



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

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

Project Report

Spread of news in a random network of agents

Blum Josef, Kazakov Dmitry,
Machacek David, Martin Kevin

Zurich
Dec 2017

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 or cited correctly and is not violating any copyright restrictions.

Blum Josef

Kazakov Dmitry

Machacek David

Martin Kevin

Contents

1	Abstract	4
2	Individual contributions	4
3	Introduction and Motivations	4
4	Description of the Model and Implementation	5
4.1	General	5
4.2	Network	5
4.3	Simulation	7
5	Simulation Results and Discussion	8
5.1	Dataset and Implementation	9
5.2	Analyzing the Friendship Network	9
5.3	Influencing Different Initial Groups	9
5.4	Analyzing the Opinion Data	11
6	Summary and Outlook	15
7	References	15

1 Abstract

The spread of information has changed drastically during the digital age. There are new ways and opportunities to precisely place advertisement, propaganda, information, news and even fake news. There might be a different motive in publishing advertisements, news or fake news, but they all share a common characteristic: They are the most effective if people share new information with their family and friends, thereby increasing the amount of people that are in contact with this new information. This work focuses on the question: How can information be spread most effective in a random network of agents with a limited budget (eg. limited number of agents, that can be influenced)?

For this reason, depending on their degree of connectivity, different groups of agents are influenced and the proliferation speed of the information is measured and compared.

2 Individual contributions

The whole group confirms that the work was shared about equally. The following listed individual contribution only state the main field of work and is not exhaustive.

Blum Josef: Implementation of own and standard network & report

Kazakov Dmitry: initializing of agents & report

Machacek David: Update of opinions, bug fixes & report

Martin Kevin: Create and interpret sensible plots & report

3 Introduction and Motivations

The massive flow of information experienced by an individual in the modern society not only limits the ability to follow, but also enhances the vulnerability to be persuaded. In fact a recent study has shown that low-quality information, such as, for instance, falsified news, are able gain far-reaching proliferation in an online social network due to information overload (1). Despite the fact that individual agents prefer quality information, they have behavioural limitations in managing a heavy information flow, which leads to the false news gaining spread and elevated attention. In such conditions formation of an opinion of each agent is strongly influenced by that of his peers. One of the key factors determining the pace of information spread

is the degree of connectivity of individuals subject to initial information placement. Here we study in a stylised model of a social network the dependence of the rate of news spread on the degree of social significance of individuals that are first targeted by the information source. We find that the rate of information spread is strongly influenced by the social activity of individual agents first targeted by the news sources.

4 Description of the Model and Implementation

4.1 General

In order to answer the question How fast can information be spread in a random system of agents by targeting different groups of agents, a numerical simulation was performed. Since an opinion is always bound to an individual, and in our simplified case only changes due to interaction with other individuals, an agent based model was chosen. Each agent has five important parameters:

1. Number of close friends of that agent
2. Number of social network friends of that agent
3. Status of influence (is this agent being influenced or not)
4. Credulity of this agent (continuous number $[0, 1]$)
5. Opinion of this agent (continuous number $[0, -1]$)

It was made sure, that every agent at least has 1 friend (close friend or social network friend) and every agents initial opinion is set to zero (no opinion) at the beginning. The opinion of the influencers (information spreaders) is -1.

Furthermore, the parameter credulity is a measure of how easily this agent adopts to opinions of his friends respectively how stubborn he is. If this value is 0 this means he only believes in himself, if its 1 this agent believes everything he hears from his friends independent of his own opinion.

The mathematical model consists of two major parts. The first one being the network generation and the second one being the actual simulation and opinion formation / change.

4.2 Network

o study opinion formation based on the opinion of social contacts, it is inevitable to create a network. Since the simulation contains two different types of friendships,

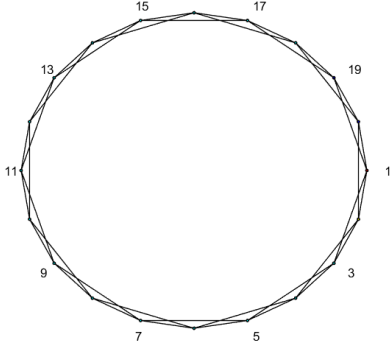


Figure 1: Representation of Strogatz network with $p=0$. Every line represents a connection between the two linked nodes. A regular grid is formed.

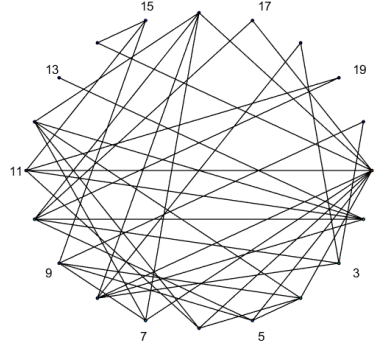


Figure 2: Representation of Strogatz network with $p=0.9$. Every line represents a connection between the two linked nodes. A random grid is formed.

namely close friends or real life friends and friends in social networks (e.g. Facebook[®]), the network generation algorithm was chosen to be as flexible as possible. This is crucial since close friends are likely to meet at regular basis. This might suggest a network in which the agents connectivity depends on a spatial aspect in the network (eg. connection to his neighbourhood). However in the case of social networks friends can be spread over large regions. Therefore a random network seems to be more suitable to simulate both close friends as well a social system friendships.

The code is built by adapting an existing script published by Pablo Blinder (2). The formation of the network is done as described by Duncan J. Watts & Steven H. Strogatz in 1998 (3). Starting with a regular network each connection has a certain probability of being split up and being rewired to a new node. With increasing probability the network becomes increasingly random (see fig. 1, 2).

The input parameters are the network size, the probability of rewiring a contact and the mean number of friends per agent. After rewiring certain connections the agents will have a chance to not build any friendships. This issue is resolved by forming a connection to a randomly chosen node for all agents that do not have at least one friend. This step is required to not change the outcome of the later simulation process in an unrealistic way. Also the existing code had to be adapted such that a friendship cannot be formed between the agent and himself.

The information of the friendship distributions in the network are stored in upper triangular matrices. For simplicity the two upper triangular matrices of the real-life and Facebook[®] networks are merged for the simulation (see fig. 3).

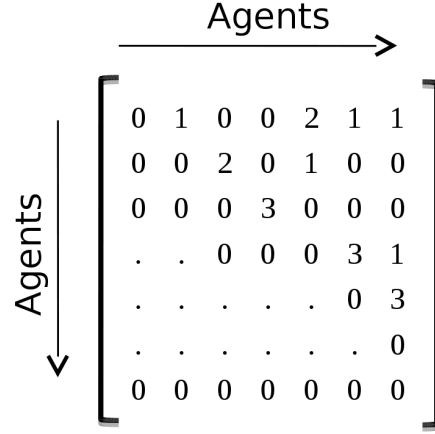


Figure 3: Upper triangular matrix for agents network. The values 1 depict a social network friendship whereas 2 denotes a close friendship, 3 accordingly means both.

4.3 Simulation

During the investigation on how to influence a social system in the most efficient way the network has to be fixed, such that the randomness of the network generation will not affect the outcome.

After that is made sure a certain number of agents will be influenced. According to the lead question of this work, a limited budget is available. This means only a limited number of agents can be influenced. Throughout the simulation this target group cannot be changed anymore. The number of agents to be influenced was chosen to be 3%, 8% or 10% of the total number of agents available in the system.

The target group was defined by the degree of connectivity of the agents (number of friendships).

The three groups were:

- The agents with the highest connectivity
- The agents with the lowest connectivity
- The agents with exactly the mean number of friendships of the underlying system

More weight was given to close friendships than to social system friendships by saying 1 close friendship is equivalent to 2 social system friendships.

To make a statement about the effectiveness of the information spread the number of iterations necessary to reach a certain overall opinion threshold (eg. -0.9) was determined.

To compare the speed of reaching this threshold for different simulation parameters, a benchmark time is necessary. It is found when simulating the opinion development a 1000 times, while choosing random agents to be influenced. For every iteration the number of simulations, that is necessary for an averaged opinion of all agents to reach the value -0.9, is chosen.

This results in 1000 simulation number, that are averaged to get the benchmark time.

The simulation loops through all agents and adapts their opinion, based on the opinions of their friends. Normally this is done by averaging the opinion of all of their friends with different weighting for real-life and social network friends. The new opinion of the current agent is then determined by changing the old opinion up to a certain percentage, that is based on the gullibility of the current agent, with the described average opinion of their friends.

$$x_{new} = (1 - g) * x_{current} + g * (k_{FB} * x_{FB} + k_{CF} * x_{CF}) / (k_{FB} + n_{FB} + k_{CF} * n_{CF})$$

x_{new} and $x_{current}$ stand for the agents new respectively current opinion. g is the gullibility of the agent and the constants n_{FB} , n_{CF} are the number of social media friends respectively close friends that agent has. k_{FB} and k_{CF} are constants that put more weight to close friends than social media friends.

The new opinion is then stored in a temporal variable in order to allow the other agents to form their opinion based on the old opinion of the all other agents.

If an agent is influenced however the opinion formation formula has to be adapted. These agents do not change their own opinion (it is fixed at -1).

After the new opinions of all agents are formed, the old opinions of the agents will be updated at the same time and the loop continues.

To find what influencing model is most efficient, one can look the rate of opinion changing. The longer it takes for the agents to reach the threshold opinion, which is close to that of the influencer, the less efficient the influencing was.

5 Simulation Results and Discussion

In this section, we present our experimental evaluation on synthetic data. Our results on synthetically generated data show that the main factor for spreading information is the amount of people that are initially convinced. We suppose it is not cost efficient to search the top-connected agents in a new random network. At some

point it becomes more economical to invest in randomly chosen agents than to invest in research to find the best-connected agents. Our results show that the time needed to convince 90% by choosing the best-connected agents compared with choosing random agents is around 50% slower if we go for the randomly chosen agents. A much better approach for reducing the time to convince 90% is by increasing the number of initial agents that are convinced by the given information.

5.1 Dataset and Implementation

The network we implemented is based on the Strogatz model which is used for generating random graphs. In our case, we need a network representing the social life of a human being. To keep it simple we reduce the social interaction into two groups: close friends and social media friends. For simplicity reasons social media friends will be referred to as Facebook[®] friends.

A random graph represents our network. Truly generated graphs should have a Gaussian distribution of number of Facebook[®] friends as well as number of close friends. Plotting both distributions should confirm that the network was randomly generated.

5.2 Analyzing the Friendship Network

As can be seen in fig. 4, the randomly generated network has a Gaussian distribution with an expected value of approximately 48 Facebook[®] friends per agent.

As fig. 5 shows, the randomly generated network has a Gaussian distribution with an expected value of approximately 9 close friends per agent.

Important: for all calculations the same Network was used!

5.3 Influencing Different Initial Groups

We compare the three different outcomes. For each, a group of agents that were influenced in the beginning and have set their opinion to -1 have to be chosen. We compare the change of the average opinion for the following three groups: Group 1, choosing the best connected Agents in the Network as the initially influenced persons, Group 2, choosing random Agents, and Group 3, choosing the Agents with the least friends. As expected, it is most efficient when initially influencing the best connected agents. We want to investigate how cost efficient it is to search for the best connected agents in a network, compared with investing in more initial influenced agents that are chosen randomly. As point of reference, we also calculated the worst case scenario, meaning, having chosen the least connected agents.

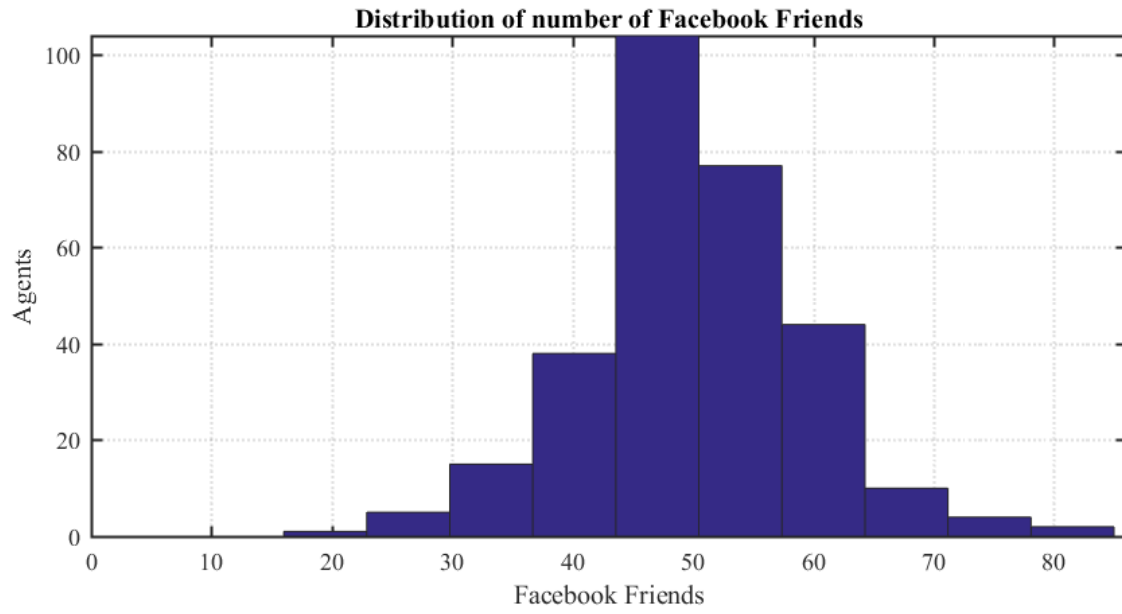


Figure 4: The distribution of Facebook[®] friends follows a Gaussian distribution.

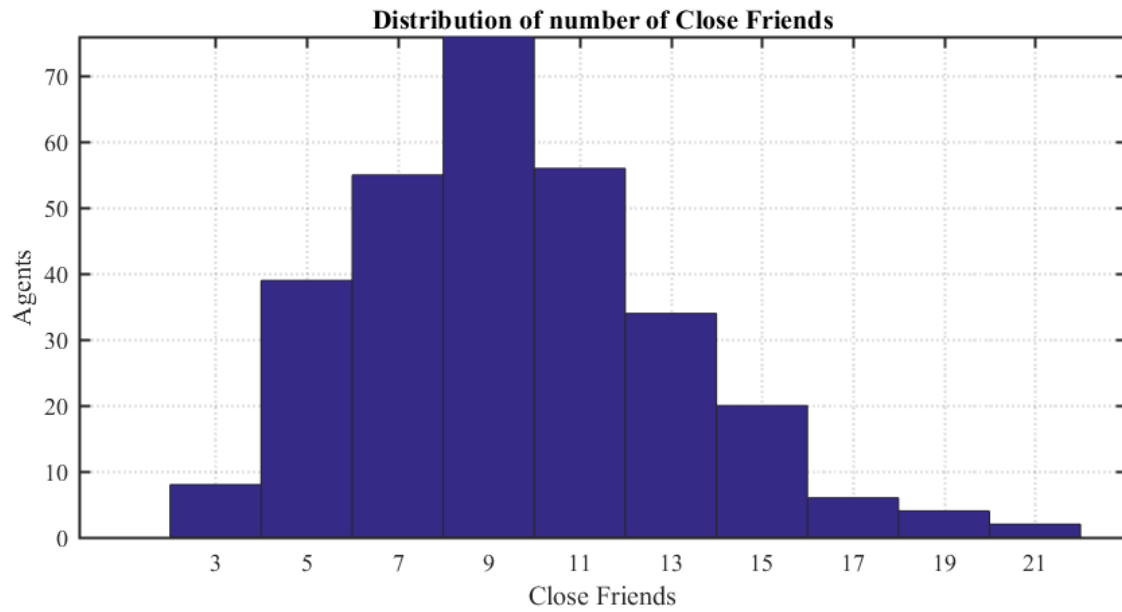


Figure 5: The distribution of close friends also follows a Gaussian distribution.

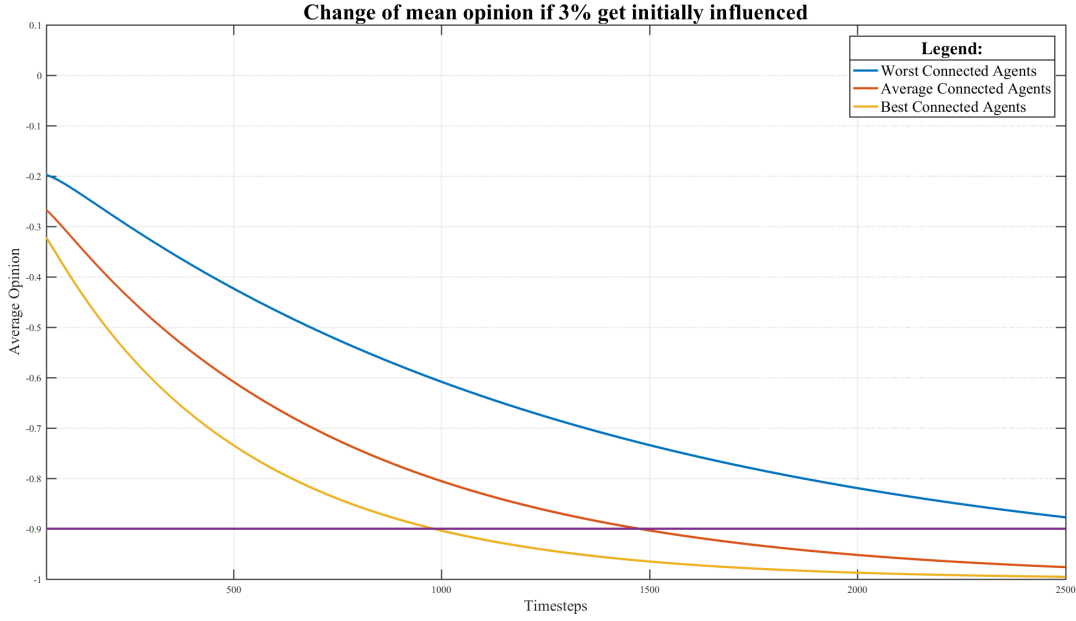


Figure 6: Influencing 3% of all existing agents one can see that the opinion drops rather slowly.

5.4 Analyzing the Opinion Data

The plots consist of 4 graphs: Worst connected agents represent the average opinion development over time if the least connected agents were chosen. Average connected agents stand for the randomly chosen Agents. Because this varies according to the randomness, thousand different outcomes for choosing random agents we calculated and the mean value of all those outcomes will be referred to as the average connected agents. The best connected agents are represented as well in the plot. As point of reference, we also plotted the value -0.9, which helps visualize the point in time when the average opinion dropped below this value.

The different plots represent the different percentages of initially influenced agents. For all calculations, the same randomly generated Network was used.

Influencing the top three percent connected agents the average opinion will drop below -0.9 after approximately 1000 timesteps. Choosing random agents, this process will take 50% more time. The worst Case scenario is almost three times longer (see fig. 6).

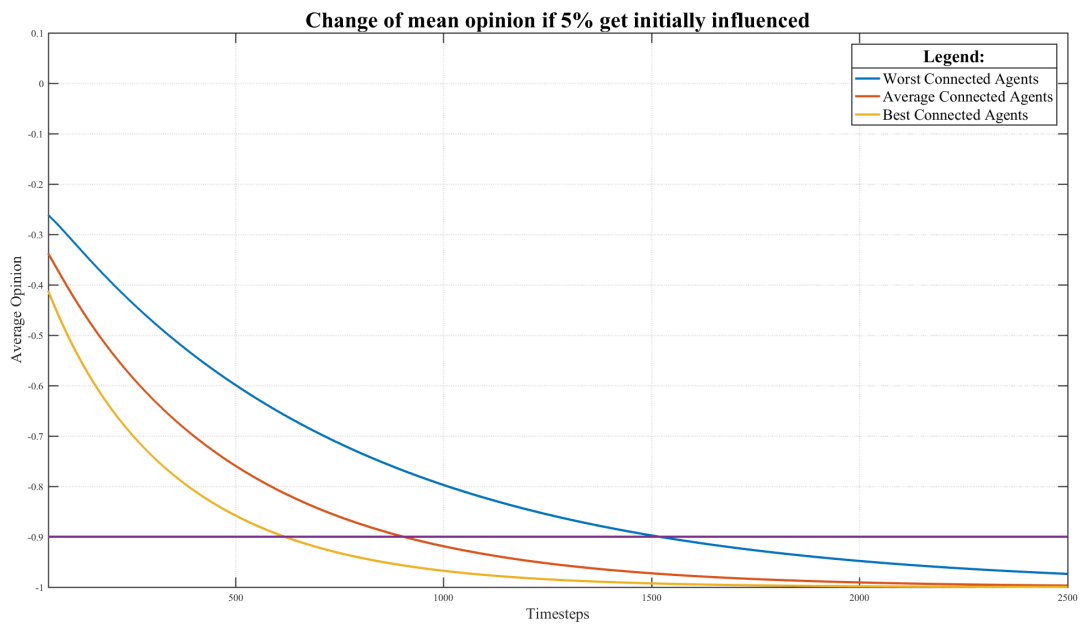


Figure 7: Influencing 5% of all existing agents one can see that the opinion drops overall opinion drops faster.

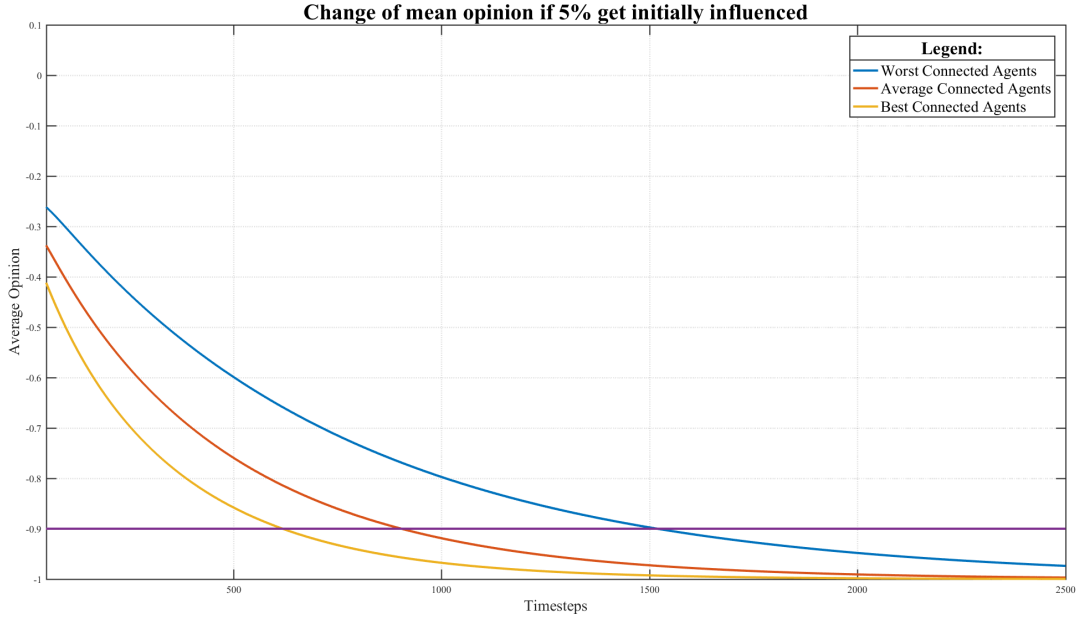


Figure 8: Influencing 8% leads to an increasingly negative slope of all graphs.

Influencing only two percent more agents, the average opinion will drop a lot faster. Interestingly, the average opinion of the randomly chosen agents reaches the reference point earlier than if one were to investigate the network for the best three connected percent of agents and influence them. This worst case scenario is as good as if three instead of five percent of random agents were chosen (see fig. 7).

Adding another three percent will again improve the average connected agents so that they can be compared to the best five percent connected agents. The last two plots show, that increasing the percentage will continuously improve but it needs to be considered that more Initial influenced agents will be costlier as well (see fig. 8,9).

All plots represent the mean value of all the agents in the network. The results show that influencing five percent of all agents by randomly choosing them is as good as investing in searching for the top three percent of agents. From this simulation it is clear that product or information placement should be spread on a wider range rather than focusing too much in searching the best placement, in a condition that one has the budget to do so.

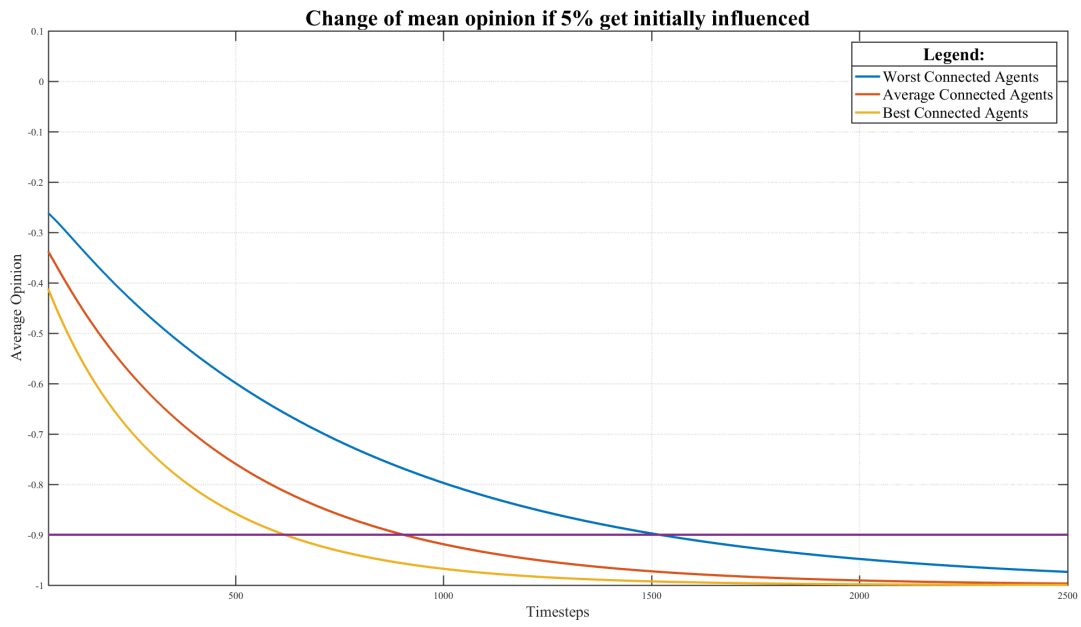


Figure 9: Influencing 10% leads once more to a faster change in opinion. It is especially beneficial for the average and worst connected agents.

6 Summary and Outlook

In conclusion, we have shown that in conditions of restricted budget the target individuals subject to initial information placement have to be carefully chosen. As a matter of fact the choice of agents with highest social capital (largest number of friends) leads to a significant decrease in time that takes the information become viral. On the other hand, we showed, that in a condition when the budget restriction is waived, the best strategy is to target a large pool of individuals without searching for those with highest social impact among their peers, as this yields the same pace of information spread without requiring additional efforts of picking the initial agents.

7 References

- [1] Xiaoyan Qiu et al, *Limited individual attention and online virality of low-quality information*, Nature Human Behaviour 1, Article number: 0132, 2017.
- [2] Pablo Blinder, <https://ch.mathworks.com/matlabcentral/fileexchange/4206-erdos-renyi-random-graph>, (State of 09.12.17)
- [3] Duncan J. Watts & Steven H. Strogatz, *Collective dynamics of small-world networks*, Nature, 1998.



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Modeling and Simulating Social Systems with MATLAB:
"Spread of news in a random network of agents"

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

Blum Josef
Kazakov
Machacek
Martin

First name(s):

Josef
Dmitry
David
Kevin

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Zürich, 15.12.2017

Signature(s)

Machacek J.
Blum
Kazakov
Martin

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.