# Algorithms and Data Structures Graphs, Depth-/Breadth-first Search, Graph-Connectivity

Prof. Dr. Rolf Backofen

Bioinformatics Group / Department of Computer Science

Algorithms and Data Structures, January 2019

## Structure

## Graphs

Introduction Implementation Application example

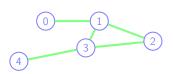
# Graphs Introduction

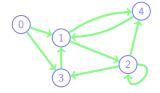
#### Graphs - Overview:

- Besides arrays, lists and trees the most common data structure (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

Introduction

### Terminology:

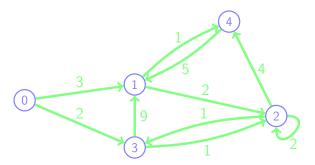




- ▶ Each graph G = (V, E) consists of:
  - A set of vertices (nodes)  $V = \{v_1, v_2, \dots\}$
  - A set of edges (arcs)  $E = \{e_1, e_2, \dots\}$
- ▶ Each edge connects two vertices  $(u, v \in V)$ 
  - ▶ Undirected edge:  $e = \{u, v\}$  (set)
  - ▶ Directed edge: e = (u, v) (tuple)
- ▶ Self-loops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

Introduction

### 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

#### Introduction

#### Example: Road network

► Intersections: vertices

► Roads: edges

Travel time: costs of the edges



Figure: Map of Freiburg © OpenStreetMap

Implementation

### How to represent this graph computationally?

1. Adjacency matrix with space consumption  $\Theta(|V|^2)$ 

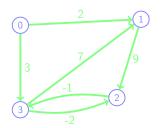


Figure: Weighted graph with |V| = 4, |E| = 6

		ena-vertice			
		0	1	2	3
<u>e</u> (	0		2		3
:-vertice	1			9	
<del>رّ</del> (	2				-1
start	3		7	-2	

Figure: Adjacency matrix

Implementation

### How to represent this graph computationally?

2. Adjacency list / fields with space consumption  $\Theta(|V| + |E|)$  Each list item stores the target vertex and the cost of the edge

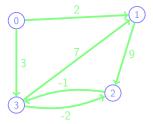


Figure: Weighted graph with |V| = 4, |E| = 6

<u>o</u> 0	1, 2	3, 3
vertice	2, 9	
.'. (2)	3, -1	
start	1, 7	2, -2

Figure: Adjacency list

Implementation

#### **Graph: Arrangement**

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

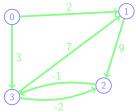


Figure: Weighted graph with |V| = 4, |E| = 6

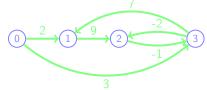


Figure: Same graph ordered by number - outer planar graph

Implementation - Python

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

Degrees (Valency)

**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 head ends adjacent to the vertex

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

► Outdegree of a vertex *u* is the number of edge tail ends adjacent to the vertex

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

Degrees (Valency)

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

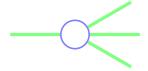


Figure: Vertex with degree of 4

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

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

#### **Paths**

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

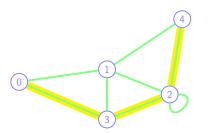


Figure: Undirected path of length 3 P = (0, 3, 2, 4)

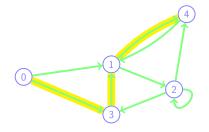


Figure: Directed path of length 3 P = (0, 3, 1, 4)

- ▶ 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**

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

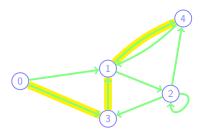


Figure: Directed path of length 3 P = (0, 3, 1, 4)

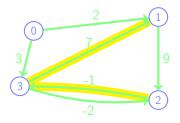


Figure: Weighted path with cost 6 P = (2,3,1)

- ► 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

**Paths** 

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

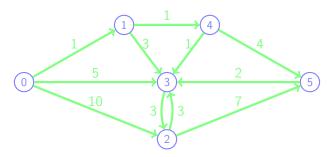


Figure: Shortest path from 0 to 2 with cost / distance d(0,2) = ?

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)

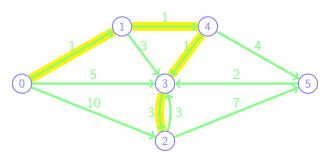


Figure: Shortest path from 0 to 2 with cost / distance d(0,2) = 6P = (0,1,4,3,2)

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

**Paths** 

Diameter of a graph: 
$$G = (V, E)$$
  $d = \max_{u,v \in V} d(u,v)$ 

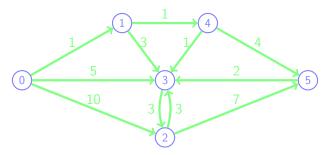


Figure: Diameter of graph is d = ?

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

Diameter of a graph: 
$$G = (V, E)$$
  $d = \max_{u,v \in V} d(u, v)$ 

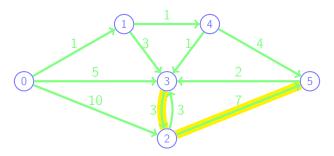


Figure: Diameter of graph is d = 10, P = (3, 2, 5)

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

#### Connected Components

## Connected components: G = (V, E)

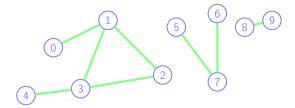


Figure: Three connected components

- 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** 

## Connected components: G = (V, E)

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

Connected Components - Graph Exploration

## **Graph Exploration:** (Informal definition)

- ▶ 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 order of the smallest distance to s
- Depth-first search: in order of the largest distance to s
- Not a problem on its own but is often used as subroutine of other algorithms
  - Searching of connected components
  - Flood fill in drawing programms

Connected Components - Breadth-First Search

#### **Breadth-First Search:**

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

#### Connected Components - Breadth-First Search

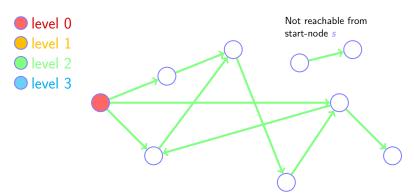


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

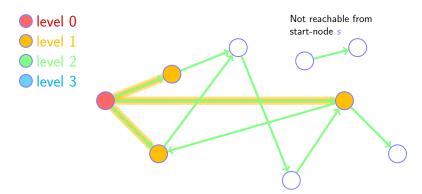


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

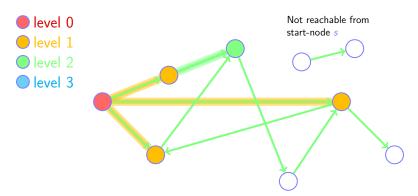


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

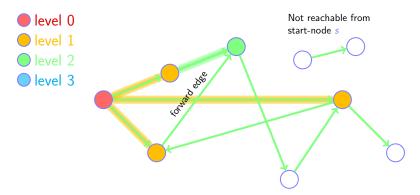


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

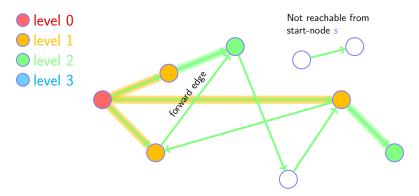


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

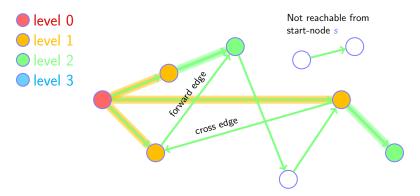


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

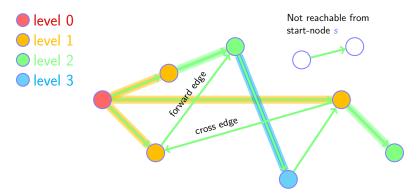


Figure: spanning tree of a breadth-first search

#### Connected Components - Breadth-First Search

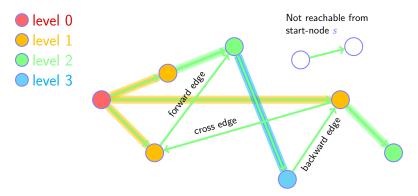


Figure: spanning tree of a breadth-first search

Connected Components - Depth-First Search

## **Depth-First Search:**

- 1. We start with all vertices unmarked and mark visited vertices
- 2. 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 and continue recursive search (reduce the recursion level by one)

Connected Components - Depth-First Search

#### Depth-first search:

- Search starts with long paths (searching with depth)
- Like breadth-first search marks 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 length from the start vertex

#### Connected Components - Depth-First Search

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

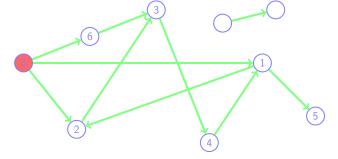


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

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

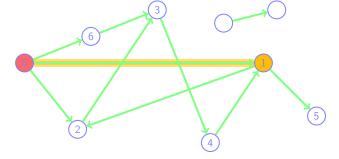


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

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

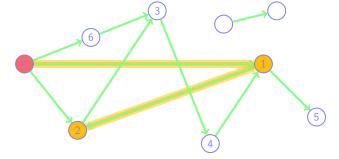


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

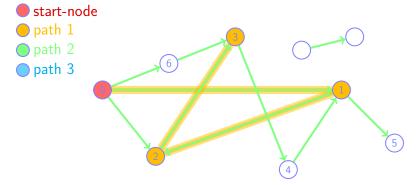


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

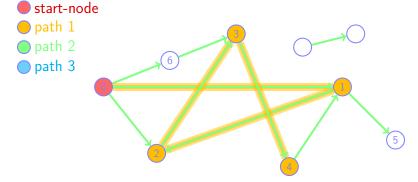


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

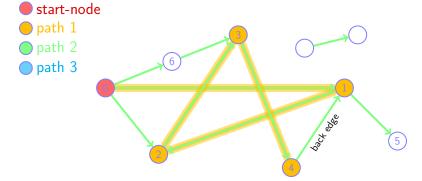


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

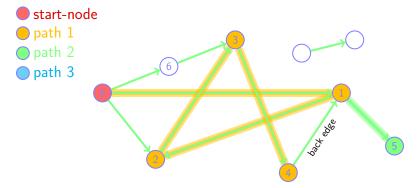


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

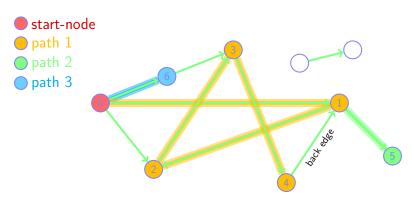


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

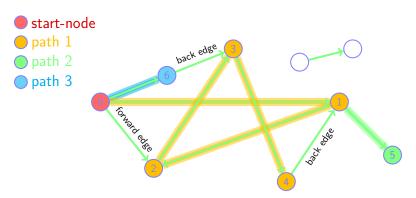


Figure: spanning tree of a depth-first search

#### Connected Components - Depth-First Search

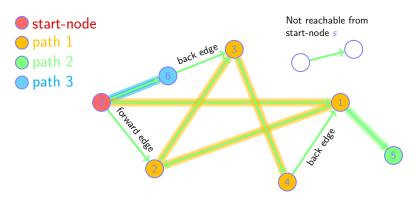


Figure: spanning tree of a depth-first search

Why is this called Breadth- and Depth-First Search?

Connected Components - Breadth-/Depth-First Search

### Runtime complexity:

- Constant costs for each visited vertex and edge
- We get a runtime complexity of  $\Theta(|V'| + |E'|)$
- $\blacktriangleright$  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

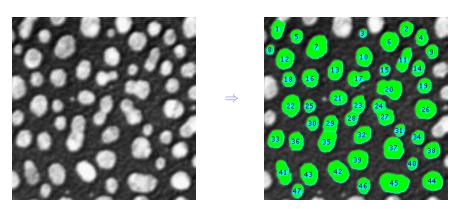


Image processing

### What is object, what is background?

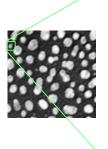


Image processing

### Convert to black and white using threshold:

value = 255 if value > 100 else 0

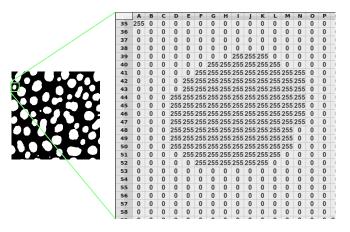
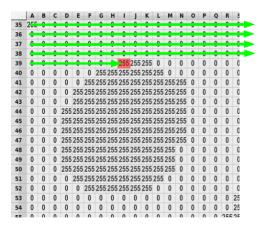


Image processing

### Interpret image as graph:

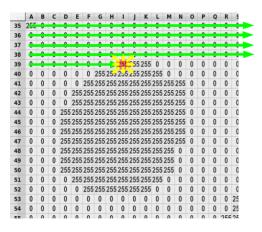
- ► Each white pixel is a node
- Edges between adjacent pixels (normally 4 or 8 neighbors)
- Edges are not saved externally, algorithm works directly on array
- Breadth- / depth-first search find all connected components (particles)

Image processing



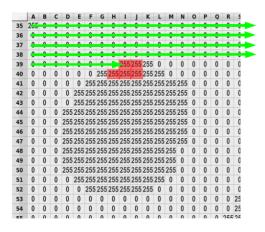
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1

Image processing



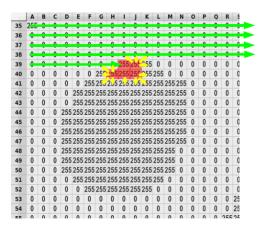
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels

Image processing



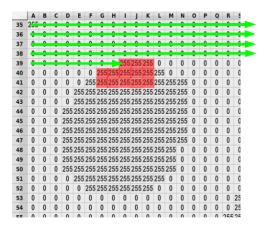
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- ► Label non-zero pixels as component 1

Image processing



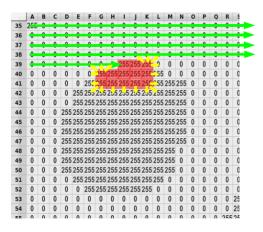
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing



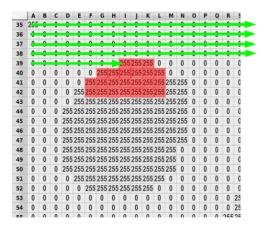
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- ► Label non-zero pixels as component 1

Image processing



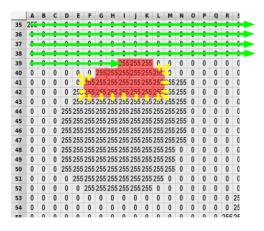
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing



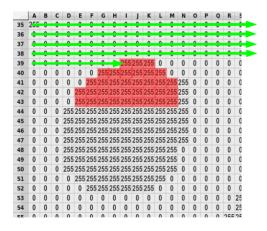
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- ► Label non-zero pixels as component 1

Image processing



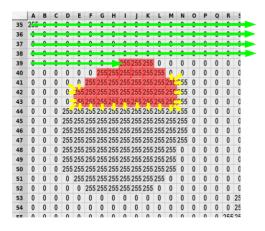
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

Image processing



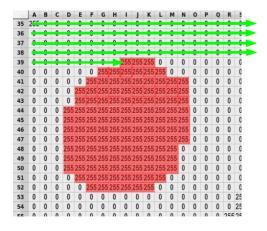
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- ► Label non-zero pixels as component 1

Image processing



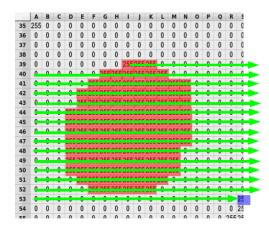
- Search pixel-by-pixel for non-zero intensity
- ► Label found pixel as component 1
- Check neighbors of all new labeled pixels
- Label non-zero pixels as component 1

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 2
- **.**.

Image processing

### Result of connected component labeling:

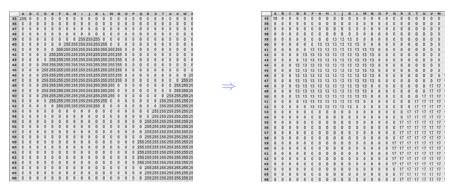


Figure: Result: particle indices instead of intensities

### Further Literature

#### General

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

### Further Literature

### Graph Search

### Graph Connectivity

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