

# **Data Mining in Emerging Decentralized Distributed Environments**

**Hillol Kargupta**

**UMBC and Agnik, LLC**

**Abstract:** Distributed data mining (DDM) deals with the problem of analyzing distributed, possibly multi-party data by paying attention to the computing, communication, storage, and human factors-related issues in a distributed environment. Unlike the conventional off-the-shelf centralized data mining products, DDM systems are based on fundamentally distributed algorithms that do not necessarily require centralization of data and other resources. DDM technology is finding increasing number of applications in many domains. Examples include data driven pervasive applications for mobile and embedded devices, grid-based large scale scientific and business data analysis, security and defense related applications involving analysis of multi-party possibly privacy-sensitive data, and peer-to-peer (p2p) data stream mining in sensor and file-sharing networks. This talk will offer a perspective of this field and then focus on data analysis applications over p2p networks. It will discuss some new algorithms, their applications, and share experimental results.