Largest Planar Matching in Random Bipartite Graphs

Marcos Kiwi* Martin Loebl †

July 19, 2000

Abstract

Given a distribution $\mathcal{G}$ over labeled bipartite (multi) graphs, $G = (W, M, E)$ where $|W| = |M| = n$, let $L(n)$ denote the size of the largest planar matching of G (here W and M are posets drawn on the plane as two ordered rows of nodes, an upper and a lower one, and a (w, m) edge is drawn as a straight line between w and m). The main focus of this work is to understand the asymptotic (in n) behavior of $L(n)$ for different distributions $\mathcal{G}$. Two well studied particular instances of this problem are Ulam's longest increasing subsequence problem and the longest common subsequence problem.

This work's main focus is in the case where $\mathcal{G}$ is the uniform distribution over the k–regular bipartite graphs on W and M. For $k = O(n^{1/5-\varepsilon})$, we establish that $L(n)/\sqrt{kn}$ tends to 2 in probability when $n \to \infty$. When $k = O(1)$ the convergence in mean to the same limit holds. It is also shown that when each of the n^2 possible edges between W and M are chosen independently with probability $0 < p < 1$, then $L(n)/n$ tends to a constant γ_p in probability and in mean when $n \to \infty$.

The problems addressed in this work can be thought of as a novel generalization of Ulam's longest increasing subsequence problem and the longest common subsequence problem.

Keywords: longest increasing subsequences, longest common subsequences, random bipartite graphs.

1 Introduction

We consider marriage problems where we are given two finite sets W and M of equal size, the first one representing women and the second one men. The role

*Dept. Ing. Matemática & Centro Modelamiento Matemático, U. de Chile, Santiago 170–3, Chile, (mkiwi@dim.uchile.cl). Gratefully acknowledges the support of Conicyt via Fondecyt No. 1981182 and Fondap in Applied Mathematics, 2000, and GACR 201/99/0242.

†Charles University, Prague, Check Republic, (loebl@ms.mff.cuni.cz). Gratefully acknowledges the support of Conicyt via Fondap in Applied Mathematics, 2000, and GAUK 158/99.

of both sets is symmetric. All men have an order of preference for the women and all women an order of preference for the men. A *matching* is an injection of W onto M. A *perfect matching* is a bijection of W onto M, i.e., a set of $|W| = |M|$ monogamous marriages between the men and the women.

We consider the matching problems when not all marriages are allowed (one can imagine that legal and social aspects forbid certain marriages). We seek solutions with the extra condition that each person be married to someone appearing on his or her allowed marriage list. Since a marriage will take place if both parties are willing to go forth with it we henceforth assume that a woman belongs to a man's allowed list of marriages if and only if the man also belongs to the woman's allowed list of marriages.

We are interested in the *size* of the largest matching for different random models of allowed marriage lists. In particular, we focus on the case where: (1) the relative rating of men and women is the same for all men and women (one can imagine a rigidly stratified society where individual preferences only reflect social status), and (2) there are no two couples where the lower ranked man marries the higher ranked woman.

We now discuss a natural graph theoretic interpretation of the problem we would like to address. Let $\mathcal{G}$ be a distribution over the set of labeled bipartite (multi) graphs (W, M, E), where $E \subseteq W \times M$. Assume $W = \{w_1, \ldots, w_n\}$ and $M = \{m_1, \ldots, m_n\}$ are such that if $j > i$ then the j–th woman is preferred over the i–th one by all men and the j–th man is preferred over the i–th one by all women. A pair (w, m) belongs to E if w and m is an acceptable marriage. We can represent the situation by a bipartite graph with W as upper nodes and M as lower nodes where an edge (w, m) is represented as a straight line between w and m (think of w and m as nails and e as a rubber band that has been stretched around both nails). The situation is depicted in Fig. 1. This way of drawing bipartite graphs will be henceforth referred to as the canonical representation of a bipartite graph. Throughout, when we refer to a bipartite graph we actually mean its canonical representation. Note that in a (canonical representation of a) bipartite graph a collection of non–intersecting edges with no common endpoints represents a matching where there are no two couples such that the lower ranked man marries the higher ranked woman. Henceforth, we refer to such a collection of edges as a *planar matching*. Observe that this notion is well defined since we have fixed the representation of the bipartite graphs we consider. Thus, the main goal of this work can be concisely stated as:

> Given a distribution $\mathcal{G}$ over labeled bipartite (multi) graphs, understand what is the size of the largest planar matching of a graph chosen according to $\mathcal{G}$.

We consider this basic question for several different choices of distributions $\mathcal{G}$.

There are particular interesting cases of our general question. Indeed, consider the following two distributions over bipartite graphs on W and M:

$\mathcal{G}_{LIS}$: picks a labeled perfect matching on W and M with equal probability,

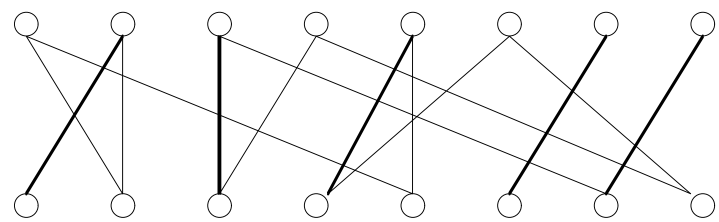

Figure 1: Women (respectively men) are represented as upper (respectively lower) nodes. Leftmost (respectively rightmost) nodes are the lowest (respectively highest) ranked nodes. Edges represent allowed marriages. Thick edges represent the largest planar matching.

$\mathcal{G}_{LCS}$: picks for every element of W and M, independently and uniformly, one of two colors, add edges between all nodes in W and M that received the same color.

There is a one–to–one correspondence between perfect matchings on W and M, $|W| = |M| = n$, and permutations of $\{1, \ldots, n\}$. Hence, picking a bipartite graph G according to $\mathcal{G}_{LIS}$ is equivalent to uniformly choosing a random permutation π_G of $\{1, \ldots, n\}$. Furthermore, note that each planar matching of G is in one–to–one correspondence with an increasing subsequences of π_G, i.e., with sequences of integers $i_1 < i_2 < \cdots < i_k$ such that $\pi_G(i_1) < \pi_G(i_2) < \cdots < \pi_G(i_k)$. Thus, determining the size of the largest planar matching of a graph chosen according to $\mathcal{G}_{LIS}$ is equivalent to determining the largest increasing subsequence of a uniformly chosen permutation of $\{1, \ldots, n\}$. This problem is known as Ulam's problem since it was apparently first raised by Ulam. It is motivated by the result of Erdös and Szekeres [8] that says that every permutation of $\{1, \ldots, n\}$ has either and increasing or decreasing subsequence of length at least $\sqrt{n}$.

Now, consider the distribution $\mathcal{G}_{LCS}$. Observe that there is a one–to–one correspondence between 2–colorings of an n element set and n bit long strings. Hence, picking a bipartite graph G according to $\mathcal{G}_{LCS}$ is equivalent to uniformly choosing two n bit long strings x_G and y_G. Furthermore, note that each planar matching of G is in one–to–one correspondence with a common subsequence of x_G and y_G, i.e., with sequences of integers $i_1 < i_2 < \cdots < i_k$ and $j_1 < j_2 < \cdots < j_k$ such that $x_{i_l} = y_{j_l}$ for every $l \in \{1, \ldots, k\}$. This problem is known as the longest common subsequence problem. It has emerged more or less independently in several remarkably disparate areas, including the comparison of versions of computer programs, cryptographic snooping, the mathematical, and molecular biology. The biological motivation of the problem is that long molecules such as proteins and nucleic acids like DNA can be schematically represented as sequences from a finite alphabet. Taking an evolutionary point of view, it is natural to compare two DNA sequences by finding their closest common ancestors. If one assumes that these molecules evolve only through the process of inserting new symbols in the representing strings, then ancestors are substrings of the string that represent the molecule. Thus, the length of the longest common subsequence of two strings is a reasonable measure of how close both strings are. In the mid 1970's, Chvátal and Sankoff [6] established that the expected length of two random k–ary sequences of length n when normalized by n converged to a constant. The value of this constant is unknown although much work has been spent in finding good upper an lower bounds for it (see for example [3] and references therein). This work is partly motivated by our desire to find alternative techniques that would allow for the determination of the value of this constant.

Related Work and Main Contributions: Increasing subsequences and longest common subsequences are well–studied combinatorial problems. As shown above, both these problems can be cast as random graph theoretic problems. The theory of random graphs was founded by Erdös and Rényi during the late 50's and early 60's. By now, it is a well developed subject area that has found many applications, specially in computer science. Informally, random graph theory studies the properties that graphs exhibit when they are chosen according to some fixed distributions. The distributions most commonly considered in the literature are the special case where $H = K_n$ of:

$\mathcal{G}(H,p)$: the binomial random graph model — a distribution over the set of subgraphs of H where for each edge $e \in H$ the events $\{e \in E(G)\}$ have probability p and are mutually independent,

$\mathcal{G}(H,N)$: the uniform random graph model — a distribution over the set of subgraphs of H where an N edge subgraph of H is uniformly chosen (provided one exists),

$\mathcal{G}_r(H,k)$: the random k–regular graph model — a distribution over the set of subgraphs of H where a k–regular subgraph of H is uniformly chosen (provided one exists).

As one might expect, in many cases $\mathcal{G}(K_n,p)$ and $\mathcal{G}(K_n,N)$ are asymptotically equivalent. In fact, in most cases, the equivalence takes place when $N \sim n^2p/2$ (see [10, Theorem 2.1, Ch. 6] for a precise statement of the result). Hence, the most interesting distributions to consider are $\mathcal{G}(K_n,p)$ and $\mathcal{G}_r(K_n,k)$.

In this work we are interested in properties of randomly chosen bipartite graphs. It is natural, given the discussion of the previous paragraph, to focus our study on the distribution $\mathcal{G}(K_{n,n},p)$ and $\mathcal{G}_r(K_{n,n},k)$.

For $0 < p < 1$ and a positive integer k let $L_p(n)$ and $L_k(n)$ be the variable representing the size of the largest planar matching of a graph chosen according to $\mathcal{G}(K_{n,n},p)$ and $\mathcal{G}_r(K_{n,n},k)$ respectively. The main result of this work is:

Theorem 1 *If $k = O(n^{1/5-\varepsilon})$, then $L_k(n)/\sqrt{kn}$ tends to 2 in probability as $n \to \infty$. Convergence in mean to the same limit holds if $k = O(1)$.*

Subadditivity techniques, in particular Kingman's Subadditive Ergodic Theorem [13] yield in a straightforward way the following:

Theorem 2 *For each $0 < p < 1$, there exists a constant $0 < \gamma_p < 1$ such that $L_p(n)/n$ tends to γ_p in probability and in mean as $n \to \infty$.*

A characteristic of limit results like the ones above is that it is often difficult to determine the limiting constant whose existence is proclaimed. One of the interesting aspects of Theorem 1 is that the limiting constant can be determined exactly. Another fact concerning Theorem 1 worth stressing is that when $k = 1$, the distribution of $L_k(n)$ is equal to the distribution of the length of the largest increasing subsequence of a randomly chosen permutation of $\{1, \ldots, n\}$. Thus, determining the size of the largest planar matching of a randomly chosen 1–regular bipartite graph is equivalent to Ulam's problem. Monte Carlo computations led Ulam to suspect that $L_1(n)$ was $\Theta(\sqrt{n})$. But it was Hammersley [12] who proved that $L_1(n)$ does tend, in probability, to $\gamma\sqrt{n}$ for some constant γ. In his paper Hammersley made ingenious use of subadditivity to establish the existence of γ and went on to show that $\pi/2 \leq \gamma \leq e$. Subsequently Kingman [13] narrowed the gap to $\sqrt{8/\pi} \leq \gamma \leq 2.49$. Later, Logan and Shepp [14] based on a result by Schensted [18] proved that $\gamma \geq 2$; finally, Vershik and Kerov [20] obtained that $\gamma \leq 2$ (a combinatorial proof of this result was supplied by Pilpel [17]) Hence,

Theorem 3 ([12, 14, 20]) *$L_1(n)/\sqrt{n}$ tends to 2 in probability and in mean as $n \to \infty$.*

Thus, Theorem 1 can be thought of as a generalization of the limiting behavior shown to hold for the increasing subsequence problem. Our proof of Theorem 1 does not directly use subadditivity techniques. Instead, it relates the limiting behavior of $L_k(n)$ to that of $L_1(n)$. The limiting constant 2 for convergence in probability of $L_1(n)/\sqrt{n}$ in Theorem 1 arises from this relation.

Although Theorem 1 addresses a rather natural generalization of Ulam's problem, to the best of our knowledge it has not been considered in the literature. Curiously, a different generalization has received considerable attention; one that for fixed d and n, considers random points $\vec{x}(1), \ldots, \vec{x}(n)$ chosen independently from the uniform distribution on the unit cube $[0,1]^d$ in Euclidean d–dimensional space. The points then form the underlying set of a random order $P_d(n)$, with partial order $\vec{x}(i) \leq \vec{x}(j)$ when $x_l(i) \leq x_l(j)$ for each $l \in \{1, \ldots, d\}$. Let $H_d(n)$ denote the height of $P_d(n)$, i.e., the number of elements in a longest chain (totally ordered subset) of $P_d(n)$. Steele [19] and Bollobás and Winkler [5] have studied the asymptotic in n behavior of $H_d(n)$ for fixed d. The $d = 2$ version of this problem is equivalent to Ulam's problem.

Standard techniques can be applied to upper and lower bound γ_p and to prove concentration bounds for $L_p(n)$. The determination of the exact value of γ_p is one of the interesting questions raised in this work.

Conventions: The set of integers $\{1, \ldots, m\}$ will henceforth be denoted by $[m]$.

Organization: In Sect. 2 we present the main contribution of this work, i.e., we study the asymptotic behavior of $L_k(n)$ for fixed k. The analysis is divided into two parts. In Sect. 2.1 we show that $L_k(n)/\sqrt{kn}$ is at least 2 when $n \to \infty$. In Sect. 2.2 we prove that the same, but reverse bound holds, i.e., $L_k(n)/\sqrt{kn}$ is at most 2, when $n \to \infty$. Finally, in Sect. 2.3, we state and prove the main results of this work, i.e., those that concern the limit behavior of $L_k(n)/\sqrt{kn}$ when $n \to \infty$. Section 3 states some facts about the asymptotic behavior of $L_p(n)$.

2 Random k–regular Graph Model

Most work on random regular graphs is based on the so called random configuration model of Bender and Canfield and Bollobás [4, Ch. II, § 4]. Below we follow this approach, but first we need to adapt the configuration model to the bipartite graph scenario. Given a fixed positive integer k and two sets of n nodes W and M, which are to be the upper and lower nodes of the graph, a k–configuration of W and M is a one–to–one pairing of $W \times [k]$ and $M \times [k]$. These kn pairs are called edges of the configuration. The natural projection of $W \times [k]$ and $M \times [k]$ onto W and M respectively (ignoring the second coordinate) projects each configuration F to a bipartite multi–graph $\pi(F)$ with W and M as upper and lower nodes. Note in particular that $\pi(F)$ may contain multiple edges (arising from sets of two or more edges in F whose end–points correspond to the same pair of vertices in W and M). However, $\pi(F)$ is a k–regular bipartite multi–graph (with multiple edges counted the natural way). In particular, if $\pi(F)$ lacks multiple edges, it is a k–regular graph. Each k–regular bipartite graph on W and M is the projection of the same number of configurations. It follows that taking the projection $\pi(F)$ of a uniformly chosen k–configuration on W and M and conditioning on it being a simple graph is equivalent to uniformly choosing a random k–regular graph, i.e., to choosing a graph according to $\mathcal{G}_r(K_{n,n}, k)$.

It will be advantageous to allow multiple edges and work with k–regular bipartite multi–graphs and afterwards condition on simple graphs. We thus let $\mathcal{G}_r^*(K_{n,n}, k)$ denote the distribution over k–regular multi–graphs obtained by randomly and uniformly chosing a k–configuration of W and M and computing $\pi(F)$. Note however, that $\mathcal{G}_r^*(K_{n,n}, k)$ does not have the uniform distribution over all k–regular bipartite multi–graphs on W and M (the probability of obtaining a given multi–graph is proportional to a weight consisting of the product of a factor $1/j!$ for each multiple edge of multiplicity j).

Throughout we will focus our attention on the distribution $\mathcal{G}_r^*(K_{n,n}, k)$ and the variable $L_k^*(n)$ defined as $L_k(n)$ but now with respect to $\mathcal{G}_r^*(K_{n,n}, k)$. The reason why this suffices for our purposes is given by the following:

Lemma 1 *Let k be a function from the positive integers into the positive integers such that $k = o(\sqrt{n})$ or $k = o(n)$ is monotonically increasing. Let Q be a graph property of bipartite regular multi–graphs. If $\neg Q$ holds with probability $o(e^{-k^2})$ for a graph randomly chosen according to $\mathcal{G}_r^*(K_{n,n}, k)$, then Q holds for almost every r–regular subgraph of $K_{n,n}$ when $n \to \infty$.*

Proof: In [15] it is shown that the probability that a graph chosen according to $\mathcal{G}_r^*(K_{n,n}, k)$ is simple is $e^{-(k-1)^2/2+o(k^2)}$ if $k = o(n)$ is monotonically increasing. In [21] it is shown that the same result holds when $k = o(\sqrt{n})$ without any assumptions on the monotonicity of k. Thus, under the hypothesis of the lemma,

$$\mathbb{P}_{\mathcal{G}_r(K_{n,n},k)}(\neg Q) \quad \leq \quad e^{(k-1)^2/2+o(k^2)} \cdot \mathbb{P}_{\mathcal{G}_r^*(K_{n,n},k)}(\neg Q) \; .$$

The desired conclusion follows by making $n \to \infty$. ∎

An additional bonus of working with $\mathcal{G}_r^*(K_{n,n}, k)$ is that it also leads to results about uniformly chosen k–regular multi–graphs. For the sake of conciseness we omit these results.

2.1 Lower Bound

The main objective of this section is to bound the probability that $L_k^*(n)/\sqrt{kn}$ is at most $2 - \chi$, for $\chi > 0$. There are a couple of technicalities involved in the formal proof of this fact. In order not to obscure the idea in which the proof relies we begin by informally justifying the above stated lower bound.

Let $W = \{w_1, \ldots, w_n\}$ and $M = \{m_1, \ldots, m_n\}$. Consider the bipartite multi–graph $G = (W, M, E)$ chosen according to $\mathcal{G}_r^*(K_{n,n}, k)$. For $\epsilon > 0$, let m be an integer such that $1 - \epsilon \leq (1 - m/n)^{k-1}$ and $t = \lfloor n/m \rfloor$, let $W_1 = \{w_1, \ldots, w_m\}$, $W_2 = \{w_{m+1}, \ldots, w_{2m}\}$, so on and so forth up to $W_t = \{w_{(t-1)m+1}, \ldots, w_{tm}\}$. Define $M_1, \ldots, M_t$ analogously with respect to M. Finally, let $G_i = (W_i, M_i, E_i)$ be the graphs induced on W_i and M_i by G. Removing from G_i all edges incident to nodes whose degree is at least 2 yields a subgraph of G_i, say G_i'. We claim that the expected size of $E(G_i')$ is at least $(1 - \epsilon)km^2/n$. Indeed, G_i contains $2m$ nodes, each one having $j \in [k]$ incident edges with probability

$$\binom{nk}{k}^{-1} \binom{mk}{j} \binom{(n-m)k}{k-j} .$$

Hence, the expected number of edges incident to a given node of G_i that are removed in order to obtain G_i' is

$$\binom{nk}{k}^{-1} \sum_{j=2}^{k} j \binom{mk}{j} \binom{(n-m)k}{k-j} \quad = \quad \frac{km}{n} \left\{ 1 - n \binom{nk}{k}^{-1} \binom{(n-m)k}{k-1} \right\} .$$

Thus, the expected number of edges removed from G_i in order to obtain G_i', i.e., $\mathbb{E}\left(|G_i| - |G_i'|\right)$ is at most

$$2\frac{km^2}{n}\left\{1 - n\binom{nk}{k}^{-1}\binom{(n-m)k}{k-1}\right\}.$$

From this, it can be shown (although it is not straightforward) that if $km/n \to 0$ as $n \to \infty$, then for $\epsilon > 0$ and large enough n, $\mathbb{E}\left(|G_i| - |G_i'|\right) \le \epsilon km^2/n$. On the other hand, it is easy to see that the expected size of $E(G_i)$ is km^2/n. Thus, the expected size of $E(G_i')$ is at least $(1-\epsilon)km^2/n$. For the sake of argument, assume that the size of $E(G_i')$ was at least $(1-\epsilon)km^2/n$ for all i. Note that, removal of the degree 0 nodes of G_i' yields a 1–regular subgraph of G_i'. By the way in which G was chosen and symmetry arguments, such subgraph, is a randomly chosen perfect matching on its set of nodes. Hence, the discussion of Sect. 1 concerning Ulam's problem implies that, if $km^2/n \to \infty$ as $n \to \infty$, then the expected size of the largest planar matching contained in G_i' (hence in G_i) is at least $2\sqrt{(km^2/n)(1-\epsilon)}$. Since the union of planar matchings contained in distinct G_i's is a planar matching contained in G, the expected size of the largest planar matching contained in G, for large enough n, is at least

$$2t\sqrt{\frac{km^2}{n}(1-\epsilon)} \;=\; 2\left\lfloor\frac{n}{m}\right\rfloor\sqrt{\frac{km^2}{n}(1-\epsilon)} \;\ge\; 2\left(1-\frac{m}{n}\right)\sqrt{kn(1-\epsilon)}\,,$$

i.e., asymptotically larger than $2(1-\epsilon)^{3/2}\sqrt{kn}$ provided $m \le \epsilon n$. Observing that when $k = o(n)$ one can choose $m = o(n)$ so $km^2/n \to 0$ while $km/n \to 0$ as $n \to \infty$, we conclude that the expected size of the largest planar matching contained in G, for large enough n, is at least $2(1-o(1))\sqrt{kn}$.

We now formalize and strengthen the given proof argument. First, we need to introduce additional notation. Let F be a k–configuration of W and M and let $G = \pi(F)$. For $w \in W_i$, $m \in M_i$, and $a, b \in [k]$, let $X^{(i)}_{w,a;m,b}$ equal 1 if $((w,a),(m,b))$ is an edge of F. Note that the number of edges of G_i is

$$|E(G_i)| \;=\; \sum_{w\in W_i, a\in[k]} Y^{(i)}_{w,a}\,, \quad \text{where} \quad Y^{(i)}_{w,a} \;=\; \sum_{m\in M_i, b\in[k]} X^{(i)}_{w,a;m,b}\,.$$

Observe that $Y^{(i)}_{w,a}$ is the indicator variable of the event "there is an edge in G_i that is the projection of an edge in F with first component (w,a)." Note also that $\mathbb{E}\left(Y^{(i)}_{w,a}\right) = m/n$. Hence, by linearity of expectation $\mathbb{E}\left(|E(G_i)|\right) = km^2/n$. We will show two things concerning G_i; (1) with overwhelming probability $|E(G_i)|$ is almost $\mathbb{E}\left(|E(G_i)|\right)$, and, (2) with overwhelming probability few of the edges of G_i have common endpoints.

Since $|E(G_i)|$ equals the sum of indicator variables $Y^{(i)}_{w,a}$, a Chernoff–Hoeffding (CH) bound would seem appropriate in order to achieve our first objective. Although the indicator variables $Y^{(i)}_{w,a}$ are not independent, for fixed i, they are negatively correlated so for purposes of stochastic bounds on their sum one can treat the variables as if they were independent (although this fact seems to have been observed in several occasions throughout the literature it has only recently been explicitly stated and exhaustively studied by Dubhashi and Ranjan [7, Proposition 7]). Specifically, the application of a CH type bound yields the following:

Lemma 2 *If $1 \leq m \leq n$, and $\epsilon > 0$, then*

$$\mathbb{P}\left(|E(G_i)| < (1-\epsilon)\frac{km^2}{n}\right) \quad \leq \quad e^{-\epsilon^2 km^2/2n} .$$

Proof: First, observe that for all $C \subseteq W_i \times [k]$, the variables $\left\{Y_{w,a}^{(i)} : (w,a) \in C\right\}$ are negatively correlated. Indeed, recall that $\mathbb{E}\left(Y_{w,a}^{(i)}\right) = m/n$ and observe that

$$\mathbb{E}\left(\prod_{(w,a)\in C} Y_{w,a}^{(i)}\right) \; = \; \prod_{j=1}^{|C|} \frac{mk-j}{nk-j} \; \leq \; \left(\frac{m}{n}\right)^{|C|} \; = \; \prod_{(w,a)\in C} \mathbb{E}\left(Y_{w,a}^{(i)}\right) .$$

To show that the CH bounds apply to the sum $|E(G_i)| = \sum_{w\in W_i, a\in[k]} Y_{w,a}^{(i)}$, we use the standard proof of the CH bound (for completeness sake, the statement and proof of the CH bounds we use throughout this work is included in Appendix 3). The only change needed is the crucial step, where one uses the fact that for $s \in \mathbb{R}$ and independent variables $X_1, \ldots, X_N$, $\mathbb{E}\left(\prod_i e^{sX_i}\right) = \prod_i \mathbb{E}\left(e^{sX_i}\right)$. For negatively correlated variables, we have, for $s \geq 0$, $\mathbb{E}\left(\prod_i e^{sX_i}\right) \leq \prod_i \mathbb{E}\left(e^{sX_i}\right)$. The rest of the proof is unchanged. ∎

We would now like to achieve our second objective, i.e., to show that few of the edges of G_i have common endpoints. Hence, with overwhelming probability G_i contains a subgraph that is a perfect matching whose size is almost $|E(G_i)|$.

Lemma 3 *Let $k \geq 2$ and $km \leq n$. Let G_i' be the edge subgraph obtained from G_i by removal of all edges incident to nodes whose degree is at least 2. If $\mu = k(m/n)\left\{1 - n\binom{nk}{k}^{-1}\binom{(n-m)k}{k-1}\right\}$, and $\beta > 1$, then*

$$\mathbb{P}\left(|E(G_i)| - |E(G_i')| \geq 2\beta m\mu\right) \quad \leq \quad 2\left(e^{\beta-1}\beta^{-\beta}\right)^{m\mu/k} .$$

Proof: For a node v either in W_i or M_i let $deg(v)$ denote the degree of v in G_i and let $X_v^{(i)}$ equal 0 if $deg(v) \leq 1$ and $deg(v)$ otherwise. Observe that $|E(G_i)| - |E(G_i')| \leq \sum_{w\in W_i} X_w^{(i)} + \sum_{m\in M_i} X_m^{(i)}$. Hence, by symmetry arguments,

$$\mathbb{P}\left(|E(G_i)| - |E(G_i')| \geq 2\beta m\mu\right) \quad \leq \quad 2\mathbb{P}\left(\sum_{w\in W_i} X_w^{(i)} \geq \beta m\mu\right) .$$

Now, for $w \in W_i$, let $Z^{(i)}_{w,a}$ be the indicator variable of the event "there are edges in G_i that are the projection of edges in F with first component (w,a) and (w,b) for some $b \in [k]$, $b \neq a$." Clearly,

$$X^{(i)}_w = \sum_{a\in[k]} Z^{(i)}_{w,a} .$$

In Appendix B we show that the variables $Z^{(i)}_{w_1,a_1}, \ldots, Z^{(i)}_{w_l,a_l}$ are negatively correlated provided $w_1, \ldots, w_l \in W_i$ are all distinct no matter what the values of $a_1, \ldots, a_l \in [k]$ are. It follows that for all $C \subseteq W_i$ the variables $\left\{X^{(i)}_w : w \in C\right\}$ are negatively correlated. Hence, as in the proof of Lemma 2, we can apply CH bounds to the sum $\sum_{w\in W_i} X^{(i)}_w$.

Note that $\mathbb{E}\left(X^{(i)}_w\right) = \mu$, since

$$\begin{aligned}
\mathbb{E}\left(X^{(i)}_w\right) &= \binom{nk}{k}^{-1} \sum_{j=2}^{k} j \binom{mk}{j}\binom{(n-m)k}{k-j} \\
&= mk\binom{nk}{k}^{-1}\left\{\sum_{j=1}^{k}\binom{mk-1}{j-1}\binom{(n-m)k}{k-j} - \binom{(n-m)k}{k-1}\right\} \\
&= \frac{km}{n}\left\{1 - n\binom{nk}{k}^{-1}\binom{(n-m)k}{k-1}\right\} .
\end{aligned}$$

Thus, since $X^{(i)}_w$ ranges over $[k]$ and $\left\{X^{(i)}_w : w \in W_i\right\}$ is a family of negatively correlated variables, a CH bound (see precise statement in Appendix 3) yields,

$$\mathbb{P}\left(\sum_{w\in W_i} X^{(i)}_w \geq \beta m\mu\right) = \mathbb{P}\left(\sum_{w\in W_i} (X^{(i)}_w - \mu) \geq (\beta-1)m\mu\right) \leq \left(e^{\beta-1}\beta^{-\beta}\right)^{m\mu/k} .$$

■

Lemma 2 and Lemma 3 put together yield

Proposition 1 *Let k, n, m, p, q, μ, ϵ, and β be as in the statements of Lemma 2 and Lemma 3 but also such that $2\beta\mu \leq \epsilon km/n$. Then,*

$$\mathbb{P}\left(|E(G'_i)| \leq (1-2\epsilon)(km^2/n)\right) \leq e^{-\epsilon^2 km^2/2n} + 2(e^{\beta-1}\beta^{-\beta})^{m\mu/k} .$$

Proof: Since $2\beta\mu m \leq \epsilon km^2/n$ we have

$$\begin{aligned}
&\mathbb{P}\left(|E(G'_i)| < (1-2\epsilon)km^2/n\right) \leq \\
&\leq \mathbb{P}\left(|E(G_i)| < (1-\epsilon)km^2/n\right) + \mathbb{P}\left(|E(G_i)| - |E(G'_i)| \geq \beta\mu m\right) .
\end{aligned}$$

Hence, Lemma 2 and Lemma 3 imply the desired conclusion. ■

Corollary 1 *For every $\chi, \varepsilon > 0$, if $k = O(n^{1/5-\varepsilon})$, then*

$$\mathbb{P}\left(\frac{L_k^*(n)}{\sqrt{kn}} \le 2-\chi\right) = e^{-k^2\omega(1)}.$$

Proof: Let m, p, q, μ, ϵ, and β be as in Lemma 2 and Lemma 3. In addition, let $1/2 > \theta$ and choose m so $m^2/n = k^{2/\theta-1}\omega(1) = k^2\omega(1)$ and $km/n = o(1)$ (this can always be done if $k = O(n^{1/5-\varepsilon})$) and let $N_k(n) = \lceil km^2/n \rceil$. Thus, if $t = \lfloor n/m \rfloor$, then

$$\mathbb{P}\left(\frac{L_k^*(n)}{\sqrt{kn}} \le 2-\chi\right) \le \sum_{i=1}^{t} \mathbb{P}\left(|E(G_i')| < N_k(n)\right) + t \cdot \mathbb{P}\left(L_1(N_k(n)) \le (2-\chi)\frac{\sqrt{kn}}{t}\right).$$

Observe that

$$n\binom{nk}{k}^{-1}\binom{(n-m)k}{k-1} = \frac{nk}{nk-(k-1)}\prod_{i=0}^{k-2}\left(1-\frac{mk}{nk-i}\right) \ge \left(1-\frac{m}{n-1}\right)^{k-1} \ge$$

$$\ge 1-\frac{km}{n-1}.$$

It follows that $\mu \le k(m/n)(km/(n-1))$. Choose $\beta = \epsilon/(4(km/n))$ and observe that since $km/n = o(1)$, for large enough n it holds that $km \le n$, $\beta \ge e^2$, and $(\epsilon/2)(km/n) \le 2\beta\mu \le \epsilon(km/n)$. Hence, Proposition 1 holds. Moreover, since $m^2/n = k^2 w(1)$, $e^{\beta-1}\beta^{-\beta} \le (1/e)^{\beta}$, and $\beta m\mu/k \ge (\epsilon/4)(m^2/n) = k^2\omega(1)$. Thus,

$$\sum_{i=1}^{t} \mathbb{P}\left(|E(G_i')| < N_k(n)\right) \le t\left(e^{-\epsilon^2 km^2/2n} + 2(e^{\beta-1}\beta^{-\beta})^{m\mu/k}\right) \le e^{-k^2\omega(1)}.$$

Also, Frieze [9] has shown that for every θ arbitrarily close to $1/2$ there is a $\theta < \alpha < 1/2$ such that for all sufficiently large N,

$$\mathbb{P}\left(|L_1(N) - \mathbb{E}\left(L_1(N)\right)| \ge N^{\alpha}\right) \le e^{-N^{\theta}}.$$

Baik, Deift, and Johansson [2] have shown that for some constant c, $\mathbb{E}\left(L_1(N)\right) = 2\sqrt{N} - cN^{1/6} + o(N^{1/6})$. Thus, for sufficiently large N, $\mathbb{P}\left(L_1(N) \le 2\sqrt{N} - N^{\alpha}/2\right) \le e^{-N^{\theta}}$. Observe that $\sqrt{N_k(n)} \ge \sqrt{kn}/t$. Thus, for large enough $N_k(n)$ we have that $(N_k(n))^{\alpha}/2 \le \chi\sqrt{N_k(n)}$ and

$$\mathbb{P}\left(L_1(N_k(n)) \le (2-\chi)\frac{\sqrt{kn}}{t}\right) \le \mathbb{P}\left(L_1(N_k(n)) \le 2\sqrt{N_k(n)} - (N_k(n))^{\alpha}/2\right) \le$$

$$\le e^{-N_k(n)^{\theta}}.$$

Hence, since $m^2/n = k^{2/\theta-1}\omega(1)$, we have $N_k(n) = k^{2/\theta}\omega(1)$ and

$$t \cdot \mathbb{P}\left(L_1(N_k(n)) \le (2-\chi)\frac{\sqrt{kn}}{t}\right) \le t \cdot e^{-N_k(n)^{\theta}} \le e^{-k^2\omega(1)}.$$

∎

2.2 Upper Bound

In this section, our main goal is to bound the probability that $L_k^*(n)/\sqrt{kn}$ is at least $2+\chi$, for $\chi > 0$. Our proof of this fact relies on a combinatorial argument based on a natural correspondence between configurations and perfect matchings. To establish this correspondence recall that given two set of nodes W and M, $W = \{w_1, \ldots, w_n\}$ and $M = \{m_1, \ldots, m_n\}$, a k–configuration F of W and M is a bijection between $W \times [k]$ and $M \times [k]$. To every such bijection, one can associate in a natural way a 1–regular subgraph $G(F)$ of $K_{kn,kn}$ by means of identifying the j–th upper (respectively lower) node of $K_{kn,kn}$ to $(w_q, r) \in W \times [k]$ (respectively $(m_q, r) \in M \times [k]$) where $q \in \{1, \ldots, n\}$ and $r \in \{1, \ldots, k\}$ are such that $j = (q-1)k+r$. The following results establishes a combinatorial relation between the largest planar matchings of $G(F)$ and $\pi(F)$.

Lemma 4 *For every configuration F, the largest planar matching in $G(F)$ is at least as large as the largest planar matching in $\pi(F)$.*

Proof: Recall that $\pi(F)$ is a k–regular multi–graph with upper and lower set of vertices W and M respectively. Let $(w_{i_1}, m_{j_1}), \ldots, (w_{i_l}, m_{j_l})$ be the largest planar matching in $\pi(F)$. Planarity implies that $i_1 < \ldots < i_l$ and $j_1 < \ldots < j_l$. By definition of $\pi(\cdot)$ there exists $k_{w_{i_1}}, \ldots, k_{w_{i_l}}$ and $k_{m_{j_1}}, \ldots, k_{m_{j_l}}$ in $\{1, \ldots, k\}$, such that the edge (w_{i_j}, m_{j_j}) of $\pi(F)$ is the projection of the pairing $((w_{i_s}, k_{w_{j_s}}), (m_{i_s}, k_{m_{j_s}}))$ of F. Note that each such pairing is also an edge of $G(F)$. Moreover, since $i_1 < \ldots < i_l$ and $j_1 < \ldots < j_l$, the collection of these l pairings is a planar matching in $G(F)$. ∎

Corollary 2 *For every $\chi, \varepsilon > 0$, if $k = O(n^{1/3-\varepsilon})$, then*

$$\mathbb{P}\left(\frac{L_k^*(n)}{\sqrt{kn}} \geq 2+\chi\right) \quad = \quad e^{-k^2\omega(1)} .$$

Proof: Since every k–regular configuration F of W and M, $|W| = |M| = n$, is in one–to–one correspondence with a 1–regular subgraph $G(F)$ of $K_{kn,kn}$, to every probability space where it makes sense to define $L_k^*(n)$ one can associate, in a natural way, a probability space over which $L_1(kn)$ is well defined. Thus, Lemma 4 yields that

$$0 \;\leq\; \mathbb{P}\left(\frac{L_k^*(n)}{\sqrt{kn}} \geq 2+\chi\right) \;\leq\; \mathbb{P}\left(\frac{L_1(kn)}{\sqrt{kn}} \geq 2+\chi\right) .$$

Recall that Frieze [9] has shown that for every θ arbitrarily close to $1/2$ there is a $\theta < \alpha < 1/2$ such that for all sufficiently large N,

$$\mathbb{P}\left(|L_1(N) - \mathbb{E}\left(L_1(N)\right)| \geq N^\alpha\right) \quad \leq \quad e^{-N^\theta} ,$$

and Baik, Deift, and Johansson [2] have shown that for some constant c for sufficiently large N, $\mathbb{E}\left(L_1(N)\right) \leq 2\sqrt{N}$. Hence, for large enough n,

$$\mathbb{P}\left(\frac{L_1(kn)}{\sqrt{kn}} \geq 2+\chi\right) \;\leq\; \mathbb{P}\left(L_1(kn) - \mathbb{E}\left(L_1(kn)\right) \geq (kn)^\alpha\right) \;\leq\; e^{-(kn)^\theta} .$$

The desired conclusion follows choosing θ sufficiently close to $1/2$. ∎

2.3 Convergence Results

In this section we present the main limiting results concerning $L_k/\sqrt{kn}$ established in this work.

Lemma 5 *If $k = O(n^{1/5-\varepsilon})$, then*

$$\mathbb{P}\left(\left|\frac{L_k^*(n)}{\sqrt{kn}} - 2\right| \geq \chi\right) \quad = \quad e^{-k^2\omega(1)} .$$

In particular, $L_k^*(n)/\sqrt{kn} \to 2$ in probability when $n \to \infty$.

Proof: Immediate from Corollary 1 and Corollary 2. ∎

Corollary 3 *If $k = O(n^{1/5-\varepsilon})$, then $L_k(n)/\sqrt{kn} \to 2$ in probability when $n \to \infty$.*

Proof: Immediate from Lemma 1 and Lemma 5. ∎

To derive convergence in mean results we rely on the fact that when convergence in probability holds, uniform integrability is equivalent to convergence in mean (see [11, Ch. 7, § 10, Theorem (3)]).

Lemma 6 *If $k = O(1)$, then $L_k(n)/\sqrt{kn} \to 2$ in mean when $n \to \infty$. In particular, $\mathbb{E}\left(L_k(n)\right)/\sqrt{kn} \to 2$ when $n \to \infty$.*

Proof: Corollary 3 insures that $L_k(n)/\sqrt{kn}$ converges to 2 in probability when $n \to \infty$. Hence, to obtain the result we seek it suffices to prove that $(L_k(n)/\sqrt{kn})_{n\geq 1}$ is uniformly integrable.

Recall that Theorem 3 says that $L_1(kn)/\sqrt{kn}$ converges in probability and in mean when $n \to \infty$. Hence, $(L_1(kn)/\sqrt{kn})_{n\geq 1}$ is uniformly integrable. But, by Lemma 4, $L_k^*(n)$ is dominated by $L_1(kn)$. It follows that $(L_k^*(n)/\sqrt{kn})_{n\geq 1}$ is uniformly integrable. In [15] it is shown that the probability that a graph chosen according to $\mathcal{G}_r^*(K_{n,n}, k)$ is simple is $e^{-(k-1)^2/2+o(k^2)}$ if $k = o(n)$ is monotonically increasing. Hence,

$$\mathbb{E}\left(\frac{L_k(n)}{\sqrt{kn}}\right) \leq e^{(k-1)^2/2+o(k^2)} \cdot \mathbb{E}\left(\frac{L_k^*(n)}{\sqrt{kn}}\right) .$$

Since $k = O(1)$ and $(L_k^*(n)/\sqrt{kn})_{n\geq 1}$ is uniformly integrable we conclude that $(L_k(n)/\sqrt{kn})_{n\geq 1}$ is uniformly integrable as desired. ∎

Finally, observe that Lemma 5 and Lemma 6 put together yield Theorem 1.

	1/8	1/4	3/8	1/2	5/8	3/4	7/8
$2/S_p$	0.446	0.584	0.681	0.757	0.823	0.883	0.940
α_p	0.632	0.773	0.853	0.905	0.942	0.969	0.988

Figure 2: Numerical values of $2/S_p$ and α_p.

3 Binomial Random Graph Model

In this section we study the behavior of $L_p(n)$. It is easy to see that $(\mathbb{E}(L_p(n)))_{n\geq 1}$ is a superadditive sequence. Thus, $\mathbb{E}(L_p(n))/n$ has a limit and the convergence is from below, that is, there exists $0 \leq \gamma_p \leq 1$ such that

$$\limsup_{n\to\infty} \frac{\mathbb{E}(L_p(n))}{n} = \sup_{n\to\infty} \frac{\mathbb{E}(L_p(n))}{n} = \gamma_p .$$

Consider now the integer lattice $\{1,2,\ldots\} \times \{1,2,\ldots\}$. For each lattice point flip independently a coin with probability p of landing heads and form the underlying set P_p of lattice points for which the outcome of the coin flip was heads. The points in P_p form a random poset, with partial order given by $(i,j) \leq (i',j')$ when $i \leq i'$ and $j \leq j'$. Note that $L_p(n)$ is the height of $P_p(n) = P_p \cap \{1,2,\ldots,n\} \times \{1,2,\ldots,n\}$, i.e., the number of elements in a longest chain (totally ordered subset) of $P_p(n)$. If $s < t$ run over the set of positive integers, the family of variables $(X_{s,t})_{s,t}$ where $X_{s,t}$ denotes the height of $P_p \cap \{s,s+1,\ldots,t\} \times \{s,s+1,\ldots,t\}$ is a stochastic superadditive process. Kingman's subadditive ergodic theorem [13] further implies that $L_p(n)/n = X_{1,n}/n \to \gamma_p$ in probability and in mean when $n \to \infty$.

We now obtain a lower bound on γ_p.

Lemma 7 *For each $0 < p < 1$ let $S_p = 2 + \sum_{s\geq 2}(1-p)^{\binom{s}{2}}$. It holds that $\gamma_p \geq 2/S_p$. (Some numerical values of $2/S_p$ are shown in Fig. 2.)*

Proof: The proof relies in a greedy construction of a chain among the points in P_p. Specifically, (i_1,j_1) is chosen from the elements of P_p so $s_1 = i_1 + j_1$ is minimal. Thereafter, (i_m,j_m) is the point satisfying $(i_m,j_m) > (i_{m-1},j_{m-1})$ for which $s_m = i_m + j_m$ is minimal. Since the probability that $\{(i,j) \in P_p : i+j < s\} = \emptyset$ is 1 if $s \leq 2$ and $(1-p)^{\binom{s-1}{2}}$ otherwise,

$$S_p = \mathbb{E}(s_1) = \sum_{s\geq 1} \mathbb{P}(s_1 \geq s) = 2 + \sum_{s\geq 3}(1-p)^{\binom{s-1}{2}} = 2 + \sum_{s\geq 2}(1-p)^{\binom{s}{2}} .$$

Now, the differences $i_m - i_{m-1}$ and $i_1 - 0$ are independent and identically distributed with mean

$$\mathbb{E}(i_1) = \frac{S_p}{2} = 1 + \frac{1}{2}\sum_{s\geq 2}(1-p)^{\binom{s}{2}} .$$

The strong Law of Large Numbers implies that $\lim_{n\to\infty}(i_n/n) = S_p/2$ with probability one. Analogously, $\lim_{n\to\infty}(j_n/n) = S_p/2$ with probability one. So, if $t(n) = \max\{i_n, j_n\}$, $L_p(t(n)) \geq n$. Hence,

$$\gamma_p = \lim_{n\to\infty} \frac{\mathbb{E}\left(L_p(n)\right)}{n} = \lim_{n\to\infty} \frac{\mathbb{E}\left(L_p(t(n))\right)}{t(n)} \geq \lim_{n\to\infty} \frac{n}{t(n)} = \frac{2}{S_p}.$$

■

We shall now upper bound γ_p.

Lemma 8 *For $0 < p < 1$ let $h_p\colon [0,1] \to \mathbb{R}$ be such that $h_p(x) = (x/\sqrt{p})^x(1-x)^{1-x}$. There is a unique $\alpha_p \in [0,1]$ such that $h_p(\alpha_p) = 1$ and $\gamma_p \leq \alpha_p$. (Some numerical values of α_p are shown in Fig. 2.)*

Proof: Note that $\lim_{x\to 0} h_p(x) = 1$, $\lim_{x\to 1} h_p(x) = 1/\sqrt{p} > 1$, and

$$\frac{dh_p(x)}{dx} = h_p(x) \log\left(\frac{x}{\sqrt{p}(1-x)}\right),$$

which is negative for $0 < x < 1 - 1/(1+\sqrt{p})$ and positive for $1 - 1/(1+\sqrt{p}) < x < 1$. Hence, for each p, there is a unique $0 < x < 1$ such that $h_p(x) = 1$ and if $x \geq \alpha_p$ then $1/h_p(x) \leq 1/h_p(\alpha_p) = 1$.

We now prove the bound on γ_p. Let $X_p(m,n)$ be the number of planar matchings of size m in a graph chosen according to $\mathcal{G}(K_{n,n}, p)$. Then,

$$\mathbb{E}\left(X_p(m,n)\right) = \binom{n}{m}^2 p^m \sim \frac{n}{2\pi m(n-m)} \left(\frac{n^n}{m^m(n-m)^{n-m}}\right)^2 p^m.$$

Let c be a constant such that $0 < c < 1$. If $m = cn$, then

$$\frac{n}{2\pi m(n-m)} = \frac{1}{2\pi c(1-c)n}, \quad \text{and} \quad \frac{n^n}{m^m(n-m)^{n-m}} = \left(\frac{1}{c^c(1-c)^{1-c}}\right)^n.$$

Hence, $\mathbb{E}\left(X_p(m,n)\right) = O(1/n)(\sqrt{p^c}/(c^c(1-c)^{1-c}))^{2n} = O(1/n)(1/h_p(c))^{2n}$, and

$$\mathbb{E}\left(\frac{L_p(n)}{n}\right) \leq c\,\mathbb{P}\left(L_p(n) \leq m\right) + \mathbb{P}\left(L_p(n) > m\right).$$

Since $L_p(n) > m$ implies that $X_p(m,n) \geq 1$, Markov's inequality yields that $\mathbb{P}\left(L_p(n) > m\right) \leq \mathbb{E}\left(X_p(m,n)\right)$. Hence, $\mathbb{E}\left(L_p(n)/n\right) \leq c + O(1/n)(1/h_p(c))^{2n}$. If $c \geq \alpha_p$, then $1/h_p(c) \leq 1$ so $\mathbb{E}\left(L_p(n)/n\right) \leq c + O(1/n) \to c$ as $n \to \infty$. Thus, $\gamma_p \leq \alpha_p$. ■

We now comment on the rate at which convergence of $L_p(n)/n$ takes place. Our observations are inspired on the work of Alexander [1]. Specifically, on the method he introduced whereby in many concrete problems of percolation theory, one can supplement subadditivity with a form of superadditivity that lead to rate results. This method is applied, in [1], to derive a rate of convergence result for the mean length of the LCS. An analogous argument yields and interpretation of $L_p(n)$ as the minimal total passage time of a dependent first–passage percolation problem. Indeed, consider the integer lattice $[2n] \times [2n]$, with horizontal and vertical bonds between nearest–neighbor sites of the lattice (that is, pairs (i, j) and (i', j') where $|i - i'| + |j - j'| = 1$) and a diagonal bond from each $(i-1, j-1)$ to (i, j), $1 \leq i, j \leq 2n$. The passage time of each horizontal and vertical bond is defined to be 1, and the passage time of the diagonal bond from $(i-1, j-1)$ to (i, j) is 0 if (w_i, m_j) is an edge of a graph chosen according to $\mathcal{G}(K_{n,n}, p)$, and ∞ otherwise. Then, $2(n - L_p(n))$ represents the minimal total passage time among all paths from $(0, 0)$ to (n, n) for which each coordinate is non–decreasing. The two characteristics satisfied by the above stated percolation problem that allows Alexander's method to go through are:

(i) Either deleting or inserting and edge in a graph chosen according to $\mathcal{G}(K_{n,n}, p)$ changes the minimal total passage time in at most 2, and,

(ii) If l_n denotes the diagonal from $(0, 2n)$ to $(2n, 0)$ and Γ a path that intersects l_n, then the two segments into which Γ is split by l_n are independent.

Hence, the analysis of [1] for the LCS is applicable here verbatim, and yields the following:

Theorem 4 *There exists a constant C such that for every $0 < p < 1$ and $n \geq 1$,*

$$\gamma_p n \ \geq \ \mathbb{E}\left(L_p(n)\right) \ \geq \ \gamma_p n - \frac{C}{\sqrt{n \log n}} .$$

Acknowledgments

We would like to thank Ricardo Baeza–Yates, Alejandro Maass, Brendan McKay, Jiri Matousek, and Jaime San Martin for interesting and helpful discussions, Michele Zito for bringing to our attention the work of Mckay and Wormald [16], and Raul Gouet for comments on an earlier draft.

References

[1] K. S. Alexander. The rate of convergence of the mean length of the longest common subsequence. *The Annals of Applied Probability*, 4(4):1074–1082, January 1994.

[2] J. Baik, P. A. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12:1119–1178, 1999.

[3] R. Baeza–Yates, G. Navarro, R. Gavaldá, and R. Schehing. Bounding the expected length of the longest common subsequences and forests. *Theory of Computing Systems*, 32(4):435-452, 1999.

[4] B. Bollobás. *Random Graphs.* Academic Press, 1985.

[5] B. Bollobás and P. Winkler. The longest chain among random points in Euclidean space. *Proceedings of the American Mathematical Society*, 103(2):347–353, June 1988.

[6] V. Chvátal and D. Sankoff. Longest common subsequences of two random sequences. *J. Appl. Prob.*, 12:306–315, 1975.

[7] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. *Random Structures & Algorithms*, 13(2):99–124, 1998.

[8] P. Erdös and G. K. Szekeres. A combinatorial problem in geometry. *Compositio Math.*, 2:463–470, 1935.

[9] A. Frieze. On the length of the longest monotone subsequence in a random permutation. *The Annals of Applied Probability*, 1(2):301–305, 1991.

[10] R. L. Graham, M. Grötschel, and L Lovász, editors. *Handbook of Combinatorics*, volume I. MIT Press/North–Holland, first edition, 1995.

[11] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes.* Oxford Science Publications, second edition, 1992.

[12] J. M. Hammersley. A few seedlings of research. In *Proc. Sixth Berkeley Sympos. Math. Stat. Prob.*, 345–394, Berkeley, Calif., 1972. Univ. of California Press.

[13] J. F. C. Kingman. Subadditive ergodic theory. *The Annals of Probability*, 1(6):883–909, 1973.

[14] B. F. Logan and L. A. Shepp. A variational problem or random Young tableaux. *Adv. in Math.*, 26:206–222, 1977.

[15] B. D. McKay. Asymptotics for 0-1 matrices with prescribed line sums. *Enumeration and Design*, Academic Press, 225–238, 1984.

[16] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, 1991

[17] S. Pilpel. Descending subsequences of random permutations. Technical Report 52283, IBM, 1986.

[18] C. Schensted. Longest increasing and decreasing subsequences. *Canad. J. Math.*, 13:179–191, 1961.

[19] M. Steele. Limit properties of random variables associated with a partial ordering of $\mathbb{R}^d$. *The Annals of Probability*, 5(3):395–403, 1977.

[20] A. M. Vershik and S. V. Kerov. Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux. *Dokl. Akad. Nauk SSSR*, 233:1024–1028, 1977.

[21] Xiaoji Wang, Asymptotic enumeration of bipartite graphs, tournaments, digraphs, and eulerian digraphs with multiple edges. Ph.D Thesis, Australian National University, 1991.

A Proof of a Chernoff–Hoeffding type bound

For completeness sake, we present in this section the statement and proof of the CH bound used throughout this work. We will need the following elementary result:

Proposition 2 *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a convex function. If Z is a variable that takes values in $[0,1]$, then*

$$\mathbb{E}\left(f(Z)\right) \quad \leq \quad \mathbb{E}\left(Z\right) f(1) + (1 - \mathbb{E}\left(Z\right)) f(0)\,.$$

Lemma 9 *Let $X_1, \ldots, X_m$ be independent identically distributed variables each ranging over $[l]$ with mean μ. If $X = X_1 + \ldots + X_m$, then for $a > 0$,*

$$\mathbb{P}\left(X - \mathbb{E}\left(X\right) \geq a\right) \quad \leq \quad e^{(a-(a+m\mu)\ln(1+a/m\mu))/l}\,,$$

and,

$$\mathbb{P}\left(X - \mathbb{E}\left(X\right) \leq -a\right) \quad \leq \quad e^{-a^2/2lm\mu}\,.$$

Proof: Let $\alpha = \ln(1 + a/m\mu)$. Then, applying Proposition 2 with $X = X_i/l$ and $f(x) = e^{\alpha x}$ yields

$$\mathbb{E}\left(e^{\alpha X_i/l}\right) \;\leq\; 1 - \frac{\mu}{l}(1 - e^{\alpha}) \;=\; 1 + \frac{a}{lm}\,.$$

Now,

$$\mathbb{E}\left(e^{\alpha X/l}\right) \;=\; \mathbb{E}\left(\prod_{i=1}^{m} e^{\alpha X_i/l}\right) \;=\; \prod_{i=1}^{m} \mathbb{E}\left(e^{\alpha X_i/l}\right) \;\leq\; \left(1 + \frac{a}{lm}\right)^m \;\leq\; e^{a/l}\,.$$

Hence, by Markov's inequality and since $\mathbb{E}\left(X\right) = m\mu$,

$$\mathbb{P}\left(X - \mathbb{E}\left(X\right) \geq a\right) \;=\; \mathbb{P}\left(e^{\alpha X/l} \geq e^{\alpha(a+m\mu)/l}\right) \;\leq\; e^{-\alpha(a+m\mu)/4\;l}\,\mathbb{E}\left(e^{\alpha X/l}\right) \;=$$

$$=\; e^{(a-(a+m\mu)\ln(1+a/m\mu))/l}\,.$$

To prove the second inequality, consider $\psi > 0$. Then, applying Proposition 2 with $X = X_i/l$ and $f(x) = e^{-\psi x}$ yields

$$\mathbb{E}\left(e^{-\psi X_i/l}\right) \;\leq\; 1 - \frac{\mu}{l}(1 - e^{-\psi}) \;\leq\; e^{\mu(e^{-\psi}-1)/l}\,.$$

We employ the inequality $e^{-\psi} \leq 1 - \psi + \psi^2/2$, valid for all $\psi > 0$, to obtain,

$$\mathbb{E}\left(e^{-\psi X_i/l}\right) \;\leq\; e^{-\mu(\psi-\psi^2/2)/l}\,.$$

Now,

$$\mathbb{E}\left(e^{-\psi X/l}\right) \;=\; \mathbb{E}\left(\prod_{i=1}^{m} e^{-\psi X_i/l}\right) \;=\; \prod_{i=1}^{m}\mathbb{E}\left(e^{-\psi X_i/l}\right) \;\leq\; e^{-m\mu(\psi-\psi^2/2)/l}\,.$$

Hence, by Markov's inequality and since $\mathbb{E}(X) = m\mu$,

$$\mathbb{P}\left(X - \mathbb{E}(X) \leq -a\right) \;=\; \mathbb{P}\left(e^{-\psi X/l} \geq e^{\psi(a-m\mu)/l}\right) \;\leq\; e^{-\psi(a-m\mu)/l}\mathbb{E}\left(e^{-\psi X/l}\right) \;\leq$$

$$\leq\; e^{\psi(m\mu\psi/2-a)l}\,.$$

Set $\psi = a/m\mu$ to optimize the inequality. ■

B Negative Correlation

Lemma 10 *Let $k \geq 2$, $m \geq l$, and $km \leq n$. Let $Z^{(i)}_{w,a}$ be the indicator variable of the event "there are edges in G_i that are the projection of edges in F with first component (w,a) and (w,b) for some $b \in [k]$, $b \neq a$." Then, for all $a_1,\ldots,a_l \in [k]$ the variables $Z^{(i)}_{w_1,a_1},\ldots,Z^{(i)}_{w_l,a_l}$ are negatively correlated provided $w_1,\ldots,w_l \in W_i$ are all distinct.*

Proof: For simplicities sake we will drop all the super–indices (i). Also we denote $T!/(T-a)!$ by $(T)_a$. Recall that $(T)_{a+b} = (T)_a(T-a)_b$. For positive integers N, M, l, and k such that $N \geq M + lk$ and $M \geq lk$, let

$$P_{j_1,\ldots,j_l}(N,M,l) \quad = \quad \frac{(M)_{j_1+\ldots+j_l+l}(N-M)_{l(k-1)-j_1-\ldots-j_l}}{(N)_{lk}}\,.$$

Claim 1 *If $N = nk$, $M = mk$, $n \geq m + l$, and $m \geq l$, then*

$$\mathbb{E}\left(\prod_{s=1}^{l} Z_{w_s,a_s}\right) \quad = \quad \sum_{j_1,\ldots,j_l=1}^{k-1}\prod_{s=1}^{l}\binom{k-1}{j_s}P_{j_1,\ldots,j_l}(N,M,l)\,. \qquad (1)$$

Proof: Let E be the event that "there are edges in G_i that are the projection of edges in F with first component $(w_1,a_1),\ldots,(w_l,a_l)$." Observe that $\mathbb{E}\left(\prod_{s=1}^{l} Z_{w_s,a_s}\right)$ corresponds to the probability that E occurs and $deg(w_1),\ldots,deg(w_l)$ are all at least 2. Moreover, the probability that E occurs and $deg(w_s) = j_s + 1$ for $s \in \{1,\ldots,l\}$, is

$$\prod_{s=1}^{l}\binom{k-1}{j_s}P_{j_1,\ldots,j_l}(N,M,l)\,.$$

■

Observe now that if $N \geq M+k$ and $M \geq k$, then $((N/M)\binom{k-1}{j}P_j(N,M,1))_{j=0}^{k-1}$ is the probability mass function of a hypergeometric distribution. Hence,

$$\sum_{j=1}^{k-1}\binom{k-1}{j}P_j(N,M,1) \quad = \quad \frac{M}{N}\left(1-\frac{(N-M)_{k-1}}{(N-1)_{k-1}}\right) . \tag{2}$$

Thus, since we want to show that

$$\mathbb{E}\left(\prod_{s=1}^{l} Z_{w_s,a_s}\right) \quad \leq \quad \prod_{s=1}^{l}\mathbb{E}\left(Z_{w_s,a_s}\right) ,$$

it will suffice, because of Claim 1, to establish that provided $N \geq M + lk$ and $M \geq lk$,

$$\sum_{j_1,\ldots,j_l=1}^{k-1}\prod_{s=1}^{l}\binom{k-1}{j_s}P_{j_1,\ldots,j_l}(N,M,l) \quad \leq \quad \frac{M^l}{N^l}\left(1-\frac{(N-M)_{k-1}}{(N-1)_{k-1}}\right)^l . \tag{3}$$

Because of Claim 1 and (2) the inequality clearly holds for $l = 1$. Assuming it holds for $l-1$ we will prove that it also holds for l. First, let $\sigma_{l-1} = \sum_{s=1}^{l-1} j_s$, and observe that

$$P_{j_1,\ldots,j_l}(M,N,l) \quad = \quad P_{j_1,\ldots,j_{l-1}}(M,N,l-1)\cdot P_{j_l}(M-\sigma_{l-1}-l+1, N-(l-1)k, 1) .$$

Applying (2) we get $\displaystyle\sum_{j_l=1}^{k-1}\binom{k-1}{j_l}P_{j_1,\ldots,j_l}(M,N,l) =$

$$P_{j_1,\ldots,j_{l-1}}(M,N,l-1)\cdot\frac{M-\sigma_{l-1}-l+1}{N-(l-1)k}\left(1-\frac{(N-(l-1)k-(M-\sigma_{l-1}-l+1))_{k-1}}{(N-(l-1)k-1)_{k-1}}\right) .$$

Moreover,

$$P_{j_1,\ldots,j_{l-1}}(M,N,l-1)\cdot\frac{M-\sigma_{l-1}-l+1}{N-(l-1)k} \quad = \quad \frac{M}{N}\cdot P_{j_1,\ldots,j_{l-1}}(M-1,N-1,l-1) ,$$

and,

$$P_{j_1,\ldots,j_{l-1}}(M-1,N-1,l-1)\cdot\frac{(N-(l-1)k-(M-\sigma_{l-1}-l+1))_{k-1}}{(N-(l-1)k-1)_{k-1}} \quad = \quad \frac{(N-M)_{k-1}}{(N-1)_{k-1}}\cdot P_{j_1,\ldots,j_{l-1}}(M-1,N-k,l-1) .$$

Hence, if we knew that $P_{j_1,\ldots,j_{l-1}}(M-1,N-k,l-1) \geq P_{j_1,\ldots,j_{l-1}}(M-1,N-1,l-1)$, we would have that

$$\sum_{j_l=1}^{k-1} \binom{k-1}{j_l} P_{j_1,\ldots,j_l}(M,N,l) \leq$$
$$\frac{M}{N} \cdot \left(1 - \frac{(N-M)_{k-1}}{(N-1)_{k-1}}\right) \cdot P_{j_1,\ldots,j_{l-1}}(M-1,N-1,l-1)\,.$$

Thus, by the inductive hypothesis, the LHS of (3) would be at most

$$\frac{M}{N} \cdot \left(1 - \frac{(N-M)_{k-1}}{(N-1)_{k-1}}\right) \cdot \frac{(M-1)^{l-1}}{(N-1)^{l-1}} \cdot \left(1 - \frac{((N-1)-(M-1))_{k-1}}{(N-2)_{k-1}}\right)^{l-1} .$$

This expression is easily seen to be upper bounded by the RHS of (3).

We still need to prove that $P_{j_1,\ldots,j_{l-1}}(M-1,N-k,l-1) \geq P_{j_1,\ldots,j_{l-1}}(M-1,N-1,l-1)$. It will suffice to show that $P_{j_1,\ldots,j_{l-1}}(M',N'-1,l-1) \geq P_{j_1,\ldots,j_{l-1}}(M',N',l-1)$ holds for $M' = M-1$ and $N' \geq N-(k-1)$. This is indeed the case, since

$$\frac{P_{j_1,\ldots,j_{l-1}}(M',N'-1,l-1)}{P_{j_1,\ldots,j_{l-1}}(M',N',l-1)} = \frac{N'}{N'-(l-1)k} \cdot \frac{N'-(l-1)k-(M'-\sigma_{l-1})}{N'-M'} .$$

The claim follows if we can show that $N'(N'-(l-1)k-(M'-\sigma_{l-1})) \geq (N'-M')(N'-(l-1)k)$, or equivalently that $N'\sigma_{l-1} \geq M'(l-1)k$. Since $M' = M-1$, $N' \geq N-k$, and $\sigma_{l-1} \geq l-1$, we only need that $(N-k)(l-1) \geq (M-1)(l-1)k$, i.e., that $N \geq Mk$. But, $N = nk$, $M = mk$, and by hypothesis $n \geq mk$, so we are done. ∎