Check for updates

DOI: 10.1111/cgf.70440

COMPUTER GRAPHICS forum

Eurographics Conference on Visualization (EuroVis) 2026
N. Andrienko, M. Chen, and B. Wang
(Guest Editors)

HamCat: Ego-Centric Relationship Exploration for Multidimensional Categorical Data

H. Balaka[1], H. Hauser[1], and L. A. Garrison[1]

[1] University of Bergen, Norway

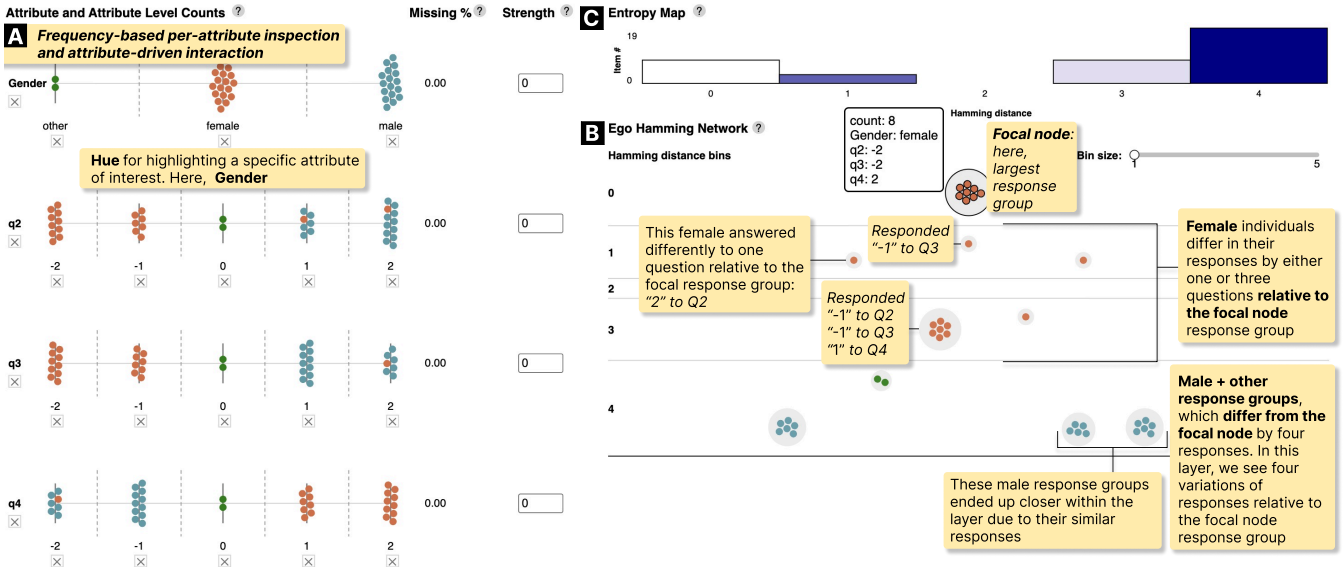

Figure 1: HamCat supports simultaneous analysis of attributes and items in multidimensional categorical survey data. (A) The attribute representation provides an overview of the attributes, attribute levels (values that an attribute can have) and counts associated with each level. (B) The item representation uses an Ego Hamming network to enable in-depth similarity analysis of the dataset from the perspective of a specific reference point, i.e., a focal node. (C) The entropy map provides an overview of the Ego Hamming network for navigation support.

Abstract

We introduce HamCat, a novel visualization method for exploring and analyzing multidimensional categorical survey data. Typical visualization approaches for multidimensional categorical data do not support simultaneous analysis of attributes and items, nor do they allow for in-depth similarity analysis of an entire dataset from the perspective of a specific reference point. HamCat, in contrast, aims to facilitate detailed analysis of multidimensional categorical data across both attributes and items. Our approach builds on the concept of a Hamming ball combined with a force-directed layout to support ego-centric, user-steered analysis of inter-item and inter-attribute relationships in multidimensional categorical survey data. In addition, our method supports the inclusion and nuanced visualization of missingness. We illustrate the value of HamCat through two case studies. The first case focuses on a survey on wellbeing collected by the European Social Survey, while the second is an expert-driven study for a survey on sense of belonging in computer science higher education. These case studies show how HamCat complements existing analysis workflows to reveal relationships and item groupings across attributes that are not easily discoverable through conventional means. Supplementary materials for our method are available at https://osf.io/uz2jv/.

CCS Concepts

• ***Human-centered computing*** → ***Visualization;***

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

1. Introduction

Categorical survey data are commonly collected in domains such as biology, healthcare and the social sciences. Such data are critical to capturing and understanding of phenomena. For instance, education surveys may seek to understand how new educational programs improve student learning outcomes; healthcare surveys may strive to discover why some patients do not adhere to treatment protocols and follow-up; social sciences surveys can uncover cross-sectional attitudes towards mental health, lifestyle, etc., to eventually shape policies. Exploring and analyzing such information can help to uncover patterns, such as relationships between attributes and between data items, or identify patterns of missingness.

The multidimensional, categorical nature of survey data presents several exploratory and analytical challenges. **Inter-attribute relations** in survey data can involve interactions between several attributes, complicating an analyst's ability to identify relevant relationships. For instance, an interaction of age group, health literacy, social factors, and education level usually influences a patient's adherence to a treatment, rather than age group alone. Multiple attributes can form individual, i.e., item, groupings in a respondent pool. This requires the successful assessment of **inter-item relations**, including the structure of the respondent pool, clustering patterns and outliers. For example, patients can form multiple groups with different disease risks based on their age, physical activity, and diet. Analyzing multidimensional categorical survey data from both attribute and item perspectives can provide not only a comprehensive view of the data, but also assists some of the challenges inherent in the analysis of such information. Categorical data do not naturally map to numeric values. This restricts, or even disqualifies, the use of otherwise common analytical methods, including many dimensionality reduction approaches [MR93, vdMH08, MHM18]. Targeted approaches to understanding (dis)similarities in multidimensional categorical data are required to enable an effective analysis of inter-item and inter-attribute relations.

Figure 1 provides an illustrative example of a four-attribute dataset with a gender question and three 5-pt Likert-type questions. Here, the analyst is interested in identifying patterns of responses between and within self-identified gender types. While a standard analysis may employ frequency- [HK81, KBH06] or set-based strategies [AMA*16], these techniques fail to reveal potentially interesting subgroups and their relations, such as subgroups of females with small variations in their Likert responses against their larger difference from male respondents, who themselves may also form subgroups of response variation. This kind of analysis is difficult, if not impossible, to perform with existing approaches, while our method quickly affords this assessment. Moreover, to our knowledge, there are no multidimensional analysis techniques that flexibly support the identification of inconsistent answers, nor similarity analysis of attitudinal directions for survey data that include nominal questions alongside Likert-type questions (negative, neutral and positive). These challenges inspired our research questions:

RQ1: How, and to what extent, can we embed multidimensional categorical survey data into 2D space, while respecting their categorical nature and enabling simultaneous exploration and analysis from both attribute and item perspectives?

RQ2: How can we incorporate user knowledge and perspective on the data in the exploration and analysis process?

RQ3: How can a 2D embedding reveal relevant inter-attribute information in the case of missing data?

To answer these questions, we contribute **HamCat**, a novel visual method for multidimensional categorical survey data. Our approach builds on a Hamming ball [AJS24] and a force-directed layout to support analysis of inter-attribute and -item relations, even in cases of incomplete data. With our method, the user can steer the analysis process and incorporate their particular knowledge to drive a meaningful analysis that is complementary to standard practices.

2. Related Work

Below, we review related work on categorical data visualization, survey data analysis, and visualization strategies for multidimensional categorical data.

Categorical data visualization includes two widespread approaches: frequency-based visualization [HK81, KBH06] and set visualization [AMA*16, FMH08]. Bar charts and pie charts are fundamental examples of frequency-based visualizations, although they do not easily capture relations between attributes or data items. Other examples are Parallel Set Plots [KBH06] and Product Plots [WH11] that follow the frequency-based paradigm but may also surface relations between attributes. However, these methods do not support exploration and analysis of inter-item relations in categorical data. In set visualization [AMA*16, FMH08], we find effective representations for certain datasets and tasks, usually limited to the analysis of the set nature of categorical data. These approaches also do not allow for item-based similarity analysis, whereas our method does. There are also methods that rely on mapping categorical data to a numerical space [Joh09, CBK09], enabling approaches designed for numerical data but violating the inherent nature of categorical data.

Survey data analysis may involve visualization, most commonly bar charts, box plots, heatmaps, and diverging stacked bar charts [HR14]. These are frequency-based techniques with corresponding limitations in detail. Other methods of survey data analysis generally entail summarizing data through descriptive statistics and applying statistical methods, such as the Mann-Whitney U test or the Kruskal–Wallis test [SAJ13, Har15]. Descriptive statistics are limited to summarizing trends, while comparative statistics, e.g., t- or chi-squared tests support group comparison [OL10]. These approaches provide only a summarized representation of survey data and do not easily capture detailed relations between multiple questions (attributes) and respondents (items).

High-dimensional data visualization is an active research area with several surveys encapsulating the state of the field [LMW*16, The08, GTC01]. Dimensionality reduction (DR), which projects high-dimensional data into lower dimensions, is a popular suite of methods [EMK*19, CG15, WVAL10] used to facilitate visualization and analysis of high-dimensional data. DR techniques for categorical data include Multiple Correspondence Analysis (MCA) [AV07], the counterpart of Principal Component Analysis (PCA) [MR93] for nominal variables. For example, Broeksema et al. [BTB13] propose an approach for visual analysis of multidimensional categorical data based on MCA. However, this method, like PCA, cannot detect

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

non-linear relationships. Multidimensional Scaling (MDS) uses dissimilarity measures appropriate for categorical variables, such as the Hamming distance [WW95]. While based on distance metrics suitable for categorical data, these respective techniques produce only a "big picture" of relationships and do not allow for in-depth similarity analysis from the perspective of a specific reference point. They furthermore do not support simultaneous analysis of both attribute and item spaces, which our method does. These techniques also require handling of missingness through imputation, exclusion, or added categories, any of which can distort the data representation.

Challenges for embedding multidimensional categorical data have encouraged the development of new visualization methods to expand analytical capabilities to such data. CatNetVis [TBL23], for example, supports the visual exploration of high-dimensional categorical data with force-directed network layouts. However, this approach focuses on semantic relations between categories and their frequencies, and does not support the exploration and analysis of similarities between items nor missingness. The Categorical Data Map [DJP*24], on the other hand, addresses the need for similarity-based analysis of data items by using a DR-based visualization for categorical data and defining the distance of two data items as the number of varying attributes. However, the number of attributes this method is able to support is limited by the number of colors, which pragmatically limits this method's scalability to a handful of dimensions [Mun14]. Neither technique supports exploration of missingness, the identification of inconsistent answers, nor the analysis of attitudinal directions, i.e., negative, neutral and positive, for Likert-type questions, each of which are central to our approach.

3. HamCat Approach

With HamCat we aim to enable analysis of multidimensional categorical survey data from both item and attribute perspectives, even in scenarios of incomplete data.

3.1. HamCat Requirements & Overview

Our approach is based on the following design requirements, which we derived from our research questions outlined in Sec. 1:

R1: Enable simultaneous exploration and analysis of all responses across different levels of granularity, both on attribute and item levels (**RQ1**). For example, an analyst identifies question-level trends to determine a product's strengths and weaknesses, while at the item level, an analyst captures and relates customers' detailed response patterns to those trends.

R2: Provide interactive means for a user to integrate their knowledge and questions in analysis (**RQ2**). For instance, an analyst may be interested in focusing on specific questions, e.g., health-related.

R3: Enable exploration and analysis of structure among the items, such as how and why items group in a respondent pool, relations between and within such groups, and the diversity of responses (**RQ1**). For example, an analyst identifies groups of patients with high disease risks formed by age, physical activity, and diet.

R4: Support response characterization at the individual item level (**RQ1**). For instance, an analyst may be interested in examining individual responses forming a certain group of participants, e.g., a high-risk disease group.

R5: Enable adjustment of attribute granularity to support analysis of both raw responses and attitudinal directions (**RQ1, RQ2**). For example, an analyst may be interested in focusing on attitudes for Likert-type questions instead of finer-granularity categories.

R6: Enable identification of a key attribute to support analysis of all responses (**RQ1, RQ2**). For example, an analyst may be interested in their survey data in relation to a specific attribute, such as gender, country, or another demographic variable.

R7: Support detection of items with inconsistent or unexpected responses (**RQ1, RQ2**). For instance, an analyst detects respondents who fail attention check questions and removes them from analysis.

R8: Support exploration of the extent of missingness at the attribute and item levels, as well as patterns of co-missing attributes (**RQ3**). For example, responses to the last block of survey questions that are often missing together can indicate survey fatigue.

These design requirements shaped our development of the HamCat approach, which we briefly summarize before discussing its key aspects in more detail.

HamCat supports a joint attribute and item analysis through two simultaneous perspectives, inspired by the Dual Analysis approach by Turkay et al. [TFH11]. The first perspective, shown in Fig. 1A, supports per-attribute frequency-based analysis, such as the prevalence of responses with neutral-valenced responses to Q2-Q4 (**R1**). The second perspective, shown in Fig. 1B, supports item and inter-attribute analysis across all or a selected subset of attributes, for instance, the degree of similarity in female responses to Q2-Q4 (**R1, R3**). In the following, we demonstrate our method using a synthetic dataset inspired by a survey on sense of belonging in computer science conducted by colleagues in our institute. The dataset contains a gender attribute with three levels (female, male, other), two self-assessment questions on one's abilities in informatics, and one question about the perception of the effect of gender on one's informatics abilities. For a demonstration of our method's interactions we refer to the video included in the supplementary materials: `https://osf.io/uz2jv/`.

3.2. Explore and Analyze Attribute Levels

Understanding individual attributes, the value range, and the value distribution of a multidimensional categorical dataset is a useful starting point for an overview and hypothesis formation on possible attribute relations before in-depth analysis. For instance, in our dataset, female aligns with a negative sense of one's abilities in informatics (Q2, Q3) and a positive, i.e., agreeing, perception that others think females are less competent in informatics (Q4).

Illustrated in Fig. 1A, we provide an overview of the data **attributes**, their **levels** (values that an attribute can have), and the attribute **levels' counts** (how many times this level appears in the data) (**R1**). While some online surveys may include tens of thousands of responses, our approach is designed to support surveys of up to approximately 1000 individuals, the standard polling size in polling agencies such as Quinnipiac [Qui23]. Inspired by Atom [PDFE17], a grammar for unit visualizations, we display responses through a set of dot plots [Wil99], chosen to signify the responses of individuals rather than the masses (**R4**). We create one dot plot per attribute, and the plots are vertically aligned and labeled with the attribute

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

levels. For each plot, one dot corresponds to one item, i.e., an individual survey respondent, and we employ a force-directed layout to center items at their respective level selections. This design allows for a facile confirmation in Fig. 1A that female respondents answer negatively to Q2 and Q3. The attribute view supports interaction techniques (**R2**), discussed in detail in Sec. 3.4, that include attribute subsetting, attribute level merging, and attribute weighting.

3.3. Explore and Analyze Items and Inter-Attribute Relations

Sparsity and redundancy are key characteristics of multidimensional categorical data [DJP*24]. While such datasets have a large possible combinatorial space, comparatively few combinations usually materialize, and even fewer are unique combinations. Our approach supports rapid identification and similarity assessment of items with unique combinations of attribute levels using grouping and positioning strategies. Inspired by the framework by Paulovich et al. [PAvdE25] that unifies dimensionality reduction and graph drawing, we display unique groupings and their relations (**R1, R3**) using a Hamming ball [AJS24]—an ego network [EPF*24] with links defined by Hamming distances [BKR02]. We use such an ego-centric approach for its ability to provide nuanced and reliable similarity exploration and analysis of all items from different perspectives (**R3**). While the traditional Hamming ball is static and limited to a specific fixed focal point, our method offers interactivity for exploration and analysis of survey data from the perspectives of different focal points (**R2**).

Ego Hamming network construction. We consider all unique level combinations in data as network **nodes**. The links between nodes, i.e., the network edges, encode their mutual Hamming distances [BKR02]. The Hamming distance between a pair of nodes is the number of attributes where these nodes have differing levels. Identical nodes, i.e., nodes with the same attribute levels, have pairwise Hamming distances of 0. Survey data often consist of both ordinal and nominal attributes. To enable similarity analysis of items represented as combinations of ordinal and nominal values (**R3**), we distill such multidimensional categorical data to their nominal properties and focus on whether responses differ, rather than on the degree of their differences (in the case where such a degree can be considered at all). Reducing the complexity of these data is a necessary compromise to support their item-similarity analysis. As in the attribute view, we encode each item as a **dot** to signify individuals' responses and to support response characterization at the individual item level (**R4**). Items with identical attribute level combinations are visually grouped together into a larger disk element, i.e., a **node**, which represents a unique set of attribute levels (**R3**). In our demonstrative dataset, each dot represents a survey participant, and each node represents a group of participants who answered the survey identically. To allow rapid identification of items in relation to the levels of a specific attribute (**R6**), the user may set a key attribute. This attribute is then double-encoded with hue, as shown in Fig. 1B where gender levels are colored. By default, this key attribute is included in the Hamming distance calculations, meaning that differences in this attribute display in the network. The user may exclude this attribute from the network calculations if desired (**R2**).

Ego Hamming network layout. To manage visual clutter and ease exploration, we do not display the network's edges explicitly and instead encode relations between nodes through a spatial layout, specifically a layered node-link representation [EPF*24] (**R3**). At the top of the network, i.e., layer 0 with a Hamming distance of 0, we place a focal node: a specific reference node selected among the network's nodes. By default, we select the group with the highest number of items as our focal node, i.e., the ego node. This immediately surfaces the most common response set from survey participants. Each subsequent layer contains nodes that increment by a Hamming distance of 1 from the focal node. As we traverse down the network, the nodes contained in the layers become increasingly dissimilar in relation to the focus node.

We support interactive user selection of a new focal node for a new similarity assessment (**R2, R3**), which subsequently rebuilds the network as shown in Fig. 2. Here, the user first selects a single female node in level 1 for similarity assessment, then pivots the space around a set of male respondents initially placed in level 4.

All Hamming layers have the same width but differ in their heights, and are separated with horizontal lines to clearly distinguish between different degrees of dissimilarity (**R3**). To optimize space, we adjust the layers' heights based on how many nodes they need to accommodate. Layers may be empty, in which case they are drawn as thin strips, as shown in layer 1 of Fig. 4. We draw the number of layers equal to the maximum Hamming distance between a focal node and the remaining nodes. As the network grows downwards and (potentially) offscreen, we include scrolling functionality to support navigation across all layers. In each layer, we use a spring embedder system and the Fruchterman-Reingold algorithm [FR91], chosen to provide parameters for the user to manipulate and steer the embedding process based on their knowledge (**R2**). Our spring embedder algorithm uses the Hamming distance to position more similar nodes, i.e., nodes with lower Hamming distances, closer together and more dissimilar nodes, i.e., nodes with higher Hamming distances, further apart (**R3**), as shown in Fig. 1B.

To show node and item metadata (**R4**), we follow the details-on-demand paradigm [Shn03] and supplement the network visualization with interactive tooltips that reveal, on hover, the number of items in the node, item ids, and attribute levels.

Handle sparsity. In the initial network, the Hamming distance from the focal node increases by one with each next layer. However, multidimensional categorical data are sparse – often very sparse. This sparsity can lead to many empty layers in a network as categorical combinations of certain Hamming distances (in relation to the focal node) are in fact absent in the data. To tackle this issue, we supplement the item representation with a binning slider shown in Fig. 3. By adjusting this slider, the user can choose the degree of similarity they want to consider (**R2**). Fig. 3 shows how by setting the slider value to two, we now consider nodes with Hamming distances of zero and one (relative to the focal node) as similar and group them together in the updated layer 0. Similarly, nodes previously placed in layers 2 and 3 are now in the updated layer 1.

Entropy map. We also supplement the items representation with an entropy map, as shown in Fig. 1C. This way, we provide an overview of the Ego Hamming network and support its navigation (**R3**). The entropy map is a horizontal bar chart with each bar representing a layer of the Ego Hamming network. The height of a bar encodes the quantity of the items belonging to the corresponding

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

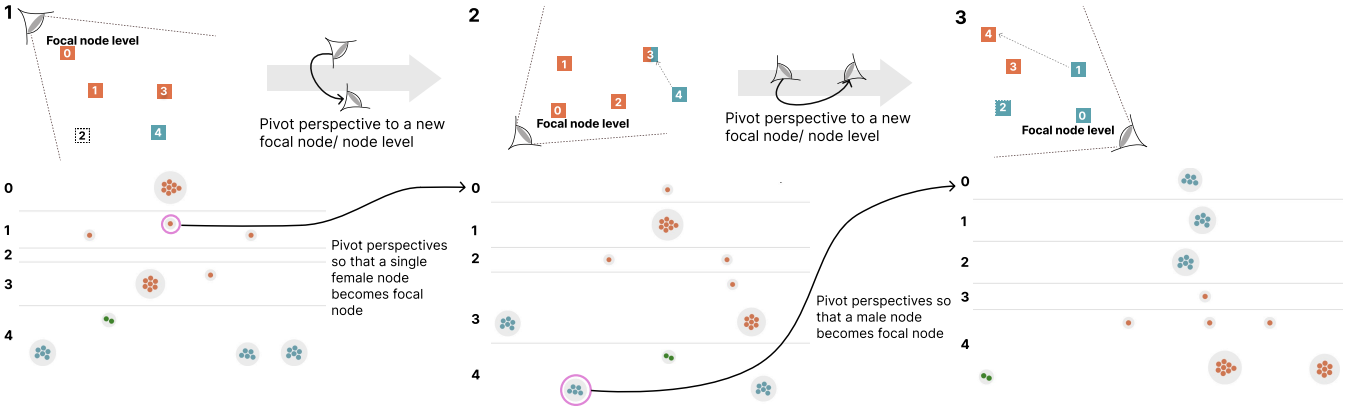

Figure 2: HamCat enables the user to select a node of interest as a focal node. Upon selection of a new node, the Ego Hamming network rebuilds revealing inter-group relationships from the perspective of the new focal node.

layer. The lightness of the bar chart conveys the entropy level of the matching layer with darker hues visualizing higher entropy. For each layer, we calculate entropy as

$$H(\text{layer}) = \frac{-\sum_i^A \sum_j^{L_i} p(l_{ij}) \log_2 p(l_{ij})}{N_{\text{items}}},$$

where A is the number of attributes, L_i is the number of levels for the i-th attribute, l_{ij} is the j-th level of the i-th attribute, and $p(l_{ij})$ is the probability of l_{ij} appearing in an item. This map allows for the detection of empty layers as well identification of layers with homogenous or heterogenous group nodes (**R3**). Additionally, it immediately shows layers with only one node, as such layers include only identical items. Furthermore, we can immediately infer the size of such a single node from the height of the matching bar.

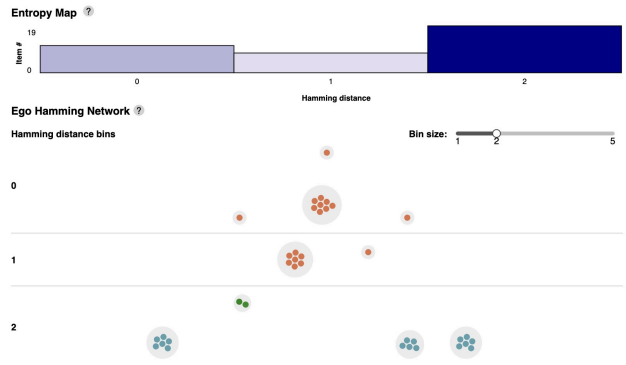

Figure 3: Binning of items results in an adjusted entropy map and Ego Hamming network. This allows the user to handle the sparsity of data and adjust the degree of similarity they want to consider.

3.4. Dynamically Adjust Attribute Levels and Strengths

The importance of integrating user knowledge in different visual tools has been emphasized in numerous works [EHR*14, SZS*17, AAA*24, etc.]. Our approach follows this idea and enables the user to bring in their expertise when studying the embedding. In the following, we describe interactions offered by our approach that allow the user to steer the embedding process (**R2**). Additionally, we demonstrate the method's interactions in the video included in the supplementary materials: https://osf.io/uz2jv/.

(De)select attributes and levels. The attribute representation displays a dataset's attributes and their levels along with each level's frequency. This view allows for rapid identification of attributes with no variance across their levels and their further omission as non-informative for an item-based similarity analysis. HamCat includes checkboxes, shown in Fig. 1A, that allow the user to select attributes to include or exclude (**R2**). The method also includes checkboxes for attribute level selection, shown in Fig. 1A (**R2**). By selecting specific levels, the user may examine particular strata of respondents in the survey. Moreover, the user can specify responses to attention-check questions that they consider inconsistent and suspect in the data to surface respondents who fail these questions (**R2, R7**). Initially, our method displays all attributes in the data. When the user deselects or selects attributes or attribute levels, the Ego Hamming network rebuilds according to this new selection. When the user deselects an attribute, the dots of the matching dot plot fade out, as shown in Fig. 4. The dots with reduced opacity serve as context and allow the user to focus on the selected attributes [Hau06]. Deselection of a key attribute does not affect the nodes' composition in the network and reflects only in the Hamming distances between them. When the user deselects an attribute level, dots encoding items with this level decrease in opacity across all dot plots, as shown in Fig. 4. This visual response reveals how items with the deselected level are distributed across all attributes' levels.

Merge levels. The user may want to merge similar attribute levels to simplify their analysis (**R5**), e.g., Likert scale responses of the same attitudinal directions ("satisfied" and "very satisfied", or "dissatisfied" and "very dissatisfied"), or adjacent age groups. HamCat supports such aggregation. For all dot plots, we position each attribute level tick within a specific region outlined by dashed lines, shown in Fig. 4. By dragging and dropping attribute level

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

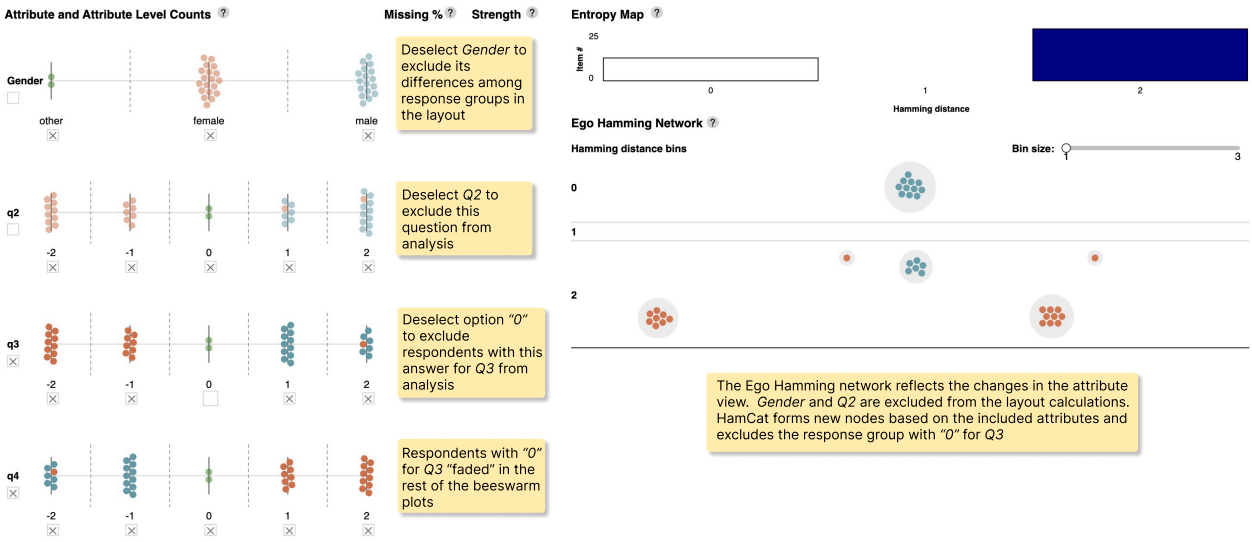

Figure 4: Gender and Q2 questions are deselected as well as "0" for Q3. The Ego Hamming network reflects this selection.

ticks, the user can place them within the same region, shown in Fig. 5 to treat the corresponding levels as one attribute level. The Ego Hamming network rebuilds accordingly. For instance, consider two nodes in the network with two distinct attribute levels for a certain attribute. Merging these attribute levels decreases the Hamming distance between these nodes by one, and the network updates accordingly. Merging can result in multiple focal nodes as in Fig. 5. Here, the Hamming distance to deeper network layers is shown relative to either focal node. This results in a network displaying more coarsely-defined relations.

Adjust attribute strength. Among all attributes that survey data offer, some may be of special interest, e.g., which define specific groups of items, such as different age groups, countries or genders, and may be main drivers of patterns in data. Our method supports emphasis of specific attributes that can be used to form and compare clusters of items (**R2, R3**). For each attribute in analysis, HamCat includes a numeric input box, as shown in Fig. 1A, to specify the attributes' strengths. These translate into forces for the Ego Hamming network layout computation. When the user adapts an attribute's strength, HamCat uses this strength to group the network's nodes based on this attribute. We base our clustering on attractive and repulsive forces. Nodes with the same level for the emphasized attribute attract while nodes with differing levels repel from each other. The definition of forces is described in more depth in Algorithm 1 in the supplementary materials.

3.5. Reveal and Manage Missingness

Our approach is flexibly designed to support missing data (**R8**), as the absence of data may provide valuable insights [BH84], such as systematic patterns of non-response or co-missing attributes. We allow the user to globally select the upper threshold of data missingness they would like to include in their analysis, ranging from 0% (no missing responses) to a theoretical 100% (all missing responses). This selection is afforded through a slider shown in Fig. 7A. By default we consider only complete items, i.e., 0% missing. We also display missingness per attribute, shown as the rightmost values in the dot plots in Fig. 7B. For example, in our dataset with 40 items (participants), questions Q3 and Q4 are missing one response each from a male participant, yielding a missingness of 2.5% for each of these attributes, as shown in Fig. 7B, lower part.

HamCat also supports discovery of missing data in items across both views. If not all data values are present, instead of a dot we use a circular sector metaphor to indicate the fraction of values present (full dot = 100%, half dot = 50%, quarter dot = 25%). For each incomplete item, we consider all possible value combinations its missing attributes could have, as this uncertainty can be important for understanding potential limitations in the conclusions drawn from attribute relations. Each combination appears with the probability $\frac{1}{|C|}$, where C denotes all possible combinations. Figure 7C shows a male participant who skipped Q3, represented in a possibility space of five nodes with circular sectors of lower opacity. Sector opacity encodes its probability, in this case 20% since Q3 is a 5-pt Likert question. Hover interactions support inspection of these metadata, as shown in Fig. 7C, and provide a highlight effect by outlining all sectors belonging to the same item, in this case the male participant with one skipped response.

4. Case Studies

We implemented our method in a web application using D3.js [BOH11] and Flask [Pal25]. The full source code is available at https://github.com/Annchovy/HamCat.git. The general workflow across our case studies begins with an inspection of attribute response distributions and an attribute–item similarity comparison in the Ego Hamming network, guided by the entropy map to identify layers of diversity and high respondent numbers relative to the focal node. The network is then adapted by merging and selecting attributes / attribute levels, or by pivoting to a new focal node for a new similarity comparison. Data missingness can be assessed to

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

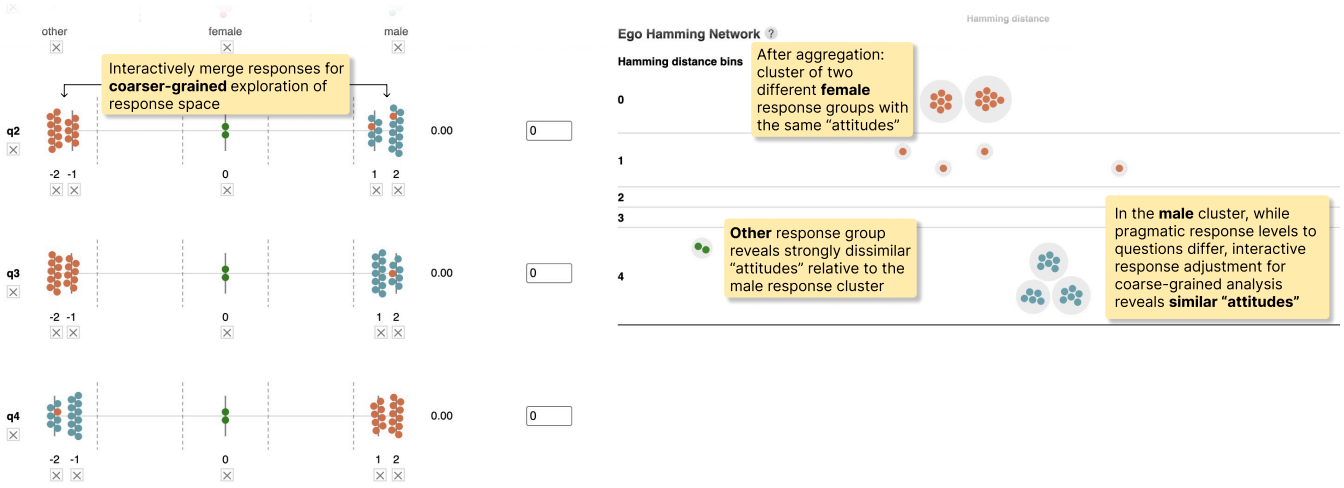

Figure 5: Illustration of merging attribute levels into coarse-grained "negative", "neutral" and "positive" attitudes (left). HamCat automatically recalculates Hamming distances between nodes and updates the network view accordingly (right) to show gender node clusters with similar attitudes. Here, the gender attribute has been excluded from network calculations and only encoded by hue.

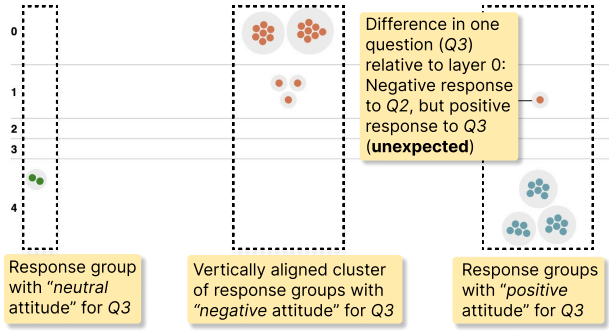

Figure 6: Adjusting the strength of Q3 to 4.0 emphasizes differences in responses to this question among response groups.

capture a more nuanced picture of the responses. We demonstrate the use and value of HamCat through two case studies. The first illustrates our method's applicability by reproducing insights from a previous study, while the second showcases HamCat's potential through an expert case study with colleagues in the UiB computer science education research group.

4.1. European Social Survey on Wellbeing

Since 2001, the European Social Survey (ESS) [ESS25] is a biennial cross-national survey on the attitudes, beliefs, and behavior patterns of populations across Europe. A recent analysis focused on dimensions of wellbeing of the survey participants [Eur23]. The analysis report [Eur23] identifies broad dimensions of wellbeing across countries, and highlights the similarity of scores for the broad community wellbeing dimension between Switzerland and Hungary. This dimension consists of five attributes with 3507 responses, i.e., items. The attribute types include four unipolar Likert questions (ascending 7- and 11-pt, where a high response is positively-valenced) and one bipolar Likert question (descending 5-pt, where a low response is positive-valenced). The report defines the scores for the community wellbeing dimension as the averaged z-scores of all five attributes, and does not discuss missing data. Our goal was to replicate the report's findings that **(1)** all five attributes yield responses similar enough to be considered a coherent, broader dimension, **(2)** and to confirm the findings that Hungary and Switzerland have mainly positive responses. Figure 8 shows key steps of our analysis.

Using our approach, with country as our (colored) key attribute as in Fig. 8A, we achieved a clear visual distinction of the respondents by country across individual attributes and the network view. We quickly confirmed a high prevalence of positive responses across all community wellbeing attributes through exploration of the attribute representation and in the Ego Hamming network where the focal node (the most abundant response set) has consistently positive answers, as in Fig. 8B (**R1**). In Fig. 8B, layer 1 contains a Hungarian response group that differs from the focal node group (Swiss) by the country attribute. Deselecting this attribute in Fig. 8C allows for a focused inspection of differences between the community wellbeing questions. Fig. 8D illustrates the result of such deselection: the Hungarian response group moved to layer 0, meaning its responses across the community wellbeing questions are identical to those of the focal node group. Further inspection of the network revealed that the subsequent layers contain different intensities of primarily positive responses for both countries (**R3**). In Fig. 8D, the entropy map shows that the positive-valenced responses in layer 0 constitute only a small part of the whole response space. Since multidimensional categorical survey data are sparse, layers with higher Hamming distances do not necessarily have a larger number of response patterns and higher entropy. Layer 4, on the other hand, contains a large and diverse group of responses (**R3**). We examined this group to understand the source of diversity. Fig. 8E shows layer 4, where we identified several responses with positive answers of varying

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

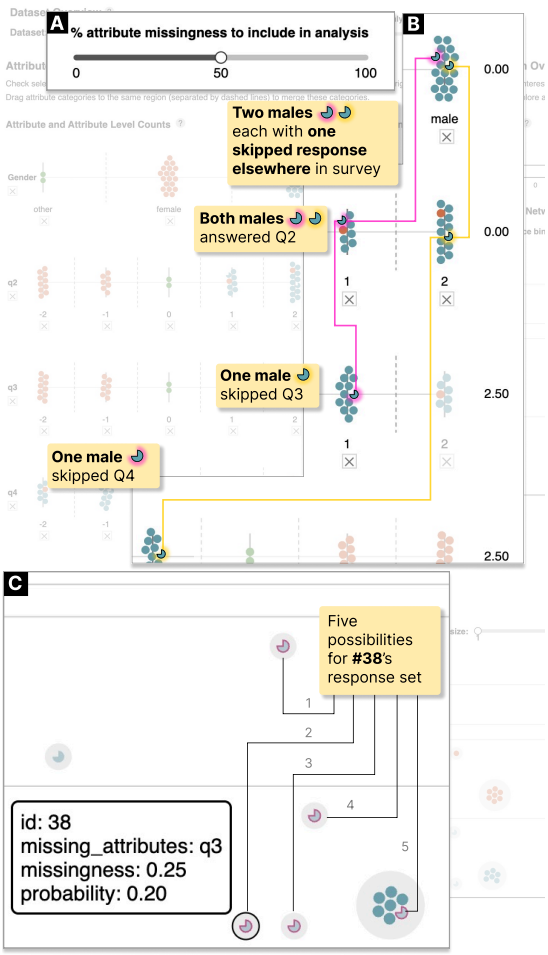

Figure 7: Missingness handling in HamCat: (A) Global selection of attribute missingness to include in analysis session, (B) trace item missingess across attributes with per-attribute missingess quantification, and (C) explore possibilities for missing items over entire attribute space in Hammingview.

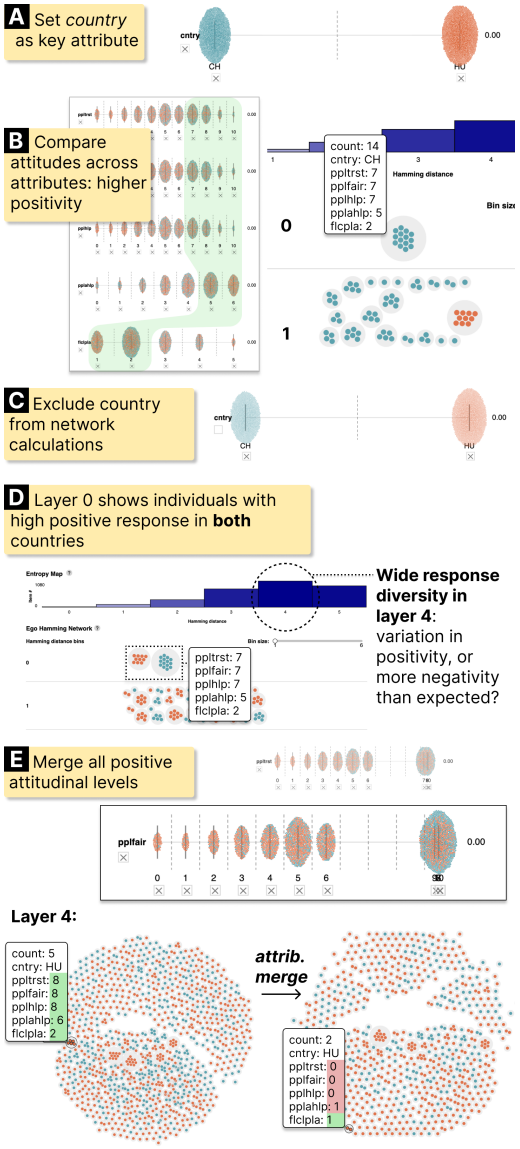

Figure 8: ESS case study comparing community wellbeing responses among Hungarian and Swiss participants [Eur23]. In (B) and (E), green highlights indicate positive responses; red negative.

strength (relative to the focal node group). After merging options into broader attitude categories in Fig. 8E, layer 0 demonstrates positive answers across all five questions, while layer 4 contains response groups with negative and neutral answers to any four of the questions. Fig. 8E shows layer 4 after the merging, where we detected inconsistent response patterns (**R7**). The ESS report aggregated the community wellbeing questions into a single score per country, obscuring response group variation. HamCat, in contrast, allows for a detailed investigation of response group heterogeneity, size, and answer consistency. With our method, we show that the responses from both countries are mainly positive. Inspecting the Ego Hamming network reveals that many respondents have consistent answers across the questions, which aligns with the treatment of these questions as one dimension. However, we also identified inconsistent response patterns, which may arise from inattention or carelessness and introduce noise into the data.

Focusing on responses from Switzerland (same five attributes, 1820 items, increased from 1493 in the complete response set) to manage computational capacity, we included all incomplete responses to gain a more nuanced understanding of the conclusion that Swiss respondents are overwhelmingly positive across questions of community wellbeing (**R8**). We explored incomplete responses in the resulting Ego Hamming network across all layers, choosing layers with the emptiest nodes, for example, layer 4 shown in Fig. 9, and inspecting their metadata. Through this inspection, we found one combination occurred more frequently than others: pplahlp ("Feel people in local area help one another") and flclpla ("Feel close to the people in local area"). Although the extent of this

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

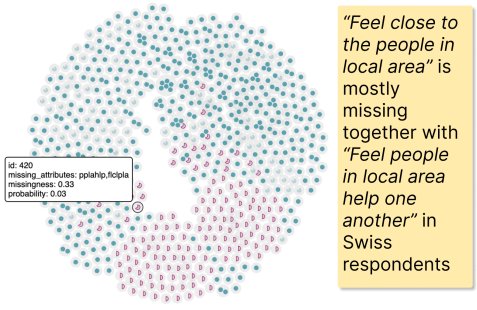

Figure 9: Comparing missing Swiss participant responses in ESS study network layer 4 [Eur23].

missingness is low, it can reveal subtle relationships between the co-missing attributes. While the overall missingness in the data is minor and does not substantially affect the analysis, its examination can contribute to greater certainty in the conclusion by presenting a more complete picture of responses.

4.2. *Sense of Belonging* in CS Higher Education

Our expert case study focuses on survey data on sense of belonging collected by the UiB computer science education research group. The survey dataset includes 22 attributes: a gender response and 21 Likert questions, and it consists of 130 items, i.e., responses from students in computer science at our university who participated in the survey. The Likert questions aim to capture the participants' comfort and sense of belonging in computer science higher education across several dimensions. The survey leaders are interested in the following questions:

T1: What is the overall response space? (**R1**)
T2: How do the participants' responses vary across genders? (**R6**)
T3: Are there inconsistent/unexpected answers? (**R7**)
T4: Given responses of individuals to specific survey questions, what are their responses to the remaining questions? (**R2**)
T5: What questions are left blank, what impact does this have? (**R8**)
T6: How do certain questions and respondents relate? (**R3**)

As a standard approach, the experts use grouped bar charts to visualize the responses to each question, with color encoding gender. This communicates trends in the responses across all questions, and how gender relates to these trends. However, this approach is limited to the analysis of two attributes at a time and does not offer insights into inter-item relationships. HamCat complements the standard approach and extends the experts' analytical capabilities.

Our study consisted of two sessions. The first session involved a discussion of data and tasks with two experts, followed by a guided walkthrough of the HamCat interface. The second session involved a one-on-one interview with an expert from the first session and lasted one hour and 30 minutes. It consisted of a paired analysis session combining hypothesis confirmation with open-ended exploratory data analysis, followed by the expert's feedback.

Analysis. For our analysis, we considered gender as the colored key attribute. Initially, the attribute and item representations provided an overview of the response space (**T1**). We first confirmed a pattern identified by the expert with their standard approach: females tend to answer negatively to questions Q5 ("I consider myself good at informatics.") and Q6 ("I am capable of excelling at informatics."). Selection of respondents with negative answers to these questions in the attribute view revealed the Ego Hamming network composed predominantly of female respondents, confirming the pattern (**T2**). Following the open-exploration process, the expert chose to explore the response space for the selected sub-group of respondents further (**T4**), focusing on self-assessment questions on one's abilities in informatics and the importance of informatics for one's future. After inspecting the resulting Ego Hamming network, the expert merged options for several questions to focus on attitudinal directions and adjusted the strength for one of those questions to explore how the respondents were distributed across the attitudes. The expert noted that respondents with negative answers to the question were more dissimilar from the focal node group—which answered positively—compared to the rest of respondents (**T6**), suggesting distinct groups worth further investigation. Additionally, the inspection of the respondent group with positive answers revealed a node with a contradictory combination of attitudes for questions Q3 ("Being good at informatics will be useful for me in my future.") and Q2 ("Doing well in informatics is important to my future success."), which the expert marked as an unexpected answer (**T3**). This prompted a focused examination of the discovered deviation in the response space, and through further interactions with HamCat, we identified more unexpected responses, as illustrated in Fig. 10.

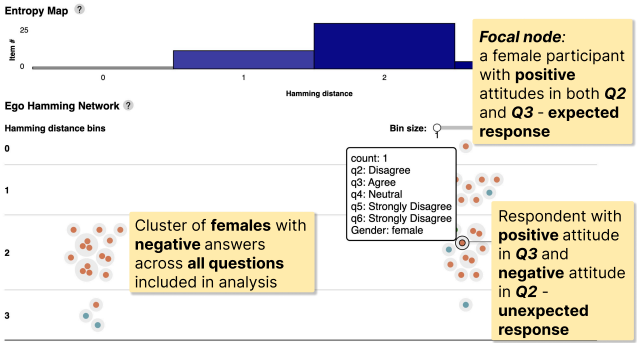

Figure 10: The Ego Hamming network resulting from the analysis workflow in the expert-driven case study. The response space contains a cluster of female respondents with negative responses to all selected questions and unexpected response patterns.

The network surfaced a cluster of female respondents with negative answers to all selected questions, which the expert marked as interesting. Moreover, the expert indicated unexpected response patterns. The expert noted that they expected to see the same attitudinal directions for both Q2 and Q3. However, we identified response groups having positive attitudes for Q3 and negative for Q2.

Expert Feedback. Although the expert noted that HamCat appears complex in the beginning, they successfully interpreted the representations, such as a focal response group and its relations to the other response groups, and used the interactive components, such as attribute strength and focal node selection, further in the

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

session. For example, for a focal node with a positive response to Q3, after adjusting the level granularity and the question's strength, the expert noted that "what' s interesting is that those who disagree [to Q3], they don't have any observations [responses] on level [layer] 1, 2." They marked this dissimilarity with regards to the focal node as more significant compared to the other response groups. In the beginning, the expert expressed their confusion about the initial, arbitrarily selected, focal node. However, later they understood its function and were "basing the selection of the focal node on a hypothesis," such as selecting a node with positive attitudes for both Q2 and Q3 (as an expected response pattern) to explore deviations from this pattern. The expert merged response options and adjusted question strengths to explore and analyze the response groups across the merged attitudes, and commented on these interactions as "very useful." Since we were working with a complete dataset, we did not have missing data to explore. We focused our analysis on a specific block of questions and were able to observe the entire Ego Hamming network at once, not requiring the entropy map for the network's overview. During the analysis session, HamCat not only confirmed the pattern discovered by the expert with their standard approach, i.e., grouped bar charts, but allowed for a detailed exploration and analysis of the response groups. With our method we were able to extend the analysis and enable inspection and identification of interesting groups of respondents, their structures and relations, e.g., participants with the unexpected response pattern. This highlights the potential of our method to generate new insights about response patterns. The expert stated, "I would find it useful for my research," and expressed interest in continuing to work with HamCat.

5. Discussion

HamCat offers exploration and analysis of inter-attribute and -item relations in multidimensional categorical survey data from the perspective of a specific reference point. With an ego-centric approach we lose the general "global picture" of relations between responses to a certain degree, but gain their detailed and reliable similarity analysis in relation to the focal node group: embedding multidimensional data into 2D inevitably results in information loss. The Ego Hamming network may lead to confusion about the dissimilarities between nodes located in differing Hamming layers 1 – N (number of attributes). However, HamCat is intentionally designed to allow the user to view response space from the perspectives of different response groups, and not for between-layer inspection of similarities. To enable item analysis of multidimensional categorical survey data, our approach uses Hamming distances and focuses on capturing response differences. However, future work can explore ways of incorporating the order of ordinal variables into the analysis as befitting the intended tasks.

Scalability poses a challenge. HamCat is designed to support up to approximately 1000 respondents, the standard polling size in polling agencies such as Quinnipiac [Qui23]. For survey data with more respondents, our method faces screen space constraints. For example, accommodating a larger number of participants in the dot plots requires reducing the size of the dots, losing readability below a certain threshold. However, our method has potential to scale with an increasing number of items, e.g., a dot in a dot plot can encode multiple items. In the Ego Hamming network, we can include zooming behavior similar to GenomeSpy [LOL*24], enabling a gradual exploration of all nodes in the network by focusing on the most important nodes at each zoom level. Additionally, we can offer interactive tools that allow the user to filter out nodes exceeding a specified Hamming distance, to ease exploration and analysis. The force-directed layout limits the number of attributes and items in the analysis. As the number of items and attributes increases, the force-directed algorithm requires more time to converge and can fail to produce intended clusters of items, demanding more advanced techniques for larger survey data. Our current approach also assumes that all combinations of missing attributes are equally likely. As a result, the number of uncertain items increases combinatorially with every missing attribute, hence, more efficient methods are needed. Future work can explore different probabilistic or constraint-based models for determining possible values for incomplete items. The expert in our sense of belonging case study noted early confusion over the arbitrarily-selected initial focal node. Future work may look into other meaningful criteria for initial focal node selection, such as a clustering coefficient [SKO*07], or interactively allowing the user to specify conditions for selection. Although HamCat does not leverage the fact that survey questions are often semantically grouped, future work can investigate their incorporation in, for instance, similarity measures for items.

6. Conclusions

We presented HamCat, a novel method employing Hamming distance-based similarity analysis for the visual exploration and analysis of multidimensional categorical survey data. Based on a Hamming ball and a force-directed layout, our method offers an ego-centric, user-driven simultaneous analysis of inter-item and inter-attribute relations. HamCat provides a nuanced visualization even when answers are missing. While the standard methods remain valuable tools for survey data analysis, these approaches offer only a summarized representation of responses and do not capture detailed inter-attribute and -item relations. The Hamming distance and the ego-centric approach show promise in extending analytical capabilities for survey data exploration and analysis. We demonstrate the potential of HamCat with two case studies: a survey on wellbeing collected by the European Social Survey and a survey on sense of belonging in CS higher education. In both cases, we could confirm earlier findings and show new information, demonstrating how HamCat complements existing methods as a new way of conducting survey data analysis, where the analysis of attributes and items is conducted simultaneously between and within views.

Acknowledgments

We are grateful for the support of Aleksandr Popov, Jan Byška, Eduard Gröller, and Nikolay Kaleyski for their invaluable insights and discussions. We thank our colleagues in the computer science education group for their time and participation in our case study. Parts of this work have been done in the context of the UiB Center for Data Science (CEDAS).

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

References

[AAA*24] Andrienko N., Andrienko G., Artikis A., Mantenoglou P., Rinzivillo S.: Human-in-the-loop: Visual analytics for building models recognizing behavioral patterns in time series. IEEE Comput Graph Appl 44, 3 (2024), 14–29. doi:10.1109/MCG.2024.3379851. 5

[AJS24] Alon N., Jin Z., Sudakov B.: The Helly number of Hamming balls and related problems. arXiv (2024). doi:10.48550/arXiv.2405.10275. 2, 4

[AMA*16] Alsallakh B., Micallef L., Aigner W., Hauser H., Miksch S., Rodgers P.: The state-of-the-art of set visualization. Comput Graph Forum 35, 1 (2016), 234–260. doi:10.1111/cgf.12722. 2

[AV07] Abdi H., Valentin D.: Multiple correspondence analysis. In Encyclopedia of Measurement and Statistics, Salkind N. J., (Ed.). Sage, 2007, pp. 651–657. 2

[BH84] Babad Y. M., Hoffer J. A.: Even no data has a value. Commun ACM 27, 8 (1984), 748–756. doi:10.1145/358198.358204. 6

[BKR02] Bookstein A., Kulyukin V. A., Raita T.: Generalized Hamming distance. Information Retrieval 5, 4 (2002), 353–375. doi:10.1023/A:1020499411651. 4

[BOH11] Bostock M., Ogievetsky V., Heer J.: D^3 Data-Driven Documents. IEEE Trans Vis Comput Graph 17, 12 (2011), 2301–2309. doi:10.1109/TVCG.2011.185. 6

[BTB13] Broeksema B., Telea A. C., Baudel T.: Visual analysis of multi-dimensional categorical data sets. Comput Graph Forum 32, 8 (2013), 158–169. doi:10.1111/cgf.12194. 2

[CBK09] Chandola V., Boriah S., Kumar V.: A framework for exploring categorical data. In Proceedings of the 2009 SIAM International Conference on Data Mining (2009), SIAM, pp. 187–198. doi:10.1137/1.9781611972795.17. 2

[CG15] Cunningham J. P., Ghahramani Z.: Linear dimensionality reduction: Survey, insights, and generalizations. J Mach Learn Res 16, 1 (2015), 2859–2900. doi:10.5555/2789272.2912091. 2

[DJP*24] Dennig F. L., Joos L., Paetzold P., Blumberg D., Deussen O., Keim D. A., Fischer M. T.: The categorical data map: A multidimensional scaling-based approach. In IEEE VDS (Los Alamitos, 2024), IEEE Comp Soc, pp. 25–34. doi:10.1109/VDS63897.2024.00008. 3, 4

[EHR*14] Endert A., Hossain M. S., Ramakrishnan N., North C., Fiaux P., Andrews C.: The human is the loop: New directions for visual analytics. J Intell Inf Syst 43, 3 (2014), 411–435. doi:10.1007/s10844-014-0304-9. 5

[EMK*19] Espadoto M., Martins R. M., Kerren A., Hirata N. S., Telea A. C.: Toward a quantitative survey of dimension reduction techniques. IEEE Trans Vis Comput Graph 27, 3 (2019), 2153–2173. doi:10.1109/TVCG.2019.2944182. 2

[EPF*24] Ehlers H., Pahr D., Filipov V., Wu H.-Y., Raidou R. G.: Me! Me! Me! Me! A study and comparison of ego network representations. Comput Graph 125, C (2024), 1–15. doi:10.1016/j.cag.2024.10412. 4

[ESS25] European Social Survey: Home | European Social Survey, 2025. https://www.europeansocialsurvey.org/ Accessed Nov 2025. 7

[Eur23] European Social Survey: Europeans' wellbeing, 2023. https://www.europeansocialsurvey.org/ Accessed Dec 2025. 7, 8, 9

[FMH08] Freiler W., Matković K., Hauser H.: Interactive visual analysis of set-typed data. IEEE Trans Vis Comput Graph 14, 6 (2008), 1340–1347. doi:10.1109/TVCG.2008.144. 2

[FR91] Fruchterman T. M., Reingold E. M.: Graph drawing by force-directed placement. Software: Practice and experience 21, 11 (1991), 1129–1164. doi:10.1002/spe.4380211102. 4

[GTC01] Grinstein G., Trutschl M., Cvek U.: High-dimensional visualizations. In Proc Visual Data Mining Workshop, KDD (2001), vol. 2, p. 120. 2

[Har15] Harpe S. E.: How to analyze Likert and other rating scale data. Curr Pharm Teach Learn 7, 6 (2015), 836–850. doi:10.1016/j.cptl.2015.08.001. 2

[Hau06] Hauser H.: Generalizing focus+context visualization. In Scientific Visualization: The Visual Extraction of Knowledge From Data. Springer, 2006, pp. 305–327. doi:10.1007/3-540-30790-7_18. 5

[HK81] Hartigan J. A., Kleiner B.: Mosaics for contingency tables. In Computer Science and S: Proc Symposium on the Interface (New York, 1981), Springer, pp. 268–273. doi:10.1007/978-1-4613-9464-8_37. 2

[HR14] Heiberger R., Robbins N.: Design of diverging stacked bar charts for Likert scales and other applications. J Stat Software 57 (2014), 1–32. doi:10.18637/jss.v057.i05. 2

[Joh09] Johansson S.: Visual exploration of categorical and mixed data sets. In Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration (2009), pp. 21–29. doi:10.1145/1562849.1562852. 2

[KBH06] Kosara R., Bendix F., Hauser H.: Parallel sets: Interactive exploration and visual analysis of categorical data. IEEE Trans Vis Comput Graph 12, 4 (2006), 558–568. doi:10.1109/TVCG.2006.76. 2

[LMW*16] Liu S., Maljovec D., Wang B., Bremer P.-T., Pascucci V.: Visualizing high-dimensional data: Advances in the past decade. IEEE Trans Vis Comput Graph 23, 3 (2016), 1249–1268. doi:10.1109/TVCG.2016.2640960. 2

[LOL*24] Lavikka K., Oikkonen J., Li Y., Muranen T., Micoli G., Marchi G., Lahtinen A., Huhtinen K., Lehtonen R., Hietanen S., et al.: Deciphering cancer genomes with GenomeSpy: A grammar-based visualization toolkit. GigaScience 13 (2024), 1–15. doi:10.1093/gigascience/giae040. 10

[MHM18] McInnes L., Healy J., Melville J.: UMAP: Uniform manifold approximation and projection for dimension reduction. JOSS 3, 29 (2018), 861. doi:10.21105/joss.00861. 2

[MR93] Maćkiewicz A., Ratajczak W.: Principal components analysis (PCA). Computers & Geosciences 19, 3 (1993), 303–342. doi:10.1016/0098-3004(93)90090-. 2

[Mun14] Munzner T.: Visualization Analysis and Design. CRC press, 2014. doi:10.1201/b17511. 3

[OL10] Ott R., Longnecker M.: An Introduction to Statistical Methods and Data Analysis. Cengage Learning Inc, 2010. 2

[Pal25] Pallets: Flask. https://flask.palletsprojects.com, 2025. 6

[PAvdE25] Paulovich F. V., Arleo A., van den Elzen S.: When dimensionality reduction meets graph (drawing) theory: Introducing a common framework, challenges and opportunities. Comput Graph Forum 44, 3 (2025), 1–12. doi:10.1111/cgf.70105. 4

[PDFE17] Park D., Drucker S. M., Fernandez R., Elmqvist N.: Atom: A grammar for unit visualizations. IEEE Trans Vis Comput Graph 24, 12 (2017), 3032–3043. doi:10.1109/TVCG.2017.2785807. 3

[Qui23] Quinnipiac University: Poll results. https://poll.qu.edu/poll-results, 2023. Accessed Dec 2025. 3, 10

[SAJ13] Sullivan G. M., Artino Jr A. R.: Analyzing and interpreting data from Likert-type scales. J Grad Med Educ 5, 4 (2013), 541–542. doi:10.4300/JGME-5-4-18. 2

[Shn03] Shneiderman B.: The eyes have it: A task by data type taxonomy for information visualizations. In The Craft of Information Visualization. IEEE Comp Soc, Los Alamitos, 2003, pp. 364–371. doi:10.1109/VL.1996.545307. 4

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[SKO*07] SARAMÄKI J., KIVELÄ M., ONNELA J.-P., KASKI K., KERTESZ J.: Generalizations of the clustering coefficient to weighted complex networks. Phys Rev E 75, 2 (2007), 027105. doi:10.1103/PhysRevE.75.027105. 10

[SZS*17] SACHA D., ZHANG L., SEDLMAIR M., LEE J. A., PELTONEN J., WEISKOPF D., NORTH S. C., KEIM D. A.: Visual interaction with dimensionality reduction: A structured literature analysis. IEEE Trans Vis ComputGraph 23, 1 (2017), 241–250. doi:10.1109/TVCG.2016.2598495. 5

[TBL23] THANE M., BLUM K. M., LEHMANN D. J.: CatNetVis: Semantic visual exploration of categorical high-dimensional data with force-directed graph layouts. In Proc EuroVis (Short Papers) (2023), pp. 91–95. doi:10.2312/evs.20231049. 3

[TFH11] TURKAY C., FILZMOSER P., HAUSER H.: Brushing dimensions – a dual visual analysis model for high-dimensional data. IEEE Trans Vis ComputGraph 17, 12 (2011), 2591–2599. doi:10.1109/TVCG.2011.178. 3

[The08] THEUS M.: High-dimensional data visualization. In Handbook of Data Visualization. Springer, Berlin, 2008, pp. 151–178. doi:10.1007/978-3-540-33037-0_7. 2

[vdMH08] VAN DER MAATEN L., HINTON G.: Visualizing data using t-SNE. J Mach Learn Res 9, 86 (2008), 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html. 2

[WH11] WICKHAM H., HOFMANN H.: Product plots. IEEE Trans Vis ComputGraph 17, 12 (2011), 2223–2230. doi:10.1109/TVCG.2011.227. 2

[Wil99] WILKINSON L.: Dot plots. Am Stat 53, 3 (1999), 276–281. doi:10.1080/00031305.1999.10474474. 3

[WVAL10] WISMÜLLER A., VERLEYSEN M., AUPETIT M., LEE J. A.: Recent advances in nonlinear dimensionality reduction, manifold and topological learning. In Proc ESANN (2010). 2

[WW95] WAGGENER B., WAGGENER W. N.: Pulse Code Modulation Techniques. Springer, New York, 1995. 3

© 2026 Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70440 by UNIVERSITY OF BERGEN, Wiley Online Library on [30/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License