



Improving the Robustness of Online Social Networks: A Simulation Approach of Network Interventions

Giona Casiraghi and Frank Schweitzer*

Chair of Systems Design, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:

Carlos Gershenson,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Radoslaw Michalski,
Wroclaw University of Science and
Technology, Poland
Domenico Rosaci,
Mediterranea University of Reggio
Calabria, Italy

*Correspondence:

Frank Schweitzer
f Schweitzer@ethz.ch

Specialty section:

This article was submitted to
Computational Intelligence in
Robotics,
a section of the journal
Frontiers in Robotics and AI

Received: 01 November 2019

Accepted: 03 April 2020

Published: 28 April 2020

Citation:

Casiraghi G and Schweitzer F (2020)
Improving the Robustness of Online
Social Networks: A Simulation
Approach of Network Interventions.
Front. Robot. AI 7:57.
doi: 10.3389/frobt.2020.00057

Online social networks (OSN) are prime examples of socio-technical systems in which individuals interact via a technical platform. OSN are very volatile because users enter and exit and frequently change their interactions. This makes the robustness of such systems difficult to measure and to control. To quantify robustness, we propose a coreness value obtained from the directed interaction network. We study the emergence of large drop-out cascades of users leaving the OSN by means of an agent-based model. For agents, we define a utility function that depends on their relative reputation and their costs for interactions. The decision of agents to leave the OSN depends on this utility. Our aim is to prevent drop-out cascades by influencing specific agents with low utility. We identify strategies to control agents in the core and the periphery of the OSN such that drop-out cascades are significantly reduced, and the robustness of the OSN is increased.

Keywords: socio-technical system, adaptability, robustness, simulations, agent-based model

1. INTRODUCTION

Self-organization describes a *collective dynamics* resulting from the *local interactions* of a vast number of system elements (Schweitzer, 1997), denoted in the following as *agents*. The macroscopic properties that *emerge* on the system level are often desired, for example, coherent motion in swarms or functionality in gene regulatory networks. But as often these self-organized systemic properties are not desired, for example, traffic jams or mass panics in social systems. Hence, while self-organization can be a very useful dynamics, we need to find ways of controlling it such that systemic malfunction can be excluded, or at least mitigated. This refers to the bigger picture of *systems design* (Schweitzer, 2019): how can we influence systems in a way that optimal states can be achieved and inefficient or undesired states can be avoided?

In general, self-organizing processes can be controlled, or designed, in different ways. On the *macroscopic* or systemic level, *global control parameters*, like boundary conditions, can be adjusted such that phase transitions or regime shifts become impossible. This can be done more easily for physical or chemical systems, where temperature, pressure, chemical concentration, etc. can be fixed. On the *microscopic* or agent level, we have two ways of controlling systems: (i) by influencing agents directly, (ii) by controlling their interactions.

Referring to socio-economic systems, we could, for example, incentivize agents to prefer certain options, this way impacting their utility function. This requires to have access to agents, which is not always guaranteed. For instance, it is difficult to access prominent agents or to influence large multi-national companies. Controlling agents' interactions, on the other hand, basically means to restrict (or to enhance) their communication, i.e., their access to information and dissemination. Restrictions can be implemented both globally and locally.

In this paper, we address one particular instance of social systems, namely *online social networks* (OSN). Prominent examples for such networks are *facebook*, *reddit*, or *Twitter*. OSN are instances of a complex system comprising a large number of interacting agents which represent users of such networks. OSN are, in fact, *socio-technical systems* because they combine elements of a social system, i.e., users communicating, with elements of a technical system, i.e., platforms, protocols, GUI (graphical user interfaces), etc. The technical component is important because it allows to *control* the access to users, as well as their communication. The term *control* refers to the fact that access and interactions are *monitored*, but also *influenced* in different ways.

In reality, it becomes very difficult to control OSN because of their large *volatility*, which has two causes. The first one is the *entry and exit* dynamics, which impacts the number of *agents*: Users enter or leave the OSN at a high frequency. The second one is the *connectivity*, which impacts the number of *interactions*: Users easily connect to and disconnect from other users or interact with lower or higher frequency. They have ample ways of interacting; thus, it becomes very difficult to shield them from certain information.

Because of this volatility, in an OSN *interactions* cannot be fully controlled. But we can certainly influence users via their *utility function*. Users join an OSN for a certain purpose, namely to socialize and to exchange information. Hence, their benefits are a function of the number of other users they interact with. Their costs, on the other hand, result from the effort of maintaining their profile, learning about the features of the graphical user interface, etc. The utility, i.e., the difference between benefits and costs, can then be increased by either increasing the benefits, e.g., by increasing their number of friends, or by decreasing their costs, e.g., by automatizing profile updates, or by a combination of both.

OSN are a paradigm for the emergence of collective dynamics and are much studied because of this. For example, the emergence of trends, fashions, social norms, or opinions occurs as a self-organized process that can sometimes be initiated but hardly be controlled. A worrying trend emerges if users decide to *leave* the social network. If their decision causes other users to leave as well, because they lost their friends, this can quickly result in large drop-out cascades and in the total collapse of the OSN (Kairam et al., 2012). This happened, for example, to *friendster*, an OSN with about 117 million users in 2011. As studied in detail (Garcia et al., 2013), less integrated users left *friendster*, this way, making it less attractive to the remaining users to further stay on the platform.

To model such a self-organized dynamics by means of an agent-based model requires us to solve a number of methodological issues. On the *agent* level, we need to model individual decisions of agents based on their perceived utility, which is to be defined. On the *system* level, we need to quantify how the drop-out of individual agents impact other agents and the whole system, in the end (Jain and Krishna, 1998, 2002). In a volatile system, agents come and go at a large rate, without threatening the stability of the system every time. Hence, we need

to define a macroscopic measure that allows quantifying whether the system is still robust.

Once these methodological issues are solved, we can turn to the more interesting question of *systems design*. This means that, by using our agent-based model, we explore possibilities to influence the system such that it becomes more robust. Our focus will be on the *microscopic* level, i.e., influencing *agents* rather than whole systems. This is sometimes referred to as *mechanism design*. But, different from designing communication, i.e., influencing *interactions*, here we influence agents via their utility functions. This leads to another methodological problem, namely how to identify those agents that are worth to be influenced, i.e., are most promising for reaching a desired system state.

This problem is for networks addressed in the so-called *controllability theory* (Liu et al., 2011), which is very much related to control theory in engineering. It allows to quantify how much of a network is controlled by a given agent, which then can be used to *rank* agents with respect to their *control capacity* (Zhang et al., 2019). To apply this formal framework, however, requires to have a static network, i.e., the interaction *topology* should *not* change on the same time scale as the interaction. So, this framework does *not* allow us to study drop-out cascades in which the network topology changes at every time step. Because of this, in our paper, we have to rely on a *computational approach*, i.e., we use our agent-based model to simulate the decision of agents to leave the network and its impact on the remaining network, while monitoring the overall robustness of the system by means of a macroscopic measure.

With these considerations, we have already specified the structure of this paper. In section 2, we model the decisions of agents and quantify the robustness of the network. In section 3, we introduce a reputation dynamics that runs on the network, to determine the benefits of the agents. In section 4 we highlight the dynamics of the OSN without any interventions, to demonstrate its breakdown. In section 5, eventually, we use our model to explore different agent-based strategies of improving the robustness of the network.

2. ROBUSTNESS OF THE SOCIAL NETWORK

2.1. Agents and Interaction Networks

2.1.1. Networks

For our agent-based model of the OSN we use the specific representation of a *complex network*. The term *complex* refers to the fact that we have a large number of interacting agents such that new system properties can *emerge* as the result of these collective interactions. The term *network* means that agents are represented by *nodes*, and their interactions by *links* of the network. This implies that all interactions are decomposed into dyadic interactions between any two agents.

Using a mathematical language, networks are denoted as *graphs*, nodes as *vertices* and links as *edges*. We can then formally define a *graph* object \mathcal{G} as an ordered pair $\mathcal{G} = \mathcal{G}(V, E)$, where V is the set of vertices of the graph, and E is the set of edges. Vertex

$i \in V$ and $j \in V$ are connected if and only if $ij \in E$. The graph is not static but changes on a time scale T , i.e., $\mathcal{G}(T)$. We call T the *network time* because agents can enter or exit the OSN, this way changing both the number of vertices and edges.

Agents are characterized by a binary state variable $s_i(T) \in \{0, 1\}$, where $s_i(T) = 1$ means that agent i at time T decides to *stay* in the OSN, whereas $s_i(T) = 0$ means that it decides to *leave* the OSN. This decision is governed by a utility function $U_i(T)$:

$$s_i(T) := \Theta[U_i(T)] ; \quad U_i(T) = B_i(T) - C_i(T) \quad (1)$$

The Heaviside function $\Theta(x)$ returns 1 if $x \geq 0$ and 0 otherwise. $B_i(T)$ and $C_i(T)$ are the benefits and the costs of agent i at time T . Only if the benefits exceed the costs, agent i will stay in the OSN, otherwise it leaves. The two functions need to be further specified, which is done in section 3.

2.1.2. Interactions

We want to model an OSN; therefore, we consider *directed interactions* between agents. Taking the example of *Twitter*, a directed interaction $i \rightarrow j$ means that agent i is a follower of agent j . Obviously, the reverse does not need to apply but can be frequently observed. Each of these interactions is represented as a directed link in the network \mathcal{G} . A formal expression for the topology of a network with N agents is the *adjacency matrix* $\mathcal{A} \in \mathbb{N}^{N \times N}$ in which the elements a_{ij} are either 0 or 1. This allows to define the *in-degree* d_i^+ and the *out-degree* d_i^- of an agent $i \in V$ as the number of incoming or outgoing links of i . We can also define the *total degree* of agent i as the sum of both in- and out-degree, $d_i = d_i^+ + d_i^-$.

Various works have proposed methods for identifying groups of agents that are stable over time in OSNs. In particular, De Meo et al. (2017) have focused on evaluating the *compactness* of such groups, i.e., the homogeneity in terms of mutual agents' similarity within groups. The concept of compactness, originally introduced in Botafogo et al. (1992), is often used to describe the cohesion of parts of the internet, collaboration networks, and OSNs (Egghe and Rousseau, 2003). Differently from this approach, in this article we aim at characterizing the robustness of the whole network, irrespectively of the stability of specific groups therein. For this reason, we begin our analysis from macroscopic quantities that allow to readily investigate the properties of a complex networks.

The degree distribution is an important macroscopic quantity to characterize a complex network. It is known that OSN have a rather broad degree distribution (Garcia et al., 2013), i.e., many agents are linked to only a few other agents, while a few agents, called hubs, have very many incoming links from other agents. Additionally, OSN often show a so-called *core-periphery structure* (Borgatti and Everett, 2000), in which well-connected agents form a core, whereas agents with only a few, or even no, connections form the periphery. Identifying such structures helps to analyze the robustness of the network. Precisely, we can assume that the OSN is robust, despite an ongoing entry and exit of agents, if the core changes, but continues to exist. This implies that the volatile dynamics mostly affects the periphery. If, however, the drop-out of a few agents is amplified into a large

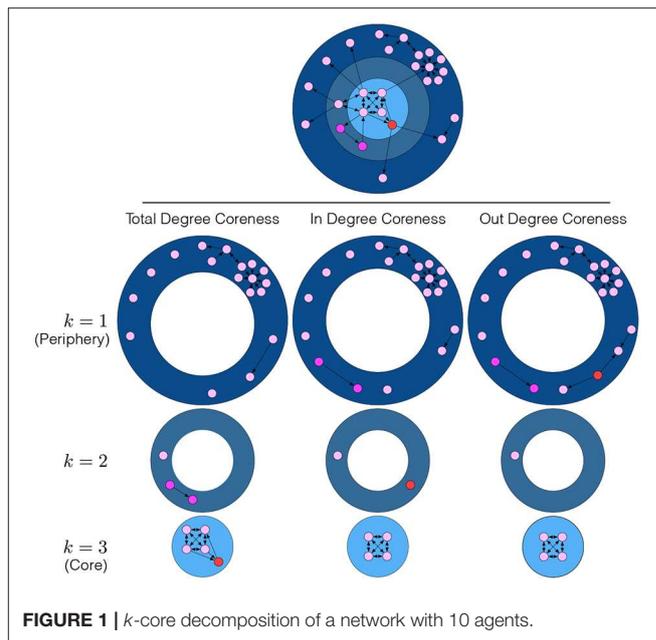


FIGURE 1 | k -core decomposition of a network with 10 agents.

drop-out cascade that affects even the core of the OSN, then the robustness of the system is very low. We need to come up with a robustness measure that reflects such a situation appropriately. This is developed in the next section.

2.2. Quantifying Robustness

2.2.1. Coreness

We decided to use the *coreness* k_i of agents as our starting point because it reflects from a topological perspective how well an agent is integrated into the network (Seidman, 1983). A coreness value k_i allows quantifying the impact on the network when removing agent i . Individual coreness values are obtained by means of a pruning procedure, which is known as *k-core decomposition*. It assigns agents to different concentric *shells* that reflect the integration of these agents in the network. Specifically, the *k-core* is identified by subsequently pruning all agents with a degree $d_i < k$. Pruning starts with $k = 1$ and stops when all the agents left have a degree greater or equal to k_{max} . The corresponding *k-shell* then consists of all agents that are in a *k-core* but not in the $(k + 1)$ -core, i.e., agents assigned to a *k-shell* have coreness value $k_i = k$.

Figure 1 provides an illustration of the *k-core decomposition* applied to a network of 10 agents. Agents with a coreness $k_i = 1$ are located in the *periphery* (dark blue), i.e., they are loosely connected with the core. Note that some of these agents have a relatively high degree, in spite of their low coreness. Agents with a coreness $k_i = 2$ are closely connected to, but *not yet* fully integrated into the core, belong to an intermediate shell (blue). The 5 agents with coreness $k_i = k_{max} = 3$ are the most densely connected ones in this sample network and belong to the innermost core (light blue). This illustrates that the higher the coreness k_i of an agent i , the stronger the impact on the network when removing i because this potentially

disconnects a large number of agents with lower coreness from the network. Conversely, removing agents with low coreness will have a weaker impact on the network because they belong to outer shells, and removing them disconnects a smaller number of agents.

In this article, we want to quantify how much the drop-out of agents will impact the *robustness* of the network. As motivated above, robustness shall be characterized by the *average coreness* of the agents:

$$\langle k \rangle = \frac{1}{N} \sum_{k=1}^{k_{\max}} k n_k ; \quad \sum_{k=1}^{k_{\max}} n_k = N \quad (2)$$

where $N = |V|$ is the total number of (connected and disconnected) agents in the network and n_k is the number of agents with a coreness value $k_i = k$. $\langle k \rangle$ will be high if either most agents have a relatively high coreness, or few agents have a very high coreness. In both cases, the core of the network is less likely to be affected by cascades that started in the periphery. So, $\langle k \rangle$ summarizes the information we are interested in. In this paper, we do not focus on the *heterogeneity* of coreness values, which could be described by the *variance* of the coreness distribution, or by *coreness centralization* (Wasserman and Faust, 1994).

2.2.2. In-Degree and Out-Degree Coreness

The above definition of coreness is based on the total degree d_i of agents, i.e., it is appropriate for *undirected* networks. For the case of a *directed* network discussed in this paper, this may give wrong conclusions about the embeddedness of agents. Therefore, we now introduce two separate measures, in-degree coreness, k_i^+ , and out-degree coreness, k_i^- , which reflect the existence of directed links via the in- and out-degrees d_i^+ , d_i^- .

The results for the different metrics and the differences between them are illustrated in the sample network of 10 agents in **Figure 1**. This network is characterized by 3 k -shells, but it is important to note that the three different coreness metrics possibly assign the same agents to very different k -shells. Take the example of the pair of purple agents that, according to total-degree coreness, are assigned to the shell $k = 2$. If we account for directionality of the links, they are now assigned to $k = 1$, i.e., to the *periphery*. Moreover, the red agent that, according to the total degree coreness, belongs to the core, $k_{\max} = 3$, is now assigned to the shell $k = 2$ if *in-degree coreness* is taken into account, and to $k = 1$, i.e., to the periphery, if *out-degree coreness* is instead considered.

This example makes clear that it very much depends on the *application* whether coreness should be calculated based on directed or undirected links, and whether in- or out-degrees should be considered. In the following we will use *in-degree coreness*, k_i^+ , to compute the average coreness $\langle k \rangle$, Equation (2), i.e., n_k is the number of agents with in-degree coreness $k_i^+ = k$. The reason for this choice comes from the benefits of agents defined in 1 and is discussed in the following section.

3. DYNAMICS ON THE SOCIAL NETWORK

3.1. User Benefits and Costs

To enable a network dynamics on the time scale T , where agents can *leave* the network according to Equation (1), we need to further specify their benefits, $B_i(T)$, and costs, $C_i(T)$. This leads to the question of why, in the real world, users join or leave an OSN. There are certainly different reasons, such as information exchange, maintaining friendship links, or receiving *attention*. From this, we can deduce that benefits should *increase* with the *in-degree* d_i^+ of an agent in a monotonous, but likely non-linear manner. For instance, on Twitter attention increases with the *number* of followers. More important, however, is not just the number, but also the *importance* of the followers. The attention for a user i can considerably increase if it has a number of important users j following. This amplifies the attention because, in an OSN, other users following the important user j this way also receive information from i .

To capture such effects in our agent-based model, we assign to each agent a second state variable, *reputation* R_i , which is continuous and positive. In real-world OSN, user reputation plays an important role and can be proxied by different measures, such as number of likes in Facebook positive votes in Amazon and Dooyoo, or retweets on Twitter. Other proxies take the activity of users into account, for example, the RG score from Researchgate, or the Karma points from Reddit. All of these measures have the drawback that they are (i) specific to the OSN, (ii) depend on the subjective judgment of other users (see e.g., Golbeck and Hender, 2004, 2006). In the existing literature, the concept of reputation often relates to that of the trust agents pose on each others (Golbeck and Hender, 2004; Guha et al., 2004; De Meo et al., 2015). Such reputation depends on the activity in the OSN of the agents, e.g., when they evaluate content posted by other agents by “liking” or “disliking” it (Liu et al., 2008; DuBois et al., 2011). In particular, De Meo et al. (2015) have shown that OSN characterized by groups of agents that have higher reputation of each other have higher compactness, and are possibly more stable over time.

Differently from these works, to express agents’ reputation we resort to so-called *feedback centrality* measures. These are prominently known from the early versions of the PageRank algorithm, in which the importance (centrality) of a node in a network entirely depends on the importance of the nodes linked to it. This choice effectively allows us to estimate agents’ reputation directly from the observed topology of the network. This leads to a set of equations for the importance of all nodes that has to be solved in a self-consistent way. While this is a crucial element to define our reputation measure, it is not enough to explain reputation. We also need to consider that reputation fades out over time if it is not continuously *maintained*. Usually, the reputation of an agent can be maintained in different ways, (i) by the own effort of the agent and (ii) by means of direct interactions with others. Such considerations have been formalized in other reputation models (Schweitzer et al., in review). Here, we only consider the increase of reputation coming from other agents, to simplify the formalization.

In the following section, we will specify our dynamics for the reputation of an agent, which leads to a stationary value of $R_i(T)$. Given that we have calculated this value, we posit that the benefit of an agent from being in the OSN comes from its reputation as a good proxy of the attention that this agent receives from others. The absolute value of R_i will also depend on the network size and the density of links. What matters in an OSN is not the *absolute* value, but the reputation of users *relative* to that of others. Therefore, we define the benefit B_i for each agent $i \in V(\mathcal{G})$ as the absolute reputation rescaled by the largest reputation value $R_{\max}(T)$ at the given time T .

$$B_i(T) := b \frac{R_i(T)}{R_{\max}(T)} = b \frac{R_i(T)}{\max_{j \in V(\mathcal{G})} R_j(T)} \quad (3)$$

The constant b allows to weight the benefits from the reputation against the costs.

To specify the costs $C_i(T)$, in our model, we consider two contributions. First, there are fixed costs per time unit, c_0 , that do not depend on the activity of the agents. They capture, in a real OSN, the minimal effort made by users to be present in the OSN, i.e., to learn about the GUI and to maintain the profile. The second contribution comes from the costly interaction with other agents. Because, for instance on *Twitter*, agent i can only control whom to follow, these costs should be proportional to the *out-degree* d_i^- of the agent, $c_i d_i^-$. In a real OSN, the costs per interaction, c_i , are not the same for all users. More prominent users have, for example, much more time constraints because of other activities that compete for their attention. Therefore, it is reasonable to assume that c_i is a non-linear function of the user's reputation, $c_i(R_i) = c_1 R_i^2$. The non-linearity induces a stronger saturation effect for more prominent users in interacting with many other users.

As with the benefits, also the costs should not depend on the absolute reputation of the agent, but on the relative one. This

leads to

$$C_i(T) := c_0 + c_1 d_i^- \left[\frac{R_i(T)}{R_{\max}(T)} \right]^2. \quad (4)$$

Denoting the relative reputation at a given time T as $r_i(T) = R_i(T)/R_{\max}(T)$, we can eventually write down the utility function of agent i , Equation (1), as:

$$U_i(T) = b r_i(T) - c_0 - c_1 d_i^- r_i^2(T) = -c_0 + [b - c_1 d_i^- r_i(T)] r_i(T). \quad (5)$$

3.2. Reputation Dynamics

After linking the utility function of agents to their reputation, we have to specify how to calculate the latter. In accordance with the above discussion, we use the following reputation dynamics:

$$\frac{dR_i(t)}{dt} = -\gamma R_i(t) + \sum_{j \in V(\mathcal{G}(T))} a_{ji} R_j(t) \quad (6)$$

Here, t denotes a *time scale* much shorter than the time scale T at which agents decide whether to stay or to leave the OSN. Hence, compared to the change of the *network*, the change of *reputation* is *fast* enough such that a stationary value $R_i(T)$ is obtained at time T .

The first term in Equation (6) expresses a continuous decay of reputation with a rate γ , to reflect the fact that reputation fades out over time if it is not maintained. The second term captures the increase of reputation coming from other agents linked to agent i , i.e., $a_{ji} = 1$. The summation is over all agents part of the OSN at time T .

Whether or not the reputation values $R_i(T)$ converge to positive stationary values very much depends on the topology of the network expressed by the adjacency matrix \mathcal{A} , as illustrated in **Figure 2**. Specifically, if an agent has no incoming links that boost its reputation, $R_i(t)$ will go to zero. Therefore, even if this agent has an outgoing link to other agents j , it cannot boost

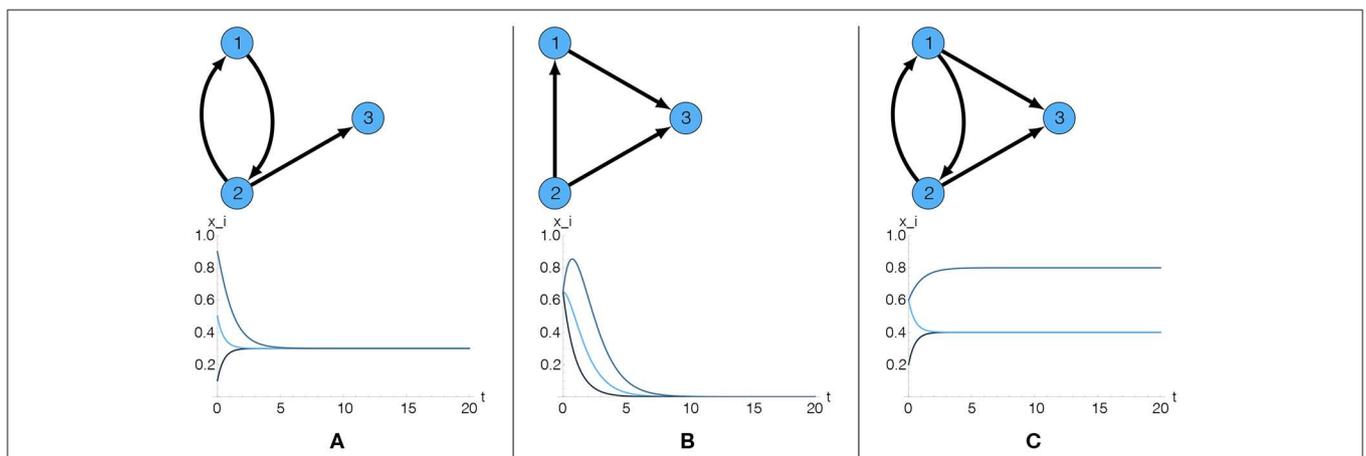


FIGURE 2 | Impact of the adjacency matrix on the reputation $R_i(t)$ of three agents. Only if cycles exist and agents are connected to these cycles, a non-trivial stationary reputation can be obtained. **(A)** The presence of one cycle guarantees a non-trivial stationary reputation, identical for all agents. **(B)** The absence of cycles results in a trivial stationary reputation for all agents. **(C)** The presence of a cycle guarantees non-trivial stationary reputations. Furthermore, two different stationary values appear when agent 3 has 2 incoming links to boost its reputation.

their reputation. Non-trivial solutions depend on the existence of *cycles*, which are formally defined as subgraphs with a closed path from every node in the subgraph back to itself. The shortest possible cycle involves two agents, $1 \rightarrow 2 \rightarrow 1$. This maps to *direct reciprocity*: agent 1 boosts the reputation of agent 2 and vice versa. Cycles of length 3 map to *indirect reciprocity*, for example $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$. In this case, there is *no* direct reciprocity between any two agents, but all of them benefit regarding their reputation because they are part of the cycle. In order to obtain a non-trivial reputation, an agent not necessarily has to be part of a cycle, but it has to be connected to a cycle.

4. DYNAMICS OF THE SOCIAL NETWORK

4.1. Entry and Exit Dynamics

We now have all elements in place to model the entry and exit dynamics of agents in the OSN. At each time step T , agents evaluate their benefits and costs according to Equations (3) and (4). This is based on their relative reputation $r_i(T)$ which has reached a stationary value at time T , according to Equation (6). They then make a (deterministic) decision to either stay or leave the OSN, according to Equation (1).

Hence, at every time T , a number $N^{\text{ex}}(T) < N$ of agents will leave the network. To compensate for this, we assume that the *same* number of new agents will enter the network at the same time, i.e., $N = \text{const.}$ all the time. One may argue that this is at odds with our research question, namely to model how cascades of users leaving impact the robustness of the OSN. But as the empirical case study of the collapse of the OSN *Friendster* has demonstrated (Garcia et al., 2013), this collapse was *not* due to the fact that no new users entered. Instead, they became *less integrated* into the social network. Signs for this trend became already visible when *Friendster* had about 80 million users. After that, it still grew up to 113 million users, until it collapsed. So, the problem of the robustness of an OSN cannot be trivially reduced to the (wrong) assumption that there is a lack of new users entering.

Therefore we have to address the question of how, *despite entering of new users*, large drop-out cascades become increasingly likely. To measure the size of the *drop-out cascades*, we will monitor $N^{\text{ex}}(T)$ over time. If this number is consistently large, it becomes evident that even with a large entry rate, new agents cannot substantially stabilize the OSN, hence its robustness is lost. We further need to study how new agents will be *integrated* in the OSN. If at any time T a varying number of $N^{\text{ex}}(T)$ agents *enter*, we have to model how they are linked to the network, to become members of the OSN. We assume that new agents do not have complete knowledge of the network; therefore, to start with, they form *random connections* to a (varying) number of members. Precisely, as in random graphs, new agents create directed links to established agents with a small probability p . Thus, their *expected number* of links is roughly Np .

Because agents leaving delete all their links and agents randomly entering create links, the topology of the network continuously changes at the time scale T . To ensure that the evolution also continues if *no* agent has decided to leave, in this case, we randomly pick one of the agents with the lowest relative

reputation, to replace it with one new agent. To measure how well new agents become integrated into the OSN, we monitor the mean coreness $\langle k \rangle(T)$, Equation (2), over time T . Large values indicate that most agents belong to the core, small values instead that most agents belong to the periphery.

4.2. Results of Computer Simulations

In the following, we discuss the simulation results for a network of fixed size, $N = 20$. Further we use fixed parameters $\gamma = 0.1$, $b = 1$, $c_0 = 0.45$, $c_1 = 0.05$, $p = 0.05$. For a discussion of parameter dependencies and optimal values, see section 5.2.

To initialize our simulations of the network dynamics, we assume that at time $T = 0$, 5 out of 20 agents initially form a fully connected cluster, as shown in **Figure 3A**. This ensures that these five agents have a non-zero reputation at $T = 1$ and thus will not leave the OSN. The remaining 15 agents with reputation zero, however, will be replaced by new agents that randomly create links to the agents in the network. This way, at $T = 50$ already a realistic network structure with a *core*, a *periphery*, different *k-shells* and a few isolated agents emerges, as shown in **Figure 3B**. **Figure 3** displays further snapshots of the network evolution, while the corresponding systemic variables to monitor the dynamics, namely the mean coreness, $\langle k \rangle(T)$, and the number of agents leaving, $N^{\text{ex}}(T)$, are shown in **Figure 4**. From the latter, we can clearly identify three different phases of network evolution.

4.2.1. (I) Build Up Phase

In this initial phase, as already mentioned, the network establishes its characteristic topology. Most agents become tightly integrated into the network, as also visible from **Figures 3B,C**. Because of this, the mean coreness quickly increases, while the number of agents leaving decreases, but both variables show considerable fluctuations.

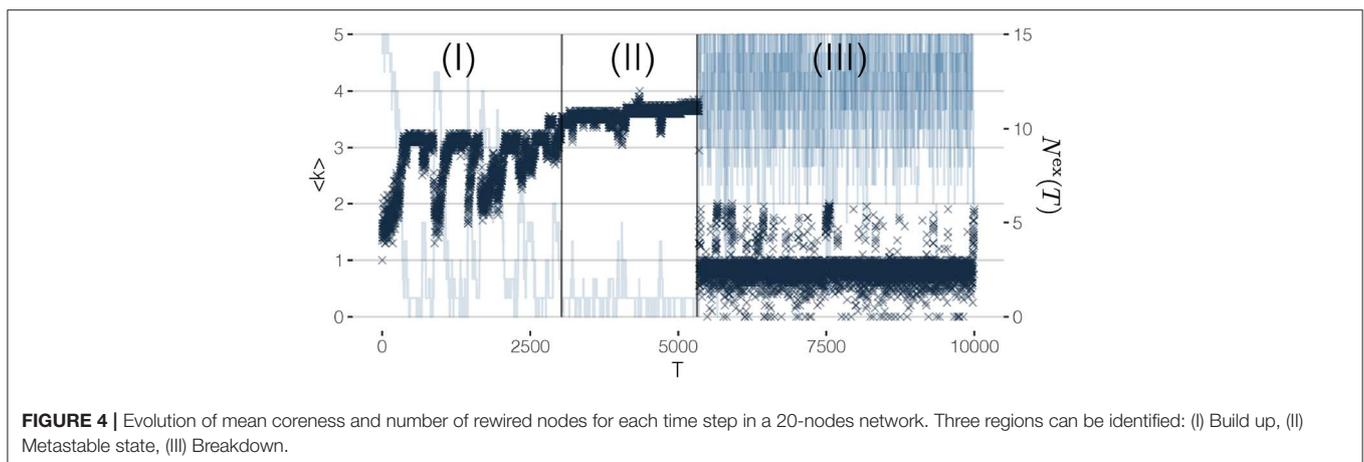
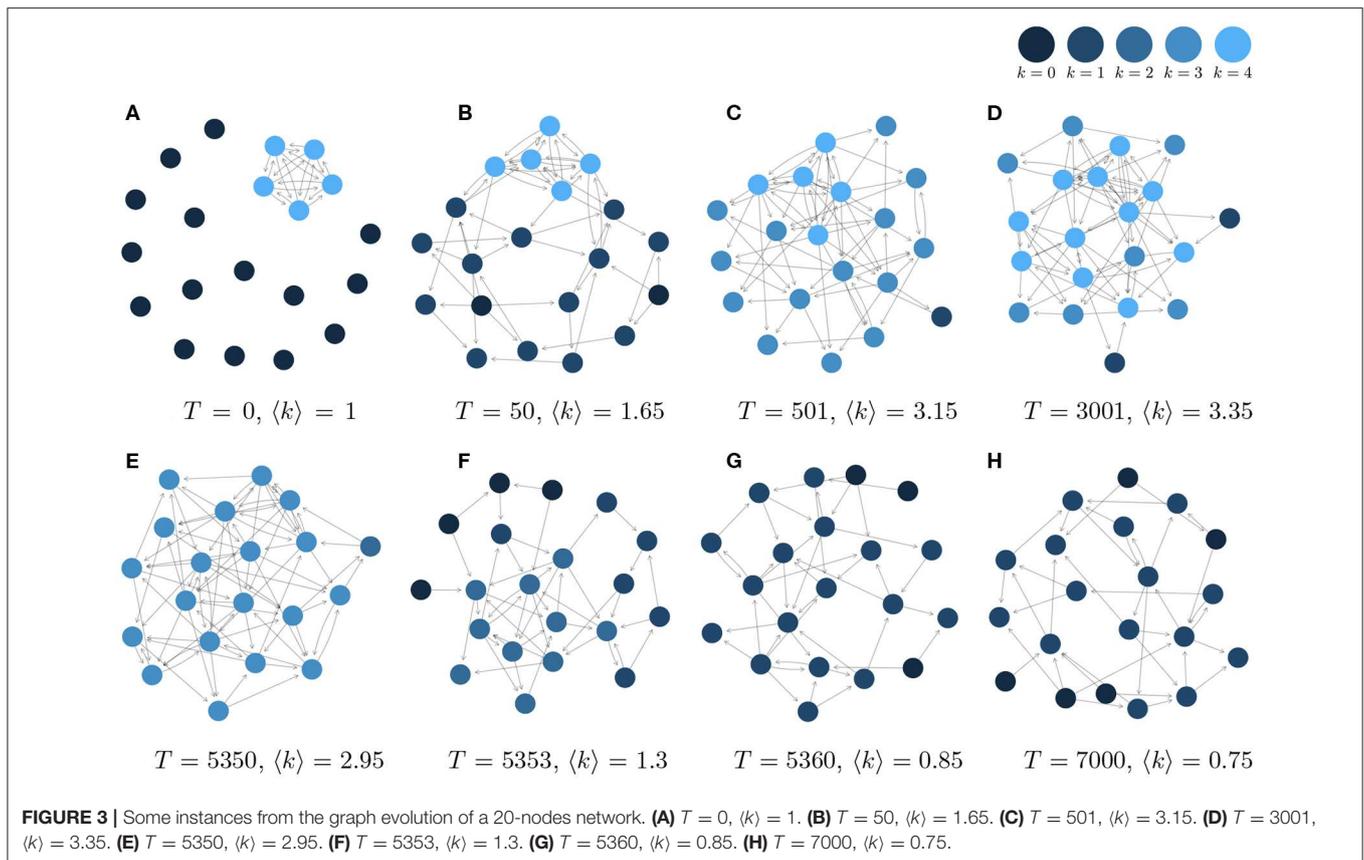
4.2.2. (II) Metastable Phase

After agents have become well-connected to the core, they tend to have higher benefits than costs. If no agent would leave the OSN, we choose one of the agents with the lowest reputation to leave, to keep the network dynamics going. Hence, $N^{\text{ex}}(T) = 1$ or very low, for most of the time, while $\langle k \rangle$ only slightly fluctuates.

Still, the status of the OSN is not stable but only *metastable*, because of the slow dynamics that is illustrated by means of **Figures 3E,F**. Agents that were earlier part of the periphery have now become part of the core, this way *decreasing* the size of the periphery. In fact, the smaller the periphery, the more likely the formation of new links to the core. The probability that a new agent i becomes part of the core Q with size $|Q|$ is given as:

$$P(i \in Q) \geq \binom{|Q|}{k_{\max}^-} p^{k_{\max}^-} \cdot \binom{|Q|}{k_{\max}^+} p^{k_{\max}^+} \quad (7)$$

where k_{\max}^- , k_{\max}^+ are the values for the in-degree and the out-degree coreness of the agents in the *core*. The two r.h.s terms stand for the probability of creating and of receiving links from the core, where p is the probability for an incoming agent to create a new link. $P(i \in Q)$ is indeed increasing with the size of the core, $|Q|$ (Łuczak, 1991).

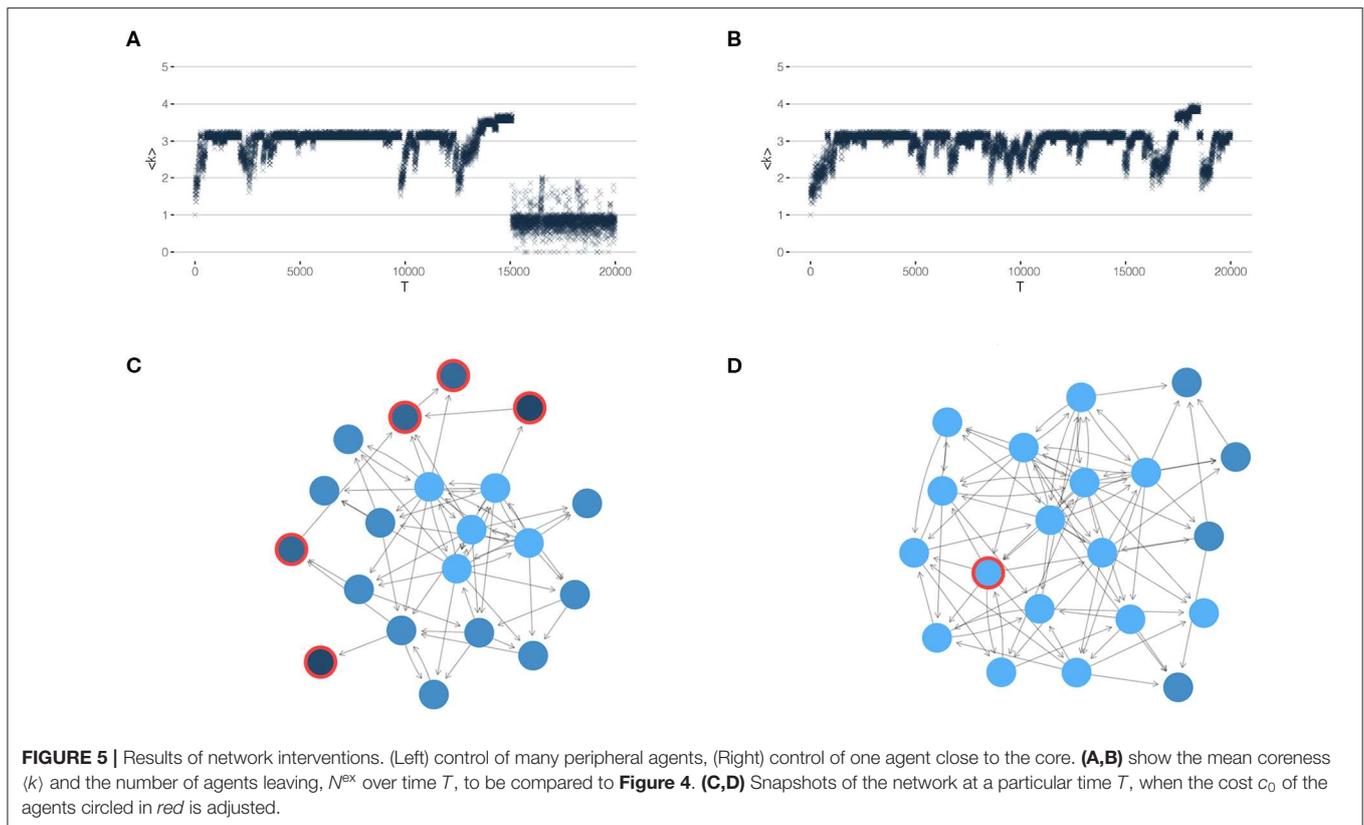


4.2.3. (III) Breakdown Phase

The slow dynamics during phase (II) leads to a point where agents from the outer shells of the in-degree core receive a higher reputation than agents in the core. If no agent decides to leave the OSN, in this situation, an agent from the core is chosen to be removed, because of the lower reputation. This then triggers whole *cascades* of agents leaving, because the drop-out of a core agent abruptly decreases the reputation of other agents in the core and the outer shells. The transition from phase (II) to phase (III) can be seen by the

increasing number of agents leaving, while the mean coreness steadily *decreases*.

Once the core has been destroyed, the OSN has no ability to recover because most agents are replaced at each time step. Nearly all links from the newly entering agents will be to agents from the periphery; thus, the probability of forming a new core is extremely low. The breakdown phase (III) can be characterized not only by the rather low mean coreness and the large number of entries and exits, but also by the much larger fluctuations of both values.



5. IMPROVING ROBUSTNESS

5.1. Network Interventions

The simulation results shown in **Figure 4** make it very clear what we mean by improving robustness: to prevent the *complete breakdown* of the OSN. This does not imply to prevent cascades, which can always happen in response to agents leaving the OSN. But we argue that a social network is *robust* if the decision of agents to leave the OSN will not trigger large cascades of leaving agents that destroy the whole core.

This requires us to influence agents in the OSN such that they decide *not* to leave the network. The trivial solution would be to reduce the costs of *all* agents to a level that always guarantees a positive utility or to increase the benefits in the same manner. A much smarter solution, however, would focus only on a *few* agents, namely those with the ability to prevent large cascades. The problem to identify those agents is addressed in research about network controllability (Liu et al., 2011; Zhang et al., 2016), which is related to control theory. The method assigns a *control signal*, i.e., an incentive to stay or to leave, to the identified agents with the most influence on the network dynamics (Zhang et al., 2019), which are called *driver nodes*. Precisely, this signal is added to the reputation dynamics, Equation (6), of the driver nodes.

We will not follow this formal procedure in our paper for several reasons. The most important one is the continuous evolution of the network topology, which is not considered in the network controllability approach. It would require us to redo the

identification of the driver nodes and the assignment of control signals at every time step T . Further, in our context of users leaving an OSN, these control signals are difficult to interpret because they change the *reputation* dynamics. Our intention instead is to influence the *decisions* of the agents, Equation (1), i.e., to apply control signals to the *costs* of staying in the OSN. Specifically, we apply two different scenarios to incentivize agents (i) from the periphery, or (ii) from the core.

The first scenario is motivated by our insight that large cascades are caused by the *disappearing periphery*. Therefore, a straightforward intervention is to choose agents with a low reputation from the periphery as drivers. These are incentivized to *stay* in the OSN, i.e., their costs are *reduced* such that their utility is increased and they decide *not* to leave. The second scenario is to choose agents close to the core, i.e., from its first outer shells, as drivers. These are incentivized to *leave* the OSN, i.e., their costs are *increased* such that they decide to *not stay*. This more subtle scenario is motivated by the insight that agents that are only close to the core will *not* trigger large cascades if they leave. But if they leave, they considerably reduce the reputation of their closest neighbors, this way *increasing* the size of the periphery. The results of these two scenarios are illustrated in **Figure 5**.

Specifically, in scenario (i), we identify at each time step T all agents from the periphery, i.e., with a coreness value $k_i = 1$. Their cost c_0 is then reduced by 10%, i.e., to $\hat{c}_0 = 0.9c_0$. As **Figure 5A** demonstrates, this scenario can only delay the

complete breakdown (in comparison to **Figure 4** without any interventions). But it cannot completely prevent large drop-out cascade, because the build-up of a large core that eventually gets destroyed is only delayed.

In scenario (ii), on the other hand, we are able to achieve the goal of preventing a complete breakdown. This scenario has remarkable differences to scenario (i): We only incentivize *one* agent, instead of many, and we choose this agent from the vicinity of the *core* instead from the periphery. Precisely, we choose the agent from the first outer shell identified by means of the directed k -core decomposition, i.e., $k_i = k_{\max} - 1$. This agent is enforced to leave by increasing its cost by 10 percent, i.e., to $\hat{c}_0 = 1.1c_0$.

As shown in **Figure 5B**, this scenario considerably improves the robustness of the network, as witnessed by the average coreness. At the same time, because one agent is chosen for control from the beginning, we also observe that the build-up phase (I) is extended in comparison to the case of no control (see **Figure 4**). But phase (II), which was called metastable before, is now considerably extended. We still notice small cascades, but no complete breakdown, i.e., the metastable phase has become a *quasistable one*.

5.2. Life-Time Before Breakdown

The above simulations are both interesting and counter-intuitive because controlling one agent close to the core leads to much better results than controlling many agents from the periphery. We, therefore, continue with a more refined discussion of the peripheral control. As shown, this kind of network intervention increases the time before the breakdown, but cannot completely prevent it. To further quantify this dynamics, we use the *life-time* Ω_Q of the core Q (measured in network time T) as an additional systemic variable (Schweitzer et al., in review). As **Figure 5A** illustrates, for scenario (i) the value of Ω_Q can be clearly obtained from the simulations because of the sharp transition toward the breakdown of the OSN. For scenario (ii), obviously $\Omega_Q \rightarrow \infty$ as **Figure 5B** shows.

We are interested in comparing the life-times of the core for peripheral control and without control (also shown in **Figure 4**). Because Ω_Q changes considerably for different simulations, we use the average life-time $\langle \Omega_Q \rangle$ taken from 100 independent runs with the same setup. We further have to consider that $\langle \Omega_Q \rangle$ depends on other system parameters, notably the system size N . We, therefore, vary N for simulations with peripheral control and without control, keeping all other parameters the same. The results are shown in **Figure 6**, from which we can deduce some interesting insights.

First, we note that for *small* networks ($N < 30$), our peripheral control strategy works very well. The life-times increased considerably in comparison to the no-control reference case. Secondly, we observe that this advantage becomes smaller if the network size increases. For networks larger than $N = 30$, there is almost no difference in life-times between the peripheral control and the no-control case. Further, for $N > 30$ in both cases, the life-time decreases almost linearly with the increasing network size.

The latter observation can be explained from the fact that, with increasing network size N , the network becomes much denser.

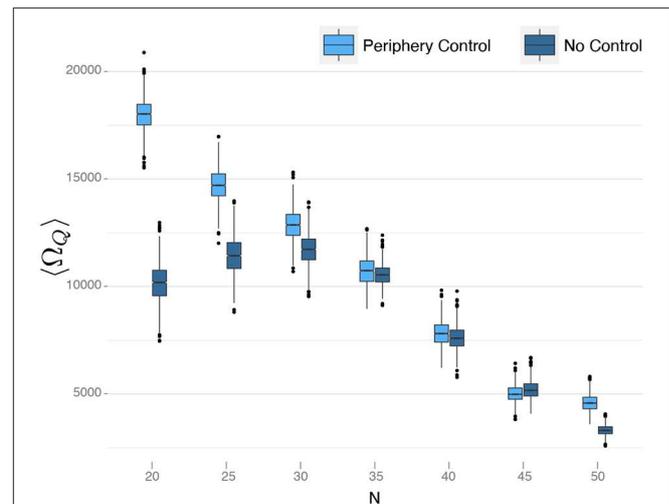


FIGURE 6 | Comparison of different periphery control approaches with fixed control signal. The effectiveness of the control method without adapting the signal to the size of the network decreases with size. In the figure are plotted bootstrap samples for $\langle \Omega_Q \rangle$ obtained from 100 simulations for each network size and each strategy. The control signal used is $u = -0.05$.

We recall that links between agents are formed such that new agents entering the OSN create links to established agents with a *fixed probability*, p . The average number of links per agent is thus Np , i.e., it increases linearly with N . The denser the network, the larger the core and the smaller the periphery. In line with our above discussion, this means less robustness of the network, i.e., the breakdown occurs earlier in time.

The non-monotonous dependence of $\langle \Omega_Q \rangle$ on the network size, for the *no-control* case, results from the fact that the model parameters are not completely independent. This fact is also obvious from Equation (5). Instead, it was already pointed out (Schweitzer et al., under review) that there is an *optimal cost level* to maximize the life-time of the network. This is understandable from our above discussions. If costs are very low, only very few agents will leave the OSN. Because of the slow dynamics described in phase (II), these agents will, at some point, reach a reputation large enough to compare to the core, and hence the core agents will leave. An intermediate cost level, on the other hand, makes sure that this evolution does not take place, or is at least considerably delayed. The optimal cost level that maximizes the life-time, however, also depends on the other parameters, b , N , γ , p .

From **Figure 6**, we can deduce that, for the fixed cost parameters chosen in our simulation, the optimal network size is $N = 30$, simply because, for this size, the life-time is maximized (kept all other parameters the same). Hence, for small networks, $N < 30$, the optimal cost level should be *lower* than what was used in the simulation. Given the suboptimal values, the life-time was also lower for the no-control case. Remarkably, the life-time in case of peripheral control is not affected by this. So, we can conclude that, at least for small networks, peripheral control also compensates for not optimal parameter choices.

For larger networks, $N > 30$, **Figure 6** suggests that there is *no difference* between peripheral control and no control. But this observation is mainly due to the fact that we have not used the optimal parameters for a given network size N . To further investigate this, we have performed an extensive optimization to determine the optimal values for c_0 and \hat{c}_0 for a given N . It then turns out that, with the optimal parameters, the life-times for the peripheral control and no-control cases are no longer the same, but differ *significantly*.

Specifically, we performed two-samples t -tests for the means and *Wilcoxon*-tests for the medians of bootstrap samples of the average life-times (Ω_Q) obtained from the simulations with and without control. As the H_0 hypothesis, we assume that the means of the life-times in both cases are equal and as alternative hypothesis that the life-times are higher in case of peripheral control. Using always the optimal parameters for both cases, we obtained p -values in the order of 10^{-12} for the alternative hypothesis, independent of the network size. This provides strong evidence for the conclusion that the peripheral control *always* improves the robustness of the network, as measured by the life-time before breakdown. For small networks, this holds already for arbitrary parameter choices, for large networks only if the optimal parameters are chosen.

In **Figure 6**, we also plot the bootstrapped 95% confidence intervals for the average life-time (Ω_Q). We note that the size of the confidence interval decreases with N . Hence, for small networks, even optimal parameter values cannot guarantee a minimal variance of Ω_Q , and in single simulations, a breakdown of the network can happen much earlier or later.

Eventually, we also tested whether reputation differences in the peripheral agents matter for the network intervention. While the above simulations assumed that *all* peripheral agents are controlled, we also considered that only peripheral agents with *high*, or with *low* reputation are influenced in their costs. These cases, however, did not generate any remarkable difference with respect to the average life-time.

6. CONCLUSIONS

After more than 35 years of *understanding* complex systems, there should be foundations enough for *managing* them in a better and more quantitative manner. Sadly, to know how systems *work* does not already imply also to know how to *influence* them such that more desired system states are obtained. This holds particularly for socio-economic systems, which are *adaptive*, which means they respond to proposed changes in both intended and unintended ways. *Systems design* (Schweitzer, 2019) therefore has to master a difficult balance: on the one hand, systems should be carefully steered toward a wanted development, on the other hand, systems should not be over-regulated, to not lose their ability to innovate and to find solutions outside the box. This balance cannot be obtained by brute force, in a top-down approach to system dynamics, it has to be found in a bottom-up approach that focuses on the system elements and their interactions.

Our paper contributes to this discussion in several ways. We study a problem of practical relevance that can hardly be solved in a top-down approach: the collapse of an online social network

(OSN) because the decision of some users to leave causes the drop-out of others at large scale. A real-world example is the collapse of the OSN *Friendster* (Garcia et al., 2013). As long as users are free to stay or to leave, the *emergence*, of such large failure cascades cannot be prevented by administrative ruling. Applying global incentives for users to stay, on the other hand, usually implies high costs and questionable efficiency.

Therefore, in this paper, we propose a bottom-up approach to influence the OSN on the level of users, i.e., agents in our model. They can be targeted in two ways: by influencing their interactions or by influencing their utility. We have argued for the latter, because of the large volatility in the dynamics of the OSN. Specifically, we propose to change the costs of particular agents such that the overall robustness of the OSN is increased. As already mentioned in the Introduction, OSN should be seen as *socio-technical systems*, and it is in fact the *technical* component that in principle allows us to influence the costs of users much easier than it would be possible in the offline world.

Improving robustness first requires us to define an appropriate measure of robustness suitable for real-world OSN. Here we propose the *average in-degree coreness*, which does not just reflect the degree of agents but quantifies how well they are *integrated* in the OSN. Next, we have to understand why robustness *decreases* in the absence of network interventions. Based on computer simulations and detailed discussions of agent benefits and costs, we show that it is the changing relation between the core and the periphery of the OSN, which eventually destabilizes the network. Our approach deviates from the one taken in De Meo et al. (2015, 2017) in the fact that we are interested in the robustness of the whole network, and not so much of separate groups. In fact, we learn that is heterogeneity within the network topology, in terms of core-periphery structure, what guarantees robustness. This is in contrast with what expected by generalizing those results obtained for separate groups, where agents' homogeneity increases stability. Moreover, our approach allows to estimate the reputation of agents in the absence of explicit data collecting active declarations of trust between agents in the OSN. To do so, we exploit so-called feedback centralities, that exploit the OSN topology. This is in contrast with common approaches that rely on the presence of likes, dislikes, or agents' ratings to provide a measure for the reputation of agents.

Based on the insights obtained from our analysis, we have proposed two different scenarios for network interventions to improve robustness. The first one targets peripheral agents and reduces their cost, to incentivize them to *stay* in the OSN. The second one targets only *one* agent from a k -shell next to the core and increases its cost, to incentivize it to *leave* the OSN. Both scenarios have in common to increase the size of the periphery, but they reach this goal in different ways. As we demonstrate by means of computer simulations, the first scenario is able to considerably *delay* the breakdown of the OSN, while the second one is able to prevent this breakdown. Dependent on the optimal choice of parameters, we could show that even the peripheral control improves the robustness of the OSN in a *statistically significant* manner. Still, we argue that the second scenario should be the preferred one because it requires (i) to only control a single

agent instead of many, and (ii) less investment because, instead of *decreasing* the costs of many agents via compensations, here the cost is increased.

Our findings are interesting and, at first sight, also counter-intuitive because they challenge our understanding of how to improve the robustness of systems. One could simply argue that the best way to increase robustness is to keep all parts of the system tightly together, to not lose anything. This may apply to mechanical or technical systems. But for socio-technical and socio-economic systems, we have to take into account their adaptivity and their ability to respond to changes in an unintended manner. Therefore, the first step for interventions is to understand the eigendynamics of these systems, i.e., their behavior in the absence of regulations or control. To achieve this understanding in the case of complex systems, agent-based modeling is the most appropriate way. Different from a complex network approach that focuses mainly on the

link topology, agent-based modeling allows also capturing the internal dynamics of the system elements, i.e., the nodes or agents, in response to interactions. Only this advanced level of modeling enables us to propose interventions targeted at specific agents and to investigate how the system as a whole responds to these network interventions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

GC and FS designed the research and wrote the manuscript. GC carried out the computer simulations.

REFERENCES

- Borgatti, S. P., and Everett, M. G. (2000). Models of core/periphery structures. *Soc. Netw.* 21, 375–395. doi: 10.1016/S0378-8733(99)00019-2
- Botafogo, R. A., Rivlin, E., and Shneiderman, B. (1992). Structural analysis of hypertexts: identifying hierarchies and useful metrics. *ACM Trans. Inform. Syst.* 10, 142–180. doi: 10.1145/146802.146826
- De Meo, P., Ferrara, E., Rosaci, D., and Sarne, G. M. L. (2015). Trust and compactness in social network groups. *IEEE Trans. Cybernet.* 45, 205–216. doi: 10.1109/TCYB.2014.2323892
- De Meo, P., Messina, F., Rosaci, D., and Sarné, G. M. (2017). Forming time-stable homogeneous groups into online social networks. *Inform. Sci.* 414, 117–132. doi: 10.1016/j.ins.2017.05.048
- DuBois, T., Golbeck, J., and Srinivasan, A. (2011). “Predicting trust and distrust in social networks,” in *2011 IEEE Third Int’l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int’l Conference on Social Computing*, 418–424.
- Egghé, L., and Rousseau, R. (2003). BRS-compactness in networks: theoretical considerations related to cohesion in citation graphs, collaboration networks and the internet. *Math. Comput. Model.* 37, 879–899. doi: 10.1016/S0895-7177(03)00091-8
- Garcia, D., Pavlin, M., and Frank, S. (2013). “Social resilience in online communities: the autopsy of Friendster,” in *Proceedings of the First ACM Conference on Online Social Networks* (New York, NY: Association for Computing Machinery), 39–50. doi: 10.1145/2512938.2512946
- Golbeck, J., and Hendler, J. (2004). “Accuracy of metrics for inferring trust and reputation in semantic web-based social networks,” in *Engineering Knowledge in the Age of the Semantic Web*, eds E. Motta, N. R. Shadbolt, A. Stutt, and N. Gibbins (Berlin; Heidelberg: Springer Berlin Heidelberg), 116–131. doi: 10.1007/978-3-540-30202-5_8
- Golbeck, J., and Hendler, J. (2006). Inferring binary trust relationships in Web-based social networks. *ACM Trans. Internet Technol.* 6, 497–529. doi: 10.1145/1183463.1183470
- Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). “Propagation of trust and distrust,” in *Proceedings of the 13th Conference on World Wide Web - WWW '04* (New York, NY: ACM Press), 403. doi: 10.1145/988672.988727
- Jain, S., and Krishna, S. (1998). Emergence and growth of complex networks in adaptive systems. *Comput. Phys. Commun.* 122:10. doi: 10.1016/S0010-4655(99)00293-3
- Jain, S., and Krishna, S. (2002). Crashes, recoveries, and core shifts in a model of evolving networks. *Phys. Rev.* 65, 26103–26104. doi: 10.1103/PhysRevE.65.026103
- Kairam, S. R., Wang, D. J., and Leskovec, J. (2012). “The life and death of online groups,” in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining - WSDM '12* (New York, NY: ACM Press), 673. doi: 10.1145/2124295.2124374
- Liu, H., Lim, E.-P., Lauw, H. W., Le, M.-T., Sun, A., Srivastava, J., et al. (2008). “Predicting trusts among users of online communities,” in *Proceedings of the 9th ACM Conference on Electronic Commerce - EC '08* (New York, NY: ACM Press), 310. doi: 10.1145/1386790.1386838
- Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of complex networks. *Nature* 473:167. doi: 10.1038/nature10011
- Łuczak, T. (1991). Size and connectivity of the k-core of a random graph. *Discrete Math.* 91, 61–68. doi: 10.1016/0012-365X(91)90162-U
- Schweitzer, F. (Ed.). (1997). *Self-Organization of Complex Structures: From Individual to Collective Dynamics. Part 1: Evolution of Complexity and Evolutionary Optimization, Part 2: Biological and Ecological Dynamics, Socio-Economic Processes, Urban Structure Formation and Traffic Dynamics*. London: Gordon and Breach.
- Schweitzer, F. (2019). “The bigger picture: complexity meets systems design,” in *Design. Tales of Science and Innovation*, eds G. Folkers and M. Schmid (Zurich: Chronos Verlag), 77–86.
- Seidman, S. B. (1983). Network structure and minimum degree. *Soc. Netw.* 5, 269–287. doi: 10.1016/0378-8733(83)90028-X
- Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Zhang, Y., Garas, A., and Schweitzer, F. (2016). Value of peripheral nodes in controlling multilayer scale-free networks. *Phys. Rev. E* 93:012309. doi: 10.1103/PhysRevE.93.012309
- Zhang, Y., Garas, A., and Schweitzer, F. (2019). Control contribution identifies top driver nodes in complex networks. *Adv. Complex Syst.* 22:1950014. doi: 10.1142/S0219525919500140

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Casiraghi and Schweitzer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.