Towards a Social Artificial Intelligence

Dino Pedreschi[2], Frank Dignum[1], Virginia Morini[2,4](✉), Valentina Pansanella[3,4], and Giuliano Cornacchia[2,4]

[1] Umeå University, Umeå, Sweden
dignum@cs.umu.se
[2] University of Pisa, Pisa, Italy
dino.pedreschi@unipi.it,
{giuliano.cornacchia,virginia.morini}@phd.unipi.it
[3] Scuola Normale Superiore, Pisa, Italy
valentina.pansanella@sns.it
[4] KDD Lab ISTI-CNR, Pisa, Italy

Abstract. Artificial Intelligence can both empower individuals to face complex societal challenges and exacerbate problems and vulnerabilities, such as bias, inequalities, and polarization. For scientists, an open challenge is how to shape and regulate human-centered Artificial Intelligence ecosystems that help mitigate harms and foster beneficial outcomes oriented at the social good. In this tutorial, we discuss such an issue from two sides. First, we explore the network effects of Artificial Intelligence and their impact on society by investigating its role in social media, mobility, and economic scenarios. We further provide different strategies that can be used to model, characterize and mitigate the network effects of particular Artificial Intelligence driven individual behavior. Secondly, we promote the use of behavioral models as an addition to the data-based approach to get a further grip on emerging phenomena in society that depend on physical events for which no data are readily available. An example of this is tracking extremist behavior in order to prevent violent events. In the end, we illustrate some case studies in-depth and provide the appropriate tools to get familiar with these concepts.

Keywords: Human-centered AI · Complex systems · Multi-agent models · Social networks · Mobility networks · Financial networks

1 Introduction

Nowadays, given the ubiquity of increasingly complex socio-technical systems - made of interacting people, algorithms, and machines, - the social dimension of Artificial Intelligence (AI) started emerging in our everyday lives. Examples range from urban mobility, with travellers helped by smart assistants, to the public discourse and economic markets, where decisions on what to see or buy are shaped by AI tools, like recommendation and filtering algorithms.

While at the individual level AI outputs could be beneficial, from a societal perspective they can lead to alarming phenomena such as traffic congestion

M. Chetouani et al. (Eds.): ACAI 2021, LNAI 13500, pp. 415–428, 2023.
https://doi.org/10.1007/978-3-031-24349-3_21

[1], radicalisation of opinions [2,3], and oligopolistic markets [4]. The current use of AI systems is often based on the hypothesis that a crowd of individuals that make "intelligent" choices would result in an intelligent crowd. However, this may be too optimistic. There are many examples of such systems giving rise to alarming phenomena at the aggregate and societal level. This tendency is well represented by the model of ethnic segregation theorized by Schelling [5]. The American economist defined an agent-based model for ethnic segregation that shows that, even when individuals are relatively open-minded and do not mind being surrounded by some people of a different ethnicity or economic background, they will still end up in segregated communities in the long run.

Therefore, to reach the goal of a human-centred AI that supports society in a positive way, there is a need to gain a better understanding of how AI can both support and affect emerging social behaviours. If we can better understand how AI interacts with social phenomena, we can employ it to help mitigate harms and to foster beneficial outcomes, oriented to social goods.

In this tutorial - part of the social simulation chapter with [6] - we approach this challenge from two sides. First, we discuss how complex human systems may experience negative consequences due to their intrinsic nature and under what circumstances Artificial Intelligence may positively or negatively impact such systems. Then, we look at the use of behavioural models as an alternative solution with respect to the data-based approach in scenarios where the human behaviour plays a key role, in order to get a further grip on emerging phenomena in society.

The **learning objectives** of this tutorial are *i)* understand and approach the emergent properties of real networks as well as their possible harmful effects on society; *ii)* leverage agent-based models to understand phenomena where human behaviour play a key role; *iii)* familiarize with the previously illustrated concepts through python libraries and tailored notebooks.

The rest of this tutorial is structured as follows. In Sect. 2, we start by introducing network effects on society and the impact of AI, providing various examples in the urban mobility, public discourse, and market domains. Then, in Sect. 3 we approach the problem of going beyond data-driven approaches – when these are not a suitable solution – in favour of behavioural models to tackle complex challenges like detecting radicalisation or predicting the overall effects of a pandemic. In Sect. 4, we describe a fully reproducible hands-on tutorial to familiarize oneself with the concepts introduced in the tutorial. In Sect. 5 we conclude by discussing limitations of the current approaches in the field of Social AI as well as setting relevant research directions.

2 Network Effects of AI and Their Impact on Society

In the following section, we explore some examples of how Artificial Intelligence may amplify intrinsic and alarming properties of real networks and worsen the wellness of society as a whole. Further, we discuss how AI can be used to better understand these phenomena and find possible solutions.

To tackle this goal, we start by describing real networks and some of the main emerging properties reported by network science literature, followed by three concrete examples of this interconnection of AI and network science.

2.1 Emergent Properties of Real Networks

The earth's climate, the human brain, an organism, a population, an economic organization are all examples of complex systems. Complex systems can be made of different constituents, but they display similar behavioural phenomena at the aggregate level, which are normally called emergent behaviours. According to complex systems theory, these emergent behaviours cannot be explained by analysing the single constituents and need a new approach to understand how they emerge.

Traditionally, social scientists tried to understand how groups of individuals behave by focusing on simple attributes. Understanding emergent behaviours, such as how a population reaches consensus around a specific topic or how a language prevails within a territory, cannot be done by focusing on the individual agents, but instead, the problem needs to be approached in terms of interaction between them. In recent years, this change of approach happened both with the advent of complex systems and network science theory, but also thanks to the availability of a huge amount of data that allow studying individual and higher-order properties of the system. Behind each complex system, there is - in fact - a network that defines the interactions between the system's components. Hence, to understand complex systems, we need to map out and explore the networks behind them. For example, in the field of disease spreading it is nearly impossible to understand how a population reaches the epidemic state without considering the very complex structure of connections between individuals.

In the remainder of this section, we are going to briefly describe some of the emerging properties that characterize many real networks and have a direct impact on real-world phenomena: connectedness, "small-world" property, hubs, and high clustering.

Connectedness. A network is said to be connected if there exists a path between any two pairs of nodes. This may not be surprising in some domains, since connectedness is essential to the correct functioning of the service built on top of the network. For example, if communication networks were not connected we could not call any valid phone number or we would not be able to send an email to any address. However, this property surprisingly emerges also in other domains. For example, online social networks, despite being very large and very sparse, present a giant connected component and any two users are very likely to belong to this component.

According to Erdős-Rényi random network model [7], the necessary and sufficient condition for the emergence of a giant connected component is that the average degree of the network (the number of arcs going out of a node) $\langle k \rangle = 1$. This critical point separates a situation where there is not yet one giant component from the situation where there is one. If the average degree $\langle k \rangle > 1$ the giant component absorbs all nodes and components and the network becomes connected.

Small-World. In real networks it holds not just that everybody is connected to everybody else, but the length of the path to get from one person to a random other person is on average very small. This property is known as "small-world phenomenon".

In the language of network science, the small world phenomenon implies that the distance between two randomly chosen nodes in a network is short, i.e. the average path length or the diameter depends logarithmically on the system size. Hence, "small" means that the average distance is proportional to $\log N$, rather than N or some power of N. In practice this means that in a town of around 100.000 people any person is connected to any other person in 3 or 4 steps. While discovered in the context of social systems, the small-world property applies beyond social networks.

Hubs. Another property that emerges in real networks is the presence of hubs. According to the Erdős-Rényi random network model [7] every node has on average the same number of links. In real-world networks, instead, the degree distribution follows a power-law, i.e. it is scale-free, meaning that there will be very few nodes that are order of magnitudes more connected than the remaining part of the nodes, namely, the hubs. In the known Barabási-Albert model [8] two factors are included to explain the presence of hubs: first, the number of nodes in the network is not fixed, but networks grow; second, there is a so-called "preferential attachment", i.e. the higher the degree of a node, the higher its attractive power to the new nodes entering the network. This second phenomenon can inevitably bring about inequalities in the distribution of resources in the system and this holds for e.g. popularity in social media or for success in the market, but also in protein-to-protein interaction networks. In socio-economical settings, this may cause an excessive amount of inequalities, unsustainable for the social outcomes that, as a society, we would like to pursue.

Clustering. The last emerging property discussed here is network clustering. To measure the degree of clustering in a network we use the local clustering coefficient C, which measures the density of links in a node's immediate neighbourhood. If $C = 0$ it means that there are no links between the node's neighbours, while if $C = 1$ each of the node's neighbours link to each other, i.e. the clustering coefficient is given by the number of triplets in the network that are closed. In real networks there is a higher clustering coefficient than expected according to the Erdős-Rényi random network model [7]. An extension of this model, proposed by Watts and Strogatz [9], addresses the coexistence of a high average clustering coefficient and the small-world property, reconciling the fact that everybody is connected and close to everybody else with the presence of segregation and communities. However, the model fails to predict the scale-free degree distribution seen in real networks mentioned in the previous paragraph.

2.2 AI Pitfalls on Real Networks

We have just seen in Sect. 2.1 that there exists an endogenous tendency of real networks to polarize and segregate that is well represented by their peculiar properties, such as the presence of hubs and the emergence of clustered communities.

In digital environments, such a tendency is further exacerbated by AI-powered tools that, using big data as fuel, make personalized suggestions to every user to make them feel comfortable and, in the end, maximize their engagement [2]. Even if, at the individual level, this kind of suggestion can be beneficial for a user, from a societal point of view it can lead to alarming phenomena in a wide range of domains. Some examples are the polarization and radicalization of public debate in social networks, congestion in mobility networks, or the "rich get richer effect" in economic and financial networks.

In the following, we discuss in detail these three different types of AI pitfalls on real networks, as concerns both their causes and effects.

Polarization, Echo Chamber and Filter Bubble on Social Networks. The rise of online social media and social networking services has drastically changed how people interact, communicate, and access information. In these virtual realms, users have to juggle a continuous, instantaneous, and heterogeneous flow of information from a wide variety of platforms and sources.

From a user perspective, several psychological studies [10,11] have observed that people, both in the online and offline world, feel discomfort when encountering opinions and ideas that contradict their existing beliefs (i.e., Cognitive Dissonance [12]). To avoid such discomfort, as stated by Selective Exposure theories, people tend to select and share contents that reinforce their opinion avoiding conflicting ones [13]. This human tendency to mainly interact with like-minded information and individuals is further strengthened by social media services. Indeed, recommendation and filtering systems play a key role in leveraging users' demographic information and past behaviors to provide personalized news, services, products, and even friends. Despite their success in maximizing user satisfaction, several studies [2,14,15] showed that such systems might lead to a self-reinforcing loop giving rise to the alarming *Filter Bubble* and *Echo Chamber* phenomena.

Even if the discussion about their definitions is still active, traditionally the term Filter Bubble (coined by Parisier [2]) refers to the ecosystem of information to which each online user is exposed, and that is driven by the recommendation algorithms of digital platforms. Similarly, but at an aggregated level, the term Echo Chamber refers to the phenomenon in which beliefs are amplified or reinforced by communication repetition inside a closed system and insulated from rebuttal [3]. In recent years, there is strong concern that such phenomena might prevent the dialectic process of "thesis-antithesis-synthesis" that stands at the basis of a democratic flow of opinions, fostering several related alarming episodes (e.g., hate speech, misinformation, and ethnic stigmatization). In this context, there is both a need for data-driven approaches to identify and analyse real situations where these phenomena take place, but also for tools to investigate causes and effects of different factors on polarizing phenomena on online and offline social networks, as well as mitigation strategies.

Congestion in Mobility Networks. Traffic congestion can cause undesired effects on a transportation network such as waste of time, economic losses to drivers, waste of energy and increasing air pollution, increased reaction time to emergencies, and dissatisfaction of the well-being of people in the urban environment [1]. To avoid such negative effects, congestion prevention is crucial to improve the overall transportation system's sustainability.

Congestion happens due to demand-supply imbalance in the road network, which can be exacerbated by AI navigation systems' advice. Indeed, while these recommendations make sense at the individual level, they can lead to collective dissatisfaction, when the same advice is given to many different drivers (e.g., the navigators suggest to all vehicles to travel across the same road to reach a certain destination) because the road links will saturate and congestions emerge.

A naive solution for traffic congestion prevention, i.e. adding an extra road to redistribute the traffic, can instead lead to the opposite effect, as stated in Braess's paradox [16]. The paradox occurs because each driver chooses whatever route minimizes their personal travel time (selfish choice). When people share a common public resource - like the road network - the lack of cooperation along with selfish behavior might lead to a stable state with a social value that is far from the social optimum, as claimed by John Nash.

In the problem of traffic congestion avoidance, there is a need for coordination, cooperation, and diversification of routes. In the very same way that we need diversification of opinions to have democracy work in our societies, we need diversification of behavior for having a better ability to travel in our cities. Therefore, we need AI systems that can forecast where a traffic congestion will occur to avoid its formation.

Disparity in Economic Networks. In just about a decade, a handful of companies have contributed to an increasingly centralized World Wide Web, contradicting the Internet's original slogan of net neutrality [17]. This is the first time in history that technology companies (e.g., Apple, Google, Microsoft, Amazon, and Facebook) have dominated the stock market, being the most valuable public businesses in the world by market capitalization [18].

However, dominance on the Internet is not limited to the digital realm, but transcends the economy as a whole, such as digital advertising and e-commerce.

Economic and financial networks are deeply characterized by the so-called *winner-takes-all markets* (WAT). This terminology refers to an economy in which the best performers can capture a considerable share of the available rewards, while the remaining competitors are left with very little [4]. Such a behavior, also known as the "rich get richer effect", is well reflected in the topology of real networks with the presence of hubs due to the law of preferential attachment that states that the growth rate is proportional to the size of the nodes [8].

As we can imagine, the prevalence of "winner-takes-all" phenomena in markets increases wealth inequalities because a selected few can capture increasing amounts of income that would otherwise be more evenly distributed throughout the population of companies [19]. Accordingly, such kind of economic polarization strongly limits the possibility of small companies emerging.

2.3 Addressing AI Pitfalls

In literature, the AI drawbacks described in the previous section have been tackled from three main perspectives: *i)* designing models to capture their dynamics and behaviors; *ii)* analyzing their emergence in real-world scenario via empirical data; *iii)* mitigating their effect through ad-hoc prevention strategies. In the following, we explore an exemplifying case study related to Polarization, Congestion, and WAT phenomena for each of these approaches.

Modeling. *How is it possible to model how opinions evolve within a population as a result of peer-to-peer interactions among people?* This kind of question can be investigated through the tools of **opinion dynamics**. Opinion dynamics models aim at understanding how opinions evolve within a population, simulating interactions between agents of the population, in a process governed by mathematical equations incorporating sociological rules of opinion formation. The perks of opinion dynamics models - and agent-based models in general - is that they allow for "what-if" scenarios analyses and to track the cause-consequence link to understand the drivers of a certain state.

In the following, we describe an opinion dynamics model that incorporates cognitive biases and explores the effects that a recommender system - creating an algorithmic bias - may have on the resulting dynamics.

Algorithmic Bias Model. The so-called *bounded confidence models* constitute a broad family of models where agents are influenced only by neighbours in the social network having an opinion sufficiently close to theirs. Specifically, the DW model [20] considers a population of N agents, where each agent i has a continuous opinion $x_i \in [0, 1]$. At every discrete time step the model randomly selects a pair (i, j), and, if their opinion distance is lower than a threshold ϵ_{DW}, $|x_i - x_j| \leq \epsilon_{DW}$, then the two agents change their opinion taking the average. The AB model [21], which extends the DW one, introduces a bias towards similar individuals in the interaction partner's choice adding another parameter to model the algorithmic bias: $\gamma \geq 0$. This parameter represents the filtering power of a generic recommendation algorithm: if it is close to 0, the agent has the same probability of interacting with all of its peers. As γ grows, so does the probability of interacting with agents holding similar opinions, while the probability of interacting with those who hold distant opinions decreases. The introduction of stronger bias causes more fragmentation, more polarization, and more instability. Fragmentation is interpreted as an increased number of clusters, while polarization is interpreted as an increasing pairwise distance among opinions and instability means a slowdown of time to convergence with a large number of opinion clusters that coexist for a certain period.

Characterizing. *How is it possible to detect and prevent congestion in mobility networks?* The detection and prediction of traffic congestions across road networks are crucial for several reasons, such as the reduction of air pollution, reduction of the travel time for the drivers, and the increase of security along roads. According to [22], the congestion detection problem requires data-driven approaches. In fact, several works use empirical data to perform their study on traffic congestion.

The pervasive presence of vehicles equipped with GPS localization systems provides a precise way to sense their movements on the road; vehicles equipped with GPS can act as mobile sensors that sense information regarding traffic conditions as well as providing a characterization of drivers' behavior that can be an indicator of congestion happening.

With proper analysis, GPS trajectories can be used for detecting and/or predicting traffic jam conditions [23] as it is possible to recognize some patterns that indicate if a driver is stuck in a traffic jam, e.g. if their speed is significantly lower than the speed allowed in that particular road and their trajectory is characterized by sudden speed changes indicating close (both in time and space) starts and braking. Vaqar et al. [23], propose a methodology that detects traffic congestion using pattern recognition.

Recently, AI models were used in the traffic congestion detection/prediction task. As an example, in [22] the authors use both Deep Learning as well as conventional Machine Learning models, trained with real vehicular GPS data, to predict the traffic congestion level. According to their results, Deep Learning models obtain higher accuracy in traffic congestion prediction compared to conventional Machine Learning models.

Mitigating. *How is it possible to mitigate the winner-takes-all effect on economic networks?* Traditionally, progressive taxes, antitrust laws, and similar legislation are typical countermeasures against centralization. However, Lera and colleagues [17] recently found that the mere limitation of the power of most dominant agents may be ineffective because it addresses only the symptoms rather than the underlying cause, i.e. an imbalanced system that catalyzes such dominance.

Following such reasoning, they designed an early warning system and then an optimal intervention policy to prevent and mitigate the rise of the WAT effect in growing complex networks. First, they mathematically defined a system of interacting agents, where the rate at which an agent establishes connections to others is proportional to its already existing number of connections and its intrinsic fitness (i.e., growth rate). Then, they found that by calibrating the system's parameters with maximum likelihood, they can monitor in real-time its distance from a WAT state. Therefore, if the system is close to a WAT state, they increase the fitness of the other economic actors of the networks.

In such a way, they have shown how to efficiently drive the system back into a more stable state in terms of the fitness distribution of all the actors involved.

3 Beyond Data-Driven Approaches: Behavioural Models to Understand Societal Phenomena

There are phenomena depending on physical events for which not all data is readily available that lead to real-world effects. One example comes from our recent experience with the Covid-19 pandemics. For example, we may imagine that after restrictions are lifted people will spend more in shops, but how can we know how much or when will this happen? How can we answer very specific questions like: will a souvenir shop survive? Surely, we need data to answer these questions, but data themselves are not enough if they are not contextualised into an underlying model. Hence, in this uncertain situation, we need models that include human behaviour to support decisions, since data-based predictions are insufficient if we want to predict human behaviour based on sociality, economics and health.

In the remainder of this section, we describe some case studies in-depth and discuss approaches to analyse them with appropriate tools to try to answer the question of how data and models can be connected.

3.1 Tracking Online Extremism that Leads to Offline Extremist Behaviours

The concept of radicalisation is by no means solid and clear, and also, when it comes to radical behaviours like terrorism, there is no universally accepted definition. Such a lack of definition inevitably makes it harder to understand the process that brings people on the path towards radicalisation and how and if people can be de-radicalised.

Theory of Radicalisation. According to Kruglanski et al. [24] the radicalisation process involves an individual moving toward believing and engaging in activities that violate important social norms, mainly because radicalised individuals are focused on only one personal goal, undermining everything else that may be important to other people, seeing radicalism as motivational imbalance. The model developed by Kruglanski et al. [24] identifies three crucial components - both personal and social - that lead to the extreme commitment to a certain goal that we can find in radicalised individuals:

1. The need for significance: the motivational component that identifies the goal to which the individual is highly committed
2. Ideology: the cultural component that defines the role of group ideologies in identifying violent means as appropriate in goal pursuit
3. The social component identifies the group dynamics through which the individual comes to endorse the group ideology. Commitment to ideology is fostered through social connections and the considerable group pressure placed on the individual when those surrounding him espouse his ideological views.

Having this (still very crude) model of extremism can support us in finding out extremist behavior on social media. We can start looking for expressions on social media platforms where people make extreme remarks just to get attention. One can also check whether people reinforce each other ideas and all these ideas can be linked to the same ideology.

Extremism on Online Social Networks. A social media platform is a powerful tool for the formation and maintenance of social groups. Crucially, social media platforms let users actively participate in groups through several mechanisms explicitly designed for the purpose. In the case of radicalisation outlined above, active participation in groups plays a crucial role: people first identify themselves as belonging to a group, and then through various types of active participation, identify increasingly closely with that group, discriminate increasingly strongly against out-groups, and eventually participate in practices that isolate the group, and instil fear of external groups. Of course, the vast majority of the time, these mechanisms help users create socially beneficial groups, and help people find, join, and participate in these groups. Nonetheless, social media systems also have minor effects in encouraging users to move along the pathway towards radicalisation alongside these socially beneficial functions.

The goal is to capture the early signals of radicalisation (on social networks) and understand the path towards such behaviour to prevent extremist behaviour in real life. We need to bridge the gap between data and models to tackle this goal.

Identify Extremism on Online Social Networks. Identifying extremist-associated conversations on Online Social Networks (OSN) is an open problem. Extremist groups have been leveraging OSN (1) to spread their message and (2) gain recruits. To determine whether a particular user engages in extremist conversation, one can explore different metrics as proxies for misbehaviour, including the sentiment of the user's published posts, the polarity of the user's ego network, and user mentions. In [25] they find that - on a Twitter dataset - combining all these features and then using different known classifiers leads to the highest accuracy. However, the recent events of the Capitol Hill riot showed how one cannot assume anything on extremist behaviours by only looking at one social network data. In fact, after the start of the Capitol Hill riot, related posts started to trend on social media. A study by Hitkul et al. [26] analysing trending traffic from Twitter and Parler showed that a considerable proportion of Parler activity was in support of undermining the validity of the 2020 US Presidential elections, while the Twitter conversation disapproved of Trump. From this simple example, one can understand that while in one social media we may not see pathways towards radicalisation, these may emerge by changing social media, so the data we need to analyse the collective behaviour is scattered between several platforms, and we need to look at the right one if we want to identify the characteristics of the phenomena. So if we want to understand radicalisation, we need to ask ourselves what is the right data for the task, whether this can be retrieved and, eventually, what is the connection between the data and reality.

A Model of Radicalisation. In order to answer the question of how people radicalise and if we can find pathways towards radicalisation, we need a model of human behaviour. The purpose of the model by Hurk and Dignum [27] - based on the theoretical framework by Kruglanski et al. [24] - is to show that the combination of a high need for significance, a radical ideology and a social group acting according to that ideology can start the process of radicalisation. Agents - connected in a social network - live in a world where the goal is to keep their significance level as high as possible. They can gain significance by performing actions and getting acknowledged by their social surrounding. How actions can increase significance is defined in two different ideologies. Every agent belongs to a group that acts according to an ideology. In extreme cases, the agent can switch to the other group with the other ideology. In this context, radical behaviour means agents that perform actions that give them a significant gain in significance, but others reject those actions. Furthermore, the population of agents will split into groups, where agents will be surrounded mainly by other agents belonging to the same group. The results show that groups of radical agents emerge, where radicalising individuals form isolated social bubbles. It shows the importance of social surroundings in order to start radicalising. Furthermore, the agent's circumstances seem to be important because not all agents with a low level of significance can gain it back. These results support understanding the radicalisation process, but they can also give insights into why de-radicalisation is not straightforward as long as one cannot escape his social group.

3.2 Multi-agent Behavioural Models for Epidemic Spreading: From Model to Data in the Covid-19 Crisis

During the Covid-19 crisis, policymakers had to make many difficult decisions with the help of models for epidemic spreading. However, classical epidemiological models do not directly translate into the interventions that can be taken to limit the spread of the virus, neither these models include economic/social consequences of these interventions. Policies may impact epidemics, economics and societies differently, and a policy that can be beneficial from one perspective can have negative consequences from another. In order to make good decisions, policymakers need to be aware of the combined consequences of the policies. There is a need for tools to support this decision-making process that enable the design and analysis of many what-if scenarios and potential outcomes. Tools should thus facilitate the investigation of alternatives and highlight the fundamental choices to be made rather than giving one solution.

Agent-Based Social Simulation for Covid-19 Crisis. The consequences of a pandemic can be addressed from different points of view, that all have some limitations when considered separately: (1) classical epidemiological models do not consider human behaviour or consequences of interventions on actions of people or - if they incorporate such things into the model parameters - they lose the cause-effect links and causes cannot be easily identified and adjusted; (2) also economic models fail to capture human behaviour always assuming perfect rationality; (3) social network theory does not say anything on how the social network will change and how people will react to a new policy. The proposed solution by Dignum et al. is to make human behaviour central and use it as a link to connect epidemics, economics, and society. ASSOCC, a model by Dignum et al. proposed in [28] is an agent-based social simulation tool that supports decision-makers gain insights on the possible effects of policies by showing their interdependencies, and as such, making clear which are the underlying dilemmas that have to be addressed. Such understanding can lead to more acceptable solutions, adapted to the situation of each country and its current socio-economic state, and that is sustainable from a long-term perspective. In this model - implemented in Netlogo - the main components are agents, places, global functions, and policies. Agents take decisions based on the container model of needs: needs are satisfied by activities and decay over time. The basic needs included in the Covid-19 simulation model are safety, belonging, self-esteem, autonomy, and survival. These needs combine health, wealth, and social well-being in the same simulation model. Agents are organised along a grid and the environment can pose constraints to physical actions and impose norms and regulations, while when interacting, agents can take other agents' characteristics (e.g., being infected).

One of the advantages of using such model is that instead of providing a single prediction, it gives support to investigate the consequences of different scenarios. Due to its agent based nature, it is also possible to explain where results come from; it allows for more fine grained analysis on the impact of interventions, and it can be used in combination with domain specific models. Having a good insight into these dependencies can provide the domain-specific models with better information to make specific optimisations or predictions for that intervention's effect. Thus, the strength of the different types of models can be combined rather than seen as competing.

4 Hands-on Tutorial

In this lecture, we also provided a three-part tutorial[1], explaining how to use different tools to simulate the effects on networks treated within this manuscript.

For this practical part of the tutorial, we employed two `Python` libraries: `scikit-mobility` [29] for the study and analysis of human mobility and `NDlib` [30] for the simulation and study of spreading phenomena on networks.

The first part of the tutorial introduces the fundamental concepts of human mobility and explains how to create a mobility network that describes the movements of individuals. In the second part of the tutorial, diffusion phenomena are simulated on a real network with different state-of-the-art algorithms. In the last section, state-of-the-art opinion dynamics algorithms are simulated over a real social network.

5 Conclusions

In this tutorial, we have discussed the social dimension of Artificial Intelligence in terms of how AI technologies can support or affect emerging social challenges.

[1] https://github.com/GiulianoCornacchia/ACAI_SAI_Tutorial.

On the one hand, the goal of this tutorial was to consider the network effects of AI and their impact on society. On the other hand, we wanted to introduce strategies, both data-driven and not, to model, characterise and mitigate AI societal pitfalls. Here we conclude by pointing out some limitations of the existing approaches and the research directions that should be addressed in the future.

First, both empirical and modelling works consider AI drawbacks too simplistically, often relying on unrealistic assumptions that do not reflect real-world phenomena' complexity. For such a reason, we claim the urgency of starting cooperation with digital platforms to understand better whether AI-powered tools exacerbate endogenous features of human beings. In this direction, ongoing projects such as the Global Partnership of Artificial Intelligence (GPAI[2]) are trying to collaborate with social media platforms in order to study, from the inside, the effects of recommendation systems on users.

Secondly, we want to stress the importance of designing AI systems that pursue goals both at the individual level and considering the whole population involved. Indeed, most harmful phenomena analysed in this tutorial emerge as group phenomena. For instance, congestion in mobility networks happens because every individual is given the same suggestion, or echo chambers emerge because recommendation systems cluster together like-minded individuals.

In conclusion, towards social AI ecosystems, we encourage readers to design AI tools that foster collective interests rather than individual needs.

References

1. Afrin, T., Yodo, N.: A survey of road traffic congestion measures towards a sustainable and resilient transportation system. Sustainability **12**(11), 4660 (2020)
2. Pariser, E.: The Filter Bubble: What the Internet is Hiding From You. Penguin UK, Westminster (2011)
3. Sunstein, C.R.: Republic. com. Princeton University Press, Princeton (2001)
4. Rycroft, R.S.: The Economics of Inequality, Discrimination, Poverty, and Mobility. Routledge, Milton Park (2017)
5. Schelling, T.C.: Models of segregation. Am. Econ. Rev. **59**(2), 488–493 (1969)
6. Lorig, F., Vanhée, L., Dignum, F.: Agent-based social simulation for policy making (2022)
7. Erdős P., Rényi, A.: On random graphs. i. Publicationes Math. **6**, 290–297 (1959)
8. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
9. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
10. Jean Tsang, S.: Cognitive discrepancy, dissonance, and selective exposure. Media Psychol. **22**(3), 394–417 (2019)
11. Jeong, M., Zo, H., Lee, C.H., Ceran, Y.: Feeling displeasure from online social media postings: a study using cognitive dissonance theory. Comput. Hum. Behav. **97**, 231–240 (2019)
12. Festinger, L.: A Theory of Cognitive Dissonance, vol. 2. Stanford University Press, Redwood City (1957)

[2] https://gpai.ai/.

13. Borah, P., Thorson, K., Hwang, H.: Causes and consequences of selective exposure among political blog readers: the role of hostile media perception in motivated media use and expressive participation. J. Inf. Technol. Polit. **12**(2), 186–199 (2015)
14. Bozdag, E.: Bias in algorithmic filtering and personalization. Ethics Inf. Technol. **15**(3), 209–227 (2013)
15. Ge, Y., et al.: Understanding echo chambers in e-commerce recommender systems. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2261–2270 (2020)
16. Braess, D.: Über ein paradoxon aus der verkehrsplanung. Unternehmensforschung **12**, 258–268 (1968)
17. Lera, S.C., Pentland, A., Sornette, D.: Prediction and prevention of disproportionally dominant agents in complex networks. Proc. Natl. Acad. Sci. **117**(44), 27090–27095 (2020)
18. Moore, M., Tambini, D.: Digital dominance: the power of Google. Facebook, and Apple. Oxford University Press, Amazon (2018)
19. Cook, P.J., Frank, R.H.: The winner-Take-all Society: Why the Few at the Top Get So Much More Than the Rest of Us. Random House, New York (2010)
20. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G.: Mixing beliefs among interacting agents. Adv. Complex Syst. **3**, 87–98 (2000)
21. Sîrbu, A., Pedreschi, D., Giannotti, F., Kertész, J.: Algorithmic bias amplifies opinion fragmentation and polarization: a bounded confidence model. PLoS ONE **14**(3), e0213246 (2019)
22. Sun, S., Chen, J., Sun, J.: Congestion prediction based on GPS trajectory data. Int. J. Distrib. Sens. Netw. **15**, 155014771984744 (2019)
23. Vaqar, S.A., Basir, O.: Traffic pattern detection in a partially deployed vehicular ad hoc network of vehicles. IEEE Wireless Commun. **16**(6), 40–46 (2009)
24. Kruglanski, A.W., Gelfand, M.J., Bélanger, J.J., Sheveland, A., Hetiarachchi, M., Gunaratna, R.K.: The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. Polit. Psychol. **35**, 69–93 (2014)
25. Wei, Y., Singh, L., Martin, S.: Identification of extremism on Twitter. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1251–1255. IEEE (2016)
26. Prabhu, A., et al.: Capitol (pat) riots: a comparative study of Twitter and parler. arXiv preprint arXiv:2101.06914 (2021)
27. van den Hurk, M., Dignum, F.: Towards fundamental models of radicalization. In: ESSA (2019)
28. Dignum, F., et al.: Analysing the combined health, social and economic impacts of the corovanvirus pandemic using agent-based social simulation. Minds Mach. **30**(2), 177–194 (2020). https://doi.org/10.1007/s11023-020-09527-6
29. Pappalardo, L., Simini, F., Barlacchi, G., Pellungrini, R.: Scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data. arXiv preprint arXiv:1907.07062 (2019)
30. Rossetti, G., Milli, L., Rinzivillo, S., Sîrbu, A., Pedreschi, D., Giannotti, F.: Ndlib: a python library to model and analyze diffusion processes over complex networks. Int. J. Data Sci. Anal. **5**(1), 61–79 (2018)