

Connecting Calgary: Problem

<<Station description>>

What combination of roads will connect all eight communities for no more than \$42 billion?

<<Station instructions>>

Pavement-eating termites have somehow chewed up every single road in Calgary, and so you must rebuild the road network from scratch. Your problem, however, is that the termites ate most of our money, too, and so you have only **\$42 billion** at your disposal. You must figure out what combination of roads will let you **travel between any two communities** without exceeding this amount. (The cost for each potential road is listed next to it in the illustration.) You don't need to build direct routes, however—for example, to go between Somerset and Tuscany, it's perfectly acceptable to use the Somerset-Oakridge-Tuscany route.

Hints:

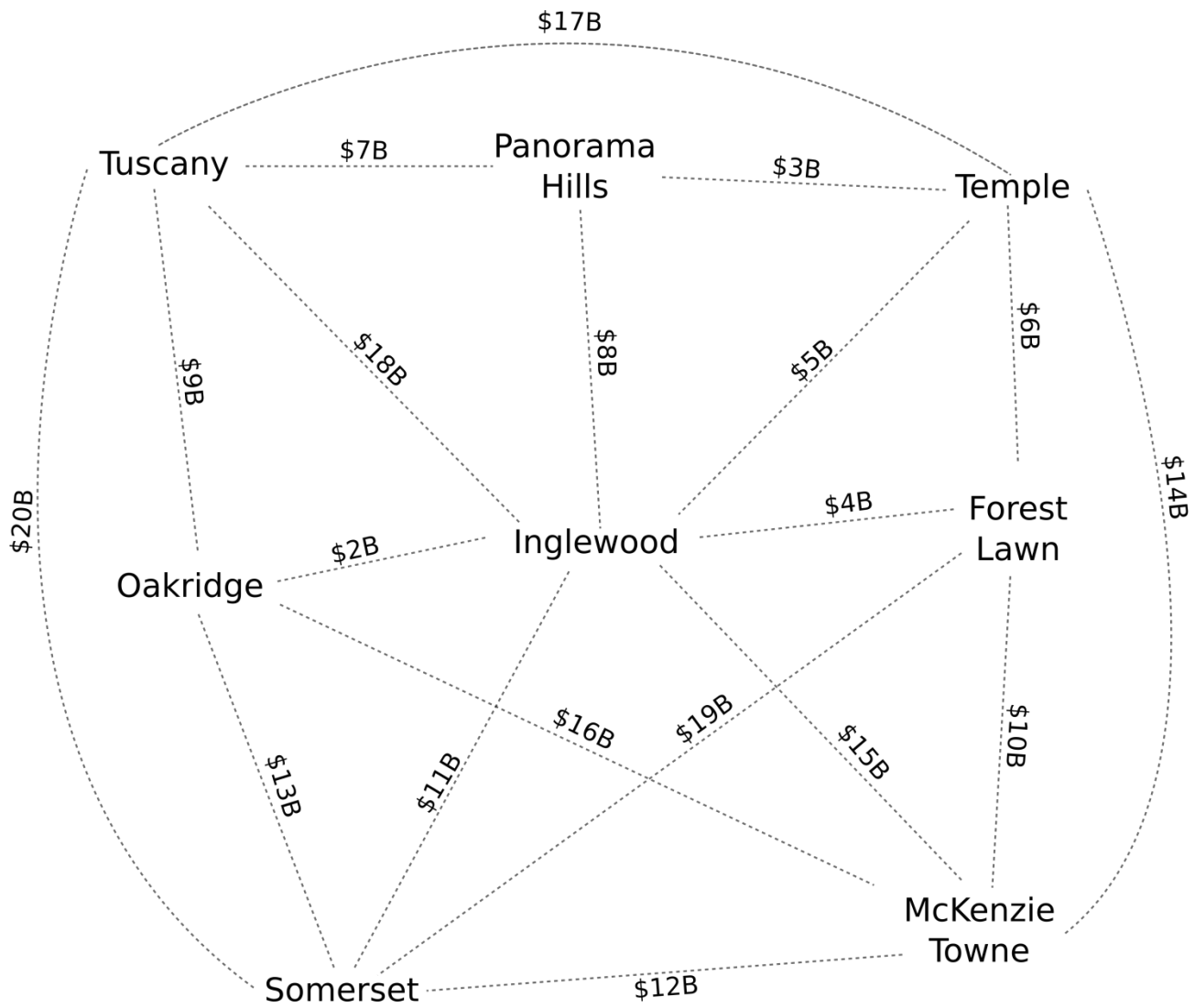
- Is it better to build a few long roads, or many short roads?
- Is redundancy a good thing to have in your road network (i.e., multiple ways to get between the same two communities)?
- What is the minimum number of roads needed to connect all eight communities?

<<Victory conditions and penalty assessment>>

This station ends once you give your answer to a SET Challenge volunteer. Penalties are assessed as follows:

- Each \$1B that your road network's cost exceeds \$42B results in a 1:00 minute penalty.
- If your road network is not fully connected—meaning that there are at least two communities that you can't travel between using any combination of roads—you will receive a 10:00 minute penalty.

What combination of roads will connect all eight communities for no more than \$42 billion?

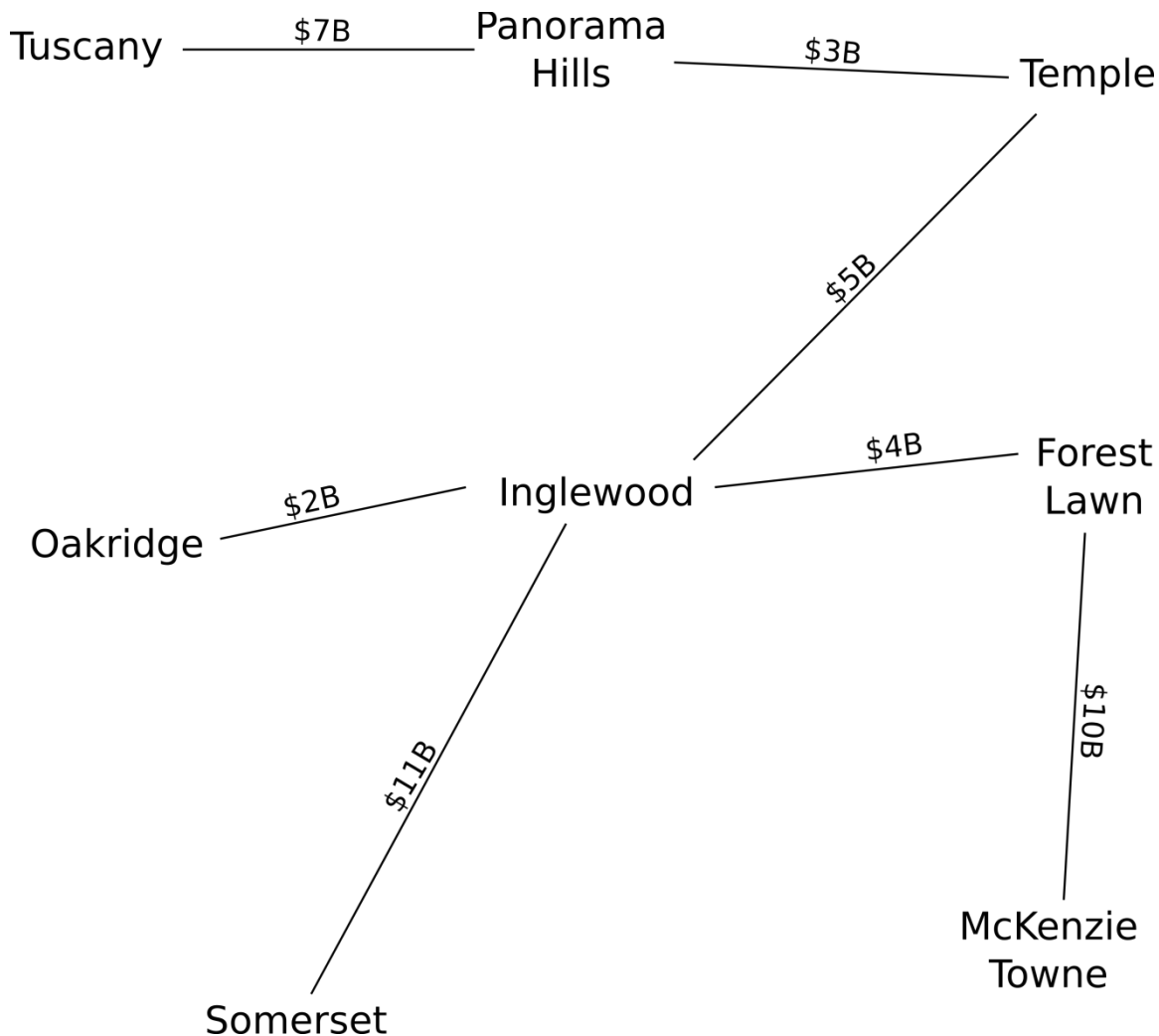


Connecting Calgary: Solution

Any correct solution will satisfy the following criteria:

- Total cost of all roads is \$42 billion or less
- If you pick any two of the eight communities, there's a way to travel between them (but not necessarily directly)
 - This means there must be **at least** $(8 - 1 = 7)$ roads

The optimal solution follows. Its total cost is \$42B. No other solution with a cost less than or equal to \$42B should be possible.



This challenge is an example of the minimum spanning tree problem ([http://en.wikipedia.org/wiki/Minimum spanning tree](http://en.wikipedia.org/wiki/Minimum_spanning_tree)), in which you must find a series of edges (“roads”) that connect all nodes (“communities”) at minimum total cost. One solution comes through Kruskal's algorithm, which Joseph Kruskal created in 1956. Kruskal's algorithm produces a minimum spanning tree by choosing the lowest-cost edge, then the next-lowest, and so forth. At each step, if choosing an edge will create a cycle—i.e., a “loop” that you can circle around repeatedly, such as Tuscany-Inglewood-Oakridge—you ignore that edge, and move on to the next-lowest. Given a total of n nodes, you thus stop choosing edges once you have $n - 1$ in total—provided you haven't created a cycle, $n - 1$ edges will connect all n nodes, and any additional edges will only add to the cost.

Kruskal's algorithm is useful as an example of a *greedy* algorithm, in which repeatedly making a locally optimal choice (choosing the lowest-cost edge that won't produce a cycle) yields a globally optimal solution (the lowest-cost tree that connects all nodes). Finding greedy algorithms to solve problems is desirable in computer science, as such algorithms tend to be simple to understand, making implementing them in a programming language straightforward, and efficient to execute, meaning the algorithm will quickly produce a solution even for large data sets (i.e., ones with a great many nodes and possible edges). The challenge with greedy algorithms such as Kruskal's is proving beyond doubt that they will always produce the best possible answer—while possible, this is considerably more difficult.

Minimum spanning trees have applications in computer networking, where they help calculate broadcast hierarchies letting you efficiently send messages to all computers on a network; in handwriting recognition, where they determine the relationship between pen strokes that compose whole characters; and in epidemiology, where they illustrate evolutionary relationships between disease strains.