

Data Mining Using Fractals and Power Laws

Christos Faloutsos

Carnegie Mellon University
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-neighbor 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.

About the Speaker

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, nine "best paper" awards, and several teaching awards. He has served as a member of the executive committee of SIGKDD; he has published over 160 refereed articles, 11 book chapters and one monograph. He holds five patents and he has given over 20 tutorials and 10 invited distinguished lectures. His research interests include data mining for streams and networks, fractals, indexing for multimedia and bio-informatics data, and database performance.