

Decision models for sustainable development goals based on a relational network approach

Junxiang Wang*, Yuxuan Cao, Xiaoqi Shu

International School, Beijing University of Posts and Telecommunications, Beijing, China

*Corresponding author: 1034696975@bupt.edu.cn

Abstract. The interrelationship and prioritisation of the 17 separate Sustainable Development Goals (SDGs) will greatly influence the overall SDGs development agenda. The models developed in this paper quantify the influencing factors of SDG, resulting in a relational network and priority ranking that can make predictions about future development. The models all perform well through fit testing and can be used to propose relationships between company's and organisation's development goals and to suggest a priority development order. Based on the Sustainable Development Goals Index (SDG-I) published by the United Nations, a network model of SDG relationships was developed through Spearman correlation analysis. An ARIMA model was constructed using time series predictive analysis on the basis of the SDG relationship network model. A revised model of SDG relevance was developed based on improvements to the SDG relational network model and a proposal to introduce mental health as a new SDG was made.

Keywords: SDGs, relationship network, Spearman correlation analysis, time series analysis, PageRank algorithm.

1. Introduction

In 2015, the United Nations General Assembly formally adopted the 2030 Agenda for Sustainable Development. Seventeen Sustainable Development Goals (SDGs) were proposed in order to promote the organisational functioning and integration of sustainable development. The integrated development of the 17 SDGs will effectively guide the new agenda of global sustainability efforts. Because of the relatively complex interrelationships among the GOALS, they can be viewed as a complex network of relationships. It is important to study the correlation and priority among the goals in order to effectively promote the implementation of the SDGs.

According to Scherera^[1], L, the SDGs play an important role in sustainable development and policy alignment^[2]. The correlation between the SDGs, while mainly demonstrating synergies, also indicates trade-offs. This implies that in some cases there is a mutually reinforcing or complementary relationship between the two goals, where the achievement of one SDG does not enable the progress of the other,^[3] and the success of one SDG is predicated on the success of the other. However, there is still room for expansion of the SDGs. For example,^[4] Des Gasper argues that some of the SDGs are missing valuable goals, such as migration.

2. SDGs relational model

Based on^[5] the Bertelsmann Stiftung and the Sustainable Development Solutions Network (SDSN) determination of 17 SDGs for 193 UN member countries (including overall results, individual goal scores, and goal trends for all countries), the SDGs correlation A relationship model was developed. The relational model was designed to map the relationships between the SDGs.

2.1. Correlation coefficient

Given the conceptual complexity and broad scope of the SDGs, translating the SDGs into measurable data is challenging. Due to some missing data in the statistical report, the paper also refers to the data in^[6]. To balance the scope of the data for cross-sectional comparisons and vertical trends over

time, the paper selected the 2019 Sustainable Development Goals Index (SDG-I) for each UN member country, which is used to monitor global progress towards the SDGs and provide a single benchmark for international comparisons. Using the SDG-I based on a K-S test of the data, the SDGs largely do not obey a normal distribution among themselves, therefore, the spearman correlation coefficients with each of the 16 SDGs other than their own were calculated separately as a way to reflect the interrelationships among the SDGs.

$$\rho = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}} \quad (1)$$

Firstly, the existence of a statistically significant relationship between X and Y was first tested to determine whether the p-value showed significance ($p < 0.05$). If significant, then a correlation exists between the two variables, and if not, then no correlation exists between the two variables. The correlation coefficients were then analysed for the positive and negative direction as well as the degree of correlation.

Then, the correlation coefficient table was obtained by performing Spearman correlation analysis on SDG-I. The table covers that for more visualization of the results, a correlation coefficient heat map is provided to visualize the correlation coefficient between SDGs. It indicates the positive, negative and magnitude of correlation coefficients between SDGs by color shades. As illustrated in Fig1.

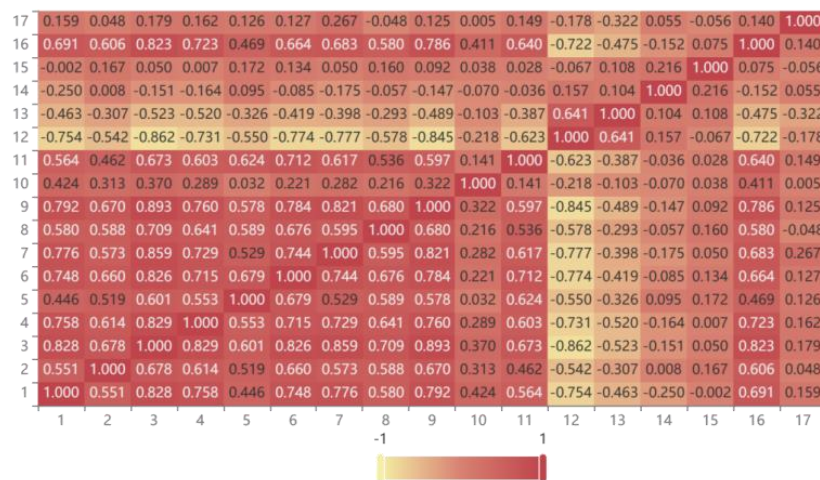


Fig 1. Correlation Coefficient Heat Map

2.2. SDG relationship network SDG

Based on the results of the analysis and graphical presentation, the paper found several sets of significant correlations between SDGs, and in order to provide a comprehensive overview of their correlations, the correlation levels were divided into five levels according to the magnitude of the correlation coefficients and were plotted as a relational network diagram^[7]. Also, based on the summation of the correlation coefficients between each SDG and the other SDGs, a weight value that reflects the degree of influence of that SDG on the other SDGs can be derived. The weight value determines the size of SDG points in the relationship network graph.^[8] Correlation levels classification and SDG relationship network are shown in the Table 1 and Fig 2.

Table 1. Correlation levels classification

Correlation Coefficient	Correlation Level	Color Grade
0.75~0.9	Strong positive correlation	Red(thick)
0.6~0.75	Weak positive correlation	Red(thin)
-0.6~0.6	Inexistent correlation	No line
-0.75~-0.6	Weak negative correlation	Green(thin)
-0.9~-0.75	Strong negative correlation	Green(thick)

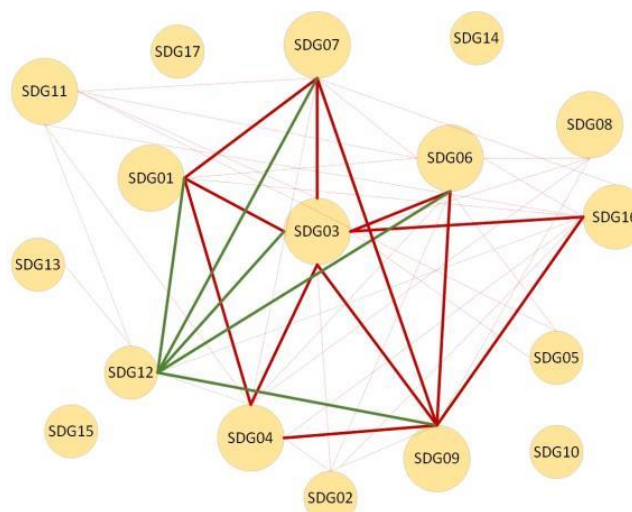


Fig 2. SDG relationship network

3. Priority

In this section, SDG priority starts to be set based on the SDG relationship network. An ARIMA model is built by a time series predictive analysis method to predict the goals that will be achieved in ten years, where a goal has been achieved.

3.1. SDG prioritization

The priority of SDGs is based on the degree of influence of each SDG on other SDGs. The principle is that the SDG correlation coefficient can reflect the positive or negative relationship and the degree of influence of SDGs on other goals, and to some extent, the marginal utility of a SDG on other goals after it is achieved. When a correlation coefficient is positive, it means that the realization of that SDG is contributing to another party. In order to facilitate the achievement of the overall SDG agenda, the SDG with the highest marginal utility to other GOALS is selected as the priority development. Therefore, based on the sum of the correlation coefficients of the 17 SDGs, the priority ranking can be calculated.

For example, based on the correlation coefficient table and SDG relationship network diagram, the sum of correlation coefficients of SDG12 is the lowest value, and if priority is given to the development of SDG12, then it may lead to inhibit the development of other SDGs. The results of priority calculation are shown in the Fig 3.

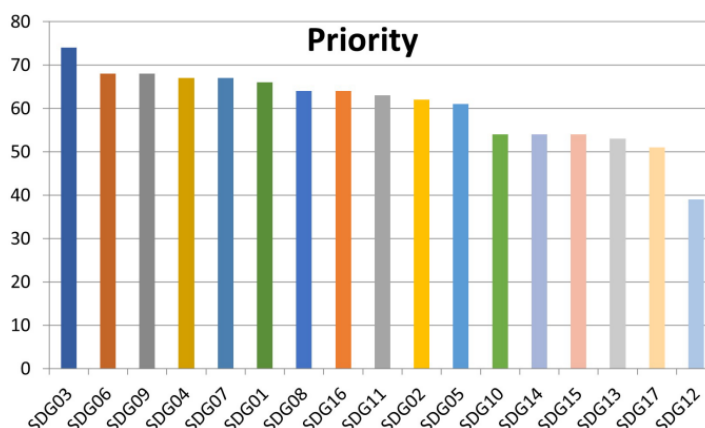


Fig 3. Priority Chart

3.2. Priority effectiveness assessment

The validity of priorities is assessed by two methods: one is direct proof and the second is logical

analysis.

Direct proof is obtained by proving that this priority ranking is derived from the actual SDG correlation, that is, the priority ranking is supported by real data that effectively shows the impact of the completion of a project on the progress of other projects, and thus proves that the priority ranking is a valid ranking with developmentality as the main objective.

According to Luis Miguel Fonseca's correlation analysis of SDG in [9], SDG2, SDG3, SDG4, SDG7, SDG8, SDG9 and SDG11 with SDG11 have the strongest positive correlations with the other SDGs, while SDG12 shows a significant negative correlation with the other SDGs. Comparing the correlation matrix and the analytical findings with the prioritization, there is little difference between the two findings if the effect of data selection differences on the results is removed. Therefore, it can be concluded that the priority ranking is valid by direct proof.

The logical analysis is to verify the validity of the priority ranking by reasoning through the accepted relationships that exist between SDGs. Take the example of the relationship between SDG13, SDG01 and SDG02. Climate change can lead to climate hazards and natural disasters, which can lead to crop destruction and food shortages, leading to food security issues and increased poverty.

To conclude, the validity of this prioritization can be verified through direct evidence based on literature and logical analysis based on speculation.

4. SDG prediction model in ten years

The prediction model is an ARIMA model based on time series analysis, by which the paper can predict the highest priority to achieve the highest priority after SDG03. The general form of the model is as follows.

$$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (2)$$

First, the smoothness of the series was verified by ADF test. The time series is analyzed for biased autocorrelation, and its p and q values are estimated according to the truncated tails. After testing the model white noise and analyzing the AIC and BIC values, the analysis is performed by model residual ACF/PACF plots. Finally, according to the model parameter table, the model equation is derived combined with the time series analysis plot for a comprehensive analysis to obtain the backward prediction order results.

4.1. Sequence smoothing test and determination of d Value

The ARIMA model requires that the series meet is a smooth time series data. Viewing the results of the ADF test, based on the analyzed t-values, it can be analyzed whether it can significantly reject the hypothesis that the series is not smooth. If it presents significance ($p < 0.05$), it means that the original hypothesis is rejected and the series is a smooth time series. If the ADF Test result is less than 1%, 5% and 10% at the same time, then it means that the hypothesis is well rejected. The following Table 2 shows the results of ADF Test for the data.

Table 2. ADF Test Form

Variable	Difference order	t	P	AIC	Threshold value		
					1%	5%	10%
Goal 3 Score	0	-3.494	0.008***	12.372	-4.138	-3.155	-2.714
	1	-3.868	0.002***	30.399	-3.809	-3.022	-2.651
	2	-4.885	0.000***	8.498	-3.924	-3.068	-2.674

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively

The results of this series test show that: At the difference into 0th order, the significance P-value is 0.008***, which presents significance at the level, and the original hypothesis is rejected, and the series is a smooth time series. The ADF hypothesis test value of the 0th order difference series is -3.494, which is less than the statistical value of three levels, and the significance p value is $0.008 < 0.05$, so

the 0th order difference can make the series become smooth, and the difference order $d=1$. The best difference series graph is shown below Fig 4.

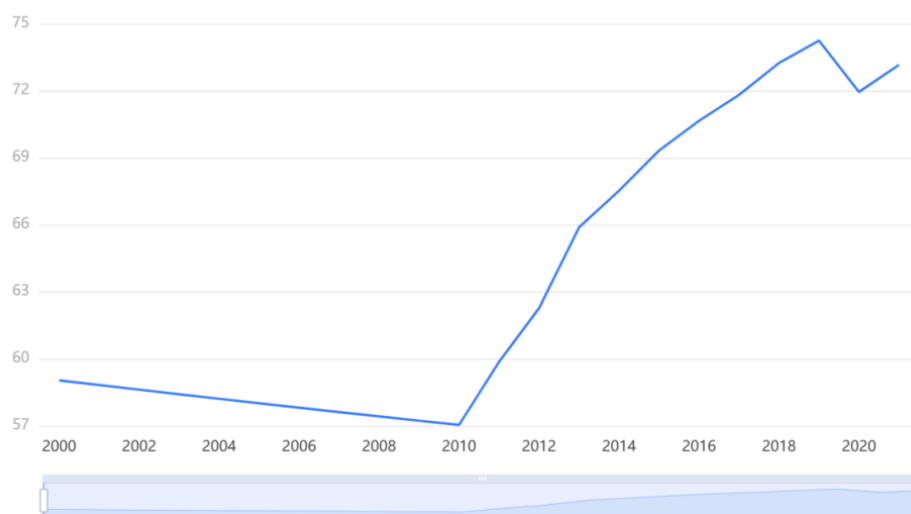


Fig 4. The Best Difference Series Graph

4.2. Determine the p and q values

The autocorrelation coefficient reflects the degree of correlation before and after the realization of SDG03 itself. The autocorrelation coefficient (ACF) plot is obtained and the q value is determined by the maximum lag point. From the autocorrelation plot, it can be seen that the autocorrelation coefficients from order 0 are all within 2 times the standard deviation, and the autocorrelation plot is judged to be a truncated tail with $q=0$.

The partial autocorrelation coefficient captures the degree of relevance of SDG03 realization to other SDGs. The partial autocorrelation coefficient (PACF) plot is obtained and the p-value is determined by the maximum lag point. From the partial autocorrelation plot, it can be seen that the autocorrelation coefficients from order 0 are all within 2 times the standard deviation, and the autocorrelation plot is judged to be truncated with $p=0$.

The autoregressive model (AR) residuals and sliding average model (MA) residuals are white noise series according to the model residual autocorrelation plot (ACF) and model residual bias autocorrelation plot.

4.3. Fitting the ARIMA Model (p, d, q)

According to Table 3, from the analysis of the Q statistic results, the paper can get: Q6 does not present significance at the level and cannot reject the hypothesis that the residuals of the model are white noise series, while the goodness of fit of the model R^2 is 0.984, the model performs excellent and the model basically meets the requirements.

Table 3. ARIMA model (0,1,0) test table

ARIMA model (0,1,0) test table		
Item	Symbol	Value
Sample size	Df Residuals	20
	N	22
Q-statistic	Q6(P-value)	0.001(0.970)
	Q12(P-value)	6.547(0.365)
	Q18(P-value)	12.44(0.411)
Information criterion	AIC	40.059
	BIC	42.148
Goodness of fit	R^2	0.984

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively

Based on the formula: $y(t)=0.687$, the model results are derived, which is the ARIMA model (0,1,0) test table and based on 0-difference data. The model parameters Table 4 is as follows .

Table 4. Model parameters table

Model parameters table						
	Coefficient	Standard deviation	t	P> t	0.025	0.975
Constant	0.687	0.125	5.512	0	0.443	0.931

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively

4.4. Projection of future values

The table shows the data projections of the time series model up to 2030. From the information in the table, it is clear that the overall SDG agenda development has steadily increased year by year after the realization of SDG03. The figure below shows the true, fitted and predicted values of the model. As can be seen from Fig 5, the trend of the fitted value is close to the trend of the true value, and the fit is good, so the prediction results are reliable.

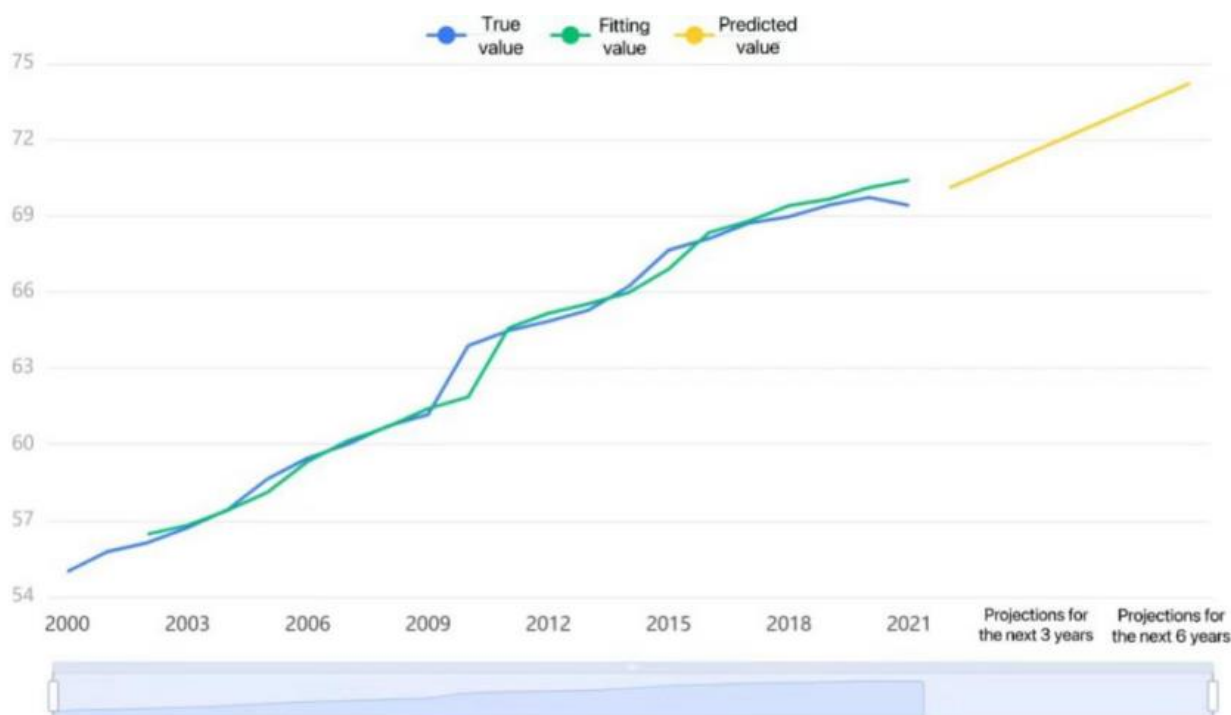


Fig 5. Time Series Chart

5. Correction of correlation matrix in case of sudden change

This paragraph describes the impact of the rapid achievement of the SDGs on the network of relationships and priorities among the SDGs, and makes recommendations regarding the inclusion of the new SDGs. The new model will consider both target correlations and the SDG index (SDG-I), and improve the calculation method of the existing correlation network.

5.1. Structure

Since the development of SDGs is influenced by complex environmental factors, the paper prefer to select SDGs that are achieved quickly and with a significant degree of development, so the paper assume a significant and sudden change in the SDG index (SDG-I) of that SDG.

In proposing a network of relationships for correlations among SDGs, a network of correlations among SDGs was established based on the correlation analysis of the actual scores of global SDGs provided by the United Nations and related foundations. By analyzing the development assessment

of global SDGs provided by the UN and related foundations, the degree of correlation between SDGs should be related to two indicators, which are the correlation between SDGs (i.e., the relationship between the development of global SDGs), and the related SDG index (SDG-I).

In proposing a network of relationships for correlations among SDGs, a network of correlations among SDGs was established based on the correlation analysis of the actual scores of global SDGs provided by the United Nations and related foundations. By analyzing the development assessment of global SDGs provided by the UN and related foundations, the degree of correlation between SDGs should be related to two indicators, which are the correlation between SDGs (i.e., the relationship between the development of global SDGs), and the related SDG index (SDG-I).

Given that the two indicators related to the degree of correlation among SDGs are homogeneous data when constructing the previous network of relationships among SDGs (the correlation among SDGs is calculated by processing the relevant SDG realities), the correlation among SDGs can be directly considered as another expression of the degree of correlation among SDGs. In the hypothesis discussed in this paragraph, there is a possibility of mutation of the SDGs, and the distribution of the SDG reality after the mutation will not be consistent with the distribution of the correlation between the SDGs that actually exist, and it is necessary to process the scores of the mutated SDGs and the correlation degree between the original SDGs to obtain the correlation degree between the new indicators. The formula is expressed as follows.

In order to ensure that the new correlation degree is relatively smooth and smooth in the change, the processed mutation data will be given 20% weight, and then summed with the original correlation degree to get the new correlation degree. After calculating and analyzing several sets of data, giving 20% weight to the processed mutation data and 80% weight to the old correlation degree can make the updated correlation degree move in the direction favorable to global development and also ensure the stability of the change of the correlation degree of sustainable development goals.

The mutation function is handled as follows.

$$\text{change}_i = (\text{target} - \text{CSAM}_{\text{goal}}) * \text{OIOCM}_{\text{goal},1} \quad (3)$$

$$\text{rate}_{i,1} = \frac{\text{change}_i}{\text{CSAM}_i} \quad (4)$$

$$\text{result}_{i,j} = \frac{\text{rate}_i}{\text{rate}_j} (\text{rate}_j > \text{rate}_i) \quad (5)$$

$$\text{result}_{i,j} = \frac{\text{rate}_j}{\text{rate}_i} (\text{rate}_i > \text{rate}_j) \quad (6)$$

$$\text{NC}_{i,j} = \text{result}_{i,j} * 0.2 + \text{OIOCM}_{i,j} * 0.8 \quad (7)$$

Metric OIOCM is the Original inter-objective correlation matrix. OIOCM is a matrix with 17 rows and 17 columns. Metric CSAM is the Current SDG Achievement Matrix. CSAM is a matrix with 17 rows and 1 column. The variable target records the fast-growing SDGs as the number of targets. The variable goal records which SDG is rapidly evolving. Change is a matrix with 17 rows and 1 column, recording the amount of change of all 17 SDGs. The rate matrix is a 17-row, 1-column matrix that holds the rate of change of the 17 SDG values.

The rate of change in target performance is the ratio obtained by dividing the predicted change in value of the non-mutant target after receiving the effect of a sudden change in the value of the mutant target (predicted using the correlation matrix) by the current (unaffected) value of the target. In setting up the model, the degree of correlation between targets is expected to reflect both the target correlation and the contribution of the target in question to the overall target development.

In reality, after a certain level of development of a target (as reflected by a certain level of achievement) the marginal benefit of continued development of the indicator will diminish as the level of development increases. Considering the overall development of the SDG, it is necessary to make some "punitive" corrections to the value of the correlation between this goal and other goals, i.e., to lower the ranking of this goal in the ranking of the SDG contribution (priority) in order to encourage

the development of those goals that are really weak (lower achievement). By doing so, the overall SDG development agenda will gain the most benefit for the same effort. Therefore, when updating the correlation, the value of data change should not be taken as a reference only at face value, but the base value of that goal should also be introduced, correcting the value of change to the rate of change of the goal. It is known that the higher the target score, the higher the "penalty", and the higher the target score, the lower the rate of change, and the lower the correlation between this target and other targets, all else being equal. At the macro level, the correlation between this target and all other SDGs will decrease. In the iterative process, the higher the score, the lower the relevance of the goal to other goals, i.e., the lower the relevance to other goals, and then the final ranking in the priority will tend to be lower.

In the first step, the paper have demonstrated the reasonableness of using the rate of change in target scores as a reference for judging the change in the degree of correlation. On this basis, in order to reflect the change in the correlation coefficient between the relevant targets, the rate of change of both scores is divided to obtain the ratio of the change in both, and this value is multiplied by the weight value to correct the original correlation coefficient.

The above discussion was verified. By analyzing the SDG index from 2000 to 2001 provided by the United Nations, the achievement of SDG9 has increased rapidly. Based on the global SDG data for 2000, with the help of the rapid growth of SDG9, the paper used the newly proposed model to forecast the data for 2021, and the forecast results fit well with the real data for 2021, and the model has credibility.

5.2. Modified relationship network

The following section will use SDG03 as an example to construct its new network structure after it has been prioritized as a development goal and rapidly achieved. According to the UN assessment of the data, a Sustainable Development Goal Index (SDG-I) of 85 or above is considered as target completion. Bringing in the relationship coefficient evaluation criteria yields a new relationship network as Fig 6 .

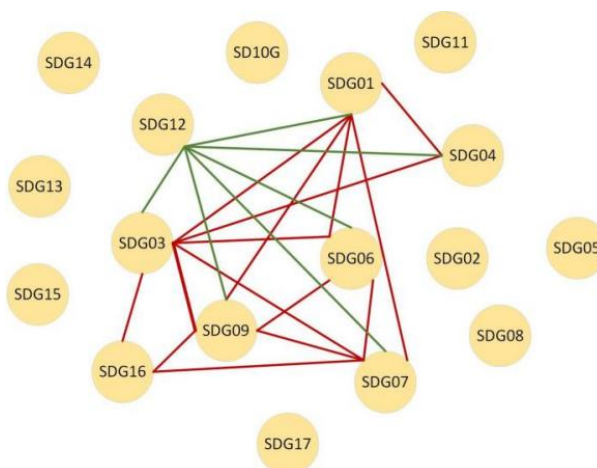


Fig 6. SDG modified relationship network

The new priority order is: SDG03, SDG09, SDG01, SDG06, SDG07, SDG16, SDG04, SDG08, SDG02, SDG11, SDG05, SDG13, SDG10, SDG14, SDG15, SDG17, SDG12.

5.3. SDG to be added

Newly added SDGs should be eligible for significant contribution to global sustainable development and positive relevance to the overall SDG agenda. Through literature review and data, mental health was suggested for addition, ^[10]mental health challenges affect millions of people worldwide, and the COVID-19 pandemic highlights the urgent need for greater investment in mental health resources. The SDGs to be added could focus on promoting mental health and well-being, which would allow more people in more countries and regions to have access to good mental health care.

6. Conclusion

The models in this paper predict the impact of a single country's relationship matrix and non-linearly changing SDGs on other goals over a given time period, providing a predictive model for other companies' and organisations' decisions. However, there are some shortcomings in the study, such as asymmetries in the interrelationships, data errors, lack of data volume.

References

- [1] Laura Scherer, Paul Behrens, Arjan de Koning, Reinout Heijungs, Benjamin Sprecher, and Arnold Tukker. Trade-offs between social and environmental sustainable development goals. *Environmental science & policy*, 90:65–72, 2018.
- [2] Prajal Pradhan, Luís Costa, Diego Rybski, Wolfgang Lucht, and Jürgen P Kropp. A systematic study of sustainable development goal (sdg) interactions. *Earth's Future*, 5(11):1169–1179, 2017.
- [3] Måns Nilsson, Dave Griggs, and Martin Visbeck. Policy: map the interactions between sustainable development goals. *Nature*, 534(7607):320–322, 2016.
- [4] Des Gasper. The road to the sustainable development goals: Building global alliances and norms. *Journal of Global Ethics*, 15(2):118–137, 2019.
- [5] Jeffrey Sachs, Guido Schmidt-Traub, Christian Kroll, David Durand-Delacre, and Katerina Teksoz. New york: Bertelsmann stiftung and sustainable development solutions network (sdsn), 2018.
- [6] Jeffrey Sachs, Guido Schmidt-Traub, Christian Kroll, and Guillaume Lafortune. Sustainable Development Report 2019: Transformations to Achieve the Sustainable Development Goals: Includes the SDG Index and Dashboards:[G20 and Large Countries Edition]. Bertelsmann Stiftung, 2019.
- [7] Sachs J D, Schmidt-Traub G, Mazzucato M, et al. Six transformations to achieve the sustainable development goals[J]. *Nature sustainability*, 2019, 2(9): 805-814.
- [8] Le Blanc D. Towards integration at last? The sustainable development goals as a network of targets[J]. *Sustainable Development*, 2015, 23(3): 176-187.
- [9] Luis Miguel Fonseca, José Pedro Domingues, and Alina Mihaela Dima. Mapping the sustainable development goals relationships. *Sustainability*, 12(8):3359, 2020.
- [10] Saloni Dattani, Hannah Ritchie and Max Roser (2021) - "Mental Health". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/mental-health>' [Online Resource]