Venue: University of Belgrade, School of Electrical Engineering
Date and Time: Thursday, September 15th 2022, 14:00 CET
Location: Room 61, main building

Title: **Multilayer Graphs and Dynamic Sampling for Classifying Highly Imbalanced and Overlapped Social Media Content**

Speaker: **Zoran Obradovic**, Laura H. Carnell Professor of Data Analytics
Data Analytics and Biomedical Informatics Center, Computer and Information Sciences Department, Statistics, Operations and Data Science Department, Temple University

Abstract: The challenges of mining social media content we address include 1) joint analysis of news articles and user-generated content, 2) unification of massive data streams from multiple sources and platforms, 3) the unpredictable quality of content, and 4) high data imbalance when modeling response to rare events. In this talk I will describe a novel framework for aligning comments and articles from imbalanced news data characterized by different degrees of annotator agreement under a constrained budget. To obtain a more complete picture of emerging events and their associated entities we developed a multi-layer graph-based approach to learn relationships between entities and topics. Such graphs enrich textual representation and enhance model performance in many downstream applications such as media bias classification and fake news detection. Existing approaches for imbalanced classification are limited in their capacity to model data characterized by highly skewed distributions and large class overlap. Our method effectively reduces the impact of these conditions by a dynamic self-paced sampling mechanism aimed to gradually transform a class distribution from imbalanced to balanced and sample instances based on their classification difficulty. Results reported in this talk are published at:

- Alshehri, J., Stanojevic, M., Khan, P., Rapp, P., Dragut, E., Obradovic, Z. "MultiLayerET: A Unified Representation of Entities and Topics Using Multilayer Graphs," *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases,* Grenoble, France, September 2022.
- Zhou, F., Gao, S., Ni, L., Pavlovski, M., Dong, Q., Obradovic, Z., Qian, W. "Dynamic Self-paced Sampling Ensemble for Highly Imbalanced and Class-overlapped Data Classification," *Data Mining and Knowledge Discovery*, June 2022.

Biography: Zoran Obradovic is a Distinguished Professor and a Center director at Temple University, an Academician at the Academia Europaea (the Academy of Europe) and a Foreign Academician at the Serbian Academy of Sciences and Arts. He mentored 45 postdoctoral fellows and Ph.D. students, many of whom have independent research careers at academic institutions (e.g.

Northeastern Univ., Ohio State Univ.) and industrial research labs (e.g. Amazon, Facebook, Hitachi Big Data, IBM T.J.Watson, Microsoft, Yahoo Labs, Uber, Verizon Big Data, Spotify). Zoran is the editor-in-chief at the Big Data journal and the steering committee chair for the SIAM Data Mining conference. He is also an editorial board member at 13 journals and was the general chair, program chair, or track chair for 11 international conferences. His research interests include data science and complex networks in decision support systems addressing challenges related to big, heterogeneous, spatial and temporal data analytics motivated by applications in healthcare management, power systems, earth and social sciences. For more details see http://www.dabi.temple.edu/zoran-obradovic