Applied Network Science

Spreading of performance fluctuations on real-world project networks

Iacopo Pozzana[1,2]* , Christos Ellinas[1], Georgios Kalogridis[1] and Konstantinos Sakellariou[1]

*Correspondence:
iacopo@nodeslink.com
[1] Nodes & Links Ltd, 107 Fleet
Street, London EC4A 2AB, UK
Full list of author information
is available at the end of the
article

Abstract

Understanding the role of individual nodes is a key challenge in the study of spreading processes on networks. In this work we propose a novel metric, the reachability-heterogeneity (RH), to quantify the contribution of each node to the robustness of the network against a spreading process. We then introduce a dataset consisting of four large engineering projects described by their activity networks, including records of the performance of each activity, i.e., whether it was timely delivered or delayed; such data, describing the spreading of performance fluctuations across activities, can be used as a reliable ground truth for the study of spreading phenomena on networks. We test the validity of the RH metric on these project networks, and discover that nodes scoring low in RH tend to consistently perform better. We also compare RH and seven other node metrics, showing that the former is highly interdependent with activity performance. Given the context agnostic nature of RH, our results, based on real-world data, signify the role that network structure plays with respect to overall project performance.

Keywords: Spreading on networks, Project management, Node vulnerability, Delay spreading, Project performance, Dynamical processes on networks, Network heterogeneity

Introduction

Spreading broadly refers to the notion of an entity propagating through a networked system, typically fueled by a dynamical process (Pastor-Satorras et al. 2015). Spreading processes are a powerful set of tools for modelling a wide-range of real-world phenomena, including the dissemination of (dis)information on social media (Vosoughi et al. 2018), the propagation of a pathogen within a population (Colizza et al. 2006), cyber attacks on computer networks (Cohen et al. 2003) and delays in transportation systems (Preciado et al. 2014). Node degree (Wasserman et al. 1994), betweenness centrality (Freeman 1977) and eigenvector centrality (Bonacich 1972) are all examples of topological metrics used to approximate the role of individual nodes in the context of spreading processes, a problem that yet remains open in the extant literature (Radicchi and Castellano 2016; Erkol et al. 2018).

The problem is further complicated by the scarcity of reliable ground truth. Datasets providing an individual-level description of a spreading process within a population are

few (Groendyke et al. 2011; Chinazzi et al. 2020), with aggregated reports being more common (Stack et al. 2013). Even when working with real-world networks, researchers often resort to simulations for what concerns the spreading dynamics itself (Mishra et al. 2016; Davis et al. 2020); and when information describing the network structure is also incomplete, the interplay between the two problems further amplifies the difficulty of the task (Gomez-Rodriguez et al. 2012).

A bountiful, yet underexploited, source of reliable data, describing both complete network structures and the fine-grained evolution of real spreading processes on them, can be found within the field of project management (Ellinas et al. 2016; Vanhoucke 2013; Santolini et al. 2020). Projects are described by schedules, time-ordered lists of interconnected activities that can be naturally modelled as directed acyclic graphs (DAGs) (Valls and Lino 2001).

Spreading can be used to describe performance fluctuations on project networks: activities completed behind or ahead of schedule can impact other activities downstream and initiate a spreading process (Ellinas et al. 2015; Guo et al. 2019). Project schedules record both planned and real starting dates for all activities, therefore providing a complete record of the performance fluctuation dynamic.

Real-world projects often perform poorly in terms of both time and cost, a fact that holds true across different countries, companies, and industries (Evrard and Nieto-Rodriguez 2004; Budzier 2011). As an example, studies have shown that, in the construction sector, almost nine out of ten projects are subject to cost overruns, for an average overrun cost estimated to be as high as 45% (Flyvbjerg et al. 2003; Flyvbjerg 2007).

Large failures in projects often start as localised phenomena, with the performance of a single activity eventually impacting the performance of the entire project. Cases have been documented where an initial disruption located in a single activity ended up affecting almost a third of the entire project (Sosa 2014), or increasing its final cost by 20 to 40% (Terwiesch and Loch 1999). In this respect, the networked structure of the schedule has been shown to play an important role (Ellinas 2019; Mihm et al. 2003).

Methodologically, most of the efforts aimed at modelling project performance through their associated networks have centered on cascade models (Wang et al. 2018), for example by focusing on how small-scale delays can trigger project-wide cascades (Ellinas 2019), [19], or by studying the role of indirect interactions between activities (Ellinas 2018). With the present study, we contribute to this line of work by developing a measure that draws a direct connection between topology and performance at the activity level, and then validate it using real performance data.

Our contribution is twofold. First, building on prior work by Estrada (2010) and by Ye and colleagues (Ye et al. 2013), we introduce a novel measure called reachability-heterogeneity (RH), which quantifies heterogeneity on DAGs. The RH is defined both at the global (how heterogeneous is a network) and local level (how much a node contributes to the heterogeneity).

Heterogeneity plays an important role in determining how vulnerable a network is with respect to spreading processes (Moreno et al. 2002). If all nodes have equal spreading power, then the network is maximally robust, not presenting any weak spots to either targeted attacks or random failures (Xiao et al. 2018). Numerous studies quantify heterogeneity by examining the distribution of some node-level measure [examples including

Pozzana *et al. Appl Netw Sci*   (2021) 6:25

Page 3 of 15

degree (Sun et al. 2016), memory (Karsai et al. 2014; Sun et al. 2015), activity potential (Perra et al. 2012; Liu et al. 2014), attractiveness (Pozzana et al. 2017), burstiness (Ubaldi et al. 2017) and modularity (Nadini et al. 2018)], and examine the relationship between such heterogeneity and the spreading dynamics.

The novelty of our contribution consists in leveraging a topological feature that is intrinsically related to the spreading process: the number of descendants and of ancestors. Due to the absence of cycles, the size of the ancestry trees plays an especially important role in DAGs; and, to the best of our knowledge, there is no study examining the relevance of its heterogeneity in spreading processes. Our analysis qualitatively verifies that the global RH score is a good indicator of the heterogeneity of the ancestry and descendancy distributions.

Our second contribution consists in the introduction of a dataset describing the networks of activities that make up four real-world, complex projects; these data provide a reliable ground truth for benchmarking spreading processes. We experimentally validate the accuracy of RH against performance records from the projects' activities. Our results show that best-performing nodes tend to score low in RH, making our metric a good tool for their identification. Furthermore, we compare the local RH to seven other node metrics by computing the mutual information between them and the activity performance; RH reports the highest (or, in one case, second- or third-highest, depending on the performance metric considered) mutual information values among all candidates. Given the context agnostic nature of RH, our results signify the role that the network structure has with respect to overall project performance, and indicate that the RH score gives computational embodiment to the notion that a network is maximally robust against spreading when all nodes contribute equally to it.

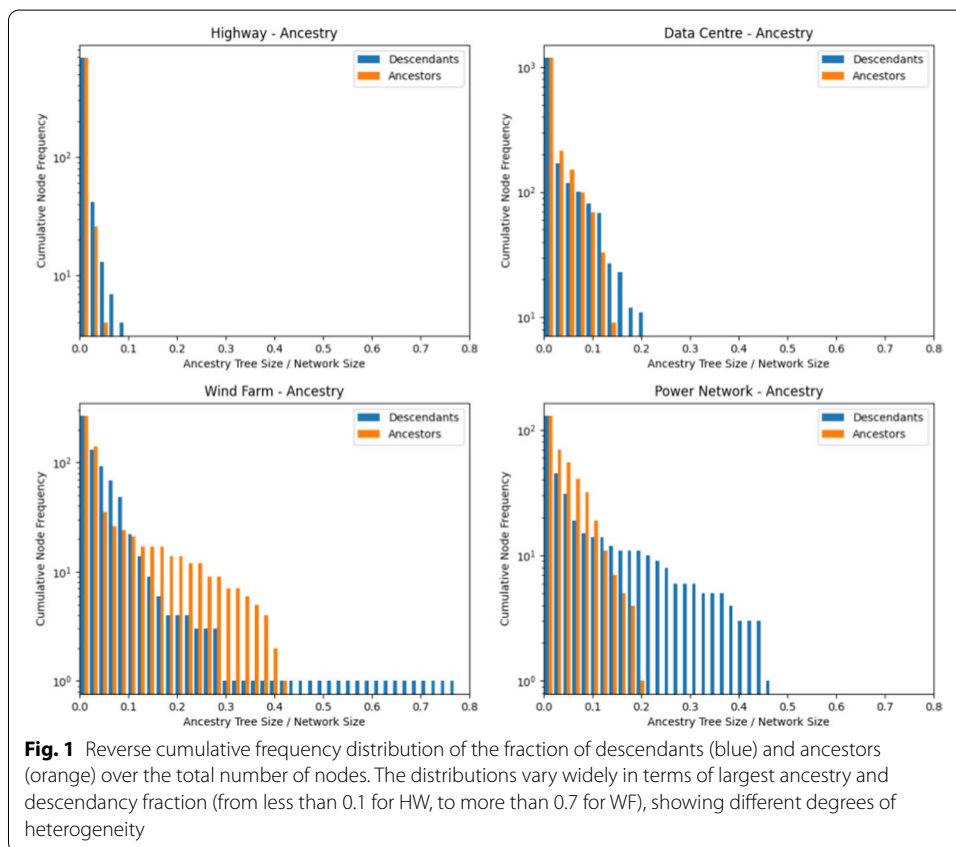## Data and methods

Project data

We use data from four complex engineering projects, where 'complex' refers to the non-triviality of underlying dependencies (Baccarini 1996; Jacobs and Swink 2011; Ellinas et al. 2016). For each project, we use the schedule to generate the respective activity network (Valls and Lino 2001). The project schedule consists of a list of activities and in a list of dependencies between them. For each activity, the schedule contains the planned and actual start and end date. Target dates for an activity correspond to its start and end date as initially planned. Actual dates, as the name suggests, correspond to the dates when the activity was actually initiated and completed.

The schedule naturally lends itself to be represented as a network, with activities taking the role of nodes and dependencies representing directed links among them (from now on, we will use the terms 'node' and 'activity' interchangeably). A link from node i to node j means that activity i must first be completed before activity j can start. At this stage, we remove from the network all isolated nodes, since they are not capable of contributing to any sort of spreading in a meaningful way. Notice that activity networks are DAGs, as cyclic dependencies between activities are not allowed.

The four projects analysed here detail the construction of different kinds of infrastructure: a highway (HW), a data centre (DC), a wind farm (WF) and a power network (PN). The number of activities and dependencies for each project ranges from

Pozzana *et al. Appl Netw Sci*   (2021) 6:25

Page 4 of 15

Table 1 For each of the four activity networks we report the number of activities (nodes), dependencies (directed links) and weakly connected components, and the size of the largest weakly connected component

Project	Activities	Dependencies	WCCs	LWCC
Highway (HW)	682	666	113	100
Data centre (DC)	1185	1510	111	440
Wind farm (WF)	266	425	1	266
Power network (PN)	129	138	10	62

Fig. 1 Reverse cumulative frequency distribution of the fraction of descendants (blue) and ancestors (orange) over the total number of nodes. The distributions vary widely in terms of largest ancestry and descendancy fraction (from less than 0.1 for HW, to more than 0.7 for WF), showing different degrees of heterogeneity

less than two hundred to more than a thousand (Table 1). Activity networks do not necessarily consist of a single component: projects may have a modular structure, being composed of independent sections. The number of weakly connected components for each network, and the size of the largest one, are also reported in Table 1. We verify that all four networks are acyclic, as expected.

Figure 1 shows the reverse cumulative distribution of the number of ancestors and descendants for each project network, divided by the network's size. The four datasets present significant differences between each other, with the most peaked (HW) having no ancestry or descendancy larger than 0.1, while WF and PN have numerous nodes with either descendancy or ancestry ranging between 0.2 and 0.5 of the entire network. In all cases the distribution of descendants has the longest tail of the

two, although in the case of WF this is caused by the presence of a single node with a large number of descendants (more than 0.7 of all nodes). Overall, the four datasets show very different degrees of heterogeneity in their ancestry and descendancy distributions.
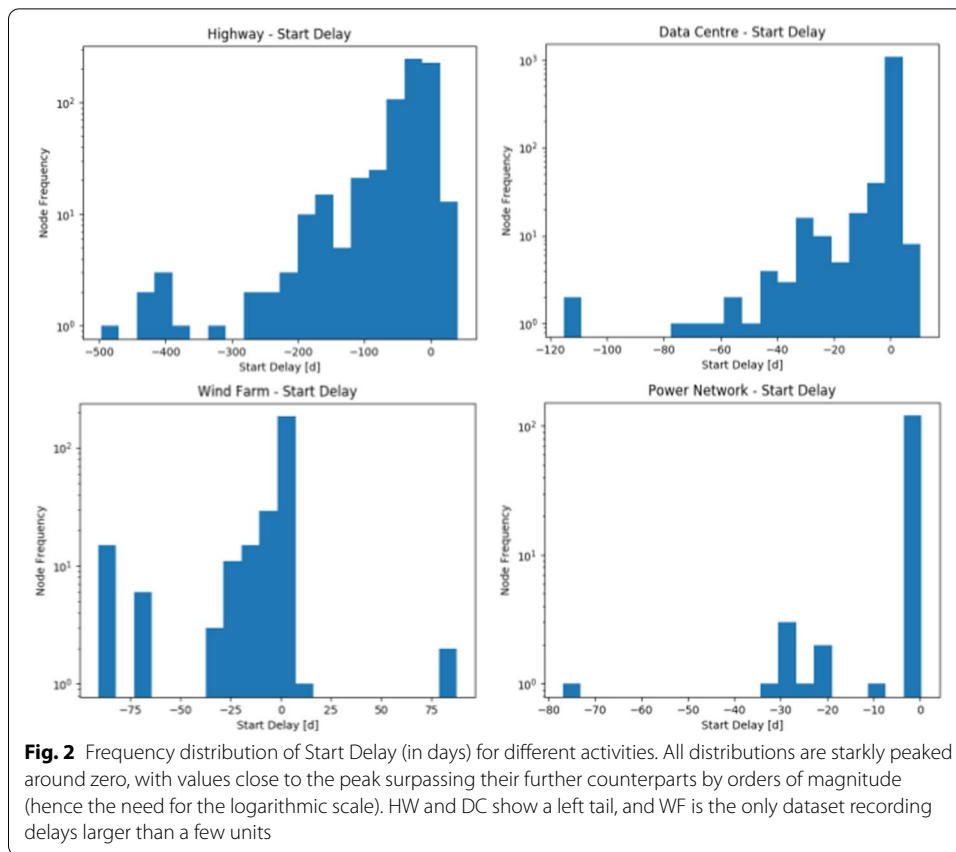
Activity performance

Performance indicators for each activity can be constructed by comparing its target with the actual start and end dates. Here we focus on a particular form of performance, the Start Delay i.e., the difference between the target and the actual start date. The advantage of this metric is that it allows us to focus on performance fluctuations that occurred upstream of an activity, separating them from fluctuations that might occur while the activity is being carried out. A possible alternative performance indicator would be represented by the End Delay, i.e., the delay in the end date of an activity; this second measure would account for fluctuations that occur while the activity is taking place too, as well as for those that took place upstream.

Suppose, for example, that the completion of activity j is dependent on the completion of activity i, and the two activities are taking place at the same time. If a delay happens in i after the start of j, the same delay might end up propagating to j as well, delaying its completion; therefore the End Delay would capture such propagation, while the Start Delay would not. However, a significant downside of the End Delay is that it also accounts for the emergence of performance fluctuations within the activity itself (endogenous fluctuations), i.e., fluctuations that would have occurred even if the activity had taken place in isolation, and that are, hence, independent of the network topology.

A third type of performance metric is represented by the Duration Difference, the difference between the actual and target duration of an activity. A significant limitation that this metric shares with the End Delay is that it does not allow to disentangle effects due to upstream activities from others native to the activity itself. Indeed, a delay occurring within an activity (and therefore increasing its duration) might very well be due to it not taking place at the originally planned time, for example when the required resources are not available in compliance with a revise schedule, causing an activity to be kept on hold.

Therefore, while all three performance metrics have their own advantages and limitations, the Start Delay is the only one that can effectively separate inherited from endogenous fluctuations, a highly desirable feature when studying the phenomenon from a spreading perspective. For this reason we choose to focus on the Start Delay as our main performance metric, while still including End Delay and Duration Difference in one of our experiments for increased robustness.

In Fig. 2, we plot the distribution of Start Delay values, measured in days. Most recorded values are negative, indicating that an activity has started ahead of schedule. Only in WF values larger than a few (positive) units appear. In all cases, the distribution peaks at zero, corresponding to activities having started as planned, and frequencies range over several orders of magnitude, warranting the use of a logarithmic scale on the y-axis. HW and DC show a distinct left tail, with the frequency of activities decreasing as the Start Delay decreases.

Fig. 2 Frequency distribution of Start Delay (in days) for different activities. All distributions are starkly peaked around zero, with values close to the peak surpassing their further counterparts by orders of magnitude (hence the need for the logarithmic scale). HW and DC show a left tail, and WF is the only dataset recording delays larger than a few units

Reachability-heterogeneity measure

To quantify the heterogeneity of a project network, we start from Estrada's heterogeneity measure (Estrada 2010), and particularly its extension to directed graphs (Ye et al. 2013):

$$\rho(G) = \frac{1}{|N| - 2\sqrt{|N| - 1}} \sum_{(i,j) \in E} \left(\frac{1}{\sqrt{k_i^{out}}} - \frac{1}{\sqrt{k_j^{in}}} \right)^2 \tag{1}$$

Above, k_i^{in} and k_i^{out} represent the in- and out-degree of node i respectively, N is the set of all edges in the network G, and the summation is taken over the set of all G's (directed) edges E.

Since activity networks are DAGs, a performance fluctuation in node i can only propagate to its descendants. In turn, node i can only be affected by performance fluctuations occurring in its ancestors. By descendant of i, we mean any node j such that a directed path from i to j exists; by ancestor of i, we mean any node j such that a directed path from j to i exists. i is a descendant of j if and only if j is an ancestor of i.

In assessing the heterogeneity of an activity network with respect to performance fluctuation spreading, we make use of the more cogent notion of ancestor (descendant) instead of predecessor (successor). The contribution of a pair to the overall score is a function of the difference between the number of ancestors and descendants of

Table 2 Global RH scores for the four activity networks

Project	Global RH
Highway (HW)	0.238
Data centre (DC)	0.332
Wind farm (WF)	0.680
Power network (PN)	0.514

The comparison with Fig. 1 shows a correspondence between higher score values and longer tail in the ancestry tree size distribution

the two nodes involved, rather than of their in- and out-degree, accounting for the impact of ancestors and descendants to the overall spreading process.

In formulae, we replace the in- and out-degree from Eq. 1 with the number of ancestors and descendants of the two nodes respectively, and we extend the summation to all pair of connected nodes, leading to the following definition:

$$RH^{global}(G) = \frac{1}{|N| - 2\sqrt{|N| - 1}} \sum_{(i,j) \in C} \left(\frac{1}{\sqrt{d_i}} - \frac{1}{\sqrt{a_j}} \right)^2 \quad (2)$$

In Eq. 2, d_i and a_i represent the number of descendants and ancestors of node i, and C is the set of all ordered pairs of connected nodes. This metric is a *global* network property that allows comparison between different topologies and quantification of their heterogeneity with respect to the size of nodal ancestry lineages. In comparison, the measure in Eq. 1 focuses exclusively on the immediate neighbourhood of the node.

In order to provide more actionable information, we introduce an additional version of the measure above, defined at the level of single nodes, in order to allow targeted interventions by project experts. Our aim in doing so is to answer the question: if a single node could be removed in order to make the topology less vulnerable, which one would be the best choice? The answer can simply be computed by taking the difference between the network scores before and after the removal:
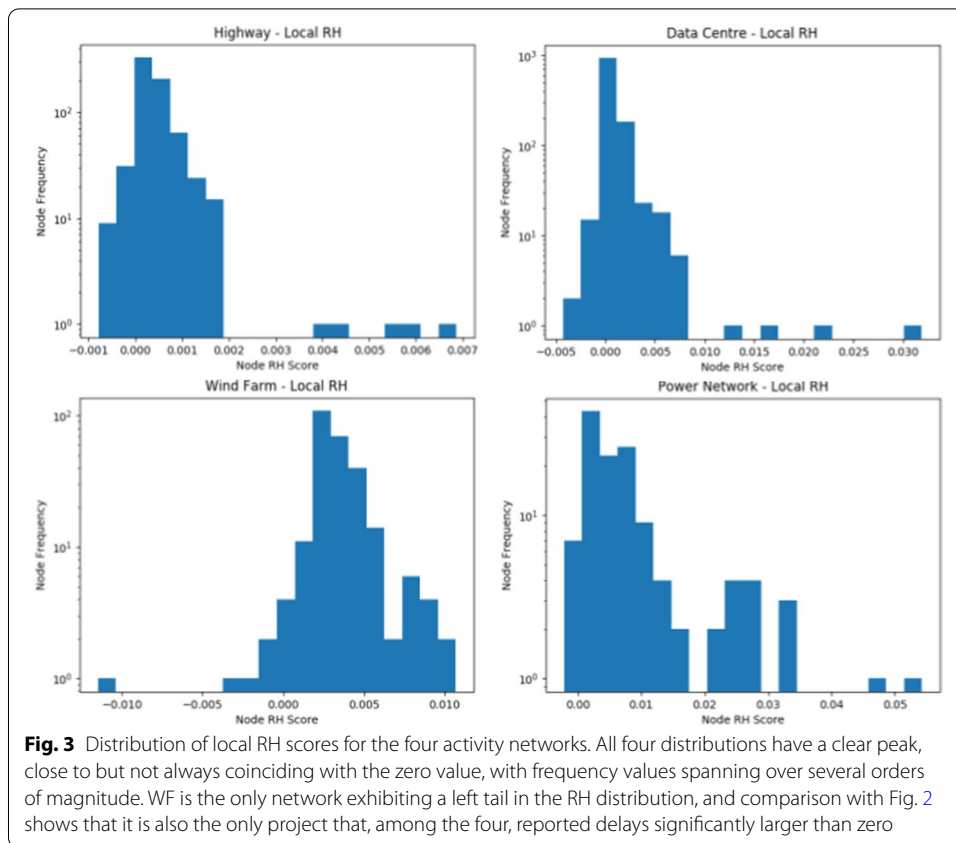
$$RH^{local}(i) = RH^{global}(G) - RH^{global}(G \setminus \{i\}) \quad (3)$$

We call this measure Reachability-Heterogeneity (RH).

Results

We first calculate the RH score for all nodes on all the four projects, as well as the four global RH scores, which are reported in Table 2. The global score provides a good characterisation of the shape of the ancestry and descendancy distributions shown in Fig. 1, with the highest RH value (WF) being assigned to the distribution with the longest tail, and the other three following in order.

The distributions of node-level RH scores for all four projects are shown in Fig. 3. All distributions show frequency values spanning over various orders of magnitude and a rather clearly identifiable peak, always close, but not always corresponding, to the zero value. HW, DC and PN bear some degree of similarity in shape, with a single-sided flat tail in the higher values, but differ in magnitude. Interestingly, WF, which is the only project to report significant positive delays (Fig. 2), is also the only project with a significant

Fig. 3 Distribution of local RH scores for the four activity networks. All four distributions have a clear peak, close to but not always coinciding with the zero value, with frequency values spanning over several orders of magnitude. WF is the only network exhibiting a left tail in the RH distribution, and comparison with Fig. 2 shows that it is also the only project that, among the four, reported delays significantly larger than zero

left tail in the RH score distribution; it is worth remarking that the RH score is based on the network structure alone, and does not account for performance data.

To assess the effectiveness of RH in quantifying node vulnerability, we first use activity performance to build our ground truth. Specifically, we use the Start Delay indicator, as described in the Methods section. To mitigate the noise, we group the nodes in bins of equal width.[1] Within every bin, we calculate the Start Delay of each node and a number of summarising statistics, namely: mean, median, 50% and 68% Confidence Intervals (CIs).

The results for each project are reported in Fig. 4, in the form of boxplots; the population and cut boundaries for each box are reported in Table 3. In general, the Start Delay value increases for greater RH,[2] showing that this newly introduced measure can provide a good indicator of activity performance. It is worth reminding that the Start Delay accounts for delays inherited from ancestors, signifying the relationship between performance and spreading (see the Data section for further discussion).

In particular, for the HW data the trend is especially evident in the mean and the lower end of the CIs. The upper end of the CIs seems to be capped at zero, as almost all Start

---

[1] We use the OptBinnig Python package to choose the number of bins: http://gnpalencia.org/optbinning/.

[2] Notice that by 'delay' we indicate here a quantity that can assume both negative and positive values, therefore an increased delay can describe an activity starting "less early".
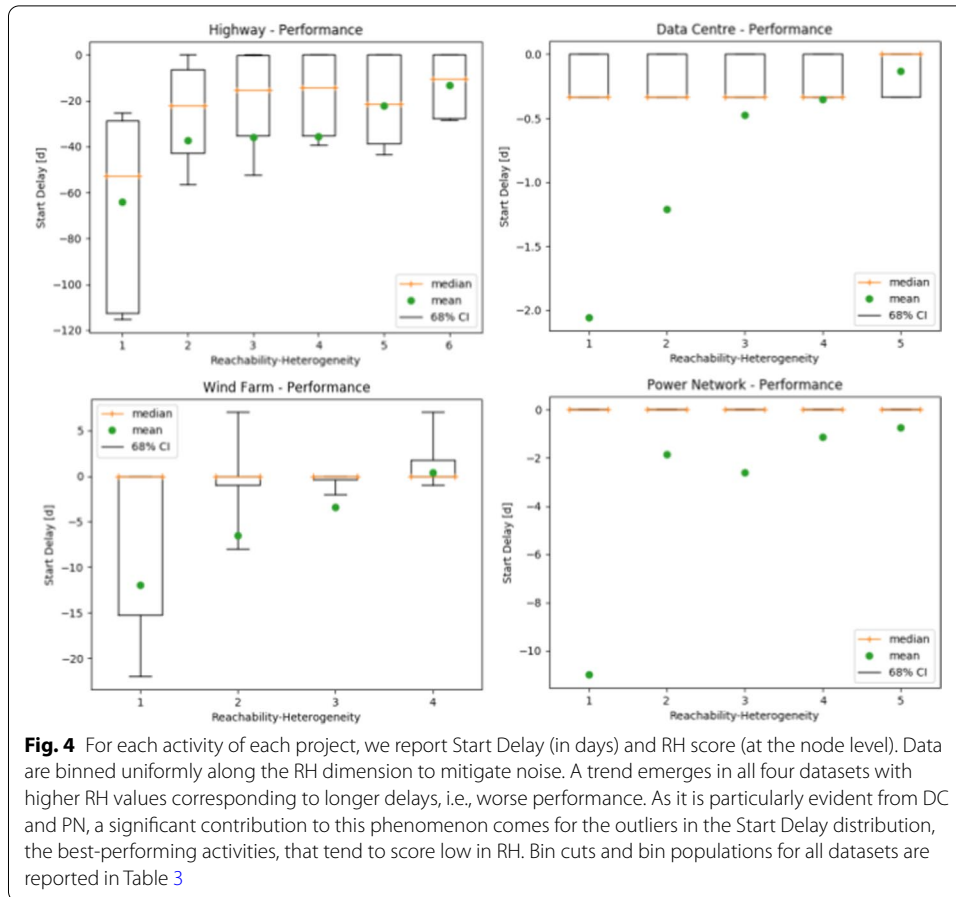
Fig. 4 For each activity of each project, we report Start Delay (in days) and RH score (at the node level). Data are binned uniformly along the RH dimension to mitigate noise. A trend emerges in all four datasets with higher RH values corresponding to longer delays, i.e., worse performance. As it is particularly evident from DC and PN, a significant contribution to this phenomenon comes for the outliers in the Start Delay distribution, the best-performing activities, that tend to score low in RH. Bin cuts and bin populations for all datasets are reported in Table 3

Table 3 Binning details for Fig. 4

Highway						
Bin cuts	− 7.87e−04	− 2.2e−05	3.61e−04	7.44e−04	0.001127	0.00151	0.00687
Bin population	–	40	327	207	64	24	20
Lower outliers	–	4	53	32	11	3	2
Data Centre						
Bin cuts	− 0.004317	0.001119	0.002931	0.004743	0.006555	0.03192
Bin population	–	948	186	23	18	10
Lower outliers	–	138	12	1	1	0
Wind farm						
Bin cuts	− 0.011488	0.002938	0.004047	0.005157	0.010705
Bin population	–	128	70	40	28
Lower outliers	–	20	10	5	5
Power network						
Bin cuts	− 0.002214	0.000612	0.003438	0.006264	0.009089	0.054302
Bin population	–	7	43	23	26	30
Lower outliers	–	1	4	2	1	1

The first and last bins cut values correspond to the minimum and maximum local RH score for the dataset. For the population and outliers rows, values in the *n*th column correspond to the bin delimited by the (*n* − 1)th and *n*th cuts. By lower outliers we designate values lower than the 16th percentile (i.e., falling below the lower bin whisker shown in the figure)

Delay values are negative (see Fig. 2). The trend is clearer for lower RH values,which then flattens towards the tail.

For the DC data, the trend is stronger in the mean. The clear separation between the mean value and the centre of the distribution confirms that Start Delay distributions within each bin are long-tailed, with longer tails in correspondence of lower RH values. Again, all Start Delay values are negative.

The WF data are the noisiest, possibly due to the smaller size of the dataset, leading to wider bins. Despite the noise, a trend, not captured by the median, can instead be seen in the CIs and mean.

Finally, in PN the same scenario as in DC is repeated, with the mean capturing a trend otherwise overlooked by the CIs, further reaffirming that low RH scores correspond to a greater presence of outliers from the (left) tail of the Start Delay distribution, the best-performing activities. Due to the extremely peaked shape of the performance distribution (Fig. 2), the small size of the CIs was indeed to be expected.

As a further step towards validating the effectiveness of the local RH score, we benchmark it against seven other node metrics: in-degree, out-degree, betweenness centrality, closeness centrality, reverse closeness (i.e., closeness centrality computed on the network with edges' direction reversed), number of descendants and of ancestors. For greater robustness, we use all the three performance quantifiers discussed in the Data section (Start Delay, End Delay and Duration Difference) as our target variables. For each of the eight metrics considered, we compute the mutual information between it and the target variable.[3]

For each of the four networks, and for each of the performance indicators, we proceed by computing a two-dimensional frequency matrix with the considered node metric as one dimension and the indicator as the other. For the purpose of computing frequencies, we group data in a number of uniform bins equal to the square root of the number of nodes, rounded down (the same number of bins is used along both dimensions). The mutual information is then computed through the frequency matrix.[4] The results, displayed in Table 4, are strongly consistent across the three performance indicators: the local RH always ranks first for all projects except DC, where it ranks in the top three (second when using End Delay and Duration Difference, third for Start Delay). Overall, the relative ranking of the eight nodes metrics remains largely consistent across the three performance metrics.

Discussion

Project performance can be understood by focusing on how fluctuations spread within the project's underlying activity network. We leverage the context agnostic nature of the approach to develop a new heterogeneity measure (RH) based on the heterogeneity measure introduced by Estrada for undirected networks in Estrada (2010). One of the main advantages of Estrada's measure is the ability to compare networks regardless of their topology, and of their degree distribution in particular. This feature, which is

---

[3] Notice that the notion of target variable has a purely methodological significance in this context: mutual information is symmetric with respect to the 'candidate' and 'target' distributions.

[4] More specifically, the mutual information is computed through the marginal and joint probability distributions for the two variables, as derived from the frequency matrix.

Table 4 Comparison between the local RH score and seven other node metrics

Node metric	Highway	Data centre	Wind farm	Power network
Start Delay				
In-degree	0.287 (8)	0.134 (7)	0.285 (8)	0.045 (8)
Out-degree	0.304 (7)	0.117 (8)	0.293 (7)	0.047 (7)
Betweenness	0.920 (4)	0.250 (6)	0.667 (3)	0.092 (6)
Closeness	1.209 (2)	**0.507 (1)**	0.653 (4)	0.106 (5)
Rev. Closeness	0.975 (3)	0.353 (4)	0.689 (2)	0.123 (4)
Descendants	0.686 (6)	0.274 (5)	0.561 (6)	0.148 (3)
Ancestors	0.812 (5)	0.382 (2)	0.586 (5)	0.149 (2)
Local RH	**1.709 (1)**	0.354 (3)	**0.821 (1)**	**0.208 (1)**
End Delay				
In-degree	0.260 (7)	0.140 (8)	0.237 (8)	0.104 (8)
Out-degree	0.253 (8)	0.174 (7)	0.308 (7)	0.135 (7)
Betweenness	0.828 (4)	0.371 (6)	0.733 (3)	0.159 (6)
Closeness	1.104 (2)	**0.730 (1)**	0.720 (4)	0.211 (5)
Rev. Closeness	0.893 (3)	0.576 (3)	0.753 (2)	0.213 (4)
Descendants	0.609 (6)	0.488 (5)	0.660 (5)	0.219 (3)
Ancestors	0.706 (5)	0.533 (4)	0.650 (6)	0.255 (2)
Local RH	**1.641 (1)**	0.629 (2)	**0.915 (1)**	**0.416 (1)**
Duration Difference				
In-degree	0.222 (7)	0.135 (8)	0.289 (8)	0.080 (8)
Out-degree	0.206 (8)	0.167 (7)	0.335 (7)	0.088 (7)
Betweenness	0.663 (4)	0.393 (6)	0.563 (5)	0.107 (6)
Closeness	0.895 (2)	**0.772 (1)**	0.661 (2)	0.157 (3)
Rev. Closeness	0.726 (3)	0.622 (3)	0.642 (4)	0.156 (4)
Descendants	0.480 (6)	0.505 (5)	0.511 (6)	0.154 (3)
Ancestors	0.519 (5)	0.559 (4)	0.643 (3)	0.196 (2)
Local RH	**1.282 (1)**	0.648 (2)	**0.830 (1)**	**0.315 (1)**

For every candidate, the three table reports its mutual information score computed with Start Delay, End Delay, and Duration Difference as a target variable respectively, and its rank in brackets (the highest ranking score is shown in bold). Local RH ranks first on all datasets minus DC, where it ranks third (Start Delay) or second (End Delay, Duration Difference)

retained in the RH, is particularly desirable when the measure is applied to real-world networks that could in principle take any shape (within their DAG-ness constraints), as in the present study. Furthermore, the importance of a network's heterogeneity in the context of spreading processes makes a measure such as Estrada's, or its extension, a natural candidate for dealing with the problem at hand, namely the analysis of delay propagation when considered as a spreading phenomenon.

Due to their being naturally embedded with a partial ordering, activity networks can be represented as DAGs, a feature which makes it possible, when defining heterogeneity, to shift the focus from first-degree neighbours only to the entirety of a node's ancestry and descendance trees. It is important to notice the particular significance of ancestry in the context of spreading, as the phenomenon at hand (in our case, performance fluctuation) can only propagate downstream; in other fields of applications, ancestry might not play an equally important part. As shown in the Methods section, from a mathematical perspective, the change from first-degree neighbours to ancestors and descendants is a rather straightforward matter when the extension of Estrada's measure to directed graphs is taken as a starting point (Ye et al. 2013).

The very fact that the measure can be used to compare networks of any topology also allows to define a local equivalent to the global RH score, as the same network can be measured before and after the removal of any node, and the two measurements compared. Thus the contribution of individual nodes is obtained "by subtraction". One significant drawback of this approach is that the ancestry trees have to be recalculated every time a node is removed, making the endeavour a computationally expensive one. Here we did not venture into a study of the computational complexity of the calculation, nor of possible ways to reduce it, and the question remains open for potential future work.

We used data from four different projects (a highway, a data centre, a wind farm, and a power network respectively) for the experimental part of our analysis. The size of the datasets varies between schedules, from 1185 for DC to 129 for PN. The networks also have very different component structure, as summarised in Table 1.

In all four cases, frequencies of ancestry size, descendancy size, and performance, take values ranging over various orders of magnitude. The global RH score (Table 2) appears to be particularly effective in quantifying the heterogeneity of the descendancy and ancestry distributions (Fig. 1), with longer-tailed distributions (i.e., more heterogeneous) corresponding to higher RH values.

The distribution of the local RH scores (displayed in Fig. 3) shows, for all networks, a peak in the proximity of the zero value and a single-sided tail (left-sided for WF, right-sided for the other three datasets) dominated by a small number of outliers falling well outside the centre. It is interesting to notice that WF is also the only project to report delays significantly larger than zero. A systematic investigation of the nature of this correspondence, as well as of the relationship between global RH and ancestry (and descendancy) size distribution discussed in the previous paragraph, is beyond the scope of this paper, and might provide the object of future works.

Our experimental results on the four datasets show that a general trend exists, according to which lower RH scores correspond to better performance (Fig. 4). Looking at these results in detail, the cases of DC and PN are particularly interesting, with the mean of the binned data showing a clear trend that the median fails to capture. A similar behaviour is apparent in the other datasets too, though not as pronounced. This is due due to the trend being driven by outliers, i.e., best-performing activities, located in the left tail of the Start Delay distribution; these are activities that take smaller RH values and hence amplify the difference between mean and median values within each bin. Such a feature might prove convenient, considering that a likely purpose of the RH measure is to identify cases of extremely high performance, although the opposite (identifying the poorly performing nodes) might also be the case in some instances. Details on the population of each bin, and on the number of outliers within each bin, are provided in Table 3.

The use of the Start Delay as a performance measure allows us to draw a direct connection between performance and vulnerability to spreading, since it accounts for delays inherited from upstream nodes (as discussed in the Data section). Three out of four projects (excluding WF) follow a similar Start Delay distribution, with a peak around zero and a tail in the negative values (corresponding to better-performing nodes).

As reported in Table 4, we run a comparison between the local RH score and seven other node metrics (in- and out-degree, betweenness centrality, closeness and reverse closeness centrality, number of descendants and of ancestors). The purpose of the

comparison is to quantify which of the candidate metrics carry the most information on node performance; for greater robustness, the same analysis is carried out using Start Delay, End Delay, and Duration Difference as a performance quantifier. To avoid making any assumption on the form of the dependency, we use mutual information, which is a non-parametric measure, capable of accounting for non-linear relationships.

The results are well consistent across the three performance proxies. With the sole exception of DC, where it ranks third or second (depending on the performance indicator considered), the local RH carries the highest mutual information of all the metrics. No other candidate shows the same consistency across datasets; closeness centrality for example, arguably the second-best candidate overall, does always rank first and second on DC and HW respectively, but ranks fourth on WF and fifth on PN by both Start and End Delay. In- and out-degree are always the two worst performing metrics, reinforcing the point that an effective performance measure must look beyond the first-degree neighbourhood, in agreement with the existing literature (Lawyer 2015).

The use of real-world data in our experiments limits our ability to enquire on what network features make the local RH a good proxy for performance, especially when compared to other node metrics. Such features could be better investigated by repeating the analyses presented here on simulated networks. Simulated networks, however, lack ground-truth performance data, an essential component of our experimental setup. A possible compromise could consist in using benchmarks obtained by randomising real-world datasets, although, it must be noted, care must be taken to maintain the DAG structure of the network. In any case, a deeper look into the nature of the relationship between RH and performance, both from an analytical perspective and via further experiments, is likely to provide significant insight towards the study of this metric, and might be the object of future studies.

Conclusions

In the present work, we tackle the question of quantifying and mitigating spreading phenomena from a topological perspective, focusing on how fluctuations in the completion time of certain activities can impact the performance of complex projects. Our contribution is twofold: first, we introduce a novel vulnerability measure that focuses on ancestry tree size, a quantity that plays a big role in spreading process across DAGs; second we apply this measure to an important but currently underrepresented domain - the delivery of complex projects - where we use ground truth data to test our proposed measure.

Using these data, we assess the effectiveness of RH in quantifying performance fluctuations of activities within projects. We show that higher values in RH correspond to worse performance, indicating its appropriateness in accounting for the propensity of such fluctuations to propagate. In addition, we compare RH with seven other node metrics, and show that RH carries the most amount of information about the activity performance on three out of four projects, strengthening its utility in identifying vulnerable nodes.

As well as introducing a new tool for the study of spreading processes on networks, and on directed acyclic graphs in particular, we hope that our work will stimulate the interest of the community in project management as a domain of application for network science.

Abbreviations
DAG: Directed acyclic graph; RH: Reachability-heterogeneity; CI: Confidence interval.

Datasets
HW: Highway; DC: Data centre; WF: Wind farm; PN: Power network.

Availability of data and materials
The datasets analysed during the current study are not publicly available, in accordance with the terms under which access was granted by their owners to the authors, but are available from the corresponding author on reasonable request.

Declarations

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

Author details
[1] Nodes & Links Ltd, 107 Fleet Street, London EC4A 2AB, UK. [2] Birkbeck, University of London, Malet Street, London WC1E 7HX, UK.

References
Baccarini D (1996) The concept of project complexity—a review. Int J Project Manag 14(4):201–204
Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. J Math Sociol 2(1):113–120
Budzier A et al (2011) Why your it project may be riskier than you think. Harv Bus Rev 89(9):23–25
Chinazzi M, Davis J.T, Ajelli M, Gioannini C, Litvinova M, Merler S, y Piontti A.P, Mu K, Rossi L, Sun K et al (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science 368(6489):395–400
Cohen R, Havlin S, Ben-Avraham D (2003) Efficient immunization strategies for computer networks and populations. Phys Rev Lett 91(24):247901
Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Nat Acad Sci 103(7):2015–2020
Davis JT, Perra N, Zhang Q, Moreno Y, Vespignani A (2020) Phase transitions in information spreading on structured populations. Nat Phys 16(5):590–596
Ellinas C (2018) Modelling indirect interactions during failure spreading in a project activity network. Sci Rep 8(1):1–12
Ellinas C (2019) The domino effect: an empirical exposition of systemic risk across project networks. Prod Oper Manag 28(1):63–81
Ellinas C, Allan N, Durugbo C, Johansson A (2015) How robust is your project? From local failures to global catastrophes: a complex networks approach to project systemic risk. PLoS ONE 10(11):0142469
Ellinas C, Allan N, Johansson A (2016) Project systemic risk: application examples of a network model. Int J Prod Econ 182:50–62
Ellinas C, Allan N, Johansson A (2016) Toward project complexity evaluation: a structural perspective. IEEE Syst J 12(1):228–239
Erkol Ş, Faqeeh A, Radicchi F (2018) Influence maximization in noisy networks. EPL (Europhys Lett) 123(5):58007
Estrada E (2010) Quantifying network heterogeneity. Phys Rev E 82(6):066102
Evrard D, Nieto-Rodriguez A (2004) Boosting business performance through programme and project management. PriceWaterhouseCoopers, London
Flyvbjerg B (2007) Cost overruns and demand shortfalls in urban rail and other infrastructure. Transp Plan Technol 30(1):9–30
Flyvbjerg B, Skamris Holm MK, Buhl SL (2003) How common and how large are cost overruns in transport infrastructure projects? Transp Rev 23(1):71–88
Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41
Gomez-Rodriguez M, Leskovec J, Krause A (2012) Inferring networks of diffusion and influence. ACM Trans Knowl Discov Data (TKDD) 5(4):1–37

Groendyke C, Welch D, Hunter DR (2011) Bayesian inference for contact networks given epidemic data. Scand J Stat 38(3):600–616

Guo N, Guo P, Dong H, Zhao J, Han Q (2019) Modeling and analysis of cascading failures in projects: a complex network approach. Comput Ind Eng 127:1–7

Jacobs MA, Swink M (2011) Product portfolio architectural complexity and operational performance: incorporating the roles of learning and fixed assets. J Oper Manag 29(7–8):677–691

Karsai M, Perra N, Vespignani A (2014) Time varying networks and the weakness of strong ties. Sci Rep 4:4001

Lawyer G (2015) Understanding the influence of all nodes in a network. Sci Rep 5(1):1–9

Liu S, Perra N, Karsai M, Vespignani A (2014) Controlling contagion processes in activity driven networks. Phys Rev Lett. https://doi.org/10.1103/PhysRevLett.112.118702

Mihm J, Loch C, Huchzermeier A (2003) Problem-solving oscillations in complex engineering projects. Manag Sci 49(6):733–750

Mishra BK, Haldar K, Sinha DN (2016) Impact of information based classification on network epidemics. Sci Rep 6(1):1–17

Moreno Y, Pastor-Satorras R, Vespignani A (2002) Epidemic outbreaks in complex heterogeneous networks. Eur Phys J B Condens Matter Compl Syst 26(4):521–529

Nadini M, Sun K, Ubaldi E, Starnini M, Rizzo A, Perra N (2018) Epidemic spreading in modular time-varying networks. Sci Rep 8(1):1–11

Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. Rev Mod Phys 87(3):925

Perra N, Gonçalves B, Pastor-Satorras R, Vespignani A (2012) Activity driven modeling of time varying networks. Sci Rep 2:469

Pozzana I, Sun K, Perra N (2017) Epidemic spreading on activity-driven networks with attractiveness. Phys Rev E 96(4):042310

Preciado VM, Zargham M, Enyioha C, Jadbabaie A, Pappas GJ (2014) Optimal resource allocation for network protection against spreading processes. IEEE Trans Control Netw Syst 1(1):99–108

Radicchi F, Castellano C (2016) Leveraging percolation theory to single out influential spreaders in networks. Phys Rev E 93(6):062314

Santolini M, Ellinas C, Nicolaides C (2020) Uncovering the fragility of large-scale engineering projects. arXiv:2009.11752

Sosa ME (2014) Realizing the need for rework: from task interdependence to social networks. Prod Oper Manag 23(8):1312–1331

Stack JC, Bansal S, Kumar VA, Grenfell B (2013) Inferring population-level contact heterogeneity from common epidemic data. J R Soc Interface 10(78):20120578

Sun K, Baronchelli A, Perra N (2015) Contrasting effects of strong ties on sir and sis processes in temporal networks. Eu Phys J B 88(12):1–8

Sun S, Wu Y, Ma Y, Wang L, Gao Z, Xia C (2016) Impact of degree heterogeneity on attack vulnerability of interdependent networks. Sci Rep 6:32983

Terwiesch C, Loch CH (1999) Managing the process of engineering change orders: the case of the climate control system in automobile development. J Prod Innov Manag Int Publ Prod Dev Manag Assoc 16(2):160–172

Ubaldi E, Vezzani A, Karsai M, Perra N, Burioni R (2017) Burstiness and tie activation strategies in time-varying social networks. Sci Rep 7:46225

Valls V, Lino P (2001) Criticality analysis in activity-on-node networks with minimal time lags. Ann Oper Res 102(1–4):17–37

Vanhoucke M (2013) An overview of recent research results and future research avenues using simulation studies in project management. Int Sch Res Not. https://doi.org/10.1155/2013/513549

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151

Wang J, Yang N, Zhang Y, Song Y (2018) Development of the mitigation strategy against the schedule risks of the r&d project through controlling the cascading failure of the r&d network. Physica A 508:390–401

Wasserman S, Faust K et al (1994) Social network analysis: methods and applications, vol 8. Cambridge University Press, Cambridge

Xiao X.-m, Jia L.-m, Wang Y.-h (2018) Correlation between heterogeneity and vulnerability of subway networks based on passenger flow. J Rail Transp Plan Manag 8(2):145–157

Ye C, Wilson RC, Comin CH, Costa LF, Hancock ER (2013) Entropy and heterogeneity measures for directed graphs. In: International workshop on similarity-based pattern recognition. Springer, pp 219–234

Publisher's Note