Towards Optimal Degree-Distributions for Left-Perfect Matchings in Random Bipartite Graphs*

Martin Dietzfelbinger and Michael Rink

Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau
{martin.dietzfelbinger,michael.rink}@tu-ilmenau.de

Abstract. Consider a random bipartite multigraph $G$ with $n$ left nodes and $m \geq n \geq 2$ right nodes. Each left node $x$ has $d_x \geq 1$ random right neighbors. The average left degree $\Delta$ is fixed, $\Delta \geq 2$. We ask whether for the probability that $G$ has a left-perfect matching it is advantageous not to fix $d_x$ for each left node $x$ but rather choose it at random according to some (cleverly chosen) distribution. We show the following, provided that the degrees of the left nodes are independent: If $\bar{\Delta}$ is an integer then it is optimal to use a fixed degree of $\bar{\Delta}$ for all left nodes. If $\bar{\Delta}$ is non-integral then an optimal degree-distribution has the property that each left node $x$ has two possible degrees, $\lfloor \bar{\Delta} \rfloor$ and $\lceil \bar{\Delta} \rceil$, with probability $p_x$ and $1 - p_x$, respectively, where $p_x$ is from the closed interval $[0, 1]$ and the average over all $p_x$ equals $\lfloor \bar{\Delta} \rfloor - \bar{\Delta}$. Furthermore, if $n = c \cdot m$ and $\Delta > 2$ is constant, then each distribution of the left degrees that meets the conditions above determines the same threshold $c^*(\bar{\Delta})$ that has the following property as $n$ goes to infinity: If $c < c^*(\bar{\Delta})$ then there exists a left-perfect matching with high probability. If $c > c^*(\bar{\Delta})$ then there exists no left-perfect matching with high probability. The threshold $c^*(\bar{\Delta})$ is the same as the known threshold for offline $k$-ary cuckoo hashing for integral or non-integral $k = \bar{\Delta}$.

1 Introduction

We study bipartite multigraphs $G$ with left node set $S$ and right node set $T$, where each left node $x$ from $S$ has $D_x$ right neighbors. The right neighbors are chosen at random with replacement from $T$, where the number of choices $D_x$ is a random variable that follows some probability mass function $\rho_x$. Let $|S| = n$ and let $|T| = m$ as well as $1 \leq D_x \leq m$ for all $x$ from $S$. For each $x$ from $S$ let $\Delta_x$ be the mean of $D_x$, that is, $\Delta_x = \sum_{l=1}^{m} l \cdot \rho_x(l)$, and let $\bar{\Delta}$ be the average mean, i.e., $\bar{\Delta} = 1/n \cdot \sum_{x \in S} \Delta_x$. We assume that the random variables $D_x, x \in S$, are independent and $\Delta$ is a given constant.

Our aim is to determine a sequence of probability mass functions $(\rho_x)_{x \in S}$ for the random variables $(D_x)_{x \in S}$ that maximizes the probability that the random

* Research supported by DFG grant DI 412/10-2.
graph \( G = G(\Delta, (\rho_x)_{x \in S}) \) has a matching that covers all left nodes, i.e., a left-perfect matching\(^1\). We call such a sequence optimal. Note that there must be some optimal sequence for compactness reasons.

### 1.1 Motivation and Related Work

Studying irregular bipartite graphs has led to major improvements in the performance of erasure correcting codes. For example in [6] Luby et al. showed how to increase the fraction of message bits that can be recovered for a fixed number of check bits by using carefully chosen degree sequences for both sides of the underlying bipartite graph. The recovery process for erased message bits translates directly into a greedy algorithm for finding a matching in the bipartite graph associated with the recovery process. This was the motivation for the authors of [1,2] to study irregularity in the context of offline \( k \)-ary cuckoo hashing. Here one has a bipartite graph with left nodes corresponding to keys and right nodes corresponding to table cells, where each key randomly chooses table cells without replacement and the aim is essentially to find a left-perfect matching. In [1] it was proven that if the degree of each left node follows some distribution with identical mean and is independent of the other nodes then it is optimal in an asymptotic sense if the degree of each left node is concentrated around its mean. This is in contrast of the following observation in [7] in analogy to [6]: an uneven distribution of the degrees of the left nodes can increase the probability for the existence of a matching that has the advantage that it can be calculated in linear time, by successively assigning left nodes to right nodes of degree one and removing them from the graph.

### 1.2 Results

We will show that for given parameters \( n, m, \) and \( \Delta \) there is an optimal sequence of probability mass functions that concentrates the degree of the left nodes around \( \lfloor \Delta \rfloor \) and \( \lceil \Delta \rceil \). Furthermore, if \( \Delta \) is an integer we can explicitly determine this optimal sequence. In the case that \( \Delta \) is non-integer we will identify a tight condition that an optimal sequence must meet.

**Theorem 1.** Let \( n \leq m \), as well as \( n, \Delta \geq 2 \), and let \( (\rho_x)_{x \in S} \) be an optimal sequence for parameters \( (n, m, \Delta) \). Then the following holds for all \( x \in S \).

(i) If \( \Delta \) is an integer, then \( \rho_x(\Delta) = 1 \).

(ii) If \( \Delta \) is non-integer, then \( \rho_x(\lfloor \Delta \rfloor) \in [0, 1] \) and \( \rho_x(\lceil \Delta \rceil) = 1 - \rho_x(\lfloor \Delta \rfloor) \).

The second statement is not entirely satisfying since it identifies no optimal solution. However, we will give strong evidence that in the situation of Theorem 1(ii) there is no single, simple description of a distribution that is optimal for all feasible node set sizes.

Since the case \( \Delta = 2 \) is completely settled by Theorem 1(i), we focus on the cases where \( \Delta > 2 \), with the additional condition that the number of left nodes

\(^1\) In the following we will use “matching” and “left-perfect matching” synonymously.