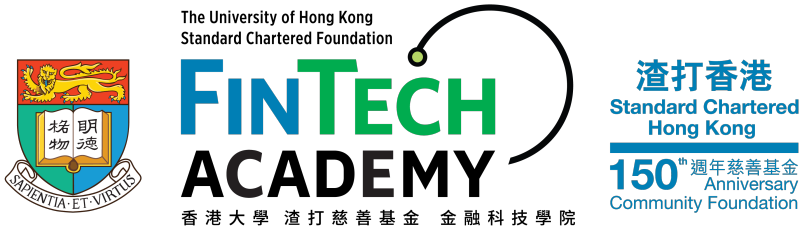
Tracing crypto-currency transactions for suspected "illegal" activities

SM Yiu
Professor, Department of Computer Science
The University of HK
Leader, "Cyber security, FinTech, blockchain research group"
Deputy Executive Director, HKU-SCF FinTech Academy
(香港大学 - 渣打香港150周年慈善基金金融科技学院)



Agenda

♡ Privacy in cryptocurrency transactions

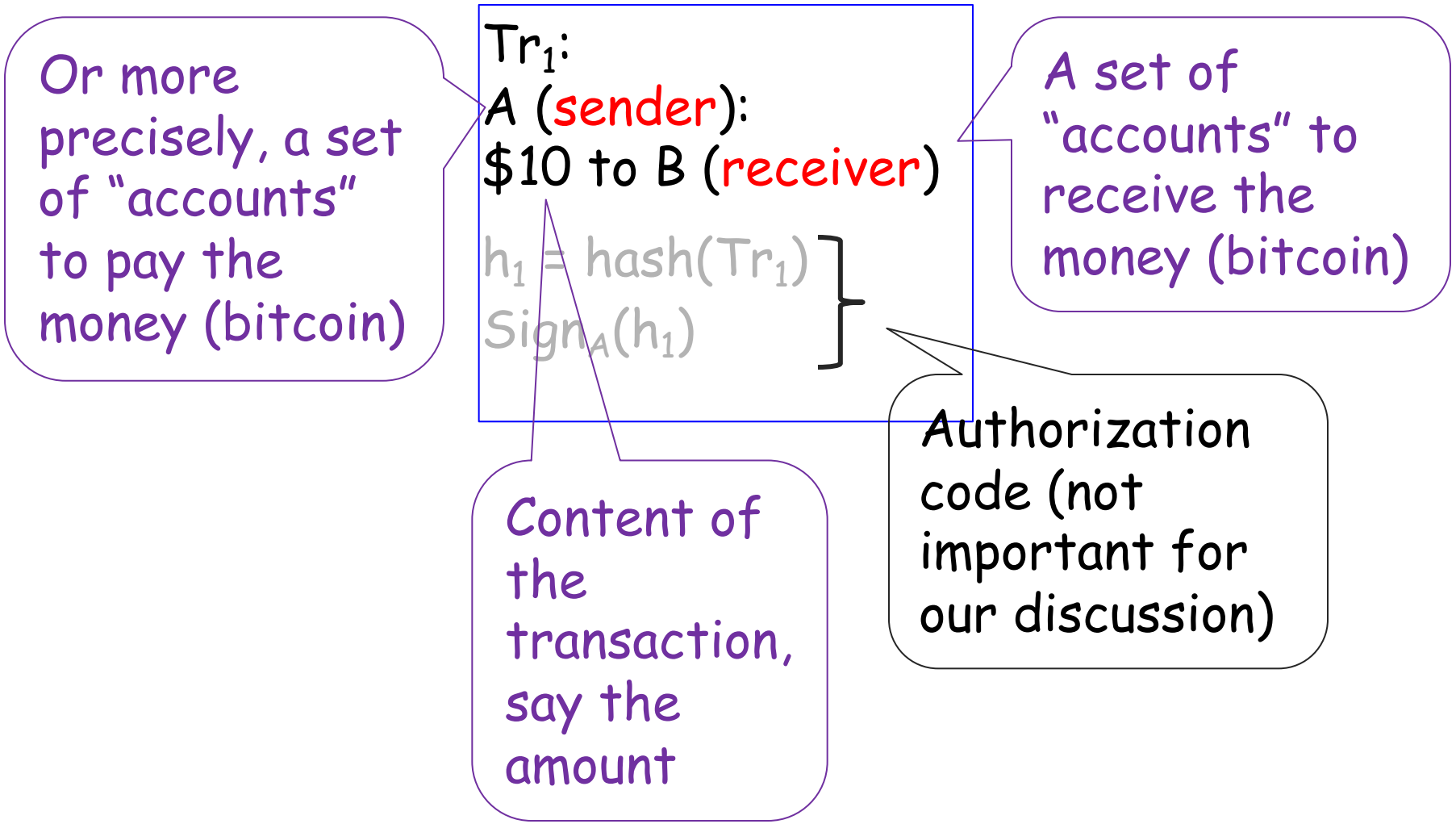
- How privacy is protected in blockchain?
- 100% guarantee for privacy?

♡ The evil side (crime cases) of cryptocurrency

♡ A brief introduction of advanced techniques for tracing cryptocurrency transactions

In the original design of blockchain (use bitcoin as an example): How to protect privacy?

High-level speaking, a transaction in bitcoin looks like:

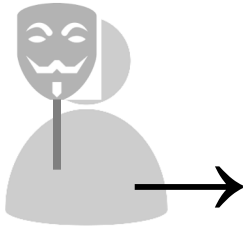


Tr 028 (id)
Input
adr 1
adr 2
Output
adr 3: 10BTC
adr 4: 5 BTC

Privacy issue:

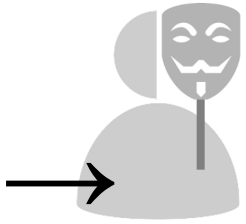
- If you know the owner of adr 1, 2, 3, 4, then you know who is the sender and who is the recipient!
- The amount involved in the transaction can be clearly known by everybody.

3 Types of Privacy:



Sender anonymity

Confidential transaction

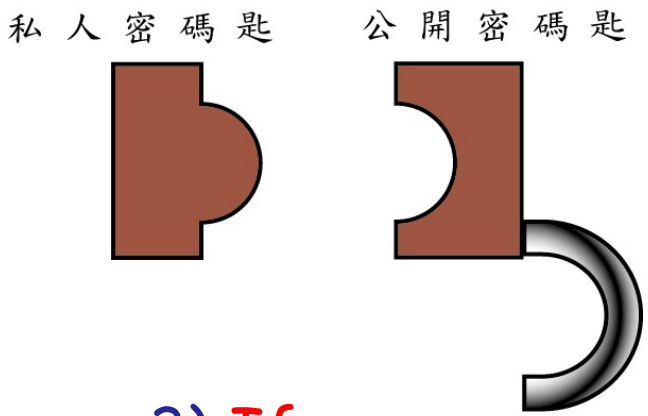


Recipient anonymity

In the original design of blockchain (i.e., bitcoin), they try to guarantee "sender anonymity" and "recipient anonymity"! (How?)

Concept of public key, private key pairs

- Public key & private key always go in pairs (like husband and wife)
- One can generate as many pairs of public and private keys as you like



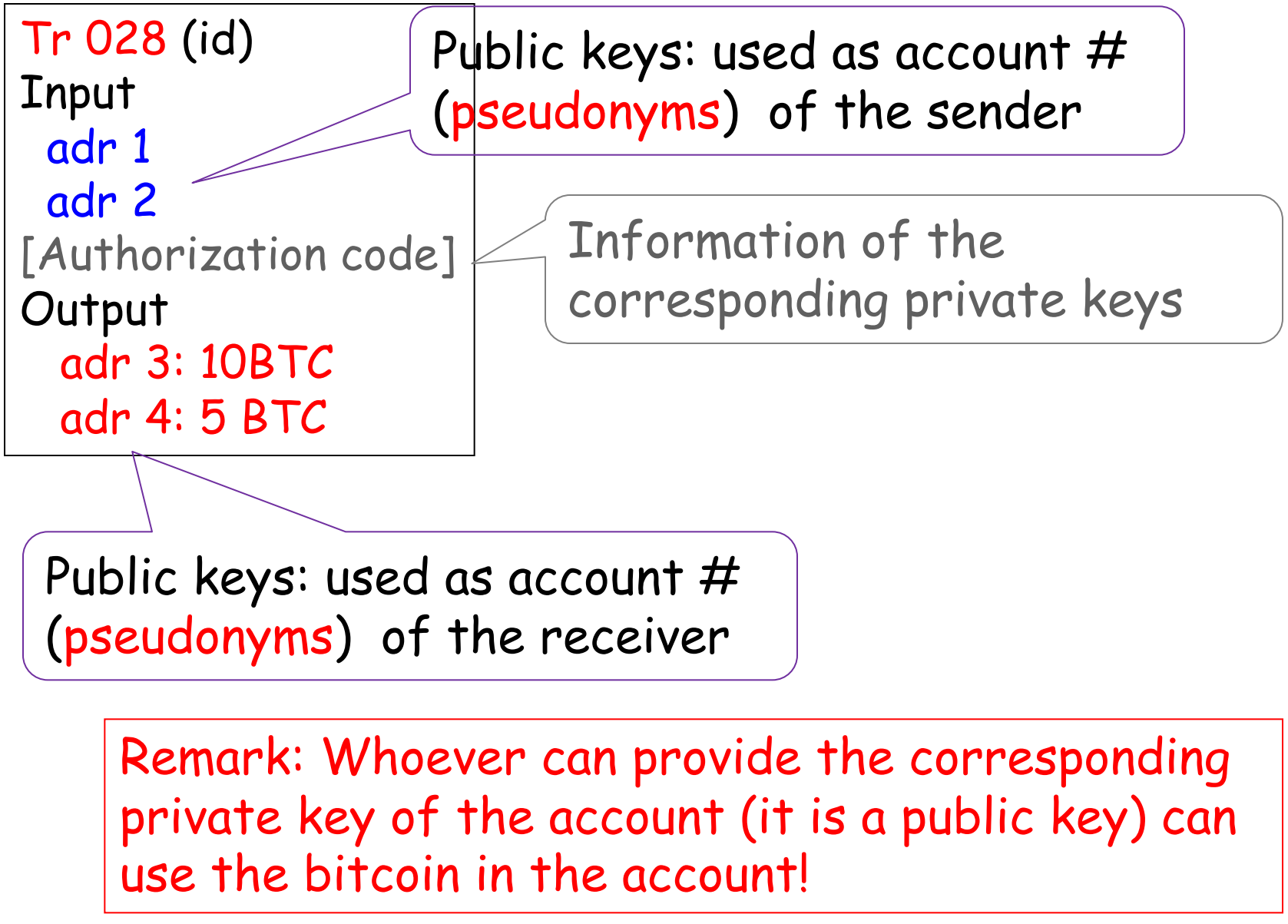
Remarks:

1) Public key and private key look like "random" numbers if you are not the owner.

2) For each public key, only the corresponding private key can "match" it.

3) If one sees a public key, it is extremely difficult to deduce the private key!

Guarantee by math (cryptograhy)


Tr 028 (id)
Input
adr 1
adr 2
[Authorization code]
Output
adr 3: 10BTC
adr 4: 5 BTC
Public keys: used as account # (pseudonyms) of the sender
Information of the corresponding private keys
Public keys: used as account # (pseudonyms) of the receiver
Remark: Whoever can provide the corresponding private key of the account (it is a public key) can use the bitcoin in the account!

A short summary

- In bitcoin, sender and recipient anonymity are guaranteed by public-private key pair (using cryptography)

- Transaction confidentiality is NOT protected, of course, in some other cryptocurrency scheme, the platform may provide additional protection to this. [Out of scope of today's talk]

Anonymity vs Unlinkability

State of being anonymous and unidentified. Seems OK

Inability to relate two observed transactions or two observed entities Not OK

A linkable example:

Tr 010 (id)
Input
adr 1
adr 2
[Authorization code]
Output
adr 3: 10BTC
adr 4: 5 BTC

- - - - - -

Tr 042 (id)
Input
adr 8
adr 9
[Authorization code]
Output
adr 1: 8BTC
adr 2: 3 BTC

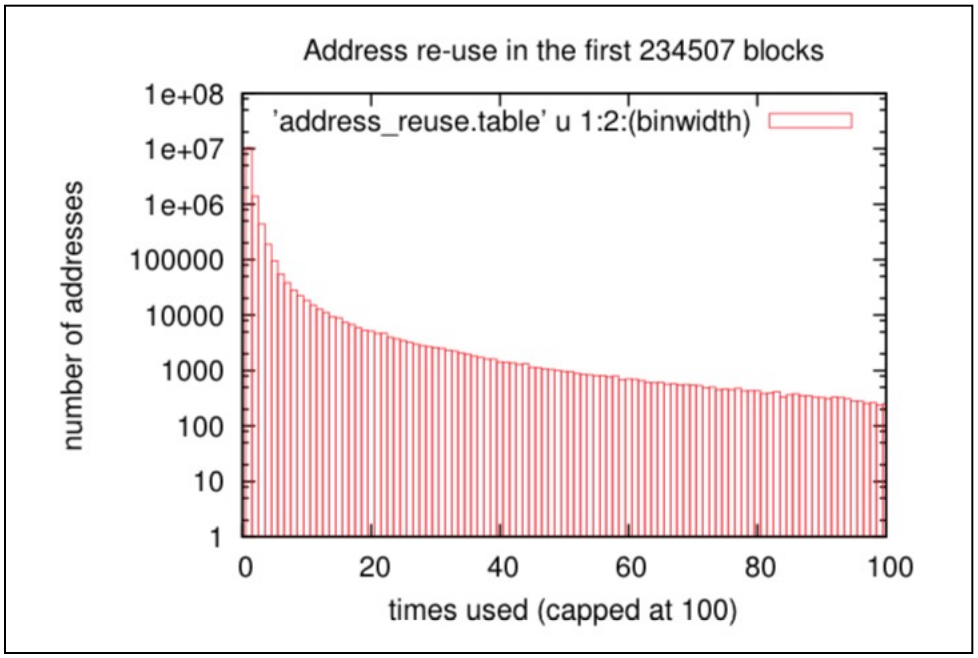
Q: what you can deduce?

Sender of Tr010 and recipient of Tr042 probably is the same person!

Original design of bitcoin: (i) recommended NOT to reuse address!

TABLE I
ADDRESS REUSE STATISTICS

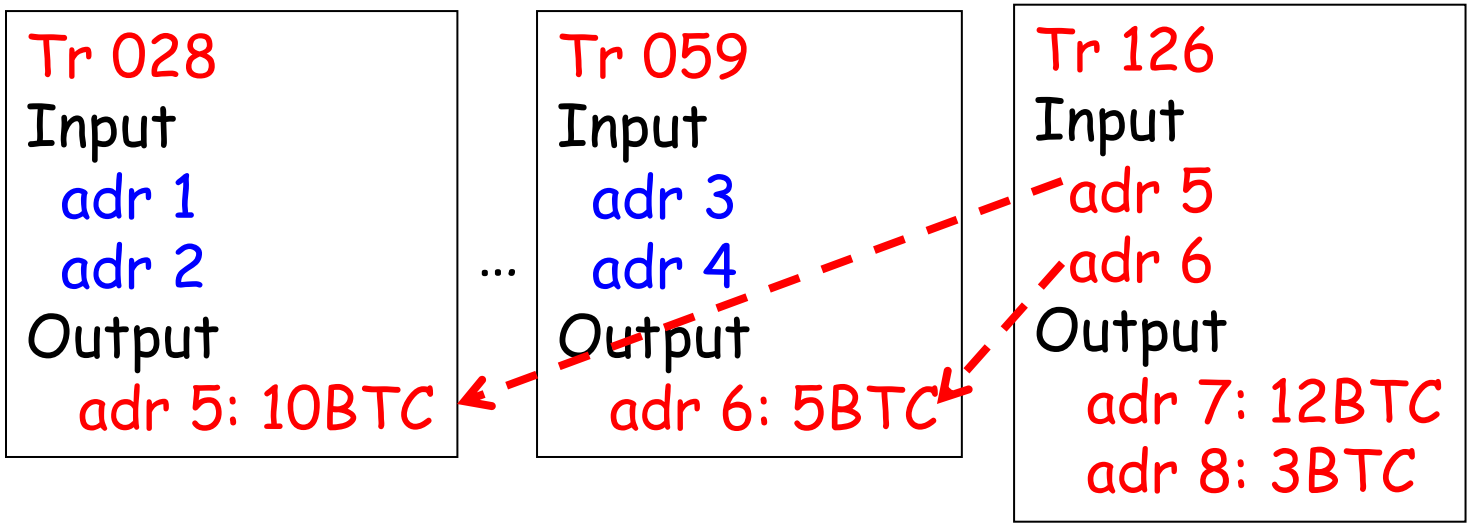
Mean	3.18
Min	1
25th perc.	1
50th perc.	1
75th perc.	1
Max	1,238,931
Number of addresses	12,963,199
Number of uses	41,244,997
Addresses used once	10,476,899
Addressed used twice	1,397,373
Used over 100 times	25,004



Jaume Barelo, "User privacy in the public bitcoin blockchain", 2007

Remark: Just looking at the addresses being reused, we may be able to link up different transactions!

E.g. Another linkable example:

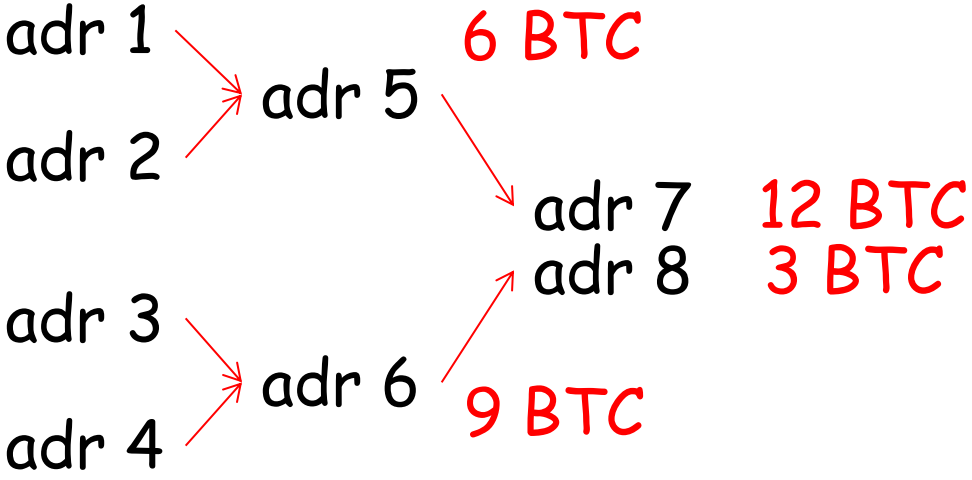


After seeing this transaction, we may predict that adr 5 and adr 6 belong to the same person.

Note: at this moment, we have no idea if adr 5 and adr 6 belong to the same person!

Q: How about adr 1 & 2; adr 3 & 4?

High chance that adr 1 & 2 belong to the same user; while adr 3 & 4 also belong to another user!



- adr 1, 2 belong to the same user
(same for adr 3, 4 & adr 5, 6)

- Usually the total input amount = total output amount.
- Most likely some of the output amount will be transferred back to the sender, e.g. adr 8
- Heuristics: small amount transferred back to sender
- If adr 8 finally used to buy pizza, then you can link up the real identity (at least the delivery address of pizza) to adr 5, 6, 8.

(ii) Another recommendation: not to include multiple addresses as input in the same transaction.

But sometimes it is difficult:
E.g. You have only 3 BTC in address A, 4 BTC in address B, but you need to pay 6BTC to an address C.

Tr 0100
Sender X
Input
adr A
adr B
Output
adr C: 6BTC
adr D: 1BTC

Then, we can deduce with high confidence that adr A, B, D belongs to the same user (the sender).

(i) At least cannot deduce easily A and B are related.

Of course, a better way (ask for 2 addresses (C and E) for receiving the money):

Tr 0101
Sender X
Input
adr A
~~adr B~~
Output
adr C: 3BTC
~~adr D: 1BTC~~

After some seconds

Tr 0104
Sender X
Input
~~adr A~~
adr B
Output
adr E: 3BTC
adr D: 1BTC

(ii) But need to ask for multiple addresses from recipient!

=> User should keep multiple addresses

Then, this is the motivation of having wallets :-P….
- Help you to keep track of all your addresses in one wallet, provide add-on services for you to use bitcoins in multiple addresses

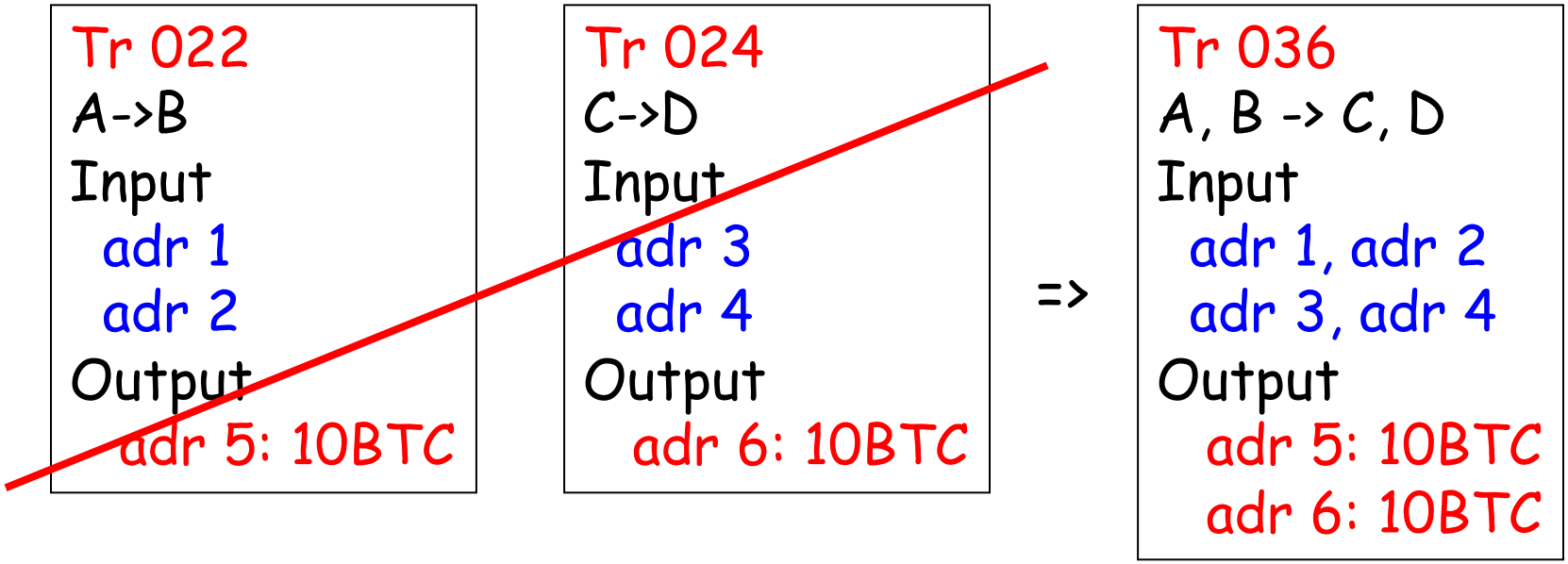
Qs:
- These wallets secure?
- Cold wallet vs hot wallet?
- How an exchange keeps track of your wallet?

Also, there exist services to help users to avoid the linkability problem (which also complicated the investigation process):

Mixer

- Idea is simple:

If A wants to send xBTC to B, and C wants to send yBTC to D, we can mix the two transactions into one!



Another short summary:

On one hand, services/functions are added and improved to to enhance anonymity and privacy;

On the other hand, from the investigation of crime cases, tracing of illegal activities become more difficult.

Also, what kind and what level of anonymity provided by each cryptocurrency may differ from one another.

- On one hand, researchers are studying how to further improve and increase the level of anonymity of these systems.
- On the other hand, crime investigators are trying their best to correlate the transactions in order to identify the suspect!

The evil side of cryptocurrencies

Bitcoin or others (e.g. Monero) is "anonymous" and difficult to trace, perfect candidates for ransom, but suspects may make mistakes + new technology to relate transactions, maybe we can still do something to trace the suspects!

- Ransomware
- Money laundry
- Cyber currencies for investment?
- Issues in cyber security exchanges

Difficult to trace once the money is converted to cryptocurrency, again we can make use of suspects' mistakes and technologies (e.g. AI) to do some.

No regulations: one news/rumors can trigger the price a lot (example? if we have time)

No regulations in most regions: many examples of problems

(ii) Money laundry: can talk about it some other time

(iii) Related to cryptocurrency exchange:

One of the instances:

Canadian cryptocurrency fund boss Gerald Cotten died – and US$190million of his investors' money may be encrypted forever

- Investors in QuadrigaCX, Canada's largest cryptocurrency exchange, have been unable to access funds since founder Gerald Cotten died in December, aged 30
- His widow says she does not know his passwords – leading some angry investors to question whether Cotten really died while opening an orphanage in India

About US$190M cannot be accessed since they claimed that only the founder (died in Dec 2018) has the access key.

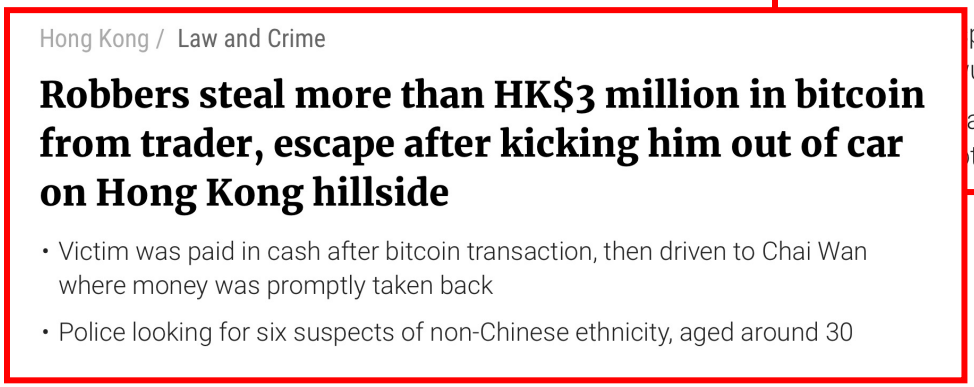
However, according to a report by Ernest & Young (court appointed monitor):
The company made a mistake transferring another 103 bitcoins to a wallet that they cannot access in Feb 2019 after the founder died!

Interestingly, some investigators claimed that all money in those wallets were emptied 8 months before CEO's death!

=> rumor: CEO faked his death and stole all money?

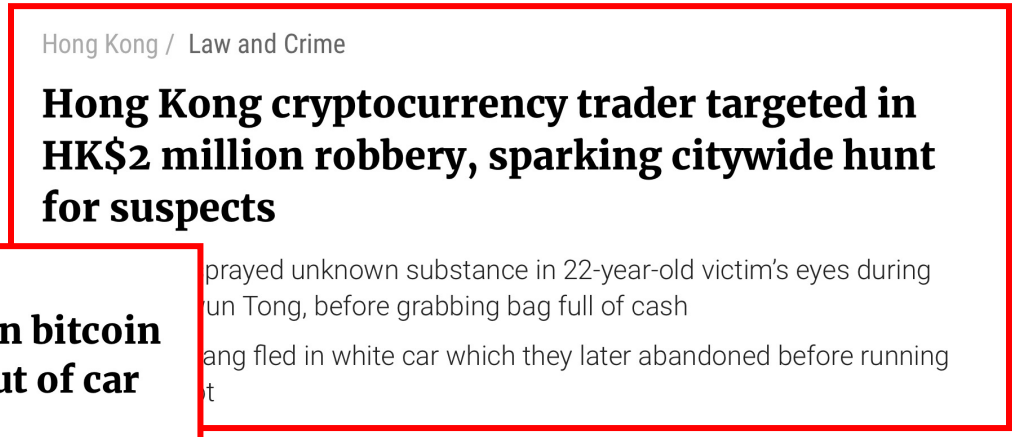
(iv) Crime cases attacking the internals of blockchain system (e.g. crypto schemes, protocols, smart contracts etc.) , talked later and must from professional hackers!

(v) Crime cases that cheats the victims to bring cash for bitcoins and rub the victims:


Hong Kong / Law and Crime

Robbers steal more than HK$3 million in bitcoin from trader, escape after kicking him out of car on Hong Kong hillside

- Victim was paid in cash after bitcoin transaction, then driven to Chai Wan where money was promptly taken back
- Police looking for six suspects of non-Chinese ethnicity, aged around 30


Hong Kong / Law and Crime

Hong Kong cryptocurrency trader targeted in HK$2 million robbery, sparking citywide hunt for suspects

prayed unknown substance in 22-year-old victim's eyes during
un Tong, before grabbing bag full of cash

ang fled in white car which they later abandoned before running
t

Some advanced techniques for tracing

(i) Analyzing wallets in a graph theoretical manner (e.g. clustering)

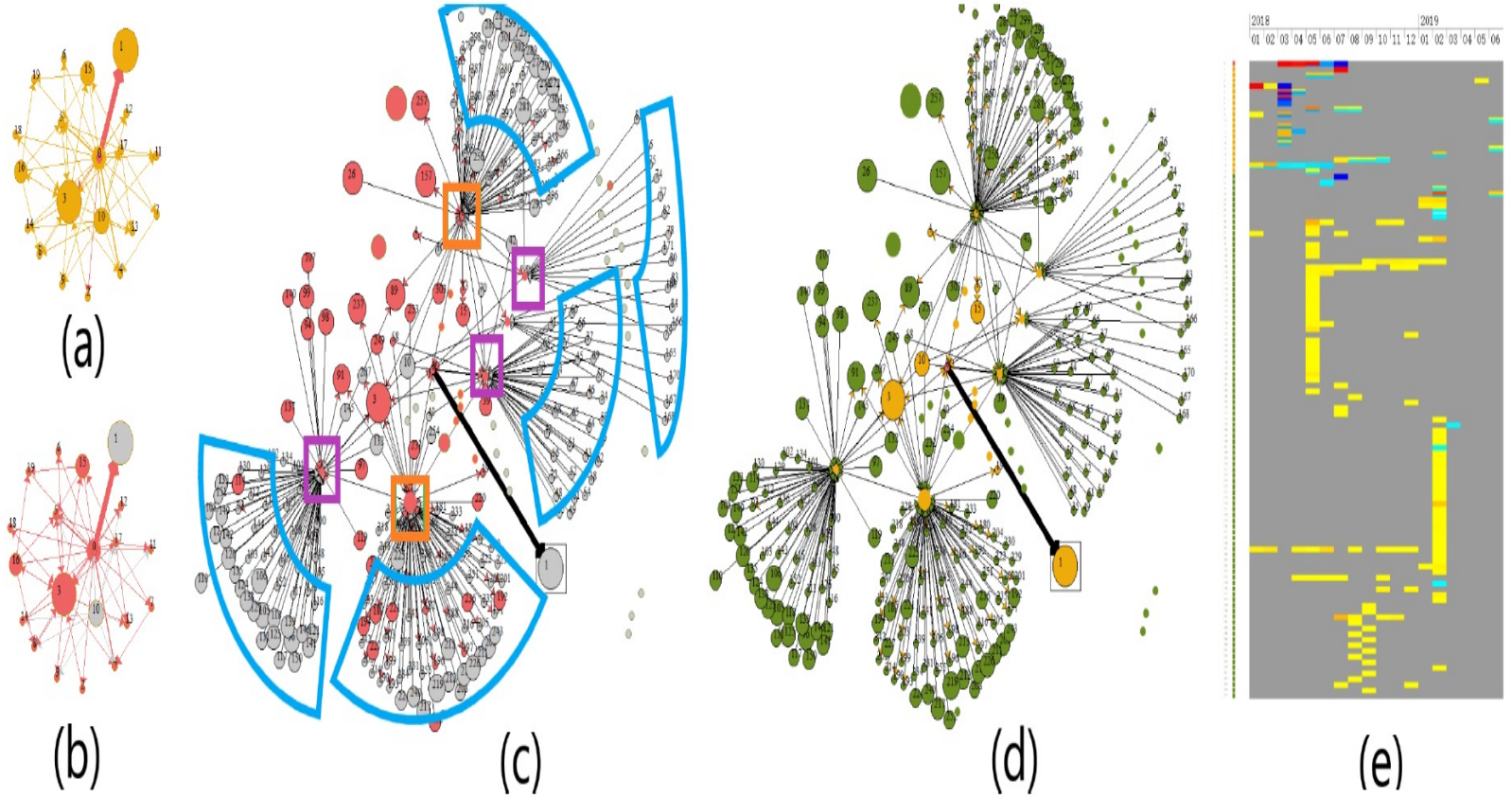


Fig. 4: Case study of whale. (a)-(d) are relationship graphs of *the whale*. (e) is the temporal analysis of all the wallets in (c)/(d). In (a) and (c), color encodes community. Red vertices are strong connected. and color grey indicates that a node does not strongly connected to any others. In (b) and (d), color encodes vertex depth. The red, yellow, blues encodes depth 0, 1, and 2, respectively.

(ii) Comparing patterns of different wallets

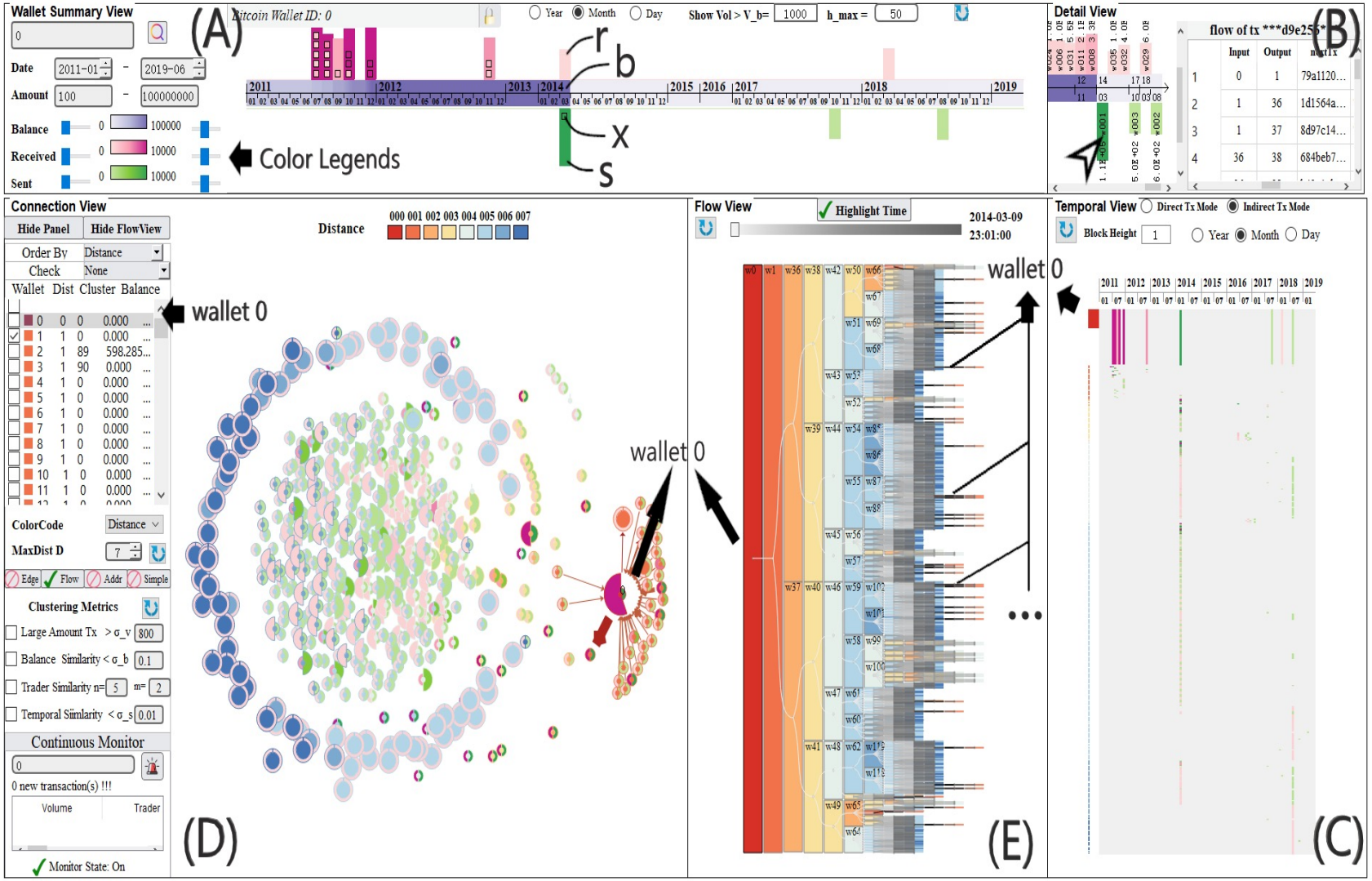
Fig. 5: A comparison between the silk road and *the whale*.

Silk Road:

A best-known wallet for selling illegal drugs!

- They both accumulated large volume of bitcoin in a period of time and then sent out most of them in one transaction.

Remark: Besides bitcoins, researchers are looking at other cryptocurrencies and mixed currency transactions!!



Our team has a new visualization system to help investigators to conduct investigation

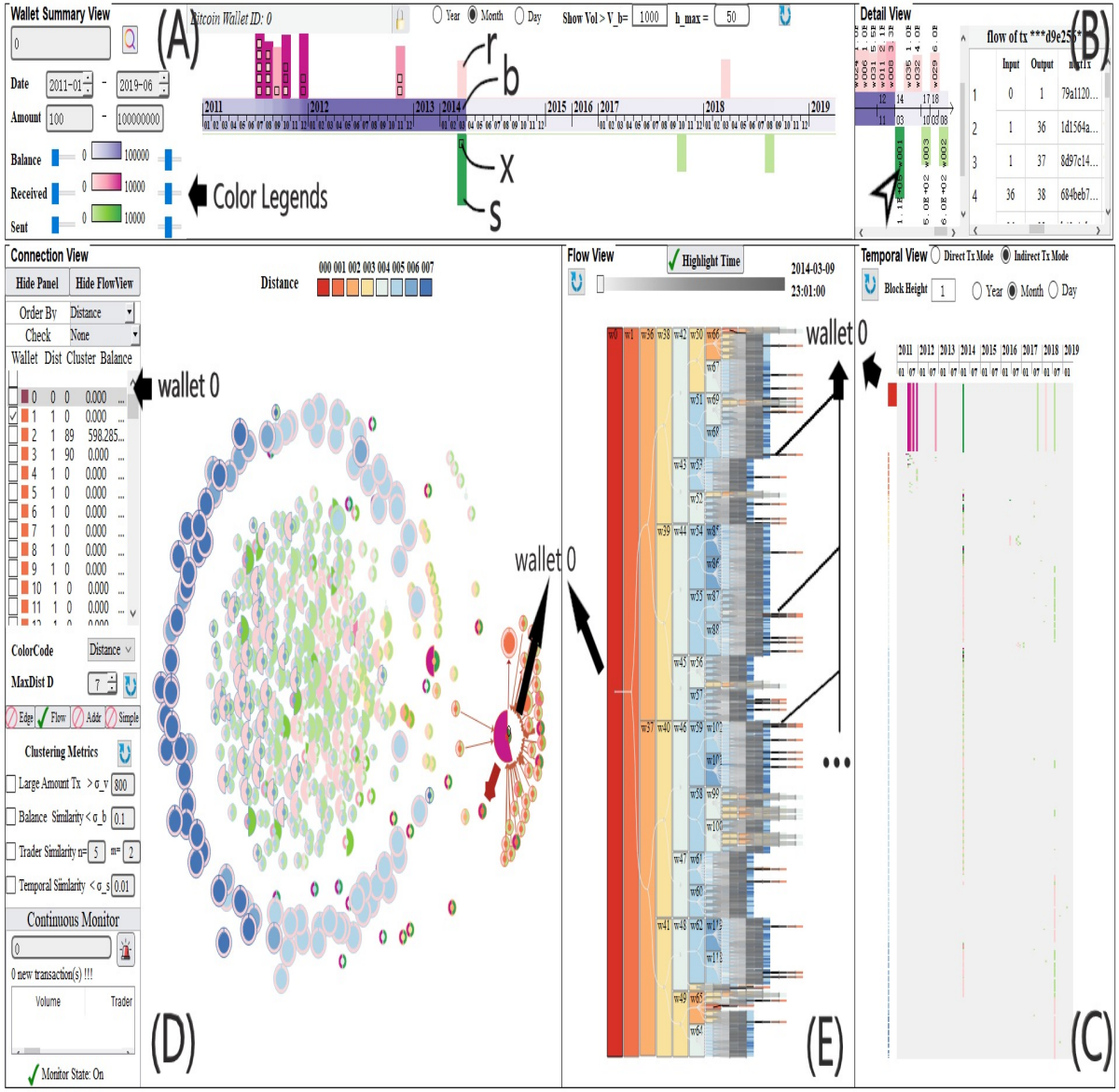
Main visualization views:

Fundamental

- A. Summary view
- B. Detail view
- C. Temporal view

Advanced

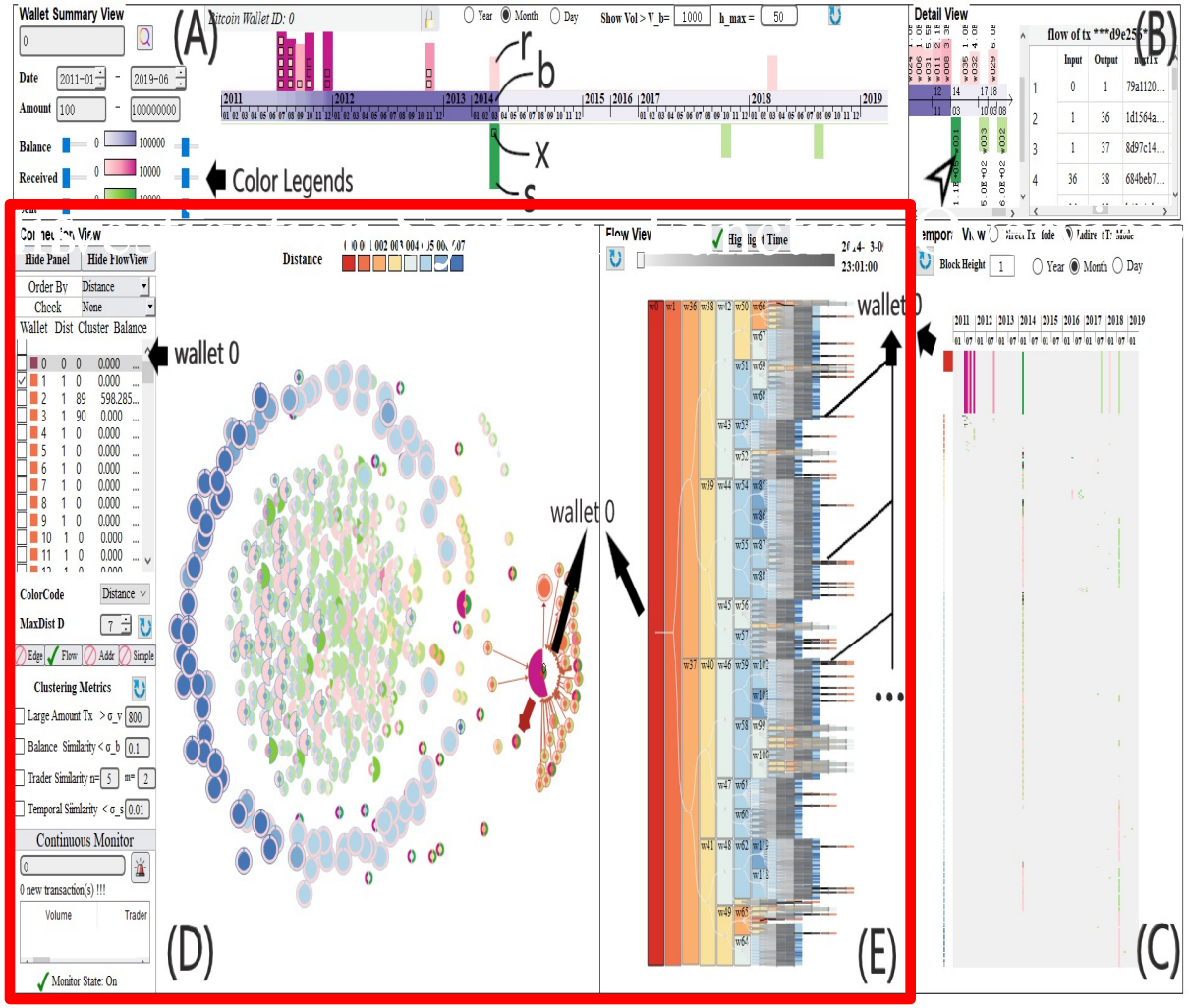
- D. Connection view
- E. Flow view



Advanced views:

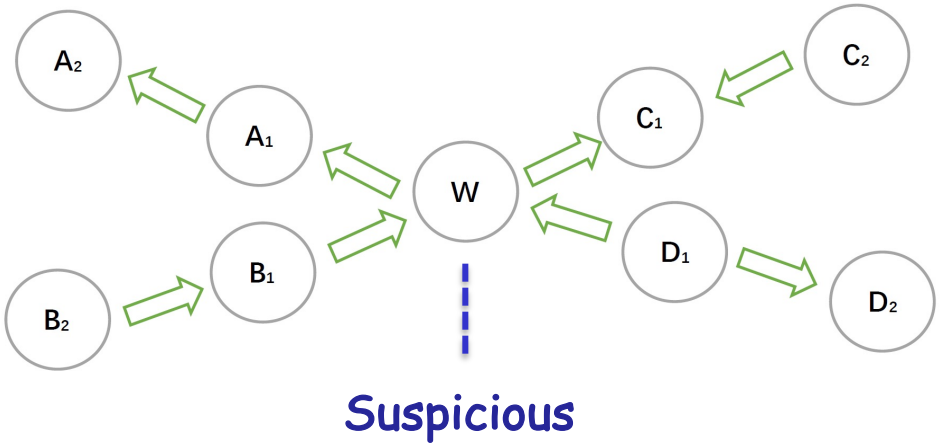
D. Connection view

E. Flow view



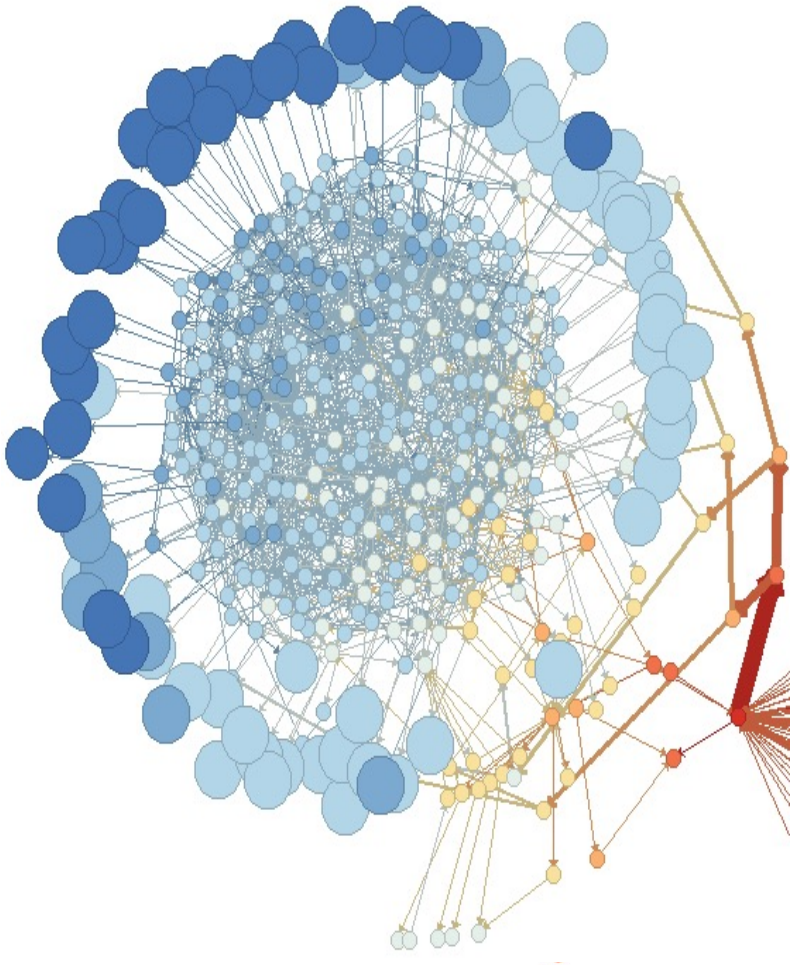
Connection Network Analysis

- **Connection network** G = {V, E} V: wallets E: transactions
- ***D-distance* connection network of wallet *w***: The network of all the wallets trading with w through D transact



2-distance connection network of w

- Circle: wallets, radius: balance
- Hide connection edges
- Easy to visualize trading distances
  - Place closer wallets trading with each other
  - Location distance correlated to trading distance
  - Color circles to indicate trading distances to wallet w



‹ Thank you ›