PS8

Problem 7. Solve Problem 6 in time O(n + m), where m is the number of roads.

Let's write the constraints of the problem:

- For each province, exactly one city must be the capital
- Each road must be covered by at least one capital

Assertion 1: Consider the interprovince roads between two provinces.

- For each pair of provinces connected by interprovince roads, if the number of connected components formed by these roads is more than two, then it is impossible to choose a single capital in each province such that all roads have at least one endpoint in a capital. This is because there would be at least one component that isn't connected to any capital, violating the problem's constraints.

- If there are 2 or less connected components then candidate capitals for each province are two or less
  - The connected component has only one road → there is a candidate capital per province
  - The connected component has multiple roads → the single endpoint completely determines the capital for one province, without putting a constraint on the other

Assertion 2: Consider now the intraprovince roads. Either:

1) a province does not have any intraprovince roads. In this case any city can be a capital, thus we are constrained by Assertion 1.
2) if a province has at least one intraprovince road, it can possess only a single internal connected component. If there were more than one connected component of intraprovince roads, then there would need to be more than one capital in the province to cover all intraprovince roads. Thus the connected component can have at most two candidates capitals (As every road has to be covered with only a single node).

Thus we can have at most two real candidates capital for each province.

Algorithm

The algorithm works as follows.
First, consider the graph without interprovinces roads and find the connected components for each province.

If any province violates assertion 2, return NO. Otherwise, remove the nodes that cannot be capitals.

Add the interprovince roads. If any interprovince road cannot be covered because the endpoints were removed, return NO.

Find connected components for each pair of provinces. If any pair violates assertion 1, return NO. Otherwise, remove the nodes that cannot be capitals.

Now each province will have at most 2 candidates (all if it is isolated, in that case just pick one randomly).

Now we can solve a 2-CNF-SAT problem where for every interprovince road ij (road with endpoints in different provinces), we create edges to represent $x_i \; OR \; x_j$. That is, we create an edge going from $\bar{x}_i \to x_j$ and another going from $\bar{x}_j \to x_j$.

We can also add the constraints $x_i\ XOR\ x_j$ that is $(x_i\ AND\ \ \overline{x_j})\ OR\ (\overline{x_i}\ AND\ x_j)$ for the 2 candidates capitals.

If the 2-CNF-SAT problems leads to a solution, the problem is solvable.

Correctness:

We have already established why the assertions are true. Thus we obtain only true candidate capitals before the execution of the 2-CNF-SAT. In problem 6, it was also proved why the 2-CNF-SAT formulated in this manner leads to a correct result (the only difference was that before we weren't removing candidate capitals).

Complexity:

1) *Intraprovince Roads - Finding Connected Components*: Can be done with DFS and would take ∑O(Vi +Ei), which is O(n + m) because ∑Vi=n and ∑Ei=m, where n is the total number of cities and m is the total number of intraprovince roads.
2) *Removal of Non-Candidate Capitals*: Removing nodes that cannot be capitals requires checking each node within each province's connected component. The complexity for this step is O(n)
3) *Checking if Interprovince Roads can be covered*: This can be done in O(m), where m is the number of interprovince roads.
4) *Connected Components for Province Pairs*: For each pair of provinces connected by interprovince roads, we need to find the connected components again. In the worst-case scenario, every province could be connected to every other province, requiring a check for each pair. This results in a complexity of O(k^2), where k is the number of provinces. However, since the number of connected components has a direct relation to the number of roads, this step is also bounded by O(m).
5) *2-CNF-SAT:* O(m) because it is bounded by the number of implications, which is based on the number of roads.

Thus the total complexity is O(n + m) where n is the total number of cities and m is the total number of roads.