Spy vs. Spy: Rumor Source Obfuscation

Peter Kairouz

University of Illinois at Urbana-Champaign

Joint work with

Giulia Fanti, Sewoong Oh, and Pramod Viswanath

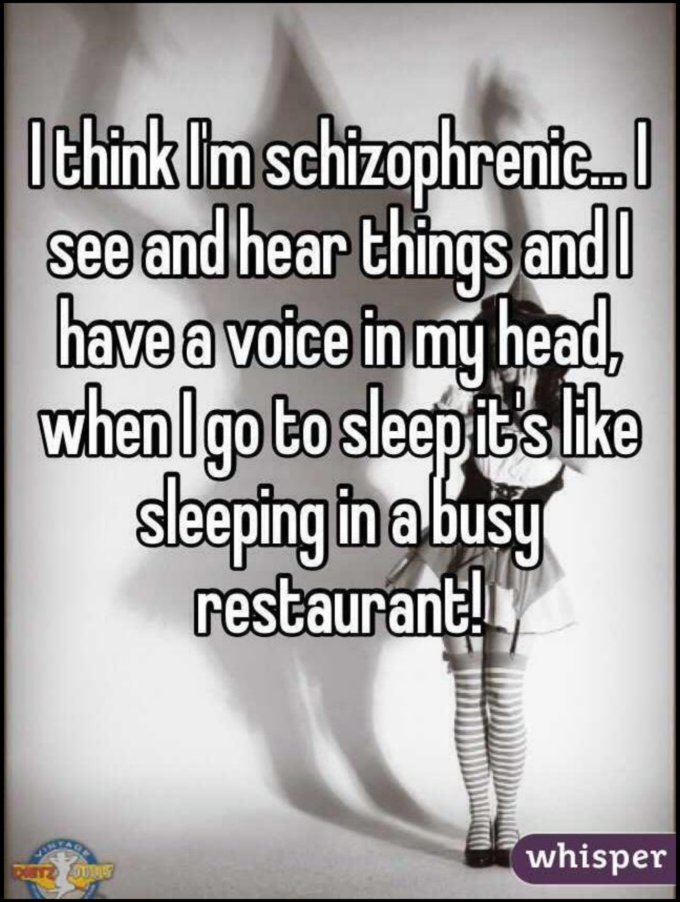
Political activism

Some people have important, sensitive things to say.

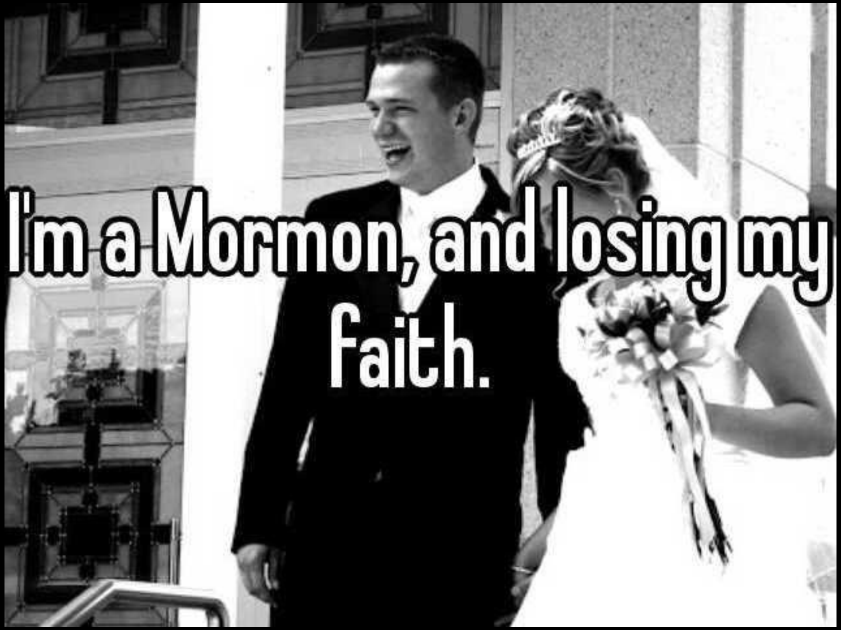
PERSONS OF THE YEAR
TIME
The Whistleblowers
CYNTHIA COOPER OF WORLDCOM
COLEEN ROWLEY OF THE FBI
SHERRON WATKINS OF ENRON

Personal confessions

Others have less important, but sensitive things to say.



I think I'm schizophrenic... I see and hear things and I have a voice in my head, when I go to sleep it's like sleeping in a busy restaurant!



I'm a Mormon, and losing my faith.

Existing anonymous messaging apps

secret

whisper

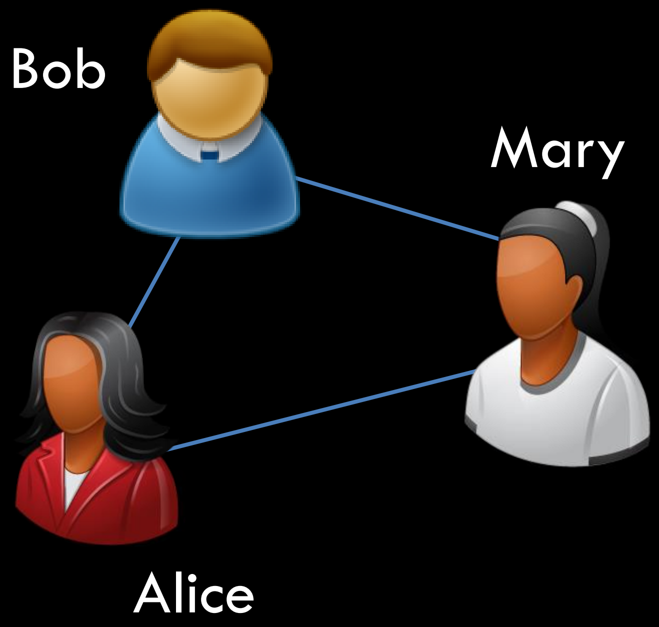
Yik Yak

Existing anonymous messaging apps

secret

whisper

Yik Yak

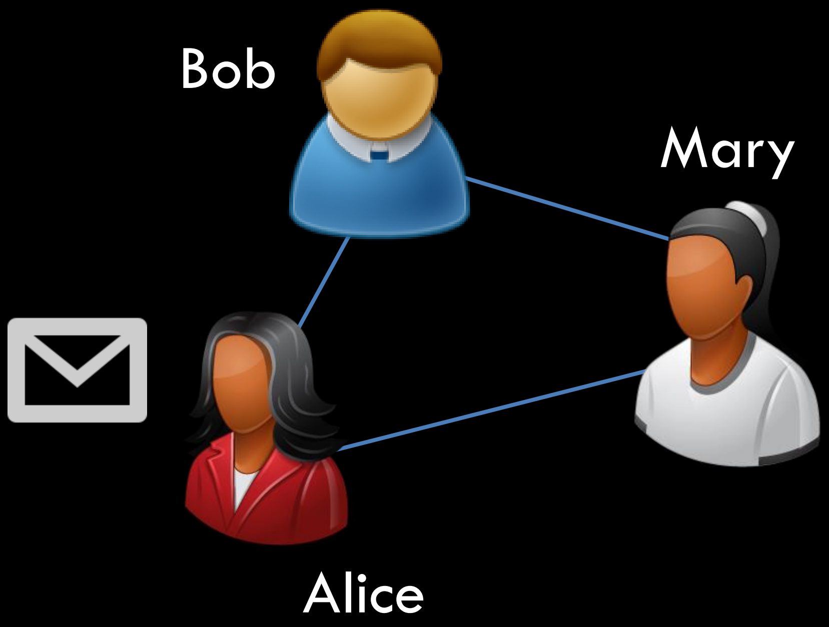


Existing anonymous messaging apps

secret

whisper

Yik Yak

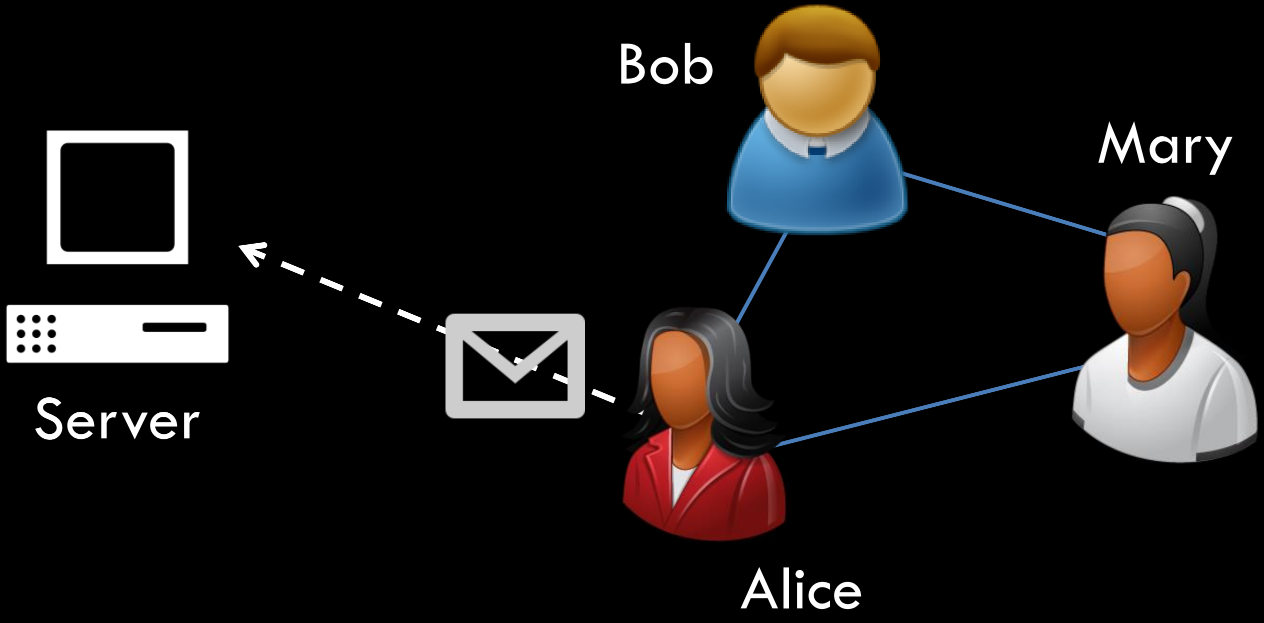


Existing anonymous messaging apps

secret

whisper

Yik Yak

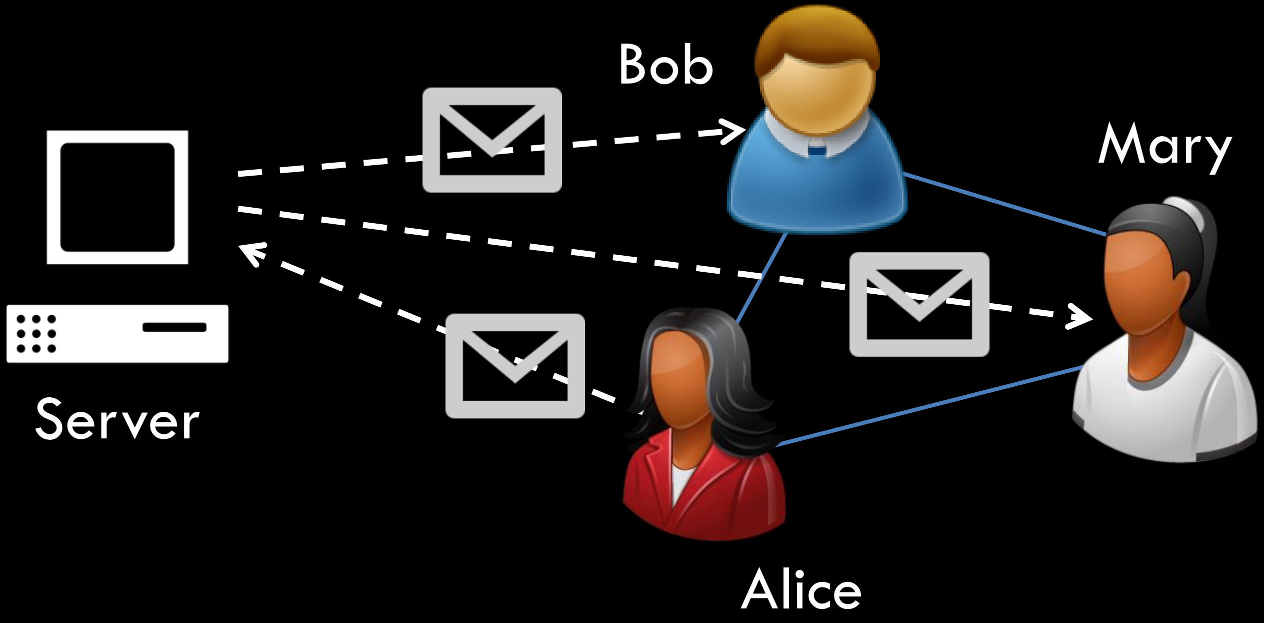


Existing anonymous messaging apps

secret

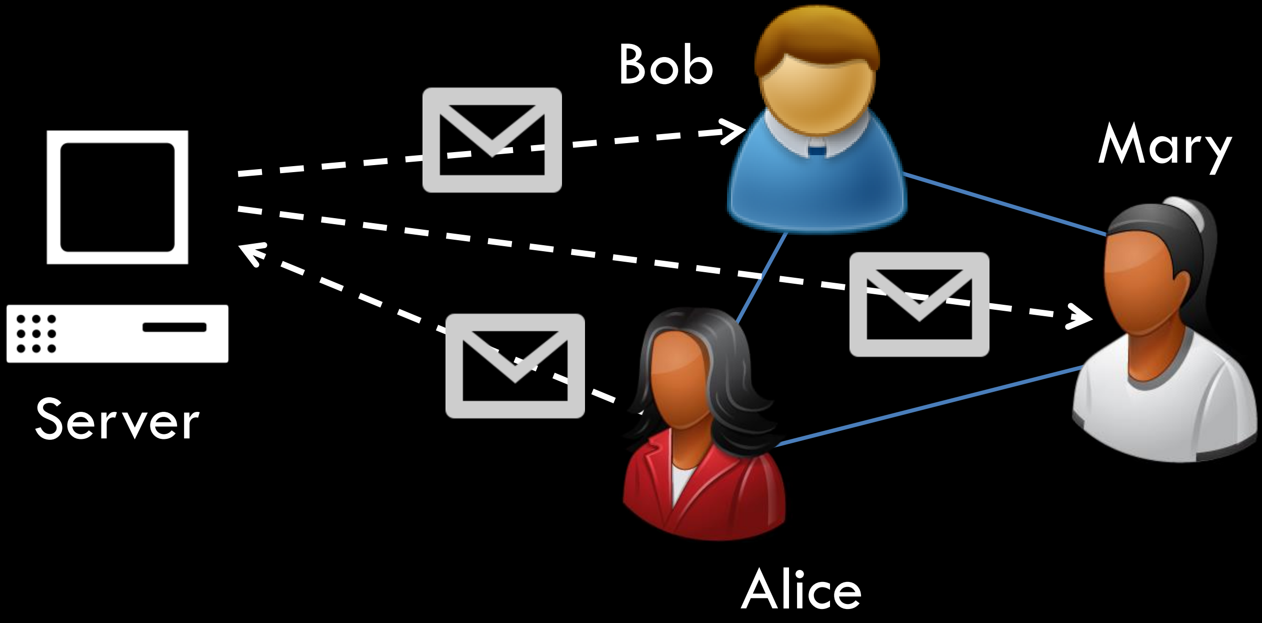
whisper

Yik Yak



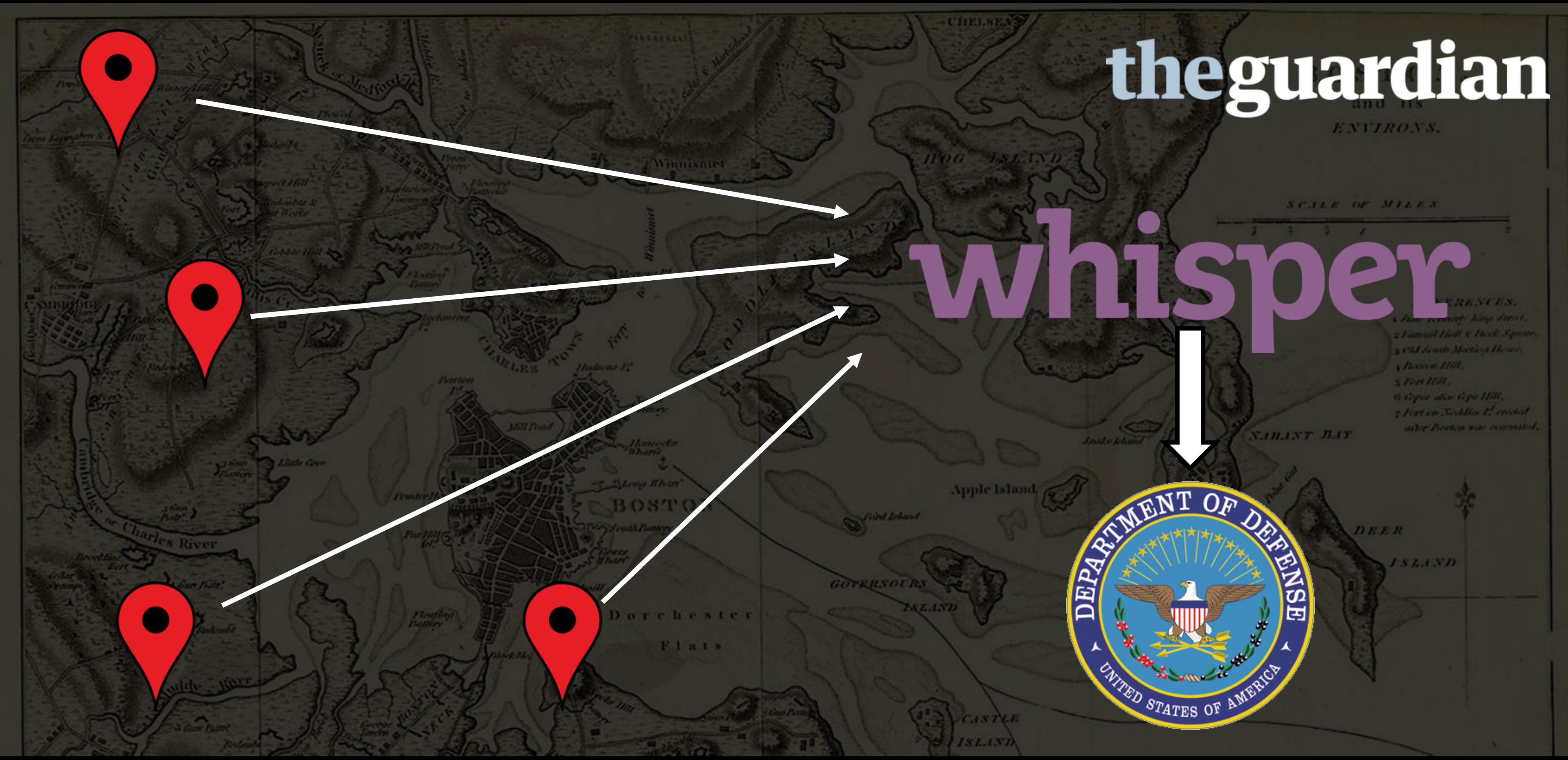
Existing anonymous messaging apps

secret whisper Yik Yak



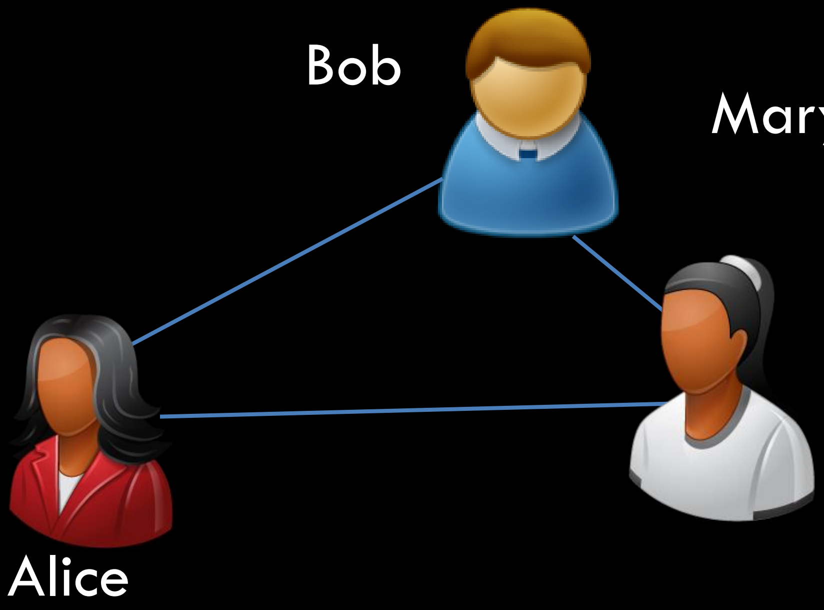
centralized networks are not truly anonymous!

Compromises in anonymity

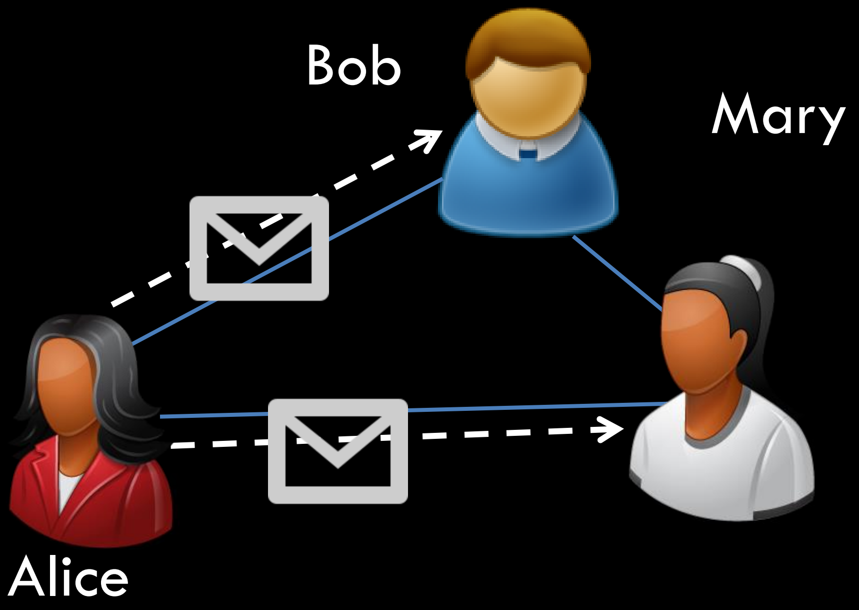


anonymity loss extends beyond the network

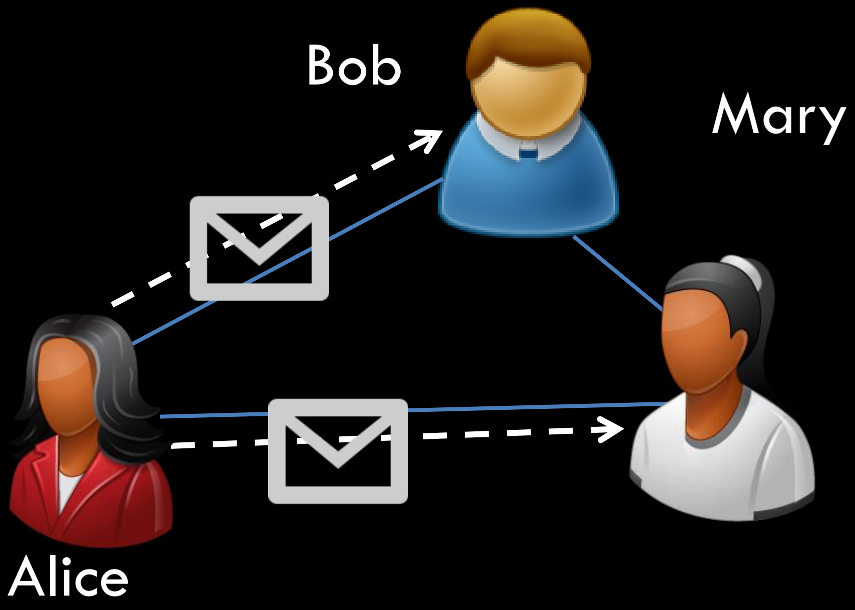
Distributed messaging



Distributed messaging

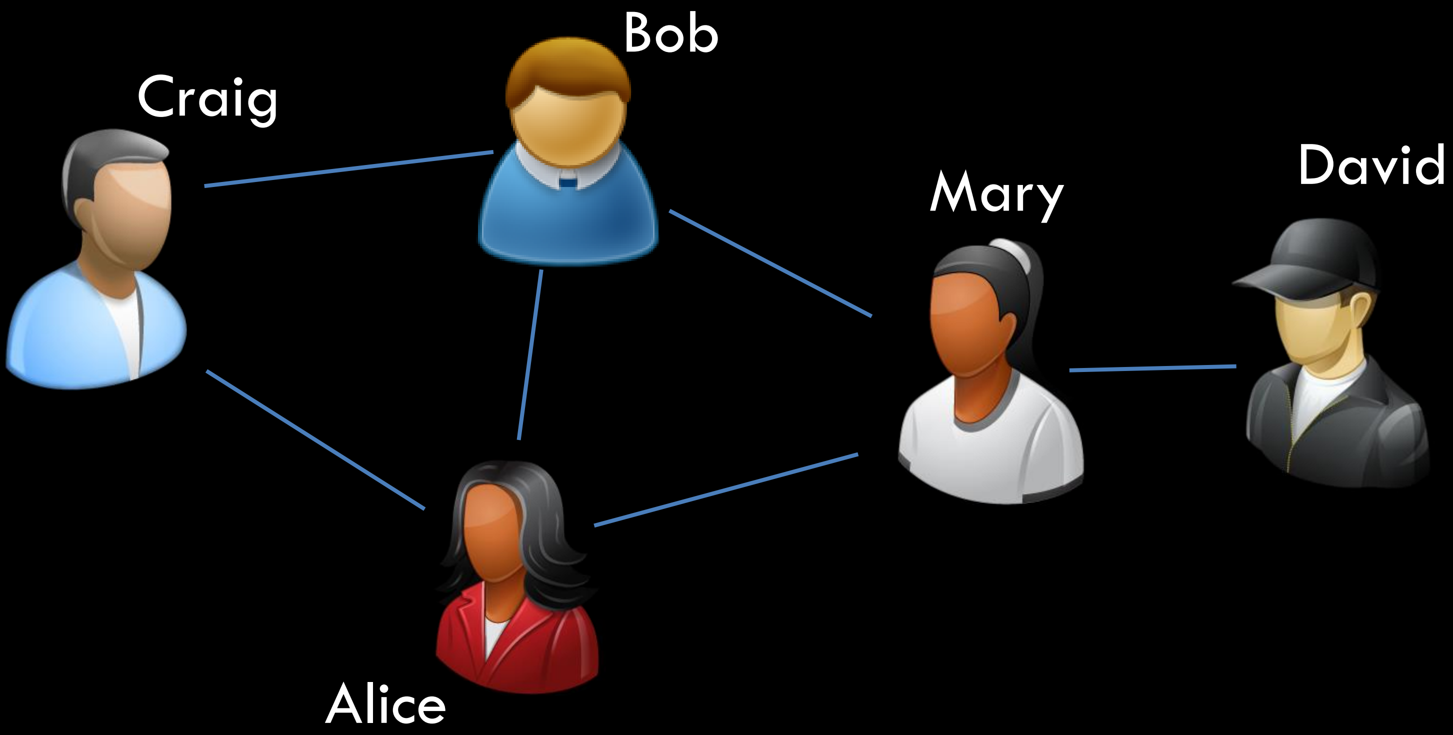


Distributed messaging

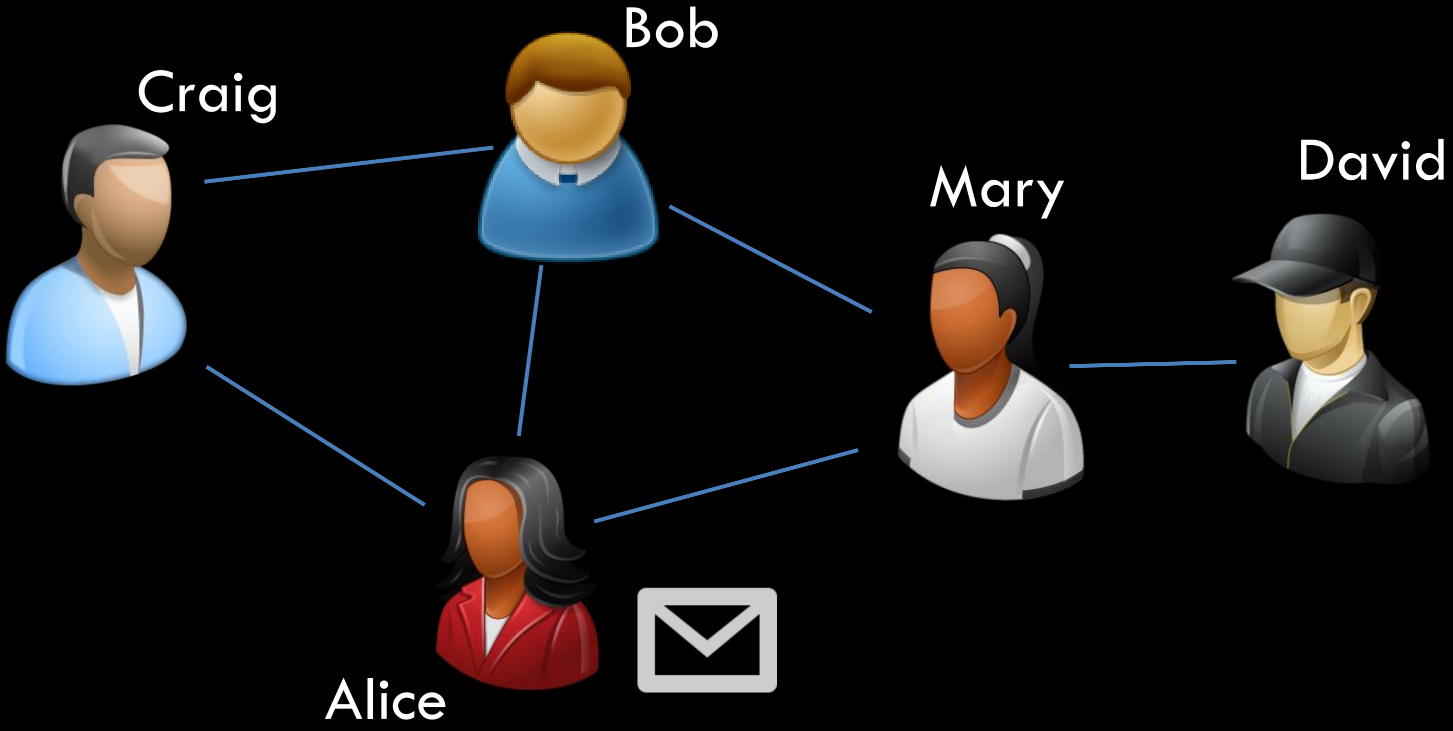


what can an adversary do?

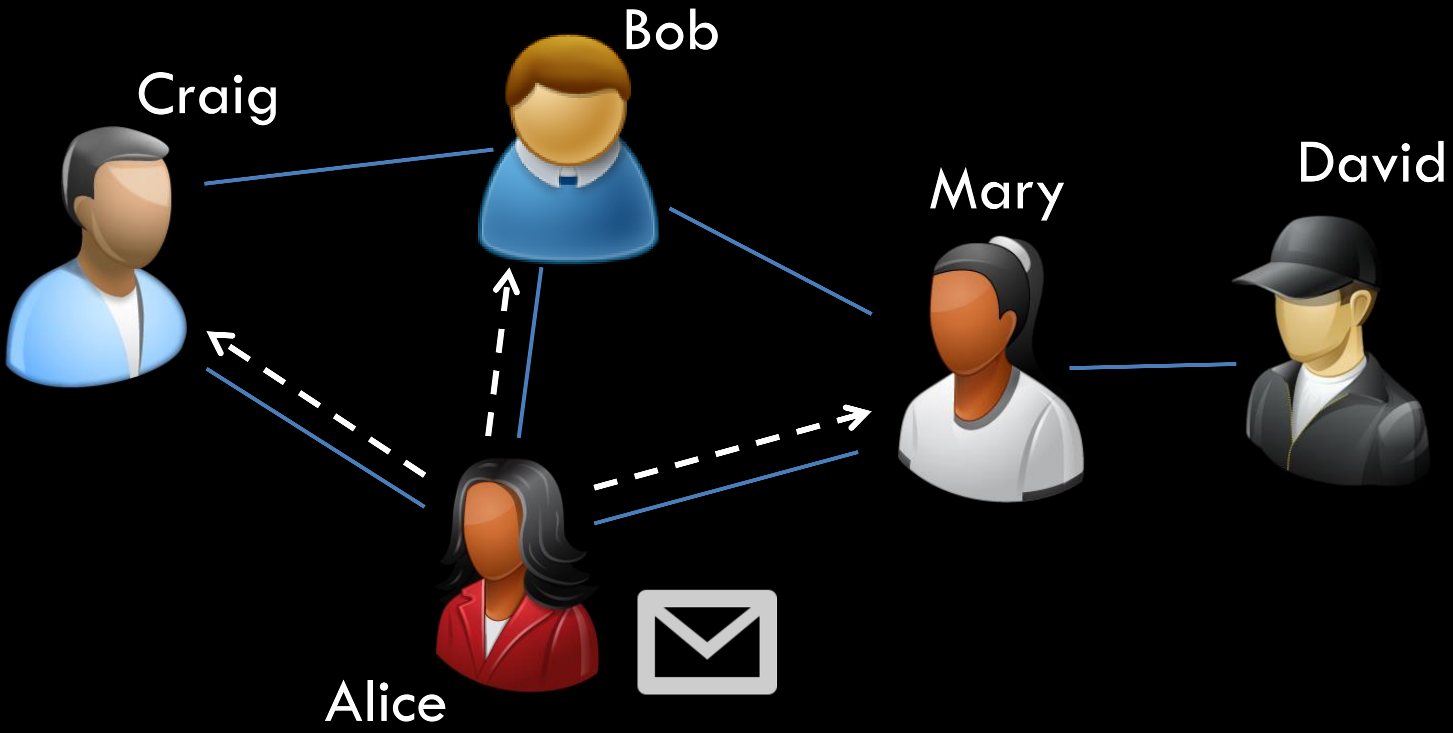
Adversarial model



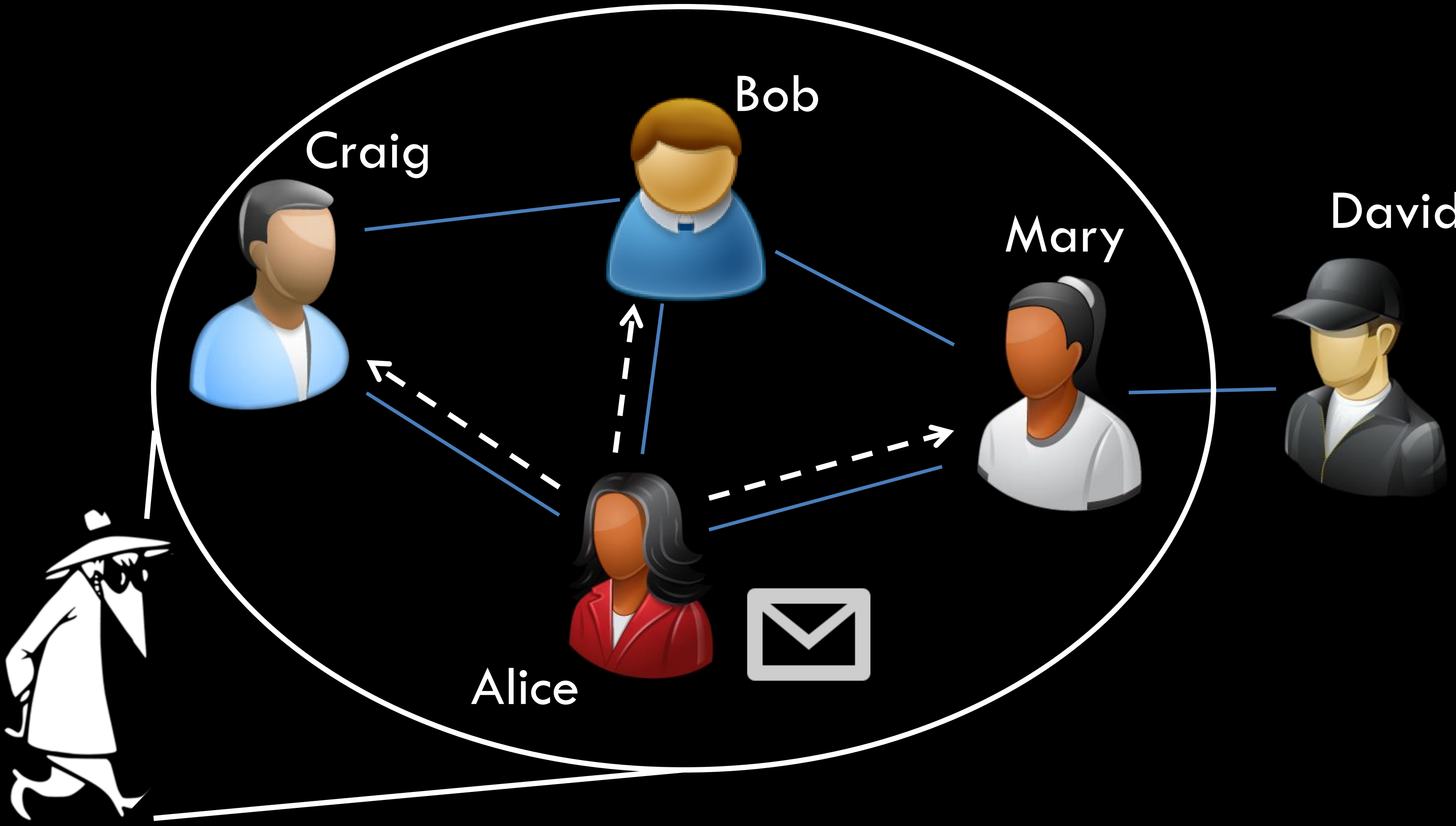
Adversarial model



Adversarial model

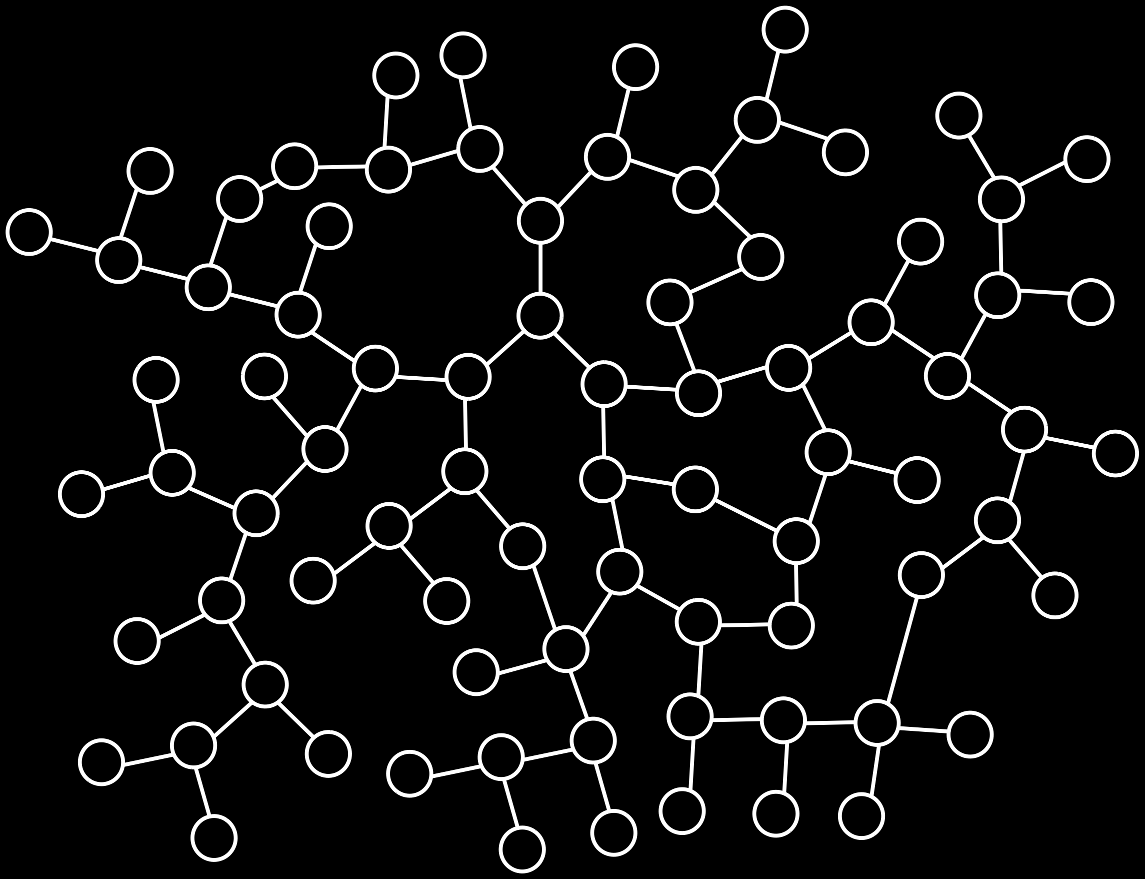


Adversarial model



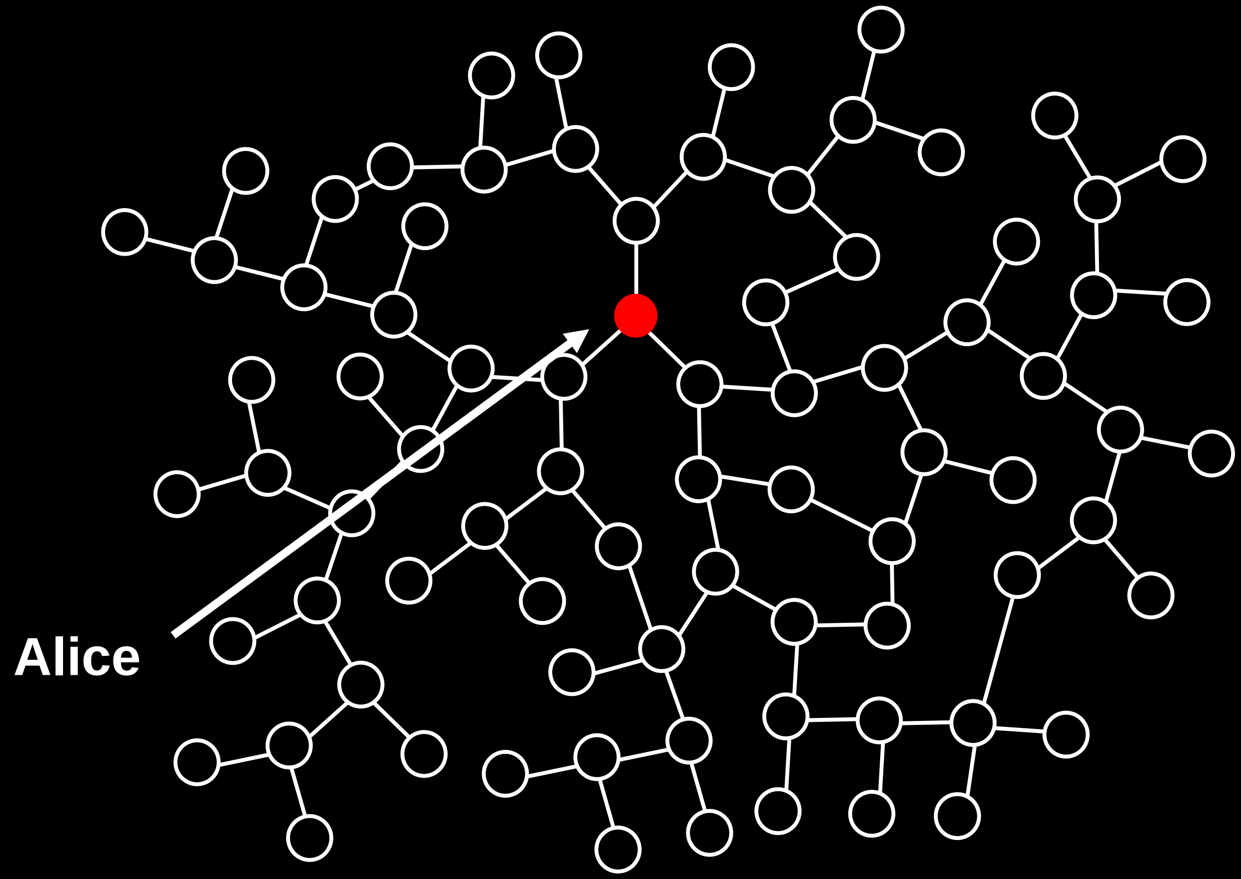
the adversary can figure out who got the message

Information flow in social networks

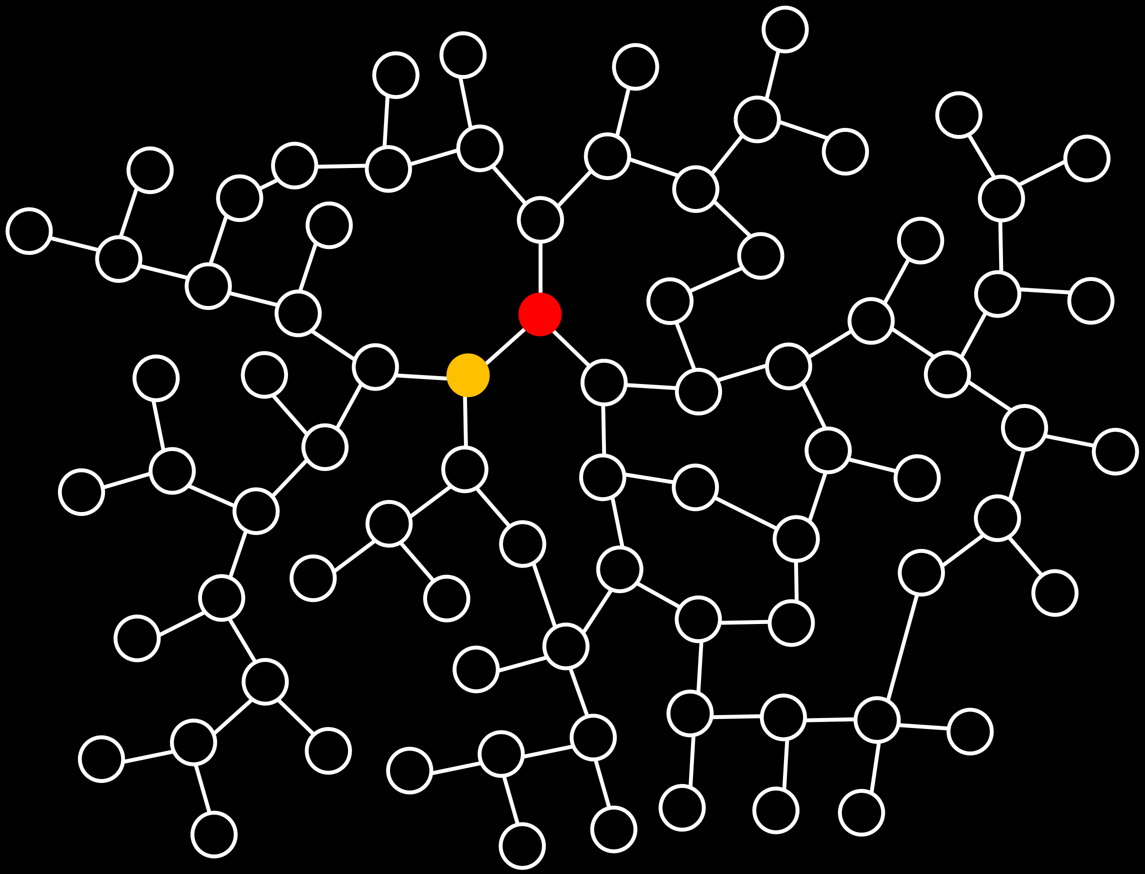


- G is the graph representing the social network

Information flow in social networks

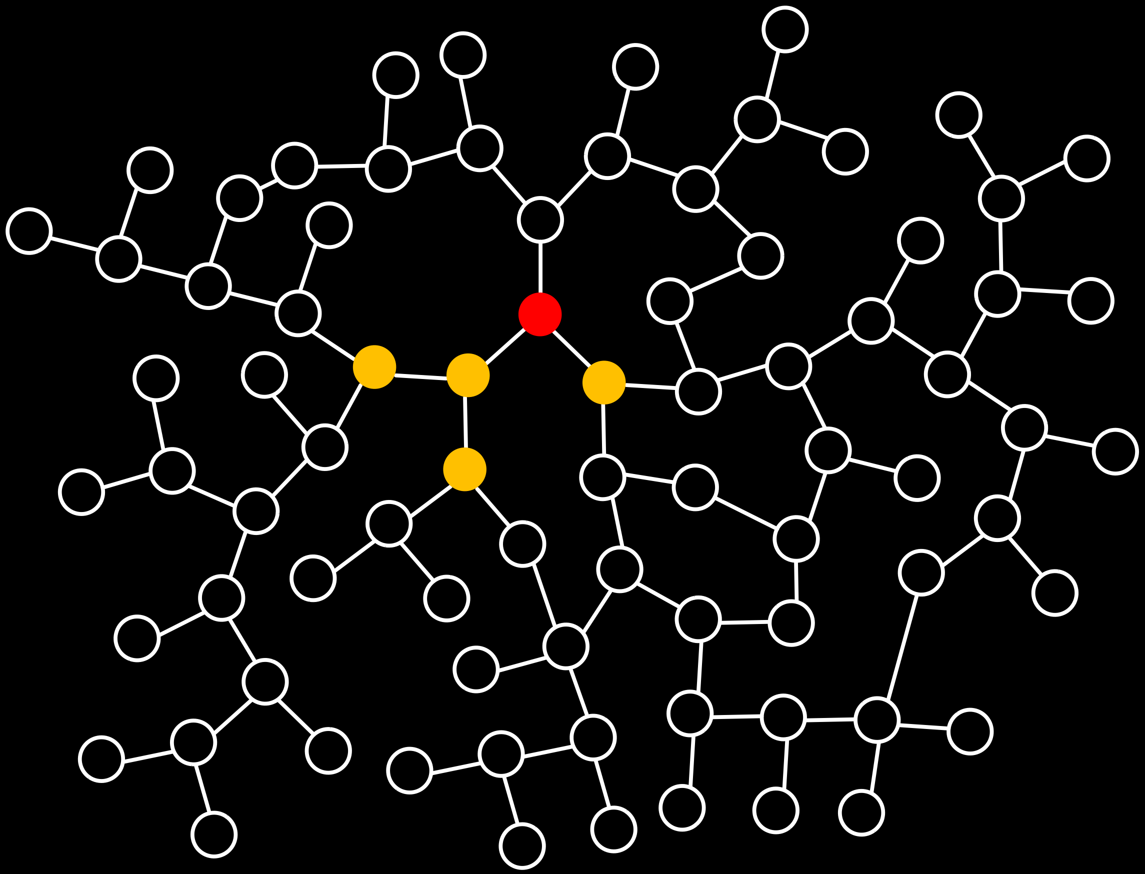


Information flow in social networks



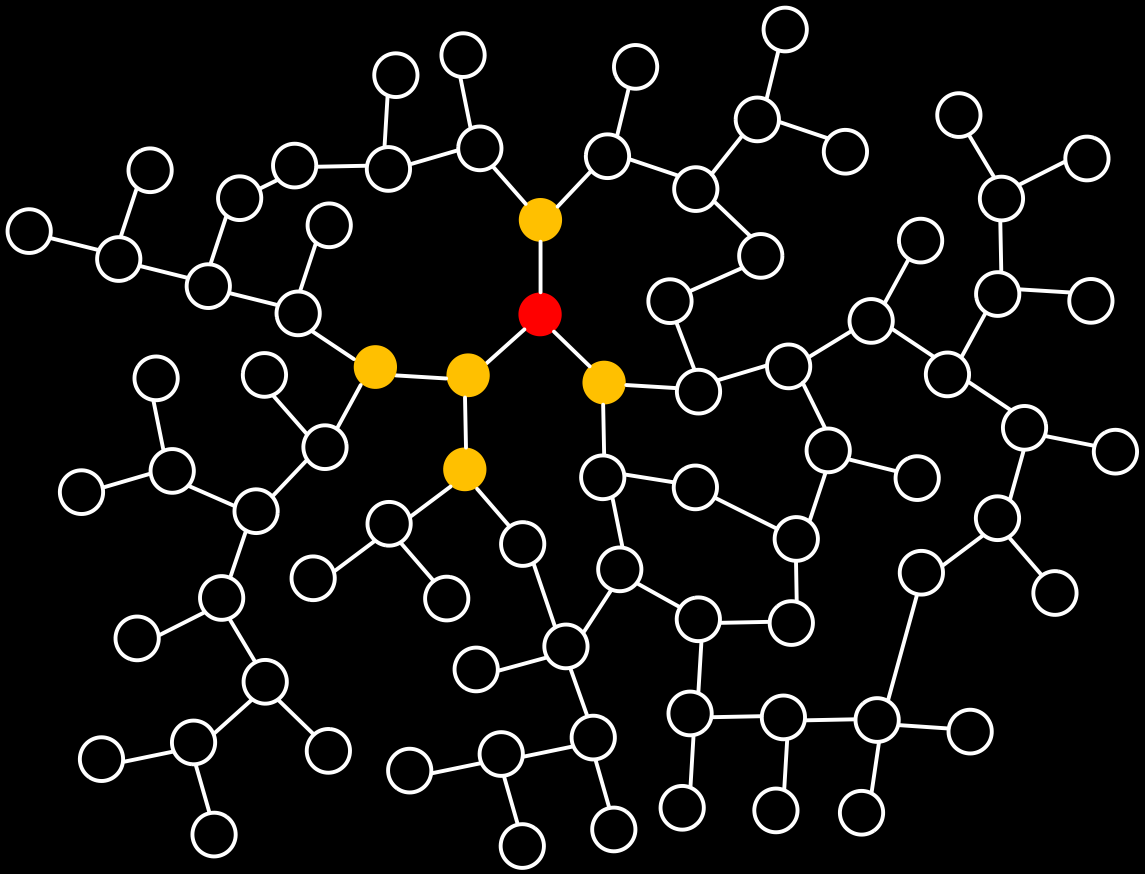
- Alice passes the message to her neighbors

Information flow in social networks



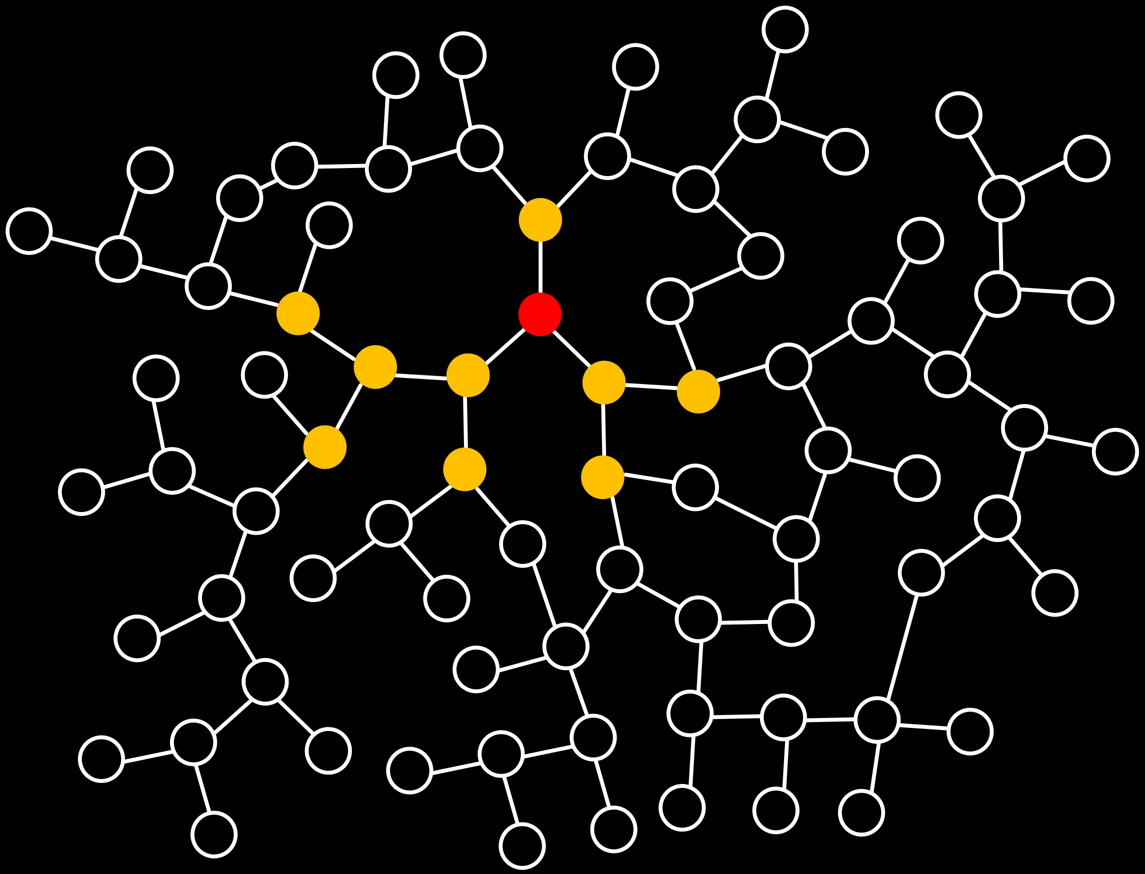
- her neighbors pass the message to theirs

Information flow in social networks



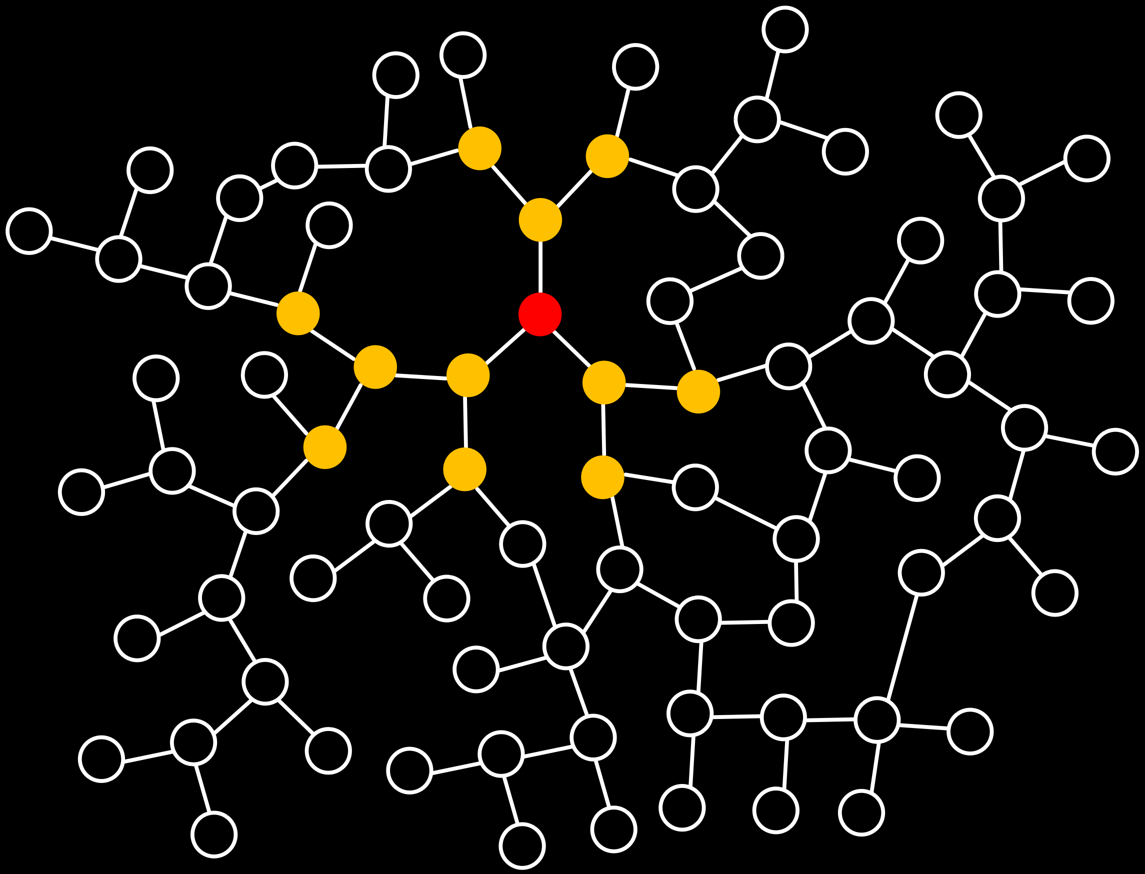
- the message spreads in **all directions** at the **same rate**

Information flow in social networks



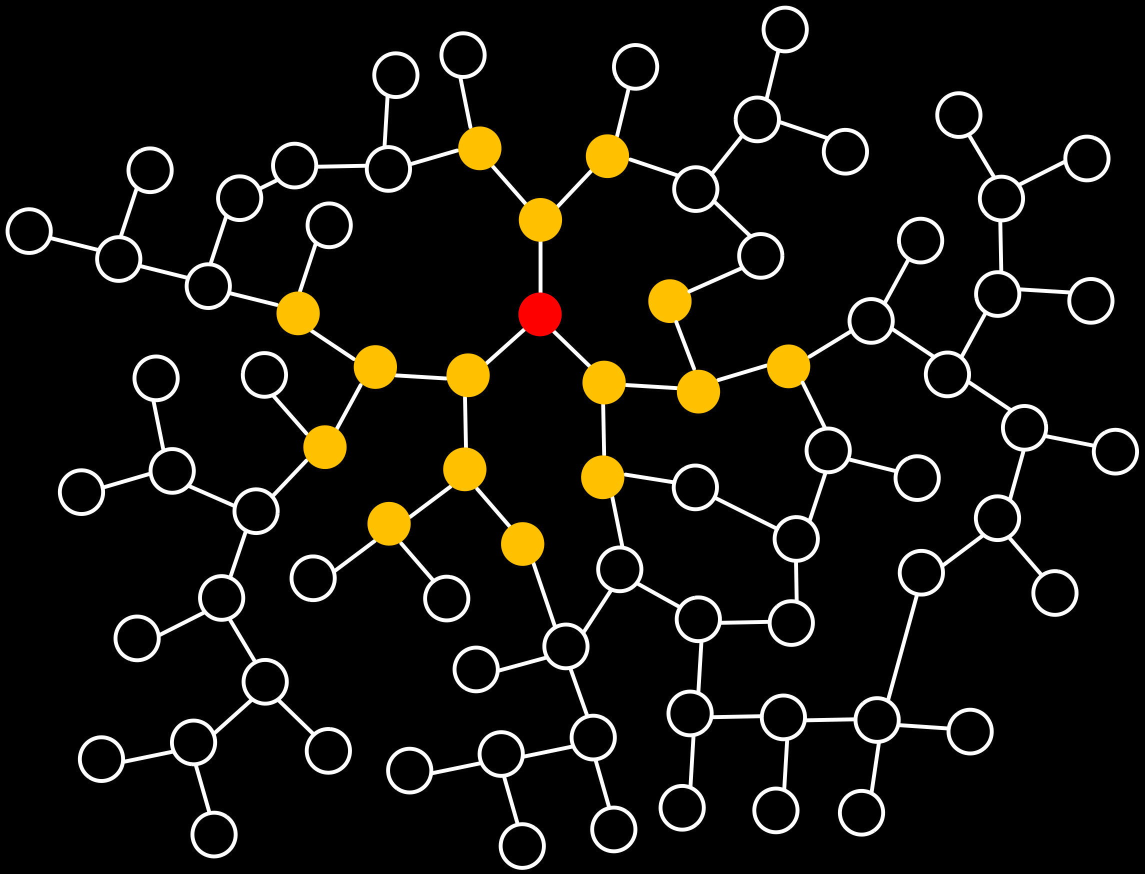
- the message spreads in **all directions** at the **same rate**

Information flow in social networks



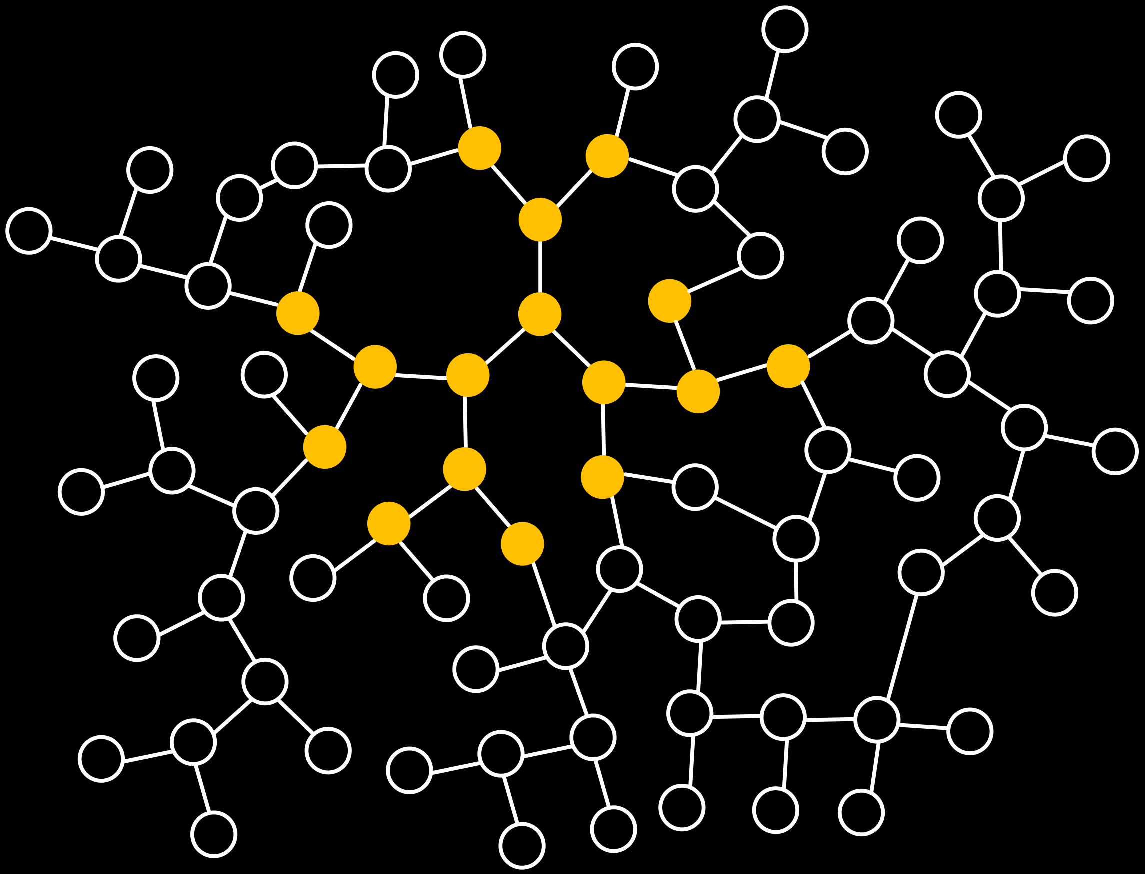
- the message spreads in **all directions** at the **same rate**

Information flow in social networks



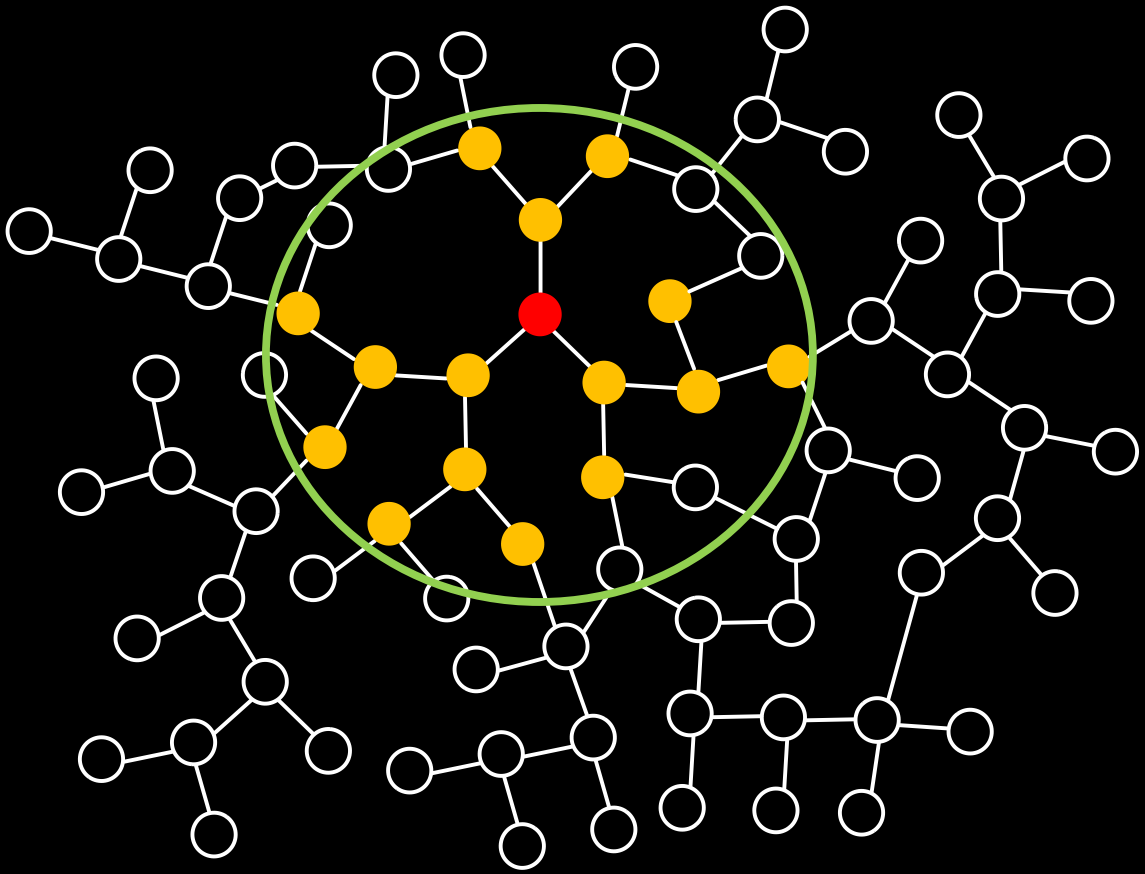
- this **spreading model** is known as the **diffusion model**

Adversary’s observation



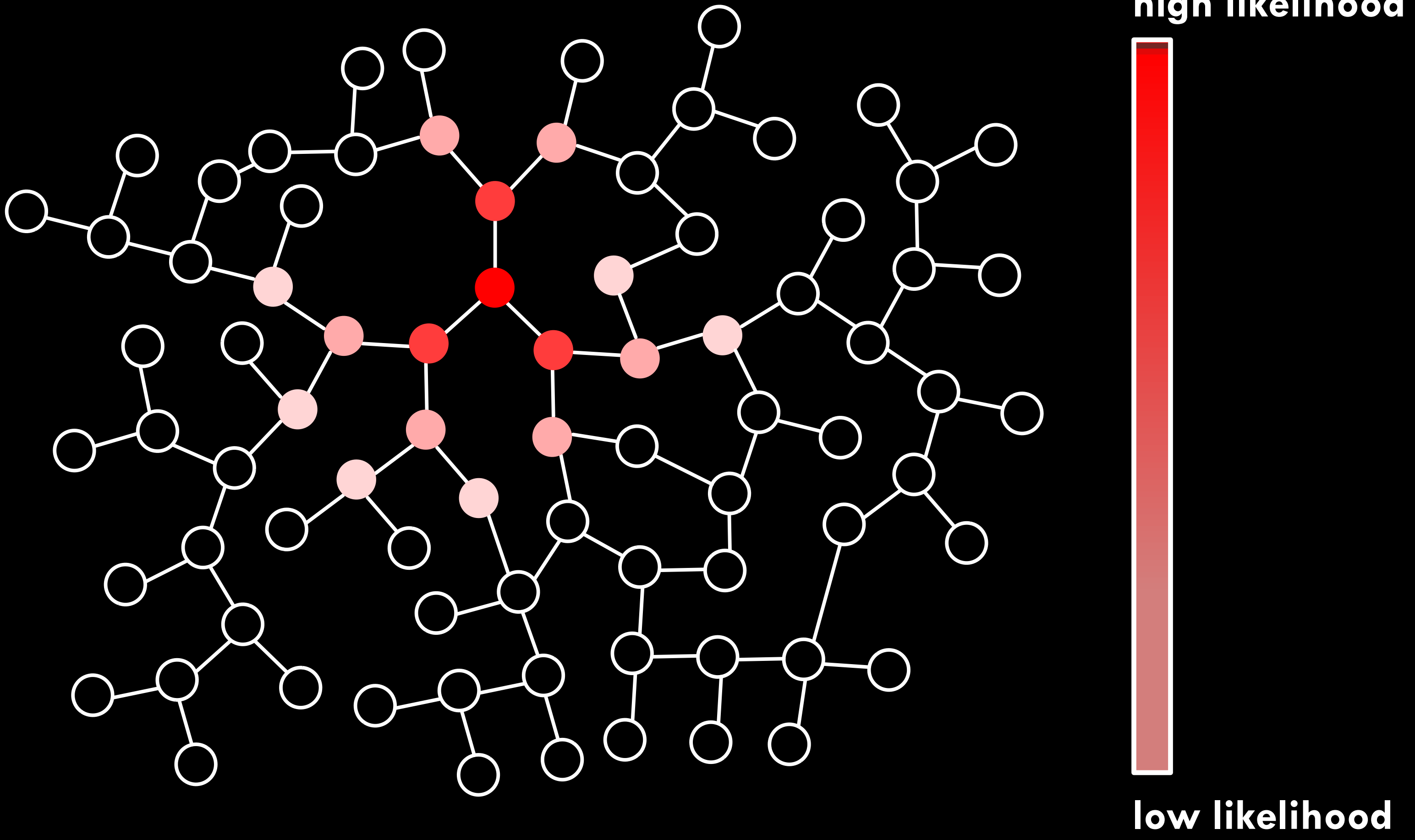
can the adversary locate the message author?

Concentration around the center



- the **message author** is in the **"center"** with high probability

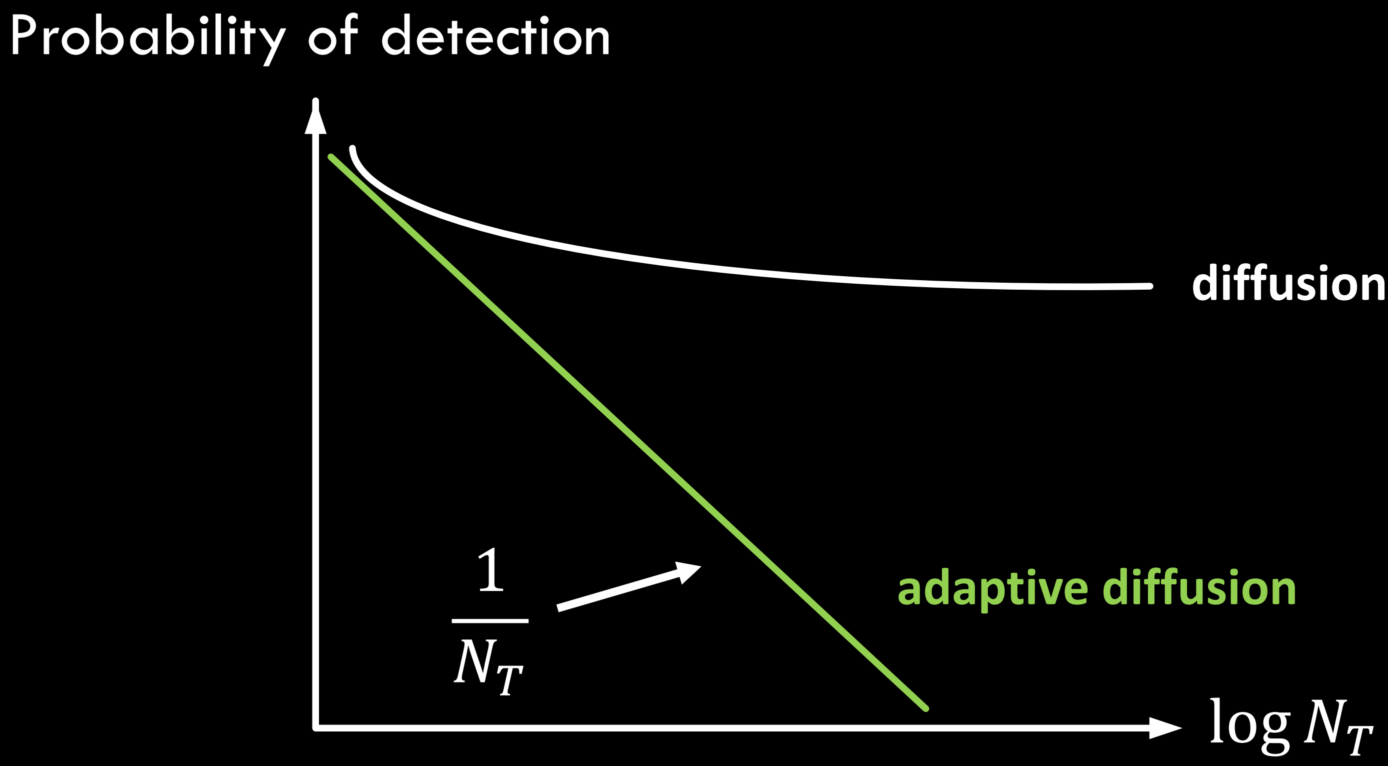
Rumor source identification



diffusion does not provide anonymity

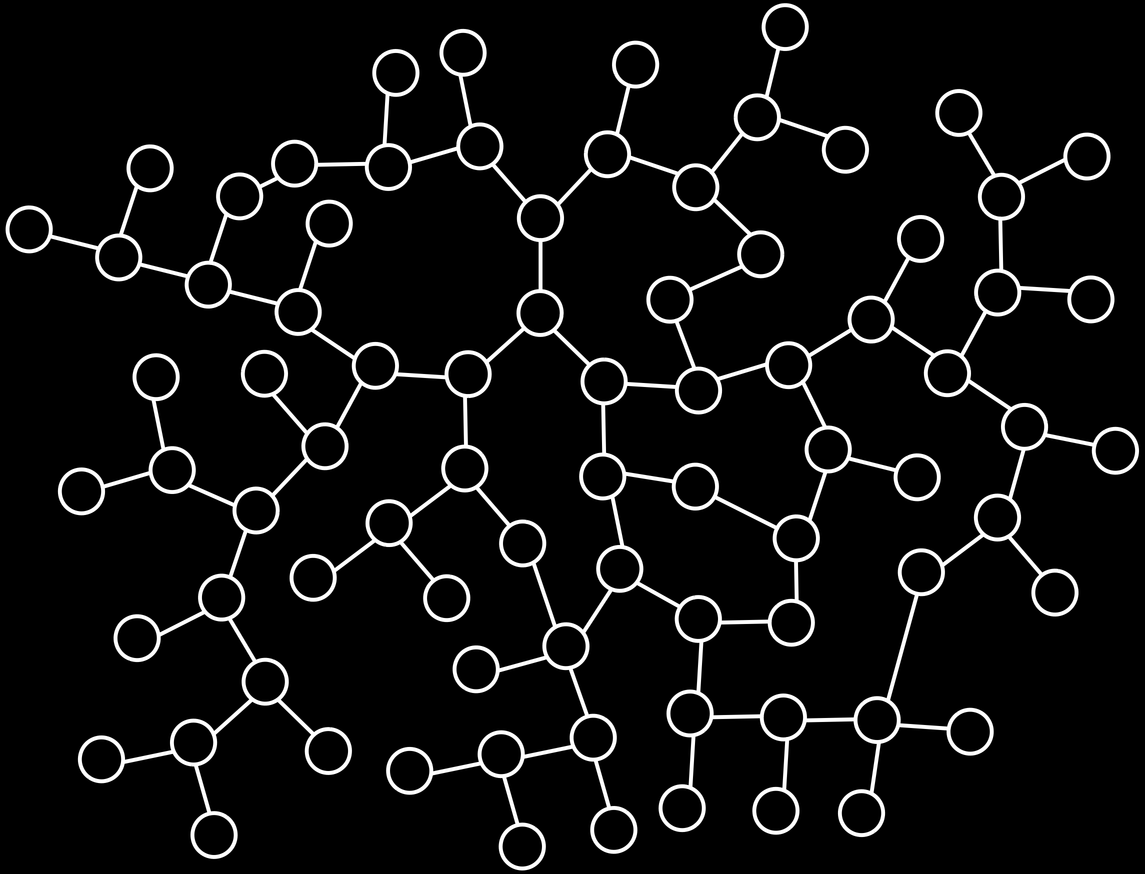
[*Shah, Zaman* 2011]

Our goal

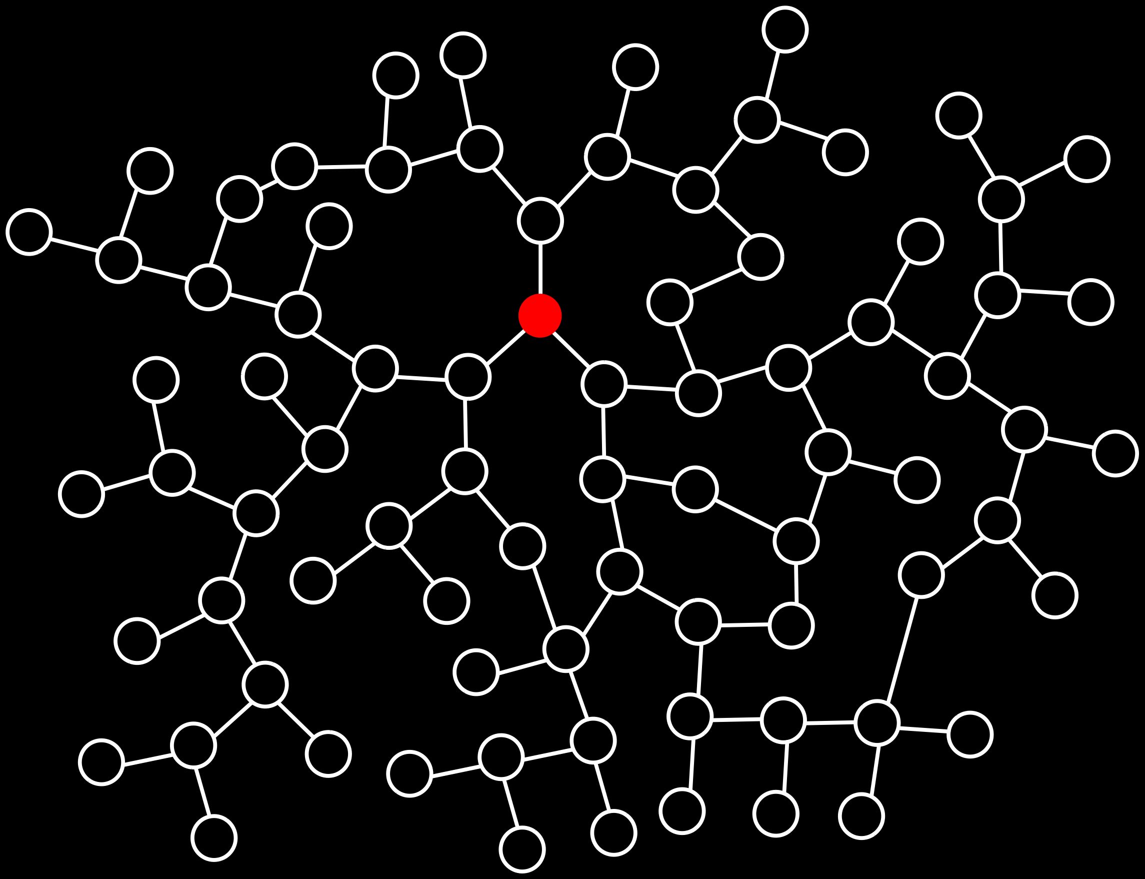


- N_T: **expected number** of nodes with the message at time T

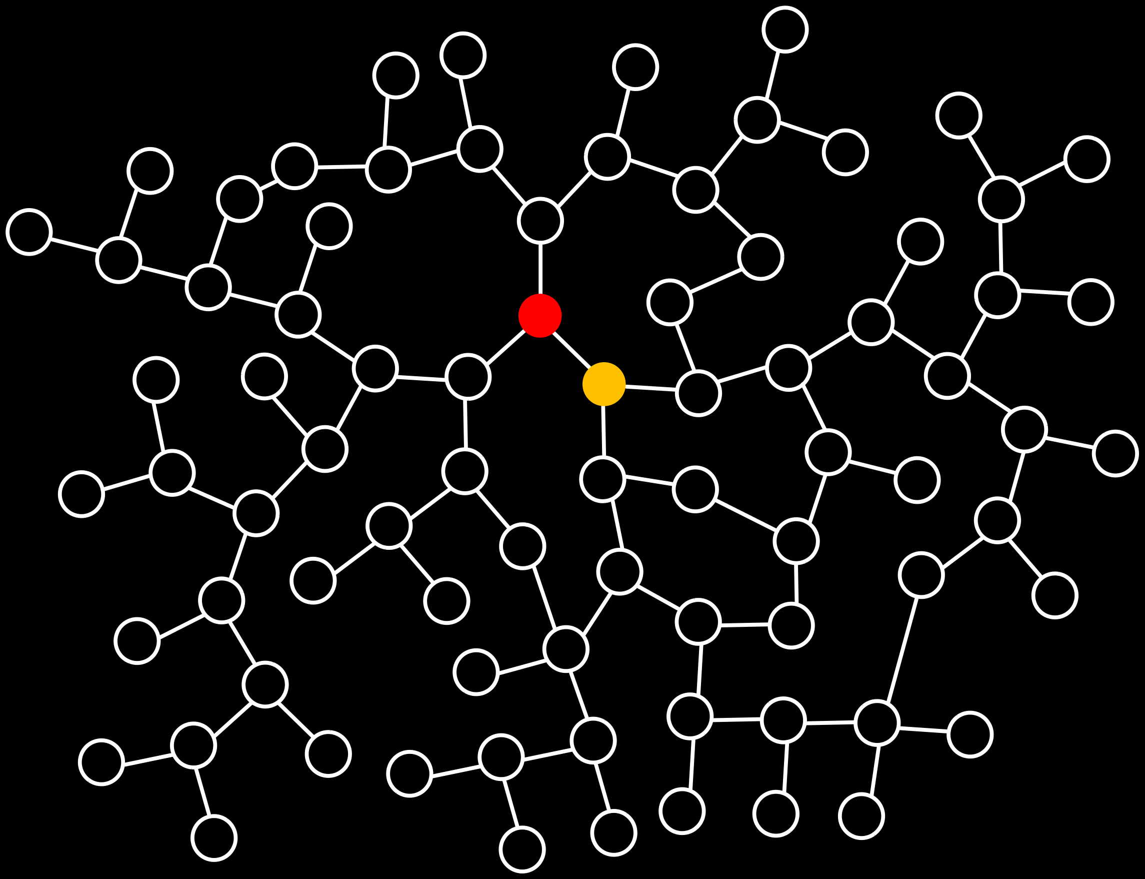
Main result: adaptive diffusion



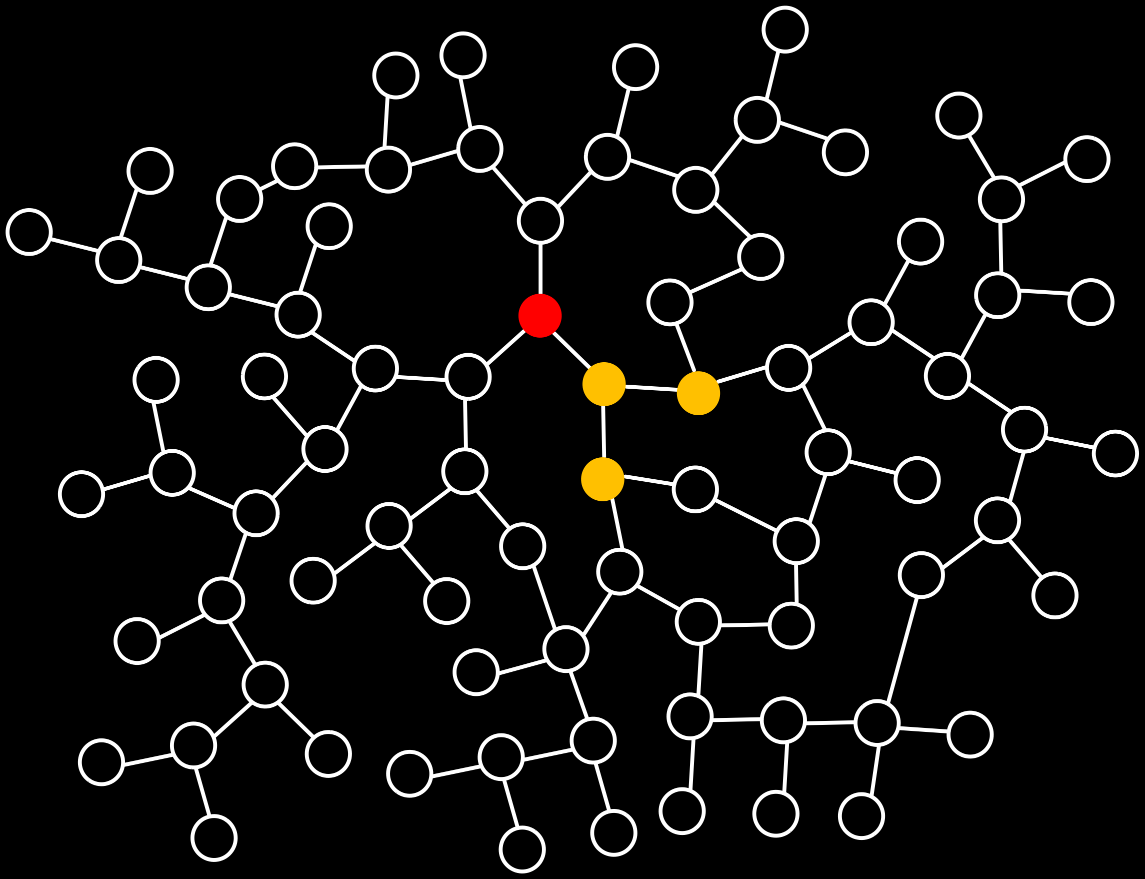
Main result: adaptive diffusion



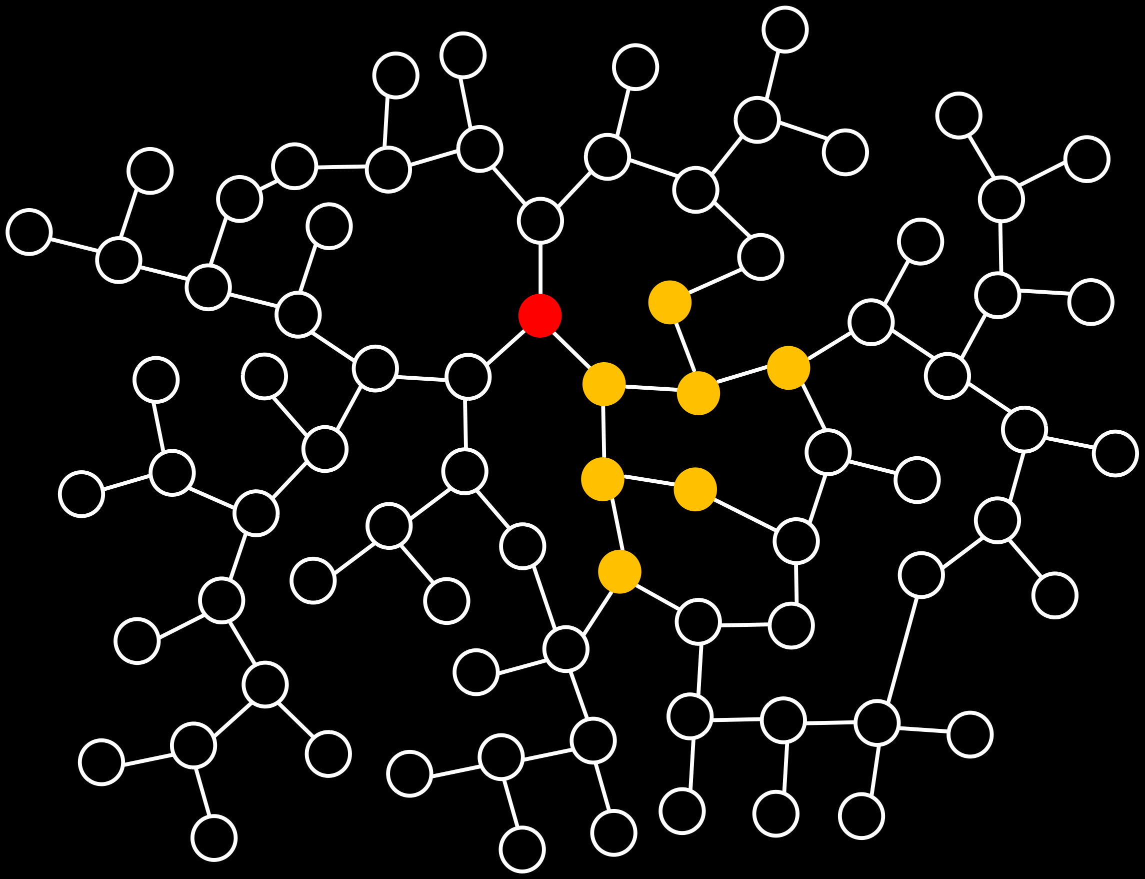
Main result: adaptive diffusion



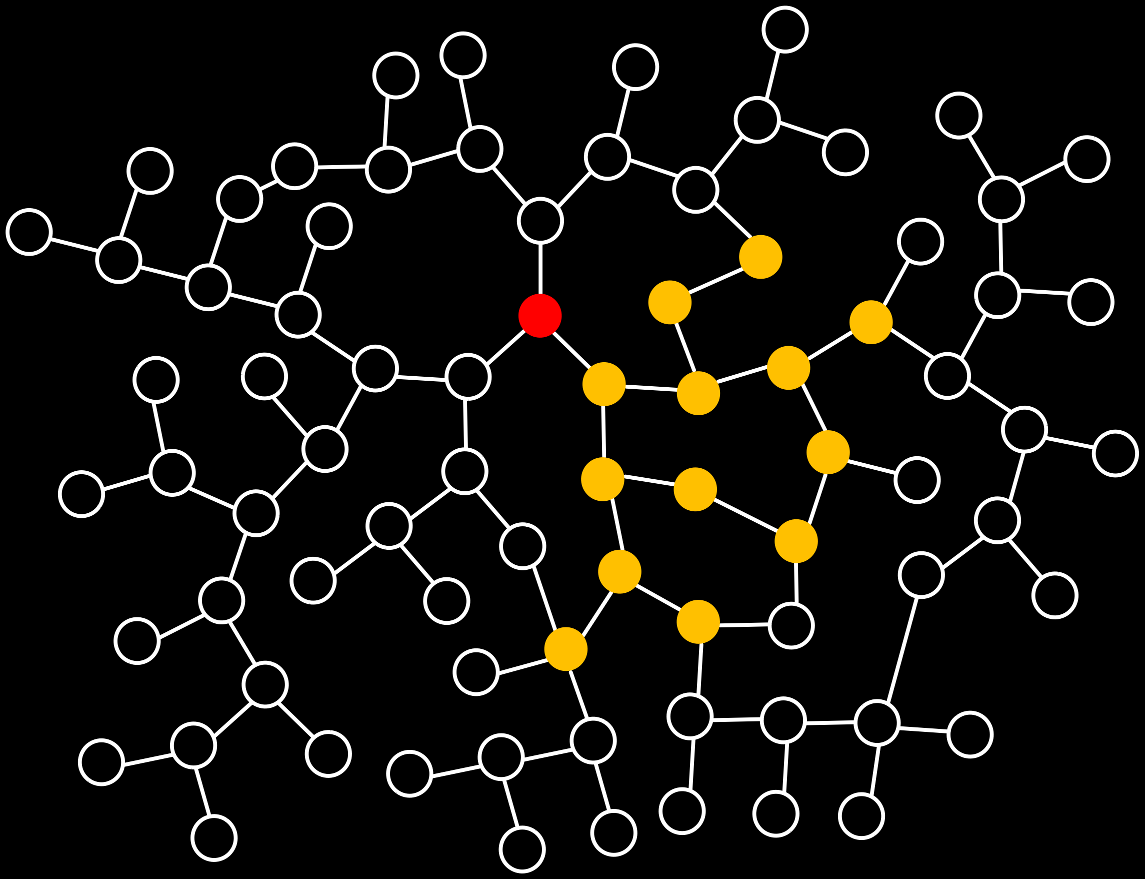
Main result: adaptive diffusion



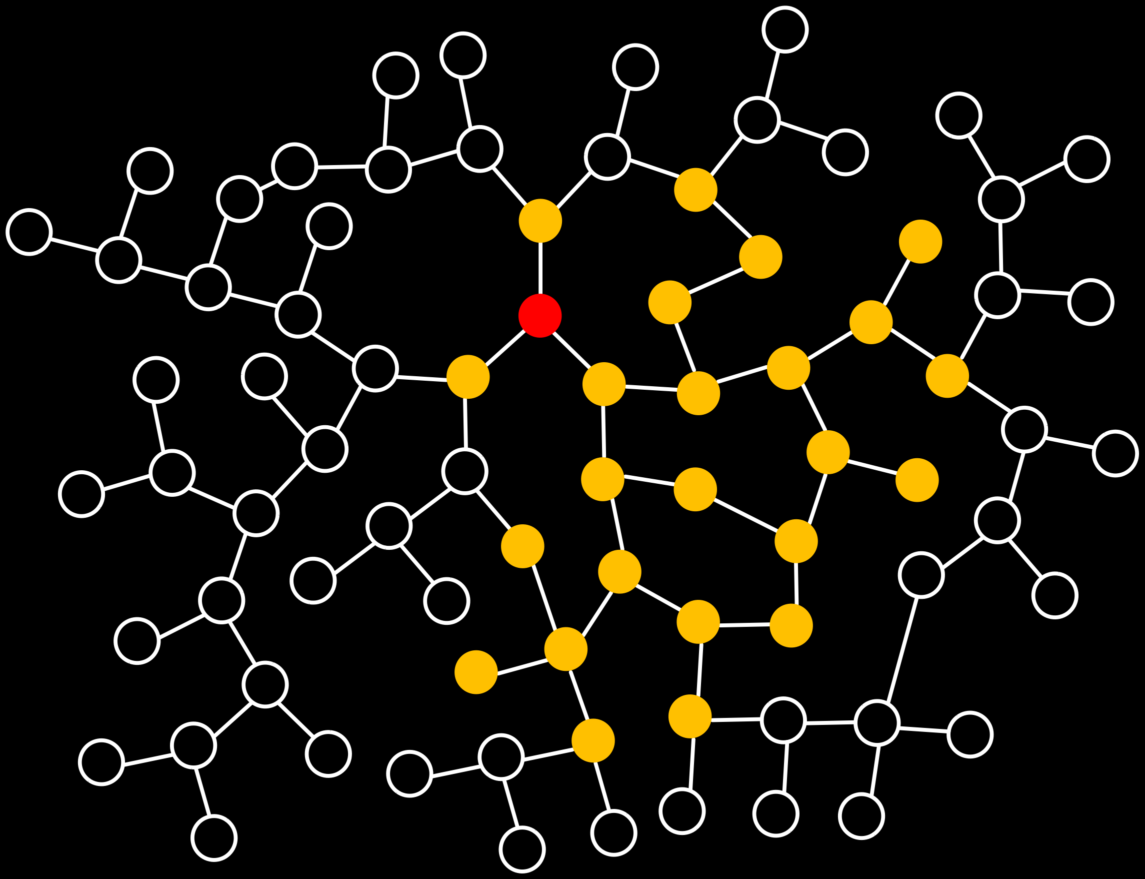
Main result: adaptive diffusion



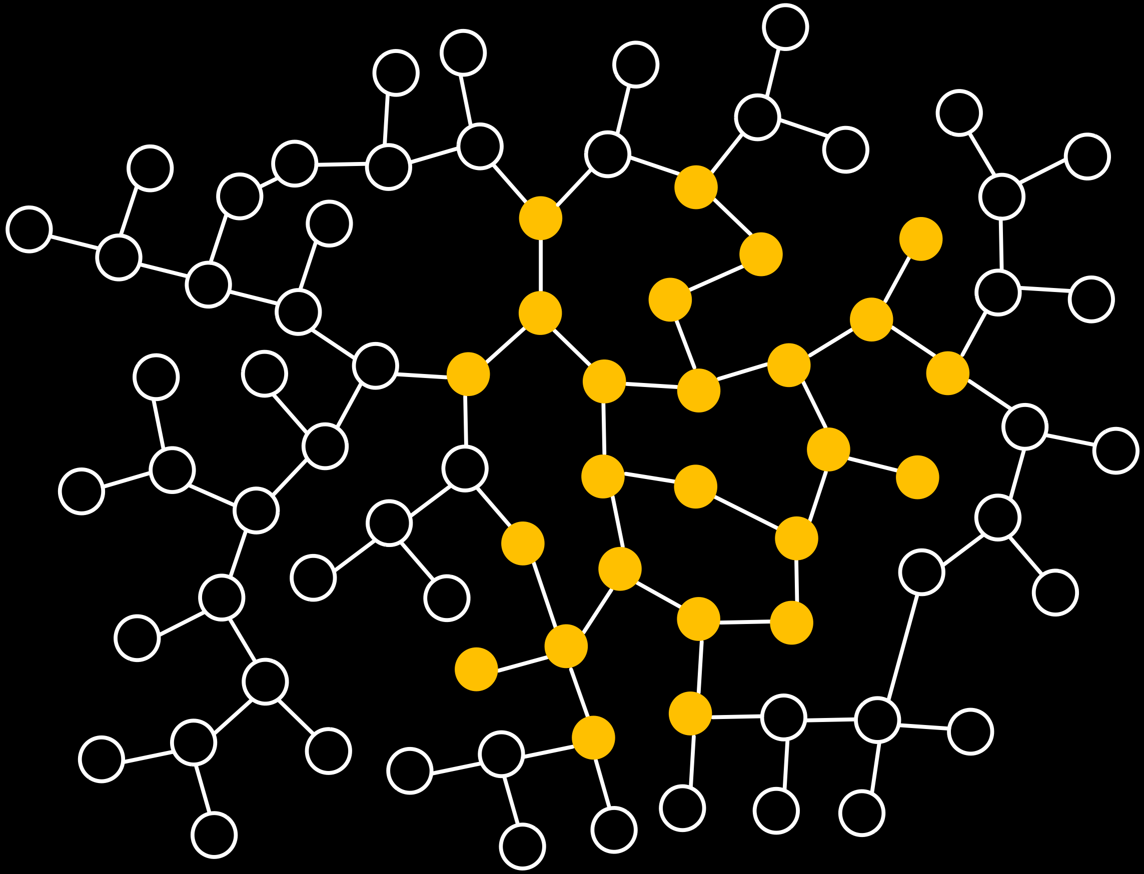
Main result: adaptive diffusion



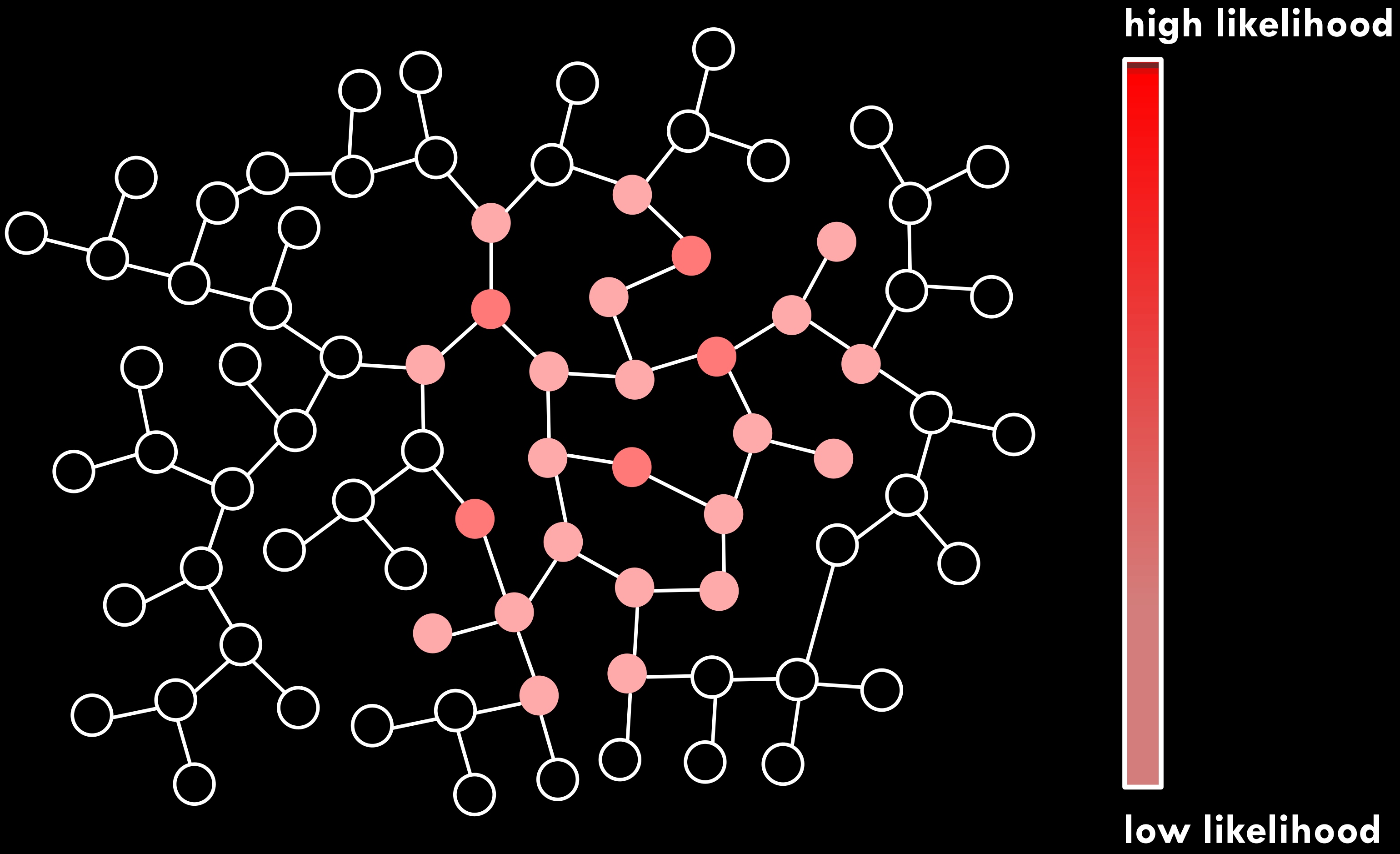
Main result: adaptive diffusion



Main result: adaptive diffusion

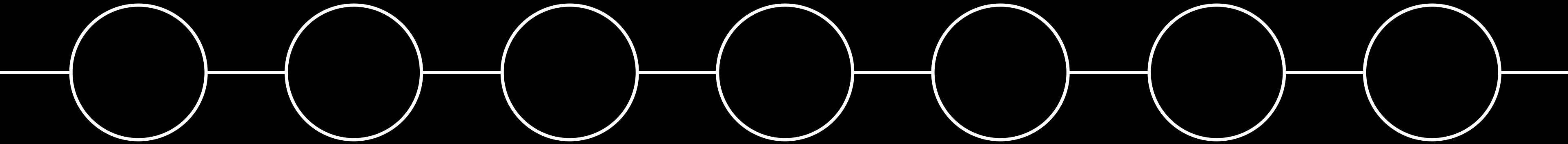


Main result: adaptive diffusion



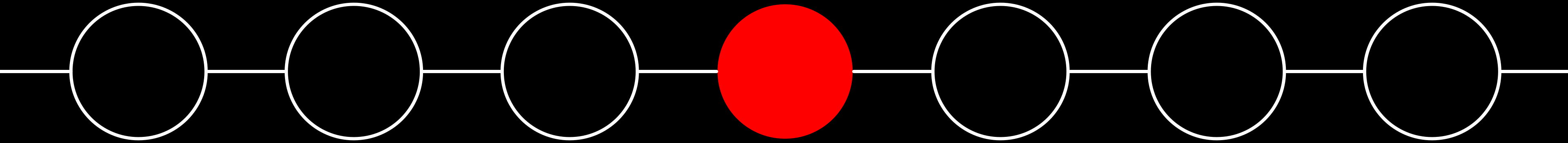
provides provable anonymity guarantees!

Line graphs



- let’s start with line graphs

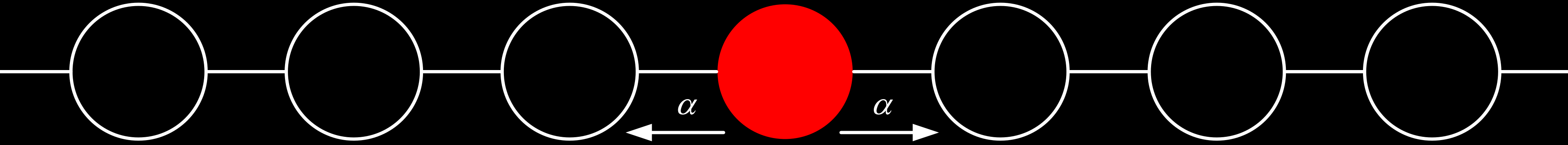
Line graphs: diffusion



$T = 0$

- the message author starts a rumor at $T = 0$

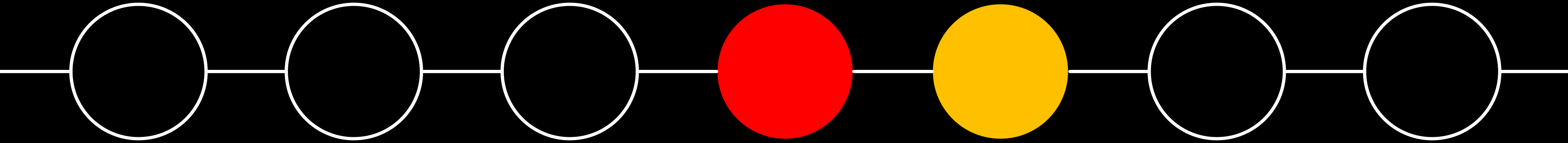
Line graphs: diffusion



$T = 1$

- with probability α, the left (right) node receives the message

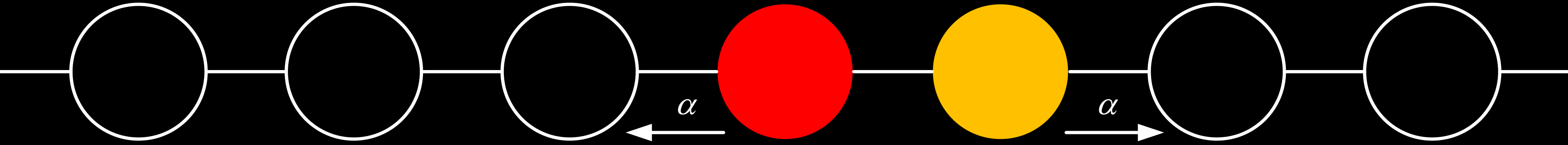
Line graphs: diffusion



$T = 1$

- the node to the right of the author receives the message

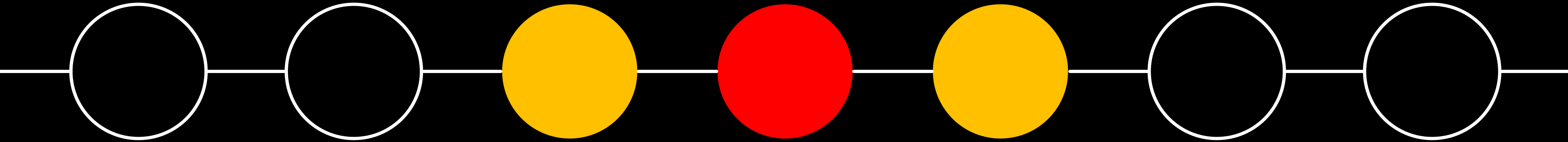
Line graphs: diffusion



$$T = 2$$

- the rumor propagates in **both directions** at the **same rate**

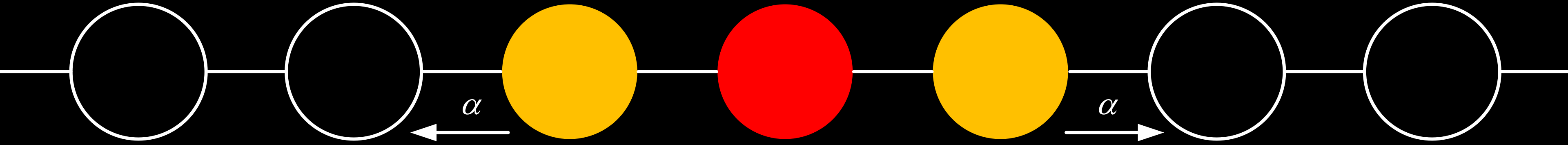
Line graphs: diffusion



$T = 2$

- the rumor propagates in **both directions** at the **same rate**

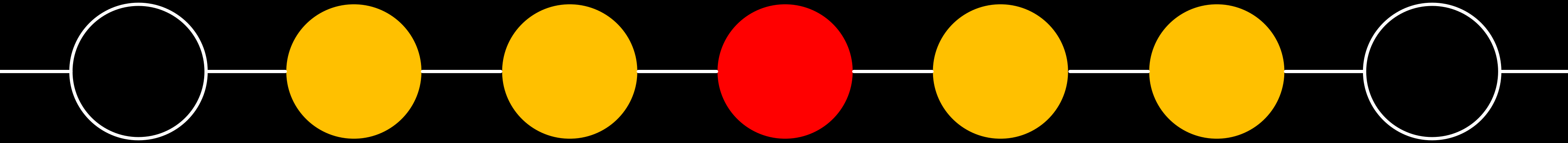
Line graphs: diffusion



$T = 3$

- α is **independent of time or hop distance** to message author

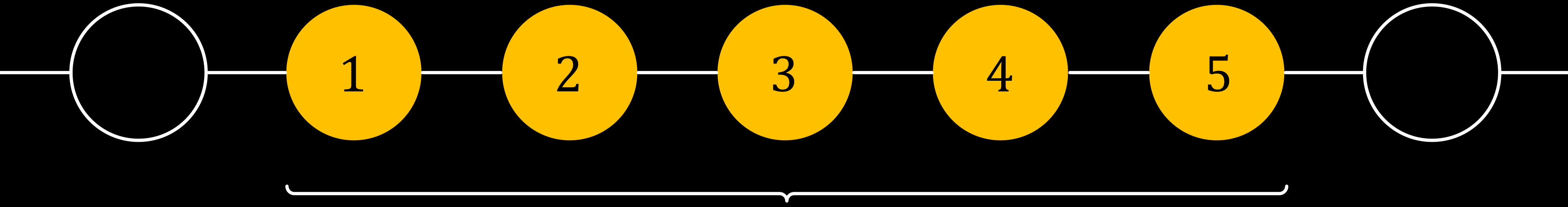
Line graphs: diffusion



$T = 3$

- diffusion on a line is equivalent to **two independent random walks**

Adversary’s observation

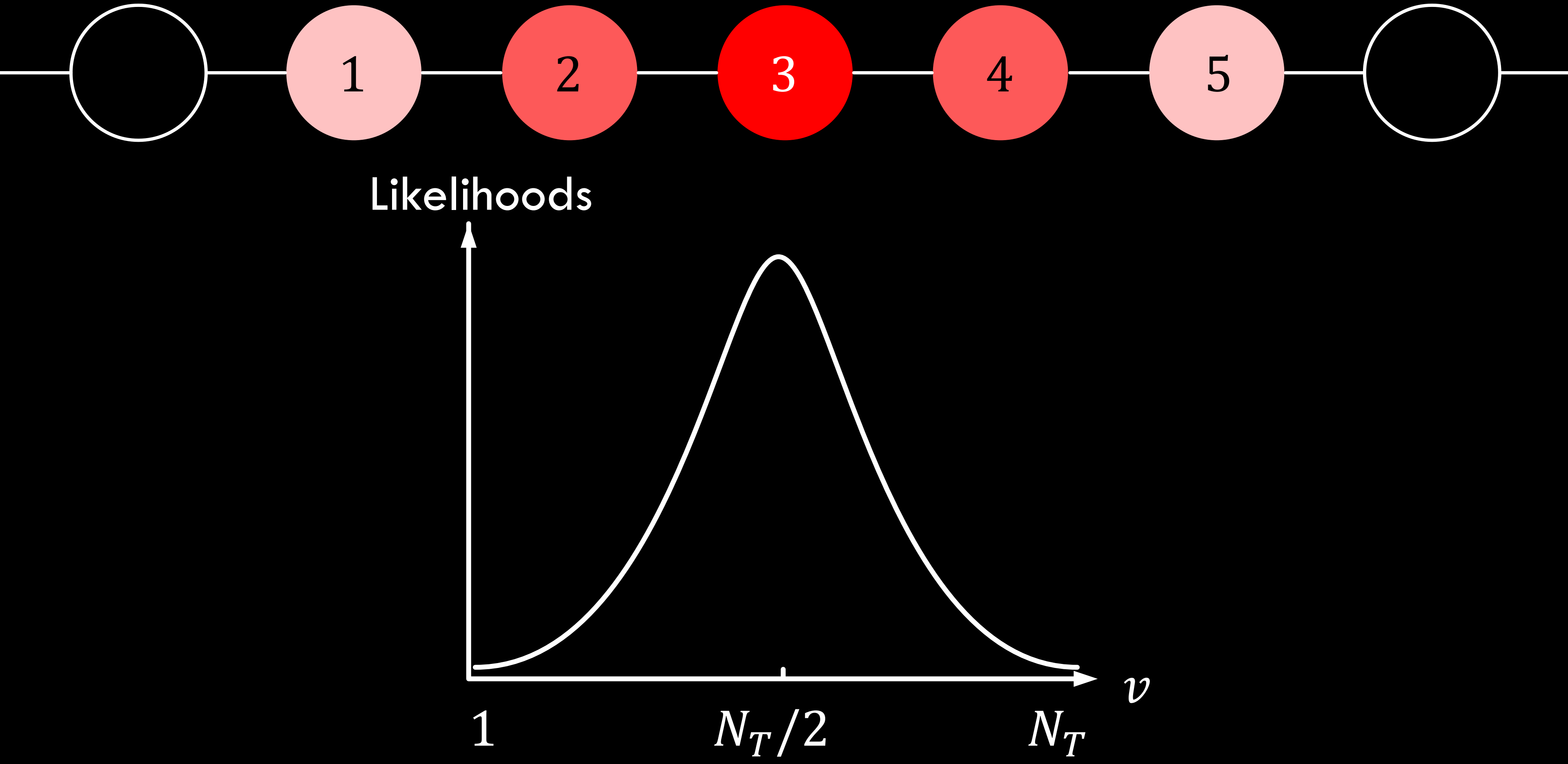


$N = 5$

nodes with the message

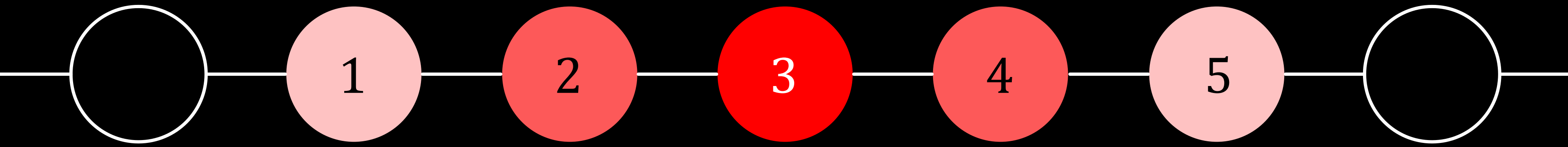
can the adversary locate the message author?

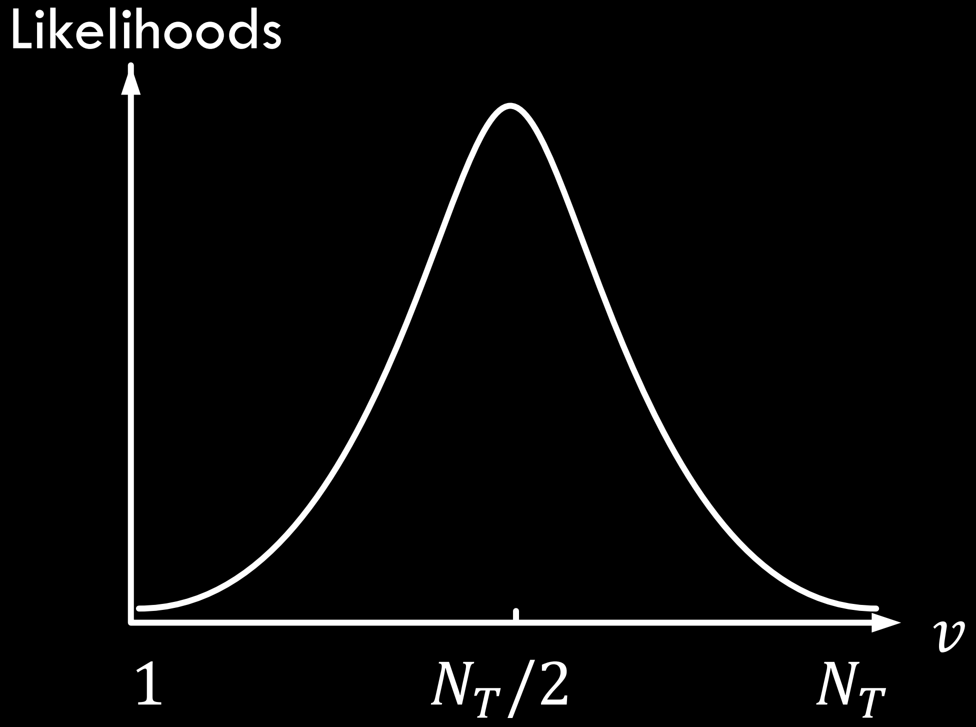
Maximum likelihood detection



- the **node in the middle** is the **mostly likely author**

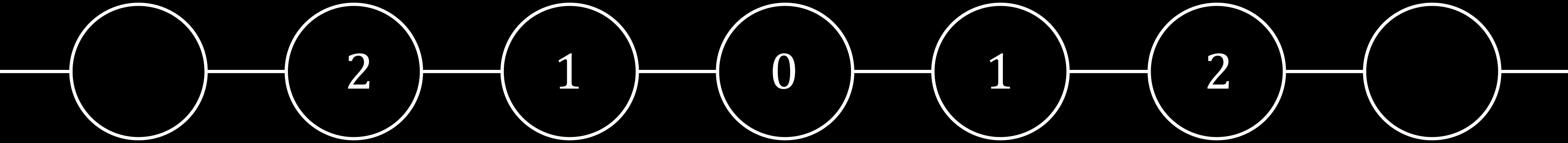
Maximum likelihood detection





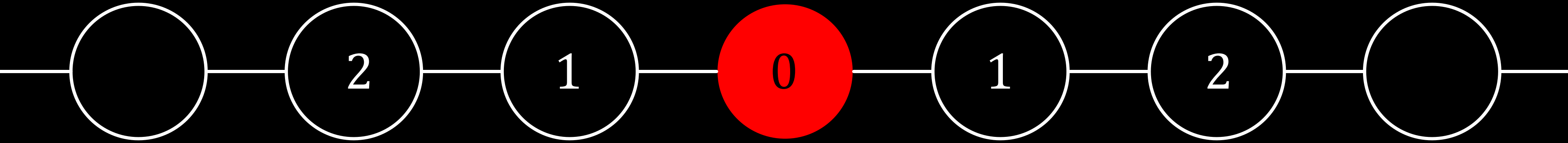
Probability of detection $\approx \frac{1}{\sqrt{N_T}}$

Line graphs: adaptive diffusion



- consider a line graph

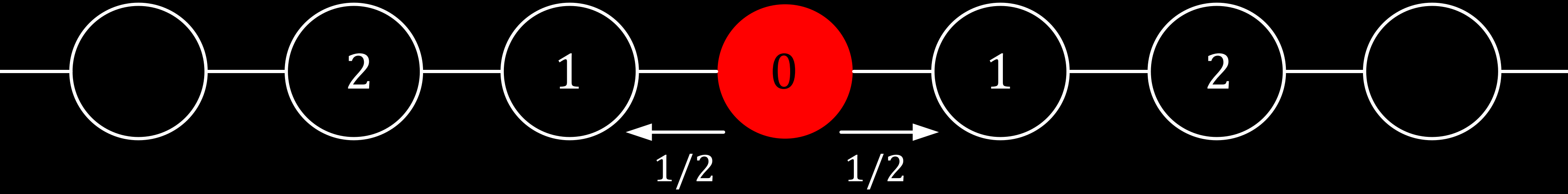
Line graphs: adaptive diffusion



$T = 0$

- node 0 starts a rumor at $T = 0$

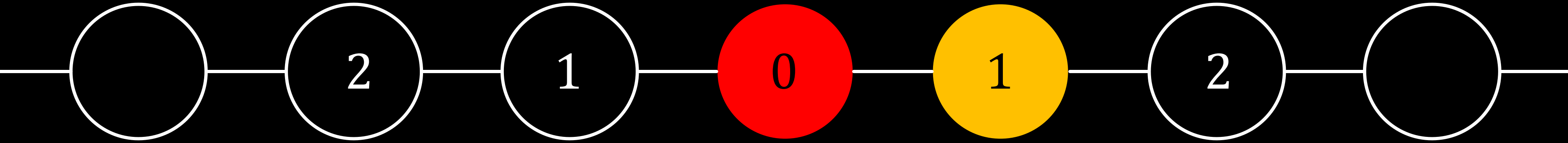
Line graphs: adaptive diffusion



$T = 1$

- with probability $1/2$, the left (right) node receives the message

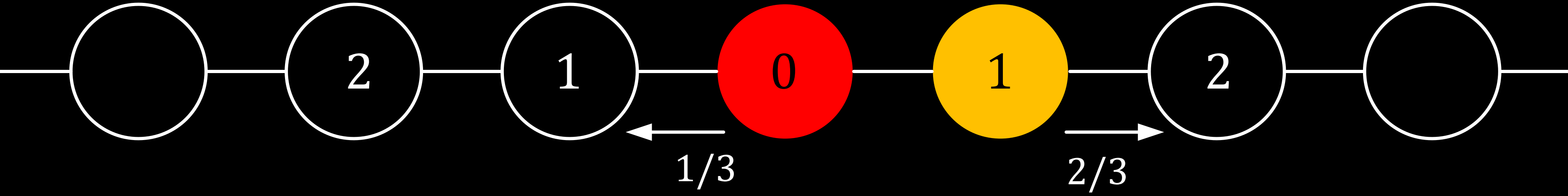
Line graphs: adaptive diffusion



$T = 1$

- right node 1 receives the message

Line graphs: adaptive diffusion



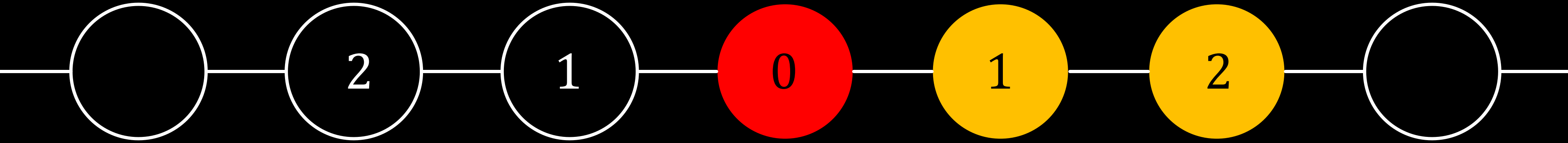
$T = 2$

hop distance to message author

- probability of passing message: $\alpha = \frac{h+1}{T+1}$

elapsed time

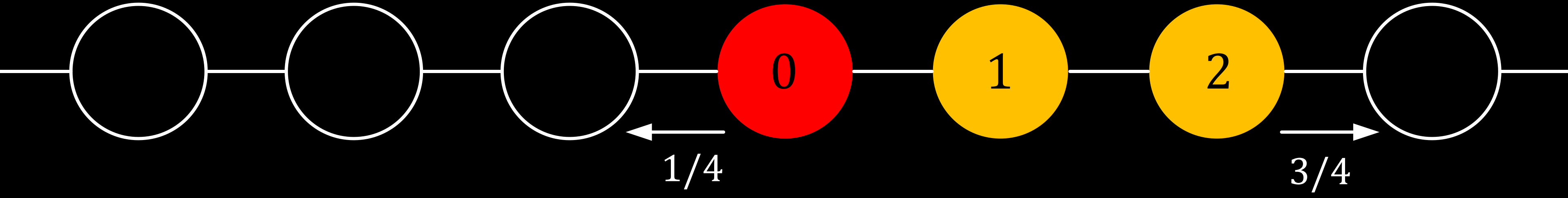
Line graphs: adaptive diffusion



$T = 2$

- right node 2 receives the message

Line graphs: adaptive diffusion



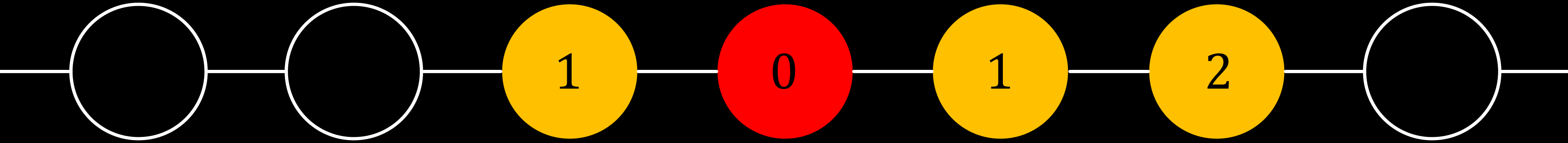
$T = 3$

- probability of passing message: $\alpha = \frac{h+1}{T+1}$

hop distance to message author

elapsed time

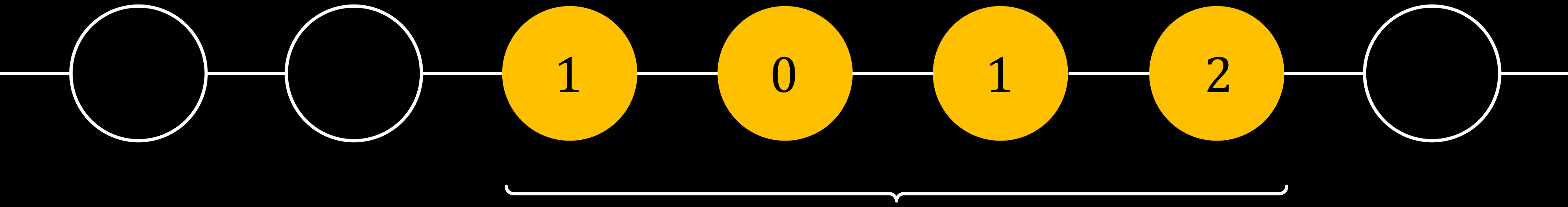
Line graphs: adaptive diffusion



$T = 3$

- left node 1 receives the message

Adversary’s observation

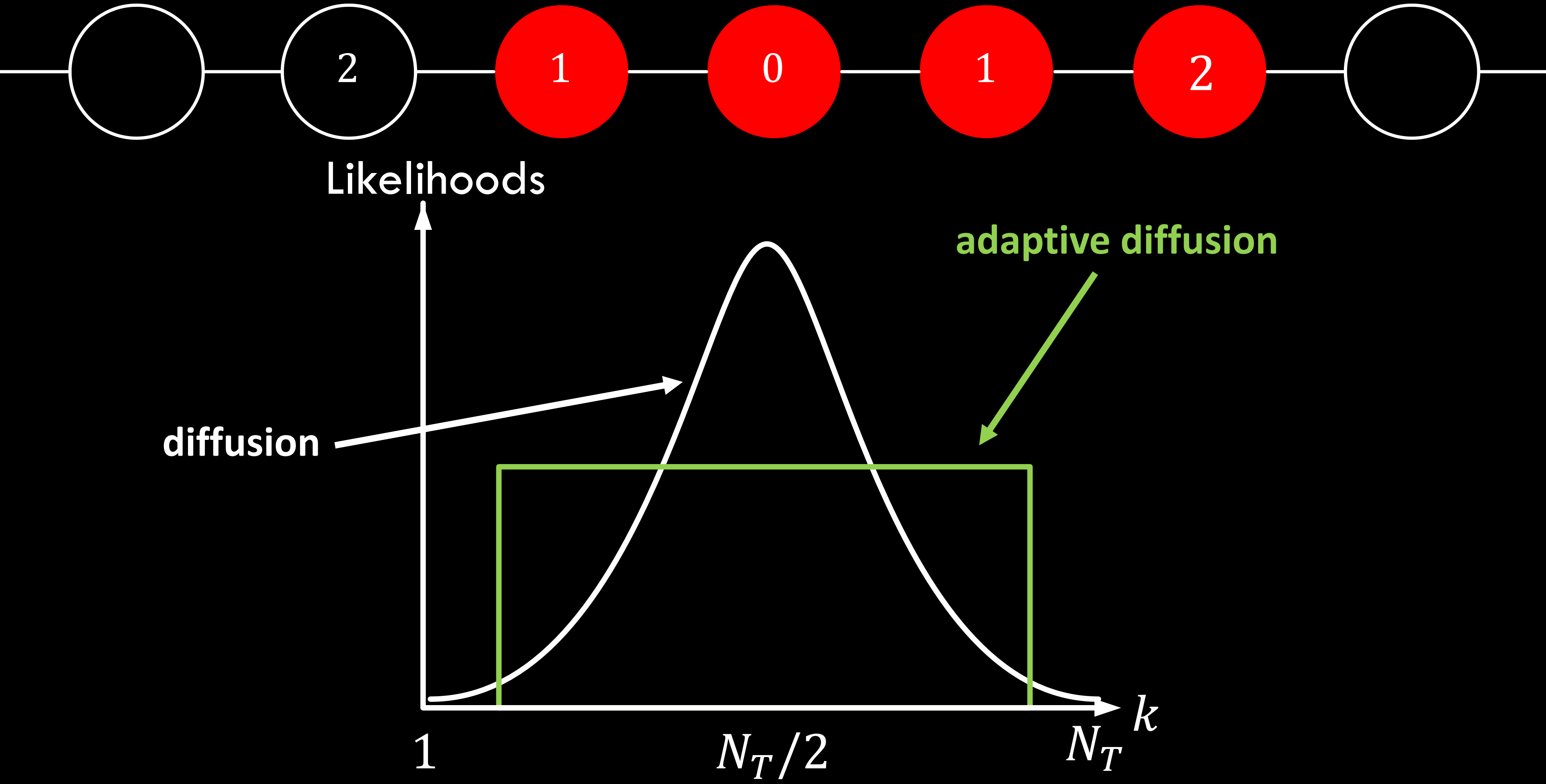


$N_T = 4$

nodes with the message

can the adversary locate the message author?

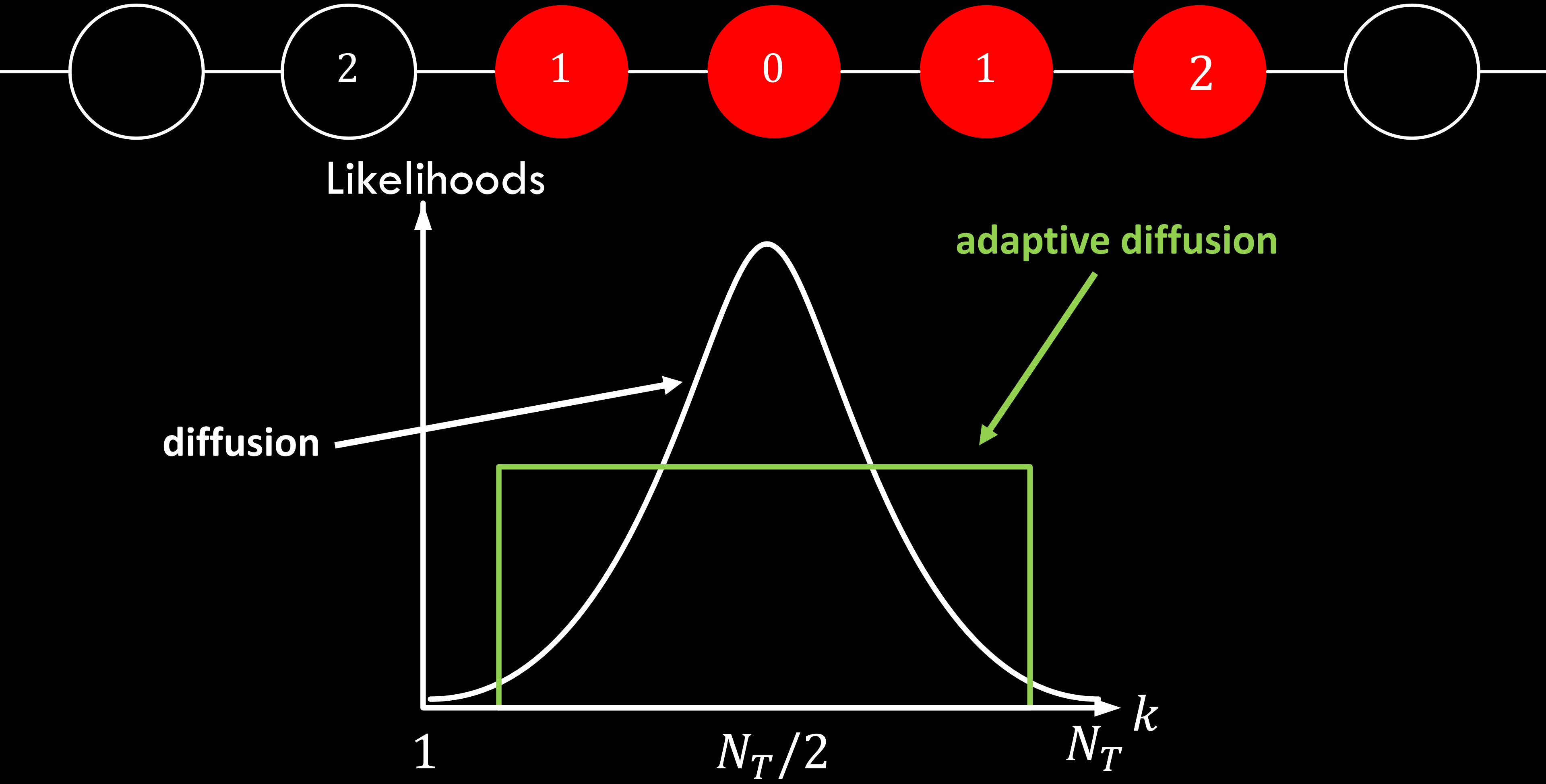
Maximum likelihood detection



diffusion

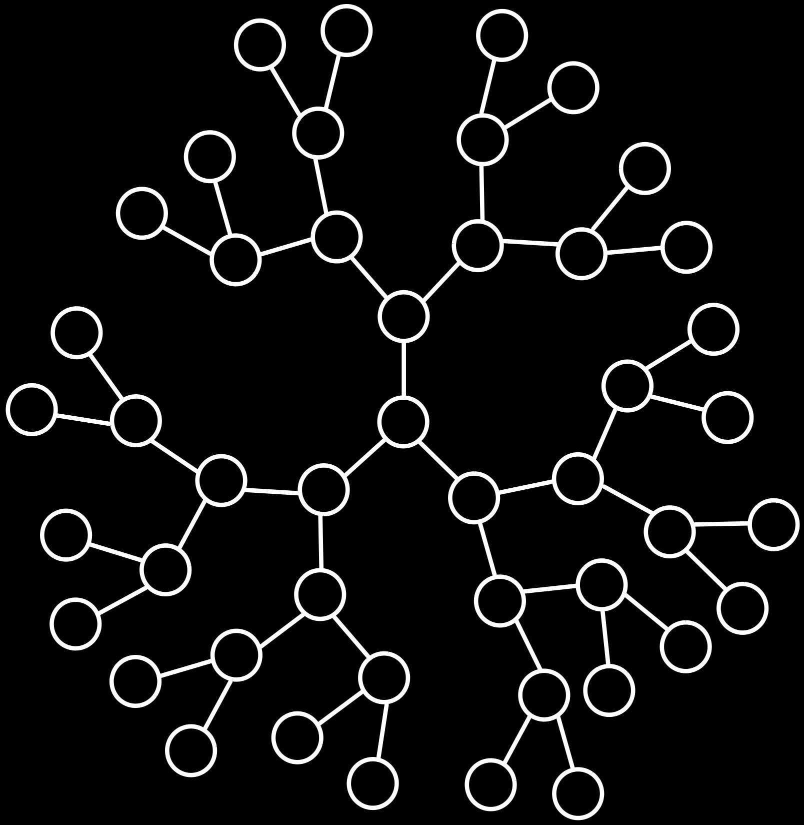
adaptive diffusion

Maximum likelihood detection



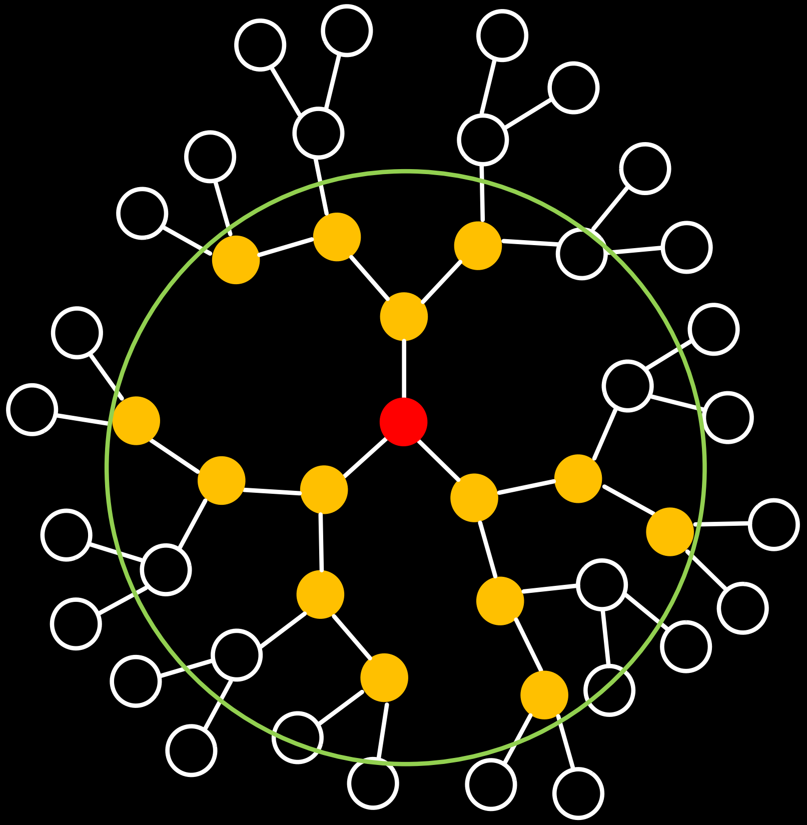
Probability of detection $\approx \frac{1}{N_T}$

d-regular trees



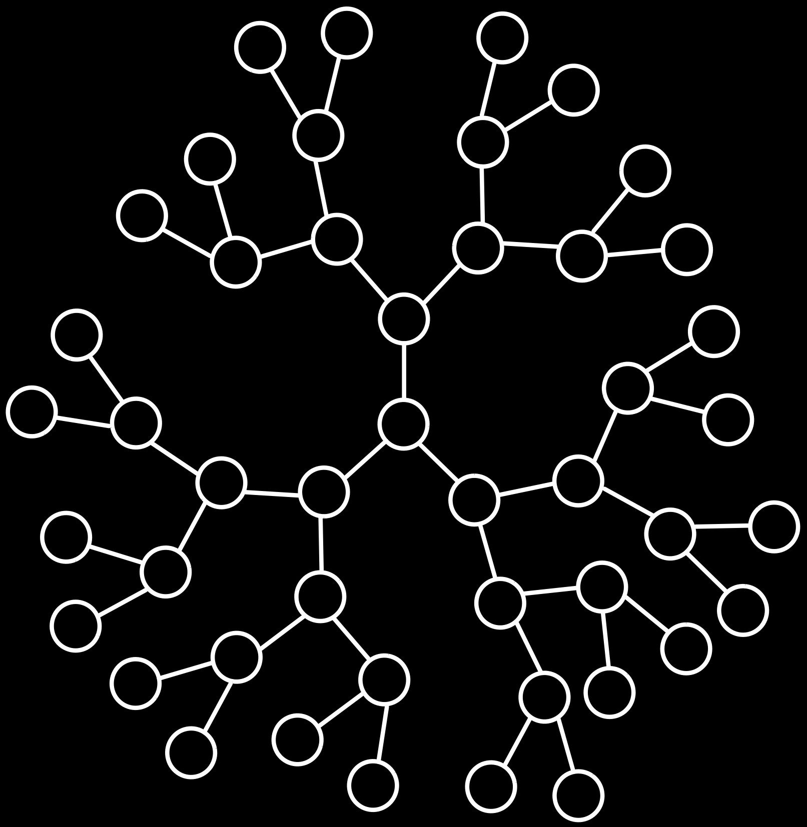
- what about d-regular trees?

d-regular trees: diffusion

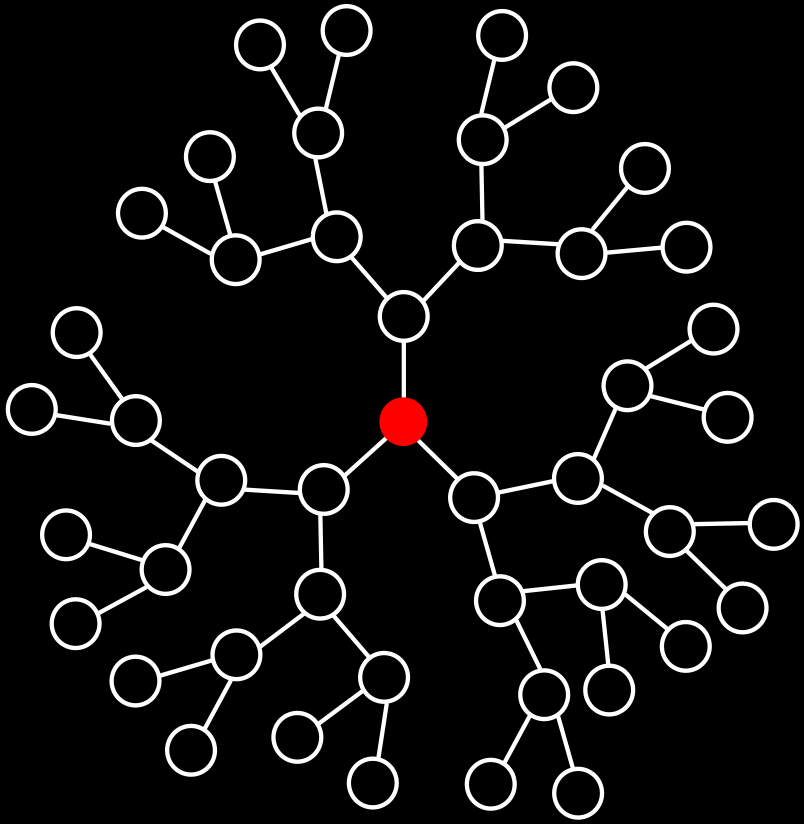


- likelihoods concentrate heavily around the "center"

d-regular trees : adaptive diffusion

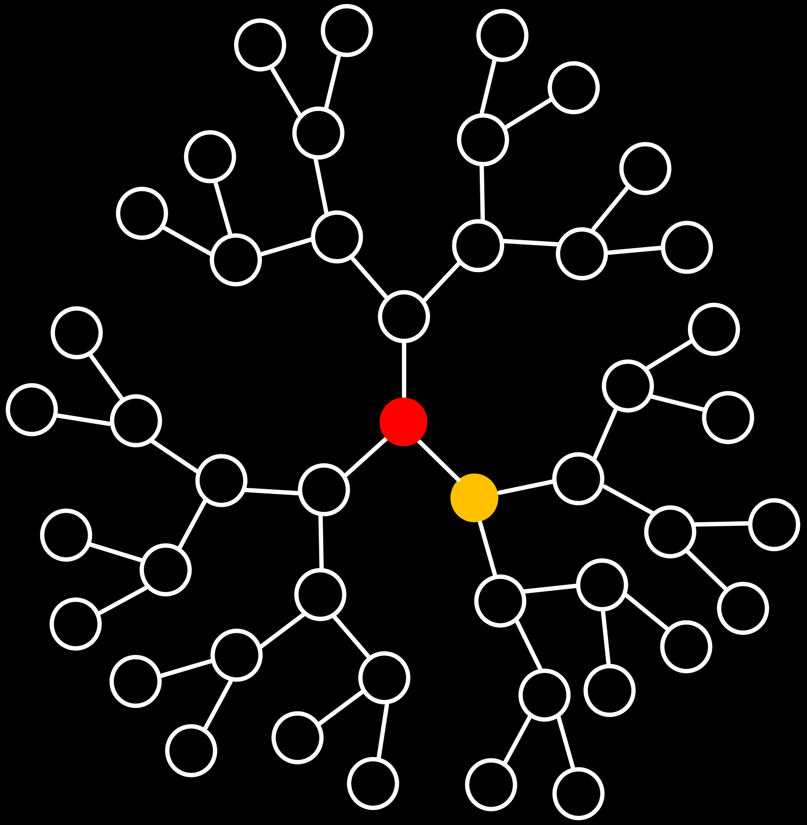


d-regular trees : adaptive diffusion



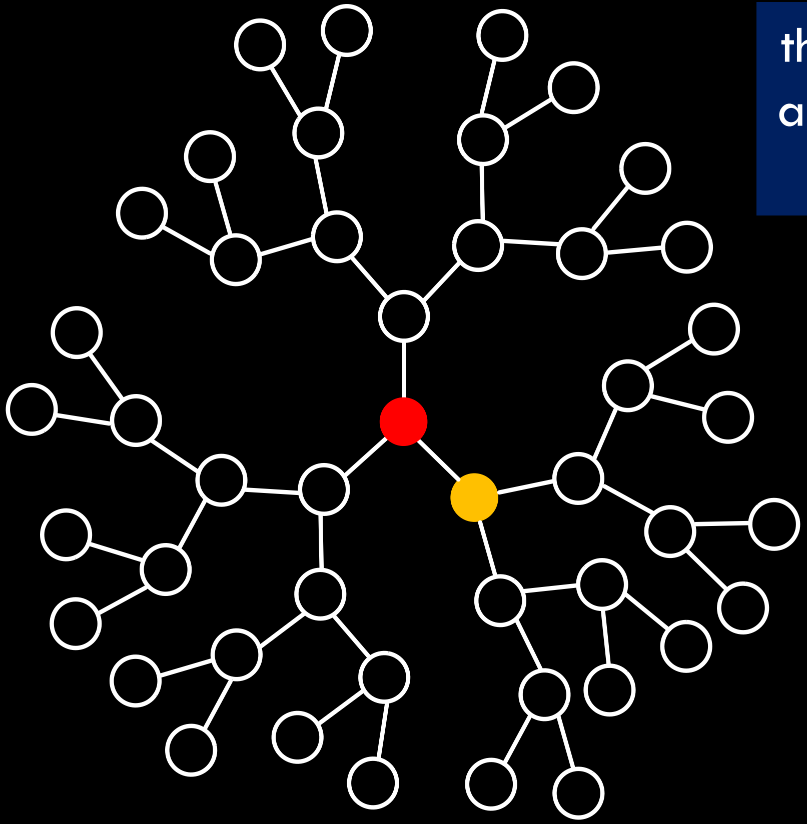
- initially, the source is also the **"virtual source"**

d-regular trees : adaptive diffusion



- at $T = 1$, the author selects one neighbor at random

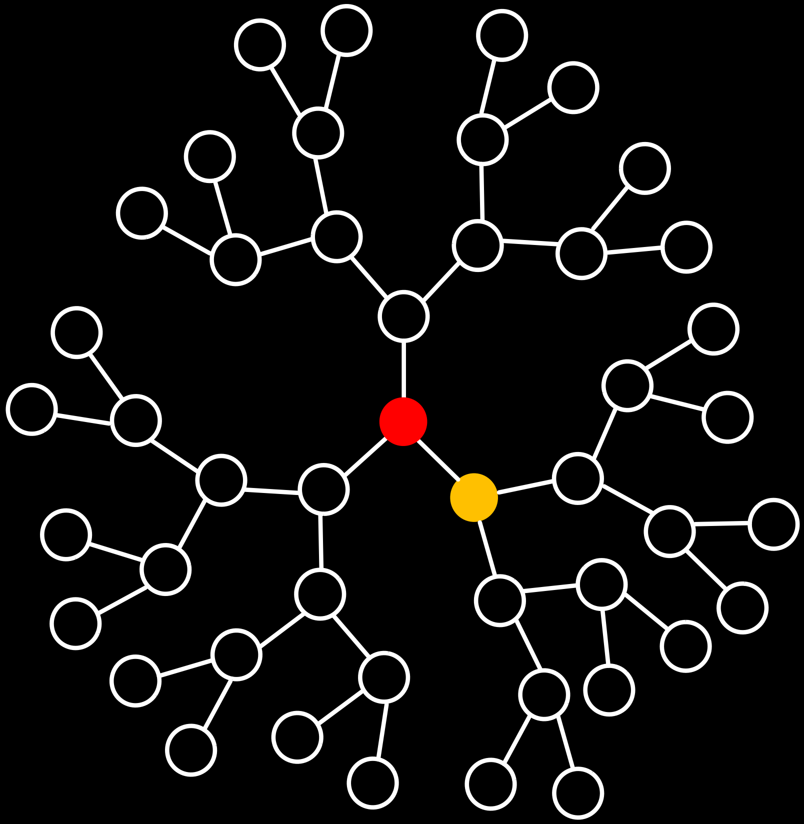
d-regular trees : adaptive diffusion



the author passes $h = 1$ and $T = 2$ to the chosen neighbor

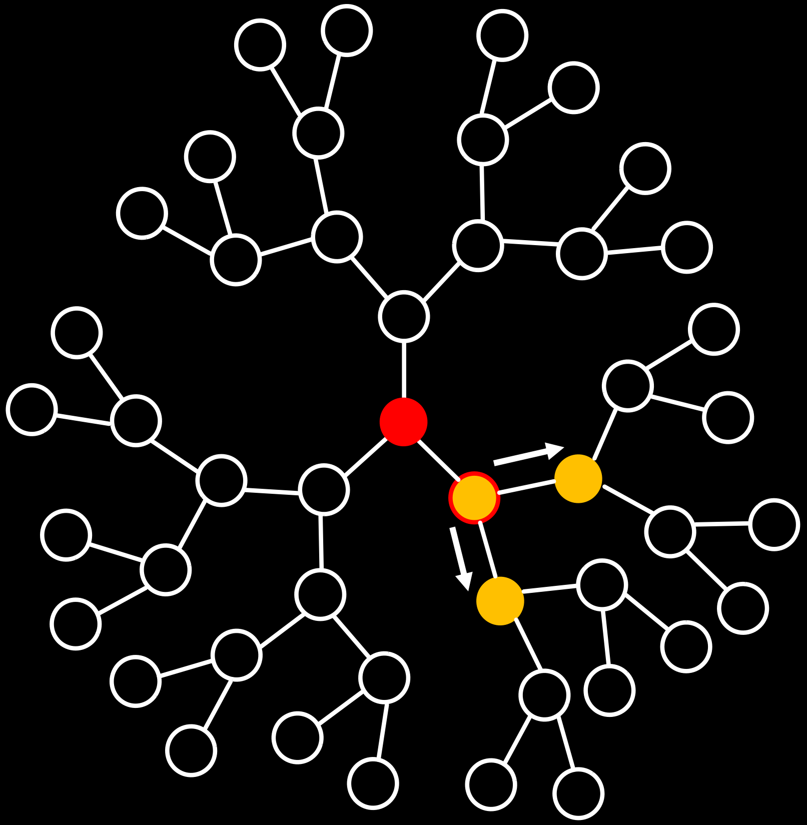
- at $T = 1$, the author selects one neighbor at random

d-regular trees : adaptive diffusion



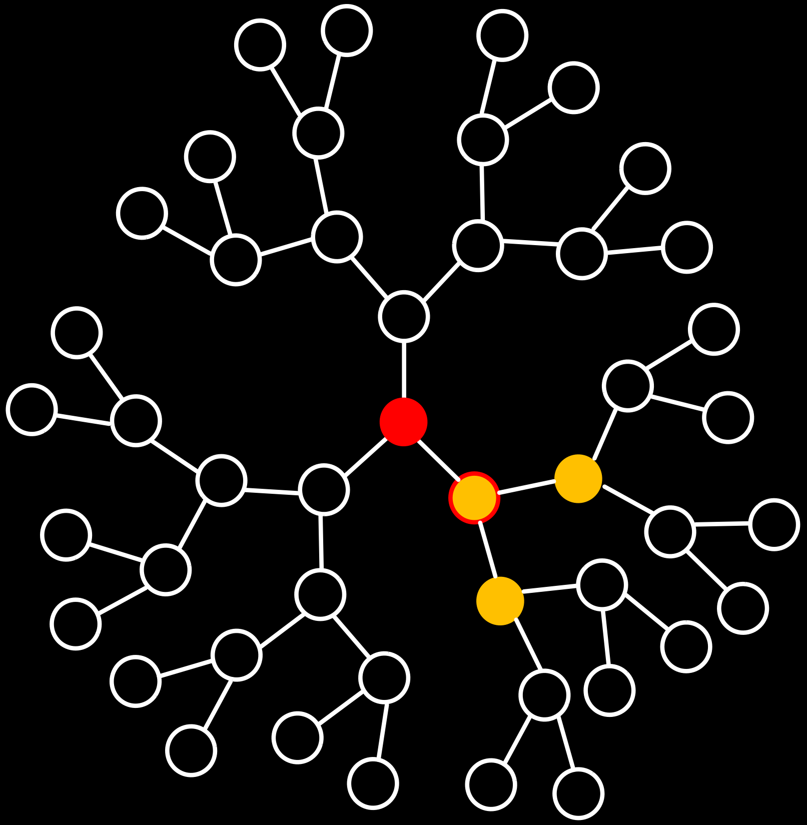
- the chosen neighbor becomes the new virtual source

d-regular trees : adaptive diffusion



- at $T = 2$, the **virtual source** passes the message to all her neighbors

d-regular trees : adaptive diffusion

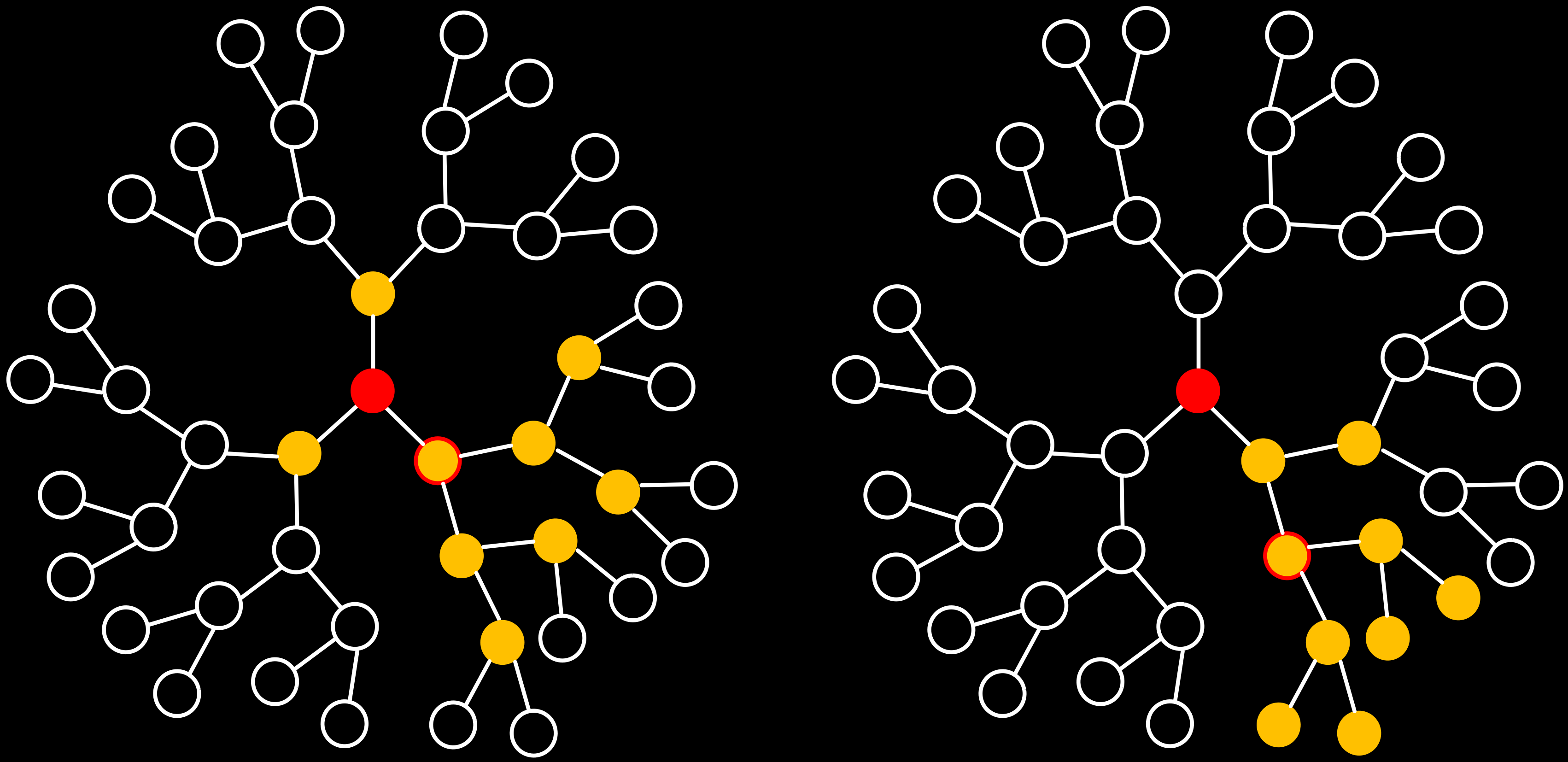


- as T transitions from even to odd, the virtual source has two options:

keeping the virtual source token

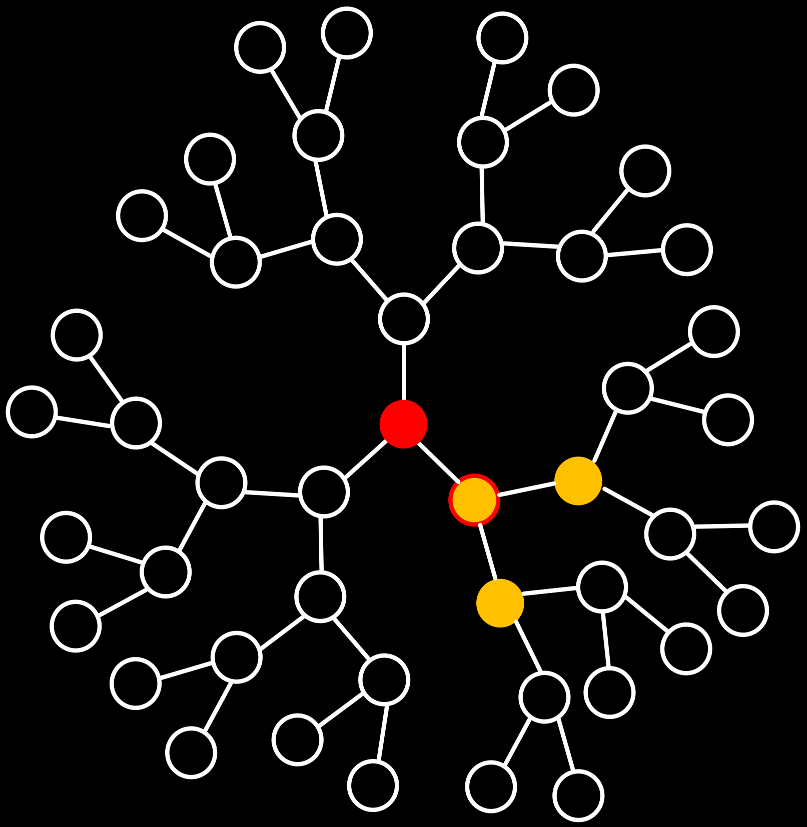
passing the virtual source token

Symmetry properties



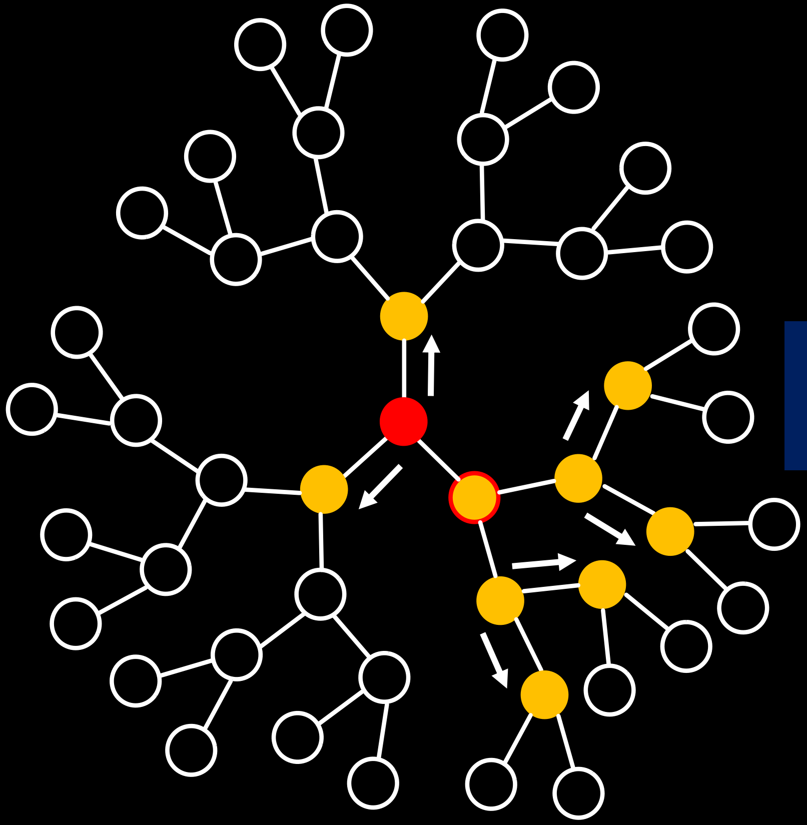
- the graph is **always symmetric** around the **virtual source**

Keeping the virtual source token



- virtual source token is kept with probability $\alpha = \frac{(d-1)^{\frac{T}{2}-h-1}-1}{(d-1)^{\frac{T}{2}+1}-1}$

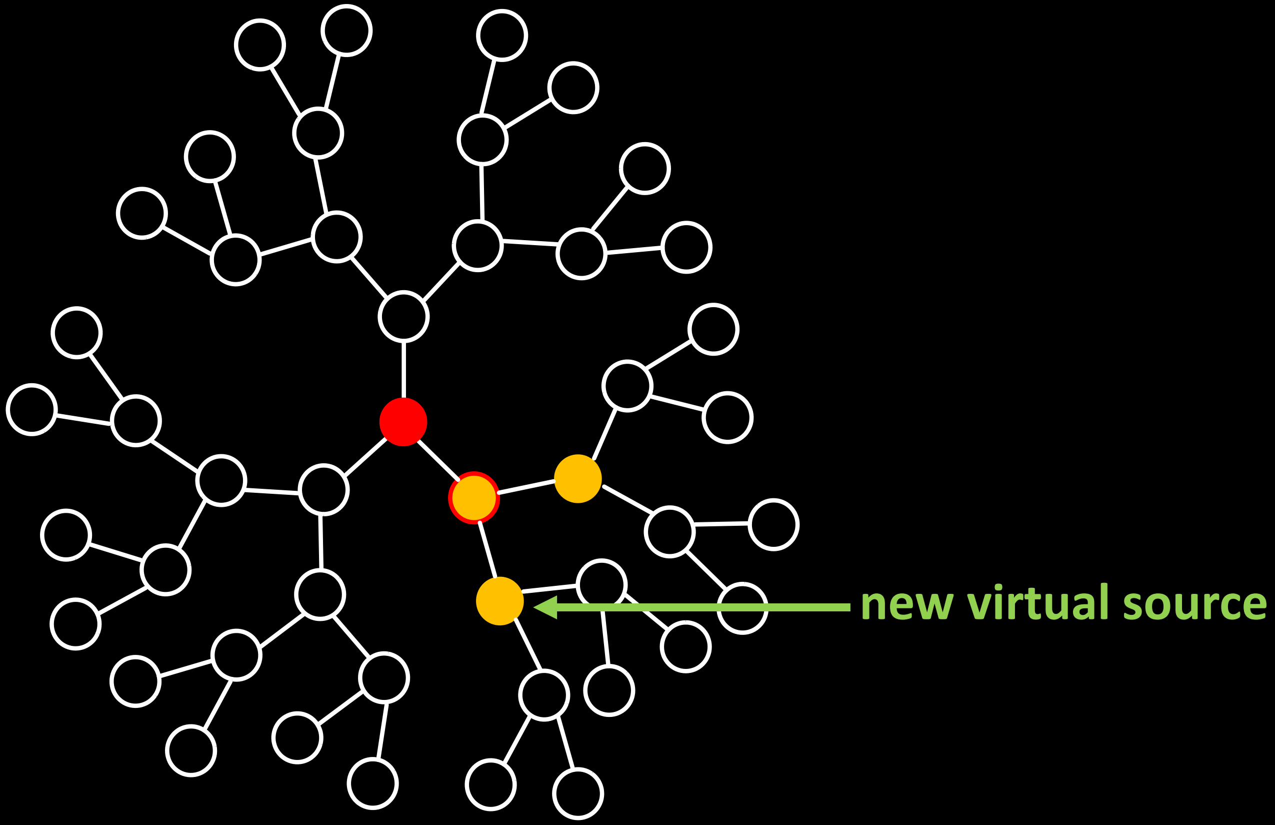
Keeping the virtual source token



happens in $T = 3$ and $T = 4$

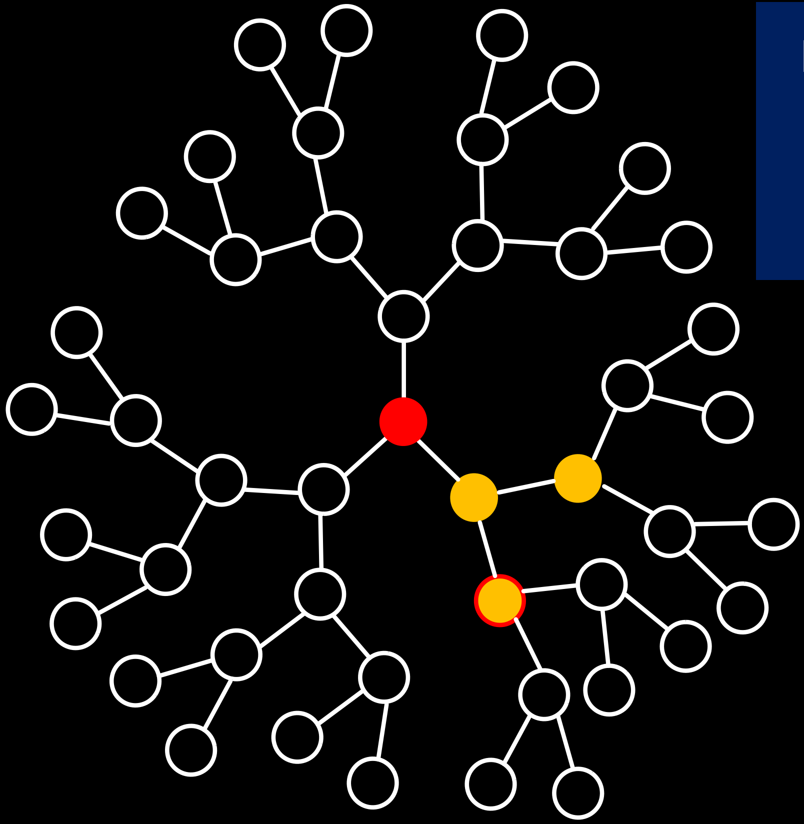
- all leaf nodes with the message pass it to their neighbors

Passing the virtual source token



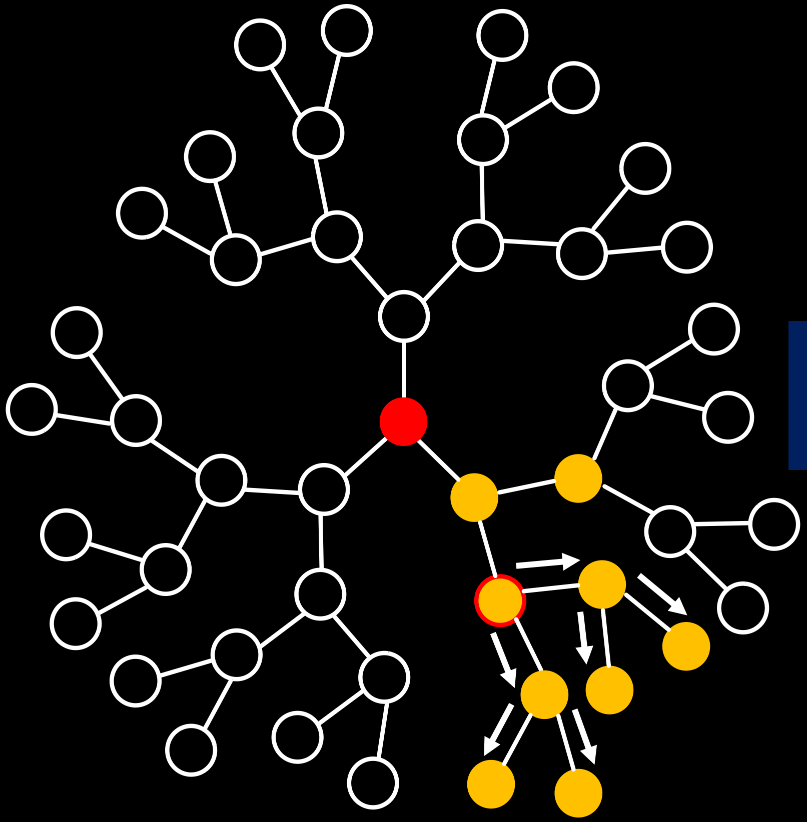
- current virtual source selects one of its neighbors at random

Passing the virtual source token



previous virtual source passes $h = 2$ and $T = 4$ to new virtual source

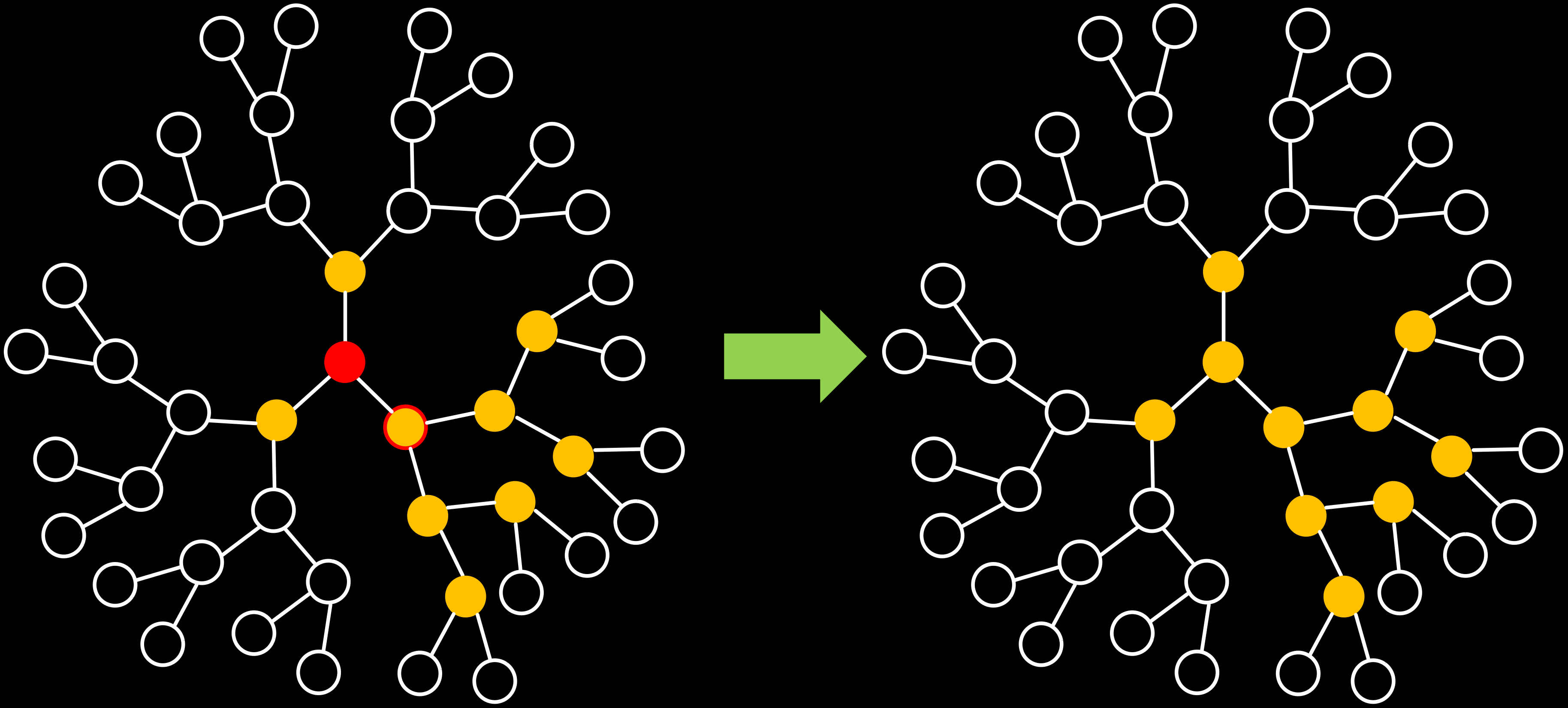
Passing the virtual source token



happens in $T = 3$ and $T = 4$

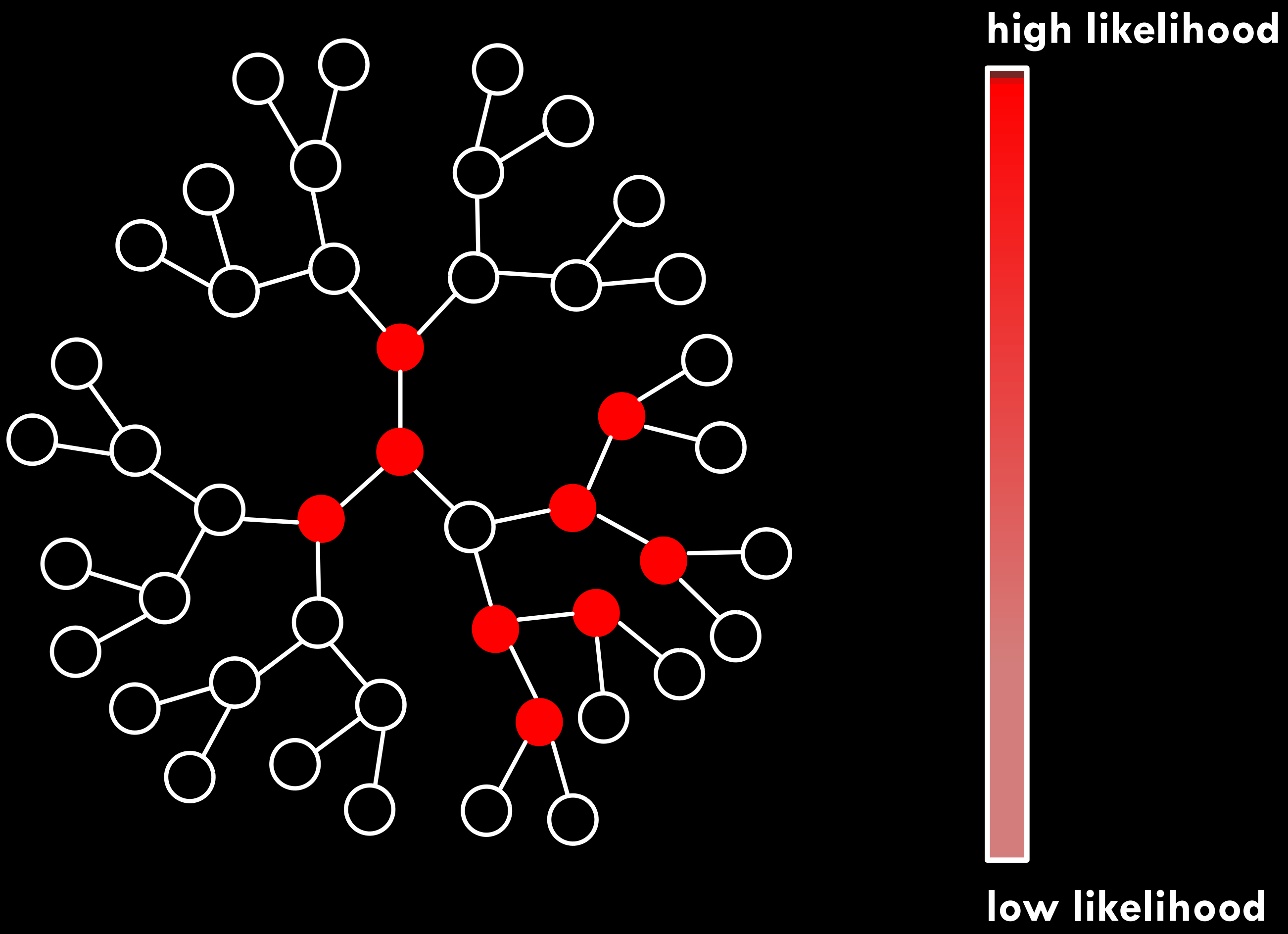
- new virtual source passes the message to its neighbors which in turn pass it to their neighbors

Adversary's observation



can the adversary locate the message author?

Maximum likelihood detection



- **all nodes** except for the final virtual source **are equally likely**

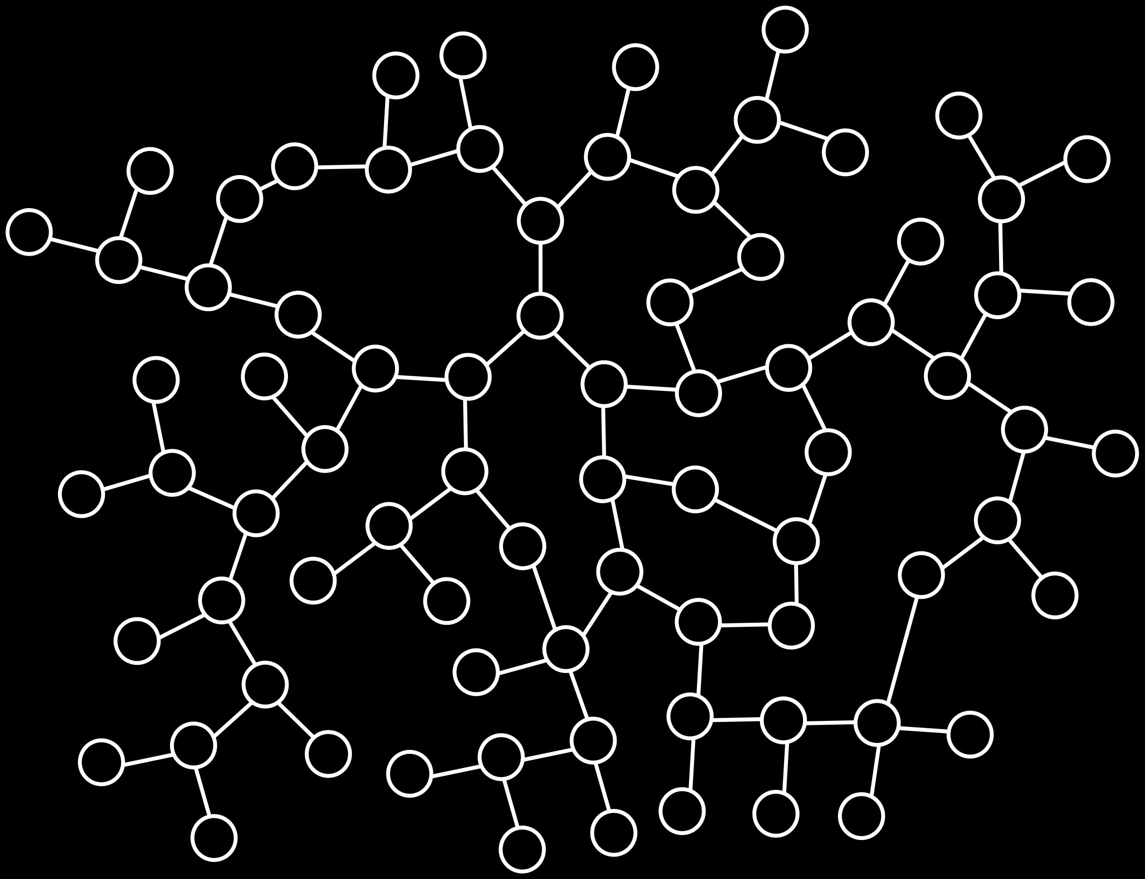
Main Theorem

1. *We spread fast:* $N_T \approx (d-1)^{\frac{T}{2}}$
2. *All nodes except for the final virtual source are equally likely to be the source, hence*

$$P(\hat{v}_{ML} = v^*) \; = \frac{1}{N_T - 1}$$

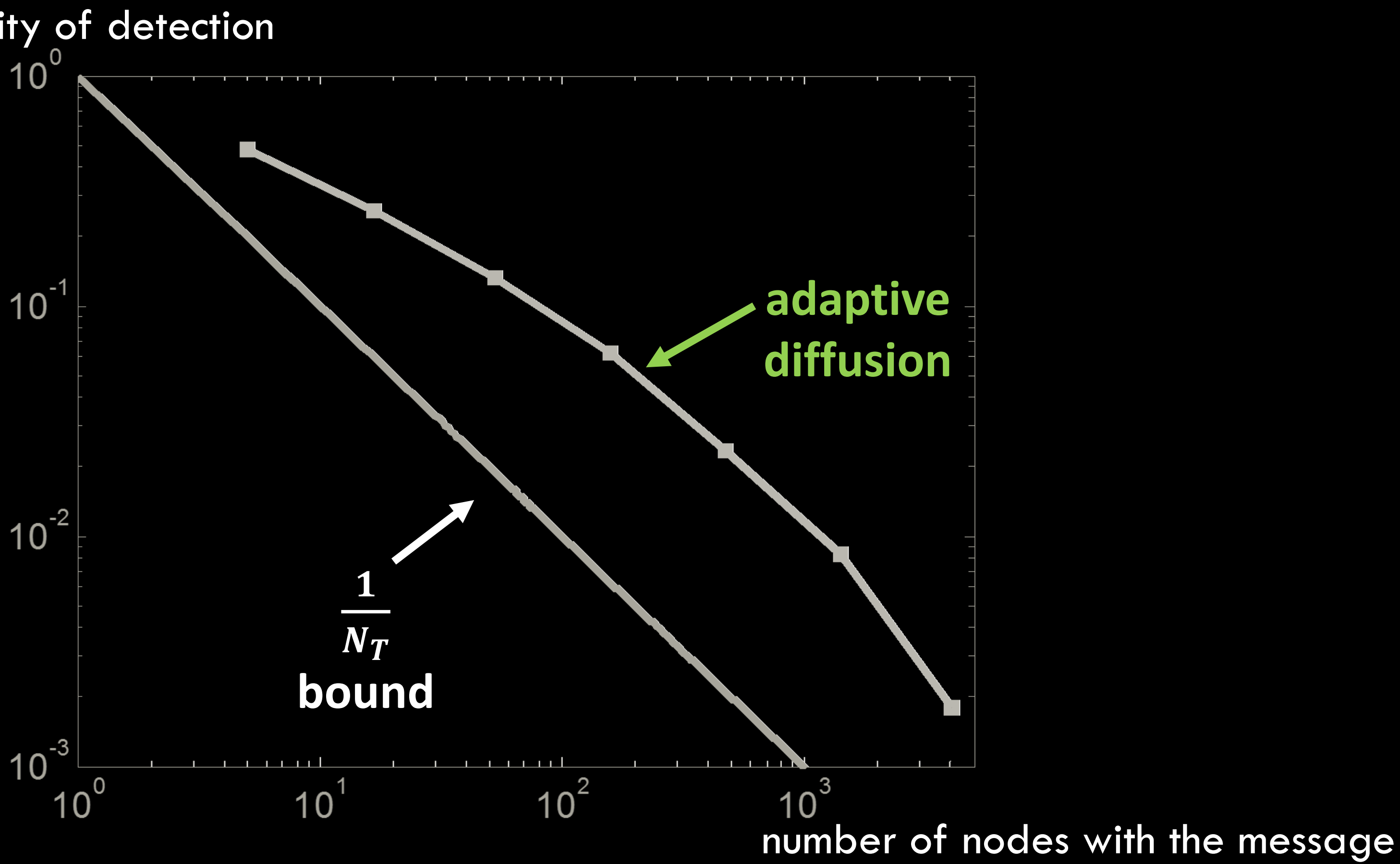
3. *The expected distance between the estimated and true source is at least* $\frac{T}{2}$.

General graphs



adaptive diffusion for general graphs?

Simulation results: Facebook graph



- likelihoods can be **approximated** numerically

Extensions and related work

Theoretical

- Adversaries with timing information
- Peer-to-peer dynamic networks
- Hiding relays
- Multiple message sources

Systems

- Cyber-bullying detection
- Anonymous video sharing
- Message caching
- Bootstrapping contacts