SCIENTIFIC REPORTS

OPEN

Trip Centrality: walking on a temporal multiplex with non-instantaneous link travel time

Silvia Zaoli, Piero Mazzarisi & Fabrizio Lillo

In complex networks, centrality metrics quantify the connectivity of nodes and identify the most important ones in the transmission of signals. In many real world networks, especially in transportation systems, links are dynamic, i.e. their presence depends on time, and travelling between two nodes requires a non-vanishing time. Additionally, many networks are structured on several layers, representing, e.g., different transportation modes or service providers. Temporal generalisations of centrality metrics based on walk-counting, like Katz centrality, exist, however they do not account for non-zero link travel times and for the multiplex structure. We propose a generalisation of Katz centrality, termed Trip Centrality, counting only the walks that can be travelled according to the network temporal structure, i.e. "trips", while also differentiating the contributions of inter- and intra-layer walks to centrality. We show an application to the US air transport system, specifically computing airports' centrality losses due to delays in the flight network.

Centrality metrics are a useful tool in network analysis to identify and rank the most important nodes or edges of a network. As different concepts of importance can be conceived, several centrality metrics have been proposed[1]. The choice of the most suited one strongly depends on the application, on the type of network, on its structure and on how signals propagate on the network. Many centrality metrics are based on walks, paths, and distances on the network. Some consider only shortest paths and minimum distances, e.g. betweenness and closeness centrality, while others considers walks of all lengths. This is the case, e.g., of Katz centrality[2], PageRank[3] and communicability[4,5], according to which a node's centrality is obtained by summing the contributions of all its outgoing (or incoming) walks, with longer walks being weighted less to reflect their smaller role in connecting the node to the rest of the network. The different metrics differ in the weighting scheme. The choice to consider walks of all lengths is motivated by the observation that processes taking place on the network, e.g. the diffusion of a signal, do not only use optimal paths[5]. Additionally, such choice permits a convenient computation of centrality based on the network's matrix description.

Many real world networks have a temporal structure[6,7], with links characterised by a time of appearance and a duration. In transportation networks, for example, a directed link between the departure and the arrival nodes appears at the time of departure and disappears at the time of arrival. Walks and paths on a temporal network must respect the temporal ordering of links, and this should be accounted for by temporal centrality metrics. While metrics for static networks can be applied to temporal networks after aggregating over time, i.e. considering all links to be present at the same time, clearly this procedure overestimates the number of walks and paths that can be travelled on the network and neglects the effect of the temporal dynamics on the network's connectivity[8,9]. In fact, the links' temporal ordering would be completely disregarded. *Ad hoc* centrality metrics for temporal networks are therefore needed, but how to perform this generalisation is not univocal. In fact, there might be different definitions of time-respecting walks and shortest paths (e.g., shortest according to the number of links used or to the time length), and different generalisation choices might be best suited for different applications. For example, different generalisations of betweenness and closeness centrality have been proposed[10,11], as well as other temporal metrics based on shortest paths and distances[9,12,13].

Metrics based on the counting of walks, like Katz centrality, are of particular interest in the temporal setting, because shortest paths might not be available at all times, therefore increasing the importance of longer itineraries. A natural way to generalise Katz centrality to the temporal case is to count only time-respecting walks. An extension based on this idea, called *Dynamic communicability*, was proposed by Grindrod *et al.*[14]. However the

Department of Mathematics, University of Bologna, Bologna, Italy. Correspondence and requests for materials should be addressed to S.Z. (email: silvia.zaoli@unibo.it)

definition of time-respecting walks used therein does not apply to cases where it takes a non-zero time to travel through a link. Yet, this is true for all transportation networks, for example, where the walk $i \to j \to k$ can be travelled only if the arrival in j (disappearance of the $i \to j$ link) precedes the departure from j to k (appearance of the $j \to k$ link), and therefore link duration should be accounted for. It is also possible to generalise Katz centrality starting from its equivalent formulation, where the centrality of a node is proportional to the sum of the centralities of neighbouring nodes, plus a constant value. This choice has been done in ref.[15], where each time interval is described by a different layer of the network, and each node is therefore a neighbour of itself at subsequent time steps. With this generalisation, however, the interpretation in terms of time-respecting walks is lost.

Additionally to the temporal structure, often networks comprise links of different types and are therefore better described as a multiplex, i.e. a network made of several layers, each containing (a copy of) all the nodes and only the links of a specific type. In transportation networks, for example, layers could represent different transportation services (e.g., bus, metro), service providers (e.g., bus operators or airlines) or lines of a transportation service. The importance of considering the multi-layer character of transportation network has been often stressed, e.g. to correctly characterise their topological properties[16] or assess their resilience in response to failures[17]. In the context of centrality metrics, this multiplex structure plays an important role because intra- and inter-layer walks might contribute differently to a node's centrality. For example, when walks represent itineraries of passengers on the airport network, considering that connections between different airlines are riskier from the passenger's point of view (and therefore less used), an airport should gain more centrality from an intra-layer walk of length n, using all flights of the same airline, rather than from an inter-layer walk of the same length. Similarly, delays propagate more probably within one layer/airline. Centrality metrics for multiplex have been proposed, see e.g. ref.[18] for a review. However, to our knowledge, none permit to weight differently inter- and intra-layer temporal walks.

Here, we propose a new centrality metric generalising Katz centrality to the case of a temporal multiplex with non-zero link travel time. Given the focus of this metric on counting only walks that can actually be travelled given the links schedule, we name it *Trip Centrality*. First, we derive the generalisation starting from the static definition of Katz centrality in terms of walks on the network. Then, we exemplify the fundamental differences between Trip Centrality and Dynamic communicability[14] by applying both metrics to two toy networks with non-zero link travel time and showing the superior performance of Trip Centrality in this case.

A primary application of Trip Centrality is to transportation networks, and in particular we will consider air transport. The description of the air transport system as a network[19,20] proved useful to study its topological characteristics[16,21], resilience[17,22], epidemic spreading[23], and delay propagation[24]. Delays change the timing of links, and their effects in terms of missed connections (and consequently of costs for airlines) depend in a complex way on the schedule. We apply Trip Centrality to the US air transport network, and by comparing the network of scheduled and realised flights we show the capability of the new metric to identify the airports' loss of centrality due to delays, something that is impossible with the static metrics and with existing temporal metrics not accounting for the non-zero link travel time. The loss of centrality of an airport quantifies the itineraries connecting that airport to the rest of the network that become unfeasible due to delays, and therefore assess the effects of delays at the network level.

Results

Definition of Trip Centrality.
Consider a static directed network of N nodes with weighted adjacency matrix A_{ij}, such that $A_{ij} = k$ if there are k links from i to j (for example, flights). The outgoing (incoming) Katz centrality of node i is given by the sum of the contributions of walks of any length outgoing from (incoming to) i, where each walk of length n contributes α^n. Then, given that A_{ij}^n is the number of walks of length n from i to j:

$$k_i^{out} = \sum_{n=0}^{\infty} \alpha^n \sum_{j=1}^{N} (A^n)_{ij} = \sum_{j=1}^{N} \left[\sum_{n=0}^{\infty} (\alpha A)^n \right]_{ij} = \sum_{j=1}^{N} [(\mathbb{I} - \alpha A)^{-1}]_{ij},$$
(1a)

$$k_i^{in} = \sum_{n=0}^{\infty} \alpha^n \sum_{j=1}^{N} (A^n)_{ji} = \sum_{j=1}^{N} \left[\sum_{n=0}^{\infty} (\alpha A)^n \right]_{ji} = \sum_{j=1}^{N} [(\mathbb{I} - \alpha A)^{-1}]_{ji},$$
(1b)

Note that the infinite sums on n in Eq. 1 converge only if α is smaller than the inverse of the largest eigenvalue of A[1].

In a temporal network, links are characterised by their time of appearance and their duration, i.e. the length of the time interval during which they are present. Here, we will take the duration to coincide with the time required for a signal to travel through the link. This assumption is inspired by transportation networks, where a link represents a route of a transportation service, appearing at the departure time and disappearing at the arrival time. A common way to treat temporal networks is to discretise time[14,15] in intervals of length $\Delta\tau$ and define an adjacency matrix $A^{[t]}$ for each time frame. The adjacency matrix $A^{[t]}$ contains only the links that either appear or disappear in the interval $[\tau, \tau + \Delta\tau)$, or are present during the entire interval. With this choice, the temporal network is represented by a series of T adjacency matrices $\{A^{[t]}\}_{t=1,\dots,T}$. The optimal length of a time frame $\Delta\tau$ depends on the dataset, as will be discussed in the following.

In the case of non-zero link travel times, however, the above formulation is not suited to count time-ordered walks. We define a time-ordered walk as a series of links such that the disappearance of a link in the series always precedes the appearance of the following link. In the air transport example, a time-ordered walk is a series of flights such that the i-th lands before the $(i+1)$-th departs. Products $A^{[t_1]} \dots A^{[t_n]}$ of adjacency matrices defined as above would count also walks such that the $(i+1)$-th links appears when the i-th is still present. In order to express the number of time-ordered walks as a product of adjacency matrices, we introduce a set of secondary
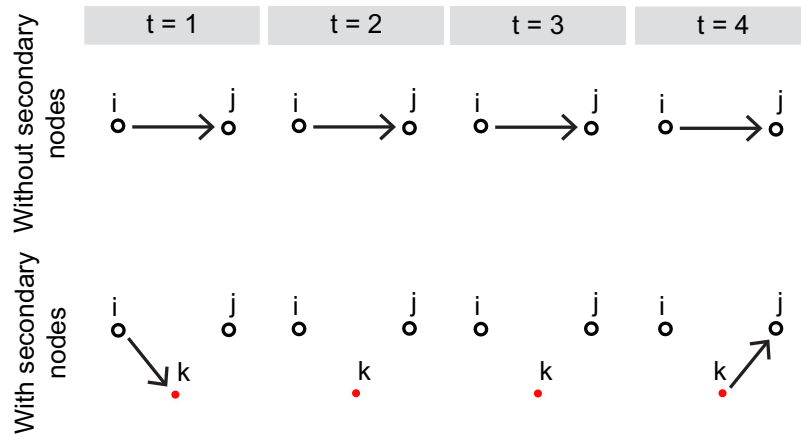
Figure 1. Example showing how secondary nodes are introduced. Above, a link between the primary nodes *i* and *j*, appearing during time frame 1 and disappearing at time frame 4. Below, the same link is represented by two stubs, one at time frame 1 between *i* and the secondary node *k*, and one at time frame 4 between *k* and *j*.

nodes, one for each of the N_l links present over the whole considered period (see Fig. 1). We therefore consider a network with $N + N_l$ nodes, of which N are the original, or primary, nodes and N_l are the secondary ones. Secondary nodes do not represent physical locations but rather the "journey" between two nodes. In this new network, a link from *i* to *j*, appearing during time frame *t* and disappearing during time frame $t' > t$, is associated with a secondary node *k* and split into two links, called 'stubs' in the following, one from *i* to *k* present during time frame *t*, and one from *k* to *j* present during time frame t'. We remark that, differently from links, stubs do not have a duration, i.e. they exist only in one time frame, and in the time frames between the appearance and disappearance of a link no stub related to that link is present in the network. The duration of the travel from *i* to *j* is therefore given by the time interval between the two time frames in which the two stubs appear. For each time frame $[\tau, \tau + \Delta\tau)$ we define an adjacency matrix $A^{[t]}$ of size $(N + N_l) \times (N + N_l)$ such that $A_{ij}^{[t]} = 1$ either if a link outgoing from node *i* appears during that time frame, and *j* is the secondary node associated with it, or if a link incoming to node *j* disappears during that time step, and *i* is the the secondary node associated with it. In the air transport example, for each flight a directed stub between the origin airport and its secondary node is present in the time frame in which the flight's departure time falls, while a directed stub between its secondary node and the destination airport is present in the time frame in which the arrival time falls. Secondary nodes ensure that matrix products of the form $A^{[t_1]} A^{[t_2]} \ldots A^{[t_n]}$, with $t_1 < t_2 < \ldots < t_n$ (thus, without repetitions, meaning that at most one link is used per time frame) count only time-ordered walks, in the sense defined above. The time frames $t_1, \ldots, t_n$ do not need to be consecutive, as a walk can pause at a node for some time frames before continuing. We note that the idea of secondary nodes to describe temporal networks with non-instantaneous transport has been suggested previously[25,26], but never applied to the computation of centrality metrics. The length $\Delta\tau$ of one time frame must be shorter than the duration of the shortest link, so that the two stubs in which the link is split belong to different time frames and can be both used in a walk. With this definition of the series of adjacency matrices, the *i*,*j*-th element of the matrix

$$Q = [(\mathbb{I} + \widetilde{\alpha} A^{[1]})(\mathbb{I} + \widetilde{\alpha} A^{[2]}) \ldots (\mathbb{I} + \widetilde{\alpha} A^{[T]}) - \mathbb{I}], \tag{2}$$

contains the contribution to centrality of all walks from *i* to *j*. In fact, *Q* contains all the time-ordered products of *n* adjacency matrices, for *n* = 1, ..., *T*. Note that $\widetilde{\alpha}$ is the weight of a one-stub walk, therefore a one-link walk, which uses two stubs, is weighted $\alpha = \widetilde{\alpha}^2$. The vectors of temporally generalised outgoing and incoming Katz centrality are then obtained summing *Q*, respectively, over columns and rows:

$$\overrightarrow{t}_S^{out} = Q \, \overrightarrow{1}_{N+N_l}, \tag{3a}$$

$$\overrightarrow{t}_S^{in} = \overrightarrow{1}_{N+N_l}^T Q, \tag{3b}$$

where $\overrightarrow{1}_{N+N_l}$ is a column vector of ones and the subscript *S* indicates that this is a single layer quantity. We refer to these centralities as *Single-Layer Trip centralities*. Note that, differently from the static case, there is no upper limit for the parameter $\widetilde{\alpha}$, as the sum is always bounded. In fact, there are no walks longer than *T*. See SI for an estimate of the computational complexity of the algorithm.

Let us now consider the multiplex structure of the temporal network. We associate one layer with each type of link existing in the network. The purpose of this further generalisation is to distinguish the walks made of links of the same type from those made of links of different types, which might give a different contribution to the centrality (length being equal), as motivated in the introduction. As the multiplex has a copy of each primary node on each layer, the adjacency matrix *A* is of size $(NN_L + N_l) \times (NN_L + N_l)$, where N_L is the number of layers. Each

secondary node exists only in the layer associated with the corresponding link. For each link of type λ a directed stub is present, in the time frame corresponding to its appearance, between the copy of the origin node on layer λ and its secondary node. A directed stub between its secondary node and the copy of the destination node on layer λ is present in the time frame corresponding to its disappearance.

Let us introduce the parameter $\varepsilon \leq 1$, such that the contribution to centrality of a walk of length n changing layer m times is $\varepsilon^m \alpha^n$. The effect of this parameter is that, the more changes of layer a walk has, the less it contributes to centrality. Let us then introduce the matrix K, of the same size of A, as the matrix with elements $K_{ii} = 1$ and $K_{ij} = \varepsilon$ if i and j are two copies of the same node on different layers. Now, the products of the form $A^{[t_1]} K A^{[t_2]} K \ldots K A^{[t_n]}$ count walks by introducing a factor ε every time there is a change of layer. Therefore, the outgoing Trip Centrality on the temporal multiplex is written as

$$\overrightarrow{t}^{out} = [(\mathbb{I} + \widetilde{\alpha} A^{[1]} K)(\mathbb{I} + \widetilde{\alpha} A^{[2]} K) \ldots (\mathbb{I} + \widetilde{\alpha} A^{[T]} K) - \mathbb{I}] K^{-1} \overrightarrow{1}_{(NN_L + N_l)}. \tag{4}$$

The incoming centrality is generalised similarly.

With this procedure, we obtain a centrality measure for each copy of each node, which sums the contributions of walks outgoing from (or incoming to) that node with a link on the layer on which that copy lies. Such *layer-specific centrality* is interesting if we want to measure the importance of a node for, e.g., a specific transportation service or service provider. However, if we are interested in the role of the node for the entire network, an aggregated measure is needed. The interpretation of centrality in terms of walks provides us with a natural way to perform such an aggregation. In fact, if we compute the *aggregated centrality* of a node by summing the centralities of all its copies, we obtain the contribution of all walks outgoing from (or incoming to) any copy of that node. Additionally, the centrality of a secondary node represents the centrality of the corresponding link. Given that links between the same two nodes having different schedules (e.g., flights between the same two airports at different times of the day) are associated with different secondary nodes, the centrality of their secondary nodes can show how the importance of a link changes depending on its schedule.

The parameter ε determines how much inter-layer walks are penalised. In the case of transportation networks, therefore, it measures the propensity of users to change layer. In particular, the case $\varepsilon = 0$ corresponds to not allowing inter-layer walks. In this case, the layer-specific centralities are those that consider only the links of one type. The aggregated centralities consider links of all layers, but only intra-layer walks contribute. When instead $\varepsilon = 1$, inter-layer walk give the same contribution of intra-layer walks. For $0 < \varepsilon < 1$, inter-layer walks are considered, but contribute less than an intra-layer walk of the same length. Note that the matrix K is invertible for every $\varepsilon \neq 1$. The final multiplication by K^{-1} in equation (4) only changes the aggregated centralities by a multiplicative factor, therefore it does not change the node's ranking (see SI, section 1). Thus, the ranking according to the aggregated centralities in the case $\varepsilon = 1$ can be obtained skipping the final multiplication by K^{-1}.

A generalisation along the same lines is also possible for PageRank, a centrality metrics developed by Google[3] which can also be defined in terms of walks. In PageRank, an additional weight is assigned to the walks, depending on the in- (or out-) degree of the nodes they cross. See section 3 of the SI for details.

Comparison with Dynamic communicability. In this section, we show the difference between the proposed Trip Centrality and Dynamic communicability, a previously proposed temporal generalization of Katz centrality[14], by applying them to two simple temporal networks. The purpose of the examples is to show that, in the case of non-zero link travel time, Trip Centrality overcomes some limitations of Dynamic communicability related to time-respecting walks and to the time-discretisation. Note that two versions of Dynamic communicability are presented in ref. [14], one allowing multiple jumps during each time frame, and one allowing only one, similarly to Trip Centrality. We consider the second version for the comparison.

First, we consider the network in Fig. 2a.I, consisting of four nodes on one layer. The links' temporal dynamics, in Fig. 2a.II, is such that it is possible to move from i to k respecting the schedule, but not from l to k. In fact, link c from l to m disappears only after the appearance of link d from m to k. In a network where the link duration coincides with the link travel time, this means that the connection between links c and d cannot be taken. Instead, it is possible to use links a and b in sequence to go from i to k. Therefore, the outgoing centrality of node i should be larger than the one of l, as it has one outgoing itinerary using one link and one using two links, while k has only one outgoing itinerary, using one link. While Trip Centrality agrees with this intuition, according to Dynamic communicability l has a larger outgoing centrality than i. In fact, this metric does not recognise that the path from l to k is not time-respecting. Therefore, the ranking reflects itineraries that cannot actually be travelled. See Methods for the exact computations.

The example in Fig. 2b, instead, shows that without secondary nodes the choice of the length $\Delta\tau$ of a time frame can affect the ranking, while with Trip Centrality it has no effect, as long as it is shorter than the shortest link duration. In the example, i is connected to j by a link of duration d and j is connected to k by a link of duration nd, appearing after an interval md from the disappearance of the first link. For simplicity, we take n and m integers. The outgoing centrality of node i should clearly be larger than the outgoing centrality on node j. However, while Trip Centrality agrees with this intuition, the ranking obtained with Dynamic communicability depends on the choice of $\Delta\tau$, if $\alpha \leq (n-1)/2n$. In particular, if $\Delta\tau$ is larger than a certain threshold, j will be ranked as more central than i. For example, if $n = 2$ and $\alpha \leq 1/4$, choosing $\Delta\tau$ shorter than $d/2$ would result in ranking j as the most central. See Methods for details.

These two simple examples make clear that, when links have a non-zero link travel time, the introduction of secondary nodes is necessary in order to compute centralities that actually reflect the itineraries that can be travelled on the network.
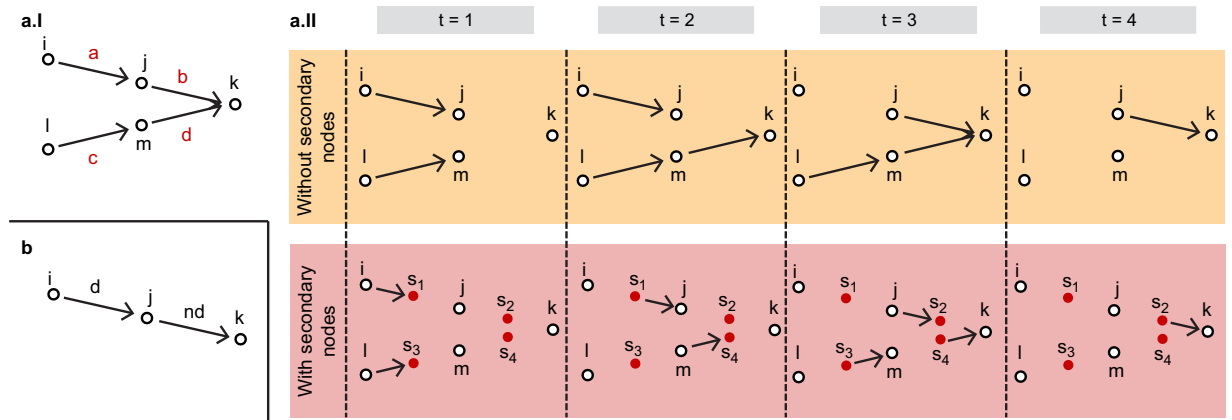
Figure 2. Two examples illustrating the difference between Trip Centrality and Dynamic communicability in ranking nodes on temporal networks with non-zero link travel time. (**a.I**) Time-aggregated representation of a single-layer temporal network. Link *a* is present during the time frames 1 and 2, link *b* during the time frames 3 and 4, link *c* from time frame 1 to time frame 3, and link *d* during the time frames 2 and 3; (**a.II**) Time-explicit representation of the network in panel a.I. Above, the representation without the use of secondary nodes, below with; (**b**) Time-aggregated representation of a single-layer temporal network. The link from *i* to *j* has a duration *d*, the link from *j* to *k* has a duration *nd* and appears after an interval *md* from the disappearance of the former.
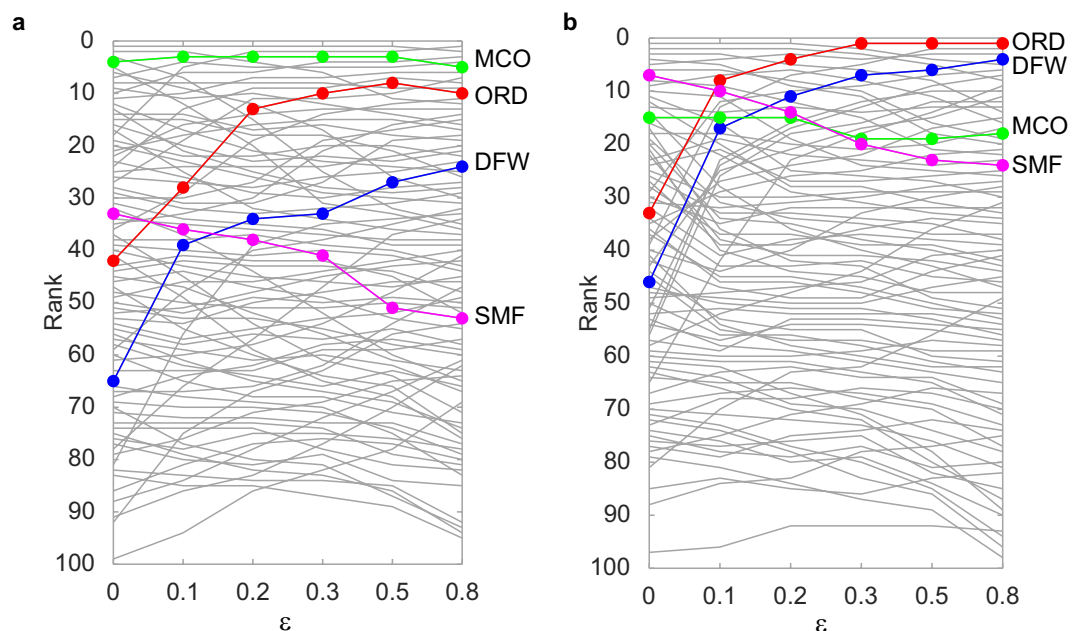

Figure 3. Evolution of the airports' ranking according to outgoing (**a**) and incoming (**b**) Trip Centrality on the scheduled network on April 1st for $\alpha = 0.2$ and different values of ε. Only the airports ranking 100 or higher are represented. Each line represents one airport, and the position on the *y*-axis indicates its rank for each ε value. Rank 1 corresponds to the most central airport. See text for comments on the highlighted lines.

Application to the air transport system. *Data.* The dataset used for the following analysis was obtained from the US Department of Transportation's (DOT) Bureau of Transportation Statistics and contains the flights operated in April 2015 by 14 major US airlines. See Methods for further details. The dataset comprises 322 airports. The temporal multiplex is therefore made of 14 layers (one per airline), each with 322 primary nodes. The length of a time frame is chosen as $\Delta \tau = 20$ min.

Comparison of airports' ranking for different values of ε. The parameter $\varepsilon \in [0, 1]$ measures the propensity of travellers to use different airlines for different legs of one trip. When ε increases, centralities always increase, as the weight of inter-layer walks is increased. However, the airports' ranking may change, as some airports rely more on inter-layer walks for their connectivity. Figure 3 shows how the airports' ranking according to outgoing and incoming Trip Centrality on April 1st changes when $\varepsilon = 0, 0.1, 0.2, 0.3, 0.5, 0.8$, for the airports ranked 100
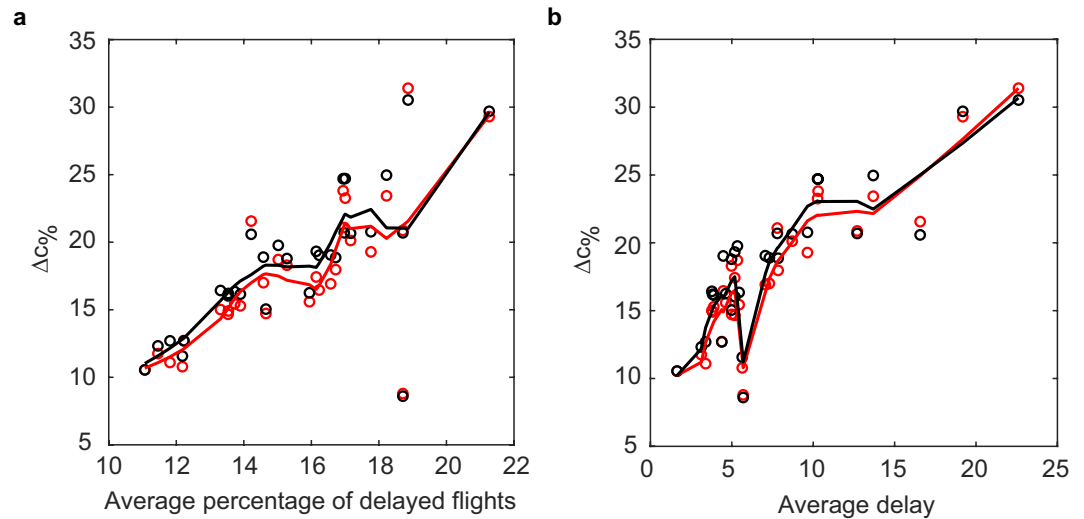
Figure 4. Percentage of centrality loss, averaged over all airports, in each day of the dataset, according to incoming Trip Centrality (red) and outgoing Trip Centrality (black) plotted against average departure delay (**a**) and average fraction of flights with departure delay in one airport (**b**). Trip centrality is computed with $\alpha = 0.2$ and $\varepsilon = 0$. Each point corresponds to one day of the dataset. The percentage centrality loss of an airport is computed as $\Delta c_\% = 100 \times (c_{sched} - c_{act})/c_{sched}$, where c_{sched} and c_{act} are the airport's centralities on the scheduled and realised network. Lines are obtained by a locally weighted smoothing (LOWESS) of the dots of the correspondent colour.

or higher (for the behaviour of all airports see Supplementary Fig. S1). Results for other days of April are qualitatively similar. Note that Trip Centrality is computed with $\alpha = 0.2$, therefore when $\varepsilon < 0.2$ a walk using n flights with one inter-layer jump is weighted less than an intra-layer walk made of $n + 1$ flights. The contrary is true when $\varepsilon > 0.2$. Some airports increase steadily their rank when ε increases. These airports would gain from an increased cooperation of airlines, making it easier for passengers to use inter-layer walks. Examples of airports gaining rank according to both incoming and outgoing Trip Centrality on every day of April are the Chicago O'Hare Airport (highlighted in red in the figures), the Dallas-Fort Worth International Airport (highlighted in blue), the George Bush Houston Airport, the J. F. Kennedy International Airport. Other airports either maintain their rank, e.g. the Orlando International Airport (green), or decrease it, e.g. the Sacramento International Airport (magenta). Note that such rank decrease is not due to a decrease of the airports' centrality, but simply to their being overtaken by others.

Comparing the network of scheduled and realised flights. In many situations, especially in transportation, there exists a scheduled and a realised network, differing due to delays and cancellations. Delays can cause the disruption of connections between flights which were feasible in the scheduled network, therefore diminishing the potential to move through the network. In this section, we show that Trip Centrality is able to quantify the loss of connectivity of a node due to the delays, differently from static centrality metrics, which do not account for the temporal structure.

For each day, we compute each airport's aggregated Trip Centrality on the scheduled and realised networks. Note that delays can also cause the opening of new walks on the realised network, which were not present in the scheduled one. While in bus or metro networks these new walks could be used by passengers, in air traffic this is not possible, therefore these 'forbidden' walks should be excluded from the computation. The procedure to do so is detailed in the Methods section. Then, the loss of centrality of airport i is computed as the difference between its centrality in the scheduled and realised network. Figure 4 plots the percentage centrality loss, averaged over all airports, for each day against two delay-related indicators characterising that day: the average delay of all flights on that day and the average fraction of delayed flights in an airport. The average percentage centrality loss increases with both indicators, meaning that, in aggregate, more or larger delays in the network cause larger centrality losses. Instead, the correlation between the rankings on the scheduled and realised networks decreases with both indicators (see Supplementary Fig. S2). These effects are due both to delays and to cancelled and diverted flights (which are increasing with both indicators). However, Supplementary Fig. S3 shows that the patterns are still present when cancelled and diverted flights are excluded from the analysis, proving that the proposed centrality metric is able to detect the changes in the network's connectivity due to delays, or, more in general, to a change in the links' temporal structure. Results are robust with respect to changes of α (see Supplementary Figs S4 and S5).

While average centrality losses tend to be larger in days with more or larger delays, at the level of a single airport they are weakly correlated with delays in that airport. Let us call an airport *distressed* when the fraction of its flights with delay larger than average is larger than the average fraction of delayed flight in an airport on the analysed day. Figure 5 shows that the percentage centrality losses of distressed airports are not larger than those
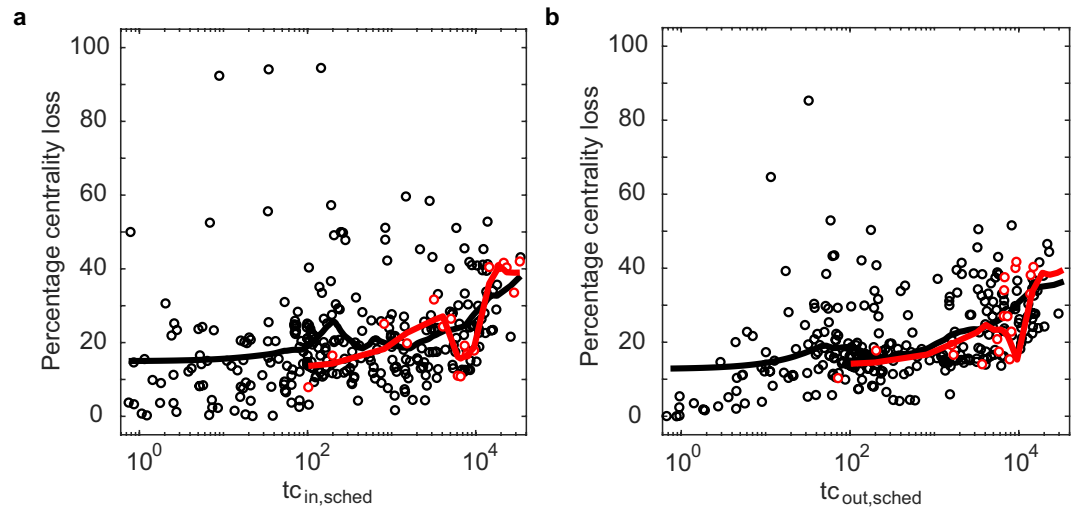
Figure 5. Percentage of Trip Centrality loss plotted against Trip Centrality in the scheduled network, for April 1st. (**a**) incoming, (**b**) outgoing. Trip centrality is computed with $\alpha = 0.2$ and $\varepsilon = 0$. Red dots represent distressed airports (see text for definition), black dots non-distressed ones. Lines are obtained by a locally weighted smoothing (LOWESS) of the dots of the correspondent colour.

of non-distressed ones. The figures refer to April 1st, results for other days are similar. Such results underline that the loss of Trip Centrality reflects the network effects of delay, which do not depend simply on the delay itself. In fact, the same delay can have a large network effect if it causes the disruption of several connections, or no effect at all if no connection is disrupted. While larger delays have a larger probability to cause a disruption, there is no simple relationship between the amount of delay and the number of missed connections, and therefore centrality loss, which depend on the schedule.

Comparison of existing centrality metrics with Trip Centrality. Here, we compare the ranking of airports according Trip Centrality with those according to Katz centrality and Dynamic communicability. Rankings are compared computing the Kendall rank correlation coefficient τ, which measures the similarity of two ranked sequences. The coefficient takes values in $[-1, 1]$, with 1 corresponding to two identical sequences and -1 to two sequences that are one the inverse of the other.

For Katz centrality we choose the largest value of α ensuring convergence, $\alpha = 0.003$. We remark that such very small value strongly penalises long walks (more than 300 walks of length 2 are needed to contribute as much as one walk of length 1), making the ranking according to Katz similar to a simple ranking according to the number of flights (average correlation coefficient 0.9, see Supplementary Fig. S6 for a comparison of the rankings for April 1st). Dynamic communicability is computed according to equation (5), where no restriction on α applies. All metrics are computed on the scheduled network.

Katz centrality and Trip Centrality produce similar rankings, on average, only when the latter is computed with a very small α (see Fig. 6a), as walks longer than one (in whose counting they differ) have a negligible weight. When Trip Centrality is computed with $\alpha = 0.003$, i.e. the same as Katz centrality, and $\varepsilon = 0$, the average correlation coefficient is 0.91 ± 0.01, both for the incoming and the outgoing case. However, they become less and less similar when the value of α used in the computation of Trip Centrality becomes larger, increasing the importance of longer walks. Increasing ε above 0.1 decreases the similarity (see Fig. 6b). Supplementary Fig. S7a shows, as an example, the comparison of the rankings according to incoming Katz and Trip Centrality for April 1st, with Trip Centrality's parameters $\alpha = 0.2$ and $\varepsilon = 0$.

Dynamic communicability and Trip Centrality, computed with the same value of α, become more similar when ε approaches 1 (see Fig. 6d, with $\alpha = 0.2$). This is understandable, since Dynamic communicability does not distinguish layers. However, for no value of ε they are very similar. For $\alpha = 0.2$, for example, they approach an average correlation of $\tau = 0.7$ when ε approaches 1. The value of α does not influence much their correlation (see Fig. 6c). Supplementary Fig. S7b shows, as an example, the comparison of the rankings according to incoming centrality for April 1st, with $\alpha = 0.2$ and Trip Centrality with $\varepsilon = 0$. Interestingly, for any parameter choice, Trip Centrality is more similar to Katz than to Dynamic communicability. This is explained by the fact that Trip and Katz centralities agree on the counting of one-flight itineraries, which give the larger contributions, while with Dynamic communicability longer flights count more than shorter ones, as they appear in more time frames.

To summarise, when itineraries of more than one flight are assigned a non-negligible weight, i.e. α not too small, accounting for the temporal multiplex structure of the network and for the non-zero link travel time does indeed make a significant difference in the ranking. The ten most central airports according to each metric are reported in Supplementary Table 1.
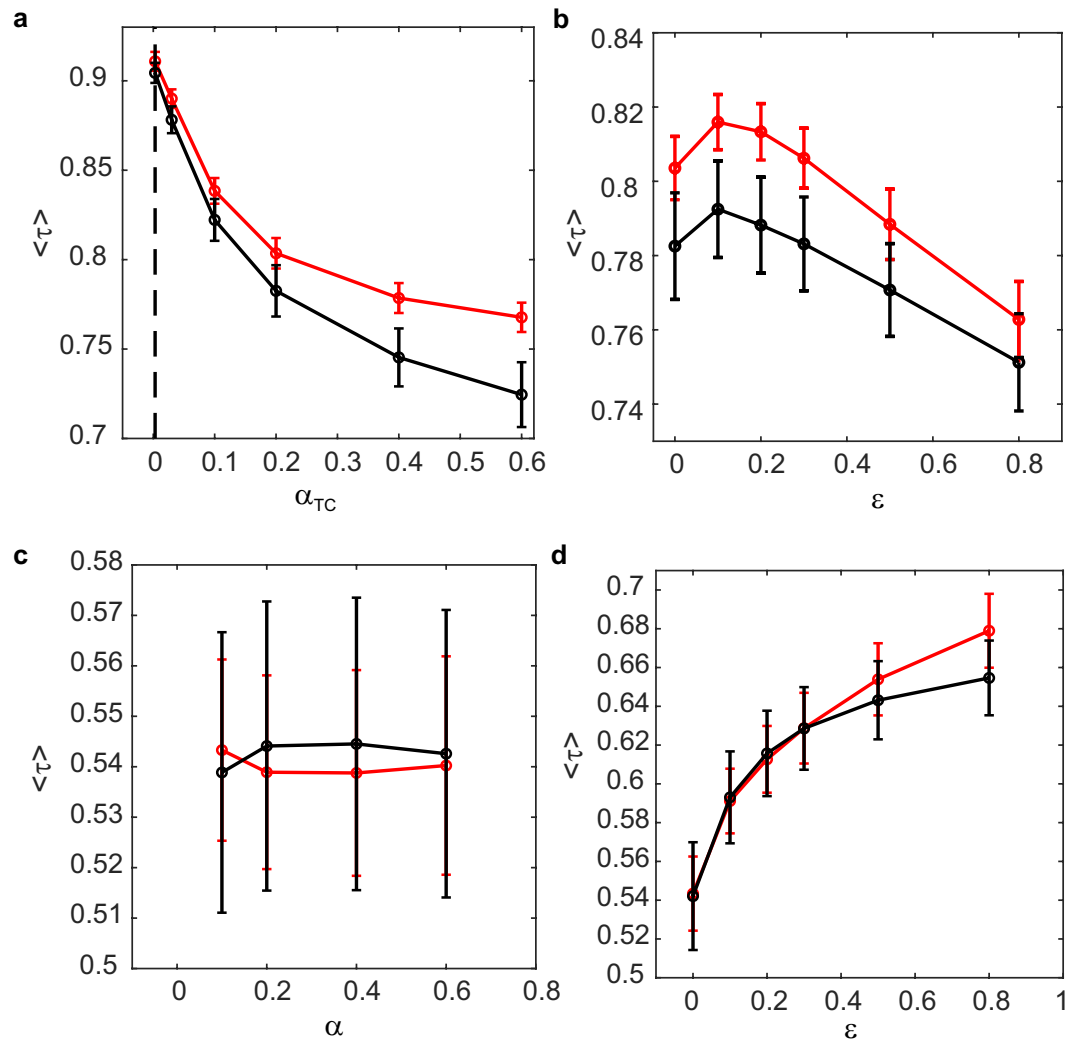
Figure 6. (**a**) Kendall correlation coefficient, averaged on all days, between the rankings according to Katz centrality with $\alpha = 0.003$ and Trip Centrality, for different values of α used in the computation of Trip Centrality. The red line corresponds to the incoming centralities and the black one to the outgoing. The dotted line marks the value of α used for Katz centrality; (**b**) Kendall correlation coefficient, averaged on all days, between the rankings according to Katz centrality with $\alpha = 0.003$ and Trip Centrality with $\alpha = 0.2$, for different values of ε used in the computation of Trip Centrality. Colors as in panel a; (**c**) Kendall correlation coefficient, averaged on all days, between the rankings according to Dynamic communicability and Trip Centrality for different values of α. Trip Centrality is computed with $\varepsilon = 0$. Colors as in panel a; (**d**) Kendall correlation coefficient, averaged on all days, between the rankings according to Dynamic communicability and Trip Centrality with $\alpha = 0.2$, for different values of ε used in the computation of Trip Centrality. Colors as in panel a. Bars represent standard errors.

Discussion

Trip Centrality generalises centrality to networks where links are dynamic, characterised by a non-zero travel time and of different types. With respect to existing centrality metrics, Trip Centrality assigns centrality to nodes according to the schedule-respecting itineraries that can be travelled on the network, with the possibility of weighting less the inter-layer ones, which use links of several types. The importance of accounting for the links' duration, through the introduction of secondary nodes, was proved by two simple examples. The metrics retain the computational convenience of the original static metrics, requiring only simple matrix multiplications. Additionally, with respect their static counterparts, there is no constraint on the value of the weight α to assign to a link, as there is no convergence issue. Therefore, α can be freely chosen as the value better reflecting the relative importance of walks of different length.

Transportation networks are the most direct application for Trip Centrality, which is however more general, being suited for any temporal network with non-zero link travel time. For example, our approach could be applied to study epidemic spreading on a temporal network when an incubation time is required before an infected individual can infect susceptible ones, or the diffusion of information by email, when emails are not read instantaneously.

The proposed approach to combine the multiplex structure with the temporal one, i.e. the use of the matrix K, can be used even in the case of instantaneous transportation by applying it without the introduction of secondary nodes. We remark that Trip Centrality, if applied on a single layer and without secondary nodes, reduces to the version of Dynamic communicability allowing only one jump per time frame.

Trip Centrality can be used to assess the change in a node's centrality following the change of the links' time dynamics, and also to explore how the interaction of different layers (parameter ε) affects the centralities. In application to the US air transport system, Trip centrality proved able to capture that, on average, the connectivity of the airports and flights network diminishes the more delays there are on the network (i.e. the average percentage centrality loss becomes larger). This could not be captured by static metrics, neglecting the time dynamics. Additionally, Trip Centrality quantified the effect of a delay on a single airport in terms of loss of network connectivity, showing that it depends in a complex, non-linear way on the delay itself. In conclusion, Trip Centrality provides a tool to assess the effect of delays on the connectivity of transportation networks at the whole network level as well as at the single node level.

Methods

Computation of centralities in the example in Fig. 6a.
According to Dynamic communicability, the vector of outgoing centralities is

$$\overrightarrow{c}^{out} = [(\mathbb{I} + \alpha A^{[1]})(\mathbb{I} + \alpha A^{[2]})...(\mathbb{I} + \alpha A^{[T]}) - \mathbb{I}] \overrightarrow{1_N}, \tag{5}$$

where $A^{[t]}$ is the $N \times N$ adjacency matrix without secondary nodes. The links present at each time frame in this adjacency matrix are represented in Fig. 6a.II. To compute the outgoing centralities of i and l we can count the outgoing walks: i has two outgoing walks of length 1 to j, counted by $A^{[1]}$ and $A^{[2]}$, and four of length 2 to k, counted by the products $A^{[1]}A^{[3]}$, $A^{[1]}A^{[4]}$, $A^{[2]}A^{[3]}$ and $A^{[2]}A^{[4]}$. Its centrality is therefore $c_i^{out} = 2\alpha + 4\alpha^2$. Instead, l has three outgoing walks of length 1, and still 4 of length 2, therefore its centrality is $c_l^{out} = 3\alpha + 4\alpha^2$, larger than the centrality of i.

Introducing secondary nodes, Trip Centrality can be computed according to equation (3), as we are working on a single layer. The adjacency matrices are now $(N+4) \times (N+4)$. The centrality of the two nodes are then, $tc_i^{out} = \widetilde{\alpha} + \widetilde{\alpha}^2 + \widetilde{\alpha}^3 + \widetilde{\alpha}^4 = \widetilde{\alpha} + \widetilde{\alpha}\alpha + \alpha + \alpha^2$ and $tc_l^{out} = \widetilde{\alpha} + \widetilde{\alpha}^2 = \widetilde{\alpha} + \alpha$, where we used $\alpha = \widetilde{\alpha}^2$. Therefore, Trip Centrality correctly finds that i has a larger outgoing centrality than l.

Computation of centralities in the example in Fig. 6b.
Let $\Delta\tau$ be the length of a time frame, such that $\Delta\tau < d$. For simplicity, let us assume that $\Delta\tau = d/\Delta$, with Δ an integer larger than 1. Therefore, in the formulation without secondary nodes, the link from i to j is present in Δ time frames and the link from j to k in $n\Delta$. It is then possible to go from i to j with Δ different walks of length 1, from j to k with $n\Delta$, and from i to k with $n\Delta^2$ different walks of length 2. Therefore we have $c_i^{out} = \alpha\Delta + \alpha^2 n\Delta^2$ and $c_j^{out} = \alpha n\Delta$. Now, $c_i^{out} > c_j^{out}$ when $\Delta > (n-1)/\alpha n$. If $\alpha > (n-1)/2n$ the condition is always realised (because $\Delta \geq 2$), and i is more central than j for any Δ. If $\alpha \leq (n-1)/2n$, instead, depending on the value of Δ, i.e. on $\Delta\tau$, i can be less or as central as j. This dependency on $\Delta\tau$ is not present in Trip Centrality, according to which the two centralities are $tc_i^{out} = \widetilde{\alpha} + \widetilde{\alpha}^2 + \widetilde{\alpha}^3 + \widetilde{\alpha}^4$ and $tc_j^{out} = \widetilde{\alpha} + \widetilde{\alpha}^2$, which do not depend on $\Delta\tau$.

US air traffic dataset.
The dataset made available by the US Department of Transportation's (DOT) Bureau of Transportation Statistics was obtained from https://www.kaggle.com/usdot/flight-delays. It includes all flights operated in 2015 by 14 major US airlines. In the multiplex of airports and flights, each airline is treated as a separate layer. Note that, in general, if two airlines belong to the same alliance they should be included in the same layer. In fact, from the passengers' point of view, connections between their flights are analogous to within-airline connections. None of the 14 airlines of the US dataset, however, belong to the same alliance.

The present analysis was performed on the month of April, which is quite heterogeneous in terms of delays, ranging from days with few small delays to days with many large ones. For each flight in the dataset, the following information is available: Date, Flight number, Tail number, Origin airport, Destination airport, Scheduled departure time, Realised departure time, Scheduled arrival time, Realised arrival time, Airline, Cancellation (1 if was cancelled, 0 otherwise), Diverted (1 if it was diverted, 0 otherwise). All times are converted from local time to Eastern Standard Time (EST). Days are considered to start at 4 AM EST, as this is the time of minimum traffic across the entire US.

Loops.
The walks counted by Trip Centrality include loops. Depending on the application, this might be wanted or not. In the air transport case, walks with loops might be important for delay propagation, but certainly not for passengers. Walks that pass a second time from their starting point can be excluded from the counting (see SI, section 4), and this leaves the ranking of US airports almost unchanged. Therefore, we claim that considering or not loops does not change significantly the results obtained.

Centrality in the realised network.
We call $\{A_{sched}^{[t]}\}_{t=1,...,T}$ and $\{A_{real}^{[t]}\}_{t=1,...,T}$, respectively, the adjacency matrices defined at each time frame for the networks of scheduled and realised flights. In the realised network the timing of links is changed due to delays, cancellations and diversions, therefore some of the time-oriented walks that existed in the scheduled network will not be present anymore, causing losses of centrality with respect to the scheduled network. Note that, in principle, delays could also create new walks, by allowing connections that were impossible in the scheduled network. In the case of air transport, differently from the bus or metro networks, such new walks cannot be used by passengers (except in the case of rerouting). Therefore, if we want to assess the

connectivity of an airport from the passengers' point of view, we should not consider the contribution of these 'forbidden' walks. The forbidden walks can appear in two cases. The first case is when a negative arrival delay (early arrival) allows a connection with a departing flight. In this case, the problem is solved by setting the negative arrival delay to zero, i.e., setting the arrival time equal to the scheduled arrival time. The second case is when the delay of a departing flight allows the connection with an incoming flight, which was originally landing too late to make the connection. A partial solution which eliminates most of the forbidden walks can be implemented and is explained in section 6 of the SI.

Data Availability

The dataset used in the analysis is available at https://www.kaggle.com/usdot/flight-delays.

References

1. Newman, M. *Networks: An Introduction* (Oxford University Press, 2010).
2. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43, https://doi.org/10.1007/BF02289026 (1953).
3. Brin, S. & Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **30**, 107–117, https://doi.org/10.1016/S0169-7552(98)00110-X (1998).
4. Estrada, E. & Hatano, N. Communicability in complex networks. *Phys. Rev. E* **77** (2008).
5. Estrada, E., Hatano, N. & Benzi, M. The physics of communicability in complex networks. *Phys. Reports* **514**, 89–119 The Physics of Communicability in Complex Networks (2012).
6. Holme, P. & Sarämaki, J. Temporal networks. *Phys. Reports* **519**, 97–125, https://doi.org/10.1016/j.physrep.2012.03.001 Temporal Networks (2012).
7. Holme, P. Modern temporal network theory: a colloquium. *The Eur. Phys. J. B* **88**, 234 (2015).
8. Tang, J., Musolesi, M., Mascolo, C. & Latora, V. Temporal distance metrics for social network analysis. *Proc. 2nd ACM workshop on Online social networks* (2009).
9. Pan, R. K. & Saramäki, J. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E* **84**, 016105, https://doi.org/10.1103/PhysRevE.84.016105 (2011).
10. Tang, J., Musolesi, M., Mascolo, C., Latora, V. & Nicosia, V. Analysing information flows and key mediators through temporal centrality metrics. *Proc. 3rd Work. on Soc. Netw. Syst. SNS'10* (2010).
11. Kim, H. & Anderson, R. Temporal node centrality in complex networks. *Phys. Rev. E* **85**, 026107 (2012).
12. Kostakos, V. Temporal graphs. *Phys. A: Stat. Mech. its Appl.* **388**, 1007–1023 (2009).
13. Holme, P. Network reachability of real-world contact sequences. *Phys. Rev. E* **71**, 046119 (2005).
14. Grindrod, P., Parsons, M. C., Higham, D. J. & Estrada, E. Communicability across evolving networks. *Phys. Rev. E* **83**, 046120 (2011).
15. Taylor, D., Myers, S., Clauset, A., Porter, M. & Mucha, P. Eigenvector-based centrality measures for temporal networks. *Multiscale Model. Simul.* **15**, 537–574 (2017).
16. Cardillo, A. *et al.* Emergence of network features from multiplexity. *Sci. Reports* (2013).
17. Cardillo, A. *et al.* Modeling the multi-layer nature of the European Air Transport Network: Resilience and passengers re-scheduling under random failures. *Eur. Phys. J. Spec. Top.* **215**, 23–33, https://doi.org/10.1140/epjst/e2013-01712-8 (2013).
18. Boccaletti, S. *et al.* The structure and dynamics of multilayer networks. *Phys. Reports* **544**, 1–122 (2014).
19. Zanin, M. & Lillo, F. Modelling the air transport with complex networks: A short review. *Eur. Phys. J. Spec. Top.* **215**, 5–21, https://doi.org/10.1140/epjst/e2013-01711-9 (2013).
20. Cook, A. *et al.* Applying complexity science to air traffic management. *J. Air Transp. Manag.* **42**, 149–158, https://doi.org/10.1016/j.jairtraman.2014.09.011 (2015).
21. Guimerà, R., Mossa, S., Turtschi, A. & Amaral, L. A. N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci.* **102**, 7794–7799, https://doi.org/10.1073/pnas.0407994102 (2005).
22. Verma, T., Araújo, N. A. M. & Herrmann, H. J. Revealing the structure of the world airline network. *Sci. Reports* (2014).
23. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.* **103**, 2015–2020, https://doi.org/10.1073/pnas.0510525103 (2006).
24. Fleurquin, P., Ramasco, J. J. & Eguiluz, V. M. Systemic delay propagation in the US airport network. *Sci. Reports* (2013).
25. Kempe, D., Kleinberg, J. & Kumar, A. Connectivity and inference problems for temporal networks. *J. Comput. Syst. Sci.* **64**, 820–842, https://doi.org/10.1006/jcss.2002.1829 (2002).
26. Zanin, M., Lacasa, L. & Cea, M. Dynamics in scheduled networks. *Chaos: An Interdiscip. J. Nonlinear Sci.* **19**, 023111 (2009).

Acknowledgements

Author Contributions

S.Z., P.M. and F.L. designed the research. S.Z. performed the research and the data analysis, with assistance from P.M. and F.L. All authors contributed to the writing and to the discussion of the results.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-47115-6.

Competing Interests: The authors declare no competing interests.

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