

The 23rd Annual International Conference on Information Security and Cryptology

ICISC 2020

December 2 (Wed) ~ December 4 (Fri), 2020 | Virtual Conference

Hosted by

Korea Institute of Information Security and Cryptology (KIISC)
National Security Research Institute (NSR)



A RDBMS-based Bitcoin Analysis Method

Hyunsu Mun, Soohyun Kim, Youngseok Lee

Data Network Lab.

*Chungnam National University,
Daejeon, Korea*











munhyunsu@cnu.ac.kr





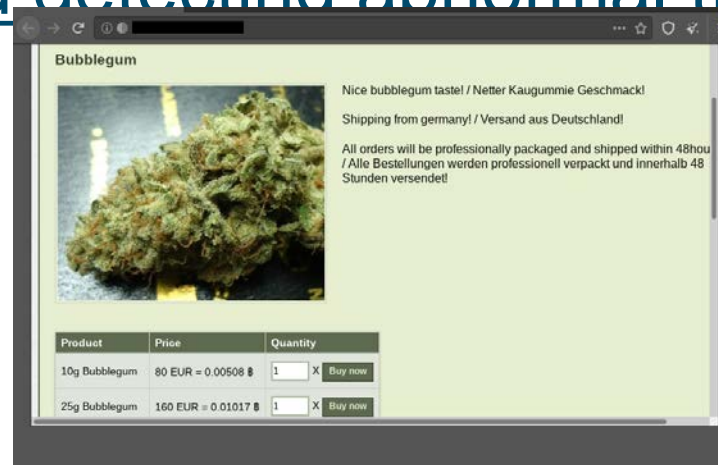
Over 1000 Cryptocurrencies

- Growing to more than \$405 billion
- Anonymity using block-chain
- **Need to monitor and analysis**
 - transaction status
 - user activity

Rank	Name	Symbol	Market Cap	Price
1	 Bitcoin	BTC	\$343,930,168,610	\$18,540.49
2	 Ethereum	ETH	\$61,127,039,116	\$538.46
3	 XRP	XRP	\$18,579,757,617	\$0.409894
4	 Tether	USDT	\$18,359,419,201	\$0.999125
5	 Chainlink	LINK	\$5,948,022,848	\$15.12
6	 Litecoin	LTC	\$5,608,520,569	\$85.08
7	 Bitcoin Cash	BCH	\$5,500,786,597	\$296.08
8	 Polkadot	DOT	\$5,023,251,257	\$5.71
9	 Binance Coin	BNB	\$4,353,982,997	\$30.15
10	 Cardano	ADA	\$3,951,075,348	\$0.126993

Monitor and Analysis Cryptocurrencies

- Cryptocurrencies with Anonymity
 - Used in the deep web selling illegal products such as drugs and weapons
- **Increasing demand** from government agencies
 - For tracking and detecting abnormal transactions early





One Line Bitcoin Analysis Method

- Bitcoin: Largest Market Size
- Example queries:
 - Increase or decrease the balance of the address for a specific period
 - Bitcoin address heuristic
 - Export Bitcoin address-transaction graph



One-Line Bitcoin Analysis Example:

- **Changes in balance** in 1st half of 2020 for address with the most bitcoins

```

sqlite> SELECT Income.value-Outcome.value AS BalanceChange
...> FROM
...> (SELECT SUM(btc) AS value
...> FROM TxOut
...> WHERE TxOut.addr = (SELECT DBINDEX.AddrID.id
...> FROM DBINDEX.AddrID
...> WHERE DBINDEX.AddrID.addr = '35hK24tcLEWcgNA4JxpvbkNkoAcDGqQPSP') AND
...> TxOut.tx IN (SELECT BlkTx.tx
...> FROM BlkTx
...> INNER JOIN BlkTime ON BlkTime.blk = BlkTx.blk
...> WHERE (SELECT STRFTIME('%s', '2020-01-01T00:00:00+00:00')) <= BlkTime.unixtime AND
...> BlkTime.unixtime <= (SELECT STRFTIME('%s', '2020-06-30T23:59:59+09:00')))) AS Income,
...> (SELECT SUM(btc) AS value
...> FROM TxIn
...> INNER JOIN TxOut ON TxIn.ptx = TxOut.tx AND TxIn.pn = TxOut.n
...> WHERE TxOut.addr = (SELECT DBINDEX.AddrID.id
...> FROM DBINDEX.AddrID
...> WHERE DBINDEX.AddrID.addr = '35hK24tcLEWcgNA4JxpvbkNkoAcDGqQPSP') AND
...> TxIn.tx IN (SELECT BlkTx.tx
...> FROM BlkTx
...> INNER JOIN BlkTime ON BlkTime.blk = BlkTx.blk
...> WHERE (SELECT STRFTIME('%s', '2020-01-01T00:00:00+09:00')) <= BlkTime.unixtime AND
...> BlkTime.unixtime <= (SELECT STRFTIME('%s', '2020-06-30T23:59:59+09:00')))) AS Outcome;
BalanceChange
0.0165592599999997

```



A RDBMS-based Bitcoin Analysis Method

- In this paper, we propose an extensible and scalable Bitcoin analysis based on RDBMS
- We solve these challenges using a RDBMS
 - Scalable Bitcoin data ingestion and storage
 - Easy interface of analytics
 - Compatibility of software integration



Previous Studies for Analysis Bitcoin

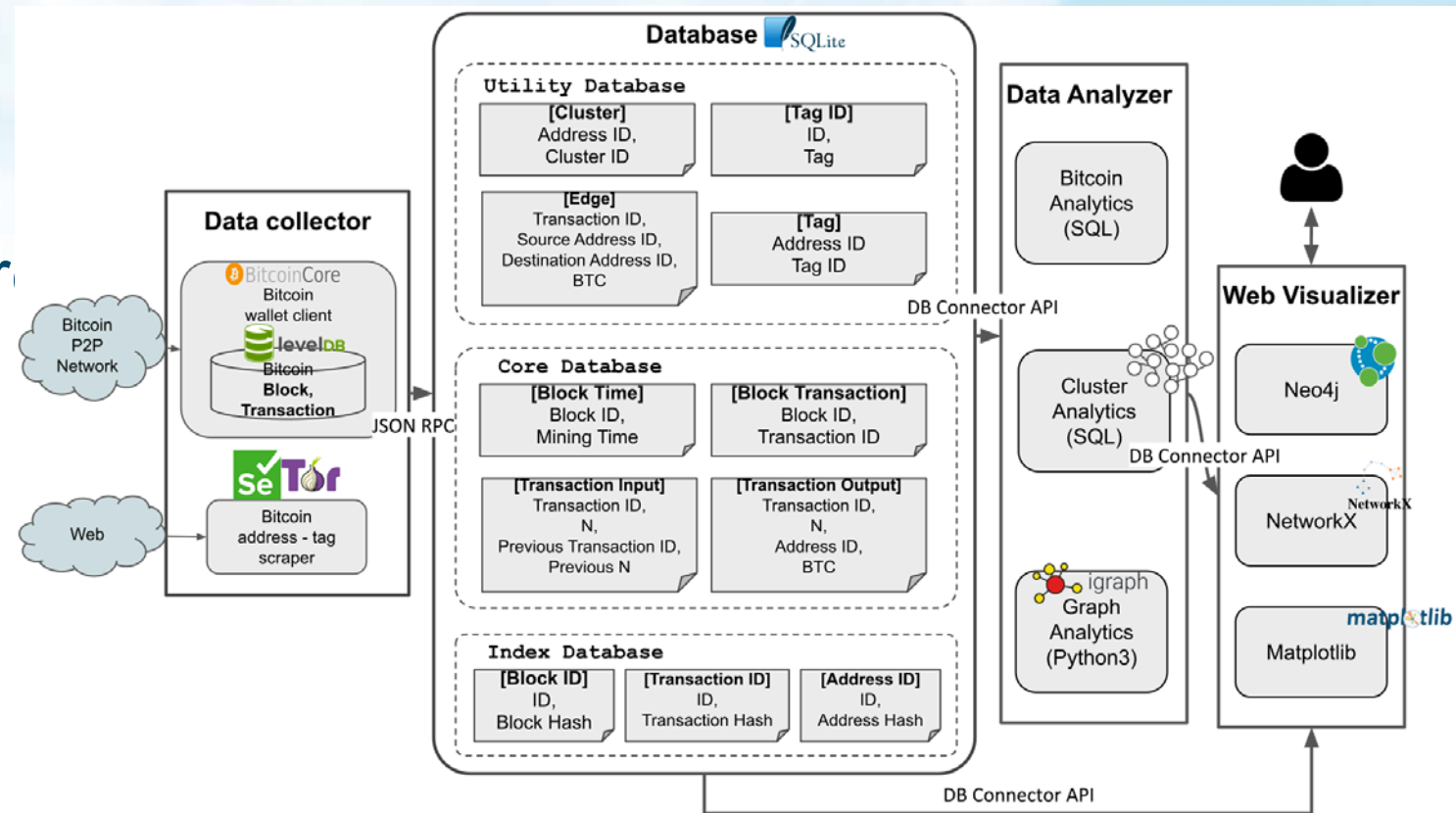
- A Users can **quickly** analyze bitcoin with **only SQL** without any dependencies

Table 1: Comparison of Bitcoin analysis tools: our method vs. previous studies.

	Ours	BlockSci	GraphSense	BTCSpark	Bitcoin-Abe	BlockAPI
Cryptocurrency	Bitcoin Ethereum	Bitcoin Bitcoin-like	Bitcoin	Bitcoin	Bitcoin	Bitcoin Ethereum
Language	SQL	C++ Python 3	Scala	Scala	C++	JAVA Scala
Dependency	-	LevelDB	BlockSci Casandra	Spark	LevelDB MySQL	LevelDB
Bitcoin Analytics	○	○	○	○	○	○
Cluster Analytics	○	○	×	×	×	×
Graph Analytics	○	×	○	×	×	×
Response Time	Low	Low	High	High	High	High

A RDBMS-based Bitcoin Analysis Architecture

- Key point:
 - **Three-layer** DB
 - Database update through JSON-RPC
 - Extensibility using DB connector



Database Schema: DB size reduction

- Three-Layer Database schema:
 - Index / Core / Utility
- The index database is useful for reducing the long Bitcoin hash value to the integer variable
 - **Decrease the size of the core table bv 65%**
(906GB → 314GB)

```
sqlite> SELECT * FROM BlkID LIMIT 3;
id|blkhash
0|000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f
1|00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048
2|000000006a625f06636b8bb6ac7b960a8d03705d1ace08b1a19da3fdcc99ddbd
sqlite> SELECT * FROM TxID LIMIT 3;
id|txid
1|4a5e1e4baab89f3a32518a88c31bc87f618f76673e2cc77ab2127b7afdeda33b
2|0e3e2357e806b6cdb1f70b54c3a3a17b6714ee1f0e68bebb44a74b1efd512098
3|9b0fc92260312ce44e74ef369f5c66bbb85848f2eddd5a7a1cde251e54ccfdd5
sqlite> SELECT * FROM AddrID LIMIT 3;
id|addr
1|1A1zP1eP5QGefi2DMPTfTL5SLmv7DivfNa
2|12c6DSiU4Rq3P4ZxziKxZrL5LmBrzjrJX
3|1HLoD9E4SDFFPDiYfNYnkBLQ85Y51J3Zb1
```

Database Schema: Maintain Bitcoin Information

- Three-Layer Database schema:
 - Index / Core / Utility
- The core database **maintain Bitcoin information**
 - Block Information: BlkTx, BlkTime
 - Transaction Information: TxIn, TxOut

```
sqlite> SELECT * FROM BlkTx LIMIT 3;
blk|tx
0|1
1|2
2|3
sqlite> SELECT * FROM BlkTime LIMIT 3;
blk|unixtime
0|1231006505
1|1231469665
2|1231469744
```

```
sqlite> SELECT * FROM TxIn LIMIT 3;
tx|n|ptx|pn
1|0|0|0
2|0|0|0
3|0|0|0
sqlite> SELECT * FROM TxOut LIMIT 3;
tx|n|addr|btc
1|0|1|50.0
2|0|2|50.0
3|0|3|50.0
```

Database Schema: Response time improvement

- Three-Layer Database schema:
 - Index / Core / Utility
- Improved response speed by **eliminating repetitive JOIN** operation
 - Save query results that can be obtained only with Index Core database
 - Especially useful for extracting address-transaction graph

```
sqlite> SELECT * FROM Edge LIMIT 3;  
tx|src|dst|btc  
172|10|172|10.0  
172|10|10|40.0  
184|10|184|10.0
```




Bitcoin Analytics

- Target Information
 - Bitcoin address, transaction, cluster information
- Condition
 - Date, Time, Clusters
- Example
 - Changes in balance in 1st half of 2020 for address with the most bitcoins

Cluster Analytics

- Target Information
 - Bitcoin **heuristic**, Tag information,
Addresses belonging to a single organization
- Condition
 - Date, Time, Clusters
- Example
 - List of Bitcoin addresses held by Huobi.com
(using Multi input heuristic)



Graph Analytics

- Target Information
 - Indegree / Outdegree of address node,
Export address-transaction graph (.csv)
- Condition
 - Date, Time, Clusters
- Example
 - The most important addresses in the exchange cluster
(using the PageRank algorithm)

RDBMS-based Bitcoin Analysis 7 Examples

- Which address has the **largest amount of Bitcoin** ?
- **Clustering** Bitcoin addresses with heuristics
- **Identify clusters** with Address-Tag information
- What is the **amount and count of transactions** to CryptoLocker addresses ?
- List the **hot wallet** addresses of a Korean exchange A
- **Graph analysis** algorithm using graph tools
- **Community detection** algorithm on Korean exchange B cluster



Bitcoin Data for Analysis

- Bitcoin database construction time: about 4 hours (Workstation PC)

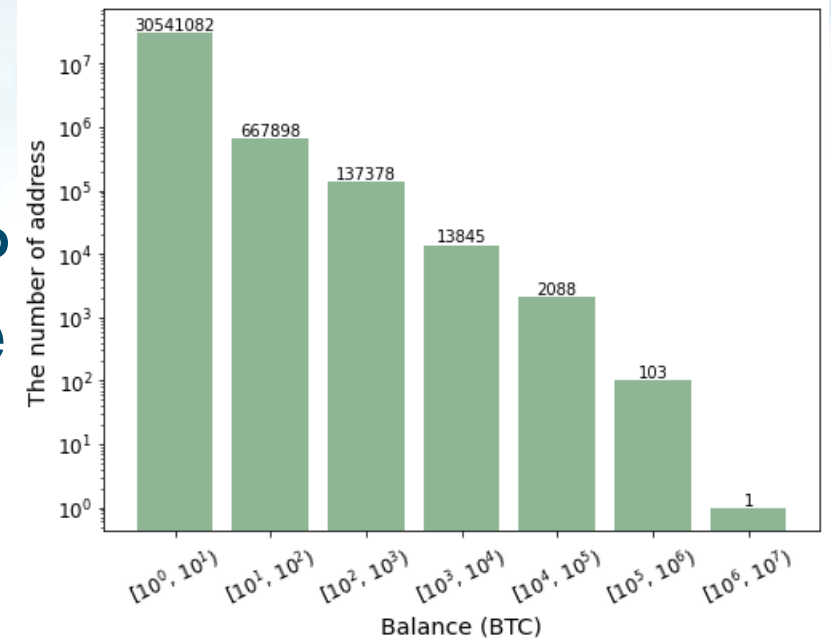
Table 2: Bitcoin data information for database construction.

Block height	0 ~ 644,806
Date	2009-01-03 06:15:05 (UTC) ~ 2020-08-22 06:24:55 (UTC)
Transactions	560,882,950
Addresses	704,688,729
Database volume	357 GB



Which address has the **largest amount of Bitcoin** ?

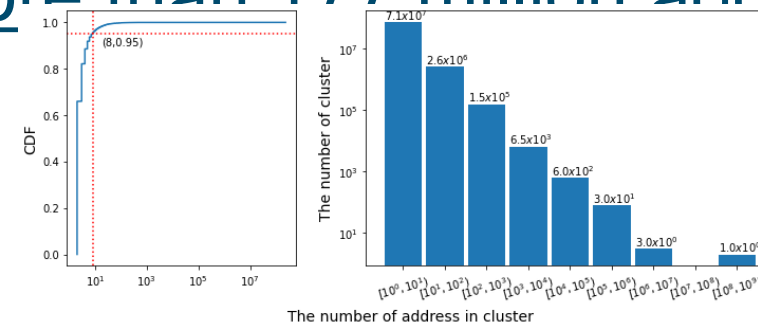
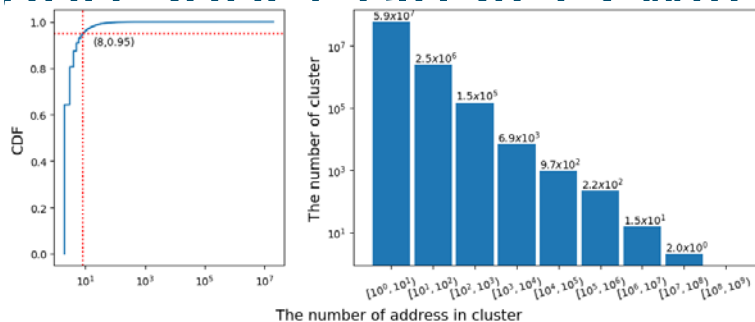
- 18467457.8422912 BTC has been generated as of 2020-08-22 06:24:55
- 31,362,395 Addresses hold Bitcoin (BTC)
- 35hK24tcLEWcgNA4JxpvbkNkoAcDGqQP sP has 355 UTXOs and 1.14% of the generated Bitcoin
- Top 10 list in the paper





Clustering Bitcoin addresses with heuristics

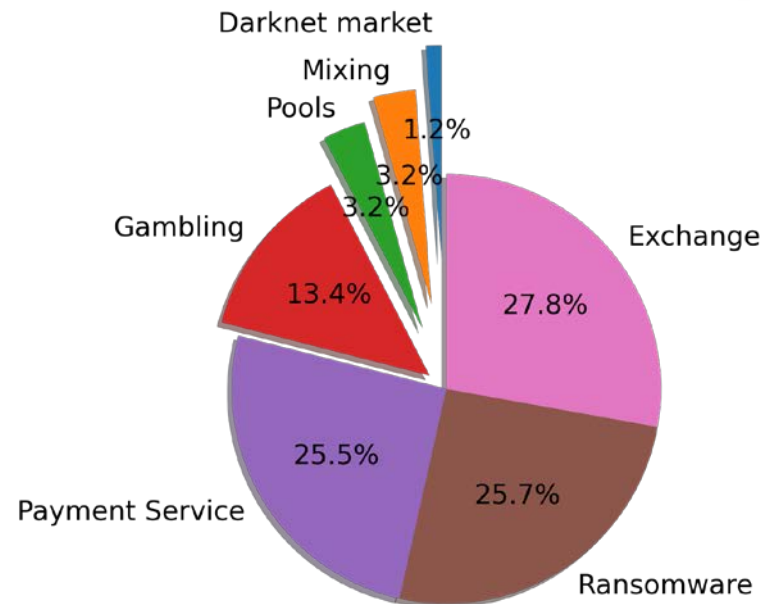
- We processed 691,806,723 addresses using Multi input
 - grouped 418,493,080 addresses into 61,918,407 clusters
 - found super-clusters with more than 10 million addresses
- One time change heuristic with multi input heuristic
 - grouped 45,980,870 addresses more then multi input only
 - found super-clusters with more than 177 million addresses





Identify clusters with Address-Tag information

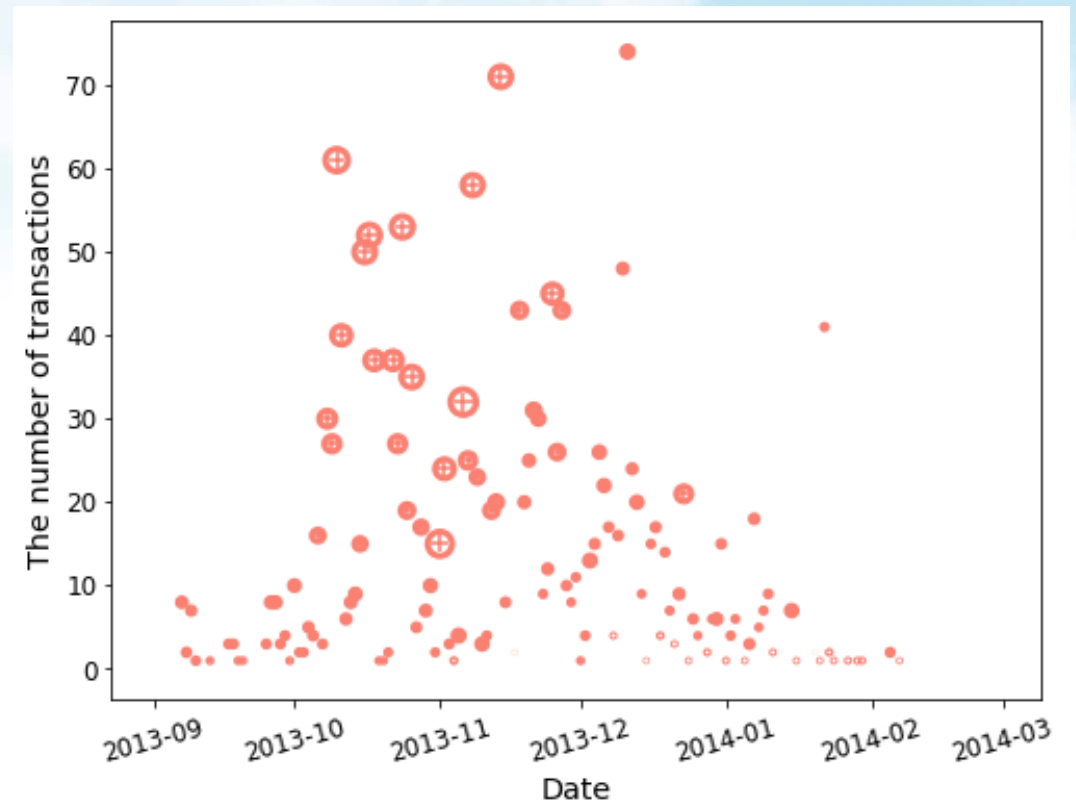
- We collected Address-Tag information in the web-site
- 86,278,071 address were identified (20.62%)
- Exchange, Ransomware, Payment Service are largest clusters





What is the **amount and count of transactions** to CryptoLocker addresses ?

- CryptoLocker Bitcoin transactions and BTC amount
 - 2013. 09. 05. ~ 2014. 05
- 2,789.15 BTC was transferred into CryptoLocker
- In October 2013, most transactions



List the **hot wallet** addresses of a Korean exchange A

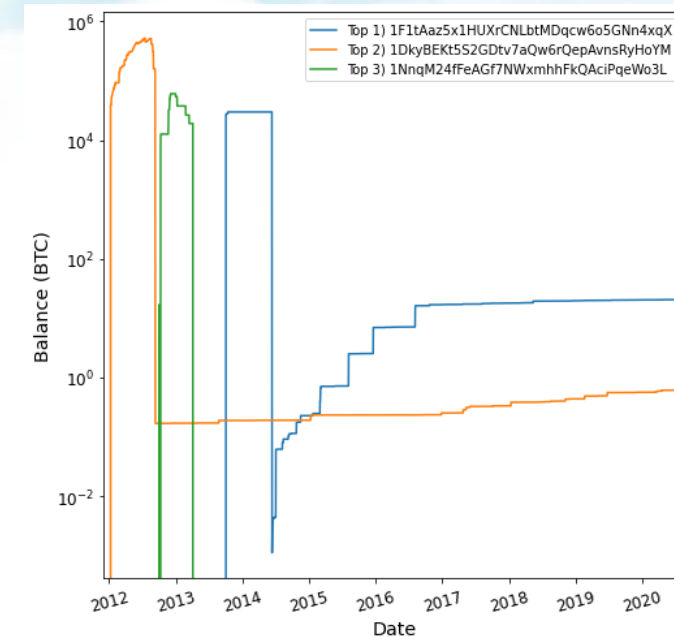
- Three types of wallets in Exchange
 - User wallet, Hot wallet, Cold wallet
- The hot wallet is the center of operation of the exchange
 - high indegree / outdegree
- We can find 12 hot wallet estimated addresses which has a degree of more than 1000 degree.

Table 4: Top 12 hot wallet list of Korean exchange A

Rank	Bitcoin address	Degree	Value
1	1En5ErLPzF9RMeP8z8hjna3.....	2,870	21,198
2	1JeyZBDJtZ5d1rfSkGqyww.....	2,520	122
3	19iGtbDzXSASmcyJFbdgCiFi.....	2,082	102
4	1KHFeyp2Sb4xXg1rjnNRi4c8.....	2,076	28,549
5	1Gxd9c2VuLcQjee1tubhHSJS.....	2,072	108
6	19Ls2qFMEztRVgSYyFFtFrE.....	1,516	14
7	1CsTzASjqs8f63pzcw5f9LJo.....	1,477	128
8	181acE6XdV4JToqMFRNmmKDq.....	1,402	70
9	1GoxkdmizZFKwndzGihHcW82.....	1,352	33
10	1QJ13PRLkWBF4s1XGKUMr9Ab.....	1,162	12,051
11	18rWxfA3Qv6uFKwKexGtskPx.....	1,130	4,842
12	1AvGPjBB3PcdhLYwy7twKFrP.....	1,116	50

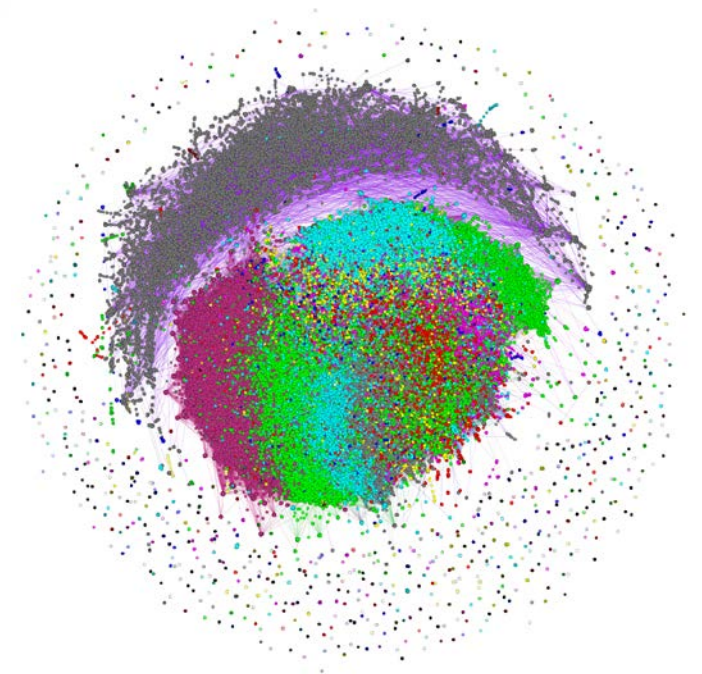
Graph analysis algorithm using graph tools

- Construct graph of Silkroad
 - Export address-transaction graph edges using query
- Investigate the main address with the PageRank algorithm
 - Using iGraph
- We can observe that the recently occurred transactions
 - Silkroad was arrested by the FBI in 2013



Community detection algorithm on Korean exchange B

- Used community detection algorithms for find cluster
- Performed Leiden community detection algorithm on 275,952 addresses
- Largest community was a 51,714 addresses
 - 7 communities over 10,000 addresses





Conclusion

- In this paper, we proposed **Bitcoin analysis based on RDBMS**
 - Three layer databases with seven bitcoin tables
- RDBMS-based analysis method supports
 - Bitcoin analytics, Cluster analytics, and Graph analytics
- We plan to build an RDBMS-based Bitcoin analysis system
 - Bitcoin address cluster analysis, machine learning, and graph analysis function



Thank you

Hyunsu Mun
munhyunsu@cnu.ac.kr



Arial, Bold, 36

- Arial, 28
- Arial, 28
- Arial, 28

Description... Arial, 24