Elegantly colored paths and cycles in edge colored random graphs

Lisa Espig∗ Alan Frieze † Michael Krivelevich‡
March 6, 2014

Abstract

We first consider the following problem. We are given a fixed perfect matching $M$ of $[n]$ and we add random edges one at a time until there is a Hamilton cycle containing $M$. We show that w.h.p. the hitting time for this event is the same as that for the first time there are no isolated vertices in the graph induced by the random edges. We then use this result for the following problem. We generate random edges and randomly color them black or white. A path/cycle is said to be zebraic if the colors alternate along the path. We show that w.h.p. the hitting time for a zebraic Hamilton cycle coincides with every vertex meeting at least one edge of each color.

We then consider some related problems and extend to multiple colors.

1 Introduction

This paper studies the existence of nicely structured objects in (randomly) colored random graphs. Our basic interest will be in what we call *zebraic* paths and cycles. We assume that the edges of a graph $G$ have been colored black or white. A path or cycle will be called *zebraic* if the edges alternate in color along the path. We view this as a variation on the usual theme of *rainbow* paths and cycles that have been well-studied already – see for example Erdős, Nešetřil and Rödl [5], Albert, Frieze and Reed [1], Cooper and Frieze [3] and Frieze and Loh [7] for the existence of rainbow Hamilton cycles in complete or randomly colored graphs.

Our first result does not at first sight fit into this framework. Let $n$ be even and let $M_0$ be an arbitrary perfect matching of the complete graph $K_n$. Now consider the random graph process $G_m = ([n], E_m)$ where $E_m = \{e_1, e_2, \ldots, e_m\}$ is obtained from $E_{m-1}$ by adding a random edge $e_m \notin E_{m-1}$, for $m = 0, 1, \ldots, N = \binom{n}{2}$.

Let

$$\tau_1 = \min \{m : \delta(G_m) \geq 1\}$$

where $\delta$ denotes minimum degree. Then let

$$\tau_H = \min \{m : G_m \text{ contains a Hamilton cycle } H \supseteq M_0\}.$$ 

**Theorem 1** \( \tau_1 = \tau_H \) w.h.p.¹

¹A sequence of events $E_n$ is said to occur *with high probability* (w.h.p.) if $\lim_{n \to \infty} \Pr(E_n) = 1$.
In actual fact there are two slightly different versions. One where we insist that \( M_0 \cap E_m = \emptyset \) and one where \( E_m \) is chosen completely independently of \( M_0 \). The theorem holds in both cases.

We note that Robinson and Wormald [10] considered a similar problem with respect to random regular graphs. They showed that one can choose \( o(n^{1/2}) \) edges at random, orient them and then w.h.p. there will be a Hamilton cycle containing these edges and following the orientations.

Theorem 1 has an easy corollary that fits our initial description. Let \( G(r)_m \) be an \( r \)-colored version of the graph process. This means that \( G(r)_{m-1} \) is obtained from \( G(r)_m \) by adding a random edge and then giving it a random color from \([r]\). Let \( E_{m,i} \) denote the edges of color \( i \) for \( i = 1, 2, \ldots, r \). When \( r = 2 \) denote the colors by \textit{black} and \textit{white} and let \( E_{m,b} = E_{m,1}, E_{m,w} = E_{m,2} \). Then let \( G_m(b) \) be the subgraph of \( G_m^{(2)} \) induced by the black edges and let \( G_m(w) \) induced by the white edges. Let

\[
\tau_{1,1} = \min \left\{ m : \delta(G_m(b)), \delta(G_m(w)) \geq 1 \right\}
\]

and let

\[
\tau_{ZH} = \min \{ m : G_m \text{ contains a zebraic Hamilton cycle.} \}.
\]

Corollary 1 \( \tau_{1,1} = \tau_{ZH} \) w.h.p.

Our next result is a zebraic analogue of rainbow connection. For a connected graph \( G \), its rainbow connection \( rc(G) \), is the minimum number \( r \) of colors needed for the following to hold: The edges of \( G \) can be \( r \)-colored so that every pair of vertices is connected by a rainbow path, i.e. a path in which no color is repeated. Recently, there has been interest in estimating this parameter for various classes of graph, including random graphs. By analogy, we say that a two-coloring of a connected graph provides a zebraic connection if there is a zebraic path joining every pair of vertices.

**Theorem 2** At time \( \tau_1 \), a random black-white coloring of \( G_{\tau_1} \) provides a zebraic connection, w.h.p.

We consider now how we can extend our results to more than two colors. Suppose we have \( r \) colors \([r]\) and that \( r \mid n \). We would like to consider the existence of Hamilton cycles where the \( i \)th edge has color \((i \mod r) + 1\). Call such a cycle \( r \)-zebraic. Our result for this case is not as tight as for the case of two colors. We are not able to prove a hitting time version. We will instead satisfy ourselves with a result for \( G(n,r) \).

**Theorem 3** Let \( \varepsilon > 0 \) be an arbitrary positive constant.

\[
\lim_{n \to \infty} \Pr(G(n,r) \text{ contains an } r \text{-zebraic Hamilton cycle}) = \begin{cases} 
0 & p \leq (1 - \varepsilon) p_r \\
1 & p \geq (1 + \varepsilon) p_r.
\end{cases}
\]
2 Notation and structure of the paper

2.1 Notation

For a graph $G = (V, E)$ and $S, T \subseteq V$ we let $e_G(S)$ denote the number of edges contained in $S$, $e_G(S, T)$ denote the number of edges with one end in $S$ and the other in $T$ and $N_G(S)$ denote the set of neighbors of $S$ that are not in $S$.

We will use certain values throughout our proofs. We list most of them here for easy reference: Let 
\[
t_0 = \frac{n}{2} (\log n - 2 \log \log n) \quad \text{and} \quad t_1 = \frac{n}{2} (\log n + 2 \log \log n) \quad \text{and} \quad p_i = \frac{t_i}{\binom{n}{2}}, \quad i = 0, 1.
\] (1)

Let 
\[
n_0 = \frac{n}{\log^2 n} \quad \text{and} \quad n_0' = \frac{n_0}{20 \log n} \quad \text{and} \quad n_1 = \frac{n}{10 \log n}.
\]

Let 
\[
n_b = \frac{n \log \log \log n}{\log \log n} \quad \text{and} \quad n_c = \frac{200n}{\log n}.
\]

Let 
\[
L_0 = \frac{\log n}{100} \quad \text{and} \quad L_1 = \frac{\log n}{\log \log n}.
\]

Let 
\[
\ell_0 = \frac{\log n}{200} \quad \text{and} \quad \ell_1 = \frac{2 \log n}{3 \log \log n} \quad \text{and} \quad \nu_L = \ell_0^{\ell_1} = n^{2/3 + o(1)}.
\]

Let 
\[
\nu_L = (n \log n)^{1/2}.
\]

2.2 Structure of the paper

3 Proof of Theorem 1

It is well known (see for example [2], [9]) that w.h.p. we have $t_0 \leq \tau_1 \leq t_1$.

Our strategy for proving this is to modify the 3-phase algorithm described in [4] to fit the current situation.

(a) We will use $\sim t_0/10$ edges to build a perfect matching $M_1$ that is independent of $M_0$. The union of $M_0, M_1$ will have $O(\log n)$ components w.h.p.

(b) $M_0 \cup M_1$ induces a 2-factor made up of alternating cycles. We use about $\sim 4t_0/5$ edges to make the minimum cycle size $\Omega(n/\log n)$.

(c) We use the final $\sim t_0/10$ edges to create a Hamilton cycle containing $M_0$.

3.1 Building $M_1$

In order to prove Corollary 1, we will need to re-define w.h.p. to mean “with probability $1 - o(n^{-0.51})$”. This remains in force until Section 3.3.

We begin with $\Psi_0 = G_{t_2}$ where $t_2 = t_0/10$. Then let $V_0$ denote the set of vertices that have degree at most $L_0$ in $\Psi_0$. Now create $\Psi_1 = ([n], E_1)$ by adding those edges in $E_{t_1} \setminus E_{t_2}$ that are incident with $V_0$. 

Let a vertex be *large* if its degree in $G_t$ is at least $L_0$ and *small* otherwise. Let $V_\lambda$ denote the set of large vertices and let $V_\sigma$ denote the set of small vertices.

The calculations for the next lemma will simplify if we observe the following: Suppose that $m = Np$. It is known that for any monotone increasing property of graphs

$$\Pr(G_m \in \mathcal{P}) \leq 3\Pr(G_{n,p} \in \mathcal{P}).$$

(2)

In general we have for not necessarily monotone properties:

$$\Pr(G_m \in \mathcal{P}) \leq 3m^{1/2}\Pr(G_{n,p} \in \mathcal{P}).$$

(3)

For proofs of (2), (3) see Bollobás [2] or Janson, Luczak and Ruciński [9].

**Lemma 2** The following holds w.h.p.:

(a) $|V_0| \leq n^{11/12}$.

(b) If $x, y \in V_\sigma$ then the distance between them in $G_t$ is at least 10.

(c) If $S \subseteq [n]$ and $|S| \leq n_0$ then $e_{G_{t_1}}(S) \leq 10|S|$.

(d) If $S \subseteq [n]$ and $|S| = s \in [n_0', n_1]$ then $|N_{\Psi_0}(S)| \geq s \log n/25$.

(e) No cycle of length 4 in $G_{t_1}$ contains a small vertex.

(f) The maximum degree in $G_{t_1}$ is less than $10 \log n$.

**Proof** (a) Suppose that the sequence $x_1, x_2, \ldots, x_{2t_2}$ is chosen randomly from $[n]^{2t_2}$ and we let $\Gamma_{t_2}$ denote the multigraph with edge-set $(x_{2i-1}, x_{2i})$, $i = 1, 2, \ldots, t_2$. After we remove repeated edges and loops we can couple what remains with a subgraph $H$ of $G_{t_2}$. Let $Z_1$ denote the number of loops and let $Z_2$ denote the number of repeated edges in $\Gamma_{t_2}$. Let $V_0'$ denote the set of vertices of degree at most $L_0$ in $\Gamma_{t_2}$. Then $|V_0| \leq Z_1 + 2Z_2 + |V_0'|$. This is because if $v \in V_0 \setminus V_0'$ then it must lie in a loop or a multiple edge.

Now $Z_1$ is distributed as $\text{Bin}(t_2, 1/n)$ and then the Chernoff bounds imply that

$$Z_1 \leq \log^2 n \text{ w.h.p.}$$

(4)

We are doing more than usual here, because we need probability $o(n^{-0.51})$, rather than just probability $o(1)$.

Now $Z_2$ is dominated by $\text{Bin}(t_2, t_2/N)$ and then the Chernoff bounds imply that

$$Z_2 \leq \log^3 n \text{ w.h.p.}$$

(5)

Now,

$$\Pr(v \in V_0') \leq \sum_{k=0}^{L_0} \binom{2t_2}{k} n^{-k} \left(1 - \frac{1}{n}\right)^{2t_2-k} \leq 2 \left(\frac{2t_2}{L_0}\right) n^{-L_0} e^{-2t_2-L_0}/n \leq 2 \left(\frac{2eL_2}{nL_0}\right) L_0^{-1/10+o(1)} n^{-1/11}. $$
It follows, that $E(|V'_0|) \leq n^{10/11}$. We now use a concentration inequality to finish the proof. Indeed, changing one of the $x_i$’s can change $|V'_0|$ by at most one. Hence, for any $u > 0$,

$$\Pr(|V'_0| \geq E(|V'_0|) + u) \leq \exp\left\{-\frac{2u^2}{t_2}\right\}.$$ 

Putting $u = n^{4/7}$ into the above and using (4), (5) finishes the proof of (a).

(b) We do not have room to apply (3) here. We need the inequality

$$\left(\frac{N-a}{t-b}\right)^t \leq \left(\frac{t}{N}\right)^b \left(\frac{N-t}{N-b}\right)^{a-b}$$

(6)

for $b \leq a \leq t \leq N$. We will now and again use the notation $A \leq B$ in place of $A = O(B)$ when it suits our aesthetic taste.

$$\Pr(\exists x, y) \leq \sum_{k=2}^{11} \binom{n}{k}! \sum_{\ell_1, \ell_2=0}^{L} \binom{n-k}{\ell_1} \binom{n-k}{\ell_2} \frac{(N-(2n-k-1))}{(N-t_1-k+1-\ell_1-\ell_2)} \leq b \sum_{k=2}^{n} \binom{n}{k} \sum_{\ell_1, \ell_2=0}^{L} \left(\frac{ne}{\ell_1}\right)^{\ell_1} \left(\frac{ne}{\ell_2}\right)^{\ell_2} \frac{(t_1+k)}{N} \frac{(N-t_1)}{N-(\ell_1+\ell_2+k-1)} \leq b \log^{k-1} n \sum_{\ell_1, \ell_2=0}^{L} \left(\frac{3\log n}{\ell_1}\right)^{\ell_1} \left(\frac{3\log n}{\ell_2}\right)^{\ell_2} n^{-2+o(1)} = o(n^{-0.51}).$$

(c) We can use (2) here. If $s = |S|$, then in $G_{n,p_1}$ where $p_1 = t_1/N$,

$$\Pr(e_G(S) > 10|S|) \leq \left(\frac{e}{10s}\right)^{10s} \leq \left(\frac{2}{10s}\right)^{10s} \leq \left(\frac{2}{10s}\right)^{10s} e^{s \log n + \log \log n} = o(n^{-0.51}).$$

So,

$$\Pr(\exists S) \leq \sum_{s=10}^{n_1} \binom{n}{s} \frac{s \log n}{n}^{10s} \leq \sum_{s=10}^{n_1} \left(\frac{ne}{s}\right)^{s} \frac{s \log n}{n}^{10s} = \sum_{s=10}^{n_1} \left(\frac{s}{n}\right)^{9} \log^{10} n = o(n^{-0.51}).$$

(d) We can use (2) here with $p_2 = t_2/N$. For $v \in V$, $\Pr(v \in N(S)) = 1 - (1 - p_2)^s \geq \frac{sp_2}{2}$ for $s \leq n_1$. So $|N(S)|$ stochastically dominates $\text{Bin}(n - s, \frac{sp_2}{2})$. Now $(n - s) \frac{sp_2}{2} \sim \frac{s \log n}{20}$ and so using the Chernoff bound,

$$\Pr(|N_{\Psi_0}(S)| < s \log n/25) \leq e^{-s \log n/1250}.$$

So,

$$\Pr(\exists S) \leq \sum_{s=n_0}^{n_1} \binom{n}{s} \frac{e^{-s \log n/1250}}{s} \leq \sum_{s=n_0}^{n_1} \left(\frac{ne}{s} \cdot n^{-1/1250}\right)^s = o(n^{-0.51}).$$

(e) The expected number of such cycles is bounded by

$$\binom{n}{4} \sum_{k=0}^{L_0} \binom{L_0}{4-k} \binom{N-n-3}{k} \frac{(N-t_1-4-k)}{N} \leq n^4 \sum_{k=0}^{L_0} \left(\frac{ne}{k}\right)^k \left(\frac{t_1+k+4}{N}\right)^{k+4} \left(\frac{N-t_1}{N-k-4}\right)^{n+k-1} \leq b \log^4 n \sum_{k=0}^{L_0} \left(\frac{e^{1+o(1)} \log n}{k}\right)^k n^{-1+o(1)} = o(n^{-0.51}).$$
(f) We apply (2) and find that the probability of having a vertex of degree exceeding $10 \log n$ is at most
\[
3n \left( \frac{n - 1}{10 \log n} \right) \left( \frac{\log n + \log \log n}{n} \right)^{10 \log n} \leq 3n \left( \frac{e^{1+o(1)}}{10} \right)^{10 \log n} = o(n^{-0.51}).
\]

We will sometimes use (f) without comment in what follows.

Lemma 2 implies the following:

**Lemma 3** W.h.p.

\[ S \subseteq [n] \text{ and } |S| \leq n/2000 \text{ implies } |N_{\Psi_1}(S)| \geq |S| \text{ in } \Psi_1, \tag{7} \]

**Proof** Assume that the conditions described in Lemma 2 hold. Let $N(S)$ refer to $\Psi_1$. We first argue that if $S \subseteq V_\lambda$ and $|S| \leq n/2000$ then
\[
|N(S)| \geq 4|S|. \tag{8}
\]

From the lemma, we only have to concern ourselves with $|S| \leq n'_0$ or $|S| \in [n_1, n/2000]$. If $|S| \leq n'_0$ and $T = N(S)$ then in $\Psi_1$ we have
\[
e(S \cup T) \geq \frac{|S| \log n}{200} \text{ and } |S \cup T| \leq n_0. \tag{9}
\]

It is important to note that to obtain (9) we use the fact that vertices in $V_0 \setminus V_\sigma$ are given all their edges in $\Psi_1$.

Equation (9) implies that $\frac{|S| \log n}{200} \leq 10 |S \cup T|$ and so (8) holds with room to spare.

If $|S| \in [n_1, n/2000]$ then we choose $S' \subseteq S$ where $|S'| = n_1$ and use
\[
|N(S)| \geq |N(S')| - |S| \geq \frac{\log n}{25} \cdot \frac{200|S|}{\log n} - |S|.
\]

This yields (8), again with room to spare.

Now let $S_0 = S \cap V_\sigma$ and $S_1 = S \setminus S_0$. Then we have
\[
|N(S)| \geq |N(S_0)| + |N(S_1)| - |N(S_0) \cap S_1| - |N(S_1) \cap S_0| - |N(S_0) \cap N(S_1)|. \tag{10}
\]

But $|N(S_0)| \geq |S_0|$. This follows from (i) $\Psi_1$ has no isolated vertices, and (ii) Lemma 2(b) means that $S_0$ is an independent set and no two vertices in $S_0$ have a common neighbor. Equation (8) implies that $|N(S_1)| \geq 4|S_1|$. We next observe that trivially, $|N(S_0) \cap S_1| \leq |S_1|$. Then we have $|N(S_1) \cap S_0| \leq |S_1|$, for otherwise some vertex in $S_1$ has two neighbors in $S_0$, contradicting Lemma 2(b.) Finally, we also have $|N(S_0) \cap N(S_1)| \leq |S_1|$. If for a vertex in $S_1$ there are two distinct paths of length two to $S_0$ then we violate one of the conditions (b), (e) Lemma 2.

So, from (10) we have
\[
|N(S)| \geq |S_0| + 4|S_1| - |S_1| - |S_1| - |S_1| = |S|.
\]

Next let $G = (V, E)$ be a graph with an even number of vertices that does not contain a perfect matching. Let $v$ be a vertex not covered by some maximum matching and let
\[
A_G(v) = \{ w : \exists \text{ a maximum matching of } G \text{ that does not cover both } v \text{ and } w. \}.
\]
Lemma 4 If $A = A_G(v)$ for some $v, G$, then $|N_G(A)| < |A|$.

Proof This follows from the Edmonds-Gallai theorem. Let $H = G - \{v\}$. Then there is a partition of $V' = V \setminus \{v\}$ into $A, B, C$ such that $A$ is the set of vertices not contained in some maximum matching of $H$ and $B = N(A)$. Every odd component of $H - B$ is contained in $A$ and there are more than $|B|$ odd components in $H - B$. $\square$

Now consider the edge set

$$E_A = E_{t_0/5} \setminus E(\Psi_1) = \{f_1, f_2, \ldots, f_\rho\}.$$ 

Lemma 5 Given $\rho$, $E_A$ is a random $\rho$-subset of $\binom{W}{2}$, where $W = \lfloor n \rfloor \setminus V_0$.

Proof This follows from the fact that if we remove any $f_i$ and replace it with any other edge from $\binom{\lfloor n \rfloor \setminus V_0}{2}$ then $V_0$ is unaffected. $\square$

Now consider the sequence of graphs $H_0 = \Psi_1, H_1, \ldots, H_\rho$ where $H_i$ is obtained from $H_{i-1}$ by adding the edge $f_i$. We claim that if $\mu_i$ denotes the size of a largest matching in $H_i$, then

$$\Pr(\mu_i \geq \mu_{i-1} + 1 \mid \mu_{i-1} < n/2, f_1, \ldots, f_{i-1}, (\Psi_1 \text{ satisfies (7)})) \geq 10^{-7}. \tag{11}$$

To see this, let $M_{i-1}$ be a matching of size $\mu_{i-1}$ in $H_{i-1}$ and suppose that $v$ is a vertex not covered by $M_{i-1}$. It follows from (7) and Lemma 4 that if $A_{H_{i-1}}(v) = \{g_1, g_2, \ldots, g_r\}$ then $r \geq n/2000$. Now consider the pairs $\{g_j, x\}$, $j = 1, \ldots, r$, $x \in A_{H_{i-1}}(g_i)$. There are at least $\binom{n/2000}{2}$ such pairs and if $f_i$ lies in this collection, then $\mu_i = \mu_{i-1} + 1$. Equation (11) follows from this and the fact that $E_A$ is a random set. In fact, given the condition in Lemma 2(a) and a maximum degree of $\leq 10 \log n$ in $G_{t_1}$, the probability in question is at least

$$\binom{n/2000}{2} - 10n^{11/12} \log n - \rho > 10^{-7}.$$ 

It follows from (11) that

$$\Pr(H_\mu \text{ has no perfect matching}) \leq o(n^{-0.51}) + \Pr\left(\binom{t_0}{5} - \binom{t_0}{10} - 10n^{11/12} \log n, 10^{-7} \right) \leq n/2) = o(n^{-0.51}). \tag{12}$$

So w.h.p. $\Psi_2 = H_\rho$ has a perfect matching $M_1$.

Remark 6 $M_1$ is uniformly random w.r.t. $M_0$ and so the inclusion-exclusion formula gives

$$\Pr(M_0 \cap M_1 = \emptyset) = \sum_{i=0}^{n/2} (-1)^i \binom{n/2}{i} \frac{(n - 2i)!}{(n/2 - i)!2^{n/2-i}} \frac{2^{n/2}(n/2)!}{n!}. \tag{13}$$

Here we use the fact that there are $(2m)!/(m!2^m)$ perfect matchings in $K_m$.

Now if $u_i$ denotes the summand in (13) then we have $u_0 = 1$ and

$$\frac{|u_{i+1}|}{|u_i|} = \frac{1}{2(i+1)} \left(1 + \frac{1}{n - 2i}\right).$$
So if \( m = \Theta(\log n) \) say then

\[
\Pr(M_0 \cap M_1 = \emptyset) \geq \sum_{i=0}^{2m+1} u_i
\]

\[
= \sum_{i=0}^{2m+1} (-1)^i \frac{1}{2^i i!} + O\left(\frac{\log n}{n}\right)
\]

\[
= e^{-1/2} + o(1).
\]

It follows that \( M_1 \) exists w.h.p. even if we insist that it be disjoint from \( M_0 \). Indeed, conditioning on \( M_0 \cap M_1 = \emptyset \) can only increase the probability of some “unlikely” event by a factor of at most \( e^{1/2} + o(1) \).

We will need the following properties of the 2-factor \( \Pi_0 = M_0 \cup M_1 \).

**Lemma 7** The following hold w.h.p.:

(a) \( M_0 \cup M_1 \) has at most \( 10 \log_2 n \) components.

(b) There are at most \( n_b \) vertices on components of size at most \( n_c \).

**Proof**

(a) The proof of Lemma 3 of [8] shows the following. If \( C \) is the cycle of \( M_0 \cup M_1 \) that contains vertex 1 then

\[
\Pr(|C| = 2k) < \frac{1}{n - 2k + 1}.
\]

Having chosen \( C \), the remaining cycles come from the union of two (random) matchings on the complete graph \( K_{n-|C|} \). It follows from this, by summing over \( k \leq n/4 \) that \( \Pr(|C| < n/2) \leq 1/2 \). Hence,

\[
\Pr(\neg(a)) \leq \Pr(Bin(10 \log_2 n, 1/2) \leq \log_2 n) = 2^{-10 \log_2 n} \sum_{i=0}^{\log_2 n} \frac{10 \log_2 n}{i} \leq 2^{-5 \log_2 n} = o(n^{-0.51}).
\]

(b) It follows from (14) that

\[
\Pr(|C| \leq n_c) \leq \frac{201}{\log n}.
\]

If we generate cycle sizes as in (a) then up until there are fewer than \( n_b/2 \) vertices left, \( \log \nu \sim \log n \) where \( \nu \) is the number of vertices that need to be partitioned into cycles. It follows that the probability we generate more than \( k = \frac{\log \log \log n \times \log n}{1000 \log \log n} \) cycles of size at most \( n_c \) up to this time is bounded by

\[
\Pr\left(Bin\left(10 \log_2 n, \frac{201}{\log n}\right) \geq k\right) \leq \left(\frac{3000e}{k}\right)^k = o(n^{-0.51}).
\]

Thus w.h.p. we have at most

\[
\frac{n_b}{2} + kn_c \leq n_b
\]

vertices on cycles of length at most \( n_b \).
3.2 Increasing cycle size

In this section, we will use the edges in

$$E_B = \{ e \in E_{9t_0/10} \setminus E_{t_0/5} : e \cap V_0 = \emptyset \}$$

to create a 2-factor that contains $M_0$ and in which each cycle has size at least $n_c$.

Note that

$$E_B \cap E(\Psi_1) = \emptyset.$$ 

Let

$$V_\tau = \{ v \in [n] \setminus V_0 : \deg_{E_B}(v) \leq L_0 \} ,$$

where for a set of edges $X$ and a vertex $x$, $\deg_X(x)$ is the number of edges in $X$ that are incident with $x$.

**Lemma 8** The following hold w.h.p.

(a) $|V_\tau| \leq n^{2/5}$.

(b) No vertex has 10 or more $G_{t_1}$ neighbors in $V_\tau$.

(c) If $C$ is a cycle with $|C| \leq n_c$ then $|C \cap V_\tau| \leq |C|/200$ in $G_{t_1}$.

**Proof**

(a) We follow a similar argument to that in Lemma 2(a). We condition on $|V_0| \leq n^{11/12}$ and maximum degree 10 log $n$ in $G_{t_0}$ and generate a random sequence from $[n-n^{11/12}]^{7t_0/10-10n^{11/12} \log n}$. The argument is now almost identical to that in Lemma 2(a).

(b) This time we can condition on $\nu = n - |V_0|$ and $\mu = |\{ e \in E_{9t_0/10} \setminus E_{t_0/5} : e \cap V_0 \neq \emptyset \}|$. We write

$$\Pr(v \text{ violates (b)}) \leq \sum_{S \in \begin{pmatrix} n-1 \end{pmatrix}} \Pr(A(v,S)) \Pr(B(v,S) \mid A(v,S))$$

where

$$A(v,S) = \{ N(v) \supseteq S, \text{ in } G_{t_1} \} ,$$

$$B(v,S) = \{ w \text{ has at most } L_0 \text{ } E_B-\text{neighbors in } [n] \setminus (S \cup \{v\}), \forall w \in S \} .$$

Applying (2) we see that $\Pr(A(v,S)) \leq 3 \left( \binom{n}{10} p_1^{10} \right)$ and then using (2) with $p = \frac{7t_0/10 - \mu}{\binom{n}{2}} \sim \frac{7 \log n}{10m}$ we see that

$$\Pr(B(v,S) \mid A(v,S)) \leq 3 \left( \sum_{k=0}^{L_0} \binom{\nu - 11}{k} p^k (1-p)^{\nu-11} \right)^{10}$$

and so

$$\Pr(v \text{ violates (b)}) \leq b \left( \binom{n}{10} p_1^{10} \left( \sum_{k=0}^{L_0} \binom{\nu - 11}{k} p^k (1-p)^{\nu-11} \right)^{10} \right)$$

$$\leq (e^{o(1)} \log n \cdot n^{-7/10+o(1)})^{10} = o(n^{-6}).$$
(c) Using (2) we have

\[
E(\# \text{ of such cycles } C) \leq \sum_{k=3}^{n_c} \binom{n}{k} k! p_1^k \left( \sum_{\ell = 0}^{L_0} \binom{\nu - k}{\ell} p^\ell (1 - p)^{\nu - \ell} \right)^{\lfloor k/200 \rfloor} \leq \sum_{k=3}^{n_c} (2n)^k \left( \frac{\log n + \log \log n}{n} \right)^k \frac{n^{-3(k/200)}}{5} = o(n^{-0.51}).
\]

Now consider the distribution of the edges in \(E_B\).

**Lemma 9** Let \(V_1 = [n] \setminus V_0\) and \(A \subseteq \binom{V_1}{2}\) with \(|A| = a = O(\log n)\). Let \(X\) be a subset of \(E_B\) that is disjoint from \(A\). Suppose that \(|X| = O(\nu^{11/12} \log n)\). Then

\[
\Pr(E_B \supseteq A \mid |E_B| = \mu = \alpha n \log n, |V_1| = \nu \geq n - n^{11/12}, E_B \supseteq X) = (1 + o(n^{-1/10})) \left( \frac{2 \alpha \log n}{n} \right)^a.
\]

**Proof** Equation (15) follows from Lemma 5. For equation (16), use the fact that in general, if \(s^2 = o(N)\) then

\[
\frac{\binom{N-s}{M-s}}{\binom{N}{M}} = \left( \frac{M}{N} \right)^s \left( 1 + O \left( \frac{s^2}{N} \right) \right).
\]

By construction, we can apply this lemma to the graph induced by \(E_B\) with

\[\alpha = \frac{7 + o(1)}{20}.\]

Let a cycle \(C\) of \(\Pi_0\) be small if its length \(|C| < n_c\) and large otherwise. Define a near 2-factor to be a graph that is obtained from a 2-factor by removing one edge. A near 2-factor \(\Gamma\) consists of a path \(P(\Gamma)\) and a collection of vertex disjoint cycles. A 2-factor or a near 2-factor are proper if they contain \(M_0\). We abbreviate proper near 2-factor to PN2F.

We will describe a process for eliminating small cycles from \(\Psi_0\). In this process we create intermediate proper 2-factors. Let \(\Gamma_0\) 2-factor and suppose that it contains a small cycle \(C\). To begin the elimiantion of \(C\) we choose an arbitrary edge \((u_0, v_0)\) in \(C \setminus M_0\), where \(u_0, v_0 \notin V_0\). This is always possible, see Lemma 8(c). We delete it, obtaining a PN2F \(\Gamma_1\). Here, \(P(\Gamma_1) \in \mathcal{P}(v_0, u_0)\), the set of \(M_0\)-alternating paths in \(G\) from \(v_0\) to \(u_0\). Here an \(M_0\)-alternating path must begin and end with an edge of \(M_0\). The initial goal will be to create a large set of PN2Fs such that each \(\Gamma\) in this set has path \(P(\Gamma)\) of length at least \(n_0\) and the small cycles of \(\Gamma\) are a strict subset of the small cycles.
of $\Gamma_0$. Then we’ll show that w.h.p. the endpoints of one of the paths in some such $\Gamma$ can be joined by an edge to create a proper 2-factor with at least one fewer small cycle than $\Pi$.

This process can be divided into two stages. In a generic step of Stage I, we take a PN2F $\Gamma$ as above with $P(\Gamma) \in \mathcal{P}(u_0, v)$ and construct a new PN2F with the same starting point $u_0$ for its path. We do this by considering edges from $E_B$ incident to $v$. Suppose $vw \in E_B$ and that the non-$M_0$ edge in $\Gamma$ containing vertex $w$ is $(w, x)$. Then $\Gamma' = \Gamma \cup (v, w) \setminus (w, x)$ is a PN2F with $P(\Gamma') \in \mathcal{P}(u_0, x)$. We say that $(v, w)$ is acceptable if $x, w \notin W$ ($W$ defined immediately below) and $P(\Gamma')$ has length at least $n_c$ and any new cycle created (in $\Gamma'$ but not $\Gamma$) has at least $n_c$ edges.

There is an unlikely technicality to be faced. If $\Gamma$ has no non-$M_0$ edge $(x, w)$, then $w = u_0$ and this is accepted if $P(\Gamma')$ has at least $n_c$ edges. This would prematurely end an iteration. The probability that we close a cycle at such a step is $O(1/n)$ and so we can safely ignore this possibility.

In addition we define a set $W$ of used vertices, where $W = V_\sigma \cup V_\tau$ at the beginning of Phase 2 and whenever we look at an edge $vw$ (that is, consider using that edge to create a new $\Gamma'$), we add both $v$ and $w$ to $W$. Additionally, we maintain $|W| = O(n^{11/12})$.

We will build a tree $T$ of PN2Fs, breadth-first, where each non-leaf vertex $\Gamma$ yields PN2F children $\Gamma'$ as above. We stop building $T$ when we have $\nu_L$ leaves. This will end Stage 1 for the current cycle $C$ being removed.

We’ll restrict the set of PN2F’s which could be children of $\Gamma$ in $T_0$ as follows: We restrict our attention to $w \notin W$ with $(v, w) \in E_B$ and $(v, w)$ acceptable as defined above. Also, we only construct children from the first $\ell_0$ acceptable $(v, w)$’s at a vertex $v$. Furthermore we only build the tree down to $\ell_1$ levels. We denote the nodes in the $i$th level of the tree by $S_i$. Thus $S_0 = \{\Gamma_1\}$ and $S_{i+1}$ consists of the PN2F’s that are obtained from $S_i$ using acceptable edges. In this way we define a tree of PN2F’s with root $\Gamma_1$ that has branching factor at most $\ell_0$. Thus

$$|S_{\ell_1}| \leq \nu_L. \quad (17)$$

On the other hand, if we let $\mathcal{E}_0$ denote the intersection of the high probability events of Lemmas 2, 7 and 8, then:

**Lemma 10** Conditional on the event $\mathcal{E}_0$,

$$|S_{\ell_1}| = \nu_L$$

with probability $1 - o(n^{-2})$.

**Proof** If $P(\Gamma)$ has endpoints $u_0, v$ and $e = (v, w) \in E_B$ and $e$ is unacceptable then i) $w$ lies on $P(\Gamma)$ and is too close to an endpoint or (ii) $x \in W$ or $w \in W$ or (iii) $w$ lies on a small cycle. Ab initio, there are at least $L_0$ choices for $w$ and we must bound the number of unacceptable choices. The probability that $\ell_0/10$ vertices are unacceptable due to (iii) is by Lemmas 7 and 9 at most

$$(1 + o(1)) \left( \frac{n_b}{\ell_0/10} \right) \left( \frac{7 \log n}{(10 + o(1))n} \right)^{\ell_0/10} \leq \left( \frac{10en_b \log n}{\ell_0 n} \right)^{\ell_0/10} \leq \left( \frac{1000e \log \log n}{\log \log n} \right)^{\ell_0/10} = O(n^{-K}) \quad (18)$$
for any constant $K > 0$.

A similar argument deals with conditions (i) and (ii).

Thus, with (conditional) probability $1 - o(n^{-2})$,

$$|S_{t+1}| \geq \left(\frac{\log n}{100} - \frac{3\log n}{200}\right)|S_t| \geq \frac{\log n}{200}|S_t|$$

for all $t$. So, with (conditional) probability $1 - o(n^{-2})$

$$|S_{\ell_1}| \geq \nu_L$$

as desired. \qed

Having built $T$, if we have not already made a cycle, we have a tree of PN2Fs and the last level, $\ell_1$ has leaves $\Gamma_i$, $i = 1, \ldots, \nu_L$, each with a path $P(\Gamma_i)$ of length at least $n_c$. Now, perform a second stage which will be like executing $\nu_L$-many Stage 1’s in parallel by constructing trees $T_i$, $i = 1, \ldots, \nu_L$, where the root of $T_i$ is $\Gamma_i$. Suppose for each $i$, $\nu_L(\Gamma_i) \in \mathcal{P}(u_0, v_i)$; we fix the vertex $v_i$ and build paths by first looking at neighbors of $u_0$, for all $i$ (so in tree $T_i$, every $\Gamma$ will have path $P(\Gamma) \in \mathcal{P}(u, v_i)$ for some $u$).

Construct these $\nu_L$ trees in the Stage 2 by only enforcing the conditions that $w \notin W$. This change will allow the PN2Fs to have small paths and cycles. We will not impose a bound on the branching factor either. As a result of this and the fact that each tree $T_i$ begins by considering edges from $E_B$ adjacent to $u_0$, the sets of endpoints of paths (that are not the $v_i$s) of PN2Fs at the same level are the same in each of the trees $T_i, i = 1, 2, \ldots, \nu_L$. That is, if $\Gamma_i'$ is a node at level $\ell$ of tree $T_i$ and $\Gamma_j'$ is a node at level $\ell$ of tree $T_j$, $P(\Gamma_i') \in \mathcal{P}(w, v_i)$ and $P(\Gamma_j') \in \mathcal{P}(w, v_j)$ for some $w \in V_0$. This can be proved by induction, see [3].

The trees $T_i, i = 1, \ldots, \nu_L$, will be succesfully constructed with probability $1 - o(1/n)$ and with similar probability the number of nodes in each tree is at most $(10\log n)^{\ell_1} = n^{2/3+o(1)}$. Here we use the fact that the maximum degree in $G_{\ell_1} \leq 10\log n$ with this probability. However, some of the trees may not follow all of the conditions listed initially, so we’ll ‘prune’ the trees by disallowing any node $\Gamma$ that was constructed in violation of any of those conditions. Call tree $T_i$ GOOD if it still has at least $L_0$ leaves remaining after pruning and BAD otherwise. Notice that

$$\Pr(\exists i : T_i \text{ is BAD}) = o\left(\frac{\nu_L}{n^2}\right) = o(n^{-1}).$$

Finally, consider the probability that there is no $E_B$ edge from any of the $n^{2/3-o(1)}$ endpoints found in Stage 1 to any of the $n^{2/3-o(1)}$ endpoints found in Stage 2. At this point we will have only exposed the edges of $\Pi_0$ incident with these endpoints. So if for some $k \leq \nu_L$ we examine the (at least) $\log n/200$ edges incident to $v_1, v_2, \ldots, v_k$ but not $W$ then the probability we fail to close a cycle and produce a proper 2-factor is at most

$$\left(1 - \frac{1}{n^{1/3-o(1)}}\right)^k \log n/200.$$

Thus taking $k = n^{1/3+o(1)}$ suffices to give a sufficently high probability. Also, this only contributes $n^{1/3+o(1)}$ to $W$. 

12
Therefore, the probability that we fail to eliminate a particular small cycle $C$ is $o(1/n)$ and then given $E_0$, the probability that Phase 2 fails is $o(\log n/n) = o(1)$.

We should check now that w.h.p. $|W| = O(n^{11/12})$ throughout Phase 2. It starts out with at most $n^{11/12} + n^{2/5}$ vertices (see Lemmas 2(a) and 8(a)) and we add $O(n^{2/3+o(1)} \times \log n)$ vertices altogether in this phase.

Lemma 11 The probability that Phase 2 fails to produce a proper 2-factor with minimum cycle length at least $n_c$ is $o(n^{-0.51})$.

\[ \square \]

### 3.3 Creating a Hamilton cycle

By the end of Phase 2, we will w.h.p. have found a proper 2-factor with all cycles of length at least $n_c$. Call this subgraph $\Pi^\ast$.

In this section, we will use the edges in

\[ E_C = \{ e \in E_{t_0} \setminus (E_{t_0}/10 \cup E(\Psi_1)) : e \cap V_0 = \emptyset \} \]

to turn $\Pi^\ast$ into a Hamilton cycle that contains $M_0$, w.h.p. It is basically a second moment calculation with a twist to keep the variance under control. We note that Lemma 9 continues to hold if we replace $E_B$ by $E_C$.

Arbitrarily assign an orientation to each cycle. Let $C_1, \ldots, C_k$ be the cycles of $\Pi^\ast$ (note that if $k = 1$ we are done) and let $c_i = |C_i \setminus W|/2$. Then $c_i \geq \frac{n_0}{2} - O(n^{11/12}) \geq \frac{99n}{\log n}$ for all $i$. Let $a = \frac{n}{\log n}$ and $m_i = 2\left\lceil \frac{c_i}{a} \right\rceil + 1$ for all $i$ and $m = \sum_{i=1}^k m_i$. From each $C_i$, we will consider choosing $m_i$ vertices $v \in C_i \setminus W$ that are heads of non-$M_0$ arcs after the arbitrary ordering of all cycles, deleting these $m$ arcs and replacing them with $m$ others to create a proper Hamilton cycle.

Re-label (temporarily) the broken arcs as $(v_i, u_i), i \in [m]$ as follows: in cycle $C_i$ identify the lowest numbered vertex $x_i \in [n]$ which loses a cycle edge directed out of it. Put $v_1 = x_1$ and then go round $C_1$ defining $v_2, v_3, \ldots, v_{m_1}$ in order. Then let $v_{m_1+1} = x_2$ and so on. We thus have $m$ path sections $P_j \in \mathcal{P}(u_{\phi(j)}, v_j)$ in $\Pi^\ast$ for some permutation $\phi$. We see that $\phi$ is an even permutation as all the cycles of $\phi$ are of odd length.

It is our intention to rejoin these path sections of $\Pi^\ast$ to make a Hamilton cycle using $E_C$, if we can. Suppose we can. This defines a permutation $\rho$ where $\rho(i) = j$ if $P_i$ is joined to $P_j$ by $(v_i, u_{\phi(j)})$, where $\rho \in H_m$, the set of cyclic permutations on $[m]$. We will use the second moment method to show that a suitable $\rho$ exists w.h.p. A technical problem forces a restriction on our choices for $\rho$. This will produce a variance reduction in a second moment calculation.

Given $\rho$ define $\lambda = \phi \rho$. In our analysis we will restrict our attention to $\rho \in R_\phi = \{ \rho \in H_m : \phi \rho \in H_m \}$. If $\rho \in R_\phi$ then we have not only constructed a Hamilton cycle in $\Pi^\ast \cup E_C$, but also in the auxiliary digraph $\Lambda$, whose edges are $(i, \lambda(i))$.

The following lemma is from [4]. The content is in the lower bound. It shows that there are still many choices for $\rho$ and it is needed to show that the expected number of possible re-arrangements of path sections, grows with $n$. 

13
Lemma 12 \((m - 2)! \leq |R_\phi| \leq (m - 1)!\)

Let \(H\) be the graph induced by the union of \(\Pi^*\) and \(E_C\).

Lemma 13 \(\Pr(H\ does\ not\ contain\ a\ Hamilton\ cycle) = o(1)\).

Proof \ Let \(X\) be the number of Hamilton cycles in \(G\) that can be obtained by removing the edges described above and rearranging the path segments generated by \(\phi\) according to those in \(\rho \in R_\phi\) and connecting the path segments using edges in \(H\).

We will use the inequality \(\Pr(X > 0) \geq \frac{\Pr(X)^2}{\Pr(X)}\) to show that w.h.p. such a Hamilton cycle exists.

The definition of \(m_i\) gives us \(2a_n - k \leq m \leq 2a_n + k\) and so \(1.99 \log n \leq m \leq 2.01 \log n\). Additionally we will use \(k \leq \frac{n}{m} = \log n \cdot \frac{200}{n^c},\ m_i \geq 199\) and \(\frac{c_i}{m_i} \geq \frac{a}{200}\) for all \(i\).

From Lemma 12, with \(\alpha = 1/(10 + o(1))\)

\[
\mathbb{E}(X) \geq (1 - o(1)) \left(\frac{2a \log n}{n}\right)^{m} (m - 2)! \prod_{i=1}^{k} \binom{c_i}{m_i} \tag{20}
\]

\[
\geq 1 - o(1) \left(\frac{2ma \log n}{en}\right)^{m} \prod_{i=1}^{k} \left(\frac{c_i e^{1-1/10m_i}}{m_i^{1+1/2m_i}}\right) \left(\frac{1 - 2m_i^2/c_i}{\sqrt{2\pi}}\right) \tag{21}
\]

where to go from (20) to (21) we have used the approximation \((m - 2)! \geq m^{-3/2}(m/e)^m\) and

\[
\left(\frac{c_i}{m_i}\right) \geq \frac{c_i m_i (1 - 2m_i^2/c_i)}{m_i!} \text{ and } m_i! \leq \sqrt{2\pi m_i} \left(\frac{m_i}{e}\right)^{m_i} e^{1/10m_i}.
\]

Explanation of (20): We choose the arcs to delete in \(\prod_{i=1}^{k} \binom{c_i}{m_i}\) ways and put them together as explained prior to Lemma 12 in at least \((m - 2)!\) ways. The probability that the required edges exist in \(E_C\) is \((1 + o(1)) \left(\frac{2a \log n}{n}\right)^{m}\), from Lemma 9.

Continuing, we have

\[
\mathbb{E}(X) \geq \frac{(1 - o(1))(2\pi)^{-m/398} e^{-k/10}}{m \sqrt{m}} \left(\frac{2ma \log n}{en}\right)^{m} \prod_{i=1}^{k} \left(\frac{c_i e}{(1.02)m_i}\right)^{m_i}
\]

\[
\geq \frac{(1 - o(1))(2\pi)^{-m/398}}{n^{1/2000}m \sqrt{m}} \left(\frac{2ma \log n}{en}\right)^{m} \left(\frac{ea}{2.01 \times 1.02}\right)^{m}
\]

\[
\rightarrow \infty. \tag{22}
\]

Let \(M, M'\) be two sets of selected edges which have been deleted in \(\Pi^*\) and whose path sections have been re-arranged into Hamilton cycles according to \(\rho, \rho'\) respectively. Let \(N, N'\) be the corresponding sets of edges which have been added to make the Hamilton cycles. Let \(\Omega\) denote the set of choices for \(M\) (and \(M'\)).

Let \(s = |M \cap M'|\) and \(t = |N \cap N'|\). Now \(t \leq s\) since if \((v, u) \in N \cap N'\) then there must be a unique \((\tilde{v}, u) \in M \cap M'\) which is the unique \(\Pi^*\)-edge into \(u\). It is shown in [4] that \(t = s\) implies \(t = s = m\) and \((M, \rho) = (M', \rho')\). (This removes a large term from the second moment calculation.)
If \((s, t)\) denotes the case where \(s = |M \cap M'|\) and \(t = |N \cap N'|\), then

\[
E(X^2) \leq E(X) + (1 + o(1)) \sum_{M \in \Omega} \left( \frac{2\alpha \log n}{n} \right)^m \sum_{\Omega, N' \cap N = \emptyset} \left( \frac{2\alpha \log n}{n} \right)^m
\]

\[
+ (1 + o(1)) \sum_{M \in \Omega} \left( \frac{2\alpha \log n}{n} \right)^m \sum_{s=2}^{m} \sum_{t=1}^{m-s-1} \left( \frac{2\alpha \log n}{n} \right)^{m-t}
\]

\[
= E(X) + E_1 + E_2 \text{ say.}
\]

Note that \(E_1 \leq (1 + o(1))E(X)^2\).

Now, with \(\sigma_i\) denoting the number of common \(M \cap M'\) edges selected from \(C_i\),

\[
E_2 \leq E(X)^2 \sum \sum_{s=2}^{m} \sum_{t=1}^{s-1} \left( \frac{s}{t} \right) \left[ \sum_{\sigma_1 + \ldots + \sigma_k = s} \prod_{i=1}^{k} \frac{\left( \frac{m_i}{\sigma_i} \right)}{\left( \frac{c_i}{m_i} \right)} \right] \left( \frac{m-t-1}{(m-2)!} \right) \left( \frac{n}{2\alpha \log n} \right)^{t}.
\]

**Some explanation:** There are \(\binom{m}{s}\) choices for \(N \cap N'\), given \(s\) and \(t\). (See the explanation for \(t \leq s\).) Given \(\sigma_i\), there are \(\binom{m_i}{\sigma_i}\) ways to choose \(M \cap M'\) and \(\binom{c_i-m_i}{\sigma_i}\) ways to choose the rest of \(M' \cap C_i\). After deleting \(M'\) and adding \(N \cap N'\) there are at most \((m-t-1)!\) ways of putting the segments together to make a Hamilton cycle.

We see that

\[
\frac{(c_i-m_i)}{(m_i)} \leq \frac{(c_i)}{(m_i)} = \frac{m_i(m_i-1) \cdots (m_i-\sigma_i+1)}{(c_i-m_i+1) \cdots (c_i-m_i+\sigma_i)} \leq (1 + o(1)) \left( \frac{2.01}{\alpha} \right)^{\sigma_i} \exp \left\{ -\frac{\sigma_i(\sigma_i-1)}{2m_i} \right\}.
\]

Also,

\[
\sum_{i=1}^{k} \frac{\sigma_i^2}{2m_i} \geq \frac{s^2}{2m} \text{ for } \sigma_1 + \ldots + \sigma_k = s
\]

and

\[
\sum_{i=1}^{k} \frac{\sigma_i}{2m_i} \leq \frac{k}{2} \text{ and } \prod_{\sigma_1 + \ldots + \sigma_k = s} \prod_{i=1}^{k} \frac{m_i}{\sigma_i} = \binom{m}{s}.
\]

Using these approximations, we have

\[
\sum_{\sigma_1 + \ldots + \sigma_k = s} \prod_{i=1}^{k} \frac{\left( \frac{m_i}{\sigma_i} \right)}{\left( \frac{c_i}{m_i} \right)} \leq (1 + o(1))e^{k/2} \exp \left\{ -\frac{s^2}{2m} \right\} \left( \frac{2.01}{\alpha} \right)^{\frac{s}{m}} \left( \frac{m-t-1}{(m-2)!} \right) \left( \frac{n}{2\alpha \log n} \right)^{t}.
\]

So we can write

\[
\frac{E_2}{E(X)^2} \leq (1 + o(1))e^{k/2} \sum_{s=2}^{m} \sum_{t=1}^{s-1} \binom{s}{t} \exp \left\{ -\frac{s^2}{2m} \right\} \left( \frac{2.01}{\alpha} \right)^{\frac{s}{m}} \left( \frac{m-t-1}{(m-2)!} \right) \left( \frac{n}{2\alpha \log n} \right)^{t}.
\]

We approximate

\[
\binom{m}{s} \frac{(m-t-1)!}{(m-2)!} \leq C_1 \frac{m^s}{s!} \left( \frac{m-t-1}{e} \right)^{m-t-1} \left( \frac{e}{m-2} \right)^{m-2} \leq C_2 \frac{m^s e^{t+3}}{s! m^{t-1}},
\]

for some constants \(C_1, C_2 > 0\).
Substituting this, we obtain, after sacrificing constants for a slightly larger exponent of \( n \) as first factor,

\[
\frac{E_2}{\mathbb{E}(X)^2} \leq n^{0.01} m \sum_{s=2}^{m} \left( \frac{2.01}{a} \right)^s \frac{m^s}{s!} \exp \left\{ -\frac{s^2}{2m} \right\} \sum_{t=1}^{s-1} \left( \frac{s}{t} \right) \left( \frac{en}{2\alpha m \log n} \right)^t
\]

\[
\leq (1 + o(1)) \left( \frac{m^3}{5en^{0.99}} \right) \sum_{s=2}^{m} \left( \frac{(2.01)en \exp \{-s/2m\}}{2\alpha a \log n} \right)^s \frac{1}{s!}
\]

\[
\leq n^{-9/10}.
\]

To see this, notice that

\[
\sum_{t=1}^{s-1} \left( \frac{s}{t} \right) \left( \frac{en}{2\alpha m \log n} \right)^t \leq m \left( \frac{en}{2\alpha m \log n} \right)^{s-1}
\]

and

\[
\sum_{s=2}^{m} \left( \frac{(2.01)en \exp \{-s/2m\}}{2\alpha a \log n} \right)^s \frac{1}{s!} \leq \sum_{s=2}^{m} \frac{30^s}{s!} \leq e^{30}.
\]

Combining things, we get

\[
\mathbb{E}(X^2) \leq \mathbb{E}(X) + \mathbb{E}(X)^2(1 + o(1)) + \mathbb{E}(X)^2 n^{-9} \quad \text{so}
\]

\[
\frac{(\mathbb{E}X)^2}{\mathbb{E}(X^2)} \geq \frac{1}{\frac{\mathbb{E}X}{\mathbb{E}X} + 1 + o(1) + n^{-9}} \rightarrow 1
\]

as \( n \to \infty \), as desired. \( \square \)

### 3.4 Proof of Corollary 1

We begin the proof by replacing the sequence \( E_0, E_1, \ldots, E_m, \ldots \) by \( E_0', E_1', \ldots, E_m', \ldots \), where the edges of \( E'_m = \{ e'_1, e'_2, \ldots, e'_m \} \) are randomly chosen with replacement. This means that \( e_m \) is allowed to be a member of \( E'_{m-1} \). We let \( G'_m \) be the graph \( ([n], E'_m) \).

If an edge appears a second time, it will be randomly re-colored. We let \( R \) denote the set of edges that get repeated. Note that if \( \tau_{1,1} = \mu \) and \( e_\mu = (v, w) \in R \) then \( v \) or \( w \) is isolated in \( G'_{m-1} \). We cannot apply this argument if both \( v, w \) are of black degree zero at this point. For then we know that there is only one choice for \( e_\mu \). On the other hand, an argument similar to that for Lemma 2(b) shows that w.h.p. there is no white edge joining \( v \) and \( w \) and so \( e_\mu \notin R \).

Explanation: The factor 4 comes from \( v \) or \( w \) has black or white degree one. Next suppose that \( e_\mu = (v, w) \) and that \( v \) has black degree zero in \( G_{\mu-1} \) and \( w \) has positive black degree in \( G'_{\mu-1} \). Then an argument similar to that given for Lemma 2(f) shows that w.h.p. the white degree of \( v \) is \( O(\log n) \) and so \( e_\mu \) has an \( O(\log n/n) \) chance of being in \( R \). There are \( n-1 \) choices for \( w \), of which \( O(\log n) \) put \( e_\mu \) into \( R \). Here we are explicitly conditioning on the fact that \( \mu = \tau_{1,1} \). We cannot apply this argument if both \( v, w \) are of black degree zero at this point. For then we know that there is only one choice for \( e_\mu \). On the other hand, an argument similar to that for Lemma 2(b) shows that w.h.p. there is no white edge joining \( v \) and \( w \) and so \( e_\mu \notin R \).
At time \( m = \tau_1 \) the graphs \( G_{m}^{(b)}, G_{m}^{(w)} \) will w.h.p. contain perfect matchings, see [6]. That paper does not allow repeated edges, but removing them enables one to use the result claimed. We choose random perfect matchings \( M_0, M_1 \) from \( G_{\tau_1}^{(b)}, G_{\tau_1}^{(w)} \).

We couple the sequence \( G_1, G_2, \ldots \), with the sequence \( G'_1, G'_2, \ldots \), by ignoring repeated edges in the latter. Thus \( G'_1, G'_2, \ldots, G'_m \) is coupled with a sequence \( G_1, G_2, \ldots, G_m \) where \( m' \leq m \). It follows from (23) that w.h.p. the coupled processes stop with the same edge. Furthermore, they stop with two independent matchings \( M_0, M_1 \). We can then begin analysing Phase 2 and Phase 3 within this context.

We will prove that
\[
\Pr(M_1 \cap R = \emptyset) \geq n^{-1/2-o(1)}.
\] (24)

Corollary 1 follows from this.

It follows from (24) and the fact that Phase 2 succeeds with probability \( 1 - O(n^{-0.51}) \) that Phase 2 succeeds w.h.p. conditional on \( M_1 \cap R = \emptyset \).

Phase 3 succeeds w.h.p. even if we avoid using edges in \( R \). We have already carried out calculations with an arbitrary set of \( O(n^{11/12}) \) edges that must be avoided. The size of \( R \) is dominated by a binomial \( \text{Bin}(O(n \log n), O(n^{-1} \log n)) \) and so \( |R| = O(\log^2 n) \) w.h.p. So avoiding \( R \) does not change any calculation in any significant way.

Finally note that the Hamilton cycle we obtain is zebraic.

**Proof of (24):** \( R \) is a random set and it is independent of \( M_1 \). Let \( t_B \) be the number of black edges then
\[
\Pr(M_1 \cap R = \emptyset \mid t_B) \geq \left( 1 - \frac{n/2}{N} \right)^{t_B-n/2} \geq \exp \left\{ -t_B \left( \frac{1}{n} + \frac{1}{n^2} \right) \right\}.
\]

But, to remove the conditioning, we take expectations and then by convexity
\[
\mathbb{E} \left( \exp \left\{ -t_B \left( \frac{1}{n} + \frac{1}{n^2} \right) \right\} \right) \geq \exp \left\{ -\mathbb{E}(t_B) \left( \frac{1}{n} + \frac{1}{n^2} \right) \right\} \geq n^{-1/2-o(1)}
\]
since \( \mathbb{E}(t_B) \sim \frac{1}{2} n \log n \). This proves (24).

### 4 Proof of Theorem 2

Let \( t_0, t_1 \) be as in (1). For a vertex \( v \in [n] \) we let its black degree \( d_b(v) \) be the number of black edges incident with \( v \) in \( G_{t_0} \). We define its white degree \( d_w(v) \) analogously. Let a vertex be large if \( d_b(v), d_w(v) \geq L_0 \) and small otherwise.

We will need the following structural properties:

**Lemma 14** The following hold w.h.p.:

(a) No set \( S \) of at most 10 vertices that is connected in \( G_{t_1} \) contains three small vertices.

(b) Let \( a \) be a positive integer, independent of \( n \). No set of vertices \( S \), with \( |S| = s \leq aL_1 \), contains more than \( s + a \) edges in \( G_{t_1} \).

(c) There are at most \( n^{2/3} \) small vertices.

(d) There are at most \( \log^3 n \) isolated vertices in \( G_{t_0} \).

17
Proof

(a) We say that a vertex is a *low color vertex* if it is incident in $G_{t_1}$ to at most $L_\varepsilon = (1+\varepsilon)L_0$ edges of one of the colors, where $\varepsilon$ is some sufficiently small positive constant. Furthermore, it follows from (2) that

$$\Pr(\text{exists a connected } S \text{ in } G_{n,t_1} \text{ with three low color vertices}) \leq b \sum_{k=4}^{10} \binom{n}{k} k^{-2} \left( \frac{N-k+1}{(N-t_1-k+1)} \right)^3 \Pr(\text{vertices 1,2,3 are low color})$$

(25)

$$\leq b \sum_{k=4}^{10} \binom{n}{k} k^{-2} \left( \frac{N-k+1}{(N-t_1-k+1)} \right)^3 \left( 2 \sum_{\ell=0}^{L_\varepsilon} \left( \frac{n-k}{\ell} \right) \left( \frac{p_1}{2} \right)^\ell \left( 1 - \frac{p_1}{2} \right)^{n-k-\ell} \right)^3$$

(26)

$$\leq b \sum_{k=4}^{10} \binom{n}{k} \left( \log \frac{n}{k} \right)^{-1} \left( \frac{n}{k} \right)^{-3} = o(1).$$

**Explanation of (25),(26):** Having chosen our tree, $\frac{N-k+1}{(N-t_1-k+1)}$ is the probability that this tree exists in $G_{t_1}$. Condition on this and choose three vertices. The final $\cdots^3$ in (26) bounds the probability of the event that 1,2,3 are low color vertices in $G_{n,p_1}$. This event is monotone decreasing, given the conditioning, and so we can use (2) to replace $G_{n,t_1}$ by $G_{n,p_1}$ here.

Now a simple first moment calculation shows that w.h.p. each vertex in $[n]$ is incident with $o(\log n)$ edges of $E_{t_1} \setminus E_0$. Hence, for (a) to fail, there would have to be a relevant set $S$ with three vertices, each incident in $G_{t_1}$ with at most $(1+o(1))L_0$ edges of one of the colors, contradicting the above.

(b) We will prove something slightly stronger. Suppose that $p = \frac{K \log n}{n}$ where $K > 0$ is arbitrary. We will show this result for $G_{n,p}$. The result for this lemma follows from $K = 1 + o(1)$ and (2). We get

$$\Pr(\exists S) \leq O(1) \sum_{s \geq 4} \frac{aL_1}{s} \binom{n}{s} \binom{s}{2} \left( \frac{n}{s+a+1} \right) p^a \langle a+1 \rangle$$

$$\leq b \sum_{s \geq 4} \frac{aL_1}{s} \left( \frac{ne^{sep}}{2} \right)^s (sep)^{a+1}$$

$$\leq b (K e^{2 \log n})^{aL_1} \left( \frac{\log n}{n} \right)^{a+1}$$

$$\leq n^{o(1)} \left( \frac{\log n}{n} \right)^{a} \log n \leq n^{1/2 + o(1)}.$$
(d) Using (2) we see that the expected number of isolated vertices in \( G_t \) is \( O(\log^2 n) \). We now use the Markov inequality.

Now fix a pair of large vertices \( x < y \). We will define sets \( S_i^{(b)}(z), S_i^{(w)}(z), i = 0, 1, \ldots, \ell_1, \) \( z = x, y \).

Assume w.l.o.g. that \( \ell_1 \) is even. We let \( S_0^{(b)}(x) = S_0^{(w)}(x) = \{x\} \) and then \( S_1^{(b)}(x) \) (resp. \( S_1^{(w)}(x) \)) is the set consisting of the first \( \ell_0 \) black (resp. white) neighbors of \( x \). We will use the notation \( S_{\leq i}(x) = \bigcup_{j=1}^{i} S_j^{(c)}(x) \) for \( c = b, w \). We now iteratively define for \( i = 0, 1, \ldots, (\ell_1 - 2)/2 \).

\[
\begin{align*}
\hat{S}_{2i+1}^{(b)}(x) &= \left\{ v \notin S_{\leq 2i}^{(b)}(x) : v \neq y \text{ is joined by a black } G_{t_0}\text{-edge to a vertex in } S_{2i}^{(b)}(x) \right\} . \\
\hat{S}_{2i+1}^{(w)}(x) &= \text{the first } \ell_0 \text{ members of } \hat{S}_{2i+1}^{(w)}(x) . \\
\hat{S}_{2i+2}^{(b)}(x) &= \left\{ v \notin S_{\leq 2i+1}^{(b)}(x) : v \neq y \text{ is joined by a white } G_{t_0}\text{-edge to a vertex in } S_{2i+1}^{(b)}(x) \right\} . \\
\hat{S}_{2i+2}^{(w)}(x) &= \text{the first } \ell_0 \text{ members of } \hat{S}_{2i+2}^{(w)}(x) .
\end{align*}
\]

We then define, for \( i = 0, 1, \ldots, (\ell_1 - 2)/2 \).

\[
\begin{align*}
\hat{S}_{2i+1}^{(w)}(x) &= \left\{ v \notin (S_{\leq i}^{(b)}(x) \cup S_{\leq 2i}^{(w)}(x)) : v \neq y \text{ is joined by a white } G_{t_0}\text{-edge to a vertex in } S_{2i}^{(w)}(x) \right\} \\
\hat{S}_{2i+1}^{(b)}(x) &= \text{the first } \ell_0 \text{ members of } \hat{S}_{2i+1}^{(w)}(x) . \\
\hat{S}_{2i+2}^{(w)}(x) &= \left\{ v \notin (S_{\leq i}^{(b)}(x) \cup S_{\leq 2i+1}^{(w)}(x)) : v \neq y \text{ is joined by a black } G_{t_0}\text{-edge to a vertex in } S_{2i+1}^{(w)}(x) \right\} . \\
\hat{S}_{2i+2}^{(b)}(x) &= \text{the first } \ell_0 \text{ members of } \hat{S}_{2i+2}^{(w)}(x) .
\end{align*}
\]

**Lemma 15** If \( 1 \leq i \leq \ell_1 \), then in \( G_{t_0} \), for \( c = b, w \),

\[
\Pr(|\hat{S}_{i+1}^{(c)}(x)| \leq \ell_0 |S_i^{(c)}(x)| \mid |S_j^{(c)}(x)| = \ell_0^j, 0 \leq j \leq i) = O(n^{-\text{any constant}}).
\]

**Proof** This follows easily from (3) and the Chernoff bounds. Each random variable \( \hat{S}^{(c)}(x) \) is binomially distributed with parameters \( n - o(n) \) and \( 1 - (1 - p_0/2)\ell_0 \). The mean is therefore asymptotically \( \frac{1}{2} \ell_0 \log n = \Omega(\log^2 n) \) and we are asking for the probability that it is much less than half its mean.

It follows from this lemma, that w.h.p., we may define \( S_0^{(b)}(x), S_1^{(b)}(x), \ldots, S_{\ell_1}^{(b)}(x) \) where \( |S_1^{(b)}(x)| = \ell_0^0 \) such that for each \( j \) and \( z \in S_j^{(b)}(x) \) there is a zebraic path from \( x \) to \( z \) that starts with a black edge. For \( S_{\ell_1}^{(w)}(x) \) we can say the same except that the zebraic path begins with a white edge.

Having defined the \( S_i^{(c)}(x) \) etc., we define sets \( S_i^{(c)}(y), i = 1, 2, \ldots, \ell_1, \) \( c = b, w \). We let \( S_0^{(b)}(y) = S_0^{(w)}(y) = \{y\} \) and then \( S_1^{(b)}(y) \) (resp. \( S_1^{(w)}(y) \)) is the set consisting of the first \( \ell_0 \) black (resp. white) neighbors of \( y \) that are not in \( S_{\leq \ell_1}^{(b)}(x) \cup S_{\leq \ell_1}^{(w)}(x) \). We note that for \( c = b, w \) we have \( |S_1^{(c)}(y)| \geq L_0 - 18 > \ell_0 \). This follows from Lemma 14(b). Suppose that \( y \) has ten neighbors \( T \) in \( S_{\ell_1}^{(w)}(x) \). Let \( S \) be the set of vertices in the paths from \( T \) to \( x \) in \( S_{\leq \ell_1}^{(w)}(x) \). If \( |S| = s \) then \( S \cup \{y\} \) contains at least \( s + 9 \) edges. This is because every additional neighbour to the first adds \( k \) vertices and \( k + 1 \) edges to the subgraph of \( G_{t_0} \) spanned by \( S \cup \{y\} \), for some \( k \leq \ell_1 \). Now \( s + 1 \leq 10\ell_1 + 1 \leq 7L_1 \) and the \( s + 9 \) edges contradict the condition in the lemma, with \( a = 7 \).
We make a slight change in the definitions of the \( \hat{S}_i^{(c)}(y) \) in that we keep these sets disjoint from the \( S_i^{(c')}(x) \). Thus we take for example

\[
\hat{S}_{2i+1}^{(u)}(y) = \left\{ v \notin (S_{\leq 2i}^{(u)}(y) \cup S_{\leq \ell_1}(x) \cup S_{\leq \ell_1}(x)) : v \text{ is joined by a black } G_{t_0}-\text{edge to a vertex in } S_{2i}^{(w)}(y) \right\}.
\]

Then we note that excluding \( o(n) \) extra vertices has little effect on the proof of Lemma 15 which remains true with \( x \) replaced by \( y \). We can then define the \( S_i^{(c)}(y) \) by taking the first \( \ell_1 \) vertices.

Suppose now that we condition on the sets \( S_i^{(c)}(x), S_i^{(c)}(y) \) for \( c = b, w \) and \( i = 0, 1, \ldots, \ell_1 \). The edges between the sets with \( c = b \) and \( i = \ell_1 \) and those with \( c = w \) and \( i = \ell_1 \) are unconditioned.

Let

\[
\Lambda = \ell_0^{2\ell_1} = n^{4/3-o(1)}.
\]

Then, for example, using (2),

\[
\Pr(\exists \text{ a black } G_{t_0} \text{ edge joining } S_{\ell_1}^{(b)}(x), S_{\ell_1}^{(b)}(y) \leq 3 \left( 1 - \frac{\log n}{(2 + o(1))n} \right)^\Lambda = O(n^{-\text{any constant}}). \quad (27)
\]

Thus w.h.p. there is a zebraic path with both terminal edges black between every pair of large vertices. A similar argument using \( S_{\ell_1}^{(w)}(x), S_{\ell_1}^{(w)}(y) \) shows that w.h.p. there is a zebraic path with both terminal edges white between every pair of large vertices.

If we want a zebraic path with a black edge incident with \( x \) and a white edge incident with \( y \) then we argue that there is a white \( G_{t_0} \) edge between \( S_{\ell_1}^{(b)}(x) \) and \( S_{\ell_1-1}^{(w)}(y) \).

We now consider the small vertices. Let \( V_1 \) be the set of small vertices that have a large neighbor in \( G_{\tau_1} \). The above analysis shows that there is a zebraic path between \( v \in V_1 \) and \( w \in V_1 \cup V_L \), where \( V_L \) is the set of large vertices. Indeed if \( v \) is joined by a black edge to a vertex \( w \in V_L \) then we can continue with a zebraic path that begins with a white edge and we can reach any large vertex and choose the color of the terminating edge to be either black or white. This is useful when we need to continue to another vertex in \( V_1 \).

We now have to deal with small vertices that have no large neighbors at time \( \tau_1 \). It follows from Lemma 14(a) that such vertices have degree one or two in \( G_{\tau_1} \) and that every vertex at distance two from such a vertex is large.

**Lemma 16** All vertices of degree at most two in \( G_{t_0} \) are w.h.p. at distance greater than 10 in \( G_{\tau_1} \).

**Proof** Simpler than Lemma 2(b). We use (3) and then

\[
\Pr(\exists \text{ such a pair of vertices}) \leq 6 \ell_1^{1/2} \sum_{k=0}^{9} n^k p_1^{k-1} \left( (1 - p_0)n^{-k-1} + (n - k)p_0(1 - p_0)n^{-k-2} \right)^2 = o(1).
\]

Let \( Z_i \) be the number of vertices of degree \( 0 \leq i \leq 2 \) in \( G_{t_0} \) that are adjacent in \( G_{\tau_1} \) to vertices that are themselves only incident to edges of one color.

First consider the case \( i = 1, 2 \). Here we let \( Z_i' \) be the number of vertices of degree \( i \) in \( G_{t_0} \) that are adjacent in \( G_{t_0} \) to vertices that are themselves only incident to edges of one color. Note that
\[ Z_i \leq Z_i' \]. Then we have, with the aid of (6),

\[
E(Z'_1) \leq n \binom{n-1}{1} \frac{N-n+1}{t_0-1} \sum_{k=0}^{n-2} \binom{n-2}{k} \left( \frac{N-2n+3}{t_0-1-k} \right) 2^{-(k-1)}. \tag{28}
\]

\[
\leq b n^2 \frac{t_0}{N} \binom{N-2n+2}{N-1} \sum_{k=0}^{n-2} \binom{n-2}{k} 2^{-k} \left( \frac{t_0}{N} \right)^k \left( \frac{N-t_0+1}{N-k} \right)^{n-2-k}
\]

\[
\leq b n \log n \exp \left\{ -\frac{(n-2)(t_0-1)}{N-1} \right\} \sum_{k=0}^{n-2} \binom{n-2}{k} \left( \frac{t_0}{2N} \right)^k \left( \frac{N-t_0+1}{N-n} \right)^{n-2-k}
\]

\[
\leq b \log^3 n \left( \frac{t_0/2}{N-n} \right)^{n-2}
\]

\[
= o(1). \tag{29}
\]

**Explanation for (28):** We choose a vertex \( v \) of degree one and its neighbor \( w \) in \( n^{(n-1)} \) ways. The probability that \( v \) has degree one is \( \binom{N-n+1}{t_0} \). We fix the degree of \( w \) to be \( k+1 \). This now has probability \( \binom{N-2n+3}{t_0-1-k} \). The final factor \( 2^{-(k-1)} \) is the probability that \( w \) only sees edges of one color.

We next eliminate the possibility of a vertex of degree two in \( G_{t_0} \) being in a triangle of \( G_{t_1} \). First, using (2), the expected number of vertices of degree two in \( G_{t_0} \) is at most

\[ 3n \binom{n-1}{2} p_0^2 (1-p_0)^{n-3} = O(\log^4 n). \]

So, w.h.p. there are fewer than \( \log^5 n \). Using (3), we see that the expected number of triangles of \( G_{t_0} \) containing a vertex of degree two is at most

\[ O\left( t_0^{1/2} \right) \times n^3 p_0^3 (1-p_0)^{n-3} = o(1). \]

So, w.h.p. there are no such triangles.

Then the probability that there is an edge of \( G_{t_1} - G_{t_0} \) that joins the two neighbors of a vertex of degree two in \( G_{t_0} \) is at most

\[ o(1) + \log^5 n \times \frac{t_1-t_0}{N} = o(1). \]
Now we can proceed to estimate $E(Z'_2)$, ignoring the possibility of such a triangle. In which case,

$$E(Z'_2) \leq b n \left( \frac{n-1}{2} \right) \left( \frac{N-n+1}{t_0-2} \right) \sum_{k,l=0}^{n-3} \binom{n-3}{k} \binom{n-3}{l} \left( \frac{N-3n+6}{t_0-2-k-l} \right) 2^{-k-l} \tag{30}$$

$$\leq n^3 \left( \frac{t_0}{N} \right)^2 \left( \frac{N-t_0}{N-2} \right)^{n-3} \sum_{k,l=0}^{n-3} \binom{n-3}{k} \binom{n-3}{l} \left( \frac{t_0-2}{N-n+1} \right)^{k+l} \left( \frac{N-t_0+2}{N-k-l} \right)^{2n-5-k-l} 2^{-k-l}$$

$$\leq b \log^4 n \left( \sum_{k=0}^{n-3} \binom{n-3}{k} \left( \frac{t_0-2}{2(N-n+1)} \right)^k \left( \frac{N-t_0+2}{N-2n} \right)^{n-3-k} \right)^2$$

$$\leq \log^4 n \left( 1 - \frac{t_0-2}{2(N-2n)} \right)^{2(n-3)}$$

$$= o(1). \tag{31}$$

Finally, consider $Z_0$. Condition on $G_{t_0}$ and assume that Properties (c), (d) of Lemma 14 hold. The first edge incident with an isolated vertex of $G_{t_0}$ will have a random endpoint. It follows immediately that

$$E(Z_0) \leq o(1) + \log^3 n \times n^{-1/3} = o(1). \tag{32}$$

Here the $o(1)$ accounts for Properties (c), (d) of Lemma 14 and $\log^3 n \times n^{-1/3}$ bounds the expected number of “first edges” that choose small endpoints.

Equations (28), (30) and (32) show that $Z_0 + Z_1 + Z_2 = 0$ w.h.p. In which case it will be possible to find zebraic paths starting from small vertices.

5 Proof of Theorem 3

The case $r = 2$ is implied by Corollary 1 and so we can assume that $r \geq 3$.

5.1 $p \leq (1 - \varepsilon)p_r$

For a vertex $v$, let

$$C_v = \{ i : v \text{ is incident with an edge of color } i \}.$$  

$$I_v = \{ i : \{ i, i + 1 \} \subseteq C_v \}.$$ 

Let $v$ be bad if $I_v = \emptyset$. The existence of a bad vertex means that there are no $r$-zebraic Hamilton cycles. Let $Z_B$ denote the number of bad vertices. Now if $r$ is odd and $C_v \subseteq \{ 1, 3, \ldots, 2\lfloor r/2 \rfloor - 1 \}$ or $r$ is even and $C_v \subseteq \{ 1, 3, \ldots, r - 1 \}$ then $I_v = \emptyset$. Hence,

$$E(Z_B) \geq n \left( 1 - \frac{\alpha_v p}{r} \right)^{n-1} = n^{\varepsilon-o(1)} \to \infty.$$ 

A standard second moment calculation shows that $Z_B \neq 0$ w.h.p. and this proves the first part of the theorem.
5.2 \( p \geq (1 + 3\varepsilon)p_r \)

Note the replacement of \( \varepsilon \) by \( 3\varepsilon \) here, for convenience.

Write \( 1 - p = (1 - p_1)(1 - p_2)^2 \) where \( p_1 = (1 + \varepsilon)p_r \) and \( p_2 \sim \varepsilon p_r \). Thus \( G_{n,p} \) is the union of \( G_{n,p_1} \) and two independent copies of \( G_{n,p_2} \). If an edge appears more than once in \( G_{n,p} \), then it retains the color of its first occurrence.

Now for a vertex \( v \) let \( d_i(v) \) denote the number of edges of color \( i \) incident with \( v \) in \( G_{n,p_1} \). Let

\[
J_v = \{ i : d_i(v) \geq \eta_0 \log n \}
\]

where \( \eta_0 = \frac{\varepsilon^2}{r} \).

Let \( v \) be poor if \( |J_v| < \beta_r \) where \( \beta_r = \lceil r/2 \rceil + 1 \). Observe that \( \alpha_r + \beta_r = r + 1 \). Then let \( Z_P \) denote the number of poor vertices in \( G_{n,p_1} \). A simple calculation shows that w.h.p. the minimum degree in \( G_{n,p_1} \) is at least \( L_0 \) and that the maximum degree is at most \( 6\log n \). Then

\[
E(Z_P) \leq o(1) + n \sum_{k=L_0}^{n-1} \binom{n-1}{k} p_1^k (1 - p_1)^{n-1-k} \sum_{l=r-\beta_r+1}^{r} \binom{r}{l} \left( \eta_0 \log n \right)^l \left( 1 - \frac{l}{r} \right)^{k-r\eta_0 \log n} \\
\leq o(1) + n \sum_{k=0}^{n-1} \binom{n-1}{k} p_1^k (1 - p_1)^{n-1-k} 2^r \left( \frac{6 \log n}{r \eta_0 \log n} \right) \left( \frac{\beta_r - 1}{r} \right)^k \left( \frac{r}{\beta_r - 1} \right)^{r\eta_0 \log n} \\
\leq o(1) + 2^r n^{1+r\eta_0 \log(6\epsilon/\eta_0)} (1 - p_1)^{n-1} \left( 1 + \frac{(\beta_r - 1)p_1}{r(1 - p_1)} \right)^{n-1} \\
\leq o(1) + 2^r n^{1+r\eta_0 \log(6\epsilon/\eta_0)} \left( 1 - \frac{\alpha_r p_1}{r} \right)^{n-1} \\
= o(1).
\]

We can therefore assert that w.h.p. there are no poor vertices. This means that

\[
K_v = \{ i : d_i(v), d_{i+1}(v) \geq \eta_0 \log n \} \neq \emptyset \text{ for all } v \in [n].
\]

The proof now follows our general 3-phase procedure of (i) finding an \( r \)-zebraic 2-factor, (ii) removing small cycles so that we have a 2-factor in which every cycle has length \( \Omega(n/\log n) \) and then (iii) using a second moment calculation to show that this 2-factor can be re-arranged into an \( r \)-zebraic Hamilton cycle.

5.2.1 Finding an \( r \)-zebraic 2-factor

We partition \([n]\) into \( r \) sets \( V_i = ([i-1]n/r + 1, in] \) of size \( n/r \). Now for each \( i \) and each vertex \( v \) let

\[
d_i^+(v) = | \{ w \in V_{i+1} : (v, w) \text{ is an edge of } G_{n,p_1} \text{ of color } i+1 \} |. \\
d_i^-(v) = | \{ w \in V_{i-1} : (v, w) \text{ is an edge of } G_{n,p_1} \text{ of color } i-1 \} |.
\]

(Here 1-1 is interpreted as \( r \) and \( r+1 \) is interpreted as 1).

We now let a vertex \( v \in V_i \) be \( i \)-large if \( d_i^+(v), d_i^-(v) \geq \eta \log n \) where \( \eta = \min \{ \eta_0, \eta_1, \eta_2 \} \) and \( \eta_1 \) is the solution to

\[
\eta_1 \log \left( \frac{e(1 + \varepsilon)}{r \eta_1 \alpha_r} \right) = \frac{1}{r \alpha_r}
\]

and \( \eta_2 \) is the solution to

\[
\eta_2 \log \left( \frac{3 \epsilon r(1 + \varepsilon)}{\eta_2 \alpha_r} \right) = \frac{1}{3 \alpha_r}.
\]
Let $v$ be large if it is $i$-large for all $i$. Let $v$ be small otherwise. (Note that $d^+_i(v), d^-_i(v)$ are defined for all $v$, not just for $v \in V_i, i \in [r]$).

Let $V_\lambda, V_\sigma$ denote the sets of large and small vertices respectively.

**Lemma 17** W.h.p., in $G_{n,p_1}$,

(a) $|V_\sigma| \leq n^{1-\theta}$ where $\theta = \frac{\varepsilon}{r\alpha_r}$.

(b) No connected subset of size at most $\log n$ contains more than $\mu_0 = r\alpha_r$ members of $V_\sigma$.

(c) If $S \subseteq [n]$ and $|S| \leq n = n/\log^2 n$ then $e(S) \leq 100|S|$.

**Proof**

(a) If $v \in V_\sigma$ then there exists $i$ such that $d^+_i(v) \leq \eta \log n$ or $d^-_i(v) \leq \eta \log n$. So we have

$$\mathbb{E}(|V_\sigma|) \leq 2rn \sum_{k=0}^{\eta \log n} \left(\frac{n}{r}\right)^k \left(\frac{p_1}{r}\right)^k \left(1 - \frac{p_1}{r}\right)^{n/r-k}$$

$$\leq 3 \left(\frac{1+\varepsilon}{r\eta\alpha_r}\right)^{\eta \log n} n^{1-(1+\varepsilon+o(1))/\alpha_r}$$

Part (a) follows from the Markov inequality. Note that we can lose the factor 2 in (33) since $d^+_i(v) = d^-_{i+2}(v)$.

(b) The expected number of connected sets $S$ of size $2 \log n$ containing $\mu_0$ members of $V_\sigma$ can be bounded by

$$\sum_{s=\mu_0}^{2\log n \log n} \left(\frac{n}{s}\right)^{s-2} p_1^{s-1} \left(\frac{s}{\mu_0}\right)^{\eta \log n} \left(\frac{n}{r-s}\right)^{k} \left(\frac{p_1}{r}\right)^k \left(1 - \frac{p_1}{r}\right)^{n/r-s-k} \mu_0$$

**Explanation:** We choose $s$ vertices for $S$ and a tree to connect up the vertices of $S$. We then choose $\mu_0$ members $A \subseteq S$ to be in $V_\sigma$. We multiply by the probability that for each vertex in $A$, there is at least one $j$ such that $v$ has few neighbors in $V_j \setminus S$ connected to $v$ by edges of color $j$.

The sum in (36) can be bounded by

$$n \sum_{s=\mu_0}^{2\log n \log n} (2\varepsilon \log n)^{s} n^{-\mu_0(1+\varepsilon+o(1))/\alpha_r} = o(1).$$

(c) This is proved in the same manner as Lemma 2(c).

For $v \in V_\sigma$ we let $\phi(v) = \min \{i : v \text{ is } i\text{-large}\}$ and then let $X_i = \{v \in V_\sigma : \phi(v) = i\}$ for $i \in [r]$. Now let

$$W_i = (V_i \setminus V_\sigma) \cup \{v \in V_\sigma : \phi(v) = i\}, \quad i = 1, 2, \ldots, r.$$ 

Suppose that $w_i = |W_i| - n/r$ for $i \in [r]$ and let $w^+_i = \max \{0, w_i\}$ for $i \in [r]$. We now remove $w^+_i$ large random vertices from each $W_i$ and then randomly assign $w_i^- = \min \{0, w_i\}$ of them to each $W_i, i \in [r]$. Thus we obtain a partition of $[n]$ into $r$ sets $Z_i, i = 1, 2, \ldots, r$, of size $n/r$ for $i \in [r]$.

Let $H_i$ be the bipartite graph induced by $W_i, W_{i+1}$ and the edges of color $i$ in $G_{n,p_1}$. We now argue that
Lemma 18 \[ H_i \] has minimum degree at least \( \frac{1}{2} \eta \log n \) w.h.p.

**Proof**\[ H_i \] It follows from Lemma 17(b) that no vertex in \( W_i \cap V_i \) loses more than \( \mu_0 \) neighbors from the deletion of \( V_\sigma \). Also, we move \( v \in V_\sigma \) to a \( W_i \) where it has large degree in \( V_{i-1} \) and \( V_{i+1} \). Its neighborhood may have been affected by the deletion of \( V_\sigma \), but only by at most \( \mu_0 \). Thus for every \( i \) and \( v \in X_i \), \( v \) has at least \( \eta \log n - \mu_0 \) neighbors in \( W_{i-1} \) connected to \( v \) by an edge of color \( i-1 \). Similarly w.r.t. \( i+1 \).

Now consider the random re-shuffling to get sets of size \( n/r \). Fix a \( v \in V_i \). Suppose that it has \( d = \Theta(\log n) \) neighbors in \( W_{i+1} \) connected by an edge of color \( i+1 \). Now randomly choose \( w_{i+1} \leq |V_\sigma| \) to delete from \( W_{i+1} \). The number \( \nu_v \) of neighbors of \( v \) chosen is dominated by

\[
\text{Bin} \left( w_{i+1}, \frac{d}{n/r-w_{i+1}} \right).
\]

This follows from the fact that if we choose these \( w_{i+1} \) vertices one by one, then at each step, the chance that the chosen vertex is a neighbor of \( v \) is bounded from above by

\[
\frac{d}{n/r-w_{i+1}}.
\]

So, given the condition in Lemma 17(a) we have

\[
\Pr(\nu_v \geq 2/\theta) \leq \left( \frac{n^{1-\theta}}{2/\theta} \right) \left( \frac{dr}{n-o(n)} \right)^{2/\theta} \leq \left( \frac{n^{1-\theta} \epsilon_\theta}{n} \right)^{2/\theta} = o(n^{-1}).
\]

We can now verify the typical existence of perfect matchings.

Lemma 19 W.h.p., each \( H_i \) contains a perfect matching \( M_i \), \( i = 1, 2, \ldots, r \).

**Proof** Fix \( i \). We use Hall’s theorem and consider the existence of a set \( S \subseteq V_i \) that has fewer than \( |S| \) \( H_i \)-neighbors in \( W_{i+1} \). Let \( s = |S| \) and let \( T = N_{H_i}(S) \) and \( t = |T| < s \). We can rule out \( s \leq n_0 = n/\log^2 n \) through Lemma 17(c). This is because we have \( e(S \cup T)/|S \cup T| \geq \frac{1}{2} \eta \log n \) in this case. Let \( n_\sigma = |V_\sigma| \) and now consider \( n/\log^2 n \leq s \leq n/2r \). Given such a pair \( S, T \) we deduce that there exist \( S_1 \subseteq S \subseteq V_i, |S_1| \geq s-n_\sigma \), and \( T_1 \subseteq T \subseteq V_{i+1} \) and \( U_1 \subseteq V_{i+1} \), \( |U_1| \leq n_\sigma \) such that there are at least \( m_s = (s \eta/2 - 6n_\sigma) \log n \) edges between \( S_1 \) and \( T_1 \) and no edges between \( S_1 \) and \( V_{i+1} \setminus (T_1 \cup U_1) \). There is no loss of generality in increasing the size of \( T \) to \( s \). We can then write

\[
\Pr(\exists S, T) \leq \sum_{s=n_0}^{n/2r} \left( \frac{n/r-n_\sigma}{s} \right)^2 \left( \frac{s^2}{m_s} \right)^p \left( 1 - p_1 \right)^{(s-n_\sigma)(n/r-s-n_\sigma)} \leq \sum_{s=n_0}^{n/2r} \left( \frac{n \sigma e}{r s} \right)^2 \left( \frac{s^2 p_1 e}{m_s} \right)^{m_s} e^{-(s-n_\sigma)(n/r-s-n_\sigma)p_1} \leq \left( \frac{s \eta \log n/3}{n} \right)^{2s} \left( \frac{3\epsilon r (1+\epsilon)}{\alpha_r \eta} \right)^{\eta \log n/2} n^{-(1-o(1))/2\alpha_r} = o(1).
\]

For the case \( s \geq n/2r \) we look for subsets of \( V_{i+1} \) with too few neighbors.

It follows from symmetry considerations that the \( M_i \) are independent of each other. Analogously to Lemma 7, we have

Lemma 20 The following hold w.h.p.:
(a) $\bigcup_{i=1}^{r} M_i$ has at most $10 \log n$ components. (Components are $r$-zebraic cycles of length divisible by $r$.)

(b) There are at most $n_b$ vertices on components of size at most $n_c$.

Proof The matchings induce a permutation $\pi$ on $W_1$. Suppose that $x \in W_1$. We follow a path via a matching edge to $W_2$ and then by a matching edge to $W_3$ and so on until we return to a vertex $\pi(x) \in W_1$. $\pi$ can be taken to be a random permutation and then the lemma follows from Lemma 7.

The remaining part of the proof is similar to that described in Sections 3.2, 3.3. We use the edges of the first copy $G_{n,p_2}$ of color 1 to make all cycles have length $\Omega(n/\log n)$ and then we use the edges of the second copy of $G_{n,p_2}$ of color 1 to create an $r$-zebraic Hamilton cycle. The details are left to the reader.

References


