**RESEARCH**

CrossMark

# Controllability of social networks and the strategic use of random information

Marco Cremonini[1]* and Francesca Casamassima[2]

*Correspondence:
marco.cremonini@unimi.it
[1] University of Milan, Via
Bramante 65, Crema, CR, Italy
Full list of author information
is available at the end of the
article

## Abstract

**Background:** This work is aimed at studying realistic social control strategies for social networks based on the introduction of random information into the state of selected driver agents. Deliberately exposing selected agents to random information is a technique already experimented in recommender systems or search engines, and represents one of the few options for influencing the behavior of a social context that could be accepted as ethical, could be fully disclosed to members, and does not involve the use of force or of deception.

**Methods:** Our research is based on a model of knowledge diffusion applied to a time-varying adaptive network and considers two well-known strategies for influencing social contexts: One is the selection of few influencers for manipulating their actions in order to drive the whole network to a certain behavior; the other, instead, drives the network behavior acting on the state of a large subset of ordinary, scarcely influencing users. The two approaches have been studied in terms of network and diffusion effects. The network effect is analyzed through the changes induced on network average degree and clustering coefficient, while the diffusion effect is based on two ad hoc metrics which are defined to measure the degree of knowledge diffusion and skill level, as well as the polarization of agent interests.

**Results:** The results, obtained through simulations on synthetic networks, show a rich dynamics and strong effects on the communication structure and on the distribution of knowledge and skills.

**Conclusions:** These findings support our hypothesis that the strategic use of random information could represent a realistic approach to social network controllability, and that with both strategies, in principle, the control effect could be remarkable.

**Keywords:** Social network, Adaptive networks, Controllability, Random information, Metrics

## Background

*Structural controllability* of networked systems [1] has been extensively studied since the end of the past decade, following the growing interest in network science [2]. The property of structural controllability is central in the study of how the dynamics of a complex system can be controlled [3]. In short, from control theory, a dynamic system is said to exhibit structural controllability if, with a suitable selection of inputs, it can be driven from one state to any other state in finite time. Inputs to the system

are represented by *driver nodes* receiving external perturbations. Liu et al. in a seminal paper demonstrated how the problem of determining the minimum set of driver nodes required for structural controllability can be mapped into a maximum matching problem [4]. Some remarkable theoretical results have been recently demonstrated for complex networks [5–8].

In complex physical, biological, and social networks, the same goal of controlling the behavior of large populations of interacting entities has been studied with models of *pinning control* regulating networks of coupled dynamical systems. In this case, nodes are dynamical systems coupled according to the network topology. Direct actions are implemented with certain strength over a subset of selected *pinned nodes*, the reactions of which propagate into the network [9, 10]. The notion of *pinning controllability*, conceptually similar to the structural controllability already mentioned, has been introduced by specifying a reference temporal evolution of the network and two parameters: a coupling gain and a vector of control gains [11].

*Adaptive* (or *coevolving*) networks are models of complex networks characterized by a mutual interaction between a time-varying network topology and nodes' own dynamics [12]. These are networks whose behavior is mostly driven by the coevolution of two feedback loops: Node dynamics modify the network topology by creating or removing edges (or changing their weight) and, at the same time, the modified topology influences the dynamics of nodes. Many real-world complex networks exhibit the coevolution between topology and node dynamics, such as in epidemics [13, 14], opinion and community formation [15], or socioeconomic networks [16] and relations among investors [17].

Adaptive networks represent the conceptual framework for this work, which is based on a network model of knowledge diffusion that presents a coevolution between nodes and network topology [18]. Nodes interact based on some local preferences and on a question–answer protocol driven by the difference of some key attribute values between neighbors. Interaction between nodes modifies the network topology, which assumes non-trivial properties. In its turn, the dynamically evolving topology changes neighborhood relations among nodes. It is in this context that we have studied the practical limitations affecting the selection of the theoretically best suited driver nodes—as we will discuss, under certain circumstances it is possible that the only realistic strategy is the opposite of what structural/pinning controllability theory would suggest—and the type of actions that could be realistically performed in a social context for controlling an adaptive network.

A relevant theoretical contribution to the problem of controlling adaptive networks, which takes into account physical and economic constraints, has been recently presented [19]. *Structural permeability* is the metric of the propensy of a network to be controlled. Specific of structural permeability is that the optimal selection of driver nodes is analyzed by considering that only some nodes are admissible as driver nodes and some other nodes should not be directly perturbed by control actions. This theoretical framework permits to generalize different problems, including ours (called *problem 2* in [19]).

However, demonstrating structural/pinning controllability for a social network is neither always necessary nor sufficient when realistic scenarios are considered [20, 21]. It is not strictly necessary because often it is not required to be able to drive a social context

from an arbitrary initial state to a specific final state. In most practical situations, there is the need to tune the dynamical evolution from one trajectory driving the population towards a negative outcome to another trajectory, possibly unknown but leading to a better outcome with respect to the welfare. So, in many practical situations, it is not an optimization problem that we should solve when social networks are considered (i.e., applying the best inputs to the smallest set of driver nodes in order to reach the optimal final state), rather it is a problem of finding heuristics for maximizing the welfare. A possibility is to perturb the system evolution for modifying the *basin of attraction* (i.e., modify the dynamics so that the system that was attracted towards a certain space of final states becomes attracted towards a different one) [22, 23]. The property of structural controllability is also not sufficient as a requirement, because often there are practical limitations to the type of perturbations that could be injected into the system through driver nodes and also limitations to the accessibility of driver nodes. In many real cases, we are neither free to choose the best inputs nor to observe and manipulate all agents. Most important of all constraints, the individuals subject of the social control observe, judge, and possibly react to the control measures if they are perceived as unethical, unfair, abusive, or oppressive. The long experience with advertising campaigns that ultimately produced adverse reactions (i.e., reactance) [24] or the many criticisms concerning the lack of ethics in the social experiment run by Facebook [25] are well known. The human individual and social rich behavior represents the fundamental difference between control theory for generic complex systems and a control theory for social networks [26].

## Goal and motivations

Different from structural permeability, in our work we consider two peculiar constraints derived from empirical observations that fundamentally change the approach. The first is on the *nature* of admissible direct actions operated on driver nodes, while the second is on the *possibility to engage* optimal driver nodes. With respect to the nature of direct actions, we assume that in a social context no explicit direct action controlling the behavior of individuals is possible unless a contract is enforced, like establishing a sponsorship or signing for a professional service. Indirect actions are instead admissible by exposing individuals to new information, either random information or recommendations, which may or may not produce a change in future activities. We have considered the effects of random information enlarging the knowledge base of driver nodes as a control strategy. With the second constraint, we question the assumption, common in structural/pinning control and network permeability, that the minimal and optimal set of driver nodes could be practically engaged for network control actions. We observe instead that there are situations where this is not the case, one of the more common is when driver nodes charge a fee for propagating control actions to their neighbors, for example if they are celebrities or so-called 'influencers' in a social network because of their centrality. In these cases, the price for enrolling a driver node could be high and budget constraints could restrict their engagement to a suboptimal number. It is even conceivable that the price tag that network celebrities would impose could be so high that in practice none of the driver nodes that control theory would prescribe can be actually enrolled. On the contrary, the opposite strategy

could become feasible: To enroll as driver nodes a large number of nodes with limited ability to propagate control actions, such as common users and lay people, which however may have unit cost almost irrelevant and could be addressable in bulk through mass campaigns. This is the common scenario for propaganda and mass manipulation through a popular media and, anecdotally, a strategy that has proved effective in several historical occasions.

The considerations regarding the limitations of structural controllability are the core motivations for the present paper that extends and improves a previous one, in which we had set the basis for the analysis and provided an initial discussion of our hypothesis for a realistically acceptable control of social networks [27]. Here we focus more explicitly on the two well-known strategies for influencing social contexts: the one based on few 'influencers' and the other based on many ordinary users.

The approaches have been studied in terms of network and diffusion effects, two concepts that we have adapted to our scenario to discuss the rich dynamics produced by the use of random information for social network control. The network effect is analyzed through the changes induced on the network average degree and the clustering coefficient, while the diffusion effect is based on two ad hoc metrics defined to measure the degree of knowledge diffusion and skill level and of the polarization of agent interests [28].

More specifically, randomness in agent behavior has been modeled as a new topic exogenously inserted in agents' state during a simulation: This event wishes to represent the typical "unsought encounter" of serendipity and modifies an agent's criterion of choice of the peers in knowledge diffusion and communication. A first study regarding how to adapt the serendipity concept to a social network model has been presented in [29].

Finally, the ultimate goal of our work is to demonstrate that exposing selected agents to random information could be practically used to modify the evolution of a social network, nudging it towards behaviors that improve the welfare, and that different strategies to achieve that goal could be realistically elaborated. Examples of problematic behaviors spontaneously emerging from social networks abound, for which the ability to exert a control on the dynamics could be justified. It is well known and documented that in several situations agents exhibit a strong *polarization* [30, 31]. Agents often form tight communities with few, if any, weak ties between them, and heterogeneity of traits and characteristics of connected agents tends to disappear with the increase of homophily [32, 33]. In practice, it was often observed how people on social networks tend to slide into "filter bubbles" [34, 35]—i.e., self-reinforcing social contexts dominated by information homogeneity with few occasions to have contacts with critical analyses, information from unaligned sources, or contrasting opinions—or, in knowledge diffusion and networked learning, how knowledge sometimes diffuses unevenly, for instance agents showing strong tendency towards specialization like when students/recipients of information grow strong interests only in a narrow set of topics disregarding the richness of the full information spectrum [36, 37].

## Model

Our model of social network is inspired by question–answer networks [38] where knowledge is shared from expert agents (with respect to a certain topic) answering questions received from less skilled ones. We assume a set of agents and a set of *topics* to be given. Each agent has a certain level of *interest* and *skill* on each topic, and both parameters could change when interacting with other agents.

More formally, let $N$ be the set of nodes in the network, $T$ be a set of topics, and $K$ be a set of time steps. Let 1 and $|K|$ indicate the first and the last time step, respectively. For each $i \in N$, $t \in T$, and $k \in K$, let $l_{it}^k \geq 0$ and $s_{it}^k \geq 0$ be the level of interest and skill, respectively, on topic $t$ for node $i$ at time $k$. For each $i, j \in N$ and $k \in K$, let $x_{ij}^k \geq 0$ be the amount of interaction between nodes $i$ and $j$ at time $k$, expressed in terms of the number of question–answer interactions.

Each agent $i$ has a state characterized by two state vectors called *Personal state* $\mathrm{PS}_i^k$ (what $i$ knows at time step $k$) and a *Friend state* $\mathrm{FS}_i^k$ (who $i$ knows at time step $k$). For each $i \in N$, $t \in T$, and $k \in K$, the *Personal state* has the form $\mathrm{PS}_i^k = \left( \bigcup_{j \in T_i^k} \left( t_j, l_{it}^k, s_{it}^k \right) \right)$, where $T_i^k \subseteq T$ is the subset of topics that agent $i$ knows at time step $k$. In other words, the *Personal state* holds the list of topics an agent knows together with the corresponding level of interest and skill. The *Friend state* has the form $\mathrm{FS}_i^k = \left( \bigcup_{j \in N_i^k} \left( x_{ij}^k \right) \right)$, where $N_i^k \subseteq N$ is the subset of agents with whom agent $i$ had interacted up to time step $k$. For each peer $j$ of agent $i$, the number of question–answer interactions $x_{ij}^k$ up to time step $k$ is recorded.

Together, Friend states define the time-varying adjacency matrix $\mathscr{A}(k)$ of the directed and weighted graph $\mathscr{G}(k) = \{\mathscr{V}, \mathscr{E}^k\}$ describing the network at time step $k$, where $\mathscr{V} := N$ is the set of nodes and $\mathscr{E}^k = \bigcup_{i \in N} FS_i^k$ is the set of weighted edges recorded by nodes' Friend states. The element of the adjacency matrix $\mathscr{A}(k)$ is computed as follows:

$$a_{ij}(k) = \begin{cases} x_{ij}^k & \text{if } (i,j) \in \mathscr{E}^k \\ 0 & \text{otherwise} \end{cases}$$

### Network construction

A network is dynamically created according to the following steps:

1. *Setup* At setup (time step 0), for each agent $i$, the *Personal state* $\mathrm{PS}_i^0$ has a number of topics $T_i^0$ and corresponding qualities $s_{it}^0$ selected randomly, while the level of interest $l_{it}^0$ is the same for all topics $T_i^0$. The *Friend state* $\mathrm{FS}_i^0$, instead, is empty, because we do not assume any preset network structure.
2. *Topic selection* At each time step, an agent $i'$ is randomly selected and for that agent a certain topic ($t^* \in T_{i'}^k$) is selected from its Personal state. The choice of the topic is a weighted random selection with values of the associated interests $l_{i't^*}^k$ as weights; this way topics with higher interest are more likely to be selected.
3. *Peer selection* Among agent $i'$ friends and friends-of-friends holding topic $t^*$, the agent $i''$ with the higher skill in topic $t^*$ is selected ($max\ (s_{jt^*}^k \forall j \in \mathrm{FS}_{i'}^k \wedge \mathrm{FS}_{\mathrm{FS}_{i'}^k}^k)$).
4. *Successful communication* The communication between the requester agent $i'$ and the selected peer $i''$ succeeds if the peer is more skilled than the requester in topic $t^*$. The condition for establishing a communication involves the skill parameter: $s_{i''t^*}^k > s_{i't^*}^k$.

5. *Failed communication, random selected peer* Otherwise, if either Step 3 or 4 fails (i.e., no peer holds topic $t^*$ or the selected peer is less skilled than the requester), then select an agent $i'''$ at random.
6. *Communication with random selected peer* If the randomly selected node holds the selected topic and has a skill greater than that of the requester agent, the communication succeeds (*WHILE* ($t^* \in T_{i'''}^k$ *AND* $s_{i'''t^*}^k > s_{i't^*}^k$) *Communication Succeed*).

For sake of clarity, it is worth noting that the inclusion of friend-of-friend agents in Step 2 (*Peer selection*) is key for network transitivity and the closure of triangles, two peculiar characteristics of social networks.

The network formation mechanism at *start-up*, instead, with agents having no friends, is purely random. In particular, considering the six steps just described, a communication attempt always fails at start-up because there is no peer to select; Step 5 (*Failed communication, random selection*) and Step 6 (*Condition for communication with random selection*) are then executed and the initial ties are created by those interactions with randomly selected peers that succeed.

### Skill and interest update

After a successful interaction, the agent that started the communication is updated. For simplicity, no change in the respondent's state is produced, assuming that knowledge, being an intangible good, does not decrease when shared, and there is no cost of processing and transmission. *Skill* and *interest* are always non-negative quantities. For the skill parameter associated to each topic an agent owns, we assumes that it simply increases in chunks calculated as a fraction of the knowledge difference between two interacting agents. This implies that in subsequent interactions between two agents the less skilled one accumulates knowledge in chunks of diminishing size. In this simple model of knowledge transmission, skill is a monotonic positive function with diminishing marginal increments. For completeness, although not specifically relevant for this work, the model also considers the effect of *trust* as an enabling factor for sharing knowledge: the more the trust between the two agents, the better the diffusion of knowledge. Here the assumption is that trust between two interacting agents is built with successful communication. In this case, when agents interact for the first time, the chunk of knowledge transferred is reduced by a discounting factor representing the absence of trust. This discount factor progressively vanishes with subsequent communications. This is a simplified form of trust (and distrust) modeling, but motivations could be found in the literature about collective behavior [39, 40] and refers both to the prevalence of ego-centrism in assimilating new information and to trust dynamics.

The function modeling how agent $i'$ improves the skill associated to a certain topic $t^*$ by interacting with agent $i''$ is as follows:

$$\delta s_{i't^*} = \frac{s_{i'',t^*} - s_{i',t^*}}{\gamma + \rho e^{-\frac{v}{\theta}}}, \tag{1}$$

where

- $s_{i'',t^*} - s_{i',t^*}$ is the difference of skill level associated to topic $t^*$ between agent $i'$ and $i''$;
- $\gamma \geq 1$ is the control factor for the size of the chunk of knowledge $\delta$ *skill* that agent $i'$ could learn from agent $i''$;
- $\rho e^{-\frac{\nu}{\theta}}$ is the discount factor representing trust;
- $\nu = x_{i'i''}^k$ represents the number of successful communications between agent $n_{i'}$ and agent $n_{i''}$ at time step $k$;
- $-\frac{\nu}{\theta}$ controls the slope of the trust function between the two agents with $\theta$ being the control factor; and
- $\rho$ is the control factor for the actual value of the trust component.

In short, the skill function says that one agent improves its skill with respect of a certain topic by learning from someone more skilled. The improvement is always a fraction of the difference of skills between the two and this passage of knowledge could be influenced by trust relations. From this derives the monotonicity with decreasing marginal gains.

The dynamics we have assumed for the *interest* associated to the topic for which the interaction takes place is slightly different with respect to skills. The difference is the assumption of *bounded total interest* an agent could have. In other words, one cannot extend indefinitely the number of topics he is interested in or the amount of interest in a certain topic, for the simple reason that time and efforts are finite quantities. Besides the common sense, motivations for this assumption can be found in cognitive science studies, which have shown the tendency of people to shift their attention and interest, rather than behave incrementally [40], and in associating the interest for a topic to the time spent dealing with that topic (studying, experimenting, etc.).

With respect to the model and how agent's state parameters are managed, this assumption means that either interests in topics are simply assumed as fixed quantities (a grossly unrealistic assumption) or when the interest associated to a certain topic increases, some other interests associated to all other topics should decrease. We have modeled this second case letting interests dynamically change, and for simplicity we assume that the increase of interest in one topic is compensated by uniformly decreasing all other interests. A full discussion of these model assumptions has been presented in [36].

The function modeling the interest function, for agent $i'$ and associated to topic $t^*$, depends on the gain in skill defined by function (1), meaning that the more one learns the more he is interested to learn with respect to that topic:

$$\delta l_{i',t^*} = \alpha \left( 1 - e^{-\frac{\delta s_{i',t^*}}{\beta}} \right) \tag{2}$$

with $\alpha > 1$ and $\beta > 1$ being the two parameters that control, respectively, the scale and the slope of the interest function, which, again, present diminishing marginal increments. A specific analysis of self-organization strategies based on a manipulation of these two parameters could be found in a previous work [36].

Finally, as mentioned before, the increase of a specific interest should be compensated by an equal decrease proportionally spread among the other interests.

Together, these skill and interest functions produce a rich dynamics with regard to knowledge diffusion on the network. In particular, agents acquire different roles with respect to communication, with a core of tightly connected agents and hubs, and a large periphery. The network tends to form a giant component, but specific setups may break it into several independent ones. In general, the diffusion of knowledge produced is typically uneven and locally skewed on just few topics. These are the basic characteristics often debated for real case studies with populations showing a strong tendency to polarization of interests, echo chambers, and forms of cultural isolation even in the presence of a social network with a high degree of connectivity, as the online space now provides.

These are also the key observations that have driven this work, which is focused on studying how, realistically, a social network with the characteristics just presented could be controlled in order to govern the process of knowledge diffusion, if the goal is to achieve a less skewed distribution of skills, a larger pool of interests shared by agents, and a more effective knowledge diffusion process.

## Control strategies with random information

### Goal and motivations

Based on the model of social network described so far, in this work we study ways to control the dynamic behavior resulting from simulations. Differently from traditional control theory that defines controllability of complex systems as the ability to drive a system from an arbitrary initial state $A$ to an arbitrary final state $B$ [1, 3], for knowledge diffusion in a social network, explicit control mechanisms are typically considered unacceptable, unethical, or severely questioned by the public opinion. This is the case of propaganda, media manipulation, or mass indoctrination. In a social network, the notion of controllability must necessarily be weaker and more nuanced than in control theory: It should be an action of nudging the system dynamics with the goal of improving the welfare, rather then an action of tuning a set of inputs that drive a system from any initial state to any desired final state within finite time [4]. In this vein, an external controller has a limited set of mechanisms and those mechanisms must be made explicit to the population; otherwise ethical concerns are likely to prevail or the recipients of those influencing attempts could revolt against them, as sometimes happened for advertising [24]. With respect to the goal of a control strategy in social networks, it is unlikely to be as precise as in control theory, which is assumed to be able to drive a system from one arbitrary state to any other one. Rather, the realistic goal could be to adjust some macro-characteristics of the system, expressed by some *metrics*, in order to drive the social network from one *basin of possible states* to another [21, 22]. This is, for instance, the way public policies are produced and enforced by public authorities for health, education, employment, or safety reasons.

This difference is important, because it permits to discriminate control strategies that are not realistic from those that could possibly be considered acceptable in a social context. Acceptability of a control in a social context depends on both the *control mechanism* (i.e., how the control is exerted) and the *control goal* (i.e., what the control wishes to achieve). In the former case, we exclude from our analysis all mechanisms that involve the use of force upon a population and all social controls based on deception and

trickery. There could be exceptional situations when the use of force is acceptable (e.g., for riots, wars, catastrophes, etc.), but they are out of the scope of this work. In our scenario, a control strategy should always be announced as a public policy to the involved population. For the latter aspect, the goal of the control in a social context should be aimed at enforcing a public policy to improve the welfare, like reducing inequalities, improving education or social inclusion, or the access to public services.

More specifically, we assume that a social control strategy is aimed at modifying the dynamics of a social network to drive it from a certain *expected basin of resulting states*, if the network is left free to evolve for a certain time and based on present configuration, to a different, *desired basin of final states* that increases the welfare.

In our case study, the control goal is to achieve a *better distribution of knowledge among the agents* by reducing polarization, augmenting the average skill level and the average number of topics held by agents.

The control mechanism, which we consider acceptable in practice, is the *strategic use of random information*. Selected agents will be exposed to new random information at certain time steps. The hypothesis is that the controlled exposure to random information, which would not happen if the network is left free to evolve autonomously, can be opportunistically managed to nudge the network evolution. Network metrics and some ad hoc diffusion metrics are monitored to evaluate the effectiveness of a control strategy. Examples of a similar idea applied to different contexts can be retrieved in recommender systems [41, 42] and search engine studies [43, 44], which have the related problem of producing recommendations or search results too heavily influenced by the profile of the requester.

## Network and diffusion metrics

To evaluate the effectiveness of control strategies, we consider how exposing a subset of agents to random information may influence some basic metrics. This represents our interpretation of *acceptable control mechanism applied to a social network*: some key parameters of the network topology and its dynamics are influenced in a predictable way and change the original evolution trajectory. The structure of the network might be modified, as well as the relevance of some agents with respect to the connectivity or the diffusion of knowledge; the polarization of agents and the average number of topics that agents learn from interactions with others could also be reduced. In short, an acceptable control mechanism should be able to influence in a predictable way (at least stochastically) the coevolution of the adaptive network.

In this work, we evaluate two classes of metrics: *network metrics* and *diffusion metrics*. Among the many network metrics defined in the literature, we focus the analysis on the *average degree* and the *clustering coefficient*. Together, these two metrics provide information about the network topology and the communication infrastructure. For instance, the type of network can be inferred, as well as the connectivity distribution between the core and periphery.[1] Some agents typically become hubs and exert a strong influence

---

[1] Network *core* and *periphery* are typically defined in terms of assortative mixing. For assortative networks, as social networks typically are [45], high-degree nodes tend to link together forming a group of higher mean degree than the network as a whole. This subnetwork is called the core. Low-degree nodes, instead, being the majority of nodes, form a generic periphery of lower mean degree than the core [46, 47].

over the communication and distribution of information in the network. The ability to modify network metrics such as the average degree and the clustering coefficient has an influence over the hub agents; for instance, their prominence could decline in favor of a more decentralized communication.

The second class of metrics is defined ad hoc. We have called them diffusion metrics because their aim is to measure some important attributes of the spread of topics and interest among the agents. Two metrics have been defined: *Average Knowledge* (*AK*) and *Knowledge Diffusion* (*KD*), which are used to evaluate different characteristics of knowledge diffusion dynamics.

*Average Knowledge* is calculated as the *average skill with respect to the topics actually owned by agents*. In this sense, *AK* contributes to measure the specialization/polarization of agents, because *AK* could be high when agents know few topics with high skill levels.

*Knowledge Diffusion*, instead, is the average skill of agents *with respect to the total number of topics in the network*. This means that, implicitly, a topic that an agent does not own is averaged with skill zero. The meaning of *KD* is to evaluate the actual diffusion of knowledge with respect to the case of perfect diffusion, when all agents hold all topics with maximum knowledge. In this sense, *KD* is relevant because it gives an indication of how effectively topics are spread over the population and with which skill level.

$$AK = \frac{\sum_{i=1}^{N} \sum_{j=1}^{|PS_i|} s_{i,j}}{\sum_{i=1}^{N} |PS_i|} \qquad (3)$$

$$KD = \frac{\sum_{i=1}^{N} \sum_{j=1}^{|PS_i|} s_{i,j}}{N \times T} \qquad (4)$$

with $|PS_i|$ being the number of topics of Personal state of agent $i$' and $T$ being the total number of topics in the population.

## Social control strategies

For the social network control based on random topics, we assume that, with a given frequency, a certain subset of agents is selected and a variable number of random topics, among the ones existing in the network, are inserted into their Personal state with an average level of interest. The goal is to force the exposure of some agents to a larger set of topics and exploit the mechanisms of topic and peer selection described in the "Network construction" section to modify the dynamic network behavior. With regard to network structure, with new random topics, the network construction mechanism strictly based on topology is occasionally bypassed and a bridge effect is produced.

Recalling the terminology typical of control theory, we call *driver agents* the agents selected for being exposed to new random topics. The key parameters for defining a social control strategy are *selection criteria of driver nodes*, *number of random new topics introduced*, and *frequency of introduction of new topics.*

The simulations and results discussed in the following will be based on different choices of these parameters. We made some assumptions with respect to the parameter

choice. For the selection criteria of driver agents, in this work we focused on just two particularly relevant options: driver agents are selected as a fraction of total agents *ranked based on the node degree, either in decreasing on increasing order*.

### Selection of high-degree driver agents

In this case, driver agents are a subset of agents with higher degree (1 or 10% is the value used in tests). This means that agents exposed to new topics and whose behavior will be directly perturbed are those that mostly centralize the communication in the network. They are the hubs and the nodes of the network core, typically densely connected. Therefore, with this choice, very intuitive and similar to strategies adopted in social media and advertising (e.g., targeting the network influencers), it is the core of the network and the hub nodes to be directly manipulated. With respect to our metrics, we expect that few nodes will have a strong influence over the network dynamics and that hub nodes will increase their degree. Overall, though, this should stimulate a better communication, with positive effects on knowledge diffusion.

### Selection of low-degree driver agents

In the other case, instead, driver agents are chosen among those with lower degree. This option may seem counterintuitive, being these agents the least connected nodes of the periphery of the network, those that typically have the least influence on communication and spreading processes. It has, however, some interesting properties. At network level, targeting the periphery of the network means to directly stimulate the closure of more triangles. This means to raise the clustering coefficient and foster a more rich decentralized communication, which may have effects on the network topology. Regarding diffusion, these nodes are likely those that had fewer connections, typically connected with other low-degree nodes (i.e., social networks are typically assortative [48], although some recent experiments have shown that during the evolution a network could transit from assortative to disassortative [49]), and thus they have a limited choice of peers for exchanging knowledge. Exposing them to random topics very likely changes their pattern of interaction, much more than what happens to hub nodes, which were and remain highly connected. On the other side, the influence that a typical low-degree node has on communication is very limited, and for this reason the number of low-degree agents that must be selected to produce a clear control effect on the network is much higher than that in the previous case. In our examples, we have shown results with 50 and 70% of agents selected as driver agents. The choice of 50 and 70% is qualitative and motivated by presentation simplicity. The goal here is to show that, in principle, targeting ordinary users instead of influencers may produce relevant effects on the coevolving network dynamics, but in one case the driver nodes are a small subset with respect to the whole network (potentially requiring a high price to be engaged), and in the other case the number of driver nodes is on a different scale, likely a large subset of nodes (possibly with low unit cost).

### Number of new topics

The second important parameter for defining a social control strategy represents the amount of perturbation that we are introducing in driver agent's Personal state. Instead

of an actual number of topics, it is expressed as a rate of new topics a node will receive to simulate the control action. Typical values we have considered are 1, 10, or 30%, defined empirically as simply the rate that, on average, gives just one new topic (i.e., 1%) and those that permit to incrementally add new topics during a simulation. For simplicity, we have assumed that all driver agents receive the same rate of new topics. Alternatives would have been possible, like simulating fixed amounts of new topics (we have tried this approach, but preliminary results were not as interesting as for the fixed rate option), or adjusting the number/rate of new topics based on contextual information for each agent (we did not explore this possibility, because it further complicates the model and the dynamics, but it is promising for developing a "personalized control strategy").

### Frequency of new topics

The considerations for the frequency of introduction of new topics are similar to those discussed for the number of new topics. In this work, for simplicity we have considered fixed time intervals (measured in number of simulation steps, called *ticks*). 1000 ticks is the value used in the examples, for simulations that typically span over 100,000 ticks. As in the previous case, the value is selected empirically, mostly for the sake of presentation. Again, similar to the previous case, more elaborate time-based strategies were possible and interesting to study. The action of inserting new topics, instead of being a fixed frequency, could be triggered by contextual information, like a particular state of the network or particular values of the metrics, for instance when the overall communication stagnates. Another, even more elaborate strategy could be to trigger the addition of new topics at agent level, when the specific driver agent exhibits a particular state or behavior (again, another form of "personalized control strategy").

## Model instances and simulation results

### Basic settings

The model of knowledge diffusion with social control mechanism is implemented in Netlogo,[2] a multi-agent programmable modeling environment, which has a proprietary script language and permits to run simulations of complex networks through the GUI or the API.

All simulations we present here have the basic setting listed in Table 1 (*Parameters shared by all instances*), chosen mainly for presentation sake, but still representative of network behaviors observed during tests with a large number of settings.

For the dynamics produced by our model, the rate between the number of agents and the number of topics $\frac{N}{T}$ is key, because it directly influences the average number of topics the agents are assigned at setup and the ability to select peers. It is to simplify the analysis and the presentation that we fixed the number of topics $T$ and the maximum number of topics assigned to agents at setup $\lambda_T$, and varied the number of agents $N$ to analyze, at least partially, how larger populations affect the results [see Table 1 (*Network sizes*)]. In this study, we focused on small- to medium-sized populations, possibly representing situations like network learning for traditional academic courses (or even local subgroups

---

**Table 1  Parameters for the model instances**

| |
|---|
| Parameters shared by all instances |
|   Total number of topics in the system: $T = 100$ |
|   Max. number of topics per agent at setup: $\lambda_T = 10$ |
|   Max. number of time steps (max #ticks): $|K| = 100{,}000$ |
|   Number of new random topics per driver agent (% over $T_i$): 30% |
|   Time interval of new random topic insertion (in #ticks): 1000 |
| Network sizes |
|   Number of agents: $N = [100, 200, 500, 1000]$ |
| High-degree agent selection |
|   Number of driver agents (% over $N$): [1, 10%] |
| Low-degree agent selection |
|   Number of driver agents (% over $N$): [50, 70%] |

of participants to large online courses interacting in dedicated forums), or online social networks built around restricted cultural or professional contexts. Very large communities represent a structurally different case study, which needs a specific analysis, out of the scope of this study. The duration of simulations, defined as the maximum number of time steps $|K|$, instead, has been decided empirically as the number of time steps sufficient for networks produced by our model to reach a steady state.

### Model instances

Model instances presented in this work are aimed at discussing how the two strategies for driver agent selection, one selecting a subset based on agents with higher node degree and the other based on agents with lower node degree, could be exploited for controlling the network behavior. The two options have been compared also with the case of network evolving without any perturbation. The choice of node degree as the attribute for selecting driver agents has several references in control theory literature and is an attribute typically analyzed in social network studies. Several alternatives to degree centrality exist for selecting driver agents, like betweenness and eigenvector centrality, or more elaborate criteria based on community detection. However, we feel that before exploring alternative strategies and study which one could optimize the outcome, the goal should be to acquire a better understanding with respect to the possibility to use random information as a social control mechanism in practical situations. The simple criteria based on node degree have the important benefit of being intuitive in practice (i.e., selection of leaders/influencers or of common users/participants) and already exploited for different applications.

Specifically, for the case of high-degree nodes, driver agents are either 1 or 10% of the population [see Table 1 (*High-degree agent selection*)]. For low-degree nodes, driver agents are either 50% or 70% of the population [see Table 1 (*Low-degree agent selection*)]. As discussed previously, the different characteristics of high- and low-degree nodes with respect to connectivity and controllability make the comparison of equally sized subsets not meaningful.

### Network and diffusion effects

With simulations, we wish to show two distinct effects that together contribute to define the effectiveness and feasibility of a social control strategy based on random topics: the *network effect* and the *diffusion effect.*

Network effect describes how the addition of new random topics permits to modify the topology structure of the social network. This effect is one of the most fundamental for a social control strategy, because by interfering with the process that produces the network topology some critical parameters of the network structure are affected, like the degree and type of assortativity, the size of independent components, the weight and number of hub nodes in centralizing the overall communication, and the relation between the core and the periphery of the network. From these parameters, many emerging behaviors and features depend, for instance, on the community structure together with the prevalence of segregation or of echo chambers, the robustness with respect to failures, attacks, or deception, and the speed and the spread of a diffusion phenomenon [50]. The network effect is measured by typical social network metrics like centrality metrics, homophily degree, density degree, and the analysis of structural holes and of tie strength among the others. In this work, we focused on two metrics among the most basic ones: *average node degree* and *clustering coefficient*, which allow to consider some fundamental aspects of network topology.

With diffusion effect, instead, we mean how the new random topics modify the specific mechanism of knowledge transmission defined by our model. Our knowledge transmission is ad hoc, but it wishes to represent a general social context where agents have different skills and different willingness to acquire new knowledge. Skills are improved based on personal one-to-one contacts supported by the network topology. This diffusion effect is the one interfering with the topic selection mechanism of driver agents, which in turn affects the peer selection criteria and consequently changes how likely is a driver agent to maximize the skills on few topics only or to distribute the interactions on an extended set of topics. Also, the diffusion effect influences the average communication efficiency of agents, because by easing the creation of new interactions, communication congestions and spikes are less likely. We measure the diffusion effect with the two previously defined ad hoc metrics, namely *AK* and *KD*. *AK* represents the average skill level of agents with respect to only the topics they possess. *KD* measures the average skill level of agents with respect to all possible topics. Then, by design, *AK* is an upper bound for *KD*.

In particular, *KD* necessarily starts very low at the beginning of a simulation because in our settings the number of topics assigned to agents at start-up is limited and the total amount of skill is bounded (i.e., in these simulations, on average, agents at start-up have 5% of the total number of topics *T*). Then *KD* monotonically increases with the exchange of knowledge among agents and the diffusion of topics, up to a value *max* (*KD*), which depends on the specific settings. It represents a measure of the *erudition of the social network* and gives an indication of how effectively skills are spread over the population and the level of knowledge. For our analysis, it is important how fast it approaches *AK*.

On the contrary, the meaning of *AK* depends on the runtime network state and it can either increase or decrease. For the same reason *KD* always starts low, *AK* always starts with a value close to the middle of the scale (i.e., at start-up agents have a fixed total

amount of skills spread over few topics). From that point, *AK* measures how much the proportion of the two quantities varies. It remains constant if the two change proportionally during the execution; it increases when agents acquire skill level faster than they acquire new topics and it decreases vice versa.

Given these ad hoc metrics, we say that an agent is *polarized* when it concentrates its best skills on just very few topics. *Polarization of the social network* is another important global characteristic, together with erudition. Taking into account *AK* and *KD* together, the degree of network polarization can be evaluated. We say that *a network tends to polarize* when the difference between *AK* and *KD* is *increasing*, because it means that agents are gaining skills mostly on the same subset of topics without a corresponding increase in erudition. On the contrary, we say that *a network tends to heterogeneity* when the difference between *AK* and *KD* decreases, because it means that agents are no longer selecting a small subset of topics, but are communicating with peers on an extended set of topics.

## Simulation results

Results of simulations are presented in Figs. 1 and 2, which show, respectively, the network and the diffusion effects of the strategic use of random topics.

For the network effect, Fig. 1 presents how the average node degree and the clustering coefficient change with respect to the benchmark represented by the typical network behavior with no injection of random topics in nodes (label **C0** in Fig. 1) for three network sizes ($N = [100, 500, 1000]$).
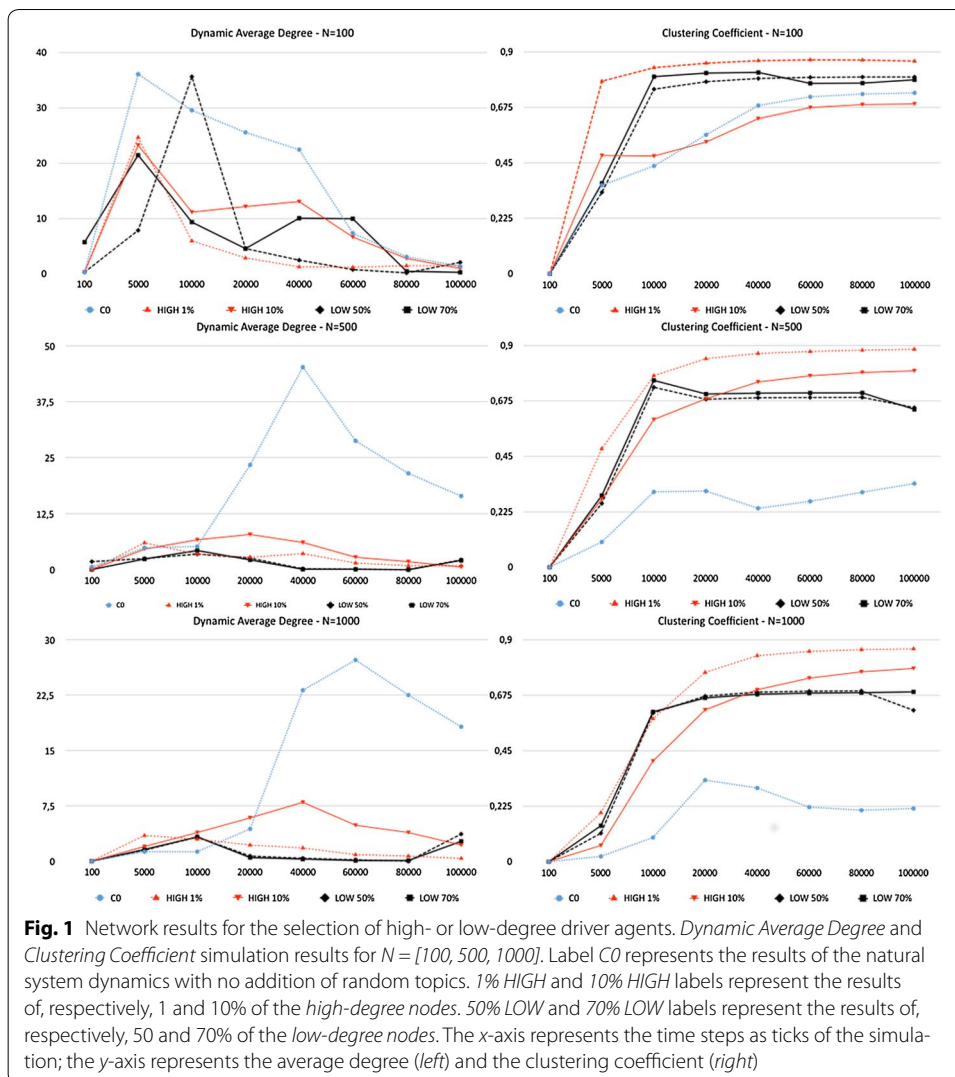
The settings involving the addition of random topics are the ones previously introduced and whose parameters are summarized in Table 1:

- Label **HIGH 1%** represents the selection of 1% of high-degree nodes as driver agents;
- Label **HIGH 10%** represents the selection of 10% of high-degree nodes as driver agents;
- Label **LOW 50%** represents the selection of 50% of low-degree nodes as driver agents;
- Label **LOW 70%** represents the selection of 70% of low-degree nodes as driver agents.

As discussed previously, the two general strategies tested could be interpreted as the selection of influencer agents the one labeled as HIGH and the selection of ordinary users the one labeled as LOW. The choice of the rate of driver agents is mostly for presentation sake. However, we note that the values of the selection of high-degree nodes are aligned with the ones for network controllability of social communication networks analyzed in [4].

### Network effect results

For the *average degree*, the results of Fig. 1 present what is called *dynamic average degree*, a representation specific for dynamic network behaviors, which calculates the average node degree taken on certain time step periods (e.g., in Fig. 1, time step periods are [0,100], [100,5000], [5000,10000], [10000,20000], [20000,40000], [40000,60000],

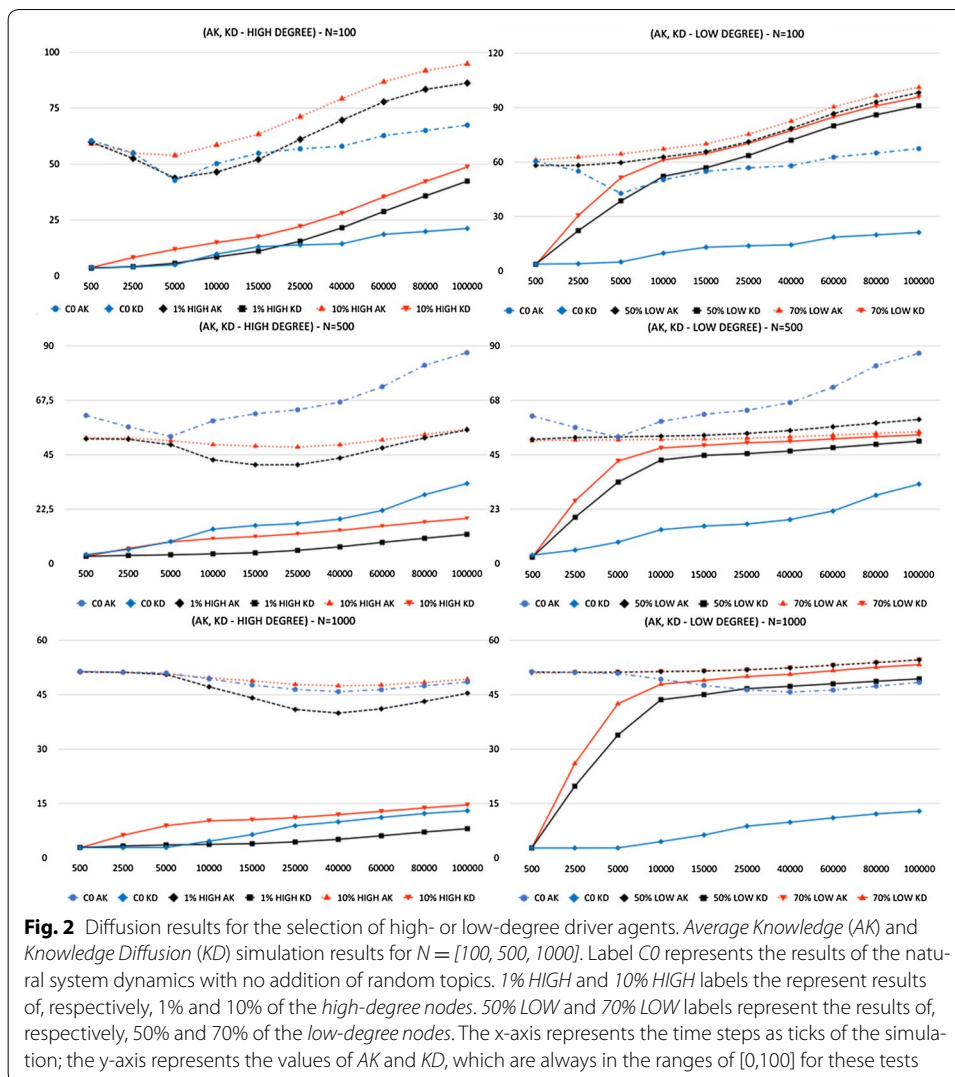**Fig. 1** Network results for the selection of high- or low-degree driver agents. *Dynamic Average Degree* and *Clustering Coefficient* simulation results for N = [100, 500, 1000]. Label *C0* represents the results of the natural system dynamics with no addition of random topics. *1% HIGH* and *10% HIGH* labels represent the results of, respectively, 1 and 10% of the *high-degree nodes*. *50% LOW* and *70% LOW* labels represent the results of, respectively, 50 and 70% of the *low-degree nodes*. The x-axis represents the time steps as ticks of the simulation; the y-axis represents the average degree (*left*) and the clustering coefficient (*right*)

[60000,80000], and [80000,100000]). The value for each time step period is the average degree corresponding to active links (new or reused) in that specific period.[3]

The form of the dynamic average degree function produces a rich representation of network behaviors. One aspect that clearly emerges is the spiky behavior of the original network *C0* that tends to quickly saturate all possible connections. The injection of random topics, instead, in all situations, even for very few driver nodes (e.g., for $N = 100$, the setting *HIGH 1%* means that the node with the highest degree is the only driver agent) is able to produce a sensible reduction of the average degree with respect to the benchmark *C0*.

In tests labeled *HIGH*, driver agents certainly include hub nodes and the large influence these nodes have on the network structure is clearly demonstrated by the fact that

---

[3] The values of the dynamic average degree are not calculated with the traditional formula $< d > = \frac{2*E}{N}$ with being $E$ the number of links and $N$ being the number of nodes in the network. Instead, for each time step period $(t_{i-1}, t_i)$ with $E_{t_i}$ being the number of active links in $(t_{i-1}, t_i)$, the dynamic average degree for each time period is $< d_{t_i} > = \frac{2*E_{t_i}}{N}$.

**Fig. 2** Diffusion results for the selection of high- or low-degree driver agents. *Average Knowledge* (*AK*) and *Knowledge Diffusion* (*KD*) simulation results for *N* = [100, 500, 1000]. Label *C0* represents the results of the natural system dynamics with no addition of random topics. *1% HIGH* and *10% HIGH* labels the represent results of, respectively, 1% and 10% of the *high-degree nodes*. *50% LOW* and *70% LOW* labels represent the results of, respectively, 50% and 70% of the *low-degree nodes*. The x-axis represents the time steps as ticks of the simulation; the y-axis represents the values of *AK* and *KD*, which are always in the ranges of [0,100] for these tests

it suffices to modify the state of a few of them to obtain a wide effect on the degree metric. The average degree drops from 50 to 80%. This is a strong network effect and potentially a powerful way of controlling a network's behavior, because it signals a reconfiguration of the communication within the network.

At first sight, a strong reduction of average node degree might be interpreted as a sensible reduction of the weight of hub nodes. However, inspecting the outcome at agent level, it could be observed that this is not the case. The average degree of hub nodes actually *increases*, which is reasonable given the fact that they receive new topics. The global reduction of average degree is driven by the average degree reduction in the large periphery of the network. The communication is even more centralized by few hub nodes and overall the network tends to become disassortative, similar to the effects discussed in [49].

With respect to the *clustering coefficient*, the tests show a general increment compared to the benchmark. This signals that more triangles have formed, a direct effect of the new connections produced by the introduction of random topics in hub nodes.

Together, the reduction of the average degree and the increase of the clustering coefficient tend to enhance the small-worldness of the network, a characteristic previously not present, which is an important effect in terms of controllability [51, 52].

Considering the different network sizes, we observe that the degree of reduction is systematic, with results being very similar for the larger networks. The case of $N = 100$ represents a peculiar situation because in that setting the communication efficiency is very high and agents establish quickly many connections. This is the reason for the initial spikes of the average degree in all situations.

Another interesting result that emerged from these tests is that with respect to the ability to influence network metrics, both strategies *HIGH* and *LOW* could be effective. However, to have comparable strength, the number of driver nodes must be necessarily very different in the two cases, as shown in our simulations. However, there is an important difference between the *LOW* and the *HIGH* strategy with respect to hub nodes. By selecting low-degree nodes as driver agents, the formation of hub nodes is more difficult and the communication becomes more decentralized and local. The effect on the communication structure of the *LOW* strategy is the opposite of the *HIGH* one. This difference appears very clearly in the following diffusion effect.

In general, these observations present an interesting practical control problem still not fully investigated in the literature. While some theoretical results have been studied for reference network topologies [4, 7, 22], few studies exist that consider mechanisms for dynamically adjusting the degree of control during the evolution of a social network.

### Diffusion effect results

With these tests, we have studied how the control strategies affected the diffusion of topics and of skills in the agent population. Here the goal is to modify the dynamics in order to reduce the polarization and increase the erudition of the network. The results are shown in Fig. 2 and two very different scenarios for *HIGH* and *LOW* strategies are described. With respect to the *HIGH* strategy, the results do not provide evidence of a clear improvement compared with the benchmark. For $N = 100$, there is a slight improvement in the actual values of *AK* and *KD*, meaning a larger diffusion of topics and skills (higher erudition), but the same high level of polarization. For $N = 500$, the outcome is opposite: Actual values are worse than the benchmark, thus a reduced diffusion of topics and skills, but also a smaller level of polarization. For $N = 1000$, the results of both 1 and 10% *HIGH* are very close to *C0* for both *AK* and *KD*. In this regard, selecting the few hub nodes as driver agents does not seem particularly effective for a better diffusion of knowledge measured as increased erudition and reduced polarization.

Opposite is the outcome of the *LOW* strategy. In this case, there appears a systematic improvement with respect to the benchmark. For all network sizes and in both settings, 50 and 70%, *KD* approaches *AK* very fast as a result of the addition of new random topics to many nodes. This is a clear sign of no polarization, meaning that agents are able to distribute their skill over the enlarged set of topics. For $N = 100$, there is also a clear improvement in absolute values of *AK* and *KD*, while for $N = 500$ and $N = 1000$ *AK* grows slowly or tends to stay flat. This behavior signals that skills increase slowly and with difficulty. The reason is that many nodes with low degree receive new random topics, but, after the initial burst of activity with neighbors, most of the time they are unable to find a peer within the

few friends or friends-of-friends. Overall, for $N = 100$ we are able to both improve erudition and have no polarization, the optimal combination for the welfare. For larger networks, instead, adding new random topics quickly removes polarization because locally the communication flows efficiently, but the positive effect is then limited by the fact that with few hub nodes in the network agents are unable to easily find new peers.

To summarize, with these tests we have shown two possible effects of a social control strategy based on the addition of new random topics to selected driver agents. The results are not conclusive, but we have presented some important differences between addressing the influencers or the ordinary users. We have also seen how important it is to define the goal of a control strategy, because the ability of manipulating some network metrics does not necessarily imply to be able to improve context-dependent features like the diffusion of knowledge, and vice versa.

## Conclusions

In this paper, we discussed a possible use of random topics as control inputs for driver nodes, based on an adaptive network model for knowledge diffusion, and two realistic control strategies, one based on high-degree nodes commonly analyzed in studies of structural/pinning controllability and the other based on low-degree nodes, which instead has received few attention in controllability studies, but has practical application in several situations. The idea of using random information in control strategies has a long tradition in control theory and could be also retrieved in recommender systems and search engine research, which share the same problem of excessive polarization of interest and the need to improve diversity that we consider for knowledge diffusion [53, 54]. However, explicitly considering random information as a control input for social networks is a new idea worth exploring, in our opinion. The two general strategies for influencing the coevolving dynamics are based, respectively, on the selection of few influencers and the manipulation of their behavior in order to drive the network, and on the selection of many ordinary users as the driver of the network as a whole. The two approaches have been studied in terms of network and diffusion effects.

The problem of controlling social networks presents striking differences with respect to the study of structural controllability for generic complex networks. The social context introduces many limitations (ethical, operational, functional) and often does not strictly require full structural controllability, rather the ability to nudge the network evolution towards a certain basin of results improving the welfare. For these reasons, the application of control theory to social networks could be a source of inspiration, but it cannot be applied as-is, because important adaptations are required, which represents a promising and still largely unexplored research strand.

There are many situations in which it would be important to know how to handle the level of random information that agents possibly receive, for instance in learning situations, in social media, journalism, knowledge diffusion, skill acquisition, experience dissemination, epidemics, and possibly risk management. In all these situations, there could be the problem of excessive polarization (of interests, attention, analyses) and lack of erudition, but the solution cannot be to simply change what individuals prefer or believe or regard as important/interesting. Increasing information heterogeneity and serendipity could be effective approaches for improving the controllability of social contexts.

Furthermore, considering that in practical situations it could be impractical to either recognize all theoretical driver nodes, accessing them with external perturbations, or sustaining the cost to engage them, we have presented some empirical solutions based on network metrics for selecting nodes that might have practical usage. The results of our work look promising to us and encourage more analyses, tests, and verifications with respect to real social networks. Model Driven Engineering (MDE), aiming to lower the competencies for dealing with complex analytics and automation processes, if applied to social network control could be another possible application scenario [55]. Other future works will include a theoretical grounding to the results, the definition of an analytical model, and a statistical analysis of a larger test set to better evaluate the performance of control strategies. Finally, a socioeconomic study of the conditions leading to prefer a strategy based on high-degree or on low-degree nodes would perhaps produce new insights with respect to practical constraints for adaptive network controllability.

**Author details**
[1] University of Milan, Via Bramante 65, Crema, CR, Italy. [2] Sogetel, Via Conca del Naviglio 18, Milano, Italy.

**Competing interests**
The authors declare that they have no competing interests.

**Availability of data and materials**
Not applicable.

**Consent for publication**
Not applicable.

**Ethics approval and consent to participate**
Not applicable.

## Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**References**
1. Liu YY, Barabási AL. Control principles of complex networks. arXiv preprint arXiv:1508.05384; 2015.
2. Newman M. Networks: an introduction. Oxford: Oxford university press; 2010.
3. Lin CT. Structural controllability. IEEE Trans Autom Control. 1974;19(3):201–8.
4. Liu YY, Slotine JJ, Barabási AL. Controllability of complex networks. Nature. 2011;473(7346):167–73.
5. Chen YZ, Wang L, Wang W, Lai YC. The paradox of controlling complex networks: control inputs versus energy requirement. arXiv preprint arXiv:1509.03196; 2015.
6. Cowan NJ, Chastain EJ, Vilhena DA, Freudenberg JS, Bergstrom CT. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. PLoS ONE. 2012;7(6):e38398.
7. Menichetti G, Dall'Asta L, Bianconi G. Network controllability is determined by the density of low in-degree and out-degree nodes. Phys Rev Lett. 2014;113(7):078701.
8. Yao P, Hou BY, Pan YJ, Li X. Structural controllability of temporal networks with a single switching controller. PLoS ONE. 2017;12(1):e0170584.
9. Grigoriev R, Cross M, Schuster H. Pinning control of spatiotemporal chaos. Phys Rev Lett. 1997;79(15):2795.

10. Wang XF, Chen G. Pinning control of scale-free dynamical networks. Phys A Stat Mech Appl. 2002;310(3):521–31.
11. Sorrentino F, di Bernardo M, Garofalo F, Chen G. Controllability of complex networks via pinning. Phys Rev E. 2007;75(4):046103.
12. Gross T, Blasius B. Adaptive coevolutionary networks: a review. J R Soc Interface. 2008;5(20):259–71.
13. Gross T, D'Lima CJD, Blasius B. Epidemic dynamics on an adaptive network. Phys Rev Lett. 2006;96(20):208701.
14. Marceau V, Noël PA, Hébert-Dufresne L, Allard A, Dubé LJ. Adaptive networks: coevolution of disease and topology. Phys Rev E. 2010;82(3):036116.
15. Iniguez G, Kertész J, Kaski KK, Barrio RA. Opinion and community formation in coevolving networks. Phys Rev E. 2009;80(6):066119.
16. Ehrhardt GC, Marsili M, Vega-Redondo F. Phenomenological models of socioeconomic network dynamics. Phys Rev E. 2006;74(3):036106.
17. DeLellis P, DiMeglio A, Garofalo F, Lo Iudice F. The evolving cobweb of relations among partially rational investors. PLoS ONE. 2017;12(2):e0171891.
18. Ceselli A, CremoniniM. Models and methods for the analysis of the diffusion of skills in social networks. In: Proceedings of international conference in operation research (OR2015). Wien, Austria; 2015.
19. Lo Iudice F, Garofalo F, Sorrentino F. Structural permeability of complex networks to control signals. Nat Commun. 2015;6. doi:10.1038/ncomms9349.
20. Gao J, Liu YY, D'Souza RM, Barabási AL. Target control of complex networks. Nat Commun. 2014;5:5415.
21. Motter AE. Networkcontrology. Chaos Interdiscip J Nonlinear Sci. 2015;25(9):097621.
22. Cornelius SP, Kath WL, Motter AE. Realistic control of network dynamics. Nat Commun. 2013;4:1942.
23. Pennacchioli D, Rossetti G, Pappalardo L, Pedreschi D, Giannotti F, Coscia M. The three dimensions of social prominence. In: international conference on social informatics. Berlin: Springer; 2013. p. 319–332.
24. Tucker CE. Social networks, personalized advertising, and privacy controls. J Mark Res. 2014;51(5):546–62.
25. Kleinsman J, Buckley S. Facebook study: a little bit unethical but worth it? J Bioeth inq. 2015;12(2):179–82.
26. Vardaman JM, Amis JM, Dyson BP, Wright PM, Van de Graaff Randolph R. Interpreting change as controllable: the role of network centrality and self-efficacy. Human Relat. 2012;65(7):835–59.
27. Casamassima F, Cremonini M. Use of random topics as practical control signals in a social network model. In: International workshop on complex networks and their applications. Berlin: Springer; 2016. p. 539–550.
28. Sun PG. Controllability and modularity of complex networks. Inf Sci. 2015;325:20–32.
29. Cremonini M. Introducing serendipity in a social network model of knowledge diffusion. Chaos Solitons Fract. 2016;90:64–71.
30. Centola D, Gonzalez-Avella JC, Eguiluz VM, San Miguel M. Homophily, cultural drift, and the co-evolution of cultural groups. J Confl Resolut. 2007;51(6):905–29.
31. Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S. Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM; 2008. p. 160–168.
32. Golub B, Jackson MO. How homophily affects the speed of learning and best response dynamics. Quart J Econ. 2012;127(3):1287–338.
33. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Ann Rev Sociol. 2001;27(1):415–44.
34. Easley D, Kleinberg J, et al. Networks, crowds, and markets: reasoning about a highly connected world. Significance. 2012;9:43–4.
35. Pariser E. The filter bubble: what the Internet is hiding from you. London: Penguin UK; 2011.
36. Allodi L, Chiodi L, Cremonini M. Self-organizing techniques for knowledge diffusion in dynamic social networks. In: Proceedings of complex networks conference 2014 (CompleNet14). Bologna, Italy; 2014.
37. Maglajlic S. Engineering social networks using the controllability approach applied to e-learning. In: 2012 IEEE 12th international conference on advanced learning technologies (ICALT). Piscataway: IEEE; 2012. p. 276–280.
38. Zhang J, Ackerman MS, Adamic L. Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th international conference on World Wide Web. New York: ACM; 2007. p. 221–230.
39. da Gama Batista J, Bouchaud JP, Challet D. Sudden trust collapse in networked societies. Eur Phys J B. 2015;88(3):55.
40. Goldstone RL, Gureckis TM. Collective behavior. Top Cognit Sci. 2009;1(3):412–38.
41. Adamopoulos P, Tuzhilin A. On unexpectedness in recommender systems: or how to better expect the unexpected. ACM Trans Intell Syst Technol TIST. 2015;5(4):54.
42. Iaquinta L, De Gemmis M, Lops P, Semeraro G, Filannino M, Molino P. Introducing serendipity in a content-based recommender system. In: Eighth international conference on hybrid intelligent systems, 2008. HIS'08. Piscataway: IEEE; 2008. p. 168–173.
43. Bozdag E. Bias in algorithmic filtering and personalization. Ethics Inf Technol. 2013;15(3):209–27.
44. White R. Beliefs and biases in web search. In: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. New York: ACM; 2013. p. 3–12.
45. Newman ME, Park J. Why social networks are different from other types of networks. Phys Rev E. 2003;68(3):036122.
46. Newman ME. Assortative mixing in networks. Phys Rev Lett. 2002;89(20):208701.
47. Newman ME. Mixing patterns in networks. Phys Rev E. 2003;67(2):026126.
48. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Nat Acad Sci USA. 2004;101(11):3747–52.
49. Li M, Guan S, Wu C, Gong X, Li K, Wu J, Di Z, Lai CH. From sparse to dense and from assortative to disassortative in online social networks. Sci Rep. 2014;4:4861.
50. Lu ZM, Li XF. Attack vulnerability of network controllability. PLoS ONE. 2016;11(9):e0162289.
51. Kossinets G, Watts DJ. Empirical analysis of an evolving social network. science. 2006;311(5757):88–90.
52. Newman ME. The structure and function of complex networks. SIAM Rev. 2003;45(2):167–256.
53. Adomavicius G, Kwon Y. Improving aggregate recommendation diversity using ranking-based techniques. IEEE Trans Knowl Data Eng. 2012;24(5):896–911.

54. Bradley K, Smyth B. Improving recommendation diversity. In: Proceedings of the Twelfth Irish conference on artificial intelligence and cognitive science, Maynooth, Ireland: Citeseer; 2001. p. 85–94.
55. Ardagna CA, Bellandi V, Ceravolo P, Damiani E, Bezzi M, Hebert C. A model-driven methodology for big data analytics-as-a-service. In: IEEE international congress on big data, BigData congress 2017, Honolulu, HI, USA, June 25–30, 2017. p. 105–112.