# Large Scale Distributed Systems

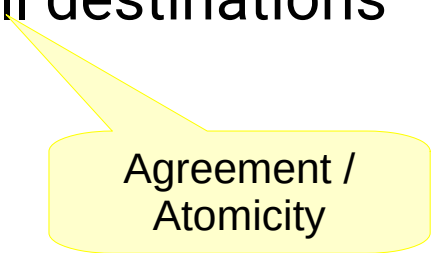
José Orlando Pereira

Departamento de Informática
Universidade do Minho

# Reliable event dissemination

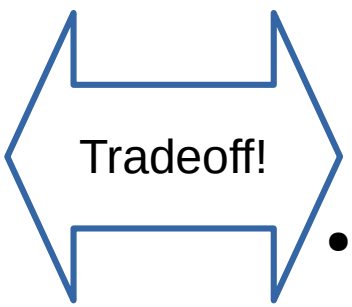
- Reliably send to multiple destinations (group)

- Informally: all destinations deliver all messages
- Senders and receivers fail: all correct destinations deliver the same messages
  - % of destinations
  - % of messages to all destinations



# Performance metrics

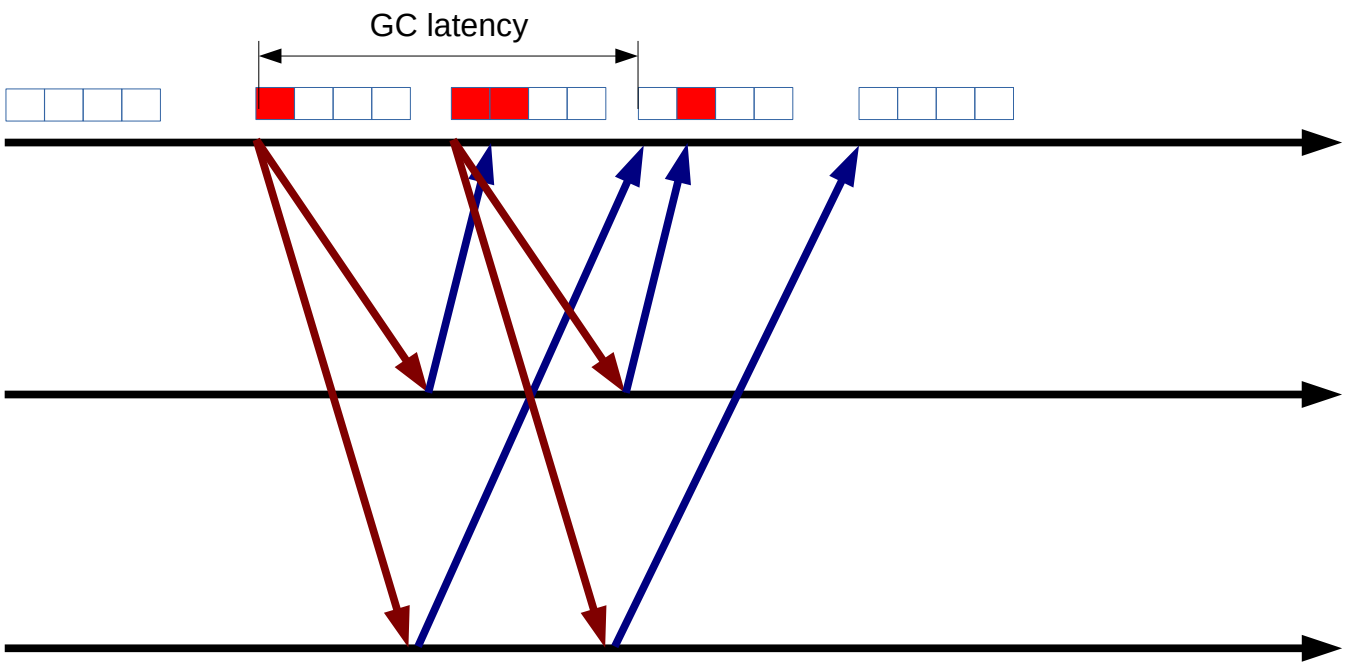
- Bandwidth
  - Payload and control messages
  - System total and maximum in one node/link




- Latency
  - Time and network hops to delivery
  - Average vs last delivery

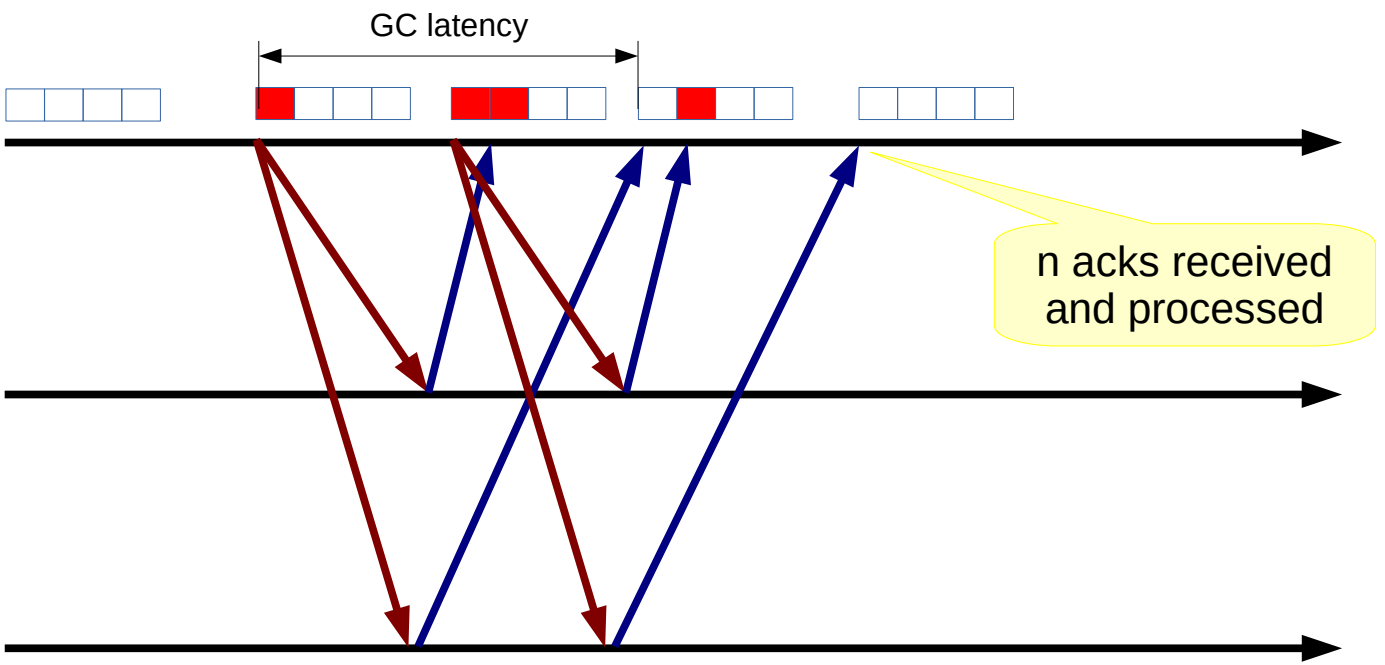
# General approach

- Buffer and retransmit until acknowledged



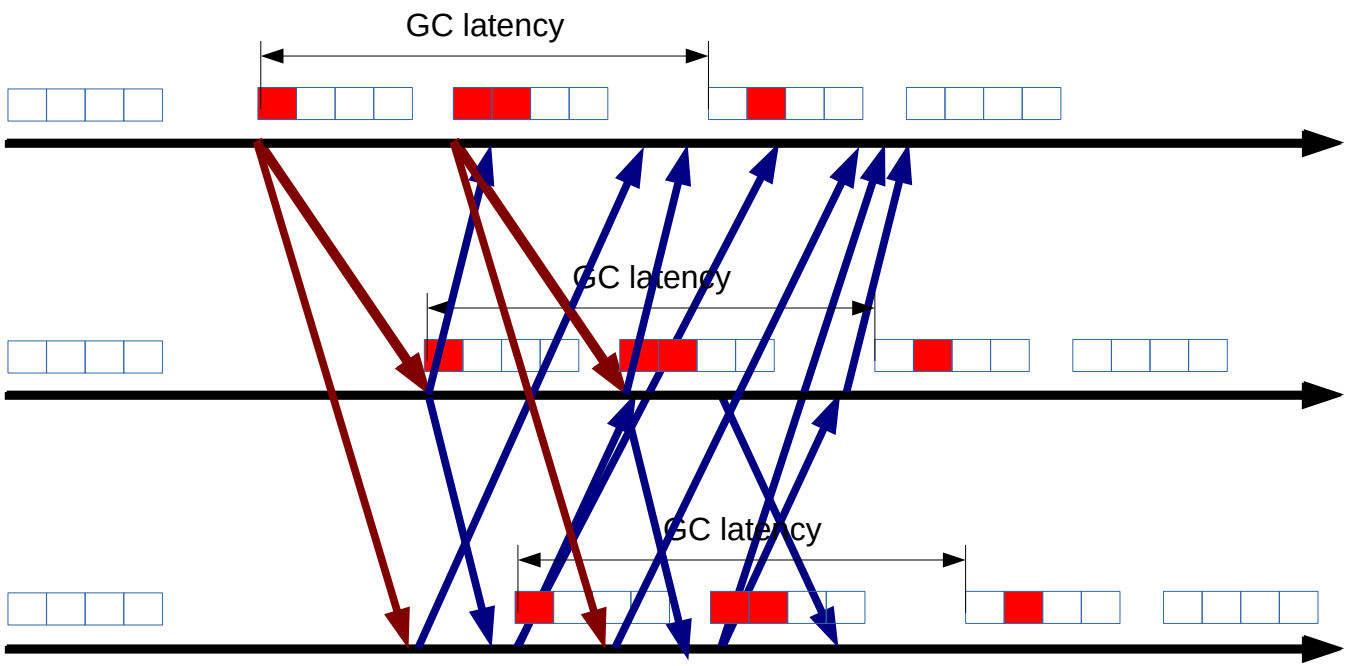
# Acknowledgments

- Not scalable to large number of destinations due to “ack implosion”:



# Agreement

- Acks need to be sent to all destinations resulting in “O(n²) ack implosion”:

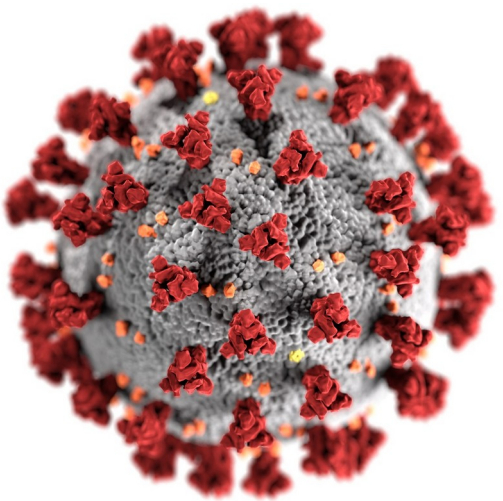


# Trees

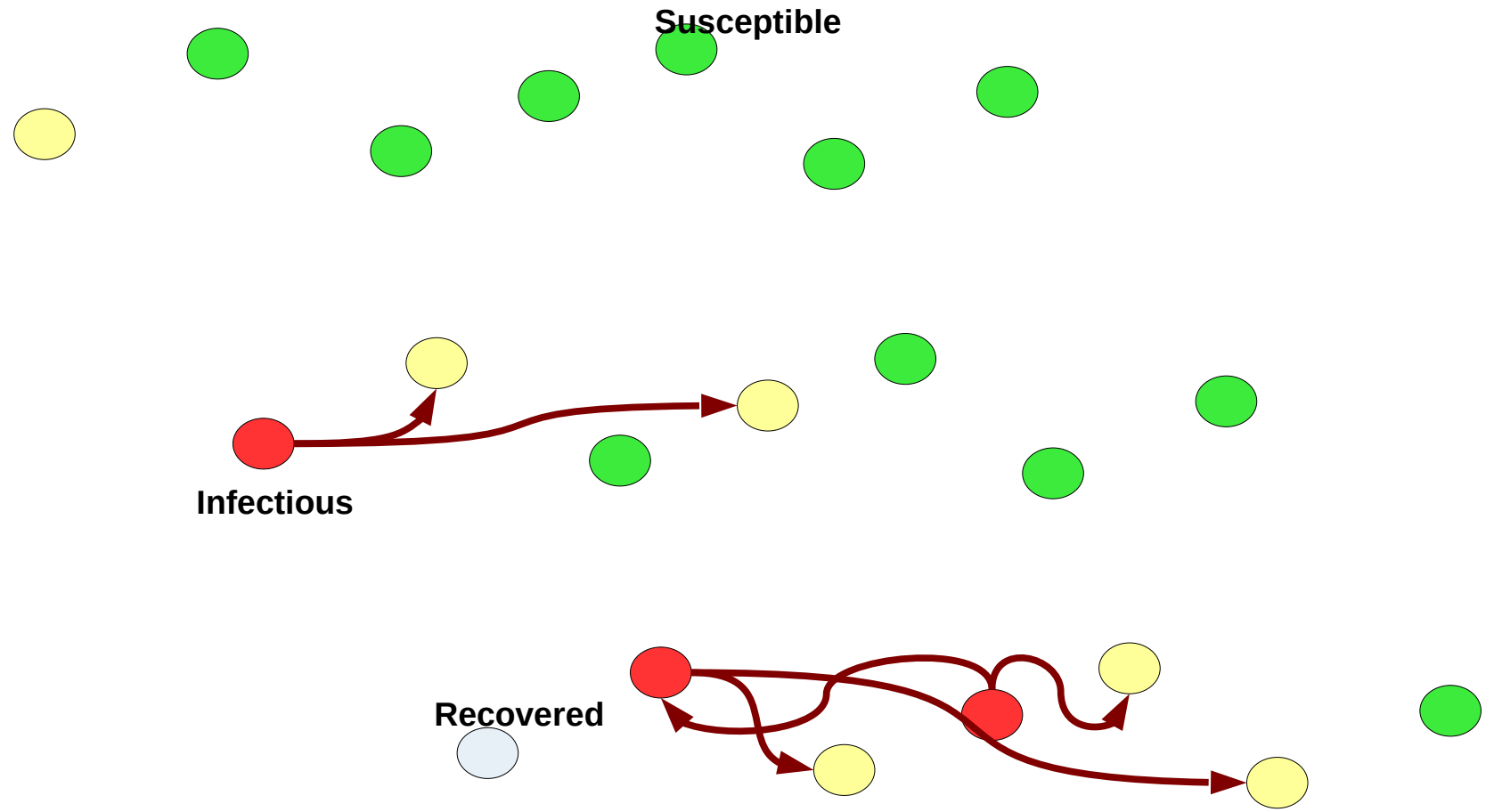
- Build a spanning tree over the destination nodes
  - Maintenance overhead
  - Brittle, as a single node/link failure leads to partitioning
- Use the tree for payload and feedback (acknowledgments)
  - Global vs local acknowledgment
- Trades latency for bandwidth!
  - More latency means more memory committed in buffers

# Epidemics

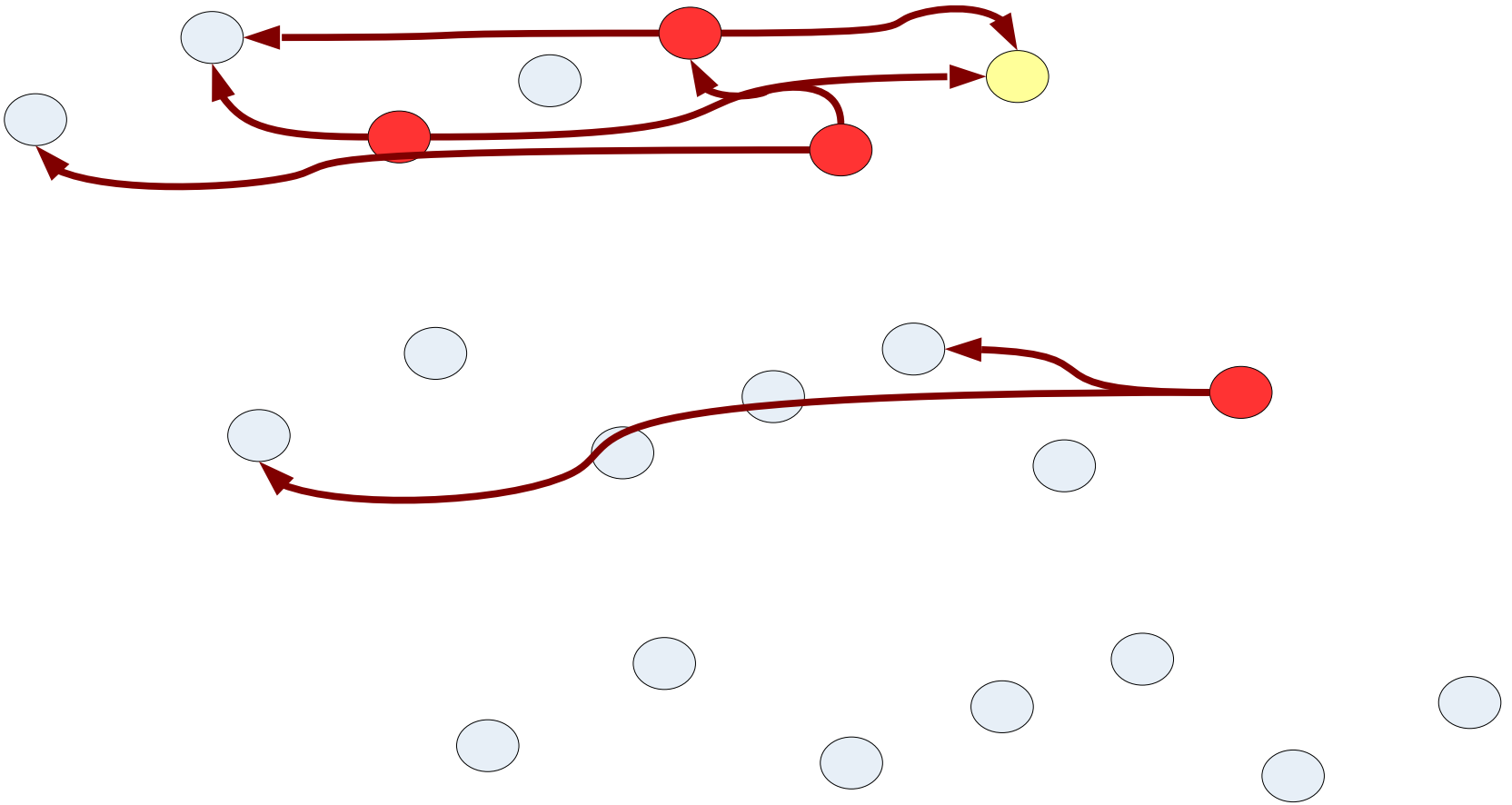
- Epidemic spread in a population
- SIR model:
  - Susceptible
  - Infectious
  - Recovered



# Epidemics



# Epidemics



# Analysis

- Probability of atomic infection p:
  - f = log(n)+c
  - p = exp(-exp(-c))
- Duration of epidemic when infecting the entire population order of log(n)

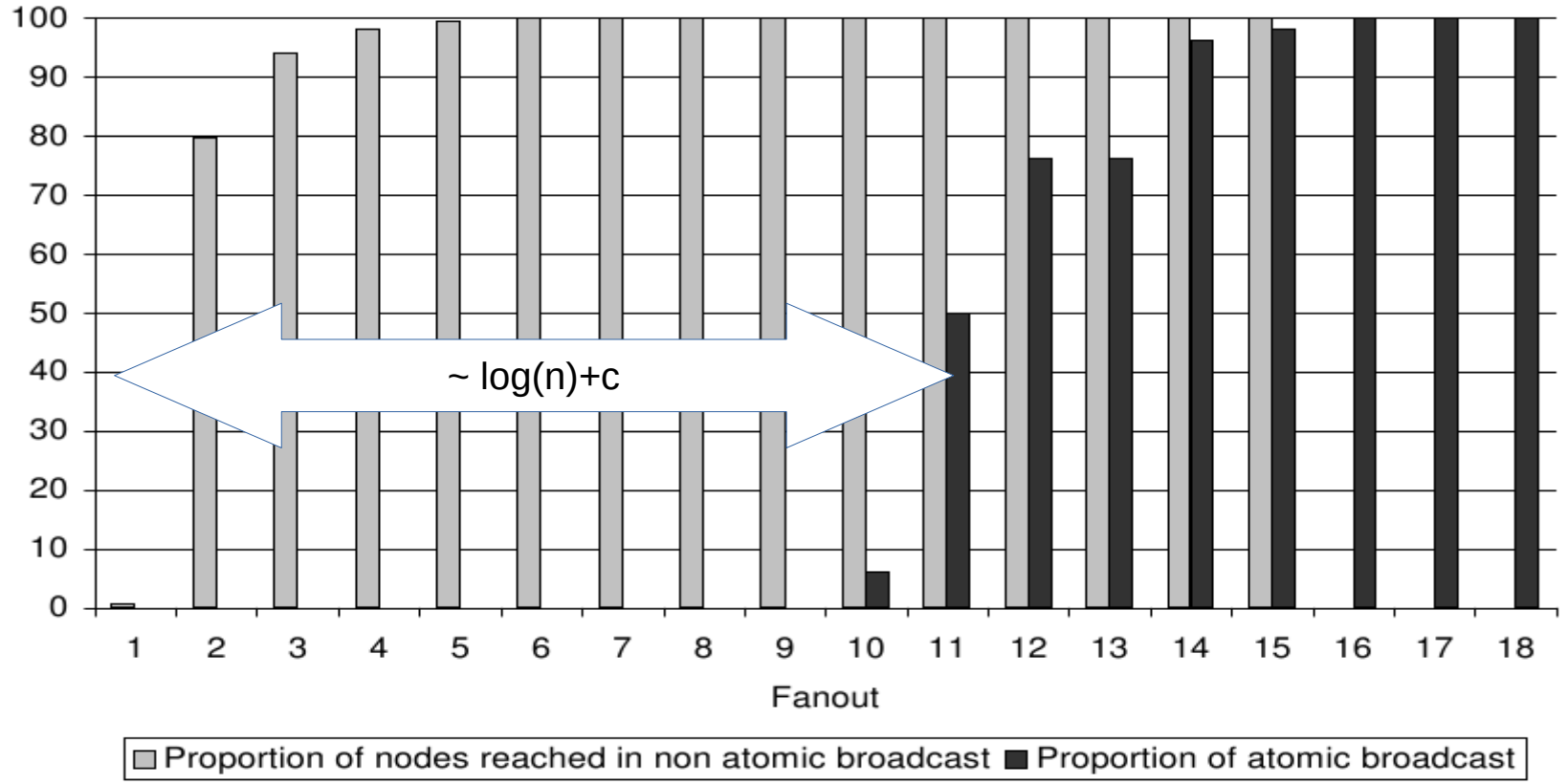
# Roadmap

- Epidemic event dissemination
- Efficiency
- Distributed aggregation

# Epidemics and Information dissemination

- Similarity with epidemics:
  - Sender = contagious = spreads rumor
  - Receiver = infected = knows rumor
  - Ignores duplicated = recovered = old news...
- Interesting parameters:
  - $n$ – size of the population
  - $f$ – number of targets

# Fanout vs Reliability




(50000 destinations)