Infection-induced cascading failures – impact and mitigation

Check for updates

Bo Li [1,2] ✉ & David Saad [2] ✉

In the context of epidemic spreading, many intricate dynamical patterns can emerge due to the cooperation of different types of pathogens or the interaction between the disease spread and other failure propagation mechanism. To unravel such patterns, simulation frameworks are usually adopted, but they are computationally demanding on big networks and subject to large statistical uncertainty. Here, we study the two-layer spreading processes on unidirectionally dependent networks, where the spreading infection of diseases or malware in one layer can trigger cascading failures in another layer and lead to secondary disasters, e.g., disrupting public services, supply chains, or power distribution. We utilize a dynamic message-passing method to devise efficient algorithms for inferring the system states, which allows one to investigate systematically the nature of complex intertwined spreading processes and evaluate their impact. Based on such dynamic message-passing framework and optimal control, we further develop an effective optimization algorithm for mitigating network failures.

Epidemic outbreaks do not only possess a direct threat to public health but also, indirectly, impact other sectors[1–3]. For instance, when many infected individuals have to rest, be hospitalized or quarantined in order to slow down the epidemic spread, this could severely disrupt public services, causing disutility even to those who are not infected. For instance, the highly interdependent supply chains can be easily disrupted due to epidemic outbreaks[4,5]. Similar concerns apply to cyber security. The spread of malware is not merely detrimental to computer networks, but can also cause failures to power grids or urban transportation networks which rely on modern communication systems[6,7]. What is even worse is that the failures of certain components of technological networks can by themselves trigger a cascade of secondary failures, which can eventually lead to large-scale outages[8]. Therefore, it is vital to understand the nature of epidemic (or malware) spreading and failure propagation on interacting networks, based on which further mitigation and control measures can be devised.

A number of previous papers address the scenario of interacting spreading processes. In the context of epidemic spreading, two types of pathogens can cooperate or compete with each other, creating many intricate patterns of disease propagation[9–15]. For interacting technological networks (e.g., communication and power networks), the failure of components in one network layer will not only affect neighboring parts within the same network, but will also influence the second network layer through the cross-layer connections. Macroscopic analyses based on simplified models show that such a spreading mechanism can easily result in a catastrophic breakdown of the whole system[16–18].

Most existing research in the area of multi-layer spreading processes employs macroscopic approaches, such as the degree-distribution-based mean-field methods and asymptotic percolation analysis, in order to obtain the global picture of the models' behavior[19]. Such methods typically do not consider specific network instances and lack the ability to treat the interplay between the spreading dynamics and the fine-grained network topology[19]. For stochastic spreading processes with specific system conditions (e.g., topology initial conditions and individual node properties), it is common to apply extensive Monte Carlo (MC) simulations to observe the evolution of the spread, based on which important policy decisions are made[20]. However, such simulations are computationally demanding on big networks and can be subject to large statistical uncertainty; as a result, they are difficult to be used for downstream analysis or optimization tasks. Therefore, researchers have been pursuing tractable and accurate theoretical methods to tackle the complex stochastic dynamics on networks[19,21].

Among the various developed theoretical approaches used, dynamic message-passing (DMP) is based on ideas from statistical physics offering a desirable algorithmic framework for approximate inference while it remains computationally efficient[22–24]. The DMP method has been shown to be more accurate than the widely adopted individual-based mean-field method, especially in sparse networks[25,26]. Moreover, the DMP approach yields a set

[1]School of Science, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China. [2]Non-linearity and Complexity Research Group, Aston University, Birmingham B4 7ET, UK. ✉e-mail: libo2021@hit.edu.cn; d.saad@aston.ac.uk

1

of closed-form equations, which is very convenient for additional parameter estimation and optimization tasks[14,27,28].

In this work, we study a scenario where the epidemic or malware spreading on one network can trigger cascading failures on another. This is relevant in the cases where epidemic outbreaks cause disruption in public services or economic activities. Similarly, it can also be applied to study the effect of malware spread on computer networks causing the breakdown of other technological networks such as the power grid. The latter phenomenon is gaining more and more attention due to the increasing interactions among various engineering networks[7]. We explore the dynamics and consequences of such infection-induced cascading failures across two-layer networks using the DMP method. Our results reveal that even relatively low infection rates can induce large-scale cascading failures, leading to widespread network disruptions. We characterized these phenomena through the derivation and analysis of DMP equations, achieving a comprehensive understanding by linking the process to combined bond and bootstrap percolation models analytically. Leveraging the analytical tractability of the DMP model, we also developed optimization algorithms that effectively mitigate these network failures. By adjusting control parameters based on the back-propagation of final state impacts, these algorithms help minimize the size of system failure.

Methods
Model and framework
The model. To study the impact of infection spread of diseases or malware and their secondary effects, we consider multiplex networks comprising two layers[29], which are denoted as layers a and b, and are represented by two graphs $G_a(V_a, E_a)$ and $G_b(V_b, E_b)$. For convenience, we assume that the nodes in both layers correspond to the same set of individuals, denoted as $V = V_a = V_b$. This can be extended to more general settings. Denote ∂_i^a and ∂_i^b as the sets of nodes adjacent to node i in layers a and b, respectively. We also define $\partial_i = \partial_i^a \cup \partial_i^b$. See Fig. 1 for an example of the network model under consideration.

Each node has states on both layers a and b. In layer a, each node assumes one of four states, susceptible (S), infected (I), recovered (R), and protected (P) at any particular time step. The infection spreading process occurs in layer a only, which is dictated by the stochastic discrete-time SIR model[19] augmented with a protection mechanism, which we term the SIRP model. Stochastic models are commonly employed for modeling the spreads of epidemics or malware[20,30,31]. The stochastic SIR model is commonly used for representing the spread of infections, wherein a susceptible individual (in state S) may become infected through contact with infected neighbors, and an infected individual (in state I) can recover, transitioning to the recovered state (R) after a certain period. The process we consider is based on the SIR model but includes one more state, P, in layer a; it admits the following state-transition rule

$$
\begin{aligned}
S(i) + I(j) &\xrightarrow{\beta_{ji}} I(i) + I(j), \\
I(i) &\xrightarrow{\mu_i} R(i) \\
S(i) &\xrightarrow{\gamma_i(t)} P(i),
\end{aligned}
\tag{1}
$$

where β_{ji} is the probability that node j being in the infected state transmits the infection to its susceptible neighboring node i at a certain time step. At each time step, an existing infected node i recovers with probability μ_i; the recovery process is assumed to occur after possible transmission activities. At time t, an existing susceptible node i turns into state P if it receives protection at time $t - 1$, which occurs with probability $\gamma_i(t - 1)$. The protection can be achieved by vaccination in the epidemic setting or special protection measures in the malware spread setting, which is usually subject to certain budget constraints. The protection probabilities $\{\gamma_i(t)\}$ will be the major control variables for mitigating the outbreaks. Note that when no protection is provided, i.e., all $\{\gamma_i(t)\}$ are zero, the SIRP model reduces to the
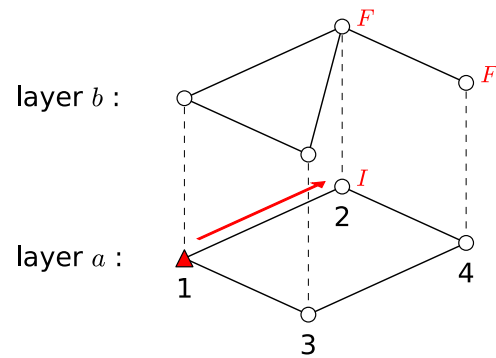

Fig. 1 | An example of the two-layer spreading process considered in this work. A node is in state I if it is infected in layer a, and a node is in state F if it fails in layer b. In this example, node 2 is infected by node 1 in layer a, therefore it turns into state F in layer b. If $b_{24} \geq \Theta_4$, then node 4 will also fail as it loses the support from node 2, even though node 4 itself has not been infected.

traditional SIR model. At initial time $t = 0$, we assume that node i has a probability $P_S^i(0)$ to be in state S, and probability $P_I^i(0) = 1 - P_S^i(0)$ to be in state I.

In layer b, each node i can either be in the normal state (N) or the failed state (F), indicated by a binary state variable x_i where $x_i = 1$ (0) denotes the 'fail' ('normal') state at a particular time step. A node i in layer b fails if (i) it has been infected, i.e., node i is in state I or R in layer a; (ii) there exists certain neighboring failed nodes such that $\sum_{j \in \partial_i^b} x_j b_{ji} \geq \Theta_i$, where Θ_i is a threshold and the influence parameter b_{ji} measures the importance of the failure of node j on node i. The latter case indicates that node i can fail due to the failures of its neighbors which it relies on, even though node i itself is not infected. In summary, the failure propagation process in layer b can be expressed as

$$
x_i =
\begin{cases}
1, & \text{either (i) node } i \text{ in state } I \text{ or } R \text{ in layer } a, \\
& \text{or (ii)} \sum_{j \in \partial_i^b} x_j b_{ji} \geq \Theta_i \text{ in layer } b; \\
0, & \text{otherwise.}
\end{cases}
\tag{2}
$$

The whole process is simulated for T time steps. As we are interested in the time scale of infection spread which is usually very fast, we do not consider any repair rule in layer b. Therefore, a failed node cannot return to normality within the time window under consideration.

Such a failure propagation mechanism is equivalent to the linear threshold model (LTM), which is commonly used in studying social contagion and other cascade processes[19,32,33]. The LTM model also offers a straightforward yet effective framework for understanding cascading failures in various systems, as it effectively encapsulates the pivotal dynamics where a component can become dysfunctional if a significant number of its dependent components fail[18,32]. Other popular models for cascading failures incorporate more details of the system functionalities[34–36]; these models require theoretical analyses specific to each case, which fall outside the scope of the current study.

Figure 1 illustrates the infection-induced cascades of our model in a simple network of 4 nodes. Node 1 is the initial infected node (or the seed) in layer a, which transmits the infection to node 2 at a certain time step. Now that node 2 is in the infected state in layer a, it also fails to function in layer b. If $b_{24} \geq \Theta_4$, then node 4 will also fail as it loses the support from node 2, even though node 4 itself has not been infected. Such additional cascade propagation needs extra care when infections spread out. Similar interacting SIR (without a protection mechanism) and LTM processes have also been considered in the social contagion setting[37].

We reiterate that the infection-spreading process (described by the SIRP model) occurs in layer a only and not the entire network, while the cascade process (described by the LTM model) occurs in layer b. Typically, a

holistic treatment of the combined two-layer processes is needed to understand their impact and develop mitigation strategies. We also remark that our model differs from the traditional settings of interdependent networks, which typically includes reciprocal dependency.

The DMP framework. We aim to use the DMP approach to investigate the two-layer spreading processes described above. The DMP equations of the usual SIR and the LTM model have been derived, based on the microscopic dynamic belief propagation equations[24,38]. As in generic belief propagation methods[39], the DMP method is exact for tree graphs, while it can constitute a good approximation for loopy graphs, particularly when short loops, such as those spanning 3 or 4 nodes, are scarce. The two-layer spreading processes combining the SIR and LTM model appear more involved, where approximations relying on uncorrelated multiplex networks were used[37]. Such approximations become less adequate when the two network layers are correlated, e.g., both layers share the same network topology.

Dynamic belief propagation. To devise more accurate DMP equations for general network models and accommodate the protection mechanism for mitigation, we start from the principled dynamic belief propagation equations of the two-layer processes. One important characteristic of our model is that state transition is unidirectional, which can only take the direction $S \to I \to R$ or $S \to P$ in layer a, and $N \to F$ in layer b. Note that layer b does not influence layer a. As a result, our model admits a reduced representation of the system's dynamical trajectories that subsequently facilitates a drastic simplification of the derivation of the DMP equations, which are exact on tree networks[24]. Nevertheless, we emphasize that the exactness of the DMP formalism for tree networks is conditioned on the unidirectional nature of the model, which no longer holds if layer b also influences layer a. Introducing reciprocal interactions between both model layers requires additional theoretical tools, which are interesting by themselves but are beyond the scope of the current study.

Following previous works[24,38], we parametrize the dynamical trajectory of each node by its state transition times. In layer a, we denote τ_i^a, ω_i^a and ε_i^a as the first time at which node i turns into state I, R and P, respectively. In layer b, we denote τ_i^b as the first time at which node i turns into state F. The quantity of interest is the probability of the trajectory of node i considered in the entire graph comprising layers a and b but having a cavity where node j is absent, denoted as $m^{i \to j}(\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b)$. Throughout the manuscript, we will refer to probabilities defined within a cavity graph as cavity probabilities. It is computed by the following dynamic belief propagation equations

$$
\begin{aligned}
& m^{i \to j}(\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b) \\
= & \sum_{\{\tau_k^a, \omega_k^a, \varepsilon_k^a, \tau_k^b\}_{k \in \partial_i}} W_{\mathrm{SIRP}}^i(\tau_i^a, \omega_i^a, \varepsilon_i^a || \{\tau_k^a, \omega_k^a, \varepsilon_k^a\}_{k \in \partial_i^a}) \\
& \times W_{\mathrm{LTM}}^i(\tau_i^b || \tau_i^a, \varepsilon_i^a, \{\tau_k^b\}_{k \in \partial_i^b}) \\
& \times \prod_{k \in \partial_i \setminus j} m^{k \to i}(\tau_k^a, \omega_k^a, \varepsilon_k^a, \tau_k^b),
\end{aligned}
$$
(3)

where $W_{\mathrm{SIRP}}^i(\cdot)$ and $W_{\mathrm{LTM}}^i(\cdot)$ are the transition kernels dictated by the dynamical rules of the SIRP and LTM model, respectively (for details see Supplementary Note 1). The marginal probability of the trajectory of node i, denoted as $m^i(\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b)$, can be computed in a similar way as Eq. (3), by replacing the product $\prod_{k \in \partial_i \setminus j}$ in the last line of Eq. (3) by $\prod_{k \in \partial_i}$. That is, the marginal probability $m^i(\cdot)$ is calculated using the entire graph, in contrast to the cavity probability $m^{i \to j}(\cdot)$ which is determined with a cavity graph where node j is absent.

The probability of node i in a certain state can be computed by summing the trajectory-level probability, which will be described in the next section.

Full node-level DMP equations

Consider the cavity probability of node i being in state S in layer a at time t (assuming node j is absent - the cavity), it is obtained by tracing over the corresponding probabilities of trajectories $m^{i \to j}(\cdot)$ in the cavity graph (assuming node j is removed)

$$
\begin{aligned}
P_S^{i \to j}(t) = & \sum_{\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b} \mathbb{I}(t < \tau_i^a < \omega_i^a) \mathbb{I}(t < \varepsilon_i^a) \\
& \times m^{i \to j}(\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b),
\end{aligned}
$$
(4)

where $\mathbb{I}(\cdot)$ is the indicator function enforcing the order of state transitions. Similarly, we denote the cavity probability of node i in state F in layer b (in the absence of node j) as $P_F^{i \to j}(t)$; it is obtained by

$$
P_F^{i \to j}(t) = \sum_{\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b} \mathbb{I}(\tau_i^b \leq t) m^{i \to j}(\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b).
$$
(5)

The marginal probabilities $P_S^i(t)$ and $P_F^i(t)$ can be computed in a similar manner, by replacing $m^{i \to j}(\cdot)$ in Eq. (4) and Eq. (5) with $m^i(\cdot)$.

DMP equations in Layer a. We note that infection spread in layer a is not influenced by cascades in layer b, while the failure time in layer b depends on the infection time and the protection time of the corresponding node in layer a. Hence, we can decompose the message $m^{i \to j}(\cdot)$ to the respective components as

$$
\begin{aligned}
m^{i \to j}(\tau_i^a, \omega_i^a, \varepsilon_i^a, \tau_i^b) = & m_a^{i \to j}(\tau_i^a, \omega_i^a, \varepsilon_i^a) \\
& \times m_b^{i \to j}(\tau_i^b | \tau_i^a, \varepsilon_i^a).
\end{aligned}
$$
(6)

where $m_a^{i \to j}(\cdot)$ and $m_b^{i \to j}(\cdot)$ denote the trajectory-level probabilities of the processes in layer a and b, respectively. Note that the messages $\{m^{i \to j}(\cdot)\}$ live in the entire network comprising layers a and b, which implies that $\{m_a^{i \to j}(\cdot), m_b^{i \to j}(\cdot)\}$ are also defined on the entire network.

Summing $m_a^{i \to j}(\cdot)$ over $\tau_i^a, \omega_i^a, \varepsilon_i^a$ up to a certain time yields the normal DMP equations of node-level probabilities for the infection spread in layer a (see details in Supplementary Note 1). They admit the following expressions for $t > 0$

$$
P_S^{i \to j}(t) = P_S^i(0) \prod_{t'=0}^{t-1} [1 - \gamma_i(t')] \prod_{k \in \partial_i^a \setminus j} \theta^{k \to i}(t),
$$
(7)

$$
\begin{aligned}
\theta^{k \to i}(t) &= \theta^{k \to i}(t-1) - \beta_{ki} \phi^{k \to i}(t-1), \\
\phi^{k \to i}(t) &= (1 - \beta_{ki})(1 - \mu_k) \phi^{k \to i}(t-1)
\end{aligned}
$$
(8)

$$
- \{ P_S^{k \to i}(t) - P_S^{k \to i}(t-1)[1 - \gamma_k(t-1)] \},
$$
(9)

where $\theta^{k \to i}(t)$ is the cavity probability that node k has not transmitted the infection signal to node i up to time t, and $\phi^{k \to i}(t)$ is the cavity probability that k is in state I but has not transmitted the infection signal to node i up to time t. Note that the messages $\{P_S^{k \to i}(t), \theta^{k \to i}(t), \phi^{k \to i}(t)\}$ are only needed for edges belonging to layer a where the SIRP model is defined.

At time $t = 0$, as we consider that each node i is either in state S with probability $P_S^i(0)$ or in state I with probability $1 - P_S^i(0)$, we have the following initial conditions for the messages

$$
\begin{aligned}
P_S^{i \to j}(0) &= P_S^i(0), \\
\phi^{i \to j}(0) &= 1 - P_S^i(0), \\
\theta^{i \to j}(0) &= 1.
\end{aligned}
$$
(10)

Upon iterating the above messages (7)-(8) starting from the initial conditions (10), the node-level marginal probabilities can be computed as

$$P_S^i(t) = P_S^i(0) \prod_{t'=0}^{t-1} [1 - \gamma_i(t')] \prod_{k \in \partial_i^a} \theta^{k \to i}(t), \tag{11}$$

$$P_R^i(t) = P_R^i(t-1) + \mu_i P_I^i(t-1), \tag{12}$$

$$P_P^i(t) = P_P^i(t-1) + \gamma_i(t-1) P_S^i(t-1), \tag{13}$$

$$P_I^i(t) = 1 - P_S^i(t) - P_R^i(t) - P_P^i(t). \tag{14}$$

The above DMP Eqs. (11)–(14) bear similarity to those of SIR model[23], except for the protection mechanism with control parameters $\{\gamma_i(t)\}$. The computational complexity for obtaining the messages for the SIRP process in layer a over a total time T is $O(|E_a|T)$, where $|E_a|$ denotes the number of edges in layer a.

DMP equations in layer b. As for the cascade process in layer b, whether node i will turn into state F (fail) also depends on the state in layer a, making it more challenging to derive the corresponding DMP equations. The key to obtaining node-level DMP equations for $P_F^{i \to j}(t)$ in Eq. (5) (and the corresponding marginal probability $P_F^i(t)$) is to introduce several intermediate quantities to facilitate the calculation; the details are outlined in Supplementary Note 1.

To summarize, the node-level failure probability $P_F^i(t)$ can be decomposed as

$$P_F^i(t) = P_I^i(t) + P_R^i(t) + P_{SF}^i(t) + P_{PF}^i(t), \tag{15}$$

where $P_{SF}^i(t)$ and $P_{PF}^i(t)$ are the probabilities that node i is in state F in layer b, while it is in state S or state P in layer a, respectively. For these two cases, the failure of node i is triggered by the failure propagation of its neighbors from layer b. A similar relation holds for the cavity probability $P_F^{i \to j}(t)$.

The probability $P_{SF}^i(t)$ admits the following iteration

$$\begin{aligned}
P_{SF}^i(t) = P_S^i(0) \prod_{t'=0}^{t-1} [1 - \gamma_i(t')] \prod_{k \in \partial_i^a \setminus \partial_i^a \cap \partial_i^b} \theta^{k \to i}(t) \\
\times \sum_{\{x_k\}_{k \in \partial_i^b}} \mathbb{I}\left(\sum_{k \in \partial_i^b} x_k b_{ki} \geq \Theta_i\right) \\
\times \prod_{\substack{k \in \partial_i^b \setminus \partial_i^a \cap \partial_i^b, \\ x_k = 1}} P_F^{k \to i}(t-1) \prod_{\substack{k \in \partial_i^b \setminus \partial_i^a \cap \partial_i^b, \\ x_k = 0}} [1 - P_F^{k \to i}(t-1)] \\
\times \prod_{\substack{k \in \partial_i^a \cap \partial_i^b, \\ x_k = 1}} \chi^{k \to i}(t) \prod_{\substack{k \in \partial_i^a \cap \partial_i^b, \\ x_k = 0}} [\theta^{k \to i}(t) - \chi^{k \to i}(t)],
\end{aligned} \tag{16}$$

where $\chi^{k \to i}(t)$ is the cavity probability that node k is in state F at time $t-1$, and it has not sent the infection signal to node i up to time t.

The cavity probability $\chi^{k \to i}(t)$ can be decomposed into

$$\chi^{k \to i}(t) = \psi^{k \to i}(t) + P_{SF}^{k \to i}(t-1) + P_{PF}^{k \to i}(t-1), \tag{17}$$

where $\psi^{k \to i}(t)$ is the cavity probability that node k is in state I or R at time $t-1$, but has not transmitted the infection signal to node i up to time t. The cavity probability $\psi^{k \to i}(t)$ can be computed as

$$\begin{aligned}
\psi^{k \to i}(t) = & \psi^{k \to i}(t-1) - \beta_{ki} \phi^{k \to i}(t-1) \\
& + [1 - \gamma_k(t-2)] P_S^{k \to i}(t-2) - P_S^{k \to i}(t-1).
\end{aligned} \tag{18}$$

Similarly, the probability $P_{PF}^i(t)$ admits the following iteration

$$\begin{aligned}
P_{PF}^i(t) = P_S^i(0) \sum_{\varepsilon=1}^{t} \gamma_i(\varepsilon - 1) \prod_{t'=0}^{\varepsilon-2} [1 - \gamma_i(t')] \\
\times \prod_{k \in \partial_i^a \setminus \partial_i^a \cap \partial_i^b} \theta^{k \to i}(\varepsilon - 1) \sum_{\{x_k\}_{k \in \partial_i^b}} \mathbb{I}\left(\sum_{k \in \partial_i^b} x_k b_{ki} \geq \Theta_i\right) \\
\times \prod_{\substack{k \in \partial_i^b \setminus \partial_i^a \cap \partial_i^b, \\ x_k = 1}} P_F^{k \to i}(t-1) \prod_{\substack{k \in \partial_i^b \setminus \partial_i^a \cap \partial_i^b, \\ x_k = 0}} [1 - P_F^{k \to i}(t-1)] \\
\times \prod_{\substack{k \in \partial_i^a \cap \partial_i^b, \\ x_k = 1}} \tilde{\chi}^{k \to i}(t, \varepsilon) \prod_{\substack{k \in \partial_i^a \cap \partial_i^b, \\ x_k = 0}} [\theta^{k \to i}(\varepsilon - 1) - \tilde{\chi}^{k \to i}(t, \varepsilon)],
\end{aligned} \tag{19}$$

where the dummy variable ε indicates the time at which node i receives the protection signal.

In Eq. (19), $\tilde{\chi}^{k \to i}(t, \varepsilon)$ is the cavity probability that node k is in state F at time $t-1$, but has not transmitted the infection signal to node i up to time ε. It can be decomposed into

$$\tilde{\chi}^{k \to i}(t, \varepsilon) = \tilde{\psi}^{k \to i}(t, \varepsilon) + P_{SF}^{k \to i}(t-1) + P_{PF}^{k \to i}(t-1), \tag{20}$$

where $\tilde{\psi}^{k \to i}(t, \varepsilon)$ is the cavity probability that node k is in state I or R at time $t-1$, but has not transmitted the infection signal to node i up to time $\varepsilon - 1$. The cavity probability $\tilde{\psi}^{k \to i}(t)$ can be computed as

$$\begin{aligned}
\tilde{\psi}^{k \to i}(t, \varepsilon) = & \psi^{k \to i}(\varepsilon - 1) + P_I^{k \to i}(t-1) + P_R^{k \to i}(t-1) \\
& - [P_I^{k \to i}(\varepsilon - 2) + P_R^{k \to i}(\varepsilon - 2)].
\end{aligned} \tag{21}$$

Note that the cavity probabilities $P_{SF}^{i \to j}(t)$ and $P_{PF}^{i \to j}(t)$ are computed using the similar formula as in Eqs. (16) and (19), but in the cavity graph where node j is removed. This closes the loop for the DMP equations in layer b. We also observe in the above equations that the node-level messages for the SIRP process only enter into the DMP equations for the LTM process through the overlapping neighbors $\partial_i^a \cap \partial_i^b$.

The initial conditions for the corresponding messages are given by

$$P_F^k(0) = P_F^{k \to i}(0) = P_I^k(0), \tag{22}$$

$$P_{SF}^k(0) = P_{SF}^{k \to i}(0) = 0, \tag{23}$$

$$P_{PF}^k(0) = P_{PF}^{k \to i}(0) = 0, \tag{24}$$

$$\psi^{k \to i}(1) = \chi^{k \to i}(1) = (1 - \beta_{ki}) P_I^k(0), \tag{25}$$

$$\tilde{\psi}^{k \to i}(1, 1) = \tilde{\chi}^{k \to i}(1, 1) = P_I^k(0). \tag{26}$$

For $t \geq 2, \varepsilon = 1$, we have

$$\begin{aligned}
\tilde{\psi}^{k \to i}(t, \varepsilon = 1) &= P_I^{k \to i}(t-1) + P_R^{k \to i}(t-1), \\
\tilde{\chi}^{k \to i}(t, \varepsilon = 1) &= P_I^{k \to i}(t-1) + P_R^{k \to i}(t-1)
\end{aligned} \tag{27}$$

$$+ P_{SF}^{k \to i}(t-1) + P_{PF}^{k \to i}(t-1). \tag{28}$$

We remark that for a total time T, the computational complexity for obtaining the messages of the cascade process in layer b is $O(|E_b|T^2)$ where $|E_b|$ denotes the number of edges in layer b, unlike the $O(|E_a|T)$ complexity for the SIRP process in layer a. This is due to the dependency of layer b on layer a, as well as the protection mechanism in layer a. The summation of the

dummy state $\{x_k\}_{k \in \partial_i^b}$ in Eq. (16) and Eq. (19) also implies a high computational demand of networks with high-degree nodes. One way to alleviate this complexity is to use the dynamic programming techniques introduced in by Torrisi et al.[40].

These DMP equations are exact if both layers are tree networks, while they are approximate solutions when there are loops in the underlying networks.

Simplification under small inter-layer overlap

If there are no overlaps between the neighbors of node i in layer a and those in layer b, i.e., $\partial_i^a \cap \partial_i^b = \varnothing$, the messages $\chi^{k \to i}, \psi^{k \to i}, \tilde{\chi}^{k \to i}$ and $\tilde{\psi}^{k \to i}$ are not needed, and the node-level probabilities $P_{SF}^i(t)$ and $P_{PF}^i(t)$ can be much simplified as

$$
P_{SF}^i(t) = P_S^i(t) \sum_{\{x_k\}_{k \in \partial_i^b}} \mathbb{I}\left(\sum_{k \in \partial_i^b} x_k b_{ki} \geq \Theta_i \right) \tag{29}
$$
$$
\times \prod_{k \in \partial_i^b, x_k = 1} P_F^{k \to i}(t-1) \prod_{k \in \partial_i^b, x_k = 0} \left[1 - P_F^{k \to i}(t-1)\right],
$$

$$
P_{PF}^i(t) = P_P^i(t) \sum_{\{x_k\}_{k \in \partial_i^b}} \mathbb{I}\left(\sum_{k \in \partial_i^b} x_k b_{ki} \geq \Theta_i \right) \tag{30}
$$
$$
\times \prod_{k \in \partial_i^b, x_k = 1} P_F^{k \to i}(t-1) \prod_{k \in \partial_i^b, x_k = 0} \left[1 - P_F^{k \to i}(t-1)\right].
$$

This is also a reasonable approximation if the two layers a and b have little correlation, which has been exploited by previous work[37]. We remark that the computational complexity of obtaining messages for the cascade process in layer b using this approximated method is $O(|E_b|T)$. In this work, we will employ this approximation when we consider the dynamics in the large time limit and devise an optimization algorithm for mitigating the cascading failures, in order to reduce computing time. In situations where inter-layer overlaps are significant and accuracy is important[41,42], one can always use the complete formulations of the DMP equations as detailed in the "Full Node-level DMP Equations" subsection above.

Results

Effectiveness of the DMP method

We firstly test the efficacy of the complete DMP equations derived in "Full Node-level DMP Equations" subsection in the Methods section, by comparing the node-level probabilities $P_S^i(t)$ and $P_F^i(t)$ to those obtained by Monte Carlo simulations. The DMP theory produces exact marginal probabilities for node activities in tree networks; this is verified in Fig. 2a, b where both layers a and b are the same binary tree network of size $N = 63$. For random regular graphs (RRG) where there are many loops, the DMP method also yields reasonably accurate solutions; this is demonstrated in Fig. 2c, d where both layers a and b are the same RRG of size $N = 100$ and degree $K = 5$. We also validate the effectiveness of the non-overlapping approximation applied to the DMP equations for the process in layer b introduced in the subsection "Simplification under Small Inter-layer Overlap" in Methods; the results are shown in Supplementary Note 2.

Impact of infection-induced cascades

The obtained DMP equations of the two-layer spreading processes allow us to examine the impact of the infection-induced cascading failures, on either a specific instance of a multiplex network or an ensemble of networks following a certain degree distribution. In this section, we do not consider the protection of nodes by setting $\gamma_i(t) = 0$, where the process in layer a is essentially a discrete-time SIR model.
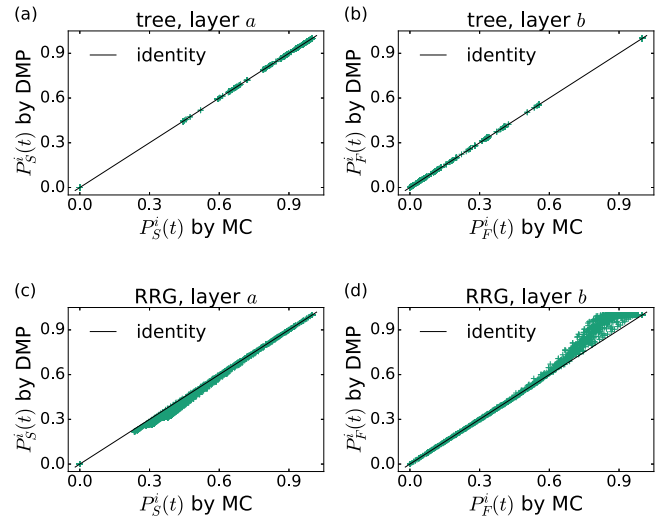
Fig. 2 | Comparison of node-level probabilities. The node-level probabilities $P_S^i(t)$ and $P_F^i(t)$ are obtained by the DMP theory and Monte Carlo (MC) simulation (averaged over 10^5 realizations). Panels **a** and **b** correspond to a binary tree network of size $N = 63$ for both layers. Panels **c** and **d** correspond to a random regular graph (RRG) of size $N = 100$ and degree $K = 5$ for both layers. The system parameters are $T = 50, \beta_{ji} = 0.2, \mu_i = 0.5, b_{ji} = 1, \Theta_i = 0.6|\partial_i^b|, \gamma_i(t) = 0$.

Impact on a specific network. For the process in layer a, we define the outbreak size at time t as the fraction of nodes that have been infected at that time

$$
\rho_I(t) + \rho_R(t) = \frac{1}{N} \sum_{i \in V_a} P_I^i(t) + \frac{1}{N} \sum_{i \in V_a} P_R^i(t). \tag{31}
$$

For the process in layer b, we define the cascade size at time t as the fraction of nodes that have failed at that time

$$
\rho_F(t) = \frac{1}{N} \sum_{i \in V_b} P_F^i(t). \tag{32}
$$

By definition, we have $\rho_F(t) \geq \rho_I(t) + \rho_R(t)$.

In Fig. 3, we demonstrate the time evolution of the infection outbreak size and the cascade size in a multiplex network where both layers are random regular graphs with size $N = 1600$. It can be observed that ρ_F is much larger than $\rho_I + \rho_R$ asymptotically, which suggests that the failure propagation mechanism in layer b significantly amplifies the impact of the infection outbreaks in layer a. In particular, the failure can eventually propagate to the whole network even though less than 70% of the population gets infected when the spread of the infection saturates. Compare to Monte Carlo simulations, the DMP method systematically overestimates the outbreak sizes due to the effect of mutual infection, but it has been shown to offer a significant improvement over the individual-based mean-field method[25,26,43].

Asymptotic properties. In the above example, the system converges to a steady state in the large time limit. The DMP approach allows us to systematically investigate the asymptotic behavior of the two-layer spreading processes.

For the process in layer a, we define an auxiliary probability

$$
p_{ij} := \frac{\beta_{ij}}{\beta_{ij} + \mu_i - \beta_{ij}\mu_i}. \tag{33}
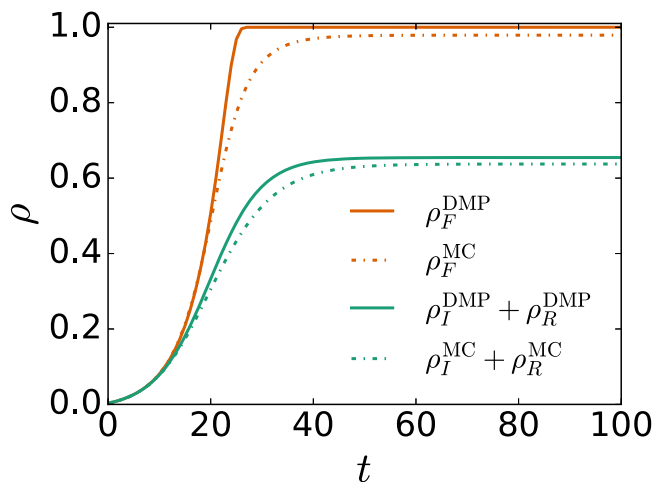$$

Fig. 3 | Evolution of the sizes of the infection outbreak in layer *a* and total failures in layer *b*. The size of infection outbreak is measured by $\rho_I + \rho_R$ (green lines), while the size of total failures is measured by ρ_F (orange lines). Both the DMP method (solid line) and MC simulation (dashed-dotted line) are considered. Layer *a* and layer *b* have different network topologies, but both are realizations of random regular graphs of size $N = 1600$ and degree $K = 5$. At time $t = 0$, there are 5 infected nodes. The system parameters are $\beta_{ji} = 0.2, \mu_i = 0.5, b_{ji} = 1, \Theta_i = 0.6|\partial_i^b|, \gamma_i(t) = 0$.

Then the messages in layer *a* admit the following expressions in the limit $T \to \infty$

$$\phi^{i \to j}(\infty) = 0,$$
$$\theta^{i \to j}(\infty) = 1 - p_{ij} + p_{ij} P_S^{i \to j}(\infty),$$
$$P_S^{i \to j}(\infty) = P_S^i(0) \prod_{k \in \partial_i^a \setminus j} \theta^{k \to i}(\infty), \tag{34}$$
$$P_S^i(\infty) = P_S^i(0) \prod_{k \in \partial_i^a} \theta^{k \to i}(\infty),$$

Details of the derivation can be found in Supplementary Note 3. The above asymptotic equations (34) suggest a well-known relationship between epidemic spreading and bond percolation[19,22,44]. The bond percolation problem involves a network where the bonds (or edges) between nodes are randomly occupied with a certain probability (denoted as λ). The main focus is to understand the formation of a giant cluster comprising connected occupied edges in the network; in large systems, this typically occurs when λ is greater than a transition point λ_c[45].

As mentioned above, it is well established that the asymptotic properties of many stochastic epidemic spreading models can be mapped to certain bond percolation problems[44,46]; we refer interested readers to two recent reviews for more details on the subject[19,45]. In the SIR model studied here (where $\gamma_i(t) = 0$), the quantity p_{ij} defined in Eq. (33) can be interpreted as the probability that an infection transmission on edge (i, j) has been realized in the long run, corresponding to an edge occupation probability in bond percolation. When the transmission probabilities $\{\beta_{ij}\}$ are large ($\{p_{ij}\}$ will also be large), a few initially infected seeds can eventually infect a significant proportion of the population and lead to a pandemic, which corresponds to the formation of a giant cluster in percolation theory. We refer readers to Supplementary Note 3 for more details of the correspondence between our model and bond percolation. Note that the edge occupation probability p_{ij} in this discrete-time SIR model differs from the continuous-time counterpart[22,44] with an additional term $\beta_{ij}\mu_i$ in the denominator. The term $\beta_{ij}\mu_i$ accounts for the simultaneous events that node i infects node j and recovers within the same time step[25].

For the process in layer *b*, we assume that layers *a* and *b* are weakly correlated due to their different topologies and adopt the approximation made in the subsection "Simplification under Small Inter-layer Overlap" in

Methods. As no protection is applied, we have $P_{PF}^i(t) = 0$. Then the messages in layer *b* admit the following expression in the limit $T \to \infty$

$$P_F^{i \to j}(\infty) = 1 - P_S^i(\infty)$$
$$+ P_S^i(\infty) \sum_{\{x_k\}_{k \in \partial_i^b \setminus j}} \mathbb{I}\left(\sum_{k \in \partial_i^b \setminus j} x_k b_{ki} \geq \Theta_i\right)$$
$$\times \prod_{k \in \partial_i^b \setminus j, x_k = 1} P_F^{k \to i}(\infty) \prod_{k \in \partial_i^b \setminus j, x_k = 0} \left[1 - P_F^{k \to i}(\infty)\right], \tag{35}$$

where a similar expression holds for $P_F^i(\infty)$ by replacing $\partial_i^b \setminus j$ with ∂_i^b in Eq. (35). The asymptotic equations for layer *b* suggest a relationship between the LTM model and bootstrap percolation[38].

Two-layer percolation in large homogeneous networks. The large-time behaviors of the two processes correspond to two types of percolation problems. To further examine the macroscopic critical behaviors of the two-layer percolation models, it is convenient to consider large-size random regular graphs of degree K (which have a homogeneous network topology), and homogeneous system parameters with $\beta_{ji} = \beta, \mu_i = \mu, b_{ji} = b, \Theta_i = \Theta$. We further assume that each node i has a vanishingly small probability of being infected at time $t = 0$ with $P_I^i(0) = 1 - P_S^i(0) \propto 1/N$. In the large size limit $N \to \infty$, we have $P_S^i(0) \to 1$.

Due to the homogeneity of the system, one can assume that all messages and marginal probabilities are identical,

$$\theta^{i \to j}(\infty) = \theta^\infty, \tag{36}$$
$$P_F^{i \to j}(\infty) = P_F^\infty, \tag{37}$$
$$P_S^i(\infty) = \rho_S^\infty, \tag{38}$$
$$P_F^i(\infty) = \rho_F^\infty. \tag{39}$$

It leads to the self-consistent equations in the large size limit ($N \to \infty$),

$$\theta^\infty = 1 - p + p \cdot (\theta^\infty)^{K-1}, \tag{40}$$

$$\rho_S^\infty = (\theta^\infty)^K, \tag{41}$$

$$P_F^\infty = 1 - \rho_S^\infty$$
$$+ \rho_S^\infty \sum_{n = \lceil \Theta \rceil}^{K-1} \binom{K-1}{n} (P_F^\infty)^n (1 - P_F^\infty)^{K-1-n}, \tag{42}$$

$$\rho_F^\infty = 1 - \rho_S^\infty$$
$$+ \rho_S^\infty \sum_{n = \lceil \Theta \rceil}^{K} \binom{K}{n} (P_F^\infty)^n (1 - P_F^\infty)^{K-n}, \tag{43}$$

where $p = \frac{\beta}{\beta + \mu - \beta\mu}$ and $\lceil x \rceil$ is the smallest integer greater than or equal to x.

We observe that $\theta^\infty = 1, \rho_S^\infty = 1, P_F^\infty = 0, \rho_F^\infty = 0$ is always a fixed point to Eqs. (40)–(43), which corresponds to vanishing outbreak sizes. When the infection probability β is larger than a critical point β_c^a, this fixed point solution becomes unstable and another fixed point with finite outbreak sizes develops.

As a concrete example, we consider random regular graphs of degree $K = 5$ and fix $\mu = 0.5, b = 1, \Theta = 3$. By solving Eqs. (40)–(43) for different β, we obtain outbreak sizes for both layers *a* and *b* under different infection strengths. The result is shown in Fig. 4, where the asymptotic theory accurately predicts the behavior of a large-size system ($N = 1600$) in the large-time limit. It is also observed that the outbreak sizes in both layers
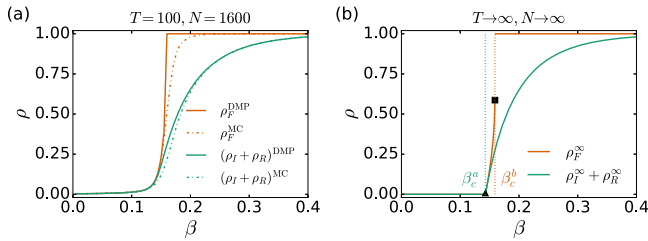
Fig. 4 | Size of infection outbreak and total failures as a function of the infection probability β. The size of infection outbreak is measured by $\rho_I + \rho_R$ (green lines), while the size of total failures is measured by ρ_F (orange lines). The limits of large system size and large time are considered. **a** Random regular graphs with $N = 1600$, $K = 5$ are considered. The spreading processes are iterated for $T = 100$ steps, where stationary states are attained. Both the DMP method (solid line) and MC simulation (dashed-dotted line) are considered. **b** Random regular graphs with $K = 5$ in the asymptotic limit $T \to \infty$, $N \to \infty$ are considered by analyzing the large-time behaviors of the DMP equations. The triangle and the square markers indicate the locations of the two transition points β_c^a and β_c^b, respectively. The system parameters are homogeneous, with $\mu = 0.5$, $b = 1$, $\Theta = 3$, $\gamma_i(t) = 0$.

become non-zero when β is larger than a critical point $\beta_c^a = \frac{1}{7}$. Furthermore, the outbreak size ρ_F^∞ in layer b exhibits a discontinuous jump to a complete breakdown ($\rho_F^\infty = 1$) when β increases and surpasses another transition point $\beta_c^b \approx 0.159$. However, at the transition point β_c^b, only about 28.6% of the population has been infected in layer a.

This example again indicates that the cascading failure propagation in layer b can drastically amplify the impact of the epidemic outbreaks in layer a. Lastly, we remark that whether layer b will exhibit a discontinuous transition or not depends on the values of K and Θ[38], as predicted by the bootstrap percolation theory[47].

Mitigation of Infection-induced Cascades

The optimization framework. The catastrophic breakdown can be mitigated if timely protections are provided to stop the infection's spread. In our model, this is implemented by assigning a non-zero protection probability $\gamma_i(t)$ to node i, after which it is immune from infection from layer a. To minimize the size of final failures, it would be more effective to take into account the spreading processes in both layers a and b when deciding which nodes to prioritize for protection.

Here, we develop mitigation strategies by solving the following constrained optimization problems

$$\min_\gamma \; \mathcal{O}(\gamma) := \rho_F(T) = \frac{1}{N}\sum_{i \in V_b} P_F^i(T), \quad (44)$$

$$\text{s. t.} \; 0 \le \gamma_i(t) \le 1 \quad \forall i, t, \quad (45)$$

$$\sum_{i \in V_b}\sum_{t=0}^{T-1} \gamma_i(t) \le \gamma^{\text{tot}}, \quad (46)$$

where the constraint in Eq. (45) ensures that $\gamma_i(t)$ is a probability, and Eq. (46) represents the global budget constraint on the protection resources. As the objective function $\mathcal{O}(\gamma)$ (the size of final failures) depends on the evolution of the two-layer spreading processes, the optimization problem is challenging. Lokhov and Saad introduced the optimal control framework to tackle similar problems, by estimating the marginal probabilities of individuals with the DMP methods[28]. The success of the optimal control approach highlights another advantage of the theoretical methods over numerical simulations[14,28,48].

In this work, we adopt a similar strategy to solve the optimization problem defined in Eqs. (44)–(46), where $P_F^i(T)$ is estimated by the DMP equations derived in Methods. As the expressions of the DMP equations have been explicitly given and only involve elementary arithmetic

operations, we leverage tools of automatic differentiation to compute the gradient of the objective function $\nabla_\gamma \mathcal{O}(\gamma)$ in a back-propagation fashion[49]. It allows us to derive gradient-based algorithms for solving the optimization problem. We remark that such a back-propagation algorithm is equivalent to optimal control with gradient descent update on the control parameters[50]. To save computing time, we adopt the approximation made in the subsection "Simplification under Small Inter-layer Overlap" in Methods for conducting the optimization; but we always use the full DMP formulations developed in the subsection "Full Node-level DMP Equations" in Methods for the evaluation of the outcomes. This is particularly suitable for networks having little inter-layer overlaps. In scenarios where significant inter-layer overlaps exist and precision is crucial, it is always possible to resort to the complete version of the DMP equations.

To handle the box constraint in Eq. (45), we adopt the mirror descent method, which performs the gradient-based update in the dual (or mirror) space rather than the primal space where $\{\gamma_i(t)\}$ live[51,52]. In our case, we use the logit function $\Psi(x) = \log(\frac{x}{1-x})$ to map the primal control variable $\gamma_i(t)$ to the dual space as $h_i(t) = \psi(\gamma_i(t)) \in \mathbb{R}$, where the gradient descent updates are performed. The primal variable can be recovered through the inverse mapping of $\Psi(\cdot)$, which is $\Psi^{-1}(h) = \frac{1}{1+\exp(-h)}$. The elementary mirror descent update step is

$$g^n \leftarrow \nabla_\gamma \mathcal{O}(\gamma^n), \quad (47)$$

$$\gamma^{n+1} \leftarrow \Psi^{-1}\left(\Psi(\gamma^n) - sg^n\right), \quad (48)$$

where n is an index keeping track of the optimization process and s is the step size of the gradient update.

In general, the above optimization process tends to increase the total resources $\sum_{i,t}\gamma_i(t)$. To prevent the violation of the constraint in Eq. (46) during the updates, we suppress the gradient component which increases the total resources when $\sum_{i,t}\gamma_i(t) \ge (1-\epsilon)\gamma^{\text{tot}}$, by shifting the gradient g^n in Eq. (48) with a magnitude b^n

$$b^n \leftarrow \frac{\sum_{t,i}\gamma_i^n(t)(1-\gamma_i^n(t))\frac{\partial}{\partial \gamma_i(t)}\mathcal{O}(\gamma^n)}{\sum_{t,i}\gamma_i^n(t)(1-\gamma_i^n(t))}, \quad (49)$$

$$g^n \leftarrow \nabla_\gamma \mathcal{O}(\gamma^n) - b^n. \quad (50)$$

The rationale for the choice of b^n is explained in Supplementary Note 4. In our implementation of the algorithm, we choose $\epsilon = 0.02$. Even though the shifted gradient method is used, it does not strictly forbid the violation of the constraint in Eq. (46). If the resource capacity constraint is violated, we project the control variables to the feasible region through the simple rescaling

$$\gamma^n \leftarrow \frac{\gamma^{\text{tot}}}{\sum_{t,i}\gamma_i^n(t)}\gamma^n. \quad (51)$$

Finally, the resource capacity constraint Eq. (46) implies that a γ^{tot} amount of protection resources can be distributed in different time steps. In some scenarios, the resources arrive in an online fashion, e.g., a limited number of vaccines can be produced every day. In these cases, there is a resource capacity constraint at each time step. Some results of such a scenario are discussed in Supplementary Note 5.

Case study in a tree network. We first verify the effectiveness of the optimization method by considering a simple problem on a binary tree network of size $N = 63$, where both layers have the same topology. The results are shown in Fig. 5, where three individuals are chosen to be the infected seeds at time $t = 0$, and the outbreak is simulated for $T = 50$ time steps. Without any mitigation strategy, *more than half of the population fail* at the end of the process.

Fig. 5 | Mitigation of the network failures in a binary tree network of size $N = 63$, where both layers have the same topology. Panels **a**–**c** correspond to the case with $\gamma^{\text{tot}} = 5$, while Panels **d**–**f** correspond to the case with $\gamma^{\text{tot}} = 4$. Panels **a** and **d** depict how the final failure size changes during the optimization process. Specifically, the control parameters $\{\gamma_i^n(t)\}$ for each optimization step n were recorded, which were fed to the DMP equations for computing $\rho_F(T)$ at step n. Panels **b** and **e** plot the histogram of the optimal decision variables $\{\gamma_i^*(t)\}$. Panels **c** and **f** show the optimal placement of resources on layer a, where green square nodes receive protection (having a high $\gamma_i^*(t)$ at time $t = 0$). The three red triangle nodes are the initially infected individuals. The system parameters are set as $\beta_{ji} = 0.5$, $\mu_i = 0.5$, $b_{ji} = 1$, $\Theta_i = 0.6|\partial_i^b|$.
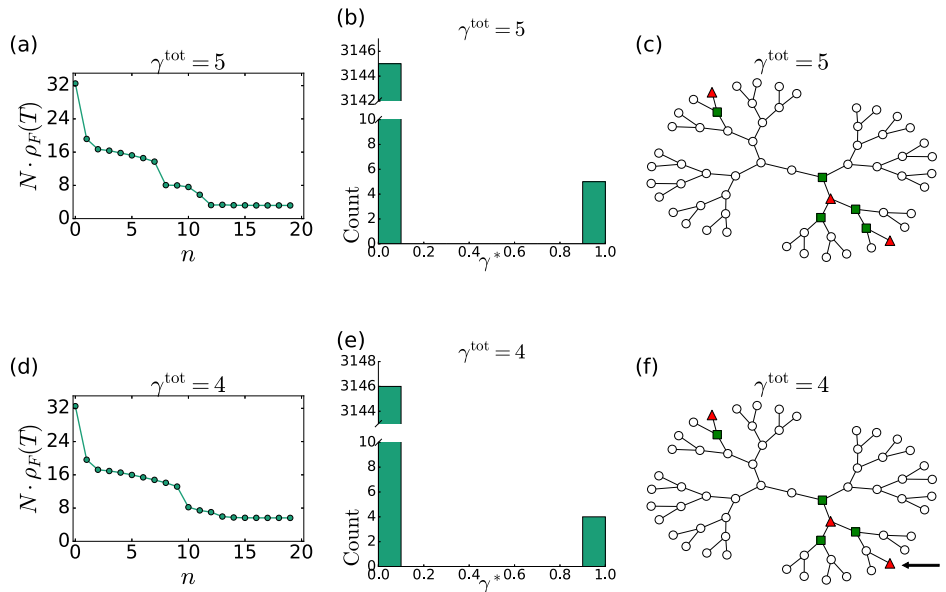
We then protect some vital nodes to mitigate the system failure, by using the optimization method proposed above. In Fig. 5a–c, we restrict the total resources to be $\gamma^{\text{tot}} = 5$. Fig. 5a shows that the optimization algorithm successfully reduces the final failure rate, which demonstrates the effectiveness of the method. We found that the optimal protection resource distribution $\{\gamma_i^*(t)\}$ mostly concentrates on a few nodes at a certain time step (as shown in Fig. 5b), which implies that we can confidently select which nodes to protect. All the nodes with high $\gamma_i^*(t)$ receive protection at time $t = 0$, which implies that the best mitigation strategy in this example is to distribute all γ^{tot} resources as early as possible to stop the infection spread. Figure 5c shows the optimal placement of resources, which can completely block the infection spread, hence minimizing the network failure. In this example, both layers a and b have the same network structure, which is depicted in Fig. 5c.

Similar phenomena are observed in the case with $\gamma^{\text{tot}} = 4$ as shown in Fig. 5d–f, except that the protections are not sufficient to completely block the infection spread. The optimization algorithm sacrifices only two nodes in the vicinity of the infected node in the lower right corner of Fig. 5f (indicated by a black arrow), leaving other parts of the network in the normal state.

In Fig. 6, we further examine the influence of the total resource availability, i.e., γ^{tot}, on the final failure size $N \cdot \rho_F(T)$ determined at the optimal solution $\gamma_i^*(t)$. It is observed that when γ^{tot} increases, the failure size (at the optimum) firstly decreases monotonically, and then saturate when γ^{tot} reaches a certain value such that there are enough protection resources to completely block the infection transmission. Another interesting observation is that for the cases with more initially infected seeds, introducing additional units of protection resource yields a less effective reduction in failure size compared to the cases with fewer initial infected seeds.

The good performance of the optimization is based on the fact that there are enough protection resources (i.e., having a large γ^{tot}) as well as being aware of the origins of the outbreak. In some cases, whether a node was infected at the initial time is not fully determined but follows a probability distribution. Such cases can be easily accommodated in the DMP framework which is intrinsically probabilistic. We investigated such a scenario with probabilistic seeding in Supplementary Note 6, and found that the optimization method can still effectively reduce the sizes of network failures.

Case study in a synthetic network. To further showcase the applicability of the optimization algorithm for failure mitigation, we consider a
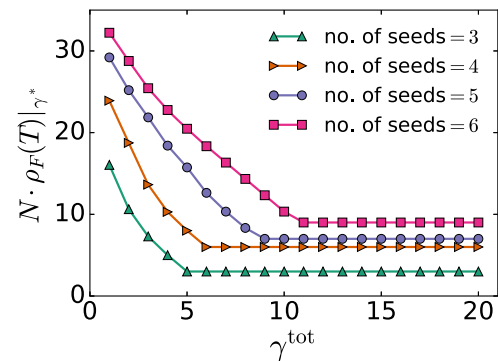

Fig. 6 | Final failure size $N \cdot \rho_F(T)$ of a binary tree network evaluated at the optimal solution $\{\gamma_i^*(t)\}$, as a function of the amount of total resources γ^{tot}. Different curves correspond to different number of initially infected seeds. The network topology and the system parameters are the same as those in Fig. 5.

synthetic technological multiplex network where layer a represents a communication network and layer b represents a power network. We consider the scenario that the communication network can be attacked by malware but can also be protected by technicians, which is modeled by the proposed SIRP model. The infection of a node in the communication network causes the breakdown of the corresponding node in the power network. The breakdown of components in a power network can trigger further failures and form a cascade, which is modeled by the proposed LTM model. We have neglected the details of the power flow dynamics in order to obtain a tractable model and an insightful simple example.

Here, we extract the network topology from the IEEE 118-bus test case to form layer b[53], which has $N = 118$ nodes. We then obtain layer a by rewiring a regular graph of the same size with degree $K = 4$ using a rewiring probability $p_{\text{rewire}} = 0.3$, which creates a Watts-Strogatz small-world network and mimics the topology of communication networks[54]. The resulting multiplex network is plotted in Fig. 7a.

As the failures in layer b are initially induced by the infections in layer a, one may wonder whether deploying the protection resources by minimizing the size of infections, i.e., minimizing $\rho_I(T) + \rho_R(T)$ instead of minimizing $\rho_F(T)$, is already sufficient to mitigate the final failures. To investigate this effect, we replace the objective function in Eq. (44) by $\mathcal{O}^a(\gamma) = \rho_I(T) + \rho_R(T)$ and solve the optimization problem using the same techniques in the subsection "The Optimization Framework". The result is shown in Fig. 7b,
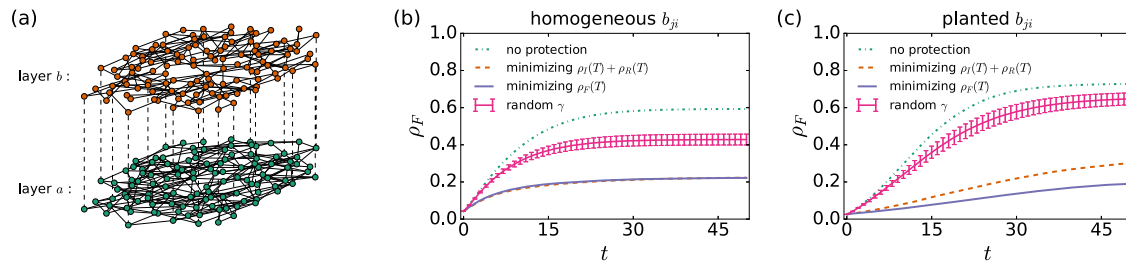
Fig. 7 | A synthetic two-layer network and the evolution of its failure rate. a The structure of the two-layer network, where each layer has $N = 118$ nodes. Layer a is a Watts–Strogatz small-world network, which mimics the topology of communication networks; it is obtained by rewiring a regular graph of degree 4 with rewiring probability $p_{rewire} = 0.3$. Layer b is a power network extracted from the IEEE 118-bus test case. **b** Evolution of the failure rate $\rho_F(t)$ under various mitigation strategies under homogeneous $\{b_{ji}\}$. The curve labeled by "random γ" corresponds to the random deployment of a γ^{tot} amount of protection resources at time $t = 0$; 20 different random realizations are considered and the error bar indicates one standard deviation of the sample fluctuations. The time window is set as $T = 50$. Most system parameters are homogeneous with $\beta_{ji} = 0.2, \mu = 0.5, b_{ji} = 1$, while $\Theta_i = 0.6|\partial_i^b|$. Five nodes are randomly chosen as the initially infected individuals, and $\gamma^{tot} = 10$ is considered. **c** Evolution of the failure rate $\rho_F(t)$ under various mitigation strategies under planted $\{b_{ji}\}$. The system parameters are $\beta_{ji} = 0.17, \mu = 0.5, \Theta_i = 0.6|\partial_i^b|$. Planted influence parameters $\{b_{ji}\}$ are considered. Three nodes are randomly chosen as the initially infected individuals, and $\gamma^{tot} = 9$ is considered. Other experiment set-ups are identical to those in Panel **b**.

which suggests that blocking the infection is as good as minimizing the original objective function in Eq. (44) for the purpose of minimizing the total failure size. Minimizing either objective function constitutes a much better improvement over the random deployment of the same amount of protection resources in this case.

The results in Fig. 7b point to the conventional wisdom that one should try best to stop the epidemic or malware spread (in layer a) for mitigating system failure. The situation will be different if there are vital components in layer b, which should be protected to prevent the failure cascade. This is typically manifested in the heterogeneity of the network connectivity or the system parameters. To showcase this effect, we manually plant a vulnerable connected cluster in layer b by setting the influence parameters b_{ji} for an edge (i, j) in this cluster as $b_{ji} > \Theta_i$, so that the failure of node j itself is already sufficient to trigger the failure of node i. Such a set-up is relevant for commercial, industrial and engineering networks, among others; e.g., supply chain networks evolve to enhance their throughput and efficiency but may operate with little redundancy and low robustness. In this case, we found that minimizing $\rho_F(T)$ yields a much better improvement over minimizing $\rho_I(T) + \rho_R(T)$ for the purpose of mitigating the system failure, as shown in Fig. 7c.

Case studies in a real-world social networks. Lastly, we examine the Kapferer's tailor shop network, a well-known social network dataset gathered by B. Kapferer in Zambia, documenting interactions among workers in a tailor shop[55,56]. This dataset records two types of interactions across two different time frames. The first interaction type is termed "sociational", which encapsulates friendship and socioemotional relationship among the workers. The second interaction type is termed "instrumental", which reflects work- and assistance-related connections among them. For our analysis, the "sociational" network observed in the initial time frame is assigned to layer a, acting as the substrate for infection transmission, while the corresponding "instrumental" network is assigned to layer b, where the failure (in terms of work accomplishment) of a node can be triggered by the malfunctioning of its neighboring nodes. These networks are treated as undirected graphs for simplicity. The resulting two-layer network is depicted in Fig. 8a.

We assign homogeneous values to the majority of system parameters without deliberately introducing any vulnerable component in the network; the set-up closely aligns with the scenario depicted in Fig. 7b and presents a stark contrast to the scenario in Fig. 7c. We select five nodes that possess the highest degrees to serve as the initially infected individuals, which can be viewed as super-spreaders in the network. We then protect the vital nodes to mitigate the system failures by using the optimization method as above, where the result is shown in Fig. 8b. Interestingly, minimizing the size of failures (i.e., $\rho_F(T)$) is evidently better than minimizing the size of infections
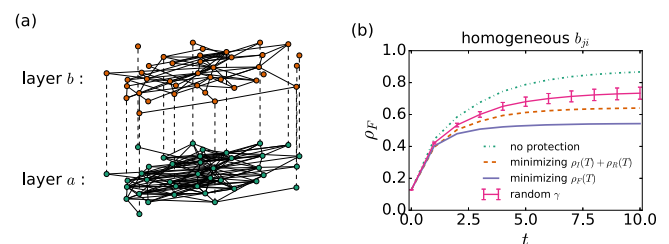

Fig. 8 | The Kapferer's tailor shop network and the evolution of its failure rate. a The structure of the Kapferer's tailor shop network, which involves interactions among 39 workers in a tailor shop in Zambia during a period of one month. Layer a represents the "sociational" relations, while layer b represents the "instrumental" relations. **b** Evolution of the failure rate $\rho_F(t)$ of the tailor shop network up to time $T = 10$ under various mitigation strategies. The "random γ" strategy has the same set-up as the one in Fig. 7b; 20 different random realizations are considered and the error bar indicates one standard deviation of the sample fluctuations.. Most system parameters are homogeneous with $\beta_{ji} = 0.07, \mu = 0.5, b_{ji} = 1$, while $\Theta_i = 0.6|\partial_i^b|$. Five nodes with the highest degrees in layer a are selected as the initially infected individuals, and $\gamma^{tot} = 10$ is considered.

(i.e., $\rho_I(T) + \rho_F(T)$) for the purpose of failure mitigation. It suggests that in this realistic and natural scenario, simply blocking the infection transmission is sub-optimal and one needs to take a holistic view of the two-layer model for optimizing the network's utility.

Conclusion
We investigate the nature of a type of two-layer spreading processes in unidirectionally dependent networks, comprising two interacting layers a and b. Disease or malware spreads in layer a, which can trigger cascading failures in layer b, leading to secondary disasters. The spreading processes in the two layers are modeled by the SIRP and LTM models, respectively. To tackle the complex stochastic dynamics in the two-layer networks, we utilized the dynamic message-passing method by working out the dynamic belief propagation equations. The resulting DMP algorithms have low computational complexity in sparse networks and allow us to perform accurate and efficient inference of the system states.

Based on the DMP method, we systematically studied and evaluated the impact of the infection-induced cascading failures. The cascade process in layer b can lead to large-scale network failures, even when the infection rate in layer a remains at a relatively low level. By considering a homogeneous network topology and homogeneous system parameters, we derive the asymptotic and large-size limits of the DMP equations. The asymptotic limit of the two-layer spreading processes corresponds to the coupling between a bond percolation model and a bootstrap percolation model,

which can be analytically solved. The infection outbreak size in layer a changes continuously from zero to non-zero as the infection probability β surpasses a transition point β_c^a, while the failure size in layer b can exhibit a discontinuous jump to the completely failed state when β surpasses another transition point β_c^b under certain conditions. All these results highlight the observation that cascading failure propagation in layer b can drastically amplify the impact of the epidemic outbreaks in layer a, which requires special attention.

Another advantage of the DMP method is that it yields a set of closed-form equations, which can be very useful for other downstream analyses and tasks. We exploited this property to devise optimization algorithms for mitigating network failure. The optimization method works by back-propagating the impact at the final time to adjust the control parameters (i.e., the protection probabilities). The mirror descent method and a heuristic gradient shift method were also used to handle the constraints on the control parameter. We show that the resulting algorithm can effectively minimize the size of system failures. We believe that our dedicated analyses provide valuable insights and a deeper understanding of the impact the infection-induced cascading failures on networks, and the obtained optimization algorithms will be useful for practical applications in systems of this kind.

Data availability

Datasets cited in this study are publicly accessible and have been referenced accordingly in the manuscript.

Code availability

Source codes of the methods and analyses used in this study are available at https://github.com/boli8/DMP-for-SIRP-LTM.

References

1. Pak, A. et al. Economic consequences of the covid-19 outbreak: the need for epidemic preparedness. *Front. Public Health* **8**, https://www.frontiersin.org/articles/10.3389/fpubh.2020.00241 (2020).
2. Chaturvedi, K., Vishwakarma, D. K. & Singh, N. Covid-19 and its impact on education, social life and mental health of students: A survey. *Children Youth Serv. Rev.* **121**, 105866 (2021).
3. Cochran, A. L. Impacts of covid-19 on access to transportation for people with disabilities. *Transp. Res. Interdiscipl. Perspect.* **8**, 100263 (2020).
4. Xu, Z., Elomri, A., Kerbache, L. & El Omri, A. Impacts of covid-19 on global supply chains: Facts and perspectives. *IEEE Eng. Manag. Rev.* **48**, 153–166 (2020).
5. Aday, S. & Aday, M. S. Impact of COVID-19 on the food supply chain. *Food Qual. Safety* **4**, 167–180 (2020).
6. Amini, M. H., Arasteh, H. & Siano, P. *Sustainable Smart Cities Through the Lens of Complex Interdependent Infrastructures: Panorama and State-of-the-art*, 45–68 (Springer International Publishing, Cham, 2019). https://doi.org/10.1007/978-3-319-98923-5_3.
7. Liu, X., Chen, B., Chen, C. & Jin, D. Electric power grid resilience with interdependencies between power and communication networks - a review. *IET Smart Grid* **3**, 182–193 (2020).
8. Guo, H., Zheng, C., Iu, H. H.-C. & Fernando, T. A critical review of cascading failure analysis and modeling of power system. *Renew. Sustain. Energy Rev.* **80**, 9–22 (2017).
9. Castillo-Chavez, C., Huang, W. & Li, J. Competitive exclusion in gonorrhea models and other sexually transmitted diseases. *SIAM J. Appl. Math.* **56**, 494–508 (1996).
10. Castillo-Chavez, C., Huang, W. & Li, J. Competitive exclusion and coexistence of multiple strains in an sis std model. *SIAM J. Appl. Math.* **59**, 1790–1811 (1999).
11. Karrer, B. & Newman, M. E. J. Competing epidemics on complex networks. *Phys. Rev. E* **84**, 036106 (2011).
12. Cai, W., Chen, L., Ghanbarnejad, F. & Grassberger, P. Avalanche outbreaks emerging in cooperative contagions. *Nat. Phys.* **11**, 936–940 (2015).
13. Wang, W., Liu, Q.-H., Liang, J., Hu, Y. & Zhou, T. Coevolution spreading in complex networks. *Phys. Rep.* **820**, 1–51 (2019).
14. Sun, H., Saad, D. & Lokhov, A. Y. Competition, collaboration, and optimization in multiple interacting spreading processes. *Phys. Rev. X* **11**, 011048 (2021).
15. Liu, J. et al. Analysis and control of a continuous-time bi-virus model. *IEEE Trans. Automatic Control* **64**, 4891–4906 (2019).
16. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
17. Bashan, A., Berezin, Y., Buldyrev, S. V. & Havlin, S. The extreme vulnerability of interdependent spatially embedded networks. *Nat. Phys.* **9**, 667–672 (2013).
18. Valdez, L. D. et al. Cascading failures in complex networks. *J. Complex Netw.* **8**, cnaa013 (2020).
19. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
20. Adam, D. Special report: The simulations driving the world's response to COVID-19. *Nature* **580**, 316–318 (2020).
21. Wang, W., Tang, M., Stanley, H. E. & Braunstein, L. A. Unification of theoretical approaches for epidemic spreading on complex networks. *Rep. Progr. Phys.* **80**, 036603 (2017).
22. Karrer, B. & Newman, M. E. J. Message passing approach for general epidemic models. *Phys. Rev. E* **82**, 016101 (2010).
23. Lokhov, A. Y., Mézard, M., Ohta, H. & Zdeborová, L. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Phys. Rev. E* **90**, 012801 (2014).
24. Lokhov, A. Y., Mézard, M. & Zdeborová, L. Dynamic message-passing equations for models with unidirectional dynamics. *Phys. Rev. E* **91**, 012811 (2015).
25. Koher, A., Lentz, H. H. K., Gleeson, J. P. & Hövel, P. Contact-based model for epidemic spreading on temporal networks. *Phys. Rev. X* **9**, 031017 (2019).
26. Li, B. & Saad, D. Impact of presymptomatic transmission on epidemic spreading in contact networks: A dynamic message-passing analysis. *Phys. Rev. E* **103**, 052303 (2021).
27. Lokhov, A. Reconstructing parameters of spreading models from partial observations. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I. & Garnett, R. (eds.) *Proceedings of the 30th International Conference on Neural Information Processing Systems*, vol. 29, 3467 – 3475 (Curran Associates Inc., 2016).
28. Lokhov, A. Y. & Saad, D. Optimal deployment of resources for maximizing impact in spreading processes. *Proc. Natl Acad. Sci.* **114**, E8138–E8146 (2017).
29. Boccaletti, S. et al. The structure and dynamics of multilayer networks. *Phys. Rep.* **544**, 1–122 (2014).
30. Balcan, D. et al. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *J. Comput. Sci.* **1**, 132–145 (2010).
31. Garetto, M., Gong, W. & Towsley, D. Modeling malware spreading dynamics. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, 3, 1869–1879 (IEEE, 2003).
32. Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl Acad. Sci.* **99**, 5766–5771 (2002).
33. Kempe, D., Kleinberg, J. & Tardos, É. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data*

Mining, KDD '03, 137–146 (Association for Computing Machinery, 2003). https://doi.org/10.1145/956750.956769.

34. Motter, A. E. & Lai, Y.-C. Cascade-based attacks on complex networks. *Phys. Rev. E* **66**, 065102 (2002).

35. Carreras, B. A., Lynch, V. E., Dobson, I. & Newman, D. E. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos* **12**, 985–994 (2002).

36. Crucitti, P., Latora, V. & Marchiori, M. Model for cascading failures in complex networks. *Phys. Rev. E* **69**, 045104 (2004).

37. Su, Z. & Kurths, J. A dynamic message-passing approach for social contagion in time-varying multiplex networks. *Europhys. Lett.* **123**, 68004 (2018).

38. Altarelli, F., Braunstein, A., Dall'Asta, L. & Zecchina, R. Large deviations of cascade processes on graphs. *Phys. Rev. E* **87**, 062115 (2013).

39. Mézard, M. & Montanari, A. *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009). https://doi.org/10.1093/acprof:oso/9780198570837.001.0001.

40. Torrisi, G., Annibale, A. & Kühn, R. Overcoming the complexity barrier of the dynamic message-passing method in networks with fat-tailed degree distributions. *Phys. Rev. E* **104**, 045313 (2021).

41. Parshani, R., Rozenblat, C., Ietri, D., Ducruet, C. & Havlin, S. Inter-similarity between coupled networks. *Europhys. Lett.* **92**, 68002 (2011).

42. Cellai, D., López, E., Zhou, J., Gleeson, J. P. & Bianconi, G. Percolation in multiplex networks with overlap. *Phys. Rev. E* **88**, 052811 (2013).

43. Shrestha, M., Scarpino, S. V. & Moore, C. Message-passing approach for recurrent-state epidemic models on networks. *Phys. Rev. E* **92**, 022821 (2015).

44. Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. *Math. Biosci.* **63**, 157–172 (1983).

45. Li, M. et al. Percolation on complex networks: Theory and application. *Phys. Rep.* **907**, 1–68 (2021).

46. Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).

47. Chalupa, J., Leath, P. L. & Reich, G. R. Bootstrap percolation on a bethe lattice. *J. Phys. C* **12**, L31 (1979).

48. Zhou, J., Zhao, Y. & Ye, Y. Complex dynamics and control strategies of SEIR heterogeneous network model with saturated treatment. *Phys. A* **608**, 128287 (2022).

49. Baydin, A. G., Pearlmutter, B. A., Radul, A. A. & Siskind, J. M. Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.* **18**, 5595–5637 (2017).

50. Li, Q., Chen, L., Tai, C. & E, W. Maximum principle based algorithms for deep learning. *J. Mach. Learn. Res.* **18**, 1–29 (2018).

51. Nemirovski, A. & Yudin, D. *Problem Complexity and Method Efficiency in Optimization* (Wiley, 1983).

52. Beck, A. & Teboulle, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* **31**, 167–175 (2003).

53. Christie, R. Power systems test case archive, university of washington. Available at: https://labs.ece.uw.edu/pstca/pf118/pg_tca118bus.htm (1993)

54. Cai, Y., Li, Y., Cao, Y., Li, W. & Zeng, X. Modeling and impact analysis of interdependent characteristics on cascading failures in smart grids. *Int. J. Electrical Power Energy Syst.* **89**, 106–114 (2017).

55. Kapferer, B. *Strategy and Transaction in an African Factory* (Manchester University Press, Manchester, 1972).

56. Kapferer tailor shop data set. Available at the UCI Network Data Repository: https://networkdata.ics.uci.edu/netdata/html/kaptail.html (1972).

Acknowledgements

Author contributions

B.L. and D.S. conceived the project and developed the theoretical framework. B.L. carried out the theoretical calculations and the numerical simulations. B.L. and D.S. discussed the results and prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-024-01638-1.

Correspondence and requests for materials should be addressed to Bo Li or David Saad.

Peer review information *Communications Physics* thanks Lenka Zdeborova, Louis Shekhtman, Davide Ghio and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions 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.