X_i)$ by Markov's inequality. Now, each bad vertex is a descendant of an edge lying in a "bad edge" in the sense of Lemma 2.1, i.e. an edge that create cycles in the exposition of the neighbourhoods of u and v. So, from the proof of Lemma 2.1, we deduce that the expected number of bad vertices in U_{i-1} is $O((d-1)^{i-i_0+\omega(n)})$, where, as n tends to infinity, the function $\omega(n)$ is allowed to tend to infinity as slow as desired.

Moreover, mimicking the calculations that lead to the expressions for $\mathbb{E}(O_{i-1} \mid S_i)$ and $\mathbb{E}(E_{i-1} \mid S_i)$ in the proof of Lemma 2.3 (equations (4) and 5)), we may easily determine that the probability that a given bad vertex in U_{i-1} is involved in a join at time i is at most

$$\frac{(d-1)^2 + (d-1)^3}{dn} |V_{i-1}| \sim \frac{1}{n}(d-1)^{i+1} \leq \frac{1}{\sqrt{n}}(d-1)^{M+1},$$

where $V_{i-1} = N_{i-1}(v)$. Clearly, the same is true for bad vertices in V_{i-1}, so that

$$\mathbb{P}(B_i) \leq \mathbb{E}(X_i) = O\left((d-1)^{M+\omega(n)} \frac{2}{\sqrt{n}} (d-1)^{M+1}\right) = o(1),$$

as M is a constant and ω may be chosen to tend to infinity sufficiently slowly.

We now evaluate $\sum_{k=-M+1}^{M-1} \mathbb{P}(\hat{Z}_l \wedge J_{i_0+k})$. Observe that each of the terms $\mathbb{P}(Z_l \wedge J_{i_0+k})$, for $-M+1 \le k \le M-1$, is equal to

$$\begin{aligned}\mathbb{P}(N_{i_0+k-1}(u) \cap N_{i_0+k-1}(v) = \emptyset)\big[\mathbb{P}(O_{i_0+k} = l \mid N_{i_0+k-1}(u) \cap N_{i_0+k-1}(v) = \emptyset) + \\ + \ \mathbb{P}(O_{i_0+k} = 0 \mid N_{i_0+k-1}(u) \cap N_{i_0+k-1}(v) = \emptyset) \times \\ \times \mathbb{P}(E_{i_0+k} = l \mid O_{i_0+k} = 0 \wedge N_{i_0+k-1}(u) \cap N_{i_0+k-1}(v) = \emptyset)\big].\end{aligned}$$

By our previous lemmas, this is asymptotic to

$$\begin{aligned}\exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)-2}}{d-2}\right)\left(\frac{\big(d(d-1)^{2k-2\gamma(n,d)-2}\big)^l}{l!}\exp(-d(d-1)^{2k-2\gamma(n,d)-2}) \ + \right.\\ \left. + \ \frac{\big(d(d-1)^{2k-2\gamma(n,d)-1}\big)^l}{l!}\exp(-d(d-1)^{2k-2\gamma(n,d)-2} - d(d-1)^{2k-2\gamma(n,d)-1})\right)\end{aligned}$$

Hence, if we let ϵ tend to zero,

$$\begin{aligned}\mathbb{P}(Z_l) \sim \sum_{k=-\infty}^{\infty} \frac{\big(d(d-1)^{2k-2\gamma(n,d)-2}\big)^l}{l!}\exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)-1}}{d-2}\right) \times \\ \times \big(1 + (d-1)^l \exp(-d(d-1)^{2k-2\gamma(n,d)-1})\big),\end{aligned}$$

as required. □

An interesting special case is when $l = 1$, since this theorem provides the probability of u and v being joined by a unique geodesic. This probability is given by

$$\begin{aligned}\sum_{k=-\infty}^{\infty} d(d-1)^{2k-2\gamma(n,d)-2}\exp\left(-\frac{d(d-1)^{2k-2\gamma(n,d)-1}}{d-2}\right) \times \\ \times \big(1 + (d-1)\exp(-d(d-1)^{2k-2\gamma(n,d)-1})\big).\end{aligned}$$

The probability here is a function of $\gamma(n,d)$ and oscillates as $\gamma(n,d)$ varies from 0 to 1.

We include some numerical results in the table below for some values of d, where *prob* is the probability of a unique geodesic as $\gamma(n,d) = 0$ and *osc* is the maximum variation with respect to $\gamma = 0$ as γ varies from 0 to 1.

d	*prob*	*osc*
3	0.7213	8.6×10^{-6}
4	0.6073	1.4×10^{-3}
5	0.5444	7.9×10^{-3}
10	0.4411	7.6×10^{-2}
100	0.3743	0.3

The magnitude of the oscillations depends on d. We justify why it is small when d is small. Note that the probability of a unique geodesic is equal to

$$\begin{aligned}\frac{d-2}{d-1} S_{(d-1)^2}\left(-\gamma(n,d) + \log_{(d-1)^2}\frac{d}{(d-1)(d-2)}\right) + \\ + \frac{d-2}{d-1} S_{(d-1)^2}\left(-\gamma(n,d) + \log_{(d-1)^2}\frac{d}{d-2}\right),\end{aligned} \tag{7}$$

where $S_c(x) = \sum_{m=-\infty}^{\infty} c^{m+x} \exp(-c^{m+x})$, a function with period 1. The classical Poisson summation formula gives us that

$$S_c(x) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} c^{t+x} \exp(-c^{t+x}) \exp(2\pi imt)\, dt.$$

Setting $z = c^{t+x}$ gives

$$S_c(x) = \frac{1}{\log c} \sum_{m=-\infty}^{\infty} \exp(-2\pi imx) \int_0^{\infty} \exp(-z + 2\pi im \log z/\log c)\, dz, \tag{8}$$

and the integral is just $\Gamma(2\pi im/\log c + 1)$.

Using properties of the gamma function (see for instance [1]), we have

$$|\Gamma(1 + yi)| = |iy\Gamma(yi)| = |y|\sqrt{\frac{\pi}{y \sinh(\pi y)}},$$

so given m in the previous summation,

$$\begin{aligned} |\Gamma(2\pi mi/\log c + 1) \exp(-2\pi imx)| &= |\Gamma(2\pi im/\log c + 1)| \\ &= \left(\frac{2\pi^2|m|}{\log c\ |\sinh(2\pi^2 m/\log c)|}\right)^{1/2}. \end{aligned} \tag{9}$$

The term for $m = 0$ in the sum (8) is independent of x, hence it yields terms independent of n in equation (7). In the special case $d = 3$, equation (9) leads to the following bounds on the other terms of the sum (8). For $|m| = 1$, the bound is approximately 4.32×10^{-3}, for $|m| = 2$, it is approximately 4.94×10^{-6}, and for larger $|m|$ the bounds are even smaller. Similar observations explain the small oscillations when d is small.

3. Almost geodesic cycles

In the proof of Theorem 1.1 we shall use the following auxiliary result.

Lemma 3.1. *Let $G \in \mathcal{G}_{n,d}$ and let u, v be vertices chosen independently at random in G. Consider functions $\alpha(n), \beta(n)$ such that $\alpha(n), \beta(n) \to \infty$, $\alpha(n) = o(\log_{d-1} n)$ and $\beta(n) = o(\alpha(n))$. Then a.a.s. every vertex at distance $\lfloor \alpha(n) \rfloor$ from u or v lies on at most one uv-path with length less than or equal to $\log_{d-1} n + \beta(n)$.*

Proof. We prove this result for vertices at distance $\lfloor \alpha(n) \rfloor$ from u, and by a similar argument the same result holds for vertices at distance $\lfloor \alpha(n) \rfloor$ from v. As in Section 2, we consider the process of exposing the neighbourhoods of u and v based on the pairing model. Here, $T_\alpha(u)$, the extended neighbourhood of u, is exposed for $\lfloor \alpha(n) \rfloor$ steps while $T_\beta(v)$, the extended neighbourhood of v, is exposed for $\lfloor \frac{1}{2} \log_{d-1} n - \beta(n) \rfloor$ steps. By Lemma 2.1, a.a.s. $T_\alpha(u)$ and $T_\beta(v)$ are both trees. Moreover, our choice of α and β imply that, given a positive integer k and n sufficiently large, we have that

$T_\alpha(u) \cap T_\beta(v) \neq \emptyset$ only if $N_{i_0-k}(u) \cap N_{i_0-k}(v) \neq \emptyset$, where, as before, $i_0 = \lfloor \frac{1}{2} \log_{d-1} n \rfloor$. By Lemma 2.3, the probability of the latter is asymptotic to

$$1 - \exp\left(-\frac{d(d-1)^{-2k-2\gamma(n,d)}}{d-2} \right),$$

which approaches zero as $k \to \infty$. Since k is arbitrary, $T_\alpha(u) \cap T_\beta(v) = \emptyset$ holds a.a.s.

Let U_α be the set of vertices at distance $\lfloor \alpha(n) \rfloor$ from u. Given a vertex $w \in U_\alpha$, let Y_w be the indicator random variable for the event that w is connected to $T_\beta(v)$ by at least two distinct paths not through U_α of length less than or equal to $\frac{1}{2} \log_{d-1} n - \alpha(n) + 2\beta(n) + 2$. Define $Y = \sum_{w \in U_\alpha} Y_w$. It is clear that this lemma follows if we prove that a.a.s. $Y = 0$. We shall do this by using

$$\mathbb{P}\big(Y \geq 1\big) \leq \sum_{w \in U_\alpha} \mathbb{P}\big(Y_w = 1\big),$$

and by showing that the right-hand side goes to zero as n tends to infinity.

For a fixed w, define the set T'_w obtained by the exposure of the neighbourhood of w for $\frac{1}{2} \log_{d-1} n - \alpha(n) + 2\beta(n) + 2$ steps. This time, however, the neighbour of w in $T_\alpha(u)$ is not added to T'_w at the first step of the process, that is, only the "new" neighbours of w are exposed. As in Lemma 2.1, we use the term "bad edges" for edges that yield cycles in T'_w. Consider the random variable X_w counting the number of "bad edges" in T'_w. Then, calculations analogous to the ones in Lemma 2.1 establish that

$$\begin{aligned} \mathbb{E}(X_w) &= \sum_{s=0}^{\lfloor \frac{1}{2} \log_{d-1} n - \alpha(n) + 2\beta(n) + 2 \rfloor} O\left(\frac{(d-1)^{2s}}{n} \right) \\ &= O\left(\frac{(d-1)^{\log_{d-1} n - 2\alpha(n) + 4\beta(n)}}{n} \right) = O\left((d-1)^{4\beta(n) - 2\alpha(n)} \right). \end{aligned}$$

Thus Markov's inequality implies

$$\mathbb{P}\big(X_w \geq 1\big) = O\left((d-1)^{4\beta(n) - 2\alpha(n)} \right).$$

Now, note that

$$\mathbb{P}\big(Y_w = 1\big) = \mathbb{P}\big(X_w \geq 1\big)\mathbb{P}\big(Y_w = 1 \mid X_w \geq 1\big) + \mathbb{P}\big(X_w = 0\big)\mathbb{P}\big(Y_w = 1 \mid X_w = 0\big).$$

We have a bound for the first term in this sum. For the second term, we use the definition of conditional probability and observe that the event $(Y_w = 1) \wedge (X_w = 0)$ occurs only if there is a pair of distinct paths joining w to $T_\beta(v)$ with length at most $\frac{1}{2} \log_{d-1} n - \alpha(n) + 2\beta(n) + 2$ and with the property that, after they first split, they do not join again.

So, a bound on $\mathbb{P}\big(Y_w = 1 \wedge X_w = 0\big)$ may be obtained by counting the number of possible pairs of distinct paths P and Q joining u_i to $T_\beta(v)$ with lengths r and s, where $r \leq s \leq \frac{1}{2} \log_{d-1} n - \alpha(n) + 2\beta(n) + 2$, and the first j vertices are shared by both paths, while the remainder of the paths are internally disjoint. So, if $i_0 = \left\lfloor \frac{1}{2} \log_{d-1} n \right\rfloor$,

$$\begin{aligned}
\mathbb{P}\big(Y_w = 1 \wedge X_w = 0\big) = &\sum_{s=1}^{\lfloor i_0-\alpha(n)+2\beta(n)+2\rfloor} \sum_{r=1}^{\lfloor i_0-s\rfloor} \sum_{j=0}^{r-1} \\
&\binom{(d-1)^{\lfloor i_0-\beta(n)\rfloor}}{2}\binom{n-o(n)}{r+s-j-2}\binom{r+s-j-2}{j}\binom{r+s-2}{r-1} \\
&\times j!(r-1)!(s-1)! O\left(\left(\frac{(d-1)}{n-o(n)}\right)^{r+s-j}\right) \\
= &\sum_{s=1}^{\lfloor i_0-\alpha(n)+2\beta(n)+2\rfloor} \sum_{s=1}^{\lfloor i_0-s\rfloor} \sum_{j=0}^{r-1} O\left(\frac{(d-1)^{2i_0-2\beta(n)}(d-1)^{r+s-j}}{n^2}\right) \\
= &\, O\left((d-1)^{2\beta(n)-2\alpha(n)}\right).
\end{aligned}$$

Note that the formula holds because there are at most $\binom{(d-1)^{\lfloor i_0-\beta(n)\rfloor}}{2}$ ways of choosing two vertices in $T_\beta(v)$ and there are $\binom{n-o(n)}{r+s-j-2}$ ways of choosing vertices in the graph to include in the two paths. Moreover, these vertices can be divided into vertices of $P \cap Q$, $P - Q$ and vertices of $Q - P$ in $\binom{r+s-j-2}{j}\binom{r+s-2}{r-1}$ ways and can then be ordered to form the paths in $j!(r-1)!(s-1)!$ ways. Finally, each edge on the path appears with probability at most $\frac{(d-1)}{n-o(n)}$ conditional on the fact that all previous edges on the path have appeared.

We conclude that

$$\begin{aligned}
\mathbb{P}\big(Y_w = 1\big) &= \mathbb{P}\big(X_w \geq 1\big)\mathbb{P}\big(Y_w = 1 \mid X_w \geq 1\big) + \mathbb{P}\big(X_w = 0\big)\mathbb{P}\big(Y_w = 1 \mid X_w = 0\big) \\
&\leq \mathbb{P}\big(X_w \geq 1\big) + \mathbb{P}\big(Y_w = 1 \mid X_w = 0\big) \\
&= O\left((d-1)^{4\beta(n)-2\alpha(n)}\right) + \frac{O\left((d-1)^{2\beta(n)-2\alpha(n)}\right)}{1 - O\left((d-1)^{4\beta(n)-2\alpha(n)}\right)} \\
&= O\left((d-1)^{4\beta(n)-2\alpha(n)}\right).
\end{aligned}$$

Now, because there are $O\left((d-1)^{\alpha(n)}\right)$ vertices at distance $\lfloor \alpha(n) \rfloor$ of u, we have

$$\begin{aligned}
\mathbb{P}\big(Y \geq 1\big) &\leq \sum_{w \in U_\alpha} \mathbb{P}\big(Y_w = 1\big) \\
&= O\left((d-1)^{\alpha(n)}\right) O\left((d-1)^{4\beta(n)-2\alpha(n)}\right) = O\left((d-1)^{4\beta(n)-\alpha(n)}\right).
\end{aligned}$$

Because $\beta(n) = o(\alpha(n))$, this term goes to zero as n tends to infinity and indeed $\mathbb{P}\big(Y \geq 1\big) \to 0$. The lemma follows. □

In preparation for proving the second part of Theorem 1.1, it is convenient to deal with the following simpler goal. To state this we need two definitions. A *k-near-geodesic* is a path that is a geodesic between the two vertices at distance k from its ends. A vertex p on a path P between vertices u and v is said to be a *midpoint* of P if $|d_P(u,p) - d_P(v,p)| \leq 1$, where d_P denotes the distance along path P.

Lemma 3.2. *Let k be a nonnegative integer. Asymptotically almost surely, for every two distinct k-near-geodesics, P and Q, between u and v with midpoints p and q, respectively,*

$$\log_{d-1}(n) - \omega(n) < d(p,q) < \log_{d-1}(n) + \omega(n).$$

Proof. For the upper bound, we know as in part (i) of the theorem that a.a.s. $d(u,v) < \log_{d-1} n + \omega(n)$, hence there is a sufficiently short path connecting p to q through u or v.

We turn to the lower bound. Given a function $\omega(n)$ satisfying $\omega(n) \to \infty$, we know that a.a.s. $d(u,v) \geq \log_{d-1} n - \omega(n)$ (see (1)). Consider distinct k-near-geodesics P and Q.

Claim 1: *P and Q a.a.s. do not have a vertex in common at distance at least $\frac{\omega(n)}{3}$ from their endpoints.*

We prove the claim by contradiction. Suppose without loss of generality that such a vertex is closer to u on P and let w be the vertex on P at distance $\lfloor \frac{\omega(n)}{3} \rfloor$ from u. Note that P and Q a.a.s. differ at some vertex or edge after w, since the set of vertices at distance at most $\frac{\omega(n)}{3}$ from u a.a.s. induces a tree. But then, w lies on at least two distinct u,v paths with length less than or equal to $\log_{d-1} n + \log \omega(n)$, which a.a.s. does not occur by Lemma 3.1 with $\alpha(n) = \frac{\omega(n)}{3}$ and $\beta(n) = \log \omega(n)$. (Note that the lengths of both P and Q are a.a.s. bounded by $\log_{d-1} n + \log \omega(n)$ because any k-near-geodesic between u and v has length at most $d(u,v) + 4k$.) This proves the claim.

Now consider the event that the midpoints p and q of P and Q are at distance at most $\log_{d-1} n - \omega(n)$. One way for this to occur is by the existence of a pq-path R of length at most $\log_{d-1} n - \omega(n)$ using vertices and edges on $P \cup Q$ only. But $d(u,v) \geq \log_{d-1} n - \omega(n)$ implies that R does not contain vertices at distance less than or equal to $\frac{\omega(n)}{3}$ from u or v. Claim 1 shows that no other vertex can be in common. Thus, a.a.s. there is no short path from p to q using edges on P and Q only.

So consider a geodesic A between p and q containing at least one edge outside $P \cup Q$. Using A oriented from p to q as a reference, let v_P denote the last vertex on $A \cap P$ and let v_Q be the first vertex on $A \cap Q$ after v_P. The vertices v_P and v_Q divide the geodesic into three parts, namely from p to v_P, from v_P to v_Q and from v_Q to q. Because P, Q are k-near-geodesics between u and v for a fixed k and A is a geodesic between p and q, we must have $d_A(p, v_P) = d_P(p, v_P)$ and $d_A(v_Q, q) = d_Q(v_Q, q)$. So, for p and q to be at distance at most $\log_{d-1} n - \omega(n)$ for some $\omega(n) \to \infty$, it must be that

$$d_A(v_P, v_Q) < \log_{d-1} n - d_P(p, v_P) - d_Q(v_Q, q) - \omega(n). \tag{10}$$

So, a short path between p and q has to be caused by a short path connecting a vertex in P to a vertex in Q which is internally disjoint from $P \cup Q$. More precisely, there must exist vertices v_P, v_Q on P and Q, respectively, and an $v_P v_Q$-path R satisfying:

$$V(R) \cap V(P \cup Q) = \{v_P, v_Q\}, \tag{11}$$

$$|R| \leq \log_{d-1} n - d(p, v_P) - d(v_Q, q) - \omega(n). \tag{12}$$

Such a configuration is illustrated by Figure 3.

We prove that a.a.s. G does not contain a path R satisfying (11) and (12). We do this by exposing the neighbourhoods of vertices along P and Q conditional on the particular paths P and Q being in the graph. By relaxing the condition that P and Q are k-near-geodesics (but retaining the condition that their length is at most $d(u,v) + 4k + \omega(n)$),

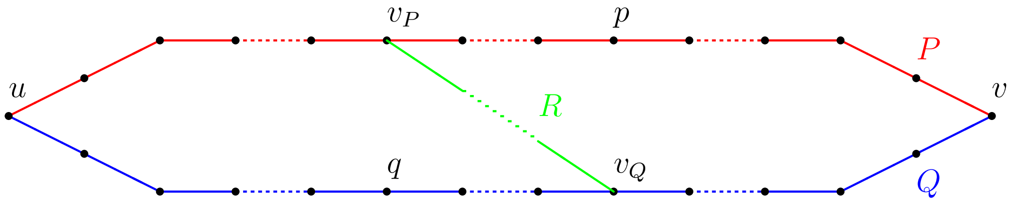

FIGURE 3. Path R

we may take the rest of the pairing to be random. We will later argue that the number of pairs of such paths P and Q is small enough for our argument to work.

Given a vertex p_r at distance r from p along P and a vertex q_s at distance s from q along Q, let X_{p_r,q_s} be the event that p_r and q_s are connected by a path of length at most $\log_{d-1} n - r - s - \omega(n)$ which is internally disjoint from P and Q. Define the random variable $Y_{P,Q} = \sum_{r=0}^{\lceil \log_{d-1} n - \omega \rceil} \sum_{s=0}^{\lceil \log_{d-1} n - \omega \rceil - r} X_{p_r,q_s}$, so that $Y_{P,Q} = 0$ only if G does not contain a path R satisfying (11) and (12) with respect to P and Q.

Once again, we look at the process in which the neighbours of p_r and q_s are exposed, then their neighbours are exposed, and so on, but we do not consider the neighbours of p_r and q_s on P or Q, so as to expose the sets $N_i(p_r)$ and $N_i(q_s)$ containing only the vertices at distance at most i from p_r and q_s that can be reached by paths internally disjoint from P and Q. Clearly, p_r and q_s are joined by a path as in (11) and (12) only if $N_i(p_r)$ and $N_i(q_s)$ join in at most $\frac{1}{2}(\log_{d-1} n - r - s - \omega(n))$ steps. The probability of this can be calculated as in the earlier sections, and we conclude that

$$\begin{aligned}
\mathbb{P}(Y_{P,Q} \geq 1) &\leq \mathbb{E}(Y) \\
&\leq \sum_{r=0}^{\lceil \frac{1}{2}(\log_{d-1} n - \omega) \rceil} \sum_{s=0}^{\lceil \frac{1}{2}(\log_{d-1} n - \omega) \rceil - r} \mathbb{P}(X_{p_r,q_s}) \\
&= 4 \sum_{r=0}^{\lceil \frac{1}{2}(\log_{d-1} n - \omega) \rceil} \sum_{s=0}^{\lceil \frac{1}{2}(\log_{d-1} n - \omega) \rceil - r} O\left((d-1)^{2\left(\frac{1}{2}(\log_{d-1} n - \omega) - r - s\right)} / n \right) \\
&= O\left((d-1)^{-\omega(n)} \right).
\end{aligned}$$

By Lemma 3.1 with $\alpha(n)$ any function tending to infinity sufficiently slowly, and $\beta(n) = o(\alpha(n))$, the number paths of length at most $\log_{d-1} n + \beta(n)$ between u and v is a.a.s. at most $2(d-1)^{\alpha(n)}$, since this is a bound on the number of vertices at distance $\lfloor \alpha(n) \rfloor$ from u. Thus, a.a.s. there are at most $\gamma(n)$ pairs of such paths between u and v, for any $\gamma(n) \to \infty$. Let Z denote this asymptotically almost sure event (for some α to be restricted shortly), and let $\bar{Z}$ be its complement.

Let $Y = \sum_{P,Q} Y_{P,Q}$, where the sum is over all pairs of distinct paths between u and v whose length is at most $d(u,v) + 4k + \omega(n)$. Then,

$$\begin{aligned}
\mathbb{P}(Y \geq 1) &= \mathbb{P}((Y \geq 1) \cap \bar{Z}) + \mathbb{P}((Y \geq 1) \cap Z) \\
&\leq \mathbb{P}(\bar{Z}) + \sum_{P,Q} \mathbb{P}((Y_{P,Q} \geq 1) \cap Z) \\
&\leq \mathbb{P}(\bar{Z}) + O\left(\gamma(n)(d-1)^{-\omega(n)} \right),
\end{aligned}$$

which tends to 0 provided $\gamma(n) = o((d-1)^{\omega(n)})$. As a consequence, a.a.s. there are no configurations satisfying the conditions in (11) and (12). Hence, a.a.s. the inequality $d(p,q) > \log_{d-1} n - \omega(n)$ holds, concluding the proof of the lemma. □

We are now ready to prove the main theorem.

Proof of Theorem 1.1. Part (i) of the theorem follows from Corollary 2.4 and equation (1).

For part (ii), we say that two points on a cycle are *diametrically opposite* if the distance between them around the cycle is $\lfloor l/2 \rfloor$, where the cycle has length l. Note that, to establish part (ii) of the theorem, it suffices to prove the following, since if there is a "short-cut" for any two vertices on a cycle, there is a short-cut for a pair of diametrically opposite ones.

Claim 2: *Asymptotically almost surely, there are d paths $P_1, \dots, P_d$ connecting u and v satisfying the following. Each P_i has length at most $\log_{d-1} n + \omega(n)$. Furthermore, the union of any two distinct paths P_i and P_j forms a cycle $C_{i,j}$ passing through u and v, whose length $l_{i,j}$ satisfies $|2\log_{d-1} n - l_{i,j}| < \omega(n)$. Moreover, for every pair of points p and q that are diametrically opposite on $C_{i,j}$, $d(p,q) \geq \log_{d-1} n - \log_{d-1}\log_{d-1} n - \omega(n)$.*

We now prove Claim 2. Note that, by (i), for all $\varepsilon > 0$, there is K sufficiently large that $\mathbb{P}(d(u,v) > \log_{d-1} n - K) > 1 - \varepsilon$ for all n sufficiently large. Let $u_1, \dots, u_d$ and $v_1, \dots, v_d$ be the neighbours of u and v, respectively. We may imitate the proof of Lemma 2.3 by first picking neighbours u' of u and v' of v, and then exploring the neighbourhoods of these vertices after deleting u and v from the graph. From the proof of Lemma 2.3 it is evident that the probability that the shortest path from u' to v' avoiding u and v has length at least $\log_{d-1} n + k$ is for sufficiently large n at most some function that tends to 0 as k increases. It follows that, for all $\varepsilon > 0$, and for every pair of neighbours u_i and v_j of u and v respectively, for some sufficiently large K, there is such a path with length at most $\log_{d-1} n + K$ with probability at least $1 - \varepsilon$ when n is sufficiently large. So, with probability at least $1 - d\varepsilon$, we can choose d such uv-paths $P_1, \dots, P_d$, where the neighbours of u and v on P_i are u_i and v_i, respectively. In each case we may select a shortest path with these specifications. Then $P_1, \dots, P_d$ must be $(K+1)$-near-geodesics. We show that these paths satisfy the conditions in the statement of this claim. First note that, by definition, these are geodesics between vertices of distance $K+1$ of its ends. In particular, their length is bounded above by $2(K+1) + (2(K+1) + \dot{(}u,v)) \leq \log_{d-1} n + 5K + 4 < \log_{d-1} n + \omega(n)$ for n sufficiently large.

Also, given any $i, j \in \{1, \dots, d\}$, with $i \neq j$, we may assume by Claim 1 that there is no vertex in the intersection of P_i and P_j at distance larger than $\omega(n)/3$ from both u and v. For ω growing slowly enough, there is a.a.s. no point in common that is at most $\omega(n)/3$ from u or v either, since Lemma 2.1 implies that a.a.s. neither u nor v is in a short cycle. As a consequence, the union $C_{i,j}$ of the paths P_i and P_j is a cycle. From the bounds on $d(u,v)$, $C_{i,j}$ has length at least $2\log_{d-1} n - 2K$ and at most $2\log_{d-1} n + 2K$.

To prove the statement about all diametrically opposite points p and q on $C_{i,j}$, we may rework the argument in Lemma 3.2. The claim proved above shows that every short path of the type we are interested in must use some edge not on $P = P_i$ or $Q = P_j$. Arguing as before, we only need to eliminate the existence of A such that (10) holds.

The same argument shows that, for any fixed such p and q, with $Y_{P,Q}$ defined as before, we again have $\mathbb{P}(Y_{P,Q} \geq 1) = O\left((d-1)^{-\omega(n)}\right)$.

Now apply this inequality to the $O(\log_{d-1} n)$ pairs of vertices p and q diametrically opposite on C. Also, put

$$\omega(n) = \log_{d-1}\log_{d-1} n + \gamma(n).$$

Then the probability that $Y_{P,Q} \geq 1$ for at least one of these choices of p and q is $O\left((d-1)^{-\gamma(n)}\right)$. Hence, if $\gamma(n) \to \infty$, we have a.a.s. for all such p and q, $d(p,q) \geq f(n) - \log_{d-1}\log_{d-1} n - \gamma(n)$. Replacing γ by ω gives the final statement in Claim 2, with probability at least $1 - 2\varepsilon + o(1)$. This statement is true for all $\varepsilon > 0$. That fact implies that the final statement in Claim 2 holds a.a.s. (This can be regarded as "letting $\varepsilon \to 0$ sufficiently slowly".) Combining this with part (i) proves Claim 2, since, although there may be different functions at the different occurrences of ω, they can be made the same. This completes the proof of Claim 2. □

4. Final remarks

In this article we have examined the "shape" of random regular graphs. This brings up related questions.

Our proof of the main theorem can be seen to give more: a.a.s. for every pair of short (i.e. bounded length) paths, one containing u and one containing v, there is an almost geodesic cycle containing both of these paths. (Note that two such paths are a.a.s. disjoint because u and v are typically far apart.) We also show that the paths referred to in the theorem each contain a geodesic between the two vertices at distance k from its ends, for any k tending to infinity with n.

Recall that a geodesic cycle C in G is a cycle in which for every two vertices u and v in C, the distance $d_G(u,v)$ is equal to $d_C(u,v)$. A significant open problem is to determine whether in a random d-regular graph, a.a.s. almost all pairs of vertices lie in a geodesic cycle. It is not even known if at least one geodesic cycle of length asymptotic to $\log_{d-1} n$ exists a.a.s.

We may also draw conclusions on how "thin" the topological triangles are in random regular graphs. Consider the proof of Lemma 2.3, which analyses the time at which two simultaneous breadth-first reaches from u and from v join each other. The proof is concerned with an accurate estimate of the probability that there are no joins by a time near i_0. It is easy to see from the ideas in the proof that for large K, the second join is quite likely to occur by time $i_0 + K$, and furthermore that the first two joins are quite likely to be in branches that diverged, in the breadth first search from u, at time less than K, and similarly from v. Let u' and v' be the points of divergence near u and v. Then the joins give two paths P and Q from u' to v', the shorter of which, say P, is geodesic, and we can choose another vertex, w, on Q, of distance K from u', such that the resulting two subpaths of Q to u' and v' from w are both geodesic. Thus u', v', w form a geodesic triangle. By Lemma 3.2 (noting P and Q are $2K$-near-geodesics from u to v), the distance between the midpoints of P and Q is a.a.s. at least $\log_{d-1}(n) - \omega(n)$, where $\omega(n)$ is any function tending to ∞. Hence the midpoint p of P a.a.s. has distance at least $\frac{1}{2}\log_{d-1} n - \omega(n)$ from the union of the other two sides of the geodesic triangle u', v', w. So this triangle is not "thin." The probability in the above statements tends to 1 if we let $K \to \infty$ sufficiently slowly.

Recall also the definition of δ-hyperbolicity from Section 1. Clearly, by taking four points almost equally spaced around a k-roughly geodesic cycle, we see that a graph containing such a cycle of length t cannot be $(t/4-k-c)$-hyperbolic, for c some small constant ($c=0$ if t is divisible by 4). This can be strengthened slightly by using the vertices $\{u',v',p\}$ defined as above, and with the midpoint of Q as a fourth vertex. This shows that a random d-regular graph, for $d\geq 3$, is a.a.s. *not* δ-hyperbolic for $\delta=(\log_{d-1}n)/2-\omega(n)$. For an upper bound, it is obviosuly δ-hyperbolic for δ equal to half of the diameter of the graph, which is $(\log_{d-1}n)/2+O(\log\log n)$ by the main result of [5].

Finally, it would be interesting to see to what extent the geometric properties we have addressed in this paper are preserved if the model of regular graphs changes. Particular models of interest might be random Cayley graphs, random lifts of regular graphs, and one can consider also some deterministic models of expanders.

Acknowledgements. The authors are indebted to anonymous referees for their valuable comments and suggestions.

References

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series, 1964.
2. O. Angel, A.E. Holroyd, D. Romik and B. Virág, Random sorting networks, *Adv. Math.* **215** (2007), no. 2, 839–868.
3. B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal of Combinatorics* **1** (1980), 311–316.
4. B. Bollobás. *Random Graphs.* Academic Press, 1985.
5. B. Bollobás and W. Fernandez de la Vega, The diameter of random regular graphs, *Combinatorica* **2** (1982), 125–134.
6. M. Bonk and O.Schramm, Embeddings of Gromov hyperbolic spaces, *Geometric and Functional Analysis* **10** (2000), 266–306.
7. A. M. Frieze and L. Zhao, Optimal construction of edge-disjoint paths in random regular graphs, *Proceedings of SODA* (1999), 346–355.
8. M. Gromov, Hyperbolic groups, S. M. Gersten (ed.), *Essays in Group Theory, MSRI Publ.* **8** Springer (1987), 75–263.
9. N. C. Wormald, The asymptotic connectivity of labelled regular graphs, *Journal of Combinatorial Theory, Series B* **28** (1986), 43–47.
10. N. C. Wormald, Models of random regular graphs, *Surveys in Combinatorics, 1999*, London Mathematical Society Lecture NoteSeries **267** (J.D. Lamb and D.A. Preece, eds) Cambridge University-Press, Cambridge,pp. 239–298, 1999.

Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel
E-mail address: itai.benjamini@weizmann.ac.il

Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo-SP, Brazil
E-mail address: choppen@ime.usp.br

Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel
E-mail address: eran.ofek@weizmann.ac.il

Department of Mathematics and Statistics, Dalhousie University, Halifax-NS, Canada B3H 3J5
E-mail address: pralat@mathstat.dal.ca

Department of Combinatorics and Optimization, University of Waterloo, Waterloo-ON, Canada N2L 3G1
E-mail address: nwormald@math.uwaterloo.ca