Epidemic Algorithms for Replicated Database Maintenance

Sindre Johansen and Andreas Løvland

Xerox

Paper is from 1987

Goal: Synchronize Clearinghouse servers on the Xerox Corporate Internet.

A name system, like DNS, but more distributed

The Problem

A large database replicated on many nodes

Database-updates are injected on one of the nodes

Eventual Consistency

The Problem

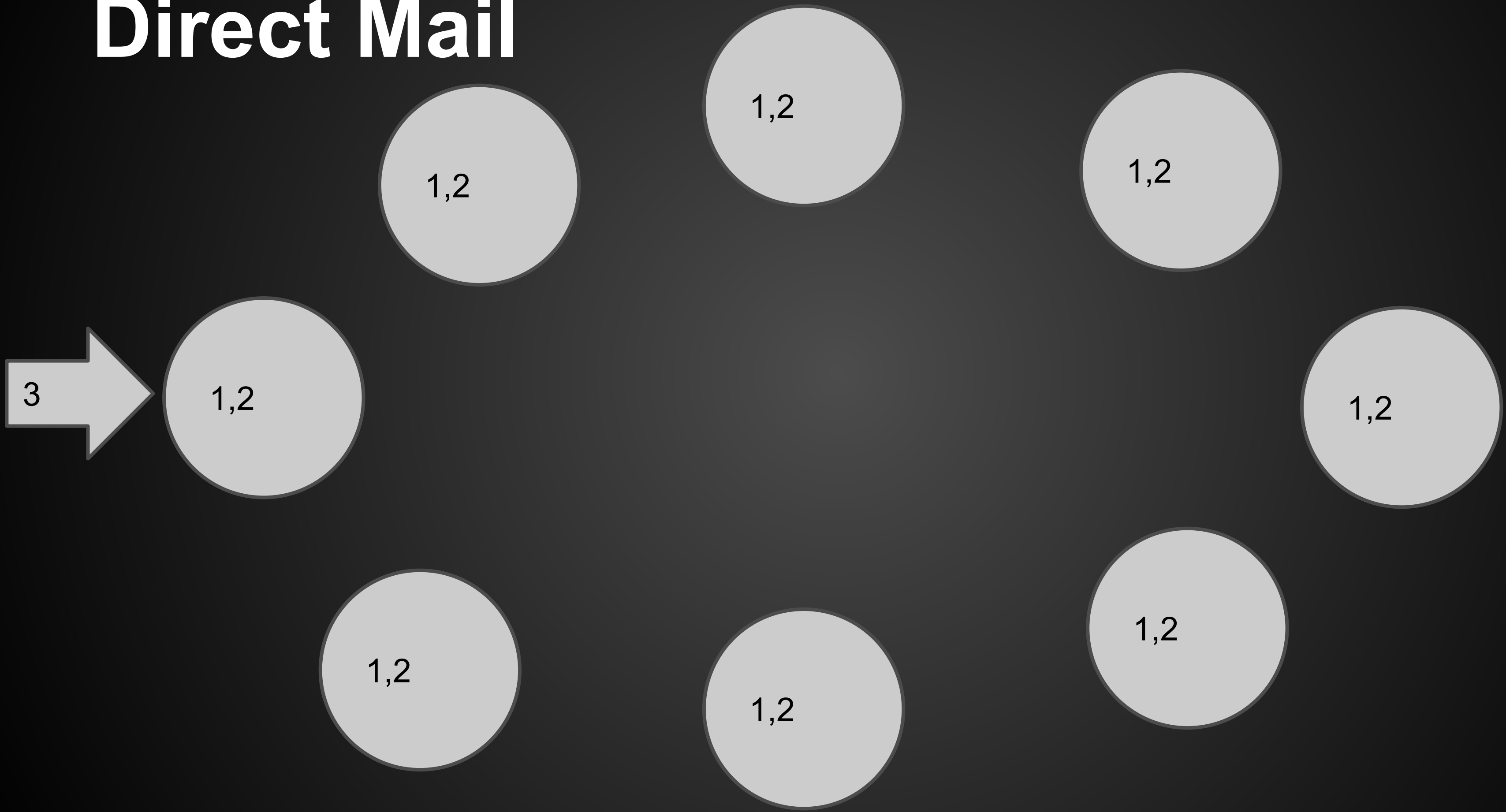
Factors:

- Time to propagate to all nodes
- Network traffic

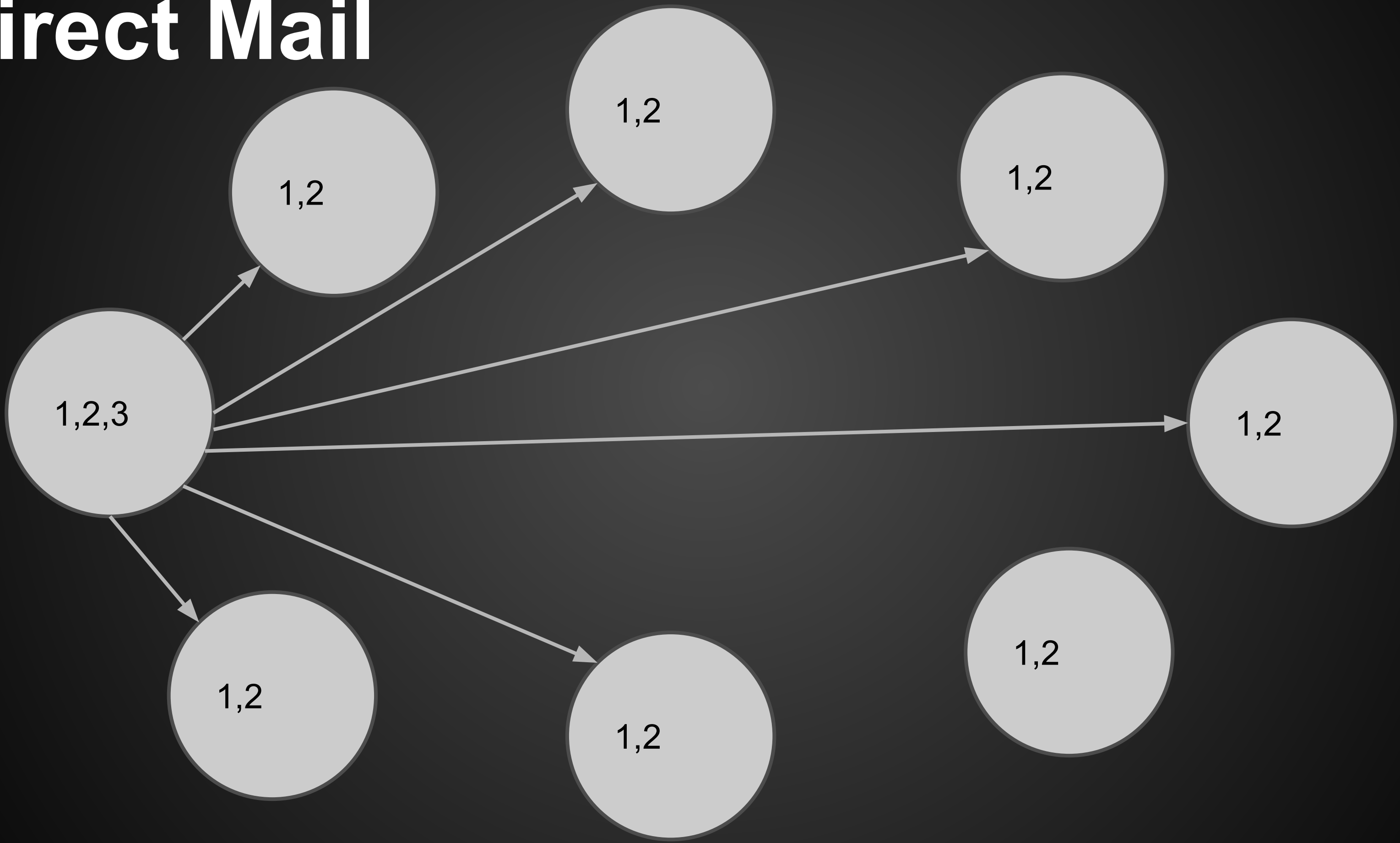
Presented Strategies

- Direct Mail
- Anti-entropy
- Rumor mongering

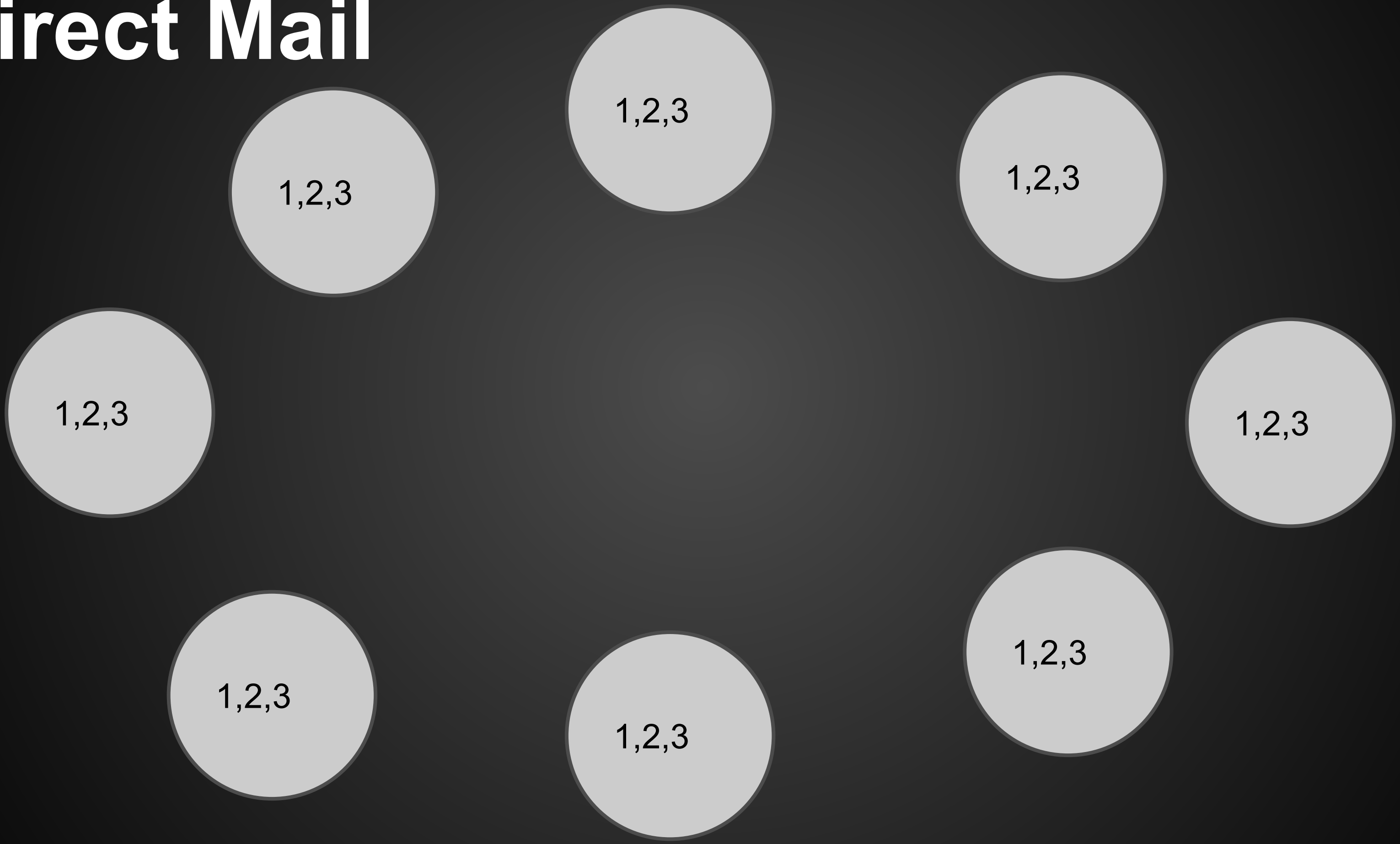
Direct Mail



Direct Mail



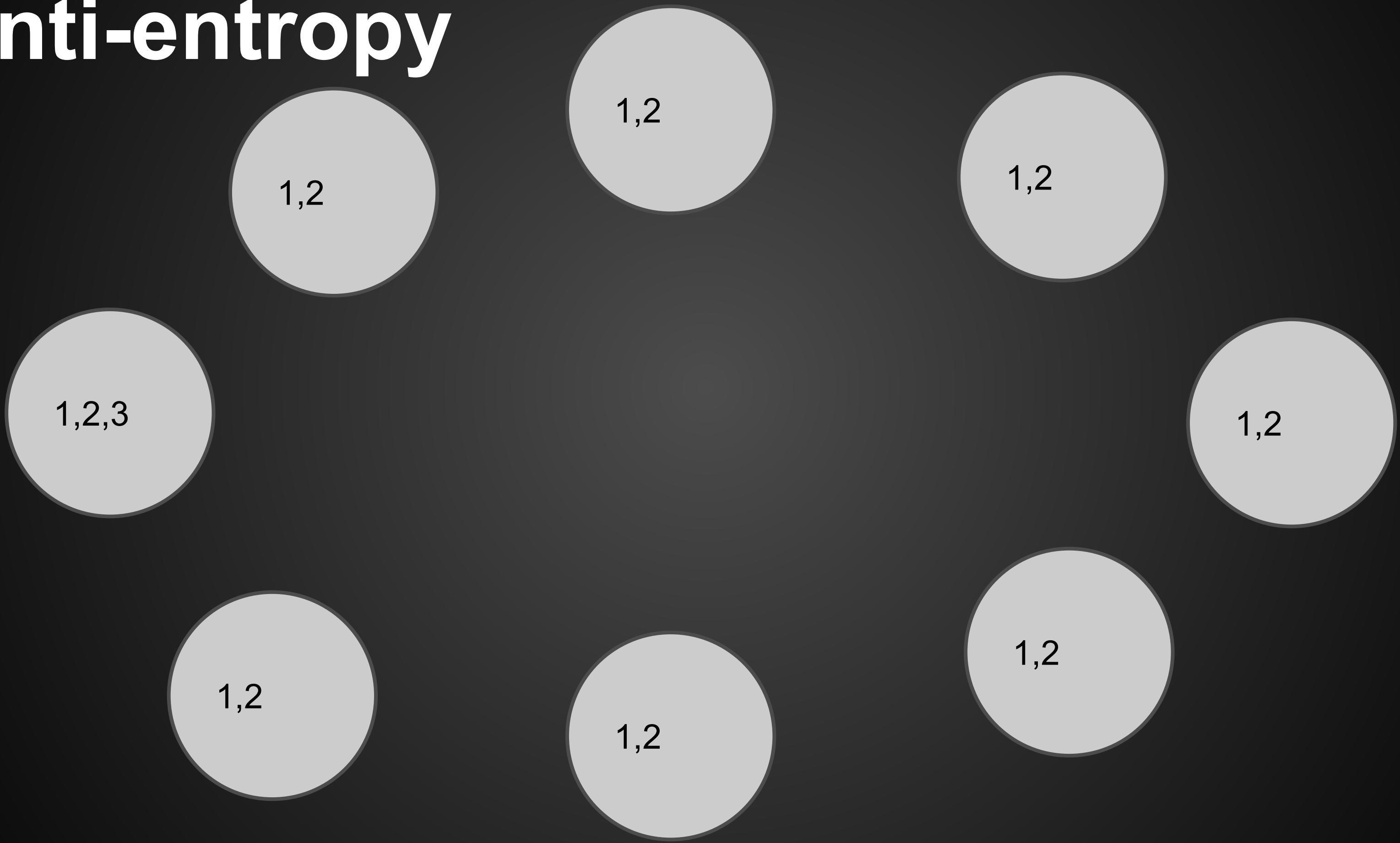
Direct Mail



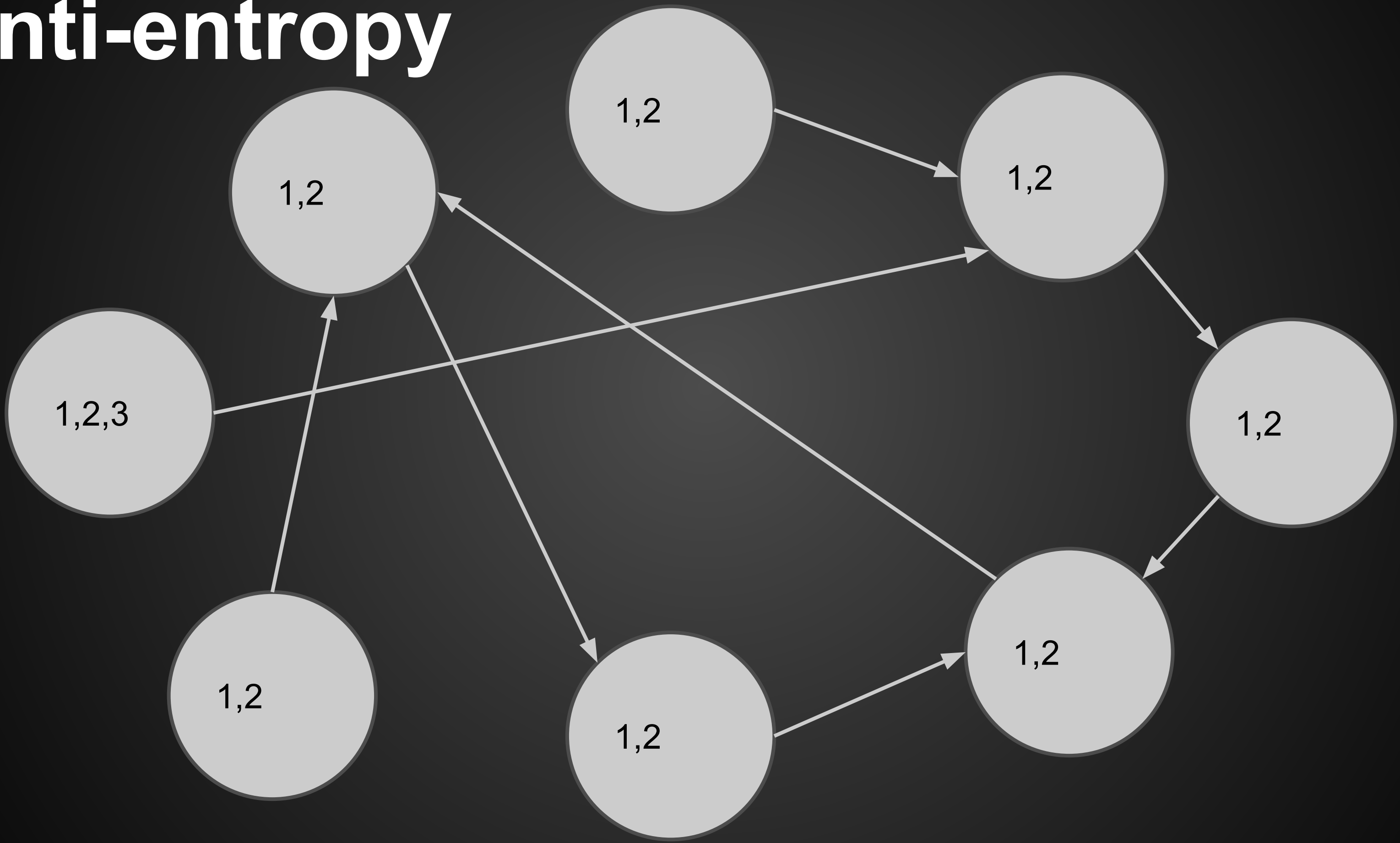
Direct Mail

+ Timely and reasonably efficient
- Needs 100% knowledge
- Mail can be lost

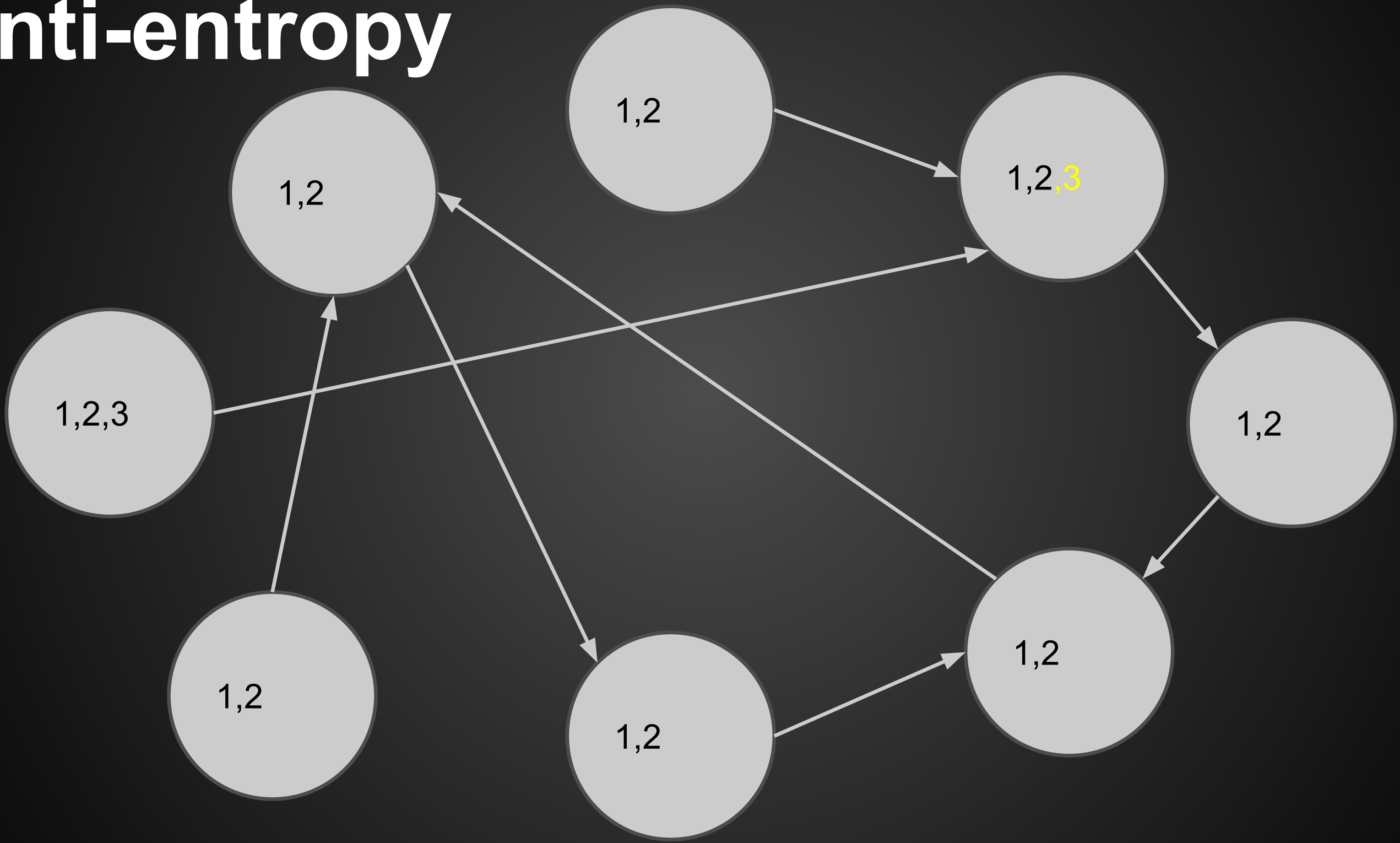
Anti-entropy



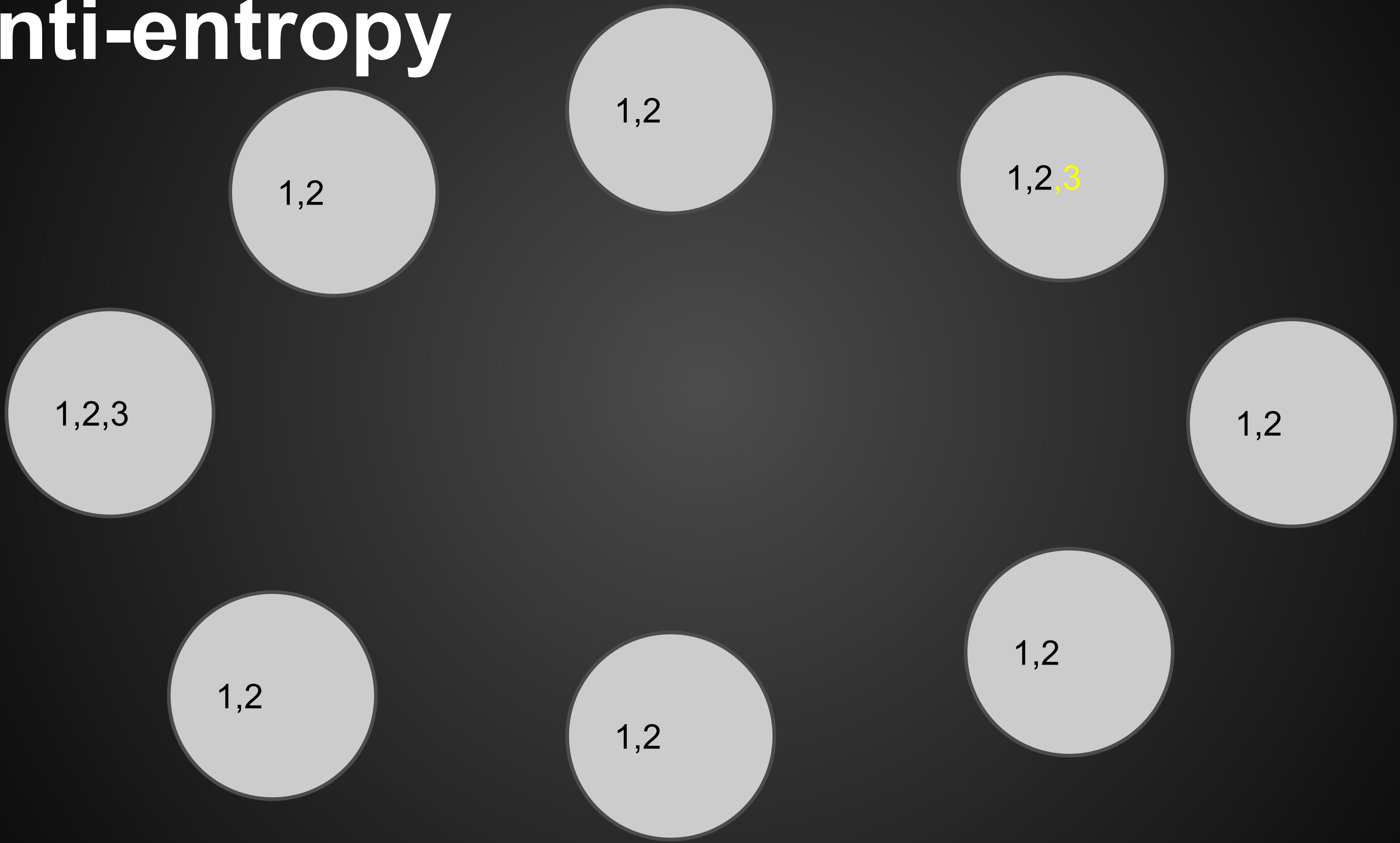
Anti-entropy



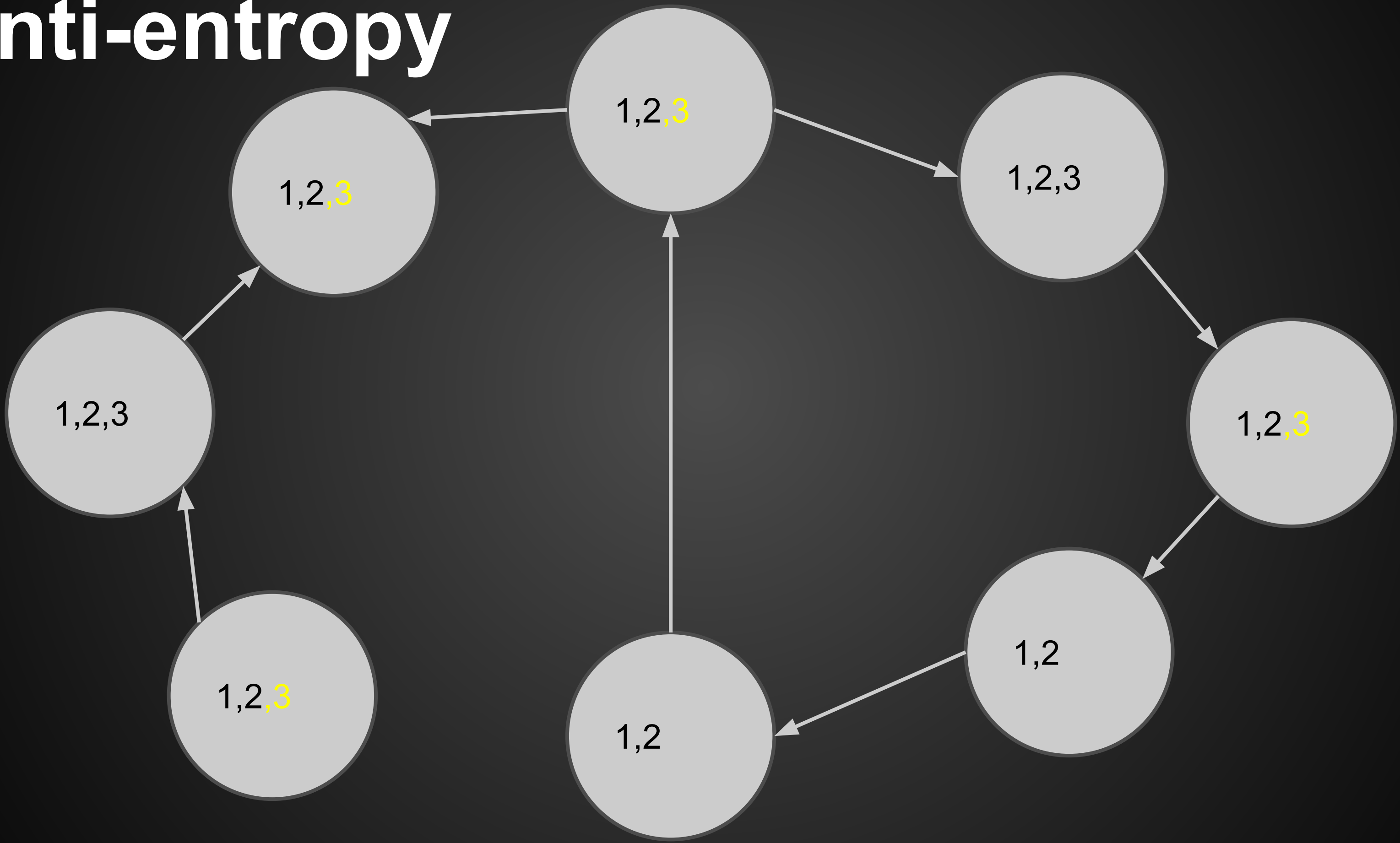
Anti-entropy



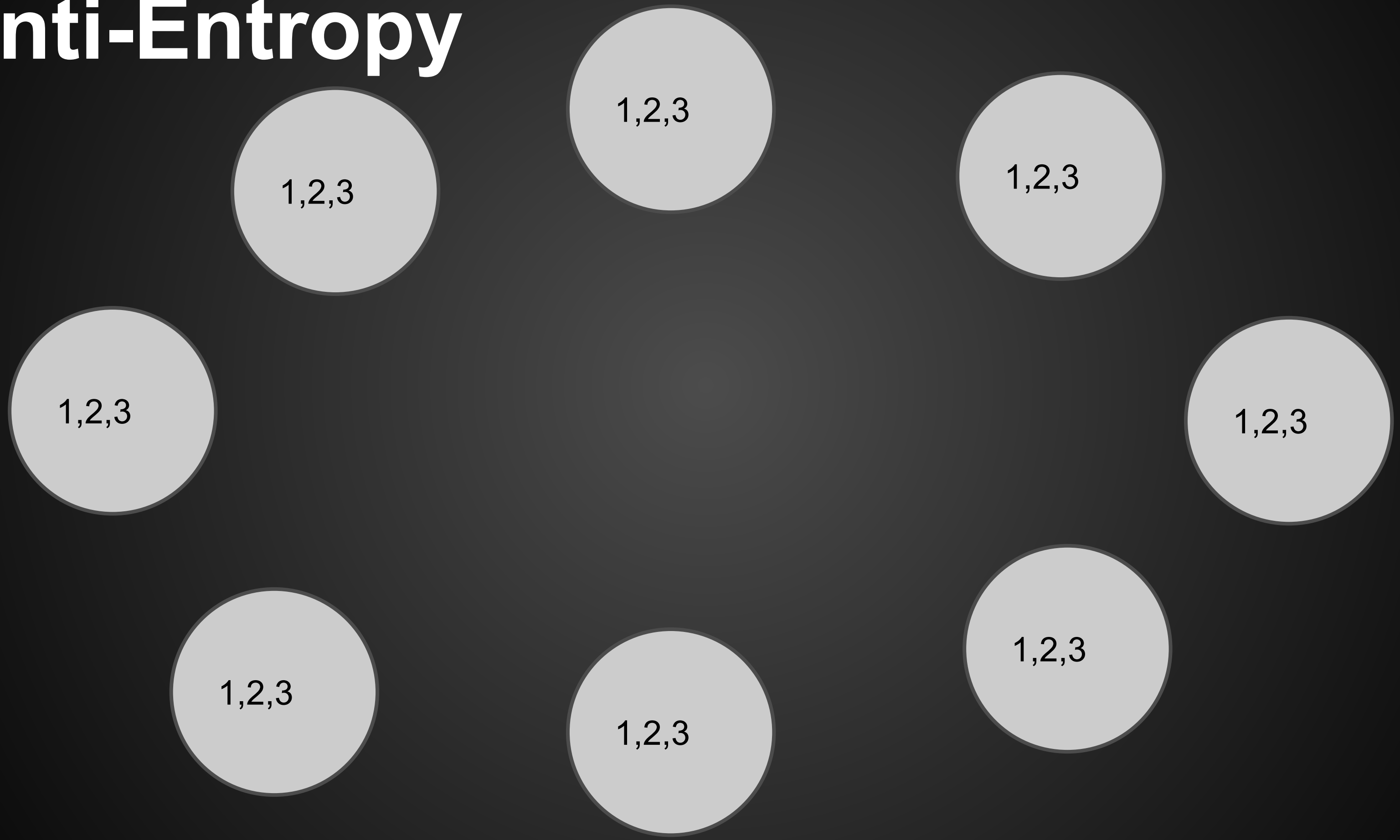
Anti-entropy



Anti-entropy



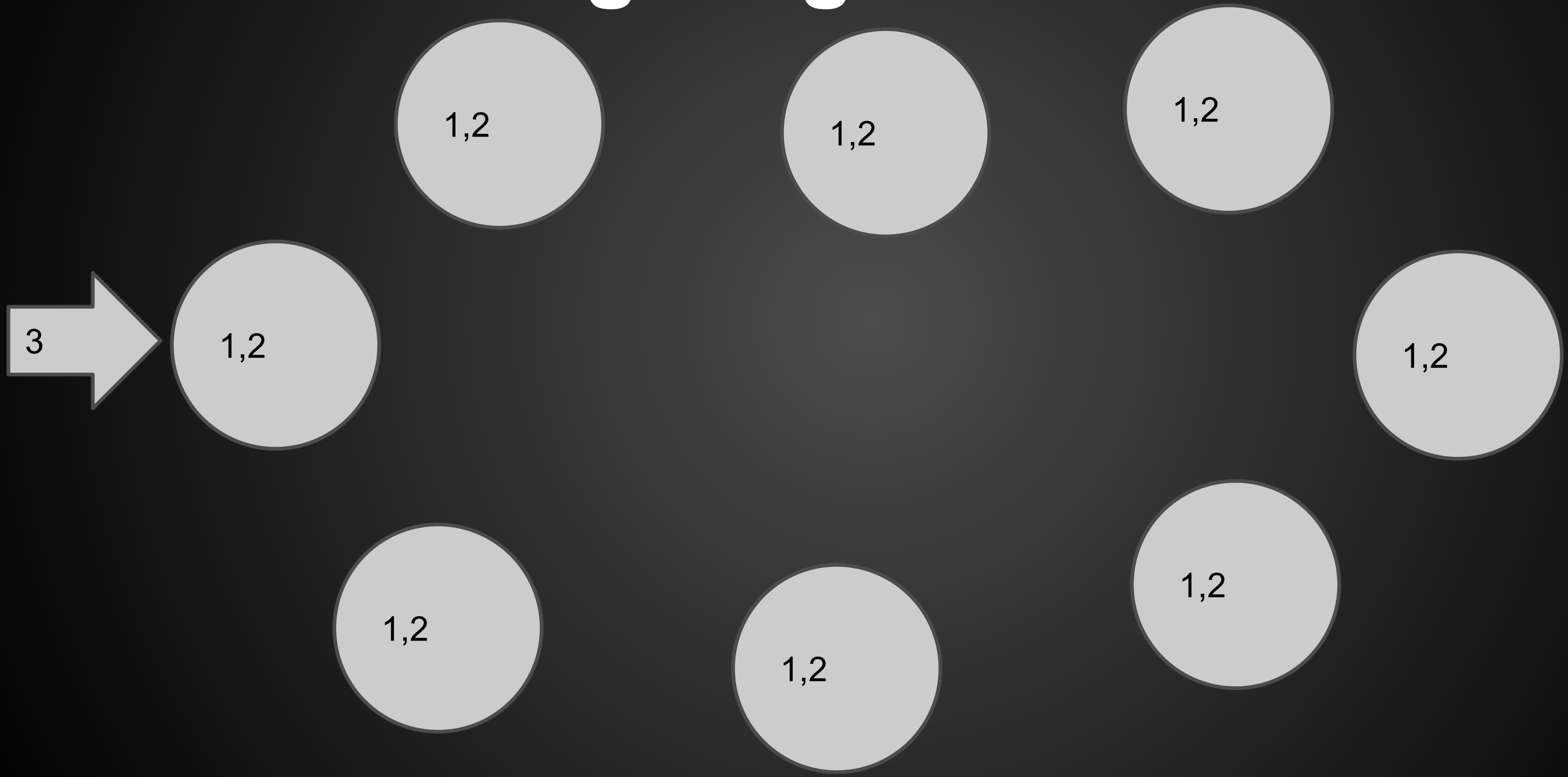
Anti-Entropy



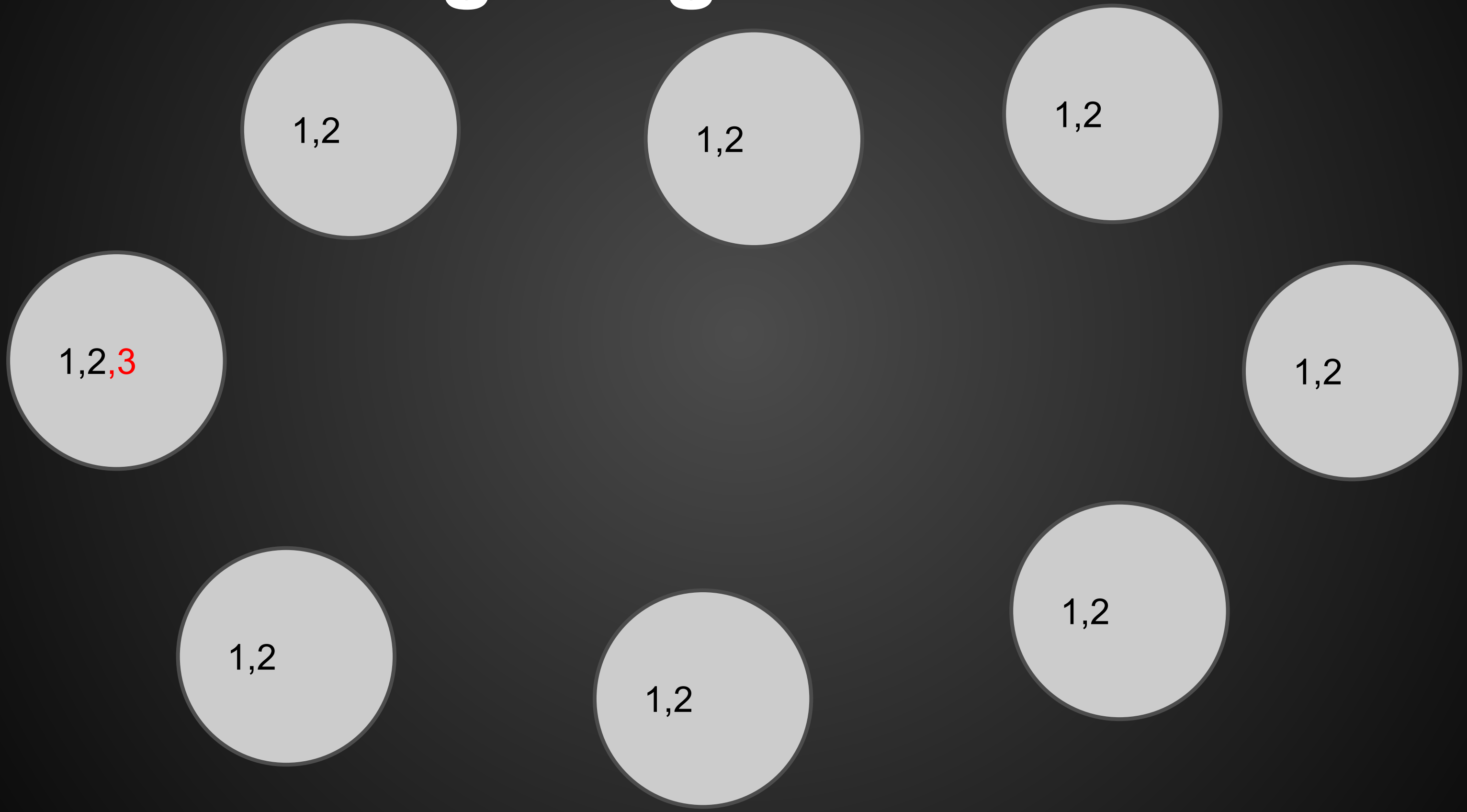
Anti-Entropy

+ Extremely reliable
- Requires lots of computing
- Needs infrequent update interval

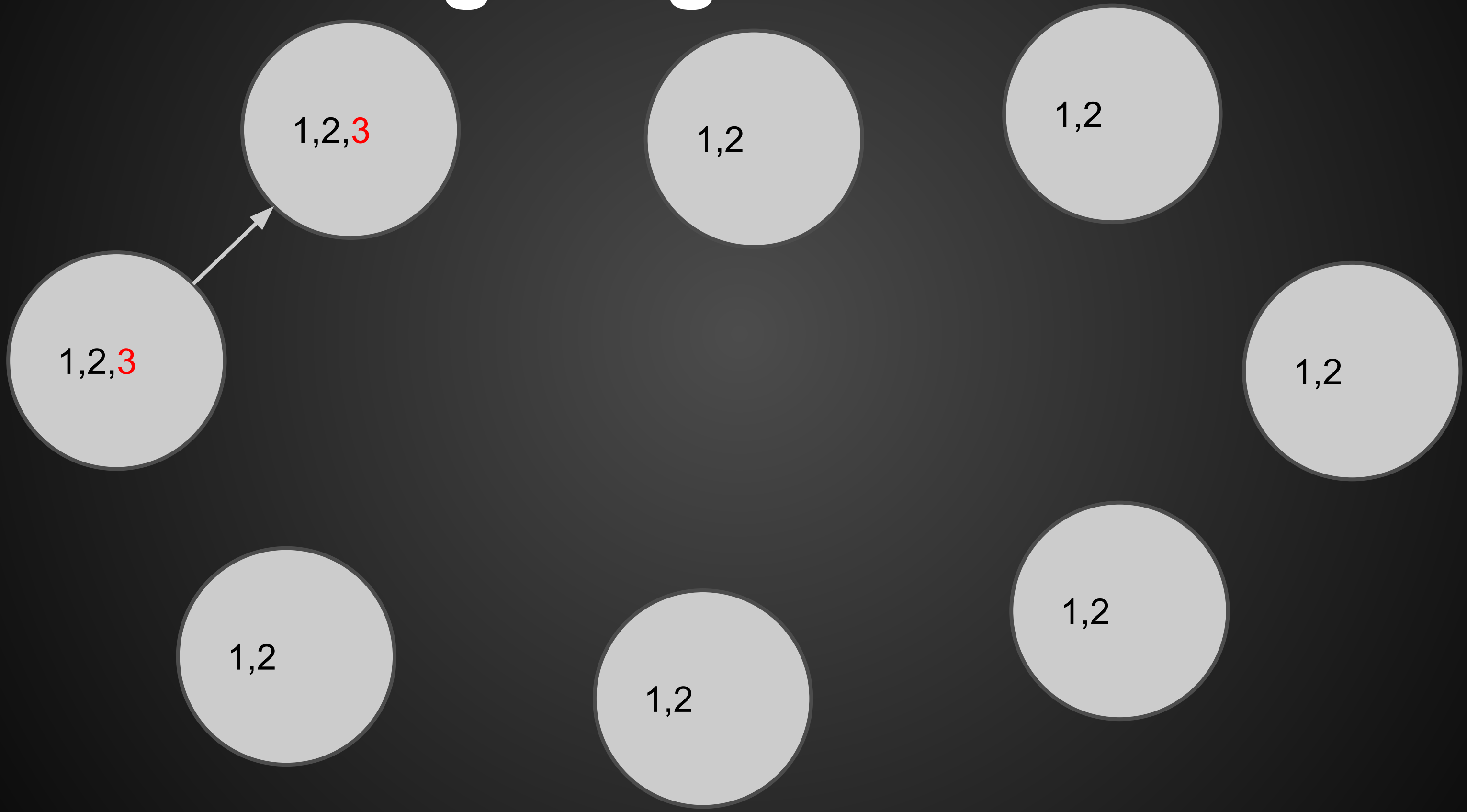
Rumor Mongering



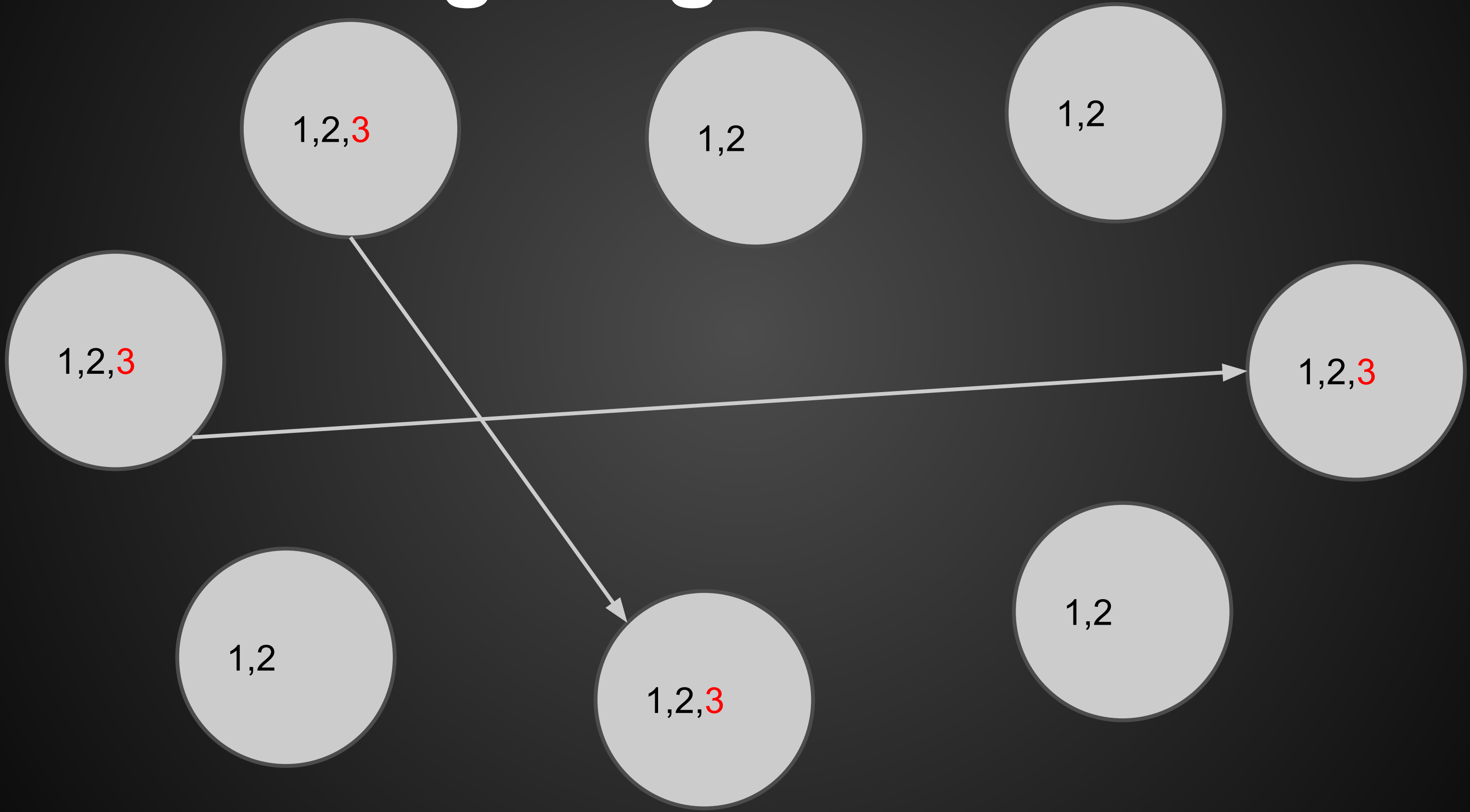
Rumor Mongering



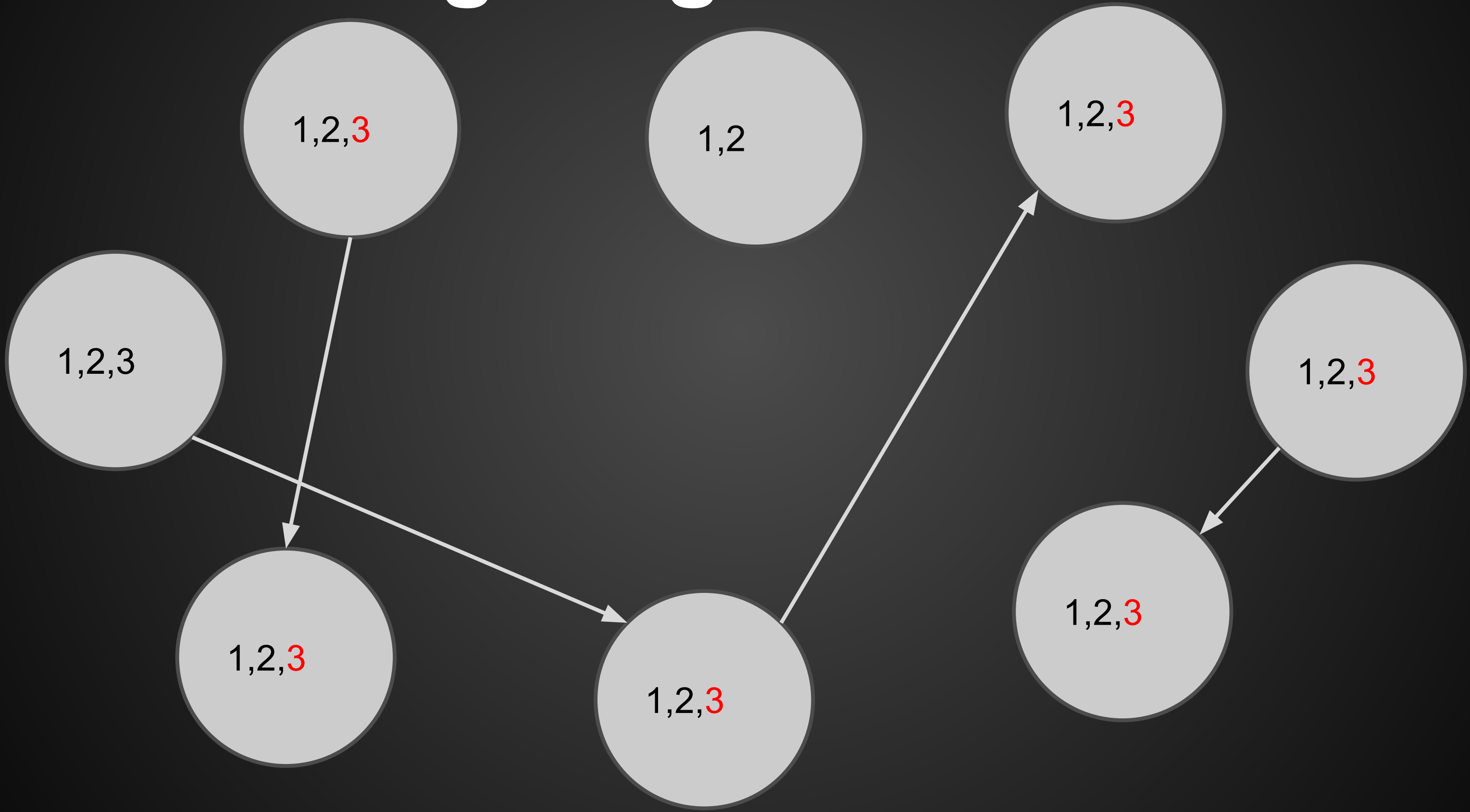
Rumor Mongering



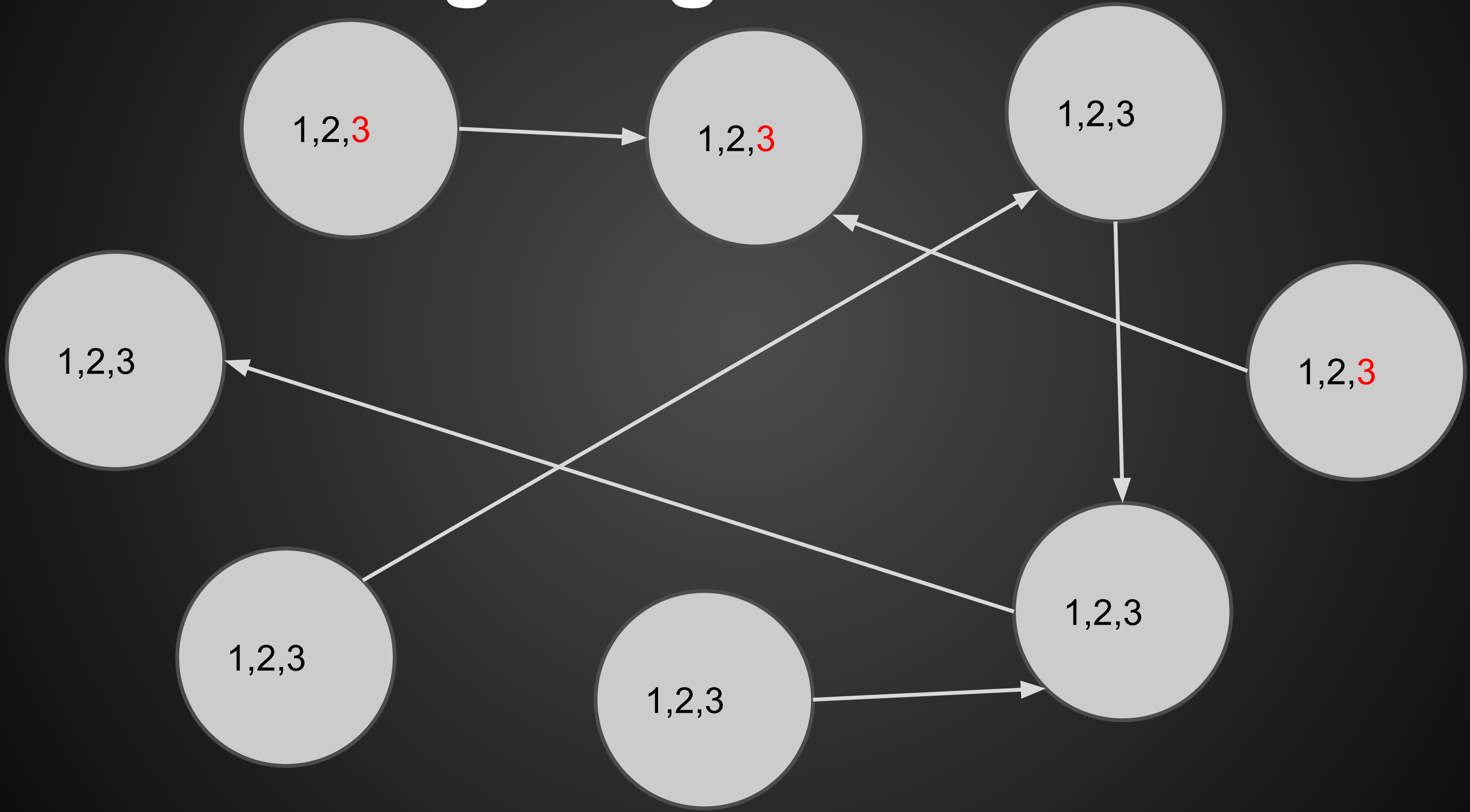
Rumor Mongering



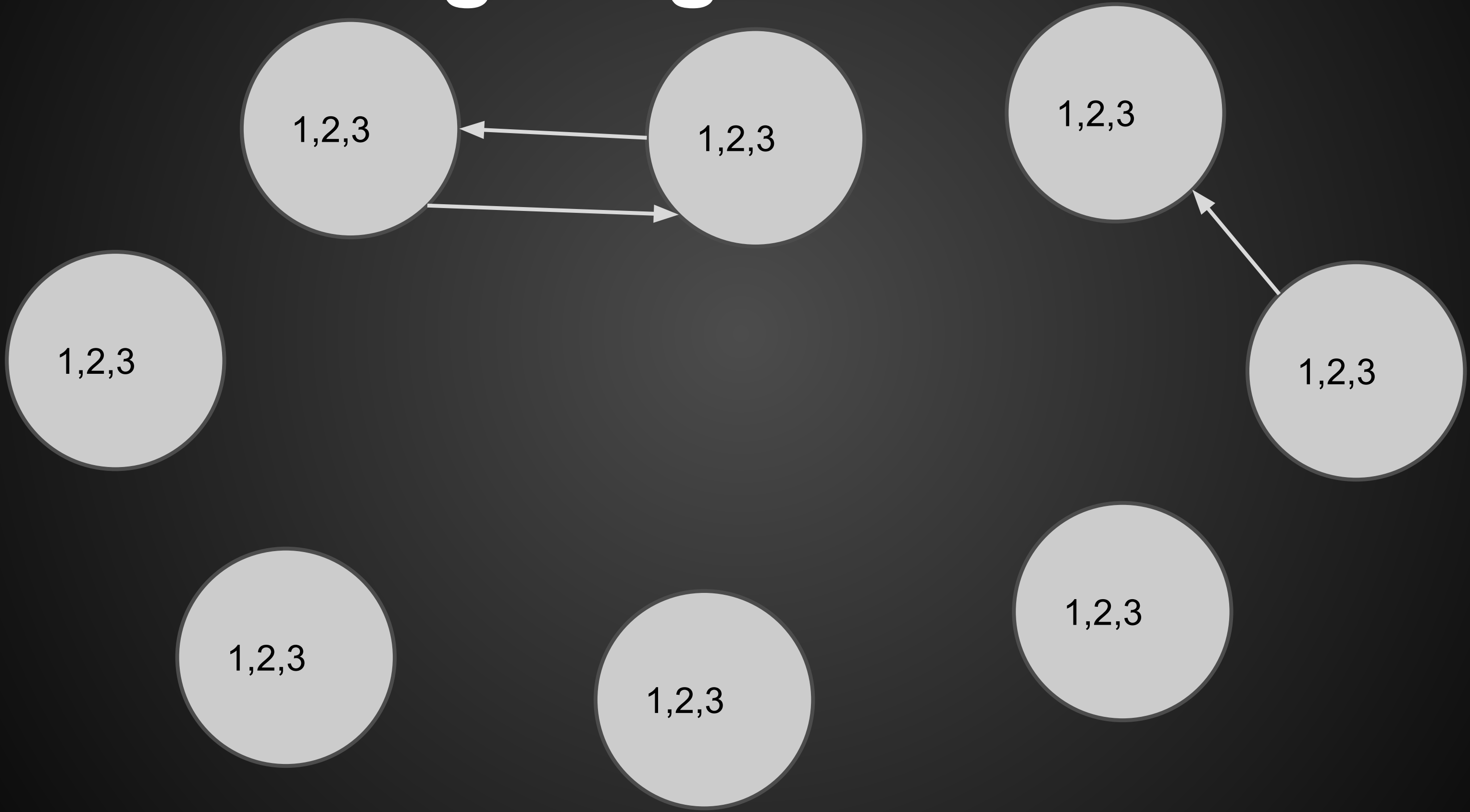
Rumor Mongering



Rumor Mongering



Rumor Mongering



Rumor Mongering

Rumor Mongering

+ Less intensive than Anti-Entropy
+ More reliable than direct mail
- Less reliable than Anti-Entropy

Epidemic processes

Anti-Entropy and Rumor Mongering

Inspired from the theory of epidemics

But we want rapid and complete spread

Susceptible - Not yet received the update

Infective - Received the update

Removed - Received update, no longer sharing

Anti-Entropy

- Proposed as a system to recover from direct mail failures.
- *Push* VS *Pull* VS *Push-Pull*
- Epidemic Theory: The infection will reach the whole population in O(log(n)) time
- p_i = probability of a node is susceptible after round i

$$p_{i+1} = (p_i)^2$$

Anti-Entropy

- Comparison of database: Very Slow
- Use a Checksum!
- Database changes rapidly, checksum will most of the time fail.
- Keep a list of recent updates, check these first, then checksum!

Complex Epidemics

- Susceptible (s) / Infective (i) / Removed (r)
- Stops spreading with prob 1/k for every unnecessary phone call
- Increasing k will decrease the residue.

Complex Epidemics

Measurements

1. Residue. Value of s when i=0
2. Traffic. m = total update traffic / number of sites
3. Delay
   - t_{ave} - average delay
   - t_{last} - receive time of last site

Variations

- Blind VS Feedback
- Counter VS Coin
- Push VS Pull

Performance of Complex Epidemics

Performance of an push epidemic on 1000 sites using feedback and counters.

Counter *k*	Residue *s*	Traffic *m*	Convergence t_{ave}	t_{last}
1	0.18	1.7	11.0	16.8
2	0.037	3.3	12.1	16.9
3	0.011	4.5	12.5	17.4
4	0.0036	5.6	12.7	17.5
5	0.0012	6.7	12.8	17.7

Performance of Complex Epidemics

Performance of an push epidemic on 1000 sites using blind and coin.

Counter ***k***	Residue ***s***	Traffic ***m***	Convergence t_{ave}	 t_{last}
1	0.96	0.04	19	38
2	0.20	1.6	17	33
3	0.060	2.8	15	32
4	0.021	3.9	14.1	32
5	0.008	4.9	13.8	32

Complex Epidemics + Anti-Entropy

- A Complex Epidemics can fail
- Combine with Anti-Entropy
- What to do if a missing update is found.
- Can also be used with peel back

Deletion and Death Certificates

- Can’t just delete a local copy of the data
- Death Certificates spread like ordinary data
    - how to delete the death certificate

Dormant Death Certificates

- propagated with 2 thresholds (T1, T2) and an activation timestamp

Anti-Entropy with Dormant Death Certificates

- What if we want to reinstate a deleted item?
- Death Certificate
    - Initiation Timestamp
    - Activation Timestamp
    - T1 (time-to-live for most sites)
    - T2 (time-to-live as dormant)

- This also works for rumor mongering

Spatial Distributions

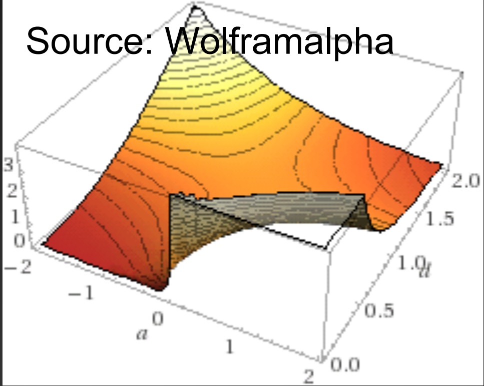
- Nearest neighbor (the one extreme):
  - O(1) link distance, O(n) time to propagate

Links

--O----------O----------O----------O----------O----------O----------O--

Sites

- Random connections (the other extreme):
  - O(n) link distance, O(log(n)) to propagate



Source: Wolframalpha

Spatial Distributions

Analysis shows that:

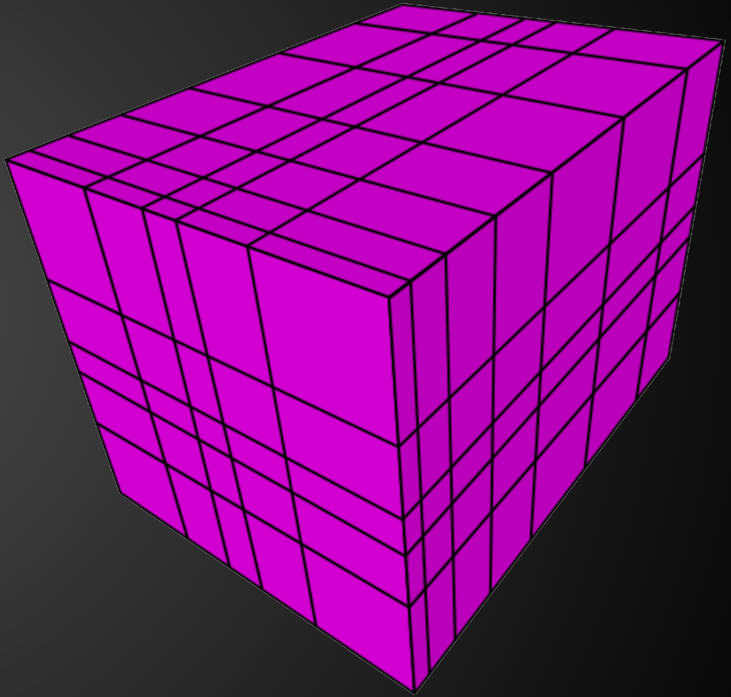
- Probability of connecting to a site at distance *d* should be proportional to d^{-a} where *a* should be 2.

This gives you O(log(n)) traffic per link

$$T(n) = \begin{cases} O(n), & a < 1; \\ O(n/\log n), & a = 1; \\ O(n^{2-a}), & 1 < a < 2; \\ \boxed{O(\log n), \quad a = 2;} \\ O(1), & a > 2. \end{cases}$$

Real networks aren't on a line

- Next paradigm: rectilinear mesh
- Each site independently chooses connections according to $1/(Q_s(d)^2)$
    - Q_s is the cumulative number of sites at distance d or less from s.



Source: wikipedia

Results - no connection limit

Table 4. Simulation results for anti-entropy, no connection limit.

Spatial Distribution	t_{last}	t_{ave}	Compare Traffic		Update Traffic	
			Average	Bushey	Average	Bushey
uniform	7.8	5.3	5.9	75.7	5.8	74.4
$a = 1.2$	10.0	6.3	2.0	11.2	2.6	17.5
$a = 1.4$	10.3	6.4	1.9	8.8	2.5	14.1
$a = 1.6$	10.9	6.7	1.7	5.7	2.3	10.9
$a = 1.8$	12.0	7.2	1.5	3.7	2.1	7.7
$a = 2.0$	13.3	7.8	1.4	2.4	1.9	5.9

Results - connection limit: 1

Table 5. Simulation results for anti-entropy, connection limit 1.

Spatial Distribution	t_{last}	t_{ave}	Compare Traffic		Update Traffic	
			Average	Bushey	Average	Bushey
uniform	11.0	7.0	3.7	47.5	5.8	75.2
$a = 1.2$	16.9	9.9	1.1	6.4	2.7	18.0
$a = 1.4$	17.3	10.1	1.1	4.7	2.5	13.7
$a = 1.6$	19.1	11.1	0.9	2.9	2.3	10.2
$a = 1.8$	21.5	12.4	0.8	1.7	2.1	7.0
$a = 2.0$	24.6	14.1	0.7	0.9	1.9	4.8

To summarize:

- Spatial distributions and anti-entropy can significantly reduce traffic on otherwise hot-spots
- The most pessimistic connection limit slows convergence but does not significantly change total amount of traffic

Spatial Distributions and Rumors

- Rumor mongering is less robust than anti-entropy

- However, we can adjust the *k* parameter to achieve almost identical results
    - this is cool because rumor mongering generates less data traffic
- Conclusion: a nonuniform spatial distribution can produce a worthwhile improvement in rumor mongering

Conclusions

- Xerox has implemented anti-entropy and has gotten good results
- randomized anti-entropy has provided impressive performance improvements
    - spatial distributions reduces link traffic by a factor of more than 4
- Neither direct mail nor anti-entropy can delete items without death certificates