

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

Deepak Bal\*      Patrick Bennett†      Tom Bohman‡      Alan Frieze§

October 26, 2017

## 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  $\kappa$  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  $\kappa(U) = \tilde{O}(n^{1/5})$ .

## 1 Introduction

In this paper we analyze the performance of a generalization of the well-known Karp-Sipser algorithm [14, 13, 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 [10]. 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 [10], 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 w.h.p. in many models with  $O(n)$  edges and  $\delta = 2$ .  $G_{3\text{-out}}$  is a random graph where each vertex chooses 3 neighbors

---

\*Department of Mathematical Sciences, Montclair State University, Montclair, NJ

†Department of Mathematics, Western Michigan University, Kalamazoo, MI. Research supported by a grant from the Simons Foundation #426894

‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA. Research supported in part by NSF Grant DMS1001638

§Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA. Research supported in part by NSF Grant CCF1013110

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 [16],[17]. Before this result, Frieze proved Hamiltonicity of  $r$ -regular graphs w.h.p. for  $r \geq 85$  using an algorithmic approach [11]. An algorithmic proof of Hamiltonicity for  $r \geq 3$  was given in [12]. In the binomial random graph  $G_{n,p}$  with  $p = c/n$ , there is no 2-factor and therefore studying the size of the largest 2-matching is an interesting problem. In the recent paper [18], an explicit asymptotic formula is given for the maximum size of a 2-matching in such graphs. The paper [15] generalizes this result to random hypergraphs.

In addition to the Hamiltonicity of  $G_{n,cn}^{\delta \geq 3}$  for  $3/2 < c < 10$ , the Hamiltonicity of random graphs  $G_{n,\mathbf{d}}$  with  $O(n)$  edges and a fixed degree sequence  $\mathbf{d}$  is a wide open question. One of the difficulties being that at present we do not know how to couple two graphs  $G_i = G_{\mathbf{d}_i}, i = 1, 2$ , where  $\mathbf{d}_1 \geq \mathbf{d}_2$  so that  $G_1 \supseteq G_2$ . 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 (which is the same as  $G_{n,cn}^{\delta \geq 3}$  for  $c = 3/2$ ).

The size of a 2-matching  $U$  is the number of edges in  $U$  and this is at least  $n - \kappa(U)$  where  $\kappa$  denotes the number of components. It was shown in [13] 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.*

Here the notation  $f(n) = \tilde{\Theta}(g(n))$  denotes  $c_1 g(n) \log^{d_1} n \leq f(n) \leq c_2 g(n) \log^{d_2} n$  for absolute constants  $c_1, c_2, d_1, d_2$  for  $n$  sufficiently large.

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  $\Theta(n^{1/5})$  components is also given in Section 4. In Section 5 we state and prove our dynamic concentration result. The proof of Theorem 1.1 is then completed in Sections 5, 6, and 7.

## 1.1 Values for the constants

Throughout the proof above, we collect various constraints on the constants in (5.16), (5.23), (5.25), (5.32), (5.35), (5.36), (5.40), (5.41), (5.42), (5.43), (5.45) and (5.46).

$$C_A = 400, \quad C_h = 500, \quad C_{p_y} = 2 \cdot 10^3, \quad C_\ell = 200, \quad C_x = 12 \cdot 10^6,$$

$$C_\zeta = 65 \cdot 10^6, \quad C_\alpha = 6 \cdot 10^8, \quad C_{v_B} = 2 \cdot 10^8, \quad C_B = 10^8, \quad C_T = 5 \cdot 10^5.$$

## 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 [10] is a generalization of Karp-Sipser for 2-matchings. Our algorithm is essentially the same as that presented in [10] 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 Bollobás [6]. Let  $W = [3n]$  be our set of *configuration points* and let  $W_i = [3(i-1) + 1, 3i]$ ,  $i \in [n]$ , partition  $W$  into 3-sets. The function  $\phi : W \rightarrow [n]$  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 all possible pairings  $\Omega$  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 multi-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  $G = G_F$ .

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  $G$ , and for each  $v \in V_\Gamma$  let  $d_\Gamma(v)$  be the degree of  $v$  in  $\Gamma$ . 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  $\Gamma$ .
- $b(v)$  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.

Note that the algorithm as written below can take any cubic (multi-)graph as input. However we intend to analyze its performance on the random cubic (multi-)graph  $G_F$ . An important aspect of our analysis is that we only reveal adjacencies (pairings) of  $G_F$  as the need arises in the

algorithm. 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  $\Gamma$  and added to  $U$ . If the selection  $w$  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  $\{w, x\}$  in the configuration adjacent to  $w$ . Call this exposure a *deletion move* and the vertex  $x$ , the *affected vertex*.

Details of the algorithm are now given.

**Algorithm 2** GREEDY:

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 unique (selected) neighbor in  $\Gamma$  is  $w$ . Remove  $\{v, w\}$  from  $\Gamma$  and add it to  $U$ . Set  $b(v) = 1$  and move  $v$  to  $Z_0$ .

Re-assign( $w$ ). ( This means place  $w$  in the set  $Z_k$  if it now has degree  $k \leq 1$  in  $U$ , or else remove  $w$  from  $\Gamma$  if it has degree 2 in  $U$ ).

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

Choose a random vertex  $v$  of  $Y_2$ . Randomly select one of the two neighbors of  $v$  in  $\Gamma$  and call it  $w$ .

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

Otherwise, remove  $\{v, w\}$  from  $\Gamma$  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  $U$ . Suppose the unique (selected) neighbor of  $v$  in  $\Gamma$  is  $w$ . Remove  $\{v, w\}$  from  $\Gamma$  and add it to  $U$ . Remove  $v$  from  $\Gamma$ .

Re-assign( $w$ ).

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

Choose a random vertex  $v$  of  $Y$ . Randomly select one of the three neighbors of  $v$  in  $\Gamma$  and call it  $w$ .

If  $w = v$ , then we remove loop  $\{v, v\}$  from  $\Gamma$  and move  $v$  to  $Y_1$ .

Otherwise, remove  $\{v, w\}$  from  $\Gamma$  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 (multi-)graph is 2-regular since  $Z$  is the set of degree 2 vertices. Put a maximum matching on this remaining (multi-)graph. Add the edges of this matching to  $U$ .

**Step 4** Return  $U$  (the algorithm terminates here).

**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  $\Gamma$ . If  $w$  was in  $Z$  prior to removal, then the removal of  $w$  from  $\Gamma$  causes an edge  $(w, w')$ , to be deleted from  $\Gamma$ . 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.

To see that this algorithm produces a 2-matching, note first that in Steps 1 and 2, only one edge (from  $\Gamma$ ) at a time is added to  $U$  and it is never a loop. Every vertex in  $\Gamma$  is adjacent to at most one edge in  $U$ . Thus the addition of such an edge can only create vertices of degree at most 2 in  $U$ . When a vertex gets degree 2 in  $U$ , it is removed from  $\Gamma$ , thus deleting all of its other edges. Immediately before Step 3, the vertices of  $Z$  have degree 1 in  $U$ , thus adding a matching among these vertices will only increase their degree to at most 2 in  $U$ .

Note that the final 2-matching  $U$  may contain cycles, since in steps 1(c) and 3, we may insert an edge that closes a cycle. However this is not a problem because a 2-matching can contain cycles. We include here cycles of length two i.e. multi-edges. Note that the expected number of multi-edges in  $G$  is  $O(1)$  and so this is not an issue.

### 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  $E_\Gamma$ . We also define the variable

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

We sometimes consider  $\zeta$  to be the set  $Y_1 \cup Y_2 \cup Z_1$ . Note that, unlike  $Y, Z$ , etc. the size of this set is not the same as  $\zeta$ , however there are  $Y_1 + 2Y_2 + Z_1$  half-edges (i.e. unpaired configuration points) in  $\Gamma$  that are adjacent to  $Y_1 \cup Y_2 \cup Z_1$ .

If  $X$  is a variable indexed by  $i$ , we define

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

#### 3.1 The sequences $\sigma, \delta$

We define two sequences  $\sigma, \delta$  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,  $\sigma$  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  $\sigma(i) = multi$  in the

following case: step  $i$  is of type 1(c),  $w = u \in Z$ , where  $u$  is the other end of the path in  $U$  that contains  $v$ . Furthermore, the edge  $\{v, u\}$  is already in  $U$  (so we have revealed a multi-edge). 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  $Y, Z, \zeta$ .

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. Using *loop*, *multi* in this way will allow us to define variables  $A, B$  whose one step changes do not depend on whether or not  $\zeta > 0$ .

$\delta$  is a sequence of the following symbols:  $Y, Z, \zeta, \emptyset$ . We will put  $\delta(i) = \emptyset$  when there is no deletion move at step  $i$  (i.e. when  $\sigma(i) \notin \{Z, \text{multi}\}$ ). Otherwise  $\delta(i)$  just indicates the type of the affected vertex that the deletion move chooses (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:

$$\begin{aligned} A &:= Y + \zeta. \\ B &:= 2Y + Z + \zeta. \end{aligned}$$

$A$  is a natural quantity to define, since Step 3 of the algorithm begins 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  $i$ . Here  $A(i), B(i), \dots$ , denote the values of the corresponding variables  $A, B, \dots$ , at the end of  $i$  iterations of 2GREEDY. We have

$$\Delta Y(i) = -\mathbb{1}_{\zeta(i)=0} - \mathbb{1}_{\sigma(i)=Y} - (\mathbb{1}_{\sigma(i)=Z} + \mathbb{1}_{\sigma(i)=\text{multi}}) \mathbb{1}_{\delta(i)=Y}. \quad (3.1)$$

To justify the above equation, note that if the selection is a  $Y$  vertex then  $Y$  decreases by 1. We may also lose a  $Y$  vertex if there is a deletion (i.e. if the selection is a  $Z$  vertex or in the event of a *multi* selection) and the affected vertex is in  $Y$ . Finally, we lose one more  $Y$  vertex whenever  $\zeta = 0$ .

The following equations are justified similarly, by considering the effect due to selections, deletions, and whether  $\zeta = 0$  and  $Y \neq 0$  (we do not consider Step 3 here).

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

Observe that if  $\sigma(i) = \text{loop}$  then we do not increase  $Z$  even though  $\zeta(i) = 0$  and so we subtract one to counter  $\mathbb{1}_{\zeta(i)=0}$ . In this case  $\zeta$  increases and this feeds into the next equation.

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

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

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

which states that we make a deletion move if and only if our selection move was  $Z$  or  $multi$ . Then we have that

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

$$\begin{aligned}\Delta B(i) &= 2\Delta Y(i) + \Delta Z(i) + \Delta\zeta(i) \\ &= -2 + \mathbb{1}_{\sigma(i)=loop} - \mathbb{1}_{\delta(i)=\zeta}.\end{aligned}\tag{3.5}$$

which do not depend on whether  $\zeta = 0$ . Note also that if we establish dynamic concentration on  $A, B, \zeta$  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 $A, B, \zeta$

In this section, 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

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

where  $t \leq n$ , since we add an edge to  $U$  each round and  $U$  is a 2-matching.

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  $\zeta$  stays “small”. We will prove later 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. Note that these represent the probabilities that a selection or deletion move is  $Z, Y$  or  $\zeta$  respectively. So for example  $E[\mathbb{1}_{\sigma(i)=Z}] = p_z$ . We can claim this because in the configuration model, we can arbitrarily change the pairing of unpaired configuration points while still being consistent with the history of the process. We are using the method of “deferred decisions”.

#### 4.1 The trajectory $b(t)$

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

$$B(i) = 2n - 2i + \sum_{j \leq i} (\mathbb{1}_{\sigma(j)=loop} - \mathbb{1}_{\delta(j)=\zeta}). \quad (4.1)$$

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

$$b(t) = 2(1 - t). \quad (4.2)$$

#### 4.2 The trajectory $a(t)$

We derive an ODE that  $a$  should satisfy. Since  $p_y = \frac{3(A-\zeta)}{2B-A} \approx \frac{3a(t)}{2b(t)-a(t)}$  and  $p_z = \frac{2(B-2A+\zeta)}{2B-A} \approx \frac{2b(t)-4a(t)}{2b(t)-a(t)}$  we should have (referring to (3.4) and ignoring all  $\zeta$ , *loop*, and *multi* events since they should be negligible)

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

Thus  $a(t)$  should satisfy

$$a' = -\frac{6a}{4 - 4t - a}. \quad (4.3)$$

**Lemma 4.1.** *The unique solution to (4.3) with boundary condition  $a(0) = 1$  is*

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

*Proof.* The substitution  $a = (1 - t)x$  yields a separable ODE:

$$(1 - t)x' = x - \frac{6x}{4 - x}$$

or

$$\frac{dt}{1 - t} = dx \left( \frac{3}{x + 2} - \frac{2}{x} \right).$$

This can be solved directly and together with  $x(0) = 1$  this gives

$$\frac{1}{1 - t} = \frac{(x + 2)^3}{27x^2}.$$

After substituting back we arrive at

$$0 = (a + 2 - 2t)^3 - 27a^2 = a^3 - (6t + 21)a^2 + 12(1 - t)^2a + 8(1 - t)^3.$$

We make the substitution  $a = r + 7 + 2t$  to obtain the equation

$$r^3 - 27(5 + 4t)r - 54(11 + 14t + 2t^2) = 0.$$



Putting  $p = 27(5 + 4t)$  and  $q = -54(11 + 14t + 2t^2)$  and using the cosine formula for the solution of a cubic equation we have three roots

$$\begin{aligned} r_k(t) &= 2\sqrt{\frac{p}{3}} \cos\left(\frac{1}{3} \arccos\left(-\frac{3q}{2p}\sqrt{\frac{3}{p}}\right) + k\frac{2\pi}{3}\right) \\ &= 6\sqrt{5 + 4t} \cos\left(\frac{1}{3} \arccos\left(\frac{11 + 14t + 2t^2}{(5 + 4t)^{3/2}}\right) + k\frac{2\pi}{3}\right), \quad k = 0, 1, 2. \end{aligned}$$

We can assume that  $0 \leq \arccos(x) \leq \pi$ .

We now have three possibilities for  $a$ , viz.  $a_k(t) = 7 + 2t + r_k(t)$ ,  $k = 0, 1, 2$ . We use the boundary condition  $a(0) = 1$  to see which choice is correct.

Putting  $t = 0$  gives

$$r_k(0) = 6\sqrt{5} \cos\left(\frac{1}{3} \arccos\left(\frac{11}{5^{3/2}}\right) + k\frac{2\pi}{3}\right), \quad k = 0, 1, 2.$$

Also, using the identity  $\cos(3\theta) = 4\cos^3(\theta) - 3\cos(\theta)$  with  $\theta = \arccos\left(\frac{1}{\sqrt{5}}\right)$  we see that

$$\cos\left(3 \arccos\left(\frac{1}{\sqrt{5}}\right)\right) = -\frac{11}{5^{3/2}}$$

from which we deduce that

$$\frac{1}{3} \left( \arccos\left(\frac{11}{5^{3/2}}\right) + \pi \right) = \frac{1}{3} \arccos\left(-\frac{11}{5^{3/2}}\right) = \arccos\left(\frac{1}{\sqrt{5}}\right)$$

and then similarly

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

It follows that  $r_2(0) = -6$  and  $a_2(0) = 1$ . Furthermore,  $r_0(0), r_1(0) = 3 \pm 6\sqrt{3}$  will be different from  $r_2(0)$  and so  $a(t) = a_2(t)$  or

$$\begin{aligned} a(t) &= 7 + 2t + 6\sqrt{5 + 4t} \cos\left(\frac{1}{3} \arccos\left(\frac{11 + 14t + 2t^2}{(5 + 4t)^{3/2}}\right) + \frac{4\pi}{3}\right) \\ &= 7 + 2t - 6\sqrt{5 + 4t} \cos\left(\frac{1}{3} \arccos\left(\frac{11 + 14t + 2t^2}{(5 + 4t)^{3/2}}\right) + \frac{\pi}{3}\right). \end{aligned} \quad (4.4)$$

□

From here we can see that  $a(t) \rightarrow 0$  as  $t \rightarrow 1^-$ . ( $t = i/n$  and  $i \leq n$  and a 2-matching has at most  $n$  edges.) More precisely,

$$\lim_{t \rightarrow 1^-} \frac{a(t)}{(1-t)^{\frac{3}{2}}} = \left(\frac{2}{3}\right)^{\frac{3}{2}}. \quad (4.5)$$

To confirm this, we use the facts that for  $\delta \rightarrow 0$

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

$$\sqrt{1-\delta} = 1 - \frac{1}{2}\delta + O(\delta^2),$$

$$\cos\left(\frac{\pi}{3} + \delta\right) = \frac{1}{2} - \frac{\sqrt{3}}{2}\delta + O(\delta^2),$$

and the fact that

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

Rewriting, we see

$$\begin{aligned} a(1-\varepsilon) &= 9-2\varepsilon - 18 \left(1 - \frac{2}{9}\varepsilon + O(\varepsilon^2)\right) \times \cos\left(\frac{1}{3} \left(\left(\frac{8\varepsilon^3}{729}\right)^{1/2} + O(\varepsilon^{5/2})\right) + \frac{\pi}{3}\right) \\ &= 9-2\varepsilon - 18 \left(1 - \frac{2}{9}\varepsilon + O(\varepsilon^2)\right) \times \frac{1}{2} \left(1 - \frac{2}{27} \left(\frac{2}{3}\right)^{1/2} \varepsilon^{3/2} + O(\varepsilon^{5/2})\right) \\ &= \left(\frac{2}{3}\right)^{3/2} \varepsilon^{3/2} + O(\varepsilon^2), \end{aligned}$$

which implies (4.5).

Additionally,

$$\begin{aligned} \frac{d}{dt} \left( \frac{a(t)}{(1-t)^{3/2}} \right) &= -\frac{6a}{4(1-t)-a} (1-t)^{-3/2} + \frac{3}{2}a \cdot (1-t)^{-5/2} \\ &= a \cdot (1-t)^{-5/2} \left( \frac{3}{2} - \frac{6(1-t)}{4(1-t)-a} \right) \\ &< 0. \end{aligned} \tag{4.6}$$

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

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

This inequality is used extensively in Section 5.

### 4.3 Downward drift of $\zeta$

We expect  $\zeta$  to be “small”, and to heuristically justify that claim we will show that whenever  $\zeta$  is positive, it is likely to decrease. In this section therefore, we are implicitly assuming that  $p_y, p_z \gg p_\zeta$ . Let us first note that this is non-trivial, i.e. it is possible for  $\zeta$  to grow. Suppose the algorithm executes Step 1(b). So  $v \in Y_2$  and  $w$  is one of its two neighbors. One possible scenario is that  $w \in Z$  which means there will be a deletion move this step (case 2 of  $\text{re-assign}(w)$ ). This deletion move may affect a vertex  $u \in Y$ . So  $v$  moved from  $Y_2$  to  $Z_1$  and  $u$  moved from  $Y$  to  $Y_2$ . Thus the net change in  $\zeta$  is  $-1 + 2 = +1$ . Assume that  $\zeta(i) > 0$ . In the following table, we once again make use of the fact that  $\delta(i) \neq \emptyset$  if and only if  $\sigma(i) \in \{Z, \text{multi}\}$ . So for example,

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

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

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

Each entry in the above table can be understood by considering the most likely way for  $\Delta\zeta$  to take the appropriate value. For example, the most likely way for  $\Delta\zeta = 1$  is for  $\sigma = Z$  and  $\delta = Y$  (even though this is not the only way for  $\Delta\zeta = 1$ ). Specifically,

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

Similarly, the most likely way for  $\Delta\zeta = 0$  is for  $\sigma = Z$  and  $\delta = Z$ . The only way to get  $\Delta\zeta = -1$  is to have  $\sigma = Y$  (in which case  $\delta = \emptyset$ ), and the only ways to get  $\Delta\zeta = -2$  involve  $\sigma = \zeta$  or  $\delta = \zeta$ .

Therefore, roughly speaking we have

$$\begin{aligned} E[\Delta\zeta(i)] &= p_z p_y - p_y + O(p_\zeta) \\ &\approx \frac{2b-4a}{2b-a} \cdot \frac{3a}{2b-a} - \frac{3a}{2b-a} = -\frac{9a^2}{(2b-a)^2} \end{aligned} \quad (4.9)$$

and are motivated to define

$$\Phi(t) := \frac{9a^2}{(2b-a)^2} = \Theta(1-t) \quad (4.10)$$

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

#### 4.4 Expected behavior of $\zeta$

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

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

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  $\mathbb{E}[\Delta X_\tau] = -(1-\tau)$  and  $\text{Var}[\Delta X_\tau] = (1-\tau)^{\frac{1}{2}}$ . After  $s$  steps, we have  $\mathbb{E}[X_\tau(s)] = -(1-\tau)s$  and  $\text{Var}[X_\tau(s)] = (1-\tau)^{\frac{1}{2}}s$ . 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 too 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  $s = \frac{1}{4}(1-\tau)^{-\frac{3}{2}}$ , where we have  $\psi(s) = \frac{1}{4}(1-\tau)^{-\frac{1}{2}}$ .

Now we reconsider the variable  $\zeta$ . 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  $\zeta$  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  $j \leq i \leq j + (1-\tau)^{-\frac{3}{2}}$ . Also, for all  $j \leq i \leq j + (1-\tau)^{-\frac{3}{2}}$  we should have  $\zeta(i) \leq \frac{1}{4}(1-\tau)^{-\frac{1}{2}}$ . 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  $j + (1-\tau)^{-\frac{3}{2}} > n$ . 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  $j \geq n - n^{\frac{3}{5}}$  and  $s \leq n-j$  we have  $\psi(s) \leq (1-\tau)^{\frac{1}{4}} s^{\frac{1}{2}} \leq n^{-1/10} \cdot n^{3/10} = n^{1/5}$ . Thus, we never expect  $\zeta$  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 which affects 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 can be increased by a deletion move that affects a vertex in  $Y_1$  or  $Y_2$ . 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,  $\mathbb{P}[\delta(i) = \zeta] = \frac{2Z}{2M-1} \cdot \frac{\zeta}{2M-3} = O\left(\frac{1}{n} \min\left\{(1-t)^{-\frac{3}{2}}, \frac{n^{\frac{1}{5}}}{1-t} \log^{O(1)} n\right\}\right)$ . So integrating, we estimate the total number of components as

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

Indeed, we will see in section 7 that a matching (up to log factors) lower bound also holds. Very roughly speaking, this is because we expect there to be a positive proportion of steps  $i$  where  $\zeta(i)$  is more than its expectation by a standard deviation.

We will now rigorously justify the above claims about the performance of the algorithm 2GREEDY.

## 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  $\zeta$  does roughly what we expect it to do. We will also introduce “error” terms for both  $A, B$  and a “correction” term  $\alpha$  for the variable  $A$ . For most of the process,  $\alpha$  will stay smaller than the error term for  $A$ . However, toward the end of the process  $\alpha$  will be significant. Using  $\alpha$  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  $A, B$ :

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

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

The definition of  $\alpha(i)$  is through a recurrence – see (5.13).

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 four following conditions fail:

$$|e_a(i)| \leq f_a(t), \quad (5.1)$$

$$|e_b(i)| \leq f_b(t), \quad (5.2)$$

$$\zeta(i) \leq f_\zeta(t), \quad (5.3)$$

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

$$\zeta(i) \leq \zeta(j) - \sum_{j \leq k < i} \Phi\left(\frac{k}{n}\right) + \ell_j(t) \quad (5.4)$$

for some as-yet unspecified error functions  $f_a, f_b, f_\zeta, \ell_j$  and absolute constant  $C_T$ . See Section 1.1 for this and subsequently introduced constants.

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), (5.4).

**Theorem 5.1.** *With high probability,*

$$T = n - C_T n^{\frac{7}{15}} \log^{\frac{6}{5}} n. \quad (5.5)$$

The remainder of this section contains the proof of Theorem 5.1. Here we define the error functions  $f_a, f_b, f_\zeta$  (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.

$$f_a(t) := C_A (1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n. \quad (5.6)$$

$$f_b(t) := C_B \cdot \begin{cases} (1-t)^{-\frac{1}{2}} \log n & : 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n. \\ -n^{\frac{1}{5}} \log^{\frac{4}{5}} n \log(1-t) & : \text{otherwise.} \end{cases} \quad (5.7)$$

$$f_\zeta(t) := C_\zeta \cdot \begin{cases} (1-t)^{-\frac{1}{2}} \log n, & : 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n. \\ n^{\frac{1}{5}} \log^{\frac{4}{5}} n & : \text{otherwise.} \end{cases} \quad (5.8)$$

## 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  $x, y \neq 0$  and  $|\frac{\varepsilon_x}{x}|, |\frac{\varepsilon_y}{y}| \leq \frac{1}{2}$ , 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.*

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

□

## 5.2 $T$ is not triggered by $A$

We define

$$A^+(i) := A(i) - na(t) - \alpha(i) - f_a(t) = e_a(i) - f_a(t) \quad (5.9)$$

and let the stopping time  $T_j := \min\{i(j), \max(j, T)\}$  where  $i(j)$  represents the least index  $i \geq j$  such that  $e_a(i)$  is not in the critical interval

$$[g_a(t), f_a(t)] \quad (5.10)$$

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) and (3.6), (3.7), (3.8) that

$$\begin{aligned}
E[\Delta A(i) | \mathcal{F}_i] &= -1 - \frac{3Y}{2M} - \frac{\zeta}{2M} + \frac{2Z}{2M} - 2 \cdot \frac{2Z}{2M} \cdot \frac{\zeta}{2M} + O\left(\frac{1}{M}\right) \\
&= \frac{-(2B - A) - 3(A - \zeta) - \zeta + 2(B - 2A + \zeta)}{2B - A} - \frac{4\zeta(B - 2A + \zeta)}{(2B - A)^2} + O\left(\frac{1}{2B - A}\right) \\
&= -\frac{6A}{2B - A} + \frac{4\zeta(A + B)}{(2B - A)^2} + O\left(\frac{1}{2B - A} + \frac{\zeta^2}{(2B - A)^2}\right) \\
&= -\frac{6(na + \alpha + e_a)}{2(nb + e_b) - (na + \alpha + e_a)} + \frac{4\zeta[(na + \alpha + e_a) + (nb + e_b)]}{[2(nb + e_b) - (na + \alpha + e_a)]^2} + O\left(\frac{1}{2B - A} + \frac{\zeta^2}{(2B - A)^2}\right) \\
&= -\frac{6a}{2b - a} + \frac{12ae_b - 12b(\alpha + e_a)}{n(2b - a)^2} + \frac{4(a + b)\zeta}{n(2b - a)^2} + O\left(\frac{1}{n(2b - a)} + \frac{\alpha^2 + f_a^2 + f_b^2 + f_\zeta^2}{n^2(2b - a)^2}\right). \quad (5.11)
\end{aligned}$$

The second line above follows from substituting the values of  $Y, Z, M$  in terms of  $A, B, \zeta$ . The last line above also follows from Lemma 5.2 (the fourth line has two fractions with error terms in the numerators and denominators. We apply Lemma 5.2 to these fractions, regarding  $e_a, e_b, \alpha$  as error terms, to arrive at the last line, making use of (5.1)-(5.3), that  $e_a + e_b + \alpha = o(2b - a)n$  and that  $a + b = O(2b - a)$ ). Also note that we only apply a crude form of Lemma 5.2 to the second fraction of the fourth line, as the lemma would allow us to put fewer of the resulting terms into the big- $O$ . 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  $y$ ,  $|xy| \leq \frac{1}{2}(x^2 + y^2)$ . Note that we have not put all the occurrences of  $\zeta$  into the big- $O$  term. While we will see that the  $\zeta$  term inside the big- $O$  is negligible, the  $\zeta$  term outside the big- $O$  may become significant towards the end of the process.

We are now motivated to cancel out the  $\zeta$  term in the last line by recursively defining

$$\alpha(0) := 0. \quad (5.12)$$

$$\begin{aligned} \alpha(i+1) &:= \alpha(i) + \frac{4(a+b)\zeta - 12b\alpha(i)}{n(2b-a)^2} \\ &\geq \alpha(i) \left( 1 - \frac{12b}{n(2b-a)^2} \right). \end{aligned} \quad (5.13)$$

We see from (4.2) and (4.7) that

$$a + b \leq 3(1-t) \leq 2b - a \leq 4(1-t). \quad (5.14)$$

So using  $T \leq n - C_T n^{\frac{7}{15}} \log^{\frac{6}{5}} n$  we have that  $\left(1 - \frac{12b}{n(2b-a)^2}\right) \geq 0$  and hence that  $\alpha(i) \geq 0$  throughout.

From (5.13) and the definition of  $f_\zeta$  and  $\zeta \leq f_\zeta$ , it follows that for  $i \leq T$ ,

$$\begin{aligned} 0 \leq \alpha(i) &\leq \sum_{j=0}^i \frac{4(a+b)f_\zeta}{n(2b-a)^2} \leq \sum_{j=0}^i \frac{4 \cdot 3 \left(1 - \frac{j}{n}\right) f_\zeta \left(\frac{j}{n}\right)}{n \left(3 \left(1 - \frac{j}{n}\right)\right)^2} \\ &= \frac{4C_\zeta}{3n} \sum_{j=0}^i \frac{1}{1 - \frac{j}{n}} \cdot \begin{cases} \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}} \log n, & : 1 - \frac{j}{n} > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n \\ n^{\frac{1}{5}} \log^{\frac{4}{5}} n & : \text{otherwise} \end{cases} \\ &\leq \frac{4C_\zeta}{3} \int_{\tau=0}^{\frac{i+1}{n}} \frac{1}{1 - \tau} \cdot \begin{cases} (1 - \tau)^{-\frac{1}{2}} \log n, & : 1 - \tau > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n \\ n^{\frac{1}{5}} \log^{\frac{4}{5}} n & : \text{otherwise} \end{cases} d\tau \\ &\leq C_\alpha \cdot \begin{cases} \log n (1-t)^{-1/2} & \text{for } i \leq n - n^{3/5} \log^{2/5} n, \\ n^{1/5} \log^{9/5} n & \text{for } n - n^{3/5} \log^{2/5} n < i \leq T, \end{cases} \end{aligned} \quad (5.15)$$

since  $C_\alpha$  and  $C_\zeta$  satisfy

$$C_\alpha > 8C_\zeta. \quad (5.16)$$

Note that we can pass from the sum on the second line to the integral on the third line since the integrand is an increasing function. Also note that in evaluating the integral, the value of the antiderivative at  $\tau = \frac{i+1}{n}$  is asymptotically the same as the value at  $\tau = t = \frac{i}{n}$ , so the last inequality holds since we chose  $C_\alpha$  large enough with room to spare.

Now let  $j \leq i < T_j$ . Note that if this holds, then by the definition of  $T_j$ ,  $i$  satisfies  $e_a(i) \in [g_a(t), f_a(t)]$ . We have the supermartingale condition

$$E[\Delta A^+(i)|\mathcal{F}_i] = E[\Delta A(i)|\mathcal{F}_i] - a'(t) - \frac{4(a+b)\zeta - 12b\alpha(i)}{n(2b-a)^2} - \frac{1}{n}f'_a(t) + O\left(\frac{1}{n}a''(t) + \frac{1}{n^2}f''_a(t)\right) \quad (5.17)$$

$$\leq -\frac{12bg_a}{n(2b-a)^2} - \frac{1}{n}f'_a(t) + O\left(\frac{af_b}{n(2b-a)^2} + \frac{1}{n(2b-a)} + \frac{\alpha^2 + f_a^2 + f_b^2 + f_\zeta^2}{n^2(2b-a)^2} + \frac{1}{n}a''(t) + \frac{1}{n^2}f''_a(t)\right). \quad (5.18)$$

Note that in the first line we have used (5.9) and (5.13), and in the last line we have used (5.11), (5.13), the fact that  $e_a \geq g_a$ , and also that  $a$  satisfies the differential equation (4.3). By taking  $g_a = C_g f_a$  where  $C_g < 1$  we guarantee that the corresponding critical interval is non-empty. We will subsequently choose  $C_g = 3/4$ , see (5.24). By (4.2), (5.6) and (5.14) we have

$$-\frac{12bg_a}{n(2b-a)^2} - \frac{1}{n}f'_a = -\Theta\left(n^{-1/2} \log^{1/2} n (1-t)^{-1/4}\right). \quad (5.19)$$

We then see that  $A^+(j), \dots, A^+(T_j)$  is a supermartingale once we prove the following claim.

**Claim 5.3.**  $-\frac{12bg_a}{n(2b-a)^2} - \frac{1}{n}f'_a$  dominates the big- $O$  term in (5.18).

*Proof.* Throughout this proof we will assume that  $(1-t) = \Omega(n^{-8/15} \log^{6/5} n)$ , see (5.5). Now (5.14) and (5.19) take care of the second big- $O$  term in (5.18). Also it is not hard to see by (5.6) that  $f''_a(t) = O(n^{1/2} \log^{1/2} n (1-t)^{-5/4}) = o(f_a^2/(2b-a)^2)$  and so the fifth big- $O$  term is taken care of.

Now consider the fourth big- $O$  term. It follows from (4.3) and (4.7) that

$$\frac{6\left(\frac{2}{3}\right)^{3/2} (1-t)^{3/2}}{4(1-t)} \leq -a'(t) = \frac{6a}{4-4t-a} \leq \frac{6(1-t)^{3/2}}{4(1-t) - (1-t)}$$

and so

$$-2(1-t)^{1/2} \leq a'(t) \leq -\sqrt{\frac{2}{3}}(1-t)^{1/2}. \quad (5.20)$$

Also we have

$$a''(t) = \frac{-6[a'(4-4t-a) + a(4+a')]}{(4-4t-a)^2}$$

and then from (4.7) and (5.20) we deduce that

$$|a''(t)| = O\left((1-t)^{-1/2}\right). \quad (5.21)$$

Thus

$$\frac{bg_a}{(2b-a)^2 a''} = \Omega\left(\left(n^{-8/15} \log^{6/5} n\right)^{1/4} n^{1/2} \log^{1/2} n\right) \gg 1$$

and this takes care of the fourth big- $O$  term.



For the first and third big- $O$  terms we must consider cases according to the value of  $t$ . First consider the case  $1-t > n^{-2/5} \log^{2/5} n$ . By (4.2), (4.7), (5.7), and our choice of  $g_a$  (again see (5.24)) we have  $af_b = O((1-t) \log n) = o(bg_a)$  since

$$\frac{bg_a}{(1-t) \log n} = \Omega \left( \left( n^{-2/5} \log^{2/5} n \right)^{3/4} n^{1/2} \log^{-1/2} n \right) \gg 1$$

and this deals with the first term. For the third term we see by (5.6)-(5.8) and (5.15) that  $\alpha^2 + f_a^2 + f_b^2 + f_\zeta^2 = O((1-t)^{-1} \log^2 n + (1-t)^{3/2} n \log n)$ . Now using the bounds on  $t$  we have

$$\frac{nb g_a}{(1-t)^{3/2} n \log n} = \Omega \left( \left( n^{-2/5} \log^{2/5} n \right)^{1/4} n^{1/2} \log^{-1/2} n \right) \gg 1$$

and

$$\frac{nb g_a}{(1-t)^{-1} \log^2 n} = \Omega \left( \left( n^{-2/5} \log^{2/5} n \right)^{11/4} n^{3/2} \log^{-3/2} n \right) \gg 1$$

and so  $\alpha^2 + f_a^2 + f_b^2 + f_\zeta^2 = o(nb g_a)$  so this deals with the third term, and finishes the case  $1-t > n^{-2/5} \log^{2/5} n$ .

The other case is  $\Omega(n^{-8/15} \log^{6/5} n) \leq 1-t \leq n^{-2/5} \log^{2/5} n$ . The only terms that change are  $f_b = -n^{1/5} \log^{4/5} n \log(1-t) = O(n^{1/5} \log^{9/5} n)$  by (5.7) and  $f_\zeta, \alpha = O(n^{1/5} \log^{9/5} n)$  by (5.8) and (5.15). So  $af_b = O(n^{1/5} \log^{9/5} n (1-t)^{3/2}) = o(bg_a)$  and  $f_\zeta^2 = O(n^{2/5} \log^{8/5} n) = o(nb g_a)$  using our bounds on  $t$ .

□

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

**Lemma 5.4.** *Let  $X_j$  be a supermartingale, such that  $-C \leq \Delta X(j) \leq c$  for all  $j$ , for  $c < \frac{C}{10}$ . Then for any  $u < cm$  we have  $\Pr(X_m - X_0 > u) \leq \exp\left(-\frac{u^2}{3cCm}\right)$*

We have by (3.4) that

$$-2 \leq \Delta A \leq 0. \tag{5.22}$$

For an absolute bound on  $a'(t)$  we have by (5.20) and the bounds on  $t$  that

$$-2 \leq a'(t) \leq -\left(\frac{2C_T}{3}\right)^{\frac{1}{2}} n^{-\frac{4}{15}} \log^{\frac{3}{5}} n.$$

Now by (5.9) we see

$$\Delta A^+ = \Delta A - a' - \Delta \alpha + O\left(\frac{1}{n} f'_a + \frac{1}{n} a''\right)$$

and before the stopping time  $T$  we have by (5.8) and (5.15) that

$$\zeta \leq C_\zeta n^{\frac{1}{5}} \log^{\frac{4}{5}} n \quad \text{and} \quad \alpha \leq C_\alpha n^{\frac{1}{5}} \log^{\frac{9}{5}} n$$

and so we have

$$\begin{aligned}
|\Delta\alpha| &\stackrel{(5.13)}{=} \left| \frac{4(a+b)\zeta - 12b\alpha(i)}{n(2b-a)^2} \right| \stackrel{(4.7)}{\leq} \frac{12(1-t)\zeta + 24(1-t)\alpha(i)}{9n(1-t)^2} \\
&\leq \frac{12C_\zeta n^{\frac{1}{5}} \log^{\frac{4}{5}} n + 24C_\alpha n^{\frac{1}{5}} \log^{\frac{9}{5}} n}{9n(1-t)} \leq 3C_\alpha n^{-\frac{4}{5}} \log^{\frac{9}{5}} n (1-t)^{-1} \\
&\leq \frac{3C_\alpha}{C_T} n^{-\frac{4}{15}} \log^{\frac{3}{5}} n
\end{aligned}$$

so

$$\left| \frac{\Delta\alpha}{a'(t)} \right| \leq \frac{3C_\alpha}{C_T^{\frac{3}{2}}}.$$

Since

$$C_\alpha > 2C_T^{3/2} \tag{5.23}$$

we get the bounds (using (5.22) and (5.20))

$$\Delta A^+ \leq - \left( a' \left( \frac{j}{n} \right) + \Delta\alpha \right) (1 + o(1)) \leq \left( \frac{3C_\alpha}{C_T^{\frac{3}{2}}} + 1 + o(1) \right) \left| a' \left( \frac{j}{n} \right) \right| \leq \frac{7C_\alpha}{C_T^{\frac{3}{2}}} \left( 1 - \frac{j}{n} \right)^{\frac{1}{2}}$$

and

$$\Delta A^+ \geq \Delta A - o(1) \geq -2 + o(1) \geq -3$$

for the supermartingale  $A^+(j) \cdots A^+(T_j)$ . Thus, if  $A$  crosses its upper boundary  $na(t) + \alpha(i) + f_a(t)$  at the stopping time  $T$ , since  $\Delta A^+ \leq \frac{7C_\alpha}{C_T^{\frac{3}{2}}}$  and this will be the first crossing there is some step  $j$  (with  $T = T_j$ ) such that

$$A^+(j) \leq g_a \left( \frac{j-1}{n} \right) - f_a \left( \frac{j-1}{n} \right) + \frac{7C_\alpha}{C_T^{\frac{3}{2}}}$$

and  $A^+(T_j) > 0$ . In this case,  $j$  is intended to represent the step when  $e_a$  enters the critical interval, (5.10). Our choice of constants in Section 1.1 allows us to apply Lemma 5.4 and see that the probability of the supermartingale  $A^+$  having such a large upward deviation has probability at most

$$\exp \left\{ - \frac{\left( f_a \left( \frac{j-1}{n} \right) - g_a \left( \frac{j-1}{n} \right) - \frac{7C_\alpha}{C_T^{\frac{3}{2}}} \right)^2}{3 \cdot 3 \cdot \frac{7C_\alpha}{C_T^{\frac{3}{2}}} n \left( 1 - \frac{j}{n} \right)^{\frac{3}{2}}} \right\}.$$

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

$$\begin{aligned}
f_a(t) &= C_A (1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n. \\
g_a(t) &= \frac{3}{4} f_a(t).
\end{aligned} \tag{5.24}$$

since the constant  $C_A$  is chosen so that

$$\frac{\left(\frac{1}{4}C_A\right)^2}{\frac{63C_\alpha}{C_T^{\frac{3}{2}}}} > 1. \quad (5.25)$$

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 condition (5.4)

Referring to (4.8) and (3.6)-(3.8), we may say that if  $\zeta(i) > 0$ ,

$$E[\Delta\zeta(i)|\mathcal{F}_i] = p_z p_y - p_y + O(p_\zeta) = -\frac{9A^2}{(2B-A)^2} + O\left(\frac{\zeta}{2B-A}\right). \quad (5.26)$$

Now, before  $T$  we have

$$\begin{aligned} & \frac{9a^2}{(2b-a)^2} - \frac{9A^2}{(2B-A)^2} = -9 \left( \frac{A}{2B-A} - \frac{a}{2b-a} \right) \left( \frac{A}{2B-A} + \frac{a}{2b-a} \right) \\ & = -9 \left[ \frac{2b(\alpha + e_a) - 2ae_b}{n(2b-a)^2} + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2}\right) \right] \\ & \quad \times \left[ 2 \left( \frac{a}{2b-a} \right) + \frac{2b(\alpha + e_a) - 2ae_b}{n(2b-a)^2} + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2}\right) \right] \\ & = \frac{36a(ae_b - b\alpha - be_a)}{n(2b-a)^3} + \\ & \quad O \left[ \frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2} \cdot \left( \frac{a}{2b-a} + \frac{b(\alpha + e_a) + ae_b}{n(2b-a)^2} + \frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2} \right) + \left( \frac{ae_b - b\alpha - be_a}{n(2b-a)^2} \right)^2 \right] \\ & = \frac{36a(ae_b - b\alpha - be_a)}{n(2b-a)^3} + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2}\right). \end{aligned} \quad (5.27)$$

On the second line we have used Lemma 5.2 and the inequalities (5.1), (5.2) and in the last step we have cleaned up the big- $O$  using the facts

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

which follow from (5.15), (5.6), (5.7), (4.7) and (5.14). For every step  $j$ , we re-define  $T_j$  to be the 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  $j \leq i < T_j$ , using (5.26) and (5.27), we have

$$E[\Delta\zeta_j^+(i)|\mathcal{F}_i] = -\frac{9A^2}{(2B-A)^2} + \frac{9a^2}{(2b-a)^2} - \frac{1}{n}h'_j(t) + O\left(\frac{\zeta}{2B-A} + \frac{1}{n^2}h''_j(t)\right) \quad (5.28)$$

$$= \frac{36a(ae_b - b\alpha - be_a)}{n(2b-a)^3} - \frac{1}{n}h'_j(t) + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2} + \frac{\zeta}{2B-A} + \frac{1}{n^2}h''_j(t)\right) \quad (5.29)$$

$$\leq \frac{36a(af_b + bf_a)}{n(2b-a)^3} - \frac{1}{n}h'_j(t) + O\left(\frac{\alpha^2 + f_a^2 + f_b^2}{n^2(2b-a)^2} + \frac{f_\zeta}{n(2b-a)} + \frac{1}{n^2}h''_j(t)\right). \quad (5.30)$$

In the last line we have used (5.1), (5.2), (5.3), and the fact that  $\alpha \geq 0$ . Note that by (5.6), (5.7) we have  $af_b = o(bf_a)$  so

$$\begin{aligned} \frac{36a(af_b + bf_a)}{(2b-a)^3} &\leq \frac{36(1-t)^{\frac{3}{2}} \cdot 2(1-t) \cdot C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n \cdot (1+o(1))}{64(1-t)^3} \\ &\leq \left(\frac{9}{8}C_A + o(1)\right) n^{\frac{1}{2}} \log^{\frac{1}{2}} n (1-t)^{\frac{1}{4}} \end{aligned} \quad (5.31)$$

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

$$C_h > \frac{9}{8}C_A. \quad (5.32)$$

One can verify, as in Claim 5.3, that the big- $O$  term is dominated by the main terms. 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 [9]:

**Lemma 5.5.** *Let  $X_i$  be a supermartingale, with  $\Delta X_i \leq C$  for all  $i$ , and  $V(i) := \sum_{k \leq i} \text{Var}[\Delta X_k | \mathcal{F}_k]$*

*Then*

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

Referring to (4.8), before  $T$  we can put

$$\begin{aligned} \text{Var}[\Delta\zeta_j^+(i)|\mathcal{F}_i] &= \text{Var}[\Delta\zeta(i)|\mathcal{F}_i] \\ &\leq E\left[(\Delta\zeta(i))^2 \mid \mathcal{F}_i\right] \\ &= 1 \cdot p_z p_y + 1 \cdot p_y + O(p_\zeta) + O\left(\frac{1}{M}\right) \\ &\leq 3p_y \end{aligned} \quad (5.33)$$

since  $p_z \leq 1$  and  $p_\zeta = o(p_y)$  before  $T$ . Note that before  $T$ , we have

$$p_y = \frac{3Y}{2M} \leq \frac{3A}{2B-A} \leq \frac{3[n(1-t)^{\frac{3}{2}} + \alpha + f_a]}{4n(1-t) - 2f_b - n(1-t)^{\frac{3}{2}} - \alpha - f_a} \leq \left(\frac{C_T^{3/2} + C_\alpha}{C_T} + o(1)\right)(1-t)^{\frac{1}{2}}. \quad (5.34)$$

Indeed, if  $t \leq 1 - n^{-2/5} \log^{2/5} n$  then by (5.15), (5.6) and (5.7) we have

$$\begin{aligned} & \frac{3[n(1-t)^{\frac{3}{2}} + \alpha + f_a]}{4n(1-t) - 2f_b - n(1-t)^{\frac{3}{2}} - \alpha - f_a} \\ & \leq \frac{3[n(1-t)^{\frac{3}{2}} + C_\alpha(1-t)^{-1/2} \log n + C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n]}{4n(1-t) - 2C_B(1-t)^{-\frac{1}{2}} \log n - n(1-t)^{\frac{3}{2}} - C_\alpha(1-t)^{-1/2} \log n - C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n} \\ & = \frac{3n(1-t)^{3/2} + o(n(1-t))}{(4 + o(1))n(1-t)}. \end{aligned}$$

Whereas if  $t \geq 1 - n^{-2/5} \log^{2/5} n$  then

$$\begin{aligned} & \frac{3[n(1-t)^{\frac{3}{2}} + \alpha + f_a]}{4n(1-t) - 2f_b - n(1-t)^{\frac{3}{2}} - \alpha - f_a} \\ & \leq \frac{3[n(1-t)^{\frac{3}{2}} + C_\alpha n^{1/5} \log^{9/5} n + C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n]}{4n(1-t) + 2C_B n^{\frac{1}{5}} \log^{\frac{4}{5}} n \log(1-t) - n(1-t)^{\frac{3}{2}} - C_\alpha n^{1/5} \log^{9/5} n - C_A(1-t)^{\frac{3}{4}} n^{\frac{1}{2}} \log^{\frac{1}{2}} n} \\ & \leq \frac{(3 + o(1))[n(1-t)^{\frac{3}{2}} + C_\alpha n^{1/5} \log^{9/5} n]}{(4 + o(1))n(1-t)}. \end{aligned}$$

Here our choice of constants  $C_\alpha, C_T, C_{p_y}$  are such that

$$C_{p_y} > \frac{C_T^{3/2} + C_\alpha}{C_T}. \quad (5.35)$$

and so  $p_y \leq C_{p_y}(1-t)^{\frac{1}{2}}$ . Also, note that

$$\Delta\zeta^+ \leq 3$$

since  $\Delta\zeta \leq 2$  and  $\Phi(t) \leq \frac{9(1-t)^3}{9(1-t)^2} \leq 1$  (by (4.10), (4.7) and (5.14)).

Suppose condition (5.4) 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 Lemma 5.5 to the supermartingale  $\zeta_j^+$  to show this event has low probability (guiding our choice for  $\ell_j$ ). Note that by (5.33), in Lemma 5.5 we can plug in the following for  $v$ :

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

So the unlikely event that condition (5.4) triggers  $T$  has probability at most

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

by Lemma 5.5. The above bound holds for each of the  $O(n^2)$  pairs  $j, i$ , but note that it is with different parameters  $\ell_j(t)$  for each  $i$  ( $\ell_j(t)$  depends on  $t$  and therefore on  $i$ ). For the union bound

to work, we'd like to make the above probability  $o\left(\frac{1}{n^2}\right)$  for each pair  $j, i$ . Towards this end we consider 2 cases.

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

$$\frac{C_\ell^2}{6C_{p_y} + 6C_\ell} > 2. \quad (5.36)$$

If  $\left(1 - \frac{j}{n}\right)^{\frac{1}{2}} (i - j) > \log n$ , then again by (5.36), 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

$$\ell_j(t) := h_j(t) + C_\ell \max \left\{ \log n, \left(1 - \frac{j}{n}\right)^{\frac{1}{4}} (i - j)^{\frac{1}{2}} \log^{\frac{1}{2}} n \right\}. \quad (5.37)$$

With this choice, w.h.p.  $T$  is not triggered by condition (5.4).

#### 5.4 An upper bound on $\zeta$

In this section we'll show that  $T$  is not triggered by condition (5.3). We will see that (5.3) actually holds deterministically, assuming  $T$  is not triggered by the other conditions

**Lemma 5.6.** *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  $j \leq j' \leq j + C_x \left(1 - \frac{j}{n}\right)^{-\frac{3}{2}} \log n$ , and
2.  $\zeta(i) \leq 40C_\ell^2 \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}} \log n$  for all  $j \leq i \leq j' - 1$ .

*Proof.* Suppose  $\zeta(j - 1) = 0$ . Note that we then have  $\zeta(j) \leq 2$ . Recall the definition (4.10) of  $\Phi$  and equations (4.3) and (4.2). Now  $\Phi(t)/(1 - t)$  is decreasing since

$$\begin{aligned} \frac{d}{dt} \left( \frac{\Phi(t)}{1 - t} \right) &= \frac{\Phi'(t)}{1 - t} + \frac{\Phi(t)}{(1 - t)^2} \\ &= \frac{18aa'}{(2b - a)^2(1 - t)} - \frac{18a^2(2b' - a')}{(2b - a)^3(1 - t)} + \frac{9a^2}{(2b - a)^2(1 - t)^2} \\ &= \frac{18a}{(2b - a)^2(1 - t)} \left( -\frac{6a}{2b - a} + \frac{a \left(4 - \frac{6a}{2b - a}\right)}{2b - a} \right) + \frac{9a^2}{(2b - a)^2(1 - t)^2} \\ &= \frac{18a}{(2b - a)^2(1 - t)} \times \frac{-4a(a + b)}{(2b - a)^2} + \frac{9a^2}{(2b - a)^2(1 - t)^2} \\ &= \frac{9a^2}{(2b - a)^2(1 - t)} \left( -\frac{8a + 8b}{(2b - a)^2} + \frac{1}{1 - t} \right) \\ &= -\frac{9a^3(8b - a)}{(1 - t)^2(2b - a)^4} \leq 0, \end{aligned}$$

where we have used  $b = 2(1 - t)$  to get the last equation and (4.7) to justify the inequality.

Also, using (4.5) and the definition of  $b$ ,

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

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  $\ell_j(t)$  from (5.37), we have that for any  $i \geq j$  such that  $\zeta(j) \dots \zeta(i)$  are all positive,

$$\zeta(i) \leq \zeta(j) - \sum_{j \leq k < i} \Phi\left(\frac{k}{n}\right) + \ell_j(t) \quad (5.38)$$

$$\begin{aligned} &\leq 2 + \frac{1}{12}nx^2 - \left[ \frac{1}{6}n \left(1 - \frac{j}{n}\right) - C_h n^{\frac{1}{2}} \log^{\frac{1}{2}} n \left(1 - \frac{j}{n}\right)^{\frac{1}{4}} \right] x \\ &\quad + C_\ell \max \left\{ \log n, \left(1 - \frac{j}{n}\right)^{\frac{1}{4}} n^{\frac{1}{2}} x^{\frac{1}{2}} \log^{\frac{1}{2}} n \right\}. \end{aligned} \quad (5.39)$$

Define

$$x_j := C_x n^{-1} \left(1 - \frac{j}{n}\right)^{-\frac{3}{2}} \log n$$

and consider (5.39) for  $i$  such that  $x = x_j$ . Since  $C_x > 1$ , we have

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

so we can evaluate the ‘‘max’’ in  $\ell_j$ . 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 (5.39) gives

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

Recall that we have assumed that  $j < n - 2C_x^{\frac{2}{5}} n^{\frac{3}{5}} \log^{\frac{2}{5}} n$  and so

$$1 - \frac{j}{n} > 2C_x^{\frac{2}{5}} n^{-\frac{2}{5}} \log^{\frac{2}{5}} n.$$

Now

$$\frac{\left(\frac{C_x}{7} - C_\ell \sqrt{C_x}\right) \log n \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}}}{\frac{C_x^2}{12} n^{-1} \log^2 n \left(1 - \frac{j}{n}\right)^{-3}} = \frac{\frac{C_x}{7} - C_\ell \sqrt{C_x}}{\frac{C_x^2}{12}} \frac{n}{\log n} \left(1 - \frac{j}{n}\right)^{\frac{5}{2}} > \frac{\frac{C_x}{7} - C_\ell \sqrt{C_x}}{\frac{C_x}{12}}$$

and so  $\zeta(j + nx_j)$  is negative for this range of  $j$  since

$$\frac{C_x}{7} - C_\ell \sqrt{C_x} > \frac{C_x}{12}. \quad (5.40)$$

Therefore,  $\zeta$  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 (5.39) 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

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

This is maximised at  $x = 0$  or when  $x = n^{-1} \log n \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}}$  and for the latter we have

$$\begin{aligned} \zeta(i) &\leq \frac{1}{12}n^{-1} \left(1 - \frac{j}{n}\right)^{-1} \log^2 n - \frac{1}{7} \left(1 - \frac{j}{n}\right)^{1/2} \log n + C_\ell \log n \\ &\leq \frac{C_x^{-2/5}}{24} n^{-3/5} \log^{8/5} n - \frac{\sqrt{2}C_x^{1/5}}{7} n^{-1/5} \log^{6/5} n + C_\ell \log n \\ &\leq C_\ell \log n. \end{aligned}$$

This deals with  $x \leq n^{-1} \log n \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}}$ .

For  $x$  larger than that, we'll put

$$\begin{aligned} \zeta(i) &\leq \frac{1}{12}nx^2 - \frac{1}{7}n \left(1 - \frac{j}{n}\right) x + C_\ell \left(1 - \frac{j}{n}\right)^{\frac{1}{4}} n^{\frac{1}{2}} x^{\frac{1}{2}} \log^{\frac{1}{2}} n \\ &\leq \frac{C_x^2}{12} n^{-1} \log^2 n \left(1 - \frac{j}{n}\right)^{-3} + \frac{7C_\ell^2}{4} \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}} \log n \\ &< 40C_\ell^2 \left(1 - \frac{j}{n}\right)^{-\frac{1}{2}} \log n. \end{aligned}$$

To justify the second line we use the fact that  $x < x_j$  as well as the inequality  $-dx + cx^{1/2} \leq \frac{c^2}{4d}$  for real numbers  $x, c, d > 0$  to handle the last two terms. And for the third inequality, we use

$$C_x \leq 400C_\ell^2. \quad (5.41)$$

□

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 5.7.** *W.h.p. for all  $i \leq T$  we have  $\zeta(i) \leq C_\zeta n^{\frac{1}{5}} \log^{\frac{4}{5}} n$ .*

*Proof.* Suppose that at step  $j' \geq n - 2C_x^{\frac{2}{5}} n^{\frac{3}{5}} \log^{\frac{2}{5}} n$  we have  $\zeta(j') = 0$ . It follows from Lemma 5.6 that w.h.p. such a  $j'$  exists. Let  $i \geq j'$  such that  $\zeta(j' - 1) \dots \zeta(i)$  are all positive. Note that we again have the bound (5.39). But now  $0 \leq x \leq \frac{n-j'}{n} \leq 2C_x^{\frac{2}{5}} n^{-\frac{2}{5}} \log^{\frac{2}{5}} n$ , and (5.39) gives  $\zeta(i) \leq \left(\frac{1}{3}C_x^{\frac{4}{5}} + 2^{\frac{3}{4}}C_\ell C_x^{\frac{3}{10}} + o(1)\right) n^{\frac{1}{5}} \log^{\frac{4}{5}} n$ . Note that the term in square brackets in (5.39) is negative here.

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

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



since

$$C_\zeta > \max \left\{ 40C_\ell^2, \frac{1}{3}C_x^{\frac{4}{5}} + 2^{\frac{3}{4}}C_\ell C_x^{\frac{3}{10}} \right\}. \quad (5.42)$$

□

## 5.5 $T$ is not triggered by $B$

Recall from (4.1) that

$$e_b(i) = \sum_{j \leq i} (\mathbb{1}_{\sigma(i)=loop} - \mathbb{1}_{\delta(j)=\zeta}).$$

First we'll bound  $\sum_{j \leq i} \mathbb{1}_{\delta(j)=\zeta}$ . Define

$$B^-(i) := - \sum_{j \leq i} \mathbb{1}_{\delta(j)=\zeta} + \frac{1}{2}f_b(t).$$

Then

$$\begin{aligned} E[\Delta B^-(i)|\mathcal{F}_i] &= -\frac{2Z}{2M-1} \cdot \frac{\zeta}{2M-3} + \frac{1}{2n}f'_b(t) + O\left(\frac{1}{n^2}f''_b(t)\right) \\ &\geq -\frac{f_\zeta}{n(2b-a)} + \frac{1}{2n}f'_b(t) + O\left(\frac{1}{n^2}f''_b(t)\right). \end{aligned}$$

We have used  $2Z \leq 2M - 5$  to get the second line. This is valid as  $Y > 0$  up until time  $T$ . Here we see that if  $Y = 1$  we must have  $Z_1 \geq 2$ , unless  $G$  contains a copy of  $K_4$ , which happens with probability  $o(1)$ .

Note that by (4.7),

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

so  $B^-$  will be a submartingale since

$$C_B > \frac{4}{3}C_\zeta. \quad (5.43)$$

Indeed, it follows from (5.7) and (5.8) that if  $1-t > n^{-2/5} \log^{2/5} n$  then

$$\begin{aligned} E[\Delta B^-(i)|\mathcal{F}_i] &\geq -\frac{C_\zeta(1-t)^{-1/2} \log n}{n(2b-a)} + \frac{C_B(1-t)^{-3/2} \log n}{4n} + O\left(\frac{(1-t)^{-5/2} \log n}{n^2}\right) \\ &\geq -\frac{C_\zeta(1-t)^{-3/2} \log n}{3n} + \frac{C_B(1-t)^{-3/2} \log n}{4n} + O\left(\frac{(1-t)^{-5/2} \log n}{n^2}\right) \\ &\geq 0. \end{aligned}$$

And if  $1-t \leq n^{-2/5} \log^{2/5} n$  then

$$\begin{aligned} E[\Delta B^-(i)|\mathcal{F}_i] &\geq -\frac{C_\zeta n^{1/5} \log^{4/5} n}{n(2b-a)} + \frac{C_B n^{1/5} \log^{4/5} n}{2n(1-t)} + O\left(\frac{(1-t)^{-2} \log^{4/5} n}{n^{9/5}}\right) \\ &\geq -\frac{C_\zeta n^{1/5} \log^{4/5} n}{3n(1-t)} + \frac{C_B n^{1/5} \log^{4/5} n}{2n(1-t)} + O\left(\frac{(1-t)^{-2} \log^{4/5} n}{n^{9/5}}\right) \\ &\geq 0. \end{aligned}$$

We'll apply Lemma 5.5 to  $-B^-$ . Note that before  $T$  we can put

$$\begin{aligned} \text{Var}[\Delta B^-(i)|\mathcal{F}_i] &= \text{Var}[\mathbb{1}_{\delta(i)=\zeta}|\mathcal{F}_i] \leq E[\mathbb{1}_{\delta(i)=\zeta}^2] \leq p_\zeta \\ &\leq \frac{f_\zeta}{4n(1-t) - na - \alpha - f_a - f_b} \leq \frac{f_\zeta}{2n(1-t)} \end{aligned}$$

using (4.7) and  $\alpha, f_a, f_b = o(n(1-t))$  for  $t \leq T$ .

And therefore, referring to  $V(i)$  as in Lemma 5.5,

$$V(i) \leq \sum_{0 \leq k \leq i} \frac{f_\zeta}{2n(1-t)} \leq C_\zeta \begin{cases} 2(1-t)^{-\frac{1}{2}} \log n & : 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n. \\ n^{\frac{1}{5}} \log^{\frac{9}{5}} n & : \text{otherwise.} \end{cases}$$

So for  $v$  we will plug in

$$v = C_{v_B} \cdot \begin{cases} (1-t)^{-\frac{1}{2}} \log n & : 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n. \\ n^{\frac{1}{5}} \log^{\frac{9}{5}} n & : \text{otherwise.} \end{cases} \quad (5.44)$$

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

$$C_{v_B} \geq 2C_\zeta. \quad (5.45)$$

Note that

$$|\Delta B^-| \leq 1 + \frac{f'_b(t)}{2n} + O\left(\frac{1}{n^2} f''_b(t)\right) = 1 + o(1),$$

so by Lemma 5.5 the probability that  $-B^-(i) > \frac{1}{2}f_b(t)$  is at most

$$\begin{aligned} \exp\left\{-\frac{\frac{1}{4}f_b^2}{(2+o(1))\left[v + \frac{1}{2}f_b\right]}\right\} &\leq \\ \begin{cases} \exp\left\{-\frac{\frac{1}{4}C_B^2(1-t)^{-1} \log^2 n}{(2+o(1))\left[C_{v_B}(1-t)^{-1/2} \log n + \frac{1}{2}C_B(1-t)^{-1/2} \log n\right]}\right\} & 1-t > n^{-\frac{2}{5}} \log^{\frac{2}{5}} n. \\ \exp\left\{-\frac{\frac{1}{4}C_B^2 n^{2/5} \log^{8/5} n \log^2(1-t)}{(2+o(1))\left[C_{v_B} n^{1/5} \log^{9/5} n + \frac{1}{2}C_B n^{1/5} \log^{4/5} n \log(1-t)\right]}\right\} & 1-t \leq n^{-\frac{2}{5}} \log^{\frac{2}{5}} n. \end{cases} \end{aligned}$$

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

$$\frac{\frac{1}{4}C_B^2}{2\left[C_{v_B} + \frac{1}{2}C_B\right]} > 1. \quad (5.46)$$

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

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

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

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

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

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  $Z_0$  (or for paths of length 0 there is only one vertex which is in  $Y_0$ ). These vertices must either have been in  $Z_1$  and had a half-edge deleted to land in  $Z_0$ ; or been in  $Y_2$  and had a half-edge deleted to land in  $Y_1$ ; or been in  $Y$  and had a loop revealed. In the first two cases the endpoints were in  $\zeta$  and had a half-edge deleted.

So to bound the number of these paths, we bound the sum

$$\sum_j \mathbb{1}_{\delta(j)=\zeta} + \mathbb{1}_{\sigma(j)=loop} = O(\log n) + \sum_j \mathbb{1}_{\delta(j)=\zeta}$$

w.h.p.. Note that in light of Section 5.5, we have the bound

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

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

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

by (5.15), and

$$na\left(\frac{T}{n}\right), f_a\left(\frac{T}{n}\right) = O\left(n^{1/5} \log^{9/5} n\right). \quad (6.1)$$

For (6.1) we use the bound on  $T$  in Theorem 5.1 and the inequalities (4.7) and (5.6).

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. ( $A$  can also decrease when a loop is found, but this only happens with small probability.) Also, the variable  $A$  is non-increasing. 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  $\Gamma$ . We point out that  $\Gamma$  is indeed distributed as a random 2-regular graph with the standard configuration model since each remaining vertex lies in  $Z$  and thus has two remaining configuration points whose edges have not yet been revealed. Let  $\nu_T \leq \sum_j \mathbb{1}_{\delta(j)=\zeta}$  be the number of paths in the 2-matching at time  $T$ . Then to bound the number of paths in the final 2-matching, we only have to add to  $\nu_T$  a bound on the number of vertices in  $\Gamma$  that are unsaturated by the matching (i.e. the number of odd cycles) in the remaining 2-regular graph  $\Gamma$ . Now, it is well-known that w.h.p. a random 2-regular graph has at most  $O(\log n)$  cycles total. (The calculations at the

end of this section show how the argument for this goes.) But the sum  $\sum_j \mathbb{1}_{\delta(j)=\zeta}$ , and therefore the number of paths in the final 2-matching, is w.h.p.  $O\left(n^{\frac{1}{5}} \log^{\frac{9}{5}} n\right)$ .

Now we bound the number of cycles in the final 2-matching. Note that at any time a single edge is added to  $U$ , 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 the  $c_j$  are independent Bernoulli random variables with

$$c_j = \begin{cases} 1 & : \text{with prob. } \frac{1}{j}. \\ 0 & : \text{with prob. } \frac{j-1}{j}. \end{cases} \quad (6.2)$$

So if we define the martingale

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

then we have  $\text{Var}[\Delta C(i)] = \frac{i}{(i+1)^2}$ , and note  $\sum_{i=1}^{3n} \frac{i}{(i+1)^2} = O(\log(3n))$ . Now, applying Lemma 5.5 to  $C(i)$  shows that w.h.p. it is always at most  $O(\log^{\frac{3}{4}} n)$ , and since  $E[C] = O(\log n)$ , we have that  $C = O(\log n)$  w.h.p. (While Lemma 5.5 may not be the simplest tool, it is to hand and we can use it).  $\square$

## 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  $\zeta$  becomes large and stays large for a significant amount of time. In this case, the algorithm will likely delete an edge adjacent to a  $\zeta$  vertex. Recall that the deletion of an edge adjacent to a  $\zeta$  vertex results in an additional component (see Section 4.5). 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  $\Omega\left(n^{\frac{1}{5}} \log^{-16/5} n\right)$  components.*

*Proof.* We show that  $\zeta$  stochastically dominates a suitably defined random walk and then apply the central limit theorem for i.i.d. sequences (see for example [8]).

**Lemma 7.2.** *Let  $X_1, X_2, \dots$  be independent and identically distributed random variables with  $E[X_i] = \mu$  and  $\text{Var}[X_i] = \sigma^2 \in (0, \infty)$ . If  $S_n = X_1 + \dots + X_n$ , then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

*the standard normal distribution.*

Let

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

In this section we will consider steps from

$$i_0 = n - n^{3/5} \text{ to } i_{end} = n - n^{3/5} + n^{3/5} \log^{-1} n \leq n - \frac{1}{2}n^{3/5}.$$

From Theorem 5.1, w.h.p.,  $T$  occurs after this time frame. Hence we have dynamic concentration on our variables. In this range,  $1 - t = \Theta(n^{-2/5})$ , so  $f_a, f_\zeta = O(n^{1/5} \log^{4/5} n)$  and  $f_b, \alpha = O(n^{1/5} \log^{9/5} n)$  and by (4.7) and (4.2),  $a(t) = \Theta(n^{-3/5}), b(t) = \Theta(n^{-2/5})$ . So using Lemma 5.2 we can say that in this range,

$$\begin{aligned} p_y(i) &= \frac{3(A(i) - \zeta(i))}{2B(i) - A(i)} \\ &= \frac{3a(i/n) + O(n^{-4/5} \log^{9/5} n)}{2b(i/n) - a(i/n) + O(n^{-4/5} \log^{9/5} n)} \\ &= w(i) - O(n^{-2/5} \log^{9/5} n). \\ p_\zeta(i) &= \frac{\zeta(i)}{2B(i) - A(i)} = O(n^{-2/5} \log^{9/5} n). \\ p_z(i) &= 1 - p_y(i) - p_\zeta(i). \end{aligned}$$

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

$$X(i) = \begin{cases} 1 & \text{with prob. } w(i) - Ln^{-2/5} \log^{9/5} n. \\ 0 & \text{with prob. } 1 - 2w(i) - Ln^{-2/5} \log^{9/5} n. \\ -1 & \text{with prob. } w(i) - Ln^{-2/5} \log^{9/5} n. \\ -2 & \text{with prob. } 3Ln^{-2/5} \log^{9/5} n. \end{cases} \quad (7.1)$$

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

$$p_z(i)p_y(i) \geq w(i) - Ln^{-2/5} \log^{9/5} n, \quad p_z(i)p_y(i) + p_z(i)^2 \geq 1 - w(i) - 2Ln^{-2/5} \log^{9/5} n$$

and

$$p_z(i)p_y(i) + p_z(i)^2 + p_y(i) \geq 1 - 3Ln^{-2/5} \log^{9/5} n.$$

Note that this is achievable since in this range  $p_z(i) = 1 - O(n^{-1/5}), p_y(i) = \Omega(n^{-1/5})$  and  $p_\zeta(i) = O(n^{-2/5} \log^{9/5} n)$ .

In this case,  $\Delta\zeta(i)$  stochastically dominates  $X(i)$ . This follows from (4.8) in the case when  $\zeta > 0$  and trivially when  $\zeta = 0$  in the sense that when  $\zeta = 0$ ,  $\Delta\zeta \geq 1$  with probability  $\approx p_z = 1 - o(1)$ . Here we need to use the fact that  $M = \Theta(n^{3/5})$  in our range.

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

$$\mu = E[X(i)] = -6Ln^{-2/5} \log^{9/5} n$$

and

$$\sigma^2 = \text{Var}[X(i)] = 2w(i) - 10Ln^{-2/5} \log^{9/5} n - 36L^2 n^{-4/5} \log^{18/5} n.$$

We will split the time range  $i_0$  to  $i_{end}$  into  $d = \log n$  many intervals of length  $\approx n^{3/5} \log^{-2} n$ . 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 random walk starting at  $i_\ell + 1$  and ending at  $i_{\ell+1}$  and let

$$S_\ell = \sum_{i=i_\ell+1}^{i_{\ell+1}} X(i).$$

We note here that these  $d$  random walks are independent, identically distributed copies of the same random walk. So for  $0 \leq \ell < d$  we have

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

and that

$$\sigma \cdot \sqrt{i_{\ell+1} - i_\ell} = \Theta \left( n^{1/5} \log^{-1} n \right)$$

because  $\sigma^2 = \Theta(w(i)) = \Theta(n^{-1/5})$  and  $i_{\ell+1} - i_\ell = n^{3/5} \log^{-2} n$ . Note that there exists an absolute constant  $c$  such that  $\sigma \cdot \sqrt{i_{\ell+1} - i_\ell} \leq cn^{1/5} \log^{-1} n$  for all  $0 \leq \ell < d$ .

Hence applying Lemma 7.2 to  $\sum_{i=i_\ell+1}^{i_{\ell+1}} X(i)$ , we see that

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

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

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

So we get that

$$\begin{aligned} \mathbb{P} \left[ \forall 0 \leq \ell < \frac{1}{2}d, \zeta(i_{\ell+1}) \leq n^{1/5} \log^{-1/5} n \right] &\leq \mathbb{P} \left[ \forall 0 \leq \ell < \frac{1}{2}d, \sum_{i=i_\ell+1}^{i_{\ell+1}} X(i) \leq n^{1/5} \log^{-1/5} n \right] \\ &\leq (1 - p_0)^{\frac{1}{2} \log n} \\ &= o(1). \end{aligned}$$

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

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

which is a martingale since  $E[X(i)|\mathcal{F}_{i-1}] = E[X(i)]$ . We use the language of martingales here only because Lemma 5.5 is already at hand. Let  $i_c = i_b + n^{3/5} \log^{-3} n < i_{end}$ . Then

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

By applying Lemma 5.5 to the negative of this martingale, we have that after  $n^{3/5} \log^{-3} n$  steps,

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

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

$$p_z(i)p_\zeta(i) \geq q_0 = (1 - o(1)) \cdot \frac{\zeta}{2M} = \Omega\left(n^{-2/5} \log^{-1/5} n\right)$$

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  $\text{Bin}(n^{3/5} \log^{-3} n, q_0)$ .

$$E[\text{Bin}(n^{3/5} \log^{-3} n, q_0)] = \Omega\left(n^{1/5} \log^{-16/5} n\right),$$

so an application of the Chernoff bound tells us that, w.h.p.,  $W = \Omega\left(n^{1/5} \log^{-16/5} n\right)$ . □

## 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] R. Durrett, *Probability: Theory and Examples*. Fourth Edition. Cambridge University Press.
- [9] David A. Freedman, *On Tail Probabilities for Martingales*, The Annals of Probability 3 (1975), 100-118.
- [10] A.M. Frieze, *On a Greedy 2-Matching Algorithm and Hamilton Cycles in Random Graphs with Minimum Degree at Least Three*, Random Structures and Algorithms 45 (2014) 443-497.
- [11] A.M. Frieze, *Finding Hamilton cycles in sparse random graphs*, Journal of Combinatorial Theory B, 44 (1988) 230-250.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] M. Lelarge, *A New Approach to the Orientation of Random Hypergraphs*, Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (2012) 251-264.
- [16] R.W. Robinson and N.C. Wormald, *Almost all cubic graphs are Hamiltonian*. Random Structures and Algorithms **3** (1992) 117-125.
- [17] R.W. Robinson and N.C. Wormald, *Almost all regular graphs are Hamiltonian*. Random Structures and Algorithms **5** (1994) 363-374.
- [18] J. Salez, *Weighted enumeration of spanning subgraphs in locally tree-like graphs*. Random Structures Algorithms **43** (2013) 377-397.