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

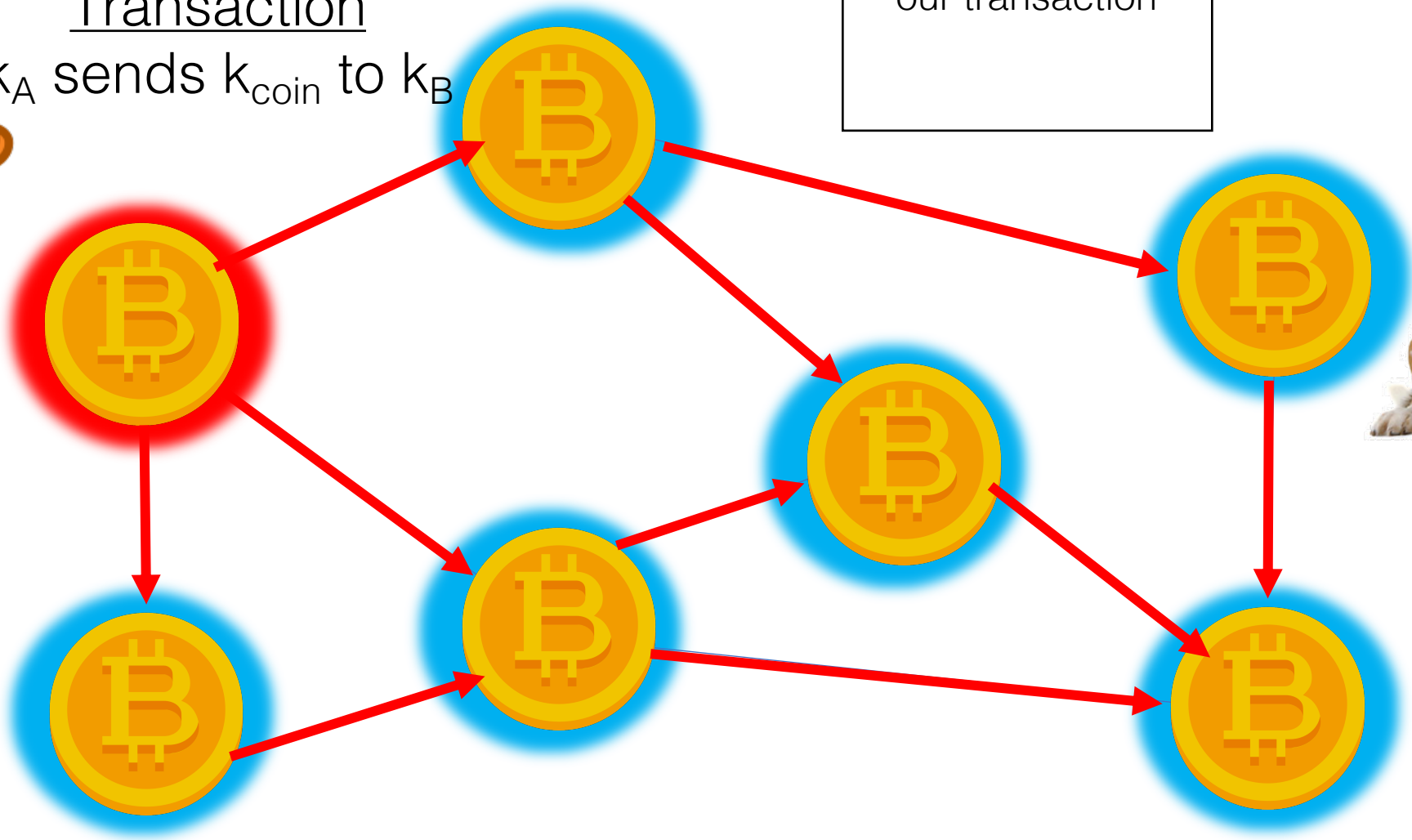
Transaction

k_A sends k_{coin} to k_B

Blockchain
sd93fjj2
pckrn29
…
our transaction

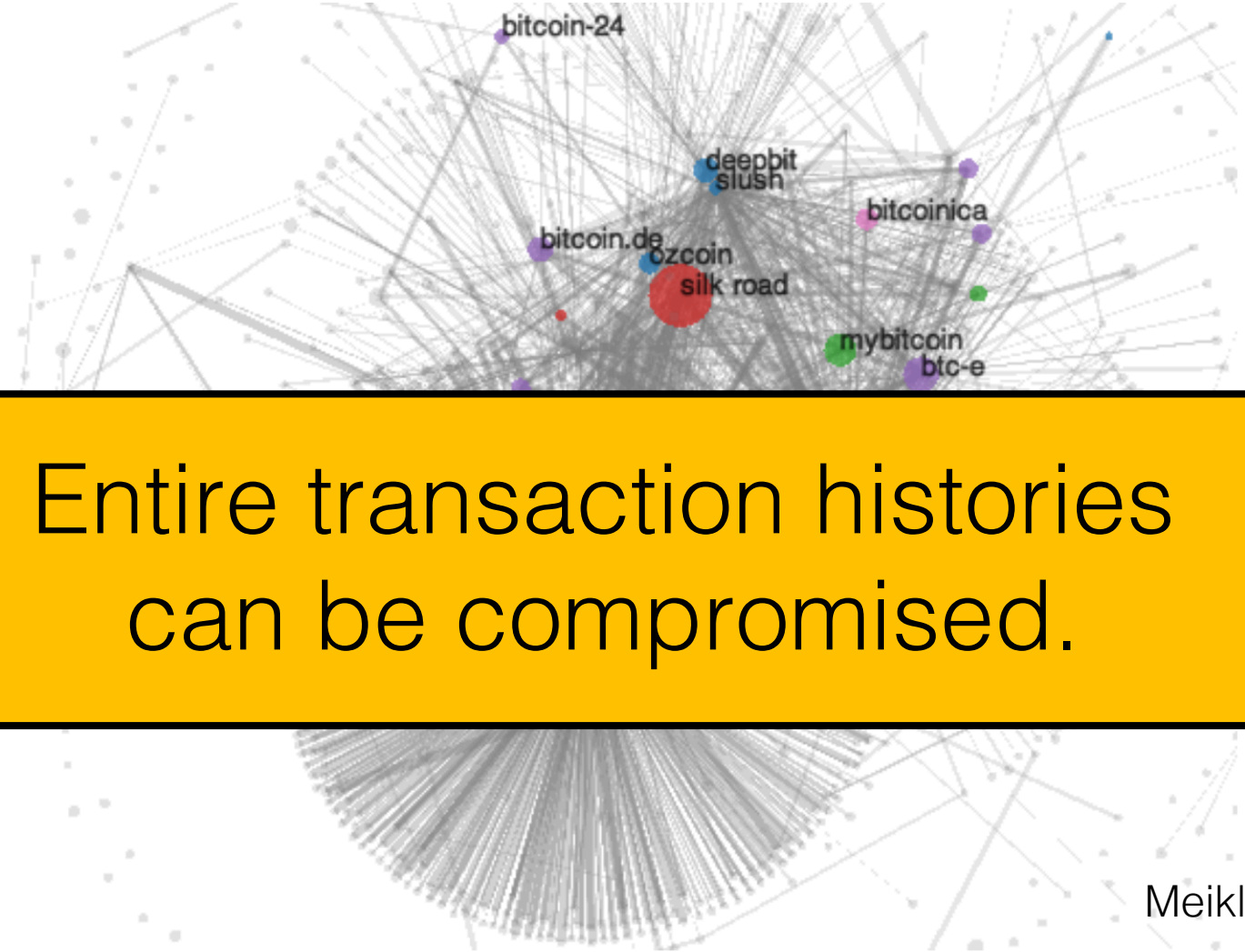
Alice
k_A

k_{coin}



Bob
k_B

How can users be deanonymized?



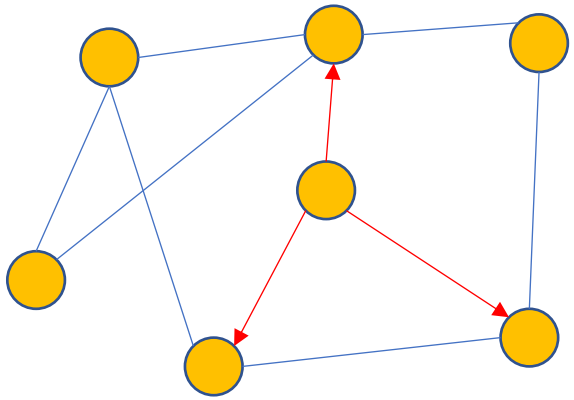
Entire transaction histories
can be compromised.

Meiklejohn et al., 2013

What about the peer-to-peer network?

Our Work

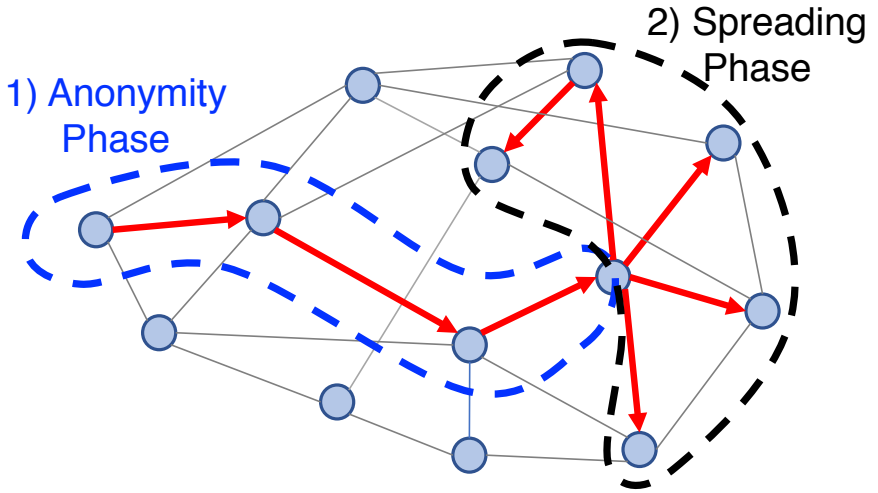
Analysis



Pr(detection)

NIPS 2017

Redesign

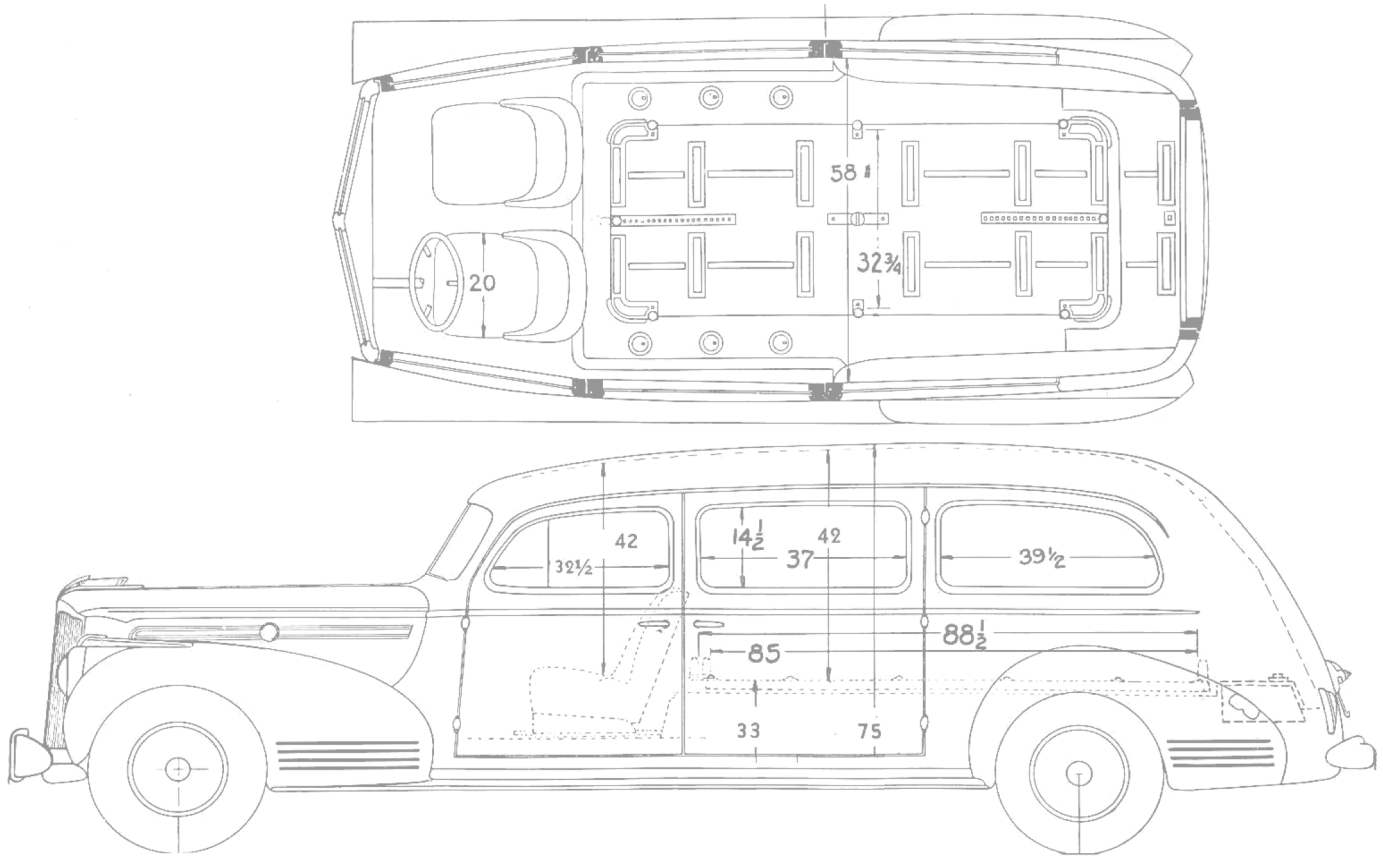


Dandelion

ACM Sigmetrics 2017,
ACM Sigmetrics 2018

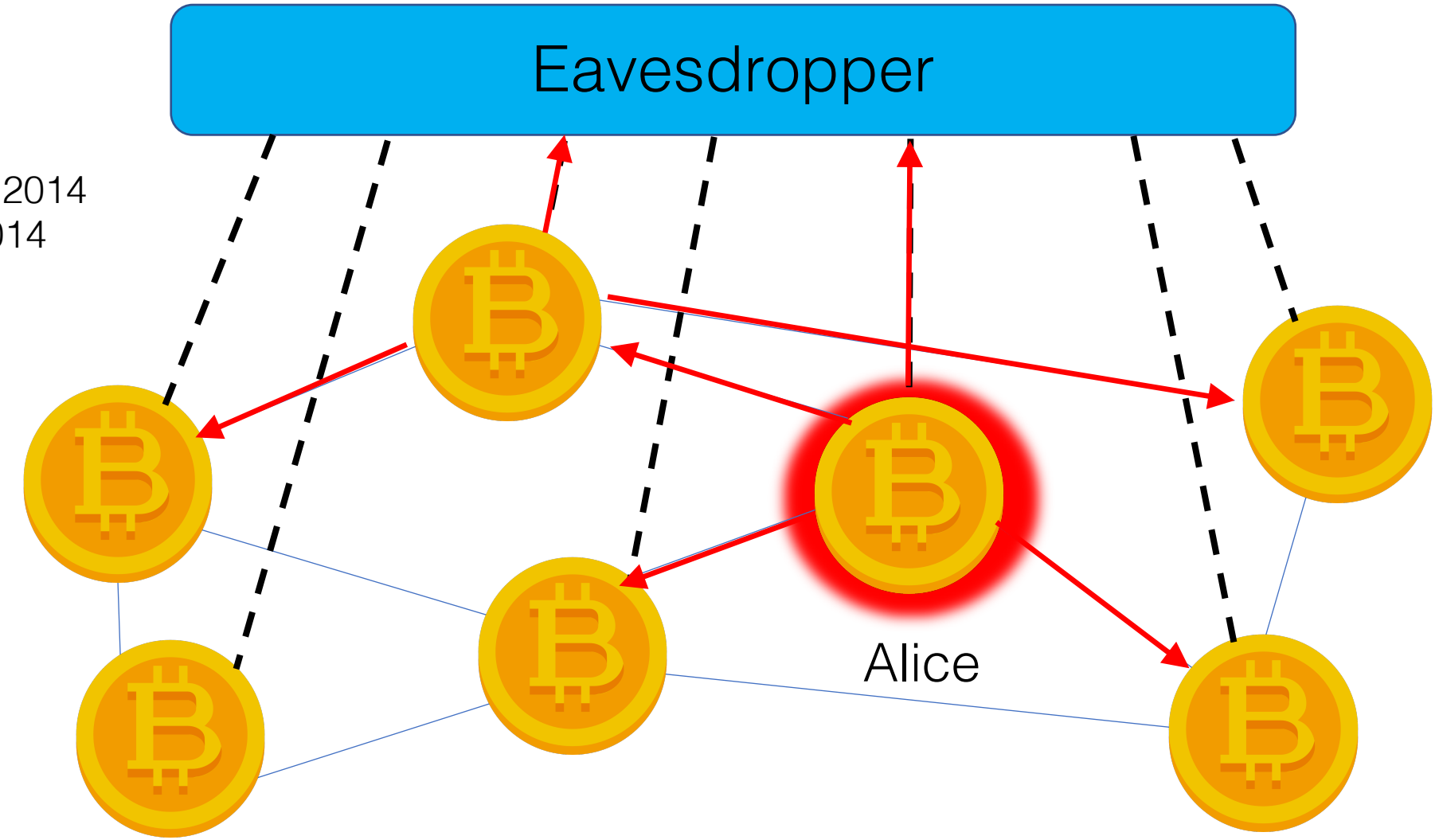
Model

Assumptions and Notation

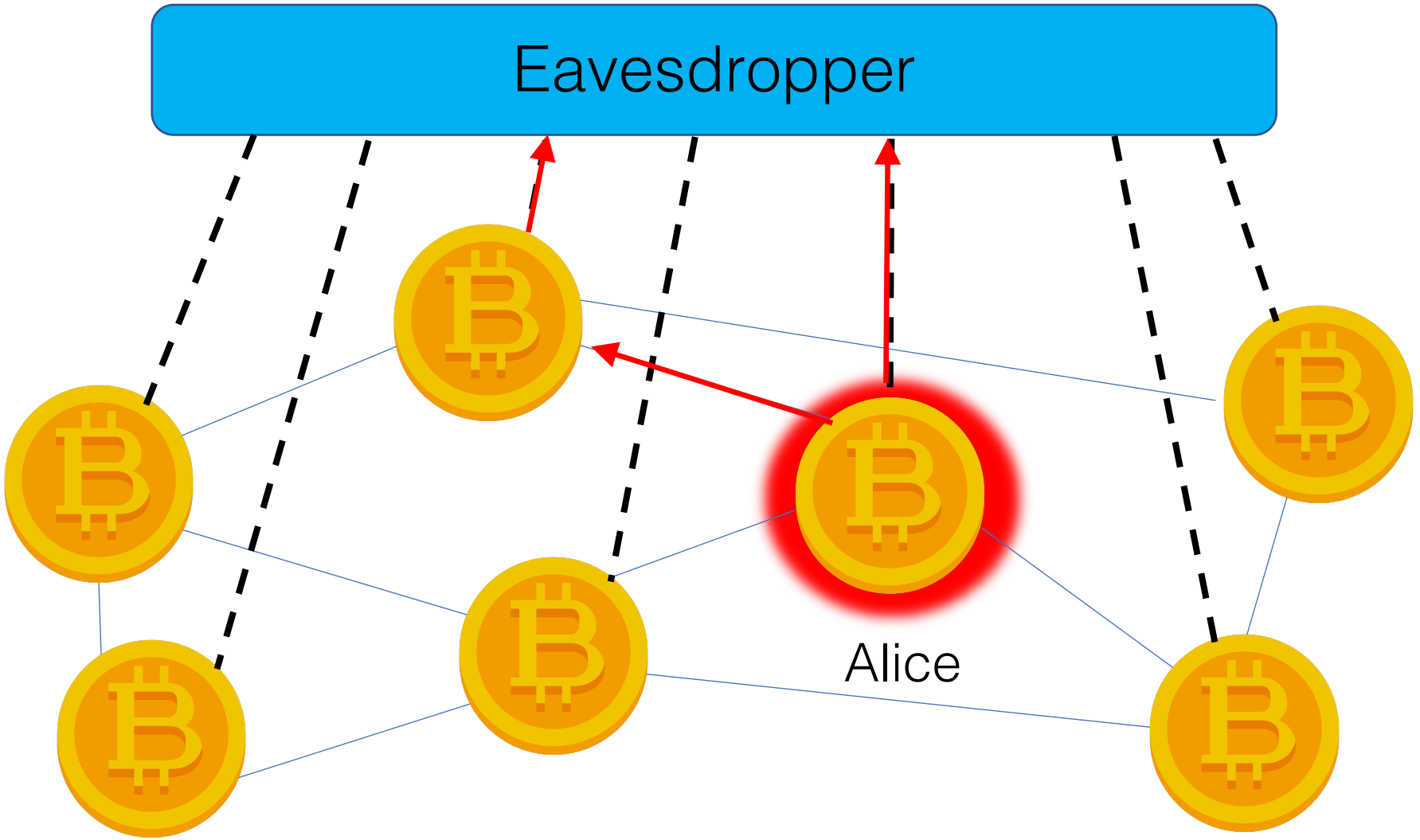


Attacks on the Network Layer

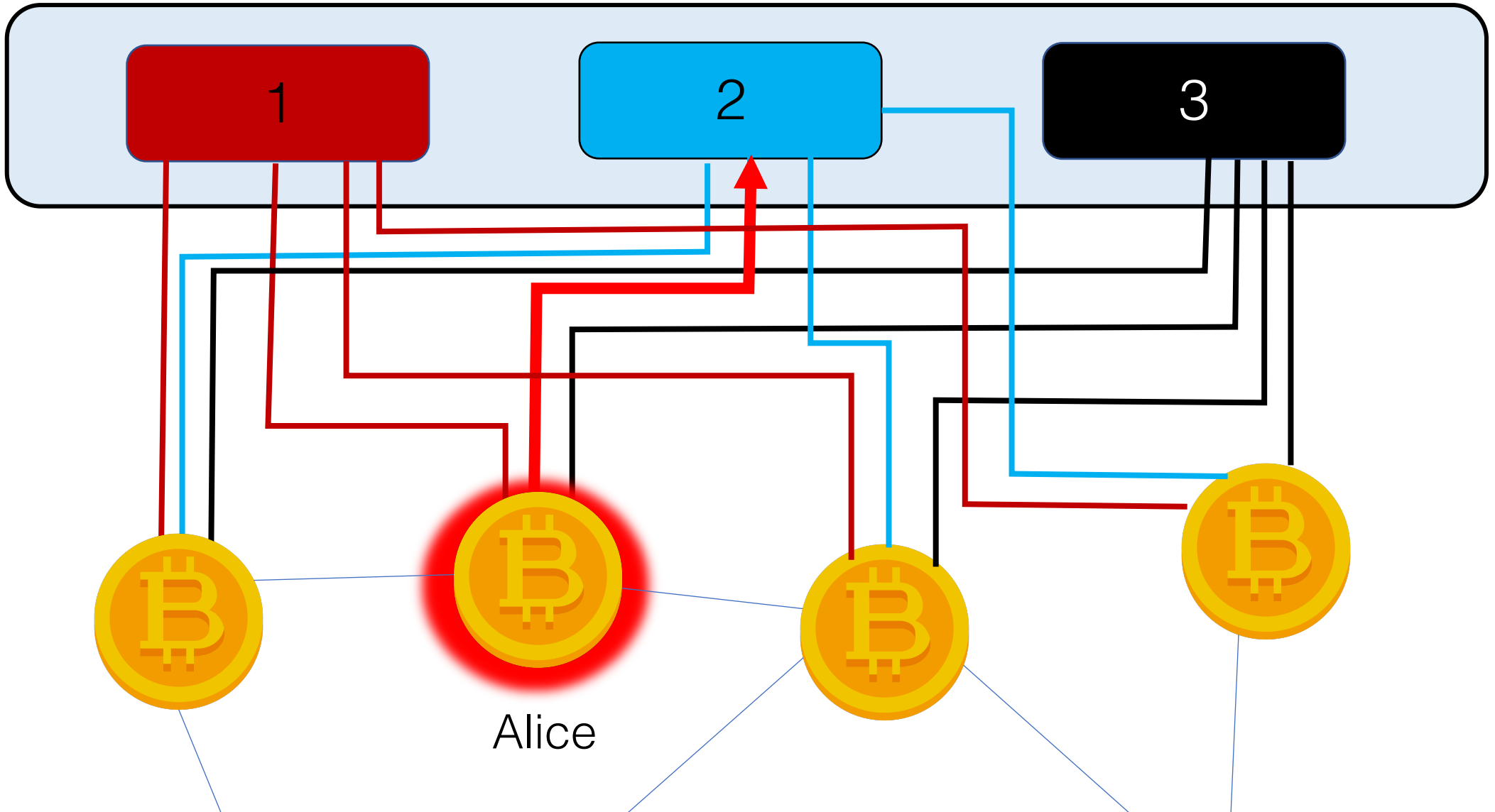
Biryukov et al., 2014
Koshy et al., 2014



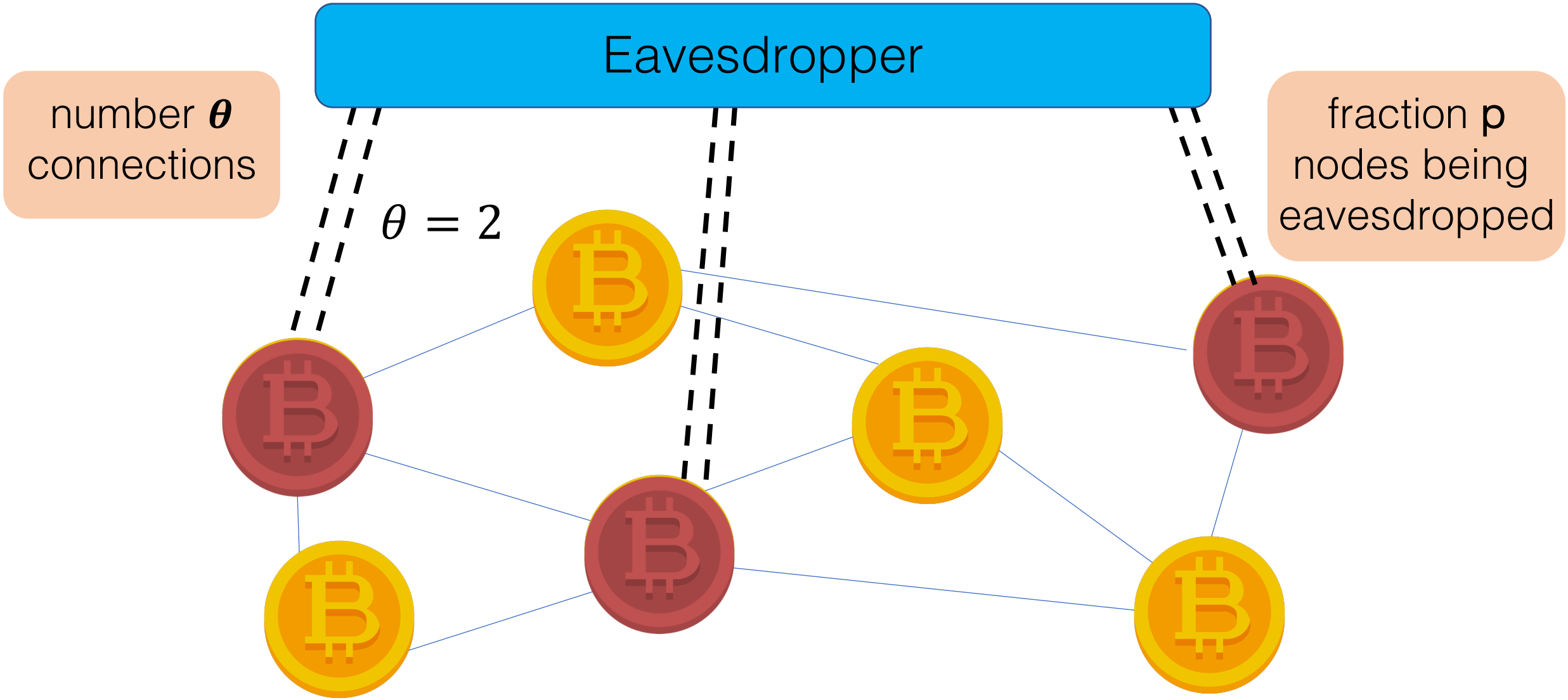
What can go wrong?

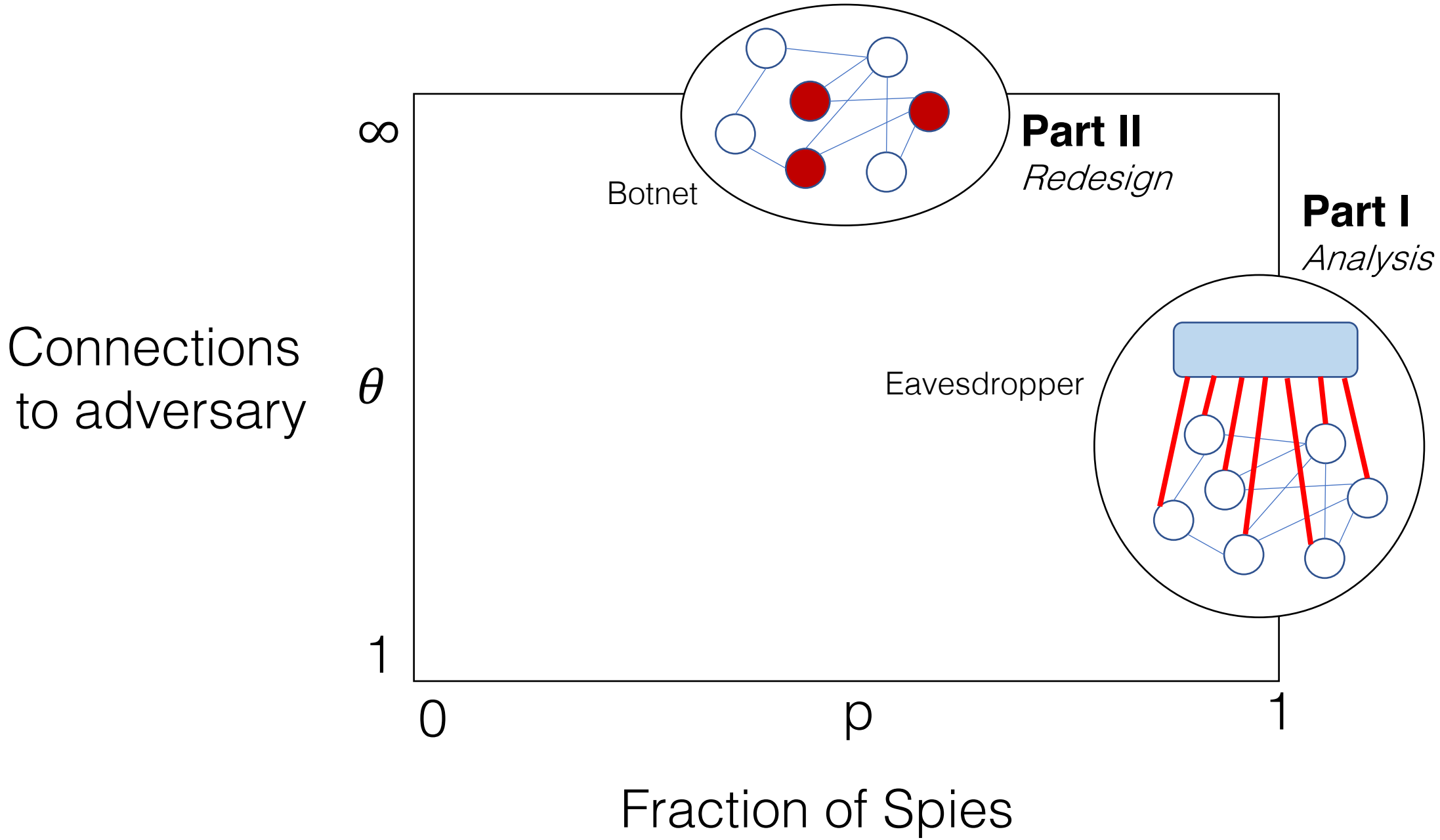


What the eavesdropper can do about it



Summary of adversarial model



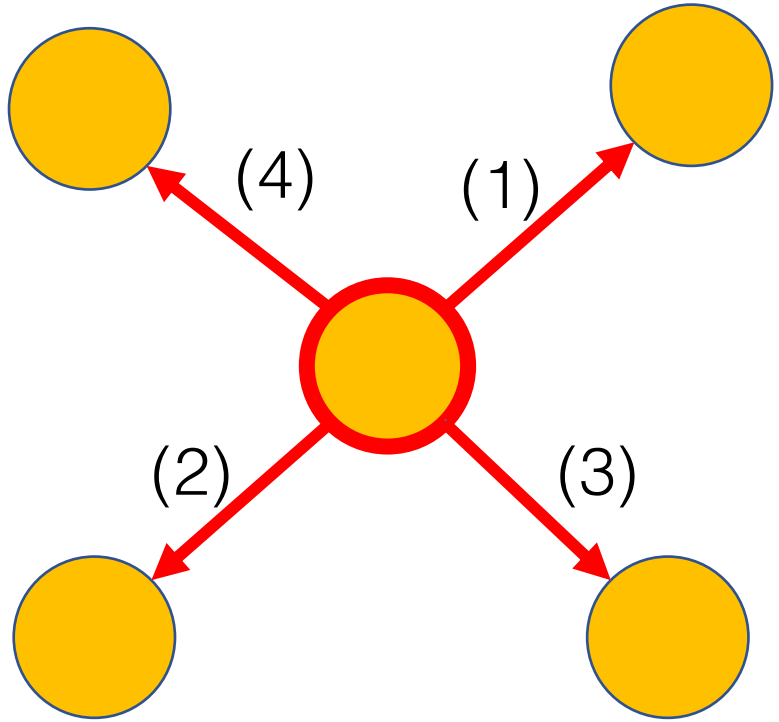

Botnet
Part II
Redesign
Part I
Analysis
Eavesdropper
∞
θ
1
0
p
1
Connections
to adversary
Fraction of Spies

Analysis

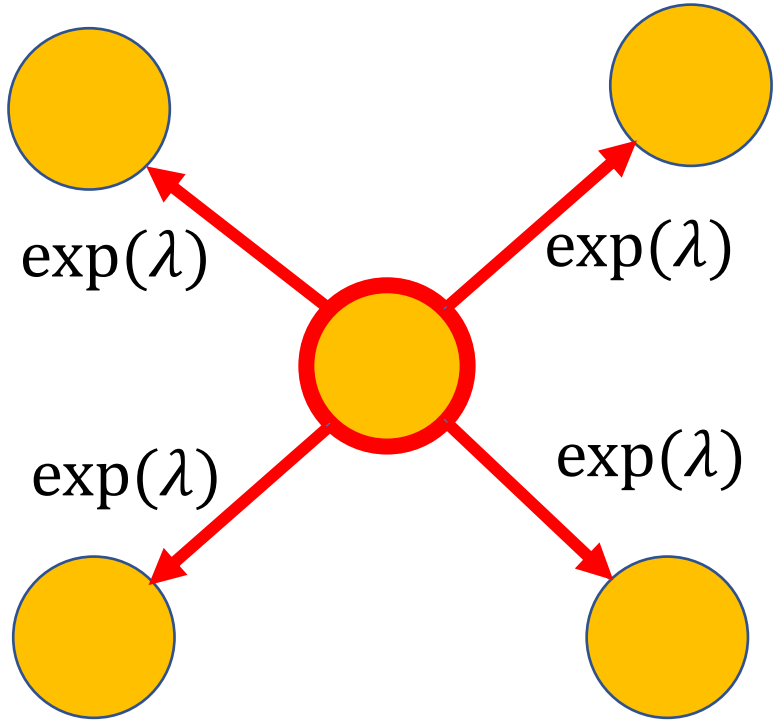
How bad is the problem?

Flooding Protocols

Trickle (pre-2015)



Diffusion (post-2015)

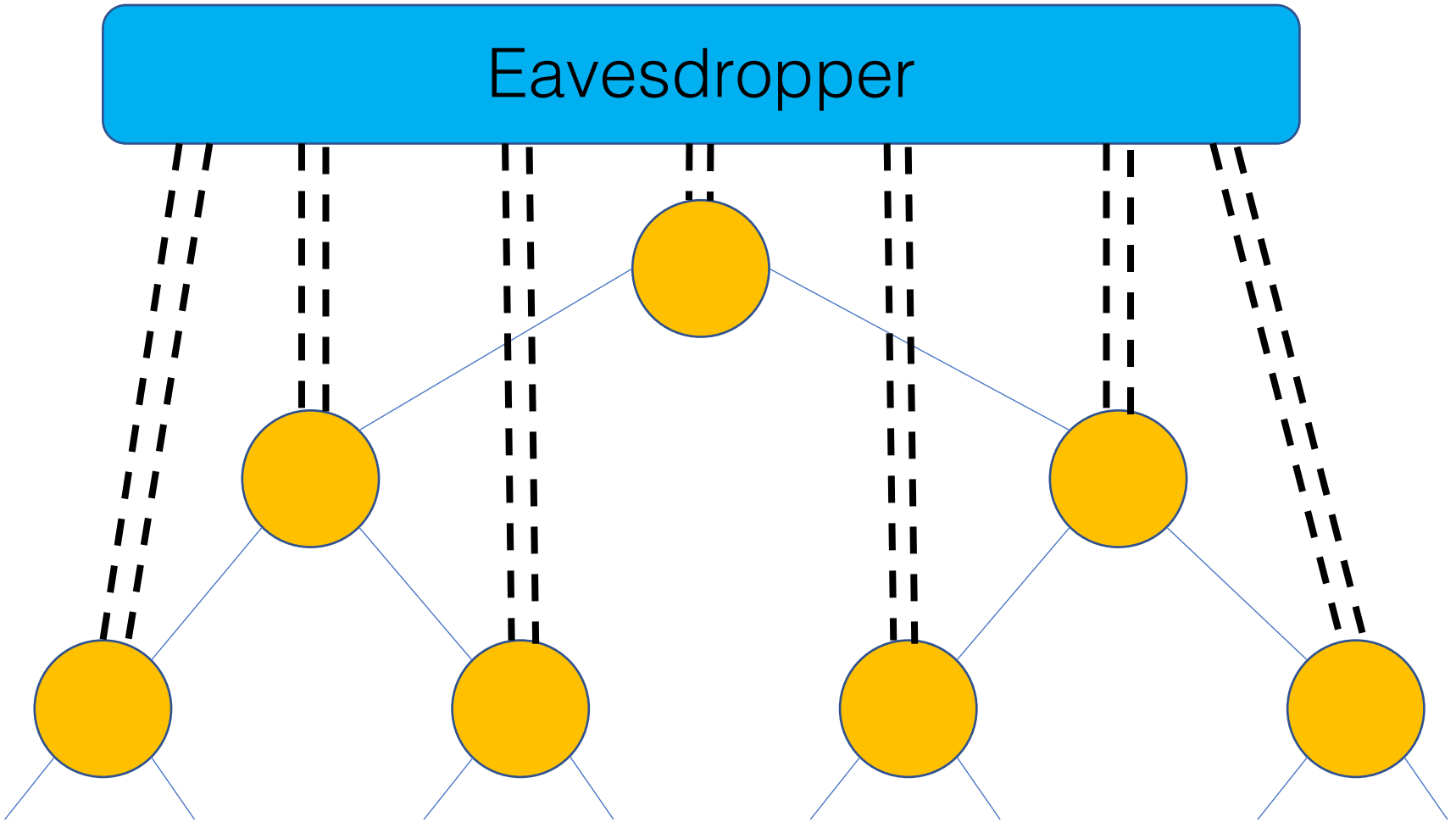


Does diffusion provide stronger anonymity than trickle spreading?

d-regular trees

Fraction of spies $p = 1$

Arbitrary number of connections θ



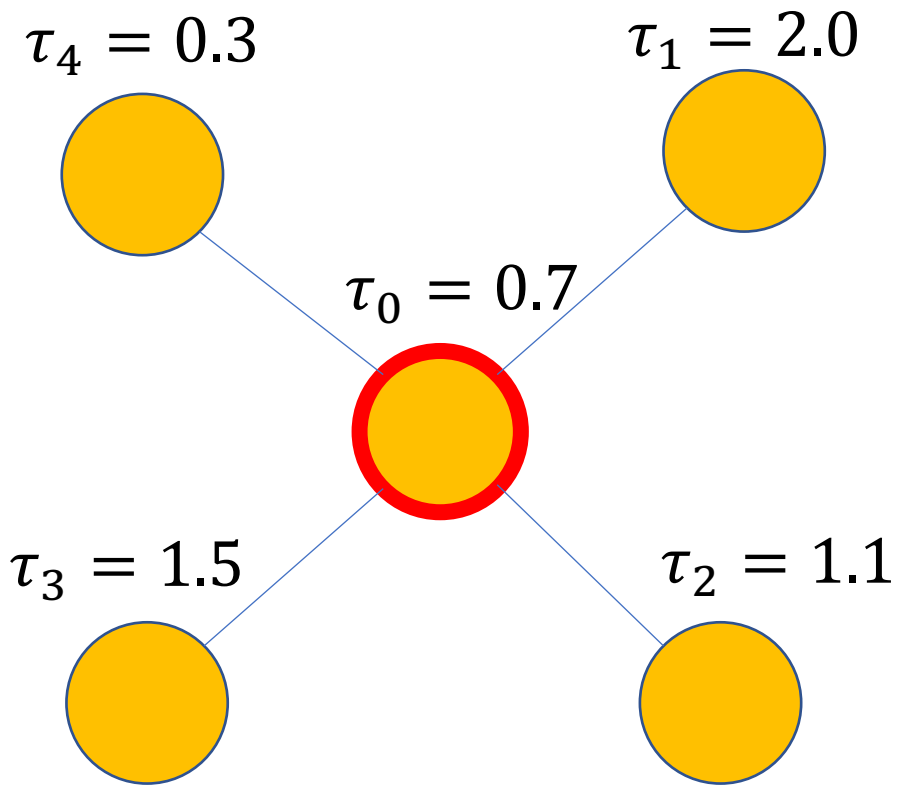
Anonymity Metric

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

timestamps

graph

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

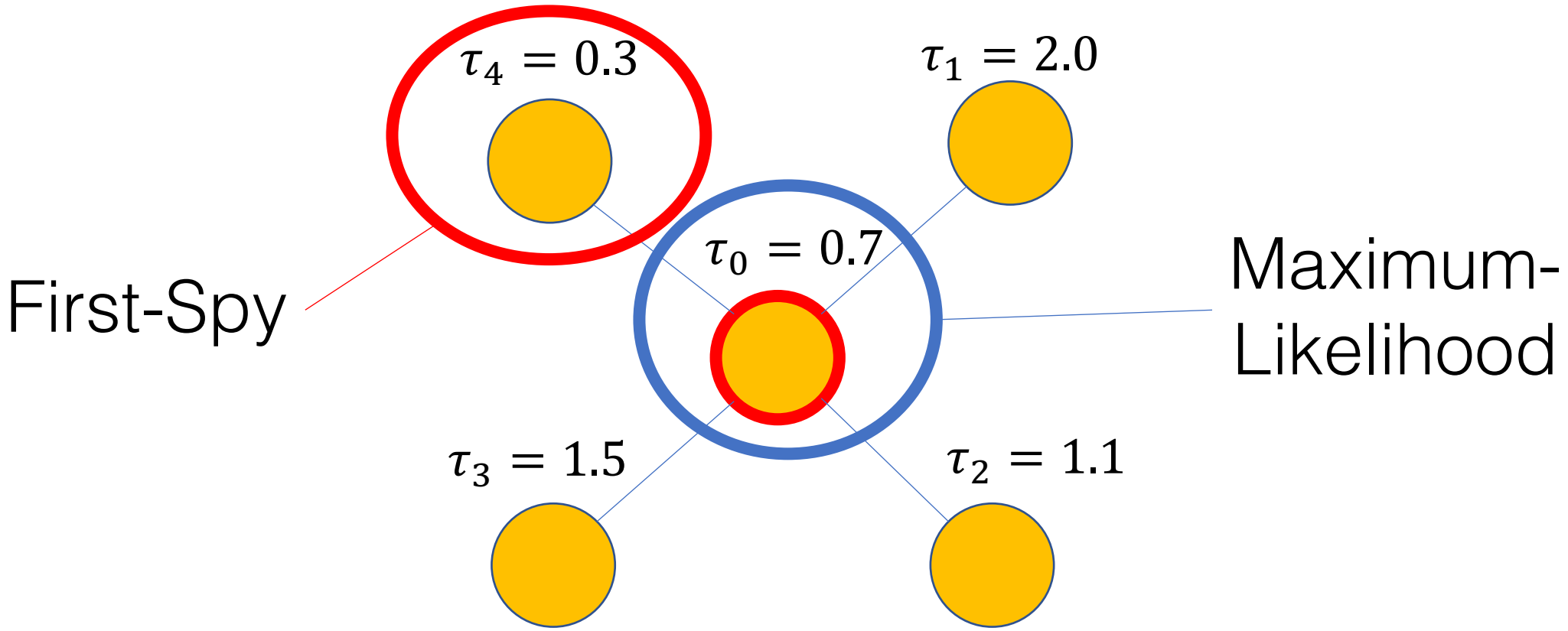


Estimators

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

timestamps

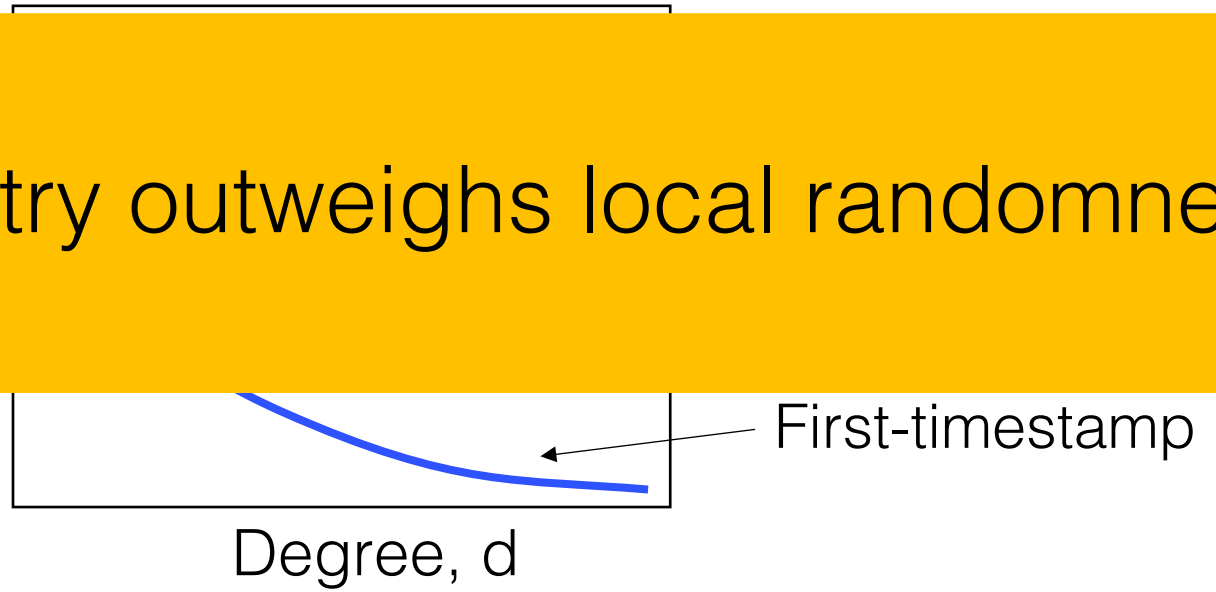
graph



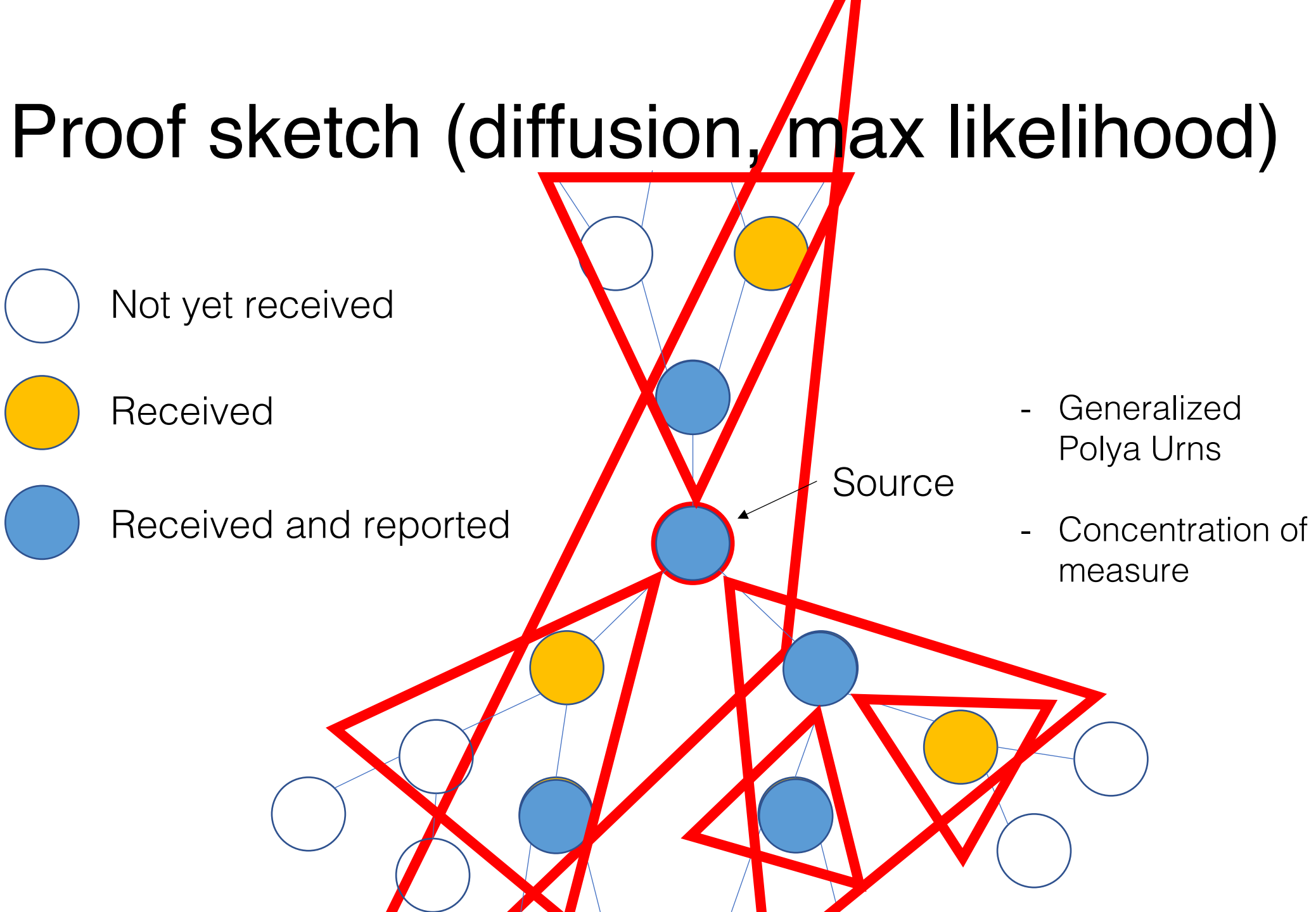
Results: d-Regular Trees

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

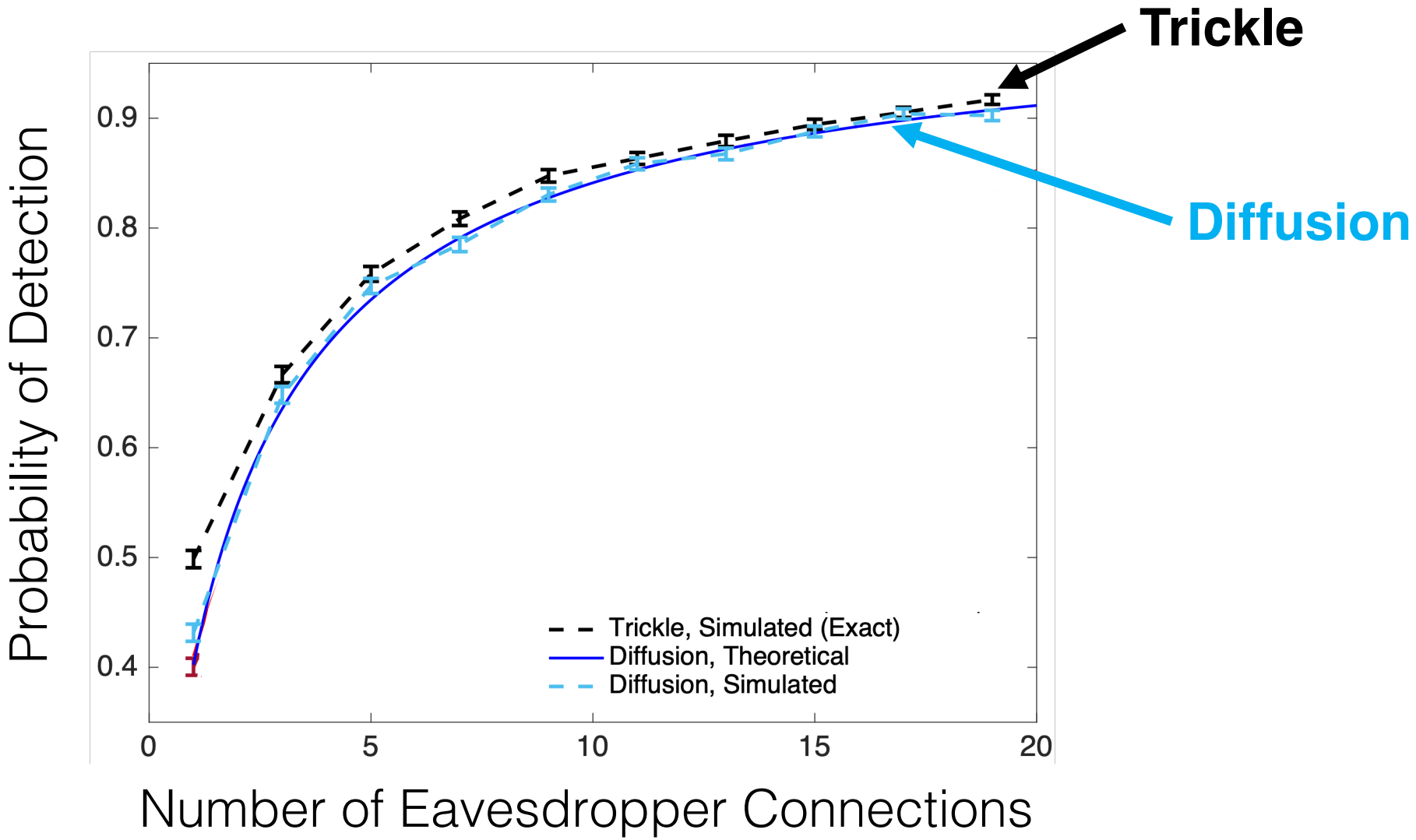
Intuition: Symmetry outweighs local randomness!



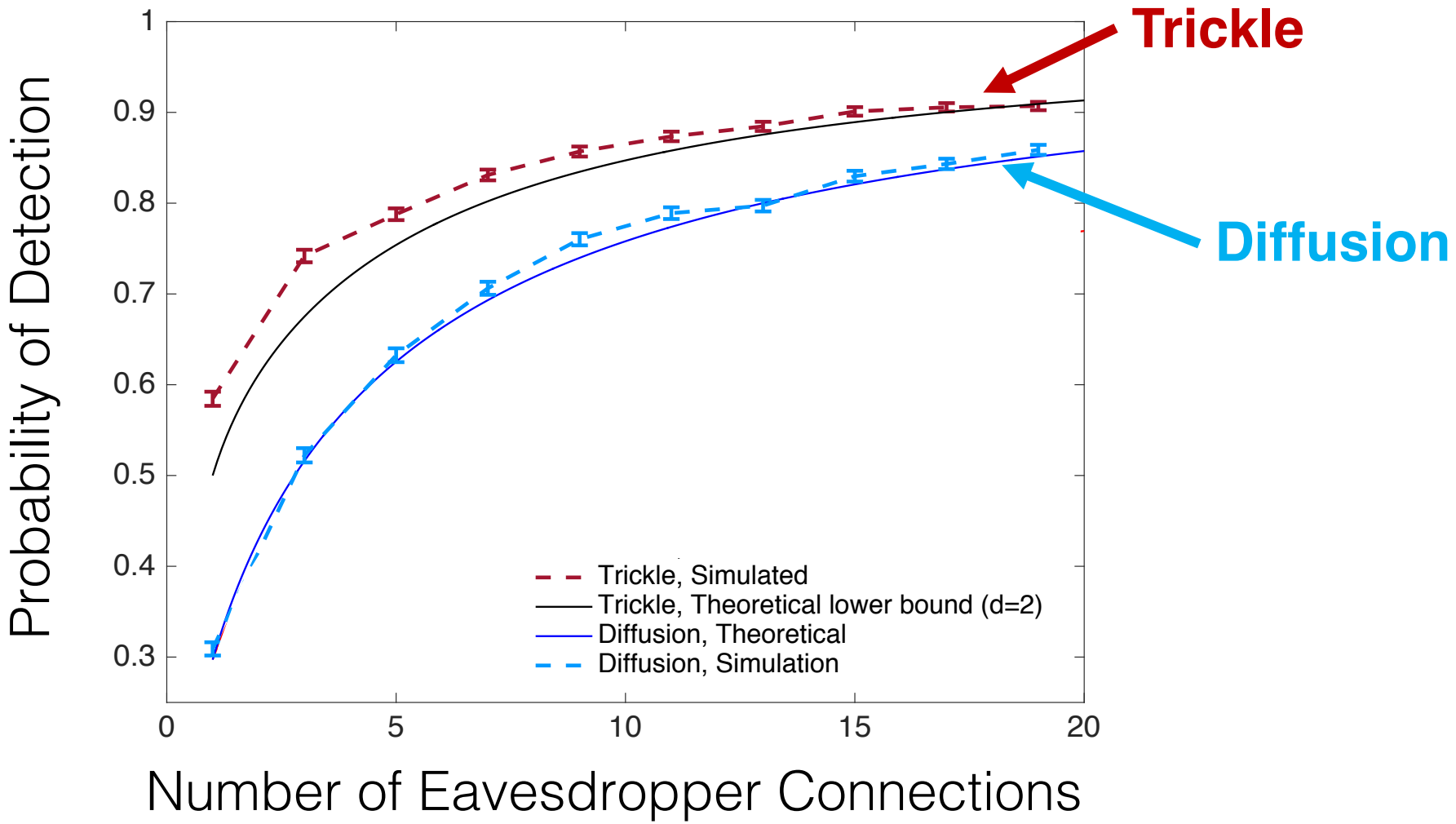
Proof sketch (diffusion, max likelihood)



Results: Trees



Results: Bitcoin Graph

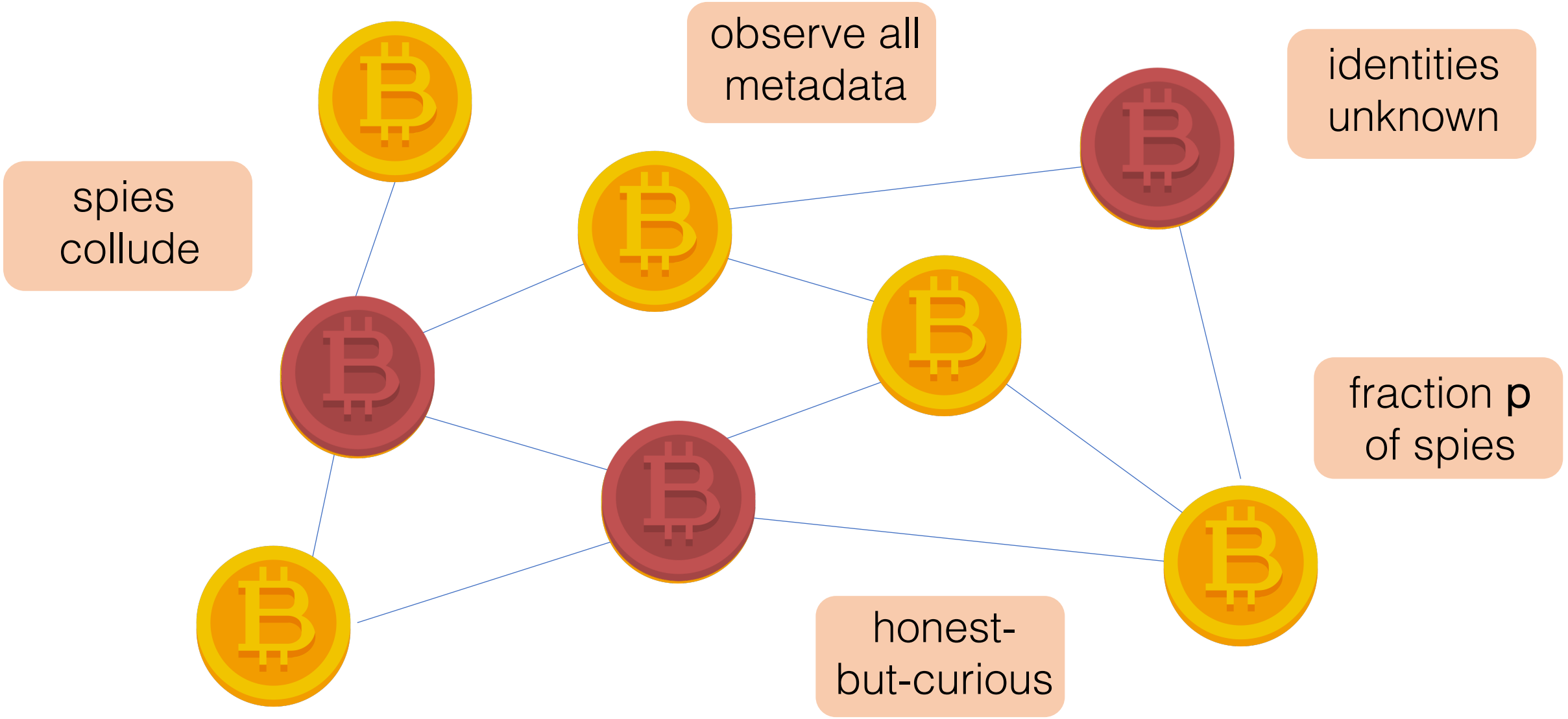


Diffusion does not have (significantly) better anonymity properties than trickle.

Redesign

Can we design a better network?

Botnet adversarial model



Metric for Anonymity

Recall

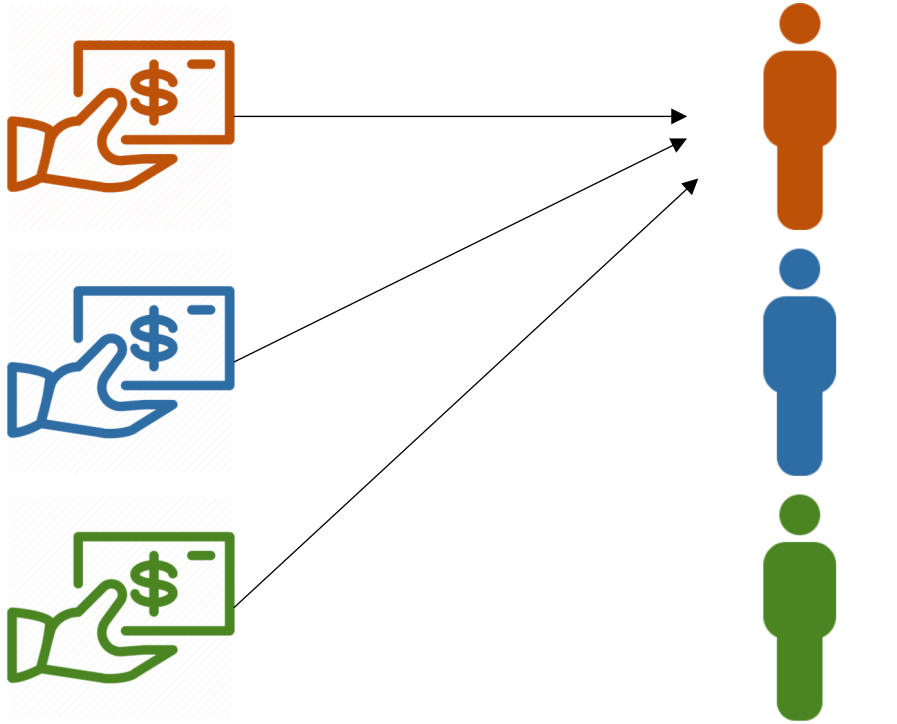
$$\frac{1}{n}\sum_{v} 1\{M(v\text{'s tx}) = v\}$$

Number honest users

User

Mapping

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



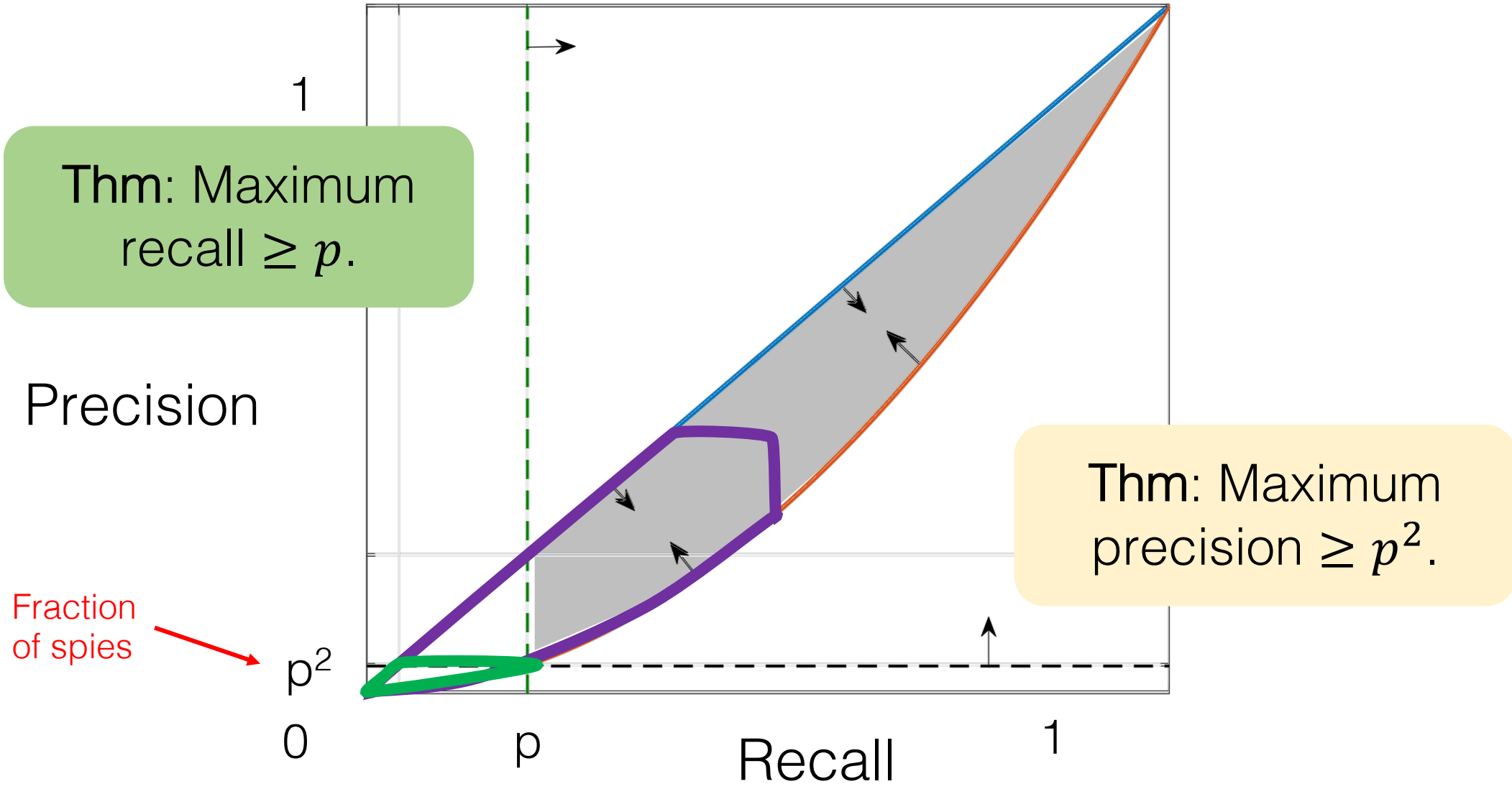
Precision

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

Goal:

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

Fundamental Limits

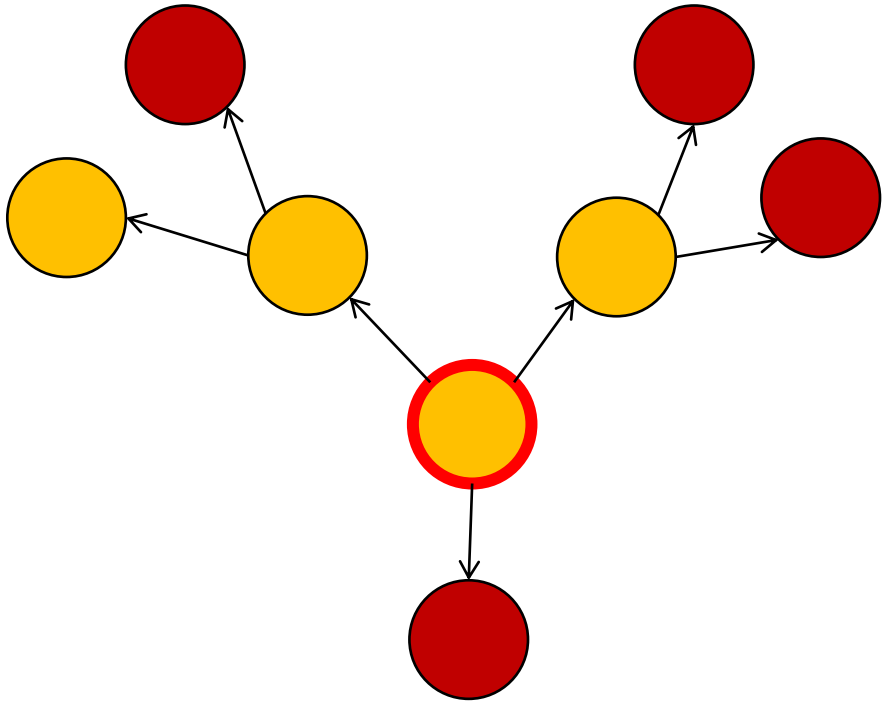


Thm: Maximum recall $\geq p$.

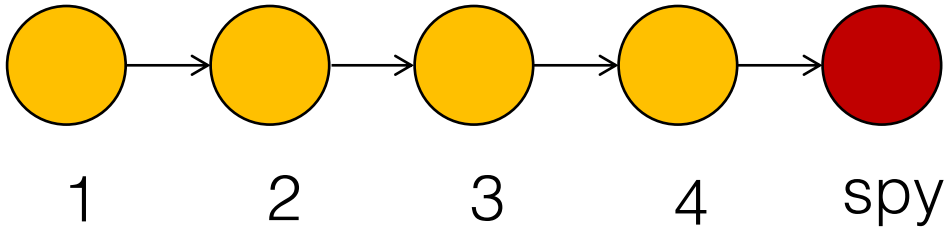
Thm: Maximum precision $\geq p^2$.

What are we looking for?

Asymmetry



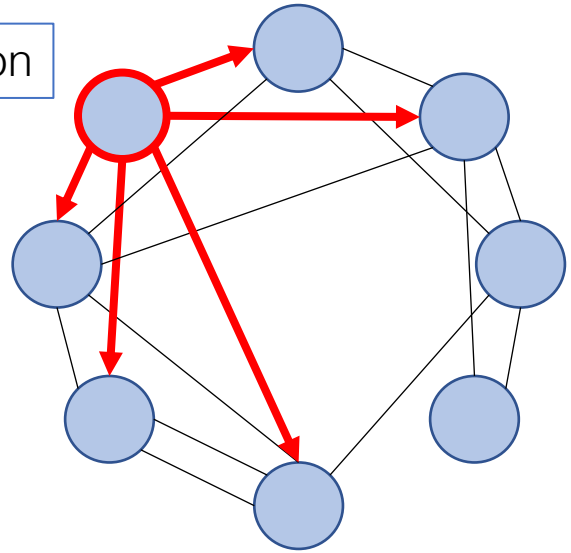
Mixing



What can we control?

Spreading Protocol

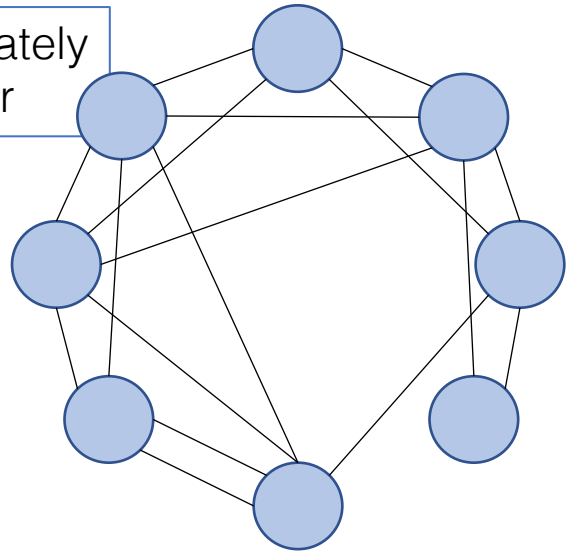
Diffusion



Given a graph, how do we spread content?

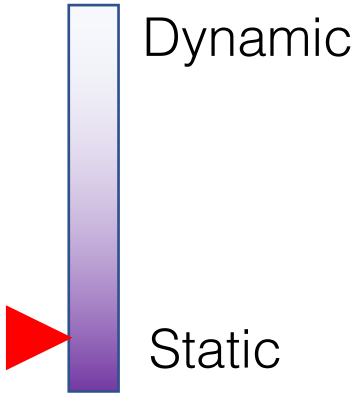
Topology

Approximately regular



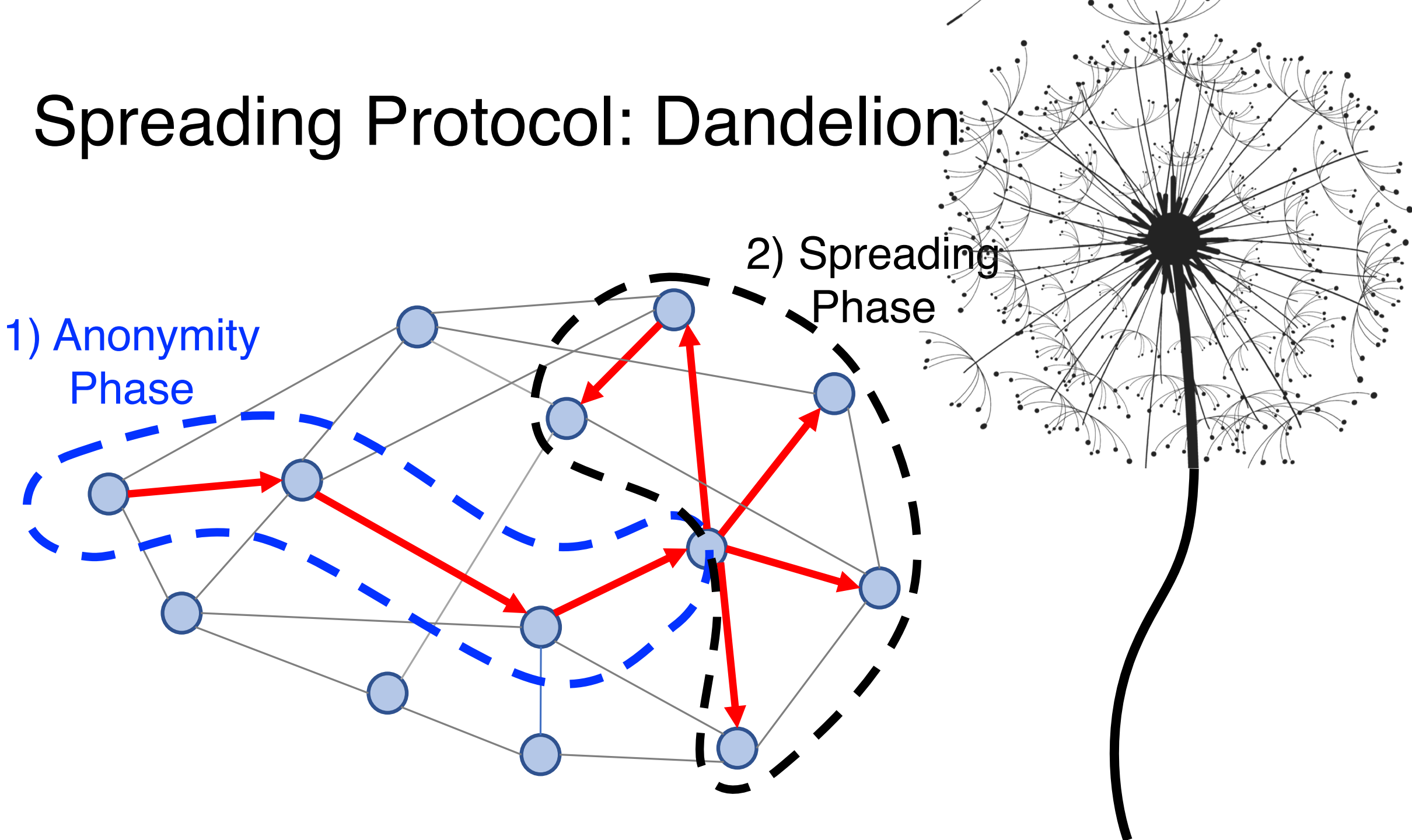
What is the underlying graph topology?

Dynamicity



How often does the graph change?

Spreading Protocol: Dandelion



Why Dandelion spreading?

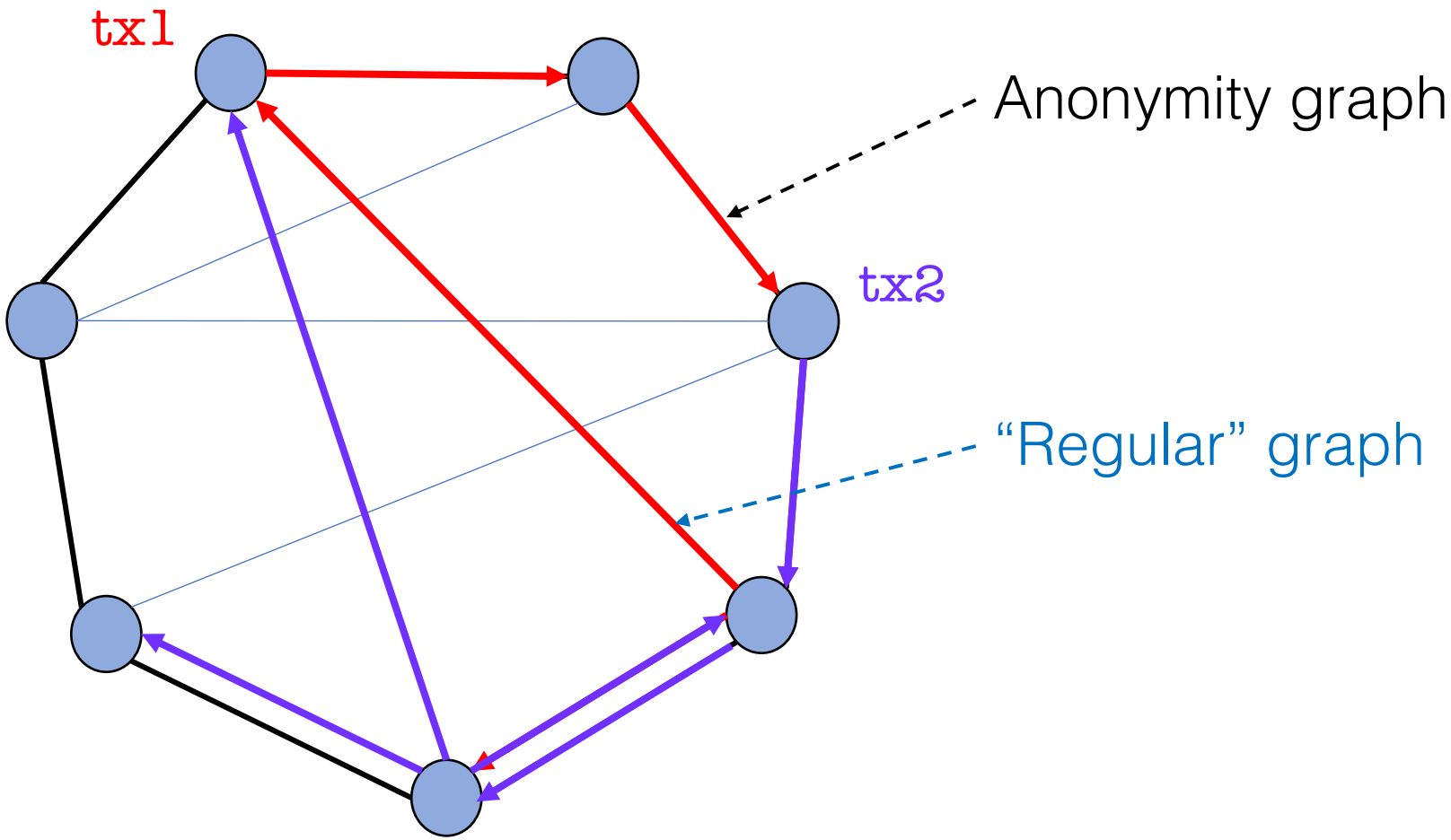
Theorem: Dandelion spreading has an optimally low maximum recall of $p + O\left(\frac{1}{n}\right)$.

Theorem: Fundamental lower bound = p

fraction of spies

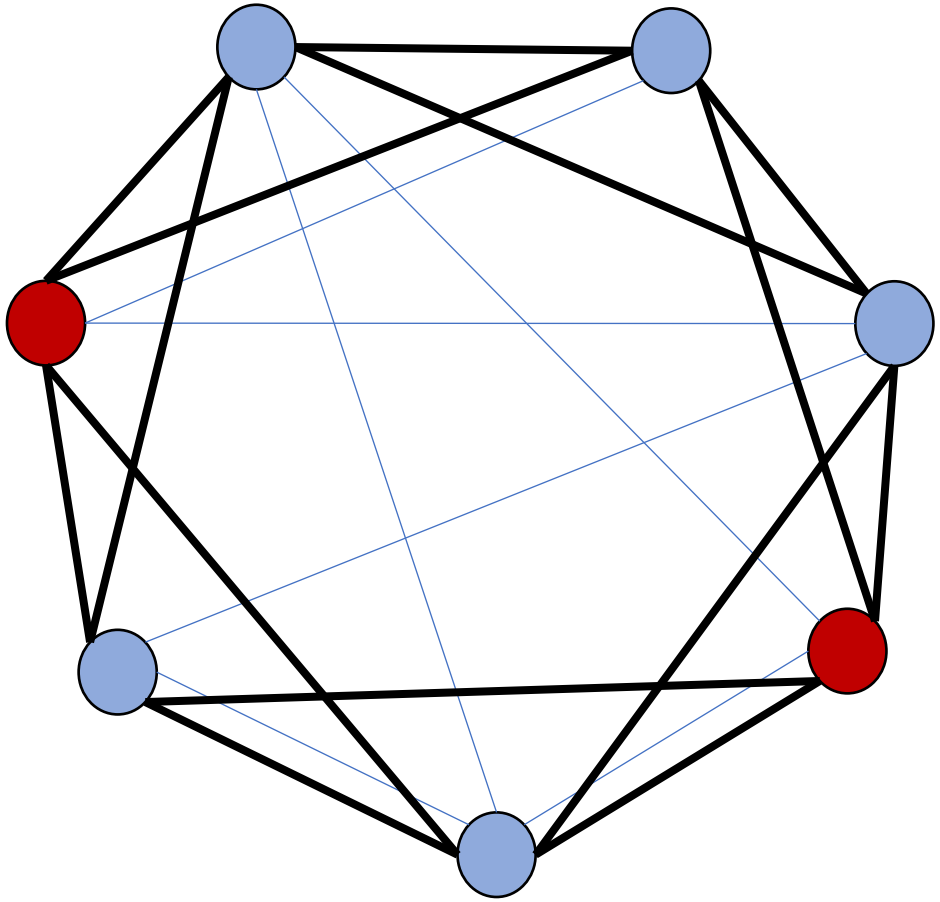
number of nodes

Graph Topology: Line



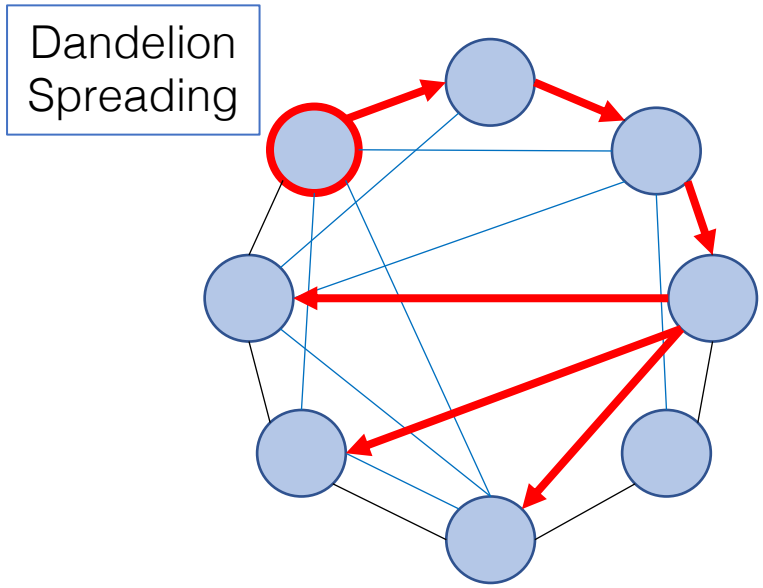
Dynamicity: High

Change the anonymity graph frequently.



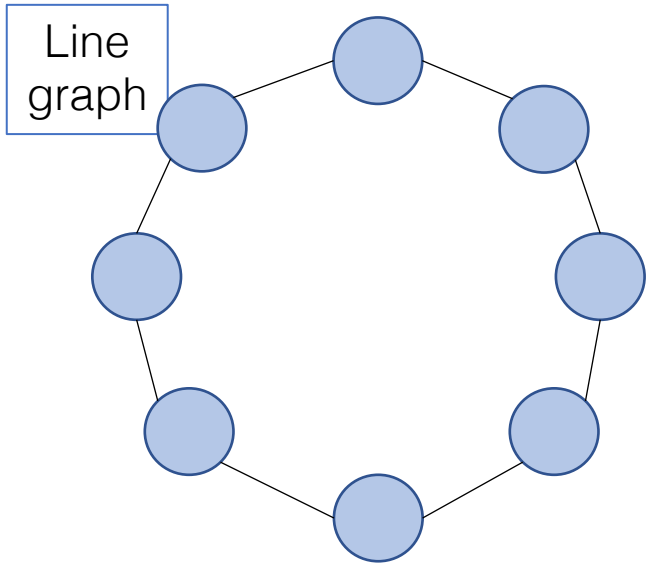
DANDELION Network Policy

Spreading Protocol



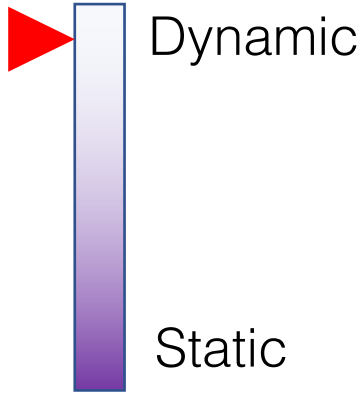
Given a graph, how do we spread content?

Topology



What is the anonymity graph topology?

Dynamicity



How often does the graph change?

Theorem: Fundamental lower bound = p^2

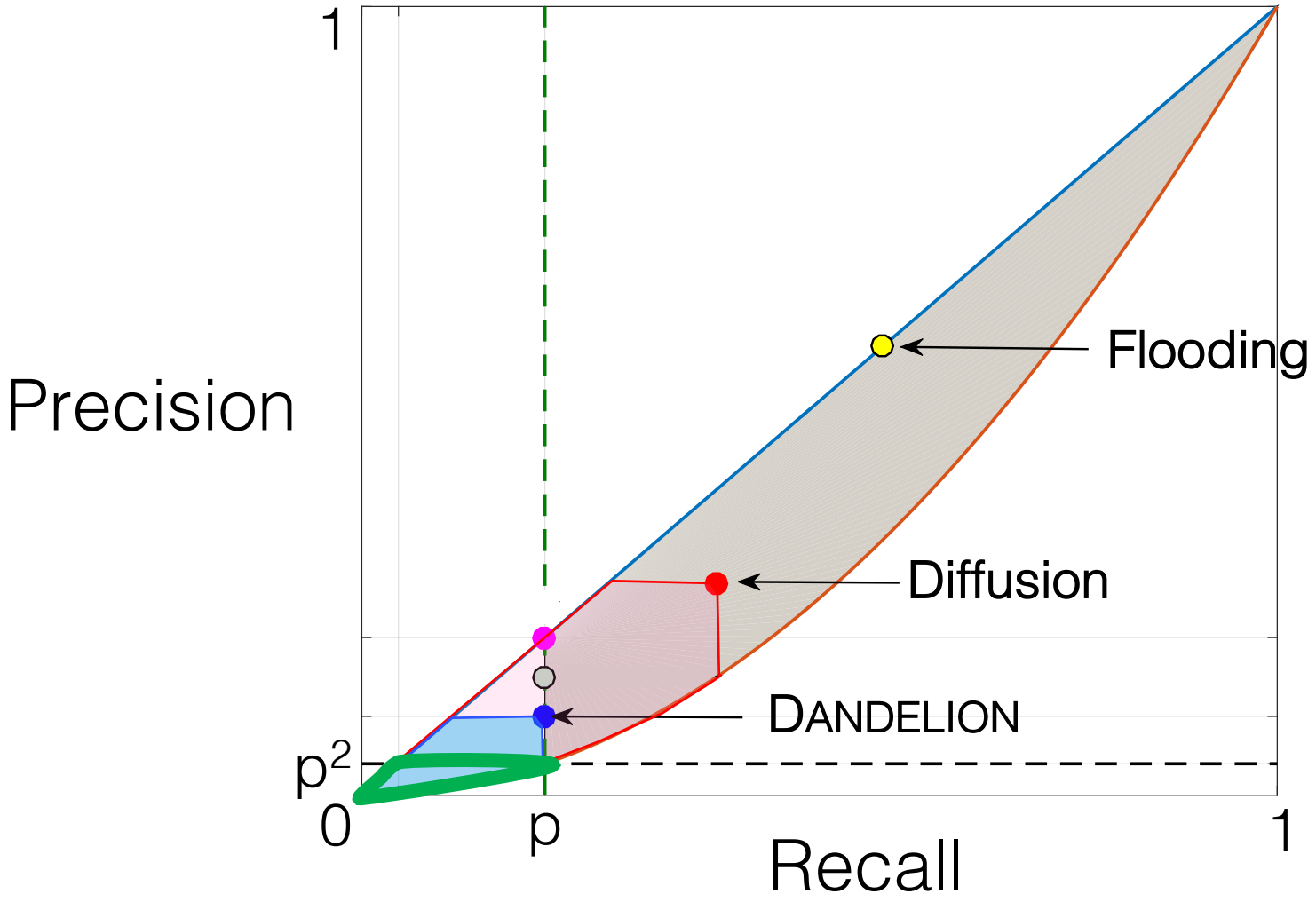
Theorem: DANDELION has a nearly-optimal maximum precision of $\frac{2p^2}{1-p}\log\left(\frac{2}{p}\right) + O\left(\frac{1}{n}\right)$.*

fraction of spies

number of nodes

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

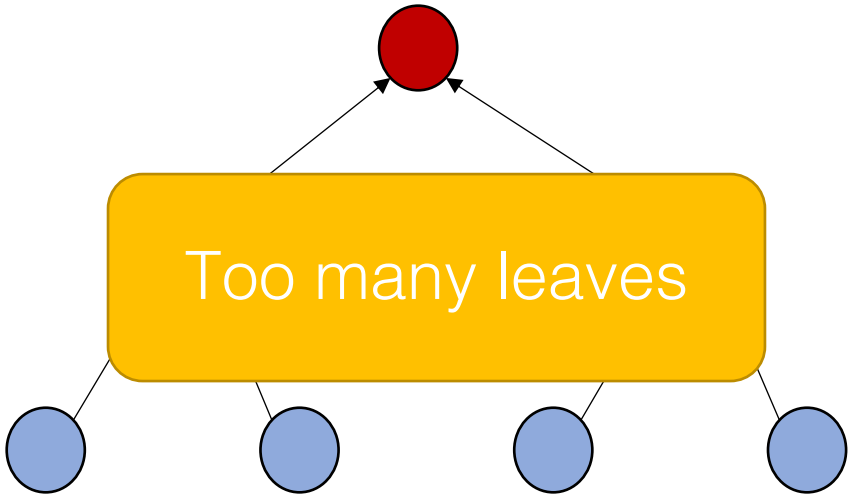
Performance: Achievable Region



Why is DANDELION good?

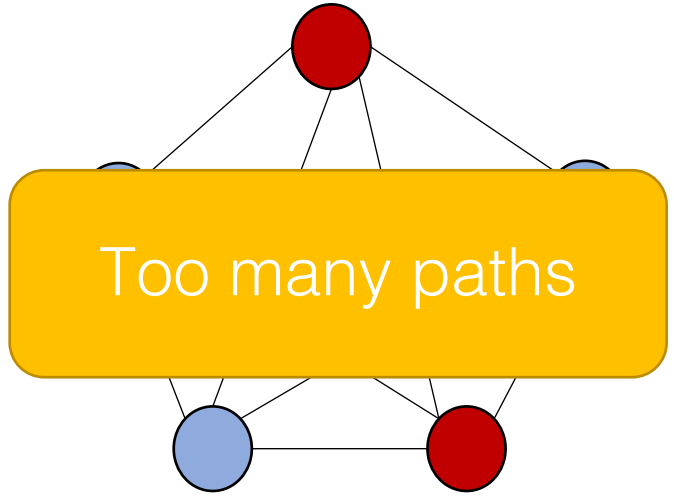
Strong mixing properties.

Tree



Precision: $O(p)$

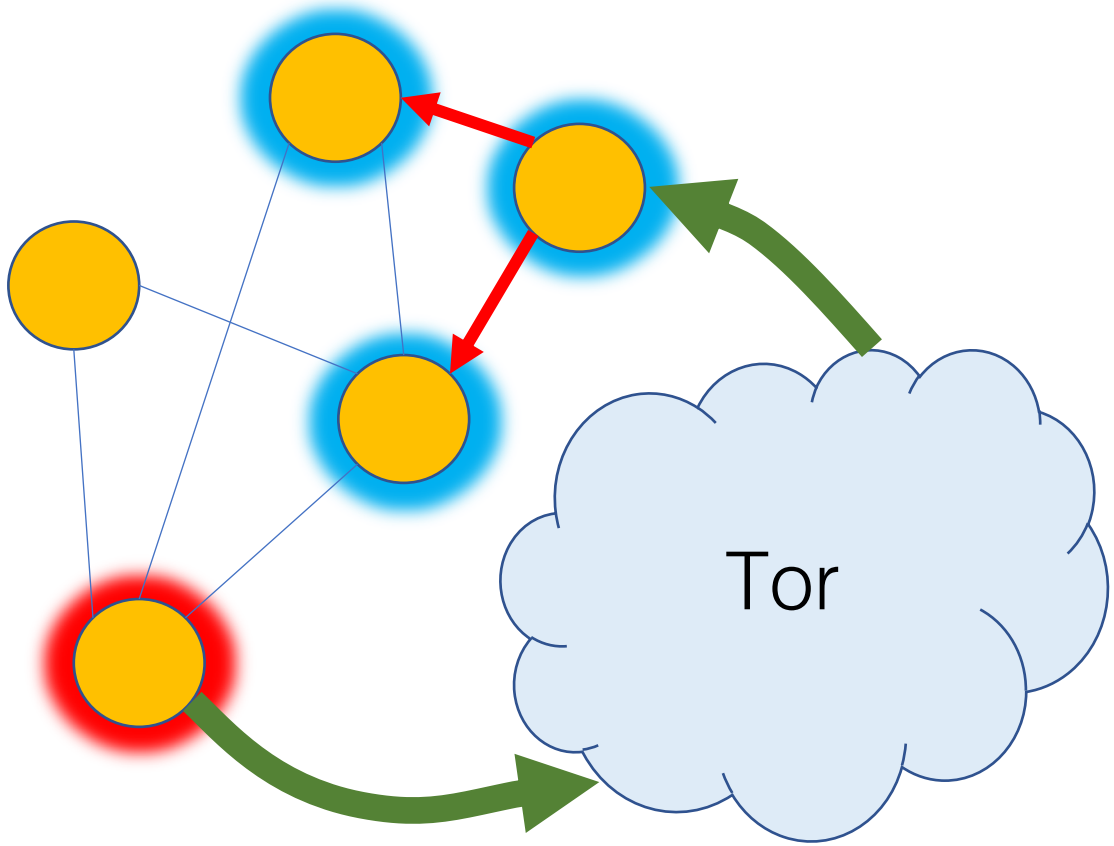
Complete graph



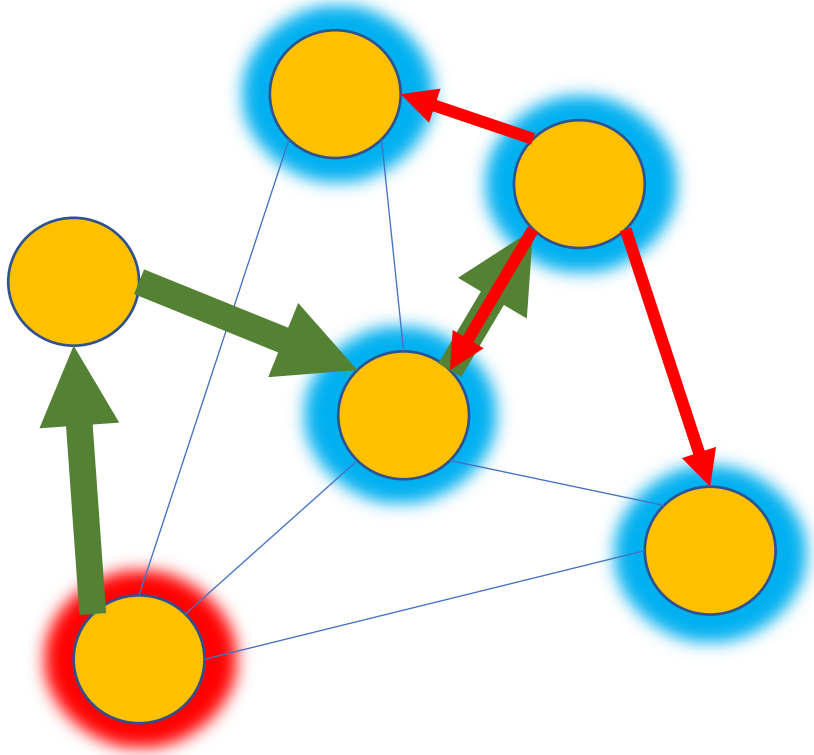
Precision: $\frac{p}{1-p}(1-e^{p-1})$

Why not alternative solutions?

Connect through Tor

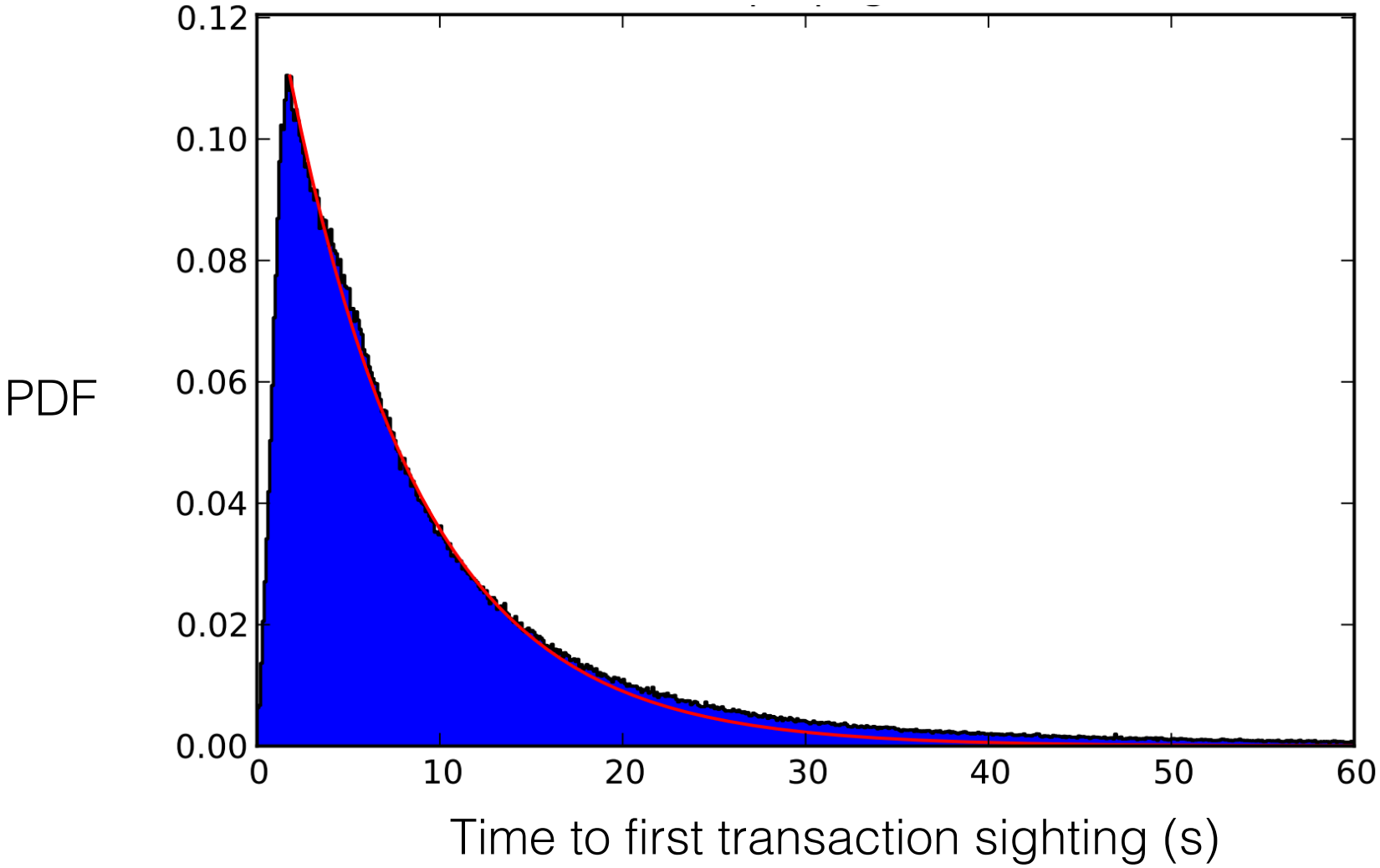


I2P Integration (e.g. Monero)



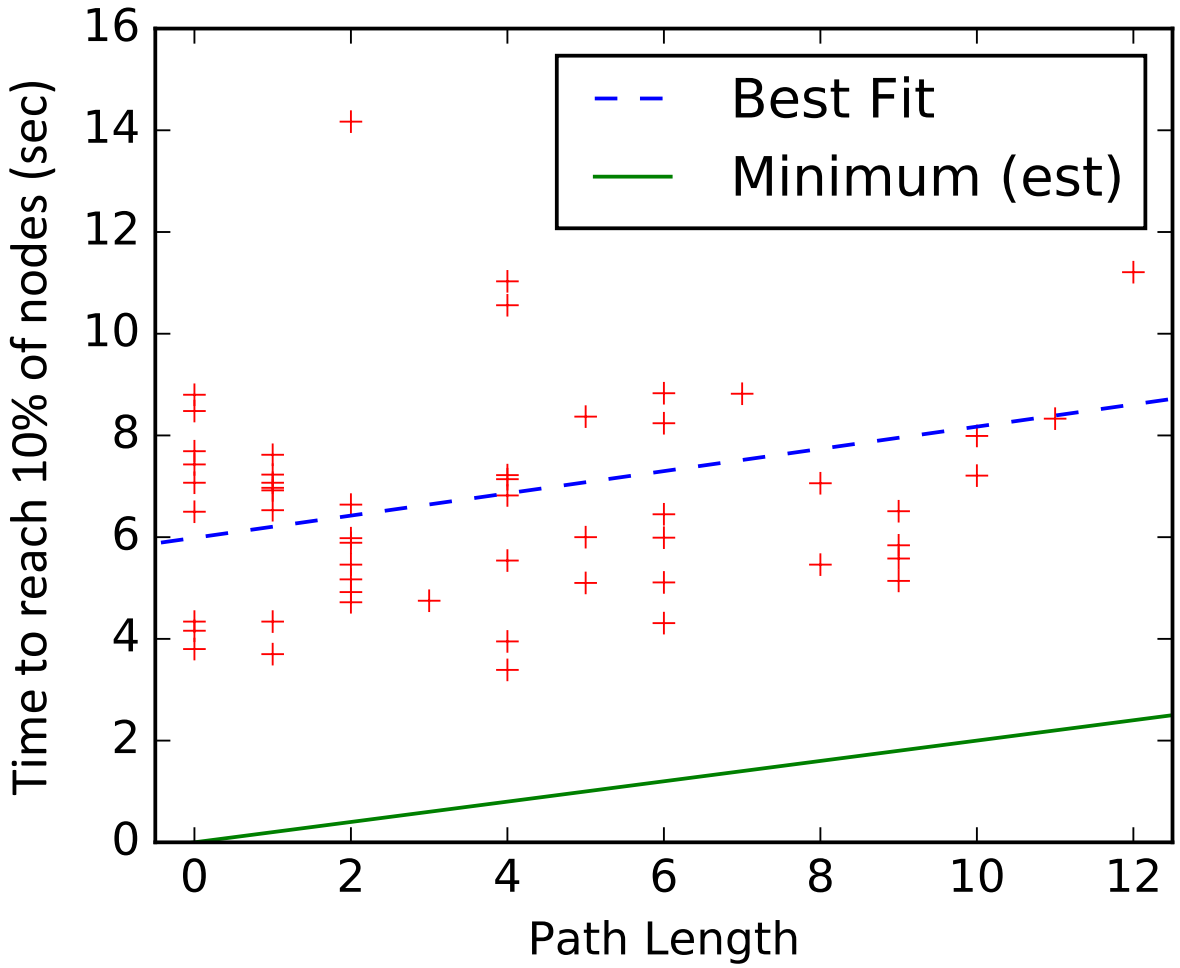
How practical is this?

Latency Overhead: Estimate

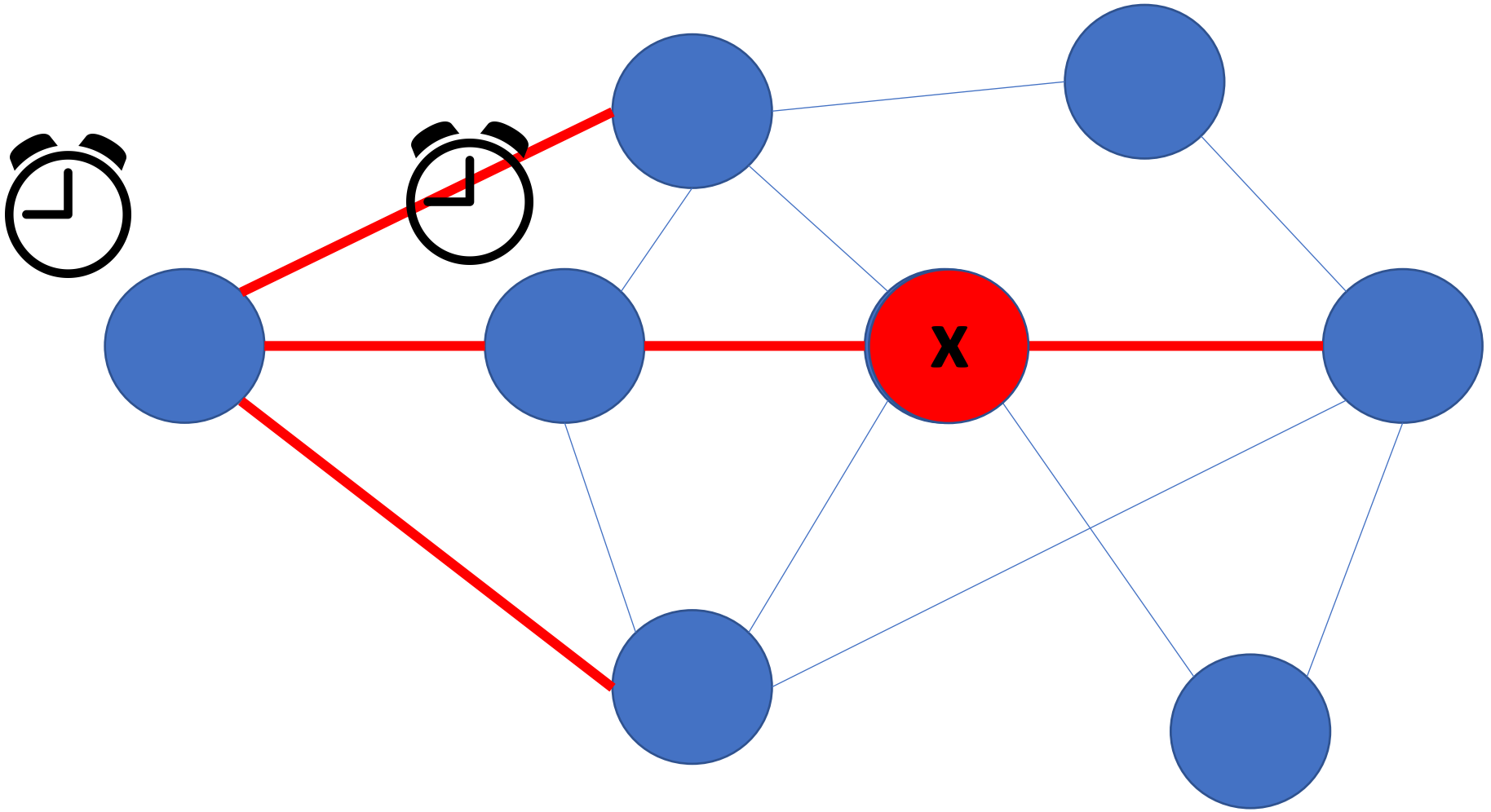


Information Propagation in the Bitcoin Network, Decker and Wattenhofer, 2013

Empirical Delay Distribution



Practical challenge: Black hole attack



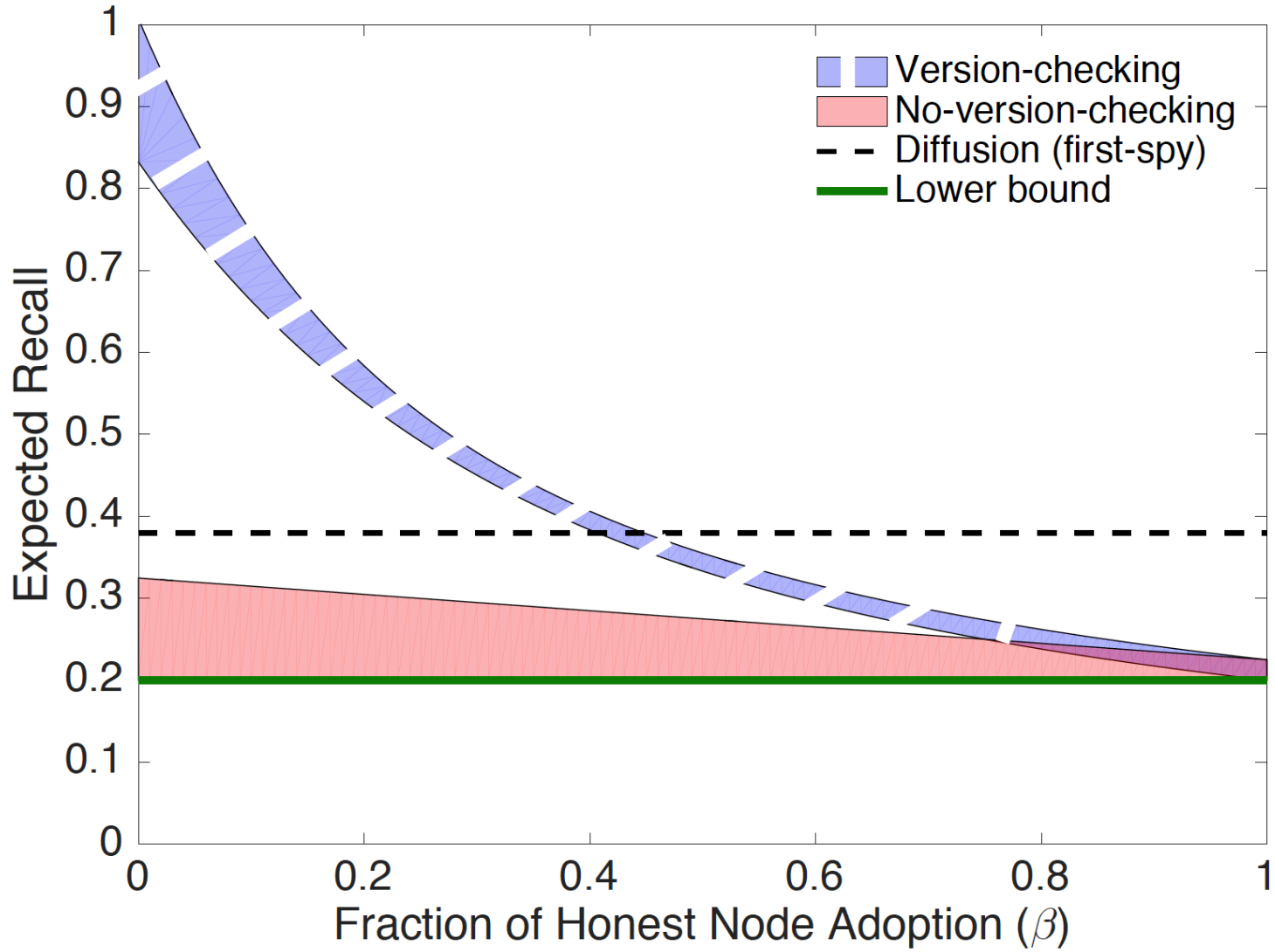
Practical Challenges: Black hole attack

When you switch a route, what happens to transactions you've already sent?

A. Could resend sent transactions on the new route

B. This makes RBF challenging

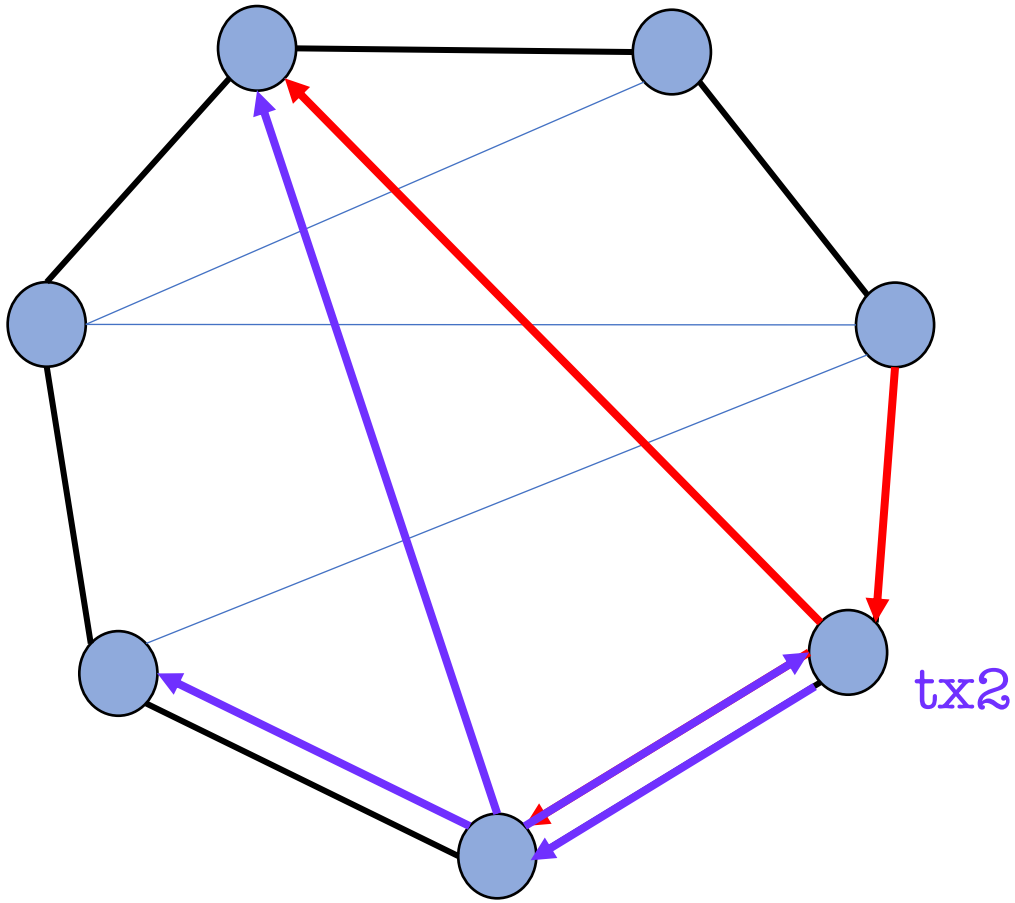
Practical Challenges: Partial deployment



End story

- Complexity/robustness seems to be a barrier

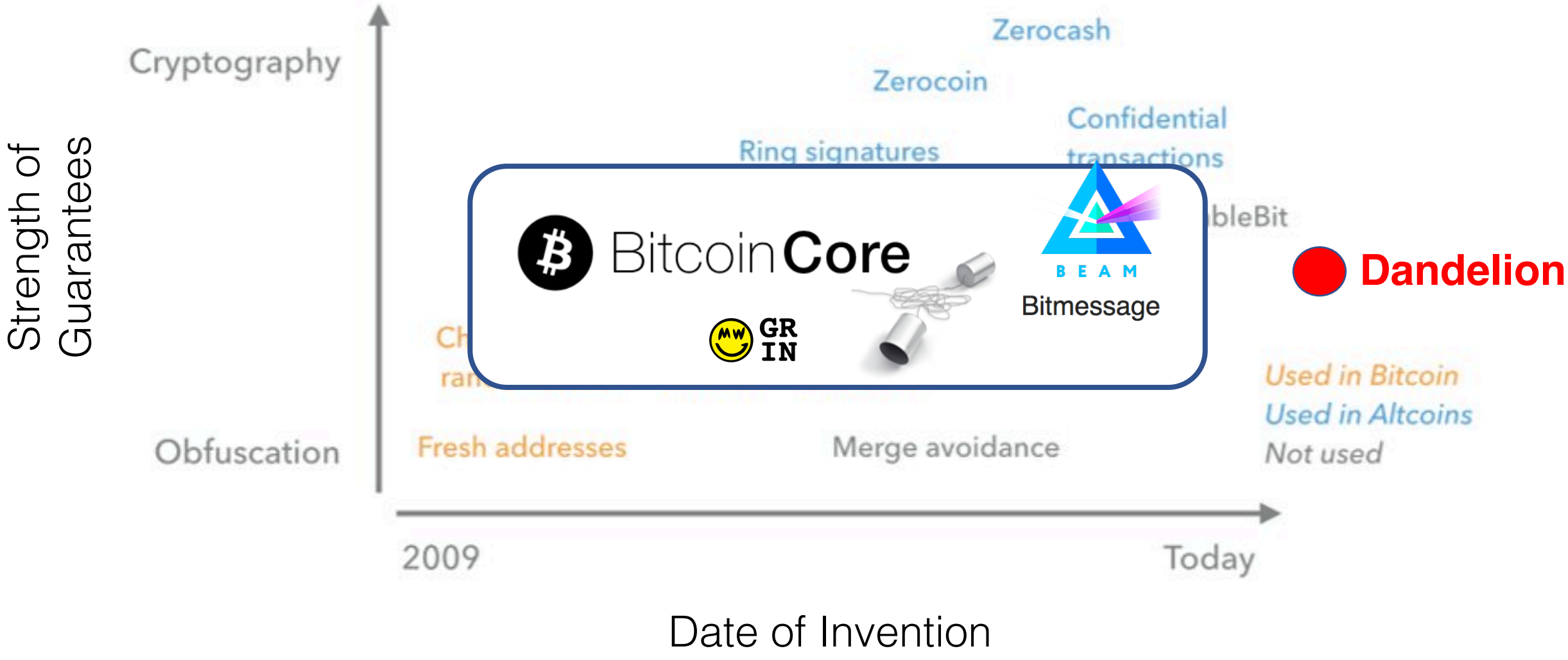
Dandelion-Lite



Only make 1 hop in the stem phase!

Dandelion-Lite: Privacy guarantees

- Similar guarantees to Dandelion when we assume that the adversary knows the graph
- Weaker guarantees when the adversary doesn't know the graph
- Still needed: simulations!!



Narayanan and Möser, 2017

Take-Home Messages

1) Bitcoin's P2P network has poor anonymity.

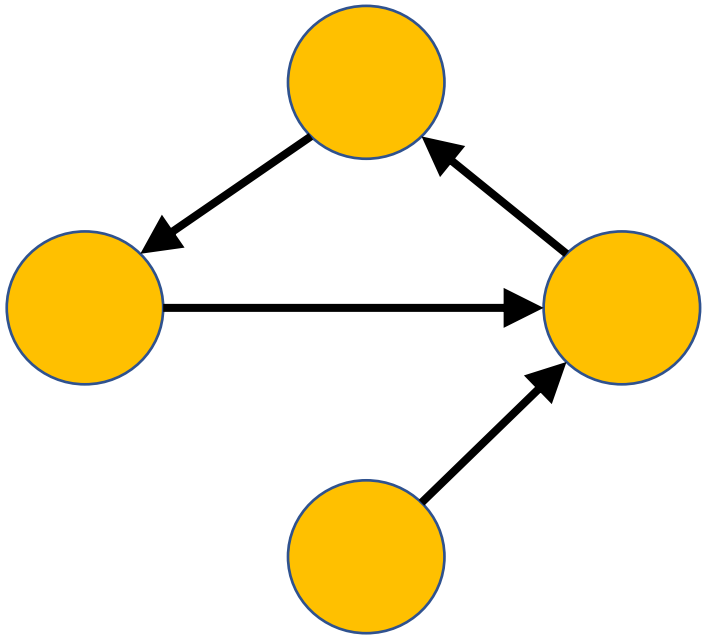
2) Moving from trickle to diffusion did not help.

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

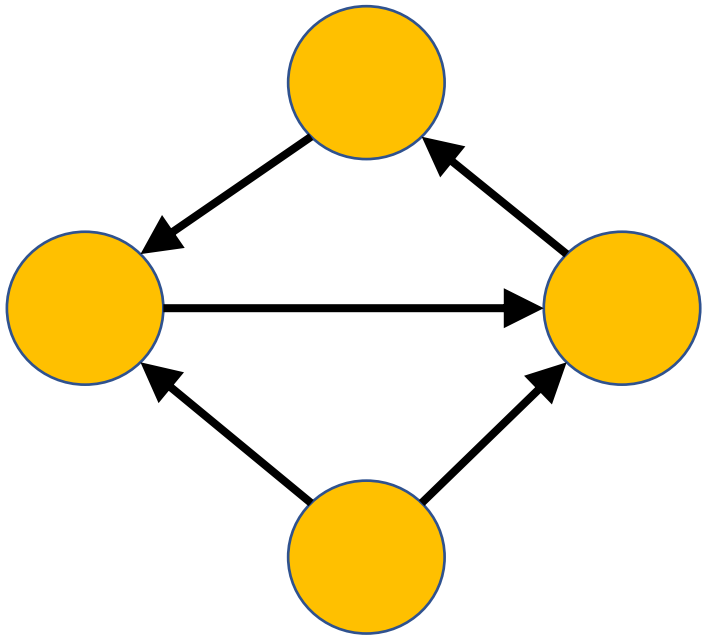
https://github.com/dandelion-org/bitcoin

BIP 156

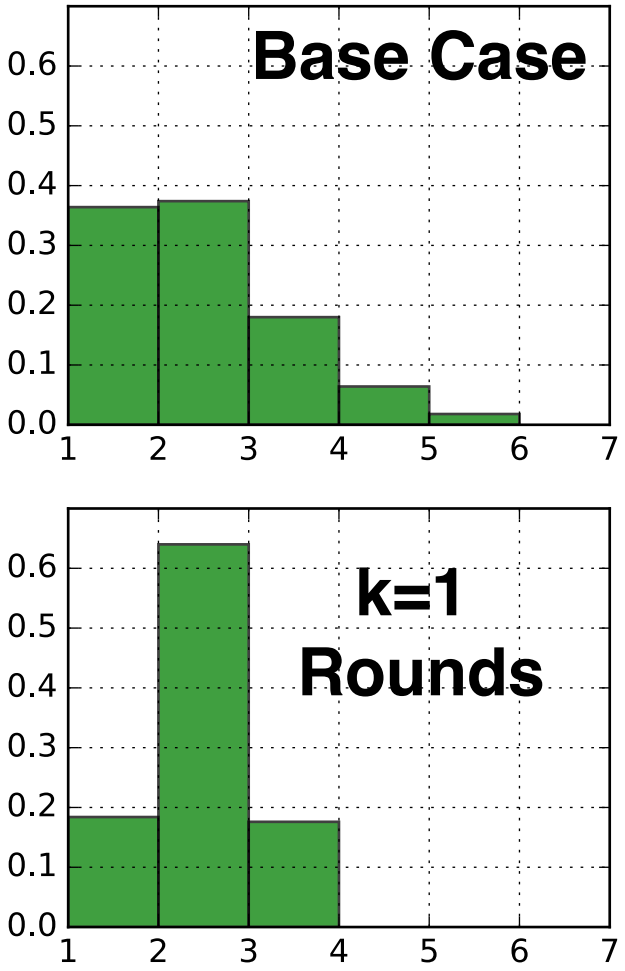
Anonymity graph construction



Base Case



k=1 rounds of Degree-Checking

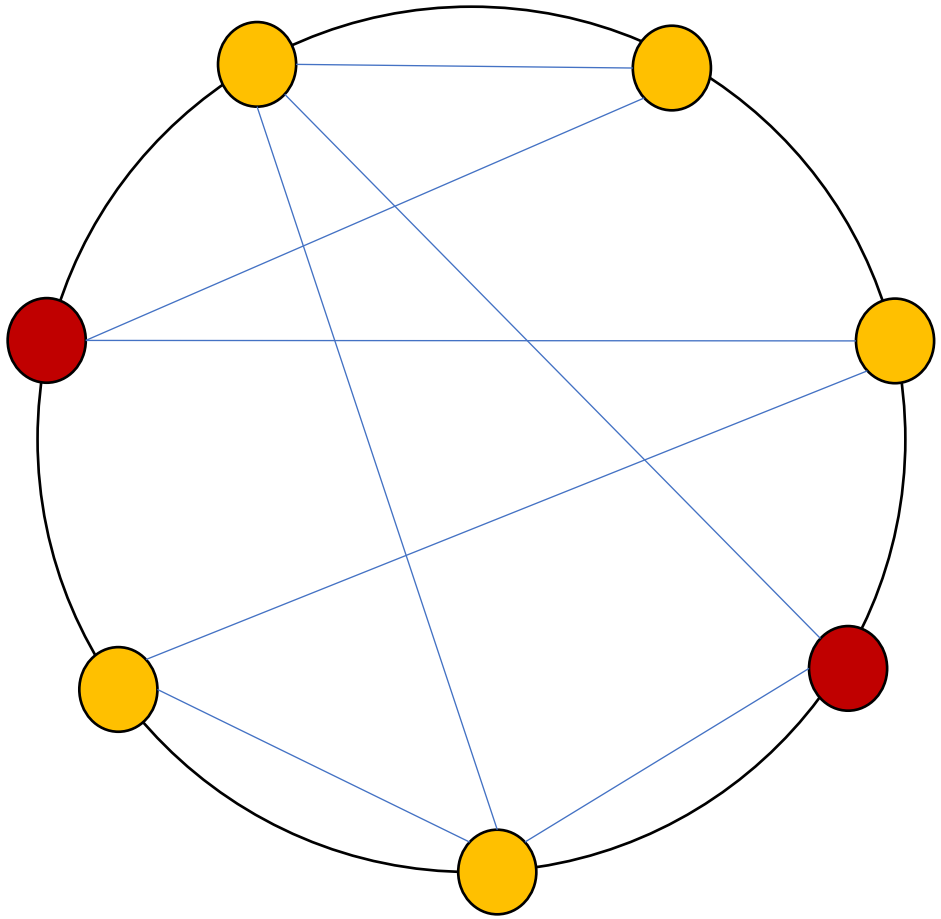


Degree

Dealing with stronger adversaries

Learn the graph	Misbehave during graph construction	Misbehave during propagation
⬇	⬇	⬇
4-regular graphs	Get rid of degree-checking	Multiple nodes diffuse

DANDELION vs. Tor, Crowds, etc.



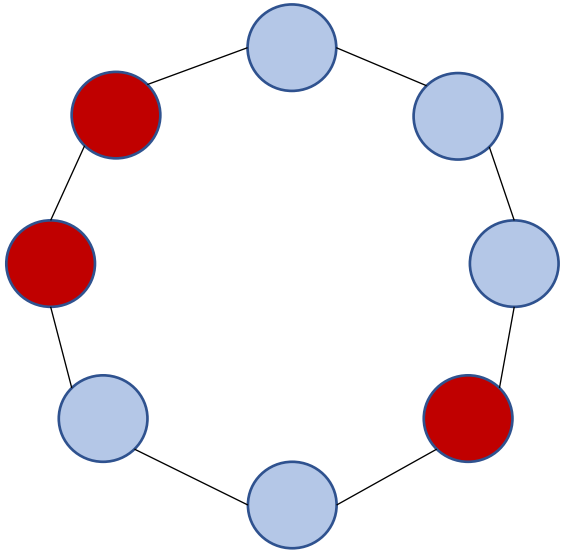
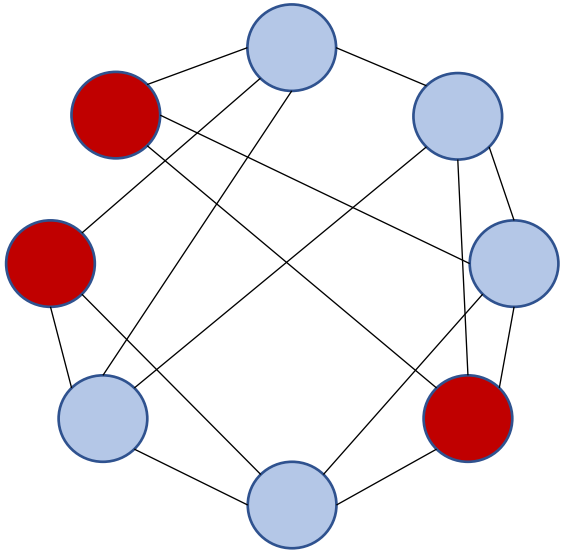
1) Messages propagate over the **same** cycle graph

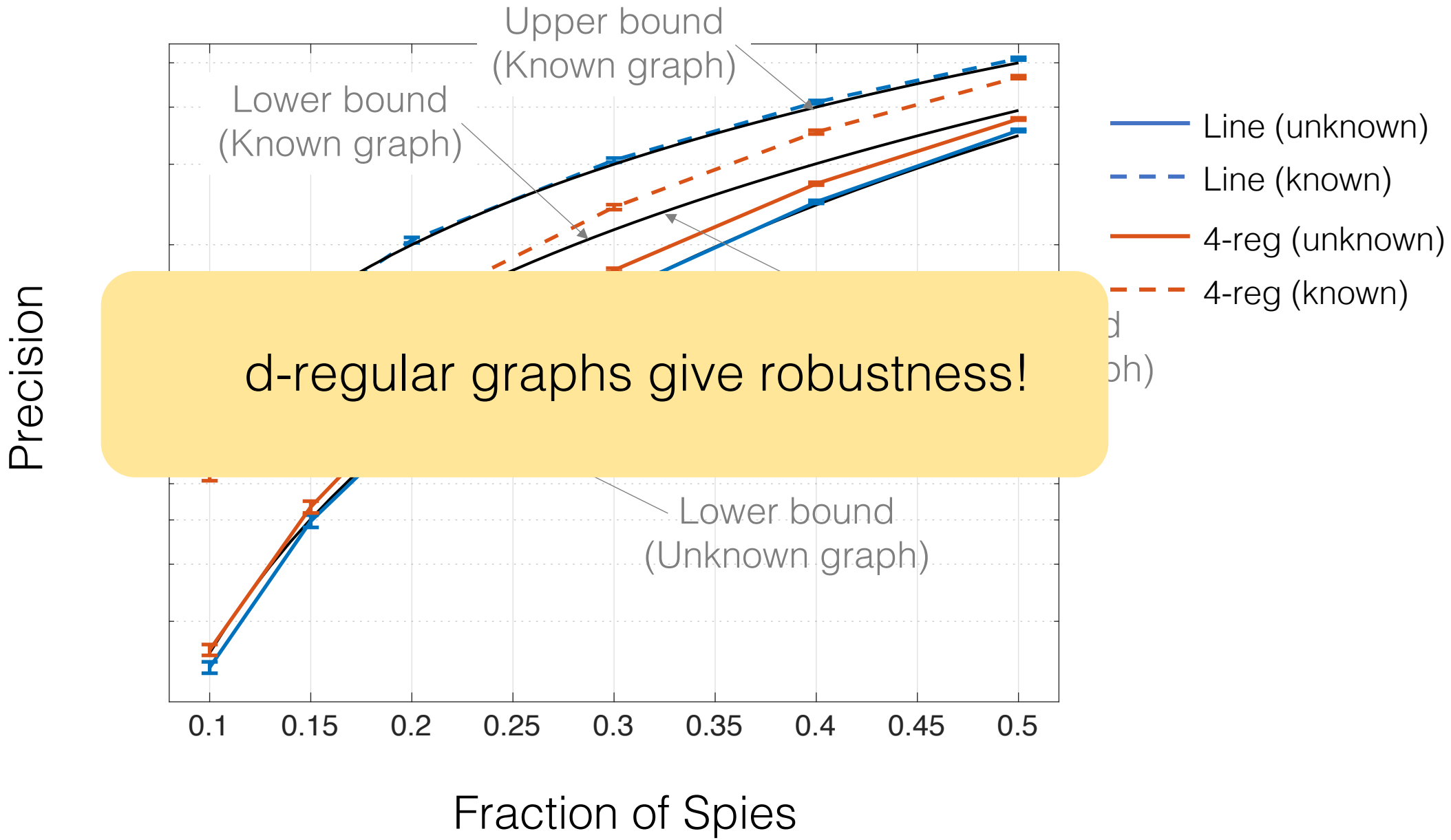
2) Anonymity graph changes dynamically.

3) No encryption required.

Learning the anonymity graph

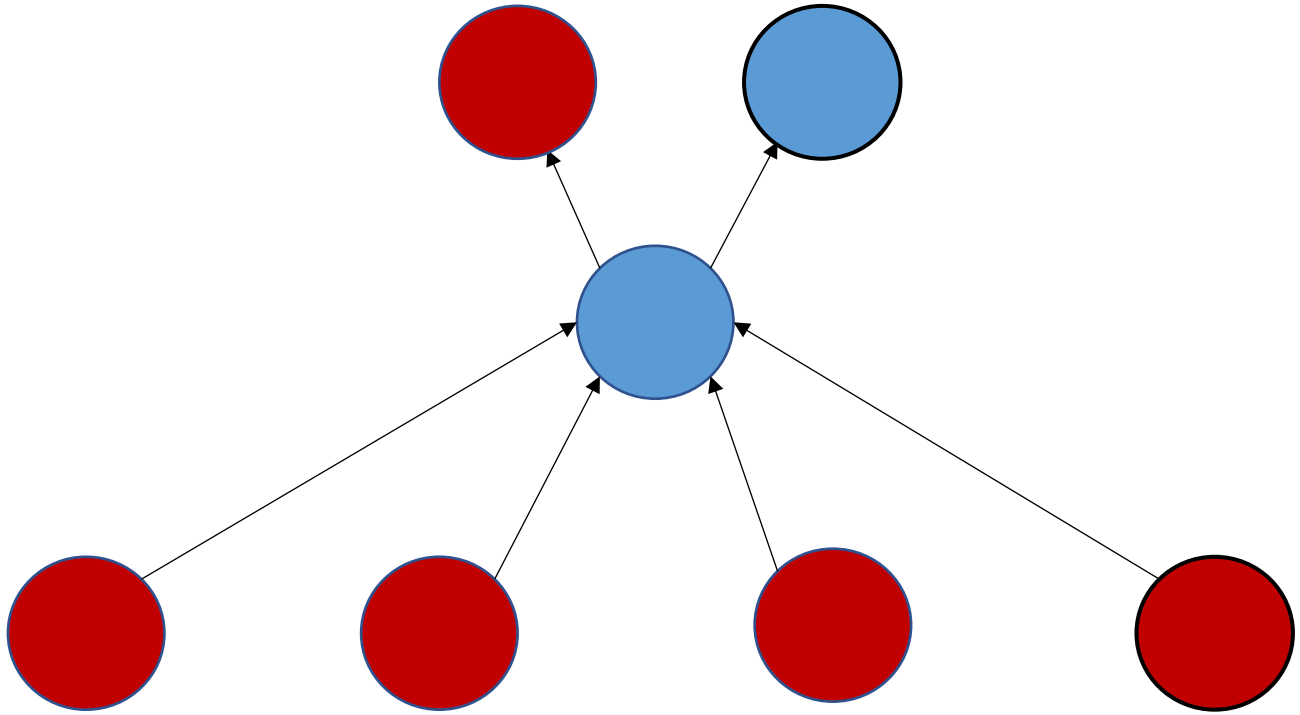
Precision

	Line	Random regular
		
Graph unknown	$O\left(\mathrm{p}^2 \log\left(\frac{1}{p}\right)\right)$	?
Graph known	$\Omega(p)$	



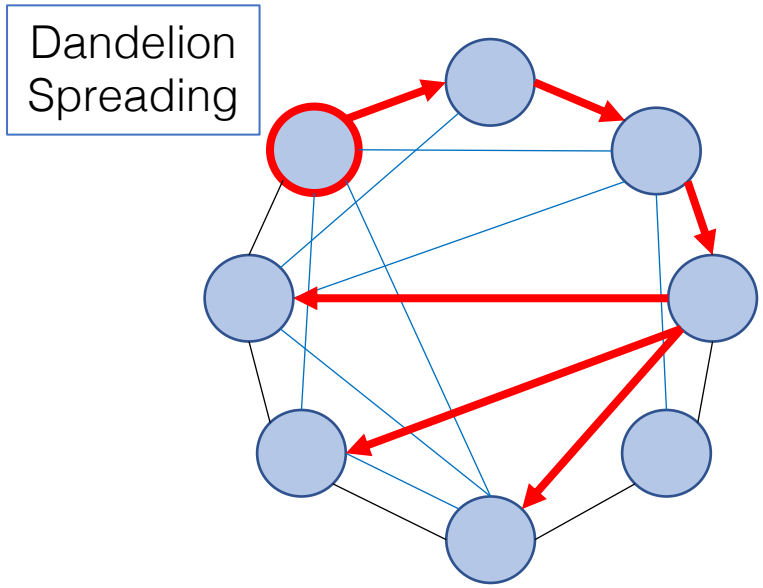
d-regular graphs give robustness!

Manipulating the anonymity graph



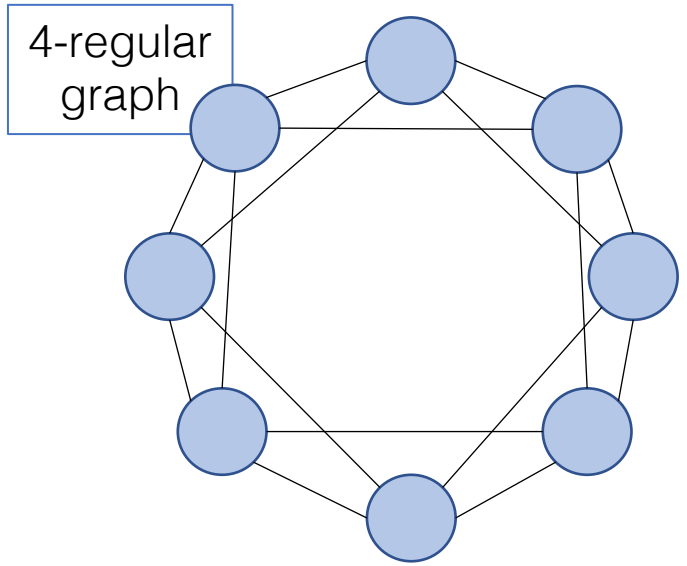
DANDELION++ Network Policy

Spreading Protocol



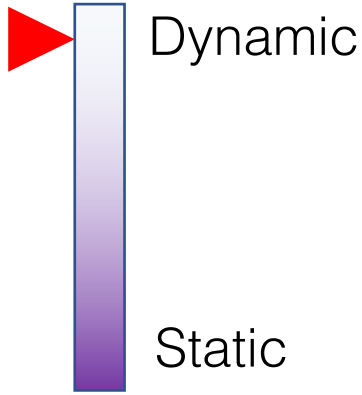
Given a graph, how do we spread content?

Topology



What is the anonymity graph topology?

Dynamicity



How often does the graph change?