**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**

***IJCSMC, Vol. 5, Issue. 1, January 2016, pg.102 – 106***

# A Survey on K- Routes Approach to Network Summarization

**Chanchla Tripathi[1], Ashish S. Sambare[2], Namrata S. Mahakalkar[3]**

[1]Research Scholar, Department of CSE, Priyadarshini Institute of Engineering & Technology, Nagpur University, Nagpur, Maharashtra, India

[2]Associate Professor, Department of CT, Priyadarshini Institute of Engineering & Technology, Nagpur University, Nagpur, Maharashtra, India

[3]Assistant Professor, Department of CSE, Priyadarshini Institute of Engineering & Technology Nagpur University, Nagpur, Maharashtra, India

[1] chanchalatripathi@gmail.com; [2] ashishsambare@yahoo.co.in; [3] namratamahakalkar@gmail.com

---

***Abstract— Data summarization is an important concept in the field of data mining for finding a compressed representation of a dataset. In any spatial network activity summarization(SNAS) , we are given a spatial network and a group of activities (e.g. , perambulator fatality reports, crime reports ) and the goal is to find the K-shortest route that summarize the activities. SNAS is essential for application where observation take place along linear paths such as roadways, train tracks, etc. SNAS is computationally challenging because of the large amount of K-subsets of shortest path in a spatial network. Nearest neighbour (NN) query is one of the most significant operations in spatial databases and their application field domains, for example location-based services and advanced traveller information systems. This addresses the problem of finding the in-route nearest neighbour (IRNN) for a query entities tuple which consists of a given path with a destination and a present location o it. The IRNN is a facility instance via which the alternative route from the original route on the way to the destination is smallest.***

***Keywords— Spatial network, Nearest neighborhood query, hot spots, hot routes, activity summarization.***

---

 

## I. INTRODUCTION

Spatial network activity summarization (SNAS) is important in numerous application domains such as disaster response, crime analysis, etc. In disaster response-related application, action is taken instantly after a disastrous incident with the aim of saving life, defending property, and dealing with immediate disruption, spoil or other effects caused by the disaster [1]. For example, transportation planners and engineers may need to recognize road segments/stretches that pose risks for pedestrians and have need of redesign[2]; crime analysts may look for concentrations of crimes along certain streets to guide law enforcement[3]; and hydrologists may try to recapitulate environmental change on water resources to understand the performance of river network and lakes [2]. Disaster response played an important responsibility in the earthquake, where there were many requests for support for example food, water and medical supplies [4]. Emergency managers need the means to review these requests so that they can better understand how to distribute relief supplies. Spatial network activity summarization(SNAS) has important application in domains where observation occur along linear paths in the network. Spatial network databases are the kernel of many vital application, including transportation planning; air traffic control; water, electric, and gas utilities; telephone network; urban management; sewer maintenance and irrigation canal management. The phenomena of interest for these application are structured as spatial network, which consist of a finite group of the points (i.e., nodes), the line-segments (i.e., edges) connecting the points, the location of the points, and the attributes of the points and line-segments. For example, a spatial network database for transportation application may store road connection points and the road segment connecting [6] the intersections .A very important query in spatial database systems and geographic information systems is the nearest neighbor(NN) search. In the nearest neighbor literature, the *Minkowski* metrics, e.g., Euclidean distance and graph path length, e.g., road distance are common distance metrics. Query entity in the literature can be of two types, namely, a point and line segments. A variant to the point-NN query is a closest pair query between two point datasets.

## II. A FRAMEWORK FOR DATA SUMMARIZATION

A Data summarization is an essential concept in data mining that entails techniques for finding a solid description or representation Summarization Framework for Various Data Genres of a dataset. The process naturally involves defining a set of groups, finding a representative for each group, and reporting a statistic for each group (e.g., sum, mean, standard deviation). These notion be different depending on the genre of the data being summarized. Table 1 presents a summarization framework for three genres of data. An example of the first, relational table summarization, is the GROUP BY clause in SQL that is used to group rows having general values to report SQL aggregation functions for example mean and standard deviation. The second genre is spatial Euclidean summarization, which includes heat maps and hotspot analysis. Heat maps gives a graphical representation of data in which individual values contained in a matrix are represented as colors. Hotspots are a special type of partitioned pattern where object in hotspot regions have high similarity in comparison to one another and are dissimilar to all the entity outside the hotspot [9]. And the third genre, spatial network summarization, which defines group based on partitioning a network and may represent groups using nodes, paths, trees, etc.

**TABLE 1**

| **Data Genre** | **Group Definition** | **Group Representation Choices** |
|---|---|---|
| Relational Table (a set of rows) | A partition of rows | Distinct values of attributes (e.g., age-group) |
| Spatial (Euclidean space) | A partition of space | Point, polygons, ellipses, line-strings |
| Spatial Network ( Neighbor relationship) | A partition of a graph | Node, path, tree, subgraph |

 

## III. LITERATURE REVIEW

Summarizing activity by grouping is an important research area in data mining. Previous techniques have generally been geometry-based[11]-[14] or network based [16]-[18]. In geometry-based summarization, partitioning of spatial data is based on clustering similar points distributed in planar space where distance is calculated by using Euclidean distance, not network distance. Such techniques focus on the discovery of the geometry (e.g., circle, ellipse) of high density areas [11] and include k-Mean[8] , k-Medoid[9], p-median[15], and Nearest Neighbor Hierarchical Clustering[13]. These methods do not consider the underlying spatial network; they group spatial entity that are close in terms of Euclidean distance but not close in terms of network distance.

In network-based summarization, spatial objects are clustered using network(e.g., road) distance. Existing method of network-based techniques for example Mean Streets[14], Maximal Subgraph Finding (MSGF), and Clumping group activity over multiple paths, a single path/subgraph , or no path at all.

*3.1 SPATIAL NETWORK ACTIVITY SUMMARIZATION*

This section describes the computational structure of SNAS. It also describes the K-Main Route(KMR) and its performance decisions Network Voronoi activity Assignment, Divide and conquer Summary Path REcomputation.

## IV. PROPOSED WORK

*4.1 ARCHITECTURE*

In the proposed system there are four modules which are as follows:
1. Create Database
2. Continuous Query Engine
3. Apply Algorithm for the Query
4. Result Analysis

Fig 1 shows the architecture of the proposed system.

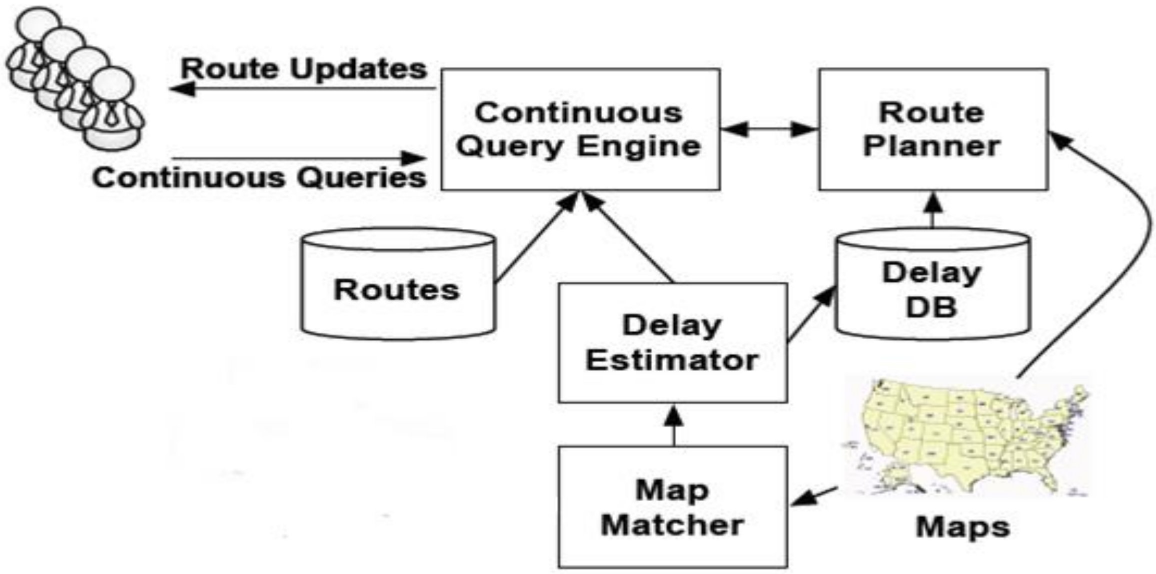


**Fig. 1. The architecture of system**

### 4.2 MODULES

*4.2.1 Create Database:*

Canal Network is digitized and delineated using the Survey of India(SOI) topo map of scale 1:25,000. Block and Chak boundaries were delineated from the features resulted from surface modeling tools, topo map and digitized canal network. The methodology is represented in fig 2 . The following layers are generated in GIS Platform:
1. Canal line;
2. Canal Node;
3. Contour and Digital Elevation Model;
4. Command Area Boundary including Block and Chak; boundary;

*4.2.2 Collection of Input*

The data includes:

a) SOI topo maps are also collected from the SOI for reference;

b) Existing Block map and other collateral data concerned command area;

c) Contour maps.

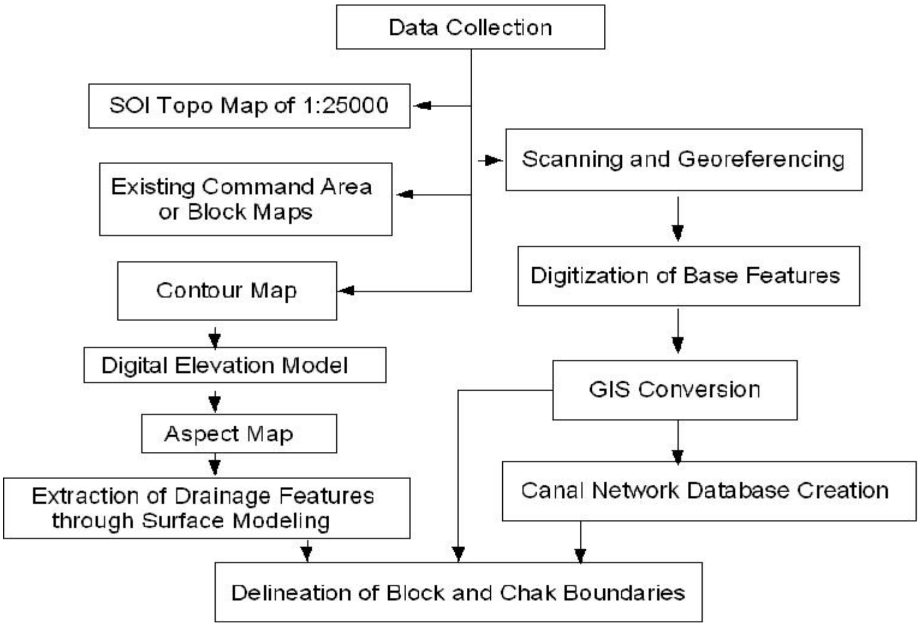


**Fig. 2: Methodology**

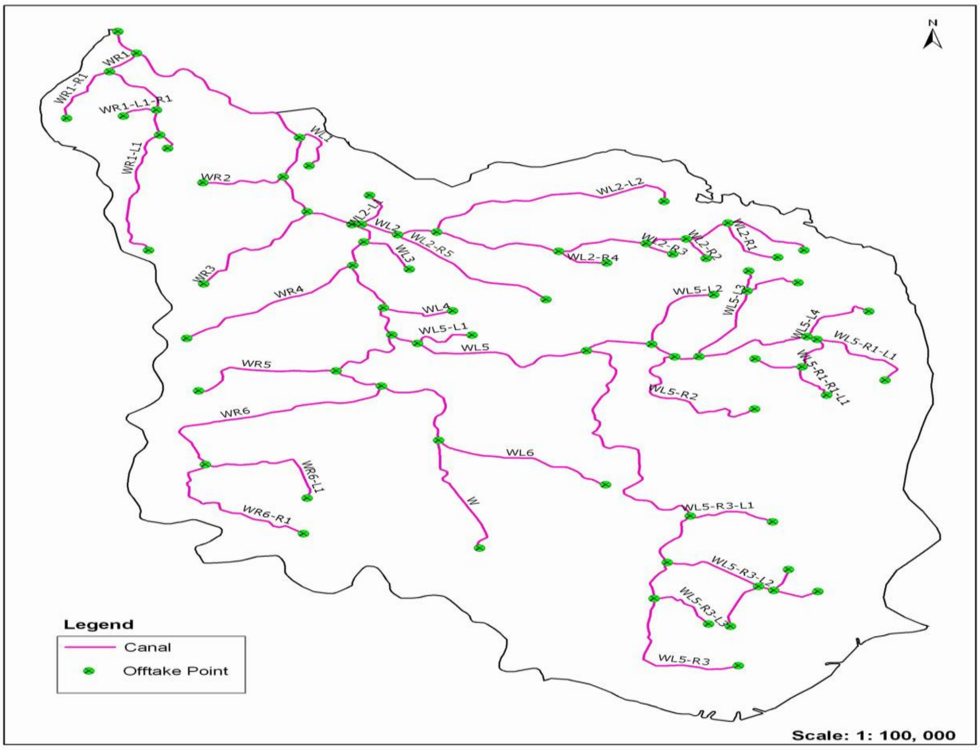


**Fig.3: Canal network representing line and nodes**

*4.2.3 Search Engine*

In this system we present a new method that allows to use a generic and generally available full text engine back end (MySQL and SphinxSearch) to generate recommendations based on a K-nearest neighborhood approach.

## V. CONCLUSION

This work explored the problem of spatial network activity summarization in relation to vital application domains for example preventing pedestrian fatalities and crime analysis.

In future work, we plan to explore another types of data that may not be associated with a point in a street (e.g., aggregated pedestrian fatality data at the zip code level). We plan to inspect a distance-based rather than coverage-based objective function. We will also generalize SNAS for all paths and explore spatial constraints (e.g., nearest neighbors).

## REFERENCES

[1] A. Barakbah and Y. Kiyoki, "A pillar algorithm for k-means optimization by distance maximization for initial centroid designation," *in Proc. IEEE Symp. CIDM, Nashville*, TN, USA, 2009.

[2] S.Borah and M. Ghose, "Perfomance analysis of AIM-k-means & K-means in quality cluster generation, *"J.Computer, Dec.2009*

[3] D. Oliver, A. Bannur , J. M. Kang, S. Shekhar , and R. Bousselaire, "AK-main routes approach to spatial network activity summarization: A summary of results," in *Proc. IEEE ICDMW , Sydney* , NSW, Australia , 2010

[4] S.Shekhar and D. Liu, "CCAM: A connectivity-clustered access method for networks and network computations," *IEEE Trans. Knowl. Data Eng., v*ol. 9 , no. 1, pp. 102-119, Jan/feb.2008

[5] W.C. Collier and R.J. Weiland. Smart Cars, Smart Highways, *IEEE Spectrum*, 1994

[6] S.Shekhar and J. S. Yoo , Processing In-Route Nearest Neighbor Queries: A Comparision of Alternative Approaches, *ACM GIS*, 2003.

[7] G. Hjaltason, H. Samet, Incemental Distance Join Algorithms for spatial Databases*, Proc. ACM Conference on Management of Data (SIGMOD), 1998*

[8] G.Hjaltason, H. Samet , Distance Browsing in spatial Databases, *Proc. ACM Transactions on Database System (TODS*), 2007.

[9] J. H. Rillings and R. J. Betsold.

Advanced Driver Information systems. *IEEE Trans.on vehicular Technology*,2001

[10] J.Zhang, N. Mamoulis, D. Papadias, Y.Tao, All-Nearest-Neighbours Queries in Spatial Databases, *Proc. IEEE Conf. on Scientific and Statistical Database Management (SSDBM),* 2011

[11] A. CORRAL, Y. Manolopoulos, Y.Theodoridis, M. Vassilakopoulos, Closes Pair Queries in Spatial Databases, *Proc. ACM Conference on Management of Data (SIGMOD), 2010.*

[12] C. Shahabi, M. R. Kolahdouzan, M Sharifzadeh, A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases, *Proc. International Symposium on Spatial and Temporal Databases (SSTD), 2011.*

[13] D. Papadias, J . Zhang, N. Mamoulis, Y.Tao, Query Processing in Spatial Network Databases, *Proc. Very Large Data Bases Conference (VLDB), 2013*.

[14] Dev Oliver, Shashi Shekhar, James M. Kang, Renee Laubscher, Veronica Carlan and Abdussalam Bannur, " A K-Main Routes Approach to Spatial Network Activity Summarization. ", *IEEE transactions on knowledge and data engineering,* june 2014.

[15] S. Shiode and N. Shiode, "Detection of multi-scale clusters in network space, "*International Journal of Geographical Information Science*, 2009

[16] M. Ernst, M.Lang, and S. Davis. (2011). Dangerous by design: Solving the epidemic of preventable pedestrian deaths. Transportation for America: *Surface Transportation Policy Partnership. Washington, DC.*

[17] J. Eck, S. Chainev , J. Cameron , M. Leitner , and R.Wilson. (2005,Aug.). Mapping Crime: *Understanding Hot Spoys, U.S. Department of Justice. Washington*, DC , USA [Online].

[ 18] D.Matthews, S.Effler, C. Driscoll, S. O'Donnell, and C. Matthews, "Electron budgets for the hypolimnion of a recovering urban lake, 1989-2004: Response to change in organic carbon deposition and availability of electron acceptors," *Limnol. Oceanogr., 2008.*

[19] S. Shekhar, M. Evanas, J.Kang, and P.Mohan , " Identifying patterns in spatial information: A Survey of methods," *WIREs Data Mining Knowl. Discov. , Apr. 2011*.

[20] X. Li, J. Han, J. Lee, and H. Gonzalez, "Traffic density-based discovery of hot routes in road networks," *in Proc. 10th SSTD, Berlin, Germany, 2007*

 