Jelena Tešić: Research Impact and Plan

The arching quest of my research work is how to provide efficient, effective, intuitive, and responsible access to and interaction with unstructured data collections for summarizing, mining, querying, analyzing, and recommendation tasks. In the process, the proposed work is addressing the challenges associated with analyzing such unstructured data, which encompasses diverse data types and formats, including multimedia data, health data records, surveys, experimental readings, sensor data, social platform content data, genomics data, and purchase reviews. In this document, I highlight the contributions of my research work at Texas State as well as planned research going forward along the main approach for representing unstructured data (i) deep and multimodal features (ii) noisy tabular data (iii) signed graphs; and (iv) multimodal data representation combining (i) - (iii).

Indexing and Summarizing Descriptor Databases

Deep descriptors, genomics descriptors, clustering, indexing, search, summarizing, ... Published: [12, 15] [**?**] Under review: [16]

High Variability Video Collections Tasks

Domain adaptation for Object classification [4] (under review); [3, 2]

Small object classification [26] and [9]

Activity recognition [23] and [22]

Multimodality ... [10]

Predictive Modeling of Noisy Tabular and Survey Data

Under review: survey teacher loss [27] and COVID learning loss [14]

Published: Covid data ... [7]

Signed Graph Analysis of Relations in Unstructured Data

Traditional graph analysis techniques are not suitable for unstructured data due to 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. We have developed a scalable and assumption-free algorithm to 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.

Published: [17] [25] [1] [24] [19]

Under review: [21], [20]

The challenge is an NP-hard problem; there is no current baseline to evaluate state-of-the-art signed graphs derived from real networks. In this paper, we propose a scalable state-of-the-art approach for the maximum balanced sub-graph detection in the network of

any size although it is still bounded by computational capability. The proposed approach builds on the graph characteristics and a scalable fundamental cycle discovery method to minimize the number of nodes discarded. We evaluate the proposed approach against state-of-the-art and demonstrate over two times higher graph size regarding the number of nodes selected of the discovered subset on an extensive signed network with millions of nodes and edges over the state-of-art in the same time frame.

[11, 5]
in submission: [18]
Unclassified: [13, 6, 8].

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] 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.

[15] 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.

[16] 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.

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

[18] 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*.

[19] 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.

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

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

[22] 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.

[23] **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.

[24] 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.

[25] 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.

[26] 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.

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