Social Networks 78 (2024) 265–278

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Towards a general method to classify personal network structures

Miguel A. González-Casado [a,*], Gladis Gonzales [b], José Luis Molina [c], Angel Sánchez [a,d]

[a] Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Leganés, 28911, Spain
[b] Loyola Behavioral Lab, Universidad Loyola Andalucía, Córdoba, 14004, Spain
[c] Grupo de investigación en Antropología Fundamental y Orientada (GRAFO), Departamento de Antropología social y cultural, Universitat Autònoma de Barcelona, Bellaterra, 08173, Spain
[d] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Zaragoza, 50018, Spain

ARTICLE INFO

Dataset link: this GitHub Repository

Keywords:
Structural typology
Personal networks
Dimensionality reduction
Clustering
Social networks analysis

ABSTRACT

In this study, we present a method to uncover the fundamental dimensions of structural variability in Personal Networks (PNs) and develop a classification solely based on these structural properties. We address the limitations of previous literature and lay the foundation for a rigorous methodology to construct a Structural Typology of PNs. We test our method with a dataset of nearly 8,000 PNs belonging to high school students. We find that the structural variability of these PNs can be described in terms of six basic dimensions encompassing community and cohesive subgroup structure, as well as levels of cohesion, hierarchy, and centralization. Our method allows us to categorize these PNs into eight types and to interpret them structurally. We assess the robustness and generality of our methodology by comparing with previous results on structural typologies. To encourage its adoption, its improvement by others, and to support future research, we provide a publicly available Python class, enabling researchers to utilize our method and test the universality of our results.

1. Introduction

The guiding idea behind the study of social networks in general, and Personal Networks (PNs) in particular, is that structure matters to explain social behavior (Bott, 1957; Mitchell, 1969; Freeman, 2004). Personal Networks are shaped by biographies, personalities, opportunities, and choices but also reflect the cultural and institutional characteristics of the broader social structures in which they are embedded. Scholars have long sought to capture such complexity by creating typologies — categorizations of PNs based on their shared features. In this work we focus on a typology based purely on *structural* characteristics. A general structural typology offers a systematic way to organize and understand the diversity of structural types, paving the way for comparisons and in-depth analyses. Recognizing the underlying structure can help identify relationships between network attributes and determine which of them are essential for describing PNs. Such understanding would also allow one to compare PNs across diverse time frames and societies. Furthermore, a solid typology can unveil the underlying mechanisms – ecological constraints, cultural nuances, or individual personality traits – that drive network-building processes. By pinpointing these factors, it may be possible to formulate general theories universally applicable to PNs, irrespective of their specific context.

Most of the work published so far has not been focused on constructing a purely *structural* typology. Instead, the features used to classify PNs were mostly tie types (familiar vs. non-kin ties, strength, role and type of the support provided by each tie, namely, emotional or economic, or frequency of interaction) and other alter attributes like country of origin, context of interaction, or alter roles (Agneessens et al., 2006). In this direction, there is a whole subset of studies focused on the assessment of support networks of old people (Fiori et al., 2006; Pelle and Pappadà, 2021; Guadalupe and Vicente, 2021, 2022; Ali et al., 2022). Regarding structural typologies, the literature is relatively recent. McCarty (2002) pioneered the use of structural variables in Personal Networks Analysis. In particular, he included variables such as Density, Degree, Closeness and Betweenness Centralities, Number of Cliques, and Number of Components. McCarty's work is particularly significant to our research for two main reasons. Firstly, he underscored the intricate link between variations in the structure of PNs and differences in personality traits (see also Kalish and Robins, 2006). For instance, some people may try to merge groups, while others maintain groups separately. This may occur not only due to different levels of extroversion or prosociality, but due to different strategies used in controlling the flow of information or managing their social support. Secondly, this paper finds that a sample of about 30 alters may account

* Corresponding author.
E-mail address: migugonz@pa.uc3m.es (M.A. González-Casado).
URL: https://miguelangel-gonzalezc.github.io (M.A. González-Casado).

https://doi.org/10.1016/j.socnet.2024.03.004
Received 25 October 2023; Received in revised form 19 February 2024; Accepted 31 March 2024
Available online 9 April 2024

for the PN's basic structural properties across cases, which may help us validate the studies developed in a single context of interaction like high schools that will be presented below.

To our knowledge, the first purely *structural* typology was developed by Bidart et al. (2018). In this paper, the authors use the following structural variables: Density, Betweenness Centralization, Number of Components, Relative size of the largest component, Modularity of the partition in communities given by the Louvain algorithm, and Diameter. They develop an empirical typology based on a previous visual analysis, and then they use statistical tests to classify networks. They find six PN types: Regular Dense, Centered Dense, Centered Star, Segmented, Pearl Collar and Dispersed types. Interestingly, they find no differences in the typology regarding gender or different waves, but they find differences regarding level of education or social origin.

Subsequently, Vacca (2020) analyzed the community structure of PNs, a dimension not completely considered by Bidart et al. (2018). As Vacca states, cohesive subgroups represent a central aspect of PN structure for theoretical, empirical, and practical considerations. The author uses three variables to develop the typology, namely Modularity of the partition given by the Girvan–Newman algorithm, number of subgroups found with 1 or 2 nodes, and number of subgroups found with three or more nodes. As a conclusion, the author suggests a typology that constitutes a spectrum from core-centered to factional PNs, and mentions that this spectrum is closely aligned with theoretical typologies proposed in the sociological literature (Pescosolido and Rubin, 2000; Rainie and Wellman, 2012; Portes and Vickstrom, 2011). Furthermore, Vacca reproduces the methodology from Bidart et al. (2018) and compares both typologies. Vacca identifies the typology from Bidart et al. (2018) in other datasets, but he find almost no overlapping between this typology and his typology based on community structure. This confirms, as we will see, that both are assessing different dimensions of structural variability in PNs. Interestingly, Vacca finds a composite type, and propose that the unclassified networks from Bidart et al. (2018) belong to this type.

Another attempt to develop a typology based on structural variables was carried out by Maya-Jariego, building on Maya-Jariego and Holgado (2015) where, using Factor Analysis, the authors find that there are three dimensions of structural variability in PNs: Cohesion, Fragmentation and Integration. The idea of performing a Factor Analysis over a set of variables to identify basic uncorrelated dimensions of structural variability is appealing if the list of structural features is comprehensive enough to cover all aspects of this structural variability. The drawback is that in this work they used only four structural features. Next, based on this result, in Maya-Jariego (2021) they use one variable representing each factor and they find four types: Dense, Intermediate, Clustered and Fragmented. All these types can find their analogous types in the classifications above. Furthermore, in Maya-Jariego et al. (2020) the authors find that there is a close relation between psychological attributes and the type of PN of an individual, even at a granular level, finding correlation between specific structural features and personality traits. This relation between personality traits and structural variation of PNs is also observed in Doeven-Eggens et al. (2008).

There are other works that combine structural properties and other features like tie types or node attributes to construct the typology (Lubbers et al., 2007; Brandes et al., 2010). Although their results are not directly comparable, the typologies they find are compatible qualitatively with the ones found in purely structural analyses. Finally, there are some works exploring some alternative methodologies, such as Maya-Jariego and González-Tinoco (2023) in which they use a hierarchical deconstruction process to unveil the nested subgroup structure of the network, finding results similar to Vacca (2020); also, in Faust and Skvoretz (2002) a motif-based analysis is presented, that can be used even with negative ties. The most informative example in the direction of alternative methodologies is Giannella and Fischer (2016). They developed a purely statistical methodology based on the usage of Random Forests (RFs). This methodology has been proven useful in posterior works like Laier et al. (2022). However, the authors do not focus on the development of a purely structural typology, and hence their work is not directly comparable to ours.

The assessment of the extant literature about PNs structural typologies offers a pathway to construct a general and robust methodology able to integrate past perspectives into a single framework. The first limitation in previous works is that researchers selected different subsets of structural features to describe PNs, most of them overlapping partially or non-overlapping at all with the rest. This fact causes a direct effect on the resultant typologies developed: either they do not overlap, or they are not comparable. In some sense, if we imagine the problem of investigating the *space* of structural variability of PNs, it is like if researchers were looking to a specific subspace embedded in this space, but missing the complete picture. We will delve into this metaphor in this work. To overcome this issue, we will include the main structural metrics explored historically in Social Network Analysis and Network Science, and we will use them all in the analysis, covering both the general structure and the nested subgroup structure. In this way, we avoid arbitrary choices in the selection in the number of variables, and, moreover, we can construct bridges between previous typologies developed in the literature, enabling direct comparison between apparently disconnected network types. This advance, nonetheless, comes at a price: By introducing a large number of structural features we increase the dimensionality of the problem. This makes the problem less interpretable, and makes more difficult its mathematical analysis. As we will see, this difficulty can be overcome, by means of reducing the dimensionality of the problem. To do so there is a wide range of techniques. There are two paradigmatic works already mentioned here: Giannella and Fischer (2016) and Maya-Jariego and Holgado (2015). The first uses Random Forests to reduce the dimensionality of the problem, and the second relies on Factor Analysis to do so. We have preferred to use Factor Analysis due to its common usage in the social sciences and its explainability, although both alternatives are adequate. By reducing the dimensionality of the problem *after* considering all possible structural features we can gain deep insights into the dimensions of variability of PNs. Basically, we can understand which variables developed historically provide redundant structural information and which is the bare minimum of variables we need to understand the structure of a PN. Once we have constructed the reduced space of dimensional variability of PNs we can construct the typology geometrically and interpret it in terms of the original structural variables afterwards. Finally, our work is also novel in the sense that one of its central purposes is to reduce as much as possible the amount of choices made when conducting the analysis, for instance, the construction of a preliminary empirical typology based on a visual analysis. Instead, we provide a blind structural typology where the relation with personal attributes or tie types could be explored afterwards.

2. Data and methods

We define the PN of an individual (from now on, 'ego') as the network composed by all the people around ego with whom ego has a personal relationship (from now on 'alters'), and the relations between them. In other words, in the PN of ego, nodes represent alters and edges represent personal relationships between alters (e.g., friendships). Notice that, in our definition of a PN, ego is not included in the network, we are only interested in how the personal relationships between alters are organized. It is important to mention at this stage that the terminology employed in Personal Networks research is sometimes not properly standardized, and concepts can become somewhat problematic. In certain sources, the term 'ego network' or 'egocentric network' is used to describe networks where ego is represented along with the individuals to whom ego is connected (alters), excluding the links between alters. Conversely, in other sources, egocentric networks encompass both ego and the links between alters but as subset of a whole network. The term 'Personal Network' has been used to denote networks where ego is absent, and only the links between alters are considered, with individuals nominating others irrespective of the social contexts in which they encountered them. An important point to bear in mind is that, in our data, ties in PNs represent exclusively friendships within the high school, leaving other types, like familiar ties, outside of the network. However, we believe that using the term 'ego network' or 'egocentric network' could lead to confusion, given that ego is not present in the network, so we will be using the term Personal Network when speaking about our data. In any case, the method is applicable independently of the ties included, as it is only dependent on the structure of the network.

2.1. Data collection

In this study, we analyzed a sample of 7969 PNs (see below for details). This corpus of PNs is derived from five different datasets, pertaining to high school students in several parts of Spain and El Salvador. These high school data were collected through surveys administered in schools via a computer interface. To elicit relationships, students were presented with a list of all other students in their respective high schools. They were then asked to select individuals with whom they had relationships, designating them as friends (+1) or best friends (+2). These designations constitute the links within the PNs. Almost every student provided this list, resulting in the extraction of a weighted directed network for the entire high school, thus allowing to straightforwardly extract the PN of each respondent. In section S5 of the Supplementary Material we explain the treatment given to the missing data, corresponding to the students who did not answer the surveys. The questionnaire included the following question: "You can now see the list of all the students in the school. Please mark those you have any relationship with by clicking 'very good relationship' or 'good relationship'. Only one choice is possible. If you do not mark any option, it will be understood to mean that you do not have a relationship with the person". Typically, it took students about 15 min to complete the survey, and they were supervised by a school teacher throughout the process. The study was integrated into the schools' academic curriculum, and our Institutional Review Boards (IRB) stipulated an opt-out procedure. Importantly, there were no opt-outs, effectively eliminating any potential selection bias. The only students who did not participate were those who were absent on the day of the experiment. More information about this data can be found in Escribano et al. (2021, 2023).

It is important to note the significance of the Name Generator used when eliciting PNs from respondents. The choice of a Name Generator has direct implications for the structural features observed in PNs and is a subject of research in its own right (see, for instance, Campbell and Lee (1991); Marin and Hampton (2007); Eagle and Proeschold-Bell (2015)). While our objective is not to comprehensively argue for one generator over another, it is crucial to ensure that the PNs being compared are elicited using similar name generator strategy. Specifically, when networks are elicited with a fixed number of alters (e.g., 30), this fact imposes constraints on the structural variability observed within the networks. Consequently, such networks cannot be directly compared with those elicited using a variable number of alters. We have exclusively used PNs elicited using a variable number of alters. Our rationale is that, as we explore the complete space of structural variability of PNs, the use of a fixed number of alters restricts the region of this space in which PNs can be placed. This limitation arises due to the correlations that exist between the number of nodes in a network and other structural features, such as centrality measures and community structure. We delve deeper into these correlations in the Supplementary Material, where we perform some Exploratory Data Analysis on our data (section S3).

2.2. Description of the data

The raw dataset for this study comprises 9070 PNs. From this initial pool of PNs, we excluded 876 networks for two reasons: either they were empty networks or composed of only 1 or 2 nodes, limiting the structural variability displayed by the networks and potentially confounding the results. The reasons for their exclusion are purely mathematical. Nonetheless, we will come back to these networks in the discussion. This removal left us with a dataset of 8194 networks. After measuring all structural variables, we identified and removed 225 outliers (details about the outlier elimination process, based on Liu et al. (2008); Breunig et al. (2000) and Schölkopf et al. (1999), can be found in the Supplementary Material, section S2), resulting in a final dataset of 7969 networks available for analysis. This data was finally standardized, subtracting the mean and dividing by the standard deviation for each metric (all plots show standardized values in all structural metrics). In section S3 of the Supplementary Material one can find some descriptive statistics for the metrics before standardization, to ease interpretability. The data description is as follows:

- **Country of residence:** 76.01% of egos reside in Spain, while 23.99% reside in El Salvador.
- **Datasets:** The data is collected from different datasets, and we will refer to them by their names during the analysis:
  Dataset A: One high school in Madrid (Spain) (4.04%).
  Dataset B: One high school in Luarca (Spain) (2.18%).
  Dataset C: Two high schools in Madrid (Spain) (7.83%).
  Dataset D: 19 high schools in Catalonia (Spain) and one high school in Andalusia (Spain) (22.60%).
  Dataset E: 13 high schools in Andalusia (Spain) and 12 high schools in different locations at El Salvador (63.35%).
- Regarding sex and gender, there is a mismatch between surveys. On the one hand, for datasets 'A', 'B' and 'C' (14.05% of the networks) we have information about the sex of the participants (in the case of the schools, it was directly obtained from the school listings, so no further question was included in the survey). Nonetheless, surveys for datasets 'D' and 'E' (85.95% of the networks) were updated with questions about the self-identified gender of the participants. Thus, regarding sex and gender, we will separate both analyses to assess both differences in sex and differences in gender in the following sections.
  For the 14.05% of networks for which we have information about sex, 53.21% of networks are from males, and 46.79% of networks are from females.
  For the 85.4% of networks for which we have information about gender, 49.74% of respondents are self-identified as men, 48.58% are self-identified as women, and 1.68% are self-identified as 'Other' or prefer not to answer.

We have not extensively studied the composition of the datasets regarding the age of the respondents. Nevertheless, it is crucial to mention that all students from datasets A-E fall within the age range of 12 to 18 years old. It is also important to point out that the PN of each person is extracted from the complete network of the respective high school. Thus, when analyzing these networks structurally, we could observe a statistical dependence between the sociocentric networks of the different schools and the individual PNs of students belonging to these schools. In this respect, the complete networks of some schools may be denser or more clustered, potentially influencing the structural characteristics of the individual PNs and possibly creating a statistical interdependence between individual observations. In any event, this is not a problem to present our methodology through the application to this dataset.

2.3. Methodology

2.3.1. Obtaining the fundamental dimensions of variability

Our point of departure is a collection of adjacency matrices representing weighted directed networks (Newman, 2018). In simple terms, for each ego we have a network in which nodes represent alters they have a relationship with, and edges represent personal relationships between these individuals. These relationships are both directed, meaning each alter specifies their relation with the others, and weighted, signifying that the relations between alters are characterized by varying degrees of intensity. To initiate our analysis, we transform each weighted directed network into an undirected and unweighted version. This transformation removes the directionality, resulting in a network where edges indicate the mere existence of a relation between two alters, with all relationships being considered of equal intensity. The main reason for this transformation is that it aligns our study with the previous research mentioned in the Introduction, which predominantly examined undirected and unweighted PNs, ensuring the comparability of our results. In practical terms, these results largely depend on the collection of structural metrics used to develop the typology, and the definition and even interpretation of these metrics change if the network is weighted and directed. We are aware that by this transformation we are losing information, but comparability with previous approaches is one of our core purposes, and therefore we leave the assessment of the weighted directed version of our PNs for a future work. There are several ways to convert a weighted directed network into an unweighted undirected version, depending on the network's representation and purpose. In our case, where we study personal relationships, we simplify the network by disregarding the intensity of relationships. In essence, we treat both friends and best friends as representing the presence of a personal relationship, in contrast to the absence of a relationship. Thus, 0 weights in the network remain as 0 weights, while +1/+2 weights are transformed into 1 weights. Regarding directionality, we consider a relationship to exist only when it is reciprocal. Consequently, bidirectional links where each alter designates the other as a friend or best friend are retained as personal relationships, while edges where only one of the alter designates the other as a friend or best friend are eliminated.

Once we have this collection of unweighted and undirected adjacency matrices representing PNs, we compute a comprehensive list of 41 network structural metrics, drawing from well-established sources in Network Theory and Social Networks, including Wasserman and Faust (1994), Estrada (2012), Estrada and Knight (2015) and Newman (2018), as well as other specific sources like Burt (1995). These metrics cover various aspects of network structure, including connectivity, closure, local and global centralities, distances, community and subgroup structures, structural holes, etc. A complete list of these variables, along with their mathematical definition and their interpretation can be found in the Supplementary material (section S1). It is important to note that while our initial list of metrics was larger, some were excluded from the analysis for four reasons: they lacked clear definitions for fragmented networks, exhibited negligible variability without significant implications for identifying network types, introduced misleading correlations, offering redundant information, or were binary in nature. Excluded variables and the reasons for their exclusion can also be found in the Supplementary material.

We calculate all these metrics for each of the PNs we have collected, constructing a dataset containing the values of these 41 variables for each PN. We can envision these networks as points in a 41-dimensional space, where each dimension represents one of the metrics we measure. This space, with each dimension corresponding to a structural metric, constitutes the 'space of structural variability of PNs': the space in which PNs are depicted as points positioned differently based on their structural 'coordinates'. However, some of the metrics we measure may provide redundant information about the structure of our PNs. This implies that although our space has 41 dimensions, not all of them are necessary. To identify this, we analyze the correlations present in our data and employ dimensionality reduction techniques. These correlations indicate that when we place all our PNs in this 41-dimensional space, points tend to accumulate in a subspace of dimension $m \leq 41$, rendering the remaining dimensions redundant. Consequently, by identifying this subspace and constructing a basis for it, we can describe our data in a simpler and more interpretable manner. This basis for the subspace comprises the fundamental dimensions required to explain the observed structural variability, built from linear combinations of the original metrics.

At this point, the mentioned limitation of previous works becomes evident: using only a small set of metrics means exploring the structure of your PNs within this restricted space of structural variation. We will link our results with each of the previously mentioned studies in the Introduction and position them within this space, allowing us to comprehend how these researchers explored different aspects of the same problem. This reveals that while their results may not be directly comparable, they do contribute to the broader picture.

Once we have the dataset with all the metrics for all the PNs we remove outliers (Supplementary Material, section S2) and we perform some Exploratory Data Analysis (Supplementary Material, section S3). Subsequently, we perform some standard tests to ensure that the variables exhibit sufficient correlation and well-behaved distributions for subsequent dimensionality reduction and clustering analyses (see Bartlett (1950); Mulaik (2009)). Now, let us delve into the dimensionality reduction process. We have employed two techniques widely used in statistics: Principal Component Analysis (PCA) (Jolliffe and Cadima, 2016) and Exploratory Factor Analysis (EFA) (Mulaik, 2009). Both methods take the covariance matrix of the original data and find a new basis in which to express this covariance matrix according to some criteria. Specifically, PCA finds a basis such that the covariance matrix, expressed in this basis, is diagonal and ranked (the elements of the diagonal, that is, the variances, appear in decreasing order). The elements of this basis are called Principal Components (PCs), and they are by definition orthogonal and uncorrelated. These PCs are the dimensions that explain the largest amount of variance by themselves in decreasing order. Hence, if we want to reduce the dimensionality of the problem to dimension m, we simply need to keep the first m PCs that express the largest amount of variance, take them as the basis of the fundamental subspace of structural variability, and the variance expressed by the rest of PCs is treated as noise. On the other hand, in EFA we want to find a basis as well, but a basis that explains only the off-diagonal elements of the covariance matrix. This technique is somewhat more sophisticated, as it assumes some intrinsic behavior to each variable apart from the behavior explained by this new basis. The elements of the basis are referred to as Factors, and they constitute again the basis of the fundamental subspace of structural variability. It is important to mention that in this work we use the *Principal Axis Factoring* method (Mulaik, 2009) to extract the factors. In principle, both methods could give very different results if the terms associated with the diagonal elements of the covariance matrix not explained by the factors are too large. Nonetheless, if the basis of factors explains not only off-diagonal elements of the covariance matrix, but also the diagonal elements, results given by both methods tend to converge. Indeed, this the case in this work, as both results are completely equivalent.

Up to this point, the concept of 'basis of fundamental dimensions of structural variability', with these dimensions being the PCs in one method and the Factors in the other method, is a bit abstract. We need to know how these dimensions are related with our original structural metrics, to be able to explain the variations in the structure of PNs in simple terms. In principle, these dimensions can be expressed as linear combinations of the original metrics. So, each metric is linearly correlated with some dimensions. To summarize these relations between metrics and dimensions we use the *loadings matrix*. Basically, it is a matrix that contains the correlation coefficients between the original metrics and the elements of the basis of the fundamental subspace (the PCs and the Factors). For instance, with this matrix we can see if the structural metrics *density*, *number of triangles* and *global efficiency* are correlated with the first Factor found with the EFA (i.e., the first element of the basis of the fundamental subspace of structural variability, or the first dimension of structural variability). With this loadings matrix we are able to group variables together and associate them with the elements of the basis of the subspace. In this way, we can understand how the data is structured without the need to assess 41 different variables, but analyzing a few number of dimensions that group these variables together, thus improving the interpretability of the result.

Although both EFA and PCA give equivalent results in our case and we have included both just for completeness, there is a practical advantage in using EFA over PCA. In principle, PCA gives a basis in which the first element explains the largest amount of variance in the data, the second element explains the second largest amount of variance, etc. Depending on the factor extraction method, this might be the case as well for EFA. Nonetheless, this might not be the best way to express the basis, and one may want to rotate this basis so the loadings matrix becomes more interpretable. Most packages including EFA techniques allow to make different rotations. We have used the *varimax* rotation (Mulaik, 2009), that rotates the basis to ensure that each dimension correlates with a few variables, and that variables correlate only with one dimension, at least as far as possible. Besides, we are aware that there are alternative, more sophisticated options available to perform non-linear dimensionality reduction, like the t-Distributed Stochastic Neighbor Embedding (Hinton and Roweis, 2002) or Kernel PCA (Schölkopf et al., 1997), but they go beyond the scope of our work, as our main purpose is to perform an informative general taxonomy of PNs that integrates previous perspectives in the field.

Using EFA or PCA we find the m dimensions that best explain how structure in PNs varies and group together the original metrics we used. Nonetheless, we have to choose the exact value of m. There is a trade-off between precision and simplicity. If we choose m to be large we may overfit the data, decreasing explainability. Conversely, if we choose m to be small we may underfit the data and lose important information. The point is to choose the amount of variance our PCs/Factors will not explain. There are different methods to choose this number, spanning from simple methods like the rule of thumb of choosing m such that the dimensions chosen explain at least 90% of the variance present in the data, the scree plot or the Kaiser criterion, to more sophisticated methods like the one used here, Horn's Parallel Analysis (Horn, 1965), which compares the correlation matrix to a null model, to determine the number of dimensions necessary for representing non-trivial correlations within the data. Specifically, we employ a non-parametric version of this test developed by Buja and Eyuboglu (1992) that avoids normality assumptions. For justification see Zwick and Velicer (1986) and Dobriban (2020).

2.3.2. Classification: clustering analysis

Once the original data is curated and projected onto a subspace that encompasses the fundamental dimensions of structural variability of PNs, we are in the proper position to develop the typology, the central purpose of this work. Basically, we want to make a classification of networks *within this subspace* and then translate the result to the original structural properties, so that each group is characterized by its structural metrics. Nonetheless, this clustering analysis involves certain subtleties. The most important point to bear in mind is that from the distribution of networks within this subspace we know that our data lacks a clear clustered structure (if it had a clear clustered structure, the diagonal distributions in Fig. 2 would be multi-modal). So, it is essential to establish our objective clearly: *we are not aiming to identify an underlying structure of clusters within the 41-dimensional space. Rather, we seek to address the question of how to classify these networks effectively*. Thus, our aim is to perform a classification we can connect with sociological, psychological and other traits, understanding that the transition between different types within the taxonomy is continuous, and that we may have mixed types, networks that are difficult to classify due to their intermediate position between clusters, etc. This fact directly bears on how we choose the algorithm to perform the classification. There are two caveats that need to be taken into account when choosing the clustering algorithm: first, we want an algorithm that gives a differentiated typology where clusters can be interpreted easily, and second, our data lacks a clustered structure. We stress that, since there is no *a priori* ground truth about the taxonomy, the selection of a different algorithm would serve to a different purpose, but that this selection is not itself *right* or *wrong*.

Considering our requirements in terms of the creation of a differentiated typology, with balanced clusters and amenable to interpretation, we decided to use the *k-means* algorithm (Ahmed et al., 2020). The drawback of this algorithm is that one needs to select manually the number of groups in which to perform the classification. When there is a ground truth, or, at least some intuition about the clustered structure of the data, there are well-known algorithms such as the Silhouette Analysis (Rousseeuw, 1987) to perform this task. Nonetheless, due to the lack of a clustered structure in the data, traditional methods for determining the appropriate number of clusters will not be as effective, as these methods are designed under the assumption of an inherent clustered structure within the data. The problem is that the objective of this work is to maintain a level of generality and avoid arbitrary decisions during the classification process. Moreover, our goal is to create an informative classification for a corpus of networks with continuously varying structures. So, the strategy we will use is to perform a classification in different number of clusters and use two criteria, *redundancy* and *consistency*, to discuss which is the most suitable number of clusters, taking into account that no number is itself a better selection but in terms of interpretability and sociological relevance, looking for non-trivial as well as non-redundant results.

To this end, we will divide the corpus of networks into a range of groups, spanning from 2 to 11. We will analyze the overall structural properties for each classification and subsequently evaluate which classification exhibits less *redundancy* and greater *consistency*. Redundancy, in this context, refers to the informativeness of the classification. For example, if clusters in an 8-cluster partition exhibit the same structural properties as those in a 7-cluster partition, except for a minor difference in one variable, we will consider the 7-cluster partition as less redundant and therefore better. Our aim is to construct a taxonomy of distinct clusters while being mindful that, given the continuous structural variability in the 41-dimensional space, there is always the potential to add an extra cluster that differentiates itself by only a single feature. On the other hand, consistency pertains to the robustness of the classification across different realizations. For each chosen number of groups, we will perform clustering multiple times and compute the normalized mutual information score (NMI) (Vinh et al., 2009). The more consistent classifications will have NMIs closer to 1.

All in all, these are the most important points regarding the methodology we developed to construct a general structural typology of PNs. We have developed a Python class that we have made available publicly (see Data availability section). This class is intended to be accessible for researchers with limited experience in these mathematical methods, allowing them to execute the code in Python and conduct similar analyses on diverse datasets. By sharing this package, our aim is to facilitate the replication and application of our analysis in various research contexts.

3. Results

In this section we explore the application of the developed methodology to our dataset of PNs belonging to high school students. We assess the generality of the results, the performance of the method and the limitations due to the specificity of the data.

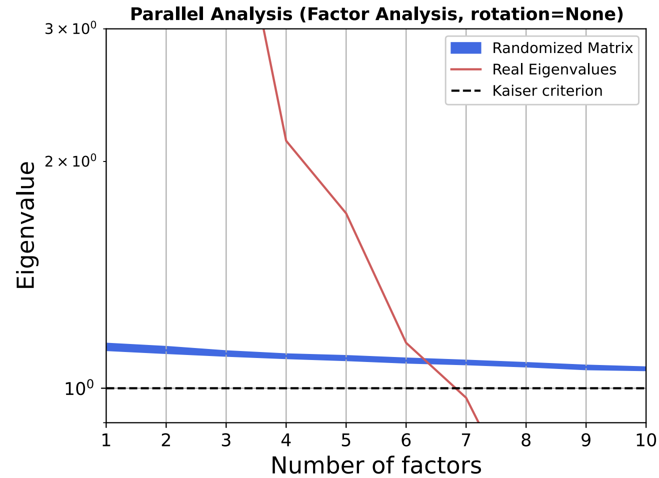

Fig. 1. Results for the Parallel Analysis to determine the number of factors to retain. In red, the eigenvalues of the matrix of data depicted in decreasing order; in blue, the eigenvalues associated to a randomized version of the data (the width of the line correspond to percentiles 95%-top and 5%-bottom of the distribution of random eigenvalues); in black, the Kaiser criterion. The Figure is zoomed in to the region of interest, that is, to the region in which red intersects the blue and black lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. Fundamental dimensions of structural variability

As we mentioned, the number of fundamental dimensions is computed using a non-parametric modification of the Horn's Parallel Analysis (HPA). Specifically, this modified HPA is a Monte Carlo based method that compares the eigenvalues obtained from the real data to those from a null model of uncorrelated random variables, obtained from permutations on the original data. A factor or component is retained if the associated eigenvalue is bigger than the 95th percentile of the distribution of eigenvalues derived from the random data. In Fig. 1, this translates into keeping the number of factors for which the red line (real eigenvalues) is above the top limit of the confidence interval of the blue line (distribution of random eigenvalues). In this case, since this intersection occurs between 6 and 7 factors, we determine that the proper number of dimensions to keep is 6. In simple terms, with this procedure we keep as many dimensions as needed to explain the non-trivial correlations present in the data, i.e., correlations that are not expected at random. In this same Figure we depict another criterion, namely, the Kaiser criterion, which poses that we keep the number of factors for which the associated eigenvalue is above one. As it is seen in the Figure, this may lead to an overestimation in the number of factors, since the red line intersects the black line after the blue line. Thus, our analysis indicates that the data can be accurately represented in a subspace of 6 dimensions. In other words, the networks, originally existing in a 41-dimensional space, can be projected onto a 6-dimensional space, accounting for approximately 90% of the variance. As stated before, this 6-dimensional subspace can be represented by any choice of a 6-dimensional basis that results from the rotation of the original basis given by the EFA or the PCA. As a result, in order to interpret what is the information conveyed in this subspace, there are multiple possible solutions, and we must make an arbitrary decision regarding the rotation of the basis to use. As stated above, the dimensions discussed here are obtained using a *varimax* rotation. The specific rotation chosen is relevant only in terms of interpretability, but it does not alter the fact that our data can be effectively described in a 6-dimensional subspace. Furthermore, the selection of a specific rotation does not impact subsequent clustering analyses.

Fig. 2 depicts the PNs in our datasets as points in the 6-dimensional fundamental subspace of structural variability along with the loadings matrix to understand how our structural metrics are related with these 6 dimensions. From these plots we can infer the relation between the 6 fundamental dimensions of structural variation and the structural metrics we measure on the PNs:

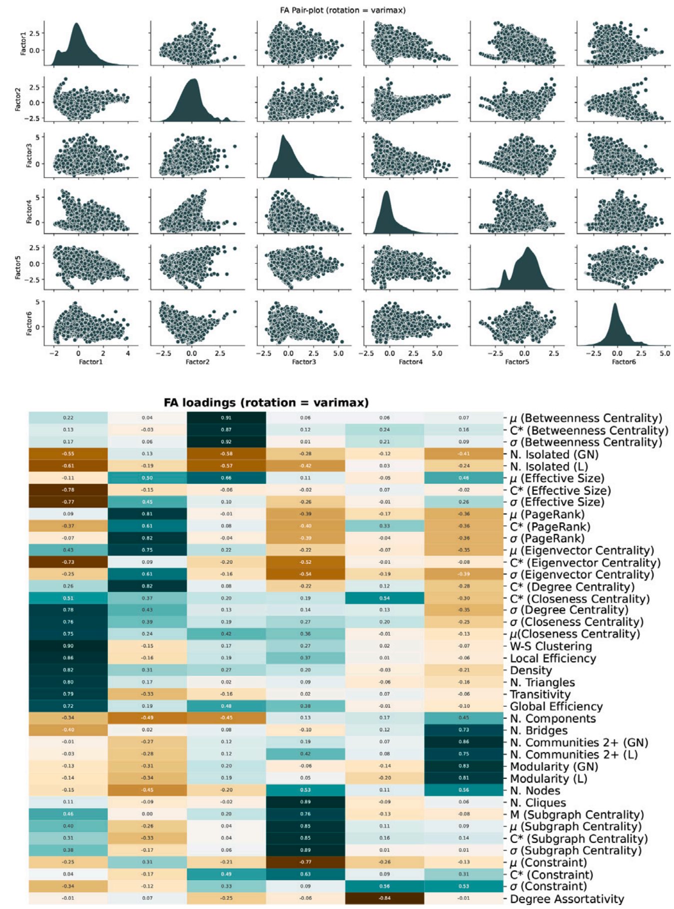

Fig. 2. Top: Pair-Plot of the data projected onto the 6-dimensional subspace. Each diagonal element represents the distribution of the data along one of the six dimensions, and the (i, j) off-diagonal elements represent the relation between the coordinates of the data in dimension i and the coordinates of the data in dimension j. Thus, each point represents a PN in the fundamental subspace of structural variability. **Bottom:** Loadings Matrix. Columns in this matrix represent the 6 dimensions of the fundamental subspace of structural variability. Rows are the structural metrics we measure in the PNs. Hence, each element of the matrix represents the correlation coefficient between each metric and each dimension. Here, M refers to the median and C* to the centralization.

- **Dimension 1** exhibits positive correlations with variables associated with network cohesion, specifically density, closure, efficiency and average closeness and degree centralities. Conversely, it displays negative correlations with measures of centralization in eigenvector and PageRank centralities, variation in effective size, and number of isolated nodes. Dimension 1 appears to capture variations in the cohesion of the network, and it accounts for 24.4% of the variance in the data.
- **Dimension 2** is positively correlated with measures of variance and centralization in degree and closeness centralities, and with mean and variance of Eigenvector and PageRank centralities. Dimension 2 seems to point towards the existence of hubs in the network. It accounts for 22.9% of the variance in the data
- **Dimension 3** correlates positively with measures of mean, variance, and centralization of Betweenness Centrality, as well as average effective size of networks. It also displays negative correlations with the number of isolated nodes. Dimension 3 appears to capture the variation in network structure from networks with

central nodes connecting groups or networks with cohesive cores and satellites to decentralized networks, and it accounts for 14.1% of the variance in the data.
- **Dimension 4** is positively correlated with all measures of subgraph centrality and the number of cliques. Furthermore, it exhibits a high negative correlation with the mean constraint (here, constraint refers to one of the 41 structural metrics associated with the concept of Structural Holes, and it is defined in the Supplementary Material), and a moderately high correlation with the centralization of the very same variable. Dimension 4 appears to capture variations in the network structure in cohesive subgroups, which should not be confused with the structure in communities, and it accounts for 17.4% of the variance in the data.
- **Dimension 5** seems to be a residual dimension, as it only correlates marginally with three unrelated variables, one of them being the Degree Assortativity which, as discussed in the Supplementary Material, might be a bit problematic due to its correlation patterns. It accounts for 4.9% of the variance in the data
- **Dimension 6** demonstrates positive correlations with the number of communities consisting of two or more nodes, the modularity of the partition in communities, the number of bridges, components and nodes. Dimension 6, seems to capture variations in the community structure of the network, and it accounts for 4.1% of the variance in the data.

In spite of the insights they provide into the different aspects of structural variability in the PNs, we refrain from assigning specific names to these dimensions to emphasize the arbitrary nature of the choice of rotation. Indeed, the usage of Factor Analysis (Furfey and Daly, 1937) has been criticized when accompanied by the assignment of a special ontological status to the factors, as if they were not mathematical constructs to understand the structure of the space of data. Instead, we limit ourselves to use factors to group metrics together and understand what they represent, as one could do with any other dimensionality reduction statistical technique.

3.2. Structural typology of personal networks

Now that we have characterized the subspace in which we can understand the structural variability of our PNs, we can proceed to the construction of the typology. As can be clearly seen in the top panel of Fig. 2, and widely discussed in previous sections, the data does not present a clear clustered structure. Thus, we need to discuss which is the appropriate number of clusters to use based on redundancy and consistency.

Let us begin the assessment of the partitions for different numbers of clusters from the perspective of redundancy. All the Figures supporting this summary of results can be found in the Supplementary Material (section S4). A **2-cluster** partition separates small networks with low cohesion and a large proportion of isolated nodes (i.e., nearly empty networks) from larger networks with diverse structures, greater cohesion, presence of communities, and so on. A **3-cluster** partition retains the group of nearly empty networks and divides the other group in two. The first group is characterized by high cohesion and subgraph centrality measures, while the second group exhibits a high number of communities, modularity, and high values in betweenness centrality measures. Moving on to the **4-cluster** classification, we identify a distinct group: networks with numerous cliques, high subgraph centrality measures, and substantial cohesion. Additionally, we still observe groups with community structure and nearly empty networks. The fourth group is marked by moderately high cohesion and eigenvector centrality metrics. As we will discuss later, by combining subgraph centrality measures, the number of communities, and modularity, we can differentiate well-connected networks (those with high cohesion values) based on their size. The **5-cluster** partition resembles the 4-cluster one, with the introduction of a new group that consistently

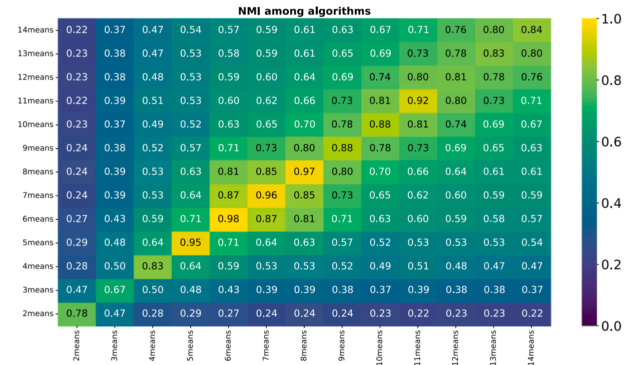

Fig. 3. Normalized mutual information score, averaged over 100 realizations, between each choice of k in the $k-means$ algorithm.

appears from this point on: a group with high values across all betweenness centrality metrics and mean effective size. When we consider a **6-cluster** partition, all previously identified types become more differentiated, and we discover a sixth group, which could be characterized as a 'composite' group with moderate values in all structural features, positioning it between networks with a diverse structure and nearly empty networks. Subsequently, as the number of clusters increases, the differences become more subtle, and larger cluster partitions tend to exhibit redundancy. For example, in a 7-group partition, the only noticeable difference is the division of nearly empty networks into smaller and larger ones, without identifying *significant* structural variations. In the 8-cluster partition, all previous groups persist, and a group sharing features with the nearly empty networks and the group with high betweenness centrality measures emerges, spatially situated between the other two. These examples are informative because, from this point on, new groups tend to fall between previously identified ones, without contributing any new insights to the spatial structure of the data. While these classifications remain informative in terms of understanding spatial distribution and transitions between cluster types, they do not introduce entirely 'novel' groups. Thus, a **9-cluster** partition situates a group between the composite and nearly empty types. In a **10-cluster** partition, a new group emerges between the community structure and composite networks. An **11-cluster** partition includes a cluster with high subgraph centrality but moderate cohesion values, and so on.

Building on this assessment of the structural variability in PNs, we can address the question of what constitutes a sociologically relevant classification of PNs by searching for the optimal number of clusters that minimizes redundancy while maximizing differentiation. To this end, let us consider consistency as well. When we evaluate NMI values across different realizations, partitions with 5, 6, 7, and 8 clusters exhibit the highest consistency. As shown in Fig. 3, NMI values along the diagonal increase as we progress from 2 clusters and reach nearly 1 for partitions with 5, 6, 7, and 8 clusters. As we continue to increase the number of clusters, NMI values decline. The above discussion indicates that a 6-cluster partition is the most informative choice in terms of redundancy, allowing to understand and describe the internal structure of the space of structural variability of PNs. On the other hand, in terms of consistency, partitions with 5, 6, 7, or 8 clusters perform best. This suggests us to focus on the 8-cluster partition, bearing in mind that it is qualitatively similar to the 6-cluster partition but includes two clusters representing nearly empty networks. These clusters occupy a region of space where structural variables are particularly sensitive to changes in structure. Thus, the 8-cluster partition will have a couple of redundant groups, but it will be the best in terms of consistency and redundancy. The corresponding characteristic features of each cluster can be found in Fig. 4. The clusters can be described as follows:

- **Large Networks**

  Networks with community structure. This is the second largest cluster (Brighter Orange in Fig. 4), encompassing just under 1/4 of the networks in the dataset. It consists of large networks with numerous components, a high number of communities with two or more nodes, high modularity, high variance of the constraint, and a significant number of bridges. However, it exhibits low values in measures of density, closure, and mean and variance of degree and closeness centralities, as well as average eigenvector and PageRank centralities.

  Well connected networks with subgroup structure. This cluster comprises around 1/8 of the networks in the dataset (Green in Fig. 4). It consists of large networks with high values in measures of subgraph centrality, a large number of cliques, and a substantial number of communities and components, but with lower modularity with respect to the previous group. Additionally, it demonstrates high values in all metrics of cohesion. Moreover, it exhibits low values in mean and deviation of eigenvector and PageRank centralities, number of isolated nodes, and variation of effective size.
- **Medium-Sized Networks**

  Composite networks. This is the largest cluster, accounting for just over 1/4 of the dataset (Darker Blue in Fig. 4). It consists of medium-sized networks which display moderate values in all variables, although characterized by a number of isolated nodes and centralization in effective size slightly above average, and mean effective size and measures of cohesion slightly below average.

  Dense BC networks. This cluster encompasses around 1/8 of the networks (Red in Fig. 4), and consists of medium-sized networks with high values in all measures of Betweenness Centrality, mean effective size, metrics of cohesion, centralization of the constraint. Additionally, it exhibits low values in the number of isolated nodes and presents some community structure.
- **Small Networks**

  Dense networks. This cluster accounts for almost 1/8 of all the networks in the dataset (Purple in Fig. 4). It is characterized by high values in measures of cohesion (density, closure, degree, and closeness centralities, local and global efficiencies). Furthermore, it exhibits high values in all metrics of eigenvector and PageRank centralities and mean constraint. On the other hand, it shows low values in the number of bridges, the number of communities with two or more nodes, modularity, number of components, all measures of Betweenness Centrality, mean and variation of the effective size and centralization of the constraint.

  Sparse BC networks. This cluster comprises only 1/16 of the networks (Yellow in Fig. 4). It is characterized by moderately large values in all metrics of Betweenness Centrality, as well as all metrics of eigenvector and PageRank centralities, deviation in degree and closeness centralities, and mean and deviation in constraint and effective size. Conversely, it displays low values in metrics of subgraph centrality, number of cliques, closure, number of communities and modularity.

  Sparse networks I and II. These are the two smallest clusters, representing together 1/6 of the networks (Lighter Blue and Dull Orange in Fig. 4). They consist of small networks characterized by a large number of isolated nodes and high measures of eigenvector and PageRank centralities. The differences between both groups derive from their differences in size and the sensitiveness of the variables due to small changes in structure in that region of the space of structural variability.

3.3. Interpretation of the types

To better understand the taxonomy presented, we depict the networks of the dataset that are closer to the centroids of each cluster in Fig. 5. Once again, it is essential to emphasize that the networks depicted in this Figure do not serve as representative centers of distinct, well-defined clusters. Instead, they are paradigmatic examples within a classification where networks exhibit *continuous* variations between clusters within the space of structural variability of PNs. This concept is easier to understand if we try to visualize the 6-dimensional subspace of structural variability of PNs. Rather than visualizing a 6-dimensional space, let us imagine that our fundamental subspace of structural variability for PNs is 3-dimensional. Specifically, this space comprises the dimensions we labeled as 1, 4, and 6. Dimension 1 represents the variation in the cohesion of the network, dimension 4 signifies variation in the subgroup structure (subgraph centrality) of the network, and dimension 6 denotes variation in the community structure of the network. In the left panel of Fig. 6, it becomes apparent how we can seamlessly transition from one network to the next, and how its appearance changes as we move within this space. To illustrate, let us focus on the dimension of 'Cohesion'. We observe a spectrum that spans from networks with low cohesion (purple, darker green, and darker blue) to those with high cohesion (red, orange, and finally yellow, lighter green, and lighter blue). Concerning community structure, we similarly witness a spectrum ranging from the light blue network, moving to orange, yellow, and lighter green, and eventually transitioning to red, darker blue, and purple. Finally, regarding subgraph centrality, our spectrum shifts from the two blue networks and the darker green one, then proceeds through purple, yellow, and orange, and ultimately arrives at red and lighter green networks. By combining these three dimensions, we can smoothly navigate from, for example, the purple network to the red network by slightly increasing cohesion, decreasing community structure, and increasing subgraph centrality. This diagram illustrates how variations in different structural metrics, aligned along their respective dimensions, impact the overall network structure. We can interpret both the typology developed and the fundamental dimensions of structural variability.

In the right panel of Fig. 6 there is another example of a three-dimensional subspace depicted to understand dimensions 2 and 3 as well, which represent variations in Eigenvector-like centrality metrics and Betweenness Centrality metrics, respectively. Probably, dimensions 2 and 3 are the most difficult to visualize, as they are less intuitive, specially dimension 2, that correlates with Eigenvector and PageRank centralities. In principle, high values in mean, variance and centralization of these two variables would help us identify networks with highly central nodes, and thus this dimension could point towards networks with highly central nodes. Nonetheless, these two variables have high correlations with other variables as well, specially a negative correlation with the number of nodes. Thus, the identification of highly central nodes will be mediated also by the size of the network. In other words, high values in these variables will help us identify networks with highly central nodes only if these networks are small. This is clearly identified in the right panel of Fig. 6. As we go to higher values in the 'PageRank' axis, we identify smaller and emptier networks, where the only connected nodes are, obviously, highly central nodes compared with the rest of isolated nodes. On the other hand, in order to understand dimension 3, i.e. the dimension that exhibits positive correlations with all metrics in Betweenness Centrality metrics, the key point is to choose a proper representation of the network. In Fig. 7 we depict three networks that display high values in Betweenness Centrality. However, we depict them in a way that makes clear what high values in Betweenness Centrality metrics mean. In these networks, high values in these metrics imply the presence of highly central nodes that mediate between cohesive groups. Depending on how the rest of the network is organized, it could happen that these nodes are placed in a cohesive core which has satellites connected to it, or placed connecting communities within the network.

With this understanding of the representatives and features of the different types, we can now present the key characteristics of our proposed taxonomy. To begin with, a clear distinction exists between

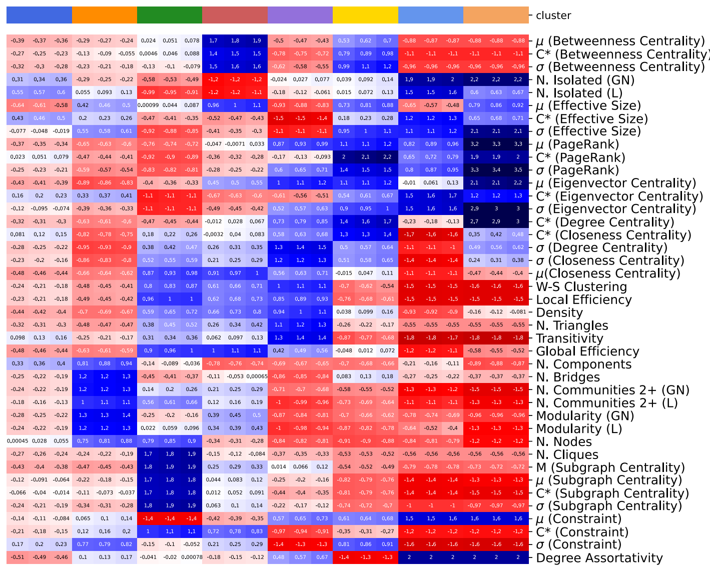

Fig. 4. Heat-Map of the values of the structural metrics that characterize each cluster in the 8-cluster partition. Each row in the heatmap represents one of the 41 metrics measured for each PN. Here, M refers to the median and C* to the centralization. The columns are divided into eight clusters (colors). Within each cluster, there are three values displayed for each row. The central value represents the mean value of the standardized metric for that particular cluster. The values on the right/left sides indicate the upper/lower limits of the 95% confidence interval. The clusters are defined as follows **Darker Blue:** Composite networks. **Brighter Orange:** Networks with community structure. **Green:** Well connected networks with subgroup structure. **Red:** Dense BC networks. **Purple:** Dense networks. **Yellow:** Sparse BC networks. **Lighter blue:** Sparse networks I **Dull orange:** Sparse networks II. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

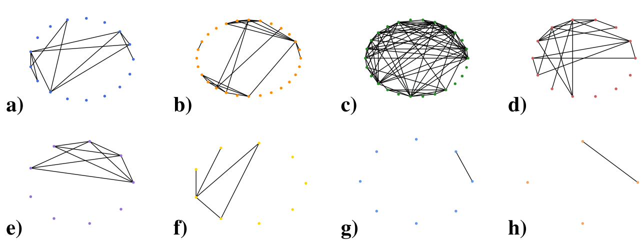

Fig. 5. Individual PNs closest to cluster centroids. **(a)** Composite networks. **(b)** Networks with community structure. **(c)** Well connected networks with subgroup structure. **(d)** Dense BC networks. **(e)** Dense networks. **(f)** Sparse BC networks. **(g)** Sparse networks I **(h)** Sparse networks II.

large and small networks. In the case of large networks, a defining characteristic is the presence of a distinct community or subgroup structure. To understand these two types of large networks, it is crucial to consider the cognitive limitations of individuals (Tamarit et al., 2018). Firstly, if every alter were connected to every other one within a large network, this would result in nodes with exceedingly high degrees. Secondly, maintaining numerous independent relations would be unfeasible for the ego. Consequently, large networks tend to exhibit a certain level of overall connectivity, allowing ego to optimize interactions while avoiding too many separate contexts of interaction. In these networks, individuals interact with some others within the network, but not necessarily with everyone, due to inherent constraints imposed by the network's size. As a result, we identify two types of networks among the larger ones. The first type is characterized by a clear community structure, featuring a limited number of interaction contexts (communities) that are largely independent of each other. However, within each community, strong connections between individuals are prevalent. The second type comprises networks that are well-connected overall, not just within communities. It is important to recognize that identifying this network type can be counter-intuitive, as the subgraph centrality metrics are not widely known or used. In these networks, individuals form dense connections within cliques or subgroups, but there are also numerous connections between these subgroups. Consequently, despite the presence of many communities, the modularity is low. The dimension of subgraph centrality serves as the distinguishing factor between the structure of large dense networks and small dense networks.

Conversely, in the case of small networks, the significance of community or subgraph structure diminishes, and the distinguishing characteristics become more related to local measures. Firstly, there are small dense networks, consisting of a single component where every individual is connected to one another. These networks exhibit high values in measures of cohesion, closure, and degree and closeness centralities. On the other hand, there are networks that display high values in measures of betweenness centrality and average effective size. As we have just seen, these networks are characterized by the presence of highly central nodes, or cohesive cores and satellites. In the intermediate zone between large and small networks, we find medium-sized networks that allow for more diverse structures. These networks can exhibit both high values in metrics of betweenness centrality and cohesion, as well as composite networks with various characteristics blending together.

3.4. Socio-demographic composition of clusters

Understanding the fundamental subspace of structural variability of PNs and how networks are positioned within it has allowed us to classify all PNs in our datasets. Although the purpose of our work is to present the potential of our methodology, we can get useful insights by analyzing the specificities of our data. In this regard, we will briefly evaluate how different socio-demographic indicators are distributed within clusters. To this end, we will assess the composition of the different clusters in terms of datasets, sexes, genders, and countries of residence.

The relevant information is summarized in Table 1. Analyzing sex and gender, focusing on the bottom table, we observe that the distribution of network types across sexes and genders deviates slightly from what would be expected if the taxonomy were independent of sex or gender. Basically, deviations from this random baseline are similar both for sex and gender, with the maximum deviation being 6% if we consider the limits of confidence intervals. Remarkably, there is a noticeable difference between sex and gender: regarding sex, the

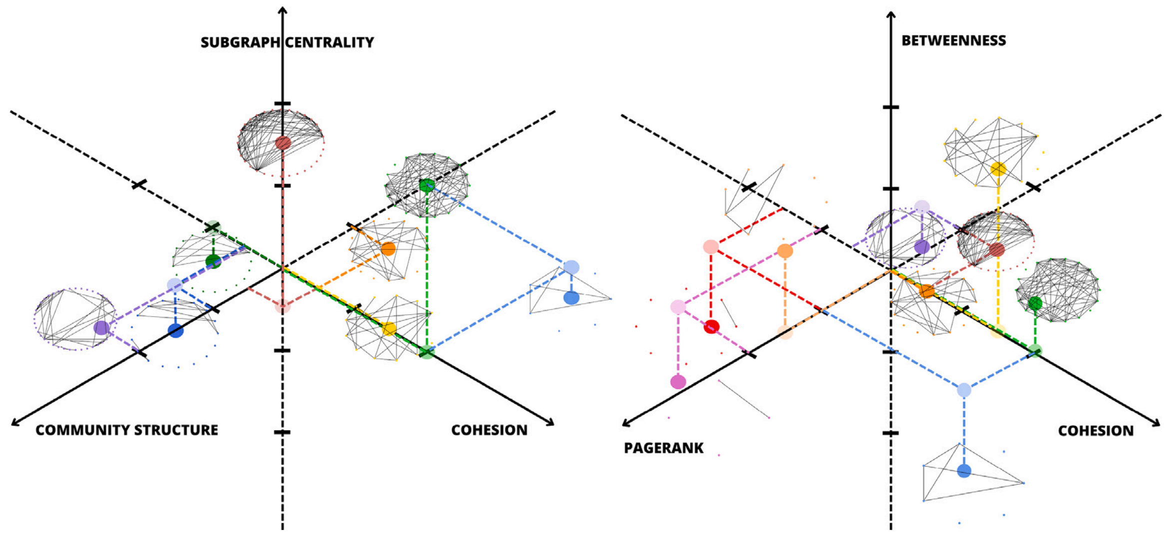

Fig. 6. Diagram illustrating how we can establish a connection between the position of a PN within the fundamental subspace of structural variability of PNs and changes in its appearance due to its structural variation. In this representation, examples of networks are situated within a three-dimensional space based on their metric values associated with cohesion, community structure, and subgraph centrality (left) or their metric values associated with cohesion, Eigenvector-like centralities, and Betweenness Centrality (right). For clarity in interpretation, each network is positioned above a point that shares the same color as its nodes. To aid in visualizing the three dimensions, dashed lines, also color-coded, extend along each dimension, connecting the networks with the corresponding axes. Additionally, for networks not lying on the horizontal plane, we have depicted their "shadows" within the horizontal plane to provide a comprehensive view of their positioning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

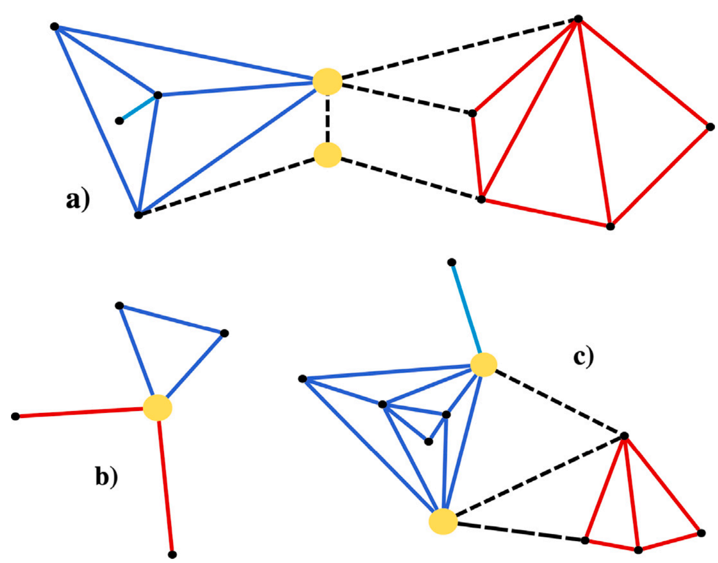

Fig. 7. Three examples of networks that display high values in Betweenness Centrality metrics. Colors are used to separate cohesive subgroups within these networks, connected between them by highly central nodes (yellow nodes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

deviation from this random baseline is not significant for most clusters, since confidence intervals contain the value 0, whereas in the case of gender this deviation is significant, since 0 falls outside the confidence intervals. Nonetheless, this result was expected if we take in to account the fact that the number of networks for which we have information about sex is much smaller and, thus, confidence intervals are wider. In summary, while there are some variations in the prevalence of certain network types among males and females, and among men and women, these deviations are not meaningful. All network types are represented in a reasonably balanced manner across sexes and genders, aligning with the findings of Bidart et al. (2018). Regarding the country of residence, we observe larger deviations, with certain network types being more common for individuals in Spain, while others being more prevalent in El Salvador. This outcome was anticipated, given that Molina et al. (2022) had previously demonstrated that different PNs' structural metrics could predict cultural backgrounds. Nevertheless, it is noteworthy that all network types are present in both countries.

It is also relevant to discuss the proportion of networks coming from the different data sources present in the different clusters. Interestingly, the distribution is not perfectly balanced. Specially, networks belonging to dataset E seem to appear unevenly in the different clusters, deviating as a consequence the proportion of networks from the rest of data sources in all the clusters. This could be the result of the cultural differences discussed in the previous paragraph (dataset E contains networks from people in El Salvador). However, there could be other reasons explaining this unbalance. For instance, as we mentioned, there could be a statistical dependence between the sociocentric networks of the different schools and the individual PNs of students belonging to these schools. In this respect, the complete networks of some schools

Table 1
In both tables, each row corresponds to one of the network types identified in our taxonomy, matching the labels specified in Fig. 5. The columns are categorized into four groups: the first group comprises five columns, each representing one of the datasets under analysis; the second group contains two columns denoting sex; the third group is composed by three columns representing gender, and the last group consists of two columns, one for each country of residence in the dataset. **Top:** The number displayed in each cell represents the proportion of networks of that type within each dataset/sex/gender/country of residence relative to all datasets/sexes/genders/countries of residence. For instance, in the first row, network type (*a*) is distributed among datasets as follows: 0.4% in dataset A, 0.1% in dataset B, 1.8% in dataset C, and so on. The same principle applies to sex, gender and country of residence. This means that in row (*a*), if we sum all five elements corresponding to the five datasets, they add up to 1; similarly, the two values for sex add up to 1, the three values for gender add up to 1, and the two values for country of residence add up to 1. The numbers in parentheses represent the error (95% confidence interval) in the last digit of the proportion displayed. **Bottom:** The value displayed in each cell is the deviation of the proportion shown in the top table from the expected proportion if network types were randomly distributed across datasets/sexes/genders/countries of residence. A value of 0 means that the observed proportion matches what we would expect randomly. Values for sex and gender are computed using only the networks for which we have that information, respectively.

	Data A	Data B	Data C	Data D	Data E	Male	Female	Man	Woman	X	ES	SV
(a)	0.004(3)	0.001(1)	0.018(6)	0.21(2)	0.76(2)	0.4(1)	0.6(1)	0.54(2)	0.44(2)	0.016(5)	0.7(2)	0.3(2)
(b)	0.021(7)	0.011(5)	0.05(1)	0.11(2)	0.81(2)	0.51(9)	0.49(9)	0.52(3)	0.46(3)	0.016(7)	0.71(2)	0.29(2)
(c)	0.09(2)	0.06(1)	0.22(3)	0.34(3)	0.28(3)	0.6(5)	0.4(5)	0.58(4)	0.41(4)	0.012(8)	0.93(2)	0.07(2)
(d)	0.14(2)	0.04(1)	0.17(2)	0.3(3)	0.35(3)	0.5(5)	0.5(5)	0.44(4)	0.55(4)	0.013(9)	0.91(2)	0.09(2)
(e)	0.023(10)	0.05(1)	0.07(2)	0.33(3)	0.52(3)	0.48(9)	0.52(9)	0.42(4)	0.56(4)	0.016(9)	0.83(3)	0.17(3)
(f)	0.03(1)	0.007(7)	0.03(2)	0.19(3)	0.74(4)	0.4(2)	0.6(2)	0.41(4)	0.58(4)	0.01(1)	0.68(4)	0.32(4)
(g)	0.006(7)	0.002(4)	0.008(8)	0.08(3)	0.9(3)	0.4(3)	0.6(3)	0.47(5)	0.49(5)	0.04(2)	0.53(5)	0.47(5)
(h)	0.03(2)	0(0)	0.02(2)	0.17(5)	0.78(6)	0.7(3)	0.3(3)	0.35(7)	0.62(7)	0.02(2)	0.59(7)	0.41(7)
	A	B	C	D	E	Male	Female	Man	Woman	X	ES	SV
(a)	−0.037(7)*	−0.021(5)*	−0.06(1)*	−0.01(3)	0.13(3)*	−0.1(2)	0.1(2)	0.04(3)*	−0.04(3)*	−0.001(8)	−0.06(3)*	0.06(3)*
(b)	−0.02(1)*	−0.011(9)*	−0.03(2)*	−0.11(3)*	0.18(3)*	−0.0(1)	0.0(1)	0.03(4)	−0.03(4)	−0.001(10)	−0.05(3)*	0.05(3)*
(c)	0.05(2)*	0.04(2)*	0.15(3)*	0.11(4)*	−0.35(4)*	0.07(6)*	−0.07(6)*	0.08(5)*	−0.08(5)*	−0.01(1)	0.17(2)*	−0.17(2)*
(d)	0.1(3)	0.02(2)	0.09(3)*	0.08(4)*	−0.29(4)*	−0.03(6)	0.03(6)	−0.06(5)*	0.06(5)*	−0.0(1)	0.15(3)*	−0.15(3)*
(e)	−0.02(1)*	0.03(2)*	−0.01(2)	0.11(4)*	−0.11(4)*	−0.05(10)	0.05(10)	−0.08(5)*	0.08(5)*	−0.0(1)	0.07(3)*	−0.07(3)*
(f)	−0.01(2)	−0.01(1)	−0.04(2)*	−0.03(4)	0.11(5)*	−0.1(2)	0.1(2)	−0.09(5)*	0.09(5)*	−0.0(1)	−0.08(5)*	0.08(5)*
(g)	−0.03(1)*	−0.02(7)	−0.07(1)*	−0.14(3)*	0.27(4)*	−0.2(4)	0.2(4)	−0.02(6)	0.01(6)	0.02(2)	−0.23(5)*	0.23(5)*
(h)	−0.01(3)	−0.022(3)*	−0.06(3)*	−0.06(6)	0.15(7)*	0.1(3)	−0.1(3)	−0.14(8)*	0.14(8)*	0.01(2)	−0.17(8)*	0.17(8)*

* We highlight with those values for which the deviation from 0 is significant (i.e., when 0 falls outside the 95% confidence interval).

may be denser or more clustered, potentially influencing the structural characteristics of the individual PNs. This could create a statistical interdependence between individual observations, potentially unbalancing the distribution between clusters. Further research is needed to address this question, along with other questions like the effect of age, cognitive capacity, prosociality and other traits in the structure of PNs. We leave that assessment for future work.

3.5. Connection to previous literature

Finally, we will connect our results with previous taxonomies from the literature to see that, even taking into account the specificity of our data, our methodology is able to integrate previous perspectives into a comprehensive framework, facilitating the interpretation of each specific result within the larger context. To begin with, Bidart et al. (2018) focused on six structural metrics: Density, Betweenness Centralization, Number of Components, Relative size of the largest component, Modularity of the partition in communities given by the Louvain algorithm, and Diameter. They identified Regular Dense, Centered Dense, Centered Star, Segmented, Pearl Collar, and Dispersed network types. Interestingly, the Pearl Collar Type, which is characterized by the diameter, is not present in our taxonomy. We do not find explicitly a Pear Collar Type because we do not include the diameter as a metric in our classification. The reason behind is that it is not properly defined for fragmented networks. In Bidart et al. (2018), they measure the diameter for the largest connected component, but we argue this is not the proper approach because, in many cases, the largest connected component is not representative of the network, and the diameter could become a deceptive metric. In simple terms, if it is only defined for certain networks and in many cases not for the complete network, it gives a lot of importance to the networks for which it is defined, due to the behavior of the techniques we use. However, this does not mean that our methodology is not able to capture such type of networks. In Bidart et al. (2018), the decision tree of their Fig. 2 shows that Pearl Collar networks are defined by a large betweenness, and large modularity plus large diameter. So, even if we do not use the diameter, it is still possible to identify this type if it is present, by searching for networks with large values in betweenness centrality metrics and large modularity. We stress that what we intend to be integrative is not the typology per se, due to the specificity of our data, but the methodological approach, and our framework is perfectly able to accommodate such a network type if it is present in the dataset. In any case, all the other types align well with our taxonomy. For example, the Regular Dense type, characterized by high and uniform density and low Betweenness Centrality, corresponds to two of our types: well-connected networks with subgroup structure and dense networks. The typology in Bidart et al. (2018) does not distinguish between these two types because they do not utilize subgraph centrality metrics. The Centered Dense and Centered Star types, characterized by high Betweenness Centrality with high and low density, respectively, correspond to our Dense BC and Sparse BC types. The Segmented type, characterized by several large and dense components with some isolates, aligns with our networks with community structure. Lastly, their Dispersed type, characterized by small groups and isolates, correspond to our two Sparse types.

The classification by Vacca (2020) used three variables: Modularity of the partition given by the Girvan–Newman algorithm, number of subgroups with 1 or 2 nodes, and number of subgroups with three or more nodes, they identified a typology ranging from core-centered to factional PNs. In our fundamental subspace of structural variability of PNs, this corresponds to observing networks that vary only along two dimensions: one correlated with metrics of cohesion and the other correlated with metrics related to the community structure of the network. By examining only these two concepts and evaluating the values for these metrics in our developed typology, we can draw connections between our types and theirs. For instance, our Dense networks align with their core-centered type. Then, both Dense BC and well-connected networks with subgroup structure begin to exhibit factional characteristics, but the modularity of these partitions is low due to the high cohesion of these networks. At the opposite end of the spectrum, we find networks with community structure. Additionally, in their typology, they differentiate between bi-factional and multi-factional networks based on the number of communities they identify. We achieve the same result by increasing the resolution of our typology, clustering into a larger number of groups. As we explained, new groups would emerge between existing ones, resulting in a finer distinction in the dimension of community structure, allowing us to differentiate between bi-factional and multi-factional networks.

Finally, considering Maya-Jariego (2021), they utilized Betweenness Centralization, number of cliques, and number of components leading to a somewhat less detailed classification in dense, intermediate, clustered, and fragmented network types. These types have direct correlates within the framework of our taxonomy.

In view of the above, we can conclude that our methodology provides a comprehensive framework that is able to integrate previous typologies into a unified perspective. Interestingly, these prior works can be seen as exploring 'projections' of networks within the fundamental subspace of structural variability we have found. For example, Bidart et al. (2018) examined networks along three dimensions correlating with Betweenness Centrality metrics, cohesion metrics, and certain community structure metrics. Vacca (2020) investigated PNs along dimensions related to cohesion and community structure, while Maya-Jariego (2021) explored PNs along dimensions linked to Betweenness Centrality metrics, cohesion, and community structure. Consequently, their typologies did not directly overlap, and their findings appeared disconnected from one another, as discussed by Vacca (2020), because they were examining different orthogonal projections within the space of structural variability of PNs. In simple terms, our methodology is general not because we find all the types other authors find. Our framework is general precisely because it is able to explain why these typologies did not overlap, and integrates these perspectives to make them comparable. There are some subtleties in this regard. Our approach shows how a typology in two types and a typology in 4 types are perfectly compatible, depending on the granularity of the partition of your space of structural variability. Furthermore, we show how metrics are structured within this space, explaining why their typologies did not overlap: they were looking at regions in this space of structural variability that are orthogonal between them, they were exploring projections. In this respect, our types do not exactly coincide because we used as an example a partition in 8 clusters. But we explained that the selection in the number of cluster serves only sociological reasons, it is not better per se. Thus, if one wants to recover with our methodology results in the previous literature, one simply needs to explore the subspace explored by this previous approach and copy the number of clusters within the same subspace. Furthermore, if then one wants to connect this typology with a completely different typology, one simply needs to go back to the complete space of structural variability, and project the analysis in another dimension explored by a different approach. In this analysis, one could see how two networks are different in one typology and equal in another typology, and understand the difference. One can even compute a 'structural distance' between both networks, and split this distance in different dimensions to understand, for instance, that two different networks in subgraph centrality may be similar in community structure. This is why our work generalizes previous approaches: we have given a geometrical integrative interpretation of how networks vary in structure, what does the creation of a typology mean, and how one can connect and understand different Personal Networks according to their structure. This is translated in the similarity between our types and the types found by previous approaches, even if these types are not completely identical. Therefore, our methodology provides a reference framework capable of integrating all previous perspectives into a unified picture, offering both an explanation and an innovative approach for comprehending and visualizing PNs.

4. Discussion

In this paper we have proposed a rigorous methodology with the purpose of systematizing the construction of a typology of PN structures based on the assessment of their fundamental dimensions of structural variability. Our approach addresses the limitations present in previous attempts of this task, and integrates these previous perspectives in a single framework that allows us to interpret and understand the variation observed in Personal Network structures. We have been able to do this by integrating all well-known fundamental metrics historically used to describe the variability of PN structures, subsequently addressing the level of redundant information they provide. Furthermore, the number of clusters is selected with respect to interpretability, keeping present the lack of an internal clustered structure in the data. Besides, to foster reproducibility, we make a publicly available code that helps to prove the robustness of the method.

We have tested the developed method with a large dataset containing almost 8000 PNs belonging to high school students. Our first key insight is that it is possible to combine the 41 structural metrics selected to characterize the structure of our PNs into 6 basic dimensions, representative of the 6 possible aspects that can meaningfully describe the structural variability of these PNs. These basic aspects are Community Structure, Subgraph Centrality and Cohesive Subgroup Structure, Cohesion, Betweenness Centrality, PageRank and Eigenvector Centralities, and a dimension needed to explain residual variance and odd-behaving variables (e.g. Degree Assortativity). We have observed how these 6 dimensions can help us identify and interpret variations in the structure of PNs in a simple, visual manner, as depicted in Fig. 6.

Secondly, we developed a typology of eight classes for these PNs, based on criteria of consistency and redundancy, bearing in mind that the purpose of this typology is not to uncover a fundamental clustered structure present in the data, but rather to develop a comprehensive, sociologically meaningful classification that allows us to understand the continuous quantitative structural variation present between qualitatively dissimilar networks. All in all, our taxonomy in 8 types can be interpreted as follows: (1) Large Networks with a clear community structure, (2) Large well-connected networks with subgroup structure, (3) Composite networks, (4) Medium-Sized Dense Networks with highly central nodes or cohesive cores, (5) Small Dense networks (6) Small sparse networks with central nodes or cohesive cores, and (7–8) Nearly empty small networks. It is important to take into account that a large group of composite indicators is not as informative as the other groups, but this limitation is not only a limitation in the typology, but a limitation in the data itself. As we see in our results, networks structural variation is continuous, and thus a lot of them occupy intermediate positions. We could arbitrarily force clusters with more extreme values to be larger and reduce this intermediate cluster, but this would not change the fact that these networks occupy intermediate positions in the space and their structure is not crystal-clear understood within general types. In this sense, what is really informative is the basis of dimensions of structural variability, that allows us to interpret in very simple terms the structural features of a network. The structural type will always be a construction to help interpret the network sociologically, and intermediate networks need to be understood by a combination of this basic dimensions. One of the powerful assets of the methodology is that in these cases, if one wants to understand in detail the features of this composite cluster, one simply needs to go to the classification in a larger number of clusters, provided in the Supplementary Material. In this case, previous clusters will be subdivided in smaller more specific clusters, giving details about how this intermediate cluster is structured.

At this point it is important to come back to the 876 PNs eliminated from the analysis due to the fact that they were empty networks or composed of only 1 or 2 nodes. They were excluded before the classification procedure for mathematical reasons, but they should be considered as a 9th type, as they constitute around 9% of the networks present in the original corpus. They could correspond to a population in situation of isolation, dependency, or high risk. This would be a particularly useful type of network for the design of psychosocial interventions or for the implementation of social integration policies targeting the most vulnerable sectors of the population.

Even if the purpose of this work is to present a general methodology and its applicability to some specific data, it is relevant to consider how the conclusions extracted from our results fit in the broader picture. To this end, it is fundamental to keep in mind the specificity of our data, that imposes limitations on the generality of the taxonomy presented. By offering high school students a list of students from the same school, we limit the universe of possibilities. Respondents cannot name family members, neighbors or other persons who are not students at their school. Ideally, it would be better to include all possible types of ties to effectively represent the complete Personal Network of the respondents. Nonetheless, in practice it is very difficult to collect such data systematically and in the proper manner in order for networks to be comparable between them if sample size is large. The restriction to the scholar environment allows to increase exponentially the amount of people we can survey. So, in the end, it is a trade-off: we exchange network specificity by sample size. By increasing the sample size to almost 8000 networks, we have increased statistical power. However, it is important to note that this does not necessarily imply the applicability of our conclusions to the general population, as the size of the sample does not inherently ensure representativeness. For instance, a random sample of 800 individuals, representative of the general population, may offer greater potential for generalizability compared to a sample of 8000 high school students. Further research is needed to assess the generalizability of our results. Bearing these points in mind, it is relevant to point out the resemblance of our results with those from Bidart et al. (2018), Vacca (2020) and Maya-Jariego (2021), even if our Personal Networks contain alters belonging to a single context of interaction. For instance, while Bidart et al. (2018) find networks with a highly central node associated with a 'significant other' they associate with a life-long partner, we also find such a type, suggesting that certain people tend to build strong special relations with 'significant others' since early stages in their life, even if they are not romantic couples. In Bellotti (2008), for instance, they find that friends can occupy the same position for single people. This result points towards a highly intriguing line for research, specifically the discovery of social fingerprints within Personal Networks, characterized by structural patterns that persist regardless of the context. In any case, we stress that the typology might not fully reflect Personal Networks of the respondents due to the specificities of the data, which have nonetheless served well to illustrate the method's potential. In summary, the power of the developed methodology, with its framework that interprets networks as points in a space of fundamental structural variability, is that it can easily accommodate the inclusion of the complete network once the data is available, and understand the results in the same terms.

One could argue that the selection of the 41 structural metrics is arbitrary, since one could always find new metrics to add to the list. Although this list is the result of a careful revision of the past literature, both in network theoretical research and in social networks analysis research, with the purpose of capturing as many dimensions of structural variability of Personal Networks as possible, we could define a new indicator not included in the list. In this regard, we stress that the purpose of this work is to develop a generalizable methodology, able to accommodate this situation, with the aim of making all results understandable within the same framework as spatial projections of the same problem, instead of having a collection of case-studies not comparable between them. The power of the methodology is that it is able to integrate new indicators in case they appear, by simply adding them to the list of variables. If the new indicator provides redundant information, it will be associated with one of the existent dimensions and the typology will not suffer modifications; instead, if the information it conveys is non-redundant, the methodology will be able to assign it a new dimension and produce a new typology in a semi-automatic way. Then, previous typologies will be understood as projections in the new higher-dimensional space, without leaving them useless, enabling us to understand integrate the new knowledge with the old knowledge in the same terms. Indeed, despite the specificity of our data, remarkably our methodology succeeds in integrating previous perspectives in the literature, even allowing us to compare results that were not previously comparable. Specifically, we find that the typologies developed by Maya-Jariego (2021), Bidart et al. (2018) and Vacca (2020) can be interpreted and compared using our method.

All in all, we acknowledge that any approach to the construction of a structural typology is biased, even if it is intended to be general. We believe any proper approach to build an unbiased methodology is to try to keep it unbiased by design. Our typology is not unbiased, but we intend it to be, and the framework established with the methodology illustrates this point: it has significantly reduced the amount of arbitrary decisions and black boxes in previous literature, it is transparent (the code is publicly available to enable other researchers to check for the existence of other black boxes and arbitrary decisions and test the results with their own data without having to replicate the methodology), and it is flexible, in the sense that is direct to build the corrections on top of this public available code.

There are some technical questions we leave open in this article we intend to discuss in future studies. The most important one concerns the projection of weighted-directed networks into unweighted-undirected networks. We are aware that, by doing so, we lose crucial information to understand the structure of PNs. Nonetheless, our central purpose in this work is to be able to connect our results with previous results, and previous results are conducted on unweighted and undirected networks. So, it is necessary as a first step to enable comparison with this previous research. It is easier to understand the structure of Personal Networks and to construct a logical discussion about it if we start from the simplest case, and we understand the unweighted-undirected structure, and then we build on top of it our knowledge about the networks including weights and directionality. With this procedure we build a natural connection with previous studies and then we can make a transition to a more complex understanding of Personal Networks. Furthermore, we also intend to find the connection between the insights obtained for individual PNs and the complete sociocentric school networks in each case, that could give us information about the importance of different socio-demographic indicators associated with each sociocentric network, thus influencing each individual PN. Finally, an interesting line of research would be to explore the fundamental dimensions of structural variability discovered from a theoretical point of view. For instance, we could investigate if it is possible to define some metrics to measure these specific dimensions without going through the dimensionality reduction procedure.

To conclude, the methodology presented and the corresponding developed typology represent a first step in a research agenda driven by three basic questions. First, is the methodology successful when analyzing new PN datasets? In this regard we do provide the necessary tools for replication (see Data availability section). Second, is there a relationship between basic PN structural types and other socio-cultural dimensions of human life like personality traits, social class, gender, life-course of specific minorities? Third, is there a relationship between different cultures and certain structural types? We hope that in future research we can successfully address these important questions.

CRediT authorship contribution statement

Miguel A. González-Casado: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gladis Gonzales:** Writing – final version, Data collection. **José Luis Molina:** Writing – final version, Supervision, Data collection, Investigation, Funding acquisition. **Angel Sánchez:** Writing – final version, Supervision, Data collection, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All necessary data and codes for replication can be found in this GitHub Repository. The code is provided in the form of a Python Class generalized to be used with any dataset of Personal Networks (see the repository for instructions), in order to encourage the adoption of our methodology, support future research and enable researchers to test the universality of the proposed typology.

Acknowledgments

M.A.G.-C. and A. S. acknowledge support from grant PID2022-141802NB-I00 (BASIC) funded by MCIN/AEI/10.13039/501100011033 and by 'ERDF A way of making Europe'. G.G. acknowledges support from Ministerio de Economía y Competitividad (PID2021-126892NB-100) & Agencia Andaluza de Cooperación Internacional para el Desarrollo (AACID-0I008/2020, AACID-2021UI004).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.socnet.2024.03.004.

References

Agneessens, F., Waege, H., Lievens, J., 2006. Diversity in social support by role relations: A typology. Soc. Netw. 28 (4), 427–441.
Ahmed, M., Seraj, R., Islam, S.M.S., 2020. The k-means algorithm: A comprehensive survey and performance evaluation. Electronics 9 (8), 1295.
Ali, T., Elliott, M.R., Antonucci, T.C., Needham, B.L., Zelner, J., Mendes de Leon, C.F., 2022. Multidimensional social network types and their correlates in older Americans. Innov. Aging 6 (1), igab053.
Bartlett, M.S., 1950. Tests of significance in factor analysis. Br. J. Psychol..
Bellotti, E., 2008. What are friends for? Elective communities of single people. Social Networks 30 (4), 318–329.
Bidart, C., Degenne, A., Grossetti, M., 2018. Personal networks typologies: A structural approach. Social Networks 54, 1–11.
Bott, E., 1957. Family and Social Network: Roles, Norms and External Relationships in Ordinary Urban Families. Tavistock Publications Limited.
Brandes, U., Lerner, J., Lubbers, M.J., McCarty, C., Molina, J.L., Nagel, U., 2010. Recognizing modes of acculturation in personal networks of migrants. Procedia-Soc. Behav. Sci. 4, 4–13.
Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. pp. 93–104.
Buja, A., Eyuboglu, N., 1992. Remarks on parallel analysis. Multivariate Behav. Res. 27 (4), 509–540.
Burt, R.S., 1995. Structural Holes: The Social Structure of Competition. Harvard University Press.
Campbell, K.E., Lee, B.A., 1991. Name generators in surveys of personal networks. Soc. Netw. 13 (3), 203–221.
Dobriban, E., 2020. Permutation methods for factor analysis and PCA. Ann. Statist. 48 (5), 2824–2847.
Doeven-Eggens, L., De Fruyt, F., Hendriks, A.J., Bosker, R.J., Van der Werf, M.P., 2008. Personality and personal network type. Pers. Individ. Differ. 45 (7), 689–693.
Eagle, D.E., Proeschold-Bell, R.J., 2015. Methodological considerations in the use of name generators and interpreters. Social Networks 40, 75–83.
Escribano, D., Doldán-Martelli, V., Lapuente, F.J., Cuesta, J.A., Sánchez, A., 2021. Evolution of social relationships between first-year students at middle school: From cliques to circles. Sci. Rep. 11 (1), 11694.
Escribano, D., Lapuente, F.J., Cuesta, J.A., Dunbar, R.I., Sánchez, A., 2023. Stability of the personal relationship networks in a longitudinal study of middle school students. Sci. Rep. 13 (1), 14575.
Estrada, E., 2012. The Structure of Complex Networks: Theory and Applications. Oxford University Press.
Estrada, E., Knight, P.A., 2015. A First Course in Network Theory. Oxford University Press, USA.
Faust, K., Skvoretz, J., 2002. Comparing networks across space and time, size and species. Sociol. Methodol. 32 (1), 267–299.
Fiori, K.L., Antonucci, T.C., Cortina, K.S., 2006. Social network typologies and mental health among older adults. J. Gerontol. Ser. B Psychol. Sci. Soc. Sci. 61 (1), P25–P32.
Freeman, L., 2004. The development of social network analysis. Study Sociol. Sci. 1 (687), 159–167.
Furfey, P.H., Daly, J.F., 1937. A criticism of factor analysis as a technique of social research. Am. Sociol. Rev. 2 (2), 178–186.
Giannella, E., Fischer, C.S., 2016. An inductive typology of egocentric networks. Social Networks 47, 15–23.
Guadalupe, S., Vicente, H.T., 2021. Social network typologies of older people: A cross-national literature review. Ciênc. Saúde Coletiva 26, 5133–5148.
Guadalupe, S., Vicente, H.T., 2022. Types of personal social networks of older adults in Portugal. Soc. Indic. Res. 160 (2–3), 445–466.
Hinton, G.E., Roweis, S., 2002. Stochastic neighbor embedding. Adv. Neural Inf. Process. Syst. 15.
Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185.
Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 374 (2065), 20150202.
Kalish, Y., Robins, G., 2006. Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. Soc. Netw. 28 (1), 56–84.
Laier, B., Hennig, M., Hundsdorfer, S., 2022. An inductive typology of egocentric networks with data from the socio-economic panel. Social Networks 71, 131–142.
Liu, F.T., Ting, K.M., Zhou, Z.-H., 2008. Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. IEEE, pp. 413–422.
Lubbers, M.J., Molina, J.L., McCarty, C., 2007. Personal networks and ethnic identifications: The case of migrants in Spain. Int. Sociol. 22 (6), 721–741.
Marin, A., Hampton, K.N., 2007. Simplifying the personal network name generator: Alternatives to traditional multiple and single name generators. Field Methods 19 (2), 163–193.
Maya-Jariego, I., 2021. Building a structural typology of personal networks: Individual differences in the cohesion of interpersonal environment. Social Networks 64, 173–180.
Maya-Jariego, I., González-Tinoco, E., 2023. Use of a hierarchical deconstruction procedure for the classification of personal networks: Exploring nested groups around you. Social Networks 73, 20–29.
Maya-Jariego, I., Holgado, D., 2015. Living in the metropolitan area. Correlation of interurban mobility with the structural cohesion of personal networks and the originative sense of community. Psychosoc. Intervent. 24 (3), 185–190.
Maya-Jariego, I., Letina, S., Tinoco, E.G., 2020. Personal networks and psychological attributes: exploring individual differences in personality and sense of community and their relationship to the structure of personal networks. Netw. Sci. 8 (2), 168–188.
McCarty, C., 2002. Structure in personal networks. J. Soc. Struct. 3 (1), 20.
Mitchell, J.C., 1969. Social Networks in Urban Situations: Analyses of Personal Relationships in Central African Towns. Manchester University Press.
Molina, J.L., Ozaita, J., Tamarit, I., Sánchez, A., McCarty, C., Bernard, H.R., 2022. Structural measures of personal networks predict migrants' cultural backgrounds: an explanation from grid/group theory. PNAS Nexus 1 (4), pgac195.
Mulaik, S.A., 2009. Foundations of Factor Analysis. CRC Press.
Newman, M., 2018. Networks. Oxford University Press.
Pelle, E., Pappadà, R., 2021. A clustering procedure for mixed-type data to explore ego network typologies: an application to elderly people living alone in Italy. Stat. Methods Appl. 30, 1507–1533.
Pescosolido, B.A., Rubin, B.A., 2000. The web of group affiliations revisited: Social life, postmodernism, and sociology. Am. Sociol. Rev. 52–76.
Portes, A., Vickstrom, E., 2011. Diversity, social capital, and cohesion. Annu. Rev. Sociol. 37, 461–479.
Rainie, H., Wellman, B., 2012. Networked: The New Social Operating System, vol. 10, MIT Press, Cambridge, MA.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Schölkopf, B., Smola, A., Müller, K.-R., 1997. Kernel principal component analysis. In: International Conference on Artificial Neural Networks. Springer, pp. 583–588.
Schölkopf, B., Williamson, R.C., Smola, A., Shawe-Taylor, J., Platt, J., 1999. Support vector method for novelty detection. Adv. Neural Inf. Process. Syst. 12.
Tamarit, I., Cuesta, J.A., Dunbar, R.I., Sánchez, A., 2018. Cognitive resource allocation determines the organization of personal networks. Proc. Natl. Acad. Sci. 115 (33), 8316–8321.
Vacca, R., 2020. Structure in personal networks: Constructing and comparing typologies. Netw. Sci. 8 (2), 142–167.
Vinh, N.X., Epps, J., Bailey, J., 2009. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 1073–1080.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.
Zwick, W.R., Velicer, W.F., 1986. Comparison of five rules for determining the number of components to retain. Psychol. Bull. 99 (3), 432.