Average-case Analysis of Algorithms for Matchings and Related Problems

*Rajeev Motwani**

Department of Computer Science
Stanford University
Stanford, CA 94305-2140

Abstract

We analyze the behavior of augmenting paths in random graphs. Our results show that in almost every graph, any non-maximum 0–1 flow admits a short augmenting path. This enables us to prove that augmenting path algorithms, which are fast in the worst case, also perform exceedingly well on the average. In particular, we show that the $O(\sqrt{|V|}|E|)$ algorithms for bipartite and general matchings run in almost linear time with high probability. It is also shown that the expected running time of the matching algorithms is $O(|E|)$ on input graphs chosen uniformly at random from the set of all graphs. We establish that the permanent of almost every bipartite graph can be approximated in polynomial time. We extend our results to the analysis of the running time of Dinic's algorithm for finding factors of graphs.

*Supported by NSF Grant CCR-9010517, and grants from Mitsubishi and OTL. Part of this work was done while the author was at the Computer Science Division, U. C. Berkeley, and was supported by NSF Grant DCR-8411954. A preliminary version of this paper was published as [Mot89], and is based on the author's PhD dissertation [Mot88].

1. Introduction

Probabilistic analysis of algorithms has traditionally been used to measure the performance of heuristic algorithms under the assumption that the input is drawn from some natural probability distribution [KLMR85]. It has proved to be much more difficult to analyze the average-case performance of algorithms which are designed to perform well on worst-case inputs. We are interested in the latter approach. We consider the problem of finding maximum matchings in bipartite and general graphs. For these problems, we show that the algorithms which are the fastest known on worst-case inputs perform surprisingly well in the average case. In fact, in certain cases these algorithms outperform the best heuristic algorithms devised for these problems. We extend our results to the problem of computing factors of graphs and the permanents of 0-1 matrices. We also show that a parallel algorithm for bipartite matchings experiences a considerable speed-up on the average.

The notion of augmenting paths has played a very significant role in the development of efficient algorithms for a large number of combinatorial problems. Perhaps the most important idea in this area has been that of always using shortest augmenting paths to incrementally construct the solution [Din70, HK73]. We analyze the lengths of augmenting paths in graphs and networks which are drawn from natural probability distributions. It turns out that, with high probability, every non-maximum flow with respect to a random graph permits a short augmenting path. It is exactly this feature of random graphs which leads to a surprisingly good average-case performance of augmenting path algorithms.

Our work was motivated in part by the empirical performance analysis of such algorithms by Derigs and Heske [DH79], Hamacher [Ham78] and Pohl [Poh77]). They observed that the augmenting path algorithms typically did not need to use augmenting paths of length greater than a small constant. For example, Derigs [DH79] found that the matching algorithm due to Even and Kariv [EK75] seems to run in time $O(|E|)$ on most input graphs. This is in contrast with its worst-case running time of $O(|V|^{2.5})$. Our results show that this algorithm, as well as the closely related algorithm of Micali and Vazirani [MV80], has expected running time $O(|E|)$ when the input graph is chosen uniformly at random. In fact, we prove the stronger result that these algorithms run in linear time on **almost every** input graph. Thus, our results provide a theoretical explanation for the empirical observation that such augmenting path algorithms rarely exhibit their worst-case performance, supporting the view that these algorithms should be the algorithm of choice in practice.

Our analysis is based on obtaining a deeper understanding of the behavior of augmenting path algorithms. We establish that the critical idea of augmenting along the shortest possible paths leads to extremely efficient algorithms even in the average case. The basic outline of our technique is as follows. We first demonstrate that sufficiently dense random graphs almost surely satisfy certain structural properties. The most important such property, the "spreading" property (see [Bol85], p. 229), is the existence of a large set of vertices which induces a subgraph with an extremely good expansion property. Next, we switch to a deterministic viewpoint and consider any fixed graph which satisfies the structural properties. We show that in such a graph a collection of extremely short paths covers a large number of vertices. This enables us to prove that there exist short augmenting paths with respect to any non-maximum 0-1 flow on this graph.

There is an intriguing connection between our proof technique and Hall's Theorem [Hal35] for bipartite matchings. It states that a bipartite graph has a perfect matching if and only if each set of k vertices has at least k neighbors. In other words, every non-perfect matching in the graph has an augmenting path if every set of vertices spreads or expands by a factor of 1. The key observation underlying our analysis is that every non-perfect matching has an augmenting path of length $O(\log n/\log\Delta)$ if all sets of at most n/Δ vertices expand by a factor of Δ. It turns out that in most random graphs almost all sets of vertices exhibit the spreading property, where Δ is close to the average degree in the graph.

The spreading property has previously been used to determine the diameter of random graphs [Bol81]. Therefore, it is not very surprising that our results show that, with high probability, a shortest augmenting path in a random graph is of length at most a constant factor greater than its diameter. We would also like to point out that similar structural properties have been used before as a tool for analyzing random graphs, most notably in the analysis of heuristic algorithms for Hamiltonian paths. Pósa [Pos76] was the first to use the spreading property in the analysis of Hamiltonian paths. The structural properties of random graphs that we will be using bear the closest resemblance to those invoked by Komlós and Szemerédi [KS83], Bollobás and Frieze [BF85], and Bollobás, Fenner and Frieze [BFF85]. Note that this is the first time such tools have been applied to the problem of matchings, and, more significantly, to the average-case analysis of non-heuristic algorithms.

The rest of this paper is organized as follows. In Section 1.1 we define the models of random graphs used in our analysis; the main results and their relation to previous work are described in Section 1.2. In Section 2 we develop some technical tools needed for dealing with matchings in graphs. It is shown in Section 3 that the bipartite matching algorithm performs very efficiently on expanders, and this provides an outline of the average case analysis which is presented in Section 4. The results are extended to a model of sparse random graphs in Section 5. Finally, in Section 6 we extend our results to non-bipartite graphs.

1.1. Random Graphs Models

We start by describing the probability distributions that will be used in our analysis. Several natural probability spaces for graphs have been defined in the theory of random graphs [ER59, Bol85]. We will draw upon the tools developed therein. In what follows, n will denote the number of vertices and m will denote the number of edges of the input graph. Also, for bipartite graphs $N = n^2$, while for general graphs $N = n(n-1)/2$, will denote the total number of edges possible.

A commonly used model of random graphs is called $\mathcal{G}_{n,p}$, where n is a positive integer and $0 \le p(n) \le 1$. The model consists of all graphs on the vertex set $V = \{1, \ldots, n\}$. In a graph drawn from $\mathcal{G}_{n,p}$, each edge (independently) is present with probability $p(n)$. In other words, a graph G with e edges has probability $p^e(1-p)^{N-e}$. A related model, $\mathcal{G}_{n,m}$, consists of all graphs on the vertex set $V = \{1, \ldots, n\}$ which have exactly m edges. Each graph on n vertices and with m edges has equal probability. Similarly, we define the models $\mathcal{B}_{n,p}$ and $\mathcal{B}_{n,m}$ which consist of bipartite graphs on the vertex sets $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$. Another model of interest is called the d-out model. The model $\mathcal{B}_{d-out}$ also consists of bipartite graphs on the vertex sets $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$. A random graph from $\mathcal{B}_{d-out}$ is selected as follows. Each vertex independently chooses exactly d neighbors in a uniform fashion. Multiple edges introduced by two vertices choosing each other as neighbors are replaced by a single edge. The probability space $\mathcal{G}_{d-out}$ is defined analogously.

It turns out that the two models $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,m}$ (or, $\mathcal{B}_{n,p}$ and $\mathcal{B}_{n,m}$) are interchangeable in a very strong sense. More precisely, consider the case where $m = Np(n)$; this is the expected number of edges in a graph drawn from $\mathcal{G}_{n,p}$. The following theorem due to Angluin and Valiant [AV79, Bol85] allows us to transfer results from one model to the other.

Theorem 1 (Angluin & Valiant): *Let $m = Np(n)$ and $\mathcal{P}$ be some monotone graph property. If $G_p \in \mathcal{G}_{n,p}$ and $G_m \in \mathcal{G}_{n,m}$ then*

$$\mathrm{Prob}[G_m \in \mathcal{P}] = O(n \cdot \mathrm{Prob}[G_p \in \mathcal{P}])$$

We will derive our results only for the $\mathcal{G}_{n,p}$ model and they will all apply to the appropriate $\mathcal{G}_{n,m}$ model too; similarly, all our results for the $\mathcal{B}_{n,p}$ model will also apply to the $\mathcal{B}_{n,m}$ model.

1.2. Main Results and Previous Work

The algorithm of Hopcroft and Karp [HK73] computes a maximum matching in a bipartite graph in $O(\sqrt{n}m)$ time. Feder and Motwani [FM91] improved the running time of this algorithm to $O(\sqrt{n}m/\log n)$, and this is currently the most efficient algorithm known for this problem. The Hopcroft-Karp algorithm is essentially the same as Dinic's flow algorithm [Din70] when applied to a special case of the 0-1 network flow problem generated by an instance of a bipartite matching problem. For random bipartite graphs, an early result due to Erdös and Rényi [ER64, ER68] gave a sharp threshold for the existence of perfect matchings.

Theorem 2 (Erdös & Rényi): *Let $p(n) = \frac{\ln n + c}{n}$, where c is a constant. Then for $B \in \mathcal{B}_{n,p}$,*

$$\lim_{n\to\infty} \mathrm{Prob}[B \text{ has a perfect matching}] = e^{-2e^{-c}}$$

A similar result holds for the probability space $\mathcal{B}_{n,m}$. Angluin and Valiant [AV79] gave a fast heuristic algorithm (the Proposal Algorithm) which works in $O(n \log n)$ time and with high probability computes a perfect matching in random graphs whose density exceeds the sharp threshold by a specific constant factor. Goldschmidt and Hochbaum [GH90] gave a greedy algorithm which improves upon the running time of the Proposal Algorithm. The greedy algorithm works for random graphs where $p > \gamma \ln n/n$ and, with high probability, computes a perfect matching in time $O(n \log(1/p))$. The only heuristic known for matchings in sparse random graphs is due to Karp and Sipser [KS81]. Their algorithm computes a near-maximum matching in random graphs with $p = c/n$, for any positive constant $c > 0$. For the more general problem of flows in networks, several heuristic algorithms have been proposed. For example, Karp, Motwani and Nisan [KMN92] have provided linear time heuristic algorithms which can compute the maximum flow with high probability on random inputs.

We show that Dinic's algorithm terminates in $O(m \log n/\log\log n)$ time with high probability when the value of $p(n)$ is at or above the threshold for the existence of perfect matchings in random bipartite graphs. This implies that Dinic's algorithm runs in almost linear time on almost every graph which has at least $n \ln n$ edges. This result improves on earlier heuristic analysis in two ways: first, the algorithm we analyze is not a heuristic algorithm but it is, in fact, the fastest algorithm known from the worst case running time point of view; secondly, this algorithm efficiently finds a perfect matching in a random bipartite graph exactly at the threshold density for the existence of these matchings, whereas the Proposal Algorithm requires the graph density to be a constant factor above the threshold. In fact, unlike the Proposal Algorithm, Dinic's algorithm will always find a perfect matching, if one exists, in any input graph. Our analysis is just concerned with the amount of time this algorithm takes in finding a maximum matching in a random graph. An important application of our result is to the case where the input graph is chosen uniformly at random from the set of all graphs on a fixed vertex set. In this case, we show that Dinic's algorithm has an expected running time of $O(m)$.

A parallel implementation of the Hopcroft-Karp algorithm was devised by Gabow and Tarjan [GT88]. Their algorithm runs in $O(\sqrt{n}m \log P/P)$ time with P processors. This algorithm has almost linear speed-up for a reasonably large number of processors. We show that the Gabow-Tarjan algorithm terminates in $O((m \log n \log P)/(P \log\log n))$ time, with high probability, when the input is a random bipartite graph.

For the model of d-out graphs, Walkup [Wal80] showed that a perfect matching almost surely exists in such graphs when $d \geq 2$, but not when $d = 1$. Karp and Rinnooy Kan [KK86] gave an algorithm which with high probability finds a perfect matching in graphs drawn from $\mathcal{B}_{d-out}$, $d \geq 2$, in $O(n \log n)$ time. We show that Dinic's algorithm terminates in $O(m \log n/\log d)$ time on such graphs, with high probability.

The situation for random non-bipartite graphs is more or less the same. Again, Erdös and Rényi [ER66] showed that the sharp threshold for the existence of perfect matchings in graphs drawn from $\mathcal{G}_{n,p}$ is centered about $p(n) = \ln n/(n-1)$. The results of Angluin and Valiant also apply to such graphs. The fastest algorithm known in terms of the worst case running time, due to Micali and Vazirani [MV80], runs in $O(\sqrt{n}m)$ time. We show that with high probability the latter algorithm terminates in $O(m \log n/\log\log n)$ time on input drawn from random graphs with density at least at the sharp threshold. If the input is chosen uniformly at random from the set of all graphs on a fixed vertex set, then this algorithm has expected running time $O(m)$.

A k-factor of a graph G is a spanning subgraph of G in which every vertex has degree k. Clearly, a perfect matching is exactly the same as a 1-factor. For fixed k, k-factors of a bipartite graph can be found via Dinic's algorithm in the same time as for matchings. It turns out that for bipartite graphs the k-factor problem can be reduced to that of finding perfect matchings [LP86]. In the case of random graphs, Shamir and Upfal [SU81] showed that the threshold for existence of k-factors is centered about $p = (\ln n + (k-1)\ln\ln n)/n$, which is also the threshold for having minimum degree k. From the algorithmic point of view, the Extension-Rotation algorithm due to Motwani [Mot92] in $O(n \log n)$ time computes a k-factor in a random graph of density a small constant factor above the threshold density. Our result is that in random bipartite graphs, Dinic's algorithm for k-factors terminates in almost linear time when $p > \ln n/(n-1)$. Our techniques extend to the k-factor problem for general graphs.

The problem of computing the permanent of an $n \times n$ matrix M with only 0-1 entries has received considerable attention of late. This problem is equivalent to that of counting the number of perfect matchings in a bipartite graph. The intractability of this problem was demonstrated by Valiant [Val79] who showed that it is complete for the class of counting problems called #P. The importance of this problem is mainly due to this result, which establishes that it is as hard as any problem that involves counting structures which are computable in NP. It turns out that even approximating the permanent is a hard problem. Recent work due to Broder [Bro86] and Jerrum and Sinclair [JS88] has led to a randomized polynomial-time approximation scheme for the permanents of extremely dense (minimum degree $n/2$) bipartite graphs. For the case of random bipartite graphs, Jerrum and Sinclair show that for graphs drawn from $\mathcal{B}_{n,p}$ (where $p(n) \geq 360 \ln n/n$) it is possible to obtain, with high probability, an approximation of the permanent.

We improve the last result to essentially the best possible. Using our results on augmenting path lengths, we show that the scheme of Jerrum and Sinclair completely resolves the problem of approximating the permanents of random bipartite graphs for any value of $p(n)$. Our results extend to the model of d-out bipartite graphs. An important implication of our results is that there is a randomized polynomial-time approximation scheme for the permanent which works for *almost every* bipartite graph.

2. Alternating Level Graphs Paths in Bipartite Graphs

In this section we review the notion of alternating and augmenting paths, and also define a generic structure called an alternating level graph. Consider a bipartite graph $B(U, V, E)$ with the vertex sets $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$. We will refer to the vertices in U as the *boy* vertices and to those in V as the *girl* vertices. Let $M \subseteq E$ be some non-maximum matching in B with $|M| = r$. An edge from M will be called a *matching* edge, while an edge which does not lie in M will be called *free* edge. Similarly, an unmatched vertex in B will be referred to as a *free* vertex.

An *alternating path* P with respect to M is a path whose edges are alternately in M and not in M. An *augmenting path* is an alternating path between two free vertices. Given an augmenting path P, we augment M by adding to it the edges in $P \setminus M$ and removing from it the edges in $P \cap M$. We then obtain a new matching M' of cardinality $r + 1$.

Fact 1: *Let M be a matching in a bipartite graph B. If M is not a maximum matching, then B contains an augmenting path for M.*

Given a non-maximum matching M, an *alternating level graph* LG with l levels is any subgraph of B which satisfies the following conditions. A level graph LG starting with free boy vertices consists of $l + 1$ disjoint sets of vertices, $V_0, V_1, \ldots, V_l$, where the set V_0 can contain only the free boy vertices. The odd-numbered vertex sets contain girl vertices from V, while the even-numbered vertex sets contain boy vertices from U. Each vertex in a vertex set V_i can only be adjacent to vertices in V_{i-1} and V_{i+1}; moreover, each vertex in V_i must be adjacent to at least one vertex in V_{i-1}. For even i, all edges between V_i and V_{i+1} must be free edges; while for odd i, every edge between V_i and V_{i+1} must be a matching edge. Alternating level graphs which start at free girl vertices are defined similarly. For $0 \leq i \leq l$, we define

$$V^{(i)} = V_0 \cup V_1 \ldots \cup V_i$$

While there is an alternating path to each vertex in LG from some free boy vertex, not every alternating path need be represented in an alternating level graph. The following important property of alternating level graphs underlies their usefulness to us. A similar fact holds for the case where LG is an alternating level graph which starts at free girl vertices.

Fact 2: *Suppose there is a free girl vertex u in the i^{th} level of LG. Then LG contains an augmenting path of length i for the matching M.*

In our analysis, we will specify two level graphs for a non-maximum matching M – LG_1 starting at free boy vertices, and LG_2 starting free girl vertices. The level graphs LG_1 and LG_2 are mere artifacts of our proof and not the ones actually constructed by the algorithm under consideration. (Recall that Dinic's algorithm constructs only a single level graph, in particular what we call a *maximal level graph*.) We present two crucial properties of any such level graphs LG_1 and LG_2.

Lemma 1: *If LG_1 and LG_2 have a common vertex within their first L levels, then there is an augmenting path for M which is of length at most $2L$.*

Proof: Consider the lowest-numbered level i in LG_1 which contains a vertex also present in the first L levels of LG_2. Let u be any such common vertex in the i^{th} level of LG_1, and let j be the level in LG_2 containing u; clearly, $i, j \leq L$. Assume, without loss of generality, that u is a boy vertex.

From the definition of an alternating level graph it follows that there is an alternating path in LG_1 which goes from a free boy vertex to u. This path, say P_1, is of length $i \leq L$ and does not contain any vertices, besides u itself, which also lie in the first L levels of LG_2. Similarly, there is an alternating path in LG_2 which goes from a free girl vertex to u. This path, say P_2, is of length $j \leq L$ and is vertex-disjoint with P_1 besides the common vertex u. It follows that the concatenation of P_1 and P_2 is an augmenting path for M which is of length $i + j \leq 2L$. □

Lemma 2: *Suppose there is a free edge in B joining a boy vertex in the first L levels of LG_1 to a girl vertex in the first L levels of LG_2. Then there is an augmenting path in B for M which is of length at most $2L+1$.*

Proof: Let $u \in U$ and $v \in V$ be in the first L levels of LG_1 and LG_2, respectively. Assume that there is a free edge in B between u and v. By Lemma 1 we know that if there is a vertex which lies in the first L levels of both LG_1 and LG_2 then we already have an augmenting path for M which is of length at most $2L$. Therefore, we may assume that there are no such common vertices.

There must be an alternating path in LG_1, say P_1, which goes from a free boy vertex to u. Similarly, there must be an alternating path in LG_2 which goes from a free girl vertex to v. Both these paths must be of length at most L since both u and v lie in the first L levels of their respective level graphs. Moreover, the two paths must be vertex-disjoint as LG_1 and LG_2 do not have any common vertices in their first L levels. It follows that the concatenation of P_1, the edge $\{u, v\}$ and P_2 forms an augmenting path for M which is of length at most $2L + 1$. □

Our definition of an alternating level graph is much more general than that used by Hopcroft and Karp in their algorithm. In fact, our definition is non-constructive and does not uniquely specify a level graph with respect to any fixed matching. As will become clear in later sections, this flexibility is very useful in our analysis. For the sake of completeness, we describe below the particular construction of a level graph used in the standard matching algorithms.

The Hopcroft-Karp algorithm constructs what we call a *maximal level graph* with respect to a non-maximum matching M. The starting level V_0 contains *all* the free boy vertices. For even i, $V_{i+1} \subseteq V$ contains all the neighbors* in B of the vertices contained in $V_i \subseteq U$ which do not already lie in $V^{(i)}$, i.e. $V_{i+1} = \Gamma_B(V_i) \setminus V^{(i)}$. All edges between V_i and V_{i+1} that are present in B are also present in LG. Note that none of these edges will be matching edges since every boy vertex is matched to some girl vertex in the *previous* level. In fact, except in V_0, a boy vertex is introduced into the level graph only if it is matched to a girl vertex already present in the level graph. For odd i, V_{i+1} is the set of vertices which are matched in M to the vertices in V_i, i.e. $V_{i+1} = \Gamma_M(V_i)$. All the matching edges between the two vertex sets are also introduced into LG. Observe that $\Gamma_M(V_i)$ must be disjoint from $V^{(i)}$. The construction terminates at the first level where we encounter free girl vertices. It follows from Fact 2 that if we encounter a free girl vertex in the set V_l then we have an augmenting path for M which is of length l.

Dinic's algorithm computes a maximum matching in bipartite graphs as follows. It builds up the maximum matching progressively in stages. In each stage, the current non-maximum matching is augmented by a maximal set of vertex-disjoint shortest augmenting paths. Each stage starts off by constructing an alternating level graph for the current non-maximum matching. The maximal set of disjoint shortest augmenting paths can then be found in $O(m)$ time by doing a depth-first traversal of the alternating level graph. The stage is completed by using these augmenting paths to construct a new matching which has a higher cardinality. An ingenious analysis due to Hopcroft and Karp [HK73] shows that even though the length of the shortest augmenting paths may reach n, the number of stages of augmentation is no more than $\sqrt{n}$. Since each stage of augmentation takes $O(m)$ time, this algorithm terminates in $O(\sqrt{n}m)$ time on all inputs.

*For any set of vertices S, we will denote its neighbors in a graph G by $\Gamma_G(S)$.

3. Bipartite Matchings and the Expansion Property

In this section we analyze the running time of Dinic's algorithm for bipartite matchings when the input graph is an *expander*. Although the analysis is fairly straightforward, we include it to explicate the connection between Hall's Theorem and our technique for analyzing Dinic's algorithm. Moreover, it also serves as an outline of the analysis for the probabilistic case. In later sections we will show that most random graphs satisfy some sort of an expansion property. We will then adapt the analysis performed in this section to extend to that class of graphs. It is important to keep in mind that the following discussion does not involve any probabilistic assumptions whatsoever. The analysis can also be extended to the case of non-bipartite graphs and to the other problems being considered.

We start by presenting the characterization [Hal35] of bipartite graphs which contain perfect matchings.

Theorem 3 (Hall): *A bipartite graph $B(U,V,E)$ has a perfect matching if and only if for each $X \subseteq U$ and $Y \subseteq V$, $|\Gamma_B(X)| \geq |X|$ and $|\Gamma_B(Y)| \geq |Y|$.*

For any $\Delta \geq 1$, a Δ-expander is defined as follows.

Definition 1: *A bipartite graph $B(U,V,E)$ is called a Δ-expander if for each $X \subseteq U$ and $Y \subseteq V$, where $|X|$, $|Y| \leq n/\Delta$, we have that $|\Gamma_B(X)| \geq \Delta|X|$ and $|\Gamma_B(Y)| \geq \Delta|Y|$. The set of all bipartite Δ-expanders on vertex sets U and V is denoted by $\mathcal{B}(\Delta)$.*

Fact 3: *Let $B \in \mathcal{B}(\Delta)$ and let $X \subseteq U$ or $Y \subseteq V$ be such that $|X|$, $|Y| > n/\Delta$. Then $|\Gamma_B(X)| > n - \Delta$ and $|\Gamma_B(Y)| > n - \Delta$.*

Proof: Let r be a positive integer such that $r = \lfloor n/\Delta \rfloor$. We will prove this fact only for some $X \subseteq U$. The proof for $Y \subseteq V$ will be similar. Let $X' \subseteq X$ be such that $|X'| = r \leq n/\Delta$. Since $B \in \mathcal{B}(\Delta)$ it follows that $|\Gamma_B(X')| \geq \Delta\lfloor n/\Delta \rfloor > n - \Delta$. Since $\Gamma_B(X') \subseteq \Gamma_B(X)$, the desired result follows. □

It is not very hard to see that Hall's Theorem can now be restated as follows.

Theorem 4 (Hall): *The set of all bipartite graphs on vertex sets U and V which contain a perfect matching is exactly $\mathcal{B}(1)$.*

In a sense, the above theorem states that any non-perfect matching in $B \in \mathcal{B}(1)$ has an augmenting path in B. We now extend this result to show that any non-perfect matching in $B \in \mathcal{B}(\Delta)$, where $\Delta > 1$, has an augmenting path of length $O(\log n / \log \Delta)$.

Theorem 5: *Let $B \in \mathcal{B}(\Delta)$, where $\Delta > 1$, and let M be any non-perfect matching in B. Then there exists an augmenting path in B for M which is of length at most $2L + 1$, where $L = 2\lceil \ln n / \ln \Delta \rceil$.*

Proof: Since $|M| < n$ it is clear that there is at least one free boy vertex and one free girl vertex in B. Suppose we construct two level graphs LG_1 and LG_2 using the construction of Dinic's algorithm, as described in the previous section. The level graph LG_1 starts at the free boy vertices and LG_2 starts at free girl vertices. We assume that the construction is carried out only up to the first L levels in each level graph.

If LG_1 contains any free girl vertices or LG_2 contains any free boy vertices then, by Fact 2, there exists an augmenting path for M of length at most L and we are done. Therefore, we only consider the case where LG_1 does not contain any free girl vertices and LG_2 does not contain any free boy vertices. Also, from Lemma 1, we have that if the first L levels of LG_1 and LG_2 have a vertex in common then there exists an augmenting path for M of length at most $2L$ and again we are done. Therefore, we further assume that LG_1 and LG_2 are disjoint.

We now restrict our attention to the level graph LG_1. Let $l = L/2$ and, for $0 \leq i \leq l$, let $X_i = V_{2i}$ be the i^{th} level of boy vertices in LG_1. We will use $X^{(i)}$ to denote the set $X_0 \cup X_1 \ldots \cup X_i$. Since there is at least one free boy vertex in B, it follows that the set $X^{(0)}$ is non-empty. We now show that each subsequent $X^{(i)}$ must rapidly expand in size.

Let $0 \leq i \leq l$ be such that $|X^{(i)}| \leq n/\Delta$. In Dinic's algorithm, the set $X_{i+1} = V_{2i+2}$ is obtained from $X_i = V_{2i}$ as follows. The set V_{2i+1} contains all those girl vertices which do not lie in $V^{(2i)}$ and are adjacent to some boy vertex in V_{2i}, i.e. $V_{2i+1} = \Gamma_B(V_{2i}) \setminus V^{2i}$. Note that each boy vertex in V_{2i} is matched to a girl in V_{2i-1}. This implies that all edges between V_{2i} and V_{2i+1} are free edges. The set X_{i+1} is exactly the set of boy vertices which are matched to the girl vertices in V_{2i+1}. Since LG_1 does not contain any free girl vertices, we have that $|X_i| = |V_{2i+1}|$.

Since $|X^{(i)}| \leq n/\Delta$, it follows that $\Gamma_B(X^{(i)}) \geq \Delta|X^{(i)}|$. Note that $\Gamma_B(X_i) \setminus V^{(2i)} = \Gamma_B(X^{(i)}) \setminus V^{(2i)}$, since if any girl vertex is adjacent to a boy vertex in $X^{(i-1)} = X^{(i)} \setminus X_i$ then it would already be present in $V^{(2i)}$. We already know that $\Gamma_B(X^{(i)}) \geq \Delta|X^{(i)}|$. Since every girl vertex in $V^{(2i)}$ must be matched to a unique boy vertex in $V^{(2i)}$, it must be the case that $V^{(2i)}$ has at most $|X^{(i)}|$ girl vertices. In fact, X_0 contains at least one free boy vertex and so $V^{(2i)}$ can contain at most $X^{(i)} - 1$ girl vertices. It then follows that $|\Gamma_B(X^{(i)}) \setminus V^{(2i)}| \geq |\Gamma_B(X^{(i)})| - (|X^{(i)}| - 1)$. This is summarized in the following claim.

Claim 1: *Suppose that $|X^{(i)}| \leq n/\Delta$. Then it must be the case that $|X_{i+1}| \geq (\Delta - 1)|X^{(i)}| + 1$ and $|X^{(i+1)}| \geq \Delta|X^{(i)}| + 1$.*

Since $|X^{(0)}| > 0$, and each subsequent $X^{(i)}$ must expand in size at least by a factor of Δ, it follows that for some $j \leq \lceil \ln n / \ln \Delta \rceil$ the set $X^{(j)}$ has more than n/Δ elements. Consider now the set $X^{(j+1)}$. By Fact 3 we have that $\Gamma_B(X^{(j)}) > n - \Delta$. By repeating the expansion argument, it can then be seen that $|X^{(j+1)}| > n - \Delta$.

Let $X = X^{(L)}$ denote that set of all boy vertices in LG_1. Since $X^{(j+1)} \subseteq X$, it must be the case that $|X| > n - \Delta$. Similarly, it can be shown that the set Y of all girl vertices in LG_2 contains at least $n - \Delta$ elements. We claim that there cannot be any matching edges between the vertex sets X and Y. If that were the case then both end-points of such a matching edge would lie in both LG_1 and LG_2. This would contradict the assumption that LG_1 and LG_2 do not have any common vertices. From Lemma 2 it follows that if there is some free edge between the vertex set X and Y, then there exists an augmenting path for M of length at most $2L + 1$ and we are done.

The only case left to consider is when there is *no* edge in B between the vertex sets X and Y. We complete the proof of the theorem by showing that this is impossible. Assume, for the moment, that $\Delta \leq n/2$; this implies that $n - \Delta \geq n/2$. If there is no edge from X to Y, then it must be the case that $\Gamma_B(X) \subseteq V - Y$. Since $|Y| > n - \Delta \geq n/2$, it follows that $|\Gamma_B(X)| < n/2$. But, Fact 3 requires that $|\Gamma_B(X)| > n - \Delta \geq n/2$ and this gives us a contradiction. In the case where $\Delta > n/2$ it is easy to see that every set of boy vertices must have more than $n/2$ girls as neighbors. The preceding argument again gives us a contradiction. □

An application of Theorem 5 to Dinic's algorithm yields the following corollary.

Corollary 1: *Let $B \in \mathcal{B}(\Delta)$. Then Dinic's algorithm for bipartite matchings terminates in $O(m \log n / \log \Delta)$ steps on input B.*

Proof: It is known [HK73] that the lengths of the shortest augmenting paths must increase at every stage of Dinic's algorithm. A trivial bound on the number of stages of augmentation performed by this algorithm is given by the maximum length of an augmenting path used by it. We have from Theorem 5 that, for any $B \in \mathcal{B}(\Delta)$, every non-perfect matching in B has an augmenting path of length $O(\log n / \log \Delta)$. Since every stage of augmentation requires $O(m)$ steps, the desired bound on the running time follows. □

The above theorem works for any $\Delta > 1$. Consider the case where $\Delta = 1$. It is easy to see that Claim 1 still applies. Claim 1 states that when $\Delta = 1$ then, for any i, $|X^{(i+1)}| \geq |X^{(i)}| + 1$. In other words, every successive level of boy vertices contains at least one element unless we already have an augmenting path for M. This means that in each $B \in \mathcal{B}(1)$ any non-perfect matching must have an augmenting path. Thus, we have proved that each $B \in \mathcal{B}(1)$ contains a perfect matching. This gives a constructive proof of Hall's Theorem. Of course, constructive proofs of Hall's Theorem have been provided earlier. We emphasize this aspect of our result only to provide some intuition about our proof technique.

4. Random Bipartite Graphs

Here we prove some structural properties of random bipartite graphs that will be used in our analysis of the various augmenting path algorithms. Let $B \in \mathcal{B}_{n,p}$ where $p > \ln n/n$. Let Δ be the expected degree of a vertex in B, i.e. $\Delta = pn$. Let Δ_m denote the maximum degree in B. Further, define a partition of the two vertex sets as follows. Let $S(U) = \{u \in U : \text{degree}(u) < \Delta_s\}$ and $L(U) = U \setminus S(U)$, where $\Delta_s = \Delta/\mu$ for some constant $\mu > 1$. The set $S(U)$ contains boy vertices with *small* degrees and $L(U)$ contains the *large* degree boy vertices; $S(V)$ and $L(V)$ are defined similarly.

Lemma 3: *Let $B \in \mathcal{B}_{n,p}$, where $p > \ln n/n$. Then, with high probability, the following properties hold for B with a suitable choice of the positive constants μ, σ and λ.*

(a) *$\Delta_m \leq \sigma\Delta$, the maximum degree in B is at most a constant factor away from the average degree.*

(b) *If $W \subseteq L(U)$ and $|W| \leq n/\Delta$ then $|\Gamma_B(W)| \geq \Delta|W|/\lambda$.*

(c) *For each $u \in U$ there is at most one $v \in S(U)$ such that u is at distance 2 from v in B. Moreover, if $v \in S(U)$ is at distance 2 from u in B then there is exactly one path of length 2 between u and v.*

(d) *Let $X \subseteq U$ and $Y \subseteq V$ be such that $|X|, |Y| \geq n/\beta$, for any constant β. Then it cannot be the case that there is no edge between X and Y in B.*

Proof: (a). Let $u \in U$ be some fixed vertex in B. Let Ψ be a random variable which denotes the number of subsets S of V such that $|S| = \sigma\Delta$ and $S \subseteq \Gamma_B(u)$. Since Ψ is a non-negative integer-valued random variable, we can make use of Markov's Inequality.

$$\text{Prob}[\Psi > 0] \leq E[\Psi] = \binom{n}{\sigma\Delta} p^{\sigma\Delta}$$

Clearly, Ψ is positive if and only if the degree of u is at least $\sigma\Delta$. Summing over all choices of u we obtain the following.

$$\text{Prob}[\Delta_m \geq \sigma\Delta] \leq n\binom{n}{\sigma\Delta} p^{\sigma\Delta} \leq n\left(\frac{e}{\sigma}\right)^{\sigma\Delta}$$

Since $\Delta \geq \ln n$, for every $\gamma > 1$ there is an $\sigma > 1$ such that the following holds.

$$\text{Prob}[\Delta_m \geq \sigma\Delta] \leq n^{-\gamma}$$

$\square$

(b). Call (X, Y), where $X \subseteq L(U)$ and $Y \subseteq V$, a *k-bottleneck* pair for B if $|X| = k < n/\Delta$, $|Y| = \Delta k/\lambda$ and $\Gamma_B(X) \subseteq Y$. It is clear that the statement of the lemma does not hold for B if and only if there exists a k-bottleneck pair in B, for $1 \leq k \leq n/\Delta$. Let Ψ be a random variable which denotes the number of k-bottleneck pairs in B.

Let (X, Y) be some fixed pair with $X \subseteq L(U)$ and $Y \subseteq V$ such that $|X| = k < n/\Delta$ and $|Y| = \Delta k/\lambda$. Since $X \subseteq L(U)$, we have that for each $x \in X$ there are at least Δ_s neighbors in V. Suppose we construct a new graph B' by including exactly Δ_s incident edges for each vertex in $L(U)$. The Δ_s edges incident on each such vertex are chosen uniformly at random from all the edges incident on that vertex. Clearly, the probability of the pair (X, Y) being a k-bottleneck pair for B is stochastically dominated by the probability of its being a k-bottleneck pair for B'. We first give an upper bound on this probability.

For any vertex $v \in L(U)$, each subset of Δ_s girl vertices is equally likely to be picked as v's neighbors in B'. Moreover, the choice of neighbors in B' is made independently for each vertex in $L(U)$. Thus, in B', the probability that (X, Y) is a k-bottleneck pair is exactly the probability that each vertex in X has all its neighbors in Y when choosing Δ_s neighbors independently and uniformly at random from Y.. This gives us the following inequality.

$$\text{Prob}[(X, Y) \text{ is } k\text{-bottleneck in } B] \leq \left(\frac{\binom{\Delta k/\lambda}{\Delta_s}}{\binom{n}{\Delta_s}}\right)^k$$

Upon summing over all possible choices of X and Y, we obtain

$$\text{Prob}[\Psi > 0] \leq \binom{n}{k}\binom{n}{\Delta k/\lambda}\left(\frac{\binom{\Delta k/\lambda}{\Delta_s}}{\binom{n}{\Delta_s}}\right)^k \leq \binom{n}{k} e^{\Delta k/\lambda}\left(\frac{\Delta k}{n\lambda}\right)^{\Delta k(1/\mu - 1/\lambda)}$$

Upon suitable algebraic manipulation, using $k < n/\Delta$ and $\Delta > \ln n$, we obtain the following upper bound. For any $\gamma > 1$, there is a choice of $\lambda > \mu > 1$ such that,

$$\text{Prob}[\Psi > 0] \leq n^{-(\gamma+1)}$$

It is easy to see that the probability that Ψ is non-zero is exactly the probability that there is a set of k vertices in $L(U)$ which do not have a sufficiently large number of neighbors in V. Summing over all values of $k < n/\Delta$ we obtain the desired result. □

(c). Let P_S denote the probability that some fixed vertex $v \in U$ is in the set $S(U)$ for B. Let E denote the event that there is a vertex $v \in U$ such that there are two distinct vertices $x, y \in S(U)$ at distance 2 from v. There are exactly two ways in which this can happen. First, there could be a path of length 4 from x to y which contains v as the middle vertex. Secondly, there could be a girl vertex $z \in V$ which is adjacent to each of v, x and y. In the first case, there are at most n^5 ways of choosing the five vertices involved in the path and it is required that four specific edges be present in B. In the second case, there are at most n^4 ways of choosing the four vertices involved and it is required that three specific edges be present in B. It is also required, in both cases, that both x and y lie in $S(U)$. The probability both x and y lie in $S(U)$ is exactly P_S and the two events are independent. This gives us the following upper bound on the expected number of structures which can cause the event E to occur. By Markov's Inequality, the probability of E cannot be more than this quantity.

$$\text{Prob}[E] \leq (n^5p^4 + n^4p^3)(P_S)^2 \leq 2n\Delta^4(P_S)^2$$

Now, P_S is the probability that in n Bernoulli trials the number of successes is no more than Δ_s, where each trial succeeds with probability p. An application of the bound on the tails of binomial distributions due to Chernoff [Che52] (see also [Bol85], p. 11) yields the following bound on P_S, where $\epsilon = (1 + \ln\mu)/\mu$.

$$\text{P}_S \leq \left(\frac{np}{\Delta_s}\right)^{\Delta_s}\left(\frac{n(1-p)}{n-\Delta_s}\right)^{n-\Delta_s} \leq e^{-(1-\epsilon)\Delta}$$

Assuming $a > 16$ and using $\Delta > \ln n$, the following inequality establishes the first part of the lemma.

$$\text{Prob}[E] \leq n\Delta^4 e^{-2\Delta(1-\epsilon)} = O(n^{-1/2})$$

We now consider the second part of the lemma. Let E' denote the event that there are two paths of length 2 in B from $u \in U$ to $x \in S(U)$. This event occurs only if there is a cycle of length 4 which contains both u and x. We may bound the probability of E by the expected number of such structures in B. This gives us the following inequality which completes the proof.

$$\text{Prob}[E] \leq n^4p^4P_S \leq \Delta^4 e^{-(1-\epsilon)\Delta} = O(n^{-1/2})$$

□

(d). Let Ψ now be a random variable which denotes the number of pairs (X, Y) such that $X \subseteq U$, $Y \subseteq V$, $|X| = |Y| = n/\beta$ and $Y \cap \Gamma_B(X) = \emptyset$. Using Markov's Inequality, we obtain

$$\text{Prob}[\Psi > 0] \leq \binom{n}{n/\beta}\binom{n}{n/\beta}(1-p)^{n^2/\beta^2}$$

Since $\Delta > \ln n$, we have that for sufficiently large n and any fixed $\gamma > 0$,

$$\text{Prob}[\Psi > 0] \leq e^{2n(\ln\beta+1)/\beta} e^{-n\Delta/\beta^2} = O(n^{-\gamma}).$$

□

By symmetry, this lemma also holds for the girl vertices, i.e. when U is interchanged with V throughout. The most crucial structural property is that of Lemma 3(b), the *expansion* property. We wish to restrict our attention to those random bipartite graphs which satisfy Lemma 3. Let $\mathcal{B}_n^\Delta$ denote the set all bipartite graphs on the vertex sets U and V which satisfy all the properties of Lemma 3. Note that $\mathcal{B}_n^\Delta$ is parametrized by the values of Δ, μ, σ and λ.

4.1. Augmenting Paths in Random Bipartite Graphs

We switch our point of view from a probabilistic one to a deterministic one. More precisely, we will consider any *fixed* graph from the set $\mathcal{B}_n^\Delta$ and establish that it has short augmenting paths, which in turn implies that various algorithms perform well on *each* one of them. By Lemma 3 it is clear that these results will also apply, with high probability, to a random bipartite graph. The proof has the following general outline. Given B and M we grow two level graphs LG_1 and LG_2, as described earlier. The graph LG_1 has all the free boy vertices in it's first level, while LG_2 has all the free girl vertices in it's first level. We make use of the *expansion* property in Lemma 3(b) to argue that these two level graphs must grow rapidly in size within a small number of levels. Finally, we show that there must be some edge joining a vertex in LG_1 to a vertex in LG_2, yielding a short augmenting path for M.

Theorem 6: *Let $B \in \mathcal{B}_n^\Delta$ and M be any non-maximum matching in B. Then there exists an augmenting path for M of length at most $2L+1$, where $L = c\ln n/\ln\Delta$ and $c > 0$ is some constant.*

Proof: Given that $B \in \mathcal{B}_n^\Delta$, it must be the case that B satisfies all the requirements of Lemma 3. Suppose we construct two alternating level graphs for M, say LG_1 and LG_2, starting with free boy vertices and free girl vertices, respectively. We will assume that these level graphs will be constructed by exactly the same method as used by Dinic's algorithm. The construction is carried out only for the first L levels of both level graphs.

As in the proof of Theorem 5, we can assume that LG_1 has no free girl vertices, LG_2 has no free boy vertices, and the two level graphs are disjoint. Since M is a non-maximum matching, Fact 1 implies that there must exist an augmenting path for M. Therefore, each level of the two level graphs must be non-empty. We now fix our attention on the first level graph LG_1. As before, let LG_1 consist of the $L+1$ vertex sets $V_0, \ldots, V_L$. Let $l = L/2$ and, for $0 \leq i \leq l$, let $X_i = V_{2i} \cap L(U)$. Each X_i contains only the large degree boy vertices in V_{2i} and $X = \cup X_i$ is the set of all the large boy vertices in LG_1. Note that at least one of the three sets, X_0, X_1, and X_2, must be non-empty. Otherwise, each of the three (non-empty) levels, V_0, V_2 and V_4, contains only small degree boy vertices. This would imply the existence of two small degree boy vertices within distance 2 of some vertex in V_2. Since $B \in \mathcal{B}_n^\Delta$ satisfies Lemma 3(c), we know that this cannot happen and, therefore, $X^{(2)}$ must be non-empty.

We first show that LG_1 must contain a large number of vertices. To this end, we wish to claim that each successive X_i must expand in size. Let $X^{(i)}$ denote the set $X_0 \cup X_1 \ldots \cup X_i$ and let $V^{(i)}$ denote the set $V_0 \cup V_1 \ldots \cup V_i$. The following definition will be used in estimating the size of $X^{(l)}$.

Definition 2: *A* good *girl vertex for LG_1 is a girl vertex in $V^{(L)}$ which is matched by M to a* large *degree boy vertex, while a* bad *girl vertex is to a small degree boy vertex.*

Equivalently, a good girl vertex is one which is matched in M to some vertex in $X^{(l)} = V^{(L)} \cap L(U)$. The important property of bad girl vertices is stated in the following claim.

Claim 2: *There is at most* one *bad girl vertex adjacent in B to each boy vertex in $L(U)$.*

We establish the validity of this claim as follows. Suppose some vertex $u \in L(U)$ had two distinct bad girl vertices as its neighbors. Each bad girl vertex is matched by M to a distinct vertex in $S(U)$ and neither of these two vertices can be u itself since $u \in L(U)$. This implies that u has two distinct small degree boy vertices within distance 2 of itself, contradicting the property of Lemma 3(c).

Suppose we fix our attention on some particular set X_{i+1}. We would like to claim that this set is much larger in size then the previous set X_i. By definition, the set V_{2i+2} is obtained from V_{2i} as follows. The set V_{2i+1} consists of all neighbors of V_{2i} which are not present in any previous level, i.e. $V_{2i+1} = \Gamma_B(V_{2i}) \setminus V^{2i}$. The set V_{2i+2} is then the set of all the boy vertices which are matched by M to the girl vertices in V_{2i+1}. The sets $X^{(i)}$ and X_{i+1} are the sets of large degree boy vertices in $V^{(2i)}$ and V_{2i+2}, respectively. By Claim 2, each vertex in $X^{(i)}$ can be adjacent to at most one bad girl vertex. Therefore, we have that the set $\Gamma_B(X^{(i)})$ can contain at most $|X^{(i)}|$ bad girl vertices. Each good girl vertex in $V^{(2i)}$ must be matched to a large boy vertex which is also present in $V^{(2i)}$. Since $V^{(2i)}$ contains exactly $|X^{(i)}|$ large degree boy vertices, it can contain at most that many good girl vertices. Thus, we have that $\Gamma_B(X^{(i)})$ can have at most $|X^{(i)}|$ bad girl vertices and at most $|X^{(i)}|$ of its good girl vertices are already present in $V^{(2i)}$. The remaining vertices in $\Gamma_B(X^{(i)})$ must, therefore, be good girl vertices which lie in V_{2i+1} and we have the following claim.

Claim 3: *The set V_{2i+1} must contain at least $|\Gamma_B(X^{(i)})| - 2|X^{(i)}|$* good *girl vertices.*

It follows from the definition that each vertex in the set $X^{(i)}$ belongs to $L(U)$. By Lemma 3(b) we then have that $|\Gamma_B(X^{(i)})| \geq \Delta|X^{(i)}|/\lambda$ when $|X^{(i)}| \leq n/\Delta$. Each good girl vertex in V_{2i+1} is matched to a large degree boy vertex which will be in the set V_{2i+2}. Since X_{i+1} contains all the large degree boy vertices in V_{2i+2}, we claim that $|X_{i+1}|$ is the number of *good* girl vertices in V_{2i+1}. By Claim 3, V_{2i+1} has at least $|\Gamma_B(X^{(i)})| - 2|X^{(i)}|$ *good* girl vertices. This gives us the following claim. Note that $X^{(i+1)}$ consists of all the vertices in $X^{(i)}$ and X_{i+1}.

Claim 4: *For each $i < l$, the set $X^{(i+1)}$ contains at least $\Delta|X^{(i)}|/\lambda - |X^{(i)}|$ large boy vertices if $X^{(i)}$ contains less than n/Δ vertices.*

We have already established that $X^{(2)}$ is non-empty. Claim 4 states that each subsequent $X^{(i)}$ must increase in size by a factor of at least $\Delta/\lambda - 1$. It follows that, for an appropriate constant $c > 0$, the set $X^{(l-1)}$ contains at least n/Δ vertices. Let Z be a subset of X_{l-1} such that $Z \cup X^{(l-2)} \subseteq X^{(l-1)}$ contains exactly $\lfloor n/\Delta \rfloor$ elements. Applying the preceding argument to $Z \cup X^{(l-2)}$, instead of $X^{(l-1)}$, shows that $X^{(l)}$ must contain at least n/λ elements.

Claim 5: *For a sufficiently large constant $c > 0$, the set $X = X^{(l)}$ must contain at least n/λ vertices, where $l = (c \ln n)/(2 \ln \Delta)$.*

We can similarly show that there are at least n/λ large degree girl vertices in the first L levels of LG_2, viz. the set $Y \subseteq L(V)$. Observe that there cannot be any matching edges between X and Y. Otherwise, both end-points of such a matching edge would lie in both LG_1 and LG_2. This leads to a contradiction since we have assumed that LG_1 and LG_2 do not have any common vertices in their first L levels. If there is a free edge between a vertex in X and a vertex in Y then Lemma 2 would imply the existence of an augmenting path for M of length less than $2L + 1$. Therefore, if B contains no short augmenting paths with respect to the matching M then there exists $X \subseteq U$ and $Y \subseteq V$ such that there is no edge between the two sets and they have a cardinality of at least n/λ each. Since $B \in \mathcal{B}_n^{\Delta}$ satisfies the property in Lemma 3(d), we then have a contradiction. □

4.2. Matchings in Random Bipartite Graphs

The result of the previous section is now applied to the analysis of Dinic's algorithm. Theorem 7 states that for any graph $B \in \mathcal{B}_n^{\Delta}$, Dinic's algorithm terminates in nearly linear time, and its proof is similar to that of Corollary 1. It is important to observe that this result does not involve any probabilistic assumptions whatsoever. In fact, the following theorem is valid for any graph which has the properties specified in Lemma 3, even if it does not contain a perfect matching.

Theorem 7: *Let $B \in \mathcal{B}_n^{\Delta}$. Then Dinic's algorithm for bipartite matchings terminates in $O(m \log n / \log \Delta)$ time on input B.*

The following corollaries derive from an application of Theorem 7 to random bipartite graphs.

Corollary 2: *Let $B \in \mathcal{B}_{n,p}$, where $p(n) > \ln n/n$. Then, with high probability, Dinic's algorithm for bipartite matchings terminates in $O(m \log n / \log \Delta)$ time on input B. Moreover, the* expected *running time of Dinic's algorithm on inputs drawn from $\mathcal{B}_{n,p}$ is also $O(m \log n / \log \Delta)$.*

Proof: From Lemma 3 it follows that, with high probability, any $B \in \mathcal{B}_{n,p}$ is in the set $\mathcal{B}_n^\Delta$. We can now apply Theorem 7 to obtain the first part of the corollary. In the case where $B \in \mathcal{B}_n^\Delta$ we know that Dinic's algorithm takes $O(m \log n / \log \Delta)$ time. From the proof of Lemma 3 it can be seen that the probability that $B \notin \mathcal{B}_n^\Delta$ is $O(n^{-1/2})$. Since when $B \notin \mathcal{B}_n^\Delta$ Dinic's algorithm takes $O(\sqrt{n}m)$ time, we have the desired bound on the expected running time. □

Corollary 3: *Let the input for Dinic's algorithm be chosen uniformly at random from the set of all bipartite graphs on U, V. Then the* expected *running time of this algorithm is $O(m)$*

Proof: Observe that choosing B uniformly at random from the set of all bipartite graphs is equivalent to choosing B from $\mathcal{B}_{n,p}$ where $p = 1/2$. In this case, we have that $\Delta = n/2$ and Corollary 2 implies the desired result. □

The parallel implementation of the bipartite matching algorithm due to Gabow and Tarjan has a speed-up of $P/\log P$ when $P \leq m/\sqrt{n} \log n$ processors are used. They show that the running time of their algorithm depends on two parameters: A, the number of stages of augmentation in the sequential algorithm, and I, the total length of all the augmenting paths used by the algorithm. Their analysis shows that the parallel implementation of the bipartite matching algorithm runs in $O((A + Im/P) \log P)$ time, where $P < m$ is the number of processors used by the parallel algorithm.

Corollary 4: *Let $B \in \mathcal{B}_{n,p}$, where $p(n) > \ln n/n$. Then, with high probability, the Gabow-Tarjan algorithm with P processors terminates in $O((n + m/P) \log n \log P / \log \Delta)$ time. It has expected running time $O((n + m/P) \log P)$ when the input is chosen uniformly at random from the set of all bipartite graphs on U, V.*

Proof: For $B \in \mathcal{B}_n^\Delta$, we have already shown that I, the number of stages of augmentation, is $O(\log n / \log \Delta)$. Since each of the n shortest augmenting paths used by this algorithm has length at most $O(\log n / \log \Delta)$, we have that A is $O(n \log n / \log \Delta)$. □

4.3. Permanents of Random Bipartite Graphs

The problem of approximating the permanent of a dense bipartite graph B was solved by Broder [Bro86] and Jerrum and Sinclair [JS88] as follows. They first construct a *Markov chain* on the vertex set corresponding to the set of all matchings of cardinality n, say M_n, and all matchings of cardinality $n-1$, say M_{n-1}, in the graph B. We refer to the matchings in M_{n-1} as *almost perfect* matchings. The vertices corresponding to a matching in M_{n-1} and a matching in M_n are adjacent if the former is contained in the latter. The vertices corresponding to two distinct matchings in M_{n-1} are adjacent if they differ in exactly one edge. These edges are the only transitions in the Markov chain and, at any vertex, all possible transitions are equally likely. It was shown that the Markov chain is *rapidly mixing* if the ratio $|M_{n-1}|/|M_n|$ is polynomially bounded. In a rapidly mixing Markov chain, after a polynomial number of transitions the final state is independent of the choice of the initial state. A result due to Broder [Bro86] shows that such a rapidly mixing Markov chain can be used to generate perfect matchings *almost* uniformly and, thus, to approximate the permanent.

We apply our results to the problem of approximating the permanent of a random bipartite graph. First, Theorem 8 shows that for graphs drawn from the set $\mathcal{B}_n^\Delta$ the Markov chain scheme succeeds in approximating the permanent.

Theorem 8: *Let $B \in \mathcal{B}_n^\Delta$. Then the Markov chain method succeeds in approximating the permanent of B.*

Proof: Consider the sets of perfect and almost perfect matchings in B, i.e. M_n and M_{n-1}, respectively. By the results of Jerrum and Sinclair, it is clear that the Markov chain scheme will succeed in approximating the permanent if the ratio $|M_{n-1}|/|M_n|$ is bounded by a polynomial in n. We will prove that this is indeed the case for any $B \in \mathcal{B}_n^\Delta$.

By Theorem 6 we have that for any almost perfect matching $M \in M_{n-1}$ there is an augmenting path in B of length $O(\log n / \log \Delta)$. We will make use of this fact to establish the result as follows. For each perfect matching $M' \in M_n$, we will count the number of almost perfect matchings which can be augmented to M' by a short augmenting path. A polynomial upper bound on the number of such paths will lead to a polynomial upper bound on $|M_{n-1}|/|M_n|$.

We associate with each almost perfect matching $M \in M_{n-1}$ a *canonical* perfect matching $M' \in M_n$. This may be done as follows. For any such $M \in M_{n-1}$, choose any one *shortest* augmenting path P in B. Then the canonical choice for M will be the perfect matching M' obtained by augmenting M by P. Since every non-maximum matching in B has an augmenting path, it is clear that with every almost perfect matching in B we will be able to associate exactly one perfect matching in B.

We claim that at most a polynomial number of matchings in M_{n-1} choose any one matching in M_n as a canonical perfect matching. To see this, let $M' \in M_n$ be a fixed perfect matching. Each $M \in M_{n-1}$ which chooses M' as a canonical perfect matching can be augmented to M' by an augmenting path of length at most $2L+1$, where $L = c \ln n / \ln \Delta$. We will compute an upper bound on the number of such augmenting paths which could result in augmenting some almost perfect matching to M'.

Let $P = v_0 v_1 v_2 \ldots v_{2r}$ be an augmenting path of length $2r+1$ which augments an almost perfect matching to M'. There are n possible choices for v_{2r}, the last vertex in P. Having chosen an even-numbered vertex, v_{2k}, for P, the previous vertex in P, v_{2k-1}, can only be the vertex matched to v_{2k} by M'. Having chosen an odd-numbered vertex, v_{2k-1}, in P, the previous vertex, V_{2k-2}, could be any of the neighbors of v_{2k-1} in B. Note that, as per Lemma 3(a), each vertex in B has degree at most $\sigma\Delta$. Therefore, in B, the number of choices of P which could augment an almost perfect matching to M' must be $n(\sigma\Delta)^r$.

Observe that the choice of M' and P uniquely specifies the almost perfect matching which can be augmented by P to M'. Thus, the number of almost perfect matchings which choose M' as a canonical matching can be no more than the number of augmenting paths of length at most $2L+1$ which result in M'. This number can be bounded by the following expression using $L = c \ln n / \ln \Delta$.

$$Ln(\sigma\Delta)^L = O(n^{c+2})$$

It follows that the ratio $|M_{n-1}|/|M_n|$ is bounded by a polynomial in n. □

Corollary 5: *Let $B \in \mathcal{B}_{n,p}$, where $p(n)$ is any arbitrary function of n. Then, with high probability, the Markov chain method succeeds in approximating the permanent of B in polynomial time.*

Proof: First of all, observe that for $p < \ln n/n$ the graph B almost surely has isolated vertices and, therefore, does not have any perfect matchings. This implies that the permanent of B is zero, which can be easily verified in polynomial time by Dinic's algorithm. We now restrict out attention to the case where $p > \ln n/n$. By Lemma 3, we have that B must lie in the set $\mathcal{B}_n^{\Delta}$, with high probability. Now Theorem 8 implies the desired result. □

Let $\mathcal{B}_n$ denote the set of all bipartite graphs on the vertex sets U and V. Further, let $\mathcal{B}^P \subseteq \mathcal{B}_n$ be the set of all graphs in $\mathcal{B}_n$ in which the ratio $|M_{n-1}|/|M_n|$ is bounded by a polynomial in n. Clearly, $\mathcal{B}^P$ is also the set of all bipartite graphs in $\mathcal{B}_n$ for which the Markov chain scheme succeeds in approximating the permanent. The following result can be obtained from Corollary 5 by considering the case where $p = 1/2$.

Corollary 6: *The set of bipartite graphs $\mathcal{B}^P$ for which the permanent can be efficiently approximated has cardinality $|\mathcal{B}^P| = (1 - o(1))|\mathcal{B}_n|$.*

4.4. k-Factors in Random Bipartite Graphs

We now consider the problem of finding k-factors in random bipartite graphs. In the following analysis we will assume that k is fixed independent of n. However, our results can easily be extended to the case where k is a function of n provided $k < \Delta/c$, where Δ denotes the expected degree and c is some large constant. It is also not very hard to see that the following results also extend to the problem of finding a k-sub-factor. This is a spanning subgraph where each vertex is required to have a fixed degree which is *at most* k. Dinic's algorithm for bipartite matchings can be extended to the problem of finding k-factors.

The notion of augmenting paths and level graphs extend to this problem in a natural fashion [LP86, Tut81, Tut82]. Let F be a k-sub-factor of a bipartite graph B. A vertex is said to be *free* in F if its degree in F is strictly less than k. An edge belonging to F will be called a matching edge, all other edges in B will be called free edges; alternating paths and augmenting paths are defined as before. To augment F by an augmenting path P, remove from F the edges in $P \cap F$ and add to F the edges in $P \setminus F$. We say that F is a non-maximal k-sub-factor of B if there exists an augmenting path for F in B. Given a non-maximal

k-sub-factor F we define an *alternating level graph* $LG \subseteq B$ exactly as in the case of bipartite matchings. It is easily seen that Facts 1 and 2, as well as Lemmas 1 and 2 apply to these level graphs too.

The k-factor problem can be seen to be a special case of the network flow problem where all edge capacities are non-negative integers which do not exceed k. Dinic's algorithm, when applied to such network flow problem instances, terminates in $O(\sqrt{n}m)$ time. The following method for constructing a maximal alternating level graph for k-factors is used in Dinic's algorithm. We describe the construction starting from free boy vertices only; the level graph constructed by starting with free girl vertices is defined analogously. Let $B \setminus F$ denote the graph obtained by removing from B all the matching edges of F. For even i, $V_{i+1} \subseteq V$ contains all the neighbors in $B \setminus F$ of V_i which do not lie in $V^{(i)}$, i.e. $V_{i+1} = \Gamma_{B\setminus F}(V_i) \setminus V^{(i)}$. All free edges between V_i and V_{i+1} that are present in B are also present in LG. For odd i, V_{i+1} is the set of vertices which are matched in F to the vertices in V_i and are not already present in $V^{(i)}$, i.e. $V_{i+1} = \Gamma_F(V_i) \setminus V^{(i)}$. All the matching edges between the two vertex sets are also introduced into LG. The construction terminates at the first level where we encounter free girl vertices. Fact 2 implies that if there is a free girl vertex in the set V_l then there is an augmenting path for F which is of length l.

Our analysis of Dinic's algorithm for k-factors is complicated by the several crucial differences between the level graphs for matchings and k-factors. First of all, in the case of matchings, each boy vertex in V_i (for even $i > 0$) is matched to precisely one vertex in V_{i-1} and cannot be matched to any other vertex. The girl vertices in V_j (for odd j) can only be matched to a single boy vertex which will be present in the set V_{j+1}. In the level graphs for k-factors, each boy vertex in V_i (for even $i > 0$) can be matched by F to up to k girl vertices in V_{i-1} and *subsequent* levels. Moreover, each girl vertex in V_j (for odd j) can be matched by F to up to k boy vertices which will be present in V_{j+1} and the *previous* levels. Finally, in the level graphs for matchings no boy vertex in V_i can be matched to a girl vertex which is not already present in $V^{(i)}$. In the level graphs for k-factors, however, a boy vertex in V_i may be matched by F to girl vertices which are not already present in $V^{(i)}$. We do not include such girl vertices in the next level, V_{i+1}, unless they are adjacent via free edges to some other vertex in V_i. In any case, all matching edges between V_i and V_{i+1}, for even i, are ignored and are not introduced into the level graph.

The following theorem establishes that there exist short augmenting paths with respect to any non-maximal k-sub-factor in B, provided B satisfies the properties of Lemma 3. Note that our notation here is exactly the same as in the proof of Theorem 6.

Theorem 9: *Let $B \in \mathcal{B}_n^\Delta$ and F be any non-maximal k-sub-factor in B. Then there exists an augmenting path for F of length at most $2L+1$, where $L = c \ln n / \ln \Delta$ and $c > 0$ is some constant.*

Proof: Given that $B \in \mathcal{B}_n^\Delta$, it must be the case that B satisfies all the requirements of Lemma 3. Once again we construct the two alternating level graphs for F, say LG_1 and LG_2, as described above. We will only be interested in the first L levels of the level graphs and it will be assumed that the construction is terminated after L levels. Since F is non-maximal, Fact 1 implies that there must exist some augmenting path for F in B. Therefore, both LG_1 and LG_2 must be non-empty. We can assume that LG_1 has no free girl vertices, LG_2 has no free boy vertices, and the two level graphs are disjoint.

Since F is non-maximal, there must exist an augmenting path for F in B. This implies that each level of the two level graphs must be non-empty. We fix out attention on the first level graph LG_1. Let $V_0, \ldots, V_L$ be the $L+1$ levels of vertices in LG_1. Let $l = L/2$ and, for $0 \leq i \leq l$, let $X_i = V_{2i} \cap L(U)$. The set of all the large degree boy vertices in LG_1 will be denoted by $X = X^{(l)}$. Once again, it is easy to see that at least one of the three sets X_0, X_1 and X_2 will be non-empty.

The following definition applies only to those girl vertices in B which are not free in F.

Definition 3: *A* k-good *girl vertex for LG_1 is a girl vertex in $V^{(L)}$ which is matched by F to k large degree boy vertices. A* k-bad *girl vertex is matched by F to at least one small degree boy vertex.*

We will refer to a k-good (k-bad) girl vertex simply as a good (bad) girl vertex in this proof. A good girl vertex in $V^{(L)}$ is matched by F to k boy vertices in $X^{(l)} = V^{(L)} \cap L(U)$. An important property of bad girl vertices is stated in the following claim.

Claim 6: *There is at most* one *bad girl vertex adjacent to each large degree boy vertex in $L(U)$.*

The validity of this claim is established as follows. Suppose some vertex $u \in L(U)$ has two distinct bad girl vertices $v, w \in V$ as neighbors. Then let $x, y \in S(U)$ be the two small degree boy vertices which are matched by F to v and w, respectively. Clearly, both x and y must be distinct from u itself since $u \in L(U)$. If x and y are distinct then there are two distinct small degree boy vertices within distance 2 of u, contradicting the property of Lemma 3(c). On the other hand, if x and y are the same vertex then u has two paths of length 2 to a small degree boy vertex, again contradicting the property of Lemma 3(c).

We wish to show that the set X_{i+1} is much larger in size than the set $X^{(i)}$. The sets X_{i+1} and $X^{(i)}$ contain all the large degree boy vertices in V_{2i+2} and $V^{(2i)}$, respectively. The set V_{2i+2} is obtained from $V^{(2i)}$ as follows. We first include in V_{2i+1} those girl vertices which are not already present in $V^{(2i)}$ and are connected by a free edge to some boy vertex in V_{2i}, i.e. $V_{2i+1} = \Gamma_{B \setminus F}(V_{2i}) \setminus V^{(2i)}$. The set V_{2i+2} is then the set of all the boy vertices which are not already present in $V^{(2i)}$ and are matched by F to some vertex in V_{2i+1}, i.e. $V_{2i+2} = \Gamma_F(V_{2i+1}) \setminus V^{(2i)}$.

Consider the set $\Gamma_{B\setminus F}(X^{(i)}) \setminus V^{(2i)}$, this is the set of the girl vertices which are introduced into V_{2i+1} by the boy vertices in X_i. By Claim 6, each vertex in $X^{(i)}$ can be adjacent to at most one bad girl vertex. This implies that $\Gamma_{B\setminus F}(X^{(i)})$ can contain at most $|X^{(i)}|$ bad girl vertices. Let g denote the number of good girl vertices in $V^{(2i)}$. Then we have that $\Gamma_{B\setminus F}(X^{(i)})$ can contain at most $|X^{(i)}|$ bad girl vertices and at most g of its good girl vertices can already be present in $V^{(2i)}$. The remaining vertices in $\Gamma_{B\setminus F}(X^{(i)})$ must be good girl vertices which will be present in V_{2i+1}. This establishes the following claim.

Claim 7: *The set V_{2i+1} must contain at least $|\Gamma_{B\setminus F}(X^{(i)})| - |X^{(i)}| - g$ good girl vertices, where g is the number of good girl vertices in $V^{(2i)}$.*

Each good girl vertex in V_{2i+1} is matched by F to k large degree boy vertices. Therefore, there are at least $k(|\Gamma_{B\setminus F}(X^{(i)})| - |X^{(i)}| - g)$ matching edges incident on the good girl vertices in V_{2i+1}. Let x be the number of these matching edges which go to the large degree boy vertices in $V^{(2i)}$. Thus, the number of matching edges going from the good girl vertices in V_{2i+1} to the large degree boy vertices in V_{2i+2} must be at least $k(|\Gamma_{B\setminus F}(X^{(i)})| - |X^{(i)}| - g) - x$.

The $|X^{(i)}|$ large degree boy vertices in $V^{(2i)}$ can have at most $k|X^{(i)}|$ matching edges incident on them. But, kg of these matching edges must connect them to the g good girl vertices in $V^{(2i)}$, since these girl vertices can be matched only to the large degree boy vertices in $V^{(2i)}$. This implies that $x \leq k|X^{(i)}| - kg$. We have that there are at least $k(|\Gamma_{B\setminus F}(X^{(i)})| - 2|X^{(i)}|)$ matching edges going from the good girl vertices in V_{2i+1} to the large degree boy vertices in V_{2i+2}. Since no vertex can be matched to more than k other vertices, we have the following claim.

Claim 8: *The set X_{i+1} contains at least $|\Gamma_{B\setminus F}(X^{(i)})| - 2|X^{(i)}|$ large degree boy vertices.*

By Lemma 3(b), we have that if $|X_{i+1}| < n/\Delta$ then $|\Gamma_B(X^{(i)})| \geq \Delta|X_{i+1}|/\lambda$. Let y be the number of vertices in $\Gamma_B(X^{(i)})$ which are connected to vertices in $X^{(i)}$ only by matching edges, i.e. $y = |(\Gamma_B(X^{(i)}) - \Gamma_{B\setminus F}(X^{(i)}))|$. It is clear that $y \leq k|X^{(i)}|$ since each vertex in $X^{(i)}$ can be matched to at most k vertices in $\Gamma_B(X^{(i)}) - \Gamma_{B\setminus F}(X^{(i)})$. Therefore, we have that $|\Gamma_{B\setminus F}(X^{(i)})| = |\Gamma_B(X^{(i)})| - y$, where $y \leq k|X^{(i)}|$ and $|\Gamma_B(X^{(i)})| \geq \Delta|X_{i+1}|/\lambda$. This establishes the following claim.

Claim 9: *For each $i < l$, the set $X^{(i+1)}$ contains at least $((\Delta/\lambda) - k - 1)|X^{(i)}|$ large degree boy vertices, when $X^{(i)}$ contains less than n/Δ vertices.*

We have already shown that $X^{(2)}$ is non-empty. Claim 9 now implies that each subsequent $X^{(i)}$ must expand in size by a factor of at least $((\Delta/\lambda) - k - 1)$. Since k is fixed and $\Delta > \ln n$, it follows that the set $X^{(l-1)}$ will contain at least n/Δ elements, for a suitably large choice of the constant c. Another stage of expansion will guarantee that $X = X^{(l)}$ will then contain at least n/λ vertices. By symmetry, it is easy to see that the set Y of the large degree girl vertices in LG_2 must also contain at least n/λ vertices.

Unlike in the case of matchings, it is possible that there is a matching edge between X and Y without the two end-points of the matching edge being present in both LG_1 and LG_2. We handle this complication as follows. Let $X' \subseteq X$ be such that $|X'| = n/2k\lambda$. Then the vertices in X' can be matched by F to at most $n/2\lambda$ vertices in Y. Therefore, there exists a set $Y' \subseteq Y$ such that $|Y'| = n/2k\lambda$ and there is no matching edges from X' to Y'. If there is a free edge between X' and Y' then, by Lemma 2, we can find an augmenting path for F of length at most $2L + 1$.

We have shown that if there is any edge going from X' to Y' then there is a short augmenting path for F in B. The only case left to consider is when there is no edge in B between X' and Y', both of which contain at least $n/2k\lambda$ vertices. Choosing $\beta = 2k\lambda$, this can be seen to contradict the property of Lemma 3(d). $\Box$

The following corollaries derive from a straightforward application of Lemma 3 and Theorem 9.

Corollary 7: *Let $B \in \mathcal{B}_n^{\Delta}$. Then Dinic's algorithm for k-factors terminates in $O(m \log n / \log \Delta)$ time on input B.*

Corollary 8: *Let $B \in \mathcal{B}_{n,p}$, where $p(n) > \ln n/n$. Then, with high probability, Dinic's algorithm for k-factors terminates in $O(m \log n/\log \Delta)$ time on input B. Moreover, the expected running time of Dinic's k-factor algorithm on inputs drawn from $\mathcal{B}_{n,p}$ is also $O(m \log n/ \log \Delta)$.*

Corollary 9: *Let the input for Dinic's k-factor algorithm be chosen uniformly at random from the set of all bipartite graphs on U and V. Then the expected running time of this algorithm is $O(m)$.*

5. The d-out Random Bipartite Graph Model

The space of d-out random bipartite graphs was introduced to enable an analysis of maximum matchings in sparse random graphs [Wal80]. The $\mathcal{B}_{n,p}$ model is not useful from this point of view because it has the same threshold density for perfect matchings as for having minimum degree 1. This trivializes the issue, especially for permanents, since the only reason why sparse $\mathcal{B}_{n,p}$ graphs cannot have perfect matchings is because they contain isolated vertices. We analyze the average-case behavior of Dinic's and related algorithm on the space of d-out graphs. Our results for this case are very similar to those obtained for graphs drawn from $\mathcal{B}_{n,p}$.

Walkup showed that for $d = 1$ almost every graph in $\mathcal{B}_{d-out}$ does not have a perfect matching. On the other hand, for $d = 2$ almost every graph in $\mathcal{B}_{d-out}$ does have a perfect matching. We are unable to prove the expansion property for graphs drawn from $\mathcal{B}_{d-out}$ when $d = 2$. However, all the following results do hold for the case where $d \geq 4$. In Lemma 4 we prove that almost every graph in $\mathcal{B}_{d-out}$ has an expanding structure.

Lemma 4: *Let $B \in \mathcal{B}_{d-out}$, where $d \geq 4$. Then, with high probability, the following properties hold for B, where $\lambda = 2.5$.*

(a) *If $W \subseteq U$ and $|W| \leq n/d$ then $\Gamma_B(W) \geq d|W|/\lambda$.*

(b) *Let $X \subseteq U$ and $Y \subseteq V$ be such that $|X|, |Y| \geq n/\beta$, for any constant β such that $1 \leq \beta \leq 2.5$. Then it cannot be the case that there is no edge between X and Y in B.*

Proof: (a). Call (X, Y), where $X \subseteq U$ and $Y \subseteq V$, a *k-bottleneck* pair for B if $|X| = k < n/d$, $|Y| = dk/\lambda$ and $\Gamma_B(X) \subseteq Y$. Let Ψ be a random variable which denotes the number k-bottleneck pairs in B. Clearly, B does not satisfy the expansion property if and only if $\Psi > 0$. Let (X, Y) be a fixed pair such that $X \subseteq U$, $Y \subseteq V$, $|X| = k < n/d$ and $|Y| = K = dk/\lambda$. The probability that the pair (X, Y) is a k-bottleneck pair is exactly the probability of the event $\Gamma_B(X) \subseteq Y$. This will happen only if each vertex in X chooses its d neighbors only from Y and each vertex in $V \setminus Y$ chooses its d neighbors only from $U \setminus X$. The probability of this event may be bounded from above by the following expression.

$$\text{Prob}[\Gamma_B(X) \subseteq Y] \leq \left(\frac{\binom{K}{d}}{\binom{n}{d}} \right)^k \left(\frac{\binom{n-k}{d}}{\binom{n}{d}} \right)^{n-K} \leq \left(\frac{K}{en}\right)^{dk} e^{dkK/n}$$

By summing over all choices of the pair (X, Y), we obtain the following from Markov's Inequality.

$$\text{Prob}[\Psi > 0] \leq \binom{n}{k}\binom{n}{K}\left(\frac{K}{en}\right)^{dk} e^{dkK/n}$$

Since $k < n/d$ and $K = kd/\lambda$, it follows that $kKd/n < K < n/\lambda$. We then obtain the following inequality.

$$\mathrm{Prob}[\Psi > 0] \leq \frac{e^{k+2K-dk}}{n^{dk-k-K}}\frac{K^{dk-K}}{k^k} \leq e^{k+2K-dk}\left(\frac{d}{\lambda}\right)^{dk-K}\left(\frac{k}{n}\right)^{dk-k-K}$$

If $k < \sqrt{n}$ then $k/n < n^{-1/2}$ and the above expression is easily seen to be appropriately small. Therefore, we only consider the case where $\sqrt{n} < k < n/d$. Now, using $\lambda = 2.5$, $d \geq 4$ and $K < dk/\lambda$, we have the following inequality for some fixed $\epsilon > 0$. This completes the proof.

$$\mathrm{Prob}[\Psi > 0] \leq \left(\frac{e4^{11}}{5^{12}}\right)^{k/5} \leq e^{-\epsilon\sqrt{n}}$$

□

(b). Let Ψ be a random variable which denotes the number of pairs (X, Y) such that $X \subseteq U$, $Y \subseteq V$, $|X| = |Y| = n/\beta$ and $Y \cap \Gamma_B(X) = \emptyset$. Clearly, the probability that Ψ is non-zero is exactly the probability that the lemma does not hold for B. Consider any fixed pair (X, Y) such that $X \subseteq U$, $Y \subseteq V$ and $|X| = |Y| = n/\beta$. The probability that $Y \cap \Gamma_B(X) = \emptyset$ is given by the probability of the event that no vertex in X chooses a vertex in Y as a neighbor and vice versa.

$$\mathrm{Prob}[Y \cap \Gamma_B(X) = \emptyset] = \left(\frac{\binom{n-n/\beta}{d}}{\binom{n}{d}}\right)^{n/\beta}\left(\frac{\binom{n-n/\beta}{d}}{\binom{n}{d}}\right)^{n/\beta}$$

Summing over all choices of X and Y, we have the following inequality.

$$\mathrm{Prob}[\Psi > 0] \leq \mathrm{E}[\Psi] = \binom{n}{n/\beta}^2\left(\frac{\binom{n-n/\beta}{d}}{\binom{n}{d}}\right)^{2n/\beta}$$

Upon suitable algebraic manipulation we obtain the following inequality.

$$\mathrm{Prob}[\Psi > 0] \leq \left(e\beta\left(1-\frac{1}{\beta}\right)^d\right)^{2n/\beta}$$

Given that $d \geq 4$ and $1 < \beta \leq 2.5$, the following inequality establishes the desired result. Here $\epsilon > 0$ is some positive constant.

$$\mathrm{Prob}[\Psi > 0] \leq \left(\frac{81e}{250}\right)^{2n/\beta} \leq e^{-\epsilon n}$$

□

By symmetry, the expansion property of Lemma 4(a) also holds for the girl vertices. As before, we restrict our attention to those d-out bipartite graphs which satisfy Lemma 4. Let $\mathcal{B}_n^d$ denote the set of all d-out bipartite graphs on the vertex sets U and V which have the properties of Lemma 4. In Theorem 10 we show that every non-maximum matching in $B \in \mathcal{B}_n^d$ will have a short augmenting path. The proof of this theorem is similar to that of Theorem 6.

Theorem 10: *Let $B \in \mathcal{B}_n^d$, where $d \geq 4$, and let M be any non-maximum matching in B. Then there exists an augmenting path of length at most $2L+1$ for M in B, where $L = c \ln n / \ln d$ and $c > 1$ is some constant.*

Proof: Consider two alternating level graphs for M, say LG_1 and LG_2, starting with free boy vertices and free girl vertices, respectively. The construction of the two level graphs is carried out for the first L levels only. We may assume that each of the first L levels of both level graphs is non-empty, the two level graphs are disjoint and have free vertices only in their starting levels.

Let LG_1 consist of the $L+1$ vertex sets $V_0, \ldots, V_L$. Let $l = L/2$ and, for $0 \le i \le l$, let $X_i = V_{2i}$ denote the i^{th} level of boy vertices in LG_1. Also, let $X^{(i)}$ denote the set $X_0 \cup X_1 \ldots \cup X_i$ and let $V^{(i)}$ denote the set $V_0 \cup V_1 \ldots \cup V_i$.

We first establish that $X^{(i+1)}$ is much larger in size than $X^{(i)}$. Observe that $V_{2i+1} = \Gamma_B(X^{(i)}) \setminus V^{(2i)}$, i.e. V_{2i+1} is the set of all girl vertices which are adjacent to the boy vertices in $X^{(i)}$ and are not already present in $V^{(2i)}$. Since there are no free girl vertices in LG_1, each girl vertex in $V^{(2i)}$ is matched to a boy vertex in $V^{(2i)}$. It then follows that there are at most $|X^{(i)}|$ girl vertices in $V^{(2i)}$. This implies that at most $|X^{(i)}|$ of the vertices in $\Gamma_B(X^{(i)})$ can already be present in $V^{(2i)}$. We conclude that the number of girl vertices in V_{2i+1} is at least $|\Gamma_B(X^{(i)})| - |X^{(i)}|$. The number of vertices in X_{i+1} (or V_{2i+2})is the same as $|V_{2i+1}|$ because each vertex in X_{i+1} is matched to a girl vertex in V_{2i+1} and no girl vertex in LG_1 is free. By Lemma 4(a), we have that $|X_{i+1}| = |V_{2i+1}| \ge |X^{(i)}|d/\lambda - |X^{(i)}|$. Thus, it is clear that $|X^{(i+1)}| \ge |X^{(i)}|d/\lambda$.

Since $X^{(0)}$ is non-empty we have that, for sufficiently large constants $c > 1$, the set $X = X^{(l)}$ contains at least n/λ boy vertices. Similarly, it can be shown that the set Y of all the girl vertices in LG_2 has at least n/λ elements. By Lemma 4(b), there must be an edge between a vertex in X and a vertex in Y. There can be no matching edge between X and Y unless LG_1 and LG_2 have common vertices in their first L levels. Since we have assumed that the two level graphs are disjoint, it must be the case that there is a free edge between X and Y. Any such edge completes an augmenting path for M of length at most $2L+1$ and this concludes the proof. □

Lemma 4 and Theorem 10 together imply the following corollaries. The proofs of these corollaries are exactly the same as those for Corollaries 2 and 4.

Corollary 10: *Let $B \in \mathcal{B}_{d-out}$, where $d \ge 4$. Then, with high probability, Dinic's algorithm for bipartite matchings terminates in $O(m\log n/\log d)$ time on input B. Moreover, the* expected *running time of Dinic's algorithm on inputs drawn from $\mathcal{B}_{d-out}$ is also $O(m\log n/\log d)$.*

Corollary 11: *Let $B \in \mathcal{B}_{d-out}$, where $d \ge 4$. Then, with high probability, the Gabow-Tarjan algorithm with P processors terminates in $O((n + m/P)\log n \log P/\log d)$ time.*

6. Non-Bipartite Graphs

We now turn our attention to the analysis of matching and flow algorithms for non-bipartite graphs. We first study the structure of graphs drawn from $\mathcal{G}_{n,p}$. It turns out that these graphs satisfy structural properties which are very similar to those of the graphs drawn from $\mathcal{B}_{n,p}$.

Let $G(V,E) \in \mathcal{G}_{n,p}$, where $p > \ln n/(n-1)$. Let Δ be the expected degree of a vertex in G, i.e. $\Delta = p(n-1)$. We partition the vertex set V into two subsets as follows. Let $S(V) = \{v \in V : \text{degree}(v) < \Delta_s\}$ and $L(V) = V \setminus S(V)$, where $\Delta_s = \Delta/\mu$ for some constant $\mu > 1$. As before, we refer to the vertices in $S(V)$ as *small* degree vertices and the vertices in $L(V)$ will be called *large* degree vertices.

Lemma 5 below specifies the structural properties of almost every graph in $\mathcal{G}_{n,p}$. These properties bear a close resemblance to the structural properties of $G \in \mathcal{G}_{n,p}$ which were exhibited by Bollobás and Frieze [BF85].

Lemma 5: *Let $G \in \mathcal{G}_{n,p}$, where $p > \ln n/(n-1)$. Then, with high probability, the following properties hold for G with a suitable choice of the positive constants μ and $\lambda > 4\mu$.*

(a) *If $X \subseteq L(V)$ and $|X| \le n/\Delta$ then $|\Gamma_G(X)| \ge \Delta|X|/\lambda$.*

(b) *No path of length 4 in G can contain more than one small degree vertex.*

(c) *Let X and Y be disjoint subsets of V such that $|X|, |Y| \ge n/\beta$, for any constant $\beta > 1$. Then there is an edge in G with one end-point in X and the other end-point in Y.*

Proof: (a). Let $X \subseteq L(V)$ be such that $|X| = k \le n/\Delta$. We say that (X,Y) is a k-bottleneck pair if $Y \cap X = \emptyset$, $|Y| = \Delta k/\lambda$ and $\Gamma_G(X) \subseteq Y$. The statement of the lemma does not hold for G if and only if there exists some k-bottleneck pair in G. To establish that, with high probability, there does not exist any k-bottleneck pair in G we consider two cases.

Case 1 : $\{ 1 \leq k \leq n/\Delta^3 \}$

Suppose (X, Y) is a k-bottleneck pair in G. Let $Z = X \cup Y$, then Z contains $z = (1+\Delta/\lambda)k$ vertices. Each vertex in X is a large degree vertex and $\Gamma_G(X) \subseteq Y$. This implies that there are at least $\Delta_s = \Delta/\mu$ edges going from each vertex in X to some vertex in $X \cup Y$. If we sum the number of such edges over all vertices in X then each edge will be counted at most twice. This establishes that there are at least $y = \Delta_s k/2$ edges which have both end-points in $Z = X \cup Y$. Thus, the existence of a k-bottleneck pair in G implies that there are z vertices in G which induce a subgraph with at least y edges.

We now show that such a dense induced subgraph is extremely unlikely to be present in G. Let Z be some fixed subset of V containing z vertices. Let $E(Z)$ denote the number of edges in G which have both end-points in Z. There can be at most $t = z(z-1)/2$ edges in $E(Z)$; each such edge is present in G with probability p. Thus,

$$\text{Prob}[|E(Z)| \geq y] \;\leq\; \sum_{r=y}^{t} \binom{t}{r} p^r (1-p)^{t-r}$$

The expected number of edges with both end-points in Z is tp which is at most $k/2\lambda^2$. Since $y = \Delta_s k/2$ is greater than tp, we can make use of Chernoff's bound [AV79, Che52] on the tail of the binomial distribution to obtain the following.

$$\text{Prob}[|E(Z)| \geq y] \;\leq\; \left(\frac{tp}{y}\right)^y \left(\frac{t(1-p)}{t-y}\right)^{t-y}$$

Let γ denote the ratio of the number of edges to the number of vertices in the subgraph of G induced by Z, i.e. $\gamma = E(Z)/z$. It follows that $\gamma \geq \lambda/2a$; since $\lambda > 4\mu$, we have that $\gamma > 2$. Moreover, $k \leq n/\Delta^3$ and this implies that $z \leq n/\lambda\Delta^2$. Summing over all choices of the set Z, we obtain the following bound on the probability that there is a set Z with $z \leq n/\lambda\Delta^2$ vertices and at least γz edges.

$$\sum_{z=1}^{n/\lambda\Delta^2} \binom{n}{z} \left(\frac{etp}{\gamma z}\right)^{\gamma z} (1-p)^{t-\gamma z} \;=\; O(n^{-(\gamma-1)})$$

Thus, we have that the probability that there exists a set of at most n/Δ^3 large degree vertices in G which does not expand by a factor of Δ/λ is $O(n^{-(\gamma-1)})$.

Case 2 : $\{ n/\Delta^3 \leq k \leq n/\Delta \}$

Let X be some fixed subset of V such that $|X| = k$. In this part of the proof we will not need to assume that X contains only large degree vertices. To prove the statement of the lemma we need to give a suitable upper bound on the probability of the event $|\Gamma_G(X)| \leq \Delta k/\lambda$.

Let E_v denote the event that some vertex $v \in V \setminus X$ lies in $\Gamma_G(X)$. The events E_u, for all $u \in V \setminus X$, are mutually independent and occur with equal probability, say P_X. Clearly, the probability $P_X = 1 - (1-p)^k$ is at most kp. We then have the following bound on the probability of $\Gamma_G(X)$ having fewer than $K = \Delta k/\lambda$ vertices.

$$\text{Prob}[|\Gamma_G(X)| \leq K] \;\leq\; \sum_{t=0}^{K} \binom{n}{t} (P_X)^t (1-P_X)^{n-t}$$

The expected number of vertices in $\Gamma_G(X)$ is at most nP_X and this is greater than $K = \Delta k/\lambda$. The Chernoff bound on the tail of the binomial distribution is applicable to the above expression and yields the following inequality.

$$\text{Prob}[|\Gamma_G(X)| \leq K] \;\leq\; \left(\frac{nP_X}{K}\right)^K \left(\frac{n(1-P_X)}{n-K}\right)^{n-K} \;\leq\; (e\lambda)^K (1-p)^{k(n-K)}$$

Summing over all choices of the set X, we obtain the following upper bound on the probability that there exists a set X with k vertices such that X has at most $K = \Delta k/\lambda$ neighbors in G. Here $\epsilon = 1 - (\ln \lambda + 2)/\lambda$

is a small positive constant.

$$\sum_{k=n/\Delta^3}^{n/\Delta} \binom{n}{k} (e\lambda)^K (1-p)^{k(n-K)} = O(e^{-\epsilon n/\Delta^2})$$

This completes the proof of the two cases. □

(b). Let Ψ denote the number of paths of length 4 in G which contain more than one small degree vertex. The expected number of paths of length 4 in G is no more than n^5p^4. Let P_S denote the probability that any two fixed vertices are of small degree in G. We can then bound the probability of Ψ being non-zero as follows.

$$\text{Prob}[\Psi > 0] \leq \text{E}[\Psi] \leq n^5 p^4 \binom{5}{2} P_S \leq 10n\Delta^4 P_S$$

We can bound P_S as follows.

$$P_S \leq \left(\sum_{i=0}^{\Delta_s} \binom{n}{i} p^i (1-p)^{n-i} \right)^2$$

An application of Chernoff's bound on the tail of a binomial distribution then yields the following inequality. Here $\epsilon = (\ln \mu + 2)/\mu$ is a small positive constant.

$$P_s \leq \left(\left(\frac{np}{\Delta_s} \right)^{\Delta_s} \left(\frac{n(1-p)}{n - \Delta_s} \right)^{n-\Delta_s} \right)^2 \leq e^{-2(1-\epsilon)\Delta}$$

Thus, we have the following bound on the probability of Ψ being non-zero.

$$\text{Prob}[\Psi > 0] \leq 10n\Delta^4 e^{-2(1-\epsilon)\Delta} = O(n^{-(1-3\epsilon)})$$

The probability that G violates the lemma is exactly that the probability that Ψ is greater than zero. Since ϵ can be made arbitrarily small by choosing a to be large, this implies the desired result. □

(c). Let Ψ be a random variable which denotes the number of pairs (X, Y) in G such that $X \cap Y = \emptyset$, $|X| = |Y| = n/\beta$ and there is no edge between X and Y. Using Markov's Inequality we obtain the following bound on the probability that Ψ is non-zero.

$$\text{Prob}[\Psi > 0] \leq \text{E}[\Psi] \leq \binom{n}{n/\beta}^2 (1-p)^{n^2/\beta^2}$$

Since $\Delta > \ln n$, we have the following for sufficiently large n and any fixed $\gamma > 0$.

$$\text{Prob}[\Psi > 0] \leq e^{2n(\ln \beta + 1)/\beta} e^{-n\Delta/\beta^2} = O(n^{-\gamma})$$

The probability that $\Psi > 0$ is exactly the probability that G does not satisfy the lemma. The above inequality establishes the desired result. □

6.1. Alternating Level Graphs for Non-bipartite Matchings

Our analysis of the lengths of augmenting paths in graphs drawn from $\mathcal{G}_{n,p}$ will require the notion of alternating level graphs for non-bipartite matchings. We will define such level graphs and present some of their properties.

Let $G(V, E)$ be a graph with the vertex set V and the edge set E. We will assume that G contains an even number of vertices. Let $M \subseteq E$ be a non-maximum matching in G, then there must exist at least two free vertices in G. The definitions of alternating and augmenting paths are as in the case of bipartite matchings. The following fact [HK73] guarantees the existence of an augmenting path for every non-maximum matching in G.

Fact 4: *Let M be a matching in a bipartite graph G. If M is not a maximum matching in G then G contains an augmenting path for M.*

The definition of an *alternating level graph* $LG \subseteq G$ for non-bipartite matchings is different from that for bipartite graphs. A level graph LG with l levels consists of $l+1$ disjoint sets of vertices $V_0, V_1, \ldots, V_l \subseteq V$. The vertex set V_0 can contain only free vertices. All edges in LG must go from a vertex in level i to a vertex in level $i+1$, for some $i \geq 0$. For even i, an edge between V_i and V_{i+1} must be a free edge. For odd i, an edge between V_i and V_{i+1} must be a matching edge. Each vertex in V_i, for $i > 0$, must be adjacent to at least one vertex in the previous level V_{i-1}.

Fact 5: *Suppose there is a free vertex in level i of LG. Then G contains an augmenting path for M which is of length i.*

Let v_1 and v_2 be two free vertices for M. In our analysis we will be constructing two level graphs for M. The first level graph, LG_1, will contain only v_1 in its initial level, while the second level graph, LG_2, will contain only v_2 in its initial level. We now present a crucial property of the two level graphs LG_1 and LG_2. The proof of the following lemma is very similar to that of Lemma 2 and is omitted.

Lemma 6: *Let LG_1 and LG_2 be two disjoint level graphs which have at most L levels each. If there is a free edge in G joining a vertex in an even-numbered level of LG_1 to a vertex in an even-numbered level of LG_2 then G contains an augmenting path for M of length at most $2L+1$.*

6.2. Augmenting Paths in Random Graphs

In this section we show that there are short augmenting paths for any non-maximum matchings in graphs which satisfy Lemma 5.

Definition 4: *The set $\mathcal{G}_n^\Delta$ consists of all graphs on n vertices which have the three properties specified in Lemma 5.*

Note that the set $\mathcal{G}_n^\Delta$ is parametrized by the value of Δ as well as the constants μ and λ. The dependence of $\mathcal{G}_n^\Delta$ on μ and λ will not be explicitly specified for the sake of notational convenience.

In Theorem 11 below we will show that for any non-maximum matching in $G \in \mathcal{G}_n^\Delta$ there exists an augmenting path of length $O(\log n/\log \Delta)$. The proof of this theorem is similar in some respects to the proof of Theorem 6. However, since we are now dealing with non-bipartite graphs, there are several obstacles to adapting the proof of Theorem 6 to the current scenario.

The first problem that we face is the following. As before, we will be constructing two level graphs LG_1 and LG_2 for a non-maximum matching M. The construction of both level graphs will be terminated after L levels have been created. Suppose a vertex v is present in both LG_1 and LG_2. In the bipartite case, Lemma 1 would then imply that there exists an augmenting path for M of length at most $2L$. This is because when v is a boy vertex then it must occur in an even-numbered level of LG_1 and an odd-numbered level of LG_2. Similarly, when v is a girl vertex then it must occur in an odd-numbered level of LG_1 and an even-numbered level of LG_2. On the other hand, in the case of non-bipartite graphs it is possible that v occurs in the even-numbered (or, odd-numbered) levels of both LG_1 and LG_2. Then we cannot guarantee the existence of a short augmenting path for M even though LG_1 and LG_2 share a vertex in their first few levels. It is important for the purposes of our analysis to ensure that LG_1 and LG_2 are disjoint unless we already have a short augmenting path for M.

There is also a second problem. Consider the level-by-level construction of a level graph LG. Suppose we have already created the first i levels of LG, where i is odd. In the bipartite case, we would then have constructed the next level, V_{i+1}, by including every vertex which is matched to some girl vertex in V_i. It is clear that no two vertices in V_i could be matched to each other since they are all girl vertices. Thus, we were able to guarantee that V_{i+1} contains exactly the same number of vertices as V_i. Unfortunately, in the case of non-bipartite graphs it is possible for any two vertices in V_i to be matched to each other. The level V_{i+1} can contain only those vertices in G which are matched to some vertex in V_i and are not already present in $V^{(i)}$. Thus, our argument that V_{i+1} contains exactly as many vertices as V_i would break down whenever any two vertices in V_i are matched to each other.

To handle these two problems we will modify the proof of Theorem 6 as follows. We will label as boys the vertices in the even-numbered levels of LG_1 and the odd-numbered levels of LG_1. Similarly, we will label as girls the vertices in the odd-numbered levels of LG_1 and the even-numbered levels of LG_2. We will then ensure that no vertex gets labeled as both a boy and a girl. Further, our construction of the level graphs will guarantee that there is no matching edge between two vertices which are both labeled as boys or which are both labeled as girls. Observe that it is not very hard to ensure that the above labeling process is consistent and satisfies the desired conditions. The hard part is to ensure that the expansion of the levels is maintained under the labeling process. We will formulate the labeling process in terms of certain invariant conditions which must be satisfied by the level graphs at each step of their construction.

Theorem 11: *Let $G \in \mathcal{G}_n^{\Delta}$ and M be any non-maximum matching in G. Then there exists an augmenting path for M of length at most $2L+1$, where $L = c \ln n / \ln \Delta$ and $c > 0$ is some constant.*

Proof: Since M is not a maximum matching in G, there must exist at least two free vertices for M. We will construct two level graphs starting at each of these two free vertices. We first provide some notation for describing the two level graphs. The actual construction of these level graphs will be described later.

The two level graphs LG_1 and LG_2 will consist of at most L levels each. The first level graph LG_1 will consist of the vertex sets $U_0, U_1, \ldots U_L$. Let $l = L/2$ and, for $0 \le i \le l$, let $X_i = U_{2i}$. We will use $U^{(i)}$ to denote $U_0 \cup U_1 \cup \ldots \cup U_i$ and $X^{(i)}$ to denote $X_0 \cup X_1 \cup \ldots \cup X_i$. Similarly, the level graph LG_2 will consist of the vertex sets $V_0, V_1, \ldots V_L$ and, for $0 \le i \le l$, let $Y_i = V_{2i}$. Again, $V^{(i)}$ will denote $V_0 \cup V_1 \cup \ldots \cup V_i$ and $Y^{(i)}$ will denote $Y_0 \cup Y_1 \cup \ldots \cup Y_i$. We will refer to the vertices in the even-numbered levels of LG_1 and the odd-numbered levels of LG_2 as *pseudo-boy* vertices. The vertices in the odd-numbered levels of LG_1 and the even-numbered levels of LG_2 will be called *pseudo-girl* vertices. In particular, the sets X_i contain only pseudo-boy vertices and the sets Y_j contain only pseudo-girl vertices.

The construction of the two level graphs will proceed in at most l phases. In phase i we will add the levels U_{2i-1} and $U_{2i} = X_i$ to LG_1 and we will add the levels V_{2i-1} and $V_{2i} = Y_i$ to LG_2. If at any point in the construction we manage to complete an augmenting path for M then the construction will be terminated. Any such augmenting path will be of length at most $2L+1$ and will satisfy the requirements of this theorem. At the end of phase i we will ensure that the following invariants hold for LG_1 and LG_2.

(1) $X^{(i)} \cap Y^{(i)} = \emptyset$.

(2) for all $j \le i$, $|X^{(j)}| = |Y^{(j)}|$.

(3) for all $j \le i$, $|X_j| \ge \lfloor d|X_{j-1}| \rfloor$ and $|Y_j| \ge \lfloor d|Y_{j-1}| \rfloor$, where $d = (\Delta/4\lambda) - 1$.

(4) $X^{(i)} \subseteq L(V)$ and $Y^{(i)} \subseteq L(V)$.

The first invariant ensures that no vertex is a pseudo-boy and a pseudo-girl at the same time. Note that this condition implies that the two level graphs are totally disjoint for the following reasons. Suppose some vertex x lies in both LG_1 and LG_2. The first invariant guarantees that x cannot lie in the even-numbered levels of both LG_1 and LG_2. If x is present in the odd-numbered levels of both LG_1 and LG_2 then it must be matched to a vertex y which lies in the even-numbered levels of both LG_1 and LG_2. It is clear that this cannot happen as then y would violate the first invariant. Finally, if x is present in an even-numbered level of LG_1 and an odd-numbered level of LG_2, or vice versa, then it would imply the existence of an augmenting path for M of length at most $2i+1$. Thus, the invariant (1) implies that $U^{(2i)} \cap V^{(2i)} = \emptyset$.

The second invariant ensures that corresponding levels in the two level graphs are equal in size. The invariant (3) guarantees that the successive levels of the two level graphs expand rapidly in size. Finally, the last invariant states that the even-numbered levels of the two level graphs contain only large degree vertices. This condition will enable us to argue that the next level will also expand in size.

Since M is not a maximum matching in G, Fact 4 implies that there must exist an augmenting path for M. Let P be such an augmenting path and let it consist of the ordered sequence of vertices $v_1 v_2 \ldots v_k$. We will assume that v_1 and v_k are large degree vertices and use these as the starting points of the level graphs LG_1 and LG_2, respectively. If v_1 is not a large degree vertex then it must be the case that v_3 is a large degree vertex. Otherwise, the first four edges in P form a path of length four which has two small degree vertices, contradicting Lemma 5(b). We can then use v_3 as the starting point of the level graph LG_1. Similarly, if v_k is not a large degree vertex then v_{k-2} must be a large degree vertex and can be used as the starting point of LG_2.

We are now ready to describe the construction of the level graphs LG_1 and LG_2. Initially, we place only v_1 in U_0 and only v_k in V_0. It is easy to verify that the four invariant conditions hold at this point of the construction. Assume that the first $2i$ levels of the two level graphs have already been constructed and satisfy the four invariant conditions. We will describe the $(i+1)^{st}$ phase of the construction. The following definitions will be necessary for the description.

Definition 5: *A vertex v is called a* good *vertex if it is matched by M to a large degree vertex, otherwise it is called a* bad *vertex.*

Definition 6: *The set $GOOD \subseteq V$ is the set of all the good vertices in V.*

Suppose a vertex v in G is adjacent to two bad vertices, say x and y. Let a and b be the two small degree vertices which are matched to x and y, respectively. A simple case analysis now establishes the existence in G of a path of length 4 which contains v, a and b. But, by Lemma 5(b), no path of length 4 in G can contain more than one small degree vertex. This implies that v could not have been adjacent to two bad vertices and gives us the following claim.

Claim 10: *There is at most one bad vertex adjacent to each large degree vertex in G.*

Let S_1 denote the set of good vertices which are adjacent to the vertices in X_i and are not already present in LG_1 and LG_2, i.e. $S_1 = (\Gamma_G(X_i) \cap GOOD) \setminus (U^{(2i)} \cup V^{(2i)})$. Similarly, define $S_2 = (\Gamma_G(Y_i) \cap GOOD) \setminus (U^{(2i)} \cup V^{(2i)})$. The set S_1 contains vertices from which we will select the vertices for U_{2i+1} and the set S_2 contains vertices from which we will select the vertices for V_{2i+1}. It is easy to see that all edges between X_i and S_1, or between Y_i and S_2, must be free edges. The following claim gives a lower bound on the size of the sets S_1 and S_2. Note that since $|X_i| = |Y_i|$ the bounds on the cardinality of the two sets are identical.

Claim 11: *$|S_1| \geq 4d|X_i|$ and $|S_2| \geq 4d|Y_i|$, where $d = (\Delta/4\lambda) - 1$.*

We establish the validity of this claim for S_1. A similar argument works for S_2. First, observe that every vertex in X_i is a large degree vertex (by invariant (4)). Lemma 5(a) implies that $|\Gamma_G(X_i)| \geq \Delta|X_i|/\lambda$. Claim 10 states that there can be at most one bad vertex adjacent to each vertex in X_i. Therefore, we have that $|\Gamma_G(X_i) \cap GOOD| \geq (\Delta|X_i|/\lambda) - |X_i|$. If a vertex in X_i is adjacent to a pseudo-girl vertex in LG_2 then we have an augmenting path for M of length at most $2i+1$ and the construction terminates. Therefore, assume that the vertices in X_i are adjacent only to the pseudo-boy vertices in LG_2. From invariants (2) and (3) we conclude that there are at most $(1+2/d)|X_i|$ pseudo-boy vertices in LG_2. Moreover, there can be at most $(1+3/d)|X_i|$ vertices in the first $2i-1$ levels of LG_1. Thus, there are at most $(2+5/d)|X_i|$ vertices in $U^{(2i)} \cup V^{(2i)}$ which can be adjacent to the vertices in X_i. This implies that $|S_1| = |(\Gamma_G(X_i) \cap GOOD) \setminus (U^{(2i)} \cup V^{(2i)})| \geq ((\Delta/\lambda) - 3 - (5/d))|X_i|$. Since $\Delta \geq \ln n$ and $d = (\Delta/4\lambda) - 1$, we have the desired bound, viz. $|S_1| \geq 4d|X_i|$.

As observed earlier, it is possible that two vertices in S_1 (or, S_2) are matched to each other. Moreover, S_1 and S_2 may share vertices without implying the existence of a short augmenting path. We handle these two problems as follows. Let S_1' be any maximal subset of S_1 such that no two vertices in S_1' are matched to each other. Similarly, define S_2' to be a maximal subset of S_2 in which no two vertices are matched to each other. Then it is easy to see that the following claim is valid.

Claim 12: *$|S_1'| \geq |S_1|/2$ and $|S_2'| \geq |S_2|/2$.*

Let $T = S_1' \cap S_2'$ and define $T_1 = S_1' \setminus T$ and $T_2 = S_2' \setminus T$. Assume, without loss of generality, that $|T_1| \leq |T_2|$. Let $T_1' = T_1$ and T_2' be any subset of T_2 which contains exactly $|T_1'|$ elements. Further, divide T into two equal-sized subsets, say T_3 and T_4. If T contains an odd number of vertices then arbitrarily remove one element from T before partitioning it into T_3 and T_4. We place each vertex in $T_1' \cup T_3$ in U_{2i+1} and each vertex in $T_2' \cup T_4$ in V_{2i+1}. It is then clear that $|U_{2i+1}| \geq |S_1'|/2$ and that $|V_{2i+1}| \geq |S_2'|/2$. Claims 11 and 12 now imply that the following is true.

Claim 13: $|U_{2i+1}| = |V_{2i+1}| \geq \lfloor ((\Delta/\lambda) - 4)|X_i|/4 \rfloor = \lfloor d|X_i| \rfloor$.

Our construction of the $(2i+1)^{st}$ levels of the two level graphs guarantees that they are disjoint, equal in size and contain only good vertices. If there is a free vertex in either U_{2i+1} or V_{2i+1} then we already have a short augmenting path for M and the construction will terminate. Otherwise, we place in U_{2i+2} all those vertices in G which are matched to the vertices in U_{2i+1}, i.e. $U_{2i+2} = \Gamma_M(U_{2i+1})$; similarly, we place in V_{2i+2} all those vertices in G which are matched to the vertices in V_{2i+1}, i.e. $V_{2i+2} = \Gamma_M(V_{2i+1})$. Note that vertices which are matched to the vertices in U_{2i+1} or V_{2i+1} cannot already be present in either LG_1 or LG_2. This completes the description of the $(i+1)^{st}$ stage of the construction.

If there is a matching edge between U_{2i+1} and V_{2i+1} then we will already have a short augmenting path for M. Therefore, it must be the case that U_{2i+2} and V_{2i+2} are disjoint and of the same cardinality as U_{2i+1} and V_{2i+1}. Since U_{2i+1} and V_{2i+1} contain only good vertices, it must be the case that U_{2i+2} and V_{2i+2} contain only large degree vertices. Thus, we have established that at the end of the $(i+1)^{st}$ stage all four invariants are satisfied by LG_1 and LG_2.

Suppose the construction continues for $l = L/2$ stages without being terminated due to the discovery of a short augmenting path. Let $X = X^{(l)}$ and $Y = Y^{(l)}$ denote the set of pseudo-boys in LG_1 and pseudo-girls in LG_2, respectively. We have already shown that $X^{(0)}$ and $Y^{(0)}$ are non-empty. The invariant (4) implies that each subsequent $X^{(i)}$ and $Y^{(i)}$ must increase in size by at least a factor of d. This gives us the following claim for $L = c \ln n / \ln \Delta$, where c is a sufficiently large positive constant.

Claim 14: *$|X| = |Y| \geq n/\beta$, where β is some positive constant.*

From invariant (2) we have that X and Y must be disjoint. Lemma 5(c) then implies that there must be an edge in G which connects a vertex in X to a vertex in Y. This edge cannot be a matching edge as that would imply that both its end-points are present in both LG_1 and LG_2, violating the invariant (1). Any free edge between X and Y will complete an augmenting path for M of length at most $2L+1$. This concludes the proof. □

6.3. Matchings in Random Graphs

We have established that for any non-maximum matching in $G \in \mathcal{G}_n^{\Delta}$ there exists a short augmenting path. We now apply Theorem 11 to the analysis of the Micali-Vazirani algorithm [MV80] for non-bipartite matchings.

Theorem 12: *Let $G(V,E) \in \mathcal{G}_n^{\Delta}$. Then the Micali-Vazirani algorithm for non-bipartite matchings terminates in $O(m \log n / \log \Delta)$ time on input G.*

Proof: The Micali-Vazirani algorithm has the same general approach as Dinic's algorithm. It computes the maximum matching in G in a number of stages. In each stage it computes a maximal set of vertex-disjoint shortest augmenting paths with respect to the current non-maximum matching. These augmenting paths are computed in at most $O(m)$ time. The current matching is then simultaneously augmented by all the augmenting paths found in the current stage. The process terminates when no more augmenting paths can be found.

The analysis due to Hopcroft and Karp [HK73] has established that the length of the shortest augmenting paths must increase at each stage. This implies that the maximum length of a shortest augmenting path with respect to any non-maximum matching in G is a trivial upper bound on the total number of stages required to reach a maximum matching. Theorem 11 states that for every non-maximum matching in $G \in \mathcal{G}_n^{\Delta}$ there is an augmenting path of length $O(\log n / \log \Delta)$. This gives us the desired result. □

The next corollary follows from Theorem 12 and Lemma 5.

Corollary 12: *Let $G \in \mathcal{G}_{n,p}$, where $p > \ln n/(n-1)$. Then, with high probability, the Micali-Vazirani algorithm terminates in $O(m \log n / \log \Delta)$ time on input G. Moreover, the* expected *running time of this algorithm on input G is also $O(m \log n / \log \Delta)$.*

Considering the case where $p = 1/2$, we obtain the following corollary.

Corollary 13: *Let the input for the Micali-Vazirani algorithm be chosen uniformly at random from the set of all graphs with n vertices. Then the* expected *running time of this algorithm is $O(m)$.*

Acknowledgements

I would like to thank Dick Karp for getting me started on this research. I am also deeply indebted to him for his constant help and encouragement, as well as numerous technical discussions.

References

[AV79] D. Angluin and L. Valiant. Fast probabilistic algorithms for Hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, vol. 19, pages 155-193, 1979.

[Bol81] B. Bollobás. The diameter of random graphs. *Transactions of the American Mathematical Society*, vol. 267, pages 41–52, 1981.

[Bol85] B. Bollobás *Random Graphs.* Academic Press, 1985.

[BF85] B. Bollobás and A.M. Frieze. On Matchings and Hamiltonian Circuits in Random Graphs. *Annals of Discrete Mathematics*, vol. 28, pages 23–46, 1985.

[BFF85] B. Bollobás, T.I. Fenner and A.M. Frieze. An Algorithm for finding Hamilton Cycles in a Random Graph. In *Proceedings of the 17th ACM Symposium on Theory of Computing*, pages 430–439, 1985.

[Bro86] A.Z. Broder. How hard is it to marry at random? (On the approximation of the permanent). In *Proceedings of the 18th ACM Symposium on Theory of Computing*, pages 50–58, 1986.

[Che52] H. Chernoff. A Measure of Asymptotic Efficiency for Tests Based on the Sum of Observations. *Annals of Mathematical Statistics*, vol. 23, pages 493–509, 1952.

[DH79] U. Derigs and A. Heske. A Computational Study on Some Methods for Solving the Cardinality Matching Problem. Report 79/2, Mathematisches Institut der Universität zu Köln, 1979.

[Din70] E.A. Dinic. Algorithm for solution of a problem of maximum flow in networks with power estimation *Soviet Math. Doklady*, vol. 11, pages 1277–1280, 1970.

[ER59] P. Erdös and A. Rényi. On random graphs. *Publ. Math. Debrecen*, vol. 6, pages 290–297, 1959.

[ER64] P. Erdös and A. Rényi. On random matrices. *Publ. Math. Inst. Hungarian Academy of Sciences*, vol. 8, pages 455–461, 1964.

[ER66] P. Erdös and A. Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.*, vol. 17, pages 359–368, 1966.

[ER68] P. Erdös and A. Rényi. On random matrices II. *Studia Sci. Math. Hungar.*, vol. 3, pages 459–464, 1968.

[EK75] S. Even and O. Kariv. An $O(n^{2.5})$-Algorithm for Maximum Matching in General Graphs. In *Proceedings of 16th IEEE Symposium on Foundations of Computer Science*, pages 100–112, 1975.

[FM91] T. Feder and R. Motwani, Clique Partitions, Graph Compression and Speeding-up Algorithms. In *23rd Annual ACM Symposium on Theory of Computing*, pages 123–133, 1991.

[GT88] H. Gabow and R. Tarjan. Almost Optimum Speed-ups of Algorithms for Bipartite Matching and Related Problems. In *Proceedings of the 20th ACM Symposium on Theory of Computing*, pages 514–527, 1988.

[GH90] O. Goldschmidt and D.S. Hochbaum. A fast perfect-matching algorithm in random graphs. *SIAM Journal on Discrete Mathematics*, vol. 3, pages 48–57, 1990.

[Hal35] P. Hall. On representatives of subsets. *Journal of London Mathematics Society*, vol. 10, pages 26–30, 1935.

[Ham78] H. Hamacher. Numerical Investigations on the Maximal Flow Algorithm of Karzanov. Report 78-7, Mathematisches Institut der Universität zu Köln, 1978.

[HK73] J.E. Hopcroft and R.M. Karp. An $O(n^{5/2})$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, vol. 2, pages 225–231, 1973.

[JS88] M. Jerrum and A. Sinclair. Conductance and the Rapid Mixing Property for Markov Chains: the Approximation of the Permanent Resolved. In *Proceedings of the 20th ACM Symposium on Theory of Computing*, pages 235–244, 1988.

[KS81] R.M. Karp and M. Sipser. Maximum matchings in sparse random graphs. In *Proceedings of the 22nd IEEE Symposium on Foundations of Computer Science*, pages 364–375, 1981.

[KLMR85] R.M. Karp, J.K. Lenstra, C.J.H. McDiarmid and A.H.G. Rinnooy Kan. Probabilistic Analysis. In M. O'hEigeartaigh, J.K. Lenstra and A.H.G. Rinnooy Kan, editors, *Combinatorial Optimization: Annotated Bibliographies*, John Wiley & Sons, 1985.

[KK86] R.M. Karp and A.H.G. Rinnooy Kan. *Personal Communication*, 1986.

[KMN92] R.M. Karp, R. Motwani and N. Nisan. Probabilistic Analysis of Network Flow Algorithms. *Mathematics of Operations Research*, to appear, 1992.

[KS83] J. Komlós and E. Szemerédi. Limit Distribution for the Existence of Hamiltonian Cycles in a Random Graph. *Discrete Mathematics*, vol. 43, pages 55–63, 1983.

[LP86] L. Lovász and M.D. Plummer. *Matching Theory*. North-Holland, 1986.

[MV80] S. Micali and V. Vazirani. An $O(|V|^{0.5}|E|)$ Algorithm for Finding Maximum Matchings in General Graphs. In *Proceedings of the 21st IEEE Symposium on Foundations of Computer Science*, pages 17–27, 1980.

[Mot88] R. Motwani. *Probabilistic Analysis of Matching and Network Flow Algorithms.* Ph. D. Thesis, University of California at Berkeley, 1988.

[Mot89] R. Motwani. Expanding Graphs and Average-case Analysis of Algorithms for Matchings and Related Problems. In *Proceedings of the 21st ACM Symposium on Theory of Computing*, pages 550–561, 1989.

[Mot92] R. Motwani. Heuristic algorithms for k-factors. *Unpublished Manuscript*, 1992.

[Poh77] I. Pohl. Improvements to the Dinic-Karzanov Network Flow Algorithm. Technical Report No. HP-77-11-001, Information Sciences, University of California at Santa Cruz, 1977.

[Pos76] L. Pósa. Hamiltonian Circuits in Random Graphs. *Discrete Mathematics*, vol. 14, pages 359–364, 1976.

[SU81] E. Shamir and E. Upfal. On factors in random graphs. *Israel Journal of Mathematics*, vol. 39, pages 296–302, 1981.

[Tut81] W.T. Tutte. Graph Factors. *Combinatorica*, vol. 1, pages 79–97, 1981.

[Tut82] W.T. Tutte. The Method of Alternating Paths. *Combinatorica*, vol. 2, pages 325–332, 1982.

[Val79] L.G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, vol. 8, pages 189–201, 1979.

[Wal80] D.W. Walkup. Matchings in random regular bipartite graphs. *Discrete Mathematics*, vol. 31, pages 59–64, 1980.