A multilayer network-based framework for investigating the evolution and resilience of multimodal social networks

A Multilayer Network-Based Framework for Investigating the Evolution and Resilience of Multimodal Social Networks

Gianluca Bonifazi[1], Francesco Cauteruccio[1], Enrico Corradini[1], Edoardo Giannelli[1], Michele Marchetti[1], Domenico Ursino[1*], and Luca Virgili[1]

[1] DII, Polytechnic University of Marche
[*] Corresponding Author
g.bonifazi@univpm.it; f.cauteruccio@univpm.it; e.corradini@univpm.it; edoardo.giannelli@outlook.it; m.marchetti@pm.univpm.it; d.ursino@univpm.it; luca.virgili@univpm.it

Abstract

In recent literature, multilayer networks are increasingly being used to model and manage complex scenarios. One of them well suited to be modeled and managed using multilayer networks is represented by multimodal social networks, in which nodes and edges can be of different types. In fact, in this case, each layer of the multilayer network can be used to model one type of nodes in the multimodal social network while the edges of the latter can be represented by intra-layer edges of the former. In this paper, we want to demonstrate the feasibility of this idea by proposing a framework for analyzing multimodal social networks through the analysis of a corresponding multilayer network. In particular, we use our framework to address two challenging issues, namely the evolution of multilayer networks and their resilience to intra- and inter-layer perturbations. After introducing the model and describing the technical details of our framework, we present several experiments that demonstrate the effectiveness of our proposal.

Keywords: Multilayer Networks; Multimodal Social Networks; Temporal Evolution of Networks; Resilience of Networks; Network Modularity; Cross-Correlation Coefficient

1 Introduction

We are in an era of unprecedented interconnection and proliferation of data, which has led to the emergence and development of increasingly complex systems. The study of such systems has become a paramount concern in several scientific disciplines. In this scenario, Network Analysis has emerged as a powerful framework for investigating and understanding the intricate relationships between entities of various domains. Examples include social interactions [24, 13, 36], biological and medical processes [17, 43, 23] and technological applications [16, 10]. Traditional network analysis approaches have mainly focused on the study of connectivity patterns and structural properties of the networks. However, as networks become increasingly multifaceted (and, thus, multimodal [42]) and interconnected, the necessity arises for more sophisticated analytical frameworks that can capture the inherent complexity and heterogeneity of these systems.

In such a scenario, the multilayer network model has gained significant attention as a powerful tool for representing and analyzing intricate complex systems characterized by multiple types of interactions or layers [7, 23, 28]. Due to its ability to integrate multiple data sources and capture the heterogeneity of interdependencies, multilayer network has proven to be an extremely effective tool in unraveling hidden patterns that enable predictive modeling and enhance decision making processes. Traditional single-layer network models are unable to capture the heterogeneous nature of relationships that characterizes many complex systems, which are often featured by multiple types of interactions or multiple layers that operate simultaneously. On the other hand, the multilayer network model is able to do this and, thus, can facilitate the identification of emerging properties and the discovery of hidden patterns in these scenarios. It also supports the development of new strategies for optimizing system performance, designing efficient interventions, and managing risks [19].

Due to the interesting features mentioned above, the characteristics and applications of multilayer networks have been intensively investigated in recent years [1, 9, 39, 18, 12, 14, 29]. However, we believe that there are two aspects of these networks that have still been understudied. They concern the ability to model temporal evolution and resilience. Regarding the first aspect, we argue that modeling time-varying interactions between entities of a multilayer network allows for an effective analysis of temporal dynamics involving it and the system modeled through it. Many real-world systems, such as evolving social networks, fluctuating transportation networks and changing communication patterns, exhibit time-dependent behavior [25, 21, 37]. By capturing the temporal evolution of interdependencies, multilayer networks can facilitate understanding how these time-dependent systems grow, evolve, adapt and reconfigure over time. Regarding resilience, there are some studies that analyze this aspect within multilayer networks when they are used in specific domains, such as engineering systems [41, 40]. However, such studies are specific to that field and can hardly be extended to very different application contexts of multilayer networks, such as multimodal social networks. We believe that a general study of the resilience of multilayer networks that does not depend on the field in which they are applied is a challenging issue. Indeed, it could help shed light on the intricate interdependencies existing within the various phenomena they model in the different circumstances in which they are applied.

This paper aims to provide a contribution in this setting. Specifically, it proposes a framework for investigating the temporal evolution and resilience of entities in a multilayer network representing a multimodal social network. The proposed framework operates on a multilayer network consisting of three layers that, in our case, represent authors, comments and topics of a multimodal social network. However, at the outset we want to point out that, albeit our framework and its associated analyses have been defined for a multilayer network representing a social network with certain characteristics, they are general enough to be applied to any multilayer network modeling any application scenario, even those other than social networks.

Our framework includes three series of analyses. The first concerns the structure of multilayer networks; it allows us to study the temporal evolution of the structural properties of the multilayer network, for example degree-degree cross-correlation. The second focuses on network organization and modularity through the study of communities. Finally, the third allows the analysis of network resilience in presence of intra-layer and inter-layer perturbations. To perform these analyses, we introduce a set of metrics that are general enough to be used for any multilayer network.

Summarizing, the main contributions of this paper are as follows:

- We propose a framework and a set of measures to investigate the temporal evolution and resilience of a multilayer network.
- We propose a method for mapping a multimodal social network into a multilayer network.
- We apply the proposed multilayer network-based framework to investigate the temporal evolution and resilience of a multimodal social network. To the best of our knowledge, this is the first attempt to study the evolution and resilience of a multimodal social network through a multilayer network.

The outline of this paper is as follows: Section 2 presents related literature. Section 3 describes the proposed framework. Section 4 presents the experimental campaign we carried out on two different datasets to assess the effectiveness of the proposed framework. In Section 5, we present a discussion on the experiments we conducted, as well as on the similarities and differences between the characteristics of our approach and those of related ones. Finally, Section 6 presents some conclusions based on the findings of this research and outlines potential future developments of it.

2 Related Literature

Multilayer networks provide a more comprehensive representation of complex systems since they can capture multiple types of interactions and relationships. For this reason, they have received significant attention in many research areas in the past, including physics [7], biology and biomedicine [23], and social science [3]. Multilayer networks offer a rich, multifaceted perspective on complex systems by allowing multiple aspects to be analyzed simultaneously. As mentioned in the Introduction, this paper aims to provide a contribution in the research on multilayer networks. More specifically, it aims to propose a framework for investigating the temporal evolution and resilience of a multilayer network that models a multimodal social network, i.e., a social network whose nodes and edges can be of different types. Actually, as we shall see, our framework is general and can be applied to any multilayer network, regardless of the complex system the latter represents. Therefore, to better position our paper against related literature, we focus on two research strands, namely: *(i)* studies that evaluate evolution, resilience, or other similar aspects in multilayer networks, and *(ii)* studies that investigate multimodal social networks and some of their properties through multilayer network-based modeling. However, we want to make it clear that, to the best of our knowledge, our approach is the first that intersects these two strands, dealing with the time evolution and resilience of multimodal social networks modeled through multilayer networks.

Having made this premise, let us begin to consider the first strand mentioned above. In the past, the investigation of measures capable of evaluating the behavior of multilayer networks has mainly concerned application contexts [39, 15, 19, 18]. However, there are also some more theory-oriented works [7, 12]. The concept of modularity is often used to study various phenomena related to multilayer networks and to propose several measures regarding them [44, 30]. Before analyzing these works in more detail, it is important to note that the term “resilience” is used differently in different communities. Much of the research on this topic has been done in general engineering systems [26]. Regarding this context, we refer the interested reader to the comparative study presented in [41] and the review introduced in [40]. In [39], the authors employ multilayer networks to study relationships in human societies across different domains. They show that coupling between layers promotes prosocial behavior and that small changes to interactions in one domain can catalyze prosociality in a different domain. This study draws insights into the ability of multilayer network analysis to study complex social relationships in human societies. In [15], the authors analyze the Chinese Airline Network using the k-core decomposition method and representing it as a multilayer network with three layers, namely Core, Bridge and Periphery. They show that the network is more robust when low-degree nodes or high-flight flow links are removed. In [19], the authors propose a multilayer satellite communication network design that considers various application requirements and balances limited resources. They develop a feature-rich network analysis strategy and multi-objective optimization models to address design variables. In [18], the authors analyze the temporal dynamics of social interactions and collective behaviors in the social spider Stegodyphus dumicola. They use multilayer network analysis to quantify changes in social networks over time and determine the corresponding consequences on individual and group changes. They find that social interactions change over time at a constant rate and that social network stability is related to individual boldness but not to group collective behavior. In [7], the authors review recent advances in network theory related to the analysis of multilayer networks considering temporal and contextual properties of interactions. Their study emphasizes the importance of considering temporal and contextual properties when analyzing complex systems. Within this study, they make a solid contribution to the concepts of resilience and percolation. In [12], the authors propose a framework for characterizing layer centrality in multilayer networks. Such a centrality can be used to describe complex systems consisting of units connected by different types of relationships. They also introduce two measures for layer centrality based on network connectivity and the assignment of centrality values to layers. They validate these measures using a real-world dataset of air transportation multilayer networks and find that the measures are capable of extracting new and useful information from the considered dataset. In [44], the authors propose an approach for community detection in multilayer networks whose entities have various types of interactions that may change over time. This approach uses a multilayer modularity function and a spectral method, called mSpec, to optimize it. The authors validate their approach on empirical multilayer networks and show that it is reliable. In [30], the authors present a general framework that uses network quality functions to analyze the structure of communities in multislice networks. The latter combine individual networks through links connecting nodes across different network slices. They allow the study of community structure in a general setting that includes networks with multiple types of links, time evolution and multiple scales.

The second research strand we analyze concerns the investigation of social networks through the use of multilayer networks. As for this strand, the authors of [8] propose a multilayer network-based approach to analyze discussions on social networks, focusing on sensitive topics like healthcare. They apply the proposed approach to a Twitter dataset related to COVID-19 vaccines and identify relevant hashtags for each line of thought. Their results show that anti-vaxxers have denser and more cohesive ego networks, leading to more interactions among them. They also show that their approach is effective in detecting influential users. In [27], the authors employ social network analysis and topic modeling to represent user interactions on Twitter as a directed, weighted and multilayer network. Then they exploit this network to identify influential users and highly connected groups. They show that a topically focused social network representing conversations returns more robust results regarding influential users, especially when it is associated with tweets related to a wide variety of discussions. In [5], the authors propose an unsupervised, parameter-free, network feature-based approach called ADOMS (Anomaly Detection on Multilayer Social Networks) to search for anomalous users in multilayer social networks based on clique/near-clique and star/near-star anomalies. ADOMS ranks users according to the similarity degree of their neighborhoods in different layers to stars- and cliques-like structures. In [33], the authors propose an approach for finding influential users on social media platforms by analyzing the content of their posts by means of a three-layer network representing users, items and keywords. The approach exploits multilinear algebra to detect users most active in stating their point of view on dominant items tagged with dominant keywords. In [34], the authors propose an approach for classifying news articles as either disinformation or mainstream news based only on their dissemination mechanisms on Twitter. In the multilayer network they use to support their analysis, each layer describes a different type of interactions. This allows the approach of [34] to quantify the impact of each layer in the classification activity.

3 Description of the proposed framework

In this section, we present our framework for studying the temporal evolution and resilience of a multimodal social network through its modeling by means of a multilayer network. Specifically, in Subsection 3.1 we illustrate the multilayer network-based model used by our framework. In Subsection 3.2, we describe some measures and analysis approaches used by it to achieve its goals.

In order to explain our framework and the associated model, we assume that the multimodal social network to be represented is a content-based network, i.e., a network for the dissemination of content and comments. We also assume that this network has three types of nodes representing authors, topics and comments, and two types of edges connecting authors to their comments and comments to the topics to which they refer. However, we want to reiterate that our framework and the associated model are generalizable to any multimodal social network.

3.1 The multilayer network-based model

Our multilayer network-based model can be represented as a quadruple:

$$\mathcal{M} = \langle A, C, T, E \rangle \tag{3.1}$$

Here, A, C and T represent three layers associated with authors, comments and topics, respectively. E denotes the set of inter-layer edges, that is, edges connecting nodes belonging to different layers. More specifically:

- $A = \langle N^A, E^A \rangle$ is the layer representing authors. There is a node $a_i \in N^A$ for each author who wrote a comment in the underlying multimodal social network. There exists an edge

$(a_i, a_j, w_{ij}^A, \tau_{ij}^A) \in E^A$ if there was at least one interaction between a_i and a_j, i.e., if a_i replied to a comment of a_j, or vice versa. The weight w_{ij}^A represents the number of interactions between a_i and a_j, while τ_{ij}^A indicates the timestamp when the last interaction occurred.

- $C = \langle N^C, E^C \rangle$ is the layer representing comments. There is a node $c_i \in N^C$ for each comment posted in the underlying multimodal social network. There exists an edge $(c_i, c_j, w_{ij}^C, \tau_{ij}^C) \in E^C$ if c_i refers to c_j, or vice versa. The weight w_{ij}^C represents the similarity degree between the content expressed in c_i and c_j, while τ_{ij}^C indicates the timestamp when the latest of the comments between c_i and c_j was posted.

- $T = \langle N^T, E^T \rangle$ is the layer representing topics. There is a node $t_i \in N^T$ for each topic discussed in the underlying multimodal social network. There exists an edge $(t_i, t_j, w_{ij}^T, \tau_{ij}^T) \in E^T$ if t_i and t_j are related to each other. In particular, t_i could be a topic in the same category as t_j (for example, both of them could refer to food), or it could be derived from t_j (for example, t_i could be "Inflation" and t_j could be "Economy"). The weight w_{ij}^T represents the strength of the relationship between t_i and t_j, while τ_{ij}^T indicates the timestamp when the latest of the topics between t_i and t_j appeared in the corresponding multimodal social network for the first time.

- E is the set of inter-layer edges, that is, edges connecting nodes belonging to different layers. In our model we can have edges between the nodes of T and the ones of C and between the nodes of C and the ones of A. In particular, there exists an edge $(t_i, c_j, \tau_{ij}^{TC}) \in E$ if the topic t_i is discussed in the comment c_j; τ_{ij}^{TC} is the timestamp when c_j was posted. There is an edge $(c_i, a_j, \tau_{ij}^{CA}) \in E$ if the comment c_i was posted by the author a_j; τ_{ij}^{CA} is the timestamp when c_i was posted.

In Figure 1, we provide a graphical representation of our model.

3.1.1 Definition of the temporal instance of a multilayer network

In order to study the temporal evolution and resilience of a multimodal social network, we need a mechanism to represent time within our model. In particular, we need to examine networks over different time intervals to understand how they change and react to perturbations. To achieve this, we introduce the concept of temporal instance of a multilayer network, which captures its structure and interactions in a specific time interval. Through it, we can study how the structure of the network and the interactions it models evolve over time both in absence and in presence of perturbations.

A temporal instance of our multilayer network-based model $\mathcal{M}$ can be represented as:

$$\mathcal{M}[\tau_1, \tau_2] = \langle A[\tau_1, \tau_2], C[\tau_1, \tau_2], T[\tau_1, \tau_2], E[\tau_1, \tau_2] \rangle \tag{3.2}$$

Here, τ_1 and τ_2 are the start and end timestamp of the time interval of interest, respectively. The layers $A[\tau_1, \tau_2]$, $C[\tau_1, \tau_2]$ and $T[\tau_1, \tau_2]$ and the set of edges $E[\tau_1, \tau_2]$ have the same structure as the layers A, C and T and the set of edges E of $\mathcal{M}$ defined above, but contain only the nodes and edges present in the time interval $[\tau_1, \tau_2]$. Clearly, if we denote by 0 the instant of origin of the multimodal social network corresponding to our multilayer network-based model and by τ_l the last time instant to which our data refer, we have that $\mathcal{M} = \mathcal{M}[0, \tau_l]$.

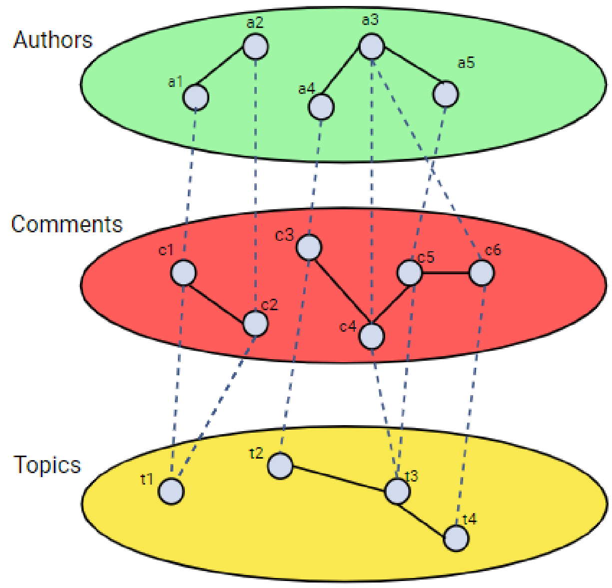

Figure 1: Representation of the multilayer network-based model adopted in our framework

For example, one instance of the model might represent the multimodal social network during a specific month or week. In this case, the layer A includes only the authors who posted comments during that period, the layer C includes only the comments posted during that period, and the layer T includes only the topics discussed during that period. Finally, the set E only includes the edges representing interactions between nodes in A and C, or between nodes in C and T, which occurred during that period.

3.2 Measures employed by our framework

Our framework supports the analysis of single layers as well as the one of the multilayer network as a whole. In addition, it supports both the static analysis of the multilayer network throughout the whole time interval of interest and its dynamic analysis taking into account its evolution over time. To ensure all these analyses, it introduces some measures. In this section, we focus on this topic; in particular, we examine the measures employed by our framework based on three aspects we want to address, namely structure (Subsection 3.2.1), communities (Subsection 3.2.2), and resilience (Subsection 3.2.3).

3.2.1 Measures regarding network structure

In order to study the structure of a multilayer network, it is necessary to perform both intra-layer and inter-layer analyses. Furthermore, to analyze the network evolution over time, we need to define a constant length for the time intervals to be associated with the various temporal instances of the multilayer network-based model. This length can be chosen based on the reference dataset, the application scenario and the type of analysis we want to conduct. We denote by $\mathcal{T} = [0, \tau_l]$ the total time period to which the data of the multimodal social network we want to investigate refers to. We must partition $\mathcal{T}$ into contiguous, non-overlapping sub-intervals $[\tau_1, \tau_2]$ such that $\tau_2 - \tau_1$ is equal to the chosen length. For example, $\mathcal{T}$ could be equal to 4 weeks; in this case, it could be divided into 4 sub-intervals each having a length of 1 week. In this case, the overall multilayer network $\mathcal{M}$ would give rise to four temporal instances, each associated with one week.

Let $\mathcal{M}[\tau_1, \tau_2]$ be a temporal instance of $\mathcal{M}$. Let L be one of the three layers of $\mathcal{M}[\tau_1, \tau_2]$ (i.e., $L \in \{A[\tau_1, \tau_2], C[\tau_1, \tau_2], T[\tau_1, \tau_2]\}$). We introduce the degree-degree correlation coefficient r_L between the nodes of the layer L. It is defined as:

$$r_L = \frac{\sum_{n_i \in N^L} \sum_{n_j \in (N^L \setminus \{n_i\})} [(\delta_i - \overline{\delta^L})(\delta_j - \overline{\delta^L})]}{\sqrt{\sum_{n_i \in N^L} (\delta_i - \overline{\delta^L})^2 \sum_{n_j \in (N^L \setminus \{n_i\})} (\delta_j - \overline{\delta^L})^2}} \tag{3.3}$$

Here: *(i)* N^L represents the set of nodes of the layer L; *(ii)* δ_i (resp., δ_j) denotes the degree of the node n_i (resp., n_j) taking only intra-layer edges into account; *(iii)* $\overline{\delta^L}$ indicates the average degree of the nodes of L taking only intra-layer edges into account. The degree-degree correlation coefficient r_L measures the assortativity of the nodes of L with respect to their degree. In other words, it is an indicator of the tendency of the nodes of L to be connected to other nodes of L with similar degree [31]. Its values range in the real interval $[-1, 1]$, where 1 indicates total assortativity, -1 represents total disassortavity and 0 denotes the absence of both assortativity and disassortativity.

We now introduce the degree-degree cross-correlation coefficient r_{L_1,L_2} between the nodes of two different layers L_1 and L_2. It can be defined as follows:

$$r_{L_1,L_2} = \frac{\sum_{n_i \in N^{L_1}} \sum_{n_j \in N^{L_2}} [(\delta_i - \overline{\delta^{L_1}})(\delta_j - \overline{\delta^{L_2}})]}{\sqrt{\sum_{n_i \in N^{L_1}} (\delta_i - \overline{\delta^{L_1}})^2 \sum_{n_j \in N^{L_2}} (\delta_j - \overline{\delta^{L_2}})^2}} \tag{3.4}$$

Here: *(i)* N^{L_1} (resp., N^{L_2}) represents the set of nodes of L_1 (resp., L_2); *(ii)* δ_i (resp., δ_j) denotes the degree of the node $n_i \in N^{L_1}$ (resp., $n_j \in N^{L_2}$); *(iii)* $\overline{\delta^{L_1}}$ (resp., $\overline{\delta^{L_2}}$) indicates the average degree of the nodes of L_1 (resp., L_2). The degree-degree cross-correlation coefficient r_{L_1,L_2} measures the correlation between the degrees of nodes of different layers. Similar to r_L, r_{L_1,L_2} measures the assortativity of the nodes of L_1 and L_2 with respect to their degree. Also r_{L_1,L_2} ranges in the real interval $[-1, 1]$, where -1 indicates the maximum disassortativity, 0 the lack of assortativity and disassortativity and 1 the maximum assortativity.

Another useful measure for investigating network structure is clustering coefficient. It measures the degree to which nodes in the same neighborhood are connected. Specifically, the clustering coefficient γ_L of a layer L of $\mathcal{M}[\tau_1, \tau_2]$ can be defined as:

$$\gamma_L = \frac{3\ N_{closed}}{N_{connected}} \tag{3.5}$$

Here: *(i)* N_{closed} indicates the number of closed triads in L; a triad is a set of three nodes; it is closed if its nodes are all connected to each other; *(ii)* $N_{connected}$ denotes the number of connected triads; a triad is connected if at least two of its nodes are connected.

Finally, a last measure that our framework adopts to evaluate the structure of a network is the average path length. It measures the average length of the shortest paths between pairs of nodes. Specifically, the average path length of a layer L of a multilayer network $\mathcal{M}[\tau_1, \tau_2]$ can be defined as:

$$\mathcal{L}_L = \frac{1}{|N^L| \ (|N^L| - 1)} \sum_{n_i \in N^L} \sum_{n_j \in (N^L \setminus \{n_i\})} \langle n_i, n_j \rangle \tag{3.6}$$

Here: *(i)* $|N^L|$ is the cardinality of N^L, and thus the number of its nodes; *(ii)* $\langle n_i, n_j \rangle$ is the length of the shortest path between the nodes n_i and n_j in the layer L.

3.2.2 Measures regarding communities

Community analysis is a widely used method for studying the organization and cohesiveness of complex networks. In the context of content-based social networks, communities can represent, for instance, groups of users who are interested in similar topics and engage in similar interactions. To perform community analysis on a multilayer network, we need to identify communities within each layer (intra-layer communities), as well as communities involving nodes from different layers (inter-layer communities).

Similar to what we did for network structure, given a multilayer network $\mathcal{M}$ and one of its temporal instances $\mathcal{M}[\tau_1, \tau_2]$, let L be one of its layers (i.e., $L \in \{A[\tau_1, \tau_2], C[\tau_1, \tau_2], T[\tau_1, \tau_2]\}$). To identify intra-layer communities, our framework employs the Louvain algorithm [6], which has been shown to be efficient and is widely used in the literature. This algorithm maximizes the modularity of the network (i.e., its tendency to be divided into cohesive and loosely coupled communities) by iteratively merging or splitting communities based on the improvement in modularity. Instead, to identify inter-layer communities, our framework adopts Infomap [35]. This is an algorithm for detecting inter-layer communities that operates by optimizing an objective function that balances the density of links within communities and the similarity between layers. Infomap can detect both overlapping and non-overlapping communities.

To measure the modularity of each layer, our framework adopts the classical modularity index [32]. According to it, the modularity index Q_L of L is defined as:

$$Q_L = \frac{1}{2\ |E^L|} \sum_{n_i \in N^L} \left[\sum_{n_j \in N^L} \left(w_{ij} - \frac{\sigma_i \ \sigma_j}{2\ |E^L|} \right) \Gamma(\mathcal{C}_i, \mathcal{C}_j) \right] \tag{3.7}$$

Here: *(i)* $|E^L|$ denotes the number of edges of L; *(ii)* N^L indicates the set of nodes of L; *(iii)* w_{ij} is the weight of the edge connecting n_i to n_j; *(iv)* σ_i (resp., σ_j) is the strength of the node n_i (resp., n_j) in L; it is defined as $\sigma_i = \sum_{n_j \in N^L} w_{ij}$ (resp., $\sigma_j = \sum_{n_i \in N^L} w_{ij}$); *(v)* $\mathcal{C}_i$ (resp., $\mathcal{C}_j$) indicates the community to which n_i (resp., n_j) belongs; *(v)* $\Gamma(\mathcal{C}_i, \mathcal{C}_j)$ is the Kronecker delta function, which is equal to 1 if $\mathcal{C}_i = \mathcal{C}_j$, while it is equal to 0 otherwise. Q_L ranges in the real interval $[0, 1]$; the higher Q_L, the stronger the subdivision of L into distinct communities.

In an analogous way, to quantify the modularity of the whole multilayer network $\mathcal{M}$, our framework employs a multilayer modularity index aiming at measuring the degree to which $\mathcal{M}$ is partitioned into distinct communities spanning multiple layers. If we denote by $\mathcal{LS}$ the set of layers of $\mathcal{M}$ (in our case $\mathcal{LS} = \{A, C, T\}$), the multilayer modularity index adopted in our framework is defined as follows:

$$Q_{\mathcal{M}} = \frac{1}{2\ \sum_{L \in \mathcal{LS}} |E^L|} \sum_{n_i \in N^{L_1}, L_1 \in \mathcal{LS}} \left[\sum_{n_j \in N^{L_2}, L_2 \in (\mathcal{LS} \setminus \{L_1\})} \left(w_{ij} - \rho\ \frac{\sigma_i\ \sigma_j}{2\ \sum_{L \in \mathcal{LS}} |E^L|} \right) \Gamma(\mathcal{C}_i, \mathcal{C}_j)\ \Gamma(L_i, L_j) \right] \tag{3.8}$$

Here:

- $|E^L|$, N^{L_1}, N^{L_2}, w_{ij}, $\mathcal{C}_i$ and $\mathcal{C}_j$ have been defined above.
- ρ is a parameter that takes into account the heterogeneity of the number of nodes in the various layers of the multilayer network. In fact, this number substantially affects the importance to be given to a community of a certain size. For example, a community of 10 nodes is very large and important in a layer that contains a total of 20 nodes, while it is negligible in a layer that contains a total of 1000 nodes. ρ ranges in the real interval $[0, 1]$; the greater ρ is, the more homogeneous the layers of the multilayer network are with each other in terms of their number of nodes.
- Since the inter-layer edges are unweighted, we assume that all of them have a weight equal to 1. Therefore, in this formula, $\sigma_i = \delta_i$ (resp., $\sigma_j = \delta_j$), where δ_i (resp., δ_j) represents the number of inter-layer edges incident on n_i (resp., n_j).

The first factor in Equation 3.8 takes intra-layer edges into account, while the second factor considers inter-layer edges.

$Q_{\mathcal{M}}$ ranges in the real interval $[-1, 1]$; values greater (resp., less) than 0 indicate that the communities into which $\mathcal{M}$ is divided are more (resp., less) cohesive than those that would be obtained by inserting the same number of intra-layer and inter-layer edges randomly.

We defined Equation 3.8 drawing inspiration from the equation on cross-modularity index introduced in [30]. Although the two formulations share some similarities, they also present several differences. This is due to the fact that they are used to measure the modularity of multilayer networks with different characteristics. In fact, the approach in [30] is designed for multislice networks, which are multilayer but do not have inter-layer edges, while our framework handles multilayer networks having both intra-layer and inter-layer edges.

By analyzing the communities within each layer and the whole multilayer network, we can identify groups of users who are interested in similar topics or engage in interactions with each other. In addition, by comparing communities across different time instances of the multilayer network, we can identify changes in its organization and modularity over time.

3.2.3 Measures regarding resilience

In order to study the resilience of a multilayer network, it is necessary to test what happens to the latter in presence of intra-layer and inter-layer perturbations. To quantitatively evaluate it, our framework adopts several metrics that we define in the next subsections.

Intra-layer perturbations Intra-layer perturbations refer to variations of nodes, edges and weights within a layer that can be caused by various factors, such as user behavior or changes in the topics being discussed. Intra-layer perturbations can be of six types, namely: *(i)* addition of new nodes and corresponding edges; *(ii)* removal of existing nodes and corresponding edges; *(iii)* addition of new edges to existing nodes; *(iv)* removal of existing edges; *(v)* increase of edge weights; *(vi)* decrease of edge weights. Intra-layer perturbations may be caused by changes in the interactions between users or in the topics they discuss. For example, a change in the behavior of a particular user or group of users may lead to the removal or addition of nodes and edges referring to them.

Let $\mathcal{M}$ be a multilayer network and let L be a layer of $\mathcal{M}$ (i.e., $L \in \mathcal{LS}$). Let $L[0,\tau_1]$ (resp., $L[0,\tau_2]$) be the occurrence of L in the temporal instance $\mathcal{M}[0,\tau_1]$ (resp., $\mathcal{M}[0,\tau_2]$) of $\mathcal{M}$. Let $\delta_M^L[0,\tau_1]$ (resp., $\delta_M^L[0,\tau_2]$) be the maximum degree of a node in $L[0,\tau_1]$ (resp., $L[0,\tau_2]$). Let δ_M be the maximum between $\delta_M^L[0,\tau_1]$ and $\delta_M^L[0,\tau_2]$. To study the resilience of $\mathcal{M}$ to the intra-layer perturbations occurring in the time interval $[\tau_1,\tau_2]$, we define the following metrics and include them in our framework:

- *Clustering Coefficient Variation* $\Delta\gamma_L[\tau_1,\tau_2]$: it measures the magnitude of the variation in the clustering coefficient γ_L of L caused by the intra-layer perturbations occurring in the time interval $[\tau_1,\tau_2]$. In other words, it measures the degree to which the intra-layer perturbations occurring in the time interval $[\tau_1,\tau_2]$ affected the structure of the network associated with L. It is defined as follows:

$$\Delta\gamma_L[\tau_1,\tau_2] = \frac{\gamma_L[0,\tau_2] - \gamma_L[0,\tau_1]}{\gamma_L[0,\tau_1]} \tag{3.9}$$

  Here, $\gamma_L[0,\tau_2]$ (resp., $\gamma_L[0,\tau_1]$) denotes the clustering coefficient of $L[0,\tau_2]$ (resp., $L[0,\tau_1]$), as defined in Equation 3.5.

- *Level-wise Degree Distribution Variation* $\Delta D_L[\tau_1,\tau_2]$: it measures the magnitude of the variation in the number of nodes $D_L(k)[0,\tau_1]$ of degree k in $L[0,\tau_1]$, $0 \leq k \leq \delta_M$, caused by the intra-layer perturbations occurring in the time interval $[\tau_1,\tau_2]$. In other words, this metric measures the degree to which the intra-layer perturbations occurring in the time interval $[\tau_1,\tau_2]$ affected the overall connectivity patterns within L. It is defined as follows:

$$\Delta D_L[\tau_1,\tau_2] = \frac{\sum_{k=0}^{\delta_M} |\Delta D_L(k)[\tau_1,\tau_2]|}{\delta_M} \tag{3.10}$$

  where:

$$\Delta D_L(k)[\tau_1,\tau_2] = \frac{D_L(k)[0,\tau_2] - D_L(k)[0,\tau_1]}{D_L(k)[0,\tau_1]} \tag{3.11}$$

The two metrics defined above allow us to analyze the resilience of the multilayer network to changes in interactions between users (resp., comments, topics) who discuss (resp., which are posted, which are discussed) in the underlying social network. For example, a significant variation $\Delta\gamma_L[\tau_1,\tau_2]$ of the clustering coefficient γ_L after some intra-layer perturbations occurring in the time interval $[\tau_1,\tau_2]$

may indicate that the behavior of the involved nodes is having a disproportionately large impact on the local network structure. On the other hand, a significant change in the level-wise degree distribution $\Delta D_L(k)[\tau_1, \tau_2]$ of L means that the user (resp., comment, topic) interactions are shifting from one or more users (resp., comments, topics) to others. Such a shift can lead to significant changes in the overall structure of the multilayer network.

Inter-layer perturbations Inter-layer perturbations are changes in the structure of connections linking nodes of different layers. This type of perturbations requires the addition or removal of nodes in at least two layers, as well as the addition or removal of inter-layer edges. For example, changes in the popularity of topics or in the interests of a user or group of users can influence the corresponding comments and the connectivity patterns between layers. Inter-layer perturbations can be of four types, namely: *(i)* addition of inter-layer edges, which also involves the addition of nodes in at least two different layers; *(ii)* removal of inter-layer edges, which also involves the removal of nodes in at least two different layers; *(iii)* addition of nodes, each having at least one inter-layer edge that connects it with an existing node or a new node in another layer; *(iv)* removal of nodes each having at least one inter-layer edge incident into it. By analyzing the resilience of a multilayer network to inter-layer perturbations, we can assess the network's ability to adapt to changes in the way topics are discussed by different groups of users.

Let $\mathcal{M}$ be a multilayer network and let $\mathcal{M}_1 = \mathcal{M}[0, \tau_1]$ and $\mathcal{M}_2 = \mathcal{M}[0, \tau_2]$ be two temporal instances of $\mathcal{M}$. Let L_1 and L_2 be two layers of $\mathcal{M}$ such that $L_1 \neq L_2$. Let $L_1[0, \tau_1]$ (resp., $L_2[0, \tau_1]$) be the occurrence of the layer L_1 (resp., L_2) corresponding to the temporal instance $\mathcal{M}_1$. Let $L_1[0, \tau_2]$ (resp., $L_2[0, \tau_2]$) be the occurrence of the layer L_1 (resp., L_2) corresponding to the temporal instance $\mathcal{M}_2$. Let $E[0, \tau_1]$ (resp., $E[0, \tau_2]$) be the occurrence of the set E of inter-layer edges corresponding to the temporal instance $\mathcal{M}_1$ (resp., $\mathcal{M}_2$). Let $\delta_M^E[0, \tau_1]$ (resp., $\delta_M^E[0, \tau_2]$) be the maximum degree of a node connected to at least one edge of $E[0, \tau_1]$ (resp., $E[0, \tau_2]$). Finally, let δ_M be the maximum between $\delta_M^E[0, \tau_1]$ and $\delta_M^E[0, \tau_2]$. To study the resilience of $\mathcal{M}$ to inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$, we define the following metrics and include them in our framework:

- *Modularity Variation* $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$: it measures the magnitude of the variation in the modularity index $Q_{\mathcal{M}}$ of the multilayer network caused by the inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$. In other words, it measures the degree to which the inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$ disrupted the community structure of the network. It is defined as follows:

$$\Delta Q_{\mathcal{M}}[\tau_1, \tau_2] = \frac{Q_{\mathcal{M}}[0, \tau_2] - Q_{\mathcal{M}}[0, \tau_1]}{Q_{\mathcal{M}}[0, \tau_1]} \tag{3.12}$$

  Here, $Q_{\mathcal{M}}[0, \tau_2]$ (resp., $Q_{\mathcal{M}}[0, \tau_1]$) denotes the multilayer modularity index relative to $\mathcal{M}[0, \tau_2]$ (resp., $\mathcal{M}[0, \tau_1]$) as defined in Equation 3.8.

- *Inter-layer Correlation Variation* $\Delta r_{L_1, L_2}[\tau_1, \tau_2]$: it measures the magnitude of the variation in the degree-degree cross-correlation coefficient r_{L_1, L_2} between the layers L_1 and L_2 of $\mathcal{M}$ caused by the inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$. In other words, it measures the degree to which the inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$ disrupted the correlation between the connectivity patterns of the layers L_1 and L_2. It is defined as follows:

$$\Delta r_{L_1,L_2}[\tau_1, \tau_2] = \frac{r_{L_1,L_2}[0, \tau_2] - r_{L_1,L_2}[0, \tau_1]}{r_{L_1,L_2}[0, \tau_1]} \tag{3.13}$$

Here, $r_{L_1,L_2}[0, \tau_2]$ (resp., $r_{L_1,L_2}[0, \tau_1]$) denotes the degree-degree cross-correlation coefficient between the layers L_1 and L_2 of $\mathcal{M}[0, \tau_2]$ (resp., $\mathcal{M}[0, \tau_1]$), as defined in Equation 3.4.

- *Multilayer Degree Distribution Variation* $\Delta D_E[\tau_1, \tau_2]$: it measures the magnitude of the variation in the number of nodes $D_E(k)[0, \tau_1]$ of degree k in E, $0 \leq k \leq \delta_M$, caused by the inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$. In other words, it measures the degree to which the inter-layer perturbations occurring in the time interval $[\tau_1, \tau_2]$ affected the overall connectivity patterns within E. It is defined as follows:

$$\Delta D_E[\tau_1, \tau_2] = \frac{\sum_{k=0}^{\delta_M} |\Delta D_E(k)[\tau_1, \tau_2]|}{\delta_M} \tag{3.14}$$

where:

$$\Delta D_E(k)[\tau_1, \tau_2] = \frac{D_E(k)[0, \tau_2] - D_E(k)[0, \tau_1]}{D_E(k)[0, \tau_1]} \tag{3.15}$$

Thanks to these metrics we can assess the resilience of the multilayer network $\mathcal{M}$ to changes in the way topics are discussed by different user groups. For example, if the modularity index $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ remains relatively stable, we can argue that the community structure of the network is robust to changes in the topics being discussed. On the other hand, if the degree-degree cross-correlation coefficient has a significant decrease (i.e., the value of $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ is negative), we can argue that the connectivity patterns of different layers are becoming less correlated, which can potentially lead to the formation of isolated sub-communities.

4 Experimental Campaign

In this section, we present the experimental campaign we conducted to test our framework. Specifically, in Subsection 4.1 we describe the data collection and preprocessing activities. In Subsection 4.2, we illustrate some exploratory analyses conducted on a first dataset. In Subsections 4.3, 4.4 and 4.5, we present our experiments related to network structure, community analysis and resilience, respectively. Finally, in Subsection 4.6 we illustrate several analyses performed on a second dataset.

Before starting the description of our experiments, we point out that the datasets and code used for them are available at `https://github.com/daisy-univpm/Resilience`.

4.1 Data Collection and Preprocessing

In order to perform our experimental analysis, we constructed a dataset obtained from Reddit, a popular social platform where users can share news, opinions and content on various topics [2]. Specifically, we derived our data of interest from the subreddit `r/worldcup`. The latter is dedicated to discussions related to the FIFA World Cup soccer tournament held in Qatar in late 2022. These discussions concern teams, players and matches. To extract data related to that subreddit, we resorted to `pushshift.io` [4], a repository of Reddit data. We collected data from this subreddit for the months of November and December 2022, which correspond to the period when the tournament was held.

Reddit provides a unique opportunity to study online social interactions and community dynamics since it allows the analysis of user-generated content and user interactions [11]. In particular, the subreddit `r/worldcup` provides a rich source of data to study the dynamics of a specific online community during a major event. The FIFA World Cup is a popular and widely watched event with a large and diverse audience, making it an ideal case for studying how online communities form, evolve, interact and disappear during a major global event [22].

The initial dataset obtained from `r/worldcup` through `pushshift.io` consisted of a set of comments. Each comment had the following fields associated with it:

- `id`, representing its unique identifier;
- `parent_id`, denoting the id of the parent comment, if it exists;
- `author`, indicating the anonymized username of its author;
- `created_utc`, representing the date and time when it was posted;
- `text`, denoting its text.

We preprocessed the dataset by removing non-English comments, empty comments, comments containing only links or images, and comments written by banned accounts. To automate the NLP tasks we used the `nltk` Python library. Due to the comment removals performed, the dataset was reduced by 10%; specifically, the final number of comments in it was 473,229.

The next step was the extraction of the topics associated with comments. These topics were to be represented in the form of keywords. To obtain them, we applied BERTopic [20] on our dataset, which returned a list of 247 unique topics discussed in the comments. Then, we assigned the topics to the comments based on the content of the latter. Upon completing this task, we added a new field, called `topics`, in our dataset, which stores, for each comment, a list of the topics discussed in it.

4.2 Exploratory analysis

To get a better understanding of our dataset, we performed an exploratory analysis on the preprocessed data. In this section, we provide a summary of the key findings obtained from that analysis.

First, we derived some basic statistics related to our dataset. They are shown in Table 1.

From the analysis of this table we can see that the number of comments is much higher than the number of authors, which, in turn, is much higher than the number of topics. In particular, the

Property	*Value*
Number of comments	473,229
Number of authors	77,655
Number of topics	247
Average number of words per comment	11.95 words
Median number of words per comment	7 words

Table 1: Basic statistics of our dataset

average number of comments per author is 6.09 and the ratio of authors to topics is about 314. It follows that the multidimensional network associated with the dataset has very unbalanced layers with regard to the number of nodes, and this must be taken into account in the next analyses (e.g., by suitably setting the parameter ρ in the definition of $Q_{\mathcal{M}}[\tau_1, \tau_2]$ - see Equation 3.8).

After obtaining the basic statistics of the dataset, we conducted its exploratory analysis. As a first task, we derived the temporal trend of comments. Figure 2 shows the number of comments posted each day during the data collection period. As we might expect, the number of comments reaches peak values on days when matches were being played, with the maximum number of comments reached on the day of the tournament final. Clearly, after the final, there is a significant decrease in the number of comments.

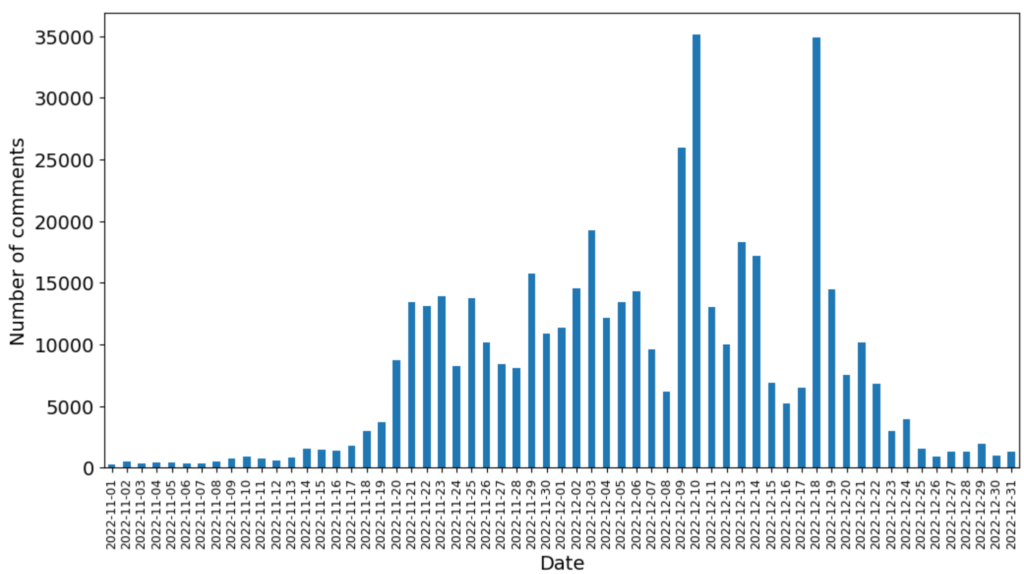

Figure 2: Temporal trends of comments in the dataset

The next analysis looked at user activity in the dataset. Figure 3 shows the distribution of comments against users in semi-log scale. From the analysis of this figure we can see that the corresponding distribution follows a power law. Specifically, the majority of users posted only one comment while a small number of users were extremely active and posted a large number of comments.

We also performed a topic analysis on the dataset using BERTopic. Figure 4 shows the distribution

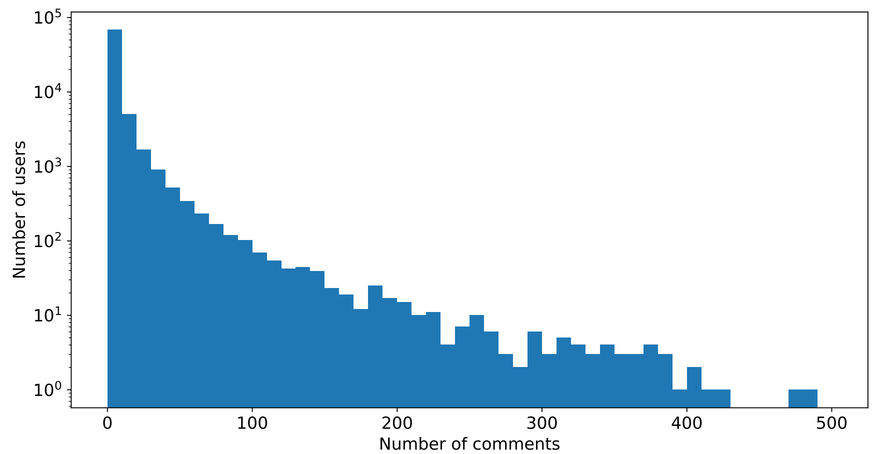

Figure 3: Distribution of comments against users (semi-log scale)

of comments against topics in semi-log scale. Also in this case, the distribution follows a power law. We found that the most discussed topic regarded match results, followed by teams and players. We also observed the presence of topics concerning the social and cultural aspects of the World Cup, e.g., the host country and fans.

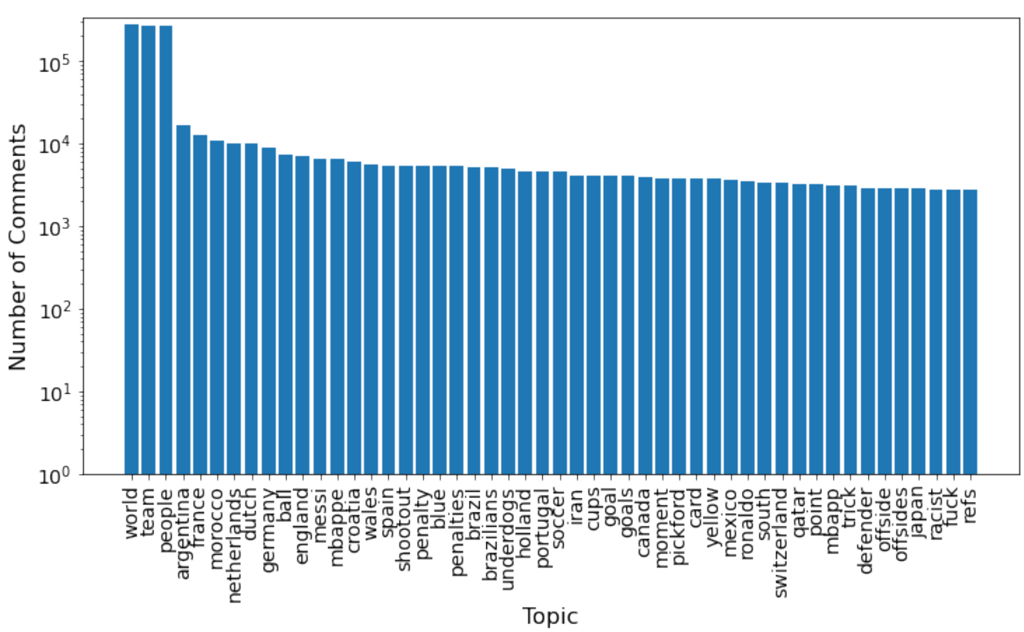

Figure 4: Number of comments per topic (semi-log scale)

4.3 Structural analysis

In this section, we present our experiments on the structural analysis of the multilayer network associated with our dataset. For this purpose, we use the measures defined in Section 3.2.1. In this activity, we consider both intra-layer and inter-layer analysis. In particular, we compute the assortativity values of each layer and of the whole multilayer network over time. This information could be useful for various applications, for example to investigate information diffusion or to design targeted interventions in the network.

As a first task, we computed the value of the degree-degree correlation coefficient r_L (Equation 3.3) for each layer of the multilayer network over time. The values obtained are shown in Figure 5. In this figure, the trend of r_L for the Authors (resp., Comments, Topics) layer is shown in blue (resp., orange, green).

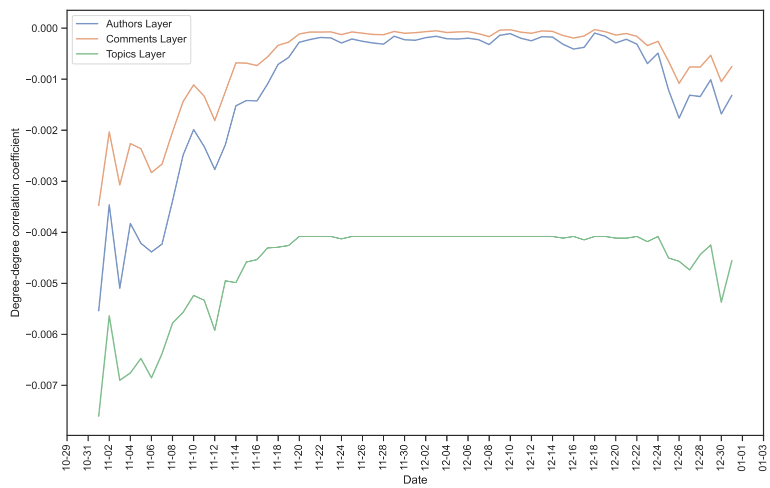

Figure 5: Values of r_L for each layer over time

From the analysis of this figure we can observe that the values of r_L are negative and increasing over time. This implies that the assortativity of the nodes of each layer in the network is negative (and thus we are in presence of disassortativity), and its absolute value decreases over time. In turn, this implies that nodes tend to be connected to nodes with significantly different degrees and that this tendency decreases over time. We can also observe that the assortativity values for the Authors and Comments layers are relatively close to each other, indicating that nodes in these layers show similar connection patterns. The presence of disassortativity is more prominent in the Topics layer than in the other layers. One possible explanation for why the Topics layer has higher disassortativity values than the Authors and Comments layers may lie in the fact that there exist edges between two topics if one is in the same category or a subcategory of the other (see Section 3.1). This also implies that there are also edges between general topics (which, therefore, have many edges and a high degree) and very specific topics related to subcategories (which, therefore, have very few edges). This behavior is not observed so strongly in the Authors and Comments layers, where each node can interact both with nodes having a degree similar to its own and with nodes having a degree very different from its own.

The decreasing trend of disassortativity values over time suggests that the network is becoming less modular, i.e., it is evolving toward a more random structure. This trend could be due to its growth; in fact, as the tournament progressed, new authors, comments and topics appeared in Reddit and new nodes were added to the multilayer network. The addition of new authors, comments and topics happened according to patterns different from the original ones. In particular, the new growth patterns turned out to be more random and less structured than the original ones. At the end of this discussion, we can see that r_L has proven to be an effective measure in distinguishing the behaviors of different layers with respect to assortativity and in extracting insights on this issue.

In addition to assortativity, another important structural measure to consider is clustering coefficient. The intra-layer clustering coefficient measures the tendency by which nodes in the same layer form closely interconnected groups. Figure 6 shows the values of the intra-layer clustering coefficient γ_L (see Equation 3.5) for the Authors, Comments and Topics layers over time.

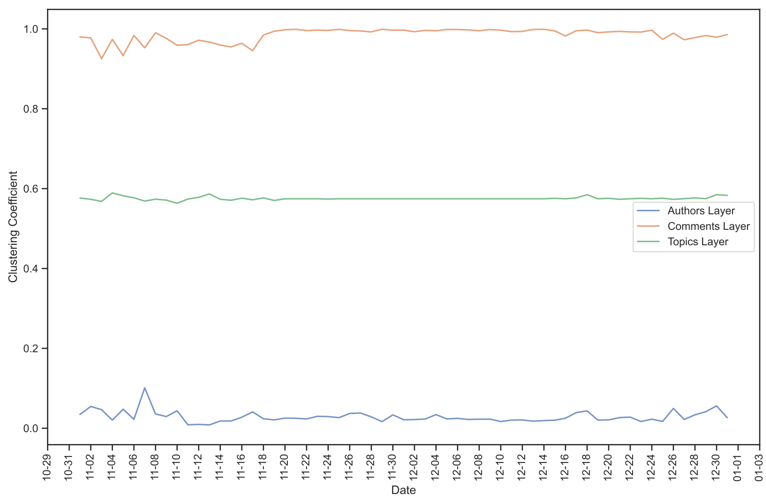

Figure 6: Values of γ_L for each layer over time

From the analysis of this figure we can observe that γ_L is lower in the Authors layer than in the other two layers, which suggests a lower tendency of authors to form cohesive groups and, therefore, not to form tightly connected communities. This may be attributed to the wide range of topics discussed by the various authors, which leads to a more scattered network structure. This trend of γ_L reflects the different perspectives held by the authors in this subreddit. In contrast, the Comments layer has a higher value of γ_L. This indicates that comments tend to form interconnected clusters, probably related to the chain of comments in which one is posted as a response to another. Finally, the topics layer shows a value of γ_L intermediate to that of the previous two layers. This can be explained by the fact that, as we have seen, comments are cohesive and clearly tend to lead to cohesive topics. On the other hand, we have seen that authors show varied interests by being part of multiple communities. Each community publishes comments that are cohesive but different from those published by another community. The contrasting influences of authors and comments on topics cause the value of γ_L of the Topics layer to be intermediate to that of the other two ones. Again, clustering coefficient also proved to be an effective measure for distinguishing the behaviors of various layers and extracting insights related to it.

Finally, we consider the average path length $\mathcal{L}_L$, which can give us important insights into the interconnections of nodes in the various layers of the network. The values of $\mathcal{L}_L$ for each layer over time are shown in Figure 7.

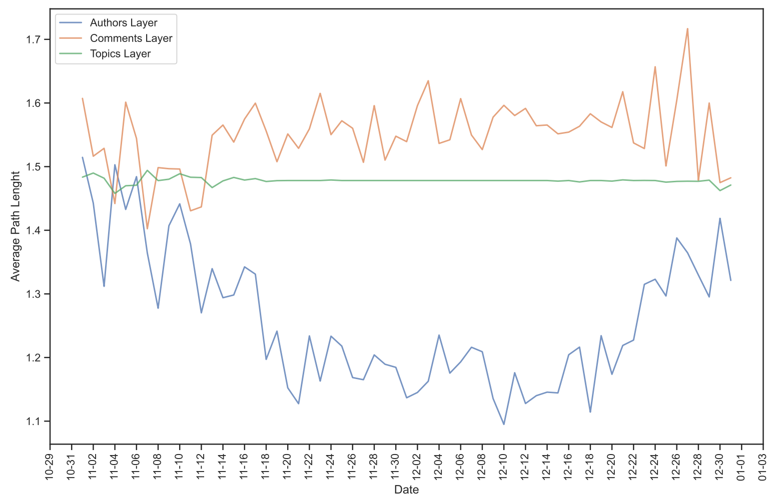

Figure 7: Values of $\mathcal{L}_L$ for each layer over time

From the analysis of this figure we can see that the values of $\mathcal{L}_L$ are very low in each layer; in fact, they are in the range $[1.12, 1.51]$ for the Authors layer, in the range $[1.40, 1.72]$ for the Comments layer, and in the range $[1.46, 1.58]$ for the Topics layer. This tells us that authors, comments and topics are well connected to each other in the corresponding layers, either directly or through very few intermediaries. Consequently, we can conclude that, in the subreddit from which our dataset was derived, communication is direct and immediate and information flows efficiently. A second insight that emerges from this figure is that the value of $\mathcal{L}_L$ is fairly constant over time for topics while it varies more markedly for authors and comments. This can be explained by considering that topics remain essentially unchanged throughout the period of interest, undergoing very little variations. In contrast, authors and comments vary considerably over time, as new authors participate to discussions, others no longer participate, and always new comments are posted every day. At the end of this analysis, we point out that $\mathcal{L}_L$, like the other two coefficients, proved effective in highlighting similarities and differences in the behaviors of layers and effective in discovering insights of interest.

As for the inter-layer analysis, we calculated the value of the degree-degree cross-correlation coefficient r_{L_1,L_2} (see Equation 3.4) for the different pairs of layers in our multilayer network over time. The values obtained are reported in Figure 8. Specifically, we show the coefficient values for the Authors-Topics pair in green, those for the Comments-Topics pair in orange and those for the Authors-Comments pair in blue.

From the analysis of this figure, we can observe that r_{L_1,L_2} is essentially equal to 0 for all pairs of layers and remains essentially constant over time. This means that there is no form of degree assortativity or degree disassortativity between authors and topics, comments and topics and authors and comments. In turn, this implies that, for example, central authors in a community might post both

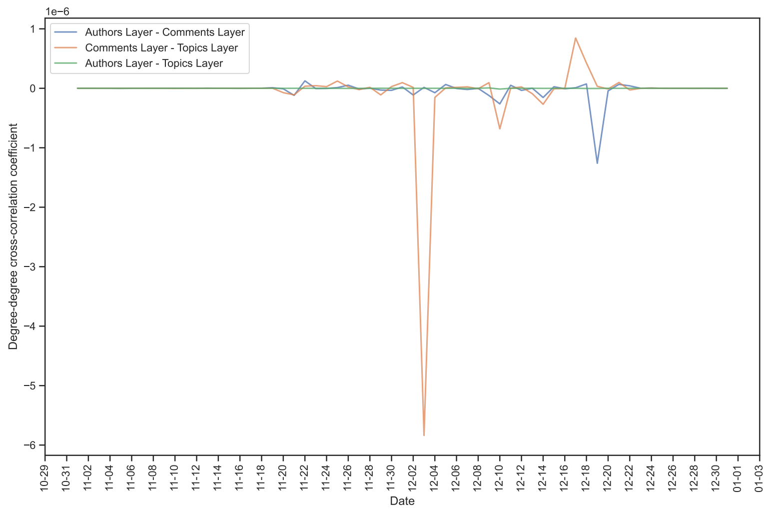

Figure 8: Values of r_{L_1,L_2} for the different pairs of layers over time

highly debated and poorly debated comments, and vice versa. Similarly, highly debated comments may involve both general topics and highly specific ones, and vice versa. Finally, central users may be interested in both highly related topics and isolated ones, and vice versa. In this case, r_{L_1,L_2} allowed us to discover the similar behavior held by the three layers of our network. This represents a useful insight for the network investigation discovered thanks to this measure of our framework.

4.4 Community analysis

In this section, we present our experiments on the community analysis of the multilayer network associated with our dataset. For this purpose, we use the measures defined in Section 3.2.2, which allow us to perform both intra-layer and inter-layer analysis.

We start the intra-layer analysis by investigating the modularity coefficient Q_L introduced in Equation 3.7. The knowledge of the value of Q_L in a layer can help identify insights characterizing the behavior of the corresponding nodes. We calculated the value of Q_L for the Authors, Comments and Topics layers. The results obtained are shown in Figure 9.

From the analysis of this figure we can derive some interesting insights. First, we observe that the Author layer has a high value of Q_L with fluctuations that are however limited. This means that authors tend to organize themselves into communities that are rather internally cohesive and weakly coupled with each other. At the other extreme, we have the Topics layer, which has low values of Q_L that remain fairly constant over time. This implies that topics are difficult to group into cohesive communities and their organization remains constant over time. The reason for this is the same that explains the trend of $\mathcal{L}_L$ for the Topics layer in Figure 7: topics were generally defined at the beginning of the observation period and are unlikely to change over time. Finally, we can observe that the value of Q_L for the Comments layer is intermediate to that associated with the other layers with very large fluctuations. This indicates that the extent to which comments can be organized into cohesive and weakly coupled clusters varies over time. On some days we observe the presence of very high peaks of

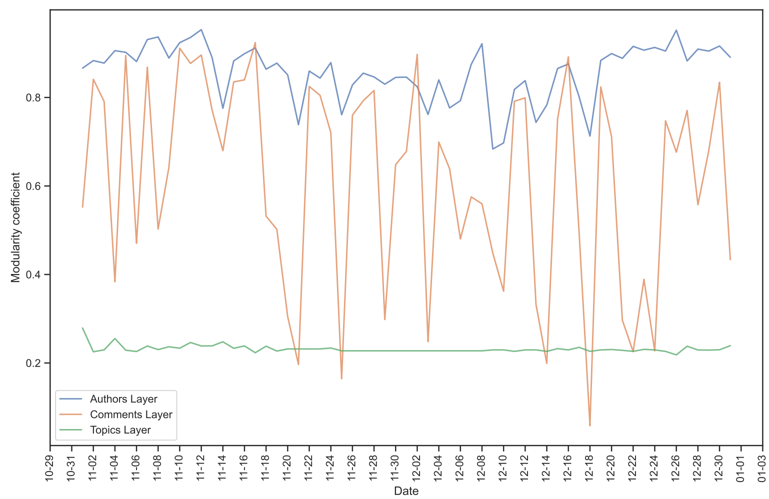

Figure 9: Values of Q_L for each layer over time

Q_L, comparable to the values of Q_L for the authors. Conversely, on other days, we notice the presence of very low peaks of Q_L, comparable to, or even lower than, the values of Q_L for the topics. We conclude this description by noting that Q_L has been effective in sketching the different behavior of layers with regards to the existence of cohesive groups of nodes within them. Indeed, thanks to it, we were able to derive useful insights on this issue.

To perform inter-layer analysis, we employed the cross-modularity coefficient $Q_{\mathcal{M}}$ introduced in Equation 3.8. We computed the values of $Q_{\mathcal{M}}$ on the multilayer network associated with our dataset. The results obtained are shown in Figure 10.

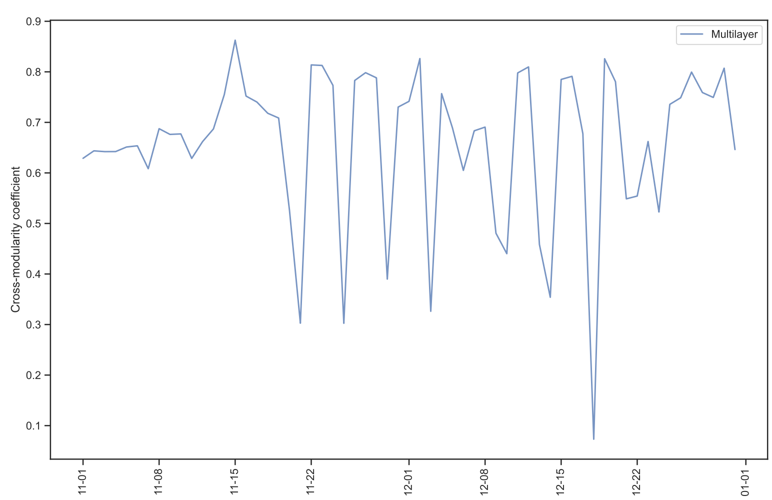

Figure 10: Values of $Q_{\mathcal{M}}$ over time

From the analysis of this figure we can observe that the values of $Q_{\mathcal{M}}$ vary considerably over time from a maximum of 0.8626 to a minimum of 0.0731. The fluctuations are considerable and frequent over time. Most of the time, the values of $Q_{\mathcal{M}}$ settle in a range between 0.4 and 0.8. This denotes some tendency of nodes of different layers to organize themselves into homogeneous clusters although this tendency is rather weak. We can hypothesize that, in many cases, users, the comments they post and the corresponding topics tend to organize together in the same community. However, this tendency is not stable or strong. This may be quite understandable if one considers that there are some specific topics (related, for example, to a player or a team), which can be enclosed in an inter-layer cluster. On the other hand, at the same time, there are as many topics that are so generic that it is virtually impossible to confine them in a single inter-layer cluster. Indeed, they may span many clusters because they are debated in many comments, perhaps posted by users belonging to completely different communities. In any case, the coefficient $Q_{\mathcal{M}}$ has also proved to be useful in supporting the analysis of the temporal evolution of the multilayer network, particularly the ability of its nodes to organize themselves into strongly cohesive and weakly coupled clusters comprising nodes belonging to different layers.

4.5 Resilience analysis

In this section, we focus on the analysis of resilience. For this purpose, we use the measures defined in Section 3.2.3. Again, we consider both intra-layer and inter-layer analysis. As we have seen in Section 3.2.3, the possible network modifications on which to study resilience are numerous. Due to lack of space, we focus only on some of them to see if our measures are effective, i.e., able to support us in extracting insights related to the resilience of the multilayer network under consideration.

We start with intra-layer analysis and the removal of existing nodes and their edges. In this case, we should consider both the Clustering Coefficient Variation $\Delta\gamma_L[\tau_1, \tau_2]$ and the Level-wise Degree Distribution Variation $\Delta D_L[\tau_1, \tau_2]$ (see Section 3.2.3). Due to lack of space, in this subsection we focus only on the latter measure, while we will employ the former in the experiments in the next subsection. During the experiments, we randomly removed 5%, 10%, 20%, 30%, 40% and 50% of the nodes and their respective edges from each layer and calculated $\Delta D_L[\tau_1, \tau_2]$ (see Equation 3.10). We repeated this task five times and computed the average values. In Figures 11 - 13 we show the results obtained for the Authors, Comments and Topics layers over time.

From the analysis of these figures we can draw some interesting insights concerning the resilience of the three layers. First, we observe that in all three cases there is a very irregular trend with the presence of sharp variations and peaks. As for the resilience values, these are generally lower for the Comments layer than for the other two layers. In all the three layers, we can observe that as long as the percentage of removed nodes is low (i.e., up to 20%), the variations of $\Delta D_L[\tau_1, \tau_2]$ are small, which is an indication of a satisfactory resilience of the three layers. When the percentage of removed nodes further increases we can see that the changes in $\Delta D_L[\tau_1, \tau_2]$ grow significantly, reaching peaks as high as 0.8 (remember that the maximum possible value is 1). In our opinion, this is correct in that each layer should be able to recognize the presence of perturbations capable of upsetting the structure and behavior of the underlying network. From this point of view, the Authors and Topics layers show the best overall performance. The performance of Comments is satisfactory but denotes some difficulty in recognizing even substantial perturbations of the network underlying the layer. From the insights drawn above, it appears that $\Delta D_L[\tau_1, \tau_2]$ has been able to give us a very clear idea of the resilience of

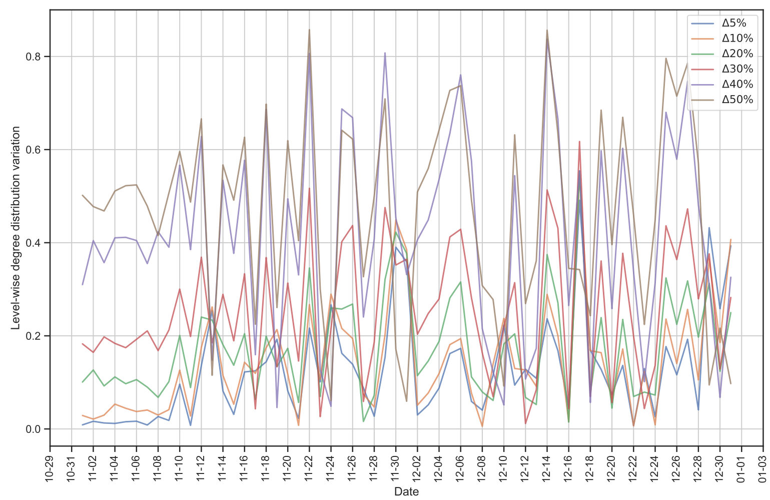

Figure 11: Values of $\Delta D_L[\tau_1, \tau_2]$ for the Authors layer over time

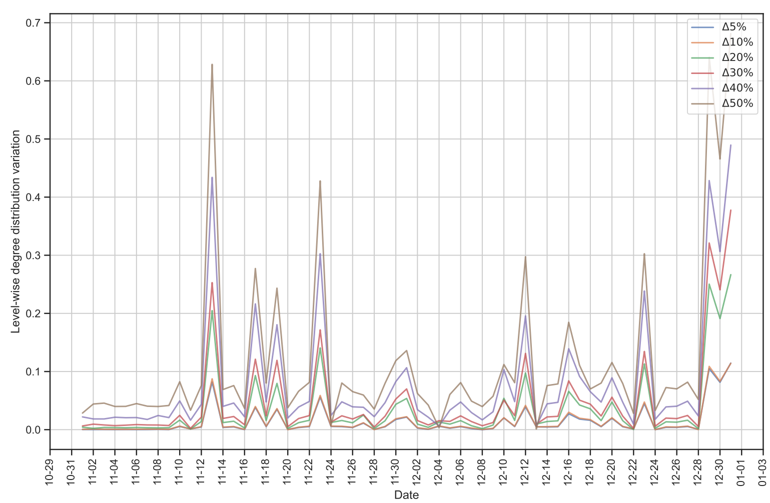

Figure 12: Values of $\Delta D_L[\tau_1, \tau_2]$ for the Comments layer over time

the various layers in presence of removals of nodes and their associated edges.

We continue the intra-layer analysis considering the removal of existing edges. In this case, due to space limitations we show only our analysis on the Clustering Coefficient Variation $\Delta\gamma_L[\tau_1, \tau_2]$ (see Equation 3.9). During the experiments, we randomly removed 5%, 10%, 20%, 30%, 40% and 50% of the edges from each layer and computed $\Delta\gamma_L[\tau_1, \tau_2]$. We repeated this task five times and calculated the average values. In Figures 14 - 16, we show the obtained results for the Authors, Comments and Topics layers over time.

From the analysis of these figures we can see that the Authors layer has a completely different behavior than the Comments and Topics layers. In particular, it is much less resilient than the other two layers when the fraction of removed edges is low. When that fraction becomes very high, the Authors layer is much more resilient than the other two. In addition, the trend of $\Delta\gamma_L[\tau_1, \tau_2]$ for the

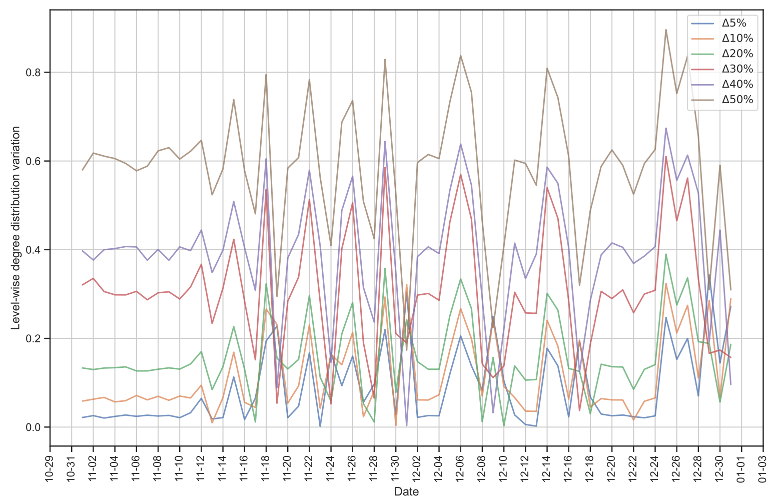

Figure 13: Values of $\Delta D_L[\tau_1, \tau_2]$ for the Topics layer over time

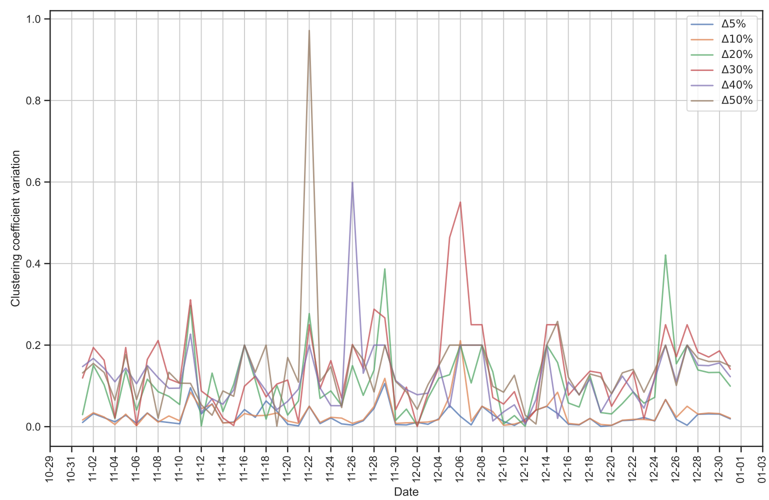

Figure 14: Values of $\Delta \gamma_L[\tau_1, \tau_2]$ for the Authors layer over time

Authors layer is much more irregular than for the other two layers. In fact, we can see the presence of both abrupt variations and some pronounced peaks in correspondence with the highest reductions in the number of edges. The other two layers show much more regular trends and are very resilient when the percentage of removed edges is low and much less resilient when that percentage is high. In our opinion, this is the ideal behavior a layer should have, i.e., it should not undergo huge variations for the loss of a few edges, while it should show "alarm bells" when such loss starts to become substantial. In any case, as far as the effectiveness of the proposed measure (i.e., $\Delta \gamma_L[\tau_1, \tau_2]$) is concerned, we observe that it has been able to give us a very clear idea of the difference in the behavior of the resilience of the various layers in presence of edge removals.

After considering the intra-layer analysis, we moved on to perform the inter-layer analysis. Again, we started with the removal of existing nodes and their edges. We should consider the three coefficients

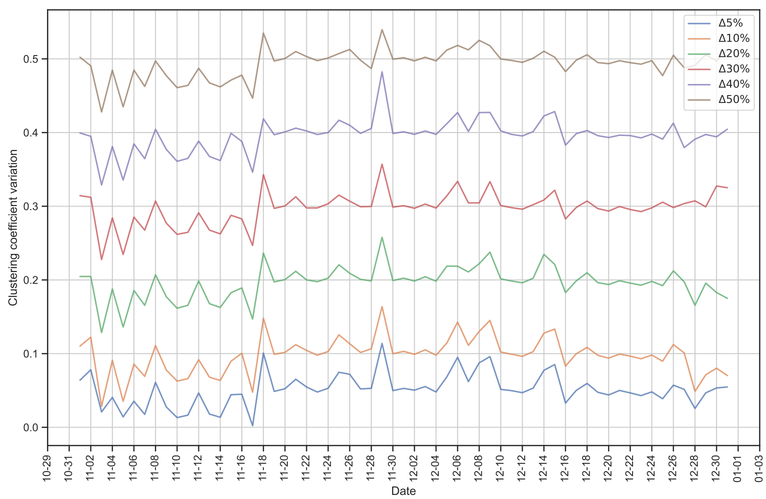

Figure 15: Values of $\Delta\gamma_L[\tau_1, \tau_2]$ for the Comments layer over time

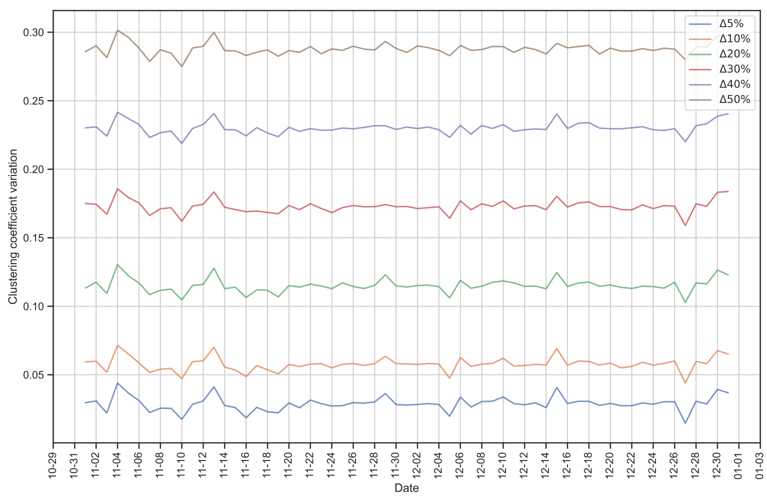

Figure 16: Values of $\Delta\gamma_L[\tau_1, \tau_2]$ for the Topics layer over time

provided for the study of resilience in this case, namely Modularity Variation $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$, Inter-layer Correlation Variation $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ and Multilayer Degree Distribution Variation $\Delta D_E[\tau_1, \tau_2]$. Due to lack of space we focus only on $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ and $\Delta D_E[\tau_1, \tau_2]$, while we will employ $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ in the analysis described in the next section. During the experiments we randomly removed 5%, 10%, 20%, 30%, 40% and 50% of the nodes and their respective inter- and intra-edges and calculated the value of $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ for the Authors and the Comments layers (see Equation 3.13) over time. We repeated this task five times and calculated the average values. In Figure 17 we show the results obtained.

From the analysis of this figure, we can observe that $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ is almost always null even in the presence of substantial perturbations and, therefore, that the multilayer network and the multimodal social network from which it is derived are substantially resilient to the removal of nodes and the

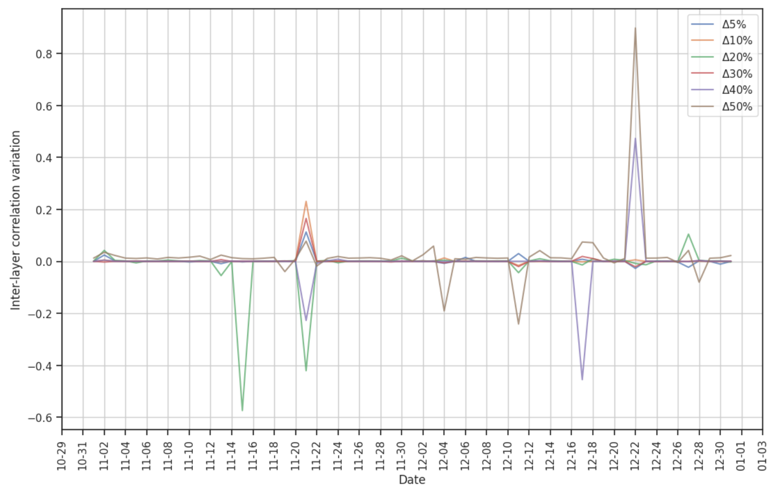

Figure 17: Values of $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ for the Authors and the Comments layers over time

corresponding intra- and inter-layer edges. This means that there is generally no degree assortativity or disassortativity between the nodes of the various layers. Because of this, for example, a very famous user might have posted both highly debated comments and comments that attracted no interest, and vice versa. In the presence of very large perturbations, it is possible to observe the existence of some peaks of assortativity or disassortativity, which last for a very short time. Interestingly, assortativity peaks are more frequent but on average smaller than disassortativity ones. The only exception is the very high assortativity peak on December 22, 2022. Finally, it is worth pointing out that there is a strong similarity between the results of inter-layer assortativity shown in Figure 17 and those of intra-layer assortativity reported in Figure 8. Actually, the results shown in Figure 17 are in full agreement with those reported in Figure 8, and vice versa. We again want to highlight that, beyond the specific dataset and the data stored in it, the index $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ introduced in our framework has proven to be effective in extracting potentially interesting insights regarding the resilience of a multilayer network in presence of inter-layer perturbations.

Proceeding in the same way, and considering the same perturbations as in the previous case, we calculated the value of $\Delta D_E[\tau_1, \tau_2]$ over time. The results obtained are shown in Figure 18.

From the analysis of this figure we can observe that the multilayer network associated with our dataset exhibits an optimal behavior with regard to resilience. In fact, in presence of a low number of perturbations, the value of $\Delta D_E[\tau_1, \tau_2]$ is small. In contrast, when the number of perturbations becomes large, the value of $\Delta D_E[\tau_1, \tau_2]$ becomes very large, acting as an "alarm bell". Thus, we can conclude that also $\Delta D_E[\tau_1, \tau_2]$ has proven effective in extracting potentially interesting insights regarding the resilience of a multilayer network in presence of inter-layer perturbations.

The inter-layer analysis continued with the investigation on the removal of existing edges. As for this case, due to space limitation, we show only the analyses performed for the Modularity Variation $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ and the Multilayer Degree Distribution Variation $\Delta D_E[\tau_1, \tau_2]$. During the experiments we randomly removed 5%, 10%, 20%, 30%, 40% and 50% of the inter-layer edges and computed the value of $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ and $\Delta D_E[\tau_1, \tau_2]$ over time. The trend of $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ is shown in Figure 19 while

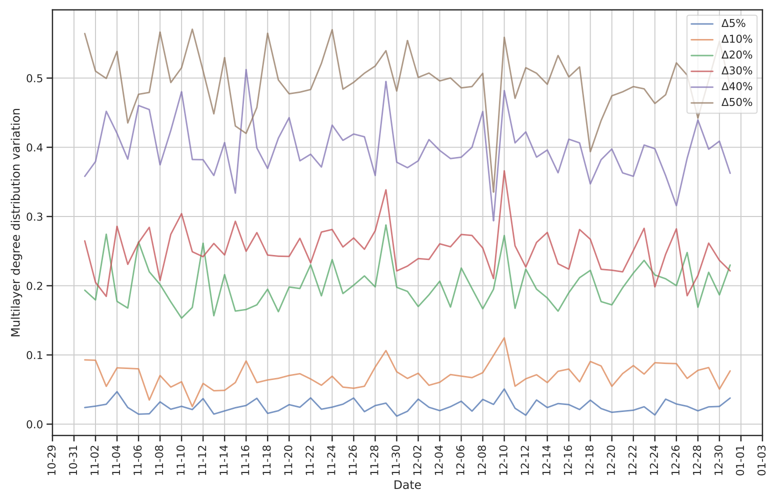

Figure 18: Values of $\Delta D_E[\tau_1, \tau_2]$ over time

the one of $\Delta D_E[\tau_1, \tau_2]$ is reported in Figure 20.

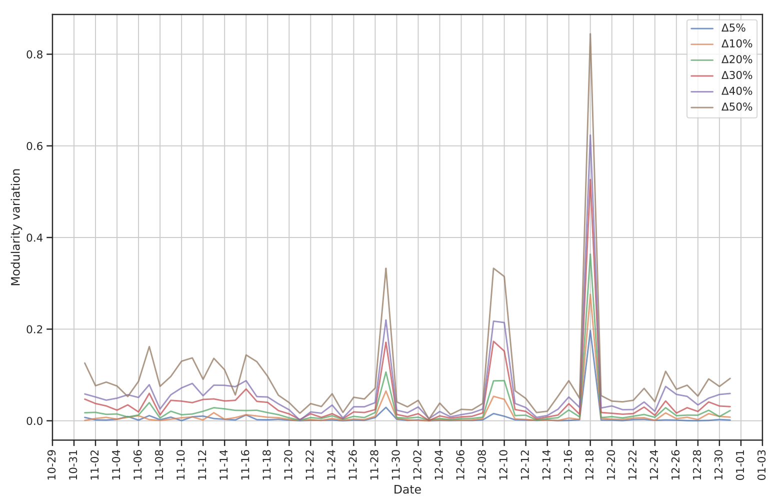

Figure 19: Values of $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ over time

From the analysis of Figure 19 we can observe that the multilayer network associated with our dataset proves to be resilient with respect to perturbations related to the removal of inter-layer edges. In some ways, the resilience evidenced through $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ appears excessive as well, since even significant perturbations cause limited changes in this measure. In Figure 19, we can observe some peaks that, however, disappear by the next day. On the other hand, the analysis in Figure 20 shows an optimal behavior of our multilayer network with respect to resilience since a small reduction in the number of inter-layer edges leads to a limited change in the value of $\Delta D_E[\tau_1, \tau_2]$. In contrast, when the number of inter-layer edges removed begins to become large, the value of $\Delta D_E[\tau_1, \tau_2]$ begins to vary substantially from its original value. Again, such variation can be used as an "alarm bell" to signal the presence of excessive perturbations that may disrupt the original network. Finally, even

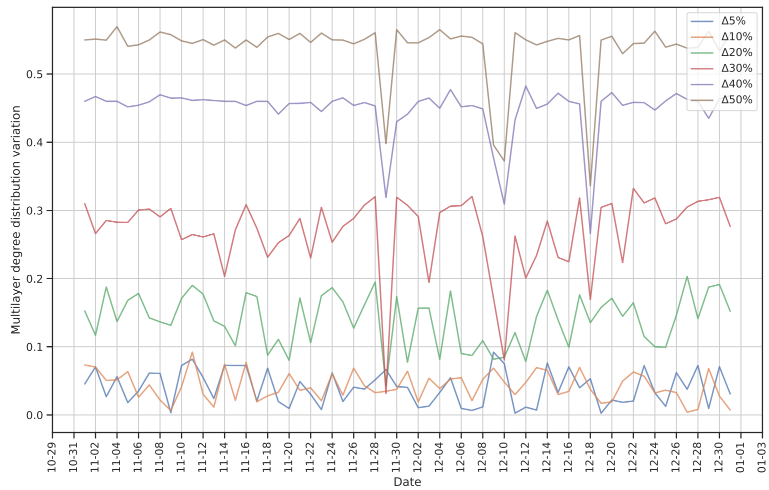

Figure 20: Values of $\Delta D_E[\tau_1, \tau_2]$ over time

for the removal of existing edges, the two coefficients $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$ and $\Delta D_E[\tau_1, \tau_2]$ proved effective in supporting the analysis of the resilience of a multilayer network and deriving insights related to it.

4.6 Experiments on a second dataset

All the experiments shown so far were conducted on the dataset described in Section 4.1. To verify the stability of the results obtained, we thought it appropriate to repeat these experiments on a second dataset, which had a similar structure to the first one but was derived from a different social network. Our choice fell on Hacker News [38], a popular social platform that focuses on technology, startups, programming, etc. It allows its users to share and discuss articles, news and ideas in the context of Information Technology, enabling them to stay informed about the latest trends in this field. Specifically, our dataset refers to the same time interval as the other one, i.e., from November 1, 2022 to December 31, 2022. Unlike Reddit, which is a generic social network, Hacker News is specific to Information Technology; therefore, this choice allowed us to test our approach on a dataset with different characteristics from those of Reddit. Also on this dataset we performed all the necessary ETL activities. Again, we used BERTopic for the extraction of topics associated with comments. In Table 2, we report some basic statistics of this dataset.

Property	*Value*
Number of comments	179,491
Number of authors	29,186
Number of topics	118
Average number of words per comment	35.49
Median number of words per comment	23

Table 2: Basic statistics of the second dataset

From the analysis of this table and Table 1 we can observe that the basic statistics of the two datasets show several important differences and several relevant similarities. Indeed, the number of comments and authors is much lower, while the average number of comments per author is similar. The ratio of the number of comments to the number of topics is about half the one of the previous dataset. Finally, the comments on this dataset are much longer than the ones on the first dataset.

In the following, for the sake of space, we cannot report the results of all the experiments we conducted in the second dataset. Therefore, we will only illustrate those that we consider most significant, while giving only a brief mention of the others.

In Figure 21, we report the values of the degree-degree correlation coefficient r_L for each layer of the multilayer network over time. In it, the trend of r_L for the Authors (resp., Comments, Topics) layer is shown in blue (resp., orange, green). Comparing this figure with Figure 5 related to the first dataset, we can observe that the results obtained (and the conclusions that can be drawn from them) are similar. In fact, the differences that can be observed are negligible.

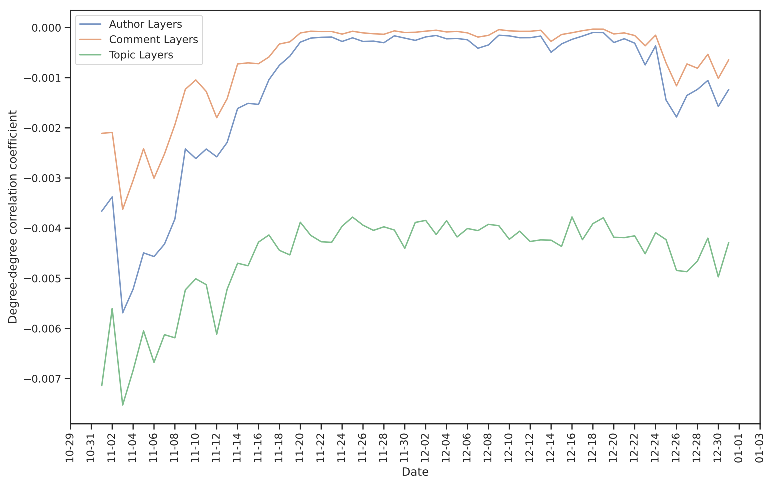

Figure 21: Values of r_L for each layer over time for the second dataset

In our experiments, we also calculated the values of the intra-layer clustering coefficient γ_L and the intra-layer average path length $\mathcal{L}_L$ for each layer of the multilayer network corresponding to the second dataset. For sake of space, we do not report the corresponding figures. However, the trends obtained are very similar to those shown in Figures 6 and 7 for the first dataset.

After that, we switched to inter-layer analysis and calculated the value of the degree-degree cross-correlation coefficient r_{L_1,L_2} for the different pairs of layers of our multilayer network over time. The values obtained are shown in Figure 22. Comparing this figure with Figure 8 related to the first dataset, we can observe that the results obtained (and the corresponding considerations that can be drawn) are similar.

At this point, we focused on community analysis and calculated the value of the modularity coefficient Q_L for each layer of the multilayer network associated with the second dataset. We also calculated the value of the cross-modularity coefficient $Q_{\mathcal{M}}$ on the same multilayer network. The trends of these two coefficients are shown in Figures 23 and 24. Comparing Figure 23 with Figure 9 and Figure 24 with Figure 10, we can observe that the trends of the two modularity parameters do

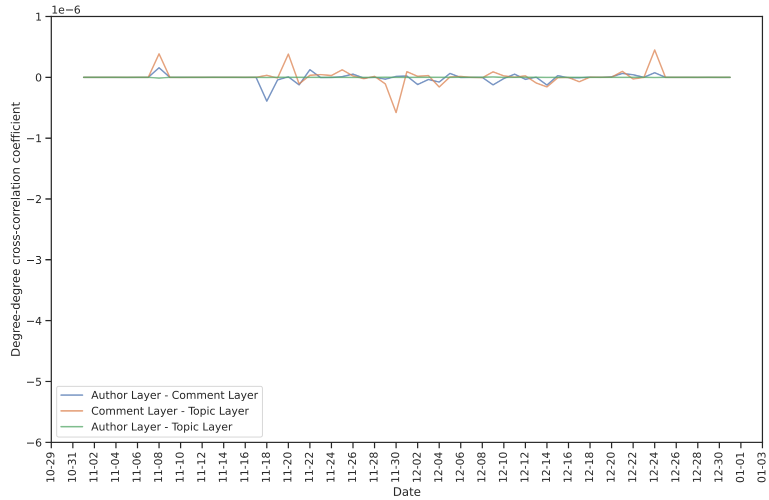

Figure 22: Values of r_{L_1,L_2} for the different pairs of layers over time for the second dataset

not differ significantly when passing from the first to the second dataset.

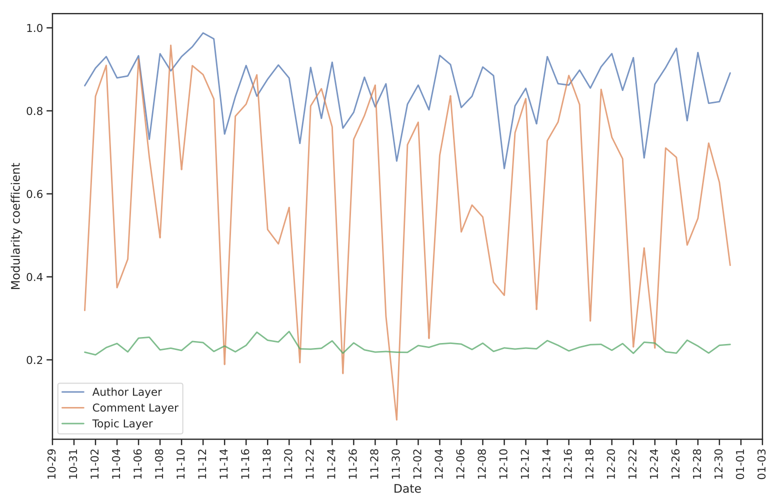

Figure 23: Values of Q_L for each layer over time for the second dataset

After analyzing modularity, we turned to testing resilience. Specifically, we started with the intra-layer resilience and calculated the trends of the Clustering Coefficient Variation $\Delta\gamma_L[\tau_1, \tau_2]$ and the Level-wise Degree Distribution Variation $\Delta D_L[\tau_1, \tau_2]$ after removing existing nodes and their edges in the multilayer network associated with the second dataset. In Figure 25, we report the values of $\Delta D_L[\tau_1, \tau_2]$ for the Authors layer over time. As can be seen by comparing this figure with Figure 11, the results obtained for the two datasets are similar. We can draw analogous conclusions by considering the variation of $\Delta D_L[\tau_1, \tau_2]$ over time for the other two layers and the variation of $\Delta\gamma_L[\tau_1, \tau_2]$ over time for all layers. In addition, we also obtained similar results for the two datasets when we computed the variation of $\Delta\gamma_L[\tau_1, \tau_2]$ and $\Delta D_L[\tau_1, \tau_2]$ after removing existing edges for the three layers under

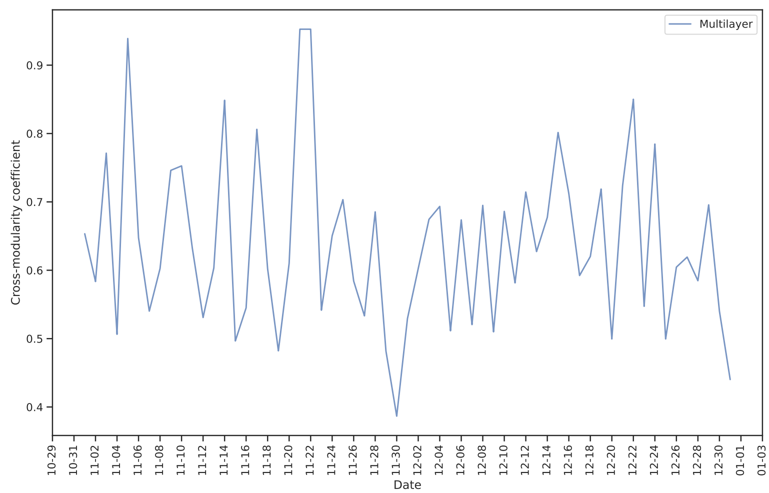

Figure 24: Values of $Q_{\mathcal{M}}$ over time for the second dataset

consideration. Once again, we cannot report the corresponding figures due to lack of space.

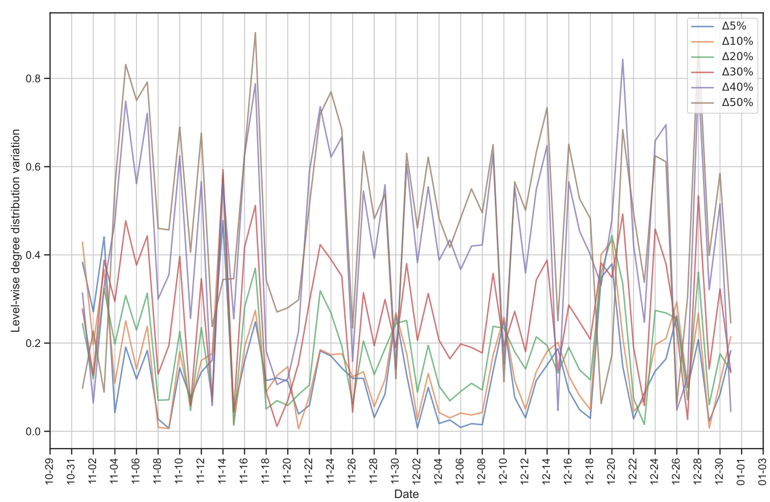

Figure 25: Values of $\Delta D_L[\tau_1, \tau_2]$ for the Authors layer over time for the second dataset

The last tests we conducted were on resilience to inter-layer perturbations. Specifically, we computed the trends of the Modularity Variation $\Delta Q_{\mathcal{M}}[\tau_1, \tau_2]$, Inter-layer Correlation Variation $\Delta r_{L_1,L_2}[\tau_1, \tau_2]$ and Multilayer Degree Distribution Variation $\Delta D_E[\tau_1, \tau_2]$ after both the removal of nodes and their edges and the removal of edges. The results obtained for this dataset show no significant change from those obtained for the first dataset. For lack of space we report only one case, specifically the variation of $\Delta D_E[\tau_1, \tau_2]$ over time after the removal of nodes and their edges. It is shown in Figure 26. The analysis of this figure and the corresponding Figure 18 for the first dataset confirms what we said about the similarity of the results obtained with the two datasets.

At the conclusion of this set of analyses we can observe that all the tests performed returned similar results for the first and the second datasets, thus showing that the results obtained are stable. This

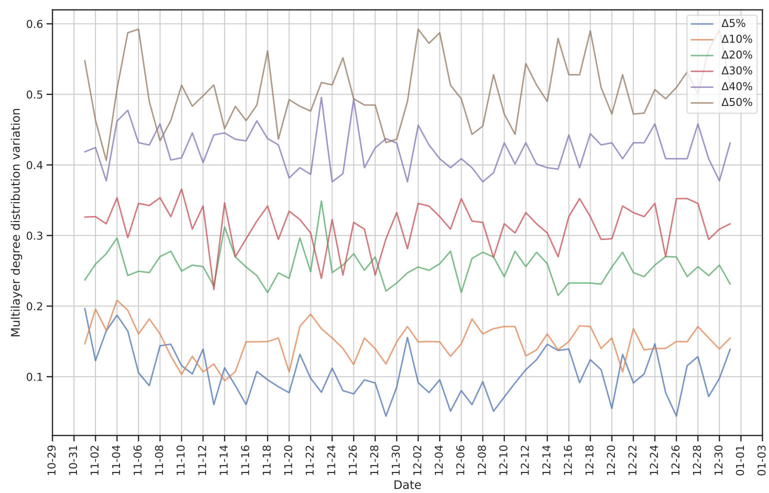

Figure 26: Values of $\Delta D_E[\tau_1, \tau_2]$ over time for the second dataset

is even more significant when we consider that the two datasets come from two very different social networks, one generic (Reddit) and one specific (Hacker News). In addition, the first dataset is related to a specific event (the FIFA World Cup Soccer Tournament held in Qatar in late 2022) while the second one includes all the discussions made on Hacker News during the period under consideration.

5 Discussion

In this section, we present a discussion regarding the proposed approach. In particular, we comment on the results obtained from our experiments (Subsection 5.1) and compare the characteristics of our approach with those of related ones (Subsection 5.2).

5.1 Discussion on experiments

After conducting all the experiments outlined in Section 4, some considerations about the work done and the results obtained are in order. First of all, we emphasize that the main objective of this paper is the definition of a framework to analyze the temporal evolution and resilience of a multilayer network and the multimodal social network modeled through it. Accordingly, the multilayer network and the associated multimodal social network built from the dataset on the World Cup soccer tournament are to be intended as tools for our experiments and not as their ultimate goal. In other words, the goal of the experiments was not finding insights into the temporal evolution and resilience of those two specific networks. This represents a side, although very interesting, result. Instead, our main goal was to understand whether our proposed framework and its analysis measures are adequate to assess the temporal evolution and resilience of a multilayer network and/or a multimodal social network under consideration.

From this point of view, the results obtained are promising. In fact, the analysis measures associated with our framework proved capable of: *(i)* highlighting the presence of different behaviors by different layers; *(ii)* highlighting the presence of assortativity, disassortativity or the absence of both of them in both individual layers and the overall multilayer network; *(iii)* understanding whether or not information flows well within each layer and in the overall multilayer network; *(iv)* understanding whether or not well-defined communities are present within each layer and in the overall multilayer network; *(v)* understanding whether or not a layer and the multilayer network as a whole are resilient to various forms of perturbations; *(vi)* understanding whether or not the multilayer network, despite its resilience, is capable of recognizing the presence of excessive perturbations that may disrupt one of its layers, a pair of layers, or the multilayer network as a whole.

With regard to temporal evolution and especially resilience, some of the analysis measures included in our framework proved more appropriate in some cases while others proved more adequate in other cases. The framework, with its overall set of measures, proved very effective in handling both the temporal evolution and resilience of a multilayer network and its associated multimodal social network. In fact, we have not encountered a single case in which at least one (and almost always more than one) of the analysis measures proposed in our framework has not proved totally effective in handling it.

5.2 Comparison of our approach with related ones

In Section 2, we have presented several related approaches and have seen that they can belong to two research strands, namely: *(i)* analysis of the properties of multilayer networks; *(ii)* study of multimodal social networks through their modeling by means of multilayer networks. We start our discussion by pointing out that our approach is the first to intersect these two strands. With that said, let us begin to make our comparison.

As far as the first strand is concerned, in Section 2 we have presented the approaches described in [39, 15, 19, 18, 7, 12, 44, 30]. The approach in [39] and our framework can be considered orthogonal. In particular, several insights discovered in the former can be used in the latter; moreover, our framework may allow us to study the relationships between prosociality and resilience. The approach in [15] studies the robustness of the multilayer network by removing nodes based on a high-degree or low-degree selection strategy. Our framework also uses the node removal technique to evaluate resilience. However, except for this similarity, the two approaches are substantially different in that their objectives are different. As for the approach presented in [19], its goals and those of our framework are very different. The work in [18] and ours share a few similarities. In particular, the former focuses on the structural reducibility of a network, which is a measure of the difference in interaction patterns in the multilayer network. In contrast, our framework does not limit its analysis to interaction patterns, but considers the overall multilayer network making use of different measures. Although the goals of the approach in [12] differ from those of our framework, there are some concepts and insights present in the former that may be useful in the latter. The approach of [44] shares some similarities with our framework. For example, both of them consider the concept of modularity, although the approach of [44] focuses on community detection while the goals of our framework are the study of temporal evolution and resilience. The framework in [30] and ours share some similarities although their objectives are different. Interestingly, this framework introduces the multilayer modularity index that is used within our framework (see Section 3.2.2).

As for the second research strand, in Section 2 we have considered the approaches described in [8, 27, 5, 34]. The approach in [8] and our framework share some similarities, such as using a multilayer network to analyze a social platform. However, their goals are different since the approach in [8] presents an empirical analysis while our approach aims to study time evolution and resilience from both theoretical and empirical perspectives. The approach in [27] can be considered orthogonal to our framework. However, although they have different goals, there are some insights from the former (e.g., the use of topic modeling) that can be included into the latter. Also the approach in [5] and our framework can be considered orthogonal. In particular, an anomaly detection task could be integrated into our framework based on the resilience measure. Finally, different insights detected in [34] could also be used in our approach. For example, resilience in our approach could be defined as a measure that indicates how easily the sharing pattern of a piece of content is carried out in the multilayer network.

6 Conclusion

In this paper, we have proposed a multilayer network-based framework for investigating the temporal evolution and resilience of multimodal social networks. In particular, we have shown how multilayer networks are a powerful enough tool to model all the characteristics of a multimodal social network, that is, a social network with different types of nodes and edges. We have initially defined the model for mapping a multimodal social network into a multilayer network having a layer for each type of nodes and whose inter-layer edges model edges connecting nodes of different types in the multimodal social network. Then, we have described the technical details of our framework and we have seen that it is based on three types of measures. The first two concern the structure of the network and communities and are used to study the temporal evolution of the network. The third type concerns resilience. Finally, we have presented a rich experimental campaign with the aim of demonstrating the effectiveness of the proposed framework in achieving the goals for which it was designed.

In summary, the main contributions of this paper are: *(i)* the definition of a framework and a set of measures to investigate the temporal evolution and resilience of a multilayer network; *(ii)* the definition of a method to map a multimodal social network into a multilayer network; *(iii)* the application of our framework and its set of measures to study the temporal evolution and resilience of a multimodal social network.

This paper should not be considered as an endpoint but rather as a starting point for further researches on this topic. In particular, we plan to extend our framework to represent and handle non-dyadic relationships, i.e., interactions involving more than two nodes, by providing support for layers representing hypernetworks. A second aspect we intend to study is user dynamism, that is, understanding how often users enter and leave the underlying network, leave communities they used to belong to, join existing communities or try to build new ones. To this end, we believe to start with the analysis measures identified in this paper and adapt them to these new goals. Finally, we think we can use our framework to investigate the engagement dynamics related to the various topics. In other words, we would like to understand how a topic can become attractive to users, or conversely ceases to be attractive to them, or ceases to be attractive to one category of users but begins to be attractive to another category. To this end, we think we can adapt the various measures introduced in this paper while enhancing the Topics layer with additional information, such as the type of content discussed in each topic.

References

[1] A. Amelio, G. Bonifazi, F. Cauteruccio, E. Corradini, M. Marchetti, D. Ursino, and L. Virgili. Representation and Compression of Residual Neural Networks through a Multilayer Network Based Approach. *Expert Systems With Applications*, 215(119391), 2023. Elsevier.

[2] K.E. Anderson. Ask me anything: what is Reddit? *Library Hi Tech News*, 32(5):8–11, 2015. Emerald.

[3] O. Artime, B. Benigni, G. Bertagnolli, V. d'Andrea, R. Gallotti, A. Ghavasieh, S. Raimondo, and M. De Domenico. Multilayer Network Science: From Cells to Societies. *Elements in Structure and Dynamics of Complex Networks*, 2022. Cambridge University Press.

[4] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift Reddit dataset. In *Proc. of the International AAAI Conference on Web and Social Media (ICWSM'20)*, volume 14, pages 830–839, Atlanta, GA, USA, 2020. AAAI Press.

[5] P.V. Bindu, P.S. Thilagam, and D. Ahuja. Discovering suspicious behavior in multilayer social networks. *Computers in Human Behavior*, 73:568–582, 2017. Elsevier.

[6] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. IOP Publishing.

[7] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics reports*, 544(1):1–122, 2014. Elsevier.

[8] G. Bonifazi, B. Breve, S. Cirillo, E. Corradini, and L. Virgili. Investigating the COVID-19 vaccine discussions on Twitter through a multilayer network-based approach. *Information Processing & Management*, 59(6):103095, 2022. Elsevier.

[9] G. Bonifazi, F. Cauteruccio, E. Corradini, M. Marchetti, D. Ursino, and L. Virgili. Applying Social Network Analysis to Model and Handle a Cross-Blockchain Ecosystem. *Electronics*, 12(1086), 2023. MDPI.

[10] M. Bujak and R. Kucharski. Network structures of urban ride-pooling problems and their properties. *Social Network Analysis and Mining*, 13(1):89, 2023. Springer.

[11] C. Buntain and J. Golbeck. Identifying Social Roles in Reddit Using Network Structure. In *Proc. of the International Conference on World Wide Web (WWW'14)*, page 615–620, Seoul, Korea, 2014. ACM.

[12] X. Chen and Z.M. Lu. Measure of layer centrality in multilayer network. *International Journal of Modern Physics C*, 29(06):1850051, 2018. World Scientific.

[13] M.M. Danziger, I. Bonamassa, S. Boccaletti, and S. Havlin. Dynamic interdependence and competition in multilayer networks. *Nature Physics*, 15(2):178–185, 2019. Nature Publishing Group UK London.

[14] F. Gamgne Domgue, N. Tsopzé, and R. Ndoundam. Correlation and dimension relevance in multidimensional networks: a systematic taxonomy. *Social Network Analysis and Mining*, 11:1–19, 2021. Springer.

[15] W.B. Du, X.L. Zhou, O. Lordan, Z. Wang, C. Zhao, and Y.-B. Zhu. Analysis of the Chinese Airline Network as multi-layer networks. *Transportation Research Part E: Logistics and Transportation Review*, 89:108–116, 2016. Elsevier.

[16] G. Dudek. Multilayer perceptron for short-term load forecasting: from global to local approach. *Neural Computing and Applications*, 32(8):3695–3707, 2020. Springer.

[17] S. El-Sappagh, J.M. Alonso, S.M. Islam, A.M. Sultan, and K.S. Kwak. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Scientific reports*, 11(1):1–26, 2021. Springer.

[18] D.N. Fisher and N. Pinter-Wollman. Using multilayer network analysis to explore the temporal dynamics of collective behavior. *Current zoology*, 67(1):71–80, 2021. Oxford University Press.

[19] S. Gao and Z. Fang. Multilayer feature-rich satellite network analysis: An application-oriented and time-evolving approach. *IEEE Transactions on Network Science and Engineering*, 8(2):1316–1327, 2020. IEEE.

[20] M. Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*, 2022. arXiv.

[21] Y. Gu, X. Fu, Z. Liu, X. Xu, and A. Chen. Performance of transportation network under perturbations: Reliability, vulnerability, and resilience. *Transportation Research Part E: Logistics and Transportation Review*, 133:101809, 2020. Elsevier.

[22] C. Gunaratne, S.K. Ray, A.C. Lourenço Alves, and M. Waldl. Exogenous Shocks Lead to Increased Responsiveness and Shifts in Sentimental Resilience in Online Discussions. In *Proc. of the International Conference of The Computational Social Science Society of the Americas (CSS'19)*, pages 57–71, Santa Fe, NM, USA, 2021. Springer.

[23] Z. Hammoud and F. Kramer. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Analytics*, 5(1):2, 2020. Springer.

[24] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron, and A. Sala. Feature-rich networks: going beyond complex network topologies. *Applied Network Science*, 4:1–13, 2019. Springer.

[25] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia. Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications*, 132:86–103, 2019. Elsevier.

[26] Q. Li, S. Dong, and A. Mostafavi. Modeling of inter-organizational coordination dynamics in resilience planning of infrastructure systems: A multilayer network simulation framework. *PloS one*, 14(11):e0224522, 2019. Public Library of Science San Francisco.

[27] A.P. Logan, P.M. LaCasse, and B.J. Lunday. Social network analysis of Twitter interactions: a directed multilayer network approach. *Social Network Analysis and Mining*, 13(1):65, 2023. Springer.

[28] F. Long, N. Ning, Y. Zhang, C. Song, P. Zhou, and B. Wu. Mining latent academic social relationships by network fusion of multi-type data. *Social Network Analysis and Mining*, 10:1–16, 2020. Springer.

[29] D. Mohapatra. A hybrid approach for pair-wise layer similarity in a multiplex network. *Social Network Analysis and Mining*, 11(1):88, 2021. Springer.

[30] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, and J. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010. AAAS.

[31] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002. APS.

[32] M.E.J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006. National Academy of Sciences.

[33] E. Oro, C. Pizzuti, N. Procopio, and M. Ruffolo. Detecting topic authoritative social media users: a multilayer network approach. *IEEE Transactions on Multimedia*, 20(5):1195–1208, 2017. IEEE.

[34] F. Pierri, C. Piccardi, and S. Ceri. A multi-layer approach to disinformation detection in US and Italian news spreading on Twitter. *EPJ Data Science*, 9(1):35, 2020. Springer.

[35] M. Rosvall, D. Axelsson, and C. T. Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, 2009. Springer.

[36] L. Scabini, L.C. Ribas, M.B. Neiva, A.G.B. Junior, A.J.F. Farfan, and O.M. Bruno. Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil. *Physica A: Statistical Mechanics and its Applications*, 564:125498, 2021. Elsevier.

[37] J. Shin, L. Jian, K. Driscoll, and F. Bar. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83:278–287, 2018. Elsevier.

[38] G. Stoddard. Popularity dynamics and intrinsic quality in Reddit and hacker news. In *Proc. of the International AAAI Conference on Web and Social Media (ICWSM'15)*, page 9(1), Oxford, UK, 2015.

[39] Q. Su, A. McAvoy, Y. Mori, and J.B. Plotkin. Evolution of prosocial behaviours in multilayer populations. *Nature Human Behaviour*, 6(3):338–348, 2022. Nature Publishing Group UK London.

[40] F. Touvet and D. Harle. Network resilience in multilayer networks: A critical review and open issues. In *Proc. of the International Conference on Networking (Networking—ICN 2001)*, pages 829–837, Colmar, France, 2001. Springer.

[41] K. S. Trivedi, D. S. Kim, and R. Ghosh. Resilience in computer systems and networks. In *Proc. of the International Conference on Computer-Aided Design (ICCAD'09)*, pages 74–77, San Jose, CA, USA, 2009. ACM.

[42] M. Tsvetovat and A. Kouznetsov. *Social Network Analysis for Startups: Finding connections on the social web.* Sebastopol, CA, USA, 2011. O'Reilly Media, Inc.

[43] M. Vaiana and S.F. Muldoon. Multilayer brain networks. *Journal of Nonlinear Science*, 30(5):2147–2169, 2020. Springer.

[44] H. Zhang, C.D. Wang, J.H. Lai, and P.S. Yu. Modularity in complex multilayer networks with multiple aspects: a static perspective. *Applied Informatics*, 4(1):1–29, 2017. Springer.