# Practical Session No. 12 – Graphs, BFS, DFS, Topological sort

## Graphs and BFS

<table>
<thead>
<tr>
<th>Graphs and BFS</th>
<th>Graphs and BFS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Graph</strong> G = (V, E)</td>
<td><strong>V(G) = V</strong> - Set of all vertices in G <strong>E(G) = E</strong> - Set of all edges (u,v) in G, where u,v in V(G)</td>
</tr>
<tr>
<td><strong>Graph Representations</strong> (V(G)={v₁,…vₙ})</td>
<td><strong>Adjacency Matrix:</strong> Dimensions</td>
</tr>
<tr>
<td></td>
<td>M[i,j] = 1 - (vᵢ, vⱼ) in E(G) M[i,j] = 0 - (vᵢ, vⱼ) not in E(G)</td>
</tr>
<tr>
<td></td>
<td><strong>Adjacency List:</strong> Each vertex v in V(G) has a list of all vertices u in V(G) where (v,u) in E(G)</td>
</tr>
<tr>
<td><strong>BFS(G, s)</strong></td>
<td><strong>Breadth-First Search</strong></td>
</tr>
<tr>
<td></td>
<td>Starting from the source node s, BFS computes the minimal distance from s to any other node v that can be reached from s. The algorithm builds a breadth-tree rooted at s with the minimal paths to nodes that can be reached from s. The algorithm is suitable for directed or undirected graphs.</td>
</tr>
<tr>
<td></td>
<td>- <strong>color[v]</strong> - the color of node v</td>
</tr>
<tr>
<td></td>
<td>o <strong>white</strong> - an unvisited node. The first time it is encountered during the search it becomes nonwhite.</td>
</tr>
<tr>
<td></td>
<td>o <strong>gray</strong> - a visited node which has white neighbors (draws a frontier between the visited and unvisited nodes)</td>
</tr>
<tr>
<td></td>
<td>o <strong>black</strong> - a node whose neighbors were visited (colored gray or black)</td>
</tr>
<tr>
<td></td>
<td>- <strong>pred[v]</strong> - the predecessor of v in the breadth-tree</td>
</tr>
<tr>
<td></td>
<td>- <strong>d[v]</strong> - the minimal distance from s to v</td>
</tr>
<tr>
<td></td>
<td>Time complexity: O((</td>
</tr>
</tbody>
</table>
### DFS,

<table>
<thead>
<tr>
<th>DFS(G)</th>
<th>Depth-First Search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Strategy: search &quot;deeper&quot; in the graph whenever possible.</td>
</tr>
<tr>
<td></td>
<td>- Edges are explored out of the most recently discovered vertex $v$ that still has unexplored edges leaving it. When all of $v$'s edges have been explored, the search &quot;backtracks&quot; to explore edges leaving the vertex from which $v$ was discovered. This process continues until we have discovered all the vertices that are reachable from the original source vertex.</td>
</tr>
<tr>
<td></td>
<td>- If any undiscovered vertices remain, then one of them is selected as a new source and the search is repeated from that source. This entire process is repeated until all vertices are discovered.</td>
</tr>
<tr>
<td></td>
<td>- The DFS may result in several trees, depending on the order of choosing the first vertex in the set of unvisited vertices.</td>
</tr>
<tr>
<td></td>
<td>- <code>color[v]</code> - the color of node $v$</td>
</tr>
<tr>
<td></td>
<td>- o white - an unvisited node</td>
</tr>
<tr>
<td></td>
<td>- o gray - a visited node</td>
</tr>
<tr>
<td></td>
<td>- o black - a node whose adjacency list has been examined completely (whose descendants were all searched).</td>
</tr>
<tr>
<td></td>
<td>- <code>pred[v]</code> - the predecessor of $v$ in the one of the depth trees</td>
</tr>
<tr>
<td></td>
<td>- <code>d[v]</code> - the time when $v$ was discovered</td>
</tr>
<tr>
<td></td>
<td>- <code>f[v]</code> - the time when $v$ was colored in black</td>
</tr>
</tbody>
</table>

A node $v$ is white before $d[v]$, gray between $d[v]$ and $f[v]$ and black after $f[v]$.

**Time complexity:** $O(|V| + |E|)$, sometimes written as $O(V+E)$
DFS can be used to classify the edges of the input graph $G=(V,E)$.

1. **Tree-edges**
   - $(u,v)$ - if $v$ was initially discovered using edge from $u$ to $v$
2. **Back-edges**
   - $(u,v)$ - if $v$ is an ancestor of $u$ in the depth-tree
3. **Forward-edges**
   - $(u,v)$ - not a tree-edge, where $u$ is $v$'s ancestor
4. **Cross-edges**
   - $(u,v)$ - All the other edges, where $u$ and $v$ are vertices in different depth-tree, $u$ is not $v$'s ancestor or $v$ is not $u$'s ancestor

It can be shown that if the graph is undirected then all of its edges are tree edges or back edges. Therefore we define Forward-edges and Cross-edges only in directed graphs.

Source of some text and images in this doc: *Intro to algorithms, MIT, 2011*, recitation 13
이용 그래프 및 벡터화 스택

**Adjacency Matrix**

\[
\begin{array}{cccccccc}
  a & b & c & d & e & f & g & h \\
  a & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  b & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  c & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
  d & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
  e & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  f & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
  g & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
  h & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

**Adjacency lists**

\[
\begin{array}{cccccccc}
  a & b & c & d & e & f & g & h \\
  a & f & & & & & & \\
  b & & g & & & & & \\
  c & & a & & & & & \\
  d & f & e & & & & & \\
  e & g & & & & & & \\
  f & h & & & & & & \\
  g & b & c & & & & & \\
  h & h & & & & & & \\
\end{array}
\]
Question 1

Given an undirected graph $G = (V,E)$ represented by an adjacency list, suggest an algorithm for creating an acyclic directed graph $G' = (V,E')$ from $G$. The algorithm directs the edges in $E(G)$, such that $|E'|=|E|$

Solution:

- Number all the vertices in $V(G)$ from 1 to $|V|$.
- Initialize $E' = \{\}$.
- For each $v_i$ in $V$ and $(v_i,v_j)$ in $E$ (found through its adjacency list) such that $i < j$
  $$E' \leftarrow E' \cup (v_i,v_j)$$

Proof:

To prove $|E'| = |E|$, we note that for every $(v_i,v_j)$ in $E$, either $(i < j)$ or $(j < i)$, and since both $(v_i,v_j)$ and $(v_j,v_i)$ will appear in adjacency list of the undirected graph, exactly one of them will appear in $E'$

To prove $G'$ is acyclic assume by contradiction that there is a cycle path in $G'$ starting from vertex $v_i$. The cycle is: $V_{i_1}, V_{i_2}, \ldots, V_{i_k}, V_i$. Then according to the way $G'$ is built: $i < i_1 < \ldots < i$. Contradiction.
Question 2

Given an undirected graph \( G = (V,E) \). \( G \) is connected. For every 2 sets of vertices \( V_1 \) and \( V_2 \) such as: \( V_1 \cup V_2 \subseteq V \) we define:

\[
distance(u,v) - \text{the length of the shortest path from } u \text{ to } v \text{ in } G.
\]
\[
distance(V_1,V_2) - \text{the length of shortest path between a vertex } v_1 \text{ in } V_1 \text{ and } v_2 \text{ in } V_2.
\]

(Note: If \( V_1 \cap V_2 \neq \emptyset \) then distance\((V_1, V_2) = 0\)).

Find distance\((V_1, V_2)\) in \( O(|V| + |E|) \) time.

Solution:

We define a new Graph \( G'=(V',E') \) from \( G = (V,E) \).

We add 2 new vertices \( s \) and \( t \) to \( V \), connect \( s \) to every vertex in \( V_1 \), connect every vertex in \( V_2 \) to \( t \) and the run BFS to find the length of the shortest path from \( s \) to \( t \). The answer we get is distance\((V_1, V_2) + 2\)

1. \( V' \leftarrow V \cup \{s,t\} \)
2. \( E' \leftarrow E \cup \{(s,u) : u \in V_1 \} \cup \{(v,t) : v \in V_2 \} \)
3. Execute BFS from \( s \) // Finding \( d(t) \).
4. return \( d(t) - 2 \)

Runtime: \( O(|V|) + O(|V|) + O(|V| + |E|) + O(1) = O(|V| + |E|) \)
Question 3

Suggest an algorithm that determines whether a given undirected graph contains a cycle (of length > 1), that runs in time $O(|V|)$. Note that runtime is independent of $E$.

Solution:

Lemma: An undirected graph is acyclic if and only if a DFS yields no back edges.

Back edge $(u,v)$ is such that $v$ has already been discovered and is $u$'s ancestor. $(u,v)$ - if $v$ is an ancestor of $u$ in the depth-tree and $(v,u)$ is not a tree-edge.

The DFS algorithm can be modified to classify edges as it encounters them. In a depth-first search of an undirected graph $G$, every edge of $G$ is either a tree edge or a back edge. If we arrive to a grey vertex (which means we arrived there through back edge) we return to a vertex we already visited and therefore the graph is cyclic.

\[ \text{Acyclic}(G=(V,E)) \]

execute DFS on the graph while:
1. counting the number of visited nodes
2. If DFS visits a grey node return "cycle"

The algorithm runs in $O(|V|)$ time, because in the worst case, the algorithm will explore $|V| - 1$ before reaching a grey node (back edge) or terminating (because an undirected graph with $|V|$ or more edges always contains a cycle).
Question 4

Definition: A super-sink is a vertex with in-degree $|V|-1$ and out-degree 0 (at most one super-sink can exist in a graph).

When an adjacency-matrix representation is used, most graph algorithms require time $O(|V|^2)$. Show that determining whether a directed graph, represented in an adjacency-matrix contains a super-sink can be done in time $O(|V|)$.

Solution:

A vertex $i$ is a super-sink if and only if $M[i,j] = 0$ for all $j$ and $M[j,i] = 1$ for all $j \neq i$.

For any pair of vertices $i$ and $j$:

- $M[i,j] = 1$ → vertex $i$ can't be a super-sink
- $M[i,j] = 0$ → vertex $j$ can't be a super-sink ($i \neq j$)

Sink($M, n$)

1. $i=0$, $j=0$, $n=|V|$.
2. while $(j<n)$
   
   if $j=i$ or $M[i,j]=0$  // $v_j$ can't be a super-sink
   
   $j \leftarrow j+1$
   
   if $M[i,j]=1$  // $v_i$ can't be a super-sink
   
   $i \leftarrow i+1$
   
3. Check row $i$ – it should be all zeros,  // $v_i$ is a super-sink
4. Check column $i$ – it should be all ones, except $M[i,i]$.
5. if both conditions hold, then $i$ is a super-sink.
6. otherwise, there is no super-sink.

Note: In the while condition it is enough to check that $j<n$, since $i<n$ anyway if we assume the graph is not reflective (there are no edges $(i,i)$), then there cannot be row of 1's in the matrix. This means that the algorithm will always be above and to the right of the main diagonal.

Time Complexity: while loop takes at most $(n+n)$ steps. Checking $i^{th}$ row and column takes $(n+n)$ time. Total: $T(n) = 4n = O(|V|)$. 
## Topological sort

<table>
<thead>
<tr>
<th>Topological Sort</th>
<th>Ordering of vertices in a directed acyclic graph (DAG) $G=(V,E)$ such that if there is a path from $v$ to $u$ in $G$, then $v$ appears before $u$ in the ordering. There could be many solutions, for example:</th>
</tr>
</thead>
</table>
|                  | 1. call DFS to compute $f[v]$  
|                  | 2. As the visit in each vertex is finished (blackened), insert it to the head of a linked list  
|                  | 3. Return the linked list of the vertices |

Time Complexity: $O(|V|+|E|)$

---

### Question 5

2 lists are given:

A - courses list. Every student must study all the courses in A.

B – Prerequisites. B contains tuples $(a, b)$ (where $a$ and $b$ are in A), indicating that course $a$ must be taken before course $b$. The prerequisites dependencies are acyclic.

Design a schedule for each of the following students.

1. A lazy student who wants to take only one course in a semester.
2. A student who wants to take all the courses in A in the minimal number of semesters. The student is willing to take any number of courses in a semester.

**Example:**

A = { Alg, Eng, Ds1, Ds2, Mat, Ph1, Ph2 }

B = { (Alg, Ds2), (Ds1, Ds2), (Mat, Ds1), (Ph1, Ph2) }

Optional output for student no. 1:

- Semester 1: Eng
- Semester 2: Mat
- Semester 3: Alg
- Semester 4: Ds1
- Semester 5: Ds2
- Semester 6: Ph1
- Semester 7: Ph2

Optional output for student no. 2:

- Semester 1: Eng, Mat, Ph1, Alg
- Semester 2: Ds1, Ph2
- Semester 3: Ds2

**Solution:**

Assume that the courses names are integers in the range $[1..n]$, $n$ is known ($n$ is not constant). The relations between the courses will be represented by a directed graph $G =$
(V, E), where V is the set of courses and if course i is a prerequisite of course j, E will contain the edge (i, j). The graph will be represented as an adjacency list.

The graph for the given example is:

![Graph Diagram]

**Algorithm for a lazy student:** Print the courses in a way that course i won't be printed before course j, if j is a prerequisite of i. In other words, sort the courses topologically. Complexity: O(|V| + |E|)

Now, let's observe another algorithm for topological sort of a DAG in O(|V|+|E|).

- Find in-degree of all the vertices and keep a list of vertices with in-degree=0 - O(|V|+|E|)
- While V is not empty do:
  - Extract a vertex v with in-degree of 0 from the list - O(1)
  - Output v and remove it from G, along with its edges,
  - Reduce the in-degree of each node u such as (v,u) was an edge in G and add to the list vertices with in-degree=0, if necessary - O(d(v))

**Algorithm for a hard working student:** A variation of the above topological sort algorithm with a slight change. In each semester k, execute the algorithm on all the nodes with degree 0 simultaneously (instead of dealing with one source at each stage, all the sources will be dealt and printed in one semester.)

(note that the topological sort is not necessary because after we find the vertexes with in-degree 0, we can go to the next vertexes through the edges.)

Complexity: O(|V| + |E|)
**Question 6**

Given a DAG (directed acyclic graph) $G=(V,E)$ with weighted edges. Suggest an $O(|V|+|E|)$ time algorithm for finding the weight of the maximum weight path in $G$.

**Solution:**

1. Topologically sort $V$ - $O(V+E)$.
   (We'll denote the topologically ordered vertex list as $(v_1, v_2, \ldots, v_n)$).
2. For every vertex $v$, create a list of all in-edges, i.e., list of all the vertices $u$, such that $(u,v) \in E$ - $O(V+E)$.
3. Define a new array $A$ for which $A[i]$ = the maximal weighted path that ends in $v_i$. By definition, $A[i] = \max\{A[j]+w(j,i) : (j,i) \in E\}$ - $O(V+E)$.
4. Start by computing $A[i]$ for all $i = 1..n$, where $v_1, v_2, \ldots, v_n$ is the topological ordering.
   Since the graph was topologically sorted, when computing $A[i]$, all the values $A[j]$ of it’s neighbors $j$ were already computed (all $i$’s neighbors appear before it in the topological ordering).
5. After computing $A[i]$ for all the nodes, find $v_k$ for which $A[k]$ is maximal.

MaxPath(G)
- $v_1, v_2, \ldots, v_n \leftarrow$ TopologicalSort(G)
- Create-in-edges-lists()
- $A \leftarrow$ new array [1..n]
- $A[1] \leftarrow 0$
- for $i \leftarrow 2$ to $n$ do
  - $A[i] \leftarrow \max\{A[j]+w(j,i) : j \in$ in_edges_list[i]\}
- return $\max\{A[1],A[2],\ldots,A[n]\}$

We design the algorithm
1. Express the solution to a problem in terms of solutions to smaller problems.
2. Solve all the smallest problems first and put their solutions in a table
3. Solve the next larger problems, and so on, up to the problem, we originally wanted to solve.
4. Each problem should be easily solvable by looking up and combining solutions of smaller problems in the table.

Time Complexity: $O(|V|+|E|)$
For further practice:

Question 1

Update the BFS algorithm so that every vertex \( v \) in the graph contains not only \( d[v] \), the length of shortest path from \( s \) to \( v \), but also the number of different paths of that length.

Solution:

Each node \( u \) in the graph will contain an extra field: \( M(u) \) - the number of different paths to \( u \) of length \( d(u) \).

1. Initialize \( M(v) \leftarrow 0 \) for all the vertices except \( s \). Initialize \( M(s) \leftarrow 1 \).
2. For each vertex \( u \), when going over the neighbors \( v \) of \( u \):
   a. if \( v \) is a white neighbor then
      \( M(v) \leftarrow M(u) \)
   b. else
      // \( v \) is not a white neighbor of \( u \)
      if \( d(v) = d(u) + 1 \)
      \( M(v) \leftarrow M(v) + M(u) \)
Question 2

A road from city A to city B is represented as a pair \((A,B)\).
Assume the roads are bi-directional.
Given a set of \(m\) roads and a set of \(n\) cities, suggest a way to preprocess the input in \(O(n+m)\) time and unlimited memory, in order to support the following operation in \(O(1)\) time:

\(\text{Reachable}(i,j)\) – Returns \(TRUE\) if it is possible to drive from city \(i\) to city \(j\) and \(FALSE\) otherwise.

Solution:

Connected Components
A connected component is defined as a sub-graph where there exists a path between any two vertices in it.

Preprocessing:
1. Represent the input as an undirected graph \(G=(V,E)\), where the nodes represent the cities and the edges represent the roads. The graph represented as an adjacency list. \(O(n+m)\) time and memory.
2. Define a new array \(C\) of size \(n\), which will save the “connected component owner” for each vertex.
3. For each vertex \(v\) in \(G\)
   a. If \(v\) has no owner, it is part of an undiscovered connected component. Execute BFS or DFS starting from \(v\) and mark in \(C\) all the vertices you discover as being owned by \(v\).
   b. Else, if \(v\) has an owner, it is part of a connected component we’ve already discovered. Ignore \(v\) and move on to the next vertex.

The runtime of the preprocessing algorithm is \(O(n+m)\) since each vertex is visited twice (once by iterating through it in the outer loop, another by visiting it in BFS/DFS), and each edge is visited once (in BFS/DFS).
Side note: unlimited memory does not really help us, if we have only a limited time for preprocessing. In $O(n+m)$ time, we can only use $O(n+m)$ memory.

Query
\textbf{Reachable}(i,j): return TRUE if and only if both $i$ and $j$ belong to the same connected component, i.e. $C[i] = C[j]$.
Runtime: $O(1)$