Abstract

Data-driven Analysis of Fractality and Other Characteristics of Complex Networks

Marcell Nagy

2018

The purpose of this work is to explore and understand the fractality moreover the common and unique structure of complex networks from different domains. We conduct a comprehensive literature review, but also highlight the contingent deficiency of mathematical rigour of related papers, proposed new methods, and point out the possible pitfalls of the wildly used techniques.

After giving a general introduction of network theory and data science, we detail the most important concepts of the topic, we acquire an extensive understanding of fractal behavior of complex networks. We present graph related box-covering algorithm, adopted from fractal theory, by which the fractality of the networks is defined. Besides the frequently used compact box burning box-covering algorithm, we present, implement and compare other novel alternative algorithms by both their running time and efficiency.

The problem of accurate identification of fractal dimension of networks, relies on the correct detection and fitting of power law distribution in the empirical data, obtained by the box-covering algorithm. We point out, that in several articles the validation of power-law distributions are carried out non-rigorously, hence we uncover a statistical framework that utilize maximum-likelihood fitting for discerning and quantifying power law behavior in data, and later in this work we apply multiple variants of this proposed technique.

Motivated by our observations, and by the contrariety of the pure small-world and pure fractal properties, we propose an extended, mathematically more rigorous definition of fractal networks, allowing a network to be locally fractal but globally small-world, and we show several real-world and model generated examples that embrace this phenomena.

Throughout this paper, we follow two main distinct approaches, firstly we attempt to understand the fractality through well-studied mathematical network models from the literature and our newly proposed models. We investigate how the graph metrics vary as the models transit from non-fractal to fractal and vice versa. We also detail a strong law of large numbers for the maximum degree of a modified version of the Barabási–Albert model.

For the second approach, we gathered a unique dataset, containing graph metrics of numerous real-world networks, and we follow data-driven analysis techniques for a rather general purpose. We do not only focus on the fractal property of the networks, but investigate a rich set of metrics that describe every aspects of the networks. We apply both traditional statistics and machine learning methods to find out how the values of the graph metrics are distributed on real-world networks, how the metrics are correlated with each other, how the correlations vary on different network domains. Furthermore, we solve classification and regression tasks, such as identify the network domains, or the fractality by the metrics, and estimate an appointed graph measurement value using different selections of the remaining metrics.