A Simplified Realization of the Hopcroft-Karp Approach to Maximum Matching in General Graphs

Norbert Blum
Informatik IV, Universität Bonn
Römerstr. 164,D-53117 Bonn, Germany
email: blum@cs.uni-bonn.de

October 26, 1999

Abstract

In [2, 3], we have reduced the problem of finding an augmenting path in a general graph to a reachability problem in a directed, bipartite graph, and we have shown that a slight modification of depth-first search leads to an algorithm for finding such paths. This new point of view enables us to give a simplified realization of the Hopcroft-Karp approach for the computation of a maximum cardinality matching in general graphs. We show, how to get an $O(n+m)$ implementation of one phase leading to an $O(\sqrt{n}m)$ algorithm for the computation of a maximum cardinality matching in general graphs.

1 Introduction and motivation

In 1973, Hopcroft and Karp [9] proved the following fact. If one computes in one phase a maximal set of shortest augmenting paths, then $O(\sqrt{n})$ such phases would be sufficient. For the bipartite case they showed that a phase can be implemented by a breadth-first search followed by a depth-first search. This led to an $O(n + m)$ implementation of one phase, and hence, to an $O(\sqrt{n}m)$ algorithm for the computation of a maximum matching in bipartite graphs.

In 1975, Even and Kariv [5, 10] presented an $O(\min\{n^2, m \log n\})$ implementation of a phase, leading to an $O(\min\{n^{2.5}, \sqrt{n}m \log n\})$ algorithm for the computation of a maximum matching in general graphs. In 1978, Bartnik [1] gave an alternative $O(n^2)$ implementation in his unpublished Ph.D. thesis (see [8]). In 1980, Micali and Vijay Vazirani [11] have presented an $O(m+n)$ implementation of a phase without the presentation of a correctness proof. In 1994, Vijay Vazirani [13] provided a correctness proof. Gabow and Tarjan [7] gave another $O(m + n)$ implementation of a phase in 1989.

We have reduced the problem of finding an augmenting path to a reachability problem in a directed, bipartite graph [3]. Moreover, we have shown how to solve this reachability problem by a modified depth-first search. We will show, how to use this latter fact for getting a simplified $O(n + m)$ implementation of one phase leading to a simplified $O(\sqrt{n}m)$ algorithm for the computation of a maximum matching in general graphs.

We assume the reader's familiarity with [3] and we do not repeat any results presented in this paper.

In Section 2, the simplified realization of one phase is described. We will prove its correctnes in Section 3, and outline an efficient implementation of an entire phase in Section 4.

2 A simplified implementation of one phase

In the bipartite case, Hopcroft and Karp [9] have described an elegant, simple $O(m + n)$ implementation of an entire phase of the matching algorithm. Let us sketch this implementation. First they have reduced the problem of finding augmenting paths to a reachability problem. Then, by performing a breadth-first search (BFS) on G_M with start node s until the target node t is reached, they have obtained a layered and directed graph $\bar{G}_M$ for which the paths from s to t correspond exactly to the shortest M-augmenting paths in G. Using depth-first search, they find a maximal set of disjoint M-augmenting paths. Whenever an M-augmenting path is found, the path and all incident edges are deleted and the depth-first search is continued. Breadth-first search and depth-first search take $O(m + n)$ time. Hence, the implementation of Hopcroft and Karp has time complexity $O(m + n)$.

With respect to general graphs, the following question suggests itself. Can we get an implementation of an entire phase by performing something like breadth-first search followed by something like depth-first search? We will give an affirmative answer to this question.

Let $G = (V, E)$ be an undirected graph, M be a matching of G, and $G_M = (V', E_M)$ be the directed graph as defined in [3]. Our goal is to construct from G_M a layered and directed graph $\bar{G}_M = (V', \bar{E}_M)$ such that

1. the lth layer contains exactly those nodes $[v, X] \in V'$ such that a shortest strongly simple path from s to $[v, X]$ in G_M has length l, and

2. $\bar{G}_M$ contains all shortest strongly simple paths from s to t in G_M.

It is clear that s is the only node in Layer 0, i.e. level$(s) = 0$. Note that by the structure of G_M, $X = B$ ($X = A$) implies level$([v, X])$ is odd (even). Since breadth-first search (BFS) on G_M with start node s finds shortest simple distances from s, and not shortest strongly simple distances, BFS cannot be used directly for the construction of $\bar{G}_M$. But we can modify BFS such that the modified breadth-first search (MBFS) finds shortest strongly simple distances. Remember that for the construction of the $(l + 1)$th level, BFS needs only to consider the nodes in Level l, and to insert into the $(l + 1)$th level all nodes w which fulfill the following properties.

1. There is a node v in the lth level with $(v, w) \in E$.

2. Level(w) has not been defined.

With respect to finding strongly simple distances from s, the construction of the $(l+1)$th level is a bit more difficult. By the structure of G_M, the level of a non-free node $[w, B]$ is well-defined by the level of the unique node $[v, A]$ with $([v, A], [w, B]) \in E_M$. Hence, the construction of odd levels is trivial.

For odd l we will describe the construction of the $(l+1)$th level under the assumption that Levels $0, 1, 2, \ldots, l$ are constructed. It is clear that similar to BFS, MBFS has to insert into the $(l+1)$th level all nodes $[w, A] \in V'$ which fulfill the following properties.

1. There is a node $[v, B]$ in Level l with $([v, B], [w, A]) \in E_M$, and there is a strongly simple path from s to $[v, B]$ of length l which does not contain $[w, B]$.

2. Level($[w, A]$) has not been defined.

Hence, the first part of Round $l+1$ of MBFS is similar to BFS.

Part 1 of Round $l+1$ of MBFS:

After the construction of the lth level, l odd, all edges $([v, B], [w, A])$ with level($[v, B]$) $= l$ are considered. We distinguish three cases.

Case 1: Level($[w, A]$) $> l$, and there is a strongly simple path P from s to $[v, B]$ of length l not containing $[w, B]$.

MBFS inserts node $[w, A]$ into the $(l+1)$th level, and adds edge $([v, B], [w, A])$ to $\bar{E}_M$.

Case 2: Level($[w, A]$) $> l$, and all strongly simple paths from s to $[v, B]$ of length l contain $[w, B]$.

MBFS does not enlarge Level $l+1$.

Case 3: level($[w, A]$) $\leq l$.

MBFS does not enlarge Level $l+1$. □

But these are not all the nodes which MBFS has to insert into Level $l+1$. Consider the example described in Figure 1. Note that level($[v_7, B]$) $= 7$, but level($[v_3, A]$) $\neq 8$, since the unique shortest strongly simple path from s to $[v_7, B]$ contains $[v_3, B]$. The unique strongly simple path P from s to $[v_3, A]$ has length 14. Hence level($[v_3, A]$) $= 14$. Furthermore, level($[v_7, A]$) $= 4$, level($[v_6, B]$) $= 5$, level($[v_{17}, A]$) $= 6$, and so on. Moreover, level($[v_3, A]$) $=$ (level($[v_{17}, B]$)+level($[v_6, B]$)+1)−level($[v_3, B]$). P is found when level($[v_6, B]$) and level($[v_{17}, B]$) have been defined. Hence, our goal will be to find such paths P at the moment when for two adjacent nodes $[v_i, A], [v_{i+1}, B]$ or $[v_i, B], [v_{i+1}, A]$, level($[v_i, X]$) and level($[v_{i+1}, X]$) are defined, for $X \in \{A, B\}$.

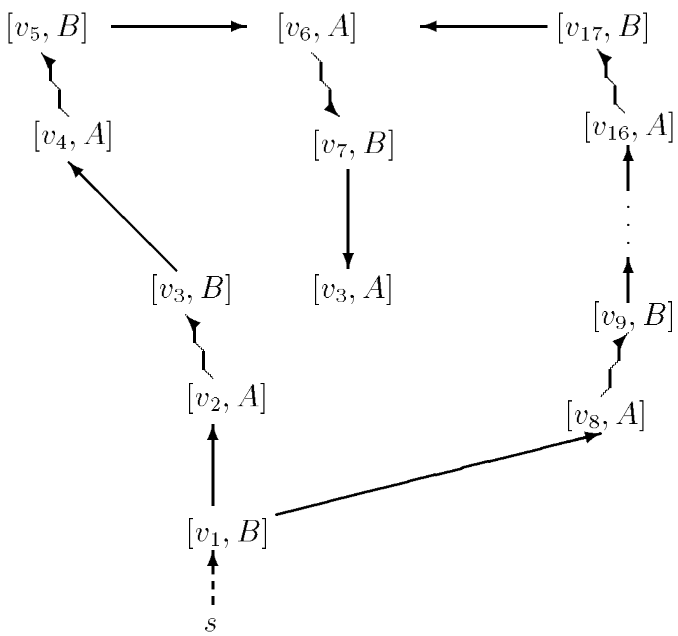

Figure 1:

Therefore, MBFS also has to insert nodes $[w, A] \in V'$ into Level $l+1$, for which there is a shortest strongly simple path $P = s, [v_1, B], \ldots, [v_l, B], [w, A]$ with level($[v_l, B]$) $< l$. Obviously, level($[w, B]$) $< l$. For the treatment of these nodes, the following notation is useful.

Let $T \subset V'$ such that level($[v, X]$) is defined, for all $[v, X] \in T$. We denote by DOM(T) the set of those nodes $[u, B] \in V'$, which satisfies:

a) All shortest strongly simple paths from s to $[v, X]$ contain $[u, B]$, for all $[v, X] \in T$;

b) Level($[u, A]$) has not been defined; and

c) Level($[w, B]$) $\leq$ level($[u, B]$) for all $[w, B] \in V'$ satisfying a) and b).

If node $[u, B]$ does not exist, then DOM(T) denotes node s. Furthermore, DOM(s) denotes node s. Note the fact that level($[u, B]$) is defined but level($[u, A]$) is not defined implies DOM($[u, B]$) $= \{[u, B]\}$. We will use DOM(T) only for subsets of V' of size at most 2. We will show that

$|\text{DOM}(T)| = 1$ later. Next, we will describe the second part of Round $l+1$ of MBFS, which is needed for even l, too.

Part 2 of Round $l+1$ of MBFS:

$\bar{G}_{l+1} = (V'_{l+1}, \bar{E}_{l+1})$ denotes the subgraph of $\bar{G}_M$, constructed after the termination of Part 1 of Round $l+1$. After the termination of Part 1, MBFS considers all pairs of nodes $[v, Z], [w, Z]$, $Z = B$ ($Z = A$) if l even (odd) such that:

i) $([v, Z], [w, \overline{Z}]) \in E_M$;

ii) level($[v, Z]$) and level($[w, Z]$) have been defined; and

iii) level($[v, Z]$) or level($[w, Z]$) is $l+1$.

Starting in $[v, Z]$, $[w, Z]$, MBFS performs a search on the backpaths of $\bar{G}_{l+1}$ until DOM($\{[v, Z], [w, Z]\}$) is reached. All visited nodes $[u, X]$ such that level($[u, X]$) has not been defined, and $[u, X] \neq$ DOM($\{[v, Z], [w, Z]\}$), are inserted into Level (level($[v, Z]$) + level($[w, Z]$) + 1 − level($[u, \overline{X}]$)). All traversed edges not in $\bar{G}_{l+1}$, and not incident to DOM($\{[v, Z], [w, Z]\}$), are added to $\bar{E}_{l+1}$. $\square$

A nice property of MBFS is that with respect to Part 1 of any round, it suffices to consider only edges $([v, B], [w, A])$ which have not been considered in Part 2 of a previous round. For the subsequence, we assume that we have changed MBFS in that way. I.e., during Part 1 of Round $l+1$ only edges $([v, B], [w, A])$ such that level($[v, B]$) $= l$ and edge $([v, B], [w, A])$ has not been traversed during an earlier round are considered.

3 The correctness proof of MBFS

Next we will prove the correctness of MBFS. First we will show $|\text{DOM}(T)| = 1$, for nonempty $T \subset V'$. This will be a simple conclusion from the following lemma.

Lemma 1 *Let $P = s, [v_1, B], [v_2, A], \ldots, [v_l, B]$ be a shortest strongly simple path from s to $[v_l, B]$, i.e. level($[v_l, B]$) $= l$. Let $[v_j, B] \in P$ be a node with level($[v_j, A]$) $> l$. Then for all odd $i < j$, level($[v_i, B]$) $<$ level($[v_j, B]$).*

Proof: Assume that the assertion does not hold. We will obtain a contradiction by the construction of a strongly simple path from s to $[v_j, A]$ of length less than l.

Consider $[v_i, B] \in P, i < j$ with level$([v_i, B]) \geq$ level$([v_j, B])$. Let $P = P_1, [v_i, B], P_2, [v_j, B], P_3$. Since level$([v_j, B]) \leq$ level$([v_i, B])$, there exists a strongly simple path $Q = Q_1, [v_j, B]$ from s to $[v_j, B]$ such that $|Q_1| \leq |P_1|$. Note that Q_1 and P_3 are not strongly disjoint. Otherwise, $R = Q_1, [v_j, B], P_3$ would be a strongly simple path from s to $[v_l, B]$ shorter than P. Let $[q, X]$ be the first node on Q_1 such that $[q, X] \in P_3$ or $[q, \overline{X}] \in P_3$. Let

$$Q_1 = Q_1', [q, X], Q_1'' \text{ and } P_3 = \begin{cases} P_3', [q, X], P_3'' & \text{if } [q, X] \in P_3 \\ P_3', [q, \overline{X}], P_3'' & \text{otherwise} \end{cases}$$

If $[q, X] \in P_3$, then $R = Q_1', [q, X], P_3''$ would be a strongly simple path from s to $[v_l, B]$ shorter than P. Hence, $[q, \overline{X}] \in P_3$. Consider $R = Q_1', [q, X], r(P_3'), [v_j, A]$. By the choice of $[q, X]$, path R is a strongly simple path from s to $[v_j, A]$. Furthermore,

$$\begin{aligned} |R| &= |Q_1'| + 2 + |P_3'| + 1 \\ &\leq |Q_1| + |P_3| + 1 \\ &< |P| \end{aligned}$$

This contradicts level$([v_j, A]) > l$. ∎

The following lemma is a simple consequence of Lemma 1.

Lemma 2 *Let $T \subseteq V'$, $T \neq \emptyset$ such that* level$([v, X])$ *is defined for all* $[v, X] \in T$. *Then the following statements hold true:*

a) $|\mathrm{DOM}(T)| = 1$.

b) Let $\mathrm{DOM}(T) = [u, B]$. *Then, after the definition of* level$([u, A])$, *always* $\mathrm{DOM}(T) = \mathrm{DOM}([u, B])$.

Proof: a) Assume that $|\mathrm{DOM}(T)| > 1$. Let $[u_1, B], [u_2, B] \in \mathrm{DOM}(T)$. By considering any shortest strongly simple path from s to $[v, X]$ for any $[v, X] \in T$, applying Lemma 1, we obtain level$([u_1, B]) <$ level$([u_2, B])$ or level$([u_2, B]) <$ level$([u_1, B])$, a contradiction.

b) is obvious by the definition of $\mathrm{DOM}(T)$. ∎

We say that, a path P *is constructed* by MBFS if all edges on P are added to $\bar{E}_M$. The correctness proof for MBFS is a direct conclusion of the following lemma.

Lemma 3 *MBFS maintains the following invariants:*

Invariant 1: Case 2 of the algorithm MBFS never occurs.

Invariant 2: For all $[u,X],[u,\overline{X}] \in V'$, $\text{level}([u,X]) < \text{level}([u,\overline{X}]$ *the following holds true:*
Level$([u,X])$ *has been defined, and all corresponding shortest strongly simple paths have been constructed after the termination of Part 1 of Round* $\text{level}([u,X])$. *Level*$([u,\overline{X}])$ *has been defined, and all corresponding shortest strongly simple paths have been constructed after the termination of Part 2 of Round* l, *where* $l = 1/2(\text{level}([u,B])+\text{level}([u,A])-1)$.

Invariant 3: If, during Part 2 of Round l, $\text{level}([u,X])$ *is defined according to the consideration of the pair of nodes* $[v,Z],[w,Z]$, *then* $\text{level}([v,Z]) = \text{level}([w,Z]) = l$.

Invariant 4: For all $[u,X] \in V'$, $\text{level}([u,X])$ *is computed correctly.*

Proof: Consider the first situation in which one of the four invariants is not maintained. Four cases are to be considered.

Case 1: Invariant 1 is not maintained.

Assume that $[v,B]$ has been inserted into Level $l-1$, edge $([v,B],[w,A])$ is considered during Part 1 of Round l, and Case 2 occurs with respect to the consideration of this edge. Note that $\text{level}([v,B]) = l-1$.

Consider any shortest strongly simple path $P = P_1,[w,B],P_2,[x,A],[v,B]$ from s to $[v,B]$. Since Case 2 of MBFS is fulfilled, the node $[w,B]$ must be on P. Consider path $P' = P_1,[w,B],[v,A]$. By construction, P' is strongly simple. Hence, $\text{level}([v,A]) \leq |P'| < l-1$. We distinguish two cases.

Case i: $([w,B],[v,A]) \in \bar{E}_M$.

Since, $\text{level}([x,A]) < l-1$, and $([x,A],[v,B])$ is an edge in E', edge $([v,B],[w,A])$ has been considered during Part 2 of a previous round. Hence, this edge cannot be considered during Part 1 of Round l, a contradiction.

Case ii: $([w, B], [v, A]) \notin \bar{E}_M$.

Then there is a strongly simple path $Q = Q', [v, A]$ from s to $[v, A]$ with $|Q| < |P'|$. Consider

$$R = Q', [v, A], [x, B], r(P_2), [w, A].$$

By construction $|R| < |P, [w, A]|$. Since we consider the first situation in which one of the four invariants is not maintained, the following holds true:

Q' and $P_2, [x, A]$ are not strongly disjoint.

Let $[z, X]$ be the first node on Q' such that $[z, X]$ or $[z, \overline{X}]$ is on $P_2, [x, A]$. Let

$$Q' = Q'_1, [z, X], Q'_2 \text{ and } P_2, [x, A] = \begin{cases} P'_2, [z, X], P''_2 & \text{if } [z, X] \text{ on } P_2, [x, A] \\ P'_2, [z, \overline{X}], P''_2 & \text{otherwise} \end{cases}$$

If $[z, X] \in P_2, [x, A]$ then $Q'_1, [z, X], P''_2, [v, B]$ would be a shorter strongly simple path from s to $[v, B]$ than P. Hence $[z, \overline{X}] \in P_2, [x, A]$. Consider

$$R = Q'_1, [z, X], r(P'_2), [w, A].$$

R is a strongly simple path from s to $[w, A]$. Furthermore, $|R| < |P|$. But this contradicts the construction. Hence, Invariant 1 cannot be the first invariant which is not maintained by MBFS.

Case 2: Invariant 2 is not maintained.

If Invariant 2 is not maintained for node $[u, B]$, then Invariant 2 is also not maintained for that node $[v, A]$ with $([v, A], [u, B]) \in E'$. Hence, it suffices to consider case $X = A$. Let P be any shortest strongly simple path from s to $[u, A]$. We will show that P is constructed by MBFS not later than during

$$\begin{cases} \text{Part 1 of Round } l & \text{if level}([u, A]) < \text{level}([u, B]) \\ \text{Part 2 of Round } l & \text{if level}([u, A]) > \text{level}([u, B]) \end{cases}$$

where $l = \min\{\text{level}([u, A]), 1/2(\text{level}([u, B]) + \text{level}([u, A]) - 1)\}$.

Let $[x, B]$ be the direct predecessor of $[u, A]$ on P. If level$([u, A]) <$ level$([u, B])$, then level$([x, B]) =$ level$([u, A]) - 1$. Since Invariant 2 has been maintained in previous rounds, edge $([x, B], [u, A])$ is considered during Part 1 of Round l. The edge $([x, B], [u, A])$ is inserted into $\bar{E}_M$, and P is constructed, a contradiction. Hence, level$([u, A]) >$ level$([u, B])$.

Let $P = P_1, [y, X], [z, \overline{X}], P_2, [u, A]$. Starting in $[u, A]$, we follow $r(P)$ until, for node $[z, X]$ which is reached, one of the following two cases arises.

a) There is a shortest strongly simple path Q from s to $[u, B]$ such that $Q, r(P_2), [z, X]$ is not strongly simple.

b) For all shortest strongly simple paths Q from s to $[u, B]$, path $R = Q, r(P_2), [z, X]$ is strongly simple, and has length $1/2(\text{level}([u, B]) + \text{level}([u, A]) - 1)$.

Case a)

If Case a) is fulfilled, then $X = A$. Since $Q, r(P_2), [z, A]$ is not strongly simple, but $Q, r(P_2)$ is strongly simple, path Q contains node $[z, B]$. Let $Q = Q_1, [z, B], Q_2$. By construction, $R = Q_1, [z, B], P_2, [u, A]$ is strongly simple. Moreover, $|R| < 1/2(\text{level}([u, A]) + \text{level}([u, B]) - 1) < \text{level}([u, A]$, a contradiction.

Case b)

Consider any shortest strongly simple path Q from s to $[u, B]$. By construction, $R = Q, r(P_2), [z, X]$ is strongly simple. By assumption, the invariant has been maintained in previous rounds. Hence, the level of every node on $r(P_2)$ has been defined. Furthermore, $\text{level}([r, B]) < \text{level}([r, A])$, for all nodes $[r, A]$ on $r(P_2)$ with larger distance from s on R than $\text{level}([r, A])$. Hence, the ingoing edges of such a node $[r, A]$ on $r(P_2)$ have been added to $\bar{E}_M$ during Part 2 of a previous round. Hence, path R has been constructed before the termination of Part 1 of Round l. Analogously, the levels of all nodes on $P_1, [y, X]$ have been computed before the termination of Part 1 of Round l, and path $P_1, [y, X]$ has been constructed, too. Therefore, path P has been constructed after the termination of Part 2 of Round l.

Case 3: Invariant 3 is not maintained.

By the construction of the algorithm MBFS, at least one of $[v, Z], [w, Z]$ has level l. W.l.o.g., let $\text{level}([v, Z]) = l$.

If $\text{level}([w, Z]) < l$, then $\text{level}([u, A]) + \text{level}([u, B]) = \text{level}([v, Z]) + \text{level}([w, Z]) + 1 < 2l + 1$. By assumption, Invariant 2 has been maintained in previous rounds. Hence, $\text{level}([u, X])$ has been defined in a previous round, a contradiction.

If $\text{level}([w, Z]) > l$, then $\text{level}([w, Z])$ has been computed during Part 2 of a previous Round l'. Since Invariant 3 has been maintained in previous rounds, $\text{level}([w, Z]) = 2l' + 1 - \text{level}([w, \overline{Z}])$. Hence, $\text{level}([w, \overline{Z}]) \leq 2l' + 1 -$

$(l+2) < l-1$. Consider any shortest strongly simple path P from s to $[w, \overline{Z}]$. Let $Q = P, [v, Z]$. Since $|Q| < l$ we obtain $[v, \overline{Z}] \in P$. Let $P = P_1, [v, \overline{Z}], P_2$. By construction, $R = P_1, [v, \overline{Z}], [w, Z]$ would be a strongly simple path from s to $[w, Z]$ shorter than l, a contradiction. Alltogether, we have obtained $\text{level}([w, Z]) = l$.

Case 4: Invariant 4 is not maintained.

If Invariant 4 is violated for node $[u, B]$, then Invariant 4 is also violated for that node $[v, A]$ with $([v, A], [u, B]) \in E'$. Hence, it suffices to consider case $X = A$.

Assume $\text{level}([u, A]) < \text{level}([u, B])$. Let $[x, B]$ be the direct predecessor of $[u, A]$ on any shortest strongly simple path from s to $[u, A]$. By the assumption that Invariant 2 and Invariant 4 have been maintained previously, $\text{level}([x, B])$ has been computed correctly after the termination of Part 1 of Round $\text{level}([x, B])$, or after the termination of Part 2 of Round $1/2(\text{level}([x, B]) + \text{level}([x, A]) - 1)$. In both cases, edge $([x, B], [u, A])$ has been considered in a previous round, and therefore, $\text{level}([u, A])$ has been computed correctly, a contradiction.

Hence, $\text{level}([u, A]) > \text{level}([u, B])$. Let $\text{level}([u, B])$ be maximal such that $\text{level}([u, A])$ is computed incorrectly during Part 2 of Round l. Let $[v, Z], [w, Z]$ be the corresponding pair of nodes. Then the following properties are fulfilled.

1. $\text{DOM}(\{[v, Z], [w, Z]\}) \neq [u, B]$.
2. MBFS inserts $[u, A]$ into Level $2l + 1 - \text{level}([u, B])$, but all strongly simple paths from s to $[u, A]$ are strictly longer.

We will show that Property 2 cannot be fulfilled. By Properties 1 and 2, there is a shortest strongly simple path from s to $[v, Z]$, containing $[u, B]$, and there is a shortest strongly simple path from s to $[w, Z]$, not containing $[u, B]$, or vice versa. W.l.o.g., let $P = P_1, [u, B], P_2, [v, Z]$ be a shortest strongly simple path from s to $[v, Z]$, and let $Q, [w, Z]$ be a shortest strongly simple path from s to $[w, Z]$, not containing $[u, B]$. By Property 2, P_2 and Q cannot be strongly disjoint. Otherwise, $R = Q, [w, Z], [v, \overline{Z}], r(P_2), [u, A]$ would be a strongly simple path from s to $[u, A]$ of length $2l+1-\text{level}([u, B])$.

Let $[q,X]$ be the first node on Q with $[q,X]$ or $[q,\overline{X}]$ on P_2. Let

$$Q = Q_1, [q,X], Q_2 \text{ and } P_2 = \begin{cases} P_{21}, [q,X], P_{22} & \text{if } [q,X] \text{ on } P_2 \\ P_{21}, [q,\overline{X}], P_{22} & \text{otherwise} \end{cases}$$

If $[q,\overline{X}]$ is on P_2, then $R = Q_1, [q,X], r(P_{21}), [u,A]$ would be a strongly simple path from s to $[u,A]$, shorter than $2l+1-\text{level}([u,B])$. Hence, $[q,X]$ is on P_2. $[q,X]$ is the *reference node*, and $[u,B]$ is the *critical node*. The path P_2 is the *critical path*, and the path Q_1 is the *reference path*.

Since $[u,A]$ is a node violating Invariant 4 with maximal $\text{level}([u,B])$, $\text{level}([p,\overline{Y}])$ has been computed correctly for all nodes $[p,Y]$ on P_2. Moreover, $\text{level}([p,Y]) + \text{level}([p,\overline{Y}]) \leq 2l+1$.

Our goal is the following. We will construct a situation such that the only case not leading to a contradiction will produce a new critical node and a new reference node such that

1. $\text{level}([u',B]) + \text{level}([u',A]) > 2l+1$, where $[u',B]$ is the critical node;
2. $\text{level}([q',X]) + \text{level}([q',\overline{X}]) \leq 2l+1$, where $[q',X]$ is the reference node;
3. for all nodes $[p,Y]$ on the critical path, $\text{level}([p,Y]) + \text{level}([p,\overline{Y}]) \leq 2l+1$;
4. the critical path is the end of a shortest strongly simple path to a node; and
5. the length of the reference path decreases strictly.

Starting with a critical node and a reference node fulfilling Properties 1–4, we will obtain a critical node and a new reference node fulfilling Properties 1–5. During the construction, we will only use Properties 1–4. Since the length of the reference path decreases strictly, after a finite number of such constructions, the case described by Properties 1–5 cannot happen, leading to a contradiction.

In the beginning, Properties 1–4 are fulfilled with respect to the critical node, the reference node, the critical path, and the reference path defined above. Let $R, [q,\overline{X}]$ be any shortest strongly simple path from s to $[q,\overline{X}]$. We distinguish two cases.

Case 1: $[u,B]$ is on R.

Consider $[u',B] \in R$ such that

a) $\text{level}([u',B]) + \text{level}([u',A]) > 2l+1$, and

b) $[u',B]$ is the node on R nearest to $[q,\overline{Z}]$ such that a) is fulfilled.

Since $[u,B] \in R$ fulfills a), node $[u',B]$ exists. Let $R = R_1, [u',B], R_2$. By Property a), R_2 and Q_1 cannot be strongly disjoint. Otherwise, $S = Q_1, [q,X], r(R_2), [u',A]$ would be a strongly simple path from s to $[u',A]$ shorter than $2l+1-\text{level}([u',B])$, a contradiction. Let $[q',X]$ be the first node on Q_1 with $[q',X]$ or $[q',\overline{X}]$ on R_2. Similarly to above, we can exclude $[q',\overline{X}] \in R_2$. Let

$$Q_1 = Q_{11}, [q',X], Q_{12} \text{ and } R_2 = R_{21}, [q',X], R_{22}.$$

Next we will define the new critical node and the new reference node, leading to a new critical path and a new reference path.

Let $[u',B]$ be the critical node, and let $[q',X]$ be the new reference node. Then, R_2 is the new critical path, and Q_{11} is the new reference path. By construction, it is easy to see that Properties 1–5 are fulfilled.

Case 2: $[u,B]$ is not on R.

Then R and P_{21} cannot be strongly disjoint. Otherwise,

$$S = R, [q,\overline{X}], r(P_{21}), [u,A]$$

would be a strongly simple path from s to $[u,A]$ of length at most $2l+1-\text{level}([u,B])$, a contradiction.

Let $[q',X]$ be the first node on P_{21} with $[q',X]$ or $[q',\overline{X}]$ on R. Similarly to above, we can exclude $[q',\overline{X}] \in R$. Let

$$P_{21} = P'_{21}, [q',X], P''_{21} \text{ and } R = R_1, [q',X], R_2.$$

Node $[u,B]$ remains the critical node and path P_2 remains the critical path, too. Let $[q',X]$ be the new reference node, and let R_1 be the new reference path. By construction, it is easy to see that properties 1–4 are fulfilled. It remains to prove that the new reference path is strictly shorter than the old reference path. We will prove the following assertions.

1. $|P_1, [u, B], P_{21}| \leq |Q_1|$, and

2. $|R_1| \leq |P_1, [u, B], P'_{21}|$.

Both assertions imply $|R_1| < |Q_1|$ directly.

The first assertion follows directly from the observation that the path $Q_1, [q, X], P_{22}, [v, Z]$ is a strongly simple path from s to $[v, Z]$ and the property that P is a shortest strongly simple path from s to $[v, Z]$.

For proving the second assertion, assume $|R_1| > |P_1, [u, B], P'_{21}|$. Consider

$$S = P_1, [u, B], P'_{21}, [q', X], R_2, [q, \overline{X}].$$

Since $|S| < |R, [q, \overline{X}]|$, and $R, [q, \overline{X}]$ is a shortest strongly simple path from s to $[q, \overline{X}]$, the paths P_1 and R_2 cannot be strongly disjoint. Let $[p, Y]$ be the first node on P_1 with $[p, Y]$ or $[p, \overline{Y}]$ on R_2. Let

$$P_1 = P_{11}, [p, Y], P_{12} \text{ and } R_2 = \begin{cases} R_{21}, [p, Y], R_{22} & \text{if } [p, Y] \text{ on } R_2 \\ R_{21}, [p, \overline{Y}], R_{22} & \text{otherwise} \end{cases}$$

If $[p, Y] \in R_2$, then $P_{11}, [p, Y], R_{22}, [q, \overline{X}]$ would be a strongly simple path from s to $[q, \overline{X}]$ shorter than $R, [q, \overline{X}]$. Hence $[p, \overline{Y}] \in R_2$. Consider

$$S = P_{11}, [p, Y], r(R_{21}), [q', \overline{X}], r(P'_{21}), [u, A].$$

By construction, S is a strongly simple path from s to $[u, A]$ shorter than $2l + 1 - \text{level}([u, B])$, a contradiction.

Altogether, we have obtained $|R_1| \leq |P_1, [u, B], P'_{21}|$. ∎

From Invariant 2 and Invariant 4, we obtain the following theorem directly.

Theorem 1 *MBFS computes correctly the graph* $\bar{G}_M = (V', \bar{E}_M)$.

4 An implementation of an entire phase

First we will describe the implementation of MBFS, and then show how to combine MBFS and MDFS for an implementation of an entire phase.

Obviously, Part 1 of all rounds can be implemented in such a way that the total time is bounded by $O(m + n)$. For the implementation of Part 2 of all rounds, we have to describe how to implement the search on the backpaths, starting in $[v, Z]$ and $[w, Z]$, until $\text{DOM}(\{[v, Z], [w, Z]\})$ is reached. Most importantly, since we do not know $\text{DOM}(\{[v, Z], [w, Z]\})$ in advance, meaning that $\text{DOM}(\{[v, Z], [w, Z]\})$ has to be computed simultaneously, we have to take care that the search does not continue beyond $\text{DOM}(\{[v, Z], [w, Z]\})$. This can be done in the following way.

1. In the front of the search, continue the search always in a node $[u, X]$ such that there is another node $[p, Y]$ in the front of the search with $\text{level}([p, \overline{Y}] \leq \text{level}([u, \overline{X}])$.

2. If the front of the search contains only one node, then stop the search.

Note that at the moment when Property 2 is fulfilled, $\text{DOM}(\{[v, Z], [w, Z]\})$ is reached on all backpaths, starting in $[v, Z]$ or $[w, Z]$. The search can be organized by performing simultaneously a depth-first search on the backpaths, starting in $[v, Z]$, and a depth-first search on the backpaths starting in $[w, Z]$. In each step, one of the depth-first searches is continued such that with respect to its head $[u, X]$ there holds that $\text{level}([u, \overline{X}]) \geq \text{level}([p, \overline{Y}])$, where $[p, Y]$ is the head of the other depth-first search. If the heads of the two depth-first searches become equal, then a backtrack with respect to one of the two depth-first searches is performed, which obtains a new head. If the heads of both depth-first searches are equal, and no backtrack is possible, then the search is terminated.

With respect to the efficiency, at the moment when the search meets a node $[u, X]$ for which $\text{level}([u, X])$ has been defined, we have to efficiently compute the next node of the search having the property that its level is not defined. By definition, this node is $\text{DOM}([u, X])$. As a consequence of Lemma 6, we can maintain these nodes by disjoint set union such that for the computation of $\text{DOM}([u, X])$ one *find* operation would suffice. Using incremental tree set union [6], we obtain a total time bound of $O(m + n)$ for the computation of the next node during the search such that its level is not defined.

The search procedure described above is exactly the double depth-first search described by Micali and Vazirani [11].

Note that edges, found during the search on backpaths in Part 2 of a round, do not need to be inserted explicitly into $\bar{E}_M$, since the corresponding reversed edge is already in $\bar{E}_M$. We only have to mark the nodes $[v, \overline{Z}]$ and $[w, \overline{Z}]$ as start nodes of a backward search.

As a consequence of Invariant 3 in Lemma 3, we can combine Part 2 of a round with Part 1 of the subsequent round.

Next, we will combine MBFS and MDFS for the implementation of an entire phase. Note that also with respect to $\bar{G}_M$, a simple path from s to t must not be strongly simple. Hence, we cannot compute a maximal set of up to s and t pairwise disjoint strongly simple paths from s to t using DFS.

Knowing $\bar{G}_M$, it is easy to compute a maximal set of shortest strongly simple paths of $\bar{G}_M$ using MDFS in $O(m+n)$ time. Every time a strongly simple path P from s to t is found, all nodes $[v, A]$, $[v, B]$ with $[v, A] \in P$ or $[v, B] \in P$ and all incident edges are deleted from $\bar{G}_M$. If a node becomes zero indegree or zero outdegree, then this node, too, and all incident edges, are deleted. Altogether, we have obtained the following theorem.

Theorem 2 *A maximum matching in a general graph $G = (V, E)$ can be computed in $O(\sqrt{n}(m+n))$ time and $O(m+n)$ space, where $|V| = n$ and $|E| = m$.*

Acknowledgment: I thank Andreas Decker, Marek Karpinski and Henning Rochow for helpful discussions and valuable hints.

References

[1] Bartnik G.W., Algorithmes de couplages dans les graphes, Thése Doctorat 3^e cycle, Université Paris VI (1978).

[2] Blum N., A new approach to maximum matching in general graphs, *17th ICALP* (1990), LNCS 443, 586–597.

[3] Blum N., Maximum matching in general graphs without explicit consideration of blossoms, Research report, Universität Bonn (1999), available at www.cs.uni-bonn.de/IV/blum/.

[4] Edmonds J., Paths, trees, and flowers, *Canad. J. Math*, **17** (1965), 449–467.

[5] Even S, and Kariv O., An $O(n^{2.5})$ algorithm for maximum matching in general graphs, *16th FOCS* (1975), 100–112.

[6] Gabow H. N., and Tarjan R. E., A linear-time algorithm for a special case of disjoint set union, *J. Comput. Syst. Sci.*, bf 30 (1985), 209–221.

[7] Gabow H. N., and Tarjan R. E., Faster scaling algorithms for general graph-matching problems, *J. ACM* **38** (1991), 815–853.

[8] Gondran M., and Minoux M., *Graphs and Algorithms*, Wiley & Sons, (1984).

[9] Hopcroft J. E., and Karp R. M., An $n^{5/2}$ algorithm for maximum matching in bipartite graphs, *SIAM J. Comput.*, **2** (1973), 225–231.

[10] Kariv O., An $O(n^{2.5})$ algorithm for maximum matching in general graphs, Ph.D. thesis, Dept. of Applied Mathematics, Weizmann Institute of Science, Rehovort, Israel (1976).

[11] Micali S., and Vazirani V. V., An $O(\sqrt{|V|} \cdot |E|)$ algorithm for finding maximum matching in general graphs, *21st FOCS* (1980), 12–27.

[12] Tarjan J. E., *Data Structures and Network Algorithms*, SIAM (1983).

[13] Vazirani V. V., A theory of alternating paths and blossoms for proving correctness of the $O(\sqrt{V}E)$ general graph maximum matching algorithm, *COMBINATORICA* **14** (1994), 71–109.