Plan:
- Introduction to TSP
- 2-approximation of TSP
- 1.5-approximation of TSP

Introduction
(We will assume all graphs are undirected)
A few terminologies, given a graph:
- Hamiltonian path: visits each vertex in $G$ exactly once (can miss edges)
- Eulerian path: visits each edge in $G$ exactly once (can visit a node many times)
- Hamiltonian/Eulerian cycle: same definition as Ham/Euler path but start node = end node

Complexity
- Given graph $G$, finding a Hamilton cycle/path are NP-complete
- Eulerian cycle exists iff every node has even degree.
- Eulerian path exists iff all but 2 nodes have even degrees. And they are the start/end points.
Travelling Salesman Problem

On an undirected, weighted graph $G$, find a Hamiltonian cycle of minimum total weight (sort of a generalization of Hamiltonian cycle problem on weighted graphs).

Just like in Hamiltonian cycle, TSP is \textit{NP-complete}.

We will work with a special variant of TSP called metric-TSP, where edge weights satisfy the triangle inequality: $W(A, B) \leq W(A, C) + W(C, B)$.

Note: If $G$ is connected, this restriction forces $G$ to be complete, since if $A, B$ are not connected by an edge, $W(A, B) = \infty$, and the triangle inequality will not be satisfied.

We will just work with complete graphs.

Metric-TSP is also \textit{NP-complete}.

\textbf{2-approximation for metric-TSP}

\underline{Step 1:} find the MST $T$ of $G$. 
Claim: $W(T) < W(OPT)$, if OPT is the optimal solution of TSP.

Proof: OPT is a cycle going through all nodes. Removing any edge $e$ from OPT yields a spanning tree $S$, whose weight is no less than the minimum spanning tree $T$.

$W(T) \leq W(S) = W(OPT) - W(e) < W(OPT)$.

Step 2: Starting from any node in $T$, do a DFS, traversing every edge in $T$ twice.

$\pi = ACDCEFECBAC$

This gives us a path $\pi$ and $W(\pi) = 2W(T) < 2W(OPT)$.

But $\pi$ is not a Ham cycle since some nodes are visited many times.

Step 3: Convert $\pi$ into a legit Ham cycle $\pi_0$ via "short cutting".

Idea: We skip the repeated nodes in $\pi$, only write down a node when we first encounter it.
\[ \pi = ACDECFCBACA \]
\[ \pi_0 = ACDDEFBA \]

Clearly, \( \pi_0 \) is a ham cycle and \( W(\pi_0) < W(\pi) \) since whenever we skip, say \( D \to E \) instead of \( D \to C \to E \), we save distance due to triangle inequality.

Thus, \( W(\pi_0) < 2 \cdot W(\text{OPT}) \), a 2-approximation.

2-approximation for metric TSP (Christofides Alg)

**Step 1** find the MST \( T \) of \( G \). Now, instead of converting \( T \) into a tour by traversing every edge twice in a DFS, we will do something better.

**Step 2** Identify nodes of odd edge degree in \( T \), call this set \( O \).

Claim: \(|O| \) must be even.
Theorem: \( \sum_{v \in \text{nodes}} \deg(v) = 2 \times \text{(number of edges)} \)](\text{even})

If there are an odd number of nodes with odd degrees, the sum will be odd.

**Step 3**: Find the minimum weight perfect matching \( M \) of \( D \) in the original complete graph.

(Recall we can find max bipartite matching w flow, general matching can also be done in poly time.) We will show \( w(M) \leq \frac{1}{2} w(\text{OPT}) \) at the end.

**Step 4**: Look at \( T + M \), all nodes now have even degree. Thus, there exists an Eulerian tour \( \pi \) with weight:

\[ w(\pi) = w(T) + w(M) \leq \frac{3}{2} w(\text{OPT}) \]

**Step 5**: Use the same trick to convert \( T + M \) into a legit ham-cycle \( \pi_0 \).

Via DFS + Short cutting:

\[ \pi_0 = \text{ACD} \text{FEC} \text{BA} \]

\[ \pi = \text{ACD} \text{F} \text{E} \text{C} \text{BA} \]

Return \( \pi_0 \) as answer.

Then \( w(\pi_0) \leq w(\pi) \leq \frac{3}{2} w(\text{OPT}) \).
Finally, we need to show $w(M) \leq \frac{1}{2} w(COPT)$.

Suppose we restrict our graph $G$ to $O$, call this restricted graph $G^o$, and find the optimal TSP solution on $G^o$. Let this restricted solution be $R$, then:

$$w(R) \leq w(COPT)$$

Why? Since $OPT$ is a ham cycle going through all nodes, it also visits all nodes in $O$. If we visit nodes in $G^o$ in the same order as they appear in $OPT$ (call this "OPT restrict to $G^o$"), the weight would only decrease, due to short cutting, so:

$$w(R) \leq w(OPT \text{ restricted to } G^o) \leq w(COPT)$$

Since $R$ is optimal in $G^o$

due to short cutting.

Now we show $w(M) \leq \frac{1}{2} w(R)$.

Again, there must be an even # of nodes in $O$, and since $G$ is complete, a perfect matching exists.
In particular, if we list the edges in $R$ in order: $e_1, e_2, \ldots, e_{2k}$ we can see that:
both $\{ e_1, e_3, e_5, \ldots \}$ and $\{ e_2, e_4, e_6, \ldots \}$ are perfect matchings.

One of the 2 perfect matchings will have weight at most $\frac{w(R)}{2}$ but that in turn is at least as big as $w(M)$, which is the min weight perfect matching.

$w(M) \leq \frac{w(R)}{2} \leq \frac{w(\text{OPT})}{2}$