A greedy algorithm for finding a large 2-matching on a random cubic graph

Deepak Bal Patrick Bennett Tom Bohman*
Alan Frieze†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213
USA

September 28, 2012

Abstract

A 2-matching of a graph G is a spanning subgraph with maximum degree two. The size of a 2-matching U is the number of edges in U and this is at least $n - \kappa(U)$ where n is the number of vertices of G and κ denotes the number of components. In this paper, we analyze the performance of a greedy algorithm 2GREEDY for finding a large 2-matching on a random 3-regular graph. We prove that with high probability, the algorithm outputs a 2-matching U with $\kappa(U) = \tilde{\Theta}\left(n^{1/5}\right)$.

1 Introduction

In this paper we analyze the performance of a generalization of the well-known Karp-Sipser algorithm [13, 12, 1, 4] for finding a large matching in a sparse random graph. A 2-matching U of a graph G is a spanning subgraph with maximum degree two. Our aim is to show that w.h.p. our algorithm finds a large 2-matching in a random cubic graph. The algorithm 2GREEDY is described below and has been partially analyzed on the random graph $G_{n,cn}^{\delta\geq 3}$, $c \geq 10$ in Frieze [9]. The random graph $G_{n,m}^{\delta\geq 3}$ is chosen uniformly at random from the collection of all graphs that have n vertices, m edges and minimum degree $\delta(G) \geq 3$. In [9], the 2-matching output by the algorithm is used to find a Hamilton cycle in $O(n^{1.5+o(1)})$ time w.h.p. Previously, the best known result for this model was that $G_{n,cn}^{\delta\geq 3}$ is Hamiltonian for $c \geq 64$ due to Bollobás, Cooper, Fenner and Frieze [7]. It is conjectured that $G_{n,cn}^{\delta\geq 3}$ is Hamiltonian w.h.p. for all $c \geq 3/2$.

The existence of Hamilton cycles in other random graph models with $O(n)$ edges has also been the subject of much research. In such graphs, the requirement $\delta \geq 3$ is necessary to avoid three vertices of degree two sharing a common neighbor. This obvious obstruction occurs with positive

*Research supported in part by NSF Grant DMS1001638
†Research supported in part by NSF Grant CCF1013110

probability in many models with $O(n)$ edges and $\delta = 2$. $G_{3\text{-out}}$ is a random graph where each vertex chooses 3 neighbors uniformly at random. This graph has minimum degree 3 and average degree 6. Bohman and Frieze proved that $G_{3\text{-out}}$ is Hamiltonian w.h.p. also by building a large 2-matching into a Hamilton cycle [3]. Robinson and Wormald proved that r-regular graphs with $r \geq 3$ are Hamiltonian w.h.p. using an intricate second moment approach [14],[15]. Before this result, Frieze proved Hamiltonicity of r-regular graphs w.h.p. for $r \geq 85$ using an algorithmic approach [10]. An algorithmic proof of Hamiltonicity for $r \geq 3$ was given in [11].

In addition to the Hamiltonicity of $G_{n,cn}^{\delta\geq 3}$ for $3/2 < c < 10$, the Hamiltonicity of random graphs with $O(n)$ edges and a fixed degree sequence is a widely open question. One natural example is the Hamiltonicity of a graph chosen uniformly at random from all the collection of all graphs with $n/2$ vertices of degree 3 and $n/2$ vertices of degree 4 (this particular question was posed by Wormald). For both $G_{n,cn}^{\delta\geq 3}$ and graphs with a fixed degree sequence one might hope to prove Hamiltonicity by first using 2GREEDY to produce a large 2-matching and then using an extension rotation argument to convert this 2-matching into a Hamilton cycle. In this paper we provide evidence that the first half of this broad program is feasible by showing that 2GREEDY finds a very large 2-matching for the sparsest of the models with minimum degree 3, the random cubic graph itself.

The size of a 2-matching U is the number of edges in U and this is at least $n - \kappa(U)$ where κ denotes the number of components. It was shown in [12] that w.h.p. the Karp-Sipser algorithm only leaves $\tilde{\Theta}(n^{1/5})$ vertices unmatched. Here we prove the corresponding result for 2GREEDY on a random cubic graph.

Theorem 1.1. *Algorithm* 2GREEDY *run on a random 3-regular graph with n vertices outputs a 2-matching U with $\kappa(U) = \tilde{\Theta}(n^{1/5})$, w.h.p.*

We prove Theorem 1.1 using the differential equations method for establishing dynamic concentration. The remainder of the paper is organized as follows. The 2GREEDY algorithm is introduced in the next Section, and the random variables we track are given in Section 3. The trajectories that we expect these variables to follow are given in Section 4. A heuristic explanation of why 2GREEDY should produce a 2-matching with roughly $n^{1/5}$ components is also given in Section 4. In Section 5 we state and prove our dynamic concentration result. The proof of Thereom 1.1 is then completed in Sections 5, 6, and 7.

2 The Algorithm

The Karp-Sipser algorithm for finding a large matching in a sparse random graph is essentially the greedy algorithm, with one slight modification that makes a big difference. While there are vertices of degree one in the graph, the algorithm adds to the matching an edge incident with such a vertex. Otherwise, the algorithm chooses a random edge to add to the matching. The idea is that no mistakes are made while pendant edges are chosen since such edges are always contained in some maximum matching. The algorithm presented in [9] is a generalization of Karp-Sipser for 2-matchings. Our algorithm is essentially the same as that presented in [9] applied to random cubic graphs. A few slight modifications have been made to ease the analysis and to account for the change in model. We assume that our input (multi-)graph $G = G([n], E)$ is generated by the configuration model of Bollobas [6]. Let $W = [3n]$ be our set of *configuration points* and let $W_i = [3(i-1)+1, 3i]$, $i \in [n]$, partition W. The function $\phi : W \to [n]$ is defined by $w \in W_{\phi(w)}$. Given a pairing F (i.e. a partition of W into $m = 3n/2$ pairs) we obtain a (multi-)graph G_F with vertex set $[n]$ and an edge $(\phi(u), \phi(v))$ for each $\{u, v\} \in F$. Choosing a pairing F uniformly at random from among all possible pairings Ω of the points of W produces a random (multi-)graph G_F. It is known that conditional on G_F being simple, i.e. having no loops or multiple edges, that it is equally likely to be any (simple) cubic graph. Further, G_F is simple with probability $(1-o(1))e^{-2}$. So from now on we work with $G = G_F$.

We only reveal adjacencies (pairings) of G_F as the need arises in the algorithm. As the algorithm progresses, it grows a 2-matching and deletes vertices and edges from the input graph G. We let $\Gamma = (V_\Gamma, E_\Gamma)$ be the current state of G. Throughout the algorithm we keep track of the following:

- U is the set of edges of the current 2-matching. The internal vertices and edges of the paths and cycles in U will have been deleted from Γ.
- $b(v)$ is the 0-1 indicator for vertex $v \in [n]$ being adjacent to an edge of U.
- $Y_k = \{v \in V_\Gamma : d_\Gamma(v) = k,\, b(v) = 0\}$, $k = 0, 1, 2, 3$.
- $Z_k = \{v \in V_\Gamma : d_\Gamma(v) = k,\, b(v) = 1\}$, $k = 0, 1, 2$.

We refer to the sets Y_3 and Z_2 as Y and Z throughout. The basic idea of the algorithm is as follows. We add edges to the 2-matching one by one, which sometimes forces us to delete edges. These deletions may put vertices in danger of having degree less than 2 in the final 2-matching. Thus, we prioritize the edges that we add to U, so as to match the dangerous vertices first. More precisely, At each iteration of the algorithm, a vertex v is chosen and an adjacent edge is added to U. We choose v from the first non-empty set in the following list: Y_1, Y_2, Z_1, Y, Z. As in the Karp-Sipser algorithm, taking edges adjacent to the vertices in Y_1, Y_2 and Z_1 is not a mistake. We will prove that by proceeding in this manner, we do not create too many components.

When a vertex v is chosen and its neighbor in the configuration is exposed it is called a *selection move*. Call the revealed neighbor, w the *selection*. The edge (v, w) is removed from Γ and added to U. If the selction w is a vertex in Z, then once (v, w) is added to U, we must delete the other edge adjacent to w. Hence we reveal the other edge in the configuration adjacent to w. Call this exposure a *deletion move*.

Details of the algorithm are now given.

Algorithm 2Greedy:

Initially, all vertices are in Y. Iterate the following steps as long as one of the conditions holds.

Step 1(a) $Y_1 \neq \emptyset$.

Choose a random vertex v of Y_1. Suppose its neighbor in Γ is w. Remove (v, w) from Γ and add it to U. Set $b(v) = 1$ and move v to Z_0.

re-assign(w).

Step 1(b) $Y_1 = \emptyset, Y_2 \neq \emptyset$.

Choose a random vertex v of Y_2. Randomly choose one of the two neighbors of v in Γ to expose and call it w.

If $w = v$ ($\{v\}$ comprises an isolated component in Γ with a loop), then remove (v, v) from Γ and move v from Y_2 to Y_0.

Otherwise, remove (v, w) from Γ and add it to U. Set $b(v) = 1$ and move it to Z_1.

re-assign(w).

Step 1(c) $Y_1 = Y_2 = \emptyset, Z_1 \neq \emptyset$.

Choose a random vertex v of Z_1. v is the endpoint of a path in U. Let u be the other endpoint of this path. Suppose the neighbor of v in Γ is w. Remove (v, w) from Γ and add it to U. Remove v from Γ.

re-assign(w).

Step 2 $Y_1 = Y_2 = Z_1 = \emptyset, Y \neq \emptyset$.

Choose a random vertex v of Y. Randomly choose one of the three neighbors of v in Γ to expose and call it w.

If $w = v$, then we remove (v, v) from Γ and move v to Y_1.

Otherwise, remove (v, w) from Γ and add it to U. Set $b(v) = 1$ and move it to Z.

re-assign(w).

Step 3 $Y_1 = Y_2 = Z_1 = Y = \emptyset, Z \neq \emptyset$

The remaining graph is a random 2-regular graph on $|Z|$ many vertices. Put a maximum matching on the remaining graph. Add the edges of this matching to U.

Subroutine re-assign(w):

1. If $b(w) = 0$:

   Set $b(w) = 1$ and move w from Y to Z, Y_2 to Z_1 or Y_1 to Z_0 depending on the initial state of w.

2. If $b(w) = 1$:

   Remove w from Γ. If w was in Z prior to removal, then the removal of w from Γ causes an edge (w, w'), to be deleted from Γ. Move w' to the appropriate new set. For example, if w' were in Z, it would be moved to Z_1; if w' were in Y, it would be moved to Y_2, etc.

3 The Variables

In this section we will describe the variables which are tracked as the algorithm proceeds. Throughout the paper, in a slight abuse of notation, we let Y, Z, etc. refer to both the sets and the size of the set. Let M refer to the size of E_Γ. We also define the variable

$$\zeta := Y_1 + 2Y_2 + Z_1.$$

If X is a variable indexed by i, we define

$$\Delta X(i) := X(i+1) - X(i).$$

3.1 The sequences σ, δ

We define two sequences σ, δ indexed by the step number i. $\sigma(i)$ will indicate what type of vertex is selected during a selection move, and $\delta(i)$ will do the same for deletion moves.

Formally, σ is a sequence of the following symbols: $Y, Z, \zeta, loop, multi$. We will put $\sigma(i) = loop$ only when step i is of type 2 and the selection move reveals a loop. We put $\sigma(i) = multi$ only when step i is of type $1(c)$, and $w = u \in Z$. The only way this happens is when $v \in Z_1$, $u \in Z$, $(v, u) \in U$, and the selection made at step i happens to select the vertex u. Otherwise we just put $\sigma(i) = Y, Z, \zeta$ according to whether the selected vertex is in Y, Z, ζ.

Note that the symbols $loop, multi$ are for very specific events, and not just any loop or multi-edge. If step i is of type $1(b)$ and our selection move reveals a loop, then we put $\sigma(i) = \zeta$. Also, if step i is of type $1(c)$ and the selection move reveals a multi-edge whose other endpoint is also in Z_1 then we put $\sigma(i) = \zeta$ as well.

δ is a sequence of symbols: $Y, Z, \zeta, \emptyset$. We will put $\delta(i) = \emptyset$ when there is no deletion move at step i (i.e. when $\sigma(i) \notin \{Z, multi\}$). Otherwise $\delta(i)$ just indicates the type of vertex that the deletion move picks (here we don't make any distinctions regarding loops or multi-edges).

3.2 The variables A, B

We will define the following two important variables:

$$A := Y + \zeta$$
$$B := 2Y + Z + \zeta.$$

A is a natural quantity to define, since the algorithm terminates precisely when $A = 0$. B is also natural because it represents the number of half-edges which will (optimistically) be added to our current 2-matching before termination. We will see that A and B are also nice variables in that their 1-step changes $\Delta A(i), \Delta B(i)$ do not depend on what type of step we take at step i. We have

$$\Delta Y(i) = -\mathbb{1}_{\zeta(i)=0} - \mathbb{1}_{\sigma(i)=Y} - \left(\mathbb{1}_{\sigma(i)=Z} + \mathbb{1}_{\sigma(i)=multi}\right) \mathbb{1}_{\delta(i)=Y} \tag{3.1}$$

$$\begin{aligned} \Delta Z(i) = {} & \mathbb{1}_{\zeta(i)=0} + \mathbb{1}_{\sigma(i)=Y} - \mathbb{1}_{\sigma(i)=Z} - \mathbb{1}_{\sigma(i)=loop} - \mathbb{1}_{\sigma(i)=multi} \\ & - \left(\mathbb{1}_{\sigma(i)=Z} + \mathbb{1}_{\sigma(i)=multi}\right) \mathbb{1}_{\delta(i)=Z} \end{aligned} \tag{3.2}$$

$$\begin{aligned} \Delta \zeta(i) = {} & -\mathbb{1}_{\zeta(i)>0} + \mathbb{1}_{\sigma(i)=loop} - \mathbb{1}_{\sigma(i)=\zeta} \\ & + \left(\mathbb{1}_{\sigma(i)=Z} + \mathbb{1}_{\sigma(i)=multi}\right) \left(-\mathbb{1}_{\delta(i)=\zeta} + \mathbb{1}_{\delta(i)=Z} + 2 \cdot \mathbb{1}_{\delta(i)=Y}\right) \end{aligned} \tag{3.3}$$

and note that these all depend on whether $\zeta = 0$ (i.e. whether step i is of type 1 or 2). However,

$$\begin{aligned} \Delta A(i) = {} & -1 - \mathbb{1}_{\sigma(i)=Y} - \mathbb{1}_{\sigma(i)=\zeta} + \mathbb{1}_{\sigma(i)=loop} - \mathbb{1}_{\sigma(i)=multi} + \mathbb{1}_{\sigma(i)=Z} \\ & - \left(\mathbb{1}_{\sigma(i)=Z} + \mathbb{1}_{\sigma(i)=multi}\right) 2 \cdot \mathbb{1}_{\delta(i)=\zeta} \end{aligned} \tag{3.4}$$

$$\Delta B(i) = -2 + \mathbb{1}_{\sigma(i)=loop} - \mathbb{1}_{\delta(i)=\zeta} \tag{3.5}$$

which do not depend on whether $\zeta = 0$. For ΔA, we have used the identity

$$\mathbb{1}_{\delta=Y} + \mathbb{1}_{\delta=Z} + \mathbb{1}_{\delta=\zeta} = \mathbb{1}_{\sigma=Z} + \mathbb{1}_{\sigma=multi}$$

which states that we make a deletion move if and only if our selection move was Z or $multi$. Note also that if we establish dynamic concentration on A, B, ζ then we implicitly establish concentration on Y, Z, M since

$$Y = A - \zeta \tag{3.6}$$
$$Z = B - 2A + \zeta \tag{3.7}$$
$$2M = 3Y + 2Z + \zeta = 2B - A. \tag{3.8}$$

4 The expected behavior of A, B, ζ

In this section, we we will non-rigorously predict the behavior of the variables and some facts about the process. Throughout the paper, unless otherwise specified, t refers to the scaled version of i, so

$$t := \frac{i}{n}.$$

Heuristically, we assume there exist differentiable functions a, b such that $A(i) \approx na(t), B(i) \approx nb(t)$. Further, we assume that ζ stays "small". We will prove that these assumptions are indeed valid. We also let

$$p_z := \frac{2Z}{2M}, \quad p_y := \frac{3Y}{2M}, \quad p_\zeta := \frac{\zeta}{2M}$$

where we have omitted the dependence on i for ease of notation.

4.1 The trajectory $b(t)$

Since $B(0) = 2n$, and recalling (3.5), we see that

$$B(i) = 2n - 2i + \sum_{j \le i} \left(1_{\sigma(j)=loop} - 1_{\delta(j)=\zeta}\right). \tag{4.1}$$

The probability that $\sigma(j) = loop$ or $\delta(j) = \zeta$ on any step j should be negligible. Thus we expect

$$B(i) \approx 2n - 2i = 2n(1 - t)$$

so we will set

$$b(t) = 2(1 - t).$$

4.2 The trajectory $a(t)$

We derive an ODE that a should satisfy:

$$a'(t) \approx E[\Delta A(i)] \approx -1 - p_y + p_z \approx -\frac{6a(t)}{2b(t) - a(t)}.$$

Note that we have used (3.4). Thus $a(t)$ should satisfy

$$a' = -\frac{6a}{4-4t-a}. \tag{4.2}$$

The substitution $x = \frac{a}{1-t}$ yields a separable ODE, which can be integrated to arrive at

$$(a+2-2t)^3 - 27a^2 = 0.$$

So $a(t)$ is given implicitly as the solution to the above cubic equation with coefficients depending on t. We may actually solve that cubic to get three continuous explicit functions $a_1(t), a_2(t), a_3(t)$ (though the formulas are nasty to look at). From the initial condition $a(0) = 1$ and the fact that $0 \le a(t) \le 1$, it's clear that the solution we want is

$$a(t) = 7 + 2t - 6\sqrt{5+4t}\cos\left(\frac{1}{3}\arccos\left(\frac{11+14t+2t^2}{(5+4t)^{\frac{3}{2}}}\right) + \frac{\pi}{3}\right).$$

From here we can see that $a(t) \to 0$ as $t \to 1^-$. More precisely,

$$\lim_{t\to 1^-} \frac{a(t)}{(1-t)^{\frac{3}{2}}} = \left(\frac{2}{3}\right)^{\frac{3}{2}}. \tag{4.3}$$

To confirm this, note that

$$\arccos(1-\varepsilon) = \sqrt{2\varepsilon} + O(\varepsilon^{3/2})$$

and

$$\frac{11+14(1-\varepsilon)+2(1-\varepsilon)^2}{(5+4(1-\varepsilon))^{3/2}} = 1 - \frac{4\varepsilon^3}{729} + O(\varepsilon^4).$$

Rewriting the cos term using the angle addition formula and Taylor expansion, we see (4.3). Additionally,

$$\frac{d}{dt}\left(\frac{a(t)}{(1-t)^{3/2}}\right) = -\frac{6a}{2b-a}(1-t)^{-3/2} + \frac{3}{2}a\cdot(1-t)^{-5/2} \tag{4.4}$$

$$= a\cdot(1-t)^{-5/2}\left(\frac{3}{2} - \frac{6(1-t)}{4(1-t)-a}\right) \tag{4.5}$$

$$< 0. \tag{4.6}$$

Since $a(0) = 1$, for all $0 \le t \le 1$ we have

$$\left(\frac{2}{3}\right)^{\frac{3}{2}}(1-t)^{3/2} \le a(t) \le (1-t)^{3/2}. \tag{4.7}$$

4.3 Downward drift of ζ

We expect ζ to be "small", and to justify that claim we will show that whenever ζ is positive, it is likely to decrease. Assume that $\zeta(i) > 0$. In the following table, we make use of the fact that $\delta(i) \ne \emptyset$ if and only if $\sigma(i) \in \{Z, multi\}$. So for example,

$$\mathbb{1}_{\delta=Y} = (\mathbb{1}_{\sigma=Z} + \mathbb{1}_{\sigma=multi})\mathbb{1}_{\delta=Y}.$$

Then from (3.3) we see that if $\zeta(i) > 0$,

$$\Delta\zeta = \begin{cases} 1 & \text{with prob. } p_z p_y + O\left(\frac{1}{M}\right) \\ 0 & \text{with prob. } p_z^2 + O\left(\frac{1}{M}\right) \\ -1 & \text{with prob. } p_y + O\left(\frac{1}{M}\right) \\ -2 & \text{with prob. } p_\zeta + p_z p_\zeta + O\left(\frac{1}{M}\right) \end{cases} \tag{4.8}$$

To illuminate this table, we provide an example.

$$\mathbb{P}\left[\Delta\zeta(i) = 1\right] = \mathbb{P}\left[\sigma(i) = Z,\, \delta(i) = Y\right] = \frac{2Z}{2M-1}\frac{3Y}{2M-3} = p_z p_y + O\left(\frac{1}{M}\right).$$

Therefore, roughly speaking we have

$$E[\Delta\zeta(i)] = p_z p_y - p_y + O\left(p_\zeta\right) \approx -\frac{9a^2}{(2b-a)^2} \tag{4.9}$$

and are motivated to define

$$\Phi(t) := \frac{9a^2}{(2b-a)^2} = \Theta(1-t)$$

to represent the downward drift of $\zeta(i)$ (if it is positive) at step i.

4.4 Expected behavior of ζ

In the last subsection we estimated $E[\Delta\zeta(i)]$ when $\zeta > 0$, using (4.8). We can also use (4.8) to estimate the variance when $\zeta > 0$. We see that

$$Var[\Delta\zeta(i)|\zeta > 0] = \Theta(p_y) = \Theta\left((1-t)^{\frac{1}{2}}\right).$$

Thus, to model the behavior of $\zeta(i)$ we consider a simpler variable: a lazy random walk $X_\tau(k)$ with $X_\tau(0) = 0$, expected 1-step change $\mathbb{E}\left[\Delta X_\tau\right] = -(1-\tau)$ and $Var[\Delta X_\tau] = (1-\tau)^{\frac{1}{2}}$. After s steps, we have $\mathbb{E}\left[X_\tau(s)\right] = -(1-\tau)s$ and $Var[X_\tau(s)] = (1-\tau)^{\frac{1}{2}}s$. There is at least constant (bounded away from 0) probability that $X_\tau(s)$ is, say, 1 standard deviation above its mean. However, the probability that $X_\tau(s)$ is very many standard deviations larger than that is negligible. In other words, it is reasonable to have a displacement as large as $X_\tau(s) = -(1-\tau)s + (1-\tau)^{\frac{1}{4}}s^{\frac{1}{2}}$, but not much larger. The quantity $\psi(s) := -(1-\tau)s + (1-\tau)^{\frac{1}{4}}s^{\frac{1}{2}}$ is negative for $s > (1-\tau)^{-\frac{3}{2}}$. Also $\psi(s)$ is maximized when $s = \frac{1}{2}(1-\tau)^{-\frac{3}{2}}$, where we have $\psi(s) = \frac{1}{4}(1-\tau)^{-\frac{1}{2}}$.

Now we reconsider the variable ζ. Roughly speaking, $\zeta(i)$ behaves like the lazy random walk considered above, so long as we restrict the variable i to a short range (so that t does not change significantly), and we have $\zeta(i) > 0$ for this range of i. We have $\zeta(0) = 0$, and ζ has a negative drift so it's likely that $\zeta(j) = 0$ for many $j > 0$. Specifically, if j is an index such that $\zeta(j) = 0$, then we expect $\zeta(i)$ to behave like $X_\tau(i-j)$ with $\tau = \frac{j}{n}$, so long as i is not significantly larger than j. Thus we expect to have $\zeta(i) = 0$ for some $j \le i \le j + (1-\tau)^{-\frac{3}{2}}$. Also, for all $j \le i \le j + (1-\tau)^{-\frac{3}{2}}$ we should have $\zeta(i) \le \frac{1}{4}(1-\tau)^{-\frac{1}{2}}$. But this rough analysis does not make sense toward the end of the process: indeed, for $j > n - n^{\frac{3}{5}}$ (i.e. for $1 - \tau < n^{-\frac{2}{5}}$), we have $j + (1-t)^{-\frac{3}{2}} > n$. However, we can still say something about what happens when j is large, since the variable s cannot be any bigger than $n - j$. Now for $j \ge n - n^{\frac{3}{5}}$ and $s \le n - j$ we have $\psi(s) \le n^{\frac{1}{5}}$. Thus, we never expect ζ to be larger than $n^{\frac{1}{5}}$, even towards the end of the process.

4.5 Why do we have $\tilde{\Theta}\left(n^{\frac{1}{5}}\right)$ many components?

At any step of the algorithm, we expect the components of the 2-matching to be mostly paths (and a few cycles). We would like the algorithm to keep making the paths longer, but sometimes it isn't possible to make a path any longer because of deletion moves. Specifically, for example, if one endpoint of a path is in Z_1, and then there is a deletion move and the deletion is that endpoint, then that end of the path will never grow. If the same thing happens to the other endpoint of the path, then the path will never get longer, and will never be connected to any of the other paths. Similarly, the number of components in the final 2-matching is increased whenever the algorithm deletes a Y_1 or a Y_2. Thus we can bound the number of components in the final 2-matching by bounding the number of steps i such that $\delta(i) = \zeta$.

Roughly, $\mathbb{P}[\delta(i) = \zeta] = \frac{2Z}{2M} \cdot \frac{\zeta}{2M} = \Theta\left(\frac{1}{n}\min\left\{(1-t)^{-\frac{3}{2}}, \frac{n^{\frac{1}{5}}}{1-t}\right\}\right)$. So integrating, we estimate the total number of components as

$$\Theta\left(\int_0^{1-\frac{1}{n}} \min\left\{(1-t)^{-\frac{3}{2}}, \frac{n^{\frac{1}{5}}}{1-t}\right\} dt\right) = \Theta\left(n^{\frac{1}{5}} \log n\right).$$

5 The stopping time T and dynamic concentration

In this section, we introduce a stopping time T, before which A and B stay close to their trajectories, and ζ does roughly what we expect it to do. We will also introduce "error" terms for both A, B and a "correction" term α for the variable A. For most of the process, α will stay smaller than the error term for A. However, toward the end of the process α will be significant. Using α in our calculations thus allows us to track the process farther. As it turns out, the variable B does not need an analogous "correction" term.

We define the following random variables which represent "actual error" in A, B:

$$e_a(i) := A(i) - na(t) - \alpha(i)$$

$$e_b(i) := B(i) - nb(t).$$

We define the stopping time T as the minimum of $n - C_T n^{\frac{7}{15}} \log^{\frac{6}{5}} n$ and the first step i such that any of the three following conditions fail:

$$|e_a(i)| \le f_a(t), \tag{5.1}$$

$$|e_b(i)| \le f_b(t), \tag{5.2}$$

and for every step $j < i$ such that ζ is positive on steps $j, \ldots, i$,

$$\zeta(i) \le \zeta(j) - \sum_{j \le k < i} \Phi\left(\frac{k}{n}\right) + \ell_j(t) \tag{5.3}$$

for some as-yet unspecified error functions f_a, f_b, ℓ_j and absolute constant C_T. Throughout the paper we will use $C_?$ to refer to unspecified but existent absolute constants. In subsection 5.6, we present actual values for these constants.

We anticipate that the conditions on ζ will imply that for some function f_ζ we have

$$\zeta(i) \le f_\zeta(t)$$

for all $i \le T$. Our goal for now is to prove that for some suitable error functions, w.h.p. T is not triggered by any of the conditions (5.1), (5.2), (5.3).

Theorem 5.1. *With high probability,*

$$T = n - C_T n^{\frac{7}{15}} \log^{\frac{6}{5}} n.$$

The remainder of this section contains the proof of Theorem 5.1. Here we define the error functions f_a, f_b, f_ζ (up to the choice of constants). While these definitions are not very enlightening at this point, they will aid the reader in confirming many of the calculations that appear below. Those same calculations will motivate the choice of these functions.

$$f_a(t) := C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n \tag{5.4}$$

$$f_b(t) := C_B \cdot \begin{cases} (1-t)^{-\frac{1}{2}} \log n & : 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n \\ -n^{\frac{1}{5}} \log^{\frac{4}{5}} n \log(1-t) & : \text{otherwise} \end{cases} \tag{5.5}$$

$$f_\zeta(t) := C_\zeta \min\left\{(1-t)^{-\frac{1}{2}} \log n, n^{\frac{1}{5}} \log^{\frac{4}{5}} n\right\}. \tag{5.6}$$

5.1 A useful lemma

We'll use the following simple lemma several times to estimate fractions.

Lemma 5.2. *For any real numbers $x, y, \varepsilon_x, \varepsilon_y$, if we have $x, y \ne 0$ and $\left|\frac{\varepsilon_x}{x}\right|, \left|\frac{\varepsilon_y}{y}\right| \le \frac{1}{2}$, then*

$$\frac{x+\varepsilon_x}{y+\varepsilon_y} - \frac{x}{y} = \frac{y\varepsilon_x - x\varepsilon_y}{y^2} + O\left(\frac{y\varepsilon_x\varepsilon_y + x\varepsilon_y^2}{y^3}\right)$$

Proof.

$$\begin{aligned}
\frac{x+\varepsilon_x}{y+\varepsilon_y} - \frac{x}{y} &= \frac{x}{y}\left\{\left(1+\frac{\varepsilon_x}{x}\right)\cdot\frac{1}{1+\frac{\varepsilon_y}{y}} - 1\right\} \\
&= \frac{x}{y}\left\{\left(1+\frac{\varepsilon_x}{x}\right)\cdot\left[1-\frac{\varepsilon_y}{y} + O\left(\frac{\varepsilon_y^2}{y^2}\right)\right] - 1\right\} \\
&= \frac{x}{y}\left\{\frac{\varepsilon_x}{x} - \frac{\varepsilon_y}{y} + O\left(\frac{\varepsilon_x\varepsilon_y}{xy} + \frac{\varepsilon_y^2}{y^2}\right)\right\} \\
&= \frac{y\varepsilon_x - x\varepsilon_y}{y^2} + O\left(\frac{y\varepsilon_x\varepsilon_y + x\varepsilon_y^2}{y^3}\right)
\end{aligned}$$

$\square$

5.2 T is not triggered by A

We define

$$A^+(i) := A(i) - na(t) - \alpha(i) - f_a(t) = e_a(i) - f_a(t)$$

and let the stopping time T_j be the maximum of j, T, and the least index $i \geq j$ such that $e_a(i)$ is not in the critical interval

$$[g_a(t), f_a(t)] \tag{5.7}$$

where $0 < g_a < f_a$ is an as-yet unspecified function of n, t. Our strategy is to show that w.h.p. A never goes above $na + \alpha + f_a$ because every time e_a enters the critical interval, w.h.p. it does not exit the interval at the top. The use of critical intervals in a similar context was first introduced in [5].

Let $\mathcal{F}_i$ be the natural filtration of the process (so conditioning on $\mathcal{F}_i$ tells us the values of all the variables, among other things).

For $i < T$, we have from (3.4) that

$$\begin{aligned}
&E[\Delta A(i)|\mathcal{F}_i] = -1 - \frac{3Y}{2M} - \frac{\zeta}{2M} + \frac{2Z}{2M} - 2 \cdot \frac{2Z}{2M} \cdot \frac{\zeta}{2M} + O\left(\frac{1}{M}\right)\\
&= -\frac{6A}{2B-A} + \frac{4\zeta(A+B)}{(2B-A)^2} + O\left(\frac{1}{2B-A} + \frac{\zeta^2}{(2B-A)^2}\right)\\
&= -\frac{6\left(na+\alpha+e_a\right)}{2(nb+e_b)-(na+\alpha+e_a)} + \frac{4\zeta\left[(na+\alpha+e_a)+(nb+e_b)\right]}{\left[2(nb+e_b)-(na+\alpha+e_a)\right]^2} + O\left(\frac{1}{2B-A} + \frac{\zeta^2}{(2B-A)^2}\right)\\
&= -\frac{6a}{2b-a} + \frac{12ae_b - 12b\left(\alpha+e_a\right)}{n(2b-a)^2} + \frac{4(a+b)\zeta}{n(2b-a)^2} + O\left(\frac{1}{n(2b-a)} + \frac{\alpha^2+f_a^2+f_b^2+f_\zeta^2}{n^2(2b-a)^2}\right)
\end{aligned}$$

The last equality follows from Lemma 5.2. Note that the lemma actually implies that the big-O term includes mixed products of terms like $\alpha \cdot f_\zeta$ for example. We have simplified by using the fact that for all real numbers x and y, $|xy| \leq \frac{1}{2}\left(x^2+y^2\right)$. We are now motivated to cancel out the ζ term in the last line by recursively defining

$$\alpha(0) := 0 \tag{5.8}$$

$$\alpha(i+1) := \alpha(i) + \frac{4(a+b)\zeta - 12b\alpha(i)}{n(2b-a)^2}. \tag{5.9}$$

From this definition and the defintion of f_ζ, it follows that for $i \leq T$,

$$0 \leq \alpha(i) \leq \sum_{j=0}^{i} \frac{4(a+b)f_\zeta}{n(2b-a)^2} \leq C_\alpha \cdot \begin{cases} \log n(1-t)^{-1/2} & \text{for } i \leq n - n^{3/5}\log^{2/5} n \\ n^{1/5}\log^{9/5} n & \text{for } n - n^{3/5}\log^{2/5} n < i \leq T. \end{cases} \tag{5.10}$$

as long as

$$C_\alpha > 8C_\zeta. \tag{5.11}$$

Now for $j \le i < T_j$, we have the supermartingale condition

$$\begin{aligned} E[\Delta A^+(i)|\mathcal{F}_i] &= E[\Delta A(i)|\mathcal{F}_i] - a'(t) - \frac{4(a+b)\zeta - 12b\alpha(i)}{n(2b-a)^2} - \frac{1}{n}f_a'(t) \\ &\quad + O\left(\frac{1}{n}a''(t) + \frac{1}{n^2}f_a''(t)\right) \qquad (5.12) \\ &\le -\frac{12bg_a}{n(2b-a)^2} - \frac{1}{n}f_a'(t) \\ &\quad + O\left(\frac{af_b}{n(2b-a)^2} + \frac{1}{n(2b-a)} + \frac{\alpha^2 + f_a^2 + f_b^2 + f_\zeta^2}{n^2(2b-a)^2} + \frac{1}{n}a''(t) + \frac{1}{n^2}f_a''(t)\right) \qquad (5.13) \end{aligned}$$

Note that in the last line we have used (5.9), the fact that $e_a \ge g_a$, and also that a satisfies the differential equation (4.2). By taking $g_a = \Omega(f_a)$, we see that $A^+(j), \ldots, A^+(T_j)$ is a supermartingale since

$$-\frac{12bg_a}{n(2b-a)^2} - \frac{1}{n}f_a' = -\Omega\left(n^{-1/2}\log^{1/2} n(1-t)^{-1/4}\right)$$

which dominates the big-O term in (5.13).

We use the following asymmetric version of the Azuma-Hoeffding inequality (for a proof see [2]):

Lemma 5.3. *Let X_j be a supermartingale, such that $-C \le \Delta X(j) \le c$ for all j, for $c < \frac{C}{10}$. Then for any $a < cm$ we have $Pr(X_m - X_0 > a) \le \exp\left(-\frac{a^2}{3cCm}\right)$*

We have

$$-2 \le \Delta A \le 0$$

and

$$-2\,(1-t)^{\frac{1}{2}} \le a'(t) \le 0.$$

This follows from analysis of the function $a(t)$. So

$$-2 \le \Delta A^+ \le 2\left(1 - \frac{j}{n}\right)^{\frac{1}{2}}$$

for the supermartingale $A^+(j) \cdots A^+(T_j)$. Thus, if A crosses its upper boundary at the stopping time T, then there is some step j (with $T = T_j$) such that

$$A^+(j) \le g_a\left(\frac{j}{n}\right) - f_a\left(\frac{j}{n}\right) + 2$$

and $A^+(T_j) > 0$. In this case, j is intended to represent the step when e_a enters the crtical interval, (5.7). Applying the lemma we see that the probability of the supermartingale A^+ having such a large upward deviation has probability at most

$$\exp\left\{-\frac{\left(f_a\left(\frac{j}{n}\right) - g_a\left(\frac{j}{n}\right) - 2\right)^2}{12n\left(1 - \frac{j}{n}\right)^{\frac{3}{2}}}\right\}.$$

As there are $O(n)$ supermartingales $A^+(j),\ldots,A^+(T_j)$, we must choose f_a, g_a to make the above probability $o\left(\frac{1}{n}\right)$. The following choice suffices:

$$f_a(t) = C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n$$

$$g_a(t) = \frac{3}{4} f_a(t).$$

as long as the constant C_A is chosen so that

$$\frac{\left(\frac{1}{4}C_A\right)^2}{12} > 1. \tag{5.14}$$

If we define

$$A^- := A - na - \alpha + f_a = e_a + f_a$$

then we may prove that A^- stays positive w.h.p. in a completely analogous fashion.

5.3 T is not triggered by ζ

Referring to (4.8), we may say that if $\zeta(i) > 0$,

$$E[\Delta\zeta(i)|\mathcal{F}_i] = p_z p_y - p_y + O(p_\zeta) = -\frac{9A^2}{(2B-A)^2} + O\left(\frac{\zeta}{2B-A}\right). \tag{5.15}$$

Now, before T we have

$$\begin{aligned}
\frac{9a^2}{(2b-a)^2} - \frac{9A^2}{(2B-A)^2} &= -9\left(\frac{A}{2B-A} - \frac{a}{2b-a}\right)\left(\frac{A}{2B-A} + \frac{a}{2b-a}\right) \\
&= -9\left[\frac{2b(\alpha+e_a) - 2ae_b}{n(2b-a)^2} + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2}\right)\right] \\
&\quad \times \left[2\left(\frac{a}{2b-a}\right) + \frac{2b(\alpha+e_a) - 2ae_b}{n(2b-a)^2} + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2}\right)\right] \\
&= \frac{36a(ae_b - b\alpha - be_a)}{n(2b-a)^3} + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2}\right).
\end{aligned} \tag{5.16}$$

In the last step we have cleaned up the big-O using the facts

$$\frac{\alpha + f_a + f_b}{n(2b-a)} = o(1) \qquad \text{and} \qquad \frac{a}{2b-a} = O(1).$$

For every step j, we define a stopping time

$$T_j := \min\{i(j), \max(j, T)\}$$

where $i(j)$ is the least index $i \geq j$ such that $\zeta(i) = 0$. Also, define a sequence $\zeta_j^+(j)\cdots\zeta_j^+(T_j)$, where

$$\zeta_j^+(i) := \zeta(i) + \sum_{j \leq k < i} \Phi\left(\frac{k}{n}\right) - h_j\left(\frac{i}{n}\right)$$

where h_j is some function we will choose that will make $\zeta_j^+(i)$ a supermartingale. Now for $j \le i < T_j$, using (5.16), we have

$$E[\Delta\zeta_j^+(i)|\mathcal{F}_i] = -\frac{9A^2}{(2B-A)^2} + \frac{9a^2}{(2b-a)^2} - \frac{1}{n}h_j'(t) + O\left(\frac{\zeta}{2B-A} + \frac{1}{n^2}h_j''(t)\right) \tag{5.17}$$

$$\le \frac{36a(af_b+bf_a)}{n(2b-a)^3} - \frac{1}{n}h_j'(t) + O\left(\frac{\alpha^2+f_a^2+f_b^2}{n^2(2b-a)^2} + \frac{f_\zeta}{n(2b-a)} + \frac{1}{n^2}h_j''(t)\right). \tag{5.18}$$

Note that

$$\frac{36a(af_b+bf_a)}{(2b-a)^3} \le \left(\frac{9}{8}C_A + o(1)\right) n^{\frac{1}{2}} \log^{\frac{1}{2}} n\,(1-t)^{\frac{1}{4}} \tag{5.19}$$

so the choice

$$h_j(t) := C_h\left(1-\frac{j}{n}\right)^{\frac{1}{4}} n^{\frac{1}{2}}\log^{\frac{1}{2}} n\left(t-\frac{j}{n}\right)$$

makes the sequence a supermartingale as long as the constant C_h is chosen so that

$$C_h > \frac{9}{8}C_A. \tag{5.20}$$

Since $h_j\left(\frac{j}{n}\right) = 0$, we will always have $\zeta_j^+(j) = \zeta(j)$.

We'll use the following supermartingale inequality due to Freedman [8]:

Lemma 5.4. *Let X_i be a supermartingale, with $\Delta X_i \le C$ for all i, and $V(i) := \sum_{k\le i} Var[\Delta X(k)|\mathcal{F}_k]$ Then*

$$P\left[\exists i : V(i) \le v, X_i - X_0 \ge d\right] \le \exp\left(-\frac{d^2}{2(v+Cd)}\right).$$

Referring to (4.8), before T we can put

$$\begin{aligned} Var[\Delta\zeta_j^+(i)|\mathcal{F}_i] &= Var[\Delta\zeta(i)|\mathcal{F}_i] \\ &\le E\left[(\Delta\zeta(i))^2 \mid \mathcal{F}_i\right] \\ &= 1\cdot p_z p_y + 1\cdot p_y + 4\cdot(p_\zeta + p_z p_\zeta) + O\left(\frac{1}{M}\right) \\ &\le 3p_y \end{aligned}$$

and note that before T, we have

$$p_y = \frac{3Y}{2M} \le \frac{3A}{2B-A} \le \frac{3[n(1-t)^{\frac{3}{2}}+\alpha+f_a]}{4n(1-t)-2f_b-n(1-t)^{\frac{3}{2}}-\alpha-f_a} \le \left(1+\frac{C_\alpha}{C_T^{\frac{3}{2}}}+o(1)\right)(1-t)^{\frac{1}{2}} \tag{5.21}$$

so we will just say $p_y \le C_{p_y}(1-t)^{\frac{1}{2}}$ for some constant C_{p_y} such that

$$C_{p_y} > 1 + \frac{C_\alpha}{C_T^{\frac{3}{2}}}. \tag{5.22}$$

Also, note that $\Delta\zeta^+ \le 2$.

Suppose the variable ζ triggers the stopping time T. Then there are steps $j < i = T$ such that $\zeta > 0$ all the way from step j to step i, and $\zeta_j^+(i) > \ell_j(t) - h_j(t)$. We'll need to apply the lemma to the supermartingale ζ_j^+ to show this event has low probability (guiding our choice for ℓ_j). Note that in the lemma we can plug in the following for v:

$$V(i) = \sum_{j \le k \le i} Var[\Delta\zeta_j^+(k)|\mathcal{F}_k] \le 3C_{p_y}\left(1 - \frac{j}{n}\right)^{\frac{1}{2}}(i-j).$$

So the unlikely event has probability at most

$$\exp\left\{-\frac{(\ell_j - h_j)^2}{2\left[3C_{p_y}\left(1-\frac{j}{n}\right)^{\frac{1}{2}}(i-j) + 2(\ell_j - h_j)\right]}\right\}.$$

As there are $O\left(n^2\right)$ pairs of steps j, i we'd like to make the above probability $o\left(\frac{1}{n^2}\right)$. Towards this end we consider 2 cases.

If $\left(1-\frac{j}{n}\right)^{\frac{1}{4}}(i-j)^{\frac{1}{2}} \le \log^{\frac{1}{2}} n$, then it suffices to put $\ell_j - h_j = C_\ell \log n$ as long as

$$\frac{C_\ell^2}{6C_{p_y} + 4C_\ell} > 2. \tag{5.23}$$

If $\left(1-\frac{j}{n}\right)^{\frac{1}{4}}(i-j)^{\frac{1}{2}} > \log^{\frac{1}{2}} n$, then it suffices to put $\ell_j - h_j = C_\ell\left(1-\frac{j}{n}\right)^{\frac{1}{4}}(i-j)^{\frac{1}{2}}\log^{\frac{1}{2}} n$. Thus we choose

$$\ell_j(t) := h_j(t) + C_\ell \max\left\{\log n, \left(1-\frac{j}{n}\right)^{\frac{1}{4}}(i-j)^{\frac{1}{2}}\log^{\frac{1}{2}} n\right\}.$$

With this choice, w.h.p. T is not triggered by ζ.

5.4 An upper bound on ζ

In this section we'll motivate our choice of the function f_ζ.

Lemma 5.5. *W.h.p. for all $j < n - 2C_x^{\frac{2}{5}} n^{\frac{3}{5}} \log^{\frac{2}{5}} n$ such that $\zeta(j-1) = 0$, we have*

1. *$\zeta(j') = 0$ for some $j \le j' \le j + C_x\left(1-\frac{j}{n}\right)^{-\frac{3}{2}} \log n$, and*
2. *$\zeta(i) \le 2C_\ell^2\left(1-\frac{j}{n}\right)^{-\frac{1}{2}} \log n$ for all $j \le i \le j'-1$*

Proof. Suppose $\zeta(j-1) = 0$. Note that we then have $\zeta(j) \le 2$. $\Phi(t)/(1-t)$ is decreasing since

$$\begin{aligned}\frac{d}{dt}\left(\frac{\Phi(t)}{1-t}\right) &= 2\left(\frac{3a}{2b-a}\right)\left(\frac{(2b-a)\cdot 3\left(-\frac{6a}{2b-a}\right) - 3a\left(-4+\frac{6a}{2b-a}\right)}{(2b-a)^2(1-t)}\right) + (1-t)^{-2}\left(\frac{3a}{2b-a}\right)^2 \\ &= -\frac{9a^3(8b-a)}{(1-t)^2(2b-a)^4} \le 0.\end{aligned}$$

Also, using (4.3) and the defintion of b,

$$\lim_{t \to 1^-} \frac{\Phi(t)}{1-t} = \frac{1}{6}.$$

Hence $\Phi(t) \geq \frac{1}{6}(1-t)$ for all $0 \leq t \leq 1$. If we substitute $x = \frac{i-j}{n}$ then

$$\sum_{j \leq k < i} \Phi\left(\frac{k}{n}\right) \geq \frac{1}{6}\left(1 - \frac{i+j-1}{2n}\right)(i-j) \geq -\frac{1}{12}nx^2 + \frac{1}{6}n\left(1-\frac{j}{n}\right)x.$$

Plugging in the value of $\ell_j(t)$, we have that for any $i \geq j$ such that $\zeta(j)\ldots\zeta(i)$ are all positive,

$$\begin{aligned} \zeta(i) &\leq \zeta(j) - \sum_{j \leq k < i} \Phi\left(\frac{k}{n}\right) + \ell_j(t) && (5.24) \\ &\leq \frac{1}{12}nx^2 - \left[\frac{1}{6}n\left(1-\frac{j}{n}\right) - C_h n^{\frac{1}{2}} \log^{\frac{1}{2}} n \left(1-\frac{j}{n}\right)^{\frac{1}{4}}\right] x && (5.25) \\ &\quad + C_\ell \max\left\{\log n, \left(1-\frac{j}{n}\right)^{\frac{1}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n x^{\frac{1}{2}}\right\} + 2. \end{aligned}$$

Consider (5.25) for $x = x_j := C_x n^{-1} \log n \left(1 - \frac{j}{n}\right)^{-\frac{3}{2}}$. As long as $C_x > 1$ and $j < n - 2C_x^{\frac{2}{5}} n^{\frac{3}{5}} \log^{\frac{2}{5}} n$, we have

$$\left(1-\frac{j}{n}\right)^{\frac{1}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n x_j^{\frac{1}{2}} > \log n$$

so we can evaluate the "max" in ℓ_j. Also note that the coefficient of x is dominated by $-\frac{1}{6}n\left(1-\frac{j}{n}\right)$, so the coefficient of x is at most, say $-\frac{1}{7}n\left(1-\frac{j}{n}\right)$. Thus (5.25) gives

$$\zeta(j + nx_j) \leq \frac{C_x^2}{12} n^{-1} \log^2 n \left(1-\frac{j}{n}\right)^{-3} - \left(\frac{C_x}{7} - C_\ell\sqrt{C_x}\right) \log n \left(1-\frac{j}{n}\right)^{-\frac{1}{2}} + 2$$

which is negative for this range of j as long as we pick C_x such that

$$\frac{C_x}{7} - C_\ell\sqrt{C_x} > 0. \tag{5.26}$$

Therefore, ζ must have hit 0 again before step $i = j + nx_j$. This proves the first part of the lemma.

To prove the second part, consider (5.25) for $j < i < j + nx_j$ (i.e. for $0 < x < x_j$). If $x \leq n^{-1} \log n \left(1-\frac{j}{n}\right)^{-\frac{1}{2}}$ then we can put

$$\zeta(i) \leq \frac{1}{12}nx^2 - \frac{1}{7}n\left(1-\frac{j}{n}\right)x + C_\ell \log n < 2C_\ell \log n$$

and for x larger than that, we'll put

$$\begin{aligned} \zeta(i) &\leq \frac{1}{12}nx^2 - \frac{1}{7}n\left(1-\frac{j}{n}\right)x + C_\ell\left(1-\frac{j}{n}\right)^{\frac{1}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n x^{\frac{1}{2}} \\ &\leq \frac{C_x^2}{12} n^{-1} \log^2 n \left(1-\frac{j}{n}\right)^{-3} + \frac{7C_\ell^2}{4}\left(1-\frac{j}{n}\right)^{-\frac{1}{2}} \log n \\ &< 2C_\ell^2 \left(1-\frac{j}{n}\right)^{-\frac{1}{2}} \log n. \end{aligned}$$

where to justify the second line we use the inequality $c\sqrt{x} - dx \le \frac{c^2}{4d}$ for real numbers $x, c, d > 0$. $\square$

We would also like to say something about $\zeta(i)$ for $i > n - 2C_x^{\frac{2}{5}} n^{\frac{3}{5}} \log^{\frac{2}{5}} n$.

Lemma 5.6. *There exists a constant C_ζ such that w.h.p. for all $i \le T$ we have $\zeta(i) \le C_\zeta n^{\frac{1}{5}} \log^{\frac{4}{5}} n$.*

Proof. Suppose step $j' \ge n - 2C_x^{\frac{2}{5}} n^{\frac{3}{5}} \log^{\frac{2}{5}} n$ with $\zeta(j') = 0$. It follows from Lemma 5.5 that w.h.p. such a j' exists. Let $i \ge j'$ such that $\zeta(j') \ldots \zeta(i)$ are all positive. Note that we again have the bound (5.25). But now $0 \le x \le \frac{n-j'}{n} \le 2C_x^{\frac{2}{5}} n^{-\frac{2}{5}} \log^{\frac{2}{5}} n$, and (5.25) gives $\zeta(i) \le \left(\frac{1}{3} C_x^{\frac{4}{5}} + 2^{\frac{3}{4}} C_\ell C_x^{\frac{3}{10}}\right) n^{\frac{1}{5}} \log^{\frac{4}{5}} n$. $\square$

So in particular we can say that for $i \le T$ we have

$$\zeta(i) \le f_\zeta(t) = C_\zeta \min\left\{(1-t)^{-\frac{1}{2}} \log n, n^{\frac{1}{5}} \log^{\frac{4}{5}} n\right\},$$

where

$$C_\zeta > \max\left\{2C_\ell^2, \frac{1}{3} C_x^{\frac{4}{5}} + 2^{\frac{3}{4}} C_\ell C_x^{\frac{3}{10}}\right\} \tag{5.27}$$

5.5 T is not triggered by B

Recall from (4.1) that

$$e_b(i) = \sum_{j \le i} \left(1_{\sigma(i)=loop} - 1_{\delta(j)=\zeta}\right).$$

First we'll bound $\sum_{j \le i} 1_{\delta(j)=\zeta}$. Define $B^-(i) := -\sum_{j \le i} 1_{\delta(j)=\zeta} + \frac{1}{2} f_b(t)$. Then

$$\begin{aligned} E[\Delta B^-(i) | \mathcal{F}_i] &= -\frac{2Z}{2M} \cdot \frac{\zeta}{2M-2} + \frac{1}{2n} f_b'(t) + O\left(\frac{1}{n^2} f_b''(t)\right) \\ &\ge -\frac{f_\zeta}{n(2b-a)} + \frac{1}{2n} f_b'(t) + O\left(\frac{1}{n^2} f_b''(t)\right). \end{aligned}$$

Note that by (4.7),

$$3(1-t) \le 2b - a \le 4(1-t),$$

so we can put

$$f_b = C_B \cdot \begin{cases} (1-t)^{-\frac{1}{2}} \log n & : 1 - t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n \\ -n^{\frac{1}{5}} \log^{\frac{4}{5}} n \log(1-t) & : \text{otherwise} \end{cases} \tag{5.28}$$

and B^- will be a submartingale as long as

$$C_B > \frac{4}{3} C_\zeta. \tag{5.29}$$

We'll apply Lemma 5.4 to $-B^-$. Note that before T we can put

$$\begin{aligned} Var[\Delta B^-(i)|\mathcal{F}_i] &= Var[1_{\delta(i)=\zeta}|\mathcal{F}_i] \\ &\leq p_\zeta \leq \frac{f_\zeta}{4n(1-t)-na-\alpha-f_a-f_b} \leq \frac{f_\zeta}{3n(1-t)} \end{aligned}$$

and therefore, referring to $V(i)$ as in Lemma 5.4,

$$V(i) \leq \sum_{0\leq k\leq i} \frac{f_\zeta}{3n(1-t)}.$$

So for v we will plug in

$$v = C_{v_B} \cdot \begin{cases} (1-t)^{-\frac{1}{2}} \log n & : 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n \\ n^{\frac{1}{5}} \log^{\frac{9}{5}} n & : \text{otherwise.} \end{cases} \tag{5.30}$$

which is an upper bound on $V(i)$ as long as

$$C_{v_B} \geq \frac{2}{3} C_\zeta \tag{5.31}$$

Note $|\Delta B^-| \leq 1$, so the probability that $-B^-(i) > \frac{1}{2} f_b(t)$ is at most

$$\exp\left\{ -\frac{\frac{1}{4} f_b^2}{2\left[v + \frac{1}{2} f_b\right]} \right\}$$

which is $o\left(\frac{1}{n}\right)$ as long as

$$\frac{\frac{1}{4} C_B^2}{2[C_{v_B} + \frac{1}{2} C_B]} > 1. \tag{5.32}$$

So w.h.p. for all $i \leq T$, we have

$$\sum_{j<i} 1_{\delta(j)=\zeta} \leq f_b(t).$$

The sum $\sum_{j<i} 1_{\sigma(j)=loop}$ presents less difficulty, since w.h.p. the configuration has at most $C_B \log n$ loops total. So we can trivially say that

$$\sum_{j<i} 1_{\sigma(j)=loop} \leq f_b(t)$$

and hence w.h.p. the stopping time T is not triggered by variable B.

5.6 Values for the constants

Throughout the proof above, we collect various constraints on the constants in (5.11), (5.14), (5.20), (5.22), (5.23), (5.26), (5.27), (5.29), (5.31) and (5.32). The reader may chack that the following values satisfy all the conditions.

$$C_A = 16, \quad C_h = 20, \quad C_{p_y} = 2, \quad C_\ell = 12, \quad C_x = 8000,$$

$$C_\zeta = 800, \quad C_\alpha = 70000, \quad C_{v_B} = 700, \quad C_B = 1200, \quad C_T = 2000.$$

This completes the proof of Theorem 5.1.

6 Upper bound on the number of components

In this section we prove the following lemma which provides the upper bound for the proof of Theorem 1.1:

Lemma 6.1. *W.h.p. the algorithm outputs a 2-matching with* $O\left(n^{\frac{1}{5}}\log^{\frac{9}{5}} n\right)$ *components.*

Proof. The components of our 2-matching at any step i consist of cycles and paths (including paths of length 0). First we'll bound the number of paths in the final 2-matching. Note that these final paths have both endpoints in Z_0, meaning that each endpoint had a half-edge in ζ that got deleted (or for paths of length 0 there is only one vertex which is in Y_0). So to bound the number of these paths, we bound the sum $\sum_j 1_{\delta(j)=\zeta}$. Note that in light of Section 5.5, we have the bound

$$\sum_{j<T} 1_{\delta(j)=\zeta} = O\left(n^{\frac{1}{5}}\log^{\frac{9}{5}} n\right).$$

Next we'll bound the terms corresponding to steps after T, but before $A=0$. By Theorem 5.1 we have w.h.p.

$$A(T) = O\left(n^{\frac{1}{5}}\log^{\frac{9}{5}} n\right)$$

since

$$0 \le \alpha(T) = O\left(n^{\frac{1}{5}}\log^{\frac{9}{5}} n\right)$$

by (5.10), and

$$na\left(\frac{T}{n}\right), f_a\left(\frac{T}{n}\right) = O\left(n^{1/5}\log^{9/5} n\right).$$

Now note that by (3.4), on each step j such that $\sigma(j) \in \{Z, multi\}$ and $\delta(j)=\zeta$, the variable A decreases by 2. Also, the variable A is nonincreasing. Therefore there can be at most $O\left(n^{\frac{1}{5}}\log^{\frac{9}{5}} n\right)$ such steps j until $A=0$.

Once we have $A=0$, the algorithm finds a maximum matching on the remaining random 2-regular graph Γ. Thus, to complete the bound on the number of paths in the final 2-matching, we'll bound the number of vertices in Γ that are unsaturated by the matching (i.e. the number of odd cycles in the remaining 2-regular graph Γ). But Γ has at most $O(\log n)$ cycles total, since it's a random 2-regular graph. Thus, the sum $\sum_j 1_{\delta(j)=\zeta}$, and therefore the number of paths in the final 2-matching, are $O\left(n^{\frac{1}{5}}\log^{\frac{9}{5}} n\right)$.

Now we bound the number of cycles in the final 2-matching. Note that at any step, the probability of closing a cycle is at most $\frac{1}{2M-1}$. Therefore, the number of cycles created for the whole process is stochastically dominated by the random variable

$$C := \sum_{j=1}^{3n} c_j$$

where

$$c_j = \begin{cases} 1 & : \text{with prob. } \frac{1}{j} \\ 0 & : \text{with prob. } \frac{j-1}{j}. \end{cases} \tag{6.1}$$

So if we define the martingale

$$C(i) := \sum_{j=1}^{i} \left(c_j - \frac{1}{j} \right)$$

then we have $Var[\Delta C(i)] = \frac{i-1}{i^2}$, and note $\sum_{i=1}^{3n} \frac{i-1}{i^2} = O(\log(3n))$. Now, applying Lemma 5.4 to $C(i)$ shows that w.h.p. it is always at most $O(\log^{\frac{1}{2}} n)$, and since $E[C] = O(\log n)$, we have that $C = O(\log n)$ w.h.p..

□

7 Lower bound on the number of components

In this section we will prove that near the end of the process, there is a non-zero probability that ζ becomes large and stays large for a significant amount of time. In this case, the algorithm will likely delete an edge adjacent to a ζ vertex. In particular, we will prove the following lemma which provides the lower bound and thus completes the proof of Theorem 1.1:

Lemma 7.1. *W.h.p. the algorithm outputs a 2-matching with* $\Omega\left(n^{\frac{1}{5}} \log^{-4} n\right)$ *components.*

Proof. We show that ζ stochastically dominates a suitably defined martingale and then apply the following central limit theorem of Freedman.

Lemma 7.2. *Let S_i be a martingale adapted to the filtration $\mathcal{F}_i$ with $X_i := S_i - S_{i-1}$, $|X_i| \le C$ for some constant C, and let $V_i := \sum_{k \le i} Var\left[X_k | \mathcal{F}_{k-1}\right]$. For each n, let $0 < \gamma_n < \gamma_n'$ be real numbers, and let σ_n be a stopping time. As $n \to \infty$, suppose $\gamma_n \to \infty$ and $\gamma_n'/\gamma_n \to 1$ and $\mathbb{P}\left[\gamma_n < V_{\sigma_n} < \gamma_n'\right] \to 1$. Then $S_{\sigma_n}/\sqrt{\gamma_n}$ converges in distribution to $\mathcal{N}(0,1)$.*

Let

$$w(i) = \frac{3a(i/n)}{2b(i/n) - a(i/n)}.$$

In this section we will consider steps from $i_0 = n - n^{3/5}$ to $i_{end} = n - n^{3/5} + n^{3/5} \log^{-1} n \le n - \frac{1}{2} n^{3/5}$. From Theorem 5.1, w.h.p., T occurs after this time frame. Hence we have dynamic concentration on our variables and can say in this range,

$$p_y(i) = w(i) - O(n^{-2/5} \log n) \tag{7.1}$$

$$p_\zeta(i) = O(n^{-2/5} \log n) \tag{7.2}$$

$$p_z(i) = 1 - p_y(i) - p_\zeta(i). \tag{7.3}$$

Note that in this range we also have $w(i) = \Theta(n^{-1/5})$. Our martingale will have independent increments given by

$$X(i) = \begin{cases} 1 & \text{with prob. } w(i) - Ln^{-2/5} \log n \\ 0 & \text{with prob. } 1 - 2w(i) - Ln^{-2/5} \log n \\ -1 & \text{with prob. } w(i) - Ln^{-2/5} logn \\ -2 & \text{with prob. } 3Ln^{-2/5} \log n \end{cases} \tag{7.4}$$

where L is a positive contstant large enough that for all $i_0 \leq i \leq i_{end}$

$$p_z(i)p_y(i) \geq w(i) - Ln^{-2/5}\log n, \quad p_z(i)p_y(i) + p_z(i)^2 \geq 1 - w(i) - 2Ln^{-2/5}\log n$$

and

$$p_z(i)p_y(i) + p_z(i)^2 + p_y(i) \geq 1 - 3Ln^{-2/5}\log n.$$

In this case, $\Delta\zeta(i)$ stochastically dominates $X(i)$. This follows from (4.8) in the case when $\zeta > 0$ and trivially when $\zeta = 0$.

For any $i_0 < i \leq i_{end}$ we have

$$E[X(i)|\mathcal{F}_{i-1}] = -6Ln^{-2/5}\log n$$

and

$$Var[X(i)|\mathcal{F}_{i-1}] = 2w(i) + 4Ln^{-2/5}\log n.$$

We will split the time range i_0 to i_{end} into $d = \log n$ many chunks of length $n^{3/5}\log^{-2} n$. Recall that $i_0 = n - n^{3/5}$ and for all $1 \leq \ell \leq d$ define

$$i_\ell = i_{\ell-1} + n^{3/5}\log^{-2} n.$$

For $0 \leq \ell < d$, we define a martingale starting at i_ℓ to be

$$S_\ell(k) = \sum_{i=i_\ell+1}^{k} \left(X(i) - E[X(i)|\mathcal{F}_{i-1}]\right).$$

Then for $0 \leq \ell < d$ we have

$$S_\ell(i_{\ell+1}) = \left(\sum_{i=i_\ell+1}^{i_{\ell+1}} X(i)\right) + 6Ln^{1/5}\log^{-1} n.$$

We also have that

$$V_\ell := \sum_{i=i_\ell+1}^{i_{\ell+1}} Var[X(i)|\mathcal{F}_{i-1}] = \Theta\left(n^{2/5}\log^{-2} n\right).$$

Further, using the fact that for the expression for V_ℓ is completely deterministic, we may choose $\gamma(n,\ell)$ such that $\gamma(n,\ell) < V_\ell < \gamma(n,\ell) + o(\gamma(n,\ell))$. By using the facts about $a(t)$ and $b(t)$ presented in Section 4, we may take $\gamma(n,\ell) = C_\ell n^{2/5}\log^{-2} n$ for come constant C_ℓ. Note here that there is an absolute constant c such that $C_\ell \leq c$ for all $0 \leq \ell < d$.

Hence applying Lemma 7.2 to S_ℓ with stopping time $i_{\ell+1}$, we see that

$$\frac{\left(\sum_{i=i_\ell+1}^{i_{\ell+1}} X(i)\right) + 6Ln^{1/5}\log^{-1} n}{\sqrt{C_\ell n^{2/5}\log^{-2} n}} \xrightarrow{d} \mathcal{N}(0,1).$$

So there exists some constant $p_0 > 0$ such that for each $0 \leq \ell < d$ (and n sufficiently large),

$$\mathbb{P}\left[\frac{\left(\sum_{i=i_\ell+1}^{i_{\ell+1}} X(i)\right) + 6Ln^{1/5}\log^{-1} n}{\sqrt{C_\ell n^{2/5}\log^{-2} n}} \geq \frac{6L+1}{\sqrt{c}}\right] \geq p_0$$

So we get that

$$\begin{aligned}\mathbb{P}\left[\forall\, 0<\ell\le d,\, \zeta(i_\ell)\le n^{1/5}\log^{-1} n\right] &\le \mathbb{P}\left[\forall\, 0\le \ell< d,\, \sum_{i=i_\ell+1}^{i_{\ell+1}} X(i)\le n^{1/5}\log^{-1} n\right]\\ &\le (1-p_0)^{\log n}\\ &= o(1).\end{aligned}$$

So we know that w.h.p. there is a point i_b where $\zeta(i_b)>n^{1/5}\log^{-1} n$. We would like to show that after $n^{3/5}\log^{-3} n$ steps, ζ has not decreased below $\frac{1}{2}n^{1/5}\log^{-1} n$. To prove this, we consider the martingale

$$S_b(k)=n^{1/5}\log^{-1} n+\sum_{i=i_b}^{k}\left(X(i)-E[X(i)|\mathcal{F}_{i-1}]\right).$$

Let $i_c=i_b+n^{3/5}\log^{-3} n$. Then

$$\sum_{i=i_b+1}^{i_c} Var[X(i)|\mathcal{F}_{i-1}]=\Theta\left(n^{2/5}\log^{-3} n\right)$$

By applying Lemma 5.4 to this martingale, we have that after $n^{3/5}\log^{-3} n$ steps,

$$\begin{aligned}\mathbb{P}\left[\exists i\,:\, i_b\le i\le i_c,\, \zeta(i)\le \frac{1}{2}n^{1/5}\log^{-1} n\right] &\le \mathbb{P}\left[\exists i\le i_c\,:\, S_b(i)\le \frac{1}{2}n^{1/5}\log^{-1} n\right]\\ &\le \exp\left(-\Omega\left(\frac{n^{2/5}\log^{-2} n}{n^{2/5}\log^{-3} n\left(1+n^{-1/5}\log^{-2} n\right)}\right)\right)\\ &\le o(1).\end{aligned}$$

So we know that whp, $\zeta(i)\ge \frac{1}{2}n^{1/5}\log^{-1} n$ for $i_b\le i\le i_c$. In this time, the algorithm is likely to delete an edge adjacent to a ζ vertex. Formally, we have that there exists some q_0 such that for all $i_b\le i\le i_c$,

$$p_z(i)p_\zeta(i)\ge q_0=\Omega\left(n^{-2/5}\log^{-1} n\right)$$

so that if W is a random variable representing the number of i between i_b and i_c when $\delta(i)=\zeta$, then W stochastically dominates $\mathrm{Bin}(n^{3/5}\log^{-3} n, q_0)$.

$$E[\mathrm{Bin}(n^{3/5}\log^{-3} n, q_0)]=\Omega\left(n^{1/5}\log^{-4} n\right),$$

so an application of the Chernoff bound tells us that, w.h.p., $W=\Omega\left(n^{1/5}\log^{-4} n\right)$. □

References

[1] J. Aronson, A.M. Frieze and B. Pittel, *Maximum matchings in sparse random graphs: Karp-Sipser revisited*, Random Structures and Algorithms 12 (1998) 111-178.

[2] T. Bohman, *The Triangle-Free Process*, Advances in Mathematics, 221 (2009) 1653-1677.

[3] T. Bohman and A.M. Frieze,*Hamilton cycles in 3-out*, Random Structures and Algorithms 35 (2009) 393-417.

[4] T. Bohman and A.M. Frieze, *Karp-Sipser on random graphs with a fixed degree sequence*, Combinatorics, Probability and Computing 20 (2011) 721-742.

[5] T. Bohman, A.M. Frieze and E. Lubetzky, *A note on the random greedy triangle-packing algorithm*, Journal of Combinatorics 1 (2010) 477-488.

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

[7] B.Bollobás, C. Cooper, T.I.Fenner and A.M.Frieze, *On Hamilton cycles in sparse random graphs with minimum degree at least k*, Journal of Graph Theory, 34 (2000) 42-59.

[8] David A. Freedman, *On Tail Probabilities for Martingales*, The Annals of Probability 3 (1975), 100-118.

[9] A.M. Frieze, *On a Greedy 2-Matching Algorithm and Hamilton Cycles in Random Graphs with Minimum Degree at Least Three*, manuscript (2011).

[10] A.M. Frieze, *Finding Hamilton cycles in sparse random graphs*, Journal of Combinatorial Theory B, 44 (1988) 230-250.

[11] A.M. Frieze, M.R. Jerrum, M. Molloy, R. Robinson and N.C. Wormald, *Generating and counting Hamilton cycles in random regular graphs*, Journal of Algorithms 21 (1996), 176-198.

[12] A.M. Frieze, J. Radcliffe and S. Suen, *Analysis of a simple greedy matching algorithm on random cubic graphs*, Combinatorics, Probability and Computing 4 (1995) 47-66.

[13] R.M. Karp and M. Sipser, *Maximum matchings in sparse random graphs*, Proceedings of the 22nd Annual IEEE Symposium on Foundations of Computing (1981) 364-375.

[14] R.W. Robinson and N.C. Wormald, *Almost all cubic graphs are Hamiltonian.* Random Structures and Algorithms **3** (1992) 117-125.

[15] R.W. Robinson and N.C. Wormald, *Almost all regular graphs are Hamiltonian.* Random Structures and Algorithms **5** (1994) 363-374.