# TempoGRAPHer: A Tool for Aggregating and Exploring Evolving Graphs

Evangelia Tsoukanara
University of Macedonia
Thessaloniki, Greece
etsoukanara@uom.edu.gr

Georgia Koloniari
University of Macedonia
Thessaloniki, Greece
gkoloniari@uom.edu.gr

Evaggelia Pitoura
University of Ioannina
Ioannina, Greece
pitoura@cse.uoi.gr

## ABSTRACT

Graphs offer a generic abstraction for modeling entities as nodes and their interactions and relationships as edges. Since most graphs evolve over time, it is important to study their evolution. To this end, we propose demonstrating TempoGRAPHer, a tool that provides an overview of the evolution of an attributed graph offering aggregation at both the time and the attribute dimensions. The tool also supports a novel exploration strategy that helps in identifying time intervals of significant growth, shrinkage, or stability. Finally, we describe a scenario that showcases the usefulness of the TempoGRAPHer tool in understanding the evolution of contacts between primary school students.

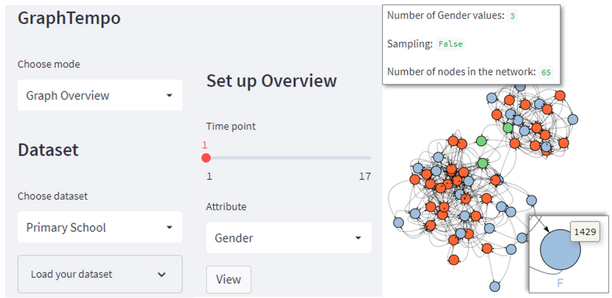



**Figure 1: Graph overview example.**

## 1 INTRODUCTION

Graphs are ubiquitous as they provide a well-described schema for several real-world problems and applications. For example, graphs can successfully capture collaborations in a co-authorship network or interactions between people in a social network. Since real-world graphs change with time, incorporating time is vital. Furthermore, real networks are often enriched with attributes that represent characteristics of a node or an edge. By aggregating a graph at different time periods and attributes, we may capture interesting patterns which can be associated with external factors. Conversely, being aware of external factors that could be associated with changes in the graph, we may focus our study on specific time periods and certain attributes. As a motivating scenario, consider a contact network, where nodes represent students and edges reflect the interactions between them during school time. Reporting an increase on the interactions between students having a specific attribute, e.g. gender, or, age, can be considered as a parameter affecting the spread of an infectious disease in the community. Taking advantage of the useful insights derived from graph aggregation and exploration, appropriate strategies can be planned to reduce the risk of disease spread.

In this paper, we introduce the TempoGRAPHer tool for exploring the evolution of attributed graphs through time. The tool builds on the GraphTempo model for graph aggregation [3]. TempoGRAPHer helps capturing the evolution of an attributed graph by aggregating on both time and attribute dimensions. The tool provides a set of temporal operators, which combined with intervals of interest, focus on different aspects of the graph and output a new temporal graph. After selecting the preferred attribute(s) and the type of aggregation, the tool returns the aggregate graph offering a high level view of the original network. Finally, TempoGRAPHer provides a novel strategy for discovering important time intervals in the evolution of the graph, such as notable shrinkage, growth, or stability. Such events are studied at different aggregation levels.



## 2 TEMPOGRAPHER TOOL

In this section, we first provide an overview of the underlying model based on the GraphTempo aggregation model [3] and then the functionality of the TempoGRAPHer tool.

### 2.1 Fundamentals

A temporal attributed graph in a set of time intervals $\mathcal{T}$ is a graph $G(V, E, \tau u, \tau e, A)$ where $V$ is the set of nodes and $E$ is the set of edges $(u, v)$ with $u, v \in V$. Each node in $V$ is associated with a timestamp $\tau u : V \longrightarrow \mathcal{T}$ where $\tau u(u)$ is the set of intervals during which $u$ exists. Similarly each edge $e$ in $E$ is associated with a timestamp $\tau e : E \longrightarrow \mathcal{T}$ where $\tau e(e)$ is the set of intervals during which $e$ exists. $A = \{A_1, A_2 \dots A_n\}$ is a set of $n$ node attributes whose values may change over time, that is, for each $u \in V$ and $t \in \tau u(u)$, there is a $n$-dimensional tuple, $A(u, t) = \{A_1(u, t), A_2(u, t) \dots A_n(u, t)\}$, where $A_i(u, t)$ denotes the value of $u$ at time $t \in \tau u(u)$ on the $i$-th attribute.

**Temporal Operators.** We define a set of temporal operators, that, given a temporal attributed graph and intervals, produce a new temporal graph.

Given $G$ and $\mathcal{T}_P \subseteq \mathcal{T}$, *project* generates a temporal attributed graph $G_P$ in $\mathcal{T}_P$, where $G_P$ includes all nodes and edges whose timestamps are in $\mathcal{T}_P$. For a set of intervals $\mathcal{T}_1$ and $\mathcal{T}_2$ *union* generates a temporal attributed graph in $(\mathcal{T}_1, \mathcal{T}_2)$, where the nodes and edges of the graph have timestamps that include either of $\mathcal{T}_1$, $\mathcal{T}_2$, while *intersection* produces a temporal attributed graph whose nodes and edges appear in both intervals. *Difference* outputs a temporal attributed graph including the nodes and edges that appear in $\mathcal{T}_1$ but not in $\mathcal{T}_2$. Last, we introduce the *evolution* graph that captures the changes occurring in a graph through time. Given $G$ and $\mathcal{T}_1$, $\mathcal{T}_2$, where $\mathcal{T}_1$ precedes $\mathcal{T}_2$, the evolution graph is defined as the combination of three graphs, the intersection graph that captures the part of the graph that remains stable, and the difference graphs $G_1 - G_2$ and $G_2 - G_1$ that represent deleted and new nodes and edges respectively.

**Graph Aggregation.** Graph aggregation allows us to study a graph at a global level and detect patterns that are not identified when observing the graph at the individual node level. Combining graph aggregation with temporal operators may reveal interesting patterns of evolution.

Graph aggregation is the operation of grouping nodes together based on one or more of their attribute values, while respecting the network structure. More precisely, given a graph $G$ in $\mathcal{T}$ and a subset of $m$ attributes, the aggregate graph is a weighted graph $G'$ where there is a node $u'$ in $G'$ for each value combination of the $m$ attributes and the weight of $u'$ is equal to the number of nodes $u$ in $G$ that have the specific attribute value combination. There is an edge between two nodes $u'$ and $v'$ in the aggregate graph if there is an edge between the corresponding nodes $u$ and $v$ in the original graph with weight the number of such edges. We distinguish between two types of aggregation: *distinct* aggregation where the weight of an aggregated node in $G'$ counts the distinct nodes in $G$ that have the corresponding attribute value combination and *non-distinct* aggregation where each appearance of an attribute value combination is counted in the weight each time it appears either on the same or on different nodes.

**Graph Exploration.** We are interested in identifying important intervals in the evolution of a graph. Specifically, given a threshold $k$, we look for pairs of intervals $(\mathcal{T}_i, \mathcal{T}_j)$ between which at least $k$ new elements are added (*growth* event), at least $k$ elements are deleted (*shrinkage* event) or at least $k$ elements remain stable (*stability* event). Elements are nodes or edges with specific attribute values. For nodes, given a set of attribute values, we look for nodes having these values, while for edges, given a pair of sets of attribute values, we look for edges connecting pairs of nodes with the corresponding values.

We consider successive pairs of intervals and gradually increase their length. In particular, starting from pairs of the shortest intervals, we keep one of the two intervals, say $\mathcal{T}_j$, fixed as a *reference point*, and gradually extend $\mathcal{T}_i$. We are interested in discovering either *maximal* or *minimal* pairs of intervals. An interval pair $(\mathcal{T}_i, \mathcal{T}_j)$ is maximal, if for a given $k$ and $\mathcal{T}_j$, there is no longer interval $\mathcal{T}_{i'}$ that contains $\mathcal{T}_i$ for which at least $k$ events occurred between $\mathcal{T}_{i'}$ and $\mathcal{T}_j$. Similarly, an interval pair is minimal if for given $k$ and $\mathcal{T}_j$, there is no shorter interval $\mathcal{T}_{i'}$ contained in $\mathcal{T}_i$ for which at least $k$ events occurred between $\mathcal{T}_{i'}$ and $\mathcal{T}_j$. For growth and shrinkage events, we look for minimal interval pairs between which the steepest changes have occurred, while for stability events, maximal interval pairs are located to discover long stability periods.

## 2.2 Tool Overview

We develop TempoGRAPHer, a tool for visualizing, aggregating, and exploring evolving attributed graphs. TempoGRAPHer has three main functionalities: (1) overview of the original graph at specific time points, (2) aggregation of temporal graphs on one or more attributes and at various time granularities, and (3) exploration of the graph for identifying intervals of significance.

First, users load their own graph or choose among existing graphs. Currently, there are three graphs, *DBLP*, a collaboration network, *MovieLens*, a co-rating dataset, both described in [3], and *Primary School* [1, 4]. In describing the tool we will use the *Primary School* dataset that is a contact network consisting of 232 students and 10 teachers of a primary school in Lyon, France. *Primary School* describes the face-to-face proximity of students and teachers. Each edge denotes a 20-seconds contact, and each

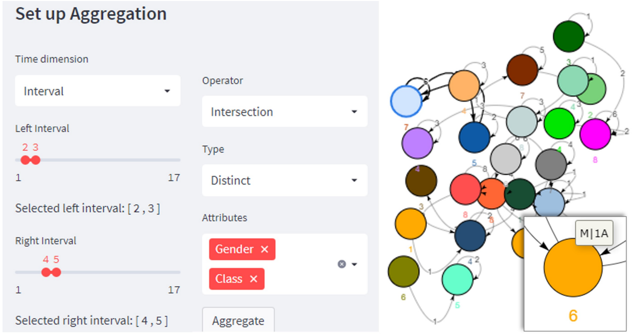



**Figure 2: Graph aggregation example.**

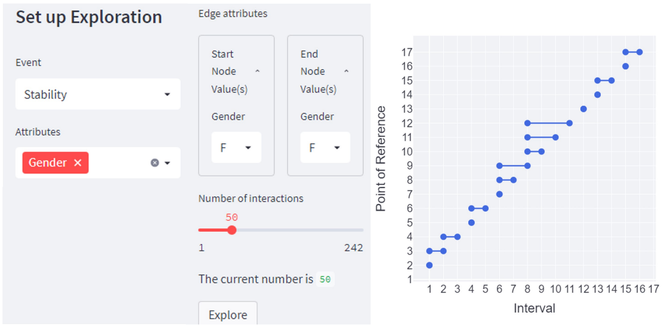



**Figure 3: Graph exploration example.**

node in the network has two static attributes, gender and class. The school has 5 grades, 1 to 5, with 2 classes each (i.e., 1A, 1B, 2A, 2B, etc) plus teachers, and 3 values for gender attribute, female (F), male (M), and unspecified (U). Our dataset covers a period of 17 hours.

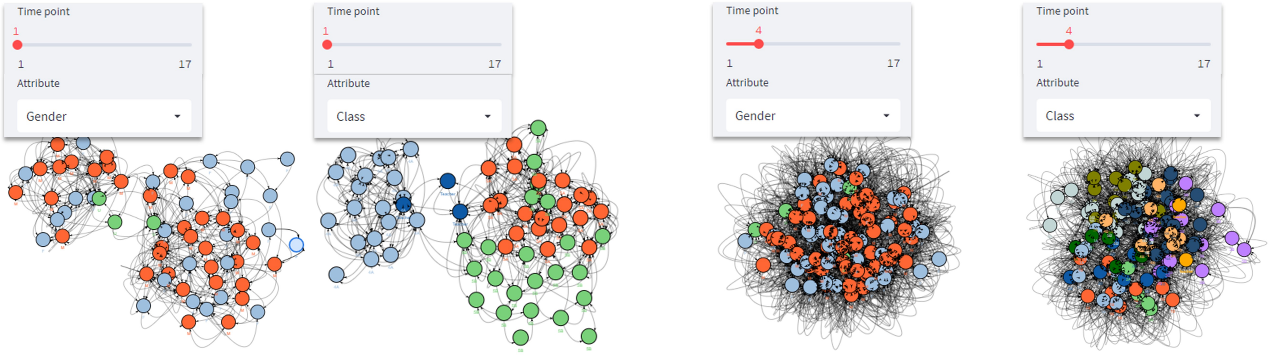
TempoGRAPHer is implemented in Python, the code is publicly available[1], and the tool can be accessed online[2].

**Graph Overview.** TempoGRAPHer provides an overview of the original attributed graph by displaying the graph at specific time points. The user selects the time point of interest and a single attribute, and in the shown graph, each node is colored by this attribute value. By hovering the mouse over a node, its id is shown. If the original graph is too large to illustrate, graph sampling is applied based on the Snowball sampler [2]. Figure 1 depicts the graph overview for the 1st time point and the gender attribute.

**Graph Aggregation.** TempoGRAPHer facilitates the aggregation of the original graph on both the time and the attribute dimensions. First the user selects the time period of interest and the preferred temporal operator (i.e., project, union, intersection, difference, or evolution), for creating the graph that corresponds to this period. Then, the user specifies the type of aggregation, distinct/non-distinct, and the set of attributes. For example, Fig. 2 shows the distinct aggregation of the gender and class attributes on the intersection graph for intervals [2,3] and [4,5]. Each attribute value combination is colored using a different color and this combination is shown by hovering over the node.

**Graph Exploration.** The third functionality of TempoGRAPHer is exploration, which offers a visual strategy for discovering parts of the graph where a significant event has occurred. The user

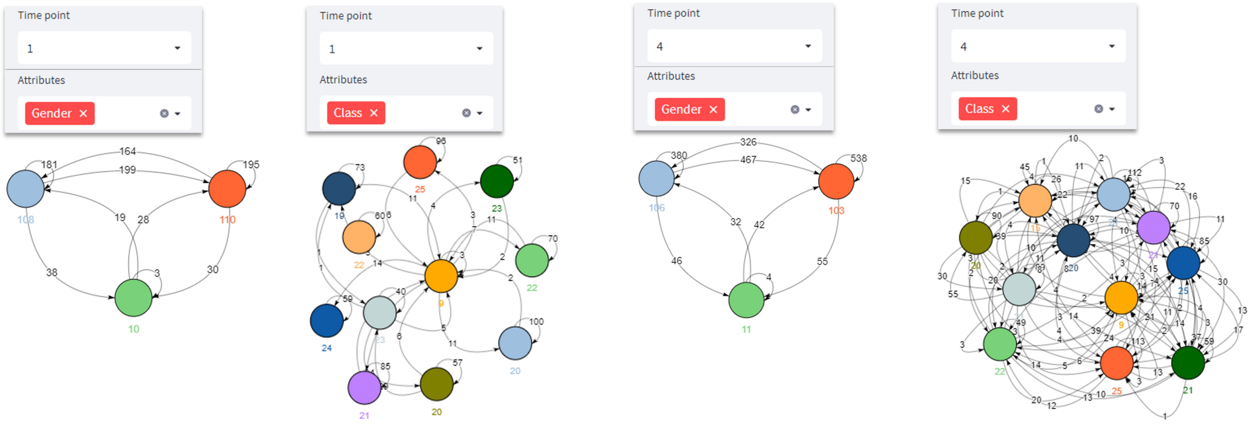
[1]https://github.com/etsoukanara/graphtempo-demo
[2]https://etsoukanara-graphtempo-demo-main-ul7qp1.streamlit.app




(a) 1st time point

(b) 4th time point

**Figure 4: Graph overview on gender and class attributes.**




(a) 1st time point

(b) 4th time point

**Figure 5: Aggregation on gender and class attributes.**


(a) Girls interactions

(b) Boys interactions

**Figure 6: Exploration results for stability for at least 100 interactions on gender.**

specifies the event of interest, e.g., growth, shrinkage, stability, selects the preferred attribute(s) and values, and provides the threshold $k$ of elements. The output is a plot illustrating for each reference point all maximal or minimal interval pairs depending on the type of the event. Figure 3 depicts the maximal intervals where at least $k = 50$ stable interactions have persisted between girls (F-F edges) in our dataset. The highest stability is observed for reference point 12 with the maximal interval pair being ([12, 12], [8, 11]), denoted as (12, [8, 11]) in the following.

## 3 DEMONSTRATION

Our demonstration shows how TempoGRAPHer helps users reveal underlying information and discover interesting aspects in the *Primary School* graph.


**(a) 1A's students interactions** **(b) 5A's students interactions**

**Figure 7: Exploration results for stability for at least 20 interactions on class.**

Suppose that we want to design a policy for containing disease spread. To this end, we want to figure out which groups of students are more active, or have less stable interactions and also when disease spread is more likely to happen.

We first run graph overview to visualize the original graph on various time points and get an idea of how the graph changes through time. Figure 4a on the left depicts the 1st time point and the gender attribute, where girls are represented with light blue, boys with orange, and unspecified with green. We observe at least two well-separated clusters in the graph showing students interactions, but clustering does not seem to depend on gender. By selecting the class attribute, we notice that there are 4 values at this time point, 4A, 5A, and 5B classes and teachers, as shown in Fig. 4a on the right. We notice that actually there are 3 clusters formed based on the class attribute, though two of them are not well-separated. We continue with the next time points and notice a change at the 4th time point shown in Fig. 4b. Compared to the 1st time point, here, we notice a more complex view, with students interacting with each other regardless of gender (on the left) or class (on the right). Thus, we can assume that the 1st time point corresponds to interactions during lessons held at the 1st hour of school, while the 4th time point shows break time in which classes are not separated. Viewing the graphs at specific time points helps us recognize lessons and breaks, but provides limited information about the volume or the duration of student interactions.

To further analyze our dataset, we proceed with distinct aggregation to quantify student interactions at lessons and breaks depending on different attribute values. On the left of Fig. 5a and Fig. 5b, we show the aggregate graphs on gender at the 1st and 4th time point respectively. Comparing the two graphs, we see a big increase in the number of interactions during breaks as we expected from our graph overview findings. While we see no clear dependence of the interactions on the gender of the students, we do observe that boys are much more active during breaks compared to girls. Regarding the class attribute, Fig. 5a on the right, refers to the 1st time point, where we notice that students mainly interact with students of the same class. On the right of Fig. 5b which corresponds to the 4th time point, we observe that in most cases students of a class have more interactions with students of other classes compared to the ones that belong to the same class. For example, students from class 5A depicted with green-yellow have 90 interactions with students of the same class and 185 interactions in total with students of other classes.

Finally, we proceed to analyze the degree of stability of the interactions of students. First, we compare the exploration results about the contacts between girls and between boys. Figure 6 depicts the exploration results for 100 stable interactions between girls (Fig. 6a) and boys (Fig. 6b). The highest stability for girls interactions is achieved in the (11, [9, 10]) and (12, [10, 11]) interval pairs, while boys have 100 stable interactions in the (6, [4, 5]), (11, [9, 10]), (12, [10, 11]), (15, [13, 14]), and (17, [15, 16]), showing that connections between boys are more stable in comparison to girls. Regarding the class attribute, first, we study the stable connections between students of junior class (1A), where we observe that there are at least 20 stable interactions in (4, [1, 3]) and (11, [8, 10]) (Fig. 7a). By selecting a senior class (5A), as shown in Fig. 7b, we see that students of 5A have at least 20 stable connections with students of their class in (11, [6, 10]) and (12, [7, 11]), showing higher stability compared to the junior class.

Summing up, we derive that boys show greater stability on their relationships at school compared to girls, though, they are much more active than girls during breaks. Concerning the class attribute, we observe that older students have more stable contacts compared to younger ones. Also, though most interactions occur between students of the same class during lessons, we notice that students of a class mostly interact with students of different classes at breaks.

## ACKNOWLEDGMENTS

Research work supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "1st Call for H.F.R.I. Research Projects to Support Faculty Members & Researchers and Procure High-Value Research Equipment" (Project Number: HFRI-FM17-1873, GraphTempo).

## REFERENCES

[1] Stehlé Juliette, Voirin Nicolas, Barrat Alain, Cattuto Ciro, Isella Lorenzo, Pinton Jean-François, Quaggiotto Marco, Van den Broeck Wouter, Régis Corinne, Lina Bruno, and Vanhems Philippe. 2011. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLOS ONE* 6, 8 (2011), e23176.

[2] Benedek Rozemberczki, Oliver Kiss, and Rik Sarkar. 2020. Little Ball of Fur: A Python Library for Graph Sampling. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20).* 3133–3140.

[3] Evangelia Tsoukanara, Georgia Koloniari, and Evaggelia Pitoura. 2023. GraphTempo: An aggregation framework for evolving graphs. In *Proceedings of the 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28-31, 2023.* OpenProceedings.org, 221–233.

[4] Gemmetto Valerio, Barrat Alain, and Cattuto Ciro. 2014. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC infectious diseases* 14, 1 (2014), 695.