Estimating nodal spreading influence using partial temporal networks

Mao, Tianrui; Zhang, Shilun; Hanjalic, Alan; Wang, Huijuan

DOI
10.1016/j.chaos.2025.117163

Publication date
2025

Document Version
Final published version

Published in
Chaos, Solitons and Fractals

Chaos, Solitons and Fractals 201 (2025) 117163

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Estimating nodal spreading influence using partial temporal networks

Tianrui Mao, Shilun Zhang, Alan Hanjalic, Huijuan Wang *

Delft University of Technology, Mekelweg 5, Delft, 2628 CD, South Holland, Netherlands

ARTICLE INFO

Keywords:
Nodal influence estimation
Spreading processes
Walk-based centrality metrics
Nodal properties
Temporal networks
Partial network information

ABSTRACT

Networks facilitate the spread of information and epidemics. The average number of nodes infected via a spreading process on a network starting from a single seed node over a given long period is called the influence of that node. Estimating nodal influence early in time is essential for the epidemic/misinformation mitigation. Influence estimation has been investigated in static networks, which identifies the relation between topological properties of a node and its influence and assumes the networks are completely known. However, the networks underlying spreading processes such as social interactions are not static but temporal networks, whose links are activated or deactivated over time. When predicting nodal influence in the long-term future, the temporal network is usually only observable till the time of prediction and only locally around the node due to data accessibility. To bridge this gap, we address the question of how to utilize the partially observed temporal network (local and of short duration) around each node, to estimate the ranking of nodes in spreading influence on the full network over a long period. This would also enable us to understand which network properties of a node, in its partially observed temporal network determine its influence. Centrality metrics (nodal properties) have been proposed recently in temporal networks. However, using such a metric derived for each node from its partial network to estimate the ranking of nodes in influence is likely to be limiting. This is because the spread of information is possibly through any time-respecting path, beyond the shortest time-respecting path considered by existing metrics. To address this disparity, we systematically propose a set of novel nodal centrality metrics that encode diverse properties of (time-respecting) walks to predict nodal influence rankings. The proposed metrics derived from partial network information, in general, outperform classic centrality metrics utilizing either full or partial temporal network information. It is found that distinct centrality metrics perform the best depending on the infection probability of the spreading process. For a broad range of the infection probability, a node tends to be influential if it can reach many distinct nodes via time-respecting walks and if these nodes can be reached early in time.

1. Introduction

Human social interactions or contacts usually occur at specific times instead of constantly. They can be represented as temporal networks, where links are activated/deactivated over time. Spreading processes have been widely used to model the propagation of epidemics, the formation of opinions, and the cascade of failures on temporal networks. The average number of nodes that are infected via a spreading process on a network starting from a single seed node at $t = 0$ over a given period $[1, \tau]$ is referred as the

* Corresponding author.
E-mail address: H.wang@tudelft.nl (H. Wang).

https://doi.org/10.1016/j.chaos.2025.117163
Received 12 February 2025; Received in revised form 21 August 2025; Accepted 30 August 2025
Available online 12 September 2025

influence of this node. Recently, centrality metrics, or equivalently network properties of a node, have been proposed for temporal networks [1–4]. The centrality metrics of the nodes derived from the entire temporal network underlying the spreading process observed within the same period $[1, \tau]$ are used to estimate the rank of the nodes in the spreading influence [5–8].

Machine learning models have shown their effectiveness in identifying influential nodes from those whose influence is unknown, provided that the actual influence of a subset of nodes is available and the full temporal (or static) network information is known [9–12]. Specifically, a statistical model trained on the set of nodes whose spreading influences are known is used to predict the influence of the rest of nodes. The statistical model maps the relation between a node's influence in the spreading process and its centrality metrics [9,10]. In addition to nodal centrality metrics, the mapping between nodal influence and node embeddings has also been explored, where node embeddings are obtained by deep learning models, based on complete network information [11,12].

In practice, it is desirable to be able to estimate the ranking of nodes in influence over the long-term $[1, \tau]$, early in time, e.g., at $\phi\tau$ where $\phi \in (0, 1)$, when the temporal network is only observable locally and within a short-term $[1, \phi\tau]$. This is motivated by the following two applications as examples. An epidemic may start to spread from several seed nodes on parallel. Knowing which seed nodes will lead to higher prevalence in the future at τ at an early time $\phi\tau$ would allow policy makers to better mitigate the spread by, e.g., prioritizing the intervention at the area around the highly influential seed and preparing for the mitigation early in time. At the prediction time $\phi\tau$, the network is observable only till $\phi\tau$, and only locally around each seed node due to data accessibility. Similarly, misinformation may spread from multiple seed nodes via an online social network on parallel. It is difficult to obtain the entire social contact network due to privacy policies and access limitations, nor to predict the future network. Leveraging partial network information to estimate nodal influence may lead to more effective interventions against the spread of misinformation and epidemics.

In this study, our objective is to utilize partially observed temporal network (local and early in time) around a node, to forecast the overall spreading influence of the node in the long term. Besides the aforementioned motivations from the perspective of applications, this objective is also motivated by its scientific importance. It would help us understand which local properties of a node in the early stages determine its long-term spreading influence.

In this work, we consider the discrete-time Susceptible–Infected (SI) spreading process [13] on a temporal network. In the SI spreading process, each node can be in one of two possible states at any time: susceptible or infected. Initially, at the discrete time step $t = 0$, a seed is infected whereas all other nodes are susceptible. At any time step t, a susceptible node could get infected by an infected node with an infection probability β if the two nodes have an interaction or contact. If a node gets infected at time t, it could infect other susceptible nodes that it contacts since time $t + 1$. The spreading influence of a node is defined as the average number of infected nodes (also known as the outbreak size) till time $t = \tau$ when the node is chosen as the only seed node of the SI epidemic spreading. We systematically explore how partial temporal network information around a node may contribute to the prediction of the ranking of nodes in spreading influence. Specifically, we address the following generalized nodal influence prediction problem: given the partial temporal network $\mathcal{G}_i(\phi, m)$ of each target node i that is the temporal network G observed in the early period $[1, \phi\tau]$ and within m hops from i, how to estimate the ranking of nodes in spreading influence on the whole temporal network G over the longer period $[1, \tau]$? The partial temporal network $\mathcal{G}_i(\phi, m)$ contains the set of nodes $\mathcal{V}_i(\phi, m)$ that includes node i and nodes that are within m hops[1] from i in the unweighted aggregated network of the temporal network $\mathcal{G}$ observed within $[1, \phi\tau]$ and all contacts of the temporal network G that occur within $[1, \phi\tau]$ and among $\mathcal{V}_i(\phi, m)$.

To solve this problem, we design three nodal centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ to predict the influence of the node. Centrality metrics are defined to quantify various network properties of a node, respectively. Centrality metrics have been proposed to capture the properties of a node in relation to the shortest paths in a static network or the shortest time-respecting paths in a temporal network. For example, temporal closeness quantifies how close a node is connected to other nodes via the shortest time-respecting paths. Such metrics could be limited in estimating nodal spreading influence. The reason is that the spread of information from a seed node to any other node is not necessarily through the shortest time-respecting path, but possibly through any time-respecting path. This motivates us to define centrality metrics that systematically capture how well the partial temporal network around a node is connected via time-respecting walks and via walks in the corresponding aggregated network $\mathcal{G}_i^w(\phi, m)$. In this way, they account all possible spreading trajectories starting from the seed node. Using centrality metrics based on partial network information to predict nodal influence also enables the exploration of how large m, thus to what extent relatively local information, is actually needed to provide a desirable prediction performance. Earlier research [14–16] has demonstrated in static networks that a centrality metric derived from the entire network, such as betweenness, which involves a high computational complexity, exhibits a significant correlation with centrality metrics obtained from local neighborhoods. Therefore, predicting nodal influence might not necessarily require extensive macroscopic neighborhood information. Centrality metrics have been proposed over time to capture diverse nodal topological properties. In contrast, our centrality metrics are specifically designed for the estimation of influence, beyond the additional nodal properties they capture.

Each proposed centrality metric will be derived for each node based on the partially observed network of the node. The ranking of nodes in the given centrality is used as the estimated ranking of nodes in nodal influence. To evaluate the performance of the proposed methods, classic centrality metrics derived from full and partial network information respectively are also used to estimate nodal influence. It is found that the proposed centrality metrics derived from partial network information, in general, outperform classic centrality metrics utilizing either full or partial temporal network information. We further explain the performance of these methods as well as their dependency on the infection probability of the spreading process. For a broad range of the infection

[1] The hopcount between two nodes in a network is the number of links contained in the shortest path between the two nodes.

probability, a node tends to be influential if it could reach many distinct nodes via time-respecting walks and if these nodes could be reached early in time.

The remainder of this paper is organized as follows. In Section 2, (partial) temporal networks and their weighted aggregated networks are defined. In Section 3, we introduce our walk-based nodal centrality metrics as well as classic centrality metrics. In Section 4, methods including datasets used to evaluate the prediction quality of aforementioned metrics are explained. In Section 5, we analyze and explain the performance of those metrics in multiple real-world networks and their randomized networks. Section 6 summarizes our findings and discusses possible future works.

2. Temporal network representation

A temporal network measured within an observation window $[1, \tau]$ at discrete times can be represented as a sequence of network snapshots $G = \{G_1, G_2, \dots, G_\tau\}$. The snapshot $G_t = (V; E_t)$ at time step t has V and E_t being the set of nodes and contacts, respectively. The number of nodes in V is represented as N. If node i and j have an interaction or a contact at time step t, $(i, j) \in E_t$. Here, we assume all snapshots share the same set of nodes, i.e., V. A temporal network can also be described by a three-dimensional binary adjacency matrix $A_{N\times N\times\tau}$, where each element $a_{i,j,t} = 1$ if there is a contact between node i and j at time step t, or else $a_{i,j,t} = 0$.

A weighted aggregated network G^w can be derived from a temporal network G by aggregating contacts over time window $[1, \tau]$. The links in the time aggregated network G^w are defined as $E = \cup_{t=1}^{\tau} E_t$. That is, a pair of nodes is connected with a link in G^w if at least one contact occurs between them in the temporal network. Each link (i, j) in G^w is associated with a weight $w_{i,j}$ counting the total number of contacts between node i and j in G. The weighted aggregated network G^w can therefore be described by a weighted adjacency matrix $W_{N\times N}$, with its element $w_{i,j} = \sum_{t=1}^{\tau} a_{i,j,t}$.

The partial temporal network $\mathcal{G}_i(\phi, m)$ observed around each target node i in the early period $[1, \phi\tau]$ and within m hops from i will be used to estimate the ranking of nodes in spreading influence on the whole temporal network G over the longer period $[1, \tau]$. Fig. 1(a)–(c) shows the example of a temporal network G, its weighted aggregated network G^w, the partial temporal network $\mathcal{G}_A(\phi, m)$ observed around node A and the corresponding weighted aggregated network $\mathcal{G}_A^w(\phi, m)$, where $\phi = 0.5$ and $m = 3$.

3. Centrality metrics

We first design three walk-based nodal centrality metrics, namely weighted degree mass, (time-scaled) temporal degree mass, and (time-scaled) temporal reachability, to capture properties of the partial temporal network observed around a node. Each metric is firstly defined based on a full temporal network G or its aggregated network G^w and then adapted for partial temporal network. Each metric derived from the partial network is used to estimate the rank of the nodes in influence. In order to evaluate the performance of the proposed walk-based metrics, we also introduce a set of classic centrality metrics and explain how these metrics could be derived from partial or full temporal networks or their aggregated networks, respectively, to estimate nodal influence.

3.1. Weighted degree mass

We firstly propose the definition of the mth-order weighted degree mass $d_i^{(m)}$ of a node i in the weighted aggregated network G^w of a temporal network G with its weighted adjacency matrix W as

$$d_i^{(m)} = \sum_{k=1}^{m} (W^k u)_i \tag{1}$$

where $u = (1, 1, \dots, 1)^T$ is the all-one vector. Each element $W_{i,j}^k$ represents the total number of distinct k-hop ($k \le m$) walks[2] between node i and j (i, j can be the same node) in G^w when interpreting the weight $w_{i,j}$ of each node pair as the number of links between the node pair. Therefore, $(W^k u)_i$ counts the total number of distinct k-hop walks starting from node i. The mth-order degree mass $d_i^{(m)}$ of a node i represents the total number of walks within m-hops starting from the node i. This definition generalizes the original degree mass definition that has been proposed for unweighted networks by replacing the unweighted adjacency matrix with a weighted one W [17].

In our context, only the partial temporal network $\mathcal{G}_i(\phi, m)$ observed around node i within a short period $[1, \phi\tau]$ is known, and its aggregated network is $\mathcal{G}_i^w(\phi, m)$ with adjacency matrix $\mathcal{W}_i(\phi, m)$. We propose to consider the mth-order weighted degree mass $d_i^{(\phi,m)}$ in the aggregated partial network $\mathcal{G}_i^w(\phi, m)$, i.e.,

$$d_i^{(\phi,m)} = \sum_{k=1}^{m} (\mathcal{W}_i^k(\phi, m)u)_i \tag{2}$$

It represents the total number of distinct walks within m-hop starting from node i in the aggregated partial network observed around i within m hops during $[1, \phi\tau]$. An example of m-hop walks in $\mathcal{G}_i^w(\phi, m)$ between two nodes is shown in Fig. 1(d). Correspondingly, we propose to use the mth-order weighted degree mass $d_i^{(\phi,m)}$ to estimate the influence of the node. For any link in G^w, its weight is larger or equal to its weight in $\mathcal{G}_i^w(\phi, m)$. Hence, $d_i^{(\phi,m)} \le d_i^{(m)}$.

[2] A k-hop walk between node n_0 and n_k in a weighted aggregated network is a succession of links $(n_0, n_1), (n_1, n_2), ..(n_k - 1, n_k)$.

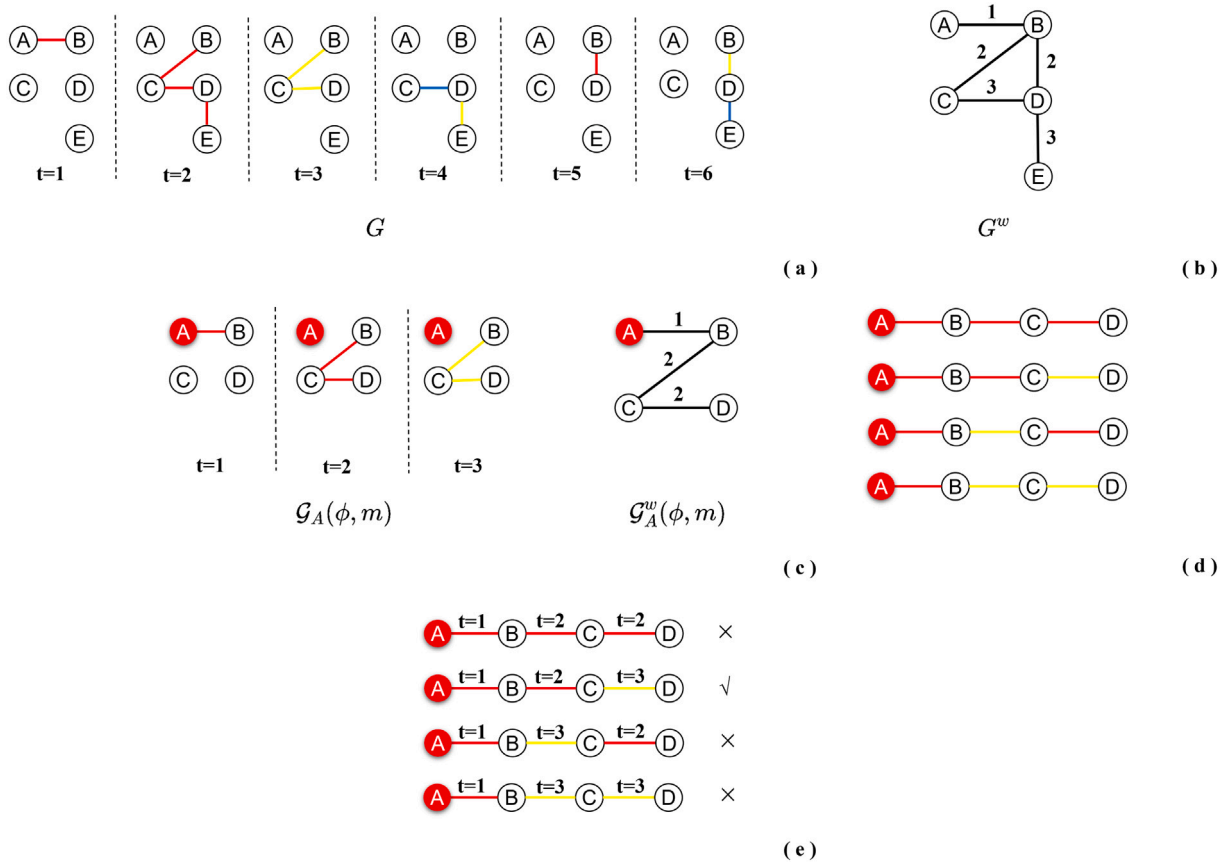


Fig. 1. (a) A temporal network G with 5 nodes and 6 time steps. The first, second, and third contacts between the same pair of nodes are marked in red, yellow, and blue, respectively. (b) The aggregated network G^w of G along with its link weight. (c) The partial temporal network $\mathcal{G}_A(\phi,m)$ where $\phi = 0.5$ and $m = 3$, observed around node A and its corresponding weighted aggregated network $\mathcal{G}_A^w(\phi,m)$. Partial network $\mathcal{G}_A(\phi,m)$ contains node A and nodes that are within $m = 3$ hops from A, and all interactions among this set of nodes occurring within $[1,3]$. The set of nodes that are within $m = 3$ hops from A are identified in the aggregated network of the temporal network G observed within $[1,3]$, which has the same topology as G^w but without link (B,D). (d) The list of all 3-hop walks between node A and D in $\mathcal{G}_A^w(\phi,m)$. To identify the walks, a link with weight e.g., 2 in $\mathcal{G}_A^w(\phi,m)$ is regarded as two parallel links. (e) For each walk listed in (d), time stamp of each link is added and the walk is marked as × if it is not a time-respecting walk and as ✓if it is time-respecting.

3.2. Temporal degree mass

Furthermore, we propose the mth-order temporal degree mass $\delta_i^{(m)}$ of a node i in a temporal network G as the total number of time-respecting walks[3] within m-hops starting from node i at time $t = 1$. An example of m-hop time-respecting walks in the partial temporal network $\mathcal{G}_i(\phi,m)$ between two nodes is shown in Fig. 1(e). The total number of time-respecting walks within m-hops starting from node i at time $t = 1$ to any node j is upper-bounded by the total number of walks within m-hops from i to j in the weighted aggregated network $\Psi_{i,j}^{(m)}$ where $\Psi^{(m)} = \sum_{k=1}^{m} W^k$, as exemplified in Fig. 1(d) and (e). We use vector $B_{i,j}^{(m)}$, a row vector of length $\Psi_{i,j}^{(m)}$ with binary elements to indicate whether each of the $\Psi_{i,j}^{(m)}$ walks in the weighted aggregated network is time-respecting or not when the time information of each link in the walk is taken into account. Hence, the mth-order temporal degree mass $\delta_i^{(m)}$ of a node i in a temporal network follows

$$\delta_i^{(m)} = \sum_{j=1}^{N} \sum_{s=1}^{\Psi_{i,j}^{(m)}} B_{i,j}^{(m)}(s) \tag{3}$$

A node i with a high temporal degree mass $\delta_i^{(m)}$, thus well connected via many time-respecting walks within m-hops to other nodes may contribute to the high influence of the node. Beyond, if these walks have a short time duration, i.e., the destination nodes

[3] A k-hop time-respecting walk from n_0 at $t = 0$ to n_k in a temporal network is a succession of contacts $(n_0,n_1,t_1),(n_1,n_2,t_2),..(n_k-1,n_k,t_k)$ that follow the time order $0 < t_1 < t_2, \ldots, < t_k$.

of these walks are reached at an early time, node i tends to be more influential. This is because of the following. Only those nodes that can be reached by time-respective walks are possibly infected in the stochastic SI spreading process starting from seed i at time $t = 0$. If a node j is infected, the spreading trajectory, i.e., the sequence of contacts via which j gets infected is a time-respecting path. Any time-respecting walk from i to j, that is also a time-respecting path, is a possible epidemic spreading trajectory. Hence, a large number of time-respecting walks within m-hops from i to j suggests a high probability for j to get infected in the spreading process starting from i. If j gets infected earlier, more contacts that occur after its infection could spread the epidemic/information further from j. Hence, if a node i has a high temporal degree mass $\delta_i^{(m)}$, thus likely a large number of time-respecting paths within m-hops to other nodes, and if these walks tend to have a short time duration such that other nodes may get infected further, it tends to have a high influence. Therefore, we propose to use the time-scaled temporal degree mass $\Delta_i^{(m)}$ of a node i to estimate the influence of the node. The time-scaled temporal degree mass $\Delta_i^{(m)}$ is defined as

$$\Delta_i^{(m)} = \sum_{j=1}^{N} \sum_{s=1}^{\Psi_{i,j}^{(m)}} B_{i,j}^{(m)}(s) \alpha^{\phi_{i,j}^{(m)}(s)} \tag{4}$$

where the vector $\phi_{i,j}^{(m)}$ records the time duration[4] of each walk from i to j within m-hops identified in the weighted aggregated network equipped with the time stamp of each link in the walk in case the walk is time-respecting and is infinity in case the walk is not time-respecting. The vector $\phi_{i,j}^{(m)}$ thus has the same length $\Psi_{i,j}^{(m)}$ as vector $B_{i,j}^{(m)}$, which equals the total number of walks from i to j within m-hops in the weighted aggregated network. Both vectors index the $\Psi_{i,j}^{(m)}$ walks in the same way. If the sth walk in the aggregated network is time-respecting in the temporal network, i.e., $B_{i,j}^{(m)}(s) = 1$, the contribution of this walk to the time-scaled temporal degree mass $\Delta_i^{(m)}$ is $\alpha^{\phi_{i,j}^{(m)}(s)}$, where $\alpha \in (0, 1]$. When $\alpha < 1$, a walk with a longer duration contributes less to the time-scaled temporal degree mass. When $\alpha = 1$, the time-scaled temporal degree mass $\Delta_i^{(m)} = \delta_i^{(m)}$ equals the temporal degree mass, counting simply the total number of time-respecting walks within m-hops from i.

In our context, we use the time-scaled temporal degree mass $\Delta_i^{(\phi,m)}$ derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ to estimate the influence of the node i.

3.3. Temporal reachability

Besides time-respecting walks, the concept of the reachable node can also be utilized for designing centrality metrics. The temporal reachability $z_i^{(m)}$ of a node i within m-hops in a temporal network G is the number of distinct nodes that could be reached via time-respecting walks starting from node i at $t = 0$ within m-hops. Mathematically,

$$z_i^{(m)} = \sum_{j=1}^{N} \mathbf{1}_{\sum_{s=1}^{\Psi_{i,j}^{(m)}} B_{i,j}^{(m)}(s) > 0} \tag{5}$$

where the condition $\sum_{s=1}^{\Psi_{i,j}^{(m)}} B_{i,j}^{(m)}(s) > 0$ is true if there is at least one time-respecting walk from i to j and the indicator function $\mathbf{1}_x$ equals one if the condition x is true or zero otherwise.

When the infection probability per contact is high, any time-respecting walk starting from node i to j within m-hops could lead to the infection of j with a high probability. In this case, the temporal reachability (the number of nodes can be reached via walks) instead of the temporal degree mass (the number of distinct walks) could be more relevant in estimating nodal influence. Similarly, if a node i has high temporal reachability $z_i^{(m)}$ and if each of these $z_i^{(m)}$ nodes has an earlier reached time,[5] node i tends to be influential. Therefore, we propose to use the time-scaled temporal reachability $Z_i^{(m)}$ of a node i to estimate the influence of the node, if the global network G is known. The time-scaled temporal reachability $Z_i^{(m)}$ is defined as

$$Z_i^{(m)} = \sum_{j=1}^{N} \mathbf{1}_{\sum_{s=1}^{\Psi_{i,j}^{(m)}} B_{i,j}^{(m)}(s) > 0} \alpha^{\min_s \phi_{i,j}^{(m)}(s)} \tag{6}$$

where the contribution of each reachable node j is scaled by $\alpha^{\min_s \phi_{i,j}^{(m)}(s)}$, depending on the shortest time $\min_s \phi_{i,j}^{(m)}(s)$ that j is reached. In our context, we use the time-scaled temporal reachability $Z_i^{(\phi,m)}$ derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ to estimate the influence of the node i. The time-scaled temporal reachability $Z_i^{(\phi,m)}$ takes into account how many nodes are reachable via time-respective walks/paths in the partial temporal network as well as when each node is reached via the fastest time-respective path.

These three proposed centrality metrics, derived from the full (aggregated) temporal network, respectively, follow

$$z_i^{(m)} \leq \delta_i^{(m)} \leq d_i^{(m)} \tag{7}$$

[4] The time duration of a time-respecting walk equals the time of the last contact in the walk minus the starting time of epidemic spreading, which is 0 in our context.

[5] The reached time equals the duration when the node is firstly reached by the seed minus the starting time of epidemic spreading, which is 0 in our context.

This is because not all the $d_i^{(m)}$ walks identified from the aggregated network are necessarily time-respecting and not each of the $\delta_i^{(m)}$ time-respecting walks reaches a unique destination node. Similarly,

$$Z_i^{(\phi,m)} \leq \Delta_i^{(\phi,m)} \leq d_i^{(\phi,m)} \tag{8}$$

when the same scaling α is considered.

3.4. Classic centrality metrics

Centrality metrics have been proposed in static networks and recently in temporal networks. We are interested in how the proposed centrality metrics, leveraging partial network information, perform in comparison with classic centrality metrics in influence estimation. We will firstly introduce 4 centrality metrics defined for the static networks. Each metric, derived from the unweighted aggregated network of temporal network G, or of partial temporal network $\mathcal{G}_i(\phi, m)$, as well as the average of this metric computed across all snapshots of G or of $\mathcal{G}_i(\phi, m)$, will be used to predict nodal influence, respectively.

* The betweenness centrality b_i [18] of a node i is the number of shortest paths between all pairs of nodes in the network that pass through the node i,

$$b_i = \sum_{s \neq i \neq d \in V} \frac{\sigma_{sd}(i)}{\sigma_{sd}} \tag{9}$$

  where $\sigma_{sd}(i)$ is the number of shortest paths that pass through node i between node s and node d, and σ_{sd} is the total number of shortest paths between s and d. Assuming that a unit packet is transmitted between each node pair via the shortest path, the betweenness b_i is the total number of packets passing through node i.
* The closeness centrality c_i [19] of a node i measures how close a node is connected to all the others via the shortest path. It is commonly defined as

$$c_i = \sum_{j \in V \setminus \{i\}} \frac{1}{H_{i,j}} \tag{10}$$

  where $H_{i,j}$ is the hopcount of the shortest path between nodes i and j.
* The eigenvector centrality x_i [20] of node i is the component of the principal eigenvector x corresponding to node i and the principal eigenvector is the eigenvector corresponds to the largest eigenvalue λ_1 of the adjacency matrix A of the static network. Hence, $x\lambda_1 = Ax$. The eigenvector centrality x_i of a node tends to be large if it has many neighbors and each neighbor has a large eigenvector centrality.
* The PageRank centrality P_i [21] of node i is the probability that node i is visited by a random walker:

$$P_i = \frac{1-\gamma}{N} + \gamma \sum_{j \in V \setminus \{i\}} \frac{A_{i,j} P_j}{k_j} \tag{11}$$

  where γ is the probability for a walker to move to a random neighbor of the current node being visited, $1-\gamma$ is thus the probability for the walker to move to a random node and k_j is the degree of node j. The parameter γ is set to 0.85, which is a common choice for calculating the PageRank centrality.

Furthermore, we introduce the temporal closeness centrality defined for temporal networks. It will be derived from the full temporal network G and partial temporal network $\mathcal{G}_i(\phi, m)$ respectively to predict nodal influence.

* Temporal closeness centrality TC_i [1] is defined analogously as closeness centrality. It measures how close node i is connected to the other nodes via time-respecting paths. Specifically, it is defined as

$$TC_i = \sum_{j \in V \setminus \{i\}} \frac{1}{TH_{i,j}} \tag{12}$$

  where $TH_{i,j}$ is the hopcount of the shortest time-respecting path[6] from node i to j, as introduced in [22].

4. Evaluation of prediction quality

In this Section, we introduce the method to evaluate the quality of an influence prediction algorithm/metric. This entails the real-world networks to be used, the parameter choice of the spreading process and the partial temporal network, and the measures that quantify the prediction quality.

[6] The shortest time-respecting path is the time-respecting path with the minimum number of hopcount, or equivalently with the minimum number of contacts.

Table 1
Basic statistics of each real-world temporal network considered: the number of nodes (N), the total number of contacts (L), the total number of timesteps (T), the type of contacts recorded, the link density (p), and the average of modularity Γ of the largest connected component of the weighted network G^w over all considered observation periods.

	N	L	T	Type	p	Γ
Workplace13	92	9827	7104	Physical	0.180	0.592
Workplace15	217	78 249	18 488	Physical	0.182	0.641
Highschool11	126	28 561	5609	Physical	0.217	0.667
Highschool12	180	45 047	11 273	Physical	0.138	0.754
Hyper-text	113	20 818	5246	Physical	0.347	0.441
SFHH	403	70 261	3509	Physical	0.118	0.536
Sms (filtered)	457	22 152	21 898	Virtual	0.006	0.912
Calls (filtered)	347	2676	2671	Virtual	0.008	0.878
Manufacturing Emails	167	82 281	57 791	Virtual	0.234	0.401

4.1. Empirical networks

The following real-world temporal networks will be considered to evaluate the aforementioned nodal spreading influence estimation methods.

* HighSchool11&12 [23] record the physical contacts between students in a high school in Marseilles, France. These two datasets incorporate two different groups of students.
* WorkPlace13&15 [24] capture the physical contacts between individuals in an office building in France. These two datasets originate from distinct sets of individuals.
* Hyper-text&SFHH [25,26] record the physical contacts among scientists during the 2009 conference of ACM Hypertext and SFHH.
* Sms&Calls [27] are obtained from the contacts via short messages and calls of the same set of mobile phones on campus, respectively.
* Manufacturing Emails [28] represents the internal email communication network between employees of a mid-sized manufacturing company.

These networks record virtual or face-to-face contacts in the context of workplace, highschool, university, and academic conference at discrete time steps. The time steps at which there is no contact in the whole network have been deleted. This pre-processing has also been used in e.g. [29]. In this way, we focus solely on those time steps relevant to information diffusion, and exclude periods without any contact, possibly resulted from the inactive period during evenings or technical errors when measuring the network. Meanwhile, we observe that the aggregated networks of Sms and Calls are not fully connected. Hence, we extract the largest connected components of these two aggregated networks respectively, and consider only nodes within each largest connected component and contacts between these nodes. Basic statistics of selected empirical temporal networks after pre-processing are shown in Table 1. The link density p is the number of links in the aggregated network of a temporal network normalized by $N(N-1)/2$, i.e., the maximum possible number of links among the same set of N nodes.

4.2. Experimental settings

Without loosing generality, we evaluate the performance of using the proposed nodal centrality metrics for influence prediction when the infection probability β of the SI spreading is systematically examined across a broad range, i.e., $\beta \in \{0.01, 0.05, 0.1, 0.25, 0.5, 1\}$. When $\beta < 1$, the SI process is stochastic, and the actual spreading influence of a node is derived as the average outbreak size over 500 independent realizations of the SI process starting from this node. When we define the spreading process and propose nodal centrality metrics in Sections 1 and 3, the starting time t_0 of the spreading is assumed to be 0. We consider the following more general case. For each dataset in Table 1, we consider a set of possible starting times, i.e., $t_0 \in \{0, T/8, T/4, 3T/8, T/2, 5T/8, 3T/4\}$. Given a starting time and a seed node, the influence of this node, i.e., the average outbreak size at $t_0 + \tau$ is considered, where $\tau = T/4$. Hence, the temporal network within $[t0, t0 + \tau]$ decides the influence of each node. The ranking of nodal influence is estimated via each of the centrality metrics we proposed based on the partial temporal network observed within $[t0, t0 + \phi\tau]$ and within m hops around each node, where ϕ is chosen as 0.25 and 0.5, respectively and $m \in [1, 2, 3]$. For each proposed centrality metric, $\alpha \in (0, 1]$ could be tuned to achieved the best performance. The average influence of a node over all seed nodes and possible starting times as a function of the infection probability β in each real-world network is shown in Fig. 2. Network Sms and Call that have the lowest link density among all networks as shown in Table 1. Correspondingly, they have the lowest average nodal influence, thus prevalence in the spreading process.

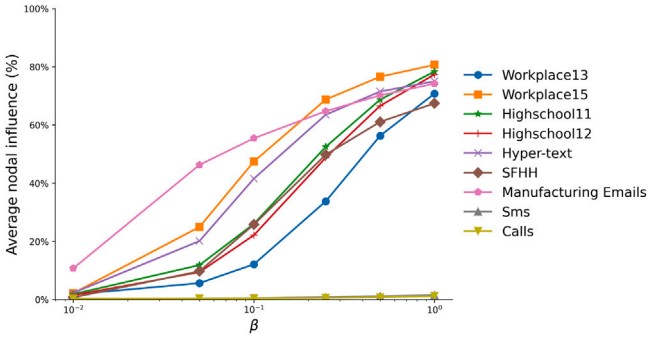


Fig. 2. Average nodal influence in each real-world network as a function of the infection probability β.

4.3. Prediction quality

We introduce two measures to evaluate the quality of each proposed centrality metric in predicting the ranking of nodal influence. The actually influence of each node is represented by a vector s, whereas $\hat{s}$ records a given centrality metric for each node, representing the predicted influence.

Kendall's correlation coefficient $Q_k(\hat{s}, s)$ [30] measures the similarity between the ranking of nodes based on the predicted nodal influence $\hat{s}$, and the ranking of nodes based on the actual nodal influence vector s obtained by SI simulation. A value of 1 for $Q_k(\hat{s}, s)$ indicates that the centrality metric gives the same node ranking as the ground truth nodal influence, while a value of -1 indicates that the two rankings are reversed. Kendall's correlation coefficient is defined as:

$$Q_k(\hat{s}, s) = \frac{n_c - n_d}{\sqrt{(n_c + n_d + O) \cdot (n_c + n_d + U)}} \tag{13}$$

where n_c and n_d are the total numbers of node pairs that are concordant and discordant respectively, based on the influence s and the predicted influence $\hat{s}$. For example, node pair (i, j) is concordant if $(\hat{s}_i - \hat{s}_j)(s_i - s_j) > 0$, and is discordant if $(\hat{s}_i - \hat{s}_j)(s_i - s_j) < 0$. The number of node pairs that have the same actual influence but different predicted influence, i.e., $s_i = s_j$, $\hat{s}_i \neq \hat{s}_j$ is denoted by O and U is the number of node pairs that have the same predicted influence but different actual influence, i.e., $\hat{s}_i = \hat{s}_j$, $s_i \neq s_j$.

Recognition rate $Q_r(\hat{s}, s, f)$ of top-$f\%$ measures the performance of a centrality metric in identifying the most influential $f\%$ nodes. It is calculated as the proportion of nodes that are present in both V_f, the set of top $f\%$ of nodes ranked by the predicted nodal influence $\hat{s}$ and R_f, the top $f\%$ of nodes ranked by the actual nodal influence s:

$$Q_r(\hat{s}, s, f) = \frac{|R_f \cap V_f|}{|R_f|} \tag{14}$$

where $|R_f| = f\%N$ is the number of nodes in R_f and we take 20% for $f\%$.

We evaluate the performance of each method via the average Kendall's correlation coefficient $\bar{Q}_k$ and the average Recognition rate $\bar{Q}_r$ over all possible starting times t_0 of the spreading process on each real-world network, as listed in 4.2.

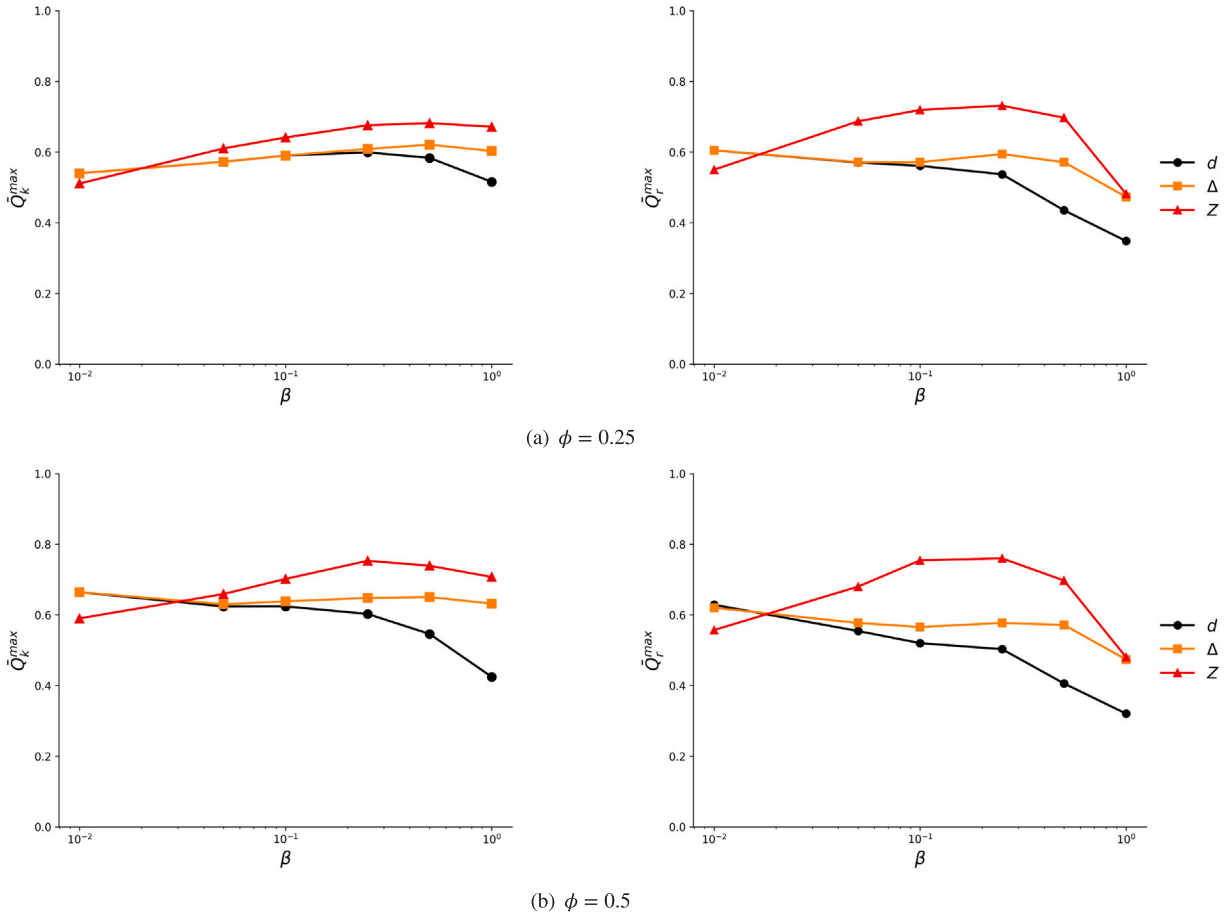
5. Performance analysis

We use the example of the HighSchool11 network to analyze the performance of the proposed influence prediction methods as similar key findings have been observed in the other real-world temporal networks. Results of the other networks are presented in the appendix.

This Section is structured as follows: first, we systematically compare and explain the prediction quality of proposed centrality metrics. Next, we compare the prediction quality of centrality metrics that we proposed based on partial temporal network information with classical centrality metrics derived from full or partial (aggregated) network information. Finally, we evaluate the prediction quality of proposed methods in randomized real-world networks.

5.1. Evaluating proposed metrics in real-world networks

First, we evaluate the prediction quality of the proposed centrality metrics, i.e., weighted degree mass (d), time-scaled temporal degree mass (Δ), and time-scaled temporal reachability (Z) in estimating the ranking of nodes in influence. For each proposed centrality metric derived from partial network information and given a prediction quality evaluation measure Kendall's correlation coefficient $\bar{Q}_k$ or Recognition rate $\bar{Q}_r$, the optimal parameter set $\{m, \alpha\}$ where $m \in [1, 2, 3]$ and $\alpha \in (0, 1]$ that leads to the best quality will be considered and the corresponding best prediction quality is denoted as $\bar{Q}_k^{max}$ (or $\bar{Q}_r^{max}$). Fig. 3 shows how the (best) prediction quality of each proposed centrality metric varies with the relative duration ϕ of the partial temporal network and the infection probability β of the diffusion process. When $\phi = 0.5$, the proposed centrality metrics tend to achieve slightly better prediction



(a) $\phi = 0.25$

(b) $\phi = 0.5$

Fig. 3. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network HighSchool11. Given each combination of ϕ and β, the best prediction quality of a metric achieved by tuning its parameters m, α is considered.

quality compared with the case when $\phi = 0.25$, regardless of the value of β. Using temporal network information observed in a longer period, the proposed centrality metrics tend to estimate nodal influence better. We find that, in general, the time-scaled temporal reachability Z performs the best whereas the weighted degree mass d performs the worst when β is relatively large. When β is small, Z performs the worst whereas the other two metrics performs equally well. These two observations can be explained as follows.

Considering the case when $\beta \to 0$. Each 1-hop neighbor of the seed node in the aggregated network G^w could get infected with a probability β times the number of contacts in between. The probability that a 2-hop neighbor gets infected is negligibly small, of order β^2. The influence of a node i is proportional to the total number of contacts $d_i^{(1)}$ of the node in the temporal network over the complete period $[1, \tau]$, which is lower bounded by $d_i^{(\phi,1)}$, the weighted degree mass $d_i^{(\phi,1)}$ derived from the partial network information. Hence, weighted degree mass in the partial network is supposed to well predict nodal influence when $\beta \to 0$. When $\alpha = 1$, $\Delta_i^{\phi,m=1} = \delta_i^{\phi,m=1} = d_i^{(\phi,1)}$. Hence, the time-scaled temporal degree mass Δ performs equally well as weighted degree mass d.

When β is large, a node j further than 1-hop away from the seed node i in the aggregated network G^w could get infected. The corresponding spreading trajectory, i.e., the set of contacts via which j gets infected, is a time-respecting path. Intuitively, the time-scaled temporal reachability Z and time-scaled temporal degree mass Δ, taking into account the time information of contacts, are supposed to perform better than the static metric d. When $\beta = 1$, the actual influence of a node i equals the number of distinct nodes that are reachable via time-respecting paths in full network G starting from i within $[1, \tau]$, i.e., $Z_i^{(\phi=1,m\geq\rho)}$, where ρ is the diameter of the aggregated network G^w, when $\alpha = 1$. This supports why metric Z, i.e., $Z_i^{(\phi<1,m\leq3)}$ which is smaller than $Z_i^{(\phi=1,m\geq\rho)}$, could perform the best when β is large. An exception is observed in the Manufacturing Emails network, where Z hardly outperforms other metrics when β is large, as is shown in Fig. 20. The Manufacturing Emails network has the largest link density p and the smallest modularity Γ of the largest connected component of the weighted network G^w, as shown in Table 1. As a consequence, the influence of most nodes approximately is equal to the size of the largest connected component in G^w when $\beta = 1$, as shown in Fig. 4. Hence, the influence of nodes is hardly distinguishable, such that the recognition rate of all proposed metrics is close to that

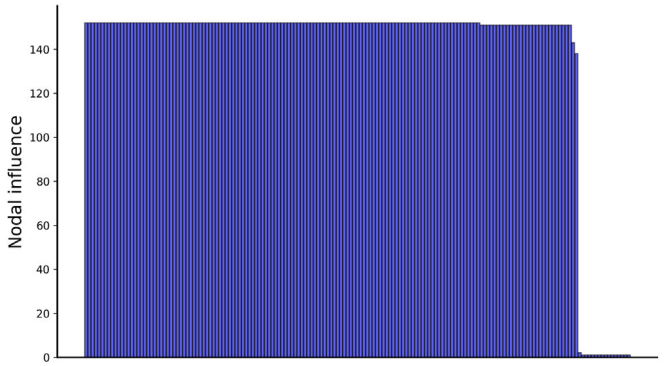


Fig. 4. The nodal influence of each node in a randomly selected observation period $[t_0 + \tau]$ in the Manufacturing Emails network.

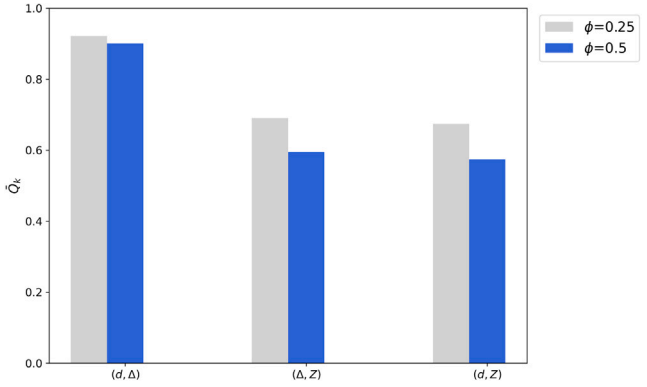


Fig. 5. Average Kendall correlation coefficient $\bar{Q}_k$ between every two proposed centrality metrics in Highschool11 with $m = 2$ and $\alpha = 1$ when $\phi = 0.25$ and 0.5, respectively.

of random guessing, i.e., 20% when $\beta = 1$. The similar nodal influence of most nodes supports why Z hardly outperforms other metrics when β is large.

We explore further the similarity, i.e., Kendall's correlation between every pair of centrality metrics of a node derived from the same partial temporal network. This will help us understand whether these three proposed metrics capture similar properties of a node in the partially observed network and whether correlated metrics perform similarly in the predicting nodal influence or not. Consider $m = 2$ and $\alpha = 1$ as an example. Fig. 5 shows the average Kendall's correlation $\bar{Q}_k$ between every two metrics averaged over all possible starting times of the diffusion process, for $\phi = 0.25$ and $\phi = 0.5$, respectively. It can be seen that all three centrality metrics are evidently and positively correlated. This is in line with their related definitions. This positive correlation $\bar{Q}_k < 1$ also suggests that these metrics capture related but different properties of nodes. The weighted degree mass d and time-scaled temporal degree mass Δ have the strongest correlation, which supports their similar prediction quality. Network Sms and Calls differ evidently from the other networks. In these two networks, the correlation between the weighted degree mass d and time-scaled temporal degree mass Δ is significantly higher than that in other networks as exemplified in Fig. 6, and these two metrics perform almost the same in influence prediction (exemplified by Fig. 18). This high correlation between these two metrics in Sms and Calls can be explained as follows. Consider the case when $m = 2$ as an example. The walks from a target node i within 2 hops on the aggregated partial networks include (1) 1-hop walks, (2) 2-hop walks that return to node i and (3) 2-hop walks reaching finally other nodes than i. There are few type (3) walks, which is partially due to the low link density in the aggregated network of Sms and Calls (see Table 1) and supported by the low prevalence in these two networks (see Fig. 2). When $\phi = 1$, $d_i^{(\phi=1,m=2)} \approx \sum_j [w(i,j) + w^2(i,j)]$, where $w(i,j)$ is the total number of contacts between i and j in partial temporal $\mathcal{G}_i(\phi = 1, m = 2)$, or equivalently in full temporal network G. The time-respecting walks from i within 2-hops on the partial temporal network include the same three types of walks. Hence, $\Delta_i^{(\phi=1,m=2)} \approx \sum_j [w(i,j) + \frac{w(i,j)(w(i,j)-1)}{2}] = \frac{1}{2}\sum_j [w(i,j) + w^2(i,j)] \approx \frac{1}{2} d_i^{(\phi=1,m=2)}$, when $\alpha = 1$. Similarly, it holds that $\Delta_i^{(\phi,m=2)} \approx \frac{1}{2} d_i^{(\phi,m=2)}$ for any ϕ when $\alpha = 1$.

Furthermore, we study the optimal parameter set $\{m, \alpha\}$ where $m \in [1, 2, 3]$ and $\alpha \in (0, 1]$ for each centrality metric to achieve the best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively. As time-scaled temporal reachability Z performs the best except for the case when β is small, we focus on metric Z. Fig. 7 shows that when ϕ is fixed and as β increases, the optimal m increases. As discussed earlier, when $\beta \to 0$, the influence of a node is equal to the infection probability β times the total number of contacts between this node and its 1-hop neighbors over $[t_0, t_0 + \tau]$. When $\beta \to 0$, the information can hardly diffuse to a node that is further than

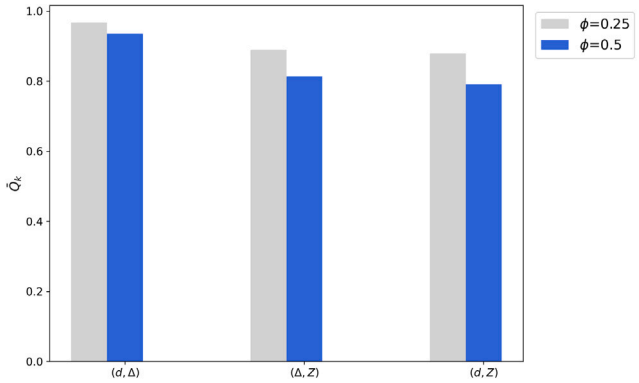
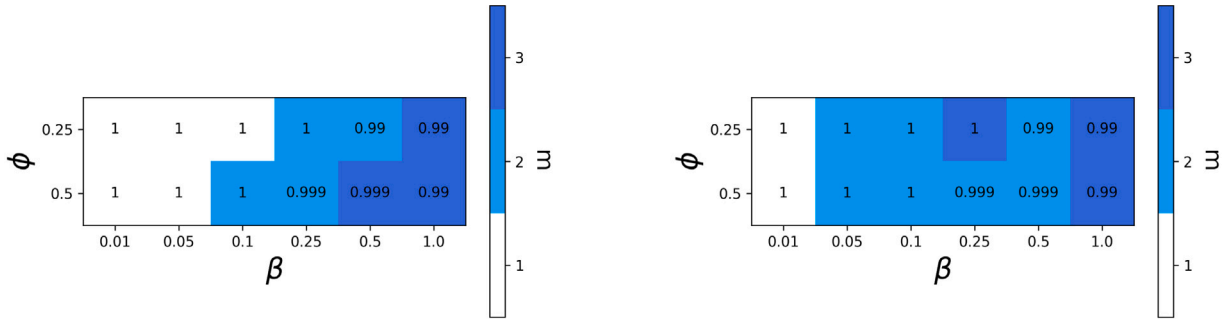


Fig. 6. Average Kendall correlation coefficient $\bar{Q}_k$ between every two proposed centrality metrics in Sms with $m = 2$ and $\alpha = 1$ when $\phi = 0.25$ and 0.5, respectively.



(a) $\bar{Q}_k$ (b) $\bar{Q}_r$

Fig. 7. The optimal m and α that lead Z to perform the best under different ϕ and β combinations based on Highschool11 evaluated by $\bar{Q}_k$ and $\bar{Q}_r$, respectively. The numbers in the heapmap grid represent the optimal α.

1-hop away from the seed node. Hence, only the partial network within 1-hop is relevant for influence prediction. As β increases, the partial network connecting nodes further than 1-hop from the seed could possibly spread the information, thus relevant for prediction.

Another finding is that when ϕ is fixed and as β increases, the optimal[7] α decreases from $\alpha = 1$. When β is large, the time-scaled temporal reachability Z is supposed to perform better than the unscaled case, i.e., $\alpha = 1$ because a node tends to be influential if it can reach many nodes and each reachable node is reached early enough for it to potentially infect other nodes further. However, this is not the case when β is small, because nodes further than 1-hop away from the seed node are unlikely to get infected. In networks Sms and Calls which have a low link density in the aggregated network, nodes further than 1-hop away from the seed node are unlikely to get infected either, even when β is large. Hence, $\alpha = 1$ leads to the best performance of Z in influence estimation, independent of β, in these two networks.

5.2. Comparison with classic centrality metrics in real-world networks

We are interested in how these proposed metrics derived from partial temporal network information perform, in comparison with classic centrality metrics, defined in 3.4, utilizing full or partial, temporal or aggregated network information. Classic centrality metrics (4 static metrics and temporal closeness) can be derived for influence prediction via 4 possible ways: (1) Full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (2) Full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (3) Partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ or (4) Partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$. Given the three parameters coming from the context of the prediction

[7] The optimal α is found via searching within {0.2, 0.4, 0.6, 0.8, 0.85, 0.9, 0.95, 0.99, 0.999, 0.9999, 1} decreases. Since it may take a long time to reach a node via the fastest time-respecting path, the metric Z with $\alpha = 0.9999$ and $\alpha = 1$ respectively could differ evidently in prediction quality.

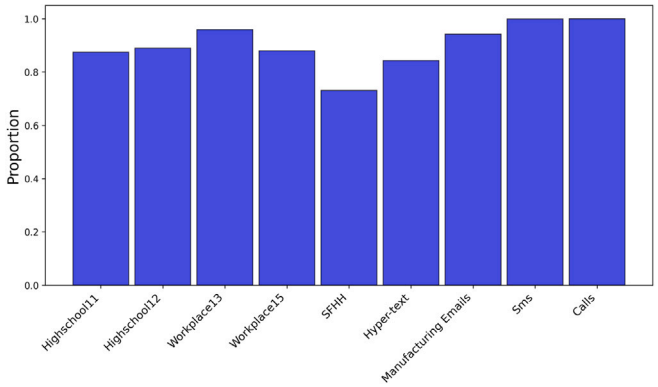


Fig. 8. Proportion of nodes that have degree 1 among nodes that have any contact in a snapshot, averaged over all snapshots of each real-world network.

Table 2
Complexity of computing each proposed centrality metric for all nodes based on each partial temporal network $\mathcal{G}_i(\phi, m)$.

d	Δ	Z
$\mathcal{O}(\mathcal{E}m)$	$\mathcal{O}(N\mathcal{L}^m)$	$\mathcal{O}(N\mathcal{L}^m)$

problem: β, ϕ, and the measure to evaluate the prediction quality, we compare the best performance achieved by each of the four classes of methods using classic metrics and the best performance achieved by our proposed centrality metrics in Fig. 9.

As shown in Fig. 9, either our proposed metrics or classic metrics derived from full temporal networks perform the best. This indicates that introducing temporal information in the design of centrality metrics could be beneficial. The centrality metrics we have proposed perform mostly the best. This suggests that those walk-related properties of a node in a partial network well indicate the influence of the node in the global network in the long term. It is found that when $\beta \to 0$, the average closeness over all snapshots of a temporal network G generally performs the best. This can be explained as follows: it is observed that in each snapshot of each temporal network, most contacts are disjoint; This is supported by high proportion of nodes that have degree 1 among nodes that have any contact in each snapshot, as illustrated in Fig. 8; Hence, the average closeness of a node over all snapshots of G approximates the average degree of this node over all snapshots, proportional to the total number of contacts of the node over all snapshots; As explained in Section 5.1, the influence of a node when $\beta \to 0$ is equal to the total number of contacts of the node in G; Hence, when $\beta \to 0$, the average closeness over all snapshots tends to provide the best estimate of nodal influence.

When β is large, the class of metrics we have proposed, actually time-scaled temporal reachability Z derived from the partial temporal network generally performs the best among all metrics. However, in the network Sms and Calls, temporal closeness centrality derived from G performs the best, when β is large. This can be attributed to the following reasons. Consider the case when $\beta = 1$. The influence of a target node in a temporal network is equal to the number of distinct nodes that are reachable by the target node via the shortest time-respecting paths. If the hop-count of the shortest time-respecting path from the target node to any node is either 1 or infinity (meaning a time-respecting path from the target to this node does not exist), the temporal closeness of the target node is equal to its influence. Fig. 10 shows that the probability $Pr[\{TH = 1\} \cup \{TH = \infty\}]$ that the hopcount of the shortest time-respecting path from a random node to another random node is either 1 or infinity is the largest, close to 1, in Sms and Calls, explaining why temporal closeness perform the best in these two networks, when β is large. This probability $Pr[\{TH = 1\} \cup \{TH = \infty\}]$ is large in Sms and Calls because these two networks have a low link density p and a large average modularity Γ of the largest connected component of the weighted aggregated network G^w, as reflected in Table 1. Nodes mainly interact within small communities, such that a node pair is likely either not connected via a time-respecting path when they belong to different communities or connected via a time-respecting path with a small hopcount when they belong to the same community.

In addition to evaluating the performance of the proposed centrality metrics in terms of predication quality, we briefly address their computational complexity. Let $\mathcal{E}$ represent the total number of contacts occurring during the partial observation period $[1, \phi\tau]$, and suppose the maximum number of contacts for any given node during this period is $\mathcal{L}$. The computational complexity of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z for all nodes derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ is shown in Table 2. The polynomial complexity of the metrics poses challenges in computing them in full temporal networks.

5.3. Evaluating proposed metrics in randomized real-world networks

The performance of an influence prediction method may depend on the properties of the underlying temporal network. The contact networks we considered manifest correlations between contacts. For example, contacts that are close in topology tend to be

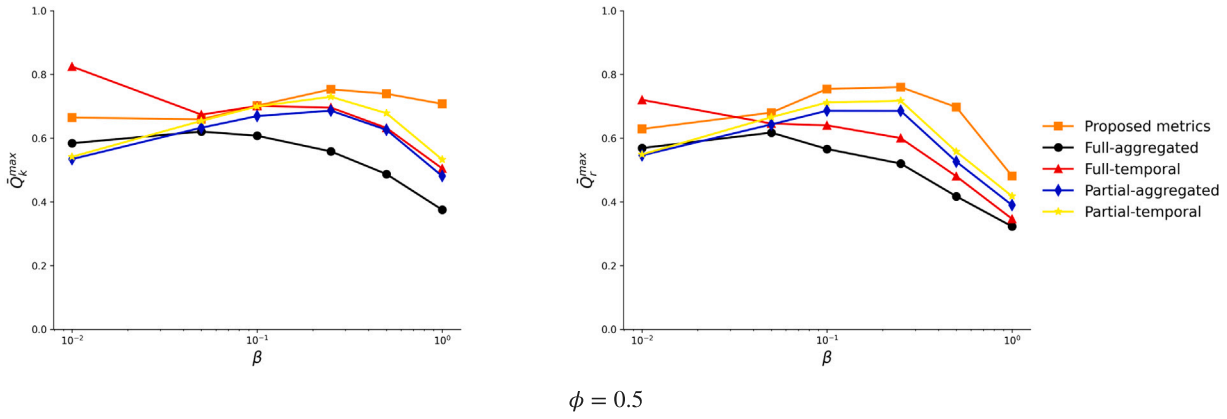


Fig. 9. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Highschool11.

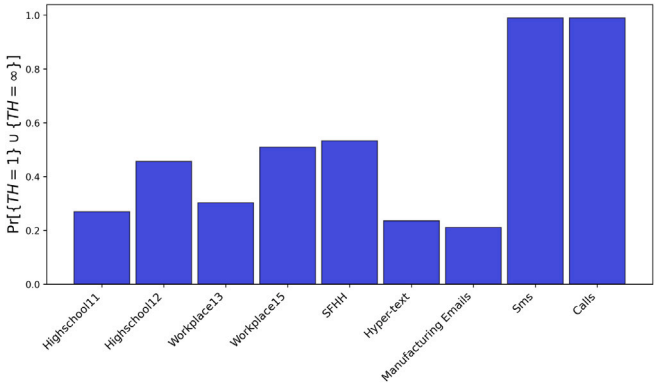
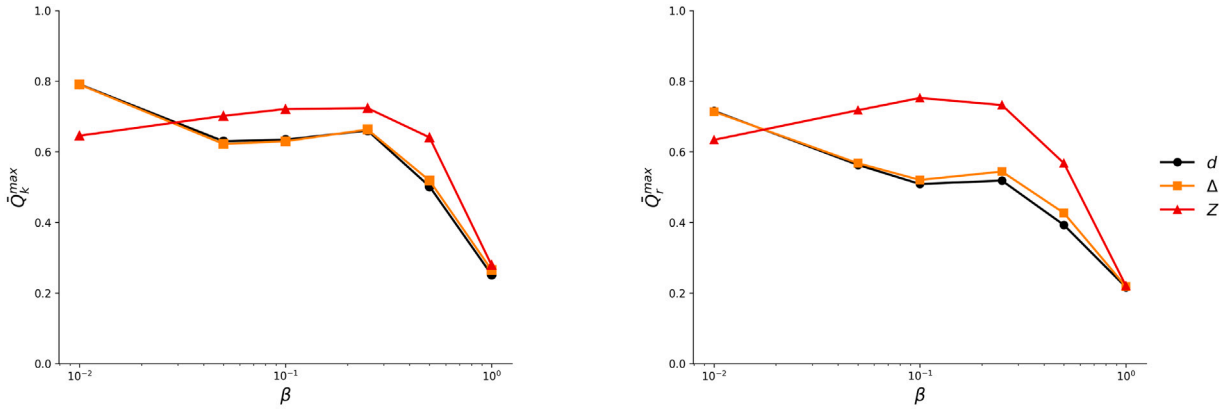


Fig. 10. Probability $Pr[\{TH = 1\} \cup \{TH = \infty\}]$ that the hopcount of the shortest time-respecting path from a random node to another random node is either 1 or infinity, averaged over all observation periods $t_0 + \tau$ of each real-world network.

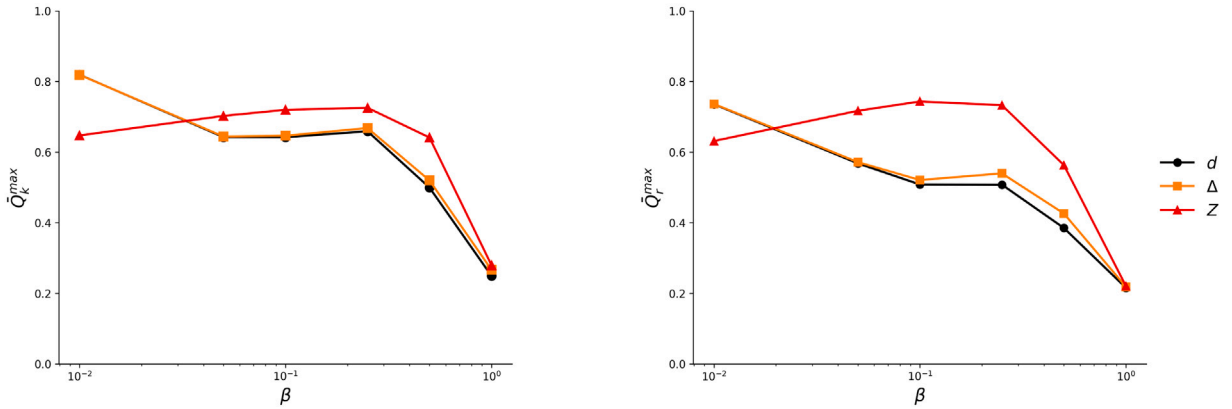
close in time. Hence, we explore further whether our findings in the performance of proposed metrics and of the classic metrics are still valid when these real-world networks are randomized, i.e., the correlation between topology and time are removed. Consider the set of contacts $l(i, j, t)$ in a temporal network G, where each contact is described by its topological location, i.e., between pair of nodes (i, j) and its time stamp t. A randomized network of G is obtained by reshuffling the time stamps of contacts in the network, without changing the topological locations of contacts [31]. This randomization does not change the number of contacts between each node pair. Only the time stamps of contacts are randomly switched.

As shown in Fig. 11, we observe the same result in real-world networks and in their randomized networks that the time-scaled reachability performs the best for a broad range of β that is not small and the other two proposed metrics perform the best when β is small. A different observation in randomized networks is that when $\beta = 1$, all proposed metrics perform similarly and badly. This is because the spreading influence of all nodes in the randomized network tends to be higher than that in the original network and tends to be identical when $\beta = 1$. For example, any seed node may lead to the infection of all the other nodes within the same connected component within $[1, \tau]$. The influence of nodes is hardly distinguishable in this case similar to the case when β is large in the Manufacturing Emails network. Meanwhile, it is observed that the Kendall's correlation between weighted degree mass d and time-scaled temporal degree mass Δ gets stronger after network randomization when $\alpha = 1$ and $m \in \{2, 3\}$. Intuitively, a node that has a large weighted degree mass d, i.e., many walks within m hops in the aggregated network of the partial network tends to have a large temporal degree mass, i.e., many time-respecting walks within m hops, because the time stamps are assigned randomly.

Furthermore, we compare the prediction quality of proposed metrics and classic metrics in the randomized real-world networks, as exemplified in Fig. 12. The same result has been found that our proposed centrality metric set performs mostly better than classic centrality metrics derived from both the full and partial temporal network, as well as their corresponding unweighted aggregated networks. The performance of our metrics tends to be robust to variations in the correlation between temporal and topological information of contacts.

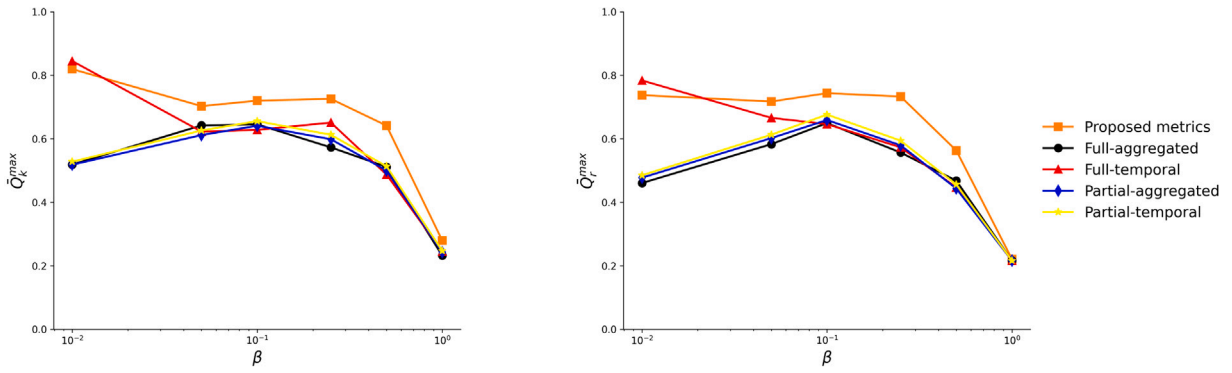


(a) $\phi = 0.25$



(b) $\phi = 0.5$

Fig. 11. The optimal prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d (denoted by black dot), time-scaled temporal degree mass Δ (denoted by orange squares), and time-scaled temporal reachability Z (denoted by red rectangle) across various ϕ and β, averaged on 10 randomized networks of Highschool11.



$\phi = 0.5$

Fig. 12. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, averaged on 10 randomized networks of Highschool11.

6. Conclusions and future work

In this work, we address the problem of using partial temporal network information, i.e., local network observed over a short period, to predict nodal spreading influence on the full temporal network over a long period. This study also aims to identify which network properties of a node in the partial network determine this node's influence. The spreading influence of a node depends on how well the node is connected to other nodes via possible spreading trajectories. The spreading trajectory from any seed node to any other node in a temporal network is not necessarily the shortest time-respecting path, but can be any time-respecting path. This motivates us to design centrality metrics that systematically capture how well a node is connected in the partial network via (time-respecting) walks. These metrics contrast with class metrics describing the connection of a node to other nodes via the shortest (time-respecting) paths. The quality of these metrics in estimating the ranking of nodes in influence is evaluated and compared against classical centrality metrics in real-world contact networks and their randomized networks. We find and explain that the proposed metrics using the partial network mostly outperform classic centrality metrics derived from the full temporal network, across a broad range of the infection probability. A node tends to be influential if it can reach many distinct nodes via time-respecting walks and if these nodes can be reached early in time.

This study has several limitations that call for further exploration. Firstly, this work estimates the influence of nodes in the SI process on temporal networks. The proposed methods can be applied and extended for other spreading processes, such as the SIR process. Epidemics and information may spread via higher-order (group) interactions or multilayer networks [32–38]. It is interesting to explore the feasibility of predicting nodal influence in diverse types of networks using partial network information. Secondly, how the parameters m and α of the proposed centrality metrics affect the prediction quality has been analyzed and explained. It is intriguing to investigate whether certain universal ranges of these parameters tend to lead to near-optimal estimation quality for certain types of networks. It is also interesting to explore how the prediction quality improves as the relative duration ϕ of observation increases, e.g., when $\phi > 0,5$. Thirdly, exploring the use of machine learning models to integrate the proposed and existing centrality metrics may further enhance the prediction of nodal influence. Finally, the polynomial computational complexity of the proposed metrics limits their scalability in large-scale temporal networks. Given their promising prediction quality, it is worthwhile to design efficient approximation algorithms.

CRediT authorship contribution statement

Tianrui Mao: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Shilun Zhang:** Methodology, Conceptualization. **Alan Hanjalic:** Supervision, Methodology. **Huijuan Wang:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank for the support of Netherlands Organization for Scientific Research NWO (project FORT-PORT no. KICH1.VE03.21.008) and the China Scholarship Council (CSC). We greatly appreciate Prof. Márton Karsai for his valuable comments and suggestions.

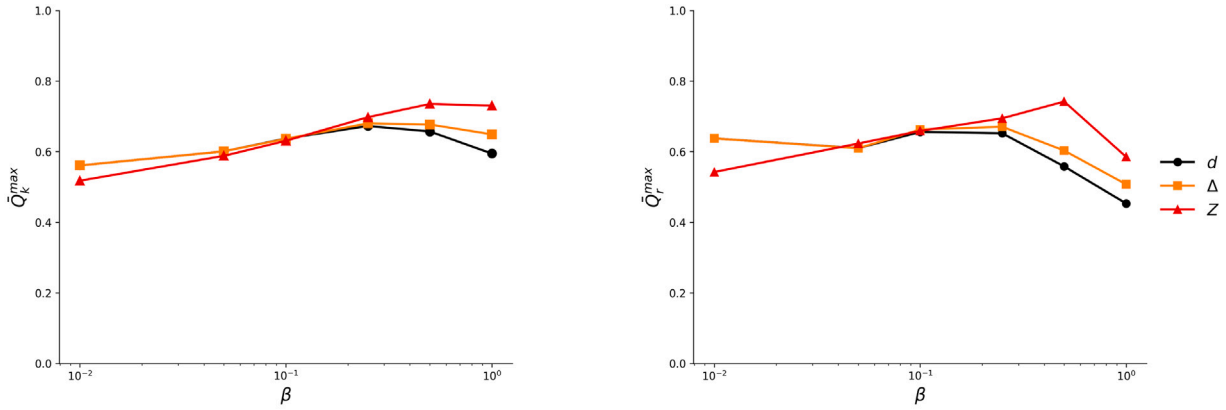
Appendix. Prediction quality of proposed metrics in other real-world networks

The appendix presents the prediction quality of three proposed centrality metrics in eight real-world networks, excluding HighSchool11 (see Figs. 13–20). Consistent with the observations in Section 5.1, we observe the same in these networks as in HighSchool11: the time-scaled temporal reachability Z generally achieves the best performance when β is not small; However, for small β, Z performs the worst whereas the other two metrics exhibit comparable performance.

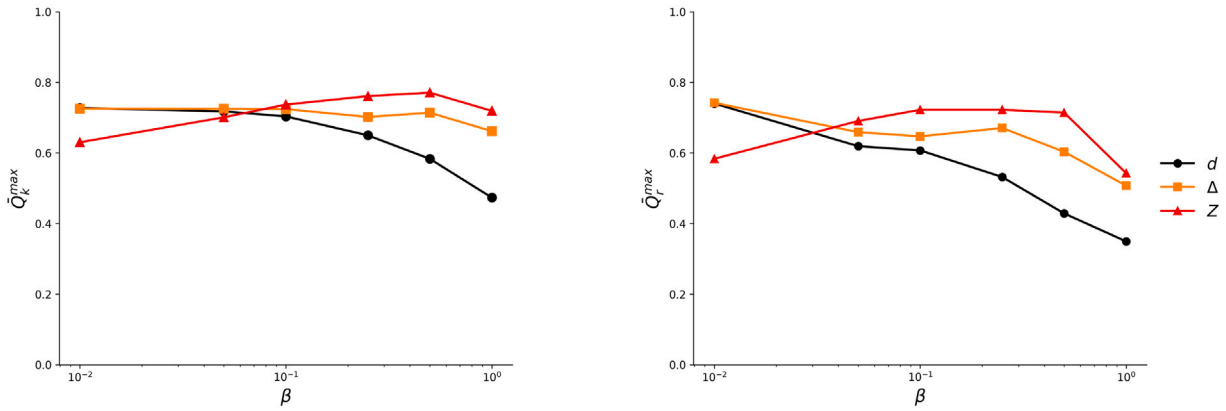
The comparison between the proposed centrality metrics and classic centrality metrics in the remaining eight real-world networks is shown in Figs. 21–28. The finding that the proposed centrality metrics mostly outperform classic centrality metrics in HighSchool11, discussed in Section 5.2, also holds for these eight real-world networks.

Data availability

Data will be made available on request.

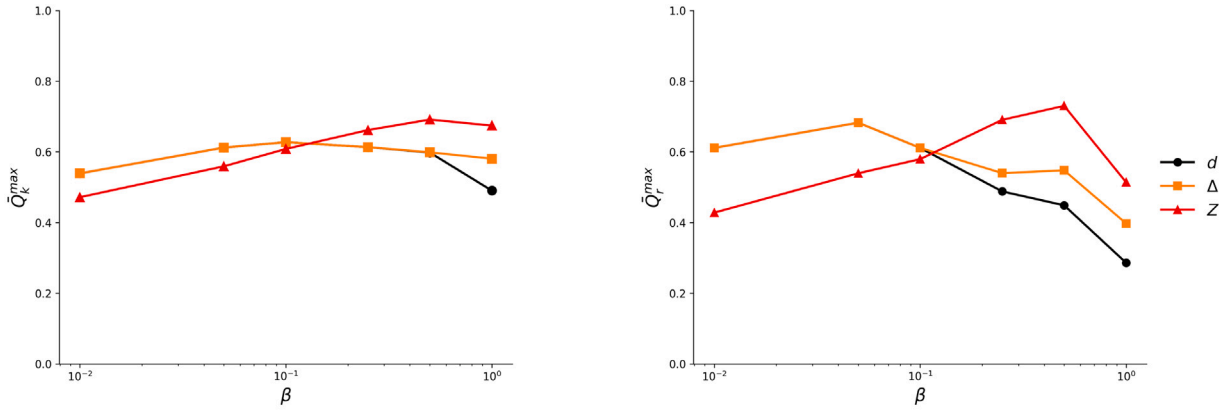


(a) $\phi = 0.25$

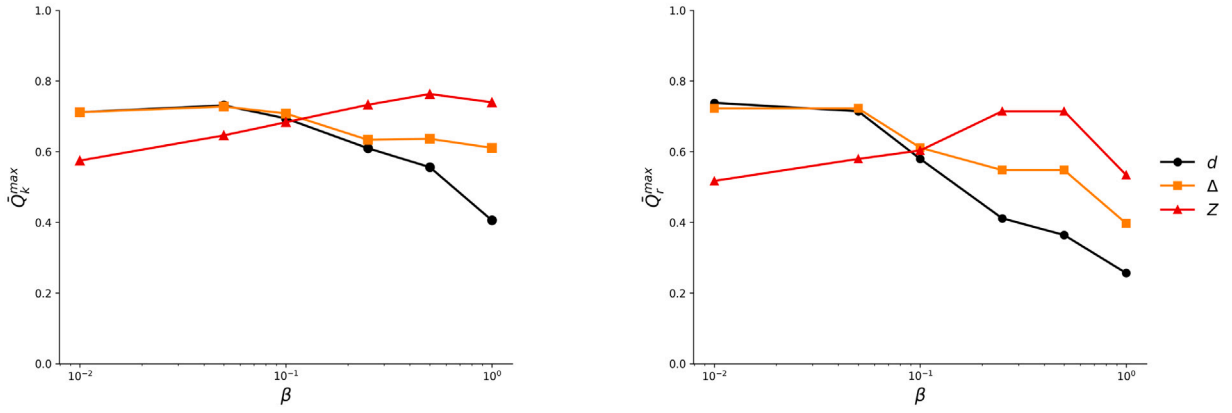


(b) $\phi = 0.5$

Fig. 13. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network HighSchool12.

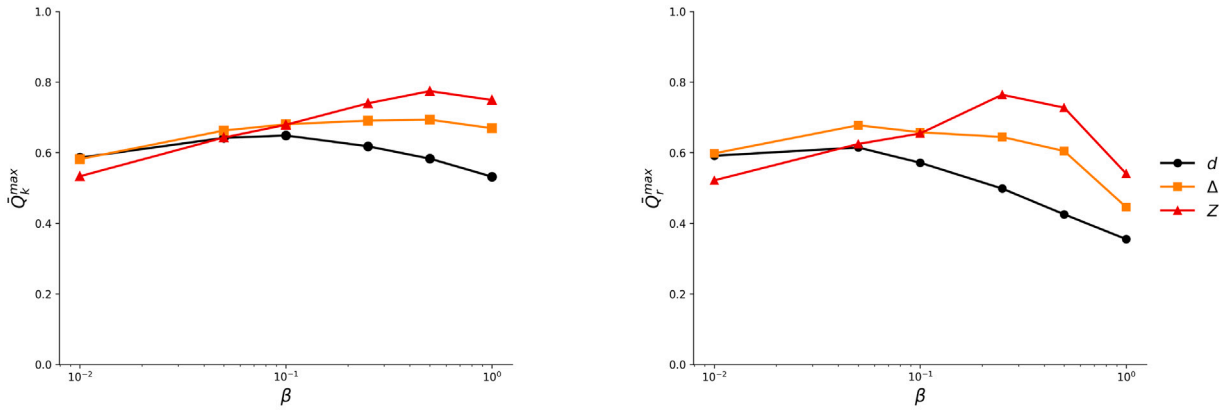


(a) $\phi = 0.25$

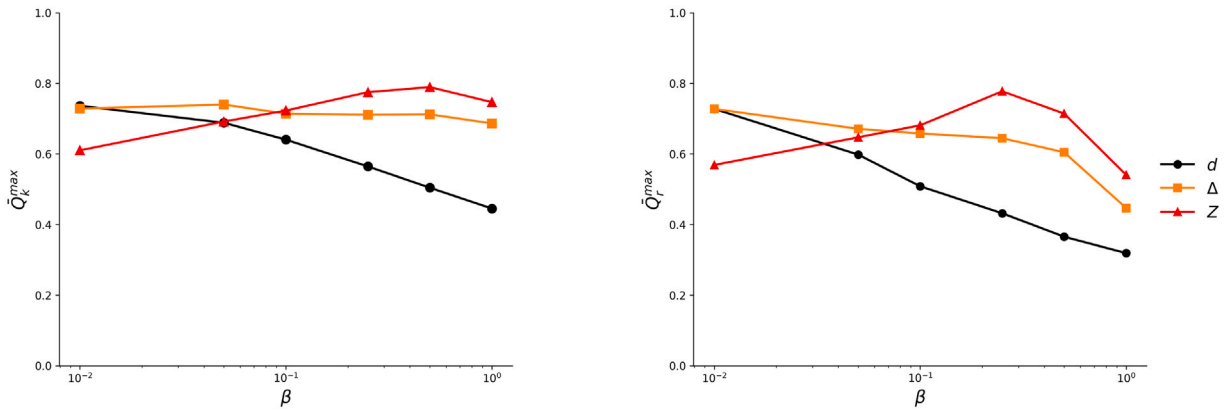


(b) $\phi = 0.5$

Fig. 14. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network Workplace13.

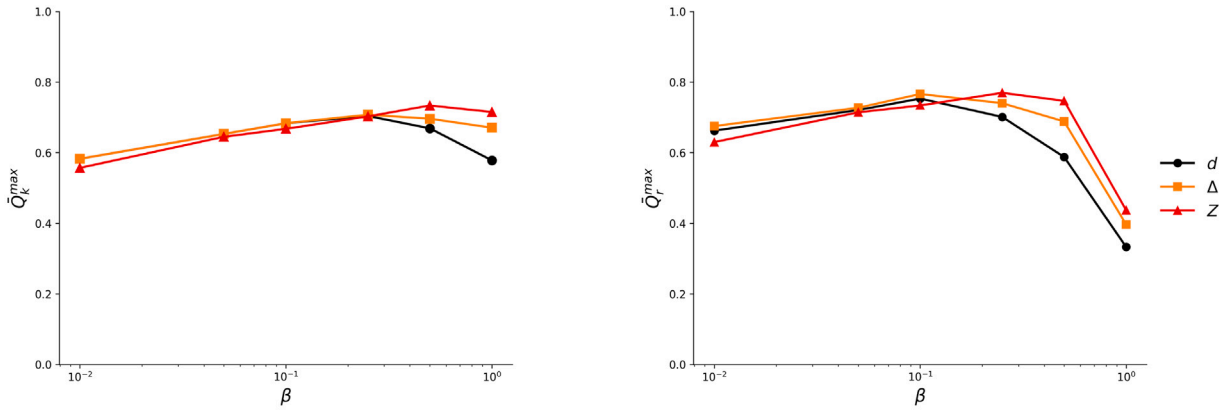


(a) $\phi = 0.25$

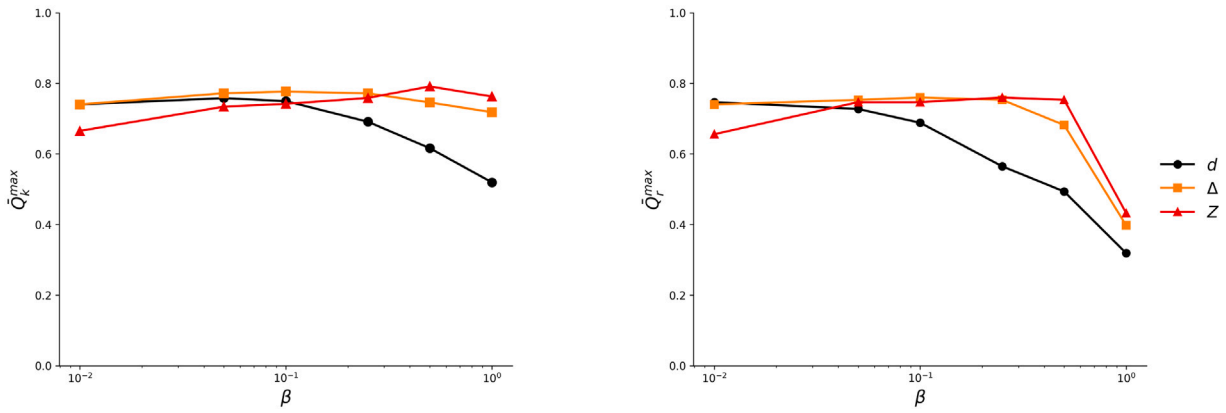


(b) $\phi = 0.5$

Fig. 15. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network Workplace15.

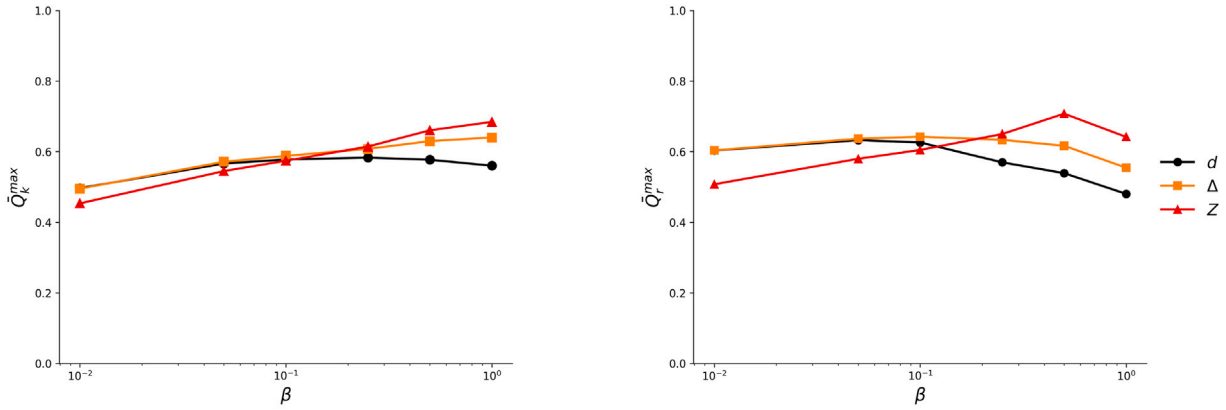


(a) $\phi = 0.25$

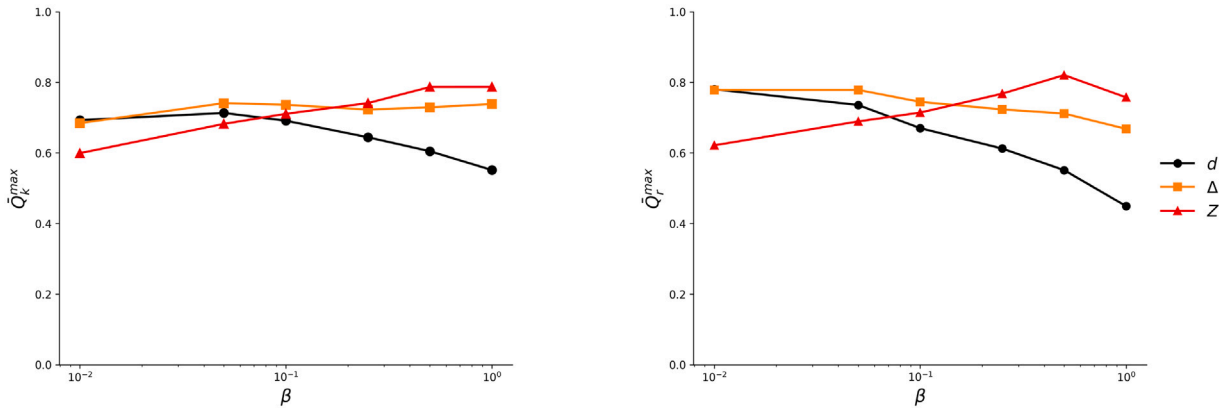


(b) $\phi = 0.5$

Fig. 16. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network Hyper-text.

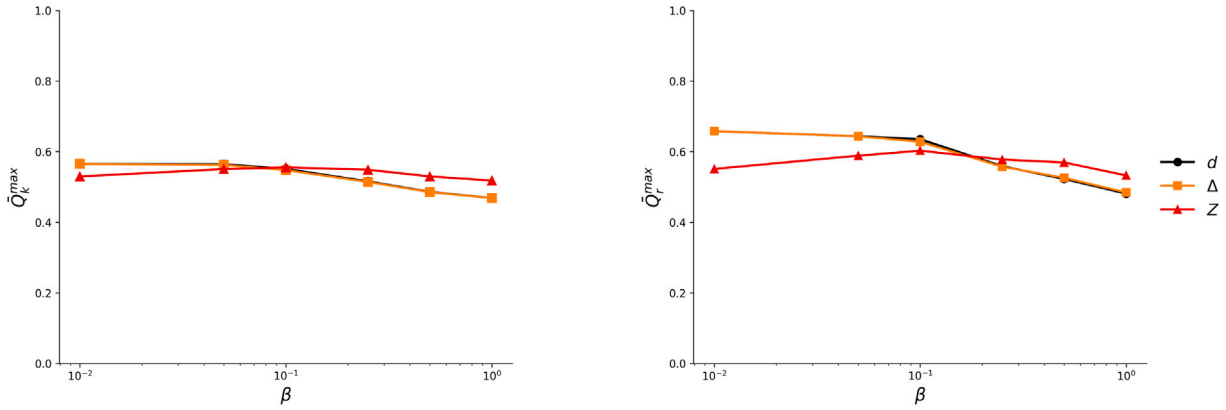


(a) $\phi = 0.25$

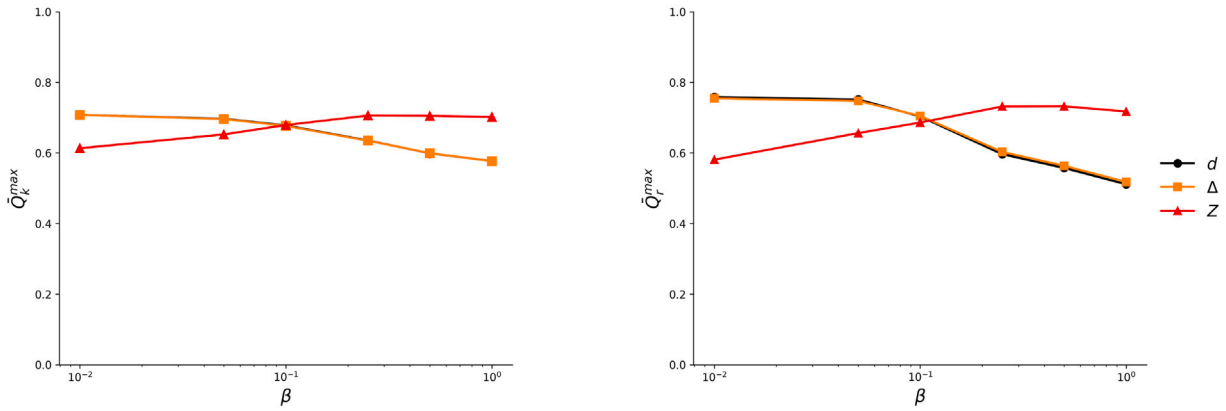


(b) $\phi = 0.5$

Fig. 17. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network SFHH.

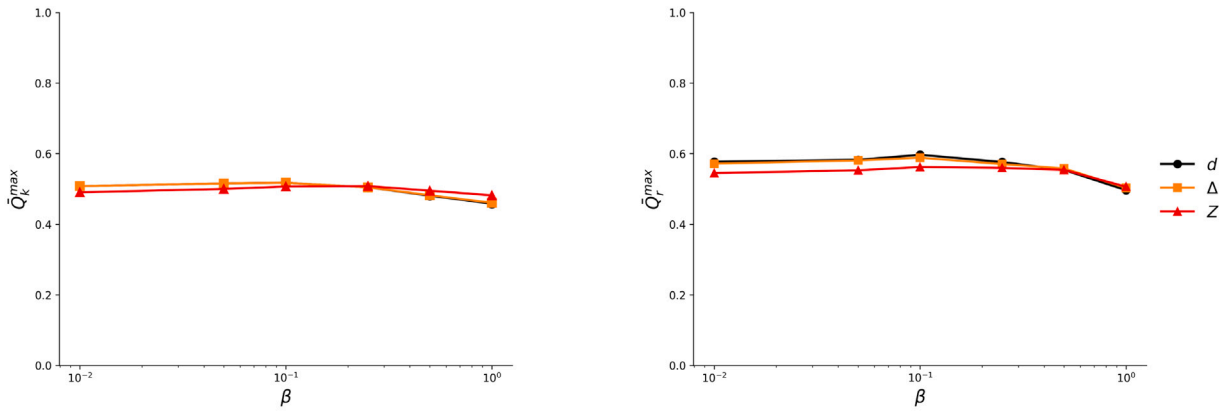


(a) $\phi = 0.25$

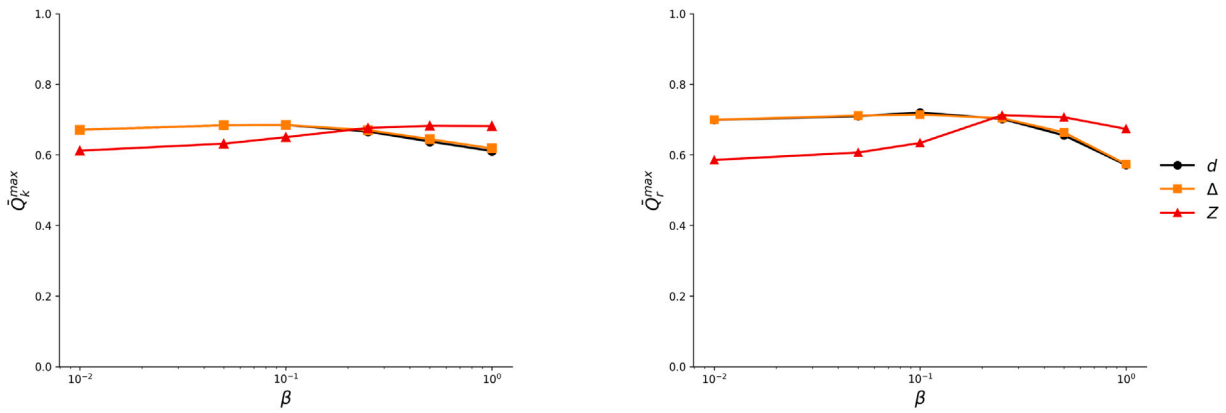


(b) $\phi = 0.5$

Fig. 18. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network Sms.



(a) $\phi = 0.25$



(b) $\phi = 0.5$

Fig. 19. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network Calls.

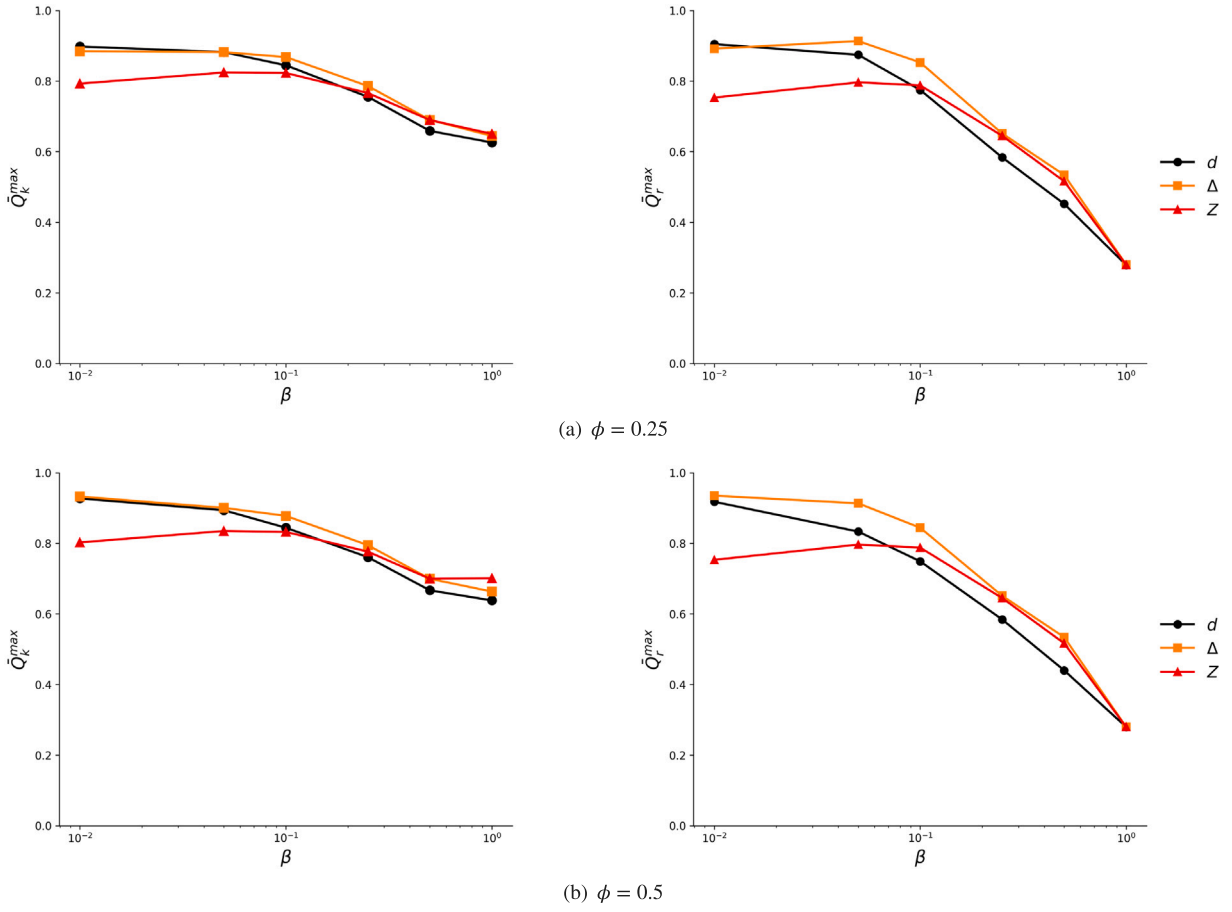


Fig. 20. The (best) prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$ of weighted degree mass d, time-scaled temporal degree mass Δ, and time-scaled temporal reachability Z, respectively, across various combination of ϕ and β, in network Manufacturing Emails.

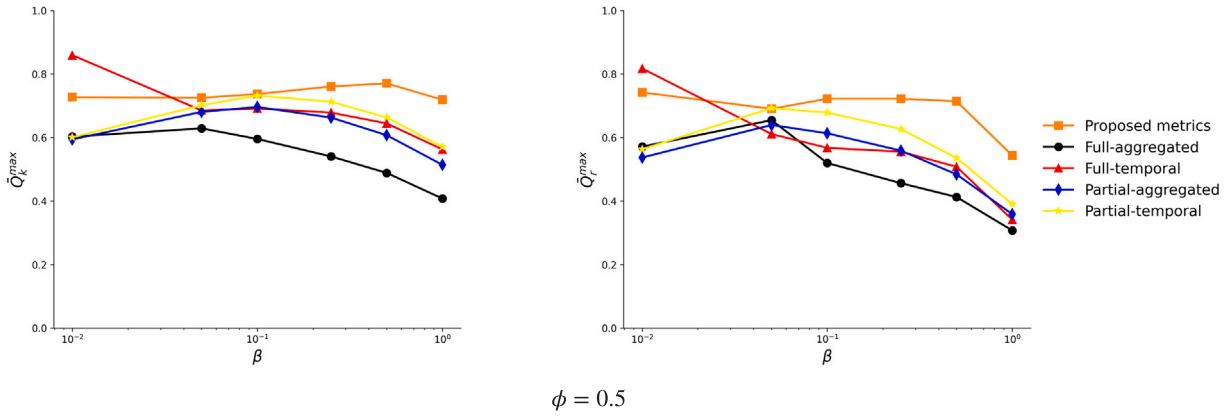


Fig. 21. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Highschool12.

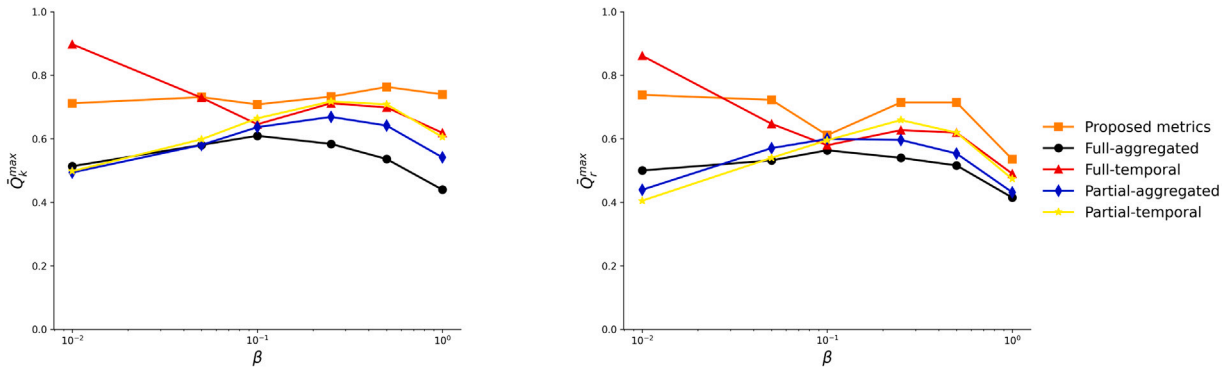


Fig. 22. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Workplace13.

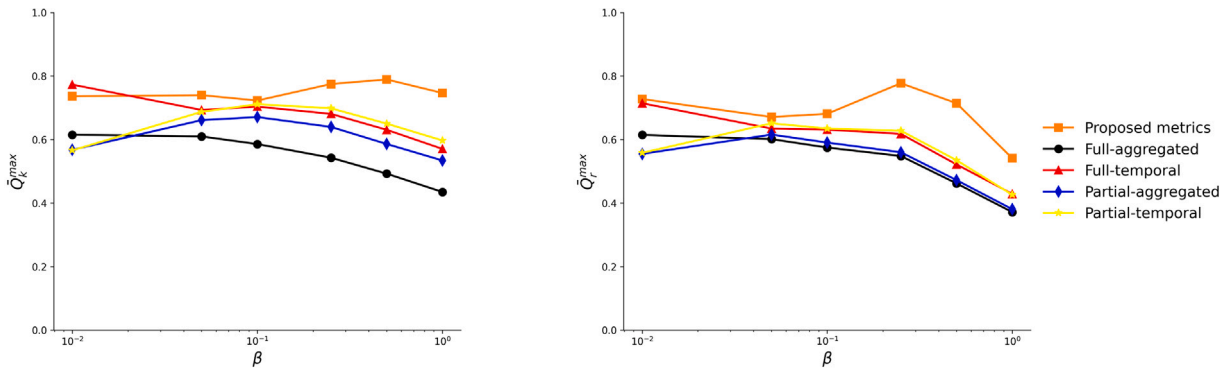


Fig. 23. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Workplace15.

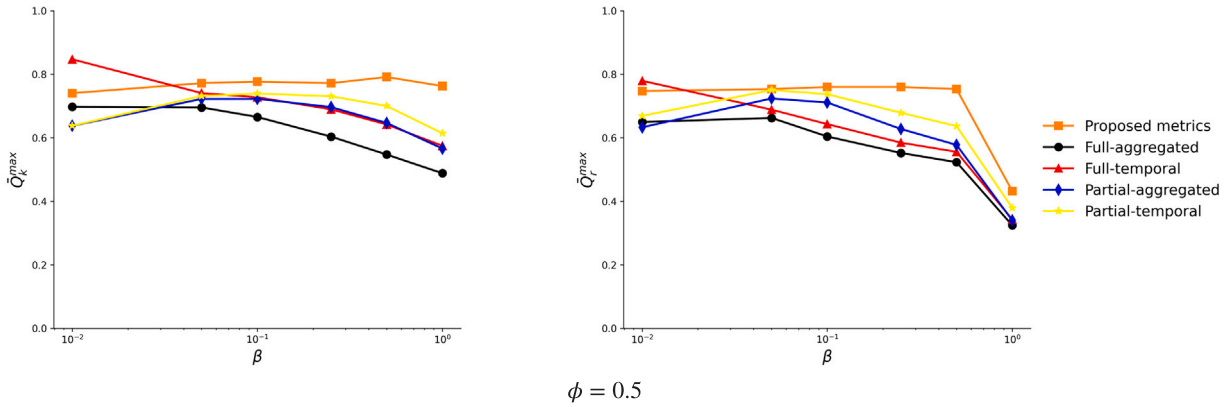


Fig. 24. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Hyper-text.

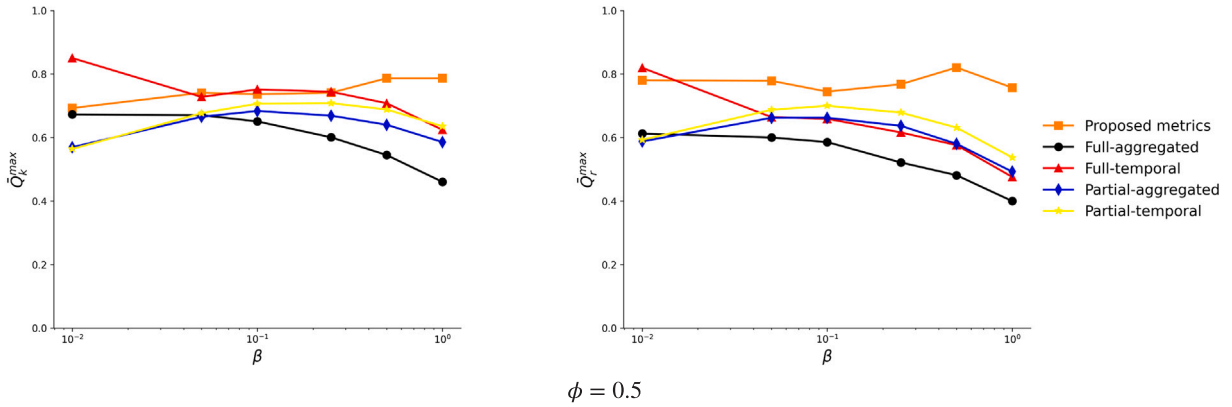


Fig. 25. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on SFHH.

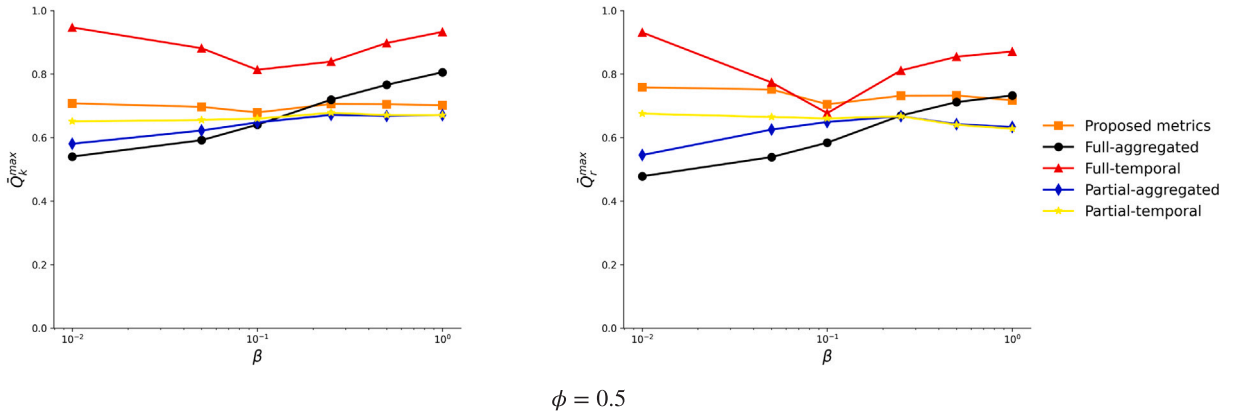


Fig. 26. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Sms.

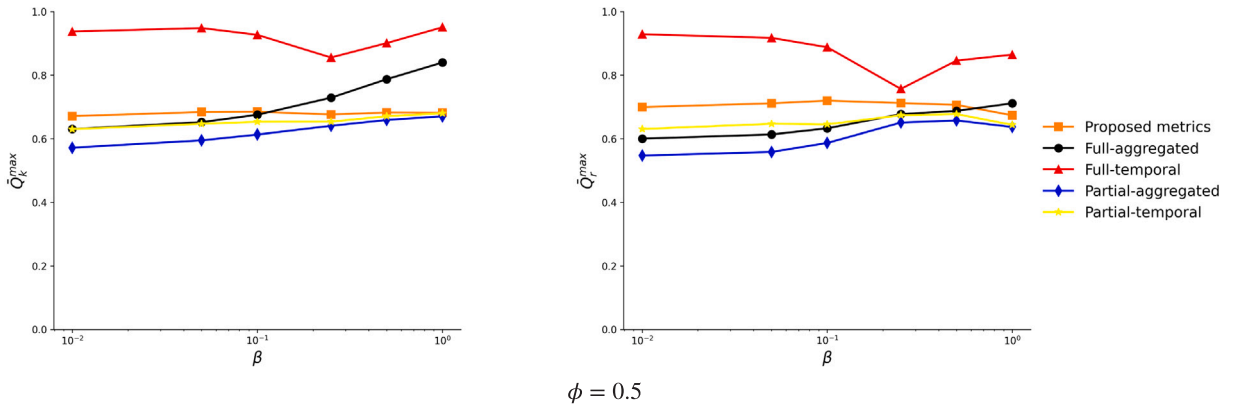


Fig. 27. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Calls.

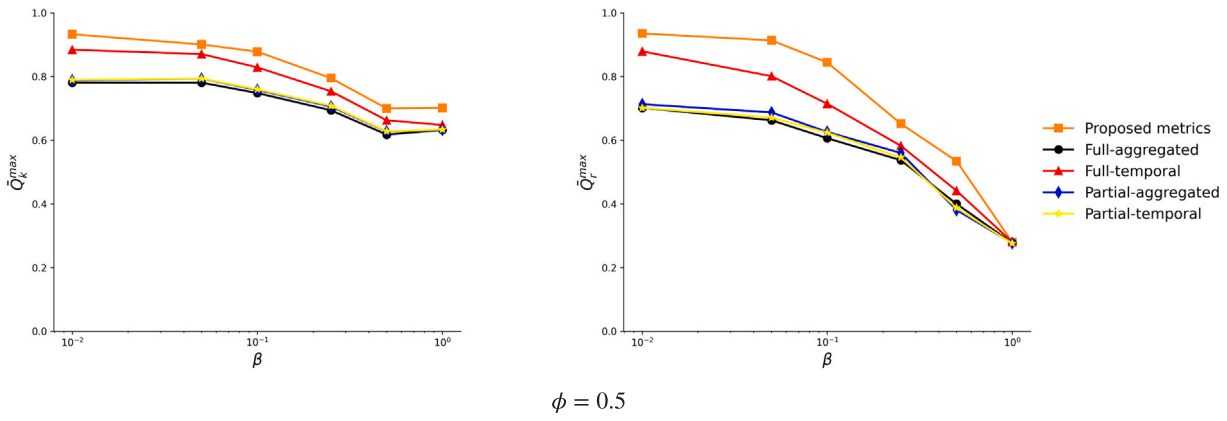


Fig. 28. The best prediction quality $\bar{Q}_k^{max}$ and $\bar{Q}_r^{max}$, respectively achieved by proposed centrality metrics derived from the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by orange squares); full-aggregated: each static centrality metric derived from the unweighted aggregated network of the full temporal network G (denoted by black dots); full-temporal: the average of each static centrality metric derived from all snapshots of G or the temporal closeness centrality derived from G (denoted by red triangles); partial-aggregated: each static centrality metric derived from the unweighted aggregated network of the partial temporal network $\mathcal{G}_i(\phi, m)$ (denoted by blue diamond); partial-temporal: the average of each static centrality metric derived from all snapshots of the partial temporal network $\mathcal{G}_i(\phi, m)$ or temporal closeness centrality derived from $\mathcal{G}_i(\phi, m)$ (denoted by yellow stars) when $\phi = 0.5$ and β varies, based on Manufacturing Emails.

References

[1] Kim H, Anderson R. Temporal node centrality in complex networks. Phys Rev E 2012;85(2):026107.
[2] Tsalouchidou I, Baeza-Yates R, Bonchi F, Liao K, Sellis T. Temporal betweenness centrality in dynamic graphs. Int J Data Sci Anal 2020;9:257–72.
[3] Ogura M, Preciado VM. Katz centrality of Markovian temporal networks: Analysis and optimization. In: 2017 American control conference. ACC, IEEE; 2017, p. 5001–6.
[4] Praprotnik S, Batagelj V. Spectral centrality measures in temporal networks. Ars Math Contemp 2015;11(1):11–33.
[5] Bi J, Jin J, Qu C, Zhan X, Wang G, Yan G. Temporal gravity model for important node identification in temporal networks. Chaos Solitons Fractals 2021;147:110934.
[6] Wu Z, He L, Tao L, Wang Y, Zhang Z. Temporal neighborhood change centrality for important node identification in temporal networks. In: International conference on neural information processing. Springer; 2022, p. 455–67.
[7] Oettershagen L, Mutzel P, Kriege NM. Temporal walk centrality: ranking nodes in evolving networks. In: Proceedings of the ACM web conference 2022. 2022, p. 1640–50.
[8] Tao L, Kong S, He L, Zhang F, Li X, Jia T, et al. A sequential-path tree-based centrality for identifying influential spreaders in temporal networks. Chaos Solitons Fractals 2022;165:112766.
[9] Bucur D. Top influencers can be identified universally by combining classical centralities. Sci Rep 2020;10(1):20550.
[10] Karoui W, Hafiene N, Ben Romdhane L. Machine learning-based method to predict influential nodes in dynamic social networks. Soc Netw Anal Min 2022;12(1):108.
[11] Yu E-Y, Fu Y, Chen X, Xie M, Chen D-B. Identifying critical nodes in temporal networks by network embedding. Sci Rep 2020;10(1):12494.
[12] Yu E, Fu Y, Zhou J, Sun H, Chen D. Predicting critical nodes in temporal networks by dynamic graph convolutional networks. Appl Sci 2023;13(12):7272.
[13] Hurley M, Jacobs G, Gilbert M. The basic SI model. New Dir Teach Learn 2006;2006(106):11–22.
[14] Meghanathan N. A computationally lightweight and localized centrality metric in lieu of betweenness centrality for complex network analysis. Vietnam J Comput Sci 2017;4:23–38.
[15] Zhang S, Hanjalic A, Wang H. Predicting nodal influence via local iterative metrics. Sci Rep 2024;14(1):4929.
[16] Zhang Q, Karsai M, Vespignani A. Link transmission centrality in large-scale social networks. EPJ Data Sci 2018;7(1):33.
[17] Li C, Li Q, Van Mieghem P, Stanley HE, Wang H. Correlation between centrality metrics and their application to the opinion model. Eur Phys J B 2015;88:1–13.
[18] Wang H, Hernandez JM, Van Mieghem P. Betweenness centrality in a weighted network. Phys Rev E 2008;77(4):046105.
[19] Sabidussi G. The centrality index of a graph. Psychometrika 1966;31(4):581–603.
[20] Bonacich P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol 1972;2(1):113–20.
[21] Gleich DF. PageRank beyond the web. Siam Review 2015;57(3):321–63.
[22] Wu H, Cheng J, Ke Y, Huang S, Huang Y, Wu H. Efficient algorithms for temporal path computation. IEEE Trans Knowl Data Eng 2016;28(11):2927–42.
[23] Fournet J, Barrat A. Contact patterns among high school students. PloS One 2014;9(9):e107878.
[24] Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Netw Sci 2015;3(3):326–47.
[25] Isella L, Stehlé J, Barrat A, Cattuto C, Pinton J-F, Van den Broeck W. What's in a crowd? Analysis of face-to-face behavioral networks. J Theoret Biol 2011;271(1):166–80.
[26] Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton J-F, Vespignani A. Dynamics of person-to-person interactions from distributed RFID sensor networks. PloS One 2010;5(7):e11596.
[27] Sapiezynski P, Stopczynski A, Lassen DD, Lehmann S. Interaction data from the copenhagen networks study. Sci Data 2019;6(1):315.
[28] Kunegis J. Konect: the koblenz network collection. In: Proceedings of the 22nd international conference on world wide web. 2013, p. 1343–50.
[29] Zou L, Wang A, Wang H. Memory based temporal network prediction. In: International conference on complex networks and their applications. Springer; 2022, p. 661–73.
[30] Kendall MG. The treatment of ties in ranking problems. Biometrika 1945;33(3):239–51.
[31] Ceria A, Havlin S, Hanjalic A, Wang H. Topological–temporal properties of evolving networks. J Complex Networks 2022;10(5):cnac041.

[32] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, et al. Networks beyond pairwise interactions: Structure and dynamics. Phys Rep 2020;874:1–92.
[33] Aktas ME, Nguyen T, Jawaid S, Riza R, Akbas E. Identifying critical higher-order interactions in complex networks. Sci Rep 2021;11(1):21288.
[34] Peters H, Ceria A, Wang H. Higher-order temporal network prediction and interpretation. PLoS One 2025;20(5):e0323753.
[35] Iacopini I, Karsai M, Barrat A. The temporal dynamics of group interactions in higher-order social networks. Nat Commun 2024;15(1):7391.
[36] Gallo L, Lacasa L, Latora V, Battiston F. Higher-order correlations reveal complex memory in temporal hypergraphs. Nat Commun 2024;15(1):4754.
[37] Zhang S-S, Xie M, Liu C, Zhan X-X. Influence maximization in multilayer networks based on adaptive coupling degree. Chaos: An Interdiscip J Nonlinear Sci 2025;35(3).
[38] Zhang H, Cao L, Fu C, Cai S, Gao Y. Epidemic spreading on multi-layer networks with active nodes. Chaos: An Interdiscip J Nonlinear Sci 2023;33(7).