Detection of Spammers in a Social Network

Ali El Abridi / 294899
Bouquet Yann / 273827
Müller Jonas / 280073
Kalim Tariq / 239239

INTRODUCTION

- More and More spammers on social networks
- Adapt their accounts: need manual detection
- Add graph theory

  for binary classification

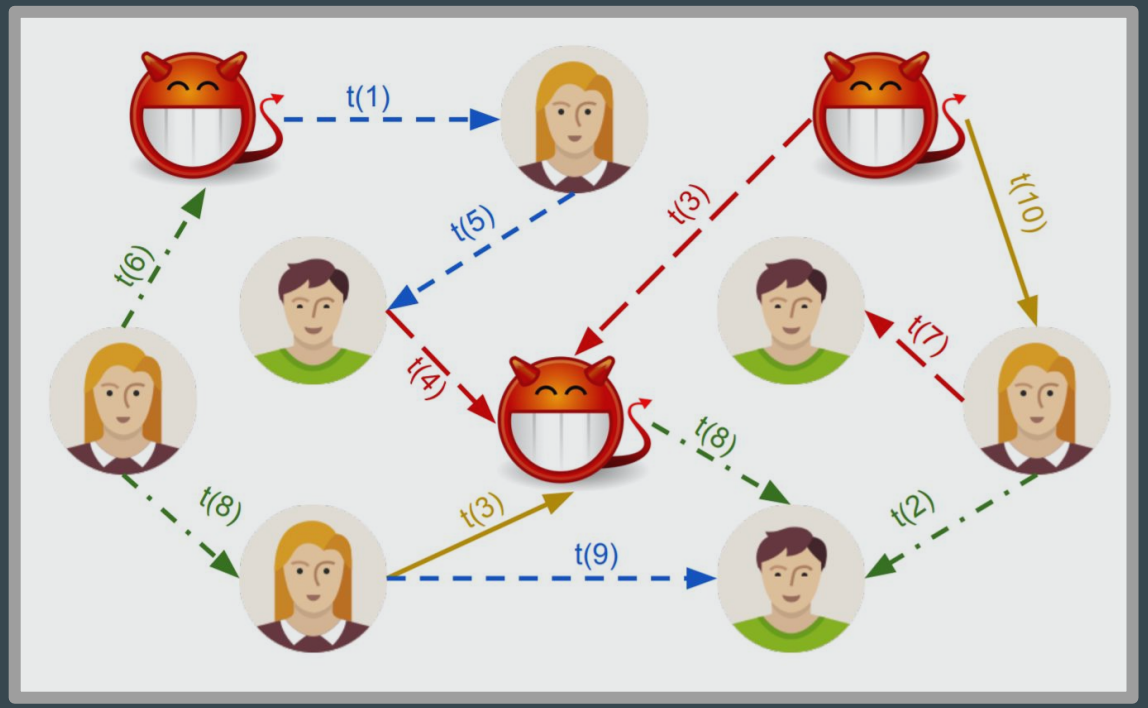


Figure from:
https://lings-data.soe.ucsc.edu/public/social_spammer/

I. **Visualization**

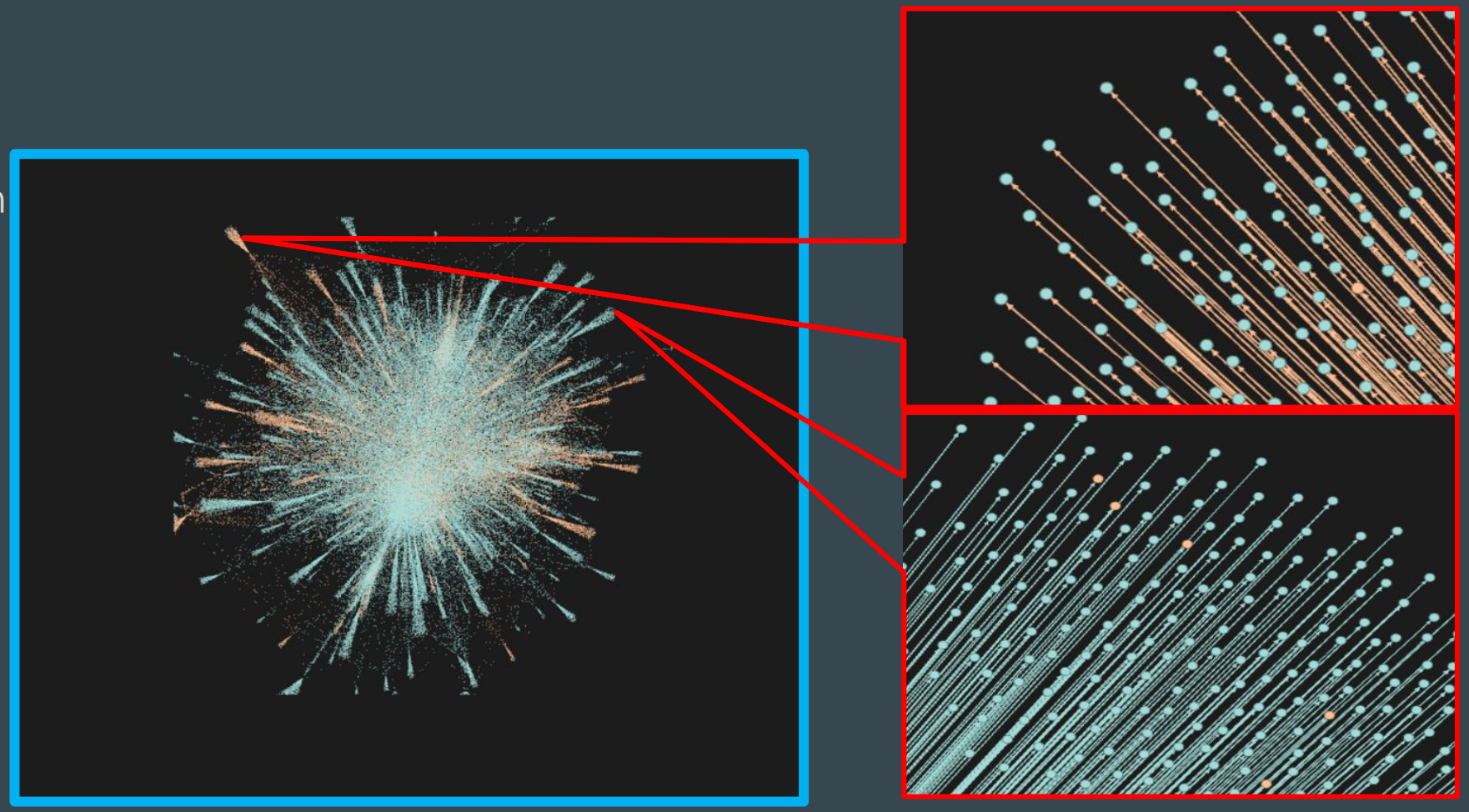
II. Exploration & Extraction

III. Exploitation

VISUALIZATION

Sampling of the relation 1 graph

- Spammer
- Not Spammer



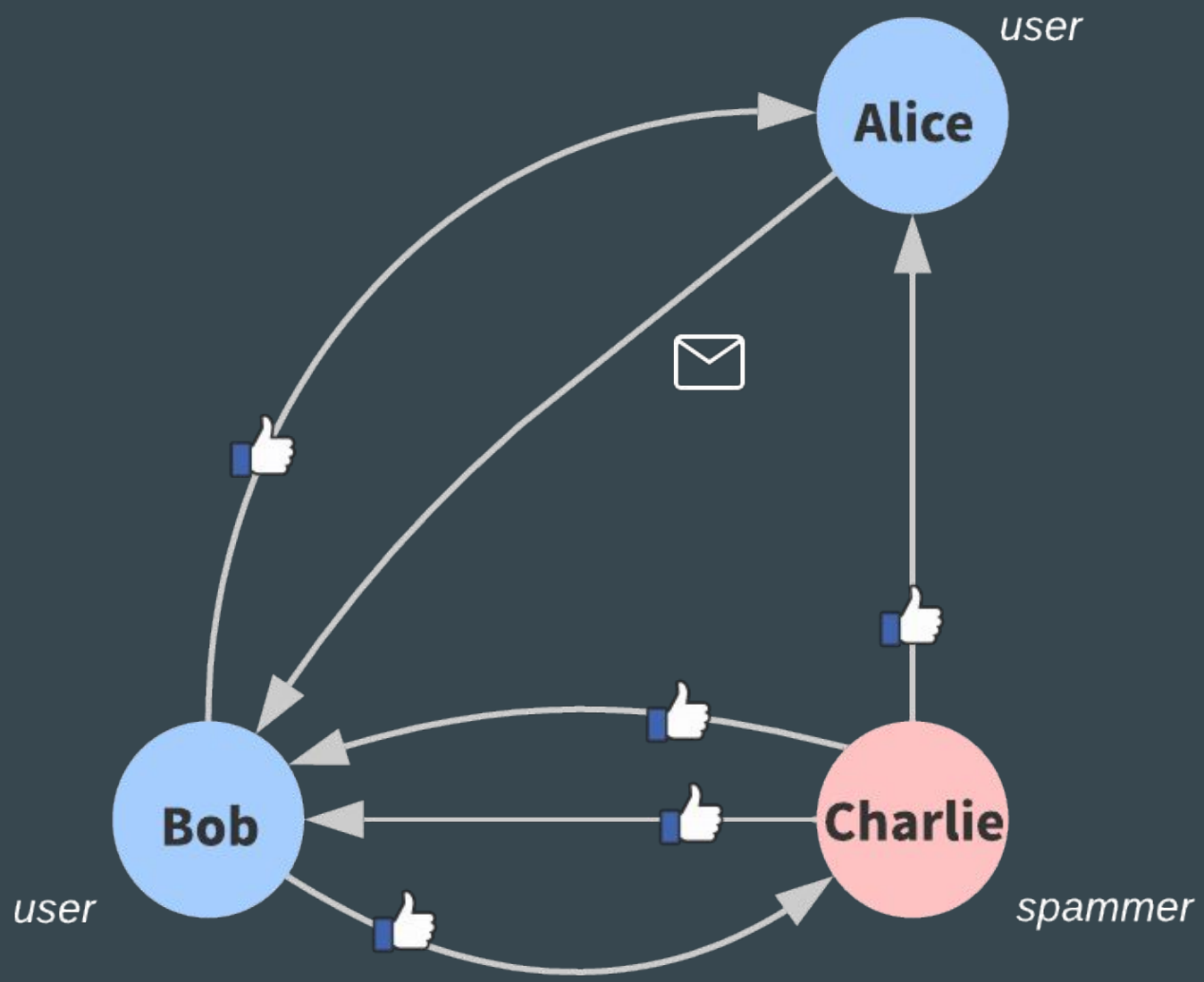
I. Visualization

II. Exploration & Extraction

III. Exploitation

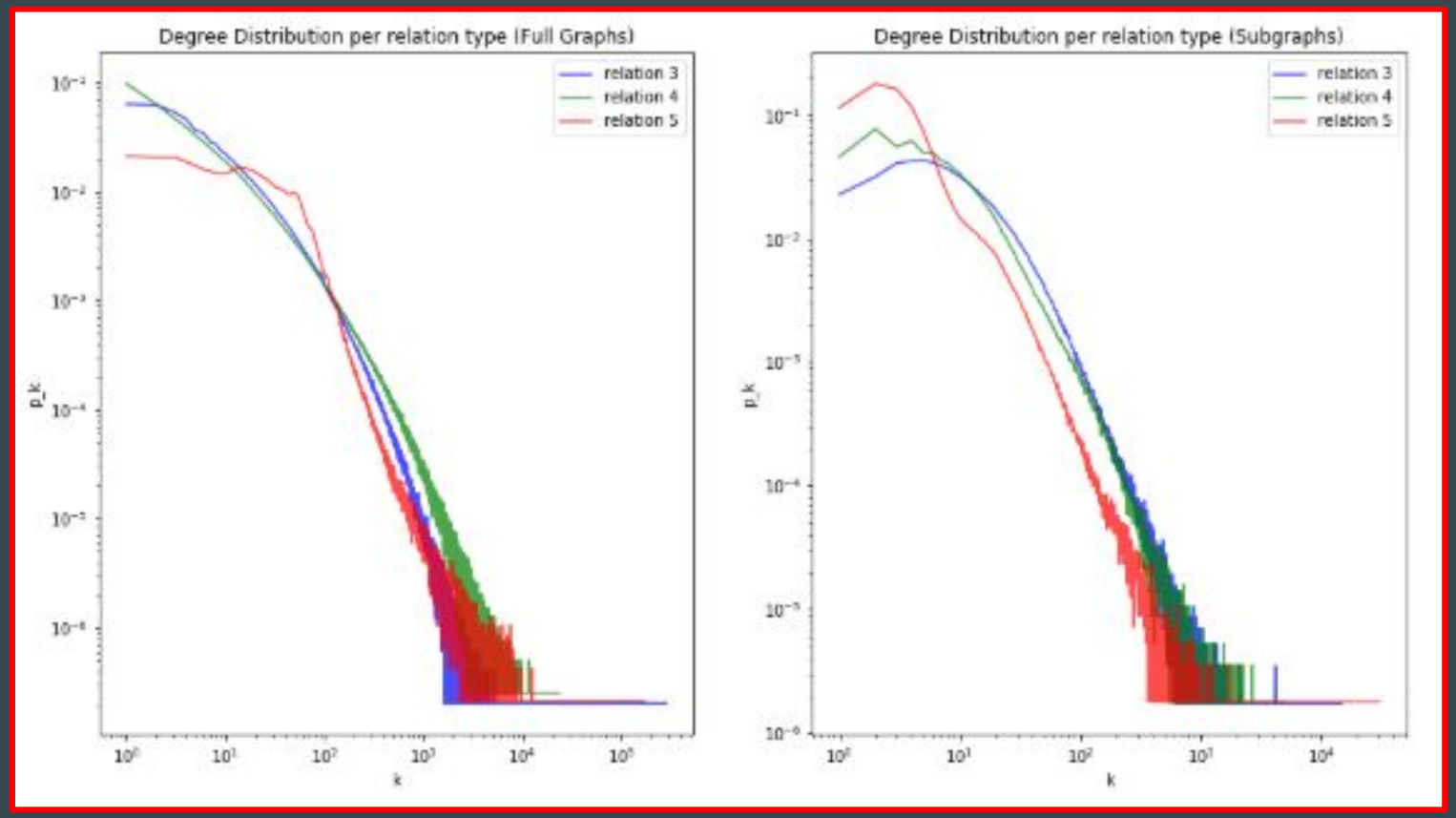
Full Dataset

- Preprocessing
- Local Features
  - General
    - **dtot_# (#=👍,✉ ...)**
    - **dout_#**
  - New
    - **duni_#**
    - **dnbi_#**
    - **n_bidir**
- Global Features*
  - Triangle count



SAMPLING

- Traversal method: breadth first search



Stretched exponential distributions for full relations 3,4 and 5. Assumption that it is equivalent for the relation 6.

Sampling :

- lack of growth
- peak in distribution

FEATURE EXTRACTION BY RELATION

- Degree, in_degree and out_degree
- Clustering coefficients
- PageRank
- Graph Coloring
- Triangle Count

I. Visualization

II. Exploration & Extraction

III. Exploitation

III. Exploitation

- Pre-processing
- Models
- Metric
- Results

Pre-processing

- The dataset is too “big” (~5M rows) to fit in memory and train models on it efficiently
- High class imbalance which is common in fraud/spam detection

Pre-processing

- Chosen solution:
  - Randomly sample 100.000 rows with balancing the labels
  - Classical standardization + One-hot encoding

Models

Various models were considered, mainly tree-based ones:

- **PCA + KNN:** Uses the K closest points in the feature space for classification
- **Random Forest:** Ensemble learning method that considers multiple decision trees for classification
- **PCA + Random forest**
- **Gradient Boosting/XGBoost:** Boosting to turn ensemble of weak classifiers into a strong one
- **Ridge regression:** Regularized linear classifier

Models

Followed methodology:

- Only keep the graph-based features at first
- Split the dataset into train/test split (80/20)
- Grid search + Cross validation
- Evaluate best model on the test split
- Add user based features and repeat

Evaluation metrics

- **F1-score**
- **AUROC:** Area under the tp_rate vs. fp_rate curve
- **AUPR:** Area under the precision vs. recall curve

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$F_1 = \left(\frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \right) = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Results: Graph-based features

- Ensemble methods have the upper hand
- “Low” scores for Ridge Regression might indicate lack of linear separability
- PCA might not be the best of ideas

Model	AUROC	AUPR	F1-score
PCA + KNN	0.78	0.80	0.71
PCA + Random Forest	0.80	0.82	0.73
Random Forest	0.81	0.83	0.73
XGBoost	0.81	0.83	0.74
Gradient Boosting	0.81	0.83	0.74
Ridge Regression	0.70	0.72	0.62

Results: Adding user data features

- Clear importance of user data
- Once again, slight advantage to boosted methods

Model	AUROC	AUPR	F1-score
PCA + KNN	0.82	0.84	0.76
PCA + Random Forest	0.83	0.85	0.78
Random Forest	0.85	0.87	0.77
XGBoost	0.85	0.87	0.77
Gradient Boosting	**0.85**	**0.87**	**0.78**
Ridge Regression	0.80	0.82	0.73

Zoom on Gradient Boosting

- Better results overall for the non-spammer class
- Low recall for the spammers

	precision	recall	f1-score	support
0	0.74	0.89	0.81	10422
1	0.84	0.66	0.74	9578
accuracy			0.78	20000
macro avg	0.79	0.77	0.77	20000
weighted avg	0.79	0.78	0.78	20000

Improvements

- Graph-Based Features
  - From whole dataset
  - Total unique neighbors (differs from sum of unique neighbors per relation)
  - N-times degenerate edges
  - Triangle count on bidirectional edges only
- Extract sequence-based features (k-gram, Markov models...) to capture the order of interactions
- Run the models on larger samples
- Extensive feature selection

Questions?