The Bitcoin Blockchain provides a fully auditable history of a peculiar economy. Nearly 300,000,000 transactions have been processed by the Bitcoin network to date. Misconceptions abounded for a long time about Bitcoin’s anonymity. Criminal behavior persists. Strange things happen.

**Overview**

- The Bitcoin Blockchain provides a fully auditable history of a peculiar economy.
- Nearly 300,000,000 transactions have been processed by the Bitcoin network to date.
- Misconceptions abounded for a long time about Bitcoin’s anonymity.
- Criminal behavior persists. Strange things happen.
- Large scale deanonymization efforts are available for analysis.

**Our Questions**

- Do users inadvertently leave a signature on their transactions?
- Can illicit behavior be separated from licit behavior in the Bitcoin Blockchain?
- Can we leverage the deanonymization efforts of others to build an analysis?

**Analysis Approach**

- Collect de-anonymized transactions and addresses
- Extract features from user’s transactions
- Use the labels in a supervised learning scheme
- Use predictions to propose bitcoin market ‘sectors’
- Do these predicted sectors match intuition?
- Are any of these sectors illicit?
- Use these sectors as the labels in an additional supervised learning scheme.
- Compare with hand labeled sectors and illicit activity.

**Feature Extraction**

- Keep it fast and simple
- 1 hop or 2 hops
- Computed 81 features
- Transaction (tx) Value in dollars, # inputs, # outputs
- 2 hop features include, min, max, mean, median and sum over the set of 1 hop features of the outputs or inputs.
- Computed typical holding time of BTC
- DID NOT include absolute time
- Transactions need not be labeled to form these feature vectors

**Neural Networks**

- Neural Networks are universal approximators, can they learn the task of separating market participants?

**Neural Network Results**

- 81 features -> 144 labels (2,000,000 txs)
- 81 features -> 2 labels (pairwise, eg HelixMixer vs SatoshiDice, 1000-100000 txs)
- Tried various architectures, settled on Relu, Relu, Relu, ... Softmax with 6 hidden layers with 125, 330, 550, 225, 152 and 147 neurons in each layer.
- Shows considerable improvement to guessing.

- Number of inputs: 81
- Number of outputs: 144
- Loss Function: 4.1594
- Accuracy: 9.9%
- Random guessing: 0.6%

**Sector Results**

- 81 features -> 6 sectors (k means sectors)
- 81 features -> 6 hand labeled sectors (casinos, exchanges, markets, mixers, pools, services)
- Considerable improvement to random guessing.

<table>
<thead>
<tr>
<th># inputs</th>
<th># outputs</th>
<th>loss function</th>
<th>accuracy</th>
<th>random guessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>6</td>
<td>1.57</td>
<td>38.9%</td>
<td>16.6%</td>
</tr>
<tr>
<td>81</td>
<td>6</td>
<td>1.46</td>
<td>38.1%</td>
<td>16.6%</td>
</tr>
</tbody>
</table>

**Illicit vs Licit**

<table>
<thead>
<tr>
<th># inputs</th>
<th># outputs</th>
<th>loss function</th>
<th>accuracy</th>
<th>random guessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>2</td>
<td>0.392</td>
<td>86.3%</td>
<td>50%</td>
</tr>
</tbody>
</table>

**Contributions to Analyst Workflow**

- Sort or Trim a list of Bitcoin Transactions
- Automatically search for transactions with various characteristics

**Acknowledgements**

Special thanks to Peter Merrill of LAS for contributions to the direction of research. I would also like to thank Melody Graves, Stephanie Beard, Matthew Schmidt, and Trey Overman from LAS and John Harer, Paul Bendich, Anastasia Deckard, and Kenneth Ball from GDA. This work was supported through contract DoD-H98230-13-D-0054 LAS D07.