# ANALYZING MOVING OBJECTS DATA IN SPATIO-TEMPORAL SPACE AND SOCIAL NETWORKS

**SONIA KHETARPAUL (NEE PRUTHI)**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

October 2017

# ANALYZING MOVING OBJECTS DATA IN SPATIO-TEMPORAL SPACE AND SOCIAL NETWORKS

by

SONIA KHETARPAUL (NEE PRUTHI)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

**Indian Institute of Technology Delhi**

**October 2017**

# Certificate

This is to certify that the thesis titled **Analyzing Moving Objects Data in Spatio-Temporal Space and Social Networks** submitted by **Ms. Sonia Khetarpaul** to the Indian Institute of Technology, Delhi, for the award of degree of **Doctor of Philosophy**, is a record of bona-fide research work carried out by her under our guidance and supervision at the Department of Computer Science and Engineering. The thesis, in our opinion, is worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute. To the best of our knowledge, the work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.


Dr. S. K. Gupta
Professor
Department of Computer Science & Engineering
Indian Institute of Technology Delhi, India

Dr. L. Venkata Subramaniam
STSM & Senior Manager
IBM Research, India

# Acknowledgments

The experience at IIT Delhi for this degree has been an enriching experience. I am indebted to several people who contributed directly or indirectly towards the completion of this thesis. I take this opportunity to thank them. First of all, I wish to express my deepest gratitude to my supervisors Prof. S. K. Gupta and Dr. L. Venkata Subramaniam for their enriching guidance and constant encouragement during the development and completion of this dissertation. I will always be thankful for their continuous support and their motivation for encouraging me to explore my research interests, with understanding and patience. I thank them for the revisions and suggestions and for setting the highest of standards for any kind of submission. It has been a privilege to work under their guidance.

I would also like to thank my research committee members Prof. Saroj Kaushik, Prof. M. P. Gupta, Prof. Prem Kalra and Prof. Sorav Bansal for their valuable guidance and support. I would like to thank Prof. K. K. Biswas for his encouragement and help. I would also like thank Prof. Naveen Garg whose constant presence in the lab has been a source of continuous encouragement for me to work hard and think deeply.

I am grateful to my friends and colleagues Aditi Kapoor, Parul Shukla and Nidhi Gupta who enriched my life during the PhD program with their discussions, encouragement and support. Our discussions along with numerous cups of tea throughout these years provided great companionship and helped in maintaining the required enthusiasm. I would further like to thank Mona Jain, Nisha Jain, Kanika Modi, Chinmay Narayan and Sunita Tiwari for their companionship. I would also like to thank the different M.Tech and summer intern students with whom I have worked during my PhD. I would like to extend my thanks to staff of Database and AI Lab, Mr. M. Rathinam and Mr. S. S. Negi for their help in all lab related issues.

I would also like to thank my family. I thank my parents-in law and brother-in law who always given me the freedom and encouragement at every point of my life with their unconditional love and support. I would especially like to express my gratitude towards my late father-in law who has always been a source of inspiration to me and who encouraged me even when he was very unwell himself. I would also like to thank my parents and brothers who made me capable enough to allow me start this journey at IIT and for their continuous support, motivation and encouragement. I would like to thank my husband Shaleen Khetarpaul for enriching my life and being a constant source of motivation and happiness and helping me in everyway possible throughout this time. I would like to dedicate this thesis to my both sons Armaan and Vivaan for their joyful presence and immense patience.

Finally, I am grateful to God for his blessings and for giving me the strength to persevere throughout this long and arduous Ph.D journey.

**Sonia Khetarpaul**

# Abstract

The advances in wireless communication and positioning devices like Global Positioning Systems (GPS) has generated significant interest in the field of analyzing and mining various patterns present in spatio-temporal data. The pervasiveness of location-acquisition technologies (GPS, GSM networks, etc.) have enabled convenient logging of location and movement histories of individuals. Also a location based social network (LBSN) not only represents social relations among actors, but also allows them to share and access their physical locations. Here, the physical location consists of instant location of an individual at a given timestamp and the location history that an individual has accumulated in a certain period. The increasing availability of large amounts of spatio-temporal data pertaining to the movement of users has given rise to a variety of application. Managing, understanding, querying and mining the collected location data are important issues for these applications.

This thesis addresses various issues related to understanding, querying and mining spatio-temporal data of moving objects and inter-connected moving objects (through social networks). We start with defining a Spatio-Temporal-Social (STS) data model and algebra which captures both non spatial and spatial properties of moving users along with their social inter-connections. It is a high level, programming language independent data model with three levels of data abstractions. At first level, raw observations are defined. Then these observations are mapped to data types and operations to capture spatio-temporal data and social network inter-connections and define the second level of the model. The third level is formalized and structured level, represents user's view of data. It shows the data as connected objects with spatial and non-spatial properties associated with them. By applying this data model and related algebra, we mine individual's location history to determine interesting locations, optimal meeting points, etc., and query social network users for their spatio-temporal locations, social inter-connections, influences, their common check-ins, etc.

Then, we address the problem of identifying interesting and temporal interesting locations in a city. GPS Trajectory datasets of various users are aggregated and analyzed by applying STS algebra and interesting locations are determined. Our next work focuses on scheduling the meeting for varying number of users who are situated in geographically distant locations of a city and have varying traveling patterns. We extend the work of previous authors in this domain and determine optimal meeting point by incorporating some real life constraints like varying travel patterns, flexible meeting point and considering road network distance.

Finally, we determine the impact of structural patterns hidden in the nodes of social networks and external environment changes on the spatio-temporal mobility patterns of the users. We explore how network and spatial properties and external factors affect the number of check-ins and influences made by users.

# सार


बेतार संचार और ग्लोबल पोजिशनिंग सिस्टम (जीपीएस) जैसे पोजिशनिंग डिवाइसेस में अग्रिमों ने स्पेसिओ-अस्थायी डेटा में मौजूद विभिन्न पैटर्नों के विश्लेषण और खनन के क्षेत्र में महत्वपूर्ण रुचि पैदा की है। स्थान-अधिग्रहण तकनीकों (जीपीएस, जीएसएम नेटवर्क आदि) की व्यापकता ने व्यक्तियों के स्थान और आंदोलन इतिहास के सुविधाजनक प्रवेश को सक्षम किया है। इसके अलावा स्थान आधारित सोशल नेटवर्क (एलबीएसएन) न केवल अभिनेताओं के बीच सामाजिक संबंधों का प्रतिनिधित्व करता है, बल्कि उन्हें अपने भौतिक स्थानों को साझा करने और उन तक पहुंचने की अनुमति भी देता है। यहां, भौतिक स्थान में किसी व्यक्ति के तत्काल स्थान दिया गया है जो किसी दिए गए टाइमस्टैम्प और उस स्थान के इतिहास में होता है जिसे किसी व्यक्ति ने एक निश्चित अवधि में जमा किया हो। उपयोगकर्ताओं के आंदोलन से संबंधित स्थैर-अस्थायी डेटा की बड़ी मात्रा में बढ़ती उपलब्धता ने विभिन्न प्रकार के आवेदनों को जन्म दिया है। एकत्रित स्थान डेटा प्रबंधित करना, समझना, पूछताछ करना और खनन करना इन अनुप्रयोगों के लिए महत्वपूर्ण समस्याएं हैं।

यह थीसिस वस्तुओं, चलने वाली वस्तुओं और अंतर-जुड़े चलती वस्तुओं (सोशल नेटवर्कों के  माध्यम से) के स्पष्टीकरण-अस्थायी डेटा को समझने, पूछताछ और खनन से संबंधित विभिन्न मुद्दों को संबोधित करता है। हम एक स्पैतिओ-टेम्परल-सोशल (एसटीएस) डेटा मॉडल और बीजगणित को परिभाषित करने से शुरू करते हैं जो उपयोगकर्ताओं को अपने सामाजिक अंतर-कनेक्शन के साथ घूमने के गैर-स्थानिक और स्थानिक गुणों को कैप्चर करते हैं। यह डेटा स्तरों के तीन स्तरों के साथ एक उच्च स्तर, प्रोग्रामिंग भाषा स्वतंत्र डेटा मॉडल है पहले स्तर पर, कच्चे अवलोकनों को परिभाषित किया जाता है फिर इन अवलोकनों को डेटा प्रकार और संचालन के लिए मैप किया जाता है ताकि स्पाटियो-अस्थायी डेटा और सोशल नेटवर्क इंटर-कनेक्शन प्राप्त हो सके और मॉडल के दूसरे स्तर को परिभाषित किया जा सके। तीसरा स्तर औपचारिक और संरचित स्तर है, डेटा के उपयोगकर्ता के दृश्य का प्रतिनिधित्व करता है। यह आंकड़ों को उनके साथ जुड़े स्थानिक और  गैर-स्थानिक संपत्तियों के साथ जुड़ी वस्तुओं के रूप में दिखाता है। इस डेटा मॉडल और संबंधित बीजगणित को लागू करने से, हम दिलचस्प स्थानों, इष्टतम मीटिंग प्वाइंट, आदि को निर्धारित करने के लिए व्यक्तिगत स्थान के इतिहास का उपयोग करते हैं, और उनके सामाजिक नियमन स्थानों, सामाजिक अंतर-कनेक्शन, उनके आम चेक-इन, रुचियों के लिए सोशल नेटवर्क उपयोगकर्ताओं से पूछते हैं , गतिविधियों, आदि

फिर, हम शहर में दिलचस्प और अस्थायी दिलचस्प स्थानों की पहचान करने की समस्या को संबोधित करते हैं। एसटीएस बीजगणित और दिलचस्प स्थानों को लागू करने से विभिन्न उपयोगकर्ताओं के जीपीएस ट्रैजिगोरी डेटासेट को एकत्रित और विश्लेषण किया जाता है। हमारा अगला काम उन उपयोगकर्ताओं की संख्या को अलग करने के लिए बैठक का निर्धारण करने पर केंद्रित है, जो शहर के भौगोलिक दृष्टि से दूर स्थानों में स्थित हैं और अलग-अलग यात्रा पैटर्न हैं। हम इस डोमेन के पिछले लेखकों के काम को विस्तारित करते हैं और कुछ वास्तविक जीवन बाधाओं को शामिल करते हुए यात्रा के पैटर्न, फ्ल एक्वाययम मीटिंग प्वाइंट और सड़क नेटवर्क दूरी पर विचार करके इष्टतम मीटिंग बिंदु निर्धारित करते हैं।

अंत में, हम  सोशल नेटवर्क के नोड्स में छिपे हुए संरचनात्मक पैटर्न के प्रभाव और उपयोगकर्ताओं के स्पेसिओ-अस्थायी गतिशीलता के पैटर्न पर बाह्य वातावरण परिवर्तन को निर्धारित करते हैं। हम पता लगाते हैं कि नेटवर्क और स्थानिक संपत्तियों और बाह्य कारकों का उपयोग एफआईपी उपयोगकर्ताओं द्वारा किए गए जांचकर्ताओं की संख्या और उनके द्वारा किए गए कार्यों के अनुसार करता है।

# Contents

# List of Figures

# List of Tables