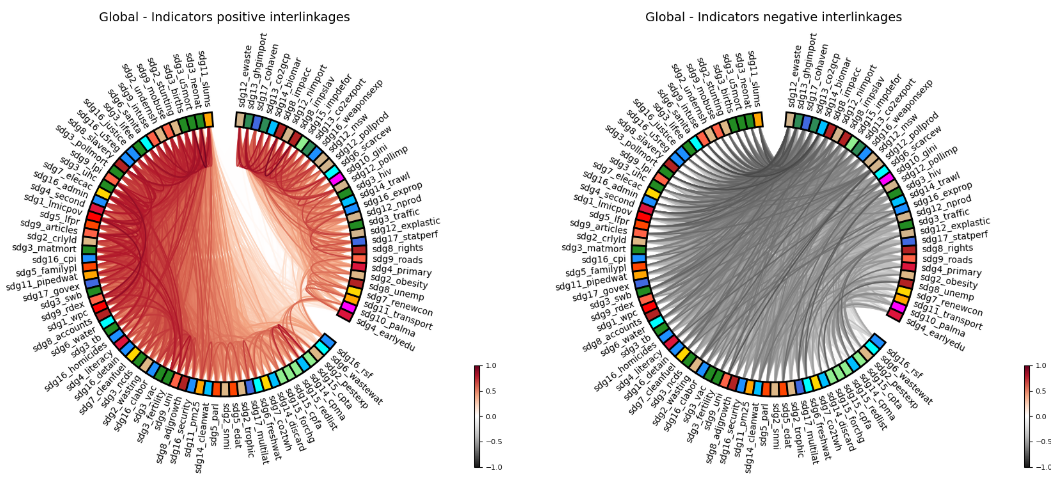
Graphical Abstract

A consistent and integrated network analysis framework for decoding interlinkages between sustainable development indicators at the global scale

Julien Ah-Pine, Elda Nasho Ah-Pine



Highlights

A consistent and integrated network analysis framework for decoding interlinkages between sustainable development indicators at the global scale

Julien Ah-Pine, Elda Nasho Ah-Pine

- Population-weighted and region-specific Kendall rank correlation for measuring SDG indicators interlinkages at the global scale.
- Integrated use of graph clustering and eigenvector centrality for consistent and interpretable SDG indicators network analysis.
- Novel visualization of SDG indicators synergies and trade-offs through structured chord diagrams.
- Decision-support tools combining network analysis and Pareto front selection to inform actionable policy design.

A consistent and integrated network analysis framework for decoding interlinkages between sustainable development indicators at the global scale

Julien Ah-Pine[a,b], Elda Nasho Ah-Pine[c]

[a] *Université Clermont Auvergne, Clermont Auvergne INP, Mines Saint-Etienne, CNRS, LIMOS, F-63000 Clermont-Ferrand, France,*
[b] *Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France,*
[c] *Université Clermont Auvergne, Clermont School of Business, IAE Clermont Auvergne, CLERMA, F-63000 Clermont-Ferrand, France,*

Abstract

Achieving the United Nations Sustainable Development Goals (SDG) is crucial to addressing urgent global challenges such as poverty, inequality, environmental degradation, and climate change. The SDG aim to foster a more equitable, resilient, and sustainable future, yet their interdependent nature means that progress in one area often influences others, creating complex patterns of synergies and trade-offs. Navigating these systemic dynamics requires robust, data-driven methods capable of capturing interactions at a granular level. This paper proposes a consistent and integrated framework for analyzing SDG indicators interlinkages at the global scale. We extend the Kendall rank correlation by incorporating population weighting and regional specificities, constructing a consistent signed weighted network of indicators. An optimal clustering method, aligned with eigenvector centrality, identifies structural groupings and systemic leverage points. We introduce a bi-criteria Pareto front selection to prioritize indicators based on influence and urgency, and an innovative visualization tool that reorganizes network diagrams for greater interpretability. Applying the framework to the SDR 2024 dataset reveals actionable synergies and tensions, particularly emphasizing the roles of governance quality, environmental management, and urban infrastructure. Our approach provides a coherent toolset for policymakers seeking to design integrated, effective interventions that balance development and sustainability goals.

Keywords: Networks, SDG interlinkages, Kendall rank correlation, Graph clustering, Eigenvector centrality

1. Introduction

The 2030 Agenda for Sustainable Development, adopted by all United Nations (UN) member states in 2015, outlines 17 Sustainable Development Goals (SDG) and 169 targets aimed at balancing social, economic, and environmental progress [55]. By design, the SDG are integrated and indivisible: achieving progress in one area often depends on simultaneous advancements in others [34, 52]. For instance, synergies between health (sdg3) and education (sdg4) can accelerate overall development, while uncoordinated efforts such as expanding industry (sdg9) without environmental safeguards, may hinder climate goals (sdg13). Effectively leveraging such synergies and managing trade-offs is therefore essential to fulfilling the Agenda's ambition [45, 6, 61, 53, 7].

Despite widespread commitment, progress remains insufficient. As the 2030 deadline approaches, less than around 20% of SDG targets are on track, with systemic setbacks exacerbated by global crises including pandemics, conflicts, economic instability, and climate disasters [49]. The slow and uneven trajectory demands not only accelerated action but also a shift toward coordinated, system-aware strategies. Recent research increasingly emphasizes that siloed, sectoral interventions are inadequate to address the systemic complexity of sustainable development [47, 27, 31, 41, 45]. Consequently, there is a growing consensus on the necessity to map and quantify interlinkages among SDG indicators to guide more coherent policy choices.

Addressing the complexity of global sustainability challenges requires tools that can support integrated, data-driven decision-making [27, 16, 53, 23, 31, 41, 45, 20]. Rigorous analysis of SDG interactions can reveal leverage points for maximizing positive spillovers and preempt unintended cross-sectoral trade-offs [44, 7]. However, navigating these complex relationships and achieving the SDG is still a daunting task, requiring a deep understanding of the synergies and trade-offs between the different goals and targets [7]. Furthermore, the complexity and structure of interlinkages between SDG indicators are inherently context-dependent, reflecting variations in socioeconomic, environmental, and institutional conditions across different geographies [33, 7].

Although SDG interlinkages are context-specific, global analyses are still essential as they offer a valuable starting point for decision-makers, reveal persistent patterns of synergies and trade-offs, and facilitate cross-context learning to accelerate progress toward the 2030 Agenda [9]. By identifying recurring interlinkages at the global level, we can better anticipate systemic dynamics, inform strategic priority-setting, and foster more coherent, scalable solutions that localized analyses alone may overlook. Moreover, because sustainability challenges are inherently global and countries are increasingly interconnected, the actions or inactions of one nation can generate significant spillover effects across borders, making it crucial to understand SDG interlinkages not only within local contexts but also through a global lens.

Accordingly, this paper aims to contribute to the global governance of the SDG. We propose a framework for investigating interlinkages between indicators in an integrated and consistent manner, utilizing complementary analytical methods while accounting for regional specificities. Our approach offers a comprehensive and systematic approach to understanding the relationships between different SDG, with a particular focus on identifying potential synergies and trade-offs between indicators. It is characterized by a set of mutually coherent and interconnected technical features, including:

- An innovative extension of Kendall correlation measure to assess the interlinkages between SDG indicators taking into account regional dependencies and population sizes.
- The application of an optimal clustering method to discover groups of indicators that are internally in positive relationships and mutually in negative relationships.
- The utilization of eigenvector centrality measure to detect which indicators are more important and influential in the SDG indicators network.
- The integration of the preceding clustering and centrality methods, shown to stem from the same optimization problem, to provide a more robust understanding of the SDG interlinkages.
- The development of an innovative visualization tool that leverages the information extracted by clustering and centrality measures to facilitate the interpretation of results.
- The use of the Pareto front to select in each cluster the optimal set of SDG indicators to focus on, ensuring that the most important and impactful indicators are prioritized.

This proposed framework aims to advance the understanding of the sustainability phenomenon and to empower decision-makers and stakeholders at global level with innovative tools to design integrated and actionable international policies that balance competing priorities and maximize positive impacts.

The remainder of this paper is structured as follows: we first review the existing literature on SDG interlinkages in the next Section 2, followed by a detailed presentation of the proposed framework in Section 3, its application to a case study with a discussion of the results and implications in Section 4, and finally, a discussion of the limitations of our work and of future research directions in Section 5.

2. Previous works

2.1. SDG interlinkages analysis

Since the launch of the 2030 Agenda in 2015, research on SDG interconnections has grown rapidly, with many studies examining the complex relationships among the 17 goals. The UN emphasizes the principles of universality, indivisibility, and interdependence, highlighting that all SDG are equally important, closely linked, and must be pursued holistically to avoid progress in one area undermining another [34]. Given this complexity, diverse methodological approaches have emerged to address the multidimensional challenges involved. These methods can be grouped into the following categories [26, 7]:

- Data analysis: examination of quantitative and/or qualitative data along with statistical and/or network analysis to identify and categorize relationships.
- Expert judgment: involves leveraging the insights and assessments of specialists, often gathered through surveys and structured interviews, to evaluate the interconnections.
- Modelling: use of computational and mathematical frameworks to simulate and assess the relationships and dynamics among different sectors to anylyze interactions in complex systems.
- Literature review: synthesizes existing research findings and data from multiple studies to identify patterns and trends among the interlinkages.

Research on SDG interlinkages has also considered various settings [26, 9, 7], including the analysis of interactions:

- at different levels, focusing on indicators, targets, or goals, which we collectively refer to as variables subsequently;
- at distinct spatial scales looking at national, regional or global tiers;
- focusing on a subset of goals such as the water-energy-food nexus;
- with additional external factors like climate.

This paper focuses on data-driven methods that rely on statistical data from official sources, using correlation and network analysis techniques, as illustrated in [46, 19, 30]. These approaches quantify the strength of relationships among variables, enabling the identification of specific synergies and trade-offs. Graph clustering methods help automatically detect groups of variables that form synergy poles, revealing meaningful structures within the complexity of sustainable development. In addition, network analysis can identify central nodes, which represent key areas for intervention that may not be evident through traditional methods. These techniques also enhance the visualization of complex interlinkages, making it easier to understand how multiple variables interact simultaneously. Although other analytical approaches, such as Multiple Factor Analysis [14], are of interest, they fall outside the scope of this paper.

Below, we provide a short review of the main tools used in this paper, including statistical association measures, clustering, and centrality measures.

2.2. Association measures

Research on SDG interlinkages often draws on statistical association measures to assess relationships between variables (indicators, targets or goals). Pearson and Spearman correlation coefficients are commonly used to quantify the direction and strength of linear or monotonic relationships among variables. These pairwise associations can be represented as weighted networks. While Pearson is used in studies such as [35, 54, 61], others prefer Spearman, as in [45, 31, 25, 5].

However, Pearson coefficient is sensitive to outliers, assumes linearity, and cannot detect non-linear patterns. Spearman measure is more robust and detects monotonic trends but is limited by sensitivity to ties and its assumption of monotonicity.

To capture more complex, non-linear, and non-parametric relationships, alternative measures such as distance correlation (dCor) [33], maximal information coefficient (MIC) [59], and revealed comparative advantage (RCA) [28] have been introduced.

Furthermore, association-based analyses of SDG interlinkages generally follow two approaches: longitudinal and cross-sectional [46]. Longitudinal studies examine how indicators evolve over time within a country or region, identifying dynamic dependencies and potential causal links [45, 54, 33]. Cross-sectional studies, by contrast, compare interactions across countries at a single time point, revealing patterns of co-development, regional disparities, and structural dependencies [31, 25, 59, 28].

2.3. Graph clustering

Graph clustering is a key tool in SDG interlinkages analysis, helping to uncover structural and functional groupings among goals, targets, or indicators [60]. By identifying tightly connected clusters, also referred to as communities or nexuses, clustering supports a more systemic understanding of synergies and trade-offs across domains such as the environment, economy, and social equity.

Among various methods, modularity-based clustering [40] is the most widely used [60, 54, 33, 28]. Techniques like the Louvain algorithm [11] optimize modularity by maximizing within-cluster connections and minimizing links between them, effectively revealing cohesive and interdependent variable groupings.

Spectral clustering, which uses eigenvalue decomposition of the graph Laplacian [57, 39], allows for detecting clusters of arbitrary shapes and was applied, for example, in [35]. In contrast, standard k-means is less suitable for SDG networks due to its rigid assumptions about cluster number and shape.

2.4. Centrality measures

Centrality measures quantify the influence of nodes based on their position in a network and are widely used across disciplines [17, 10]. In SDG interlinkages analysis, they help identify the most impactful goals, targets, or indicators [18, 19], guiding policymakers in prioritizing high-leverage actions.

Different measures capture different aspects of influence. Degree centrality identifies variables with the most direct connections [44], while betweenness centrality highlights those acting as bridges between otherwise disconnected goals [28]. Eigenvector centrality, by contrast, considers not only a node's connections but also the importance of its neighbors, offering a more nuanced view of systemic influence [19].

Although betweenness can reveal bottlenecks, eigenvector centrality is generally preferred in SDG analysis due to its ability to identify structurally dominant variables and its greater stability under network changes [12].

2.5. Limitation of current approaches

To summarize the previous points, Figure 1 provides a generic overview of commonly used network analysis tools and a general workflow for studying SDG interlinkages.

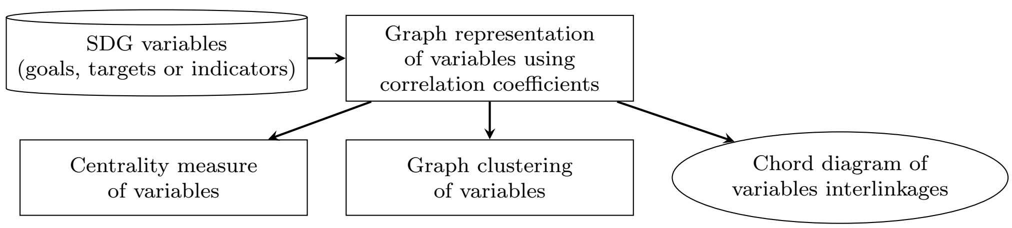


Figure 1: Conventional network analysis procedures for SDG interlinkages.

Within this scope, we identify the following research gaps:

- Existing association measures often lack interpretability for policymakers and overlook key factors such as population size and regional dependencies, limiting global representativity.
- Modularity is widely used in graph clustering but is not the only relevant criterion. Modularity-based algorithms often rely on heuristics, introducing non-determinism and inconsistent community structures.
- While eigenvector centrality has been interpreted for directed signed graphs [18, 19], a clear understanding for undirected networks in SDG analysis, is still missing.
- Clustering and centrality are typically applied independently, resulting in fragmented insights. Their integration is needed for a coherent understanding of complex systems.
- Chord diagrams, though common, become cluttered with many variables or edges, limiting their utility for interpreting complex SDG interlinkages.

- Despite increasing research, there is still no comprehensive, operational framework that enables policymakers to design actionable, context-specific interventions.

In the following section, we introduce our methodology that aims to address the aforementionned issues.

3. Proposed framework

3.1. Overview

Our proposal involves employing complementary tools and methods that ultimately provide a consistent and integrated framework for a robust and enriched analysis of SDG interconnections at a global scale. In Figure 2, we depict the different components of our approach, emphasizing in bold and red the specificities of our methodology in comparison to the base workflow exposed in Figure 1.

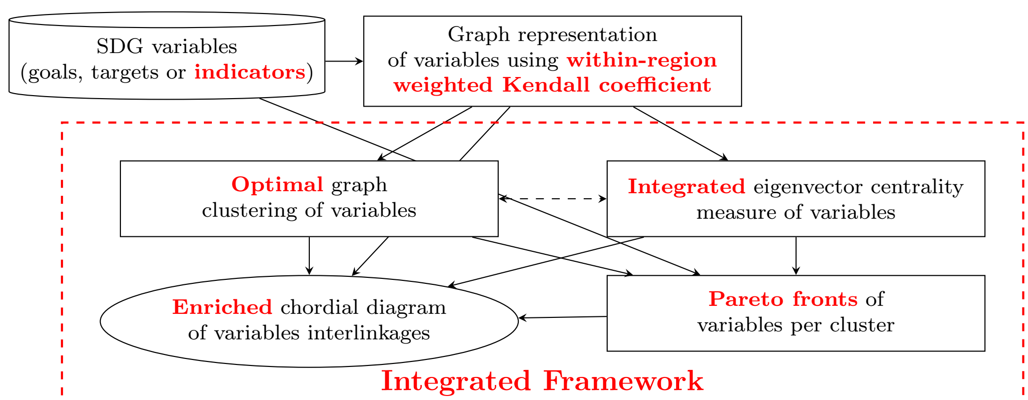


Figure 2: Our consistent and integrated framework for SDG interlinkages analysis and visualization.

In a nutshell, the main features of our approach are the following ones:

- We study interconnections between SDG at the level of indicators rather than at the level of goals or targets. This approach provides a more granular and precise analysis, allowing for a better understanding of synergies and trade-offs, ultimately informing more actionable policy design.

- We apply a cross-sectional approach that emphasizes the comparison between countries at a given point in time, with the goal of extracting global structural dependencies among indicators.
- We utilize Kendall rank coefficient to assess the type and strength of the relationship between each pair of indicators. Moreover, we take into account the population size of each country as a weight to ensure that the results are representative of the majority of the world's population.
- We target global SDG governance, but it is essential to ensure that regional specificities align with the global strategy. We address this concern by using an extension of the weighted Kendall association measure that accounts for the regional dependencies between countries, thereby enhancing the consistency between global and regional analyses.
- We apply an optimal clustering model to the consistent within-region weighted Kendall correlation network to discover clusters of indicators that show strong intra-synergies and strong inter-trade-offs simultaneously. This clustering approach does not require the pre-setting of the number of clusters.
- We use the principal eigenvector of the within-region weighted Kendall correlation graph as a centrality measure. This approach is fully aligned with the aforementioned clustering procedure. We show that the leading eigenvector is the solution to a relaxed version of this clustering problem.
- In our proposal, the clusters and centrality measures are closely associated, and we suggest using these outputs to represent the indicators in a chord diagram in a structured manner. This feature provides an innovative and informative visualization of the interactions between SDG.
- Studying the interactions at the indicator level allows for a fine-grained analysis, but the large number of indicators can be cumbersome for decision-makers. As a result, we propose determining the Pareto front in each cluster by considering two criteria: the percentage of performance (or achievement) and the eigenvector centrality measure.

In the following paragraphs, we formally introduce and further explain the motivations behind each component of our integrated framework, as well as provide additional details on the complementarities they share.

3.2. Within-region weighted Kendall rank correlation

We use the Kendall rank correlation to assess the direction and strength of association between SDG indicators. Unlike Pearson correlation, Kendall coefficient is non-parametric and captures general monotonic relationships. It offers an intuitive interpretation by comparing the number of concordant and discordant pairs, rather than rank differences as in Spearman measure, making it more accessible for decision-makers.

To account for global priorities, we introduce a population-weighted version of Kendall correlation. This approach ensures that indicator associations more accurately represent the current global landscape by giving greater weight to countries with larger populations, thus supporting more equitable and impactful policy decisions.

Let X^j and $X^{j'}$ be two indicators, measured across n countries. Denote their value vectors as $\mathbf{x}^j = (x_1^j, \ldots, x_{\mathrm{n}}^j)$ and $\mathbf{x}^{j'} = (x_1^{j'}, \ldots, x_{\mathrm{n}}^{j'})$. For two countries X_i and $X_{i'}$, we introduce the following definitions:

- $(X_i, X_{i'})$ is concordant if $(x_i^j < x_{i'}^j \wedge x_i^{j'} < x_{i'}^{j'})$ or $(x_i^j > x_{i'}^j \wedge x_i^{j'} > x_{i'}^{j'})$. This indicates a positive association between indicators X^j and $X^{j'}$.
- $(X_i, X_{i'})$ is discordant if $(x_i^j < x_{i'}^j \wedge x_i^{j'} > x_{i'}^{j'})$ or $(x_i^j > x_{i'}^j \wedge x_i^{j'} < x_{i'}^{j'})$: they are ranked in opposite orders, suggesting a negative correlation.

Then, to compare all $\frac{\mathrm{n}(\mathrm{n}-1)}{2}$ country pairs, we use the following quantities:

- $\mathrm{Conc}(\mathbf{x}^j, \mathbf{x}^{j'})$ denotes the total number of concordant pairs.
- $\mathrm{Disc}(\mathbf{x}^j, \mathbf{x}^{j'})$ denotes the total number of discordant pairs.

The basic Kendall τ_a coefficient (no ties) is given by:

$$\tau_a(\mathbf{x}^j, \mathbf{x}^{j'}) = \frac{\mathrm{Conc}(\mathbf{x}^j, \mathbf{x}^{j'}) - \mathrm{Disc}(\mathbf{x}^j, \mathbf{x}^{j'})}{\frac{\mathrm{n}(\mathrm{n}-1)}{2}}. \tag{1}$$

τ_a ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation). To account for ties, we introduce the following definitions:

- $(X_i, X_{i'})$ is tied in $\mathbf{x}^j$ if $x_i^j = x_{i'}^j$.
- $\mathrm{Tied}(\mathbf{x}^j)$ denotes number of tied pairs in $\mathbf{x}^j$.

Accordingly, the adjusted Kendall τ_b is:

$$\tau_b(\mathbf{x}^j, \mathbf{x}^{j'}) = \frac{\mathrm{Conc} - \mathrm{Disc}}{\sqrt{\left(\frac{\mathrm{n}(\mathrm{n}-1)}{2} - \mathrm{Tied}(\mathbf{x}^j)\right)\left(\frac{\mathrm{n}(\mathrm{n}-1)}{2} - \mathrm{Tied}(\mathbf{x}^{j'})\right)}}. \tag{2}$$

Let $\mathbf{S}^j = (s^j_{ii'})$ be the sign matrix of $\mathbf{x}^j$, where:

$$s^j_{ii'} = \begin{cases} -1, & x^j_i < x^j_{i'}, \\ 0, & x^j_i = x^j_{i'}, \\ 1, & x^j_i > x^j_{i'}. \end{cases} \tag{3}$$

Then τ_b can be rewritten as:

$$\tau_b(\mathbf{S}^j, \mathbf{S}^{j'}) = \frac{\sum_{i,i'} s^j_{ii'} s^{j'}_{ii'}}{\sqrt{\sum_{i,i'} (s^j_{ii'})^2 \sum_{i,i'} (s^{j'}_{ii'})^2}}. \tag{4}$$

In this paper, we aim to weight the concordance and discordance counts with respect to population sizes. Weighted versions of the Kendall coefficient have been studied by several papers [50, 2, 56]. We opt for the following definition denoted by τ_w:

$$\tau_w(\mathbf{S}^j, \mathbf{S}^{j'}) = \frac{\sum_{i,i'=1}^{\mathrm{n}} w_i w_{i'} s^j_{ii'} s^{j'}_{ii'}}{\sqrt{\sum_{i,i'=1}^{\mathrm{n}} w_i w_{i'} (s^j_{ii'})^2 \sum_{i,i'=1}^{\mathrm{n}} w_i w_{i'} (s^{j'}_{ii'})^2}}, \tag{5}$$

where, for all $i = 1, \ldots, \mathrm{n}$, w_i is the positive weight of X_i.

The weighted Kendall τ_w coefficient lies in the interval $[-1, 1]$, similar to Kendall τ_a and τ_b, and shares the same interpretations [56].

Next, we extend τ_w in order to integrate regional specificities and make our linkage measure context-dependent. When measuring the interconnection between two indicators X^j and $X^{j'}$ using the weighted Kendall association measure, we propose to consider only pairs of countries within the same region, rather than all possible country pairs. Our assumption is that assessing the probability of concordance and discordance between indicators is more meaningful within a shared regional context, as cross-regional comparisons lack interpretive relevance.

Suppose that the n countries are partitionned into m regions, $\mathbb{R} = \{\mathbb{R}_1, \ldots, \mathbb{R}_{\mathrm{m}}\}$. we define the within-region weighted Kendall coefficient denoted $\tau_{w,\mathbb{R}}$ by:

$$\tau_{w,\mathbb{R}}(\mathbf{S}^j, \mathbf{S}^{j'}) = \frac{\sum_{l=1}^{\mathrm{m}} \sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} s^j_{ii'} s^{j'}_{ii'}}{\sqrt{\sum_{l=1}^{\mathrm{m}} \sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} (s^j_{ii'})^2 \sum_{l=1}^{\mathrm{m}} \sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} (s^{j'}_{ii'})^2}}, \tag{6}$$

where the subscript $i, i' \in \mathbb{R}_l$ is shorthand for $i, i' = 1, \ldots, \mathrm{n}$, such that $X_i \in \mathbb{R}_l \wedge X_{i'} \in \mathbb{R}_l$.

Let us denote by $\mathbf{S}^{j,l} = (s^j_{ii'})_{i,i' \in \mathbb{R}_l}$, the block submatrix of size $|\mathbb{R}_l| \times |\mathbb{R}_l|$ extracted from $\mathbf{S}^j$ and containing only rows and columns corresponding to countries in region $\mathbb{R}_l$. Then, it is not difficult to see that (6) is similar to:

$$\tau_{w,\mathbb{R}}(\mathbf{S}^j, \mathbf{S}^{j'}) = \sum_{l=1}^{\mathrm{m}} \omega_l \tau_w(\mathbf{S}^{j,l}, \mathbf{S}^{j',l}), \tag{7}$$

where ω_l is the positive weight of region $\mathbb{R}_l$ defined by:

$$\omega_l = \frac{\sqrt{\sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} (s^j_{ii'})^2 \sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} (s^{j'}_{ii'})^2}}{\sqrt{\sum_{l=1}^{\mathrm{m}} \sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} (s^j_{ii'})^2 \sum_{l=1}^{\mathrm{m}} \sum_{i,i' \in \mathbb{R}_l} w_i w_{i'} (s^{j'}_{ii'})^2}}. \tag{8}$$

Accordingly, $\tau_{w,\mathbb{R}}(\mathbf{S}^j, \mathbf{S}^{j'})$ provides a measure of the direction and intensity of the interaction between X^j and $X^{j'}$ at the global level, while explicitly accounting for regional dependencies between countries. It can be interpreted as a linear combination over regions $\mathbb{R}_l$, for $l = 1, \ldots, \mathrm{m}$, of the weighted Kendall association measures $\tau_w(\mathbf{S}^{j,l}, \mathbf{S}^{j',l})$ defined in (5), where each term is weighted by ω_l, which is proportional to the total population of the countries within $\mathbb{R}_l$. We provide a concrete example in paragraph 4.2 to illustrate the rationale behind our context-informed global linkage measure.

In the following paragraphs, we assume a set of p SDG indicators $\{X^1, \ldots, X^{\mathrm{p}}\}$ with observed values over a set of countries $\{X_1, \ldots, X_{\mathrm{n}}\}$, represented by the set of vectors of size n, $\{\mathbf{x}^1, \ldots, \mathbf{x}^{\mathrm{p}}\}$. Our framework is built upon the interlinkages network represented by the signed weighted adjacency matrix $\mathbf{K} = (k_{jj'})$ of size $\mathrm{p} \times \mathrm{p}$, where each entry is defined by the within-region weighted Kendall coefficient given in (6). Specifically, $\forall j, j' = 1, \ldots, \mathrm{p}$:

$$k_{jj'} = \tau_{w,\mathbb{R}}(\mathbf{S}^j, \mathbf{S}^{j'}). \tag{9}$$

3.3. Mathematical Relational Analysis clustering

The within-region weighted Kendall correlation matrix $\mathbf{K}$ captures synergies and trade-offs as positive and negative values. These are represented as a complete graph, with indicators as nodes and signed edge weights reflecting correlation directions and strengths.

Figure 3 in page 19 presents a conventional visualization of the SDG indicator network. While informative, this dense representation with numerous nodes and edges reflects the complexity of sustainability without offering clear insights. To reveal underlying structures, we apply network analysis tools aimed at identifying two key patterns: positively interconnected clusters representing win-win relationships, and negatively linked groups indicating trade-offs requiring careful policy balancing.

To extract these patterns, we introduce a graph clustering method that explicitly incorporates both positive and negative associations. It searches for an optimal partition maximizing intra-cluster synergies and inter-cluster tensions. Unlike heuristic approaches, our method formulates the clustering as a binary integer linear program (0-1 ILP), ensuring rigorous and exact solutions.

Our approach builds on the Mathematical Relational Analysis (MRA) framework, originally developed in [38, 37], which combines graph theory, statistics, and optimization to model binary relations such as preferences and equivalences. In our case, the goal is to approximate a graph via an equivalence relation over its nodes. Related applications of MRA can be found in [36, 3, 15], and similar questions are also addressed in recent literature on correlation clustering [8, 21, 58].

In the MRA notations, a binary relation R over a set of p items is represented by a graph with an adjacency matrix, also called a relational matrix, $\mathbf{R} = (r_{jj'})$ of size p $\times$ p, with the general term:

$$r_{jj'} = \begin{cases} 1, & \text{if } X^j R X^{j'}, \text{ that is } X^j \text{ is in relation with } X^{j'}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

A clustering (or partition) of the nodes of the graph is exactly the same as an equivalence relation over the set of nodes. This type of binary relations satisfies the following properties:

- Reflexivity: $X^j R X^j$, for all $j = 1, \ldots, \mathrm{p}$.
- Symmetry: $X^j R X^{j'} \Leftrightarrow X^{j'} R X^j$, for all $j, j' = 1, \ldots, \mathrm{p}$, such that $j \neq j'$.
- Transitivity: $X^j R X^{j'} \wedge X^{j'} R X^{j''} \Rightarrow X^j R X^{j''}$, for all $j, j', j'' = 1, \ldots, \mathrm{p}$, such that $j \neq j' \neq j''$.

In [38], the authors show that all properties can be expressed as linear constraints of terms of the relational matrix $\mathbf{R}$:

- Reflexivity: $r_{jj} = 1$, for all $j = 1, \ldots, \mathrm{p}$.
- Symmetry: $r_{jj'} = r_{j'j}$, for all $j, j' = 1, \ldots, \mathrm{p}$, such that $j \neq j'$.

- Transitivity: $r_{jj'} + r_{j'j''} - r_{jj''} \leq 1$, for all $j, j', j'' = 1, \ldots, \mathrm{p}$, such that $j \neq j' \neq j''$.

We aim to approximate $\mathbf{K}$ using an unknown equivalence relation represented by a relational matrix $\mathbf{R}$, obtained by solving the following 0-1 ILP:

$$\max_{\mathbf{R}\in\{0,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'} r_{jj'} \tag{11}$$
$$\text{s.t.} \begin{cases} r_{jj} = 1, & \forall j = 1, \ldots, \mathrm{p}, \\ r_{jj'} = r_{j'j}, & \forall j, j' = 1, \ldots, \mathrm{p}; j \neq j', \\ r_{jj'} + r_{j'j''} - r_{jj''} \leq 1, & \forall j, j', j'' = 1, \ldots, \mathrm{p}; j \neq j' \neq j''. \end{cases}$$

Although clustering is NP-hard [4], the0-1 ILP in (11) enables efficient and exact solutions for problems with p in the hundreds, as in our case.

We further detail why Problem (11) enables the detection of indicator clusters that specifically reveal synergies and trade-offs within the SDG network. For simplicity, and without loss of generality, assume that there are no tied values in any of the vectors $\mathbf{x}^j$, $j = 1, \ldots, \mathrm{p}$. In this case, $k_{jj'}$ from (9) can be written as the difference between the estimated probability that X^j and $X^{j'}$ are concordant and the estimated probability that they are discordant:

$$k_{jj'} = \widehat{P}(\mathrm{Conc}(X^j, X^{j'})) - \widehat{P}(\mathrm{Disc}(X^j, X^{j'})). \tag{12}$$

Then, it is not difficult to see that:

$$\max_{\mathbf{R}\in\{0,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'} r_{jj'}$$
$$\Leftrightarrow \max_{\mathbf{R}\in\{0,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} \widehat{P}(\mathrm{Conc}(X^j, X^{j'})) r_{jj'} + \sum_{j,j'=1}^{\mathrm{p}} \widehat{P}(\mathrm{Disc}(X^j, X^{j'}))(1 - r_{jj'}).$$

Consequently Problem (11) is equivalent to:

$$\max_{\mathbf{R}\in\{0,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} \Big(\widehat{P}(\mathrm{Conc}(X^j, X^{j'})) r_{jj'} + \widehat{P}(\mathrm{Disc}(X^j, X^{j'}))(1 - r_{jj'}) \Big) \tag{13}$$
$$\text{s.t.} \begin{cases} r_{jj} = 1, & \forall j = 1, \ldots, \mathrm{p}, \\ r_{jj'} = r_{j'j}, & \forall j, j' = 1, \ldots, \mathrm{p}; j \neq j', \\ r_{jj'} + r_{j'j''} - r_{jj''} \leq 1, & \forall j, j', j'' = 1, \ldots, \mathrm{p}; j \neq j' \neq j''. \end{cases}$$

From this formulation, we observe the following:

- Indicators X^j and $X^{j'}$ are likely to be in the same cluster ($r_{jj'} = 1$) when $\widehat{P}(\text{Conc}(X^j, X^{j'}))$ exceeds $\widehat{P}(\text{Disc}(X^j, X^{j'}))$, reflecting strong positive associations and potential synergies.
- Conversely, when $\widehat{P}(\text{Disc}(X^j, X^{j'}))$ dominates, X^j and $X^{j'}$ tend to be assigned to different clusters ($r_{jj'} = 0$), indicating negative associations and trade-off dynamics.

3.4. Eigenvector centrality

In addition to clustering, centrality measures are other tools borrowed from network analysis to extract meaningful information from the SDG interactions graph. They allow for the exhibition of the relative importance of each indicator in the network, highlighting those that are most strongly connected to others and, thus, most likely to have a ripple effect on the achievement of multiple SDG. Moreover, they can reveal indicators that serve as crucial hubs or bridges between different clusters or poles, exerting significant influence on the overall network dynamics.

Let $\mathbf{K} = (k_{jj'})$ be a given correlation matrix between the indicators $\{X^1, \ldots, X^{\mathrm{p}}\}$. We denote by $\mathbf{v} = (v_j)$ the vector of size p that gives the centrality measure for each indicator X^j with $j = 1, \ldots, \mathrm{p}$. We focus on eigenvector centrality (or eigencentrality) which quantifies the global influence of a node X^j across the entire network structure. It is formally defined as follows for any indicator X^j:

$$v_j = \frac{1}{\lambda} \sum_{j'=1}^{\mathrm{p}} k_{jj'} v_{j'}, \tag{14}$$

where $\lambda > 0$ is interpreted as a scaling factor.

Eigenvector centrality is based on the idea that the importance of a node X^j, denoted by v_j, is determined by the importance of its neighboring nodes. Specifically, the contribution of each neighbor $X^{j'}$, with importance $v_{j'}$, is weighted by the strength of their connection, represented by $k_{jj'}$.

From a mathematical standpoint, the correlation matrix $\mathbf{K}$ being a real symmetric square matrix of size $\mathrm{p} \times \mathrm{p}$, it has p real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\mathrm{p}}$ associated to p eigenvectors $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^{\mathrm{p}}$ that are mutually orthogonals and by definition we have, $\forall j = 1, \ldots, \mathrm{p}$:

$$\mathbf{K}\mathbf{v}^j = \lambda_j \mathbf{v}^j.$$

Then, the eigenvector centrality $\mathbf{v}$ and the scaling factor λ in (14) are exactly the leading eigenvector and eigenvalue $\mathbf{v}^1$ and λ_1: $\mathbf{v} = \mathbf{v}^1$ and $\lambda = \lambda_1$.

In the context of SDG interlinkages, eigencentrality is typically applied to binary or non-negative weighted adjacency matrices, where a non-zero weight indicates a connection between nodes. In such cases, the Perron–Frobenius theorem ensures that both the leading eigenvalue λ_1 and its corresponding eigenvector $\mathbf{v}^1$ have non-negative entries, allowing for a straightforward ranking of nodes by importance. This approach, which yields non-negative eigencentrality scores, has been used in SDG correlation networks in [54, 33].

In the more general case where $\mathbf{K}$ contains both positive and negative values, it is not guaranteed that λ_1 and $\mathbf{v}^1$ contain non-negative values. In this situation, if some nodes have a negative centrality measure, the definition of eigencentrality provided by (14) becomes more difficult to interpret. In [13], the authors address this concern in the context of social network analysis and show that, in the case of a signed adjacency matrix with "like" $(+1)$ *versus* "dislike" (-1) relationships, the leading eigenvector $\mathbf{v}^1$ reveals two groups of nodes: those with positive values in $\mathbf{v}^1$, which are opposed to those with negative values, forming two contrasting cliques.

In our case, we apply eigenvector centrality to the within-region weighted Kendall correlation network $\mathbf{K}$ which includes both positive and negative values. As can be seen from (5), $\mathbf{K}$ is a Gram matrix and therefore its eigenvalues are non-negative. In what follows, we demonstrate that the leading eigenvector $\mathbf{v}^1$ corresponds to the solution of a relaxed version of Problem (11), where the number of clusters is fixed at two.

Let $\tilde{\mathbf{R}} = (\tilde{r}_{jj'})$ be a signed relational matrix associated to the binary relational matrix $\mathbf{R} = (r_{jj'})$ through the following relation, $\tilde{r}_{jj'} = 2r_{jj'} - 1$. Thus, we have, $\forall j, j' = 1, \dots,$ p:

$$\tilde{r}_{jj'} = \begin{cases} 1, & \text{if } X^j R X^{j'}, \\ -1, & \text{otherwise.} \end{cases} \tag{15}$$

If R is an equivalence relation, it is not difficult to see that the characteristic properties translate into similar linear constraints as those for $\mathbf{R}$ in the case of $\tilde{\mathbf{R}}$:

- Reflexivity: $\tilde{r}_{jj} = 1$, for all $j = 1, \dots,$ p.
- Symmetry: $\tilde{r}_{jj'} = \tilde{r}_{j'j}$, for all $j, j' = 1, \dots,$ p, such that $j \neq j'$.
- Transitivity: $\tilde{r}_{jj'} + \tilde{r}_{j'j''} - \tilde{r}_{jj''} \leq 1$, for all $j, j', j'' = 1, \dots,$ p, such that $j \neq j' \neq j''$.

Moreover, we have the following relation regarding the objective function of Problem (11):

$$\max_{\tilde{\mathbf{R}} \in \{-1,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'}\tilde{r}_{jj'} \Leftrightarrow \max_{\mathbf{R} \in \{0,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'}(2r_{jj'}-1) \Leftrightarrow \max_{\mathbf{R} \in \{0,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'}r_{jj'}.$$

As a consequence, Problem (11) is equivalent to:

$$\begin{aligned} &\max_{\tilde{\mathbf{R}} \in \{-1,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'}\tilde{r}_{jj'} \\ &\text{s.t. } \begin{cases} \tilde{r}_{jj} = 1, & \forall j = 1,\dots,\mathrm{p}, \\ \tilde{r}_{jj'} = \tilde{r}_{j'j}, & \forall j,j' = 1,\dots,\mathrm{p}; j \neq j', \\ \tilde{r}_{jj'} + \tilde{r}_{j'j''} - \tilde{r}_{jj''} \leq 1, & \forall j,j',j'' = 1,\dots,\mathrm{p}; j \neq j' \neq j''. \end{cases} \end{aligned} \tag{16}$$

Next, suppose that $\tilde{\mathbf{R}}$ encodes a partition with only two clusters. Then we can equivalently formulate $\tilde{\mathbf{R}}$ with respect to a vector $\mathbf{z} \in \{-1,1\}^{\mathrm{p}}$ as follows, $\forall j,j' = 1,\dots,\mathrm{p}$:

$$\tilde{r}_{jj'} = z_j z_{j'}, \tag{17}$$

where $z_j = 1$ if X^j is in the 1st group and $z_j = -1$ if it is in the 2nd group.

Note that through (17), vector $\mathbf{z}$ automatically implies a signed relational matrix $\tilde{\mathbf{R}}$ that satisfies the linear constraints of reflexivity, symmetry and transitivity in Problem (16). As a result, in the particular case of a partition with only two groups, we can reduce the representation of the equivalence relation from a matrix to a vector. This specific property makes it possible to simplify the MRA clustering problem as follows:

$$\begin{aligned} &\max_{\tilde{\mathbf{R}} \in \{-1,1\}^{\mathrm{p}\times\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'}\tilde{r}_{jj'} \text{ s.t. } \begin{cases} \tilde{r}_{jj} = 1, & \forall j, \\ \tilde{r}_{jj'} = \tilde{r}_{j'j}, & \forall j,j'; j \neq j', \\ \tilde{r}_{jj'} + \tilde{r}_{j'j''} - \tilde{r}_{jj''} \leq 1, & \forall j,j',j''; j \neq j' \neq j'', \\ \tilde{\mathbf{R}} \text{ encodes 2 clusters.} & \end{cases} \\ &\Leftrightarrow \max_{\mathbf{z} \in \{-1,1\}^{\mathrm{p}}} \sum_{j,j'=1}^{\mathrm{p}} k_{jj'} z_j z_{j'} \\ &\Leftrightarrow \max_{\mathbf{z} \in \{-1,1\}^{\mathrm{p}}} \mathbf{z}^\top \mathbf{K} \mathbf{z}. \end{aligned} \tag{18}$$

The discrete problem $\max_{\mathbf{z} \in \{-1,1\}^{\mathrm{p}}} \mathbf{z}^\top \mathbf{K}\mathbf{z}$ is NP-hard. A standard relaxation replaces the discrete domain with a continuous one: $\max_{\mathbf{z} \in \mathbb{R}^{\mathrm{p}}, \|\mathbf{z}\|^2 \leq 1} \mathbf{z}^\top \mathbf{K}\mathbf{z}$

where $\|\mathbf{z}\|$ is the Euclidean norm. It is well-known that the optimal solution of this convex optimization problem is $\mathbf{z} = \mathbf{v}^1$, the leading eigenvector of $\mathbf{K}$. This mathematical result aligns with the work presented in [13], which shows that eigenvector centrality $\mathbf{v} = \mathbf{v}^1$ provides, in the case of a signed network, a division of the nodes into two cliques.

Moreover, we can infer an importance order within each cluster by ranking their members with respect to their eigencentrality scores absolute values. Additionally, the two nodes with the most opposite centrality measures tend to be the most adversarial. Overall, eigenvector centrality applied to a signed weighted graph provides richer insights compared to an unsigned weighted network, especially for discovering the two main opposing poles.

It is important to mention that the previous development is reminiscent of several research works in spectral graph theory, and in data science for dimension reduction and clustering [24, 22, 57]. Nonetheless, from our perspective, this result allows us to innovate by integrating two typically separate network analysis tools into a unified framework for SDG interlinkages, enabling decision-makers to identify synergies, trade-offs, and key indicators for more coherent and actionable policy design.

3.5. Visualization tool

In Figure 3, we provide illustrations of chord diagrams of the SDG interconnections graph between a set of 96 indicators measured by our approach. The intensity of the edge color indicates the strength of the relationship given by the within-region weighted Kendall correlation measure, with darker colors representing stronger positive or negative links between the nodes. On the left hand side, only the positive correlation measures are illustrated while, on the right hand side, solely the negative association values are depicted.

However, in Figure 3, the indicators are organized in circle with respect to the 17 goals they belong to from sdg1 to sdg17. Clearly, from these graphs, it is difficult to infer any meaningful insights because there are too many edges that cross each other, creating a cluttered representation that obscures the underlying relationships.

In contrast, the connection we have established previously between MRA clustering and eigenvector centrality can serve as a framework to reorganize the indicators and provide a more informative visualization of the SDG interlinkages using chord diagrams:

- Firstly, we can group indicators based on the clusters identified by the

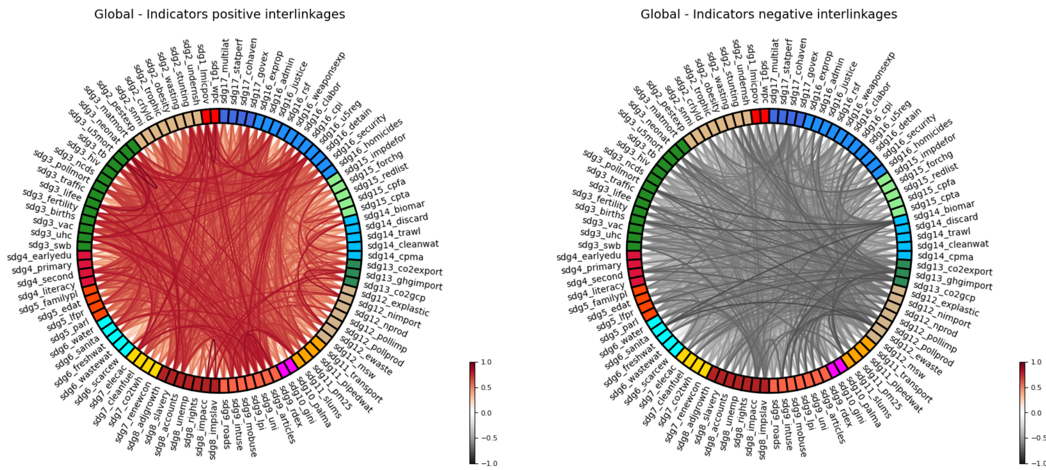


Figure 3: Chord diagrams of the SDG interlinkages of the 96 indicators organized with respect to their goals, with positive correlations on the left and, negative correlations on the right.

MRA clustering.

- Secondly, we can arrange the clusters in a circular order, starting from the top and moving counterclockwise, according to the mean eigencentrality measures.
- Thirdly, within each cluster, we can order the indicators according to their eigenvector centrality scores. As a consequence, indicators in the top left and the ones in the top right positions are those that tend to be the most in trade-off with each other.

Therefore, Figure 4 presents a reorganized view of the same graph shown in Figure 3, using the proposed clustering and eigencentrality ordering. The left diagram highlights two synergy clusters, with strong positive intraconnections and darker edges at the top indicating the most central indicators. The right diagram shows negative correlations, making trade-offs between the two clusters more visible, especially for top-ranked indicators positioned on opposite sides.

Overall, our approach uncovers meaningful groupings and rankings within the within-region weighted Kendall graph, enhancing the clarity of chord diagram visualizations. It highlights synergies and trade-offs among indicators

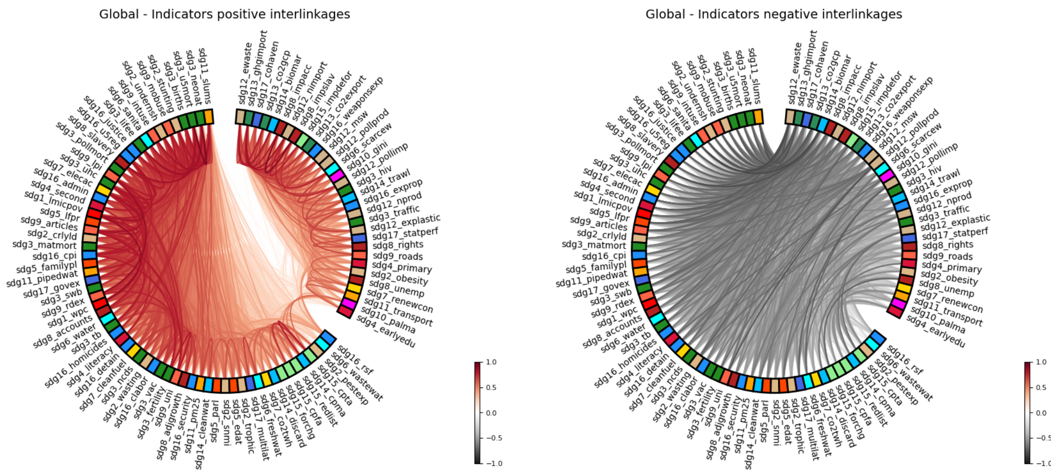


Figure 4: Chord diagrams of the SDG interlinkages of the 96 indicators organized with respect to their two clusters and eigenvector centrality scores provided by our integrated framework, with positive correlations on the left and, negative correlations on the right.

and goals, emphasizing the importance of integrated policy responses. More detailed interpretations based on Figure 4 are provided in Section 4.

3.6. Decision aiding using Pareto fronts

Our focus is on exploring the relationships between SDG through their individual indicators rather than through broader goals or targets. This detailed perspective enables us to gain a clearer insight into the synergies and trade-offs involved. A drawback of this approach however, is that decision-making necessitates the manipulation of numerous variables, which can be overwhelming and complicate the analysis process.

To address this issue, we propose focusing on two complementary criteria to select the indicators that decision-makers should prioritize. The first criterion U is the percentage of performance (or achievement) of indicators and the second criterion V is the eigenvector centrality measure that we already discussed in paragraph 3.4.

Below, we provide further details on the first criterion, the performance percentage U, as defined in [49, 32] and included in the SDR 2024 dataset. This dataset serves as the primary data source for our empirical analysis and supports the application of our integrated framework, described in Section 4.

For any numerical indicator X^j, we assume two thresholds: w^j and b^j, representing the worst- and best-case scenarios. Depending on the indicator, these may correspond to upper and lower bounds, or *vice versa*. For example, for $X^j =$ sdg3_neonat (Neonatal mortality rate), the worst case is at the upper bound ($w^j = 40/1000$) and the best at the lower bound ($b^j = 1.1/1000$). Conversely, $X^{j'} =$ sdg3_births (Births attended by skilled health personnel) has its worst case at the lower bound ($w^{j'} = 23\%$) and best at the upper bound ($b^{j'} = 100\%$).

The initial value of indicator X^j for country X_i, denoted d_i^j, is converted into a performance percentage x_i^j using the transformation defined in [49, 32]:

- First, the data distributions across countries are censored, meaning that any value d_i^j falling outside the interval $[\min(w^j, b^j), \max(w^j, b^j)]$ is set to the nearest bound.
- Next, the data are normalized to obtain a percentage of performance x_i^j that increases with proximity to the best-case scenario. For all indicators, higher values of x_i^j (closer to 100%) indicate better performance and greater progress toward achieving the target associated with indicator X^j. This normalization is given by, $\forall j = 1, \dots, \mathrm{p}$ and $\forall i = 1, \dots, \mathrm{n}$:

$$x_i^j = \frac{|d_i^j - w^j|}{|b^j - w^j|} 100. \tag{19}$$

Since our analysis is conducted at the global scale, the first criterion for indicator selection is based on the weighted mean of values across all countries, using population size as the weighting scheme.

Then, to facilitate a more focused and interpretable analysis, we reduce the number of indicators using the concept of the Pareto front. Widely used in multi-objective optimization, the Pareto front represents the set of non-dominated solutions that balance multiple criteria. In our case, we apply this concept from a bi-objective perspective, selecting indicators that are simultaneously influential and underperforming.

Specifically, in the case of the two criteria, U and V, a solution is said to be Pareto optimal if there is no other solution that improves U without degrading V, or *vice versa*. Let us denote $\mathbf{u} = (u_j)$ with $u_j = \sum_{i=1}^{\mathrm{n}} w_i x_i^j$, the vector of the weighted means of percentages of performance of all indicators and, $\mathbf{v} = (v_j)$ the vector of eigenvector centrality measures. Regarding U, an indicator X^j should be prioritized over another indicator $X^{j'}$ if $u_j < u_{j'}$, because in that case, X^j has a low level of advancement and requires more urgent progress. On the contrary, in the case of V, X^j should be prioritized over $X^{j'}$ if $v_j > v_{j'}$, since this indicates that X^j is more influential. Accordingly, we say that the indicator X^j is Pareto optimal if the two following conditions are verified:

$$\begin{cases} (u_j \leq u_{j'}) \wedge (v_j \geq v_{j'}), & \forall j' = 1, \ldots, \mathrm{p}, \text{ such that } j' \neq j, \\ (u_j < u_{j'}) \vee (v_j > v_{j'}), & \forall j' = 1, \ldots, \mathrm{p}, \text{ such that } j' \neq j. \end{cases} \tag{20}$$

By considering the percentage of performance, we direct our attention to indicators that require urgent upgrades, which aligns with the indivisible nature of the SDG as emphasized by the UN. Additionally, by incorporating the eigencentrality measure, we aim to support overall SDG progress by prioritizing indicators that exert significant influence across the network.

In our approach, we determine the Pareto front for each cluster rather than for the overall set of indicators. This procedure aligns with the structured insights provided by our MRA clustering and integrated eigenvector centrality, offering robustness and significant advantages. First, it allows for broader coverage of the main poles of synergies, capturing unique interconnections within each cluster. Second, this approach enhances our ability to identify critical trade-offs between clusters. By focusing on cluster specific Pareto fronts, we can effectively reduce the number of indicators, simplifying decision-making and enabling targeted strategies that optimize synergies while minimizing conflicts across the SDG.

In Figure 5, we present the SDG interlinkages chordal graphs shown in Figure 4, but restricted to the Pareto fronts of each cluster. In this case, the number of indicators is reduced from 96 to 13.

4. Application to the SDR 2024 dataset

4.1. The SDR 2024 dataset

Similarly to [31, 25, 61], we analyze SDG interlinkages using the dataset provided by the Sustainable Development Solutions Network (SDSN), a global initiative launched by the UN to promote practical solutions for sustainable development. The SDSN publishes an annual report, the Sustainable Development Report (SDR), which assesses the progress of countries toward achieving the SDG. This report provides data, analysis, and recommendations, highlighting best practices and identifying areas where further efforts are needed.

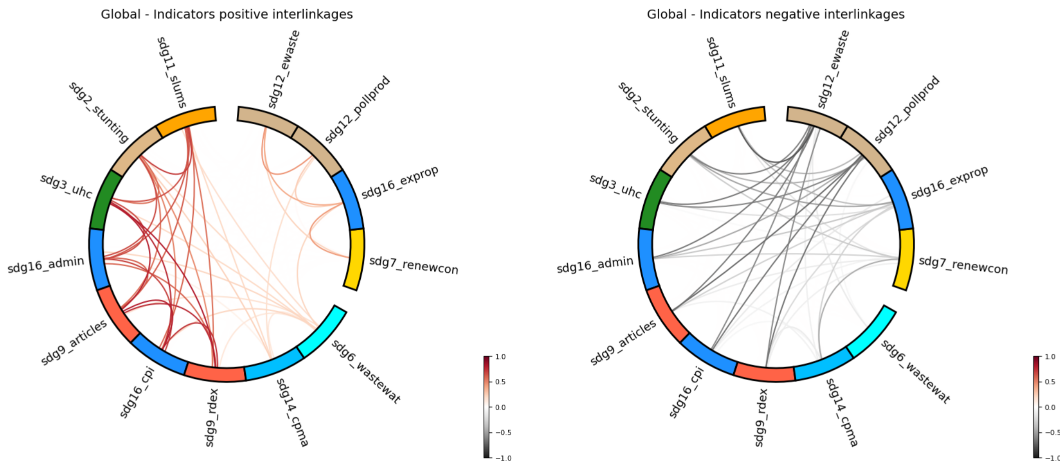


Figure 5: Chord diagrams of the SDG interlinkages of the indicators in the Pareto fronts organized with respect to their two clusters and eigenvector centrality scores provided by our integrated framework, with positive correlations on the left and, negative correlations on the right.

In this paper, we analyze the SDR 2024 dataset, which serves as the foundation for the SDSN report of 2024 [49]. This dataset is freely available at `https://dashboards.sdgindex.org/downloads`. We apply our consistent and integrated framework for global analysis. To integrate context-specific dependencies in our within-region weighted Kendall association measure given by (6), we considered the 7 regions as defined in the SDR 2024, which consist of the following: Eastern Europe & Central Asia, East & South Asia, Latin America and the Caribbean, Middle East and North Africa, Oceania, OECD members, Sub-Saharan Africa.

The SDR 2024 dataset countains 100 global indicators and 193 countries. We work with the censored and normalized version of these variables as described in paragraph 3.6 and defined by (19). Therefore, all indicators range from 0 to 100 and they are all in ascending order. These motivated data preprocessing steps performed by the SDSN ensure the homogeneity of variables, significantly enhancing the clarity and reproducibility of the analyses conducted. However, there are many missing values. We implemented the following extra procedures to address this issue:

- If an indicator has more than 40% missing values, it is dropped.

- Then, if a country has more than 30% missing values, it is removed.

After this first round of filtering, we were left with only two countries from the region of Oceania. We decided to remove these two cases for statistical consistency reasons. Consequently, our final dataset consists of 96 indicators and 165 countries, distributed across 6 regions. Descriptions of all indicators, along with their corresponding worst-case and best-case scenario thresholds, are provided in [49] and in the SDR 2024 dataset spreadsheet file.

4.2. Empirical results

Let $\mathbf{X} = (x_i^j)$ be the data matrix of size 165×96, where $i = 1, \ldots, 165$ and $j = 1, \ldots, 96$. For all pairs of vectors $\mathbf{x}^j = (x_i^j)$ and $\mathbf{x}^{j'} = (x_i^{j'})$, we applied (6) to assess the interlinkage between indicators X^j and $X^{j'}$, resulting in the within-region weighted Kendall correlation matrix $\mathbf{K}$ of size 96×96.

To illustrate the impact of the within-region weighted extension of the Kendall rank coefficient, we consider the example of $X^j =$ sdg1_wpc, representing "Poverty headcount ratio at \$2.15/day (2017 PPP, %)", and $X^{j'} =$ sdg3_traffic, which refers to "Traffic deaths (per 100,000 population)". For this pair of indicators, the distribution of values $\mathbf{x}^j$ and $\mathbf{x}^{j'}$ across 165 countries yields a Pearson correlation coefficient of 0.50 and a weighted Kendall rank correlation of 0.24. In contrast to these positive associations, the within-region weighted Kendall correlation is negative, with a value of -0.48.

Examining the region specific weighted Kendall scores reveals substantial variation: 0.18 for Eastern Europe & Central Asia, -0.60 for East & South Asia, 0.06 for Latin America and the Caribbean, -0.10 for Middle East and North Africa, 0.03 for OECD members, -0.19 for Sub-Saharan Africa.

Region specific scores show sharp contrasts: -0.60 for East & South Asia and -0.19 for Sub-Saharan Africa, *versus* weakly positive values in other cases. This suggests that population weighted regional perspectives can reveal divergent patterns hidden in global averages. In heavily populated regions like East & South Asia, large middle-income countries such as China and India exhibit high traffic deaths alongside reductions in extreme poverty. These outliers drive a negative association when population size is accounted for. Hence, when weighting countries by population, a negative global correlation is arguably more reasonable, reflecting the realities of development where rising incomes do not uniformly translate into improved road safety.

Figure 3 shows the positive and negative interactions among the 96 indicators using standard chord diagrams, with variables grouped by SDG. To structure this signed weighted network, we solve the MRA clustering Problem (11). The optimal solution yields two clusters. As detailed in Section 3.4, the sign of the entries in the leading eigenvector $\mathbf{v}$ of $\mathbf{K}$ aligns exactly with this partition, linking centrality and clustering outcomes. We then reorder the indicators based on the MRA clustering and eigencentrality, following the procedure in Section 3.5. This produces a clearer and more informative chord diagram of both positive and negative interlinkages, shown in Figure 4. All indicator acronyms, descriptions, and eigencentrality scores are provided in the supplementary materials.

Despite improved visualization and partial simplification, the large number of indicators still challenges decision-making. To address this, we identify the Pareto front for each cluster (see paragraph 3.6) to retain only the most relevant indicators. Figure 5 displays the interlinkage subgraph between these selected indicators. This refinement reduces visual clutter and enhances clarity, highlighting key priorities and facilitating more targeted interventions.

The within-region weighted Kendall correlation submatrix for the indicators on the Pareto front of each cluster is presented in Table 1, while their descriptions, population weighted mean performance percentages U, and eigenvector centrality measures V are provided in Table 2.

Acronym	**sdg6_wastewat**	**sdg14_cpma**	**sdg9_rdex**	**sdg16_cpi**	**sdg9_articles**	**sdg16_admin**	**sdg3_uhc**	**sdg2_stunting**	**sdg11_slums**	sdg7_renewcon	sdg16_exprop	sdg12_pollprod	sdg12_ewaste
sdg6_wastewat	1.00	0.12	0.13	0.22	0.16	0.22	0.20	0.20	0.18	0.00	0.22	-0.05	-0.13
sdg14_cpma	0.12	1.00	-0.12	-0.05	-0.08	0.08	0.01	0.23	0.20	-0.45	-0.27	0.17	-0.10
sdg9_rdex	0.13	-0.12	1.00	0.72	0.74	0.58	0.78	0.47	0.55	-0.10	-0.23	-0.67	-0.58
sdg16_cpi	0.22	-0.05	0.72	1.00	0.79	0.68	0.76	0.50	0.59	-0.08	-0.14	-0.65	-0.58
sdg9_articles	0.16	-0.08	0.74	0.79	1.00	0.60	0.78	0.55	0.61	-0.12	-0.28	-0.69	-0.64
sdg16_admin	0.22	0.08	0.58	0.68	0.60	1.00	0.66	0.66	0.63	-0.27	-0.33	-0.47	-0.64
sdg3_uhc	0.20	0.01	0.78	0.76	0.78	0.66	1.00	0.56	0.64	-0.04	-0.21	-0.60	-0.65
sdg2_stunting	0.20	0.23	0.47	0.50	0.55	0.66	0.56	1.00	0.69	-0.31	-0.48	-0.43	-0.62
sdg11_slums	0.18	0.20	0.55	0.59	0.61	0.63	0.64	0.69	1.00	-0.17	-0.37	-0.41	-0.77
sdg7_renewcon	0.00	-0.45	-0.10	-0.08	-0.12	-0.27	-0.04	-0.31	-0.17	1.00	0.43	0.02	0.23
sdg16_exprop	0.22	-0.27	-0.23	-0.14	-0.28	-0.33	-0.21	-0.48	-0.37	0.43	1.00	0.24	0.37
sdg12_pollprod	-0.05	0.17	-0.67	-0.65	-0.69	-0.47	-0.60	-0.43	-0.41	0.02	0.24	1.00	0.44
sdg12_ewaste	-0.13	-0.10	-0.58	-0.58	-0.64	-0.64	-0.65	-0.62	-0.77	0.23	0.37	0.44	1.00

Table 1: Within-region weighted Kendall correlation network of indicators in the Pareto front of each cluster. Pareto-Cluster 1 is in **bold**.

In the following paragraph, we briefly interpret these results with refer-

Acronym	Description	U	V
sdg6_wastewat	Anthropogenic wastewater that receives treatment (%)	13.13	0.024
sdg14_cpma	Mean area that is protected in marine sites important to biodiversity (%)	30.36	0.041
sdg9_rdex	Expenditure on research and development (% of GDP)	31.27	0.114
sdg16_cpi	Corruption Perceptions Index (worst 0-100 best)	35.76	0.117
sdg9_articles	Articles published in academic journals (per 1,000 population)	37.22	0.121
sdg16_admin	Timeliness of administrative proceedings (worst 0 - 1 best)	46.01	0.123
sdg3_uhc	Universal health coverage (UHC) index of service coverage (worst 0-100 best)	47.39	0.124
sdg2_stunting	Prevalence of stunting in children under 5 years of age (%)	52.17	0.129
sdg11_slums	Proportion of urban population living in slums (%)	70.17	0.133
sdg7_renewcon	Renewable energy share in total final energy consumption (%)	20.07	-0.055
sdg16_exprop	Expropriations are lawful and adequately compensated (worst 0 - 1 best)	27.34	-0.087
sdg12_pollprod	Production-based air pollution (DALYs per 1,000 population)	57.64	-0.095
sdg12_ewaste	Electronic waste (kg/capita)	70.86	-0.130

Table 2: Acronyms, descriptions, weighted means of percentages of performance (U), and eigencentrality scores (V) for indicators in the Pareto front of each cluster. Pareto-Cluster 1 is in **bold**.

ence to the cluster specific subsets of Pareto optimal indicators.

4.3. Results interpretations and implications

Pareto-Cluster 1:. This cluster includes indicators related to governance, R&D, health and urbanism, forming a strongly synergistic group with multiple high correlations ($\tau_{w,\mathbb{R}} > 0.6$). Key relationships include:

- Governance, Innovation, Health: Good governance, measured by low corruption (sdg16_cpi) and high administrative efficiency (sdg16_admin), strongly correlates with R&D investment (0.72 and 0.68), scientific output (0.79 and 0.60), and universal health coverage (0.76 and 0.66), indicating that effective institutions support innovation and health [1, 48]. Additionally, R&D expenditure and scientific output each correlate at 0.78 with universal health coverage, suggesting that research investment drives better health outcomes.
- Urban Development and Nutrition: The share of urban population living in slums (sdg11_slums) strongly correlates with child stunting (sdg2_stunting) at 0.69, showing how poor urban conditions hinder nutrition and health [51]. With only 13.13% of anthropogenic wastewater treated (sdg6_wastewat), urgent improvements in water infrastructure are needed to support sustainable development and reduce inequalities.

Pareto-Cluster 2:. In our analysis, this cluster appears to link environmental sustainability, circular economy, and institutional capacity. Although the correlations are moderate (around 0.4), they remain consistently positive:

- Pollution and Waste: Countries that reduce industrial pollution (sdg12_pollprod) tend to manage electronic waste (sdg12_ewaste) more effectively, as shown by a positive correlation of 0.44. Despite its specificity, e-waste stands out as a critical leverage point, with a centrality score of –0.13, due to its strong structural connections across multiple sustainability dimensions.
- Rule-of-law: Strong rule-of-law (sdg16_exprop) is positively associated with improved e-waste recycling (correlation 0.37) and greater renewable energy use (correlation 0.43), exposing the importance of legal certainty for environmental outcomes [29]. Despite being Pareto optimal, both indicators underperform globally (20.07% for renewable energy, 27.34% for property rights), making them urgent priorities for policy action.

Cross-Cluster Trade-offs. Negative correlations between clusters highlight several development dilemmas:

- Urban Development vs. Waste: The strong negative correlation of -0.77 between slum prevalence (sdg11_slums) and e-waste generation (sdg12_ewaste) reveals that reducing slums often coincides with increased electronic consumption and waste. This highlights the need to align urban development with circular economy strategies to mitigate environmental trade-offs [43].
- Innovation vs. Pollution: Scientific output (sdg9_articles) and research investments (sdg9_rdex) show strong negative correlations with industrial pollution (sdg12_pollprod), at -0.69 and -0.67 respectively. This suggests that innovation-driven industrial growth may increase pollution unless accompanied by strict environmental regulations.
- Health vs. Waste: Universal health coverage (sdg3_uhc) is negatively correlated with e-waste generation (sdg12_ewaste) at -0.65, suggesting that improved health systems often coincide with higher consumption and electronic waste. This reflects the growing use of medical technologies, which, despite their benefits, contribute significantly to e-waste [42].

Policy Implications. To advance the SDG, strengthening governance should be a core priority. Reducing corruption and improving administrative efficiency can trigger cascading benefits such as boosting R&D, enhancing health services, and improving urban conditions. Integrating urban infrastructure with public health, clean water, and sanitation initiatives is key to reducing inequalities and fostering well-being.

At the same time, sustainability efforts must address both upstream pollution and downstream waste. The positive link between industrial emission control and e-waste management underscores the need for comprehensive environmental policy. Strengthening rule-of-law such as through secure property rights and legal certainty, is vital for effective recycling and renewable energy adoption. Environmental governance thus depends on broader institutional quality.

Finally, policymakers must anticipate trade-offs between socio-economic gains and environmental sustainability. Urban upgrades and tech access can increase waste and emissions. Responses should pair development with circular economy strategies such as urban mining and sustainable consumption, while enforcing strong environmental regulations. Holistic, system-wide approaches are needed to align growth with long-term ecological responsibility.

5. Discussions and future works

This paper presents a robust, data-driven framework for analyzing interlinkages among the SDG at the indicator level, offering a granular and context-sensitive perspective on sustainability dynamics. By extending Kendall rank correlation to incorporate both population weighting and regional specificity, and by integrating clustering and centrality measures within a unified system, the framework efficiently captures both synergies and trade-offs across indicators. This consistent and integrated methodology generates more actionable insights, helping to guide policymaking at global level that balances urgency with influence. Furthermore, the framework's use of Pareto fronts and advanced visualization tools equips decision-makers with practical mechanisms to prioritize high-impact indicators, thereby supporting more coherent and effective global sustainability governance.

Empirical results based on the SDR 2024 dataset reveal strong interlinkages between governance quality, innovation capacity, and health outcomes, suggesting that effective institutions foster both research ecosystems and equitable health services. The analysis also highlights links between urban conditions, nutrition, and access to clean water, reinforcing the foundational role of infrastructure for human development. A second set of interconnections ties environmental sustainability to institutional strength, showing that better rule-of-law frameworks support pollution control, waste management, and renewable energy uptake. However, critical trade-offs emerge, as improvements in living standards often correlate with increased electronic waste and industrial emissions, underscoring the need for integrated strategies that balance socio-economic progress with environmental stewardship.

Naturally, several limitations must be acknowledged. Firstly, our analysis relies on a single global dataset, the SDR 2024 database [49]. Although this source is widely used in SDG research [31, 25, 61], it is possible that some of our findings reflect dataset specific biases or gaps. Issues related to data availability, consistency, and comparability remain persistent challenges in the field of SDG governance [46, 7]. Future work will apply our methodology to additional global databases, such as those provided by the UN and the World Bank, which feature a broader set of indicators and could help robustify and extend our findings.

Secondly, like most studies on SDG interconnections [54, 61, 5, 28], our framework identifies correlational rather than strictly causal relationships. While the network of synergies and trade-offs offers valuable insights into how indicators tend to move together, it does not establish causality. Consequently, deeper analyses informed by local expertise are essential to support our findings. Moving forward, we intend to explore causal dynamics by using time series data and concepts such as Granger causality, following methodologies outlined in [54] and [33].

Thirdly, although our framework accounts for regional context through weighting, the analysis remains primarily global. As such, the clusters and priority indicators identified may not directly translate to national realities. For example, a trade-off that appears significant globally might be mitigated more easily in certain countries due to different policy frameworks or institutional capacities. To address this, future research will extend our method to analyze SDG interlinkages at national or regional levels such as in [14].

Acknowledgment

This work was partly supported by the Agence Nationale de la Recherche of the French government through the program “Investissements d’Avenir” ANR-10-LABX-14-01.

References

[1] Acemoglu, D. and Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity, and poverty.* Crown Currency.

[2] Ah-Pine, J. (2013a). A general framework for comparing heterogeneous binary relations. In *International Conference on Geometric Science of Information*, pages 188–195. Springer.

[3] Ah-Pine, J. (2013b). Graph clustering by maximizing statistical association measures. In *Advances in Intelligent Data Analysis XII: 12th International Symposium, IDA 2013, London, UK, October 17-19, 2013. Proceedings 12*, pages 56–67. Springer.

[4] Aloise, D., Deshpande, A., Hansen, P., and Popat, P. (2009). Np-hardness of euclidean sum-of-squares clustering. *Machine learning*, 75:245–248.

[5] Anderson, C. C., Denich, M., Warchold, A., Kropp, J. P., and Pradhan, P. (2022). A systems model of sdg target influence on the 2030 agenda for sustainable development. *Sustainability science*, 17(4):1459–1472.

[6] Asadikia, A., Rajabifard, A., and Kalantari, M. (2021). Systematic prioritisation of SDGs: Machine learning approach. *World Development*, 140:105269.

[7] Assubayeva, A. and Marco, J. (2024). Methodological approaches on synergies and trade-offs within the 2030 agenda. *iScience*, 27(11).

[8] Bansal, N., Blum, A., and Chawla, S. (2004). Correlation clustering. *Machine learning*, 56:89–113.

[9] Bennich, T., Persson, Å., Beaussart, R., Allen, C., and Malekpour, S. (2023). Recurring patterns of sdg interlinkages and how they can advance the 2030 agenda. *One Earth*, 6(11):1465–1476.

[10] Bloch, F., Jackson, M. O., and Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare*, 61(2):413–453.

[11] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

[12] Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social networks*, 29(4):555–564.

[13] Bonacich, P. and Lloyd, P. (2004). Calculating status with negative relations. *Social networks*, 26(4):331–338.

[14] Cling, J.-P. and Delecourt, C. (2022). Interlinkages between the sustainable development goals. *World Development Perspectives*, 25:100398.

[15] Conde-Céspedes, P., Marcotorchino, J.-F., and Viennet, E. (2017). Comparison of linear modularization criteria using the relational formalism, an approach to easily identify resolution limit. *Advances in Knowledge Discovery and Management: Volume 6*, pages 101–120.

[16] Dai, S. (2023). Variable selection in convex quantile regression: L1-norm or l0-norm regularization? *European Journal of Operational Research*, 305(1):338–355.

[17] Das, K., Samanta, S., and Pal, M. (2018). Study on centrality measures in social networks: a survey. *Social network analysis and mining*, 8:1–11.

[18] Dawes, J. (2022). SDG interlinkage networks: Analysis, robustness, sensitivities, and hierarchies. *World Development*, 149:105693.

[19] Dawes, J. H. (2023). Network analysis of sdg interlinkages. In *Interlinkages between the Sustainable Development Goals*, pages 94–128. Edward Elgar Publishing.

[20] De Brucker, K., Macharis, C., and Verbeke, A. (2013). Multi-criteria analysis and the resolution of sustainable development dilemmas: A stakeholder management approach. *European journal of operational research*, 224(1):122–131.

[21] Demaine, E. D., Emanuel, D., Fiat, A., and Immorlica, N. (2006). Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187.

[22] Ding, C. and He, X. (2004). K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, page 29.

[23] Ezbakhe, F. and Pérez-Foguet, A. (2021). Decision analysis for sustainable development: The case of renewable energy planning under uncertainty. *European Journal of Operational Research*, 291(2):601–613.

[24] Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305.

[25] Fonseca, L. M., Domingues, J. P., and Dima, A. M. (2020). Mapping the sustainable development goals relationships. *Sustainability*, 12(8):3359.

[26] Fronza, V., Barbero Vignola, G., Borchardt, S., Valentini, S., Buscaglia, D., Maroni, M., and Marelli, L. (2023). Uncovering sdg interlinkages: interconnection at the core of the 2030 agenda. *Publications Office of the European Union.*

[27] Fuso Nerini, F., Mazzucato, M., Rockström, J., Van Asselt, H., Hall, J. W., Matos, S., Persson, Å., Sovacool, B., Vinuesa, R., and Sachs, J. (2024). Extending the sustainable development goals to 2050—a road map. *Nature*, 630(8017):555–558.

[28] Gong, M., Yu, K., Xu, Z., Xu, M., and Qu, S. (2024). Unveiling complementarities between national sustainable development strategies through network analysis. *Journal of Environmental Management*, 350:119531.

[29] Hoicka, C. E., Lowitzsch, J., Brisbois, M. C., Kumar, A., and Camargo, L. R. (2021). Implementing a just renewable energy transition: Policy advice for transposing the new european rules for renewable energy communities. *Energy Policy*, 156:112435.

[30] Issa, L., Mezher, T., and El Fadel, M. (2024). Can network analysis ascertain SDGs interlinkages towards evidence-based policy planning? a systematic critical assessment. *Environmental Impact Assessment Review*, 104:107295.

[31] Kroll, C., Warchold, A., and Pradhan, P. (2019). Sustainable development goals (SDGs): Are we successful in turning trade-offs into synergies? *Palgrave Communications*, 5(1).

[32] Lafortune, G., Fuller, G., Moreno, J., Schmidt-Traub, G., and Kroll, C. (2018). Sdg index and dashboards detailed methodological paper.

[33] Laumann, F., von Kügelgen, J., Uehara, T. H. K., and Barahona, M. (2022). Complex interlinkages, key objectives, and nexuses among the sustainable development goals and climate change: a network analysis. *The Lancet Planetary Health*, 6(5):e422–e430.

[34] Le Blanc, D. (2015). Towards integration at last? the sustainable development goals as a network of targets. *Sustainable Development*, 23(3):176–187.

[35] Lusseau, D. and Mancini, F. (2019). Income-based variation in sustainable development goal interaction networks. *Nature Sustainability*, 2(3):242–247.

[36] Marcotorchino, F. (1986). Cross association measures and optimal clustering. In *COMPSTAT: Proceedings in Computational Statistics, 7th Symposium held in Rome 1986*, pages 188–194. Springer.

[37] Marcotorchino, J.-F. and Michaud, P. (1981). Aggregation of similarities in automatic classification. *Revue de statistique appliquée*, 30(2):21–44.

[38] Michaud, P. and Marcotorchino, F. (1979). Modèles d'optimisation en analyse des données relationnelles. *Mathématiques et Sciences humaines*, 67:7–38.

[39] Nascimento, M. C. and De Carvalho, A. C. (2011). Spectral methods for graph clustering–a survey. *European Journal of Operational Research*, 211(2):221–231.

[40] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

[41] Nilsson, M. and Weitz, N. (2019). Governing trade-offs and building coherence in policy-making for the 2030 agenda. *Politics and Governance*, 7(4):254–263.

[42] Ogunseitan, O. A. (2022). Side effects of the electronic health care revolution: Toxic e-waste. *World Neurosurgery*, 167:2–3.

[43] Ouro-Salim, O. (2024). Urban mining of e-waste management globally: Literature review. *Cleaner Waste Systems*, page 100162.

[44] Pham-Truffert, M., Metz, F., Fischer, M., Rueff, H., and Messerli, P. (2020). Interactions among sustainable development goals: Knowledge for identifying multipliers and virtuous cycles. *Sustainable development*, 28(5):1236–1250.

[45] Pradhan, P., Costa, L., Rybski, D., Lucht, W., and Kropp, J. P. (2017). A systematic study of sustainable development goal (SDG) interactions. *Earth's Future*, 5(11):1169–1179.

[46] Pradhan, P. and Warchold, A. (2023). Quantitative approaches to explore synergies and trade-offs among sustainable development goals (sdgs). In *Interlinkages between the Sustainable Development Goals*, pages 70–93. Edward Elgar Publishing.

[47] Pradhan, P., Weitz, N., Daioglou, V., Abrahão, G. M., Allen, C., Ambrósio, G., Arp, F., Asif, F., Bennich, T., Benton, T. G., et al. (2024). Three foci at the science-policy interface for systemic sustainable development goal acceleration. *Nature communications*, 15(1):8600.

[48] Rocha, L. A. (2019). Corruption, bureaucracy and other institutional failures: the "cancer" of innovation and development. *Economics bulletin*, 39(3).

[49] Sachs, J. D., Lafortune, G., and Fuller, G. (2024). *The SDGs and the UN summit of the future. Sustainable development report 2024*. Dublin: Dublin University Press.

[50] Shieh, G. S. (1998). A weighted kendall's tau statistic. *Statistics & probability letters*, 39(1):17–24.

[51] Silva, J. M., Vieira, L., Abreu, A. M., de Souza Fernandes, E., Moreira, T., da Costa, G. D., and Cotta, R. M. (2023). Water, sanitation, and hygiene vulnerability in child stunting in developing countries: a systematic review with meta-analysis. *Public Health*, 219:117–123.

[52] Spaiser, V., Ranganathan, S., Swain, R. B., and Sumpter, D. J. (2017). The sustainable development oxymoron: quantifying and modelling the incompatibility of sustainable development goals. *International Journal of Sustainable Development & World Ecology*, 24(6):457–470.

[53] Swain, R. B. and Min, Y. (2023). Interlinkages and interactions among the sustainable development goals. In *Interlinkages between the Sustainable Development Goals*, pages 1–15. Edward Elgar Publishing.

[54] Swain, R. B. and Ranganathan, S. (2021). Modeling interlinkages between sustainable development goals using network analysis. *World Development*, 138:105136.

[55] United Nations (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015 (A/RES/70/1).

[56] Vigna, S. (2015). A weighted correlation index for rankings with ties. In *Proceedings of the 24th international conference on World Wide Web*, pages 1166–1176.

[57] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17:395–416.

[58] Wahid, D. F. and Hassini, E. (2022). A literature review on correlation clustering: cross-disciplinary taxonomy with bibliometric analysis. In *Operations Research Forum*, volume 3, page 47. Springer.

[59] Warchold, A., Pradhan, P., and Kropp, J. P. (2021). Variations in sustainable development goal interactions: Population, regional, and income disaggregation. *Sustainable Development*, 29(2):285–299.

[60] Weitz, N., Carlsen, H., Nilsson, M., and Skånberg, K. (2018). Towards systemic and contextual priority setting for implementing the 2030 agenda. *Sustainability science*, 13:531–548.

[61] Wu, X., Fu, B., Wang, S., Song, S., Li, Y., Xu, Z., Wei, Y., and Liu, J. (2022). Decoupling of sdgs followed by re-coupling as sustainable development progresses. *Nature Sustainability*, 5(5):452–459.