1. Topology methods on heterogeneous graphs: Mostly used for homogeneous graphs, are there any work to adapt them to heterogeneous graphs? If so, what specific adaptations they proposed. If not, can you find out the reasons?
    a. https://arxiv.org/pdf/2104.02478.pdf
        i. Uses unsupervised methods to learn graph topological features.
        ii. Tested on five public datasets
            1. Cora, Citeseer, PubMed, Amazon Photo, and Amazon Computer
        iii. For undirected graphs
2. SEAL
    a. https://proceedings.neurips.cc/paper_files/paper/2018/file/53f0d7c537d99b3824f0f99d62ea2428-Paper.pdf
        i. Heuristic methods – Computes a heuristic node based on scores of the likelihoods of links
        ii. Semi-supervised learning technique
        iii. Can be applied to labeled data such as the MAG dataset.
        iv. Specific to the environment it is being applied to.
        v. Uses a self-training followed by fine-tuning process
3. Does the edge direction in a graph affect the model prediction on a heterogeneous graph?
    a. https://arxiv.org/abs/2011.14867
        i. This is a survey of popular state-of-the-art heterogeneous graph embeddings.
        ii. Focuses primarily on ecommerce and cybersecurity.
        iii. Challenges with changing edge prediction
            1. Preserving the structure of the graph
            2. Capturing properties of the graph
            3. Struggles with deep learning on heterogeneous graph
            4. Finding a reliable way to do so
4. Temporal Graph Learning Augmentation
    a. https://arxiv.org/pdf/2211.01214.pdf
        i. Emphasizes the point that many graphs are continuously changing/growing with time.
            1. Dynamic graphs
        ii. Looks for changes in real-world relationships overtime
        iii. Implemented for both link prediction and node classification
            1. Can be applied to the MAG dataset?
    b. https://proceedings.neurips.cc/paper_files/paper/2021/file/0b0b0994d12ad343511adfbfc364256e-Paper.pdf
    c. https://arxiv.org/pdf/2006.10637.pdf
        i. Both of these papers explore the importance of analyzing the changes in the graph over time.

1. Similar to number 4.

ii. I am particularly interested in further exploring this and its applications for data augmentation.

The most intriguing area discussed in these papers is the analysis of the graphs as they change overtime. This is an area that I would be interested in further exploring and potentially finding a way to use it to augment the dataset. For example, we could track the changes in the graph every year (such as the changes in the edges/nodes), find a node that was previously predicted incorrectly and then was later correctly predicted, and analyze the changes. One of the primary changes to check for could be the change in the number of citation edges and where these edges come from, since this is the only edge that will actively be added to it as the year's advance. We could also potentially analyze the topology of the area around each of these nodes that were previously incorrectly predicted, but then were accurately predicted.

This project would expand on the work done by a few of the paper's above (the ones on temporal graph learning). This would be specifically focused on the performance of the SeHGNN model and the MAG dataset and how it evolves/improves as more data is given to it over time. We can review the changes to the graph over the years and find trends in what leads to improvements in the model's accuracy.