geographic gossip: efficient aggregation for sensor networks

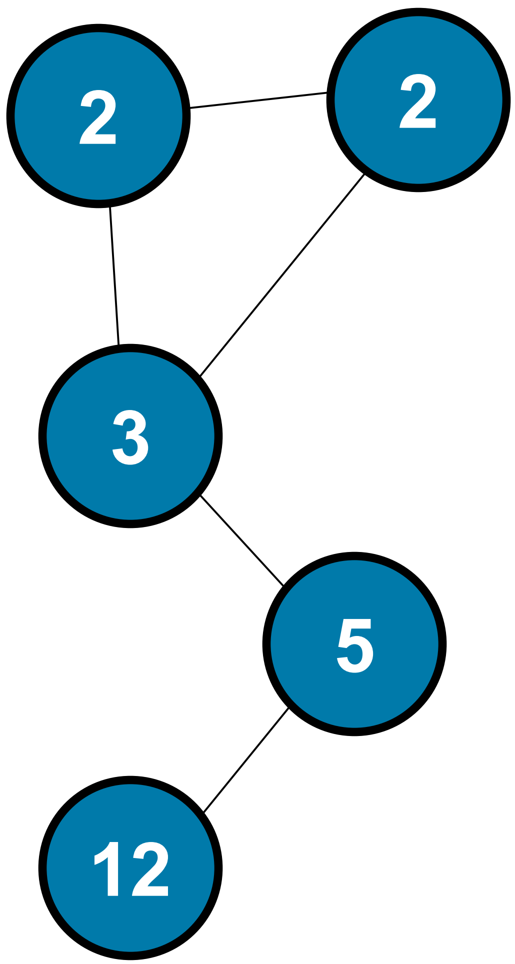
Alex Dimakis, Anand Sarwate, Martin Wainwright

BASiCS group,
EECS and Statistics departments,
UC Berkeley

Outline

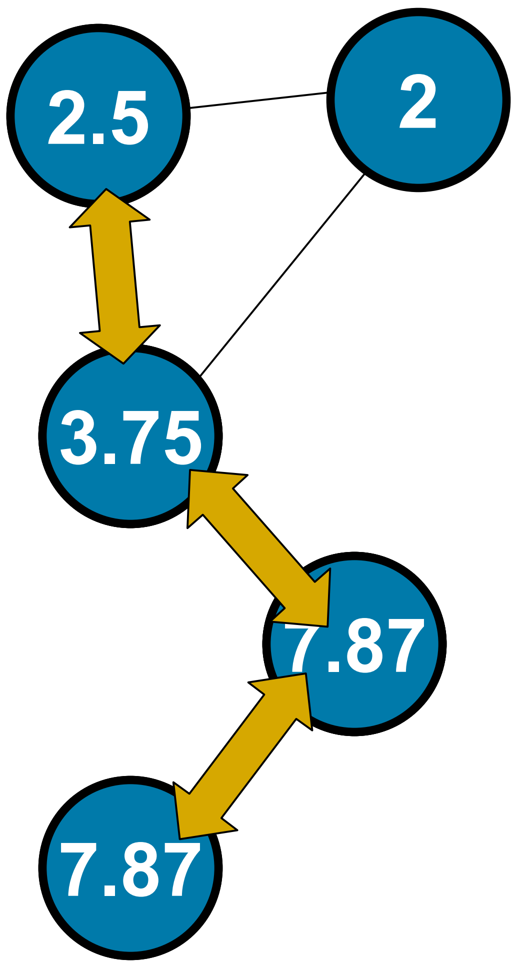
- Introduction: Gossip algorithms for aggregation
- The problem: Gossip is slow
- Random Target routing: How to find a random node
- Solution: Geographic Gossip
- Outline of proof and techniques

Problem: distributed aggregation



- Every node has a number (i.e. sensing temperature)
- **Every** node wants access to **global** average
- Want a truly **distributed**, **localized**, **robust** algorithm to compute averages.

Gossip algorithms for aggregation



- Start with initial measurement as an estimate for the average and update
- Each node interacts with a random neighbor and both compute **pairwise** average
- Converges to true average
- Useful building block for more complex problems

Related work: Alanyali et al. , Boyd et al, Byers et al, Kempe at al, Rabbat et al, Spanos et al, Xiao et al.

How many interactions?

- ε-averaging time: First time where x(k) is ε-close to the normalized true average with probability greater than 1-ε.

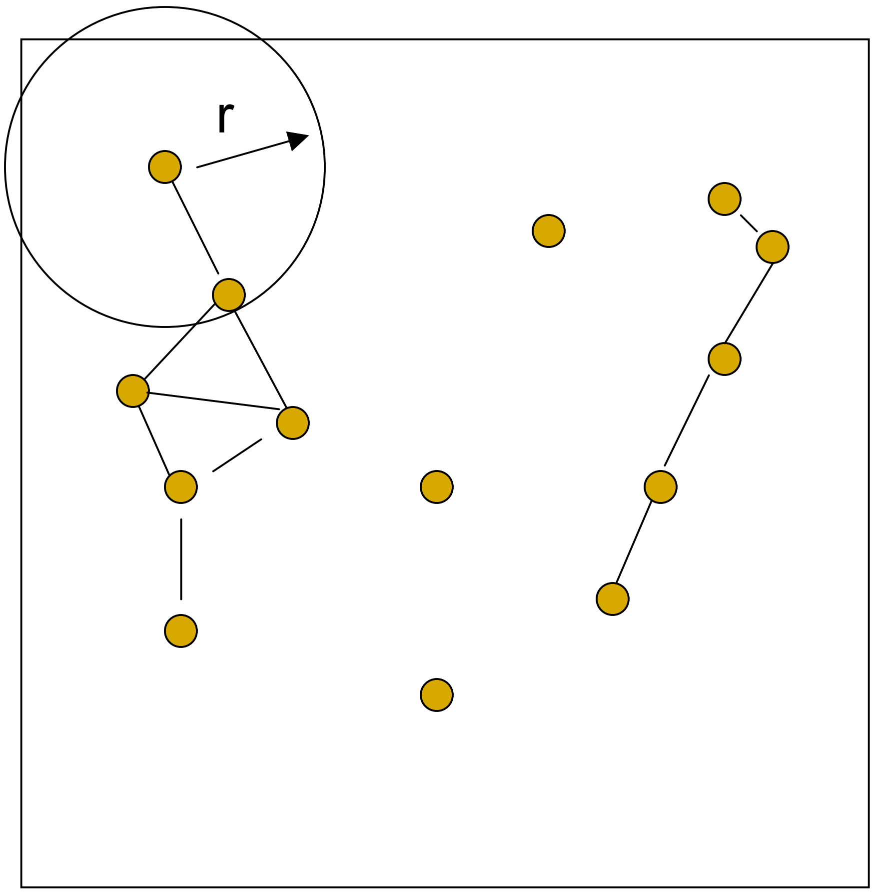
$$T_{ave}(n,\varepsilon) = \sup_{x(0)} \inf \left\{ k : P\left(\frac{\| x(k) - x_{ave}\vec{1} \|}{\| x(0) \|} \geq \varepsilon\right) \leq \varepsilon \right\}$$

- Averaging time connected with mixing time [Boyd et al]:

$$T_{ave}(\varepsilon, n) = \Theta(n(\log n + T_{mix}(\varepsilon)))$$

- Relevant Cost: **Number of Radio transmissions** (fixed Tr radius) , Proportional to total energy spent

Random Geometric Graphs



$$T_{ave}(\varepsilon, n) = \Theta(n(\log n + T_{mix}(\varepsilon)))$$

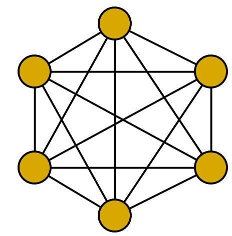
- Depends on graph and the transition probabilities **Realistic sensor network model** (Gupta & Kumar):
- Random Geometric Graph G(n,r): n random points, connect if distance<r

$$r(n) = \Theta\left(\sqrt{\frac{\log n}{n}}\right)$$

Cost of Standard Gossip

$$T_{ave}(\varepsilon, n) = \Theta(n(\log n + T_{mix}(\varepsilon)))$$

- Depends on graph and the transition probabilities:
- Complete graph: $T_{mix}=\Theta(1)$ so $T_{ave}=\Theta(n \log(n))$



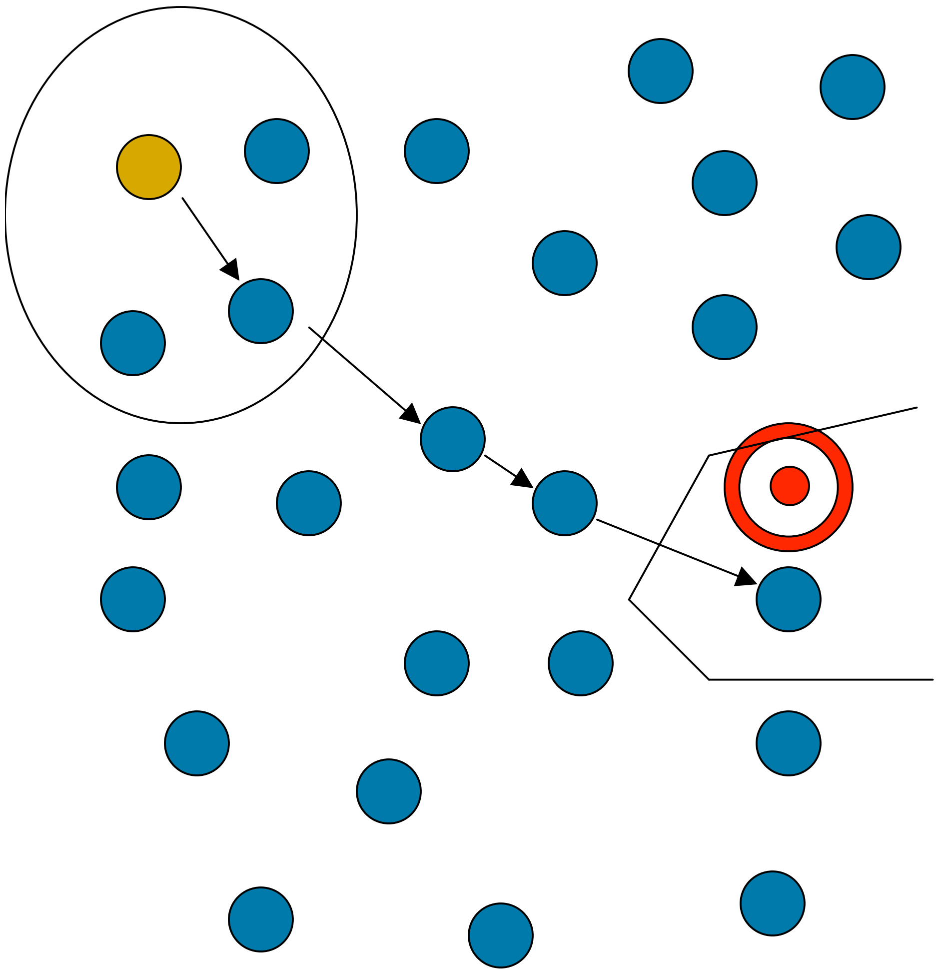
- Small World/Expander: $T_{mix}=\Theta(\log(n))$ so $T_{ave}=\Theta(n \log(n))$
- Random Geometric Graph[Boyd et al]:

$$T_{ave}=\Theta(\log n / r^2) = \boldsymbol{\Theta(n^2)}$$

Cost of standard Gossip

- Standard Gossip algorithms **require a lot of energy**.
  (For realistic sensor network topologies)
- Why: useful information performs random walks, diffuses slowly
- Can we save energy with extra information?
- Idea: Add a random directions to gossip, to diffuse faster.
- Assume each node knows its location and locations of 1-hop neighbors.

Random Target Routing: How to find an (almost) random node



- Node picks a random location (="target")
- Greedy routing towards the target
- Probability to receive ~ Voronoi cell area

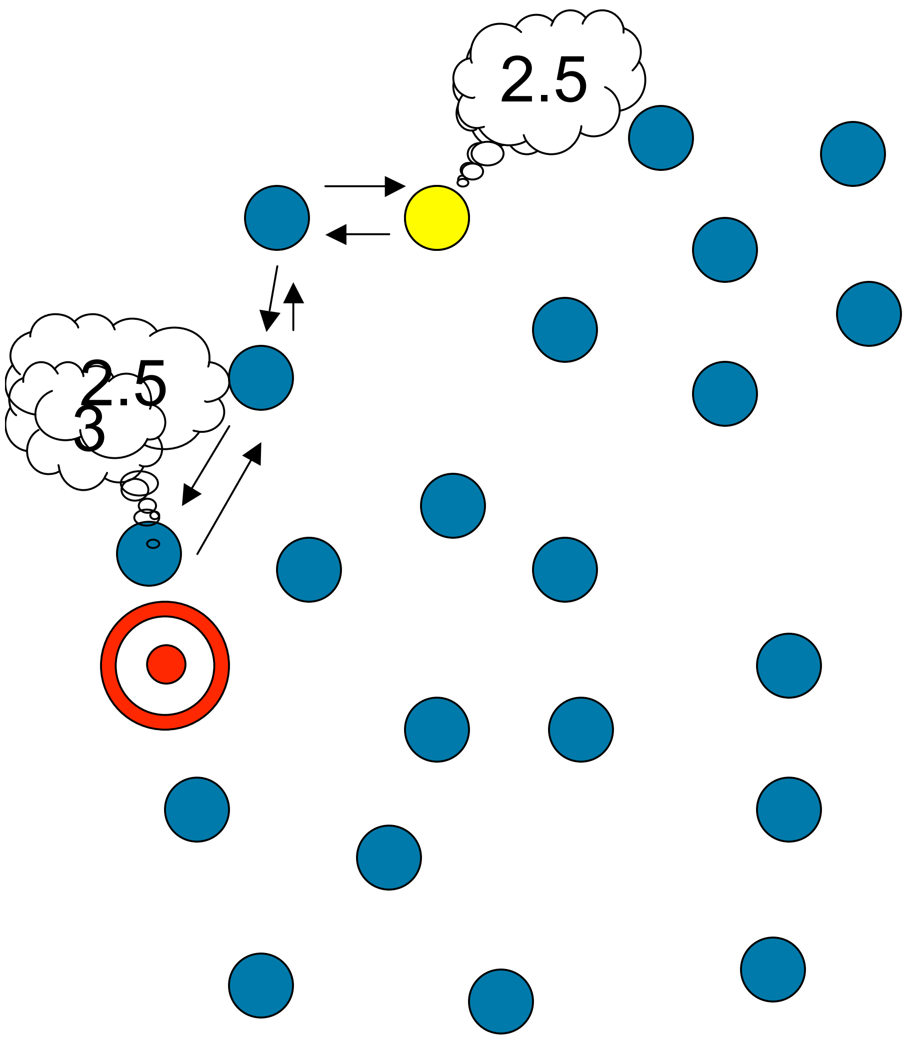
Random Routing: How to find an (almost) random node

- Lemma: if $r(n) \geq \sqrt{10 \frac{\log n}{n}}$

Then by random routing on G(n, r) the following are true with high probability:

- Random routing will transport the packet to the node closest to the random target
- The number of hops will be: $O(\sqrt{\frac{n}{\log n}}) = O(\frac{1}{r(n)})$
- Probability for each node to receive proportional to its Voronoi cell area

Geographic Gossip



- Nodes use random routing to gossip with nodes far away in the network
- Each interaction costs

$$O\left(\sqrt{\frac{n}{\log n}}\right) = O\left(\frac{1}{r(n)}\right)$$

- But faster mixing

Geographic Gossip

- Main Theorem: if $r(n) \geq \sqrt{10 \frac{\log n}{n}}$

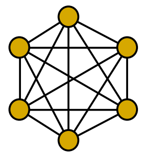
Then, Geographic gossip on G(n, r) has

- Expected averaging cost of $O(n^{3/2}\sqrt{\log n})$
- w.h.p. the averaging cost is bounded by $O(n^{3/2}(\log n)^{3/2})$

- For grid, geographic gossip has
  expected averaging cost of $O(n^{3/2} \log n)$

Sketch of Proof 1: Intuition

- **Assume Voronoi cell areas were uniform**. Each node was selected with uniform probability.
- **Equivalent overlay graph**:



- Complete connectivity
- Equivalent overlay graph, each edge has cost: $O(\sqrt{\frac{n}{\log n}}) = O(\frac{1}{r(n)})$
- Total (Expected) Cost:

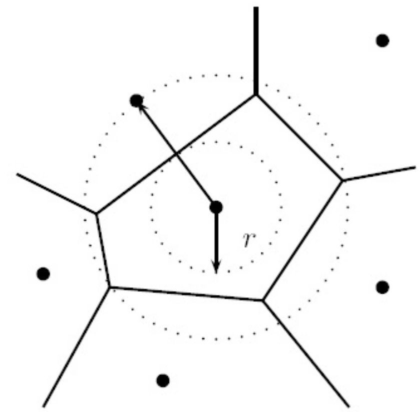
$$T_{ave}(\varepsilon, n) \sim O(\sqrt{\frac{n}{\log n}} n(\log n + T_{mix}(\varepsilon)))$$

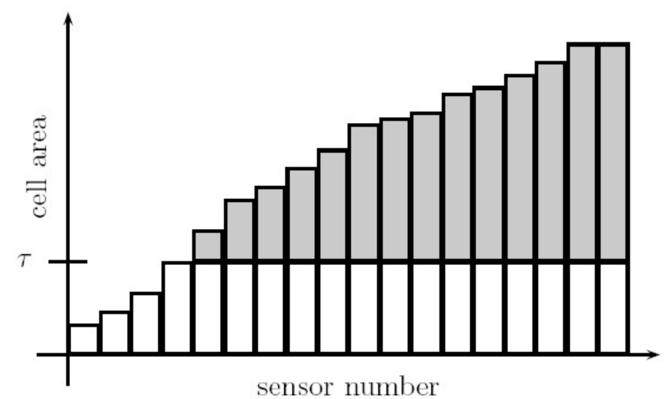
- But $T_{mix} = O(1)$
- Therefore cost should scale like: $T_{ave}(\varepsilon, n) \sim O(n^{3/2}\sqrt{\log n})$

Sketch of Proof 2: Rejection sampling

- **Unfortunately Voronoi cell areas are not uniform**.

- Rejection Sampling: Nodes with large Voronoi areas reject the packet. [Byers et al]
- Lemma: Voronoi areas are not too uneven, rejection probability constant.





Conclusions

- Geographic gossip saves a factor of $n^{1/2}$ in energy required for aggregation. For realistic graph topologies
- Achieves with location information. Only localized distributed operations. **distributed**, **localized**, **robust.**
- Can be combined with other related consensus algorithms