# Modelling and Analysis of Social Contagion Processes with Dynamic Networks

Alexei Sharpanskykh and Jan Treur

VU University Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
http://www.few.vu.nl/~{sharp,treur} {sharp, treur}@few.vu.nl

**Abstract.** In this paper an agent-based social contagion model with an underlying dynamic network is proposed and analysed. In contrast to the existing social contagion models, the strength of links between agents changes gradually rather than abruptly based on a threshold mechanism. An essential feature of the model – the ability to form clusters – is extensively investigated in the paper analytically and by simulation. Specifically, the distribution of clusters in random and scale-free networks is investigated, the dynamics of links within and between clusters are determined, the minimal distance between two clusters is identified, and the convergence speed of networks is analysed.

## 1 Introduction

Social contagion models have been extensively applied to represent and analyse social decision making, opinion formation, spread of diseases and innovation [1, 2, 4, 5, 7, 8, 12, 14]. Such models describe an evolution of states of individual agents under influence of their neighbouring agents by mutual contagion of these states. In many models [4, 9, 16, 17] the links between agents and their neighbors are constant. In some other models [1, 2, 4, 7] such links may disappear abruptly when states of interacting agents are considered to be too different from each other compared to some threshold. In this paper a social contagion model for social decision making with an underlying network of agents with variable link strengths is proposed and analysed. The strength of the links in the model reflects the degree of influence of one agent to another. The higher the influence to an agent, the higher the extent to which information provided by that agent is used in the decision making; sometimes this also is related to the notion of trust (e.g., [6]). In contrast to the existing models [2, 4, 7, 8], the strength of the links changes gradually in a continuous manner, rather than in a discontinuous manner based on a threshold mechanism. Such a mechanism is supported by sociological literature; e.g., [10], in which much evidence exist that relations between individuals develop continuously.

Many experimental evidences exist that influence correlates positively with similarity of agents; e.g., [3, 11], either in a static sense or in a dynamic sense. This has led to the principle that the closer the opinions of the interacting agents, the higher the mutual degrees of influence of the agents is (static perspective) or will become (dynamic perspective). Such an assumption underlies most of the existing models of social influence [2, 4, 7, 8, 12, 14].

Inspired by these findings, the dynamics of the links in the proposed model is defined based on the dynamic variant of this principle: closeness of opinions leads to a positive change of connection strength over time. An important feature of the proposed model is that for certain ranges of parameter values clusters of agents emerge that are isolated from each other. A *cluster* is a set of connected agents (i.e., a connected graph) with the same states (e.g., opinions).

The dynamics of social decision making based on the model was analysed by simulation and by mathematical analysis. In particular, the formation and dynamics of clusters of agents was investigated. Both simulation and analytical findings show that the links between the agents within a cluster become stronger over time, and the corresponding degrees of influence tend to 1 (i.e., the highest strength value). At the same time, the strength of the links between the agents in different clusters degrade, and the corresponding degrees of influence tend to 0 (the lowest strength value equivalent to the absence of a connection between the agents). Furthermore, it turns out that different emerging clusters have a certain minimal distance, which was determined analytically. Cluster size distributions in random and scale-free networks were investigated by simulation. The rate of convergence of agent states to equilibrium were investigated both by simulation and analytically and are discussed in the paper. Furthermore, the cluster formation and convergence properties of the proposed model are compared with the corresponding properties of the well-cited threshold-based model developed by Hegselmann and Krause [8].

The paper is organized as follows. In Section 2 a social contagion model for social decision making with an underlying dynamic network is proposed. Results of the model analysis analytically and by simulation are presented in Section 3. Section 4 concludes the paper.

## 2 The Dynamical Model

The model describes dynamics of decision making by agents in a group as a process of social contagion. The opinion $q_{s,i}$ of an agent *i* for a decision option *s* is expressed by a real number in the range [*0*, *1*], reflecting the degree of the agent's support for the option. For each option each agent communicates its opinion to other agents. Agents communicate only with those agents to which they are connected in a social network. In this study two network topologies are considered:

- a *scale-free network topology*: a connected graph with the property that the number of links originating from a given node representing an agent has a power law distribution. In such networks the majority of the agents have one or two links, but a few agents have a large number of links;
- a *random network topology*: a graph, in which links between nodes occur at random. Only connected graphs are considered in this study.

To compare the dynamics in both types of networks, the networks used in this study were generated with 5000 agents and the same average node degree equal to 4.5. This value is close to the average node degree of real social networks.

It is assumed that the agents are able to both communicate and receive opinions to/from the agents, to which they are connected (i.e., the links are bidirectional). Furthermore, a weight $\gamma_{i,j} \in [0,1]$, indicating the degree of influence of agent *i* on agent *j*, is associated with each link for each direction of interaction. This weight determines to which extent the opinion of agent *i* is taken into account in the update of the opinion of agent *j* for each option. These weights may or may not be symmetric.

It is assumed that the agents interact with each other synchronously, i.e., at the same time (parallel interaction mode). For a quantitative comparison of the dynamics of social contagion models with the parallel interaction mode with models with the sequential interaction mode, please refer to [16, 17].

In the parallel mode, the opinion states of the agents are updated at the same time point $t$ as follows:

$$q_{s,i}(t+\Delta t) = q_{s,i}(t) + \eta_i \, \delta_{s,i}(t)\Delta t \qquad (1)$$

Here $\eta_i$ is an agent-dependent parameter within the range [$0,1$], which determines how fast the agent adjusts to the opinion of other agents, and

$$\delta_{s,i}(t) = \sum_{j\in AG} \gamma_{j,i}(t)(q_{s,j}(t) - q_{s,i}(t)) / \sum_{j\in AG} \gamma_{j,i}(t)$$

is the amount of change of the agent $i$'s opinion; $AG$ is the set of all agent names.

The normalization by $\sum_{j\in AG} \gamma_{j,i}(t)$ has the effect that the agent balances by a relative comparison its own self-influence $\gamma_{i,i}(t)$ (i.e., self-assurance that its own opinion is correct) with the influences of other agents.

The degrees of influence $\gamma_{i,j}$ also change over time based on the principle: the closer the opinions of the interacting agents, the higher the mutual degrees of influence of the agents will become. This dynamic principle may be formalised by different functions as follows:

$$\gamma_{i,j}(t+\Delta t) = \gamma_{i,j}(t) + f_{i,j}(\gamma_{i,j}(t), q_{s,i}(t), q_{s,j}(t)) \, \Delta t \qquad (2)$$

where for function $f_{i,j}(X, Y, Z)$ the main example used is:

$$f_{i,j}(X, Y, Z) = \text{Pos}(\alpha_{ij} (\beta_{ij} - (Y-Z)^2))(1-X) - \text{Pos}(-\alpha_{ij} (\beta_{ij} - (Y-Z)^2))X \qquad (3)$$

with $\text{Pos}(x) = (|x|+x)/2$, $\alpha_{ij}$ is a speed parameter and $\beta_{ij}$ is a threshold or tolerance parameter.

Other alternatives for $f_{i,j}(X, Y, Z)$ are:

$$X \, \alpha_{ij} (\beta_{ij} - (Y-Z)^2)(1-X) \qquad\qquad X \, \lambda_{ij}(1-|Y-Z|)(1-X) - \zeta_{ij} \, |Y-Z|$$

Here $\lambda_{ij}$ is an amplification parameter and $\zeta_{ij}$ is an inhibition parameter. Note that (1) and (2) are expressed in difference equation format. In Section 3 they are also considered in differential equation format.

A threshold-based model with abruptly changing links and threshold $\tau$ as described in [8] can be obtained by defining $f_{i,j}(X, Y, Z)$ as follows:

$$f_{i,j}(X, Y, Z) = 1-X, \quad \text{when } |Y-Z| \leq \tau \qquad (4)$$
$$f_{i,j}(X, Y, Z) = -X, \quad \text{when } |Y-Z| > \tau \qquad (5)$$

# 3 Model Analysis

In this section first formal analytical results for the model are presented. After that the model is analysed by simulation.

## 3.1 Mathematical Analysis

For a mathematical analysis, as a point of departure the following differential equations were derived from (1), (2) and (3) in Section 2. For $\gamma_{i,j}(t)$:

$$\mathbf{d}\gamma_{i,j}(t)/\,\mathbf{d}t = \text{Pos}(\alpha_{ij} (\beta_{ij} - (q_{s,i}(t) - q_{s,j}(t))^2))(1-\gamma_{i,j}(t)) - \text{Pos}(-\alpha_{ij} (\beta_{ij} - (q_{s,i}(t) - q_{s,j}(t))^2)) \, \gamma_{i,j}(t)$$

The differential equations for the $q_{s,i}(t)$ are:

$$\mathbf{d}q_{s,i}(t)/\,\mathbf{d}t = \eta_i \sum_{j\in AG} \gamma_{j,i}(t)(q_{s,j}(t) - q_{s,i}(t)) / \sum_{j\in AG} \gamma_{j,i}(t)$$

### *Equilibrium values for connection strengths $\gamma_{i,j}(t)$*

First, the equilibrium values $\underline{\gamma}_{i,j}$ for $\gamma_{i,j}(t)$ are addressed. The standard approach is to derive an equilibrium equation from the differential equation by putting $\mathbf{d}\gamma_{i,j}(t)/\,\mathbf{d}t = 0$. For the specific case for the function $f_{i,j}(X, Y, Z)$ this is

$$\text{Pos}(\alpha_{ij}\,(\beta_{ij} - (q_{s,i}(t)\text{-}\,q_{s,j}(t))^2))(1\text{-}\gamma_{i,j}(t)) \\ - \text{Pos}(-\alpha_{ij}\,(\beta_{ij} - (q_{s,i}(t)\text{-}\,q_{s,j}(t))^2))\,\gamma_{i,j}(t) = 0$$

The following lemma is used:

**Lemma 1**

For any numbers $\alpha$ and $\beta$ the following are equivalent:

(i) $\alpha\,\text{Pos}(x) + \beta\,\text{Pos}(-x) = 0$

(ii) $\alpha\,\text{Pos}(x) = 0$ and $\beta\,\text{Pos}(-x) = 0$

(iii) $x = 0$ or $x>0$ and $\alpha = 0$ or $x<0$ and $\beta = 0$. ■

Using Lemma 1 it is found that the above equilibrium equation has three solutions

$|\underline{q}_{s,i} - \underline{q}_{s,j}| = \sqrt{\beta_{ij}}$

$|\underline{q}_{s,i} - \underline{q}_{s,j}| > \sqrt{\beta_{ij}}$ and $\underline{\gamma}_{i,j} = 0$

$|\underline{q}_{s,i} - \underline{q}_{s,j}| < \sqrt{\beta_{ij}}$ and $\underline{\gamma}_{i,j} = 1$

More can be found about the circumstances under which such equilibria can occur, and for a wider class of functions $f_{i,j}(X, Y, Z)$. The following symmetry properties are relevant.

**Definition**

The network is called *weakly symmetric* if for all nodes *i* and *j* at all time points it holds $\gamma_{i,j} = 0 \;\Leftrightarrow\; \gamma_{j,i} = 0$ or, equivalently: $\gamma_{i,j} > 0 \;\Leftrightarrow\; \gamma_{j,i} > 0$. The network is called *fully symmetric* if $\gamma_{i,j} = \gamma_{j,i}$ for all nodes *i* and *j* at all time points.

Note that the network is fully symmetric if the initial values for $\gamma_{i,j}$ and $\gamma_{j,i}$ are equal and $f_{i,j}(X, Y, Z)=f_{j,i}(X, Z, Y)$ for all $X$, $Y$, $Z$; the latter condition is fulfilled for the specific case if $\alpha_{i,j}=\alpha_{j,i}$ and $\beta_{i,j}=\beta_{j,i}$. The following lemma is used to obtain Theorem 1.

**Lemma 2**

a) If for some node *i* at time *t* for all nodes *j* with $q_{s,j}(t) > q_{s,i}(t)$ it holds $\gamma_{j,i}(t) = 0$, then $q_{s,i}(t)$ is decreasing at *t*: $\mathbf{d}q_{s,i}(t)/\mathbf{d}t \leq 0$.

b) If, moreover, a node *k* exists with $q_{s,k}(t) < q_{s,i}(t)$ and $\gamma_{k,i}(t) > 0$ then $q_{s,i}(t)$ is strictly decreasing at *t*: $\mathbf{d}q_{s,i}(t)/\mathbf{d}t < 0$.

**Proof**: a) From the expressions for $\delta_{s,i}(t)$ it follows that $\delta_{s,i}(t) \leq 0$, and therefore $\mathbf{d}q_{s,i}(t)/\mathbf{d}t \leq 0$, so $q_{s,i}(t)$ is decreasing at *t*.

b) In this case it follows that $\delta_{s,i}(t) < 0$ and therefore $\mathbf{d}q_{s,i}(t)/\mathbf{d}t < 0$, so $q_{s,i}(t)$ is strictly decreasing.■

**Theorem 1 (Equilibrium values $\underline{\gamma}_{i,j}$)**

Suppose the network is weakly symmetric, and $f_{i,j}(X,Y,Y) > 0$ for all $X$, $Y$ with $0 < X < 1$. Then in an equilibrium state for any two nodes *i* and *j* it holds $\underline{\gamma}_{i,j} = 0$ or $\underline{\gamma}_{i,j} = 1$. More specifically, the following hold:

a) In an equilibrium state with $\underline{q}_{s,i} \neq \underline{q}_{s,j}$ it holds $\underline{\gamma}_{i,j} = 0$.

b) In an equilibrium state with $\underline{\mathbf{q}}_{s,i} = \underline{\mathbf{q}}_{s,j}$ it holds $\underline{\gamma}_{i,j} = 0$ or $\underline{\gamma}_{i,j} = 1$. If $q_{s,i}(t) = q_{s,j}(t)$ and $0 < \gamma_{i,j}(t) < 1$, then $\gamma_{i,j}(t)$ is strictly increasing at time $t$: $\mathbf{d}\gamma_{i,j}(t)/\mathbf{d}t > 0$.

**Proof**: a) Suppose in an equilibrium state $\underline{\mathbf{q}}_{s,i} \neq \underline{\mathbf{q}}_{s,j}$ and $\underline{\gamma}_{i,j}, \underline{\gamma}_{j,i} > 0$ for some nodes $i$ and $j$. Take the node $i$ with this property with highest value $\underline{\mathbf{q}}_{s,i}$. Then for all nodes $j$ with $\underline{\mathbf{q}}_{s,j} > \underline{\mathbf{q}}_{s,i}$ it holds $\underline{\gamma}_{i,j} = \underline{\gamma}_{j,i} = 0$. Now apply Lemma 2 to this node $i$. It follows that $\mathbf{d}q_{s,i}(t)/\mathbf{d}t < 0$, so $q_{s,i}(t)$ is not in equilibrium. This contradicts the assumption. Therefore $\underline{\gamma}_{i,j} = 0$ for all nodes $i$ and $j$ with $\underline{\mathbf{q}}_{s,i} \neq \underline{\mathbf{q}}_{s,j}$.

b) If $0 < \gamma_{i,j}(t) < 1$, then from $q_{s,i}(t) = q_{s,j}(t)$ it follows that $f_{i,j}(\gamma_{i,j}(t), q_{s,i}(t), q_{s,j}(t)) > 0$. From this it follows that $\mathbf{d}\gamma_{i,j}(t)/\mathbf{d}t > 0$: $\gamma_{i,j}(t)$ is strictly increasing and is not in equilibrium. Therefore in an equilibrium state with $\underline{\mathbf{q}}_{s,i} = \underline{\mathbf{q}}_{s,j}$ it holds $\underline{\gamma}_{i,j} = 0$ or $\underline{\gamma}_{i,j} = 1$.■

Note that the criterion on the function $f_{i,j}(X, Y, Z)$ in Theorem 1 is satisfied for the specific function $f_{i,j}(X, Y, Z) = \text{Pos}(\alpha_{ij}\ (\beta_{ij} - (Y\text{-}Z)^2))(1\text{-}X)\text{-}\ \text{Pos}(-\alpha_{ij}\ (\beta_{ij} - (Y\text{-}Z)^2))X$ if and only if $\alpha_{ij}, \beta_{ij} > 0$, which is the case.

### *Equilibrium values for* $\mathbf{q}_{s,i}(t)$

In an equilibrium of the network not only the $\gamma_{i,j}$ are in an equilibrium $\underline{\gamma}_{i,j}$ but also the $q_{s,i}$ . From the differential equations for the $q_{s,i}$ it follows that the equilibrium values $\underline{\mathbf{q}}_{s,i}$ for $q_{s,i}\ (t)$ have to satisfy $\sum_{j\in AG} \underline{\gamma}_{j,i}\ (\underline{\mathbf{q}}_{s,j} - \underline{\mathbf{q}}_{s,i}) = 0$.

When $\underline{\gamma}_{j,i} = 0$ for all $j$, then from the differential equation it follows that $q_{s,i}$ is in equilibrium irrespective of what value it has. Suppose at least one node $j$ exists with $\underline{\gamma}_{j,i} \neq 0$. Then the equilibrium equations can be rewritten as

$$\underline{\mathbf{q}}_{s,i} = \sum_{j\in AG} (\underline{\gamma}_{j,i} / \sum_{k\in AG} \underline{\gamma}_{k,i})\ \underline{\mathbf{q}}_{s,j}$$

This provides a system of linear equations for the $\underline{\mathbf{q}}_{s,i}$ that could be solved, unless they are trivial or dependent. To analyse this, suppose $S_i$ is the cluster (of size $s_i$) of nodes with same equilibrium value as $\underline{\mathbf{q}}_{s,i}$ :

$$S_i = \{\ j \mid \underline{\mathbf{q}}_{s,j} = \underline{\mathbf{q}}_{s,i}\ \} \qquad\qquad s_i = \#\,(S_i)$$

In Theorem 1a) above it has been found that $\underline{\gamma}_{j,i} = 0$ if $j \notin S_i$. Therefore

$$\sum_{j\in AG} \underline{\gamma}_{j,i}\ \underline{\mathbf{q}}_{s,j} = \sum_{j\in S_i} \underline{\gamma}_{j,i}\ \underline{\mathbf{q}}_{s,j} = \sum_{j\in S_i} \underline{\gamma}_{j,i}\ \underline{\mathbf{q}}_{s,i}$$

Then the equilibrium equation for $\underline{\mathbf{q}}_{s,i}$ becomes:

$$\underline{\mathbf{q}}_{s,i} = \sum_{j\in AG} \underline{\gamma}_{j,i}\ \underline{\mathbf{q}}_{s,j} / \sum_{j\in AG} \underline{\gamma}_{j,i} = \sum_{j\in S_i} \underline{\gamma}_{j,i}\ \underline{\mathbf{q}}_{s,j} / \sum_{j\in S_i} \underline{\gamma}_{j,i} = \underline{\mathbf{q}}_{s,i}$$

Thus these equations do not provide a feasible way to obtain information about the equilibrium values $\underline{\mathbf{q}}_{s,i}$ . However, by different methods at least some properties of the equilibrium values $\underline{\mathbf{q}}_{s,i}$ can be derived, as is shown below.

The following conditions on the function $f_{i,j}(X, Y, Z)$ are assumed:

**Definition**

The function $f_{i,j}(X, Y, Z)$ has a *threshold* $\tau$ for $Y - Z$ if

a) For all $Y$ and $Z$ *it holds*

$$f_{i,j}(0, Y, Z) \geq 0 \qquad\qquad f_{i,j}(1, Y, Z) \leq 0$$

b) For all X with $0 < X < 1$ and all $Y$ and $Z$ it holds

$$f_{i,j}(X, Y, Z) > 0 \quad \text{iff} \quad |Y - Z| < \tau$$
$$f_{i,j}(X, Y, Z) = 0 \quad \text{iff} \quad |Y - Z| = \tau$$

$$f_{i,j}(X, Y, Z) < 0 \quad \text{iff} \quad |Y - Z| > \tau$$

Note that (given that $\alpha_{ij} > 0$ is assumed) the function

$$f_{i,j}(X, Y, Z) = \mathrm{Pos}(\alpha_{ij}\,(\beta_{ij} - (Y\text{-}Z)^2))(1\text{-}X) - \mathrm{Pos}(-\alpha_{ij}\,(\beta_{ij} - (Y\text{-}Z)^2))X$$

satisfies these conditions for threshold $\sqrt{\beta_{ij}}$.

**Theorem 2 (Distance between equilibrium values $\underline{\mathbf{q}}_{s,i}$)**
Suppose the network is weakly symmetric, the function $f_{i,j}(X, Y, Z)$ has a threshold $\tau$, and the network reaches an equilibrium state with values $\underline{\mathbf{q}}_{s,i}$ for the different nodes $i$. Then for every two nodes $i$ and $j$ if their equilibrium values $\underline{\mathbf{q}}_{s,i}$ and $\underline{\mathbf{q}}_{s,j}$ are distinct, and the initial values for $\gamma_{i,j}$ and $\gamma_{j,i}$ are nonzero, they have a distance of at least $\tau$ : $|\underline{\mathbf{q}}_{s,i} - \underline{\mathbf{q}}_{s,j}| \geq \tau$. In particular, when all initial values for $\gamma_{i,j}$ and $\gamma_{j,i}$ are nonzero, there are at most $1 + 1/\tau$ distinct equilibrium values $\underline{\mathbf{q}}_{s,i}$ .
**Proof**: Suppose two nodes are given with distinct equilibrium values $\underline{\mathbf{q}}_{s,i}$ and $\underline{\mathbf{q}}_{s,j}$ with distance less than $\tau$. Then $|\underline{\mathbf{q}}_{s,i} - \underline{\mathbf{q}}_{s,j}| = \tau - \delta$ for some $\delta > 0$. Without loss of generality it can be assumed that $\underline{\mathbf{q}}_{s,j} < \underline{\mathbf{q}}_{s,i}$ . Because $q_{s,i}(t)$ converges to $\underline{\mathbf{q}}_{s,i}$ and $q_{s,j}(t)$ converges to $\underline{\mathbf{q}}_{s,j}$ it follows that there exists a $t$ such that for all $t'$ with $t' \geq t$ it holds

$$|q_{s,i}(t') - \underline{\mathbf{q}}_{s,i}| < \tfrac{1}{2}\delta \quad \text{and} \quad |q_{s,j}(t') - \underline{\mathbf{q}}_{s,i}| < \tfrac{1}{2}\delta$$

Therefore for all $t' \geq t$ it holds (by the triangle inequality)

$$\begin{aligned}
|q_{s,i}(t') - q_{s,j}(t')| &= |(q_{s,i}(t') - \underline{\mathbf{q}}_{s,i}) - (q_{s,j}(t') - \underline{\mathbf{q}}_{s,j}) + (\underline{\mathbf{q}}_{s,i} - \underline{\mathbf{q}}_{s,j})| \\
&\leq |q_{s,i}(t') - \underline{\mathbf{q}}_{s,i}| + |q_{s,j}(t') - \underline{\mathbf{q}}_{s,j}| + |\underline{\mathbf{q}}_{s,i} - \underline{\mathbf{q}}_{s,j}| \\
&< \tfrac{1}{2}\delta + \tfrac{1}{2}\delta + \tau - \delta = \tau
\end{aligned}$$

So, $|q_{s,i}(t') - q_{s,j}(t')| < \tau$ for all $t' \geq t$.

Since the function $f_{i,j}(X, Y, Z)$ has threshold $\tau$, from this it follows that for all $t' \geq t$ when $0 < \gamma_{i,j}(t') < 1$ it holds

$$f_{i,j}(\gamma_{i,j}(t'), q_{s,i}(t'), q_{s,j}(t')) > 0$$

From the differential equation for $\gamma_{i,j}(t)$ it follows that when $0 < \gamma_{i,j}(t') < 1$ it holds that $\mathbf{d}\gamma_{i,j}(t')/\mathbf{d}t > 0$ for all $t' \geq t$. From Theorem 1a) it follows that the equilibrium value for $\gamma_{i,j}(t)$ is $\underline{\boldsymbol{\gamma}}_{i,j} = 0$. Taking into account that always $\gamma_{i,j}(t') \geq 0$, and that $\mathbf{d}\gamma_{i,j}(t')/\mathbf{d}t > 0$ when $0 < \gamma_{i,j}(t') < 1$ for all $t' \geq t$ this equilibrium value $\underline{\boldsymbol{\gamma}}_{i,j} = 0$ can only be reached when $\gamma_{i,j}(t) = 0$ for all $t$, which contradicts the fact that the initial value for $\gamma_{i,j}$ is nonzero. Summarising, the assumption that $|\underline{\mathbf{q}}_{s,i} - \underline{\mathbf{q}}_{s,j}| < \tau$ has been falsified, so the distance between two distinct equilibrium values $\underline{\mathbf{q}}_{s,i}$ and $\underline{\mathbf{q}}_{s,j}$ is at least $\tau$: $|\underline{\mathbf{q}}_{s,i} - \underline{\mathbf{q}}_{s,j}| \geq \tau$. The last statement of the theorem follows since the interval [$0$, $1$] can be divided in at most $1/\tau$ subintervals of length $\tau$. ■

In case the network is fully symmetric (i.e., $\gamma_{j,i} = \gamma_{i,j}$ for all $i$ and $j$) the equilibrium values $\underline{\mathbf{q}}_{s,i}$ can be related to the initial values $q_{s,i}(t)$. In this case the sum $\Sigma_i\, q_{s,i}(t)$ is preserved: $\Sigma_{i \in AG}\, q_{s,i}(t) = \Sigma_{i \in AG}\, q_{s,i}(t')$ for all $t$ and $t'$. From $\gamma_{j,i} = \gamma_{i,j}$ this can be established as follows:

$$\mathbf{d}\,\Sigma_{i \in AG}\, q_{s,i}(t)/\mathbf{d}t = \Sigma_{i \in AG}\, \mathbf{d}\, q_{s,i}(t)/\mathbf{d}t = \eta_i\, \Sigma_{i \in AG} \sum_{j \in AG} \gamma_{j,i}(t)(q_{s,j}(t) - q_{s,i}(t))/\sum_{j \in AG} \gamma_{j,i}(t)$$
$$= \eta_i\, [\Sigma_{k \in AG}\, \Sigma_{i \in AG}\, \gamma_{i,k} q_{s,k}(t) - \Sigma_{k \in AG}\, \Sigma_{i \in AG}\, \gamma_{i,k}\, q_{s,k}(t)]/\sum_{j \in AG} \gamma_{j,i}(t) = 0$$

The fact that $\Sigma_{i\in AG}\ q_{s,i}(t)$ is preserved can be applied to compare the equilibrium values $\underline{\boldsymbol{q}}_{s,i}$ to the initial values $q_{s,i}(t_0)$. Let $\underline{\boldsymbol{S}}$ be the set of clusters of equilibria:

$\underline{\boldsymbol{S}} = \{S_i \mid i \text{ any node}\}$

For $C\in\underline{\boldsymbol{S}}$ define

$\underline{\boldsymbol{q}}_{s,C} = \underline{\boldsymbol{q}}_{s,j}$ for any $j\in C$

$s_C = \#(C) = s_j$ for any $j\in C$

Then from the preservation it follows

$$\Sigma_{i\in AG}\ \underline{\boldsymbol{q}}_{s,i} = \Sigma_{i\in AG}\ q_{s,i}(t_0)$$

Therefore

$$\Sigma_{C\in\underline{\boldsymbol{S}}}\ \Sigma_{i\in C}\ \underline{\boldsymbol{q}}_{s,i} = \Sigma_{i\in AG}\ q_{s,i}(t_0)$$

$$\Sigma_{C\in\underline{\boldsymbol{S}}}\ (s_C/n)\ \underline{\boldsymbol{q}}_{s,C} = \Sigma_{i\in AG}\ q_{s,i}(t_0)/n$$

with $n=\#(AG)$ the total number of nodes. So, the weighted average over the clusters (with as weights the fraction of the total number of nodes in the cluster) is the average of the initial values $q_{s,i}(t_0)$. These are summarised in the following theorem:

**Theorem 3 (Equilibria $\underline{\boldsymbol{q}}_{s,i}$ in fully symmetric case)**
Suppose the network is fully symmetric. Then the sum $\Sigma_{i\in AG}\ q_{s,i}(t)$ is preserved over time. Moreover, the weighted average of the equilibrium values for the clusters, with the fraction of the total number of nodes in the cluster as weights, is the average of the initial values:

$$\Sigma_{C\in\underline{\boldsymbol{S}}}\ (s_C/n)\ \underline{\boldsymbol{q}}_{s,C} = \Sigma_{i\in AG}\ q_{s,i}(t_0)/n \quad \blacksquare$$

Because of the space limitations, analysis of the model behaviour around equilibria is described in an online appendix at http://iccci13.9k.com/app.pdf

## 3.2 Analysis by Simulation

In this section two model variants from Section 2 are analysed by simulation: model *M1* with continuously changing links (equation (3)) and a threshold-based model *M2* with abruptly changing links (equations (4) and (5)). Both models have the same threshold $\tau=\sqrt{\beta_{ij}}$. The models were simulated in Matlab.

To compare the models, 10 different random network topologies with 5000 agents and 10 different scale-free network topologies with 5000 agents were generated. The scale-free networks were obtained using the Complex Networks Package [13] with scale-free degree distribution of $\alpha$=-2.2 (as in many real social networks). The average node degree of such networks with 5000 agents equals 4.5. The random networks were generated with the same average node dergree.

The agents formed opinions on some topic *s*. The parameters of the agents and of the links were uniformly distributed as follows: $\eta_i\in[0.5, 1]$; $q_{s,i}(0)\in[0,1]$; $\gamma_{i,j}(0)\in(0, 1]$ (in model *M2* $\gamma_{i,j}(0)=1$, if there was a link between *i* and *j*, and *0* otherwise). These distributions are assumed to represent the diversity that naturally occurs in real-world agent populations.

The simulation time was 300 time points and $\Delta t=1$.

In the previous section 3.1 it was proven that when all initial values $\gamma_{i,j}(0)$ in the population of agents are nonzero, then at most $1 + 1/\sqrt{\beta_{ij}}$ clusters can be formed in the model with threshold $\tau=\sqrt{\beta_{ij}}$. The minimal distance between two clusters is $\sqrt{\beta_{ij}}$,

**Table 1.** The minimal distances between the clusters in 10 random and 10 scale-free networks determined by simulation and analytically

| Parameter settings | β=0.001 | | | β=0.0025 | | | β=0.01 | | |
|---|---|---|---|---|---|---|---|---|---|
| | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 |
| **Scale-free** | 0.034 | 0.041 | 0.038 | 0.053 | 0.059 | 0.07 | 1 cluster | | |
| **Random** | 0.041 | 0.034 | 0.033 | 0.1 | 0.084 | 0.2 | 0.4 | 0.15 | 0.23 |
| **Analytical** | 0.032 | | | 0.05 | | | 0.1 | | |

provided that these two clusters were not disconnected initially. In this section we investigate how parameters $\alpha_{ij}$ and $\beta_{ij}$ influence the number and size of the clusters emerging in the scale-free and random networks. Furthermore, the rate of convergence of the agent opinions is determined for different parameter settings.

In the simulation study three values for $\beta_{ij}$: 0.001, 0.0025, 0.01 and three values for $\alpha_{ij}$: 1, 10, 20 were used. According to the findings from Section 3.1, at most 33 clusters could emerge for $\beta_{ij}$=0.001, 21 clusters for $\beta_{ij}$=0.0025 and 11 clusters for $\beta_{ij}$=0.01. In the networks used in the simulation less clusters were formed, as these networks were not fully connected. For $\beta_{ij} > 0.01$, in most cases only one cluster was formed containing all the agents. The minimal distances between the clusters in the simulated networks were greater than $\sqrt{\beta_{ij}}$. (Table 1).

In the tables with the results small clusters have the size up to 101 agents; medium clusters contain more than 100 but less than 1001 agents, and large clusters comprise of more than 1000 agents.

Besides $\beta_{ij}$, parameter $\alpha_{ij}$ also influences the number of clusters with the limit $1 + 1/\sqrt{\beta_{ij}}$ (Tables 2, 3), however in a more intricate manner. For example, in the random networks the largest number of clusters emerges when $\alpha_{ij}$ takes intermediate values (around 10), whereas in the scale-free networks many clusters tend to form with low values of $\alpha_{ij}$.

In the random networks simulated with *M1* model only small and large size clusters tend to form (Table 4). Only one large size cluster emerges in the *M1*-based simulations, which contains the great majority of the agents in the population. Model *M2* produces also medium size clusters, but their amount is less than the number of the small size clusters. This can be partially explained by the absence of central agents or

**Table 2.** The mean and standard deviation values (in parentheses) of the numbers of small ( ≤ 100 agents), medium ( >100 and ≤ 1000 agents) and large ( >1000 agents) size clusters emerging in 10 random networks

| Parameter settings | β=0.001 | | | β=0.0025 | | | β=0.01 | | |
|---|---|---|---|---|---|---|---|---|---|
| | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 |
| **Small *M1*** | 1.1 (1) | 4 (1.2) | 2.9 (1.2) | 0.6 (0.9) | 1.9 (0.7) | 1 (0.8) | 0.2 (0.4) | 0.3 (0.6) | 0.2 (0.4) |
| **Small *M2*** | 5.1 (1.1) | | | 2.9 (0.9) | | | 0.6 (0.5) | | |
| **Medium *M1*** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Medium *M2*** | 3.8 (0.9) | | | 2 (0.8) | | | 0 | | |
| **Large *M1*** | 1 | 1 | 1.1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **Large *M2*** | 2.3 (0.6) | | | 1.3 (0.5) | | | 1 | | |

**Table 3.** The mean and standard deviation values (in parentheses) of the numbers of small ( ≤ 100 agents), medium ( >100 and ≤ 1000 agents) and large ( >1000 agents) size clusters emerging in 10 scale-free networks

| **Parameter settings** | **β=0.001** | | | **β=0.0025** | | | **β=0.01** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **α=1** | **α=10** | **α=20** | **α=1** | **α=10** | **α=20** | **α=1** | **α=10** | **α=20** |
| **Small *M1*** | 5.9 (0.9) | 2.4 (0.8) | 0.6 (0.6) | 1.8 (0.9) | 0.8 (0.7) | 0.1 (0.3) | 0 | 0 | 0 |
| **Small *M2*** | 3.1 (1.3) | | | 1.6 (0.8) | | | 0.1 (0.3) | | |
| **Medium *M1*** | 0 | 2.4 (0.8) | 2.8 (0.6) | 0 | 1.6 (0.5) | 1.6 (0.7) | 0 | 0 | 0 |
| **Medium *M2*** | 6.6 (1.3) | | | 4.1 (1) | | | 1.6 (0.5) | | |
| **Large *M1*** | 1 | 1 | 1 | 1 | 1.1 (0.3) | 1.2 (0.4) | 1 | 1 | 1 |
| **Large *M2*** | 1.6 (0.5) | | | 1.5 (0.5) | | | 1 | | |

hubs with many connections in the random networks. Such agents would be able to attract large groups of other agents and influence their opinions so that the whole agent population may become polarized in larger opposing clusters.

More medium size clusters tend to emerge in the scale-free networks simulated with both *M1* and *M2* models (Tables 3 and 5). The scale-free network topology contains hub agents. When such agents have opposing opinions, they form the basis for future clusters in which the whole population is divided. Because of such agents, who group others around themselves, the small and medium size clusters in the scale-free networks are on average larger than the ones emerging in the random networks (Tables 4 and 5).

According to the analytical results from Section 3.1, the links between the agents within a cluster in model *M1* become stronger and the corresponding degrees of

**Table 4**. The mean and standard deviation values (in parentheses) of the sizes of small ( ≤ 100 agents), medium ( >100 and ≤ 1000 agents) and large ( >1000 agents) clusters emerging in 10 random networks

| **Parameter settings** | **β=0.001** | | | **β=0.0025** | | | **β=0.01** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **α=1** | **α=10** | **α=20** | **α=1** | **α=10** | **α=20** | **α=1** | **α=10** | **α=20** |
| **Small *M1*** | 1.1 (0.3) | 2.2 (1.6) | 2.1 (1.9) | 1 | 1.7 (1.2) | 1.6 (0.8) | 1 | 1.3 (0.6) | 1 |
| **Small *M2*** | 17.5 (22.1) | | | 19.4 (25.2) | | | 5.7 (3.4) | | |
| **Medium *M1*** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Medium *M2*** | 405.7 (253.8) | | | 439.7 (271.8) | | | 0 | | |
| **Large *M1*** | 4998 (1) | 4996 (3) | 4539 (1026) | 4999 (1) | 4997 (2) | 4998 (1) | 4999 (0.4) | 4999 (1) | 4998 (0.4) |
| **Large *M2*** | 1465 (427) | | | 3126 (1280) | | | 4997 (3) | | |

**Table 5.** The mean and standard deviation values (in parentheses) of the sizes of small ( ≤ 100 agents), medium ( >100 and ≤ 1000 agents) and large ( >1000 agents) clusters emerging in 10 scale-free networks

| Parameter settings | β=0.001 | | | β=0.0025 | | | β=0.01 | | |
|---|---|---|---|---|---|---|---|---|---|
| | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 |
| **Small *M1*** | 1.7 (0.9) | 47.5 (35.9) | 1 | 1.7 (0.7) | 53.3 (34.6) | 1 | 0 | 0 | 0 |
| **Small *M2*** | 30.4 (30) | | | 36.7 (33) | | | 48 | | |
| **Medium *M1*** | 0 | 222 (136) | 429 (166) | 0 | 276 (173) | 450 (192) | 0 | 0 | 0 |
| **Medium *M2*** | 424.8 (242) | | | 462.6 (271) | | | 543.5 (241) | | |
| **Large *M1*** | 4990 (3) | 4351 (228) | 3797 (246) | 4997 (1.8) | 4105 (1051) | 3566 (1200) | 5000 | 5000 | 5000 |
| **Large *M2*** | 1314 (165) | | | 2030 (727) | | | 4126 (131) | | |

influence $\gamma_{i,j}(t)$ gradually tend to *1*, whereas the links between the agents from different clusters gradually disappear. In contrast to *M1*, links in *M2* appear and disappear instantaneously, depending on the states of the agents, which leads to faster and more abrupt cluster formation. Intuitively, this may be compared to instantaneous decision making (e.g., under high stress) without more detailed consideration and discussion. Because of this, in *M2* agents break links more easily than in *M1*, and thus, more clusters emerge (Tables 2, 3).

In general, the convergence time of a network decreases with the increase of $\beta_{ij}$. (Table 6). This is because the agents become more tolerant to each other's opinions, and less clusters tend to form. The random networks converged faster than the scale-free networks, because on average more clusters were formed in the scale-free networks.

## 4 Conclusions and Discussion

In this paper an agent-based social decision making model based on social contagion with a dynamic network is proposed. In contrast to the existing models [2, 4, 7, 8], similarity in agent states or opinions has a dynamic effect on the strengths of the links between agents: they change gradually over time, rather than that a more static effect

**Table 6.** The mean convergence time (in simulation time points) of 10 random network and 10 scale-free networks

| Parameter settings | β=0.001 | | | β=0.0025 | | | β=0.01 | | |
|---|---|---|---|---|---|---|---|---|---|
| | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 | α=1 | α=10 | α=20 |
| **Scale-free *M1*** | 68.3 | 83.5 | 76.7 | 56.7 | 59.6 | 51.8 | 20.2 | 18.2 | 17.3 |
| **Scale-free *M2*** | 12.7 | | | 12.8 | | | 10.2 | | |
| **Random *M1*** | 44 | 38.9 | 85.6 | 23.8 | 27 | 28.8 | 16.9 | 12.1 | 10.4 |
| **Random *M2*** | 12 | | | 11.9 | | | 8.4 | | |

is used based on a threshold mechanism. The model was analysed analytically and by simulation. Cluster formation has been extensively investigated in the paper for the models with gradually and abruptly changing links: the distribution of clusters in random and scale-free networks was investigated, the dynamics of links within and between clusters were determined, the minimal distance between two clusters was identified, and the convergence speed was analysed.

In the paper simulation results were presented for the networks with the average node degree 4.5. However, also experiments for more dense networks were performed. In general, the higher the density of a network, the higher its convergence speed to an equilibrium state and the less number of clusters are formed.

Besides the parallel mode of interaction we also performed simulation for the sequential interaction mode. In the latter mode two randomly chosen agents interacted in each iteration. Small, medium and large clusters emerged in this case as well, however their numbers were lower than in the parallel case, and the convergence speed was higher.

Simulations were also performed for model variants, in which agents exchanged opinions on multiple topics at the same time. For these models two alternatives exist: 1) to introduce a separate degree of influence for each topic, or 2) to use one degree of influence for all topics. Also clusters can be considered for each topic separately or by introducing a similarity measure of the agents combining all the topics (e.g., based on the Euclidean distance). In all these cases small, medium and large clusters emerged. However, the convergence speed was significantly slower than in the case with one topic.

An important future step is validation of the proposed model. To this end data from a news discussion web site gathered for several years will be used. On this site users vote for news and for each other's comments to the news (like/dislike). Voting for news would reflect the opinions of the agents in our model and voting for comments would be related to the degrees of influence between the agents. The proposed model will be used to predict the dynamics of the network and formation of clusters over time.

## References


[1] R. Axelrod. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. Journal of Conflict Resolution 41: 2023-226, 1997.

[2] V.D. Blondel, J. M. Hendrickx, and J.N. Tsitsiklis. On Krause's consensus formation model with state-dependent connectivity. IEEE Trans. on Automatic Control, vol. 54(11), pp. 2506-2517, 2009.

[3] D. Byrne, The attraction hypothesis: Do similar attitudes affect anything? Journal of Personality and Social Psychology, 51(6), 1986, 1167-1170.

[4] G. Deffuant, D. Neau, F. Amblard, and G.Weisbuch. Mixing beliefs among interacting agents. Adv Complex Syst, 3. pp. 87-98, 2000.

[5] J.R., French. A Formal Theory of Social Power. Irvington Publishers, 1993.

[6] J. Golbeck. Computing and Applying Trust in Web-based Social Networks. PhD Thesis, University of Maryland, 2005.

[7] M. Granovetter. Threshold Models of Collective Behavior. American Journal of Sociology, 83(6):1420:1443, 1978.

[8] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: models, analysis and simulation, J.of Artificial Societies and Social Simulation 5, 3, 2002.

[9] M. Hoogendoorn, J. Treur, C.N. van der Wal, and A. van Wissen. An Agent-Based Model for the Interplay of Information and Emotion in Social Diffusion. In Proceedings of IAT'10. IEEE Computer Society Press, 439-444, 2010.

[10] K. Lewin. Group Decision and Social Change. New York: Holt, Rinehart and Winston, 1958.

[11] M. McPherson, L. Smith-Lovin, J. M. Cook (2001). Birds of a feather: homophily in social networks. Annu. Rev. Sociol. 27, 415–444 (2001).

[12] M. Macy, J. A. Kitts, and A. Flache. Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure. In Dynamic Social Network Modeling and Analysis. Washington, DC: National Academies Press, 162-173, 2003.

[13] Lev Muchnik, Royi Itzhack, Sorin Solomon, and Yoram Louzoun, Self-emergence of knowledge trees: Extraction of the Wikipedia hierarchies, Phys. Rev. E 76, 016106 (2007)

[14] H.V.D. Parunak, T.C. Belding, R. Hilscher, and S. Brueckner. Modeling and Managing Collective Cognitive Convergence. In Proceedings of AAMAS 2008, pp. 1505-1508, ACM Press, 2008.

[15] O. Reingold. Undirected connectivity in log-space. Journal of the ACM 55 (4), pp. 17- 24, 2008

[16] A. Sharpanskykh. Managing the Complexity of Large-Scale Agent-based Social Diffusion Models with Different Network Topologies. In Proceedings of the 12th Pacific Rim International Conference on Artificial Intelligence (PRICAI'12), Springer Verlag, Lecture Notes in Computer Science, Vol. 7458, 2012, pp. 540-551.

[17] A. Sharpanskykh, J. Treur. Group Abstraction for Large-Scale Agent-Based Social Diffusion Models. In: Zhan, J., et al. (eds.), Proceedings of the Third International Conference on Social Computing, SocialCom'11. IEEE Computer Society Press, 2011, pp.830-837.