Evaluation of Lower Bound on Quasi-Random Rumor Spreading Protocols

Charles Drake Poole

Undergrad; Department of Mathematics,
University of Massachusetts Dartmouth
cdrakep@gmail.com

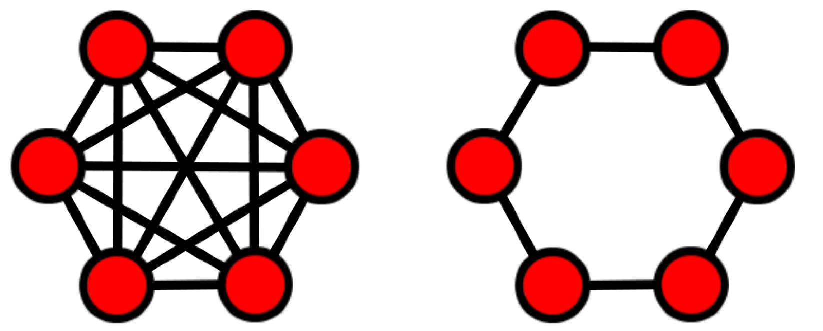
Abstract

Quasi randomized rumor spreading protocols are used to spread information across a network. Doerr-Huber-Lavavi showed that a quasi random rumor spreading protocol is as robust as the randomized protocol. They estimate the time steps required for robustness on a complete graph and use this estimation for every spanning subgraph. It is easily demonstrable that the spread of information across a graph is not dependent only on the node count, but also the edge distribution. Through simulation we have shown there exists a subnetwork of a network with an average degree of connectivity exceeding 13% with other nodes that represents a the minimum spread in complexity, with only an increase in average rounds of 14%. Our results show there exists a lower bound than the one proposed for complexity with the given probability of transmission failure when connectivity is considered.

Objectives

THE OBJECTIVE 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.

Using stationary fixed connection agents the total computation is directly related with the degree of connectivity; so we focus on efficiency in terms of agent degree.



Complete graph vs Sparse graph.

The quasi-random protocol explored has no formal mechanism for stopping the flow of information. The current algorithm will continue n number of rounds, where n is a number of rounds determined by a probability that it will have spread to the entire graph. With the quasi-random protocol proposed by DHL, connectivity between agents is ignored. We look at adding methods for minimizing the number of time steps including stop checks and connectivity modifiers.

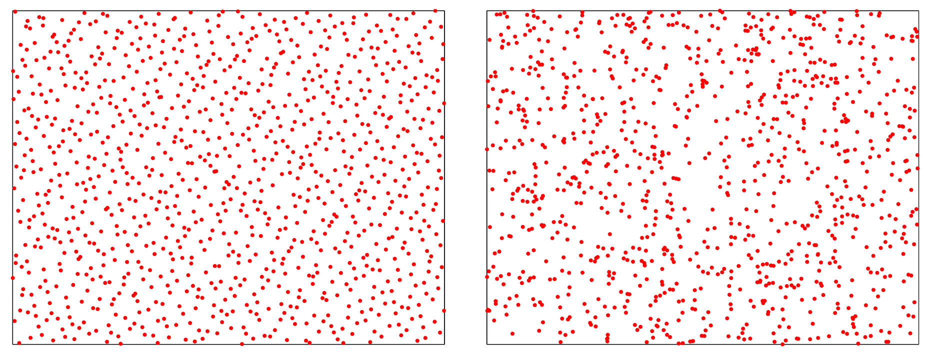
In short, it will be shown (numerically) that the lower bounds set by Doerr-Huber-Lavavi [?] for a rumor to spread through a graph were higher than necessary in most cases. It will also be shown the distribution of edges on a graph is as important as the node count for setting an upper and lower bound for completion. Also considered is the impact of transmission failure and showing that given a certain probability, p, of transmission failure, the lower and upper bounds for completion are still as dependent on edge distribution as on node count.

Description of Work

RUMOR SPREADING has been employed in a variety of applications. 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 picture random walks and rumor spreading as the traditional disease model.

Quasi vs Pseudo Random

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

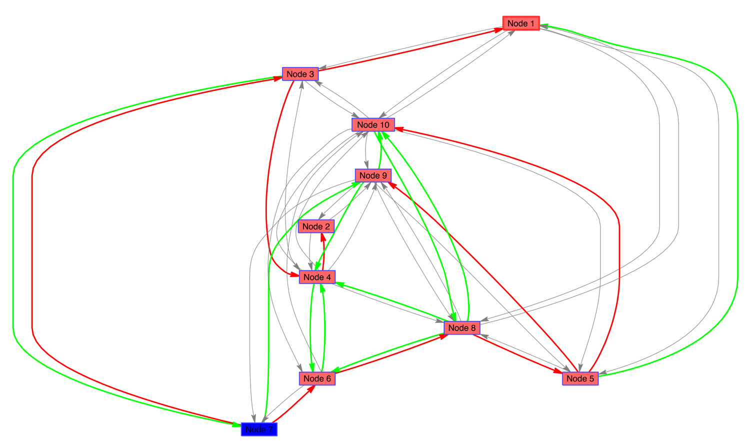


Quasi Random vs Pseudo Random.

The difference between quasi-random points and pseudo-random points leads to a different spread rate. A quasi-random protocol will spread information more uniformly, with few if any resends. Pseudo-random, however, could frequently resend data to the same node.

Stationary Fixed Connection Agents

- Quasi-random rumor protocols have 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,

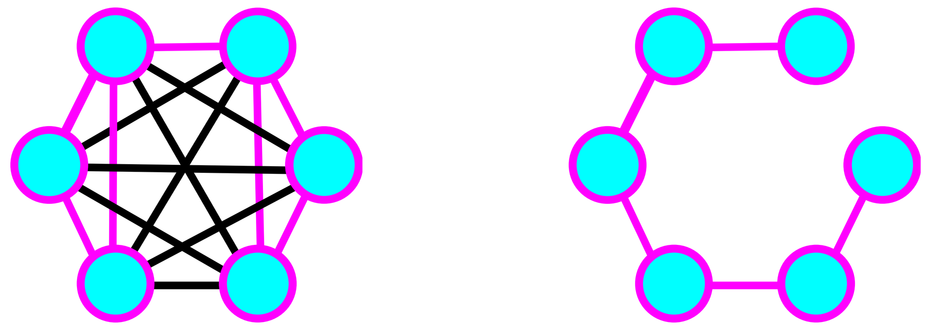
Based solely on the number of nodes we can show that a rumor **should** be spread across an entire graph after $(1+\varepsilon)(\frac{1}{\log_2(1+p)}\log_2 n + \frac{1}{p}\ln n)$ rounds with probability $p = 1 - n^{\frac{-p\varepsilon}{40}}$

Drawbacks on Demonstrated Model

While the number of iterations required for robustness can be calculated using a probability of coverage equation, we focused on optimizing for speed without loss of robustness. Our research concerned the following:

- 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?

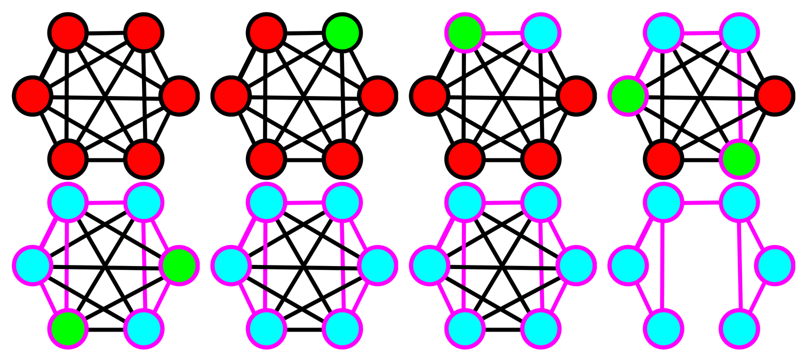
Connectedness becomes important when testing for efficiency. A complete graph will require more computation per time step than a less dense graph.



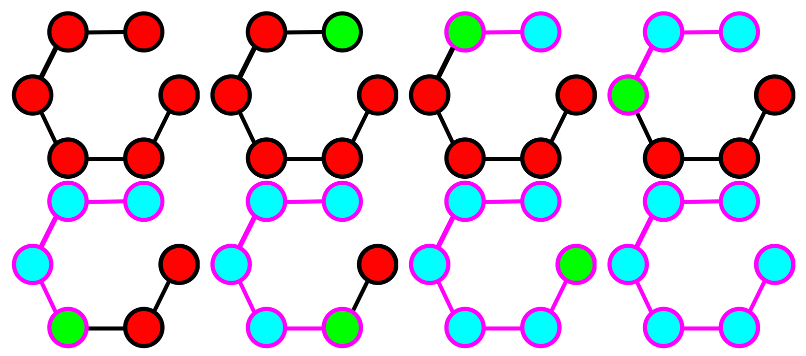
Spread on a Complete Graph vs Sparse Graph

The unused edges in the complete graph are black and stand as an obvious indicator that there is increased pathing available in a complete graph. Here is a visual representation of the "ideal" spread of information,

Complete Graph:



Sparse Graph:



Spread of a rumor across a complete and a sparse graph. Note that the complete graph finishes two rounds before the sparse graph in this example.

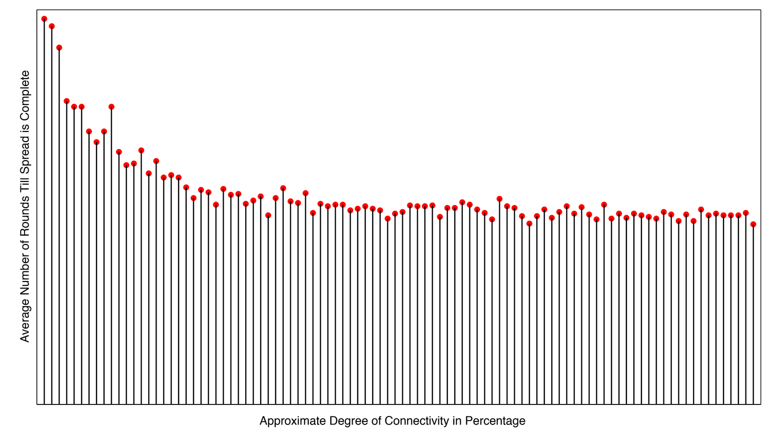
Visually, and using a graph small enough to watch all at once, it is clear that edge distribution plays a role in the spread of a information with a given protocol. Numerically this can be shown through simulation.

Results and Research

MUCH OF THE CURRENT 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 concerning stationary agents has been finished, and the our focus is now transient agents.

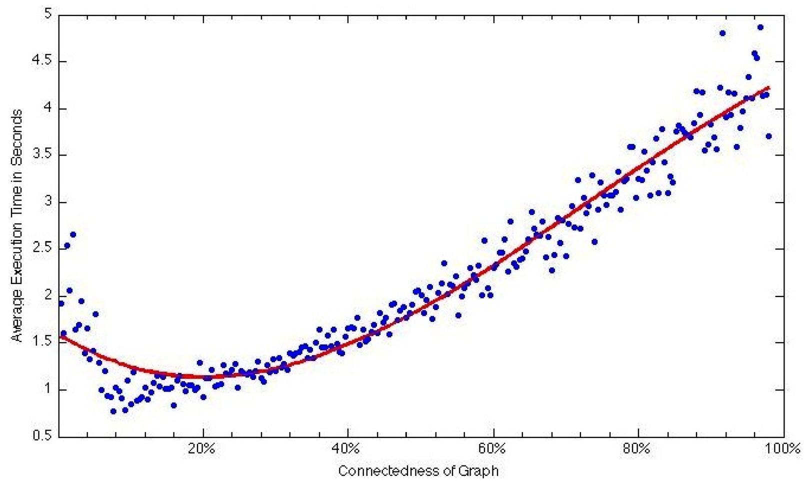
Stationary Agents

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



Number of Rounds against the Average Connectivity

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



Approximation of Complexity using Time

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

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
- 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%.

The Model

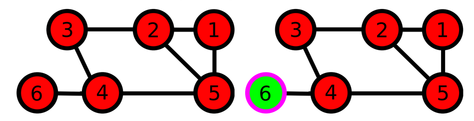
The model would be Agents sit on a graph, when they are informed start going through the protocol we're testing. We skipped right to the randomized protocol. We programmed everything with Matlab.

How it works

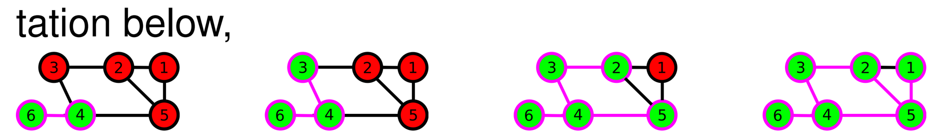
We start with an adjacency matrix that looks something like this,

$$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix} \tag{1}$$

This describes a graph that looks like this,



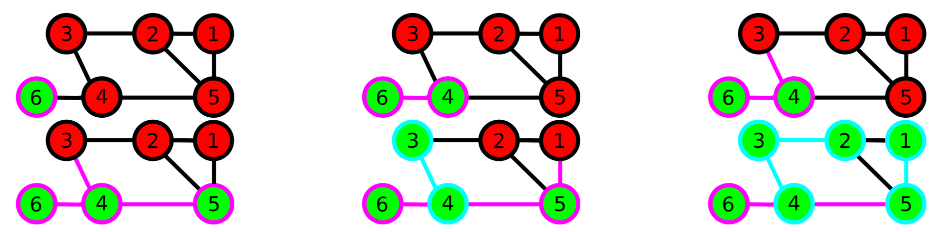
Node six only has one choice, and lets assume 0% of transmission failure, so we see the rumor spread to 4, with the newly informed matrix representation below,



.

Note the only unused edge is $2 \rightarrow 1$

Same as above with Transmission Failure



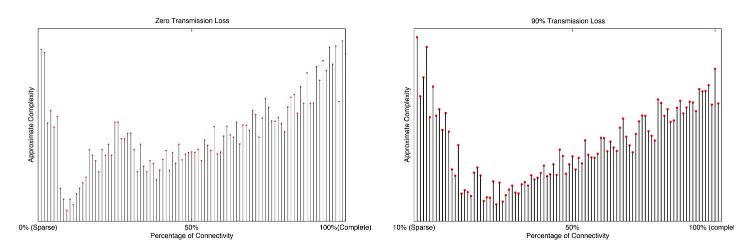
In the third image 4 fails to transmit to 3 and uses the next time step to transmit to 5. We're using blue here to demonstrate the path that involves a failure in the chain at some point. 4 successfully transmits to 3 on the 4th timestep with 5 failing to transmit to 1. It all works out in the end with the same number of over all times steps with 3 informing 2 and 5 informing 1.

For example with a transmission failure of 0% you find the optimum degree of connectivity is approximately 13% but with a transmission failure rate of 90% you get an optimum connectivity of about 30%.

Some Numerical Results

When node count, connectivity, and transmission failure are all accounted for you get a computationally optimum solution for connectivity with relation to the node count and the probability of transmission failure.

With a transmission failure of 0% you find the optimum degree of connectivity is approximately 13%, with a transmission failure rate of 90% you get an optimum connectivity of about 30%.



NAMIAM 2010, First North American Meeting on Industrial and Applied Mathematics, December 7 - 10, 2010, Universidad del Mar, Huatulco, Oaxaca Mxico