Chapter 1 - Vertex Coloring

Problem 1.1 - Vertex Coloring
Given an undirected graph $G = (V, E)$, assign a color cu to each vertex $u \in V$ such that the following holds: $e = (v, w) \in E \Rightarrow cv \neq cw$.

Definition 1.2 - Node Identifiers
Each node has a unique identifier, e.g., its IP address. We usually assume that each identifier consists of only log n bits if the system has n nodes.

Definition 1.3 - Chromatic Number
Given an undirected Graph $G = (V, E)$, the chromatic number $\chi(G)$ is the minimum number of colors to solve Problem 1.1.

Algorithm 1 - Greedy Sequential *(see page 6)*

- **Vertex Coloring**
- Non-distributed
- Centralized
- **Theorem 1.5:** is correct and terminates in n "steps". The algorithm uses at most $\Delta + 1$ colors.

Definition 1.4 - Degree
The number of neighbors of a vertex v, denoted by $\delta(v)$, is called the degree of v. The maximum degree vertex in a graph G defines the graph degree $\Delta(G) = \Delta$.

Definition 1.6 - Synchronous Distributed Algorithm
n a synchronous al- gorithm, nodes operate in synchronous rounds. In each round, each processor executes the following steps:

1. Do some local computation (of reasonable complexity).
2. Send messages to neighbors in graph (of reasonable size).
3. Receive messages (that were sent by neighbors in step 2 of the same round).

Algorithm 3 - Reduce *(see page 7)*

- **Theorem 1.8:** is correct and has time complexity n. The algorithm uses at most $\Delta + 1$ colors.

Definition 1.7 - Time Complexity
For synchronous algorithms (as defined in 1.6) the time complexity is the number of rounds until the algorithm terminates.

Lemma 1.9
$\chi(Tree) \leq 2$

Algorithm 4 - Slow Tree Coloring *(see page 8)*

- Time Complexity: Height of tree (up to n)
- Does not need to be synchronous

Definition 1.10 - Asynchronous Distributed Algorithm
In the asynchronous model, algorithms are event driven ("upon receiving message ..., do..."). Processors cannot access a global clock. A message sent from one processor to another will arrive in finite but unbounded time.

Definition 1.11 - Time Complexity
For asynchronous algorithms (as defined in 1.6) the time complexity is the number of time units from the start of the execution to its completion in the worst case (every legal input, every execution scenario), assuming that each message has a delay of at most one time unit.

Definition 1.12 - Message Complexity
The message complexity of a syn- chronous or asynchronous algorithm is determined by the number of messages exchanged (again every legal input, every execution scenario).

Theorem 1.13 - Slow Tree Coloring
Algorithm 4 (Slow Tree Coloring) is correct. If each node knows its parent and its children, the (asynchronous) time complexity is the tree height which is bounded by the diameter of the tree; the message complexity is $n - 1$ in a tree with n nodes.

Definition 1.14 - Log-Star
$\forall x \leq 2 : log * x := 1 \forall x > 2 : log * x := 1 + log * (log(x))$

Algorithm 5 - "6-Color" *(see page*

- *Time Complexity:* $O(log^*(n))$

) 10

Theorem 1.15 - 6-Color
Algorithm 5 ("6-Color") terminates in $log^*(n)$ time.

Algorithm 6 - Shift Down *(see page 11)*

- **Lemma 1.16:** Preserves coloring legality: also siblings are monochromatic

Algorithm 7 - Six-2-Three *(see page 11)*

- **Theorem 1.17:** colors a tree with three colors in $O(log^*(n))$.

Chapter 2 - Leader Election

Setup:

- Ring topology

Problem 2.1 - Leader Election
Each node eventually decides whether it is a leader or not, subject to the constraint that there is exactly one leader.

Definition 2.2 - Anonymous
A system is anonymous if nodes do not have unique identifiers.

Definition 2.3 - Uniform
An algorithm is called uniform if the number of nodes n is not known to the algorithm (to the nodes, if you wish). If n is known, the algorithm is called non-uniform.

Lemma 2.4
After round k of any deterministic algorithm on an anonymous ring, each node is in the same state s_k.

Theorem 2.5 - Anonymous Leader Election
Deterministic leader election in an anonymous ring is impossible.
(Also holds for other symmetric network topologies (e.g., com- plete graphs, complete bipartite graphs), but does not hold for randomized algorithms.)

Algorithm 8 - Clockwise *(see page 17)*

- **Theorem 2.6:** is correct. The **time complexity** is $O(n)$. The **message complexity** is $O(n2)$.

Algorithm 9 - Radius Growth *(see page 18)*

- **Theorem 2.7:** is correct. The **time complexity** is $O(n)$. The **message complexity** is $O(nlogn)$.
- Asynchronous
- Uniform

Definition 2.8 - Execution
Definition 2.8 (Execution). An execution of a distributed algorithm is a list of events, sorted by time. An event is a record (time, node, type, message), where type is "send" or "receive".

Definition 2.9 - Open Schedule
A schedule is an execution chosen by the scheduler. An open (undirected) edge is an edge where no message traversing the edge has been received so far. A schedule for a ring is open if there is an open edge in the ring.

Lemma 2.10
Given a ring R with two nodes, we can construct an open schedule in which at least one message is received. The nodes cannot distinguish this schedule from one on a larger ring with all other nodes being where the open edge is.

Lemma 2.11
By gluing together two rings of size $n/2$ for which we have open schedules, we can construct an open schedule on a ring of size n. If $M(n/2)$ denotes the number of messages already received in each of these schedules, at least $2M(n/2) + n/4$ messages have to be exchanged in order to solve leader election.

Lemma 2.12
Any uniform leader election algorithm for asynchronous rings has at least message complexity $M(n) \geq n(log(n) + 1)$.

Theorem 2.13 - Asynchronous Leader Election Lower Bound
Any uniform leader election algorithm for asynchronous rings has $\Omega(n * log(n))$ message complexity.

Algorithm 10 - Synchronous Leader Election *(see page 21)*

- Non-uniform (i.e. the ring size n is known)
- Every node starts at same time
- Mimimum identifies (integers) becomes the leader
- **Time Complexity:** $m * n$, where m is the minimum identifier
- **Message Complexity:** n.

Chapter 3 - Tree Algorithms

Definition 3.1 - Broadcast
A broadcast operation is initiated by a single processor, the source. The source wants to send a message to all other nodes in the system.

Definition 3.2 - Distance, Radius, Diameter
The **distance** between two nodes u and v in an undirected graph G is the number of hops of a minimum path between u and v. The **radius of a node** u is the maximum distance between u and any other node in the graph. The **radius of a graph** is the minimum radius of any node in the graph. The **diameter** of a graph is the maximum distance between two arbitrary nodes.
Relation between radius and diameter: $R \leq D \leq 2R$

Theorem 3.3 - Broadcast Lower Bound
The message complexity of broadcast is at least $n - 1$. The source's radius is a lower bound for the time complexity.

Definition 3.4 - Clean
A graph (network) is clean if the nodes do not know the topology of the graph.

Theorem 3.5 - Clean Broadcast Lower Bound
For a clean network, the num- ber of edges is a lower bound for the broadcast message complexity.

Algorithm 11 - Flooding *(see page 24)*

-

Algorithm 12 - Echo *(see page 25)*

- **Message Complexity:** $n - 1$
- Together with flooding the message complexity is $O(m)$, where $m = |E|$ is the number of edges in the graph.
- **Time Complexity:** Depth of the spanning tree.

Algorithm 13 - Dijkstra BFS *(see page 26)*

- **Spanning Tree**
- Always add the closest node.
- **Theorem 3.6:** The **time complexity** is $O(D2)$, the **message complexity** is $O(m + nD)$, where D is the diameter of the graph, n the number of nodes, and m the number of edges.

Algorithm 14 - Bellman-Ford BFS *(see page 26)*

- **Spanning Tree**
- Simply keep the distance to the root accurate.
- Border Gateway Protocol
- **Theorem 3.7:** The **time complexity** is $O(D)$, the **message complexity** is $O(nm)$, where D is the diameter of the graph, n the number of nodes, and m the number of edges.

Definition 3.8 - MST (Minimum spanning tree)
Given a weighted graph $G = (V, E, \omega)$, the MST of G is a spanning tree T minimizing $\omega(T)$, where $\omega(G') = \sum_{e \in G'} \omega_e$ for any subgraph $G' \subseteq G$.

Definition 3.9 - Blue Edge
Let T be a spanning tree of the weighted graph G and $T' \subseteq T$ a subgraph of T (also called a fragment). Edge $e = (u, v)$ is an outgoing edge of T' if $u \in T'$ and $v \notin T'$ (or vice versa). The minimum weight outgoing edge $b(T')$ is the so-called blue edge of T'.

Lemma 3.10
For a given weighted graph G (such that no two weights are the same), let T denote the MST, and T' be a fragment of T. Then the blue edge of T' is also part of T, i.e., $T' \cup b(T') \subseteq T$.

Algorithm 15 - GHS (Gallager-Humblet-Spira) *(see page 28)*

- **Theorem 3.11:** The **time complexity** is $O(nlogn)$, the **message complexity** is $O(mlogn)$.
- Directly solves leader election in general graphs: The leader is simply the last surviving root.

Chapter 4 - Distributed Sorting

Definition 4.1 - Sorting
We choose a graph with n nodes v1, ..., vn. Initially each node stores a value. After applying a sorting algorithm, node vk stores the kth smallest value.
Star Topology: Sorting takes $O(1)$.

Definition 4.2 - Node Contention
n each step of a synchronous algorithm, each node can only send and receive $O(1)$ messages containing $O(1)$ values, no matter how many neighbors the node has.
Star Topology: Sorting takes $O(n)$.

Algorithm 16 - Odd/Even Sort *(see page 32)*

- **Theorem 4.4:** sorts correctly in n steps.

Lemma 4.3
0-1 Sorting Lemma: If an oblivious comparison-exchange algorithm sorts all inputs of $0's$ and $1's$, then it sorts arbitrary inputs.

Algorithm 17 - Shearsort *(see page 33)*

- **Theorem 4.5:** sorts n values in $\sqrt(n)(logn + 1)$ time in snake-like order.

Definition 4.6 - Sorting Network
A comparator is a device with two inputs x, y and two outputs x', y' such that $x' = min(x, y)$ and $y' = max(x, y)$. We construct so-called comparison networks that consist of wires that connect comparators (the output port of a comparator is sent to an input port of another comparator). Some wires are not connected to comparator outputs, and some are not connected to comparator inputs. The first are called input wires of the comparison network, the second output wires. Given n values on the input wires, a sorting network ensures that the values are sorted on the output wires. We will also use the term width to indicate the number of wires in the sorting network.

Definition 4.7 - Depth
he depth of an input wire is 0. The depth of a comparator is the maximum depth of its input wires plus one. The depth of an output wire of a comparator is the depth of the comparator. The depth of a comparison network is the maximum depth (of an output wire).

Definition 4.8 - Bitonic Sequence
A bitonic sequence is a sequence of numbers that first monotonically increases, and then monotonically decreases, or vice versa.

Algorithm 18 - Half Cleaner *(see page 35)*

- **Lemma 4.9:** Feeding a bitonic sequence into a half cleaner, the half cleaner cleans (makes all-0 or all-1) either the upper or the lower half of the n wires. The other half is bitonic.

Algorithm 19 - Bitonic Sequnence Sorter *(see page 35)*

- **Lemma 4.10:** A bitonic sequence sorter of width n sorts bitonic sequences. It has depth $logn$.

Algorithm 20 - Merging Network *(see page 36)*

- **Lemma 4.11:** A merging network of width n merges two sorted input sequences of length $n/2$ each into one sorted sequence of length n.

Algorithm 21 - Batcher's "Bitonic" Sorting Network *(see page 36)*

- **Theorem 4.12:** A sorting network sorts an arbitrary sequence of n values. It has depth $O(log^2 n)$.
- Can be simulated by a Butterfly network and other hypercubic networks.

Definition 4.13 - Distributed Counting
A distributed counter is a variable that is common to all processors in a system and that supports an atomic test-and-increment operation. The operation delivers the system's counter value to the requesting processor and increments it.

Definition 4.14 - Balancer
A balancer is an asynchronous flip-flop which forwards messages that arrive on the left side to the wires on the right, the first to the upper, the second to the lower, the third to the upper, and so on.

Algorithm 22 - Bitonic Counting Network *(see page 38)*

- **Theorem 4.20:** In a quiescent state, the w output wires of a bitonic counting network of width w have the step property.
- **Lemma 4.23:** The bitonic counting network is not linearizable.

Definition 4.15 - Step Property
A sequence y_0,y_1,...,y_{w-1} is said to have the step property,if $0 \leq y_i - y_j \leq 1$, for any $i < j$.

Fact 4.16
For a balancer, we denote the number of consumed messages on the i^{th} input wire with x_i, $i = 0, 1$. Similarly, we denote the number of sent messages on the ith output wire with y_i , $i = 0, 1$. A balancer has these properties:

1. A balancer does not generate output-messages; that is, $x_0 + x_1 \geq y_0 + y_1$ in any state.
2. Every incoming message is eventually forwarded. In other words, if we are in a quiescent state (no message in transit), then $x_0 + x_1 = y_0 + y_1$.
3. The number of messages sent to the upper output wire is at most one higher than the number of messages sent to the lower output wire: in any state $y_0 = \lceil (y_0 + y_1)/2 \rceil$ (thus $y_1 = \lfloor (y_0 + y_1)/2 \rfloor$).

Fact 4.17
If a sequence $y_0, y_1, \ldots, y_{w-1}$ has the step property,

1. then all its subsequences have the step property.
2. then its even and odd subsequences satisfy
   $\sum_{i=0}^{w/2-1} y_{2i} = \lceil \frac{1}{2} \sum_{i=0}^{w-1} y_i \rceil$ and $\sum_{i=0}^{w/2-1} y_{2i+1} = \lfloor \frac{1}{2} \sum_{i=0}^{w-1} y_i \rfloor$

Fact 4.18
If two sequences $x_0, x_1,\ldots,x_{w-1}$ and $y_0, y_1, \ldots, y_{w-1}$ have the step property,

1. and $\sum_{i=0}^{w-1} x_i = \sum_{i=0}^{w-1} y_i$, then $x_i = y_i$ for $i = 0,\ldots,w-1$.

2. and $\sum_{i=0}^{w-1} x_i = \sum_{i=0}^{w-1} y_i + 1$ then there exists a unique j (j = 0,1,...,$w-1$) such that $x_j = y_j + 1$, and $x_i = y_i$ for i = 0,...,$w-1$, $i \neq j$.

Lemma 4.19
Let M[w] be a Merger of width w. In a quiescent state (no mes- sage in transit), if the inputs x_0, x_1, ..., $x_{w/2-1}$ resp. $x_{w/2}$, $x_{w/2+1}$, ..., x_{w-1} have the step property, then the output y_0, y_1, ..., y_{w-1} has the step property.

Theorem 4.21 - Counting vs. Sorting
If a network is a counting network then it is also a sorting network, but not vice versa.

Definition 4.22 - Linearizable
A system is linearizable if the order of the values assigned reflects the real-time order in which they were requested. More formally, if there is a pair of operations o_1, o_2, where operation o_1 terminates be- fore operation o_2 starts, and the logical order is "o_2 before o_1", then a distributed system is not linearizable.

Chapter 5 - Maximal Independent Set

Definition 5.1 - Independent Set
Given an undirected Graph $G = (V, E)$ an independent set is a subset of nodes $U \subseteq V$, such that no two nodes in U are adjacent. An **independent set is maximal** if no node can be added without violating independence. An independent set of maximum cardinality is called maximum.

Algorithm 23 - Slow MIS *(see page 46)*

- **Theorem 5.2:** features a **time complexity** of $O(n)$ and a **message complexity** of $O(m)$.

Corollary 5.3
Given a coloring algorithm that needs C colors and runs in time T, we can construct a MIS in time C + T. *(We first choose all nodes of the first color. Then, for each additional color we add "in parallel" (without conflict) as many nodes as possible.)*

Algorithm 24 - Fast MIS *(see page 47)*

- **Lemma 5.4 - Joining MIS:** A node v joins the MIS in Step 2 of the Fast MIS with probability $p \geq 1$.
- **Lemma 5.5 - Good Nodes:** A node v is called good if
  $\sum_{w \in N(v)} \frac{1}{2d(w) \geq \frac{1}{6}}$
  Otherwise we call v a bad node. A good node will be removed in Step 3 with probability $p \geq \frac{1}{36}$.
- **Lemma 5.6 - Good Edges:** An edge $e = (u, v)$ is called bad if both u and v are bad; else the edge is called good. The following holds: At any time at least half of the edges are good.
- **Theorem 5.8:** terminates in expected **time** $O(logn)$ (, even with high probabilty).

Lemma 5.7
A bad node has outdegree (number of edges pointing away from bad node) at least twice its indegree (number of edges pointing towards bad node).

Algorithm 25 - Fast MIS 2 *(see page 50)*

- **Lemma 5.10:** In a single phase, we remove at least half of the edges in expectation.
- **Theorem 5.11:** terminates after time at most $3log_{4/3}m + 1 \in O(logn)$ phases in expectation.
  Corollary 5.14: terminates w.h.p. in O(log n) time.

Theorem 5.9 - Linearity of Expectation
Let X_i, $i = 1, \ldots, k$ denote random variables, then
$E[\sum_i X_i] = \sum_i E[X_i]$.

Definition 5.12 - With High Probability (W.h.p.)
We say that an algorithm terminates w.h.p. (with high probability) within $O(t)$ time if it does so with probability at least $1 - 1/n^c$ for any choice of $c \geq 1$. Here c may affect the constants in the Big-O notation because it is considered a "tunable constant" and usually kept small.

Chapter 6 - Locality Lower Bounds

Chapter 7 - All-to-All Communication

Chapter 8 - Social Networks

Chapter 9 - Shared Memory

Chapter 10 - Shared Objects

Chapter 11 - Wireless Protocols

Chapter 12 - Synchronization

Chapter 13 - Peer-to-Peer Computing