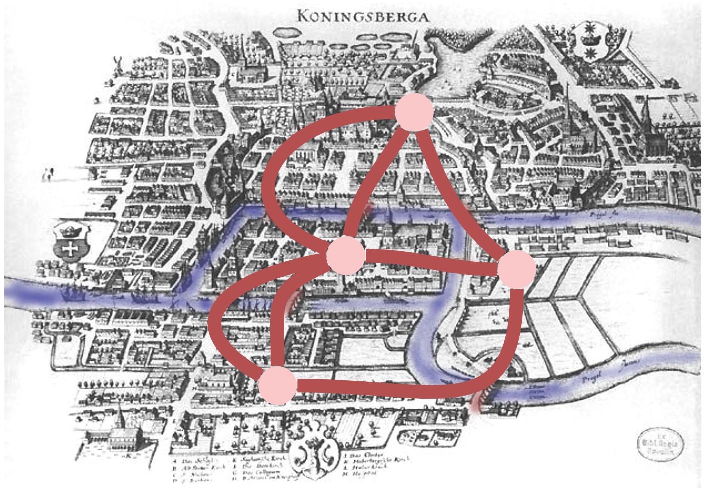
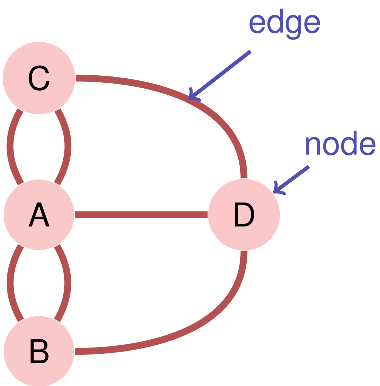
22. Graphs

Reflexive transitive closure, Graph Traversal (DFS, BFS), Connected components, Topological Sorting Ottman/Widmayer, Kap. 9.1 - 9.4,Cormen et al, Kap. 22

Königsberg 1736



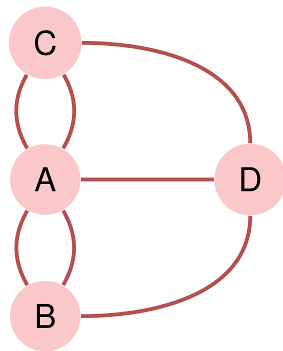
Graph



Cycles

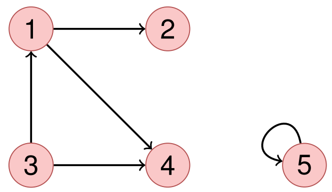
- Is there a cycle through the town (the graph) that uses each bridge (each edge) exactly once?
- Euler (1736): no.
- Such a *cycle* is called *Eulerian path*.
- Eulerian path $\Leftrightarrow$ each node provides an even number of edges (each node is of an *even degree*).

  '$\Rightarrow$" ist straightforward, "$\Leftarrow$" ist a bit more difficult



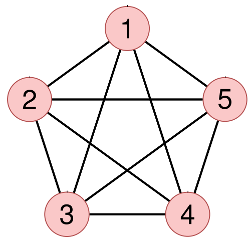
Notation

A *directed graph* consists of a set $V = \{v_1, \dots, v_n\}$ of nodes (*Vertices) and a set* $E \subseteq V \times V$ *of* Edges. The same edges may not be contained more than once.



Notation

An *undirected graph* consists of a set $V = \{v_1, \dots, v_n\}$ of nodes a and a set $E \subseteq \{\{u, v\} | u, v \in V\}$ of edges. Edges may bot be contained more than once.[31]



a complete undirected graph

[31] As opposed to the introductory example – otherwise call it multi-graph.

Notation

A graph $G = (V, E)$ with E comprising all edges is called *complete*.

A graph where V can be partitioned into disjoint sets U and W such that each $e \in E$ provides a node in U and a node in W is called *bipartite*.

A *weighted graph* $G = (V, E, c)$ is a graph $G = (V, E)$ with an *edge weight function* $c : E \to \mathbb{R}$. $c(e)$ is called *weight* of the edge e.

Notation

For directed graphs $G = (V, E)$

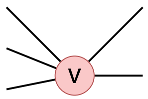
- $w \in V$ is called adjacent to $v \in V$, if $(v, w) \in E$
- *Predecessors* of $v \in V$: $N^-(v) := \{u \in V | (u, v) \in E\}$.
  Successors: $N^+(v) := \{u \in V | (v, u) \in E\}$
- *In-Degree*: $\deg^-(v) = |N^-(v)|$,
  Out-Degree: $\deg^+(v) = |N^+(v)|$

$\deg^-(v) = 3, \deg^+(v) = 2$ $\quad$ $\deg^-(w) = 1, \deg^+(w) = 1$

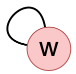
Notation

For undirected graphs $G = (V, E)$:

- $w \in V$ is called *adjacent* to $v \in V$, if $\{v, w\} \in E$
- *Neighbourhood* of $v \in V$: $N(v) = \{w \in V | \{v, w\} \in E\}$
- *Degree* of v: $\deg(v) = |N(v)|$ with a special case for the loops: increase the degree by 2.



$\deg(v) = 5$



$\deg(w) = 2$

Relationship between node degrees and number of edges

For each graph $G = (V, E)$ it holds

1. $\sum_{v \in V} \deg^-(v) = \sum_{v \in V} \deg^+(v) = |E|$, for G directed
2. $\sum_{v \in V} \deg(v) = 2|E|$, for G undirected.

Paths

- *Path*: a sequence of nodes $\langle v_1, \ldots, v_{k+1} \rangle$ such that for each $i \in \{1 \ldots k\}$ there is an edge from v_i to v_{i+1} .
- *Length* of a path: number of contained edges k.
- *Weight* of a path (in weighted graphs): $\sum_{i=1}^{k} c((v_i, v_{i+1}))$ (bzw. $\sum_{i=1}^{k} c(\{v_i, v_{i+1}\})$)
- *Simple path*: path without repeating vertices
- *Connected*: undirected graph where for each pair $v, w \in V$ there is a connecting path.

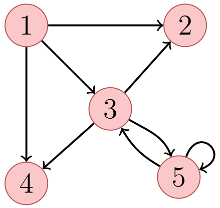
Cycles

- *Cycle*: path $\langle v_1, \ldots, v_{k+1} \rangle$ with $v_1 = v_{k+1}$
- *Simple cycle*: Cycle with pairwise different $v_1, \ldots, v_k$, that does not use an edge more than once.
- *Acyclic*: graph without any cycles.

Conclusion: undirected graphs cannot contain cycles with length 2 (loops have length 1)

Representation using a Matrix

Graph $G = (V, E)$ with nodes $v_1 \dots, v_n$ stored as *adjacency matrix* $A_G = (a_{ij})_{1 \le i,j \le n}$ with entries from $\{0, 1\}$. $a_{ij} = 1$ if and only if edge from v_i to v_j.

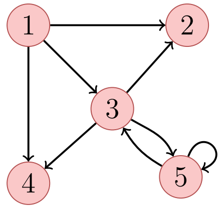


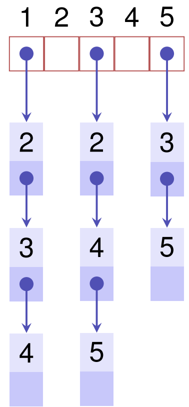
$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Memory consumption $\Theta(|V|^2)$. A_G is symmetric, if G undirected.

Representation with a List

Many graphs $G = (V, E)$ with nodes $v_1, \dots, v_n$ provide much less than n^2 edges. Representation with *adjacency list*: Array $A[1], \dots, A[n]$, A_i comprises a linked list of nodes in $N^+(v_i)$.



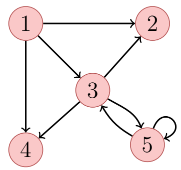


Memory Consumption $\Theta(|V| + |E|)$.

Runtimes of simple Operations

Operation	Matrix	List
Find neighbours of $v \in V$	$\Theta(n)$	$\Theta(\deg^+ v)$
find $v \in V$ without neighbour	$\mathcal{O}(n^2)$	$\mathcal{O}(n)$
$(u, v) \in E$?	$\mathcal{O}(1)$	$\mathcal{O}(\deg^+ v)$
Insert edge	$\mathcal{O}(1)$	$\mathcal{O}(1)$
Delete edge	$\mathcal{O}(1)$	$\mathcal{O}(\deg^+ v)$

Adjacency Matrix Product



$$B := A_G^2 = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 \end{pmatrix}$$

Interpretation

Theorem

Let $G = (V, E)$ be a graph and $k \in \mathbb{N}$. Then the element $a_{i,j}^{(k)}$ of the matrix $(a_{i,j}^{(k)})_{1 \le i,j \le n} = A_G^k$ provides the number of paths with length k from v_i to v_j .

Proof

By Induction.
Base case: straightforward for $k = 1$. $a_{i,j} = a_{i,j}^{(1)}$.
Hypothesis: claim is true for all $k \le l$
Step ($l \to l+1$):

$$a_{i,j}^{(l+1)} = \sum_{k=1}^{n} a_{i,k}^{(l)} \cdot a_{k,j}$$

$a_{k,j} = 1$ iff egde k to j, 0 otherwise. The sum above counts the number of nodes having a direct connection to v_j where a path of length l exists from v_i i.e. all paths with length $l+1$.

Shortest Path

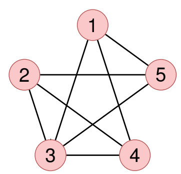
Question: is there a path from i to j? How long is the shortest path?

Answer: exponentiate A_G until for some $k < n$ it holds that $a_{i,j}^{(k)} > 0$. k provides the path length of the shortest path. If $a_{i,j}^{(k)} = 0$ for all $1 \le k < n$, then there is no path from i to j.

Number triangles

Question: How many triangular path does an undirected graph contain?

Answer: Remove all cycles (diagonal entries). Compute A_G^3. $a_{ii}^{(3)}$ determines the number of paths of length 3 that contain i. There are 6 different permutations of a triangular path. Thus for the number of triangles: $\sum_{i=1}^{n} a_{ii}^{(3)}/6$.



$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}^3 = \begin{pmatrix} 4 & 4 & 8 & 8 & 8 \\ 4 & 4 & 8 & 8 & 8 \\ 8 & 8 & 8 & 8 & 8 \\ 8 & 8 & 8 & 4 & 4 \\ 8 & 8 & 8 & 4 & 4 \end{pmatrix} \Rightarrow 24/6 = 4$$

Dreiecke.

Graphs and Relations

Graph $G = (V, E)$ with adjacencies $A_G \widehat{=}$ Relation $E \subseteq V \times V$ over V

- *reflexive* $\Leftrightarrow a_{i,i} = 1$ for all $i = 1, \dots, n$.
- *symmetric* $\Leftrightarrow a_{i,j} = a_{j,i}$ for all $i, j = 1, \dots, n$ (undirected)
- *transitive* $\Leftrightarrow (u, v) \in E, (v, w) \in E \Rightarrow (u, w) \in E$.

Equivalence relation $\Leftrightarrow$ collection of complete, undirected graphs where each element has a loop.

Reflexive transitive closure of G $\Leftrightarrow$ *Reachability relation* E^*: $(v, w) \in E^*$ iff $\exists$ path from node v to w.

Computation of the Reflexive Transitive Closure

Goal: computation of $B = (b_{ij})_{1 \le i,j \le n}$ with $b_{ij} = 1 \Leftrightarrow (v_i, v_j) \in E^*$

Observation: $a_{ij} = 1$ already implies $(v_i, v_j) \in E^*$.

First idea:

- Start with $B \leftarrow A$ and set $b_{ii} = 1$ for each i (Reflexivity.).
- Iterate over i, j, k and set $b_{ij} = 1$, if $b_{ik} = 1$ and $b_{kj} = 1$. Then all paths with lenght 1 and 2 taken into account.
- Repeated iteration $\Rightarrow$ all paths with length $1 \dots 4$ taken into account.
- $\lceil \log_2 n \rceil$ iterations required.

Improvement: Algorithm of Warshall (1962)

Inductive procedure: all paths known over nodes from $\{v_i : i < k\}$. Add node v_k.

Algorithm ReflexiveTransitiveClosure(A_G)

Input : Adjacency matrix $A_G = (a_{ij})_{i,j=1}^n$
Output : Reflexive transitive closure $B = (b_{ij})_{i,j=1}^n$ of G

```
B ← A_G
for k ← 1 to n do
    a_kk ← 1                                  // Reflexivity
    for i ← 1 to n do
        for j ← 1 to n do
            b_ij ← max{b_ij, b_ik · b_kj}     // All paths via v_k
return B
```

Runtime $\Theta(n^3)$.

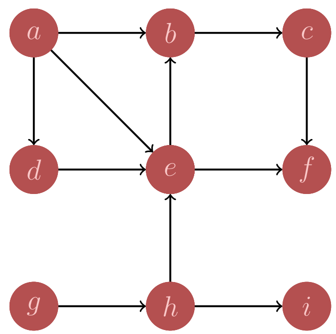
Correctness of the Algorithm (Induction)

Invariant (k): all paths via nodes with maximal index $< k$ considered.

- Base case ($k = 1$): All directed paths (all edges) in A_G considered.
- Hypothesis: invariant (k) fulfilled.
- Step ($k \to k + 1$): For each path from v_i to v_j via nodes with maximal index k: by the hypothesis $b_{ik} = 1$ and $b_{kj} = 1$. Therefore in the k-th iteration: $b_{ij} \leftarrow 1$.

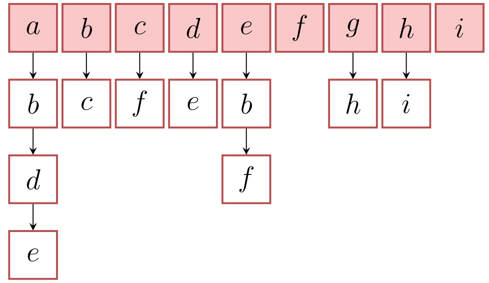
Graph Traversal: Depth First Search

Follow the path into its depth until nothing is left to visit.



Order $a, b, c, f, d, e, g, h, i$

Adjazenzliste



Algorithm Depth First visit DFS-Visit(G, v)

Input : graph $G = (V, E)$, Knoten v.

Mark v visited
foreach $(v, w) \in E$ **do**
  if $\neg(w$ visited$)$ **then**
    DFS-Visit(w)

Depth First Search starting from node v. Running time (without recursion): $\Theta(\deg^+ v)$

Algorithm Depth First visit DFS-Visit(G)

Input : graph $G = (V, E)$

foreach $v \in V$ **do**
  if $\neg(v$ visited$)$ **then**
    DFS-Visit(G,v)

Depth First Search for all nodes of a graph. Running time: $\Theta(|V| + \sum_{v \in V}(\deg^+(v) + 1)) = \Theta(|V| + |E|)$.

Problem with recursion?
With large graphs a stack overflow can happen.

Iterative DFS-Visit(G, v)

Input : graph $G = (V, E)$

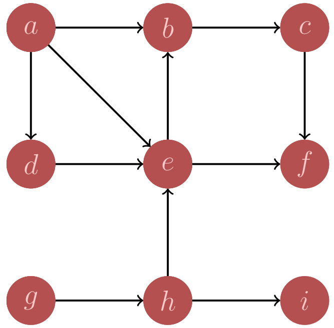
Stack $S \leftarrow \emptyset$; push(S, v)
while $S \neq \emptyset$ **do**
  $w \leftarrow$ pop(S)
  if $\neg(w$ visited$)$ **then**
    mark w visited
    foreach $(w, c) \in E$ **do** // (in reverse order, potentially)
      if $\neg(c$ visited$)$ **then**
        push(S, x)

Stack size up to $|E|$, for each node an extra of $\Theta(\deg^+(w) + 1)$ operations. Overal: $\mathcal{O}(|V| + |E|)$

Including all calls from the above main program: $\Theta(|V| + |E|)$

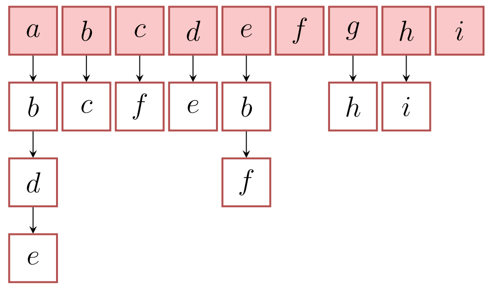
Graph Traversal: Breadth First Search

Follow the path in breadth and only then descend into depth.



Order $a, b, d, e, c, f, g, h, i$

Adjazenzliste



Iterative BFS-Visit(G, v)

Input : graph $G = (V, E)$

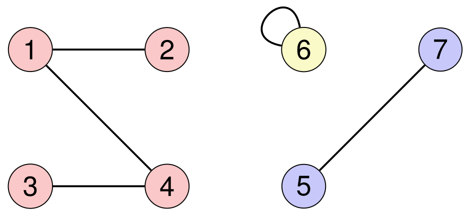
Queue $Q \leftarrow \emptyset$
Mark v as active
enqueue(Q, v)
while $Q \neq \emptyset$ **do**
  $w \leftarrow$ dequeue(Q)
  mark w visited
  foreach $(w, c) \in E$ **do**
    if $\neg(c$ visited $\vee$ c active$)$ **then**
      Mark c as active
      enqueue(Q, c)

- Algorithm requires extra space of $\mathcal{O}(|V|)$.(Why does that simple approach not work with DFS?)
- Running time including main program: $\Theta(|V| + |E|)$.

Connected Components

Connected components of an undirected graph G: equivalence classes of the reflexive, transitive closure of G. Connected component = subgraph $G' = (V', E')$, $E' = \{\{v, w\} \in E | v, w \in V'\}$ with

$\{\{v, w\} \in E | v \in V' \vee w \in V'\} = E = \{\{v, w\} \in E | v \in V' \wedge w \in V'\}$



Graph with connected components $\{1, 2, 3, 4\}$, $\{5, 7\}$, $\{6\}$.

Computation of the Connected Components

- Computation of a partitioning of V into pairwise disjoint subsets $V_1, \ldots, V_k$
- such that each V_i contains the nodes of a connected component.
- Algorithm: depth-first search or breadth-first search. Upon each new start of DFSSearch(G, v) or BFSSearch(G, v) a new empty connected component is created and all nodes being traversed are added.

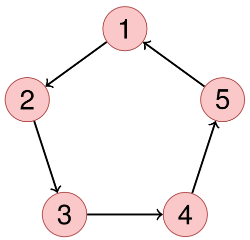
Topological Sorting

Topological Sorting of an acyclic directed graph $G = (V, E)$:
Bijective mapping

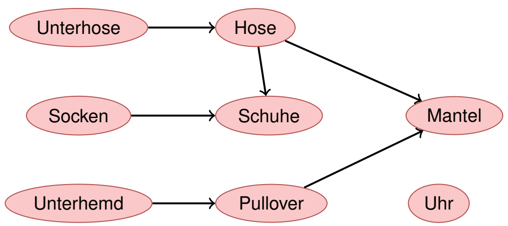
$$\mathrm{ord} : V \to \{1, \ldots, |V|\} \quad | \quad \mathrm{ord}(v) < \mathrm{ord}(w) \ \forall\, (v, w) \in E.$$

Can identify i with v_i. Topological sorting $\widehat{=} \langle v_1, \ldots, v_{|V|} \rangle$.

(Counter-)Examples



Cyclic graph: cannot be sorted topologically.



A possible toplogical sorting of the graph:
Unterhemd,Pullover,Unterhose,Uhr,Hose,Mantel,Socken,Schuhe

Observation

Theorem

A directed graph $G = (V, E)$ permits a topological sorting if and only if it is acyclic.

Proof "$\Rightarrow$": If G contains a cycle it cannot permit a topological sorting, because in a cycle $\langle v_{i_1}, \ldots, v_{i_m} \rangle$ it would hold that $v_{i_1} < \cdots < v_{i_m} < v_{i_1}$.

Inductive Proof Opposite Direction

- Base case ($n = 1$): Graph with a single node without loop can be sorted topologically, set$\mathrm{ord}(v_1) = 1$.
- Hypothesis: Graph with n nodes can be sorted topologically
- Step ($n \to n + 1$):
  1. G contains a node v_q with in-degree $\deg^-(v_q) = 0$. Otherwise iteratively follow edges backwards – after at most $n + 1$ iterations a node would be revisited. Contradiction to the cycle-freeness.
  2. Graph without node v_q and without its edges can be topologically sorted by the hypothesis. Now use this sorting and set $\mathrm{ord}(v_i) \leftarrow \mathrm{ord}(v_i) + 1$ for all $i \neq q$ and set $\mathrm{ord}(v_q) \leftarrow 1$.

Preliminary Sketch of an Algorithm

Graph $G = (V, E)$. $d \leftarrow 1$

1. Traverse backwards starting from any node until a node v_q with in-degree 0 is found.
2. If no node with in-degree 0 found after n stepsm, then the graph has a cycle.
3. Set $\mathrm{ord}(v_q) \leftarrow d$.
4. Remove v_q and his edges from G.
5. If $V \neq \emptyset$, then $d \leftarrow d + 1$, go to step 1.

Worst case runtime: $\Omega(|V|^2)$.

Improvement

Idea?

Compute the in-degree of all nodes in advance and traverse the nodes with in-degree 0 while correcting the in-degrees of following nodes.

Algorithm Topological-Sort(G)

Input : graph $G = (V, E)$.
Output : Topological sorting ord

```
Stack S ← ∅
foreach v ∈ V do A[v] ← 0
foreach (v, w) ∈ E do A[w] ← A[w] + 1 // Compute in-degrees
foreach v ∈ V with A[v] = 0 do push(S, v) // Memorize nodes with in-degree 0
i ← 1
while S ≠ ∅ do
    v ← pop(S); ord[v] ← i; i ← i + 1 // Choose node with in-degree 0
    foreach (v, w) ∈ E do // Decrease in-degree of successors
        A[w] ← A[w] − 1
        if A[w] = 0 then push(S, w)
if i = |V| + 1 then return ord else return "Cycle Detected"
```

Algorithm Correctness

Theorem

Let $G = (V, E)$ be a directed acyclic graph. Algorithm TopologicalSort(G) computes a topological sorting ord for G with runtime $\Theta(|V| + |E|)$.

Proof: follows from previous theorem:

1. Decreasing the in-degree corresponds with node removal.
2. In the algorithm it holds for each node v with $A[v] = 0$ that either the node has in-degree 0 or that previously all predecessors have been assigned a value $\mathrm{ord}[u] \leftarrow i$ and thus $\mathrm{ord}[v] > \mathrm{ord}[u]$ for all predecessors u of v. Nodes are put to the stack only once.
3. Runtime: inspection of the algorithm (with some arguments like with graph traversal)

Algorithm Correctness

Theorem

Let $G = (V, E)$ be a directed graph containing a cycle. Algorithm TopologicalSort(G) terminates within $\Theta(|V| + |E|)$ steps and detects a cycle.

Proof: let $\langle v_{i_1}, \ldots, v_{i_k} \rangle$ be a cycle in G. In each step of the algorithm remains $A[v_{i_j}] \geq 1$ for all $j = 1, \ldots, k$. Thus k nodes are never pushed on the stack und therefore at the end it holds that $i \leq V + 1 - k$.

The runtime of the second part of the algorithm can become shorter. But the computation of the in-degree costs already $\Theta(|V| + |E|)$.