

# Towards Personalized Wellness: Detection and Analysis of Homogeneous Communities in Multi-Modal Biomedical Networks

Rama Krishna Thelagathoti and Hesham H. Ali  
*College of Information Science and Technology*  
*University of Nebraska Omaha*  
 Omaha, NE 68182, USA  
 email: rthelagathoti, hali@unomaha.edu

**Abstract**—In recent years, personalized wellness has attracted significant attention by many research groups. However, distinguishing between various groups associated with different health conditions, which is critical to achieve personalized healthcare, remains a complex challenge. The driving force behind this study is to harness the potential of biomedical networks to reveal homogenous groups with a high degree of similarity. Our primary goal is to uncover hidden relationships that may not be evident using traditional data analysis by discerning interaction patterns and similarities between nodes in the network. To construct a biomedical graph, we integrated mobility data and medical data, mainly Heart Rate Variability (HRV) data, for each subject, representing them as nodes. The edges between nodes reflect the similarity between subjects in both mobility patterns and HRV responses. Leveraging advanced network analysis techniques, we applied an appropriate network features to identify strongly connected homogenous groups within the graph. These groups represent subsets of subjects displaying highly similar mobility characteristics and HRV patterns. Our analysis identified five highly similar groups within the biomedical network based on the subjects mobility and HRV data. Remarkably, mobility features played a more pronounced role in the formation of these groups compared to HRV data, suggesting strong ties and similarities in mobility characteristics. These findings significantly contribute to our understanding of the interplay between movement behaviors, offering promising avenues for personalized interventions and enhanced clinical decision-making in the pursuit of advanced healthcare and improved well-being.

**Index Terms**—Homogeneous groups; Network Analysis; Mobility Analysis; Personalized interventions.

## I. INTRODUCTION

Understanding homogeneous groups within biological networks is essential for advancing personalized therapies, designing targeted interventions, and developing individualized medicine [1] [2]. These homogeneous groups represent subsets of interconnected individuals or elements within a network who share similar characteristics, behaviors, or functional relationships [3] [4]. In other words, they are said to be cohesive in nature within the biomedical network. By analyzing the structure and dynamics of these groups, we can gain valuable insights into the complex systems of biological networks, uncover hidden patterns, and identify key factors

driving their formation and functioning [5]. Cohesive groups play a pivotal role in personalized medicine and healthcare [6]. They offer a means to stratify patients into subgroups based on shared characteristics, enabling the development of tailored treatments that address the specific needs and responses of individuals within each group. Moreover, homogeneous groups can provide insights into disease subtypes, treatment response prediction, and the identification of novel therapeutic targets [6].

This study incorporates data from two modalities: mobility data collected from physical activity and medical data collected from HRV. These modalities are utilized to construct the biomedical network. Mobility data, including step counts, distance traveled, and activity patterns, offer insights into individuals' daily routines, physical activity levels, and overall mobility behaviors [7] [8]. HRV data, on the other hand, captures the variability in the time intervals between successive heartbeats, reflecting autonomic nervous system activity, physiological stress, and overall cardiovascular health [9]. The integration of mobility and HRV data allows for a comprehensive assessment of individuals' behavioral and physiological characteristics within homogeneous groups. By considering both modalities, we can gain a deeper understanding of the interplay between physical activity, physiological responses, and health outcomes. This holistic approach facilitates the identification of homogeneous groups that share not only mobility patterns but also physiological similarities, enhancing our ability to personalize interventions and treatments [10].

In this study, our primary objectives are three-fold. Firstly, we aim to construct a network representation that integrates mobility and HRV data. Secondly, we seek to apply an appropriate network threshold that defines the strength of connections within the graph. By adjusting the threshold, we can identify strongly connected homogeneous groups that exhibit distinct patterns and interrelationships. Lastly, we aim to analyze the mobility patterns and HRV characteristics within these groups. The rest of the document is organized as follows. Section II presents the detailed methodology while Section III contains the obtained results. Discussion of the

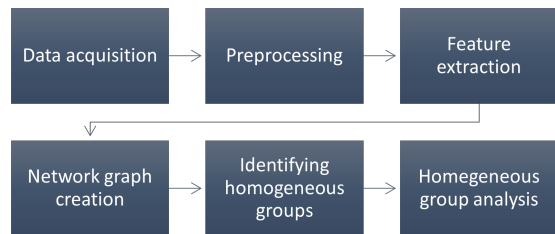


Fig. 1. The overview of the methodology.

results is presented in Section IV and Section V lists some of the limitations and future work of the study.

## II. METHOD

### A. Overview

Our study follows a five-step approach, as shown in Figure 1, to identify homogeneous groups in a biomedical network based on mobility and HRV data. We acquired multimodal data for the subjects, performed data preprocessing to clean and prepare the data, and extracted relevant features. Using these features, we constructed the network, representing subjects as nodes with edges reflecting their similarities. By applying a network threshold, we identified homogeneous groups, revealing strong ties based on mobility characteristics. The analysis emphasized the influential role of mobility features in the formation of homogeneous groups, providing valuable insights for personalized healthcare interventions.

### B. Data Acquisition

The dataset utilized in this study is the HYPERAKTIV dataset, which is publicly available and can be accessed from a public repository [11]. The dataset comprises mobility data collected using wearable sensors, comprehensive diagnostic assessments, and demographic characteristics. It consists of two primary categories: individuals with Attention-Deficit/Hyperactivity Disorder (ADHD) and individuals without ADHD. But most of them were also diagnosed with other psychiatric disorders such as ADD (Attention Deficit Disorder), unipolar depression, bipolar depression, and anxiety. However, the assessment of these psychiatric disorders was made using self-reported feedback. Furthermore, the dataset includes a total of 85 patients, with 45 diagnosed with ADHD and 40 were diagnosed with other disorders in addition to ADHD. The mobility data in the dataset is captured through a wrist-worn actigraph device (Actiwatch), with a sampling frequency of 32 Hz. The motor activity data includes measurements of movement intensity in the x, y, and z-axes. In addition to the activity data, the dataset also contains HRV data. HRV data is ECG-based and recorded using a small chest-worn battery-driven device Actiheart (Cambridge Neurotechnology Ltd, England), allowing free movement and long recordings.

### C. Preprocessing and Feature Extraction

The collected mobility data from the Actiwatch device was processed and organized into individual files for each participant. Each file contained raw activity measurements

recorded over a span of 6 days for the ADHD group and 7 days for the non-ADHD group. However, for consistency, we considered 6 days of mobility data for all participants. The combined mobility data files were carefully examined for missing values or anomalies, and appropriate techniques were employed to handle them effectively. To gain insights into the participants' motor activity patterns, two types of features were extracted: hour-wise features and day-wise features. These features provide valuable information about the characteristics and variations in motor activity over different time intervals. The hour-wise features reveal intra-daily variations, capturing details such as activity levels during specific hours, peak activity hours, and temporal variability. On the other hand, the day-wise features capture longer-term patterns, trends, and summary statistics of activity levels over the 6-day duration.

In addition to features extracted from mobility data, we have also extracted time domain features from HRV data such as mean and median. Table 3 summarizes the list of features extracted, including mobility and HRV features. This comprehensive set of features offers a detailed representation of the participants' activity patterns at various temporal scales. They serve as a foundation for analyzing mobility patterns, exploring relationships with demographic parameters, and investigating associations with diagnostic assessment outcomes. The extracted hour-wise and day-wise features provide a robust and comprehensive representation of the participants' activity patterns, enabling in-depth analyses of their mobility behaviors.

### D. Network Generation

In our study, the generation of the correlation network involves a series of steps to establish relationships and interconnections between the participants based on their mobility data. To achieve this, we employ a population analysis approach, without relying on supervised label information, to identify groups of persons with similar motor activity profiles. First, we denote all 55 participants as nodes in the correlation graph, representing each subject within the network. The objective is to connect two participants with an edge if they possess a similar motor activity profile. The similarity between subjects is quantified using the 48 hour-wise features extracted during the feature extraction phase. Specifically, we compute the pairwise Pearson correlation coefficient ( $\rho$ ) between each pair of participants, which measures the linear dependence between them [12] [13].

To construct the correlation graph, we utilize the computed  $\rho$  values to create a correlation matrix that captures the degree of similarity between all pairs of users. The  $\rho$  value ranges from -1 to +1, where -1 indicates a negative correlation and +1 represents a strong positive correlation [14] [15]. Subsequently, we introduce a predefined threshold, denoted as 'k,' to identify strongly correlated pairs, thus forming a significance matrix. The significance matrix takes a value of 1 if the ( $\rho$  value between a pair of users is greater than or equal to the threshold 'k'; otherwise, it takes a value of 0. The significance matrix serves as the adjacency matrix for the

TABLE I: FEATURE LIST

Feature set	Feature	Count	Description
Hour-wise Features	h-m-0 to h-m-23	24	Mean (average) of activity measured by every hour (0 – 23 hours)
	h-sd-0 to h-sd-23	24	Standard deviation (SD) of activity measured by every hour (0 – 23 hours)
Day-wise Features	d-m-1 to d-m-6	6	Mean (average) of activity measured by a day (1 – 6 days)
	d-sd-1 to d-sd-6	6	Standard deviation (SD) of activity measured by a day (1 – 6 days)
HRV features	hrv1 to hrv5	5	Mean, Median, SD, Min, and Max of HRV
	Total	65	

correlation graph. In this graph, two participants ( $P_i, P_j$ ) are connected by an edge if their correlation coefficient exceeds or equals 'k.' This process creates a network of strongly interconnected nodes, representing the relationships between participants based on their motor activity.

To uncover the hidden clusters or communities within the correlation graph, we apply the Louvain Clustering technique, a popular unsupervised clustering algorithm suitable for extracting clusters in biological networks [16]. The Louvain algorithm employs a random walk property to categorize nodes into different communities. These clusters represent groups of persons with similar motor activity profiles. The clustering process ensures that each community demonstrates homogeneity, with participants within the same cluster showing similar motor activity characteristics. Additionally, the clusters exhibit separation, meaning that persons in different communities exhibit distinct mobility profiles [17]. In summary, the generation of the correlation network involves computing the pair-wise Pearson correlation coefficients between subjects' motor activity data, setting a threshold 'k' to identify significant correlations, and constructing the correlation graph based on the adjacency matrix. The application of the Louvain algorithm reveals the hidden clusters representing groups of individuals with similar motor activity profiles. The resulting network provides valuable insights into the relationships and homogeneous patterns among subjects based on their mobility data.

#### E. Detection and Analysis of Homogeneous Groups

Following the generation of the correlation network, the next step is to identify homogeneous groups within the network and analyze their mobility and HRV data. To achieve this, we perform a two-step process: first, determining a correlation threshold that results in the complete disintegration of the network, and second, tweaking the network threshold to create sub-networks with cohesive connections. We initially explore the correlation threshold at which the network becomes en-

tirely disconnected. This threshold serves as an upper limit for the subsequent network threshold adjustments. By carefully selecting a value below this disintegration threshold, we create a sub-network in which participants retain cohesive connections, thus ensuring the identification of meaningful homogeneous groups. Next, we employ the Louvain clustering algorithm on the sub-network to detect homogeneous groups of interconnected nodes. The Louvain algorithm is an efficient and widely used community detection method that optimizes the modularity of the network, effectively partitioning nodes into non-overlapping communities based on their interconnections.

Once the homogeneous groups are identified, we proceed to analyze the mobility and HRV data for each group separately. This analysis involves examining the mobility characteristics, such as average steps taken, distance traveled, and activity patterns, as well as HRV metrics, including time domain, frequency domain, and non-linear features, extracted during the feature extraction phase. Comparing the mobility and HRV data within each homogeneous group allows us to gain insights into their shared characteristics and behavior patterns. By understanding how these groups differ in their motor activity and physiological responses, we can uncover unique features that may distinguish them from other groups and potentially indicate underlying health conditions or specific behavioral traits. Furthermore, we investigate whether there are significant correlations or associations between mobility and HRV data within each homogeneous group. Such correlations could offer valuable information on how motor activity influences autonomic nervous system activity and vice versa, contributing to a deeper understanding of the interplay between physiological and behavioral aspects.

### III. RESULTS

After preprocessing the initial dataset of 85 subjects, 33 subjects were excluded due to insufficient clinical and mobility data, leaving a final dataset of 52 subjects for analysis. To construct a meaningful network, we selected a correlation threshold of 0.5, resulting in a network with a reasonable level of connectivity among the subjects. However, when the threshold was increased to 0.82, the network disintegrated completely. At the initial threshold of 0.5, all subjects were connected in the network, forming the first network as shown in Figure 2(a). Subsequently, we chose a threshold of 0.8, just before the network disintegrated, to create a second network as shown in Figure 2(b). The hypothesis was that the groups identified at this resolution would represent strongly connected groups compared to other nodes. These groups demonstrated resilience even at the maximum threshold resolution, indicating significantly stronger connectivity compared to other nodes. In contrast, other nodes appeared to be either loosely or moderately connected. The groups that proved to be strongly connected at the 0.8 network resolution were identified as homogeneous groups. Their robust interconnections suggested shared behavior and characteristics within these groups, making them distinguishable from other nodes. These homogeneous groups are of particular interest as they potentially

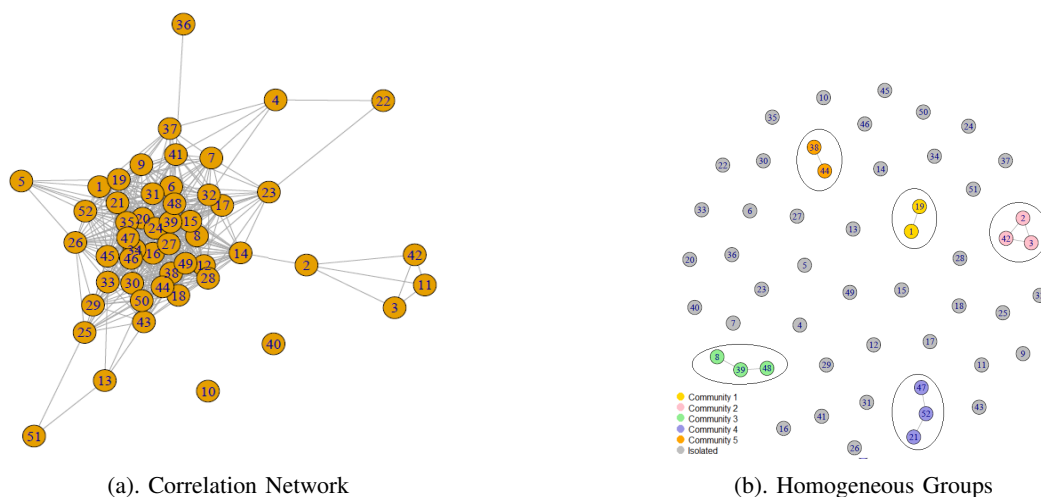


Fig. 2. Identification of homogenous groups

represent distinct subpopulations with unique mobility patterns and HRV responses. Analyzing the mobility patterns and HRV data within each homogeneous group revealed rich and vital information. These insights offer valuable opportunities for designing personalized interventions and tailoring healthcare strategies based on the distinct characteristics exhibited by each homogeneous group. Understanding the interplay between mobility and HRV data within these groups can lead to targeted and effective approaches to improving health outcomes and well-being for individuals within each group.

The results of our analysis shed light on the mobility patterns observed within the five homogeneous communities or clusters, represented in Figure 3(a) to Figure 3(e), at an hourly granularity level. Additionally, the HRV patterns are illustrated in Figure 3(f). These homogeneous communities are formed based on various factors, including hourly mobility patterns, sleep and wakeup patterns, average mobility acceleration (in m/s<sup>2</sup>), and overall day-wise mobility acceleration (in m/s<sup>2</sup>). Analyzing the hour-wise mobility patterns within each homogeneous community reveals distinct sleep, wakeup, and daily routine characteristics. Each community exhibits unique patterns in their sleep schedules. For instance, subjects in community 1 maintain a consistent sleep routine between midnight and 5 am, while community 3 displays a consistent sleep pattern from midnight to 6 am. On the other hand, community 5 follows a sleep routine from 2 am to 7 am, and community 4’s sleep routine is observed between 1 am and 6 am. These variations in sleep patterns contribute to the formation of distinct homogeneous groups within the network.

Furthermore, the wakeup and day start routines also exhibit noticeable differences among the communities. As individuals wake up, there is a marked increase in their mobility levels. In community 1, there is a significant rise in mobility between 5 am and 7 am, while community 4 shows an increase between 7 am and 9 am. Community 3 exhibits an upsurge in mobility between 6 am and 8 am, and community 5 demonstrates a slightly delayed wakeup schedule between 9 am and 12 pm.

These distinct patterns in wakeup times further differentiate the homogeneous communities, highlighting the importance of considering individual sleep and morning routine behaviors. During daytime hours, most of the communities engage in relatively active mobility activities, with little variation in their average mobility acceleration, except for community 2. Community 1 stands out with the highest average acceleration compared to the other groups, indicating a more intense level of mobility. On the other hand, communities 3 and 5 exhibit slightly lower average mobility acceleration, suggesting a comparatively less intense but still active level of mobility. Interestingly, community 2 demonstrates the lowest mobility acceleration among the groups, accompanied by abnormal mobility patterns. This finding indicates the presence of unique and potentially abnormal mobility characteristics within community 2, which warrants further attention and validation.

In terms of HRV patterns, the total average heart rate shows no significant difference among the subjects, but there is noticeable variation in the maximum heart rate measured over the total data collection period. Specifically, subjects ID47 and ID52 demonstrate significantly lower HRV compared to other individuals within the homogeneous groups. However, it is observed that mobility patterns exert a stronger influence in the formation of communities compared to HRV patterns.

In conclusion, our analysis of the mobility patterns within the five homogeneous communities reveals distinctive sleep, wakeup, and daily routine characteristics among the subjects. The variations in these patterns contribute to the formation of homogeneous groups, highlighting the relevance of individual sleep behaviors and daytime mobility activities. Additionally, the identification of abnormal mobility patterns within community 2 underscores the importance of considering such anomalous behaviors for further investigation and validation. These findings provide valuable insights into the mobility behaviors and daily routines within each community, facilitating the design of personalized interventions and healthcare strategies tailored to the specific characteristics exhibited by



Fig. 3. Mobility and HRV data Analysis

these homogeneous groups.

IV. DISCUSSION

Identifying homogeneous groups within biological networks is a critical task with significant implications for developing personalized therapies and individualized medicines. In many cases, universal medications and group therapies do not provide optimal results for all individuals. Therefore, the development of personalized and targeted interventions is essential. The fundamental objective of identifying homogeneous groups in biological networks is to enable the customization of treatments and interventions based on the unique characteristics of each group. Each homogeneous group identified in our analysis exhibits two important characteristics: homogeneity within the group and distinctiveness from other homogeneous groups. This means that individuals within a homogeneous group share similar patterns and behaviors, while also differing significantly from individuals in other groups. Understanding these distinctions allows for the design of interventions that specifically cater to the needs and characteristics of each group.

A unique aspect of our methodology is that it does not rely on the inclusion of class labels to detect homogeneous

groups. This data-driven approach eliminates the need for manual human annotation of class labels, saving considerable time and effort during the data collection process. By solely relying on the data itself, our methodology efficiently identifies homogeneous groups based on the inherent patterns and relationships within the data. This data-driven approach enhances the objectivity and accuracy of the analysis. Additionally, our methodology has the potential to identify subject groups with abnormal or peculiar mobility and HRV patterns. An example of such an abnormal group is community 2, where individuals exhibit minimal and abnormal mobility patterns throughout the day. The identification of such abnormal groups is crucial as it can highlight individuals who may require specific attention, interventions, or further investigation due to their unique characteristics. By recognizing these abnormalities, healthcare professionals can tailor interventions to address the specific needs of these individuals, potentially improving their health outcomes.

V. LIMITATIONS AND FUTURE WORK

There are several limitations to be acknowledged in our study. First, the sample size of our dataset was relatively small, consisting of only 52 subjects. This limited sample

size may impact the generalizability of our findings and limit the scope of interpretations. A larger and more diverse dataset would provide a more comprehensive understanding of homogeneous groups and their mobility patterns within the biomedical network. Another limitation arises from the selection of the correlation threshold. The choice of threshold can introduce bias into the analysis, as it determines the strength of connections within the network. Different thresholds may yield different results, and selecting the optimal threshold is subjective. Future studies should explore robust methods to determine an appropriate threshold or consider alternative approaches that are less dependent on this parameter.

To overcome the limitations mentioned above, future work should focus on expanding the dataset to include a larger number of subjects. A more extensive and diverse sample would enhance the representativeness and generalizability of the findings. Additionally, incorporating longitudinal data would provide insights into the temporal dynamics of homogeneous groups and their mobility patterns over time. Furthermore, it would be valuable to integrate clinical parameters and additional biomarkers into the analysis. Including clinical data, such as medical history, comorbidities, and medication usage, can provide a more comprehensive understanding of the factors influencing the formation of homogeneous groups. The incorporation of other biomarkers, such as genetic information or biochemical markers, could offer deeper insights into the underlying physiological and molecular mechanisms related to mobility patterns. Moreover, future research could explore advanced network analysis techniques to gain a more nuanced understanding of the relationships within the biomedical network. For example, community detection algorithms beyond the Louvain clustering algorithm could be employed to uncover finer-grained subgroups within homogeneous groups. Additionally, considering the dynamic nature of the data, time-evolving network analysis methods could be employed to capture changes and transitions in homogeneous group formations over time.

## VI. CONCLUSION

In this study, we developed a graph model to take advantage of mobility data and medical data to identify groups with potential similar properties. Our approach focused on the identification of homogeneous groups within a biomedical network using multi-modal data. By constructing a correlation network, we identified homogeneous groups that exhibited strong interconnections based on their mobility patterns. The analysis revealed distinct patterns and characteristics within these homogeneous groups, providing valuable insights for personalized therapies and individualized interventions. Furthermore, the analysis of hour-wise mobility patterns highlighted unique sleep, wake-up, and daily routine characteristics within each group. Additionally, the examination of average mobility acceleration revealed variations among the homogeneous groups, with some demonstrating higher or lower levels of mobility. The data-driven approach used in our methodology eliminated the need for manual class label

annotations, enhancing the efficiency and objectivity of the analysis. Moreover, the identification of subjects that didn't belong to any of highly similar groups emphasized the potential of our methodology to detect individuals with different mobility characteristics, warranting further investigation and personalized attention. For future directions, we plan to expand our analysis by incorporating additional clinical parameters and other medical information for each individual in the study.

## REFERENCES

- [1] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [2] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics reports*, vol. 659, pp. 1–44, 2016.
- [3] J. Yang, J. McAuley, and J. Leskovec, "Detecting cohesive and 2-mode communities in directed and undirected networks," in *Proceedings of the 7th ACM international conference on Web search and data mining*, 2014, pp. 323–332.
- [4] R. Rotabi, K. Kamath, J. Kleinberg, and A. Sharma, "Detecting strong ties using network motifs," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 983–992.
- [5] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [6] I. D. Yalom and K. Rand, "Compatibility and cohesiveness in therapy groups," *Archives of General Psychiatry*, vol. 15, no. 3, pp. 267–275, 1966.
- [7] E. A. Finkelstein et al., "Effectiveness of activity trackers with and without incentives to increase physical activity (trippa): a randomised controlled trial," *The lancet Diabetes & endocrinology*, vol. 4, no. 12, pp. 983–995, 2016.
- [8] Y. Yuan, *Characterizing human mobility from mobile phone usage*. University of California, Santa Barbara, 2013.
- [9] B. M. Saykrs, "Analysis of heart rate variability," *Ergonomics*, vol. 16, no. 1, pp. 17–32, 1973.
- [10] A. Sallaberry, Y.-c. Fu, H.-C. Ho, and K.-L. Ma, "Contact trees: Network visualization beyond nodes and edges," *PloS one*, vol. 11, no. 1, p. e0146368, 2016.
- [11] S. A. Hicks et al., "Hyperaktiv: An activity dataset from patients with attention-deficit/hyperactivity disorder (adhd)," in *Proceedings of the 12th ACM Multimedia Systems Conference*, 2021, pp. 314–319.
- [12] I. Cohen et al., "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.
- [13] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008.
- [14] R. Thelagathoti and H. Ali, "A population analysis approach using mobility data and correlation networks for depression episodes detection," *Ann Depress Anxiety*, vol. 9, no. 1, p. 1112, 2022.
- [15] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.
- [16] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized louvain method for community detection in large networks," in *2011 11th international conference on intelligent systems design and applications*. IEEE, 2011, pp. 88–93.
- [17] R. K. Thelagathoti and H. H. Ali, "The comparison of various correlation network models in studying mobility data for the analysis of depression episodes." in *BIO SIGNALS*, 2022, pp. 200–207.