

Quasar: A Probabilistic Publish-Subscribe System for Social Networks

Bernard Wong and Saikat Guha

Cornell University

IPTPS 2008

Contribution

Quasar

A simple efficient topic-based publish-subscribe system for unreliable data

Problem Statement

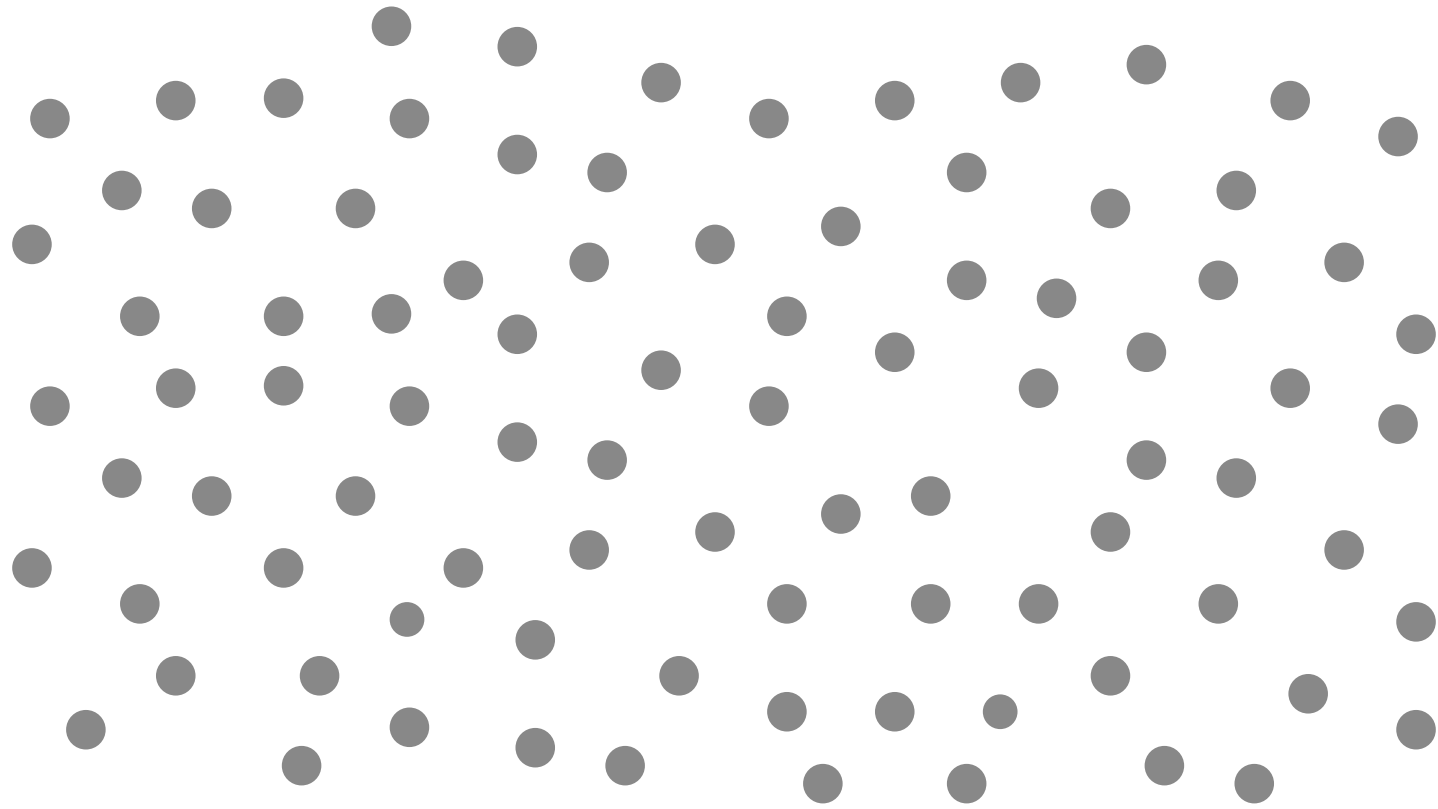
Topic-based pub-sub routing that . . .

- ▶ more efficient than flooding, simpler than DHT
 - ▶ trades off reliability
- ▶ works with many overlays
 - ▶ Latency-aware, unstructured, social-networks, . . .
 - ▶ Good for small-world network
- ▶ does not use rendezvous nodes
 - ▶ Avoid hot spots, single points of failure
- ▶ supports a large number of topics
 - ▶ Cheap subscribe and unsubscribe process

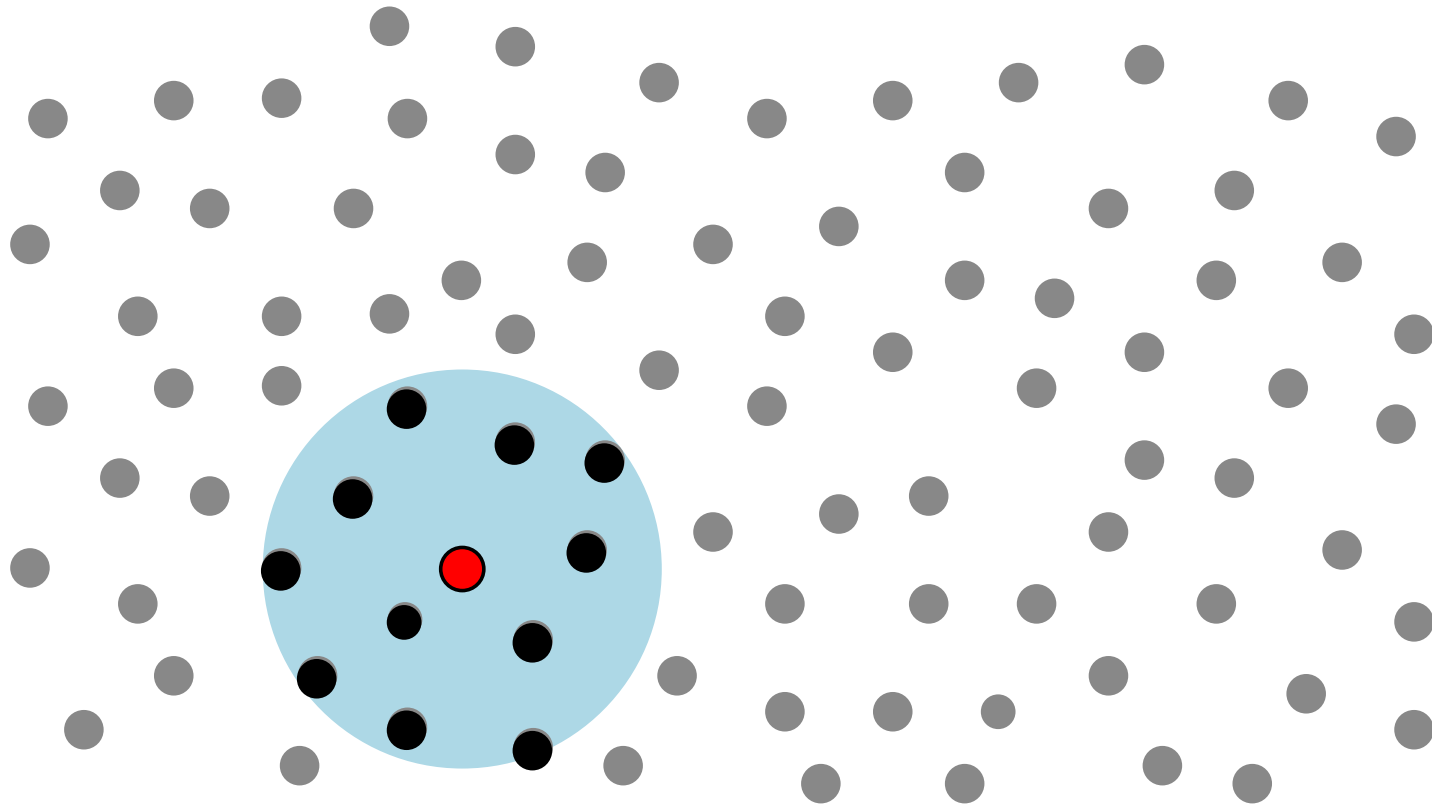
Quasar: Summary

- ▶ Limited proactive dissemination of positive information
- ▶ Per-message negative information
- ▶ Combination of parallel random walks and directed routing

Quasar: Big Picture

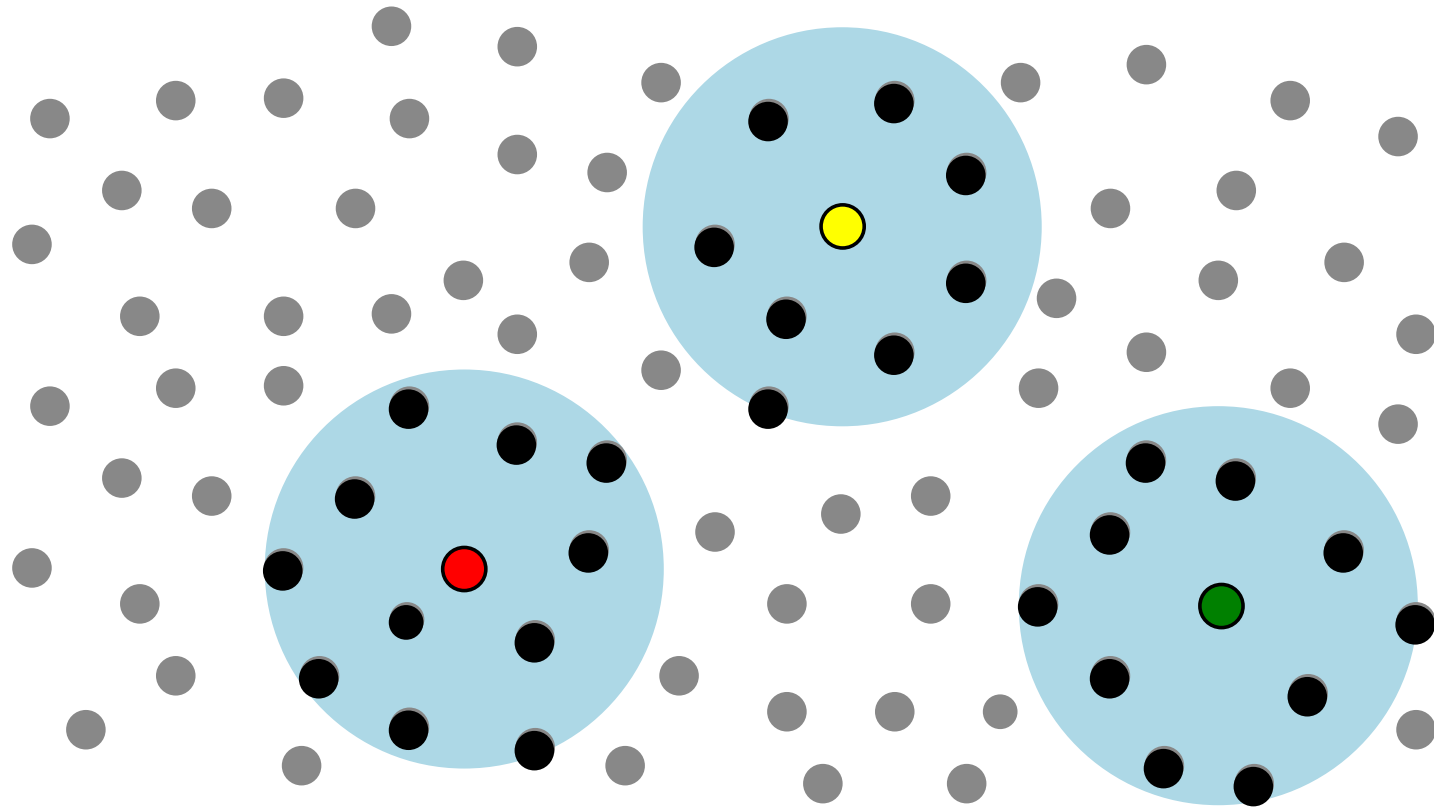


Quasar: Big Picture



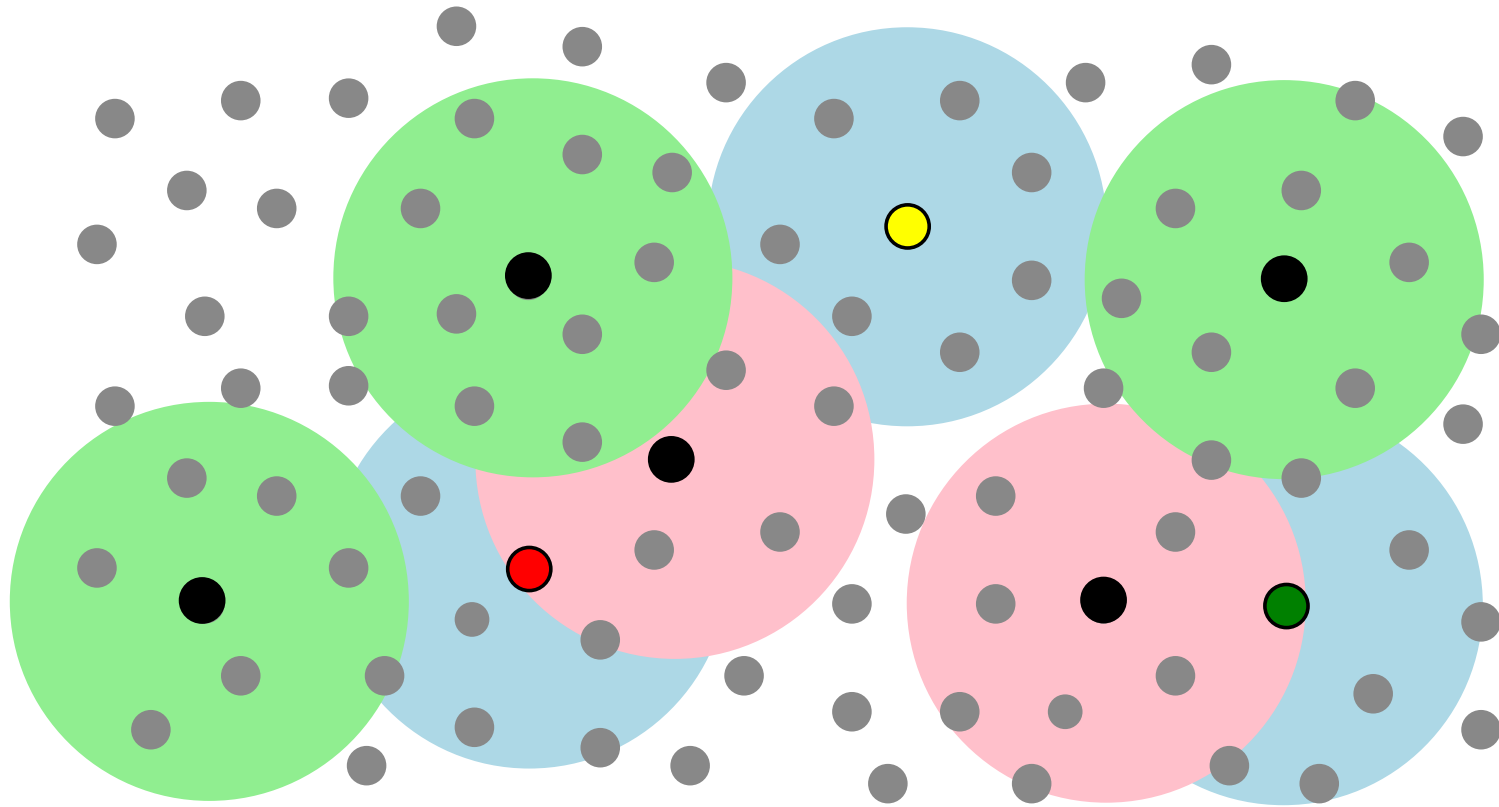
Subscriptions disseminated proactively to a fixed depth:
gravity well

Quasar: Big Picture



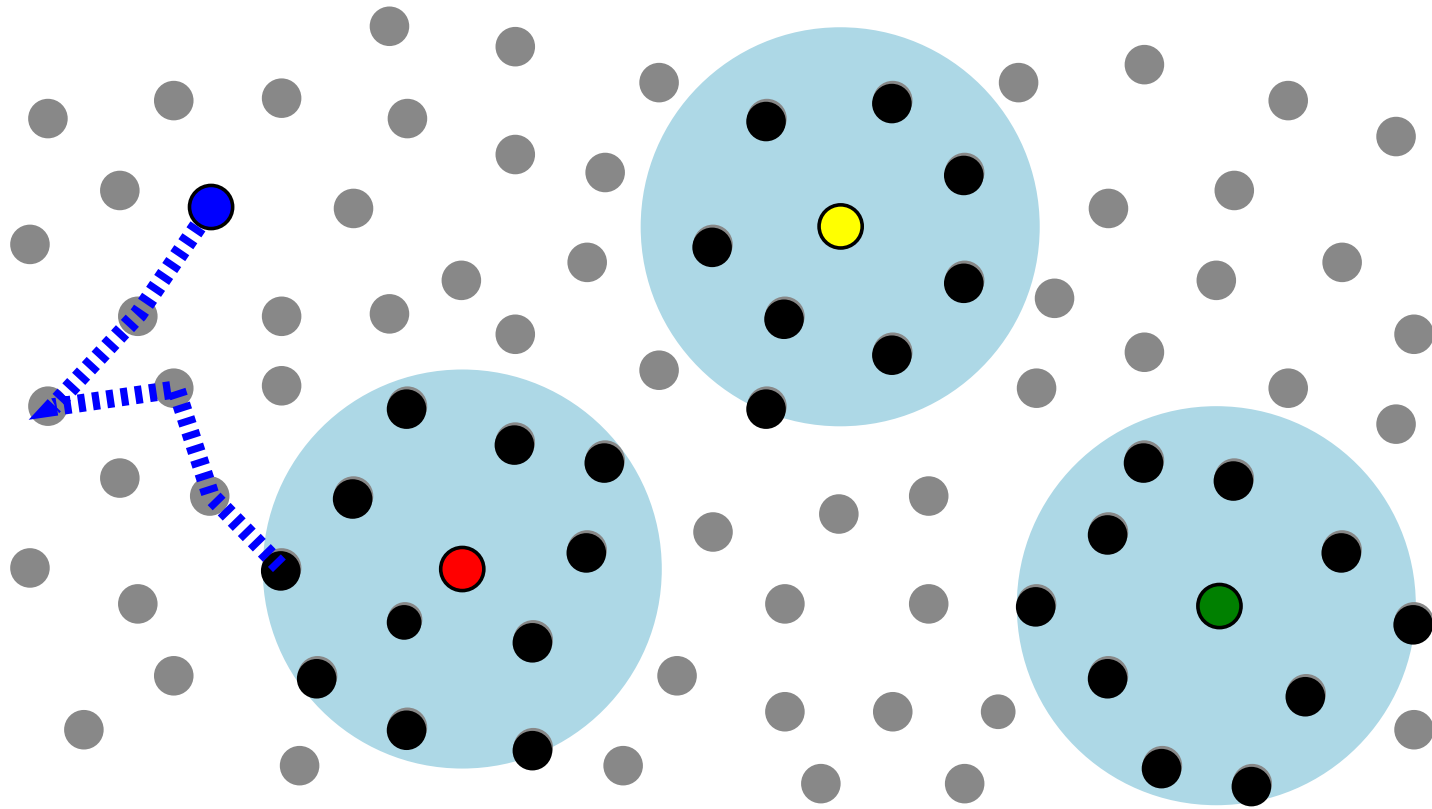
Subscriptions disseminated proactively to a fixed depth:
gravity well

Quasar: Big Picture



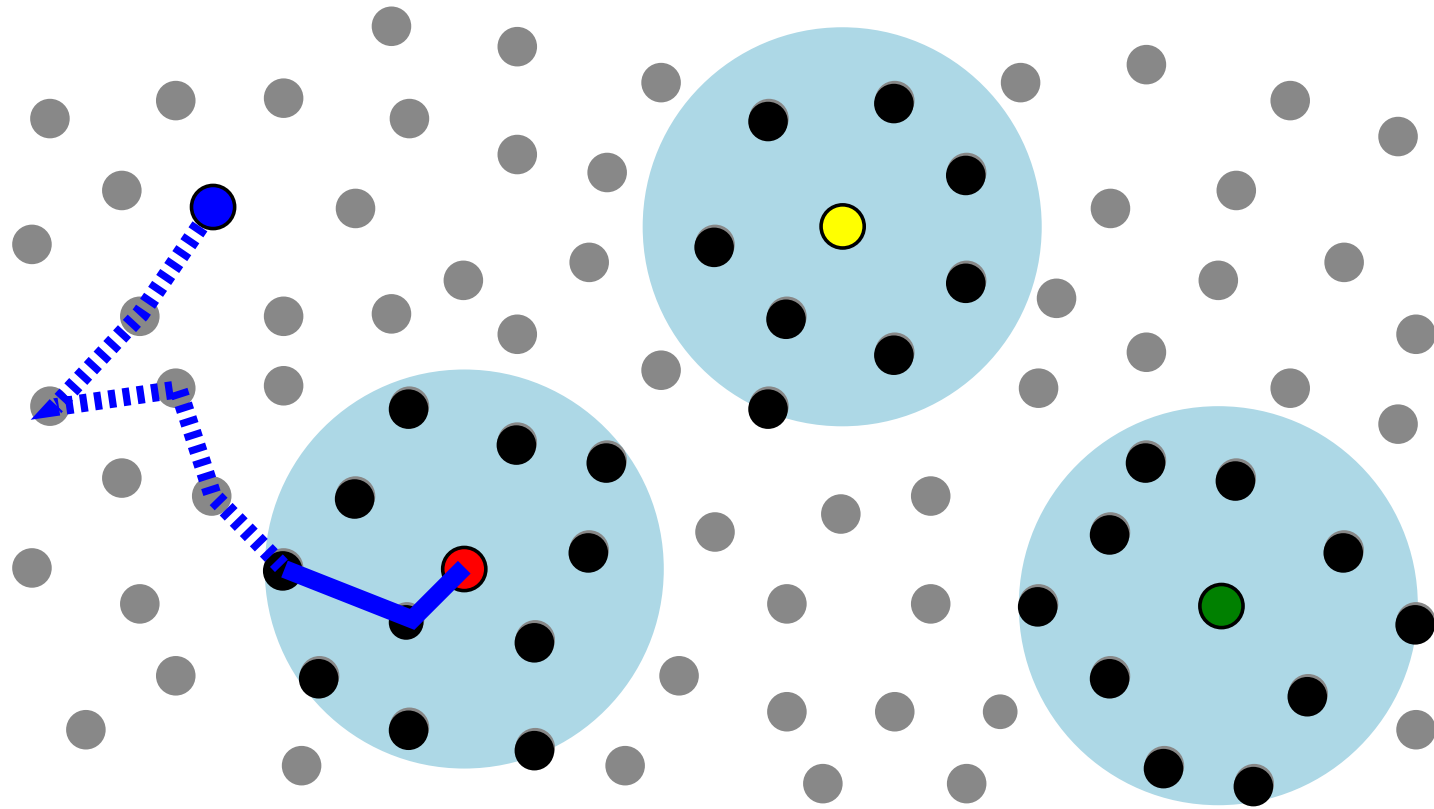
Different topics create own gravity wells

Quasar: Big Picture



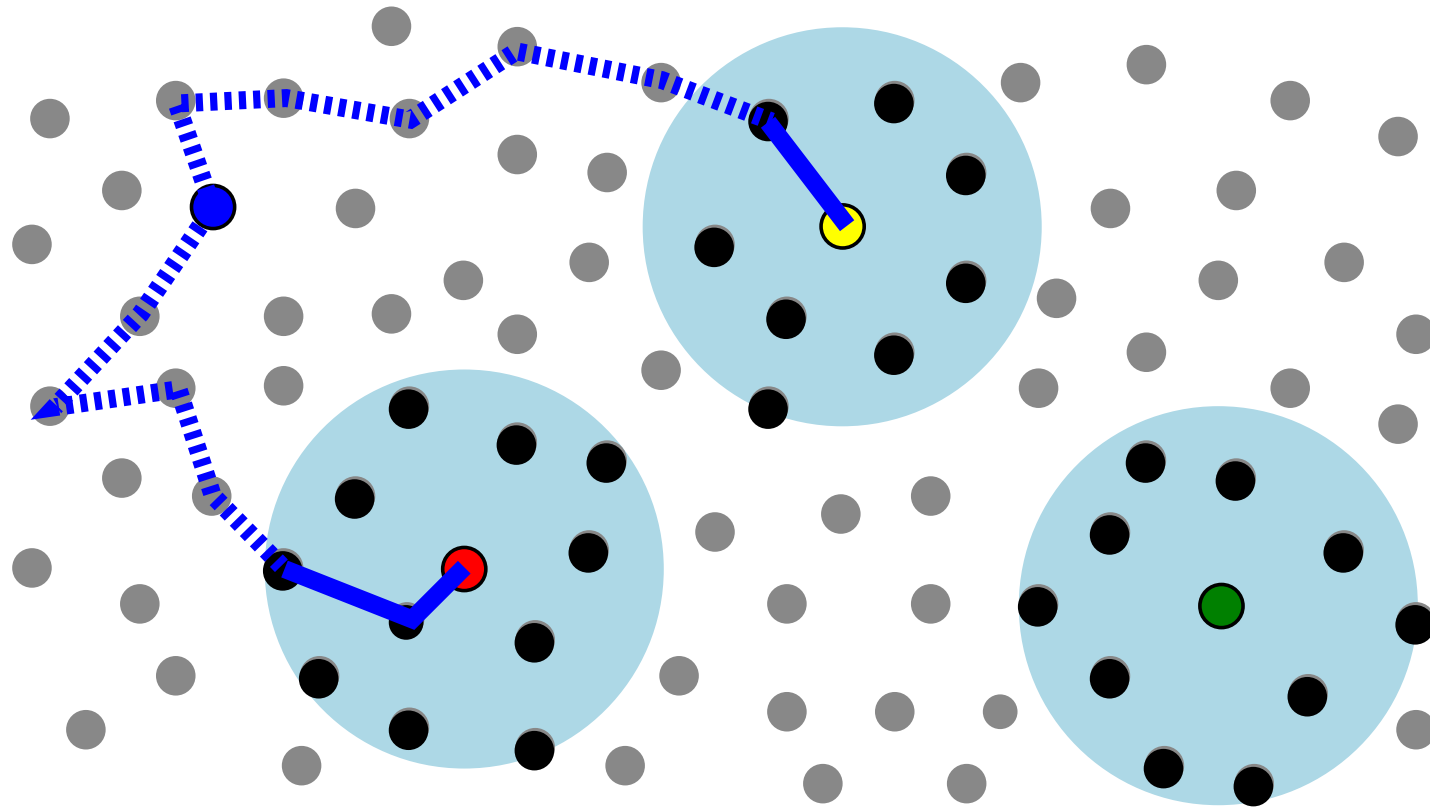
Message routed by random walk until reaching a gravity well

Quasar: Big Picture



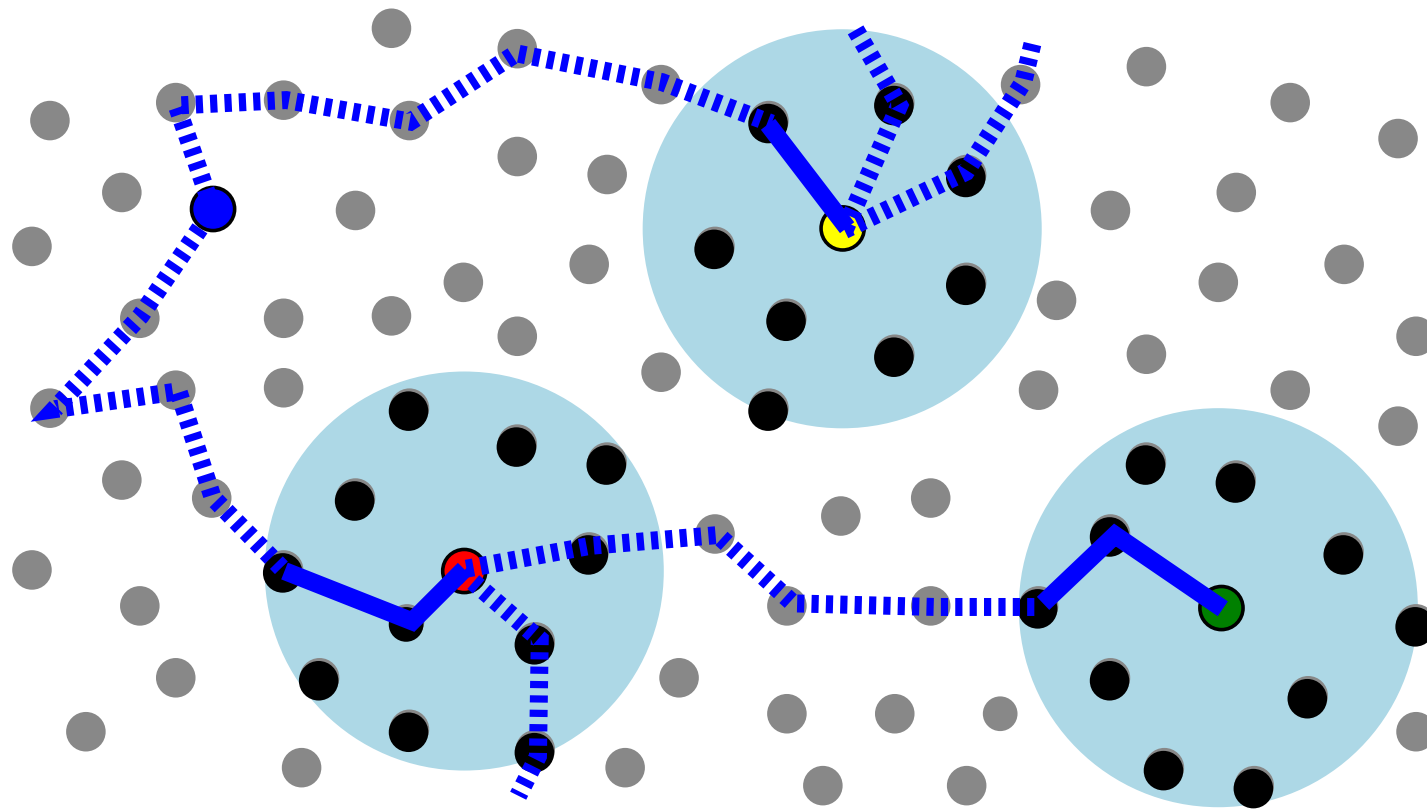
Message routed progressively closer to node in gravity well

Quasar: Big Picture



Multiple walks issued simultaneously

Quasar: Big Picture



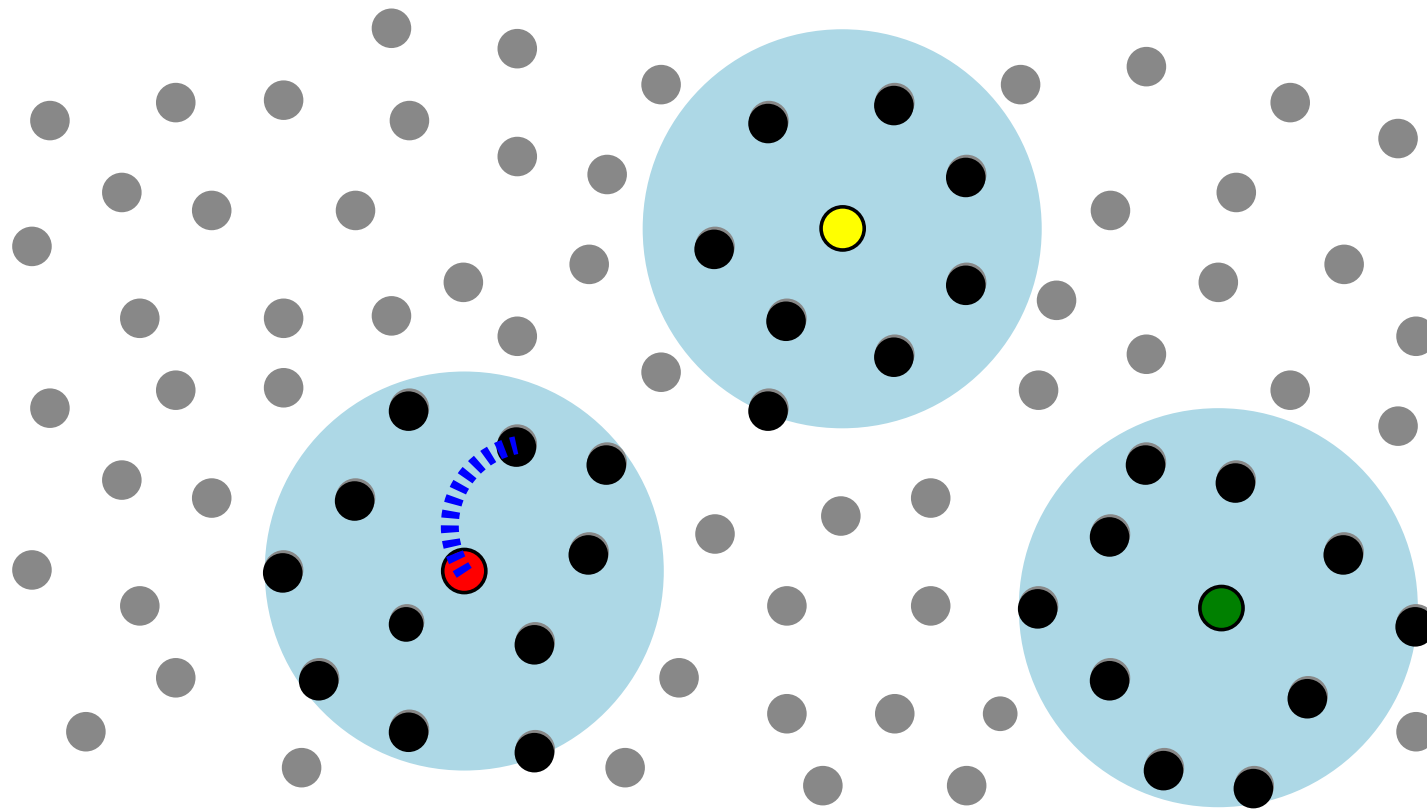
Messages republished for greater coverage

Quasar: Messages

- ▶ Duplicates suppressed (Unique message ID)
- ▶ Random-walk length limited (TTL)
- ▶ Republishing limited (Generation count)

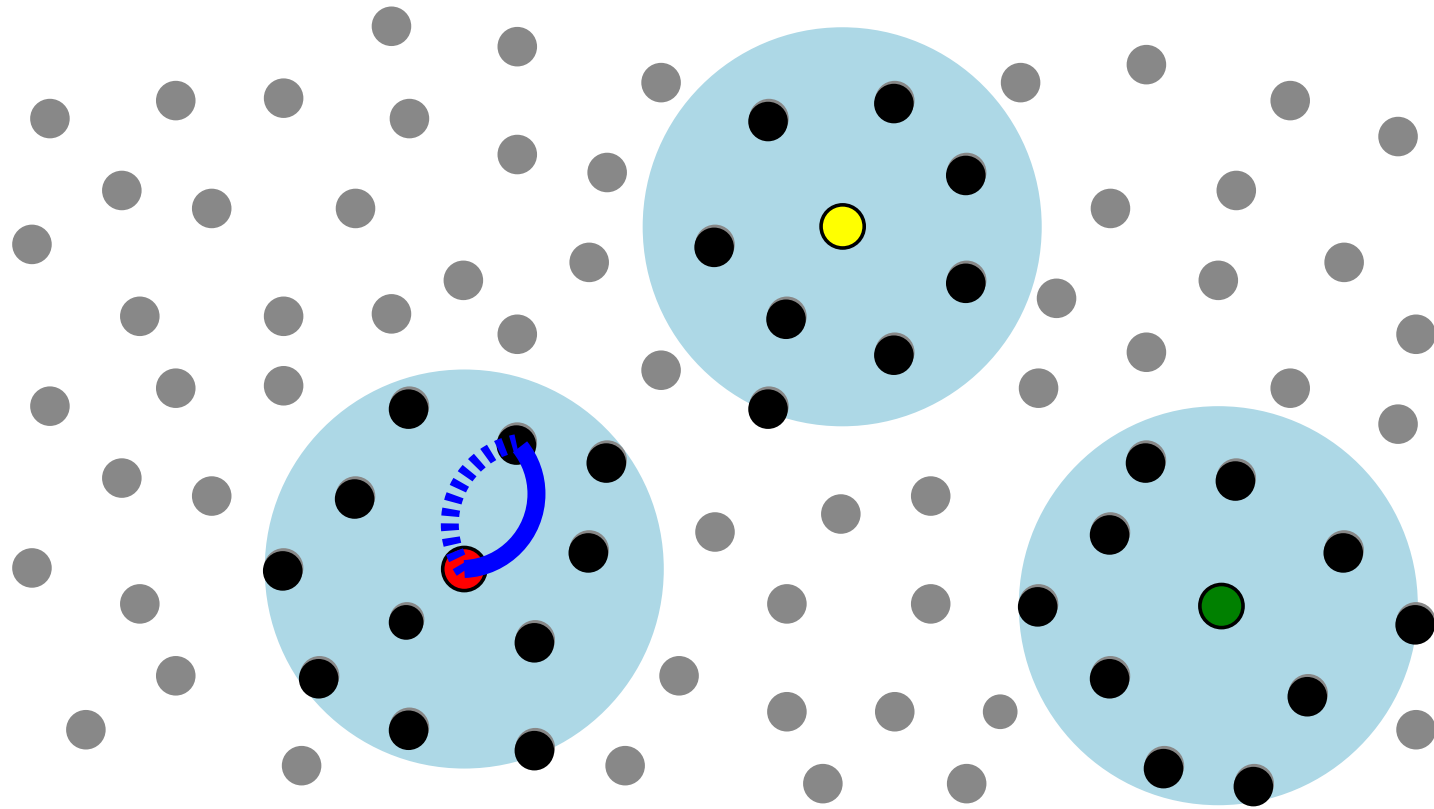
... and, of course, topic ID and negative information

Quasar: Big Picture



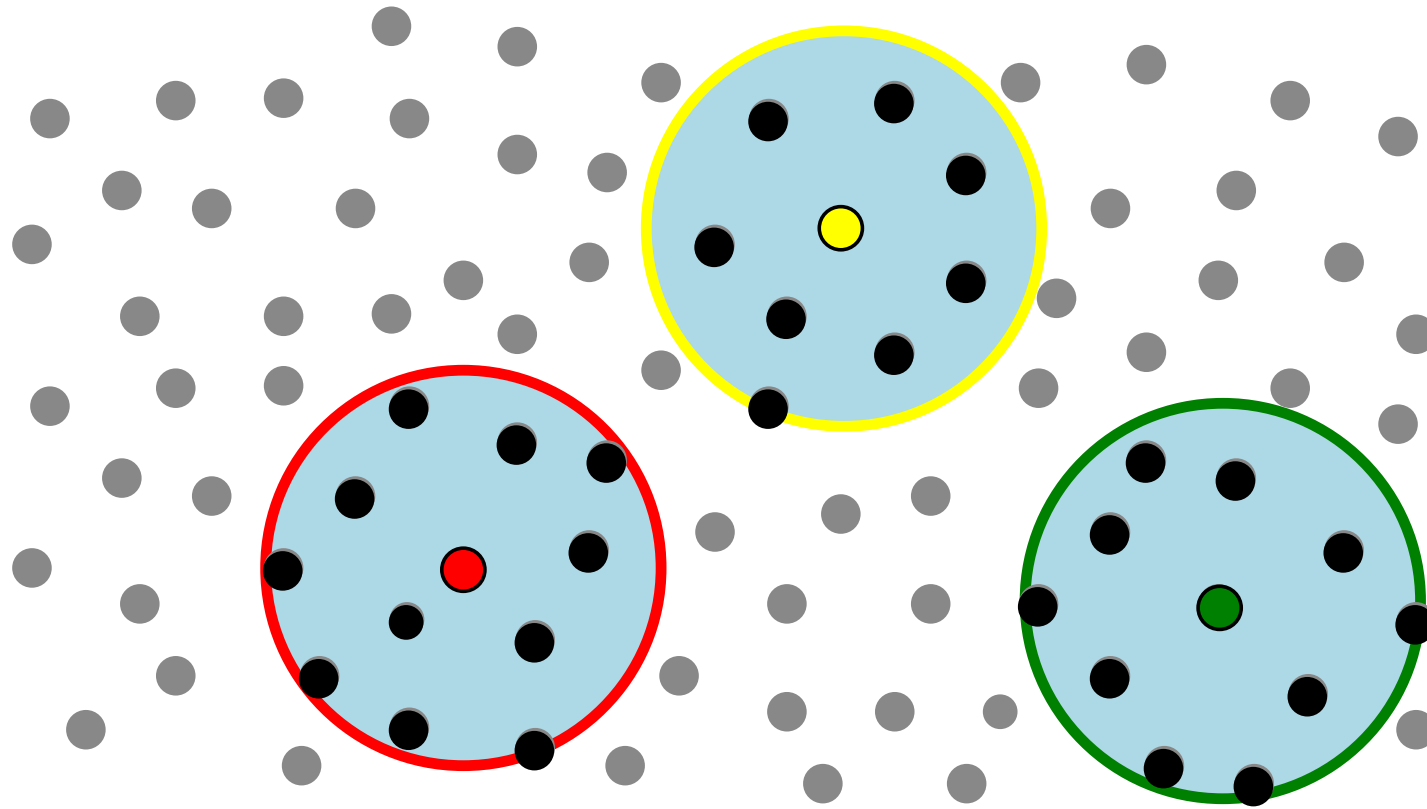
Problem: Self-loops

Quasar: Big Picture



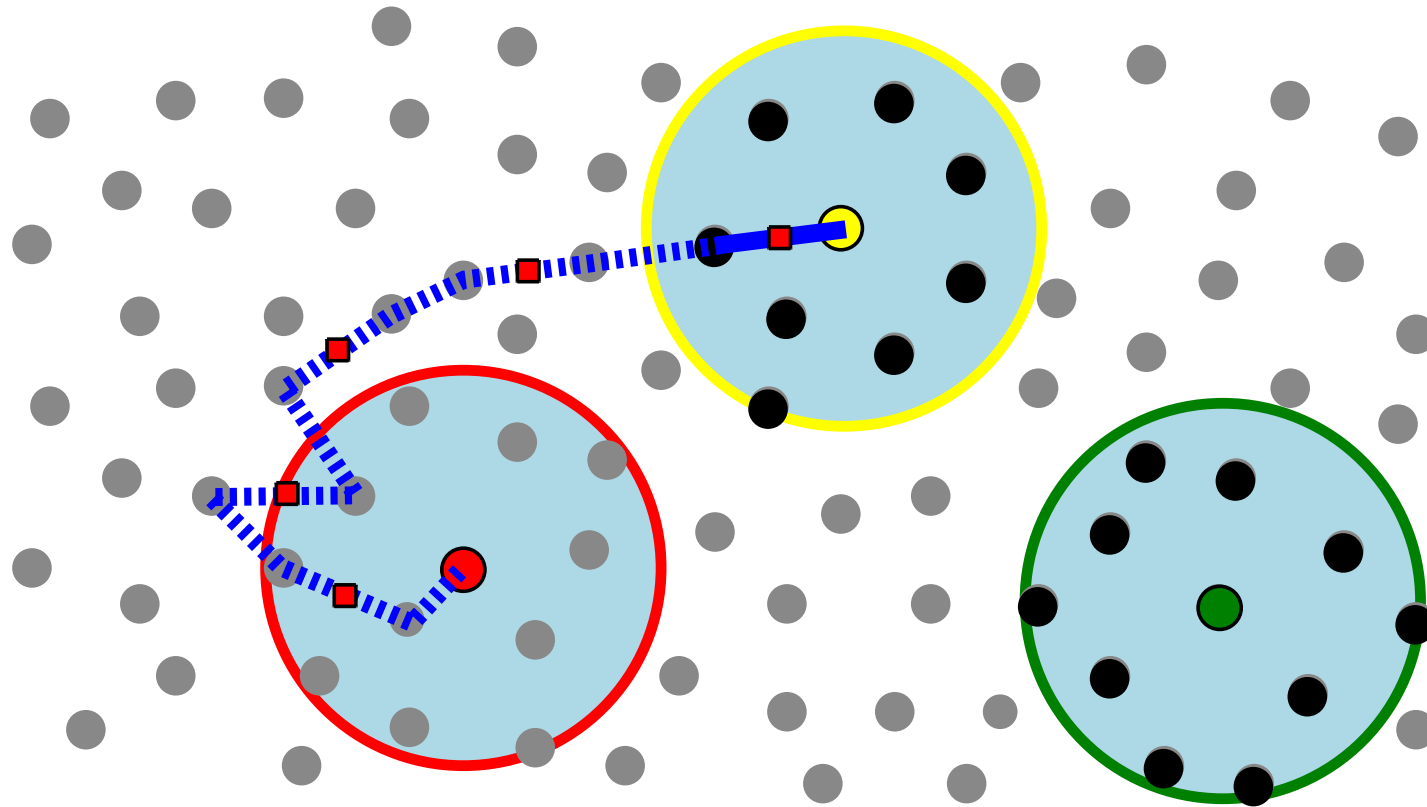
Problem: Self-loops

Quasar: Big Picture



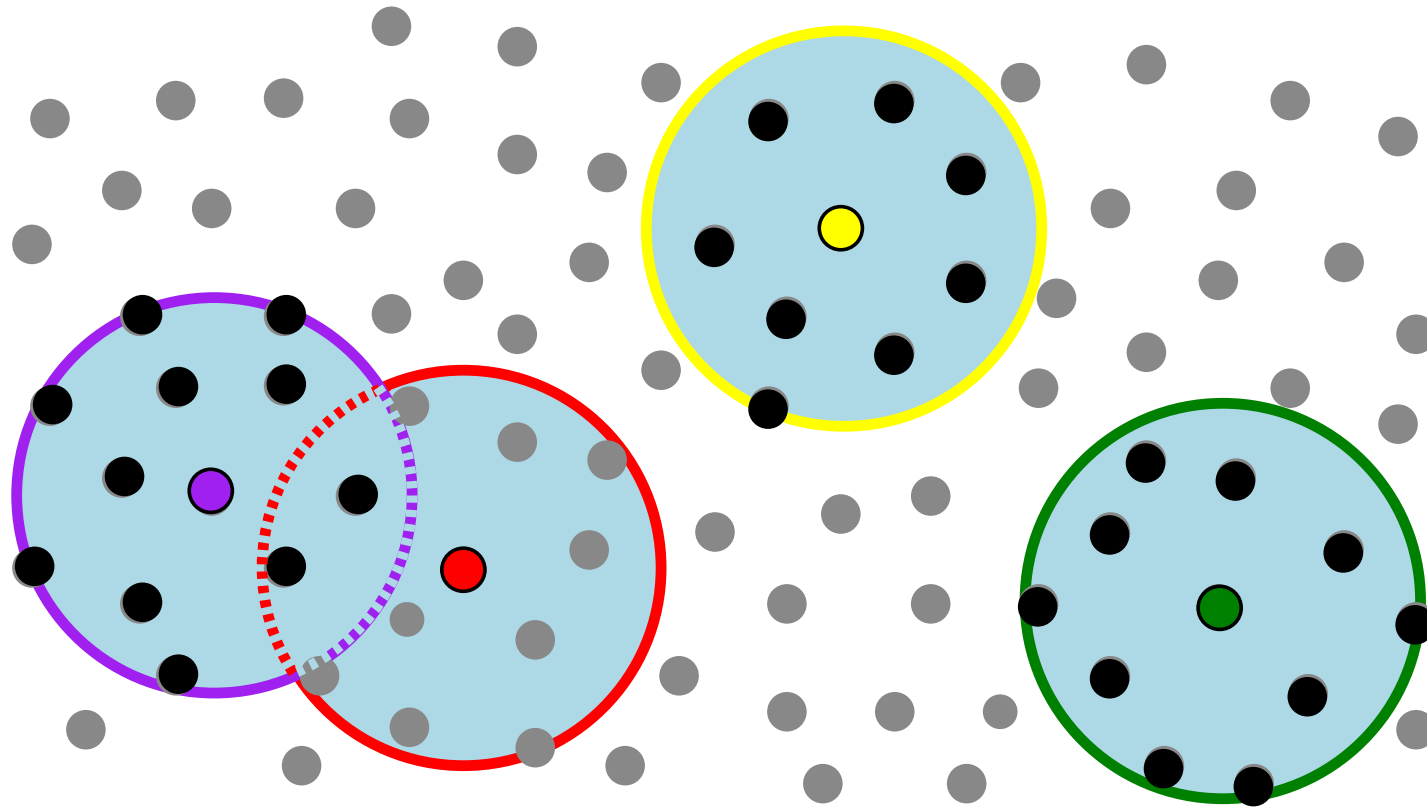
Gravity wells indexed by topic ID as well as node ID

Quasar: Big Picture



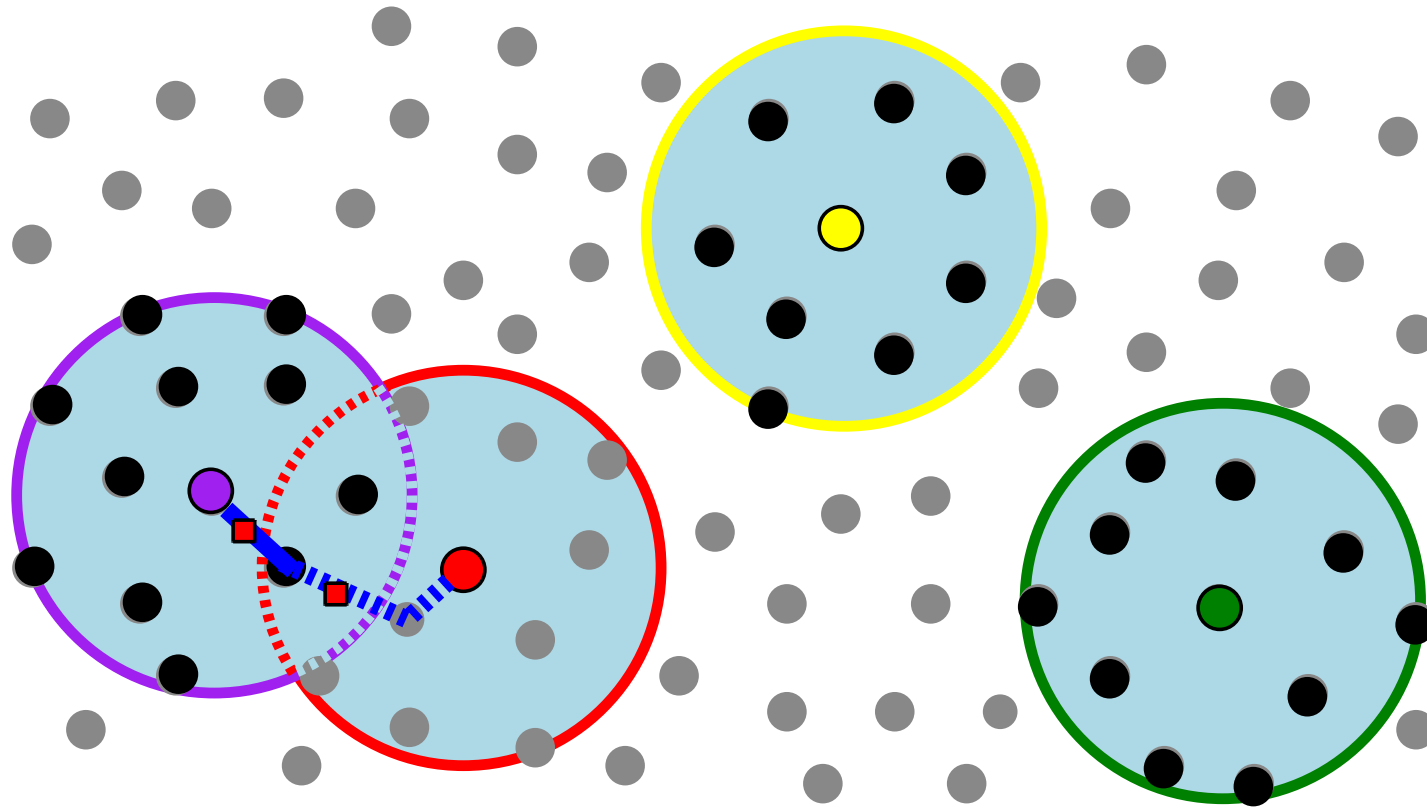
Effects of gravity well nullified per message

Quasar: Big Picture



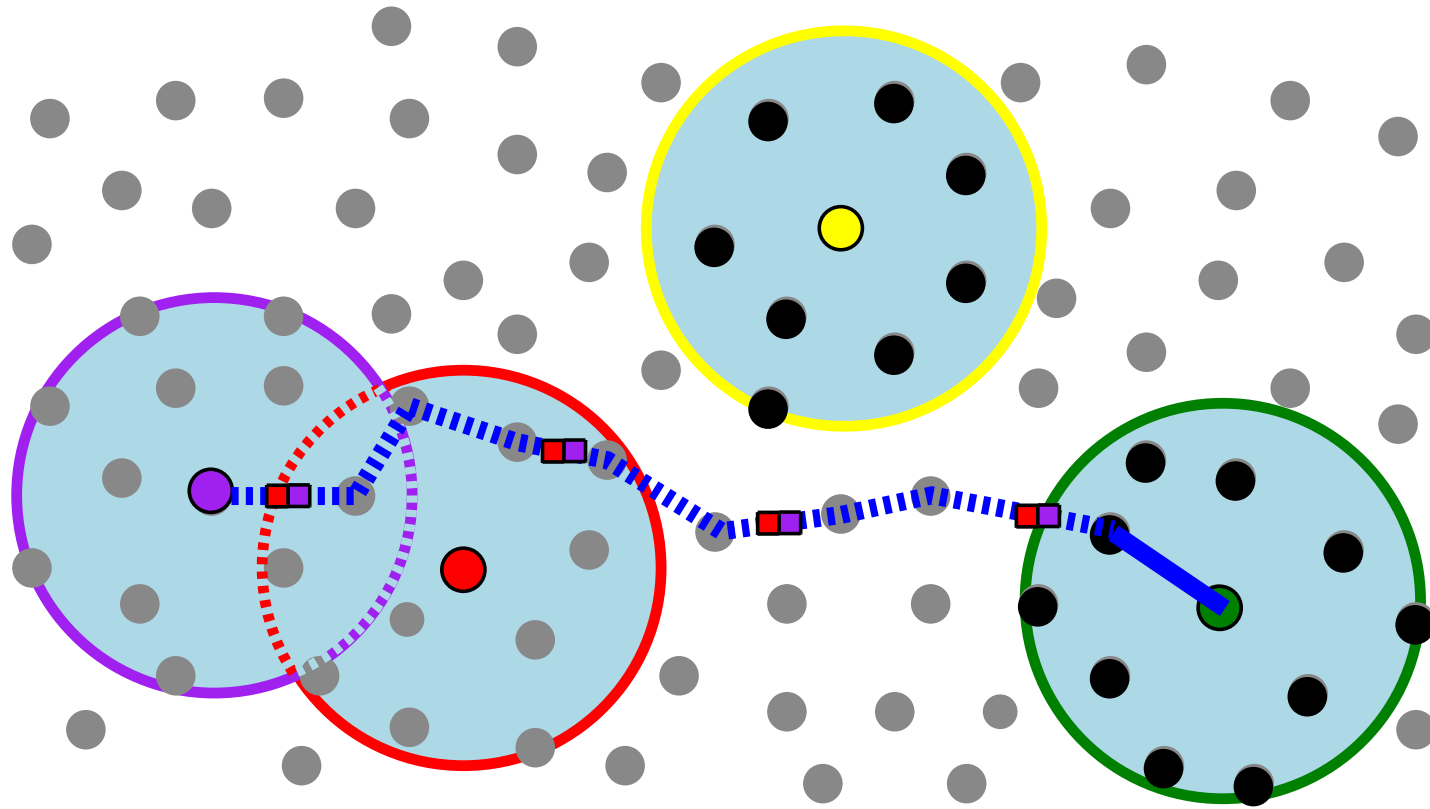
Nullifying a gravity well does not affect overlapping gravity wells

Quasar: Big Picture



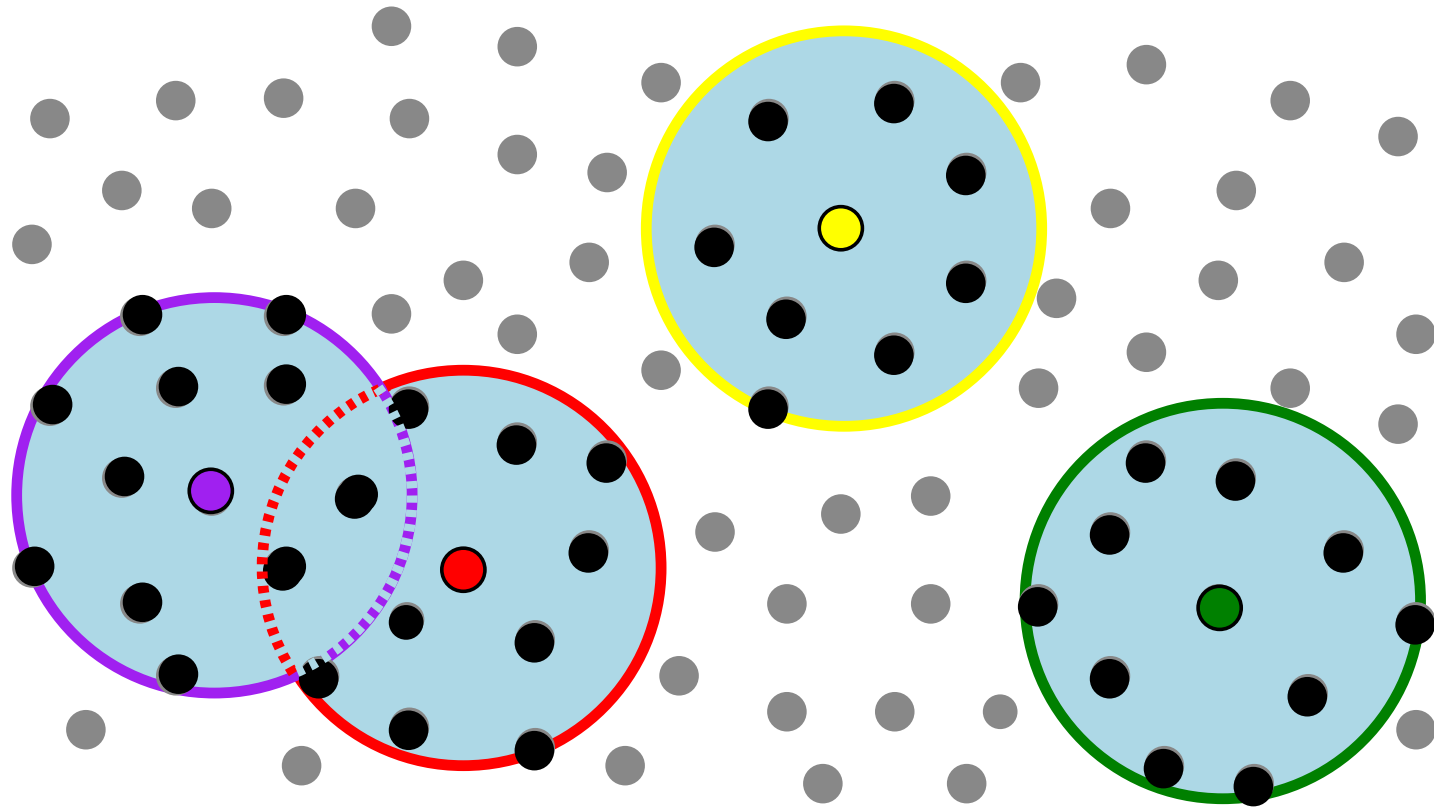
Nullifying a gravity well does not affect overlapping gravity wells

Quasar: Big Picture



Multiple gravity wells can be nullified for a message

Quasar: Big Picture



Details: soft state, attenuated bloom filters. See paper

Properties

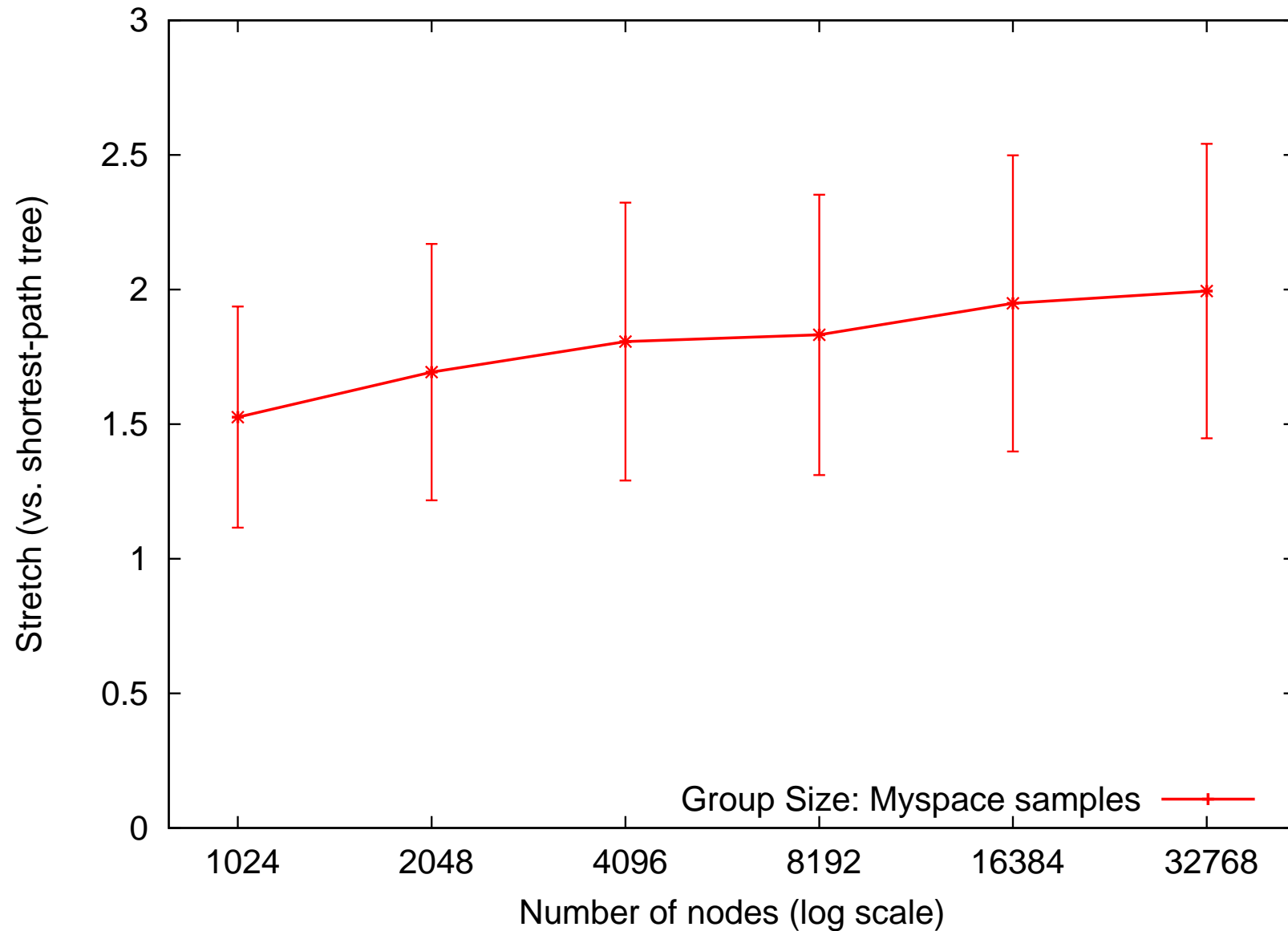
- ▶ Per-Node State
 - ▶ Depends on the density of subscriptions
 - ▶ Not on the total number of nodes or topics
 - ▶ Can be controlled by choice of overlay
- ▶ Stretch
 - ▶ Ratio of hops in Quasar to a dissemination tree
 - ▶ Depends on the density of topic subscribers
 - ▶ Can be controlled by higher layers
- ▶ Coverage
 - ▶ Fraction of topic subscribers receiving the message
 - ▶ Seems to be independent* of system size!
 - ▶ Perhaps dependent on the underlying overlay?

*Based on our simulations

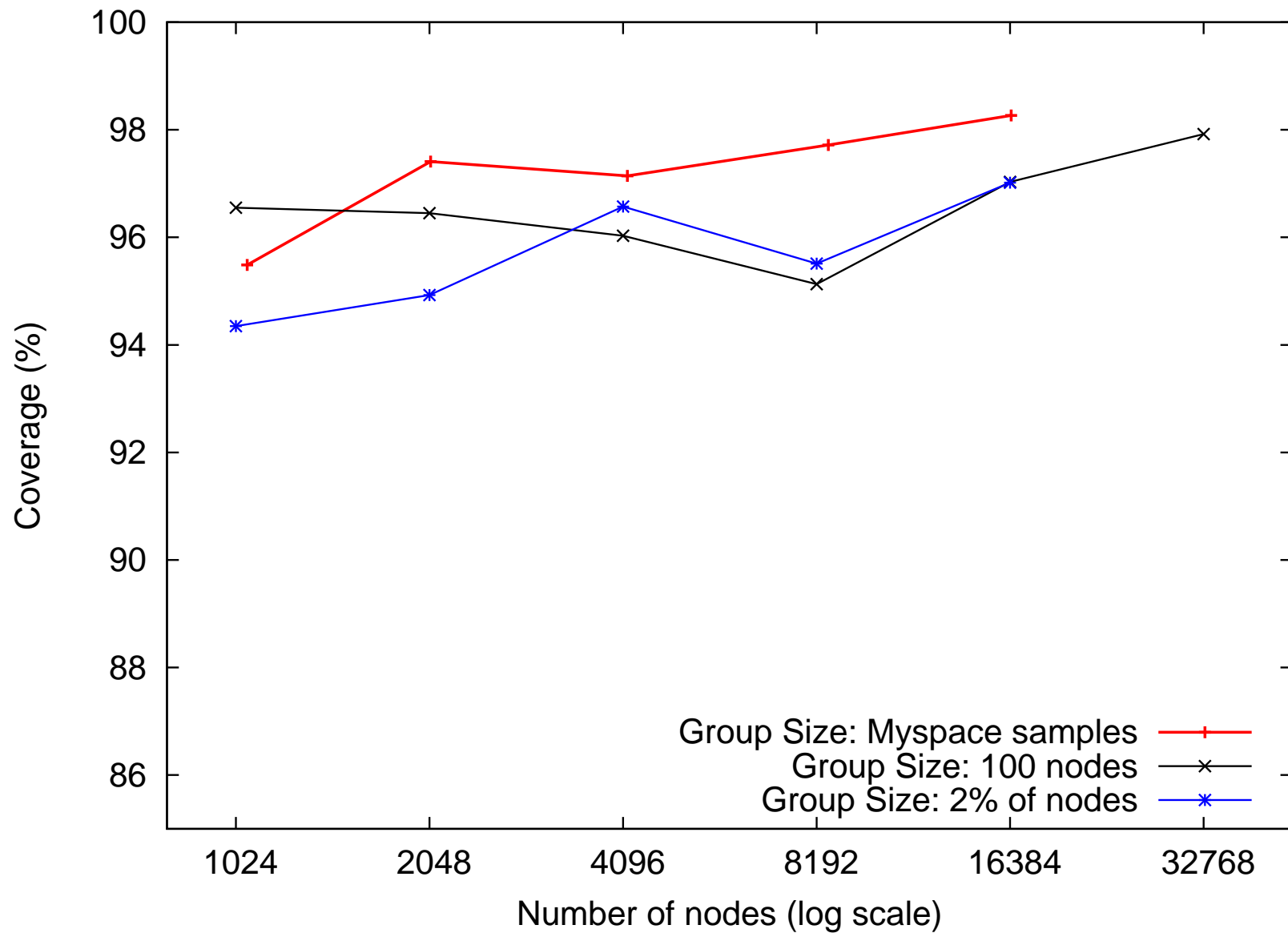
Evaluation

- ▶ Custom simulator
- ▶ Number of nodes between 1K–32K
- ▶ Topic sizes sampled from MySpace
 - ▶ By scraping public profiles
 - ▶ Mostly individuals
 - ▶ Some communities
- ▶ Approx 17.5K users
- ▶ Topic size: $\sim 1.5K$ average, ~ 37 online
 - ▶ 33% topics have < 10 online users
- ▶ Also simulated synthetic topic sizes
 - ▶ Constant size
 - ▶ Size proportional to network size

Stretch



Coverage



Road Ahead

Quasar, only a small piece of the puzzle

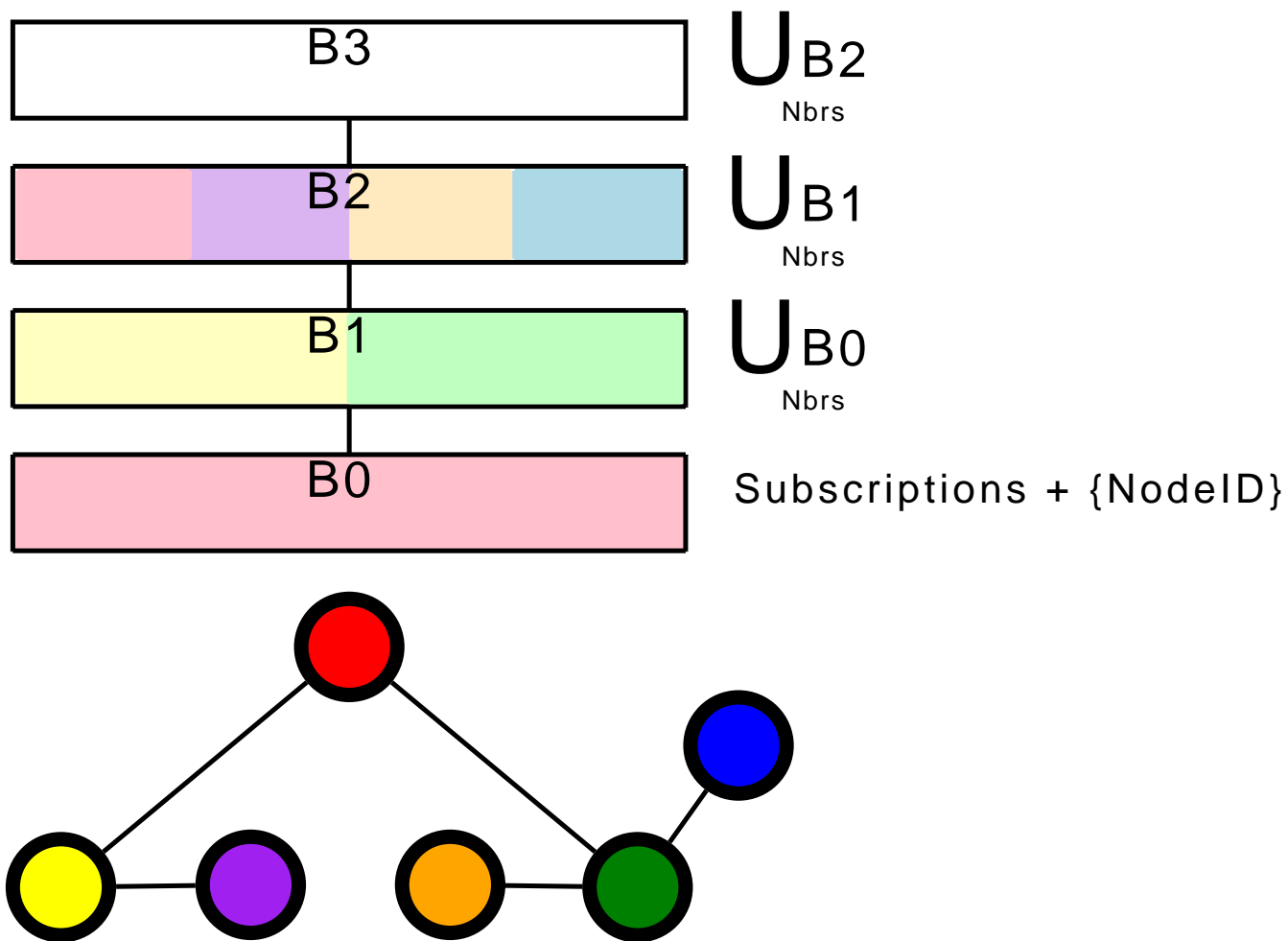
- ▶ Reliability layer
 - ▶ Reactive or periodic polling
 - ▶ Erasure coded messages
- ▶ Privacy and Security layer
 - ▶ Can we do onion routing on top of Quasar?
- ▶ Persistence layer
 - ▶ A distributed object store on top of Quasar
 - ▶ One topic per object
 - ▶ Replication and reliability under the hood
 - ▶ Peer-to-peer photo sharing (ala Flickr)

Summary

- ▶ Quasar: publish-subscribe routing primitive
 - ▶ Best effort (probabilistic)
 - ▶ Overlay agnostic (unstructured)
 - ▶ Scalable (number of nodes and topics)
- ▶ Building P2P applications by composing layers
- ▶ Coming soon: reliability and persistence layers

This is not the slide you are looking for.

Attenuated Bloom Filters



Node State and Routing

