Some notes on graphs

This summarize some important concepts that are spread out a bit in the book.

Directed graphs ("riktade grafer")

A directed graph G is a pair (V,E) of a finite set of nodes (V) and a set of edges (E). Nodes V can be anything. Edges are (ordered) pairs of nodes (u,v), where u and v are both in V. When we have an edge (u,v) we say that u "points to" v.

This means that:

(1) for any pair of nodes u and v, we either have an edge from u to v or not. We cannot have multiple edges from u to v. We can have one edge from u to v and one edge from v to u.

(2) we can have an edge from a node u to itself, this is represented by the edge (u,u).

The indegree of a node v (indeg(v)) is the number of edges that point to it. The outdegree of a node v (outdeg(v)) is the number of edges that point away from it.

Simple graphs (undirected graphs) ("enkla grafer")

A simple graph G is a pair (V,E) of a finite set of nodes (V) and a set of edges (E). Nodes V can be anything. Edges are subsets of nodes {u,v} of size 2, where u and v are both in V.

This means that:

(1) for any pair of nodes u and v, we either have an edge between u and v, or not. We cannot have multiple edges between u and v.

(2) we cannot have an edge from a node u to itself; there cannot be an edge representing this, because {u,u} does not have size 2.

The degree of a node v (deg(v)) is the number of edges that are attached to it.

Walks, paths, closed walks, cycles ("vägar", "enkla vägar", "slutna vägar", "cykler")

We have seen the following 4 concepts which exist for both directed graphs and simple graphs:

A **walk** **("väg")** is a sequence of nodes u_1, u_2, …, u_n, such that there exists an edge (u_i,u_{i+1}) (or $\{u_i,u_{i+1}\}$) between every pair of adjacent nodes in the sequence. The **length** of a walk is the number of edges. A walk of length 0 (the empty walk) is thus a sequence with only one node u.

A **path** **("enkel väg")** is a walk where no node appears more than once.

A **closed walk** **("sluten väg")** is a walk that begins and ends with the same node. A special example of a closed walk is the empty walk.

A **cycle** **("cykel")** is a closed walk where no node appears more than once, except for the beginning and the end node. In a directed graph, we require a cycle to involve at least 1 edge. In an undirected graph, a cycle has to have at least 3 edges. (The empty walk is thus never a cycle.)

In these notes, I will use the following notation for paths:

u --p-- v means that there exists a walk/path p from u to v.

u --p-- v --q-- w means that there exists a walk/path p from u to v and a walk/path q from v to w.

Some lemmas involving walks, paths, cycles

These are all proved in the book. These all hold for directed as well as simple graphs.

* **Lemma:** If there exists a walk from u to v, there also exists a path from u to v.

Proof sketch: By strong induction on the length of the walk. Base case is easy. Step case: if the walk is already a path, we are done. Otherwise, the walk must use one point twice, so it looks like u --p-- w --q-- w --r-- v, where u and v are the beginning and end node, where w is the node we use twice, and where q is a non-empty walk. The walk u --p-- w --r-- v is shorter than the walk we started off with, and so we can invoke the induction hypothesis.

* **Lemma:** If there exists two different paths from u to v, there must be a cycle in the graph.

Proof sketch: By contradiction. Let us assume we have a graph G with two points u and v, two different paths from u to v, and no cycles. Let us also assume that the lengths of those two paths together is the smallest we can find in the whole graph. There are two cases.

1. The two paths have no node in common except for u and v. Then together those two paths form a cycle, which contradicts that we do not have a cycle.

2. The two paths have at least one node w in common. Then the two paths look like this:

path 1: u --p1-- w --p2-- v
path 2: u --q1-- w --q2-- v

Since those two paths are different, either p1 and q1 must be different, or p2 and q2 are different. (If both of these were the same, than both our paths would be the same.)

If p1 and q1 are different, then we have found two points (u and w) with two different paths p1 and q1 between them, which together are shorter than the paths we started with. Contradiction.

If p2 and q2 are different, then we have found two points (w and v) with two different paths p2 and q2 between them, which together are shorter than the paths we started with. Contradiction.

(For a different proof, using strong induction, see the end of this document.)

Trees

A simple graph is **connected** iff. for every pair of nodes u and v, there exists a walk from u to v.

A simple graph is called a **tree** if it is connected and does not contain any cycles.

There are other characterizations of trees, notably:

* **Lemma:** A connected graph with n nodes is a tree iff. it has n-1 edges

Proof: In the book. But also exercise 9 and 10 (week 7) together, plus assignment 1 (week 8).

* **Lemma:** A simple graph is a tree iff. for every pair of nodes u and v, there exists a unique path from u to v.

Proof: (1) G is a tree => for every u,v there exists a unique path: Assume G is a tree, then it is connected so there exists a walk from u to v. By the earlier lemma, there must also be a path from u to v. Assume there are two different paths from u to v, then according to an earlier lemma there exists a cycle in the graph, so it is not a tree. So any path from u to v is the only path from u to v.

(2) for every u,v there exists a unique path => G is a tree. Since there exists a path between every pair of nodes, G must be connected. If G had a cycle, we would have two nodes with two different paths between them (that together form the cycle), so G cannot have a cycle. Therefore, G is a tree.

Euler tour

Given a simple graph G.

An **Euler tour** is a closed walk where every edge of G appears exactly once.

A graph has an Euler tour iff. it is connected and every node has an even degree.

Topological ordering

Given a directed graph G.

A **topological ordering** is a sequence of nodes where every node of G appears exactly once, such that for every edge (u,v) in the graph, u appears before v in the ordering.

A graph has a topological ordering iff. there are no cycles in the graph.

(See other notes about topological ordering for more information.)

Proof of lemma by strong induction

*** Lemma:** If there exists two different paths from u to v, there must be a cycle in the graph.

Proof: By strong induction on the sum of the lengths of the two paths.

Let P(n) = “for any two nodes u, v, if we have two different paths from u to v, and the length of these paths together equals n, then the graph has a cycle”. We are going to prove by strong induction that P(n) holds for any natural number n, which implies that the lemma holds.

Base case: P(0) = “for any two nodes u, v, if we have two different paths from u to v, and the length of these paths together equals 0, then the graph has a cycle”

There cannot be two different paths of length 0 from u to v. OK.

Step case: (P(0) /\ P(1) /\ … /\ P(n)) => P(n+1)

Assume: “for any 0 <= k <= n: for any two nodes u, v, if we have two different paths from u to v, and the length of these paths together equals k, then the graph has a cycle” (I.H.)

Show: “for any two nodes u, v, if we have two different paths from u to v, and the length of these paths together equals n+1, then the graph has a cycle”

So, we have two nodes u and v, two different paths from u to v, and the lengths of those paths together is n+1. Now, there are two cases.

Case 1: The two paths have no other nodes in common than u and v. This means that those paths fused together form a cycle.

Case 2: The two paths have at least one node w in common. This means that the two paths look like this:

path 1: u --p1-- w --p2-- v
path 2: u --q1-- w --q2-- v

Since the two paths are different, either p1 and q1 must be different, or p2 and q2 are different. (If both of these were the same, than both our paths would be the same.)

If p1 and q1 are different, then we have found two points (u and w) with two different paths p1 and q1 between them, which together are shorter than the paths we started with (<n+1). We invoke the I.H. to find a cycle in the graph.

If p2 and q2 are different, then we have found two points (w and v) with two different paths p2 and q2 between them, which together are shorter than the paths we started with (<n+1). We invoke the I.H. to find a cycle in the graph.