



HAL
open science

Semantic Centrality for Temporal Graphs

Landy Andriamampianina, Franck Ravat, Jiefu Song, Nathalie
Vallès-Parlangeau

► **To cite this version:**

Landy Andriamampianina, Franck Ravat, Jiefu Song, Nathalie Vallès-Parlangeau. Semantic Centrality for Temporal Graphs. 27th European Conference on Advances in Databases and Information Systems (ADBIS 2023), Sep 2023, Barcelona, Spain. pp.163-173, 10.1007/978-3-031-42941-5_15 . hal-04240413

HAL Id: hal-04240413

<https://hal.science/hal-04240413v1>

Submitted on 13 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic Centrality for Temporal Graphs

Landy Andriamampianina^{1,2}[0000-0001-9636-7007], Franck Ravat¹[0000-0003-4820-841X], Jiefu Song¹[0000-0002-2066-7051], and Nathalie Vallès-Parlangeau³[0000-0002-4463-5177]

¹ IRIT-CNRS (UMR 5505) - Université Toulouse Capitole, 2 Rue du Doyen Gabriel Marty F-31042 Toulouse Cedex 09, France

`firstname.lastname@irit.fr`

² Activus Group, 1 Chemin du Pigeonnier de la Cépière, 31100 Toulouse, France

`firstname.lastname@activus-group.fr`

³ Université de Pau et des Pays de l'Adour, France

`firstname.lastname@univ-pau.fr`

Abstract. Centrality metrics in graphs, such as degree, help to understand the influence of entities in various applications. They are used to quantify the influence of entities based on their relationships. Time dimension has been integrated to take into account the evolution of relationships between entities in real-world phenomena. For instance, in the context of disease spreading, new contacts may appear and disappear over time between individuals. However, they do not take into account the semantics of entities and their relationships. For example, in the context of a disease spreading, some relationships (such as physical contacts) may be more important than others (such as virtual contacts). To overcome this drawback, we propose centrality metrics that integrate both temporal and semantics aspects. We carry out experimental assessments, with real-world datasets, to illustrate the efficiency of our solution.

Keywords: Centrality Metrics · Degree · Influential Entities · Temporal Graphs.

1 Introduction

Understanding the influence of entities is a major issue in various real-world phenomena such as social, biological, epidemiological phenomena [8]. For example, in a social network, it is important to identify the individuals who have the greatest influence to understand the spread of information [4]. To do so, several centrality metrics from Graph Theory quantify the influence of vertices based on the graph structure, such as the degree, betweenness or closeness centrality [8]. For instance, in Fig. 1 a), the vertex A is the most influential because it has the highest degree, i.e., the highest number of direct relationships. Yet, real-world phenomena may be dynamic in nature. So the influence of entities may also evolve. For instance, in Fig. 1 b), the most influential vertex changes between the two timestamps t_1 and t_2 because of the addition and removal of certain relationships. Current research is therefore aimed at extending classic metrics

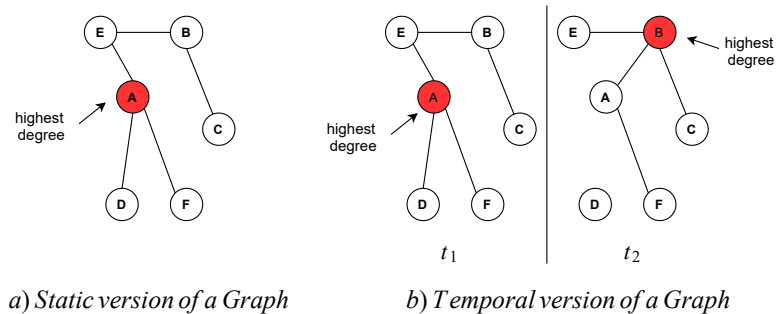


Fig. 1: Degree Centrality in static graphs VS temporal graphs.

by integrating temporal aspects (e.g. timestamps, time intervals, and temporal paths) [3]. However, these temporal metrics do not take into account the semantics of their application. They do not distinguish classes of entities and types of relationships. For instance, in a social network, some classes of entities (e.g., such as students, professionals) and some relationship types (e.g., family, friends, co-workers) may have different levels of influence. Therefore, our research question is the following: how can we improve the concept of centrality to consider both temporal and semantic aspects of an application?

The contribution of our work is the new concept of Semantic Temporal Centrality. It extends the basic concept of centrality to include both temporal and semantic aspects of an application. We use the degree metric as a simple baseline for understanding the importance of both aspects in centrality [8]. In this paper, we first review the calculation methods of degree in dynamic applications (Section 2). Then, we define our new concept using as an instantiation example, a graph model including both temporal evolution and semantics (Section 3). Finally, we evaluate the efficiency of our solution through experiments (Section 4).

2 Related Work

Existing calculation methods for centrality metrics in dynamic applications depend on the chosen metric and on how the time dimension is integrated in the graph. In the case of degree metric, the classic calculation method is the snapshot-based approach: the graph is modelled through a sequence of static graphs observed at different time points [7]. The centrality degree is calculated for each snapshot. The calculated degrees are generally exploited to produce a centrality ranking or distribution of vertices for each snapshot [5].

Another calculation method of degree metric is the time-aggregated approach. It consists in aggregating the time-series of degree values (calculated according to the snapshot-based approach) to obtain a single degree for each vertex. The aggregated degree value summarizes the overall influence of a vertex over time [4,5]. Generally, the average of degree values is used, but some research

works propose also to use the minimum or maximum degree [6]. Therefore, instead of generating several rankings or distributions as in the snapshot-based approach, the aggregated-approach provides a single result.

The above-mentioned degree metrics have several limitations. On the one hand, they are only calculated at a single time point because graph snapshots are based on discrete-time. They do not integrate calculation over a time interval to consider their evolution over a time. On the other hand, current degree calculation methods do not take into account the semantics of vertices and edges because most graph models do not. Yet, vertex and edge semantics help distinguish the levels of influence of entity classes and relationship types from a qualitative perspective. We therefore extend the current methods for calculating degrees by introducing a time interval into the temporal parameter and the concept of vertex and edge label to represent the semantics in an application.

3 Proposition

3.1 Preliminaries

We propose to use the model of temporal graph presented in [1] to exemplify our concept of Semantic Temporal Degree. The temporal graph $TG = \langle E, R \rangle$ is composed of sets of entities E and relationships R . It represents the semantics of entities and relationships by a set of labels L_E describing entity classes and a set of labels L_R describing relationship types. Each entity has an explicit semantic described by a label $l_{E'}$ and, similarly, each relationship with a label $l_{R'}$.

The temporal graph captures the evolution of entities and relationships over time. An entity is then described by an identifier id and a set of states $\{s_1, \dots, s_m\}$. A relationship r_i links a couple of entity states (s_k, s_j) and is composed by a set of states $\{s_1, \dots, s_n\}$. Each state s of an entity or relationship is associated to a valid time interval $T^s = [t_b, t_f[$, which indicates the stability period of the state. The set of relationships incident to an entity e_i is denoted, $R(e_i) = \{r_1, \dots, r_p\}$. The function $\Sigma : r_i \rightarrow \{s_1, \dots, s_n\}$ returns for a relationship r_i all of its states.

Some operators allow querying the temporal graph to select the semantics and time window of interest in the degree calculation [2]. On the one hand, the operator $matching_{predicate} : TG_{input} \times w \rightarrow TG_{output}$ allows extracting a subgraph representing the entities and relationships valid during a user-defined time-window w . On the other hand, the operator $matching_{pattern} : TG_{input} \times (l_{E_i} \times l_{R_j} \times l_{E_k}) \rightarrow TG_{output}$ allows extracting a subgraph respecting user-defined conditions on the semantics of entities and relationships.

Example 1. We consider the temporal graph in Figure 2 to analyse the number of interactions (degree) between individuals of a university in the context of disease spreading over a time period of 6 days. In the university, we distinguish students from teachers. Each entity is then represented by a grey vertex labelled with its semantics (*student* or *teacher*). Moreover, we distinguish two types of interactions between students and teachers: virtual contact (*call* relationship) and physical contact (*socialize* or *attend class* relationship). Taking into account the

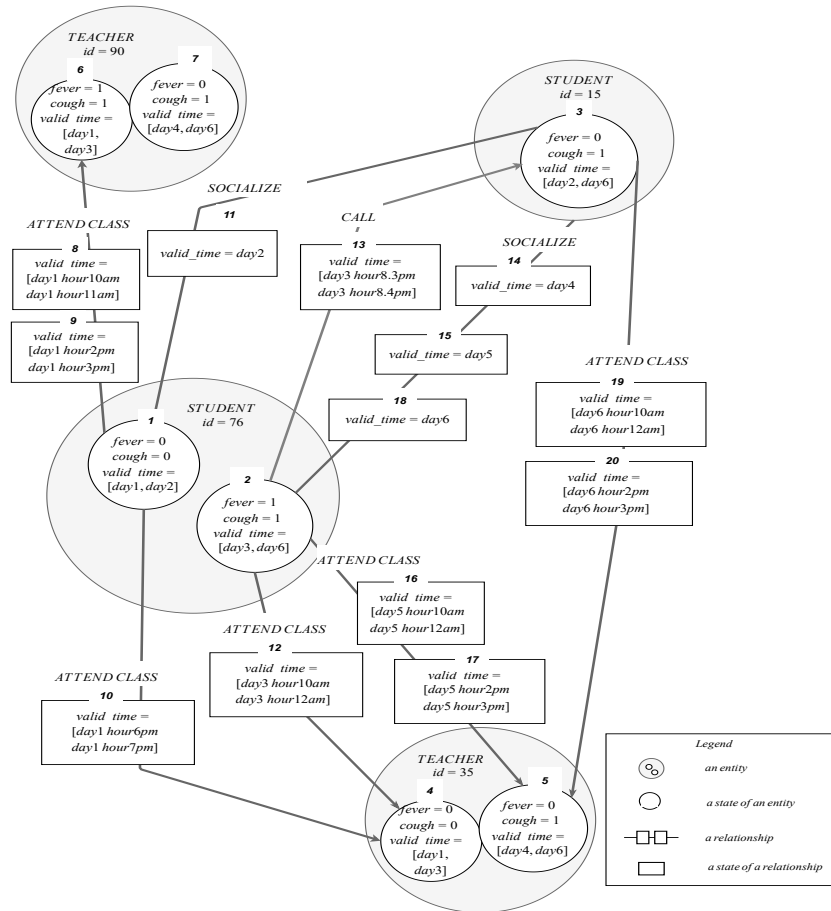


Fig. 2: An example of temporal graph

semantics during analyses can make a great difference. In this context, we can ignore *virtual contacts* and put more emphasis on *physical contacts* of teachers and students, who generally meet more people during classes. The health status of students and teachers can change over time due to the disease transmission. Each entity (grey vertex) is then composed of one or several states (white vertices). For instance, the student 76 has two different health status: the state 1 in which he has no fever and no cough from day 1 to day 2, and the state 2 in which he has fever and cough from day 3 to day 6. Moreover, new interactions between students and teachers may appear and disappear over time. Each interaction represents a state of a labelled relationship between two entity states and is illustrated by a white rectangle. For instance, when the student 15 is in the state 3 and the student 76 is in the state 2, they socialized three days in the period (rectangles numbered 14, 15 and 18).

3.2 Centrality Metrics

Existing research works only consider the degree metric at a specific time point and do not consider the semantics of their application (Section 2). We therefore propose three centrality metrics extending the current degree metric with the possibility of calculating its evolution within a time interval and by including the semantics of entities and relationships.

Definition 1 (semantic temporal degree). *The semantic temporal degree of an entity e_i in a temporal graph TG is calculated as a function of a time window w (a time point or a time interval), and as a function of a type $l_{R'}$ of the incident relationships to the entity. The semantic temporal degree of an entity at a time point $w = t_x$ is a single value representing the number of relationship states having the type $l_{R'}$, incident to e_i , and valid at the time point. The semantic temporal degree of an entity during a time interval $w = [t_{start}, t_{end}]$ is a series of values describing the changes in the number of relationship states having the type $l_{R'}$, incident to e_i , and valid during the time interval. If the entity does not exist during the time window w , the semantic temporal degree of the entity is not calculated. It is then denoted as follows:*

$$deg(e_i, w, l_{R'}) = \begin{cases} |\cup_{r_j \in R(e_i)} \Sigma(r_j)|, & \text{where } \forall r_j, l_{r_j} = l_{R'}, \forall s \in \Sigma(r_j), T^s \circ w, \text{ if } w = t_x \\ \{y_1, \dots, y_m\}, & \text{where } y = deg(e_i, t_x, l_{R'}), t_x \in w, \text{ if } w = [t_{start}, t_{end}] \\ \text{not defined,} & \text{otherwise} \end{cases}$$

The calculation of semantic temporal degree is detailed in Table 1. It involves the extraction of a subgraph respecting a given semantic pattern using the $matching_{pattern}$ operator and the extraction of a subgraph which is valid in the given time window using the $matching_{predicate}$ operator.

Example 2. Consider the temporal graph in Fig 2. We want to know how evolve the interactions of the student 76 over the whole time period (6 days). To do so, we can compute its temporal degree: $deg(76, [day1, day6], \emptyset) = \{deg(76, day1, \emptyset), deg(76, day2, \emptyset), deg(76, day3, \emptyset), deg(76, day4, \emptyset), deg(76, day5, \emptyset), deg(76, day6, \emptyset)\} = \{3, 1, 2, 1, 3, 1\}$. It is graphically illustrated in Fig 3 a). The student 76 has peaks of interactions at the days 3 and 5. This could indicate that on those days the student is at a higher risk of contracting the disease or that they play an important role in spreading the disease to others. However, we do not know through what type of interactions. Therefore, we compute the semantic temporal degree of the student 76 according to the different types of interactions from *day1* to *day6*. In Fig 3 b), we observe that the student actively socialized and attended class over time, but called rarely. Moreover, the number of socializations of the student are almost the same over time. The spikes and drops of degree values observed in Fig 3 a) are therefore not caused by calls and socializations. We should then focus on the other semantics of relationships: *attend class*. As we notice in Fig 3 b), the student can attend multiple classes throughout the day. This could indicate that class interactions of the student

Table 1: Semantic Temporal Degree

Metric:	$deg(e_i, w, l_{R'})$
Input:	a temporal graph TG an entity e_i a time window w a relationship label $l_{R'}$
Output:	a single value or a set of values or an empty set
Actions:	<ol style="list-style-type: none"> 1. $TG_1 \leftarrow matching_pattern(TG, l_{E'}, l_{R'}, \emptyset)$ where $l_{E'}$ is the label of the entity e_i 2. If $w = t_x$ 3. $TG_2 \leftarrow matching_predicate(TG_1, \{(l_{E'}.id = id^{e_i})$ $AND(T^{l_{R'}} \circ t_x)\})$ 4. Apply to TG_2: $deg(e_i, w, l_{R'}) = \cup_{r_j \in R(e_i)} \Sigma(r_j)$ 5. Else if $w = [t_{start}, t_{end}]$ 6. $Y \leftarrow \emptyset$ 8. For each $t_x \in w$ 9. $TG_2 \leftarrow matching_predicate(TG_1, \{(l_{E'}.id = id^{e_i})$ $AND(T^{l_{R'}} \circ t_x)\})$ 10. $Y \leftarrow Y \cup \{deg(e_i, t_x, l_{R'}) \text{ in } TG_2\}$ 11. $deg(e_i, w, l_{R'}) = Y$

play an important role in spreading the disease rather than other interaction types.

As discussed in Section 2, existing research work uses time-aggregated calculation methods to summarize the degree metric values for a time interval into a single value. We provide the same calculation possibility for the semantic temporal degree during a time interval by computing its average.

Definition 2 (average semantic temporal degree). *The average semantic temporal degree of an entity e_i in a temporal graph TG during the time interval w is the average where each value is the semantic temporal degree of the entity weighted by its duration δ in the time interval. It is denoted as follows:*

$$deg_{average}(e_i, w, l_{R'}) = \frac{\delta_1 \times deg_1(e_i, w_1, l_{R'}) + \dots + \delta_n \times deg_n(e_i, w_n, l_{R'})}{\sum_{p \in [1, n]} \delta_p}$$

Example 3. Let us compute the average semantic temporal degree of the student 76 over the entire period for its interactions with the semantics *attend class*. We have: $deg_{average}(76, [day1, day6], ATTENDCLASS) = \frac{1day \times 3 + 1day \times 1 + 1day \times 2 + 3days \times 0}{6days} = 1$. In average, the student 76 attends one class in average in the week.

Existing research work generally focuses solely on the degree of a vertex. We enable the calculation of the distribution of a class of vertices through the *semantic temporal degree distribution* metric. The advantage of the latter is that it highlights the level of heterogeneity of the semantic temporal degree within a class of entities, which cannot be observed with the simple centrality ranking. Specifically, the semantic temporal degree of a chosen entity can be positioned in the distribution of its entity class to understand its relative importance.

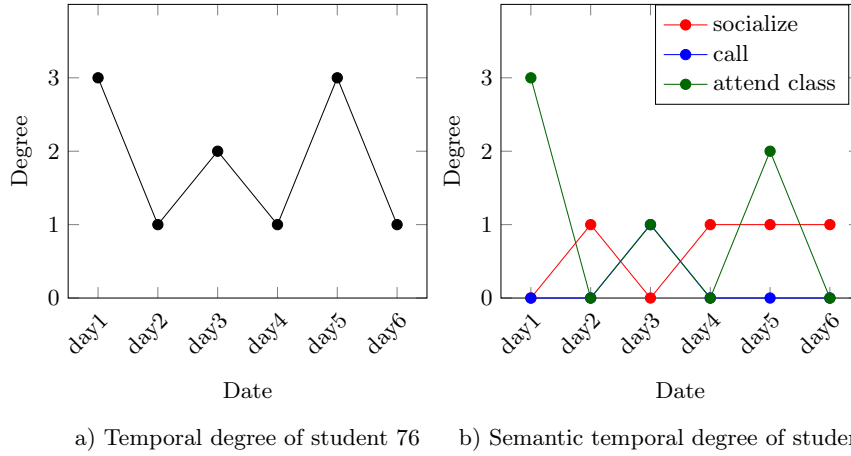


Fig. 3: Centrality degree of student 76 in Fig 2

Definition 3 (semantic temporal degree distribution). *The semantic temporal degree distribution of an entity class E' labelled $l_{E'}$ in a temporal graph TG is a set of five values, namely the minimum (\min), the three quartiles (q_x) and the maximum (\max) of the whole set of average semantic temporal degree values of entities with the semantics $l_{E'}$ during a time window w . The 1st quartile, denoted as q_1 , represents the average semantic temporal degree at which 25% of entities falls below and 75% falls above. The 2nd quartile or median, denoted as q_2 , represents the average semantic temporal degree at which 50% of entities falls below and 50% falls above. The 3rd quartile, denoted as q_3 , represents the average semantic temporal degree at which 75% of entities falls below and 25% falls above. The semantic temporal degree distribution of an entity class E' is then denoted as follows:*

$$deg_{distribution}(l_{E'}, w, l_{R'}) = \{\min(\cup_{e_i \in E'} deg_{average}(e_i, w, l_{R'})), q_1(\cup_{e_i \in E'} deg_{average}(e_i, w, l_{R'})), q_2(\cup_{e_i \in E'} deg_{average}(e_i, w, l_{R'})), q_3(\cup_{e_i \in E'} deg_{average}(e_i, w, l_{R'})), \max(\cup_{e_i \in E'} deg_{average}(e_i, w, l_{R'}))\}$$

It is graphically represented by a box plot (Fig 4). The bottom and top of the box represent the first and third quartiles (q_1 and q_3), respectively. The line inside the box represents the median (q_2). The minimum and maximum values are represented by the end points of the whiskers.

Example 4. The blue box plot in Fig 4 presents the semantic temporal degree distribution of students during one week for their class interactions ($deg_{distribution}(STUDENT, week1, ATTENDCLASS)$). It shows that 25% of students have in average less than 2 class interactions in a week. 75% of students have in average less than 7 class interactions in a week. Half of students has in average more than 3 class interactions in a week and the other half has in average less than 3 class interactions in a week. We position the average semantic

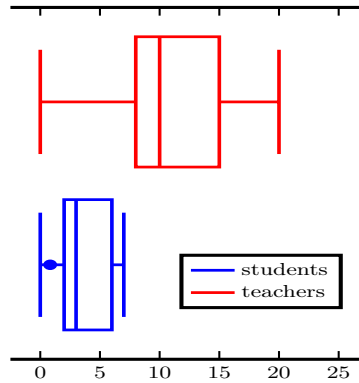


Fig. 4: Semantic Temporal Degree Distribution

temporal degree of the student 76 (calculated in the Example 3) as a blue dot on the blue box plot. We observe that the average semantic temporal degree of the student is very low compared to the rest of students. He may therefore have no influence on the disease transmission. The red box plot in Fig 4 presents the distribution of teachers during the same week as students. We notice that teachers have many more interactions than students and are therefore probably at greater risk of contracting or transmitting the disease.

4 Experimental Evaluation

In this section, we describe our experimental assessment which aims at evaluating the efficiency of our proposal. Details of our experimental assessment are available on the website <https://gitlab.com/2573869/degreecentrality>. A usability evaluation of the proposed centrality metrics in a real application is also available in the website.

4.1 Technical Environment

The experiments are conducted on a PowerEdge R630, 16 CPUs x Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40Ghz, 63.91 GB. One virtual machine installed on this hardware, that has 32 GB in terms of RAM and 100 GB in terms of disk size, is used for our experiments. The Neo4j graph database is installed in the virtual machine to store and query temporal graph datasets. The programming language used to implement our metrics is Python 3.7.

4.2 Datasets

The first real-world dataset we use is the Social experiment dataset⁴. The dataset traces the changes in the students' symptoms and in their interactions over time

⁴ <http://realitycommons.media.mit.edu/socialrevolution.html>

to study epidemiological contagion at an university dormitory. Students may have different types of interactions: physical interactions (such as *proximity*) and virtual interactions (such as *calls*) which may occur several times each day and other interaction types (such as *socialize*). The number of interactions of students can be analysed to identify potential "spreaders" of the epidemic. The second real-world dataset we use is a dataset from an E-commerce website⁵. The dataset presents customers' actions on items (view, add to cart and buy) which are made frequently within an hour. It also traces the changes in items' characteristics which are more rare over time. The number of actions done on items can be analysed to identify the popular items that can be recommended to customers.

4.3 Experimental Methodology and Results

Methodology. First, we transform each dataset into the temporal graph model presented in Section 3.1. The Social Experiment dataset contains 33 934 vertices and 2 168 270 edges. The E-commerce dataset is composed of 4 315 375 vertices and 4 447 430 edges. All transformation details are available in [1]. Then, we evaluate the computation time of the basic application of centrality metrics: ranking entities of a dataset for a chosen time interval. In our case, we choose to compute the average semantic temporal degree metric for each entity of each dataset. For this metric, we choose to put the parameters that make the computation the longest: we integrate all relationship labels and the whole time span of each entity. We define 5 scale factors by varying the number of edges involved in the metric calculation. From 1 to 5, the scale factor is 20%, 40%, 60%, 80%, 100% of the number of edges of each dataset respectively. We run the metric calculation 10 times for each scale factor and make the average elapsed computation time over the 10 executions. To sum up, our experiments account for a total of 100 executions: 2 datasets \times 5 scale factors \times 10 executions.

Results. Fig 5 shows the computation time to rank the number of interactions of students in the Social Experiment dataset (in blue) and the computation time to rank the number of actions on items in the E-commerce dataset (in red) based on the average semantic temporal degree metric. We observe that the computation time of the analysis increases linearly with the number of edges for both datasets.

5 Conclusion and Future Work

In this paper, we proposed three degree-based metrics to understand the centrality of entities in a dynamic context. From the conceptual point of view, we complete the classic degree metrics by adding both temporal and semantic aspects. An instantiation of the proposed metrics is done through a temporal

⁵ https://www.kaggle.com/retailrocket/ecommerce-dataset?select=item_properties_part2.csv

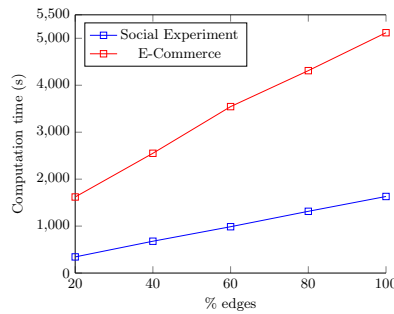


Fig. 5: Computation time of ranking the average semantic temporal degree values

graph model. To validate their efficiency, we applied the proposed metrics on two real datasets. We observed that the computation time of the metrics increases linearly with the edge number. In our future work, we plan to make new applications of our metrics by using new datasets and to extend other graph metrics such as betweenness centrality or local clustering coefficient.

References

1. Andriamampianina, L., Ravat, F., Song, J., Vallès-Parlangeau, N.: Graph data temporal evolutions: From conceptual modelling to implementation. *Data & Knowledge Engineering* **139**, 102017 (May 2022). <https://doi.org/10.1016/j.datak.2022.102017>
2. Andriamampianina, L., Ravat, F., Song, J., Vallès-Parlangeau, N.: Querying Temporal Property Graphs. In: *Advanced Information Systems Engineering: 34th International Conference, CAiSE 2022, Leuven, Belgium, June 6–10, 2022, Proceedings*. pp. 355–370. Springer-Verlag, Berlin, Heidelberg (Jun 2022). https://doi.org/10.1007/978-3-031-07472-1_21
3. Ghanem, M., Magnien, C., Tarissan, F.: Centrality Metrics in Dynamic Networks: A Comparison Study. *IEEE Transactions on Network Science and Engineering* **6**(4), 940–951 (Oct 2019). <https://doi.org/10.1109/TNSE.2018.2880344>
4. Ishfaq, U., Khan, H.U., Iqbal, S.: Identifying the influential nodes in complex social networks using centrality-based approach. *Journal of King Saud University - Computer and Information Sciences* **34**(10, Part B), 9376–9392 (Nov 2022). <https://doi.org/10.1016/j.jksuci.2022.09.016>
5. Meng, Y., Qi, Q., Liu, J., Zhou, W.: Dynamic Evolution Analysis of Complex Topology and Node Importance in Shenzhen Metro Network from 2004 to 2021. *Sustainability* **14**(12), 7234 (Jan 2022). <https://doi.org/10.3390/su14127234>
6. Rost, C., Gomez, K., Christen, P., Rahm, E.: Evolution of Degree Metrics in Large Temporal Graphs (2023). <https://doi.org/10.18420/BTW2023-23>
7. Uddin, S., Piraveenan, M., Chung, K.S.K., Hossain, L.: Topological Analysis of Longitudinal Networks. In: *2013 46th Hawaii International Conference on System Sciences*. pp. 3931–3940. IEEE, Wailea, HI, USA (Jan 2013). <https://doi.org/10.1109/HICSS.2013.556>
8. Wan, Z., Mahajan, Y., Kang, B.W., Moore, T.J., Cho, J.H.: A Survey on Centrality Metrics and Their Network Resilience Analysis. *IEEE Access* **9**, 104773–104819 (2021). <https://doi.org/10.1109/ACCESS.2021.3094196>