**INTRODUCTION**

AN effective health insurance system plays a significant role in managing healthcare resources, enhancing life quality for people, and maintaining social stability. More than 1.3 billion people have enrolled in the National Basic Medical Insurance in China1. However, increasing health insurance fraud events have become a severe social problem. According to the inspection conducted by the National Healthcare Security Administration and the Ministry of Public Security in China, nearly half of 815,000 health institutes have improper or even illegal funds costs in 2020, leading to an economic loss of more than 22.3 billion yuan ($3.4 billion)2. The emerging collusive fraud is the most serious and urgent among these events [1]. Fraudsters collude to purchase drugs with insurance reimbursement and cash the drugs out. The massive amount of fraud brings serious consequences. There is an urgent need for efficient and effective detection methods to quickly identify collusive fraud and prevent further loss.

Detecting collusive fraud in health insurance faces two challenges. First, it is difficult to distinguish the medical visits behavior of fraudsters from those of normal patients. Typically, fraudsters frequently buy large quantities of easily marketable drugs. However, due to the need to maintain long-term medication, patients with chronic diseases and those requiring Traditional Chinese medicine (TCM) treatment have similar purchasing behaviors to fraudsters’. Second, manual auditing is necessary but laborious. Misidentification is unacceptable for fraud detection because a patient has to bear legal responsibility after being recognized as a fraudster. Verifying fraud requires auditors to synthesize a large amount of contextual information, such as the amount of reimbursement, the degree to which the patient’s disease and drugs match, and the time of visits.

\

Existing collusive fraud detection methods can hardly handle these challenges. Existing methods focus on modeling the relationships between fraudsters by graphs and detecting fraudulent groups through statistical or machine learning (ML) methods. Statistical approaches use structural and attribute features [2], [3], [4], or spectral analysis [5] to detect anomalous substructures (i.e., fraud groups/events). However, audit experts told us that these methods are prone to false positives, due to the ambiguous nature of collusive fraud in health insurance. Excluding false positives is time-consuming for auditors and can significantly reduce detection efficiency. ML methods mainly use graph neural network (GNN) models to detect collusive fraud [6], [7], [8]. Fraudsters and their associations are constructed as homogeneous or heterogeneous graphs. GNNs trained on labeled data can yield the representation of fraudsters and be further applied to judge unlabeled individuals. Unfortunately, large amounts of labeled data are indispensable for high-performance GNNs. Without sufficient labeled fraud, GNN models are not applicable in our scenario.

To address these challenges, we propose a novel visual analytics approach to help health insurance audit experts identify suspicious groups, investigate the visit behavior of suspicious patients, and validate collusive fraud results. We propose a co-visit network to represent the relationship among patients. The weights of the edges are calculated by extracting the characteristics of collusive fraudsters, such as the time gap and number of visits. Suspicious groups with multiple simultaneous visits to the same location can then be identified by a weighted community detection algorithm. The algorithm is integrated into a prototype system, Fraud Auditor, that supports experts in interactively browsing and improving model detection results. Fraud Auditor can help experts quickly locate and examine fraud by observing co-visit links in visualizations of patient medical behavior. Combined with contextual information such as disease, drug, and fee information, false positive groups can be verified and excluded. We provide case studies and expert interviews in real health insurance scenarios to validate the effectiveness of the proposed approach.

The contributions in this work include:

\_ A problem characterization that summarizes the requirements of collusive fraud detection in the scenario of health insurance.

\_ A novel three-stage visual analytics approach to detect collusive fraud in health insurance that considers the visit pattern of fraud groups and expert knowledge.

\_ An interactive prototype system, Fraud Auditor, to facilitate the identification, examination, and validation of suspicious collusive fraud groups.