Yikun Ban

Computer Science, Peking University, Beijing, China +86 15652658218
Research Interests: Data Mining, Machine Learning, Graph Algorithms banyikun@pku.edu.cn

EDUCATION:

- M.S. in Computer Science, **Peking University**, Beijing, China. Major GPA: 3.53/4 **Sep. 2016 - Present**
  Core Course: Data Mining **(95/100, Top 1**), Deep Learning (90/100), Machine Learning (87), Algorithm Analysis (86/100)
- B.S. in Spatial-information & Digitalized Technology, **Wuhan University**, Wuhan, China. **Sep. 2012- Jun. 2016**
  Ranking: 2/51 (overall ranking), GPA: 3.34/4

RESEARCH EXPERIENCE:

Data Ming, Machine Learning, Graph Algorithms. Research Assistant **Nov. 2017- Present**
Advisors: Wei Xu (Assistant Prof.), Yitao Duan (Visiting Prof.), and Ling Huang (Visiting Prof.). Tsinghua University

Publication 1 : Yikun Ban, Xin Liu, Ling Huang, Yitao Duan, Xue Liu, and Wei Xu. No Place to Hide: Catching Fraudulent Entities in Tensors. Submitted to **WWW 2019** (arXiv:https://arxiv.org/pdf/1810.06230.pdf). Contributions:

- [Graph Model]. We propose the novel Information Sharing Graph(ISG) model, which converts every value sharing in a tensor to the representation of weighted edges or nodes. Furthermore, our graph model considers diverse data types and their corresponding distributions based on information theory, to automatically prioritize multiple features.
- [Algorithm]. We propose the D-Spot algorithm, which is able to find multiple densest subgraphs in one run. And we theoretically prove that the multiple subgraphs found by D-Spot must contain some subgraphs that are at least 1/2 as dense as the optimum. In real-world graphs, D-Spot is up to 11× faster than the state-of-the-art competing algorithm.
- [Effectiveness]. In an N-dimensional tensor, the entity group found by the ISG+D-Spot is at least 1/2 of the optimal group regarding density, compared with the 1/N guarantee ensured by competing methods. In experiments using eight public real-world datasets, ISG+D-Spot detected fraudulent entities more accurately than conventional methods.

Publication 2 : **Yikun Ban**, Yitao Duan, and Ling Huang. On Finding Dense Subgraphs in Bipartite Graphs: Linear Algorithms with Applications to Fraud Detection. Submitted to **ICDE 2019** (arXiv:https://arxiv.org/pdf/1810.06809.pdf). Contributions:

- [A New Graph Problem]. We introduce a new restricted biclique problem, Maximal Half Isolated Biclique (MHI Biclique), and show that the problem finds immediate applications in fraud detection.
- [A Linear Algorithm]. We propose a novel data structure, S-tree, and a mining algorithm, that solves the MHI Biclique problem in linear time. The algorithm is effective for detecting unbalanced dense sub-bipartite graphs. We provide theoretical proofs regarding the algorithm's effectiveness and robustness against adversarial perturbations.
- [Practical Algorithms for Fraud Detection]. Based on the S-tree-based MHI Biclique problem solver, we introduce a new algorithm that detects near bicliques and can be used to catch a wide range of fraud groups.
- [Automatic Feature Prioritization]. We further extend S-tree to S-forest to handle multimodal data. S-forest can utilize both structural or graph data representing the relations between objects and attribute data characterizing individual objects. Furthermore, S-forest can automatically combine and prioritize multiple features, reducing the need for feature engineering.
- [Effectiveness on Real-world Data]. We conducted extensive experiments on thirteen real-world datasets, including twelve public datasets and one proprietary data collected from the production system log of a major e-commerce vendor. Our solution outperforms strong rivals across all configurations, becoming the new state of the art in fraud detection.

Publication 3 : Yikun Ban, Jiao Sun, Xin Liu, Ling Huang, Yitao Duan, and Wei Xu. FraudTrap: Catching Loosely Synchronized Behavior in Face of Camouflage. Submitted to **WWW 2019** (arXiv:https://arxiv.org/pdf/1810.08885.pdf). Contributions:

- [Metric C]. We build Object Similarity Graph (OSG) by a novel similarity metric, C-score, which transforms the sparse subgraphs induced by fraud groups in the bipartite graph into the much denser subgraphs in OSG, by merging information from unlabeled and labeled (if available) data.
- [Algorithm LPA-TK]. We propose a similarity-based clustering algorithm, LPA-TK, that perfectly fits in OSG and outperforms the baseline (LPA) in face of noise edges (camouflage).
- [Metric F]. Given candidate groups returned by C + LPA-TK, we propose an interpretable suspiciousness metric, F-score, meeting the all basic "axioms" proposed in CrossSpot.
- [Effectiveness]. Our method FraudTrap (C+LPA-TK+F) can operate in two modes: unsupervised and semi-supervised. The unsupervised mode outperforms other state-of-the-art methods for catching synchronized behavior in face of camouflage. Semi-supervised mode naturally takes advantage of partially labeled data to further improve the performance.

Computer Vision. Research Assistant. **Mar. 2017- Aug. 2017**
Advisor: Ge Li (Prof.) . Peking University

- Implemented an effective system to semantic reasoning for autonomous driving.
- Grouped classification, detection and semantic segmentation via a unified architecture where the encoder consisting of convolution and pooling layers from VGG network is shared amongst the three tasks.

AWARDS

- **First Prized (Top 0.1%, 2859 teams involved).** The 4th "China Software Cup" Software Design Competition for College Students (National Competition). - Aug. 2015
- **Second Prized (Top 0.2%, 5000+ teams involved)** in National Finals, **First Prized** in Provincial Finals. "Challenge Cup" China College Student's Entrepreneurship Competition in 2014. - Nov. 2014
- **The Best Financial Creative Award**, "CITI CUP" Citi Financial Innovation Application Competition in 2014. - Sep. 2014

HONERS

- Excellent Students Scholarship of Wuhan University
- Outstanding Students of Peking University

SKILLS

- Programming: Python, C++, Java, Scala, Swift, Matlab