CAREER: From Art to Science: A General Framework for Anomaly Mining

Leman Akoglu, *Department of Computer Science, Stony Brook University*

Project Summary

Anomaly mining finds applications in numerous domains, such as security, finance, astronomy, medicine, and so on. Despite this immense popularity, addressing anomaly mining tasks in the real world remains a frustrating and dissatisfying experience due to the relative simplicity with which the anomaly mining problems are currently modeled. In fact research for the most part focused on a few standard problems, often defined only on a single data source: outliers in multi-dimensional space, change points in time series, etc. Most efforts have been put into optimizing algorithms for better speed and accuracy in place of defining new problems for modern real-world applications, leading to approaches that do not directly meet their needs.

The goal of this proposal is to create a new anomaly mining framework that utilizes multiple data sources and techniques in a corroborative fashion to fundamentally reframe our understanding and ability to **define, detect, and describe real-world anomalies**. We will study how different information sources can be utilized to model new data representations and new problem formulations, as well as how evidences from different formulations and algorithms can be merged into a coherent framework for detecting and describing the anomalies. Our main goal is to dramatically improve the usability, effectiveness, and interpretability of anomaly mining techniques in the real world. We plan to address this goal through three broad challenges:

(1) Complex anomaly definitions and algorithms: We will invent new problem definitions on unified data models – by harnessing multiple heterogeneous data sources into enriched graph representations, and will design descriptive detection algorithms that will also assist in characterizing the anomalies.

(2) Ensembles for anomaly mining: We will leverage different detection algorithms to reduce noise and improve performance – by building ensemble techniques to assess, select, and combine multiple outputs.

(3) Multi-view anomaly mining: We will investigate a co-detection approach for anomaly mining – by enabling heterogeneous detectors to communicate and revise their findings. We will also build multi-view visualization tools and construct a taxonomy of problem formulations for easy consumption by end users.

Further, we will leverage our unique collaborations with Northrop Grumman, ObjectVideo, and the Stony Brook Hospital to apply our techniques to three real-world problems in surveillance and healthcare. It will get us closer to realizing our vision of bridging the gap between anomaly mining research and practice.

Intellectual Merit. This proposal aims to push the boundaries of anomaly mining as a field through a quest for principled foundations and practices. Research will create previously unstudied classes of data representations that unify heterogeneous data sources, and build on them to formulate novel anomaly mining problems. We will invent new, descriptive algorithms for complex anomaly detection and characterization, that will also explore and exploit ensemble and multi-view approaches. The proposed research will give rise to a comprehensive framework for anomaly mining; through a deeper understanding of the space of problems and objectives, new models and algorithms, and systematic techniques to harness them.

Broader Impacts. *Societal:* Proposed research will take the essential steps to mature anomaly mining into a valuable contributor to the larger world. It will have direct significance to many concrete problems (e.g., outsider threat, fraud, intrusion) important for the government, industry, and the society. The project will build a web-based platform that hosts a repository of formulations, algorithms, tools, and datasets, for the research community and the public to leverage. We will collaborate with industry and hospital partners to shepherd our innovations into deployed technology, with tangible impact on security and healthcare. *Educational:* The PI is committed to developing an education plan that: trains students to think creatively in formulating and solving problems, enhances undergraduate training by involving Honors thesis students in proposed research, promotes campus wide synergism for students across departments, and increases the role of women in Computer Science through mentoring and open house events for women in the community.