Anomalous Edge Detection in Edge Exchangeable Social Network Models

Rui Luo RUILUO@CITYU.EDU.HK
City University of Hong Kong, Kowloon Tong, Hong Kong
Cornell University, Ithaca, USA
Buddhika Nettasinghe BUDDHIKA-NETTASINGHE@UIOWA.EDU
University of Iowa, Iowa City, USA
Vikram Krishnamurthy VIKRAMK@CORNELL.EDU
Cornell University, Ithaca, USA

Editor: Harris Papadopoulos, Khuong An Nguyen, Henrik Boström and Lars Carlsson

Abstract

This paper studies detecting anomalous edges in directed graphs that model social networks. We exploit edge exchangeability as a criterion for distinguishing anomalous edges from normal edges. Then we present an anomaly detector based on conformal prediction theory; this detector has a guaranteed upper bound for false positive rate. In numerical experiments, we show that the proposed algorithm achieves superior performance to baseline methods.

Keywords: Edge Exchangeable Model, Conformal Prediction, Anomaly Detection, Social Networks.

1. Introduction

The rapid growth of online social networks raises security concerns, including coordinated spam attacks (Wagner et al., 2012), cyber-bullies (Bindu and Thilagam, 2016), and dissemination of misinformation (Nguyen et al., 2012). The problem we address is: how do we spot such activities in a network? Specifically, we represent them as anomalous edges in a graph, which show significant deviations from normal connection patterns in the network. In this regard, we aim to develop algorithms to detect anomalous edges in social networks.

A number of novel graph-based approaches have been proposed for anomaly detection in networks, which account for the correlations of objects in networks (Akoglu et al., 2015). Among the approaches focusing on detecting anomalous edges, some share a similar perspective with link prediction algorithms, using the proximity of adjacent nodes (Yu et al., 2018; Akoglu et al., 2010) or factorization of the adjacency matrix (Tong and Lin, 2011; Chakrabarti, 2004) as evidence of abnormal connections. Other approaches analyze anomalous edges by aggregating them into graph snapshots (Shin et al., 2018; Eswaran et al., 2018). One issue is that these graph- or snapshot-based approaches require high memory and often prohibitive time lag as well.

More recently, Ranshous et al. (2015); Eswaran and Faloutsos (2018); Chang et al. (2021) have proposed to detect individually surprising edges in the edge stream which have superior memory and time efficiency. However, due to the non-i.i.d. nature of graph-structured data, they fail to obtain provable guarantees for these algorithms. Bhatia et al. (2020) develops a hypothesis testing framework and has shown guarantees on the false positive rate. However, it focuses on detecting micro-clusters of edges.

In this paper, we use edge exchangeability as a criterion for detecting anomalous edges in networks. We then propose an algorithm which guarantees the false positive rate by combining edge exchangeability with conformal prediction. Conformal prediction generates set-valued predictions (the set covers the ground truth with high probability) and has been used successfully in several machine learning applications, including image classification (Shafer and Vovk, 2008) and anomalous trajectory detection (Laxhammar and Falkman, 2010). Specifically, we fit an edge exchangeable model to the network data. Then, we construct an algorithm (Conformal detector) based on conformal prediction to detect anomalous edges which break exchangeability.

Main Results and Organization:

(1) Section 2 reviews the edge exchangeable models and justifies that edge exchangeability captures the nature of many real-world networks.

(2) Section 3.1 presents the edge exchangeable model and related inference approach according to the asymmetric Dirichlet network distribution (ADND) (Williamson, 2016). Section 3.3 gives the definition of anomalous edges regarding edge exchangeability. Section 3.3 proposes the anomalous edge detection algorithm (Conformal detector) based on conformal prediction.

(3) Section 4 presents numerical results that demonstrate the performance of the proposed Conformal detector in anomalous edge detection using transaction data in Ethereum blockchain.

(4) Finally, the appendix proves that the proposed Conformal detector guarantees the upper bound of false positive rate (detecting normal edges as an anomaly).

1.1. Related Work

Previous related works can be considered under three categories:

(1) *Anomalous edge detection in graphs:* Existing anomalous edge detection methods focus on the following three aspects.

- Graph compression: Chakrabarti (2004) encoded an input graph by alternatively arranging the nodes into disjoint partitions and splitting the partitions; then spots the anomalous edges whose removal reduces the total encoding cost significantly. Tong and Lin (2011) factorized the adjacency matrix of the input graph and flag edges with high reconstruction error as anomalies. Aggarwal et al. (2011) constructed a structural connectivity model of the input graph to define anomalous edges.

- Node and subgraph similarity: Yu et al. (2018) learned a dynamic node embedding which minimizes the pairwise distance of node embeddings along each walk and computes the abnormality of edges according to the Hadamard product of node embeddings. Akoglu et al. (2010) extracted egonet-based (1-hop neighborhood around a node) features and spots anomalous edges according to abnormal pairwise correlation patterns. Shin et al. (2018) and Eswaran and Faloutsos (2018) aggregated edges into graph snapshots and detect dense subgraphs with anomalous nodes and edges.

- Edge frequency: Ranshous et al. (2016) proposed an edge probability function which scores each edge based on occurrence frequency, common neighbors and popularity of two endpoints. Eswaran et al. (2018) identified edge anomalies as bursty occurrence which connect parts of the graph which are sparsely connected. Bhatia et al. (2020) proposed a probabilistic framework to detect microcluster anomalies (suddenly arriving of groups of suspiciously similar edges). Chang et al. (2021) used a frequency-factorization technique to model the distributions of edge frequencies and flags edges with low likelihood of the observed frequency as anomaly.

(2) *Exchangeable graph models:* Vertex exchangeable random graph models assume that the distribution of a graph is invariant to the labeling of its vertices (Wang and Wong, 1987; Hoff et al., 2002; Kemp et al., 2006; Miller et al., 2009). An extensive list of vertex exchangeable random graph models can be found in (Lloyd et al., 2012). A limitation of vertex exchangeable random graph models is that they yield dense graphs (Lloyd et al., 2012). Edge exchangeable random graph models, on the other hand, are capable to produce sparse graphs and small-world behavior of real-world networks (Janson, 2018). Several edge exchangeable models (Cai et al., 2016; Crane and Dempsey, 2018; Williamson, 2016) have been proposed for modeling networks as exchangeable sequences of edges. These models assume that the distribution of a graph's edge sequence is invariant to the order of the edges (Cai et al., 2016). In addition, Caron and Fox (2017) proposed representing a network as an exchangeable point process and using completely random measures to construct sparse graphs.

(3) *Conformal prediction:* Conformal prediction algorithms provide valid measures of confidence for individual predictions made by machine learning algorithms (Gammerman and Vovk, 2007). It is based on the assumption that the samples are probabilistically exchangeable (Shafer and Vovk, 2008), which is slightly weaker than i.i.d. assumption. Different machine learning algorithms have previously been proposed and investigated for conformal prediction, e.g., the k-nearest neighbours algorithm (Vovk et al., 2005), SVM (Vovk et al., 2005), and neural networks (Papadopoulos et al., 2007). The computational efficiency of conformal prediction is considerably improved using a modification known as inductive conformal prediction (Papadopoulos, 2008), which avoids training the algorithm one time for each sample in the dataset. Conformal prediction has been used for testing exchangeability of a sequence of data (Vovk et al., 2003; Fedorova et al., 2012; Volkhonskiy et al., 2017; Cai and Koutsoukos, 2020). In addition, Laxhammar and Falkman (2010) applied conformal prediction in detecting anomalous trajectories in surveilance applications; Ishimtsev et al. (2017) proposed an anomaly detection algorithm for univariate time series by combinning k-nearest neighbors algorithm with the conformal prediction framework. Recently, (Nettasinghe et al., 2022) extended conformal prediction to the setting of a Hidden Markov Model, which greatly relaxes the i.i.d. assumption.

2. Edge Exchangeable Social Network Model

In this section, we first introduce the edge exchangeable models and their connections with the vertex exchangeable models. We then justify how these models capture some important properties of real world social networks, including sparsity and power-law degree

Table 1: Glossary of Symbols Used in This Article

Symbols	Description
G	the directed graph representing the network
V	the set of nodes
E	the set of edges
N	the number of edges in E
e_n	the edge in $E, n = 1, \cdots, N$
e_{N+1}	the new edge to be tested
u_n	the sender node of edge e_n
v_n	the receiver node of edge e_n
α_n	the non-conformity score of edge e_n
p_n	the p-value of edge e_n being normal
ϵ	the significance level
Anom_n^ϵ	indicator of whether edge e_n is anomalous or not at the significance level ϵ
A	The sender-specific topic distribution
B	The receiver-specific topic distribution
H	The base distribution for A and B
τ	concentration parameters of A and B
γ	concentration parameters of H
$\overset{d}{=}$	equality in distribution

distribution. We further illustrate via several real world examples that edge exchangeability is a useful assumption in modeling social networks, which guides the definition of anomalous edges in Section 3.3 and the anomaly detection algorithm in Section 3.3.

2.1. Edge Exchangeable Model

Consider a directed graph $G = (V, E)$, where $V = \{v_1, \cdots, v_M\}$ is the node set with M nodes and $E = \{e_1, \cdots, e_N\}$ is the edge set with N edges. A sequence of random variables is (finitely) exchangeable if their joint distribution is invariant under permutations (Niepert and Domingos, 2014). Exchangeability reflects the assumption that the variables do not depend on their indices although they may be dependent among themselves. Exchangeability[1] is a weaker assumption than iid; iid variables are automatically exchangeable (Orbanz and Teh, 2010).

This paper deals with edge exchangeable networks. We now briefly discuss vertex exchangeable networks. In graphs, the *vertex exchangeability*, defined as follows, means that the distribution of the graph is invariant to relabelings of the vertices.

Definition 1 (Vertex Exchangeability) *Consider a graph $G = (V, E)$, where $V = \{1, \cdots, M\}$ and $E = \{e_1, \cdots, e_N\} \subseteq V \times V$. Such networks can be expressed as an adjacency matrix*

[1] The simplest type of exchangeable process is sampling without replacement of an iid process (Orbanz and Teh, 2010).

$\boldsymbol{y} = (y_{ij})_{i,j \in V}$ *with*

$$y_{ij} = \begin{cases} 1, & (i,j) \in E, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Let permutation $\pi : V \to V$*. Relabeling vertices according to* π *leads to the adjacency matrix* $\boldsymbol{y}^\pi = (y_{\pi(i)\pi(j)})_{i,j \in V}$*. Then,* G *is vertex exchangeable if for every permutation* π*,* $\boldsymbol{y}^\pi \overset{d}{=} \boldsymbol{y}$*. Here,* $\overset{d}{=}$ *denotes equality in distribution, that is,*

$$P(Y = \boldsymbol{y}) = P(Y = \boldsymbol{y}^\pi), \tag{2}$$

where $Y = (Y_{ij})_{i,j \in V}$ *is a random matrix defined on* $\{0,1\}^{V \times V}$*.*

The assumption of vertex exchangeability underlies many well-known network models, including the stochastic block model (SBM) (Wang and Wong, 1987), latent space model (Hoff et al., 2002), infinite relational model (IRM) (Kemp et al., 2006), latent feature relational model (LFRM) (Miller et al., 2009) and many others. However, they fall under the framework of Aldous-Hoover representation theorem (Aldous, 1981; Hoover, 1979), which guarantees that these graphs are either empty or dense with probability one (Orbanz and Roy, 2015). Such models cannot produce sparsity in real-world graphs.

In comparison to vertex exchangeable networks, edge exchangeable networks (Cai et al., 2016; Crane and Dempsey, 2018; Williamson, 2016) considered in this paper capture important properties of large-scale social networks such as sparsity, community structure and power law degree distributions. Roughly speaking, for an edge exchangeable graph, the joint distribution of the edge sequence is invariant to the order of the edges. We define this formally as follows:

Definition 2 (Edge Exchangeability) *Consider a graph* $G = (V, E)$*, where* $V = \{1, \cdots, M\}$ *and* $E = \{e_1, \cdots, e_N\} \subseteq V \times V$*. Each edge is represented as a sender-receiver pair* $(u, v) \in V \times V$*. Let permutation* $\pi : \{1, \cdots, N\} \to \{1, \cdots, N\}$*. Then,* G *is edge exchangeable if for every permutation* π*, the edge sequence* $(e_1, \cdots, e_N)$ *is exchangeable, i.e.,*

$$(e_1, e_2, \cdots, e_N) \overset{d}{=} (e_{\pi(1)}, e_{\pi(2)}, \cdots, e_{\pi(N)}), \tag{3}$$

for every permutation π*. Here,* $\overset{d}{=}$ *denotes equality in distribution, that is,*

$$P(e_n = (u_n, v_n)_{,n=1,\cdots,N}) = P(e_n = (u_{\pi(n)}, v_{\pi(n)})_{,n=1,\cdots,N}), \tag{4}$$

for every permutation π*.*

2.2. Why Edge Exchangeability

In this subsection, we demonstrate how edge exchangeability accounts for the complex behavior of social networks. First, in situations where the focus is on the interactions between social network members, e.g., communications, collaborations, and relationships, it is reasonable to represent edges as the statistical units of the network (Crane, 2018). Second, edge exchangeability assumes that the observed edges are a representative sample of the population of all edges (Crane and Dempsey, 2018). In the phone call database example

(Crane and Dempsey, 2018), each phone call has the same probability to be sampled, leading to callers who make a large number of phone calls appearing in the database with a high frequency. In contrast, vertex exchangeability assumes that sampled vertices are representative of the population of all vertices, which neglects the different frequencies of callers in making phone calls. Another advantage is that edge exchangeable models suffice to model interactions driven by edge growth.

Although exchangeability often fails to hold for time series data due to the dependence structure in the data, exchangeability holds for networks where the distribution of edges is invariant under the addition of new edges[2]. We justify edge exchangeability in social networks via the following examples. In the first two examples, edges are sampled without replacement from a fixed distribution, and the permutation of the edges will not affect their joint distribution.

Example 1. Online discussion thread: A discussion thread gathers people interested in the same topics and their messages on such topics (e.g., sports, politics). In the thread, people have different roles such as opinion leaders, followers, and moderators, and they each have a regular pattern of participating in the discussion. If one person u's post is commented on by another person v, u will have a directed edge to v, forming an ordered pair (u, v). On the online thread, individuals develop a consistent pattern of interacting with each other. Following the definition in Crane and Dempsey (2018), we denote $\mathcal{P}$ as a set of participants in the discussion thread. Now we define a distribution $f = (f_e)_{e\in \text{fin}_2(\mathcal{P})}$ on $\text{fin}_2(\mathcal{P})$, where $\text{fin}_2(\mathcal{P})$ is the set of all ordered pairs of $\mathcal{P}$,

$$P((u,v) = e|f) = f_e, \quad e \in \text{fin}_2(\mathcal{P}), \tag{5}$$

where

$$(f_e)_{e\in \text{fin}_2(\mathcal{P})} : f_e \geq 0 \quad \text{and} \quad \sum_{e\in \text{fin}_2(\mathcal{P})} f_e = 1. \tag{6}$$

Let the edges $(u_1, v_1), \cdots, (u_n, v_n)$ be independent, identically distributed random multisets of size 2 drawn from (5). This generating mechanism results in an exchangeable edge sequence.

In addition to the above examples, (Crane and Dempsey, 2018) and (Williamson, 2016) have verified that the edge exchangeable model outperforms vertex exchangeable model counterparts (IRM, SBM) in predictive likelihood on several real world datasets.

On the other hand, anomalous edges will likely lie an abnormal distance from other edges in the population and therefore causing the edge sequence nonexchangeable. Regarding Example 3 below, a situation where the new edges are anomalous and lead to nonexchangeable edge sequence is as follows.

Example 2. (Continued from Example 1) Nonexchangeable online discussion thread: A hacker h wants to attack the thread by spreading some fake news. He either joins several (trending) existing discussions by commenting about fake news in a post or

[2] In other words, new edges are sampled without replacement from the same distribution of the existing edges.

starts a discussion by posting fake news. The first action results in him having in-edges with people in the existing discussions, e.g., (u, h); the second action results in him having out-edges with people deluded by the fake news, e.g., (h, v). Either action signifies an abnormal pattern where the edges are not representative samples of the edge population $\mathrm{fin}_2(\mathcal{P})$. Thus if we permute the order of these edges in the existing edge sequence, the resulting edge sequence is nonexchangeable.

Example 3. University email network (Williamson, 2016): A university's email network records email communications among entities in the university. The email network follows a consistent communication pattern: administrators may send out a large number of group emails but receive a relatively small number of emails–that is, they often operate in the "sender" role but not the "receiver" role; faculties in collaboration may send and receive emails between each other on a regular basis–they operate as the "sender" role and the "receiver" role both at a moderate frequency. Each email from the email network, (u, v), corresponds to an email from u to v. Denote P_U as a distribution of the sender and P_V as a distribution of the receiver. The emails $(u_1, v_1), \cdots, (u_n, v_n)$ are exchangeable as a sequence in the space of the joint distribution $P_U \times P_V$.

3. Anomalous Edge Detection

In this section, we fit the social network edges using an Asymmetric Dirichlet Network Distribution (ADND) model whose prior is an exchangeable, two-level hierarchical Dirichlet process (HDP). The model accounts for the hierarchical nature of networks, where nodes are grouped into groups that are further divided into subgroups, as well as the asymmetric role that senders and receivers play in forming edges. We use a variational inference approach to approximate the posterior distribution of the latent variables and compute the log likelihood of new edges. The anomalous edges with regard to edge exchangeability are then defined. Additionally, based on conformal prediction, we propose Algorithm 1 for detecting the anomalous edges, which has a false positive rate guarantee (Proposition 5).

3.1. Asymmetric Dirichlet Network Distribution (ADND) Model

Consider a graph $G = (V, E)$, where V is the node set and $E = \{e_1, \cdots, e_N\}$ is the edge set with N edges. We adopt the asymmetric Dirichlet network distribution (ADND) model (Williamson, 2016) to fit the network edges. Here, each edge is represented as a tuple $e_n = (u_n, v_n), n = 1, \cdots, N$, where the nodes u_n, v_n denotes a sender-receiver pair.

The fundamental concept behind the ADND model involves using Bayesian nonparametrics to define the prior and likelihood for sampling edges from a countably infinite population of nodes. Denote the size of the node population as W, the ADND model can be described using a two-level Hierarchical Dirichlet Process (HDP) as follows:

$$\begin{aligned}
&\Theta \sim \mathrm{Dir}(\boldsymbol{\eta}), \text{ where } \boldsymbol{\eta} = (\eta, \cdots, \eta) \in \mathbb{R}^{W+1}, \\
&H \sim \mathrm{DP}(\gamma\Theta), \\
&A \sim \mathrm{DP}(\tau H), \quad B \sim \mathrm{DP}(\tau H),
\end{aligned} \tag{7}$$

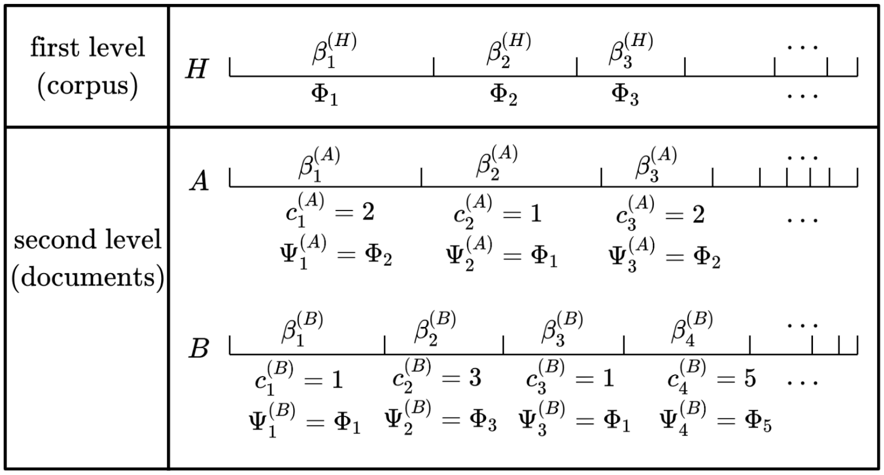

Figure 1: Illustration of the two-level stick-breaking representation of the hierarchical Dirichlet process. In the first level, $\phi_i \sim \Theta$ and $\boldsymbol{\beta}^{(H)} \sim \text{GEM}(\gamma)$; in the second level, $\boldsymbol{\beta}^{(A)}, \boldsymbol{\beta}^{(B)} \sim \text{GEM}(\tau)$, $c_t^{(A)}, c_t^{(B)} \sim \text{Categorical}(\boldsymbol{\beta}^{(H)})$ and $\psi_t^{(A)} = \phi_{c_t^{(A)}}, \psi_t^{(B)} = \phi_{c_t^{(B)}}$.

where $\gamma, \tau > 0$ are the concentration parameters of the first- and second-level Dirichlet process. The first level of the HDP, H, represents the topic distribution, where each topic is a distribution over nodes. H is itself Dirichlet process distributed – its base distribution Θ is a symmetric $(W+1)$-dimensional Dirichlet distribution with concentration parameter $\boldsymbol{\eta}$, where the $(W+1)$-th dimension is reserved for unseen nodes. The sender distribution A and the receiver distribution B are topic distributions that share the same set of topics with H, but weigh them using specific proportions $\boldsymbol{\beta}^{(A)}$ and $\boldsymbol{\beta}^{(B)}$.

We follow Wang et al. (2011)'s stick-breaking representation of the HDP in (7), which enables closed form variational updates in Section 3.2. For the corpus-level DP draw, this representation is

$$\begin{aligned} &\beta_i^{(H)'} \sim \text{Beta}(1, \gamma), \quad \beta_i^{(H)} = \beta_i^{(H)'} \prod_{l=1}^{i-1} (1 - \beta_l^{(H)'}), \\ &\phi_i \sim \Theta, \\ &H = \sum_{i=1}^{\infty} \beta_i^{(H)} \delta_{\phi_i}. \end{aligned} \tag{8}$$

Thus, H is discrete and has support at the atoms $\boldsymbol{\phi} = (\phi_i)_{i=1,\cdots,\infty}$ with weights $\boldsymbol{\beta}^{(H)} = (\beta_i^{(H)})_{i=1,\cdots,\infty}$. The distribution for $\boldsymbol{\beta}^{(H)}$ is also written as $\boldsymbol{\beta}^{(H)} \sim \text{GEM}(\gamma)$ (”GEM” stands for Griffiths, Engen and McCloskey).

To represent the document-level atoms $\boldsymbol{\psi}^{(A)} = (\psi_t^{(A)})_{t=1,\cdots,\infty}$ and $\boldsymbol{\psi}^{(B)} = (\psi_t^{(B)})_{t=1,\cdots,\infty}$, we introduce two series of indicator variables, $\mathbf{c}^{(A)} = (c_t^{(A)})_{t=1,\cdots,\infty}$ and $\mathbf{c}^{(B)} = (c_t^{(B)})_{t=1,\cdots,\infty}$, which are drawn i.i.d.,

$$c_t^{(A)}, c_t^{(B)} \sim \text{Categorical}(\boldsymbol{\beta}^{(H)}) \tag{9}$$

Then let

$$\psi_t^{(A)} = \phi_{c_t^{(A)}}, \quad \psi_t^{(B)} = \phi_{c_t^{(B)}}, \tag{10}$$

By doing this, we can avoid having to explicitly express the document atoms $\boldsymbol{\psi}^{(A)}$ and $\boldsymbol{\psi}^{(B)}$.

Given the representation in Eq. (8, 9, 10), the generative process for the n-th edge, (u_n, v_n), is as follows,

$$\begin{aligned} &z_n^{(A)} \sim \text{Categorical}(\boldsymbol{\beta}^{(A)}), \quad z_n^{(B)} \sim \text{Categorical}(\boldsymbol{\beta}^{(B)}), \\ &u_n \sim \text{Categorical}(\phi_{c^{(A)}_{z_n^{(A)}}}), \quad v_n \sim \text{Categorical}(\phi_{c^{(B)}_{z_n^{(B)}}}). \end{aligned} \tag{11}$$

The indicator variables $z_n^{(A)}, z_n^{(B)}$ select document-level topics $\psi_t^{(A)}$ and $\psi_t^{(B)}$, respectively, according to the document-level stick weights $\boldsymbol{\beta}^{(A)}$ and $\boldsymbol{\beta}^{(B)}$. $\psi_t^{(A)}, \psi_t^{(B)}$ map to one of corpus-level topics $\boldsymbol{\phi}$ through the indicators $c_t^{(A)}$ and $c_t^{(B)}$. We describe the process of drawing senders of edges as follows: for the n-th edge, first choose the sender's topic indicator $z_n^{(A)} = t$; then choose the corpus-level topic indicator $c_t^{(A)} = i$. In this way, ϕ_i is the sender's chosen topic. We have,

$$u_n | z_n^{(A)} = t, c_t^{(A)} = i, \boldsymbol{\phi} \sim \text{Categorical}(\phi_i) \tag{12}$$

Similarly, for the receiver,

$$v_n | z_n^{(B)} = t, c_t^{(B)} = i, \boldsymbol{\phi} \sim \text{Categorical}(\phi_i) \tag{13}$$

Discussion of the ADND model: As defined in Section 2.1, vertex exchangeable models such as the SBM and IRM assume a fixed node set and model the network by clustering the nodes. The edges are represented as nonzero-valued entries of the corresponding adjacency matrix. In contrast, ADND models the network as a sequence of edges, and clusters edges into a hierarchy of topics, rather than nodes. Each node can be associated with multiple topics based on its incoming and outgoing edges. Further, by placing a nonparametric prior distribution over the topics, the ADND model can incorporate previously unseen nodes and is flexible for predicting unseen edges.

3.2. Variational Inference

In the above representation of ADMD model, the observed data is $x = (u_n, v_n)_{n=1,\cdots,\infty}$, the hidden variables are $z = \{(z_n^{(A)})_{n=1,\cdots,\infty}, (z_n^{(B)})_{n=1,\cdots,\infty}, (c_t^{(A)})_{t=1,\cdots,\infty}, (c_t^{(B)})_{t=1,\cdots,\infty},$ $(\beta_t^{(A)})_{t=1,\cdots,\infty}, (\beta_t^{(B)})_{t=1,\cdots,\infty}, (\beta_i^{(H)})_{i=1,\cdots,\infty}, (\phi_i)_{i=1,\cdots,\infty}\}$. Due to the difficulty in directly computing $p(z|x)$, we use a fully factorized variational distribution for approximation,

$$\begin{aligned} &q(\boldsymbol{\beta}^{(H)'}, \boldsymbol{\beta}^{(A)'}, \boldsymbol{\beta}^{(B)'}, \boldsymbol{c}^{(A)}, \boldsymbol{c}^{(B)}, \boldsymbol{z}^{(A)}, \boldsymbol{z}^{(B)}, \boldsymbol{\phi}) \\ =&q(\boldsymbol{\beta}^{(H)'})q(\boldsymbol{\beta}^{(A)'})q(\boldsymbol{\beta}^{(B)'})q(\boldsymbol{c}^{(A)})q(\boldsymbol{c}^{(B)})q(\boldsymbol{z}^{(A)})q(\boldsymbol{z}^{(B)})q(\boldsymbol{\phi}) \end{aligned} \tag{14}$$

which further factorizes into

$$q(\boldsymbol{c}^{(A)}) = \prod_{t=1}^{K^{(A)}} q(c_t^{(A)}|\varphi_t^{(A)}), \quad q(\boldsymbol{c}^{(B)}) = \prod_{t=1}^{K^{(B)}} q(c_t^{(B)}|\varphi_t^{(B)}), \tag{15a}$$

$$q(\boldsymbol{z}^{(A)}) = \prod_{n} q(z_n^{(A)}|\zeta_n^{(A)}), \quad q(\boldsymbol{z}^{(B)}) = \prod_{n} q(z_n^{(B)}|\zeta_n^{(B)}), \tag{15b}$$

$$q(\boldsymbol{\phi}) = \prod_{i=1}^{K^{(H)}} q(\phi_i|\lambda_i), \tag{15c}$$

where the variational parameters are $\varphi_t^{(A)}$ (multinomial), $\varphi_t^{(B)}$ (multinomial), $\zeta_n^{(A)}$ (multinomial), $\zeta_n^{(B)}$ (multinomial), and λ_i (Dirichlet). We set the corpus-level topic truncation $K^{(H)}$ and the document-level truncation $K^{(A)}$ and $K^{(B)}$. $K^{(H)}$ is usually much bigger than $K^{(A)}$ or $K^{(B)}$ because in practice, the number of topics needed for each document is far fewer than that required for the corpus. The factorized forms of the corpus-level stick proportions and bottom-level stick proportions are

$$q(\boldsymbol{\beta}^{(H)'}) = \prod_{i=1}^{K^{(H)}} q(\beta_i^{(H)'}|a_i^{(H)}, b_i^{(H)}), \tag{16a}$$

$$q(\boldsymbol{\beta}^{(A)'}) = \prod_{t=1}^{K^{(A)}} q(\beta_t^{(A)'}|a_t^{(A)}, b_t^{(A)}), \quad q(\boldsymbol{\beta}^{(B)'}) = \prod_{t=1}^{K^{(B)}} q(\beta_t^{(B)'}|a_t^{(B)}, b_t^{(B)}), \tag{16b}$$

where $(a_i^{(H)}, b_i^{(H)})$, $(a_t^{(A)}, b_t^{(A)})$, and $(a_t^{(B)}, b_t^{(B)})$ are parameters of beta distributions.

The evidence lower bound (ELBO) in variational theory (Blei et al., 2017) lower bounds the marginal log-likelihood of the observed data. ELBO is defined as follows.

$$\begin{aligned}
&\mathbb{E}_q[\log\frac{p(x,z)}{q(z)}] = \mathbb{E}_q[\log(p(x,z)] + H(q) \\
&= \mathbb{E}_q\Big[\log p(\mathbf{u}|\mathbf{c}^{(A)}, \mathbf{z}^{(A)}, \phi)p(\mathbf{c}^{(A)}|\boldsymbol{\beta}^{(H)})p(\mathbf{z}^{(A)}|\boldsymbol{\beta}^{(A)}) \\
&p(\boldsymbol{\beta}^{(A)})p(\mathbf{v}|\mathbf{c}^{(B)}, \mathbf{z}^{(B)}, \boldsymbol{\phi})p(\mathbf{c}^{(B)}|\boldsymbol{\beta}^{(H)})p(\mathbf{z}^{(B)}|\boldsymbol{\beta}^{(B)})p(\boldsymbol{\beta}^{(B)})\Big] \\
&+ H(q(\mathbf{c}^{(A)})) + H(q(\mathbf{z}^{(A)})) + H(q(\boldsymbol{\beta}^{(A)})) + H(q(\mathbf{c}^{(B)})) + H(q(\mathbf{z}^{(B)})) + H(q(\boldsymbol{\beta}^{(B)})) \\
&+ \mathbb{E}_q\Big[\log p(\boldsymbol{\beta}^{(H)})p(\boldsymbol{\phi})\Big] + H(q(\boldsymbol{\beta}^{(H)})) + H(q(\boldsymbol{\phi})),
\end{aligned} \tag{17}$$

where $H(\cdot)$ is the entropy of the variational distribution. Taking derivatives of this lower bound with respect to each variational parameter leads to the coordinate ascent updates.

Document-level Updates: At the document level, we update the following parameters,

$$a_t^{(A)} = 1 + \sum_n \zeta_{nt}^{(A)}, \quad b_t^{(A)} = \tau + \sum_n \sum_{s=t+1}^{K^{(A)}} \zeta_{ns}^{(A)}, \tag{18a}$$

$$a_t^{(B)} = 1 + \sum_n \zeta_{nt}^{(B)}, \quad b_t^{(B)} = \tau + \sum_n \sum_{s=t+1}^{K^{(B)}} \zeta_{ns}^{(B)}, \tag{18b}$$

$$\varphi_{ti}^{(A)} \propto \exp(\sum_n \zeta_{nt}^{(A)} \mathbb{E}_q[\log p(u_n|\phi_i)] + \mathbb{E}_q[\log \beta_i^{(H)}]) \tag{18c}$$

$$\zeta_{nt}^{(A)} \propto \exp(\sum_{i=1}^{k^{(H)}} \varphi_{ti}^{(A)} \mathbb{E}_q[\log p(u_n|\phi_i)] + \mathbb{E}_q[\log \beta_t^{(A)}]) \tag{18d}$$

$$\varphi_{ti}^{(B)} \propto \exp(\sum_n \zeta_{nt}^{(B)} \mathbb{E}_q[\log p(u_n|\phi_i)] + \mathbb{E}_q[\log \beta_i^{(H)}]) \tag{18e}$$

$$\zeta_{nt}^{(B)} \propto \exp(\sum_{i=1}^{k^{(H)}} \varphi_{ti}^{(B)} \mathbb{E}_q[\log p(u_n|\phi_i)] + \mathbb{E}_q[\log \beta_t^{(B)}]) \tag{18f}$$

Corpus-level Updates: At the corpus level, we update the following parameters,

$$a_i^{(H)} = 1 + [\sum_{t=1}^{K^{(A)}} \phi_{ti}^{(A)} + \sum_{t=1}^{K^{(B)}} \phi_{ti}^{(B)}], \tag{19a}$$

$$b_i^{(H)} = \gamma + [\sum_{t=1}^{K^{(A)}} \sum_{l=i+1}^{K^{(H)}} \varphi_{tl}^{(A)} + \sum_{t=1}^{K^{(B)}} \sum_{l=i+1}^{K^{(H)}} \varphi_{tl}^{(B)}], \tag{19b}$$

$$\lambda_{iw} = \eta + [\sum_{t=1}^{K^{(A)}} \varphi_{ti}^{(A)} (\sum_n \zeta_{nt}^{(A)} \mathbf{1}(u_n = w)) + \sum_{t=1}^{K^{(B)}} \varphi_{ti}^{(B)} (\sum_n \zeta_{nt}^{(B)} \mathbf{1}(v_n = w))] \tag{19c}$$

The expectations involved are taken under the variational distribution q and can be found in (Wang et al., 2011).

Holdout Likelihood of New Edges: Given the variational inference for the HDP, we can compute the log likelihood of the new edge $(u_{\text{new}}, v_{\text{new}})$ as follows:

$$\log P((u_{\text{new}}, v_{\text{new}})) = \log(\sum_{i=1}^{K^{(H)}} \bar{\beta}_i^{(H)} \bar{\lambda}_{iu_{\text{new}}} \; \bar{\beta}_i^{(H)} \bar{\lambda}_{iv_{\text{new}}}) \tag{20}$$

where $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^{K^{(H)} \times (W+1)}$ is the variational expectation of the corpus-level topic distribution $\boldsymbol{\lambda}$ given the observation $(u_n, v_n)_{n=1,\cdots,\infty}$, and $\bar{\boldsymbol{\beta}}^{(H)} \in \mathbb{R}^{K^{(H)}}$ is the variational expectation of the corpus-level stick lengths measuring the expected proportions of the corpus-level topics.

3.3. Detecting Anomalous Edges with Conformal Prediction

The previous subsections gave a setup of the edge exchangeable model. In this subsection, we define an anomalous edge that is significantly different from the remaining edges which were generated to be edge-exchangeable. We formulate the anomaly detection problem as a hypothesis test of the exchangeability assumption. We then propose an algorithm, a Conformal detector, which has a guaranteed upper bound for the false positive rate.

Definition 3 (Anomalous Edges) *An anomalous edge e_{N+1} is an edge which breaks the exchangeability of an edge sequence $e_1, \cdots, e_N$ if*

$$(e_1, \cdots, e_N) \overset{d}{=} (e_{\pi_1}, \cdots, e_{\pi_N}), \tag{21}$$

for any permutation $\pi : \{1, \cdots, N\} \to \{1, \cdots, N\}$, and

$$(e_1, \cdots, e_{N+1}) \overset{d}{\neq} (e_{\pi_1^+}, \cdots, e_{\pi_{N+1}^+}) \tag{22}$$

for some permutation $\pi^+ : \{1, \cdots, N+1\} \to \{1, \cdots, N+1\}$.

Suppose an existing edge sequence $e_1, e_2, \cdots, e_N$ is either constructed or approximated using the ADND model in Section 3.1, then the edge sequence is exchangeable, i.e.,

$$(e_1, e_2, \cdots, e_N) \overset{d}{=} (e_{\pi_1}, e_{\pi_2}, \cdots, e_{\pi_N}), \tag{23}$$

According to Definition 3, an edge e_{N+1} is anomalous if it breaks edge exchangeability, i.e.,

$$(e_1, e_2, \cdots, e_{N+1}) \overset{d}{\neq} (e_{\pi_1}, e_{\pi_2}, \cdots, e_{\pi_{N+1}}), \tag{24}$$

for some permutation $\pi^+ : \{1, \cdots, N+1\} \to \{1, \cdots, N+1\}$. We next demonstrate how we use such a definition of anomalous edges to construct an anomaly detection algorithm.

Conformal Anomalous Edge Detection: Our algorithm is based on conformal prediction (Shafer and Vovk, 2008). Conformal prediction uses labeled training data to determine the confidence in new predictions. Its output is a prediction set that is guaranteed to contain the true label with a specified confidence level as long as the data is exchangeable (Zeni et al., 2020). To this end, conformal prediction estimates a p-value for each possible label for a new example. In our application, we have training edges that are labeled as normal edges, and we aim to use conformal prediction to determine the confidence that new edges are also normal. Specifically, the conformal detection algorithm we propose (Algorithm 1) detects whether an edge e_{N+1} is anomalous or not by using conformal prediction to estimate the p-value for the null hypothesis:

$$H_0 : e_1, e_2, \cdots, e_{N+1} \text{ forms an exchangeable sequence.} \tag{25}$$

In order to estimate the p-values for the null hypothesis H_0, we employ a *non-conformity measure* (NCM) (Shafer and Vovk, 2008) to measure how different an observation is relative to other observations. In detecting anomalous edges, to measure how different an edge e_n is relative to other edges in the edge set $E = \{e_1, \cdots, e_{N+1}\}$, we define the NCM as the negative likelihood of edge e_n given other edges in the edge set E, which can be computed according to Eq. (20) as follows:

$$\alpha_n = -P(e_n | E \setminus \{e_n\}) \tag{26}$$

α_n in (26) constitutes the test statistic for our detector, which determines the corresponding p-value as follows:

$$p_n = \frac{1}{N+1} \big|\{i = 1, \cdots, N+1 \mid \alpha_n > \alpha_i\}\big| + \frac{u}{N+1} \big|\{i = 1, \cdots, N+1 \mid \alpha_n = \alpha_i\}\big|, \tag{27}$$

where $u \sim \text{Unif}(0, 1)$. Given a significance level or an *anomaly threshold*, ϵ, if $p_n \leq \epsilon$, then e_n is classified as an anomalous edge: $\text{Anom}_n^\epsilon = 1$. Otherwise, e_n is classified as normal: $\text{Anom}_n^\epsilon = 0$.

Note that we will need to fit the ADND model to compute the log likelihood in Eq. (20) for any $E \setminus \{e_n\}, n = 1, \cdots, N+1$. To avoid such computation burden, we take advantage of the inductive conformal predictor (Papadopoulos, 2008). We split[3] the existing edge set (excluding the new edge to be tested) into a training set E_{train} and a calibration set E_{calib}. We fit the ADND model based on E_{train} and compute the non-conformity scores for each observation in E_{calib}. The p-value of the new edge is computed by comparing its non-conformity score with the ones of E_{calib}:

$$p_n = \frac{\big|\{i \in E_{\text{calib}} \cup \{N+1\} \mid \alpha_n > \alpha_i\}\big|}{|E_{\text{calib}}| + 1} + \frac{u\big|\{i \in E_{\text{calib}} \cup \{N+1\} \mid \alpha_n = \alpha_i\}\big|}{|E_{\text{calib}}| + 1}, \tag{28}$$

We present the complete anomaly detection algorithm (Conformal detector) in Algorithm 1. Presented below is our main finding: we demonstrate that Algorithm 1 has a false positive rate (Definition 4), meaning that it incorrectly detects normal edges as anomalies, that is no greater than the predetermined anomaly threshold of ϵ.

Definition 4 (False Positive Rate) *A false positive rate refers to the proportion of non-anomalous data points that are incorrectly labeled as anomalous or abnormal. For the conformal anomaly detection algorithm, the false positive rate represents the proportion of normal edges that are incorrectly detected as anomalous (Definition 3).*

Proposition 5 *Assuming that a new edge e_{N+1} is normal in relation to the existing edges, i.e., the edges $e_1, e_2, \cdots, e_{N+1}$ form an exchangeable sequence. For the choice of anomaly score in Eq. (26), the specified anomaly threshold ϵ corresponds to an upper bound of the false positive rate (Definition 4) of Algorithm 1:*

$$P(Anom_{N+1}^\epsilon = 1) \leq \epsilon \tag{29}$$

Proof See Appendix. ■

[3] Due to the edge-exchangeability, the split can be done randomly.

Algorithm 1: (Conformal Detector) Anomalous edge detection based on inductive conformal prediction

```
Input: anomaly threshold $\epsilon$, training edge set $E_{\text{train}}$, calibration edge set $E_{\text{calib}}$, and a
       new edge $e_{N+1}$ to be tested.
Output: Indicator variable $\text{Anom}^{\epsilon}_{N+1} \in \{0, 1\}$ with performance guarantee
        (Proposition 5).
Initialize the corpus-level topic distributions $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_i)_{i=1,\cdots,K^{(H)}}$, the parameters of the
  beta distribution $\boldsymbol{a}^{(H)} = (a_i^{(H)})_{i=1,\cdots,K^{(H)}-1}$, $\boldsymbol{b}^{(H)} = (b_i^{(H)})_{i=1,\cdots,K^{(H)}-1}$.
/* Approximate the training edges using the ADND edge distribution      */
while stopping criterion is not met do
    Compute $\boldsymbol{a}^{(A)}, \boldsymbol{b}^{(A)}, \boldsymbol{a}^{(B)}, \boldsymbol{b}^{(B)}, \boldsymbol{\psi}_t^{(A)}, \boldsymbol{\zeta}_n^{(A)}, \boldsymbol{\psi}_t^{(B)}$ and $\boldsymbol{\zeta}_n^{(B)}$ using variational inference
      through document-level updates, Eq. (18).
    Update $\boldsymbol{\lambda}, \boldsymbol{a}^{(H)}, \boldsymbol{b}^{(H)}$ using Eq. (19).
end
/* Compute the non-conformity scores of the calibration edges           */
for $j \leftarrow 1$ to $|E_{calib}|$ do
    Compute $\alpha_j = -P(e_j|E_{\text{train}})$ according to Eq. (20).
end
/* Compute the $p$-value that the test edge is normal                   */
Compute $p_{N+1}$ according to Eq. (28).
if $p_{N+1} \leq \epsilon$ then
    $\text{Anom}^{\epsilon}_{N+1} \leftarrow 1$
else
    $\text{Anom}^{\epsilon}_{N+1} \leftarrow 0$
end
```

4. Numerical Examples of Anomalous Edge Detection

In this section, we illustrate how the proposed Conformal detector (Algorithm 1) can detect anomalous transactions in the Ethereum blockchain transaction network.

Dataset. Decentralized exchanges (DEXs), which enable the non-custodial trade of digital assets, are implemented on the Ethereum blockchain. The most popular DEX that uses an automated liquidity protocol is Uniswap (Adams et al., 2021). Since Uniswap V2, it has allowed the creation of arbitrary ERC20/ERC20[4] pairs that store the pooled reserves of two different assets. Uniswap V3, the most recent version, debuts in May 2021. Compared to top centralized exchanges (CEXs), Uniswap V3 is more liquid (Liao and Robinson, 2022).

In Uniswap, traders can exchange assets using a swap function that involves adding one asset to a liquidity pair while removing the other. To improve routing efficiency, traders can perform multi-hop swaps, such as exchanging token X for Y and then exchanging Y for Z. However, if a liquidity pair for tokens X and Z does not exist, this can be treated as a potential pricing anomaly and an arbitrage opportunity. In our experiment, we use this scenario as the ground truth anomaly to be detected.

To do this, we gathered all the multi-hop swap transactions in Uniswap V3. We accessed the Ethereum Mainnet using the Python library *Web3.py*[5]. We searched all blocks with a height between 12500000 and 12503000, which corresponds to about three weeks after the debut of Uniswap V3 on May 5, 2021. A directed edge from the source token to the destination token is used to model each transaction. There are 140 anomalous edges in our collection of 307 test edges, 363 training edges, and 363 calibration edges.

Baseline. We use the RHSS method (Ranshous et al., 2015) as the baseline, which is an edge stream anomaly detector applicable to directed multigraphs. RHSS estimates the edge probabilities by combining three individual scores, i.e., sample, preferential attachment, and homophily scores. We assign equal weighting coefficients to the individual scores.

Evaluation metrics. Both the Conformal detector and the RHSS method output a score per edge (the Conformal detector outputs a p-value; the RHSS method outputs an estimate of edge probabilities. Both are lower for more anomalous edges). We compute the precision and recall following the method outlined in (Eswaran and Faloutsos, 2018): sorting the edges in ascending order of their scores, we count the number of edges c_k flagged correctly as anomalous among the top k edges, for every cut-off rank $k \in \mathbb{N}$. Let C be the total number of ground truth anomalies, we compute: precision@$k = c_k/k$ and recall@$k = c_k/C$. In addition, we compute the true positive rate (proportion of anomalous edges correctly classified as anomalies) and the false positive rate (proportion of normal edges mistakenly classified as anomalies) for a range of threshold values. We also report AUC (area Under the ROC curve).

Fig. 2 shows the token swaps collected from Uniswap V3. The normal edges are blue; the anomalous edges are red. The left of Fig. 3 shows the precision vs. recall curves, while

[4] The ERC20 (*Ethereum Request for Comment #20*) stands for fungible tokens on the Ethereum blockchain.

[5] https://web3py.readthedocs.io/

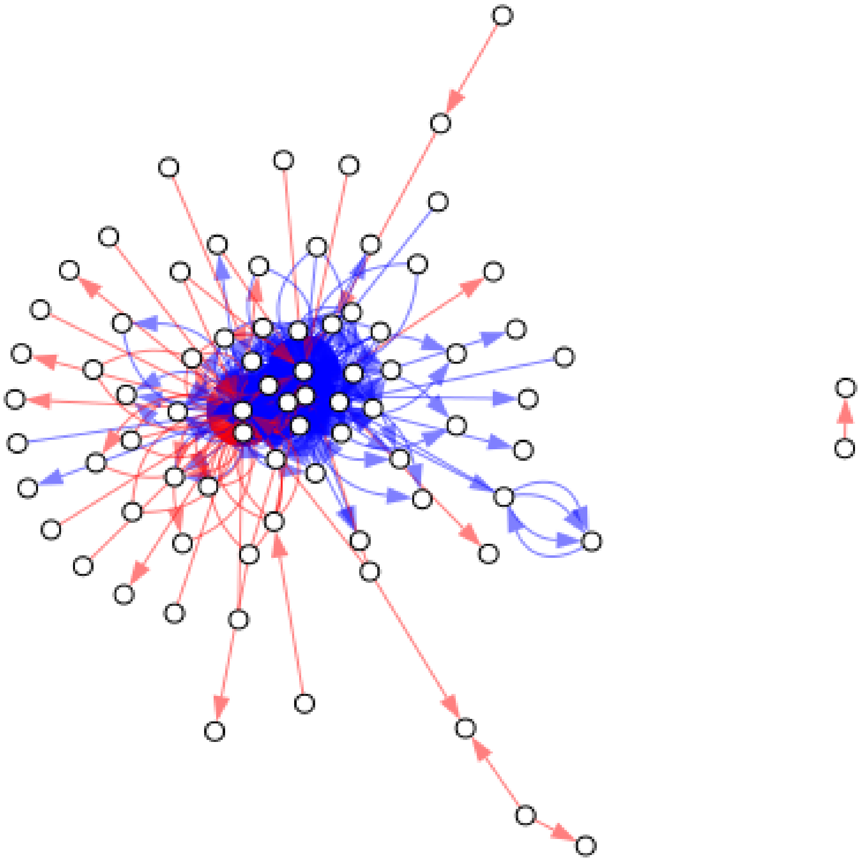

Figure 2: The Uniswap V3 token swap network collected as described in Section 4. The edges that are normal are blue, whereas the anomalous ones are red. The normal edges are approximated using an exchangeable ADND model, and we aim to detect the anomalous edges that represent tokens swaps without existing liquidity pools. The normal edges and anomalous edges suggest different generating properties.

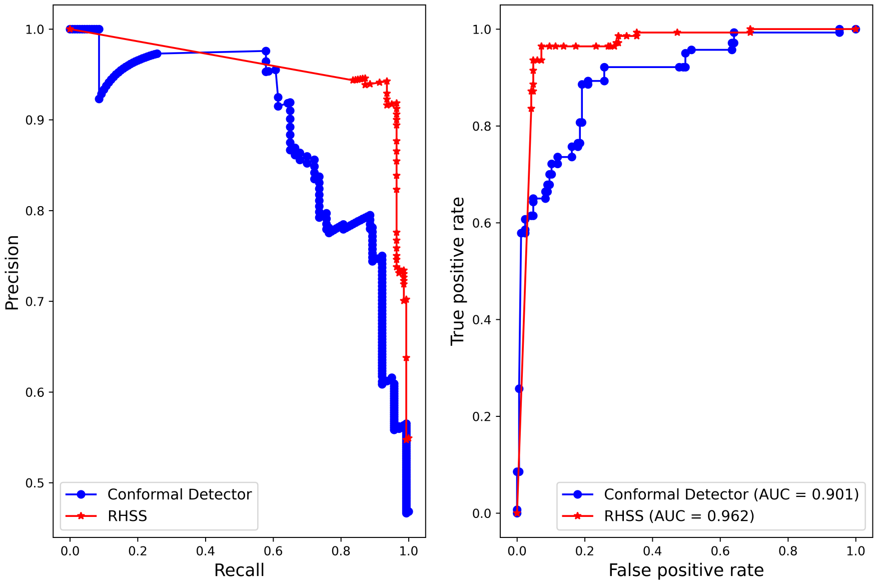

Figure 3: The precision-recall curve and the ROC curve of the Conformal Detector (blue circles), and the RHSS method (red stars). The Conformal Detector is comparable to the RHSS method.

the right shows the ROC curve and the AUC computed. The performance of the Conformal detector is comparable to that of the RHSS method. Overall, the comparison result indicates that the Conformal detector is comparable with the RHSS method in anomalous edge detection. Moreover, the relatively small performance gap illustrates the robustness of the Conformal detector to misspecified models.

5. Conclusions and Extensions

Conclusions: This paper addressed detecting anomalous edges in social networks and transaction graphs based on edge exchangeability. We motivated edge exchangeability in social networks via real-world examples. We proposed anomalous edge detection algorithm (Conformal detector) by combining conformal prediction with the edge exchangeable model. We proved that Conformal detector guarantees a false positive rate (flagging normal edges as anomaly) lower than a pre-specified threshold. In the numerical experiments, we verified that edge exchangeability is a valid criterion for distinguishing anomalous edges from normal edges using Ethereum blockchain transaction dataset. In addition, we demonstrated that the proposed Conformal detector achieves a comparable anomaly detection performance to the baseline method on synthetic datasets.

Limitations and Extensions: There are several limitations of the proposed method that open up future research directions. Firstly, the assumption of edge exchangeability may not hold for real-world networks whose properties change over time as they expand. (Ghalebi et al., 2019a,b) have incorporated distance-dependent Chinese restaurant process to model nonexchangeable edge sequence by preferentially assigning edges to communities that have been recently active. Secondly, although the validity of conformal prediction (e.g. false positive rate in the proposed Conformal detector algorithm) is guaranteed, the efficiency of conformal prediction (i.e., the prediction region of normal edges) depends on the non-conformity measure. It is worthwhile to investigate cascades of predictors to improve the efficiency (Fisch et al., 2020). Finally, combining the proposed conformal prediction framework with graph neural networks or embedding methods remains an interesting future direction. However, conformal prediction crucially depends on the assumption of exchangeability which requires verification of exchangeability in the graph-structured data.

Acknowledgments

We would like to acknowledge support for this project from City University of Hong Kong New Faculty Start-up Grant, the U. S. Army Research Office under grant W911NF-21-1-0093, and the National Science Foundation under grant CCF-2112457.

Proof of Proposition 5

Definition 6 (Rank, ascending) *For a set of n real numbers $X = \{x_1, \cdots, x_n\}$, define the rank of x_i among X as*

$$rank(x_i; X) = \begin{cases} |\{j \in \{1, \cdots, n\} : x_j \leq x_i\}| \\ \qquad \textit{if all elements of } X \textit{ are distinct;} \\ |\{j \in \{1, \cdots, n\} : x_j + \xi U_j \leq x_i + \xi U_i\}| \\ \qquad \textit{if some elements of } X \textit{ are equal.} \end{cases} \tag{30}$$

where $\xi > 0$ is a small number such that

$$\xi \leq \min_{\forall i,j \in \{1, \cdots, n\}: x_i \neq x_j} \frac{|x_i - x_j|}{2} \tag{31}$$

and $U_1, \cdots, U_n \sim Unif(-1, 1)$ are iid uniform random variables. Because $U_1, \cdots, U_n$ are almost surely distinct, $x_i + \xi U_i$ are also distinct. The addition of $U_1, \cdots, U_n$ breaks potential ties among the original sequence $x_1, \cdots, x_n$.

Theorem 1. **(Kuchibhotla, 2020)** If $W_1, \cdots, W_n$ are exchangeable random variables, then

$$P(\text{rank}(W_i; \{W_1, \cdots, W_n\}) \leq m) = \frac{\lfloor m \rfloor}{n}, \tag{32}$$

where $\text{rank}(W_i; \{W_1, \cdots, W_n\})$ represents the rank of W_i among $\{W_1, \cdots, W_n\}$, $\lfloor m \rfloor$ represents the largest integer smaller than or equal to m. Moreover, the random variable $\text{rank}(W_i; \{W_1, \cdots, W_n\})/n$ is a valid p-value, i.e.,

$$P(\text{rank}(W_i; \{W_1, \cdots, W_n\})/n \leq \alpha) \leq \alpha \quad \forall \alpha \in [0, 1]. \tag{33}$$

Theorem 2. **(Kuchibhotla, 2020)** Suppose $W = (W_1, \cdots, W_n) \in \mathcal{W}^n$ is a vector of exchangeable random variables. Fix a transformation $G : \mathcal{W}^n \to (\mathcal{W}')^m$. If for each permutation $\bar{\pi} : \{1, \cdots, m\} \to \{1, \cdots, m\}$ there exists a permutation $\underline{\pi} : \{1, \cdots, n\} \to \{1, \cdots, n\}$ such that

$$\bar{\pi} G(w) = G(\underline{\pi} w), \quad \forall w \in \mathcal{W}^n, \tag{34}$$

then $G(.)$ preserves exchangeability of W.

Proposition 5. **(Section 3.3)** Assume that $e_1, e_2, \cdots, e_{N+1}$ form a sequence of exchangeable random variables. For the NCM as defined in Eq. (26)), the specified anomaly threshold ϵ corresponds to an upper bound of the probability of the event that e_{N+1} is classified as an anomalous edge by Algorithm 1:

$$P(\text{Anom}_{N+1}^{\epsilon} = 1) \leq \epsilon \tag{35}$$

Proof. Let $e_1, e_2, \cdots, e_{N+1} \in \mathcal{E}$ where $\mathcal{E}$ represents the set of all possible edges. For any permutation $\pi : \{1, \cdots, N+1\} \to \{1, \cdots, N+1\}$,

$$\begin{aligned}&\pi\Bigg(-P(e_1|E_{N+1} \setminus \{e_1\}), \cdots, -P(e_{N+1}|E_{N+1} \setminus \{e_{N+1}\})\Bigg)\\=&\Bigg(-P(e_{\pi_1}|E_{N+1} \setminus \{e_{\pi_1}\}), \cdots, -P(e_{\pi_{N+1}}|E_{N+1} \setminus \{e_{\pi_{N+1}}\})\Bigg)\end{aligned} \tag{36}$$

Theorem 2 indicates that the negative likelihood of an edge, $\alpha : \mathcal{E} \to \mathbb{R}$, is a transformation that preserves exchangeability. Thus $\alpha_1, \cdots, \alpha_{N+1}$ are real-valued exchangeable random variables.

The right hand side of Eq. (27) is the rank of α_{N+1} among $\{\alpha_1, \cdots, \alpha_{N+1}\}$. According to Theorem 2, the p-value constructed from Eq. (27) is valid, i.e., p_{N+1} is distributed uniformly on $[0, 1]$ when $e_1, \cdots, e_{N+1} \in \mathcal{E}$ be exchangeable random variables. This implies that $P(\text{Anom}^{\epsilon}_{N+1} = 1 \mid e_{N+1} \text{ is normal})$, i.e., the false alarm rate of Algorithm 1, is less than or equal to ϵ at any specified anomaly threshold ϵ.

References

Hayden Adams, Noah Zinsmeister, Moody Salem, River Keefer, and Dan Robinson. Uniswap v3 core. *Tech. rep., Uniswap, Tech. Rep.*, 2021.

Charu C Aggarwal, Yuchen Zhao, and S Yu Philip. Outlier detection in graph streams. In *2011 IEEE 27th international conference on data engineering*, pages 399–409. IEEE, 2011.

Leman Akoglu, Mary McGlohon, and Christos Faloutsos. Oddball: Spotting anomalies in weighted graphs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 410–421. Springer, 2010.

Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery*, 29(3):626–688, 2015.

David J Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Siddharth Bhatia, Bryan Hooi, Minji Yoon, Kijung Shin, and Christos Faloutsos. Midas: Microcluster-based detector of anomalies in edge streams. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3242–3249, 2020.

PV Bindu and P Santhi Thilagam. Mining social networks for anomalies: Methods and challenges. *Journal of Network and Computer Applications*, 68:213–229, 2016.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 4249–4257, 2016.

Feiyang Cai and Xenofon Koutsoukos. Real-time out-of-distribution detection in learning-enabled cyber-physical systems. In *2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS)*, pages 174–183. IEEE, 2020.

François Caron and Emily B Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 79(5):1295, 2017.

Deepayan Chakrabarti. Autopart: Parameter-free graph partitioning and outlier detection. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 112–124. Springer, 2004.

Yen-Yu Chang, Pan Li, Rok Sosic, MH Afifi, Marco Schweighauser, and Jure Leskovec. F-fade: Frequency factorization for anomaly detection in edge streams. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 589–597, 2021.

Harry Crane. *Probabilistic foundations of statistical network analysis.* CRC Press, 2018.

Harry Crane and Walter Dempsey. Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, 113(523):1311–1326, 2018.

Dhivya Eswaran and Christos Faloutsos. Sedanspot: Detecting anomalies in edge streams. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 953–958. IEEE, 2018.

Dhivya Eswaran, Christos Faloutsos, Sudipto Guha, and Nina Mishra. Spotlight: Detecting anomalies in streaming graphs. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1378–1386, 2018.

Valentina Fedorova, Alex Gammerman, Ilia Nouretdinov, and Vladimir Vovk. Plug-in martingales for testing exchangeability on-line. *arXiv preprint arXiv:1204.3251*, 2012.

Adam Fisch, Tal Schuster, Tommi Jaakkola, and Regina Barzilay. Efficient conformal prediction via cascaded inference with expanded admission. *arXiv preprint arXiv:2007.03114*, 2020.

Alexander Gammerman and Vladimir Vovk. Hedging predictions in machine learning. *The Computer Journal*, 50(2):151–163, 2007.

Elahe Ghalebi, Hamidreza Mahyar, Radu Grosu, Graham W Taylor, and Sinead A Williamson. A nonparametric bayesian model for sparse dynamic multigraphs. *arXiv preprint arXiv:1910.05098*, 2019a.

Elahe Ghalebi, Hamidreza Mahyar, Radu Grosu, Graham W Taylor, and Sinead A Williamson. Sequential edge clustering in temporal multigraphs. *arXiv preprint arXiv:1905.11724*, 2019b.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.

Douglas N Hoover. Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2, 1979.

Vladislav Ishimtsev, Alexander Bernstein, Evgeny Burnaev, and Ivan Nazarov. Conformal k-nn anomaly detector for univariate data streams. In *Conformal and Probabilistic Prediction and Applications*, pages 213–227. PMLR, 2017.

Svante Janson. On edge exchangeable random graphs. *Journal of statistical physics*, 173 (3):448–484, 2018.

Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, volume 3, page 5, 2006.

Arun Kumar Kuchibhotla. Exchangeability, conformal prediction, and rank tests. *arXiv preprint arXiv:2005.06095*, 2020.

Rikard Laxhammar and Göran Falkman. Conformal prediction for distribution-independent anomaly detection in streaming vessel data. In *Proceedings of the first international workshop on novel data stream pattern mining techniques*, pages 47–55, 2010.

Gordon Liao and Dan Robinson. The dominance of uniswap v3 liquidity. https://uniswap.org/blog/uniswap-v3-dominance, 2022. Accessed: 2022-10-30.

James Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *Advances in Neural Information Processing Systems*, 25:998–1006, 2012.

Kurt Miller, Michael Jordan, and Thomas Griffiths. Nonparametric latent feature models for link prediction. *Advances in neural information processing systems*, 22:1276–1284, 2009.

Buddhika Nettasinghe, Samrat Chatterjee, Ramakrishna Tipireddy, and Mahantesh Halappanavar. Extending conformal prediction to hidden markov models with exact validity via de finetti's theorem for markov chains, 2022. URL https://arxiv.org/abs/2210.02271.

Nam P Nguyen, Guanhua Yan, My T Thai, and Stephan Eidenbenz. Containment of misinformation spread in online social networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 213–222, 2012.

Mathias Niepert and Pedro Domingos. Exchangeable variable models. In *International Conference on Machine Learning*, pages 271–279. PMLR, 2014.

Peter Orbanz and Daniel M Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):437–461, 2015.

Peter Orbanz and Yee Whye Teh. Bayesian nonparametric models. *Encyclopedia of machine learning*, 1, 2010.

Harris Papadopoulos. *Inductive conformal prediction: Theory and application to neural networks*. INTECH Open Access Publisher Rijeka, 2008.

Harris Papadopoulos, Volodya Vovk, and Alex Gammerman. Conformal prediction with neural networks. In *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, volume 2, pages 388–395. IEEE, 2007.

Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, and Nagiza F Samatova. Anomaly detection in dynamic networks: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3):223–247, 2015.

Stephen Ranshous, Steve Harenberg, Kshitij Sharma, and Nagiza F Samatova. A scalable approach for outlier detection in edge streams using sketch-based approximations. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 189–197. SIAM, 2016.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

Kijung Shin, Tina Eliassi-Rad, and Christos Faloutsos. Patterns and anomalies in k-cores of real-world graphs with applications. *Knowledge and Information Systems*, 54(3):677–710, 2018.

Hanghang Tong and Ching-Yung Lin. Non-negative residual matrix factorization with application to graph anomaly detection. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 143–153. SIAM, 2011.

Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov, Alexander Gammerman, and Vladimir Vovk. Inductive conformal martingales for change-point detection. In *Conformal and Probabilistic Prediction and Applications*, pages 132–153. PMLR, 2017.

Vladimir Vovk, Ilia Nouretdinov, and Alexander Gammerman. Testing exchangeability online. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 768–775, 2003.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

Claudia Wagner, Silvia Mitter, Christian Körner, and Markus Strohmaier. When social bots attack: Modeling susceptibility of users in online social networks. In *# MSM*, pages 41–48, 2012.

Chong Wang, John Paisley, and David Blei. Online variational inference for the hierarchical Dirichlet process. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 752–760. JMLR Workshop and Conference Proceedings, 2011.

Yuchung J Wang and George Y Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987.

Sinead A Williamson. Nonparametric network models for link prediction. *The Journal of Machine Learning Research*, 17(1):7102–7121, 2016.

Wenchao Yu, Wei Cheng, Charu C Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang. Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2672–2681, 2018.

Gianluca Zeni, Matteo Fontana, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *arXiv preprint arXiv:2005.07972*, 2020.