PhD subject: Advanced machine learning methods for characterizing animal welfare

Advisor: Marc Plantevit & Céline Robardet

1 Context

The proposed PhD thesis is part of the PEPR project known as WAIT4, focusing on *"Artificial Intelligence and New Technologies for Tracking Key Indicator Traits in Animals Facing Challenges of the Agro-Ecological Transition."* The primary objective of this project is to enhance animal welfare (AW), acknowledged as a crucial factor for the sustainability of livestock production systems.

The agroecological (AE) transition is anticipated to significantly impact AW, presenting both positive outcomes, such as increased space allowance for animals and greater freedoms, as well as specific vulnerabilities including less optimized and more diverse feeding regimens, fluctuating environments, and the emergence of pathogens. Environmental conditions in the upcoming years are expected to be more variable due to the consequences of global warming, exerting broader effects on animals raised in AE systems characterized by less optimization and outdoor conditions.

In light of these challenges, there is a heightened need for the development of new assessment tools for AW and decision-making tools. This becomes even more crucial considering the variable environmental conditions and the potential benefits of adopting AE practices to enhance AW. Strategies aimed at improving AW heavily rely on the measurement of traits related to physiology, behavior, and the animal's perception when encountering diverse challenges.

2 PhD objective

The WAIT4 project leverages the capabilities of digital technologies to continuously monitor various aspects of animal welfare (AW) in real time. The primary objective of the PhD thesis is to develop artificial intelligence algorithms capable of extracting the most pertinent indicators (proxies) of AW from extensive and diverse datasets collected through these devices positioned on, in, or around the animals.

The proposed work in this thesis aims to create models of the activities, behaviors, and dynamics within a herd of animals. These methods will be designed to be generic and be applicable to other datasets structured as heterogeneous graphs, such as social media data, for instance. The project started over a year ago, and the data is currently accessible and prepared for analysis.

3 State of the art

Heterogeneous information networks, as described by Sun and Han in their work [SH12], provide a framework for learning intricate models from linked data based on spatial, temporal, or other relationships. Utilizing a multigraph representation (where a graph $G = (V, E)$ can feature multiple edges between a given pair of nodes and incorporates labeled vertices and edges), semantic relationships can be translated into sets of paths connecting groups of entities. These path sets, known as meta-paths, can be effectively leveraged by algorithms, as highlighted in previous studies [FLL17, CW17, ZC22].

The LIRIS DM2L team has achieved significant results in recent years regarding the analysis of such networks [Rob09, DPRB13, PPRB13, BPR16, KPZ+17, BML+20, BPRA21]. Additionally, the team has developed expertise in Graph Neural Networks (GNN) as a proxy for extracting key characteristics of a graph [VKD+22, VFKD+22].

The data describing activities and social interactions between animals are at the same time, relational, heterogeneous and dynamics. The tools for analyzing such data are few and need to be adapted to the specific questions related to the study of animal welfare, taken into consideration that phenomena occurred at multiple time and space scales. The candidate will develop graph based unsupervised and supervised approaches for analyzing interaction data and finding regularities in relation to AW. The researcher will also identify discriminant subgraphs [CDV22] that are specific to negative or positive conditions experienced by the animal. When dealing with spatial data, particularly GPS data, the optimal level of granularity can be automatically identified to pinpoint the most pertinent descriptions of spatial interaction patterns. In instances where exact matching between subgraphs is unfeasible due to excessive noise, metrics between graphs will be defined using optimal transport concepts, as outlined in [VCT+19]. Subsequently, standard machine learning and data mining methods that rely on a distance measure between studied objects will be applied.

To further enhance the analysis, discriminant analysis will be performed to understand how different spatial interaction patterns contribute to AW.

4 Expected theoretical contributions

The envisioned research perspective involves utilizing Graph Neural Networks (GNNs) as a proxy to extract essential features from a graph. While some avenues may already be under consideration, the following list is not exhaustive and represents potential directions for exploration:

- Node Embeddings and Graph Representations: Investigate techniques within GNNs for generating embeddings of nodes and overall graph representations, aiming to capture meaningful information about the structure and relationships within the graph.

- Attention Mechanisms in GNNs: Explore the integration of attention mechanisms within GNN architectures to enhance the model's focus on relevant nodes and edges, particularly in scenarios where certain parts of the graph may be more influential than others.

- Graph Classification and Regression: Apply GNNs to address graph classification or regression tasks, with a focus on predicting key characteristics or labels associated with the graph, such as animal welfare indicators.

- Exploit the different levels of data aggregation resulting in a hierarchy to identify the most informative level [BLP+19];

- Define metrics between graphs using optimal transport concepts in order to be able to use standard methods of machine learning and data mining requiring a distance measure between the studied objects;

- Propose a gradual approach to quantify the importance of some sub-graphs in relation to numerical variables expressing animal welfare.

5 Expected application contributions

The research faces several challenges from the application point of view, each pivotal in advancing our understanding of animal welfare through the extraction of key features from diverse data sources. These challenges encompass:

- Quantifying activity time budget: The task involves developing methodologies to precisely quantify the activity time budget of animals using biometric data collected in various situations. This requires sophisticated algorithms capable of discerning and categorizing different activities, considering the dynamic nature of animal behavior across diverse contexts.

- Describing inter-related animal positions: Understanding the spatial dynamics of animals is crucial for assessing space occupation and social behavior. The challenge is to devise techniques that describe and analyze inter-related animal positions, considering factors such as proximity, movement patterns, and group interactions. This involves capturing the complexity of spatial relationships within a given environment.

- Proposing recognition systems for emotions: Developing recognition systems capable of classifying positive and negative emotions is a complex task. This challenge requires the creation of models that can interpret biometric signals to identify emotional states in animals. It involves bridging the gap between biometric data and emotional responses, considering nuances in expression and behavior associated with various emotional states.

- Identifying relevant patterns under different husbandry conditions: Recognizing and understanding relevant patterns, specifically activity profiles, in relation to different husbandry conditions is essential. This challenge entails uncovering associations between animal behavior and environmental factors, such as variations in feeding regimens, space availability, and other husbandry practices. The goal is to identify patterns that serve as indicators of animal welfare across diverse contexts.

Addressing these challenges requires a multidisciplinary approach, integrating expertise in artificial intelligence, data science, and animal behavior. The development of robust algorithms, the incorporation of advanced machine learning techniques, and a nuanced understanding of animal physiology and psychology are crucial components of overcoming these challenges. Successfully tackling these issues will contribute significantly to the development of new machine learning techniques and effective tools for assessing and improving animal welfare in a variety of contexts.

References

[BLP+19] Anes Bendimerad, Jefrey Lijffijt, Marc Plantevit, Céline Robardet, and Tijl De Bie. Contrastive antichains in hierarchies. In Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis, editors, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pages 294–304. ACM, 2019.

[BML+20] Anes Bendimerad, Ahmad Mel, Jefrey Lijffijt, Marc Plantevit, Céline Robardet, and Tijl De Bie. Sias-miner: mining subjectively interesting attributed subgraphs. *Data Min. Knowl. Discov.*, 34(2):355–393, 2020.

[BPR16] Ahmed Anes Bendimerad, Marc Plantevit, and Céline Robardet. Unsupervised exceptional attributed sub-graph mining in urban data. In *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, pages 21–30, 2016.

[BPRA21] Anes Bendimerad, Marc Plantevit, Céline Robardet, and Sihem Amer-Yahia. User-driven geolocated event detection in social media. *IEEE Trans. Knowl. Data Eng.*, 33(2):796–809, 2021.

[CDV22] Corinna Coupette, Sebastian Dalleiger, and Jilles Vreeken. Differentially describing groups of graphs. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 3959–3967. AAAI Press, 2022.

[CW17] Yuxin Chen and Chenguang Wang. HINE: heterogeneous information network embedding. In K. Selçuk Candan, Lei Chen, Torben Bach Pedersen, Lijun Chang, and Wen Hua, editors, *Database Systems for Advanced Applications - 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part I*, volume 10177 of *Lecture Notes in Computer Science*, pages 180–195. Springer, 2017.

[DPRB13] Elise Desmier, Marc Plantevit, Céline Robardet, and Jean-François Boulicaut. Trend mining in dynamic attributed graphs. In *Machine Learning and Knowledge Discovery in Databases*, pages 654–669. Springer, 2013.

[FLL17] Tao-Yang Fu, Wang-Chien Lee, and Zhen Lei. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In Ee-Peng Lim, Marianne Winslett, Mark Sanderson, Ada Wai-Chee Fu, Jimeng Sun, J. Shane Culpepper, Eric Lo, Joyce C. Ho, Debora Donato, Rakesh Agrawal, Yu Zheng, Carlos Castillo, Aixin Sun, Vincent S. Tseng, and Chenliang Li, editors, *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 1797–1806. ACM, 2017.

[KPZ+17] Mehdi Kaytoue, Marc Plantevit, Albrecht Zimmermann, Anes Bendimerad, and Céline Robardet. Exceptional contextual subgraph mining. *Machine Learning*, pages 1–41, 2017.

[PPRB13] Adriana Prado, Marc Plantevit, Céline Robardet, and Jean-François Boulicaut. Mining graph topological patterns: Finding co-variations among vertex descriptors. *IEEE Trans. Knowl. Data Eng.*, 25(9):2090–2104, 2013.

[Rob09] Céline Robardet. Constraint-based pattern mining in dynamic graphs. In *ICDM 2009, The Ninth IEEE International Conference on Data Mining, Miami, Florida, USA, 6-9 December 2009*, pages 950–955, 2009.

[SH12] Yizhou Sun and Jiawei Han. Mining heterogeneous information networks: a structural analysis approach. *SIGKDD Explor.*, 14(2):20–28, 2012.

[VCT+19] Titouan Vayer, Nicolas Courty, Romain Tavenard, Laetitia Chapel, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *ICML*, pages 6275–6284, 2019.

[VFKD+22] Luca Veyrin-Forrer, Ataollah Kamal, Stefan Duffner, Marc Plantevit, and Céline Robardet. On GNN explainability with activation rules. *Data Mining and Knowledge Discovery*, pages 1–35, October 2022.

[VKD+22] Luca Veyrin-Forrer, Ataollah Kamal, Stefan Duffner, Marc Plantevit, and Céline Robardet. What does my GNN really capture? on exploring internal GNN representations. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 747–752. ijcai.org, 2022.

[ZC22] Xinyi Zhang and Lihui Chen. mshine: A multiple-meta-paths simultaneous learning framework for heterogeneous information network embedding. *IEEE Trans. Knowl. Data Eng.*, 34(7):3391–3404, 2022.