Near–perfect non-crossing harmonic matchings in randomly labeled points on a circle

József Balogh[1†] and Boris Pittel[1‡] and Gelasio Salazar[2§]

[1] *Department of Mathematics, The Ohio State University, Columbus, OH 43210*
[2] *Instituto de Física, UASLP. Alvaro Obregon 64, San Luis Potosi, SLP, 78210 Mexico*

Consider a set S of points in the plane in convex position, where each point has an integer label from $\{0, 1, \ldots, n-1\}$. This naturally induces a labeling of the edges: each edge (i, j) is assigned label $i + j$, modulo n. We propose the algorithms for finding large non–crossing *harmonic* matchings or paths, i. e. the matchings or paths in which no two edges have the same label. When the point labels are chosen uniformly at random, and independently of each other, our matching algorithm with high probability (w.h.p.) delivers a nearly–perfect matching, a matching of size $n/2 - O(n^{1/3} \ln n)$.

Keywords: Graceful, harmonious labeling, noncrossing, harmonic graph, convex position, matching, algorithm, average case behavior

---

1 Introduction

We are motivated by the concepts of graceful labelings and harmonious graphs introduced by Graham and Sloane [6] (see [5] for a comprehensive survey on these problems). Our interest is in the problem of existence of the large substructures (subsets of edges or subgraphs) such that all the edges involved have different labels. Typically, an edge label is a function of the labels of the endvertices, e.g. the absolute value of their difference (graceful labelings), or their sum modulo some n (harmonious graphs).

For the point set in the plane it is natural to seek the large substructures (matchings, trees) that meet certain geometric conditions. One popular *non–crossing* condition requires that no two edges in the substructure cross each other. For a sample of diverse results in this area of combinatorial geometry we refer the reader to see [1, 2, 7, 8, 9].

To describe the results of this paper, we need some terminology and notations. Following [3], let S be a set of points in the plane in a convex position. Assume that each point has an integer label from $\{0, \ldots, n-1\}$. If p, q are distinct points (also called *vertices*) in S, then we let (p, q) denote the straight segment (or *edge*) that has p and q as its endvertices. This naturally induces a (complete) *geometric graph* G_S. In general, we let $E(K)$ denote the set of edges of a graph K. A subset E' of $E(G_S)$ is *non–crossing* if no two edges in E' intersect in a point other that a common endvertex. A subgraph H of G_S is *non–crossing* if $E(H)$ is non–crossing.

As for the edge labels, we use the sum rule; it assigns to each edge (p, q) a number equal to the sum of labels of p and q modulo n. One such rule assigns to each edge the sum (modulo n) of the labels of its endpoints. In this geometric setting, the central problem is to find conditions for existence of large non–crossing subgraphs whose edge labels are all distinct.

While the paper [3] dealt exclusively with the worst–case instances of the labeled set S, our goal is to study the average (likely) case behavior under assumption that the labels of points in S are random. More specifically, we assume that each of the n points is labeled with an integer drawn uniformly at random from $\{0, 1, 2, \ldots, n-1\}$, independently of all other labels.
How many edges are there typically in a maximum size harmonic non–crossing matching in G_S?

---

† Supported by NSF Grant DMS–0302804.
‡ Supported by NSF Grant DMS–0406024.
§ Supported by FAI–UASLP

We have found a greedy matching algorithm (HMATCHING) that w.h.p. delivers a nearly perfect matching, of size $n/2 - O(n^{1/3}\ln n)$, with the number of unmatched vertices of likely order $n^{1/3}\ln n$, at most. Thus the maximum matching number w.h.p. is $n/2 - O(n^{1/3}\ln n)$ at least. The matching process works by going through the cyclically ordered sequence of vertices, starting from an arbitrary initial point. The probability that the resulting matching is perfect is not too small, of order $\Omega(n^{-1/3}\ln^{-1} n)$ at least, i. e. the expected number of the "lucky" starting points is $\Omega(n^{2/3}\ln^{-1} n)$. We conjecture that the number itself is likely to be that large as well, so that w.h.p. there exists a perfect matching! In Section 2 we present HMATCHING, and in Section 3 we give the experimental results that allowed us to predict the likely behavior of the algorithm. In Section 4 we provide a rigorous analysis which confirms—within the logarithmic factors— the conjectured bounds. Despite appearance of the fractional powers $1/3, 2/3$, that are also prevalent in the asymptotic behavior of the near-critical random graph $G(n,m)$, Bollobás [4], we do not see any connection between the two schemes.

How many vertices are there typically in a maximum size harmonic non–crossing tree or forest in G_S?

We conjecture, that the answer is $(1-\varepsilon_n)n$, where $\varepsilon_n \to 0$.

2 HMATCHING: the algorithm

Recall that we are assuming that we have a collection S of n points in convex position. No relevant geometrical information is lost if we assume that all the points lie on a circle. Therefore, we may denote the points as $p_0, p_1, \ldots, p_{n-1}$, according to the cyclic (counter-clockwise) order in which they appear on the circle. Further each point p_i gets a label $A[i]$, and the n labels are drawn independently from the uniform distribution on $\{0,1,\ldots,n-1\}$. Given the point labels, each edge (p_i, p_j) gets the label $A[i,j] :\equiv A[i] + A[j] (\mathrm{mod}\, n)$.

HMATCHING takes as the input an array $(A[0], A[1],\ A[2],\ \ldots, A[n-1])$ and its output is a (non–crossing, harmonic) matching on S. At each step we have a current matching, both non–crossing and harmonic , to which we add a new edge to get a larger matching that meets the same requirements. Formally, we maintain the current matching $\mathcal{M}$ as a collection of ordered pairs (i,j) with $i<j$, where (i,j) represents (p_i,p_j). Clearly the edge set $\mathcal{M}$ satisfies the following conditions:

(a) if (i,j) and (i',j') are different pairs in $\mathcal{M}$, then $\{i,j\}\cap\{i',j'\}=\emptyset$ ($\mathcal{M}$ is a matching);

(b) if (i,j) and (i',j') are different pairs in $\mathcal{M}$, then $A[i]+A[j] \not\equiv A[i']+A[j'] (\mathrm{mod}\, n)$ ($\mathcal{M}$ is harmonic);

(c) if (i,j) and (i',j') are different pairs in $\mathcal{M}$, with $i<i'$, then either $i<j<i'<j'$ or $i<i'<j'<j$ ($\mathcal{M}$ is non–crossing).

The pseudocode for HMATCHING is the following.

Input : An array $(A[0], A[1], \ldots, A[n-1])$, such that $A[i] \in \{0,1,\ldots,n-1\}$ for every i.
Output : The size of a set $\mathcal{M}$ of pairs (i,j), with $i<j$, that satisfies (a), (b), and (c).
Procedure :

```
S = ∅; M = ∅; L = 0; k = 0
while k ≤ n − 1
      do
            if S ≠ ∅
                  then if A[max S] + A[k](mod n) ∉ L
                              then M ← M ∪ {(max S, k)}
                                    L ← L ∪ {A[max S] + A[k](mod n)}
                                    S ← S \ {max S}
                              else   S ← S ∪ {k}
                  else S ← {k}
            k ← k + 1
return |L|
```

The action of the algorithm is illustrated in Figure 1.

In this example, $A[0]=7, A[1]=6, A[2]=4, \ldots, A[9]=3$. In the first step we explore $A[0]$ and add 0 to S. In the second step, we explore $A[1]$, and check if $A[0]+A[1]$ (mod 10) is in L. Since it is not, the edge $(A[0], A[1])$ is added to $\mathcal{M}$, and since $A[0]+A[1]\ = 7+6\ \equiv 3$ (mod 10), L becomes $\{3\}$, and S goes back to $\emptyset$. In the third step we explore $A[2]$, and since there is no stack, we add 2 to S (so that S becomes $\{2\}$, since it was empty) and move on to the fourth step, where we explore $A[3]$; since $A[2]+A[3]=4+9\equiv 3$ (mod 10) is already in L, we must now set $S=\{2,3\}$. In the fifth step we explore $A[4]=5$. Since 3 is the largest integer in S, we check if $A[3]+A[4]\equiv 4$ (mod 10) is in L. Since

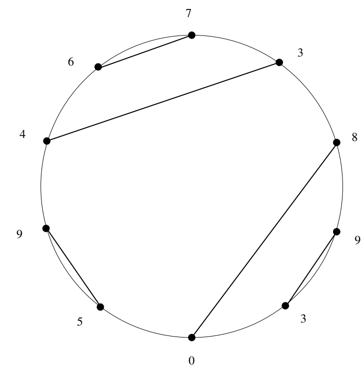

Fig. 1: Illustration of HMATCHING.

it is not, then we add $(A[3], A[4])$ to $\mathcal{M}$, 4 to L, and remove 3 from S. At the end, we obtain the matching shown, which happens to be perfect.

3 Performance of HMATCHING: empirical results

There are two natural parameters to measure the performance of HMATCHING: (i) the expected size of the matching obtained by running HMATCHING, and (ii) the probability that HMATCHING delivers a perfect matching.

(i) We ran our algorithm 10^6 times for each of the following values of n: $5000, 10000, 15000, 20000$, 25000, $30000, 35000, 40000, 45000$, and 50000. For each such n, we computed the average of the 10^6 experiments. Using `Gnuplot`©, we plotted the results and obtained a curve $n^{1/3}/1.46$ that fitted the data quite well. In view of our experiments, we conjecture that the expected number of vertices left unmatched is $\Theta(n^{1/3})$, or in other words, the expected size of the matching is $n/2 - \Theta(n^{1/3})$.

(ii) We ran 10^6 experiments for each $n = 5000, 10000, \dots, 50000$, and computed the proportion of experiments for which HMATCHING yielded a perfect matching. The data fit the curve $n^{-1/3}$ so well that we are led to the conjecture: the probability that HMATCHING delivers a perfect matching is of order $\Theta(n^{-1/3})$. It is tempting to state an even stronger conjecture: the probability that the resulting matching is perfect is asymptotic to $n^{-1/3}$. In the next section we prove a slightly weaker result, namely that this probability is between $c_1 n^{-1/3} \ln^{-1} n$ and $c_2 n^{-1/3} \ln n$. We also show that the likely size of the terminal matching is between $n/2 - c_3 n^{1/3} \ln n$ and $n/2 - c_4 n^{1/3} \ln^{-1} n$, which again is within the logarithmic factors from the conjectured formula $n/2 - \Theta(n^{2/3})$. Consequently, on average, the number of the starting points for which the algorithm finds a perfect matching is of an empirical order $\Theta(n^{2/3})$, and of a provable order $\Omega(n^{2/3} \ln^{-1} n)$. This suggests the following.

Conjecture 1 *W.h.p. there is a perfect (non–crossing, harmonic) matching, and it can be found by running* HMATCHING n *times, selecting each of the* n *points as a starting point.*

In our computer experiments, with n up to 10^5 and 10^6 problem instances, we always found a perfect matching by running the algorithm for sufficiently many starting points.

4 Analysis of HMATCHING

4.1 The matching algorithm as a Markov Chain.

Consider the generic, k-th, step of the matching algorithm. Before this step the vertices $p_1, \dots, p_{k-1}$ have been explored, and some of them have been matched. Let $\mathcal{M}$ be the current (non-crossing, harmonious) matching and S be the current set (stack) of all unmatch ed points whose labels have been explored. Then $2|\mathcal{M}| + |S| = k - 1$. Suppose first that $S \neq \emptyset$. Assume inductively that there are no triples (p_a, p_b, p_c), $a < b < c$, such that $(p_a, p_c) \in \mathcal{M}$ and $p_b \in S$. This condition means that no edge (p_a, p_b), such that $p_a \in S$ and $b > b^* = \max\{c : p_c \in S\}$, crosses an edge from $\mathcal{M}$. In particular, we can and do add to $\mathcal{M}$ the edge (p_{b^*}, p_k) if the label of this edge is not in L, the label set of the edges in $\mathcal{M}$, i. e. if $A[b^*] + A[k] (\mathrm{mod}\, n) \notin L$. The last condition restricts the value $A[k]$ to a subset of $\{0, \dots, n-1\}$ of cardinality $n - |L| = n - |\mathcal{M}|$. Since $A[k]$ is uniform on $\{0, \dots, n-1\}$, and

independent on $A[0],\dots,A[k-1]$, the (conditional) probability that (p_{b^*},p_k) is added to $\mathcal{M}$ in the k–th iteration step is $1-|L|/n=1-|\mathcal{M}|/n$. In this case $\mathcal{M}+\{(p_{b^*},p_k)\}$ and $S\setminus\{p_{b^*}\}$ are the next matching set and the next stack respectively. Alternatively, with the probability $|\mathcal{M}|/n$ the matching set remains the same, but the stack grows to $S\cup\{p_k\}$. If $S=\emptyset$, then the matching set $\mathcal{M}$ remains the same, and the next S is $\{p_k\}$. In all cases the new matching $\mathcal{M}$ and the new stack S meet the same non-crossing condition as the previous $\mathcal{M}$ and S. Clearly the sequence $\{\mathcal{M}_k,S_k\}_{k\le n}$, $(\mathcal{M}_0=\emptyset,\ S_0=\emptyset)$, is a Markov chain. The chain terminates once $2|\mathcal{M}_k|+|S_k|$ reaches n, that is when there are no unexplored points left. Remarkably, the transition probabilities and the termination rule depend only on $|\mathcal{M}_k|$. So there is a reduction of $\{\mathcal{M}_k,S_k\}$ to a much simpler Markov chain $\{m_k,s_k\}$ on the set of pairs (m,s), $m=|\mathcal{M}|$, $s=|S|$, with termination condition $2m_k+s_k=n$.
Here is the formal definition of the reduced Markov chain.
Markov Process 1 (MP$_1$) *Each state is a pair (m,s), where m and s are nonnegative integers, and $2m+s<n$, where n is a fixed integer given in advance. The initial state is $(0,0)$. The transition rules are :*
If $s=0$, then the next state is

$$(m,s+1)=(m,1).$$

If $s>0$, then the next state is

$$\begin{aligned}(m+1,s-1), &\quad \text{with probability } 1-m/n,\\ (m,s+1), &\quad \text{with probability } m/n.\end{aligned}$$

4.2 The likely size of the terminal matching.

According to our reduction, to study the size of the terminal matching is the same as to study Z_n, the terminal value of m in the Markov chain MP$_1$.

Theorem 2 *1. Given $a>0$, set $\alpha=2\sqrt{a(1+a)}$.*

$$Pr(Z_n>n/2-\alpha n^{1/3}\ln n)=1-O(n^{-a}). \tag{1}$$

2.

$$Pr(Z_n\le n/2-n^{1/3}\ln^{-2}n)=1-O(\ln^{-1}n). \tag{2}$$

3. Let $P_n=Pr(Z_n=n/2)$, n even, and $P_n=Pr(Z_n=(n-1)/2)$, n odd. Then, for some constants $\alpha,\beta>0$,

$$\alpha n^{-1/3}\ln^{-1/2}n\le P_n\le\beta n^{-1/3}\ln n. \tag{3}$$

For the proof we need the following statements.

Proposition 3 *Define the random variable $X_{p,t}$ as the time it takes for the $(p,1-p)$–random walk to reach the zero state from the state t. Then, for all $r>0$,*

$$Pr(X_{p,t}\ge r)\le\frac{1}{(2p)^t(4p(1-p))^{-r/2}}.$$

Proposition 4 *For each even $t\ge 0$, let $R_p(t)$ denote the probability of being at 0 at the time step t in the (p,q)–random walk on $\{0,1,2,\dots\}$ with the repellent 0 state. Then*

$$\begin{aligned}R_{1/2}(t)&\ \sim\ c_1t^{-1/2},\ \text{and}\\ R_p(t)&\ <\ 3(1-2p)+(\pi t)^{-1/2},\ \text{ if }\ p<1/2.\end{aligned}$$

Lemma 5 *Let $a>0$. With probability $1-O(n^{-a})$, there exists k such that*

$$m_k\in(n/2-(1+a)n^{2/3}\ln n,\ n/2-0.5(1+a)n^{2/3}\ln n),\quad s_k=0,$$

with $\Theta(n^{2/3}\ln n)$ points remaining to be explored.

Proof of Lemma 5. Given $m < n/2$, let $T_m = \min\{k : m_k = m\}$ and set $T_m = n$, if no such k exists. Introduce $H_m = s_{T_m}$, the stack size at this moment. By the definition of MP_1, for $j < k$ and $s_j > 0$, the conditional probability of the transition $(m_j, s_j) \to (m_{j+1}, s_{j+1}) = (m_j, s_j + 1)$, which leads to increase of the stack by 1, is m/n at most. And the alternative transition leads to the stack size $s_j - 1$. For $s_j = 0$, we have $s_{j+1} = 1$. These observations imply that H_m is stochastically dominated by W_m, the maximum of the simple asymmetric random walk $\{\xi_j\}_{j \le n}$ on $\{0, 1, 2, \dots\}$, defined as follows: $\xi_0 = 0$,

$$\begin{aligned}
&\Pr(\xi_{j+1} = \xi_j + 1 \,|\, \xi_j) = p := m/n, \quad (\xi_j \ge 1),\\
&\Pr(\xi_{j+1} = \xi_j - 1 \,|\, \xi_j) = q := 1 - m/n, \quad (\xi_j \ge 1),\\
&\Pr(\xi_{j+1} = 1 \,|\, \xi_j = 0) = 1.
\end{aligned}$$

Furthermore, for each integer $w > 0$, $\Pr(W_m > w) \le n\Pr(\mathcal{W}_m > w)$, where $\mathcal{W}_m$ is the maximum of ξ_j for j between 0 and the first moment $t > 0$ when $\xi_t = 0$. Using the classic gambler's ruin formula, we have

$$\Pr(\mathcal{W}_m > w) = \frac{q/p - 1}{(q/p)^{w+1} - 1} \le (p/q)^w.$$

Then, introducing $m_i = \frac{n}{2} - [a_i n^{2/3} \ln n]$ and $p_i = m_i/n$, $i = 1, 2$, with $a_2 = a_1/2$, we have

$$\begin{aligned}
\Pr\bigl(W_{m_i} > n^{1/3}\bigr) &\le 2n \left(\frac{m_2}{n - m_2}\right)^{n^{1/3}} < 3n\bigl(1 - 4a_2 n^{-1/3} \ln n\bigr)^{n^{1/3}}\\
&< 3n \exp(-4a_2 \ln n) = \frac{3}{n^{2a_1 - 1}} \to 0,
\end{aligned}$$

provided $a_1 > 1/2$.
Now, since $2m_k + s_k = k$ at each step, we have

$$m = \frac{T_m - H_m}{2} \ge \frac{T_m - W_m}{2}.$$

Applying this to $m = m_1, m_2$, we see that

$$\Pr\left\{\bigcap_{i=1}^{2} \bigl(2m_i \le T_{m_i} \le 2m_i + n^{1/3} \text{ and } H_{m_i} \le n^{1/3}\bigr)\right\} \ge 1 - O(n^{-a}), \quad a = 2a_1 - 1.$$

Therefore

$$\Pr\left\{(T_{m_2} - T_{m_1} = a_1 n^{2/3} \ln n + O(n^{1/3})) \cap (H_{m_1} \le n^{1/3})\right\} \ge 1 - O(n^{-a}).$$

Denote the event in this bound by A. Let

$$B = A \cap \{s_k \text{ becomes zero at some } k \in [T_{m_1}, T_{m_2}]\}.$$

We want to show that $\Pr(A \setminus B) \le n^{-b}$, $\forall\, b > 0$, for n large enough. Let $t_1 \in [0, n-1]$. Suppose that $m_{t_1} \le m_2$, and $0 < s_{t_1} \le n^{1/3}$. These conditions certainly hold if $t_1 = T_{m_1}$. Let $\mathcal{T} = \mathcal{T}(t_1)$ be the first $t > t_1$ such that either $m_t = m_2$, or $s_t = 0$. As before, $\{s_t\}_{t < \mathcal{T}}$ is dominated by the asymmetric walk $\{\xi_j\}_{j \ge t_1}$, $\xi_{t_1} = \lfloor n^{1/3} \rfloor$, with $p = m_2/n$. Therefore $\mathcal{T} - t_1$ is dominated by $X_{p, \lfloor n^{1/3} \rfloor}$, where $X_{p,s}$ is the first time the random walk hits 0, if $\xi_0 = s$. Since by Proposition 3

$$\Pr(X_{p,s} \ge r) \le \frac{(4pq)^{-r/2}}{(2p)^s},$$

it follows that

$$\Pr\Bigl(X_{\frac{m_2}{n}, \lfloor n^{1/3} \rfloor} \ge \lfloor n^{2/3} \rfloor\Bigr) \le \frac{\left(1 - \frac{4\lfloor a_2 n^{2/3} \ln n \rfloor}{n^2}\right)^{\lfloor n^{2/3} \rfloor}}{\left(1 - \frac{2\lfloor a_2 n^{2/3} \ln n \rfloor}{n}\right)^{\lfloor n^{1/3} \rfloor}} \le \exp(-a_2 \ln^2 n).$$

Therefore $\mathcal{T}(t_1) - t_1 \le n^{2/3}$ quite surely (q.s. in short), i.e. with probability $1 - n^{-b}$, for every $b > 0$, uniformly for all t_1. Thus $\mathcal{T}(T_{m_1}) - T_{m_1} \le n^{2/3}$ q.s. as well. Since $T_{m_2} - T_{m_1}$ is of order $n^{2/3}\ln n \gg n^{2/3}$ on A, we conclude that indeed $\Pr(A \setminus B) \le n^{-b}$, for every $b > 0$. So the Markov process $\{m_k, s_k\}$ reaches a state $(m_0, 0)$, where $n/2 - m_0 \in (0.5a_1 n^{2/3}\ln n, a_1 n^{2/3}\ln n)$, with probability $1 - O(n^{-a})$, $a = 2a_1 - 1$. ∎

Proof of Theorem 2, part 1. Let T be the first k such that

$$m_k \in (n/2 - (1+a)n^{2/3}\log n,\ n/2 - 0.5(1+a)n^{2/3}\log n), \quad s_k = 0.$$

By Lemma 1, T is well defined with probability $1 - O(n^{-a})$. Let ℓ be the number of the remaining unexplored points after T steps; clearly

$$(1+a)n^{2/3}\ln n \le \ell \le 2(1+a)n^{2/3}\ln n.$$

The additional increase of m_k during the remaining $n - T$ steps is $(\ell - s_n)/2$, where s_n is the terminal stack size. So $Z_n = m_n$ is given by

$$Z_n = \frac{n-\ell}{2} + \frac{\ell - s_n}{2} = \frac{n}{2} - 0.5 s_n.$$

Thus we need to show that w.h.p. $s_n = O(n^{1/3}\log n)$. Since $m_k \le n/2$ for all k, s_n is dominated ξ_ℓ, where $\{\xi_j\}$ is the simple symmetric random walk with $p = q = 1/2$, and $\xi_0 = 0$. We need to find a likely upper bound for ξ_ℓ. First of all, for each integer $x \ge 0$,

$$\Pr(\xi_\ell = x) = \sum_{2t+\mu=\ell} \mathcal{P}_t \mathcal{Q}_\mu(x); \tag{4}$$

here $\mathcal{P}_t = \Pr(\xi_{2t} = 0)$, the probability that the walk returns to 0 after $2t$ steps; $\mathcal{Q}_\mu(0) = \delta_{\mu,0}$, and $\mathcal{Q}_\mu(x)$, $x > 0$, is the probability that the walk, that starts at 0, reaches x after μ steps without ever returning to 0. We will need the full strength of this formula later, but for now we are content with its weak corollary, namely

$$\Pr(\xi_\ell = x) \le \sum_{\substack{\mu,t\ge 0\\ 2t+\mu=\ell}} \mathcal{Q}_\mu(x). \tag{5}$$

As for $\mathcal{Q}_\mu(x)$, recall that, by the ballot theorem, the total number of ways to reach the point x from the point 0 by making μ (± 1)-moves, without returning to 0, is

$$\frac{x}{\mu}\binom{\mu}{(\mu+x)/2}, \quad \mu \ge x,$$

$(\mu + x)/2$ being the total number of right moves. Therefore, for the (p, q)-simple walk,

$$\mathcal{Q}_\mu(x) := \frac{x}{\mu}\binom{\mu}{(\mu+x)/2} p^{(\mu+x)/2-1} q^{(\mu-x)/2};$$

(the probability of the first move, from 0 to 1, is 1, each of the other $\mu - 1$ moves has probability p.) Using Stirling's formula and $4pq \le 1$, we obtain a simple estimate

$$\mathcal{Q}_\mu(x) \le c_0 x \frac{\exp(-x^2/2\mu)}{\sqrt{\mu(\mu^2 - x^2 + \mu)}}, \quad x > 0 \tag{6}$$

where c_0 is some constant. (We will continue to use c's for various absolute constants.) Combining (5) and (6), we have

$$\Pr(\xi_\ell = x) \le c_0 x e^{-x^2/2\ell} \sum_{x\le\mu\le\ell} \frac{1}{\sqrt{\mu(\mu^2 - x^2 + \mu)}} \le c_1 x e^{-x^2/2\ell}.$$

(That the last sum is uniformly bounded follows from considering separately $\mu \ge 2x$ and $x \le \mu \le 2x$.) Then

$$\Pr(\xi_\ell \ge \alpha n^{1/3}\ln n) \le c_1 \sum_{x\ge\alpha n^{1/3}\ln n} x e^{-x^2/2\ell} \le c_2 \exp\left(-\frac{\alpha^2 n^{2/3}\ln^2 n}{4(1+a)n^{2/3}\ln n}\right) = c_2 n^{-a}, \tag{7}$$

as $\ell \le 2(1+a)n^{2/3}\ln n$, and $\alpha = 2\sqrt{a(1+a)}$. ∎

Proof of Theorem 2, part 2. As in the proof of part 1,

$$Z_n = \frac{n}{2} - 0.5 s_n,$$

so we need to show that w.h.p. $s_n \ge \nu_n := 2n^{1/3}\ln^{-2} n$. Clearly s_n stochastically dominates ξ_ℓ for the (p,q)-walk, where

$$p = \frac{m_T}{n} = \frac{1}{2} - \frac{\ell}{n}, \qquad \ell = n - 2m_T \in [(1+a)n^{2/3}\ln n, 2(1+a)n^{2/3}\ln n].$$

Thus

$$\Pr(s_n \le \nu_n) \le \Pr(\xi_\ell \le \nu_n) = \sum_{x\le \nu_n} \Pr(\xi_\ell = x),$$

with $\Pr(\xi_\ell = x)$ given by (4). This time we need a sharp bound for $\mathcal{P}_t$, which is

$$\mathcal{P}_t \le c((1-2p) + (t+1)^{-1/2}) = c\left(\frac{\ell}{n} + t^{-1/2}\right), \tag{8}$$

see Proposition 4. For $t \in [\ell/2, \ell]$, the first summand dominates since $\ell^{3/2} \gg n$, and the bound simplifies to $\mathcal{P}_t \le 2c(\ell/n)$. Break the sum in (4) into two parts, $\mu \ge \ell/2$ and $\mu < \ell/2$. Since $x \le \nu_n \ll \ell$, it follows from (6) and (8) that, for $x > 0$,

$$\sum_{\substack{2t+\mu=\ell\\ \mu\ge \ell/2}} \mathcal{P}_t \mathcal{Q}_\mu(x) \le c' x \left[\sum_{\mu \ge \ell/2} \mu^{-3/2}\big(\ell/n + (\ell-\mu+1)^{-1/2}\big)\right] \le c'' x\big((\ell/n)\ell^{-1/2} + \ell^{-1}\big) = O\big(x\ell^{1/2}/n\big),$$

as $\ell^{3/2} \gg n$. Therefore

$$\sum_{0<x\le\nu_n} \sum_{\substack{2t+\mu=\ell\\ \mu\ge \ell/2}} \mathcal{P}_t \mathcal{Q}_\mu(x) = O(\nu_n^2 \ell^{1/2} n^{-1}) = O(\ln^{-3/2} n). \tag{9}$$

Let $\mu \le \ell/2$ now. Since $2t + \mu = \ell$, it follows that $t \ge \ell/4$, and so $\mathcal{P}_t = O(\ell/n)$. Then, using (6), we obtain

$$\sum_{\substack{t+\mu=\ell\\ \mu\le \ell/2}} \mathcal{P}_t \mathcal{Q}_\mu(x) \le \hat{c}\ell n^{-1} x \left(\int_x^\infty \frac{e^{-x^2/2y}}{\sqrt{y(y^2-x^2)}}\,dy\right). \tag{10}$$

Substituting $y = x/z$, we transform the last integral into

$$x^{-1/2}\int_0^1 \frac{e^{-xz/2}}{\sqrt{z(1-z^2)}}\,dz = \sqrt{\frac{2}{x}}\,(J_1 + J_2),$$

with J_1, J_2 corresponding to integration over $[0,1/2]$ and $[1/2,1]$, respectively. Then, substituting $w = xz/2$,

$$J_1 \le \frac{2}{\sqrt{3}}\int_0^{1/2} z^{-1/2} e^{-xz/2}\,dz \le \frac{2}{\sqrt{3}} x^{-1/2}\int_0^\infty w^{-1/2}e^{-w}\,dw = \hat{c}_1 x^{-1/2},$$

and

$$J_2 \le e^{-x/4}\int_{1/2}^1 \frac{dz}{\sqrt{z(1-z^2)}} = \hat{c}_2 e^{-x/4}.$$

Therefore the bound (10) becomes

$$\sum_{\substack{2t+\mu=\ell\\ \mu\le \ell/2}} \mathcal{P}_t \mathcal{Q}_\mu(x) = O(\ell n^{-1} x (x^{-1/2})^2) = O(\ell/n), \quad x > 0.$$

Consequently

$$\sum_{0<x\le\nu_n}\sum_{\substack{t+\mu=\ell\\ \mu\le\ell/2}}\mathcal{P}_t\mathcal{Q}_\mu(x)=O(\ell n^{-1}\nu_n)=O(\ln^{-1}n). \qquad (11)$$

Combining (9) and (11), we obtain

$$\sum_{0<x\le\nu}\sum_{2t+\mu=\ell}\mathcal{P}_t\mathcal{Q}_\mu(x)=O(\ln^{-1}n).$$

Finally

$$\sum 2t+\mu=\ell\mathcal{P}_t\mathcal{Q}_\mu(0)=\mathcal{P}_{\ell/2}=O(\ell/n)=O(n^{-1/3}\ln n).$$

So

$$\begin{aligned}\Pr(\xi_\ell\le\nu_n)&=\sum_{0\le x\le\nu_n}\sum_{2t+\mu=\ell}\mathcal{P}_t\mathcal{Q}_\mu(x)\\ &=O(n^{-1/3}\ln n)+O(\ln^{-1}n)=O(\ln^{-1}n).\end{aligned}$$

Since $Z_n=n/2-0.5s_n$, and s_n dominates ξ_ℓ, the statement follows. ∎

Proof of Theorem 2, part 3. First of all, for n even, $Z_n=n/2$ iff $s_n=0$, and, for n odd, $Z_n=(n-1)/2$ iff $s_{n-1}=0$. Consider, for instance, even n. We know that, conditioned on the event in Lemma (call it $\mathcal{A}$), s_n is dominated by $\xi_\ell(1/2)$ of the walk $(\{\xi_j\}_{j\le\ell})$ with $p=1/2$, and dominates ξ_ℓ of the walk with $p=p_n:=1/2-\ell/n$. Then, using (3),

$$\begin{aligned}\Pr(s_n=0\,|\,\mathcal{A})\le&\ \Pr(\xi_r(p_n)=0)|_{r=\ell}\\ &=O((1-2p_n)+\ell^{-1/2})=O(\ell/n)=O(n^{-1/3}\ln n).\end{aligned} \qquad (12)$$

On the other hand, again using (3),

$$\begin{aligned}\Pr(s_n=0\,|\,\mathcal{A})\ge&\ \Pr(\xi_r(1/2))|_{r=\ell}\\ &=\Omega(\ell^{-1/2})=\Omega(n^{-1/3}\ln^{-1/2}n).\end{aligned} \qquad (13)$$

Since $\Pr(\mathcal{A}^c)=O(n^{-a})$, picking $a>1/3$ we conclude that unconditionally

$$\alpha n^{-1/3}\ln^{-1/2}n\le\Pr(s_n=0)\le\beta n^{-1/3}\ln n,$$

for some absolute constants $\alpha,\beta>0$. The case n odd is similar. This completes the proof of the theorem.
∎

References

[1] Manuel Abellanas, A. Garcí a, Ferrán Hurtado, and J. Tejel. Caminos alternantes. In *X Encuentros de Geometría Computacional, Sevilla, España*, 2003.

[2] Noga Alon, S. Rajagopalan, and S. Suri. Long non–crossing configurations in the plane. *Fund. Inform.*, 22, no. 4:385–394, 1995.

[3] Gabriela Araujo, József Balogh, Ruy Fabila, Gelasio Salazar, and Jorge Urrutia, 2005. (submitted).

[4] Bela Bollobás. *Random Graphs*. Cambridge University Press, Cambridge, 2nd. edition, 2001.

[5] Joseph A. Gallian. A dynamic survey of graph labeling. *Electronic Journal of Combinatorics*, 5, 1998.

[6] Ronald Graham and Neil Sloane. On additive bases and harmonious graphs. *SIAM J. Alg. Discrete Meth.*, 1, 1980.

[7] A. Kaneko and M. Kano. *Discrete geometry on red and blue points in the plane — a survey*, pages 551–570. Algorithms Combin. 25. Springer–Verlag, Berlin, 2003.

[8] A. Kaneko, M. Kano, and K Yoshimoto. Alternating hamiltonian cycles with minimum number of crossings in the plane. *International Journal of Computational Geometry and Applications*, 10:73–78, 2000.

[9] J. Kỳncl, J. Pach, and G. Tóth. Long alternating paths in bicolored point sets. In J. Pach, editor, *Graph Drawing 2004*, Lecture Notes in Computer Science 3383, pages 340–348. Springer–Verlag, 2004.