

# Network Based Modeling for the Spread of Scientific Ideas

Mayra Bermúdez, Sara Grimm & Rzgar Hosseini

ETH Zurich

December 16, 2012

## Spread of scientific ideas

Spread of scientific ideas

Introduction and Motivations

# Spread of scientific ideas



In this slide I' would like to add a figure very similar to the last one but with some labels and different color in the nodes. A draft of it is the Figure1beamer. I haven't manage to put here! Hopefully I'll do it later!

# Introduction and Motivations

- But how do ideas spread?

# Introduction and Motivations

- But how do ideas spread?
- Research has shown that not only does the nature of information or innovation influence the diffusion of it, but also that the structure of a network influences the diffusion dynamics.

# Introduction and Motivations

- But how do ideas spread?
- Research has shown that not only does the nature of information or innovation influence the diffusion of it, but also that the structure of a network influences the diffusion dynamics.



- Here, we try to simulate the spread of scientific ideas in different networks.

- Here, we try to simulate the spread of scientific ideas in different networks.
- The model presented is based on two studies: one that investigated critical parameter values for complex contagion and another that investigated critical values of a rewiring parameter.

- Here, we try to simulate the spread of scientific ideas in different networks.
- The model presented is based on two studies: one that investigated critical parameter values for complex contagion and another that investigated critical values of a rewiring parameter.

# Fundamental Questions

The main goals of this simulation study are to investigate how network structure influences the distribution of ideas, and how the distribution of ideas influences network structure.

## Some terminology

I'll do a link to this slide, in this way we can go back to this slide every time that be necessary. I'll do that the table fix here, but I'm not sure what do you think if this information should be in a table? or in other way?

Table : List of terminology

	Term
Neighborhood index	The fraction of the holders of the s who are neighbor as well averaged over
Intra-idea distance	The average distance between holders of the same idea in the n
Dominant frequency	The frequency of the dominant in the network at each time steps of th
Average dominance time	The average number of time s in which the dominant idea keeps its
Novelty index	The fraction of newly generated

During the study we vary : **Parameters**

- Probability of rewiring  $\phi$
- Rate of innovation  $\alpha$
- Complex contagion threshold  $\delta$  respectively)

**Network structures**

- Random
- Scale free
- Small world
- Caveman

**Idea distributions**

# Before simulations: our thoughts

## Effects of Network Structure on Idea Distribution

- Given a starting network and a random idea distribution, how do different network structures (see Section ) affect the distance between nodes that have the same idea (intra-idea distance)?
- How do they affect the neighbourhood index?
- How do they change the emergence of dominant ideas and their time of dominance? How do their effects depend on the values of  $\phi$ ,  $\alpha$  and  $\delta$ ?

# Effects of Idea Distribution on Network Structure

Given a starting idea distribution and a caveman network structure, how do different idea distributions affect the average path length and diameter of the network? Do they change the number of connected components in the network? Do clusters form differently, and how does the clustering coefficient change? What does the distribution of node degree looks like? How do these effects depend on the values of  $\phi$ ,  $\alpha$ , and  $\delta$ ? -random

-parallell

-nonparallell



## Description of the Model

The model used here is based on a study by Each simulation begins with a specified network structure as well as a distribution of the 'idea' (or state) of the nodes. At each time step a node either changes its idea to that of one of its neighbours' ideas if its frequency surpasses a defined threshold, rewires to connect with a node that has the same idea, or generates a novel idea (this is the innovation parameter).

Given the network structure and node states, three parameters are introduced:  $\phi$  (probability of rewiring),  $\alpha$  (probability of innovation), and  $\delta$  (complex contagion threshold). As in [citeHN2006](#),  $\phi$  is a value from zero to one, and is the probability that one of the edges of a randomly chosen node  $i$  will be changed to connect to another node  $j$  that  $i$  is unconnected with. We decided to add one more criterion to this definition: node  $j$  is a node that has the same idea as node  $i$ . This encourages the simulations to reflect a common tendency of individuals to seek out others who think like them. If there is no such node  $j$ , then the chosen node will do nothing.

At each time step a node may 'come up with a new idea' with a probability of  $\alpha$ . This value is small to reflect that novel ideas are not frequently observed.

We introduced a node threshold  $\delta$  to the general model in order to investigate the behaviour of complex contagion as opposed to simple contagion. This was motivated by a study by [citet\\*CM2007](#). Simple contagion is well suited for modeling the spread of diseases since they may often be passed on by a single contact with an infected individual. However, as our intuition may suggest, other kinds of innovations raise questions about the legitimacy and credibility of the innovations themselves, and may thus require exposure to multiple sources of the innovation. This is called complex contagion. Models usually represent it in two ways regarding the number of connected sources that must have adopted an innovation in order to influence the agent (or individual) in question: it is either a fixed number (greater than one) or a fraction (between zero and one, inclusive). Our simulation used the second formulation since the degree of individual nodes varied across network structures.

# Networks

For the purposes of our simulations, we used four network structures. These structures can be characterized by properties such as average shortest path lengths, clustering coefficients, and the degree of connectivity (see Table 1 for definitions). Below are short descriptions of each network structure. In order to compare between different structures, the network structure parameters were chosen such that the mean degree of the networks were similar (approximately 25). For further details about parameter values, see Table ??

free scale

caveman

random

smallworld

Ideas

# Implementation

Our simulation comprises mainly two parts: the phase 1 corresponding to the study of the effect of network structure on the idea distribution, and phase 2 where we study how the distribution of the ideas affect the topology of the network. These two phases are implemented in the same MATLAB file **mainscript.m**.

Implementation of both phases is divided into three steps.

**Step 1:** Network structure is chosen and we generate the adjacency matrix corresponding to that structure. The initial idea distribution is also generated.

**Step 2:** The updating process is done onto the chosen network structure.

**Step 3:** At this step a series of functions are called to get the results.

# Step 1: Generating network structures and initial idea distribution

Step 2: Update

Here it suppose to be the algorithm

## Step 3: Getting results

In the phase 1 we want to observe the influence of a certain network structure on the distribution of ideas after the updating process (step 2). For this, it is necessary to search for features that reflect the final distribution of ideas. In this study we observed the following features:

??

## Discussion

As previously mentioned, we observed that some features of the idea distribution varied with the type of network structure. We also observed that some features of the network structure varied with the kind of idea distribution implemented. Some of these results were intuitive and expected, while others were not.

Network structure mainly influenced two features of the idea distribution: the intra-idea distance and the neighbourhood index. The intra-idea distance was larger for the caveman and small world networks, as anticipated. However, contrary to expectations, the neighbourhood index was larger for these two structures, and the caveman structure did not decrease the frequency of dominance (and neither did any other structure). Additionally, the values of  $\phi$  and  $\delta$  did correlate with changes in the values of the intra-idea distance (by decreasing it) and the neighbourhood index (with different effects depending on the structure), but not always:  $\delta$  did not vary the results of the intra-idea distance or the neighbourhood index of the random and scale-free networks, and  $\phi$  did not vary the results of the intra-idea distance of the small world network, the results of the neighbourhood index of the random and scale-free networks, and the results of the frequency of dominance of the small world network.



## Summary and Outlook

To conclude our simulation study, our results support the general idea that network structure and qualities of the ideas held in them may mutually influence each other. The more rigid the structure of a network was, the more likely that the intra-idea distance and neighbourhood index of the network was larger. These two features varied more with the innovation rate  $\alpha$  for less structured networks, and varied more with the complex contagion threshold  $\delta$  for the most rigid structure - the caveman network. The average dominance time seemed to vary more with parameter values ( $\phi$ ,  $\delta$ , and  $\alpha$ ) than with the network structures, whereas the frequency of dominance of ideas did not decrease on average in the long run with any network structure.

A network in which the pattern of ideas held by the nodes are 'in accord' within the clusters of the caveman network maintained more of the structure features of a caveman network. These values varied with values of rewiring probabilities  $\phi$  and complex contagion thresholds  $\delta$ . Networks with a random pattern of ideas or with the pattern of ideas matching well within the clusters maintained

paragraphRewiring criteria Future simulations may compare the emerging features of idea distributions not just between network structures, but also between different rewiring criteria. It is not clear how much of the features of the idea distribution in this simulation study was a result of the network structure or of the 'preference' that nodes had in rewiring to like-minded nodes. Therefore, these results could be compared to (1) random rewiring or, to allow structure to play a larger role, (2) to allow random rewiring to nodes that are at most a distance of three nodes away. This is somewhat more realistic since most people connect with individuals who are somewhat in their vicinity through mutual connections.

paragraphComplex contagion and innovation Considering complex contagion, novel ideas in our model were at a disadvantage for spreading in the network. Allowing novel ideas to have a larger influence weight may be one way to grant the ability of them to spread to other nodes. Alternatively, the complex contagion threshold for novel ideas may be lowered.

paragraphRandom idea adoption Future simulations may