Efficiency in Randomized Rumor Spreading Protocols

Charles Poole

Department of Mathematics,
University of Massachusetts Dartmouth
cpoole@umassd.edu

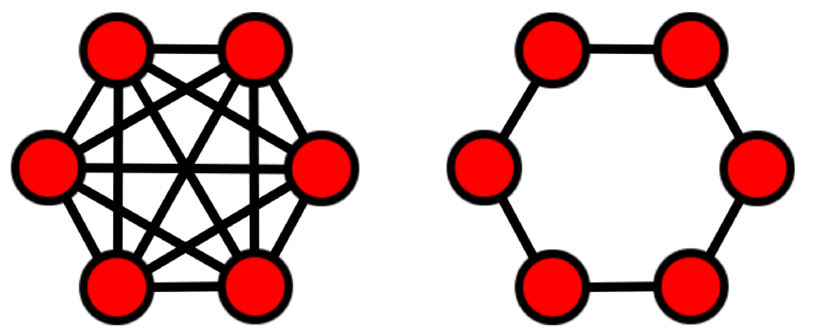
Abstract

Randomized rumor spreading protocols are classically used to spread information across a network. We start with the model proposed by Doerr-Huber-Levavi where each node must follow certain rules to maintain the robustness of a random protocol to insure against transmission failure. We will explore the wastefulness of algorithms with regard to the topology of the network. Also examined is rumor spreading in a non-stationary environment where graphs are formed and reformed based on agent movement. Overall, stationary fixed connection agent, and dynamic transitory agent performance are compared in different contexts.

Objectives

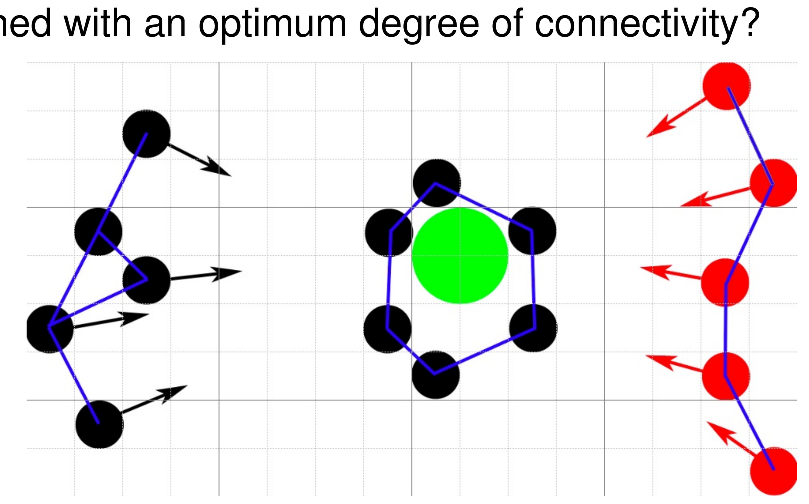
THE OBJECTIVE here is to optimize for wall time. We explore adding a back channel for information and ignoring the previous formula for robustness by node. The increase in complexity might both guarantee robustness and minimize total wall time.

In specific, with stationary fixed connection agents the total computation was directly related with the degree of connectivity, so we focus on efficiency in terms of agent degree.



Complete graph vs Sparse graph.

Currently, there is no mechanism for stopping the flow of information - the current algorithm will just continue for all time. The only thing that could be used to stop it is the above method of obtaining what should be the max number of rounds to guarantee robustness to a given probability. Would such a stop check be more computationally expensive than continuing with a minimum number of rounds when combined with an optimum degree of connectivity?



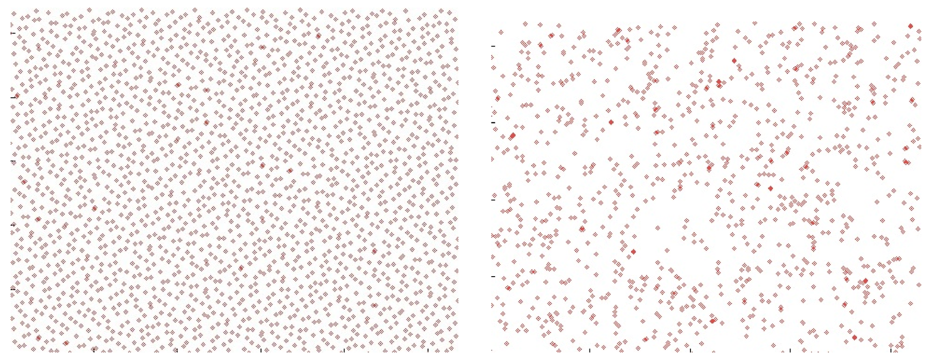
With our transient agent model we're examining grouping and preconditions to make predictions about completion time. In addition we examine methods of releasing information in a manner which minimizes total spread time.

Description of Work

RUMOR SPREADING can describe a number of situations. Here we describe agents as mobile and fixed computing platforms. A wired network is analogous to stationary agents. A mobile platform, laptops or cellphones, would be an example of a transient network. Additionally, we could use the traditional disease model to examine random walks and rumor spreading.

Quasi vs Pseudo Random

First, a quick aside. Quasi-random and Pseudo-Random lead to very different results. It was not immediately apparent. Here's a quick visual demonstration of the difference.

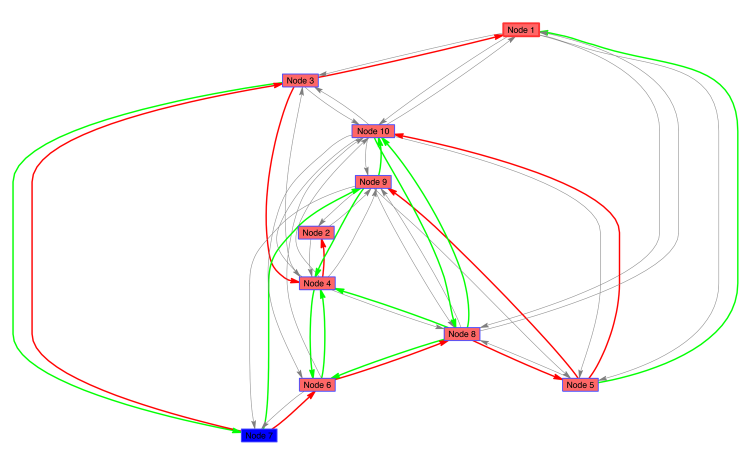


Quasi Random vs Pseudo Random.

It may not seem that important, but the difference between quasi-random points and pseudo-random points leads to a different spread rate. Quasi will spread more uniformly - there will be little to no resends, while pseudo-random will frequently resend data to the same node.

Stationary Fixed Connection Agents

- Quasi-Random rumor spread has rules beyond spray and pray (Classic "Randomized rumor spread method")
- How do we know it's penetrated the whole network? No absolute assurance.
- Are these protocols incredibly wasteful?



Red is Original Inform Vector,
Green is Re-informing,
Grey is unused.

Based solely on the number of nodes we can show that a rumor **should** be spread across an entire graph after $(1+\varepsilon)(\frac{1}{\log_2(1+p)}\log_2 n + \frac{1}{p}\ln n)$ rounds with a probability of $1 - n^{\frac{-p\varepsilon}{40}}$

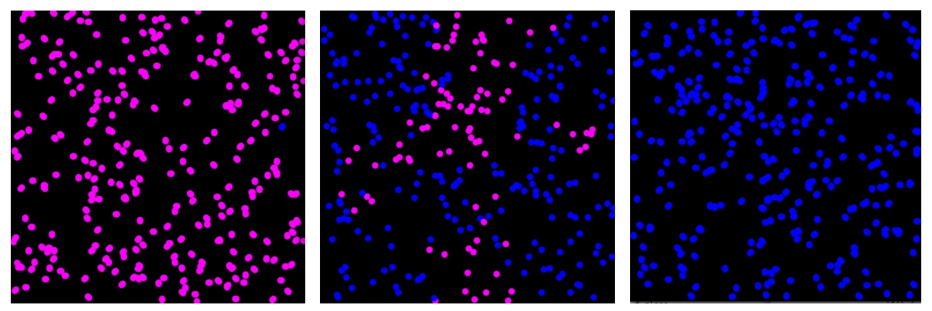
Drawbacks on Demonstrated Model

That's great, for robustness. A couple questions beyond just robustness.

- Can we minimize the number of time steps?
- Does the degree of connectivity matter in this context?
- Can we optimize the amount of computation per time step in terms of connectivity?

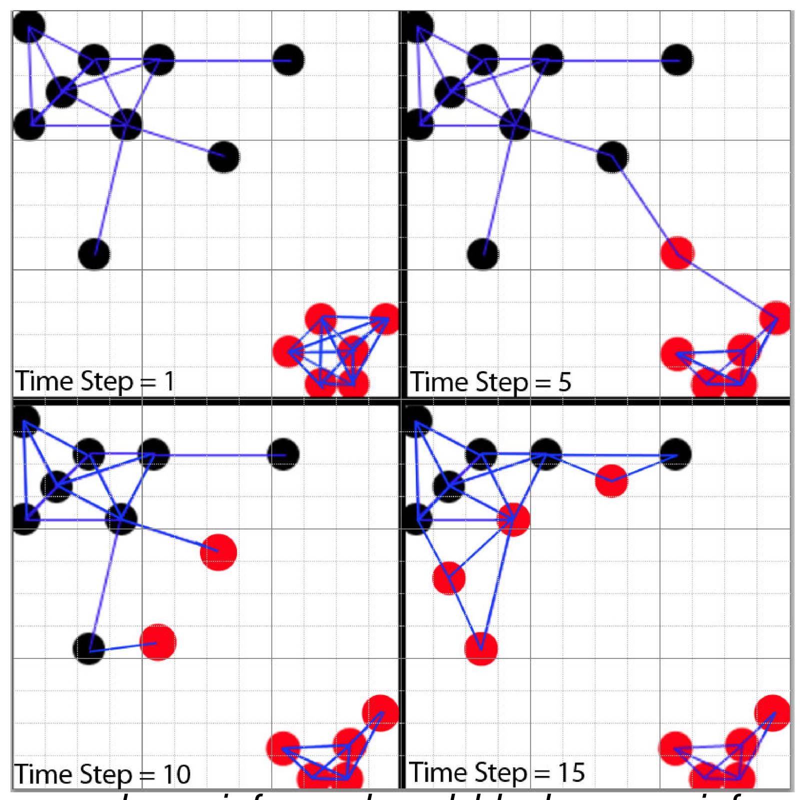
Connectedness becomes important when testing for efficiency. A complete graph will require more computation per time step than a sparser graph.

Transitory Dynamic Connection Agents



Progression of Transitory Agent Model

Within this model all agents were allowed to move unconstrained about a grid on a torus. Agent movement, dynamic graphing, and remote sensing were allowed to wrap on the edges. Allowing for dynamically formed graphs in an agent based model, we allow agents more chances to be informed as total informed increases.



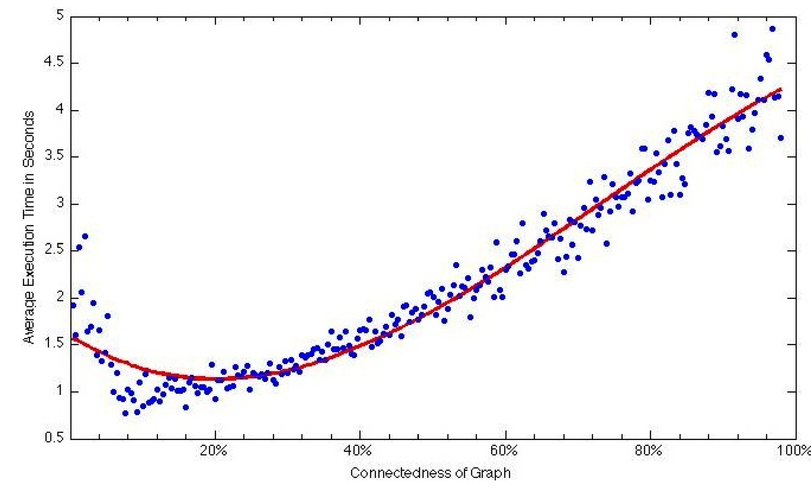
Here red are informed and black are uninformed. The blue lines represent the connections based on proximity.

The above is a simplified example of the actual process. All agents are moving randomly and form their connections based on proximity. As they connect in a quasi-random spread they will immediately pass on the rumor to any agent with which they connect. When coming into proximity with multiple nodes they will use the same selection process as the stationary model. The agent will select one node in a quasi-random manner and pass on the rumor.

Results and Research

MUCH OF THE CURRENT optimization work we have done has focused on the stationary model. With fixed agents and fixed connections between them, rapid evaluation of the model is much simpler than rapid evaluation of the dynamic agent model. With that in mind, most of the work we were interested in concerning stationary agents as been finished, and the new focus is on transient agents.

Stationary Agents



Approximation of Complexity using Time

This graph represents the execution time vs the average number of connections per node. The minimal execution time can be explained by realizing a sparser graph will have less total computation per round ($\mathcal{O}(n)$), but takes on average a larger number of rounds to be fully informed.

The scheme for Quasi-Random rumor spreading is extremely wasteful for the sake of robustness.

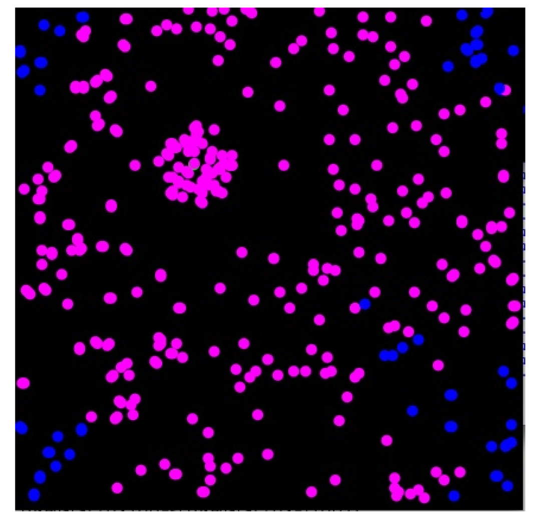
- For a fully connected graph, every round requires $\mathcal{O}(n^2)$ calculations, where n is the number of nodes.
- For a very sparse graph, every round requires $\mathcal{O}(n)$ calculations, where n is the number of nodes.

Experimentally we have shown that there is an optimal total execution time.

- Exists where the number of edges between nodes is enough that there is a reasonable chance for each node to be informed
- Not so large that the number of steps per round approaches our upper bound of $\mathcal{O}(n^2)$.
- This optimal value is expressed in connectedness of a static graph.
- The value represents the average percentage of other nodes to whch each node is connected, and is around 12.8%.

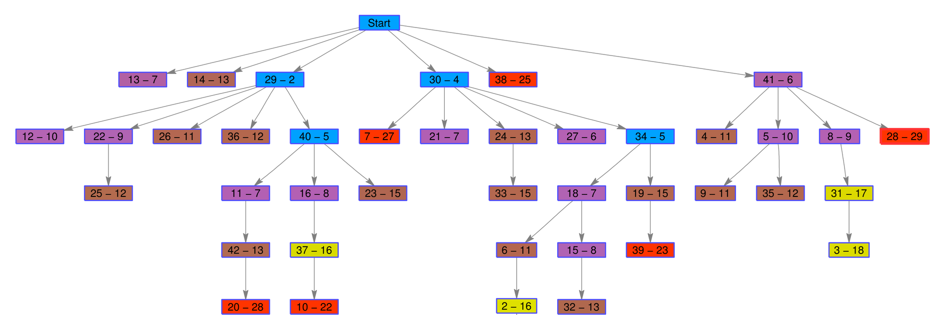
This value was found by repeated simulation of graphs of increasing size. As total number of nodes increase the average converged to having approximately 12.8 percent of nodes connected to each node.

Transient Agents



Transient Agent Model, featuring attractor.

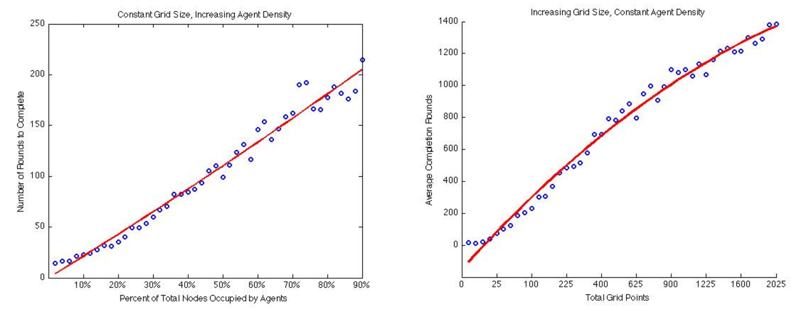
Currently we are tracking the path of a rumor through a transient network. The most important thing tracked at the moment is who informed whom, and at what time step.



Flow of Rumor through Agents by Number

The above graph shows the spread by time step and informer. There is very little apparent structure in the flow of information, so there seems to be little that can be done at that step to increase the completion rate and minimize computation.

We have performed some experiments related to grid size and agent density to get an idea of what happens when these vary.

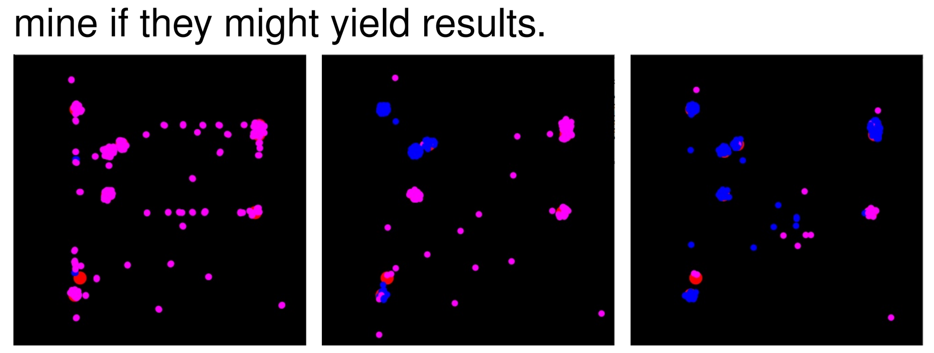


Increasing Density, and Increasing Grid Size

The results so far are pretty basic. When the number of agents for a given grid is increased the increase in completion time is linear. When the grid size is increased and the number of agents or agent density is kept static, the increase in time is logarithmic.

Future

Currently, we are working on agent structure and attractors. Little can be done with the movement of agents or the way they interact. Initial conditions and gathering places are being investigated to determine if they might yield results.



Demonstration of selective attraction

SIAM AN10, Society for Industrial and Applied Mathematics Annual Meeting, 12 - 16 July 2010, Pittsburgh, Pennsylvania