Lecture 17: Shortest Paths I: Properties, Dijkstra's Algorithm, Breadth-first Search

Pre-listening

1) What do you know about the shortest path problem?

2) What is a graph composed of?

3) What kinds of graphs do you know?

Listening. Listen to and watch the lecture and answer questions.

1) Why did the lecturer mention the Star Wars?

....

2) Why did the lecturer mention the dynamic programming?

....

3) Why did he mention Alderon and Cambridge?

....

4) What is directed graph?

....

5) What is edge weight?

....

6) How do we denote paths?

....

7) What is directed path?

....

8) What is the weight of a path?

....

9) What is the property of a simple graph?

....

10) How can you find the shortest path from u to v?

....

Travelling salesman problem
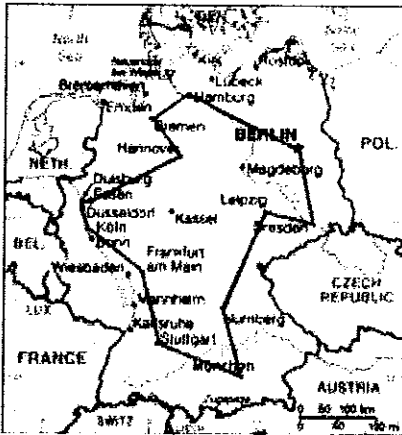
An optimal TSP tour through Germany's 15 largest cities. It is the shortest among 43 589 145 600 possible tours visiting each city exactly once.

1. Read the text and answer the Qs.
   a) What kind of problem TSP is?
   b) What does it mean: it is used as a benchmark......
   c) What are heuristic methods?
   d) What are the applications of the TSP?
   e) Which two basic concepts are there?
   f) What makes the problem hard?
   g) What are NP-complete problems?
   h) What are CPU years?

The **Travelling Salesman Problem (TSP)** is a problem in combinatorial optimization studied in operations research and theoretical computer science. Given a list of cities and their pairwise distances, the task is to find a shortest possible tour that visits each city exactly once.

The problem was first formulated as a mathematical problem in 1930 and is one of the most intensively studied problems in optimization. It is used as a benchmark for many optimization methods. Even though the problem is computationally difficult, a large number of heuristics and exact methods are known, so that some instances with tens of thousands of cities can be solved.

The TSP has several applications even in its purest formulation, such as planning, logistics, and the manufacture of microchips. Slightly modified, it appears as a sub-problem in many areas, such as DNA sequencing. In these applications, the concept *city* represents, for example, customers, soldering points, or DNA fragments, and the concept *distance* represents travelling times or cost, or a similarity measure between DNA fragments. In many applications, additional constraints such as limited resources or time windows make the problem considerably harder.

In the theory of computational complexity, the decision version of TSP belongs to the class of NP-complete problems. Thus, it is assumed that there is no efficient algorithm for solving TSPs. In other words, it is likely that the worst case running time for any algorithm for TSP increases exponentially with the number of cities, so even some instances with only hundreds of cities will take many CPU years to solve exactly.

Description - As a graph problem
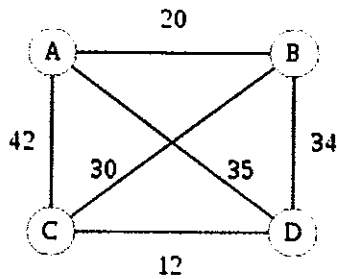
2. Fill in the missing words.

The TSP can be modeled as a 1)..........., such that cities are the graph's 2) ...., paths are the graph's 3) ...., and a path's distance is the edge's length. A TSP tour becomes a 4) ...., and the optimal TSP tour is the shortest Hamiltonian cycle. Often, the model is a 5) ....(i.e. an edge connects each pair of vertices). If no path exists between two cities, adding an arbitrarily long edge will complete the graph without affecting the optimal 6) ....

With metric distances

3. Read the next part and decide whether the statements are true or false.

   a) The connection from A to B cannot be longer than the detour through C.
   b) The Manhattan distance is the difference of x and y coordinates.
   c) The edge length is defined by a metric on the set of vertices
   d) The city block metrics is a rectilinear TSP.
   e) The drilling machine in maximum metric moves slowly to the next hole.

In the *metric TSP* the intercity distances satisfy the triangle inequality. This can be understood as "no shortcuts", in the sense that the direct connection from A to B is never longer than the detour via C.

$$c_{ij} \leq c_{ik} + c_{kj}$$

These edge lengths define a metric on the set of vertices. When the cities are viewed as points in the plane, many natural distance functions are metrics.

- In the Euclidian TSP the distance between two cities is the Euclidean distance between the corresponding points.

- In the Rectilinear TSP the distance between two cities is the sum of the differences of their x- and y-coordinates. This metric is often called the Manhattan distance or *city-block metric*.

- In the maximum metric, the distance between two points is the maximum of the differences of their x- and y-coordinates.

The last two metrics appear for example in routing a machine that drills a given set of holes in a printed circuit board. The Manhattan metric corresponds to a machine that adjusts first one co-ordinate, and then the other, so the time to move to a new point is the sum of both movements. The maximum metric corresponds to a machine that adjusts both co-ordinates simultaneously, so the time to move to a new point is the slower of the two movements.

4. Have a look at the picture and try to solve the problem.

Problem: **Find the cycle of minimum cost visiting all of the vertices of *G* exactly once.**