The Islands Approach to Nearest Neighbor Querying in Spatial Networks

Xuegang Huang, Christian S. Jensen, and Simonas Šaltenis

Department of Computer Science, Aalborg University, Fredrik Bajers Vej 7E, DK-9220, Aalborg, Denmark
{xghuang, csj, simas}@cs.aau.dk

Abstract. Much research has recently been devoted to the data management foundations of location-based mobile services. In one important scenario, the service users are constrained to a transportation network. As a result, query processing in spatial road networks is of interest. We propose a versatile approach to \( k \) nearest neighbor computation in spatial networks, termed the Islands approach. By offering flexible yet simple means of balancing re-computation and pre-computation, this approach is able to manage the trade-off between query and update performance. The result is a single, efficient, and versatile approach to \( k \) nearest neighbor computation that obviates the need for using several \( k \) nearest neighbor approaches for supporting a single service scenario. The experimental comparison with the existing techniques uses real-world road network data and considers both I/O and CPU performance, for both queries and updates.

1 Introduction

An infrastructure is emerging that enables location-based mobile services, and we are witnessing substantial efforts in the research community to establish fundamental data management support for such services. Mobile services typically involve service users and so-called points of interest. We consider the scenario where these are located within a spatial network or, more specifically, a road network \[12,13,14,17,21,29\]. The movements of the users, often termed moving objects, are constrained by the network, and the points of interest can only be visited by traveling along the network. The relevant notion of distance is network distance based on shortest-path computation.

Existing approaches to \( k \) nearest neighbor (\( k \)NN) computation in spatial networks can be divided into two types: approaches that compute \( k \)NN queries by incrementally scanning the network until \( k \) neighbors are found, and approaches that apply some form of pre-computation and “compute” \( k \)NN queries by looking up data collected in pre-computed data structure. Both types of approaches assume that the spatial network is represented by graph-like data structures.

The first type of approach, denoted as “online computation,” naturally captures the dynamic aspects of the network, e.g., the emergence or disappearance of points of interest, and applies some form of network expansion-based search. This type of approach is able to output the network distances and paths to each \( k \)NN, as these are computed as part of the process. The data structures used in online computation capture the connectivity of the network and are easily updated. When compared to online computation,
the second type of approach, termed “pre-computation,” typically has better query performance, but has difficulty in coping with frequent updates of the road network and the points of interest.

We consider the performance of queries as well as updates, as both efficient querying and update are important for location-based mobile services. In particular, we propose a novel approach, termed the Islands approach, to \(k\)-NN processing in spatial networks. This approach computes the \(k\)-NNs along with the distance to each, but does not compute the corresponding shortest paths. The rationale for this design decision is that a mobile user is expected to only be interested in the actual path to one nearest neighbor selected from the \(k\)-NN result, and so the path computation is better left to a subsequent processing step.

The Islands approach is designed with the assumption that the overall I/O cost of queries and updates is the main performance evaluation criterion, and the approach aims to be efficient for varying frequencies of queries and updates, which yields broad applicability. The versatility of the approach is demonstrated by an experimental comparison with two other approaches that covers the cases these two are optimized for.

The paper makes three main contributions.

– The Islands approach offers an attractive generalization of the existing \(k\)-NN query processing techniques for spatial networks. It employs a relatively simple data structure and an intuitive search algorithm. And it is applicable to a broad range of mobile service scenarios, thus avoiding the need for using more specialized algorithms for different scenarios.
– The Islands approach offers a direct and elegant way of controlling the amount of pre-computation performed and thus also the trade-off between query and update performance. This enables the approach to accommodate varying densities of points of interest and varying query versus update frequencies.
– The paper presents an experimental evaluation that is significantly more comprehensive than previous evaluations. Specifically, this is the first evaluation that covers both online computation and pre-computation and that considers both query and update performance in a setting with real road network data. The paper thus offers new insight into relative merits of the existing approaches.

In Section 2, we proceed to introduce related work. Section 3 presents the Islands approach and its variations. This is followed by a section that compares the Islands approach with existing \(k\)-NN techniques. Section 5 then presents the empirical performance study that characterizes the Islands approach as well as compares it with the existing algorithms. The last section summarizes and offers directions for future research.

2 Related Work

Nearest neighbor computation is a classical topic. Many existing algorithms assume an indexing structure, e.g., an R-tree, and search in a branch-and-bound manner [10,18]. Many extensions and applications of \(k\)-NN computation have also been proposed [1,5,11,15,20,23,24,27,28].