



Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain

By:

Cuneyt Akcora (University of Manitoba)

From the proceedings of

The Digital Forensic Research Conference

DFRWS USA 2021

July 12-15, 2021

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment.

As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

<https://dfrws.org>

BitcoinHeist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain

Akcora, Gel, Kantarcioglu



**University
of Manitoba**




Ransomware is a type of malware that infects a victim's data and resources and demands ransom to release them.



Image: gdatasoftware.com

In two main types, ransomware can lock access to resources or encrypt their content.

After the ransom payment, a decryption tool is delivered to the victim.

 MUST READ: [What is neuromorphic computing? Everything you need to know about how it is changing the future of computing](#)

Why ransomware is still so successful: Over a quarter of victims pay the ransom

Organisations are paying an average of \$1m to cyber criminals to restore their networks after falling victim to ransomware.

The combination of strong and well-implemented **cryptographic techniques** to take files hostage, the **Tor protocol** to communicate anonymously, and the use of a **cryptocurrency** to receive unmediated payments provide altogether a high level of impunity for ransomware attackers.

Paquet-Clouston, “Ransomware payments in the Bitcoin ecosystem (2019)”

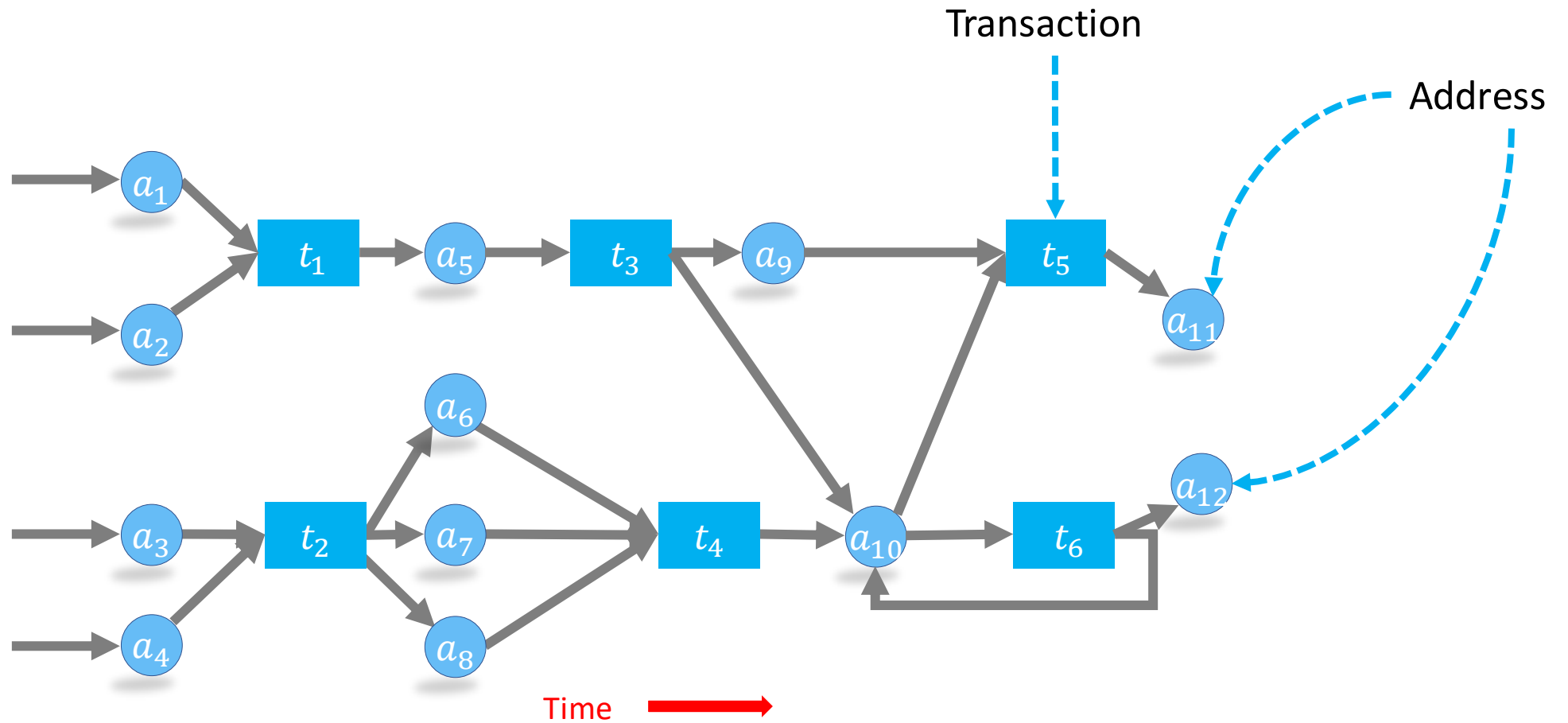
There is more bad news..

41% of the victims pay the ransom to recover their files.

Only 3% was conjectured by Symantec, 0.4% by Dell SecureWork.

Hernendex-Castro, The 2nd kent cyber security survey (2014)

Bitcoin transaction network is public – we can see all coin transfers.



Can we identify ransomware victims automatically?

Our two tasks!

Can we discover new
ransomware families?

On Bitcoin



Our ransomware dataset is a union of datasets from three widely adopted studies:

Montreal, Princeton and Padua.

The combined dataset contains 24,486 addresses from **27 ransomware** families.

Huang, D.Y., Aliapoulios, M.M., Li, V.G., Invernizzi, L., Bursztein, E., McRoberts, K., Levin, J., Levchenko, K., Snoeren, A.C. and McCoy, D., 2018, May. Tracking ransomware end-to-end. In *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 618-631). IEEE.

Paquet-Clouston, M., Haslhofer, B. and Dupont, B., 2019. Ransomware payments in the bitcoin ecosystem. *Journal of Cybersecurity*, 5(1).

Conti, M., Gangwal, A. and Ruj, S., 2018. On the economic significance of ransomware campaigns: A Bitcoin transactions perspective. *Computers & Security*, 79, pp.162-189.

We divide the Bitcoin network into 24-hour long windows by using the UTC-6 timezone as reference.

On the Bitcoin network, an address may appear **multiple times**.

An address u that appears in a transaction at time t can be denoted as a_u^t .

Let $\{a_u\}_{u \in \mathcal{Z}^+}$ be a set of addresses and let each address a_u be associated with a pair (\vec{x}_u, y_u) , where $\vec{x}_u \in \mathcal{R}^D$ is a vector of its features and y_u is its label.

The label y_u can designate a **white** (i.e., non-ransomware) address or a **ransomware** address.

Let f_1, \dots, f_n be labels of known ransomware families which have been observed until time point t .

We set f_0 to be the label of addresses which are **not known** to belong to any ransomware family, and we assume them to be **white addresses**.

Assumption: those addresses that we do not know as ransomware are white (non-ransom) addresses.

On the heterogeneous Bitcoin network, in each snapshot we extract the following six features for an address:

Income of an address u is the total amount of coins output to u : $I_u = \sum_{t_n \in \Gamma_u^o} A_u^o(n)$.

Neighbors of an address u is the number of transactions which have u as one of its output addresses: $|\Gamma_u^i|$.

Income and neighbors do not consider position of the address in the network!

We define the next four address features by using address position in a defined 24-hour time window.

For each window, we first locate the set of transactions that **do not receive outputs** from any earlier transaction within the studied window t , i.e.,

$$\mathbb{TX} = \{\forall tx_n \in TX, s.t., \Gamma_n^i = \{a_1^{t^0}, \dots, a_z^{t^n}\}, t^0 \leq t^n < t\}.$$

$$\mathbb{TX} = \{\forall tx_n \in TX, s.t., \Gamma_n^i = \{a_1^{t^0}, \dots, a_z^{t^n}\}, t^0 \leq t^n < t\}.$$

These transactions consume outputs of transactions that have been generated in previous windows.

For simplicity, we refer to a transaction $tx \in \mathbb{TX}$ as a **starter transaction**.

Weight of an address u , W_u , is defined as the sum of the fraction of coins that originate from a starter transaction and reach u .

Length of an address u , L_u , is the number of non-starter transactions on its longest chain.

A length of zero implies that the address is an output address of a starter transaction.

Count of an address u , C_u is the number of starter transactions which are connected to u through a chain (path).

Loop of an address u , O_u is the number of starter transactions which are connected to u with more than one directed path.

Putting address features together!

Before time point t , if we observe l addresses a_1, \dots, a_l , then we form their $D \times l$ -matrix of features $X_t = \{\vec{x}_1, \dots, \vec{x}_l\}$ and a vector of labels $Y_t = \{y_1, \dots, y_l\} \in \{f_0, f_1, \dots, f_n\}$.

Table 1: Most frequent feature values in ransomware addresses.

Len	Wei	Nei	Cou	Loo	Inc	# addresses	OverallRank
0	0.5	2	1	0	1	327	1
0	0.5	2	1	0	1.2	250	113
0	1	2	1	0	1	189	4
0	1	1	1	0	0.5	178	9
0	0.5	2	1	0	0.8	160	116
0	1	1	1	0	1	146	3
0	1	2	1	0	1.2	127	121
0	0.5	2	1	0	1.25	119	327
0	0.5	1	1	0	0.5	118	6
0	1	1	1	0	2	117	18

Most payments are N-1 or N-2!

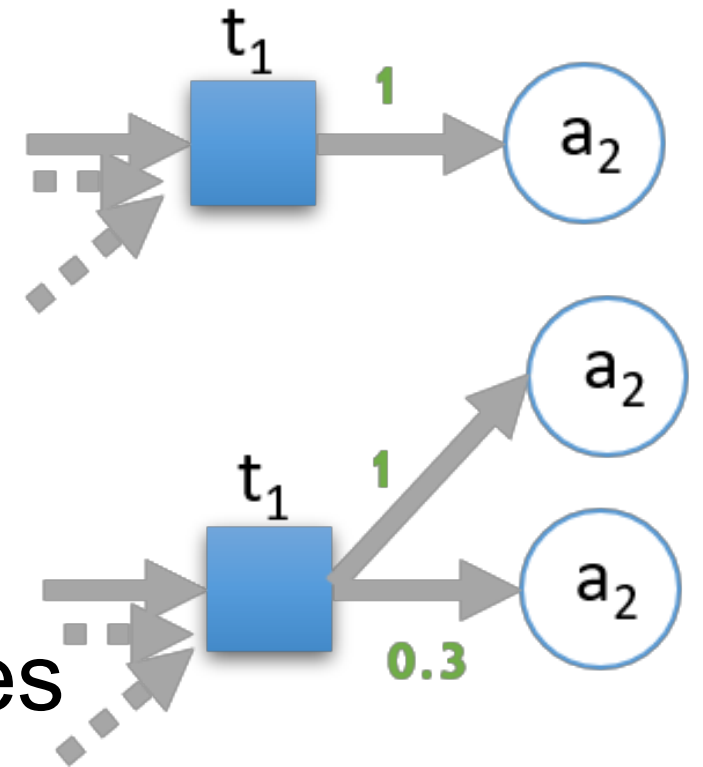
Length 0: The first transaction involving these coins in the day.

Weight 1: All output goes into the address.

Neighbor 1: One transaction makes a payment into the address.

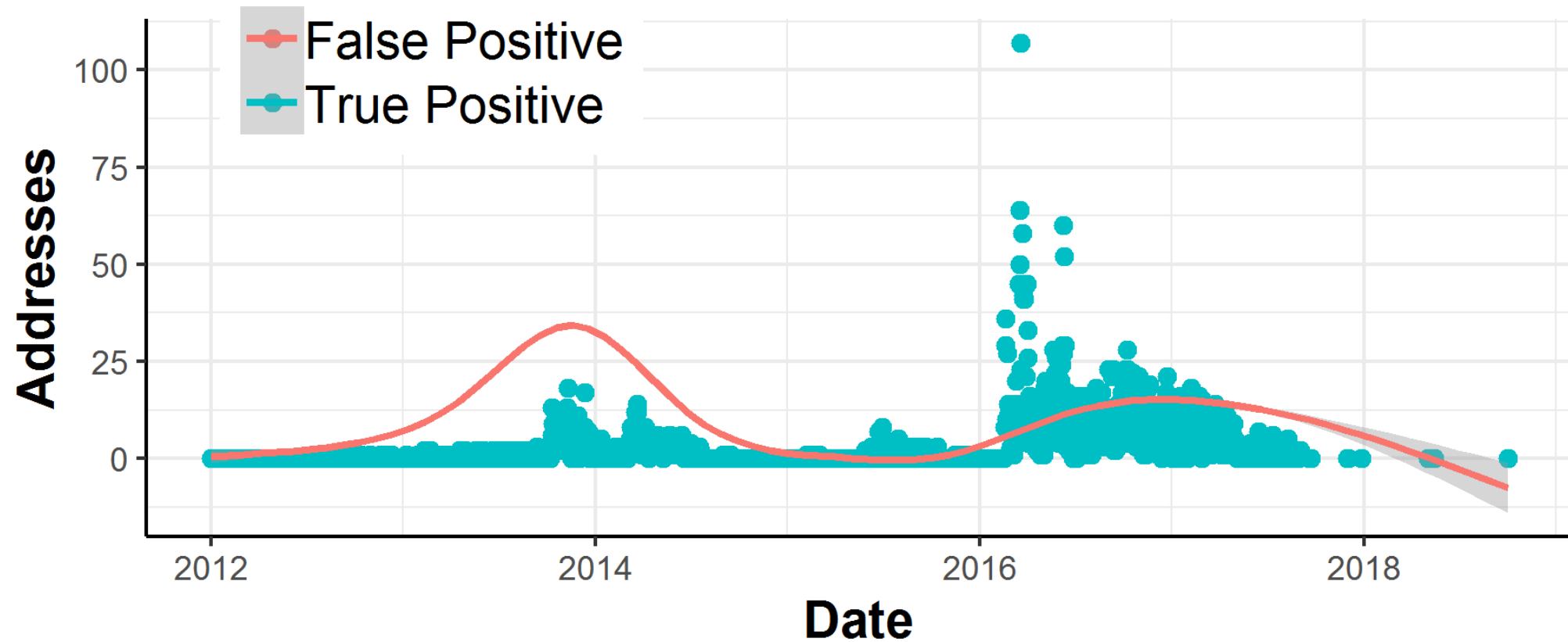
Count 1: One starter transaction reaches the address.

Loop 0: No obfuscation, coins are directly paid.

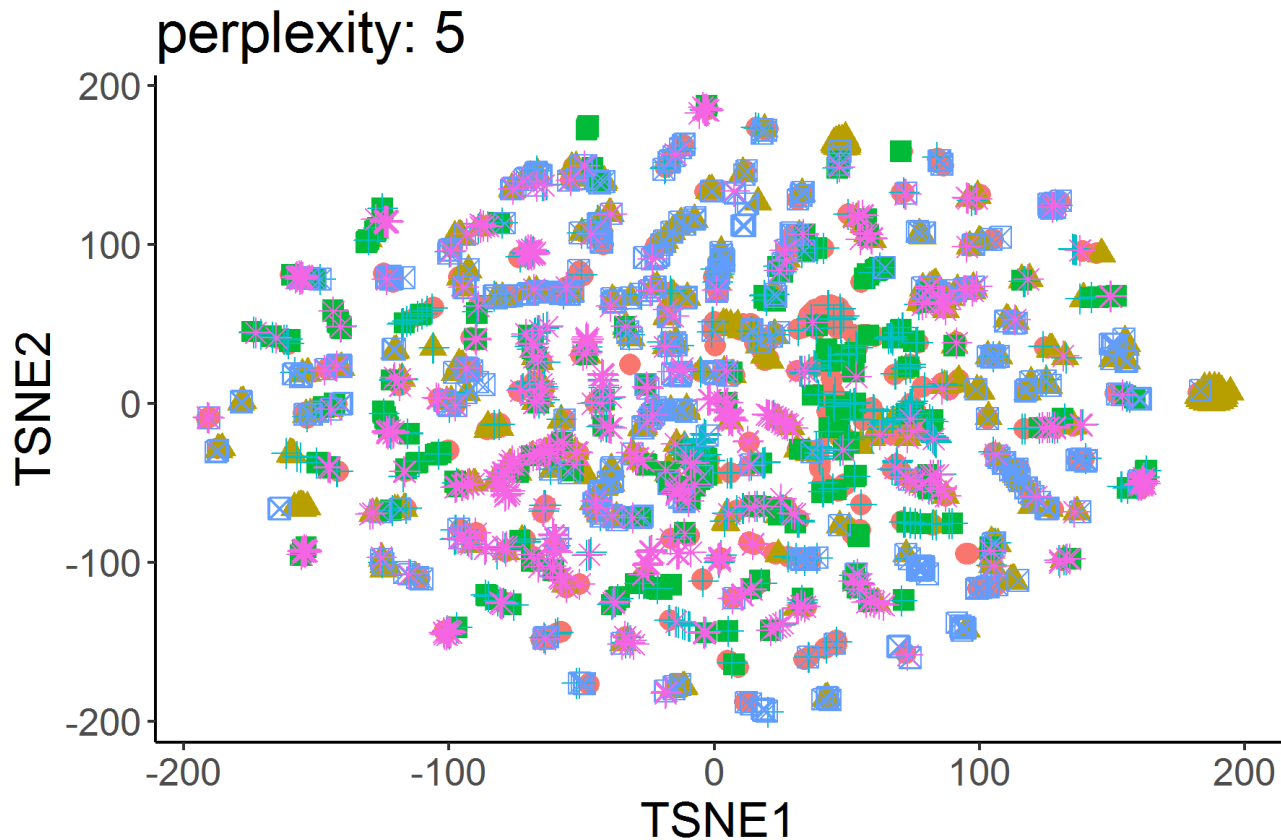


Experiment 1: Detecting undisclosed payments

Naïve approach: Similarity search all history. Not so bad!

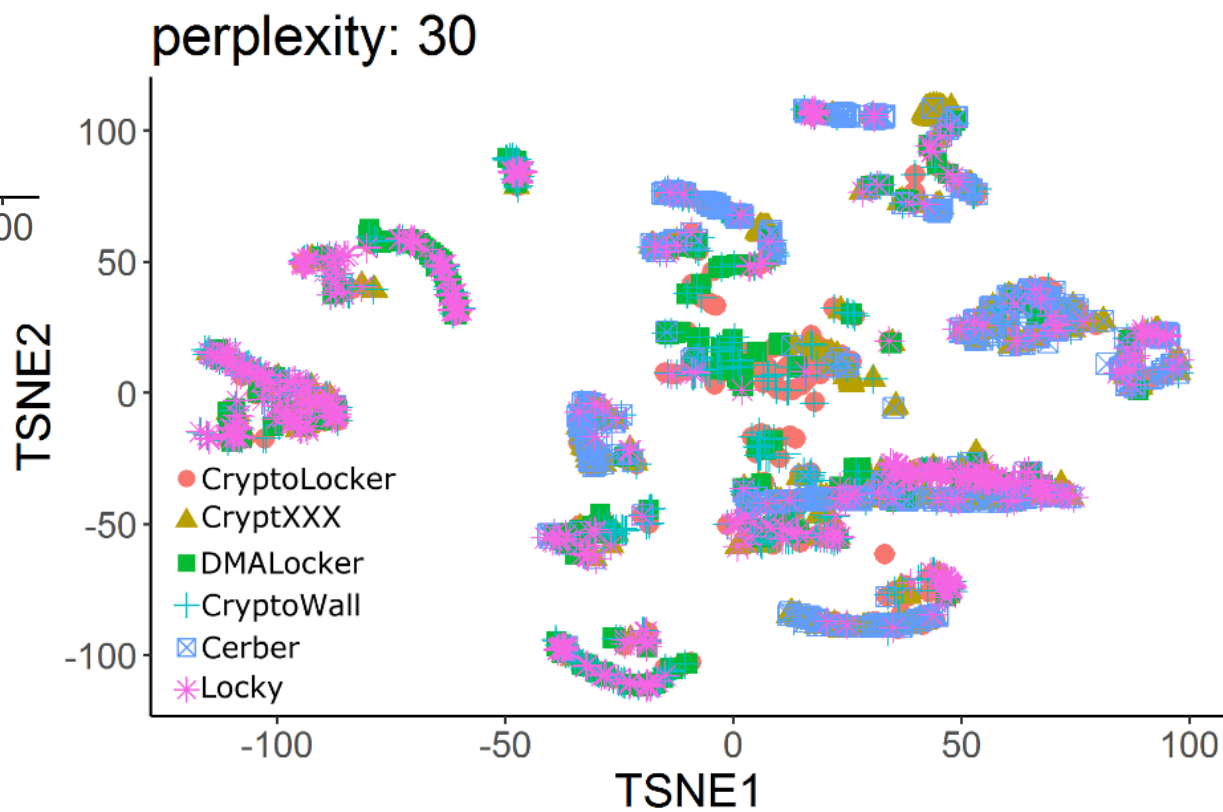


However, this naive approach creates 21,371 FP addresses overall.

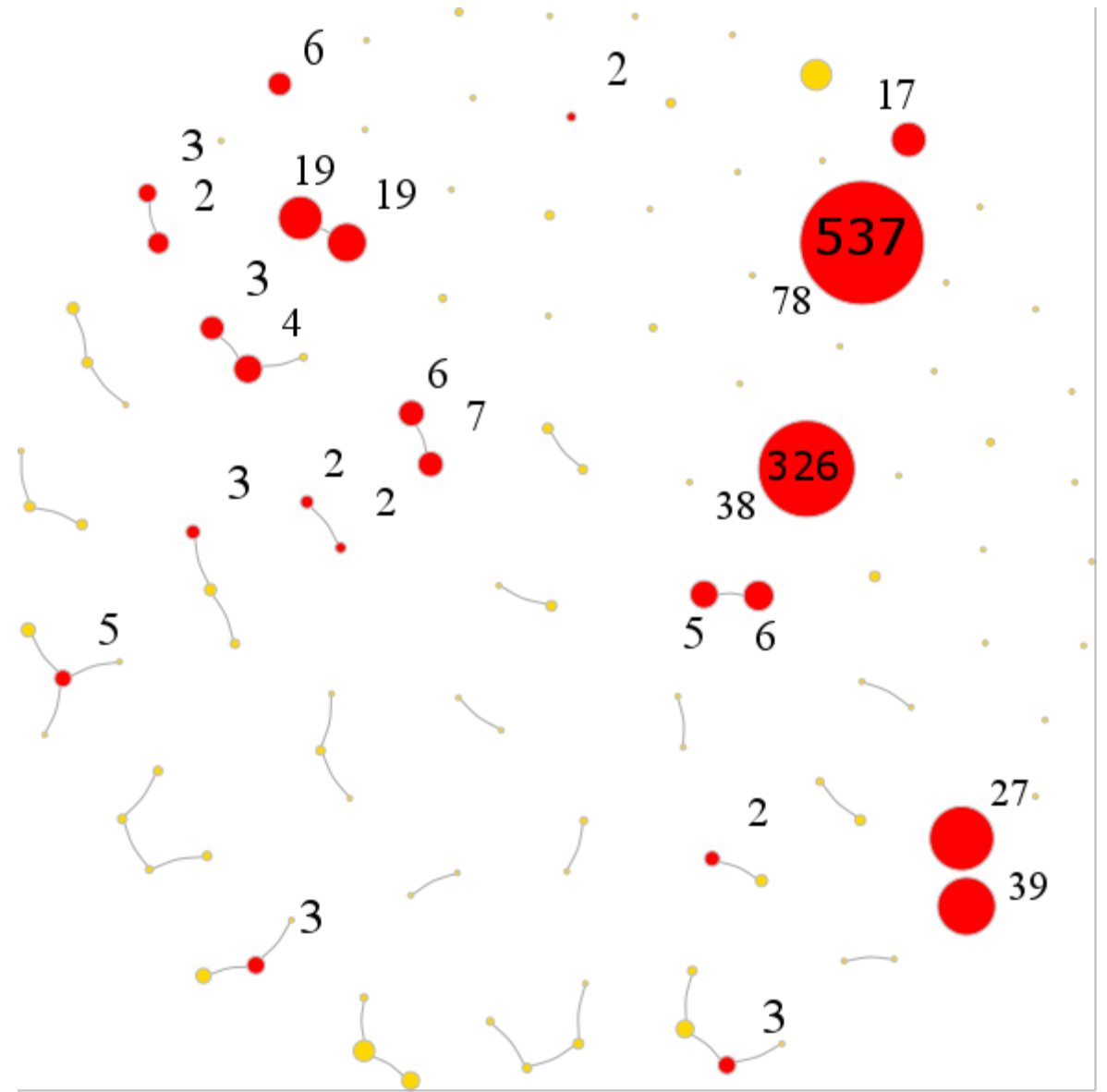


Address patterns
are diverse!

T-Stochastic neighbor
embeddings of
ransomware
addresses



We apply Topological Data Analysis for ransomware payment detection and compare our node classification results to ML techniques.



Problem: Network **node classification** with past labeled data.

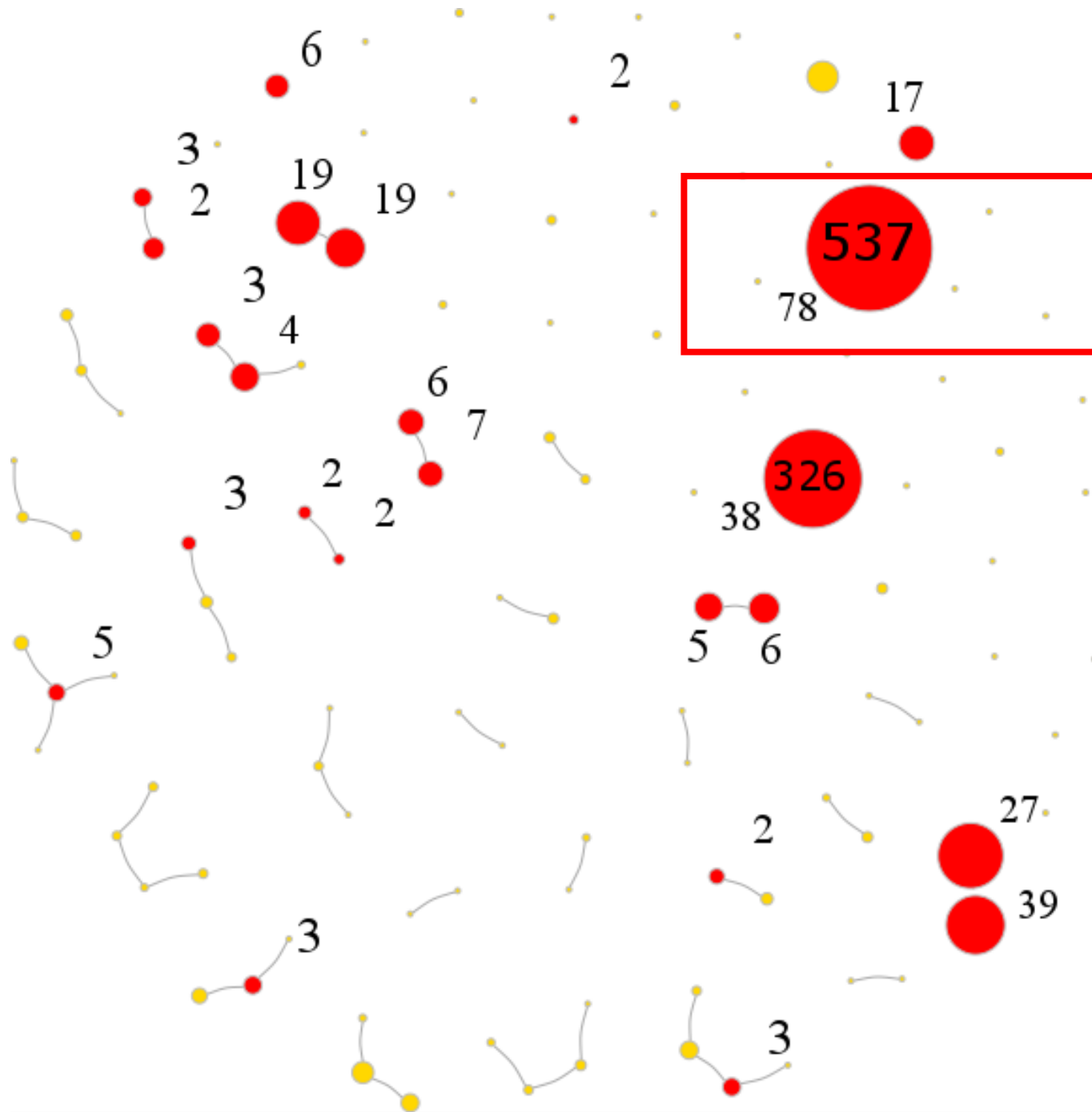
1. Naïve Cosine similarity search
2. Transition and co-spending heuristics
3. Tree based methods: XGBoost, Random Forest
4. Clustering: DBSCAN, K-means

5. TDA Mapper

The key idea behind Mapper is the following:

- Let U be a total number of observed addresses and $\{\vec{x}_u\}_{u=1}^U \in \mathcal{R}^D$ be a data cloud of address features.
- Select a **filter function** $\xi: \{\vec{x}_u\}_{u=1}^U \rightarrow \mathbb{R}$.
- Let I be the range of ξ , that is, $I = [m, M] \in \mathbb{R}$, where $m = \min_u \xi(\vec{x}_u)$ and $M = \max_u \xi(\vec{x}_u)$.

- Now place data into overlapping bins by dividing the range I into a set S of smaller **overlapping** intervals of uniform length.
- Let $u_j = \{u: \xi(\vec{x}_u) \in I_j\}$ be addresses corresponding to features in the interval $I_j \in S$.
- For each u_j perform a single linkage clustering to form clusters $\{u_{jk}\}$.



537 addresses
78 of which are
known past
ransomware
addresses.

In BitcoinHeist, we did not consider
the edge information of the network.

- If current addresses are contained in clusters that also **contain** many **past** known **ransomware** addresses, by **association**, we deem these current addresses potential ransomware addresses.
- We filter the TDA mapper graph by using each of our six graph features. As a result, we get six filtered graphs $\mathcal{CT}_1, \dots, \mathcal{CT}_6$ for each time window.
- Afterwards, we **assign** a suspicion, or **risk score** to an address a_u .

Experiment 1: Detecting undisclosed payments

- ML Methods: TDA gives the best F1. For each ransomware family, we predict 16.59 false positives for each true positive.
- In turn, this number is 27.44 for the best non-TDA models.

RS	Method	l	N	TP	FP	FN	TN	#w	Prec	Rec	F1	PLR
Locky	TDA $_{.9}^{.8 .5}$	240	300	451	2350	50	8221	11	0.161	0.900	0.273	0.192
	COSINE	90	300	2395	41681	3990	146369	194	0.054	0.375	0.095	0.057
Crypto Wall	TDA $_{.9}^{.8 .65}$	240	600	217	3087	155	11200	15	0.066	0.583	0.118	0.070
	DBSCAN $_{.2}$	240	600	728	18960	794	16913	59	0.037	0.478	0.069	0.038
Crypto Locker	TDA $_{.9}^{.65 .65}$	240	300	439	9686	212	22129	34	0.043	0.674	0.081	0.045
	DBSCAN $_{.15}$	60	300	935	42771	295	11316	67	0.021	0.760	0.042	0.022
Cerber	TDA $_{.9}^{.5 .35}$	120	300	187	5174	459	23027	29	0.035	0.289	0.062	0.036
	XGBOOST	240	300	1606	47307	7279	374169	436	0.033	0.181	0.056	0.034
Crypt XXX	TDA $_{.9}^{.35 .35}$	90	300	77	2460	271	11057	14	0.030	0.221	0.053	0.031
	COSINE	30	600	589	20872	610	42952	65	0.027	0.491	0.052	0.028

Experiment 2: Predicting a new family

RS	Method	Prec	Rec	TN	FP	TP	FN	PLR
CryptXXX	$\text{TDA}_{0.9}^{0.2 0.2}$	0.500	0.026	917	1	1	37	1.0
	COSINE	0.046	0.342	654	264	13	25	0.049
Locky	COSINE	0.098	0.138	795	37	4	25	0.108
	$\text{TDA}_{0.9}^{0.05 0.95}$	0.047	0.586	489	343	17	12	0.049
CryptoWall	$\text{TDA}_{0.9}^{0.05 0.95}$	0.0625	0.500	810	165	11	11	0.067
	$\text{TDA}_{0.9}^{0.35 0.8}$	0.061	0.500	805	170	11	11	0.0647
Cerber	$\text{TDA}_{0.9}^{0.05 0.95}$	0.029	0.214	849	100	3	11	0.030
	$\text{TDA}_{0.9}^{0.35 0.8}$	0.023	0.642	570	379	9	5	0.023
DMALocker	$\text{DBSCAN}_{0.2}$	0.019	0.875	120	367	7	1	0.019
	$\text{DBSCAN}_{0.15}$	0.015	0.875	4	459	7	1	0.015

In CryptXX we catch two addresses, one is a TP!

In general, we predict 27.53 false positives for each true positive

Through some ~~black magic~~ Topological Data Analysis methods

In locating ransomware addresses

We predict **16.59 false positive** ransom addresses
for each true positive

In identifying new ransomware families

We predict **27.53 false positive ransom** addresses
for each true positive

Among 600K Bitcoin addresses daily!

Dataset



BitcoinHeistRansomwareAddressDataset

Download: [Data Folder](#), [Data Set Description](#)

Abstract: BitcoinHeist datasets contains address features on the heterogeneous Bitcoin network to identify ransomware payments.

Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	2916697	Area:	Computer
Attribute Characteristics:	Integer, Real	Number of Attributes:	10	Date Donated	2020-06-17

Cuneyt.Akcora@umanitoba.ca

Cuneyt G. Akcora¹, Yitao Li², Yulia R. Gel², Murat Kantarcioglu²

¹University of Manitoba, Canada

²University of Texas at Dallas, USA