Understanding Bitcoin’s Network Topology

Andrew Miller
Scaling Bitcoin
Sep 2015, Montreal

Tools to study Bitcoin’s p2p network

- Shadow-Bitcoin

  Scalable bitcoin simulator framework

- Coinscope

  Active/passive network measurement station

Simulating Bitcoin - Approaches

- Customized “model” (e.g., simbit)

  May differ from actual node behavior

- Local private network: (see following talk)

  Not deterministic/repeatable

  Less fine grain control (must avoid slowdown)

- Simulator/emulator

  Run the real code, with a simulated network stack

Shadow-Bitcoin (CSET ‘15)

Shadow (Rob Jansen’s PhD thesis):

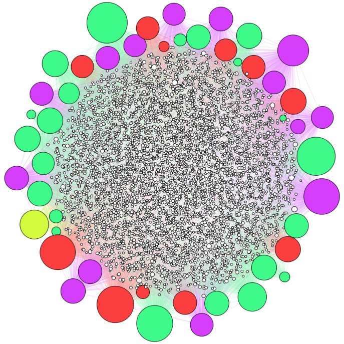
- Framework for simulator/emulator
- Previously used to study Bittorrent, Tor

Challenges: support multithreaded applications

Result: up to 6k nodes on a server, 1/14 of realtime

Caveat: how do we form the network graph?

https://github.com/shadow/shadow-plugin-bitcoin



Coinscope
Network Measurement

James Litton, Andrew Pachulski, Neal Gupta,
Dave Levin, Neil Spring, Bobby Bhattacharjee.

Periodic scans (every 4 hours)

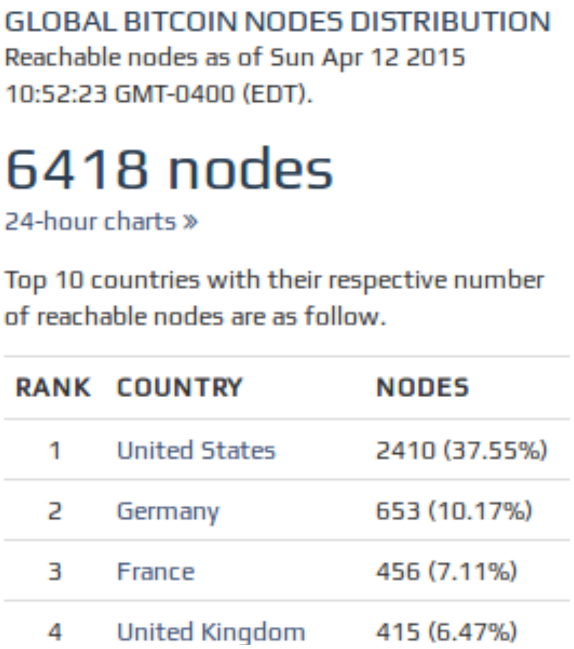
Our focus is network health (not “deanonymization”)

Safety disclaimers: /UMDCoinscope/, 1 outgoing connection

Extended version of getaddr.bitnodes.io

What do we know about the network?

- How many&where, but not their connections



GLOBAL BITCOIN NODES DISTRIBUTION

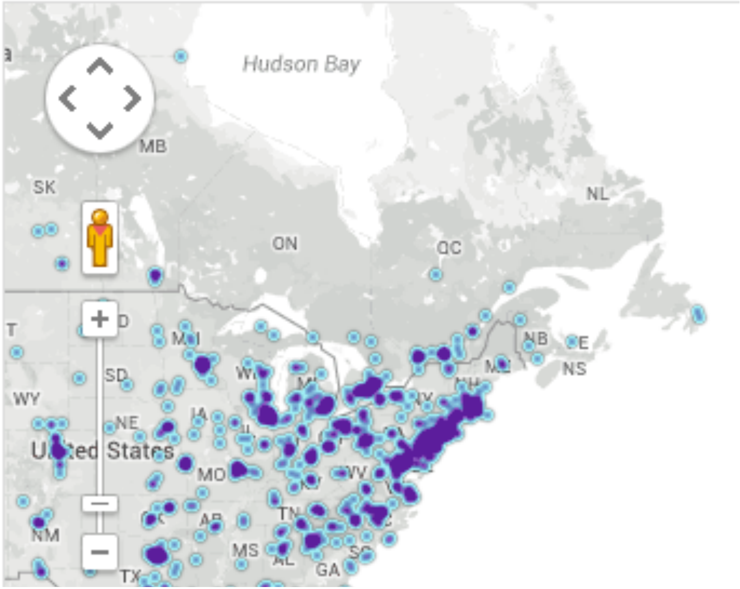
Reachable nodes as of Sun Apr 12 2015 10:52:23 GMT-0400 (EDT).

6418 nodes

24-hour charts »

Top 10 countries with their respective number of reachable nodes are as follow.

RANK	COUNTRY	NODES
1	United States	2410 (37.55%)
2	Germany	653 (10.17%)
3	France	456 (7.11%)
4	United Kingdom	415 (6.47%)



Bitcoin strives for a random graph

- Form 8 outgoing connections
- Allow up to 117 incoming connections
- Store and propagate info about peers

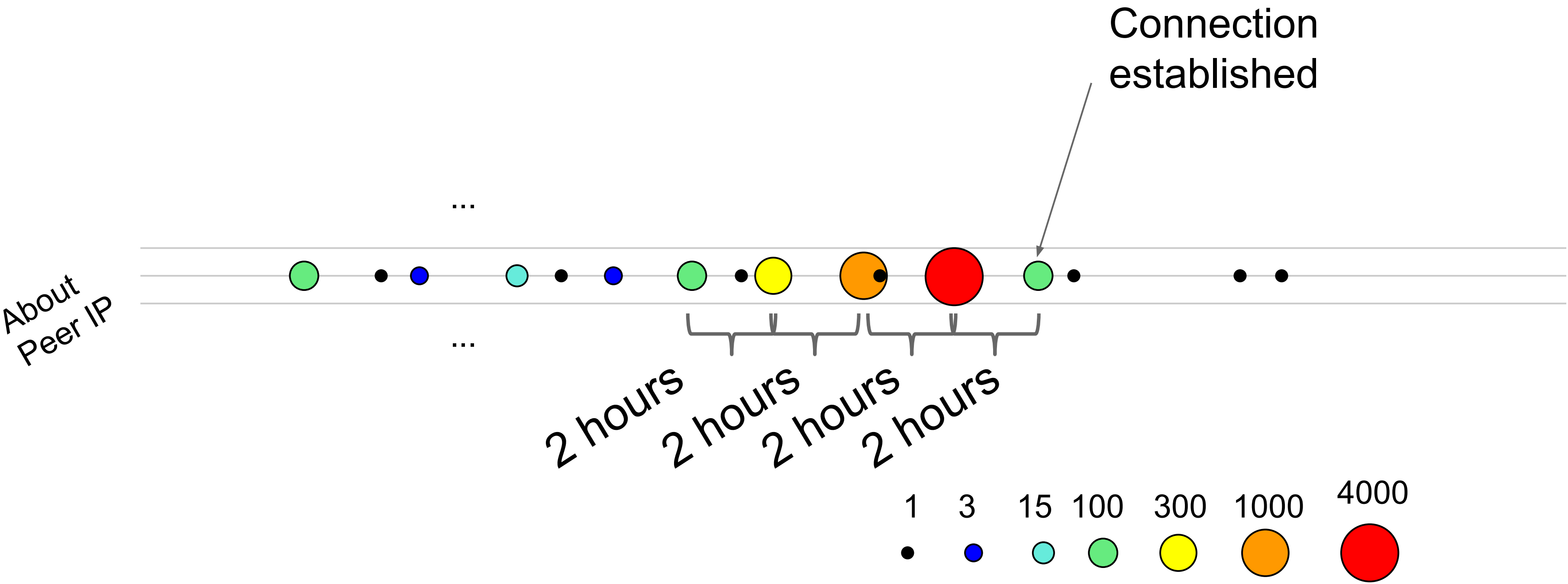
  AddrMan: Addresses and (last seen) timestamps

We scrape the AddrMan from each node and use it to infer the network topology

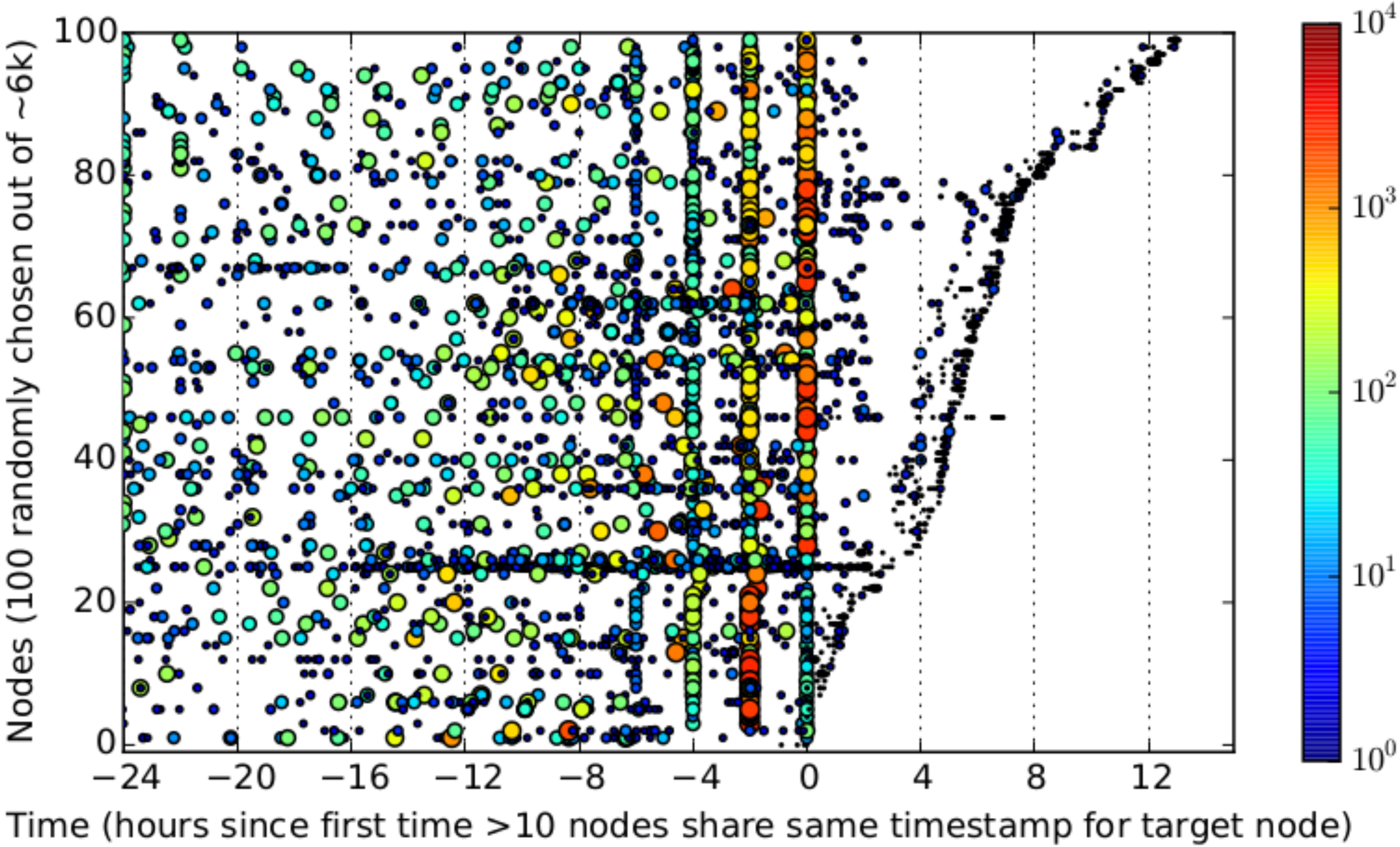
How addresses propagate

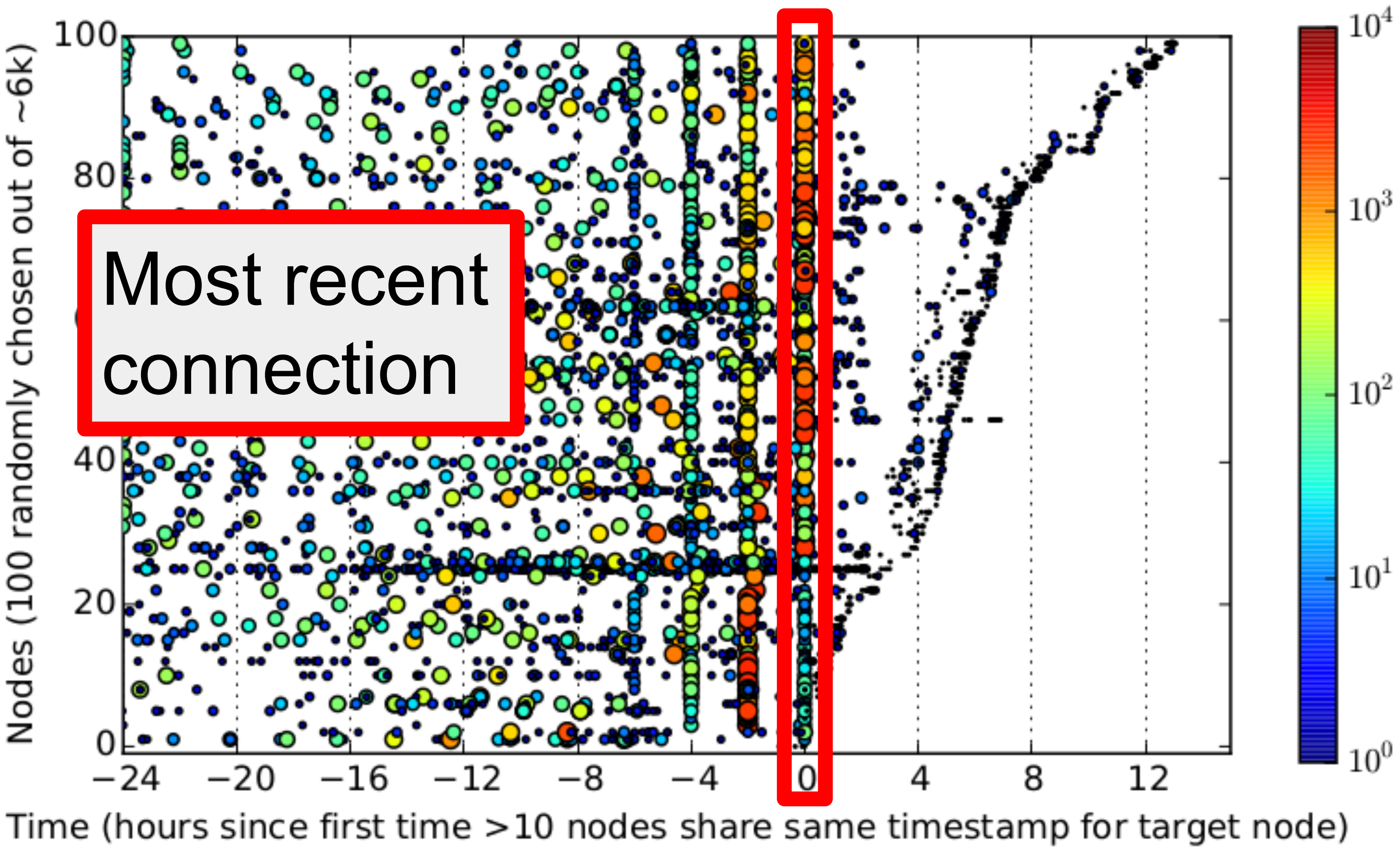
- **Relay**
    - Upon new connection (initiator only)
    - Every 24 hours
- **In response to “GetAddr”**
    - 2500 exchanged at a time
    - Upon new connection

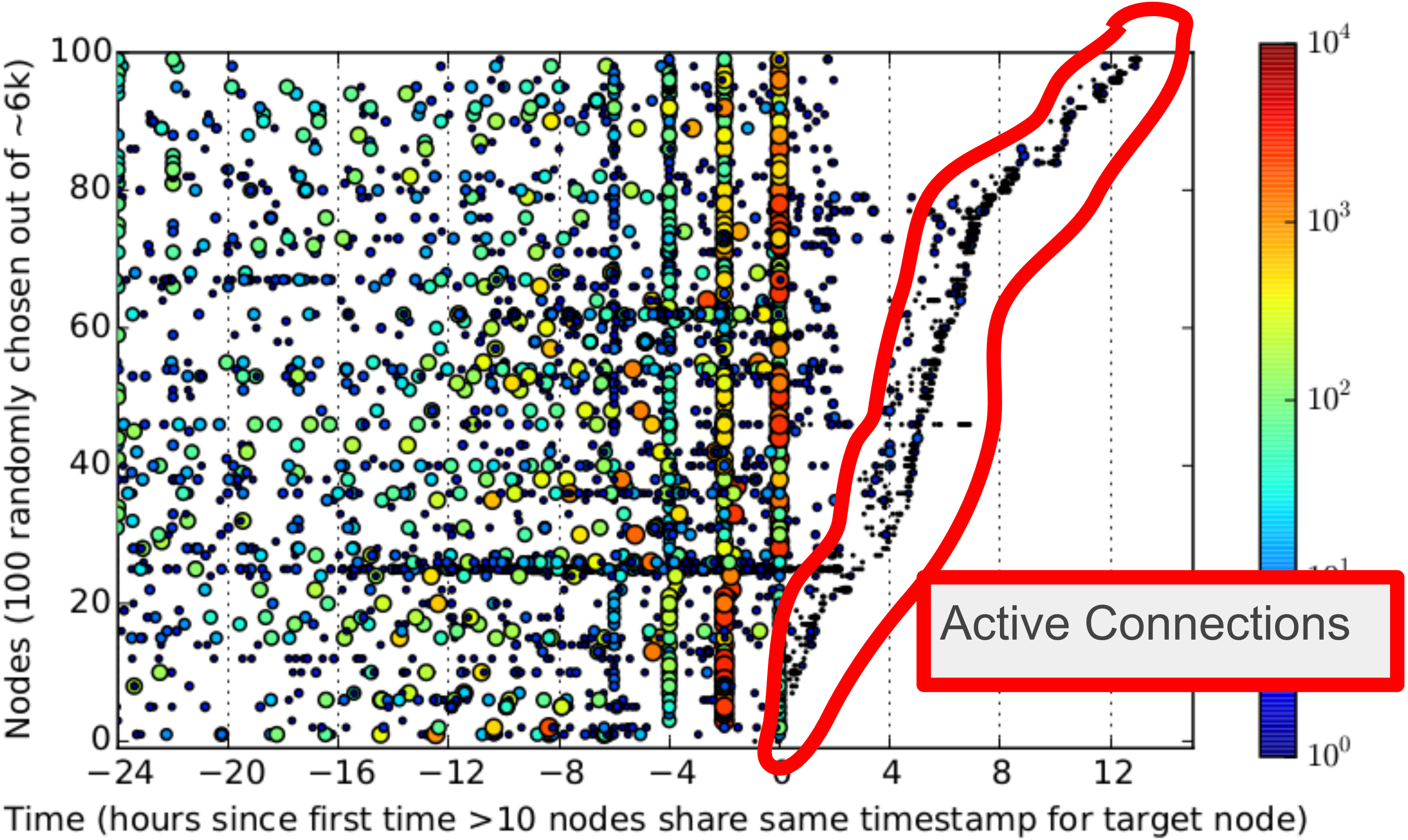
Echoes of prior connection events



Number of “from” nodes reporting the same timestamp about a node.


Nodes (100 randomly chosen out of ~6k)
100
80
60
40
20
0
−24
−20
−16
−12
−8
−4
0
4
8
12
Time (hours since first time >10 nodes share same timestamp for target node)
10^4
10^3
10^2
10^1
10^0


Most recent connection
Nodes (100 randomly chosen out of ~6k)
100
80
40
20
0
−24
−20
−16
−12
−8
−4
0
4
8
12
Time (hours since first time >10 nodes share same timestamp for target node)
10^4
10^3
10^2
10^1
10^0



Active Connections

Results

Only the “reachable” subgraph

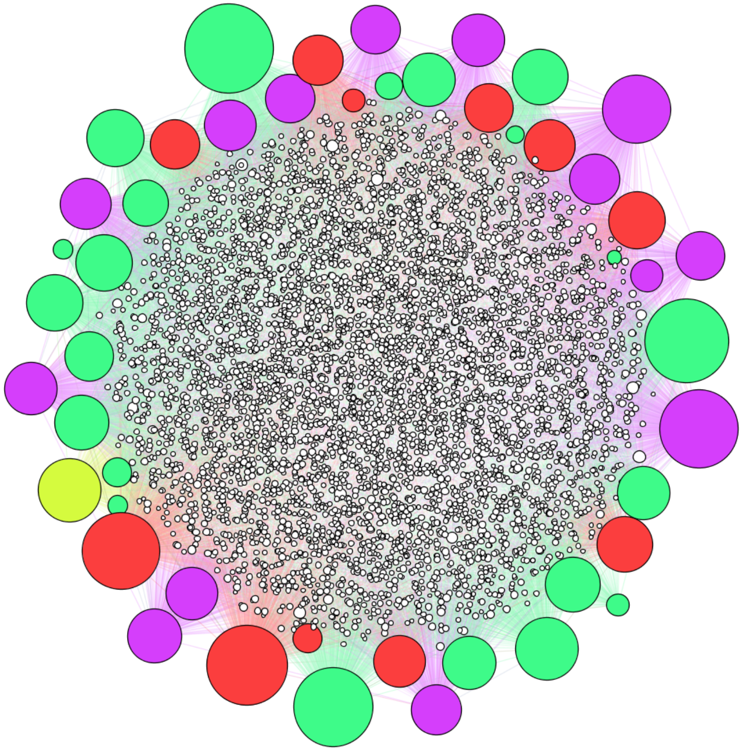
Mostly random, mostly low degree

Super nodes detected:

“bitcoinaffiliate” miners:

~40 nodes with 1k+ connections

More in paper….



Bitcoin avoids measurement

Patch in v0.10.1 breaks AddrProbe

- after a “deanonymization”-themed report

Backup technique: TxProbe

- (invasive, expensive, we don’t do it)

Visible network may be irrelevant anyway

- Private miner peering
- BlueMatt’s optimized miner relay network

Conclusions

- Let’s make measurement an active goal
  - Attackers will use invasive techniques (researchers won’t)
    - Tor has privacy preserving usage stats collection
    - Statoshi
- Fortifying the P2P network is essential,
  will affect other technical decisions

[PRE-ANN] Ledger Journal

- Main goal: useful, efficient peer review
  bridging academia & Bitcoin dev

- Open access (no @#^% paywalls)

- Reviews are published along with articles

- Articles signed and timestamped