Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-023-01998-z OPEN

Social networks, disinformation and diplomacy: a dynamic model for a current problem

Alfredo Guzmán Rincón [1]✉, Sandra Barragán Moreno[2], Belén Rodríguez-Canovas[3], Ruby Lorena Carrillo Barbosa[4] & David Ricardo Africano Franco [4]

The potential of social networks for the circulation of disinformation as a strategy of diplomacy has been of great interest to the academic community, but the way in which it is propagated and modelled is still in its beginnings. This article aimed to simulate the propagation of disinformation in social networks derived from the diplomacy strategy, based on the elements of the system. The main research question that was opened up was how do the elements of disinformation derived from the social media diplomacy strategy interact to affect a susceptible population? For the design of the simulation model, system dynamics was used as the main technique in the re-search methodology in conjunction with statistical analysis. Five computational simulations were run for the adoption methods of susceptible and uninformed population, misinformation techniques and echo chamber. The model developed found that the diplomacy disinformation agent is able to spread its message efficiently through the bot outreach mechanism and only a part of the susceptible population unsubscribes to the disinformation agent's account. Significant differences were identified in the absence of paid outreach, bots and trolls in the propagation of information, and in the variation in the timing of disinformation propagation. Consequently, the developed model allows the understanding of the problem of disinformation as a strategy of diplomacy from international rather than local dynamics, as well as the effects of the use of each element in the system.

[1] Corporación Universitaria de Asturias, Bogota, Colombia. [2] Universidad de Bogotá Jorge Tadeo Lozano, Bogota, Colombia. [3] Universidad Complutense de Madrid, Madrid, Spain. [4] Universidad de Ciencias Aplicadas y Ambientales U.D.C.A, Bogota, Colombia. ✉email: alfredo.guzman@asturias.edu.co

Introduction

In recent years, social networks have positioned themselves as the preferred means of communication for connecting citizens and governments (Jahng, 2021; Guzmán & Rodríguez-Cánovas, 2021; Lazer et al., 2018; Wang, 2006), as they facilitate, mediate and speed up the interactions, which makes this type of network a space for the circulation of information of a massive nature, in which the expression and exchange of ideas and opinions is allowed in a generalised manner (Carlo Bertot et al., 2012), breaking traditional paradigms of communication between states and stakeholders (e.g., citizens and businesses) by moving from a one-way to a two-way approach (Guzmán et al., 2020), which has influenced all state functions, including diplomacy, in a cross-cutting manner (Manor & Segev, 2020). Social media communication has been widely adopted in diplomacy, understood as a systematised process in which international actors seek to achieve foreign policy objectives (Cull, 2011) resulting in closer contact between the international sender and the local receiver of information, thereby providing individuals with the possibility of communicating with diplomatic actors (Graffy, 2009).

In this context, the potential of social networks as a communication channel for diplomacy has been recognised, as they make it possible to build loyal communities by bringing senders and receivers closer together (Graffy, 2009); the achievement of effective and efficient communication with the stakeholders (Gebhard, 2016); budget optimisation as it is associated with lower costs and investments compared to traditional methods (Fjällhed, 2021); among others. However, at the edge of this potential, some governments have made use of this channel and the direct relationship with citizens online to systematically propagate disinformation and thus meddle in national issues of other sovereign states, influencing the opinion of citizens in order to benefit their own interests and fulfil some of their foreign policy objectives (Lazer et al., 2018; Cull, 2016).

As an example of this, the elections in the United States of America (USA) in 2016 can be mentioned, in which the Russian government, through its agencies, intermediaries, paid advertising campaigns, paid users, trolls and state-funded media, discredited the Democratic candidate Hillary Clinton in key USA election states. Initially, it was determined by the Office of the Director of National Intelligence (2017) that Russian intervention had the potential to swing the election in favour of Donald Trump and Moscow's interests; however, recent studies have indicated that the impact of disinformation in this election campaign would not have had such an impact (Guess et al., 2020). This is because the disinformation campaign focused on the already disinformed population and not on other susceptible populations (Guess et al., 2020; Gunther et al., 2019). More recently, disinformation continues to permeate social media for diplomatic purposes, as Agarwal and Alsaeedi (2020) identified how the Russian media RT and Sputnik initially accused NATO and the USA of creating the COVID-19 virus and using it to destabilise China's economy. Hence, disinformation as a strategy of diplomacy has regained relevance in the field of international relations (Fjällhed, 2021), and has become one of the main problems for the defence of states, as it develops in a new scenario such as social networks, in which information is disseminated at great speed and whose origin is difficult to trace, in addition to the intervention of new mechanisms for disseminating the messages that are specific to this type of network (McGonagle, 2017; Pamment et al., 2017).

Thus, studies related to the use of the strategy of disinformation from diplomacy and social networks have focused mainly on the documentation of cases, with the aim of understanding the elements involved in the dissemination of this type of information and the effects it has on citizens (e.g.: Lanoszka, 2019), There are many gaps in the understanding of the use of this strategy, due to the lack of previous experience in the field of international relations (Fjällhed, 2021), the lack of confirmation of its use by states, and the difficulty of finding declassified (uncensored) information from the governments concerned. Hence, authors such as La Cour (2020) recognise that, although progress has been made in understanding how this type of information is spread from other areas of knowledge, it is important to establish an approach directly related to diplomacy, due to the fact that local dynamics cannot fully explain how this information is disseminated at the international level which involves monetary resources and actors that go beyond traditional disinformation campaigns. In addition to the above, it is necessary to establish the patterns generated by disinformation as a strategy of diplomacy, based on the behaviour of individuals and the elements of the system itself, in order to generate strategies to mitigate the effects caused by this phenomenon, which affect multiple aspects of citizens' lives, such as the influence on their opinions and beliefs, the generation of disturbances, among others (Fjällhed, 2021; Lanoszka, 2019; La Cour, 2020).

This article aimed to simulate the propagation of disinformation in social networks derived from the strategy of diplomacy, based on the elements of the system documented in the literature. Thus, from the approach of modelling and diplomacy, we sought to provide a first approximation to the answer to the following question: how do the elements of disinformation derived from the social media diplomacy strategy interact to affect a susceptible population? With the answer to this research question, we gain an understanding of the dynamics and impact of disinformation generated through diplomacy on social media, focusing on how these elements influence people's opinions and beliefs, as well as the generation of disturbances in society. In addition to this, the response provides a comprehensive analysis of the mechanisms and consequences of disinformation in this context from a diplomatic perspective, offering a more complete view of how diplomatic actors strategically use social media to achieve their objectives. Furthermore, an approach was also sought to address the following research questions:

- What impact do bots and trolls, as elements of the digital world, have on the spread of disinformation on social media as a strategy of diplomacy?
- What is the impact of social media in delaying the activation of disinformation mechanisms as a strategy of diplomacy?
- What are the effects of the echo chambers that social media algorithms foster on diplomacy-generated disinformation?

By fulfilling the aim and answering the research questions, two contributions are made to the study of disinformation on social media from the perspective of diplomacy. Firstly, a simulation model based on system dynamics is presented, with which specialists in international relations can generate scenarios that approximate the way in which diplomacy agents used this medium to achieve their objectives, eliminating, to a certain extent, possible biases in their conclusions due to not having all the information available in terms of time. Secondly, a holistic approach to disinformation in social networks is presented, incorporating elements that interact at the same time (for example, bots, trolls and the payment of campaigns to promote disinformation) and that had not been addressed in studies related to diplomacy, which allows for a more realistic view of the behaviour of the disinformation system in social networks where agents of diplomacy intervene.

Accordingly, this article is structured into four main sections. The first section conceptualises disinformation, the use of this strategy of social media diplomacy and the elements of the system involved in such a strategy; the second one sets out the

methodology used for the development of the dynamic model and the corresponding simulations to solve the research questions; the third section presents the model, together with the results of the computational simulation defined in the methodology; and the fourth one presents the discussion and conclusions.

Theoretical framework and background

Conceptual delimitation of disinformation. The term disinformation has become common in journalistic contexts and political language in recent years (Rodríguez, 2018), relating as a current phenomenon derived from web-based technologies; however, the conceptualisation of this term occurred at the beginning of the 20th century, having its origins in the political sphere, when it was used by the French after the First World War to refer to actions directed from inside and outside the country to prevent the consolidation of the communist regime in France (Durandin, 1993; Jacquard, 1988) by discrediting its political and economic systems, based on the propagation of false information. Since that time the term has evolved to refer to any deviant information that has the intent and effect of distorting and misleading a target audience in a predetermined way (Innes, 2020).

It is necessary to clarify that disinformation, being a colloquial expression, is often misinterpreted by social actors, assigning conceptualisations and characteristics that do not correspond to its scope (Fallis, 2015), hence the need for a conceptual delimitation of the term. The first delimitation relates to the intentionality with which it is recognised that such information is not the result of a mistake but is specifically intended to deceive (Fallis, 2015; Fallis, 2011), exerting influence and control over the receptors to make them act according to the sender's intentions, therefore it is clearly a deliberate phenomenon (Van Dijk, 2006). The second one corresponds to the lack of truth, because disinformation can be by commission, in which a falsehood is knowingly transmitted (Rodríguez, 2018; Durandin, 1993), or by omission, when relevant data is concealed so that it is not possible to obtain the veracity (McGonagle, 2017). Having stated that, the misinforming's operation focuses on giving the appearance of truth to an event that is not true, so that the receiver trusts the information and takes it as real (McGonagle, 2017).

The third description is closely related with the channels of communication, because the sender uses them in order to massify the disinformation (Agarwal & Alsaeedi, 2020); hence, the intention to misinform it is not only enough, but an effective intermediation is required resulting in accordance with the point of view of the creator of the disinformation content (Rodríguez, 2018). While the emitters of disinformation had relied on traditional means of communication, which have been widely documented at the time (Desantes-Guanter, 1976; Chiais, 2008), the internet, with its ability to disseminate both true and false facts, has changed the landscape, in which communicators can reach out directly to users and amplify the message to a larger target group (Lazer et al., 2018). And the fourth delimitation of this concept and the point of intersection between the intention, the creation of the message (lack of truth) and the communication channels is the organisation in which it is planned, how the activities related to disinformation will be executed, ranging from the definition of the target audience to the evaluation of the efficiency of the misinformative message, represented in the opinions and actions created in the citizenship (Jacquard, 1998).

However, in the field of diplomacy, disinformation should not be confused with propaganda, given the existence of a fine line between the two concepts. Thus, propaganda is associated with a message in order to keep the receiver under control, benefiting the sender in the medium and long term (Desantes-Guanter, 1976). This is exemplified in the case of dictatorial or absolutist regimes. Disinformation from diplomacy seeks objectives that do not lead to this type of control over the population, but rather seeks to unbalance one or several states in the short term.

Social media disinformation as a strategy for diplomacy. Disinformation as a strategy of diplomacy aims to spread false information to unbalance foreign states by confusing and misleading their citizens (Agarwal & Alsaeedi, 2020; Gerrits, 2018), in this way, the state sending the message benefits from the disagreement generated in the society, the change of policies due to pressure from citizens on governments, as well as increasing its international presence and power, and fulfilling its international policy objectives (Fjällhed, 2021; Cull, 2016).

In this context, it is acknowledged that the use of this strategy is not a recent development in diplomacy, since the US and its allies, as well as the Soviet Union, began to broadcast disinformation about its rival during the Cold War (Chiais, 2008; Gerrits, 2018), making use of traditional channels of communication such as television, radio and newspapers. However, like any strategy, whatever its scope, it has evolved and incorporated new elements from a changing environment, hence disinformation has started to spread on internet-based communication media channels such as social media. The digitalisation of disinformation and its transmission on this type of network has resulted in a change in its potential, since what is new is not the message or the change of channel, but the speed at which it is spread and the impact that false information disseminated in this medium can have on the population, hence the importance of analysing disinformation on this channel (Vériter et al., 2020).

Therefore, disinformation as a strategy of diplomacy in recent years has concentrated its efforts on social networks, due to the mechanisms they have for the amplification of the message (e.g., echo chambers, bots, trolls, etc.) and, which allow a larger number of users to be exposed to disinformation (Bjola, 2018). Hence, there is growing interest in the study of the use of this strategy by both governments and the academic community. Thus, advances in diplomatic understanding have focused on documenting countries' use of disinformation, concentrating on Russia and China (e.g.: La Cour, 2020; Lupion, 2018; Mölder & Sazonov, 2018) because of its foreign policy towards Western countries, especially the US and those in Western and Southern Europe, which have shown the potential to interfere in democratic processes such as elections (La Cour 2020; Bayer et al., 2019); the possibility of polarising citizens' opinions through the spread of conspiracy theories, the exacerbation of radical and supremacist (racist) thinking (Faris et al., 2017); and the diminishing credibility of traditional media and mainstream institutions (Bennett & Livingston, 2018).

Despite the advances described in the literature, the analysis of disinformation as a strategy of diplomacy has been rather limited, focusing on the description of case studies related to the effect of the implementation of the strategy and the evaluation of citizens' perceptions. This is largely due to the difficulties involved in the study of this strategy, especially in terms of tracing the origin of disinformation, making it impossible to determine the attribution factor and the study from the origin of the issuer (Gerrits, 2018). Therefore, there is a need to explore other aspects of disinformation and its use in diplomacy, such as its diffusion, building on existing theory and thus proposing models and new scenarios that allow for new insights that have not been addressed.

Propagation of disinformation and elements of diplomacy's use of this strategy in social networks. The propagation of

disinformation in many ways is similar to the way in which an epidemic spreads as there are a number of uninformed (infected) individuals who seek to affect a susceptible population by transmitting the message with false information, thus models of the spread of disinformation are based on the SIR (Susceptible-Infected-Recovered) model (e.g.: Zhao & Wang, 2013a; Rapoport & Rebhun, 1952). Subsequent studies have complemented the basis of this model, including and eliminating elements, such as the SIRaRu model, which allowed us to understand the behaviour of disinformation in homogeneous and heterogeneous communities (Wang et al., 2014), the SEIR model (Susceptible-Exposed-Infectious-Recovered), which established the possibility of quantifying the duration of the disinformation outbreak (Di et al., 2020), the SIR model for complex social networks (Zhao & Wang, 2013a), among others.

While the above models explain the spread of misinformation, they have generally focused on traditional communication channel mechanisms, and therefore do not incorporate the characteristic elements of social media such as types of reach (organic, paid and by invitation) or level of engagement. Advances in models of the spread of disinformation in social networks have been more recent, focusing on pattern detection and incorporating context for predicting misinformation dissemination behaviour (Bian et al., 2020; Ma et al., 2015) and maximising user influence, where an individual with many followers can generate a massive disinformation cascade (Li et al., 2020).

In view of these developments, models of disinformation propagation have focused on other areas of knowledge not directly related to diplomacy, so that the construction of these models lacks some elements that are incorporated in the use of this strategy by governments, thus varying the overall behaviour of the propagation system. It is worth remembering that disinformation is intentional (Gerrits, 2018), which is why its use in diplomacy obeys strategic planning, seeking to maximise the effects of the message on a population (Vosoughi et al., 2018). Therefore, the social media profiles of the disinformation agent seek to attract the greatest number of target audiences (Hollenbaugh & Ferris, 2014) and therefore make use of organic, paid and invitation-based outreach to attract the target population and convert them into a population susceptible to viewing the disinformation message (Buchanan & Benson, 2019).

With the linking of the susceptible population to the disinformation profiles, the process of sending the message through the various media begins, highlighting organic reach (Buchanan & Benson, 2019), paid reach (Bodine-Baron et al., 2016), bots (Helmus et al., 2018) and trolls (Starbird, 2019), exposing the message in a systematic way to establish the misinformed population. However, this is done once there is a consolidated susceptible population, there is a delay between the susceptible population and the moment when they are disinformed, as the disinformation agent seeks to amplify the effect of the disinformation, taking advantage of the possible reactions and comments to the message sent. The delay in sending the disinformation is only justified if one wants to maximise the organic reach in the first stage. Regarding the means available to the misinforming agent, it should be noted that organic and paid reach are typical of the dynamics of social networks, facilitated by the algorithm, and in which the misinforming message is subject to the rules of the social network. Otherwise, Bots and Trolls are used to amplify the message in parallel to the dynamics of social networks. These last two elements were incorporated into Russia's diplomatic disinformation strategy in the US elections (Helmus et al., 2018).

Under the systematic exposure of the biased message, inhich the misinformed population is involved, it has been shown that, by constantly interacting with the message, an echo chamber is generated, which reinforces it (Bessi et al., 2015; Garrett, 2009). This leads to a higher level of interaction of the uninformed population with the message (engagement level), which hinders exposure to truthful content, resulting in the uninformed population not becoming the informed population (Quattrociocchi et al., 2016), thus achieving one of the ultimate goals of disinformation as a strategy of diplomacy. However, the ability of the uninformed population to seek additional information in media other than social media is recognised as a final element, which translates into a correction rate, leading to a reduction in it (Chiang & Knight, 2011; Entman, 2007). In this scenario, the now-informed population must make the decision to stop following the misinforming agent's profile(s), or to continue to be in contact with them and remain part of the susceptible population. Table 1 summarises the elements identified in the literature that relate to the strategy of disinformation in diplomacy.

Methodology

Design. In order to fulfil the proposed objective and answer the research questions, this article was based on the development of a computational simulation model whose main technique was System Dynamics, considering Bala et al. (2017), Forrester (2013) and Sterman (2012) as theoretical references. Thus, the choice of this computational modelling and simulation method is based on the recognition of the complexity of the disinformation propagation system because of the diplomacy strategy, in which multiple elements are involved, and whose behaviour is non-linear, multi-causal and time-lagged (Bal et al., 2017). Thus, for the development of the model, the elements identified in the literature (Table 1), which are employed in diplomacy to propagate disinformation, were used. With these elements, we proceeded to conceptualise the model and its formal construction, following the procedure suggested by Bala et al. (2017).

In this sense, the diagram of flows and levels of the model was constructed, understanding this as the underlying physical structure of the system, where the stocks represent the state or condition of the system in a defined period, while the flows represent the change in function of the decisions taken in the system. In this phase, the variables that allow the system's behaviour to be represented must be defined. Subsequently, the differential equations representing the cause-effect relationships between the variables were established. With these equations, the parameters were determined, assigning numerical values to each of the variables. Thus, the parameters were based on the US Senate Select Committee on Intelligence reports on Russian interference in the 2016 US presidential election, and on previously developed studies on the elements of the system. In addition, estimates were made for the variables using disaggregation, aggregation and multiple equation techniques. Finally, the internal consistency of the model was tested to establish that the representation of the system was adequate within the scope of the study's purpose.

The proposed model. Figure 1 presents the proposed model of flows and levels based on the SIR model and advances in other fields of knowledge related to the propagation of disinformation, as well as the characteristics of this diplomacy strategy. This model was designed with seven levels: five measured in number of persons, one in number of B and one in number of T.

The model also considered other variables in addition to those defined in Table 1 that are required for the functioning of the disinformation system as a diplomacy strategy, and which together regulate the levels of the model, as presented in Table 2.

The structure of the model allowed us to understand how disinformation spreads as a strategy of diplomacy based on three

Table 1 Elements of disinformation as a strategy of diplomacy.

Element	Abbreviation	Conceptualisation
Target population	PO	The set of individuals targeted by disinformation on social media. This has specific demographic, socio-economic, psychological and behavioural attributes, which are analysed by the disinformation agent to define the ways and means of disinformation.
Susceptible population to misinformation	PS	People who had a relationship with the disinformation agent's social media accounts, and who are now part of his network of contacts.
Misinformed population	PD	A portion of the population susceptible to misinformation that encountered the misinforming message, and that in a first state may identify with the message or reject it to become an informed population.
Informed population	PIn	Misinformed population who encountered truthful information and accepted it, reinforcing, or changing their ideas and beliefs in positive way.
Unsubscribed population	PU	Informed population that ceased to be in contact with the social media accounts of the disinformation agent.
Organic outreach	ao_n	Number of users who, through the algorithm's free distribution methods, encounter posts from an account, allowing them to subscribe to a relationship with the account or to access the content generated. In this case, n. represents the number of times the variable will be used in the system with different values.
Paid Scope	ap	Number of users who by paid methods (cost per click or per thousand) encounter publications from an account, and which allow them to subscribe to a relationship with the account or to view the content generated.
Outreach by invitation	ai	Number of individuals who encounter an account, through a direct invitation to join the network of contacts.
Bots	*B*	Computer-driven automated accounts that systematically spread disinformation through their organic reach which can be deactivated by the social network when detected.
Trolls	*T*	Anonymous accounts that post the misinforming message or comment on it to amplify the disinformation. These accounts are controlled by a user on the website and can be blocked through reports made by users in accordance with the social network's terms and conditions.

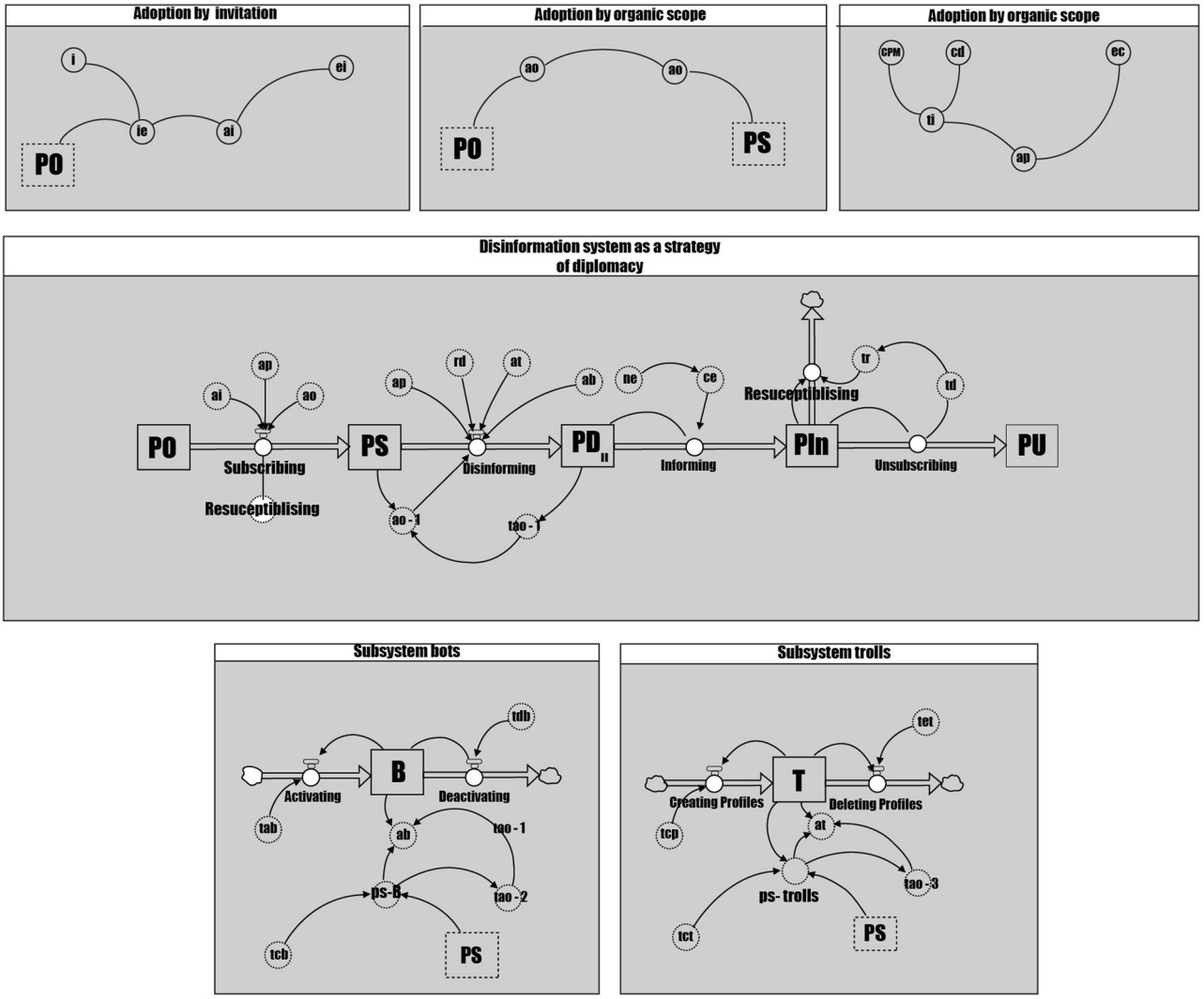

Fig. 1 . Model of flows and levels of disinformation as a strategy of diplomacy.

Table 2 Other variables required for model development.

Element	Abbreviation	Conceptualisation
Invitation fee	i	Percentage of POs that are contacted by the disinformation agent via direct invitation to be part of their network of contacts.
Effectiveness of invitation	ei	Corresponds to the effectiveness of the acceptance of the invitation sent by the disinformation agent.
Organic reach rate	tao_n	Percentage of publications displayed by the algorithm distribution methods. This rate is defined according to the number of PS contacted. The "n" represents the different types of organic outreach rates. For the purposes of the article, there are three.
Costs per mille	CPM	Constant representing a thousand impressions paid for by the disinformation agent. The term does not represent a monetary value but the number of impressions paid by the advertiser.
Distortion campaigns	cd	Number of paid advertisements (selected method is CPM) by the disinformation agent to display to both PO and PS in a period t.
Effectiveness of campaign	ec	Effectiveness of the campaign carried out, representing the acceptance of the contact with the disinformation agent or of the message sent.
Resusceptibility rate	tr	Percentage of PIn who do not unsubscribe from the disinformation agent's accounts after having encountered truthful information.
Bots contact rate	tcb	Percentage of PS that have contact with Bots.
Trolls contact rate	tct	Percentage of PS that have contact with Trolls.
Delayed disinformation	rd	Delay in the start of disinformation. This corresponds to the initial t at which the message starts propagating. This delay is developed under the delay function.
Level of engagement	ne	Refers to the rate of user interaction with the disinformation message, usually represented in likes, comments, etc.
Echo chamber	ce	Corresponds to overexposure to disinformation, as a result of the level of engagement of the social network user. This variable is a result of the level of engagement the susceptible population has with disinformation. It ranges from 0 to 1.
Bots activation rate	tab	Rate at which new Bots are activated (created) in a period t.
Bots deactivation rate	tdb	Rate at which Bots are deactivated (removed) in a period of time t. This event occurs when the social network detects the fake profile.
Troll profile creation rate	tcpt	Rate at which new troll profiles are created over a period of time t.
Troll profile removal rate	tet	Rate at which troll profiles are deleted or blocked over a period of time t. When this event happens are reported by the PD or PIn
Disengagement rate	td	Percentage of PIn who remove the disinformation agent from their social media contacts.
Bots outreach	ab	Corresponds to the number of people who have come into contact with disinformation as a result of the operation of the bots.
Trolls outreach	at	Corresponds to the number of people who have come into contact with disinformation as a result of the troll operation.

assumptions. The first was that OP was fixed, so it did not increase or decrease due to effects other than PS formation. Second, that cd was the same in both the susceptibility adoption process and the disinformation process. Third, the model defines the growth of the amount of B and T with exponential growth. In this sense, it is assumed that its growth is not a function of the amount of monetary resources of the disinforming agent, but of the agent's need to have as many B and T as possible to spread disinformation. To eliminate this assumption, in this section the model can be adapted to the mechanism described by Guzmán et al. (2022). Under the technical conditions of non-negativity of the variables (i.e. their domain is restricted to 0 or positive numbers) and that $t = 0, 1, 2\ldots, 180$, the model was represented by the following system of differential equations.

Target population:

$$PO_{(t)} = \left[PO_{(t-1)} - \left[\left(PO_{(t-1)} \times i \times ei\right) + \left(PO_{(t-1)} \times tao\right) + (CPM \times cd \times ec)\right]\right] dt \quad (1)$$

Susceptible population:

$$PS_{(t)} = \left[PS_{(t-1)} + \left[\left(PO_{(t-1)} \times i \times ei\right) + \left(PO_{(t-1)} \times tao\right) + (CPM \times cd \times ec) + \left(PIn_{(t-1)} \times tr\right)\right] - \left[f\left(x_t, x_{t-\tau}, t\right)dt; t \geq t_0\right]\right] dt \quad (2)$$

It is worth noting that $f\left(x_t, x_{t-\tau}, t\right)dt; t \geq t_0$ mathematically describes the delay of an action, for our case of the onset of disinformation propagation. The above apply to Eqs. 2 and 3. Where x_t is equal to:

$$x_t = \left[\left(PS_{(t-1)} \times tao_1\right) + (CPM \times cd \times ec) + \left(PS_{(t-1)} \times tcb \times tao_2 \times B\right) + \left(\begin{matrix} PS_{(t-1)} \times tct \\ \times tao_3 \times T \end{matrix}\right)\right] dt \quad (2.1)$$

In turn:

$$B_{(t)} = \left[B_{(t-1)} + \left(B_{(t-1)} \times tab\right) - \left(B_{(t-1)} \times tdb\right)\right] dt \quad (2.1.1)$$

$$T_{(t)} = \left[T_{(t-1)} + \left(T_{(t-1)} \times tcpt\right) - \left(T_{(t-1)} \times tet\right)\right] dt \quad (2.1.2)$$

Disinformed population:

$$PD_{(t)} = \left[PD_{(t-1)} + \left[f\left(x_t, x_{t-\tau}, t\right)dt; t \geq t_0\right] - \left(PD_{(t-1)} \times ce\right)\right] dt \quad (3)$$

Informed population:

$$PIn_{(t)} = \left[PIn_{(t-1)} + \left(PD_{(t-1)} \times ce\right) - \left[\left(PIn_{(t-1)} \times td\right) + \left(PIn_{(t-1)} \times tr\right)\right]\right] dt \quad (4)$$

where tr is equal to:

$$tr = [1 - td]dt \quad (4.1)$$

The value of ce depends on the value of ne, being this represented in a graphical function (see Table 3), the above is

represented:

$$ce = f(ne)dt \tag{4.2}$$

Unsubscribed population:

$$PU_{(t)} = \left[PU_{(t-1)} + \left(PIn_{(t-1)} \times td\right)\right]dt \tag{5}$$

Having said that, the initial parameters of the dynamic model are presented in Table 3.

Model validation. Regarding the validation process, Schwaninger and Groesser (2020) recognise that system dynamics-based models can be validated using both quantitative and qualitative methods. Thus, three major categories of validation are distinguished: model context, model structure and model behaviour. In the case of this article, the validation of the model was based on the model structure category. Therefore, the model structure tests aim to increase confidence in the structure of the theory created about the mode of behaviour of interest. In this sense, structure tests evaluate whether the logic of the model is in line with the corresponding structure in the real world (Schwaninger & Groesser, 2020). The test used was sensitivity analysis to parameter changes.

This validation method "evaluates changes in the model's behaviour by systematically varying input parameters" (Schwaninger & Groesser, 2020). This validation test reveals the parameters to which the model is highly sensitive, through numerous simulations with changes in parameters randomly within a range defined by the modelling. Thus, a model is considered valid when the numerical values of the simulation results change, but the model's behaviour remains consistent.

This validation test can reveal the degree of robustness in the model's behaviour and, therefore, indicate to what extent the conclusions based on the model could be affected by uncertainty in parameter values (Schwaninger & Groesser, 2020). For the purposes of this study, the following variables were modified by ± 10% of their initial parameter values (see Table 3): ec, i, rd, tab, tdb, ne, tcpt and tet. For the variables cd, the modification was made in the range of 0–20, and for rd, between 65 and 85 days. If modifying a variable resulted in negative values, the minimum value for sensitivity analysis was set to 0. A total of 100 scenarios were simulated with a uniform distribution for all variables.

Sensitivity analysis was performed on the model's stocks of PO, PS, PD, Pin and PU, as shown in Fig. 2. Numerical sensitivity was observed in the analysed stocks, indicating that the values change significantly with the parameters; however, the system's behaviour remains consistent for all stocks.

Based on calculations using a 95% confidence interval (CI), it is estimated that the number of people exposed to misinformation in the PO category, at time $t = 180$, will range between 0 and 757,000 individuals (Fig. 2a). Similarly, within the same interval and period, in the PS category (Fig. 2b), the susceptible population is expected to be between 0 and 923,000 people. As for the PD category (Fig. 2c), the number of misinformed individuals is estimated to range from 227 to 203,000. Furthermore, with a 95% CI and for $t = 180$, it is projected that the informed population (Pin, Fig. 2d) will range from 138 to 69,200 individuals. Finally, regarding the number of unsubscribed individuals (PU, Fig. 2e), it is estimated that the values will be within a range of 1,060 to 76,000.

However, the behaviour of the system after day 150 is explained by the fact that the target population of the disinformation agent has reached its limit, as shown in Fig. 2a, b. In the case of PD, PIn and PU stocks, the behaviour is derived from the confluence of the variables involved in the model flows. For these three variables the behaviour presents peaks and troughs due to the extreme conditions, being this represented in the quartiles simulated in the sensitivity analysis, changing only the numerical value of the stocks.

Simulations and data analysis. With the proposed model, we proceeded to establish the effect of the different elements of the system through computer simulation, for which modifications were made to the parameters established in the initial model (see Table 3). It should be noted that in the execution of the

Table 3 Initial parameters of the model variables.

Element	Type	Initial value	Units
PO	Stock	1,000,000	People
PS	Stock	1	People
PD	Stock	0	People
PIn	Stock	0	People
PU	Stock	0	People
B	Stock	1	Bots
T	Stock	10	Trolls
ce	Variable	Graph(ne)(0.00,1.00),(0.100,0.67),(0.20,0.44),(0.30,0.30)...(0.90,0.02),(1.00,0.01)	NA
i	Variable	5	%
ei	Variable	10	%
tao - n	Variable	Graph (PO o PS) (0,0.000042) ...(10,000,0.000042)...(11,000,0.000013)...(100,000,0.000013) ...(101,000,0.000003)...(1,000,000,0.000003)	NA
CPM	Variable	1000	Impressions
cd	Variable	10	campaigns / day
ec	Variable	15	%
tcb	Variable	20	%
tct	Variable	40	%
rd	Variable	70	days
ne	Variable	15	%
tab	Variable	3	%
tdb	Variable	0.1	%
tcpt	Variable	3	%
tet	Variable	0.08	%
td	Variable	8	%

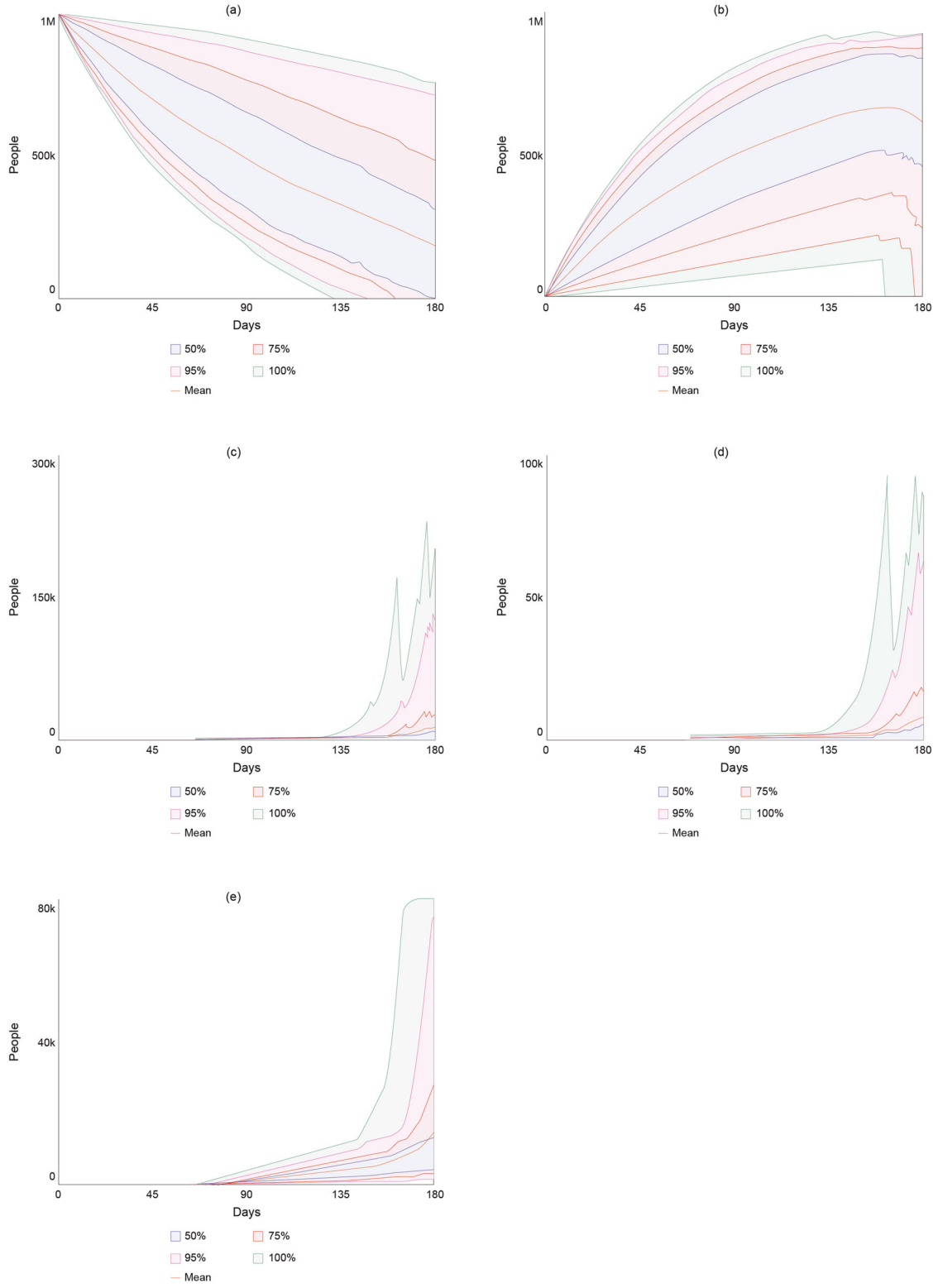

Fig. 2 Model sensitivity analysis. a PO sensitivity analysis. **b** Sensitivity analysis of PS. **c** Sensitivity analysis of PD. **d** Sensitivity analysis of PIn. **e** Sensitivity analysis of PU.

simulations only the parameter indicated in Table 4 was modified, and the others retained their initial values shown in Table 3, and the results on the levels of the system were named with the simulation code assigned in Table 4, followed by the name given to the level. RQ1 was answered by the model and RQ2, RQ3 and RQ4 were answered by the simulations.

Based on the results of the developed simulations, system dynamics-based models can be either deterministic or stochastic. In the case of the present model, it is deterministic because it does not consider variables with random parameters. Therefore, it is assumed that the causal relationships between the system variables are known and constant over time. In other words,

Table 4 Computer simulations.

Code	Simulation	Modified parameters	Units
Sim - 1	Adoption methods susceptible population and misinformation	cd = 0	campaigns/days
Sim - 2	Method of misinformation	$B = 0$	Bot
Sim - 3	Method of misinformation	$T = 0$	Troll
Sim - 4	Method of misinformation	rd = 30	Days
Sim - 5	Echo chamber	ne = 5% y ne = 40%	%

the behaviour of the system is fully determined by the rules and relationships established in the model. This means that if the simulation is run multiple times with the same parameters and initial conditions, the same result will be obtained each time without random variations. Hence, for the subsequent statistical analyses described, it is not necessary to run the simulations multiple times.

Thus, to test for statistically significant differences between the initial behaviour of the system and those generated with the modified parameters, the average levels of the model were compared. The Kolmogorov-Smirnov statistic was applied to check whether the data fit a normal distribution (p-value > 0.05), and it was found that the data did not follow a normal distribution. In this way, to establish the difference in the medians between the behaviour of the system with the initial parameters and the modified parameters, the Wilcoxon test was used, considering this difference with a p-value < 0.05. In this way, it was possible to answer RQ2.

Finally, the computational work on the model and simulations was developed in Stella Architect software version 3.3. The following model settings were considered: initial time = 0, final time = 180, $\Delta t = 1/10$, time units in days and selected Euler integration method. SPSS software version 25 was used for the statistical analyses.

Results

Under the initial conditions of the model, it was observed that, in the 180 days simulated, the PO decreased by 84.3%, so that 843,000 people were susceptible to being uninformed, however, the final PS was 691,722 people (Fig. 3a). The diplomacy 's disinformation agent managed to spread the message to a total of 267,275 people, 135,463 of whom had previously been misinformed. Thus, the PIn during the 180 days was 148,117 people of whom only 11,779 (PU) took the decision to cancel their subscriptions to the disinformation agent 's accounts. On the other hand, on average since the start of the disinformation activity, the agent managed to impact 1476 people each day, with the 176th day being the day of greatest growth with 3335 people (Fig. 3b). Similarly, a growth in PIn was evidenced (Fig. 3b), which represented a decrease in the difference between this population and PD, being 4.69 times at $t = 71$ to 1.70 at $t = 180$, however, the value of this difference on average was 1.83 times.

The behaviour of B and T showed an exponential growth of B and T from one to 411,036 ≈ 412 and 314, respectively (Fig. 3c). Regarding the dissemination methods used by diplomacy to disinform, it was shown that in the case of ap, for any value of t, it is constant disinforming 1200 people per day, compared to the other disinformation mechanisms for $t = 180$, ao 1 managed to misinform 29 people, at 261 and finally ab 4,580. In the case of the decrease of ab for day 100 (Fig. 3d), it is a consequence of the tao-2 effect, since the more followers the bots have, the more the organic reach of the publications they make is limited. This is due to the fact that the more susceptible population the bot has in the social network, the fewer people who can see the disinformation. This strategy is used by social networks to force accounts with a large reach to pay for users to see their publications. Figure 3d shows the behaviour of the disinformation methods.

With regard to the comparison of the behaviour of the original system and simulation one (Sim-1), it was found that there are statistically significant differences in the absence of cd, which is represented in that the levels of PO and Sim-1 PO ($z = -11.63$, p-value < 0.001); PS and Sim-1 PS ($z = -11.63$, p-value < 0.001); PD and Sim-1 PD ($z = -9.10$, p-value < 0.001); PIn and Sim-1 PIn ($z = -9.10$, p-value < 0.001); and, PU and Sim-1 PU ($z = -9.10$, p-value < 0.001) changed between the run simulations. Thus, for $t = 180$, which resulted in the number of uninformed, informed and unsubscribed people in the disinformation agent's account decreasing by 1,355,864 and 10,506 people, respectively. All this behaviour is presented in Fig. 4b.

For Sim-2, statistically significant differences were found in the absence of B in the propagation of disinformation as a strategy of diplomacy. Thus, the levels of PO and Sim-2 PO ($z = -6.95$, p-value < 0.001); PS and Sim-2 PS ($z = -9.06$, p-value < 0.001); PD and Sim-2 PD ($z = -9.06$, p-value < 0.001); PIn and Sim-2 PIn ($z = -9.02$, p-value < 0.001); and PU and Sim-2 PU ($z = -8.81$, p-value < 0.001) changed between the run simulations. In this scenario, for $t = 180$, it was established that PO was lower by 14,000 persons (Fig. 4c), that is, in the absence of the misinforming element PD, PIn and PU decreased by 514,360 and 1346 persons, respectively (Fig. 4d).

However, in the case of Sim-3, statistically significant differences were established in the absence of T. The levels of PO and Sim-3 PO ($z = -6.92$, p-value < 0.001); PS and Sim-3 PS ($z = -9.06$, p-value < 0.001); PD and Sim-3 PD ($z = -9.06$, p-value < 0.001); PIn and Sim-3 PIn ($z = -9.02$, p-value < 0.001; and, PU and Sim-3 PU ($z = -8.81$, p-value < 0.001) changed between the run simulations. In this way, it was determined that in $t = 180$, The PS was lower by 13,000 persons (Fig. 4e), and that the levels of PD, PIn and PU decreased by 497,343 and 1252 persons respectively, as shown in Fig. 4f.

For Sim-4, statistically significant differences were found for the variation of rd, i.e. the time at which the disinformation agent initiates the propagation of the message. Thus, the levels of PO and Sim-4 PO ($z = -10.55$, p-value < 0.001); PS and Sim-4 PS ($z = -10.62$ p-value < 0.001); PD and Sim-4 PD ($z = -2.86$, p-value < 0.001); PIn and Sim-4 PIn ($z = -3.03$, p-value < 0.001); and PU and Sim-4 PU ($z = -10.62$, p-value < 0.001) changed between the run simulations. In this scenario, in $t = 180$, the PS increased by 15,000 people (Fig. 4g), while PD and PIn levels decreased by 186 and 184 people, respectively, while PU increased by 3,026 people (Fig. 4h).

Finally, compared to the scenarios presented in Sim-5, statistically significant differences were found with both the increase and decrease of ne in the system levels as shown in Table 5, whereby the levels changed between the run simulations. Thus, for the case of ne = 5% at $t = 180$, the PO decreased by 12,000 persons (Fig. 5a), which meant that for the PD, PIn and PU levels it decreased by 1,215,331 and 1151 persons, respectively (Fig. 5b) regarding SIM-1 . When one equals 40% for the same t, a decrease in PS by 15,000 persons was observed (Fig. 5c), however,

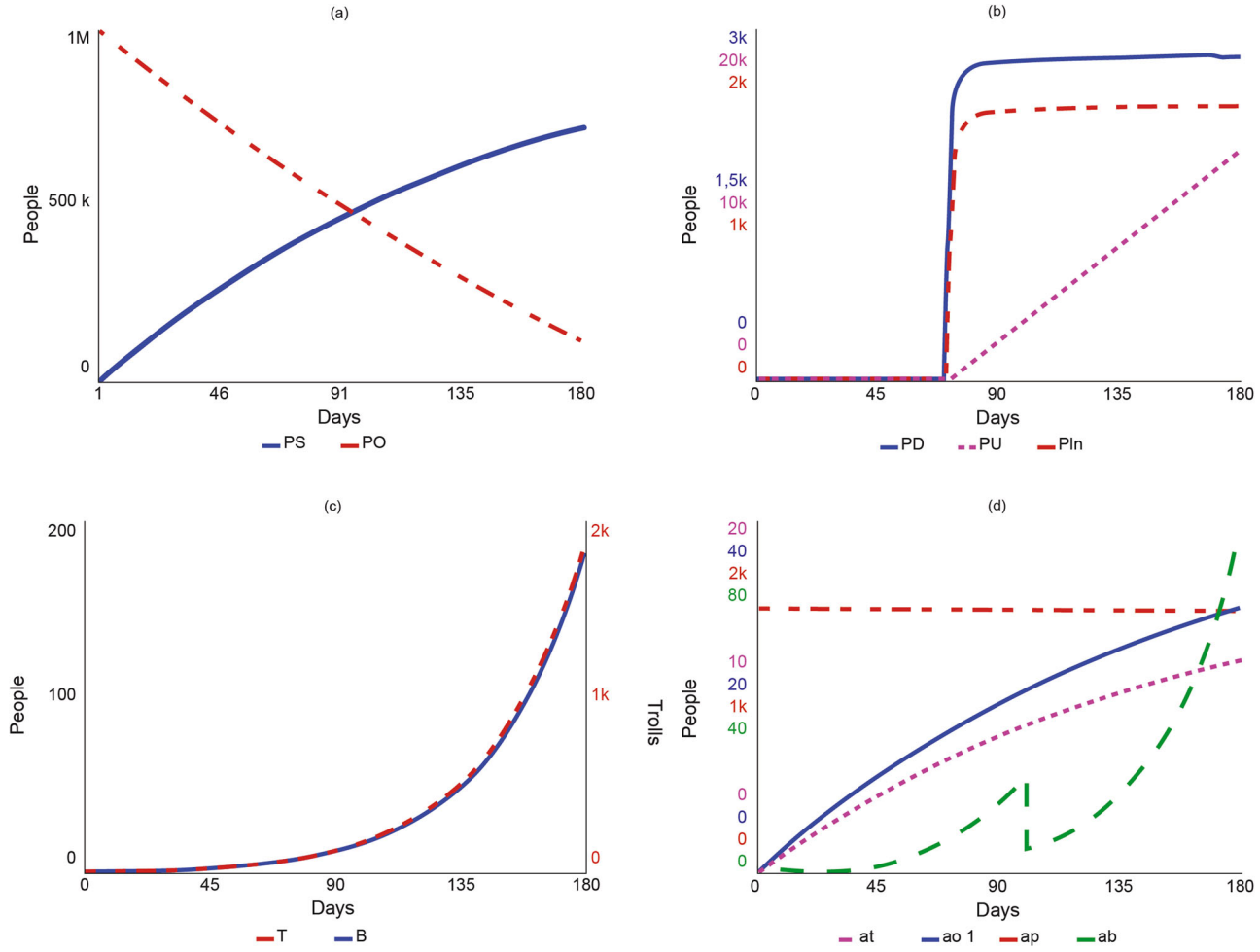

Fig. 3 Simulation results of the model with initial parameters. a System behaviour at PO and PS levels. **b** System behaviour at PD, PU and PIn levels. **c** System behaviour at B and T levels. **d** Behaviour of variables at, ao 1, ap and ab.

Table 5 Initial parameters of the model variables.

Variables	Statistic	ne = 5%	ne = 40%
PO y Sim-5 PO	*z*	−5.38	−9.12
	p-value	<0.001	<0.001
PS y Sim-5 PS	*z*	−9.10	−8.25
	p-value	<0.001	<0.001
PD y Sim-5 PD	*z*	−9.10	+9.10
	p-value	<0.001	<0.001
PIn y Sim-5 PIn	*z*	−7.60	−9.10
	p-value	<0.001	<0.001
PU y Sim-5 PU	*z*	−3.56	−9.10
	p-value	<0.001	<0.001

PD increased by 3648 persons while PIn and PU decreased by 307 and 1538 persons respectively (Fig. 5d).

Discussion and conclusions

The study aimed to simulate the propagation of disinformation in social networks derived from the strategy of diplomacy, based on the elements of the system. In this sense, and in accordance with the results presented above, it was possible to provide an initial approximation to answering the research questions through modelling and diplomacy. A conceptual, mathematical and simulation model was established to understand how disinformation spreads on social networks as a diplomacy strategy, taking the SIR model as a basis and modifying it to include the elements of this diplomacy strategy documented in the literature (e.g., paid, and organic reach, bots and trolls). It is important to highlight that the model is adaptable in parameters to any social network, or multiple in case of using layer or array-based modelling.

Compared to the original model proposed by Rapoport and Rebhun (1952), and to the models of disinformation in social networks in contexts other than diplomacy, such as those of Bian et al. (2020), Li et al. (2020) or Guzmán et al. (2022), the model proposed here differs in two aspects: the first one relates to the target population, which is defined by the agent of international diplomacy, given that it focuses its efforts on a limited audience with specific characteristics, which it seeks to influence through the disinformation message, this aspect was not taken into account in other non-diplomatic models, which assumed that the uninformed population would grow without limit; the second concerns linking the different elements of the disinformation system as a strategy of diplomacy, as previous research has focused on analysing each of these separately, as exemplified by Buchanan and Benson (2019), Starbird (2019), Helmus et al. (2018) and Entman (2007), Hence, this model makes it possible to understand the impact of each of the elements identified in the literature by integrating them into a single system, and, in line with La Cour (2020), the proposed model provides an explanation for this problem from a macro and not a local dynamic, by involving a greater number of elements and the

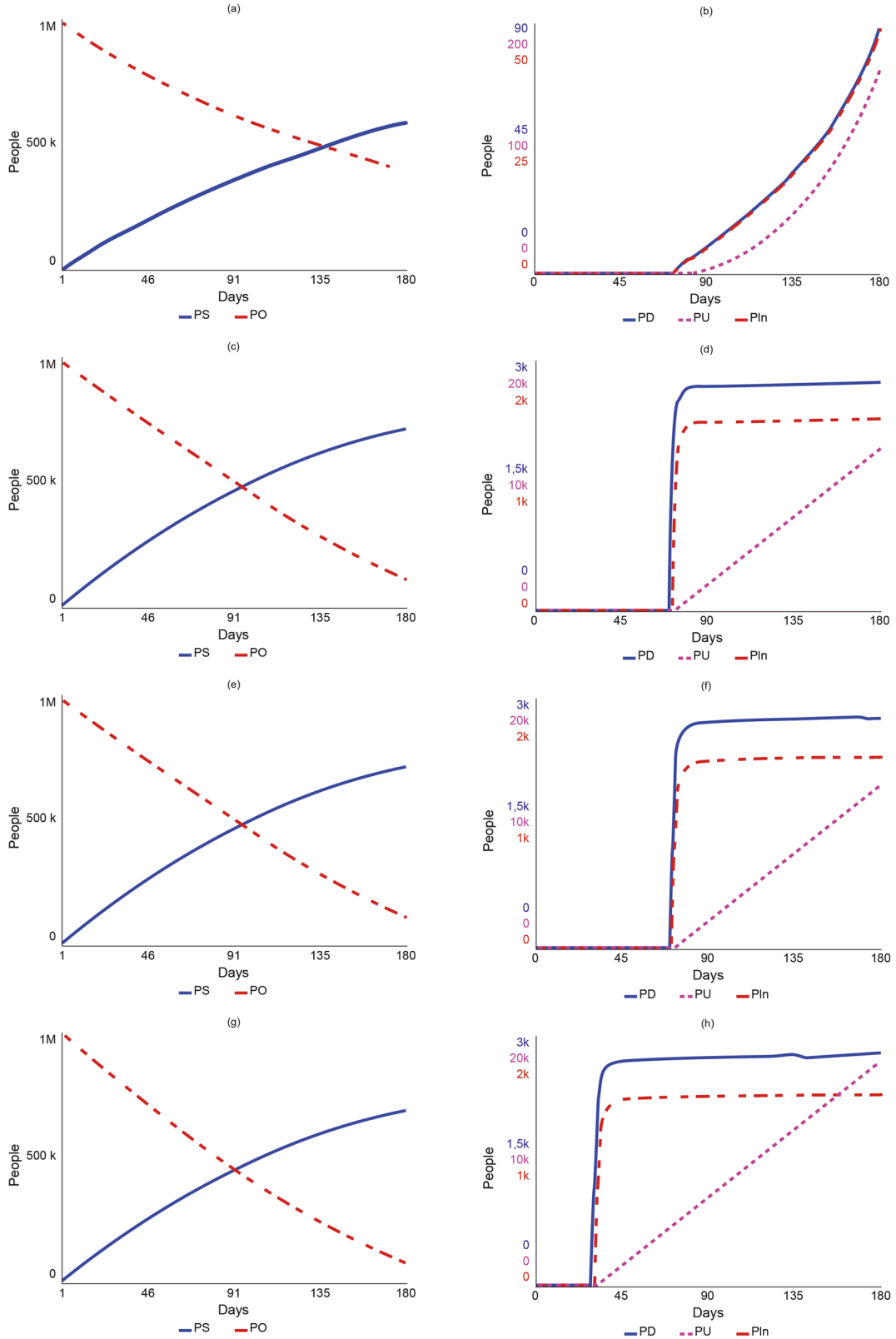

Fig. 4 Simulation results of the model with parameters set for Sim-1, Sim-2, Sim-3 and Sim-4. a, **c**, **g**, **e** System behaviour at PO and PS levels. **b**, **d**, **f**, **h** System behaviour at PD, PU and PIn levels.

possibility of executing monetary resources to intensify disinformation work.

Regarding the behaviour of the system in the case of suppressing some of the elements that compose it or modifying the established parameters such as the level of engagement, statistically significant differences were found that increase or decrease the levels of PS, PO, PD, PU and PIn, as shown in the state of the levels at $t = 180$ and in figures three and four. Thus, in the absence of paid outreach, the PS of disinformation was reduced by 38.02%, which means that paying for the linking of the

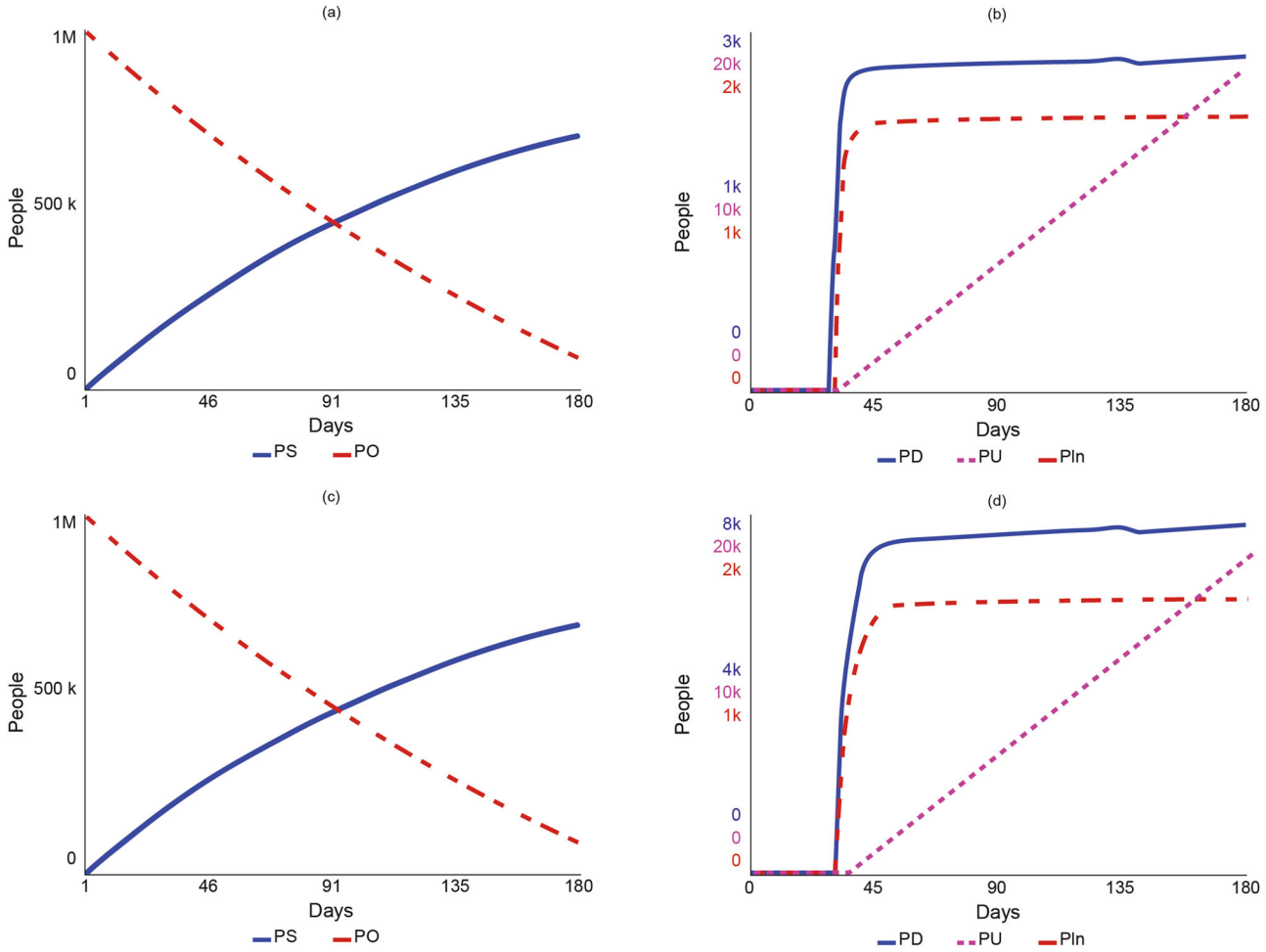

Fig. 5 Simulation results of the model with parameters set for SIM-5. a, **b** system behaviour at PO and PS levels with ne = 5%. **c**, **d** system behaviour at PO and PS levels with ne = 40%.

target population to the disinformation agent's accounts, as well as the propagation of the message on the social network, are of vital importance for the action in this strategy of international diplomacy. The absence of this element in the system changes the behaviour of the disinformation system, affecting fewer people in the target population, so the role of social networks and this mechanism to control the spread of disinformation should be evaluated. This generates a new scenario that should be incorporated into the study of the phenomenon of disinformation, especially in diplomacy, which evaluates the double standards of social networks in wanting to prevent the propagation of the disinformation message, but at the same time profit from this activity, as was shown in the case of the US elections and documented by the Office of the Director of National Intelligence (2017).

Thus, in the absence of bots and trolls, the amount of uninformed population decreases, but not to the same extent as in the absence of paid reach. This behaviour can be explained for three reasons: the first is related to the limited number of parameterised bots and trolls hired at the initial moment of the propagation of the disinformation; the second is related to the limited reach they have, as their activity is concentrated exclusively on the organic reach defined by the social network in which they disinform; and the third is related to the effectiveness of the mechanisms that these types of networks have to deactivate the bots and eliminate the troll accounts.

Regarding the onset of disinformation, the simulation showed that the early beginning of the propagation of the disinformation message has the capacity to increase the susceptible population, as well as to increase the number of people disengaging from the disinformation agent's accounts; however, the number of uninformed and informed people did not show a major change (0.06% and 0.12%, respectively) compared to the results of the initial behaviour of the system. Finally, the simulation of the level of engagement showed that its decrease generates a decrease in PS, although less interaction with the disinformation message does not generate a greater number of informed, uninformed and unsubscribed people. Furthermore, the increase in the level of citizen interaction with the disinformation message results in an increase in PS and the misinformed population.

Given the results and discussion presented here, the model developed sheds light on how disinformation spreads on social media as a result of the strategy of diplomacy, providing a novel new picture that links the highly theoretical component of the study of this phenomenon from international relations, and the documentation of cases. It is recognised that the study of disinformation remains complex, especially in diplomacy, because of the difficulty of tracing the origin of disinformation and the exact use of the elements of the system, and therefore the academic community and states are widely encouraged to use the model presented here to continue the analysis of this strategy of diplomacy.

Now, in view of the limitations of the study, it should be taken into account that the simulations only modified one parameter during their execution, so the results presented here are based on the Ceteris Paribus criterion, so that the modification of several parameters will result in a change in the behaviour of the system.

Randomisation of some of the parameters should also be considered to determine possible changes in the behaviour of the disinformation system. On the other hand, it is recommended that the academic community use various techniques to evaluate the model with different techniques associated with system dynamics, in order to provide additional evidence of its robustness. Additionally, the proposed model was based on the current elements used by diplomacy to misinform on social media, so if a new element is introduced as a result of the evolution of both the platforms and the strategy, it should be incorporated.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study. The simulation model is available at https://exchange.iseesystems.com/models/editor/alfredoguzmanrincon/modelo-de-flujos-y-niveles-de-la-desinformacion-como-estrategia-de-la-diplomacia.

Received: 4 November 2022; Accepted: 27 July 2023;
Published online: 15 August 2023

References

Agarwal NK, Alsaeedi F (2020). Understanding and fighting disinformation and fake news: Towards an information behavior framework. Proceedings of the Association for Information Science and Technology 57. https://doi.org/10.1002/pra2.327

Bala BK, Arshad FM, Kusairi MN (2017). System Dynamics. Springer Texts in Business and Economics. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-2045-2

Bayer J, Bitiukova N, Bárd P, Szakács J, Alemanno A, Uszkiewicz E (2019) Disinformation and propaganda. European Parliament

Bennett WL, Livingston S (2018) The disinformation order: disruptive communication and the decline of democratic institutions. Eur J Commun 33:122–139. https://doi.org/10.1177/0267323118760317

Bessi A, Zollo F, Del Vicario M, Scala A, Caldarelli G, Quattrociocchi W (2015) Trend of narratives in the age of misinformation. PLoS ONE 10:e0134641. https://doi.org/10.1371/journal.pone.0134641

Bian T, Xiao X, Xu T, Zhao P, Huang W, Rong Y, Huang J (2020) Rumor detection on social media with bi-directional graph convolutional networks. Proceeding AAAI conference on Artificial Intelligence. Vol. 34, p. 549–556. https://doi.org/10.1609/aaai.v34i01.5393

Bjola C (2018) The ethics of countering digital propaganda Ethics & International Affairs. Vol. 32, p. 305–315. https://doi.org/10.1017/S0892679418000436

Bodine-Baron E, Helmus TC, Magnuson M, Winkelman Z. 2016. Examining ISIS support and opposition networks on twitter. 1st edn. Santa Monica, Rand Corporation

Buchanan T, Benson V (2019) Spreading Disinformation on Facebook: Do Trust in Message Source, Risk Propensity, or Personality Affect the Organic Reach of "Fake News"? Social Media + Society 5:205630511988865. https://doi.org/10.1177/2056305119888654

Carlo Bertot J, Jaeger PT, Grimes JM (2012) Promoting transparency and accountability through ICTs, social media, and collaborative e-government Edited by Soon Ae Chun. Transforming Government: People, Process and Policy. Vol. 6, p. 78–91. https://doi.org/10.1108/17506161211214831

Chiais M (2008). Menzogna e propaganda: Armi di disinformazione di massa. 1st edn. Rome, Lupetti

Chiang C-F, Knight B (2011) Media bias and influence: evidence from newspaper endorsements. Rev Econ Stud 78:795–820. https://doi.org/10.1093/restud/rdq037

Cull NJ (2011) WikiLeaks, public diplomacy 2.0 and the state of digital public diplomacy. Place Brand Public Dipl 7:1–8. https://doi.org/10.1057/pb.2011.2

Cull NJ(2016) Engaging foreign publics in the age of Trump and Putin: Three implications of 2016 for public diplomacy. Place Brand Public Dipl 12:243–246. https://doi.org/10.1057/s41254-016-0052-4

Desantes-Guanter JM (1976). La verdad en la información. 1st edn. Madrid, Servicio de Publicaciones de la Diputación Provincial

Di L, Gu Y, Qian G, Yuan GX (2020) A Dynamic Epidemic Model for Rumor Spread in Multiplex Network with Numerical Analysis. Arvix.org

Durandin G (1993). L'information, la désinformation et la réalité. 1st edn. Presses universitaires de France

Entman RM (2007) Framing bias: media in the distribution of power. J Commun 57:163–173. https://doi.org/10.1111/j.1460-2466.2006.00336.x

Fallis D (2015) What is disinformation. Library Trends 63:401–426. https://doi.org/10.1353/lib.2015.0014

Fallis D (2011) Floridi on Disinformation. Etica and Politica/Ethics and Politics: 201–214

Faris RM, Roberts H, Etling B, Bourassa N, Zuckerman E, Benkler Y (2017) Partisanship, propaganda, and disinformation: Online media and the 2016 U.S. Presidential election. 1st edn. Washinton, Berkman Klein Center for Internet & Society Research Paper

Fjällhed A (2021) Managing disinformation through public diplomacy. In *Public Diplomacy and the Politics of Uncertainty*, ed. Pawel Surowiec and Ilan Manor. Palgrave Macmillan Series in Global Public Diplomacy. Cham, Springer International Publishing. https://doi.org/10.1007/978-3-030-54552-9

Forrester JW (2013) Industrial Dynamics. Eastford, Martino Publishing

Garrett RK (2009) Echo chambers online?: Politically motivated selective exposure among Internet news users. J Comput Mediat Commun 14:265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Gebhard C (2016) One world, many actors. In International Relations, ed. S. McGlinchey, 1st edn. Bristol, E-International Relations Publishing

Gerrits, André WM (2018) Disinformation in international relations: how important is it. Secur Hum Rights 29:3–23. https://doi.org/10.1163/18750230-02901007

Graffy C (2009) Public diplomacy: a practitioner's perspective. Am Behav Sci 52:791–796. https://doi.org/10.1177/0002764208326524

Guess AM, Nyhan B, Reifler J (2020) Exposure to untrustworthy websites in the 2016 US election. Nat Hum Behav 4(5):472–480. https://doi.org/10.1038/s41562-020-0833-x

Gunther R, Beck PA, Nisbet EC (2019) "Fake news" and the defection of 2012 Obama voters in the 2016 presidential election. Elect Stud 61:102030. https://doi.org/10.1016/j.electstud.2019.03.006

Guzmán RA, Barbosa RLC, Segovia-García N, Franco DRA (2022) Disinformation in social networks and bots: simulated scenarios of its spread from system dynamics. Systems 10:34. https://doi.org/10.3390/systems10020034

Guzmán A, Rodríguez-Cánovas B, Valencia LI, Ramírez DA (2020) Comunicación de las políticas públicas en redes sociales: caso Colombia. In Comunicación especializada: historia y realidad actual, 783–802. Madrid, McGraw-Hill Interamericana de España

Guzmán A, Rodríguez-Cánovas B (2021). Disinformation propagation in social networks as a diplomacy strategy: analysis from system dynamics. JANUS.-NET e-journal of International Relations DT. https://doi.org/10.26619/1647-7251.DT21.3

Helmus TC, Bodine-Baron E, Radin A, Magnuson M, Mendelsohn J, Marcellino W, Bega A, and Winkelman Z (2018) Russian Social Media Influence: Understanding Russian Propaganda in Eastern Europe. 1st edn. Santa Monica, Rand Corporation

Hollenbaugh EE, Ferris AL (2014) Facebook self-disclosure: examining the role of traits, social cohesion, and motives. Comput Hum Behav 30:50–58. https://doi.org/10.1016/j.chb.2013.07.055

Innes M (2020) Techniques of disinformation: Constructing and communicating "soft facts" after terrorism. Br J Sociol 71:284–299. https://doi.org/10.1111/1468-4446.12735

Jacquard R (1988). La desinformación, una manipulación del poder. 1st edn. Madrid, Espasa Calpe

Jahng MR (2021) Is fake news the new social media crisis? examining the public evaluation of crisis management for corporate organizations targeted in fake news. Int J Strateg Commun 15:18–36. https://doi.org/10.1080/1553118X.2020.1848842

La Cour C (2020) Theorising digital disinformation in international relations. Int Politics 57:704–723. https://doi.org/10.1057/s41311-020-00215-x

Lanoszka A (2019) Disinformation in international politics. Eur J Int Secur 4:227–248. https://doi.org/10.1017/eis.2019.6

Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ et al. (2018) The science of fake news. Science 359:1094–1096. https://doi.org/10.1126/science.aao2998

Li J, Cai T, Deng K, Wang X, Sellis T, Xia F (2020) Community-diversified influence maximization in social networks. Inf Syst 92:101522. https://doi.org/10.1016/j.is.2020.101522

Lupion M (2018) The gray war of our time: information warfare and the kremlin's weaponization of russian-language digital news. J Slav Mil Stud 31:329–353. https://doi.org/10.1080/13518046.2018.1487208

Ma, J, Gao W, Wei Z, Lu Y, Wong K-F (2015). Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. In Proceedings of the 24th ACM International on Conference on Information and Knowledge

Management, p. 1751–1754. Melbourne Australia, ACM. https://doi.org/10.1145/2806416.2806607
Manor I, Segev E (2020) Social media mobility: leveraging twitter networks in online diplomacy. Glob Policy 11:233–244. https://doi.org/10.1111/1758-5899.12799
McGonagle T (2017) "Fake news": false fears or real concerns? Netherlands Quarterly of Human Rights 35:203–209. https://doi.org/10.1177/0924051917738685
Mölder H, Sazonov V (2018) Information warfare as the hobbesian concept of modern times — the principles, techniques, and tools of russian information operations in the donbass. J Slav Mil Stud 31:308–328. https://doi.org/10.1080/13518046.2018.1487204
Office of the Director of National Intelligence (2017). Assessing Russian activities and intentions in recent US elections. U.S. Senate select Committee on Intelligence
Pamment J, Olofsson A, Hjorth-Jenssen R (2017) The response of Swedish and Norwegian public diplomacy and nation branding actors to the refugee crisis. J Commun Manag 21:326–341. https://doi.org/10.1108/JCOM-03-2017-0040
Quattrociocchi W, Scala A, Sunstein CR (2016). Echo Chambers on Facebook. SSRN Electron J. https://doi.org/10.2139/ssrn.2795110
Rapoport A, Rebhun LI (1952) On the mathematical theory of rumor spread. Bull Math Biophys 14:375–383. https://doi.org/10.1007/BF02477853
Rodríguez AR (2018) Fundamentos del concepto de desinformación como práctica manipuladora en la comunicación política y las relaciones internacionales. Hist Comun Soc 23:231–244. https://doi.org/10.5209/HICS.59843
Schwaninger M, Groesser, S (2020). System Dynamics Modeling: Validation for Quality Assurance. In System Dynamics, ed. Brian Dangerfield, 119–138. New York, NY, Springer USA. https://doi.org/10.1007/978-1-4939-8790-0_540
Starbird K (2019) Disinformation's spread: bots, trolls and all of us. Nature 571:449–449. https://doi.org/10.1038/d41586-019-02235-x
Sterman J (2012) Business Dynamics: Systems Thinking And Modeling For A Complex World. Massachusetts Institute of Technology, Massachusetts
van Dijk T (2006) Discurso y manipulación: discusión teórica y algunas aplicaciones. Rev. Signos 39:49–74. https://doi.org/10.4067/S0718-09342006000100003
Vériter SL, Bjola C, Koops JA (2020) Tackling COVID-19 disinformation: internal and external challenges for the European union. Hague J Dipl. 15:569–582. https://doi.org/10.1163/1871191X-BJA10046
Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359:1146–1151. https://doi.org/10.1126/science.aap9559
Wang J (2006) Managing national reputation and international relations in the global era: public diplomacy revisited. Public Relat Rev 32:91–96. https://doi.org/10.1016/j.pubrev.2005.12.001
Wang J, Zhao L, Huang R (2014) SIRaRu rumor spreading model in complex networks. Phys A: Stat Mech Appl 398:43–55. https://doi.org/10.1016/j.physa.2013.12.004
Zhao X, Wang J (2013) Dynamical model about rumor spreading with medium. Discrete Dyn Nat Soc 2013:1–9. https://doi.org/10.1155/2013/586867

Author contributions

A.G.R., R.L.C.B., B.R.C., D.R.A.F. and S.B.M. contributed to conception and design of the study. A.G.R. organised the database. A.G.R. and R.L.C.B. performed quantitative and qualitative analysis. A.G.R., B.R.C. and S.B.M. wrote the first draft of the manuscript. A.G.R., S.B. and D.R.A.F. reviewed and edited. S.B.M. and B.R.C. supervised both the development of the research and the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Alfredo Guzmán Rincón.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023