Time-critical decentralised situational awareness in emergencies: an adversarial biosecurity scenario

Jason Bassett[*] , Niccolò Pescetelli, Alex Rutherford and Manuel Cebrian

*Correspondence:
bassett@mpib-berlin.mpg.de
Center for Humans
and Machines, Max Planck
Institute for Human
Development, Berlin,
Germany

Abstract

Crises in a global setting of interdependencies call for time-critical coordinated responses. However, it is often the case that the mechanisms responsible for these actions do not agree across all their hierarchies. This can be roughly attributed to personal estimations of the situation and to social influence. An ensuing lack of consensus against crises can be dire and echo across entire populations. One such instance is the case of biosecurity threats. A particularly interesting class of threats lie within urban environments, which tend to fall within the scope of bad actors. With this work we aim to computationally contribute to the understanding of the dynamics of perceived danger formation among agents responsible for responding to ongoing biological attacks in urban settings. We assume this perception is a function of a personal estimation of local information about the danger and of social influence stemming from the agents in question framed in an agent-based model. The simulations point towards a high dependence of perceived dangers on the personal estimations of the agents. The conditions under which the perceived dangers deviate from the real ones are explored over a range of assumptions on personal measurements and several dispositions towards the influencing environment. The insight provided by these results at the individual and collective level set the tone for further investigation on such behavioural phenomena, providing a flexible computational framework addressing generic threats (true dangers) in a time-critical context.

Keywords: Agent-based modelling, Danger perception, Estimation bias, Social influence, Urban environments

Introduction

The main goal in a crisis in undoubtedly its effective and timely resolution. However, in the modern interconnected and complex world it is often the case that the parties involved for a crisis' resolution have significantly different perceptions of the situation at hand. This can be for a number of reasons, pertinent to the nature of the crisis as well as to the responsive mechanisms in place to face it. In addition, this intricate systemic complexity renders large-scale, unprecedented threats increasingly likely (Taleb 2007).

Biosecurity threats consist of a class of crises of massive impact where expert knowledge does not always coincide with the situational awareness of the officials in the first

lines of response. With past instances of dispatched weaponised agents against a naive population (Wein et al. 2003; Alibek 2008; Kaszeta 2020) up to recent examples of malevolent cybersecurity infiltration on essential infrastructure (The New York Times 2021), the imminence of biosecurity threats is increasingly becoming a tangible liability. Furthermore, statistical analysis on past adversarial events classified as terrorist attacks testifies to a likely periodical time horizon in which they can be anticipated (Clauset et al. 2013). More broadly, along with cybersecurity threats excessively large-scale ones being specified as increasingly recurring (Flyvbjerg 2020), it is imperative to have emergency and effective response mechanisms readily deployable.

In this context, momentum was gained after the US 9/11 attacks on biosecurity attacks with the seminal work of Wein, Craft and Kaplan in 2003 on a hypothetical Anthrax airbourne point delivery in an urban environment. This modelled various scenarios on the impact of the attack and the corresponding response mechanisms (Wein et al. 2003). Further, biosecurity topics have occupied instances of the vast epidemics literature inasmuch social behavioural implications are involved in mitigating adversarial biological attacks in the general population (Del Valle et al. 2005). Notwithstanding, the attention on the matter culminated with the multiagent model of Carley et al. published in 2006 and modelling a wide range of social and economic factors, which contribute to the spread and mitigation of maliciously introduced diseases in urban centres. The novelty of the work lay especially in the network and agent behavioural patterns availing themselves to the formulation of tracking and surveillance protocols by policy makers.

Within the time-criticality context, previous work has established methods and limitations of mobilisation mechanisms to unequivocally detect targets (Rutherford et al. 2013a, b). This would be the first step to activate a network of response mechanisms in a coordinated fashion against a globally distributed biosecurity threat, i.e. clarifying the mechanisms' situational awareness. Nevertheless, the mechanisms (essentially administrators), being human and systemic, are not infallible, are characterised by bounded rationality and often exhibit biases and delays in the realisation of a crisis, ranging from the individual to the collective level (Simon 1956; Tversky and Kahneman 1971, 1974; Pescetelli et al. 2016; Rosenberg et al. 2016).

Aiming to tackle this situational awareness dilemma encumbering policy makers, with this work we aspire to provide a computational framework modelling the range of the corresponding perceived dangers on behalf of policy making organisations and their executive branches, as dynamical individual and collective cognitive processes. We assume that the organisations in question are spatially distributed in different urban settings but each one forms their own social network, not necessarily belonging in the same environment [akin to a multilayer representation (Kivelä et al. 2014; Strano et al. 2015)]. The behavioural insight of the work constitutes a first step towards preparing in a coordinated fashion for biological attacks but also more broadly for attacks or threats of an unexpected nature (Madnick 2020). Lastly, it is noteworthy that the cognitive problem we are tackling has recently drawn increased attention from the epidemics community (Granell et al. 2013, 2014; Agaba et al. 2017), signifying its importance across disciplines.
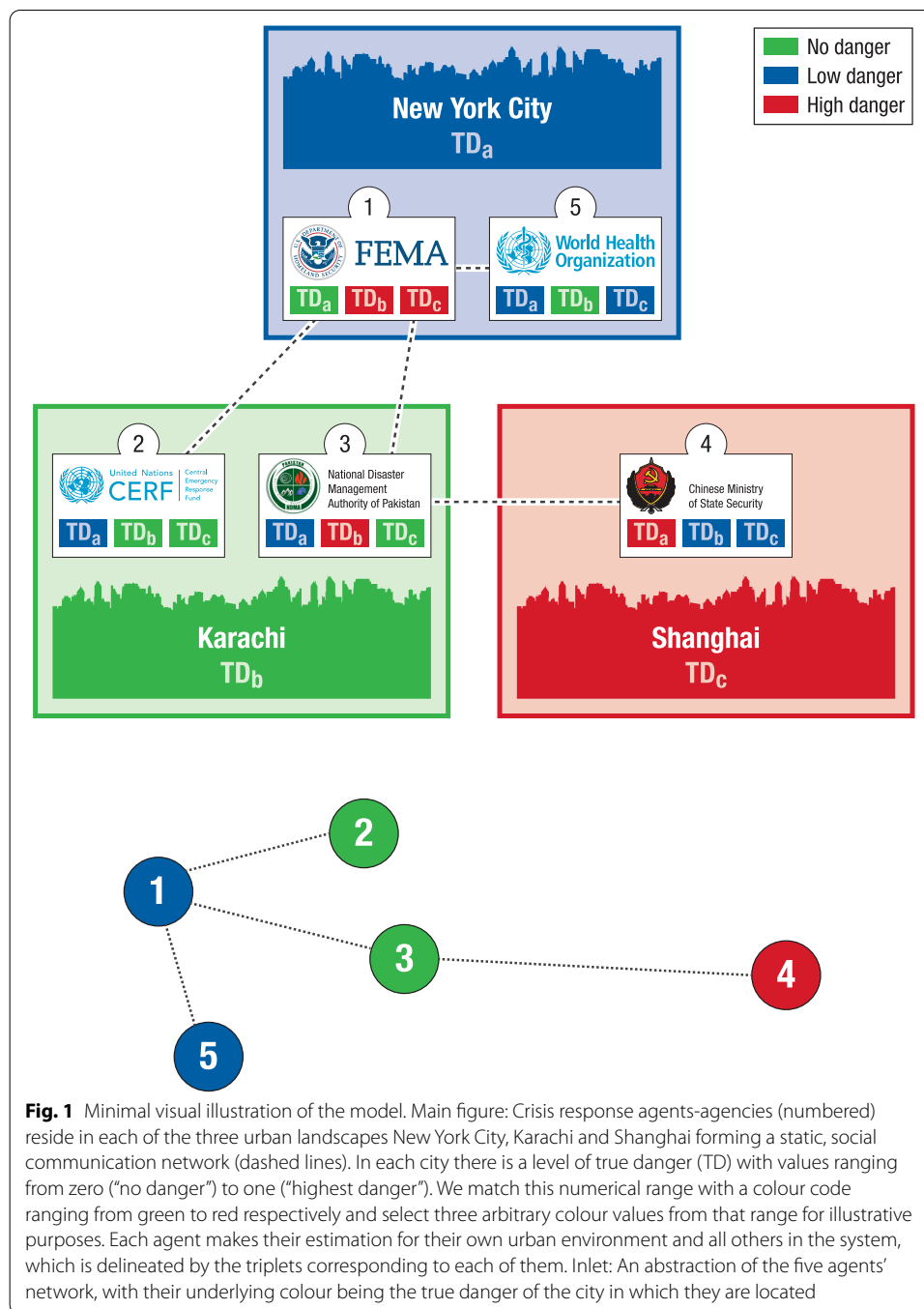
Methods

For the purpose of our study we formulated an agent-based model of discrete time distributed over 18 cells representing urban metropolitan areas, from here onwards to be referred to as cities. The constraint was to choose urban agglomerations with a metropolitan population of at least 10^7 (Citypopulation 2021) but otherwise the selection was such as to introduce a small global urban ecosystem for illustrative purposes and accounts for the flexibility of the model (the list of cities can be found in Table 1 of the Appendix). The agents themselves represent policy-making organisations and affiliated branches with a vital role in the response to biosecurity threats. We assumed that the number of agents within each city was proportional to the population of the city and that the number of the total agents was 1000 (see Table 1 in the Appendix for an overview of the scaling factor's effect on the population). For this latter figure, to the extent of our knowledge, there does not exist any specific single dataset, tailored to explicitly substantiate it and as such remains an educated guess. As we will see in the following however, these agents-organisations form an information exchange network and can therefore be regarded under a transnational scope. In this sense and drawing from Vitali and Battiston's work on the global corporate network, we can expect to be at least in the right order of magnitude, assuming a 1:43 ratio of organisations encumbered with emergency responses to biological attacks against the 43,060 transnational corporations in the dataset of the aforementioned authors (Vitali and Battiston 2014).

The agents do not act in isolation, rather form a network of information exchange on situational awareness of biologically related dangers for each and every one of the network's cities. Therefore, each agent (node of the network) has a situational awareness for 18 locations-cities. We do not take any agent mobility into account, ergo, the network is static. Simultaneously the agents make personal measurements within their own city (i.e. with the information locally available), which they then jointly assess with the available (social) information from their network. In this way a dynamical process takes place with which the agents make a final personal judgment for the current degree of danger in their own locale and in every other city. A minimal, invented example of three cities (New York City, Karachi and Shanghai) with their agent-agencies' social information network as well as a snapshot of the underlying processes is illustrated in Fig. 1, while in the inlet between the main figure and the colour code scale lies an abstraction of the upper.

Agent network

For the connectivity of the network, which determines the strength of influence in the aforementioned dynamical process, we examined three of the prototype networks (empirical networks are beyond the scope of this work), namely the Erdős–Rényi (ER) (Erdős and Rényi 1960; Newman 2018), the small-world introduced by Watts and Strogatz (WS) (Newman 2018; Watts and Strogatz 1988) and the preferential attachment one generated by the Barabási–Albert (BA) model (Newman 2018; Barabási and Albert 1999). A constraint that we set for all three of the networks was that they have the same

Fig. 1 Minimal visual illustration of the model. Main figure: Crisis response agents-agencies (numbered) reside in each of the three urban landscapes New York City, Karachi and Shanghai forming a static, social communication network (dashed lines). In each city there is a level of true danger (TD) with values ranging from zero ("no danger") to one ("highest danger"). We match this numerical range with a colour code ranging from green to red respectively and select three arbitrary colour values from that range for illustrative purposes. Each agent makes their estimation for their own urban environment and all others in the system, which is delineated by the triplets corresponding to each of them. Inlet: An abstraction of the five agents' network, with their underlying colour being the true danger of the city in which they are located

average edge density $\langle \rho \rangle$ over 100 realisations. Consequently and with 1000 nodes in each network this constraint narrowed the range of possibilities for different topologies to a single set of control parameters for each network,[1] so that $\langle \rho \rangle \approx 1000$. We summarise the values for each case in the following:

---

[1] For N nodes the BA graph cannot have less than $N - 1$ edges and likewise the WS.

- ER connection probability: $p_{\text{ER}} = 2 \times 10^{-3}$.
- WS: nearest-neighbours $k_{\text{WS}} = 2$ and rewiring probability $p_{\text{WS}} = 0.5$.
- BA: edges to attach from a new node to existing ones: $k_{\text{BA}} = 999$ or $k_{\text{BA}} = 1$ (single hub-topology and cliques respectively).

Note that for $k_{\text{BA}} = 1$ we retrieve the same density as $k_{\text{BA}} = 999$. The difference lies in the topology. In the first case all nodes are connected to a single one, while in the latter several nodes share the property of being a hub, distributing (unevenly in general) the amount of attached nodes among each other.

Danger perceived by agents

Based on previous work demonstrating the distinction between personal and social components forming perceptions (Pescetelli et al.2016), we define two mechanisms dynamically determining an agent's perception of a given danger. In particular, these mechanisms are dependent either on the information available to the agent at hand or to the information conveyed through their network of acquaintances. Hence, we set the dynamical progression for each agent to be

$$\mathbf{PD}_{i,t+1} = \text{Mean}\big(\mathbf{f}\big(\text{TD}_{x,i,t+1}\big), \mathbf{g}\big(\mathbf{PD}_{i,t}, \mathbf{PD}_{j \neq i,t}\big)\big) \tag{1}$$

with $\text{PD}_{i,t+1}$ denoting the perceived danger of an agent i at time $t + 1$, $\mathbf{f}$ a vector function mapping a given true danger $\text{TD}_{x,i,t+1}$ of city x where agent i resides at time t to a personal estimation (e.g. by measurement) for the true danger by the agent, and $\mathbf{g}$ a vector function mapping the perceived danger of agent i and that of every other agent $j \neq i$ to an overall influence for the danger value exerted on agent i at time t. The boldface font marks a vector with components being the perceived danger values of the agent in question for every city. Thus for 18 cities in the system, every agent i has 18 perceptions of danger and consequently there are 18 corresponding values of the true danger for every time step t. Both PD as well as TD take values in [0, 1], hence assuming two extremes of a danger degree, where 0 stands for complete lack of danger and 1 for total danger (the maximum degree). A special note needs to be made for the time indices of the two terms in (1). The personal estimation term is happening at the time of updating the perceived danger (thus introducing time-criticality to the personal assessment task via the latest local information available), while the influence is coming from the exact previous time step due to a minimal time lag in communications and cognitive processing. Lastly, the "Mean" in (1) refers to the arithmetic average (coinciding with the median since we are referring to two arithmetic quantities). However, its purpose is to introduce a measure of coestimation of the personal estimation and influence quantities and can therefore assume in principle any other plausible form, introducing, for instance, bias. We will consider only an unbiased estimation here.

Personal estimations

In the absence of data for the scope of our study, we simulate the personal measurement function *f* with draws from a number of plausible representational probability

distributions, while the influence function g from established influence models. Firstly and for the personal estimation we set f to draw a random variable X from the triangular distribution, the Gaussian, the Laplacian and a power-law for every time step, parametrised by $TD_{x,i,t}$ (the locally available information) as follows:

1. Triangular: $Tr(X; a, \text{mode}, b) = Tr(X; 0, TD_{x,i,t}, 1)$. As such the agent makes a personal estimation of the true danger for every city x with information about the true danger in their city of residence serving as the mode of the distribution.

2. Gaussian: $|N(X; \mu, \sigma)| = |N(X; TD_{x,i,t}, 1)|$. In this case, the agent's personal estimation for every city x is drawn from an absolute valued Gaussian, centered on the true danger of the agent's city of residence and with a scale parameter being the maximum possible fluctuation for the danger estimation they can have. If values are drawn beyond 1, then we assume that the agent made a far-fetched personal estimation of the danger and therefore uniformly, randomly assign to this estimation the extreme value of either 0 or 1.

3. Laplacian: $|L(X; \mu, b)| = |L(X; TD_{x,i,t}, 1)|$. The personal estimation in this instance is similar to the case of the Gaussian estimation in terms of the location and scale parameter of the distribution, as well as for the assigned values in case of draws beyond 1. The sole difference lies in the form of the distribution (pronounced center and fat tails).

4. Power-law: $P(X; a) = P(X; TD_{x,i,t})$. For this case, the agent's personal estimate follows a power-law distribution of the form $\alpha X^{\alpha-1}$ with $\alpha = TD_{x,i,t} \in (0, 1]$ in the role of the exponent. In case of a zero value of the true danger, then the personal estimate is also set to zero.

The rationale for the selection of the previously mentioned distributions is to examine a broad range of generative personal estimations of the danger degree via a single measurement (in the context of time-criticality), in a reference time and given only local information for an arbitrary population. The selection is by no means exclusive and its ultimate justification could come from some data proxy (e.g. by means of lexicographical analyses on social media platforms (Zhang et al. 2020; Aiello et al. 2020)].

Social influence

It is by now well understood that in a social environment, a personal view on a topic is not just based on personal estimations (e.g. via measurements) rather is weighed against the range of other views held on the same topic by the corresponding social network (Sherif et al. 1965). The phenomenon is called *social judgment* and is a major driver in the established literature of influence models (Flache et al. 2017).

In this context and in regard to the information exchange mechanism among the agents about the true danger in the various cities, we utilise three models from the social dynamics literature to represent the coupling in the danger estimation among agents in different cities. In particular, the one due to Friedkin and Johnson (F–J) accounting for assimilative social influence (consensus) (Flache et al. 2017; Friedkin and Johnsen1990),

the one due to Deffuant and collaborators accounting for embracement of opinions with a selective bias (the emergence of fractions) (Flache et al. 2017; Deffuant et al. 2000), and the one due to Jager and Amblard (J–A) accounting for a repelling effect to persuasive views beyond a certain threshold (polarisation) (Flache et al. 2017; Jager and Amblard 2005). With an adjusted form of these models we formulate the *g* function in the dynamic perceived danger relation (1). The exact form of each model as utilised is as follows:

1. F–J:

$$\mathbf{g}_{i,t}\big(\mathbf{PD}_{i,t}, \mathbf{PD}_{j\neq i,t}\big) = \mathbf{PD}_{i,t} + \mu \mathbf{D}_{1/2} \sum_{j\neq i} \mathbf{w}_{i,j}\big(\mathbf{PD}_{j,t} - \mathbf{PD}_{i,t}\big), \tag{2}$$

where $0 < \mu \leq 1$ is the rate of opinion convergence (we chose $\mu = 0.5$), $\mathbf{D}_{1/2}$ the distance matrix with inverse entries for all non-zero elements and $\mathbf{w}_{i,j}$ is a weight of opinion importance ascribed to an agent's network. We assume this weight to be subject to the constraint $\sum_j \mathbf{w}_{i,j} = 1, \forall i$. We generate the weights from a uniform random distribution in [0, 1) and normalise them. Note that the boldface (vector notation) applies to the weights as well, since each agent's perceived danger involves all the 18 cities in the simulation.

2. Deffuant:

$$\mathbf{g}_{i,t}\big(\mathbf{PD}_{i,t}, \mathbf{PD}_{j\neq i,t}\big) = \mathbf{PD}_{i,t} + \mathbf{D}_{1/2} \sum_{j\neq i} \mathbf{h}_w\big(\mathbf{PD}_{i,t}, \mathbf{PD}_{j,t}\big)\big(\mathbf{PD}_{j,t} - \mathbf{PD}_{i,t}\big) \tag{3}$$

with the weight function being

$$\mathbf{h}_w = \begin{cases} \mu, & \text{if } \big|\mathbf{PD}_{i,t} - \mathbf{PD}_{j,t}\big| \leq \epsilon \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$
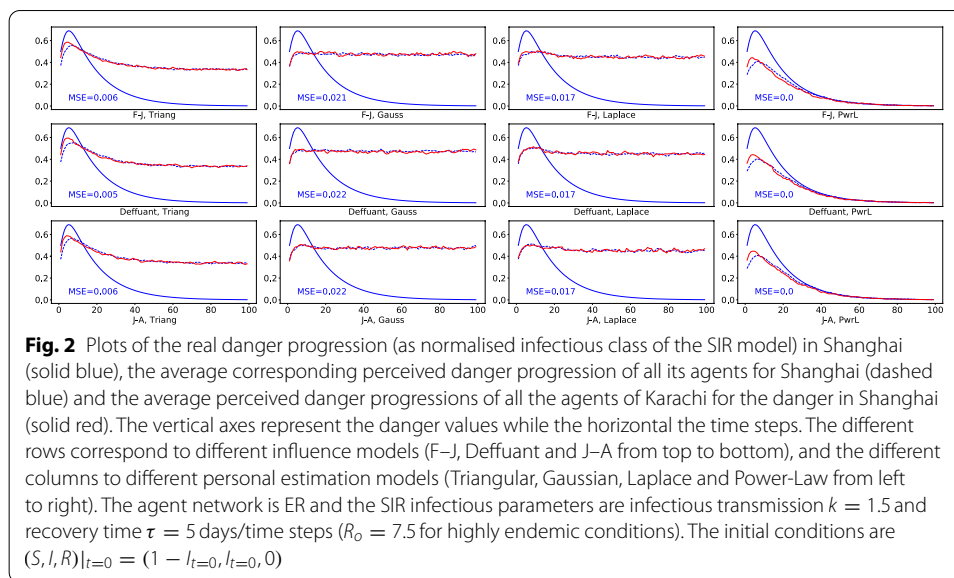
The quantity ϵ (we choose $\epsilon = 0.2$) is simply the tolerance up to which a danger estimation other than that of the agent in question (*i*) will influence them (in this case $0 < \mu \leq 0.5$ and we select $\mu = 0.5$ again). Beyond that tolerance they simply ignores different estimates.

3. J–A: This is as the Deffuant model but the weight function assumes the form

$$\mathbf{h}_w = \begin{cases} \mu, & \text{if } \big|\mathbf{PD}_{i,t} - \mathbf{PD}_{j,t}\big| \leq \epsilon \\ \mu\big(1 - 2\big|\mathbf{PD}_{i,t} - \mathbf{PD}_{j,t}\big|\big), & \text{otherwise.} \end{cases} \tag{5}$$

Thus, the agent can be persuaded of the danger estimation of a peer up to a certain tolerance. Beyond that however, instead of ignoring the estimation of the other agent as in the Deffuant model the agent is thrown to the other side of the spectrum in respect to that estimation. For example, a moderately high estimation of danger by a peer will be in this case interpreted by the (influenced) agent as a moderately low danger.
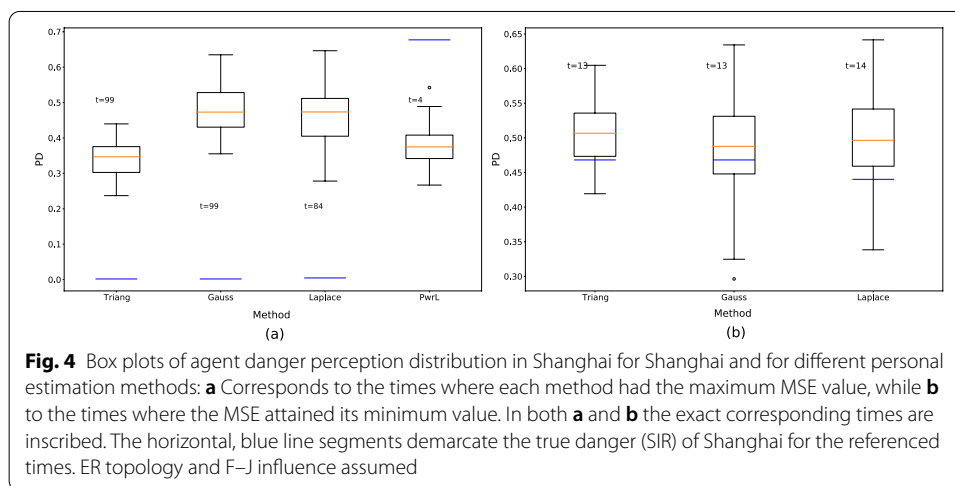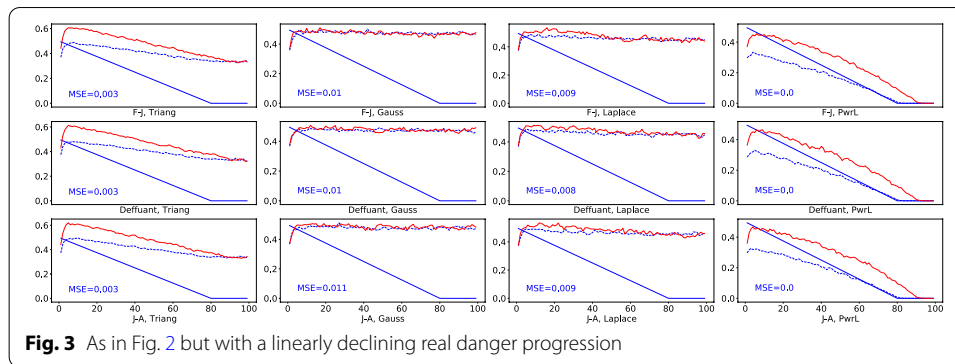
Fig. 2 Plots of the real danger progression (as normalised infectious class of the SIR model) in Shanghai (solid blue), the average corresponding perceived danger progression of all its agents for Shanghai (dashed blue) and the average perceived danger progressions of all the agents of Karachi for the danger in Shanghai (solid red). The vertical axes represent the danger values while the horizontal the time steps. The different rows correspond to different influence models (F–J, Deffuant and J–A from top to bottom), and the different columns to different personal estimation models (Triangular, Gaussian, Laplace and Power-Law from left to right). The agent network is ER and the SIR infectious parameters are infectious transmission $k = 1.5$ and recovery time $\tau = 5$ days/time steps ($R_o = 7.5$ for highly endemic conditions). The initial conditions are $(S, I, R)|_{t=0} = (1 - I_{t=0}, I_{t=0}, 0)$

True danger progression

Finally, for the true danger degree value's evolution with time in a particular city we can use any arbitrary function to represent it. Indeed, following the example of Wein et al. on a point air-delivery of Anthrax (Wein et al. 2003) we have built in our model various functions with which the impact of danger can progress. This need not necessarily be decaying. Our purpose here is to unveil the interplay between the perceived danger, the social influence and the actual magnitude of a given danger so there is no reason to use anything overly elaborate or to exhaust all the possible underlying true danger progression possibilities in a sensitivity analysis. Therefore, we will assume two cases: firstly, one typical of an epidemic spread (SIR) of highly endemic conditions (Kermack and McKendrick 1927; Keeling and Rohani 2011) and secondly, a linearly decaying danger. Specifically, the second assumes that the true danger within a city decays (e.g. because the authorities intervened for its mitigation and the projected deaths or hospitalisations decrease) in a linear fashion with time down to a zero value at a specified point. The first merely identifies the danger with the normalised infectious class of the SIR model. These two approaches do not forbid the usage of any other function with which the true danger evolves, including periodic or increasing.

Results

The implementation of the simulation was executed in Python 3.8, in which the packages NetworkX 2.4 and NumPy 1.18.5 were employed. The simulation code as well as that to generate the graphs presented in this work can be found at https://github.com/Yperidis/ThreatPerception.

The initial conditions for relation (1), i.e. the initial perceived dangers for each agent, were determined by a draw from a triangular distribution with extremities 0 and 1 and mode the initial danger of the agent's city of residence.
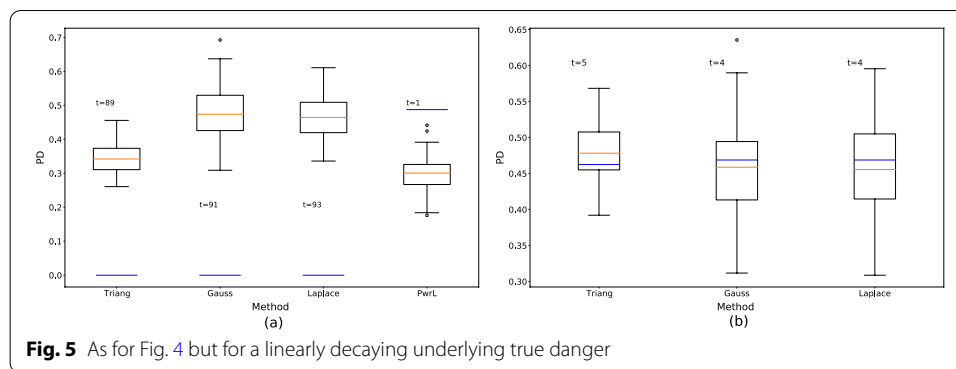
Fig. 3 As in Fig. 2 but with a linearly declining real danger progression

Fig. 4 Box plots of agent danger perception distribution in Shanghai for Shanghai and for different personal estimation methods: **a** Corresponds to the times where each method had the maximum MSE value, while **b** to the times where the MSE attained its minimum value. In both **a** and **b** the exact corresponding times are inscribed. The horizontal, blue line segments demarcate the true danger (SIR) of Shanghai for the referenced times. ER topology and F–J influence assumed

Citywide metrics

Although the simulation can probe at the individual level of an agent, for the evolution of danger and the associated processes we aggregate the agents over cities. In this way we can have an overview of all the danger processes at the scale of urban environments, while concurrently averaging the danger perceptions of individual outliers. Moreover, although the model enables us to predict the perceived danger for every city included in the simulation, for illustrative reasons we will focus on the collective perceptions of two arbitrarily chosen cities, Shanghai and Karachi, for one of the two (Shanghai). The initial danger in each is set to be $T_{\mathrm{TD,in,Sh}} = 0.5$ and $T_{\mathrm{TD,in,Ka}} = 0.9$ and the time at which the danger vanishes $T_{\mathrm{TD,null,Sh}} = 80$ and $T_{\mathrm{TD,null,Ka}} = 90$ for a total runtime of 100 time steps.

In Figs. 2 and 3 we present two cases of evolution of the true danger for Shanghai, one for the danger representing the normalised population of the infectious class of an SIR model and one for linear decay of the danger respectively. In the same graphs we overlay the aggregate danger perceptions of Shanghai's agents and those of Karachi for each of the two instances of the true danger progressions for the city of Shanghai. The results

Fig. 5 As for Fig. 4 but for a linearly decaying underlying true danger

examine each case of personal estimation (columns) and influence model (rows) as outlined in the section of methods. Each time-series was averaged over ten runs to lessen stochastic effects. We have also inscribed in every plot the mean squared error (MSE) of each time-series between the actual danger in Shanghai and the average perceived danger of its agents, i.e. that of the solid and dashed blue time-series lines.

The most striking result seems to be the minimal effect of the influence models on the perceived danger and hence of the network architecture (see Appendix for the formation of the perceived danger). Secondly and according to the MSE values, there is a ranking among the different personal estimation methods, with the power-law outperforming all of them. Second is the triangular distribution method, third the Laplace and the worst performance is exhibited by the Gaussian.

Individual agent metrics

We so far explored the collective perceived danger at the city level under a certain range of personal estimations and influence models. We would now wish to refer to the accuracy of individual agents within a single city, leading to their collective estimation of danger in it. One way of representing this accuracy is by distributing the perceived dangers of the agents around the true value for selected times, so as to estimate the maximum and minimum perceived danger deviations from the real value. The box plots of Figs. 4 and 5 do exactly this for the city of Shanghai, sticking to our example. We choose the F–J influence model and ER topology for the demonstration but the result is robust across any constant influence model and topology.

As the agents' perceptions are dynamical processes we can only study their dispersion around a mean and the true value at a snapshot. Two meaningful points in time to do this are when the MSE of the agents aggregated over Shanghai assume their maximum and minimum values. This is what is depicted in panels (a) and (b) of Figs. 4 and 5 respectively. Incidentally, one can notice the same ranking among the methods, regardless of the corresponding MSE value, as observed in the time-series of Figs. 2 and 3.

Discussion

We have presented the results of a purely computational spatially distributed agent-based model emulating cognitive formation processes in estimating a threat parameter of varying degree under time pressure (uncertainty), limited information and social influence. The context of the threat was of a biological nature and assuming its release in urban environments. The utility of the model lies in its capacity to provide behavioural insight for a broad range of intertwined individual estimation biases of agents based on local information and their drifts due to peer pressure from their underlying social network.

Although only alluded in the presentation so far, the model can be regarded as a multilayer network with its different layers representing the agent and the city networks, as well as the corresponding processes taking place on them (perceived and true dangers respectively). This alternative way of representation offers a summary overview of aggregate measures in an integrated framework and has a rich presence in the literature (Kivelä et al. 2014; Strano et al. 2015; De Domenico et al. 2013, 2015, 2016). It would be especially useful if the true danger processes of one city would be coupled to those of other cities.

The model appears to be robust to the influence among its agents, meaning that across the different opinion dynamics' models and studied topologies the results remain largely unaltered for the presented accuracy. This is not an obvious finding and the explanation lies in a better understanding of the mechanism dictated by the governing equation (1). As witnessed in Figs. 2, 3, 5 and 4 the personal estimation of the agents appears to be the pivotal force driving the perceived danger, which is the first term of the dynamic relation (1). The second term, which is the social influence, serves as an interference on the measurement taking place at the time of the danger's perception. Thus, with a lag of one time step, the agents average their current personal estimation (a measurement for example) with the collective perceived danger of their network. This mechanism determines their final danger perception (see also the Appendix on the perceived danger perception). However, we expect that the initial, spatial agent opinion distribution, combined with non-standard [e.g. core-periphery (Rombach et al. 2014)] or empirical network architectures connecting agents of diverse, non-cancelling opinions at the city level and between different cities (when looking at aggregate measures) would reveal a more significant effect of the influence term.

Reducing the system size to a total of 50 agents with the given heterogeneities (see the city populations in Table 1 in the Appendix) did not introduce any significant effects on the influence term across different topologies and models. Hence the role of the influence is to provide a regression to the mean of the agents' perceived danger on the previous time step. Basically this implicates the effect of a lag in forming the perception of a danger on a given time. However, as the literature indicates (Flache et al. 2017), there are many ways to tune the opinion thresholds ϵ on the selective bias (Deffuant) and repelling (J–A) models, as well as the rates of convergence μ across all models so as to amplify the effect of the social interference to the personal estimation. The limits of these effects would be an interesting addition to the results so far, belonging to a sensitivity analysis.

An additional note on the time horizon selection of one time step for the effect of the neighbouring agents (i.e. the influence term of Eq. (1)) is that we chose it so as to reflect an emergency. We envision an emergency being a chaotic situation of modern digital, rapid communications, where the data changes (smoothly) from one time stamp of communication to the next, rendering previous information obsolete as a limiting case. Nevertheless, a more realistic approach would consider at least several instances (steps) of previous information as well. Towards such a direction recent graph message-passing approaches (Kapoor et al. 2020; Panagopoulos et al. 2020) could be a promising way to extend the model exhibited with Eq. (1). The data utilised to prepare (train) such a scheme could stem from inter-city [e.g. freight ship trade (Kaluza et al. 2010)] or individual [e.g. post or email traffic (Hristova et al. 2016)] transactional flows, which would serve as a proxy for the strength of influence at the city and agent level respectively. This kind of data could also form the basis for an empirical network structure among cities and agents in a future study.

Interestingly and in regard to aggregating agent perceptions over cities, these are not necessarily contained within the boundaries of a single urban environment. Affiliated organisations or different branches of the same organisation, represented as agents in the simulation, might not form a consensus cluster within their common spatial location. Important as this observation might be to identify these clusters in a data-driven analysis, it is however out of scope within the limits of our study, as we have assumed indistinguishable agents.

To end with, the model's flexibility in personal and social estimation of a danger renders it suitable for mixture of the models investigated in either component, thus attempting to capture more pragmatic behavioural features of the agents. Indeed such an approach could introduce higher levels of realism but should be of course based on evidence and thus justified by indicators stemming from data (see again Kaluza et al. 2010, Hristova et al. 2016). Furthermore, since the personal estimations act as the principal driver of the model's dynamics, a data-driven approach focused on coordinating reports of measurement rather than opinions would be of more significance as the model has been framed. Under the scope of coordination does the full magnitude of situational awareness manifest itself as an invaluable ingredient for mitigating biosecurity threats within the time-critical context that they demand. Needless to say, the model can be fitted to any threat as long as the elements of time-criticality and coordinated response involving personal estimations and social influences are present.
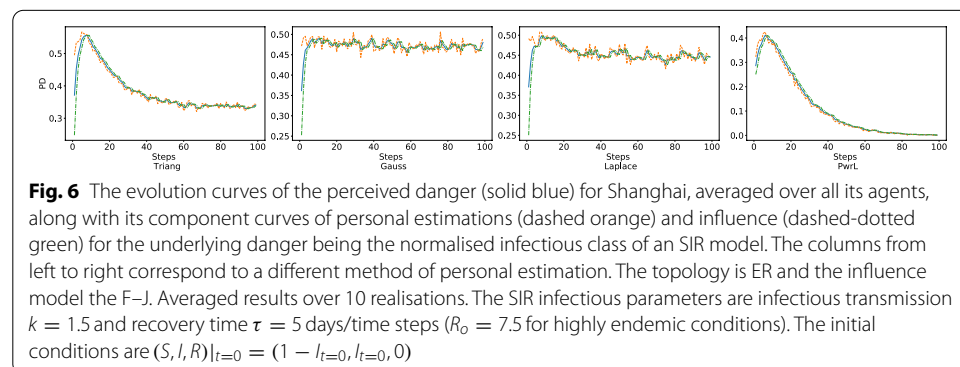
Appendix

City populations and scaling

The scaling factor to bring the total urban populations in question down to 1000 agents was 2.631×10^{-6} (Table 1).

Table 1 The list of cities, their population and their corresponding agents in ascending order, as set in the simulation

City	Population ($\times 10^6$)	Agents
Paris	11.4	30
Johannesburg	13.9	37
London	14.8	39
Tehran	15.3	40
Istanbul	16.0	42
Buenos Aires	16.4	43
Moscow	17.3	46
Karachi	17.8	47
Lagos	19.4	51
Dhaka	20.2	53
Cairo	21.0	55
New York City	22.1	58
São Paulo	22.4	59
Mexico City	23.0	61
Seoul	24.8	65
Delhi	30.3	80
Shanghai	33.6	88
Tokyo	40.4	106
Total	380.1	1000

Fig. 6 The evolution curves of the perceived danger (solid blue) for Shanghai, averaged over all its agents, along with its component curves of personal estimations (dashed orange) and influence (dashed-dotted green) for the underlying danger being the normalised infectious class of an SIR model. The columns from left to right correspond to a different method of personal estimation. The topology is ER and the influence model the F–J. Averaged results over 10 realisations. The SIR infectious parameters are infectious transmission $k = 1.5$ and recovery time $\tau = 5$ days/time steps ($R_o = 7.5$ for highly endemic conditions). The initial conditions are $(S, I, R)|_{t=0} = (1 - I_{t=0}, I_{t=0}, 0)$

Formation of perceived danger

As alluded to in the main text, the influence is posited to average the exaggerated personal estimation of the running time step in regard to the influence of the previous one. Due to the initialisation of influence and the dominance of the personal estimation term as per relation (1), the effect of influence (and thus the effect of the network) on the perceived danger formation of any agent is significantly different than the personal estimation only on the first time step. This effect is exhibited in Fig. 6 for an arbitrarily chosen topology (ER) and influence model (F–J), for illustrative purposes, for the city of Shanghai (case of linearly decaying underling real danger). The perceived danger (blue solid curve) results as an average of the personal estimations of all the agents of Shanghai for Shanghai (orange dashed curve) and of the average influence exerted on them by each of their social networks (green dashed-dotted curve) for all the combinations of different

personal estimation methods. Equivalently said, the three curves represent the perceived danger as an average of the personal estimation and the influence on a "super agent" representing all the agents in the city of Shanghai. Given that the columns of the panels in Fig. 6 reflect a different personal estimation method, the trends draw similarities to those across the columns of the panels of Figs. 2 and 3, while providing an explanation for their particular form.

Declarations

Ethics approval and consent to participate
Not applicable.

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

References
Agaba G, Kyrychko YN, Blyuss KB (2017) Mathematical model for the impact of awareness on the dynamics of infectious diseases. Math Biosci 286:22–30

Aiello LM, Quercia D, Zhou K, Constantinides M, Šćepanović S, Joglekar S (2020) How epidemic psychology works on social media: evolution of responses to the covid-19 pandemic. arXiv:2007.13169

Alibek K (2008) Biohazard. Random House, New York City

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Citypopulation.de: Major Agglomerations of the World (2021). https://www.citypopulation.de/en/world/agglomerations/. Accessed 24 Feb 2021

Clauset A, Woodard R et al (2013) Estimating the historical and future probabilities of large terrorist events. Ann Appl Stat 7(4):1838–1865

Deffuant G, Neau D, Amblard F, Weisbuch G (2000) Mixing beliefs among interacting agents. Adv Comp Syst 3(01n04):87–98

De Domenico M, Granell C, Porter MA, Arenas A (2016) The physics of spreading processes in multilayer networks. Nat Phys 12(10):901–906

De Domenico M, Nicosia V, Arenas A, Latora V (2015) Structural reducibility of multilayer networks. Nat Commun 6(1):1–9

De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y, Porter MA, Gómez S, Arenas A (2013) Mathematical formulation of multilayer networks. Phys Rev X 3(4):041022

Del Valle S, Hethcote H, Hyman JM, Castillo-Chavez C (2005) Effects of behavioral changes in a smallpox attack model. Math Biosci 195(2):228–251

Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5(1):17–60

Flache A, Mäs M, Feliciani T, Chattoe-Brown E, Deffuant G, Huet S, Lorenz J (2017) Models of social influence: towards the next frontiers. J Artif Soc Soc Sim 20(4)

Flyvbjerg B (2020) The law of regression to the tail: how to survive covid-19, the climate crisis, and other disasters. Environ Sci Policy 114:614–618

Friedkin NE, Johnsen EC (1990) Social influence and opinions. J Math Soc 15(3–4):193–206

Granell C, Gómez S, Arenas A (2013) Dynamical interplay between awareness and epidemic spreading in multiplex networks. Phys Rev Lett 111(12):1287011

Granell C, Gómez S, Arenas A (2014) Competing spreading processes on multiplex networks: awareness and epidemics. Phys Rev E 90(1):012808

Hristova D, Rutherford A, Anson J, Luengo-Oroz M, Mascolo C (2016) The international postal network and other global flows as proxies for national wellbeing. PLoS ONE 11(6):0155976

Jager W, Amblard F (2005) Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. Comput Math Organiz The 10(4):295–303

Kaluza P, Kölzsch A, Gastner MT, Blasius B (2010) The complex network of global cargo ship movements. J R Soc Interface 7(48):1093–1103

Kapoor A, Ben X, Liu L, Perozzi B, Barnes M, Blais M, O'Banion S (2020) Examining covid-19 forecasting using spatio-temporal graph neural networks. arXiv:2007.03113

Kaszeta D (2020) Toxic: a history of nerve agents, from Nazi Germany to Putins Russia. Oxford University Press, Oxford

Keeling MJ, Rohani P (2011) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton

Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc Lond A 115(772):700–721

Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. J Comput Netw 2(3):203–271

Madnick S (2020) How do you prepare for the unexpected cyber attack? Available at SSRN 3555451

Newman M (2018) Networks. Oxford University Press, Oxford

Panagopoulos G, Nikolentzos G, Vazirgiannis M (2020) Transfer graph neural networks for pandemic forecasting

Pescetelli N, Rees G, Bahrami B (2016) The perceptual and social components of metacognition. J Exp Psychol Gen 145(8):949

Rombach MP, Porter MA, Fowler JH, Mucha PJ (2014) Core-periphery structure in networks. SIAM J Appl Math 74(1):167–190

Rosenberg L, Baltaxe D, Pescetelli N (2016) Crowds vs swarms, a comparison of intelligence. In: 2016 swarm/human blended intelligence workshop (SHBI). IEEE, pp 1–4

Rutherford A, Cebrian M, Dsouza S, Moro E, Pentland A, Rahwan I (2013a) Limits of social mobilization. Proc Natl Acad Sci 110(16):6281–6286

Rutherford A, Cebrian M, Rahwan I, Dsouza S, McInerney J, Naroditskiy V, Venanzi M, Jennings NR, DeLara JR, Wahlstedt E et al (2013b) Targeted social mobilization in a global manhunt. PLoS ONE 8(9):74628

Sherif CW, Sherif M, Nebergall RE (1965) Attitude and attitude change: the social judgment-involvement approach. Saunders, Philadelphia

Simon HA (1956) Rational choice and the structure of the environment. Psychol Rev 63(2):129

Strano E, Shai S, Dobson S, Barthelemy M (2015) Multiplex networks in metropolitan areas: generic features and local effects. J R Soc Interface 12(111):20150651

Taleb NN (2007) The black swan: the impact of the highly improbable. Random House, New York City

The New York Times: "Dangerous Stuff": Hackers Tried to Poison Water Supply of Florida Town (2021). www.nytimes.com/2021/02/08/us/oldsmar-florida-water-supply-hack.html?campaign_id=9&emc=edit_nn_20210208&instance_id=26913&nl=the-morning&regi_id=97308334&segment_id=51272&te=1&user_id=2699fcd2b975bc1895466f4268a58cf7. Accessed 20 Feb 2021

Tversky A, Kahneman D (1971) Belief in the law of small numbers. Psychol Bull 76(2):105

Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131

Vitali S, Battiston S (2014) The community structure of the global corporate network. PLoS ONE 9(8):104655

Watts DJ, Strogatz SH (1988) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442

Wein LM, Craft DL, Kaplan EH (2003) Emergency response to an anthrax attack. Proc Natl Acad Sci 100(7):4346–4351

Zhang JS, Keegan BC, Lv Q, Tan C (2020) A tale of two communities: characterizing reddit response to covid-19 through/r/china_flu and/r/coronavirus. arXiv:2006.04816

Publisher's Note

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