group 8: categorical range queries on spatial networks

rahul saladi, xiaofei zhao and akash agrawal

(reviewed by akash agrawal and atanu roy)

problem motivation: This project is a combination of both applied research and basic research. In terms of basic research the speaker has addressed the problem of doing range query on spatial network with network distance instead of traditional euclidean. In the applied side, the speaker showed the example of locating facilities in spatial network within a given query network distance which is a practical and very useful problem in scenarios like finding hotels, restaurants, movie theatres, hospitals or any other facilities when a person is visiting new place/city.

problem statement: This project addresses following research problem:
“Preprocess the given a spatial/road network with a set of facilities, which lie along the edges of the network and can be categorized into disjoint groups $F_1, F_2, \ldots, F_l$, such a way that given a query in the form of a location $q$, network distance $d_q$ and facility group $F_i$, report all the facilities of $F_i$ which are within $d_q$ distance from $q$.”

The listed constraints that facilities lie on the edge on network is reasonable because if some facility do not lie on any edge, we can always map it to nearest edge of network because to reach to that facility we have to follow some edge of network otherwise that facility is unreachable and makes no significance in reporting. The input and output was well defined and explained with the help of an example.

challenges: The challenge in this problem is the computational cost involved. Although the challenges are well articulate and explained in this presentation, there is a scope for improvement. It would be better if the speaker talks about the basic reason for the computational cost for example the range query based on network distance is a difficult problem because the basic problem with this kind of query is establishing mapping between network distance and euclidean distance, which highly depends on underling network, and hence cannot be unified for general networks.

proposed approach: The talk explains the key elements like neighbourhood of a node and facility, list structure and facility neighbourhood graph very well with the help of a first principle example. The proposed approach honours all the constraints listed in problem description except that solution is presented for only one facility group, but also the speaker mentions that the same solution can be extended for multiple facility groups by making one list structure and facility neighbourhood graph for each facility group. The proposed approach is as follows:

preprocess step:

1. For each non-facility node find the neighbouring facility and store in sorted list with the node.
2. For each facility, find neighbouring facility and create facility neighbourhood graph.

query step:

1. For query location $q$ find nearest node.
2. Initialize the priority queue for Dijkstra algorithm.
3. For each nearest node, search the list structure and populate the queue with facilities within query distance.
4. Run Dijkstra with start point as this queue and stop when a distance travelled is more than $d_q$.

**Novel/Better**: This presentation do not talk about the previous work and the classification of related work. Although the novelty of this work is clear from the motivation, problem statement and approach presented, but comparison with related work is missing. Traditionally a lot of work has been done on range query, but all these works consider euclidean distances also many problems related to some kind of query with network distance such as nearest neighbour, $k$-nearest neighbour etc has been solved. It would be helpful in bringing out the novelty if the speaker show the classification of either problem or approach.

**Validation**: The speaker has shown asymptotic analysis for this algorithm and has explained the different situations which can lead to best case and worst case performance.

**Presentation Critique**: The talk was very well organized and easily understandable. Although it covers and explains the key idea very well with the help of real life examples, but a running example would have helped more to understand the approach. Not many questions were asked during this presentation. We would rate this talk 8 on scale of 10 due to missing related work and running example. For each of the 6 elements the score is as follows:

- **Motivation**: 10.
- **Problem statement**: 10.
- **Challenges**: 9, lack of explanation for basic challenge.
- **Proposed Approach**: 9, works good for average case but can have huge runtime and memory footprint in worst case.
- **Novel/Better**: 5, Although this is problem is not addressed before, but related problems have been solved. Speaker should mention that.
- **Validation**: 10.