Algorithms and Datastructures Graphs, Depth-/Breadth-first Search, Graph-Connectivity



Albert-Ludwigs-Universität Freiburg

Prof. Dr. Rolf Backofen

Bioinformatics Group / Department of Computer Science Algorithms and Datastructures, January 2017

Structure



Feedback

Exercises Lecture

Graphs

Introduction Implementation Application example

Feedback from the exercises



The upcoming exercise sheet 12 and 13 will be merged together (finding largest connected component + Dijkstra)

Some people were asking for more solution sheets for the exercises

We are working on it.

Feedback from the lecture



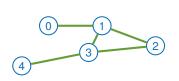
Code in the lecture will be a little bit different from exercise sheet.

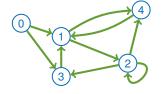
One person asked for additional explanations regarding proofs.

Graphs - Overview:

- Besides arrays, lists and trees the most common datastructure (Trees are a special type of graph)
- Representation of graphs in the computer
- Breadth first search (BFS)
- Depth first search (DFS)
- Connected components of a graph

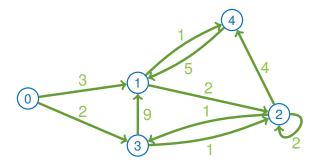
Terminology:





- Each Graph G = (V, E) consists of:
 - A set of vertices (nodes) $V = \{v_1, v_2, ...\}$
 - A set of edges (arcs) $E = \{e_1, e_2, ...\}$
- Each edge connects two vertices $(u, v \in V)$
 - Undirected edge: $e = \{u, v\}$ (set)
 - Directed edge: e = (u, v) (tuple)
- Self-loops are also possible: e = (u, u) or $e = \{u, u\}$

Weighted graph:



- Each edge is marked with a real number named weight
- The weight is also named length or cost of the edge depending on the application



- Intersections: vertices
- Roads: edges
- Travel time: costs of the edges



Figure: Map of Freiburg © OpenStreetMap

How to represent this graph computationally?

- Two classic variants
 - Adjacency matrix with space consumption $\Theta(|V|^2)$

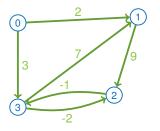
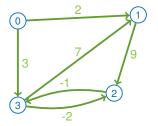


Figure: Weighted graph with |V| = 4, |E| = 6

	end-vertice			
	0	1	2	3
ice 0		2		3
start-vertice			9	
£ 2				-1
sta 3		7	-2	

Figure: Adjacency matrix

- Two classic variants
 - 2 Adjacency list / fields with space consumption $\Theta(|V| + |E|)$
 - Each list item stores the target vertice and the cost of the edge



<u>8</u> (0)	1, 2	3, 3
start-vertice	2, 9	
£ 2	3, -1	
sta 3	1, 7	2, -2
		•

Figure: Weighted graph with

$$|V| = 4$$
, $|E| = 6$

Figure: Adjacency list

Graph: Arrangement

- Graph is fully defined through the adjacency matrix / list
- The arrangement is not relevant for visualisation of the graph

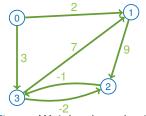


Figure: Weighted graph with |V| = 4, |E| = 6

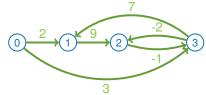


Figure: Same graph ordered by number - outer planar graph

```
class Graph:
    def init (self):
        self.vertices = []
        self.edges = []
    def addVertice(self, vert):
        self.vertices.append(vert)
    def addEdge(self, fromVert, toVert):
        self.edges.append((fromVert, toVert))
```



Degree of a vertex: Directed graph: G = (V, E)



Figure: Vertex with in- / outdegree of 3 / 2

Indegree of a vertex u is the number of edge heads adjacent to the vertex

$$\deg^+(u) = |\{(v, u) : (v, u) \in E\}|$$

Outdegree of a vertex u is the number of edge tails adjacent to the vertex

$$\deg^{-}(u) = |\{(u, v) : (u, v) \in E\}|$$

Degree of a vertex: Undirected graph: G = (V, E)



Figure: Vertex with degree of 4

Degree of a vertex u is the number of vertices adjacent to the vertex

$$deg(u) = |\{\{v, u\} : \{v, u\} \in E\}|$$

Paths in a graph: G = (V, E)

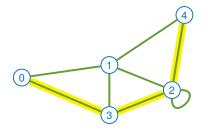


Figure: Undirected path of length 3 P = (0,3,2,4)

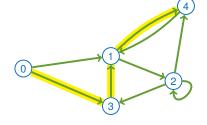
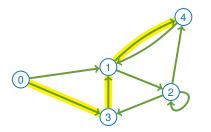


Figure: Directed path of length 3 P = (0,3,1,4)

- A path of *G* is a sequence of edges $u_1, u_2, ..., u_i \in V$ with
 - Undirected graph: $\{u_1, u_2\}, \{u_2, u_3\}, \dots, \{u_{i-1}, u_i\} \in E$
 - Directed graph: $(u_1, u_2), (u_2, u_3), \dots, (u_{i-1}, u_i) \in E$

Paths in a graph: G = (V, E)



3 -1 2

Figure: Directed path of length 3 P = (0,3,1,4)

Figure: Weighted path with cost 6 P = (2,3,1)

- The length of a path is: (also costs of a path)
 - Without weights: number of edges taken
 - With weights: sum of weigths of edges taken

Shortest path in a graph: G = (V, E)

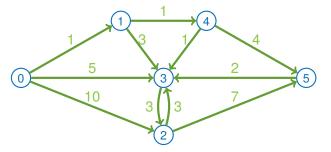


Figure: Shortest path from 0 to 2 with cost / distance d(0,2) = ?

The shortest path between two vertices u, v is the path P = (u, ..., v) with the shortest length d(u, v) or lowest costs

Shortest path in a graph: G = (V, E)

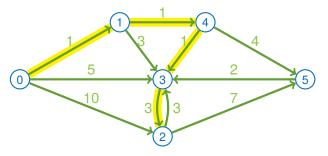


Figure: Shortest path from 0 to 2 with cost / distance d(0,2) = 6P = (0,1,4,3,2)

The shortest path between two vertices u, v is the path P = (u, ..., v) with the shortest length d(u, v) or lowest costs

Diameter of a graph: G = (V, E)

$$d = \max_{u,v \in V} d(u,v)$$

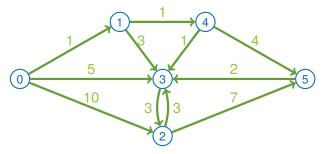


Figure: Diameter of graph is d = ?

The diameter of a graph is the length / the costs of the longest shortest path

Diameter of a graph: G = (V, E)

$$d = \max_{u,v \in V} d(u,v)$$

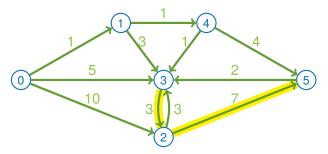


Figure: Diameter of graph is d = 10, P = (3, 2, 5)

The diameter of a graph is the length / the costs of the longest shortest path

Connected components: G = (V, E)

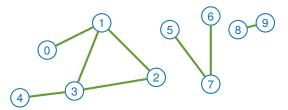


Figure: Three connected components

- Undirected graph:
 - All connected components are a partition of V

$$V = V_1 \cup \cdots \cup V_k$$

Two vertices u, v are in the same connected component if a path between u and v exists

Connected components: G = (V, E)

- Directed graph:
 - Named strongly connected components
 - Direction of edge has to be regarded
 - Not part of this lecture

- Let G = (V, E) be a graph and $s \in V$ a start vertex
- We visit each reachable vertex connected to s
- Breadth-first search: in sequence of the smallest distance to s
- Depth-first search: in sequence of the largest distance to s
- Not a problem on its own but is often used as subroutine of other algorithms

ldea:

- We start with all vertices unmarked and mark visited vertices
- Mark the start vertex s (level 0)
- Mark all unmarked connected vertices (level 1)
- Mark all unmarked vertices connected to a level 1-vertex (level 2)
- 5 Iteratively mark reachable vertices for all levels
- All connected nodes are now marked and in the same connected component as the start vertex s

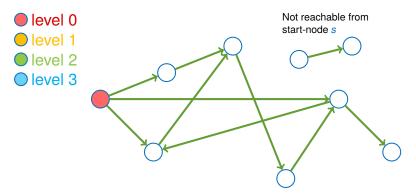


Figure: spanning tree of a breadth-first search

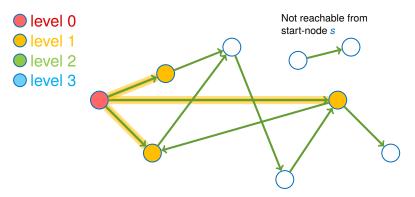


Figure: spanning tree of a breadth-first search

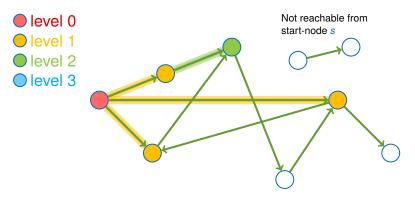


Figure: spanning tree of a breadth-first search

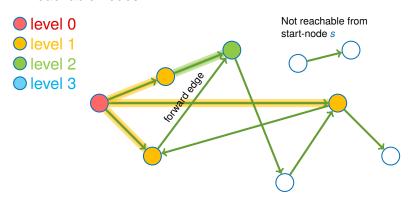


Figure: spanning tree of a breadth-first search

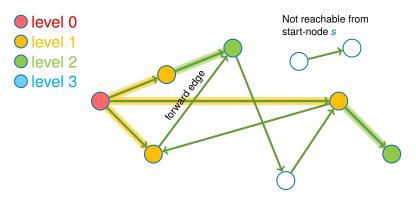


Figure: spanning tree of a breadth-first search

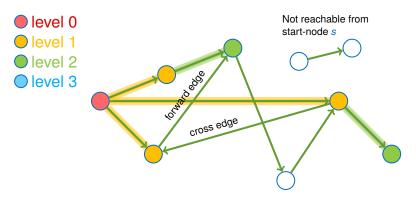


Figure: spanning tree of a breadth-first search

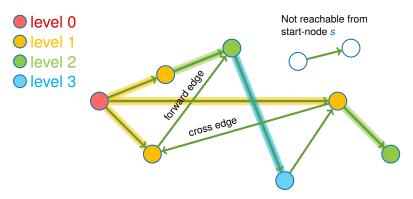


Figure: spanning tree of a breadth-first search

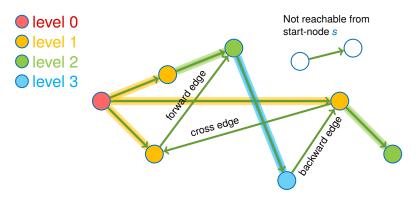


Figure: spanning tree of a breadth-first search

Idea:

- We start with all vertices unmarked and mark visited vertices
- Mark the start vertex s
- Pick an unmarked connected vertex and start a recursive depth-first search with the vertex as start vertex (continue on step 2)
- If no unmarked connected vertex exists go one vertex back (reduce the recursion level by one)

Depth-first search:

- Search starts with long paths (searching with depth)
- Marks like breadth-first search all connected vertices
- If the graph is acyclic we get a topological sorting
 - Each newly visited vertex gets marked by an increasing number
 - The numbers increase with path from the start vertex

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- opath 1
- path 2
- opath 3

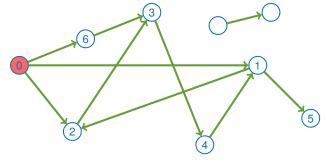


Figure: spanning tree of a depth-first search

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- path 1
- path 2
- opath 3

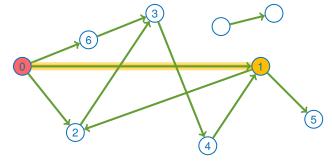


Figure: spanning tree of a depth-first search

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- path 1
- path 2
- opath 3

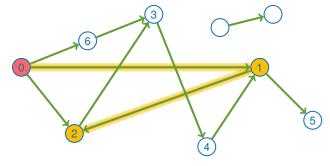


Figure: spanning tree of a depth-first search

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- opath 1
- path 2
- opath 3

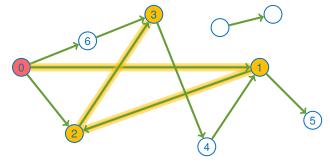


Figure: spanning tree of a depth-first search

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- opath 1
- path 2
- opath 3

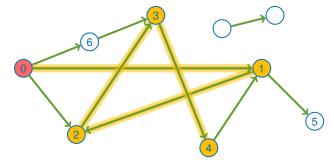


Figure: spanning tree of a depth-first search

The marked vertices create a different spanning tree containing all reachable nodes

- start-node
- opath 1
- path 2
- opath 3

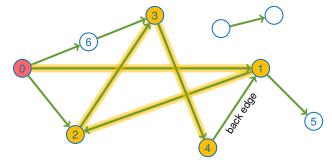


Figure: spanning tree of a depth-first search

The marked vertices create a different spanning tree containing all reachable nodes

- start-node
- path 1
- path 2
- opath 3

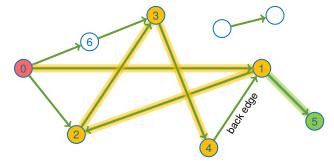


Figure: spanning tree of a depth-first search

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- opath 1
- path 2
- opath 3

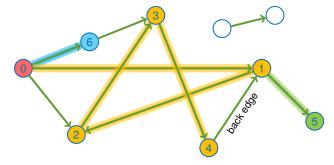


Figure: spanning tree of a depth-first search

- The marked vertices create a different spanning tree containing all reachable nodes
- start-node
- opath 1
- path 2
- opath 3

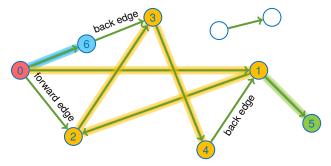


Figure: spanning tree of a depth-first search

The marked vertices create a different spanning tree containing all reachable nodes

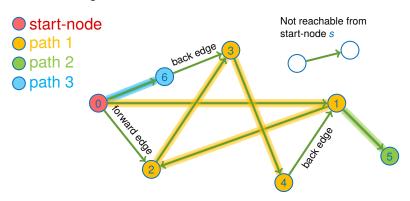


Figure: spanning tree of a depth-first search

Graphs

Why is this called Breadth - and Depth First Search?



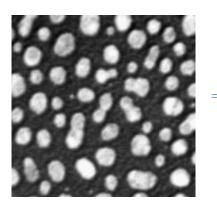
Runtime complexity:

- Constant costs for each visited vertex and edge
- We get a runtime complexity of $\Theta(|V'| + |E'|)$
- Let V' and E' be the reachable vertices and edges
- All vertices of V' are in the same connected component as our start vertex s
- This can only be improved by a constant factor

Image processing



- Connected component labeling
- Counting of objects in an image



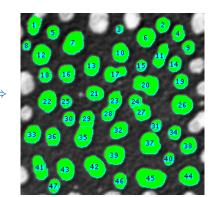
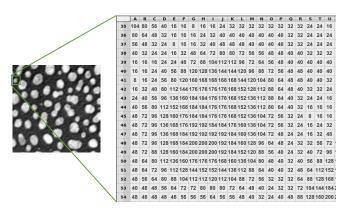


Image processing



What's object, what's background?





Convert to black white using threshold:

value = 255 if value > 100 else 0

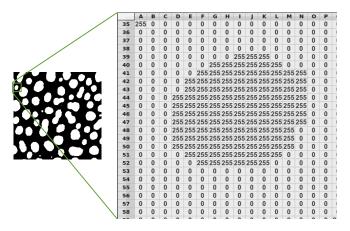


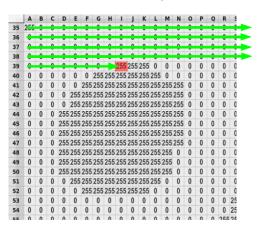
Image processing

Interpret image as graph:

- Each white pixel is a node
- Edges between adjacent pixels (normally 4 or 8 neighbors)
- Edges are not saved externally, algorithm works directly on array
- Breadth- / depth-first search find all connected components (particles)

Image processing

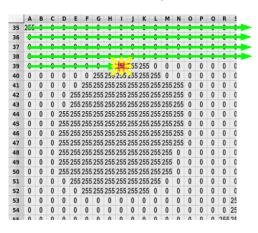




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1

Image processing

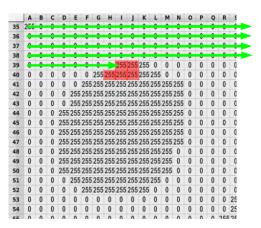




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels

William Willia

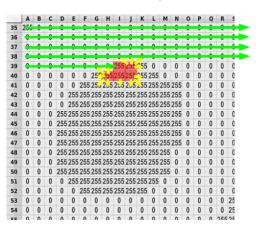
Image processing



- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing

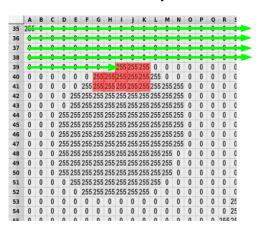




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing

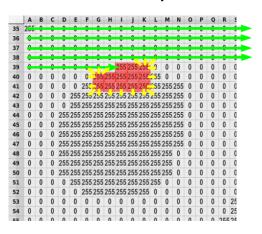




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing

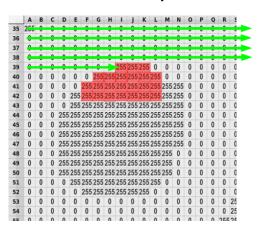




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing

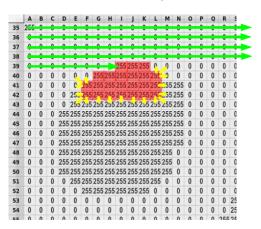




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing

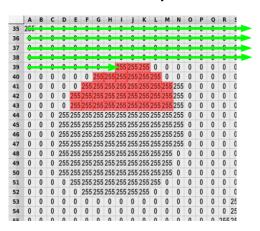




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing

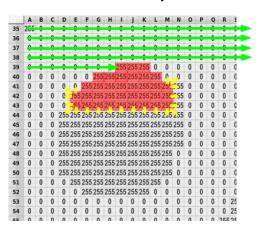




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

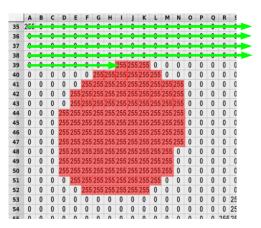
Image processing





- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

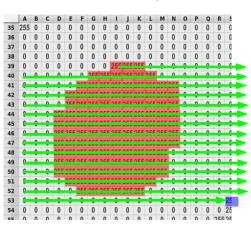




- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing





- Search pixel-by-pixel for non-zero intensity
- Label found pixel as component 2
- ...

Result of connected component labeling:

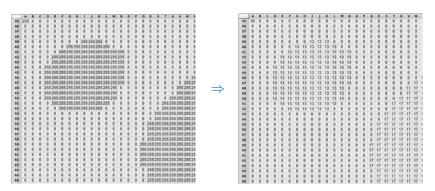


Figure: Result: particle indices instead of intensities

General

- [CRL01] Thomas H. Cormen, Ronald L. Rivest, and Charles E. Leiserson. Introduction to Algorithms. MIT Press, Cambridge, Mass, 2001.
- Kurt Mehlhorn and Peter Sanders. [MS08] Algorithms and data structures, 2008. https://people.mpi-inf.mpg.de/~mehlhorn/

ftp/Mehlhorn-Sanders-Toolbox.pdf.

■ Graph-Search

Graph-Connectivity

```
[Wik] Connectivity (graph theory)
    https://en.wikipedia.org/wiki/Connectivity_
    (graph_theory)
```