Anonymity in the Bitcoin Peer-to-Peer Network

Giulia Fanti

Joint work with: Shaileshh Bojja Venkatakrishnan, Surya Bakshi, Brad Denby, Shruti Bhargava, Andrew Miller, Pramod Viswanath

Carnegie Mellon University

“Untraceable Bitcoin”

Teenagers using untraceable currency Bitcoin to buy dangerous drugs online

Fears have been raised as children as young as 14 are getting parcels of legal highs delivered to their home

Mirror

This is false.

Bitcoin Primer

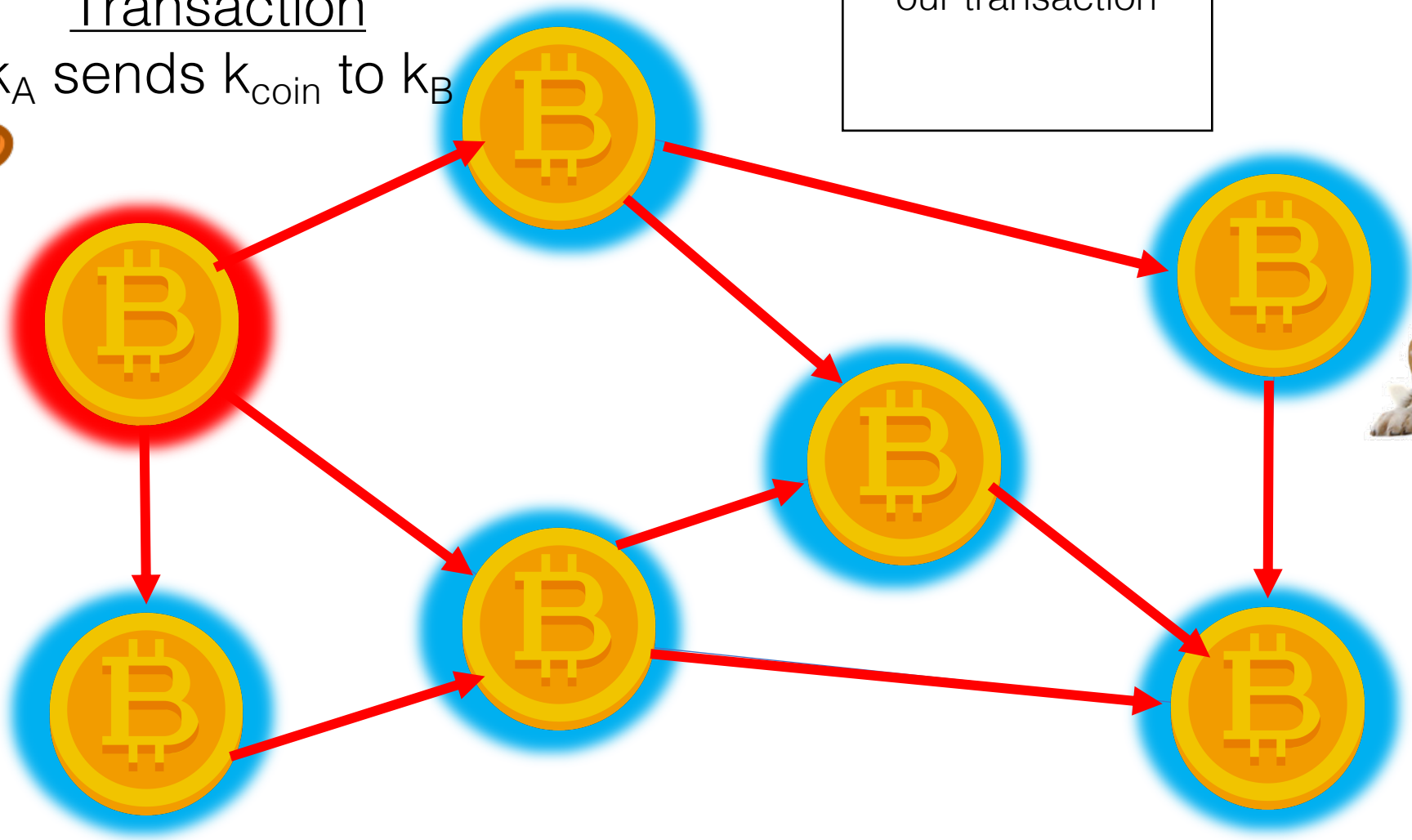
Transaction

k_A sends k_{coin} to k_B

Blockchain
sd93fjj2
pckrn29
…
our transaction

Alice
k_A

k_{coin}



Bob
k_B

Multiple Identities

Alice

Used nowhere

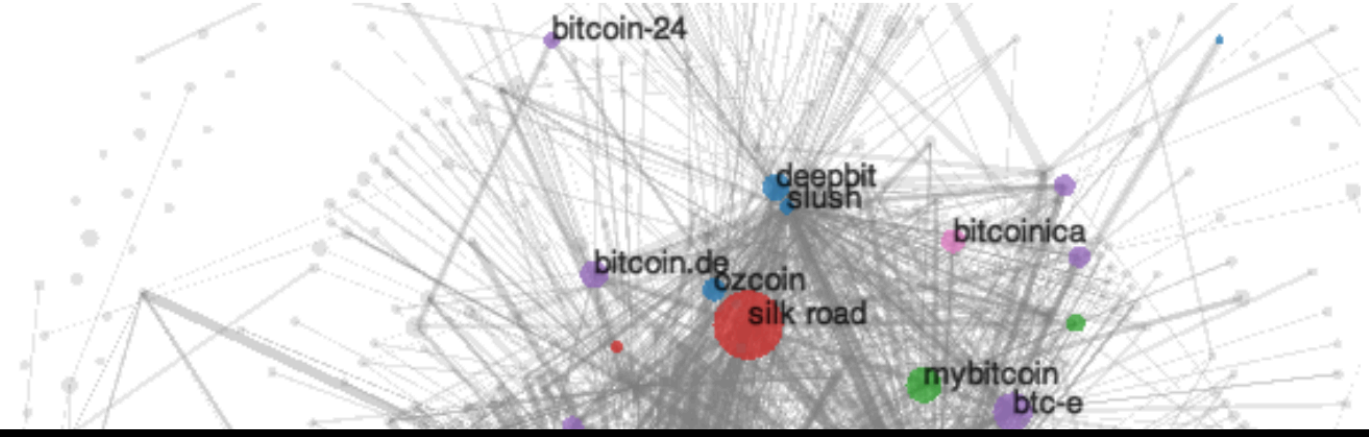
Used in the Blockchain

Public Key

Used in the P2P Network

IP Address

How can users be deanonymized?



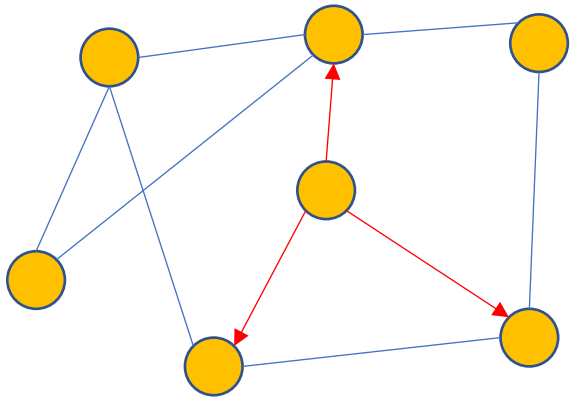
Entire transaction histories
can be compromised.

Meiklejohn et al., 2013
Ober et al., 2013

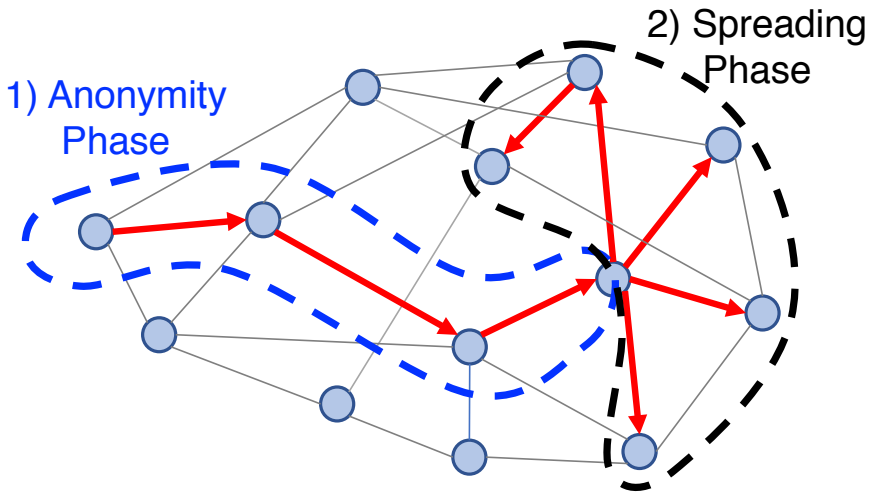
What about the peer-to-peer network?

This Talk

How to break privacy



How to fix it

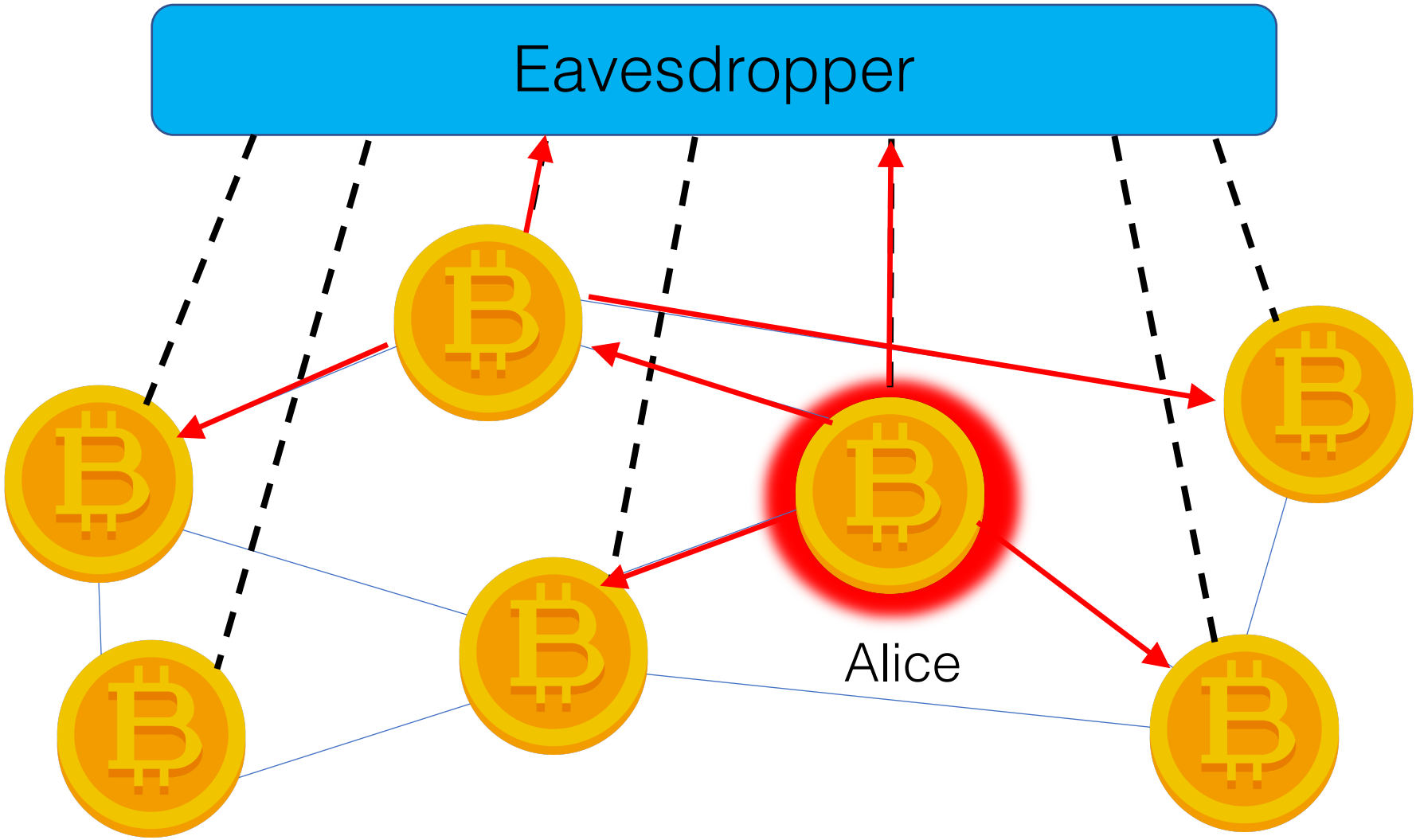


Early attacks

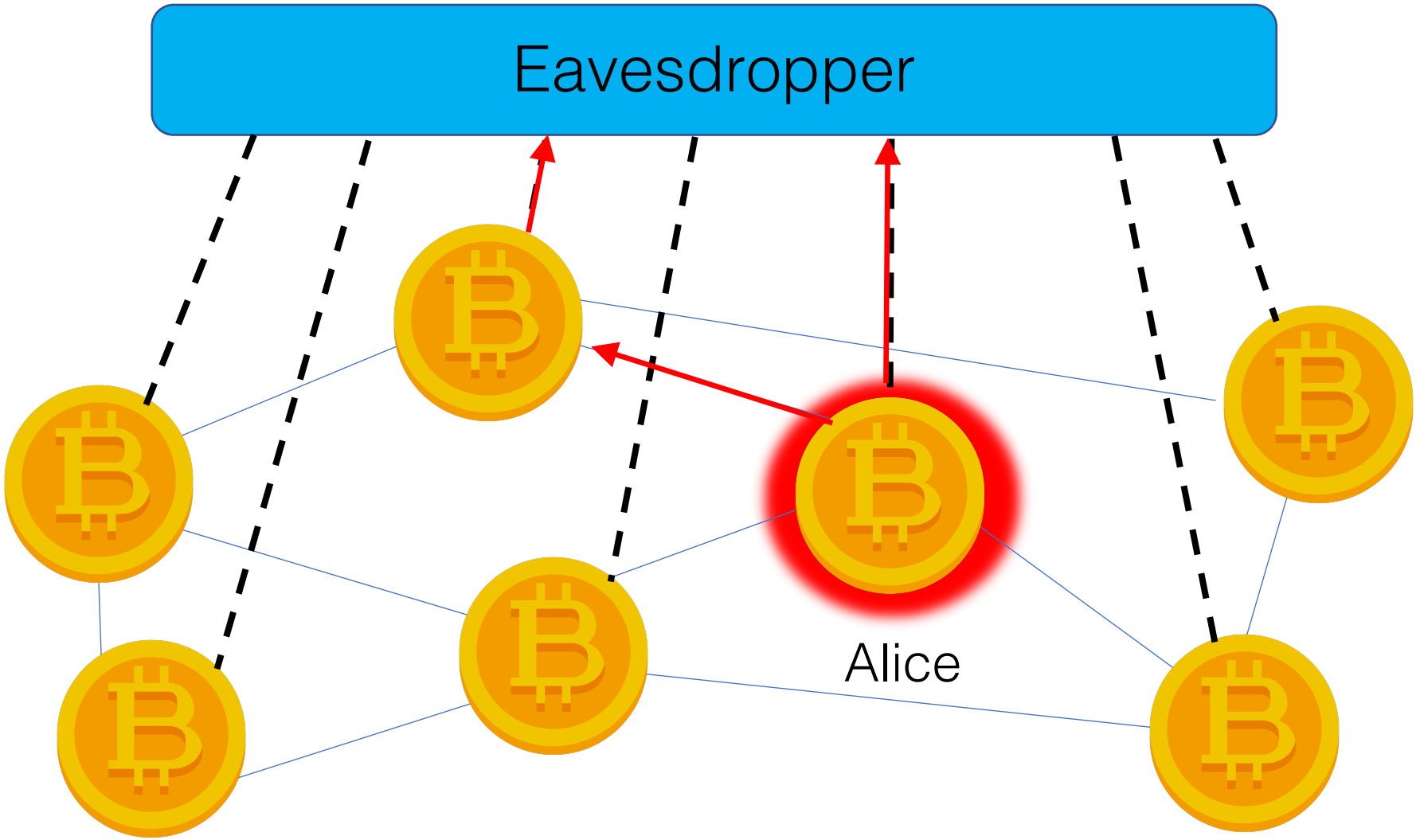
- A. Biryukov, D. Khovratovich, I. Pustagurov, *"Deanonymisation of clients in Bitcoin P2P network"*, CCS 2014

- P. Koshy, D. Koshy, P. McDaniel, *"An analysis of anonymity in Bitcoin using P2P network traffic"*, Financial Crypto 2014

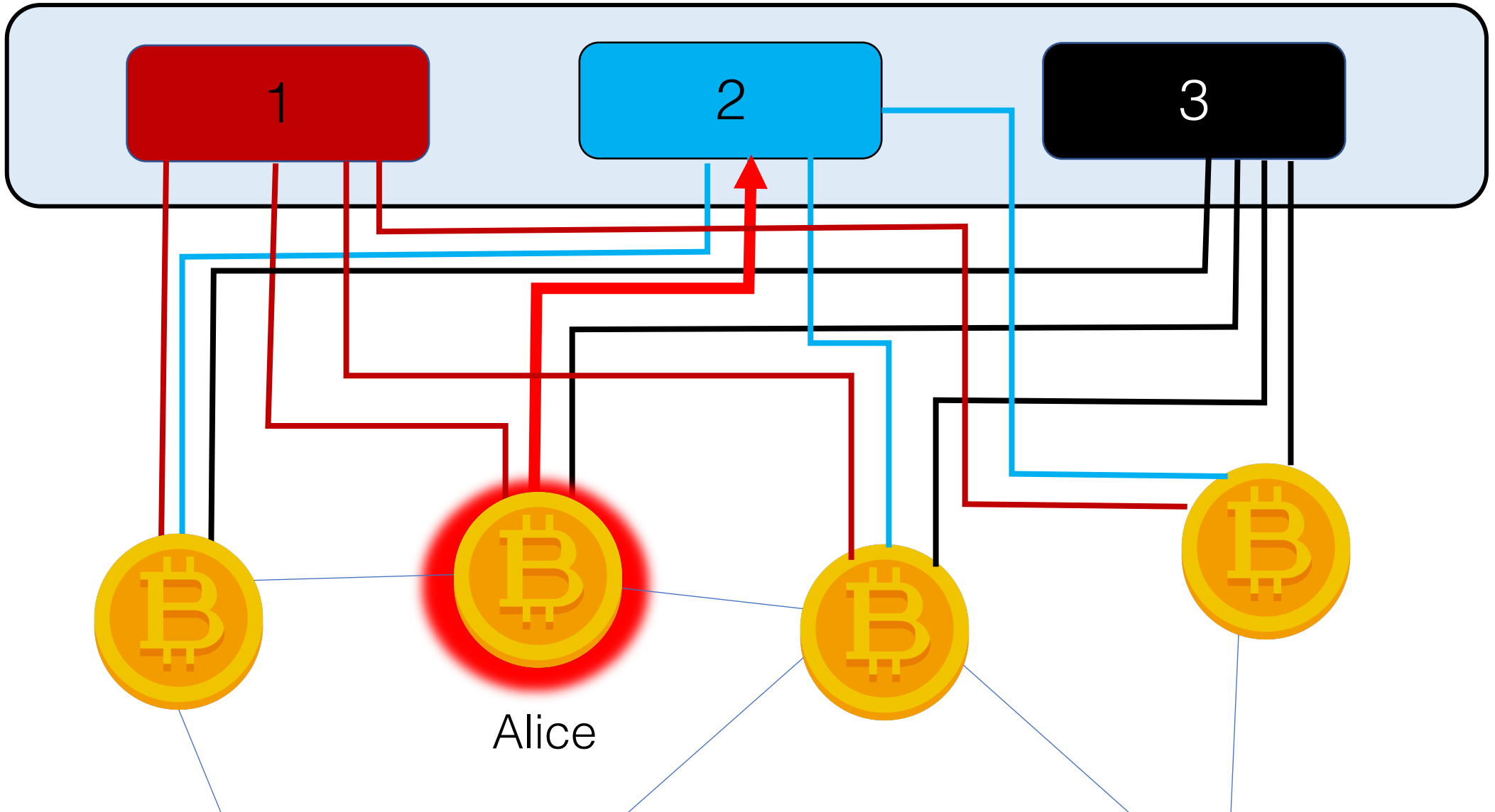
Attacks on the Network Layer



What can go wrong?



What the eavesdropper can do about it



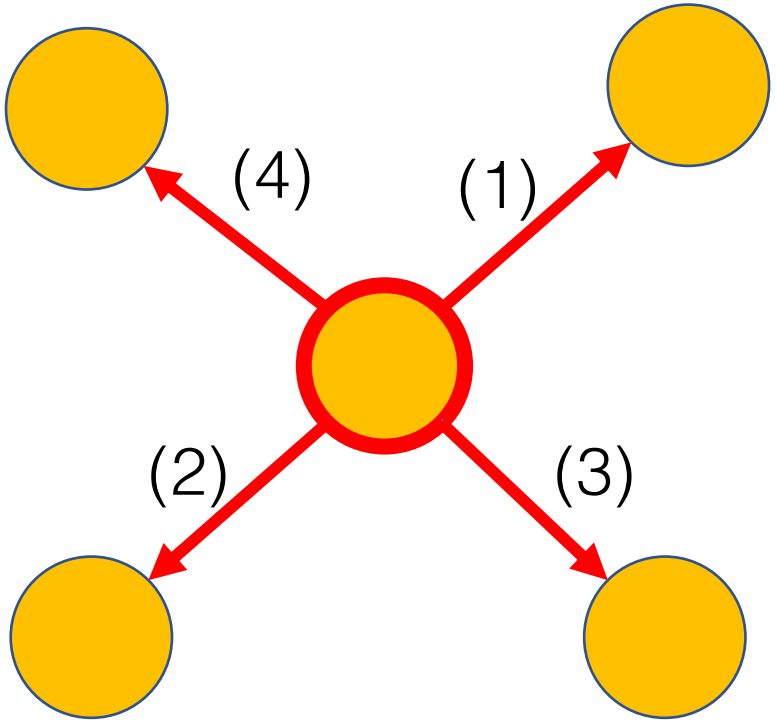
Key Results

- Make $\theta \approx 50$ connections per node

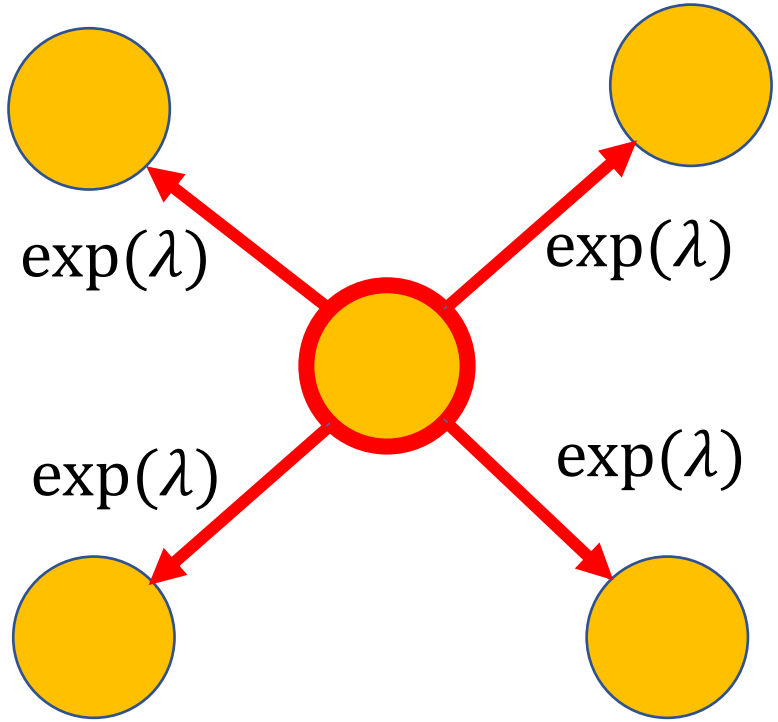
- Between **11-34% of users deanonymized,** *even behind NAT!*

Bitcoin Core Responds

Trickle (pre-2015)



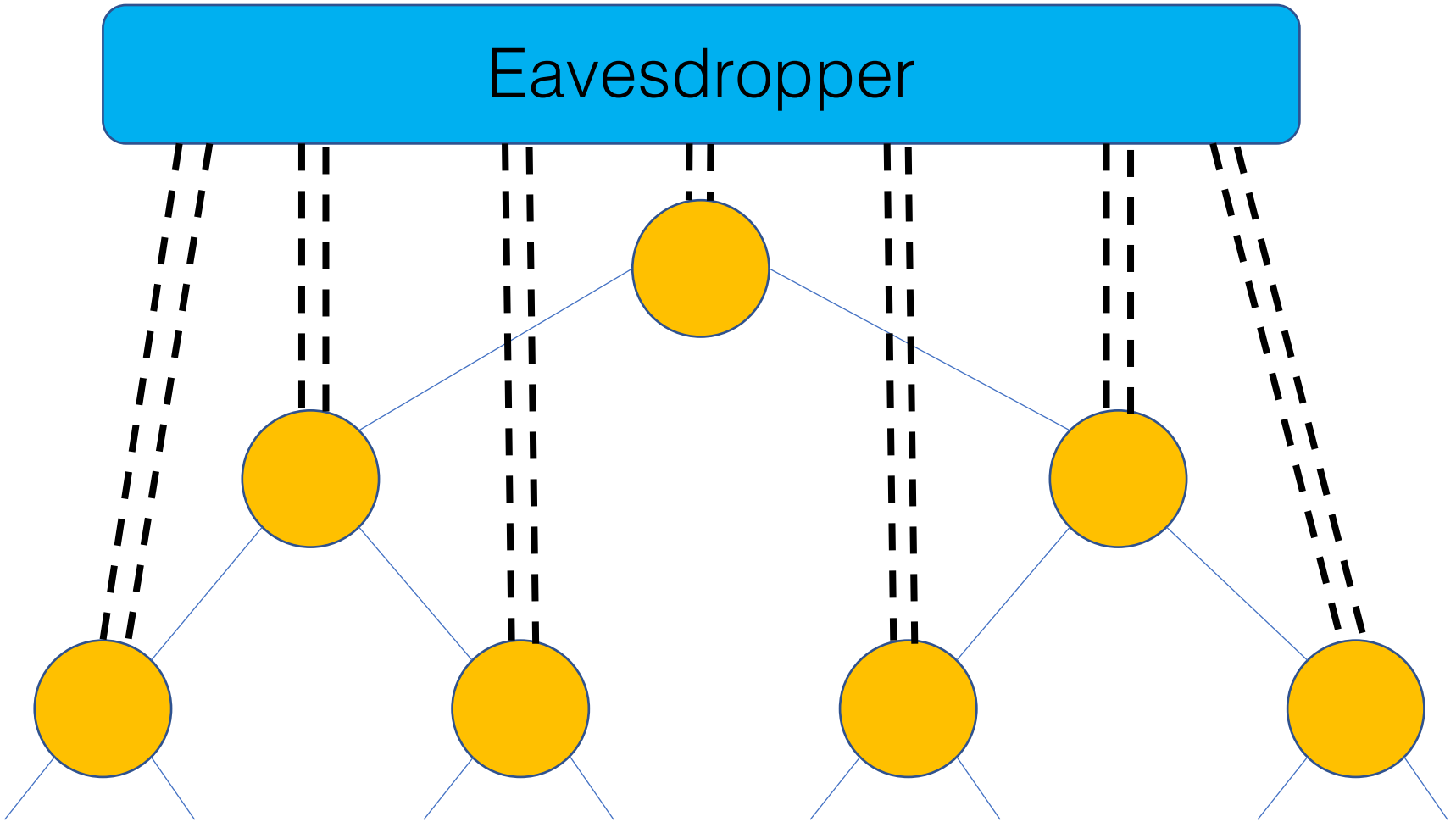
Diffusion (post-2015)



Does diffusion provide stronger anonymity than trickle spreading?

G. F., P. Viswanath, *“Anonymity in the Bitcoin P2P Network”*, NeurIPS 2017

d-regular trees



Arbitrary number of connections θ

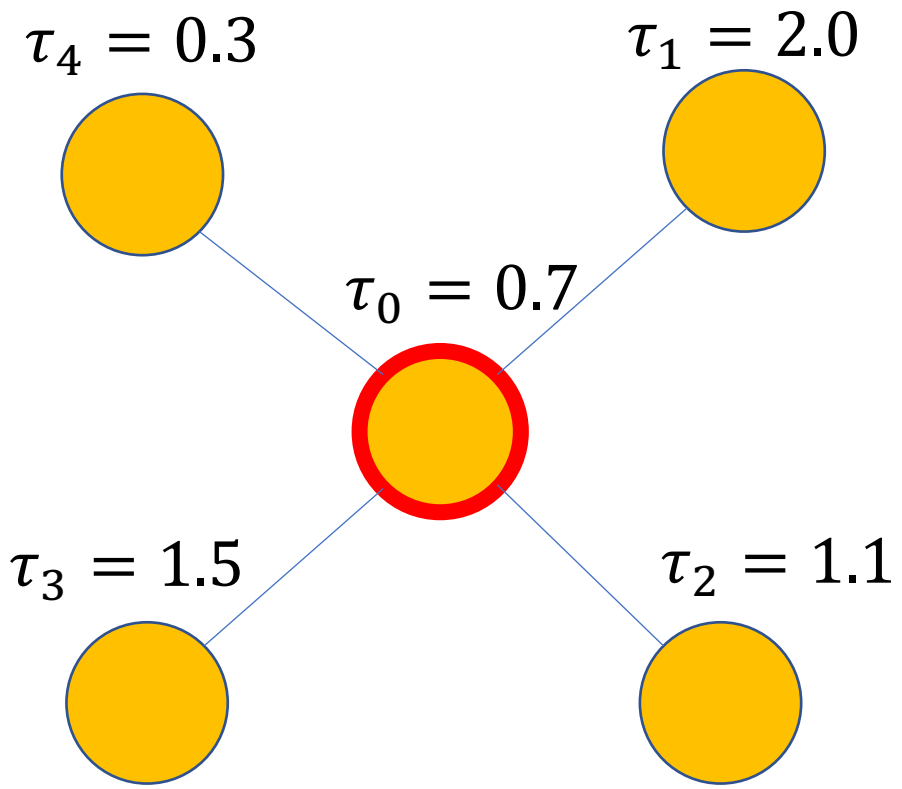
Anonymity Metric

$$P(\text{detection}|\boldsymbol{\tau}, G)$$

timestamps

graph

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \dots \\ \tau_n \end{bmatrix}$$

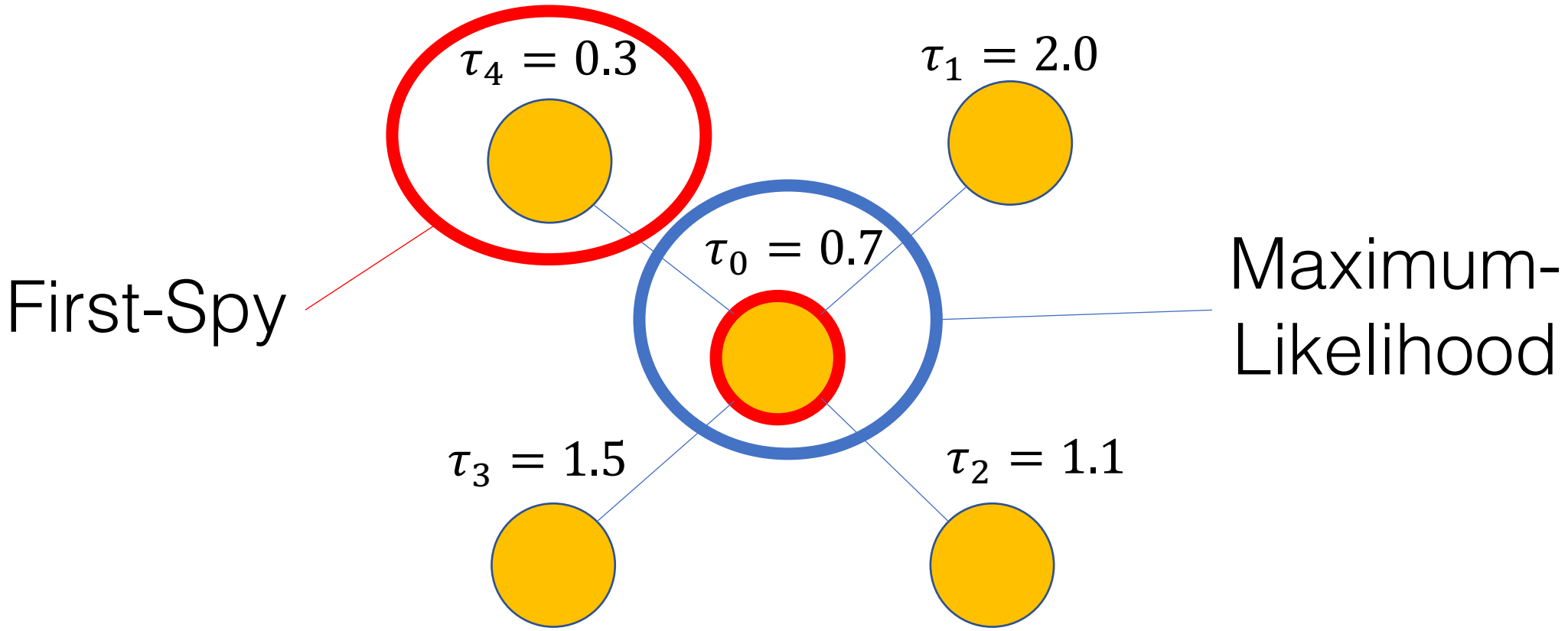


Estimators

$$P(\text{detection}|\boldsymbol{\tau}, G)$$

timestamps

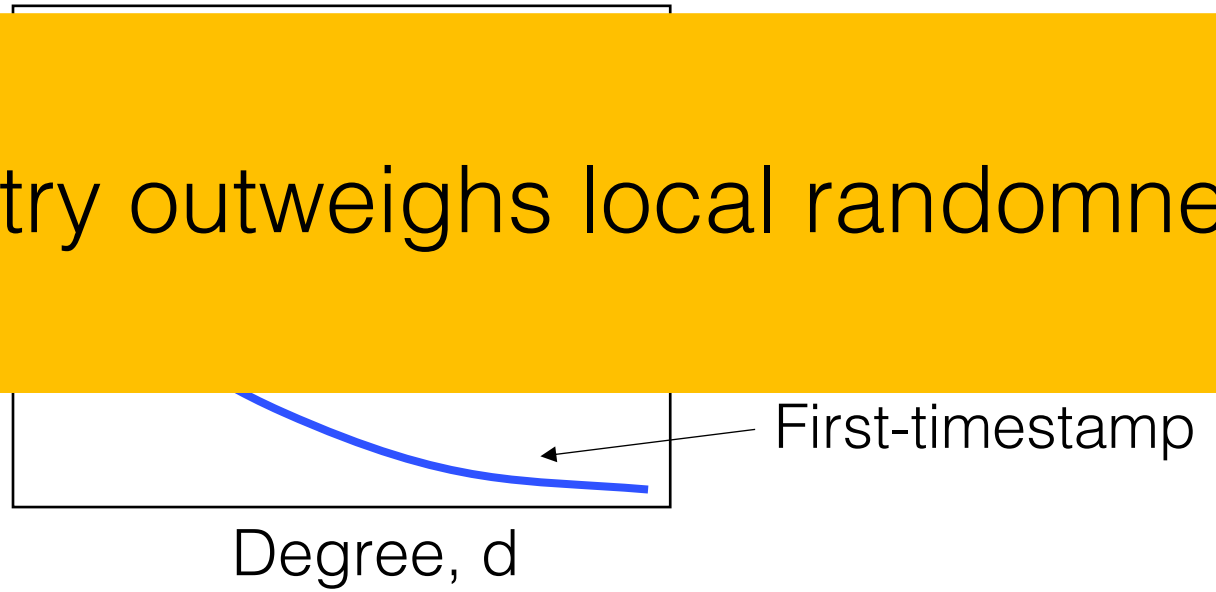
graph



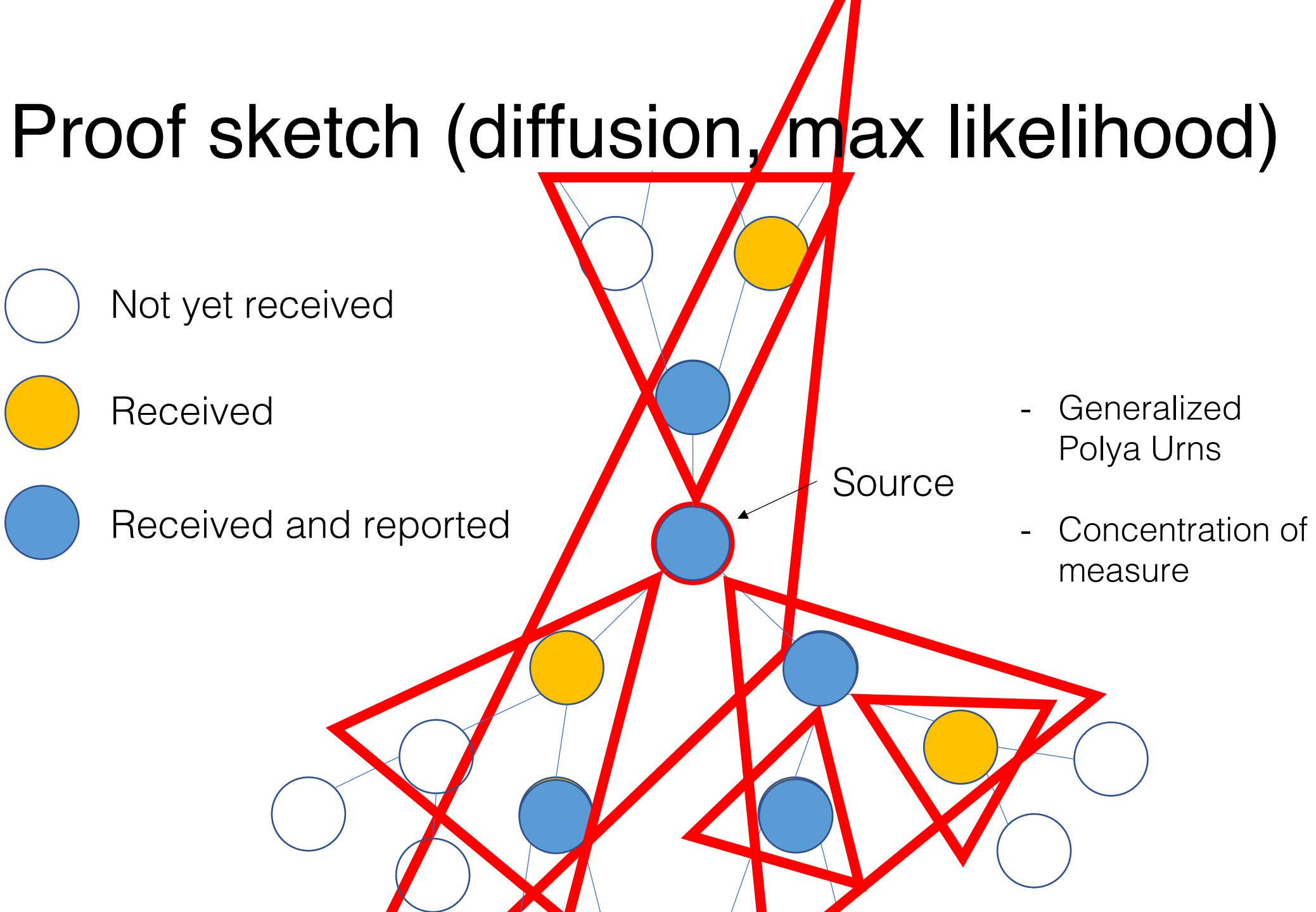
Results: d-Regular Trees

	Trickle	Diffusion
First-Timestamp	$o\left(\frac{\log d}{d}\right)$	$o\left(\frac{\log d}{d}\right)$
Maximum-Likelihood	$\Omega(1)$	$\Omega(1)$

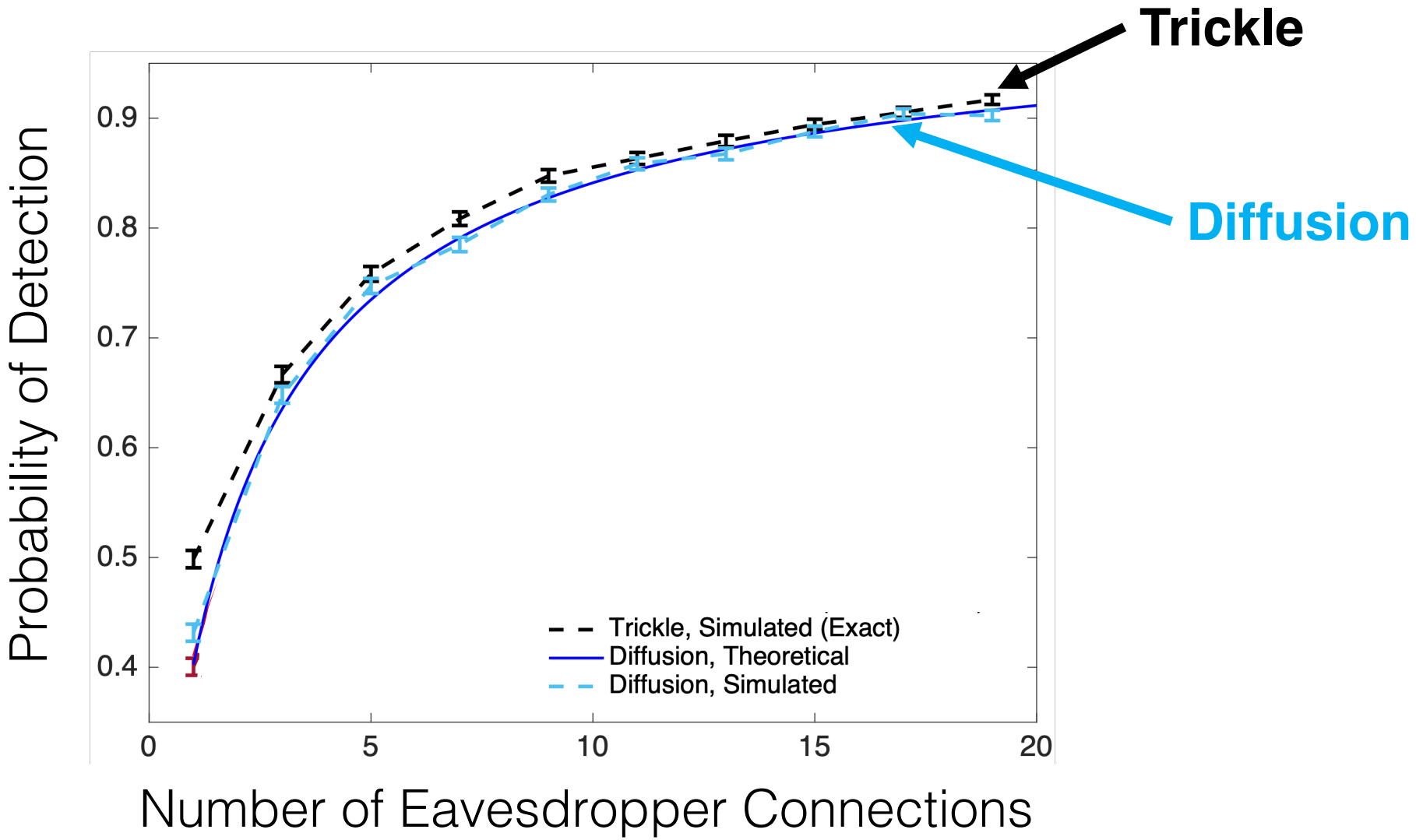
Intuition: Symmetry outweighs local randomness!



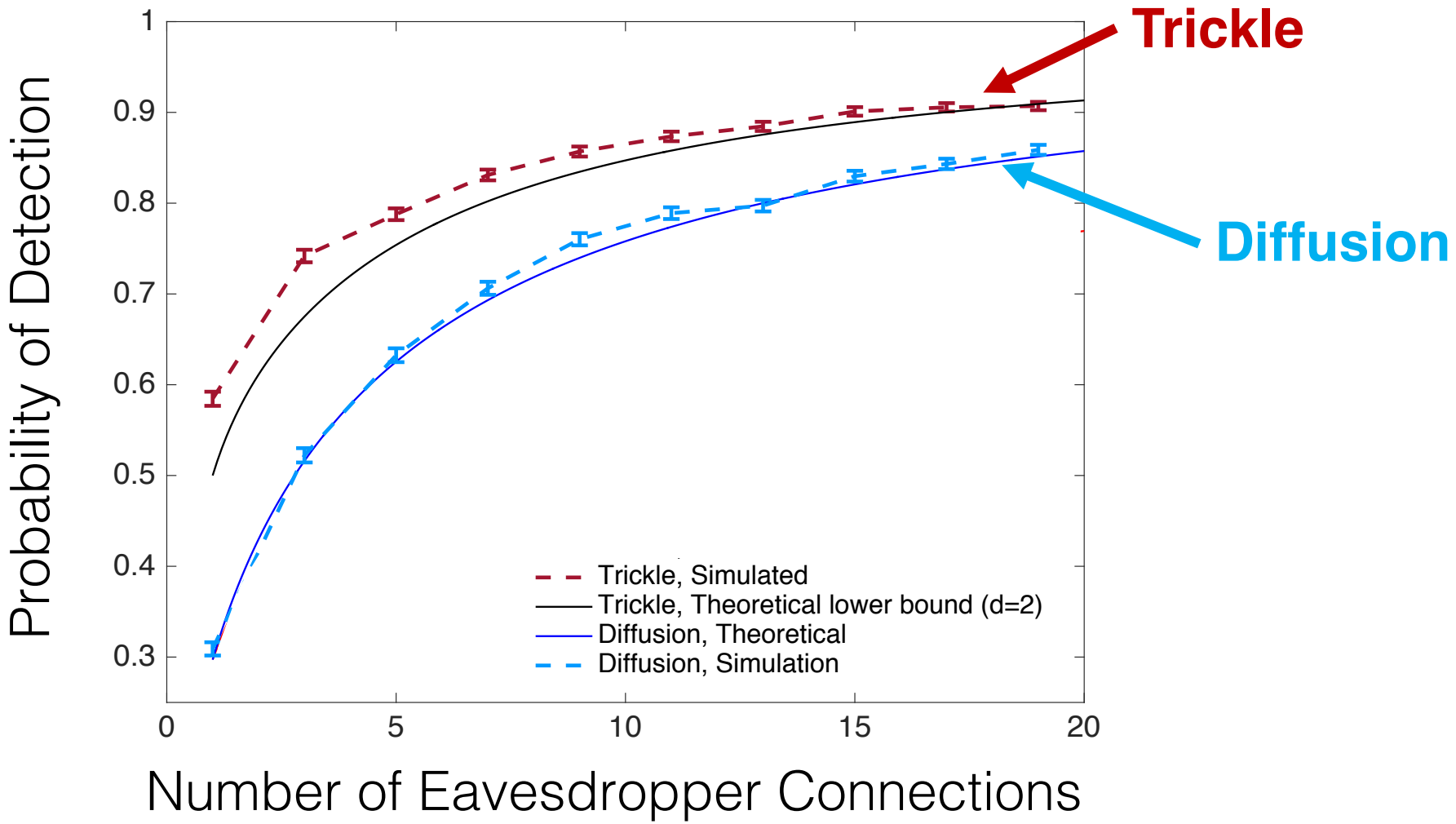
Proof sketch (diffusion, max likelihood)



Results: Trees



Results: Bitcoin Graph



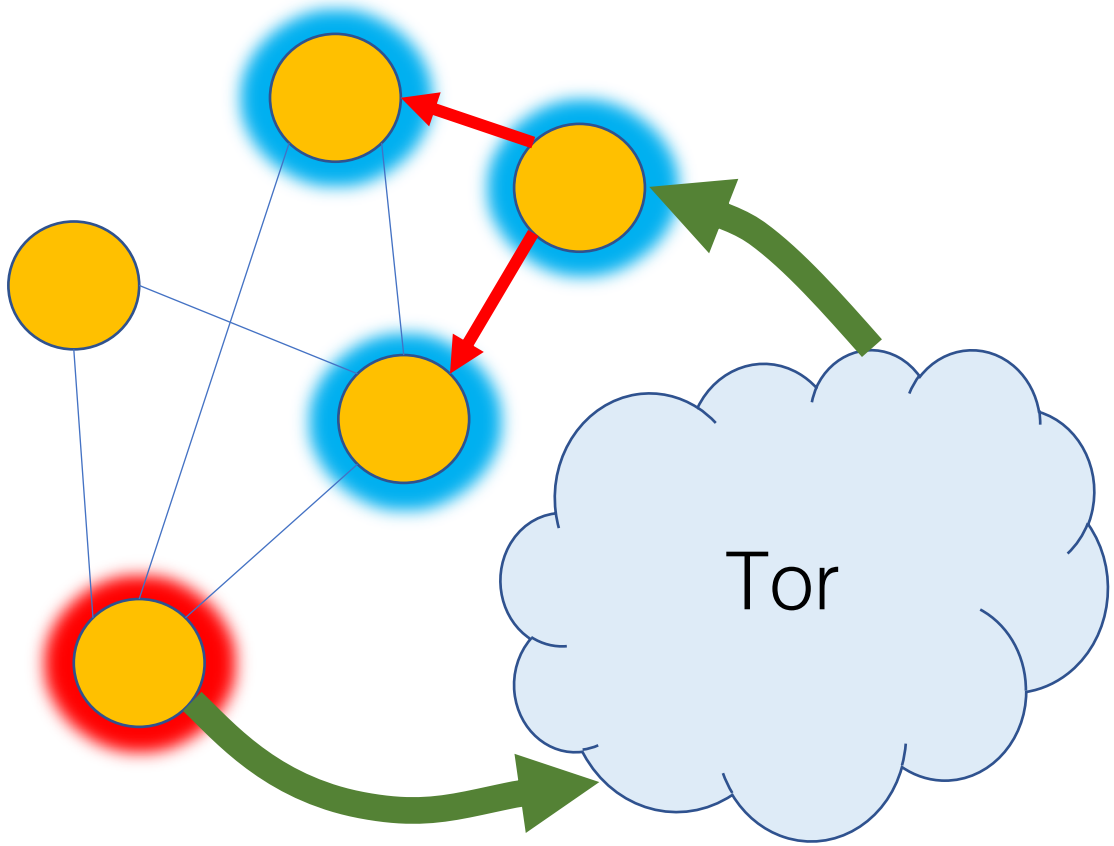
Diffusion does not have (significantly) better anonymity properties than trickle.

Redesign

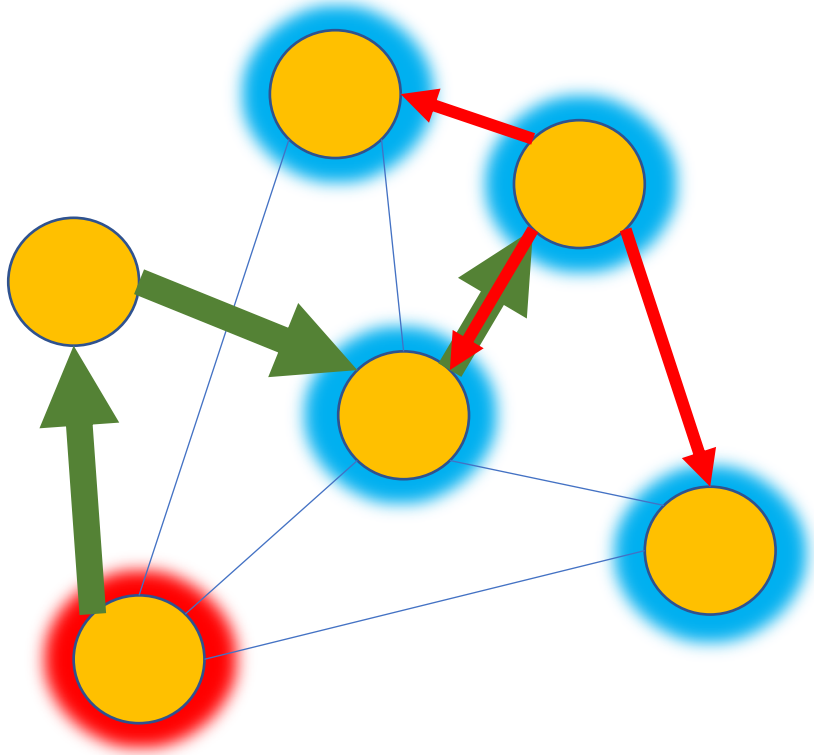
Can we fix this problem?

First-order solutions

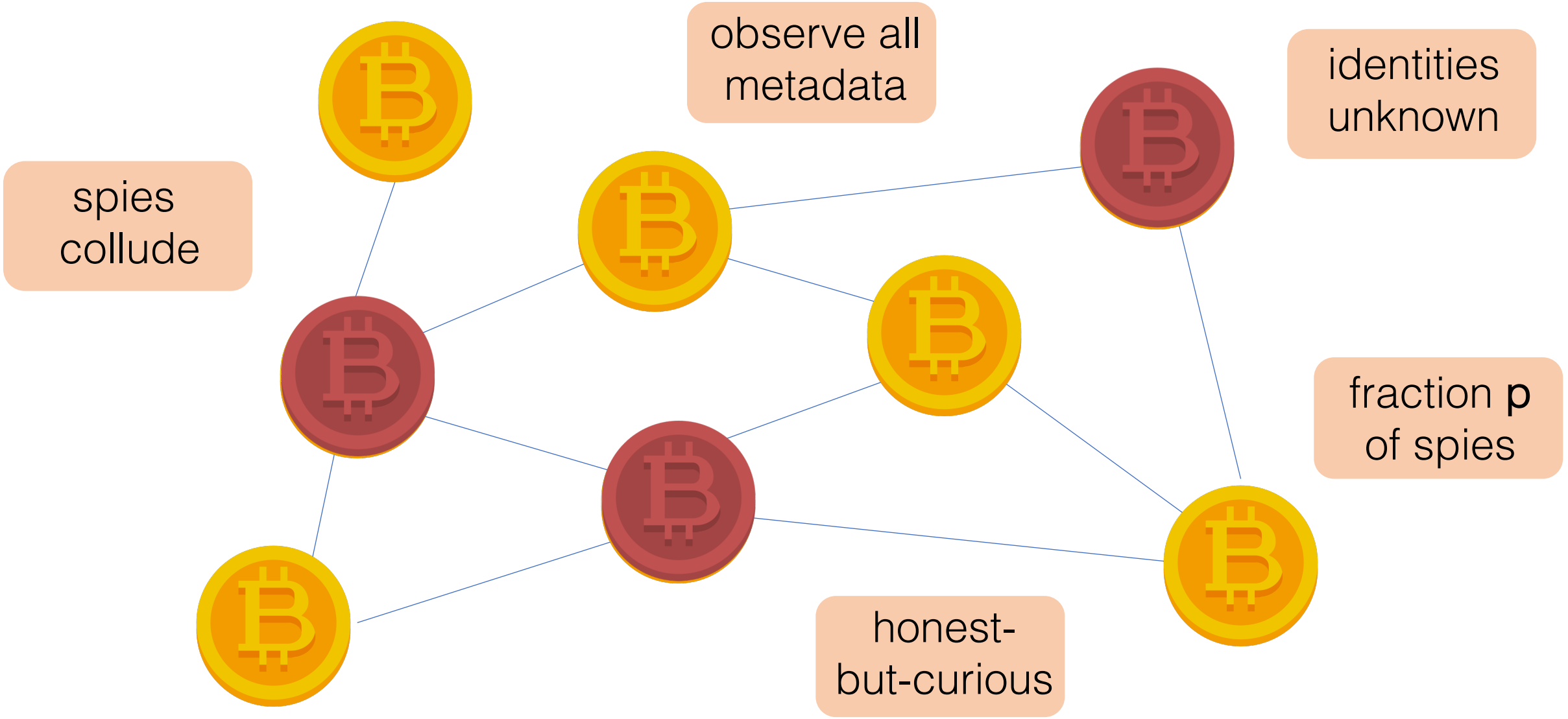
Connect through Tor



I2P Integration (e.g. Monero)



Botnet adversarial model



Metric for Anonymity

Recall

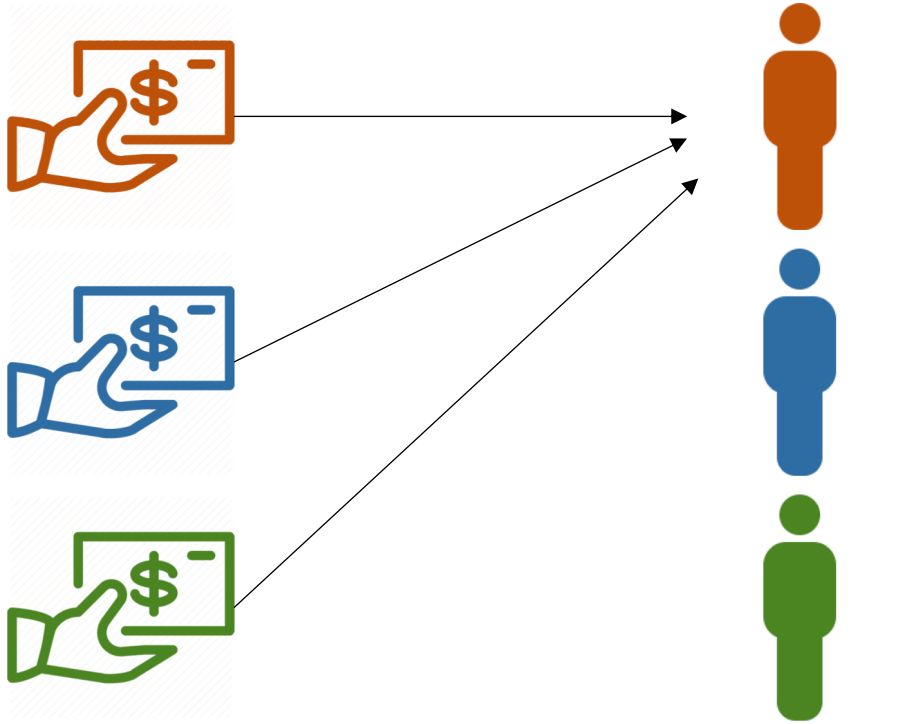
$$\frac{1}{n}\sum_{v} 1\{M(v\text{'s tx}) = v\}$$

Number honest users

User

Mapping

$\mathbb{E}[\text{Recall}]$ = Probability of Detection



Precision

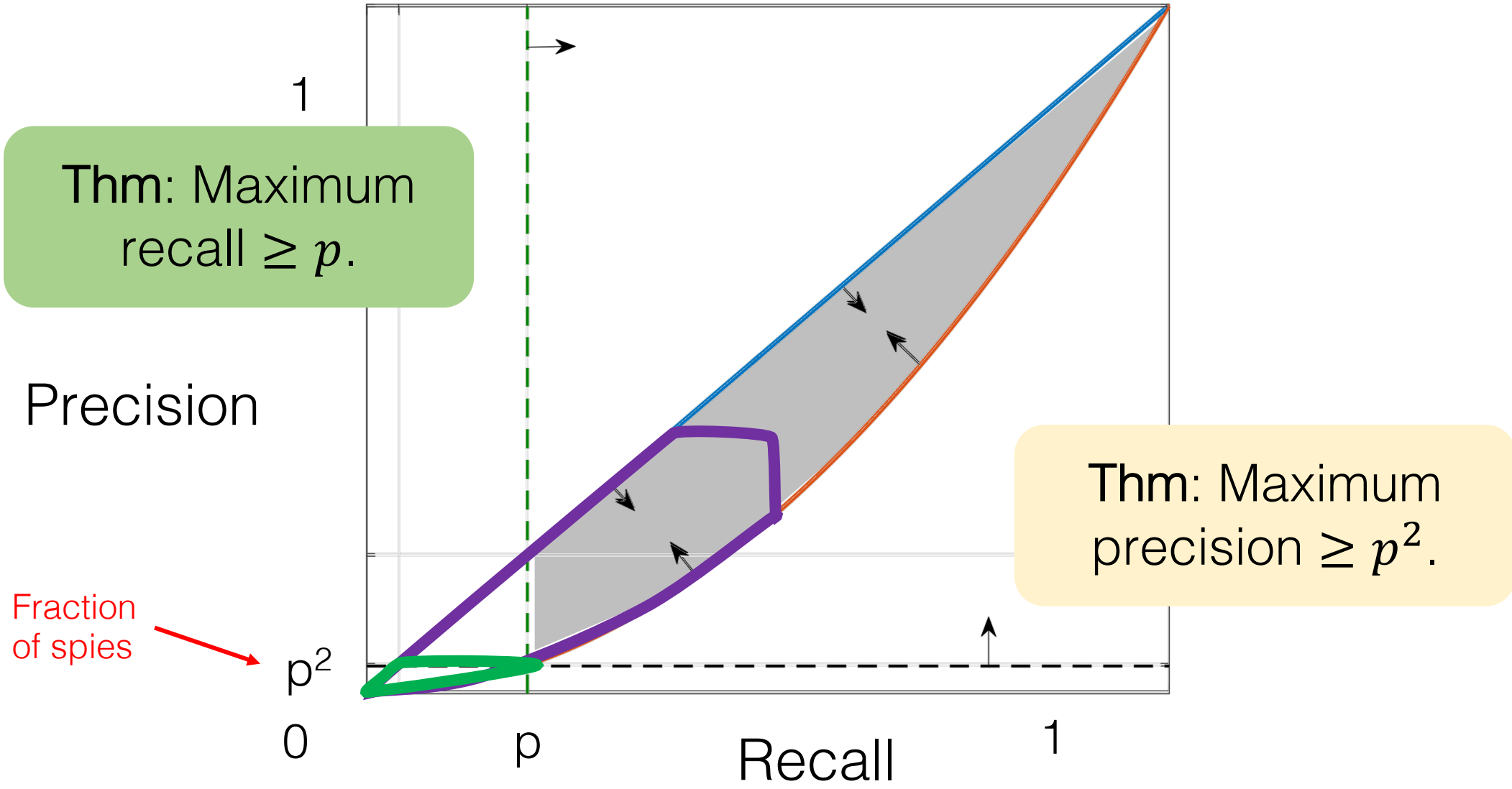
$$\frac{1}{n}\sum_{v} \frac{1\{M(v\text{'s tx}) = v\}}{\#\text{ tx mapped to v}}$$

Goal:

Design a distributed flooding protocol that minimizes the maximum precision and recall achievable by a computationally-unbounded adversary.

S. B. Venkatakrishnan, G. F., P. Viswanath, *"Dandelion: Redesigning the Bitcoin Network for Anonymity "*, Sigmetrics 2017

Fundamental Limits

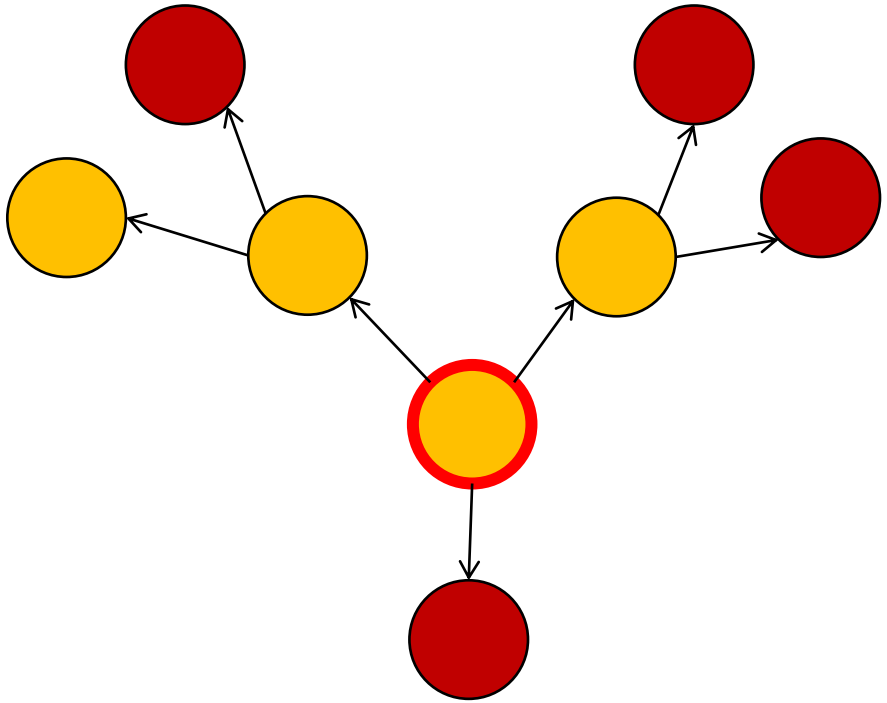


Thm: Maximum recall $\geq p$.

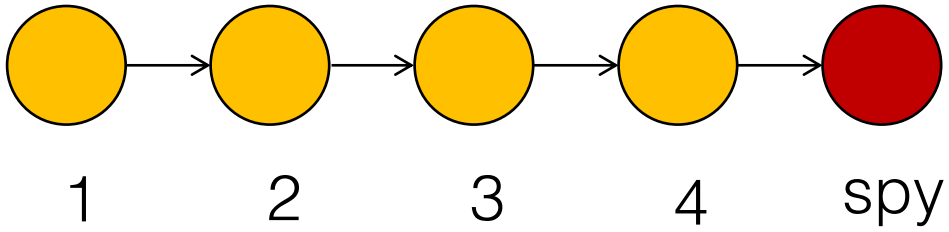
Thm: Maximum precision $\geq p^2$.

What are we looking for?

Asymmetry



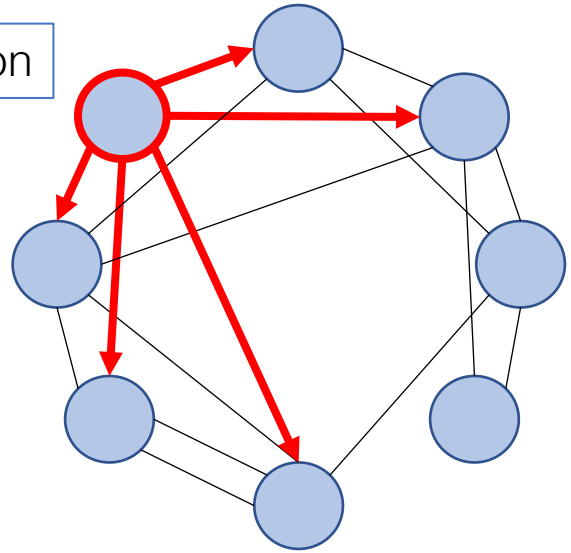
Mixing



What can we control?

Spreading Protocol

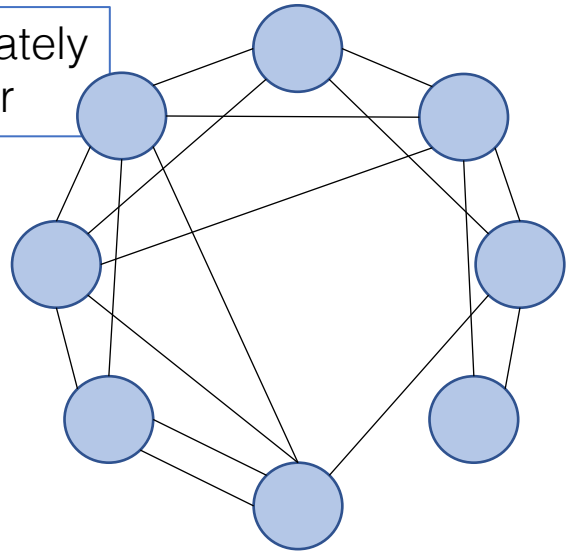
Diffusion



Given a graph, how do we spread content?

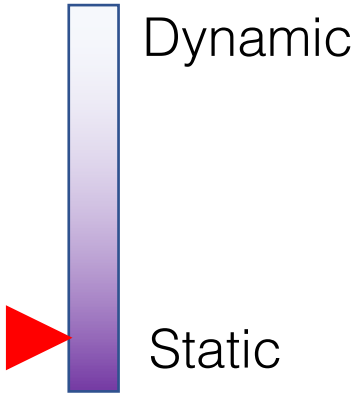
Topology

Approximately regular



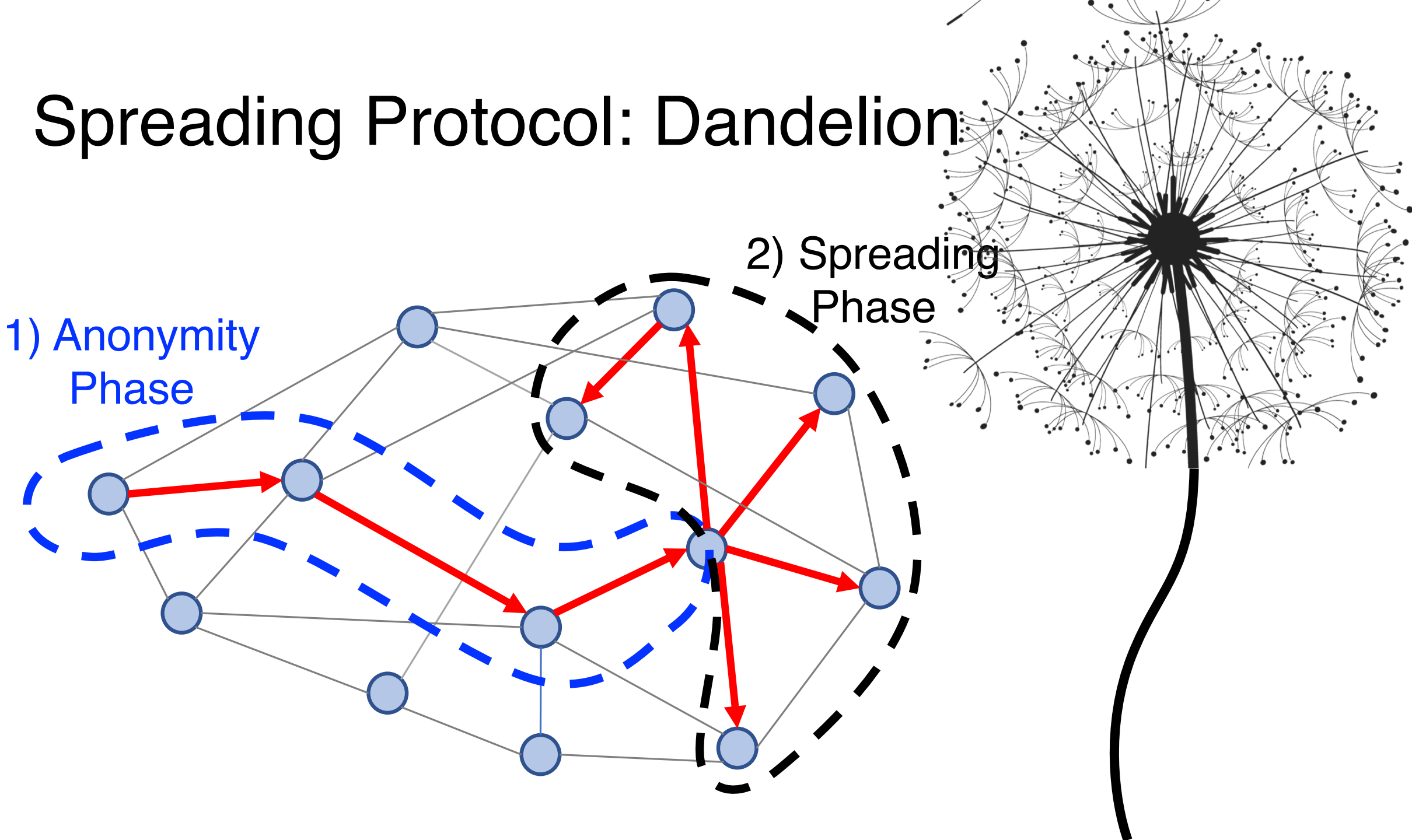
What is the underlying graph topology?

Dynamicity



How often does the graph change?

Spreading Protocol: Dandelion



Why Dandelion spreading?

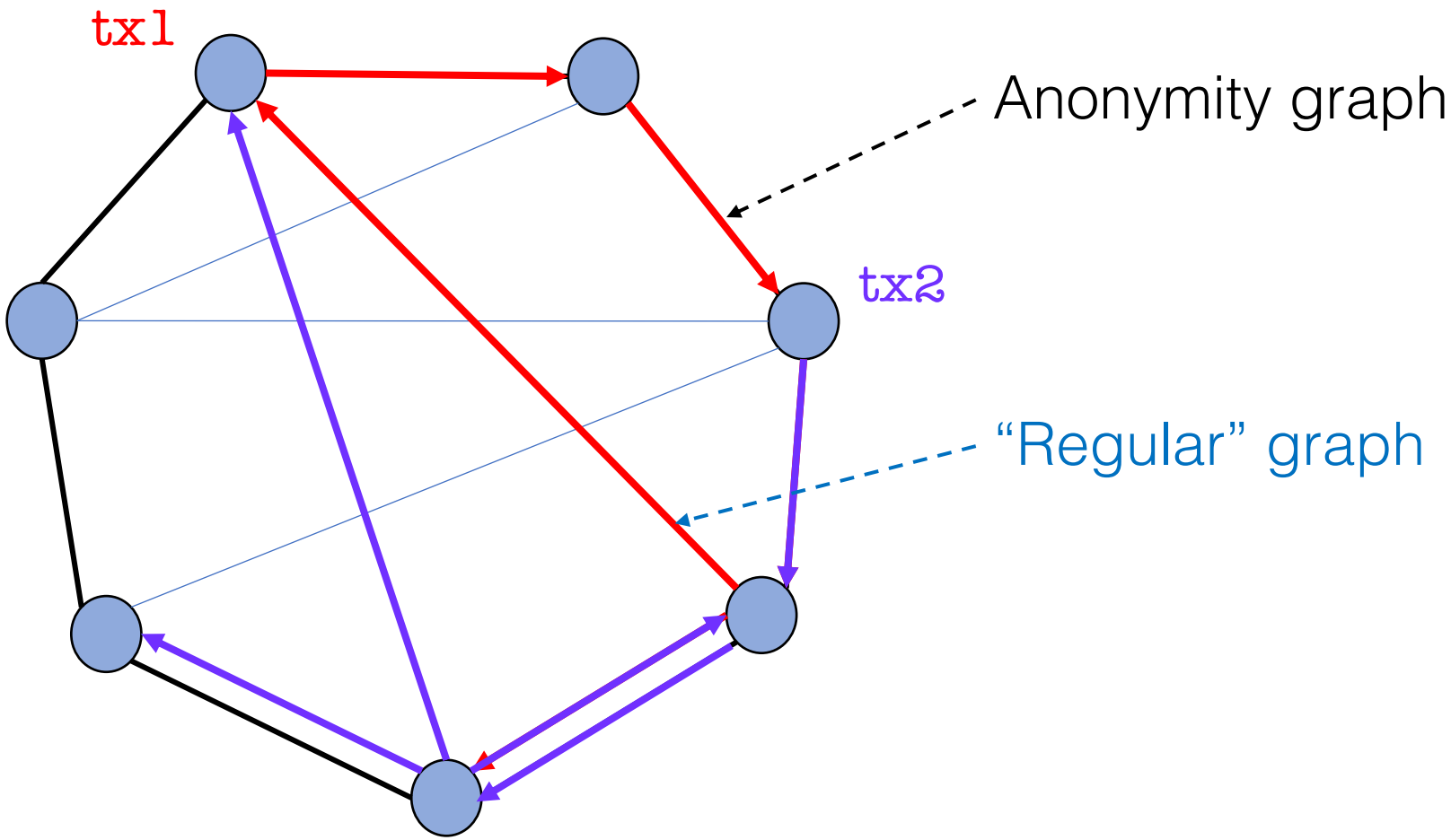
Theorem: Dandelion spreading has an optimally low maximum recall of $p + O\left(\frac{1}{n}\right)$.

Theorem: Fundamental lower bound = p

fraction of spies

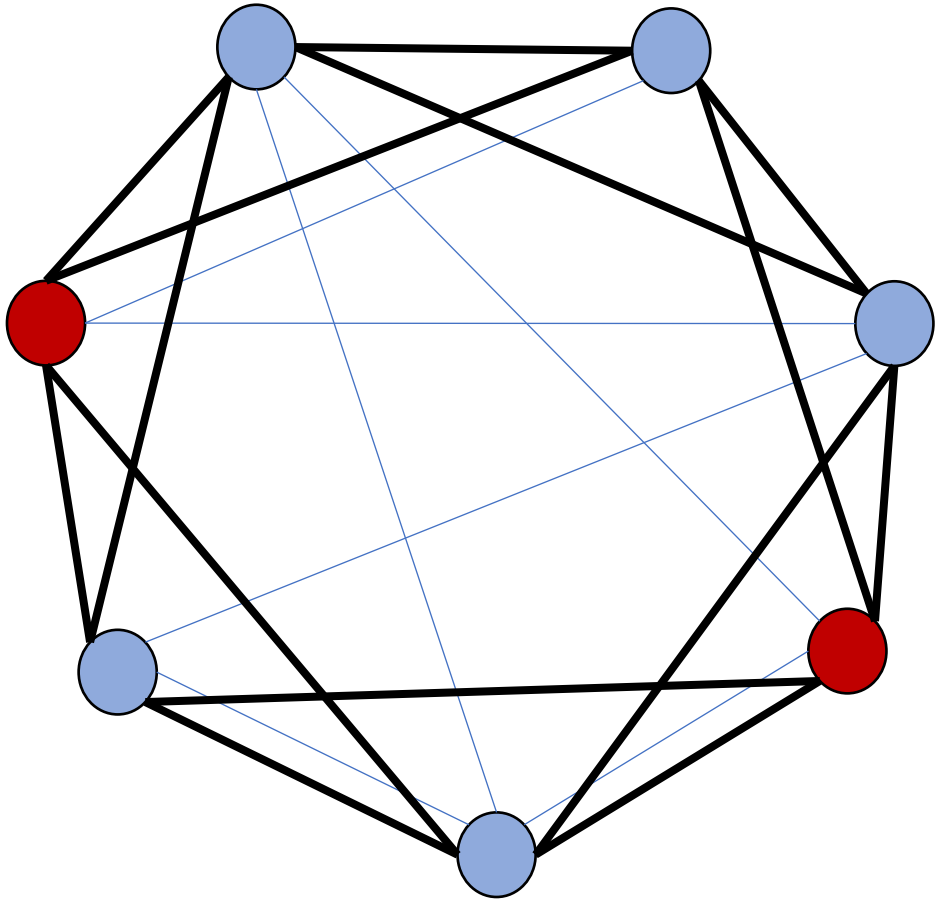
number of nodes

Graph Topology: Line



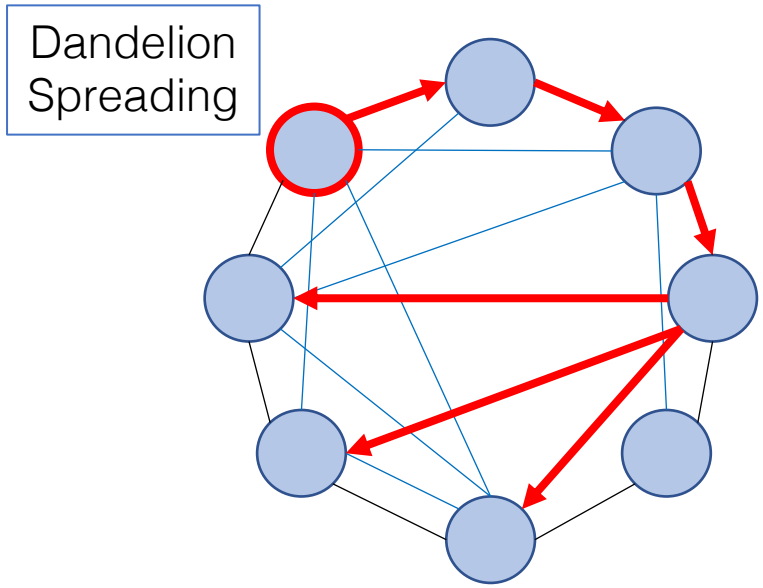
Dynamicity: High

Change the anonymity graph frequently.



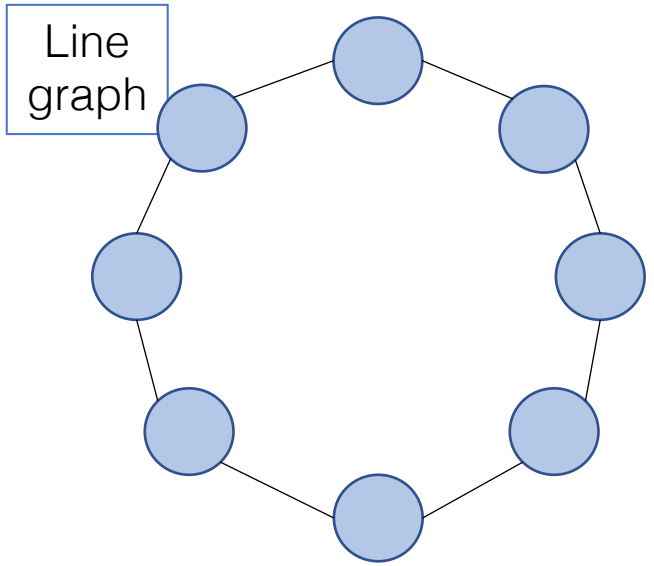
DANDELION Network Policy

Spreading Protocol



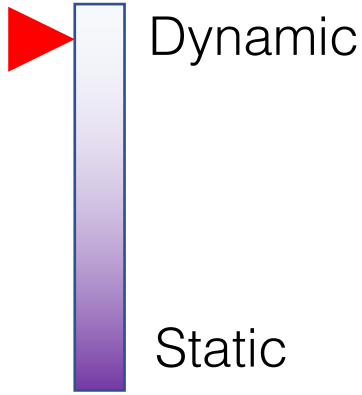
Given a graph, how do we spread content?

Topology



What is the anonymity graph topology?

Dynamicity



How often does the graph change?

Theorem: Fundamental lower bound = p^2

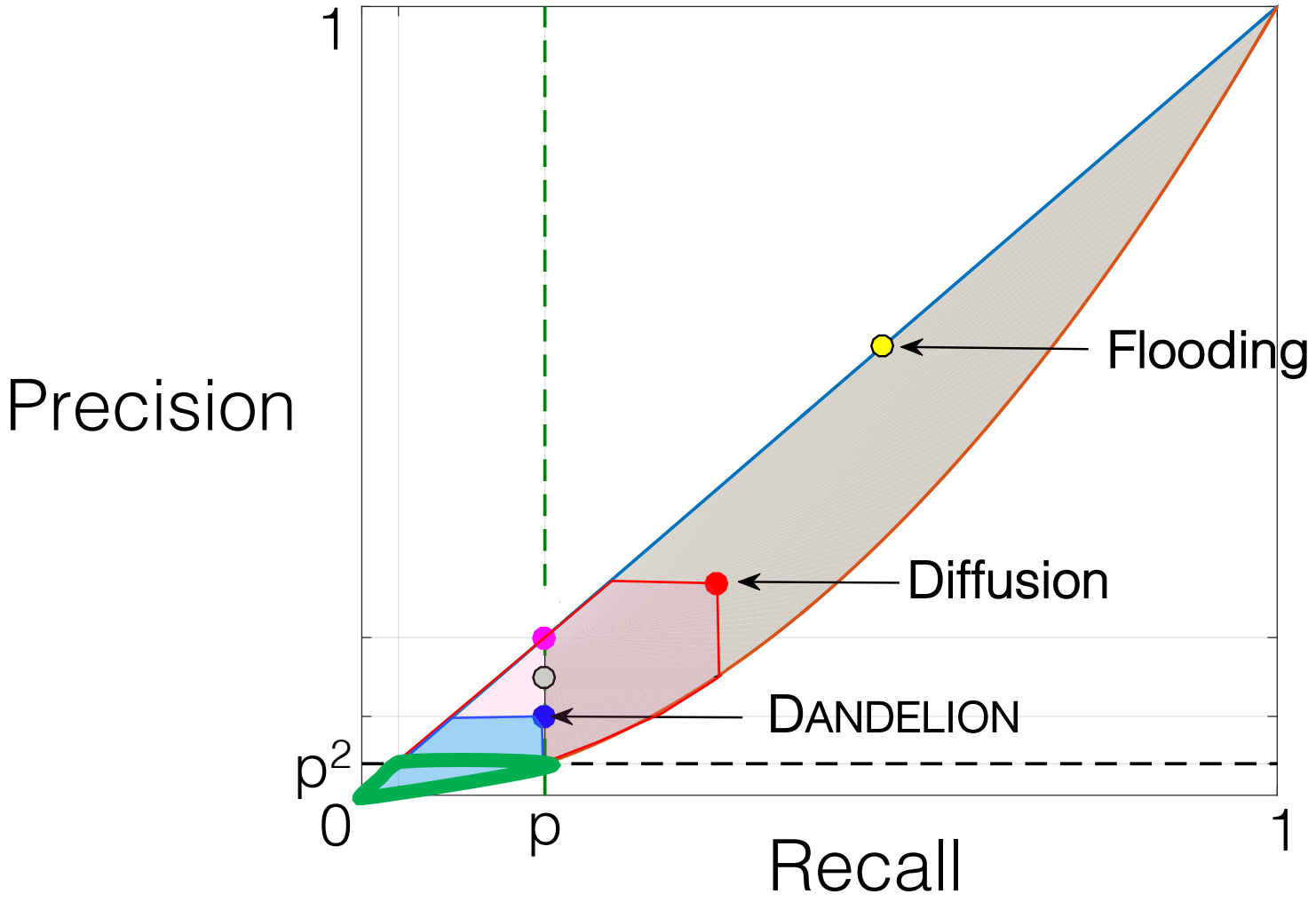
Theorem: DANDELION has a nearly-optimal maximum precision of $\frac{2p^2}{1-p}\log\left(\frac{2}{p}\right) + O\left(\frac{1}{n}\right)$.*

fraction of spies

number of nodes

*For $p < \frac{1}{3}$

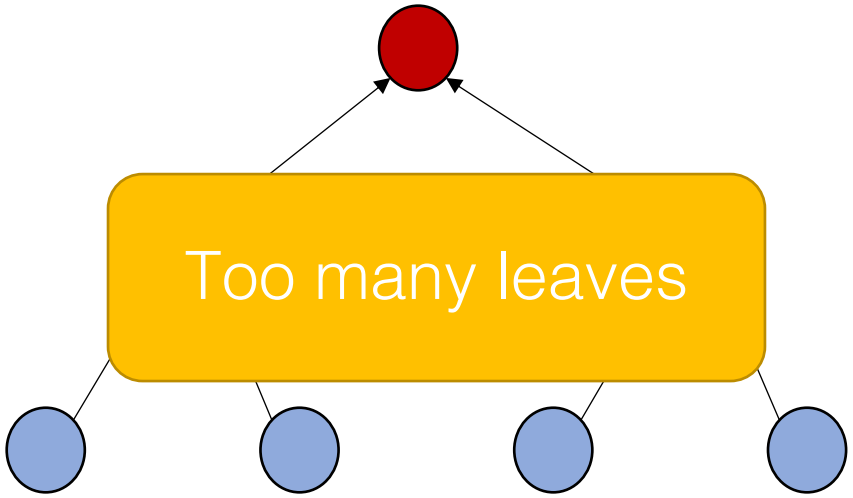
Performance: Achievable Region



Why is DANDELION good?

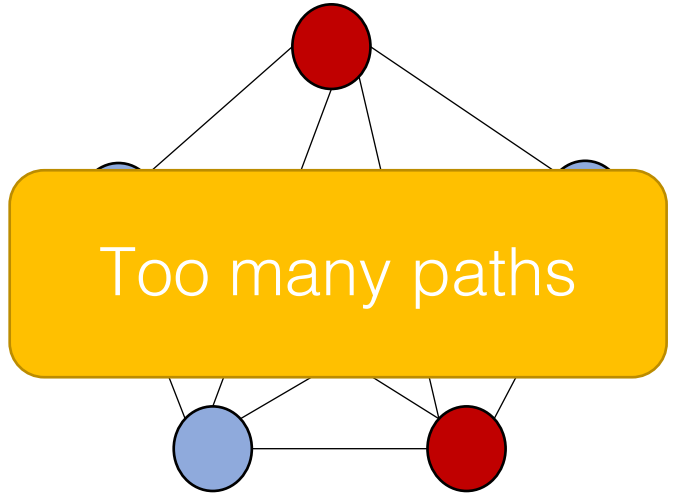
Strong mixing properties.

Tree



Precision: $O(p)$

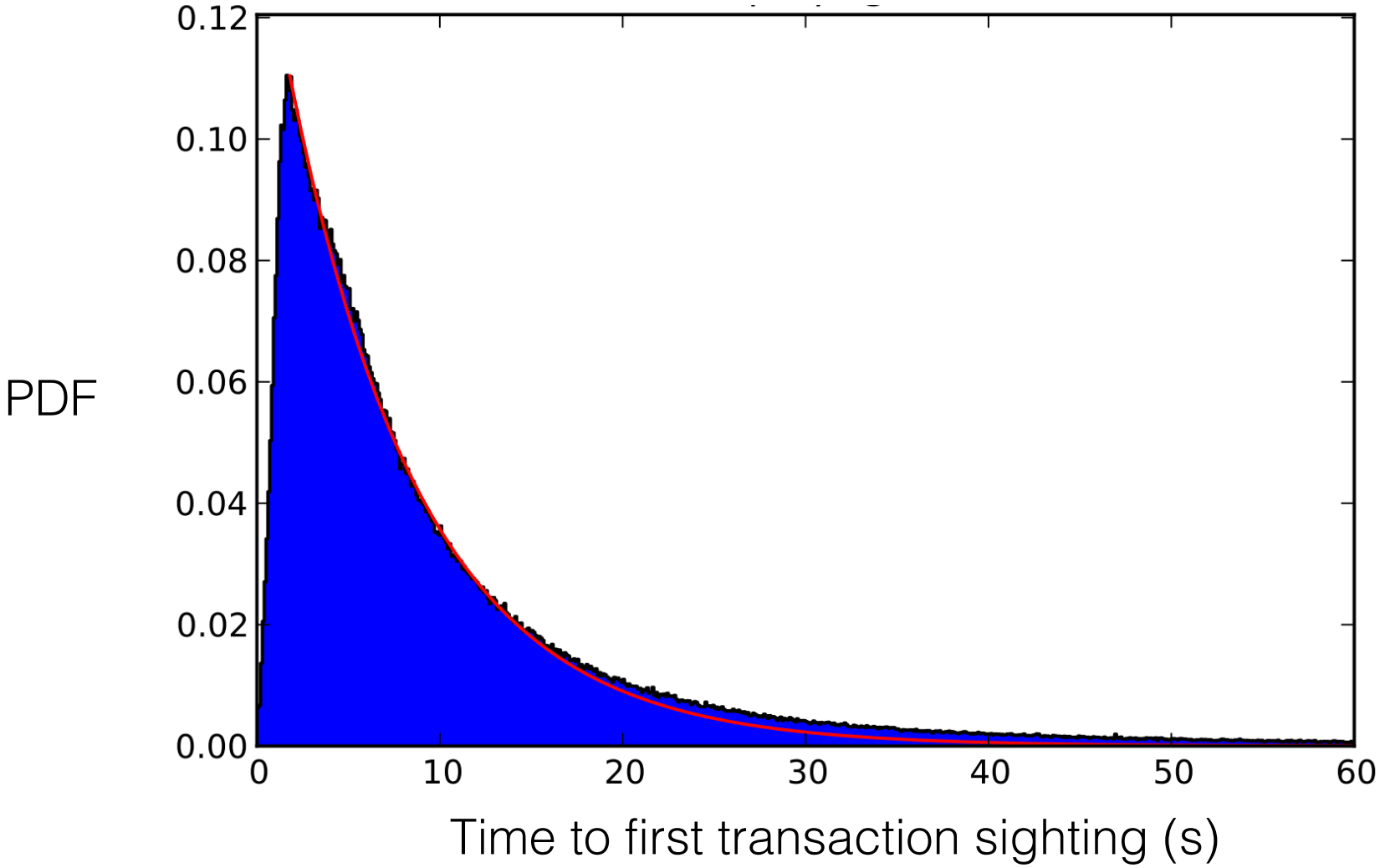
Complete graph



Precision: $\frac{p}{1-p}(1-e^{p-1})$

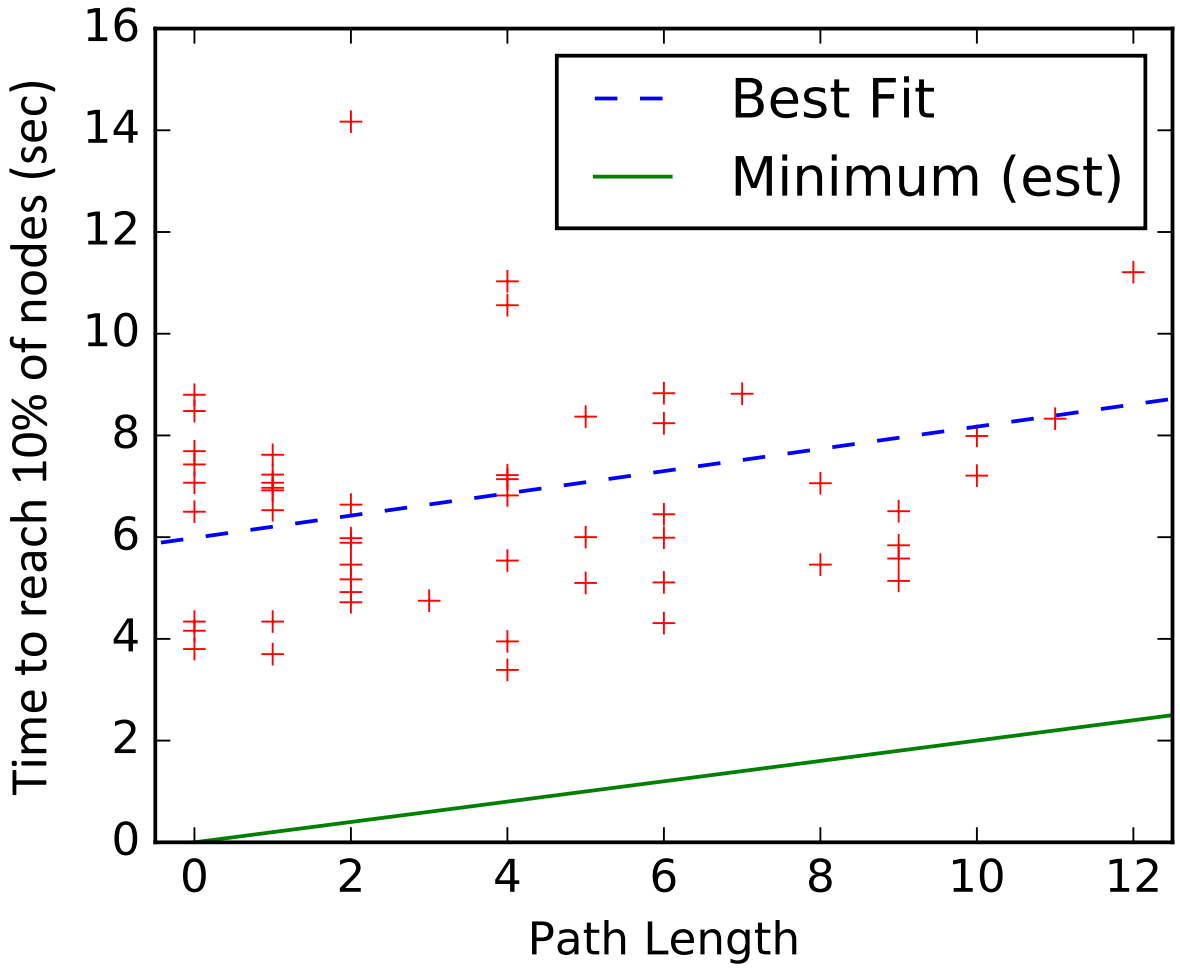
How practical is this?

Latency Overhead: Estimate

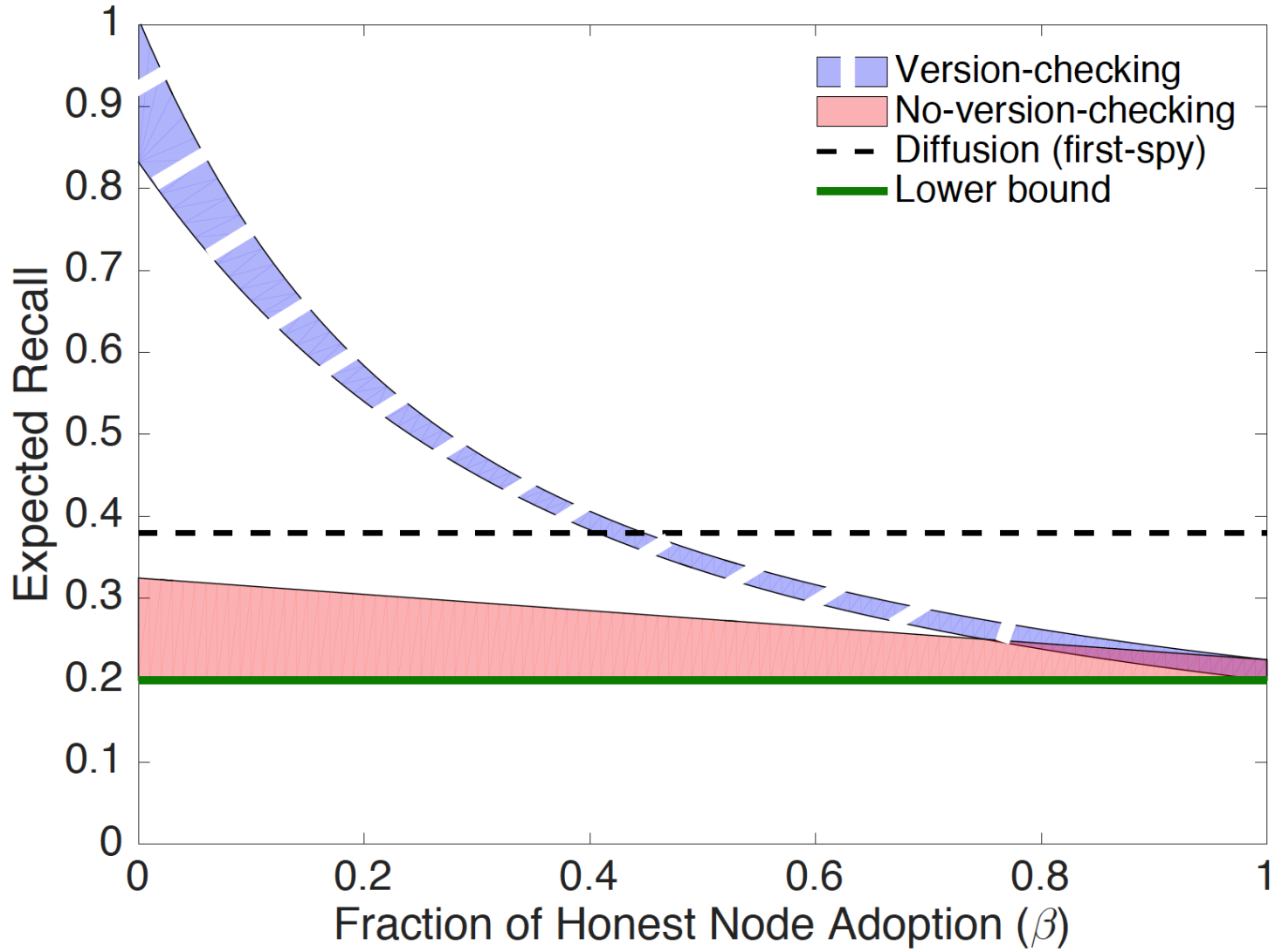


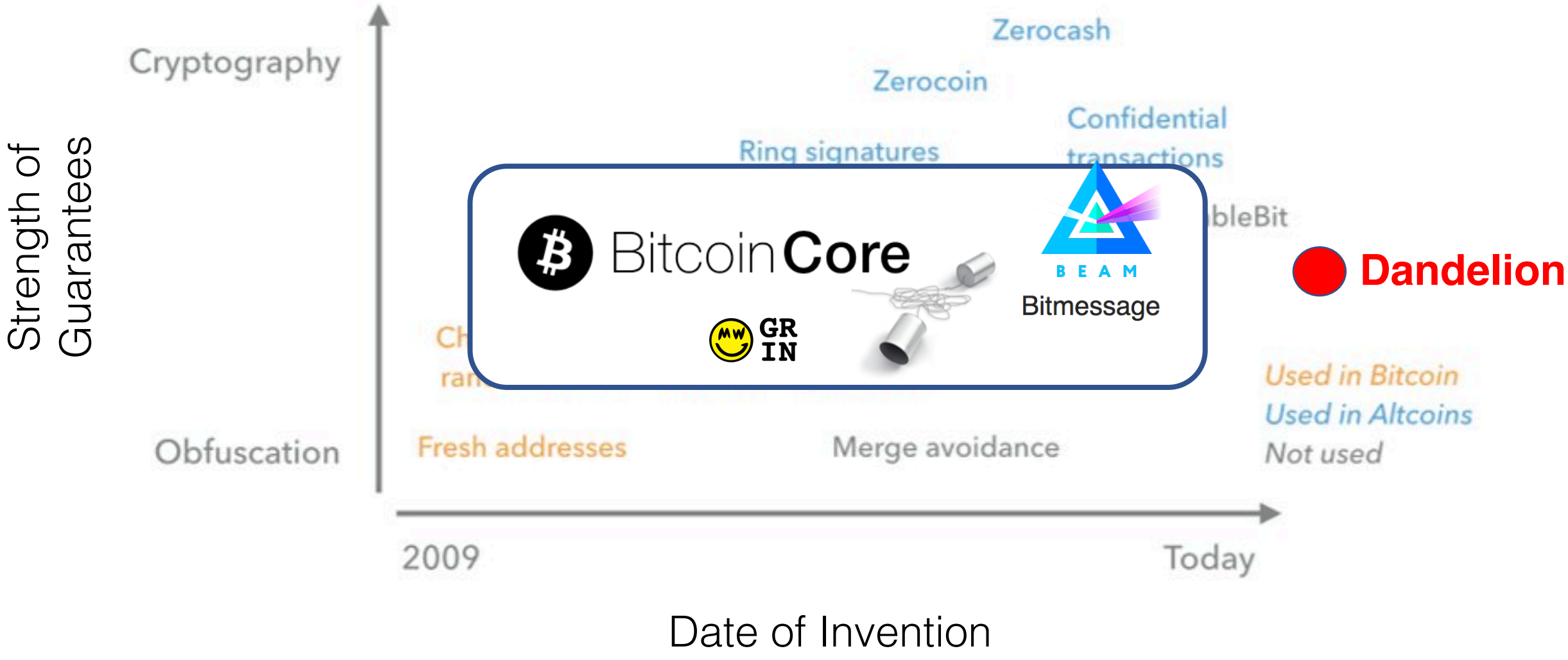
Information Propagation in the Bitcoin Network, Decker and Wattenhofer, 2013

Empirical Delay Distribution



Practical Challenges: Partial deployment





Narayanan and Möser, 2017

Take-Home Messages

1) Bitcoin's P2P network has poor anonymity.

2) Moving from trickle to diffusion did not help.

3) DANDELION may be a lightweight solution for certain classes of adversaries.

https://github.com/dandelion-org/bitcoin

BIP 156