LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND STATISTIK
INSTITUT FÜR INFORMATIK

LEHRSTUHL FÜR DATENBANKSYSTEME
UND DATA MINING

# On the topological property of dynamic transaction graph

Master Thesis

Yuhao Wang

22.03.2021

Scientific Supervisor: **Dr. Yunpu Ma**
Responsible professor: **Prof. Dr. Volker Tresp**
Company**: secunet Security Networks AG**

DBS

# Motivation

- Bitcoin transaction
- Cryptocurrencies
- Dynamic graph
- Money laundering

# Background

- Blockchain

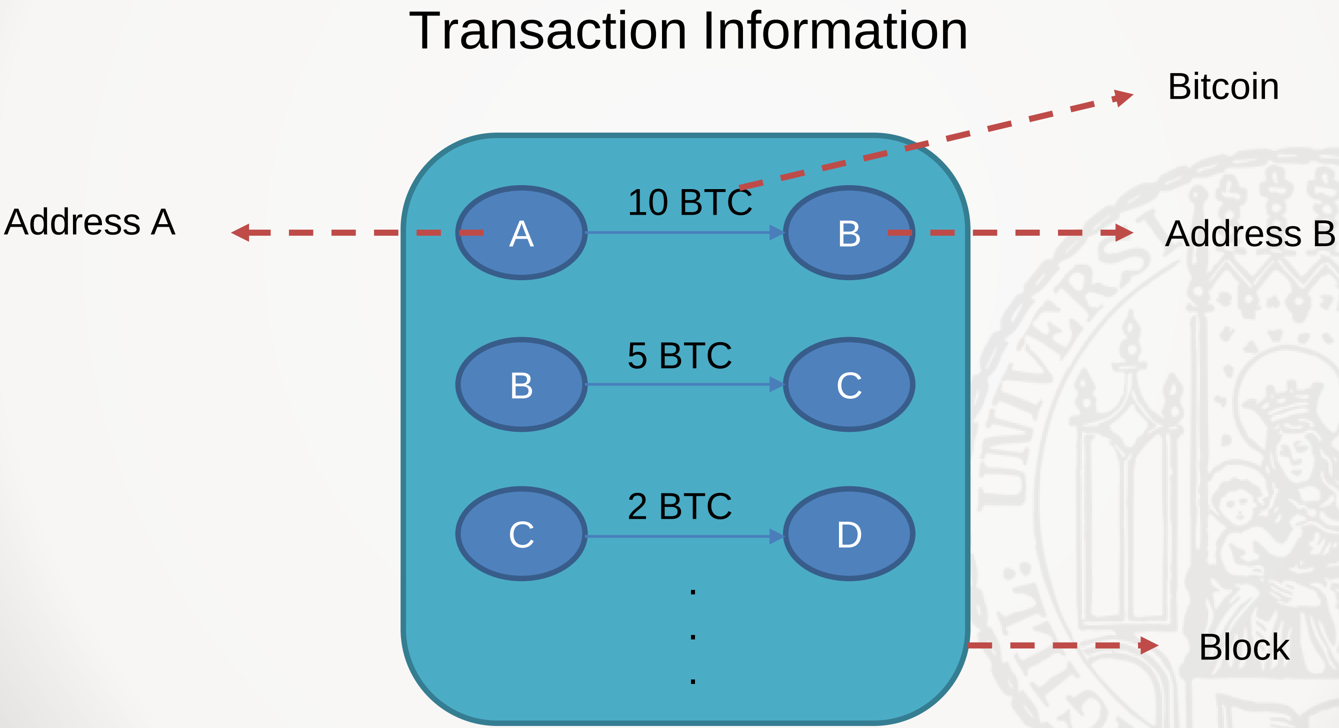
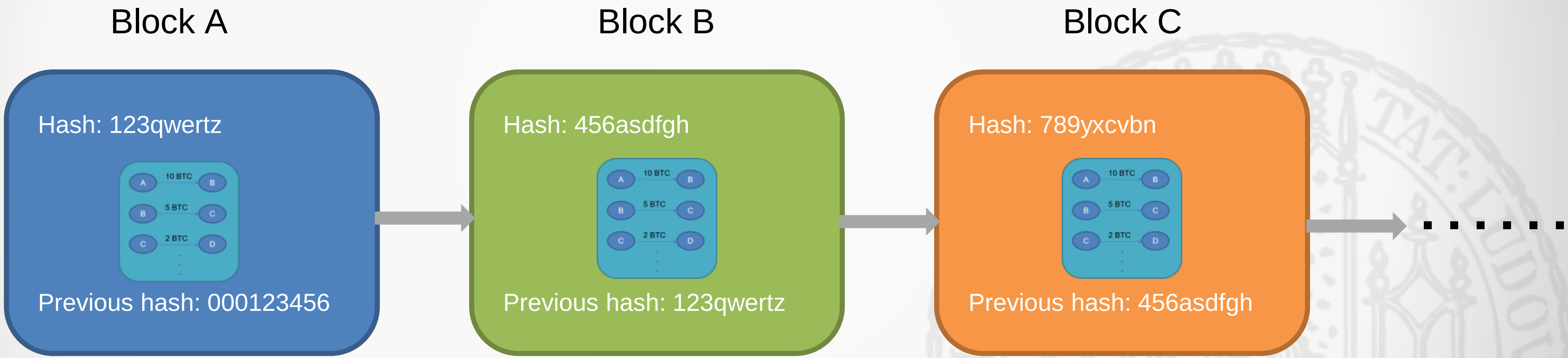


Fig. 1: A example of Block

# Background

- Blockchain



Block A

Hash: 123qwertz

Previous hash: 000123456

Block B

Hash: 456asdfgh

Previous hash: 123qwertz

Block C

Hash: 789yxcvbn

Previous hash: 456asdfgh

Fig. 2: A example of a BlockChain with three blocks; A,B and C. The system produces a new block every 10 minutes.

# Methodology

## A. Learning Graph Representations

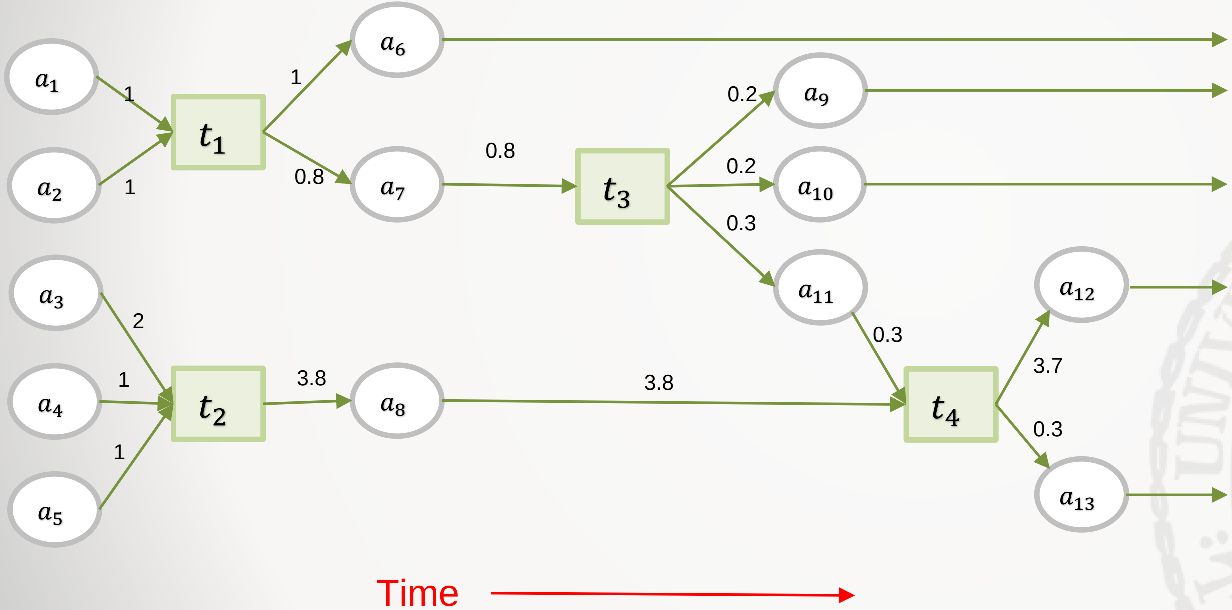


Fig. 3: A transaction-address graph representation of the Bitcoin network.

Denote:

- A: address
- T: transaction
- Value: the amount of the BTC

1. Cuneyt el all., Forecasting bitcoin price with graph chainlets, PAKDD, 2018

# Methodology

## A. Learning Graph Representations

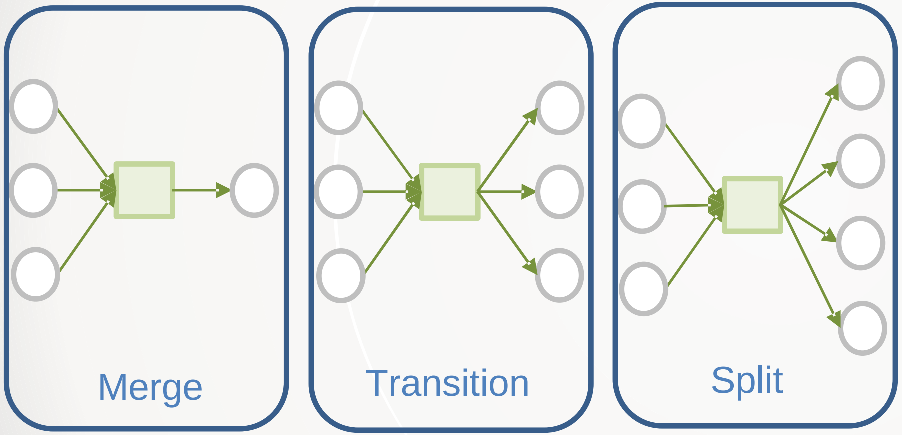


Fig. 4: Merge ($C_{3\to1}$), Transition ($C_{3\to3}$) and Split ($C_{3\to4}$) chainlets for 3 inputs. [1]

Denote:

$C_{x\to y}$ refers to x inputs and y outputs:

- Merge: x > y
- Transaction: x = y
- Split: x < y

---

1. Cuneyt el all., Forecasting bitcoin price with graph chainlets, PAKDD, 2018

# Methodology

## A. Learning Graph Representations

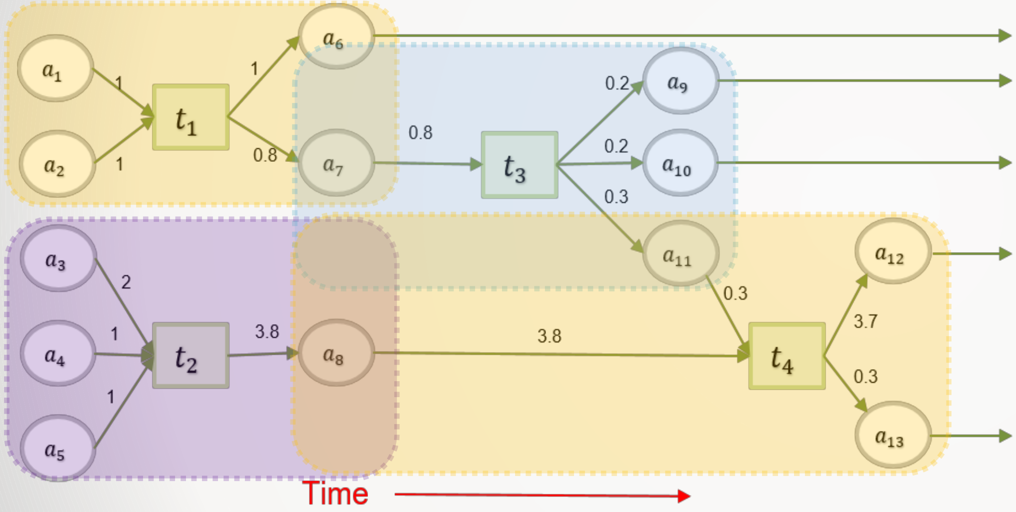


Fig. 5: A transaction-address graph representation of the Bitcoin network. [1]

1. Occurrence and Amount Matrices:

$$O = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 0 & 0.8 \\ 0 & 6.1 & 0 \\ 4 & 0 & 0 \end{bmatrix}$$

Occurence matrices:

- Row: the amount of the input
- Column: the amount of the output
- Value: the amount of the transaction with specific input and output

Amount matrics:

- Row: the amount of the input
- Column: the amount of the output
- Value: the value of the transfered BTC with specific input and output

---

1. Nazmiye el all., ChainNet: Learning on Blockchain Graphs with Topological Features, ICDM, 2019

# Methodology

## A. Learning Graph Representations

Graph Filtration (FL):

- Given the amount and occurrence information, a natural combination of them entails filtering the occurrence matrix with user defined thresholds on amounts, or filtering the amount matrix with user defined thresholds on occurrences.[1]

**Algorithm 1 FL: Graph Filtration**

```
Input: $\mathcal{G}$: Blockchain graph, time t, $\epsilon_{1,\ldots S}$: set of S filtration scales.
for $\epsilon \in \epsilon_{1,\ldots S}$ do
    $\mathcal{O}^{\epsilon} \leftarrow []$ // initialize occurrence matrix
for chainlet $\mathbb{C}_{i \to j} \in \mathcal{G}_t$ do
    for each scale $\epsilon \in \epsilon_{1,\ldots S}$ do
        if $\epsilon \leq amout\,(\mathbb{C}_{i \to j})$ then
            $\mathcal{O}_{ij}^{\epsilon} \leftarrow 1 + \mathcal{O}_{ij}^{\epsilon}$
return $x_t = [\mathcal{O}^{\epsilon_1 \cdots \epsilon_S}]$ // concatenated occ. matrices
```

---

1. Nazmiye el all., ChainNet: Learning on Blockchain Graphs with Topological Features, ICDM, 2019

# Methodology

## A. Learning Graph Representations: Absorbing random walks (ARW)

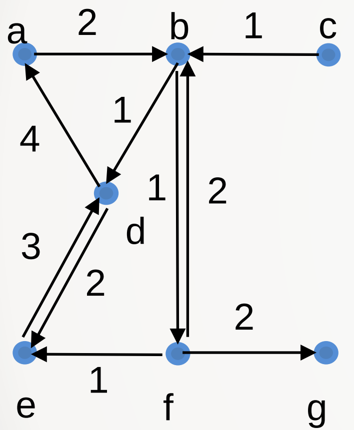


Fig. 6: A visualization of the transaction network.

| n | $L_{arw}(n)$ |
|---|---|
| A | 0,09 |
| B | 0,06 |
| C | 0,24 |
| D | 0,08 |
| E | 0,12 |
| F | 0,12 |
| G | 0,27 |

Tab. 1: The corresponding 1-ARW-betweenness centrality scores. To calculate the scores, the 1-ARW algorithm (Algorithm 1) is used, which will be explained further on.

$$N = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \end{bmatrix}$$

Matrix 1: The matrix representation of the ARW neighbor-probability distributions.[1]

---

1. Anouk van Schetsen el all., Impact of graph-based features on Bitcoin prices, Delft University of Technology, 2019

# Methodology

## B. Learning Topological Representations

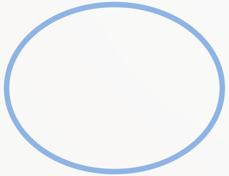
**Betti number:**

the *k*th Betti number refers to the number of *k*-dimensional *holes* on a topological surface.

Examples:

$b_0$ is the number of connected components;<br>
$b_1$ is the number of one-dimensional or "circular" holes;<br>
$b_2$ is the number of two-dimensional "voids" or "cavities".

$b_0$ = 1 0D loop
$b_1$ = 0 1D loop
$b_2$ = 0 2D loop



$b_0$ = 1 0D loop
$b_1$ = 1 1D loop
$b_2$ = 1 2D loop

$b_0$ = 1 0D loop
$b_1$ = 2 1D loop
$b_2$ = 1 2D loop

$b_0$ = 1 0D loop
$b_1$ = 0 1D loop
$b_2$ = 1 2D loop

---

1. Nazmiye el all., ChainNet: Learning on Blockchain Graphs with Topological Features, ICDM, 2019

# Methodology

## B. Learning Topological Representations

1. **Betti Sequences for a Blockchain Network.**
   1) $a'$ = log(1 + a/$10^8$), a is an amount of Satoshis.
   2) a k-th q-quantile, k = 0, 1, . . . , q, is the amount Q(k) such that:

      $\sum_{i=1}^{\tau} \mathbb{1}_{y_i<Q(k)} \approx \frac{\tau k}{q}$ and $\sum_{i=1}^{\tau} \mathbb{1}_{y_i>Q(k)} \approx \frac{\tau(q-k)}{q}$

      $$d_{ij} = \sqrt{\sum_{k=0}^{q} \left[Q_i(k) - Q_j(k)\right]^2}$$
   3) construct a sequence of scales 1 < 2 < . . . < S covering a range of distances during the entire 365- day period, the filtration of VR complexes $VR_1 \subseteq VR_2 \subseteq . . . \subseteq VR_S$.
   4) $x_t$ = {$\beta_0$ ($\epsilon_1$), . . . , $\beta_0$ ($\epsilon_S$), ; $\beta_1$ ($\epsilon_1$), , . . . , $\beta_1$ ($\epsilon_S$) }

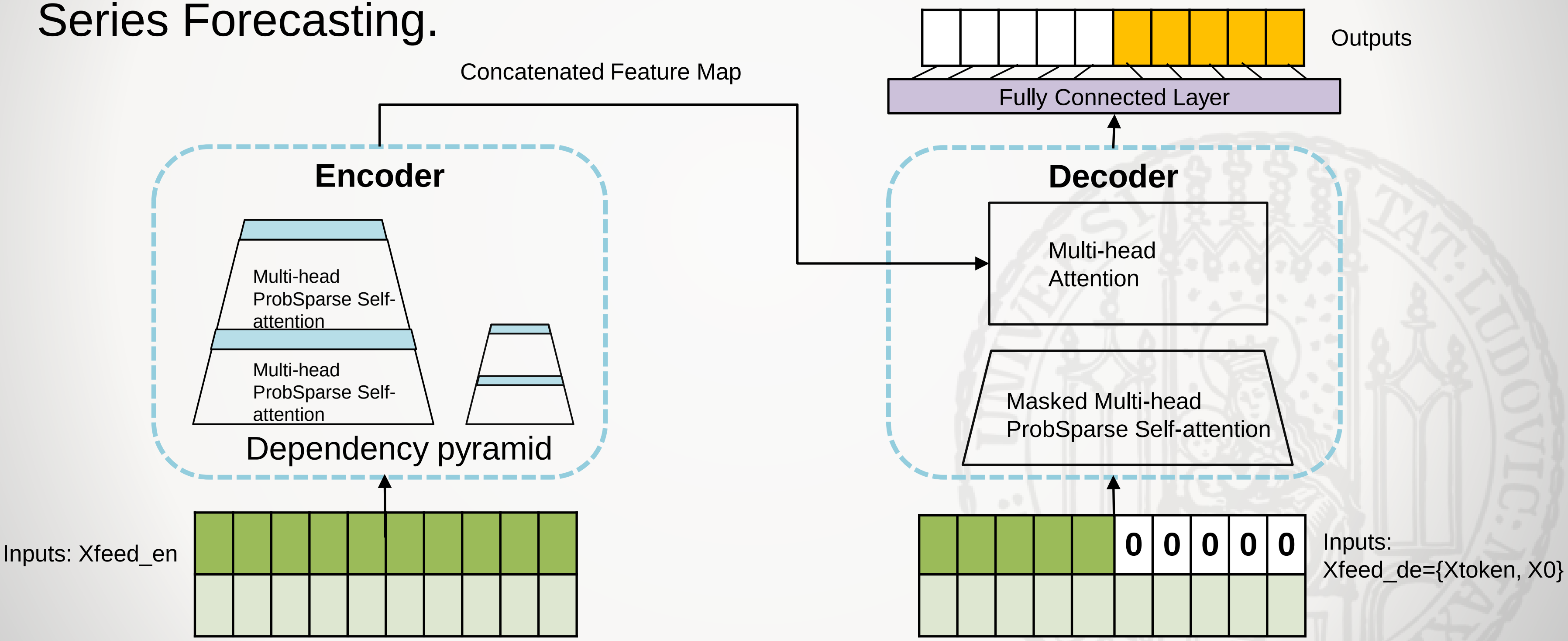
2. **Betti derivatives.**

   1) $\Delta^{\ell} \beta_{\mathrm{p}} (\epsilon_k) = \Delta^{\ell-1} \beta_{\mathrm{p}} (\epsilon_{k+1}) - \Delta^{\ell-1} \beta_{\mathrm{p}} (\epsilon_k)$

1. Nazmiye el all., ChainNet: Learning on Blockchain Graphs with Topological Features, ICDM, 2019

# Methodology

## C. Informer: a beyond efficient Transformer for Long Sequence Time-Series Forecasting.

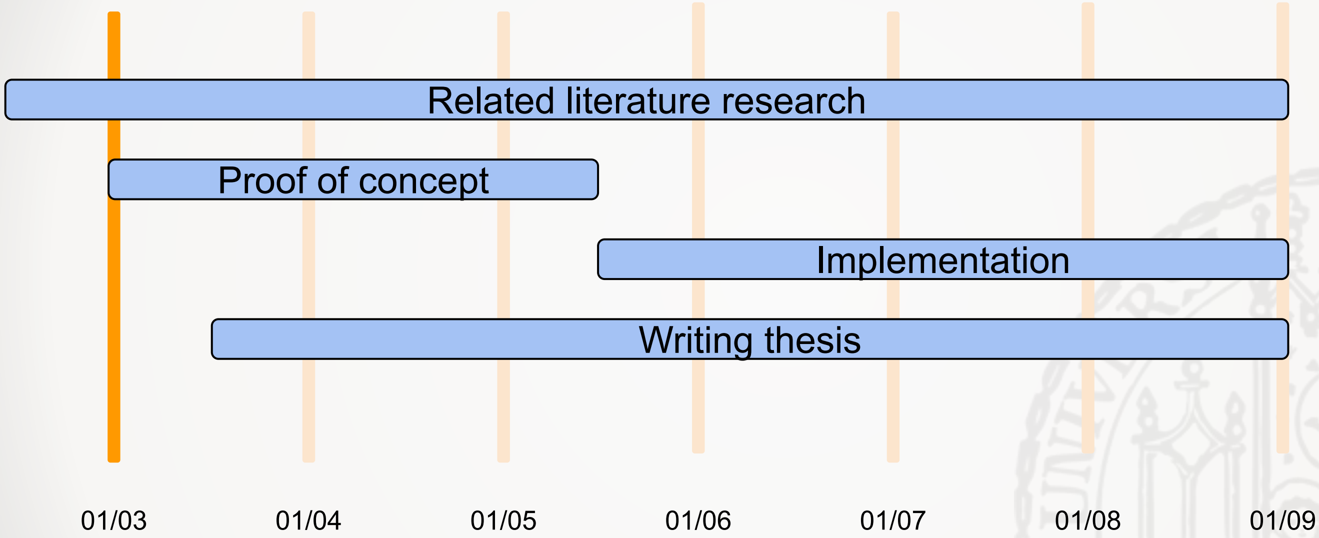


1. Haoyi el.all, Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. AAAI, 2021.

# Research goals

- Research the persistent effectiveness of the topological properties on the Blockchain after the explosion of the cryptocurrency trading.
- Explore the benefit of the Informer on the dynamic transaction graph.
- Examine the dynamic property of topological features on transaction graph.

# Time plan

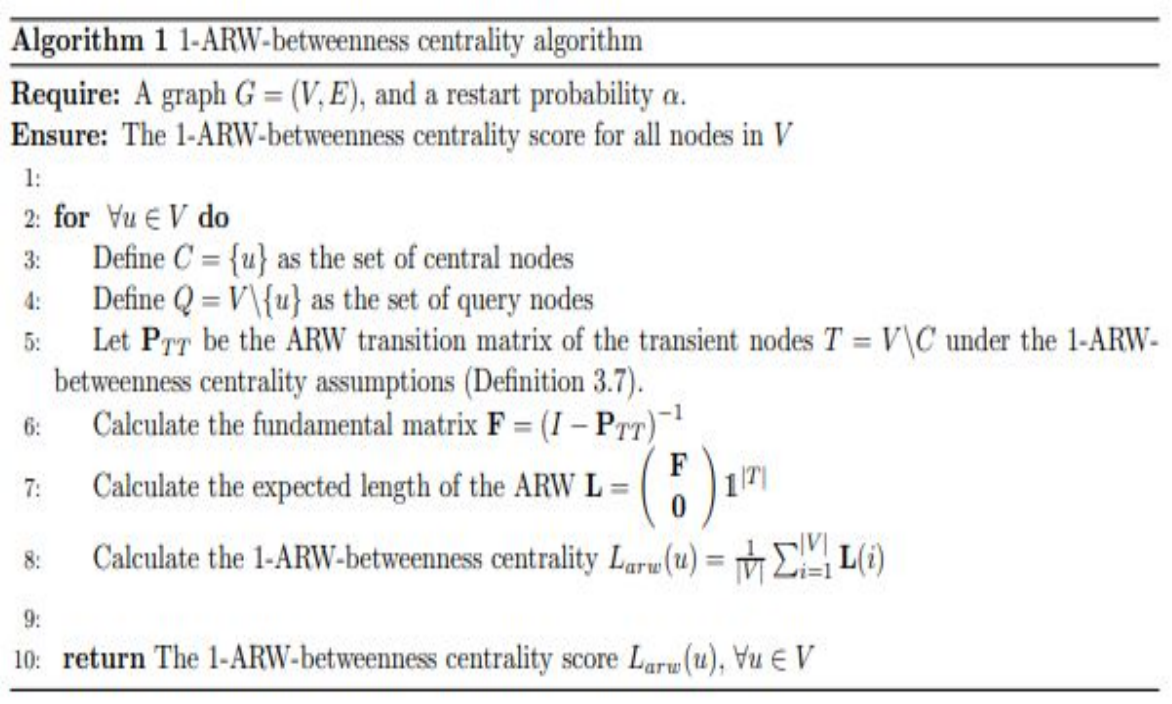

Related literature research
Proof of concept
Implementation
Writing thesis
01/03
01/04
01/05
01/06
01/07
01/08
01/09

# Thank you

1. Cuneyt el all., Forecasting bitcoin price with graph chainlets, PAKDD, 2018

# Backup

# 1-ARW-betweeness centrality



**Algorithm 1** 1-ARW-betweenness centrality algorithm

**Require:** A graph $G = (V, E)$, and a restart probability $\alpha$.

**Ensure:** The 1-ARW-betweenness centrality score for all nodes in $V$

```

for $\forall u \in V$ do
    Define $C = \{u\}$ as the set of central nodes
    Define $Q = V \backslash \{u\}$ as the set of query nodes
    Let $\mathbf{P}_{TT}$ be the ARW transition matrix of the transient nodes $T = V \backslash C$ under the 1-ARW-betweenness centrality assumptions (Definition 3.7).
    Calculate the fundamental matrix $\mathbf{F} = (I - \mathbf{P}_{TT})^{-1}$
    Calculate the expected length of the ARW $\mathbf{L} = \begin{pmatrix} \mathbf{F} \\ \mathbf{0} \end{pmatrix} \mathbf{1}^{|T|}$
    Calculate the 1-ARW-betweenness centrality $L_{arw}(u) = \frac{1}{|V|} \sum_{i=1}^{|V|} \mathbf{L}(i)$

return The 1-ARW-betweenness centrality score $L_{arw}(u)$, $\forall u \in V$
```

1. Anouk van Schetsen el all., Impact of graph-based features on Bitcoin prices, Delft University of Technology, 2019

# Feature Engineering

A. Covariance matrix of features.

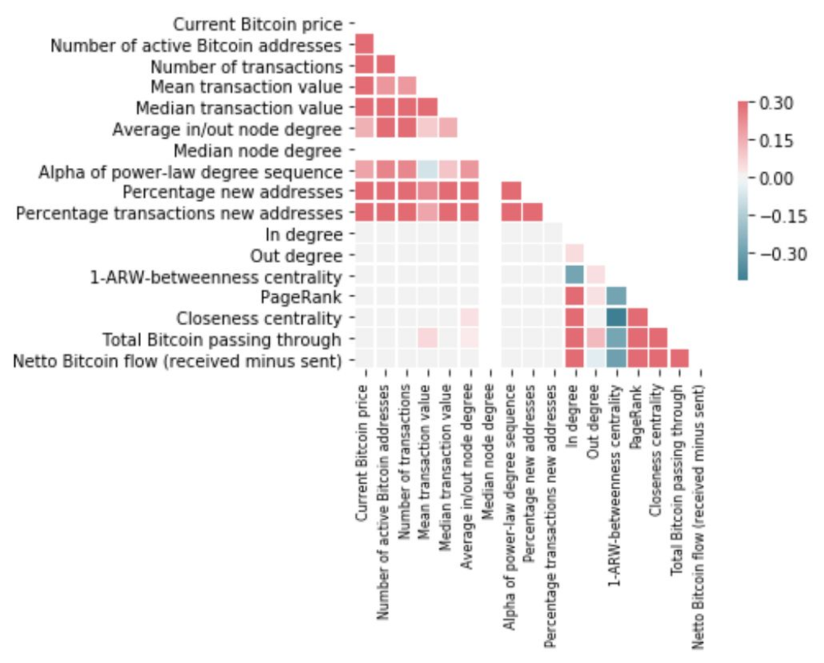


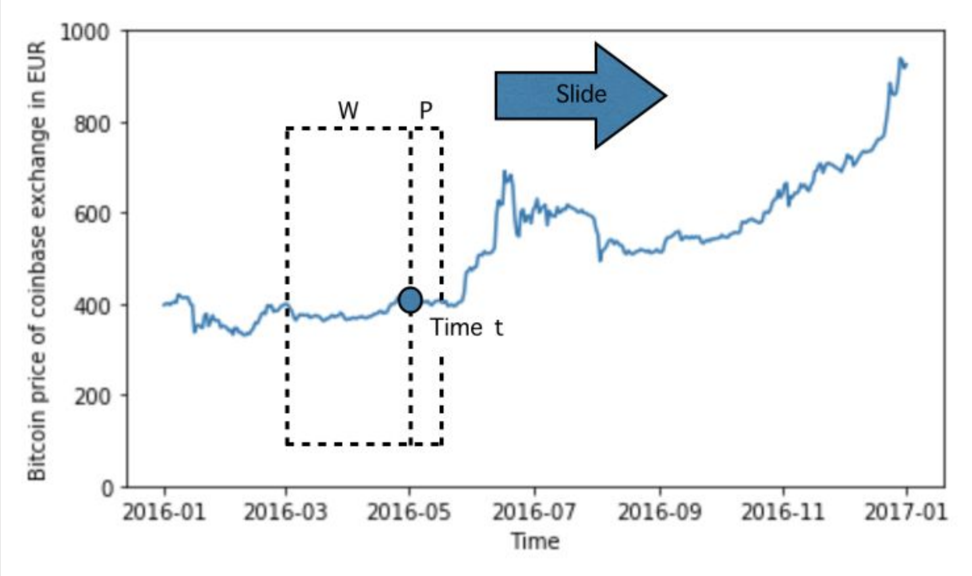
Fig. 6: Covariance matrix of features. [1]

---

1. Anouk van Schetsen el all., Impact of graph-based features on Bitcoin prices, Delft University of Technology, 2019

# Data preparation

## A. Window sampling



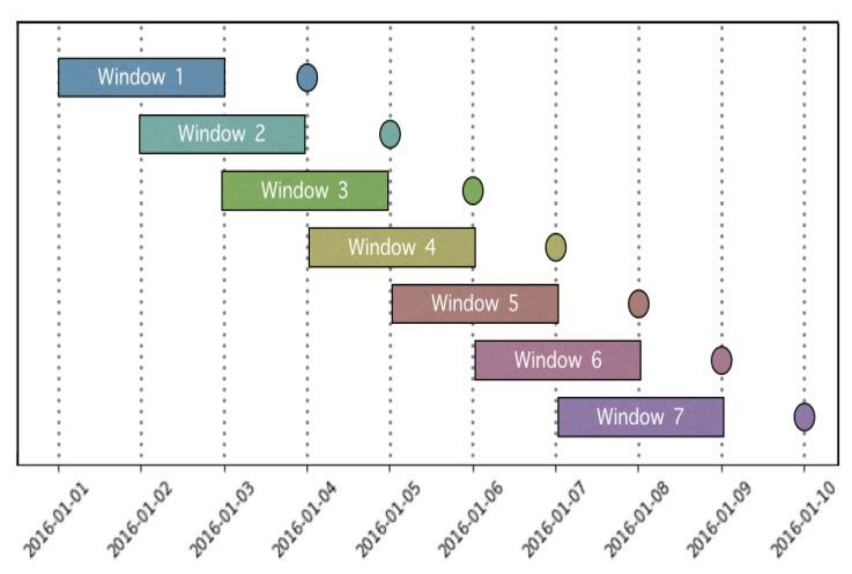


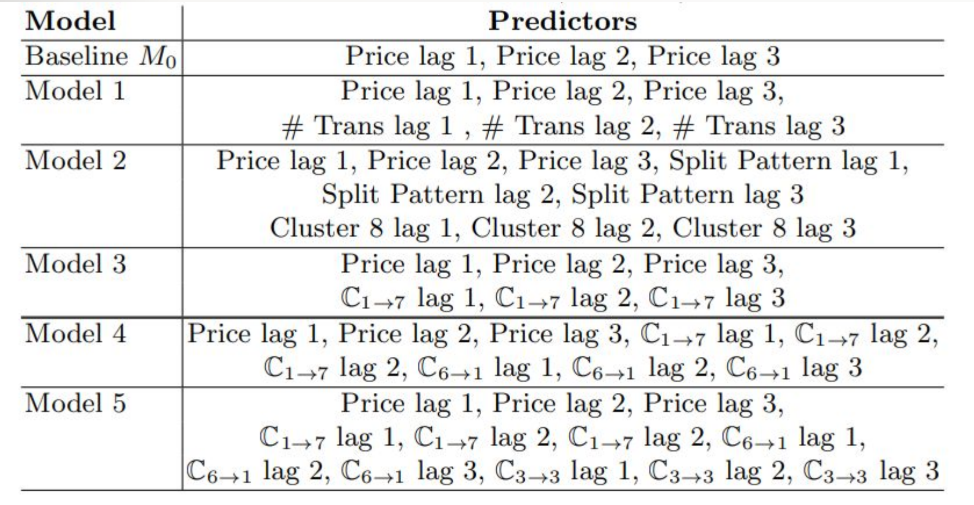
Fig. 11: Visualization of a single window, with w the window size and p the prediction interval. [1]

Fig. 12: Example of 7 windows (squares) and their prediction points (circles) [1]

---

1. Anouk van Schetsen el all., Impact of graph-based features on Bitcoin prices, Delft University of Technology, 2019

# ChainLet



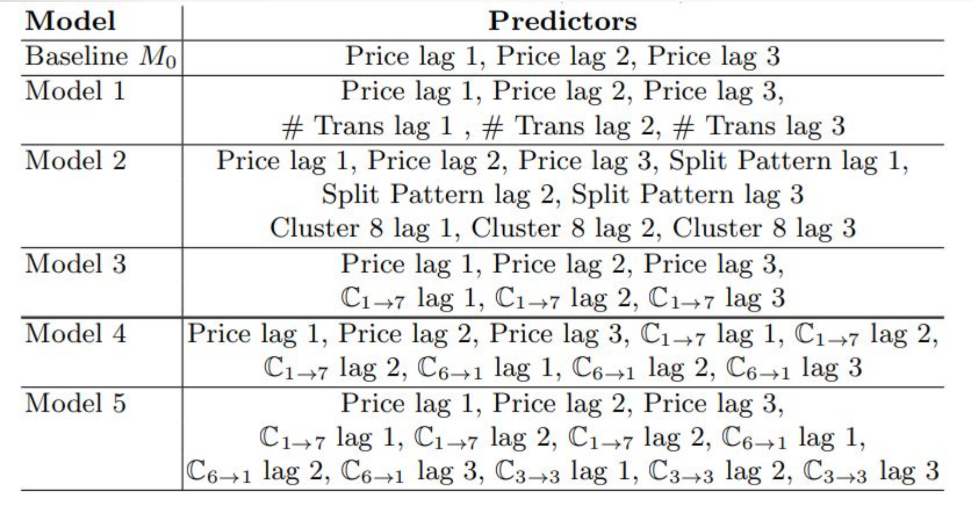
| Model | Predictors |
|---|---|
| Baseline $M_0$ | Price lag 1, Price lag 2, Price lag 3 |
| Model 1 | Price lag 1, Price lag 2, Price lag 3,<br># Trans lag 1 , # Trans lag 2, # Trans lag 3 |
| Model 2 | Price lag 1, Price lag 2, Price lag 3, Split Pattern lag 1,<br>Split Pattern lag 2, Split Pattern lag 3<br>Cluster 8 lag 1, Cluster 8 lag 2, Cluster 8 lag 3 |
| Model 3 | Price lag 1, Price lag 2, Price lag 3,<br>$\mathbb{C}_{1\to7}$ lag 1, $\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{1\to7}$ lag 3 |
| Model 4 | Price lag 1, Price lag 2, Price lag 3, $\mathbb{C}_{1\to7}$ lag 1, $\mathbb{C}_{1\to7}$ lag 2,<br>$\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{6\to1}$ lag 1, $\mathbb{C}_{6\to1}$ lag 2, $\mathbb{C}_{6\to1}$ lag 3 |
| Model 5 | Price lag 1, Price lag 2, Price lag 3,<br>$\mathbb{C}_{1\to7}$ lag 1, $\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{6\to1}$ lag 1,<br>$\mathbb{C}_{6\to1}$ lag 2, $\mathbb{C}_{6\to1}$ lag 3, $\mathbb{C}_{3\to3}$ lag 1, $\mathbb{C}_{3\to3}$ lag 2, $\mathbb{C}_{3\to3}$ lag 3 |

Table 1: Model description for Bitcoin price (response) and varying predictors. [1]

---

1. Cuneyt el all., Forecasting bitcoin price with graph chainlets, PAKDD, 2018

# ChainLet



| Model | Predictors |
|---|---|
| Baseline $M_0$ | Price lag 1, Price lag 2, Price lag 3 |
| Model 1 | Price lag 1, Price lag 2, Price lag 3,<br># Trans lag 1 , # Trans lag 2, # Trans lag 3 |
| Model 2 | Price lag 1, Price lag 2, Price lag 3, Split Pattern lag 1,<br>Split Pattern lag 2, Split Pattern lag 3<br>Cluster 8 lag 1, Cluster 8 lag 2, Cluster 8 lag 3 |
| Model 3 | Price lag 1, Price lag 2, Price lag 3,<br>$\mathbb{C}_{1\to7}$ lag 1, $\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{1\to7}$ lag 3 |
| Model 4 | Price lag 1, Price lag 2, Price lag 3, $\mathbb{C}_{1\to7}$ lag 1, $\mathbb{C}_{1\to7}$ lag 2,<br>$\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{6\to1}$ lag 1, $\mathbb{C}_{6\to1}$ lag 2, $\mathbb{C}_{6\to1}$ lag 3 |
| Model 5 | Price lag 1, Price lag 2, Price lag 3,<br>$\mathbb{C}_{1\to7}$ lag 1, $\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{1\to7}$ lag 2, $\mathbb{C}_{6\to1}$ lag 1,<br>$\mathbb{C}_{6\to1}$ lag 2, $\mathbb{C}_{6\to1}$ lag 3, $\mathbb{C}_{3\to3}$ lag 1, $\mathbb{C}_{3\to3}$ lag 2, $\mathbb{C}_{3\to3}$ lag 3 |

Table. 2: Model description for Bitcoin price (response) and varying predictors. [1]

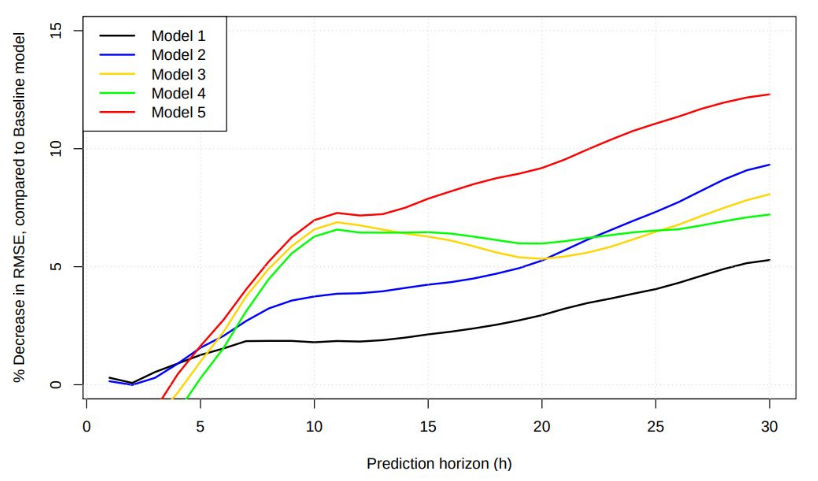
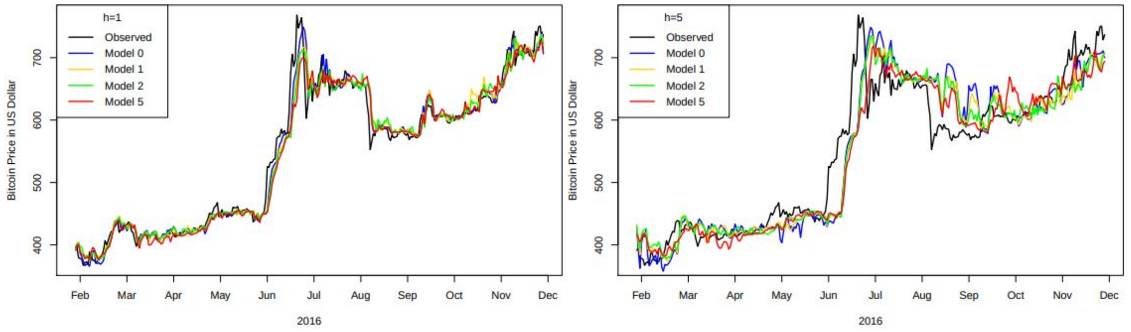


Fig. 7: % Change (decrease) in RMSE compared to the baseline model. [1]

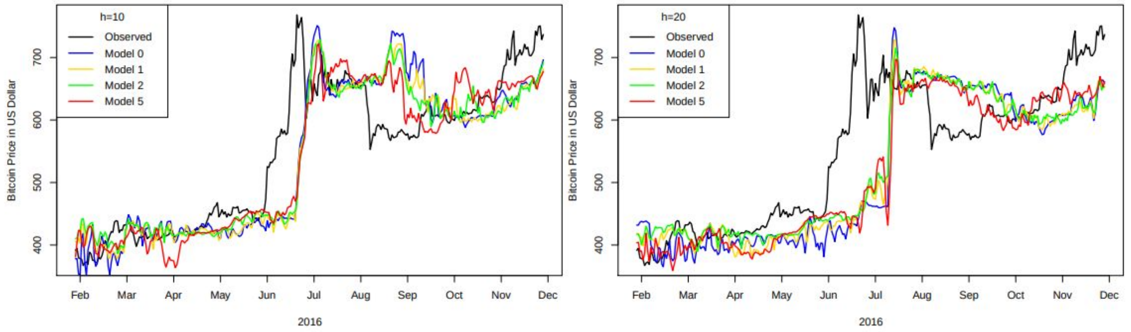
---

1. Cuneyt el all., Forecasting bitcoin price with graph chainlets, PAKDD, 2018

# ChainLet



(a) h=1 day. (b) h=5 days.



(c) h=10 days. (d) h=20 days.

Fig. 8: Price prediction for 2016 with 1, 5, 10 for 20 day horizons. [1]

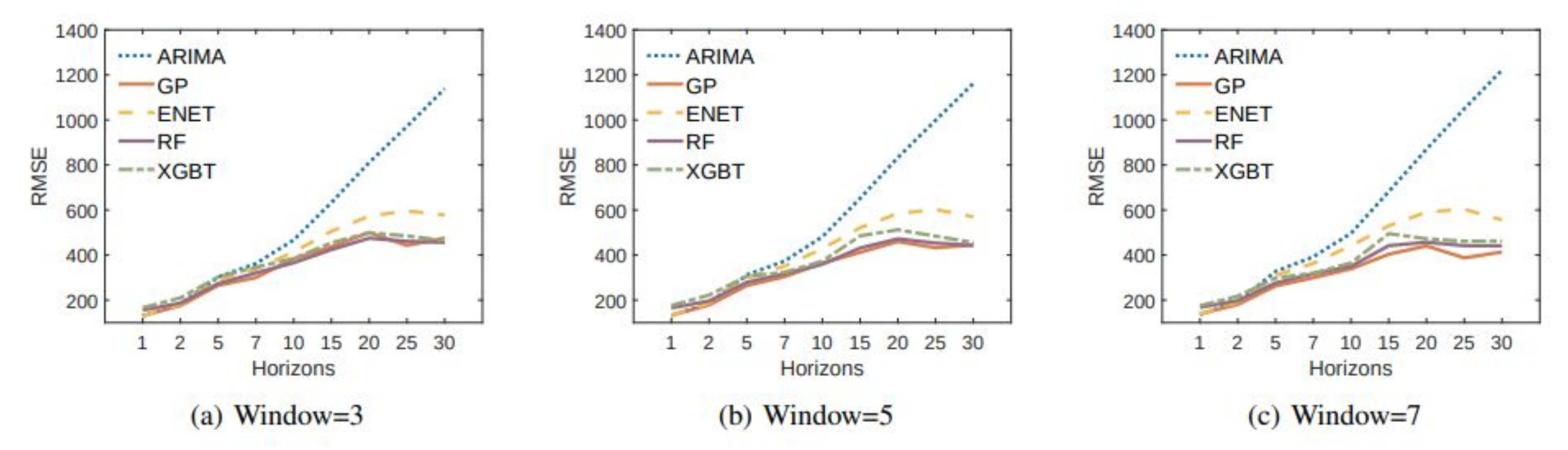
---

1. Cuneyt el all., Forecasting bitcoin price with graph chainlets, PAKDD, 2018

# ChainNet

## Result

1. RMSE



(a) Window=3 (b) Window=5 (c) Window=7

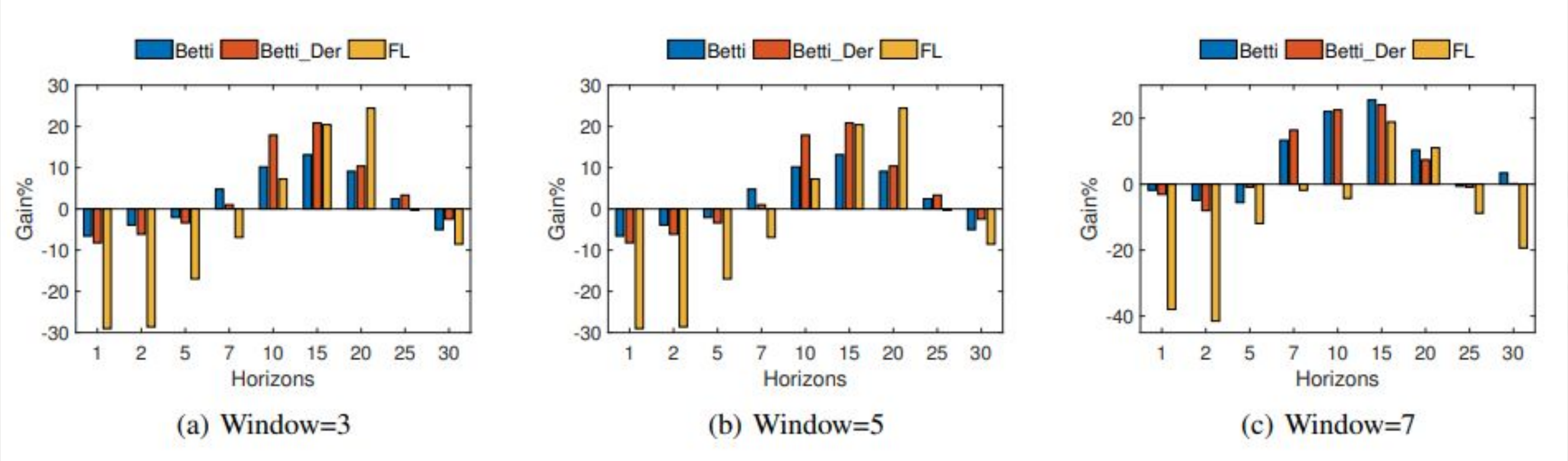
Fig. 9: RMSE of sliding window based predictions of 2017 Bitcoin prices in different window and horizon values. [1]

[1] Nazmiye el all., ChainNet: Learning on Blockchain Graphs with Topological Features, ICDM, 2019

# ChainNet

## Conclusion

1. RMSE



(a) Window=3 (b) Window=5 (c) Window=7

Fig. 10: Random Forest Performance.[1]

[1] Nazmiye el all., ChainNet: Learning on Blockchain Graphs with Topological Features, ICDM, 2019

# ANMELDUNG EINER MASTERARBEIT

Name: **Yuhao Wang**

Matrikelnummer: **11863170**

Aufgabensteller: **Dr. Yunpu Ma**

Anmeldedatum: **2021-02-15**

Titel der Arbeit: **Topologische Merkmale im dynamischen Transaktionsdiagramm**

Englischer Titel der Arbeit: **Topological features on the dynamic transaction graph**