Key Sources and Summary

1. **EXPLORATORY ANALYSIS OF SPATIAL DATA**


3. *Detecting graph-based spatial outliers: algorithms and applications (a summary of results)*, Shashi Shekhar, Chang-Tien Lu, Pusheng Zhang

4. *Research track: Mining distance-based outliers in near linear time with randomization and a simple pruning rule*, Stephen D. Bay, Mark Schwabacher

5. *Detecting region outliers in meteorological data*, Jiang Zhao, Chang-Tien Lu, Yufeng Kou

6. *Efficient algorithms for mining outliers from large data sets*, Sridhar Ramaswamy, Rajeev Rastogi, Kyuseok Shim


8. *A unified approach for mining outliers*, Edwin M. Knorr, Raymond T. Ng

9. *Distance-based outliers: algorithms and applications*, Edwin M. Knorr, Raymond T. Ng, Vladimir Tucakov

10. *LOF: identifying density-based local outliers*, Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, Jörg Sander
Summary of key readings:

A spatial outlier is a spatially referenced object whose non-spatial attribute values are significantly different from those of other spatially referenced objects in their spatial neighborhoods [2, 3, 7]. Detecting spatial outliers is useful in many applications of geographic information systems, including transportation, ecology, public safety, public health, climatology, and location based services [2]. Shekhar et al. introduced a method for detecting spatial outliers in graph data set based on the distribution property of the difference between an attribute value and the average attribute value of its neighbors [3]. Shekhar also proposed an algorithm to find all outliers in a dataset, which replace many statistical discordance tests, regardless of any knowledge about the underlying distribution of the attributes [7]. Stephen D. Bay et al. introduced a simple nested loop algorithm to detect spatial outlier, which gives linear time performance when data is in random order and a simple pruning rule is used [4]. Existing methods for finding outliers can only deal efficiently with two dimensions/attributes of a dataset. A distance-based detection method was introduced by Sridhar Ramaswamy et al., which ranks each point on the basis of its distance to its \( k \)th nearest neighbor and
declares the top n points in this ranking to outliers. A highly efficient partition-based algorithm was also introduced in this paper [6]. Edwin M. Knorr et al. proposed another distance-base outlier detection method that can be done efficiently for large datasets, and for k-dimensional datasets with large value of k [9]. Spatial outliers are most time represented as point data, but they are frequently represented in region, i.e., a group of point. Jiang Zhao et al. proposed a wavelet analysis based approach to detect region outlier [5]. Markus M. Breunig et al. showed a different approach to detecting spatial outliers; it was done by assigning to each object a degree of being an outlier, the degree, which was called the local outlier factor of an object, depends on how isolated the object is with respect to the surrounding neighborhood [10].