Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Lecture with Computer Exercises:
Modelling and Simulating Social Systems with MATLAB

Project Report

Network-Based Modeling for the Spread of Scientific Ideas ...

Mayra Bermúdez, Sarah Grimm & Rzgar Hosseini

Zurich
December 14th, 2012

Agreement for free-download

We hereby agree to make our source code for this project freely available for download from the web pages of the SOMS chair. Furthermore, we assure that all source code is written by ourselves and is not violating any copyright restrictions.

Mayra Bermúdez Sarah Grimm Rzgar Hosseini

Contents

1 Abstract

2 Individual contributions

This report represents a group effort by all members.

3 Introduction and Motivation

We live in a time in which aspects of our lives and of our world become more and more connected with each other. To understand one aspect, we must understand many other aspects, which all together form a large, complex system, or network. Globalization has changed the meaning of 'distance' and communication, allowing seemingly unrelated and unconnected individuals to share more than they ever could before. Fields of studies are overlapping with each other, creating new interdisciplinary domains and building a diverse playground for the sharing of ideas. But how do ideas spread? This question is especially interesting with the increase of technology that allows us to record and visualize the networks that connect individuals and ideas, and in particular the ability to 'see' how they change. This can lead to insight about why and when ideas spread and into the complexity of the matter. 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. The model presented is based on two studies: one that investigated critical parameter values for complex contagion (Centola and Macy, 2007) and another that investigated critical values of a rewiring parameter (Holme and Newman, 2006).

3.1 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. For a list of the terminology that will be used throughout the paper, please refer to ****TABLE DEFINITION OF TERMS*****. By varying three parameters - probability of rewiring, rate of innovation, and complex contagion threshold (ϕ, α, and δ respectively) - and using different network structures and idea distributions, we observed how network structure characteristics changed. Similarly, we observed how the distribution of ideas and the connections between them changed. Below we describe our questions more specifically.

3.1.1 Effects of Network Structure on Idea Distribution

Given a starting network and a random idea distribution, how do different network structures 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 ϕ, α and δ?

More rigid network structures (those with less randomness, such as the caveman and the small world networks) may make it more difficult for ‘like-minded’ nodes to connect and may thus have smaller neighbourhood indexes and larger intra-idea distances than the more random network structures (such as the random and scale-free networks). Their effects may be more sensitive to the values of ϕ (because this affects how likely it is for their structure to change) and to values of δ because being restricted to a more closed group of nodes makes it difficult to reach a threshold necessary to become similar to surrounding nodes. Values of α may decrease the neighbourhood indexes by creating larger diversity among neighbouring nodes.

If more rigid network structures do make it more difficult for like-minded nodes (that is, nodes with the same idea) to connect, then it would be more difficult for a dominant idea to emerge in these networks. These effects may be smaller for larger values of ϕ since these values would allow for the structure to change more. For larger values of ϕ therefore one could expect that the effects of the network structures on the characteristics of the idea distribution are more similar since allowing to change the structure removes their initial influence.

3.1.2 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 look like? How do these effects depend on the values of ϕ, α, and δ?

If the starting idea distribution is parallel to the caveman network structure (see ******TABLE DEFINITIONS OF TERMS*** for definitions), then like-minded nodes will already be connected and thus rewiring will probably not change much of the average path length, nor will it change the diameter. Similarly, the clustering coefficient will probably remain high just like the starting value. The distribution of the node degree will also not change (nodes will have one of two values for their degree). In other words, if the idea distribution is parallel to the network structure, the structure will not change much. Changing ϕ and δ will not change these effects, and perhaps increasing α will decrease the clustering coefficient and will increase the number of connected components because nodes will disconnect from nodes with novel ideas and will rewire to nodes with the same idea.

If the starting idea distribution is random, then the network's caves will disintegrate as nodes will rewire with other nodes outside of their caves. This will change the degree distribution by increasing its variance (nodes will have a variety of different degree values). Depending on the value of δ this disintegration may be reduced because nodes have a higher chance of forming dominant ideas within caves. Similarly, increasing ϕ will increase the disintegration of caves. Thus for this idea distribution the parameter values may play a larger role.

If the starting idea distribution is anti-parallel, nodes within each cave will initially be connected with nodes that do not hold the same idea as them. Therefore the threshold δ will not likely be met in order for nodes to change their ideas, and they will rewire with other nodes outside of their cave. The clustering coefficient and the number of connected components will likely decrease, and the diameter and the average path length will decrease as well since the structure will change significantly. The degree distribution will increase in variance. Increasing ϕ and α and decreasing δ will probably increase the magnitude of these effects. Thus, having an anti-parallel idea distribution will probably cause the most changes in the characterstics of the network structure that are in question.

4 Description of the Model

The model used here is based on a study by Holme and Newman (2006). 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: ϕ (probability of rewiring), α (probability of innovation), and δ (contagion threshold). As in Holme and Newman (2006), ϕ 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.

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

We introduced a node threshold δ 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 Centola and Macy (2007). 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, and as studies have shown, the spread of other kinds of innovations require several exposures before they are adopted by individuals; these situations refer to complex contagion.

4.1 Networks

For the purposes of our simulations, we used four network structures. *******NETWORK STRUCTURE FIGURE****** illustrates them. These structures can be characterized by properties such as average shortest path lengths, clustering coefficients, and the degree of connectivity (see ********TABLE DEFINITIONS OF TERMS****** 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 30). For further details about parameter values, see *********TABLE WITH PARAMETERS ******.

4.1.1 Random Graph

A variant of the Erdős-Rényi random graph model (Erdős and Rényi, 1960), and implemented by Brugger and Schwirzer (2011). Random graphs have a short average path length. The graph is defined by the total number of nodes, and by the probability of any two nodes to be connected. Thus all connections are random. They typically have a small clustering coefficient.

4.1.2 Caveman Graph

As defined by Watts (2003). The caveman structure is defined as having k isolated and fully connected 'cliques' from which one link is changed to connect one clique to another, rendering all cliques to be connected. Thus, relative to random graphs, they have a high clustering coefficient and a large average shortest path length.

4.1.3 Small World Graph

Defined by Watts and Strogatz (1998), and implemented by Brugger and Schwirzer (2011). Small world graphs have characteristics that lie in between random graphs and highly clustered graphs (such as caveman graphs): they have a high clustering coefficient similar to the latter, but also have a small average shortest path similar to the former. Many real-world networks have been observed to have a small world structure, and thus we included it in our simulations.

4.1.4 Scale-Free Graph

Defined by Barabási and Albert (1999), and implemented by Brugger and Schwirzer (2011). Scale-free network structures are often found where new nodes are constantly being added, and they are connected to already well-connected nodes. Such a structure displays a scale-free power-law distribution of the degree (connectivity) of nodes. Thus, there are few nodes that are highly connected, and more nodes that are moderately or mildly connected. Compared to random graphs, they have a smaller average shortest path.

4.2 Ideas

After choosing a starting structure for our model, we then chose a distribution for the starting ideas (states) of nodes. Each node was randomly assigned one of these ideas, thus allowing for multiple nodes to have the same idea. For the caveman structure, however, there were two other options: to either distribute the starting ideas 'parallel' to the structure, i.e. such that all nodes in a cave shared the same idea, or 'antiparallel' such that all nodes in a cave had a different idea. This was used for the analysis of the effect of the idea distribution on the network structure. ******FIGURE OF IDEA DISTRIBUTION****** illustrates these idea distributions. Why was the caveman structure investigated? There were two reasons: firstly, it was straightforward how to define idea distributions that are in accord or disaccord with the connections of nodes in the network. Secondly, in reality research teams are often made up of closely-connected members that are only weakly connected to other research teams, and within these teams, members may or may not be interested in the same ideas for research.

As previously mentioned, each node had a small probability α of adopting a novel idea from a virtually unlimited number of new ideas.

5 Implementation

6 Simulation Results and Discussion

6.1 Effects of Network Structure on Idea Distribution

For this analysis, ideas were assigned randomly onto the nodes of four different network structures (caveman, small world, random, and scale-free). Simulations were run for each network structure for 27 different parameter combinations (three values for each of ϕ, δ and α). The effects of each network structure on the idea distribution was evaluated on five features: the intra-idea distance, the neighbourhood index, the frequency of dominance of an idea, the average dominance time for an idea, and the novelty index. We also investigated how these effects vary with the three parameters (ϕ, δ, α).

We observed that the caveman and small world network structures had a similar influence on the intra-idea distance and neighbourhood index, and the random and scale-free network structures had a different - yet similar to each other - influence on these features.

6.1.1 Intra-Idea Distance and Neighbourhood Index

For any parameter combination, the caveman and small world structures resulted in larger intra-idea distances (respectively) than those of the random and scale-free structures, which were very similar to each other (***FIGURE 1.1****). Additionally, a similar influence was found on the neighbourhood index: the caveman structure held the largest index regardless of parameter combination, followed by the small world, random, and scale-free structures (****FIGURE 1.2****). It seems that the caveman structure encourages nodes to be within the direct neighbourhood of like-minded nodes (nodes with the same idea) and at a farther distance from like-minded nodes that are not in their direct neighbourhood, whereas the random and scale-free structures have a tendency to keep like-minded nodes in each other's direct neighbourhood but to also keep those like-minded nodes not in their direct neighbourhood at a shorter distance. This could be a result of the general larger average path distance that caveman networks have when compared to random and scale-free networks.

Dependence on α For all network structures, increasing the level of innovation α decreased the intra-idea distance. This effect was more pronounced in the scale-free and random networks (***FIGURE 1.5-1.8****). Does innovation bring people together in scienitific communities? Perhaps it does not: under complex contagion, nodes that adopted a novel idea could not influence other nodes since they were the sole holders of these ideas. Therefore, their intra-idea distance for these 'novel nodes' would have been zero.

Higher values of α also increased the neighbourhood index for all network structures. This correlation was again most visible for the scale-free network, followed by the random network (***FIGURE 1.9-1.12***). One possible explanation for this correlation is that the more novel ideas nodes create, the less likely it is that the contagion threshold is met for other ideas, and thus nodes will only be rewiring instead of also changing their ideas. Thus more like-minded nodes will be connected, and the index increases.

Dependence on ϕ By increasing ϕ, the intra-idea distance decreased for all network structures. This correlation is quite an intuitive result since ϕ is the probability of deleting a connection between two nodes with different ideas and the formation of a new connection between two nodes with the same idea, and thus, by increasing ϕ the distance between like-minded nodes decreases. Unlike the α parameter, the effects of ϕ are more pronounced for the caveman and small world structures rather than for the random and scale-free networks (*****FIGURES 1.13-1.16****).

On the other hand, th effects of ϕ on the neighborhood index varied between networks. Increasing ϕ increased the neighbourhood indexes of the random and scale-free networks, while it decreased the neighbourhood index of the caveman network and had no correlation with changes in the small world network (***FIGURES 1.17-1.20***).

Dependence on δ Increasing values of the complex contagion threshold δ decreased the intra-idea distance for all network structures (****FIGURE 1.21-1.24***). Using the δ value of 0.05 resulted in slightly lower intra-idea distances than the two smaller values (which behaved very similarly) for the caveman and small world network structures. WHY?

The neighbourhood index for the caveman network increased as δ increased (****FIGURE 1.25***). This is intuitive since requiring more like-minded nodes to be in the direct neighbourhood for a node to adopt their idea automatically increases the number of like-minded nodes in the neighbourhood (neighbourhood index). However, increasing δ could have had the opposite effect: if the threshold was too high, nodes would not have adopted their neighbhours' ideas and thus the neighbourhood index would have remained small. No correlation was found between the neighbourhood index of the small world, scale-free, and random network structures and values of δ. WHY only for caveman?

6.1.2 Frequency of Dominance

This feature is interesting if one considers scientific society. How dominant are dominant ideas in the scientific community? Does the structure of this community influence this dominance? For all parameter combinations and for all network structures in our simulations, the frequency of dominance of ideas increased with time. This may suggest that none of these structures impede the adoption of new ideas, and that, not surprisingly, society tends to adopt dominant ideas. More surprisingly, however, the increase in dominance frequency progressed more quickly for the caveman network structure (***FIGURE 1.3***). Could it be that this structure encourages nodes to adopt dominant ideas more easily? This may not be generalizable because the results are quite sensitive to parameter values due to the stochasticity of the simulations. For example, ****FIGURE 1.4*** shows a different combination of parameters, and here the faster increase in the dominance is not observed for the caveman structure.

6.1.3 Average Dominance Time and Novelty Index

Because the results of the average dominance time and the novelty index were too dependent on the parameter values, we did not include their results. The network structure does not seem to play a strong enough role over all parameter combinations in influencing these features. It is not surprising, however, that the novelty index was highly dependent on the value of α, but it is not so intuitive why the average dominance time was not correlated with α at all: increasing the number of novel ideas decreases the possible number of 'followers' for already-established ideas. Perhaps the value of α would need to be increased to observe this behaviour.

6.2 Effects of Idea Distribution on Network Structure

Here we investigated the opposite direction of effects: how changing the idea distribution affects the resulting network structure. To do this we applied three different idea distributions (random, parallel, and antiparallel to the structure) to a caveman network. There were again 27 different parameter combinations for each idea distribution, as in the previous analysis. The five features of the network structure that we evaluated were: the clustering coefficient, the degree distribution of the nodes, the number of connected components, the average path length, and the network diameter.

We observed that the influence of applying a parallel idea distribution on the features of the network structure was quite different to the influence of the random and antiparallel idea distributions. While the networks always remained fully connected (**FIGURE 2.4****), the remaining features changed. The parameter α did not change these effects for any of the idea distributions (see ***FIGURES 2.6-2.14***). Given these simulation results, perhaps a caveman-structured scientific community would also manage certain levels of innovativity without changing its fundamental structure.

6.2.1 Clustering Coefficient

The clustering coefficient of resulting networks varied with the type of idea distribution. When nodes in the same cave shared the same idea (parallel distribution), the clustering coefficient was larger (***FIGURE 2.3***). Additionally, the clustering coefficients for both the random and the anti-parallel idea distributions decreased in a similar manner regardless of the parameter combinations. Both of these results are intuitive since less rewiring would have taken place for the parallel distribution case, and the high clustering coefficient of the caveman structure would have been conserved, whereas more rewiring would have ocurred for the two other distributions, thus decreasing the clustering coefficients.

Dependence on Parameters Increasing values of the rewiring parameter ϕ decreased the clustering coefficient for all starting distributions, especially for the random and antiparallel distributions (***FIGURES 2.21-2.23***). This again is intuitive since ϕ increases the chances of changing connections in a highly clustered network. Increasing values of δ, however, only increased the clustering coefficient when using a parallel idea distribution, and only slightly (***FIGURES 2.30-2.32***). WHY?

6.2.2 Average Path Length

The average path length of resulting networks was larger when a parallel idea distribution was used (***FIGURE 2.1****). This is not surprising, since the structure of the caveman network was more conserved (because most nodes were already connected to like-minded nodes and did not need to rewire), and its average path length is larger than that of more randomized networks, such as the ones resulting from a larger amount of rewiring.

Dependence on Parameters Increasing ϕ, regardless of the idea distribution, decreased the average path length. This effect was more prononced for the random and anti-parallel idea distribution cases (**FIGURE 2.15-2.17***). This is another intuitive result since more rewiring naturally disturbs the rigid structure of a caveman network, thereby decreasing its large average path length; rewiring also occurs less frequently in the parallel idea distribution case because most nodes are already connected to like-minded nodes. Changing values of δ (which requires more than one neighbouring node to have the same idea in order to influence the chosen node) did not change the effects of idea distributions on the average path length (***FIGURES 2.24-2.26***). This is natural, since nodes were either already connected to a significant number of nodes with the same idea (in the case of the parallel distribution), not connected to any other node with the same idea (the antiparallel case), or connected to nodes with random assignment of ideas. Thus the threshold δ would either already be fulfilled from the start, would definitely not be fulfilled, or would have a very small chance of being fulfilled, respectively.

6.2.3 Network Diameter

Similar to the clustering coefficient and the average path length, the network diameter was larger when the parallel idea distribution was applied (***FIGURE 2.2*****). This distribution encouraged the structure of the caveman network to remain mostly unchanged, and thus the farthest distance between two nodes was larger than for the case of random or anti-parallel distributions, where more rewiring occured.

Dependence on Parameters As ϕ increased, the network diameter decreased for all idea distributions, and slightly less for the parallel distribution (**FIGURES 2.18-2.20*****). Rewiring a caveman network intuitively may decrease the network diameter by connecting more of its caves. Similar to the average path length, values of δ did not change the behaviour of the network diameter given the idea distributions (***FIGURES 2.27-2.29***).

6.2.4 Degree Distribution

Random graphs have a somewhat normally distributed degree distribution. Scale-free graphs, on the other hand, have a degree distribution that follows a scale-free power-law. We observed that the degree distribution of the networks depended on the idea distribution (***FIGURE 2.5****). A parallel idea distribution resulted in a degree distribution similar to that of a scale-free graph, which is not surprising. Caveman graphs have two or three different degrees for their nodes, and allowing for some rewiring would 'smooth' out this discrete distribution. Similar to previous results, the random and antiparallel idea distributions behaved similarly: their degree distribution was similar to that of random graphs. It is interesting to see such a visible difference in these distributions over relatively few time steps (1000 steps), regardless of the parameter combinations.

6.3 Discussion

Expected results vs actual results. network structure affects the ditirbution of ideas –¿ distribution of scientific idea depends on structure of scientific community delta effect with caveman on index random and antiparallel more dependent on delta distribution changed a lot, but the others didn'ttoo much.

Future Study intra-idea distance = dont' include ideas that are held by only one node.

7 Summary and Outlook

References

Barabási, A. and R. Albert (1999). Emergence of scaling in random networks. *science 286*(5439), 509–512.

Brugger, S. and C. Schwirzer (2011). Opinion formation by "employed agents" in social networks. Project work done in the course Lecture with Computer Exercises: Modelling and Simulating Social Systems with MATLAB, May 2011.

Centola, D. and M. Macy (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology 113*(3), 702–734.

Erdős, P. and A. Rényi (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl 5*, 17–61.

Holme, P. and M. Newman (2006). Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E 74*(5), 056108.

Watts, D. (2003). *Small worlds: the dynamics of networks between order and randomness*. Princeton university press.

Watts, D. J. and S. H. Strogatz (1998, June). Collective dynamics of 'small-world' networks. *Nature 393*(6684), 440–442.