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, and therefore they tend to become more complex. 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. The model presented in this simulation study 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 here (and its definitions), 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 described 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 indices 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 indices 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

four structures, random idea distribution, three ϕ three δ and three α −¿ 27 combinations for each structure. five features

neighbourhood index larger for caveman and small world smaller for random and scale-free intra-idea distance larger for caveman and small world random and scale-free are the same and lower dominant frequency for all structures and for all parameter combinations the frequency increased −¿ adoption of dominant idea. more rapid increase for caveman. but could be stochastic and dependent on parameter combination average dominance time too dependent on parameter values not dependent on alpha. novelty index too dependent on parameter values dependent on alpha obviously.

alpha for all structures: increasing it will decrease the intra-idea distance. !!! Could be because of more single unique ideas that have no 'distance'. more visible effect for random network and scale-free for all structures: increasing it decreased neighbourhood index. more visible effect for random and scale-free. phi for all structures: increasing it will decrease the intra-idea distance. more visible effect for caveman and small world. for neighbourhood index: in random and scale-free increasing phi increased it, and in caveman increasing phi decreased it, and for small world no effect on index. delta for all structures: increasing it decreased intra-idea distance. look at plot: certain spikes - combinations for neigbourhood index: in caveman increasing delta increases index. for other structures no connection between delta and index. obvious why it would, but not clear why only for caveman

caveman and small world keep nodes with same idea close, but those wiht same idea that are not in direct neighbourhood are far (as compared with random and scale-free).

6.2 Effects of Idea Distribution on Network Structure

three idea distributions on caveman network, three ϕ three δ and three α −¿ 27 combinations. examining five features of structure

clustering coefficient higher for parallel degree distribution for parallel ended up similar to scale-free graph distribution for random and antiparallel the distribution was similar to that of a random graph connected components all the same: 1. always all reachable average path length larger for parallel diameter larger for parallel

alpha no effect phi same effect for all distributions: increasing it will decrease path length, diameter, clustering coefficient. less effect for parallel delta for parallel only: increasing it increased the clustering coefficient.

Parallel is different to random and antiparallel: - diameter, average shortest path and clustering coefficient are HIGHER than other two distributions'.

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.