Optimal Construction of Edge-Disjoint Paths in Random Graphs

Andrei Z. Broder* Alan M. Frieze† Stephen Suen‡
Eli Upfal§

Abstract

Given a graph $G = (V, E)$ with n vertices, and m edges, and a family of κ pairs of vertices in V, we are interested in finding for each pair (a_i, b_i), a path connecting a_i to b_i, such that the set of κ paths so found is edge-disjoint. (For arbitrary graphs the problem is $\mathcal{NP}$-complete, although it is in $\mathcal{P}$ if κ is fixed.)

We present a polynomial time randomized algorithm for finding the optimal number of edge disjoint paths (up to constant factors) in the random graph $G_{n,m}$, for all edge densities above the connectivity threshold. (The graph is chosen first, then an adversary chooses the pairs of endpoints.) Our results give the first tight bounds for the edge disjoint paths problem for any non-trivial class of graphs.

1 Introduction

Given a graph $G = (V, E)$ with n vertices, and m edges, and a set of κ pairs of vertices in V, we are interested in finding for each pair (a_i, b_i), a path connecting a_i to b_i, such that the set of κ paths so found is edge-disjoint.

*Digital Systems Research Center, 130 Lytton Avenue, Palo Alto, CA 94301, USA. E-mail: broder@src.dec.com.

†Department of Mathematics, Carnegie-Mellon University, Pittsburgh, PA 15213. A portion of this work was done while the author was visiting IBM Almaden. Supported in part by NSF grants CCR9024935 and CCR9225008. E-mail: af1p@euler.math.cmu.edu.

‡Department of Mathematics, Carnegie-Mellon University, Pittsburgh, PA 15213.

§IBM Almaden Research Center, San Jose, CA 95120, USA, and Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel. Work at the Weizmann Institute supported in part by the Norman D. Cohen Professorial Chair of Computer Science, by a MINERVA grant, and by the Israeli Academy of Science. E-mail: eli@wisdom.weizmann.ac.il.

For arbitrary graphs the related decision problem is $\mathcal{NP}$-complete, although it is in $\mathcal{P}$ if κ is fixed – Robertson and Seymour [RS]. Nevertheless, this negative result can be circumvented for certain classes of graphs. Peleg and Upfal [PU] presented a polynomial time algorithm for the case where G is a (sufficiently strong) bounded degree expander graph, and $\kappa \leq n^\epsilon$ for a small constant ϵ that depends on the expansion property of the graph. (A precise upper bound for ϵ was not computed, but it is clearly less that 1/3). This result has recently been improved by Broder, Frieze, and Upfal [BFU]: G still has to be a (sufficiently strong) bounded degree expander but κ can now grow as fast as $n/(\ln n)^\theta$, where θ depends only on the expansion properties of the input graph, but is at least 7.

For the vertex-disjoint paths problem Kleinberg and Tardos [KT] come within an $O(\log n)$ factor of the maximum possible κ for a class of planar graphs. In random graphs Shamir and Upfal have shown in [SU] that any set of up to $O(\sqrt{n})$ pairs can be connected via vertex-disjoint paths; similar results using efficient flow techniques were also obtained by Hochbaum [H]. These results were proved for graphs with $m \geq Kn \log n$ random edges, where K is a sufficiently large constant.

Let D be the median distance between pairs of vertices in G. Clearly it is not possible to connect more than $O(m/D)$ pairs of vertices by edge-disjoint paths, for all choices of pairs, since some choice would require more edges than all the edges available. In the case of bounded degree expanders, this absolute upper bound on κ is $O(n/\log n)$. The results mentioned above use only a vanishing fraction of the set of edges of the graph, thus are far from reaching this upper bound. In contrast, in this work we show that for the basic models of random graphs, $G_{n,m}$ and $G_{n,p}$, the absolute upper bound is achievable within a constant factor, and we present an algorithm that constructs the required paths in polynomial time.

As usual, let $G_{n,p}$ denote a random graph with vertex set $\{1, 2, \ldots, n\} = [n]$ in which each possible edge is included independently with probability p, and let $G_{n,m}$ denote a random graph also with vertex set $[n]$ and exactly m edges, all sets of m edges having equal probability. The degree of a vertex v is denoted by $d_G(v)$.

Our main result is formulated in the following theorem.

Theorem 1 *Let $m = m(n)$ be such that $d = 2m/n \geq (1 + o(1)) \ln n$. Then, as $n \longrightarrow \infty$, with probability $1 - o(1)$, the graph $G_{n,m}$ has the following property: there exist positive constants α and β such that for all sets of pairs of vertices $\{(a_i, b_i) \mid i = 1, ..., \kappa\}$ satisfying:*

(i) $\kappa = \lceil \alpha m \ln d / \ln n \rceil$,

(ii) for each vertex v, $|\{i : a_i = v\}| + |\{i : b_i = v\}| \leq \min\{d_G(v), \beta d\}$,

there exist edge-disjoint paths in G, joining a_i to b_i, for each $i = 1, 2, \ldots, \kappa$. Furthermore, there is an $O(nm^2)$ time randomized algorithm for constructing these paths.

A similar result holds for $G_{n,p}$, with $d = np$, and $\kappa = \lceil \alpha n^2 p \ln d/(2 \ln n) \rceil$.

This result is the best possible up to constant factors. For (i) note that the distance between most pairs of vertices in G is $\Omega(\log n/\log d)$, and thus with m edges we can connect at most $O(m \log d/\log n)$ pairs. For (ii) note that a vertex v can be the endpoint of at most $d_G(v)$ different paths. Furthermore suppose that $d \geq n^\gamma$ for some constant $\gamma > 0$ so that $\kappa \geq \lceil \alpha\gamma nd/2 \rceil$. Let $\epsilon = \alpha\gamma/3$, $A = [\epsilon n]$, and $B = [n] \setminus A$. Now with probability 1-o(1) there are less than $(1 + o(1))\epsilon(1 - \epsilon)nd$ edges between A and B in $G_{n,m}$. However almost all vertices of A have degree $(1+o(1))d$ and if for these vertices we ask for $(1 - \epsilon/2)d$ edge-disjoint paths to vertices in B then the number of paths required is at most $(1 + o(1))\epsilon(1 - \epsilon/2)nd < \kappa$, but, without further restrictions, this many paths would require at least $(1 - o(1))\epsilon(1 - \epsilon/2)nd > (1 + o(1))\epsilon(1 - \epsilon)nd$ edges between A and B which is more than what is available. This justifies an upper bound of $1 - \epsilon/2$ for β of Theorem 1.

We note that we have proved similar optimal results for the vertex disjoint paths problem in random graphs [BFSU].

The construction of $n/(\ln n)^\theta$ edge-disjoint paths on expander graphs that was described in [BFU], was achieved through the use of the Lovász Local Lemma [EL]. Sets of possible paths were constructed for each pair, and the Local Lemma was applied to prove that there is a global choice of one path per set such that all the choices are edge-disjoint. However, this approach can only be used when the total number of edges in the final set of disjoint paths is a vanishing fraction of the number of edges in the graph; inherently, it does not lead to optimal bounds.

Here we address the problem in a different way. After a randomization phase, similar to the one in [BFU], the disjoint paths are constructed one after the other, and all the edges seen during the construction are deleted from the graph. The paths connecting each pair are chosen through a "random walk" type process. The crux of the analysis is to show that after a number of pairs have already been connected, the remaining graph is sufficiently connected to continue with this process. To prove that, we use a good estimate on the eigenvalues of the intermediate graphs generated by the algorithm. (Since we cannot throw logarithmic factors at our trouble spots, the proofs are rather intricate, although the algorithm itself is quite simple.) Eventually the number of pairs not yet connected becomes small enough that we can use [BFU] directly.

The disjoint paths problem has numerous algorithmic applications. One that has received increased attention in recent years is in the context of communication networks. The only efficient way to transmit high volume communication, such as in multimedia applications, is through disjoint paths that are dedicated to one pair of processors for the duration of the communication. To efficiently utilize the network one needs a very simple algorithm that with minimum overhead constructs a large number of edge disjoint paths between a given set of requests. The algorithm we study is simple and easy to implement (after eliminating some steps that are needed only for the proof), and thus suggests some possibly good practical heuristics.

In Section 3 we present a very brief overview of the algorithm. The details of the algorithm are exposed in Section 4. The remainder of the paper gives the analysis.

2 Preliminaries

The paper contains a number of unspecified constants of which α and β above are the first. Exact values could be given but it is easier for us *and the reader* if we simply give the relations between them. New constants will be introduced as $C_0, C_1, \ldots$ without further comment. Furthermore, specific constants have been chosen for convenience, we made no attempt to optimize them, and, in general, we only claim that inequalities dependent on n hold for n sufficiently large.

For a graph $G = (V, E)$ we use $\delta(G)$ and $\Delta(G)$ to denote the smallest and largest degrees respectively. For a set $S \subseteq V$ we define its neighbour set, $N(S, G)$, as

$$N(S,G) = \{v \in V \setminus S : \exists w \in S \text{ such that } \{v,w\} \in E\}.$$

For $S \subseteq V$, we use $G[S]$ to denote the subgraph of G induced by S.

The *Chernoff bounds* on the tails of the Binomial $\mathrm{Bin}(n, \theta)$ that we use are

$$\mathbf{Pr}(\mathrm{Bin}(n,\theta) \le (1-\epsilon)n\theta) \le e^{-\epsilon^2 n\theta/2}, \tag{1}$$

$$\mathbf{Pr}(\mathrm{Bin}(n,\theta) \ge (1+\epsilon)n\theta) \le e^{-\epsilon^2 n\theta/3}, \tag{2}$$

valid for $0 \le \epsilon \le 1$.

3 Overview of the algorithm

Our algorithm divides naturally into the five phases sketched below.

Phase 1: Partition G into five edge-disjoint graphs $G_i = (V_i, E_i)$, $1 \leq i \leq 5$. Phase 2 will use only the graph G_1; Phase 3 will use only the graph G_2; Phase 4 will use only the graph G_3 and Phase 5 will use only the graphs G_4 and G_5. The partition is such that $V_1 = V$ but $V_2 = V_3 = V_4 = V_5 \subseteq V$ with $|V_2| = n - o(n)$.

Phase 2: Choose a random multiset $Z = \{z_1, \ldots, z_{2\kappa}\}$ of 2κ points in V_2. Connect the endpoints $\{(a_i, b_i) \mid i = 1, ..., \kappa\}$ to the newly chosen points in an arbitrary manner via edge-disjoint paths in G_1 using a flow algorithm. Let $\tilde{a}_i$ (resp. $\tilde{b}_i$) be the vertex connected to a_i (resp. b_i). The original problem is now reduced to finding edge-disjoint paths from $\tilde{a}_i$ to $\tilde{b}_i$ for each i. (This randomization was used in [BFU] and has its roots in Valiant's routing algorithm [VB].)

Phase 3: For each $z \in Z$ in turn, we do a random walk of length $\tau = \lceil C_0 \ln n / \ln d \rceil$ in G_2, starting at z. We remove the edges of the j'th walk before embarking on the $j+1$'st. This keeps the paths constructed edge-disjoint. The terminating endpoint of the walk starting at $\tilde{a}_i$ (resp. $\tilde{b}_i$) will be denoted by $\hat{a}_i$ (resp. $\hat{b}_i$) for $1 \leq i \leq \kappa$. The analysis below shows that in almost every $G_{n,p}$, the (multi)-set of vertices $\hat{a}_1, \ldots, \hat{a}_\kappa, \hat{b}_1, \ldots, \hat{b}_\kappa$ is *not too far* from being independently, uniformly distributed.

Phase 4: For each i in turn, we repeatedly do a certain type of random walk in G_3 starting from $\hat{a}_i$ until one of these walks ends at $\hat{b}_i$. We keep the last walk as our path from $\hat{a}_i$ to $\hat{b}_i$ and remove from G_3 all edges seen in these walks. (The analysis below promises that this process will succeed **whp**[1] for most i.) Not every pair $(\hat{a}_i, \hat{b}_i)$ will be successfully connected in this phase but the final path for each pair that succeeds is the concatenation of the paths from a_i to $\tilde{a}_i$, and from b_i to $\tilde{b}_i$ found in Phase 2, the paths from $\tilde{a}_i$ to $\hat{a}_i$ and $\tilde{b}_i$ to $\hat{b}_i$ found in Phase 3, and the path from $\hat{a}_i$ to $\hat{b}_i$ found here.

Phase 5: At the end of Phase 4, **whp**, there will be at most $n^{1-\epsilon}$ pairs $(\hat{a}_i, \hat{b}_i)$, for a constant $\epsilon > 0$, which have not been joined by paths. We use the algorithm of [BFU] to join them by edge disjoint paths, using only the edges of G_4 and G_5, and then construct the final paths as above.

[1] In this paper, an event $\mathcal{E}_n$ is said to occur **whp** (with high probability) if $\mathbf{Pr}(\mathcal{E}_n) = 1 - o(n^{-9/10})$ as $n \longrightarrow \infty$. For reasons explained in Section 7, the usual 1-o(1) does not suffice here.

To prove Theorem 1 it suffices to show that for almost every $G_{n,p}$:

- Phases 1 and 2 will succeed for *all* choices of $a_1, \ldots, b_\kappa$ and *almost every* choice of $z_1, \ldots, z_{2\kappa}$.
- Phases 3, 4, and 5 are successful for *almost every* choice of $z_1, \ldots, z_{2\kappa}$ and *any* mapping $\{\tilde{a}_1, \ldots, \tilde{a}_\kappa, \tilde{b}_1, \ldots, \tilde{b}_k\} \leftrightarrow \{z_1, \ldots, z_{2\kappa}\}$

Note that to prove these facts we have to consider only one experiment, namely choose $G_{n,p}$ or $G_{n,m}$ at random and then $z_1, \ldots, z_{2\kappa}$ at random. From this we can deduce that almost every $G_{n,p}$ or $G_{n,m}$ is such that for *all* choices of $a_1, \ldots, b_\kappa$ and almost every choice of $z_1, \ldots, z_{2\kappa}$, we can find edge-disjoint paths $a_i - \tilde{a}_i - \hat{a}_i - \hat{b}_i - \tilde{b}_i - b_i$ for $1 \leq i \leq \kappa$.

4 Description of the algorithm

The input to our algorithm is a random graph $G_{n,p}$ and a set of pairs of vertices $\{(a_i, b_i) \mid i = 1, ..., \kappa\}$ satisfying the premises of Theorem 1. The output is a set of κ edge-disjoint paths, $P_1, \ldots, P_\kappa$ such that P_i connects a_i to b_i.

4.1 Phase 1.

We start by partitioning G into five edge-disjoint graphs $G_i = (V_i, E_i)$, for $1 \leq i \leq 5$. Phase 2 will use only G_1; Phase 3 will use only G_2; Phase 4 will use only G_3; Phase 5 will use only G_4 and G_5. The partition is such that $V_1 = V$ but $V_2 = V_3 = V_4 = V_5 \subseteq V$ with $|V_1| = n - o(n)$.

In this construction, we use the notion of a k-core. The k-core of a graph H is the largest $S \subseteq V(H)$ which induces a subgraph of minimum degree at least k. It is unique and can be found by repeatedly removing vertices of degree less than k until what remains is empty or has minimum degree k.

The algorithm below starts by constructing preliminary versions of these graphs, denoted G_i' for $1 \leq i \leq 5$. Then edges and vertices are deleted from $G_2', \ldots, G_5'$ in order to achieve certain minimum degree properties.

algorithm SPLIT
begin
  Divide E into E_i', $1 \leq i \leq 5$ by placing each edge of E independently with probability $5/6$ in E_1', and with probability $1/24$ into each of E_i' for $2 \leq i \leq 5$.
  For $1 \leq i \leq 5$ set $G_i' \leftarrow (V, E_i')$
  $K \leftarrow \lfloor d/2 \rfloor$-core of G_1'
  For $2 \leq i \leq 5$ set $G_i \leftarrow (K, E_i' \cap (K \times K))$
  while $\exists v \in K$ such that $\min\{d_{G_i}(v) : 2 \leq i \leq 5\} < d/30$ **do**
    For $2 \leq i \leq 5$ remove v and its adjacent edges from G_i.
    $K \leftarrow K \setminus \{v\}$
  od
  For $2 \leq i \leq 5$ set $V_i \leftarrow V(G_i)$ and set $E_i \leftarrow E(G_i)$
  $G_1 \leftarrow (V, E \setminus (E_2 \cup E_3 \cup E_4 \cup E_5))$
end SPLIT

We will show later (Lemma 1) that **whp** this algorithm terminates with $|K| = n - o(n)$. Note that SPLIT ensures that

- The final graphs G_i, $2 \leq i \leq 5$ have the same vertex set K.
- Every $v \in K$ has degree at least $\lfloor d/2 \rfloor$ in G_1 and at least $\lfloor d/30 \rfloor$ in each of G_i, $2 \leq i \leq 5$.
- If $v \in V \setminus K$ then $d_{G_1}(v) = d_G(v)$.

4.2 Phase 2.

Choose $z_1, z_2, \ldots, z_{2\kappa} \in V_2$ uniformly and randomly with replacement. Let Z denote the multiset $\{z_1, z_2, \ldots, z_{2\kappa}\}$. We are going to replace the problem of finding paths from a_i to b_i by that of finding paths from $\tilde{a}_i$ to $\tilde{b}_i$, where $\{\tilde{a}_1, \tilde{b}_1, \tilde{a}_2, \tilde{b}_2, \ldots, \tilde{a}_\kappa, \tilde{b}_\kappa\} = Z$ as multisets. Let A denote the multiset $\{a_1, b_1, a_2, b_2, \ldots, a_\kappa, b_\kappa\}$.

We connect A to Z via edge-disjoint paths in the graph G_1 using network flow techniques. We construct a network as follows

- Each undirected edge of G_1 gets capacity 1.

- Each member of A becomes a source and each member of Z becomes a sink.
- If a vertex occurs r times in A then it becomes a source with supply r, and if a vertex occurs s times in Z, then it becomes a sink with demand s.

Then we find a flow from A to Z that satisfies all demands. Since the maximum flow has integer values, it decomposes naturally into $|A|$ edge-disjoint paths (together perhaps with some cycles). If a path joins a_i to $z \in Z$, then we let $\tilde{a}_i = z$. Similarly, if a path joins b_i to $z \in Z$, then we let $\tilde{b}_i = z$.

Thus Phase 2 finds edge-disjoint paths $P_i^{(1)}$ from a_i to $\tilde{a}_i$ and $P_i^{(5)}$ from $\tilde{b}_i$ to b_i, $1 \leq i \leq \kappa$, where the vertices $\tilde{a}_1, \tilde{b}_1, \tilde{a}_2, \tilde{b}_2, \ldots, \tilde{a}_\kappa, \tilde{b}_\kappa \in V_2$ are chosen uniformly at random with replacement. (Some of these paths may of course be single vertices.) On the other hand there may be some difficult conditioning involved in the pairing of $\tilde{a}_i$ with $\tilde{b}_i$, $1 \leq i \leq \kappa$. We deal with this in Phase 3.

4.3 Phase 3.

We construct paths $P_i^{(2)}$, $P_i^{(4)}$ in G_2 with start vertices $\tilde{a}_i, \tilde{b}_i$ respectively for $1 \leq i \leq \kappa$. Each path is constructed by simulating a random walk of length $\tau = \lceil C_0 \ln n / \ln d \rceil$ from each start point. The endpoints of $P_i^{(2)}$, $P_i^{(4)}$ are $\hat{a}_i, \hat{b}_i$ respectively. The edges of a walk are deleted from G_2 before the next one starts. This keeps the paths edge-disjoint. We construct these walks with start points Z in the *random* order $z_1, z_2, \ldots, z_{2\kappa}$ (This random order is helpful in the proof of (21) below). W_i denotes the walk started at z_i; it ends at $\hat{z}_i$. Γ_i denotes the state of G_2 after the edges of $W_1, W_2, \ldots, W_{i-1}$ have been deleted.

A *random walk* on an undirected graph (or multigraph) $G = (V, E)$ is a Markov chain $\{X_t\}$ on V associated with a particle that moves from vertex to vertex according to the following rule: The probability of a transition from vertex v, of degree d_v to a vertex w is $1/d_v$ if $\{v, w\} \in E$ and 0 otherwise. (For multigraphs, each edge out of a vertex is an equally likely exit; loops are counted as two exits.) Its stationary distribution, denoted by π or $\pi(G)$, is given by $\pi_v = d_v/(2|E|)$. A trajectory W of length τ is a sequence of vertices $[w_0, w_1, \ldots, w_\tau]$ such that $\{w_t, w_{t+1}\} \in E$ for $1 \leq t < \tau$. The Markov chain induces a probability distribution on trajectories in the usual way. We use $P_G^{(\tau)}(a, b)$ to denote the probability that a random walk in G of length τ starting at a terminates at b.

4.4 Phase 4.

The problem now is to find edge-disjoint paths $P_i^{(3)}$ joining $\hat{a}_i$ to $\hat{b}_i$, for $1 \leq i \leq \kappa$. We use only the edges of G_3 to avoid conflict with paths already chosen in $G_1 \cup G_2$. Thus eventually we can take P_i to be the path (after removing cycles if necessary) that joins a_i to $\tilde{a}_i$ via $P_i^{(1)}$, $\tilde{a}_i$ to $\hat{a}_i$ via $P_i^{(2)}$, $\hat{a}_i$ to $\hat{b}_i$ via $P_i^{(3)}$, $\hat{b}_i$ to $\tilde{b}_i$ via $P_i^{(4)}$ and $\tilde{b}_i$ to b_i via $P_i^{(5)}$. (Actually, this will only be true for *most* i. If $d = O(\ln n)$ then a fifth phase may be necessary to find paths for some indices i.)

The paths $P_i^{(3)}$ are again found by simulating a random walk. The reader might expect us to choose a random walk from those with endpoints $\hat{a}_i, \hat{b}_i$. The main problem with this is that the distribution of $\hat{b}_i$ may be significantly different from the steady state distribution of a walk from $\hat{a}_i$ in G_3. If we choose a walk in this manner then deleting it will condition the graph in a way which is complex to analyse, especially as we have to repeat the procedure κ times.

We overcome this by choosing a set of random walks and use rejection sampling to make the final walk have the correct distribution. There is still the complication that the $\hat{b}_i$ are chosen before we do the walks. This leads to the subroutine WALK described next. WALK$(\hat{a}_i, \hat{b}_i, \hat{\Gamma}_i, \Gamma_j, z_j)$ generates a series of random walks of length τ in $\hat{\Gamma}_i$ starting from $\hat{a}_i$. The graph $\hat{\Gamma}_i$ is such that $\hat{\Gamma}_i \subseteq G_3$ with $V(\hat{\Gamma}_i) = V(G_3) = V(\Gamma_j)$, and j is defined by $z_j \equiv \tilde{b}_i$. The last walk generated ends at $\hat{b}_i$ which, by the construction used in the previous phase, has the distribution

$$\hat{p}_v = P_{\Gamma_j}^{(\tau)}(z_j, v).$$

The somewhat strange method used to generate these walks will be further explained in Section 7.

We maintain an array of counters, $S[v]$, for $v \in V_3$, initially all 0. The counter $S[v]$ shows how many times v was used as a start point of a walk. No vertex is allowed to be the start of more than $d/120$ walks, thus there is a chance (in fact, only when $d = O(\ln n)$) that for some pairs of vertices Phase 4 will not connect them. The indices of these pairs are kept in a set L and considered in the last phase.

subroutine WALK($\hat{a}_i, \hat{b}_i, \hat{\Gamma}_i, \Gamma_j, z_j$)
begin
/* By construction, $z_j = \tilde{b}_i$. */
$p_v \leftarrow P^{(\tau)}_{\hat{\Gamma}_i}(\hat{a}_i, v)$ for $v \in V_3$
$\hat{p}_v \leftarrow P^{(\tau)}_{\Gamma_j}(z_j, v)$ for $v \in V_2$ (the distribution of $\hat{b}_i$.)
$p_{\min} \leftarrow \min\{p_v : v \in V_3\}$
$\hat{p}_{\max} \leftarrow \max\{\hat{p}_v : v \in V_2\}$
Choose r from the geometric distribution with probability
    of success $s = p_{\min}/\hat{p}_{\max}$
if $S[\hat{a}_i] + r \geq d/120$ **then**
        $L \leftarrow L \cup \{i\}$
        exit WALK
else
        $S[\hat{a}_i] \leftarrow S[\hat{a}_i] + r$
fi
for k **from** 1 **to** $r - 1$ **do**
        Choose x_k according to
            $\mathbf{Pr}(x_k = v) = (p_v - \hat{p}_v p_{\min}/\hat{p}_{\max})/(1 - s)$
od
$x_r \leftarrow \hat{b}_i$
for k **from** 1 **to** r **do**
        Pick a walk $\hat{W}_k$ of length τ in $\hat{\Gamma}_i$ according to the
            distribution on trajectories, conditioned on
            start point $= \hat{a}_i$ and end point $= x_k$
od
output $\hat{W}_1, \hat{W}_2, \ldots, \hat{W}_r$
end WALK

The distributions p_v and $\hat{p}_v$ can be computed in $O(nm\tau)$ time by computing powers of the transition matrix, after which a random walk can be found in $O(n\tau)$ time. (For details see [BFU].) The analysis will show that in the range of interest, **whp**, s is bounded away from zero by a constant, hence the expected total running time of WALK is $O(nm\tau)$.

We can now describe the complete algorithm for Phase 4.

algorithm GenPaths
begin
Let $E(W)$ denote the edge set of a walk W.
$\hat{\Gamma}_1 \leftarrow G_3$
for $i = 1$ **to** κ **do**
Define j such that $z_j \equiv \tilde{b}_i$.
Execute Walk$(\hat{a}_i, \hat{b}_i, \hat{\Gamma}_i, \Gamma_j, z_j)$
if $i \notin L$ **then**
$P_i^{(3)} \leftarrow \hat{W}_r$
$\hat{\Gamma}_{i+1} \leftarrow \hat{\Gamma}_i \setminus \left(\cup_{j=1}^r E(\hat{W}_j) \right)$
fi
od
end GenPaths

The expected running time of GenPaths is $O(\kappa n m \tau) = O(nm^2)$.

4.5 Phase 5.

Use (a slight modification of) the algorithm of [BFU] to find edge disjoint paths in $G_4 \cup G_5$ from $\hat{a}_i$ to $\hat{b}_i$, for $i \in L$.

5 Analysis of Phase 1

In Lemmas 1,2,3 we calculate with $G_{n,p}$ and deduce the result for $G_{n,m}$ via

$$\mathbf{Pr}(G_{n,m} \in \mathcal{P}) \leq O(n^{1/2})\, \mathbf{Pr}(G_{n,p} \in \mathcal{P})$$

for any graph property $\mathcal{P}$, assuming $m = \binom{n}{2} p \geq n$.

Our immediate task regarding this phase is to prove

Lemma 1 *With high probability, the vertex set $K = V_i$, $2 \leq i \leq 5$ satisfies*

$$|K| = n - o(n).$$

Proof: Let K_0 denote the value of K immediately prior to the execution of the while loop of Split, that is, K_0 is the $\lfloor d/2 \rfloor$ core of G_1'. The final K is the largest subset S of K_0 for which $\delta(G_i'[S]) \geq d/30$, $2 \leq i \leq 5$.

Let $A_1 = \{v \in V : d_{G_1'}(v) \leq 2d/3\}$ and $A_i = \{v \in V_i : d_{G_i'}(v) \leq d/27,\ 2 \leq i \leq 5\}$. Let $A = \bigcup_{i=1}^5 A_i$. We show that **whp**

(i) $|A| = o(n)$.

(ii) $v \in V \setminus A$ implies that v has at most 50000 G-neighbors in A.

It follows from (ii) and the definition of A that for d sufficiently large $K \supseteq V \setminus A$ and then (i) implies the lemma.

We start from the fact that for any $k \geq 1$

$$\mathbf{Pr}(|A_1| \geq k) \leq \binom{n}{k} \mathbf{Pr}(\mathrm{Bin}(n-k, 5p/6) \leq 2d/3)^k.$$

Putting $k = k_1 = \lceil n^{61/62} \rceil$ and using (1) with $\epsilon = 1/5$, we obtain

$$\mathbf{Pr}(|A_1| \geq k_1) \leq \left(\frac{ne}{k_1}\right)^{k_1} e^{-k_1 d/61} = \left(\frac{ne^{1-d/61}}{k_1}\right)^{k_1} = o(n^{-2}). \tag{3}$$

Also, for any fixed $k \geq 300$,

$$\begin{aligned}\mathbf{Pr}(\exists v \in V : |N(v, G_{n,p}) \cap A_1| \geq k)& \\ &\leq n\binom{n}{k} p^k \mathbf{Pr}(\mathrm{Bin}(n-k-1, 5p/6) \leq 2d/3)^k \\ &\leq nd^k e^{-kd/61} = o(n^{-2}).\end{aligned} \tag{4}$$

Similarly for any $k \geq 1$ and $i \geq 2$

$$\mathbf{Pr}(|A_i| \geq k) \leq \binom{n}{k} \mathbf{Pr}(\mathrm{Bin}(n-k, p/24) \leq d/27)^k.$$

Now, putting $k = k_2 = \lceil n^{3999/4000} \rceil$ and using (1) with $\epsilon = 1/9$, we obtain

$$\mathbf{Pr}(|A_i| \geq k_2) \leq \left(\frac{ne}{k_2}\right)^{k_2} e^{-k_2 d/3900} = \left(\frac{ne^{1-d/3900}}{k_2}\right)^{k_2} = o(n^{-2}), \tag{5}$$

and for any fixed $k \geq 12000$,

$$\begin{aligned}\mathbf{Pr}(\exists v \in V : |N(v, G_{n,p}) \cap A_i| \geq k)& \\ &\leq n\binom{n}{k} p^k \mathbf{Pr}(\mathrm{Bin}(n-k-1, p/24) \leq d/27)^k \\ &\leq nd^k e^{-kd/3900} = o(n^{-2}).\end{aligned} \tag{6}$$

From (3) and (5) we conclude that **whp** $A = o(n)$, and from (4) and (6) we conclude that **whp**, no vertex in G has more than 50000 neighbors in A. □

6 Analysis of Phase 2

In this section we show that if our input graph $G = (V, E)$ is $G_{n,p}$ then **whp**, after we run SPLIT, we can find in G_1 edge-disjoint paths from a_i to $\tilde{a}_i$, and b_i to $\tilde{b}_i$, for $1 \leq i \leq \kappa$, for *any* choice for $a_1, \ldots, b_\kappa$ consistent with the premises of Theorem 1, and *almost every* choice for $\tilde{a}_1, \ldots, \tilde{b}_\kappa$.

Let A and Z be as defined in section 4.2. For $v \in V$, let $\alpha(v)$ be the multiplicity of $v \in A$ and $\xi(v)$ be the multiplicity of $v \in Z$. For $S \subseteq V$, let $\alpha(S) = \sum_{v \in S} \alpha(v)$ and $\xi(S) = \sum_{v \in S} \xi(v)$. For sets $S, T \subseteq V$, let $e_{G_1}(S, T)$ denote the number of edges of G_1 with an endpoint in S and the other endpoint in T. It suffices to prove that

$$e_{G_1}(S, \bar{S}) \geq \alpha(S) - \xi(S), \qquad \forall S \subseteq V. \tag{7}$$

We can then apply a theorem of Gale [G] (or see Bondy and Murty [BM], Theorem 11.8) to deduce the existence of the required flow in G_1 for the successful run of Phase 2b. (We must of course demonstrate (7) for all A satisfying the premises of Theorem 1 and almost all Z.)

We next prove three lemmas instrumental in proving Lemma 5 below:

Lemma 2 **Whp**, *for any* $v \in V_2$

$$\xi(v) \leq \beta d_{G_1}(v). \tag{8}$$

Proof: Observe that $\xi(v)$ has the distribution $\mathrm{Bin}(2\kappa, |V_2|^{-1})$. Thus

$$\mathbf{Pr}(\xi(v) > \beta d_{G_1}(v)) \leq \binom{2\kappa}{\beta d/2} |V_2|^{-\beta d/2} \leq \left(\frac{4e\kappa}{\beta d} \cdot \frac{1 + o(1)}{n} \right)^{\beta d/2}$$
$$\leq \left(\frac{12\alpha \ln d}{\beta \ln n} \right)^{\beta d/2} = o(n^{-3}),$$

provided that

$$\alpha \leq \beta e^{-6/\beta}/12. \tag{9}$$

□

Lemma 3 *(a)* G_1 *has the following property* **whp***: If* $S \subseteq V$ *and* $n_0 = ne^{-d/10} \leq |S| \leq n/2$ *then* $e_{G_1}(S, \bar{S}) \geq d|S|/5$*;*

(b) $G_{n,p}$ *has the following property* **whp***: If* $S \subseteq V$ *and* $|S| \leq n_0$ *then* $e_G(S, S) \leq 2|S|$.

Proof: (a) Note that G_1' is distributed as $G_{n,5p/6}$ and $G_1' \subseteq G_1$. But

$$\begin{aligned}
&\mathbf{Pr}(G_{n,5p/6} \text{ does not satisfy the property (a)}) \\
&\quad \leq \sum_{k=n_0}^{n/2} \binom{n}{k} \mathbf{Pr}(\mathrm{Bin}(k(n-k), 5p/6) \leq kd/5) \\
&\quad \leq \sum_{k=n_0}^{n/2} \binom{n}{k} \mathbf{Pr}(\mathrm{Bin}(k(n-k), 5p/6) \leq \frac{1}{2}k(n-k)\frac{5}{6}p) \\
&\quad \leq \sum_{k=n_0}^{n/2} \left(\frac{ne}{k}\right)^k \exp\left(-\frac{5}{48}k(n-k)p\right) = o(n^{-2}).
\end{aligned}$$

(b) Note that property (b) holds trivially for $|S| \leq 5$ or $d \geq 10 \ln n$, which implies $n_0 \leq 1$. Assume $d \leq 10 \ln n$ and $|S| \geq 6$. Thus,

$$\begin{aligned}
&\mathbf{Pr}(G_{n,p} \text{ does not satisfy the property in (b)}) \\
&\leq \sum_{k=6}^{n_0} \binom{n}{k} \binom{k(k-1)/2}{2k} p^{2k} \leq \sum_{k=6}^{n_0} \left(\frac{ne}{k}\right)^k \left(\frac{k^2 ep}{2k}\right)^{2k} = o(n^{-2}).
\end{aligned}$$

□

Lemma 4 *Let $I = \{v \in V : d_{G_1}(v) \leq 2\beta d\}$. Then* ***whp****, no two (distinct) vertices in I are within distance of two or less in G_1.*

Proof: Observe first that $I = \emptyset$ if $d \geq C \ln n$ for C sufficiently large. We can thus assume that $d = O(\log n)$ for the rest of the proof of this lemma. If $v \in I$ then either $d_G(v) \leq 2\beta d$ or $d_{G_1}(v) \neq d_G(v)$. The latter cannot be true, since it implies that $v \in V_2$, and then $d_{G_1}(v) \geq \lfloor d/2 \rfloor$. Thus,

$$\mathbf{Pr}(I \text{ contains an edge}) \leq n^2 p \left(\sum_{k=0}^{2\beta d} \binom{n-2}{k} p^k (1-p)^{n-k} \right)^2.$$

But

$$\begin{aligned}
&\sum_{k=0}^{2\beta d} \binom{n-2}{k} p^k (1-p)^{n-k} \\
&\quad = O\left(\binom{n}{2\beta d} p^{2\beta d} (1-p)^n \right) = O\left(\left(\frac{e}{2\beta}\right)^{2\beta d} e^{-d} \right) = O(n^{-.99}),
\end{aligned}$$

provided that

$$2\beta(1 - \ln 2\beta) \le 1/100. \tag{10}$$

Thus

$$\mathbf{Pr}(I \text{ contains an edge}) = o(n^{-9/10}).$$

A similar calculation deals with the case of a path of length two joining two vertices of I.

The rather tedious calculation for $G_{n,m}$ is left to the interested reader – see [BFF] for details of a similar calculation. □

Now inequality (7) will follow easily from

Lemma 5 *Define for every* $v \in V$

$$\theta(v) = \min\{d_{G_1}(v), \beta d\}$$

Then **whp** *for every* $S \subseteq V$ *satisfying* $1 \le |S| \le n/2$,

$$e_{G_1}(S, \bar{S}) \ge \theta(S) \tag{11}$$

where $\bar{S} = V \setminus S$ *and* $\theta(S) = \sum_{v \in S} \theta(v)$.

Proof: Since $\beta < 1/5$, Lemma 3(a) implies that for $S \subseteq V$ satisfying $n_0 \le |S| \le n/2$,

$$e_{G_1}(S, \bar{S}) \ge d|S|/5 \ge \theta(S).$$

Suppose next that $|S| \le n_0$. Let

$$I_1 = \{v \in V : d_{G_1}(v) \le 2\beta d\}$$
$$I_2 = \{v \in V : 2\beta d < d_{G_1}(v)\}$$

Let $S_i = S \cap I_i$, for $i = 1, 2$. Then

$$e_{G_1}(S, \bar{S}) = e_{G_1}(S_1, \bar{S}_1) + e_{G_1}(S_2, \bar{S}_2) - 2e_{G_1}(S_2, S_1).$$

But by Lemma 4 $G_1[S_1]$ has no edges, so that

$$e_{G_1}(S_1, \bar{S}_1) \ge \theta(S_1), \qquad \textbf{whp},$$

and using Lemma 3(b),

$$e_{G_1}(S_2, \bar{S}_2) \ge (2\beta d - 4)|S_2|, \qquad \textbf{whp},$$

and since Lemma 4 implies that **whp**, no vertex in S_2 is adjacent to two or more vertices in S_1, we have also

$$e_{G_1}(S_2, S_1) \leq |S_2|, \qquad \textbf{whp},$$

It thus follows that **whp**,

$$e_{G_1}(S, \bar{S}) \geq \theta(S_1) + (2\beta d - 6)\, |S_2| \geq \theta(S),$$

where the last inequality holds for sufficiently large n so that $\beta d > 6$. This shows (11). $\square$

We now show that the Lemma above implies equation (7). First note that condition (ii) in Theorem 1 implies $\alpha(v) \leq \theta(v)$, for all $v \in V$.

Second observe that SPLIT guarantees that for $v \in Z$, $\theta(v) = \beta d$, assuming that $\beta < 1/2$, since $Z \subseteq V_2$ and every $v \in V_2$ has degree at least $\lfloor d/2 \rfloor$ in G_1. Thus,

$$\begin{aligned}&\mathbf{Pr}(\exists v \in V \text{ such that } \xi(v) > \theta(v) \mid |V_2|)\\ &\leq |V_2| 2 \binom{\kappa}{\beta d} |V_2|^{-\beta d} \leq 2n \left(\frac{(1+o(1))e\alpha m \ln d}{n \beta d \ln n} \right)^{\beta d} = o(n^{-2}),\end{aligned}$$

provided that α is sufficiently small. We can thus assume that **whp** $\xi(v) \leq \theta(v)$ for all $v \in V$.

To complete the proof of equation (7), note first that for $|S| \leq n/2$, by Lemma 5,

$$e_{G_1}(S, \bar{S}) \geq \theta(S) \geq \alpha(S) \geq \alpha(S) - \xi(S);$$

and for $|S| \geq n/2$

$$e_{G_1}(S, \bar{S}) = e_{G_1}(\bar{S}, S) \geq \theta(\bar{S}) \geq \xi(\bar{S}) \geq \xi(\bar{S}) - \alpha(\bar{S}) = \alpha(S) - \xi(S).$$

7 Analysis of Phase 3

If a vertex $v \in V_2$ has degree $d_v^{(i)}$ in Γ_i then the steady state probability of a random walk in Γ_i being at v is

$$\pi_v^{(i)} = \frac{d_v^{(i)}}{\sum_{w \in V_2} d_w^{(i)}}.$$

The main thrust of our analysis is to show that the joint distribution of the $\hat{z}_i$ is close to that of independent samples from $\pi^{(i)}$, for $1 \leq i \leq 2\kappa$, that is **whp**, for $v \in V_2$ and $1 \leq i \leq 2\kappa$,

$$\mathbf{Pr}(\hat{z}_i = v \mid \Gamma_i,\, \hat{z}_j,\, j \neq i) = (1 + o(1))\pi_v^{(i)}. \tag{12}$$

In this case, when we come to join $\hat{a}_i$ to $\hat{b}_i$ then we can argue that $\hat{b}_i$ is (essentially) independent of $\hat{a}_i$. It is difficult to argue this for $\tilde{a}_i, \tilde{b}_i$ since they have been "chosen" as pairs by a flow algorithm. This is why we need Phase 3.

Let $\mathcal{E}_0$ denote the intersection of the events previously shown to hold **whp**. Let $P^{(i)}$ denote the transition probability matrix of a random walk on Γ_i. Let $\lambda^{(i)}$ be the second largest eigenvalue of $P^{(i)}$. We will prove later

Theorem 2 *For $1 \leq i \leq \kappa$ let $\mathcal{E}_i$ be the event that*
(a) the maximum degree $\Delta^{(i)}$ in Γ_i satisfies

$$\Delta^{(i)} \leq C_1 d; \tag{13}$$

and
(b) the minimum degree $\delta^{(i)}$ in Γ_i satisfies

$$\delta^{(i)} \geq d/C_2. \tag{14}$$

If $d \leq n^{1/10}$ then there exists a constant $\gamma = \gamma(C_1, C_2) > 0$ such that if $\mathcal{F}_i$ denotes the event that

$$\lambda^{(i)} \leq \gamma/\sqrt{d}. \tag{15}$$

and $\mathcal{U}_i = \mathcal{F}_i \cap \mathcal{E}_i \cap \cdots \cap \mathcal{F}_1 \cap \mathcal{E}_1 \cap \mathcal{E}_0$, then

$$\mathbf{Pr}(\mathcal{F}_i \mid \mathcal{E}_i, \mathcal{U}_{i-1}) = \mathbf{Pr}(\mathcal{F}_i \mid \mathcal{E}_i) = 1 - O(n^{-3}). \tag{16}$$

Proof: See Section 10. □

The reader will notice the bound $d \leq n^{1/10}$ in the theorem above. If $d > n^{1/10}$ we can randomly split the edge set of G into $r = \lceil 2d/n^{1/10} \rceil$ subsets $E_1, E_2, \ldots, E_r$, each of size roughly $m' = m/r$. We can similarly split the set of κ pairs into r roughly equal sets K_i. We can then use the graph $G_i = (V, E_i)$ to find paths for the pairs in K_i. Every vertex of every G_i will have degree roughly d/r **whp**. Hence, since

$$\frac{\kappa}{r} \leq \alpha m' \frac{\ln n}{\ln d}$$

we can apply Theorem 1 to each G_i, which implies that we succeed **whp** on each K_i, and thus we will succeed overall[2] with probability $1 - o(1)$. Therefore, without loss of generality, we can assume from now on that $d \leq n^{1/10}$.

[2]This is the reason for our definition of **whp**. The number r of subgraphs G_i is $O(n^{9/10})$ and we succeed with probability $1 - o(n^{-9/10})$ on each.

We now return to the analysis of Phase 3. We start by assuming that

$$\mathcal{E}_i \text{ and } \mathcal{F}_i \text{ hold for every } i. \tag{17}$$

It is well known that the second eigenvalue determines the rate of convergence of a Markov chain to its steady state. An explicit form of this result was obtained by Jerrum and Sinclair [SJ]: if $P^{(t)}_{\Gamma_i}(u,v)$ denotes the probability that a random walk of length t in Γ_i which starts at u will end at v, then assuming $\mathcal{E}_i$, we have

$$|P^{(t)}_{\Gamma_i}(u,v) - \pi^{(i)}_v| \leq \left(\lambda^{(i)}\right)^t \sqrt{\frac{\pi^{(i)}_v}{\pi^{(i)}_u}} \leq \left(\lambda^{(i)}\right)^t \sqrt{C_1C_2} \leq \frac{\gamma(C_1,C_2)^t}{d^{t/2}}\sqrt{C_1C_2}. \tag{18}$$

Since in the algorithm we take $t = \tau = \lceil C_0 \ln n / \ln d \rceil$, this implies (12).

We now proceed to show that the assumption (17) is indeed correct **whp**. We take $C_1 = 5$ and $C_2 = 60$. Since Γ_i is a subgraph of $G_{n,p}$, inequality (13) holds for all i **whp**, and since $\Gamma_1 = G_2$ and by construction $\delta(G_2) \geq d/30$, inequality (14) holds for Γ_1, thus $\mathcal{E}_1$ holds. Applying Theorem 2, we see that $\mathcal{F}_1$ holds **whp**. We continue by showing inductively that for $i \geq 0$,

$$\mathbf{Pr}(\mathcal{U}_i \mid \mathcal{E}_0) = 1 - O(in^{-3}).$$

Since

$$\mathbf{Pr}(\mathcal{U}_{i+1} \mid \mathcal{U}_i) = \frac{\mathbf{Pr}(\mathcal{U}_{i+1})}{\mathbf{Pr}(\mathcal{U}_i)} = \mathbf{Pr}(\mathcal{F}_{i+1} \mid \mathcal{E}_{i+1}, \mathcal{U}_i)\,\mathbf{Pr}(\mathcal{E}_{i+1} \mid \mathcal{U}_i),$$

and

$$\mathbf{Pr}(\mathcal{U}_{i+1} \mid \mathcal{E}_0) = \mathbf{Pr}(\mathcal{U}_{i+1} \mid \mathcal{U}_i)\,\mathbf{Pr}(\mathcal{U}_i \mid \mathcal{E}_0),$$

and given Theorem 2, we only need to prove that

$$\mathbf{Pr}(\mathcal{E}_{i+1} \mid \mathcal{U}_i) = 1 - O(n^{-3}), \tag{19}$$

which reduces to proving that given $\mathcal{U}_i$, the removal of the walks $W_1, \ldots, W_i$ from G_2 does not reduce the degree of any vertex to less than $d/60$.

Now assume $\mathcal{U}_i$. Consider the walk W_i on Γ_i. For $v \in V_2$, let $Z_{i,v}$ denote the number of edges incident with v that are covered by W_i, and let $N_{i,v}$ be the number of visits to v during W_i. Let $q_k = \mathbf{Pr}(N_{i,v} = k \mid \mathcal{U}_i)$ for $k \geq 1$. We claim that independently of $W_1, W_2, \ldots, W_{i-1}$, there exist constants C_3 and C_4 so that

$$q_k \leq \frac{C_4 C_3^{k-1} \ln n}{d^{k-1} n \ln d}. \tag{20}$$

To prove (20) for $k = 1$, fix Γ_i, and let $h_v(t)$ be the probability that the walk is at v at time t. Then

$$h_v(0) = 1/|V_2| \leq C_1 C_2 \pi_v^{(i)} \tag{21}$$

since the walk starts from z_i which is a vertex chosen uniformly at random in $|V_2|$. (The last inequality follows assuming that $\mathcal{U}_i$ occurs, and thus $\mathcal{E}_i$ occurs.)

We next show inductively that for all $v \in V_2$, we have $h_v(t) \leq C_1 C_2 \pi_v^{(i)}$. This follows from stationarity equations and

$$h_v(t+1) = \sum_{w \in N(v:\Gamma_i)} \frac{h_w(t)}{d_w^{(i)}} \leq C_1 C_2 \pi_v^{(i)}. \tag{22}$$

Hence, since $\tau = \lceil C_0 \ln n / \ln d \rceil$ and $\pi_v^{(i)} \leq C_1 C_2 / n$, there is a constant C_4 so that

$$q_1 \leq \sum_{t=0}^{\tau} h_v(t) \leq \frac{C_4 \ln n}{n \ln d}.$$

We next prove (20) for $k \geq 2$. Fix Γ_i and for vertex v let ρ_v be the probability that a random walk of length τ from v ever returns to v. Since a return to v requires at least two steps, we obtain from equation (18), that there exists a constant C_3 such that

$$\rho_v \leq \tau \pi_v^{(i)} + \sqrt{C_1 C_2} \sum_{t \geq 2} \frac{\gamma(C_1, C_2)^t}{d^{t/2}} \leq \frac{C_3}{d}. \tag{23}$$

This gives (20) since

$$q_k \leq (\rho_v)^{k-1} \sum_{t=1}^{\tau} h_v(t).$$

We now show that (20) implies (19). First (20) implies that for any constant c

$$\begin{aligned} \mathbf{E}(e^{2cN_{i,v}} \mid \mathcal{U}_i, W_1, \ldots, W_{i-1}) &\leq 1 + \sum_{k \geq 1} e^{2ck} \frac{C_4 C_3^{k-1} \ln n}{d^{k-1} n \ln d} \\ &\leq 1 + \frac{2C_4 e^{2c} \ln n}{n \ln d} \end{aligned}$$

Clearly $Z_{i,v} \le 2N_{i,v}$. Thus for any constant $c > 0$ and any $t > 0$,

$$\begin{aligned}
&\mathbf{Pr}\Big(\sum_{j=1}^{i} Z_{j,v} \ge t \;\Big|\; \mathcal{U}_i\Big)\\
&\qquad \le e^{-ct}\,\mathbf{E}\Big(\exp\Big(2c\sum_{j=1}^{i} N_{j,v}\Big) \;\Big|\; \mathcal{U}_i\Big)\\
&\qquad \le e^{-ct}\Big(1+\frac{2C_4 e^{2c}\ln n}{n\ln d}\Big)\,\mathbf{E}\Big(\exp\Big(2c\sum_{j=1}^{i-1} N_{j,v}\Big) \;\Big|\; \mathcal{U}_i\Big)\\
&\qquad \le e^{-ct}\exp\Big(\frac{2C_4 e^{2c}\ln n}{n\ln d}\Big)\,\mathbf{E}\Big(\exp\Big(2c\sum_{j=1}^{i-1} N_{j,v}\Big) \;\Big|\; \mathcal{U}_{i-1}\Big)\frac{\mathbf{Pr}(\mathcal{U}_{i-1})}{\mathbf{Pr}(\mathcal{U}_i)}\\
&\qquad \le \exp\Big(-ct + i\frac{2C_4 e^{2c}\ln n}{n\ln d}\Big)\frac{1}{\mathbf{Pr}(\mathcal{U}_i)}\\
&\qquad \le \exp\Big(-ct + 2\kappa\frac{2C_4 e^{2c}\ln n}{n\ln d}\Big)(1+O(in^{-3}))\\
&\qquad \le 2\exp(-ct + 4\alpha C_4 e^{2c} d)
\end{aligned}$$

Taking $t = d/60$, $c = 240$, and $\alpha \le (4C_4e^{480})^{-1}$, we obtain that

$$\mathbf{Pr}\Big(\sum_{j=1}^{i} Z_{j,v} \ge \frac{d}{60} \;\Big|\; \mathcal{U}_i\Big) \le 2n^{-3},$$

and since the minimum degree in G_2 is at least $d/30$, this proves (19). (Recall that $C_2 = 60$.)

8 Analysis of Phase 4

We start by discussing the subroutine WALK. Consider a modification of WALK defined as follows:

subroutine WALK1$(\hat{a}_i, \hat{\Gamma}_i, \Gamma_j, z_j)$
begin
/* By construction, $z_j = \tilde{b}_i$. */
$p_v \leftarrow P^{(\tau)}_{\hat{\Gamma}_i}(\hat{a}_i, v)$ for $v \in V(\hat{\Gamma}_i)$
$\hat{p}_v \leftarrow P^{(\tau)}_{\Gamma_j}(z_j, v)$ for $v \in V(\hat{\Gamma}_i)$ (the distribution of $\hat{b}_i$.)
$p_{\min} \leftarrow \min\{p_v : v \in V(\hat{\Gamma}_i)\}$
$\hat{p}_{\max} \leftarrow \max\{\hat{p}_v : v \in V(\hat{\Gamma}_i)\}$
$\bar{r} \leftarrow 0$
forever do
    $\bar{r} \leftarrow \bar{r} + 1$
    $S[\hat{a}_i] \leftarrow S[\hat{a}_i] + 1$
    if $S[\hat{a}_i] \geq d/120$ **then**
        $L \leftarrow L \cup \{i\}$
        exit WALK1
    fi
    Pick a walk $\bar{W}_{\bar{r}}$ of length τ according to the distribution on trajectories, conditioned on start point $= \hat{a}_i$
    Let $\bar{x}_{\bar{r}}$ be the terminal vertex of $\bar{W}_{\bar{r}}$
    With probability $\hat{p}_{\bar{x}_{\bar{r}}} p_{\min} / (p_{\bar{x}_{\bar{r}}} \hat{p}_{\max})$ accept $\bar{W}_{\bar{r}}$ and **exitloop**
od
output $\bar{W}_1, \bar{W}_2, \ldots, \bar{W}_{\bar{r}}$
end WALK1

Lemma 6 *In* WALK1, $\bar{x}_{\bar{r}}$ *is chosen according to the distribution* $\hat{p}$.

Proof: The probability s that a walk is accepted at the last step in the loop is given by

$$s = \sum_{v \in V(G)} p_v \frac{\hat{p}_v p_{\min}}{p_v \hat{p}_{\max}} = \frac{p_{\min}}{\hat{p}_{\max}}. \tag{24}$$

(Observe that $\hat{p}_{\max} \geq 1/|V(G)| \geq p_{\min}$.) Thus if S_0 is the value of $S[\hat{a}_i]$ at the start of WALK1 and $k_0 = d/120 - S_0$ then

$$\begin{aligned}&\mathbf{Pr}(\bar{x}_{\bar{r}} = v \mid \text{Step 14 is not executed}) \\ &\qquad = \frac{1}{1-(1-s)^{k_0}} \sum_{k=0}^{k_0-1} (1-s)^k p_v \frac{\hat{p}_v p_{\min}}{p_v \hat{p}_{\max}} = \hat{p}_v. \end{aligned} \tag{25}$$

Also, $\mathbf{Pr}$(Step 14 is executed) is equal to $(1-s)^{k_0}$ in both procedures. □

Hence $\bar{W}_{\bar{r}}$ is a random walk to a vertex chosen with distribution $\hat{p}$. Furthermore, since the minimum degree of any graph in which a walk is constructed is at least $d/60$ and the maximum degree is at most $5d$, we find

$$s \geq \frac{1}{300} \stackrel{\text{def}}{=} \sigma \tag{26}$$

and therefore the expected number of generated walks is constant.

There is a minor problem in that we want to choose the endpoints before we do the walks. This leads to the algorithm WALK described before. We now turn to its analysis.

Lemma 7 *Suppose that $\hat{b}_i$ is chosen from $V(G)$ with distribution $\hat{p}$. Then the walks $\bar{W}_1, \ldots, \bar{W}_{\bar{r}}$ in WALK1$(\hat{a}_i, G, \Gamma_j, z_j)$, and the walks $\hat{W}_1, \ldots, \hat{W}_r$ in WALK$(\hat{a}_i, \hat{b}_i, G, \Gamma_j, z_j)$ have the same distribution.*

Proof: Note first from the proof of Lemma 6 that $\bar{r}$ and r have the same truncated geometric distribution. Also, we have from Lemma 6 that $\bar{x}_{\bar{r}}$ and $x_r = \hat{b}_i$ have the same distribution. Consider next that for $v_1, v_2, \ldots, v_i \in V(G)$,

$$\begin{aligned}&\mathbf{Pr}(\bar{x}_1 = v_1, \ldots, \bar{x}_i = v_i \text{ and } \bar{r} > i) \\ &\quad = \prod_{j=1}^{i} \left(\left(1 - \frac{\hat{p}_{v_j} p_{\min}}{p_{v_j} \hat{p}_{\max}}\right) p_{v_j} \right) = \prod_{j=1}^{i} \left(p_{v_j} - \frac{\hat{p}_{v_j} p_{\min}}{\hat{p}_{\max}} \right) \\ &\quad = (1-s)^i \prod_{j=1}^{i} \left(\frac{p_{v_j} - \hat{p}_{v_j} p_{\min}/\hat{p}_{\max}}{1-s} \right) \\ &\quad = \mathbf{Pr}(x_1 = v_1, \ldots, x_i = v_i \text{ and } r > i).\end{aligned}$$

Thus $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{\bar{r}}$ and $x_1, x_2, \ldots, x_r$ have the same distribution. Finally, the lemma follows from the fact that the distribution of $\bar{W}_j$ conditional on $\bar{x}_j = v$ is clearly equal to that of $\hat{W}_j$ conditional on $x_j = v$. □

The net effect of GenPaths is to run Walk κ times. In the light of Theorem 2 we only need show that the minimum degree in G_3 is not made too small by the deletion of paths generated by Walk. This requires a slightly more complicated analysis than for Phase 3. The main problem in extending the analysis of Phase 3 is that we cannot argue now that (21) holds independently of previous walks. Each execution of Walk (or equivalently Walk1) involves a set of walks with the same starting point. The initial vertex of each set is chosen nearly randomly, but it is the same for each walk.

For the purpose of the analysis we relate to Walk1. Fix $v \in V_3$ and $1 \leq i \leq \kappa$ and let $\bar{W}_1, \bar{W}_2, \ldots, \bar{W}_r$ denote the walks made while trying to connect $\hat{a}_i$ to $\hat{b}_i$. We shall refer to these walks as the i'th *bundle*, B_i. We shall follow closely the line of proof used in the analysis of Phase 3, with all the events now referring to $\hat{\Gamma}_i$ rather than Γ_i. As before, the proof reduces to showing that given $\mathcal{U}_i$, the removal of the bundles $B_1, \ldots, B_i$ from G_3 does not reduce the degree of any vertex in G_3 to less than $d/60$.

The stationary distribution on $\hat{\Gamma}_i$ is denoted $\hat{\pi}^i$.

Lemma 8 *Assuming $\mathcal{U}_i$, the probability that a fixed vertex v is visited by bundle i is less than*

$$\frac{C_5 \ln n}{n \ln d}.$$

Proof: Assume for a moment that the probability of a walk being accepted is decreased to exactly σ (See (26)). This can only increase the number of visits to v but the number of walks is now independent of the start point. For every walk in the bundle we can show via (20) and (21) applied to $\hat{\Gamma}_i$ that the expected number of visits to v is less than $C_1 C_2 \tau \hat{\pi}_v^{(i)}$, thus the expected total number of visits is less than

$$\frac{C_1 C_2}{\sigma} \tau \hat{\pi}_v^{(i)} = \frac{C_5 \ln n}{n \ln d}.$$

□

Lemma 9 *Assume $\mathcal{U}_i$ and consider a random walk of length τ in $\hat{\Gamma}_i$ starting from vertex v. Then*
(a) The probability that the walk returns k times to v is less than $(C_3/d)^k$.
(b) For any vertex $u \neq v$, the probability that u is visited k times is less than $(C_3/d)^k$.

Proof: As before let ρ_v be the probability that a random walk of length τ from v ever returns to v. From (23) applied to $\hat{\Gamma}_i$,

$$\rho_v \leq \frac{C_3}{d}.$$

For part (a) notice that the probability of k returns to v is bounded by

$$\sum_{t=2}^{\tau}\left(\hat{\pi}_v^{(i)} + \frac{\gamma(C_1,C_2)^t}{d^{t/2}}\right)\rho_v^{k-1} \leq \left(\frac{C_3}{d}\right)^k.$$

For part (b) the probability of k visits to u is bounded by

$$\frac{C_2}{d}\rho_u^{k-1} + \sum_{t=2}^{\tau}\left(\hat{\pi}_u^{(i)} + \frac{\gamma(C_1,C_2)^t}{d^{t/2}}\right)\rho_u^{k-1} \leq \left(\frac{C_3}{d}\right)^k.$$

(The first term deals with the case when u is a neighbour of v.) □

We are ready now to evaluate the number of visits to a fixed vertex v. To this goal we will distinguish between *free* visits and *start* visits. If $v = \hat{a}_i$, then v undergoes $|B_i|$ visits, as the start point of all the walks in the bundle. All other visits to v are free visits. In particular a return visit to $\hat{a}_i$ is a free visit.

Analogously to the analysis of Phase 3, let $N_{i,v}$ be the number of *free* visits to v during B_i and let $q_k = \mathbf{Pr}(N_{i,v} = k \mid \mathcal{U}_i)$ for $k \geq 1$. We claim that independently of $B_1, B_2, \ldots, B_{i-1}$, there exists a constant C_6 so that

$$q_k \leq \frac{C_5 C_6^{k-1} \ln n}{d^{k-1} n \ln d}. \tag{27}$$

To simplify notation view the r walks in bundle B_i as a single walk X_t that restarts from $\hat{a}_i$ every τ steps. Let $h_v(t)$ be the probability that this walk is at v at time t. Then

$$q_k \leq \sum_{1 \leq t \leq \tau} h_v(t)\, \mathbf{Pr}(k-1 \text{ free visits to } v \text{ after } t \mid \hat{a}_i, X_t = v).$$

Now given r, the number of walks in bundle B_i, the $k-1$ free visits to v can be distributed among the r walks in at most $\binom{k+r-2}{r-1}$ ways. So in view of Lemma 9 we have

$$\mathbf{Pr}(k-1 \text{ free visits to } v \text{ after } t \mid \hat{a}_i, X_t = v, r) \leq \binom{k+r-2}{r-1}\left(\frac{C_3}{d}\right)^{k-1}.$$

From Lemma 8

$$\sum_{1\le t\le\tau} h_v(t) \le \frac{C_5 \ln n}{n\ln d},$$

and using equation (26) we finally obtain that

$$q_k \le \frac{C_5 \ln n}{n \ln d} \sum_{1\le r\le\infty} \binom{k+r-2}{r-1}\left(\frac{C_3}{d}\right)^{k-1} \sigma(1-\sigma)^{r-1} = \frac{C_5\ln n}{n\ln d}\left(\frac{C_3}{\sigma d}\right)^{k-1},$$

which proves (27). From here we can proceed exactly as in the analysis of Phase 3 to show that the decrease in degree due to free visits is no more than $d/120$ **whp**, provided that α is small enough. By construction the reduction in degree due to start visits is at most $d/120$, so that the total reduction in degree during Phase 4 is at most $d/60$ as required. It remains to show that not too many pairs are deferred to phase 5.

9 Analysis of Phase 5

We start by bounding the number of pairs not connected in Phase 4. Recall that a pair $\hat{a}_i, \hat{b}_i$ is not connected iff the total number of walks started from $\hat{a}_i$ would have exceeded $d/120$.

Fix $v \in V_3$. From (22) and the discussion that follows it, we have that for every i

$$\mathbf{Pr}(\hat{a}_i = v) \le \frac{C_1^2 C_2^2}{n} \stackrel{\text{def}}{=} p.$$

Thus in view of (26) the number of starts from v is dominated by a random variable with the following probability generating function:

$$\sum_i \binom{\kappa}{i} p^i (1-p)^i \frac{\sigma x}{1-x(1-\sigma)} = \left(\frac{\sigma p x}{1-x(1-\sigma)} + 1 - p\right)^{\kappa}.$$

In general, given a probability generating function $f(x)$ for the random variable $X \ge 0$, and an integer $a \ge 1$, we have

$$\mathbf{Pr}(X \ge a) \le \frac{f(x)}{x^a}, \qquad \rho > x \ge 1$$

where ρ is the radius of convergence of f. So let X be the random variable that counts starts from v. Choosing

$$x = \frac{2+\sigma}{2+\sigma-\sigma^2} = 1 + \frac{\sigma^2}{2+\sigma-\sigma^2},$$

we obtain that

$$
\begin{aligned}
\mathbf{Pr}\left(X \geq \frac{d}{120}\right) &\leq \left(1+\frac{\sigma p}{2}\right)^{\kappa}\left(1+\frac{\sigma^2}{2+\sigma-\sigma^2}\right)^{-d/120} \\
&\leq \exp\left(\frac{\sigma p}{2}\kappa - \frac{\sigma^2}{3}\frac{d}{120}\right) \\
&= \exp\left(\frac{C_1^2 C_2^2 \sigma \alpha n d \ln d}{2n \ln n} - \frac{\sigma^2 d}{360}\right) \leq \exp\left(-\frac{\sigma^2 d}{400}\right)
\end{aligned}
$$

for α small enough.

At the end of Phase 4 we will be left with a set L of indices of pairs $(\hat{a}_i, \hat{b}_i)$ for which Phase 4 failed to find a path. The discussion above shows that $|L|$ is dominated in distribution by $\mathrm{Bin}(n, \exp(-\sigma^2 d/500))$, so **whp** $L = \emptyset$ if $d \geq 1000\sigma^{-2}\log n$ and otherwise $|L| \leq n^{1-\epsilon}$, for a *constant* $\epsilon > 0$. So assume that $d = O(\log n)$.

We join the pairs in L, using a modification of the algorithm of [BFU]. That algorithm starts by splitting the edges of an expander graph to form two disjoint expanding subgraphs. This is unnecessary here as G_4 and G_5 will suffice for the two expander graphs, namely G_4 can be used for the flow phase of [BFU] and then G_5 can be used for the random walks phase of [BFU]. The algorithm is capable of joining $\Omega(n/(\ln n)^c)$ pairs for some constant $c > 0$, provided the graph in the flow phase has edge-expansion at least one and the second eigenvalue of the graph used in the random walks phase has second eigenvalue bounded away from 1. Here **whp** we have fewer than $n^{1-\epsilon}$ pairs, the graph G_4 has an edge expansion $\Omega(\ln n)$ and the graph G_5 has a second eigenvalue of size $O(1/\sqrt{\ln n})$. So from this point of view there is room to spare.

On the other hand [BFU] only deals with the case where the required path endpoints are distinct. We will replace the flow phase of [BFU] with the following procedure. Suppose $v \in V_3$ is required to be an endpoint $\lambda(v)$ times. We have

$$\sum_{v \in V_3} \lambda(v) = 2|L| \leq 2n^{1-\epsilon}.$$

Furthermore $\lambda(v)$ is dominated in distribution by $\mathrm{Bin}(\kappa, (1+o(1))n)$, hence $\mathbf{E}(\lambda(v)) = O(d \log d/\log n) = O(\log\log n)$ and **whp**

$$\lambda(v) \leq C_7 \log n/\log\log n \qquad \text{for all } v \in V_3.$$

We start Phase 5 by constructing for each $v \in V$, and $1 \leq i \leq \lambda(v)$ a set of $2/\epsilon$ random walks of length τ with start point v. We delete the edges of previous walks before beginning the next walk. The analysis of Phase 3 shows that we will succeed in constructing these walks **whp**, since in Phase 3 the average number of walks per start point was $O(d \log d/logn)$ and the maximum was βd while the corresponding numbers are now $o(1)$ and $O(d/\log d)$.

The probability that k such walks all end at the endpoints already visited is bounded by

$$\binom{4\epsilon^{-1}|L|}{k} O(n^{-k}) = O(n^{-\epsilon k}),$$

so **whp** for each v and i at least one of the $2/\epsilon$ random walks ends up at a previously unvisited point. Thus we can associate to each $v \in V_3$ a set of $\lambda(v)$ endpoints of walks started from v, and all these sets are disjoint. From here we can continue with the second phase of [BFU] on G_5.

It only remains to prove Theorem 2.

10 Proof of Theorem 2

Now it is not too difficult to verify that the second eigenvalue of the walk on Γ_1 is not too large. There is however a technical problem in the fact that we are deleting the edges of a random graph by a process that conditions the distribution. We overcome this by considering graphs with a fixed degree sequence and consequently the configuration model of multigraphs. (We need now only consider $G_{n,m}$. To handle $G_{n,p}$ we simply condition on the number of edges being close to the expected number.)

10.1 Configuration model

The graphs G_i, $2 \leq i \leq 5$ will be random given their degree sequences. This is because the executions of lines 5 and 8 of SPLIT do not condition the remaining graphs, once we are given their degree sequences. This idea has been used several times previously, see for example Bollobàs, Fenner, and Frieze [BFF].

The simplest model for graphs with a fixed degree sequence is the *configuration model* of Bollobás [B2] which is a probabilistic interpretation of the counting formula of Bender and Canfield [BC]. Let $\mathbf{d} = \{d_1, d_2, \ldots, d_\nu\}$ denote a degree sequence, $D_i = \{1, \ldots, d_i\} \times \{i\}$ for $1 \leq i \leq \nu$ and $D = \cup_{i=1}^{\nu} D_i$. Let $\Omega = \Omega(D)$ be the set of partitions of D into pairs. If $F \in \Omega$ then the multigraph $M = M(F)$ is defined as follows: $V(M) = [\nu]$ and there is an edge $\{i, j\}$ for every pair in F of the form $\{(x, i), (y, j)\}$ (for some x and

y). It is unfortunate that we have to introduce multi-graphs, but the salient properties of M are:

Lemma 10 **(a)** *If M is simple, then it is equally likely to be any simple graph with degree sequence* $\mathbf{d}$.

(b) $\mathbf{Pr}$(*M is simple*) $= \exp\{-O(\mu^2/\nu^2)\}$ *where $\mu = |D|/2$ is the number of edges in M – hence $2\mu/\nu$ is the average degree of M.*

We consider the probability space of multigraphs $M(F, \phi)$ where F is chosen randomly from Ω. We are interested in the case where

$$\delta = \min \mathbf{d} \geq d/C_2,$$
$$\Delta = \max \mathbf{d} \leq C_1 d.$$

It will be useful to think of F as being constructed sequentially.

algorithm CONSTRUCT
begin
    $F_0 \leftarrow \emptyset$; $R_0 \leftarrow W$
    for $t = 1$ **to** m^* **do**
        Choose $u_t \in R_{t-1}$ *arbitrarily*
        Choose v_t *randomly* from $R_{t-1} \setminus \{u_t\}$
        $F_t \leftarrow F_{t-1} \cup \{\{u_t, v_t\}\}$; $R_t \leftarrow R_{t-1} \setminus \{u_t, v_t\}$
    od
output F
end CONSTRUCT

It is important to observe that for any $t > 0$, $F \setminus F_t$ is a random member of $\Omega(R_t)$.

An important consequence of the above observation is that if we start with $M = M(F)$, then the multigraph obtained by removing from M the edges of a random walk W remains random. Indeed, we may imagine CONSTRUCT as performed in parallel with our walk W. Suppose our walk makes a transition from a vertex x and the current value of R_t in CONSTRUCT is R. The transition from x is equivalent to choosing a random member $u = u_t$ of D_x. If $u \in R$, then we perform one step of CONSTRUCT and pair u with a point $v = v_t \in R \setminus \{u\}$. If $v \in D_y$ for some y, then the walk makes a transition from x to y. If $u \notin R$ then v is the point already paired with u. Thus, since $F \setminus F_t$ is random, we see that removing from M the edges of a random walk results in a multigraph from a *random* configuration.

10.2 Random walks on configurations

We only discuss G_3 since the situation for G_2 is identical. Suppose G_3 has degree sequence $\mathbf{d}' = (d'_1, d'_2, \ldots, d'_\nu)$, where $\nu = n(1 - o(1))$. As observed, G_3 is random given its degree sequence. In our analysis, we want to consider G_3 as of the form $M(F)$ conditional on it being simple.

Each of the κ iterations deletes some pairs from F. Suppose $F^{(i)}$ denotes the remaining pairs at the start of iteration i and $D^{(i)} = \bigcup F^{(i)}$. If we ignore the condition that $M(F)$ is simple then $F^{(i)}$ is a random member of $\Omega(D^{(i)})$. This requires a little justification. Our algorithm produces paths by choosing $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_\kappa$ and $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_\kappa$ at random and by applying WALK. As observed in Lemma 7, this is equivalent to just applying WALK1 a number of times. By our arguments of the previous section, deleting edges in the walks produced by WALK1 leaves a random configuration.

Thus, we may imagine that initially we have a multigraph M_1 Then for $i \geq 2$ we apply GENPATHS to M_{i-1} and eventually produce M_i. In which case M_i is a multigraph from a random configuration (when its degree sequence is given).

All that remains now is to show that (15) holds with suitably high probability for M_i, $i \geq 1$, conditioned on it being simple.

10.3 Eigenvalues

We will prove (15) by imitating the proof of Kahn and Szemerédi [KS].

Let $\mathbf{d} = d_1, d_2, \ldots, d_n$ be a degree sequence with maximum $\Delta = o(n^{1/2})$ and minimum $\delta > 0$ such that $\Delta/\delta < \theta$ for some constant θ. (Strictly speaking we should be concerned with $\mathbf{d} = d_1, d_2, \ldots, d_\nu$ but $\nu = n - o(n)$ **whp** and n is "friendlier"). Let $M = M(F)$ be the multigraph on $[n]$ formed from a random configuration $F \in \Omega(\mathbf{d})$. Use e_{uv} to denote the number of edges joining vertices u and v. Consider the Markov chain of a random walk on M. The transition matrix of the chain is

$$P_{uv} = \frac{e_{uv}}{d_u}.$$

Note that since the Markov chain is reversible, all eigenvalues of P are real and the largest eigenvalue of P equals 1. The eigenvalues are denoted by

$$1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n.$$

We need to show that conditional on M being simple, with probability $1 - O(n^{-3})$

$$\rho^* = \max\{|\lambda_2|, |\lambda_n|\} \leq \gamma/\sqrt{d}, \tag{28}$$

where $d = \sum_{i=1}^{n} d_i/n$ and $\gamma = \gamma(\theta)$.

Lemma 11 *Let A be the matrix*

$$A_{uv} = \frac{e_{uv}}{d_u d_v},$$

and let

$$\rho_1 = \max\left\{|y^t A y| : \sum_u y_u = 0, \sum_u y_u^2 = 1\right\}.$$

Then $\rho^* \leq \Delta\rho_1$.

Proof: Let Q be the matrix

$$Q_{uv} = \frac{e_{uv}}{d_u^{1/2} d_v^{1/2}}.$$

Note that Q and P are similar, that is, $Q = DPD^{-1}$ where D is a diagonal matrix with diagonal elements $d_1^{1/2}, \ldots, d_n^{1/2}$ and so (λ, v) is an eigenvalue-eigenvector pair of P iff (λ, Dv) is an eigenvalue-eigenvector pair of Q. Since the largest eigenvalue of Q is 1 with eigenvector $(d_1^{1/2}, d_2^{1/2}, \ldots, d_n^{1/2})$, the Rayleigh quotient principle gives that

$$\rho^* = \max\left\{\frac{|\sum_{u,v} x_u Q_{uv} x_v|}{\sum_u x_u^2} \,\middle|\, \sum_u x_u d_u^{1/2} = 0\right\}.$$

Since

$$\sum_{u,v} x_u Q_{uv} x_v = \sum_{u,v} x_u d_u^{1/2} A_{uv} x_v d_v^{1/2}$$

and

$$\sum_u x_u^2 \geq \frac{1}{\Delta} \sum_u x_u^2 d_u,$$

we have, on putting $y_u = x_u d_u^{1/2}$,

$$\rho^* \leq \Delta \max\left\{|y^t A y| \,\middle|\, \sum_u y_u = 0, \sum_u y_u^2 = 1\right\} = \Delta\rho_1.$$

□

Following Kahn and Szemerédi, choose a real $\epsilon \in (0,1)$ (eventually ϵ will be fixed in equation (38)), and let

$$T = \left\{x \in \left(\frac{\epsilon}{n^{1/2}}\boldsymbol{Z}\right)^n \,\middle|\, \sum_u x_u = 0, \sum_u x_u^2 \leq 1\right\}.$$

where $\boldsymbol{Z}$ denotes the set of integers. Then, by considering the total volume of cubes of side $\epsilon/\sqrt{n}$ which have their centres in T, we see that

$$\begin{aligned}
|T| &\leq \left(\frac{n^{1/2}}{\epsilon}\right)^n \boldsymbol{V}\mathrm{ol}\left(\left\{x \in \boldsymbol{R}^n \;\middle|\; \sum_u x_u^2 \leq \left(1+\frac{\epsilon}{2}\right)\right\}\right) \\
&= \left(\frac{(2+\epsilon)n^{1/2}}{2\epsilon}\right)^n \frac{\pi^{n/2}}{\Gamma(n/2+1)} \leq \left(\frac{(2+\epsilon)n^{1/2}}{2\epsilon}\right)^n \frac{\pi^{n/2}e^{n/2}}{(n/2)^{n/2}\sqrt{\pi n}} \qquad (29) \\
&\leq \left(\frac{(2+\epsilon)\sqrt{2\pi e}}{2\epsilon}\right)^n.
\end{aligned}$$

Lemma 12 *Let ρ_1 be defined as in Lemma 11. We claim that*

$$\rho_1 \leq (1-\epsilon)^{-2}\rho$$

where

$$\rho = \max\{|x^t Ay| \mid x, y \in T\}.$$

Proof: Let $S = \{x \in \boldsymbol{R}^n \mid \sum_u x_u = 0, \sum_u x_u^2 \leq 1\}$. We first show that for every $x \in S$, there is a $y \in T$ such that $x - y \in \epsilon S$ and $\|x-y\| \leq \epsilon$. Suppose that for $i = 1, 2, \ldots, n$,

$$x_i = \epsilon m_i n^{-1/2} + f_i, \qquad m_i \in \boldsymbol{Z}, \qquad f_i \in [0, \epsilon n^{-1/2}).$$

Note that since $\sum_u x_u = 0$, we have $\sum_i f_i = \epsilon f n^{-1/2}$ where f is a non-negative integer less than n. Rearrange subscripts so that $m_i \leq m_j$ whenever $i \leq j$. Define a vector $y \in \boldsymbol{R}^n$ so that

$$y_u = \begin{cases} \epsilon(m_u+1)n^{-1/2}, & \text{if } u \leq f; \\ \epsilon m_u n^{-1/2}, & \text{if } u > f. \end{cases}$$

Then we have

$$\begin{aligned}
&(a) \quad \sum_u y_u = \sum_u x_u = 0. \\
&(b) \quad \sum_u y_u^2 \leq \sum_u x_u^2 \leq 1. \\
&(c) \quad \|x-y\| \leq \epsilon \quad (\text{since } |x_u - y_u| \leq \epsilon n^{-1/2}).
\end{aligned}$$

Thus, y is in T and has the required property. It follows that one can apply the above construction to obtain that for any $x \in S$, there are $x^{(0)}, x^{(1)}, \ldots$ in T such that

$$x = \sum_i \epsilon^i x^{(i)},$$

and therefore for any $x \in S$, there are $x^{(i)} \in T$ such that

$$x^t A x = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left(x^{(i)}\right)^t A x^{(j)} \epsilon^{i+j} \leq (1-\epsilon)^{-2} \max\{|y^t A z| : y, z \in T\}.$$

The lemma now follows. □

Now write

$$\rho = \max\{|x^t M y| : x, y \in T\}.$$

Our aim is to find a probabilistic upper bound for ρ of order $O(\Delta^{-3/2})$ which will verify (28). This is done by considering the random variables $X = X(x, y) = \sum_{u,v} x_u A_{uv} y_v$ where $x, y \in T$. Note that for any two distinct points in the configuration, the probability that the two points are joined by an edge is $1/(2m-1)$, where $2m = \sum_{i=1}^{n} d_i$. Thus for $u \neq v$,

$$\mathbf{E}[e_{uv}] = \frac{d_u d_v}{2m-1},$$

and

$$\mathbf{E}[e_{uu}] = \frac{d_u(d_u-1)}{2(2m-1)}.$$

Fix $x, y \in T$ and define

$$B = \left\{(u,v) \;\middle|\; 0 < |x_u y_v| < \Delta^{1/2}/n\right\}. \tag{30}$$

Let

$$X' = \sum_{(u,v)\in B} x_u A_{uv} y_v \quad \text{and} \quad X'' = \sum_{(u,v)\notin B} x_u A_{uv} y_v,$$

so that $X = X' + X''$.

10.4 Estimating X'

Note that

$$\mathbf{E}[X'] = \sum_{(u,v)\in B} \frac{x_u y_v}{2m-1} + \sum_{(u,u)\in B} \frac{x_u y_u (d_u - 1)}{2(2m-1)d_u}.$$

Write S_1 and S_2 for the first and second sums in the above equation. Then

$$|S_2| \leq \sum_{(u,u)\in B} \frac{|x_u y_u|(d_u-1)}{2(2m-1)d_u} \leq \frac{\Delta^{1/2}}{4m}. \tag{31}$$

For S_1, we follow Lemma 2.4 in Kahn and Szemerédi. Since $\sum_u x_u = \sum_v y_v = 0$ we have $\sum_{u,v} x_u y_v = 0$ and so

$$\Big|\sum_{(u,v)\in B} x_u y_v\Big| = \Big|\sum_{(u,v)\notin B} x_u y_v\Big|.$$

Now,

$$\Big|\sum_{(u,v)\notin B} x_u y_v\Big| \le \sum_{|x_u y_v|\ge \Delta^{1/2}/n} \frac{x_u^2 y_v^2}{|x_u y_v|} \le \frac{n}{\Delta^{1/2}} \sum_{u,v} x_u^2 y_v^2 \le \frac{n}{\Delta^{1/2}}.$$

Hence

$$|\,\mathbf{E}[X']| \le \frac{n}{(2m-1)\Delta^{1/2}} + \frac{\Delta^{1/2}}{4m} = (1+o(1))\frac{n}{2m\Delta^{1/2}}. \qquad (32)$$

We next show that X' is concentrated around its mean. For this we need some more notation. Recall that a configuration is a perfect matching F of the set $W = \cup_{i=1}^n \{i\}\times[d_i]$. We call the elements in W points and assume that the points in W are ordered lexicographically. For $\alpha \in W$, let $v(\alpha)$ denote the first component of α, and for a pair $e = \{\alpha, \beta\}$, in a configuration with $\alpha < \beta$, we write $t(e)$ for $v(\alpha)$ and $h(e)$ for $v(\beta)$. For real x, define

$$\chi(x) = \begin{cases} x, & \text{if } |x| < \Delta^{1/2}/n; \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$X' = \sum_{e\in F} \frac{\chi(x_{t(e)} y_{h(e)})}{d_{t(e)} d_{h(e)}} + \sum_{e\in F} \frac{\chi(x_{h(e)} y_{t(e)})}{d_{t(e)} d_{h(e)}} = X'_a + X'_b, \quad \text{say.} \qquad (33)$$

We next write $F = F_1 \cup F_2 \cup F_3$ where

$$\begin{aligned} F_1 &= \{e \in F : |x_{t(e)}| > n^{-1/2}/\epsilon\}, \\ F_2 &= \{e \in F : |y_{h(e)}| > n^{-1/2}/\epsilon, |x_{t(e)}| \le n^{-1/2}/\epsilon\}, \\ F_3 &= \{e \in F : |y_{h(e)}| \le n^{-1/2}/\epsilon, |x_{t(e)}| \le n^{-1/2}/\epsilon\}. \end{aligned}$$

Then let

$$X_i = \sum_{e\in F_i} \frac{\chi(x_{t(e)} y_{h(e)})}{d_{t(e)} d_{h(e)}}, \qquad \text{for } i = 1,2,3,$$

so that

$$X'_a = X_1 + X_2 + X_3.$$

Recall that $\Delta/\delta < \theta$, a constant. We claim

Lemma 13 *There are constants $B_i = B_i(\theta) > 0$, for $i = 1, 2, 3$, such that for any $t > 0$*

$$\mathbf{Pr}(|X_1 - \mathbf{E}[X_1]| \geq t\Delta^{-3/2}) \leq 2\exp(-tn + B_1 n) \tag{34}$$
$$\mathbf{Pr}(|X_2 - \mathbf{E}[X_2]| \geq t\Delta^{-3/2}) \leq 2\exp(-tn + B_2 n). \tag{35}$$
$$\mathbf{Pr}(|X_3 - \mathbf{E}[X_3]| \geq t\Delta^{-3/2}) \leq 2\exp(-tn + B_3 n) \tag{36}$$

Proof: We first prove (34). Assume without loss of generality that $|x_i| \geq |x_{i+1}|$ for all i and let the pairs $\{\alpha_i, \beta_i\}$ for $1 \leq i \leq m$ that compose F be ordered such that $\alpha_i < \beta_i$ and $\alpha_i < \alpha_{i+1}$. Recall that the order among points is lexicographic, thus $v(\alpha_i) \leq v(\alpha_{i+1})$ and $|x_{v(\alpha_i)}| \geq |x_{v(\alpha_{i+1})}|$.

Let $\equiv_k$ be the equivalence relation on Ω such that $F \equiv_k F'$ if and only if the sequences of the first k pairs in F and F' are identical. Write Ω_k for the set of equivalence classes, and $\mathcal{F}_k$ for the corresponding σ-algebra. Define $Y_k = \mathbf{E}[X_1|\mathcal{F}_k]$; that is, Y_k is a function from Ω to $\boldsymbol{R}$ so that $Y_k(F)$ equals the expected value of X_1 conditional on the first k pairs being exactly equal to the first k pairs in F. Now $Y_0, Y_1, \ldots, Y_m$ is a Doob martingale with $Y_0 = \mathbf{E}[X_1]$ and $Y_m = X_1$. Define $Z_k = Y_k - Y_{k-1}$. Note that as in Lemma 2.7 in [KS], if there is $f_k(\zeta)$ such that $Z'_k = \mathbf{E}[\exp(\zeta^2 Z_k^2)|\mathcal{F}_{k-1}] \leq f_k(\zeta)$, then for all t and $\zeta > 0$,

$$\mathbf{Pr}(|X_1 - \mathbf{E}[X_1]| \geq t) \leq 2e^{-\zeta t} \prod_{k=1}^{m} f_k(\zeta). \tag{37}$$

We next write down the distribution of Z_k. Define

$$\hat{\chi}(x, y) = \begin{cases} xy, & \text{if } |xy| < \Delta^{1/2}/n \text{ and } |x| > n^{-1/2}/\epsilon; \\ 0, & \text{otherwise.} \end{cases}$$

For a pair $e = \{\alpha, \beta\}$ in F with $\alpha < \beta$, write

$$q(e) = \frac{\hat{\chi}(x_{v(\alpha)}, y_{v(\beta)})}{d_{v(\alpha)} d_{v(\beta)}}.$$

Then

$$X_1(F) = \sum_{e \in F} q(e).$$

Note that we can express

$$Z_k(F) = \frac{2^{m-k}(m-k)!}{(2m-2k)!} \left\{ \sum_{F' \equiv_k F} X_1(F') - \frac{1}{2m-2k+1} \sum_{F'' \equiv_{k-1} F} X_1(F'') \right\}.$$

Let $\{\alpha, \beta\}$ be the k-th pair in F with $\alpha < \beta$ and let J be the set of points contained in the first k pairs in F. For $\eta \notin J - \{\beta\}$ and for $F' \equiv_k F$, we define F'_η as follows. Suppose that η is matched with γ in F'. Write $e = \{\alpha, \beta\}$, $f = \{\eta, \gamma\}$, $e' = \{\alpha, \eta\}$, $f' = \{\gamma, \beta\}$. Then F'_η is defined to be $(F' - \{e, f\}) \cup \{e', f'\}$, giving $F'_\eta \equiv_{k-1} F$ and $F'_\beta = F'$. Note also that $\{\{F'_\eta \mid \eta \notin J - \{\beta\}\} \mid F' \equiv_k F\}$ is a partition of $\{F'' \mid F'' \equiv_{k-1} F\}$. Thus,

$$Z_k(F) = \frac{2^{m-k}(m-k)!}{(2m-2k)!} \frac{1}{2m-2k+1} \sum_{F' \equiv_k F} \sum_{\eta \notin J} \left(X_1(F') - X_1(F'_\eta) \right).$$

Also, since

$$X_1(F') - X_1(F'_\eta) = q(e) + q(f) - q(e') - q(f'),$$

we have

$$Z_k(F) = \sum_{\eta \notin J} \sum_{\gamma \notin J, \gamma \neq \eta} \frac{q(\{\alpha, \beta\}) + q(\{\gamma, \eta\}) - q(\{\alpha, \eta\}) - q(\{\gamma, \beta\})}{(2m-2k+1)(2m-2k-1)}.$$

Note that since $\sum x_u^2 \leq 1$, there at most $n\epsilon^2$ indices u such that $|x_u| > n^{-1/2}/\epsilon$. Thus $Z_k = 0$ if $k \geq \epsilon^2 \Delta n$; Otherwise

$$2m - 2k - 1 \geq 2m - 2\epsilon^2 \Delta n - 1 \geq \delta n,$$

if we choose ϵ so that

$$\frac{\epsilon^2 \Delta}{\delta} = \frac{1}{3}. \tag{38}$$

Therefore

$$\begin{aligned} &|Z_k(F)| \\ &\leq \frac{1}{(\delta n)^2} \sum_{\eta \notin J} \sum_{\gamma \notin J, \gamma \neq \eta} \{|q(\{\alpha, \beta\})| + |q(\{\gamma, \eta\})| + |q(\{\alpha, \eta\})| + |q(\{\gamma, \beta\})|\}. \end{aligned}$$

Let

$$y^\alpha = \frac{1}{|x_{v(\alpha)}|} \min\{|yx_{v(\alpha)}|, \Delta^{1/2}/n\}.$$

Note that $x_{v(\alpha)} \geq \max\{x_{v(\beta)}, x_{v(\gamma)}, x_{v(\eta)}\}$ and that $|x_{v(\eta)}| \leq |x_{v(\alpha)}|$ implies $|x_{v(\eta)}|y^{\eta} \leq |x_{v(\alpha)}|y^{\alpha}$. Therefore

$$\begin{aligned}
|q(\{\alpha,\beta\})| &\leq \delta^{-2}|x_{v(\alpha)}|y^{\alpha}_{v(\beta)}, \\
|q(\{\gamma,\eta\})| &\leq \delta^{-2}(|x_{v(\gamma)}|y^{\gamma}_{v(\eta)} + |x_{v(\eta)}|y^{\eta}_{v(\gamma)}) \\
&\leq \delta^{-2}(|x_{v(\alpha)}|y^{\alpha}_{v(\eta)} + |x_{v(\alpha)}|y^{\alpha}_{v(\gamma)}), \\
|q(\{\alpha,\eta\})| &\leq \delta^{-2}|x_{v(\alpha)}|y^{\alpha}_{v(\eta)}, \\
|q(\{\gamma,\beta\})| &\leq \delta^{-2}(|x_{v(\gamma)}|y^{\gamma}_{v(\beta)} + |x_{v(\beta)}|y^{\beta}_{v(\gamma)}) \\
&\leq \delta^{-2}(|x_{v(\alpha)}|y^{\alpha}_{v(\beta)} + |x_{v(\alpha)}|y^{\alpha}_{v(\gamma)}).
\end{aligned}$$

Next, observe that since $\sum y_u^2 \leq 1$ implies $\sum |y_u| \leq n^{1/2}$, we have for example

$$\sum_{\eta \notin J} y^{\alpha}_{v(\eta)} \leq \sum_{\eta \notin J} |y_{v(\eta)}| \leq \Delta \sum_{w=1}^{n} |y_w| \leq \Delta n^{1/2}.$$

Thus, we have

$$|Z_k(F)| \leq 4\Delta^2\delta^{-4}|x_{v(\alpha)}|(y^{\alpha}_{v(\beta)} + n^{-1/2}).$$

Writing $Z'_k = \mathbf{E}[\exp(\zeta^2 Z_k^2)|\mathcal{F}_{k-1}]$, we have

$$Z'_k(F) \leq \frac{1}{2m-2k-1} \sum_{\nu \notin J-\{\beta\}} \exp(16\zeta^2\Delta^4\delta^{-8}(x_{v(\alpha)})^2(y^{\alpha}_{v(\nu)} + n^{-1/2})^2).$$

Take

$$\zeta = \Delta^{3/2} n, \tag{39}$$

which means that the expression

$$\begin{aligned}
&\zeta^2\Delta^4\delta^{-8}(x_{v(\alpha)})^2(y^{\alpha}_{v(\nu)} + n^{-1/2})^2 \\
&\quad = \zeta^2\Delta^4\delta^{-8}\Big\{(x_{v(\alpha)}y^{\alpha}_{v(\nu)})^2 + 2(x_{v(\alpha)}y^{\alpha}_{v(\nu)})x^{\alpha}_{v(\alpha)}n^{-1/2} + (x_{v(\alpha)})^2 n^{-1}\Big\}
\end{aligned}$$

is bounded by $4\theta^8\epsilon^{-2}$.

(Here we use $x_{v(\alpha)}y^{\alpha}_{v(\nu)} \leq \sqrt{\Delta}/n$ and $y^{\alpha}_{v(\nu)} \geq \epsilon/\sqrt{n}$ (since $y_{v(\nu)} \neq 0$) to get $x_{v(\alpha)} \leq \sqrt{\Delta}/(\epsilon\sqrt{n})$).

Hence, putting $B = \exp\{64\theta^8\epsilon^{-2}\}$ and using $e^x \leq 1 + xe^x$ for $x \geq 0$,

$$\begin{aligned}
Z'_k(F) &\leq 1 + \frac{B}{2m-2k-1} \sum_{\nu \notin J-\{\beta\}} \zeta^2\Delta^4\delta^{-8}(x_{v(\alpha)})^2(y^{\alpha}_{v(\nu)} + n^{-1/2})^2 \\
&\leq 1 + B\zeta^2\Delta^4\delta^{-9}n^{-1}(x_{v(\alpha)})^2 \sum_{\nu}(y^2_{v(\nu)} + 2|y_{v(\nu)}|n^{-1/2} + n^{-1}) \\
&\leq \exp(4B\zeta^2\Delta^5\delta^{-9}n^{-1}(x_{v(\alpha)})^2).
\end{aligned}$$

Writing $r(k) = \lceil k/\Delta \rceil$, we have for any $F \in \Omega$,

$$Z'_k(F) \leq \exp(4B\zeta^2\Delta^5\delta^{-9}n^{-1}(x_{r(k)})^2)$$

Thus, using (37), we have

$$\begin{aligned}\mathbf{Pr}(|X_1 - \mathbf{E}[X_1]| \geq t\Delta^{-3/2}) &\leq 2e^{-\zeta t\Delta^{-3/2}} \exp\left(\sum_{k=1}^{m} 4B\zeta^2\Delta^5\delta^{-9}n^{-1}(x_{r(k)})^2\right) \\ &\leq 2\exp\left(-tn + 4B\Delta^8\delta^{-9}n\sum_{k=1}^{m}(x_{r(k)})^2\right) \\ &\leq 2\exp\left(-tn + 4B\Delta^9\delta^{-9}n\right). \qquad (40)\end{aligned}$$

This proves (34).

The proof of (35) is almost identical even though F_2 has a slightly different definition to F_1. We simply re-order F according to $y_{v(\beta)}$, and go through the proof above, without using the condition $|x_{v(\alpha)}| \leq n^{-1/2}/\epsilon$.

The proof of (36) is much simpler. We use the more usual martingale argument (Alon and Spencer [ASE], Bollobás [B3], McDiarmid [M]); for now if $Y_k = \mathbf{E}(X_3 \mid \mathcal{F}_k)$ then $|Y_k - Y_{k-1}| \leq 4/(\epsilon^2 n\delta^2)$. Since we took (in (38)), $\epsilon = 1/\sqrt{3\theta}$, we have

$$\begin{aligned}\mathbf{Pr}(|X_3 - \mathbf{E}[X_3]| \geq t\Delta^{-3/2}) &\leq 2\exp\left(\frac{-t^2\epsilon^4n^2\delta^4}{32\Delta^3 m}\right) \\ &\leq 2\exp\left(\frac{-t^2\epsilon^4 n}{32\theta^4}\right) \leq 2\exp\left(\frac{-t^2 n}{288\theta^6}\right).\end{aligned}$$

□

Note that the lemma above shows that there is a constant $B > 0$ such that

$$\mathbf{Pr}(|X'_a - \mathbf{E}[X'_a]| \geq t\Delta^{-3/2}) \leq 6\exp(-tn + Bn).$$

Clearly the same result holds for the second sum X'_b in (33). Thus, we have that for any $\hat{\xi} < \xi \in (0,1)$, there is a $K = K(\theta, \xi) > 0$ such that

$$\begin{aligned}\mathbf{Pr}(|X' - \mathbf{E}[X']| \geq K\Delta^{-3/2} \mid M \text{ is simple}) &\leq 2^{O(d^2)}\hat{\xi}^n \\ &\leq \xi^n. \qquad (41)\end{aligned}$$

Note that we should multiply the RHS of (41) by $\kappa \leq n^2$ to account for the probability there exists M_i for which X' is large.

10.5 Estimating X''

In view of (41), it remains to show that $X'' = O(\Delta^{-3/2})$ with suitably high probability. We shall first prove a preliminary result showing that the random graph G with degree sequence $\mathbf{d}$ is unlikely to have a *dense* subgraph. It will be enough to consider the case $G = G_3$ and argue an immediate implication for its subgraphs.

Lemma 14 *Let G be chosen randomly from the set $\mathcal{G}(\mathbf{d})$ of simple graphs with vertex set $[n]$ and degree sequence $\mathbf{d}$. For $A, B \subseteq [n]$, let $e(A, B)$ be the number of edges joining a vertex in A to a vertex in B and $\mu(A, B) = \theta|A||B|\Delta/n$, where $\theta > \Delta/\delta$ is sufficiently large. For every constant $K > 0$ there is a constant $C = C(\theta, K)$ such that with probability $1 - o(n^{-K})$ every pair of $A, B \subseteq [n]$, with $|A| \leq |B|$, satisfies at least one of the following:*

(I) $e(A, B) \leq C\mu(A, B)$,

(II) $e(A, B) \ln \frac{e(A,B)}{\mu(A,B)} \leq C|B| \ln \frac{n}{|B|}$,

Proof: Write $a = |A|$, $b = |B|$, and let d_A and d_B be the sums of degrees in A and in B respectively. Condition (I) clearly holds *deterministically* if b is at least a constant fraction of n since $e(A, B) \leq a\Delta$. Assume then that $a, b \leq n/(4\theta)$.

We prove later that for any set of possible edges S, $|S| \leq n\Delta/(4\theta) \leq n\delta/2$, we have

$$\mathbf{Pr}(G \text{ contains } S) \leq \left(\frac{\Delta^2}{m}\right)^{|S|}. \tag{42}$$

Thus, the probability that there exists a pair (A, B) with $e(A, B) = t$ is at most

$$\binom{n}{b}\binom{n}{a}\binom{ab}{t}\left(\frac{\Delta^2}{m}\right)^t$$
$$\leq \left(\frac{ne}{b}\right)^{2b}\left(\frac{abe\Delta^2}{mt}\right)^t \leq \left(\frac{ne}{b}\right)^{2b}\left(\frac{\mu(A,B)}{t}\right)^t e^t.$$

Now consider a value x that satisfies

$$x \ln\left(\frac{x}{\mu(A,B)}\right) > Cb\ln\left(\frac{n}{b}\right) \geq \frac{1}{2}Cb\ln\left(\frac{en}{b}\right)$$
$$x > C\mu(A,B)$$
$$x \geq (\ln n)^2$$

Then clearly

$$\mathbf{Pr}(\exists A,B:\ e(A,B)=x)\le n^{-\ln n},$$

and therefore

$$\begin{aligned}&\mathbf{Pr}\left(\exists A,B:\ (e(A,B)\ge(\ln n)^2)\ \&\ \neg(\mathrm{I})\ \&\ \neg(\mathrm{II})\right)\\&=\sum_{(\ln n)^2\le x\le n^2}\mathbf{Pr}\Big(\exists A,B:\ (e(A,B)=x)\ \&\ \neg(\mathrm{I})\ \&\ \neg(\mathrm{II})\Big)\le n^2n^{-\ln n}.\end{aligned}$$

It remains to deal with $\mathbf{Pr}\left(\exists A,B:\ (e(A,B)<(\ln n)^2)\ \&\ \neg(\mathrm{I})\ \&\ \neg(\mathrm{II})\right)$. If $e(A,B)<(\ln n)^2$ and (II) does not hold then

$$2e(A,B)\ln n > Cb\ln\left(\frac{n}{b}\right)\ge Cb\ln(4\theta) \tag{43}$$

and so $b\le e(A,B)\ln n\le(\ln n)^3$, which in turn, from the first inequality in (43), implies that $e(A,B)>Cb/3$. But the probability that $e(A,B)\ge Cb/3$, for C sufficiently large, and $b\le(\ln n)^3$ can be bounded by

$$\begin{aligned}\binom{n}{b}\binom{n}{a}\sum_{t=Cb/3}^{\Delta(\ln n)^3}\binom{ab}{t}\left(\frac{\Delta^2}{m}\right)^t &\le 2\binom{n}{b}\binom{n}{a}\binom{ab}{Cb/3}\left(\frac{\Delta^2}{m}\right)^{Cb/3}\\&\le 2\left(\frac{ne}{b}\right)^{2b}\left(\frac{3ea\Delta^2}{Cm}\right)^{Cb/3}\\&\le 2(n^{2-3C/10}b^{C/3-2})^b.\end{aligned}$$

This yields the conclusion of the lemma.

Proof of (42): Let $S=\{e_1,e_2,\dots,e_s\}$, $\mathcal{G}_0=\mathcal{G}(\mathbf{d})$ and $\mathcal{G}_i=\{G\in\mathcal{G}(\mathbf{d}):\ G$ contains $\{e_1,e_2,\dots,e_i\}$ for $1\le i\le s$. It is sufficient to prove that for $0\le i<s$:

$$\frac{|\mathcal{G}_{i+1}|}{|\mathcal{G}_i|}\le\frac{\Delta^2}{2m-2\Delta^2-2s}\le\frac{\Delta^2}{m}, \tag{44}$$

where the second inequality follows from our bound on s. To prove the first inequality we consider

$$X=\{(H_1,H_2):\ H_1\in\mathcal{G}_i\setminus\mathcal{G}_{i+1},H_2\in\mathcal{G}_{i+1},\ H_1\sim H_2\},$$

where $H_1\sim H_2$ means that there is some 4-cycle with edges $f_1=e_{i+1},f_2,f_3,f_4$ such that H_2 is obtained from H_1 by adding f_1,f_3 and deleting f_2,f_4. The first inequality in (44) follows immediately from the following:

(i) A particular $H_1\in\mathcal{G}_i\setminus\mathcal{G}_{i+1}$ appears in at most Δ^2 pairs of X.

(ii) A particular $H_2 \in \mathcal{G}_{i+1}$ appears in at least $2m - 2\Delta^2 - s$ pairs of X.

Let $e_{i+1} = (x, y)$. For (i) observe that there are at most Δ^2 choices for f_2, f_4 – one is incident with x and one is incident with y. For (ii), given $H_2 \in \mathcal{G}_{i+1}$ we chose an *oriented* edge $f_3 = (u, v) \in H_2$ not incident with e_{i+1}. Let $f_2 = (x, u)$ and $f_4 = (y, v)$. At most $2(\Delta - 1)^2$ choices of f_3 are forbidden because at least one of f_2, f_4 are already in H_1 and at most $s - 1$ choices are disallowed because $f_3 \in S$. □

We now explain why it suffices just to consider G_3 (and G_2) for the large pairs and not their subgraphs $\hat{\Gamma}_j$ (and Γ_j). Indeed, if one of the conditions (I) or (II) hold for G_3 then at least one holds for any of its subgraphs Γ. If condition (I) was true for G_3 then it is true a fortiori for Γ. Similarly, if condition (I) fails, $C \geq 1$, and condition (II) holds for G_3 then it holds a fortiori for Γ.

Lemma 15 *Given the assertions in Lemma 14,* $X'' = \sum_{(u,v) \notin B} x_u A_{uv} y_v$*, where* $B = \{(u, v) \mid 0 < |x_u y_v| < \Delta^{1/2}/n\}$*, satisfies*

$$X'' = O(\Delta^{-3/2})$$

for every *pair* $x, y \in T$.

Proof: Given $x \in T$, we write

$$S_i(x) = \{u : \epsilon^{2-i} n^{-1/2} \leq |x_u| < \epsilon^{1-i} n^{-1/2}\}, \qquad i \in I,$$

where $I = \{i : S_i(x) \neq \emptyset\}$. Define J and $S_j(y)$ analogously. Also, for $S \subseteq [n]$ and $x \in T$, write

$$(x_S)_u = \begin{cases} x_u, & \text{if } u \in S; \\ 0, & \text{otherwise.} \end{cases}$$

Given $x, y \in T$, we write $A_i = S_i(x)$, $B_j = S_j(y)$, $a_i = |A_i|$, $b_j = |B_j|$. Let

$$\mathcal{C} = \{(i, j) \mid i, j > 0,\ \epsilon^{2-i-j} > \sqrt{\Delta},\ a_i \leq b_j\},$$
$$\mathcal{C}' = \{(i, j) \mid i, j > 0,\ \epsilon^{2-i-j} > \sqrt{\Delta},\ a_i > b_j\}.$$

Since

$$X'' = \sum_{|x_u y_v| \geq \Delta^{1/2}/n} x_u A_{uv} y_v,$$

it is sufficient to show that

$$\sum_{(i,j) \in \mathcal{C}} (x_{A_i})^t A y_{B_j} = O(\Delta^{-3/2}),$$

or equivalently, if $e_{i,j} = e(A_i, B_j)$,

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{C}} \frac{e_{i,j}}{\epsilon^{i+j}} = O(\sqrt{\Delta}).$$

The sum on $\mathcal{C}'$ follows by symmetry.

Note that since $\sum x_u^2 \le 1$ for $x \in T$, we have

$$\sum_{i\in I} a_i/\epsilon^{2(i-2)} \le n, \qquad \sum_{j\in J} b_j/\epsilon^{2(i-2)} \le n.$$

Next, partition $\mathcal{C}$ into $\mathcal{C}_I \cup \mathcal{C}_{II}$, where $\mathcal{C}_x$ is the set of $(i,j) \in \mathcal{C}$ such that $e(A_i, B_i)$ satisfies assertion x in Lemma 14. First, using the definition of $\mathcal{C}$, we have

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{C}_I} \frac{e_{i,j}}{\epsilon^{i+j}} = O\left(\frac{1}{n^2} \sum_{(i,j)\in\mathcal{C}_I} \frac{a_i b_j \Delta}{\epsilon^{i+j}}\right) = O\left(\frac{\Delta}{n^2} \sum_{(i,j)\in\mathcal{C}_I} \frac{a_i b_j}{\epsilon^{2(i+j)}\sqrt{\Delta}}\right) = O(\sqrt{\Delta}).$$

It therefore remains to show

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{C}_{II}} \frac{e_{i,j}}{\epsilon^{i+j}} = O(\sqrt{\Delta}). \tag{45}$$

For $k = 1, \ldots, 5$, let $\mathcal{D}_k$ be the set of $(i,j) \in \mathcal{C}_{II}$ satisfying (k) below but not (k') if $k' < k$.

(1) $\epsilon^j \ge \epsilon^i \sqrt{\Delta}$,

(2) $e_{i,j} \le \mu_{i,j}/(\epsilon^{i+j}\sqrt{\Delta})$, where $\mu_{i,j} = \mu(A_i, B_j)$,

(3) $\ln(e_{i,j}/\mu_{i,j}) \ge \frac{1}{4} \ln(n/b_j)$,

(4) $n/b_j \le \epsilon^{-4j}$,

(5) $n/b_j > \epsilon^{-4j}$.

Then equation (45) follows if for $k = 1, \ldots, 5$,

$$H_k = \frac{1}{n} \sum_{(i,j)\in\mathcal{D}_k} \frac{e_{i,j}}{\epsilon^{i+j}} = O(\sqrt{\Delta}).$$

Start by noting that since $(i,j) \in \mathcal{C}_{II}$, we have

$$e_{i,j} \ln(e_{i,j}/\mu_{i,j}) \le C b_j \ln(n/b_j), \tag{46}$$

For $k = 1$, from the trivial inequality $e_{i,j} \leq a_i \Delta$, we have

$$H_1 \leq \frac{1}{n} \sum_i \sum_{j:\epsilon^j \geq \epsilon^i \sqrt{\Delta}} \frac{a_i \Delta}{\epsilon^{i+j}} = O\left(\frac{1}{n} \sum_i \frac{a_i \sqrt{\Delta}}{\epsilon^{2i}}\right) = O(\sqrt{\Delta}).$$

For $k = 2$, we have

$$H_2 \leq \frac{1}{n} \sum_{i,j} \frac{\mu_{i,j}}{\sqrt{\Delta} \epsilon^{2(i+j)}} = O\left(\frac{\sqrt{\Delta}}{n^2} \sum_{i,j} \frac{a_i b_j}{\epsilon^{2(i+j)}}\right) = O(\sqrt{\Delta}).$$

For $k = 3$, equation (46) implies that

$$e_{i,j} = O(b_j),$$

and so using $(i,j) \notin \mathcal{D}_1$, that is $\epsilon^j < \epsilon^i \sqrt{\Delta}$,

$$H_3 = O\left(\frac{1}{n} \sum_j \sum_{i:\epsilon^i > \epsilon^j / \sqrt{\Delta}} \frac{b_j}{\epsilon^{i+j}}\right) = O\left(\frac{1}{n} \sum_j \frac{\sqrt{\Delta} b_j}{\epsilon^{2j}}\right) = O(\sqrt{\Delta}).$$

For $k = 4$, using $(i,j) \notin \mathcal{D}_3$, we have

$$\frac{e_{i,j}}{\mu_{i,j}} \leq \frac{1}{\epsilon^j}.$$

Also, using $(i,j) \notin \mathcal{D}_2$, we have

$$\frac{e_{i,j}}{\mu_{i,j}} \geq \frac{1}{\epsilon^{i+j} \sqrt{\Delta}},$$

thus giving

$$\epsilon^{-i} \leq \sqrt{\Delta}.$$

From (46), we also have $e_{i,j} = O(j b_j)$ (using also $e_{i,j} \geq C \mu_{i,j}$). Thus,

$$H_4 = O\left(\frac{1}{n} \sum_j \sum_{i:\epsilon^{-i} \leq \sqrt{\Delta}} \frac{j b_j}{\epsilon^{i+j}}\right) = O\left(\frac{\sqrt{\Delta}}{n} \sum_j \frac{j b_j}{\epsilon^j}\right).$$

Since $\sum_{j \in J} b_j / (n \epsilon^{2j}) = O(1)$ we have

$$H_4 = O(\sqrt{\Delta}).$$

For $k = 5$, since $b_j < n \epsilon^{4j}$, we have from (46) that

$$e_{i,j} \leq C n \epsilon^{4j} \ln \epsilon^{-4j} = O(n j \epsilon^{4j}).$$

Also, since $(i,j) \notin \mathcal{D}_1$, we have $\epsilon^j < \epsilon^i \sqrt{\Delta}$, thus

$$H_5 = O\left(\sum_j \sum_{i:\epsilon^i > \epsilon^j / \sqrt{\Delta}} j \epsilon^{3j-i}\right) = O\left(\sqrt{\Delta} \sum_j j \epsilon^{2j}\right) = O(\sqrt{\Delta}).$$

□

Observe finally that for future reference we have in fact proven the following

Lemma 16 *Let* $\mathbf{d} = d_1, d_2, \ldots, d_n$ *be a degree sequence with maximum degree* $\Delta = o(n^{1/2})$ *and minimum degree* δ *such that* $\Delta/\delta < \theta$ *for some constant* $\theta > 0$. *Let* G *be chosen randomly from the set of* simple *graphs with degree sequence* $\mathbf{d}$. *Let* $0 < c < 1$ *be an arbitrary constant and* $\mathcal{G}$ *be the set of vertex induced subgraphs* H *of* G *which have degree at least* $c\delta$. *Let* $K > 0$ *be an arbitrary constant. Then with probability* $1 - O(n^{-K})$ every *graph* H *in* $\mathcal{G}$ *has second eigenvalue at most* $\gamma/\sqrt{\Delta}$ *where* $\gamma = \gamma(\theta, c, K)$.

Proof: We can handle "small pairs" by using multigraphs and pass to simple graphs as above. We observe that only the failure probability (41) now needs to be inflated by $2^{n+O(d^2)}$ and this is handled by making ξ small enough or γ large enough. The case of "large" pairs is handled as before by deducing it from what happens in G. $\square$

There are no lower bounds explicitly stated for δ, but our results are not useful for small minimum degree. It follows from (40) that γ is at least $4\theta^9 \exp\{192\theta^9\}$. Thus say for $\delta \leq 10^6$ we will have $\gamma \geq \Delta$ and so the estimate for the second eigenvalue will exceed one, the largest eigenvalue.

Acknowledgment: It is a pleasure to acknowledge the work of the anonymous reviewers. In particular, we would like to thank the referee who pointed out a small gap in the originally submitted version of the paper.

References

[ASE] N. Alon, J. Spencer and P. Erdős, *The probabilistic method*, John Wiley and Sons, 1992.

[BC] E. A. Bender and E. R. Canfield, *The asymptotic number of labelled graphs with given degree sequences*, Journal of Combinatorial Theory (A) 24, (1978) 296-307.

[B] B. Bollobás, *Random graphs*, Academic press, 1985.

[B2] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European Journal on Combinatorics 1, (1980) 311-316.

[B3] B. Bollobás, *Martingales, isoperimetric inequalities and random graphs*, Coll. Math. Soc. J. Bolyai **52**, 113-139 (1987).

[BFF] B. Bollobás, T. I. Fenner, and A. M. Frieze, *Hamilton cycles in random graphs with minimal degree at least k*, in "A tribute to Paul Erdős," edited by A. Baker, B. Bollobás, and A. Hajnal (1990) 59-96.

[BM] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, North-Holland 1976.

[BFU] A. Z. Broder, A. M. Frieze, and E. Upfal, *Existence and construction of edge disjoint paths on expander graphs*, Proceedings of the 24'th Annual ACM Symposium on Theory of Computing, (1992) 140-149.

[BFSU] A. Z. Broder, A. M. Frieze, S. Suen, and E. Upfal, *Optimal construction of vertex-disjoint paths in random graphs*, in preparation.

[G] D. Gale, *A theorem on flows in networks*, Pacific Journal of Mathematics 7 (1957) 1073-1082.

[EL] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In A. Hajnal et al., editors, *Infinite and Finite Sets*, volume 11 of *Colloq. Math. Soc. J. Bolyai*, pages 609–627. North Holland, 1975.

[H] D. S. Hochbaum. *An exact sublinear algorithm for the max flow, vertex-disjoint paths, and communication problems on random graphs*, Operations Research 40 (1992) 923–935.

[KS] J. Kahn and E. Szemerédi, *On the second eigenvalue in random regular graphs - Section 2*, Proceedings of the 21'st Annual ACM Symposium on Theory of Computing (1989) 587-598.

[KT] J. Kleinberg and E. Tardos, *Approximations for the disjoint paths problem in high diameter planar networks*, Proceedings of the 27'th Annual ACM Symposium on Theory of Computing, (1995) 26-35.

[M] C. J. H. McDiarmid, *On the method of bounded differences*, in Surveys in Combinatorics, Proceedings of the Twelfth British Combinatorial Conference (J. Siemons, Ed.), 1989, pp. 148-188.

[PU] D. Peleg and E. Upfal, *Constructing disjoint paths on expander graphs*, Combinatorica 9, (1989) 289-313.

[RS] N. Robertson and P. D. Seymour, *Graph minors-XIII: The disjoint paths problem*, to appear.

[SU] E. Shamir and E. Upfal. *A fast construction of disjoint paths in networks.* Annals of Discrete Mathematics 24 (1985) 141–154.

[SJ] A. Sinclair and M. Jerrum, *Approximate counting, uniform generation, and rapidly mixing Markov chains*, Information and Computation 82 (1989) 93-133.

[VB] L. G. Valiant and G. J. Brebner, *Universal schemes for parallel computation*, Proceedings of the 13'th Annual ACM Symposium on Theory of Computing (1981) 263-277.