A QUANTITATIVE ANALYSIS OF THE RELATIONSHIPS BETWEEN THE SUSTAINABLE DEVELOPMENT GOALS[1]

Viviana Amati, Marta Nai Ruscone

Abstract. The 2030 Agenda for Sustainable Development has the seventeen Sustainable Development Goals (SDGs) at its heart and describes the path to ensure well-being, reduce inequality, encourage economic growth, and preserve the environment. Goals and targets are interconnected by construction. A standard approach to quantify synergies and trade-offs relies on the computation of positive and negative correlation coefficients between the indicators. In this work, we propose a method based on copulas to analyze SDGs interactions, offering a practical and innovative solution. Copulas allow studying more specific types of dependence beyond the linear association described by the correlation coefficients. We illustrate the practical application of this approach by analyzing the interactions between the goals' targets related to health and well-being (Goal 3) and those related to biodiversity (Goals 6,13,14,15).

1. Introduction

The 2030 Agenda for Sustainable Development was adopted by all the members of the United Nations in 2015. Its origin dates to the United Nations Conference on Sustainable Development (Rio+20) that took place in Rio de Janeiro (Brazil) in June 2012, when the nations recognized that they must act in collaborative partnership to achieve sustainable development. The 2030 Agenda is a pivotal document with the seventeen Sustainable Development Goals (SDGs) at its core. Those goals build upon and replace the Millennium Development Goals2, milestones that the countries must have achieved in 2015 to eradicate poverty and improve the life on Earth.

The concept of sustainable development is multidimensional, and its dimensions concern the social, economic and environment domains (Redclift, 1991). It is well-established that countries should deploy strategies to eradicate poverty and deprivations. Those strategies are successful only if the countries simultaneously adopt strategies fostering economic growth, addressing social needs — such as

[1] The Authors contributed equally to this work
[2] https://research.un.org/en/docs/dev/2000-2015

health, education, and inequality — and preserving the environment (United Nations, 2015).

The seventeen SDGs align with the areas defining the concept of development. They are classified into three groups (Boar *et al*., 2020). One group relates to the social area and includes the Goals 1 (no poverty), 2 (zero hunger), 3 (good health and well-being), 4 (quality education), 5 (gender equality), 7 (affordable and clean energy), 11 (sustainable cities and communities) and 16 (peace, justice and strong institutions). Another group concerns the economic area and comprises Goals 8 (decent work and economic growth), 9 (industry, innovation and infrastructure), 10 (reduced inequalities) and 12 (responsible consumption and production). The last group relates to the environment domain and collects Goals 6 (clean water and sanitation), 13 (climate action), 14 (life below water) and 15 (life on land). Goal 17 does not belong to any group since it regards the governance and the countries' commitment to act to monitor and achieve the Goals.

Each Goal includes targets whose progress is measured by a set of indicators created by the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs). Goals and targets are interconnected by construction (Griggs et al., 2014; Le Blanc, 2015) since the sociological, economic and environment area are interdependent. Several approaches have been used to describe the interactions among the goals, such as synergies — progress in one Goal triggers the advancement in another — and trade-offs — improvement in one Goal determines a worsening in another. Understanding this pattern of relationships is vital for policymakers to prepare efficient strategies to move and reach, given the urgency of the 2030 Agenda for Sustainable Development.

A first approach relies on the computation of positive and negative correlation coefficients between the indicators collected for a subset of countries for which data are available (e.g., Kroll *et al.*, 2019; Kostetckaia *et al.*, 2022; Zhao *et al.*, 2023). The sign of the correlation coefficient indicates whether two goals move in the same direction and foster each other (positive coefficient) or move in the opposite direction and progress in one goal might lead to a regression of the other (negative coefficient). Another class of studies (e.g., Pham-Truffert *et al.*, 2020; Allen *et al.*, 2019; Ehrensperger *et al.*, 2019; Griggs *et al.*, 2017) aims at understanding the causality between the goals using the information provided by experts. The experts have to evaluate the influence that progress in a target has on the progress of the other targets. The answers are then aggregated in a point scale (Nilsson et al., 2017) from negative to positive values, where positive values indicate synergies and negative values indicate trade-offs.

In this paper, we align with the studies based on correlation and propose a method that employs copulas to analyze the interactions between SDGs. While correlation and Kendall's tau measure a global association between variables, copulas allow studying more specific types of dependence beyond linear association. To illustrate the method, we consider the SDGs related to health and well-being (Goal 3) and those related to biodiversity (Goals 6,13,14,15), whose links are well described theoretically. We select the corresponding SDG indicators and study the dependence among the targets of these goals, thereby deepening the insights of most of the current studies looking at the goal level. In line with other studies (e.g., Lusseau and Mancini, 2020), we study the dependence by distinguishing the countries into two groups based on their economies (lower and middle-lower vs middle-upper and high) to determine heterogeneities among countries.

The remainder of the paper is organized as follows: We describe the data in Section 2 and briefly introduce copulas in Section 3. In Section 4, we provide an illustrative example of using copulas to study relationships among the Goals. We conclude by discussing the results, limitations of the analysis, and future steps.

2. Data

We consider the Tier 1 class indicators as of 6 March 2024, defined by the IAEG-SDGs group.[3] An indicator belongs to the Tier I group if it "is conceptually clear, has an internationally established methodology and standards are available, and data are regularly produced by countries for at least 50 per cent of countries and of the population in every region where the indicator is relevant." (IAEG-SDGs).

The indicators are supposed to be collected yearly and made available for each country by the UN Statistic Divisions and the World Bank. We considered the most recent values dated to 2022. If the values in 2022 were not available, we imputed the missing data using the most recent value before 2022. For Goal 3, the proportion of missing data was 49.4% and reduced to 9.2% when imputing data using the value observed in 2019 and 2020 and to 5.2% when considering the value observed in 2015. For Goal 6, the proportion of missing data was 59.4% and reduced to 1.7% when imputing data using the value observed in 2020. For Goal 13, the 39.4% of missing data in 2022 was reduced to 19.1% and to 9.7% when considering the values in 2020 and 2017, respectively. For Goal 15, the 5.7% of missing data was reduced to 3.5% and to 1.6% when considering the values in 2020 and 2017, respectively.

The total number of indicators is 61, of which 20 were selected because they are available for at least 75% of the countries. All the indicators for the targets of Goal 14 did not meet the selection criterion; therefore, there are no indicators related to the use of oceans, seas, and marine resources. Table 1 describes the full list of indicators considered in the study.

[3] https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/

Table 1 – *Goals and indicators used in the study.*

Goal 3: Ensure healthy lives and promote well-being for all at all ages
3.1.1 Maternal mortality ratio
3.2.1 Under-5 mortality rate
3.2.2 Neonatal mortality rate
3.3.4 Hepatitis B incidence per 100,000 population
3.3.5 Number of people requiring interventions against neglected tropical diseases
3.4.1 Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease
3.4.2 Suicide mortality rate
3.9.3 Mortality rate attributed to unintentional poisoning
Goal 6: Ensure availability and sustainable management of water and sanitation for all
6.4.1 Change in water-use efficiency over time
6.4.2 Level of water stress: freshwater withdrawal as a proportion of available freshwater resources
6.5.1 Degree of integrated water resources management
6.6.1 Change in the extent of water-related ecosystems over time
Goal 13: Take urgent action to combat climate change and its impacts
13.1.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population
Goal 15: Protect, restore and promote sustainable use of terrestrial ecosystems, sustainable manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
15.1.1 Forest area as a proportion of total land area
15.1.2 Proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, by ecosystem type
15.2.1 Progress towards sustainable forest management
15.4.1 Coverage by protected areas of important sites for mountain biodiversity
15.5.1 Red List Index
15.6.1 Number of countries that have adopted legislative, administrative and policy frameworks to ensure fair and equitable sharing of benefits
15.8.1 Proportion of countries adopting relevant national legislation and adequately resourcing the prevention or control of invasive alien species

The indicators are based on a different scale, and the interpretation of the values sometimes goes in different directions. For most indicators, high values indicate a small development, while low values indicate a high development with respect to a target. We reverse the scale of the indicators whose interpretation did not adhere to this direction by using the complement to 100 when the indicators represent percentages and the difference between the maximum value observed and the indicator's value when the unit of measure has no upper boundary. Finally, we normalized the indicators by using the min-max normalization so that we control for the range of values.

The following analysis focuses on the 111 countries for which the values for all the indicators was available. Among the selected countries, 43 are classified as lower or middle-lower and 58 as medium-upper or high economies countries according to the World Bank country classification by income level.[4]

3. Methods

We study the dependence between target indicators using bivariate copulas. A bivariate copula is a function that joins the bivariate distribution function to their marginal distributions. It describes the dependence structure existing across pairwise marginal random variables. Sklar's theorem (see Nelsen, 2013) shows that every bivariate/multivariate distribution can be defined via copula representation.

Let (X_1, X_2) be a bivariate random variable with marginal cumulative distribution functions $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ and joint cumulative distribution function $F_{X_1,X_2}(x_1, x_2; \theta)$. Sklar's theorem affirms that it exists a copula function $C\big(F_{X_1}(x_1), F_{X_2}(x_2); \theta\big)$ with $C: I^2 \rightarrow I$ such that

$$F_{X_1,X_2}(x_1, x_2; \theta) = C\big(F_{X_1}(x_1), F_{X_2}(x_2); \theta\big), \;\; x_1, x_2 \in \mathrm{I} \subseteq R \; .$$

Copula functions are helpful tools for handling multivariate continuous distributions with given univariate marginals (Nelsen, 2013). They are applied to describe the dependence structure between the marginal distributions of an arbitrary joint distribution. As proven by Sklar's theorem (Sklar, 1959), we can factorize an arbitrary joint distribution in the product of its marginal distributions and dependence structure captured by the copula distribution. Thus, applying copulas allows for separately modeling the marginals and the dependence structure.

By changing the copula function, we can construct new bivariate distributions with different dependence structures. The association parameter of the copula function indicates the strength of the dependence, which may also be different from the linear one that characterizes the multivariate normal distribution.

[4] https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries

Most of the measures of association characterizing the relationship between two variables X_1 and X_2 can be computed using copula functions. For example, the coefficient Kendall's τ is defined as

$$\tau(X_1, X_2) = 4\int_0^1\int_0^1 C(u_1, u_2)dC(u_1, u_2) - 1 .$$

The equation above shows that Kendall's τ depends only on the underlying copula since it is invariant with respect to marginal distributions.

Similarly, tail dependence coefficients can also be defined using copulas. Tail dependence coefficients concern the level of dependence between more extreme values in the upper, lower, or both quadrant tails of a bivariate distribution.

Considering the probability of the joint occurrence of extremely small or large values, the upper and lower tail dependence coefficients are defined as

$$\lambda_L = \lim_{t\to 0^+} P\left(X_2 \le F_2^{-1}(t) \middle| X_1 \le F_1^{-1}(t)\right) = \lim_{t\to 0^+} \frac{C(t,t)}{t}$$

$$\lambda_U = \lim_{t\to 1^-} P\left(X_2 > F_2^{-1}(t) \middle| X_1 > F_1^{-1}(t)\right) = \lim_{t\to 1^-} \frac{1 - 2t + C(t,t)}{1 - t}$$

in case the limits exist.

The coefficients λ_L and λ_U take different values according to the copula chosen. For instance, the Gaussian and Frank copulas do not exhibit tail dependence, i.e., $\lambda_U = \lambda_L = 0$. The Student's T copula has symmetric tail dependence, i.e., $\lambda_U = \lambda_L$. The Clayton and Gumbel copulas are characterized only by lower λ_L or upper λ_U tail dependence.

These parameters λ_L and λ_U measure the dependence in the tails of the joint distribution, i.e. low/high values of one variable are associated with high/low values of the other one. They represent the probability that one variable is extreme given that the other is extreme too. The tail dependence parameters are directly associated to the parameters of some copula families. Further examination of copulas and measures of dependence can be found in Nelsen (2013) and Joe (1997).

In the following, we use the Clayton copula since we are interested in the left tail of the joint distribution of two indicators. The choice of the copula is based on the expectations that biodiversity and health (Griggs et al., 2017) are synergetic, meaning they work together in a way that an improvement in one goal implies an improvement in the other goal. This relationship manifests itself in low values of the indicators.

We represent the dependence structure implied by the Clayton Copula as a dependence graph, a graph in which the nodes are the indicators and the ties denote the presence and strength of the dependence as measured via a suitable copula-based tail coefficient describing the extreme of the co-movements of the dependence in the lower tail.

Finally, we utilize the dependence structure to cluster the indicators by applying the Louvain clustering method for community detection (Blondel *et al.*, 2008). This method allows us to extract non-overlapping sets of synergetic indicators. The Louvain method clusters the nodes in groups that are internally well-connected and weakly connected to other sets of nodes. The chosen node partition is generated by an algorithm that iteratively maximizes the modularity coefficient (Clauset *et al.*, 2004), and the best partition is the one with the maximum modularity.

4. Results

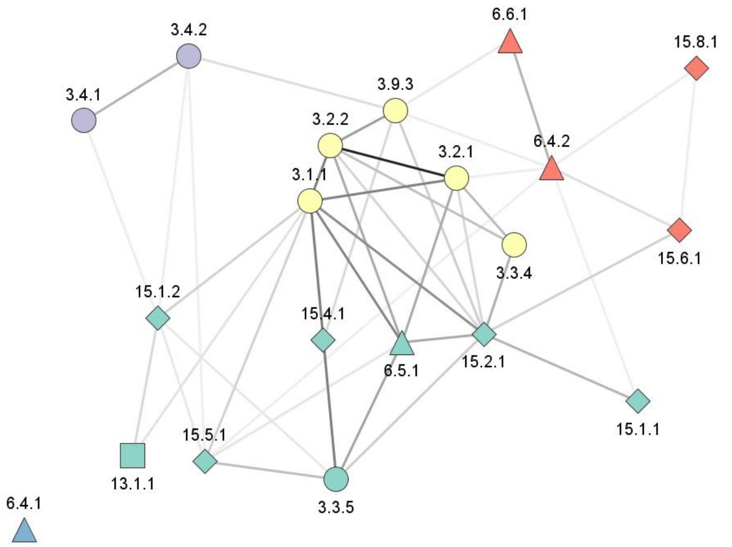
Figures 1 to 3 represent the dependence graphs. Nodes are the indicators. The node shape depicts the goal, and the color indicates the membership to the clusters obtained with the Louvain method.

We start by commenting on the dependence graph computed on all the 111 countries (Figure 1). The clustering identified 5 groups, each with its unique characteristics and significant implications for global development and sustainability.

The first group contains the indicators of target 6.4.1, which is independent from all the other targets. The second group (violet) includes the indicators of target 3.4, which is weakly connected to the other indicators. The third group (yellow) contains 5 indicators concerning health, specifically maternal and child mortality, hepatitis B incidence, and mortality due to unintentional poisoning. The fourth group (orange) has 4 indicators and relates to the legislative, administrative, and policy framework for sharing benefits, controlling invasive alien species, and managing water resources. The last group refers to forest preservation, management of water resources, and interventions against neglected tropical diseases.

Figure 2 shows the dependence graph for 58 countries classified as high and middle-upper economies. The clustering identified 6 groups. Among those, two groups include only one indicator, the one for targets 6.4.1 and 15.8.1, indicating weak dependences on all the other targets. The third group (orange) contains the indicators of target 3.4, which are also dependent on mountain biodiversity and interventions against neglected tropical diseases. The fourth group (yellow) includes

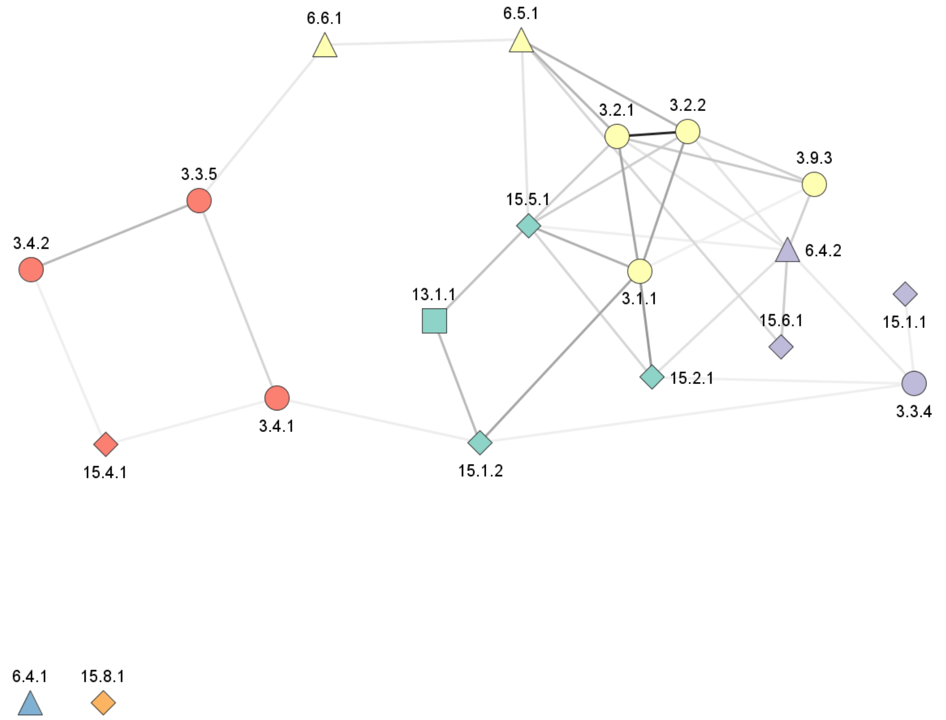
Figure 1 – *Dependence graphs for Goals 3, 6, 13, and 15 for the 111 countries for which the data are available. The node colour indicates the membership in the group, the shape the goal, the intensity of the link the dependence strength as measured by the Clayton copula.*



six indicators concerning health (specifically maternal and child mortality and the mortality due to unintentional poisoning) and the sustainable management of water and sanitation. The fifth group (green) has 4 indicators and relates to the presence of forests and deaths due to natural disasters. The last group (violet) has 4 indicators measuring the legislative, administrative, and policy framework for sharing benefits, forest management, water stress, and hepatitis B incidence.

Finally, Figure 3 reports the results for the 43 countries classified as low and middle-low economies. The first group (red) includes 8 indicators relating to the management of forests, changes in the water-related ecosystems, and the deaths for united poisoning. The second group (green) has 6 indicators all related to mortality, specifically maternal child, cardiovascular disease, cancer, diabetes and chronic respiratory disease, and suicide mortality rate. The third group (violet) contains 3 indicators concerning water resource management, water-use efficiency changes, and biodiversity. The last group (yellow) includes four indicators concerning hepatitis B incidence, interventions against neglected tropical diseases, deaths due to natural disasters, and water stress.

Figure 2 – *Dependence graphs for Goals 3, 6, 13, and 15 for the 58 middle-high and high economies countries for which the data are available. The node colour indicates the membership in the group, the shape the goal, the intensity of the link the dependence strength as measured by the Clayton copula.*

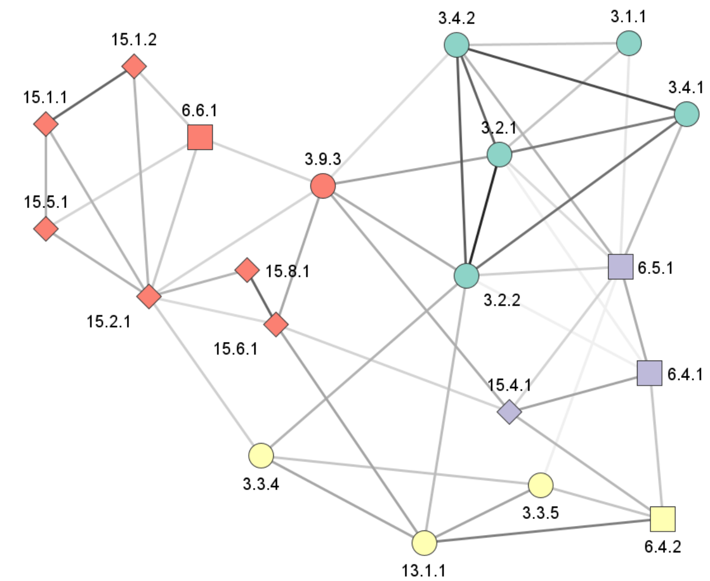


5. Discussion and conclusions

In this paper, we propose using copulas to deepen the analysis of trade-offs and synergies among the sustainable development goals and their targets. We illustrate the method by analyzing the interconnections among the targets of Goals 3, 6, 13, and 14 related to health and biodiversity. Given the documented synergies between the Goals, we delved into the dependencies, describing how an improvement in one goal implies an improvement in the other goal using the Clayton copula.

The clustering analysis indicates that targets related to maternal and child mortality and improvement in sanitation are strongly associated. Similarly, the forest and the water resources management show a strong lower tail dependence. Those results are in line with theoretical expectations (Griggs et al., 2017)

Figure 3 – *Dependence graphs for the Goals 3, 6, 13, and 15 for the 43 countries classified as low and middle-low economies for which the data are available. The node colour indicates the membership in the group, the shape the goal, the intensity of the link the dependence strength as measured by the Clayton copula.*



We performed the analysis on three groups, including all the 111 countries that did not present any missing data, the subset of the 58 countries classified as middle-upper and high economies, and the 43 countries classified as low and middle-low economies for which the data are available. The results are stable, though a few indicators with weak dependencies on the others were classified in different clusters across the three contingents of countries. The main difference is that the indicator of target 6.4.1 did not show any dependence on the other indicators when all the countries and those classified as middle-upper and high economies are considered. However, target 6.4.1 is associated with mountain biodiversity and the management of water resources when focusing on low and middle-low economy countries.

It is important to point out that our study is an illustration and the results are far from being complete and cannot be interpreted in a causal sense. Results are incomplete because of the high number of missing data and the focus on a limited set of targets and goals. Results cannot be interpreted in a causal sense since they only describe the dependencies among indicators and the targets they refer to, and this dependence describes how improvement in one goal dimension is associated with improvements in other goal dimensions.

The next natural step is to extend the analysis to all seventeen goals, use different copulas to describe various types of dependencies, and complement the results on the correlation coefficients. This future research holds the promise of further enhancing our understanding of the complex interconnections among sustainable development goals and their targets. Results can guide policymakers and practitioners in their efforts toward sustainable development.

References

ALLEN C., METTERNICHT G., WIEDMANN T. 2019. Prioritising SDG targets: Assessing baselines, gaps and interlinkages, *Sustainability Science*, Vol. 14, pp. 421-438.

BOAR A., BASTIDA R., MARIMON F. 2020. A systematic literature review. Relationships between the sharing economy, sustainability and sustainable development goals, *Sustainability*, Vol. 12, No. 17, pp. 1-14.

BLONDEL V. D., GUILLAUME J. L., LAMBIOTTE R., LEFEBVRE E. 2008. Fast unfolding of communities in large networks, *Journal of statistical mechanics: theory and experiment*, Vol. 2008, No. 10, pp. 1-12.

EHRENSPERGER A., DE BREMOND A., PROVIOLI I., MESSERLI P. 2019. Land system science and the 2030 agenda: exploring knowledge that supports sustainability transformation, *Current Opinion in Environmental Sustainability*, Vol. 38, pp. 68-76.

GRIGGS D. J., SMITH M. S., ROCKSTRÖM J., ÖHMAN M. C., GAFFNEY, O., GLASER G., KANIE N., NOBLE I., STEFFEN W., SHYAMSUNDAR P. 2014. An integrated framework for sustainable development goals, *Ecology and society*, Vol. 19, No. 4, pp. 1-24.

GRIGGS D. J., NILSSON M., STEVANCE A., MCCOLLUM D. 2017. *A guide to SDG interactions: from science to implementation*. International Council for Science, Paris.

JOE H. 1997. *Multivariate models and multivariate dependence concepts*. CRC press.

KOSTETCKAIA M., HAMENTNER M. 2022. How sustainable development goals interlinkages influence European Union countries' progress towards the 2030 agenda, *Sustainable Development*, Vol. 30, No. 5, pp. 916-926.

KROLL C., WARCHOLD A., PRADHAN P. 2019. Sustainable Development Goals (SDGs): Are we successful in turning trade-offs into synergies? *Palgrave Communications*, Vol. 5, No. 1, pp. 1-11.

LE BLANC D. 2015. Towards integration at last? The sustainable development goals as a network of targets, *Sustainable Development*, Vol. 23, pp. 176–187.

LUSSEAU D., MANCINI F. 2019. Income-based variation in Sustainable Development Goal interaction networks, *Nature Sustainability*, Vol. 2, No. 3, pp. 242-247.

NELSEN R. B. 2013. *An introduction to copulas*. Springer.

NILSSON M, GRIGGS D, VISBECK M, RINGLER C. 2017. A framework for understanding sustainable development goal interactions. In GRIGGS D. J., NILSSON M., STEVANCE A., MCCOLLUM D. (Eds) *A Guide to SDG Interactions: From Science to Implementation*, International Council for Science: Paris, pp. 18-31.

PHAM-TRUFFERT M., METZ F., FISHER M., RUEFF H., MESSERLI P. 2020. Interactions among Sustainable Development Goals: Knowledge for identifying multipliers and virtuous cycles, *Sustainable development*, Vol. 28, No. 5, pp. 1236-1250.

REDCLIFT M. 1991. The multiple dimensions of sustainable development. *Geography*, Vol. 76, No. 1, pp. 36-42.

SKLAR A. 1959. Fonctions de Répartition à n Dimensions et Leurs Marges, *Annales de l'Institut Statistique de l'Université de Paris*, Vol. 8, No. 3, pp. 229-231.

UNITED NATIONS. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. *United Nations: New York, NY, USA*.

ZHAO M., MENG Z., XU B., YU D., WU S., LI R. 2023. Prior goal setting of SDGs based on the interaction correlation, *Frontiers in Sustainability*, Vol. 4, pp. 1-8.

Viviana AMATI, Department of Statistics and Qualitative Methods, University of Milano- Bicocca, viviana.amati@unimib.it

Marta NAI RUSCONE, Department of Economics, University of Genoa, marta.nairuscone@unige.it