COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **27** (6) (2007) 721–736
DOI: 10.1007/s00493-007-2190-z

SPECTRAL NORM OF RANDOM MATRICES

VAN H. VU*

Received February 23, 2005

In this paper, we present a new upper bound for the spectral norm of symmetric random matrices with independent (but not necessarily identical) entries. Our results improve an earlier result of Füredi and Komlós.

1. Introduction

Let a_{ij}, $1 \le i \le j \le n$, be independent (but not necessarily identical) random variables with the following properties:

- $|a_{ij}| \le K$ for all $1 \le i \le j \le n$;
- $\mathbf{E}(a_{ij}) = 0$, for all $1 \le i < j \le n$;
- $\mathbf{Var}(a_{ij}) = \sigma^2$, for all $1 \le i < j \le n$.

For most part of the paper, we assume that σ and K are fixed (in the last section, we will discuss the case when these parameters are functions of n). We assume that n is large, whenever needed. We say that an event holds almost surely if it holds with probability tending to 1 as n tends to infinity.

Define $a_{ji} = a_{ij}$ and consider the symmetric random matrix $A = (a_{ij})_{i,j=1}^n$. A quantity which plays a significant role in various areas of mathematics, including combinatorics, mathematical physics, and theoretical computer science (see e.g., [1,5]), is the spectral norm of A, defined as follows

$$\lambda(A) = \sup_{\mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\|=1} \left| \mathbf{v}^T A \mathbf{v} \right| .$$

Mathematics Subject Classification (2000): 05D40, 15A52

* Research supported by an NSF CAREER award and by an Alfred P. Sloan fellowship.

The most well-known estimate on $\lambda(A)$ is perhaps the following, stated by Füredi and Komlós in 1981 [3].

Theorem 1.1. *For a random matrix A as above there is a positive constant $c=c(\sigma,K)$ (depending on σ and K but not on n) such that*

$$2\sigma\sqrt{n}-cn^{1/3}\ln n \le \lambda(A) \le 2\sigma\sqrt{n}+cn^{1/3}\ln n,$$

holds almost surely.

It is of considerable interest, from both theoretical and practical point of view, to sharpen this estmate. In [4], Krivelevich and the author showed that $\lambda(A)$ is concentrated very strongly around its mean.

Theorem 1.2. *For a random matrix A as above there is a positive constant $c=c(K)$ such that for any $t>0$*

$$\mathbf{P}(|\lambda(A)-\mathbf{E}(\lambda(A))| \ge ct) \le 4e^{-t^2/32}.$$

Notice that in this theorem c does not depend on σ, in fact for this theorem we do not have to assume anything about the variances.

Theorem 1.2 implies, among others, that the variance of $\lambda(A)$ is $O(1)$ (from Theorem 1.1 one would guess $O(n^{1/3}\ln n)$). It also follows from this concentration result that for some constant c [2]

$$2\sigma\sqrt{n}-c\ln n \le \lambda(A).$$

This improves the lower bound in Theorem 1.1. In fact, the lower bound in Theorem 1.1 was not verified rigorously in [3] and we will comment about this point in the last paragraph of Section 2.

The main goal of this paper is to improve the upper bound, which is usually more useful in applications. We are able to prove:

Theorem 1.3. *For a random matrix A as above there is a positive constant $c=c(\sigma,K)$ such that*

$$\lambda(A) \le 2\sigma\sqrt{n}+cn^{1/4}\ln n,$$

holds almost surely.

In many situations σ and K may depend on n. In this case, we have the following result:

Theorem 1.4. *There are constants C and C' such that the following holds. Let a_{ij}, $1\le i\le j\le n$ be independent random variables, each of which has* *mean 0 and variance at most σ^2 and is bounded in absolute value by K, where $\sigma \geq C' n^{-1/2} K \ln^2 n$. Then almost surely*

$$\lambda(A) \leq 2\sigma\sqrt{n} + C(K\sigma)^{1/2} n^{1/4} \ln n.$$

Remark 1.5. For the case when the entries of A are i.i.d. symmetric random variables, there are sharper bounds. The best current bound we know of is due to Soshnikov [7], which showes that the error term in Theorem 1.3 can be reduced to $n^{-1/6+o(1)}$. For the exact statement, we refer to [7].

In the next section, we present the general trace method for proving the upper bound. The core of this method is an estimate on the number of walks of a certain kind on a complete graph with n vertices. One can obtain such an estimate via a coding argument. The coding scheme, together with the proof, are presented in Section 3 and Section 4. The proof of Theorem 1.4 is similar to that of Theorem 1.3 and is left as an exercise.

In the proofs, we will make an extra (and convenient) assumption that the diagonal entries are zero. This is possible since switching all diagonal entries to zero changes the spectral norm by at most K, which is negligible compared to the upper bounds. (We would like to thank Krivelevich for pointing out this.)

2. Wigner's trace method

The common way to derive an upper bound for $\lambda(A)$ is to use Wigner's trace method, initiated in [6]. A standard fact in linear algebra tells us that for any positive integer k

$$\sum_{i=1}^{n} \lambda_i(A)^k = \mathbf{Trace} A^k,$$

which, via linearity of expectation, implies

$$\sum_{i=1}^{n} \mathbf{E}\big(\lambda_i(A)^k\big) = \mathbf{E}\big(\mathbf{Trace} A^k\big).$$

If k is even, then $(\lambda_i A)^k$ are non-negative, so $\sum_{i=1}^{n} \mathbf{E}(\lambda_i(A)^k) \geq \mathbf{E}(\lambda(A)^k)$, and we have

$$\mathbf{E}\big(\lambda(A)^k\big) \leq \mathbf{E}\big(\mathbf{Trace} A^k\big).$$

If we can find positive constants δ, c_1 and c_2 so that for $k = c_1 n^{\delta}$

$$\mathbf{E}\big(\mathbf{Trace} A^k\big) \leq c_2 n (2\sigma\sqrt{n})^k, \tag{1}$$

then we have

$$\mathbf{E}\big(\lambda(A)^k\big) \le c_2 n (2\sigma\sqrt{n})^k, \tag{2}$$

which, via Markov's inequality, implies that

$$\begin{aligned}\mathbf{P}\big(\lambda(A) \ge 2\sigma\sqrt{n} + cn^{1/2-\delta}\ln n\big) &\le c_2 n\Big(\frac{2\sigma\sqrt{n}}{2\sigma\sqrt{n} + cn^{1/2-\delta}\ln n}\Big)^k\\ &= c_2 n\Big(1 - \frac{cn^{1/2-\delta}\ln n}{2\sigma\sqrt{n} + cn^{1/2-\delta}\ln n}\Big)^k\\ &\le c_2 n \exp\Big(-\frac{c}{3\sigma} n^{-\delta} k \ln n\Big)\\ &= c_2 n \exp\Big(-\frac{c}{3\sigma} c_1 \ln n\Big) = o(1),\end{aligned}$$

for all sufficiently large c.

In [3], one set $\delta = 1/6$ and thus $1/2 - \delta = 1/3$, resulting in the upper bound in Theorem 1.1. In order to verify (1), observe that

$$\mathbf{Trace} A^k = \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}, \tag{3}$$

which implies

$$\mathbf{E}\big(\mathbf{Trace} A^k\big) = \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n \mathbf{E}(a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}). \tag{4}$$

For $1 \le p \le k$, denote by $E(n,k,p)$ the sum of $\mathbf{E}(a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1})$ over all sequences $i_1, \dots, i_k$ where $|\{i_1, \dots, i_k\}| = p$ (not counting multiplicities). Since the expectation of a_{ij} is zero, if some a_{ij} in the product $a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}$ has multiplicity one, then the expectation of the whole product is zero. By the pigeon hole principle, if $p > \frac{k}{2} + 1$, then there must be an entry with multiplicity one. Therefore, $E(n,k,p) = 0$ for $p > \frac{k}{2} + 1$ and the left hand side of (4) is $\sum_{p=1}^{k/2+1} E(n,k,p)$.

Notice that a product $a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}$ defines a closed walk

$$(i_1 i_2)(i_2 i_3) \dots (i_{k-1} i_k)(i_k i_1)$$

of length k on the complete graph K_n on $\{1, \dots, n\}$ (we allow loops in K_n). If a product has a non-zero contribution in the expectation of the trace, then any edge in the walk should appear at least twice. Thus, the estimation of $E(n,k,p)$ relies on a bound on the number of walks with this property.

Let $W(n,k,p)$ be the number of walks in K_n using k edges and p vertices where each edge in the walk is used at least twice. The heart of the matter is the following estimate, claimed in [3]

$$W(n,k,p) \le n(n-1)\cdots(n-p+1)\binom{k}{2p-2}p^{2(k-2p+2)}2^{2p-2}. \tag{5}$$

Remark. In fact, a slightly stronger bound was claimed in [3]. Technically speaking, one can replace 2^{2p-2} by $\frac{1}{p}\binom{2p-2}{p-1}$, which is better by a factor $\Theta(p^{3/2})$. This factor, however, does not play any significant role and it is more convenient to work with the simpler formula in (5). In fact, as we take the kth root in (2), any factor of order $\exp(o(k))$ is negligible.

To see that (5) implies the theorem, notice that when the walk has exactly k edges and p vertices then

$$\mathbf{E}(a_{i_1i_2}a_{i_2i_3}\cdots a_{i_{k-1}i_k}a_{i_ki_1}) \le \sigma^{2p-2}K^{k-(2p-2)},$$

since $\mathbf{E}(|a_{ij}|^l) \le K^{l-2}\sigma^2$ for any $l \ge 2$. This implies that

$$\begin{aligned} E(n,k,p) &\le \sigma^{2p-2}K^{k-(2p-2)}W(n,k,p) \\ &\le \sigma^{2p-2}K^{k-(2p-2)}n(n-1)\cdots(n-p+1)\binom{k}{2p-2}p^{2(k-2p+2)}2^{2p-2} \\ &= S(n,k,p). \end{aligned}$$

It is easy to show that

$$S(n,k,p-1) \le \frac{K^2k^6}{4\sigma^2 n}S(n,k,p). \tag{6}$$

Thus, for $k = (\frac{\sigma}{K})^{1/3}n^{1/6}$, $S(n,k,p-1) \le \frac{1}{2}S(n,k,p)$. So

$$\begin{aligned} \sum_{p=1}^{k/2+1} E(n,k,p) \le \sum_{p=1}^{k/2+1} S(n,k,p) &\le 2S(n,k,k/2+1) \\ &= 2\sigma^k n(n-1)\cdots(n-k/2)2^k = 2n(2\sigma\sqrt{n})^k. \end{aligned}$$

For a walk W with p different vertices, let $V(W) = v_1,\dots,v_p$ be the (ordered) sequence formed by its vertices in the order they appear in W. There are $n(n-1)\cdots(n-p+1)$ ways to fix an ordered sequence of p different vertices of K_n. Denote by $W'(k,p,n)$ the number of walks corresponding a fixed sequence (by symmetry, this number does not depend on the sequence). Inequality (5) is a consequence of the following:

Lemma 2.1. *We have*

$$W'(k,p,n) \leq \binom{k}{2p-2} p^{2(k-2p+2)} 2^{2p-2}.$$

The main task now is to prove Lemma 2.1. To verify this lemma, the general strategy is to code a walk by a sequence of symbols in an injective way and estimate the number of codewords from above. The coding scheme we gave here was is a refinement of the one from [3]. It also fixes a small problem in this previous scheme (a detailed discussion is given in the extended abstract of this paper, appeared in STOC 2005).

To conclude this section, let us mention that the approach discussed so far concerns only the upper bound for $\lambda(A)$. As pointed out in [3], a lower bound $2\sigma\sqrt{n}+o(\sqrt{n})$ follows immediately from Wigner's semicircle law. However, the finer bound $2\sigma\sqrt{n}-cn^{1/3}\ln n$ claimed in Theorem 1.1 does not. In [3], it was mentioned that this lower bound can be proved in a similar fashion via Markov's inequality. We do not see any way to materialize this idea. The main trouble is that in the previous arguments we bound the norm of $\lambda(A)$, but not its variance. Thus, there is no symmetry between the upper tail and the lower tail like when one applies Chebyshev's inequality.

The lower bound in Theorem 1.3 was obtained via Talagrand's inequality, which is more recent than Theorem 1.1 and has little to do with the trace method.

3. A new coding scheme

The scheme which leads to the sharper bound in Theorem 1.3 is a little bit too technical, so we are going to present it in two steps. In this section, we present a simpler scheme which reproves Lemma 2.1 (and with it the bound in Theorem 1.1). In the next section, we refine this scheme to obtain the bound in Theorem 1.3.

The case $p=1$ is trivial, so we assume that p is at least 2. Given a walk W, we define a tree $T=T(W)$ with the following rules.

- The first vertex of T is the starting of the walk.
- Consider the walk edge by edge. If in an edge uv, the right end point v has not appeared in the walk, then add v to the tree and draw an edge between u and v. (Notice that u us already in the tree.)

The first three steps of our scheme are motivated by the scheme in [3] and are as follows.

- Mark an edge uv in W with a plus $(+)$ sign if it appears in T for the first time (in other words, v is a new vertex of the walk).
- Mark an edge uv in W with a minus $(-)$ sign if it appears in T for the second time.
- If uv is neutral then we mark uv by v.

Let us recall that an edge is neutral if it is neither plus nor minus. If W has k edges and p vertices, then the number of plus edges and the number of minus edges are both $p-1$. Let N denote the number of neutral edges. We have

$$N = k - (2p-2).$$

With this notation, the bound in Lemma 2.1 becomes

$$W'(k,p,n) \leq \binom{k}{2p-2} p^{2N} 2^{2p-2}. \tag{7}$$

The above three steps are, of course, not enough to make the coding scheme injective. In order to make the scheme complete, we are going to assign extra symbols to certain minus edges. The rule is somewhat complicated and before describing it let us give the reader a motivation.

Call a codeword obtained from a walk W by the above three steps a *preliminary* codeword. In the following, we try to decode a preliminary codeword and see what kind of information is missing. Since we have fixed the order in which the vertices appear, we know the starting point of the walk. Assume now that we have determined the first j vertices $v_1, \dots, v_j$ of the walk for some $j \geq 1$. Consider the jth symbol from the (preliminary) codeword; this symbol represents the jth edge $v_j v_{j+1}$ in the walk. If this symbol is a plus, then v_{j+1} is the next new vertex and we have no problem determining it. If the symbol in question is v, then $v_{j+1} = v$ and $v_j v_{j+1}$ is a neutral edge. The troublesome case is when the symbol in question is a minus.

Here is a situation when one can decode a minus sign with no further information. We call an interval of (consecutive) symbols in a code word *redundant* if it has the form $++\cdots+--\cdots-$, where the number of pluses and the number of minuses are equal. It is easy to see that if we know the left end vertex of the first plus edge, then we can decode the whole interval. The minus edges are exactly the plus edges in the reversed order. Thus we can remove redundant intervals from our codeword and consider the condensed codeword instead. A codeword is condensed if it does not contain a redundant interval. Together with the condensed codeword, we can define the condensed walk similarly.

We would like to mention that the redundant intervals are removed consequentially, and the removing of one redundant interval may create a new one. It is not hard to show, however, that the order in which we remove them does not matter, and so the condensed codeword is well defined.

Example. Let us consider the codeword $+++--+--uv$. If we first remove the first redundant interval $++--$ we are left with $++--uv$ (so a new redundant interval is created). After removing $++--$ again, we obtain uv. If we remove the last $+-$ first, we are left with $+++---uv$. Removing $+++---$, we obtain the same condensed codeword uv.

It is trivial that if we can decode the condensed codeword, then we can decode the original one, providing the positions of the redundant intervals. We call the edges which do not appear in the condensed codeword redundant.

The above situation is a special case of the following more general one. As before, assume that we have determined $v_1, \ldots, v_j$ and now face a minus edge $v_j v_{j+1}$. This means that $v_j v_{j+1}$ is an edge of the (partial) tree $T_j(W)$ created by the walk so far. Moreover, this edge has been used exactly once, namely, in its previous occurrence $v_j v_{j+1}$ was marked by a plus. Thus, if in the tree $T_j(W)$, v_j is adjacent to exactly one plus edge, then one can determine v_{j+1}. (Notice that $T_j(W)$ and the symbols on its edges are given by the sequence $v_1, \ldots, v_j$.)

The only case when we cannot decode is when v_j is adjacent to at least two plus edges in $T_j(W)$. We call such a v_j a *critical* vertex. Given a walk W, the set of its critical vertices is well defined.

Here is a naive solution to the situation. Once we arrive at a critical vertex v_j and see a minus, simply give the name of the next vertex (which is the right end of the minus edge starting from v_j). This solution is, however, too wasteful and we can do better as follows. Consider the maximal sequence of consecutive minuses from v_j in the condensed codeword. (By the definition of critical vertices, the minus edge under consideration is not redundant and appears in the condensed codeword.) It is sufficient to give the name of the right end of the last edge of this sequence. Indeed, the edges between v_j and this vertex (say v) are all minus, so they form a path between v_j and v in the tree $T_j(W)$. But in a tree, there is a unique path connecting any two vertices. The redundant edges can be inserted later at no cost.

The last edge in the sequence of minuses discussed above plays an essential role in our study and we call it an *important* edge.

Definition 3.1. An edge is important if it is the last edge in the maximal sequence of minus edges in the condensed walk starting from a critical point.

Examples. Assume that a is a critical point and the next six steps (starting from a) in the walk are $e_1, \ldots, e_6$.

- If these edges are coded as $-+a_1---$ then e_1 is important.
- If these edges are coded as $-++--a_1$ then again e_1 is important, since in the condensed codeword we have $-a_1$.
- If these edges are coded as $-+---a_1$ then e_5 is important, since in the condensed codeword we have $---a_1$ and the last minus corresponds to e_5.
- If these edges are coded as $---+a_1a_2$ then e_3 is important.

Now we are ready to describe the last, and perhaps most critical, rule of our coding scheme:

- If a minus edge uv is important, then remark it by the double symbol $(-,v)$.

The reader should keep in mind that we grow the spanning tree gradually with the decoding process. As pointed out, there is no problem with decoding a plus or a neutral edge. Once we arrive to a minus symbol which cannot be decoded trivially, then the corresponding edge must be a minus edge adjacent to a critical point. Consider the sequence of minuses starting from this symbol (to the right) in the condensed codeword. The last minus corresponds to an important edge and has an extra symbol v, which tells us which branch of the tree we should follow in the walk.

Now let us bound the number of codewords. It is more convenient to separately bound C_i, the number of codewords with exactly i important edges, and then sum up over i. To build a codeword, we first build a preliminary one by filling in the pluses, minuses and the symbols for neutral edges. Next, we fill in the extra symbols for the important edges. So we are going to first estimate the number of preliminary codewords and next estimate the number of ways to extend a preliminary codeword to an authentic one.

To estimate the number of preliminary codewords, notice that

- The number of pluses and minuses is $2p-2$ and there are $\binom{k}{2p-2}$ ways to choose the places for them.
- Once $2p-2$ slots have been chosen, there are at most 2^{2p-2} ways to fill them (in each slot one may put a plus or a minus).
- For each neutral edge, we mark it using one of the vertices. There are p vertices, so there are at most p^N ways to mark the neutral edges.

Remark. As pointed out in [3], one can replace 2^{2p-2} by $\frac{1}{p}\binom{2p-2}{p-1}$ by noticing that in any initial segment, the number of pluses is at least the number of minuses.

It follows that the number of preliminary codewords is at most

$$\binom{k}{2p-2}2^{2p-2}p^N. \qquad (8)$$

Now we estimate the number of ways to extend a preliminary codeword. Here is an essential observation: the locations of the first critical vertex and the first important edge are determined by the preliminary codeword. Given its location, we have at most p ways to put an extra symbol on the first important edge. Once this symbol is given, the locations of the next critical point and the next important edge are determined and so on. Thus, we have no freedom in choosing the locations of the important edges. Therefore, given the preliminary codeword, the number of ways one can extend it is at most p^i (as we assume before hand that there are i important edges).

Given (8), we conclude that the number of codewords is upper bounded by

$$S = \sum_{i=1}^{I} C_i = \binom{k}{2p-2} 2^{2p-2} p^N \sum_{i=1}^{I} p^i,$$

where I is the maximum number of important edges in a walk.

The rest of the proof relies on the following observation.

Fact 3.2. *The number of important edges in a walk is at most* $\max\{0, N–1\}$, *where N is the number of neutral edges.*

A special case is when there are no neutral edges, i.e., $p = k/2+1$. In this case every edge is redundant and there is no important edge and (7) is trivial. In what follows, we assume $N \geq 1$. By Fact 3.2 and the assumption that $p \geq 2$, the number of possible codewords is at most

$$\binom{k}{2p-2} 2^{2p-2} p^N \sum_{i=0}^{I} p^i \leq \binom{k}{2p-2} 2^{2p-2} p^N \sum_{i=0}^{N-1} p^i \leq \binom{k}{2p-2} 2^{2p-2} p^{2N},$$

proving (7).

To conclude the proof, it remains to verify Fact 3.2. We can assume that there is at least one important edge. Fact 3.2 is an immediate consequence of the following three simple observations:

(i) There must be a neural edge prior to the first important edge.

(ii) There must be a neutral edge between two consecutive important edges.

(iii) There must be a neutral edge after the last important edge.

To see (i), notice that if an initial interval of a codeword consists of only pluses and minuses, then all minuses are redundant. To see (ii), notice that if between two important edges e_1 and e_2 there is no neutral edges then after the removing of redundant intervals, there are only minus edges left. This means that in the condensed codeword there are only minuses between e_1 and e_2. This contradicts the definition of e_1 as the last edge of a minus sequence. The reader is left to verify (iii).

4. A finer code and a sharper bound

In this section, we are going to present a finer coding scheme which leads to the proof of the sharper upper bound claimed in Theorem 1.3. Our goal is to show

Lemma 4.1. *We have*

$$W(n,k,p) \leq \binom{k}{2p-2} 2^{k+N+1} p^N (N+2)^N.$$

Plugging this bound into the estimates in Section 2 would result in the upper bound in Theorem 1.3. Indeed, redefining (in the obvious way) $S(n,k,p)$ using Lemma 4.1, one can show, instead of (6), that

$$S(n,k,p-1) \leq \frac{K^2k^4}{C\sigma^2 n} S(n,k,p)$$

for some constant C independent of σ and K. Thus, by setting $k = a(\frac{\sigma}{K})^{1/2} n^{1/4}$, for some properly chosen constant a, one can guarantee that $S(n,k,p-1) \leq \frac{1}{2} S(n,k,p)$ and continue as in Section 2. The critical gain here is that we can replace $n^{1/6}$ in the definition of k by $n^{1/4}$ and this results in the claimed bound.

In order to make the presentation less technical, instead of 4.1, we will first prove another bound

$$W(n,k,p) \leq \binom{k}{2p-2} 2^{k+N+1} p^{3N/2}, \tag{9}$$

and next modify the argument to verify Lemma 4.1.

Proof of inequality (9). Our first observation is that in a certain situation, we can be more economical when marking the important edges. Let us go back to the process of decoding the preliminary code in the last section. Assume that we arrive at a critical vertex, say a. The important edge following a is $e=(u,v)$ and we marked it with the double symbol $(-,v)$. Now assume that a is adjacent with exactly two plus edges (so there are two possibilities for a minus edge from a). The vertex v can be reached from a by a path consisting of plus edges. Assume furthermore that all inner vertices of this path are adjacent to exactly two plus edges. In this case, we say that e is *simple* and propose a new way to mark e. Instead of using $(-,v)$, we use the double symbol $(-,dir)$ where dir (shorthand for direction) is either left or right and represents the direction we should take from a in order to get to e. Once the correct direction is given, there is no chance for mistake as there is a unique path of pluses in this direction.

The double symbol $(-, dir)$ is a lot cheaper than the double symbol $(-, v)$ since there are only two choices for dir but p for v. If we could prove that all important edges are simple, then the upper bound would become

$$\binom{k}{2p-2} 2^{2p-2} p^N 2^N,$$

and we would gain a significant factor $(p/2)^N$.

It is, of course, not true that all important edges are simple. However, it turns out that the notion of simple important edges is useful.

Another heuristic to improve upon the bound is to show that I, the maximum number of important edges, is significantly less than N. In the previous proof, we showed that between two consecutive important edges, there must be a neutral edge. The reason, roughly, is that we must use a neutral edge to terminate the first important edge before getting to the second one. The intuition behind the improvement is that an important edge not only needs to be terminated, but also needs to be created, and so an important edge would cost two neutral edges. Therefore, the ratio between I and N would be at most $1/2$, i.e., $I \leq N/2$.

Unfortunately, this argument is not rigorous, either. The main trouble is that there could be overlaps as the terminator of one important edge might serve as the creator of another. The extra and critical observation which links the above two proposals and makes the proof works is that there is a trade-off between the number of overlaps and that of simple edges. Basically, whenever an overlap occurs, we obtain a new simple edge. This allows us to bound the number of non-simple important edges by $N/2$.

Lemma 4.2. *The number of non-simple important edges is at most $N/2$.*

Based on this lemma, our new coding scheme is as follows. The first three steps are as before, but we modify the last step to the following:

- If $e = (u, v)$ is a simple important edge, then mark it with $(-, dir)$. If $e = (u, v)$ is a non-simple important edge, then mark it with $(-, v)$.

Now we are going to prove (9) via Lemma 4.2. For $0 \leq i \leq N/2$, let C_i denote the number of walks with exactly i simple important edges. We first estimate C_i. Once we get to an important edge, there are two possibilities: the edge is simple or not. This amounts to (at most) another factor 2^N as there are less than N important edges. If the edge is simple, then we have two choices for the extra symbol (either left or right). This contributes a factor at most 2^N. If the edge is not simple, then we have at most p choices and this contribute a factor at most p^i. So

$$C_i \le \binom{k}{2p-2} 2^{2p-2} p^N 2^N 2^N p^i = \binom{k}{2p-2} 2^{k+N} p^{N+i}.$$

It follows, via Lemma 4.2, that for $p \ge 2$ (for $p=1$, the claim of the lemma is trivial)

$$\begin{aligned} W'(n,k,p) &= \sum_{i=0}^{\lfloor N/2 \rfloor} C_i \\ &\le \binom{k}{2p-2} 2^{k+N} \sum_{i=0}^{\lfloor N/2 \rfloor} p^{N+i} \\ &\le \binom{k}{2p-2} 2^{k+N+1} p^{\frac{3N}{2}}, \end{aligned}$$

proving (9).

It remains to prove Lemma 4.2. Consider a walk $W = v_0 v_1 \dots v_k$. As usual, we grow a spanning tree together with W; $T(v_j)$ is spanning tree when the walk reaches v_j. The edges which have been used exactly once up to time j form a subforest of $T(v_j)$. We call this forest the *fundamental* forest at v_j and denote it by $F(v_j)$. If v_j is a critical vertex, then it is adjacent to at least two edges which have been used exactly once and so it must be an inner vertex of a connected component of the forest. The important edge e following v_j is simple if in $F(v_j)$, v_j has degree two and every inner vertex on the path from v_j to the right end of e also has degree exactly two.

Assume that the fundamental forest $F(v_j)$ has l connected components $F_1, \dots, F_l$. We define the *energy* of the walk at v_j as follows

$$\mathrm{Ener}(v_j) = \sum_{i=1}^{l} (L(F_i) - 2),$$

where $L(F_i)$ is the number of leaves in F_i. If $F(v_j)$ is empty, we set $\mathrm{Ener}(v_j) = 0$.

Claim 4.3. *Let c_1 and c_2 be two consecutive critical vertices. Let e be the (unique) important edge between c_1 and c_2 and m be the number of neutral edges between c_1 and c_2. Then one of the following two cases must hold*

- $\mathrm{Ener}(c_2) - \mathrm{Ener}(c_1) \le m - 2$;
- $\mathrm{Ener}(c_2) - \mathrm{Ener}(c_1) = m - 1$ *and e is simple.*

Let us now deduce Lemma 4.2 from Claim 4.3. Consider the sequence $c_0, w_1, \dots, c_l$, where c_0 is the first vertex of the walk and $c_1, \dots, c_l$ are (all) the critical vertices. Let m_i be the number of neutral edges between c_i and c_{i+1}. It is easy to see that the energy at c_1 is zero. Moreover,

$$m_0 + m_1 + \cdots + m_{l-1} + m_l = N. \tag{10}$$

On the other hand, by the claim

$$\sum_{i=1}^{l-1} \mathrm{Ener}(c_{i+1}) - \mathrm{Ener}(c_i) \geq \sum_{i=1}^{l-1} m_i - 2(l-1) + l',$$

where l' is the number of simple important edges. The left hand side is $\mathrm{Ener}(c_l) - \mathrm{Ener}(c_1) = \mathrm{Ener}(c_l) \geq 0$. Thus the right hand side is also at least 0. Using (10) we can deduce that

$$\sum_{i=1}^{l-1} m_i - 2(l-1) + l' = (N - m_0 - m_l + 2) - (2l - l') \geq 0.$$

Since $2l - l'$ is at least twice the number of non-simple important edges, and both m_0 and m_l are at least one, we are done.

Proof of Claim 4.3. Consider the fundamental forest $F(c_1)$ at c_1. This forest contains a path from c_1 whose last edge is $e = (u, w)$. Trivially

$$\mathrm{Ener}(c_2) - \mathrm{Ener}(c_1) = D_1 + D_2,$$

where $D_1 = \mathrm{Ener}(w) - \mathrm{Ener}(c_1)$ and $D_2 = \mathrm{Ener}(c_2) - \mathrm{Ener}(w)$.

Notice that since e an important edge, all neutral edges between c_1 and c_2 appear after e. We denote the right end points of these edges by $y_1, \dots, y_m$. It is easy to see that

$$\begin{aligned} \mathrm{Ener}(y_1) - \mathrm{Ener}(w) &\leq 1 \\ \mathrm{Ener}(y_2) - \mathrm{Ener}(y_1) &\leq 1 \\ &\vdots \\ \mathrm{Ener}(y_m) - \mathrm{Ener}(y_{m-1}) &\leq 1 \\ \mathrm{Ener}(c_2) - \mathrm{Ener}(y_m) &\leq 0. \end{aligned}$$

Summing up these inequalities, we have that

$$D_2 \leq m.$$

To estimate D_1, we have to consider several cases, depending on the degrees of c_1 and w in the forest $F(c_1)$. Recall that since c_1 is critical, the degree of c_1 is at least 2.

Case 1. $\deg c_1 \geq 3, \deg w \geq 3$. Assume first that the inner vertices in the path from c_1 to w all have degree two. In this case, after erasing the path from c_1 to w, the number of leaves does not change, but the number of components increases by one. Thus the energy decreases by 2 and so $D_1 = -2$. Now if an inner vertex v has degree larger than two, then we have one more component and at most one more leave (this new leave appears if the degree of v is exactly 3). This only decreases the value of D_1. In both cases, $D_1 + D_2 \leq -2 + m = m - 2$.

Case 2. $\deg c_1 = 2, \deg w \geq 3$. Again assume first that the inner vertices in the path from c_1 to w all have degree two. In this case, after erasing the path from c_1 to w, the number of components increases by one and the number of leaves increase by one (the new leave is c_1). Thus, $D_1 = -1$. Moreover, $e = (u, v)$ is simple by definition. Now if any inner vertex has degree larger than two, then arguing as above, we see that D_1 should decrease further and so $D_1 \leq -2$. So $D_1 + D_2 \leq m - 2$.

Case 2'. $\deg c_1 \geq 3, \deg w = 2$. By symmetry, this case can be treated exactly as the previous one.

Case 3. $\deg c_1 = 2, \deg w = 2$. The new observation in this case is that the first neutral edge (with right end point y_1) does not increase the energy, i.e., $\mathrm{Ener}(y_1) = \mathrm{Ener}(w) = 0$. Thus, $D_2 \leq m - 1$. Now assume (as usual) that the inner vertices in the path from c_1 to w all have degree two. Thus e is simple and the erasing increases both the number of components and the number of leaves by two (the two new leaves are c_1 and w). So $D_1 = 0$ and $D_1 + D_2 \leq m - 1$. If there is an inner vertex with degree larger than two, then $D_1 \leq -1$ and $D_1 + D_2 \leq m - 2$.

Case 4. $\deg c_1 = 2, \deg w = 1$. By the same reason, we have $D_2 \leq m - 1$. If the inner vertices in the path from c_1 to w all have degree two, then the erasing increases the number of leaves by one (the new leave is c_1) but the number of components remain the same (now w is no longer in the forest). In this case e is simple and $D_1 = 0$ so $D_1 + D_2 \leq m - 1$. If there is an inner vertex with degree larger than two, then $D_1 \leq -1$ and $D_1 + D_2 \leq m - 2$. The proof of the claim is complete. ∎

We are now going to prove the stronger bound claimed in Lemma 4.1

$$W(n, k, p) \leq \binom{k}{2p - 2} 2^{k+N+1} p^N (N + 2)^{N/2}.$$

Let us write $p^{3/2N}$ as $p^N p^{N/2}$. The first term p^N comes from marking the neutral edges. We reconsider the second term $p^{N/2}$. This term arises as we have at most $N/2$ non-simple important edges and we mark them by their right end point. The obvious bound for the number of choices for this right end point is p. Now we are going to argue that the number of choices for this right end point is at most $N+2$. Thus, we can replace $p^{N/2}$ by $(N+2)^{N/2}$.

Notice that once we arrive to a critical vertex c, the position of the next important edge e is determined in the preliminary codeword. Thus, we know the distance l from c to the right end point w of e in the fundamental forest $F(c)$. (Both c and e belong to this forest.) Thus, the number of choices for w is at most the number of vertices of distance l from c in $F(c)$. On the other hand, the latter is bounded from above by the number of leaves in $F(c)$. To create the first two leaves one may not need any neutral edge, but for any of the remaining leaves one needs at least one neutral edge. Therefore, the number of leaves in $F(c)$ is at most the number of neutral edges used up to c plus two, which is at most $N+2$. This completes the proof. The forth step in our coding scheme remains the same as before, the only difference is that the number of choices for v has changed. ∎

Remark. One can perhaps reduce $(N+2)^{N/2}$ to $(N+2)(N-1)\cdots(N-N/2+3)$, but this refinement does not effect the bounds significantly.

References

[1] N. Alon: Spectral techniques in graph algorithms, in: *LATIN'98: theoretical informatics (Campinas, 1998)*, 206–215; Lecture Notes in Comput. Sci., **1380**, Springer, Berlin, 1998.

[2] N. Alon, M. Krivelevich and V. H. Vu: On the concentration of eigenvalues of random symmetric matrices, *Israel J. Math.* **131** (2002), 259–267.

[3] Z. Füredi and J. Komlós: The eigenvalues of random symmetric matrices, *Combinatorica* **1(3)** (1981), 233–241.

[4] M. Krivelevich and V. H. Vu: Approximating the independence number and the chromatic number in expected polynomial time, *J. Comb. Optim.* **6(2)** (2002), 143–155.

[5] M. L. Mehta: *Random matrices*, Second edition, Academic Press, Inc., Boston, MA, 1991.

[6] E. Wigner: On the distribution of the roots of certain symmetric matrices, *The Annals of Mathematics* **67** (1958), 325–327.

[7] A. Soshnikov: Universality at the edge of the spectrum in Wigner random matrices, *Comm. Math. Phys.* **207(3)** (1999), 697–733.

Van H. Vu

Department of Mathematics, Rutgers
Piscataway, NJ 08854-8019, USA
vanvu@math.rutgers.edu