



Community Detection in Weighted Time-Variant Social Network

Himansu Sekhar Pattanayak
 himansu.sekhar@mail.jiit.ac.in
 Jaypee Institute of Information
 Technology
 Noida, U.P., India

Bhawna Saxena
 bhawna.saxena@jiit.ac.in
 Jaypee Institute of Information
 Technology
 Noida, U.P., India

Adwitiya Sinha
 mailto:adwitiya@gmail.com
 Jaypee Institute of Information
 Technology
 Noida, U.P., India

ABSTRACT

Community detection in time-variant networks is a challenging task. Owing to the dynamic nature of the interactions between the participant entities, the network structure tends to evolve continuously. There are two different approaches for community detection in time-variant networks, namely snapshot-based and evolutionary-based. The former technique is prone to instabilities, while the latter is expensive. This work investigates the aggregated snapshot-based technique based on a high school interaction data set. The whole time-stamped networked data is partitioned into five temporal instances. We have extended the Louvain community detection algorithm for weighted time-variant networks. On analyzing the data collected from community detection, we found that each individual instance has similar characteristics as compared to the whole network for the detected communities.

CCS CONCEPTS

- Information systems → Social networking sites; • Mathematics of computing → Graph algorithms.

KEYWORDS

Social network analysis, community detection, time-variant networks, social physical network

ACM Reference Format:

Himansu Sekhar Pattanayak, Bhawna Saxena, and Adwitiya Sinha. 2023. Community Detection in Weighted Time-Variant Social Network. In *2023 Fifteenth International Conference on Contemporary Computing (IC3-2023) (IC3 2023), August 03–05, 2023, Noida, India*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3607947.3608077>

1 INTRODUCTION

The graph-theoretic model of a social network consists of the participants that are represented as nodes, and the interactions between them are considered as network edges. A graph $G(V, E)$ is defined as a social network consisting of N nodes or entities, and V refers to the set of N nodes and E is a set of edges. A time-variant network is defined as $G_t = (V, E_t)$, where G_t is the graph structure at a particular-time step t , $E_t \subseteq V \times V$ the edges of the graph change with each time step t . In the time variant network, the graph G

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IC3 2023, August 03–05, 2023, Noida, India

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0022-4/23/08...\$15.00

<https://doi.org/10.1145/3607947.3608077>

is represented as the set of snapshots, i.e., the set of network interactions recorded at different time steps t spanning from 1 to T . $G = G_1(V, E_1), G_2(V, E_2) \dots G_t(V, E_t), \dots G_T(V, E_T)$. The adjacency matrix A_t is defined as the interactions at the particular time-step t , $A = \text{agg}\{A_1, A_2 \dots A_T\}$. $A_t[i][j] = 1$ if there is interaction between node $v[i]$ and $v[j]$ at time-step t , otherwise 0.

One of the most important areas of network analysis study is the detection of inherent communities in the underlying time-varying connected structure [1]. This process is also called temporal community detection, which exhibits similar structural patterns evolving dynamically over time in a social network. In such networks, community detection requires time-aware diffusion of information, thus maximizing content outreach [13, 22]. Modularity is considered a popular method for detecting communities that can be extended to time-varying networks by incorporating temporal connectivity information. This technique focuses on maximizing the modularity of the network while considering the network dynamics. There are several dynamic clustering methods that explicitly consider the temporal dimension of the network to cluster nodes having similar connectivity patterns. An instance of a dynamic clustering algorithm is the Louvain method which is extensible for temporal networks. Other algorithms include Infomap, Temporal Walklets, Graph2Vec, dynamic stochastic block models, etc.

It divides the whole network into functional parts solely based on the network structure. These functional units are strongly linked, yet they have fewer connections to the outer nodes. The whole network G is partitioned into l communities as $G = C^1 \cup C^2 \cup C^3 \dots \cup C^l$, where C^i is the i^{th} community. If $C^i \cap C^j = \emptyset$, then the communities are non-overlapping, otherwise overlapping. However, in a time-variant network G_t , the network is partitioned as: $G_t = C_t^1 \cup C_t^2 \cup C_t^3 \dots \cup C_t^l$ at a time-step t . The construction of aggregated networks is performed by the snapshot technique. This technique is a series of static networks at different points in time scale, which are aggregated into a single network with possible interconnections, thereby enabling the discovery of inherent communities in a time-variant network. There are several static and dynamic community identification algorithms that are used to sequence the multiple time-variant networks, thus leading to the formation of an aggregated network.

In this work, we are comparing the time-variant network instances of a social physical network with its aggregated structure by using community detection. The social physical network depicts the interactions of students for five days in the real-physical world. We compare all the network instances in terms of different network statistics. It was found that the network instances are of equivalent sizes and have nearly identical communities, and most of the entities participated in interactions in each of the networks. We also aim

to examine the correlation amongst different network properties, based on the interactions from all the time-variant networks.

2 RELATED WORKS

There are several research contributions for detecting communities in social networks with time-varying subnets. As a significant contribution, some authors have provided a snapshot approach to find the community structure [3, 20]. In a time-variant network, the snapshot approach of clustering is straightforward to implement. Comparing communities at various points in time, on the other hand, is complex and subject to frequent changes in community organization across time. Because static community identifications must often be performed, the computational cost rises; this approach entirely disregards the information gleaned in earlier stages. The authors have computed the similarity between the communities discovered in various time steps and summarized the results in the form of comparative analysis [2].

In contrast to the snapshot-based technique, the evolutionary approach takes into consideration the community structure in prior time steps [4, 5, 7, 11]. The number of modifications to the community structure is limited to a minimal level. The homogeneity of the identified community may cause the partition's accuracy to deteriorate. Further, for time-variant networks, the stochastic block model (SBM) is a suitable approach. The transition probability paradigm is helpful to characterize the relationship between the community structure in a succession of time-steps[8]. The random-walk approach is also commonly used for community discovery, starting at any node and moving to the neighboring nodes at random. Long-distance random walk tends to trap inside a community. The likelihood of the nodes being visited by the random walk is used to calculate their similarity. Similar nodes group together to form communities in a network [19]. However, the length of a random walk needs to be chosen carefully.

The existing literature has deployed Susceptible- Infectious- Recovered (SIR) model the spreading dynamics in recent time-variant community detection and reveal the possible networked structures [10, 12, 14, 16, 23, 24]. In each time step, an infectious disease spreads from a single node to adjacent nodes. Nodes that have not yet been affected are susceptible to infection and may be infected at a future stage. The infected nodes are recovered with a probability. A node that has been recovered cannot be infected again. The recovered node is no longer involved in the spreading process. Susceptibility, infection, and recovery are all defined in terms of probability. Nodes with comparable recovery probability are clustered together for partitioning by using any of the available clustering techniques. The downside of this approach is that it is difficult to determine the duration of the spreading process. Another significant research revolves around the computation of link-threshold value to decide the evolving communities. The maximum diameter of any community is predicted using a probability-based approach [18], which would restrict the information flow process and lowers the computing cost. The information reach probability of nodes emanating from numerous source nodes is compared to determine their similarity. The structure of the network is deduced from the similarity data using a clustering approach.

In a relevant study conducted in [9], the authors have provided an extensive assessment of synchronisation during time-varying network formation. The study proposed two paradigmatic frameworks which includes networked systems with structural dynamics due to temporal and spatial changes. The first framework involves those systems where the connections between nodes are time-dependent due to adaptation, outside influence, or any other process impacting each link or interaction. The second framework, caters to the change in structural scenario of the graph due to spatial movement of nodes, thus causing connections to be continuously turned on and off over time. Another work involves probabilistic generative modeling that considers hidden variables to incorporate communities and reciprocity as structural network information changing over time, thereby forming a time-variant social network [21]. In witnessing reciprocal data, the model presupposes a fundamental order in which an edge is noticed in relation to its reciprocated edge in the past. The transition matrices of the social network from several timestamps are built using a Markovian technique, and parameters are inferred using an expectation-maximization algorithm that takes advantage of the dataset's sparsity to achieve high computational efficiency. The authors have validated the model using citation networks, email datasets, and synthetic dynamical networks.

In another recent study, some authors intend to identify the evolution characteristics of product structural qualities based on the time-dependent network model in order to provide product development direction for designers [15]. The cross-generation product structures are represented by weighted time-dependent networks, and topological similarities linking product structures with complex networks are used to analyze complexity, significance, and robustness. The results reveal that when product structures evolve, the model complexity decreases while key components get more complicated due to the scale-free feature. In general, numerous solutions exist for addressing the problem of community detection, but the majority of them are constrained to static circumstances. Such systems need to be extended to tailor the dynamic social networks. To address this issue, the Actor-Critical for Community Detection (AC2CD) architecture is provided in [6] for social networks evolving dynamically with time. The authors have employed local optimization of the modularity density function, and the deep reinforcement learning technique is applied to the architecture for handling the changing characteristics of vast networks.

3 METHODOLOGY

This section describes the methodology followed for detecting communities in weighted time-variant social physical networks. The complete outline of the workflow is presented in Figure 1.

- **Dataset Collection:** The dataset used for the presented work represents the "contact network of students of a high school in Marseilles, France [17]". The dataset contains the contact information of the students from real-physical world, as recorded during 5 days in December 2013. The data was captured and quantified using sensors embedded in wearable badges, which exchange radio packets to detect the proximity of students. Two students are considered to be in contact

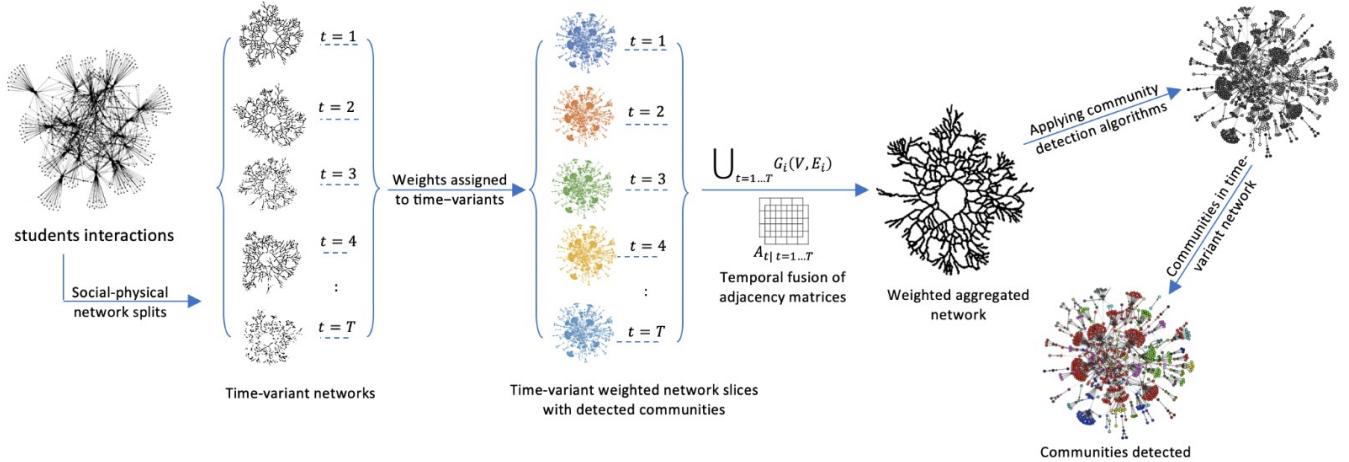


Figure 1: Outline of the workflow

if, during a 20-second time interval, the sensors in their wearable badges exchange at least one packet, and if no packet gets exchanged over a 20-second interval, then the contact is considered to be over.

- **Network Generation:** Firstly, five time-variant network instances were generated by dividing the timestamped dataset into five-time slices, wherein each time slice corresponded to one day. Each of the five time-variant networks depicts the interactions carried out by the entities (nodes) during the respective time slice. Then, for each of these time-variant network instances, edge weights were computed in terms of the number of interactions carried out between each pair of nodes. Thereafter, the five weighted time-variant network instances were combined into a single weighted aggregated interaction network. Hence, a total of six networks were generated from the collected dataset, as mentioned above.
- **Community Detection:** Weighted community detection using the Louvain method was performed on each of the five weighted time-variant network instances as well as the combined weighted aggregated interaction network. Louvain is a widely used community detection algorithm that utilizes the concept of graph density. It comprises two phases: Modularity Optimization (MO) and Community Aggregation (CA). Under MO, a unique community is assigned to each node in the graph. Thereafter, each node is designated a different community and the graph modularity is recalculated. This is repeated for multiple iterations till no significant increment in modularity is recorded. Thereafter, in CA, the nodes belonging to the same community are merged together, say into one big node. All big nodes are connected through links that are the summation of the links between the nodes from the corresponding unmerged communities. Self-loops on the big nodes represent the summation of links present within a community, before getting merged. Both these phases get

executed until no further gain is observed and maximum modularity is achieved.

- **Result Analysis:** Finally, statistical analysis of the obtained results was done.

4 RESULTS AND DISCUSSION

4.1 Experimental Set-up

As mentioned in the above section, data pertaining to the contact network of students of a high school was used in the presented work. There were a total of 327 participants in the student contact network. The collected data (high school student interaction network) was divided into five time-variant interaction network instances such that each network instance represented the interaction network for one day. The edge weights were computed in terms of the frequency of interaction between the two entities for the particular instance of the network. Louvain method was then applied for detecting communities in these weighted network instances. Thereafter, the Louvain community detection method was applied on the weighted aggregated interaction network, formed by combining the five weighted time-variant network instances. Gephi 0.9.2 was used for network analysis and visualization. Furthermore, network statistics like network diameter, graph density, modularity, average degree, average weighted degree, average clustering coefficient, and average path length were also computed for all six time-variant networks.

4.2 Results

Community detection in time-variant networks is usually done using snapshot-based or evolutionary-based approaches. The snapshot-based approach may not fully reveal the network structure due to limited data and evolutionary approaches are relatively expensive as they consider the whole data along with the time-variant network dynamics. On the contrary, an instance-based approach will consider only a part of the whole data and can be considered as an aggregation of snapshots. Moreover, the data collection duration

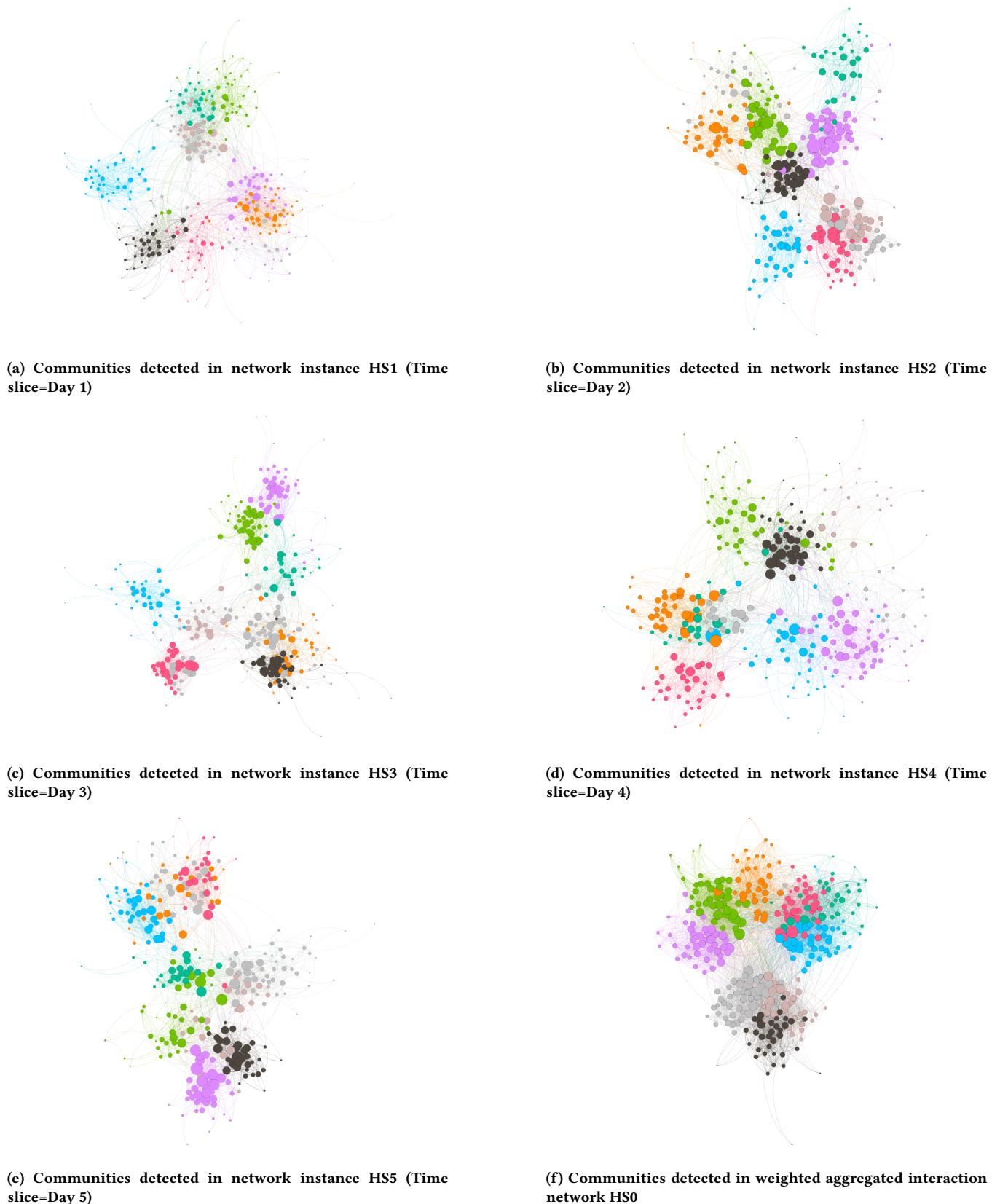


Figure 2: Time-variant communities detected using Louvain method

Table 1: Network properties of the generated networks

Network instance	Number of communities	Average degree	Average weighted degree	Network diameter	Density	Modularity	Average clustering coefficient	Average path length
HS1	12	14.683	236.539	6	0.046	0.825	0.419	2.904
HS2	12	16.29	243.329	6	0.053	0.813	0.412	2.78
HS3	13	13.184	244.052	8	0.043	0.841	0.444	3.23
HS4	11	14.425	246.497	5	0.047	0.807	0.391	2.834
HS5	14	14.92	246.039	6	0.049	0.831	0.433	2.947
HS0	11	35.584	1192.954	4	0.109	0.812	0.504	2.159

Table 2: Correlation coefficient values between network properties

	Number of communities	Density	Network diameter	Modularity	Average clustering coefficient	Average path length
Number of communities	-	-0.0708	0.5204	0.7893	0.8569	0.5166
Modularity	0.7893	-0.5869	0.8616	-	0.9709	0.8894
Average path length	0.5166	-0.7956	0.8967	0.8894	0.8052	-

can be shortened, thereby making the process comparatively less expensive than the evolutionary approach. Figures 2(a-e) depict the visual representation of the communities detected in the five weighted time-variant network instances (labeled HS1 to HS5) and figure 2f depicts the visual representation of the detected communities in the weighted aggregated interaction network labeled as HS0.

Network properties like modularity, density, network diameter, average path length, etc., were computed for the generated time-variant networks. Table 1 shows the values obtained for all six networks. HS1, HS2, HS3, HS4, and HS5 are the five weighted time-variant network instances and HS0 is the weighted aggregated interaction network. The number of nodes in each of the six networks has similar values except for HS1 having 319 nodes and HS0 having 327 nodes. Out of these 327 participants, the majority of the nodes must have interacted in all the network instances, as the minimum number of nodes is 306 in the case of HS4. Though the number of edges varies from 2207 to 2525, the variation is not very significant, thereby indicating that the communication pattern of the participants doesn't vary much across the five time-variant networks. Except for HS0, which is the aggregated network, the density, average degree, and average weighted degree of each of the five time-variant network instances is also found to be similar. The number of detected communities varies from 11 to 14 across HS1 to HS5, and HS0 has 11 communities. It can further be observed that the community structure of HS1 to HS5 is nearly similar with modularity value ranges between 0.807 to 0.841. This suggests that even though they are time-variant networks, the network structure is not significantly altered in this type of interaction network. Hence, community detection on a single instance of the time-variant network can help reveal the structure of the aggregated network.

Furthermore, if we focus on the number of communities, modularity, clustering coefficient, network diameter, and average path

length, we can see a correlation. As the network diameter value is decreased, the number of detected communities also decreases. Similarly, along with the decrease in modularity and clustering coefficient values, there is a decrease in the number of detected communities. Average path length signifies network compactness, and with an increase in the number of detected communities, the average path length also increased. Table 2 presents the correlation between the different network properties. Data from HS1, HS2, HS3, HS4, and HS5 are only considered for calculating the correlation coefficients. These networks have identical number of nodes, edges, graph density and average node degrees. HS0 is not considered for the correlation analysis as it differs from the other networks with regard to graph density, number of edges, and average node degrees. As can be seen from Table 2, the number of communities and graph density are not correlated. The number of communities is moderately positively correlated with network diameters and average path length. The number of communities is strongly positively correlated with the average clustering coefficient and modularity. Modularity is strongly positively correlated with the number of communities, network diameter, average clustering coefficient, and average path length. However, it is moderately negatively correlated with graph density. The average path length is strongly negatively correlated with graph density, but strongly positively correlated with network diameter and average clustering coefficient.

5 CONCLUSION

Community detection in the time-variant social physical network is performed using snapshot based and evolutionary-based. In this work, an aggregated snapshot-based technique is used for community detection. The high school interaction of students from the physical world is taken to build a social-physical network. It is used to analyze by creating five weighted temporal network instances. The Louvain community detection method is used on all

five networks and the whole network by extending it to a weighted version. Nearly all the entities were found to participate in the interactions during all five days. The number of detected communities is found to be similar, with slight variance along with the number of communities detected from the whole network. Therefore, the aggregated snapshot-based approach may be a faster and lesser expensive alternative for the time-variant interaction networks. We also conducted a correlation analysis of different attributes of the collected data. We found that an increase in the number of communities results in an increased modularity score, average path length, average clustering coefficient, and network diameter, provided that the networks have nearly equal numbers of nodes, edges, and graph density.

REFERENCES

- [1] Norah Alotaibi and Delel Rhouma. 2022. A review on community structures detection in time evolving social networks. *Journal of King Saud University - Computer and Information Sciences* 34, 8 (2022), 5646–5662. <https://doi.org/10.1016/j.jksuci.2021.08.016>
- [2] Alessia Amelio and Clara Pizzuti. 2017. Correction for closeness: Adjusting normalized mutual information measure for clustering comparison. *Computational Intelligence* 33, 3 (2017), 579–601. <https://doi.org/10.1111/coin.12100>
- [3] Piotr Bródka, Stanisław Saganowski, and Przemysław Kazienko. 2013. GED: the method for group evolution discovery in social networks. *Social Network Analysis and Mining* 3, 1 (2013), 1–14. <https://doi.org/10.1007/s13278-012-0058-8>
- [4] Deepayan Chakrabarti, Ravi Kumar, and Andrew Tomkins. 2006. Evolutionary clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 554–560. <https://doi.org/10.1145/1150402.1150467>
- [5] Yun Chi, Xiaodan Song, Dengyong Zhou, Koji Hino, and Belle L. Tseng. 2007. Evolutionary spectral clustering by incorporating temporal smoothness. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 153–162. <https://doi.org/10.1145/1281192.1281212>
- [6] Aurélio Ribeiro Costa and Célia Ghedini Falha. 2023. AC2CD: An actor-critic architecture for community detection in dynamic social networks. *Knowledge-Based Systems* 261 (2023), 110202. <https://doi.org/10.1016/j.knosys.2022.110202>
- [7] Harry Crane and Walter Dempsey. 2015. Community detection for interaction networks. *CoRR* abs/1509.09254, 3 (2015). arXiv:1509.09254 <http://arxiv.org/abs/1509.09254>
- [8] Amir Ghasemian, Pan Zhang, Aaron Clauset, Christopher Moore, , and Leto Peel. 2016. Detectability Thresholds and Optimal Algorithms for Community Structure in Dynamic Networks. *Physical Review X* 6, 3 (2016), 031005. <https://doi.org/10.1103/PhysRevX.6.031005>
- [9] Dibakar Ghosh, Mattia Frasca, Alessandro Rizzo, Soumen Majhi, Sarbendu Rakshit, Karin Alfaro-Bittner, and Stefano Boccaletti. 2022. The synchronized dynamics of time-varying networks. *Physics Reports* 949 (2022), 1–63. <https://doi.org/10.1016/j.physrep.2021.10.006>
- [10] Clara Granell, Sergio Gómez, and Alex Arenas. 2013. Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks. *Physical Review Letters* 111 (2013), 128701. <https://doi.org/10.1103/PhysRevLett.111.128701>
- [11] Robert Görke, Pascal Maillard, Andrea Schumm, Christian Staudt, and Dorothea Wagner. 2013. Dynamic graph clustering combining modularity and smoothness. *ACM Journal of Experimental Algorithms* 18 (2013), 1.1–1.29. <https://doi.org/10.1145/2444016.2444021>
- [12] Kazuki Kuga Jun Tanimoto, and Marko Jusup. 2019. To vaccinate or not to vaccinate: A comprehensive study of vaccination-subsidizing policies with multi-agent simulations and mean-field modeling. *Journal of Theoretical Biology* 469 (2019), 107–126. <https://doi.org/10.1016/j.jtbi.2019.02.013>
- [13] Aditya Lahiri, Yash Kumar Singhal, and Adwitiya Sinha. 2021. TODD: Time-aware Opinion Dynamics Diffusion Model for Online Social Networks. In *Proceedings of International Conference on Artificial Intelligence and Applications (Advances in Intelligent Systems and Computing, Vol. 1164)*, P. Bansal, M. Tushir, V. Balas, and R. Srivastava (Eds.). Springer, 235–245. https://doi.org/10.1007/978-981-15-4992-2_23
- [14] Hui-Jia Li and Lin Wang. 2019. Multi-scale asynchronous belief percolation model on multiplex networks. *New Journal of Physics* 21 (2019), 015005. <https://doi.org/10.1088/1367-2630/aaf775>
- [15] Yupeng Li, Yongbo Ni, Na Zhang, Qinming Liu, and Jin Cao. 2022. Towards the evolution characteristics of product structural properties based on the time-dependent network. *Journal of Engineering Design* 33, 3 (2022), 207–233. <https://doi.org/10.1080/09544828.2022.2030683>
- [16] Yang Liu, Yong Deng, Marko Jusup, , and Zhen Wang. 2016. A biologically inspired immunization strategy for network epidemiology. *Journal of Theoretical Biology* 400 (2016), 92–102. <https://doi.org/10.1016/j.jtbi.2016.04.018>
- [17] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. 2015. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. *PLOS One* 10, 9 (2015), e0136497. <https://doi.org/10.1371/journal.pone.0136497>
- [18] Himansu Sekhar Pattanayak, Harsh K. Verma, and Amrit Lal Sangal. 2022. Gravitational community detection by predicting diameter. *Discrete Mathematics, Algorithms and Applications* 14, 4 (2022), 2150145. <https://doi.org/10.1142/S1793830921501457>
- [19] Pascal Pons and Matthieu Latapy. 2005. Computing Communities in Large Networks Using Random Walks. In *International Symposium on Computer and Information Sciences - ISCIS 2005 (Lecture Notes in Computer Science, Vol. 3733)*, P. Yolum, T. Güngör, F. Gürgen, and C. Özturen (Eds.). Springer, 284–293. https://doi.org/10.1007/11569596_31
- [20] Martin Rosvall and Carl Bergstrom. 2010. Mapping change in large networks. *PLOS One* 5, 1 (2010), e8694. <https://doi.org/10.1371/journal.pone.0008694>
- [21] Hadiseh Saefi, Martina Contisciani, and Caterina De Bacco. 2021. Reciprocity, community detection, and link prediction in dynamic networks. *Journal of Physics: Complexity* 3, 1 (2021), 015010. <https://doi.org/10.1088/2632-072X/ac52e6/pdf>
- [22] Bhawna Saxena and Vikas Saxena. 2022. Hurst exponent based approach for influence maximization in social networks. *Journal of King Saud University - Computer and Information Sciences* 34, 5 (2022), 2218–2230. <https://doi.org/10.1016/j.jksuci.2019.12.010>
- [23] Zhen Wang, Michael A. Andrews, Zhi-Xi Wu, Lin Wang, and Chris T. Bauch. 2015. Coupled disease-behavior dynamics on complex networks: A review. *Physics of Life Reviews* 15 (2015), 1–29. <https://doi.org/10.1016/j.plrev.2015.07.006>
- [24] Peican Zhu, Xiangfeng Dai, Xuelong Li, Chao Gao, Marko Jusup, and Zhen Wang. 2019. Community detection in temporal networks via a spreading process. *Europhysics Letters* 126, 4 (2019), 48001. <https://doi.org/10.1209/0295-5075/126/48001>