Topological and Scaling Analysis of Geospatial Big Data

Ding Ma
Abstract

Geographic information science and systems face challenges related to understanding the instinctive heterogeneity of geographic space, since conventional geospatial analysis is mainly founded on Euclidean geometry and Gaussian statistics. This thesis adopts a new paradigm, based on fractal geometry and Paretian statistics for geospatial analysis. The thesis relies on the third definition of fractal geometry: A set or pattern is fractal if the scaling of far more small things than large ones recurs multiple times. Therefore, the terms fractal and scaling are used interchangeably in this thesis. The new definition of fractal is well-described by Paretian statistics, which is mathematically defined as heavy-tailed distributions. The topology of geographic features is the key prerequisite that enables us to see the fractal or scaling structure of the geographic space. In this thesis, topology refers to the relationship among meaningful geographic features (such as natural streets and natural cities).

The thesis conducts topological and scaling analyses of geographic space and its involved human activities in the context of geospatial big data. The thesis utilizes the massive, volunteered, geographic information coming from LBSM platforms, which are the global OpenStreetMap database and countrywide, georeferenced tweets and check-in locations. The thesis develops geospatial big-data processing and modeling techniques, and employs complexity science methods, including heavy-tailed distribution detection and head/tail breaks, along with some complex network analysis. Head/tail breaks and the induced ht-index are a powerful tool for geospatial big-data analytics and visualization. The derived scaling hierarchies, power-law metrics, and network measures provide quantitative insights into the heterogeneity of geographic space and help us understand how it shapes human activities at city, country, and world scales.

Keywords: Third definition of fractal, scaling, topology, power law, head/tail breaks, ht-index, complex network, geospatial big data, natural cities, natural streets