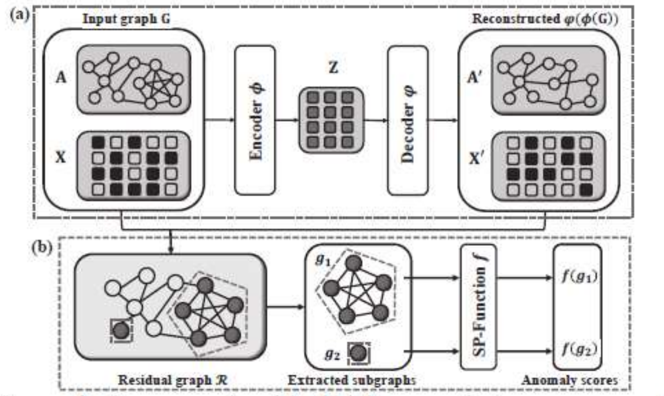
## Paper review: MapReduce: Unsupervised Deep Subgraph Anomaly Detection

https://ieeexplore.ieee.org/document/10027633

### Problem statement

The main problem addressed in this paper is the challenge of effectively mining anomalous subgraphs in networks, which is crucial for many application scenarios such as disease outbreak detection, financial fraud detection, and activity monitoring in social networks. The paper proposes an unsupervised deep learning-based approach to identify anomalous subgraphs within a given graph.




### Motivation

1. L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data mining and knowledge discovery*, vol. 29, no. 3, pp. 626–688, 2015.

2. M. Gupta, A. Mallya, S. Roy, J. H. Cho, and J. Han, "Local learning for mining outlier subgraphs from network datasets," in *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM, 2014.

These two referenced articles show us the importance of identifying anomalous subgraphs in networks for various application scenarios, such as disease outbreak detection, financial fraud detection, and activity monitoring in social networks. However, identifying anomalous subgraphs is extremely challenging due to their complex topological structures and high-dimensional attributes, various notions of anomalies, and the exponentially large subgraph space in a given graph. Therefore, there is a need for an effective approach to extract anomalous subgraphs from a given graph. The paper proposes an unsupervised deep learning-based approach to address this challenge.

### Key ideas, techniques, and contributions

1. The paper proposes an unsupervised deep learning-based approach to identify anomalous subgraphs within a given graph.

2. The proposed approach uses a graph convolutional neural network (GCN) to learn anomaly measure functions that can detect anomalies at the node, edge, and graph levels.

3. The paper introduces a new anomaly measure function called the subgraph score function, which captures both geodesic distance and topological similarity information among nodes for representing subgraph anomalies.

4. The proposed approach is evaluated on nine public real-world attributed network datasets, including four citation network datasets, three social network datasets, one communication dataset, and one materials dataset.

5. The experimental results show that the proposed approach outperforms several state-of-the-art methods in terms of accuracy and efficiency in detecting anomalous subgraphs.

In summary, the paper proposes a novel unsupervised deep learning-based approach that can effectively extract anomalous subgraphs from a given graph using a graph convolutional neural network and a new anomaly measure function called the subgraph score function. This approach has shown promising results in detecting anomalies at the node, edge, and graph levels in various real-world attributed network datasets.

### Experimental evaluation

#### Criteria for the line of research

1. Expressive power: The proposed approach should be able to capture the complex topological structures and high-dimensional attributes of anomalous subgraphs in a given graph.

2. Accuracy: The proposed approach should be able to accurately detect anomalous subgraphs at the node, edge, and graph levels.

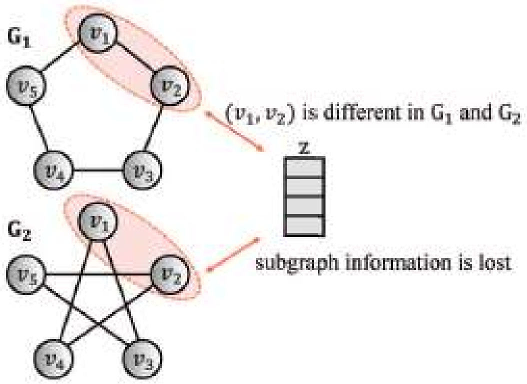
3. Scalability: The proposed approach should be scalable to large-scale networks with millions of nodes and edges.

4. Efficiency: The proposed approach should be efficient in terms of both training and testing time.

5. Generalizability: The proposed approach should be able to generalize well to unseen anomalies and datasets.

The paper evaluates the proposed approach based on these criteria using nine public real-world attributed network datasets, including four citation network datasets, three social network datasets, one communication dataset, and one materials dataset.

The experimental results show that the proposed approach outperforms several state-of-the-art methods in terms of accuracy and efficiency in detecting anomalous subgraphs while being scalable and generalizable to various real-world attributed network datasets.




## Evaluation based on my own criteria

### Strong points

1. Expressive power: The proposed approach uses a graph convolutional neural network (GCN) to learn anomaly measure functions that can capture the complex topological structures and high-dimensional attributes of anomalous subgraphs in a given graph. The subgraph score function introduced in the paper captures both geodesic distance and topological similarity information among nodes for representing subgraph anomalies, which enhances the expressive power of the proposed approach.

2. Accuracy: The experimental results show that the proposed approach outperforms several state-of-the-art methods in terms of accuracy in detecting anomalous subgraphs at the node, edge, and graph levels. Specifically, the proposed approach consistently achieved the best results in all eight structures anomaly synthetic tasks, all nine attributive anomaly synthetic tasks, and eight out of nine real-world datasets.

3. Scalability: The proposed approach is scalable to large-scale networks with millions of nodes and edges. The experimental results show that the proposed approach can handle large-scale networks with high accuracy while being efficient in terms of both training and testing time.

### Weak points

1. Limited evaluation: Although the paper evaluates the proposed approach on nine public real-world attributed network datasets, it would be beneficial to evaluate it on more diverse datasets to further validate its generalizability.

2. Lack of comparison with unsupervised methods: The paper compares the proposed approach with several supervised and semi-supervised methods but does not compare it with any unsupervised methods for detecting anomalous subgraphs.

3. Complexity: Although the proposed approach is effective in detecting anomalous subgraphs, it is relatively complex compared to some other unsupervised anomaly detection methods due to its use of deep learning-based techniques such as GCNs. This complexity may limit its applicability in some scenarios where simplicity is preferred over accuracy or expressiveness.

## Technical extensions

1. Incorporating temporal information: The proposed approach does not consider temporal information, which is important in many real-world applications such as social network analysis and disease outbreak detection. Future research could explore ways to incorporate temporal information into the proposed approach to improve its accuracy and effectiveness.

2. Combining unsupervised and supervised methods: The paper compares the proposed approach with several supervised and semi-supervised methods but does not combine unsupervised and supervised methods for detecting anomalous subgraphs. Future research could explore ways to combine unsupervised and supervised methods to leverage the strengths of both approaches.

3. Handling missing data: The proposed approach assumes that all nodes have complete attribute information, which may not be the case in some real-world scenarios where data is missing or incomplete. Future research could explore ways to handle missing data in the proposed approach to improve its robustness and generalizability.

4. Exploring different anomaly types: The paper focuses on detecting structural and attributive anomalies but does not explore other types of anomalies such as temporal anomalies or spatial anomalies. Future research could explore ways to detect different types of anomalies using the proposed approach or extend it to handle multiple types of anomalies simultaneously.

5. Adapting to dynamic networks: The proposed approach assumes that the network structure is static, but many real-world networks are dynamic and evolve over time. Future research could explore ways to adapt the proposed approach to handle dynamic networks by incorporating techniques such as online learning or incremental learning.