

# Data Mining Using Fractals and Power Laws

Christos Faloutsos

Department of Computer Science, Carnegie Mellon University, USA  
`christos@cs.cmu.edu`

**Abstract.** What patterns can we find in a bursty web traffic? On the web or on the internet graph itself? How about the distributions of galaxies in the sky, or the distribution of a company's customers in geographical space? How long should we expect a nearest-neighbour search to take, when there are 100 attributes per patient or customer record? The traditional assumptions (uniformity, independence, Poisson arrivals, Gaussian distributions), often fail miserably. Should we give up trying to find patterns in such settings? Self-similarity, fractals and power laws are extremely successful in describing real datasets (coast-lines, rivers basins, stock-prices, brain-surfaces, communication-line noise, to name a few). We show some old and new successes, involving modeling of graph topologies (internet, web and social networks); modeling galaxy and video data; dimensionality reduction; and more.