BIOGRAPHICAL SKETCH

NAME: **Zoran Obradovic**

eRA COMMONS USER NAME:

POSITION TITLE: L.H. Carnell Professor of Data Analytics / Director, Data Analytics and Biomedical Informatics / Professor, Computer and Information Science Department

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of Belgrade, Belgrade, Serbia	B.A.	12/85	Applied Mathematics
University of Belgrade, Belgrade, Serbia	M.S.	08/87	Math/Computer Science
Penn State University, State College, PA	Ph.D.	05/91	Computer Science

A. Personal Statement

I am an internationally recognized leader in data science and biomedical informatics (about 400 articles and 25,400 citations with H-index 58). My research focuses on improving predictive modeling and decision support through data-driven discovery and modeling of hidden patterns in complex data. I was the principal investigator in pioneering research on the prediction and functional analysis of intrinsically disordered proteins that has provided new insight into how protein structure establishes function. Software developed by my team was the best-rated predictor of intrinsic disorder at three consecutive international competitions organized by protein structure prediction assessment community (CASP 5-7). Much of my research has been based on large clinical databases, and I have a history of collaboration with clinical colleagues, including currently serving as Primary Mentor to an NIDDK K23 awardee. My ***research objective*** is to collaborate on developing effective method to facilitate integration of qualitative domain knowledge and learning from data towards discovering novel interactions in high-dimensional complex networks. My ***teaching goal*** is to train data scientists how to extract value from data by integrating statistical and computational techniques into a larger framework while understanding data context, appreciating the responsibilities involved in using private and public data and clearly communicating on what a dataset can and cannot tell.

B. Positions and Honors

Positions and Employment

1991 - 2000	Assist/Assoc Prof, Electrical Eng. and Computer Science, Washington State University
1997 - 2000	Associate Prof., Electrical Eng. and Computer Science, Washington State University
2000 -	Professor, Computer and Information Sciences, Temple University
2000 -	Director, Center for Data Analytics and Biomedical Informatics, Temple University
2013 -	Professor (secondary appointment), Statistics Dept., Temple University
2013 -	L.H. Carnell Professor of Data Analytics, Temple University

Honors

2002-	Keynote speaker at 19 international conferences
2002-2006	Team leader for the best predictor in disorder category at the 5th, 6th, and 7th critical assessments of protein structure prediction experiments (CASP 5, 5 and 7)
2008	Distinguished award for excellence in research, College of Science& Technology, Temple U.
2009	President's outstanding faculty research award, Temple University
2014-2015	Elected Chair, SIAM Activity Group on Data Mining and Analytics (SIAG/DMA)
2015-	Research Professor, School of Medicine, Univ. of Belgrade
2015-	Elected Foreign Academician at the Serbian Academy of Sciences and Arts
2015-	Elected Academician at the Academia Europaea, The Academy of Europe

Editorial Boards.

- **Journal Editor-in-Chief:**
    - *Big Data Mining*, 2018 - Present

- **Journal Executive Editor:**
    - *Statistical Analysis and Data Mining*, 2009 - 2018
- **Conference Steering Committee Chair:**
    - *SIAM International Conference on Data Mining*, 2017 – Present
- **Elected Society Chair:**
    - *SIAM Activity Group on Data Mining and Analytics (SIAG/DMA),* 2014-2015
- **Journal Editorial Board Member:**
    - *Computers in Biology and Medicine, 2020 – Present*
    - *IEEE Transactions on Big Data, 2016 – Present*
    - *PeerJ Computer Science, 2015-Present*
    - *Management, 2013-Present*
    - *PeerJ, 2012-Present*
    - *Intrinsically Disordered Proteins, 2012-Present*
    - *Network Modeling and Analysis in Health Informatics and Bioinformatics, 2011- Present*
    - *Int'l J. of Computational Intelligence in Bioinformatics and Systems Biology*, 2009 - Present
    - *International Journal of Computational Models and Algorithms in Medicine*, 2009 - Present
    - *Journal of Biomedicine and Biotechnology*, 2008 - Present
    - *Advances in Bioinformatics*, 2008 - Present
    - *International Journal of Parallel, Emergent and Distributed Systems*, 2006 - Present
    - *International Journal of Data Mining and Bioinformatics*, 2005 - Present
    - *Multiple Valued Logic, An International Journal*, 1995 - Present
    - *Statistical Analysis and Data Mining*, 2006 - 2009

Guest Editor:

- *ACM Transactions on Knowledge Discovery from Data (TKDD),* Special Issue on Connected Health at Big Data Era (co-edited with F. Wang, H. Tong, M. De Choudhury), 2015
- *Mathematical Problems in Engineering,* Special Issue on Advances in High Performance Computing and Related Issues (co-edited with V. Milutinovic and B. Furht), 2015
- *Data Mining and Knowledge Discovery*, Special Issue on Data Mining for Medicine and Healthcare (co-edited with F. Wang, G. Stiglic and I. Davidson), 2014
- *Statistical Analysis and Data Mining*, the Best of SIAM SDM'09 Issue, 2(5-6), 2009
- *BMC Bioinformatics,* Text Mining in Bioinformatics, 8 (9), 2007
- *Knowledge and Info Systems*, Distributed and Parallel Knowledge Discovery, 3(4) 2001

Conference General Chair:

- *13th SIAM International Conference on Data Mining,* Austin, TX 2013
- *14th SIAM International Conference on Data Mining,* Philadelphia, PA 2014

Conference Program Chair or Co-Chair:

- *IEEE Big Data Conference,* Boston, Dec. 2017
- *KDD Big Chat Workshop: Connected Health at Big Data Era,* Sidney, Australia, Aug. 2015
- The *Cross-Cutting Computational Methods and Bioinformatics Infrastructure* at *the IEEE International Conference on Bioinformatics and Biomedicine,* Washington DC, Nov 2015
- *KDD Big Chat Workshop: Connected Health at Big Data Era,* New York, NY, Aug. 2014
- *4thInt'l Workshop on Mining Multiple Information Sources, IEEE ICDM,* Sydney, Australia, 2010
- *9th SIAM International Conference on Data Mining***,** Reno, NV, 2009
- *IEEE 2007 International Conf. Bioinformatics and Biomedicine*, San Jose, CA, 2007
- *39th Symp. Interface Statistics, Computing Science and Applications,* Philadelphia, PA, 2007
- *ACM First Int'l Workshop on Text Mining in Bioinformatics,* Arlington, MD, 2006

Conference Track Chair or Program Committee Member:

- Track chair at 13 international conferences including *ACM KDD, SIAM SDM, IEEE ICDM, CIKM and IEEE ICBB and program committee member at more than 60 international conferences.*

Steering Committee Member:

- *IEEE 2009 International Conference on Bioinformatics and Biomedicine.*
- *2010 Conference on Intelligent Data Understanding, NASA Ames Research Center.*

C. Contributions to Science

1. Protein Disorder Characterization (PI)

My laboratory has carried out a number of "first of" studies on intrinsically disordered proteins. In 1997 we discovered that thousands of proteins are disordered under physiological conditions and their functions depend on the unstructured rather than structured state. In 2002, we found that intrinsic disorder is critical for molecular recognition and that such proteins are commonly involved in cancer. These were revolutionary discoveries as for more than 100 years such proteins were not studied much since the dominant view was that a fixed structure is needed for binding and for protein function. We obtained these paradigm-shifting results by developing the first protein disorder predictor that learns from sequence-based attributes. In follow up studies we showed that disordered proteins are especially common among the eukaryotic proteomes. We found that many significant biological functions depend directly on, or are importantly associated with, the unfolded or partially folded state and characterized relationship. In 2007 we made a leap jump in answering a challenging question of understanding relationship between protein disorder and protein function by discovering that 238 Swiss-Prot functional categories are strongly positively correlated with predicted long intrinsically disordered regions. We performed a large scale analysis of intrinsic disorder in genes implicated in Human Disease Network and found significant differences in disorder content distributions for genes related to 20 human disease classes. We found that intrinsic disorder in disease genes is mainly involved in protein-protein interactions and together with alternative splicing it likely plays an important role in many diseases.

<u>Selected Publications Related to Protein Disorder Characterization</u>

(a) Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradovic, Z. (2002) "Intrinsic Disorder and Protein Function," *Biochemistry*, vol. 41, issue 21, pp. 6573 - 6582. *Findings: Using our PONDR VL-XL protein disorder predictor in this work we found that intrinsically unstructured proteins play key roles in cell-signaling, regulation and cancer.*

(b) Obradovic, Z, Peng, K, Vucetic, S., Radivojac, P., Brown, C., and Dunker, A.K. (2003) "Predicting Intrinsic Disorder from Amino Acid Sequence," *Proteins: Structure, Function and Genetics,* vol. 53 Suppl. 6, pp. 566-72. *Findings: Our disorder predictor was ranked as the most accurate at the fifth Critical Assessment of Structure Prediction (CASP 5) experiment.*

(c) Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., Obradovic, Z. (2006) "Length-Dependent Prediction of Intrinsic Protein Disorder," *BMC Bioinformatics,* vol. 7 (1), 208. *Findings: Our VSL2 protein disorder prediction that was ranked as the most accurate at the sixth Critical Assessment of Structure Prediction (CASP 6) experiment is described in this article.*

(d) Ghalwash, M., Dunker A.K., Obradovic, Z. (2012) "Uncertainty Analysis in Protein Disorder Prediction," *Molecular BioSystem,* 2012, 8 (1), 381 – 391. *Findings: A method is developed to estimate uncertainty associated with protein disorder prediction based on analysis of uncertainty in the reference model itself and the uncertainty in data.*

2. Diagnostics and Multiple Blood Cleansing Interventions in Sepsis (PI)

Sepsis is recognized as one of the leading causes of in-hospital death. The results obtained on my DARPA funded dialysis-like therapy project provide evidence that spectacular mortality rate reduction can be achieved by early interpretable diagnostics provided with uncertainty estimates followed by a combined therapy that includes precisely optimized pathogen reduction and hemoadsorption based blood cleansing interventions.

<u>Selected Publications Related to Early Intervention and Therapy Optimization in Sepsis</u>

(a) Ghalwash, M, Obradovic, Z. (2012) Early Classification of Multivariate Temporal Observations by Extraction of Interpretable Shapelets," *BMC Bioinformatics.* Aug 8;13(1):195. *Findings: The method is shown to provide accurate and easy to interpret diagnostics results much earlier in several challenging clinical applications.*

(b) Radosavljevic, V., Ristovski, K., Obradovic, Z. (2013) "A Data-Driven Acute Inflammation Therapy," *BMC Medical Genomics*, 6(Suppl 3):S7. *Findings: A therapy optimization method for sepsis learned on patients data saved 88% of patients that would otherwise die within a week, while previously the best method saved 73% of patients.*

(c) Ghalwash, M., Ramljak, D., Obradovic, Z. (2015) "Patient-Specific Early Classification of Multivariate Observations," *International Journal of Data Mining and Bioinformatics*, vol. 11, no. 4, pp. 392-411. *Findings: The proposed early diagnostics model for multivariate time*

series outperformed baselines in several challenging applications including prediction of response to drug therapy in multiple sclerosis and to sepsis therapy.

(d) Stojkovic, I, Ghalwash, M., Cao, X.H., Obradovic, Z. (2016) "Effectiveness of Multiple Blood-Cleansing Interventions in Sepsis, Characterized in Rats," *Nature Scientific Reports*, 2016, April 21, 6:24719. *Findings: A mathematical model of sepsis is utilized to find computationally that a combination of hemoadsorption therapy and pathogen reduction therapy synergistically enhances the positive effects beyond the superposition of the benefits of two approaches for sepsis treatment.*

3. **Fusion of Qualitative Knowledge and Big Data in Complex Networks (PI)**

The objective of my ONR project was to unify data-driven and knowledge-driven approaches for structured network-based predictive analytics in order to broaden the scope of possible applications. The approach consists of extending our structured regression models to explore in detail the hypothesis that a unified approach of integrating big data with high-level knowledge (ontologies, domain-based constraints etc.) is beneficial for predictive modeling of complex phenomena.

Publications Related to Fusion of Qualitative Knowledge and Big Data in Complex Networks

(a) Stiglic, G., Brzan, P.P., Fijacko, N., Fei, W., Delibasic, B., Kalousis, A., Obradovic, Z., (2015) "Comprehensible Predictive Modeling Using Regularized Logistic Regression and Comorbidity Based Features," *PLOS ONE*, Dec 8, 2015, doi: 10.1371/journal.pone. 0144439. *Findings: An interpretable method is developed that helped understand the origin and evolution of medical complications by analyzing comorbidities in a discharge database.*

(b) Stojanovic, J., Gligorijevic, Dj., Radosavljavic, V., Djuric, N., Grbovic, M., Obradovic, Z., (2016) "Modeling Healthcare Quality via Compact Representations of Electronic Health Records," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, July 14. *Findings: We learn low-dimensional representations of patient conditions and clinical procedures from electronic health records, and generate feature vectors of hospitalized patients useful for predicting their length of stay, total incurred charges and mortality rates.*

(c) Feldman, K., Stiglic, G., Dasgupta, D., Kricheff, M., Obradovic, Z., Chawla, N. (2016) "Insights into Population Health Management Through Disease Diagnoses Networks," *Nature Scientific Reports,* 2016, July 27, 6:30465 doi: 10.1038/srep30465. *Findings: The fundamental question is addressed of how to better, more efficiently and cost effectively target resources/care plans/treatment options for population health management.*

(d) Gligorijevic Dj., Stojanovic J., Djuric N., Radosavljevic V., Grbovic M., Kulathinal R., Obradovic Z. (2016) "Large-Scale Discovery of Disease-Disease and Disease-Gene Associations," *Nature Scientific Reports*, 2016, Aug. 31, 6:32404 doi: 10.1038/srep32404. *Findings: An effective method is developed to extract knowledge from disease comorbidities in big electronic health records databases.*

4. **Prospective Analysis of Large and Complex Partially Observed Networks (PI)**

Attributes and links in complex networks evolve over time and are inextricably dependent on each other. In addition, the network is partially observed, multiple kinds of links exist among nodes, various nodes have different temporal dynamics and environmental influence can be both positive and negative. My project closely examined the hypothesis that a unified approach of jointly modeling these and related problems is beneficial for prospective analysis of large-scale partially observed temporal multi-layer networks.

Selected Publications Related to Prospective Analysis of Large and Complex Networks

a. Gligorijevic, Dj., Stojanovic, J., Obradovic, Z., (2016) "Disease Types Discovery from a Large Database of Inpatient Records: A Sepsis Study," *Methods*, 111, 45-55. S1046-2023(16)30232-8. *Findings: We map a very large Electronic Health records database into a low dimensional space where diseases with similar phenotypes have similar representations which allows an effective segmentation of diseases.*

b. Glass, J., Obradovic, Z. (2017) "Structured Regression on Multi-Scale Networks," *IEEE Intelligent Systems*, Vol. 32, Issue 2, Mar-April, 2017, pp. 23-30. *Findings: A log-scale time and space complexity method for structured regression in multiscale Kronecker networks is developed in this study.*

c. Jordanski, M, Radovic, M., Milosevic, Z., Filipovic, N., Obradovic, Z. (2018) “Machine Learning Approach for Predicting Wall Shear Distribution for Abdominal Aortic Aneurysm and Carotid Bifurcation Models,” *IEEE Journal of Biomedical and Health Informatics, Vol 22, Issue 2, pp. 537-544. Findings: A structured regression model is developed that successfully predicts wall shear stress distribution at different cardiac cycle time points.*

d. *Gligorijevic,* J., Gligorijevic,Dj., Pavlovski, M., Milkovitz, E., Glass,L., Grier, K., Vankireddy, P., Obradovic, Z, (2019) “Optimizing Clinical Trials Recruitment via Deep Learning” *Journal of the American Medical Informatics Association,”* June 12. *Findings: Devpeloed an effective ranking method that learns from both investigator and trial-related heterogeneous data sources for more efficient clinical trials execution.*

Complete Bibliography (>380 publications; H-index 58; >25,400 citations)
http://scholar.google.com/citations?user=iuNSO00AAAAJ&hl=en&oi=ao
http://www.dabi.temple.edu/dabi/people/zoran/publications/publications.htm

D. Ongoing Research Support

US Department of Energy (DE-OE0000913, subaward M2000781**)** (PI: Kezunovic, M, Texas A&M Univ.) **Big Data Analysis of Synchrophasor Data**	9/19 – 3/21	
OBJECTIVE: The objective is to analyze the waveforms, extract the signature features and implement pattern analysis techniques most suitable for recognizing and differentiating a variety of events recorded by Phasor Measurement Units in parts of the Eastern and Western Interconnection, and Texas.		
National Science Foundation (IIS-8142183) (PI: Obradovic, Z.) *EAGER: Assessing Influence of News Articles on Emerging Events*	9/18 – 8/21	
OBJECTIVE: Transform the streams of social media chatter at hundreds of news outlets into data signals from which to predict the imminence of an (important) event.		
Air Force Research Laboratory (PI: Luzzi, D., Northeastern Univ.) *Modeling UAS Detection and Counter UAS R&D*	9/18 – 4/20	
OBJECTIVE: *Develop machine learning support for spatio-temporal prediction at multiple scales on limited computational resources.*		
KAUST’s Center Partnership Fund Program (PI: Obradovic, Z.) *Characterizing Diversity of Demographics and Health Conditions Affecting Risk for Comorbidities and Health Conditions Affecting Risk for Comorbidities Development in Cancer Patients*	5/19 – 4/21	
OBJECTIVE: Developing a machine learning algorithm to address cancer treatment, prognosis and follow up care more accurately and more reliably than existing alternatives.		
National Science Foundation (SES-1447670) (PI: Yoo, Y., Case Western University) *BIGDATA: Multi-Level Predictive Analytics and Motif Discovery Across Massive Dynamic Spatio-Temporal Networks in Complex Socio-Technical Systems: An Organizational Genetics*	2/2015 – 1/2020	
OBJECTIVE: Co-Principal Investigator on a big data analytics study aimed to develop a method to analyze large volumes of digital trace data using evolutionary social ontology to build a multi-level model of complex socio-technical systems.		
Verizon Media Faculty Research and Engagement Program (PI: Obradovic, Z.) **Modeling Temporal Dynamics of User Behavior for Improved Advertising**	8/19 – 7/21	
OBJECTIVE: Principal Investigator on a study aimed to developed effective predictive methods to facilitate sponsored search and purchase recommendations from big streaming data.		