SCIENTIFIC REPORTS

OPEN

Modelling indirect interactions during failure spreading in a project activity network

Christos Ellinas

Spreading broadly refers to the notion of an entity propagating throughout a networked system via its interacting components. Evidence of its ubiquity and severity can be seen in a range of phenomena, from disease epidemics to financial systemic risk. In order to understand the dynamics of these critical phenomena, computational models map the probability of propagation as a function of direct exposure, typically in the form of pairwise interactions between components. By doing so, the important role of indirect interactions remains unexplored. In response, we develop a simple model that accounts for the effect of both direct and subsequent exposure, which we deploy in the novel context of failure propagation within a real-world engineering project. We show that subsequent exposure has a significant effect in key aspects, including the: (a) final spreading event size, (b) propagation rate, and (c) spreading event structure. In addition, we demonstrate the existence of 'hidden influentials' in large-scale spreading events, and evaluate the role of direct and subsequent exposure in their emergence. Given the evidence of the importance of subsequent exposure, our findings offer new insight on particular aspects that need to be included when modelling network dynamics in general, and spreading processes specifically.

Recent years have witnessed a flurry of work on spreading processes[1,2], ranging from empirical expositions on the impact of such spreading (e.g. existence of large-scale spreading events[3–5], properties of 'super-spreaders')[6–10] to methodological developments that map the underlying dynamics (e.g. spreading mechanisms[11–14], modelling frameworks)[15–19]. Pairwise interactions are the perceived centrepiece in understanding the evolution of spreading processes[20], since they capture the *direct exposure* of each node to the prospect of switching its state, for example from 'non-affected' to 'affected'. Importantly, the impact of direct exposures can be supplemented by *subsequent* exposures which may arise due to global or local network features. Past work has focused on the impact of these indirect effects by considering the *global* topology of the network (e.g. distribution of shortest paths)[21,22] and how it is influenced by particular mechanisms (e.g. flow redistribution)[23,24]. However, such indirect effects can also arise from *local*, non-trivial structures (e.g. particular network motifs such as the feed-forward loop)[25,26]. Despite the intuitive importance of these local structures[26], little attention has been paid in evaluating their impact to the overall spreading process, largely due to the particularities of the spreading models typically deployed to study these processes.

In particular, spreading models can be classified to two broad categories[12,27], depending on the incorporated mechanisms: (i) epidemiological models, where the spreading process is viewed as an *independent* event across different pairs of nodes (e.g. the probability of node *i* to affect its neighbour *j* is independent of other interactions), and (ii) sociological models, where the spreading process is viewed as an *interdependent* event (e.g. the probability of node *i* to affect its neighbour *j* depends on the state of node's *j* neighbours). Despite this distinction, both model categories are grounded on the same fundamental premise, in which direct exposure is the principal factor that determines the dynamics of the spread. Yet, the implication of this premise can be non-trivial, as subsequent exposures can interfere with the spreading process and affect key outcomes, even in simple examples like the one discussed below.

Consider the toy network in Fig. 1, where each node can switch states irreversibly from 'non-affected' to 'affected' with a given probability *p*. In the case of an epidemiological spreading model, this premise corresponds to the widely used 'Susceptible'-'Infected' model[28], where nodes and links correspond to individuals and infection pathways (e.g. social interactions), respectively. In this case, the ability of node *i* to affect its immediate neighbours

Engineering Mathematics, University of Bristol, Bristol, UK. Correspondence and requests for materials should be addressed to C.E. (email: ce12183@bristol.ac.uk)
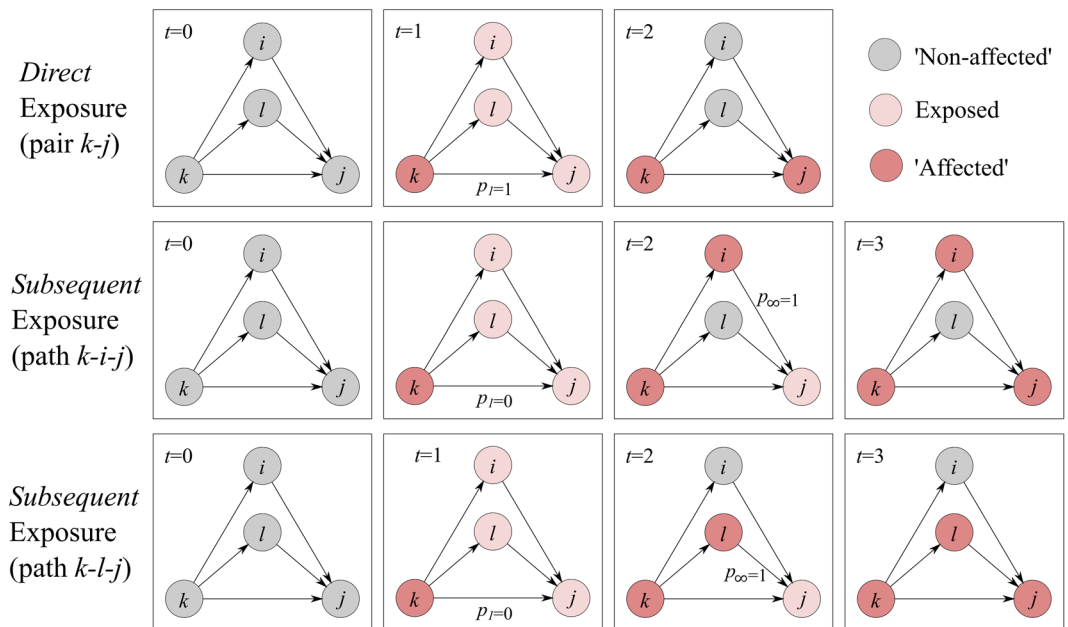
Figure 1. Example to highlight the distinction between the direct and subsequent exposure of node *j* to the failure of node *k* (time runs from left to right). Top panel focuses on the direct case, where node *j* is directly exposed to node *k*'s failure; rest illustrate the case where node *j* is subsequently exposed, via node *i* (middle panel) and node *l* (bottom panel).

is assessed independently across all possible pairs (pair $k \rightarrow i$; pair $k \rightarrow l$; pair $k \rightarrow j$). At this point, node *j* is directly exposed to the infection of node *k*, through the directed link from node *k* to node *j*, and indirectly, through paths $k \rightarrow i \rightarrow j$ and $k \rightarrow l \rightarrow j$ (see Methods). As a result, the possibility for node *j* to switch state is evaluated at three distinct points during the evolution of the spread (one direct link; two indirect paths), compared to the single evaluation that takes place in the case of node *i* and node *l* (due to their direct link). Hence, node *j* is three times more likely to switch state, compared to node *i* and *l*, despite that fact that the probability of them changing state is uniformly set. We can trace this effect to the implicit assumption that direct and subsequent exposures are of equal importance, which is appropriate when dealing with a typical disease spreading.

Fleshing out this assumption, the probability of node *j* to switch states, either due to its contact with node *k*, or due to its contact with node *k* (as part of path $k \rightarrow i \rightarrow j$), is exactly the same. This suggests that the underlying pathogen that drives the spread has remained unchanged (and therefore, the probability for node *k* to infect its neighbours, including node *j*, is exactly the same as the probability of node *i* to infect its own neighbours, including node *j*). However, consider an alternative context where the pathogen is replaced by a defect, which spreads across a network of activities, where nodes correspond to technical activities[29] (e.g. specify the design of an engineering component) and links to functional dependencies (e.g. activity *j* can start if, and only if, its predecessor task *i* finishes; i.e. manufacturing an engineering component can start only if its specifications have first been specified)[30,31]. In this case, the interpretation of direct and subsequent exposures is distinct, where activity *j* may be particularly susceptible (or immune) to the particular defect that caused activity *i* to fail. For example, consider the case where activity *k* corresponds to 'specify component', activity *i* to 'manufacture component' and activity *j* to 'install component'. In the case where activity *k* fails, the probability of affecting activity *j* may be low (e.g. due to the availability of prefabricated, standardised components). Yet, if activity *i* is affected by the failure of activity *k*, the probability of activity *j* to be affected is now higher – in this case the defect propagating across the activity network has evolved from 'failure to specify' to 'failure to specify and manufacture'. Therefore, the fact that activity *j* is additionally exposed to the defect in an indirect way (through node *i* and node *l*; Fig. 1) directly interferes with the likelihood of activity *j* changing its state; yet typical spreading models – like the aforementioned SI model – would assume that such subsequent exposure does not interfere.

In conjunction with the fact that activity networks have a high concentration of local, non-trivial structures[32] – and hence, experience pronounced levels of subsequent exposure (see Supplementary Fig. 1) - the likelihood of obtaining misleading results through the use of typical spreading models is high. Consequently, a range of decisions that may be driven by those results can be significantly affected, from estimating an appropriately sized contingency budget to developing appropriate (project) risk management framework(s)[31]

In response, we propose a simple spreading model that allows us to map the impact of direct and subsequent exposures (see Methods). It does so by disentangling the probability of a defect to spread (p), to two distinct components:p_1, which controls the probability of node *j* being affected by the failure of its predecessor, node *k* (Fig. 1, top panel) and p_∞, which controls the probability of node *j* being affected by additional, subsequent exposures to the failure of node *k*, through node *i* (Fig. 1, middle panel) or node *l* (Fig. 1, bottom panel). We distinguish between the two by keeping track of node *k*'s successor at time *t*, say node *j*, and assess whether node *j* has been encountered before. If so, then node *j* has been directly exposed to the failure of a *different* predecessor before *t*

(Fig. 1 middle/bottom panel at $t = 2$), and hence the probability of node j failing is controlled by p_∞; if node j is encountered for the first time, the probability of is controlled by p_1 (Fig. 1 middle/bottom panel at $t = 1$).

We formulate two models, M_0 where $p_1 \in [0, 1]$ and $p_\infty = 0$, and M where both $p_1 \in [0, 1]$ and $p_\infty \in [0, 1]$. Hence, any difference between results obtained under M_0 and M reflect the impact of subsequent exposure, with p_∞ controlling the magnitude of the effect. We deploy both model variants M and M_0 to a real-world network of activities (i.e. a *project*)[33,34], and explore two key quantities that characterise spreading – the spreading event size (S_n) and the rate by which spreading propagates (S_r) –where p_∞ has an important effect. We subsequently focus on particular structural features of the pathways used to sustain these spreading events, and specifically on the different impact that P_1 and p_∞ have. Finally, we explore the topological properties of nodes involved in large-scale spreading events. In agreement with recent studies[6,35] on information spreads, we report the presence of 'hidden influentials' (i.e. nodes with average topological properties which play a key role in sustaining large-scale spreading events), with their existence being increasingly pronounced at higher P_1 and/or p_∞ values.

Results

We first establish the importance of subsequent exposure (parameter p_∞) by illustrating its effect on the spreading event size through a comparative analysis between the results of spreading model M and M_0. At this point we provide evidence on the link between subsequent exposure and clustering, where an increase in the latter (clustering) controls the magnitude of the former (subsequent exposure). We subsequently use model M to highlight the contrasting impact that direct (parameter P_1) and subsequent exposures have on propagation rate. We then focus on the relationships between the structural characteristics of these spreading events and their size/rate, and how P_1 and p_∞ affect them. Finally, we provide insight on the topological features of nodes capable of fuelling large-scale spreading events, and how P_1 and p_∞ influence them.

Spreading event size and propagation rate.   We first define the ratio of the largest spreading event size obtained using the M model, over the largest spreading event size obtained under the M_0 model, as $r^{\max} = \frac{M(S_n^{\max})}{M_0(S_n^{\max})}$. In a similar fashion, we denote the ratio of the average spreading event sizes as r^{avg}. With $r^{\max}$ and r^{avg} being functions of both P_1 and p_∞, we can examine their isolated effect by considering their corresponding averages: $\tilde{r}^{\max}(p_1) = \frac{1}{|p_\infty|}\sum_{p_\infty=0}^{1} r^{\max}(p_1, p_\infty)$ and $\tilde{r}^{\max}(p_\infty) = \frac{1}{|p_1|}\sum_{p_1=0}^{1} r^{\max}(p_1, p_\infty)$. By applying the same approach to r^{avg}, we obtain $\tilde{r}^{\mathrm{avg}}(p_1)$ and $\tilde{r}^{\mathrm{avg}}(p_\infty)$.

Increasing parameter p_∞ leads to values of $r^{\max}$ (Fig. 2a) and r^{avg} (Fig. 2b) being significantly higher than 1, demonstrating the augmenting effect that subsequent exposures have on the spreading event size. This result suggests that this activity network contains a high enough number of non-trivial subgraphs – such as the ones included in Fig. 1 – which allows p_∞ to have a significant impact on the progression of the spreading process. If the converse where to be true (i.e. $r^{\max} \cong r^{\mathrm{avg}} \cong 1$), it would suggest that the activity network could be approximated by 'locally tree-like' network, where subsequent exposure would have had no effect over the spreading process, with results resembling that of a tree network (results from a tree network are shown in Supplementary Fig. 2 for reference).

Taken in isolation, P_1 and p_∞ show qualitatively different characteristics in terms of their impact in the spreading process, further highlighting the non-trivial interaction between direct and subsequent exposure in the context of a spreading process. The concave relationship between both $\tilde{r}^{\max}$ and $\tilde{r}^{\mathrm{avg}}$, with respect to P_1 demonstrates the principal role of P_1 in sustaining the spreading process, with both $\tilde{r}^{\max}$ and $\tilde{r}^{\mathrm{avg}}$ converging to 1 at the two extreme ends of P_1 (Fig. 2c). On one hand, if $p_1 = 0$ no spreading occurs and therefore the effect of p_∞ is nullified, with both M and M_0 converging to identical spreading events; when $p_1 = 1$ then direct exposure successfully switches the state of *all* nodes and therefore, no nodes are left for p_∞ to affect. Interestingly, high $\tilde{r}^{\max}$ values are preserved up to relatively high P_1 values ($p_1 \leq 0.7$), indicating the strong influence of p_∞ even under unfavourable conditions i.e. under $p_1 = 0.7$, a node is much more likely to switch state due to a direct rather than subsequent exposure, and therefore one would naturally expect that the influence of p_∞ would be limited, giving rise to a low $\tilde{r}^{\max}$ value. Finally, the intuitive expectation of p_∞ having an ever-increasing effect in terms of both $\tilde{r}^{\max}$ and $\tilde{r}^{\mathrm{avg}}$ is supported by the monotonically increasing trends shown in Fig. 2d.

This insight is consistent when we consider the cumulative probability distribution of S_n, focusing on (i) the probability of observing a spreading event of a given size, and (ii) the magnitude of the largest event. In this case, the augmenting role of p_∞ is particularly pronounced at the tail of the distribution, where higher p_∞ increases both (i) and (ii) (Fig. 3b). This is in contrast to the direct impact of P_1, where highger P_1 increase both (i) and (ii) across the entire range of S_n (Fig. 3a). Taken in conjunction, this behaviour demonstrates the subtle impact of indirect interactions with respect to the emergence of small-scale spreading events – which are largely driven by direct interactions – and their marked influence with respect to large-scale spreading events, both in terms of their probability of their emergence and their absolute size.

The extent by which subsequent exposures control the spreading event size depends on the clustering of the network, as captured by the clustering coefficient (C)[36]. To demonstrate this, we have deployed the Watts-Strogatz model to generate artificial networks whilst varying the rewiring probability β, moving progressively from clustered ($\beta = 0.1$, $C = 0.3$) to random networks ($\beta = 1$, $C = 0.03$). In doing so, we find that an increase in β (and hence, drop in C) decreases the difference between spreading event sizes obtained by M and M_0, for both r^{avg} and $r^{\max}$, with results being qualitatively similar to Fig. 2 (see Fig. 4). This finding suggests that the increased impact that subsequent exposure has to the spreading process in general – and to the activity network in particular – relies, at least partly, to increased clustering.
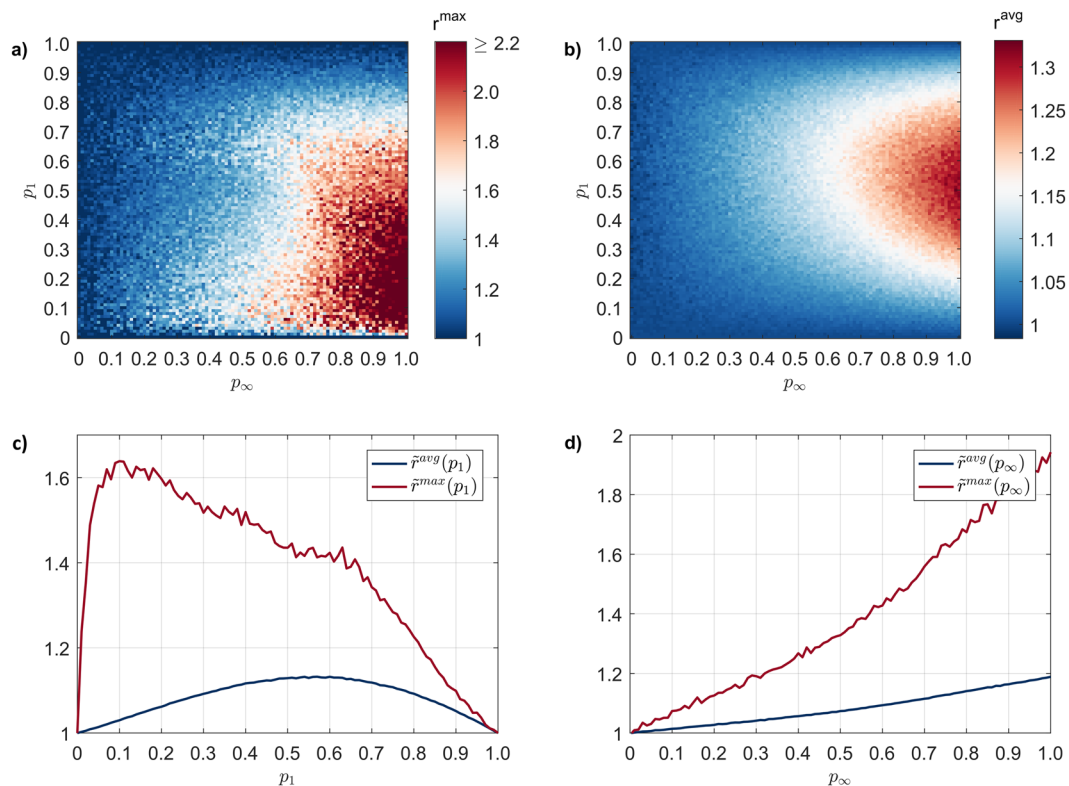
Figure 2. Difference in spreading event size between spreading model M_0 and M (**a**) Ratio of the largest spreading event sizes (r^{max}) under the entire spectrum of p_1 an p_∞ (largest $r^{max} = 3.2$); (**b**) like (**a**), focusing on the ratio of average spreading event size (r^{avg}) under parameter p_1 and p_∞; (**c**) ratio of largest (red) and average (blue) spreading event size, averaged across p_∞, as a function of p_1; (**d**) ratio of largest (red) and average (blue) spreading event size, averaged across p_1, as a function of p_∞.
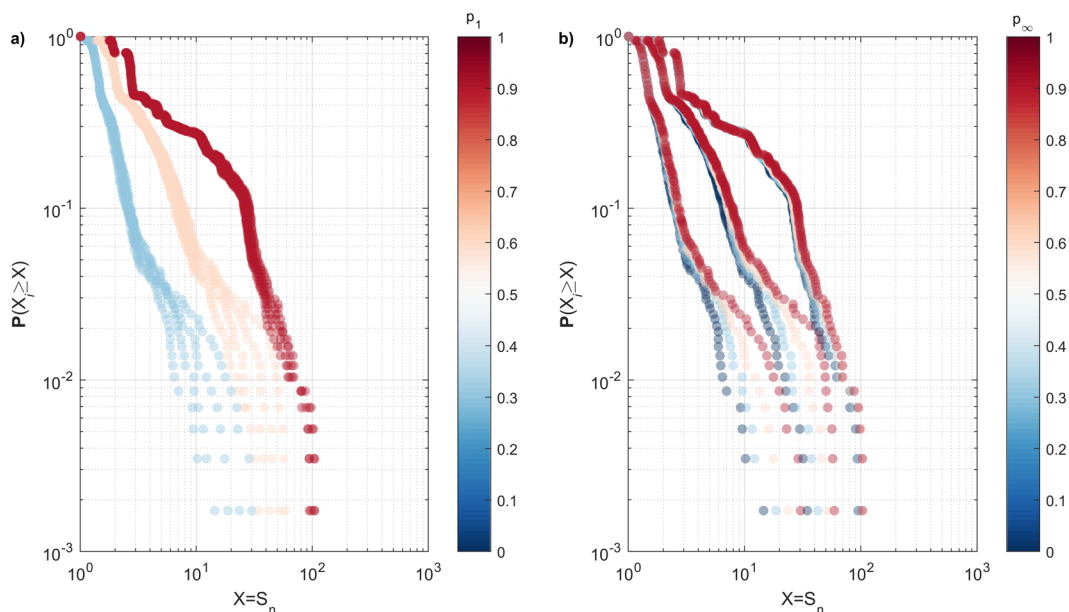

Figure 3. Cumulative probability distribution of spreading event size (S_n), where marker colour corresponds to (**a**) p_1 and (**b**) p_∞. For clarity, p_1 and p_∞ are sampled in 0.3 step intervals, taking the values of [0, 0.3, 0.6, 0.9].

We now focus on the rate by which a given spreading event propagates across the network, which we quantify as $S_r = \frac{S_n}{t}$, where S_n refers to the spreading event size and t refers to the average number of simulation steps needed for all nodes affected (within that spreading event) to switch state from 'non-affected' to 'affected' (which
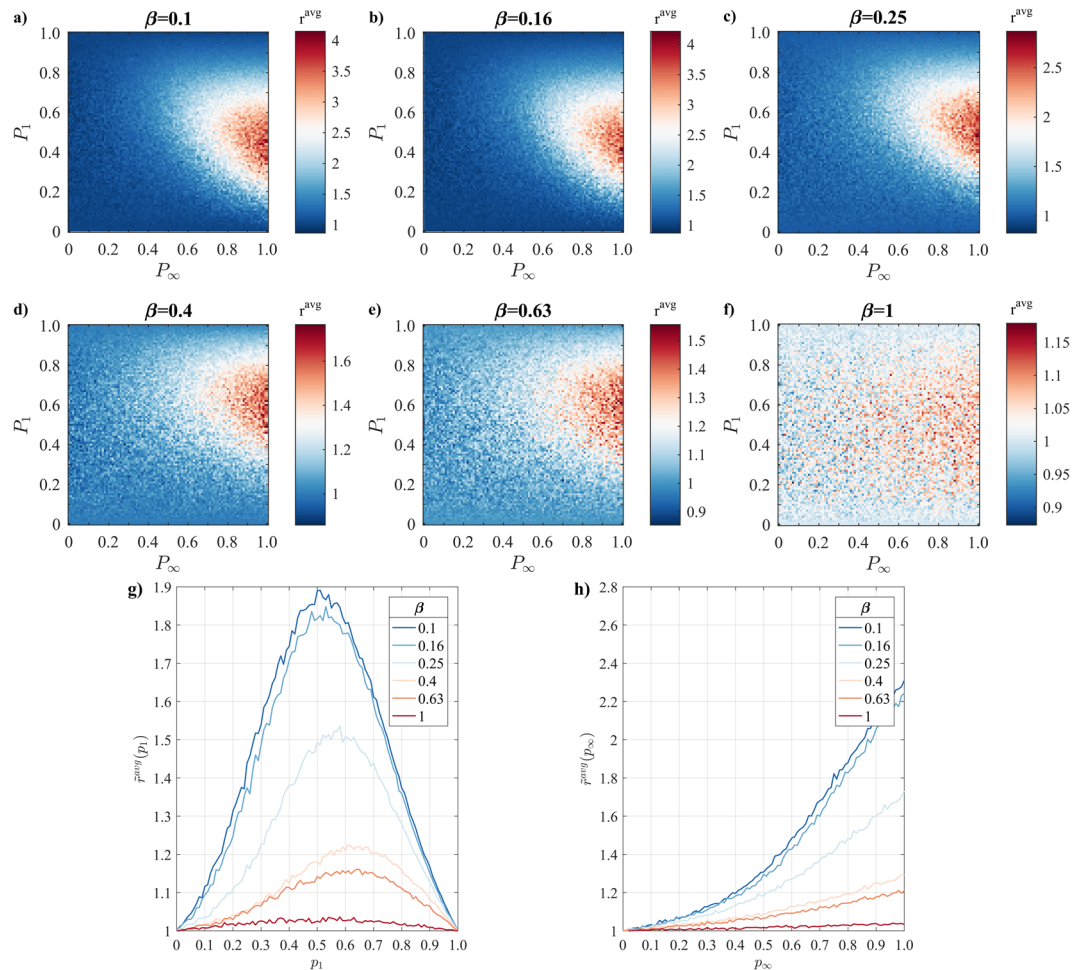
Figure 4. Difference in spreading event size between spreading model M_0 and M for a set of artificial, 'small-world' networks with increasingly probability of rewiring (β), ranging from increasingly clustered (**a**) β = 0.1) to random (**f**) β = 1) topologies; (**g**) ratio of average (blue) spreading event size, averaged across p_∞, as a function of p_1, across the β spectrum; (**h**) ratio of average (blue) spreading event size, averaged across p_1, as a function of p_∞, across the β spectrum.

is a variation of survival probability)[37]. As such, we define the average propagation rate, $\widetilde{S}_r^{\,avg}$ as the propagation rate for each spreading event, averaged across all events, and the maximum propagation rate, $\widetilde{S}_r^{\,max}$, as the propagation rate for the single largest spreading event.

We find that direct and subsequent exposures have the converse effect with respect to the propagation rate, both in terms of $\widetilde{S}_r^{\,avg}$ and $\widetilde{S}_r^{\,max}$. In particular, we find a positive relationship between P_1 and the propagation rate, in terms of both $\widetilde{S}_r^{\,avg}$ (blue marker) and $\widetilde{S}_r^{\,max}$ (red marker), which corresponds to the intuitive expectation where increased direct exposure eases the way in which spreading progresses, enhancing the overall propagation rate (Fig. 5a). However, a negative relationship exists between p_∞ and propagation rate, both in terms of $\widetilde{S}_r^{\,avg}$ and $\widetilde{S}_r^{\,max}$ (Fig. 5b). This is due to the elaborate topology of the pathways deployed by the spreading process, where higher p_∞ increases the likelihood of utilising wider spreading pathways (i.e. high S_w; see *Topology of spreading pathways*) which are more likely to involve a higher number of links to be traversed for affecting the same number of nodes (increasing the denominator of S_r) eventually delaying the overall spreading process. Considering both effects, this behaviour suggests that propagation rate is conflated by contrasting dynamics, where direct exposure provides immediate – and thus, faster – pathways for spreading to propagate, while subsequent exposure unlocks slower pathways which supress the overall in propagation rate (even though they may increase the overall spreading event size, as seen in Fig. 2d).

Topology of spreading pathways.   We characterise the structure of a spreading event by considering the maximum depth and width of the underlying pathways that have sustained it. We define the maximum depth of a spreading event (S_d) as the shortest path between the initial seed node and the farthest node involved in the event[6]. In addition, we define its maximum width (S_w) as the maximum number of nodes affected whilst being at the same distance from the seed node[38]. As such, we characterise the structure of each spreading event as $S_{d/w} = \frac{S_d}{S_w}$, which provides a continuous measure for the overall shape of the underlying pathways, where a high
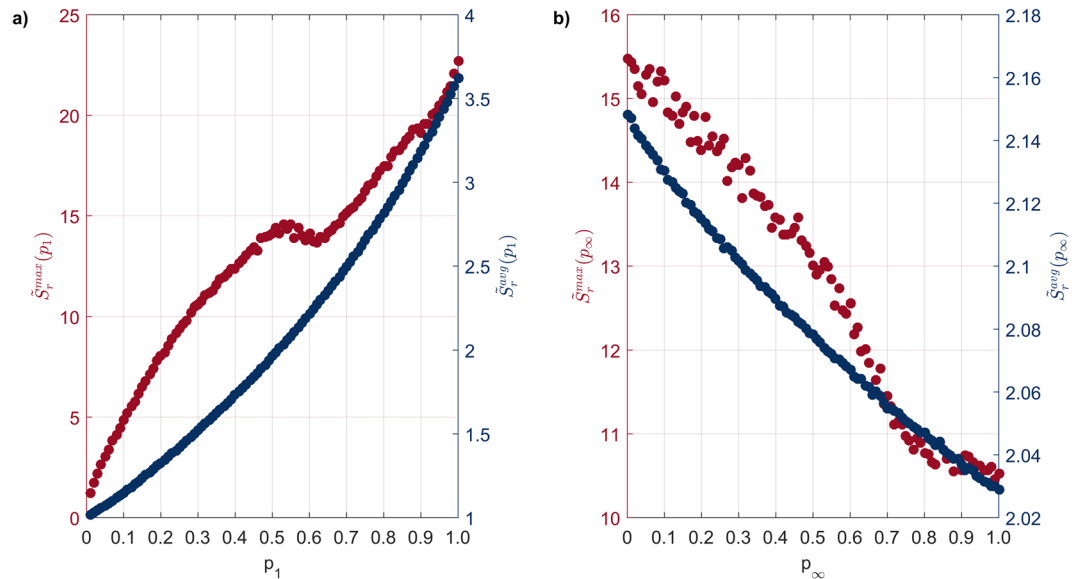
Figure 5. Propagation rate as a function of (**a**) direct exposure, p_1, and (**b**) subsequent exposure, p_∞. Blue and red markers correspond to the average rate across all spreading events, $\tilde{S}_r^{\text{avg}}$, and propagation rate for the largest spreading event, $\tilde{S}_r^{\text{max}}$, respectively.

value of $S_{d/w}$ corresponds to long and narrow pathways, whilst a low value of $S_{d/w}$ corresponds to short and wide pathways (see Supplementary Fig. 3 for examples). In that way, $S_{d/w}$ is maximised when the spread is composed of a single linear chain, and minimised when the spread resembles the structure of a star-shaped network.

We first focus on the relationship between the largest spreading event size, S_n^{max} as a function of $S_{d/w}$, where we identify a non-trivial relationship roughly composed of two opposing trends, see Fig. 6 (S_n^{max} is normalised over the total number of nodes, N). The first trend dominates the small to medium sized events, where the spreading event size increases in step with $S_{d/w}$, demonstrating the reliance of the spreading process to long and narrow pathways. However, as spreading events become larger than a given threshold (in this case, when $\frac{S_n^{\text{max}}}{N} \geq \approx 0.05$) the positive relationship between $\frac{S_n^{\text{max}}}{N}$ and $S_{d/w}$ reverses, with the spreading process enlisting an increasingly high number of relatively shorter and wider pathways. This switch suggests the existence of an upper bound in the number of long and narrow sequence of consecutive tasks within the activity network. These sequences are largely composed of low-out degree nodes, and given the finite size of the network, pose a limit to the growth of the spread. To surpass this limit, and to further fuel the growth of the spreading event size, spreading utilises additional pathways which emerge through the inclusion of occasional high out-degree nodes, which allow for the spreading process to branch out in order to increase in size, resulting in relatively wider spreads.

The rate by which the spreading event size increases depends on the spreading event structure. In the case where the spreading event size is negatively correlated with $S_{d/w}$, the rate by which the spreading event size grows is roughly 3 times faster compared to the case where the spreading event size is positively correlated with $S_{d/w}$ (gradient is roughly -0.09 and $+0.03$, respectively). This result emphasizes the multiplicative effect that wider structures can provide, which in turn enhances the number of nodes that can be reached, and in turn, affected.

With respect to the impact of direct and subsequent exposure, P_1 and p_∞ show district trends, demonstrated by the marker colour patterns in Fig. 6a,b, respectively. Focusing on the impact of P_1, the transition in marker colour, from blue to red, is accompanied with a smooth increase in the spreading event size (Fig. 6a,c). This result is somewhat expected, since P_1 plays a key role in the progression of the overall spreading process. In addition, P_1 has an important role in determining the structure of the resulting spreading structure, albeit in a non-trivial manner. In particular, short and wide structures (low $S_{d/w}$) can occur at both extremes of P_1, with the shape slowly converging to the highest attainable $S_{d/w}$ values as P_1 approaches ~0.5. Shifting focus to the impact of p_∞, we observe that the entire range of $S_{d/w}$ is obtainable under any given value of p_∞, indicating the limited role of subsequent exposure in determining the structure of the pathways used by the spreading process (Fig. 6b). The subtle impact of p_∞ on $S_{d/w}$ is further highlighted in the limited range of $S_{d/w}$ obtained reported Fig. 6d, which is significantly lower than the corresponding impact of P_1 in Fig. 6c.

Note that these results are robust when we consider the relationship between S_n^{avg} (instead of S_n^{max}) and $S_{d/w}$, as a function of P_1 or p_∞, see Supplementary Fig. 4.

We now focus on the relationship between the propagation rate of the largest spreading event, S_r^{max}, and the structure of the pathways used to sustain it, $S_{d/w}$, as a function of P_1 (Fig. 7a) and p_∞ (Fig. 7b). Similar to Fig. 7, we identify a non-trivial relationship roughly composed of two distinct behaviours, where the propagation rate initially increases in step with spreading pathways becoming increasingly long and narrow (higher $S_{d/w}$). This overall increase in S_r^{max} is the result of two conflicting effects – an increase in propagation rate, driven by higher P_1
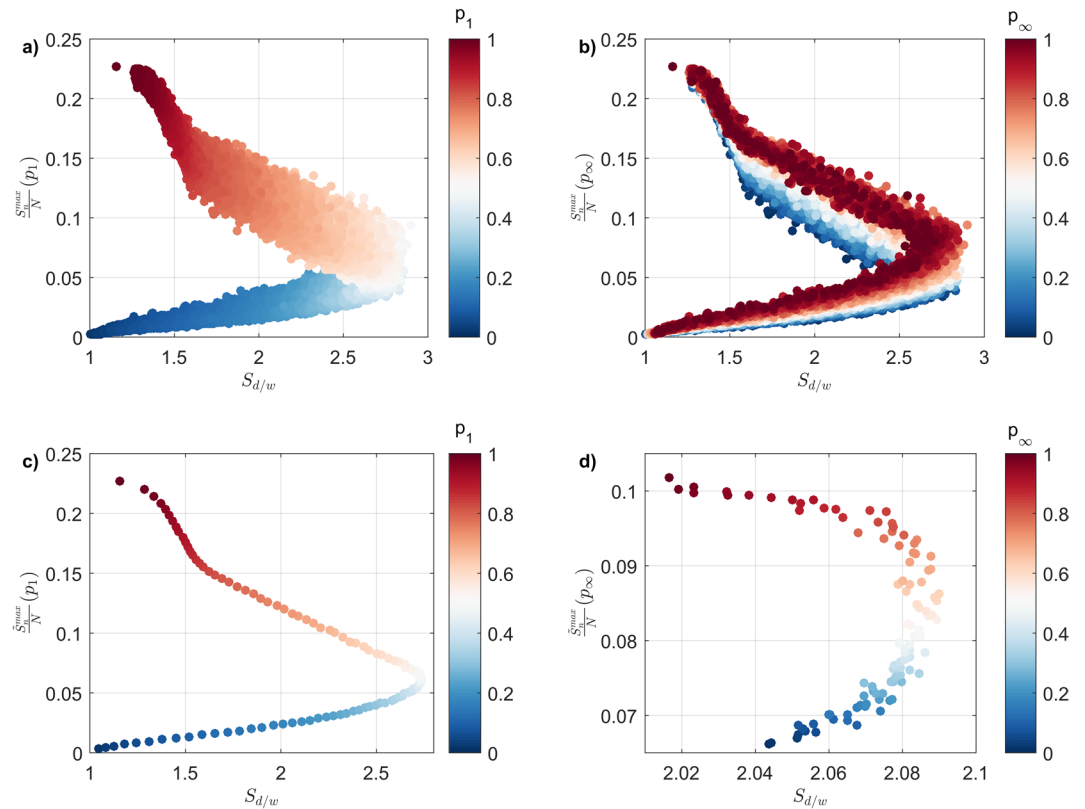
Figure 6. Parameter space of the largest spreading event size ($S_n^{\max}$), normalised over the total number of nodes (N), and its underlying structure ($S_{d/w}$), as a function of (**a**) p_1 and (**b**) p_∞; (**c**) the relationship between the largest spreading events, averaged over p_∞ and mapped as a function of p_1, and (**d**) the relationship between the largest spreading events, averaged over p_1 and mapped as a function of p_∞.

values (Fig. 7a,c), combined with a decrease in propagation rate driven by higher p_∞ values (Fig. 7b,d). This result demonstrates the conflicting nature of P_1 and p_∞, where the former relies on direct interactions which are faster to affect, while the latter introduces additional indirect pathways that take longer to evaluate completely due to their non-trivial nature (similar to the ones depicted in Fig. 1).

Once $S_r^{\max}$ reaches a given threshold (in this case, $S_r^{\max} \approx 12$), its positive relationship with $S_{d/w}$, reverses to a negative relationship, which essentially reflects the need to utilise wider pathways (and hence, triggers a decrease in $S_{d/w}$), in order to increase the propagation rate further. This switch in behaviour is induced when P_1 grows over ~0.5; as soon as this reversal in the relationship between $S_r^{\max}$ and $S_{d/w}$ takes place, the fork-like shape of Fig. 7 indicates that two possible trajectories are available. Importantly, the similar marker colouring within both trajectories (Fig. 7a) suggests that P_1 has a limited role in determining which of the two trajectories is followed. Yet the distinct marker colouring shown in Fig. 7b indicates that p_∞ is the key parameter in determining which of the two trajectories is followed. Specifically, the primary trajectory is accessible under the entire range of p_∞, and the secondary trajectory, accessible under a limited range of p_∞ values, roughly ranging from 0 to 0.6. This demonstrates the complex nature of subsequent exposure: on one hand, low-to-medium subsequent exposure means that the topology of the utilised failure pathways depends on the topology of the network itself; on the other hand high subsequent exposure always results in wider pathways being utilised (since any auxiliary path stemming from the main failure pathway is pursued to its full length, triggering an increase in S_w, and hence a decrease in $S_{d/w}$). Both of these aspects are further reinforced by isolating the results obtained at equal P_1 and p_∞ increments, with p_∞ controlling the emergence of the second trajectory (see Supplementary Fig. 5 for P_1, and Supplementary Fig. 6 for p_∞).

'Hidden influentials' and the effect of p_1 and p_∞. Large-scale spreading events are typically associated with extra-ordinary topological characteristics of the node that initially triggers them (i.e. the seed node), the simplest of which being the (out) degree[39] – our results confirm this common conjecture, albeit with certain strong caveats. Specifically, we find that the spreading event size is highly correlated with node out-degree, as reported in[6,40] (Fig. 8a). However, this correlation deteriorates as P_1 increases in size, which suggests that the spreading process is shifting from being a local-driven process (and hence, dominated by the properties of the seed node) to a globally-driven process, where the characteristics of the intermediate nodes eventually dilute the correlation between spreading event size and the topological characteristics of the seed node. This correlation deteriorates faster once the effect of p_∞ is introduced, as additional intermediate nodes are employed early on during the spreading. Recent empirical work on information spreads has identified a similar effect, where
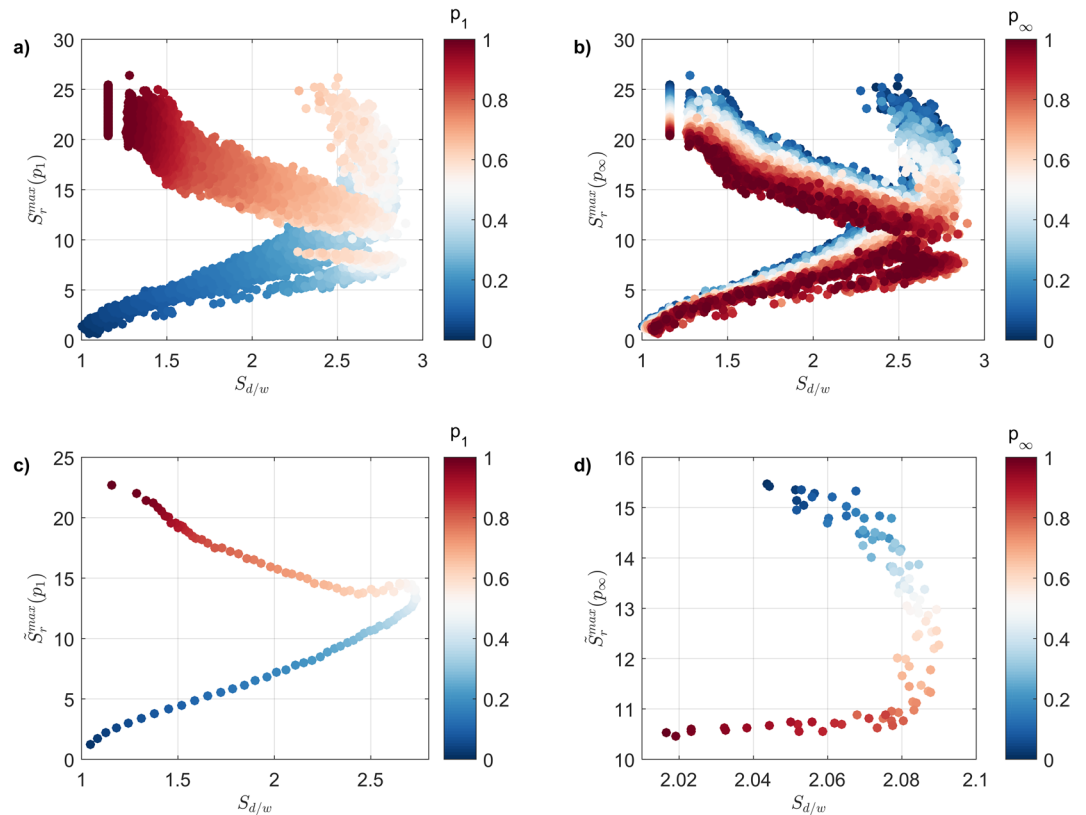
Figure 7. The relationship between the propagation rate of the largest spreading events across the entire range of p_1 and p_∞, (S_r^{max}), and its underlying structure ($S_{d/w}$), as a function of (**a**) p_1 and (**b**) p_∞; (**c**) the relationship between the propagation rate of the largest spreading events, averaged over p_∞ and mapped as a function of p_1, and (**d**) the relationship between the propagation rate of the largest spreading events, averaged over p_1, and mapped as a function of p_∞.

large-scale spreading events are largely sustained by intermediate nodes with no special topological features – the so-called 'hidden influentials'[6,41].

Following the work of Baños, *et al.*[6], we evaluate whether these 'hidden influentials' exist in the activity network by comparing the average out-degree of nodes involved in a spreading event, excluding that of the seed node, ($\tilde{k}_{out}^{avg}$) with the network average (k_{out}^{avg}), and the relationship between $\tilde{k}_{out}^{avg}$ and the spreading event size. Notably, $\tilde{k}_{out}^{avg}$ converges to k_{out}^{avg} as the spreading event size increases, confirming the presence of these 'hidden influentials' (Fig. 8b). This behaviour demonstrates that the existence of extra-ordinary nodes (e.g. hubs) is not a necessary condition for large-scale spreading to occur, and hints for other non-local topological properties that may characterise the nature of these 'hidden influentials'[9,42]. More generally, this result highlights the intrinsic challenge in containing spreading in general, where system-wide spreading events are sustained by merely typical nodes, which themselves are hard to identify *a priori*.

We now focus on exploring the impact of P_1 and p_∞ on the emergence of these 'hidden influentials', by considering the largest $\tilde{k}_{out}^{avg}$, averaged across the entire p_∞ and P_1 values, respectively. Notably, an overall decreasing trend is noted as both P_1 and p_∞ increase, where an increase in P_1 triggers a rapid decrease in $\tilde{k}_{out}^{avg}$ (Fig. 8c), converging towards k_{out}^{avg}, while an increase in p_∞ triggers a linear decrease in $\tilde{k}_{out}^{avg}$ (Fig. 8d). This behaviour corresponds to an increase in the role of 'hidden influential' in sustaining the spreading process as immediate failure becomes more likely. In conjunction with the fact that spreading event sizes increases with larger P_1 and/or p_∞, these results further suggest that larger spreading events may be harder to contain than smaller ones, simply because larger ones are increasingly reliant on the existence of these 'hidden influentials'.

Discussion

In this paper, we have introduced a simple model which allows us to decouple the effect of subsequent exposure from the overall spreading process, and comparatively examine its impact on key quantities, including spreading event size (Figs 2,3) and propagation rate (Fig. 5). Our results highlight the conflating nature of spreading, where subsequent exposure increases the number of nodes affected whilst reducing the rate in which the spread progresses. With subsequent exposure being a derivative of clustered networks, our results clarify broader discussions within the literature, which typically focus in providing high-level insight[43–46]. For example, ref.[44] focuses on identifying a (largely) positive link between clustering and spreading event size. Yet, the question of *why* clustering enhances spreading event size remains unexplored, with cases of no effect being treated as some sort of outliers. Our results
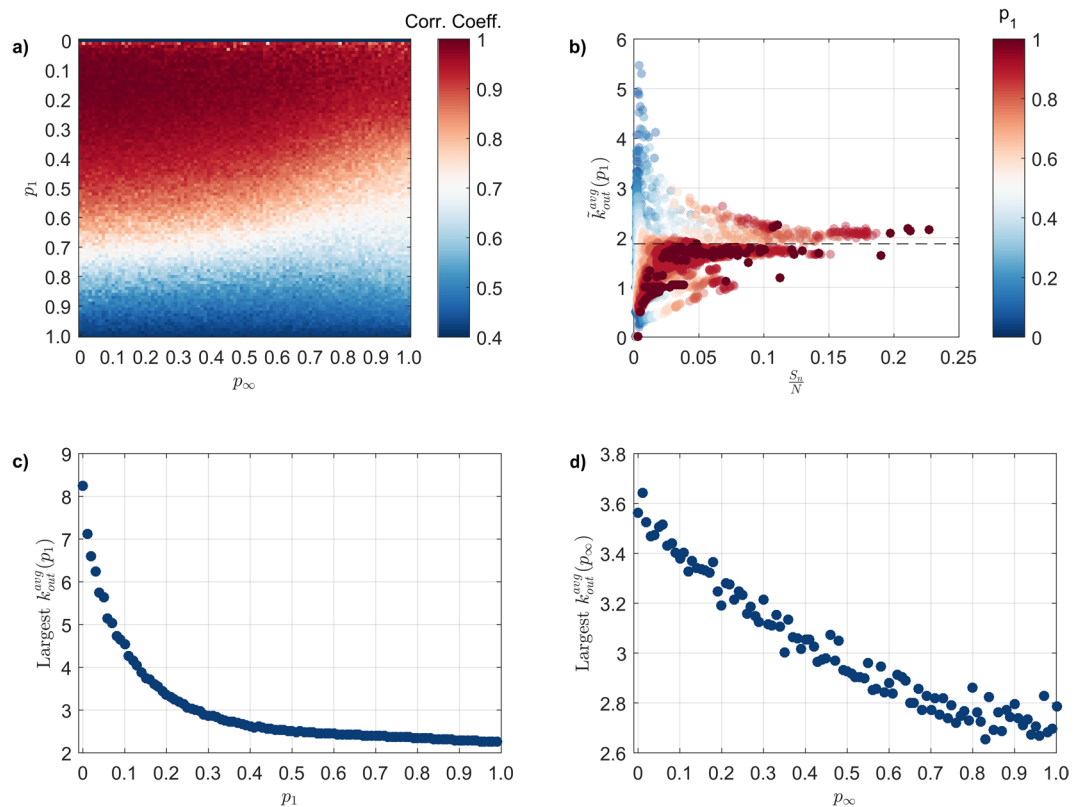
Figure 8. (**a**) Correlation coefficient between spreading event size and the out-degree of the initial seed node, as a function of p_1 and p_∞; (**b**) average out-degree of nodes involved in a spreading event, excluding that of the initial seed node, $(\tilde{k}_{out}^{avg})$ as a function of the normalised spreading event size and p_1; dotted line corresponds to the network average out-degree, k_{out}^{avg}; (**c**) largest $\tilde{k}_{out}^{avg}$ as a function of p_1, where a lower value indicates an increasingly important role for typical node in sustaining spreading, converging to k_{out}^{avg}; (**d**) largest $\tilde{k}_{out}^{avg}$ as a function of p_∞.

suggest that subsequent exposure is one possible avenue by which the relationship between clustering and enhanced spreading event size depends on, allowing for additional aspects to be explored in a similar fashion.

From a methodological standpoint, our results demonstrate the need to explicitly account for, and control the effects of, subsequent exposure when modelling spreading-like processes. For example, consider the frequent use of 'locally tree-like' approximations, typically used to deploy analytically tractable expositions into various network dynamics[1,47–49]. Despite the valuable insights that these approximations provide, the eventual nullification of subsequent exposure – and its effect on the spreading process – clouds the real difference between these models and the respective real-world systems they represent, skewing our confidence and biasing results in a non-trivial manner. The results presented within this paper serve as additional motivation to recently emerging lines of inquiry[50,51] which focus on relaxing the 'locally-tree like assumption', integrating the effect of subsequent exposure to the overall spreading process. More generally, relaxing these approximations has the potential for uncovering dynamical properties that are shared across a range of real-world systems, similar in spirit to the work of Barzel and Barabási[20].

In terms of applications, our work provides the grounds for a dialogue between researchers in the network science and project management, where hotly-researched, domain challenges (e.g. project complexity evaluation)[52–57] can be treated as network-related problems[31,32]. For example, increased susceptibility to the spreading of failures can be reasonably interpreted as a contributing factor to project complexity. Hence, the relationship between spreading event size (or propagation rate) and the structure of the underlying pathways (Figs 6 and 7 respectively) can serve as an objective, quantitative measure for project complexity (assuming that the activity network is an up-to-date reflection of the actual project plan). Similarly, the proportion (and identity) of 'hidden influentials' within an activity network (Fig. 8) could be used to support the overall project risk mitigation scheme, where activities with limited connections (yet increased probability of being 'hidden influentials') receive adequate attention.

Methods

Data. The data comprises of a real-world engineering project which captures a set of planned activities that need to be completed in order to deliver a definitive commercial product in the area of defence. The overall duration of the project is 577 days, and is composed of 578 distinct tasks with 1,085 dependencies. Note that some tasks are used as planning instruments (e.g. milestones)[58] and thus, include no dependencies – these tasks are excluded from the analysis (8 tasks in total).

The delivery of each activity is typically conditional to a number of other activities. We refer to these interactions as functional dependencies, since they effectively control the function of each activity e.g. the start of activity *j* depends on the completion of activity *i*. The directionality of each dependency dictates the functional role of each activity i.e. whether it acts as a predecessor (activity *i* proceeds activity *j*) or a successor (activity *k* succeeds activity *j*) to a subsequent task (leaf activities also exist, with no successor activities).

The set of tasks and dependencies was subsequently converted to an activity network, defined as a directed network $G = \{V, \ E\}$, where V is the set of nodes and E is that of directed edges. Every activity is abstracted as a node, where a functional dependency between activity *i* and *j* is captured in the form of a directed link from node *i* to node *j*, denoted by $e_{ij} \in E$. The number of successors and predecessors each activity has corresponds to its out-degree and in-degree respectively. The cumulative probability distribution of out-degree (red) and in-degree (blue) is shown in Supplementary Fig. 7– note its heavy-tail nature, evident by the straight line formed under the log-log axes.

Spreading Model Formulation.

Every node *j* of the network at time *t* is characterised by a dynamic variable $s_j(t) \in \{0, 1\}$, where '0' and '1' correspond to the 'non-affected' and 'affected' state, respectively. During the spreading process, node *j* may irreversibly switch from the 'non-affected' to the 'affected' state at time *t* if: (i) node *j* has at least one predecessor, node *i*, and (ii) at least one node *i* was in the 'affected' state at *t*-1. Then, we artificially switch the state of some seed node at $t = 0$, from 'non-affected' to 'affected' and track the progression of the spreading process as time increases at discrete increments of 1.

In order to distinguish between direct and subsequent exposures, we keep track of node *i*'s successors at time *t*, say node *j*, and assess whether node *j* has been encountered before. If so, then node *j* has been directly exposed to the failure of a *different* predecessor at some time >*t*, but did not switch states during that time. Therefore, the probability of node *j* to switch states at now, at time *t*, is controlled by p_∞ (Fig. 1, middle panel). However, if this is the first time node *j* has been encountered, then it is the first time node *j* is exposed to failure in general and therefore the probability to switch states at time *t* is controlled by p_1 (e.g. Fig. 1, top panel).

Note the broad nature of the term 'affected', acknowledging the fact that failure can mean very different things, depending on the context of the project. For example, 'failure' can mean 'structural defect' in a construction project, or something much less tangible such as a 'contaminated' or 'compromised' in a cyber-security project.

Spreading Model Implementation.

The model is implemented as follows. First, an initialisation phase is implemented, where simulation time *t* is set to 0 and the state of all nodes is set to '0'. In addition, an empty set *B* is created in order to record all successor nodes encountered during time *t* The spreading process is initiated by externally switching the state of node *i* from '0' to '1'. We then identify all successors of node *i*, node(s) *j* (if no neighbours exist, the process terminates). For each node *j*, we record index *j* in set $B(t)$, and then check whether index *j* was already present in set $B(t - 1)$. If index *j* was not present, the interaction between node *i* and *j* is the result of direct exposure; hence, the probability of node *j* to switch states, under both model *M* and M_0, is equal to p_1. However, if index *j* was already present in set $B(t - 1)$, the interaction between node *i* and *j* is the result of subsequent exposure; hence, the probability of node *j* to switch states, under model *M*, is equal to p_∞ (in the case of M_0, p_∞ is always set to 0). Once all node(s) *j* have been tested with respect to the prospect of changing states, we record the total number of state changes up to, and including, time *t*, $S_n(t)$, and increase *t* by 1. The process repeats until the total number of state changes remains constant i.e. $S_n(t) = S_n(t - 1)$. Finally, the process is reiterated for each node *i*, in order to evaluate the total number of state changes the failure of every possible seed node. Finally, this process is repeated for 48 independent runs, with results presented herein being the average (number of runs determined in order to minimise the standard error of the mean).

Formally, the condition by which node *j* changes state depends on the spreading model used. The condition for spreading model M_0 is determined by eq. 1:

$$s_j(t) = \begin{cases} 1, & \text{if } j \notin B(t-1) \wedge \mathrm{P}_1 \geq \Theta, \mathrm{P}_1 \in [0, 1] \\ & s_j(t-1) \text{ otherwise} \end{cases} \qquad (1)$$

and the condition for spreading model *M* is determined by eq. 2:

$$s_j(t) = \begin{cases} 1, & \text{if } j \notin B(t-1) \wedge \mathrm{P}_1 \geq \Theta, \quad \mathrm{P}_1 \in [0, 1] \\ 1, & \text{if } i \in B(t-1) \wedge \mathrm{P}_\infty \geq \Theta, \quad \mathrm{P}_\infty \in [0, 1] \\ & s_j(t-1) \text{ otherwise} \end{cases} \qquad (2)$$

where variable Θ is uniformly drawn at random from U(0,1) for both eqs 1 and 2.

Data availability.

The datasets generated and/or analysed during the current study, and related source code, are available from the corresponding author on reasonable request.

References

1. Wang, W., Tang, M., Stanley, H. E. & Braunstein, L. A. Unification of theoretical approaches for epidemic spreading on complex networks. *Reports on Progress in Physics* **80**, 036603 (2017).
2. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Reviews of modern physics* **87**, 925 (2015).

3. Goel, S., Anderson, A., Hofman, J. & Watts, D. J. The Structural Virality of Online Diffusion. *Management Science* **62**, 180–196, https://doi.org/10.1287/mnsc.2015.2158 (2016).
4. Roukny, T., Bersini, H., Pirotte, H., Caldarelli, G. & Battiston, S. Default cascades in complex networks: Topology and systemic risk. *Scientific reports* **3**, 2759 (2013).
5. Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)* **1**, 5 (2007).
6. Baños, R. A., Borge-Holthoefer, J. & Moreno, Y. The role of hidden influentials in the diffusion of online information cascades. *EPJ Data Science* **2**, 6, https://doi.org/10.1140/epjds18 (2013).
7. Kitsak, M. *et al*. Identification of influential spreaders in complex networks. *Nature physics* **6**, 888–893 (2010).
8. Lawyer, G. Understanding the influence of all nodes in a network. *Scientific Reports* **5**, 8665, https://doi.org/10.1038/srep08665 (2015).
9. Morone, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65–68, https://doi.org/10.1038/nature14604 (2015).
10. Watts, D. J. & Dodds, P. S. Influentials, networks, and public opinion formation. *Journal of consumer research* **34**, 441–458 (2007).
11. Crucitti, P., Latora, V. & Marchiori, M. Model for cascading failures in complex networks. *Physical Review E* **69**, 045104 (2004).
12. Dodds, P. S. & Watts, D. J. Universal behavior in a generalized model of contagion. *Physical review letters* **92**, 218701 (2004).
13. Lorenz, J., Battiston, S. & Schweitzer, F. Systemic risk in a unifying framework for cascading processes on networks. *The European Physical Journal B* **71**, 441–460 (2009).
14. Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* **99**, 5766–5771 (2002).
15. Gleeson, J. P. Binary-State Dynamics on ComplexNetworks: Pair Approximation and Beyond. *Physical Review* **X3**, 021004 (2013).
16. Gleeson, J. P. & Porter, M. A. Message-Passing Methods for Complex Contagions. *arXiv preprint arXiv* **1703**, 08046 (2017).
17. Masuda, N. & Holme, P. In *Temporal Network Epidemiology* (eds Naoki Masuda & Petter Holme) 1–16 (Springer Singapore, 2017).
18. Masuda, N., Klemm, K. & Eguíluz, V. M. Temporal networks: slowing down diffusion by long lasting interactions. *Physical Review Letters* **111**, 188701 (2013).
19. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Physical Review Letters* **86**, 3200 (2001).
20. Barzel, B. & Barabási, A.-L. Universality in network dynamics. *Nature physics* **9**, 673 (2013).
21. Motter, A. E. & Lai, Y.-C. Cascade-based attacks on complex networks. *Physical Review E* **66**, 065102 (2002).
22. Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Physical Review E* **65**, 056109 (2002).
23. Zhao, J., Li, D., Sanhedrai, H., Cohen, R. & Havlin, S. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nature communications* **7**, 10094 (2016).
24. Daqing, L., Yinan, J., Rui, K. & Havlin, S. Spatial correlation analysis of cascading failures: congestions and blackouts. *Scientific reports* **4**, 5381 (2014).
25. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences* **100**, 11980–11985 (2003).
26. Milo, R. *et al*. Network motifs: simple building blocks of complex networks. *Science* **298**, https://doi.org/10.1126/science.298.5594.824 (2002).
27. Dodds, P. S. & Watts, D. J. A generalized model of social and biological contagion. *Journal of theoretical biology* **232**, 587–604 (2005).
28. Keeling, M. J. & Eames, K. T. Networks and epidemic models. *Journal of the Royal Society Interface* **2**, 295–307 (2005).
29. Shenhar, A. J. One size does not fit all projects: Exploring classical contingency domains. *Management Science* **47**, 394–414 (2001).
30. Ellinas, C., Allan, N., Durugbo, C. & Johansson, A. How Robust Is Your Project? From Local Failures to Global Catastrophes: A Complex Networks Approach to Project Systemic Risk. *PLoS ONE* **10**, e0142469, https://doi.org/10.1371/journal.pone.0142469 (2015).
31. Ellinas, C., Allan, N. & Johansson, A. Project systemic risk: Application examples of a network model. *International Journal of Production Economics* **182**, 50–62, https://doi.org/10.1016/j.ijpe.2016.08.011 (2016).
32. Ellinas, C., Allan, N. & Johansson, A. Exploring Structural Patterns Across Evolved and Designed Systems: A NetworkPerspective. *Systems Engineering* **19**, 179–192, https://doi.org/10.1002/sys.21350 (2016).
33. Ellinas, C., Allan, N. & Johansson, A. Toward Project Complexity Evaluation: A StructuralPerspective. *IEEE Systems Journal* **PP**, 1–12, https://doi.org/10.1109/JSYST.2016.2562358 (2016).
34. Williams, T. Towards realism in network simulation. *Omega* **27**, 305–314 (1999).
35. Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J. In *Proceedings of the fourth ACM international conference on Web search and data mining*. 65–74 (ACM).
36. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *Nature* **393**, 440–442 (1998).
37. Porter, M. A. & Gleeson, J. P. *Dynamical Systems on Networks*. (Springer, Cham, 2016).
38. Zang, C., Cui, P., Song, C., Faloutsos, C. & Zhu, W. Structural patterns of information cascades and their implications for dynamics and semantics. *arXiv preprint arXiv* **1708**, 02377 (2017).
39. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics* **74**, 47 (2002).
40. Borge-Holthoefer, J., Baños, R. A., González-Bailón, S. & Moreno, Y. Cascading behaviour in complex socio-technical networks. *Journal of Complex Networks* **1**, 3–24 (2013).
41. González-Bailón, S., Borge-Holthoefer, J. & Moreno, Y. Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist* **57**, 943–965 (2013).
42. Teng, X., Pei, S., Morone, F. & Makse, H. A. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Scientific reports* **6**, 36043 (2016).
43. Newman, M. E. Properties of highly clustered networks. *Physical Review E* **68**, 026121 (2003).
44. Stegehuis, C., van der Hofstad, R. & van Leeuwaarden, J. S. Epidemic spreading on complex networks with community structures. *Scientific reports* **6** (2016).
45. Serrano, M. Á. & Boguná, M. Clustering in complex networks. II. *Percolation properties. Physical Review E* **74**, 056115 (2006).
46. Gleeson, J. P. Bond percolation on a class of clustered random networks. *Physical Review E* **80**, 036107 (2009).
47. Gleeson, J. P. Cascades on correlated and modular random networks. *Physical Review E* **77**, 046117 (2008).
48. Gleeson, J. P. & Cahalane, D. J. Seed size strongly affects cascades on random networks. *Physical Review E* **75**, 056103 (2007).
49. Karrer, B. & Newman, M. E. Message passing approach for general epidemic models. *Physical Review E* **82**, 016101 (2010).
50. Durrett, R. Some features of the spread of epidemics and information on a random graph. *Proceedings of the National Academy of Sciences* **107**, 4491–4498 (2010).
51. Radicchi, F. & Castellano, C. Beyond the locally treelike approximation for percolation on real networks. *Physical Review E* **93**, 030302 (2016).
52. Whitty, S. J. & Maylor, H. And then came Complex Project Management (revised). *International Journal of Project Management* **27**, 304–310, https://doi.org/10.1016/j.ijproman.2008.03.004 (2009).
53. Oehmen, J., Thuesen, C., Ruiz, P. P. & Geraldi, J. Complexity Management for Projects, Programmes, and Portfolios: An Engineering Systems Perspective. (Project Management Institute, PMI, 2015).
54. Baccarini, D. The concept of project complexity—a review. *International Journal of Project Management* **14**, 201–204 (1996).
55. Williams, T. M. The need for new paradigms for complex projects. *International Journal of Project Management* **17**, 269–273 (1999).

56. Geraldi, J., Maylor, H. & Williams, T. Now, let's make it really complex (complicated) A systematic review of the complexities of projects. *International Journal of Operations & Production Management* **31**, 966–990 (2011).
57. Qazi, A., Quigley, J., Dickson, A. & Kirytopoulos, K. Project Complexity and Risk Management (ProCRiM): Towards modelling project complexity driven risk paths in construction projects. *International Journal of Project Management* **34**, 1183–1198 (2016).
58. PMI. *Project Management Body of Knowledge (PMBOK®GUIDE).* (2001).

Acknowledgements

Author Contributions

C.E. obtained funding, designed the research, performed numerical simulations, analysed the results and prepared the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-22770-3.

Competing Interests: The author declares no competing interests.

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