Group Abstraction for Large-Scale Agent-Based Social Diffusion Models

Alexei Sharpanskykh
Artificial Intelligence Department
VU University Amsterdam
Amsterdam, the Netherlands
sharp@cs.vu.nl

Jan Treur
Artificial Intelligence Department
VU University Amsterdam
Amsterdam, the Netherlands
treur@cs.vu.nl

Abstract— **In this paper an approach is proposed to handle complex dynamics of large-scale multi-agents systems modelling social diffusion processes. Based on local properties of the individual agents and their connections, groups and dynamic properties of these groups are identified. To determine such dynamic group properties two abstraction methods are proposed: determining a group invariant and approximation of group processes by weighted averaging of interactions. This enables simulation of the multi-agent system at a more abstract level by considering groups as single entities substituting a large number of interacting agents. In this way the scalability of large-scale simulation can be improved significantly. Computational properties of the developed approach are addressed in the paper. The approach is illustrated for a collective decision making model.**

Keywords – social contagion, group abstraction, agent-based simulation, large-scale multi-agent systems

I. INTRODUCTION

Social diffusion models describe spread and changes of attitudes in a group or community of agents under the impact of social interaction. Such models have been extensively used to study diverse social processes: such as dynamics of social power [3], polarization of opinions of individuals in a group [8,2], and spread of innovation [11]. Although the local behaviors of an agent may be simple, global patterns that emerge from interaction between the agents in large-scale social diffusion systems are far from trivial. Such patterns are difficult to infer directly from the local dynamic properties of the agents. A high computational complexity of large-scale multi-agent systems hinders automated analysis of such systems by simulation and verification.

In this paper an approach is proposed to handle complex dynamics of large-scale agent-based social diffusion models by using abstraction. The approach is based on identifying groups of interacting agents with similar states (e.g., opinions on an issue) in a society of agents. The idea is that an approximate form of simulation is obtained by using such groups as single entities representing abstractions of large numbers of interacting agents. In such a way the scalability of large-scale multi-agent simulation can be improved. The obtained abstracted process provides an approximation with a behavioural error that can be estimated.

To determine global emerging dynamic properties of groups based on local dynamic properties of the interacting group members, two group abstraction methods are proposed. In the first method relative degrees of importance of the agents in a group are determined. The degree of importance of an agent is in proportion to the strength of the agent's influence on the group. The aggregated state of the group is determined by the weighted average of the states of the group members with the weights defined by the relative degrees of importance of the members. In the paper this method is called *abstraction by weighted averaging*.

The second group abstraction method is based on identifying an invariant for a state of a group: an expression in terms of states of group members that does not change over time, as long as no input from outside the group occurs. The existence of an invariant can be considered to provide a kind of preservation law for the (collective) support for an opinion in a group. An invariant determines a state of a group and how this state depends on the states of the group members. In the paper the abstraction method based on considering an invariant instead of the individual states is called *invariant-based abstraction*.

The proposed group abstraction approaches are illustrated by a case of a collective decision making model for scenarios, which may exist in real social systems. The approximation errors and time complexity of the proposed abstraction methods applied for this case were evaluated.

The paper is organized as follows. An agent-based collective decision making model is described in Section II. The proposed methods for group abstraction are explained in Section III. Some simulation results are discussed in Section IV. In Section V the proposed abstraction methods applied to the collective decision making model are evaluated. Section VI concludes the paper.

II. THE COMPUTATIONAL MODEL

In the computational model collective decision making is modeled as a process of social diffusion of opinions of agents on decision options. Without loss of generality the agents are assumed to consider two different decision options *s1* and *s2* for one issue (e.g., two exits of a burning building).

In most existing social diffusion models (e.g., [8,10,4]), opinions of agents are represented by binary variables, which reflect the opposite attitudes of an agent towards an issue. The choice for binary variables is well motivated for models, which focus on attitudes of agents towards highly salient events, for which strong opinions are common (e.g., in voting). However continuous variables are suited better than binary variables for representing doubts of agents, e.g., when they are situated in uncertain environments with scarce information. Furthermore, the change of the agent's opinion to the opposite one occurs gradually, through a number of phases [7]. This can be captured better by a continuous variable than by a binary variable. In contrast to these models, the opinions of an agent in the model used here are described by continuous variables within the range [*0,1*]. These variables reflect the degrees of support of an agent for the decision options *s1* and *s2*.

The initial values of the opinions of the agents on both options are uniformly distributed in the interval [*0,1*]. By interaction the agents start to influence each other's opinions. The strength of social influence of an agent *i* on another agent *j* is determined by parameter $\gamma_{i,j}$ within the range [*0, 1*]. This parameter may be refined, e.g., by distinguishing expressiveness of the sender (agent *i*), openness of the receiver (agent *j*), the strength of the channel between *i* and *j* [5]. This parameter may also refer to a distance between *i* and *j* in 'social' space. For simplicity $\gamma_{i,j}$ will be used without refinement.

Two modes of interaction of agents are considered:

- *parallel mode*, in which agents interact synchronously with each other, and all states of the agents are updated in parallel, e.g., as in [10];
- *sequential mode*, in which at each time point at most one agent is chosen at random to interact with its neighbours, e.g., as in [6].

Take $q_{s,j}(t)$ to denote the strength of support of agent *j* of decision option *s*. The strength of the social influence of the other group members *j* on agent *i* with respect to decision option *s* at time *t* is determined by:

$$\delta_{s,i}(t) = \sum_{j \neq i} \gamma_{j,i}(q_{s,j}(t) - q_{s,i}(t)) / \sum_{j \neq i} \gamma_{j,i} \qquad \text{when } \sum_{k \neq i} \gamma_{k,i} \neq 0$$

and

$$\delta_{s,i}(t) = 0 \qquad \text{when } \sum_{k \neq i} \gamma_{k,i} = 0$$

The update of the strength of support of agent *i* for *s* is determined by:

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

Here η_i is an agent-dependent (openness) parameter within the range [*0,1*], which determines how fast the agent adjusts to the opinion of other agents.

In the parallel mode the states of all agents are updated at every time point, whereas in the sequential mode the states of one randomly chosen agent is updated at every time point.

First an initial consolidation phase takes place during the interval [*0*, t_{end_init}], in which the agents exchange opinions on the options. After this phase the whole population of agents is divided into two groups *G1* and *G2* depending on which from two options *s1* or *s2* is preferred:

$$G1 = \{i \mid q_{s1,i}(t_{end\_init}) \geq q_{s2,i}(t_{end\_init})\}$$

$$G2 = \{i \mid q_{s2,i}(t_{end\_init}) > q_{s1,i}(t_{end\_init})\}$$

Each group can be viewed as a connected directed graph $G=<V, E>$ with a set of vertices *V* representing agents and a set of directed edges *E* representing influence relations between the agents. It is assumed that there are less interactions between members of different groups than within the groups. This assumption is partially supported by social studies [2].

In the paper scenarios will be addressed based on the following topologies of groups (see Figure 1).

Definition

(a) A subset *S* of *G* is called *isolated from impact by others* if it is nonempty and not *G* and for all group members $i, j \in G$ with $i \in S$ and $j \notin S$ it holds $\gamma_{j,i} = 0$. Intuitively, the agents in set *S* are not influenced by any agent outside set *S*.

(b) A group *G* is called *cohesive* if no subset *S* of *G* is isolated from impact by others. Thus, in a cohesive group each agent is necessarily influenced by at least one another agent. However, there are could be agents in a cohesive group, which do not influence any other agents.

(c) A group *G* is called *fully connected* when $\gamma_{j,i} > 0$ for all group members $i, j \in G$ with $i \neq j$. In such a group each agent influences at least one another agent and is influenced by at least one another agent.

Note the following:

- Every fully connected group is cohesive, but not conversely
- Applied to a singleton subset $S = \{i\}$ for any *i* the cohesiveness criterion implies that $\sum_{k \neq i} \gamma_{k,i} > 0$.

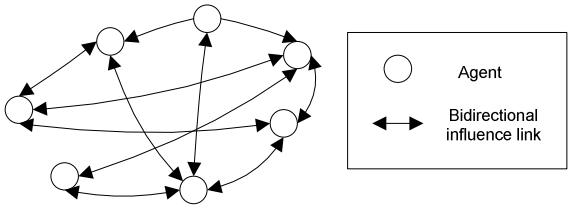

Figure 1. An example of a group topology considered in the paper

For cohesive groups the following theorem on equilibria holds.

Theorem

A cohesive group is in an equilibrium state if and only if all members have equal values.

This can be shown as follows. From the equations it is immediately clear that any state in which for any *s* the $q_{s,i}(t)$ are equal for all *i* is an equilibrium. Conversely, for a cohesive

group it can easily be shown that *all* equilibria have equal values for all of the members. First note that

$$dq_{s,i}(t)/dt \geq 0 \text{ iff } \sum_{j \neq i} (\gamma_{j,i} / \sum_{k \neq i} \gamma_{k,i}) q_{s,j}(t) \geq q_{s,i}(t)$$

In particular, for i with lowest $q_{s,i}(t)$ (i.e., $q_{s,i}(t) \leq q_{s,j}(t)$ for all j) it follows that $q_{s,i}(t)$ is monotonically increasing at t (similarly a highest $q_{s,i}(t)$ is monotonically decreasing). This can include nonstrict monotonicity: remaining equal. However, suppose not all values $q_{s,j}(t)$ are equal; due to the cohesiveness condition, within the set S_{lowest} of agents i with lowest $q_{s,i}(t)$ at least one agent $i \in S_{lowest}$ exists that is affected by an agent $m \notin S_{lowest}$ with nonlowest value, i.e., with $\gamma_{m,i} > 0$. Then by

$$\sum_{j \neq i} (\gamma_{j,i} / \sum_{k \neq i} \gamma_{k,i}) q_{s,j}(t) > q_{s,i}(t)$$

it follows that strict monotonicity occurs. From this it follows that no equilibria can occur when the values are not all equal. Therefore the only equilibria are when all values are equal. Note that this is also implied by general mathematical theorems as described, for example, in [9].

The abstraction approaches introduced apply to cohesive groups, but as illustration the fully connected case is used in the simulations.

Usually groups do not function in isolation from agents outside the group. Every now and then members of a group receive messages from diverse external sources via peer-to-peer communication. The degree of influence of external source k (e.g., an agent from another group) on a group member i is represented by the already introduced parameter $\gamma_{k,i}$. When impact from other agents is assumed not present, based on the states of k and i concerning option s, agent i updates its state as follows:

$$q_{s,i}(t+\Delta t) = q_{s,i}(t) + \eta_i \gamma_{j,i}(q_{s,k}(t) - q_{s,i}(t))\Delta t$$

If after interaction with an external source, agent i from group *G1* supporting option *s1* changes its preference from *s1* to *s2*, and $q_{s2,i}(t) - q_{s1,i}(t) > threshold$, then an agent is considered to leave *G1* and become a member of *G2* supporting *s2*. In the scenarios considered in the paper *threshold=0.3*. By sensitivity analysis similar outcomes for all *threshold* values from *(0,1)* were obtained.

III. Methods for Group Abstraction

To model abstracted states of a group two group abstraction methods are proposed in this section: abstration based on weighted averaging, and invariant-based abstraction.

A. A Method based on Weighted Averaging

The first method is based on an estimation of an aggregated group state by determining the contribution of each group member to this state as follows. It is assumed that the contribution of an agent is in proportion to the strength of influence of the agent on the group. An agent may influence another agent directly or indirectly through other agents. In the direct case the strength of influence of i on j is determined by $\gamma_{i,j}$. If i influences k through j, the strength of indirect influence of i on k via j is determined by $\gamma_{i,j}\, \gamma_{j,k}$, and in total by

$$\Sigma_{j \neq i, j \neq k} \gamma_{i,j}\, \gamma_{j,k}$$

In the general case, the strength of influence of an agent on any other agent via an arbitrary number of mediating agents (hops) can be calculated recursively. Thus, for each agent a network of influence can be identified, through which an agent exerts influence and is influenced by other agents. In such a network the strength of influence of an agent i (soi_i) is calculated as follows:

$$soi_i = \sum_{i \neq j} \gamma_{i,j}(1 + \sum_{k \neq i,\, k \neq j} \gamma_{j,k}(1+\ldots)) / (1 + \sum_{i \neq j} \gamma_{j,i}(1 + \sum_{k \neq i,\, k \neq j} \gamma_{k,j}(1+\ldots)))$$

The denominator contains the term *1* to ensure that it is not equal to *0* for the agents isolated from impact by others. The precision of estimation of the group state depends on the number of hops in a network of influence for which indirect influences are calculated. However, the more hops are taken the more intensive computation is required for abstraction by this method. In this paper two hops in a network of influence are used. In the single-hop variant of the method (called *first-order weighted averaging*) soi_i is calculated as:

$$soi_i = \sum_{i \neq j} \gamma_{i,j} / (1 + \sum_{i \neq j} \gamma_{j,i})$$

For the two-hop variant (called *second-order weighted averaging*) soi_i is:

$$soi_i = \sum_{i \neq j} \gamma_{i,j}(1 + \sum_{k \neq i, k \neq j} \gamma_{j,k}) / (1 + \sum_{i \neq j} \gamma_{j,i}(1 + \sum_{k \neq i,\, k \neq j} \gamma_{k,j}))$$

Initially and after each interaction of an agent from group G with an external agent, the aggregated state of group G for option s is estimated by the following weighted average:

$$q_{s,G}(t+\Delta t) = q_{s,G}(t) + \nu \left[\sum_{i \in G} soi_i\, q_{s,i}(t) / \sum_{i \in G} soi_i - q_{s,G}(t) \right] \Delta t$$

This state $q_{s,G}(t)$ represents a common opinion of all agents in the group for decision option s. It persists until a new interaction with an external agent occurs. Then, the formula for $q_{s,G}(t+\Delta t)$ is applied again. The parameter ν is a speed parameter which represents how fast such an update takes place.

B. An Invariant-Based Method

In this section it is discussed how an invariant can be found and used to serve as an aggregate group state and to determine equilibrium values.

For given initial values $q_{s,i}(0)$ for $i = 1, ..,$ only one of the possible equilibria with equal values will be actually reached both in the parallel and sequential modes of interaction (see Figure 2). How this equilibrium value depends on the initial values can be described by an *invariant*: an expression in terms of the $q_{s,i}(t)$ for $i = 1, ..,$ that does not change over time. In this case an invariant inv_s as a weighted sum

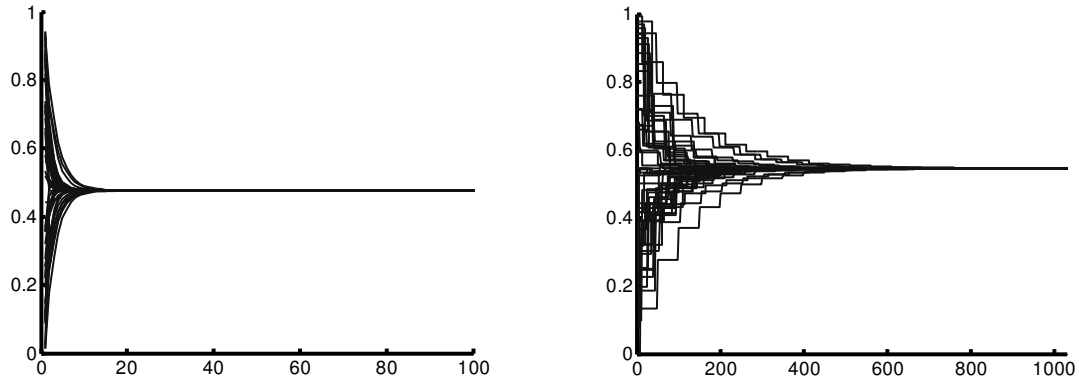

Figure 2. Convergence of the decision states of 50 agents in a group with the parallel (left) and sequential (right) modes of interaction; no external messages are provided to the group. The horizontal axis is time and the vertical axis is the degree of support for the first decision option

$$inv_s = inv_s(t) = \sum_i \lambda_{s,i}\, q_{s,i}(t)$$

can be obtained where the weights $\lambda_{s,i}$ depend on the coefficients η_i and $\gamma_{j,i}$ (and not on initial values).

These weights can be taken (normalised) with $\sum_i \lambda_{s,i} = 1$, so that when all $q_{s,i}(t) = 1$ for all i, also the invariant is 1. Below it will be discussed how an invariant can be found.

Using an invariant to determine an equilibrium

Given an invariant, for the cohesive case an equilibrium can be determined easily from this invariant and the initial values as follows (taking into account that the equilibrium concerns equal values $q_{s,i}(t) = q_s$):

$$\sum_i \lambda_{s,i}\, q_{s,i}(0) \;= inv_s \;= \sum_i \lambda_{s,i}\, q_{s,i}(t) = \; q_s \sum_i \lambda_{s,i} = \; q_s$$

Therefore the equilibrium value is obtained as a weighted sum of the initial values:

$$q_s = \sum_i \lambda_{s,i}\, q_{s,i}(0)$$

Note that agents i with higher $\lambda_{s,i}$ contribute more to the group's equilibrium. This shows that the coefficients $\lambda_{s,i}$ of the invariant can be interpreted as the relative importance of agent i for the group's collective support of s. Note that this shows that the equilibria depend on the initial values, and therefore cannot be determined by solving equilibrium equations obtained from the differential equations. Indeed these equilibrium equations obtained from the differential equations are

$$\eta_i\, \delta_{s,i}(t) = 0 \quad \text{where} \quad \delta_{s,i}(t) = \sum_{j \neq i} \gamma_{j,i}(q_{s,j}(t) - q_{s,i}(t)) / \sum_{j \neq i} \gamma_{j,i}$$

These equations are fulfilled always when all values $q_{s,i}(t)$ are equal, but do not give any information about a specific value, and how that depends on the initial values. An invariant provides exactly this connection to the initial values.

Using an invariant for group abstraction

An invariant can be considered as a kind of preservation law for the (collective) support for option s in the considered group. By internal (intragroup) interactions this collective support for s can be redistributed over persons in the group, but this does not change the collective amount. During time intervals where no external interaction is coming in, the group's collective support for s will not change. This provides an interesting use of the invariant: as a means of abstraction from the internal processes by using a descriptor at the group level. In contrast, incoming external (intergroup) interaction can change the collective support of the group for s. This can be done as follows. Suppose agent i with level $q_{s,i}(t)$ within the group gets incoming interaction from agent j external from the group, with level $q_{s,j}(t)$. Then the value of agent i's level is updated as follows:

$$q_{s,i}(t+\Delta t) = q_{s,i}(t) + \eta_i\, \gamma_{j,i}\, (q_{s,j}(t) - q_{s,i}(t))\, \Delta t$$

Accordingly the group's invariant inv_s is updated as follows:

$$inv_s(t+\Delta t) = inv_s(t) + \lambda_{s,i}\, \eta_i\, \gamma_{j,i}\, (q_{s,j}(t) - q_{s,i}(t))\, \Delta t$$

Note that for the general case in the time intervals where no external interaction comes in, internal interactions take place that may lead to an equilibrium. The internal group dynamics is assumed to take place with higher frequencies than external interactions, so that each time a group equilibrium is approximated before a next external interaction occurs.

Determining an invariant

The weights $\lambda_{s,i}$ for the invariant inv_s can be determined from the differential equations: from

$$\sum_i \lambda_{s,i}\, q_{s,i}(t+\Delta t) = \sum_i \lambda_{s,i}\, q_{s,i}(t)$$

it follows

$$\sum_i \lambda_{s,i}\, \eta_i\, [\, \sum_{j \neq i} \gamma_{j,i}(q_{s,j}(t) - q_{s,i}(t)) / \sum_{k \neq i} \gamma_{k,i}\,] = 0$$

This can be rewritten as:

$$\sum_i \lambda_{s,i}\ \eta_i \sum_{j\neq i} \gamma_{j,i} q_{s,j}(t)/\sum_{k\neq i} \gamma_{k,i} = \sum_i \lambda_{s,i}\eta_i \sum_{j\neq i} \gamma_{j,i}\ q_{s,i}(t)/\sum_{k\neq i} \gamma_{k,i}$$

$$\sum_i \sum_{j\neq i} \lambda_{s,i}\ \eta_i\ \gamma_{j,i} q_{s,j}(t)/\sum_{k\neq i} \gamma_{k,i} = \sum_i \lambda_{s,i}\eta_i\ q_{s,i}(t)$$

$$\sum_j \sum_{i\neq j} \lambda_{s,i}\ \eta_i\ \gamma_{j,i} q_{s,j}(t)/\sum_{k\neq i} \gamma_{k,i} = \sum_i \lambda_{s,i}\eta_i\ q_{s,i}(t)$$

$$\sum_m \sum_{i\neq m} \lambda_{s,i}\ \eta_i\ \gamma_{m,i} q_{s,m}(t)/\sum_{k\neq i} \gamma_{k,i} = \sum_m \lambda_{s,m}\eta_m\ q_{s,m}(t)$$

$$\sum_m \left(\sum_{i\neq m} \left(\lambda_{s,i}\ \eta_i\ \gamma_{m,i}/\sum_{k\neq i} \gamma_{k,i}\right)\right) q_{s,m}(t) = \sum_m \lambda_{s,m}\eta_m\ q_{s,m}(t)$$

One way to satisfy this is by taking the coefficients of $q_{s,m}(t)$ in the above expression on both sides equal; this provides the following set of linear equations for the $\lambda_{s,i}$ for all m:

$$\sum_{i\neq m} \left(\eta_i\ \gamma_{m,i}/\sum_{k\neq i} \gamma_{k,i}\right) \lambda_{s,i} = \eta_m \lambda_{s,m}$$

Taking into account that $\sum_i \lambda_{s,i} = 1$ this can be rewritten into:

$$\sum_{i\neq m} \left(\eta_i\ \gamma_{m,i}/\sum_{k\neq i} \gamma_{k,i}\right) \lambda_{s,i} = \eta_m (1 - \sum_{i\neq m} \lambda_{s,i})$$

$$\sum_{i\neq m} \left(\eta_i\ \gamma_{m,i}/\sum_{k\neq i} \gamma_{k,i}\right) \lambda_{s,i} + \eta_m \sum_{i\neq m} \lambda_{s,i} = \eta_m$$

$$\sum_{i\neq m} \left(\eta_i\ \gamma_{m,i}/\sum_{k\neq i} \gamma_{k,i}\right) + \eta_m \Big) \lambda_{s,i} = \eta_m$$

$$\sum_{i\neq m} \left(\eta_i\ \gamma_{m,i}/\eta_m \sum_{k\neq i} \gamma_{k,i}\right) + 1 \Big) \lambda_{s,i} = 1$$

Thus a system of linear equations

$$\sum_{i\neq m} \mu_{m.i}\ \lambda_{s,i} = 1$$

is found with coefficients

$$\mu_{m.i} = \left(\eta_i\ \gamma_{m,i}/\eta_m \sum_{k\neq i} \gamma_{k,i}\right) + 1 \geq 1$$

This system can be described in matrix form as

$$\mathbf{A}\boldsymbol{\lambda_s} = \boldsymbol{1}$$

where $\boldsymbol{1}$ is the vector with all components 1,

$$\boldsymbol{\lambda_s} = (\lambda_{s,1}, \ldots)$$

and **A** is a square matrix with only zeros at the diagonal and all other entries ≥ 1 (expressed in η_i and $\gamma_{m.i}$). When it is assumed that the determinant $\mathbf{det(A)} \neq 0$, then this system has a unique solution. Indeed, for the general case this condition is fulfilled, and the weights $\lambda_{s,i}$ of the invariant can be obtained as a solution; this is illustrated in Sections IV and V.

IV. Simulation

The methods described in Section III were implemented in Matlab. Simulation time was 1030 with the initial stabilization interval [0, 30]. For each simulation setting 50 iterations were executed. The number of agents was varied across simulation runs: 50, 100, 200, and 500. The initial states of each agent for the strengths of support for the two decision options *s1* and *s2* were uniformly distributed in the interval [0,1].

In addition to the agents external sources were used, which number was 10 times less than the number of agents. The average time between two subsequent messages provided by each external source to a randomly chosen agent was varied across simulation runs: 1, 2, 5, and 10.

Each average time value can also be interpreted as a ratio of the time scale of the group's internal dynamics to the time scale of the external dynamics. The impact of these ratios on approximation errors was investigated. The parameters γ and η of each agent were taken from the uniform distribution in the interval]0,1]. Moreover, $\Delta t = 1.25$ and $\nu = 0.8$.

In the simulation the first and second-order weighted averaging methods and the invariant-based method were used for abstraction of the model. The simulation was performed both for the parallel (Figure 3) and sequential (Figure 4) interaction modes of agents. The peaks in the graphs indicate incoming messages from external sources. In the parallel interaction mode, after receiving each message the group quickly reaches a new stable state (Figure 3). In the sequential mode the relaxation time is much longer (Figure 4). Because of this, the approximation of the group dynamics by the developed abstraction techniques is less precise in comparison with the parallel case. A detailed evaluation of efficiency and quality of the proposed abstraction methods is considered in the following Section V.

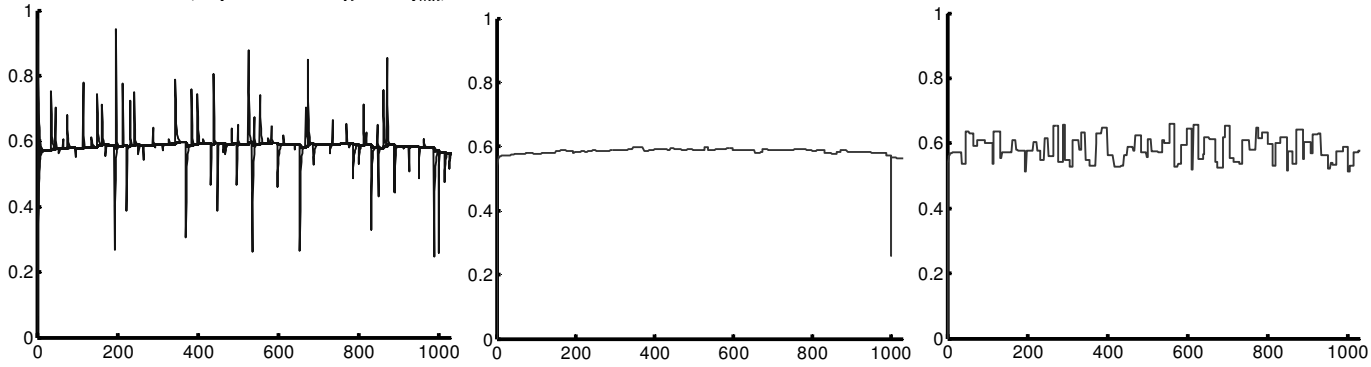

Figure 3. The dynamics of valuation of option 1 by 50 agents in a group in the parallel interaction mode (left) and the abstraction of the group dynamics obtained by the invariant-based method (center) and by the 2nd order weighted averaging (right); the average time between messages is 10. The horizontal axis is time and the vertical axis is the degree of support for option 1.

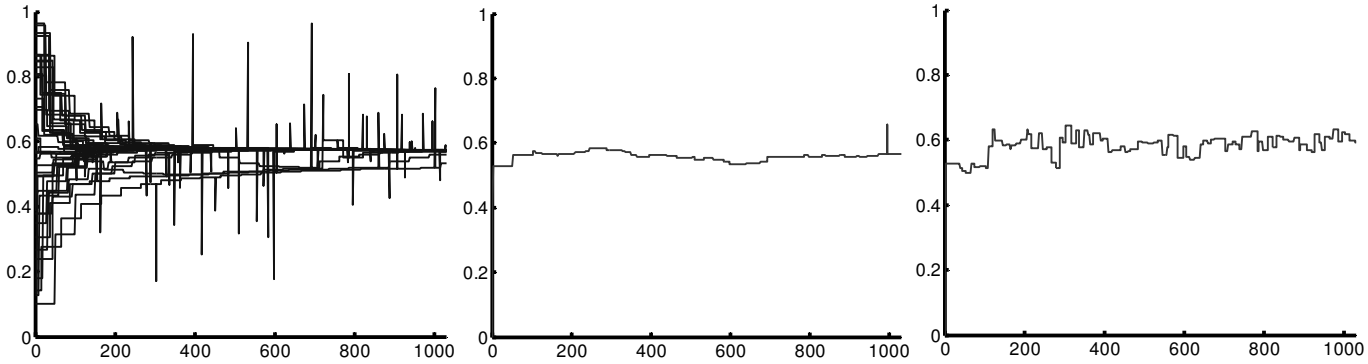

Figure 4. The dynamics of valuation of option 1 by 50 agents in a group in the sequential interaction mode (left) and the abstraction of the group dynamics obtained by the invariant-based method (center) and by the 2nd order weighted averaging (right); the average time between messages is 10. The horizontal axis is time and the vertical axis is the degree of support for option 1.

Table I. Mean simulation time in seconds for the original and abstracted models

# of agents	**200**				**500**			
Average time between messages	**1**	**2**	**5**	**10**	**1**	**2**	**5**	**10**
Original model (parallel)	96.44	93.49	87.9	82.72	383.7	383.2	365.3	350.8
Original model (sequential)	0.67	0.63	0.59	0.58	2.14	1.88	1.12	1.03
Invariant-based abstraction	4.33	3.8	3.37	3.16	15.7	14.2	12.5	11.8
Abstraction by 1st order weighted averaging	3.32	3.18	3.05	3.01	12.9	12.3	11.9	11.1
Abstraction by 2nd order weighted averaging	3.71	3.39	3.14	3.06	14.5	13.4	12.4	11.7

Table II. Mean simulation time in seconds for the original and abstracted models

# of agents	**50**				**100**			
Average time between messages	**1**	**2**	**5**	**10**	**1**	**2**	**5**	**10**
Original model (parallel)	5.87	5.46	4.98	4.72	23.67	22.75	20.55	19.25
Original model (sequential)	0.18	0.14	0.11	0.08	0.31	0.27	0.17	0.11
Invariant-based abstraction	0.30	0.26	0.24	0.23	1.05	0.95	0.85	0.81
Abstraction by 1st order weighted averaging	0.25	0.22	0.20	0.20	0.88	0.83	0.78	0.76
Abstraction by 2nd order weighted averaging	0.27	0.24	0.21	0.20	0.96	0.88	0.80	0.78

V. Evaluation of the Abstraction Methods

In this section the time complexity and approximation errors of the methods are considered.

A. *Time Complexity Results*

The mean time complexity for the original model from Section II and for the proposed abstraction methods is provided in Tables I and II. The variances of these results are very low (of the order of 10^{-5}).

B. *Time Complexity Results*

The mean time complexity for the original model from Section II and for the proposed abstraction methods is provided in Tables I and II. The variances of these results are very low (of the order of 10^{-5}).

The developed abstraction methods increase the computational efficiency of the simulation of the parallel case of the original model significantly. The acceleration factor grows with the number of agents: for smaller numbers (*~50*) it is of the order *20* to *25*, for larger numbers (*~500*) it grows to the order of *25* to *31*. Note that the simulation time of the sequential variant of the model is low, and depends weakly on the number of agents. This is because the state of only one agent is updated at each time point in this case.

The fastest simulation models are obtained by the abstraction by first-order weighted averaging. The slowest are the models obtained by the invariant-based abstraction. However, as one can see from Tables I and II, the ratio of the simulation time of the slowest to the fastest abstraction method is less than *1.3* for all cases. The impact of the number of messages on the simulation time is stronger for the invariant-based method than for the weighted averaging methods. This is because (large) systems of linear equations need to be solved in the former method every time when the structure of a group changes. The greatest decrease of the acceleration rate for both methods for the settings considered in the paper is of the order of *1.4*.

The time ratio between first and second order weighted averaging is at most *1.1*. Thus, adding the second level in the network of influence for each agent does not result in a significant increase of simulation time, even for larger numbers of agents.

C. Approximation errors

The error of approximation of the original model by a group abstraction method is defined as

$$\sum_{t \in [31,\ 1031]} (|G1^{o,t} \cup G1^{a,t}| - |G1^{o,t} \cap G1^{a,t}|)/1000,$$

where $G1^{o,t}$ is the group comprising the agents supporting decision option *s1* at time point *t* according to the original model, and $G1^{a,t}$ is the group of the agents supporting *s1* at time point *t* according to the abstracted model.

A comparison of the mean approximation errors for the proposed abstraction methods is provided in Figures 5 and 6 (for 50 agents). The variances of the errors are low (of the order 10^{-6}); they are depicted by small error bars in Figures 5 and 6.

The invariant-based method outperforms the weighted averaging abstraction methods. In the parallel interaction mode (Figure 5), the error of the invariant-based abstraction is at most $3*10^{-4}$.

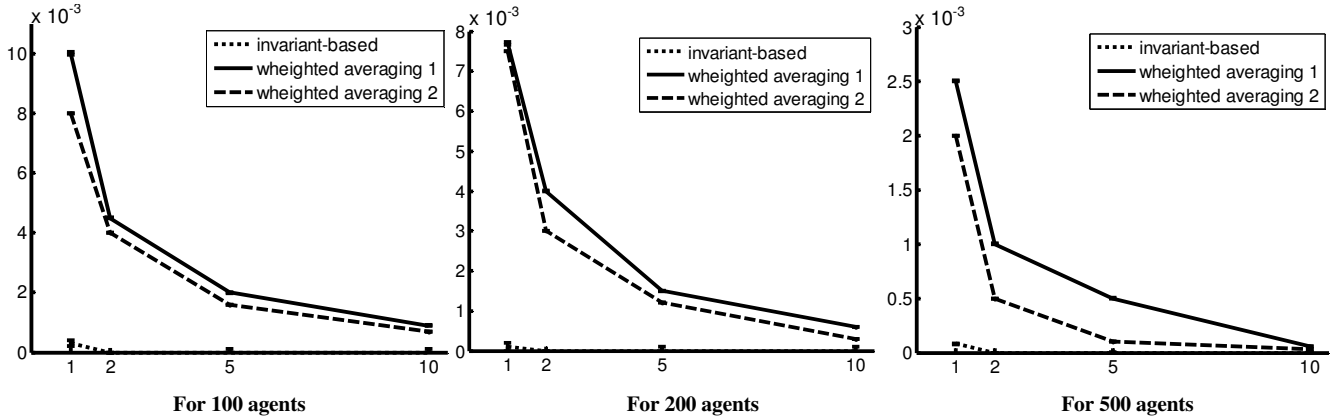

Figure 5. Mean approximation errors for the proposed abstraction methods for 100, 200 and 500 agents with the parallel interaction mode; the horizontal axis is the average time between messages.

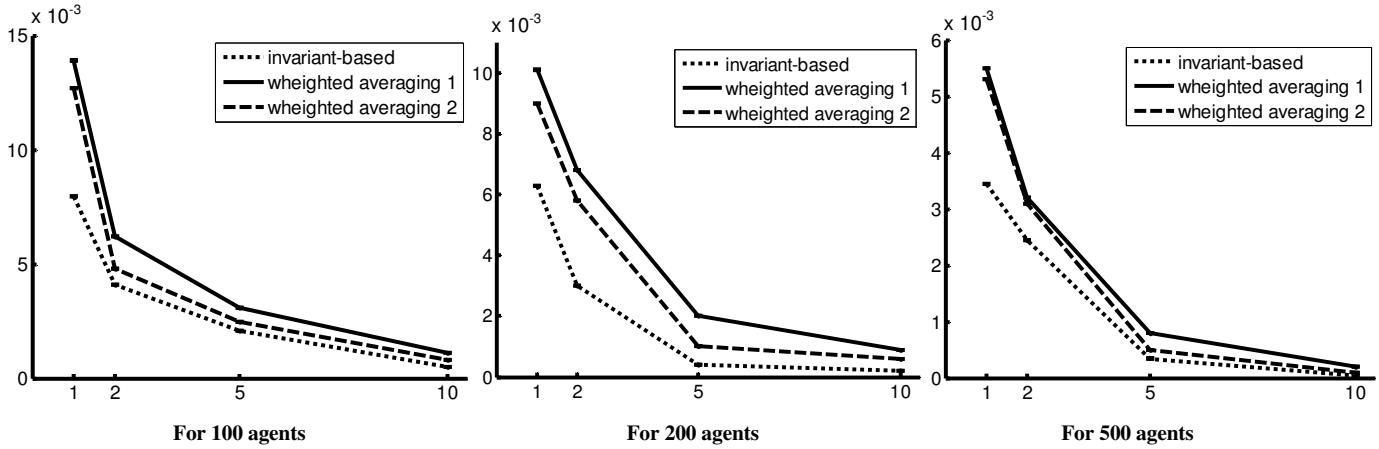

Figure 6. Mean approximation errors for the proposed abstraction methods for 100, 200 and 500 agents with the sequential interaction mode; the horizontal axis is the average time between messages.

In the sequential interaction mode the rate of convergence of the system is significantly slower than in the parallel mode (see Figure 2). This slower convergence rate in the presence of (frequent) external perturbations (i.e., messages from external sources) causes a high variance of the decision states of the agents in a group (see Figure 4). The higher the variance of the states of the agents in a group, the higher approximation errors of the abstraction methods proposed, as these methods rely on reaching a congruent stable state by the group. In contrast to the sequential case, in the parallel case groups reach a common stable state rapidly after every external disturbance (see Figure 3). Thus, the approximation errors of the abstraction by the proposed methods in the parallel case are significantly lower than in the sequential case (see Figures 5 and 6).

Both invariant-based and weighted averaging abstraction methods are sensitive to the average time between messages from external sources. The greater the time between messages, the more closely a group approaches its equilibrium state, thus the smaller the approximation error of the abstraction. When the average time is high (10), the error is the lowest. However, when the external world interacts with a group every time point, the errors of the abstraction methods grow, 15 times in the worst case.

As expected, the abstraction by the second-order weighted averaging is more precise than the abstraction by the first-order weighted averaging. The difference in precision between the weighted averaging methods depends on the density of connections in the topology of a group: in general, the higher the density, the less the error difference between both variants. This is because the density determines how many direct neighbors an agent has, and thus, how many agents are influenced directly by one-hop message propagation of new information. The more densely a graph is connected, the more agents in a group new information reaches by one-hope propagation, and the more fully the new group's state can be captured by the first-order weighted averaging. The less the graph's density, the more information about the group dynamics each additional hop provides. In a sparsely connected graph, one-hope propagation reaches only a small number of agents, thus, only partial information about the group dynamics can be extracted by the first-order weighted averaging. In this case the difference between the first- and second-order weighted averaging may be significant. For the experiments considered in this paper densely connected groups were used.

The abstraction by both the invariant-based method and the weighted averaging methods becomes more precise with the increase of the number of agents.

VI. CONCLUSIONS

In the paper an approach is proposed to handle complex dynamics of large-scale agent-based social diffusion models. On the one hand this approach allows identifying global, emergent properties of groups of agents. On the other hand, it enables a significant increase of the computational efficiency of automated analysis of large-scale social diffusion models (up to 31 times for larger numbers of agents (~500) in the case of parallel interaction of agents).

The approach comprises two methods dedicated for abstraction of the dynamics of social groups: invariant-based abstraction, (first- and second-order) weighted averaging. For both parallel and sequential interaction modes of agents the invariant-based abstraction method is the best choice. In the parallel interaction mode it has the approximation error close to *0*. The abstraction by the second-order weighted averaging is more precise than the abstraction by the first-order weighted averaging. In general, the higher the ratio of the time scale of the external world dynamics to the time scale of the group's internal dynamics, the less the approximation error of the abstraction methods proposed.

In the case of sequential interaction a group reaches an equilibrium state slowly. When the ratio of the time scale of the group's internal dynamics to the time scale of the external dynamics is high (10), then such a group is able to reach a congruent decision state. However, when the dynamics of such a group is disturbed frequently, the variance of the decision states of the agents becomes high. Because of this high variance, the proposed abstraction methods based on similarity of opinions of agents are less efficient and precise. In this situation, when the agents have constantly diverging opinions on the decision options, one may question the applicability of the concept of a group in general. Moreover, in [10] it is stated that the dynamics of group interaction is captured more accurately when many individuals are allowed to interact simultaneously than by sequential pairwise interaction.

Note that in many applications the sizes of dynamic groups, which could be numerous, are (much) smaller than the total number of agents. The developed abstraction techniques were applied in a large-scale crowd evacuation study (~10000 agents) [12]. Although the number of agents was significant, the maximal size of emergent dynamic groups was 174.

Social diffusion models have been studied extensively [2,3,10,8]. A common research question of these studies is about the existence of equilibrium states of a model for different topologies. In contrast to the continuous model considered in the paper, most of other studies consider binary, threshold-based models.

Currently several techniques for abstraction of models based on hybrid automata and differential equations [1] exist. However, such approaches can be applied efficiently for systems described by sparse matrixes. Social diffusion models represent tightly connected systems, which do not allow a significant reduction of the state space using such techniques. In particular, a previous study showed that common model reduction techniques such as balanced truncation [1] do not allow decreasing the rank of the matrix describing the model from Section II.

REFERENCES

[1] A.C. Antoulas and D. C. Sorensen. Approximation of large-scale dynamical systems: An overview. Int. J. Appl. Math. Comp. Sci., 11:1093-1121, 2001.

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

[3] John R. French. A Formal Theory of Social Power. Irvington Publishers, 1993.

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

[5] 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.

[6] B.A. Huberman and N.S. Glance. Evolutionary Games and Computer Simulations. PNAS, USA 90:7716-7718, 1993.

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

[8] Michael Macy, James A. Kitts, and Andreas 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.

[9] L. McKenzie. Matrices with Dominant Diagonals and Economic Theory. In Mathematical Methods for Social Sciences. Stanford University Press.

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

[11] E.M. Rogers. Diffusion of innovations. New York, Free Press, 2003.

[12] Sharpanskykh, A. and Zia, K. Grouping behaviour in AmI-enabled crowd evacuation. In Proceedings of the 2nd International Symposium on Ambient Intelligence, ISAmI'11, Springer Verlag, 233-240, 2011.