Graphs to the Rescue: Revealing Hidden Relationships in Survey Data

Publisher: IEEE

Bruno M. F. de Resende ; Alexandre C. B. Delbem ; Eric K. Tokuda All Authors

Abstract:
Extracting meaningful insights from tabular data remains a fundamental challenge in machine learning, as traditional methods often struggle to capture complex feature interactions. In this work, we propose a novel graph-based approach for analyzing tabular datasets by leveraging spectral graph theory and community detection techniques. Our method represents tabular data as a weighted directed graph, where edges encode feature dependencies based on SHAP values. To enhance interpretability, we apply a sparsification technique that retains only the most significant connections. We further analyze the structural properties of the resulting graph using the deformed magnetic Laplacian, which captures directional dependencies among features. Additionally, we employ a nonparametric stochastic block model (nSBM) to uncover hierarchical modular structures and use tabular embeddings (tab2vec) to reveal fine-grained relationships in feature space. Our framework is validated on the PeNSE dataset, a large-scale survey on adolescent health, demonstrating its ability to reveal hidden structures and improve feature interpretability. Results show that spectral analysis provides an effective way to categorize features into meaningful clusters, identify redundant variables, and highlight key relationships that may be overlooked by conventional techniques. This approach offers a powerful alternative for exploring complex tabular datasets, with potential applications in various domains such as healthcare, finance, and social sciences.

Published in: 2025 International Joint Conference on Neural Networks (IJCNN)

Date of Conference: 30 June 2025 - 05 July 2025

Date Added to IEEE *Xplore*: 14 November 2025

ISBN Information:

ISSN Information:

Electronic ISSN: 2161-4407

Print on Demand(PoD) ISSN: 2161-4393

DOI: 10.1109/IJCNN64981.2025.11227576

Publisher: IEEE

Conference Location: Rome, Italy

Graphs to the Rescue: Revealing Hidden Relationships in Survey Data

Bruno M. F. de Resende
Instituto de Física de São Carlos (IFSC)
Universidade de São Paulo, Brazil

Alexandre C. B. Delbem
Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo, Brazil

Eric K. Tokuda
Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo, Brazil

2025 International Joint Conference on Neural Networks (IJCNN) | 979-8-3315-1042-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/IJCNN64981.2025.11227576

***Abstract*—Extracting meaningful insights from tabular data remains a fundamental challenge in machine learning, as traditional methods often struggle to capture complex feature interactions. In this work, we propose a novel graph-based approach for analyzing tabular datasets by leveraging spectral graph theory and community detection techniques. Our method represents tabular data as a weighted directed graph, where edges encode feature dependencies based on SHAP values. To enhance interpretability, we apply a sparsification technique that retains only the most significant connections. We further analyze the structural properties of the resulting graph using the deformed magnetic Laplacian, which captures directional dependencies among features. Additionally, we employ a nonparametric stochastic block model (nSBM) to uncover hierarchical modular structures and use tabular embeddings (*tab2vec*) to reveal fine-grained relationships in feature space. Our framework is validated on the PeNSE dataset, a large-scale survey on adolescent health, demonstrating its ability to reveal hidden structures and improve feature interpretability. Results show that spectral analysis provides an effective way to categorize features into meaningful clusters, identify redundant variables, and highlight key relationships that may be overlooked by conventional techniques. This approach offers a powerful alternative for exploring complex tabular datasets, with potential applications in various domains such as healthcare, finance, and social sciences.**

***Index Terms*—Socioeconomic surveys, Explainable ML, Spectral analysis, Feature embedding**

I. Introduction

Tabular data is a fundamental representation in machine learning, appearing in domains such as healthcare, finance, and social sciences. Despite its ubiquity, extracting meaningful insights from high-dimensional tabular datasets remains a challenge. Traditional feature selection and transformation techniques, such as principal component analysis (PCA) [1] and autoencoders [2], often struggle to capture complex, non-linear relationships between features. Recently, graph-based approaches have emerged as powerful alternatives for structuring and analyzing tabular data [3], [4].

Graph representation learning has gained significant attention in recent years due to its ability to model relational data. Methods such as graph neural networks (GNNs) [5], [6] and graph embedding techniques like node2vec [7] have demonstrated their effectiveness in capturing structured dependencies. In the context of tabular data, graphs provide a way to encode relationships between features, facilitating interpretability and dimensionality reduction [8]. Existing works have leveraged graph structures to improve feature selection [9], detect hidden correlations [10], [11], and enhance predictive models [12]. However, many of these approaches rely on predefined structures or external domain knowledge, which may not always be available.

A recent study [13] explored graph-based models in public health, constructing a graph from PeNSE variables with edges inferred via conditional dependency metrics. Their goal was to identify potential confounders in adolescent health analyses, showing how hidden dependencies could bias regression results. This study served as an important motivation for our work. While both approaches leverage graph theory to analyze tabular data, our methodology diverges in focus and technique. Rather than refining causal inference, we aim to uncover latent structure through spectral analysis. We construct a directed, weighted graph using SHAP-derived feature importance scores [14], aligning edge definitions with machine learning explainability. In contrast to traditional graph measures such as centrality or clustering coefficients, we apply spectral techniques—specifically the deformed magnetic Laplacian [15], [16]—to capture directional dependencies. We further enhance interpretability by using the nonparametric stochastic block model (nSBM) [17] for hierarchical clustering and tabular embeddings (tab2vec) to reveal fine-grained relationships.

Our methodology is validated through a case study on the PeNSE dataset, a large-scale survey on adolescent health conducted in Brazil [18]. By leveraging spectral graph analysis, we demonstrate how this approach can uncover meaningful feature clusters, identify redundant attributes, and highlight key relationships that might be overlooked by conventional techniques. Unlike previous work focused on confounder detection, our study provides a broader framework for structuring and interpreting complex tabular data through graph-based representations. Preliminary results from this line of investigation were previously reported in [11].

II. Methods

In this section, we present the methodology used to analyze tabular data through graph-based representations. The proposed approach consists of four main stages: (a) Graph Representation, where the tabular dataset is mapped to a weighted directed graph using SHAP-based feature dependencies; (b) Group Analysis, which employs spectral methods and community detection techniques to uncover structural relationships among features; (c) Individual Analysis, where centrality measures and feature embeddings (*tab2vec*) are used to assess the relevance and similarity of individual features; and (d) Multilevel Analysis, which refines the interpretability by focusing on selected subsets of features. Together, these stages provide a structured framework for uncovering hidden relationships and improving the interpretability of complex tabular datasets.

In Fig.1, we present an overview of the proposed method, which consists of four key stages: graph representation, group analysis, individual analysis, and multilevel column analysis. The process begins by constructing a weighted directed graph from the tabular dataset, where vertices correspond to columns and edges represent their relationships, weighted using SHAP values. To enhance interpretability, we apply an edge filtering technique to remove weak connections, completing the preprocessing phase and enabling further analysis of the dataset's structure.

To achieve this, we leverage spectral information extracted from the deformed magnetic Laplacian operator and employ the hierarchical modular structure derived from nSBM. The nSBM framework allows us to categorize columns into distinct groups, while the spectral information helps refine the results iteratively, focusing on a subset of the graph with increasing granularity. In addition to these group-level analyses, we incorporate techniques that assess feature relevance (using centrality measures) and represent columns as vectors in a latent space (tab2vec). In the following sections, we provide a detailed discussion of each step in this framework.

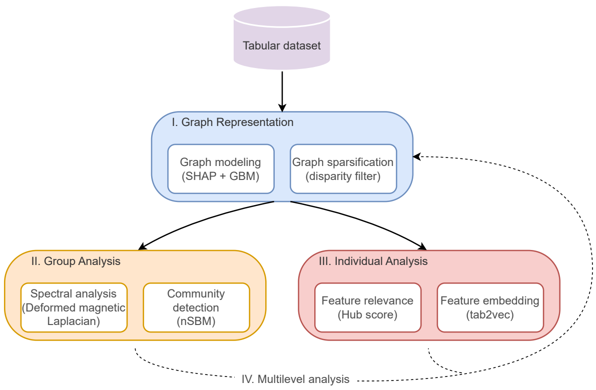


Fig. 1: Flow diagram of the proposed approach. The tabular dataset is initially mapped to a weighted directed graph. The graph is then sparsified to remove weak connections, allowing for (b) group-level analysis via spectral methods and community detection. Additionally, (c) individual-level analysis (centrality measures) and pairwise comparisons (feature embeddings) are performed. These procedures can be further refined (d) by considering only a selected subset of columns. The algorithms used at each step are listed in parentheses.

A. Graph Modeling

To analyze tabular data using a graph-based approach, we represent the dataset as a weighted directed graph. Formally, a weighted directed graph is defined as a tuple (V, E, w), where:

- V is the set of vertices, each representing a feature (column) in the dataset.
- E is the set of directed edges that capture relationships between features.
- $w : E \mapsto \mathbb{R}^+$ is a weight function that quantifies the strength of these relationships.

Constructing the Graph. Each feature (column) in the tabular dataset is mapped to at least one vertex in the graph. The edges between features are assigned weights based on their predictive influence.

1) Selecting a Target Feature: a column $c \in C$ is randomly chosen as the target variable to be predicted.
2) Predicting the Target Feature: the remaining columns serve as input features to train a gradient boosting machine (GBM) model, which estimates the values of c. The subset of features used in this prediction is denoted by $\bar{V}_c$.
3) Defining Edge Weights: once the GBM model is trained, we evaluate how much each feature contributes to predicting c. A directed edge (u, v_c) is added to the graph, where $w(u, v_c)$ represents the importance of feature u in predicting c.

This procedure is repeated for every feature in the dataset, ultimately constructing a fully connected weighted directed graph.

Computing Edge Weights. A key challenge is determining the contribution of a feature u to predicting another feature v. This contribution should reflect the predictive power of the trained GBM model.

We define the in-degree of a vertex v_c as the sum of incoming edge weights:

$$k_{in}(v_c) = \sum_{u \in \bar{V}_c} w(u, v_c) = Acc(v_c), \tag{1}$$

where $Acc(v_c)$ denotes the predictive accuracy of the GBM model for feature c. If a feature has weak or no predictive relationships with others, its in-degree will be low, reducing its influence in the graph.

The weight of an edge (u, v) is then computed as:

$$w(u, v) = Acc(v) \frac{\epsilon(u \to v)}{\sum\limits_{z \in V} \epsilon(z \to v)}. \tag{2}$$

where $\epsilon(u \to v)$ represents the contribution of feature u to the prediction of feature v.

Using SHAP Values for Edge Weights. To quantify feature contributions, various methods exist in the literature [3]. In this work, we adopt the SHapley Additive exPlanations (SHAP) method [14], a technique rooted in cooperative game theory [19] that measures the marginal contribution of each feature to the prediction task.

Since SHAP values are computed for individual instances, they provide fine-grained insights into feature relationships. However, to construct a single aggregated interpretability graph, we average SHAP values across all instances:

$$w(u,v) = Acc(v)\frac{\mathbb{E}[|\mathrm{SHAP}_i(u \to v)|]}{\sum\limits_{z\in V}\mathbb{E}[|\mathrm{SHAP}_i(z \to v)|]}. \quad (3)$$

This ensures that the edge weights reflect global feature importance rather than individual instance-specific relationships.

B. Graph sparsification

By construction, the interpretability graph is complete, which poses challenges for further processing. One major issue is the high computational cost associated with handling the entire graph. Additionally, the large number of connections can obscure meaningful patterns, making it difficult to extract relevant insights [20].

A straightforward approach to reducing the number of edges and improving interpretability in graph visualizations is to apply a naive threshold to edge weights, retaining only the strongest connections. However, selecting an appropriate threshold value is not trivial and lacks a clear justification [20]. Furthermore, this method can result in a fragmented graph with many disconnected components.

To address these challenges, various graph filtering techniques, also known as graph sparsification methods, have been developed over the past decade [20]. In this work, we adopt the disparity filter criterion introduced in [21] to selectively remove edges while preserving the structural backbone of the graph.

Let $s(u) = \sum\limits_{v\in V|(u,v)\in E} w(u,v)$ denote the out-degree of a feature associated with node u in the interpretability graph. This value quantifies the total contribution of feature u in explaining the outputs of other features. The relative importance of an edge (u,v) is given by $p(u,v) = w(u,v)/s(u)$, which measures how much feature u contributes to predicting feature v relative to its total explanatory power. Using this, we define an edge filtering criterion based on the disparity filter:

$$w_\alpha(u,v) = 1 - (k_{out}(u) - 1) \int\limits_0^{p(u,v)} (1-x)^{k_{out}(u)-2}\mathrm{d}x. \quad (4)$$

Edges with w_α exceeding a given threshold $\alpha \in [0,1]$ are removed. This method enables edge filtering while preserving the key structural relationships in the graph, ensuring that the backbone of the network remains intact [21].

C. Spectral Analysis

Once the interpretability graph is constructed, we can analyze its structure to uncover meaningful feature relationships. One powerful approach is to study the spectral properties of the *magnetic Laplacian*, which helps reveal clusters of interdependent features.

From Directed to Undirected Graph Representation. Since the interpretability graph is directed and weighted, we begin by decomposing the edge weight function into $w_s(u,v)$ and $w_a(u,v)$, symmetric and asymmetric $w_a(u,v)$ components respectively, capturing mutual relationships between features and directional dependencies:

$$w_s(u,v) = \frac{w(u,v)+w(v,u)}{2}, \quad w_a(u,v) = \frac{w(u,v)-w(v,u)}{2}. \quad (5)$$

Using this decomposition, we define the *flow function* at vertex v due to u as:

$$a(v,u) = 2w_a(u,v). \quad (6)$$

This transformation allows us to construct an undirected counterpart of the original directed graph, denoted as $G_s = (V, E_s, w_s)$.

Combinatorial Laplacian. The undirected graph G_s is associated with the *combinatorial Laplacian* operator L, which is defined as:

$$(Lf)(u) = f(u)d(u) - \sum_{v\in V} w_s(u,v)f(v), \quad (7)$$

where $d(u) = \sum\limits_{v\in V} w_s(u,v)$ represents the degree of vertex u.

Since L is symmetric, it provides valuable insights into the structure of undirected graphs. However, it does not incorporate the directional nature of feature relationships. To address this, we introduce the *magnetic Laplacian*, which incorporates phase perturbations.

Introducing Directionality. To retain directionality in the spectral analysis, we modify the combinatorial Laplacian by introducing a phase perturbation to edge weights:

$$\gamma_q(u,v) = e^{2\pi \mathrm{i} q a(v,u)}. \quad (8)$$

This phase term encodes directional dependencies into the spectral representation. Substituting this into Eq. (7), we obtain the magnetic Laplacian $\mathcal{L}_q$:

$$(\mathcal{L}_q f)(u) = f(u)d(u) - \sum_{v\in V} w_s(u,v)\gamma_q(u,v)f(v), \quad (9)$$

where $q \in [0,1]$ is a parameter known as the *charge* [22], controlling the influence of directionality.

Normalized Magnetic Laplacian. For practical analysis, we define a normalized version of the magnetic Laplacian, $\mathcal{H}_q$, given by:

$$(\mathcal{H}_q f)(u) = f(u) - \frac{\sum\limits_v w_s(u,v)\gamma_q(u,v)f(v)}{d(u)}. \quad (10)$$

Unlike the standard combinatorial Laplacian, the magnetic Laplacian is represented by a Hermitian matrix [23], making it particularly useful for spectral analysis. Additionally, it is a *positive semi-definite operator*, meaning its eigenvalues and eigenvectors can be leveraged to analyze graph structure.

Spectral Interpretation: Feature Clustering. The eigenvectors of the normalized magnetic Laplacian $\mathcal{H}_q$ provide valuable insights into the organization of features:

Circular Dependencies and Group Synchronization. The eigenvector corresponding to the smallest eigenvalue of $\mathcal{H}_q$ helps approximate a group synchronization problem, capturing cyclic dependencies in feature interactions [24]. Mathematically, this problem minimizes:

$$\eta_c(\theta) = \frac{1}{2\mathrm{vol}(G_s)} \sum_{u,v \in V} w_s(u,v) \left| e^{\mathrm{i}\theta(u)} - \gamma_q(u,v) e^{\mathrm{i}\theta(v)} \right|^2, \quad (11)$$

where $\mathrm{vol}(G_s) = \sum_{u \in V} d(u)$ represents the total degree sum of the graph.

Graph Partitioning via Eigenvector Phases. The phase angles of eigenvectors, denoted as $\mathbf{v}_q^{(l)} \in \mathbb{C}^{|V|}$, reveal natural partitions within the dataset. The second smallest eigenvector of $\mathcal{H}_q$ provides *an approximate solution to a graph-cut problem*, helping to identify clusters of strongly related features [24], [25].

These spectral properties enable an interpretable decomposition of feature relationships, uncovering structures that conventional methods may overlook.

D. Community Detection

Features with similar interpretability characteristics should naturally form communities within the interpretability graph. To explore these relationships effectively, it is crucial to determine a robust method for community identification. One traditional approach is modularity optimization [26], but it has limitations, including the tendency to detect communities even in random graphs [27], leading to unreliable feature groupings.

To address this issue, we adopt the nested Stochastic Block Model (nSBM) [28], a non-parametric Bayesian approach that hierarchically clusters graph communities. Unlike the standard Stochastic Block Model (SBM) [29], which partitions graphs into predefined groups, nSBM constructs a hierarchy of nested communities, improving the detection of small-scale structures [28].

Mathematically, SBM applies Bayesian inference to estimate graph partitions by considering block sizes and intra- and inter-block connection probabilities. Let b represent the partitioning of vertices and θ denote the parameters of the generative model for a given graph G.

By leveraging nSBM's hierarchical framework, we can efficiently infer the modular organization of the graph, uncovering intricate feature relationships. This allows us to better understand the dataset's structure and identify meaningful groups of interrelated variables.

E. Feature Relevance

Beyond understanding feature communities, it is also essential to quantify the importance of individual features within the dataset. Inspired by [30], we assess feature importance using centrality measures. Various centrality metrics exist [31] and each provides a different perspective on feature relevance within the interpretability graph. In this work, we focus on hub and authority scores due to their interpretability and computational efficiency.

Originally introduced for ranking web pages [32], hub and authority scores distinguish between two types of nodes: *hubs*, which serve as connectors, and *authorities*, which represent key informational sources. In the context of our interpretability graph, highly ranked hubs indicate features strongly linked to other important features, while high-authority nodes represent the most influential features in prediction tasks.

Unlike the force-directed layout and nSBM, which require graph sparsification, hub and authority scores can be efficiently computed on the complete interpretability graph. This ensures that we retain all relationships during the analysis, offering a more comprehensive evaluation of feature relevance without discarding potentially useful connections.

F. Feature Embedding (tab2vec).

While community detection groups features into broad clusters, it does not quantify their pairwise similarity. To capture fine-grained relationships, we employ node2vec [7], generating a low-dimensional vector representation for each feature, referred to as *tab2vec*.

The interpretability graph is used as input, where node2vec simulates biased random walks to learn structural dependencies. The transition probabilities are controlled by two hyperparameters: p, which biases the walk toward local neighbors, and q, which encourages exploration of distant nodes. The sampled walks are then used to train a skip-gram model [33], optimizing feature embeddings so that features appearing in similar graph contexts have similar representations.

Each feature is mapped to a d-dimensional vector $\mathbf{z}_i \in \mathbb{R}^d$, where similarity between features u and v is computed using cosine similarity:

$$\text{Similarity}(u,v) = \frac{\mathbf{z}_u \cdot \mathbf{z}_v}{\|\mathbf{z}_u\| \|\mathbf{z}_v\|}. \quad (12)$$

These embeddings improve feature selection by identifying redundant variables, enhance visualization, and support downstream tasks such as clustering and anomaly detection. Unlike PCA, which captures variance, tab2vec retains relational structure, making it well-suited for structured data analysis.

G. Multilevel analysis

The previous analyses can be repeated for a selected subset of columns. This step is conceptually similar to graph filtering, as it aims to refine the interpretability of feature relationships by reducing complexity. However, it differs in approach: instead of filtering edges within an existing graph, it constructs a new graph based on a subset of columns, effectively modifying the prediction task by excluding certain features.

Unlike simply extracting a subgraph from the original interpretability graph, this method ensures that only relevant features are considered in the new analysis. To define this subset of columns, we leverage the magnetic eigenmaps of the interpretability graph, which allow for feature separation as described in [24]. This approach is analogous to using the eigenvectors of the combinatorial Laplacian for image segmentation [25].

III. CASE STUDY: PENSE

Adolescence is a formative stage that strongly influences adult life, prompting extensive research into adolescent health and behavior [34]. The PeNSE (National Survey of Scholar's Health) [35], conducted by IBGE with the Ministries of Health and Education, offers detailed insights into the health and risk factors of Brazilian teenagers.

Carried out in 2009, 2012, and 2015, PeNSE surveyed 9th-grade students—typically around fourteen years old—from public and private schools, following international ethical standards. We focus on the 2015 edition, which collected responses from nearly 130,000 students nationwide.

The electronic questionnaire covered domains such as socioeconomic context, parental education, mental and oral health, eating habits, family life, and school conditions.

Prior research has explored PeNSE from diverse angles. For instance,[36], [37] addressed issues like bullying and chronic illness.

Force-directed layout and the effect of the disparity filter

We first explore how our method can reveal groups of related questions in the PeNSE survey. To achieve this, we construct the interpretability graph following the previously described approach and apply the disparity filter to remove weaker edges.

In Fig.2, we present force-directed visualizations of both the complete graph (Fig.2(a)) and the sparsified graph obtained after applying edge filtering (Fig.2(b)). The complete graph exhibits a *hairy-ball* structure, which hinders direct interpretation. However, after applying the disparity filter with a threshold of $\alpha = 0.1$, underlying group structures become apparent. A visual inspection suggests that questions related to physical activity form two distinct clusters.

Despite the usefulness of force-directed layouts, it is well known that their interpretation can be subjective. Therefore, any observations made from these visualizations should be validated using more rigorous analytical methods. In the following sections, we further investigate the clustering behavior of these features using spectral analysis and community detection.

Hierarchical categorization of the features

Community detection is a challenging problem, partly because there is no universally agreed-upon definition of what constitutes a community [28]. The nSBM approach addresses

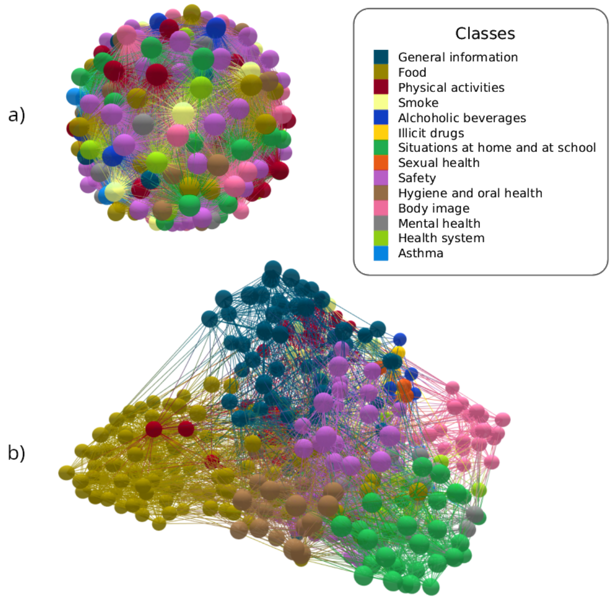


Fig. 2: Interpretability graph of the PeNSE dataset. Nodes represent features, and edges indicate relationships between feature pairs based on our proposed approach. In (b), the graph was filtered using a disparity filter, as defined in (2), with a parameter of 0.1. The node layout follows a force-directed algorithm, with vertex and edge colors corresponding to feature groups in the dataset. A strong agreement between *spatial* communities and predefined *categories* is visible, such as the brown group in the lower section of the figure.

this challenge by providing a statistically principled method to infer modular structures. In this work, we utilize the graph-tool[1] implementation of nSBM [29], [38].

Fig.3 presents a circular visualization of the filtered interpretability graph from the PeNSE survey, as inferred by nSBM. The gray vertices and edges represent the hierarchical structure of the detected communities, with vertex positions determined by the modular structure of the graph. The color of each node and edge corresponds to the predefined class of the respective question in the survey, as originally assigned by the survey designers. Consequently, communities where vertices share the same color indicate alignment between the inferred modular structure and the survey's original classification.

This hierarchical visualization (Fig.3) allows for multiple analyses, with two being particularly relevant to our study. First, we examine how the detected communities align with the divisions proposed in the survey. Second, we investigate the connections between different feature groups, identifying dominant clusters and their interrelations.

In Fig.3, we observe a strong correspondence between the inferred and predefined groupings for at least two categories: *Food* (mustard) and *Body Image* (pink). The *Safety* (violet)

[1]https://graph-tool.skewed.de/

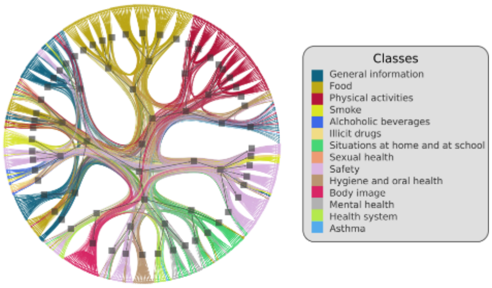
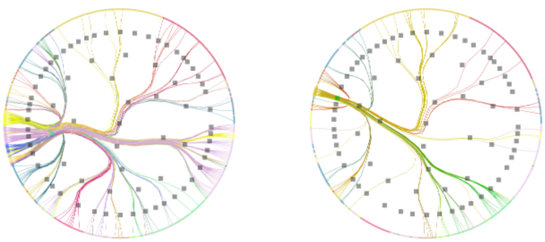


Fig. 3: Circular visualization [28] of the filtered interpretability graph with edge bundling. The outer vertices represent features, while directed edges illustrate relationships between them. Nodes are grouped according to the inferred modular structure, and their colors correspond to their respective categories in the survey. The overlaid hierarchical structure reveals the community hierarchy.

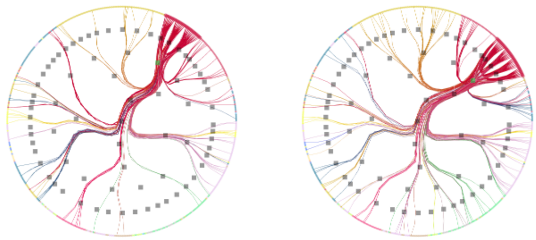
category also exhibits reasonable alignment, though some features are positioned separately on the left side of the circle, forming a cluster with questions related to drug use (Fig.4(a)). This suggests that an alternative classification could categorize these features under *Illicit Drugs*. It is important to emphasize that the nSBM approach is entirely data-driven and non-subjective, relying solely on the observed response patterns in the survey.



(a) (b)

Fig. 4: Distinct feature groupings in the hierarchical community structure. In (a), a subset of safety-related features is positioned separately, suggesting a more refined categorization. In (b), the highlighted orange features have strong connections to the green vertices, reflecting an expected relationship based on the survey's original classification.

Fig.4(b) highlights a small cluster in mustard, positioned on the left side of the circle. This group exhibits strong connectivity with the green cluster at the bottom. The orange category corresponds to *Food*, with the highlighted nodes representing questions about eating meals with parents. The nodes at the bottom, corresponding to the category *Situations at home and at school*, in contrast, relate to questions about the respondent's relationship with their parents. This suggests a strong correlation between family interactions and shared meals, potentially revealing an alternative way to categorize these questions in the survey.



(a) (b)

Fig. 5: Questions originally classified under *Physical Activities*. Although located near each other in the circular layout, they are divided into two distinct subgroups. In (a), they pertain to recreational sports, while in (b), they are related to mobility constraints driven by socioeconomic factors, such as walking or cycling to school.

The hierarchical nature of nSBM enables a more granular categorization of features. As shown in Fig.5, most questions related to *Physical Activities* (in wine) are positioned within the same region but are subdivided into two distinct groups. Upon closer inspection, we observe that the first group (Fig.5(a)) consists of recreational activities, such as playing football or dancing. The second group (Fig.5(b)) consists of mobility-related activities, such as walking or cycling to school, which are often dictated by socioeconomic conditions. The following questions, with the highest hub scores in the community highlighted in Fig.5(b), while categorized under *Physical Activities*, are strongly linked to socioeconomic conditions.

- *During the last 7 days, on how many days did you walk or ride a bicycle to school?*
- *During the last 7 days, on how many days did you return from school on foot or by bicycle?*
- *When you travel to school on foot or by bicycle, how long does the journey take?*
- *When you return from school on foot or by bicycle, how long does the journey take?*

This pattern aligns with findings in developing countries, where mobility choices are often influenced by economic factors [39].

Features with a similar interpretation structure as revealed by a tab2vec approach

The hierarchical structure inferred from nSBM provides a mesoscale view of feature relationships, enabling the grouping of related questions and assessing their level of association. However, for certain tasks—such as detecting potential data leakage or investigating specific factors—it is useful to identify features with highly similar roles in the dataset.

One way to address this is by embedding features into a vector space using a word-embedding approach, such as node2vec [40]. Table I shows an example where we applied node2vec to the interpretability graph and computed cosine similarities between feature embeddings. The table lists the top four questions most similar to *"At school, have you* *ever received pregnancy prevention counseling?"*. Notably, the question *"At school, have you ever received advice on how to get condoms for free?"* has a cosine similarity of 0.99, indicating that their embeddings are nearly identical. This result is intuitive, as discussions on pregnancy prevention often include information about condoms. Such findings suggest that embedding-based approaches can be useful for detecting relationships between survey questions.

TABLE I: Most similar questions (cosine similarity) to *"At school, have you ever received pregnancy prevention counseling?"*.

Sim.	**Question**
0.99	*At school, have you ever received advice on how to get condoms for free?*
0.98	*At school, have you ever received advice about AIDS or other sexually transmitted diseases?*
0.87	*Have you heard about the vaccination campaign against the HPV virus?*
0.52	*In the last twelve months, how many times did you get into a physical fight?*

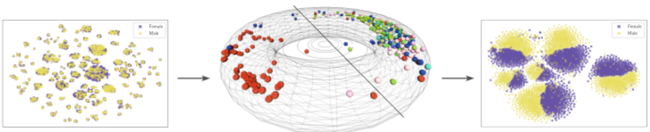


Fig. 6: Embedding of the rows from the PeNSE survey, color-coded based on responses to the survey question *What is your gender?*. Meanwhile, the images at the left display the UMAP embedding, weighted by the hub score of the interpretability graph that incorporates all the survey questions. The image at the right show the embedding results for a subset of questions, excluding those from the left of the figure of the toroidal embedding, and obtained without considering the physical activity-related questions.

Zooming In on Feature Sets: Figure 6 shows a UMAP projection of survey respondents, colored by their answers to a specific question. This embedding was computed using cosine similarity between rows and all available features. Notably, the projection exhibits no clear clustering, suggesting that irrelevant or noisy features may be diluting meaningful structure. One strategy to enhance the quality of such embeddings is to weight the features based on their relevance—an approach successfully applied in previous work [41] using feature importance scores. Here, we use the hub score derived from the interpretability graph as a feature weighting scheme. However, as shown on the left of Figure 6, this strategy alone still results in many small, fragmented clusters. In contrast, by segmenting the features based on the toroidal space induced by the interpretability graph and reapplying UMAP using only one of these feature subsets, we obtain more coherent and semantically meaningful clusters—particularly with respect to the gender variable. This illustrates how focused feature selection, informed by graph structure, can substantially improve the expressiveness of embedding spaces.

IV. Conclusions

In this work, we introduced a novel framework for analyzing tabular data by leveraging graph-based representations, spectral analysis, and community detection techniques. By constructing a weighted directed graph where edges encode predictive relationships between features, our approach provides a structured way to uncover hidden patterns in high-dimensional datasets. The integration of spectral methods, particularly the deformed magnetic Laplacian, allowed us to extract meaningful insights from directional dependencies. Furthermore, the application of nonparametric stochastic block modeling (nSBM) revealed hierarchical structures, while feature embeddings through tab2vec helped capture fine-grained relationships among variables.

The case study on the PeNSE dataset demonstrated the effectiveness of this methodology. The results highlighted how spectral graph analysis can refine interpretability, revealing clusters of features that traditional techniques may overlook. The sparsification process improved the clarity of feature interactions while preserving the structural integrity of the data. Additionally, the embedding-based analysis provided an alternative perspective, identifying redundant or highly correlated variables that could inform feature selection or dimensionality reduction.

Despite its advantages, our method has some limitations. The reliance on SHAP values for defining graph edges means that the results are influenced by the predictive power and potential biases of the underlying model. Additionally, while the disparity filter effectively reduces noise, it requires careful parameter tuning to balance interpretability and information retention. Future work will explore the extension of this approach to dynamic datasets where feature relationships evolve over time, the use of alternative weighting schemes such as mutual information or causal inference methods, and its application to other domains, including healthcare, finance, and social sciences, to assess its generalizability.

Overall, our results suggest that spectral graph analysis provides a powerful toolkit for enhancing the interpretability of tabular data. By structuring feature interactions as a graph and leveraging spectral techniques, we offer a new perspective on data analysis that uncovers complex relationships hidden under traditional paradigms.

Acknowledgment

The authors of this work would like to thank the Center for Artificial Intelligence (C4A-IUSP) and the support from the São Paulo Research Foundation (FAPESP grant #2019/07665-4, #2020/16578-5) and from the IBM Corporation; the CEPID-CeMEAI/ICMC-USP (CEPID, FAPESP grant #2013/07375-0); the FAPESP grant #2024/16994-0; CNPq (grant 307085/2018-0); and CAPES. The authors also appreciate the suggestions and fruitful discussions with Joao Ricardo Sato, Suzana de S. Santos, Filipi N. Silva and Thomas Peron. The authors also want to disclose that ChatGPT (OpenAI) was used for language editing and grammar enhancement throughout the manuscript. All intellectual content and interpretations are solely those of the authors.

REFERENCES

[1] I. Jolliffe, *Principal Component Analysis*. Springer, 2002.

[2] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[3] C. Molnar, *Interpretable Machine Learning*, 2019.

[4] M. A. Serrano, M. Boguna, and A. Vespignani, "Extracting the multiscale backbone of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6483–6488, 2009.

[5] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.

[6] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2017.

[7] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 855–864.

[8] Y. Liu, J. Wu, C. Yang, and J. Wu, "Elastic graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2357–2368, 2021.

[9] P. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu, "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.

[10] X. Wang, A. De, S. Sra, and S. Jegelka, "Learning robust representations with graph denoising policy networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20 865–20 879, 2021.

[11] B. M. F. de Resende, E. K. Tokuda, and L. d. F. Costa, "Unraveling the graph structure of tabular data through bayesian and spectral analysis," *arXiv preprint arXiv:2110.01421*, 2021.

[12] Y. Feng, H. You, Z. Lin, R. Cheng, Q. Liu, and X. Chang, "Graph random neural networks for semi-supervised learning on graphs," *Advances in Neural Information Processing Systems*, vol. 33, pp. 22 092–22 103, 2020.

[13] A. Y. Ambriola Oku, G. A. Zimeo Morais, A. P. Arantes Bueno, A. Fujita, and J. R. Sato, "Potential confounders in the analysis of brazilian adolescent's health: a combination of machine learning and graph theory," *International journal of environmental research and public health*, vol. 17, no. 1, p. 90, 2020.

[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 4768–4777.

[15] M. Fanuel, C. M. Alaíz, Ángela Fernández, and J. A. Suykens, "Magnetic eigenmaps for the visualization of directed networks," *Applied and Computational Harmonic Analysis*, vol. 44, no. 1, pp. 189–199, 2018.

[16] B. M. F. de Resende and L. da F. Costa, "Characterization and comparison of large directed networks through the spectra of the magnetic laplacian," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 7, p. 073141, 2020.

[17] T. P. Peixoto, "Nonparametric bayesian inference of the microcanonical stochastic block model," *Physical Review E*, vol. 95, no. 1, p. 012317, 2017.

[18] M. M. de Oliveira, M. O. Campos, M. A. R. de Andreazzi, and D. C. Malta, "Characteristics of the national adolescent school-based health survey-pense, brazil," *Epidemiologia e Serviços de Saúde*, vol. 26, pp. 605–616, 2017.

[19] H. W. Kuhn and A. W. Tucker, Eds., *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press, Dec. 1953.

[20] M. Hashemi, S. Gong, J. Ni, W. Fan, B. A. Prakash, and W. Jin, "A comprehensive survey on graph reduction: Sparsification, coarsening, and condensation," *arXiv preprint arXiv:2402.03358*, 2024.

[21] M. A. Serrano, M. Boguna, and A. Vespignani, "Extracting the multiscale backbone of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6483–6488, Apr. 2009.

[22] M. A. Shubin, "Discrete Magnetic Laplacian," *Communications in Mathematical Physics*, vol. 164, no. 2, pp. 259–275, Aug. 1994.

[23] B. M. F. de Resende and L. da F. Costa, "Characterization and comparison of large directed networks through the spectra of the magnetic Laplacian," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 7, p. 073141, Jul. 2020.

[24] M. Fanuel, C. M. Alaíz, Á. Fernández, and J. A. Suykens, "Magnetic Eigenmaps for the visualization of directed networks," *Applied and Computational Harmonic Analysis*, vol. 44, no. 1, pp. 189–199, Jan. 2018.

[25] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Comput. Soc, 1997, p. 731–737.

[26] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, Feb 2004.

[27] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, "Modularity from fluctuations in random graphs and complex networks," *Physical Review E*, vol. 70, no. 2, p. 025101, Aug 2004.

[28] T. P. Peixoto, "Hierarchical Block Structures and High-Resolution Model Selection in Large Networks," *Physical Review X*, vol. 4, no. 1, p. 011047, Mar. 2014.

[29] ——, "Nonparametric bayesian inference of the microcanonical stochastic block model," *Physical Review E*, vol. 95, no. 1, p. 012317, 2017.

[30] A. Y. Ambriola Oku, G. A. Zimeo Morais, A. P. Arantes Bueno, A. Fujita, and J. R. Sato, "Potential Confounders in the Analysis of Brazilian Adolescent's Health: A Combination of Machine Learning and Graph Theory," *International Journal of Environmental Research and Public Health*, vol. 17, no. 1, p. 90, Dec. 2019.

[31] M. Newman, *Networks*. Oxford university press, 2018.

[32] J. M. Kleinberg *et al.*, "Authoritative sources in a hyperlinked environment." in *SODA*, vol. 98. Citeseer, 1998, pp. 668–677.

[33] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2013.

[34] J. A. Grunbaum, L. Kann, S. Kinchen, J. Ross, J. Hawkins, R. Lowry, W. A. Harris, T. McManus, D. Chyen, and J. Collins, "Youth Risk Behavior Surveillance - United States, 2003 (Abridged)," *Journal of School Health*, vol. 74, no. 8, pp. 307–324, Oct. 2004.

[35] M. M. de Oliveira, M. O. Campos, M. A. R. de Andreazzi, and D. C. Malta, "Characteristics of the National adolescent school-based health survey-PeNSE, Brazil," *Epidemiologia e Serviços de Saúde*, vol. 26, pp. 605–616, 2017.

[36] D. C. Malta, M. A. I. Silva, F. C. M. de Mello, R. A. Monteiro, L. M. V. Sardinha, C. Crespo, M. G. O. de Carvalho, M. M. A. da Silva, and D. L. Porto, "Bullying in Brazilian schools: Results from the National School-based Health Survey (PeNSE), 2009," *Ciência & Saúde Coletiva*, vol. 15, pp. 3065–3076, Oct. 2010.

[37] D. C. Malta, M. A. R. de Andreazzi, M. Oliveira-Campos, S. S. C. d. A. Andrade, N. N. B. de Sá, L. de Moura, A. J. R. Dias, C. D. Crespo, and J. B. da Silva Júnior, "Trend of the risk and protective factors of chronic diseases in adolescents, National Adolescent School-based Health Survey (PeNSE 2009 e 2012)," *Revista Brasileira de Epidemiologia*, vol. 17, no. 1, pp. 77–91, 2014.

[38] T. P. Peixoto, "Merge-split markov chain monte carlo for community detection," *Physical Review E*, vol. 102, no. 1, p. 012305, Jul 2020.

[39] A. N. R. Da Silva, M. da Silva Costa, and M. H. Macedo, "Multiple views of sustainable urban mobility: The case of brazil," *Transport Policy*, vol. 15, no. 6, pp. 350–360, 2008.

[40] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug 2016, p. 855–864.

[41] B. I. Grisci, M. J. Krause, and M. Dorn, "Relevance aggregation for neural networks interpretability and knowledge discovery on tabular data," vol. 559, p. 111–129, Jun 2021.

SECTION I.

Introduction

Tabular data is a fundamental representation in machine learning, appearing in domains such as healthcare, finance, and social sciences. Despite its ubiquity, extracting meaningful insights from high-dimensional tabular datasets remains a challenge. Traditional feature selection and transformation techniques, such as principal component analysis (PCA) [1] and autoencoders [2], often struggle to capture complex, non-linear relationships between features. Recently, graphbased approaches have emerged as powerful alternatives for structuring and analyzing tabular data [3], [4].

Graph representation learning has gained significant attention in recent years due to its ability to model relational data. Methods such as graph neural networks (GNNs) [5], [6] and graph embedding techniques like node2vec [7] have demonstrated their effectiveness in capturing structured dependencies. In the context of tabular data, graphs provide a way to encode relationships between features, facilitating interpretability and dimensionality reduction [8]. Existing works have leveraged graph structures to improve feature selection [9], detect hidden correlations [10], [11], and enhance predictive models [12]. However, many of these approaches rely on predefined structures or external domain knowledge, which may not always be available.

A recent study [13] explored graph-based models in public health, constructing a graph from PeNSE variables with edges inferred via conditional dependency metrics. Their goal was to identify potential confounders in adolescent health analyses, showing how hidden dependencies could bias regression results. This study served as an important motivation for our work. While both approaches leverage graph theory to analyze tabular data, our methodology diverges in focus and technique. Rather than refining causal inference, we aim to uncover latent structure through spectral analysis. We construct a directed, weighted graph using SHAP-derived feature importance scores [14], aligning edge definitions with machine learning explainability. In contrast to traditional graph measures such as centrality or clustering coefficients, we apply spectral techniques—specifically the deformed magnetic Laplacian [15], [16]—to capture directional dependencies. We further enhance interpretability by using the nonparametric stochastic block model (nSBM) [17] for hierarchical clustering and tabular embeddings (tab2vec) to reveal fine-grained relationships.

Our methodology is validated through a case study on the PeNSE dataset, a large-scale survey on adolescent health conducted in Brazil [18]. By leveraging spectral graph analysis, we demonstrate how this approach can uncover meaningful feature clusters, identify redundant attributes, and highlight key relationships that might be overlooked by conventional techniques. Unlike previous work focused on confounder detection, our study provides a broader framework for structuring and interpreting complex tabular data through graph-based representations. Preliminary results from this line of investigation were previously reported in [11].

SECTION II.
Methods

In this section, we present the methodology used to analyze tabular data through graph-based representations. The proposed approach consists of four main stages: (a) Graph Representation, where the tabular dataset is mapped to a weighted directed graph using SHAP-based feature dependencies; (b) Group Analysis, which employs spectral methods and community detection techniques to uncover structural relationships among features; (c) Individual Analysis, where centrality measures and feature embeddings (*tab2vec*) are used to assess the relevance and similarity of individual features; and (d) Multilevel Analysis, which refines the interpretability by focusing on selected subsets of features. Together, these stages provide a structured framework for uncovering hidden relationships and improving the interpretability of complex tabular datasets.

In Fig. 1, we present an overview of the proposed method, which consists of four key stages: graph representation, group analysis, individual analysis, and multilevel column analysis. The process begins by constructing a weighted directed graph from the tabular dataset, where vertices correspond to columns and edges represent their relationships, weighted using SHAP values. To enhance interpretability, we apply an edge filtering technique to remove weak connections, completing the preprocessing phase and enabling further analysis of the dataset's structure.

To achieve this, we leverage spectral information extracted from the deformed magnetic Laplacian operator and employ the hierarchical modular structure derived from nSBM. The nSBM framework allows us to categorize columns into distinct groups, while the spectral information helps refine the results iteratively, focusing on a subset of the graph with increasing granularity. In addition to these group-level analyses, we incorporate techniques that assess feature relevance (using centrality measures) and represent columns as vectors in a latent space (tab2vec). In the following sections, we provide a detailed discussion of each step in this framework.

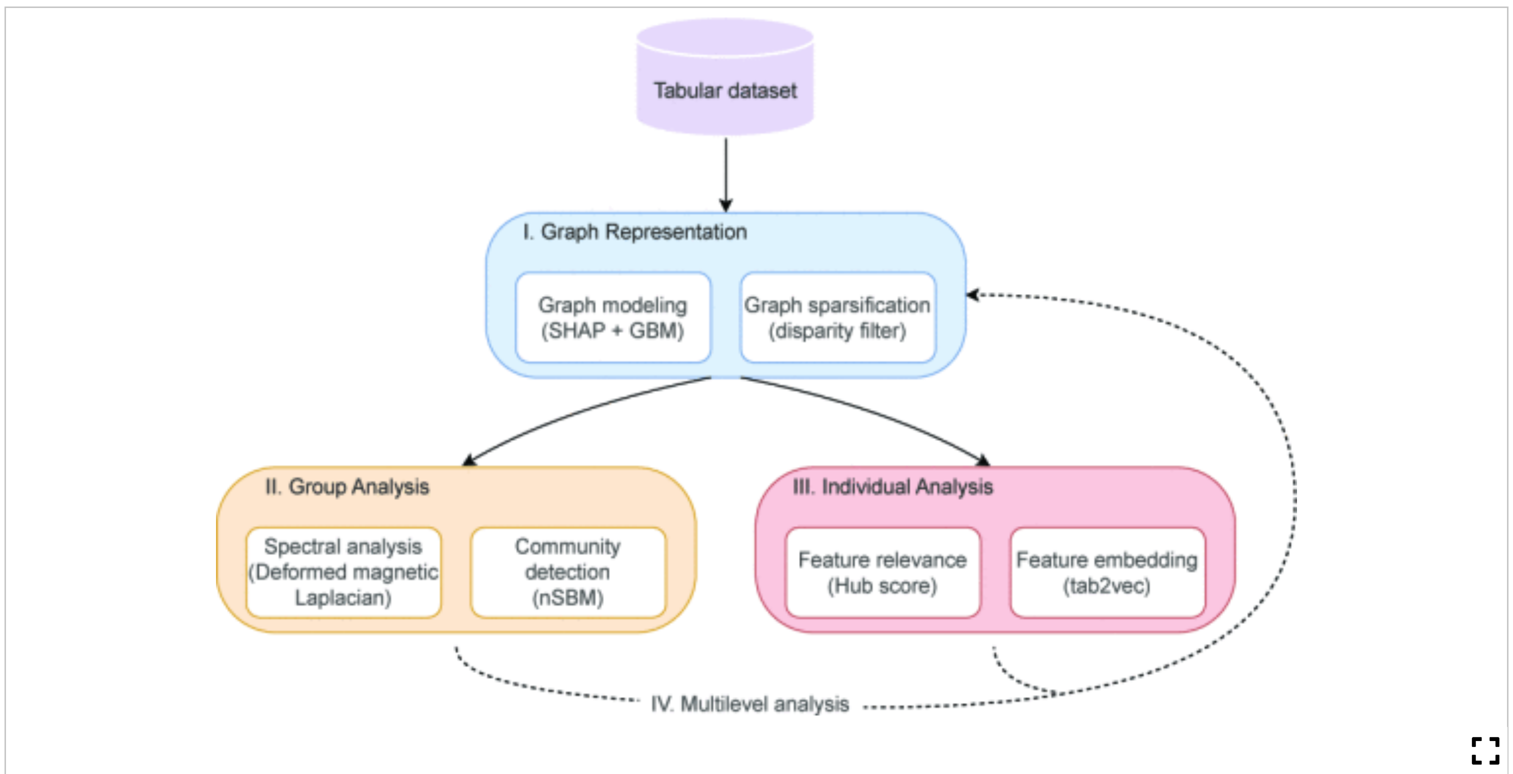


Fig. 1:
Flow diagram of the proposed approach. The tabular dataset is initially mapped to a weighted directed graph. The graph is then sparsified to remove weak connections, allowing for (b) group-

level analysis via spectral methods and community detection. Additionally, (c) individual-level analysis (centrality measures) and pairwise comparisons (feature embeddings) are performed. These procedures can be further refined (d) by considering only a selected subset of columns. The algorithms used at each step are listed in parentheses.

A. Graph Modeling

To analyze tabular data using a graph-based approach, we represent the dataset as a weighted directed graph. Formally, a weighted directed graph is defined as a tuple (V,E,w), where:

- V is the set of vertices, each representing a feature (column) in the dataset.
- E is the set of directed edges that capture relationships between features.
- $w: E \mapsto \mathbb{R}^{+}$ is a weight function that quantifies the strength of these relationships.

Constructing the Graph. Each feature (column) in the tabular dataset is mapped to at least one vertex in the graph. The edges between features are assigned weights based on their predictive influence.

1) Selecting a Target Feature: a column $c \in C$ is randomly chosen as the target variable to be predicted.

2) Predicting the Target Feature: the remaining columns serve as input features to train a gradient boosting machine (GBM) model, which estimates the values of c. The subset of features used in this prediction is denoted by $\bar{V}_c$.

3) Defining Edge Weights: once the GBM model is trained, we evaluate how much each feature contributes to predicting c. A directed edge (u,v_c) is added to the graph, where $w(u,v_c)$ represents the importance of feature u in predicting c.

This procedure is repeated for every feature in the dataset, ultimately constructing a fully connected weighted directed graph.

Computing Edge Weights. A key challenge is determining the contribution of a feature u to predicting another feature v. This contribution should reflect the predictive power of the trained GBM model.

We define the in-degree of a vertex v_c as the sum of incoming edge weights:

$$k_{in}\left(v_c\right) = \sum_{u \in \bar{V}_c} w\left(u, v_c\right) = Acc\left(v_c\right), \tag{1}$$

View Source

where $Acc(v_c)$ denotes the predictive accuracy of the GBM model for feature c. If a feature has weak or no predictive relationships with others, its in-degree will be low, reducing its influence in the graph.

The weight of an edge (u,v) is then computed as:

$$w(u,v) = Acc(v)\frac{\varepsilon(u \to v)}{\sum\limits_{z \in V} \varepsilon(z \to v)}. \tag{2}$$

View Source

where $\epsilon(u \to v)$ represents the contribution of feature u to the prediction of feature v.

Using SHAP Values for Edge Weights. To quantify feature contributions, various methods exist in the literature [3]. In this work, we adopt the SHapley Additive exPlanations (SHAP) method [14], a technique rooted in cooperative game theory [19] that measures the marginal contribution of each feature to the prediction task.

Since SHAP values are computed for individual instances, they provide fine-grained insights into feature relationships. However, to construct a single aggregated interpretability graph, we average SHAP values across all instances:

$$w(u,v) = Acc(v)\frac{\mathbb{E}\left[|\mathrm{SHAP}_i(u \to v)|\right]}{\sum\limits_{z \in V} \mathbb{E}\left[|\mathrm{SHAP}_i(z \to v)|\right]}. \tag{3}$$

View Source

This ensures that the edge weights reflect global feature importance rather than individual instance-specific relationships.

B. Graph sparsification

By construction, the interpretability graph is complete, which poses challenges for further processing. One major issue is the high computational cost associated with handling the entire graph. Additionally, the large number of connections can obscure meaningful patterns, making it difficult to extract relevant insights [20].

A straightforward approach to reducing the number of edges and improving interpretability in graph visualizations is to apply a naive threshold to edge weights, retaining only the strongest connections. However, selecting an appropriate threshold value is not trivial and lacks a clear justification [20]. Furthermore, this method can result in a fragmented graph with many disconnected components.

To address these challenges, various graph filtering techniques, also known as graph sparsification methods, have been developed over the past decade [20]. In this work, we adopt the disparity filter criterion introduced in [21] to selectively remove edges while preserving the structural backbone of the graph.

Let $s(u) = \sum\limits_{v \in V|(u,v) \in E} w(u,v)$ denote the out-degree of a feature associated with node u in the interpretability graph. This value quantifies the total contribution of feature u in explaining the outputs of other features. The relative importance of an edge (u,v) is given by $p(u,v) = w(u,v)/s(u)$, which measures how much feature u contributes to predicting feature v relative to its total explanatory power. Using this, we define an edge filtering criterion based on the disparity filter:

$$w_\alpha(u,v) = 1 - (k_{out}(u) - 1)\int_0^{p(u,v)} (1-x)^{k_{out}(u)-2} dx. \tag{4}$$

View Source

Edges with w_α exceeding a given threshold $\alpha \in [0, 1]$ are removed. This method enables edge filtering while preserving the key structural relationships in the graph, ensuring that the backbone of the network remains intact [21].

C. Spectral Analysis

Once the interpretability graph is constructed, we can analyze its structure to uncover meaningful feature relationships. One powerful approach is to study the spectral properties of the *magnetic Laplacian*, which helps reveal clusters of interdependent features.

From Directed to Undirected Graph Representation. Since the interpretability graph is directed and weighted, we begin by decomposing the edge weight function into $w_s(u,v)$ and $w_a(u,v)$, symmetric and asymmetric $w_a(u,v)$ components respectively, capturing mutual relationships between features and directional dependencies:

$$w_s(u,v) = \frac{w(u,v) + w(v,u)}{2}, \quad w_a(u,v) = \frac{w(u,v) - w(v,u)}{2}. \tag{5}$$

View Source

Using this decomposition, we define the *flow function* at vertex v due to u as:

$$a(v,u) = 2w_a(u,v). \tag{6}$$

View Source

This transformation allows us to construct an undirected counterpart of the original directed graph, denoted as $G_s = (V, E_s, w_s)$.

Combinatorial Laplacian. The undirected graph G_s is associated with the *combinatorial Laplacian* operator L, which is defined as:

$$(Lf)(u) = f(u)d(u) - \sum_{v \in V} w_s(u,v)f(v), \tag{7}$$

View Source

where $d(u) = \sum_{v \in V} w_s(u,v)$ represents the degree of vertex u.

Since L is symmetric, it provides valuable insights into the structure of undirected graphs. However, it does not incorporate the directional nature of feature relationships. To address this, we introduce the *magnetic Laplacian*, which incorporates phase perturbations.

Introducing Directionality. To retain directionality in the spectral analysis, we modify the combinatorial Laplacian by introducing a phase perturbation to edge weights:

$$\gamma_q(u,v) = e^{2\pi \mathrm{i} q a(v,u)}. \tag{8}$$

View Source

This phase term encodes directional dependencies into the spectral representation. Substituting this into Eq. (7), we obtain the magnetic Laplacian $\mathscr{L}_q$:

$$\left(\mathcal{L}_q f\right)(u) = f(u)d(u) - \sum_{v \in V} w_s(u,v)\gamma_q(u,v)f(v), \tag{9}$$

View Source

where $q \in [0, 1]$ is a parameter known as the *charge* [22], controlling the influence of directionality.

Normalized Magnetic Laplacian. For practical analysis, we define a normalized version of the magnetic Laplacian, $\mathcal{H}_q$, given by:

$$\left(\mathcal{H}_q f\right)(u) = f(u) - \frac{\sum_{v} w_s(u,v)\gamma_q(u,v)f(v)}{d(u)}. \tag{10}$$

View Source

Unlike the standard combinatorial Laplacian, the magnetic Laplacian is represented by a Hermitian matrix [23], making it particularly useful for spectral analysis. Additionally, it is a *positive semi-definite operator*, meaning its eigenvalues and eigenvectors can be leveraged to analyze graph structure.

Spectral Interpretation: Feature Clustering. The eigenvectors of the normalized magnetic Laplacian H_q provide valuable insights into the organization of features:

Circular Dependencies and Group Synchronization. The eigenvector corresponding to the smallest eigenvalue of $\mathcal{H}_q$ helps approximate a group synchronization problem, capturing cyclic dependencies in feature interactions [24]. Mathematically, this problem minimizes:

$$\eta_c(\theta) = \frac{1}{2\,\mathrm{vol}(G_s)} \sum_{u,v \in V} w_s(u,v)\left|e^{\mathrm{i}\theta(u)} - \gamma_q(u,v)e^{\mathrm{i}\theta(v)}\right|^2, \tag{11}$$

View Source

where $\mathrm{vol}(G_s) = \sum_{u \in V} d(u)$ represents the total degree sum of the graph.

Graph Partitioning via Eigenvector Phases. The phase angles of eigenvectors, denoted as $\mathbf{v}_q^{(l)} \in \mathbb{C}^{|V|}$, reveal natural partitions within the dataset. The second smallest eigenvector of $\mathcal{H}_q$ provides *an approximate solution to a graphcut problem*, helping to identify clusters of strongly related features [24], [25].

These spectral properties enable an interpretable decomposition of feature relationships, uncovering structures that conventional methods may overlook.

D. Community Detection

Features with similar interpretability characteristics should naturally form communities within the interpretability graph. To explore these relationships effectively, it is crucial to determine a robust method for community identification. One traditional approach is modularity optimization [26], but it has limitations, including the tendency to detect communities even in random graphs [27], leading to unreliable feature groupings.

To address this issue, we adopt the nested Stochastic Block Model (nSBM) [28], a non-parametric Bayesian approach that hierarchically clusters graph communities. Unlike the standard Stochastic Block Model (SBM) [29], which partitions graphs into predefined groups, nSBM constructs a hierarchy of nested communities, improving the detection of small-scale structures [28].

Mathematically, SBM applies Bayesian inference to estimate graph partitions by considering block sizes and intra- and inter-block connection probabilities. Let b represent the partitioning of vertices and θ denote the parameters of the generative model for a given graph G.

By leveraging nSBM's hierarchical framework, we can efficiently infer the modular organization of the graph, uncovering intricate feature relationships. This allows us to better understand the dataset's structure and identify meaningful groups of interrelated variables.

E. Feature Relevance

Beyond understanding feature communities, it is also essential to quantify the importance of individual features within the dataset. Inspired by [30], we assess feature importance using centrality measures. Various centrality metrics exist [31] and each provides a different perspective on feature relevance within the interpretability graph. In this work, we focus on hub and authority scores due to their interpretability and computational efficiency.

Originally introduced for ranking web pages [32], hub and authority scores distinguish between two types of nodes: *hubs*, which serve as connectors, and *authorities*, which represent key informational sources. In the context of our interpretability graph, highly ranked hubs indicate features strongly linked to other important features, while high-authority nodes represent the most influential features in prediction tasks.

Unlike the force-directed layout and nSBM, which require graph sparsification, hub and authority scores can be efficiently computed on the complete interpretability graph. This ensures that we retain all relationships during the analysis, offering a more comprehensive evaluation of feature relevance without discarding potentially useful connections.

F. Feature Embedding (tab2vec)

While community detection groups features into broad clusters, it does not quantify their pairwise similarity. To capture fine-grained relationships, we employ node2vec [7], generating a low-dimensional vector representation for each feature, referred to as *tab2vec*.

The interpretability graph is used as input, where node2vec simulates biased random walks to learn structural dependencies. The transition probabilities are controlled by two hyperparameters: p, which biases the walk toward local neighbors, and q, which encourages exploration of distant nodes. The sampled walks are then used to train a skip-gram model [33], optimizing feature embeddings so that features appearing in similar graph contexts have similar representations.

Each feature is mapped to a d-dimensional vector $\mathbf{z}_i \in \mathbb{R}^d$, where similarity between features u and v is computed using cosine similarity:

$$\text{Similarity}(u, v) = \frac{\mathbf{z}_u \cdot \mathbf{z}_v}{\|\mathbf{z}_u\| \, \|\mathbf{z}_v\|}. \tag{12}$$

View Source

These embeddings improve feature selection by identifying redundant variables, enhance visualization, and support downstream tasks such as clustering and anomaly detection. Unlike PCA, which captures variance, tab2vec retains relational structure, making it well-suited for structured data analysis.

G. Multilevel analysis

The previous analyses can be repeated for a selected subset of columns. This step is conceptually similar to graph filtering, as it aims to refine the interpretability of feature relationships by reducing complexity. However, it differs in approach: instead of filtering edges within an existing graph, it constructs a new graph based on a subset of columns, effectively modifying the prediction task by excluding certain features.

Unlike simply extracting a subgraph from the original interpretability graph, this method ensures that only relevant features are considered in the new analysis. To define this subset of columns, we leverage the magnetic eigenmaps of the interpretability graph, which allow for feature separation as described in [24]. This approach is analogous to using the eigenvectors of the combinatorial Laplacian for image segmentation [25].

SECTION III.
Case study: PeNSE

Adolescence is a formative stage that strongly influences adult life, prompting extensive research into adolescent health and behavior [34]. The PeNSE (National Survey of Scholar's Health) [35], conducted by IBGE with the Ministries of Health and Education, offers detailed insights into the health and risk factors of Brazilian teenagers.

Carried out in 2009, 2012, and 2015, PeNSE surveyed 9thgrade students—typically around fourteen years old—from public and private schools, following international ethical standards. We focus on the 2015 edition, which collected responses from nearly 130,000 students nationwide.

The electronic questionnaire covered domains such as socioeconomic context, parental education, mental and oral health, eating habits, family life, and school conditions.

Prior research has explored PeNSE from diverse angles. For instance,[36], [37] addressed issues like bullying and chronic illness.

Force-directed layout and the effect of the disparity filter

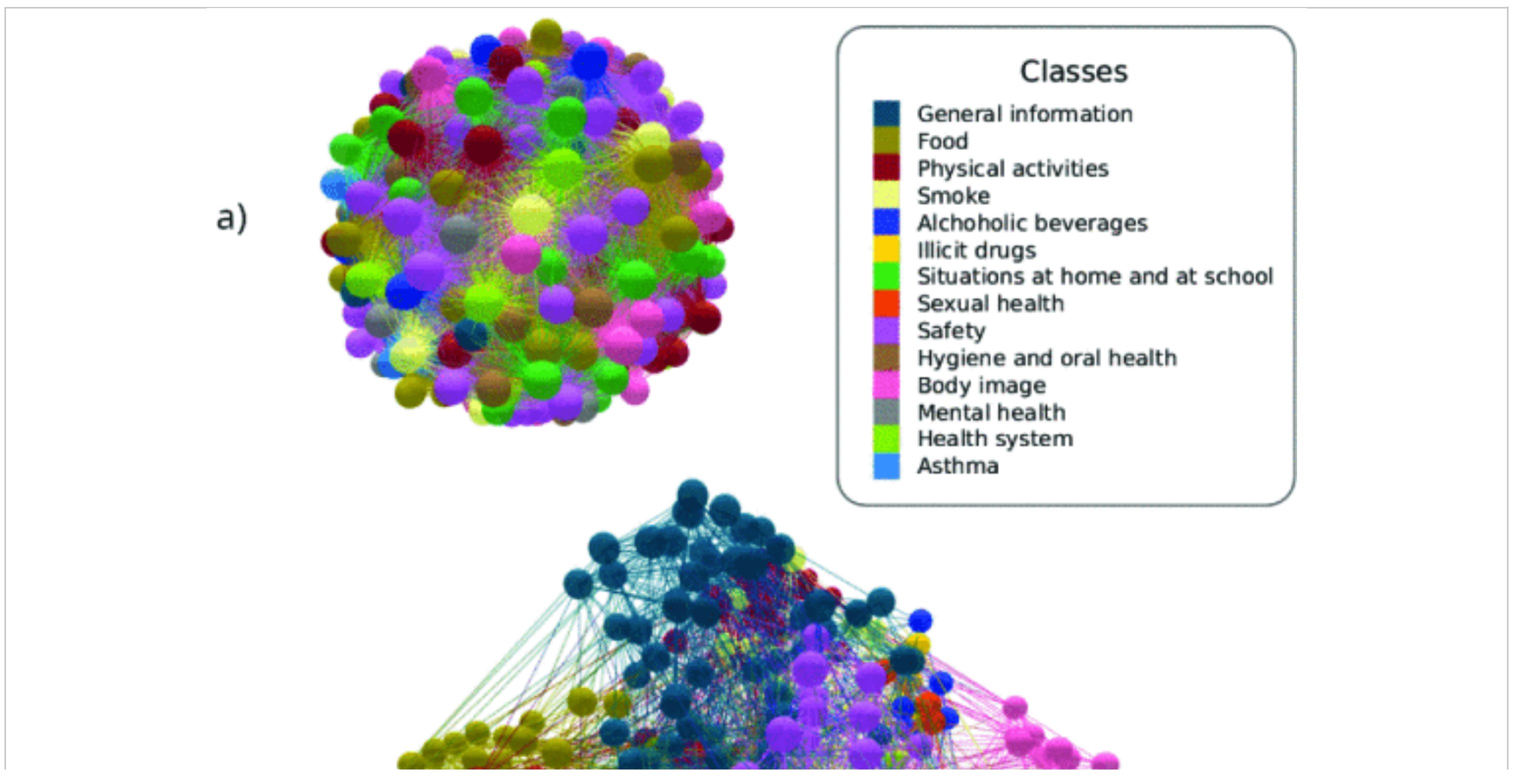
We first explore how our method can reveal groups of related questions in the PeNSE survey. To achieve this, we construct the interpretability graph following the previously described approach and apply the disparity filter to remove weaker edges.

In Fig. 2, we present force-directed visualizations of both the complete graph (Fig. 2(a)) and the sparsified graph obtained after applying edge filtering (Fig. 2(b)). The complete graph exhibits a *hairy-ball* structure, which hinders direct interpretation. However, after applying the disparity filter with a threshold of $\alpha = 0.1$, underlying group structures become apparent. A visual inspection suggests that questions related to physical activity form two distinct clusters.

Despite the usefulness of force-directed layouts, it is well known that their interpretation can be subjective. Therefore, any observations made from these visualizations should be validated using more rigorous analytical methods. In the following sections, we further investigate the clustering behavior of these features using spectral analysis and community detection.

Hierarchical categorization of the features

Community detection is a challenging problem, partly because there is no universally agreed-upon definition of what constitutes a community [28]. The nSBM approach addresses this challenge by providing a statistically principled method to infer modular structures. In this work, we utilize the graphtool [1] implementation of nSBM [29], [38].



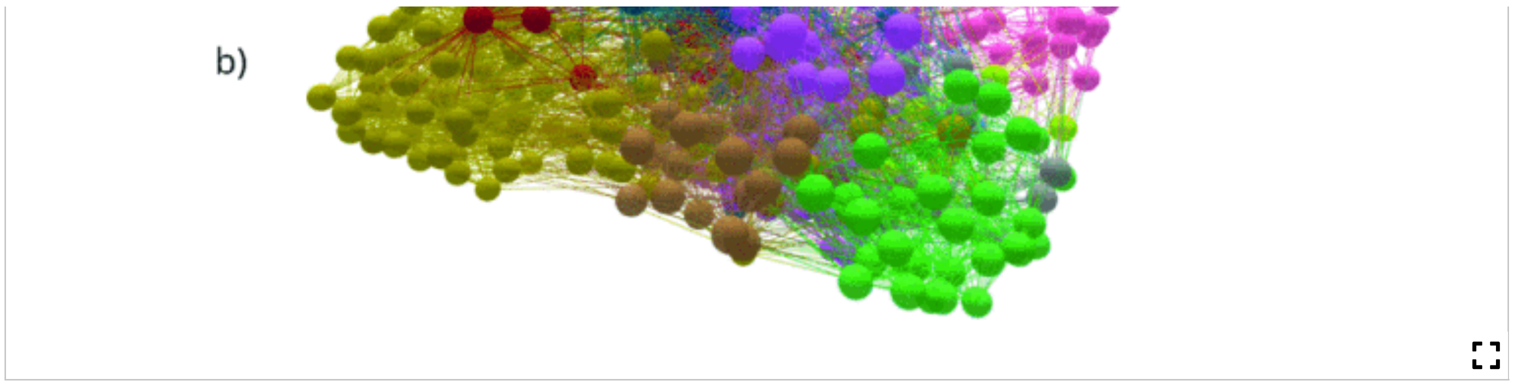


Fig. 2:
Interpretability graph of the PeNSE dataset. Nodes represent features, and edges indicate relationships between feature pairs based on our proposed approach. In (b), the graph was filtered using a disparity filter, as defined in (2), with a parameter of 0.1. The node layout follows a force-directed algorithm, with vertex and edge colors corresponding to feature groups in the dataset. A strong agreement between *spatial* communities and predefined *categories* is visible, such as the brown group in the lower section of the figure.

Fig. 3 presents a circular visualization of the filtered interpretability graph from the PeNSE survey, as inferred by nSBM. The gray vertices and edges represent the hierarchical structure of the detected communities, with vertex positions determined by the modular structure of the graph. The color of each node and edge corresponds to the predefined class of the respective question in the survey, as originally assigned by the survey designers. Consequently, communities where vertices share the same color indicate alignment between the inferred modular structure and the survey's original classification.

This hierarchical visualization (Fig. 3) allows for multiple analyses, with two being particularly relevant to our study. First, we examine how the detected communities align with the divisions proposed in the survey. Second, we investigate the connections between different feature groups, identifying dominant clusters and their interrelations.

In Fig. 3, we observe a strong correspondence between the inferred and predefined groupings for at least two categories: *Food* (mustard) and *Body Image* (pink). The *Safety* (violet) category also exhibits reasonable alignment, though some features are positioned separately on the left side of the circle, forming a cluster with questions related to drug use (Fig. 4(a)). This suggests that an alternative classification could categorize these features under *Illicit Drugs*. It is important to emphasize that the nSBM approach is entirely data-driven and non-subjective, relying solely on the observed response patterns in the survey.

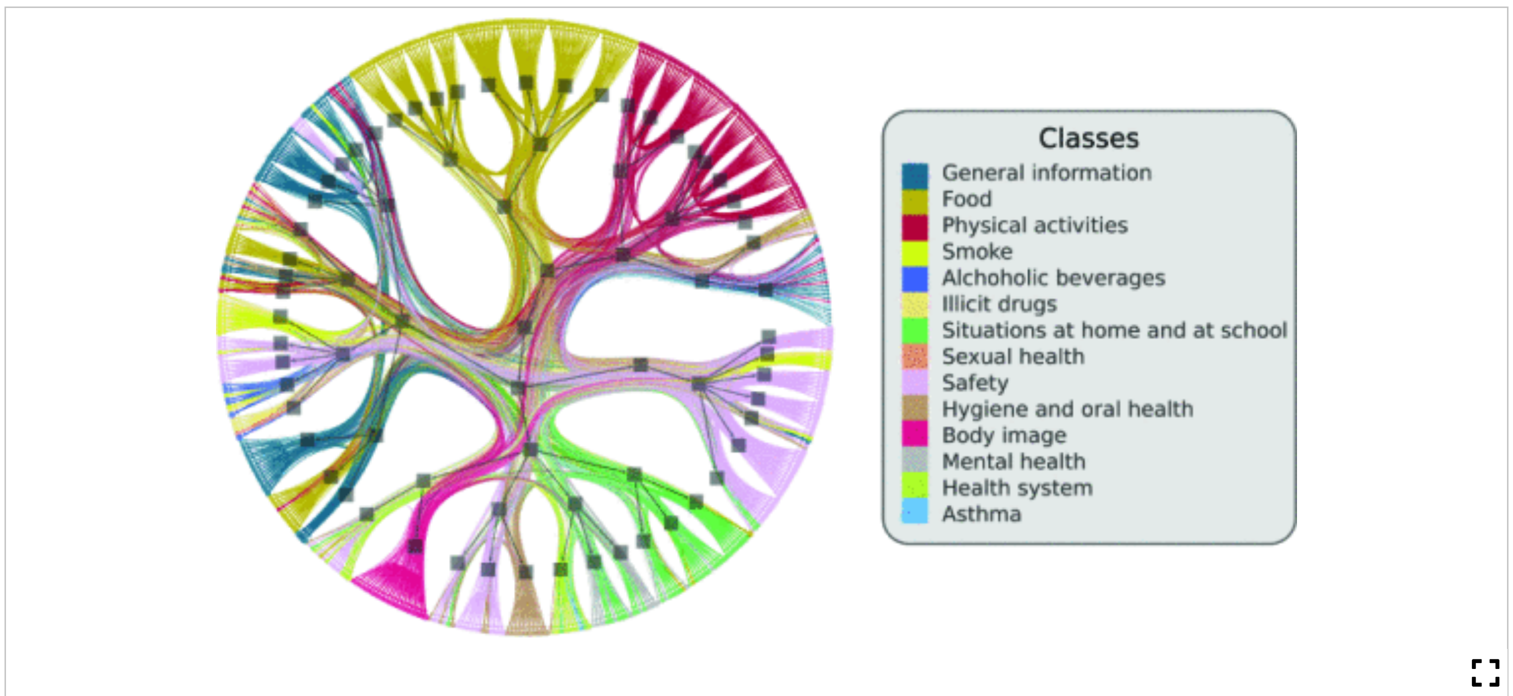
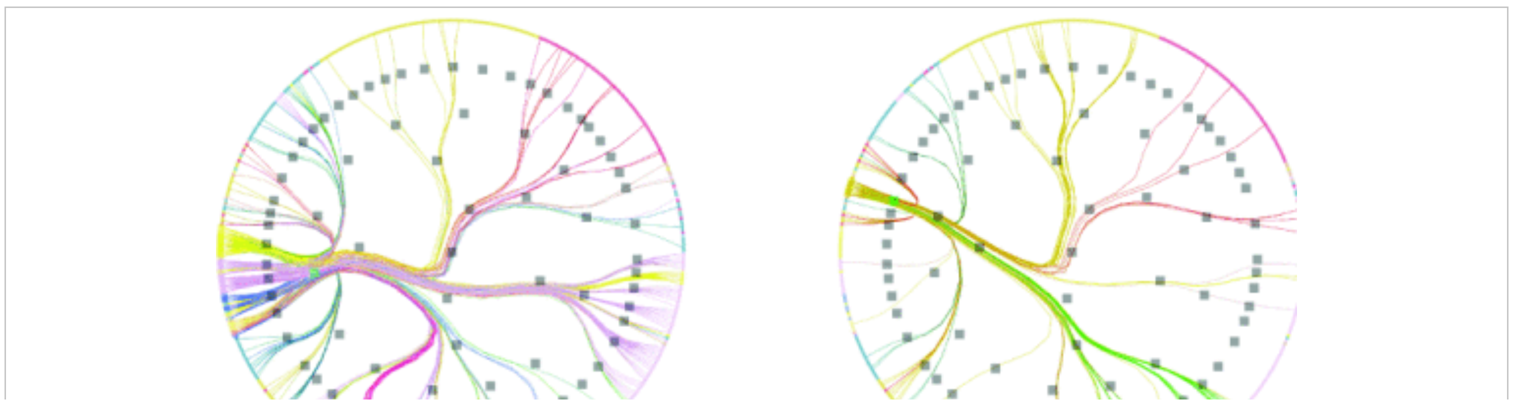


Fig. 3:
Circular visualization [28] of the filtered interpretability graph with edge bundling. The outer vertices represent features, while directed edges illustrate relationships between them. Nodes are grouped according to the inferred modular structure, and their colors correspond to their respective categories in the survey. The overlaid hierarchical structure reveals the community hierarchy.



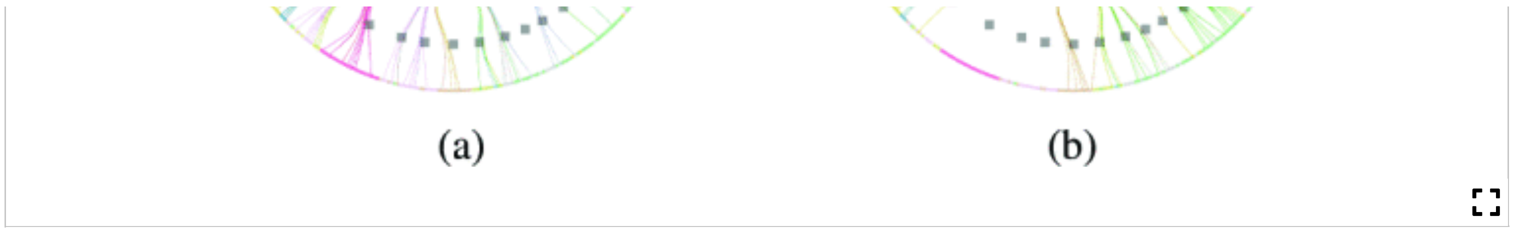


Fig. 4:
Distinct feature groupings in the hierarchical community structure. In (a), a subset of safety-related features is positioned separately, suggesting a more refined categorization. In (b), the highlighted orange features have strong connections to the green vertices, reflecting an expected relationship based on the survey's original classification.

Fig. 4(b) highlights a small cluster in mustard, positioned on the left side of the circle. This group exhibits strong connectivity with the green cluster at the bottom. The orange category corresponds to *Food*, with the highlighted nodes representing questions about eating meals with parents. The nodes at the bottom, corresponding to the category *Situations at home and at school*, in contrast, relate to questions about the respondent's relationship with their parents. This suggests a strong correlation between family interactions and shared meals, potentially revealing an alternative way to categorize these questions in the survey.

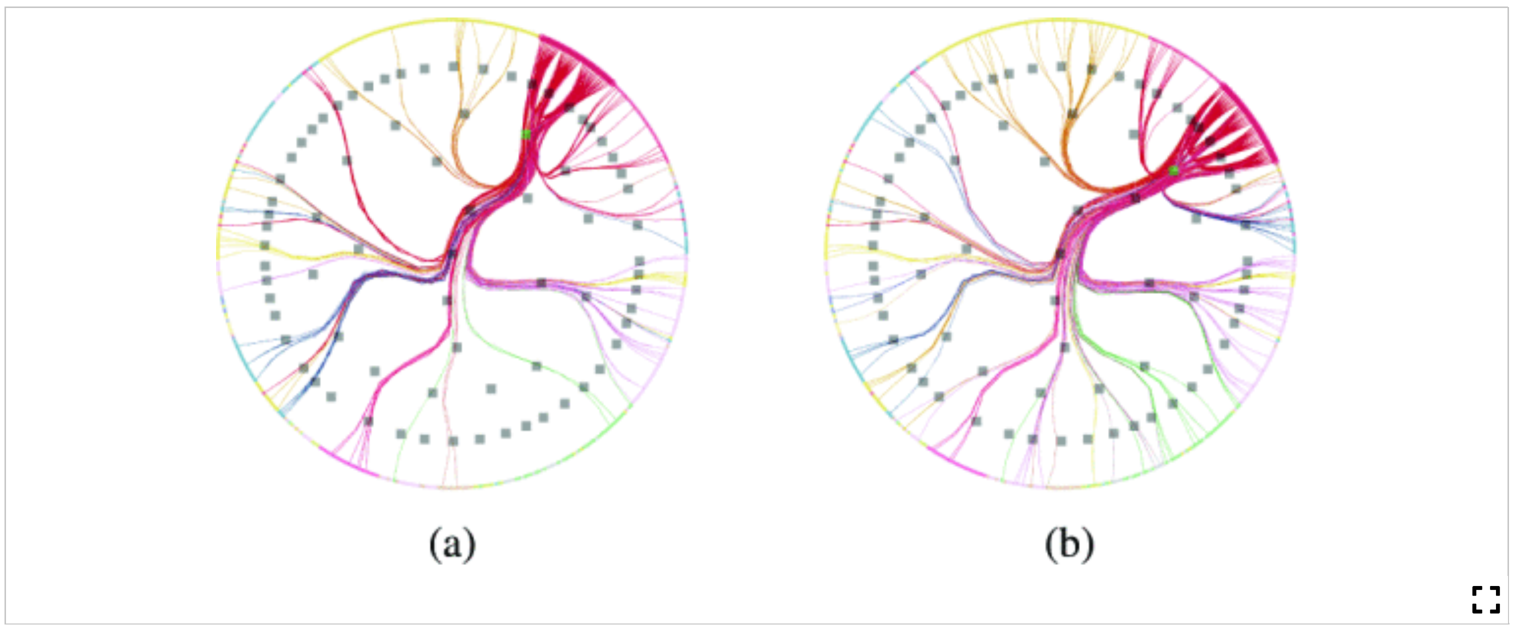


Fig. 5:
Questions originally classified under *Physical Activities.* Although located near each other in the circular layout, they are divided into two distinct subgroups. In (a), they pertain to recreational sports, while in (b), they are related to mobility constraints driven by socioeconomic factors, such as walking or cycling to school.

The hierarchical nature of nSBM enables a more granular categorization of features. As shown in Fig. 5, most questions related to *Physical Activities* (in wine) are positioned within the same region but are subdivided into two distinct groups. Upon closer inspection, we observe that the first group (Fig. 5(a)) consists of recreational activities, such as playing football or dancing. The second group (Fig. 5(b)) consists of mobility-related activities, such as walking or cycling to school, which are often dictated by socioeconomic conditions. The following questions, with the highest hub scores in the community highlighted in Fig. 5(b), while categorized under *Physical Activities*, are strongly linked to socioeconomic conditions.

- *During the last 7 days, on how many days did you walk or ride a bicycle to school?*
- *During the last 7 days, on how many days did you return from school on foot or by bicycle?*
- *When you travel to school on foot or by bicycle, how long does the journey take?*
- *When you return from school on foot or by bicycle, how long does the journey take?*

This pattern aligns with findings in developing countries, where mobility choices are often influenced by economic factors [39].

Features with a similar interpretation structure as revealed by a tab2vec approach

The hierarchical structure inferred from nSBM provides a mesoscale view of feature relationships, enabling the grouping of related questions and assessing their level of association. However, for certain tasks—such as detecting potential data leakage or investigating specific factors—it is useful to identify features with highly similar roles in the dataset.

One way to address this is by embedding features into a vector space using a word-embedding approach, such as node2vec [40]. Table I shows an example where we applied node2vec to the interpretability graph and computed cosine similarities between feature embeddings. The table lists the top four questions most similar to *“At school, have you ever received pregnancy prevention counseling?”*. Notably, the question *“At school, have you ever received advice on how to get condoms for free?”* has a cosine similarity of 0.99, indicating that their embeddings are nearly identical. This result is intuitive, as discussions on pregnancy prevention often include information about condoms. Such findings suggest that embedding-based approaches can be useful for detecting relationships between survey questions.

TABLE I: Most similar questions (cosine similarity) to “At school, have you ever received pregnancy prevention counseling?”.

Sim.	Question
0.99	*At school, have you ever received advice on how to get condoms for free?*
0.98	*At school, have you ever received advice about AIDS or other sexually transmitted diseases?*
0.87	*Have you heard about the vaccination campaign against the HPV virus?*
0.52	*In the last twelve months, how many times did you get into a physical fight?*

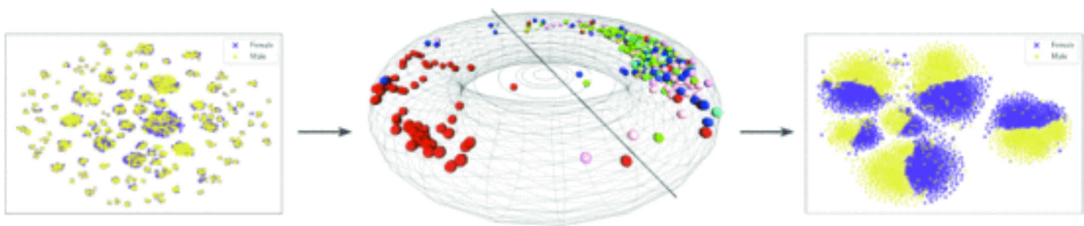


Fig. 6:
Embedding of the rows from the PeNSE survey, colorcoded based on responses to the survey question *What is your gender?.* Meanwhile, the images at the left display the UMAP embedding, weighted by the hub score of the interpretability graph that incorporates all the survey questions. The image at the right show the embedding results for a subset of questions, excluding those from the left of the figure of the toroidal embedding, and obtained without considering the physical activity-related questions.

Zooming In on Feature Sets: Figure 6 shows a UMAP projection of survey respondents, colored by their answers to a specific question. This embedding was computed using cosine similarity between rows and all available features. Notably, the projection exhibits no clear clustering, suggesting that irrelevant or noisy features may be diluting meaningful structure. One strategy to enhance the quality of such embeddings is to weight the features based on their relevance—an approach successfully applied in previous work [41] using feature importance scores. Here, we use the hub score derived from the interpretability graph as a feature weighting scheme. However, as shown on the left of Figure 6, this strategy alone still results in many small, fragmented clusters. In contrast, by segmenting the features based on the toroidal space induced by the interpretability graph and reapplying UMAP using only one of these feature subsets, we obtain more coherent and semantically meaningful clusters—particularly with respect to the gender variable. This illustrates how focused feature selection, informed by graph structure, can substantially improve the expressiveness of embedding spaces.

SECTION IV.
Conclusions

In this work, we introduced a novel framework for analyzing tabular data by leveraging graph-based representations, spectral analysis, and community detection techniques. By constructing a weighted directed graph where edges encode predictive relationships between features, our approach provides a structured way to uncover hidden patterns in high-dimensional datasets. The integration of spectral methods, particularly the deformed magnetic Laplacian, allowed us to extract meaningful insights from directional dependencies. Furthermore, the application of nonparametric stochastic block modeling (nSBM) revealed hierarchical structures, while feature embeddings through tab2vec helped capture fine-grained relationships among variables.

The case study on the PeNSE dataset demonstrated the effectiveness of this methodology. The results highlighted how spectral graph analysis can refine interpretability, revealing clusters of features that traditional techniques may overlook. The sparsification process improved the clarity of feature interactions while preserving the structural integrity of the data. Additionally, the embedding-based analysis provided an alternative perspective, identifying redundant or highly correlated variables that could inform feature selection or dimensionality reduction.

Despite its advantages, our method has some limitations. The reliance on SHAP values for defining graph edges means that the results are influenced by the predictive power and potential biases of the underlying model. Additionally, while the disparity filter effectively reduces noise, it requires careful parameter tuning to balance interpretability and information retention. Future work will explore the extension of this approach to dynamic datasets where feature relationships evolve over time, the use of alternative weighting schemes such as mutual information or causal inference methods, and its application to other domains, including healthcare, finance, and social sciences, to assess its generalizability.

Overall, our results suggest that spectral graph analysis provides a powerful toolkit for enhancing the interpretability of tabular data. By structuring feature interactions as a graph and leveraging spectral techniques, we offer a new perspective on data analysis that uncovers complex relationships hidden under traditional paradigms.

ACKNOWLEDGMENT

The authors of this work would like to thank the Center for Artificial Intelligence (C4A-IUSP) and the support from the São Paulo Research Foundation (FAPESP grant #2019/07665-4, #2020/16578-5) and from the IBM Corporation; the CEPID-CeMEAI/ICMC-USP (CEPID, FAPESP grant#2013/07375-0); the FAPESP grant #2024/16994-0; CNPq (grant 307085/2018-0); and CAPES. The authors also appreciate the suggestions and fruitful discussions with Joao Ricardo Sato, Suzana de S. Santos, Filipi N. Silva and Thomas Peron. The authors also want to disclose that ChatGPT (OpenAI) was used for language editing and grammar enhancement throughout the manuscript. All intellectual content and interpretations are solely those of the authors.

Authors

Bruno M. F. de Resende
Instituto de Física de São Carlos (IFSC), Universidade de São Paulo, Brazil

Alexandre C. B. Delbem
Instituto de Ciências Matemáticas e de Computação (ICMC), Universidade de São Paulo, Brazil

Eric K. Tokuda
Instituto de Ciências Matemáticas e de Computação (ICMC), Universidade de São Paulo, Brazil

Figures

References

Keywords

Metrics

Footnotes