Jelena Tešić: Research Impact and Plan

Unstructured data is one name to encompass diverse data types and formats, such as video data, electronic health data records, surveys, experimental readings, sensor data, social platform content data, genomics data, and purchase reviews. The arching quest of my research is to provide useful, efficient, effective, intuitive, and responsible algorithms for addressing the challenges of analytics tasks when applied to such extensive unstructured data collections in the wild. The research I lead in the Data Lab (DataLab12.github.io) at Texas State is grouped in the following sections by the tasks it solves.

1 Network Science for Unstructured Data

1.1 Scaling Signed Graph Balancing Tasks through Fundamental Cycle Basis

In network science, signed graph representation of relations provides more information than unsigned graph networks do. Spectral pollution, computational complexity, inherent bias, and training requirements are the main hurdles that have slowed the adoption of the signed network as data representation and signed graph algorithms for modern data analysis. We took a different approach, stemming from mathematical sociology, and focus on balance theory to solve the signed network tasks at scale. First, we have expanded the balance theory to signed social network graph analysis. We propose a frustration cloud view of the signed graph where we quantified vertex and edge in terms of frustration cloud statistics and validated this novel social network graphs analysis approach [18]. Next, I worked with the Ph.D. student and her advisor to develop an assumption-free algorithm to efficiently compute the fundamental cycle basis in large, unstructured graphs without requiring training data or relying on spectral computation assumptions to scale the findings to signed graphs constructed from social networks and recommendation data [1]. There was no current baseline to evaluate state-of-the-art tasks on signed graphs derived from real networks, so we have created the most extensive assumption-free comparison of community discovery on signed networks in [26] in terms of efficacy, efficiency, scalability, and reproducibility of existing methods, and propose frustration cloud-based approach for cluster boosting for high modular signed graphs [27]. Two master and five undergraduate students contributed and were supported by this research project.

The research plan is to use a fundamental cycle basis to scale the solution of other NP-hard tasks to large signed networks using the efficient fundamental cycle computation approach [21] and propose algorithms to solve the task of computing frustration and the balanced state of the graph at scale [23] and the task of finding the largest balanced subgraph in any network [22] for graphs with millions of nodes and edges. Next, we plan to propose a scalable community discovery algorithm and extend the analysis using a fundamental cycle basis to the recommendation and anomaly detection tasks, as well as to extend the application to sensor and agent networks and signed gene networks. The project has evolved into a major part of the Ph.D. thesis work for one student.

1.2 Modeling Social Network Relations and Improving Label Propagation

A large number of users, messages, and tags makes it difficult to separate exciting and meaningful conversation threads from spreading fake news, malicious accounts, background noise, or irrelevant trolling on Twitter. First, we have introduced a scalable end-to-end Twitter network data management pipeline that gathers, stores, and models

rich relationships from Twitter networks [14]. Next, we compared and contrasted the analysis results of millions of Twitter data using multiple graph construction processing approaches [5].

We seek to answer how well the content of Tweets can be classified based on modeling relationships from interactions alone and how well community classification can predict the content class. We unveil compelling evidence demonstrating that community-based modeling techniques yield precision, recall, and accuracy comparable to lexical classifiers. Remarkably, these promising results are achieved even without considering the textual content of tweets beyond hashtags. We show that the combined community and lexical approach, which consistently delivers the most robust outcomes and exhibits the highest performance measures for network graphs constructed based on Twitter interactions related to the COVID-19 pandemic [20]. One undergraduate, one master, and one Ph.D. student participated in the research work for this research project. The next step is to model social network connections using signed extended graphs (edge does not have to have a sign) and explore multiple graph network representations to provide an assumption-free and bias-free baseline for evaluating graphical neural network performance.

2 Advancing Computer Vision tasks in the Wild

2.1 Solving Computer Vision tasks for Overhead Videos

Domain adaptation for Object classification [4] (under review); [3, 2] Small object classification [28] and [9] Activity recognition [25] and [24]

Segmentation [**?**]

2.2 Deep Descriptor Database Indexing Search and Retrieval

Deep descriptors, genomics descriptors, clustering, indexing, search, summarizing, ... Published: [19], [16, 12] Under review: [17]

3 Predictive Modeling of Noisy Tabular and Survey Data

Under review: survey teacher loss [29] and COVID learning loss [15] Published: Covid data ... [7]

4 Summary

Summary on other applications [10] [11] [13, 6, 8]. Plan is to

References Cited

[1] Ghadeer Alabandi, **Jelena Tešić**, Lucas Rusnak, and Martin Burtscher. Discovering and balancing fundamental cycles in large signed graphs. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '21, New York, NY, USA, 2021. Association for Computing Machinery.

[2] Debojyoti Biswas, Toufik Rahman, Ziliang Zong, and **Jelena Tešić**. Improving the energy efficiency of real-time dnn object detection via compression, transfer learning, and scale prediction. In *The IEEE 16th International Conference on Networking, Architecture, and Storage (NAS 2022)*, September 2022.

[3] Debojyoti Biswas and **Jelena Tešić**. Small object difficulty (sod) modeling for objects detection in satellite images. In *2022 IEEE 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 125–130, 2022.

[4] Debojyoti Biswas and **Jelena Tešić**. Domain adaptation with contrastive learning for object detection in satellite imagery (under review). *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[5] Lia Nogueira de Moura and **Jelena Tešić**. pytwanalysis: Twitter data management and analysis at scale. In *2021 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2021.

[6] Noah Dunstatter, Alireza Tahsini, Mina Guirguis, and **Tešić, Jelena**. Solving cyber alert allocation markov games with deep reinforcement learning. In Tansu Alpcan, Yevgeniy Vorobeychik, John S. Baras, and György Dán, editors, *Decision and Game Theory for Security*, pages 164–183, Cham, 2019. Springer International Publishing.

[7] Mirna Elizondo, Rasim Musal, June Yu, and **Jelena Tešić** on behalf of N3C. Long covid challenge: Predictive modeling of noisy clinical tabular data. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, 2023.

[8] Blake W Ford, Apan Qasem, Jelena Tešić, and Ziliang Zong. Migrating software from x86 to arm architecture: An instruction prediction approach. In *2021 IEEE International Conference on Networking, Architecture and Storage (NAS)*, pages 1–6, 2021.

[9] David Heyse, NicholasWarren, and **Jelena Tešić**. Identifying maritime vessels at multiple levels of descriptions using deep features. In Tien Pham, editor, *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006, pages 423 – 431. International Society for Optics and Photonics, SPIE, 2019.

[10] Andreas Lommatzsch, Benjamin Kille, Özlem Özgöbek, Yuxiao Zhou, **Jelena Tešić**, Cláudio Bartolomeu, David Semedo, Lidia Pivovarova, Mingliang Liang, and Martha Larson. Newsimages: Addressing the depiction gap with an online news dataset for text-image rematching. In *Proceedings of the 13th ACM Multimedia Systems Conference*, MMSys '22, page 227–233, New York, NY, USA, 2022. Association for Computing Machinery.

[11] Andrew Magill, Lia Nogueira de Moura, Maria Tomasso, Mirna Elizondo, and **Jelena Tešić**. Enriching content analysis of tweets using community discovery graph analysis. In *Proceedings of the MediaEval 2020 Workshop*, volume 2882, 2020.

[12] M M Mahabubur Rahman and **Jelena Tešić**. Hybrid approximate nearest neighbor indexing and search (hannis) for large descriptor databases. In *2022 IEEE International Conference on Big Data*, pages 3895–3902, 2022.

[13] Taylor Mauldin, Anne H. Ngu, Vangelis Metsis, Marc E. Canby, and **Jelena Tešić**. Experimentation and analysis of ensemble deep learning in iot applications. *2019 VLDB DMAH*, 5(1):133–149, 2019.

[14] Lia Nogueira de Moura. Social network analysis at scale: Graph-based analysis of twitter trends and communities. Master's thesis, Texas State University, 2020.

[15] Daniel Payan, June Yu, Li Feng, and **Jelena Tešić**. Data driven intervention for covid learning loss in texas public school districts (under review). *IEEE Transactions on Learning Technologies*, 2023.

[16] M M Mahabubur Rahman and **Jelena Tešić**. Evaluating hybrid approximate nearest neighbor indexing and search (hannis) for high-dimensional image feature search. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6802–6804, 2022.

[17] MMM Rahman, Debojyoti Biswas, and **Jelena Tešić**. Evirec: Efficient visual indexing and retrieval for edge crowd-sensing (under review). In *2023 IEEE/ACM 8th Symposium on Edge Computing (SEC)*, 2023.

[18] Lucas Rusnak and **Jelena Tešić**. Characterizing attitudinal network graphs through frustration cloud. *Data Mining and Knowledge Discovery*, 6, November 2021.

[19] Hanie Samimi, Jelena Tešić, and Anne Hee Hiong Ngu. Patient centric data integration for improved diagnosis and risk prediction. In *Heterogeneous Data Management, Polystores, and Analytics for Healthcare*, pages 185–195, Cham, 2019. Springer International Publishing.

[20] Muhieddine Shebaro, Lia Nogueira de Moura, and **Jelena Tešić**. Multimodal mining of twitter networks for improved label propagation (under review). *Social Network Analysis and Mining*.

[21] Muhieddine Shebaro and **Jelena Tešić**. Identifying stable states of large signed graphs. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion)*, 2023.

[22] Muhieddine Shebaro and **Jelena Tešić**. Abcd: Algorithm for balanced component discovery in signed networks. In *Under Submission*, 2023.

[23] Muhieddine Shebaro and **Jelena Tešić**. Scaling frustration index and corresponding balanced state discovery for real signed graphs. In *Under Submission*, 2023.

[24] George E. Strauch, Jiajian Jax Lin, and **Jelena Tešić**. Overhead projection approach for multi-camera vessel activity recognition. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 5626–5632, 2021.

[25] **J. Tešić**, D. Tamir, S. Neumann, N. Rishe, and A. Kandel. Computing with words in maritime piracy and attack detection systems. In Dylan D. Schmorrow and Cali M. Fidopiastis, editors, *Augmented Cognition. Human Cognition and Behavior*, pages 434–444, Cham, 2020. Springer International Publishing.

[26] Maria Tomasso, Lucas Rusnak, and **Jelena Tešić**. Advances in scaling community discovery methods for signed graph networks. *Journal of Complex Networks*, 10(3), 06 2022. cnac013.

[27] Maria Tomasso, Lucas Rusnak, and **Jelena Tešić**. Cluster boosting and data discovery in social networks. In *Proceedings of the 37th ACM/SIGAPP Symposium On Applied Computing (SAC)*, 2022.

[28] Nicholas Warren, Ben Garrard, Elliot Staudt, and **Jelena Tešić**. Transfer learning of deep neural networks for visual collaborative maritime asset identification. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 246–255, Oct 2018.

[29] June Yu, Li Feng, and **Jelena Tešić**. Mitigating u.s. public school teacher attrition crisis: A data science approach (under review). *Information Processing & Management*.