Chapter 1

Introduction

With the maturity of database technologies, nowadays applications collect data in all domains at an unprecedented scale. For example, billions of social network users and their activities are collected in the form of *graphs*; Thousands of sensor reports are collected in the form of *streams*; Hundreds of millions of temporal-locations are collected as *trajectories*, etc. Although modern data management systems are able to collect and store such tremendous amount of data, there are lack of study on providing useful and efficient analytics for various data domains.

Traditional SQL-based analytics including ranking, aggregation, and window functions, has seen a great success in supporting data-based decision-makings on relational data. However, when applying SQL-based analytics on other data domains, it often involves expensive joins which are hard to optimize without leveraging the domain knowledge. For instance, when computing the K-hop neighborhoods of vertexes in a graph, SQL-based traversal of graph involves multi-rounds of joins, which is inefficient than search-based traversal []. Or, when searching for a group of objects that travels together for a certain time, SQL-based solution would involves recursive joins and chain joins.

See from the limitation of SQL-analytic on those domains, in this thesis, we explore

the neighborhood based analytics in various data domains. In particular, we address three issues. First, we define the neighborhoods on various domains. Second, we showcase the usefulness of neighborhood analytics on data models. Third, we address the efficiency issues on applying the neighborhood analytics in different data domains for large data.

1.1 Neighborhood Data Analytic

Neighborhood analytic is the counterpart of global analytic, which aims to providing analytics over each objects' vicinity. In SQL-analytic, neighborhood analytic is in the form of window function. Specifically, window function is launched with the keyword "over". For example XXXXX. In this example, each tuple in attached by a window based on the order on attribute YYY. Basically, the window contains the most related tuples to this tuple. Then, an aggregate function is applied on each tuples' window, resulting in a neighborhood analytic. To generalize,.....

1.2 Motivation and Challenge

1.2.1 Graph

Neighborhood is defined by different distance function

1.2.2 Data Stream

Nested neighborhoods.

1.2.3 Spatial-temporal Data

Reflexive neighborhoods.

1.3 Contribution and Impact

1.4 Organization

Chapter 2

Literature Review

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 3

Sample Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Appendix A

Sample Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.