

# Network Based Modeling for the Spread of Scientific Ideas

Mayra Bermúdez, Sara Grimm & Rzgar Hosseini

ETH Zurich

December 17, 2012

## Spread of scientific ideas

## Spread of scientific ideas

### Motivations

Spread of scientific ideas

Motivations

Definitions needed

## Spread of scientific ideas

### Motivations

### Definitions needed

### Elements of the model

## Spread of scientific ideas

### Motivations

### Definitions needed

### Elements of the model

### Model

## Spread of scientific ideas

### Motivations

### Definitions needed

### Elements of the model

### Model

### Questions

Spread of scientific ideas

Motivations

Definitions needed

Elements of the model

Model

Questions

Simulation



Spread of scientific ideas

Motivations

Definitions needed

Elements of the model

Model

Questions

Simulation

Results

# 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

What other definitions should I put here? Paralell, nonparallel???

**Neighborhood index:** The fraction of the holders of the same idea who are neighbor as well averaged over all ideas.

**Neighborhood index:** The fraction of the holders of the same idea who are neighbor as well averaged over all ideas.

**Intra-idea distance:** The average distance between holders of the same idea in the network.

**Neighborhood index:** The fraction of the holders of the same idea who are neighbor as well averaged over all ideas.

**Intra-idea distance:** The average distance between holders of the same idea in the network.

**Dominant frequency :** The frequency of the dominant idea in the network at each time steps of the simulation.

**Neighborhood index:** The fraction of the holders of the same idea who are neighbor as well averaged over all ideas.

**Intra-idea distance:** The average distance between holders of the same idea in the network.

**Dominant frequency :** The frequency of the dominant idea in the network at each time steps of the simulation.

**Average dominance time:** The average number of time steps in which the dominant idea keeps its dominance.

**Neighborhood index:** The fraction of the holders of the same idea who are neighbor as well averaged over all ideas.

**Intra-idea distance:** The average distance between holders of the same idea in the network.

**Dominant frequency :** The frequency of the dominant idea in the network at each time steps of the simulation.

**Average dominance time:** The average number of time steps in which the dominant idea keeps its dominance.

**Novelty index:** The fraction of newly generated ideas.



**Neighborhood index:** The fraction of the holders of the same idea who are neighbor as well averaged over all ideas.

**Intra-idea distance:** The average distance between holders of the same idea in the network.

**Dominant frequency :** The frequency of the dominant idea in the network at each time steps of the simulation.

**Average dominance time:** The average number of time steps in which the dominant idea keeps its dominance.

**Novelty index:** The fraction of newly generated ideas.

**Average shortest path :** The average number of steps along the shortest paths for all possible pairs of network nodes.

**Average shortest path :** The average number of steps along the shortest paths for all possible pairs of network nodes.

**Clustering coefficient:** A measure of degree to which nodes in a graph tend to cluster together.

**Average shortest path :** The average number of steps along the shortest paths for all possible pairs of network nodes.

**Clustering coefficient:** A measure of degree to which nodes in a graph tend to cluster together.

**Degree of connectivity :** The number of edges incident to the vertex.

**Average shortest path :** The average number of steps along the shortest paths for all possible pairs of network nodes.

**Clustering coefficient:** A measure of degree to which nodes in a graph tend to cluster together.

**Degree of connectivity :** The number of edges incident to the vertex.

**Connected component:** A subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in the supergraph.

**Average shortest path :** The average number of steps along the shortest paths for all possible pairs of network nodes.

**Clustering coefficient:** A measure of degree to which nodes in a graph tend to cluster together.

**Degree of connectivity :** The number of edges incident to the vertex.

**Connected component:** A subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in the supergraph.

**Diameter of network:** The longest of all the calculated shortest paths in a network.

## During the study we vary:

### Parameters

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

## 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



## 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

- Random
- Parallel
- Nonparallel

# Parameters

$\phi$ : 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.

## Parameters

$\phi$ : 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.

$\alpha$ : Is the probability that at each time step a node may 'come up with a new idea' .

## Parameters

$\phi$ : 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.

$\alpha$ : Is the probability that at each time step a node may 'come up with a new idea' .

$\delta$ : The node threshold to the general model in order to investigate the behaviour of **complex contagion**.

## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).

## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).  
Simple contagion well suited for modeling the spread of diseases

## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).  
Simple contagion well suited for modeling the spread of diseases  
But is enough to hear about a new scientific idea for the first time  
to adopt it?

## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).  
Simple contagion well suited for modeling the spread of diseases  
But is enough to hear about a new scientific idea for the first time  
to adopt it?  
⇒ Complex contagion



## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).  
Simple contagion well suited for modeling the spread of diseases  
But is enough to hear about a new scientific idea for the first time  
to adopt it?

⇒ Complex contagion

How to implement it in the model?

## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).  
Simple contagion well suited for modeling the spread of diseases  
But is enough to hear about a new scientific idea for the first time  
to adopt it?

⇒ Complex contagion

How to implement it in the model?

- It is either a fixed number (greater than one)

## What is complex contagion?

We included this element motivated by a study by [citet\\*CM2007](#).  
Simple contagion well suited for modeling the spread of diseases  
But is enough to hear about a new scientific idea for the first time  
to adopt it?

⇒ Complex contagion

How to implement it in the model?

- It is either a fixed number (greater than one)
- **It is a fraction**

## Network structures

I'll include here some images to illustrate them

# Idea distributions

also images with cave man model

## Description of the Model

The model used here is based on a study by [citet\\*HN2006](#). Each simulation begins with a specified network structure as well as a distribution of the 'idea' (or state) of the nodes.

## Description of the Model

The model used here is based on a study by [citet\\*HN2006](#).  
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 selected at random either:

## Description of the Model

The model used here is based on a study by [citet\\*HN2006](#).

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 selected at random either:

- Changes its idea to that of one of its neighbours' ideas if its frequency surpasses a defined threshold



## Description of the Model

The model used here is based on a study by [citet\\*HN2006](#).

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 selected at random 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

## Description of the Model

The model used here is based on a study by [citet\\*HN2006](#).

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 selected at random 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).

# 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)?

# 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?

# 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 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?

# 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?



## 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?

## 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?

## 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  $\delta$ ?

## 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  $\delta$ ?

should we mention something about the initial idea distribution in phase 2?? (parallel, nonparallel, random)

# Implementation

**Phase 1:** Corresponding to the study of the effect of network structure on the idea distribution.

# Implementation

**Phase 1:** Corresponding to the study of the effect of network structure on the idea distribution.

**Phase 2:** Where we study how the distribution of the ideas affect the topology of the network.

# Implementation

**Phase 1:** Corresponding to the study of the effect of network structure on the idea distribution.

**Phase 2:** Where we study how the distribution of the ideas affect the topology of the network.

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



**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 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.

We have too many results and I don't think that we have time enough to mention everything, what do you think? In case you think the same, which ones do you think are the most interesting?

## Summary and Outlook

### Conclusions

- Our results support the general idea that network structure and qualities of the ideas held in them may mutually influence each other.

# Summary and Outlook

## Conclusions

- 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.

## Summary and Outlook

### Conclusions

- 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.
- 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.

# Summary and Outlook

## Conclusions

- 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.
- 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.
- Networks with a random pattern of ideas or with a pattern with more 'disaccord' within the clusters resulted in a structure that began to resemble a random graph more than a caveman network

# Summary and Outlook

## Conclusions

- 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.
- 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.
- Networks with a random pattern of ideas or with a pattern with more 'disaccord' within the clusters resulted in a structure that began to resemble a random graph more than a caveman network



Thus, in addition to the structure of networks and the pattern of ideas in them, complex contagion thresholds, innovation rates, and the probability of creating new connections (while reflecting 'preferences' of being connected with like-minded others) tend to influence features of the network.

## Further ??

Several extensions to the proposed model could be investigated to better understand the relationships between network structure and idea distribution.

## Further ??

Several extensions to the proposed model could be investigated to better understand the relationships between network structure and idea distribution.

- Rewiring criteria

## Further ??

Several extensions to the proposed model could be investigated to better understand the relationships between network structure and idea distribution.

- Rewiring criteria
- Complex contagion and innovation

## Further ??

Several extensions to the proposed model could be investigated to better understand the relationships between network structure and idea distribution.

- Rewiring criteria
- Complex contagion and innovation
- Random idea adoption