On the Anonymity of Peer-To-Peer Network Anonymity Schemes Used by Cryptocurrencies

Piyush Kumar Sharma, Devashish Gosain & Claudia Diaz

KU LEUVEN

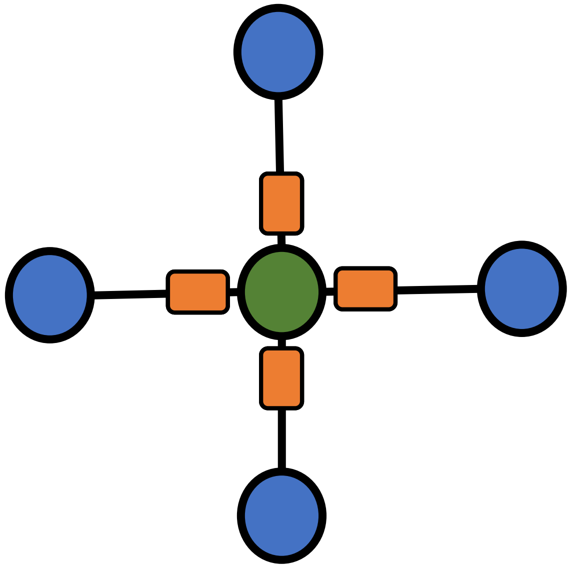
max planck institut informatik

NYM

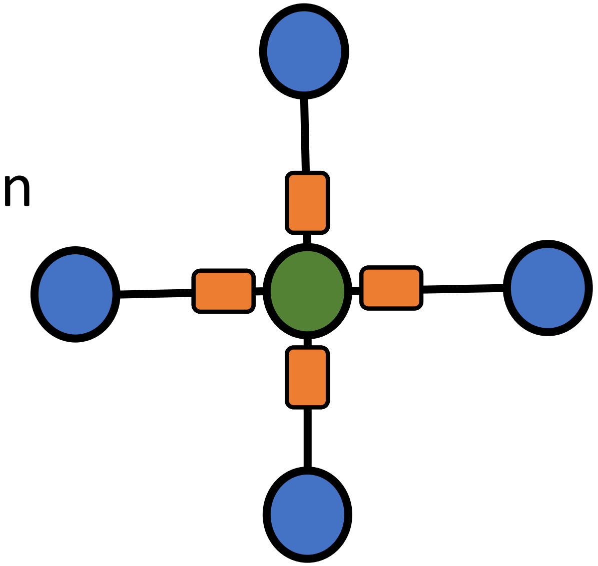
Background

- Cryptocurrency systems function broadly at two layers:
  - **Application layer** is responsible for **performing transactions**, mining blocks etc.
  - **Network Layer** (a distributed p2p network of nodes) is responsible for **broadcasting transactions**, node discovery etc.
- Anonymity required both on the chain and in the p2p network
- Interested in the network layer anonymity of such systems
  - Idea is to map the network identity (e.g., IP address) of nodes to transactions observed in the network
  - The routing algorithms can leak information about the originators (or receivers)

P2P routing changes (Bitcoin)



- Default mechanism to broadcast transaction: **flood** (before 2015)
  - Vulnerable to deanonymization attacks *[A. Biryukov et al.]*



- Mechanism updated: **diffusion** (since 2015)
  - Wait for a random time before broadcast
  - Vulnerable to attacks that analyse the symmetry of transaction propagation in the p2p network *[G. Fanti & P. Viswanath]*

Newer approaches

- To facilitate network layer anonymization, new anonymity enhancing schemes were proposed

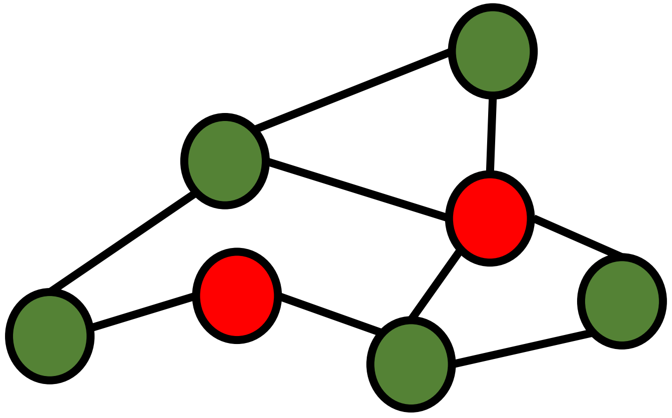
- Hop by hop routing:
  - Dandelion (2017) *[S. Venkatakrishnan et al.]*
  - Dandelion++ (2018) *[G. Fanti et al.]*

- Source routing:
  - Lightning Network (proposed:2015 live:2017) *[J. Poon and T. Dryja]*

Objective

- Research goal: To develop a generic framework to measure the anonymity of such p2p systems

- Adversary model: Passive adversary that controls a fraction of nodes in the p2p network

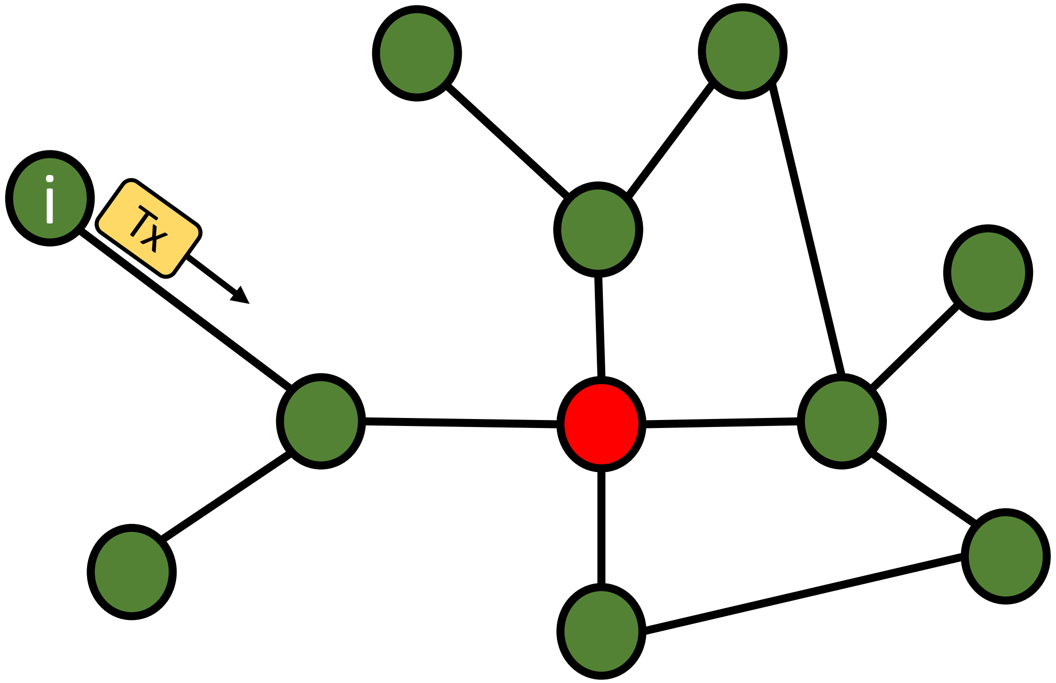


Approach

- High-level idea:
  - **Input:** network, routing scheme and observations from the network
  - **Output:** anonymity set of the transactions seen by the adversary
- Use a Bayesian framework to model anonymity

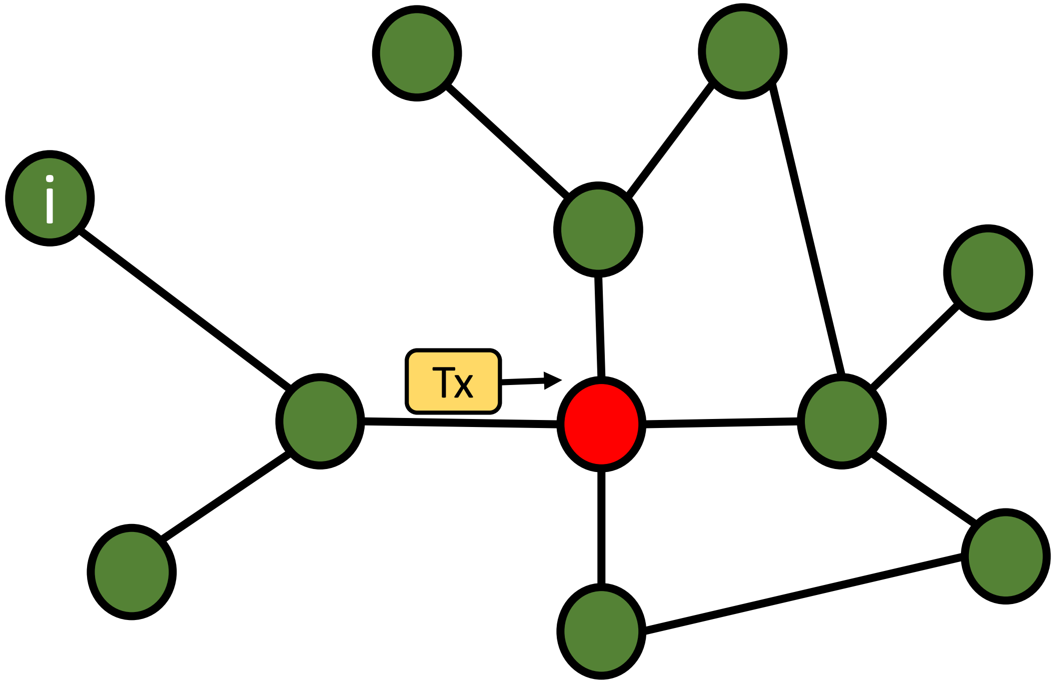
Bayesian framework

- B_i: event that a benign node ‘i’ generates a transaction (Tx)



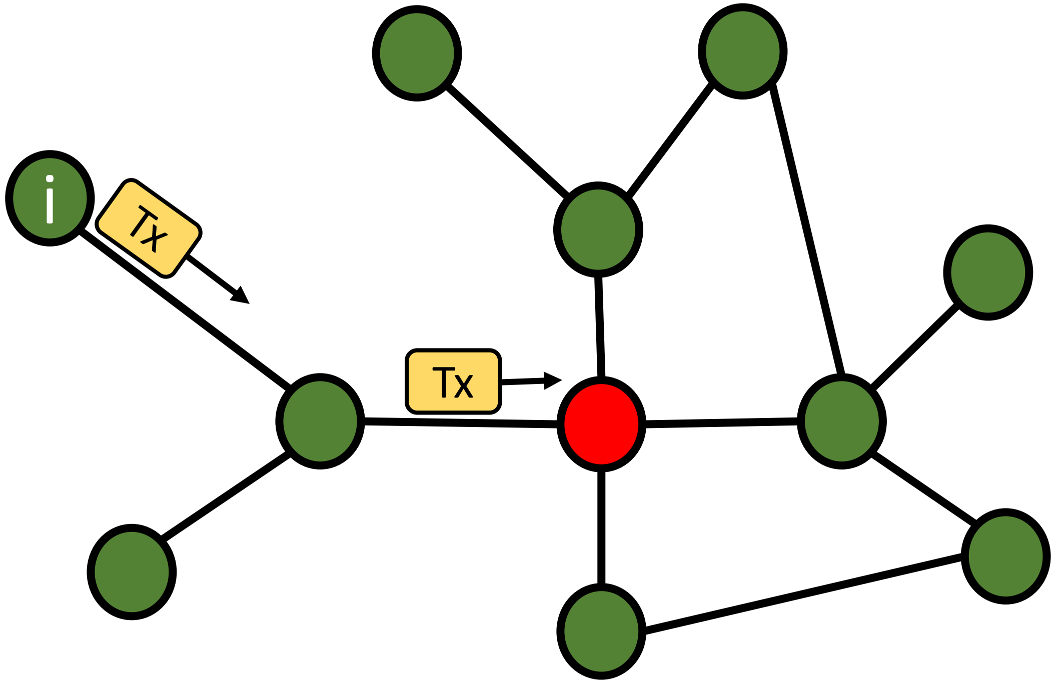
Bayesian framework

- A: event that an adversary node receives a Tx



Bayesian framework

- **GOAL: To compute P(B_i|A)**
  - Prob. of node 'i' generating a message given that an adversary node received it



Bayesian framework

- $\mathbf{P(B_i | A) = \frac{P(B_i) * P(A|B_i)}{P(A)}}$, [for all i]
  - $P(B_i)$ = Prob. that a benign node 'i' generated a Tx
  - P(A) = Prob. that an adversary node receives a Tx
  - $P(A|B_i)$ = Prob. of an adversary node receiving a Tx given that node 'i' generated it

Approach

- Let N = total nodes, C = number of adversary nodes
- $P(B_i) = 1/(N\text{-}C)$
- $P(A) = \sum_{i=1}^{N-C} P(Bi) * P(A|B_i)$
- $P(B_i|A) = \frac{P(B_i) * P(A|B_i)}{P(A)} = \frac{P(B_i) * P(A|B_i)}{\sum_{k=1}^{N-C} P(B_k) * P(A|B_k)} = \frac{\mathbf{P(A|B_i)}}{\sum_{k=1}^{N-C} \mathbf{P(A|B_k)}}$

Approach

- Use entropy as a metric for anonymity

$$H = -\sum_{i=1}^{N-C} P(B_i|A) * \log_2[P(B_i|A)]$$

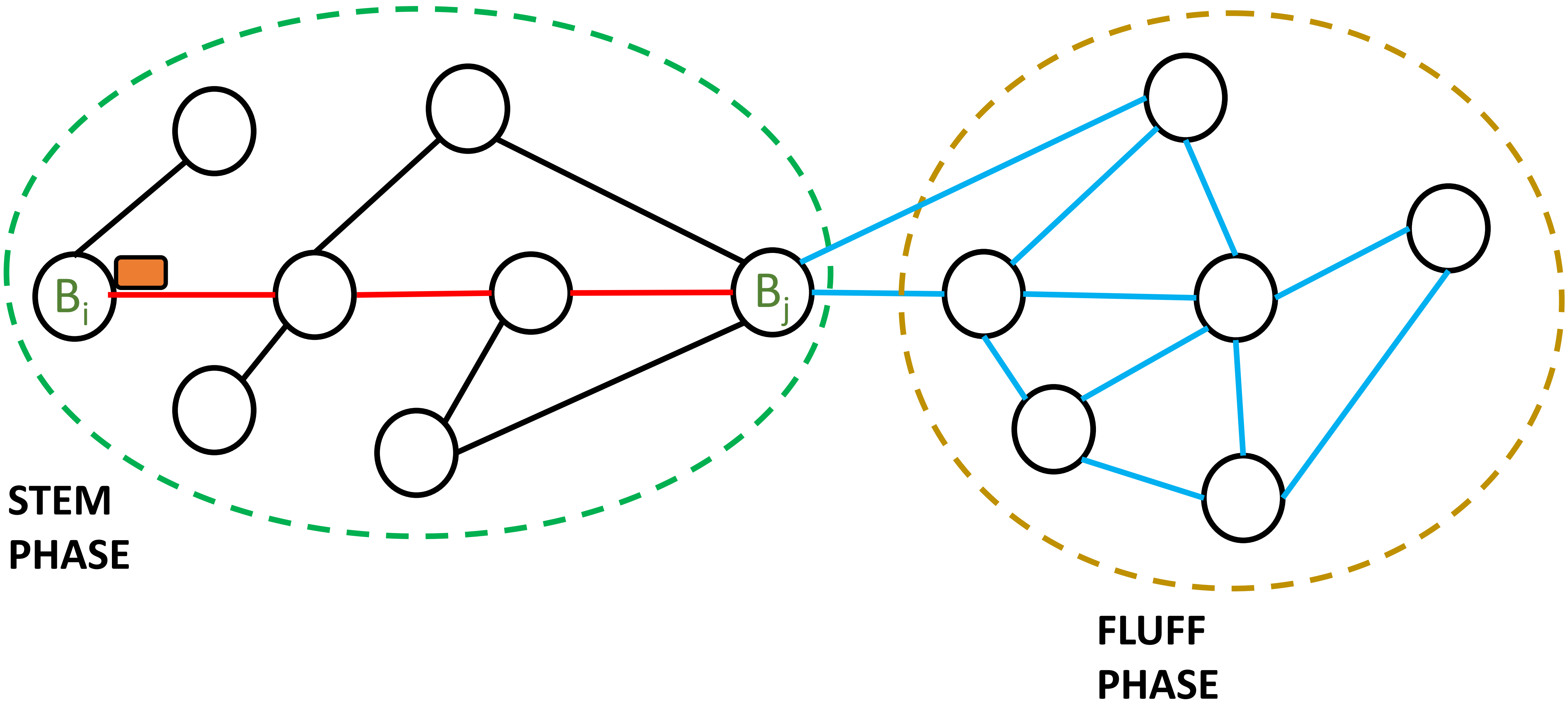
- Entropy measured in bits
  - An entropy of x bits imply an anonymity set of 2^x nodes
  - If entropy is 0 bits -> completely deanonymized

Hop by Hop

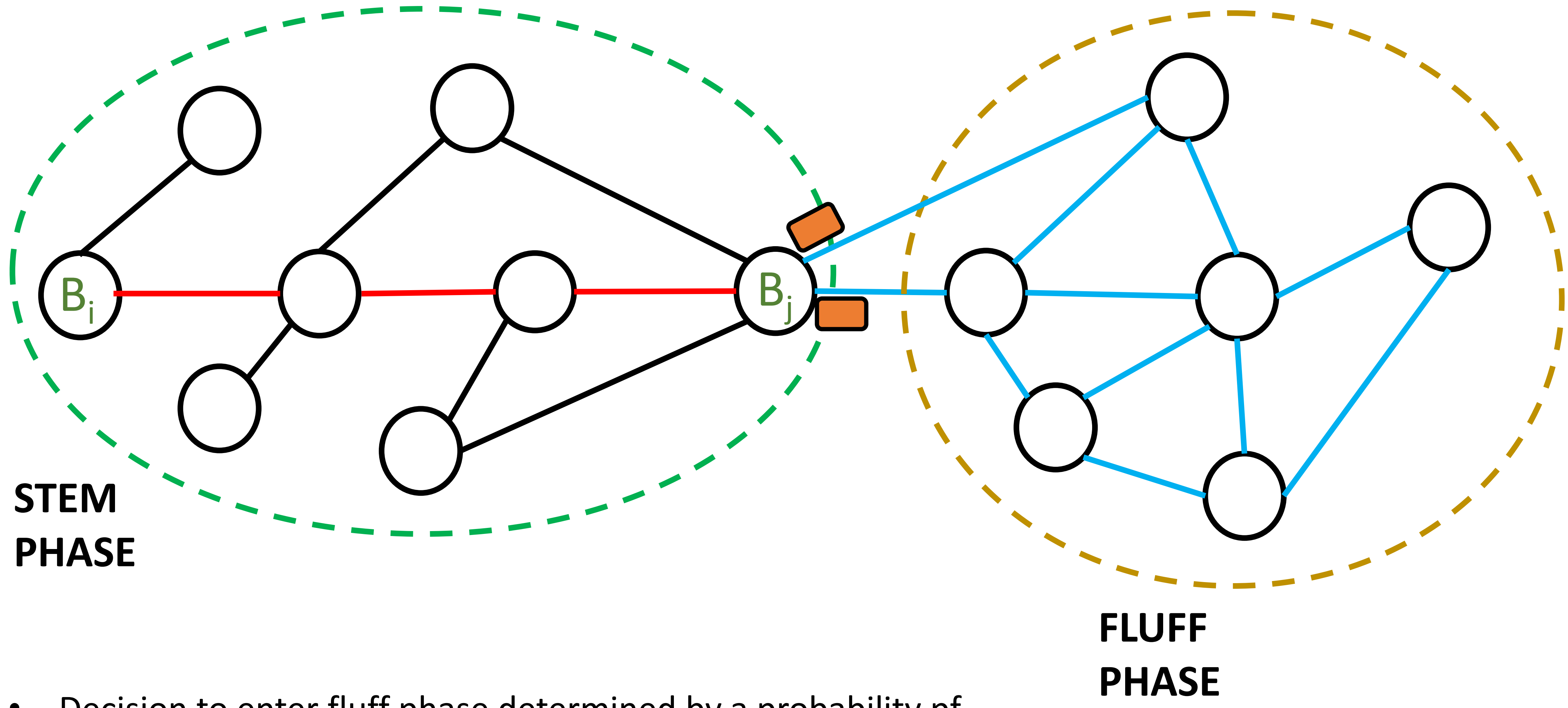
Dandelion design

- Two sets of graph:
  - Bitcoin p2p graph
  - Privacy subgraph: A line graph covering all the nodes

- Two phased operation:
  - **Stem phase**
  - **Fluff phase**



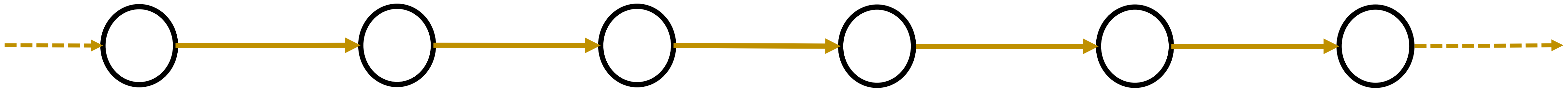
B_i
B_j
STEM PHASE
FLUFF PHASE



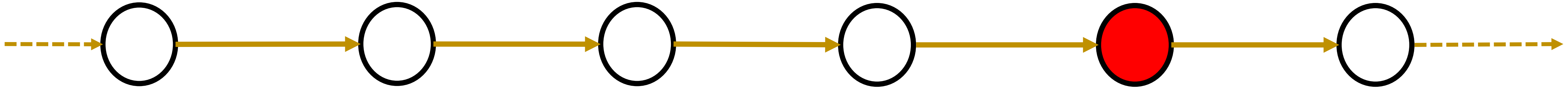
- Decision to enter fluff phase determined by a probability pf
- Anonymity provided only in the stem phase

Modelling Dandelion

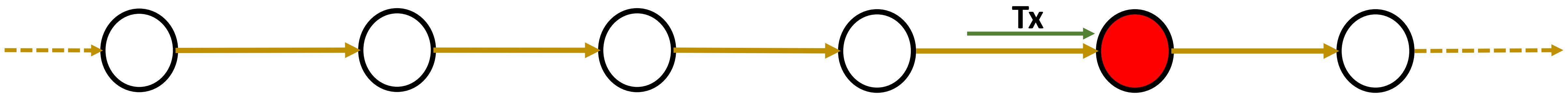
- Remember, we need to calculate the value of $P(A|B_i)$ to obtain the probability distribution



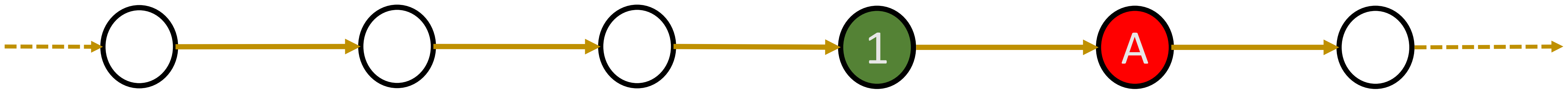
Modelling Dandelion



Modelling Dandelion

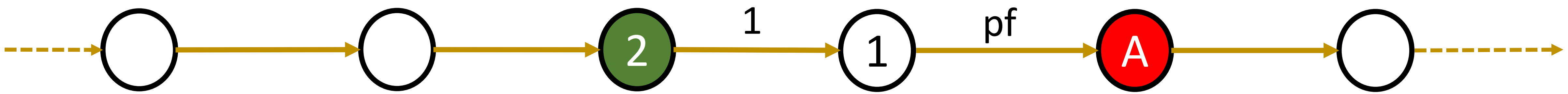


Modelling Dandelion



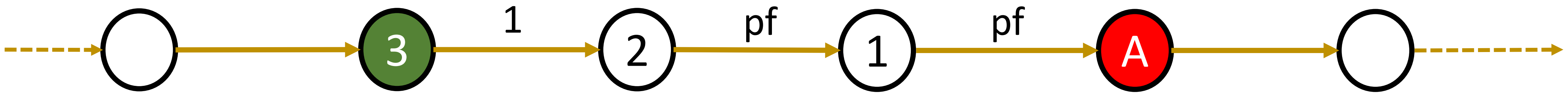
$P(A|B_1) = 1$

Modelling Dandelion



$P(A|B_2) = 1*pf$

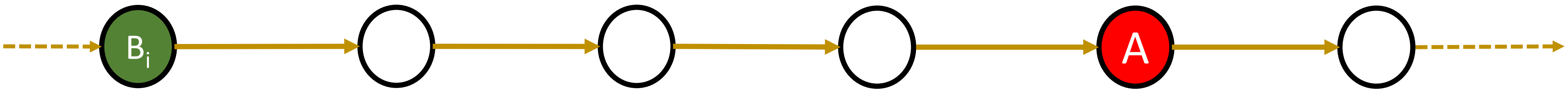
Modelling Dandelion



$P(A|B_3) = 1*pf*pf$

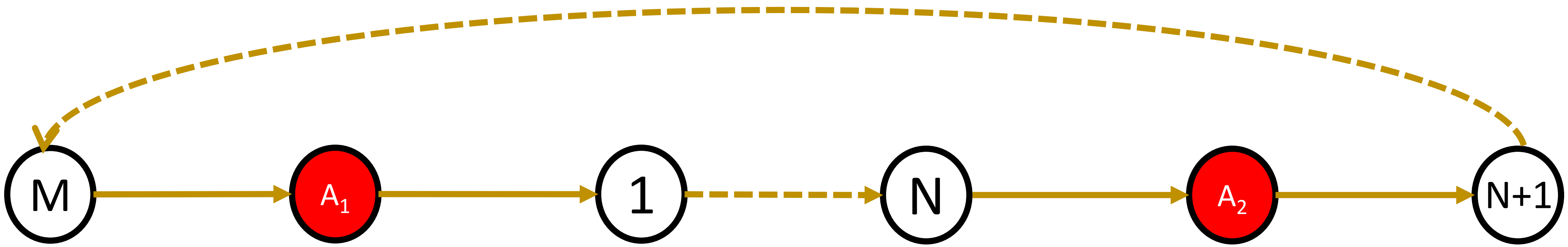
Modelling Dandelion

- To generalize, for any benign node i



- $P(A|B_i) = (pf)^{h_i-1}$
  - pf = forwarding probability
  - h_i = number of hops between benign node B_i and adversary node A

Can we do better?

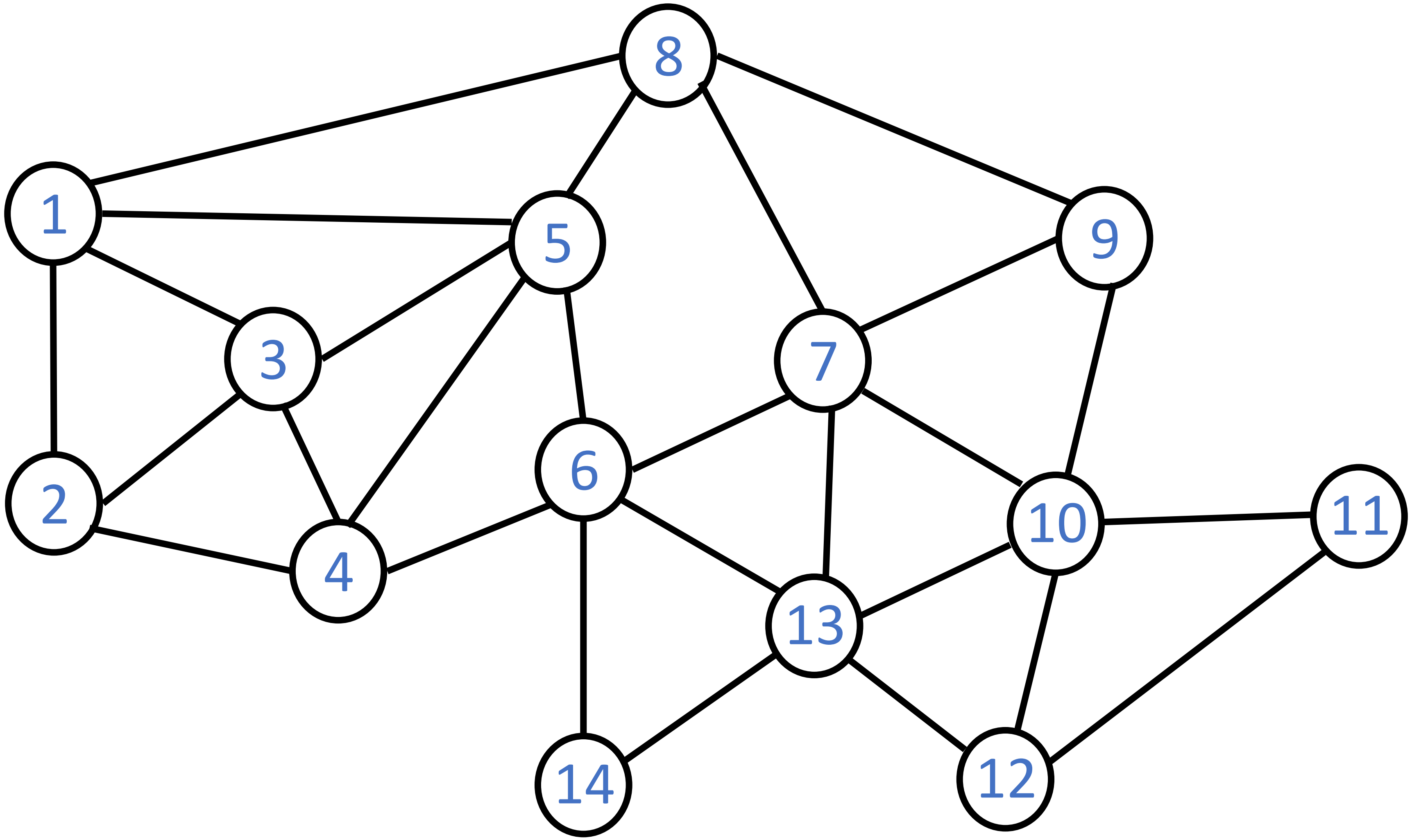


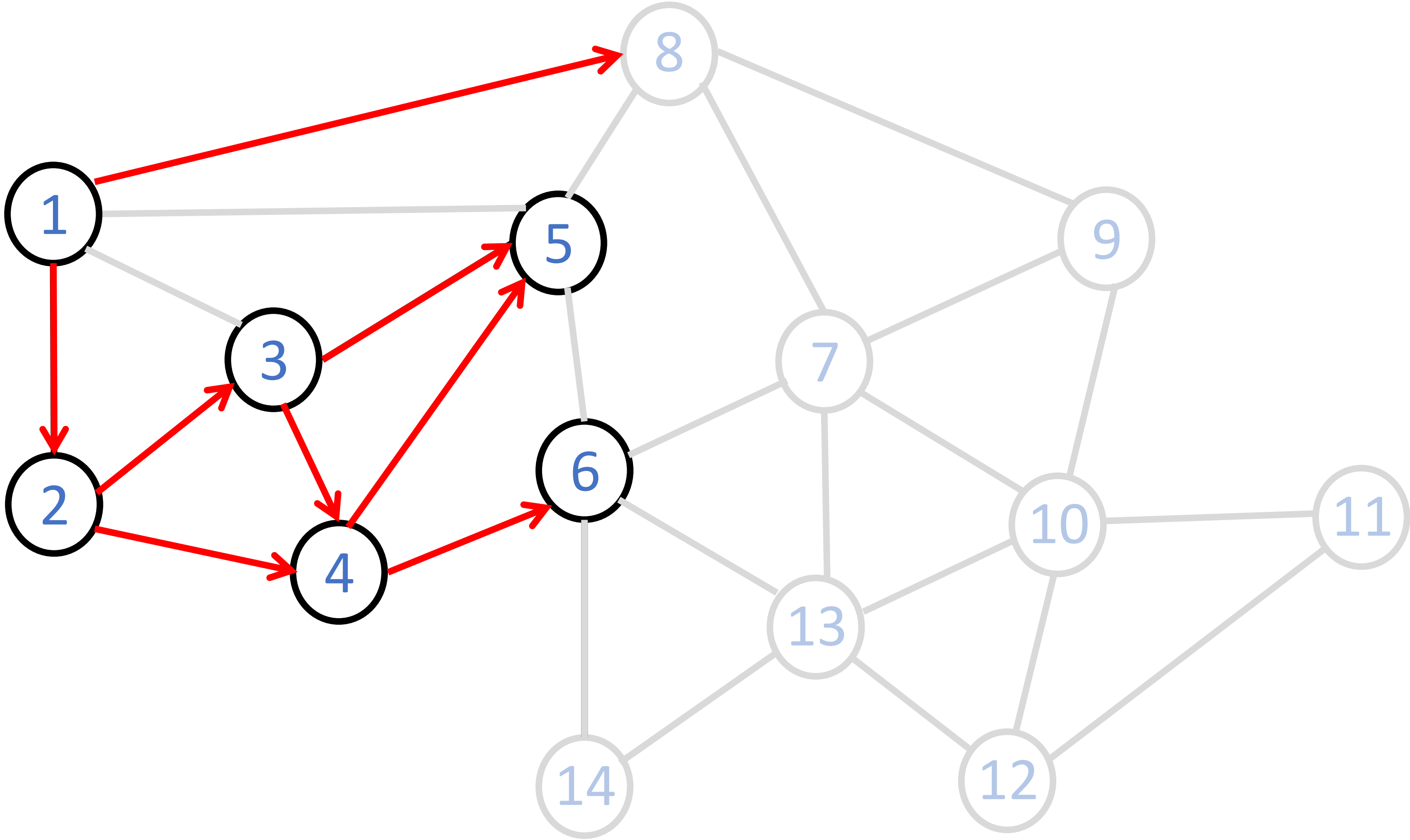
- Multiple adversaries can know their position in the privacy subgraph
- The line graph will then be divided into multiple partitions
- The potential originators will then be limited to just the partition

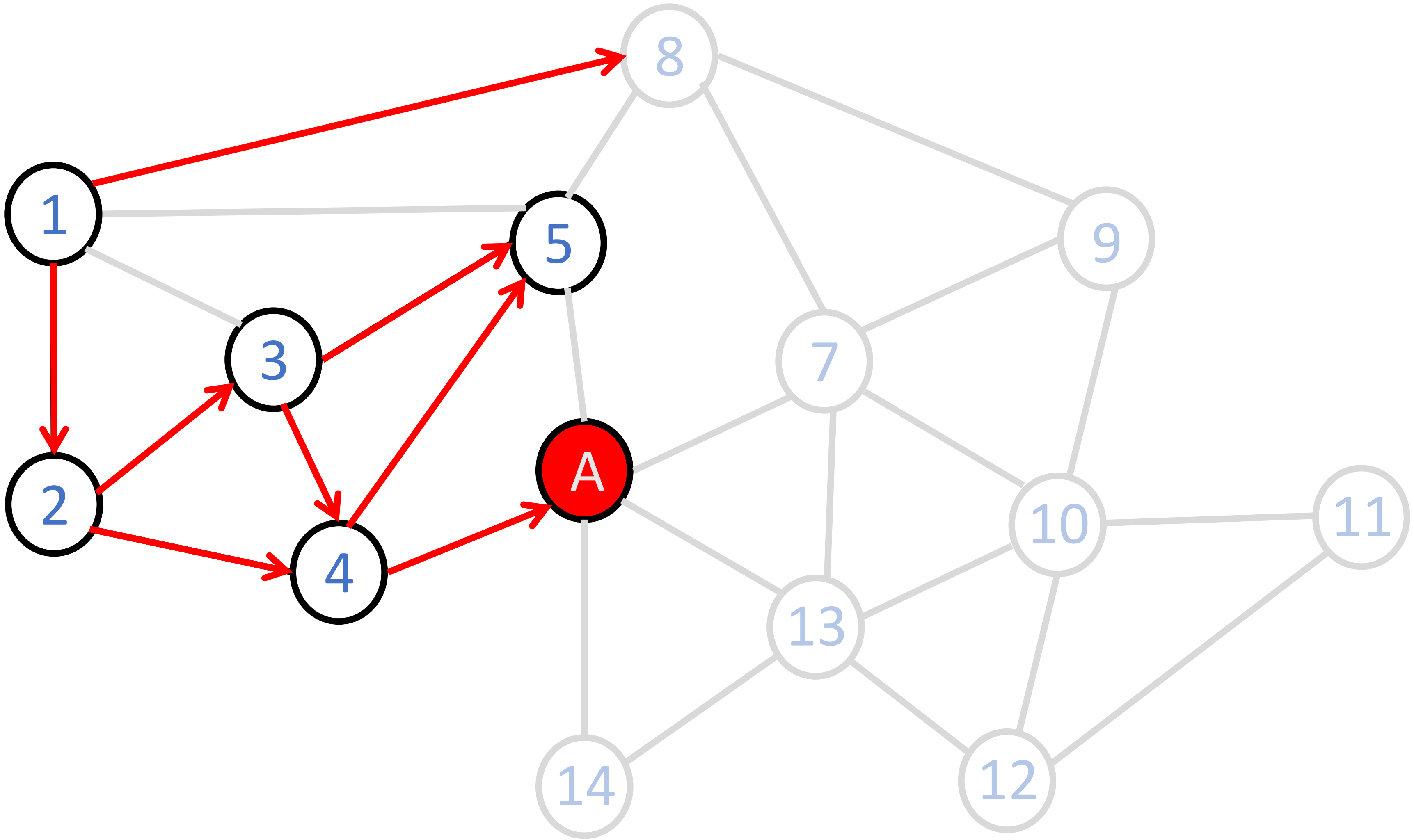
Dandelion++

- Similar to Dandelion
  - Stem phase
  - Fluff phase

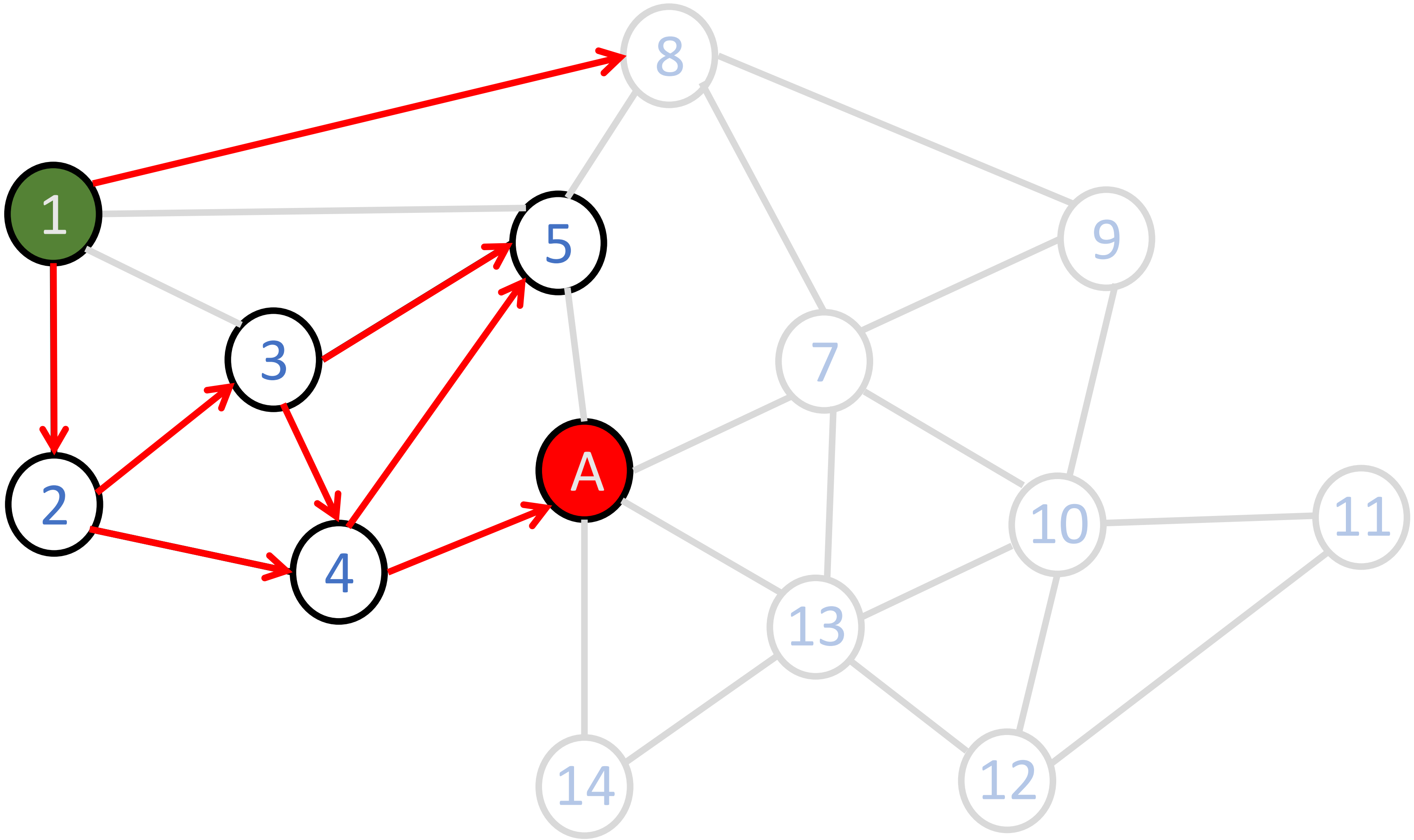
- Privacy subgraph is 4-regular instead of line graph
  - For each node outdegree will be 2

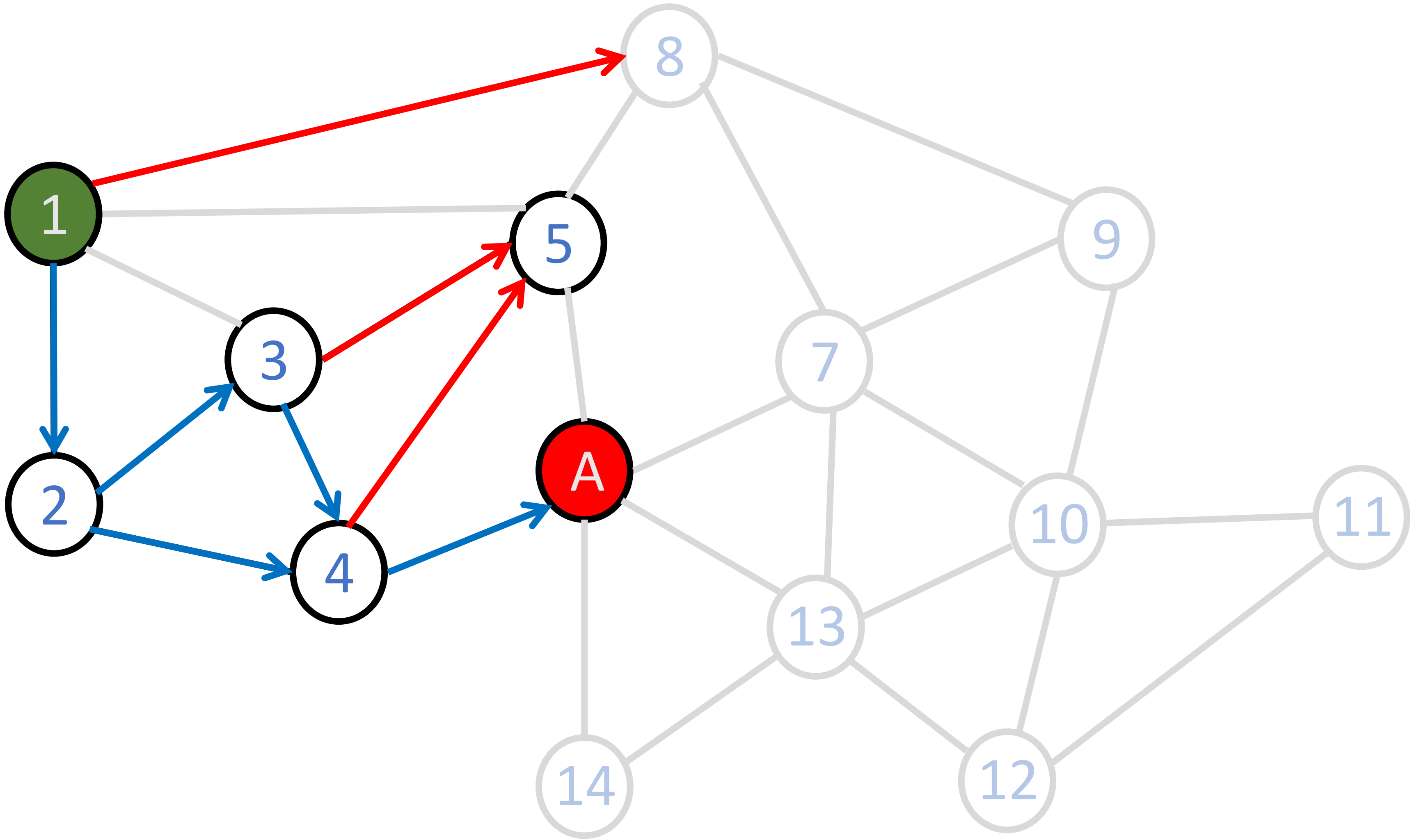

8
1
5
9
3
7
2
6
10
11
4
13
14
12


1
2
3
4
5
6
7
8
9
10
11
12
13
14

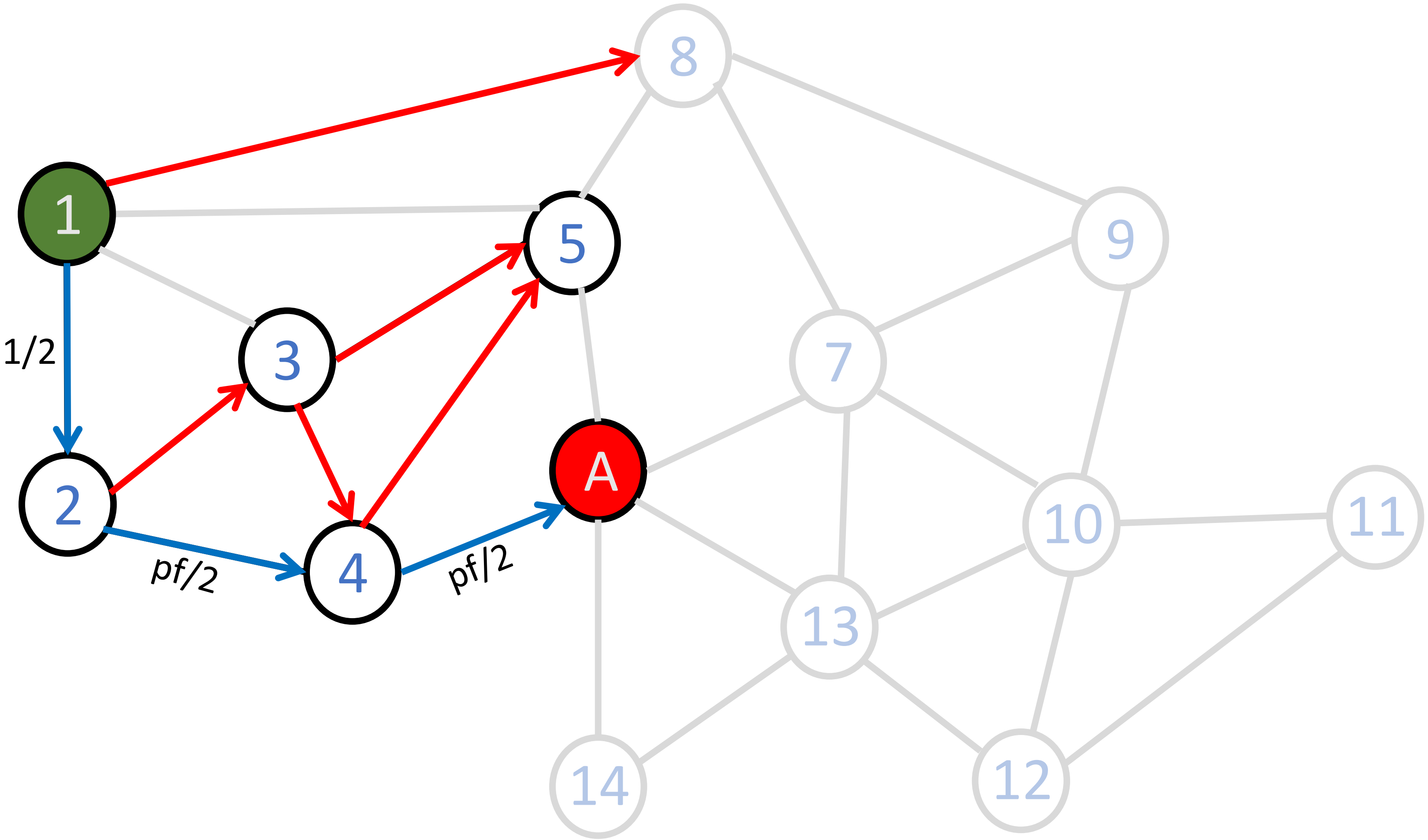


8
1
5
9
3
7
A
2
10
11
4
13
14
12

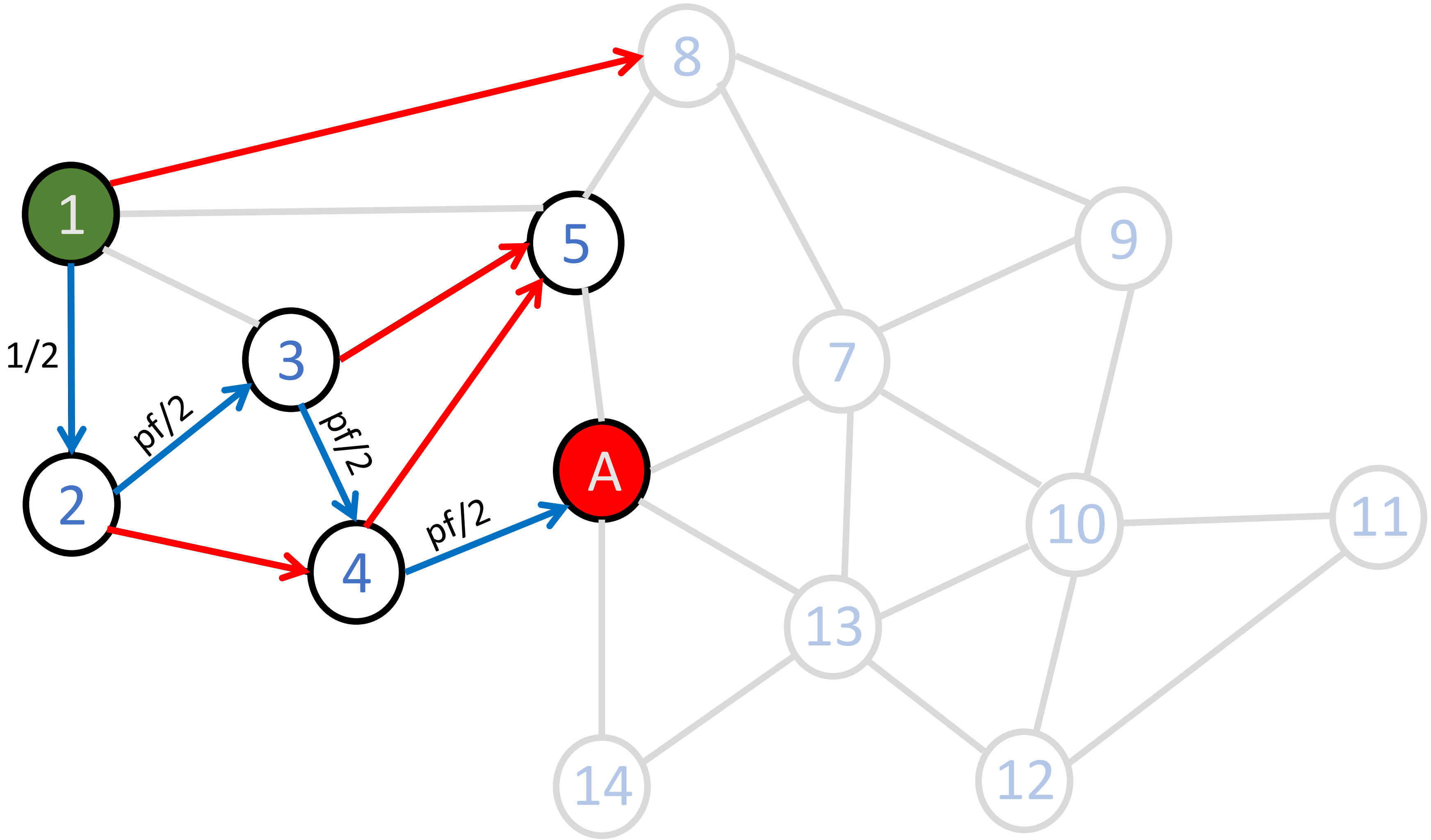

1
8
5
9
3
7
2
A
10
11
4
13
14
12



- Two paths from 1 to A:
  - 1 -> 2 -> 4 -> A
  - 1 -> 2 -> 3 -> 4-> A



- For path: 1 -> 2 -> 4 -> A
  - P = ½*(pf*1/2)*(pf*1/2)



- For path: 1 -> 2 -> 3 -> 4 -> A
  - P = ½*(pf*1/2)*(pf*1/2)*(pf*1/2)

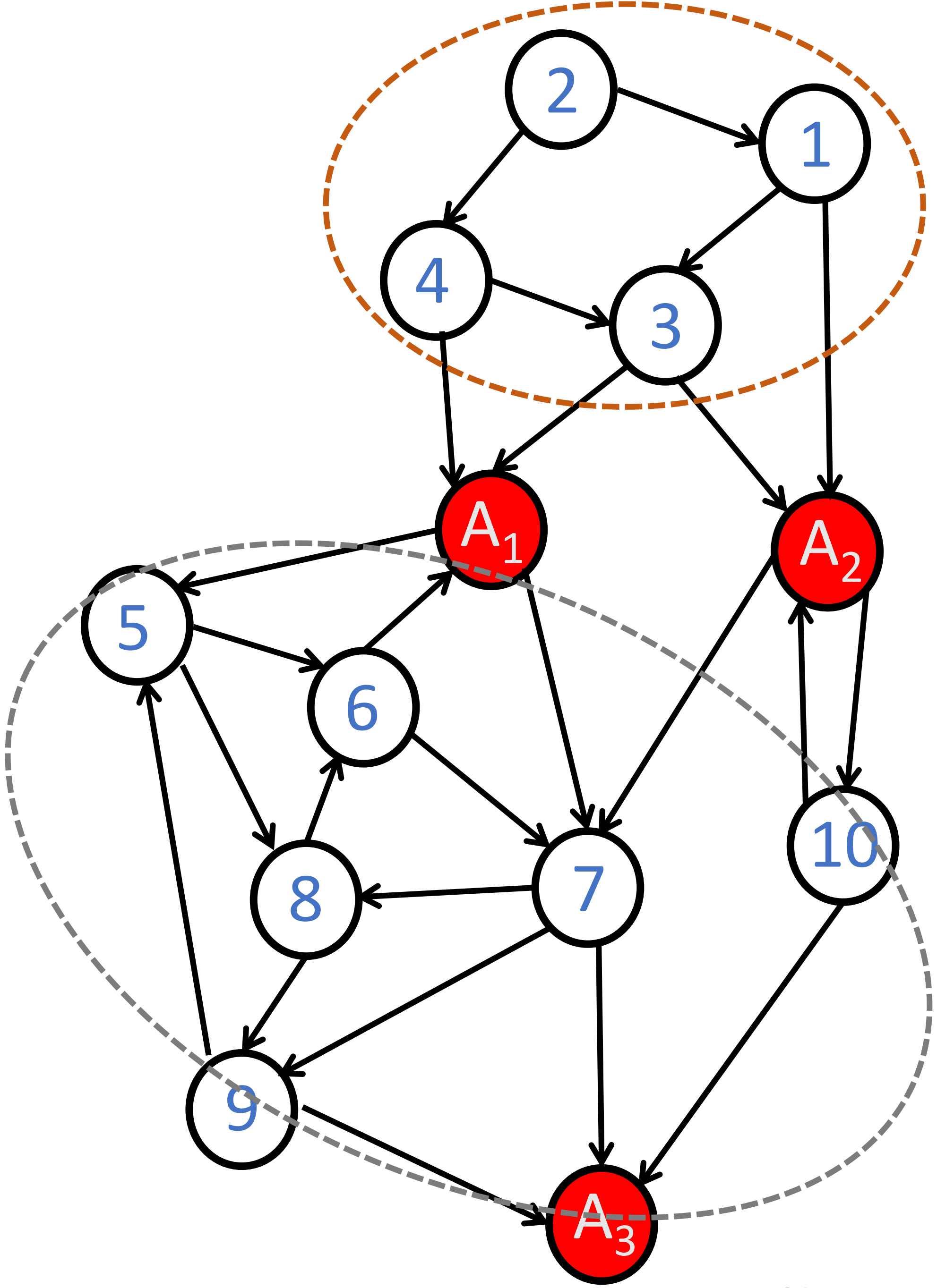
Modelling Dandelion++

- P (A|B_i) = $\sum^{Tp} \frac{1}{2} * \left(\frac{\text{pf}}{2}\right)^{hi-1}$ for each path Tp, where

- Tp = total number of paths between node i and adversary A

Can we do better?

- **Combine information** from multiple adversary nodes
- If A_3 receives a transaction not observed by A_1 and A_2, the originator set is reduced to:
  - 5-10
- Helps reducing the anonymity set significantly



Evaluation

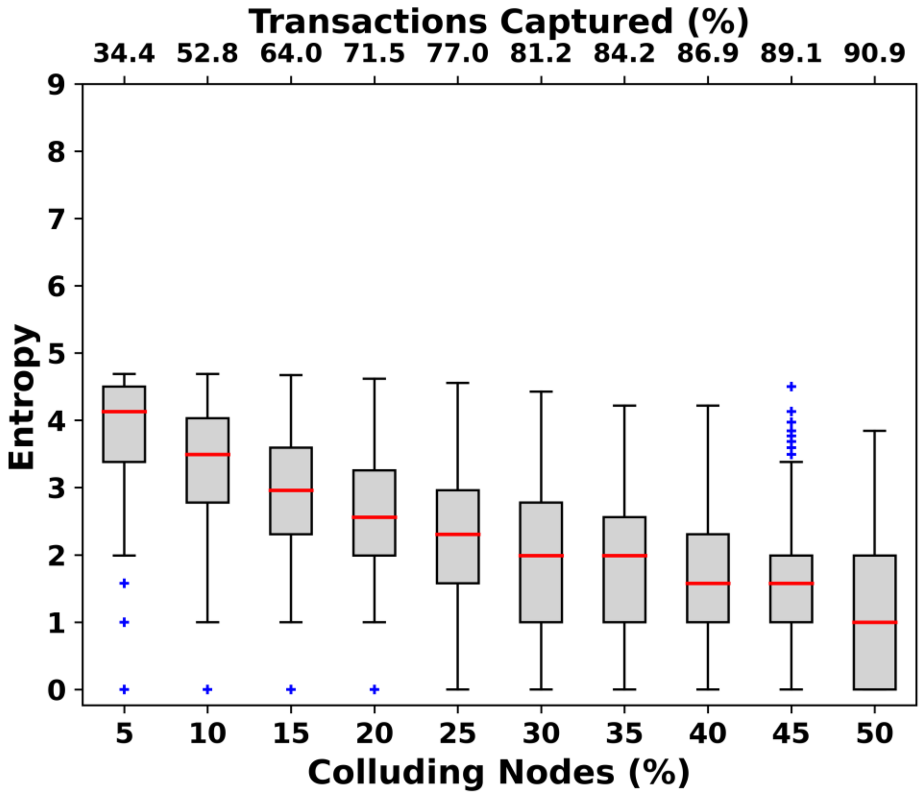
- Incorporated the modelling approach in a simulator
  - Construct p2p graph: line graphs for Dandelion and 4-regular for Dandelion++
  - Assume adversary nodes: randomly
  - Calculate probability distributions and entropy

- Measured the entropy for different parameters (pf, C, N)
  - Selected the adversary nodes randomly 1000 times

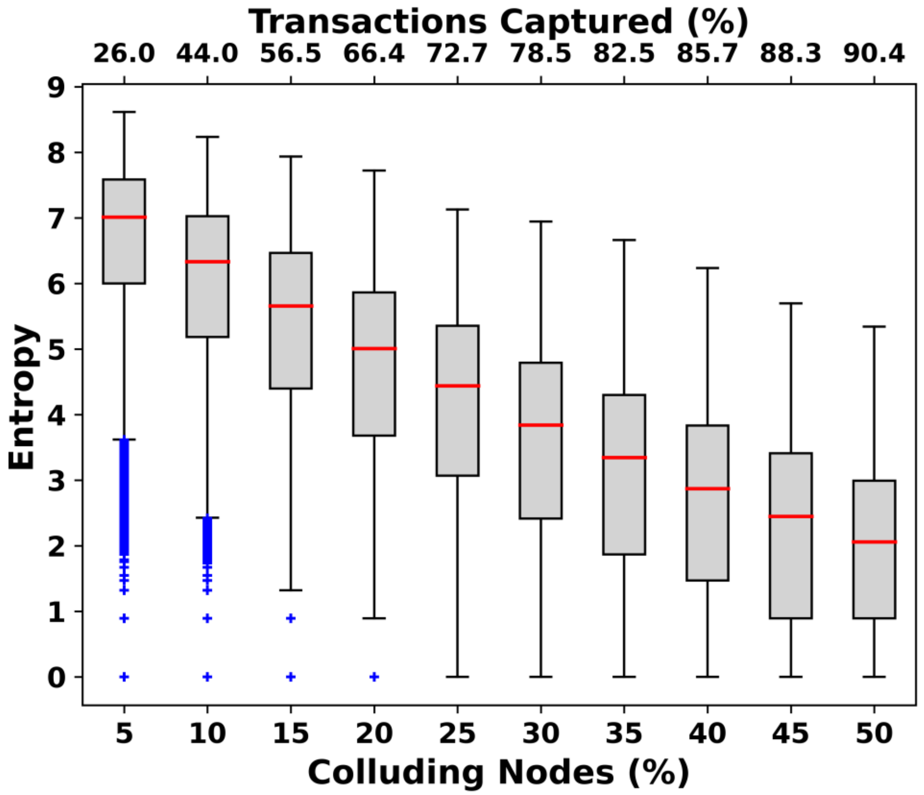
Results (hop-by-hop)

Entropy with increasing adversary nodes (pf = 0.9, N = 1000)

Dandelion



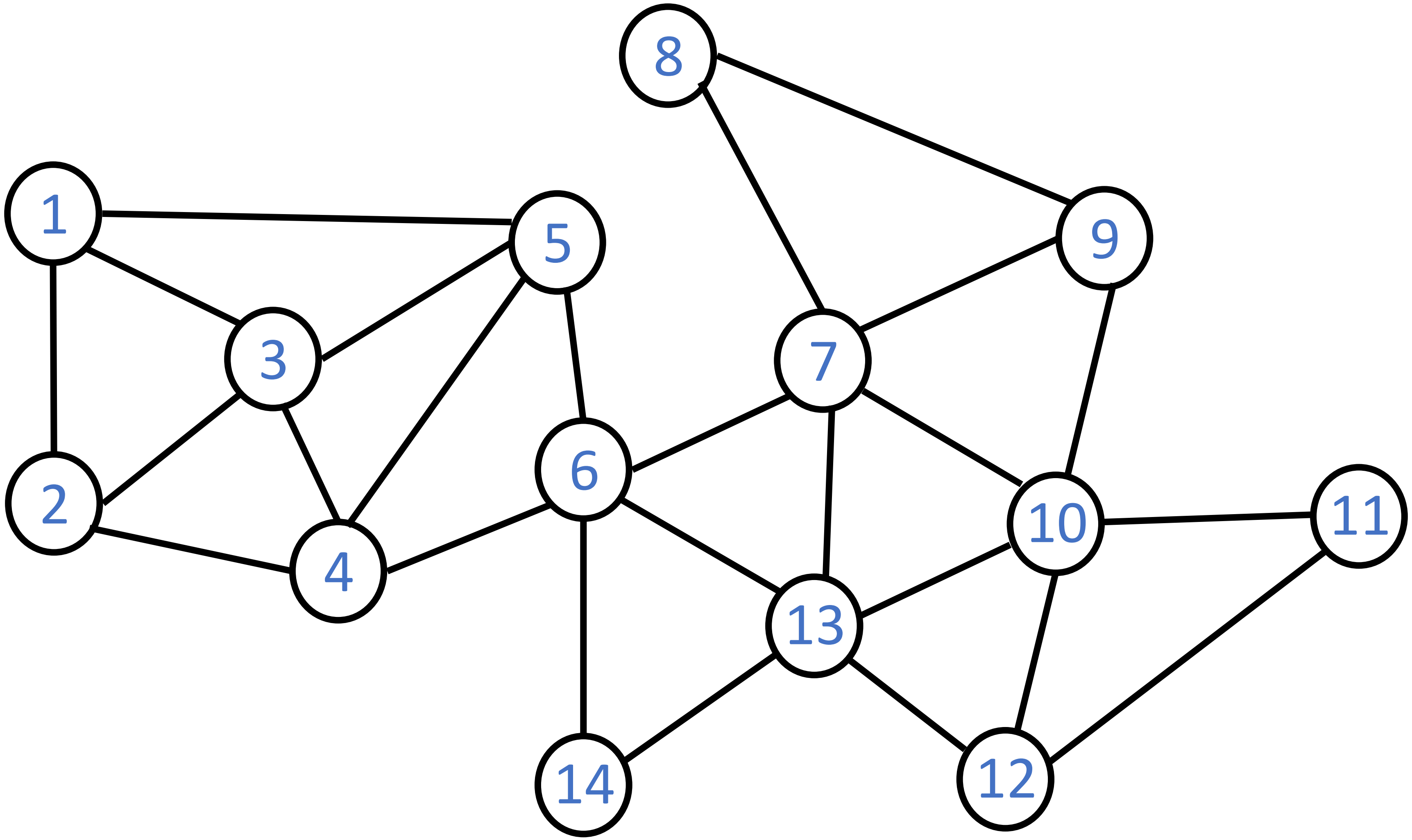
Dandelion++

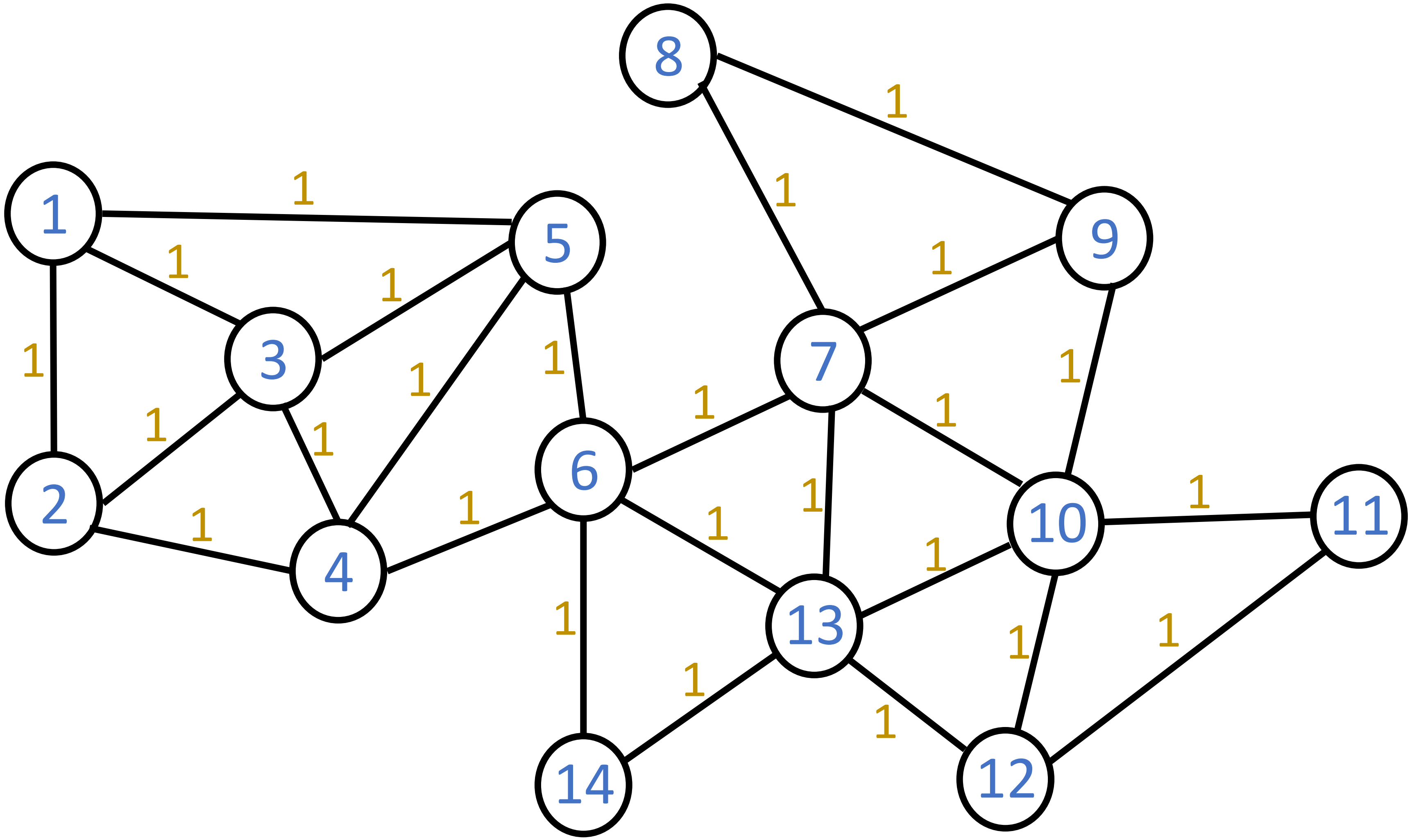


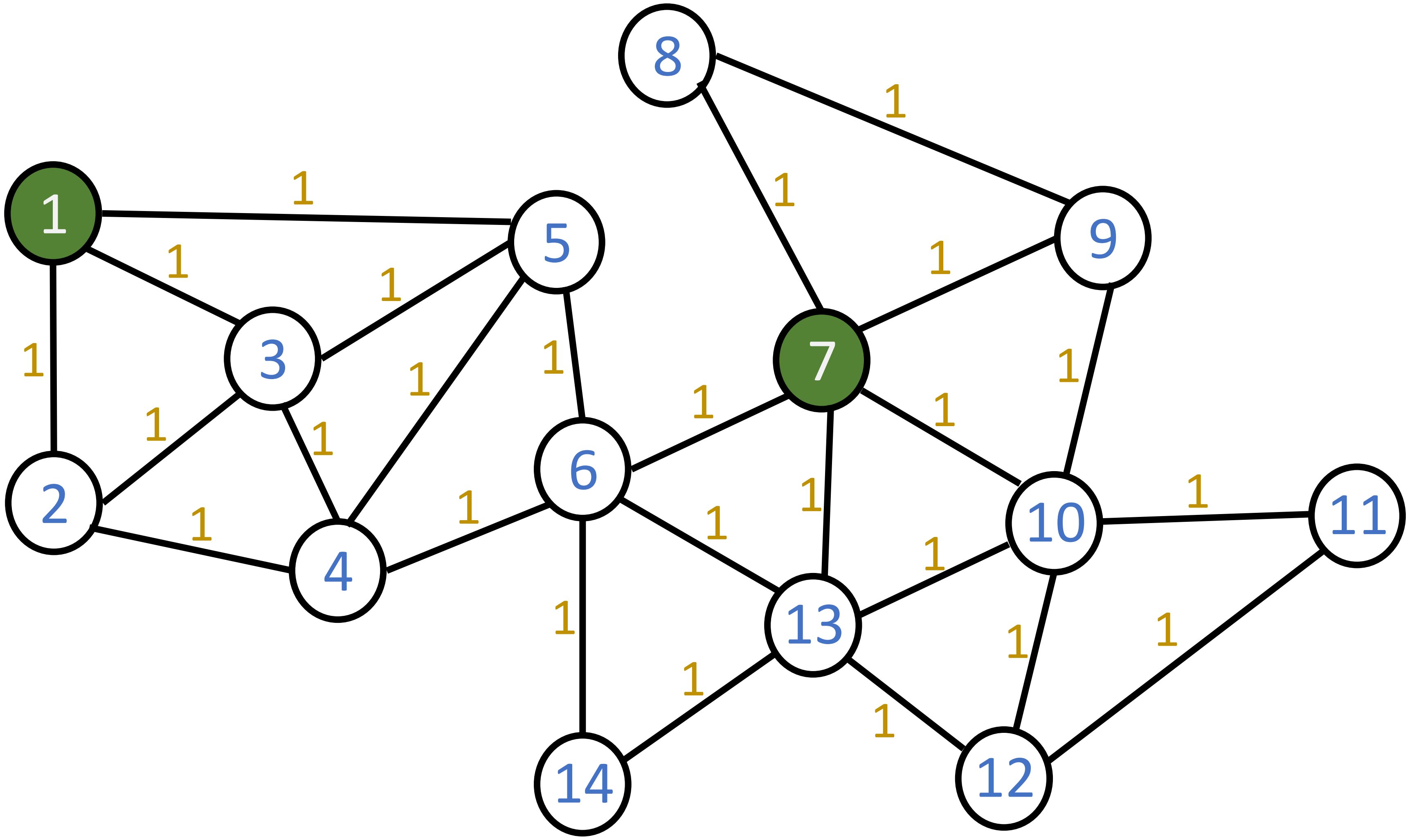
Source Routing

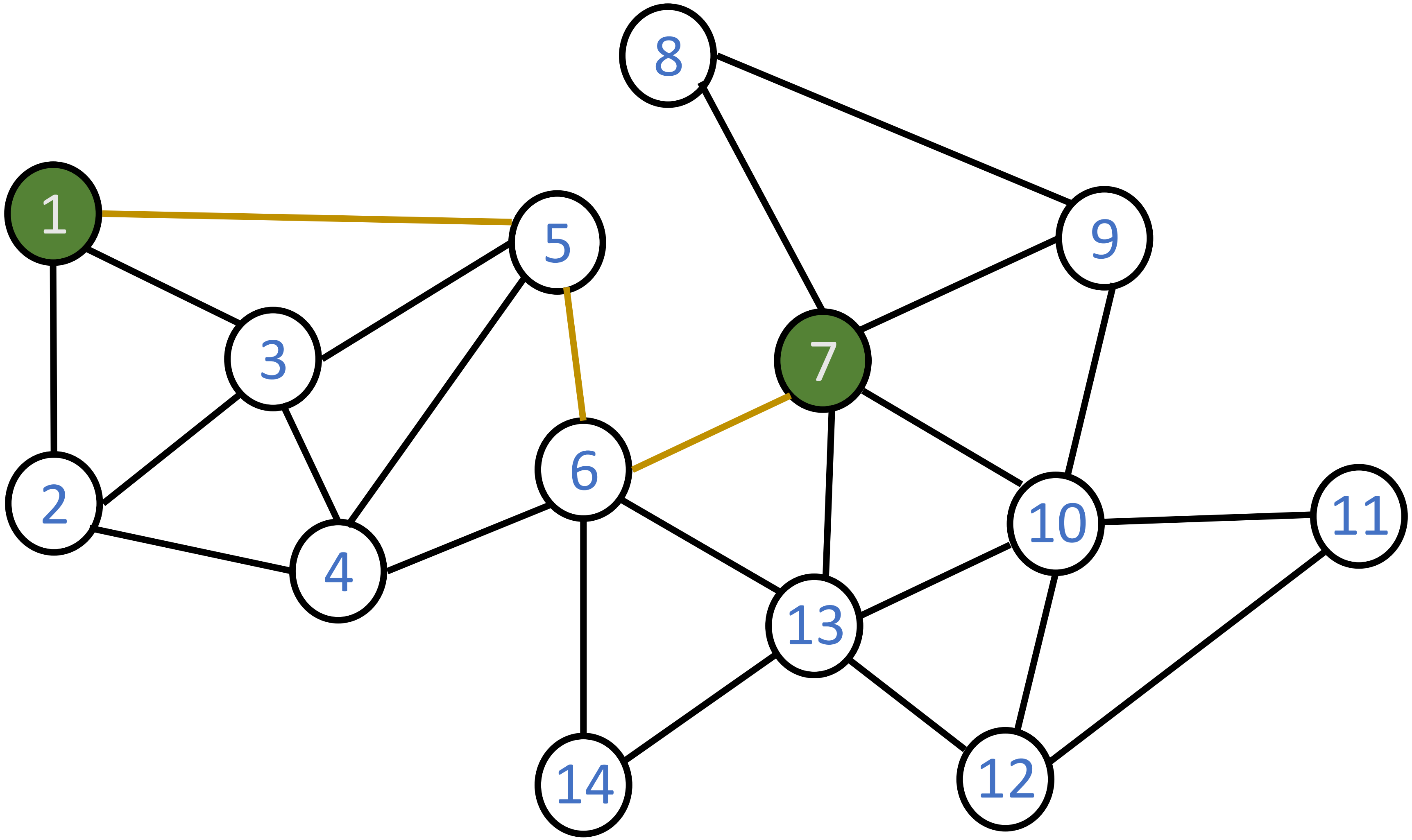
Lightning Network

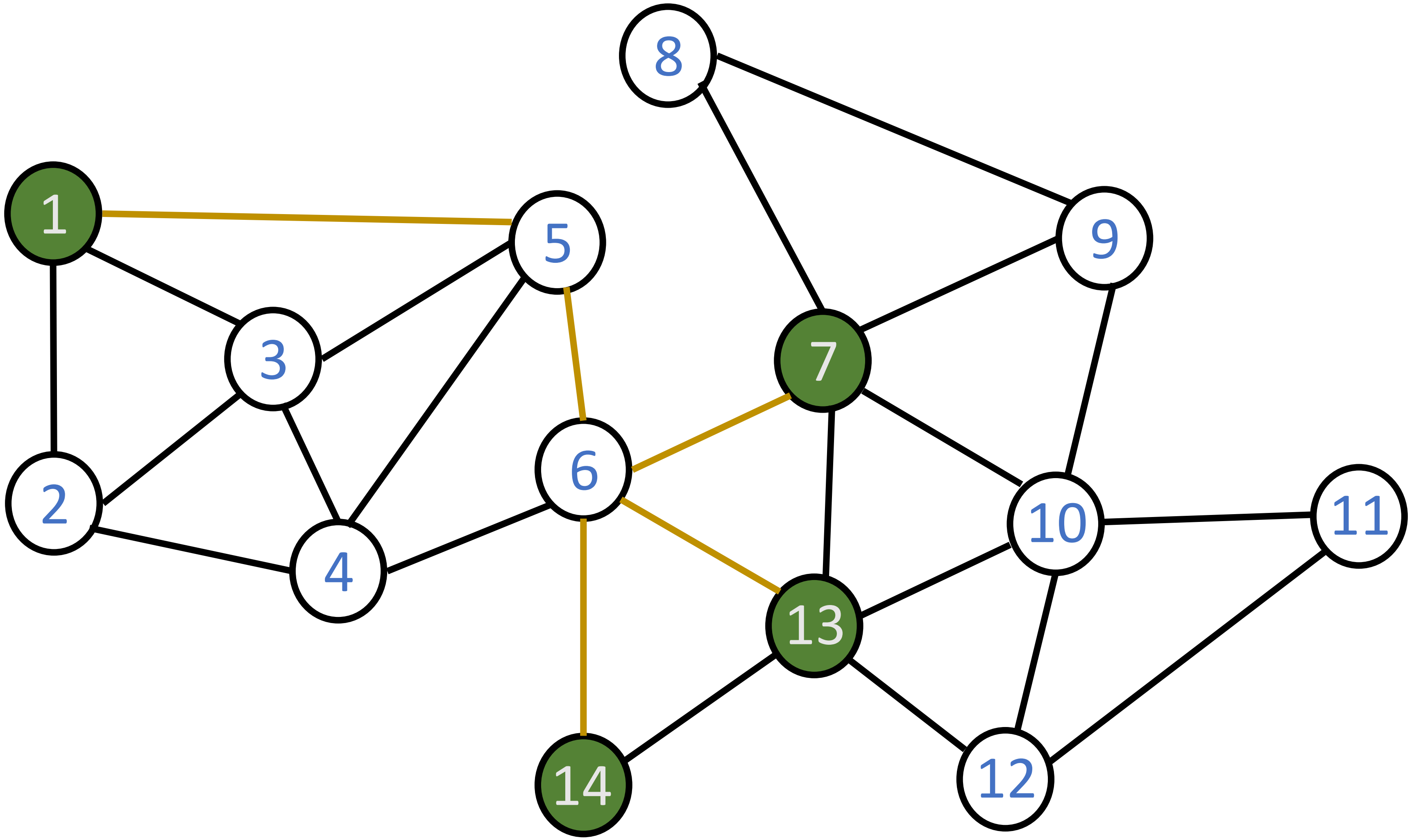
- An overlay payment channel network to help provide scalability and anonymity to Bitcoin
- Uses onion routing: forwarder has no knowledge of the originator or receiver
- Transaction are performed with the help of intermediate "channels"
  - Routing payment through a channel incurs cost
  - Best path is constructed by minimizing the overall cost
- Complete network is public and known to all nodes

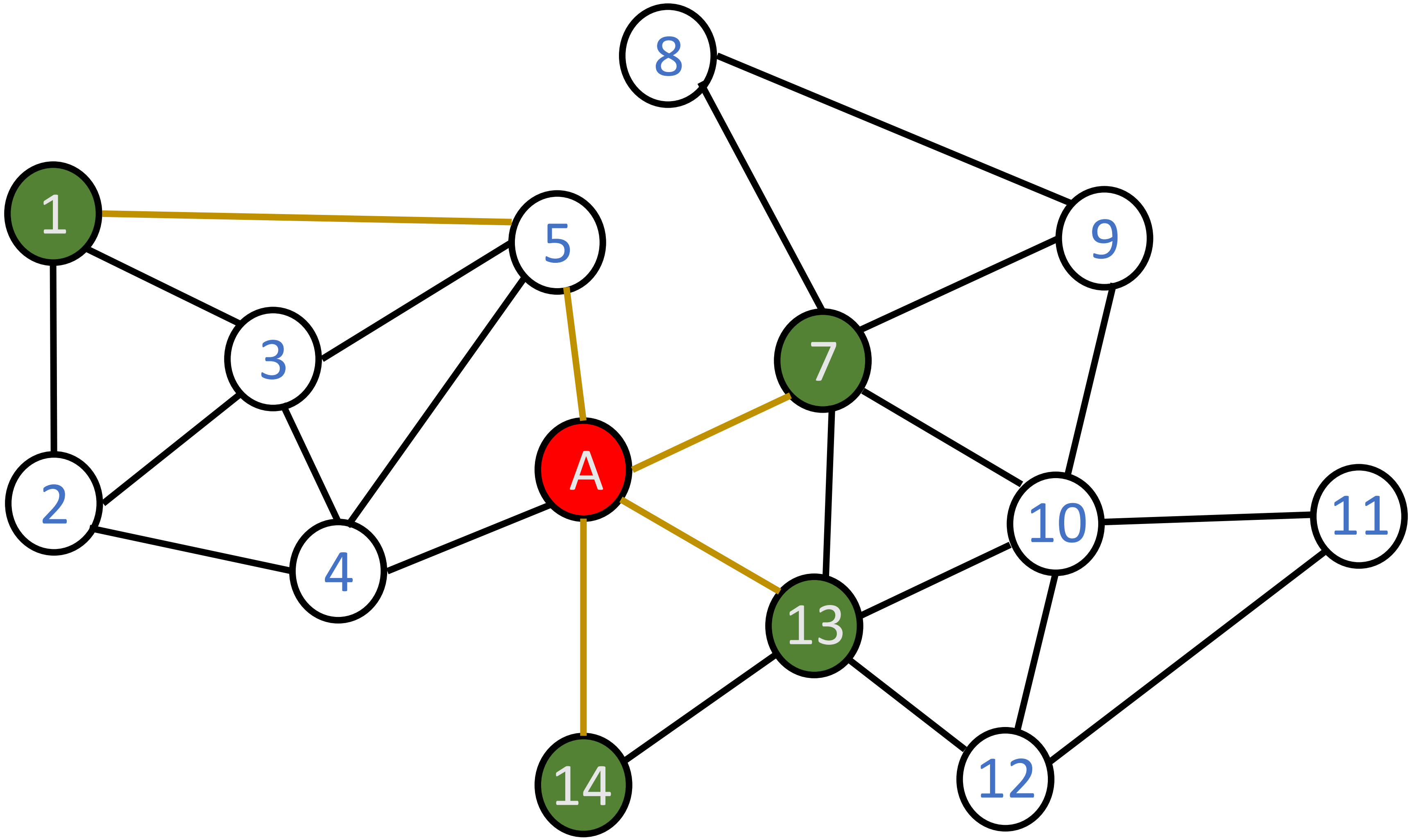

1
2
3
4
5
6
7
8
9
10
11
12
13
14


1
2
3
4
5
6
7
8
9
10
11
12
13
14
1


8
1
5
9
3
7
2
6
10
11
4
13
14
12


1
2
3
4
5
6
7
8
9
10
11
12
13
14


8
1
5
9
3
7
6
2
10
11
4
13
14
12


8
1
5
9
3
7
2
A
10
11
4
13
14
12

Modelling Lightning Network

- Two steps are involved

STEP I:

- Calculate all source--destination pair *paths* (based on min. weight) with public knowledge of network graph and their weights

STEP II:

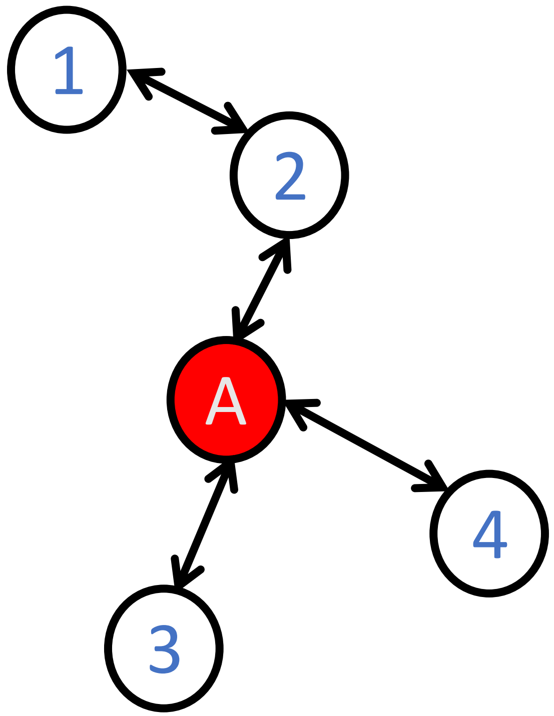
- Calculate probability distribution using info from STEP I

Modelling Lightning Network

- $P(A|B_i) = SP_{iA} / SP_i$

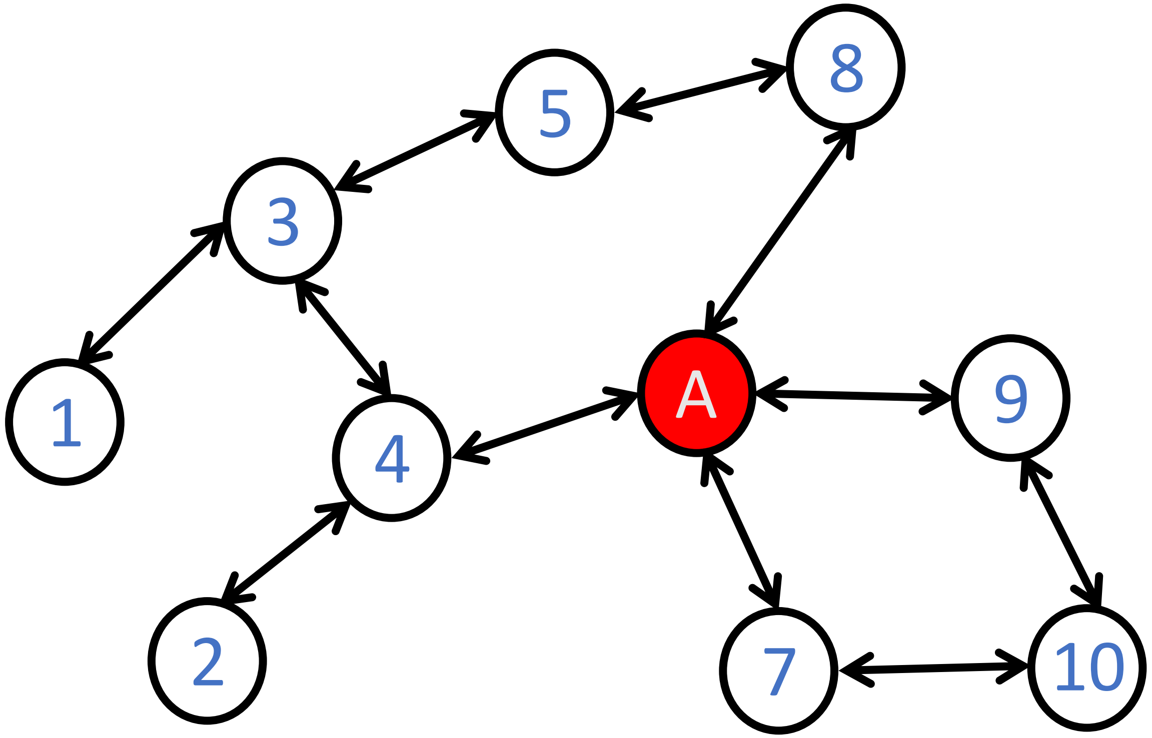
- SP_{iA} = No. of paths from i passing through A

- SP_i = No. of paths from i



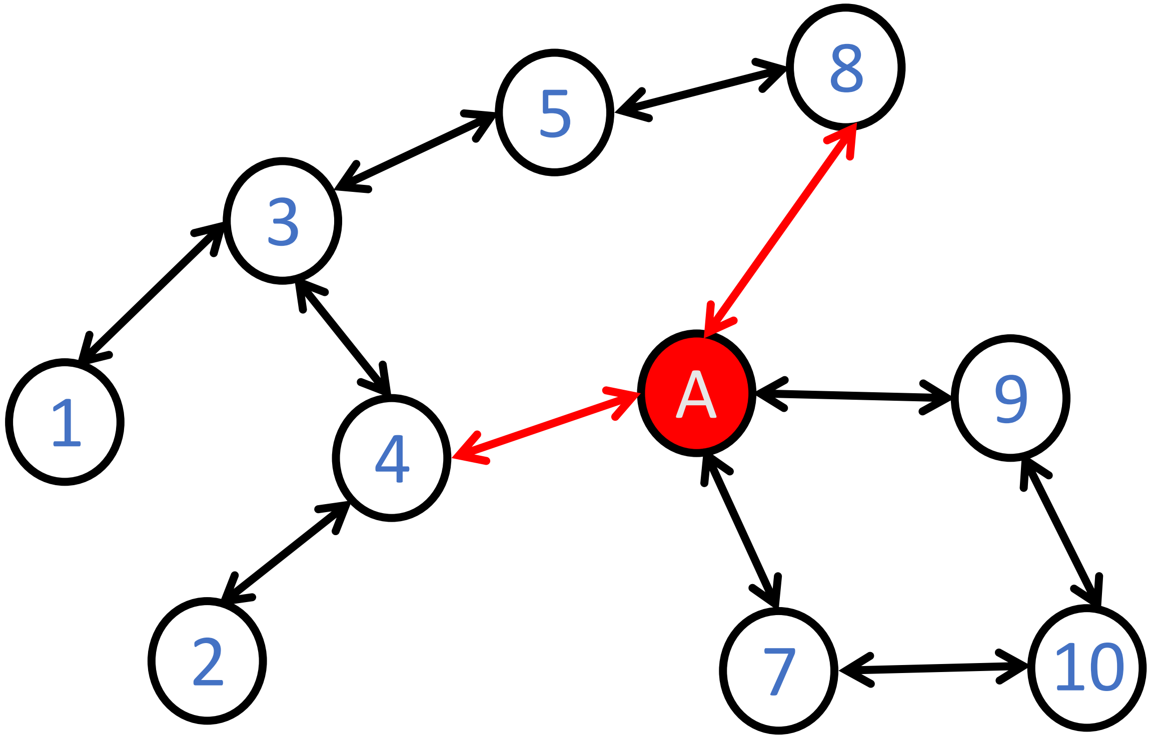
Can we do better?

- The **predecessor and successor** for a node are predefined in a source routed scheme
- Can help to reduce the potential set of originators



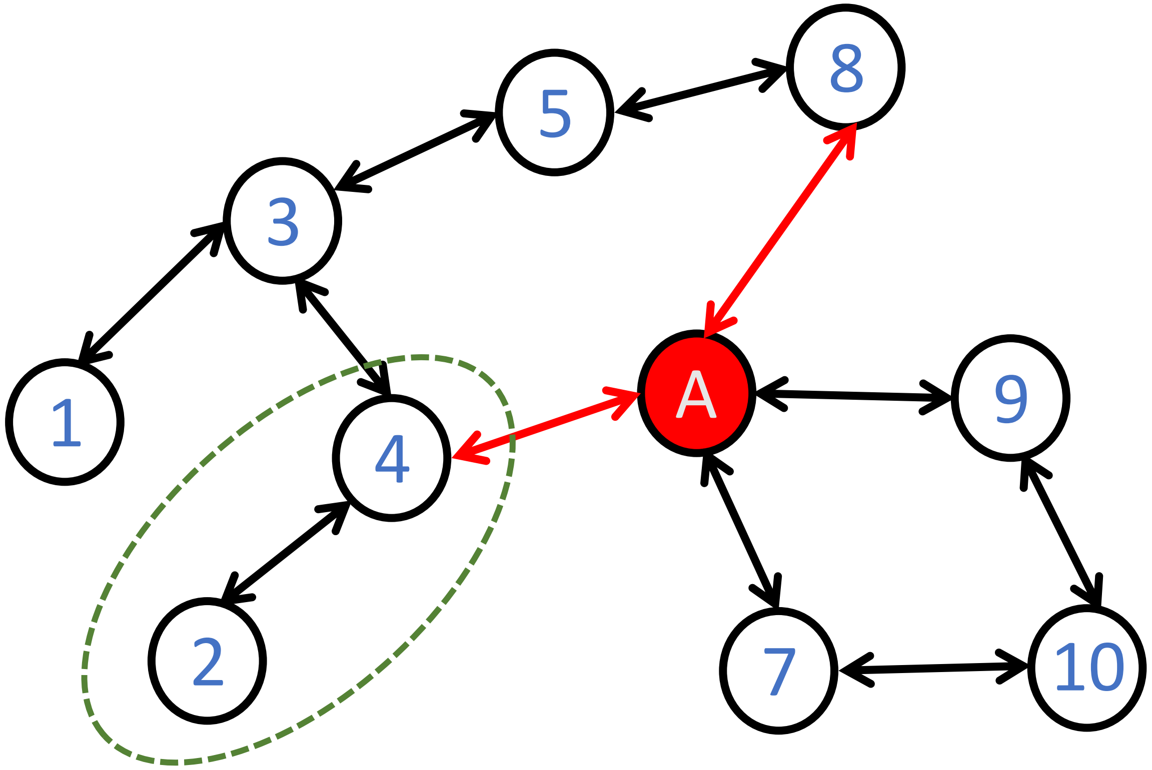
Can we do better?

- The **predecessor and successor** for a node are predefined in a source routed scheme
- Can help to reduce the potential set of originators
- For subpath 4 -> A -> 8, only nodes 2 and 4 are the potential originators



Can we do better?

- The **predecessor and successor** for a node are predefined in a source routed scheme
- Can help to reduce the potential set of originators
- For subpath 4 -> A -> 8, only nodes 2 and 4 are the potential originators



Evaluation of LN

- Incorporated the modelling approach in a simulator
  - Obtain public topology snapshot of LN
  - Assign adversary nodes based on different strategies
  - Calculate probability distributions for Tx observed by adversary nodes
  - Measured the entropy

Evaluation

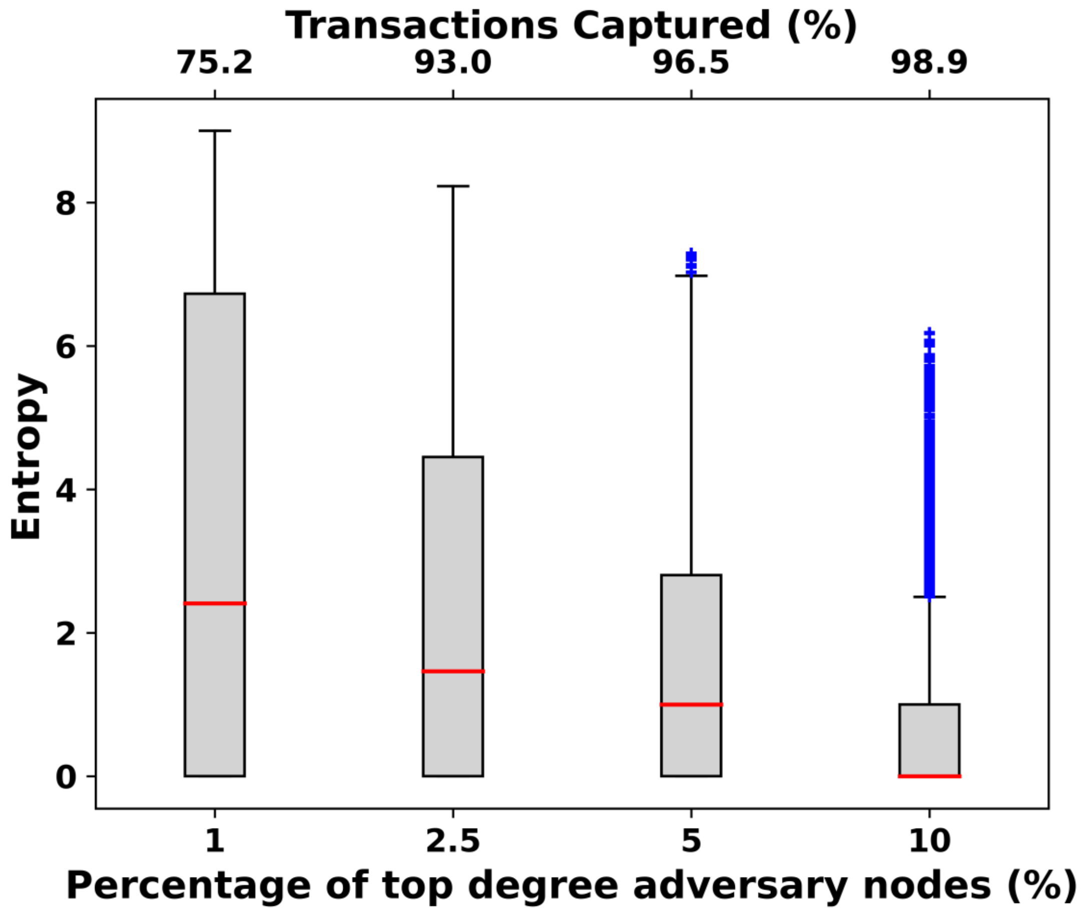
- Strategic selection: High node-degree
- Longitudinal analysis with the strategic selection
- Select randomly among best-k paths instead of only the best one.
- Transaction amount

Evaluation

- Strategic selection: High node-degree
- Longitudinal analysis with the strategic selection
- Select randomly among best-k paths instead of only the best one.
- Transaction amount

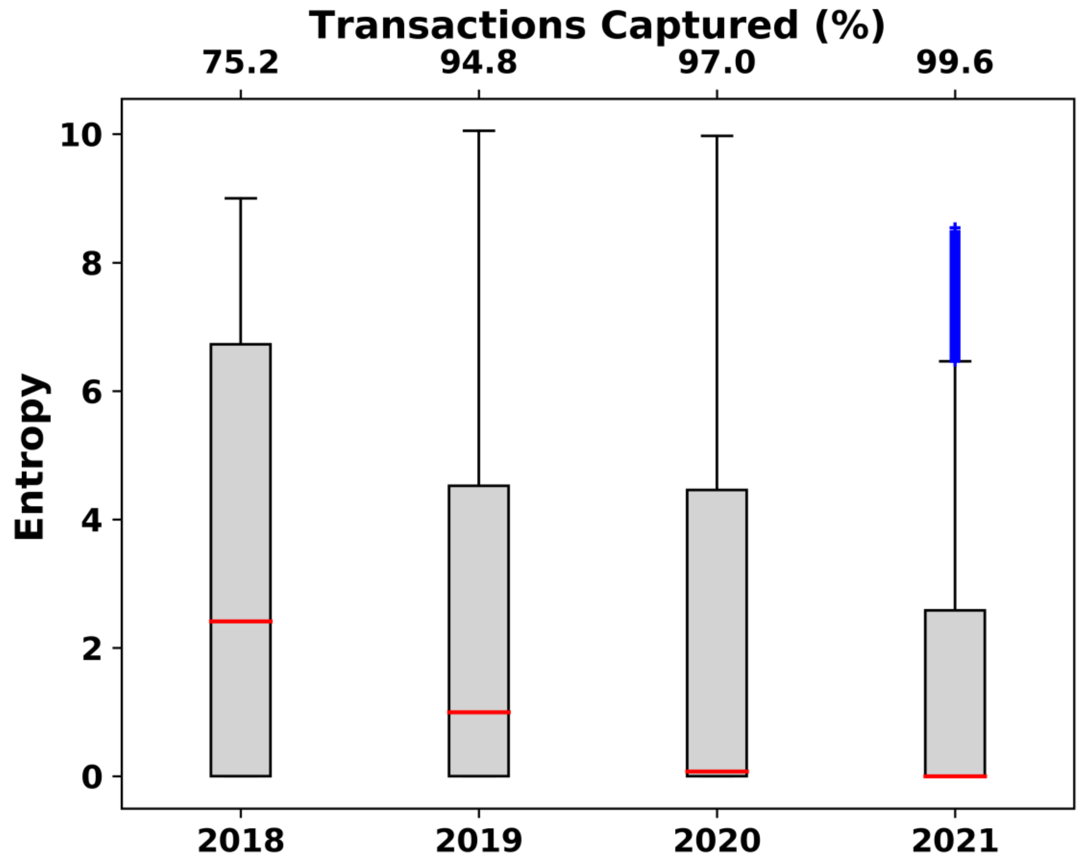
Strategic selection based on node degree

- We select top-degree nodes as adversary
  - Analysis for the topology with 2018 (1200) nodes with varying adversary node fraction



Longitudinal analysis

- We select top-1% degree nodes as adversary
  - Perform analysis for 2018 (1200 nodes), 2019, 2020 and 2021 (9000 nodes) topology



Summary

- We proposed a generic Bayesian framework to evaluate network-level anonymity in peer-to-peer networks
- We modelled and evaluated three schemes proposed or deployed to support transaction anonymity in Bitcoin
  - Hop by hop: Dandelion, Dandelion++
  - Source routed: Lightning Network
- We present a detailed evaluation of the schemes and observe that they do not generally offer high anonymity to transactions
- To encourage reproducibility we make the source code of the simulator and analysis public
  - https://netanoncrypt.github.io

Thanks