

SDM'14 Tutorial

Node similarity, graph similarity and matching: Theory and Applications

Danai Koutra
Carnegie Mellon University
danai@cs.cmu.edu

Tina Eliassi-Rad
Rutgers University
tina@eliassi.org

Christos Faloutsos
Carnegie Mellon University
christos@cs.cmu.edu

Tutorial Website: <http://www.cs.cmu.edu/~dkoutra/tut/sdm14.html>

Given a graph how can we find similarities among its nodes? Given a sequence of graphs, can we spot similarities or differences between them and detect temporal anomalies? How can we match the nodes in different graphs in order to reveal similarities between them? In a microscopic level, the nodes can be similar in terms of their proximity or their structural features and roles. In a macroscopic level, two graphs are similar if the structures within them are similar, or equivalently their nodes are connected in similar ways. How can we find the similarity between two graphs when the node correspondence is known or unknown? When the correspondence is unknown, how can we efficiently match the two graphs, i.e., infer the node mapping? The objective of this tutorial is to provide a concise and intuitive overview of the most important concepts and tools, which can detect similarities between nodes of a static graph and different snapshots of dynamic graphs or distinct networks.

We review the state of the art in three related fields: (a) node similarity and role discovery, (b) graph similarity, and (c) graph matching. The emphasis of this tutorial is to give the intuition behind these powerful mathematical concepts and tools, as well as case studies that illustrate their practical use in data mining.

Outline of the Tutorial

- Part 1: Node similarity: Roles and proximity [40 minutes]
 - What are roles
 - Roles and communities
 - Roles and equivalences (from sociology)
 - Proximity: Guilt-by-association techniques (RWR, BP,SSL)
- Part 2: Graph Similarity [25 minutes]

- With node correspondence
- Without node correspondence

- Part 3: Graph matching and de-anonymization [25 minutes]
 - Unipartite graphs
 - Bipartite graphs
 - Challenges, and heuristics
- Part 4: Case Studies [30 minutes]
 - Role discovery and transfer learning
 - Brain graphs classification and clustering
 - Anomalies in time-evolving graphs
 - Re-identification

Presenters' Bios

- *Danai Koutra* is a Ph.D. student at the Computer Science Department at Carnegie Mellon University. She earned her diploma in ECE at the National Technical University of Athens. Her research interests include large-scale graph mining, graph similarity and matching, graph summarization, and anomaly detection. Danai's research has been applied mainly to social, collaboration and web networks, as well as brain connectivity graphs. She holds 7 patents on bipartite graph alignment.
- *Tina Eliassi-Rad* is an Associate Professor of Computer Science at Rutgers University. She earned her Ph.D. from the University of Wisconsin-Madison. Within data mining and machine learning, Tina's research has been applied to the World-Wide Web, text corpora, large-scale scientific simulation data, complex networks, and cyber situational awareness. She has given over 80 invited presentations. In 2010, she received an Outstanding Mentor Award from the US DOE Office of Science.

- *Christos Faloutsos* is a Professor at Carnegie Mellon University. He has received the Research Contributions Award in ICDM 2006, the Innovations award in KDD10, 18 “best paper” awards, and several teaching awards. He has given over 30 tutorials and over 10 invited distinguished lectures. His research interests include data mining for graphs and streams, fractals, and database performance.