Research Statement

Tyler Derr
http://www.TylerDerr.com

My research focuses on the general area of data science/mining. As we rapidly push forward into the Information Age, the rate at which data is produced and collected has created an unprecedented demand for novel methods to both effectively and efficiently utilize this data for extracting insightful patterns. These patterns can then be used to understand the past, make predictions about the future, and ultimately take actionable steps towards improving our society. My research strives to help meet this demand; due to the fact that much of today's big data can be represented as graphs, my emphasis has primarily been on harnessing this natural structure of data through network analysis. More specifically, my research spans across the following three main areas: (1) **signed network analysis**, i.e., network analysis with negative links, where positive links represent friends/trust and negative links represent foes/distrust; (2) **deep learning on graphs**, which attempts to harness deep learning for graph structured data; and (3) **data science for social good**, where my work has spanned across a wide variety of domains such as Education, Health Informatics, and Political Science.

While the outcomes of applying data science towards social good can be very fruitful, there are many fundamental challenges for each of my research directions. For example, when handling social networks consisting of both positive and negative links (i.e., signed networks), there are two overarching issues: (i) negative links exhibit distinctly different properties from positive links such as transitivity and reciprocity [8] and (ii) the fundamental social principles and theories of signed networks are substantially different from those of unsigned networks (i.e., networks only consisting of positive links). One direction in applying deep learning to graph data is that of graph neural networks (GNNs), where we face two inherent challenges. The first, (i) scalability of GNNs, since unlike traditional deep neural networks, the samples in graph structured data are inherently linked (i.e., not independent) and node/graph representations are learned through recursive expansion of node neighborhoods across the GNN layers. The second, (ii) it is non-trivial on how to utilize domain specific knowledge of certain graph types when attempting to use and further improve upon traditional GNNs for graphs that have additional complexities, such as signed networks. However, in my research I embrace these challenges and use them as opportunities to continue pushing the frontier in these domains. For example, in signed networks, I have developed multiple principled methodologies for measuring, modeling, and mining these networks while successfully incorporating negative links through signed network social theories, such as balance theory. For example, the development of deep learning approaches for graph data through GNNs specifically designed for signed networks to further improve the performance in learning node representations that can be utilized for a plethora of downstream tasks (e.g., predicting an unknown link sign between nodes).

The novelty and innovation of my research is apparent through the numerous prestigious awards it has received, such as the **Best Student Poster Award** of SDM2019; **"People's Choice" Award** for the 3 Minute Thesis Competition at Michigan State University (MSU); **2nd Prize** at University of Michigan's Postdoctoral Symposium; my advisor receiving his **NSF CAREER Award based on my research** in social network analysis with negative links; and **Student Travel Awards** to CIKM2019, SDM2019, ICDM2018, CIKM2018, KDD2017, and SDM2017. My research is also of high interest to the scientific community as I will present a tutorial at AAAI2020 (which contains some of my work) and my work has been incorporated into PyTorch Geometric, which is a geometric deep learning extension for Facebook's PyTorch. In addition, my research was featured in MSU's Institute for Cyber-Enabled Research (ICER) Student Highlights, while also having the British technology news website, The Register, write an article covering my work.

Furthermore, my research goal of utilizing data science towards having a positive impact on society has led to my research being quite interdisciplinary. I've had both the pleasure and privilege of collaborating with experts from other disciplines such as Psychology (e.g., Dr. Kenneth Frank) and Social Science (e.g., Dr. Russell Bernard), while also in my own domain of Network Science (e.g., Dr. Jiejun Xu) and Data Mining (e.g., Dr. Charu Aggarwal).

Research Contributions

As previously mentioned, my three research directions can be summarized as (1) signed network analysis; (2) deep learning on graphs; and (3) data science for social good. Furthermore, (1) and (2) can be seen as more fundamental research where I address the unique problem challenges through the development of novel methodologies, whereas (3) is more of a practical and applied data science research direction with emphasis on solving problems that are likely to have a more direct impact on our society; with many of my projects spanning across one or more of the above directions.

Signed Network Analysis

Most existing network analysis research has focused on unsigned networks (or networks with only positive links). However, in many real-world social systems, relations between two nodes can be represented as signed networks with positive and negative links, where negative links can denote their foes (e.g., Slashdot), distrust (e.g., Epinions), "unfriended" and "unfollowed" friends (e.g., Facebook and Twitter), and blocked users (e.g., Snapchat and WeChat). The introduction of negative links in signed networks not only increases the complexity of the representation but also poses tremendous challenges for traditional unsigned network analysis, particularly across the three foundational areas of modeling, measuring and mining. Hence, signed network analysis requires innovative and dedicated efforts. Therefore, I have conducted a comprehensive and systematic investigation in signed network analysis in terms of modeling [2, 6], measuring [8, 10] and mining [3, 4, 5, 7, 17].

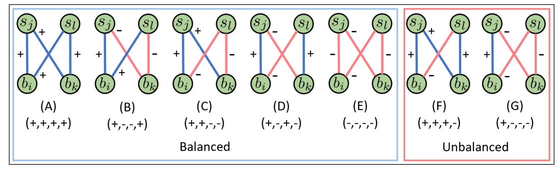


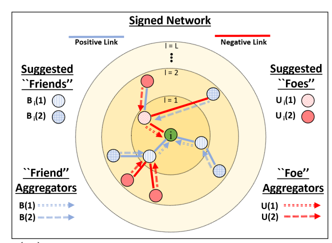
Figure 1: Signed Butterfly Isomorphism Classes.

Most existing unsigned network models fail to model unique properties of signed networks. Thus, I introduced the first signed generative model [6], which can preserve certain important properties of signed networks such as the degree distribution, the ratio of positive links and the ratio of balanced triangles. Furthermore, I have extended and validated balance theory to signed bipartite networks and verified its usefulness in the development of multiple dedicated link sign prediction methods showing superior performance when utilizing new network structures I developed - signed butterflies [2]. Figure 1 visualizes the structures that adhere to the social theory (i.e., balanced) and those that do not (i.e., unbalanced). As an example, these structures could model the relations between two buyers and two sellers online (e.g., Amazon Marketplace). Meanwhile, I introduced a variety of local and global measurements [8], which allow us to measure node relevance with negative links and have demonstrated that signed relevance measurements can measure relevance remarkably better. In addition, I have utilized a deep learning approach to merge the two fundamental signed social theories of balance and status into a signed centrality measurement that can better differentiate between "famous" and "infamous" users online [10]. Simultaneously I have also developed multiple application and mining advancements using signed networks, such as analyzing the underlying social relations behind congressional votes [5], state of the art signed network embedding by pairing signed social theories with GNNs [4], predicting congressional votes where some of the influential features are extracted from the signed bipartite network formed from the representative and the bills they've previously voted on [3], and using cross-domain techniques to jointly learn an improved model for predicting the polarity of both user links and interactions [7].

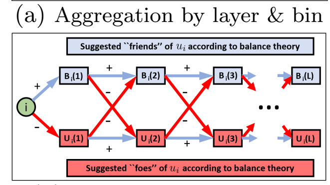
My work in signed networks has largely enriched existing research and opened the way for a new chapter of network analysis. Thanks to my breakthrough leadership in this field, I have recently started receiving offers of collaboration from those seeking my guidance on signed network related research. Furthermore, my advisor's NSF CAREER Award was based on my research in network analysis with negative links.

Deep Learning on Graphs

Recently much of the big data that is being created inherently has an underlying structure (such as social networks, physical systems, and knowledge graphs). Thus, a growing trend is to try and harness this structure by representing the data in the form of graphs, so that more meaningful analysis and predictions can be made from the data. Simultaneously, graph neural networks (GNNs), which are adaptations from the classical deep neural networks (DNNs) for graph structured data, have seen rapid development due to their utilization for learning node/graph representations/embeddings that have been shown to obtain state of the art performance across many applications.



(a) Aggregation by layer & bin



(b) Paths for binning nodes

Figure 2: Signed Graph Convolutional Networks.

I have been working on GNNs since 2017, while they were still in the early stages of development. At the time, the GNN model was unable to handle different types of graphs outside of the traditional simple graphs. Thus, especially with the ever growing popularity of online social media (which inherently allows for the creation of implicit and/or explicit negative relations), I proposed a new kind of principled model for signed graphs. As seen in Figure 2b I first defined the notion of binning a node's local neighborhood, at various path distances, where each bin is either balance or unbalanced, based on whether balance theory would suggest them as a friend or foe, respectively. Then, based on this categorization, as seen in Figure 2a the aggregation process happens according to the binned neighborhoods where they are then merged according to balance theory, which resulted in a state of the art signed network embedding methodology.

More recently, with researchers wanting to extend the application of GNNs to larger real-world graphs, there appeared the need for sophisticated methods of applying mini-batch style training, and although some methods had been proposed they were rather ad-hoc. Thus, I developed Epidemic Graph Convolutional Networks [1]. It uses the first principled sampling methods combining both local and global sampling, which are constructed based on both network diffusion and epidemic models, respectively, and has led to increased performance on node classification.

I have also developed the first advanced methodology for recommendations in customer-to-customer ecommerce (e.g., eBay and Amazon Marketplace) utilizing a deep learning approach on the heterogenous graph consisting of buyers and items, along with the novel introduction of allowing numerous sellers for a given item [16]; developed an adversarial social recommendation framework that adopts a bidirectional mapping to transfer users' information between the user-user social graph domain and the user-item graph domain [11]; a GNN based approach that uses zero-shot learning and a knowledge graph to help alleviate the cold-start item recommendation problem [20]; opening new doors for the field of network alignment by introducing an adversarial based method that utilizes generative adversarial networks (GANs) to minimize the global graph embedding distributions followed by an efficient node alignment stage achieving state of the art performance in aligning real world networks [9]; pushing the frontier of deep learning interpretability through a comprehensive analysis of the decision boundary via an two-stage adversarial based methodology finding realistic instances along the decision boundary [19]; and designing a novel reinforcement learning based attack strategy using edge rewiring, having theoretical guarantees that the rewiring is limited to only performing small perturbations on the graph, and showing the vulnerabilities of current GNN models [22].

Data Science for Social Good

After having strengthened my technical abilities through working on fundamental research in signed network analysis and deep learning on graphs, I have recently started to focus in the direction of applying data science techniques for social good. More specifically, my focus has primarily been in seeking out interesting problems that can either directly have an impact on society in a positive way, providing an insight towards future work I plan to pursue in these directions, or leading to caution against possible harm. This area of my research is inherently interdisciplinary by nature and I've truly enjoyed pursuing these practical and applied research efforts across the domains of Education, Health Informatics, and Political Science.

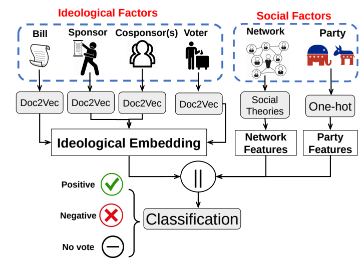


Figure 3: Multi-factor Congressional Vote Prediction.

In the direction of Education, I am currently an active member of the Teachers in Social Media (TISM) project at MSU (`http://www.teachersinsocialmedia.com`). In TISM we focuses on applying data science techniques towards improving education research with the addition of social media, since "Social media is a reflection of how teachers view themselves as teachers, how they view their practice, and how they view their pedagogical choices or aspirations." as stated by Dr. Kaitlin Torphy (lead researcher and founder of TISM). I have recently published a work pioneering this new direction [12]. More specifically, it provides a comprehensive roadmap that was specifically designed to guide education researchers using tips/tricks for harnessing online social media in their research, discussing everything from interdisciplinary team-building techniques to introducing the technical fundamentals of popular data science techniques through examples.

In the direction of Political Science, I first became interested when seeing how influential data science can be in both understanding and impacting election and the outcomes (especially in terms of policy) that permeate throughout our society as a result. Hence, I harnessed my developed methods from signed bipartite network analysis [2] to investigate the balance and properties of the signed bipartite network formed from the U.S. Congress representatives and the bills they have voted on, where their votes represent the signed edges [5]. Then, based on these findings, I developed the first thorough congressional vote prediction framework (seen in Figure 3) incorporating both ideological and social factors. My framework can not only achieve excellent prediction accuracy, but also provide an understanding as to which features are contributing/influencing a representative's vote the most [3]. This has been very well received and resulted in winning a "People's Choice" Award and 2nd Prize when presenting this work.

Furthermore, I've developed an effective method for the prediction and understanding of weight-loss through the use of online social media [13] and developed a novel reinforcement learning based reverse dialogue generation model that is the first to show the dark side of state of the art deep neural chit-chat dialogue systems being susceptible to manipulation [14, 21]. The latter has essentially shown input utterances can be automatically generated that can trick a neural dialogue model to utter any particular desired response.

Future Research Directions

As previously stated, my long-term research goal is primarily focused on effectively and efficiently utilizing the plethora of data currently being generating to both develop new and apply existing data science methodologies towards improving our society. More specifically, my short-term plan is to remain focused in solving some of the many remaining challenging research directions related to my current three research directions, in addition to focusing on security and privacy in machine learning. However, my long-term plan is to remain flexible keeping an adaptive research agenda and staying abreast to interesting challenges that need to be overcome across a wide variety of domains. This will allow for the expansion of my collaboration network, while simultaneously providing opportunities for the procurement of funding from various agencies and industries. Therefore, below is the summarization of my currently known research directions.

Network Analysis. Although previously I had mostly focused on signed network analysis, here I seek to expand my direction into other graph types, such as heterogeneous and dynamic/streaming. I want to balance my advanced knowledge in signed network analysis to continue efforts in that domain with exploration into new directions. More specifically in signed network analysis, I plan to extend and investigate the applicability of using strong triadic closure to signed networks, then seek to utilize the new developed theory towards the construction of a GNN approach to infer tie strength using user features and advanced attention mechanisms. Further directions include negative link prediction by using domain adaptation and transfer learning to apply knowledge gained from signed networks to predict negative links in unsigned networks; and providing the first comprehensive investigation of negative links in popular social media (e.g., Twitter), which can be modeled in various ways (e.g., an "unfollowing" or blocked user), and the comparison of why/how negative links form online versus offline (i.e., in the physical world).

Deep Learning on Graphs The field of deep learning on graphs has become ubiquitous and due to its ability to advance so many domains of research, I plan to continue efforts in this direction at both fundamental and application levels. Similar to traditional deep models, GNNs lack interpretability. Without understanding and verifying the inner working mechanisms, GNNs cannot be fully trusted. Thus, to allow their use in applications pertaining to fairness and safety, I will work towards the development of novel innovations for explaining GNNs. This will include example-level and model-level methods, which will explain predictions of individual examples and provide an understanding of the feature importances in the model, respectively; extending my existing work on explaining the behavior of traditional deep neural networks [18, 19]. Also, in the direction of network alignment, there are many uses for deep learning extending my previous work into domains such as aligning multiple, heterogeneous, attributed, etc. networks.

Interdisciplinary Research using Data Science for Social Good One of the keystones of good researchers is their ability to collaborate, especially having successful engagement in interdisciplinary research, since these types of endeavours have historically led to many break-through advancements. Furthermore, given the fact that my expertise in data science can be applicable in almost every research domain, my long term goal is towards building fruitful interdisciplinary collaborations with specific emphasis on topics for social good. I first plan to utilize my gained experiences in Political Science, Health Informatics, and Social Sciences in general. For instance, continued efforts in the political domain will first focus on the discrepancy in terms of sentiment and overall discussion of political candidates between the online social community and the potentially biased news outlets. In pioneering efforts for computational health, although the obesity epidemic has been thoroughly studied (especially through personal connections in the physical world), I plan to pursue if similar effects are taking place for weight loss, specifically in online applications, where I have already acquired data from one of the world's most popular weight loss applications. In terms of education, I plan to continue my collaboration with the TISM project at MSU.

Security and Privacy in Machine Learning Although machine learning has been able to transform and improve our lives in many ways, recently the research community has been discovering a wide range of ethical concerns in its usage. Hence, in an effort of wanting to still be able to pair machine learning with both the data of today and tomorrow, I plan to focus heavily in the direction of security and privacy in machine learning. For example, in GNNs, although they might be able to obtain increased performance for recommendation of friends, items, pages, etc., to online users, this could be at a cost of being susceptible to attack. More specifically, a user could become a victim of a targeted attack; if a GNN is utilized for prediction the attacker could infer information about them. Thus, I plan to first show these vulnerabilities and simultaneously (to avoid any exploitation in the meantime) propose robust deep learning models that have theoretical privacy guarantees (e.g., differential privacy in GNNs). Another direction is ensuring privacy to avoid graph denonymization, since a growing trend is to share data across online platforms (e.g., from Facebook) that utilize cross-domain recommender systems [15] for improved performance, but while theoretically proving these systems are securely sharing our data. Furthermore, my work on chatbot manipulation and GNN rewiring attacks have paved the way into other fruitful security directions in machine learning. In the recent words of Facebook's CEO Mark Zuckerberg, "The future is private."

References

[1] **T. Derr**, Y. Ma, W. Fan, X. Liu, C. Aggarwal, and J. Tang. Epidemic graph convolutional network. In *WSDM*, 2020.

[2] **T. Derr**, C. Johnson, Y. Chang, and J. Tang. Balance in signed bipartite networks. In *CIKM*, 2019.

[3] **T. Derr***, H. Karimi*, A. Brookhouse, and J. Tang. Multi-factor congressional vote prediction. In *ASONAM*, 2019.

[4] **T. Derr**, Y. Ma, and J. Tang. Signed graph convolutional networks. In *ICDM*, 2018.

[5] **T. Derr**, and J. Tang. Congressional vote analysis using signed networks. In *ICDMW*, 2018.

[6] **T. Derr**, C. Aggarwal, and J. Tang. Signed network modeling based on structural balance theory. In *CIKM*, 2018.

[7] **T. Derr**, Z. Wang, and J. Tang. Opinions power opinions: joint link and interaction polarity predictions in signed networks. In *ASONAM*, 2018.

[8] **T. Derr**, C. Wang, S. Wang, and J. Tang. Signed node relevance measurements. In *MLG*, 2018.

[9] **T. Derr**, H. Karimi, X. Liu, J. Xu, and J. Tang. Deep adversarial network alignment. In *arXiv*, 2019. (submitted to AAAI 2020)

[10] **T. Derr**, Y .Chang, C. Aggarwal, and J. Tang. Deep centrality measurement for signed networks. In *arXiv*, 2019. (submitted to SDM 2020)

[11] W. Fan, **T. Derr**, Y. Ma, Q. Li, J. Tang, and J .Wang. Deep adversarial social recommendation. In *IJCAI*, 2019.

[12] H. Karimi, **T. Derr**, K. Torphy, K. Frank, and J. Tang. A roadmap for incorporating online social media in educational research. In *TCRecord*, 2019.

[13] Z. Wang, **T. Derr**, D. Yin, and J. Tang. Understanding and predicting weight loss with mobile social networking data. In *CIKM*, 2017.

[14] H. Liu, **T. Derr**, Z. Liu, and J. Tang. Say what I want: towards the dark side of neural dialogue models. In *arXiv*, 2019. (submitted to SDM 2020)

[15] J. Dacon, **T. Derr**, and J. Tang. Cross-domain recommendation systems: a clearer view. In *arXiv*, 2019. (submitting to CSUR)

[16] W. Fan, **T. Derr**, Y. Ma, J. Wang, J. Tang, Q. Li. Deep customer-to-customer recommendations. In *arXiv* , 2019. (submitted to AAAI 2020)

[17] A. Javari, **T. Derr**, P. Esmalian, J. Tang, and K. Chang. ROSE: role-based signed network embedding. 2019. (submitted to WWW 2020)

[18] H. Karimi, **T. Derr**, and J. Tang. Explaining the behavior of deep neural networks through the lens of decision boundary In *arXiv*, 2019. (submitted to WWW 2020)

[19] H. Karimi, **T. Derr**, and J. Tang. Characterizing the decision boundary of deep neural networks. In *arXiv* , 2019. (submitted to AAAI 2020)

[20] W. Fan, Y. Ma, **T. Derr**, J. Wang, Y. Gu, S. Wang, D. Yin, J. Tang, Q. Li. Zero-shot item recommendations with knowledge graph. In *arXiv*, 2019. (submitted to AAAI 2020)

[21] H. Liu, Z. Wang, **T. Derr**, Z. Liu, and J. Tang. Chat as expected: manipulating black-box neural dialogue models. In *arXiv* , 2019. (submitted to AAAI 2020)

[22] Y. Ma, S. Wang, **T. Derr**, L. Wu, and J. Tang. Attacking Graph Convolutional Networks via Rewiring. In *arXiv*, 2019. (submitted to ICLR 2020)