Socio-Economic Planning Sciences 105 (2026) 102496

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

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

Behind the screen: A comprehensive framework for digital work metrics and data integration[☆,☆☆]

F. Demaria [a,1,*], A. Papana Dagiasis [b,1], M. Cavicchioli [c,1], T. Fabbri [d,1]

[a] Department of Communication and Economics, University of Modena and Reggio Emilia, Viale A. Allegri 9, Reggio Emilia, 42121, Italy
[b] Department of Mathematics and Statistics, Cleveland State University, 2121 Euclid Avenue, RT 1515, Cleveland, 44115, OH, United States
[c] Department of Economics Marco Biagi, University of Modena and Reggio Emilia and ReCent, Via J. Berengario 51, Modena, 41121, Italy
[d] Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Via J. Berengario 51, Modena, 41121, Italy

ARTICLE INFO

Keywords:
Digital work metrics
Digitalization
Random Forest
sidClustering
Visual tools
Social sustainability

ABSTRACT

The digital transformation and the subsequent datafication of work generate rich electronic traces that can be transformed into actionable insights through modern analytics. In this context, this study proposes a data-driven framework that leverages *sidClustering* and unsupervised Random Forest (RF) for clustering and feature selection, constructs composite indicators via multiple aggregation strategies, and employs visual tools to enhance the interpretability of results. A key feature of the framework is the integration of two data sources: click metadata from Microsoft 365 and employee attitudes measured through a questionnaire. This integration enables the analysis of digital work behavior (DWB) in relation to employee sentiment. We apply the framework to a highly digitalized Italian consulting company. The analysis identifies two employee clusters, 'Operational' and 'Coordination', and yields two synthetic digital work metrics, work *Quantity* and *Complexity*. Overall, the study introduces a scalable methodological framework that combines tree-based learning, composite indicators, and visual tools, representing one of the first empirical integrations of digital work metadata with employee attitudes. The resulting indicators offer early-warning capabilities for assessing the impact of technology adoption on employee outcomes and provide decision support for HR analytics and policy.

1. Introduction

Over the past decade, we have witnessed a significant integration of digital technologies into work processes and workplace relationships. The digital revolution compels organizations to transform their operations, shifting from labor-intensive models to more technology-driven approaches to work organization [2,3]. In this sense, the digital transformation of work translates into an increasing number of actions and decisions at every logical or empirical level – including decision-making, coordination, control, and execution – that are performed digitally, i.e., using digital information, within digital workflows, hosted on corporate digital premises, accessible at any time from anywhere via digital and mobile devices [4]. A central outcome of this transformation is the emergence of the digital workplace [5], broadly defined as the set of connected technologies employees use daily to perform their jobs, including intranets, communication platforms, and business systems [6,7]. Beyond new tools, this transformation has also led to the datafication of work, i.e., the continuous creation of digital records that capture employee activities in digital workspaces, resulting in time-stamped logs as indirect products of worker behaviors. These metadata enhance the behavioral visibility of work activities, which can be conceptualized as the "sociomaterial enactment of behaviors" observable by others [1]. As by-products of workers' behaviors, passively recorded and stored, these metadata are literally "exhausts" and can be accessed and analyzed to generate data representations of work behaviors and inform HR theory [8]. Clearly, the volume of data generated is directly proportional to an organization's level of digitalization, with more digitally advanced organizations producing big data [9]. Furthermore, the recent shift towards extensive remote working, accelerated by the Covid-19 pandemic, has likely increased

☆ This article is part of a Special issue entitled: 'Statistical Measurement of Economic Analysis' published in Socio-Economic Planning Sciences.
☆☆ A preliminary version of this article was delivered at the 3rd Italian Conference on Economic Statistics, held in Naples, February 13–14, 2025. The authors gratefully acknowledge the participants for their valuable comments and discussion.

* Corresponding author.
E-mail address: fabio.demaria@unimore.it (F. Demaria).

[1] All authors contributed to the writing of the manuscript.
[2] According to Leonardi and Treem [1], not only are sound quantitative representations necessary, but their sharing within the organization and a certain degree of agreement upon their meaningfulness are also required for work to become truly 'visible', an aspect not addressed here.

https://doi.org/10.1016/j.seps.2026.102496
Received 18 September 2025; Received in revised form 5 March 2026; Accepted 12 April 2026
Available online 24 April 2026

the coverage of digital exhaust, capturing a broader spectrum, if not the entirety, of work behaviors [10]. There are two main implications of digital transformation for organizations. First, the growing volumes of digital data require advances in learning algorithms to provide organizations with new analytical capabilities to process and analyze such data and, when possible, to facilitate the automatic detection of patterns within both structured and unstructured data [11]. Second, work has become potentially observable with unprecedented timeliness and detail. However, potential observability does not equate to actual visibility: in order to generate a meaningful representation of work or some of its dimensions, clicks need to be extracted, selected, combined, and analyzed.[2] Therefore, data analytics is needed to craft meaningful representations of work to be used for updating managerial practices to digital work, in the never-ending joint quest for organizational performance and workers' well-being. Moreover, being able to exploit this new availability of data may enable organizations to pursue social sustainability goals (e.g., Sustainable Development Goal 8 – *"Decent work and economic growth"*) [12,13], monitor employees, enhance organizational performance, and derive policy implications. While existing studies have primarily focused on isolated technologies or specific outcomes in the digital workplace, a comprehensive understanding of employees' holistic experience of digital work, spanning everyday practices such as logging in, emailing, messaging, and navigating multiple applications, and its relationship to the diverse negative psychological effects that may arise remains underexplored [7].

In view of this gap, our work proposes a data-driven methodology based on unsupervised learning and visual tools to analyze digital work behavior (DWB) – defined as work performed within digital environments and captured through quantitative metrics (i.e., metadata) – and to examine how it can be linked to employee attitudes measured through a survey, thereby combining different data sources in line with the approach proposed by Thanos et al. [14]. We apply this methodological framework to an Italian digital workplace that forms part of a research project and has provided primary data. Data-driven methods aim to extract information and discover patterns, in accordance with the concept of data-driven science, where data inductively inform hypotheses and theories [15]. This is exemplified by Exploratory Data Analysis (EDA), which seeks diverse insights from data [16]. In this context, Thanos et al. [14] highlight the value of integrating multiple data types to reveal unique information and create knowledge. Furthermore, the methodology of data-driven science is advanced by combining inductive approaches with deductive techniques, giving rise to a new category of big data experiments designed to foster knowledge generation [14,17].

The contribution of this work is threefold. First, it introduces a methodological framework comprising unsupervised tree-based methods, composite indicators, and visual tools to facilitate interpretation and knowledge extraction [18–21]. The proposed methodology is grounded in cutting-edge statistical learning approaches and enables the analysis of large-scale data to uncover actionable insights through metrics and visualizations that extend beyond traditional ones. Furthermore, the framework allows researchers to integrate data from different sources, potentially generating novel patterns and yielding new insights [14]. The second contribution is the introduction of new digital work synthetic indicators. These novel composite indicators serve as quantitative descriptors of work in the digital era and may function as early-warning indicators of the impact of technology adoption on worker outcomes, with potential implications for organizational policy and decision-making. Third, this study is among the first to combine two distinct data sources from a digital workplace. In doing so, it contributes to the literature on digital work and organizational theory for highly digitalized companies and provides insights into the observability of employees' work. In fact, despite the potential benefits, privacy concerns surrounding employee data and the need for a multidisciplinary approach that includes organizational, computational, and statistical modeling expertise have so far limited empirical research exploiting this unprecedented data availability [22–24].

The remainder of the paper is structured as follows. Section 2 presents the framework and outlines the methods employed for data analysis. In particular, we describe the statistical framework, which comprises unsupervised RF, *sidClustering*, composite indicators, and visual tools. Section 3 illustrates the application of the proposed methodology in the context of an Italian digital workplace. Section 4 presents the results, discusses conclusions and implications, and highlights limitations and future directions.

2. Methodological framework

In this study, to address the need to shed light on employees' holistic experience of digital work, we propose a data-driven statistical framework based on state-of-the-art techniques for analyzing two distinct data sources: digital work (i.e., click metadata) and employee attitudes (i.e., survey data). Specifically, with reference to digital work metadata, we leverage Random Forest (RF)-based methods to (i) cluster units, (ii) identify the most important digital work features, (iii) construct new synthetic metrics of digital work, and (iv) examine the interaction between these metrics and employee attitudes. Finally, we graphically represent this interaction using advanced visual tools to investigate whether specific aggregations of digital behavioral patterns are associated with particular employee attitudes. A flowchart of the proposed framework is presented in Fig. 1. In the following, we detail each technique that forms part of the methodological framework.

2.1. Random forest

Machine learning can be broadly categorized into supervised and unsupervised learning. Supervised learning involves models that predict known responses, while unsupervised learning discovers structures from features without a target variable [25]. In the unsupervised setting, clustering algorithms are prominent [26–29], yet they often lack feature selection, crucial in high-dimensional spaces, and struggle with mixed numerical and categorical data. RF offers feature selection, handles mixed data types, and scales well with large datasets, proving robust against outliers [20]. The RF algorithm is a tree-based ensemble learning technique that enhances prediction accuracy by averaging outputs from multiple decision trees [30]. It employs bootstrap aggregating, or bagging, to construct individual trees using bootstrap samples from the training dataset, thereby mitigating overfitting issues [31]. Developed by Breiman [30], the genesis of RF's methodology can be traced back to his bagging approach (1996), the feature selection techniques of Amit and Geman [33] and Ho [34], and Dietterich's random split selection approach (2000). RF, along with other machine learning algorithms, can model nonlinear relationships between predictors and outcomes effectively [36]. Moreover, RF is particularly advantageous for handling high-dimensional data in contexts with small sample sizes [37,38]. By combining bagging with random feature selection, RF intrinsically mitigates the risk of overfitting that typically affects standard models under such conditions, yielding robust and competitive performance even with limited data [38–40]. As a nonparametric method, RF is widely applicable across various fields for its superior performance in classification and prediction tasks [41,42].

Unsupervised RF. Many properties of RF can be leveraged in the unsupervised context, where they are traditionally used for clustering observations [43,44]. Unsupervised RF often outperforms conventional clustering algorithms as it not only produces a distance matrix that extends beyond traditional measures but also enables feature selection and uncovers nonlinear interactions in a fully nonparametric way [20,45]. It can be implemented by transforming unsupervised tasks into supervised ones, creating an artificial dataset to build the forest (i.e., pseudo-outcomes) using the multivariate splitting rule, and extracting the

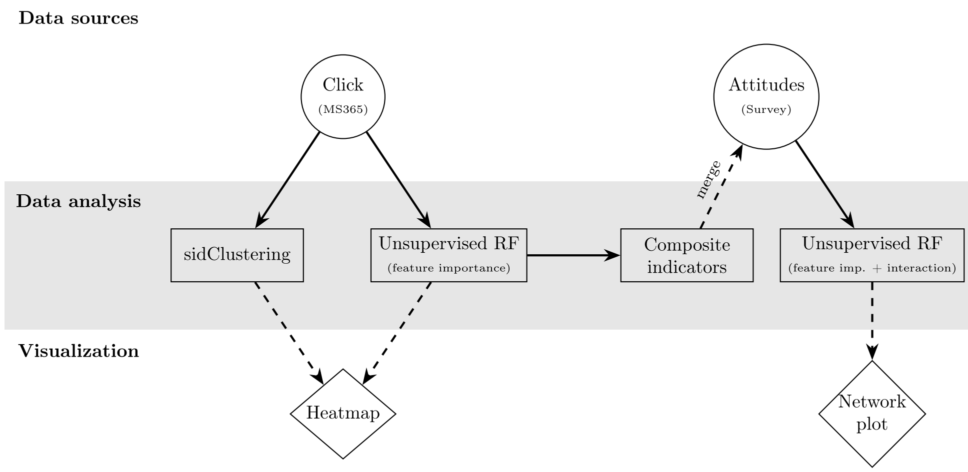

Fig. 1. Methodological framework flowchart.

resulting proximity matrix [46]. The multivariate splitting rule combines contributions from continuous and categorical outcomes into a single composite measure. Specifically, the composite standardized split rule is defined as:

$$\Theta(s,t) = E_q^*(s,t) + G_r^*(s,t) \tag{1}$$

where s denotes a candidate split variable and t denotes a potential split node. The term $E_q^*(s,t)$ is the standardized mean-squared error statistic aggregated across the q continuous outcome dimensions, and $G_r^*(s,t)$ is the standardized Gini impurity decrease statistic averaged over the r categorical outcome dimensions. Splits are chosen by maximizing this composite criterion $\Theta(s,t)$, providing an efficient and unified way to handle mixed-type multivariate outcomes [46].

2.1.1. *sidClustering*

Building on Breiman's RF method (2001), Mantero and Ishwaran [20] presented an innovative clustering algorithm, *sidClustering*. This method creates an artificial dataset through 'sidification' of raw data and utilizes multivariate RF [46] to calculate distances in the newly formed data space, resulting in a tree-based distance matrix. For a comprehensive understanding of *sidClustering*, we refer to Mantero & Ishwaran (2021). Let $\mathbf{X} = (X_1, \ldots, X_p)^T$ be the p-dimensional features and $\mathcal{L} = (\mathbf{X}_i)_{i=1}^n$ be the learning data (i.e., the set of features). The procedure starts with the 'sidification' of the original variables to obtain the sidified data $\mathcal{L}^s = (\mathbf{Z}_i, \mathbf{Y}_i)_{i=1}^n$, thereby creating the 'enhanced feature space' (i.e., artificial data). Specifically, let X_j and X_k be coordinates of $\mathbf{X}$, the SID main effects $\mathbf{Y} = (Y_1, \ldots, Y_p)^T$, derived from X_j and X_k, are coordinates Y_j and Y_k defined by:

$$\begin{aligned} Y_j &= \delta_j + X_j \\ Y_k &= \delta_k + X_k \end{aligned} \tag{2}$$

where $\delta_j, \delta_k > 0$ so that Y_j, Y_k are positive and distinct. This process is referred to as 'staggering'. Let $\mathbf{Z}$ be the SID interaction features, procured by forming all pairwise interactions of the SID main effects. The SID interaction corresponding to features X_j and X_k is a coordinate of $\mathbf{Z}$ denoted by $X_j \star X_k$ and is defined as the product of Y_j and Y_k:

$$X_j \star X_k = Y_j \times Y_k = (\delta_j + X_j)(\delta_k + X_k) \tag{3}$$

The SID interaction features $\mathbf{Z}$ are utilized in the multivariate regression to predict the main features $\mathbf{Y}$. Upon data sidification, a multivariate RF is fitted using the SID interaction features $\{\mathbf{Z}_i\}_{i=1}^n$ to forecast the SID main features $\{\mathbf{Y}_i\}_{i=1}^n$. Subsequently, the RF distance matrix is extracted from the multivariate forest, and the distance between observations is computed. The *sidClustering* algorithm introduces a novel forest distance, diverging from the conventional proximity measure used by RF. Unlike proximity, which relies on terminal node membership to assess the similarity between observations, the new distance employs a metric based on the tree's topology to provide a more nuanced measure of dissimilarity. Let T_b denote the bth tree in a forest. The forest distance is applied to SID interaction features $\mathbf{Z}$. For each pair of observed data points $\mathbf{Z}_i$ and $\mathbf{Z}_l$, we define $S(\mathbf{Z}_i, \mathbf{Z}_l, T_b)$ to be the minimum number of splits on the path from the terminal node containing $\mathbf{Z}_i$ to the terminal node containing $\mathbf{Z}_l$ in T_b, such that the path includes at least one common ancestor node of $\mathbf{Z}_i$ and $\mathbf{Z}_l$. Similarly, we define $R(\mathbf{Z}_i, \mathbf{Z}_l, T_b)$ as the minimum number of splits on the path from $\mathbf{Z}_i$ to $\mathbf{Z}_l$ that passes through the root in T_b. The forest distance between $\mathbf{Z}_i$ and $\mathbf{Z}_l$ is defined as:

$$D(\mathbf{Z}_i, \mathbf{Z}_l) = \frac{1}{B} \sum_{b=1}^{B} \frac{S(\mathbf{Z}_i, \mathbf{Z}_l, T_b)}{R(\mathbf{Z}_i, \mathbf{Z}_l, T_b)} \tag{4}$$

where B is the total number of trees in the forest. Consequently, when two observations fall into the same terminal node, $D(\mathbf{Z}_i, \mathbf{Z}_l) = 0$. Finally, observations are clustered based on the resulting distance matrix using a clustering algorithm such as k-means [26,29], hierarchical clustering [27], or partitioning around medoids (PAM) [28]. In our study, we employ k-means clustering, which partitions the data into k clusters and identifies centroids representing the centers of these clusters. The process involves specifying k, selecting initial centroids, assigning observations to the nearest centroid, recalculating centroids by averaging points within each cluster, and iterating until centroids stabilize and the variation within clusters is minimized.

2.1.2. *Feature importance*

Feature selection is an important task in many data analytics problems, especially when dealing with a large number of features. In RF-based models, feature importance provides information about the predictive ability of each feature and is often assessed using traditional methods like mean decrease impurity and permutation importance, the latter also known as Breiman-Cutler importance [30,47]. Permutation importance is calculated by the increase in prediction error after permuting a variable's out-of-bag (OOB) data [30]. Recent methods for estimating variable importance, such as minimal depth, offer a more direct assessment of a variable's predictive power and enable features selection in the unsupervised setting. The minimal depth method measures the distance from a variable to the root of the tree, with a smaller minimal depth indicating a stronger predictive variable [48]. It is worth noting that traditional feature selection metrics, such as permutation importance, can be strongly penalized by dependence among covariates, as correlated variables can mask each other's permutation effect [49]. Minimal depth overcomes this issue because it is rooted in fundamental tree-based principles and relies strictly on the topological structure of the tree rather than on the prediction error. As a result, this method is advantageous not only because it provides a direct and robust measure of a variable's predictive ability, even in the presence of highly correlated covariates, but also because it is applicable to various forest types [48,50]. Such a method is formulated in terms of maximal subtree, which is computed considering the topology of a tree. Specifically, the first-order maximal subtree for a variable v is defined as the largest subtree where v is used for the initial split at the root. The depth d of a node is its distance to the root, with $d \in \{0, 1, \dots, D(T)\}$, where $D(T)$ represents the tree's maximum depth. The minimal depth D_v of v is the distance from the tree root to the root of the nearest maximal subtree where v first splits the tree, serving as an indicator of v's predictive power. A graphical representation can be found in [18]. Ishwaran et al. [48] propose using the mean of the minimal depth distribution as a threshold to identify relevant features. Variables above this threshold are considered noisy. The minimal depth distribution for a weak variable v is given by:

$$P\{D_v = D(T) | v \text{ is weak}\} = 1 - \sum_{d=0}^{D(T)-1} \left[\mathbb{P}(D_v = d | v \text{ is weak})\right] \tag{5}$$

where

$$\mathbb{P}(D_v = d | v \text{ is weak}) = \left[1 - \left(1 - \frac{1}{p}\right)^{\ell_d}\right] \prod_{j=0}^{d-1} \left(1 - \frac{1}{p}\right)^{\ell_j}, \tag{6}$$

ℓ_d is the number of nodes at depth d, and p is the total number of variables. The mean of this distribution serves as a threshold for selecting important variables. Upon the selection of the most relevant features, they can be used to construct a composite indicator, a valuable tool for investigating and summarizing complex phenomena that are not directly measurable [51]. The selected variables can be weighted according to their significance within the unsupervised RF model, using weights proportional to their minimal depth. Different aggregation procedures can then be applied to construct composite indicators, as described in the dedicated section.

2.1.3. Feature interaction

To address the need to examine variable interdependencies, tree-based methods such as RF offer the capability to discover potential interactions between variables [20,48]. In supervised settings, the H-statistic, developed by Friedman and Popescu (2008), uses partial dependence properties to measure interactions. For function $f(\mathbf{x})$, an interaction exists between variables x_j and x_k if changes in x_j alter $f(\mathbf{x})$ differently depending on x_k's value. In unsupervised settings, interactions are discerned through tree structures [48], with the minimal depth concept extending to second-order maximal subtrees [52] to analyze variable relationships. Let v and w be two variables used for splitting along the same branch of a tree. For variable v, a second-order maximal subtree $T_{v,w}$ within a maximal v-subtree T_v indicates an interaction: the closer w appears to the root of v in $T_{v,w}$, the stronger their association. To quantify this, the minimal depth q of $T_{v,w}$ is normalized by the depth of the corresponding subtree $D_{v,w}$ to yield a score in the interval $[0, 1]$, which is then averaged across the entire forest. Smaller scores signify stronger interactions.

In unsupervised RF, using maximal subtrees and minimal depth, a $p \times p$ interaction matrix ($\boldsymbol{\Gamma}$) is constructed, where p is the variable count. The main diagonal (Γ_{jj}) stores the importance score of each variable, while off-diagonal entries Γ_{jk} represent normalized minimal depths of variable k relative to the maximal subtree for variable j. The resulting interaction matrix can be used to visualize patterns in the data structure in terms of both variable importance and variable interactions.

2.2. Composite indicators

Composite indicators, also referred to as synthetic indices or performance indices, are constructed by aggregating individual metrics into a single index that captures a multidimensional concept [53]. These indices are particularly valuable for quantifying complex, multidimensional phenomena, such as competitiveness, well-being, or sustainability, which cannot be captured adequately by a single metric [54–56]. Over the past two decades, composite indicators have gained significant attention across various policy domains including economics, environment, and social sciences, serving purposes ranging from benchmarking and ranking to policy monitoring and public communication [51,57]. Their appeal lies in their capacity to condense large datasets into a single figure, facilitating interpretation, comparison, and decision-making processes for policymakers and stakeholders [56,58]. Composite indicators are frequently developed through multivariate statistical techniques, such as Principal Component Analysis (PCA), to reduce dimensionality and derive meaningful latent structures that underpin the indicator framework [59–61]. Nonetheless, the construction of a composite index entails a series of theoretical and methodological assumptions that must be carefully considered to ensure analytical robustness and avoid misleading interpretations [53].

2.2.1. Data-driven weighting

One of the most critical aspects in the development of composite indicators is the weighting scheme adopted for the sub-indicators, as this directly affects the reliability and interpretability of the resulting index [62,63]. While traditional approaches often involve subjective weight assignments based on expert judgment or stakeholder consultation, data-driven weighting techniques have emerged as a more 'objective' alternative, allowing the data itself to determine the relative importance of each indicator [54,57]. Data-driven methods avoid explicit value judgments and instead rely purely on the statistical properties of the data matrix $\mathbf{X}$. Techniques such as correlation or regression analysis, PCA, Factor Analysis (FA), and Data Envelopment Analysis (DEA) are employed to derive weights based on the internal structure of the data [64,65]. The perceived objectivity of these techniques has made them particularly appealing in the context of social and economic measurement, where concerns about arbitrariness and manipulation are prevalent. Nevertheless, data-driven methods are not without limitations. For example, weights derived from PCA depend on the correlation structure of the data, which may not reflect the true relevance of indicators to the underlying construct [53]. Moreover, these techniques often lack transparency, complicating their communication and interpretability for non-technical stakeholders [54]. A more recent advancement in this domain involves the use of RF to derive weights based on variable importance measures. Unlike linear regression, RF does not assume linear relationships and can capture complex, non-linear interactions. Furthermore, they avoid the constraint of relying solely on correlation, a limitation commonly observed in traditional statistical methods [62].

2.2.2. Aggregation strategies

Once weights are defined, the final stage in composite indicator construction involves aggregation. Aggregation rules can be broadly classified into three categories: linear, geometric, and multi-criteria approaches [53]. Another widely used classification distinguishes between compensatory, non-compensatory, and mixed strategies, each with specific methodological implications [54,58]. To assess whether our results are robust to alternative aggregation assumptions, we compare a compensatory linear approach, a non-compensatory outranking method (i.e., ELECTRE III), and the mixed Mazziotta–Pareto Index (MPI) [62,66]. Each method entails clear strengths and limitations; therefore, using them in combination helps balance their respective trade-offs and reduces reliance on any single set of aggregation assumptions. Specifically, linear aggregation offers high transparency and easy decomposition, but its fully compensatory nature allows strong performances to mask critical deficits [58]. Non-compensatory outranking methods such as ELECTRE III prevent such full compensation and can incorporate meaningful preference thresholds, but they require additional modeling choices (e.g., thresholds, preference functions, veto/discordance settings) and are less intuitive for non-technical audiences [67,68]. Finally, a mixed strategy such as the Mazziotta–Pareto Index can be attractive because it retains the simplicity of additive indices while penalizing unbalanced performances across dimensions, thereby reducing excessive compensability, though it remains sensitive to normalization and penalty specifications [69]. Taken together, these methods enable triangulation across compensatory and non-compensatory models, thereby strengthening robustness checks, improving diagnostic interpretation of divergences, and supporting more defensible policy prioritization under uncertainty. Below, we provide further details on these methods.

Compensatory strategies. In compensatory schemes, such as linear or geometric aggregation, a deficit in one dimension can be offset by a surplus in another [53]. Weighted linear aggregation, the most common method, assumes full compensability, where the trade-off rate between indicators equals the ratio of their weights [58]. Given a set of standardized variables z_j $(j = 1, \ldots, J)$, with corresponding weights w_j, the general form of the composite indicator for unit i can be defined as:

$$\mathrm{CI}_i = \sum_{j=1}^{J} z_{ij} \cdot w_j \tag{7}$$

where J is the number of variables to be aggregated in the composite indicator. Geometric aggregation reduces compensability by penalizing imbalances more severely, representing a step towards addressing concerns about indicator substitution [57].

Non-compensatory strategies. Non-compensatory methods reject the notion of trade-offs among dimensions and are therefore more appropriate when indicators are considered essential and non-substitutable. These methods include ELECTRE [68,70] and PROMETHEE [67,71], which use pairwise comparisons and outranking procedures to establish preference relations among alternatives. Weights in this context are interpreted as importance coefficients, and aggregation involves building outranking matrices and deriving partial or complete rankings [72]. The ELECTRE III method is part of the ELECTRE family methods and it is designed to produce a partial or complete ranking of a set of alternatives $A = \{A_1, A_2, \ldots, A_n\}$ evaluated on multiple criteria $G = \{g_1, g_2, \ldots, g_p\}$ [68,70]. For each criterion g_j, ELECTRE III incorporates three discrimination thresholds: an indifference threshold q_j, a preference threshold s_j, and a veto threshold v_j, allowing for the modeling of imperfect knowledge and small or unacceptable differences in performance. For each pair of alternatives (A_i, A_l), the method computes a partial concordance index $c_j(A_i, A_l) \in [0, 1]$ on each criterion, which is then aggregated into a global concordance index $C(A_i, A_l)$ representing the degree of support for the statement "A_i is at least as good as A_l". Simultaneously, discordance indices $d_j(A_i, A_l) \in [0, 1]$ are used to assess whether large disadvantages on any criterion contradict this outranking. The final strength of the outranking relation is expressed by the credibility index $\sigma(A_i, A_l) \in [0, 1]$, which combines concordance and discordance and quantifies the degree to which alternative A_i is at least as good as A_l. Once $\sigma(A_i, A_l)$ is computed for all pairs of alternatives, we obtain a credibility matrix, which will be used to construct ascending and descending preorders using a two-phase distillation procedure: the *descending* distillation, which iteratively identifies the best-performing alternatives, and the *ascending* distillation, which iteratively identifies the worst-performing ones. Typically, the final ranking is obtained by intersecting the two preorders resulting from the distillation processes. For a comprehensive review of ELECTRE methods, we refer to [68].

Mixed strategies. To address the limitations of full compensability without adopting the complexity of non-compensatory techniques, hybrid methods such as the Mazziotta–Pareto Index (MPI) have been proposed [73]. MPI modifies the arithmetic mean by introducing a penalty component that reflects the internal variability among indicators for each unit. This penalty – computed as the product of the standard deviation and coefficient of variation – penalizes unbalanced profiles and favors consistent performances across dimensions [74]. While the use of an arithmetic mean implies residual compensability, the adjustment introduced by the MPI ensures that units with erratic performances do not score disproportionately well, thereby balancing interpretability with conceptual rigor. To compute the MPI, the first step concerns data normalization. In particular, given a data matrix ($\mathbf{X}$), raw data must be normalized as follows:

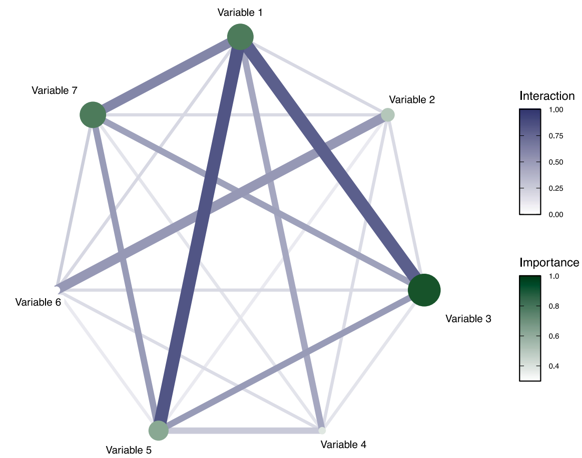

Fig. 2. Network plot.

$$z_{ij} = 100 \pm \frac{(x_{ij} - \mathrm{M}_{x_j})}{\mathrm{S}_{x_j}} \cdot 10 \tag{8}$$

where M_{x_j} and S_{x_j} are the mean and the standard deviation of indicator j, respectively.[3] This transformation ensures that all indicators are made dimensionless and comparable by bringing them onto the same scale and centered at 100, with standard deviation of approximately 10. Once the standardized matrix $\mathbf{Z} = \{z_{ij}\}$ is obtained, the aggregation phase involves computing the mean, the standard deviation, and the coefficient of variation of each unit i across all J indicators:

$$\mathrm{M}_{z_i} = \frac{\sum_{j=1}^{J} z_{ij}}{J}, \qquad \mathrm{S}_{z_i} = \sqrt{\frac{\sum_{j=1}^{J} (z_{ij} - \mathrm{M}_{z_i})^2}{J}}, \qquad \mathrm{CV}_i = \frac{\mathrm{S}_{z_i}}{\mathrm{M}_{z_i}} \tag{9}$$

Finally, the MPI is expressed as:

$$\mathrm{MPI}_i^{+/-} = \mathrm{M}_{z_i} \pm \mathrm{S}_{z_i} \cdot \mathrm{CV}_i \tag{10}$$

where the sign $\pm$ reflects whether the indicators are intended to be maximized or minimized.

2.3. Visual tools

In our research, to complement the numerical analysis, we generate visual representations to elucidate the data's inherent structure, quality, and multi-dimensional relationships [75]. In particular, we leverage graphical tools such as heatmaps and network plots to illustrate the relationships among variables and to underscore the patterns present within the data structure. An illustrative example of a network plot is depicted in Fig. 2.

This network plot enables the visualization of two distinct types of information: the importance of features and their interactions with other features. More important and intensely colored (green) nodes signify a higher importance of features. The interactions between pairs of features are represented by (violet) connecting lines, where the thickness and color intensity of the lines denote stronger interaction values. To improve the clarity of our visual tools, we employ dendrogram

[3] The sign $\pm$ depends on the nature of the indicator.

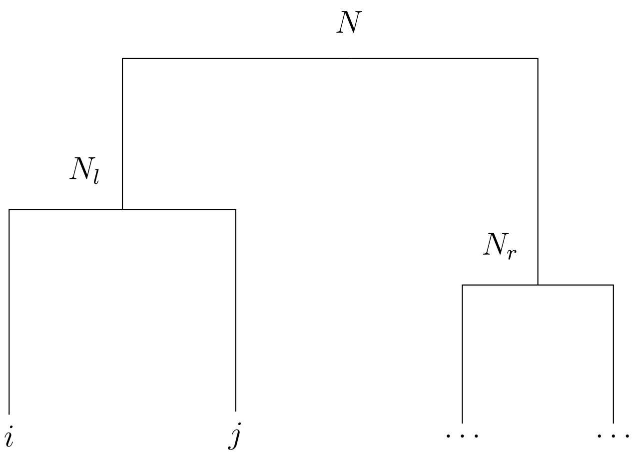

Fig. 3. Dendrogram node and leaves.**Note**: N_l: left node, N_r: right node.

seriation, namely the process of rearranging variables to place similar objects closer together in the sequence [19,76–78]. In the literature, seriation is defined as a systematic reordering of data objects aimed at uncovering structural information [77]. The arrangement of elements in the visualization of multiple variables or instances significantly influences the visualization's interpretability and efficacy [76]. In fact, numerous studies have shown that the systematic reordering of data components substantially enhances data visualization [76,79–81].

2.3.1. Matrix reordering: dendrogram seriation

The foundational dendrogram seriation technique was introduced by Gruvaeus and Wainer in 1972. For any two observations i and j, associated with a seriation weight $s_{w_{jk}}$, this algorithm iteratively permutes the dendrogram's leaves to identify the smallest configuration of seriation weights. The corresponding dendrogram configuration is then selected based on this criterion. In our study, we implement a leaf sort seriation algorithm (DendSer), which integrates hierarchical clustering with a variable sorting mechanism [76]. For each pair of variables j and k (where $j, k \in p$), a seriation weight $s_{w_{jk}}$ is assigned, reflecting the significance of their comparison in the visualization context. Specifically, let N (with $N = 1, \ldots, N_{n-1}$) represent the nodes of a dendrogram Δ. Let $\tilde{\Pi}(N; \Delta)$ denote all possible leaf permutations of Δ, and $\mathcal{F}$ be the function assessing the cost of these permutations. A dendrogram node's graphical illustration is shown in Fig. 3.

The objective of dendrogram seriation is to identify the leaf permutation π^* that minimizes $\mathcal{F}$:

$$\pi^* = \underset{\pi \in \tilde{\Pi}(N_{n-1};\Delta)}{\arg\min} \ \mathcal{F} \tag{11}$$

Rather than evaluating all possible permutations, which would be factorial in number ($n!$), the algorithm employs a node operator, such as reflection or translation, to narrow down the permutation set $\Pi(N; \Delta) \subset \tilde{\Pi}(N; \Delta)$. Here, translation involves swapping the left and right subnodes of a node N, while reflection inverts the order of a node's leaves. Throughout the process, the `DendSer` algorithm sequentially examines the nodes N_i of Δ, updating Δ with the optimal permutation at each step [76]. This iterative process continues until no further changes in π are observed. This methodology positions pairs of variables with low weights adjacently or in close proximity, enhancing the visualization's interpretability.

Seriation: weighting scheme. In our study, we introduce a modification of the seriation algorithm that features a data-driven weighting scheme based on the interaction matrix ($\boldsymbol{\Gamma}$), which is treated as a similarity matrix. To align with our objective of prioritizing variables based on both their individual importance and their interaction scores, we implement a weighting scheme following the approach of Inglis et al. [19]. Specifically, for each pair of variables (j, k), we define a seriation weight $s_{w_{jk}}$ that incorporates both importance and interaction measures:

$$s_{w_{jk}} = \lambda_1 \gamma_{jk} + \lambda_2 \left(\gamma_{jj} + \gamma_{kk} \right) \tag{12}$$

where γ_{jj} and γ_{kk} are the importance scores of variables j and k, respectively, γ_{jk} is the interaction measure between them, and λ_1, λ_2 are scaling factors. These scaling factors normalize variable importance and interaction to the unit interval, accounting for their potentially different orders of magnitude. This weighted sorting scheme arranges the dendrogram leaves such that the weights $s_{w_{jk}}$ are in descending order, optimizing the positioning of high-interacting variable pairs in the visual sequence and enhancing the visualization's interpretability.

3. Application to digital workplaces

In this section, we present an illustrative example of digital work patterns within an Italian digital organization to implement and test the methodological framework discussed in previous sections. Firstly, we applied *sidClustering* and unsupervised RF to inductively identify distinct groups of workers and extract a concise list of the most discriminant digital work features. These features are then used to build DWB composite indicators using different aggregation strategies. In the second stage, we developed and employed visual tools to link the emerging configurations of digital work with employee attitudes. All analyses were performed using R packages `randomForestSRC` [82], `DendSer` [76], `MCDA` [83], and `vivid` [19].

3.1. Sample and measures

Data were collected from the entire workforce ($n = 39$) of a highly digitalized consulting company situated in Northern Italy. The organization operates in the field of AI solutions and applications for business (e.g., marketing, logistics, energy) and is fully digital. In fact, almost 100% of the working activity is performed on the Microsoft365 platform.

As for the sample characteristics, the workforce demonstrates a male predominance, with men constituting approximately 72% of the total employees (28), while females represent 28% (11). A significant majority of the workforce, approximately 92% (36), hold university degrees, reflecting a highly educated workforce, with only 8% (3) having completed only secondary education. The organization's demographics are relatively young, with 69% (27) of employees aged between 26 and 35 years. The average organizational tenure is 2.62 years and the average number of children per employee is 0.56. Regarding employment status, the majority, 31 employees (79%), are on permanent contracts, complemented by 3 freelancers (8%) and 5 interns (13%). All employees are employed on a full-time basis. In terms of job roles, specialists account for 67% of the workforce, while managerial positions comprise 21%. Leadership roles, including heads, constitute 13%. This demonstrates a concentration of employees in specialized roles, with fewer in managerial or leadership positions. The overall workforce age distribution is skewed towards younger employees, with only 2 individuals (5%) aged above 45 and a single employee (3%) in the 18–25 age group. A detailed overview of the sample is in Table 1.

The final dataset consists of two main data sources: (i) 35 click variables from Microsoft365 collected in the period from October 2021 to May 2022; (ii) 93 attitude variables obtained through a self-compiled survey sent in May 2022. A detailed description of the variables is provided in Table 2. For the click variables (i), we sourced and anonymized data from the Microsoft365 platform, which generates time-stamped logs documenting every activity carried out by employees within the platform. Examples of the metrics extracted include "Total number of actions taken by the user", "Maximum number of projects a user works on daily", and "Number of meetings attended by the user during the specified period".

Table 1
Demographic characteristics of the sample.

	Respondents (n = 39)	
	Count (%)	Mean (St. dev.)
Gender		
Female	11 (28.21)	
Male	28 (71.79)	
Age		
18–25	1 (2.56)	
26–35	26 (66.67)	
36–45	10 (25.64)	
>45	2 (5.13)	
Education		
Secondary education	3 (7.69)	
University degree	36 (92.31)	
Role		
Technical	19 (48.72)	
Leader/Head	11 (28.21)	
Manager	9 (23.08)	
Employment status		
Permanent	31 (79.49)	
Freelancer	3 (7.69)	
Intern	5 (12.82)	
Organizational tenure		2.62 (2.05)
N. children		0.56(0.94)

Regarding the attitude variables (ii), we measured respondents through an online survey utilizing a 7-point Likert scale, ranging from strongly disagree (1) to strongly agree (7). The survey assessed a series of work-related attitudes across the following dimensions:

a. Job satisfaction (js): this encompasses various aspects such as satisfaction with empowerment (js_e), job fulfillment (sat_jf), work group (sat_wg), work facilitation (sat_wf), pay (sat_pay), and job security (sat_sec). [84].
b. Embeddedness (po_fit): focused on the fit between the person and the organization [85].
c. Engagement: evaluated through three dimensions: vigor (vig), dedication (ded), and absorption (abs) [86].
d. Work-related stress: This includes assessments of both psychological (psy_d) [87] and physical discomfort (phys_d) [88].
e. Work-life interface: concerning work-life conflict (wl_c) [89] and work-life enrichment (wl_e) [90].
f. Organizational communication: assessed through the quality of communication climate (comm) and the extent of horizontal communication (h_comm) [91].
g. Innovative behavior (inn): this measures the level of organizational innovation and the innovative actions of employees [92].

These items were based on validated scales and drawn from existing literature. A comprehensive list of the attitude variables is reported in Appendix. The descriptive statistics of both data sources are presented in Table 3.

3.2. sidClustering and feature selection

The first stage of the analysis aims to identify clusters of units and construct new metrics for DWB. To this end, we employed *sidClustering* and unsupervised RF on click metadata from Microsoft365. To determine the optimal number of clusters k, we utilized the Gap statistic, which compares the within-cluster dispersion with its expectation under a null reference distribution that assumes no apparent clustering [44]. The null reference dataset is created by Monte Carlo simulations, uniformly generating values across the range of each variable x_j. Let x_{ij} represent the observations, clustered into k clusters. We denote D_r as the sum of pairwise Euclidean distances within cluster r,

Table 2
Click variables from Microsoft365.

Label	Description
n.act	Total number of actions (any type) taken by the user during the period.
role	Role of the user within the company.
peak.act	Maximum number of actions performed by the user in a working day.
avg.act	Average number of actions taken by the user in a workday (Tot. actions/Working days).
extra.act	Number of days in which the user took more actions than his/her average daily actions.
max.out	Maximum number of daily actions taken by the user outside working hours.
d.out	Average number of actions taken by the user in a working day outside working hours.
days_out	Number of days on which the user performed actions outside working hours (9-18).
n.out	Total number of actions performed by the user outside working hours in the period.
dept.peak	Maximum number of departments a user worked for in a day.
d_dept	Average number of departments a user works with on a daily basis.
proj.peak	Maximum number of projects on which a user worked in a day.
d_proj	Average number of projects on which a user works daily.
a.act	Number of active actions (i.e. add, create, update, modify, edit) taken by the user on a file in the period.
task.comp	Number of tasks completed in the period by the user.
task.create	Number of tasks created in the period by the user.
f_task.comp	Number of tasks completed by field in the period by the user.
r.meet	Percentage of meetings done remotely.
out.meet	Number of meetings outside of working hours.
not.met	Average number of hours of notice with which the user was informed of meetings.
hrs.meet	Total hours of meetings the user attended.
peak.hrs.meet	Maximum number of hours of weekly meetings the user attended.
avg.hrs.meet	Average number of hours per week of user's meetings.
overlap.meet	Number of overlapping meetings in the user's period.
n.meet	Total number of meetings the user attended in the period.
peak.meet	Maximum number of weekly meetings in which the user participated.
w_avg.meet	Average number of weekly meetings in which the user participated.
d_peak.meet	Maximum number of daily meetings in which the user participated
d_avg.meet	Average number of daily meetings in which the user participated.
n.team	Number of teams to which the user belongs during the period.
p.rank	Centrality measure from 'Pagerank' algorithm: user's average level of cooperation with other users.
d.score	Centrality measure from 'Degree centrality' algorithm: average number of unidirectional coworking relationships of the user.
p.rank.rel	Centrality measure from 'Pagerank' algorithm: average relational pattern score among users (n. of actions on the same item).
d.score.rel	Centrality measure from 'Degree centrality' algorithm: average number of connections in the relational pattern of the user.
rel.pat	Centrality measure from 'Eigenvector centrality' algorithm: relative influence of the user in the relational pattern network.

and n_r as the number of points in cluster r. The within-cluster sum of squared distances is defined as:

$$W_k = \sum_{r=1}^{k} \left(\frac{1}{2n_r} D_r \right) \tag{13}$$

The Gap function is calculated as the difference between the expected log-transformed within-cluster dispersion of the null reference $E_n^*[\log(W_k^*)]$ and the log-transformed within-cluster dispersion of the dataset $\log(W_k)$:

$$\text{Gap}_n(k) = E_n^*[\log(W_k^*)] - \log(W_k) \tag{14}$$

A graphical representation of the Gap results is displayed in Fig. 4. As can be seen from the graph, the optimal number of clusters is $k = 2$.

Confirmatory results were obtained using the Silhouette method, which evaluates the clustering quality [93]. This method measures how well each object fits within its cluster for various numbers of possible clusters k. The results indicate that the optimal clustering occurs when k is between 2 and 3. Thus, we can conclude that the best separation

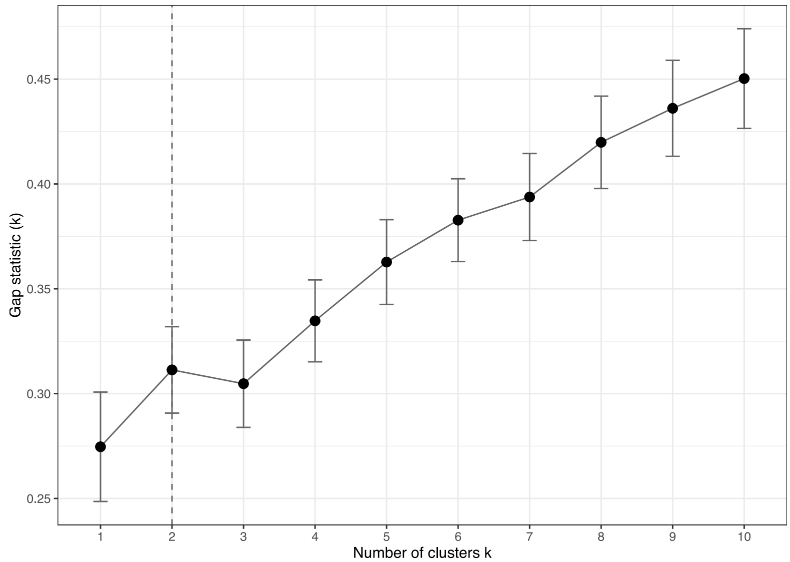

Fig. 4. Gap statistic for k clusters.

of the data is in $k = 2$ groups. Moreover, the Hopkins statistic (0.71) suggests that the data are suitable for cluster analysis [94]. Accordingly, we ran the *sidClustering* algorithm with the following parameters: `k` = 2, `mtry` = 6, `nodesize` = 2, `ytry` = 6, and `ntree` = 500. At the same time, we implemented an unsupervised RF using the multivariate splitting rule in Eq. (1), setting the same hyperparameters (i.e., `mtry`, `nodesize`, `ytry`, and `ntree`) to perform feature selection. These hyperparameters were carefully selected to suit the dataset's dimensions ($n = 39$, $p = 35$). Specifically, an `mtry` of 6, which closely approximates the standard $\sqrt{p}$ heuristic, was chosen to ensure sufficient tree decorrelation [95,96]. A small `nodesize` of 2 allows the trees to grow deep enough to capture fine-grained proximity structures and complex interactions [50], while intentionally avoiding absolute "node purity". The `ytry` parameter, which dictates the number of pseudo-responses evaluated at each split in the multivariate forest, was symmetrically set to 6 to maintain a balanced exploration of the artificial feature space. Finally, `ntree` was set to 500, a standard value sufficiently high to guarantee the convergence and stability of the forest without incurring unnecessary computational costs [95,96].

Fig. 5 illustrates the output from the *sidClustering* algorithm, which includes the processes of data 'sidification' and multivariate impurity splitting. In the graph, rows (i.e., variables) are sorted according to seriation algorithm, taking as a weight a metric incorporating both variable importance and variable interaction (12). Conversely, columns (i.e., observations) are sorted according to the *sidClustering* distance matrix. The ten most relevant features, extracted based on minimal depth from the unsupervised RF, are highlighted in bold. By analyzing the heatmap, we uncover valuable insights into the characteristics and variables associated with each cluster. Cluster 1, displayed on the left-hand side of the heatmap and marked by the red bar, demonstrates higher scores in actions-related variables such as daily average and outwork, as well as task-related variables such as tasks created and completed. These findings align with the predominantly technical roles within this cluster, leading us to designate it as the 'Operational group'. In contrast, Cluster 2, represented by the blue bar, exhibits characteristics aligned with meeting-related variables, including numbers and averages of meetings attended, as well as higher degrees of relational pattern/centrality measures and involvement in multiple departments. These findings correspond to the managerial job roles within this cluster, emphasizing coordination efforts. Consequently, we refer to this cluster as the 'Coordination group'. Both groups display significance in the context of remote working, reflecting the prevalent work arrangement within the environment under study.

3.3. Digital work indicators

In line with the methodology proposed by Boccuzzo and Gianecchini [97], we employ composite indicators to capture job-related phenomena. While prior research has focused primarily on job quality [98], we extend this framework to assess DWB, with a specific focus on two dimensions: job quantity and job complexity. The composite indicators developed in this paper are designed for application at the individual level, thereby accounting for the heterogeneous nature of work experiences and digital behaviors across different worker profiles. As such, distinct sets of indicators are constructed for individuals with varying characteristics, ensuring that the measurement framework remains context-sensitive and representative of the underlying behavioral patterns.

Table 4 reports the feature selection results based on the minimal depth criterion in the unsupervised RF. The 'Min. depth' column shows the average minimal depth across the forest, and the 'Weight' column reflects the inverted min–max normalized minimal depth across variables. As can be seen from the Table, applying the minimal depth thresholding scheme (5) resulted in the selection of 10 key features. The findings highlight the relevance of actions- and meetings-related variables, along with two measures of relational network, namely the number of departments and a centrality measure.

These variables, representing DWB, were categorized according to the traditional dimensions of work analysis and design into two groups: work *Quantity* (denoted as q, with $Q = 8$) and work *Complexity* (denoted as c, with $C = 2$). Work quantity (or workload) refers to the amount of work performed per unit of time (i.e., saturation) and the total volume of work carried out over time (i.e., overflow). Work complexity (or quality) is characterized by the heterogeneity and mutability of work activities.

Compensatory vs. non-compensatory strategies. Based on the findings of the feature selection stage, we constructed a composite indicator for each category, following the approach of Han et al. [99] and Hanadé Houmma et al. [100], employing the importance scores from the unsupervised RF as weights and normalizing them to the $[0, 1]$ range [73]. With reference to Eq. (7), the weighted arithmetic mean for each category is calculated as the sum of the product between each normalized variable z_{ik} and the corresponding weight w_k:

$$\text{Quantity}_i = \sum_{q=1}^{Q} (z_{iq} \times w_q) \tag{15}$$

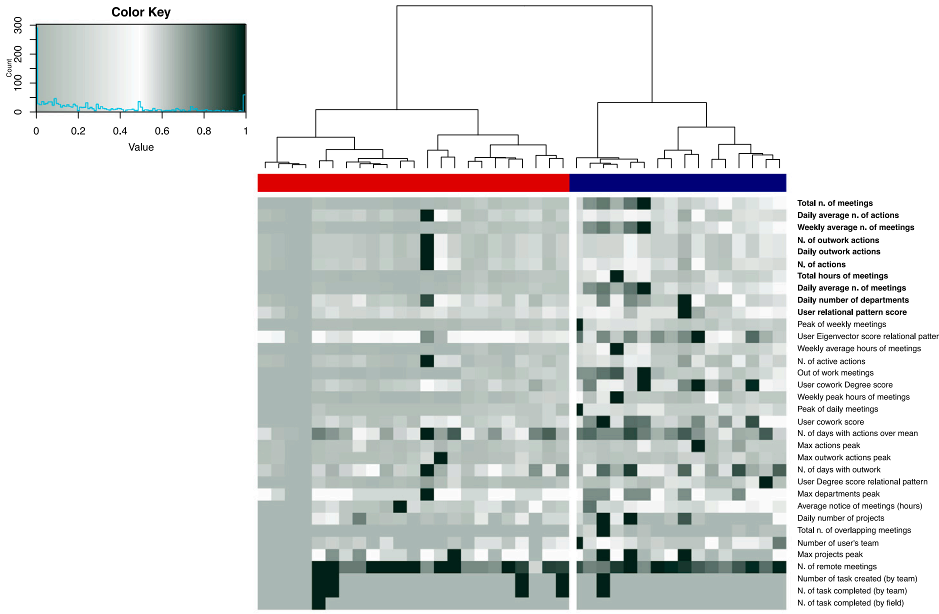

Fig. 5. *sidClustering* heatmap on click metadata.
Note: Data were scaled to the $[0, 1]$ interval to improve interpretability. The color range extends from gray (0) to dark green (1).

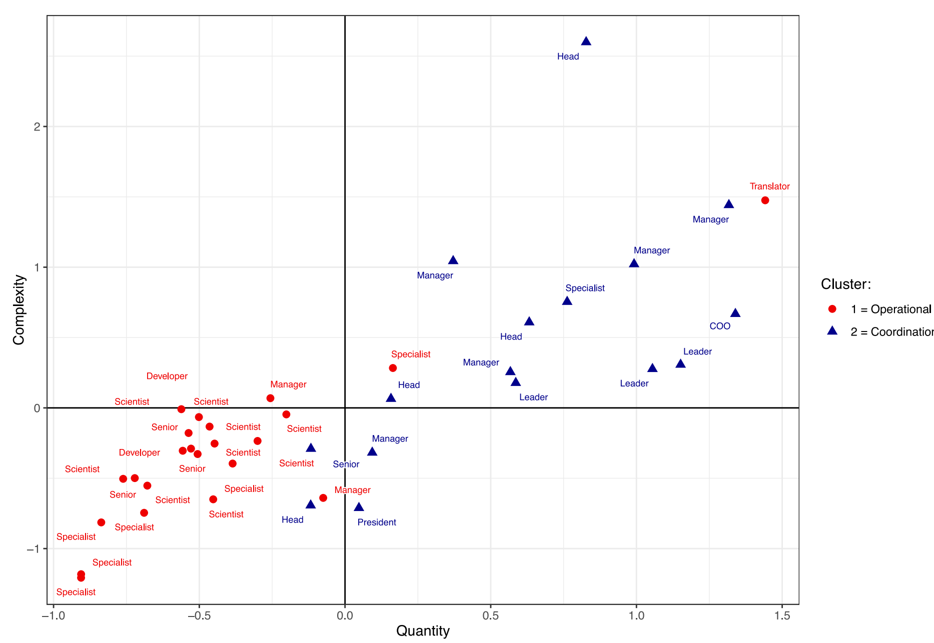

Fig. 6. Quantity vs Complexity – Linear aggregation.

$$\text{Complexity}_i = \sum_{c=1}^{C} (z_{ic} \times w_c) \tag{16}$$

We can now represent graphically the positioning of the study units along the two dimensions (i.e., composite indicators) representing a quantitative representation of the DWB. As can be seen from Fig. 6, which depicts each observation's position relative to the weighted *Quantity* and *Complexity* scores, the clustering produced by *sidClustering* clearly separates the units in the two-dimensional space. Specifically, we can see that each cluster features specific combinations of *Quantity* and *Complexity*: **Cluster 1** is characterized by low quantity and complexity; **Cluster 2**, on the contrary, is characterized by high quantity and complexity.

To confirm these findings and add robustness to our statistical analysis, we constructed the same composite indicators using a non-compensatory strategy, namely the MPI. In particular, we first normalized the raw data $\mathbf{X}$ as in Eq. (8). Then, after computing the mean (M_{x_j}), the standard deviation (S_{x_j}), and the coefficient of variation (CV_i) of each unit i across all q and c indicators, we get:

$$\text{MPI}_i^{\text{Q}} = \text{M}_{z_i}^{\text{Q}} + \text{S}_{z_i}^{\text{Q}} \cdot \text{CV}_i^{\text{Q}} \tag{17}$$

$$\text{MPI}_i^{\text{C}} = \text{M}_{z_i}^{\text{C}} + \text{S}_{z_i}^{\text{C}} \cdot \text{CV}_i^{\text{C}} \tag{18}$$

which represent the two MPI composite indicators for *Quantity* and *Complexity*, respectively. The results of the MPI are reported in Fig. 7. As can be noted from the figure, the two different aggregation

Table 3
Descriptive statistics.

Variable type	Label	Mean	Median	St.Dev.	Min.	Max.
Click – (source: Microsoft365)	n.act	2761.05	2219	2218.33	5	10 543
	peak.act	210.97	134	244.78	2	1138
	avg.act	18.58	14.94	14.93	0.03	70.96
	extra.act	42.31	46	14.66	4	65
	max.out	57.49	35	72.64	0	418
	d.out	3.86	3.26	3.41	0	18.3
	days_out	83.33	89	39.81	0	155
	n.out	573.85	484	506.31	0	2719
	dept.peak	1.77	2	0.87	0	4
	d_dept	0.41	0.32	0.35	0	1.35
	proj.peak	1.38	1	1.29	0	4
	d_proj	0.15	0.06	0.21	0	0.73
	a.act	698.74	495	733.22	0	3698
	task.comp	0.13	0.04	0.34	0	1
	task.create	0.13	0.04	0.34	0	1
	f_task.comp	0.03	0	0.16	0	1
	r.meet	0.79	0.89	0.3	0	1
	out.meet	6.64	4	9.06	0	32
	not.met	138.28	127.22	111.69	0	506.87
	hrs.meet	43.81	24	61.57	0	300
	peak.hrs.meet	13.03	6.5	20.41	0	100
	avg.hrs.meet	1.47	0.81	2.07	0	10.1
	overlap.meet	0.59	0	2.14	0	13
	n.meet	29.56	13	35.58	0	130
	peak.meet	5.28	4	6.81	0	40
	w_avg.meet	0.99	0.44	1.2	0	4.38
	d_peak.meet	3.13	2	3.33	0	19
	d_avg.meet	0.2	0.09	0.24	0	0.88
	n.team	3.15	2	2.28	1	11
	p.rank	0.97	0.81	0.68	0.15	2.57
	d.score	7062.92	5368	5819.35	0	19 731
	p.rank.rel	1	0.93	0.48	0.15	2.75
	d.score.rel	2142.49	845	3353.04	0	16 746
	rel.pat	0.11	0.11	0.04	0	0.2
Attitude – (source: survey)	js	5.68	6	1.06	3	7
	js_e	5.35	5.71	1.26	1.43	7
	sat_jf	5.46	6	1.24	2.67	7
	sat_wg	5.94	6	0.82	4	7
	sat_wf	5.28	5.2	1.14	1.8	7
	sat_pay	4.73	5	1.18	2	7
	sat_sec	5.69	6	1.14	1.5	7
	po_fit	5.57	5.8	1.12	2.2	7
	vig	5.41	5.5	0.89	3.17	7
	ded	5.18	5.4	1.23	2.4	7
	abs	5.16	5.33	1.2	2.17	7
	psy_d	4.37	4.33	0.67	3.17	5.75
	phys_d	2.5	2.22	1.08	1	5.11
	wl_c	3.05	2.8	1.52	1	5.8
	wl_e	4.23	4.5	1.35	1	6.5
	comm	4.66	5	1.56	1	7
	h_comm	4.72	5	0.95	2.4	6.2
	inn	4.89	4.83	1.16	1.83	7

strategies produce very similar results in terms of units positioning in the two-dimensional graph, adding robustness to the results.

Lastly, to complete the analysis of the units' positioning with respect to the two categories of DWB, we applied the ELECTRE III method separately for each category to generate unit rankings. We set the indifference threshold at $0.05r$, the preference threshold at $0.15r$, and the veto threshold at $0.75r$, where $r = (\mathbf{X}_{\max} - \mathbf{X}_{\min})$ represents the data range [101]. The results are in Table 5.

Since aggregating ascending and descending preorders by intersection would result in preferential, indifferent, or incomparable relations, we applied the Borda rule to obtain the final ranking [102]. In this approach, the descending ranks were reversed to align with the direction of the ascending ranks, where higher values indicate better performance. Consequently, the final rank of each alternative corresponds to the average of its positions in the two distillation procedures. As shown in the table, the top 10 units ranked in the *Quantity* category mainly belong to the coordination group (i.e., Manager, Head, Leader). These units also exhibit perfect rank intersection, meaning that their ascending and descending distillation ranks coincide. Similarly, in the *Complexity* category, the top 10 units are also drawn from the coordination group, with the majority (6 out of 7) showing rank intersection. These results are consistent with the findings based on composite indicators constructed using both compensatory and non-compensatory aggregation methods, which also indicate that units with high *Quantity* and *Complexity* scores primarily occupy managerial or coordination roles.

3.4. Digital work and employee attitudes

As a final step, we investigated the relationship between employee attitudes and the digital work scores of *Quantity* and *Complexity* within each cluster identified through our clustering process. In this phase, we linked the findings from the analysis of DWB, represented by the composite constructed through linear aggregation, to employee attitudes measured by the questionnaire. This approach is in line with Thanos et al. [14], who emphasizes the benefits of integrating data from multiple sources to foster knowledge creation. Prior to conducting the analysis, we aggregated the scores for each employee within each attitude dimension to simplify the investigation of the relationships among variables. We then combined the aggregated attitude variables with the two composite indicators derived through linear aggregation and implemented an unsupervised RF using the following parameters: `mtry` = 5, `nodesize` = 2, `ytry` = 5. The interaction matrices coming from the unsupervised RF and used to build the network plots underwent the leaf sort seriation process (i.e. variable reordering) specified in Section 2.3.1. This procedure strategically positions variables of high importance and significant interaction in a clockwise arrangement, initiating from the top. The network plots, representing the outcome of the unsupervised RF analysis in terms of variable importance and interaction, are presented in Figs. 8 and 9.

As for the **Operational** group (Fig. 8), the most important features pertain to the domains of *Job Satisfaction*, *Engagement* (i.e., Vigor, Absorption, Dedication), and *Communication climate*. Employees in this group show a strong connection between *Dedication* – defined as experiencing meaning, enthusiasm, pride, and inspiration at work – and *Job Satisfaction*, particularly through *Satisfaction with job fulfillment*, which reflects the use of skills, personal accomplishment, and appreciation of work content. Robust interactions are also observed among satisfaction-related attitudes such as *Satisfaction with empowerment* (involvement in decision-making, encouragement for skill development, and opportunity for innovation), *Satisfaction with work facilitation* (access to training, information, and favorable working conditions), and overall *Job satisfaction*. These interconnections suggest that, in a highly digitalized work environment, employees' perceptions of job meaningfulness, empowerment, and support through adequate resources are tightly linked. Two additional relationships deserve attention. First, the link between *Dedication* and *Psychological discomfort* indicates that higher engagement may be accompanied by psychological strain. This could reflect the challenges typical of digitalized work settings, such as increased performance pressure, connectivity demands, and blurred boundaries between work and private life [7]. This resonates with concerns in the literature on *technological ill-being*, where constant device use and application switching can contribute to stress, overload, and even burnout [103–105]. At the same time, the positive links to job satisfaction align with studies emphasizing that technology can enrich jobs and increase motivation when it supports meaningful work [106]. Second, the connection between *Dedication* and *Communication climate* underscores the importance of transparent, timely, and motivating communication in fostering employees' commitment. In digitalized organizations, where much interaction occurs through mediated channels, an effective communication climate seems critical to reinforcing a sense of belonging and shared purpose, consistent with

Table 4
Minimal depth rank and variable weights.

Variable	Category	Description	Min. depth	Weight
n.meet	Quantity	Total number of meetings the user attended.	4.322	1
avg.act	Quantity	Average number of actions taken by the user in a workday.	4.392	0.939
w_avg.meet	Quantity	Average number of weekly meetings in which the user participated.	4.414	0.920
n.out	Quantity	Total number of actions performed by the user outside working hours.	4.420	0.915
d.out	Quantity	Average number of actions taken by the user in a working day outside working hours.	4.438	0.899
n.act	Quantity	Total number of actions (any type) taken by the user.	4.450	0.889
hrs.meet	Quantity	Total hours of meetings the user attended.	4.482	0.861
d_avg.meet	Quantity	Average number of daily meetings in which the user participated.	4.482	0.861
d_dept	Complexity	Average number of departments a user works with on a daily basis.	4.528	0.821
p.rank.rel	Complexity	Centrality measure from 'Pagerank' algorithm: average relational pattern score among users.	4.534	0.815

Note: Weights are computed via min–max normalization of minimal depths D_v. The filtering threshold is defined in Eq. (5).

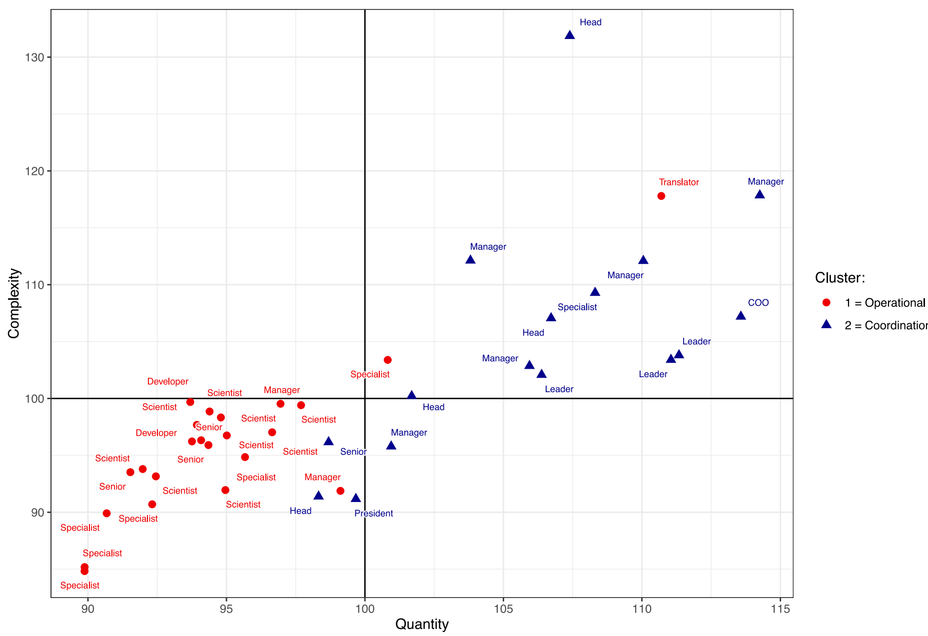

Fig. 7. Quantity vs Complexity – MPI.

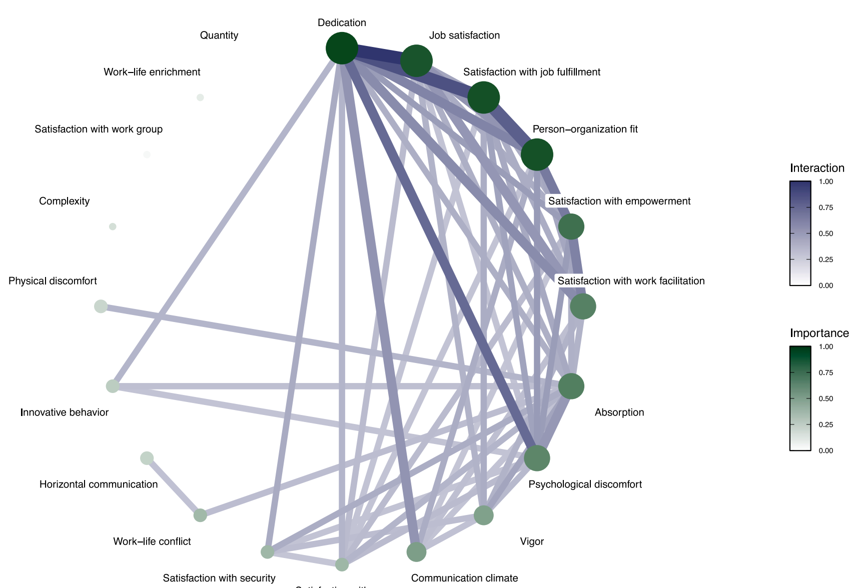

Fig. 8. Network plot: Operational group. **Note**: Node size represents variable importance; edge thickness represents the strength of variable interaction.

Table 5
ELECTRE III ranking.

Quantity					Complexity				
Alternative	Ascending	Descending	Average[a]	Int.[b]	Alternative	Ascending	Descending	Average[a]	Int.[b]
Head	5	4	20.5	0	Manager	5	4	20.5	0
Scientist	11	10	20.5	0	Specialist	6	5	20.5	0
Manager	1	1	20	1	Specialist	9	8	20.5	0
Head	6	6	20	1	Head	1	1	20	1
Manager	7	7	20	1	Translator	2	2	20	1
Manager	8	8	20	1	Manager	3	3	20	1
Senior	9	9	20	1	COO	6	6	20	1
Scientist	10	10	20	1	Head	7	7	20	1
Manager	1	2	19.5	0	Leader	10	10	20	1
Leader	1	2	19.5	0	Leader	11	11	20	1
COO	1	2	19.5	0	Manager	4	5	19.5	0
Manager	11	12	19.5	0	Manager	8	9	19.5	0
Specialist	13	14	19.5	0	Leader	9	10	19.5	0
Scientist	17	18	19.5	0	Head	11	12	19.5	0
Scientist	17	18	19.5	0	Scientist	12	14	19	0
Specialist	19	20	19.5	0	Scientist	14	16	19	0
Specialist	19	20	19.5	0	Scientist	15	17	19	0
Specialist	1	3	19	0	Specialist	18	20	19	0
Leader	1	3	19	0	Specialist	23	25	19	0
Scientist	13	15	19	0	Specialist	24	26	19	0
Developer	15	17	19	0	Specialist	24	26	19	0
Senior	15	17	19	0	Manager	10	13	18.5	0
Specialist	15	17	19	0	Developer	12	15	18.5	0
Specialist	16	18	19	0	Scientist	12	15	18.5	0
Senior	17	19	19	0	Scientist	16	19	18.5	0
Specialist	18	20	19	0	Senior	16	19	18.5	0
Leader	2	5	18.5	0	Scientist	22	25	18.5	0
Manager	2	5	18.5	0	Specialist	22	25	18.5	0
Manager	3	6	18.5	0	Head	22	25	18.5	0
Scientist	13	16	18.5	0	President	22	25	18.5	0
Scientist	14	17	18.5	0	Manager	21	25	18	0
Senior	14	17	18.5	0	Scientist	13	18	17.5	0
Scientist	12	17	17.5	0	Developer	16	21	17.5	0
Developer	12	17	17.5	0	Senior	17	22	17.5	0
Head	1	8	16.5	0	Manager	18	23	17.5	0
Head	6	13	16.5	0	Scientist	20	25	17.5	0
President	6	13	16.5	0	Senior	20	25	17.5	0
Specialist	4	12	16	0	Scientist	19	25	17	0
Translator	1	11	15	0	Senior	17	24	16.5	0

Note:
[a] Computed using the Borda rule.
[b] Ranking intersection: 1=yes; 0=no.

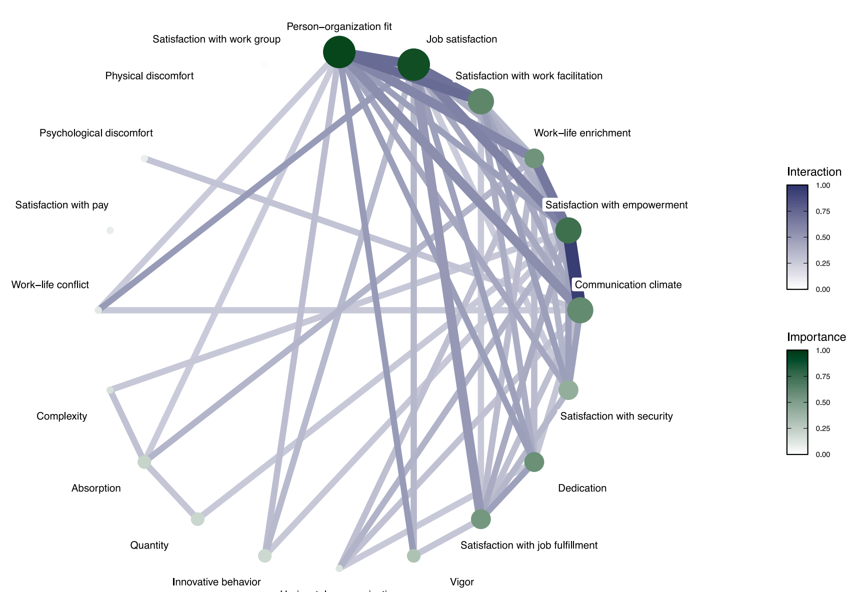

Fig. 9. Network plot: Coordination group. **Note**: Node size represents variable importance; edge thickness represents the strength of variable interaction.

earlier findings that ICT resources can mitigate isolation and enhance satisfaction [107,108].

The **Coordination** group (Fig. 9), on the contrary, displays a distinct pattern. Here, *Person-organization fit*, reflecting the perception of value congruence, cultural alignment, and role compatibility with the organization, emerges as the dominant variable. *Person-organization fit* is strongly linked to *Job satisfaction*, *Satisfaction with work facilitation*, and *Work-life enrichment*. The latter refers to the positive transfer of skills, attitudes, and experiences from work to home life, enhancing personal and family functioning. These associations suggest that managerial or coordination roles may benefit from a strong sense of alignment with organizational values, access to resources, and a healthy interface between work and private life. This is consistent with research highlighting how digitalization can promote work-life balance through flexibility and blended working arrangements [109]. *Satisfaction with empowerment* also plays a crucial role in this group, consistent with the autonomy and influence typically afforded by managerial positions. Moreover, *Communication climate*, namely the extent to which communication is open, motivating, and effective, shows a tight connection with *Person-organization fit*, highlighting the centrality of communication in reinforcing organizational alignment and identity. Additionally, *Job Satisfaction* is closely tied to both *Satisfaction with empowerment* and *Satisfaction with job fulfillment*, indicating that satisfaction in these roles may stem not only from the nature of the tasks performed but also from autonomy, empowerment, and a sense of purpose. Interestingly, no work-related stress variables emerge as central or highly connected in this group. This may indicate that, for employees in managerial or coordination roles, potential stressors are counterbalanced by job resources such as empowerment, communication quality, and person-organization fit, which enhance resilience and mitigate the impact of job demands.

Across both groups, satisfaction with the work group does not emerge as a key predictor. However, we interpret this pattern cautiously and consider three non-mutually-exclusive explanations. First, it may reflect a broader feature of increasingly digitalized and remote work: coordination is more technology-mediated and often asynchronous, potentially reducing the impact of team social dynamics relative to individual or job design factors [110,111]. Second, it may be organization-specific: organizational design choices (e.g., standardized procedures, individualized performance management, or limited task interdependence) can reduce reliance on peer coordination and thereby dampen the observable role of work-group climate [112]. Third, a methodological explanation is plausible: work-group satisfaction shows high central tendency and limited dispersion in our sample, consistent with restricted range/ceiling effects that can attenuate observed associations and reduce the likelihood that a variable is selected early in trees.

Taken together, these results illustrate a dual effect of digitalization. On the one hand, digital work supports engagement, satisfaction, and flexibility, consistent with prior studies on autonomy, job enrichment, and work-life balance [106,109]. On the other hand, it also carries risks of psychological discomfort and overload, echoing concerns about technological ill-being in the digital workplace [7,104,105]. This duality supports the argument that understanding employees' holistic digital work experience, spanning both enabling resources and potential stressors, is crucial to balancing the benefits of digitalization with its adverse effects, as emphasized by Marsh et al. [7].

4. Discussion and conclusion

This era is characterized by the ubiquity of work, which can occur anytime and anywhere, and by the replacement of on-premise, direct supervision during working hours with remote, "digital" observability enabled by the datafication of work. In organizational contexts, work datafication involves the continuous creation of data documenting employee behavior, thereby enhancing the visibility of such behaviors [1]. As work processes become increasingly digitalized, work behaviors generate a wealth of digital traces, offering unprecedented opportunities to analyze work behaviors [113,114]. Such data enable the application of sophisticated statistical techniques to extract insights, identify patterns, and advance HR theory and research, potentially transforming HRM into an evidence-based, data-driven practice [4,8,115].

To address the need for extracting actionable insights from the increasing availability of work data, we propose a data-driven methodology to discover patterns and types of digital work in digital workplaces. Analyzing such complex data requires a suite of sophisticated algorithms and statistical methods to make the content functionally discernible [1,44,116]. In this context, we present a methodological framework that leverages RF-based models, composite indicators with diverse aggregation strategies, and graphical representations designed to enhance interpretability and facilitate knowledge extraction [19,20,30,76]. These techniques enable researchers to: (i) group observations based on a tree-based distance matrix; (ii) extract the key variables representing DWB; (iii) develop new synthetic digital work metrics using composite indicators; and (iv) investigate the interaction between the new digital work metrics and attitude data, and represent this interaction graphically using tailored visual tools. The implementation of such methods supports inductive exploration of DWB, potentially leading to new theoretical and practical insights [14,15,117].

While the proposed framework is effective with small sample sizes [38,39], from a computational perspective, it is also highly scalable to larger datasets. Tree-based ensemble methods like RF and *sidClustering* are intrinsically parallelizable and efficiently handle large numbers of observations [30,118]. Similarly, the computation of composite indicators via linear aggregation and MPI scales linearly with the sample size. The ELECTRE III outranking method, however, scales quadratically due to exhaustive pairwise comparisons [119] and may require prior segmentation for very large n. Lastly, visual tools can be easily tailored to accommodate larger datasets, for example, by applying stricter thresholds to display only the most important features or interactions.

To test our framework, we collected data on a sample of employees from a highly digital consulting company. The dataset combined two distinct sources: (i) click data from Microsoft365 (35 variables) capturing daily work activity within the platform, and (ii) 93 work attitudes measured on a 7-point Likert scale and collected through an online questionnaire. The joint implementation of *sidClustering* and the extraction of the most important features through an unsupervised RF made it possible to identify groups of units and extract the most relevant features to develop synthetic representations of DWB: work *Quantity* and work *Complexity*. The emerging patterns reveal two distinct clusters, classifying employees into an 'Operational' group and a 'Coordination' group. In the second step, these new digital work configurations were linked to employee attitudes using tailored visual tools to explore the relationship between digital work and employee attitudes. Our findings reveal two distinct patterns in the attitudes of employees working in a highly digitalized environment. In the 'Operational' group, employees' experience is primarily shaped by the interplay between *Job Satisfaction*, *Engagement*, and *Communication Climate*. High levels of *Dedication*, expressed as enthusiasm and pride in one's work, are closely linked to job fulfillment and are supported by empowerment, access to resources, and meaningful communication. However, this engagement is not without risks, as it is also associated with psychological discomfort, reflecting the tension between engagement and strain in digital work settings. Conversely, the 'Coordination' group is characterized by the centrality of *Person-organization fit*, underscoring the importance of value alignment, organizational support, and empowerment, especially in managerial roles. In this group, positive organizational attitudes seem to buffer potential stressors, emphasizing the role of empowerment and effective communication in sustaining satisfaction and mitigating strain. Across both groups, *Satisfaction with the work group* does not emerge as a significant factor, suggesting that employees tend to work individually rather than in tightly integrated teams, a dynamic possibly accentuated by the highly digitalized work environment.

Table A.1
Survey items.

Dimension	Attitude	Questionnaire item
JOB SATISFACTION [84]	Overall job satisfaction (js)	– Considering everything, how satisfied are you with your job? – How would you rate the organization as a company to work for compared to other companies? – Considering everything, how would you rate your overall satisfaction with the organization at the present time?
	Satisfaction with empowerment (js_e)	– How satisfied are you with your involvement in the decisions that affect your work? – Sufficient effort is made to get the opinions and thinking of people who work at the organization. – How satisfied are you with the information you receive from management regarding what is going on in the company? – How satisfied are you with the opportunity to get a better job at the company? – I am given a real opportunity to improve my skills in the company. – I feel encouraged to come up with new and better ways of doing things. – Overall, how good a job do you feel is being done by your immediate supervisor/manager?
	Satisfaction with job fulfillment (sat_jf)	– I like the kind of work I do. – My work gives me a feeling of personal accomplishment. – My job makes good use of my skills and abilities.
	Satisfaction with work group (sat_wg)	– How would you rate the overall quality of work done in your work group? – The people I work with cooperate to get the job done.
	Satisfaction with work facilitation (sat_wf)	– The company is making the changes necessary to compete effectively. – How satisfied are you with the training you received for your present job? – I have enough information to do my job well. – Conditions at my job allow me to be about as productive as I could be. – How satisfied are you with your physical working conditions?
	Satisfaction with pay (sat_pay)	– In comparison with people in similar jobs in other companies my pay is – How do you rate the amount of pay you get on your job?
	Satisfaction with security (sat_sec)	– How do you rate this company in providing job security for people like yourself? – How do you rate your total benefits program?
EMBEDDEDNESS – [85]	Person-organization fit (po_fit)	–The company utilizes my skills and talents well. – I feel like I am a good match for the organization. – I fit with the organization's culture. – I like the authority and responsibility I have at the company. – My values are compatible with the organization's values.
ENGAGEMENT [86]	Vigor (vig)	– At my work, I felt bursting with energy. – At my job, I felt strong and vigorous. – When I got up in the morning, I felt like going to work. – I could continue working for very long periods at a time. – At my job, I was very resilient, mentally. – At my work I always persevered, even when things did not go well.
	Dedication (ded)	– I found the work that I did full of meaning and purpose. – I was enthusiastic about my job. – My job inspired me. – I was proud of the work that I did. – To me, my job was challenging.
	Absorption (abs)	– Time flew when I was working. – When I was working, I forgot everything else around me. – I felt happy when I was working intensely. – I was immersed in my work. – I got carried away when I was working. – It was difficult to detach myself from my job.

(*continued on next page*)

Table A.1 (*continued*).

Dimension	Attitude	Questionnaire item
WORK-RELATED STRESS [87,88]	Psychological discomfort (psy_d)	– Were you able to concentrate on whatever you were doing? – Did you feel like you were playing a useful part in the things that happened to you? – Did you feel capable of making decisions about the things that happened to you? – Were you able to enjoy your day-to-day activities? – Were you able to face your problems? – Did you suffer from loss of sleep over worries? – Did you feel constantly under strain? – Did you feel able to overcome the difficulties? – Have you felt unhappy or depressed? – Have you lost confidence? – Have you thought of yourself as worthless? – Did you feel reasonably happy?
	Physical discomfort (phys_d)	– I have been feeling headache and concentration difficulties. – I have been feeling stomach pain, gastritis. – I have been feeling nervousness, restlessness, anxiety. – I have been feeling a sense of excessive fatigue. – I have been feeling asthma, breathing difficulties. – I have been feeling muscle and joint pain. – I have been feeling difficulties to fall asleep, insomnia. – I have been feeling a sense of depression. – I have been feeling visual impairment.
WORK-LIFE INTERFACE [89,90]	Work-life conflict (wl_c)	– Did you feel like you were playing a useful part in the things that happened to you? – Due to work-related duties, I had to make changes to my plans for family activities. – The demands of my work interfered with my home and family life. – My job produced strain that made it difficult to fulfill family duties. – The amount of time my job took up made it difficult to fulfill family responsibilities. – Things I wanted to do at home did not get done because of the demands my job put on me.
	Work-life enrichment (wl_e)	– The things I did at work helped me deal with personal and practical issues at home. – The things I did at work made me a more interesting person at home. – The skills I used on my job were useful for things I had to do at home. – Having a good day on my job made me a better companion when I got home.
ORGANIZATIONAL COMMUNICATION [91]	Communication climate (comm)	– Extent to which communication motivated and stimulated enthusiasm for meeting organization goals. – Extent to which the people at the company had great ability as communicators. – Extent to which the communication made me identify with or made me feel a vital part of the organization. – Extent to which I received on time the information needed to do my job. – Extent to which conflicts were handled appropriately through proper communication channels.
	Horizontal communication (h_comm)	– Extent to which the grapevine was active. – Extent to which horizontal communication with other employees was accurate and free-flowing. – Extent to which communication practices were adaptable to emergencies. – Extent to which my work group was compatible. – Extent to which informal communication was active and accurate.
INNOVATION [92]	Innovative behavior (inn)	– At work, I generate creative ideas. – At work, I promote and champion my creative ideas. – At work, I seek new technology, process technique and/or product ideas. – At work, I develop adequate plans and schedules for the implementation of new ideas. – At work, I investigate and secure the funding and resources needed to implement new ideas. – Overall, I consider myself a creative member of my team.

4.1. Implications

From a methodological perspective, this study advances the field by presenting a data-driven and scalable framework grounded in tree-based learning techniques. By combining unsupervised tree-based methods, composite indicators, and visual tools, the framework enables the extraction of actionable knowledge through synthetic metrics and graphical representations that extend beyond traditional approaches. This provides practical support for HR analytics and evidence-based policy development [19–21]. In addition, the proposed framework also facilitates the integration of diverse data sources (i.e., digital work metadata and survey-based measures of employee attitudes), thereby fostering the discovery of novel patterns and insights [14]. This study represents one of the first empirical applications of combined datasets in a digital workplace, directly addressing a key limitation of existing research, where privacy concerns and disciplinary silos have hampered the use of such data [22–24]. From a theoretical perspective, the empirical findings extend the literature on digital work and organizational theory in highly digitalized settings by clarifying how distinct roles within organizations are associated with different attitudes. The results serve a dual purpose: uncovering patterns in digital work behavior and generating evidence to inform the social dimension of organizational sustainability. The composite indicators developed here act as quantitative descriptors of work in the digital era and function as early-warning tools to anticipate the consequences of technology adoption on employee outcomes. This has important implications for HR theory, organizational policy, and managerial decision-making [8]. More broadly, these indicators can inform the design of employee policies targeting social sustainability goals, such as SDG 8 (Decent Work and Economic Growth) [12,13]. For example, they could be applied to monitor the effectiveness of policy interventions not only within organizations but also across industries and institutions, positioning them as valuable instruments for linking micro-level digital work practices to macro-level sustainability agendas.

4.2. Limitations and future research

This study is subject to certain limitations. From a theoretical perspective, the explanatory power of the new metrics for work quantity and complexity should be further tested across diverse organizational settings. However, the main restriction lies in their applicability exclusively to jobs performed entirely within digital workplaces, limiting their generalizability to non-digital or hybrid work contexts. Future research could apply the proposed framework to larger organizational contexts to evaluate the scalability of the statistical methods and extend the model to longitudinal data in order to track the effects of policy interventions.

CRediT authorship contribution statement

F. Demaria: Formal analysis, Data curation. **A. Papana Dagiasis:** Methodology, Conceptualization. **M. Cavicchioli:** Supervision, Methodology. **T. Fabbri:** Project administration, Investigation.

Code availability

The R code used for the statistical analyses in this study is available from the corresponding author upon reasonable request.

Funding

Open access funding provided by University of Modena and Reggio Emilia within the CRUI-CARE Agreement.

Declaration of competing interest

The authors have no financial or proprietary interests in any material discussed in this work.

Acknowledgments

Cavicchioli, Demaria, and Fabbri gratefully acknowledge financial support from the research grant FAR 2021 "*Work datafication and behavioral visibility in the digital workplace*", provided by University of Modena and Reggio Emilia, Italy.

Appendix

The complete list of attitude variables included in the questionnaire is detailed in Table A.1.

Data availability

Data will be made available on request.

References

[1] Leonardi PM, Treem JW. Behavioral visibility: A new paradigm for organization studies in the age of digitization, digitalization, and datafication. Organ Stud 2020;41:1601–25. http://dx.doi.org/10.1177/0170840620970728.

[2] Loebbecke C, Picot A. Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. J Strateg Inf Syst 2015;24(3):149–57. http://dx.doi.org/10.1016/j.jsis.2015.08.002.

[3] Oztemel E, Gursev S. Literature review of industry 4.0 and related technologies. J Intell Manuf 2020;31(1):127–82. http://dx.doi.org/10.1007/s10845-018-1433-8.

[4] Fabbri T. Digital work: an organizational perspective. In: Ales E, Curzi Y, Fabbri T, Rymkevich O, Senatori I, Solinas G, editors. Working in digital and smart organizations. Cham: Springer International Publishing; 2018, p. 29–38. http://dx.doi.org/10.1007/978-3-319-77329-2.

[5] Vial G. Understanding digital transformation: A review and a research agenda. J Strat Inf Syst 2019;28(2):118–44. http://dx.doi.org/10.1016/j.jsis.2019.01.003.

[6] Attaran M, Attaran S, Kirkland D. The need for digital workplace: Increasing workforce productivity in the information age. Int J Enterp Inf Syst 2019;15(1):1–23. http://dx.doi.org/10.4018/IJEIS.2019010101.

[7] Marsh E, Vallejos EP, Spence A. The digital workplace and its dark side: An integrative review. Comput Hum Behav 2022;128:107118. http://dx.doi.org/10.1016/j.chb.2021.107118.

[8] McAbee ST, Landis RS, Burke MI. Inductive reasoning: The promise of big data. Hum Resour Manag Rev 2017;27(2):277–90. http://dx.doi.org/10.1016/j.hrmr.2016.08.005.

[9] McAfee A, Brynjolfsson E. Big data: The management revolution. Harv Bus Rev 2012;90:60–8.

[10] Jiang Y, Xie Y, Shao Q. How did internet usage affect life satisfaction before and after COVID-19? Mediating effects and heterogeneity analysis. Socio-Economic Plan Sci 2024;95:102007.

[11] von Krogh G. Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. Acad Manag Discov 2018;4:404–9. http://dx.doi.org/10.5465/amd.2018.0084.

[12] United Nations General Assembly. Transforming our world: The 2030. Agenda for sustainable development. 2015, URL.

[13] Wang S, Zhang H. Enhancing SMEs sustainable innovation and performance through digital transformation: Insights from strategic technology, organizational dynamics, and environmental adaptation. Socio-Economic Plan Sci 2025;98:102124.

[14] Thanos C, Meghini C, Bartalesi V, Coro G. An exploratory approach to data driven knowledge creation. J Big Data 2023;10:1–15. http://dx.doi.org/10.1186/s40537-023-00702-x.

[15] Kitchin R. Big data, new epistemologies and paradigm shifts. Big Data & Soc 2014;1. http://dx.doi.org/10.1177/2053951714528481.

[16] Ho Yu C. Exploratory data analysis in the context of data mining and resampling.. Int J Psychol Res 2010;3:9–22. http://dx.doi.org/10.21500/20112084.819.

[17] Bakshy EB, Eckles D, Bernstein M. Designing and deploying online field experiments. In: 23rd world wide web international conference. New York; 2014, p. 1–10.

[18] Demaria F, Cavicchioli M. Employees' well-being, work-related factors, and digitalization: a decision-tree approach for imbalanced data. Ann Oper Res 2025;1–29. http://dx.doi.org/10.1007/s10479-025-06702-9.

[19] Inglis A, Parnell A, Hurley CB. Visualizing variable importance and variable interaction effects in machine learning models. J Comput Graph Statist 2022;31:766–78. http://dx.doi.org/10.1080/10618600.2021.2007935.

[20] Mantero A, Ishwaran H. Unsupervised random forests. Stat Anal Data Min: ASA Data Sci J 2021;14:144–67. http://dx.doi.org/10.1002/sam.11498.

[21] Paruolo P, Saisana M, Saltelli A. Ratings and rankings: Voodoo or science? J R Stat Soc Ser A: Stat Soc 2013;176:609–34. http://dx.doi.org/10.1111/j.1467-985X.2012.01059.x.

[22] Giermindl LM, Strich F, Christ O, Leicht-Deobald U, Redzepi A. The dark sides of people analytics: reviewing the perils for organisations and employees. Eur J Inf Syst 2022;31:410–35. http://dx.doi.org/10.1080/0960085X.2021.1927213.

[23] Greasley K, Thomas P. HR analytics: The onto-epistemology and politics of metricised HRM. Hum Resour Manag J 2020;30:494–507. http://dx.doi.org/10.1111/1748-8583.12283.

[24] Leonardi P, Contractor N. Better people analytics. Harv Bus Rev 2018;2018. Cited by: 73.

[25] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York, NY: Springer US; 2021, http://dx.doi.org/10.1007/978-1-0716-1418-1.

[26] Hartigan JA, Wong MA. A K-means clustering algorithm. Appl Stat 1979;28:100. http://dx.doi.org/10.2307/2346830.

[27] Johnson SC. Hierarchical clustering schemes. Psychometrika 1967;32:241–54. http://dx.doi.org/10.1007/BF02289588.

[28] Kaufman L, Rousseeuw PJ. Wiley series in probability and statistics. Hoboken, NJ: John Wiley & Sons, Inc; 1990, p. 68–125. http://dx.doi.org/10.1002/9780470316801.ch2.

[29] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth berkeley symposium on mathematical statistics and probability. Berkeley, CA: University of California Press; 1967, p. 281–97.
[30] Breiman L. Random forests.. Mach Learn 2001;45:5–32. http://dx.doi.org/10.1023/A:1010933404324.
[31] Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J: Promot Commun Stat Stata 2020;20:3–29. http://dx.doi.org/10.1177/1536867X20909688.
[32] Breiman L. Bagging predictors. Mach Learn 1996;24:123–40. http://dx.doi.org/10.1007/BF00058655.
[33] Amit Y, Geman D. Shape quantization and recognition with randomized trees. Neural Comput 1997;9:1545–88. http://dx.doi.org/10.1162/neco.1997.9.7.1545.
[34] Ho TK. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 1998;20:832–44. http://dx.doi.org/10.1109/34.709601.
[35] Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Mach Learn 2000;40:139–57. http://dx.doi.org/10.1023/A:1007607513941.
[36] Speer AB, Perrotta J, Jacobs RR. Supervised construct scoring to reduce personality assessment length: A field study and introduction to the short 10. Organ Res Methods 2024;27(2):223–64. http://dx.doi.org/10.1177/10944281221145694.
[37] Biau G, Scornet E. A random forest guided tour. TEST 2016;25:197–227. http://dx.doi.org/10.1007/s11749-016-0481-7.
[38] Qi Y. Random forest for bioinformatics. In: Zhang C, Ma Y, editors. Ensemble machine learning. New York, NY: Springer New York; 2012, p. 307–23. http://dx.doi.org/10.1007/978-1-4419-9326-7_11.
[39] Han S, Williamson BD, Fong Y. Improving random forest predictions in small datasets from two-phase sampling designs. BMC Med Inform Decis Mak 2021;21:322. http://dx.doi.org/10.1186/s12911-021-01688-3.
[40] Shaikhina T, Lowe D, Daga S, Briggs D, Higgins R, Khovanova N. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. Biomed Signal Process Control 2019;52:456–62. http://dx.doi.org/10.1016/j.bspc.2017.01.012.
[41] Zaklouta F, Stanciulescu B. Real-time traffic-sign recognition using tree classifiers. IEEE Trans Intell Transp Syst 2012;13:1507–14. http://dx.doi.org/10.1109/TITS.2012.2225618.
[42] Zhang Y, Haghani A. A gradient boosting method to improve travel time prediction. Transp Res Part C: Emerg Technol 2015;58:308–24. http://dx.doi.org/10.1016/j.trc.2015.02.019.
[43] Liu B, Xia Y, Yu PS. Clustering through decision tree construction. In: Proceedings of the ninth international conference on information and knowledge management. New York, NY, USA: ACM; 2000, p. 20–9. http://dx.doi.org/10.1145/354756.354775.
[44] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York, NY: Springer New York; 2009, http://dx.doi.org/10.1007/978-0-387-84858-7.
[45] Breiman L. Manual on setting up, using, and understanding random forests v3.1. 2002, p. 1–58.
[46] Tang F, Ishwaran H. Random forest missing data algorithms. Stat Anal Data Min: ASA Data Sci J 2017;10:363–77. http://dx.doi.org/10.1002/sam.11348.
[47] Hapfelmeier A, Hothorn T, Ulm K, Strobl C. A new variable importance measure for random forests with missing data. Stat Comput 2014;24:21–34. http://dx.doi.org/10.1007/s11222-012-9349-1.
[48] Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. J Amer Statist Assoc 2010;105:205–17. http://dx.doi.org/10.1198/jasa.2009.tm08622.
[49] Ishwaran H. Variable importance in binary regression trees and forests. Electron J Stat 2007;1(none). http://dx.doi.org/10.1214/07-ejs039.
[50] Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. Stat Anal Data Min: ASA Data Sci J 2011;4:115–32. http://dx.doi.org/10.1002/sam.10103.
[51] Becker W, Saisana M, Paruolo P, Vandecasteele I. Weights and importance in composite indicators: Closing the gap. Ecol Indic 2017;80:12–22. http://dx.doi.org/10.1016/j.ecolind.2017.03.056.
[52] Lu M, Sha Y, Silva TC, Colaprico A, Sun X, Ban Y, Wang L, Lehmann BD, Chen XS. LR hunting: A random forest based cell–cell interaction discovery method for single-cell gene expression data. Front Genet 2021;12. http://dx.doi.org/10.3389/fgene.2021.708835.
[53] Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E. Handbook on constructing composite indicators: Methodology and user guide. OECD; 2008, http://dx.doi.org/10.1787/9789264043466-en.
[54] Decancq K, Lugo MA. Weights in multidimensional indices of wellbeing: An overview. Econometric Rev 2013;32(1):7–34. http://dx.doi.org/10.1080/07474938.2012.690641.
[55] Mazziotta M, Pareto A. Principal component analysis for constructing socio-economic composite indicators: theoretical and empirical considerations. SN Soc Sci 2024;4:114. http://dx.doi.org/10.1007/s43545-024-00920-x.
[56] Somarriba N, Pena B. Synthetic indicators of quality of life in europe. Soc Indic Res 2009;94:115–33. http://dx.doi.org/10.1007/s11205-008-9356-y.
[57] Zhou P, Ang BW, Zhou DQ. Weighting and aggregation in composite indicator construction: a multiplicative optimization approach. Soc Indic Res 2010;96:169–81. http://dx.doi.org/10.1007/s11205-009-9472-3.
[58] Munda G, Nardo M. Constructing consistent composite indicators: The issue of weights, EUR 21834 EN. 2005. JRC32434. Technical Report, Joint Research Centre; 2005.
[59] Ram R. Composite indices of physical quality of life, basic needs fulfilment, and income: A 'principal component' representation. J Dev Econ 1982;11(2):227–47. http://dx.doi.org/10.1016/0304-3878(82)90005-0.
[60] Slottje DJ. Measuring the quality of life across countries. Rev Econ Stat 1991;73(4):684–93.
[61] Li T, Zhang H, Yuan C, Liu Z, Fan C. A PCA-based method for construction of composite sustainability indicators. Int J Life Cycle Assess 2012;17:593–603. http://dx.doi.org/10.1007/s11367-012-0394-y.
[62] Saisana M, Saltelli A, Tarantola S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. J R Stat Soc: Ser A (Statistics Society) 2005;168(2):307–23. http://dx.doi.org/10.1111/j.1467-985X.2005.00350.x.
[63] Jiménez-Fernández E, Sánchez A, Ortega-Pérez M. Dealing with weighting scheme in composite indicators: An unsupervised distance-machine learning proposal for quantitative data. Socio-Economic Plan Sci 2022;83:101339.
[64] Ray AK. Measurement of social development: an international comparison. Soc Indic Res 2007;86:1–46. http://dx.doi.org/10.1007/s11205-007-9097-3.
[65] Grupp H, Mogee ME. Indicators for national science and technology policy: how robust are composite indicators? Res Policy 2004;33(9):1373–84. http://dx.doi.org/10.1016/j.respol.2004.09.007, What do we know Innovation? Selected papers from an International Conference in honour of Keith Pavitt.
[66] Greco S, Ishizaka A, Tasiou M, Torrisi G. On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. Soc Indic Res 2019;141:61–94. http://dx.doi.org/10.1007/s11205-017-1832-9.
[67] Brans J, Vincke P. Note – a preference ranking organisation method. Manag Sci 1985;31:647–56. http://dx.doi.org/10.1287/mnsc.31.6.647.
[68] Roy B. The outranking approach and the foundations of electre methods. Theory and Decision 1991;31:49–73. http://dx.doi.org/10.1007/BF00134132.
[69] Alkire S, Foster J. Counting and multidimensional poverty measurement. J Public Econ 2011;95(7–8):476–87.
[70] Figueira JR, Greco S, Roy B, Słowiński R. An overview of ELECTRE methods and their recent extensions. J Multi-Criteria Decis Anal 2013;20(1–2):61–85. http://dx.doi.org/10.1002/mcda.1482.
[71] Brans J-P, Mareschal B. Multiple criteria decision analysis: State of the art surveys. New York: Springer-Verlag; 2005, p. 163–86. http://dx.doi.org/10.1007/0-387-23081-5_5.
[72] Munda G, Nardo M. Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. Appl Econ 2009;41(12):1513–23. http://dx.doi.org/10.1080/00036840601019364.
[73] Mazziotta M, Pareto A. Methods for constructing composite indices: One for all or all for one? Riv Ital di Econ Demogr E Stat 2013;67–80.
[74] De Muro P, Mazziotta M, Pareto A. Composite indices of development and poverty: An application to MDGs. Soc Indic Res 2011;104:1–18. http://dx.doi.org/10.1007/s11205-010-9727-z.
[75] Knaflic CN. Storytelling with data. Wiley; 2015, http://dx.doi.org/10.1002/9781119055259.
[76] Earle D, Hurley CB. Advances in dendrogram seriation for application to visualization. J Comput Graph Statist 2015;24:1–25.
[77] Gruvaeus G, Wainer H. Two additions to hierarchical cluster analysis. Br J Math Stat Psychol 1972;25:200–6. http://dx.doi.org/10.1111/j.2044-8317.1972.tb00491.x.
[78] Hahsler M, Hornik K, Buchta C. Getting things in order: An introduction to the r package seriation. J Stat Softw 2008;25. http://dx.doi.org/10.18637/jss.v025.i03.
[79] Chen C. Generalized association plots: information visualization via iteratively generated correlation matrices. Statist Sinica 2002;12:7–29.
[80] Friendly M, Kwan E. Effect ordering for data displays. Comput Statist Data Anal 2003;43:509–39. http://dx.doi.org/10.1016/S0167-9473(02)00290-6.
[81] Hurley CB. Clustering visualizations of multidimensional data. J Comput Graph Statist 2004;13:788–806. http://dx.doi.org/10.1198/106186004X12425.
[82] Ishwaran H, Kogalur U. Fast unified random forests for survival, regression, and classification (RF-SRC). 2023, URL , R package version 3.2.3.
[83] Bigaret S, Hodgett RE, Meyer P, Mironova T, Olteanu A-L. Supporting the multi-criteria decision aiding process: R and the mcda package. EURO J Decis Process 2017;5(1):169–94. http://dx.doi.org/10.1007/s40070-017-0064-1.
[84] Schneider B, Hanges PJ, Smith DB, Salvaggio AN. Which comes first: Employee attitudes or organizational financial and market performance? J Appl Psychol 2003;88:836–51. http://dx.doi.org/10.1037/0021-9010.88.5.836.
[85] Ng TWH, Feldman DC. How broadly does education contribute to job performance? Pers Psychol 2009;62:89–134. http://dx.doi.org/10.1111/j.1744-6570.2008.01130.x.

[86] Schaufeli WB, Bakker AB, Salanova M. The measurement of work engagement with a short questionnaire. Educ Psychol Meas 2006;66:701–16. http://dx.doi.org/10.1177/0013164405282471.

[87] Goldberg DP, Williams P. A user's guide to the General Health Questionnaire. NferNelson; 1988.

[88] Avallone F, Paplomatas A. Salute organizzativa. Psicologia del benessere nei contesti lavorativi. Raffaello Cortina Editore; 2005.

[89] Netemeyer RG, Boles JS, McMurrian R. Development and validation of work–family conflict and family–work conflict scales. J Appl Psychol 1996;81:400–10. http://dx.doi.org/10.1037/0021-9010.81.4.400.

[90] Grzywacz JG, Marks NF. Reconceptualizing the work–family interface: An ecological perspective on the correlates of positive and negative spillover between work and family. J Occup Health Psychol 2000;5:111–26. http://dx.doi.org/10.1037/1076-8998.5.1.111.

[91] Akkirman AD, Harris DL. Organizational communication satisfaction in the virtual workplace. J Manag Dev 2005;24:397–409. http://dx.doi.org/10.1108/02621710510598427.

[92] Scott SG, Bruce RA. Determinants of innovative behavior: a path model of individual innovation in the workplace. Acad Manag J 1994;37:580–607. http://dx.doi.org/10.2307/256701.

[93] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65. http://dx.doi.org/10.1016/0377-0427(87)90125-7.

[94] Lawson RG, Jurs PC. New index for clustering tendency and its application to chemical problems. J Chem Inf Comput Sci 1990;30:36–41. http://dx.doi.org/10.1021/ci00065a010.

[95] Liaw A, Wiener M. Classification and regression by randomforest. R News 2002;2(3):18–22, URL https://CRAN.R-project.org/doc/Rnews/.

[96] Probst P, Wright MN, Boulesteix A-L. Hyperparameters and tuning strategies for random forest. WIREs Data Min Knowl Discov 2019;9(3):e1301. http://dx.doi.org/10.1002/widm.1301.

[97] Boccuzzo G, Gianecchini M. Measuring Young graduates' job quality through a composite indicator. Soc Indic Res 2015;122(2):453–78.

[98] Stefana E, Marciano F, Rossi D, Cocca P, Tomasoni G. Composite indicators to measure quality of working life in europe: A systematic review. Soc Indic Res 2021;157:1047–78. http://dx.doi.org/10.1007/s11205-021-02688-6.

[99] Han H, Bai J, Yan J, Yang H, Ma G. A combined drought monitoring index based on multi-sensor remote sensing data and machine learning. Geocarto Int 2021;36:1161–77. http://dx.doi.org/10.1080/10106049.2019.1633423.

[100] Hanadé Houmma I, Gadal S, Mansouri LE, Garba M, Gbetkom PG, Barkawi MBM, Hadria R. A new multivariate agricultural drought composite index based on random forest algorithm and remote sensing data developed for sahelian agrosystems. Geomatics, Nat Hazards Risk 2023;14. http://dx.doi.org/10.1080/19475705.2023.2223384.

[101] Rogers M, Bruen M. Choosing realistic values of indifference, preference and veto thresholds for use with environmental criteria within ELECTRE. European J Oper Res 1998;107(3):542–51. http://dx.doi.org/10.1016/S0377-2217(97)00175-6.

[102] Liao H, Wu X, Mi X, Herrera F. An integrated method for cognitive complex multiple experts multiple criteria decision making based on ELECTRE III with weighted borda rule. Omega 2020;93:102052. http://dx.doi.org/10.1016/j.omega.2019.03.010.

[103] Leclercq-Vandelannoitte A. Is employee technological "ill-being"' missing from corporate responsibility? The foucauldian ethics of ubiquitous IT uses in organizations. J Bus Ethics 2019;160:339–61. http://dx.doi.org/10.1007/s10551-019-04202-y.

[104] Puranik H, Koopman J, Vough HC. Pardon the interruption: An integrative review and future research agenda for research on work interruptions. J Manag 2020;46(6):806–42. http://dx.doi.org/10.1177/0149206319887428.

[105] Stich J-F, Tarafdar M, Cooper CL. Electronic communication in the workplace: boon or bane? J Organ Eff: People Perform 2018;5(1):98–106. http://dx.doi.org/10.1108/JOEPP-05-2017-0046.

[106] Liu X, Wu D, Zewdie GK, Wijerante L, Timms CI, Riley A, Levetin E, Lary DJ. Using machine learning to estimate atmospheric ambrosia pollen concentrations in tulsa, OK. Environ Health Insights 2017;11. http://dx.doi.org/10.1177/1178630217699399.

[107] Hislop D, Axtell C, Collins A, Daniels K, Glover J, Niven K. Variability in the use of mobile ICTs by homeworkers and its consequences for boundary management and social isolation. Inf Organ 2015;25(4):222–32. http://dx.doi.org/10.1016/j.infoandorg.2015.10.001.

[108] van Zoonen W, Sivunen AE. The impact of remote work and mediated communication frequency on isolation and psychological distress. Eur J Work Organ Psychol 2022;31(4):610–21. http://dx.doi.org/10.1080/1359432X.2021.2002299.

[109] Lee D-J, Joseph Sirgy M. Work-life balance in the digital workplace: The impact of schedule flexibility and telecommuting on work-life balance and overall life satisfaction. In: Coetzee M, editor. Thriving in digital workspaces: emerging issues for research and practice. Cham: Springer International Publishing; 2019, p. 355–84. http://dx.doi.org/10.1007/978-3-030-24463-7_18.

[110] Lane JN, Leonardi PM, Contractor NS, DeChurch LA. Teams in the digital workplace: Technology's role for communication, collaboration, and performance. Small Group Res 2024;55(1):139–83.

[111] Wang B, Liu Y, Qian J, Parker SK. Achieving effective remote working during the COVID-19 pandemic: A work design perspective. Appl Psychol 2021;70(1):16–59.

[112] Sun N, Li Y, Ning Y. Organic coordination in project design: Complemented or substituted by standardization? Int J Proj Manage 2025;102738.

[113] Bernhardt A, Kresge L, Suleiman R. The data-driven workplace and the case for worker technology rights. ILR Rev 2023;76:3–29. http://dx.doi.org/10.1177/00197939221131558.

[114] Sánchez-Monedero J, Dencik L. The datafication of the workplace.. 2019, URL https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-The-datafication-of-the-workplace.pdf.

[115] Gal U, Jensen TB, Stein M-K. Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. Inf Organ 2020;30:100301. http://dx.doi.org/10.1016/j.infoandorg.2020.100301.

[116] Constantiou ID, Kallinikos J. New games, new rules: Big data and the changing context of strategy. J Inf Technol 2015;30:44–57. http://dx.doi.org/10.1057/jit.2014.17.

[117] Tonidandel S, King EB, Cortina JM. Big data methods. Organ Res Methods 2018;21:525–47. http://dx.doi.org/10.1177/1094428116677299.

[118] Louppe G. Understanding random forests: From theory to practice. 2014, arXiv:1407.7502.

[119] Figueira J, Greco S, Ehrogott M. Multiple criteria decision analysis: State of the art surveys. vol. 78, New York, NY: Springer New York; 2005, http://dx.doi.org/10.1007/b100605.

Fabio Demaria, PhD Postdoctoral researcher and lecturer in Economic Statistics at the University of Modena and Reggio Emilia. His research interests lie in statistical learning, stochastic models, and the application of statistics to sustainability-related topics. His current work focuses on the development and use of statistical models to analyze and visualize complex data structures – such as hierarchical and tensor data – and on regime-switching modeling for non-normal distributions.

Ariadni Papana Dagiasis, PhD Associate Professor in the Department of Mathematics & Statistics at Cleveland State University. She earned her Ph.D. in Statistics from Case Western Reserve University in 2008, with a dissertation titled "Tools for Comprehensive Analysis of Microarray Data". Her research interests include high-dimensional data analysis, especially in genomics and proteomics, variable selection, dimension reduction, graphical methods, ensemble learning, GenAI and statistics education. Her work has appeared in international journals such as Bioinformatics, Operations & Supply Chain Management, Journal of Modern Applied Statistical Methods, among others.

Maddalena Cavicchioli, PhD Associate Professor of Economic Statistics at the University of Modena and Reggio Emilia, Italy. Her research interests are in the areas of statistics and econometrics, with particular focus on nonlinear models. She has published in high-profile international journals such as Econometric Theory, Journal of Financial Econometrics, Journal of Business & Economic Statistics, Scandinavian Journal of Statistics, Journal of Multivariate Analysis, Journal of Statistical Planning and Inference, Electronic Journal of Statistics, among others.

Tommaso Fabbri, PhD Full Professor of Organization and Human Resources Management and Dean of the Department of Economics at the University of Modena and Reggio Emilia. Scientific Coordinator and Board Member at Fondazione Marco Biagi, and faculty member at Bologna Business School. Member of Associazione Italiana di Organizzazione Aziendale (ASSIOA). He has been Visiting Professor at Pennsylvania State University, Smeal College of Business Administration. His interdisciplinary research, grounded in a processual view of organization, focuses on change, learning, and well-being. He also consults on organizational design, HR management, and digital transformation for private companies and public institutions.