Research Summary Jelena Tešić

Overview

This research summary highlights the significant contributions and impacts of my work in uncovering relationships in unstructured data through graph analysis. The research focuses on addressing the challenges associated with analyzing unstructured data, which encompasses diverse data types and formats, including records, surveys, experimental readings, tweets, and purchase reviews.

Signed Graph Analysis

The main intellectual merit of my research lies in developing scalable and assumption-free algorithms for graph analysis in unstructured data. Traditional graph analysis techniques are not suitable for unstructured data due to the inherent ambiguity and lack of clear underlying structures. By formalizing the sign and weight of relationships in a graph representation for unstructured data, my research provides a solution to this challenge.

The proposed algorithms effectively compute the fundamental cycle basis in large, unstructured graphs without requiring training data or relying on spectral computation assumptions. This enables the discovery of key patterns, communities, and recommendations, contributing to more accurate and reliable analysis in various domains. The research also emphasizes the importance of mapping unstructured data relations to edge signs and weights, providing valuable insights into the assumptions and limitations of graph analysis.

In terms of broader impacts, my research has made significant contributions to the field of network science and recommender systems. I have developed course materials on network science and recommender systems, which have facilitated the dissemination of knowledge and improved understanding in these areas. Additionally, my work has played a pivotal role in the growth of the network science branch at the Data Science Center at Texas State University, fostering research and educational advancements.

The research outcomes have also had a positive impact on the scientific community. I have provided a comprehensive graph representation guide that assists domain data scientists in making informed decisions regarding graph construction and analysis. The release of open-source analysis tools and algorithms has enabled researchers and practitioners to leverage these resources for their own investigations, enhancing the rigor and reliability of graph analysis on unstructured data.

Furthermore, my research has actively promoted diversity and inclusion by engaging students from underrepresented groups. Many dozens of students have been involved in the network science project, gaining valuable research experience and mentorship opportunities. This engagement has contributed to fostering diversity within the field of data science and nurturing the next generation of researchers.

In conclusion, my research on uncovering relationships in unstructured data through graph analysis has made significant intellectual contributions and broader impacts. By developing scalable and assumption-free algorithms and providing valuable resources to the scientific community, I have advanced the field of network science and improved the evaluation and effectiveness of graph analysis on unstructured data. The engagement of

students from underrepresented groups has further promoted diversity and inclusivity in the field. These achievements demonstrate my expertise, leadership, and commitment to excellence, making me a strong candidate for tenure and promotion.

Data science is an ever-evolving umbrella that pushes advances in a wide range of disciplines as it combines novel algorithms and methods from fundamental sciences with sophisticated engineering approaches to solve challenges presented by societal, economic, and scientific needs. I have named my education and outreach target group the "New Kids On the Block" (NKOTB): this is an umbrella term for all underrepresented student and domain expert researchers I have encountered in my career in Computer Science: racial and national minorities, first-generation students, women/LGBTQIA +, neurodivergent people and people that want to do science but have no connections, mentors, or means to start the path and are likely to quit or fail. My **goal** for this project is to introduce network science to NKOTB and to educate them in all aspects on their paths to become data scientists.

EO1. Course development My plan is to introduce sign graph network science for real data and integrate tasks **A1**, **A2**, and **E1** into the student project in the Ph.D. Data Science (CS7311) course. Students become more motivated in data science when there is a clear new problem to be solved, and they will be given the opportunity to continue work as part of the independent research study course (CS7387) and thesis work. I have taught CS7311 since its inception in 2018, and class projects led to thesis chapters, teams that won the benchmark competition, and peer-reviewed publications [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]. The enrollment in the class is 10 CS Ph.D. students, 5 CS M.Sc. students, and 5 Ph.D./M.Sc. students from other departments. The expected outcome is to upskill up to one hundred graduate students at Texas State in network science and to expand this course and introduce separate focused sessions for College of Science and Engineering non-CS students and for McCoy College of Business Administration students. In my very popular object-oriented (CS3354) and machine learning (CS4347/CS5369L) undergraduate and graduate classes, half of the student body can be categorized as NKOTB. Such students are reluctant to join coding clubs as it re-emphasizes the stereotypes, and my plan is to engage them in coding projects in the classes I teach. First, they get to practice object-oriented coding using simple signed graph data for the CS3354 project (Tasks **G1** and **A1**), and then they learn the data science pipeline and signed graph modeling as part of the CS4347 project (Tasks **G2** and **A2**). I and my Doctoral Instructional Assistants instruct 100 per year in CS3354, and 80 per year in CS4347. The expected result is to get 900 students exposed and interested in data science, and 20 mentored, as described in **EO2**.

EO2. Mentoring Development Most of the students who choose to work in the Data Lab [**?**] can be classified as NKOTB, and the graduate students working on this project are not different. Although their challenges are diverse, they have had a similar effect on their career progression: no real guidance or support system; no networking skills; afraid to ask questions or to ask for help; no job interview skills. My multisemester educational pipeline allows NKOTB students to improve their data science and coding skills and to add research project to their resume. My path of mentoring has been shaped by my experience teaching and mentoring students at graduate school, the industrial research laboratory,

and TXST. The support for NKOTB's unique needs has been slow and incidental, and one of the reasons I returned to academia after 13 years in the industry is to speed up this process and support underrepresented talent in their career paths. As a high functioning female with social anxiety who faced prejudice in the US due to her nationality at the beginning of graduate school and who has had corporate startup and government job experiences, I am in a good position to help bridge the gap in the successful computer science research and representation, and this project will help me in that effort. I have mentored 20 undergraduate students for independent research courses, NSF REU and SURE summer projects, and thesis and sponsored projects, 70% were in the NKOTB. 20+ NKOTB undergraduate students (**I** label in Sec. **??**) will participate in this research as part of undergraduate research and thesis courses, and 2 Ph.D. students funded (**Sa,Sb** labels in Sec. **??**). I will setup weekly Data Lab meets to help students succeed in the career path and cover topics on research, ethics, paper reading, writing, how to present your work, events logistics, job interviews, and resume polishing, how to network when you feel less than or have high social anxiety, and coding reviews.

EO3. Development of outreach My outreach plan is to validate the results of tasks **G1** and **A3** based on current findings to identify areas of resilience research where data-driven methods have the greatest impact in my role as a fellow at the TXST Translational Health Research Center. The algorithms and methods developed in this project will help me identify different decision-making processes when approving or denying treatments, identify highly effective health teams and practices, and validate data integration methods in the use of health data. I am also the founding member and lead investigator of predictive methods at the newly established university-level TXST Data Science Center. As the Center will be led by a group of senior assistant and junior associate professors, this project will help me to find interdisciplinary teams and data sets to apply methods **A1, A2** developed in the project. Next, the plan in years 4 and 5 is to extend project integration with CS7311 and create a new interdisciplinary course where the material is tailored to a specific (science, engineering, business) domain-expert audience within the Center. The collaboration plan is listed in the "Facilities, Equipment and Other Resources" section of the proposal, and the affected audience is estimated to be six faculty and 20 students at TXST, and two faculty and eight students in Europe in the first two years of the project. That number will grow to hundreds of TXST students when the cross-disciplinary data science courses are offered in Fall 2025.