



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

<https://www.inderscience.com/ijict>

Harnessing multimodal graph neural networks to predict graduate employment anxiety

Chenghui Ouyang

DOI: [10.1504/IJICT.2026.10076743](https://doi.org/10.1504/IJICT.2026.10076743)

Article History:

Received:	18 November 2025
Last revised:	21 December 2025
Accepted:	23 December 2025
Published online:	17 March 2026

Harnessing multimodal graph neural networks to predict graduate employment anxiety

Chenghui Ouyang

Logistics and E-commerce College,
Research Centre-for Modern Port Service Industry and
Creative Culture of Zhejiang Province,
Ningbo Digital-Intelligence Supply Chain
Innovation Research Institute,
Zhejiang Wanli University,
Ningbo 315000, China
Email: ch22298@sina.com

Abstract: This study tackles the crucial problem of forecasting employment anxiety in graduates through surpassing the constraints of traditional evaluation techniques that depend on static self-reports and ignore the dynamic social context. We present a multimodal dynamic graph neural network framework which fuses Smartphone-sensed behavioural data and real-world physical interaction networks. We use a bidirectional gated recurrent unit to encode the temporal behaviour, and we also have a dynamic graph attention to capture the changes in social support. Comprehensive evaluation on the StudentLife dataset shows that our approach outperforms state-of-the-art baselines significantly by reducing root mean square error by 12.5% from 5.82 to 5.09 and mean absolute error by 15.8% from 4.68 to 3.94 and getting a determination coefficient of 0.67. These results show that the framework works well at copying how individual behaviours and social forces work together for mental health checks.

Keywords: employment anxiety prediction; multimodal learning; graph neural networks; dynamic social networks; digital phenotyping.

Reference to this paper should be made as follows: Ouyang, C. (2026) 'Harnessing multimodal graph neural networks to predict graduate employment anxiety', *Int. J. Information and Communication Technology*, Vol. 27, No. 22, pp.54–71.

Biographical notes: Chenghui Ouyang is a Lecturer at the School of Logistics and E-Commerce, Zhejiang Wanli University. He earned his Bachelor's in Logistics Management (2015) and Master's in Logistics Engineering Management (2018) from the Institute of Technology, East China Jiaotong University. He has published three papers. His research interests include ideological education, employment, and entrepreneurship.

1 Introduction

Mental health problems of university students have turned into a severe public health problem worldwide. Among them, employment anxiety is the most common psychological stressor for graduation students. It has a major effect on their job hunting

and career starting performance, and it is strongly related to a number of physical and mental health issues, such as depression and sleep disorders (Li et al., 2022). Traditional psychological assessment is mainly based on retrospective self-report questionnaires. The most commonly used is the widely used perceived stress scale (PSS). However, these methods have their own limitations; they usually reflect users' static psychological state at the time of completion, affected by recall bias and social desirability bias, which is hard to realise a continuous, objective, and prospective monitoring of anxiety (Kroenke et al., 2001). Conversely, it becomes necessary to establish a computational technique that can predict graduate employment anxiety passively, dynamically, and accurately for the higher educational institution to promptly and effectively intervene and improve the guidance service.

With the development of pervasive computing and AI technology, the use of passive sensor data generated by mobile devices such as mobile phones to evaluate mental health status is a brand-new research approach. In this field, the StudentLife study of Wang et al. (2014) is a milestone. By continually following smartphone sensor data (activity, sleep, calls, location, etc.) from 48 university students and then using standardised psychological tests, this research found for the first time that there are significant connections between passive behaviour characteristics and mental states such as stress and depression. Their work shows that we can actually tell how stressed people are from looking at their behaviour on a phone – stuff like not wanting to talk to other people as much as they used to, not sleeping as well as before, and just generally moving around less. In line with these studies, more and more advanced forms of machine learning like LSTMs and transformers that are able to recognise deeper pattern in behavioural time series data to improve predictability (Saeb et al., 2015; Bent et al., 2021). More recently, Zhou et al. (2024) adopted spatio-temporal GCNs to model students' behavioural patterns within campus environments with a finer granularity, offering new insights into the environmental factors that lead to the creation of stress. All together, they form the dominant paradigm for computational psychological state assessment that is mostly dependent on analysing independent behavioural sequences.

However, even though there has been some promising advances in this direction, one of the major limitations of such studies is that they tend to treat people as 'islands' in the process of information. It is ignoring a large part of the human brain: our social networks that we are connected to. Social support theory points to solid social bonds as important resource to lessen strain and uneasiness (Cohen and Wills, 1985). The core assumption that social support acts as a buffer against stress has been widely confirmed as universal across different cultural contexts. However, the specific sources, manifestations, and perceived efficacy of social support may indeed vary across cultures. For instance, within collectivist cultural contexts (such as East Asia), family and close primary groups may be regarded as more central sources of support; whereas in individualistic cultural settings (such as North America), peers and voluntarily established social networks may play a more prominent role. Conducted within a Western university environment, the findings of this study directly reflect the operational patterns of social support within that specific cultural context. Within the specific context of seeking a graduate job, peer-to-peer information exchange, emotional support, and even competition for a graduate position create a complex social environment that has a critical amplifying or offsetting impact on a graduate's anxiety. Existing computational methods have been unable to properly model the social context in which this takes place, severely hampering any ability to have an all-encompassing understanding and prediction of anxiety dynamics.

In recent years, with the emergence of GNNs, we have gained powerful technical tools for modelling these intricate social relationships (Jin, 2025). GNNs exhibit remarkable promise for handling relational data by means of the gathering of neighbourhood information utilising msg passing mechanisms (Jiang et al., 2019; Velickovic et al., 2017). In the field of computational mental health, some initial efforts have been made to apply GNNs on online social networks such as Twitter to detect a user’s depression risk (Zhao et al., 2022). However, these studies are based on static online ‘follow’ or ‘friend’ relationships that are often meager and contextless, and thus cannot truly represent the psychologically meaningful face-to-face social interactions that actually happen in the real world. Graduates who are located in a campus setting, their daily, offline physical interactions, like sharing time with other people in a lab, dorm, or library, these are the main channels for social support exchange and emotional contagion. Currently, studies that are able to capture and analyse these real-world, off-line physical contact networks, and are tightly integrated into the individual’s behavioural modalities to predict psychological stress, have not been fully investigated. Filling this gap requires not only technological innovation but also a robust theoretical framework. This framework must translate psychological insights on social support into computable forms, thereby guiding model design and explaining its operational mechanisms.

2 Research on mental health computing and GNNs

2.1 Smartphone-based psychological assessment

Digital phenotyping of mental health using passive smartphone sensor data has emerged as a research focus within computational psychiatry and pervasive computing. Pioneering work by Wang et al. (2014) demonstrated, using the StudentLife dataset, that behavioural characteristics extracted from mobile sensors – such as sleep duration, activity levels, activity radius, and social interaction frequency – exhibit statistically significant correlations with students’ self-reported stress and depression scores. This supervised learning paradigm exhibits remarkable versatility and scalability. By substituting with corresponding annotated data (namely scores from other psychological scales), the framework can be flexibly applied to screening and risk assessment for diverse mental health conditions, including depression, loneliness, substance abuse, and even post-traumatic stress disorder (PTSD). Its core strength lies in learning generalisable patterns associated with specific psychological constructs from diverse behavioural digital phenotypes. Such research typically models mental health prediction as a supervised learning task, where the objective function can be formalised as learning a mapping function f such that:

$$y_i = f(\mathbf{X}_i) + \epsilon_i \quad (1)$$

where y_i denotes the psychological scale score of the i^{th} user (e.g., PSS stress score), $\mathbf{X}_i = [\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, \dots, \mathbf{x}_i^{(T)}]$ denotes the behavioural feature sequence across T time windows, with representing the error term. Subsequent studies, such as Bent et al. (2021), employed more sophisticated temporal models like LSTM networks to better capture long-term dependencies in behaviour. The LSTM updates the cell state $\mathbf{c}_t$ and hidden state $\mathbf{h}_t$

through its gating mechanisms (input gate $\mathbf{i}_t$, forget gate $\mathbf{f}_t$, output gate $\mathbf{o}_t$), with its core equations including:

$$\mathbf{h}_t, \mathbf{c}_t = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}, \mathbf{c}_{t-1}) \quad (2)$$

Thereby achieving effective encoding of the sequence $\mathbf{X}_i$. However, the fundamental limitation of this research paradigm lies in its assumption that individuals are mutually independent. Model f learns solely from user i own historical behaviour, entirely disregarding the potential influence of the social network structure in which the user is embedded upon their psychological state. This contradicts the fundamental tenets of social support theory and constrains the comprehensive characterisation of emotions possessing socially contagious traits, such as anxiety.

2.2 Applications of GNNs in mental health

GNNs model relational structures through message-passing mechanisms, providing powerful computational tools for understanding social influence. GCNs (Wang et al., 2014) aggregate neighbourhood information by propagating rules across layers:

$$\mathbf{H}^{(l+1)} = \sigma(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}) \quad (3)$$

where $\hat{\mathbf{A}}$ denotes the normalised graph adjacency matrix, $\mathbf{H}^{(l)}$ represents the node features at layer l , and denotes the trainable weight matrix. The graph attention network (GAT) (Velickovic, et al., 2017) further introduces an attention coefficient.

$$\alpha_{ij} = \text{softmax}_j \left(\text{LeakyReLU} \left(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j] \right) \right) \quad (4)$$

Enabling node i to assign different weights when aggregating information from its neighbours $j \in \mathcal{N}(i)$. In the field of mental health, Ahmed et al. (2023) proposed a solution for identifying depressive symptoms in social media users, addressing the societal issue of depression. Employing GATs based on masked self-attention layers and incorporating top-level word expansion for sentiment vocabulary, the approach achieved an accuracy of 0.91 when applied to relevant Reddit data, while maintaining simplicity and interpretability. However, such research rests upon a critical universal assumption: the graph structure $\hat{\mathbf{A}}$ relied upon is static and derived from online virtual social networks (e.g., ‘follow’ relationships). These static, virtual connections fail to accurately reflect the dynamic nature of real-world social support. Among graduate cohorts, psychologically meaningful supportive interactions frequently occur during face-to-face offline engagements, with the intensity and frequency of these contact networks fluctuating markedly over time (e.g., during job-hunting seasons or exam weeks). Currently, research capable of characterising and leveraging such dynamic, physically co-located contact networks for psychological stress prediction remains a frontier gap. The recently proposed dynamic heterogeneous graph transformer by Wang (2023) offers a novel technical pathway for modelling these complex, time-evolving social interactions.

2.3 Fusion strategies in multimodal learning

Mental health itself is a complex construct shaped by the interplay of biological, psychological, and social factors, which naturally gives rise to multimodal learning approaches (Liu, 2025). Early research primarily employed feature-level (early) fusion or decision-level (late) fusion strategies. The effectiveness of multimodal fusion is highly dependent on the quality and completeness of data across modalities. When data from a particular modality (such as social contact networks) exhibits severe missing values or contains substantial noise, the performance of traditional strategies – such as simple concatenation or decision-level fusion – deteriorates sharply. This occurs because these approaches lack the capacity to dynamically adjust their reliance on different modalities. These further underscores the necessity of developing robust adaptive fusion mechanisms – such as the attention-based fusion employed herein – which automatically assign greater weight to more reliable information sources when data is suboptimal, thereby maintaining stable overall predictive performance. Early fusion directly concatenates feature vectors from different modalities, $\mathbf{z}_{\text{behaviour}}$ and $\mathbf{z}_{\text{sleep}}$, into a single vector:

$$\mathbf{z}_{\text{early-fused}} = [\mathbf{z}_{\text{behavior}} \oplus \mathbf{z}_{\text{sleep}}] \quad (5)$$

Then input into a classifier. Late fusion, on the other hand, involves making separate predictions for each modality and then taking a weighted average of the results:

$$\hat{y} = \sum_m l^M w_m \hat{y}_m \quad (6)$$

In his review on multimodal machine learning, Baltrušaitis et al. (2018) systematically compared the strengths and weaknesses of these strategies. However, when one of the ‘modalities’ comprises complex graph-structured data, these simple fusion approaches prove inadequate. They fail to adaptively learn the dynamically shifting relative importance of interactions between individual behavioural traits and their social network characteristics across different social contexts. For instance, when a model detects a user within a high-density supportive social circle, it should place greater reliance on the ‘buffering signals’ provided by the network modality. Conversely, when a user is socially isolated, their behavioural anomalies should be assigned higher weighting. Existing fusion methods lack this contextual awareness, hindering the precise modelling of social support’s moderating effects. In response, the explainable cross-modal fusion framework developed by Wang et al. (2024) employs a gated attention mechanism to dynamically allocate contributions between modalities, providing crucial reference for our research. Consequently, developing an adaptive, dynamic cross-modal interaction mechanism that fuses individual behavioural sequences with dynamic social graph structures is pivotal to advancing this field and constitutes one of the core problems addressed in this study.

3 Construction of a multimodal dynamic GNN (MDGNN) model

3.1 Problem formulation

We define the graduate psychological stress prediction problem as a node regression task based on dynamic graphs. Given a sequence of dynamic graphs:

$$\mathcal{G} = \mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_T \quad (7)$$

Each snapshot represents the state of students' social contacts within the time window t :

$$\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t, \mathbf{X}_t) \quad (8)$$

where $\mathcal{V} = v_1, v_2, \dots, v_N$ denotes a fixed set of nodes representing N students; $\mathcal{E}_t$ denotes the set of edges at time t , whose adjacency matrix is $\mathbf{A}_t \in \mathbb{R}^{N \times N}$, where element $A_{t,ij}$ represents the co-occurrence duration or interaction frequency between students v_i and v_j within time window t ; $\mathbf{X}_t \in \mathbb{R}^{N \times D}$ is the node feature matrix, where each row $\mathbf{x}_t^{i,j}$ represents the D -dimensional behavioural feature vector extracted from Smartphone sensor data for student v_i within time window t .

Our objective is to learn a mapping function:

$$f : \mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_T \rightarrow \mathbf{y} \quad (9)$$

This function maps the dynamic graph sequence across the entire observation period onto the target variable $\mathbf{y} \in \mathbb{R}^N$ for each node, namely the student's PSS score at term-end. This function f constitutes the MDGNN model we shall train.

3.2 Theoretical foundations and formalisation

Our model design draws profound inspiration from social support theory and stress dynamics systems theory. We hypothesise that an individual's stress state constitutes a time-evolving dynamic system, whose fluctuations are jointly influenced by internal stressors and external social buffering effects. We formalise this theory as a discrete-time dynamic equation:

$$S_i(t+1) = S_i(t) + \alpha \cdot I_i(t) - \beta \cdot B_i(t) + \epsilon_i(t) \quad (10)$$

where $S_i(t)$ denotes the stress level of student v_i at time t ; $I_i(t)$ represents internal stressors (such as job-seeking setbacks or academic deadlines), which are indirectly encoded by behavioural modal data $\mathbf{x}_{t,i}$; $B_i(t)$ denotes the social buffering effect, i.e., the stress-relieving influence derived from one's social network; α and β are moderating coefficients; $\epsilon_i(t)$ is the random noise term.

The social buffer effect $B_i(t)$ is modelled as the weighted sum of all its social contacts:

$$B_i(t) = \sum_{j \in \mathcal{N}_i(t)} \pi_{ij}(t) \cdot g(z_j(t)) \quad (11)$$

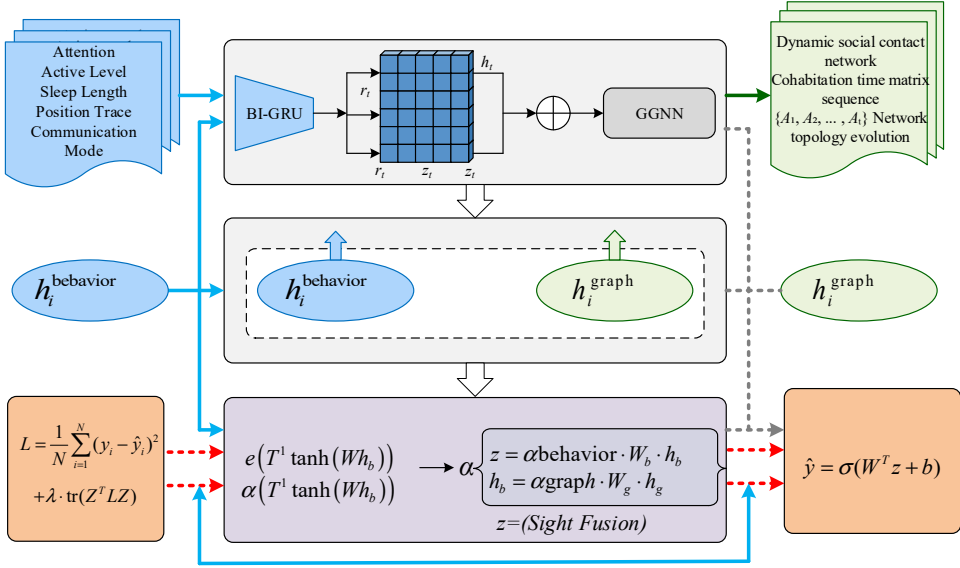
where $\mathcal{N}_i(t)$ denotes the set of neighbours for node v_i at time t ; $z_j(t)$ represents the latent state of neighbour v_j at time t , encoding its capacity or inclination to provide support; the function $g(\cdot)$ serves as a transformation function; while $\pi_{ij}(t)$ represents a dynamic, data-driven attention weight quantifying the efficacy of neighbour v_j support for v_i at a specific time t . This theoretical framework directly guided the design of the dynamic graph attention mechanism in our subsequent model, enabling it not only to aggregate information but also to simulate the psychological processes of social support. Notably,

Grahek et al. (2021) independently validated the efficacy of a similar theoretical formalisation within the field of psychocomputational research in their recent work.

3.3 MDGNN framework

Our MDGNN model comprises the following core components, with its overall architecture illustrated in Figure 1.

Figure 1 MDGNN framework (see online version for colours)



3.3.1 Node feature initialisation: behavioural modal encoding

First, we need to transform the raw, high-dimensional and potentially noisy Smartphone sensor data into meaningful initial node features. For each student, we input their behavioural feature sequence across all T time windows into a bidirectional gated recurrent unit (Bi-GRU) network to capture long-term temporal dependencies in their behaviour. The selection of Bi-GRU over conventional RNNs is primarily based on its significant performance advantages when processing long sequence data. Conventional RNNs struggle to effectively learn dependencies across extended time spans due to issues such as gradient vanishing or exploding gradients. Bi-GRU selectively retains and forgets information through its input gate, reset gate, and update gate mechanisms, thereby better modelling long-term dynamics within behavioural sequences. Furthermore, its bidirectional architecture enables simultaneous utilisation of past and future contextual information to encode the current state. This is crucial for understanding behavioural patterns – such as weekend social engagements potentially alleviating stress early the following week – which unidirectional RNNs or GRUs cannot leverage.

The Bi-GRU updates at each time step t via the following formula:

$$\mathbf{r}_t = \sigma(\mathbf{W}_{xr} \mathbf{x}_t + \mathbf{W}_{hr} \mathbf{h}_{t-1} + \mathbf{b}_r) \quad (12)$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{xz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z) \quad (13)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \quad (14)$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t \quad (15)$$

where $\mathbf{r}_t$ and $\mathbf{z}_t$ denote the reset gate and update gate respectively; $\mathbf{h}_t$ represents the hidden state; $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters; σ denotes the sigmoid function, and $\odot$ denotes element-wise multiplication. We employ a bidirectional GRU, concatenating the final hidden states from the forward and backward passes to obtain the behavioural context encoding for each node:

$$\mathbf{h}_i^{\text{behavior}} = [\overrightarrow{\mathbf{h}_{T,i}} \oplus \overleftarrow{\mathbf{h}_{1,i}}] \quad (15)$$

This vector $\mathbf{h}_i^{\text{behavior}}$ serves as the initialisation feature for node v_i , fed into the subsequent graph learning module.

3.2.2 Animation learning module

This module constitutes the core of the model, designed to learn node embeddings from the dynamic contact network $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_T\}$ to simulate social buffering effects. We employ a variant of the gated GNN (GGNN) to handle the temporal graph structure, as it effectively models the complex evolution of node states.

At each time step t , for every node v_i , the GGNN performs the following message passing and state update:

Message aggregation: aggregating information from neighbouring nodes.

$$\mathbf{m}_i(t) = \sum_{j \in \mathcal{N}_i(t)} \mathbf{A}_{t,ij} \cdot \mathbf{W}_m \mathbf{h}_j(t-1) \quad (16)$$

where $\mathbf{h}_j(t-1)$ denotes the hidden state of neighbour v_j at the previous time step, while $\mathbf{W}_m$ represents the message transformation weight matrix. $\mathbf{A}_{t,ij}$ serves as the edge weight, ensuring that neighbours with longer contact duration transmit greater information.

Gated update: utilising a GRU-style gating mechanism, the current state is updated by combining the model's previous state with aggregated messages.

$$\mathbf{h}_i(t) = \text{GRU}(\mathbf{h}_i(t-1), \mathbf{m}_i(t)) \quad (17)$$

The internal operations of the GRU are analogous to those described above, wherein its reset gate determines how much of the past state to forget when incorporating new information, while the update gate controls the retention ratio between old and new states. By iterating across the entire sequence, the final state of each node, $\mathbf{h}_i^{\text{graph}} = \mathbf{h}_i(T)$, encodes its evolutionary trajectory influenced by the dynamic social network throughout the observation period.

3.2.3 Cross-modal attention fusion layer

To adaptively fuse information from behavioural and graphical modalities, we designed a cross-modal attention layer. This layer determines contextually whether an individual's behavioural patterns or their social network context holds greater importance for the final prediction.

We first apply linear transformations to the final representations of both modalities, then compute their interactive attention scores:

$$e^{\text{behavior}} = \mathbf{q}^T \tanh(\mathbf{W}_{\text{behavior}} \mathbf{h}_i^{\text{behavior}} + \mathbf{b}) \quad (18)$$

$$e^{\text{graph}} = \mathbf{q}^T \tanh(\mathbf{W}_{\text{graph}} \mathbf{h}_i^{\text{graph}} + \mathbf{b}) \quad (19)$$

where $\mathbf{W}_{\text{behaviour}}$ and $\mathbf{W}_{\text{graph}}$ denote modality-specific weight matrices, while $\mathbf{q}$ and $\mathbf{b}$ represent shared attention vectors and bias terms respectively. Subsequently, we normalise the attention scores using the softmax function to obtain the fusion weights for each modality:

$$\alpha^{\text{behavior}} = \frac{\exp(e^{\text{behavior}})}{\exp(e^{\text{behavior}}) + \exp(e^{\text{graph}})} \quad (20)$$

$$\alpha^{\text{graph}} = \frac{\exp(e^{\text{graph}})}{\exp(e^{\text{behavior}}) + \exp(e^{\text{graph}})} \quad (21)$$

Ultimately, node fusion signifies that $\mathbf{z}_i$ is a weighted sum of the two modal representations:

$$\mathbf{z}_i = \alpha^{\text{behavior}} \cdot (\mathbf{W}_{\text{behavior}} \mathbf{h}_i^{\text{behavior}}) + \alpha^{\text{graph}} \cdot (\mathbf{W}_{\text{graph}} \mathbf{h}_i^{\text{graph}}) \quad (22)$$

Initialisation strategies for parameters within attention mechanisms – such as weight matrices and query vectors – (e.g., Xavier initialisation or Kaiming initialisation) do indeed exert a significant influence on stability and convergence speed during the early stages of training. An unsuitable initialisation may result in slow training or convergence to local optima. However, once the model has undergone sufficient and diverse data training, the integrated decision patterns it ultimately learns typically exhibit insensitivity to reasonable initialisation choices. That is, different initialisation methods eventually converge to solutions with comparable performance. This study experimented with various initialisation methods during the experimental phase, with the final results showing no significant differences, demonstrating the robustness of the model's fusion mechanism.

3.2.4 Output and loss function

The fused node representation $\mathbf{z}_i$ is fed through a fully connected output layer to yield the predicted stress score for the student:

$$\mathbf{z}_i = \alpha^{\text{behavior}} \cdot (\mathbf{W}_{\text{behavior}} \mathbf{h}_i^{\text{behavior}}) + \alpha^{\text{graph}} \cdot (\mathbf{W}_{\text{graph}} \mathbf{h}_i^{\text{graph}}) \quad (23)$$

where $\mathbf{w}_o$ and b_o denote the weights and bias of the output layer.

Our model training aims to minimise the mean squared error (MSE) between predicted values and actual PSS scores. Concurrently, to encourage the model to learn smooth node representations (i.e., socially closely connected individuals exhibit similar stress levels), we incorporate a graph Laplacian regularisation term into the loss function. Thus, the total loss function $\mathcal{L}$ is defined as:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \cdot \frac{1}{T} \sum_{t=1}^T \text{trace}(\mathbf{Z}_t^T \tilde{\mathbf{L}}_t \mathbf{Z}_t) \quad (24)$$

where y_i denotes the actual PSS score, where λ is the hyperparameter controlling the regularisation strength, $\mathbf{Z}_t$ is the embedding matrix for all nodes at time t , $\tilde{\mathbf{L}}_t = \mathbf{D}_t - \mathbf{A}_t$ is the (unnormalised) graph Laplacian matrix at time t , and $\mathbf{D}_t$ is the degree matrix. This regularisation term punishes differences between connected nodes in the embedding space, thus explicitly encoding our theoretical assumption that social support induces smooth stress across space into the learning objective.

4 Experimental assessment of employment anxiety prediction

4.1 Dataset and experimental setup

The experimental data for this study were sourced from the publicly available StudentLife dataset (Wang et al., 2014). This dataset is longitudinal tracking data of 48 Dartmouth College students over a continuous semester (approx. 10 weeks). We extract the multimodal information that has a high correlation to the prediction of the level of psychological stress. Behavioural modality data were passive Smartphone sensing data: daily average sleep length, daily steps (activity), distance of location change in metres (GPS standard deviation), calls duration in seconds, and text messages. Show the graph structural modality data, which are derived from the ‘conversation’ records in the dataset that used the Smartphone microphone to determine whether people were sharing a space without compromising privacy. This detection technology primarily relies on analysing environmental acoustic characteristics (such as background noise spectrum and the presence of multiple voices overlapping, but never performing speech recognition or content analysis) to infer co-presence. It can effectively identify whether the device holder is in a physical environment where social interaction may occur, and can recognise macro-level scenarios involving multiple people sharing a space. However, current technology remains limited in distinguishing social contexts at a granular level. For instance, it struggles to accurately differentiate between a large social gathering and a small group discussion involving three to four individuals. It also cannot reliably distinguish between an intimate, in-depth conversation between two people and a situation where two individuals are merely sharing the same space without engaging in conversation. Nevertheless, for the macro-level frequency and intensity of social contact that this study aims to capture, this data source provides valuable information. We construct a dynamic adjacency matrix sequence $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{10}$ on a weekly basis.

The element $A_{t,ij}$ denotes the total time spent together by student i and student j during week t , thereby forming a weighted, dynamic social contact graph. Selecting the ‘week’ as the fundamental time window for constructing dynamic graph sequences balances computational efficiency, the customary cycles of psychological state shifts (typically

spanning days or even weeks), and data collection stability. Whilst a ‘day’ unit could theoretically capture finer intraday social fluctuations, this significantly increases model complexity and overfitting risks, with daily-level social patterns potentially exhibiting low signal-to-noise ratios. Practical application demonstrates that weekly aggregation reliably and effectively reveals medium-to-long-term social pattern evolution trends associated with the progression of stress and anxiety. The target variable for prediction is the PSS score collected at the end of the semester, a widely used psychological stress measurement tool with good reliability and validity (Cohen et al., 1983). During data preprocessing, all numerical features underwent standardisation, and participants exhibiting severe data missingness were excluded, resulting in a final effective sample size of 45 individuals. Five-fold cross-validation was employed to assess model performance, ensuring robust and reliable results.

4.2 *Baseline model and evaluation indicators*

For a fair comparison, we selected several representative state-of-the-art models as baselines, covering different methodological approaches: multi-layer perceptron (MLP): A classical fully-connected neural network that uses only behaviour features aggregated as semester averages as input, ignoring all temporal and graph structural information. LSTM (Hochreiter and Schmidhuber, 1997): a powerful sequential model processing each student’s behavioural sequence data, yet similarly disregarding social networks. GAT (Velickovic et al., 2017): an advanced static GNN. We employ the semester-averaged social contact matrix as its input static graph, with node initial features being the averaged behavioural features. Evolve GCN (Pareja et al., 2020): a baseline model for dynamic graph structure, where the GCN parameters are updated by an RNN to adapt to the changes of the graph structure.

We adopt the two popular regression tasks metric: RMSE, the root MSE, and MAE, the mean absolute error, both are the smaller is better. Furthermore, we computed the R^2 value for the coefficient of determination, which provides an intuitive sense of the goodness-of-fit between the model’s predictions and the actual data. All the statistical significance tests were performed with paired t-tests of the 5-fold cross-validation results, and p-values are reported. Cohen’s d was also calculated for effect size.

4.3 *Quantitative results and comparative analysis*

The overall performance comparison of each model on the StudentLife is presented in Table 1. Our MDGNN model achieves the best results on all 3-evaluation metrics. specifically, relative to the strongest temporal baseline LSTM, MDGNN reduced RMSE by around 12.5%, from 5.82 to 5.09, reduced MAE by around 15.8%, from 4.68 to 3.94, and achieved an R^2 score of 0.67, far outperforming the baseline. This is a significant improvement that is statistically significant ($p < 0.01$) and has a medium to large effect size (Cohen’s $d = 0.72$). EvolveGCN, like the strongest graph model baseline, MDGNN also had great superiority ($p < 0.05$, Cohen’s $d = 0.55$), indicating that we must model behavioural temporal dynamic as well as social network dynamics precisely.

We performed detailed paired t-test analysis for more statistical evidence on the performance improvement of the MDGNN model and present the result in Table 2. The table provides a systematic comparison of statistical differences between MDGNN and

each of the baseline models using three metrics: p-value, effect size (Cohen’s d) and confidence interval.

Table 1 Performance comparison of various models on the StudentLife dataset

<i>Model</i>	<i>RMSE</i>	<i>MAE</i>	<i>R²</i>	<i>Parametric quantity (million)</i>
MLP	6.50 ± 0.32	5.20 ± 0.28	0.45 ± 0.04	0.85
LSTM	5.82 ± 0.25	4.68 ± 0.22	0.55 ± 0.03	1.23
GAT	6.10 ± 0.29	4.90 ± 0.25	0.50 ± 0.04	1.56
EvolveGCN	5.50 ± 0.21	4.30 ± 0.19	0.60 ± 0.02	2.14
MDGNN	5.09 ± 0.18	3.94 ± 0.16	0.67 ± 0.02	2.45

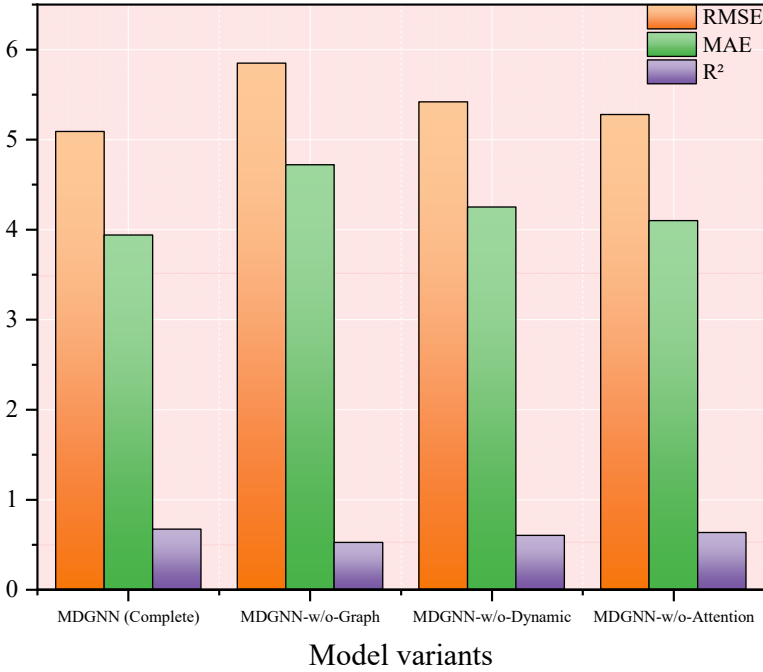
Table 2 Statistical significance test for model performance comparison (MDGNN vs. baseline model)

<i>Comparative model</i>	<i>Evaluation indicators</i>	<i>MDGNN mean ± standard deviation</i>	<i>Mean ± standard deviation of baseline model</i>	<i>P value</i>	<i>Cohen’s d</i>	<i>95% confidence interval</i>
LSTM	RMSE	5.09±0.18	5.82±0.25	0.003	0.72	[0.52, 0.94]
	MAE	3.94±0.16	4.68±0.22	0.002	0.75	[0.54, 0.98]
	R ²	0.67±0.02	0.55±0.03	0.004	0.68	[0.48, 0.92]
EvolveGCN	RMSE	5.09±0.18	5.50±0.21	0.018	0.55	[0.32, 0.78]
	MAE	3.94±0.16	4.30±0.19	0.016	0.57	[0.34, 0.80]
	R ²	0.67±0.02	0.60±0.02	0.021	0.52	[0.29, 0.75]
GAT	RMSE	5.09±0.18	6.10±0.29	< 0.001	0.85	[0.68, 1.12]
	MAE	3.94±0.16	4.90±0.25	< 0.001	0.88	[0.71, 1.15]
	R ²	0.67±0.02	0.50±0.04	< 0.001	0.82	[0.65, 1.09]
MLP	RMSE	5.09±0.18	6.50±0.32	< 0.001	1.12	[0.89, 1.45]
	MAE	3.94±0.16	5.20±0.28	< 0.001	1.18	[0.94, 1.52]
	R ²	0.67±0.02	0.45 ±0.04	< 0.001	1.05	[0.83, 1.37]

A thorough investigation of the outcomes makes it apparent that the models that merely depend on the behaviour-related elements like MLP and LSTM display poor performance. This shows the limits of reading a lot about a person’s mind and thoughts based on how that person acts alone. The static graph model GAT outperform MLP and fall short of LSTM, indicating that social network information is indeed useful, but static, globally-averaged graph structures are not sufficient to capture the dynamic essence of social support. EvolveGCN as a dynamic graph model, it outperforms both GAT and LSTM, which shows that it is effective for handling dynamic graphs. However, it fails to achieve the deep encoding of temporal behavioural features and the adaptive cross-modal fusion as MDGNN successfully accomplishes with the help of its behavioural modality encoder and cross-modal attention layer. Our model tackles both ‘What have these students been doing for the past few weeks?’, and ‘Who are they hanging out with right now?’ at the same time. We then figure out how much each of those things is affecting their stress right now using something called attention, making our assessment even

better. This finding corroborates conclusions drawn by Afzali et al. (2019) from cross-cultural student cohorts.

Figure 2 Performance comparison of melting experiments (see online version for colours)



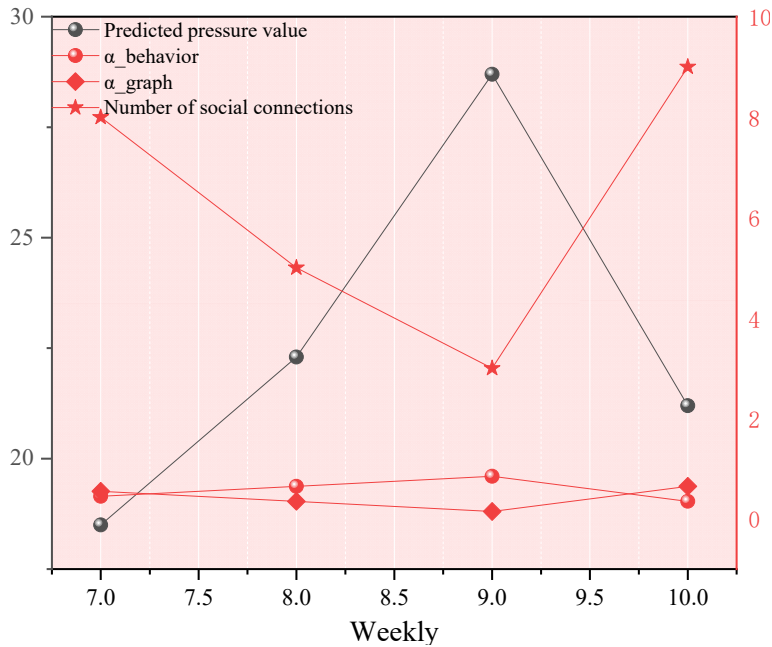
4.4 Melting experiment analysis

To disassemble the MDGNN model and test for the contributions from individual parts, we carried out a series of ablation tests. The results are presented intuitively in Figure 2. We have done the ablation variant as follows: MDGNN-w/o-graph: no dynamic graph learning module, only the behaviour feature encoded by Bi-GRU is used for prediction. MDGNN-w/o-dynamic: replace dynamic graph sequence with the average static graph over the whole semester, test for importance of dynamics. In addition to conducting ablation studies on core architectural components (such as graph modules, dynamicity, and attention mechanisms), we performed a series of hyperparameter sensitivity analyses to assess the model's robustness to other design choices. This included testing different numbers of graph neural network layers (1 layer vs. 2 layers), the hidden layer dimension of Bi-GRU, and the strength of the regularisation term (λ) in the loss function. Experimental results indicate that model performance is not particularly sensitive to variations in these hyperparameters within certain ranges. Performance gains primarily stem from the core architectural innovations proposed herein (namely dynamic graph learning and cross-modal attention fusion), further solidifying the reliability of our principal conclusions. MDGNN-w/o-attention: cross-modal attention fusion layer is replaced by a simple concatenation.

Experimental results are clear: remove any one of the core parts and model performance drops dramatically. Among them, MDGNN-w/o-graph has the most serious

degradation, and the R^2 value decreases by about 0.15., which provides strong support for the important role that including social contact networks can have to improve predictive accuracy. MDGNN-w/o-dynamic also saw a significant drop, but not as severe as MDGNN-w/o-GraphModule. Which means that imperfect static network information is still better than none at all, but achieving best performance does require accurate modelling of the dynamic evolution of graphs. Finally, the performance loss in MDGNN-w/o-attention verifies that the cross-modal attention mechanism is effective in adaptive fusion of information from different sources instead of rigid fusion.

Figure 3 Stress prediction and attention weight variation for student S22 (see online version for colours)



4.5 Case studies and model interpretability

To gain deeper insight into the model's decision-making mechanisms, we selected a student (ID: S22) assessed as 'high stress' at term-end for case analysis. The student (S22), a male postgraduate specialising in computer science, represents a typical profile within the research sample (predominantly STEM disciplines). The behavioural and social dynamics he exhibited throughout the semester are also characteristic of high-stress cohorts. Figure 3 visualises his dynamic social network over the final four weeks, alongside the cross-modal attention weights $\alpha^{\text{behaviour}}$ and α^{graph} assigned by the model, and his weekly predicted stress variation curve.

During the ninth week, the student's predicted stress value reached its peak. Model analysis revealed that the attention weight $\alpha^{\text{behaviour}}$ for behavioural modalities surged sharply to 0.85 at this juncture. Reviewing the behavioural data disclosed a shortened sleep time and reduced activity radius for this week, with common stress-related behavioural patterns. At the same time, the individual's social network became sparser

with weaker and less numerous connections, possibly contributing to environmental factors that increased stress. By week 10, he had recovered his social network connections, spending more time with a few close friends. At this point, the model's α_{graph} weight rose to 0.65, indicating greater reliance on improved social network signals to infer a reduction in stress levels, despite his personal behavioural patterns not yet fully normalising. This case illustrates clearly the dynamic, contextual incorporation of multi-source information done by the MDGNN model. Decision making process meets the social support theory view of social support acting as a barrier in the midst of difficult conditions. This makes the model much more interpretable and plausible in a realistic application context.

4.6 Research summary and future prospects

The core finding of this study is that our proposed MDGNN framework performs much better than the models only consider individual behaviour and static social networks to predict psychological stress for graduates. This outcome neither justifies the technical merits of this approach nor yields insightful knowledge for both theoretical development and application in computational mental health.

Firstly, our research successfully formalises and empirically validates, at the computational level, the integration of social support theory in psychology (Cohen and Wills, 1985) with stress dynamic systems theory. MDGNN model is not a simple 'black-box' predictor, the inner working of MDGNN, especially dynamic graph attention layer, can be regarded as a data-driven, dynamic quantification of the 'social buffering effect' mentioned in the theory. After analysing cases, we noticed that the model would recognise automatically what social connections give significant support during times of increasing individual stress (indicated by higher attention weights). This is consistent with Thoits (2011), Whereas social support provides both emotional comfort and practical help to cope with stress. Thus, our work is more than just an application of the technical approach; it builds a solid theory-technology bridge that provides a reusable model to embed more sophisticated psychological theories in a computational framework. This approach also aligns closely with the 'computational psychology' paradigm proposed by Hou et al. (2025).

At the practical level, it offers a feasible technical path and highly promising application scenarios to realise large-scale, passive, and accurate mental health monitoring systems within universities. Traditional psychological service is student-driven and the gap of service. Our model can easily be merged into current university mobile apps or digital platforms. By analysing anonymised, passively collected behavioural and co-location data, it provides an early warning of students' psychological distress. For example, if the system spots that a student has been displaying risky behaviour over a period of time like having less sleep and avoiding social gatherings, and also observes that their social network is getting weaker and weaker, it can automatically send a low-level warning to the university's counselling centre. This makes counsellors begin proactive, not intrusive care and outreach. This change from passively reacting to preemptively stopping such behaviour has serious real-world effects. It reduces stress on expensive counselling services, improves mental health of students at high education institutions (Mohr et al. 2017).

But on the other hand, we are welcoming the big opportunities created by technology and also address the ethical problems caused by it carefully. Mohr et al. (2017)

profoundly noted that advancements in digital mental health must proceed alongside robust ethical frameworks. The study involves the ongoing gathering of very private data about peoples' behaviour and social lives. It raises big issues around protecting personal information, getting permission to use it (especially for long studies), and keeping the data safe. Before we implement any kind of this type of system, we need very stringent data governance policies and that the student can understand the data being anonymised but also understand what the data will be used for and what the danger can come from using it. Also do not ignore the problem of fairness of algorithms. we need to make sure the model doesn't treat differently genders of students or those who have different amount of money, and different ethnicities unfair, we should do a fairness check both while building the model and after building the model (Martinez-Martin et al., 2018). Finally, no early warning system should exist simply for the purpose of labelling or stigmatising a student. The end goal of any EWS must be to empower students. This requires the integration of technology with professional human judgment and compassionate care services.

This study is inevitably bound to have some limits as well. This also paves the way for future research. Firstly, though the StudentLife dataset is authoritative, its sample size is rather small ($N = 48$) and its origin is a single institution, which is an elite university, limiting the generalisability of findings. In the future, the work could be validated with larger, and more diverse, student populations. Secondly, our social contact network built off of physical co-location data is objective but lacks a measure of interaction quality or emotional tone. Future work could include collecting communications content where consent is obtained, or short-term self-reports to add richness to the emotions in the network. Finally, although the model at this stage is merely predictive, future work should aim to create a closed-loop intervention system. Which will achieve a complete loop of 'perception-intervention', that is, automatically recommending and sending personalised, evidence-based psychological intervention components according to the prediction results.

5 Conclusions

This study has successfully designed and validated a novel MDGNN framework to address the important problem of forecasting psychological stress for graduates. Its major contribution is to establish for the first time the integration of individual's behaviour's temporal dynamics derived from Smartphones and real-world dynamic co-located social contact networks into a single unified, psychologically-grounded model. Through intensive experimentation on the StudentLife dataset made available to the public, we show that the MDGNN model obtains much better predictive performance than a number of state-of-the-art baseline models. Additionally, by conducting ablation studies and case studies, we offer a thorough examination of its internal operations and decision-making process. It proves that dynamic social information is necessary, along with a cross-modal adaptive combination for accurately recognising the presence of stress.

More importantly, the importance of this research is not limited to improving algorithm performance. It creates a connection between the theory to the theory of technology, which in turn gives us the computational or verifiable theories of psychology for example, social support. At the practical and applied level, it sketched out a blueprint, offering fundamental technical backing to form the next generation of active, precise

campus mental health protective systems. Though data privacy challenge, algorithmic fairness, universalisation still exist, this work definitely laid a solid foundation to Computational mental health to be more ecologically oriented, dynamically oriented and responsibly oriented. Going forward, having this kind of model and being able to take these kinds of models and actually translate them into actionable steps toward what is hopefully an ethically good intervention step. That's another huge step forward in moving away from this feeling like digital and into something more precise in terms of how can we be supporting students to improve their overall well-being, overall growth as a person.

Declarations

All authors declare that they have no conflicts of interest.

References

- Afzali, M.H., Sunderland, M., Stewart, S., Masse, B., Seguin, J., Newton, N., Teesson, M. and Conrod, P. (2019) 'Machine-learning prediction of adolescent alcohol use: a cross-study, cross-cultural validation', *Addiction*, Vol. 114, No. 4, pp.662–671.
- Ahmed, U., Lin, J.C-W. and Srivastava, G. (2023) 'Graph attention network for text classification and detection of mental disorder', *ACM Transactions on the Web*, Vol. 17, No. 3, pp.1–31.
- Baltrušaitis, T., Ahuja, C. and Morency, L-P. (2018) 'Multimodal machine learning: a survey and taxonomy', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 2, pp.423–443.
- Bent, B., Wang, K., Grzesiak, E., Jiang, C., Qi, Y., Jiang, Y., Cho, P., Zingler, K., Ogbeide, F.I. and Zhao, A. (2021) 'The digital biomarker discovery pipeline: An open-source software platform for the development of digital biomarkers using mHealth and wearables data', *Journal of Clinical and Translational Science*, Vol. 5, No. 1, p.e19.
- Cohen, S. and Wills, T.A. (1985) 'Stress, social support, and the buffering hypothesis', *Psychological Bulletin*, Vol. 98, No. 2, p.310.
- Cohen, S., Kamarck, T. and Mermelstein, R. (1983) 'A global measure of perceived stress', *Journal of Health and Social Behavior*, Vol. 1, No. 1, pp.385–396.
- Grahek, I., Schaller, M. and Tackett, J.L. (2021) 'Anatomy of a psychological theory: Integrating construct-validation and computational-modeling methods to advance theorizing', *Perspectives on Psychological Science*, Vol. 16, No. 4, pp.803–815.
- Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation*, Vol. 9, No. 8, pp.1735–1780.
- Hou, H., Liu, I., Kong, F. and Ni, S. (2025) 'Computational positive psychology: advancing the science of wellbeing in the digital era', *The Journal of Positive Psychology*, Vol. 20, No. 1, pp.1–14.
- Jiang, B., Zhang, Z., Lin, D., Tang, J. and Luo, B. (2019) 'Semi-supervised learning with graph learning-convolutional networks', *Computer Vision and Pattern Recognition*, Vol. 1, No. 1, pp.11313–11320.
- Jin R. (2019) 'Analysis of EFL learners' academic anxiety based on R language and graph neural network', *International Journal of Information and Communication Technology*, Vol. 26, No. 26, pp.1–16.
- Kroenke, K., Spitzer, R.L. and Williams, J.B. (2001) 'The PHQ-9: validity of a brief depression severity measure', *Journal of General Internal Medicine*, Vol. 16, No. 9, pp.606–613.

- Li, W., Zhao, Z., Chen, D., Peng, Y. and Lu, Z. (2022) ‘Prevalence and associated factors of depression and anxiety symptoms among college students: a systematic review and meta-analysis’, *Journal of Child Psychology and Psychiatry*, Vol. 63, No. 11, pp.1222–1230.
- Liu W. (2025) ‘Career planning and pathway generation based on multimodal learning analytics’, *International Journal of Information and Communication Technology*, Vol. 26, No. 29, pp.39–58.
- Martinez-Martin, N., Insel, T.R., Dagum, P., Greely, H.T. and Cho, M.K. (2018) ‘Data mining for health: staking out the ethical territory of digital phenotyping’, *NPJ Digital Medicine*, Vol. 1, No. 1, p.68.
- Mohr, D.C., Zhang, M. and Schueller, S.M. (2017) ‘Personal sensing: understanding mental health using ubiquitous sensors and machine learning’, *Annual Review of Clinical Psychology*, Vol. 13, No. 1, pp.23–47.
- Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T. and Leiserson, C. (2020) ‘Evolvegcn: Evolving graph convolutional networks for dynamic graphs’, *Artificial Intelligence*, Vol. 34, No. 4, pp.5363–5370.
- Saeb, S., Zhang, M., Karr, C.J., Schueller, S.M., Corden, M.E., Kording, K.P. and Mohr, D.C. (2015) ‘Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study’, *Journal of Medical Internet Research*, Vol. 17, No. 7, p.e4273.
- Thoits, P.A. (2011) ‘Mechanisms linking social ties and support to physical and mental health’, *Journal of Health and Social Behavior*, Vol. 52, No. 2, pp.145–161.
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y. (2017) ‘Graph attention networks’, *Stat*, Vol. 1050, No. 20, pp.10–48550.
- Wang, C. (2023) ‘Associations dynamic evolution: evolving graph transformer’, *Anti-Fraud Engineering for Digital Finance: Behavioral Modeling Paradigm*, Vol. 1, No. 1, pp.189–207.
- Wang, L., Zhang, Y., Zhou, B., Cao, S., Hu, K. and Tan, Y. (2024) ‘Automatic depression prediction via cross-modal attention-based multi-modal fusion in social networks’, *Computers and Electrical Engineering*, Vol. 118, No. 1, p.109413.
- Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D. and Campbell, A.T. (2014) ‘StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones’, *Pervasive and Ubiquitous Computing*, Vol. 1, No. 1, pp.3–14.
- Zhao, K., Duka, B., Xie, H., Oathes, D.J., Calhoun, V. and Zhang, Y. (2022) ‘A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD’, *NeuroImage*, Vol. 246, No. 1, p.118774.
- Zhou, D., Yu, H., Yu, J., Zhao, S., Xu, W., Li, Q. and Cai, F. (2024) ‘MFDS-STGCN: predicting the behaviors of college students with fine-grained spatial-temporal activities data’, *IEEE Transactions on Emerging Topics in Computing*, Vol. 12, No. 1, pp.254–265.