Research Statement

Enyan Dai (emd5759@psu.edu)

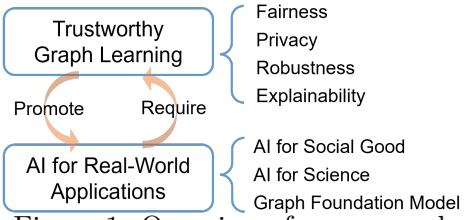

Figure 1: Overview of my research.

Graph-structured data is ubiquitous in various domains, ranging from social networks and energy transportation systems to molecular and protein structures. Numerical real-world applications can be facilitated with the investigation on graph mining. For example, friend recommendation on the social network can be regraded as the link prediction task. Chemical molecular properties prediction can be treated as graph classification. **Given the prevalence of graph-structured data across domains, my overarching research goal is to utilize AI in a responsible manner to serve humanity.** Guided by this principle, as the Fig. 1 shows, my contributions and future plans focus on trustworthy graph learning and AI for real-world applications.

Trustworthy Graph Learning. Graph Neural Networks (GNNs) have emerged as powerful methods for modeling graph-structured data. Despite their capabilities, GNNs have been found to exhibit various trustworthiness issues. These include private information leakage, vulnerability to adversarial attacks, classification biases, and lack of explainability. This largely hinders the adoption of GNNs in high-stake applications. Therefore, my research mainly focuses on trustworthy graph neural networks to ensure the allure of AI can be enjoyed without concerns. I have already conducted foundation research in graph learning algorithms for each aspect of the trustworthiness, resulting a number of high impact publications in top-tier conferences (e.g., KDD, WWW, ICLR, NeurIPS, WSDM, and CIKM). In the future, I plan to consolidate these into a unified framework of trustworthy graph neural networks to facilitate their deployment in high-stake applications. I also aim to extend the trustworthiness to the future foundation model on graph-structured data. And these investigations in trustworthy graph learning will promote the graph-augmented AI for real-world applications, which is the other research direction of mine.

Graph-Augmented AI for Real-World Applications. In many real-world application scenarios, we often encounter relational data that can provide pivotal information for the target task. For instance, fake news detection can benefit from the news propagation network. The ability of GNNs in encoding relational information paves the way to enhance AI on these applications. Therefore, I am also interested in developing graph-augmented AI for real-world challenges. Specifically, some influential works for social good such as fake news detection and power grid anomaly detection have been investigated. In scientific problems, graph-structured data such as molecular and protein also can encode vital knowledge. In the coming years, I aim to broaden the scope of GNN applications in scientific problems. I plan to achieve this by collaborating with researchers from diverse domains. This includes computer science subfields such as system and network, as well as other scientific disciplines like biology, chemistry, and physics. One promising technical path to achieve the aforementioned goal is building graph foundation models for domains such as network anomaly detection and protein analysis. In addition, the trustworthiness will be considered in the future exploration.

Research Contributions

Foundation Research in Trustworthy Graph Learning From the computational perspective, the trustworthiness encompasses four main dimensions: *fairness* [7, 16, 15], *privacy* [9, 3], *robustness* [1, 3, 4, 5, 14], and *explainability* [8, 10, 11]. Next, I outline my contributions in each aspect.

- **Fairness**: Many applications of graph neural networks (e.g. credit estimation with financial network) require the model predictions to be fair for different individuals and demographic groups. However, my analysis in [7] first verifies that the societal bias in graph-structured data can be magnified with the message-passing mechanism of GNNs, resulting severe discrimination towards protected sensitive attributes such as races and genders. To address discrimination in graph learning, I constructed three graphs containing sensitive attributes which have become widely recognized benchmarks for fair graph

learning [7]. In addition, I developed FairGNN [7] which guarantees group fairness with adversarial debiasing and fairness constraints. Acknowledging the difficulty of collecting users' sensitive attribute information, FairGNN proposes to utilize the estimated sensitive attributes to constrain the GNN models for fairness. The effectiveness of FairGNN is demonstrated by thoroughly theoretical analysis and experiments. Following this, I also collaborated with fellow researchers to explore fair model training without sensitive attributes [16, 15].

- **Privacy**: It has been proven that AI models have a risk of leaking private information of training data. To protect the privacy of sensitive attributes, FairGNN was extended to comply with local differential privacy which provides strong theoretical privacy guarantee [9]. Specifically, differentially private sensitive attributes are produced and collected in users' own devices by randomly flipping their true sensitive attributes [9]. This approach, in theory, safeguards against the disclosure of sensitive information from GNN models. Furthermore, the overfitting of GNNs can reveal the membership of training samples. Therefore, I established a graph information bottleneck (GIB) framework that can protect the membership privacy [3]. The proposed graph information bottleneck can constrain the mutual information between node representations and labels on the training set, which narrows the gap between training and test sets to protect the membership privacy.
- **Robustness**: Various types of noises and adversarial attacks can significantly degrade the performance of GNNs. To address these challenges, I firstly developed a novel NRGNN framework [1] that is resistant to different types of noises in annotations. Secondly, I designed approaches that can eliminate structural noises and malicious links in graphs [3, 4]. Finally, I further investigated the vulnerability of GNNs under the backdoor attack [5], which is a new form of attack method. The proposed approach can unnoticeablely backdoor models trained on large-scale graphs with over 100K nodes. Our results indicate that GNNs (even many existing robust GNNs) can be easily injected backdoors. And we are going to investigate how to defend against this backdoor attack in the future.
- **Explainability**: Providing explanations from models enhances trust in the predictions made by AI systems. The extracted explanations can also facilitate the discovery of knowledge from data. However, initial efforts to interpret GNN predictions largely rely on auxiliary post-hoc explainers, which could misrepresent the true inner working mechanism of the target GNN. Therefore, I have studied self-explainable GNNs that can simultaneously give accurate predictions and corresponding explanations [8, 10]. The self-explainable GNN in [8] will identify K-nearest labeled samples with interpretable similarity metric for both explaining and predicting. Our work in [10] introduces a prototype-based self-explainable GNN that learns prototype graphs capturing representative patterns of the corresponding class.

Graph-Augmented AI for Social Good. To promote social good with GNNs, I have conducted research in *fake health news detection with social networks* [6] and *power grid anomaly detection* [2].

- **Fake Health News Detection**: Nowadays, Internet has been a primary source of attaining health information. However, the propagation of fake health news over the Internet has become a severe threat to public health. Motivated by this, I constructed and released the first comprehensive fake health news dataset repository, which includes thousands of labeled health news with explanations and rich social contexts [6]. The extracted explanations can guide the training of explainable fake health news identification. Over 500K related tweets and 200K engaged user profiles are collected to facilitate the fake health news detection with social context. Released in 2019, this repository enabled research on GNN-based fake health news detection using social network information, which was particularly valuable during the COVID pandemic.
- **Power Grid Anomaly Detection**: In spatial-temporal data such as power grid, time series at nearby locations are often correlated and their behavior may be causal under cascading effects. To model such interdependencies, I developed a novel graph-augmented anomaly detector. This framework learns a

Bayesian Network to model relationship of nodes in power grid graph, which facilitates the identification of anomalies, such as power failures.

Future Research Directions

My short term goal is to explore frontiers of trustworthy GNNs. In the longer term, I plan to incorporate the trustworthy model for science problems and build graph foundation models for critical domains.

Unified Framework for Trustworthy GNNs. High-stake applications, such as financial analysis, often require multiple dimensions of trustworthiness. However, existing research in trustworthy GNNs mainly focuses on a single dimension of trustworthiness. Methods designed for distinct trustworthy aspects can be conflict with each other in model design, resulting in poor performance. Consequently, a unified framework for trustworthy GNNs is necessary. I have already conducted a preliminary work in unifying adversarial robustness and membership privacy within the graph information bottleneck (GIB) framework [3]. As the principle of GIB is to extract minimal sufficient information, it is promising to further utilize GIB to remove sensitive attribute information for fairness. The extracted minimal sufficient information can be analyzed to interpret the model's decision-making processes, promoting the explainability of graph learning. Another potential direction is to enhance the robustness and fairness with model explanations. I have already worked on self-explainable GNNs [8, 10]. Moving forward, my aspiration is to align the self-explanations generated by models with the rationale provided by humans when making decisions that are both robust and unbiased. To achieve this, research in data construction, techniques of alignment, and self-explainable frameworks for various applications would be investigated.

Trustworthy Model for Science. Graph-structured data is also very pervasive in scientific research such as the studies on molecules, proteins, and gene graphs. Capturing useful information from such complex topology is a crucial challenge. For instance, the 3D topology of a drug molecule plays a vital role in drug analysis. Hence, adopting trustworthy graph models to process these data could bring breakthroughs to scientific problems in critical areas. I have conducted works in explaining the predictions on molecule graphs with prototype molecules which capture key patterns of molecules with the target property [10]. These explanations could assist researchers in designing chemical compounds. In addition, an explainable GNN that aligns biomedical prior knowledge with the model logic is developed. The preprint is in submission. In the future, I plan to further collaborate with researchers in biomedical, chemistry, physics, and other fields. Leveraging the intersection of these disciplines with my expertise in trustworthy GNNs, I envision significant contributions to pivotal areas such as drug discovery and personalized medicine.

Domain-Specific Graph Foundation Model Learning. Recent breakthroughs in large language models demonstrate the power of large foundation models trained on massive data. Foundation models trained on large-scale image/text data can consistently enhance the performance. Inspired by this success, learning graph foundation models on vast graph data is promising to attain a level of expertise ready for the deployment in production environments. However, different from images and text, graph-structured data comes from various domains which exhibit distinct distributions. The knowledge learned from these distinct distributions is often non-transferable. For example, a traffic network would not be able to benefit from learning about proteins. Therefore, graph foundation models should be constructed in a domain-specific way. In particular, there are three major research questions to be answered for domain-specific graph foundation model learning.

- *What architectures should the graph foundation model adopt for different domains?* Given that data distributions vary across domains, the key challenge of this research question would be developing specialized architectures to capture essential topological patterns. I already have designs of GNN for heterophilic graphs [12] and spatial-temporal data such as energy network [2, 13]. These preliminary works can guide me to complete the architecture with large size, enabling fully leverage of the massive data in corresponding domains.Apart from the aforementioned areas, my focus will broaden to cover

the domain-specialized neural architectures for modeling the 3D topology of molecules and proteins.

- *What tasks should we deploy in training to encode domain knowledge in the graph foundation model?* I have already done several works can be referred for the investigation of tasks for graph foundation model training. Firstly, I have investigated the contrastive learning task on graph-structured data [8], which is a general pre-training task. A generative task is employed in [2] to train a model on around 500 GB power grid data for anomaly detection. With this experience, I plan to explore pre-training tasks for graph foundation models for more critical domains.
- *How to ensure the trustworthiness of the graph foundation model?* Given the wide applicability of foundational models across tasks, ensuring their trustworthiness is paramount. For the robustness aspect, I've already contributed to certifying robustness in graph contrastive learning as a preliminary work [14]. In the future, I plan to thoroughly investigate the four dimensions of trustworthiness, i.e., fairness, robustness, privacy and explainability, regarding graph foundation models.

References

[1] E. Dai, C. Aggarwal, and S. Wang. Nrgnn: Learning a label noise resistant graph neural network on sparsely and noisily labeled graphs. In *SIGKDD*, pages 227–236, 2021.

[2] E. Dai and J. Chen. Graph-augmented normalizing flows for anomaly detection of multiple time series. In *ICLR*, 2022.

[3] E. Dai, L. Cui, Z. Wang, X. Tang, Y. Wang, M. Cheng, B. Yin, and S. Wang. A unified framework of graph information bottleneck for robustness and membership privacy. *SIGKDD*, 2023.

[4] E. Dai, W. Jin, H. Liu, and S. Wang. Towards robust graph neural networks for noisy graphs with sparse labels. In *WSDM*, pages 181–191, 2022.

[5] E. Dai, M. Lin, X. Zhang, and S. Wang. Unnoticeable backdoor attacks on graph neural networks. In *WWW*, pages 2263–2273, 2023.

[6] E. Dai, Y. Sun, and S. Wang. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *ICWSM*, volume 14, pages 853–862, 2020.

[7] E. Dai and S. Wang. Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In *WSDM*, pages 680–688, 2021.

[8] E. Dai and S. Wang. Towards self-explainable graph neural network. In *CIKM*, pages 302–311, 2021.

[9] E. Dai and S. Wang. Learning fair graph neural networks with limited and private sensitive attribute information. *IEEE TKDE*, 2022.

[10] E. Dai and S. Wang. Towards prototype-based self-explainable graph neural network. *arXiv preprint arXiv:2210.01974*, 2022.

[11] E. Dai, T. Zhao, H. Zhu, J. Xu, Z. Guo, H. Liu, J. Tang, and S. Wang. A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability. *arXiv preprint arXiv:2204.08570*, 2022.

[12] E. Dai, S. Zhou, Z. Guo, and S. Wang. Label-wise graph convolutional network for heterophilic graphs. In *Learning on Graphs Conference*, pages 26–1. PMLR, 2022.

[13] Y. Hu, X. Cheng, S. Wang, J. Chen, T. Zhao, and E. Dai. Times series forecasting for urban building energy consumption based on graph convolutional network. *Applied Energy*, 307:118231, 2022.

[14] M. Lin, T. Xiao, E. Dai, X. Zhang, and S. Wang. Certifiably robust graph contrastive learning. In *NeurIPS*, 2023.

[15] T. Zhao, E. Dai, K. Shu, and S. Wang. Towards fair classifiers without sensitive attributes: Exploring biases in related features. In *WSDM*, pages 1433–1442, 2022.

[16] H. Zhu, E. Dai, H. Liu, and S. Wang. Learning fair models without sensitive attributes: A generative approach. *Neurocomputing*, page 126841, 2023.