\documentclass[10pt,letterpaper]{article}

\usepackage[top=0.85in,left=2.75in,footskip=0.75in]{geometry}

% amsmath and amssymb packages, useful for mathematical formulas and symbols

\usepackage{amsmath,amssymb}

% Use adjustwidth environment to exceed column width (see example table in text)

\usepackage{changepage}

% Use Unicode characters when possible

\usepackage[utf8x]{inputenc}

% textcomp package and marvosym package for additional characters

\usepackage{textcomp,marvosym}

% cite package, to clean up citations in the main text. Do not remove.

\usepackage{cite}

% Use nameref to cite supporting information files (see Supporting Information section for more info)

\usepackage{nameref,hyperref}

% line numbers

\usepackage[right]{lineno}

% ligatures disabled

\usepackage{microtype}

% color can be used to apply background shading to table cells only

\usepackage[table]{xcolor}

% array package and thick rules for tables

\usepackage{array}

\usepackage{subfig}

% \usepackage[demo]{graphicx}

% Remove comment for double spacing

\usepackage{setspace}

%\doublespacing

\usepackage{csquotes}

\usepackage{dirtytalk}

\usepackage[english]{babel}

\usepackage{hyperref}

\hypersetup{

colorlinks=true,

citecolor=blue,

linkcolor=blue,

urlcolor=blue,

pdftitle={Algorithmic Bias Media Model},

pdfpagemode=FullScreen}

\urlstyle{same}

% Bold the 'Figure #' in the caption and separate it from the title/caption with a period

% Captions will be left justified

\usepackage[aboveskip=1pt,labelfont=bf,labelsep=period,justification=raggedright,singlelinecheck=off]{caption}

% Use the PLoS provided BiBTeX style

\bibliographystyle{plos2015}

% Header and Footer with logo

\usepackage{lastpage,fancyhdr,graphicx}

\usepackage{subfig}

\usepackage{epstopdf}

%\pagestyle{myheadings}

\pagestyle{fancy}

\fancyhf{}

%\setlength{\headheight}{27.023pt}

%\lhead{\includegraphics[width=2.0in]{PLOS-submission.eps}}

\rfoot{\thepage/\pageref{LastPage}}

\renewcommand{\headrulewidth}{0pt}

\renewcommand{\footrule}{\hrule height 2pt \vspace{2mm}}

\fancyheadoffset[L]{2.25in}

\fancyfootoffset[L]{2.25in}

\lfoot{\today}

%% Include all macros below

\newcommand{\lorem}{{\bf LOREM}}

\newcommand{\ipsum}{{\bf IPSUM}}

%% END MACROS SECTION

\begin{document}

%

\title{Mean-field analysis of the effects of mass media presence on Deffuant-Weisbuch model with Algorithmic Bias.}

%

%\titlerunning{Abbreviated paper title}

% If the paper title is too long for the running head, you can set

% an abbreviated paper title here

%

% Author Orchid ID: enter ID or remove command

% \author{Frank Dignum $^{1}$\orcidA{}, Dino Pedreschi $^{2}$\orcidB{}, Giuliano Cornacchia $^{2,4}$\orcidC{}, \\

% Virginia Morini$^{2,4}$\orcidD{}, Valentina Pansanella$^{3,4}$\orcidE{}}

%\author{Frank Dignum\inst{1}\orcidA{} \and

%Dino Pedreschi\inst{2,3}\orcidID{0000-0003-4801-3225} \and

%Giuliano Cornacchia\inst{3}\orcidID{0000-0003-2263-7654} \and

%Virginia Morini\inst{4}\orcidID{0000-0002-7692-8134} \and

%Valentina Pansanella\inst{5}\orcidID{0000-0001-8106-7677}}

%

% \authorrunning{F. Dignum et al.}

% First names are abbreviated in the running head.

% If there are more than two authors, 'et al.' is used.

% %

% \institute{Umeå University, Sweden \\ \email{dignum@cs.umu.se} \\ \and University of Pisa, Italy \\ \email{dino.pedreschi@unipi.it} \\

% \email{\{giuliano.cornacchia,virginia.morini\}@phd.unipi.it} \\ \and

% Scuola Normale Superiore, Pisa, Italy \\

% \email{valentina.pansanella@sns.it}\\ \and

% KDD Lab ISTI-CNR, Pisa, Italy}

%\url{http://www.springer.com/gp/computer-science/lncs} }

%ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany\\

%\email{\{abc,lncs\}@uni-heidelberg.de}}

%

\maketitle % typeset the header of the contribution

%

\begin{abstract}

\end{abstract}

%

%

%

\section{Introduction}

Opinions and beliefs shape both the individual behaviour, in turn driving human actions, and the collective behaviour of a society, hence impacting on choices concerning politics, public health, the environment, etc. Changes in public opinion\footnote{An aggregate of the individual views, attitudes, and beliefs about a particular topic, expressed by a significant proportion of a community. The influence of public opinion is not restricted to politics and elections. It is a powerful force in many other spheres, such as culture, fashion, literature and the arts, consumer spending, and marketing and public relations.} - even the creation of committed minorities\footnote{\say{A committed few can influence the many and sweep away social conventions, new research shows}. \url{https://bit.ly/3tMYV25}} - may profoundly affect decision-makers and politics: a great example is the temporary suspension of Oxford-AstraZeneca vaccine during March 2021\footnote{Reuters \say{Decision by Germany, France and Italy to suspend AstraZeneca's COVID-19 shots after several countries reported possible serious side-effects is a \say{political one}, the general director of Italy's medicines authority AIFA said on Tuesday}. \url{https://bit.ly/3ki4Wk7}}, which costed a slowdown in the vaccination strategy.

Opinion exchange is boosted by online social networks and media, which are increasingly used as news and information sources. Media clearly play a role in the process of opinion formation, deploying a critical ingredient that is: information.

This essential ingredient in the process of opinion formation is often missing in the first generation of opinion dynamics models \cite{McKeown2006MassMA, Carletti2006HowTM, Timothy2017HowDP, Pineda2015MassMA, Huang2016ModelingAS, Brooks2019AMF, Glass2021OpinionDO, Quattrociocchi2011OpinionsMM}: while it is true that social influence plays a significant role in shaping our culture, ideology and opinion spectrum, it is also true that reading news or being the target of mass political propaganda \cite{Bornhuser2006GermanTI, Weatherall2020HowTB} may also affect our belief system and should not be disregarded when addressing the study of the dynamics of opinions.

Also, external agents (i.e., a government, a company, a group of terrorists) may be interested in shifting the public’s opinion concerning a specific topic. Propaganda or other actions can be exploited to try and promote one opinion over the others \cite{Franke2015StrategiesOP}, or to achieve a particular value for the consensus opinion \cite{Lorenz2007AboutTP}, or maybe to prevent people from reaching more extreme opinions \cite{Timothy2017HowDP} to avoid the consequent risks.

However, how this information is presented is argued to be highly biased towards pre-existing beliefs by recommender systems. There is theoretical evidence that algorithmic bias fosters opinion fragmentation/polarisation. Even if their initial intent was to maximise platform usage and users' engagement, these algorithms end up working as a reinforcement bias for online users' opinions, neglecting them access/confront with narratives different from their own. A deeper understanding of these dynamics is essential to pave the way for developing mitigation strategies.

The field of opinion dynamics aims to understand opinion formation, evolution, and eventually stabilisation in groups of interacting agents. Typically, such models consider a finite number of interacting people (% usually called

\emph{agents}); each agent has its own opinion, discrete or continuous, which can vary over time,

% Opinion Dynamics models establish

according to rules to explain the change due to interactions with other agents.

The present work aims to understand the role of mass media in an environment with a recommender system, and algorithmic bias with an opinion dynamics approach, assuming a mean-field context (e.g., all individuals can interact among them without any social restrictions).

Adopting such controlled environments, used to simulate the social structure among a population of interacting individuals, we analyse the behaviours of the Algorithmic Bias model\cite{Srbu2019AlgorithmicBA} (e.g., an extension of the Deffuant-Weisbuch\cite{DW00} opinion dynamics model) with the presence of mass media modelled as stubborn agents.

\smallskip

The paper is organised as follows.

In section \ref{sec:modeldefinition} we introduce the algorithmic bias model, and we describe the change in the interaction rules due to the presence of mass media.

Section \ref{sec:results} discusses the main finding of our simulations, finally section \ref{sec:conlcusion} concludes the paper while opening to future investigations.

\section{Model definition}

\label{sec:modeldefinition}

Considering such a complex scenario, the present work aims to deepen the AB model's behavior analysis and understand how a biased interactions with mass media may influence the outcomes of the dynamical process.

Nowadays, online social networks have become the primary source of information and an excellent platform for

discussions and

opinion exchanges.

However, the flow of content that each user sees is organized by algorithms that are built to maximize platform usage: from this, it comes to the hypothesis that there is an algorithmic bias (also called algorithmic segregation) since these contents are selected based on users' precedent actions on the platform or the web,

reinforcing the human tendency to interact with content confirming their beliefs (confirmation bias).

It is a human tendency to prefer interaction with information/content/discussions that confirm their narrative (a phenomenon known as confirmation bias), and so users tend to interact more with this kind of content rather than go and face evidence that puts in discussion their beliefs.

Therefore, it is in the service provider's interest to deliver its contents in a targeted way.

To introduce in the study of opinion dynamics the idea of a recommender system generating an algorithmic bias, we started from a recent extension of the well-known Deffuant-Weisbuch model (DW-model henceforth), proposed in \cite{Srbu2019AlgorithmicBA} (Algorithmic Bias model or AB model, henceforth).

In the Algorithmic Bias model, we have a population of $N$ agents, where each agents $i$ has a continuous opinions $x\_{i} \in [0,1]$.

At every discrete time step the model randomly select a pairwise $(i, j)$, and, if their opinion distance is lower than a threshold $\epsilon$, $|x\_{i} - x\_{j}| \leq \epsilon$, then the two agents change their opinion according to the following rule:

\begin{equation}

\begin{aligned}

\label{eq:updateDW}

x\_{i}(t+1) &= x\_{i} + \mu(x\_{j}-x\_{i}) \\

x\_{j}(t+1) &= x\_{j} + \mu(x\_{i}-x\_{j}).

\end{aligned}

\end{equation}

The parameter $\epsilon \in [0,1]$ models the population's confidence bound, and it is assumed to be constant among all the agents.

A low $\epsilon$ creates a close-minded population, where the individuals can only be influenced by those whose opinions are very similar to theirs; a high $\epsilon$, instead, creates an open-minded population since two agents can influence each other even if their initial opinions are very distant.

The parameter $\mu \in (0, 0.5]$ is a convergence parameter, modeling the strength of the influence the two individuals have on each other or, in other words, how much they change their opinion after the interaction.

Even if there is no reason to assume that $\epsilon$ should be constant across the population or at least symmetrical in the binary encounters, this parameter is always considered equal for every agent.

The numerical simulations of this model show that the qualitative dynamic is mainly dependent on $\epsilon$: as $\epsilon$ grows, the number of final opinion clusters decreases.

As for $\mu$ and $N$, these parameters tend to influence only the time to convergence and the final opinion distribution width.

The AB model introduces 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 its peers. As $\gamma$ grows, so does the probability of interacting with agents holding similar opinions, while the probability of interacting with those who hold distant opinions decreases.

Therefore, this extended model modifies the rule to choose the interacting pair $(i, j)$ to simulate a filtering algorithm's presence. An agent $i$ is randomly picked from the population, while $j$ is chosen from $i$'s peers according to the following rule:

\begin{eqnarray}

\label{eq:algbias}

p\_{i}(j)=\frac{d\_{ij}^{-\gamma}}{\sum\_{k \ne i}d\_{ik}^{-\gamma}}

\end{eqnarray}

where $d\_{ij} = |x\_{i}-x\_{j}|$ is the opinion distance between agents $i$ and $j$, so that for $\gamma = 0$ the model goes back to the DW-model.

When two agents interact, their opinions change if and only if the distance between their opinions is less than the parameter $\epsilon$, i.e. \(|x\_{i}-x\_{j}| \leq \epsilon \), according to Eq.~\ref{eq:updateDW}.

\begin{figure}

\centering

\includegraphics[width=\textwidth]{figures/example.PNG}

\caption{\textbf{Example interaction.}}

\label{fig:example}

\end{figure}

Our goal is to understand the effects of the presence of biased media outlets that promote a certain, fixed, opinion during the whole time period. Therefore, media were modelled as stubborn agents (or zealots) never changing their opinion and connected to everyone, to simulate a news source accessible to everyone in the population. In figure \ref{fig:example} an example of a possible interaction both peer to peer and agent to media and its effects on the node’s opinion is illustrated.

\section{Analysis and results}

\label{sec:results}

We performed simulations of this model on a complete network of $100$ nodes where the initial opinions are initially uniformly distributed across the population and averaged the results over $100$ runs. Like in \cite{Srbu2019AlgorithmicBA}, to avoid undefined operations in equation \ref{eq:algbias}, when $d\_{ik} = 0$ we use a lower bound $d\_{\epsilon} = 10^{-4}$.

The simulations are designed to stop when the population reaches an equilibrium, i.e., the cluster configuration will not change anymore, even if the agents keep exchanging opinions.

We also set an overall maximum number of iterations at $10^5$.

To better understand the differences in the final state concerning the different topologies considered, we study the model on all networks for different combinations of the parameters.

We are interested in whether, parameters being equal, the different number and positioning of mass media in the opinion spectrum and the growing probability to interact with such fixed agents influences the final clustered configuration enhancing or reducing fragmentation and radicalising individuals towards more extreme opinions.

We replicated the work of \cite{Srbu2019AlgorithmicBA} by setting a null probability to interact with the media, to define a reliable baseline for comparison.

In the simulations, we tested the model on every possible combination of the parameters over the following values:

\begin{itemize}

\item $p\_{m}$ takes values in $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, where for $p\_{m}=0$ the model becomes the AB model.

\item $\epsilon$ takes value in $\{0.1, 0.2, 0.3, 0.4, 0.5, 1.0\}$.

\item $\gamma$ takes value in $\{0.0, 0.5, 0.75, 1.0, 1.25, 1.5\}$ where for $\gamma = 0$ the model becomes the DW-model.

\item $\mu = 0.5$, so whenever two agents interact, if their opinions are close enough, they update to the average opinion of the pair.

\end{itemize}

We analysed two scenarios to understand the effects of (i) one extreme media with an opinion $m\_{1}=0.0$ and (ii) three media with opinions $m\_{1}=0.05, m\_{2}=0.5, m\_{3}=0.95$.

The AB model with mass media implementation used to carry out our experiments is the one provided by the NDlib \cite{rossetti2018ndlib} Python library\footnote{NDlib: \url{http://ndlib.rtfd.io}}.

\subsection{Experimental results}

\subsubsection{Media and bounded confidence}

To understand the results of the presence of mass media in a non-biased environment, we analysed the number of opinion cluster and the value of entropy as a function of both $p\_{m}$ and $\epsilon$, setting $\gamma=0.0$.

The behaviour of the original Deffuant-Weisbuch model is defined by the parameter $\epsilon$, when this is large enough an initially uniformly random population converge to consensus, while as $\epsilon$ decreases cluster emerge in the population and their number can be approximated by $\lfloor{\frac{1}{2\epsilon}\rfloor}$, ignoring minor clusters that appear in some situations.

%devo dire come ho fatto il clustering?

To compute the effective number of clusters, accounting for the presence of major and minor ones, we use the cluster participation ratio, as in \cite{Srbu2019AlgorithmicBA}:

\begin{eqnarray}

\label{eq:ncluster}

C = \frac{(\sum\_{i}{c\_{i}})^{2}}{\sum\_{i}{c\_{i}^{2}}}

\end{eqnarray}

where $c\_{i}$ is the dimension of the $i$th cluster, i.e. the fraction of population we can find in that cluster. In general, for n clusters, the maximum value of the participation ratio is n and is achieved when all clusters have the same size, while the minimum can be close to 1, if one cluster contains most of the population and a very small fraction is distributed among the other $n − 1$.

To account for situations where proper clusters may not emerge in the population - hence the number of clusters may fail to capture the features of the distribution - we computed the entropy of the final opinion distribution, using Shannon entropy on the discrete probability distribution, dividing the opinion spectrum into $k$ intervals of equal width and computing the probability of a node to be in the interval $i$, $p(i)$. The Shannon's entropy is therefore:

\begin{eqnarray}

\label{eq:entropy}

S = {-\sum{p(i)\ln p(i)}}

\end{eqnarray}

\begin{figure}

\centering

\includegraphics[width=0.8\textwidth]{figures/hm media nobias mo0.05;0.5;0.95 new\_avg\_ncluster groupedby\_media\_op.png}

\caption{\textbf{Average number of clusters as a function of $\epsilon$ and $p\_m$ in a non-biased environment.} In the figure the average number of clusters as a function of $\epsilon$ and $p\_m$ is showed for both settings: with one media with opinion $m\_1 = 0.0$ (left) and with three media with opinions $m\_1=0.05, m\_2 = 0.5, m\_3=0.95$ (right). The average number of clusters grows with the increase in the probability to interact with the one extreme media, but the growth of $\epsilon$ counters this effect (left). When the medias are $3$ the increase in the level of bounded confidence has the opposite effect and exacerbates fragmentation when combined with agent-to-media interactions.}

\label{fig:gammazero}

\end{figure}

In figure \ref{fig:gammazero} we can see the average number of clusters obtained in our simulations. The first row of both heatmaps shows the behaviour of the original Deffuant-Weisbuch model, which is then compared to the results of an increasing probability to interact with the media present in the model. These plots shows how increasing $p\_m$ the average number of clusters increases, however it is clear that there are two very different behaviour in the two different settings. If we have only one media, consensus is always reached for $\epsilon \geq 0.4$ and for $\epsilon \geq 0.5$ interacting with the media has no effect on the number of clusters. When we consider the three media setting, instead, a higher $\epsilon$ results in a higher fragmentation and the average number of clusters increases, instead of decreasing, with respect to lower values of $\epsilon$. This happens because while $\epsilon \leq 0.3$ the majority of the nodes will mainly interact with only one of the media - the one closest to their present opinion - and only a few nodes will move to the opposite side of the opinion spectrum; however, when $\epsilon \ge 0.3$ many nodes at any time step can be influenced by more than one media. This creates a very unstable dynamics, because nodes keep moving within the opinion spectrum and hardly cluster.

\subsubsection{Biased media interactions}

% Also the effects of the algorithmic bias, i.e. $\gamma > 0$, are different in the two different settings.\\ \\

Biasing both peer-to-peer and agent-to-media interactions, namely setting $\gamma > 0$, has very different effects in the two settings which we will briefly describe in the reminder of this sections, using entropy, the average number of clusters, the average pairwise distance, the average number of iterations at convergence, the mean opinion and the percentage of agents attracted by each media to give a full description of the dynamics and its results.

\textbf{One extreme media}

\begin{figure}

\centering

\includegraphics[width = 0.8\textwidth]{figures/hm media mo0.0 0.01MS\_avg\_ncluster groupedby\_eps.png}

\caption{\textbf{Average number of clusters with one extreme media.} The figure shows the average number of clusters for different values of the confidence bound $\epsilon$ as a function of $\gamma$ and $p\_m$.}

\label{fig:00ncluster}

\end{figure}

In the case we have one extreme media, namely $m\_1 = 0.0$ a higher probability to interact with such media and a higher bias both enhance fragmentation, like we can see from figure \ref{fig:00ncluster}, where the average number of clusters of the final distribution is plotted for different values of the confidence bound $\epsilon$ as a function of $\gamma$ and $p\_m$. We can see how in this case not only the number of clusters can grow due to the bias, but also due to a higher number of interactions with a single fixed external source of information, which is able to radicalise a portion of the population.

\begin{figure}

\centering

\includegraphics[width=0.8\textwidth]{figures/hm media mo0.0 perc\_00 groupedby\_eps.png}

\caption{\textbf{Percentage of nodes in the extremist media cluster.}}

\label{fig:extrnodes}

\end{figure}

If we consider the percentage of agents that hold opinions in the range $[0.0, 0.0+\lambda]$ we can see from figure \ref{fig:extrnodes} that this percentage is already around $20-30\%$ with a probability to interact with the media of only $10\%$. Moreover, while a high confidence bound is able to restore a situation of consensus, the mean opinion tends to go towards 0, until there is a full consensus around the extreme opinion for very high $\epsilon$ values and $p\_m$ values. In this case a higher bias means still a higher fragmentation, but, it also means that a small part of the population is able to separate themselves from the extremist consensus. The presence of one single media agent also fastens the convergence (the number of iterations at convergence slightly reduces as $p\_m$ grows) whether the final state is fragmented, polarised or a consensus and - differently from the baseline AB model - the presence of an algorithmic bias does not seem to slow down the convergence in a significant way. \\ \\

\textbf{Three media}

\begin{figure}

\centering

\includegraphics[width=0.8\textwidth]{figures/hm media mo0.05;0.5;0.95 0.01MS\_avg\_ncluster groupedby\_eps.png}

\caption{\textbf{Average number of clusters for different values of $\epsilon$}}

\label{fig:3mncluster}

\end{figure}

% \begin{figure}

% \centering

% \includegraphics[width=0.8\textwidth]{figures/hm media mo0.05;0.5;0.95 avg\_opinion groupedby\_eps.png}

% \caption{\textbf{Average mean opinion.}}

% \label{fig:3maverageop}

% \end{figure}

Adding multiple media outlets has different results on the outcomes of the opinion dynamics. Since active interactions can happen only with agents within one's confidence bound, for $\epsilon \leq 0.3$ - since media have a $0.45$ distance between them - the dynamics follows the baseline model behavior: fragmentation increases with both $p\_m$ and $\gamma$, but this increase is not strong: as we can see from \ref{fig:3mncluster} we go from two to three clusters for $\epsilon=0.2$ and from 1 to 3 clusters for $\epsilon=0.3$. Moreover,

% as we can see from figure \ref{fig:3maverageop}

agents tend to cluster around the media opinions, so the media \say{attract} the population and rapidly fragments them. When $\epsilon \geq 0.4$, however, an increase in $p\_m$ enhances fragmentation: agents remain sparser in the opinion spectrum, distributing themselves between the opinions of two media and - in some cases - maintaining an uniform distribution between 0.05 and 0.95; moreover fragmentation is also enhanced by a huge slowing down in convergence, or, more precisely, the absence of convergence, since nodes keep on interacting with different fixed opinions - the media ones - and with these interactions being active they keep shifting their position in the opinion spectrum, enhancing instability of the system; the mean opinion at the end of the dynamics is very variable, while in all other cases it's pretty stable around 0.5, and it depends on the fact that sometimes all the agents are attracted in the space between 0.05 and 0.5, while in other cases they are attracted between 0.5 and 0.95, while in others they converge in a few clusters. This different final state is visible also by comparing the average number of clusters with the average pairwise distance. While in the baseline model the number of clusters and the pairwise distances followed the same behaviour, in this case we have a very high number of clusters with a very low pairwise distance because agents are closer to each-other than a polarised situation, but their distribution have a very higher entropy, i.e. number of clusters. Pheraps counterintuively, if we also increase $\gamma$ - while increasing the probability to interact with media - this actually reduces the level of fragmentation in the system and brings it to consensus around the mean opinion (0.5), as we can see from the percentage of nodes in $[0.5-\lambda, 0.5+\lambda]$, probably due to ???

\section{Conclusions}

\label{sec:conlcusion}

A bounded confidence model with algorithmic bias and mass media agents was presented. We found that the presence of one extremist media is able to always radicalise a portion of the population, this portion depending on the probability to interact with this media and the filtering power of the algorithm. A higher number of media interactions brings a higher portion of the population towards the extreme, while increasing the filtering power of the algorithm fragments the population and actually reduces the number of final extremists, since it makes less probable for the agents on the opposite side of the opinion spectrum to interact with the media at all. When the number of media in the system is higher than one, this can highly change the dynamics: in the case of a low \say{open-mindedness} - i.e. $\epsilon \leq 0.3$ the population clusters around the media opinions on average, while for $\epsilon > 0.3$, the dynamics is very unstable up to a certain value of $\gamma$, because agents keep on changing their opinions due to the agent-to-media interactions, while a strong filtering power actually enhances consensus, when combined with a high probability of agent-to-media interactions.

\bibliography{references}

\end{document}