Jagiellonian University
Department of Theoretical Computer Science

Rafał Kilar

Comparison of maximum weight matching algorithms on general graphs

Master's Thesis

Supervisor: dr hab. inż. Krzysztof Turowski

September 2024

Contents

1 Introduction **2**
- 1.1 Definitions . . . 2
- 1.2 Maximum matching algorithms . . . 5

2 The Blossom Algorithm **8**
- 2.1 Edmonds' blossom algorithm . . . 8
  - 2.1.1 Augmenting path search . . . 12
  - 2.1.2 Complexity . . . 16
  - 2.1.3 Numerical accuracy . . . 17
  - 2.1.4 Correctness . . . 18
  - 2.1.5 Finding a maximum weight perfect matching . . . 19
- 2.2 Gabow's algorithm . . . 20
- 2.3 Galil, Micali & Gabow algorithm . . . 21
  - 2.3.1 Data structures . . . 21
  - 2.3.2 Search procedure . . . 24
  - 2.3.3 Complexity . . . 26

3 Scaling algorithms **27**
- 3.1 Gabow's scaling algorithm . . . 27
- 3.2 Path procedure . . . 29
  - 3.2.1 The path augmentation procedure . . . 30
  - 3.2.2 The shell search procedure . . . 31
- 3.3 Shell search implementation . . . 33
- 3.4 Splitting list . . . 42
- 3.5 Complexity . . . 46
- 3.6 Correctness . . . 50

4 Experimental results **51**
  - 4.0.1 Conclusions . . . 56

Chapter 1

Introduction

The graph matching problem is among the most extensively researched topics in combinatorial optimization. Initial studies into the subject were motivated by practical issues including minimization of transportation costs [25] or optimally assigning personnel to tasks [33]. Over time, matching algorithms have found their use in scheduling, approximation algorithms and network switching among other problems. They play a crucial role in various other optimization algorithms including undirected shortest paths [27], planar maximum cut [24], metric traveling salesman problem [3], and Chinese postman tours [10].

1.1 Definitions

We first give some basic definitions in graph theory borrowed from [2].

Definition 1.1 (graph)**.** *A* graph G *is an ordered pair of disjoint sets* (V, E) *such that* E *is a subset of the set* $\binom{V}{2}$ *of unordered pairs of* V*.*

We only consider finite graphs, that is, V and E are always finite. The set V is the set of *vertices* and E is the set of *edges*. If G is a graph, then $V = V(G)$ is the *vertex set* of G, and $E = E(G)$ is the *edge set* of G. If v is a vertex of G we will sometimes write $v \in G$ instead of $v \in V(G)$.

We denote $n_G = |V(G)|$ and $m_G = |E(G)|$ for a graph $G = (V, E)$. We will drop the subscripts for brevity when G is clear from the context.

An edge xy is said to *join* the vertices x and y Thus, xy and yx mean exactly the same edge, the vertices x and y are the *endvertices* of this edge. If $xy \in E(G)$, then x and y are *adjacent*, *neighboring* or *connected*, and the vertices x and y are *incident* with the edge xy. Two edges are *adjacent* if the have exactly one common endvertex.

Definition 1.2 (subgraph)**.** *We say that* $G' = (V', E')$ *is a* subgraph *of* $G = (V, E)$ *if* $V' \subseteq V$ *and* $E' \subseteq E$*. In this case we write* $G' \subseteq G$*.*

If G' *contains all edges of* G *that join two vertices in* V' *then* G' *is said to be a subgraph* induced *or* spanned *by* B' *and is denoted* $G[V']$*.*

For a subgraph $G' = G[W]$ spanned by a vertex subset $W \subseteq V(G)$, we write $n_W = |W|$ and $m_W = |E(G')|$.

If $W \subseteq V(G)$, then $G - W = G[V \setminus W]$ is the subgraph of G obtained by deleting the vertices of W and all edges incident with them. Similarly, if $E' \subseteq E(G)$, then $G - E' = (V(G), E(G) \setminus E')$. If $W = \{w\}$ and $E' = \{xy\}$ for some vertex $w \in V(G)$ and edge $xy \in E(G)$, then the notation is simplified to $G - w$ and $G - xy$ respectively. Similarly, if x and y are nonadjacent vertices of G, then $G + xy = (V(G), E(G) \cup \{xy\})$.

Definition 1.3 (path)**.** *A* path *is a graph P of the form*

$$V(P) = \{x_0, x_1, \ldots, x_l\}$$
$$E(P) = \{x_0x_1, x_1x_2, \ldots, x_{l-1}x_l\}$$

This path P is usually denoted by $x_0x_1 \ldots x_l$. The vertices x_0 and x_l are the *ends* of P and the value $l = |E(P)|$ is the *length* of P. We say that P goes from x_0 to x_l.

Definition 1.4 (connected graph)**.** *A graph is* connected *if for every pair $\{x, y\}$ of distinct vertices there is a path from x to y.*

A maximal connected subgraph is a *component* of a graph.

Definition 1.5 (cycle)**.** *A* cycle *is a graph C of the form*

$$V(C) = \{x_0, x_1, \ldots, x_l\}$$
$$E(C) = \{x_0x_1, x_1x_2, \ldots, x_{l-1}x_l, x_lx_0\}$$

This cycle C is denoted by $x_0x_1 \ldots x_lx_1$. The value $l+1 = |E(C)| = |V(C)|$ is the *length* of C.

Definition 1.6 (forest, tree)**.** *A graph without any cycles is a* forest*, or an* acyclic *graph. A* tree *is connected forest.*

Definition 1.7 (bipartite graph)**.** *A graph G is a* bipartite *graph with vertex classes V_1 and V_2 if $V(G) = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$ and every edge joins a vertex of V_1 to a vertex of V_2.*

A simple observation is that a graph is bipartite if and only if it does not contain an odd-length cycle.

A set of vertices (edges) is *independent* if it has no two adjacent elements.

Definition 1.8 (matching)**.** *A set of independent edges is called a* matching*. A matching M is* perfect *if every vertex is adjacent to exactly one edge in M.*

Definition 1.9 (exposed vertices)**.** *A vertex v is* exposed *for a matching M if it is not adjacent to any edge in M. If a vertex is not exposed it is* matched*.*

If a vertex v is matched in a matching M, then we call the vertex u, such that $uv \in M$, the *mate* or *matched vertex* of v and the edge uv the *matched edge* of v.

Definition 1.10 (alternating path)**.** *A path $P = x_0x_1 \dots x_l$ is alternating for a matching M if for each $i \in \{0, \dots, l-2\}$, $x_ix_{i+1} \in M$ if and only if $x_{i+1}x_{i+2} \notin M$.*

Definition 1.11 (augmenting path)**.** *An alternating path $x_0x_1 \dots x_l$ is* augmenting *if the vertices x_0 and x_l are both exposed.*

Definition 1.12 (weighted graph)**.** *A* weighted *graph is a graph $G = (V, E)$ along with a* weight function *$w : E \to \mathbb{R}$, which assigns a real valued* weight *$w(e)$ to each edge of G.*

For a set of edges $S \subseteq E$, we define the *weight* of S to be $w(S) = \sum_{e \in S} w(e)$.

We denote $N_G = \max_{e \in E} w(e)$ for a graph $G = (V, E)$, which we shorten to N when G is clear from the context. In this work, we consider only graphs with non-negative weights.

The algorithms for the maximum matching problems are usually divided into groups based on the classes of graphs they operate on and whether the graph is weighted. The graphs can be either bipartite or non-bipartite. When the graphs are unweighted, the algorithms tries to find matchings with maximum number of edges. When they are weighted, a matching with maximum possible weight is sought. In the case of weighted graphs we can also restrict our search to perfect matching, looking for the one with the highest weight. In this work we consider the following variants of the maximum matching problem on general graphs:

- Maximum Cardinality Matching (MCM) Find a matching in a graph G with maximum number of edges,
- Maximum Weight Matching (MWM) Find a matching in a weighted graph G with maximum weight,
- Maximum Weight Perfect Matching (MWPM) Find a perfect matching in a weighted graph G with maximum weight.

Theorem 1.1.1. *The* MWM *and* MWPM *problems are reducible to each other in general graphs.*

Proof. For an instance $G = (V, E)$ of MWM, define a new graph $G' = (V', E')$ where $V' = V_1 \cup V_2$ consists of two copies of V and the edge set E' contains two copies of E along with zero-weight edges between each corresponding pair of vertices in the two copies of V. A maximum weight perfect matching M' on G' can be used to obtain a maximum weight matching M on G by restricting the matching to only edges contained in V_1. If a vertex in V_1 is matched to its copy in V_2, it is unmatched in M. It is easy to see that M is a maximum weight matching on G as a matching with higher weight could be used to create a perfect matching on G' with weight higher than M'.

In the other direction, let a graph $G = (V, E)$ with weight function w be an instance of MWPM. Construct a weight function $w'(e) = w(e) + nN$. A maximum weight matching on the graph $G' = G$ with weight function w' must have the maximum possible number of edges as the nN term in the definition w' ensures that any matching with more edges has a higher weight. □

In the case of perfect matchings, sometimes the problem is defined as the Minimum Weight Perfect Matching. It is easy to see that it is equivalent to Maximum Weight Perfect Matching. To reduce an instance of one of the problems consisting of a graph $G = (V, E)$ with a weight function w to an instance of the other, simply create a new weight function $w'(e) = N - w(e)$. Similar reduction can be used when the instance of Minimum Weight Perfect Matching contains negative weights, we just need to take into account the difference between the minimum and maximum weights.

1.2 Maximum matching algorithms

The first polynomial algorithm for maximum weight matching in general graphs was given by Edmonds in [8]. It uses the primal-dual method and relies on his previous work on an algorithm for maximum cardinality matching [9]. It is called the *blossom algorithm* after blossoms, which are certain odd length cycles, a feature that notably does not appear in bipartite graphs and is a major source of complexity in algorithms for general graphs. The running time given by Edmonds was $O(n^2m)$. It was then independently improved to $O(n^3)$ by Lawler [27] and Gabow [20]. Many implementations of Edmonds' blossom algorithm followed with improved theoretical time complexity. The current best one due to [15] runs in $O(nm + n^2 \log n)$ due to, which is in some sense optimal: the blossom algorithm works in $O(n)$ phases, a single of which can be used to sort n numbers and therefore requires $\Omega(m+n \log n)$ time in a comparison-based model of computation.

Other algorithms use an approach called scaling to solve the problem for graphs with integer edge weights. The weights are exposed one bit a time starting from the most significant one. In the i-th scale, the algorithm computes the optimal solution for weights w_i consisting of i significant bits of w. The solution is then used to more efficiently solve the next scale. The first algorithm based on this approach was presented in [17] and runs in time $O(n^{3/4}m \log N)$. This was later improved to $O(m\sqrt{n\alpha(m,n) \log n} \log(nN))$ in [18] and lately $O(m\sqrt{n} \log(nN))$ in [7].

Additionally, there exists an algebraic randomized algorithm by [5] running in $O(n^\omega N)$ with high probability for the matrix multiplication exponent ω.

A summary of existing algorithms is presented in Table 1 and Table 2.

We will take a closer look at a selection of maximum weight matching algorithms, namely the original $O(n^2m)$ blossom algorithm of Edmonds [8], the $O(n^3)$ algorithm of Gabow [20], $O(nm \log n)$ algorithm of Galil, Micali and Gabow [21] and the $O(n^{3/4}m \log N)$ scaling algorithm running of Gabow [13].

Year	Authors	Ref	Running time
1965	Edmonds	[9]	$O(n^2m)$
1976	Gabow	[14]	$O(nm)$ or $O(n^3)$
1976	Lawler	[27]	$O(nm)$ or $O(n^3)$
1980	Micali & Vazirani	[29]	$O(\sqrt{n}m)$
1991	Gabow & Tarjan	[29]	$O(\sqrt{n}m)$
1989	Rabin & Vazirani	[32]	$O(n^{\omega+1})$
2004	Goldberg & Karzanov	[23]	$O(\sqrt{n}m/\kappa)$, $\kappa = \frac{\log n}{\log(n^2/m)}$
2004	Mucha & Sankowski	[30]	$O(n^\omega)$

Table 1: Maximum cardinality matching (MCM) algorithms.

Year	Authors	Ref	Running time
1965	**Edmonds**	[8]	$O(n^2m)$
1974	**Gabow**	[20]	$O(n^3)$
1976	Lawler	[27]	$O(n^3)$
1985	**Gabow**	[13]	$O(n^{3/4}m \log N)$
1986	**Galil, Micali & Gabow**	[21]	$O(nm \log n)$
1989	Gabow, Galil & Spencer	[19]	$O(nm \log\log\log_{2+\frac{m}{n}} + n^2 \log n)$
1990	Gabow	[15]	$O(nm + n^2 \log n)$
1991	Gabow & Tarjan	[18]	$O(m\sqrt{n\alpha(m,n) \log n} \log(nN))$
2012	Cygan, Gabow & Sankowski	[5]	$O(n^\omega N)$
2018	Duan, Pettie & Su	[7]	$O(m\sqrt{n} \log(nN))$

Table 2: Maximum weight matching (MWM) algorithms. Implemented algorithms are marked in bold.

We implement all of them in C++20 as part of the open-source KOALA NetworKit library [22]. In chapter 2 and chapter 3 we describe how they work and how to implement them. In chapter 4, we present results of computational tests and compare the performance of our implementations.

Chapter 2

The Blossom Algorithm

2.1 Edmonds' blossom algorithm

We take special consideration of specific sets of vertices of odd size which we will refer to as *blossoms*. We define blossoms recursively.

Definition 2.1 (trivial blossom)**.** *For each vertex v, the singleton $\{v\}$ is a* trivial *blossom.*

An edge sequence $e_0, e_1, \ldots, e_{n-1}$ where $e_i = (u_i, v_i)$ is *alternating* for blossoms $B_0, B_1, \ldots, B_n$ if $u_i \in B_i$ and $v_i \in B_{i+1}$ and $e_i \in M$ if and only if $e_{i+1} \notin M$. Such a sequence is *augmenting* if $e_0, e_{n-1} \notin M$.

Definition 2.2 (non-trivial blossom, subblossoms)**.** *Consider a sequence of blossoms $B_0, B_1, \ldots, B_n$ where n is odd, $B_0 = B_n$ with an alternating path of odd length $e_0, e_1, \ldots e_{n-1}$ where $e_0, e_{n-1} \notin M$. The blossoms $B_1, \ldots, B_n$ combine to form a* non-trivial *blossom B and are called the* subblossoms *of B. The edges $e_0, e_1, \ldots e_{n-1}$ are called the* blossom edges *of B. We call the list of pairs $(B_1, e_1), \ldots, (B_{n-1}, e_{n-1}), (B_n, e_0)$ the* subblossom list *of B.*

Definition 2.3 (base)**.** *For every blossom B we designate on of its vertices as the base of B. When a blossom is trivial the sole vertex is the base. For non-trivial blossom B defined as above the base of B is equal to the base of B_n.*

We say that a blossom B is *exposed* in some matching if the base of B is exposed.

Consider a blossom B with base b, subblossoms $B_1, B_2, \ldots, B_n$ and blossom edges $e_0, e_1, \ldots e_{n-1}$. We observe some useful facts about blossoms.

Fact 2.1. *Any vertex c of B, $c \neq b$, is matched to another vertex in B other than b.*

This directly implies the following

Fact 2.2. *If the base b is matched, then its mate is outside B.*

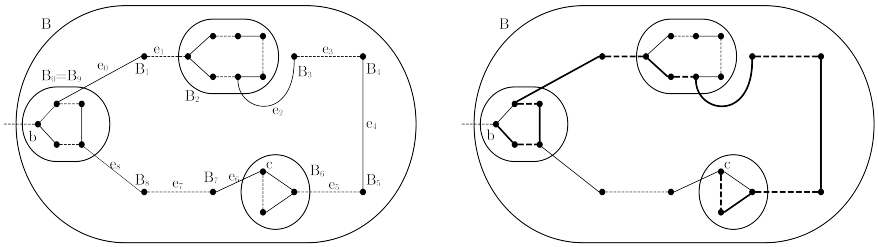

Figure 1: Example of a blossom B with subblossom list $(B_1, e_1), \ldots, (B_8, e_8), (B_9, e_0)$ and a base b. Edges in the matching are indicated with dashed lines. On the right an alternating path of even length between the base b of B and an example vertex c is highlighted.

Fact 2.3. *For every subblossom B_i, $0 < i < n$, the sequences $e_0, e_1, \ldots, e_{i-1}$ and $e_n, e_{n-1}, \ldots, e_i$ are alternating paths from B_0 to B_i. One of them is of odd length and the other one is of even length.*

Fact 2.4. *There exists an even length alternating path between b and any vertex $c \in B$.*

Proof. We use induction. If B is trivial the path is empty. If it is non-trivial, find the subblossom B_i of B such that $c \in B$ and choose the even length path from those described in Fact 2.3. For any subblossom on the path B_i exactly one of e_{i-1} or e_i is in M, and it is adjacent to the base b_i of B_i according to Fact 2.1 and Fact 2.2. The other of the two edges is adjacent to some $d \in B_i$. By induction, there is an even length alternating path from b_i to d. Inserting said path between e_{i-1} and e_i preserves the fact that the path is of even length and alternating. Doing so for each subblossom on the alternating path yields the desired path between b and c. □

Definition 2.4 (structure tree)**.** *The structure of a blossom B can be represented by a tree T_B. The root of T_B corresponds to B. When B is non-trivial, the children of the root of T_B are trees corresponding to subblossoms $B_1, \ldots, B_n$ of B. The leaves of T_B correspond to individual vertices that comprise B. We refer to T_B as the* structure tree *of B.*

Definition 2.5 (blossom list)**.** *The order of leaves in B structure tree implies an order on vertices of B. We call the list $L(B)$ of vertices of B in the order of their corresponding leaves in T_B the* blossom list *of B.*

Notice that a blossom list for any of the subblossoms of B or other any of their descendants is a substring of $L(B)$.

During the execution of the algorithm we will find and construct new blossoms and shrink all nodes and edges of the blossom into a single node. We might also expand blossoms and return their subblossom to the state they were

in before the expanded blossom was created. We call the blossoms which are not currently subblossoms of any other blossom *proper*. At any point in the algorithm we refer to the graph whose vertices correspond to proper blossoms as the *current graph*. For each vertex of the original graph we refer to the unique proper blossom it belongs to as its *current blossom*.

Edmonds' maximum weight matching algorithm from [8] is based on the primal-dual method [27].

The maximum weight matching problem in a graph $G = (V, E)$ can be expressed as an integer linear program with variables x_e for each edge $e \in E$

$$\begin{aligned} \text{maximize} \quad & \sum_{e \in E} x_e w(e) & \\ \text{subject to} \quad & \sum_{v} x_{uv} \leq 1 & \text{for each } u \in V \\ & x_e \geq 0 & \text{for each } e \in E \\ & x_e \in \{0, 1\} & \text{for each } e \in E \end{aligned}$$

It is well known that for bipartite graphs the integer constraint is not needed. This is not the case for general graphs. Consider the above linear program without the $x_e \in \{0, 1\}$ constraint. Take a triangle $G = (\{a, b, c\}, \{ab, bc, ca\})$ with weights 1 on all edges. The maximum weight matching in G consists of a single edge and has a weight of 1. In contrast, the optimal solution to the resulting linear program sets $x_v = 1/2$ for all $v \in V(G)$ and gives a value of $3/2$. In order to get rid of the integer constraint, Edmonds introduced an exponential number of constraints, one for each odd-sized vertex subsets with more than one vertex. We denote the set of such vertex subsets as $\mathcal{O}$.

$$\begin{aligned} \text{(MWM)} \qquad \text{maximize} \quad & \sum_{e \in E} x_e w(e) & \\ \text{subject to} \quad & \sum_{v} x_{uv} \leq 1 & \text{for each } u \in V \\ & \sum_{u,v \in B} x_{uv} \leq \left\lfloor \frac{1}{2} n_B \right\rfloor & \text{for each } B \in \mathcal{O} \\ & x_e \geq 0 & \text{for each } e \in E \end{aligned}$$

To show that the resulting linear program models the maximum weight matching problem he proved the following.

Theorem 2.1.1 ([8, Theorem(P)])**.** *Any optimal solution x to* (MWM) *contains only values of* 0 *and* 1 *which corresponds to a matching in G.*

The primal linear program (MWM) has a corresponding dual program with variables y_v for each vertex $v \in V$ and z_B for each vertex set $B \in \mathcal{O}$ which looks as follows:

$$
\begin{aligned}
(\overline{\text{MWM}}) \qquad \text{minimize } & \sum_{v\in V} y_v + \sum_{B} z_B \left\lfloor \frac{1}{2} n_B \right\rfloor & \\
\text{subject to } & y_u + y_v + \sum_{u,v\in B} z_B \geq w(uv) & \text{for each } uv \in E \\
& z_B \geq 0 & \text{for each } B \in \mathcal{O} \\
& y_v \geq 0 & \text{for each } v \in V
\end{aligned}
$$

Definition 2.6 (slack)**.** *For each edge $e = (u, v) \in E$ we introduce a function* slack(e) *defined as follows:*

$$slack(e) = \pi_e = y_u + y_v - w(e) + \sum_{\substack{B\in\mathcal{O}:\\ u,v\in B}} z_B$$

We say that an edge e is *tight* if $slack(e) = 0$. Moreover, we say that dual weights y and z are *dominating* if $slack(e) \geq 0$ for each $e \in E$.

Theorem 2.1.2. *Given dual variables y and z, a matching M is a maximum weight matching if the following conditions hold:*

(1) $y_v, \pi_e, z_B \geq 0$ *for each* $v \in V, e \in E, B \in \mathcal{O}$,

(2) $\pi_e = 0$ *for each edge* $e \in M$,

(3) $y_v = 0$ *for each exposed vertex* v,

(4) $|\{e\colon e \subseteq B, e \in M\}| = \left\lfloor \frac{1}{2} n_B \right\rfloor$ *for each blossom* B *where* $z_B > 0$.

Proof. Consider an arbitrary matching M'. We show that M has a higher weight:

$$
\begin{aligned}
\sum_{e\in M'} w(e) &= \sum_{uv\in M'} \left(y_u + y_v + \sum_{u,v\in B} z_B \right) && \text{from } \pi_{uv} \geq 0 \\
&\leq \sum_{v\in V} y_v + \sum_{B\in\mathcal{O}} z_B \left\lfloor \frac{1}{2} n_B \right\rfloor && \text{from } y_v, z_B \geq 0 \\
&= \sum_{uv\in M} \left(y_u + y_v + \sum_{u,v\in B} z_B \right) && \text{from (2) and (3)} \\
&= \sum_{e\in M} w(e) && \text{from (1)}
\end{aligned}
$$

□

The algorithm starts with a matching M and dual weights y and z that meet conditions (1), (2) and (4) of Theorem 2.1.2 and continually decreases the number of exposed vertices v with $y_v > 0$ by either matching them or decreasing their dual variable to 0. At the same time, we maintain that $z_B > 0$ only for blossoms B of the current graph.

The initial state of the algorithm consists of an empty matching M_0, dual weight $y_v = \max_{e \in E} w(e)/2$ and no blossoms, which trivially meet the desired constraints.

The algorithm works in phases we call *stages*. Each stage we find an augmenting path between two exposed blossoms comprised of solely tight edges and augments the matching along the path. The execution ends when there are no more exposed vertices or all dual variables y_v for all exposed vertices v have been reduced to 0.

We describe the basic data structures used to maintain the current state. Each vertex is associated with a unique number from $\{0, 1, \ldots, n-1\}$, which implies an order on vertices and allows us to perform comparisons $u < v$ between two vertices u and v by comparing their associated numbers. Each edge has a unique ID from $\{0, 1, \ldots, m\}$. Three arrays are used to represent the current matching M: MATE$[v]$ stores the mate of each vertex v, MATCHED_EDGE$[v]$ stores the ID of the matched edge of v and IN_MATCHING$[e]$ stores a boolean value indicating whether the edge e belongs to M.

In order to be able to iterate over them, we maintain a list of all proper blossoms. Each blossom contains a pointer to its position in said list which can be used to remove it. Additionally, each blossom B stores its blossom list $L(B)$.

2.1.1 Augmenting path search

The main part of the algorithm is concerned with finding an augmenting path between two exposed blossoms in the current graph. We will build trees comprised of alternating paths rooted in exposed blossoms which we will call *search trees*. Every blossom can be labeled with one of three values: *even*, *odd* or *free*. Free blossoms are outside of any search tree, while those that have can be reached are labeled based on whether the alternating path through which they have been reached has an even or odd length. We refer to blossoms labeled with even as *even blossoms* and all their vertices *even vertices* with analogous names for the remaining labels. At the beginning, all exposed blossoms are marked as even with the rest of blossoms marked as free.

Besides the label, for each blossom we store the edge called the *backtrack edge* by which the blossom has been reached. We refer to the collection of search trees as the *search structure*.

In order to expand the search structure we will look for so-called *useful edges*.

Definition 2.7 (useful edges)**.** *An edge $uv \in E$ is called* useful *if it is tight, that is $slack(e) = 0$, and one of following conditions hold:*

- *u is even and v is free,*

- *u and v are even.*

Based on which of the two conditions hold and whether the search trees of the two even blossoms are distinct, one of three steps can take place: *grow*, *blossom* or *augment*. Which of the steps happen is decided by the CONSIDER_EDGE and BACKTRACK procedures. At some point there may be any useful edges in which case a *dual weight adjustment* step is performed, during which dual weights are adjusted. The adjustment maintains optimality conditions and reveals new useful edges or forces blossom expansion. The intervals between dual weight adjustments are called *substages*. The BLOSSOM_SEARCH procedure is presented in Algorithm 1.

Algorithm 1 The blossom search procedure

```
1:  procedure BLOSSOM_SEARCH
2:      for each blossom B do
3:          if B is exposed then
4:              label B as even
5:          else
6:              label B as free
7:          end if
8:      end for
9:
10:     while an augmenting path has not been found do
11:         while a useful edge uv exists do
12:             CONSIDER_EDGE(uv)
13:         end while
14:         DUALWEIGHTADJUSTMENT()
15:     end while
16: end procedure
17:
18: procedure CONSIDER_EDGE(u, v)
19:     if v is free then                ▷ u is even from definition of useful edges
20:         GROW(u, v)
21:     else                                                        ▷ v is even
22:         BACKTRACK(uv)
23:     end if
24: end procedure
```

Grow step

In the case when the useful edge connects an even vertex u to a free vertex v, the grow step takes place. Let B_v be the blossom of v. We mark B_v as odd and set its backtrack edge to uv. Let p be the base of B_v. Because B_v was previously free, p has to be matched to some vertex q and its blossom B_q. We

mark B_q as even and set its backtrack edge to pq. Vertices of B_q become even for the first time so some of their adjacent edges might become useful.

Algorithm 2 The grow step procedure

```
1: procedure GROW(u, v)
2:     Let B_v be the current blossom of v
3:     b ← base(B_v)
4:     c ← MATE[b]
5:     Let B_c be the current blossom of c
6:     Label B_v as odd
7:     backtrack(B_v) ← (u, v)
8:     Label B_c as even
9:     backtrack(B_c) ← (b, c)
10: end procedure
```

Backtrack procedure

When considering a tight edge $e = uv$ between two even blossom B_u and B_v we backtrack using the previously described backtrack edges. We either find an augmenting path between two exposed blossoms at the roots of B_u and B_v's search trees and perform the augment step or the two blossoms are in the same search tree, in which case we execute the blossom step.

There is one important detail we need to take into account to ensure that the procedure does not waste too much time. Imagine that B_u and B_v are in the same search tree with root B_r. Let B_q be the lowest common ancestor of B_u and B_v in the search tree. The paths from B_u and B_v to B_r can be as long as $O(n)$. Meanwhile, the paths to B_q can be as short as a 1. If we backtracked along the whole path from B_u to B_r before backtracking from B_v we would spend $O(n)$ time.

To avoid that, we advance simultaneously along the two paths one step at a time and mark the blossom as visited while checking if the paths have met. Doing that we only visit $2\{l_u, l_v\}$ blossoms where l_u and l_v are the lengths of paths from B_u and B_v to B_q.

Augment step

If an augmenting path between two exposed blossoms is found, the augment step is performed and the current stage comes to an end. The augmentation consists of swapping the edges along the found path in the current matching M. We have previously described how to recursively find alternating paths in blossoms. After such an augmentation the bases of blossoms along the paths change and with them the order of subblossom lists, which get shifted cyclically to put the blossom containing the new base at the end.

Notice we do not actually need to perform the augmentation inside the blossoms and instead can do it lazily. For a blossom we only need to know its

base and its vertices. Instead of swapping edges along the full augmenting path in the original graph, we only swap the edges between proper blossom. The base of a blossom B is its sole vertex whose mate is outside B. When swapping edges between blossom we also need to change their bases. Lazy augmentation saves time when paths are repeatedly augmented. The augmentation is only done when the blossom is expanded. We do so by swapping edges along the even path from the initial base of the blossom to its current base. We again only do so for the blossom edges and only update the bases of subblossoms.

At the end of the algorithm we expand all blossoms to reveal the final matching. The only issue is that lazy augmentation makes it harder to maintain the matching. The value of the MATE array may not be accurate inside a blossom and is only guaranteed to be consistent for bases of blossoms. When expanding a blossom, besides swapping edges along the even path we also fix the MATE array for remaining edges using the information contained in the IN_MATCHING array, which is not affected by lazy augmentation.

Blossom step

When the two backtrack paths meet at some blossom B_q, a new blossom B is formed. The base of the new blossom is the base q of B_q. As the paths split in B_q, it has to be an even blossom which is why we also label B as such. The backtrack edge for B is the same as the backtrack edge of B_q, which is also the edge matched to q, or it is empty if B_q is a root of a search tree. Notice that vertices in odd subblossoms of B become even for the first time in the current stage, so some of their adjacent edges might turn useful.

Dual weight adjustment step

Whenever there are no more useful edges, a dual weight adjustment step is performed. For a chosen $\delta > 0$ we adjust the values of dual variables as follows:

- $y_v \leftarrow y_v - \delta$ for all even vertices v,
- $y_v \leftarrow y_v + \delta$ for all odd vertices v,
- $z_B \leftarrow z_B + 2\delta$ for all even blossoms B,
- $z_B \leftarrow z_B - 2\delta$ for all odd blossoms B.

We choose a maximum value of δ that preserves the (1), (2) and (4) constraints of Theorem 2.1.2. In order to do that we compute 4 candidate values $\delta_1, \delta_2, \delta_3$ and δ_4 and choose $\delta = \min\{\delta_1, \delta_2, \delta_3, \delta_4\}$. We define them as follows:

- $\delta_1 = \min_u y_u$ for even vertices u to preserve $y_u \geq 0$,
- $\delta_2 = \min_{uv} \pi_{uv}$ for edges $uv \in E$ with even u and free v to preserve $\pi_{uv} \geq 0$,

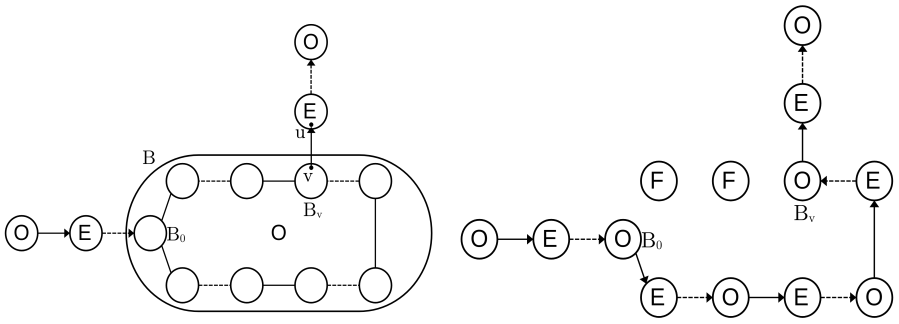

Figure 2: An example odd blossom B before and after expansion. The arrows indicate backtrack edges for each blossom. Blossoms are marked based on their label: E for even, O for odd and F for free blossoms.

- $\delta_3 = \frac{1}{2} \min_{uv} \pi_{uv}$ for edges $uv \in E$ with even vertices u and v to preserve $\pi_{uv} \geq 0$,
- $\delta_4 = \frac{1}{2} \min_B z_B$ for odd blossoms B to preserve $z_B \geq 0$.

Based on which of the above values we choose for δ a few different things can happen.

When $\delta = \delta_1$, all even vertices v have their dual variables reduced to $y_v = 0$. This includes all exposed vertices which meet the constraint (3). With all optimality constraints satisfied, the current matching M is maximum and the algorithm is done.

When $\delta = \delta_2$ or $\delta = \delta_3$ the edges for which the minimum was achieved become tight as their slack reaches 0, meaning they're now useful and we can continue with the search.

When $\delta = \delta_4$ we expand all odd blossoms B for which $z_B = 0$ after the dual weight adjustment. Let b be the base of B and uv be B the backtrack edge of B. Let B_v be the subblossom of B that contains v and B_0 the one that contains the base b. We find an alternating path of even length over subblossoms of B from B_0 to B_v and label the subblossom on the path even and odd in turns while setting their backtrack edges accordingly. The remaining subblossoms become free. An example expansion is pictured in Figure 1. As some vertices become even for the first time, some edges might become useful as a result of this process.

At the end of the stage all even blossoms B with $z_B = 0$ are expanded to preserve the constraint (4).

2.1.2 Complexity

All new blossoms are even and even blossoms are only expanded at the end of the stage, meaning all even blossoms correspond to some node in a structure tree of one of the proper blossoms at the end of the stage. Similarly, all odd

blossoms correspond to nodes in structure trees at the beginning of the stage. Overall, there are at most $O(n)$ different blossoms during a stage.

We count the number of dual weight adjustments. We get $\delta = \delta_1$ at most once as it leads to the end of the algorithm. Whenever $\delta = \delta_2$, the newly useful edges lead to grow steps during which new blossoms are labeled as odd and even. When $\delta = \delta_3$, either the blossom step is performed and a new even blossom is added or the augment step is executed and the stage is finished. Finally, each time $\delta = \delta_4$, an odd blossoms is expanded. We see that there are at most $O(n)$ dual weight adjustments in a single stage as their number is bounded by the number of unique blossoms. We call the section of the algorithm between each dual weight adjustment a *substage*.

The Edmonds' original algorithm spends most of its time calculating δ for each dual weight adjustment and finding new useful edges. It maintains a doubly linked lists of vertices for each proper blossom, which are split and concatenated whenever blossoms are created or expanded. For each vertex its current blossom is maintained in a simple array and updated together with the vertex lists. As there are $O(n)$ unique blossoms, it takes $O(n)$ time to maintain vertex lists and $O(n^2)$ to update the current blossom array in each stage. We also maintain a doubly linked list of all proper blossoms which we can update in constant time each time a new blossom appears by storing a pointer to its position in said list for each blossom.

Once a vertex becomes even, it stays that way, so we only need to scan its edges once either at the start of the stage, after a grow step or after an odd blossom expansion. All of that takes $O(m)$ time per stage. We calculate δ by iterating over all vertices and edges. If $\delta = \delta_2$ or $\delta = \delta_3$ we also check which edges achieved the minimum and become useful. Each dual weight adjustment step takes $O(n + m) = O(m)$ time for a total cost of $O(nm)$ per stage. With $O(n)$ stages the total time complexity of Edmonds' algorithm comes down to $O(n^2m)$ or $O(n^4)$ if we bound m by n^2, which is how he reported it.

2.1.3 Numerical accuracy

The algorithm is defined for rational numbers, but for practical reasons we would like to use integer data types. This is complicated by the presence of division which we do at two points – when calculating the starting values of y and when finding δ_3. In [20], it is shown that if the weights $w(e)$ are even for all edges $e \in E$ all of these calculations yield integer values.

At the start of the algorithm, we initialize the dual weight y_v for all v to the same value $\frac{1}{2}\max_{e\in E} w(e)$, which is obviously an integer when all weights are even.

Remember that we defined

$$\delta_3 = \frac{1}{2} \min_{\substack{uv\in E \\ u,\, v \text{ even}}} \pi_{uv} = \frac{1}{2} \min_{\substack{uv\in E \\ u,\, v \text{ even}}} \left(y_u + y_v - w(e) + \sum_{u,v\in B} z_B \right)$$

As dual variables z_B start at 0 and are changed by 2δ during dual weight adjustment, we can ignore them along with the edge weights. It is then sufficient to prove that weights y_v are of the same parity for all even vertices v. All exposed vertices start with the same dual weights which are adjusted by $-\delta$ with each dual weight adjustment, so their dual weight are all equal. It is enough for us to show that for any even $v \in V$ the parity of y_v is the same as y_u where u is the base of the exposed blossom at the root of v's search tree.

Let B_v and B_u be the current blossoms of u and v. There exist an even length alternating path P between v and u. This path can be obtained by finding the even length path from B_v to B_u that follows backtrack edges. The path is then expanded in similar way to how finding an even length path from a vertex to its blossom base. Let M be the current matching. We can swap edges along P and obtain a new matching M'. All edges in P are tight, so we can write the weight of M as

$$\sum_{uv \in M} w(uv) = \sum_{uv \in M} y_u + y_v + \sum_{u,v \in B} z_B$$

We can ignore the weights z_B as they're all even. The sums for M and M' share all vertices except for u and v so the difference between the weights of M and M' is of the same parity as $y_u - y_v$. As all edge weights are even, so are the weights of the two matchings meaning that $y_u - y_v$ is even and y_u and y_v are of the same parity.

2.1.4 Correctness

It is easy to verify that the obtained matching is optimal using Theorem 2.1.2. We've shown that during the execution we maintain conditions (1), (2) and (4) and the algorithm only finished when (3) is satisfied. What remains to show is that the algorithm terminates.

The progress of the algorithms is first measured by the number of exposed vertices v for which $y_v > 0$ which violate (3). Each stage finishes either by finding an augmenting path and decreasing the number of vertices violating (3) by making two of them no longer exposed or during the dual weight adjustment step when $\delta = \delta_1$ in which case all exposed vertices have their dual weights reduced to 0.

The progress of the BLOSSOM_SEARCH procedure is measured by the number of edges that have been added to the search structure (including those that have been absorbed into new blossoms). An edge is only added to the search structure at most once. When it becomes absorbed into a new even blossom, it remains in this state as even blossoms are only expanded at the end of a stage. If a useful edge exists, it can be added to the search structure which may result in an augmenting path being found or a new blossom being created. If no useful edges exist the dual weight adjustment is performed. We consider the possible cases that arise during dual weight adjustment:

1. If $\delta = \delta_1$, then the algorithm as a whole finishes.

2. If $\delta = \delta_2$, then some edge uv between an even vertex u and free vertex v becomes useful and can be added to the search structure.

3. If $\delta = \delta_3$, then some edge uv between even vertices u and v becomes useful and can be added to the search structure.

4. If $\delta = \delta_4$, then some odd blossoms B have their dual weights z_B reach 0 and are expanded. There no new odd blossoms, so this happens at most $O(n)$ times.

In all cases the dual weight adjustment results in new edges being added to the search structure. As there is a finite number of edges, this process has to come to an end, meaning the BLOSSOM_SEARCH procedure always terminates.

2.1.5 Finding a maximum weight perfect matching

The blossom algorithm can be used to find a maximum weight perfect matching with very few modifications. The MAXIMUM WEIGHT PERFECT MATCHING (MWPM) problem can be expressed as a linear program very similar to the one for (MWM). The only difference is that the inequality constraint $\sum_v x_{uv} \leq 1$ is replaced with an equality that ensures that each vertex is matched:

$$\begin{aligned}
\text{(MWPM)} \qquad \text{maximize } & \sum_{e \in E} x_e w(e) & \\
\text{subject to } & \sum_{v} x_{uv} = 1 & \text{for each } u \in V \\
& \sum_{u,v \in B} x_{uv} \leq \left\lfloor \frac{1}{2} n_B \right\rfloor & \text{for each } B \in \mathcal{O} \\
& x_e \geq 0 & \text{for each } e \in E
\end{aligned}$$

Naturally, the corresponding dual program is also very close to ($\overline{\text{MWM}}$). The only difference is that the equality constraint causes the corresponding dual variables y_v to be unrestricted:

$$\begin{aligned}
\text{($\overline{\text{MWPM}}$)} \qquad \text{minimize } & \sum_{v \in V} y_v + \sum_{B} z_B \left\lfloor \frac{1}{2} n_B \right\rfloor & \\
\text{subject to } & y_u + y_v + \sum_{u,v \in B} z_B \geq w(uv) & \text{for each } uv \in E \\
& z_B \geq 0 & \text{for each } B \in \mathcal{O}
\end{aligned}$$

The conditions need for a matching to be the maximum weight perfect matching are very similar and are summarized in the following theorem analogous to Theorem 2.1.2:

Theorem 2.1.3. *Given dual variables y and z, a matching M is a maximum weight perfect matching if the following conditions hold:*

1. $\pi_e, z_B \geq 0$ *for each* $v \in V, e \in E, B \in \mathcal{O}$,
2. $\pi_e = 0$ *for each edge* $e \in M$,
3. M *is a perfect matching,*
4. $|\{e\colon e \subseteq B, e \in M\}| = \left\lfloor \frac{1}{2} n_B \right\rfloor$ *for each blossom* B *where* $z_B > 0$.

The algorithm again maintains conditions (1), (2) and (4) and runs until the matching M is perfect. Each stage increases the number of edges in M. The only difference from the maximum weight matching algorithm is that we no longer calculate $\delta_1 = \min_{\text{even } u} y_u$ as the dual weights y_u are unrestricted.

The lack of restriction on y_u allows us more freedom in choosing initial weights. This was not possible in the described maximum weight matching algorithm, as it relies on the fact that all exposed vertices have the same dual weight. If the weights differed, at some point one of the exposed vertices v would reach $y_v = 0$ while others would still be positive. In this case, we would not be able to perform dual weight adjustment as we've described it. One initialization strategy is the *greedy* initialization. We start by assigning to each weight y_v half of the maximum weight of any edges adjacent to v. We then iterate over all the edges and add any tight edge to the initial matching M (only if none of its endvertices have been matched). This brings some problems with numerical accuracy. To ensure that all weights for even vertices start out with the same parity, we multiply the weights by 4 instead of 2.

2.2 Gabow's algorithm

Most of the time in Edmonds' original implementation is spent calculating δ_2 and δ_3 as well as finding useful edges during each dual weight adjustment. These particular issues are addressed by Gabow in his implementation from [20].

For each non-even vertex v we maintain $edge(v)$ which we define as the edge vu with the smallest possible slack that connects v to an even vertex u.

For each even blossom B we maintain a list $edges(B)$ consisting of edges $v_1u_1, \ldots, v_nu_n$ sorted such that $u_1 < u_2 < \cdots < u_n$ where for each $i = 1, \ldots, n$: $v_i \in B$, $u_i \notin B$, u_i is even and v_i became even after u_i and v_iu_i has the smallest slack among such edges vu_i where $v \in B$. We also store the one edge in $edges(B)$ with the smallest slack and denote it $edge(B)$.

With those values we can perform a dual weight adjustment in time $O(V)$ by calculating:

$$\delta_2 = \min_{\substack{uv \in E \\ u \text{ even},\ v \text{ free}}} slack(uv) = \min_{u \text{ even}} slack(edge(u))$$

$$\delta_3 = \min_{\substack{uv \in E \\ u,\ v \text{ even}}} \frac{1}{2} slack(uv) = \min_{B \text{ even}} \frac{1}{2} slack(edge(B))$$

We can find new useful edges after a dual weight adjustment in $O(n)$ by checking which edges achieved the minimum. Notice that the minimum edge in $edge(v)$ and $edge(B)$ does not change with a weight adjustment as all relevant edges have their slack changed by the same amount.

Whenever a blossom B becomes even either at the start of a stage, after a grow step or during odd blossom expansion, we check all edges adjacent to vertices of B. For each $v \in B$ we iterate over its neighbors u in ascending order. If u is not even, we update $edge(u)$ if uv has smaller slack. If u is even, we merge it into $edges(B)$ if it has smaller slack than any other edge $\{v', u\}$ in the list. To do this efficiently we take advantage of the fact that the list and neighbors are sorted by utilizing a pointer to an element of $edges(B)$ that we reset to the beginning for each $v \in B$ and move along the list according to the value of u. We only scan each vertex once during a stage, so we spend at most $O(n^2)$ time per stage doing this.

The only other time the lists *edges* change is when a new blossom B is found. We first perform the previously described scan for all odd subblossoms of B and then merge all the lists together. We perform at most $O(n)$ merges per stage which we can each do in $O(n)$ using the classic sorted list merge algorithm costing us $O(n^2)$ time per stage.

To sum up, we spend $O(n^2)$ time maintaining the new values and $O(n^2)$ doing dual weight adjustment per stage. All the other calculations stay the same as in Edmonds' algorithm and cost $O(n^2)$ per stage giving us a final running time of $O(n^3)$.

2.3 Galil, Micali & Gabow algorithm

Implementations of the blossom algorithm described above spend most of their time doing dual weight adjustments and finding useful edges. The authors of [29] showed that it is possible to perform dual weight adjustment in constant time by taking advantage of the fact that large groups of weights are changed by the same amount. We start by describing the specialized priority queues they have devised to do so and we show how they are used in the search procedure of the blossom algorithm.

2.3.1 Data structures

We define a specialized priority queue we call pq_1 which operates on integers from a set $\{0, \ldots, n-1\}$ for a specified n. Elements start outside the queue and can be inserted into or deleted from it. Elements inside the queue are called *current* elements. Each current element has an associated priority. The data structure suppors following operations:

- INSERT(i, p) inserts an element i with priority p or update the priority of i if p is smaller than the current priority of i in time $O(\log n)$,
- DELETE(i) deletes the element i in time $O(\log n)$,

- FIND_MIN() finds the element with the lowest priority in time $O(1)$,
- CURRENT_PRIORITY(i) returns the current priority of i in time $O(1)$
- DECREASE(δ) decreases the priorities of all current elements by δ in time $O(1)$.

Let Δ be the sum of all priority decreases. We make use of a priority queue Q which supports insertion, deletion, finding minimum and updating priority which in our case is implemented with an array heap. We do not store the exact priorities of the elements Q. Instead, we use what we call their *modified priorities*. The modified priority of an element is calculated at the moment of insertion into the priority update. When the priority of an element is set to p we set its modified priority to the value $p+\Delta$ and the time of the change and store it as such in Q. Additionally, we store the modified priority in an array to support checking the current priority of a given element in CURRENT_PRIORITY. The current priority of any element can be calculated by adding Δ to its modified priority. The order of all the modified priorities is the same as the order of actual priorities as they're all just shifted by Δ.

We will also make use of what we call *concatenable queues*. Each queue contains a list of elements in a specific order. Each element has a corresponding priority. The queues can be concatenated together or split at a specified element. Elements are referenced using handles which are preserved by split and concatenate operations. An elements handle allows one to find its current queue.

- INIT() creates an empty queue,
- APPEND(q, i, p) appends an element i with priority p to the end of the queue q in $O(\log n)$,
- DELETE(i) deletes the element i from its queue in time $O(\log n)$,
- FIND_MIN(q) finds the element in queue q with the lowest priority in time $O(1)$,
- CONCAT(q_1, q_2) concatenate queues q_1 and q_2 in time $O(\log n)$,
- SPLIT(q, i) splits q into two new queues: one that contains all elements in q up to i and one that contains all elements after i in q in time $O(\log n)$,
- FIND_QUEUE(i) returns the current queue of i in time $O(\log n)$.

We implement concatenable queues using 2–3 trees which support splitting and concatenation as described in [1]. Elements are stored in the leaves of the tree with their order determined by the order of the leaves. Each inner node stores the element with the lowest priority among its descendants along with said priority, meaning the minimum can be easily accessed by checking the root. The elements are referenced using pointers to their corresponding leaves. We

take care to maintain said pointers during splits and concatenations. The root of the tree stores a pointer to its queue which allows us to implement FIND_QUEUE with a simple walk from its corresponding leaf. To differentiate between queues each one stores an ID in its root which can be set during INIT, CONCAT or s SPLIT operations. The ID is what is returned by FIND_QUEUE.

The last data structure we need is priority queue pq_2. Just like pq_1 it operates on elements from an integer set $\{0, \ldots, n-1\}$ which are divided into groups. All elements have an associated priority. Within a given group the elements are ordered. Each group can be either *active* or *inactive*. We call elements active if they are inside an active group. We support modifying the priorities of all active elements by a provided value. The status of a group can be changed at any time and their elements can be split to create two new groups. Lastly we can retrieve the active element with the smallest priority. To summarize, the queue supports following operations:

- CREATE_GROUP() creates a new empty group,
- APPEND(g, i, p) adds an element i with priority p to the end of the group g in $O(\log n)$,
- UPDATE(i, p) updates the priority of i to p if it is smaller than the current one in time $O(\log n)$,
- FIND_MIN() finds the active element with the lowest priority in time $O(1)$,
- SPLIT(g, i) splits g into two new groups: one that contains all elements in q up to i and one that contains all elements after i in q in time $O(\log n)$,
- CHANGE(g, s) changes the status of g to s in time $O(1)$,
- DELETE(g) deletes the group g,
- DECREASE(δ) decreases the priorities of all active elements by δ.

Additionally, we store the sum of all changes from the DECREASE function in a variable Δ.

For each group g we sum up all the changes to the priorities of its elements in a variable Δ_g. This value is updated each time we refer to g. To do this we also store the value Δ_{last} which corresponds to the value of Δ last time we looked at g. If during that time g was active, we increase Δ_g by $\Delta - \Delta_{last}$. We then update Δ_{last} to the current value of Δ regardless of the status of g. The elements of g are stored in a concatenable queue Q_g. The priorities of the elements in Q_g again use modified priorities similarly to pq_1 but using Δ_g instead.

In order to find a minimum active element we maintain a pq_1 called $Q_{\min}$ which for each active group stores its smallest element according to priority. Whenever a new element is added or a priority is changed in an active group we check if we need to update its corresponding entry in $Q_{\min}$. When a group is set to inactive or deleted we remove said entry. Similarly, during a split if the

group is active we need to replace its entry with entries for the two resulting groups. We need to make sure to store the elements' actual priorities in $Q_{\min}$. The decrease function simply calls decrease on $Q_{\min}$.

2.3.2 Search procedure

We use concatenable queues to maintain the current blossoms for each vertex. Each proper blossom B has an associated concatenable queue Q_B. The elements of Q_B are vertices of B in the order of its blossom list $L(B)$. When a new blossom B' is formed create a new concatenable queue by performing the CONCAT operation with queues corresponding to subblossoms of B'. When a blossom is expanded, we perform SPLIT operations to create queues for its subblossoms that now become proper.

Maintaining the queues costs us $O(n \log n)$ per stage. The IDs of queues store references to their corresponding blossoms allowing us to retrieve the current blossom of any given vertex in $O(\log n)$.

To maintain dual weights y_v we make use of two queues pq_1 we call y_{even} and y_{odd} which we use to maintain values y_v for correspondingly even and odd vertices. The values for free vertices are stored in an array y_{free}. Similarly, for weights z_B two queues pq_1 z_{even} and z_{odd} for even and odd blossoms are used. Then B is free, its dual weight z_B is stored int a field in the blossom struct. To retrieve the current value of y_v or z_B for a vertex v or blossom B (mainly for the purpose of calculating the slack of an edge) the algorithm checks their label and refers to the appropriate queue.

In order to calculate the value of δ_3, we use a pq_1 called Q_{good} in which we store edges between two different even blossoms which we refer to as *good edges*. The priority of each edge in Q_{good} corresponds to its current slack. Over the course of the algorithm, as new even blossoms are created, some of those edges might become contained within a single even blossom. We do not have enough time to detect it at the time it happens, so instead we remove them in a lazy manner each time we want to check the minimum by removing the smallest edge as long as it is no longer good.

To calculate the value of δ_2 efficiently we make use of a pq_2 called Q_{even}. Each non-even blossom B has a corresponding group g_B. The elements of g_B are vertices of B stored in blossom order. The priority of $v \in B$ in Q_{even} corresponds to the minimum slack of an edge $uv \in E$ such that u is even. We also remember the edge for which the minimum is achieved. Group g_B is active when B is free and inactive when it is odd. Free blossoms may become odd and odd blossoms might be expanded into even, odd or free subblossoms, so we need to change the status of a group and perform splits whenever necessary.

Beginning of a stage

For free blossom we initialize groups in Q_{even} and set all priorities to inf. For even blossoms we scan outgoing edges to update Q_{even} and Q_{good} and start tracking the dual variables for even vertices using y_{even}.

Grow step

Two blossoms are added to the search structure, one even and one odd. We scan edges adjacent to the newly even vertices to update Q_{good} and Q_{even}. To maintain the y weights we add the blossoms' vertices to y_{even} and y_{odd} according to their labels.

Blossom step

We concatenate all vertex queues for all subblossoms. For each odd subblossom we scan its outgoing edges to update Q_{good} and Q_{even}, move its vertices from y_{odd} to y_{even} and delete its corresponding group in Q_{even}. We remove the subblossoms from z_{even} and z_{odd} and insert the new blossom into z_{even}.

Odd blossom expansion

Let B be the odd blossom that is being expanded. We split the group B in Q_{even} and assign the new groups to corresponding subblossoms. We delete the groups for even subblossom as they're no longer in use. For non-even ones we set the status according to the label – active for free subblossoms and inactive for odd. We also swap the way we track y according to the new labels.

Dual weight adjustment step

We can calculate δ using our FIND_MIN methods for corresponding queues:

- y_{even} to calculate δ_1,
- Q_{good} to calculate δ_2,
- Q_{even} to calculate δ_3,
- z_{odd} to calculate δ_4.

To adjust dual weights we use the DECREASE methods to change priorities in all the queues:

- by δ in y_{even},
- by $-\delta$ in y_{odd},
- by -2δ in z_{even},
- by 2δ in z_{odd},
- by 2δ in Q_{good},
- by δ in Q_{even}.

End of a stage

After the search finishes we clear all the queues after first moving the current priorities from queues y_{even}, y_{odd}, z_{even} and z_{odd} into the y_{free} array and the z field to keep the dual weights up to date for the next stage.

2.3.3 Complexity

There are $O(n)$ stages in the algorithm. During each stage we consider each edge at most twice – once each time one of its endvertices becomes even and its adjacent edges are scanned. When considering an edge we make a constant number of calls to various queues – to check the current blossoms of its endvertices or to update edge queues. Each queue operation takes at most $O(\log n)$ time.

Maintaining queues for blossoms by executing splits and concatenations when necessary takes $O(n \log n)$ time per stage. We previously showed that there are at most $O(n)$ different proper blossom during a stage. Each unique proper blossom corresponds to at most one SPLIT operation when it becomes proper as a result of an expansion of its parent and at most one CONCAT operation when it becomes a subblossom of a newly created blossom.

The $O(n)$ dual adjustments each take constant time. Summing it all up, each stage takes $O((n+m)\log n) = O(m \log n)$ time for a total running time of $O(nm \log n)$.

Chapter 3

Scaling algorithms

One of the limiting factors in the running time of algorithms based on the blossom algorithm is the fact that they find augmenting path one at a time. The most efficient maximum cardinality matching algorithms do so in batches. One technique that has been successfully used to achieve that is *scaling* first introduced in [11] and has been applied to among others to weighted bipartite matching in [19] and various geometric problems in [16].

The idea of the scaling approach is to recursively solve the problems for reduced weights, representing i most significant binary digits, and use the returned solution to more efficiently solve the original instance. It has been first used in [13] to achieve a running time of $O(n^{3/4}m \log N)$. The algorithm works in $O(\log N)$ phases each of which finds the maximum weight perfect matching. Later work from [18] showed that it is enough to find an approximate solution in each scale that differs from optimal by at most $O(n)$, but it requires $O(\log(nN))$ scales to be executed, each of which runs in $O(m\sqrt{n\alpha(m,n)\log n})$ time. The algorithm of [7] further reduces the requirements by finding near-optimal and near-perfect matching in each scale to achieve a running time of $O(m\sqrt{n}\log(nN))$. We take a closer look and describe an implementation of the $O(n^{3/4}\log N)$ algorithm from [13].

3.1 Gabow's scaling algorithm

All the existing scaling algorithms solve the maximum weight perfect matching problem. When solving the maximum weight matching problem, a reduction described in Theorem 1.1.1 is used.

We define a function $yz(S)$ for a subset of vertices $S \subseteq V$ as follows:

$$yz(S) = \sum_{v \in S} y_v + \sum_{B \subset S} z_B \left\lfloor \frac{1}{2} n_B \right\rfloor + \sum_{B \supseteq S} z_B \left\lfloor \frac{1}{2} n_S \right\rfloor$$

Notice that $yz(V)$ corresponds to the objective function of the dual program ($\overline{\text{MWPM}}$) and $yz(e) = slack(e)$.

Definition 3.1 (blossom tree)**.** *For a matching M on a graph G with corresponding weights y and z and proper blossoms $B_1, \ldots, B_k$, we define its* blossom tree *to be a tree whose root is G with children $T_1, \ldots, T_k$ where T_i is the structure tree of B_i for $i = 1, \ldots, k$.*

The algorithm works using recursion. The SCALE(w) procedure receives even weights $w(e)$ and returns a maximum weight perfect matching for w along with corresponding dual weights and blossom tree. We first calculate smaller weights $w'(e) = 2\lfloor w(e)/4 \rfloor$ that remain even. We then perform a recursive call SCALE(w') which returns a perfect matching M' with dual weights y' and z' along with a blossom tree T'. The returned values are used to solve the problem for the original weights $w(e)$. We start by calculating new weights y^0 and z^0 which are then used by the MATCH procedure.

Algorithm 3 The SCALE procedure

```
1: procedure SCALE(w)
2:     if w(e) = 0 for all e ∈ E then
3:         return perfect matching M, dual weights y_v = 0 and no blossoms
4:     end if
5:     w'(e) ← 2⌊w(e)/4⌋ for all e ∈ E
6:     M', y', z', T' ← SCALE(w')
7:     y^0_v ← 2y'_v + 1 for all v ∈ V
8:     z^0_B ← 2z'_B for all old blossoms B ∈ T'
9:     MATCH(y^0, z^0, T')
10:    return perfect matching M with weights y, z and blossom tree T
11: end procedure
```

The base case of the recursion happens when $w(e) = 0$ for all $e \in E$. The algorithm returns a perfect matching M which can be found by a maximum cardinality matching algorithm. Technically, unless all weights in the graph are zero, the matching gets immediately discarded. Along with the matching we return dual weights $y_v = 0$ and an empty blossom tree (consisting only of root G).

The depth of recursion is $O(\log N)$. The formula $y^0_v \leftarrow 2y'_v + 1$ accounts for the bit lost in calculating w'. The weights y^0 and z^0 are feasible and close to optimal, which is show by a following fact:

Theorem 3.1.1. *For any maximum weight perfect matching N:*

$$y^0 z^0(V) \geq w(N) \geq y^0 z^0(V) - n \tag{3.1}$$

Proof. The first equality follows from complementary slackness and the second one from the fact that the duals returned from the recursive call were optimal:

$$\begin{aligned}
y^0 z^0(V) &= 2yz(V) + n && \text{from the definition of } y^0 \text{ and } z^0 \\
&= 2w'(M) + n && \text{from the recursive call } yz(V) = w'(M) \\
&= 2 \sum_{e \in M} 2 \left\lfloor \frac{w(e)}{4} \right\rfloor + n && \text{from the definition of } w' \\
&\leq \sum_{e \in M} w(e) + n && \text{from the fact that } 4 \left\lfloor \frac{x}{4} \right\rfloor \leq x \\
&\leq w(N) + n && \text{from the definition of } N
\end{aligned}$$

□

The problem is that the edges in blossoms in T' are not tight for the new weights y^0 and z^0. We want to make use of the new weights while getting rid of the old blossoms. We say a blossom has *dissolved* when its dual weight has been reduced to 0. One way to do that is to *distribute* the dual weight z_B of an old blossom B. The distribution consists of decreasing z_B by a value δ while increasing y_v by $\delta/2$ for all $v \in B$. While the distribution maintains dominance it also increases the dual objective function by $\delta/2$.

The match procedure starts with an empty matching and weights y^0 and z^0 and builds a *current matching* with corresponding blossoms. We work with two types of blossoms. We refer to the ones inherited from the recursive call as *old blossoms* and blossoms created during this scale as *current blossoms*. When calculating the $yz(S)$ function we now take into account both old and current blossoms.

In order to efficiently dissolve old blossoms the algorithm uses heavy path decomposition. Let B be a blossom in a blossom tree T. We call a subblossom C of B its *heavy child* if $n_C > \frac{1}{2} n_B$. Any blossom has at most one heavy child. We can partition T into *heavy paths* which are maximal paths in T in which the next blossom on the path is the heavy child of the previous blossom. If a blossom is not the heavy child of its parent (or it is the root of the tree) it is a root of its own heavy path.

The MATCH function works by partitioning the old blossom tree T' into heavy paths and then calling PATH(B) on roots B of heavy path in postorder.

Additionally, in order to maintain the dominance on the edges outgoing from the currently processed old blossom, the algorithm maintains the invariant $y_v \geq y_v^0$ for all vertices $v \in G$.

3.2 Path procedure

We refer to a matching M with weights dual weight y and z and blossoms that meet conditions (1), (2) and (4) of Theorem 2.1.3 as a *structured matching*.

Definition 3.2 (shell)**.** *Consider a perfect structured matching with blossom B and its descendant C we refer to the graph induced by $B \setminus C$ as a* shell.

The procedure PATH(B) takes as an input a root B of a heavy path P_B. All old blossoms contained in B that are not a part of P_B have been dissolved prior to calling PATH(B).

Two consecutive blossoms on P_B form a shell. All the undissolved blossoms on path P_B form a sequence of shells. We also include the deepest blossom C in P_B which corresponds to a shell $C \setminus \emptyset$. This is not a shell as we have previously defined them and as such has different properties.

The procedure dissolves all old blossoms on P_B while building up the current matching and associated blossoms and maintaining the optimality constraints. Over time as the old blossoms are dissolved the shells merge together. One exception is when $B = G$, in which case the top shell $G \setminus C$ never dissolves. The procedure ends when all the old blossoms are dissolved or when $B = G$ and the current matching is perfect.

Algorithm 4 The PATH procedure

```
1:  procedure PATH(B)
2:      while there are undissolved blossoms in P_B or B = G and M is not
        perfect do
3:          AUGMENT_PATHS()
4:          Let L be a list of shells with at least 1 exposed vertex
5:          Sort shells in L according to descending number of exposed vertices
6:          for each shell S in L do
7:              if S has not been searched or dissolved then
8:                  SHELL_SEARCH(S)
9:              end if
10:         end for
11:     end while
12: end procedure
```

3.2.1 The path augmentation procedure

The matching is augmented by the AUGMENT_PATHS procedure. The procedure builds a graph G' by shrinking current blossoms and using only tight edges contained within a single shell as edges. It translates the current matching M to a matching M' on G'. Using M' as a starting point, it finds the maximum cardinality matching in G' by calling a maximum cardinality matching algorithm. The resulting matching is used to augment the matching M on G.

In the AUGMENT_PATHS procedure we make use of a maximum cardinality matching algorithm. In our case it is specifically the Micali-Vazirani algorithm introduced in [29] which we will refer to as the MV algorithm. To efficiently contract the blossoms and create a new graph we need some information about the shells and blossoms. For each shell we maintain a list of proper blossoms

contained within it and a list of vertices. These lists get concatenated as old blossoms dissolve. At the start of the iteration we record for each vertex its current shell and blossom in an array to be able to efficiently check it. We do not delete the structure for the root of the path even after it was dissolved, we just mark it as such. At the end of the PATH procedure, after all the old blossoms have been dissolved it will contain a list of all blossoms and vertices in the path. These lists are then concatenated to the lists of the parent of B in the blossom tree, so that it contains the information about current blossoms.

We also need to efficiently calculate a slack of an edge to check whether it is tight. Notice that we only need to check an edge's slack if it is contained in a single shell and connects vertices in different blossoms. If we want to calculate the slack of such an edge uv, we only need to know y_u, y_v, $w(uv)$ and the sum of dual weights of all old blossoms that contain it. For each blossom B on the current path we define $z_{path}(B)$ as a sum of dual weights z_P for all ancestors P of B on the current path including B. Before starting PATH(B) we also calculate the value z_{outer} which is the sum of dual weights for all old blossoms containing B in the old blossom tree. We do that while performing the heavy path decomposition by summing up the values along the way. Let B be the smallest old blossom one P_B that contains uv. We can calculate the slack of uv as

$$slack(uv) = y_v + y_u - w(uv) + z_{path}(B) + z_{outer}$$

We explicitly build the shrunk graph G' before passing it to the MV algorithm. To do that, we iterate over all blossoms and associate them with consecutive integers. When adding edges, we remember which original edge they come from. The MV algorithm returns the matching as a mapping from a vertex to its mate. We iterate over the returned matching and augment our current matching according to it. If a blossom B was matched to C by the MV algorithm, we first iterate over the adjacency list of B in G' to find the original edge through which the two were matched. Just like in the blossom algorithm, the augmentation is lazy – we only change the bases of B and C (along with some other indicators that describe the matching).

3.2.2 The shell search procedure

The main part of the procedure consists of calls to the SHELL_SEARCH($B \setminus C$) procedure, which takes as an input a shell and executes a variant of the blossom algorithm search. We search for an augmenting path over tight edges in the shell defined by two consecutive undissolved old blossoms. We call these two old blossoms the *boundaries* of the searched shell and refer to them as *outer* and *inner* where the outer blossom contains the inner one. During the search some of these old blossoms might get dissolved. When that happens the boundaries of the shell change and the searched graph grows. When searching it is the case that the innermost shell the inner boundary does not exist.

Note that apart from blossom dissolution the high level search algorithm is

almost identical to the search procedure of the perfect matching version of the blossom algorithm. The main difference is in the dual weight adjustment step. When calculating δ we have to take into the account the weight corresponding to the inner and outer blossom of the shell. When adjusting dual weights by δ we distribute 2δ units for the two boundaries of the shell. The two exceptions are when the inner blossom does not exist and when the outer blossom is the whole graph G for which we do not perform any distributions. When calculating δ we make sure that the distribution does not decrease the dual weight of a boundary below 0.

After dual weight adjustment the dual weight corresponding to one of the boundaries might be brought down to 0. When that happens that old blossom is dissolved and the boundaries are moved. When the outer blossom is dissolved we move the boundary to its parent in the path P_B and when it is the inner one, we move the boundary to its child. We say that the current shell *dissolves into* the shell defined by the dissolved boundary and the new boundary. If said shell has already been involved in a previous shell search during this iteration, we finish our search. When the outer boundary is dissolves, and it happens to be the outermost non-dissolved blossom in P_B, we also halt our search. If none of these cases happen the search finishes when it finds an augmenting path over tight edges.

The details of implementing the SHELL_SEARCH procedure differ in major ways from the previously described approaches and are described in the next section.

Lemma 3.2.0.1. *The* SHELL_SEARCH *procedure maintains the dominance on all edges, tightness on blossom edges and the property* $y_v \geq y_v^0$.

Proof. The dominance and tightness on blossom edges are ensured by the blossom search algorithm.

The distributions on C and D maintain the dominance and tightness on edges contained within $C \setminus D$ as the decrease of z_C cancels out the increase of y.

For edges uv such that $u \in C \setminus D$ and $v \in D$ only dominance needs to be maintained, as they are not part of the matching. The slack of such an edge uv changes by -2δ from the change to z_C, by δ and 2δ from distributions to y_u and y_v and at the worst by $-\delta$ from a dual adjustment to y_u, depending on the label of u. This means that the slack of uv does not decrease and remains non-negative. □

Similar investigation of dual adjustments shows how the value of the yz function changes with each adjustment:

Lemma 3.2.0.2. *Let f be the number of exposed vertices in the current shell.*

In the search of a shell $C \setminus D$ other than the last shell of B, a dual adjustment with the value of δ decreases $yz(B)$ by $\delta(f-2)$ and $yz(C \setminus D)$ by δf where f is the number of exposed vertices in $C \setminus D$.

In the search of the last shell $C \setminus \emptyset$, both $yz(B)$ and $yz(C)$ are decreased by $\delta(f-1)$.

3.3 Shell search implementation

The authors of the algorithm leave out the details of implementing the SHELL_SEARCH procedure except for key data structures used. We rely on a description given in [7] with some additions.

To take advantage of integer weights instead of employing tree based priority queues that work in logarithmic time we can make use of an array based queue. It is convenient for us to reframe the dual adjustment in terms of *time*. We call steps of the search algorithm *events*. We say that an event happens at time t if happens after the sum of dual adjustments reaches t. A simple implementation of an array queue stores events in an array of linked lists and supports scheduling events at set times. It stores the current time and a pointer to current event. To retrieve the next event it moves the pointer to the next event in the current list or moves the time counter until the corresponding list is non-empty.

There are 4 types of events that happen during the search:

- $grow(v, e)$ – happens when the blossom B_v containing the vertex v is added to the search structure through edge e. The value of e can be equal to $\emptyset$, which happens when B_v is exposed and it becomes the root of a new search tree,
- $blossom(e)$ – happens when a new blossom is created after the addition of edge e to the search structure,
- $dissolve(B)$ – happens when an odd blossom B expands after its corresponding dual weight z_B reaches 0,
- $dissolve\_shell(B)$ – happens when one of the boundaries of the search dissolves after its weight reaches 0.

Another problem we need to solve is maintaining vertices' membership to blossoms. We have previously used concatenable queues which we split and concatenated as blossom were created and expanded. We can do better by taking advantage of the order blossoms change. Notice that once a blossom becomes even it can no longer expand and stays even, it can only become a part of a new even blossom. Only odd blossoms expand after which some of their subblossoms become even, odd and free. A free blossom can either become even or odd as it gets added to the search structure. We can model this behavior using two data structures – one for splitting and one for joining sets of vertices.

The first data structure used is the well known UNION-FIND, which operates on a universe of elements from $\{0, 1, \ldots, n-1\}$. The elements are divided into sets. Each set additionally has an ID. It supports the following operations:

- INIT(u, id) – initializes the set of u to a singleton $\{u\}$ with the provided ID,
- JOIN(u, v) – joins the sets containing u and v into a new set with the provided ID,

- FIND(u) – returns the ID of the set of u.

The last data structure we need is the SPLIT-FINDMIN data structure also called the splitting list data structure introduced in [13]. In the next section we described in detail its implementation. The splitting list data structure operates on a universe of elements from $\{0, 1, \ldots, n\}$. Every element x can be contained in at most one list $L(x)$ and has an associated cost $c(x)$ which can be infinite. The cost $c(L)$ of a list is the smallest cost of an element in L. Every list has an id. It supports the following operations:

- INITIALIZE($x_1, \ldots, x_l$) initializes a list of elements,
- DECREASE_COST(x, d) update $c(x)$ to $\min\{d, c(x)\}$,
- SPLIT(x) split $L(X)$ into two lists L_1 and L_2 where L_1 contains all elements of $L(x)$ until and including x and L_2 contains the remaining elements in the order they appear in $L(x)$,
- FIND_MIN(L) return $c(L)$ and the element of L which achieves the minimum,
- FIND_LIST(x) return the current list of x.

With these two data structures we can keep track of which blossom each vertex belongs to. Each even blossom will correspond to a set in the UNION-FIND data structure. Meanwhile, each non-even blossom will correspond to a list in the SPLIT-FINDMIN data structure. When a blossom is no longer proper $L(v)$ will correspond to the last proper blossom x was a part of before it became even.

When a blossom is labeled as even through a grow step or after becoming a root of a search tree, we iterate over its vertices to join them into a single set with ID pointing to the blossom. After a new blossom with base b is created we iterate over all newly even vertices v and call JOIN(b, v). For all even subblossoms with base c we call JOIN(b, c). After these join operations all vertices of the new blossom are in a single set with ID corresponding to the blossom. It is important for us to change the label of odd subblossoms to even. At the beginning of the search and after the searched graph expands we create lists in the SPLIT-FINDMIN for all blossoms with their vertices in the blossom order. When an odd blossom expands we call SPLIT to divide its list into lists corresponding to the new blossoms. In order to find which blossom a vertex v belongs to we first call FINDLIST(v). If the blossom corresponding to the returned list is non-even it is the current blossom of v. If the blossom is even we need to call FIND(v) on the UNION-FIND data structure. Additionally, we will store costs associated with all non-even vertices which we will describe later.

During the search we maintain the following values:

- t_{now} – the current time, maintained by the array queue,
- $z_0(B)$ – the value of z_B at the time it became a proper blossom,

- $t_{proper}(B)$ – the time B became a proper blossom,
- $t_{even}(B)$ – the time B became an even blossom,
- $t_{odd}(B)$ – the time B became an odd blossom,
- $\Delta(B)$ – the sum of dual weight adjustment experienced by the vertices of B when they were a part of an odd blossom before B became proper or even,
- $y_0(v)$ – the value of y_v before all searches,
- $t_{search}(v)$ – the time the vertex v was added to the searched graph
- $\Delta(v)$ – the sum of distributions to vertex v before it was added to a s,earch graph,
- t_{outer} – the time the current outer boundary became the boundary,
- t_{inner} – the time the current inner boundary became the boundary,
- t_{whole} – the time the outer boundary became G – a special case,
- $z_{boundary}$ – the sum of weights of old blossom on the path containing the outer boundary (including itself) at the time it became the boundary.

For each shell we maintain a list of blossoms contained inside it to allow us to iterate over them. Each proper blossom has a pointer to its position in said list. We remove and add blossoms from said list whenever dissolve and blossom events happen. When a shell dissolves into another we join their lists.

With these values we can calculate the values of y_v and z_B at a time t_{now}. Let

$$D(v) = \Delta(v) + \max\{0, \min\{t_{now}, t_{whole}\} - t_{search}(v)\}$$

be the sum of old blossom distributions to v since the beginning of the iteration – consisting of distributions before it was added to search graph $\Delta(v)$ and distributions from the outer boundary after v was added to the search graph. Assuming $B_v = \text{FINDLIST}(v)$, we can calculate

$$y(v) = y_0(v) + D(v) + \begin{cases} \Delta(B) + t_{now} - t_{odd}(B) & \text{if } B_v \text{ is odd,} \\ \Delta(B) - (t_{now} - t_{even}(B)) & \text{if } B_v \text{ is even,} \\ \Delta(B) & \text{if } B_v \text{ is free.} \end{cases}$$

$$z(B) = z_0(B) + \begin{cases} -2(t_{now} - t_{odd}(B)) & \text{if } B \text{ is odd,} \\ 2(t_{now} - t_{even}(B)) & \text{if } B \text{ is even,} \\ 0 & \text{if } B \text{ is free.} \end{cases}$$

When calculating the slack of an edge in the search graph not contained within a blossom we need to include the weights of old blossoms containing it z_{old}. W

$$z_{old} = z_{outer} + \begin{cases} 0 & \text{if } G \text{ is the boundary,} \\ z_{boundary} - 2(t_{now} - t_{outer}) & \text{otherwise.} \end{cases}$$
$$slack(uv) = y(u) + y(v) - w(uv) + z_{old}$$

During each iteration, the shells are searched in descending order based on the number of exposed vertices they contain. Distributions from old blossom higher in the path influence the vertices of all lower blossoms. To calculate $\Delta(v)$ we need to know the sum of all distributions from old blossom containing v in previous searches. We use a data structure ADD-PREFIXSUM capable of calculating a prefix sum of an array and adding a value to an element at a specified index. In our case it is implemented as a Fenwick tree [12], which accomplishes both of these operations in $O(\log n)$ for an array of size n.

Before performing any searches we index undissolved shells with consecutive integers $0, 1, \ldots$ which we use as indices in the ADD-PREFIXSUM data structure. We denote the number of all distributions done to B and its ancestors on the current path since the start of the current iteration as $dist(B)$.

The costs in the SPLIT-FINDMIN data structure will help us maintain the slack of edges between even and non-even vertices. We want to keep the invariant that for each non-even vertex v the value $c(v)$ is equal to $\min_{u \text{ even}} slack(vu)$ shifted by some offset shared by all vertices of the blossom $B = L(v)$. Specifically, for each non-even vertex u in blossom $B = L(v)$ we maintain the following invariant:

$$\min_{u \text{ even}} slack(vu) = \begin{cases} c(v) - (t_{odd}(B) - \Delta(B)) & \text{if } B \text{ is odd} \\ c(v) - (t_{now} - \Delta(B)) & \text{if } B \text{ is free} \end{cases}$$

When v is odd the slack of edges connecting it to even vertices does not change with dual adjustments and when v is free it decreases by 1 with each dual adjustment, meaning that the invariant is maintained after dual adjustment without making changes to $c(v)$. We additionally remember which edge is responsible for the current cost $c(v)$.

We show how to update all the counters and data structures throughout the search starting with the beginning of the search.

The start of the search

First we calculate $z_{boundary}$. Let B be the outer boundary. We already know the sum of the dual weight before any searches and it is equal to $z_{path}(B)$. To include the distributions that happened to the ancestors of B in any searches that took place since then, we query the ADD-PREFIXSUM data structure and call the result Δ. The value of $z_{boundary}$ is equal to $z_{path}(B) - 2\Delta$.

We iterate over all blossom B in the searched graph to initialize their lists in the SPLIT-FINDMIN data structure and set $\Delta(B) \leftarrow 0$. If B is exposed we schedule a $grow(v, \emptyset)$ event at time 0 for the base v of B. For all vertices v we set $t_{search}(v) \leftarrow 0$ and $\Delta(v) \leftarrow \Delta$. For each boundary B we queue up the $dissolve\_shell(B)$ event at time $z_B/2$ as each dual adjustment decreases the value of z_B by 2. We do not do that if the inner boundary does not exist or the outer boundary is the whole graph G. If the outer boundary is G, we set $t_{whole} = 0$, otherwise we initialize it to ∞.

The $grow(v, e)$ event

Let B be the current blossom of v. If B is not free, it is already in the search structure, and we have nothing else to do. We proceed based on whether e is in the current matching.

If e is not provided or if it is currently part of the matching, B becomes even. We set e to be the backtrack edge of B and assign

$$t_{root}(B), t_{even}(B) \leftarrow t_{now}$$

In the UNION-FIND data structure, we link all vertices of B to v, which is the base of B according to our definition of the grow event. For every $u \in B$ we call SCHEDULE(u) to queue up event associated with these vertices.

If e is not $\emptyset$ and it is not in the matching, we label B as odd and the set values

$$t_{root}(B), t_{odd}(B) \leftarrow t_{now}$$

After which we call SCHEDULE(u) where u is the base of B.

The schedule procedure

The SCHEDULE(u) procedure schedules events associated with a vertex u which is either even or a base of an odd blossom. Let B_u be the blossom containing u.

If u is odd, we schedule the $dissolve(B_u)$ event at time $t_{now} + z(B_u)/2$, which is when z_B reaches 0. As B_u is odd, it is not exposed and as such u is matched to some vertex v belonging to B_v. We can continue the search by scheduling either a $grow(v, uv)$ or $blossom(uv)$ event based on whether v is free or even.

Now assume that u is even. We iterate over all neighbors v of u. Let B_v be the current blossom of v. We do different things depend on how u is labeled.

- If v is even we schedule an event $blossom(uv)$ at time $t_{now} + slack(uv)/2$ as that is when uv becomes tight.
- If v is odd, we update the cost $c(v)$ in the SPLIT-FINDMIN by calling DECREASE_COST($v, slack(uv) + (t_{odd}(B_v) - \Delta(B_v))$).

- When v is free, we try to decrease $c(v)$ to $slack(uv) + (t_{now} - \Delta(B_v))$ and if we're successful, we schedule $grow(v, uv)$ at $t_{now} + slack(uv)$ as that is when uv becomes tight and the search tree can be extended to B_v.

An implementation of SCHEDULE(u) is presented in Algorithm 5.

Algorithm 5 The SCHEDULE procedure

```
 1: procedure SCHEDULE(u)
 2:     if u is odd then
 3:         Let B_u be the current blossom of u
 4:         if B_u is not trivial then
 5:             Schedule dissolve(B_u) at time t_now + z(B_u)/2
 6:         end if
 7:         v ← MATE[v]
 8:         if v is free then
 9:             Schedule grow(v, uv) at time t_now
10:         else
11:             Schedule blossom(uv) at time t_now
12:         end if
13:     else                                                  ▷ u is even
14:         for neighbors v of u in the shell search graph do
15:             if FIND(u) = FIND(v) then
16:                 continue              ▷ u and v are in the same even blossom
17:             end if
18:             edge_slack ← slack(uv)
19:             if v is even then
20:                 Schedule blossom(uv) at time t_now + edge_slack/2
21:             else if v is odd then
22:                 DECREASE_COST(v, edge_slack + (t_odd(B_v) − Δ(B_v)))
23:             else
24:                 Let B_v be the current blossom of v
25:                 min_slack ← FIND_MIN(B_v) − (t_now − Δ(B_v))
26:                 if edge_slack < min_slack then
27:                     DECREASE_COST(v, edge_slack + (t_now − Δ(B_v)))
28:                     Schedule grow(v, (u, v)) at time t_now + edge_slack
29:                 end if
30:             end if
31:         end for
32:     end if
33: end procedure
```

The *blossom*(e) event

The BLOSSSOM(e) procedure, invoked for *blossom* events is presented in Algorithm 6. We follow the backtrack edges the same way we did in the blossom

algorithm by simultaneously moving two pointers until we either visit the same blossom twice or reach two distinct roots. If we reached different roots, an augmenting path is found and the SHELL_SEARCH procedure is finished. If not, a new blossom is created.

Let B be the new blossom and $B_1, \ldots, B_k$ its subblossoms. For each subblossom B_i we store its current dual weight which will be its starting weight when it becomes proper by setting $z_0(B_i) \leftarrow z(B_i)$. For previously odd subblossoms B_i we set $t_{even}(B_i) \leftarrow t_{now}$ and update the value $\Delta(B_i)$ by adding to it $(t_{now} - t_{odd}(B_i))$, which is how much time past since B_i last became odd. We link all vertices of B_i to its base and the base to the new base of B in the UNION-FIND data structure. We call SCHEDULE(v) for all vertices v of B_i which become even for the first time but only after updating all counters and data structures.

Algorithm 6 The blossom event implementation

```
1:  procedure BLOSSOM(uv)
2:      backtrack from u and v
3:      if an augmenting path has been found then
4:          Finish the SHELL_SEARCH procedure
5:      end if
6:      Otherwise a new blossom with subblossoms B_1, ..., B_k has been found
7:      Let B be the new blossom formed from B_1, ..., B_k
8:      Let b be the base of B
9:      Let to_schedule be an empy list
10:     for each B_i do
11:         z_0(B_i) ← z(B)
12:         t_even(B_i) ← t_now
13:         Let b_i be the base of B_i
14:         if B_i is odd then
15:             Label B_i as even
16:             Δ(B_i) ← Δ(B_i) + (t_now − t_odd(B_i))
17:             for each v ∈ B do
18:                 Add v to to_schedule
19:                 LINK(v, b_i)
20:             end for
21:         end if
22:         LINK(b_i, b)
23:     end for
24:     for each v in to_schedule do
25:         SCHEDULE(v)
26:     end for
27: end procedure
```

The $dissolve(B)$ event

We split the lists in the SPLIT-FINDMIN data structure so that they correspond to the vertices of subblossoms of B. The subblossoms are labeled and incorporated into the search tree the same way as we've done in the blossom algorithm search. Let C be a subblossom of B that now becomes proper. We set $t_{proper}(C) \leftarrow t_{now}$ and update the count $\Delta(C) \leftarrow \Delta(B) + (t_{now} - t_{odd}(B))$ to account for the dual adjustments while B was a proper odd blossom. We set t_{odd} or t_{even} to t_{now} if b was labeled accordingly.

If C has become odd, we schedule a $dissolve(b)$ event at $t_{now} + z_0(b)/2$. If it was labeled as even, we link all vertices v of C to its base in the UNION-FIND data structure and call SCHEDULE(v) to queue event associated with the newly even vertices. We do that at the end for all even vertices after updating counters and data structures for all new proper blossoms. If C has become free we find the edge with the smallest slack connecting a vertex v of C to some even vertex u using the SPLIT-FINDMIN data structure. According to our invariant we can calculate $slack(uv) = c(v) - (t_{now} - \Delta(C))$. With that we can schedule a $grow(v, uv)$ event at time $t_{now} + slack(uv)$.

The $dissolve\_shell(B)$ event

Assume that B is the outer boundary. The steps for the inner boundary are analogous except for a few corner cases. We start by recording the distributions to B in the ADD-PREFIXSUM data structure by adding $t_{now} - t_{outer}$ at the index corresponding to B.

The old blossom B is marked as dissolved, so it can be deleted at the end of the iteration. We check if B is the outermost undissolved old blossom. To be able to do that we maintain a pointer to said blossom. If B matches that pointer, the SHELL_SEARCH procedure is done. We follow the heavy children along the current path to find the new highest undissolved blossom and update our pointer.

Assume now that B has an undissolved ancestor C in the path. If the shell $C \setminus B$ has already been searched, we finish the search. If it is not the case, we expand the searched graph by adding vertices of $C \setminus B$. For each vertex $v \in C \setminus B$, we set $t_{search}(v) \leftarrow t_{now}$ and $\Delta(v) \leftarrow dist(C)$.

For all blossom D contained in $C \setminus B$ we set $\Delta(D) \leftarrow 0$ and create a corresponding list in the SPLIT-FINDMIN data structure.

For each exposed vertex v of $C \setminus B$ we schedule a $grow(v, \emptyset)$ event that happens at current time and establishes the blossom of v as a root of a new search tree.

Next we scan all new edges vu where $v \in B \setminus C$ and u is already a part of the searched graph. If u is even we schedule a $grow(v, uv)$ event that happens at time $t_{now} + slack(uv)$, which is when uv becomes tight.

If C is not equal to G, we schedule a $dissolve\_shell(C)$ event at time $t_{now} + z_C/2$. Otherwise, we set $t_{whole} \leftarrow t_{now}$. We set $t_{outer} \leftarrow t_{now}$ and recalculate $z_{boundary}$ as described above.

No matter which of the above cases happens, we append the lists of shell vertices and blossoms of B to the corresponding lists of C. We also remove B from the path by reassigning heavy child/parent pointers.

The end of the search

For each vertex v, we store the current value of y_v by setting $y'(v) \leftarrow y(v)$. This value might no longer be accurate by the end of the iteration, as more distributions take place in subsequent searches. To take that into the account we store the current value $\Delta(v) \leftarrow (\min\{t_{now}, t_{whole}\} - t_{search}(v))$. When all the searches have finished we can calculate the final value $y_v \leftarrow y'(v) + dist(B) - \Delta(v)$, where B is the lowest old blossom containing v at the start of the iteration.

For each blossom B we delete its corresponding list in the SPLIT-FINDMIN data structure and store the final value $z_B \leftarrow z(B)$. If its dual weight z_B is equal to 0, we expand it.

We update the dual weights for the current boundaries to reflect the distributions that took place. For the current outer boundary B, we update $z_B \leftarrow z_B - 2(t_{now} - t_{outer})$ and add $t_{now} - t_{outer}$ at B's index in the ADD-PREFIXSUM data structure. We do the same for the inner boundary.

Time complexity

The SPLIT-FINDMIN data structure can perform m DECREASE_COST operations in $O(m\alpha(m, n))$ time as we will show in Theorem 3.4.2.

The running time of the array priority queue with k events and maximum time $t_{\max}$ is trivially $O(k + t_{\max})$. It can be shown that a search of shell $C \setminus D$ decreases the value $yz(C \setminus D)$ by $n_C - n_D$ with a proof similar to that of Lemma 3.5.1.2. Together with Lemma 3.2.0.2 this shows that $t_{\max} = O(n_C - n_D)$.

The time needed to maintain the ADD-PREFIXSUM data structure during a single iteration of PATH(B) is $O(n_B \log n_B)$, which is enough for the desired time bound. This time is not included in the original analysis [16] suggesting the author had a different implementation in mind. I suspect such an algorithm most likely takes advantage of the fact that a distribution of an old blossom keeps the slack of edges it contains the same while increasing the slack of edges that cross between the inside of the blossom and outside. Each shell is only searched once during an iteration and during a single execution of the SHELL_SEARCH procedure, we only check the slack of edges within some set of consecutive shells. Instead of calculating how the exact value of y changes as a result of distributions, we can calculate the slack of an edge by checking how much time it spent connecting a vertex of the current graph and one outside it. We can do that using the t_{search} values while also accounting for the case when G is the outer boundary. The actual value of y is only calculated at the end of iteration. Distributions to shells are stored in an array for which prefix sums are calculated. Those sums are used to update the value of y. Personally, I have

found approach using ADD-PREFIXSUM easier to follow as it explicitly calculates the current values of y.

With that in mind the searches of all shells in a single iteration of PATH(B) except for the innermost shell C can be done in time $O(m_B\alpha(m_B, n_B))$.

All that remains is the time needed to search the innermost shell C. When it has at least 2 exposed vertices the same array priority queue is sufficient. When that number is equal to one, similar calculations as in Lemma 3.5.1.2 show the blossom C dissolves after n_C dual adjustments. As the number of old blossoms can reach $O(n_B)$, it results in quadratic running time in the worst case. When $m = \Omega(n^{5/4})$ this is enough for our desired time bound. In general, it is enough to perform the search of the C shell in $O(\sqrt{n}m\alpha(m, n))$ which can be done with search procedures of various blossom algorithm implementations.

For simplicity, our implementation omits this case and uses the same simple array priority queue in all searches. This introduces some issues with setting the size of the array. Initially, we have used a dynamic array which resizes when events at later times are scheduled. Experiments have shown that this approach causes large running time spikes at higher values of N as times can get quite large. The events at larger times never happen as the search finishes earlier. Another idea is to use the mentioned bounds on the number of steps to limit the size of the array, but we have been unsuccessful in doing so. Instead, we divide events into buckets based on their time. The i-th bucket contains events with times from $[(i-1)n, in-1]$. We move from bucket to bucket as needed. The events from the current bucket are put into an array of size at most n just like in the simple array queue. When we reach the end of the current bucket, we clear it and fill it with events from the next bucket. This approach proved to be faster and the running time no longer spiked at higher values of N, instead it scaled with $\log N$ as we expected.

3.4 Splitting list

We now present an implementation of the previously used splitting list data structure as described by [13]. Remember that the splitting list data structure operates on elements from a set $\{0, 1, \ldots, n\}$. Every element x can be contained in at most one list $L(x)$ and has an associated cost $c(x)$, which can be infinite. The cost $c(L)$ of a list is the smallest cost of an element in L. It supports following operations:

- INITIALIZE($x_1, \ldots, x_l$) initialize a list of elements,
- DECREASE_COST(x, d) update $c(x)$ to $\min\{d, c(x)\}$,
- SPLIT(x) split $L(X)$ into two lists L_1 and L_2 where L_1 contains all elements of $L(x)$ until and including x and L_2 contains the remaining elements in the order they appear in $L(x)$,
- FIND_MIN(L) return $c(L)$ and the element of L which achieves the minimum,

- FIND_LIST(x) return the current list of x.

For our analysis, we define the Ackermann's function $A(i, j)$ along with inverse functions $a(i, n)$ and $\alpha(m, n)$ as follows:

$$\begin{aligned}
A(i,0) &= 2 && \text{for } i \geq 1 \\
A(1,j) &= 2^j && \text{for } j \geq 1 \\
A(i,j) &= A(i-1, A(i,j-1)) && \text{for } i \geq 2 \text{ and } j \geq 1 \\
a(i,n) &= \max\{j : 2A(i,j) \leq n\} && \text{for } n \geq 4 \\
\alpha(m,n) &= \min\left\{i : A\left(i, \left\lfloor \frac{m}{n} \right\rfloor\right) \geq n\right\} && \text{for } m \geq n
\end{aligned}$$

It is handy for us to be able to calculate $A(i, j)$ in constant time. We only need to calculate values below the n. Cases when $j = 0$ or $i = 1$ can be calculated on the fly by simply returning 2 or calculating 2^j using a bit shift operation. For the remaining cases, we store all values $A(i, j) < 10^9$ as there are only 5 such cases.

The splitting list data structure works recursively. Each list has a set *level* $i \in \{1, 2, \dots\}$. A list L is divided into a *head* and a *tail*, where the head stores a starting fragment of L and the tail the stores the remaining elements. Either of the fragments can be empty.

Both the head and the tail are divided into *superelements*. A superelement of *rank* $j \geq 0$ in a list L of level i consists of $2A(i, j)$ consecutive elements of L. A superelement e of the head has a maximum rank $a(i, n_e)$ where n_e is the number of elements from the beginning of the head to the last element of e. This means that the head consists of superelements of non-decreasing ranks and at most three remaining elements at the start of the head which do not belong to any superelement which we call *singletons*. We call a maximal sequence of superelements of the same rank a *sublist* of L. We say that a sublist containing superelements of rank j is of that rank. A sublist L' of a level i list L is represented by a list of level $i-1$. The tail is partitioned similarly with the exception that the ranks of elements are non-increasing.

Along with the size n of the universe of elements, when initializing the splitting list data structure we provide the chosen level of lists used to store the original elements which we denote as $i_{\max}$. Performing operations on the splitting list data structure consists of interacting with these top level lists, which then refer to their lower level sublists.

For every list we store:

- the lists ID for top level lists,
- its level,
- the list of all its elements,
- its current cost $c(L)$,

- two lists of sublists for the head and tail,
- two lists of singletons for the head and tail,
- pointer to a sublist if it represents one.

For every sublist we store:

- its rank,
- pointer to its list,
- the splitting list containing its superelements,
- pointer to its position int the list of sublists of its parent.

In order to index arrays using superelements we associate them with one of their elements (in the head it is their last element and in the tail the first one). We make use of two-dimensional arrays of size $i_{\max} \cdot n$. For each element x at its corresponding level we store:

- the value $e(x)$ of its superelement if it belongs to one or -1 if x is a singleton,
- the cost $c(x)$ of the element – for superelement this is the minimum cost of one of its elements,
- the pointer to the list $L(x)$ if x is a singleton,
- a doubly-linked list of all elements comprising the superelement if x corresponds to one.

We maintain the value of $c(L)$ so that it is always correct, allowing us to execute FIND_MIN(L) in constant time.

We now describe how to implement the splitting list operations. The function FIND_LIST(x) for an element of a level i list checks if x is a singleton in which case it has a pointer to its list. If x is part of a superelement we recursively call FIND_LIST($e(x)$) at level $i-1$. The returned list has a pointer to the sublist containing $e(x)$, which itself points to the list $L(x)$. Total time taken is $O(i_{\max})$.

The function DECREASE_COST(x, d) works similarly to FIND_LIST(x) – it returns the list containing x. If x is a singleton it simply updates $c(x)$ and $c(L(x))$. When x is a part of a superelement $e(x)$ it calls DECREASE_COST($e(x), d$) recursively and uses the returned pointer to $L(x)$ to update the costs. The time complexity is again $O(i_{\max})$.

One of two internal functions INITIALIZE_HEAD or INITIALIZE_TAIL can be used to initialize a list L. The function INITIALIZE_HEAD works by scanning elements from the right to left and dividing them into superelements of maximum rank. These superelements are then divided into sublists which are initialized by recursively calling INITIALIZE_HEAD at the lower level. The scan at one level can be done in linear time. If the list has l elements, then all of its sublists have

no more than $\frac{l}{4}$ elements which means the total time for INITIALIZE_HEAD(L) is $O(l)$. The INITIALIZE_TAIL function works similarly with the exception that the scan proceeds from left to right.

The last operation to implement is SPLIT(x). If x is a singleton, we first check if it is in the head or tail. Assume x is a head singleton. We split the head singleton list at x and create a new list which consists solely of head singletons up to x. Similarly, if x is a tail singleton, we create a new list with just tail singletons.

Assume now that x is not a singleton in a level i list and belongs to a superelement $e(x)$ in sublist S inside the list L. We perform two splits on S to divide it into three parts: S_1 containing superelements before $e(x)$, S_2 containing elements after $e(x)$ and a sublist containing solely $e(x)$ which we can discard. Assume S is in the head, the case when S is part of the tail is analogous. We create two new lists L_1 and L_2. The head of L_1 consists of the sublists in the head of L before S along with S_1. The tail of L_1 is created by calling INITIALIZE_TAIL on elements of $e(x)$ until x. The tail of L_2 is just the tail of L and its head is initialized with a call to INITIALIZE_HEAD on elements of $e(x)$ after x. We then update the cost of L_1 and L_2 by checking costs of all sublists and singletons. It is easy to see that the new lists are partitioned consistently with the previously described rules.

Theorem 3.4.1. *The time taken to perform all split operations is $O(na(i, n))$.*

Proof. First we estimate the time at the top level. A single list has at most $a(i, n)$ sublists as each has a stores superelements of different ranks and $a(i, n)$ is the maximum rank possible. The number of singletons in a list is at most 6. Using the maintained pointers splits on sublist lists, element lists take constant time. Updating pointers in sublists and calculating cost for new lists takes time $a(i, n)$. All that remains is the time spent in INITIALIZE_HEAD and INITIALIZE_TAIL and on splitting lists of elements comprising superelements. We perform the split in linear time by finding x in the list of elements comprising $e(x)$. The initialization also takes linear time. Each time an element x is in an initialization the rank of its superelement decreases or it becomes a singleton meaning time needed for all initialization calls is $O(na(i, n))$.

We show by induction that time needed for splits on level i lists on a universe of k elements is $O(ka(i, k))$. For $i = 1$ all sublists contain at most one superelement, so every recursive split takes constant time, meaning the total time for all splits is $O(ka(1, k))$.

Suppose $i > 1$. A rank j sublist of a level i list contains at most

$$\frac{2A(i, j + 1)}{2A(i, j)} \leq A(i, j + 1)$$

as there are fewer than $2A(i, j + 1)$ elements remaining at the time of partition as otherwise a rank $j + 1$ elements would have been created. The time spent in recursive calls per a level j superelement is

$$\begin{aligned} a(i-1, A(i, j+1)) &= a(i-1, A(i-1, A(i, j))) \\ &= \max\{j : 2A(i-1, j) \leq A(i-1, A(i, j))\} \\ &< A(i, j) \end{aligned}$$

There are $n/2A(i, j)$ rank j superelements at level i, so the total time spent on them is $O(n)$. As the maximum rank of a superelement is $a(i, n)$, the total time complexity is $O(na(i, n))$.

□

Theorem 3.4.2. *Assuming we know the number $m \geq n$ of* DECREASE_COST *operations on universe of n elements in advance, we can choose the value $i_{\max}$ such that the total running time of the splitting list data structure is $O(m\alpha(m, n))$.*

Proof. Choose $i_{\max} = \alpha(m, n)$. Total time needed is $O(mi_{\max} + na(i_{\max}, n))$. From the definition of α, we know that $A(\alpha(m, n), \lfloor \frac{m}{n} \rfloor) \geq n$. This means that $a(i_{\max}, n) = \max\{j : 2A(\alpha(m, n), j) \leq n\} < \lfloor \frac{m}{n} \rfloor$ and the final complexity is $O(m\alpha(m, n))$. □

Our splitting list data structure implementation differs from the one described here in that it stores an additional value associated with a cost that can be specified during a DECREASE_COST operation and is returned in FIND_MIN instead of the element. In our case the value corresponds to some edge and the cost is determined by its slack.

Let us mention that there exist solutions to the SPLIT-FINDMIN problem with better theoretical running time. The data structure from [31] works in $O(m \log \alpha(m, n) + n)$ and the one from [34] solves it in $O(m + n)$ for integer costs.

3.5 Complexity

We now provide a proof of complexity adapted from [16].

Theorem 3.5.1. *The time complexity of the* MATCH *procedure is $O(n^{3/4}m)$.*

Proof. We first show the time complexity for PATH(B). To do that we need a few lemmas with the first one describing a special property of shells.

Lemma 3.5.1.1. *(Shell lemma) Assume we have a perfect structured matching with dual weights y and z. Let $B \setminus C$ be a shell.*

The maximum weight perfect matching in a shell $B \setminus C$ exists and has weight $yz(B) - yz(C)$. If C is a subblossom of B then $yz(B) - yz(C) = yz(B \setminus C)$.

Proof. First assume that C is a trivial blossom for vertex v. Construct a graph B' by adding a vertex v' to B connected with an edge vv' with zero weight.

The structured matching implies that there exists a perfect matching on $B \setminus v$ which can be obtained by changing the base of B to v. We can extend it to B' by adding vv'. Call the resulting matching M'.

We construct weights y' and z' which are identical to y and z with the exception that:

$$y'_{v'} = -y_v$$
$$z'_B = \sum_{B \subseteq D} z_D$$

The new duals are tight and dominating and with M' constitute a perfect structured matching on B', meaning that M' is a maximum weight perfect matching on B' with weight $y'z'(B') = yz(B) - y_v$. By removing vv' from M' we obtain a perfect matching M on $B \setminus v$ which is a maximum weight perfect matching (otherwise we could find a perfect matching on B' which weighs more than M').

In case when C is non-trivial start by choosing a vertex $v \in C$. We can again obtain a perfect matching on $B \setminus v$ by using the blossom structure. The resulting matching is also perfect on $C \setminus v$ and $B \setminus C$ which can be seen by examining the matching after changing the base of B to v – no matched edges cross between C and $B \setminus C$ and v is not matched to any of vertices in B.

The resulting perfect matching on $B \setminus C$ weighs

$$(yz(B) - y_v) - (yz(C) - y_v) = yz(B) - yz(C)$$

A larger matching would give a larger matching on $B - v$.

If C is a child of B then

$$\begin{aligned} yz(B) - yz(C) = &\left(\sum_{v \in B} y_v + \sum_{S \subset B} z_S \left\lfloor \frac{n_S}{2} \right\rfloor + \sum_{S \supseteq B} z_S \left\lfloor \frac{n_B}{2} \right\rfloor \right) - \\ &\left(\sum_{v \in C} y_v + \sum_{S \subset C} z_S \left\lfloor \frac{n_S}{2} \right\rfloor + \sum_{S \supseteq C} z_S \left\lfloor \frac{n_C}{2} \right\rfloor \right) \\ = &\sum_{v \in B \setminus C} y_v + \sum_{S \subset B \setminus C} z_S \left\lfloor \frac{n_S}{2} \right\rfloor + z_C \left\lfloor \frac{n_C}{2} \right\rfloor + \\ &\sum_{S \supseteq B \setminus C} z_S \left\lfloor \frac{n_{B \setminus C}}{2} \right\rfloor - z_C \left\lfloor \frac{n_C}{2} \right\rfloor \\ = &\, yz(B \setminus C) \end{aligned}$$

□

Lemma 3.5.1.2. *During the execution of* PATH(B) *the value of* $yz(B)$ *decreases by at most* n_B.

Proof. If $B = G$, it follows from the fact that the duals are almost optimal as shown by Equation (3.1). Assume now that B is an old blossom. The dual objective never increases according to Lemma 3.2.0.1. At the start of PATH(B) the value of $yz(B)$ is at most $y^0z^0(B)$. It is sufficient for us to show that at the end of PATH(B) the dual weight satisfy:

$$yz(B) \geq y^0z^0(B) - n_B$$

We choose a vertex $v \in B$. Let M^* be a maximum weight perfect matching on the shell $B \setminus v$. We get

$$w(M^*) \geq y^0z^0(B) - y^0_v - n_B$$

which follows from the Lemma 3.5.1.1 and reasoning analogous to proof of Equation (3.1). By constructing dual weights similarly to the proof of Lemma 3.5.1.1 it can be shown that

$$yz(B) - y_v \geq w(M^*)$$

Using the fact that $y_v \geq y^0_v$ we get

$$yz(B) - y_v \geq w(M^*) \geq y^0z^0(B) - y^0_v - n_B \geq y^0z^0(B) - y_v - n_B$$

By adding y_v to both sides we obtain the desired inequality. □

Lemma 3.5.1.3. *For any $\varepsilon > 0$ the number of iterations in* PATH(B) *with at least n^ε_B exposed vertices after the* AUGMENT_PATHS *procedure is $O(n^{1-\varepsilon}_B)$.*

Proof. Assume B is a blossom. Let f be the number of exposed vertices after the AUGMENT_PATHS procedure of an iteration. We call a shell *small* when it has at most 2 exposed vertices and *big* otherwise. We consider an iteration with $f \geq n^\varepsilon_B$.

If at least $\frac{1}{2}f$ exposed vertices are in big shells. There are no augmenting paths in any of the shells after the AUGMENT_PATHS procedure. This means that any search that finds an augmenting path performs at least one dual adjustment. This adjustment decreases the dual weight of all exposed vertices. A shell might not have the dual weights of its exposed vertices adjusted if another shell dissolved into after the same dual adjustment that found an augmenting path. Because we search the shell in the order of decreasing number of exposed vertices, such a shell has to be adjacent to a shell with at least as many exposed vertices that has been found to contain an augmenting path. This means that at least one third of exposed vertices in big shells had their dual weights adjusted by at least 1. The big shells where dual adjustments took place contain at least $\frac{1}{6}f$ exposed vertices. By Lemma 3.2.0.2, the function $yz(B)$ decreases by at least $\frac{1}{12}f \geq \frac{1}{12}n^\varepsilon_B$. By Lemma 3.5.1.2, the number of such iterations is at most $O(n^{1-\varepsilon}_B)$.

In the remaining case, at least $\frac{1}{2}f$ exposed vertices are in small shells – those with at most 2 exposed vertices. Assume that that the iteration we're considering is not the first iteration. Any shell that had exposed vertices at the start

of the iteration had to contain an augmenting path by the stopping condition of SHELL_SEARCH. The procedure ends when an augmenting path is found or the shell dissolves either into another one that has already been searched or the outer boundary is the outermost undissolved old blossom and the shell ceases to exist. This means that all the shells with exposed vertices had their number of exposed vertices decreased by at least 2. There are no less than $\frac{1}{4}f$ such shells. This means that at least $\frac{1}{2}f$ were matched in the AUGMENT_PATHS procedure. Meaning the number of exposed vertices was multiplied by a number less than $\frac{2}{3}$. This means that the number of such iterations is $O(\log n_B)$. □

We can now show how much time is needed for PATH(B).

Lemma 3.5.1.4. *The time complexity of* PATH(B) *is* $O(n_B^{3/4} m_B)$.

Proof. The algorithm executes $O(\sqrt{n_B})$ iterations. From the Lemma 3.5.1.3 there are $O(\sqrt{n_B})$ iterations with at least $\sqrt{n_B}$ exposed vertices. Since each iteration matches at least one edge there are $O(\sqrt{n_B})$ iterations with fewer than $\sqrt{n_B}$ exposed vertices. As we've discussed before, all executions of the SHELL_SEARCH procedure can be accomplished in $O(m_B \alpha(m_B, n_B))$ per iteration except the last shell, which itself can be dissolved in desired time.

We now focus on the augmentation step. The contraction of the graph and augmentation can be done in $O(n_B + m_B)$ the way we've described. According to Lemma 3.5.1.3 there are $O(n_B^{1/4})$ iterations with at least $n_B^{3/4}$ exposed vertices. The MV algorithms runs in $O(\sqrt{n}m)$ time [29]. This means that we spend $O(n_B^{3/4} m_B)$ time in these iterations. In the iterations with less than $n_B^{3/4}$ exposed vertices less than $n_B^{3/4}$ new vertices are matched. We make use of the fact that the MV algorithm works in phases each of which takes $O(m)$ time and matches at least 1 new vertex. By providing a partial matching consistent with current matching, we can complete augmentation in these cases in $O(n_B^{3/4} m_B)$. □

Knowing that the execution of PATH(B) takes $O(n_B^{3/4} m_B)$ time we show that the *match* procedure runs in $O(n^{3/4}m)$ time.

Let i be a non-negative integer. Consider the set $\mathcal{B}_i$ of the roots B of major paths in T' such that $\frac{n}{2^{i-1}} > n_B \geq \frac{n}{2^i}$.

There are no two roots C and D in $\mathcal{B}_i$ such that D is a descendant of C. Assume otherwise and let P_C be the heavy path starting in C. D has to be a descendant of a non-heavy child of some blossom C' in P_C so $n_D \leq \frac{n_{C'}}{2} \leq \frac{n_C}{2} < \frac{n}{2^i}$ which contradicts the definition of $\mathcal{B}_i$.

This means that any vertex v belongs to at most one blossom in $\mathcal{B}_i$ and any edge is contained in at most one of such blossoms so $\sum_{B \in \mathcal{B}_i} m_B \leq m$. We sum up the time spent in PATH(B) for all $B \in \mathcal{B}_i$:

$$\sum_{B \in \mathcal{B}_i} n_B^{3/4} m_B < \sum_{B \in \mathcal{B}_i} \frac{n^{3/4}}{2^{3(i-1)/4}} m_B = \frac{n^{3/4}}{2^{3(i-1)/4}} \sum_{B \in \mathcal{B}_i} m_B \leq \frac{n^{3/4} m}{2^{3(i-1)/4}}$$

By summing over i we obtain the final complexity $O(n^{3/4}m)$.

3.6 Correctness

The MATCH procedure finishes when all old blossoms have been dissolved and the current matching M is perfect. It also produces dual weights y and z with corresponding blossoms that together meet the conditions of Theorem 2.1.3, meaning M is a maximum weight perfect matching.

To show that MATCH terminates it is enough to show that each call to PATH(B) terminates. We show that all old blossoms on the heavy path P_B dissolve. Each shell that contains exposed vertices gets searched. The search ends when it is found to contain an augmenting path, dissolves into another shell with an augmenting path or dissolves as the outermost undissolved shell. The shells with augmenting paths have their number of exposed vertices decreased in the next iteration by the AUGMENT_PATHS procedure. Because the edges of the matching are contained within blossoms and each blossom has an odd number of vertices, as long as there are undissolved old blossoms, there are exposed vertices. This means that in each iteration of PATH some shell search takes place that results in either old blossoms dissolving or the number of exposed vertices decreasing. As there is a finite number of old blossoms and vertices, this can not go on forever and PATH has to finish.

□

Chapter 4

Experimental results

All the algorithms described in the previous chapter have been implemented as part of the KOALA library [22] built on top of NetworKit(TODO cite).

The implementations are contained in following files:

- `include/matching/MaximumMatching.hpp` contains declarations of classes corresponding to maximum matching algorithms, namely:
  - `MaximumWeightMatching` – the base class for MWM and MWPM algorithms,
  - `BlossomMaximumMatching` – an abstract base class for implementations of the blossom algorithm. Responsible for the common operations shared by all implementations – building the search tree, creating blossoms, augmenting paths, expanding blossom and others. It relies on call virtual methods implemented by children classes to perform implementation-specific parts of the algorithm. One such part is calculating the δ_i values during dual weight adjustment. Another one is finding useful edges to perform the next step. After each step of the algorithm, specific virtual methods are called to allow the implementations to update their specific data structures.
  - `EdmondsMaximumMatching` – the Edmonds' original $O(n^2m)$ version of the blossom algorithm [8],
  - `GabowMaximumMatching` – the Gabow's $O(n^3)$ implementation of the blossom algorithm [20],
  - `GalilMicaliGabowMaximumMatching` – the $O(nm \log n)$ implementation of the blossom algorithm of [21],
  - `GabowScalingMatching` – the $O(n^{3/4}m \log N)$ scaling algorithm [16],
  - `MaximumCardinalityMatching` – the base class for MCM algorithms,
  - `MicaliVaziraniMatching` – the class for the implementation of the Micali-Vazirani maximum cardinality matching algorithm [29].

- `include/matching/structures` directory contains the implementations of various data structures utilized across the implemented algorithms:
  - `ArrayPriorityQueue.hpp` – contains the implementation of a simple array priority queue used by the scaling algorithm,
  - `ConcatenableQueue.hpp` – contains the implementation of concatenable queues,
  - `FenwickTree.hpp` – contains an implementation of the ADD-PREFIXSUM data strucutre based the Fenwick trees,
  - `HeapWithRemove.hpp` – contains a simple implementation of an array based heap with ability to remove elements, used by the pq_1 priority queue implementation,
  - `PriorityQueues.hpp` – contains the implementations `PriorityQueue1` and `PriorityQueue2` of the pq_1 and pq_2 priority queues,
  - `SplitFindMin.hpp` – contains the splitting list implementation of the SPLIT-FINDMIN data structure,
  - `UnionFind.hpp` – contains the implementation of the UNION-FIND data structure.
- `cpp/matching` contains the implementation of the classes defined in the `MaximumMatching.hpp` header file,
- `test/testMaximumMatching.cpp` contains simple unit tests for the implemented algorithms,
- `benchmark/benchmarkMaximumMatching.cpp` contains the program used to benchmark the various algorithms on provided graphs with various options.

All tests were conducted on a personal computer equipped with an Intel Core i7-6700HQ processor running at a clock speed of 2.60 GHz, using the Ubuntu 22.04.4 LTS operating system.

We've made use of test generators provided as part of the 1st DIMACS implementation challenge [6].

Random instances

The first class of graphs we've tested are random sparse and dense graphs generated using the `random.c` generator from the DIMACS challenge written by C. McGeoch. The program generates a random graph with a provided number of nodes and edges with weights uniformly chosen from a specified range $[1, N]$. It is also possible to specify a seed used to generate random numbers. For all of our test we've set the maximum cost at 2^{16}. For each pair of values n and m we've generated 3 different graphs using random seeds and then average the running time.

n	m	E	G	M	S
1000	2000	0.15	0.17	**0.11**	0.20
1000	4000	0.28	0.25	**0.19**	0.25
1000	8000	0.57	0.39	0.33	**0.29**
2000	4000	0.59	0.76	**0.43**	0.44
2000	8000	1.17	1.22	0.77	**0.59**
2000	16000	2.14	1.75	1.37	**0.69**
4000	8000	2.39	3.33	1.86	**1.29**
4000	16000	4.54	5.35	3.47	**1.72**
4000	32000	8.61	8.15	6.56	**2.73**
8000	16000	9.48	14.91	11.64	**4.05**
8000	32000	18.30	27.09	22.72	**5.41**
8000	64000	34.24	50.38	41.95	**8.77**
16000	32000	43.03	93.19	73.50	**11.72**
16000	64000	88.41	171.60	145.75	**16.00**
16000	128000	149.67	294.83	264.03	**33.65**
32000	64000	232.85	569.88	410.78	**44.63**
32000	128000	414.49	1026.40	790.05	**52.79**
32000	256000	611.80	1652.81	1355.59	**125.09**

Table 3: Results on random sparse graphs.

Our set of sparse graphs consists of graphs with varying values of n and $m = an$ where $a \in \{2, 4, 8\}$. The results are presented in Table 3. We use the letter E to refer to our implementation of Edmonds' algorithm, G to Gabow's blossom algorithm, M to Galil, Micali and Gabow's algorithm and S to Gabow's scaling algorithm. We will refer to them using those letters. As our set of dense graphs we've chosen random graphs with $n \in \{1000, 2000, 3000, 4000\}$ and containing approximately 20%, 40% and 60% possible edges with results presented in Table 4. In both cases the scaling algorithm performs the best. Among our implementations of the blossom algorithm, M runs fastest for some smaller test, but gets overtaken by E for larger instances. The G implementation is consistently the slowest for sparse graphs and on par with M for dense instances.

We plot the running time of tested algorithms in relation to n for sparse graphs with $m = 4n$ and dense graphs containing roughly 20% of possible edges. For sparse graphs, the running time of blossom algorithm implementation is roughly quadratic, indicating that that the G and E variant perform better than their theoretical worst case complexity would suggest. Meanwhile, the running time of the scaling algorithm S seems to align with its $n^{7/4}$ complexity. For dense instances, gauging the asymptotic running time seems to be harder with the blossom implementations taking roughly n^3 time.

n	m	E	G	M	S
1000	100000	5.65	3.51	4.08	**0.82**
1000	200000	7.35	7.26	6.87	**1.62**
1000	300000	8.04	10.26	10.32	**2.26**
2000	400000	25.09	40.97	43.96	**4.47**
2000	800000	31.16	60.80	55.41	**8.14**
2000	1200000	40.67	87.33	87.49	**12.77**
3000	900000	82.91	149.13	160.21	**12.36**
3000	1800000	84.38	205.11	214.92	**17.59**
3000	2700000	102.76	286.73	278.40	**50.79**
4000	1600000	139.25	346.93	383.64	**23.99**
4000	3200000	151.17	505.38	503.95	**38.57**
4000	4800000	233.07	700.78	775.19	**99.25**

Table 4: Results on random dense graphs.

Perfect matching

We repeat the same tests, this time running the perfect matching versions of our implementations. The tests are generated similarly as in the previous section, with the difference that we make sure that the generated instances contain perfect matchings by running our implementation of the Micali-Vazirani maximum cardinality algorithm. The approach was infeasible for some larger sparse instances, in which case we've used our own random test generator that ensures that the generated graph contains a perfect matching by first generating a random permutation $v_1, v_2, \dots v_n$ of vertices $[1, n]$ and adding edges $v_1v_2, v_3v_4, \dots, v_{n-1}v_n$, with the rest of edges chosen randomly.

We've tested our blossom algorithm implementation with both the empty and greedy initialization strategies. The results using the greedy initialization are marked with a G subscript. Additionally, all algorithms are marked with * to indicate that the perfect matching variant is used.

The results of our experiments on random sparse and dense graphs containing perfect matchings are presented in Table 5. As in the case of non-perfect matchings, the scaling algorithm S comes out on top. The margin is even larger most likely to the fact it is designed to find maximum perfect matching and finding a non-perfect matching requires a reduction that doubles the number of nodes and edges. For all blossom algorithm implementations, the greedy initialization resulted in a shorter running time. The speedup ranged from around 30% for the E algorithm and as much as 2-4 times for the G and M implementations.

The influence of edge weight ranges

The previous tests have neglected to account for one of the main contributing factors in the running time of the scaling algorithm - the maximum edge weight N. To test how different values of N affect the running time of the S algorithm,

n	m	E^*	E^*_G	G^*	G^*_G	M^*	M^*_G	S^*
1000	4000	0.32	0.22	0.27	0.13	0.19	**0.08**	0.13
1000	8000	0.54	0.36	0.36	0.16	0.33	**0.13**	**0.13**
2000	8000	1.36	0.90	1.25	0.65	0.80	0.33	**0.29**
2000	16000	2.11	1.43	1.63	0.74	1.35	0.51	**0.30**
4000	16000	5.52	3.91	5.66	3.07	3.55	1.38	**0.73**
4000	32000	9.15	5.91	7.75	3.42	6.61	2.45	**0.93**
8000	32000	21.34	15.33	29.52	13.81	22.47	8.85	**2.24**
8000	64000	35.71	23.37	48.09	16.29	42.31	15.19	**3.31**
16000	64000	100.73	74.28	184.69	87.47	147.05	54.72	**7.66**
16000	128000	156.35	106.55	298.43	107.97	266.19	93.96	**11.38**
32000	128000	512.00	354.36	1166.83	594.43	767.65	283.84	**26.70**
32000	256000	707.58	542.16	1655.12	727.57	1381.63	528.16	**21.80**
n	m	E^*	E^*_G	G^*	G^*_G	M^*	M^*_G	S^*
1000	100000	4.79	3.75	3.54	0.74	4.21	1.27	**0.41**
1000	200000	6.49	4.47	7.11	1.67	6.60	2.11	**0.70**
1000	300000	8.14	5.68	10.98	2.72	10.29	3.32	**0.89**
2000	400000	28.21	22.42	41.09	10.21	44.07	13.51	**1.93**
2000	800000	33.37	26.92	62.93	16.37	56.31	18.27	**2.83**
2000	1200000	40.55	30.74	86.72	22.38	84.48	26.57	**4.20**
4000	1600000	134.92	110.14	347.76	87.46	389.26	114.10	**10.89**
4000	3200000	157.18	111.22	494.36	127.89	504.29	147.19	**16.26**
4000	4800000	218.07	143.22	718.58	180.23	786.43	227.36	**32.61**

Table 5: Results on random sparse and dense graphs when looking for a perfect matching.

we've conducted a series of tests on random sparse graphs with $m = 4n$ and values of N ranging from 1 to 10^7. We've compared both the perfect and non-perfect variant of S to the E algorithm.

The results can be found in Table 6. We can observe that the running time of the blossom algorithm implementation E is influenced slightly by the value of N but seems to level off at higher values of N. Intuitively, we speculate that the initial increase in weight range causes a rise in complexity, but as the weights get larger, especially compared to the number of edges, the higher weight range doesn't meaningfully increase the complexity. As we expect from its time complexity, the running time of the S algorithm seems to scale with $\log N$, which can be seen in Figure 4.

DIMACS instances

We next turn to more structured instances. The tests were generated using the following programs:

- `t.f` written by N. Ritchey and B. Mattingly. For a parameter k it generates a sequence of k triangles. Consecutive triangles are connected with a single edge. The configuration tends to generate a lot of blossoms according to the comments provided with the generators.
- `tt.f` written by N. Ritchey and B. Mattingly. Similar to `t.f` with the difference that each pair of consecutive triangles is connected with 3 separate edges.
- `hardcard.f` written by B. Mattingly. It generates graphs that have been shown in [14] to be hard for Edmonds' maximum cardinality matching algorithm.

All of these instances contain perfect matching, so we've run both variants of the algorithms. For perfect matching we've opted to only test with greedy initialization. We ran each algorithm 3 times and average the time spent.

The results for `t.f` are presented in Table 7. Once again, S is by far the fastest. Of the blossom algorithm implementations E is the clear winner with G and M following. The results are similar for `tt.f` with the exception that the greedy initialization tends to find the perfect matching immediately.

In the case of `hardcard.f`, as observed in Table 8, the S algorithm once again comes out on top. What is unusual is that G outperforms both E and M, while in most of the previous tests it was consistently the slowest.

4.0.1 Conclusions

We have found our implementation of the scaling algorithm to be superior to our various implementations of the blossom algorithm. We have found that better theoretical running time doesn't necessarily result in faster execution in practice. The asymptotic time of of the blossom implementation is better than the theoretical one, suggesting the worst cases rarely happen.

n	m	N	E	S	E^*_G	S^*
1000	4000	1	0.07	**0.01**	**0.01**	**0.01**
1000	4000	10	**0.06**	0.07	**0.03**	0.04
1000	4000	100	**0.08**	0.11	0.07	**0.06**
1000	4000	1000	0.19	**0.16**	0.17	**0.09**
1000	4000	10000	0.28	**0.21**	0.23	**0.11**
1000	4000	100000	0.29	**0.24**	0.24	**0.12**
1000	4000	1000000	0.30	**0.26**	0.29	**0.15**
1000	4000	10000000	**0.29**	0.33	0.22	**0.16**
2000	8000	1	0.27	**0.03**	**0.02**	**0.02**
2000	8000	10	0.23	**0.15**	0.11	**0.09**
2000	8000	100	**0.25**	0.27	0.22	**0.15**
2000	8000	1000	0.55	**0.41**	0.54	**0.24**
2000	8000	10000	1.05	**0.50**	0.85	**0.25**
2000	8000	100000	1.17	**0.59**	0.90	**0.29**
2000	8000	1000000	1.19	**0.68**	1.02	**0.33**
2000	8000	10000000	1.20	**0.70**	1.06	**0.36**
4000	16000	1	1.12	**0.07**	0.09	**0.04**
4000	16000	10	0.90	**0.42**	0.44	**0.26**
4000	16000	100	0.91	**0.78**	0.68	**0.36**
4000	16000	1000	1.64	**1.14**	1.65	**0.50**
4000	16000	10000	3.75	**1.51**	3.17	**0.63**
4000	16000	100000	4.72	**1.75**	3.96	**0.75**
4000	16000	1000000	4.79	**1.97**	4.04	**0.83**
4000	16000	10000000	4.74	**2.33**	3.82	**0.87**
8000	32000	1	4.76	**0.24**	0.36	**0.12**
8000	32000	10	3.68	**1.23**	1.90	**0.73**
8000	32000	100	3.55	**2.35**	2.67	**1.07**
8000	32000	1000	5.19	**3.45**	6.93	**1.58**
8000	32000	10000	13.16	**4.49**	13.62	**1.70**
8000	32000	100000	18.89	**5.28**	14.74	**2.05**
8000	32000	1000000	19.66	**6.06**	16.81	**2.36**
8000	32000	10000000	19.70	**6.76**	16.62	**2.75**

Table 6: Results on random graphs with varying maximum weights for both perfect and non-perfect variants.

k	n	m	E	G	M	S
100	300	399	**0.00**	0.01	0.01	**0.00**
200	600	799	0.02	0.02	0.03	**0.00**
400	1200	1599	0.06	0.10	0.09	**0.01**
1000	3000	3999	0.35	0.60	0.50	**0.04**
2000	6000	7999	1.38	2.40	2.31	**0.12**
4000	12000	15999	6.01	10.38	14.08	**0.39**
k	n	m	E^*_G	G^*_G	M^*_G	S^*_G
100	300	399	**0.00**	**0.00**	**0.00**	**0.00**
200	600	799	**0.00**	**0.00**	0.01	**0.00**
400	1200	1599	**0.01**	0.02	0.04	**0.01**
1000	3000	3999	0.05	0.10	0.16	**0.02**
2000	6000	7999	0.19	0.40	0.74	**0.06**
4000	12000	15999	0.74	1.76	4.29	**0.20**

Table 7: Results on the `t.f` DIMACS instances with varying k.

k	n	m	E	G	M	S
50	300	20000	0.11	0.08	0.10	**0.02**
100	600	80000	0.90	0.60	0.75	**0.06**
200	1200	320000	8.60	5.56	7.32	**0.24**
400	2400	1280000	68.91	44.53	63.66	**0.97**
800	4800	5120000	529.36	351.33	595.68	**4.08**
k	n	m	E^*_G	G^*_G	M^*_G	S^*
50	300	20000	0.03	**0.01**	0.04	**0.01**
100	600	80000	0.21	0.07	0.25	**0.03**
200	1200	320000	2.55	0.69	2.56	**0.13**
400	2400	1280000	19.67	5.43	23.45	**0.55**
800	4800	5120000	139.87	41.25	244.50	**2.28**

Table 8: Results on the `hardcard.f` DIMACS instances with varying k.

Our implementations of the various iterations of the blossom algorithm are reasonably close to how they were described originally. The variants of the blossom algorithm mainly differ by how they implement distinct aspects of the search procedure: representing blossoms, maintaining dual weights, performing adjustments and finding tight edges. One method that was found to lead to good performance is the variable δ approach introduced in [4] that perform dual adjustment by choosing a different value of δ for each search tree using a heuristic. The authors of [28] showed an efficient variant of the $O(nm \log n)$ algorithm that uses a combination of concatenable queues and union-find similarly to how splitting lists and union-find are used by us in the scaling algorithm. The current best performing implementation presented in [26] is based on the variable δ approach and uses priority queues to maintain edge slacks. The authors don't give a proof of a theoretical time complexity but believe the worst-case to be $O(n^3 m)$.

We compared the performance of our scaling algorithm implementation to the implementation of [26] and found it to be significantly slower. While there are some ideas from existing blossom algorithm implementations that could improve the performance of our scaling algorithm, we doubt it would be enough to outperform the best implementations of the blossom algorithm. The main disadvantages of the scaling approach is that it is sensitive to sizes of weight and does a lot of repeated work by finding a new matching in each scale.

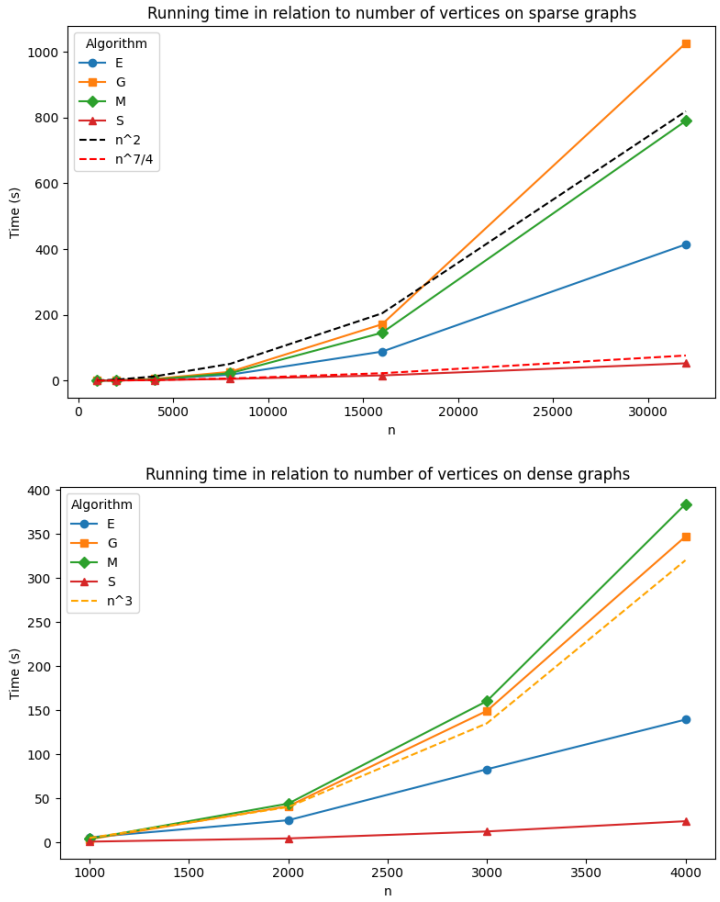

Figure 3: Running time of tested algorithms on sparse and dense random instances.

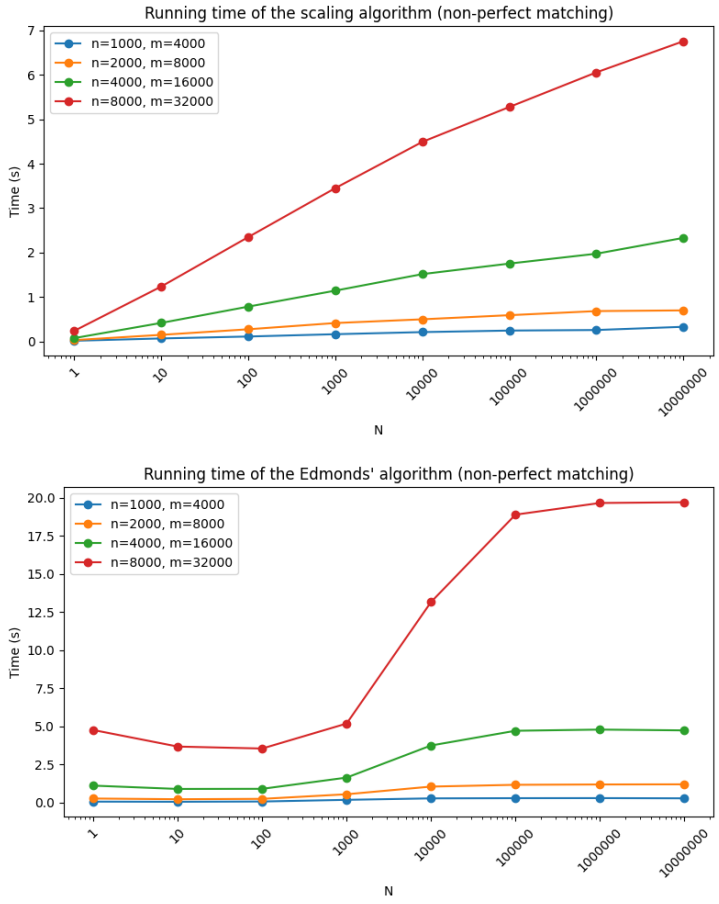

Figure 4: Running time of the scaling algorithm and the Edmonds' algorithm in relation to the maximum edge weight N.

Bibliography

[1] Alfred Aho and John Hopcroft. *The design and analysis of computer algorithms*. Pearson Education India, 1974.

[2] B. Bollobás. *Modern Graph Theory*. Graduate Texts in Mathematics. Springer New York, 1998. ISBN: 9780387984889. URL: https://books.google.pl/books?id=SbZKSZ-1qrwC.

[3] Nicos Christofides. "Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem". In: *Operations Research Forum* 3 (1976).

[4] William Cook and Andre Rohe. "Computing minimum-weight perfect matchings". In: *INFORMS journal on computing* 11.2 (1999), pp. 138–148.

[5] Marek Cygan, Harold Gabow, and Piotr Sankowski. "Algorithmic applications of Baur-Strassen's theorem: Shortest cycles, diameter, and matchings". In: *Journal of the ACM (JACM)* 62.4 (2015), pp. 1–30.

[6] *DIMACS :: Implementation Challenges — dimacs.rutgers.edu*. http://dimacs.rutgers.edu/programs/challenge/. [Accessed 29-08-2024].

[7] Ran Duan, Seth Pettie, and Hsin-Hao Su. "Scaling algorithms for weighted matching in general graphs". In: *ACM Transactions on Algorithms (TALG)* 14.1 (2018), pp. 1–35.

[8] Jack Edmonds. "Maximum matching and a polyhedron with 0, 1-vertices". In: *Journal of research of the National Bureau of Standards B* 69.125-130 (1965), pp. 55–56.

[9] Jack Edmonds. "Paths, trees, and flowers". In: *Canadian Journal of mathematics* 17 (1965), pp. 449–467.

[10] Jack Edmonds and Ellis Johnson. "Matching, Euler tours and the Chinese postman". In: *Mathematical Programming* 5 (1973), pp. 88–124.

[11] Jack Edmonds and Richard M Karp. "Theoretical improvements in algorithmic efficiency for network flow problems". In: *Journal of the ACM (JACM)* 19.2 (1972), pp. 248–264.

[12] Peter Fenwick. "A new data structure for cumulative frequency tables". In: *Software: Practice and Experience* 24 (1994), pp. 327–336.

[13] Harold Gabow. "A scaling algorithm for weighted matching on general graphs". In: *26th Annual Symposium on Foundations of Computer Science (FOCS 1985)*. 1985, pp. 90–100.

[14] Harold Gabow. "An efficient implementation of Edmonds' algorithm for maximum matching on graphs". In: *Journal of the ACM (JACM)* 23.2 (1976), pp. 221–234.

[15] Harold Gabow. "Data structures for weighted matching and nearest common ancestors with linking". In: *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*. 1990, pp. 434–443.

[16] Harold Gabow, Jon Louis Bentley, and Robert Tarjan. "Scaling and related techniques for geometry problems". In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. 1984, pp. 135–143.

[17] Harold Gabow, Zvi Galil, and Thomas Spencer. "Efficient implementation of graph algorithms using contraction". In: *25th Annual Symposium on Foundations of Computer Science, (FOCS 1984)*. IEEE. 1984, pp. 347–357.

[18] Harold Gabow and Robert Tarjan. "Faster scaling algorithms for general graph matching problems". In: *Journal of the ACM (JACM)* 38.4 (1991), pp. 815–853.

[19] Harold Gabow and Robert Tarjan. "Faster scaling algorithms for network problems". In: *SIAM Journal on Computing* 18.5 (1989), pp. 1013–1036.

[20] Haroldeil Gabow. *Implementation of algorithms for maximum matching on nonbipartite graphs.* Stanford University, 1974.

[21] Zvi Galil, Silvio Micali, and Harold Gabow. "An $O(EV \log V)$ algorithm for finding a maximal weighted matching in general graphs". In: *SIAM Journal on Computing* 15.1 (1986), pp. 120–130.

[22] *GitHub - krzysztof-turowski/koala-networkit: KOALA library fork built on NetworKit — github.com*. `https://github.com/krzysztof-turowski/koala-networkit`. [Accessed 10-09-2024].

[23] Andrew Goldberg and Alexander Karzanov. "Maximum skew-symmetric flows and matchings". In: *Mathematical Programming* 100 (2004), pp. 537–568.

[24] Frank Hadlock. "Finding a maximum cut of a planar graph in polynomial time". In: *SIAM Journal on Computing* 4.3 (1975), pp. 221–225.

[25] Frank Hitchcock. "The distribution of a product from several sources to numerous localities". In: *Journal of mathematics and physics* 20.1-4 (1941), pp. 224–230.

[26] Vladimir Kolmogorov. "Blossom V: a new implementation of a minimum cost perfect matching algorithm". In: *Mathematical Programming Computation* 1 (2009), pp. 43–67.

[27] Eugene Lawler. *Combinatorial optimization: networks and matroids*. Courier Corporation, 2001.

[28] Kurt Mehlhorn and Guido Schäfer. "Implementation of O (nm log n) weighted matchings in general graphs: the power of data structures". In: *Journal of Experimental Algorithmics (JEA)* 7 (2002), p. 4.

[29] Silvio Micali and Vijay Vazirani. "An $O(\sqrt{V}E)$ algorithm for finding maximum matching in general graphs". In: *21st Annual Symposium on Foundations of Computer Science (FOCS 1980)*. IEEE. 1980, pp. 17–27.

[30] Marcin Mucha and Piotr Sankowski. "Maximum matchings via Gaussian elimination". In: *45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004)*. 2004, pp. 248–255.

[31] Seth Pettie. "Sensitivity analysis of minimum spanning trees in sub-inverse-Ackermann time". In: *International Symposium on Algorithms and Computation*. Springer. 2005, pp. 964–973.

[32] Michael Rabin and Vijay Vazirani. "Maximum matchings in general graphs through randomization". In: *Journal of Algorithms* 10.4 (1989), pp. 557–567.

[33] Robert Thorndike. "The problem of classification of personnel". In: *Psychometrika* 15.3 (1950), pp. 215–235.

[34] Mikkel Thorup. "Equivalence between priority queues and sorting". In: *Journal of the ACM (JACM)* 54.6 (2007), pp. 28–55.