Midterm Report

Student name	Student ID
Tran Ngo Gia Bao	20127121
Le Quoc Trung	20127369
Lê Nguyễn Nguyên Anh	20127438
Nguyễn Quốc Thắng	20127627

Task 1:

Student name	Topic
20127121	graph classification
20127369	embedding entire graphs
20127438	community detection
20127627	frequent graph mining

Task 2:

2.1 Overview

Background

- Identifying anomalies in dynamic networks is a crucial task that has numerous applications. However, it presents significant challenges due to the intricate nature of anomalies, the absence of definitive knowledge, and the intricate and ever-changing interactions within the network. Many current approaches focus on networks with a single type of connection between vertices, but in numerous applications, the interactions between objects are diverse, resulting in multiplex networks.

Motivation

- In the realm of network analysis, the detection of anomalies plays a pivotal role in ensuring the robustness, security, and efficiency of various systems. It encompasses diverse domains such as cybersecurity, social network analysis, transportation networks, biological networks, and more. Anomalies can manifest in various forms, including but not limited to unusual patterns of behavior, unexpected connections, malicious activities, or significant deviations from the norm. Effectively identifying and characterizing these anomalies is crucial for understanding system behavior, predicting potential risks, and ultimately enabling proactive decision-making.

Contribution

- This paper have 4 main contribution:
    - A novel layer-aware node embedding approach in multiplex dynamic networks, Snapshot Encoder, which uses an attention mechanism to incorporate both temporal and structural information on different relation types.
    - ANOMULY, ageneral end-to-end unsupervised learning method for anomalous edge detection in multiplex dynamic networks, using a GRU cell to incorporate the outputs of Snapshot Encoder for different snapshots.
    - Demonstrate a new application of edge-anomaly detection in dynamic multiplex networks and present a case study on brain networks of people living with attention deficit hyperactivity disorder 2(ADHD). Our results show the effectiveness and usefulness of A NO M ULY in identifying abnormal connections

of different ROIs of the human brain. This approach could be employed as a new tool to understand abnormal brain activity that might reveal a disease or disorder.

- Conduct extensive experiments on nine real-world multiplex and simple networks. Results show the superior performance of A NO M ULY in both single-layer and multiplex networks.

Application

- Brain Network:
    - Monitoring the functional systems in the human brain is a fundamental task in neuroscience. Brain networks are represented by nodes that correspond to regions of interest (ROIs) responsible for specific functions, while edges indicate high functional correlation between two ROIs. Typically, a temporal brain network is constructed using functional magnetic resonance imaging (fMRI) to measure the statistical association between ROI functionalities over time.However, individual brain networks generated from fMRI data can be noisy and incomplete. To address this, prior approaches have relied on averaging brain networks across multiple individuals. Nevertheless, these methods overlook the intricate relationships within each individual's brain. By considering the network as a multiplex (dynamic) network, with each layer representing an individual's brain network
- Fraud Detection in Multiple Blockchain Networks:
    - To address this challenge, employing an edge anomaly detection approach within multiplex networks proves crucial in accurately detecting suspicious transactions and identifying criminal activities across various blockchain transaction networks. By considering the interconnectedness and interactions between these networks, this method enhances the precision and efficacy of anomaly detection, enabling the detection of emerging criminal patterns that may otherwise go unnoticed in isolated analyses.
    - The utilization of multiplex networks and edge anomaly detection techniques provides a powerful toolset to combat the ever-evolving tactics employed by cryptocurrency criminals. By leveraging the holistic view of interconnected blockchain networks, this approach enhances the ability to uncover and thwart illicit activities, contributing to the overall security and integrity of financial systems built on blockchain technology.

2.2. Statement

2.2.1. Input:

- Dataset can be seen as a MULTIPLEX DYNAMIC NETWORKS that change overtime

$$\mathcal{G} = \{\mathcal{G}^{(t)}\}_{t=1}^{T}$$

- Multiplex dynamic graph can be demonstrated as sequence of multiplex network snapshot. Each _snapshot_ is a static multiplex graph

$$\mathcal{G}^{(t)} = \left\{\mathcal{G}_r^{(t)}\right\}_{r=1}^{\mathcal{L}} \equiv \left(\mathcal{V}^{(t)}, \mathcal{E}^{(t)}, \mathcal{X}^{(t)}\right)$$

- Aggregate process is defined as

$$\mathrm{AGG}^{(\ell)}\left(\left\{m^{\ell}_{r_{(v\to u)}} | v \in \Lambda_r(u)\right\}\right)$$

- and the Update process of entire framework:

$$\hat{H}_r^{(t)^{(\ell)}} = \mathrm{GRU}_r\left(\tilde{H}_r^{(t)^{(\ell)}}, H_r^{(t-1)^{(\ell)}}\right)$$

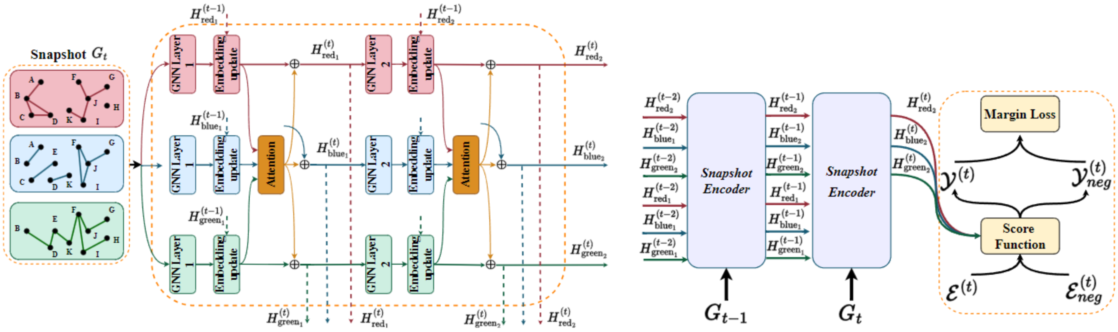
2.2.2. Output:

- Anomalous score which compute the score for each $(u,v)\in{\mathcal{E}}_{r}$ at time t

$$\varphi_r^{(t)}(u,v) = \sigma\left(\eta.\left(||\mathbf{a}\odot\mathbf{h}_{r_u}^{(t)} + \mathbf{b}\odot\mathbf{h}_{r_v}^{(t)}||_2^2 - \mu\right)\right)$$

2.2.3. Framework

- As the mention above. ***ANOMULY*** apply Snapshot Encoder which utilize ***attention mechanism*** to capture the temporal and structural information and ***GRU*** to incorporate the outputs of Snapshot Encoder of distinctive snapshots


Snapshot G_t
A
B
C
D
E
F
G
H
I
J
K
GNN Layer 1
Embedding update
Attention
GNN Layer 2
$H^{(t-1)}_{\text{red}_1}$
$H^{(t-1)}_{\text{red}_2}$
$H^{(t-1)}_{\text{blue}_1}$
$H^{(t-1)}_{\text{green}_1}$
$H^{(t)}_{\text{red}_1}$
$H^{(t)}_{\text{blue}_1}$
$H^{(t)}_{\text{red}_2}$
$H^{(t)}_{\text{blue}_2}$
$H^{(t)}_{\text{green}_2}$
$H^{(t)}_{\text{green}_1}$
$H^{(t-2)}_{\text{red}_2}$
$H^{(t-2)}_{\text{blue}_1}$
$H^{(t-2)}_{\text{green}_2}$
$H^{(t-2)}_{\text{red}_1}$
$H^{(t-2)}_{\text{blue}_1}$
$H^{(t-2)}_{\text{green}_1}$
Snapshot Encoder
$H^{(t-1)}_{\text{red}_2}$
$H^{(t-1)}_{\text{blue}_2}$
$H^{(t-1)}_{\text{green}_2}$
$H^{(t-1)}_{\text{red}_1}$
$H^{(t-1)}_{\text{blue}_1}$
$H^{(t-1)}_{\text{green}_1}$
Snapshot Encoder
G_{t-1}
G_t
Margin Loss
$\mathcal{Y}^{(t)}$
$\mathcal{Y}^{(t)}_{neg}$
Score Function
$\mathcal{E}^{(t)}$
$\mathcal{E}^{(t)}_{neg}$

(a) Architecture of the *Snapshot Encoder*

(b) ANOMULY Framework

Figure 1: Framework and design of ANOMULY model.

Challenge and Limitation

- The task of anomaly detection in dynamic networks is far from trivial. The complex nature of anomalies makes their identification challenging, as they can exhibit different temporal, spatial, and relational patterns. Moreover, the absence of ground truth knowledge further complicates the problem, as the true nature of anomalies is often unknown or ambiguous. This lack of definitive information necessitates the development of sophisticated and intelligent algorithms that can leverage available data to infer and detect deviations.

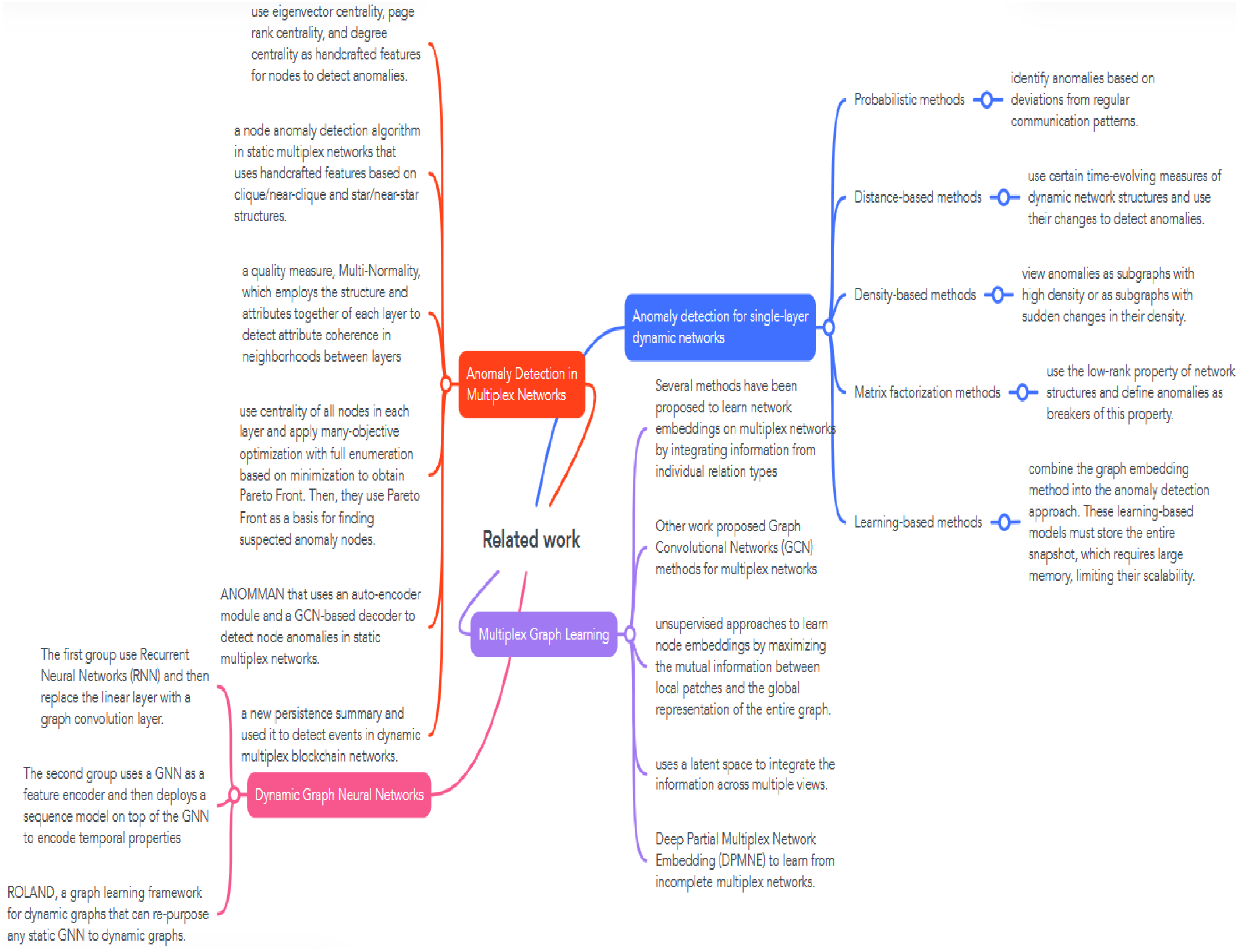
2.3 Related work and comparison to other work

Comparison table

Methods	Bitcoin		Amazon - S		DBLP - S	
Anomaly %	1%	5%	1%	5%	1%	5%
GOUTLIER	0.7143	0.7091	0.6923	0.6614	0.7108	0.6995
CM-SKETCH	0.7146	0.7015	0.7049	0.6621	0.7084	0.6877
NETWALK	0.8375	0.8367	0.7483	0.7302	0.7779	0.7590
ADDGRAPH	0.8534	0.8416	0.7872	0.7828	0.7911	0.7932
ANOMULY	**0.8707**	**0.8661**	**0.8014**	**0.7943**	**0.8129**	**0.8236**

Table 1: Performance comparison in simple networks (AUC).

Mindmap



Related work

Anomaly Detection in Multiplex Networks

use eigenvector centrality, page rank centrality, and degree centrality as handcrafted features for nodes to detect anomalies.

a node anomaly detection algorithm in static multiplex networks that uses handcrafted features based on clique/near-clique and star/near-star structures.

a quality measure, Multi-Normality, which employs the structure and attributes together of each layer to detect attribute coherence in neighborhoods between layers

use centrality of all nodes in each layer and apply many-objective optimization with full enumeration based on minimization to obtain Pareto Front. Then, they use Pareto Front as a basis for finding suspected anomaly nodes.

ANOMMAN that uses an auto-encoder module and a GCN-based decoder to detect node anomalies in static multiplex networks.

a new persistence summary and used it to detect events in dynamic multiplex blockchain networks.

Anomaly detection for single-layer dynamic networks

Probabilistic methods: identify anomalies based on deviations from regular communication patterns.

Distance-based methods: use certain time-evolving measures of dynamic network structures and use their changes to detect anomalies.

Density-based methods: view anomalies as subgraphs with high density or as subgraphs with sudden changes in their density.

Matrix factorization methods: use the low-rank property of network structures and define anomalies as breakers of this property.

Learning-based methods: combine the graph embedding method into the anomaly detection approach. These learning-based models must store the entire snapshot, which requires large memory, limiting their scalability.

Multiplex Graph Learning

Several methods have been proposed to learn network embeddings on multiplex networks by integrating information from individual relation types

Other work proposed Graph Convolutional Networks (GCN) methods for multiplex networks

unsupervised approaches to learn node embeddings by maximizing the mutual information between local patches and the global representation of the entire graph.

uses a latent space to integrate the information across multiple views.

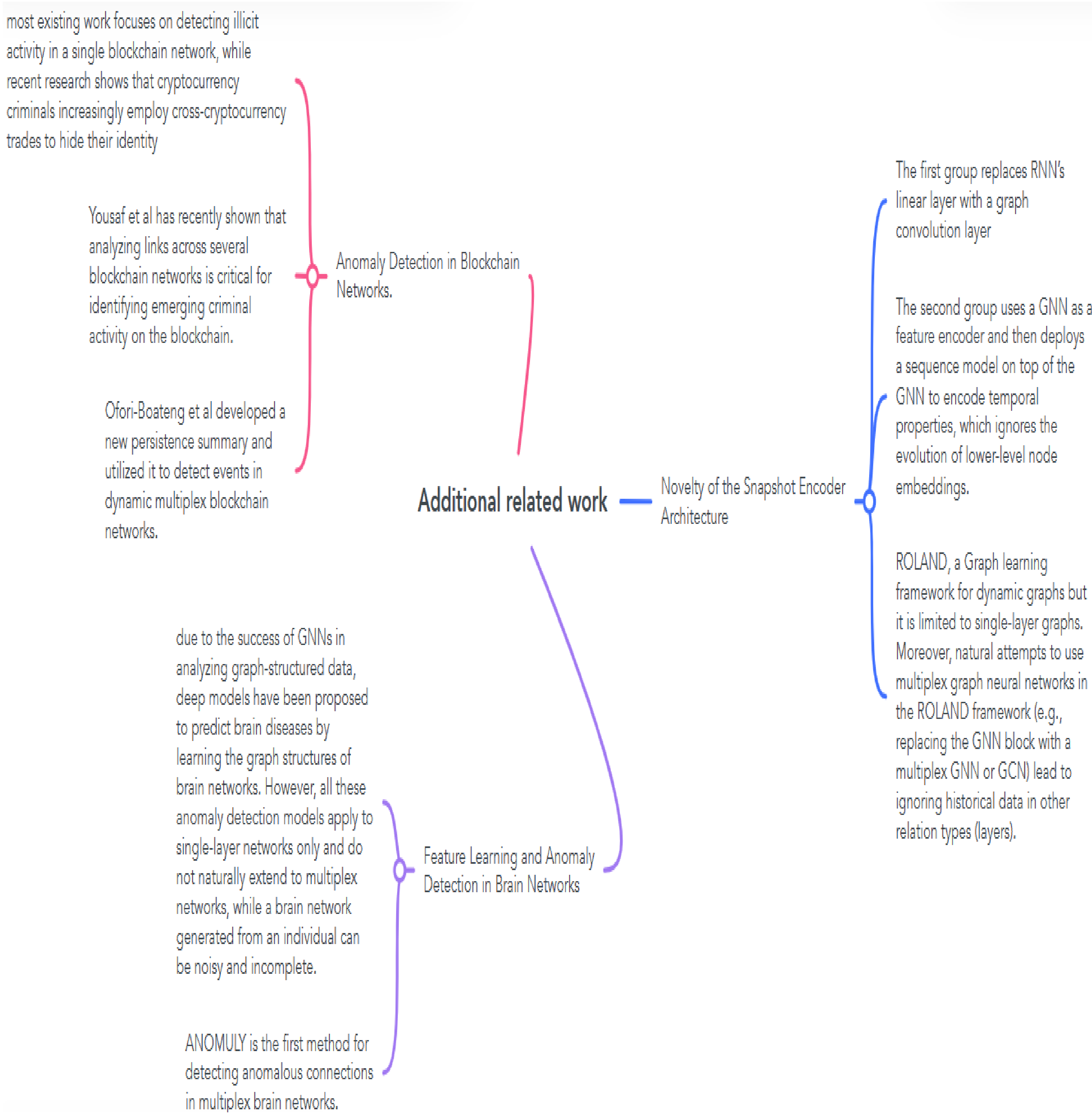
Deep Partial Multiplex Network Embedding (DPMNE) to learn from incomplete multiplex networks.

Dynamic Graph Neural Networks

The first group use Recurrent Neural Networks (RNN) and then replace the linear layer with a graph convolution layer.

The second group uses a GNN as a feature encoder and then deploys a sequence model on top of the GNN to encode temporal properties

ROLAND, a graph learning framework for dynamic graphs that can re-purpose any static GNN to dynamic graphs.



Additional related work

Anomaly Detection in Blockchain Networks.

- most existing work focuses on detecting illicit activity in a single blockchain network, while recent research shows that cryptocurrency criminals increasingly employ cross-cryptocurrency trades to hide their identity
- Yousaf et al has recently shown that analyzing links across several blockchain networks is critical for identifying emerging criminal activity on the blockchain.
- Ofori-Boateng et al developed a new persistence summary and utilized it to detect events in dynamic multiplex blockchain networks.

Novelty of the Snapshot Encoder Architecture

- The first group replaces RNN's linear layer with a graph convolution layer
- The second group uses a GNN as a feature encoder and then deploys a sequence model on top of the GNN to encode temporal properties, which ignores the evolution of lower-level node embeddings.
- ROLAND, a Graph learning framework for dynamic graphs but it is limited to single-layer graphs. Moreover, natural attempts to use multiplex graph neural networks in the ROLAND framework (e.g., replacing the GNN block with a multiplex GNN or GCN) lead to ignoring historical data in other relation types (layers).

Feature Learning and Anomaly Detection in Brain Networks

- due to the success of GNNs in analyzing graph-structured data, deep models have been proposed to predict brain diseases by learning the graph structures of brain networks. However, all these anomaly detection models apply to single-layer networks only and do not naturally extend to multiplex networks, while a brain network generated from an individual can be noisy and incomplete.
- ANOMULY is the first method for detecting anomalous connections in multiplex brain networks.