Blockchain: A Graph Primer

CUNEYT GURCAN AKCORA, University of Texas at Dallas

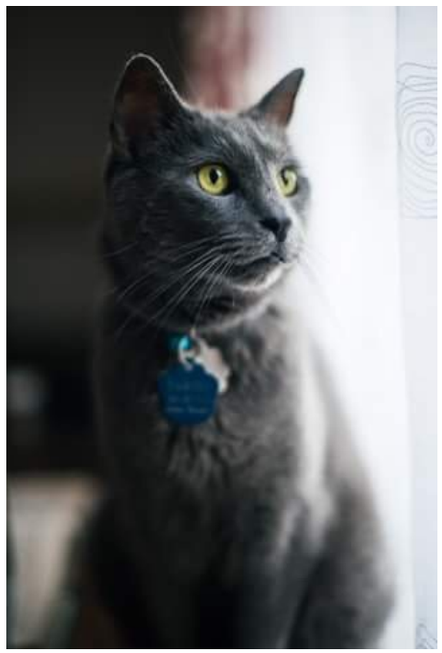
YULIA R. GEL, University of Texas at Dallas

MURAT KANTARCIOGLU, University of Texas at Dallas

Presented By **Jacob Kobernik**

Me — Jacob Kobernik

+ Software Engineer at Skuid
+ I like Lisp!
+ I spend lots of time away from my computer



Put a cat picture here… a nice bagheera one

Why this paper?

+ Cryptonomicon
+ Satoshi Nakamoto
+ My parents
+ IMF Head

- I tend to pick bad papers but its hard to know what you are going to get until you dig in!
- Just wanted to learn about blockchains and such
- “Mom and I are watching. We bought some last week. Kinda exciting…”
- Christine Lagarde (Head of International Monetary Fund) chastised her colleagues for not embracing the future of visual money — Thanks Matt Hardwick!

Cuneyt Gurcan Akcora

+ Postdoctoral Fellow at University of Texas at Dallas

+ PhD Computer Science from Italy

Featured Skills & Endorsements

Machine Learning · 17

Zonghu
this ski

Algorithms · 12

Murat A

LaTeX · 10

Mustaf
skill

Yulia R. Gel

+ Professor in the Department of Mathematical Sciences at the University of Texas at Dallas

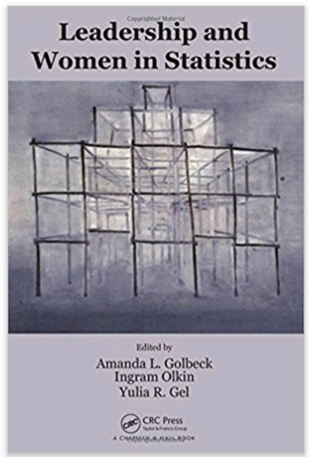
+ Interests in ML and graph mining

Leadership and Women in Statistics 1st Edition

by Amanda L. Golbeck (Editor), Ingram Olkin (Editor), Yulia R. Gel (Editor)

1 customer review

Look inside



ISBN-13: 978-1482236446
ISBN-10: 1482236443
Why is ISBN important?

Have one to sell? Sell on Amazon

eTextbook	Hardcover	Paperback	Other Sellers
$69.95	$45.56	$68.90	See all 3 versions

Buy new $45.56

Only 4 left in stock (more on the way).
Ships from and sold by Amazon.com. Gift-wrap available.

Want it Wednesday, Oct. 4? Order within **14 hrs** and choose **One-Day Shipping** at checkout. Details

List Price: ~~$69.95~~ Save: $24.39 (35%)
9 New from $45.56
FREE Shipping.

Qty: 1

Add to Cart

Turn on 1-Click ordering

Ship to:
Select a shipping address:

More Buying Choices

9 New from $45.56 | 4 Used from $77.87

13 used & new from $45.56
See All Buying Options

prime student College student? Get FREE shipping and exclusive deals LEARN MORE

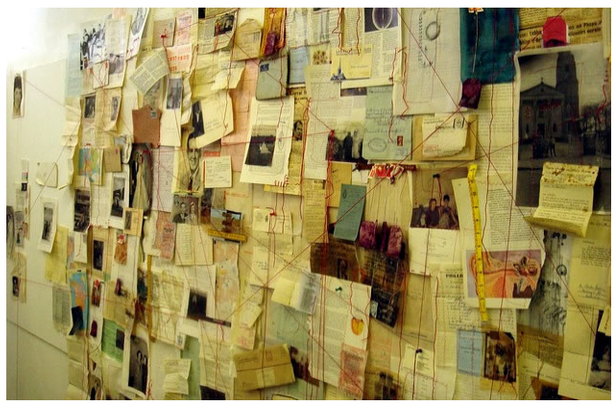
Murat Kantarcioglu

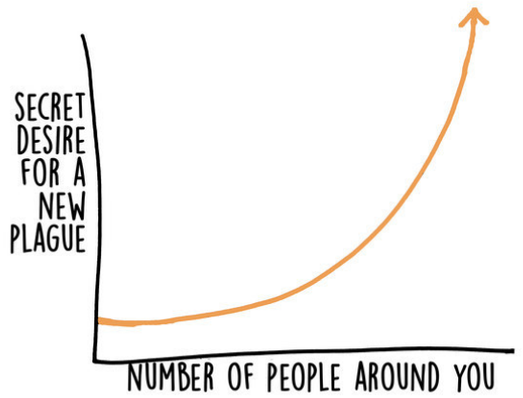
+ Professor of Computer Science, University of Texas at Dallas

+ Interests in cloud computing and "Privacy-preserving data mining"


Dr. Murat
Kantarcioglu

Graph not Graph





+ This paper's purpose is to lay a basis for people who want to understand the blockchain for its graph mining implications
+ Graph Theory: Euler's solution to the Seven Bridges of Konigsberg https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Seven Bridges of Königsberg

Euler's analysis [edit]

First, Euler pointed out that the choice of route inside each land mass is irrelevant. The only important feature of a route is the sequence of bridges crossed. This allowed him to reformulate the problem in abstract terms (laying the foundations of graph theory), eliminating all features except the list of land masses and the bridges connecting them. In modern terms, one replaces each land mass with an abstract "vertex" or node, and each bridge with an abstract connection, an "edge", which only serves to record which pair of vertices (land masses) is connected by that bridge. The resulting mathematical structure is called a graph.

Since only the connection information is relevant, the shape of pictorial representations of a graph may be distorted in any way, without changing the graph itself. Only the existence (or absence) of an edge between each pair of nodes is significant. For example, it does not matter whether the edges drawn are straight or curved, or whether one node is to the left or right of another.

Next, Euler observed that (except at the endpoints of the walk), whenever one enters a vertex by a bridge, one leaves the vertex by a bridge. In other words, during any walk in the graph, the number of times one enters a non-terminal vertex equals the number of times one leaves it. Now, if every bridge has been traversed exactly once, it follows that, for each land mass (except for the ones chosen for the start and finish), the number of bridges touching that land mass must be *even* (half of them, in the particular traversal, will be traversed "toward" the landmass; the other half, "away" from it). However, all four of the land masses in the original problem are touched by an

https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

+ Points (nodes, vertices)
+Lines (edges) with “weights” describing attributes of the relationships

Contents

The Blockchain

+A secure distributed database

+Proposed in a paper by Satoshi Nakamoto in 2008

+Applied to fields outside cryptocurrency

+Bitcoin's blockchain is a good learning-example

+ Secure by design

+Although the version we are talking about was proposed in 2008, a chain of blocks had been talked about since the early 90s. The difference was making it distributed.

+Transaction history and flow of diamonds!

+Why Bitcoin for this paper?

- Bitcoin transactions demonstrate how the relevant features of the blockchain work
- Bitcoin's blockchain is the longest in use and has the most practical experience

+Bitcoin's blockchain is currently over 100gb in size. Half of that has happened between January 2016 and January 2017!

"Nakamoto may have been the mother of Bitcoin, but it is a child of many fathers: David Chaum's blinded coins and the fateful compromise with DNB, e-gold's anonymous accounts and the post-9/11 realpolitik, the cypherpunks and their libertarian ideals, the banks and their industrial control policies, these were the whole cloth out of which Nakamoto cut the invention."

— Ian Grigg

A Brief History

+ The first digital currencies attempted to set themselves up as central authorities

+ The issue Bitcoin's immediate predecessors were trying to solve was the **central authority problem**

+ Decentralization had issues: Communication

+ The issue Bitcoin's immediate predecessors were trying to solve was the central authority problem by being decentralized but
    1. Networks are faulty and unreliable
    2. Malicious actors compounded communication issues

+ *HashCash* was an agreement between email senders and email providers that solved a spam problem

+ Nakamoto solved this by using *HashCash*'s Proof-of-Work concept in his blockchain design

+ The solution embedded in the blockchain was so successful that it has spawned hundreds of **alt-coins**

+ PoW makes lying about the blockchain nearly impossible
+ PoW forces the transaction network to slow down so that everyone can reach consensus

Building Blocks of the Blockchain

+ Addresses
+ Transactions
+ Payment Verification and Confirmation

+ Addresses = Nodes
+Transactions = Edges

1. Addresses

+ A unique string of 26-35 characters created from a public key

+ Keys => Addresses are designed to be one-way

+ Each user publishes their own address for the purpose of entering into transactions

+ PubkeyHash

+ ScriptHash

+ Users are encouraged to have many addresses
+ PubkeyHash
    1. Starts with a “1”
    2. Single private key is used to spend bitcoins
    3. Original address type and the practical standard for most transactions
+ ScriptHash
    1. Starts with a “3”
    2. Added later to support m-of-n multi signature transactions

1. Addresses

+ Graph perspective

- Either address type is suitable, both can be used for input or output to a transaction
- ScriptHash is interesting because "it allows users to participate in spending decisions". This behavior has mining implications.

+ Users are encouraged to have many addresses
+ PubkeyHash
    1. Starts with a "1"
    2. Single private key is used to spend bitcoins
    3. Original address type and the practical standard for most transactions
+ ScriptHash
    1. Starts with a "3"
    2. Added later to support m-of-n multi signature transactions

2. Transactions

+ Bitcoin allows transfers from multiple to multiple addresses
+ Each input address requires 3 pieces of data
  - ID of previous transaction for the asset
  - Index number of the output in the previous transaction
  - Amount to be transferred
+ Each input is signed to authorize the transaction
+ The receiving address need not sign the transaction.

+ Users are encouraged to have many addresses
+ PubkeyHash
    1. Starts with a “1”
    2. Single private key is used to spend bitcoins
    3. Original address type and the practical standard for most transactions
+ ScriptHash
    1. Starts with a “3”
    2. Added later to support m-of-n multi signature transactions

2. Transactions

+ **Transaction Fees**

- Senders do not specifically indicate a transaction fee, fees are usually just the difference between the sum of all inputs and the sum of all outputs
- Transaction fees are not required but help to motivate miners to include your transaction in a new block
- Transaction fees also help reduce the number of spam transactions

+ Users are encouraged to have many addresses
+ PubkeyHash
    1. Starts with a “1”
    2. Single private key is used to spend bitcoins
    3. Original address type and the practical standard for most transactions
+ ScriptHash
    1. Starts with a “3”
    2. Added later to support m-of-n multi signature transactions

2. Transactions

+ **Value amounts are Transaction Bound**
  - Addresses can receive outputs of multiple transactions
  - Community practice is to spend the total output of one transaction as the total input of the next transaction
  - Left-over value can either be a transaction fee OR be sent back to the owner by using a multiple-output transaction

+ Users are encouraged to have many addresses
+ PubkeyHash
    1. Starts with a "1"
    2. Single private key is used to spend bitcoins
    3. Original address type and the practical standard for most transactions
+ ScriptHash
    1. Starts with a "3"
    2. Added later to support m-of-n multi signature transactions
+ Online Wallets are websites created for the purpose of automating the tedious process of manual transactions

3. Verification and Confirmation

+ Payment Verification

- "Does the spender have enough money to satisfy the transaction?"
- "Does the spender want to spend this money?"
- The spender presents their public key to claim the address named as an input to the transaction
- The spender lists the ID of the previous transaction to the address
- The spender signs the transaction input with their private key

+ The spender presents their public key to claim the address named as an input to the transaction
    + Proves that they own the value at that address
+ The spender lists the ID of the previous transaction to the address
    + Proves that they have enough value at that address
+ The spender signs the transaction input with their private key
    + Proves their intent to spend this value in a transaction

3. Verification and Confirmation

+ Payment Confirmation

- Created transactions are immediately broadcasted to the network, but these transactions are tentative; they need to be **confirmed**
- **Double-Spending Problem**
- This problem is what Nakamoto's 2008 paper addresses: Blockchain
  - Each block contains a hash and ID of the previous block, so the chain is sequential
  - Blocks are limited to 1MB to so as to not clog the network
  - Bitcoin manipulates the PoW intending only 1 block being created every 10 min

+ Confirmed means the transaction has been closed-over in a confirmed block
+ Double Spending Problem: Without consensus of transaction history spenders could easily spend the same bitcoins more than once (or twice, or three times)
+

3. Verification and Confirmation

+ Block creation

- A miner is a user who works to find a confirmed block
- Miners listen for transactions on the network and collect transactions into a new block
- They add the hash of the previous block in the block chain, as well as the current ts
- The miner begins hashing the block contents with a random number (**nonce**) and checking its output against a predetermined pattern (**difficulty**)
- The miner stops hashing his block when **a)** the pattern matches, or **b)** he hears that another block that matches the pattern has been found for his blockchain

3. Verification and Confirmation

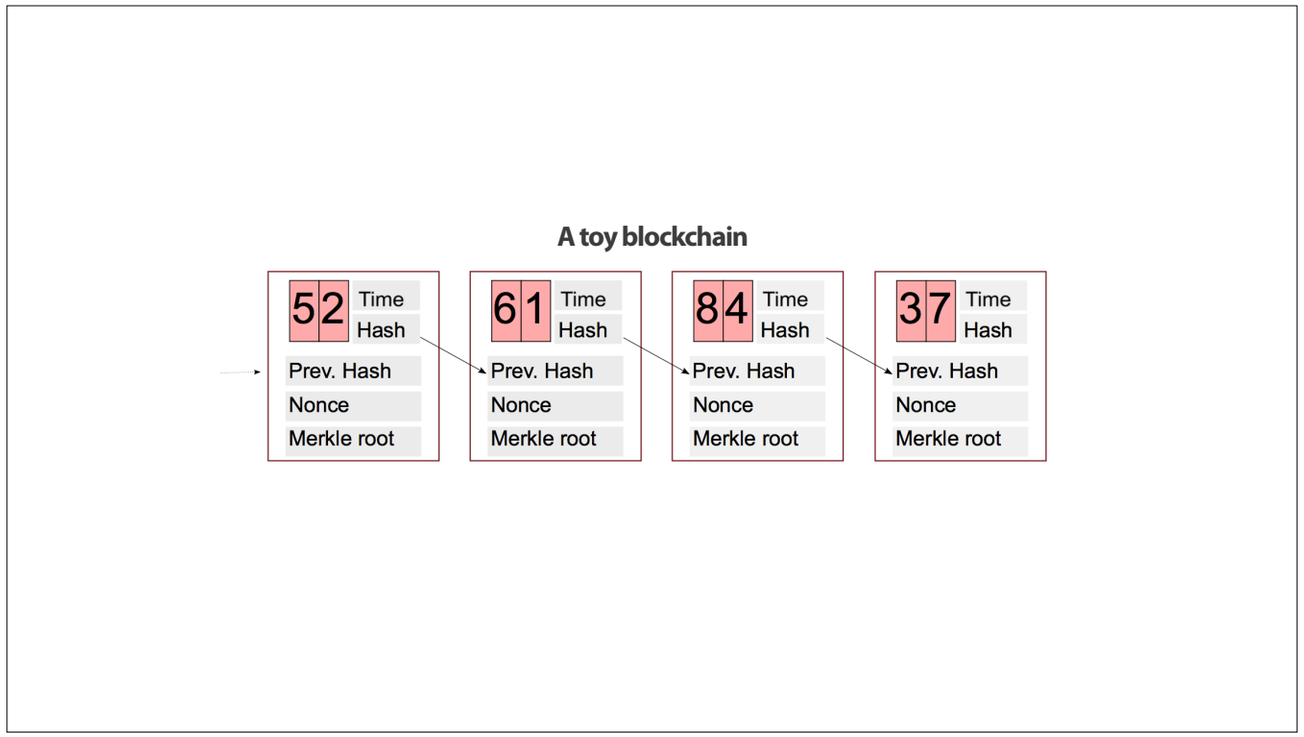
+ The newly-minted block is broadcasted to the network and miners update their blockchain accordingly

  - Miners remove transactions that are confirmed in the new block
  - They update the difficulty (if it has been changed)
  - They replace the previous block's hash with the hash of the new block

+ A user considers the block with the greatest effort the valid next block in the chain

  - Forks happen as miners finish and broadcast blocks at the same time
  - "effort": greater difficulty, failing back to the longest chain

+ All miners continue mining with the longest blockchain they have heard about, but they still hold and receive updates to another fork
+ Eventually the network will find consensus about forward progress of the longest chain


A toy blockchain
52
Time
Hash
Prev. Hash
Nonce
Merkle root
61
Time
Hash
Prev. Hash
Nonce
Merkle root
84
Time
Hash
Prev. Hash
Nonce
Merkle root
37
Time
Hash
Prev. Hash
Nonce
Merkle root

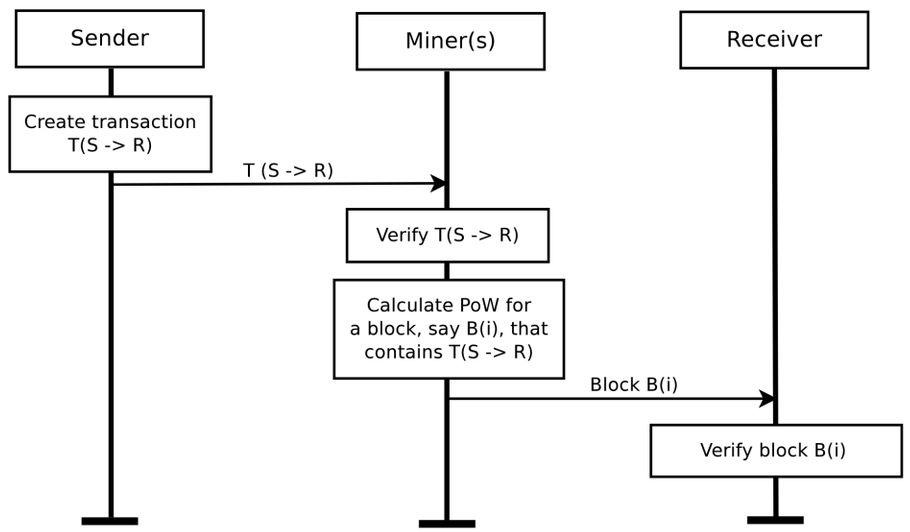


Fig. 3. Bitcoin transaction execution process: A high-level view

3. Verification and Confirmation

+ A transaction is considered tentatively confirmed after it appears in one block but in practice it isn't considered secure until after it has been buried 6 blocks deep in the chain

+ Proof-of-Work ensures that a bad actor cannot write their own transaction history and replace the blockchain except at great expense

+ Despite the protections to confirm payment there are still scams that persist to take advantage of wallets and users alike

- "Survey on Security and Privacy Issues of Bitcoin" (https://arxiv.org/pdf/1706.00916.pdf)
- An example of a simple wallet scam involves **transaction malleability**

+ Proof-of-Work Protections
    + They would have to be able to out mine the rest of community combined, which would require at least 51% of the computing power
    + In an eclipse-attack a malicious user takes over all the peers of a victim and feeds them a mutated blockchain for their own benefit
+ An example of a simple wallet scam involves transaction malleability
    + User buys BTC from a wallet and sends them to an address
    + User captures the transaction when the wallet broadcasts it
    + User changes the id of the transaction and rebroadcasts it
    + If the duplicate transaction gets picked up and put in a block then miners will reject the original
    + This can lead to the wallet refunding the original BTC value

3. Verification and Confirmation

+ In addition to transaction fees, miners receive a mining reward when they generate new blocks

- This is called the **coinbase** transaction and is usually the first transaction in the block
- The coinbase reward started at 50 coins for each block and halves every 210k blocks using a geometric series
- This series insures that there will be a maximum of 21 million bitcoins in circulation

Changes and Improvements

+ Assets over cryptocurrency

- Any hashable value can be used in transactions on the blockchain
- Asset-based blockchains still need to use currency to motivate mining
- Graph perspective: asset transactions are particularly interesting because of the diversity of assets a user may transfer

+ Any hashable value can be used in transactions on the blockchain
    + Legal documents
    + Diamonds

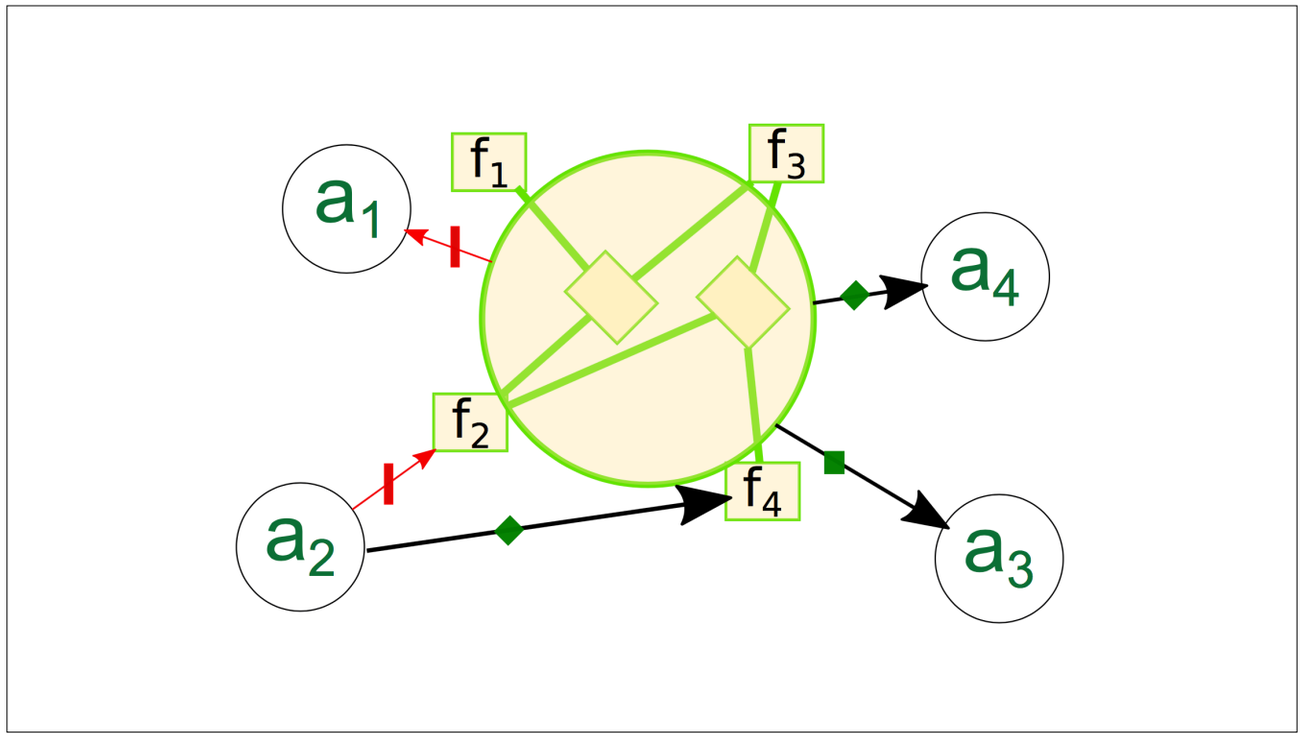
+ This would best be described by a multi-layer network where nodes are connected through networks of differing assets
    + Implications are asset network interaction (supply/demand between networks, etc)

Changes and Improvements

+ Smart Contracts

- Introduced by Ethereum
- Contracts (code) wrap a transaction and are put to an address
- The contract is publicized and confirmed in the blockchain
- A contract can be analyzed by the public and has a clear description of its guarantees
- Graph perspective: Smart Contracts can be thought of as "edge templates"; they are predefined but constrained by conditions

+ Smart Contracts are expected to have a large impact in IoT


a1
f1
f3
a4
f2
f4
a2
a3

Changes and Improvements

+ Decentralized Autonomous Organizations
  - Organizations are mostly held together by contracts, replacing these with smart contracts automates much of the organization and removes the bureaucratic overhead
  - Corporations (DAC) could be small entities that live entirely on the blockchain and yet safely provide trust and guarantees to investors
  - Ethereum's creator Vitalik Buterin holds this view:
    - DAO is an entity that lives on the internet and exists autonomously
    - DAO relies on hiring to perform certain tasks that the automaton itself cannot do
    - DAO contains some kind of internal property that is valuable in some way

+ "The DAO" was the largest example that unfortunately ended in disaster

Changes and Improvements

+ Fork Issues

- Two types of blockchain forks: **Soft** and **Hard**
- The most famous hard fork involves Ethereum and "the DAO" which was hacked in 2016
  - The disagreement was over returning the stolen money
- Bitcoin was hard forked in August of 2017 resulting in Bitcoin and Bitcoin Cash
  - The disagreement was over increasing transaction throughput

+ Soft Forks
  + Backwards compatible
  + Small updates to features and functionality
  + Considered to be improvements or extensions
+ Hard Forks
  + Split in the main chain
  + Two continuing chains are maintained by different groups
  + Often results from a divergence in ideology

Changes and Improvements

+ Tokens
  - **Initial Coin Offerings** or **ICO**
+ Hyperledger Fabric: Enterprise Blockchain
+ Scalability Issues
  - Bitcoin is capped to a maximum of 7 transactions per second (pre August 2017)
  - Other coins have fiddled with the numbers so as to allow more transactions

+ Compare that to the VISA network which does 2000 transactions per second on average

Changes and Improvements

+ Proof-of-X

- An umbrella term that covers Proof-of-Work alternatives
  - **Proof-of-Stake**: takes into consideration the age of the coins in a miner's block
  - **Proof-of-Burn**: Miners sacrifice coins to win the transaction fees and coinbase
- No Proof-of-Work alternative has shown to be as good at discouraging miners from supporting every fork they hear of

+ In 2014 it was estimated that a single bitcoin cost 15.9 gallons of gasoline
+ "Peer-to-peer heat engine"

Blockchain Analysis

+ Theoretically, you can trace the history of every bitcoin in existence

- **Taint Analysis** is used to find whether a unit of currency is tainted because it is acquired through a crime
  - Implications on **fungibility** of bitcoins
- Analysis spread and now covers many topics from predicting the value of bitcoins to scalability studies

"Based on the aforementioned discussion in Section V, it is evident that the public nature of the blockchain poses a significant threat to the privacy of Bitcoin users. Even worse, since funds can be tracked and tainted, no two coins are equal, and fungibility, a fundamental property required in every currency, is at risk."

— A Survey on Security and Privacy Issues of Bitcoin

Blockchain Analysis

+ Application (content) Graphs
  - Transaction Graphs
    - Acyclic
  - Address Graphs
    - Cyclic
  - Entity/User Graphs
    - Concerned with using clustering to link addresses

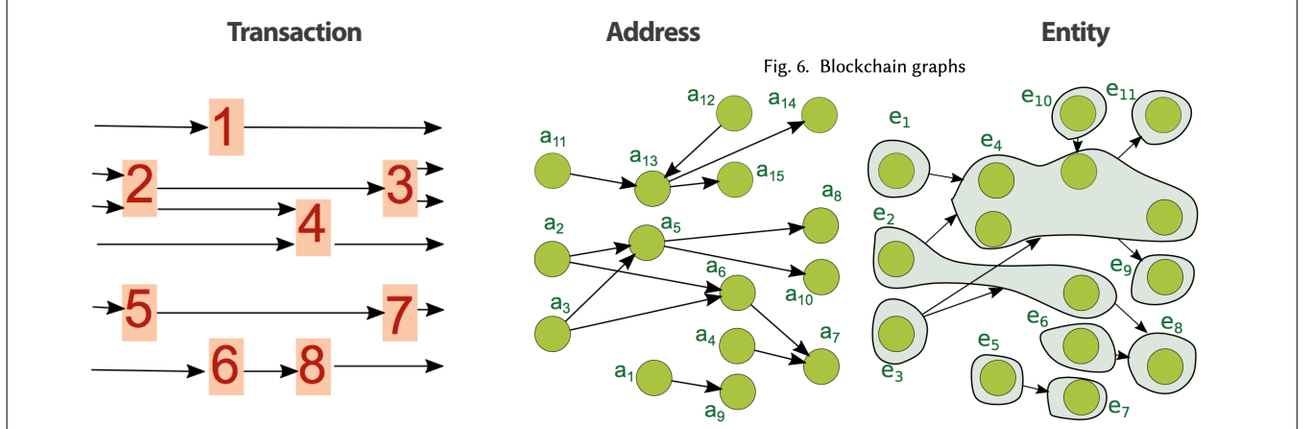


Fig. 6. Blockchain graphs

Blockchain Analysis

+ Uncovering identities behind addresses
  - No clear-cut method of linking addresses
  - Common heuristics for clustering, all error prone
  - Community practices complicate matters
    - **CoinJoin** and mixing inputs
    - Wallets trade asset ownership without transactions
  - Privacy concerns vs finding criminal entities are ongoing discussions

Blockchain Analysis

+ Communication Graphs
  - Persistent TCP peer-to-peer connection
  - Topology of network is organic
  - Relay of new blocks and transactions are priority, even when mining
  - **Trickling** of communication
    - User sends hash of transaction id, peer responds if it needs details
    - Also how new transactions originate

+ Trickling: a random 1/4 of transactions from their pool is sent to their peers

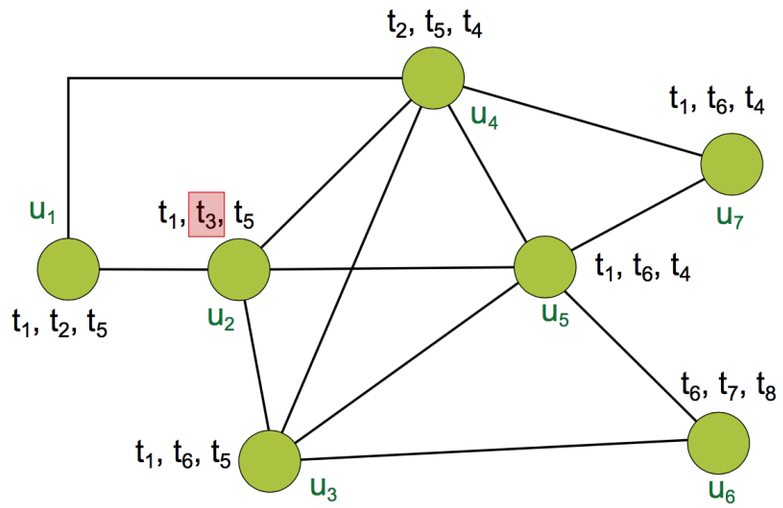
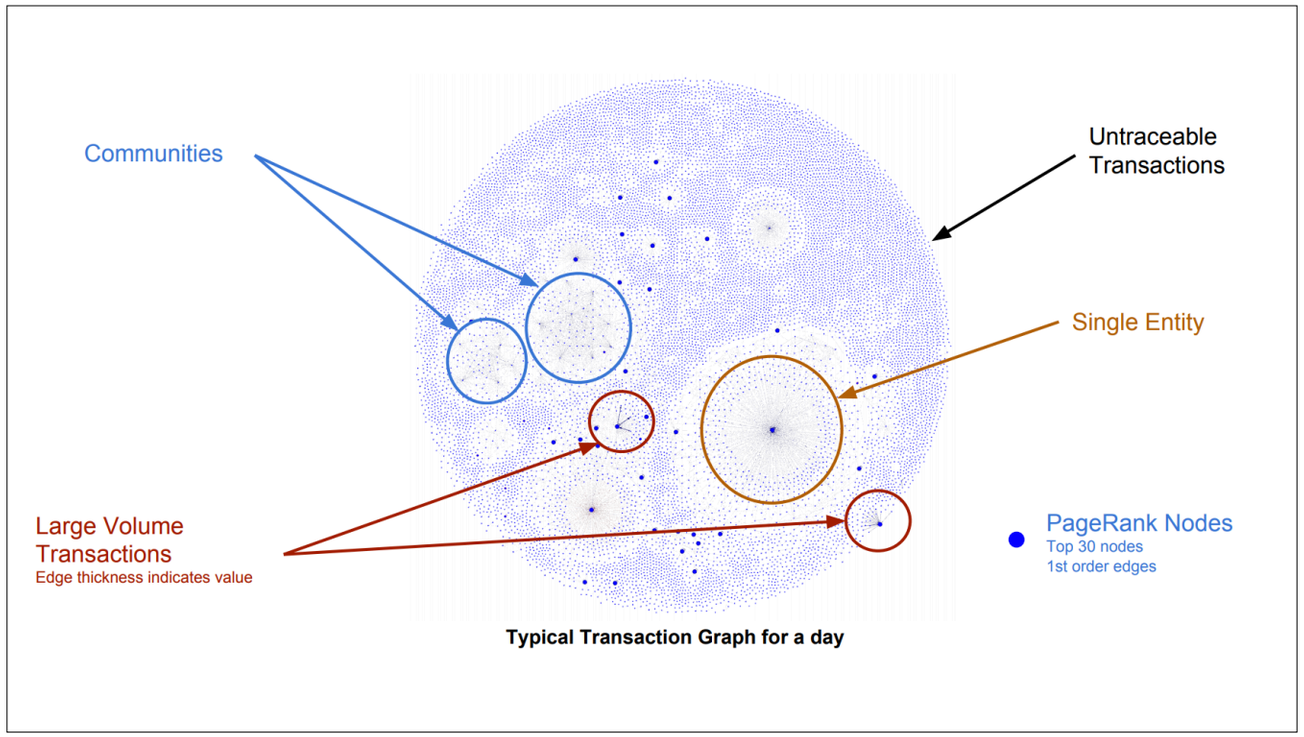
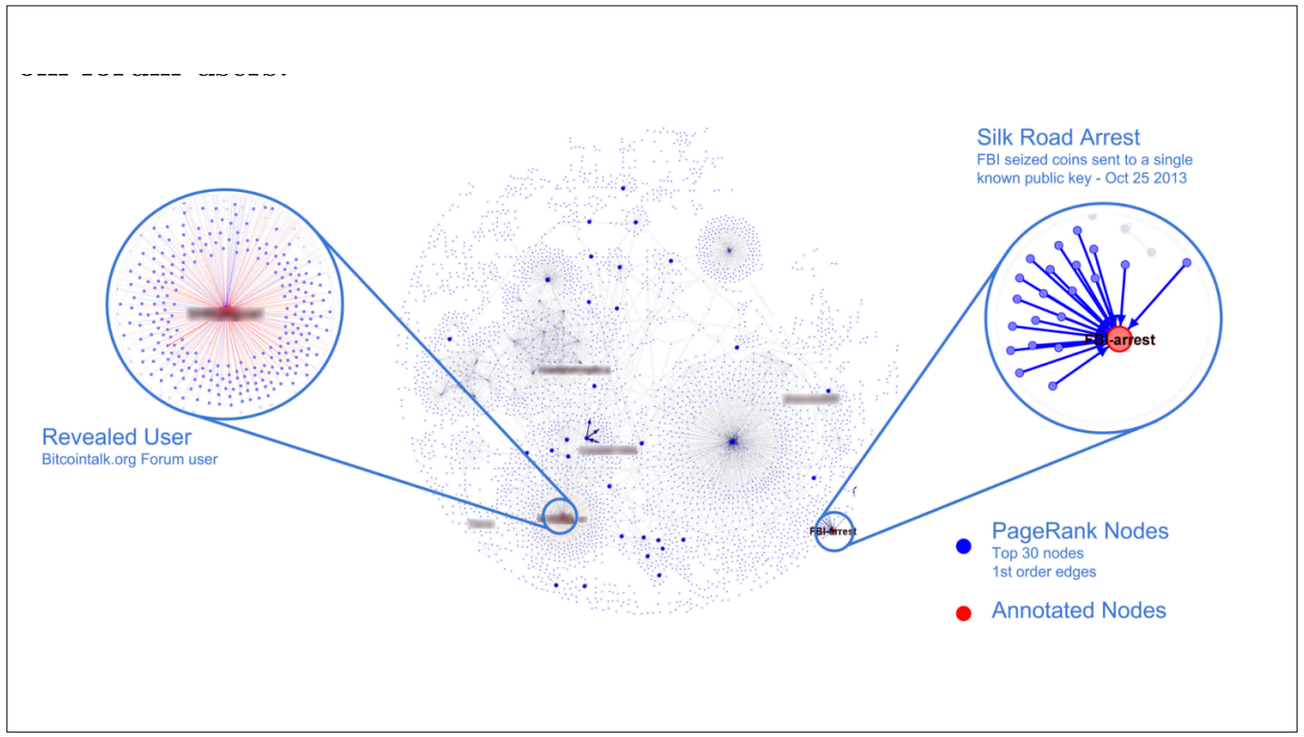


Fig. 8. A graph with seven peers and three trickled transactions at each peer.



Typical Transaction Graph for a day


Silk Road Arrest
FBI seized coins sent to a single known public key - Oct 25 2013
FBI-arrest
Revealed User
Bitcointalk.org Forum user
FBI-arrest
PageRank Nodes
Top 30 nodes
1st order edges
Annotated Nodes

Thanks!