Problem Chosen	2023	Team Control Number
D	MCM/ICM	2317835
	Summary Sheet	

Would tomorrow be better? A Brief Analysis of SDG Network

Summary

The United Nations (UN) has set 17 Sustainable Development Goals (SDG) for the well-being of all humans. These 17 goals can have both positive and negative correlations. In this study, we tend to discuss these correlations and view these goals wholly as a flowing system, which can be influenced by entries and exits of SDGs and external disturbance.

First, the urgency of SDGs stands for node value of the network. We divide urgency into two indicators: the media exposure index and the keywords index. **TOPSIS evaluation model** is used in the calculation of media exposure index. We also obtain keyword index in UN news by the crawler

Second, bilateral correlation between SDGs stands for edge value of the network. We divide bilateral correlation into three components: urgency level, correlation of indicators, and semantic similarity. We use 17 target representative timeline sequence data to calculate correlation. Furthermore, **Canonical Correlation Analysis (CCA) method** is used to analyze multidimensional timeline sequence data. As for semantic similarity, we use **sentence-transformers model, a powerful NLP method** to calculate cosine similarity to quantify this index. **Analytic Hierarchy Process(AHP)** is finally applied to obtain the weight of the three indexes and finally obtain bilateral correlation between each pair of SDG.

Third, we build up the fluid SDG network. Based on the node value and edge value of the network, we quantify the effect of our SDG network based on the urgency index and directional correlation between every pair of SDGs. Assuming that flow of the network occurs every month, we can list the prioritization of the 17 SDGs and what can be reasonable to achieve in the next 10 years if our priorities are initiated.

Lastly, entries and exits of SDGs can have an impact on the network. Due to the interaction between problems, some problems have been solved, and others still exist. Thus, we analyze the new priority order of the network after eliminating hunger. In addition, some new problems that may emerge in the future are found based on regression fitting curve of relevant indicators. We add new nodes to the network and perform new operations to analyze their impact. Some external disturbances including technological advances can impact system node. In this case, we again use the **sentence-transformers model** to quantify its influence on urgency level of SDGs. In use of our model, the United Nations can well analyze the correlation between problems and make better decisions.

Moreover, our network approach can help other organizations set priorities. First, we use **k-means algorithm** to divide countries into three categories (GINI index as an example). The classification result indicates that different strategies should by applied according to features of a country. Second, value of both node and correlations should be taken into consideration when it comes to prioritization issues.

Keywords: SDG, NLP, AHP model, TOPSIS, network

Contents

I. Introduction

1.1 Background

With the expansion of population and increasing excavation of resources, the concept of sustainable development has been in the limelight and focus of multiple strategic policies in many countries. Taking joint action to achieve sustainability around the globe is more urgent than ever.

In pursuit of a more sustainable future for all humans, the United Nations(UN) has set 17 Sustainable Development Goals (SDGs) to eradicate poverty, protect the planet and improve the lives and future of all people around the world. These 17 goals are not independent of each other, instead, there may be significant or negative correlations between these goals. As a result, countries face trade-off when moving the work of the UN forward and this can cause diverging results. Therefore, identifying potential prioritization and correlations among them is critical for national planning.

1.2 Literature Review

There has been a certain amount of research on the issue of prioritization and interactions between the 17 SDGs. For example, Pradhan, P. et al.[1] made use of both country and country-disaggregated data and multiple time-series to explore positive and negative correlations between SDGs. Allen et al.[2] did a case study for 22 countries in the Arab region and adopted a multi-criteria analysis decision framework to assess and prioritize SDGs from three dimensions. Yang, S. et al.[3] did a survey in 66 countries and grouped the answers into three macro-region to identify SDG priorities and SDG-ES relations.

However, most of the literature uses perceptual indicators to rank the 17 SDGs and doesn't dig into the objective interrelationships between them. Also, few study they rarely evaluate the importance of the 17 SDGs from a global level.

1.3 Our works

Based on the analysis of the problem, we construct the model framework shown in figure 1, which is mainly composed of the following parts:

Urgency Index Determination: By observing data on UN's official website, we complete the construction of media exposure index model with the help of **TOPSIS model.** Also, counts of occurrence are taken into consideration to decide on one SDG's urgency index.

Correlation Analysis: We assume that positive or negative correlations between a pair of SDGs are derived from three factors: Urgency, Correlation of Indicators, and Semantic Similarity. **Canonical Correlation Analysis (CCA) method** is used to quantify correlations of indicators from multiple dimensions. **Sentence-transformers model**, a powerful **NLP method** is applied to analyze semantic similarity. **Finally, we use Analytic Hierarchy Process(AHP)** to obtain the weight of the three indexes and finally obtain bilateral correlations between each pair of SDG**.**

Network Flow Model: Based on the flow principle of Graph Theory, we quantify the effect of our SDG network based on the urgency index and directional correlation between every pair of SDGs. Assuming that flow of the network occurs every month, we can list the prioritization of the 17 SDGs and what can be reasonable to achieve in the next 10 years if our priorities are initiated.

Entry and Exits of SDGs: When one of the SDGs is achieved, prioritization will change. Based on extensive data collection, we assign Information Security problem and Aging Problem as

new SDGs. We make assumption about new SDGs's urgency index. Also, indicators are chosen to calculate the correlation of indicators for new SDGs. For semantic similarity, we use their definitions on Wikipedia to analyze. Then the established network flow model will reset the prioritization for us. For disturbance from the external environment, we again use **the sentence-transformers model** to quantify the influence that this disturbance has on urgency level of SDGs.

Inspiration for Prioritization Issue: First, we use **k-means algorithm** to divide countries with SDGs into three categories and take GINI index as an example. It can be concluded that different strategies should be applied in different countries. Second, positive and negative correlations should be taken into consideration when it comes to prioritization issues.

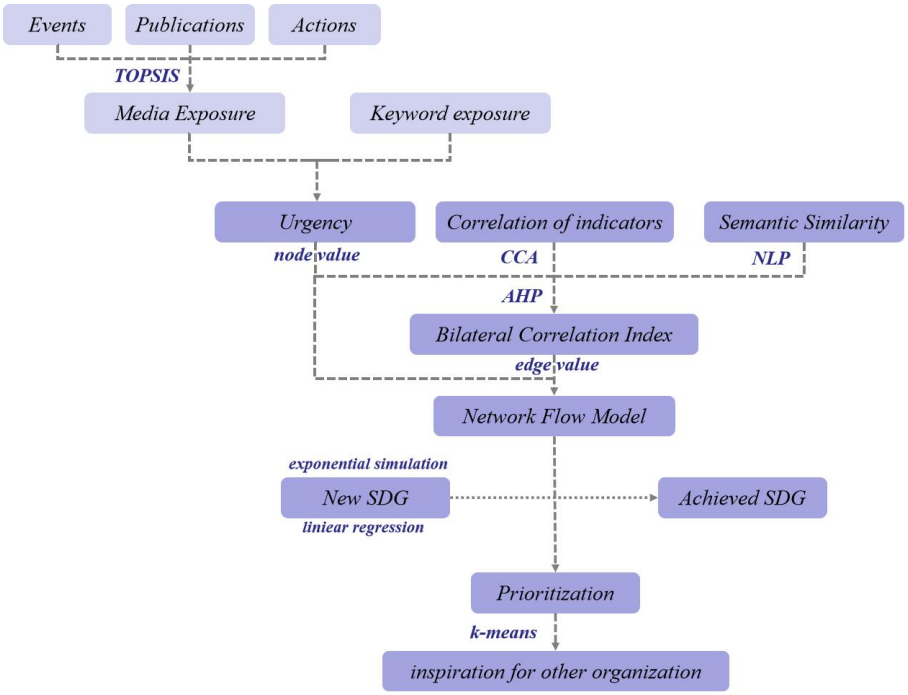

Figure1. structure of our work

II. Preparation of the Models

2.1 Basic Assumptions

Assumption 1: assume that the number of events, publications, and actions can reflect the SDG's urgency.

Reason: this assumption is made to ensure that media data disclosed on the official website can have an impact on the SDGs's urgency.

Assumption 2: assume that the data from the official website and authorized database is reliable.

Reason: this assumption is made to ensure the accuracy of the model construction and solution.

Assumption 3: assume that financial support from the UN is limited.

Reason: this assumption is based on the fact that fluid process of SDG network has funding limitations and other national and international priorities on its edges.

2.2 Notation

Symbols	*Definition*
m_i	Media exposure index of the i-th SDG
n_i	Keyword exposure index of the i-th SDG
u_i	Urgency index of the i-th SDG
$corr_{ij}$	correlation of indicators of the i-th SDG
s_{ij}	semantic similarity index between SDG-i and SDG-j
c_{ij}	The intensity of impact i-th SDG has on i-th SD
ur_i	The remaining sum of urgency index when SDG-i is supported first
si_{ij}	semantic similarity index between i-th external disturbance and SDG-j

III. Models

3.1 Urgency Index Determination

3.1.1 TOPSIS Evaluation Model for Media Exposure Index

TOPSIS method was first proposed by C.L.Hwang and K.Yoon in 1981. It is a method of compensatory aggregation that compares a set of alternatives, normalizing scores for each criterion and calculating the geometric distance between each alternative and the ideal alternative, which is the best score in each criterion.[4]

Next, we use TOPSIS method to evaluate importance of 17 SDGs from media data released on the official website of UN, which can be grouped into three types: number of events, number of publications and number of actions.

Step 1: Create and the decision matrix

Let $S = [S_1, S_2, \dots S_{17}]$ be the set of SDGs, $P = [P_1, P_2, P_3]$ be the set of evaluation indexes, $X = (x_{ij})_{m*n}$ be the decision matrix, and x_{ij} represents the value of the i-th SDG on the index j.

Step 2: Normalize evaluation matrix

To eliminate the influence of different indicator dimensions, we should first normalize the evaluation matrix before data analysis:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} {x_{ij}}^2}}$$

We call the dimensionless decision matrix $Z = (Z_{ij})_{m*n}$

Step 3: Evaluate based on the TOPSIS method

First, calculate the positive ideal solution Z^+ and the negative ideal solution Z^-:

$$Z^+ = (Z_1^+, Z_2^+, Z_3^+)$$
$$= (\max\{z_{11}, z_{21}, z_{31}\}, \max\{z_{12}, z_{22}, z_{32}\}, \dots, \max\{z_{1,17}, z_{2,17}, z_{3,17}\})$$
$$Z^- = (Z_1^-, Z_2^-, Z_3^-)$$

$$= (\min\{z_{11}, z_{21}, z_{31}\}, \max\{z_{12}, z_{22}, z_{32}\}, \ldots, \max\{z_{1,17}, z_{2,17}, z_{3,17}\})$$

Then, Calculate the distance between each sample and the optimal scheme D_i^+ and the worst scheme D_i^-:

$$\begin{cases} D_i^+ = \sqrt{\sum_{j=1}^{m} (Z_j^+ - z_{ij})^2} \\ D_i^- = \sqrt{\sum_{j=1}^{m} (Z_j^- - z_{ij})^2} \end{cases}$$

Finally, we can evaluate the degree of closeness between each sample and the optimal scheme and define it as the media exposure index m_i:

$$m_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

Let $MI = [m_1, m_2, \ldots, m_{17}]$. In this case, we record these 17 SDGs as $S_1, S_2, \ldots S_{17}$ respectively. Using the TOPSIS method, we can calculate $MI =$ [15.82845047, 16.45081949, 10.05025103, 17.69212932, 17.74135749, 48.25421134, 12.39294267, 42.58506202, 30.06107444, 95.02563012, 104.64372261, 10.746764, 31.50222957, 35.48243554, 9.91130964, 4.01674227, 197.22222222]. It can be seen from this result that among the 17 SDGs, S_{17} is of the greatest importance and S_{16} is of the least importance.

3.1.2 Keyword Exposure Index

Besides number of events, publications and actions published on UN's official website, we consider the attributes and composition of news content that UN released also crucial in deciding the importance of one SDG. Because number of times that relevant words appears can represent emphasis that UN lay on this SDG. To quantify this kind of emphasis, we use code to automatically grab text to on UN's news website obtain data in batches.

Specific steps are as follows:

First, we obtain 43 news reports published between October 2021 and February 2023 from UN's news website. Next, we choose 1-3 representative words for each SDG. The matching rules is shown in table 1:

Table 1 Representative Words for SDGs

SDG	*Representative Words*		
SDG1：NO POVERTY	poor	poverty	
SDG2：ZERO HUNGER	hunger	food	nutrition
SDG3：GOOD HEALTH AND WELL-BEING	health	well-being	
SDG4：QUALITY EDUCATION	educate+educational	school	
SDG5：GENDER EQUALITY	gender		
SDG6：CLEAN WATER AND SANITATION	sanitation	water	
SDG7：AFFORDABLE AND CLEAN ENERGY	energy		

SDG8：DECENT WORK AND ECONOMIC GROWTH	economic+economy+ economical	employment	
SDG9：INDUSTRY, INNOVATION AND INFRASTRUCTURE	industry+industrial	infrastructur e	
SDG10: REDUCED INEQUALITIES	equal+equality		
SDG11：SUSTAINABLE CITIES AND COMMUNITIES	city	settle	house+household
SDG12：RESPONSIBLE CONSUMPTION AND PRODUCTION	produce+product	consume+ consumer	
SDG13：CLIMATE ACTION	climate		
SDG14：LIFE BELOW WATER	fish	marine	
SDG15：LIFE ON LAND	animal	terrestrial	
SDG16：PEACE, JUSTICE AND STRONG INSTITUTIONS	peace	justice	
SDG17：PARTNERSHIPS FOR THE GOALS	partner		

Second, for each representative word, number of articles that this word appears is recorded. For i-th SDG, let the word that appeared the most time be R_I^* and number of times that it appears be n_i. The result is shown in table2:

Table 2 Count of Representative Words for SDGs

SDG	R_I^*	n_i
SDG1：NO POVERTY	poor/poverty	1
SDG2：ZERO HUNGER	food	6
SDG3：GOOD HEALTH AND WELL-BEING	health	9
SDG4：QUALITY EDUCATION	educate+educational	2
SDG5：GENDER EQUALITY	gender	1
SDG6：CLEAN WATER AND SANITATION	water	29
SDG7：AFFORDABLE AND CLEAN ENERGY	energy	5
SDG8：DECENT WORK AND ECONOMIC GROWTH	economic+economy+eco nomical	28
SDG9：INDUSTRY, INNOVATION AND INFRASTRUCTURE	industry+industrial	2
SDG10: REDUCED INEQUALITIES	equal+equality	37
SDG11：SUSTAINABLE CITIES AND COMMUNITIES	city	37
SDG12：RESPONSIBLE CONSUMPTION AND PRODUCTION	produce+product	3
SDG13：CLIMATE ACTION	climate	13
SDG14：LIFE BELOW WATER	marine	6
SDG15：LIFE ON LAND	animal/terrestrial	0
SDG16：PEACE, JUSTICE AND STRONG	peace	2

INSTITUTIONS		
SDG17：PARTNERSHIPS FOR THE GOALS	partner	37

3.1.3 Urgency Index Determination

After quantifying the 17 SDGs' importance in both media data and news context, we can take the two index together into consideration.

$$m_i^{'} = \frac{m_i}{max\{m_1, m_2, \ldots, mi_{17}\}}$$

$$n_i^{'} = \frac{n_i}{max\{n_1, n_2, \ldots, n_{17}\}}$$

Define me_i as the media Urgency Index of the i-th SDG:

$$u_i = m_i^{'} + n_i^{'}$$

Finally, we can calculate $UI =$ [15.82845047, 16.45081949, 10.05025103, 17.69212932, 17.74135749, 48.25421134, 12.39294267, 42.58506202, 30.06107444, 95.02563012, 104.64372261, 10.746764 , 31.50222957, 35.48243554, 9.91130964, 4.01674227, 197.22222222].

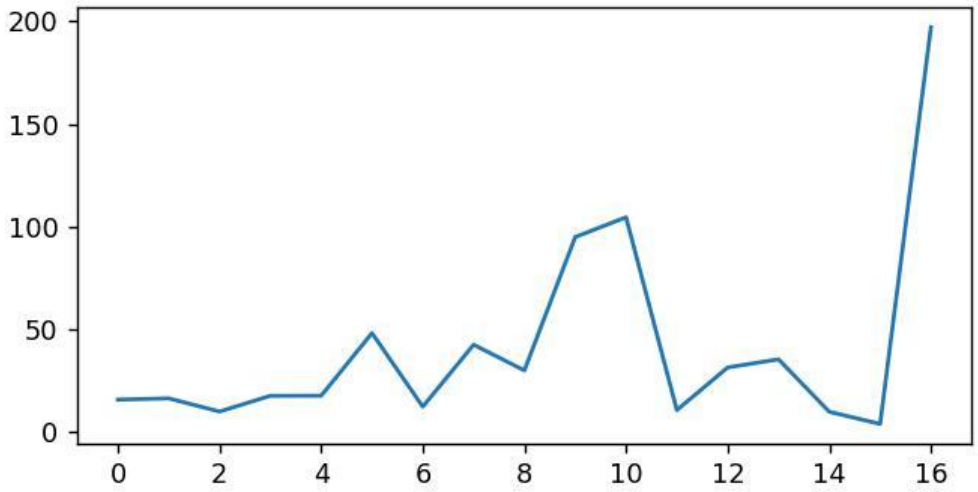

Figure2. urgency index of SDGs

It can be seen from this result that among the 17 SDGs, SDG17 is of the greatest urgency and SDG16 is of the least urgency. And the urgency index of all SDGs are between 0 and 200.

3.2 Correlation Analysis

3.2.1 Urgency Index of SDGs

For each pair of SDG, we think their interactions rely not just on how strong their connection is. Urgency of the two SDGs themselves also exerts influence on the process of interaction. The more urgent one SDG is, the more influence it will exert on the other SDG. Therefore, in terms of urgency of the two SDGs, we assume that the unidirectional correlation(positive or negative) that *i*-th SDG exerts on *j*-th SDG be c_{ij}.

3.2.2 Correlation of Indicators

In the case of interactions between SDGs, we conduct simple linear regression to calculate the correlation between representative indexes of each SDG.

First, to quantify the degree of completion and the trend of each SDG in recent years, we designate representative index. The matching rules are shown in the table3:

Table3 SDG indicator

SDG	*Indicator*	*Units*	*+/-*
SDG1：NO POVERTY	15 a day (2017 PPP)	% of population	+
SDG2：ZERO HUNGER	Number of people with insufficient food	/	+
SDG3：GOOD HEALTH AND WELL-BEING	Life expectancy at birth, total	(years)	-
SDG4：QUALITY EDUCATION	Primary completion rate, total	(% of relevant age group)	-
SDG5：GENDER EQUALITY	CPIA gender equality rating	(1=low to 6=high)	-
SDG6： CLEAN WATER AND SANITATION	Renewable internal freshwater resources per capita	(cubic meters)	-
SDG7：AFFORDABLE AND CLEAN ENERGY	Access to electricity	(% of population)	-
SDG8：DECENT WORK AND ECONOMIC GROWTH	Employment to population ratio, 15+, total	(modeled ILO estimate)	-
SDG9:INDUSTRY, INNOVATION AND INFRASTRUCTURE	Industrial design applications, resident, by count		-
SDG10: REDUCED INEQUALITIES	Gini Index		+
SDG11：SUSTAINABLE CITIES AND COMMUNITIES	Households and NPISHs Final consumption expenditure	(current US$)	-
SDG12： RESPONSIBLE CONSUMPTION AND PRODUCTION	CPIA policy and institutions for environmental sustainability rating	(1=low to 6=high)	-
SDG13：CLIMATE ACTION	Climate Change	(CO2 emissions (kt))	+
SDG14：LIFE BELOW WATER	Marine protected areas	(% of territorial waters)	-
SDG15：LIFE ON LAND	Terrestrial protected areas	(% of total land area)	-
SDG16：PEACE, JUSTICE AND STRONG INSTITUTIONS	Refugee population by country or territory of asylum		+
SDG17：PARTNERSHIPS FOR THE GOALS	Exports of goods and services	(current US$)	-

Then, we calculate correlation between each pair of single factors. The result is shown in table4:

Table4 The Result of single indicator correlation

	No poverty	Zero Hunger	Good Health	Quality Education	Gender Equality	Sanitation	Energy	Economic Growth	Infrastructure	Inequalities	Sustainable cities	Consumption
No poverty	1.000000	-0.554237	NaN	0.840278	NaN	-0.916093	0.982959	NaN	0.821131	0.775016	0.818643	0.795611
Zero Hunger	-0.554265	1.000000	NaN	-0.596824	-0.657653	0.663834	-0.611072	0.968819	-0.754354	NaN	NaN	-0.818587
Good Health	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.716441	NaN
Quality Education	0.840277	-0.148719	NaN	1.000000	0.814081	-0.946849	0.839081	-0.606499	0.715249	NaN	NaN	0.839816
Gender Equality	NaN	-0.657653	NaN	0.814081	1.000000	-0.734743	NaN	NaN	0.856124	NaN	NaN	0.892271
Sanitation	-0.916093	0.604635	NaN	-0.946849	-0.734743	1.000000	-0.952194	0.701986	-0.623029	-0.800565	NaN	-0.712415
Energy	0.982958	-0.611072	NaN	0.839083	NaN	-0.952194	1.000000	NaN	0.761017	0.645901	0.809803	0.842059
Economic Growth	NaN	0.968819	NaN	-0.609080	NaN	0.701986	NaN	1.000000	NaN	NaN	NaN	-0.662209
Infrastructure	0.821132	-0.754359	NaN	0.716016	0.856124	-0.623029	0.761017	NaN	1.000000	0.357817	NaN	0.801585
Inequalities	0.775016	NaN	NaN	NaN	NaN	-0.800584	0.645901	NaN	0.357821	1.000000	0.876215	0.668326
Sustainable cities	0.818638	NaN	0.716452	NaN	NaN	NaN	0.809803	NaN	NaN	0.876215	1.000000	NaN
Consumption	0.795642	-0.818651	NaN	0.846005	0.892271	-0.712459	0.842059	-0.662209	0.801586	0.668326	NaN	1.000000
Climate	-0.640707	0.943814	NaN	-0.655447	-0.604565	0.890195	-0.764749	0.654197	-0.846353	NaN	NaN	-0.756511
Life Under Water	0.928604	-0.632389	NaN	0.708804	NaN	-0.888360	0.924297	-0.100333	NaN	0.747027	NaN	0.692503
Life on Land	0.958414	-0.757569	NaN	0.702137	NaN	-0.752492	0.926194	-0.193797	0.403268	0.727207	0.686242	0.733574
good governance	-0.889373	NaN	NaN	NaN	NaN	0.764619	-0.899073	NaN	NaN	-0.484697	-0.859268	NaN
Partnerships	NaN	NaN	0.815099	NaN	NaN	NaN	NaN	NaN	NaN	0.688689	0.932964	NaN

Second, in order to make the results more credible, we expand the dimension of the indicators into two and use the **Canonical Correlation Analysis (CCA) method** to calculate correlation between multiple factors. The specific steps are as follows:

Step1: Calculate the variance of variable X as σ_x^2, the variance of variable Y as σ_y^2, and the covariance between X and Y as S_{xy}.

Step2: Calculate the matrix:

$$M = S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2}$$

Step3: Perform singular value decomposition on matrix *M* to obtain maximum singular value ρ and correspondent left and right singular vector.

Step4: Calculate the linear coefficient vectors a and b of X and Y:

$$a = S_{XX}^{-1/2} u, b = S_{YY}^{-1/2} v$$

In this case, we use CCA to analyze every pair of SDG indicators. Part of the result is shown in table5:

Table5 The Result of CCA Analysis

	No poverty	Zero Hunger	Good Health	Quality Education	Gender Equality	Sanitation	Energy	Economic Growth	Infrastructure	Inequalities	Sustainable cities	Consumption	Climate
No poverty	1.000000	0.967583	0.991480	0.999887	0.997251	0.996827	0.998202	0.998606	0.999579	0.974590	0.998007	0.992802	0.961660
Zero Hunger	0.967589	1.000000	0.984192	0.941411	0.978698	0.960893	0.974412	0.999970	0.992215	0.970554	0.996967	0.990408	0.999865
Good Health	0.991480	0.984192	1.000000	0.993896	0.992348	0.997642	0.973539	0.993423	0.971771	0.996571	0.991650	0.919491	0.974579
Quality Education	0.999887	0.791305	0.993896	1.000000	0.991922	0.998014	0.996046	0.953905	0.971718	0.923729	0.999465	0.982482	0.953707
Gender Equality	0.997251	0.978698	0.992348	0.991922	1.000000	0.999710	0.999996	0.946774	0.995361	0.927075	0.998540	0.994423	0.979346
Sanitation	0.996827	0.949739	0.997642	0.998014	0.999710	1.000000	1.000000	0.976553	0.959996	0.999862	0.999712	0.969202	0.988027
Energy	0.998202	0.974412	0.973539	0.996046	0.999942	1.000000	1.000000	0.958842	0.984686	0.954884	0.999969	0.997233	0.979827
Economic Growth	0.998606	0.999970	0.993423	0.954411	0.946773	0.976553	0.958843	1.000000	0.972214	0.654277	0.999886	0.985407	0.958764
Infrastructure	0.999579	0.992216	0.971771	0.971848	0.995361	0.959996	0.984686	0.972214	1.000000	0.917186	0.975351	0.985727	0.996472
Inequalities	0.974590	0.970554	0.996571	0.923729	0.927075	0.999865	0.954884	0.654277	0.917187	1.000000	0.990900	0.985512	0.930728
Sustainable cities	0.998006	0.996967	0.991652	0.999465	0.998540	0.999711	0.999969	0.999886	0.975351	0.990900	1.000000	0.921239	0.992309
Consumption	0.992807	0.990418	0.919492	0.983384	0.994423	0.969210	0.997233	0.985407	0.985727	0.985512	0.921239	1.000000	0.982590
Climate	0.961660	0.999865	0.974579	0.953707	0.979346	0.988027	0.979827	0.958764	0.996471	0.930710	0.992309	0.982588	1.000000
Life Under Water	0.999018	0.983212	0.962373	0.988943	0.997511	0.996179	0.999822	0.784299	0.871058	0.985610	0.998556	0.997232	0.988927
Life on Land	0.999962	0.998004	0.974394	0.978927	0.990511	0.971592	0.996916	0.848950	0.933086	0.977907	0.997420	0.990829	0.939682
good governance	0.995870	0.979785	0.998039	0.957746	0.997986	0.995796	0.999728	0.987407	0.873067	0.925696	0.999773	0.867306	0.999111
Partnerships	0.999954	0.987695	0.992241	0.999877	0.996814	0.996315	0.997045	0.989187	0.998504	0.987668	0.999991	0.981245	0.960245

Let the single indicator correlation between SDG-i and SDG-j be $corrs_{ij}$ and the CCA result of SDG-i and SDG-j be $corrc_{ij}$. To reduce the absolute value of the data and eradicate the impact of different dimensions, let $corr_i = corrs_{ij} * corrc_{ij}$. The result is shown in table6:

Table6 Correlation of Indicators

	No poverty	Zero Hunger	Good Health	Quality Education	Gender Equality	Sanitation	Energy	Economic Growth	Infrastructure	Inequalities	Sustainable cities	Consumption
No poverty	1.000000	-0.554237	NaN	0.840278	NaN	-0.916093	0.982959	NaN	0.821131	0.775016	0.818643	0.795611
Zero Hunger	-0.554265	1.000000	NaN	-0.596824	-0.657653	0.663834	-0.611072	0.968819	-0.754354	NaN	NaN	-0.818587
Good Health	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.716441	NaN
Quality Education	0.840277	-0.148719	NaN	1.000000	0.814081	-0.946849	0.839081	-0.606499	0.715249	NaN	NaN	0.839816
Gender Equality	NaN	-0.657653	NaN	0.814081	1.000000	-0.734743	NaN	NaN	0.856124	NaN	NaN	0.892271
Sanitation	-0.916093	0.604635	NaN	-0.946849	-0.734743	1.000000	-0.952194	0.701986	-0.623029	-0.800565	NaN	-0.712415
Energy	0.982958	-0.611072	NaN	0.839083	NaN	-0.952194	1.000000	NaN	0.761017	0.645901	0.809803	0.842059
Economic Growth	NaN	0.968819	NaN	-0.609080	NaN	0.701986	NaN	1.000000	NaN	NaN	NaN	-0.662209
Infrastructure	0.821132	-0.754359	NaN	0.716016	0.856124	-0.623029	0.761017	NaN	1.000000	0.357817	NaN	0.801585
Inequalities	0.775016	NaN	NaN	NaN	NaN	-0.800584	0.645901	NaN	0.357821	1.000000	0.876215	0.668326
Sustainable cities	0.818638	NaN	0.716452	NaN	NaN	NaN	0.809803	NaN	NaN	0.876215	1.000000	NaN
Consumption	0.795642	-0.818651	NaN	0.846005	0.892271	-0.712459	0.842059	-0.662209	0.801586	0.668326	NaN	1.000000
Climate	-0.640707	0.943814	NaN	-0.655447	-0.604565	0.890195	-0.764749	0.654197	-0.846353	NaN	NaN	-0.756511
Life Under Water	0.928604	-0.632389	NaN	0.708804	NaN	-0.888360	0.924297	-0.100333	NaN	0.747027	NaN	0.692503
Life on Land	0.958414	-0.757569	NaN	0.702137	NaN	-0.752492	0.926194	-0.193797	0.403268	0.727207	0.686242	0.733574
good governance	-0.889373	NaN	NaN	NaN	NaN	0.764619	-0.899073	NaN	NaN	-0.484697	-0.859268	NaN
Partnerships	NaN	NaN	0.815099	NaN	NaN	NaN	NaN	NaN	NaN	0.688689	0.932964	NaN

3.2.3 Semantic Similarity Index

Since the UN has given clear definition of the 17 SDGs, it is of great importance to dig out the similarity between these definitions. And here we accomplish this task by applying the **sentence-transformers model**, a powerful NLP method for text approximation analysis. It maps sentences & paragraphs to a 384 dimensional dense vector space and can be used for tasks like clustering or semantic search.[5] After importing this model, we can calculate the **Cosine Similarity** between two targeted context. Higher cosine similarity characterize higher similarity between these two words. For example, if we want to quantify the semantic similarity between word v and word w:

$$cosine(v,w) = \frac{v * w}{|v||w|} = \frac{\sum_{i=1}^{N} v_i * w_i}{\sqrt{\sum_{i=1}^{N} d_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

In this case, we can calculate the cosine similarity between definitions of every two SDGs, the result is shown in figure3:

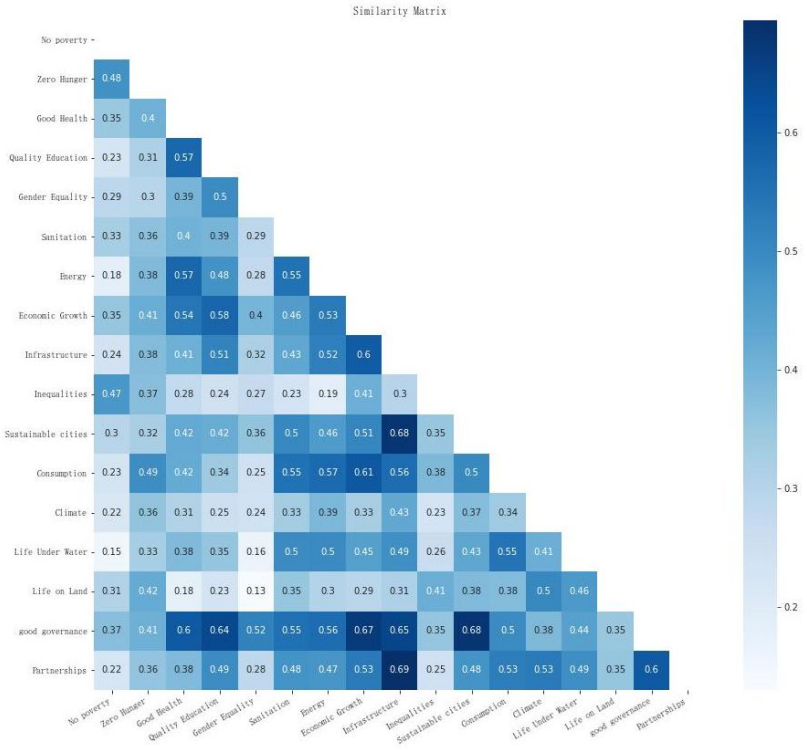

Figure3. semantic similarity between SDGs

3.2.4 Bilateral Correlation Index

With media exposure index, correlation of indicators and semantic similarity index prepared, we can now take the **Analytic Hierarchy Process (AHP)** to obtain the weight of the three indexes.

Step1: decide on the objective, criterion and plan of AHP analysis.

In this case, we choose the comprehensiveness, pertinence, and accuracy of the three indicators as the criterion. The analysis structure is shown in figure4:

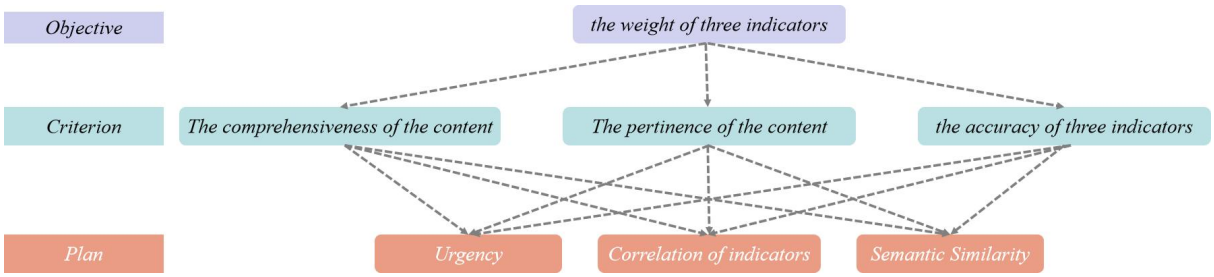

Figure4. semantic similarity between SDGs

Step2: build the decision matrix and complete the consistency check

Determine the importance score under each evaluation criterion, and the importance score of the evaluation criterion:

Table7: the decision matrix for AHP analysis

	comprehensiveness	***pertinence***	***accuracy***
comprehensiveness	1	1	1

pertinence	1	1	1
accuracy	1	1	1

Then we will do the consistency check for this decision matrix. Define the maximum eigenvalue of this matrix as γ and number of dimensions in this matrix as *n*.

Calculate CI, RI and CR:

$$CI = \frac{\gamma - n}{n - 1}$$

$$RI = 0.52$$

$$CR = \frac{CI}{RI} = 0 < 0.1$$

Therefore, the consistency of the decision matrix is acceptable.

Step3: build the scoring matrix

Based on the three indicators, we can give each of the three criterion a scoring matrix:

Table8: the scoring matrix

Comprehensiveness	***Urgency***	***Correlation of Indicators***	***Semantic Similarity***
Urgency	1.0	0.5	0.25
Correlation of Indicators	2.0	1.0	0.5
Semantic Similarity	4.0	2.0	1.0
Pertinence	***Urgency***	***Correlation of Indicators***	***Semantic Similarity***
Urgency	1.0	2.0	2.0
Correlation of Indicators	0.5	1.0	1.0
Semantic Similarity	0.5	1.0	1.0
Accuracy	***Urgency***	***Correlation of Indicators***	***Semantic Similarity***
Urgency	1.0	2.0	2.0
Correlation of Indicators	0.5	1.0	1.0
Semantic Similarity	0.5	1.0	1.0

Step4: decide the weight for each indicator

Based on comprehensive consideration of three evaluation criteria, we can get weight for each of the three indicators:

Table9: result of AHP analysis

indicator	***weight***
Urgency	0.380952
Correlation of Indicators	0.261905
Semantic Similarity	0.357143

Let c_{ij} be the intensity of impact i-th SDG has on j-th SDG, we define c_{ij} as a weighted sum of urgency index, correlation of indicators and semantic similarity index:

$$c_{ij} = 0.38 * u_i + 0.26 * corr_{ij} + 0.36 * s_i$$

The result is shown in figure5:

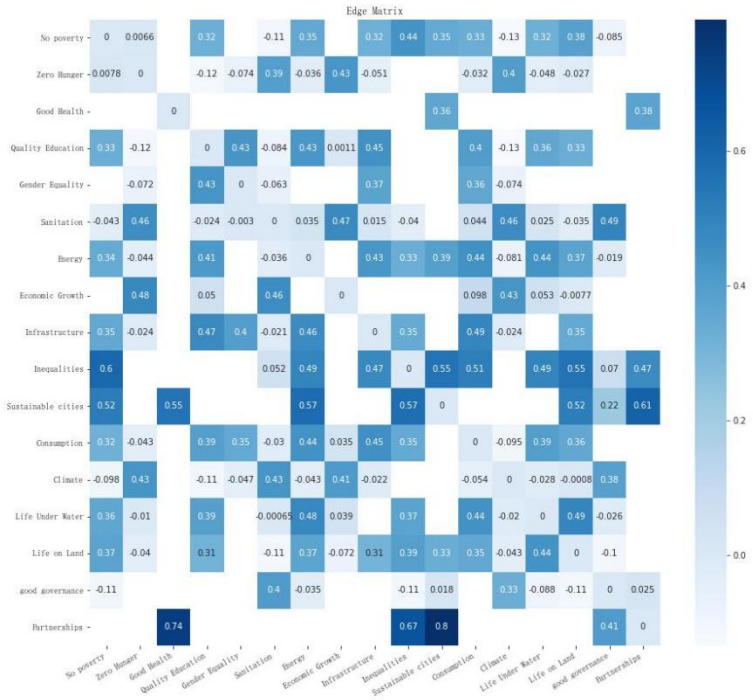

Figure5. Bilateral Correlation Index

In order to screen for pairs of SDGs that are significantly connected, we removed SDG pairs whose single indicator index is less than 0.7. This screening mechanism is displayed in the figure as blank areas.

As can be seen from the figure, there are several strong positive correlation that deserve attention: SDG-3→SDG-1；SDG10→SDG1；SDG11→SDG1. Among them, SDG-11 exerts the most positive influence on SDG1.

Representative synergy pairs ***Representative trade-off pairs***

Figure6. Representative pairs of SDGs

Finally, we can create a network of the relationships between the 17 SDGs:

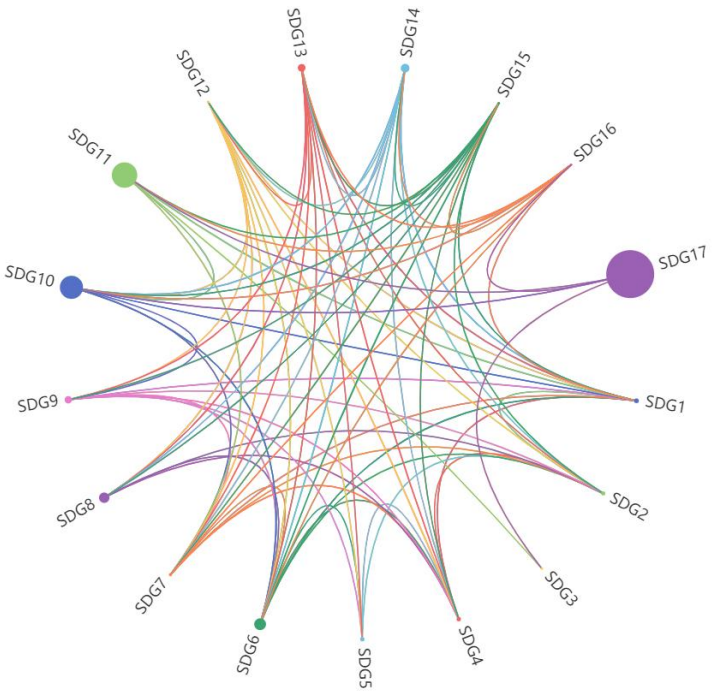

Figure7. a network of the relationships between the 17 SDGs.

3.3 Network Flow Model

3.3.1 Model Preparation

Define u_i^t as the remaining u_i of the i-SDGi at the beginning of the t-th month, Δu_i^t as the variation of u_i in the t-th month:

$$\Delta u_i^{t+1} = \sum_{j=1}^{i-1} \Delta u_j^t * c_{ji} + \sum_{j=i+1}^{i-17} \Delta u_j^t * c_{ji}$$

For model preparation, we do the following processing:

① In case the network become a self-expansion system, for each SDG, if it has significant connections with j SDG, we divide its Bilateral Correlation Index by j.

② In consideration of funding limitations and other national and other international priorities, we assume that based on UN's financial resource in one year, at most 10% urgency of one SDG can be reduced（or sum of percentage of variation in all SDGs). To illustrate:

$$\sum_{i=1}^{17} \Delta u_i \leq 10\% \sum_{i=1}^{17} u_i$$

③ Interactions between SDGs happen at the beginning of every month.

④ In the first month, only the s-th SDG's urgency change. Therefore, result of the first interaction is that urgency the other 16 SDGs decrease $\Delta m_s^1 * c_{si}$ (i=1,2,...17) respectively.

⑤ When the urgency of one SDG is reduced to zero, we consider this SDG as achieved. The lower the remaining urgency is, the more effective this initial financial support is. In this case, we set the remaining urgency index sum after one year of flow as the benchmark for prioritization.

3.3.2 Model establishment

Assume that flow of the network start with SDG-s, calculate the remaining urgency index:

$$\begin{cases} u_i^2 = u_i^1 - \Delta u_s^1 * c_{si} \quad (j=17) & t = 2 \\ u_i^t = u_i^{t-1} - \sum_{j=1} \Delta u_j^1 * c_{ji} & t \geq 3 \end{cases}$$

3.3.3 Results

If the disturbance starts from SDG-i, let the remaining sum of urgency after one year be ur_i. Let $URI = [ur_1, ur_2, \ldots, ur_{17}]$. We can calculate $URI = [14.99187336,\ 16.50638271,\ 5.477729,\ 17.54859682,\ 17.70231155,\ 48.3228148,\ 11.54776622,\ 42.60703491,\ 29.73458498,\ 90.75971403,\ 98.8999935,\ 10.37228581,\ 31.53262244,\ 35.15015219,\ 9.10828852,\ 3.04341282,\ 175.66809528]$.

The result is shown in figure8:

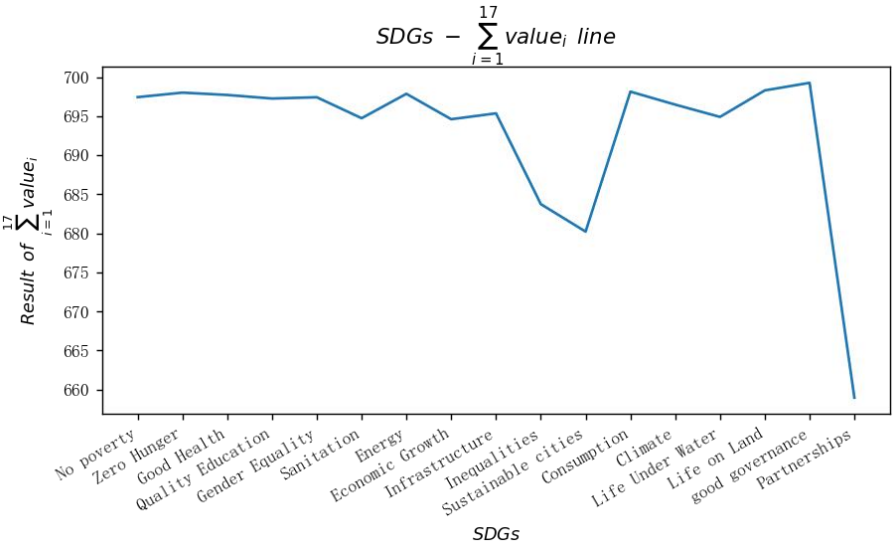

Figure8. remaining urgency index after one year

Second, we can quantify what can be reasonably achieved in 10 years. Let f_i be the financial support that SDG-i receives and f be the total financial support. We assume that financial support is distributed in proportion to the urgency index of one SDG:

$$f_i = \frac{u_i}{\sum_{i=1}^{17} u_i} * f$$

At the beginning of every year, we give an initial financial support to all SDG. After that, the flow of the network system happens once at the beginning of the month. As a result, sum of the urgency index will decrease. The specific decreasing trend is shown in figure9:

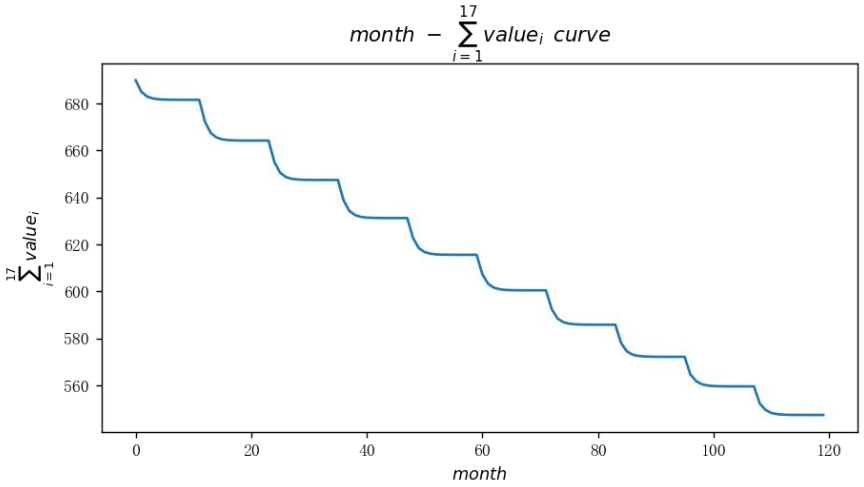

Figure9. changing trend of remaining urgency index in 10 years

It can be seen that in the following months, sum of urgency will automatically show a downward trend in this network system. To sum up , the resolution process presents a periodic decreasing trend.

3.4 Entry and Exit of SDGs

3.4.1 One SDG achieved

Under the flow of the system, some SDGs may be achieved. Take SDG1 as an example, when SDG1 is achieved, some SDGs have been achieved, and some SDGs is still urgent. The prioritization of the 17 SDGs after the elimination of hunger is shown in figure10:

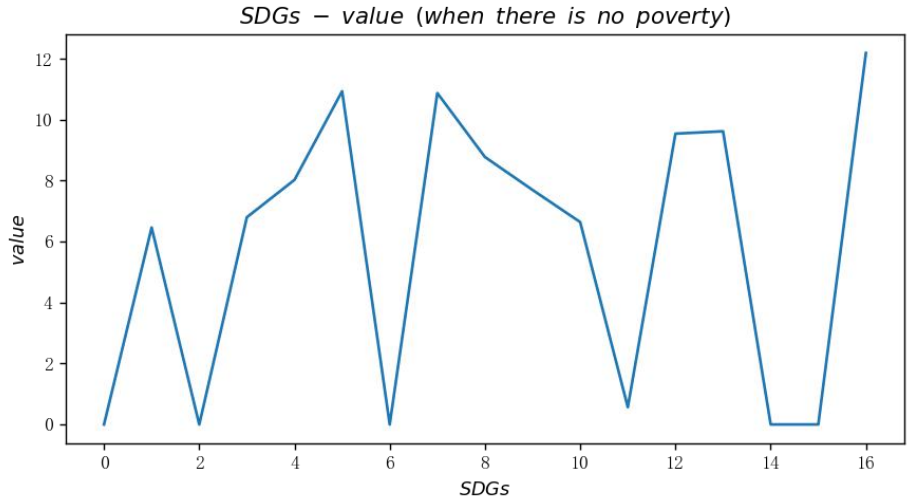

Figure10. urgency level when SDG1 is achieved

It can be seen that the prioritization has changed. SDG1, SDG2, SDG6, SDG11, SDG14, SDG15 is already achieved. Among the rest SDGs, the new prioritization is: SDG16, SDG5, SDG7, SDG13, SDG12, SDG8, SDG4, SDG3, SDG10, SDG1, SDG11.

3.4.2 Entry of new SDGs

Step 1： choose new SDG

Through extensive investigation and comparison, we choose two topics as the possible new SDGs in the future: **Information Security problem and Aging Problem**. The reason are as follows:

①These two topics are relatively new in the spectrum of sustainable development. With the popularization and development of the Internet, information security implicates not only personal information, but also a country's national defense system. Also, as the decline in world fertility continues, labor force in major counties will become a big issue.

② In spite of qualitative analysis, we also find quantitative evidence to prove these two emerging problem. To illustrate this, we choose indicator for the two new SDG and plot two keyword curves.

For information security problem, we choose annual number of ransomware attacks worldwide as the indicator:

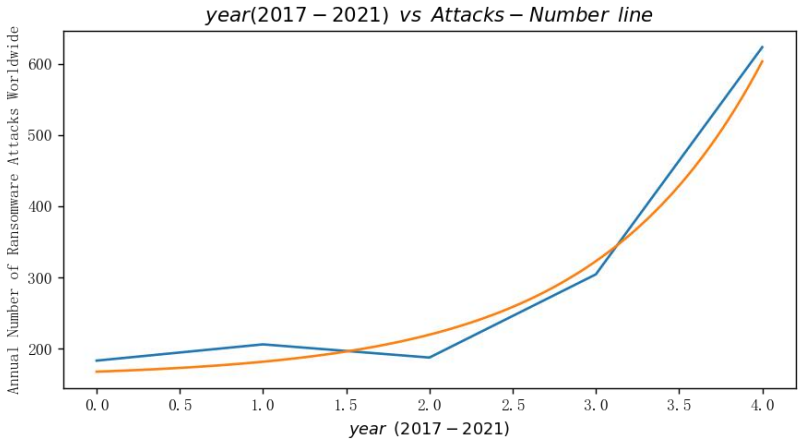

Figure11. Annual Number of Ransomware Attacks in 2017-2021

For aging problem, we choose population ages 65 and above (% of total population) as the indicator:

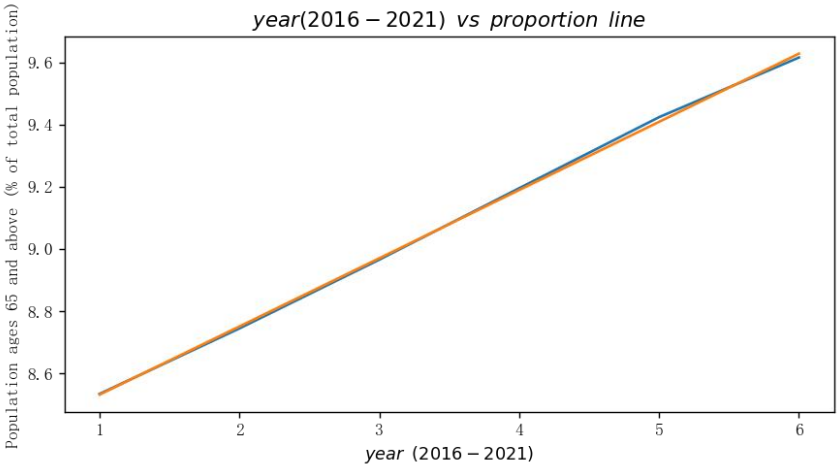

Figure12. Population ages 65 and above (% of total population) in 2017-2021

As can be seen from the figure, both indicators are proliferating in recent years, which proves our hypothesis on the new SDGs.

Step 2: calculate the urgency index

We believe that the UN's current SDGs is very comprehensive. The reason why some emerging problem is not included in current SDG list is that they are not urgent enough. Therefore, we set urgency indexes for the two new SDGs as the minimum urgency index of the initial 17 SDGs.

Step 3: calculate the bilateral correlation index

To calculate the bilateral correlation index, we replicate the model in 3.2 and the result in shown in figure13:

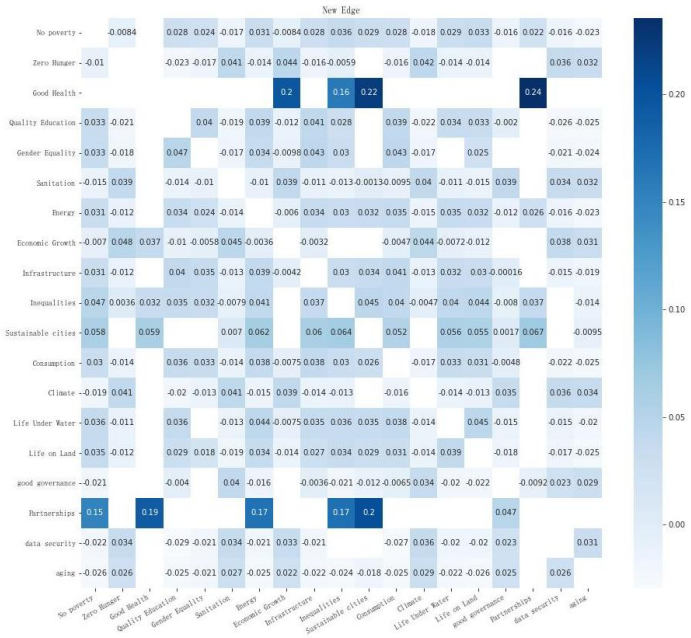

Figure13. Bilateral Correlation Index With New SDGs

Step 4: rebuild the network

With urgency level and bilateral correlation determined, we can plot the new network:

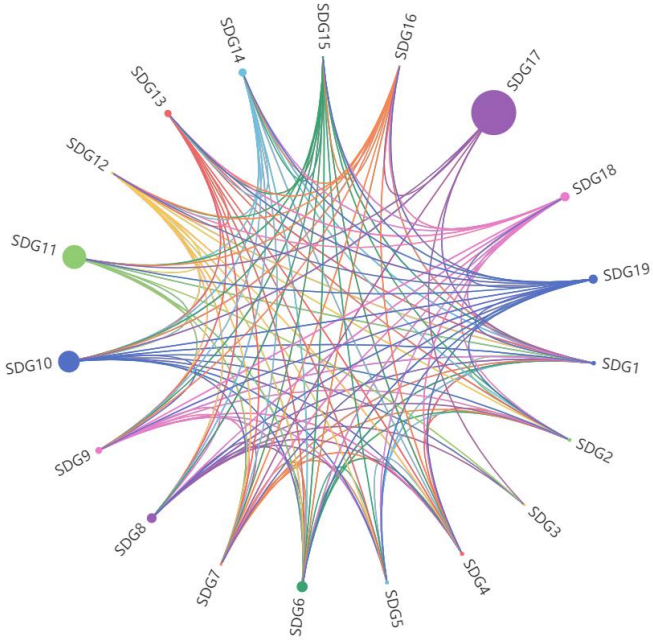

Figure14. a network of the relationships between the 17 SDGs (new SDG included)

3.4.3 Network under external disturbance

First, we list the six external impacts that may happen: technological advances, global pandemics, climate change, regional wars, refugee movements, international crises. We divided the six external disturbances into two classes: positive disturbance and negative disturbance. Define $u_i^{'}$ as the urgency index after external disturbance.

By applying the sentence-transformers model, we can calculate the cosine similarity between these six external disturbance and the 17 SDGs. Define semantic similarity index between i-th external disturbance and SDG-j as si_{ij}.

Positive disturbance includes technological advances. Calculate $u_i^{'}$:

$$u_i^{'} = u_i * (1 - si_{ij})$$

Negative disturbance includes global pandemics, climate change, regional wars, refugee movements and international crises. Calculate $u_i^{'}$:

$$u_i^{'} = u_i * (1 + si_{ij})$$

Take technological advances as an example. If technology is advanced significantly, the new prioritization is shown in figure15:

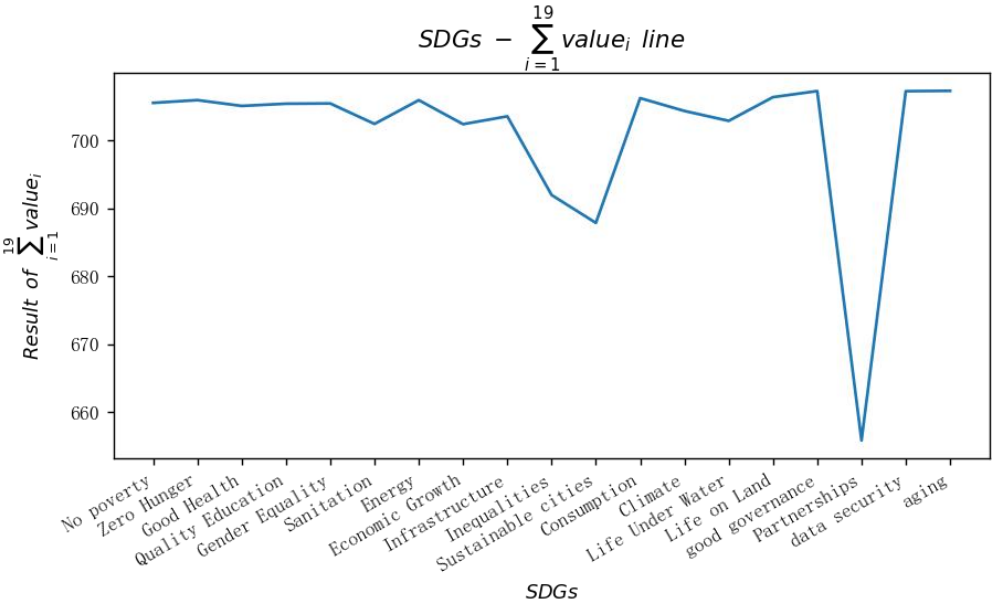

Figure15. a network of the relationships between the 17 SDGs (new SDG included)

It can be seen that both the structure of the network and the priority sequence has changed. External disturbance not only influence the prioritization of the 17 SDGs, but also change their urgency level.

3.5 Inspiration for Prioritization Issue

The most important issue when do prioritization is that both the urgency of the problem itself and connection between problems should be taken into consideration. Also, the connection can be bilateral.

Next, different areas may have different prioritization emphasis. Here we use K-means clustering method and GINI index as an example to illustrate this.

K-means clustering is a method of vector quantization that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.[6]

The existing data set C is composed of $c^{(1)}, c^{(2)}, \ldots, c^{(m)}$. the model needs to classify the data cluster into $k = 3$ clusters: $\{C\} = [C_1, C_2, C_3]$. Take GINI index as an example, the clustering result:

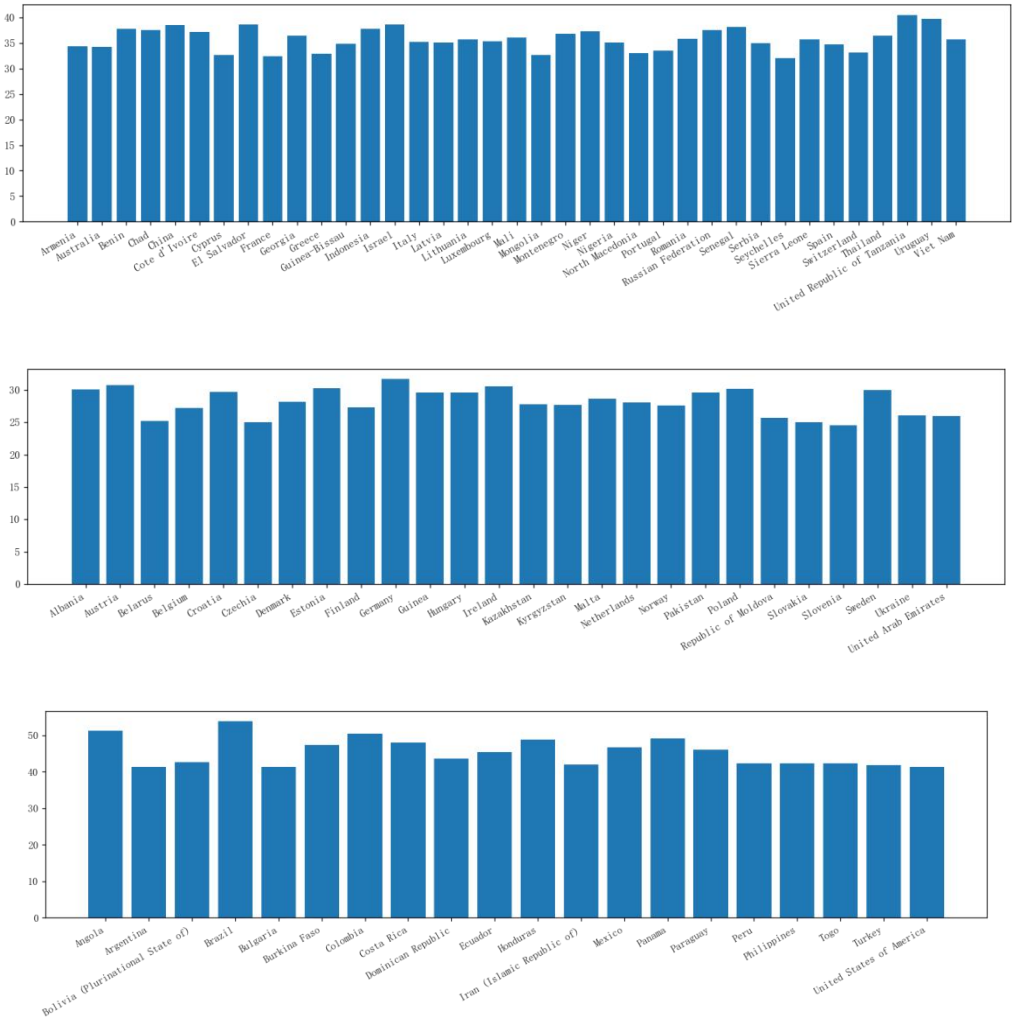

Figure16. Countries cluster bar chart

As shown in figure 16 , countries can be divided into three groups. This means that SDG indicators can demonstrate different attributes in different countries. If we want to help other companies and organizations set priorities of their goals, effect will be better if different characteristics of departments is explored and corresponding strategy is taken.

IV. Error Analysis and Sensitivity Analysis

4.1 Effectiveness Analysis

First, it can be made clear that SDG will hopefully be achieved one day because sum of the urgency index displays an overall declining trend.

Also, it should be noticed that our network system displays excessive effect after the initial financial support. To be more specific, the initial sum of urgency index is 699.61. after we apply the optimal strategy for achieving SDGs for one year (give the initial financial support in proportion to urgency level), the sum of urgency index is 658.97. It can be calculated that:

$$\sum_{i=1}^{i=17} im_i = 658.61$$

$$the\ percentage\ of\ reduction\ resulting\ from\ the\ network = \frac{699.61 - 658.97}{699.61} = 5.81\%$$

$$the\ percentage\ of\ human\ effort = \frac{197.22 * 10\%}{699.61} = 2.80\%$$

It is clear that $5.81\% > 2.80\%$. Therefore, conclusion can be made that the network system that we defines can have excess returns, which is also in line with reality.

4.2 Sensitivity Analysis

In the model above, we assume that at most 10% urgency of one SDG can be reduced（or sum of percentage of variation in all SDGs) in support of finance from the UN. To test the sensitivity of the initial financial support, now we let this percentage be 10%, 20%,..., 100% respectively. $\sum_{i=1}^{17} ur_i$ after one year of flow can be calculated:

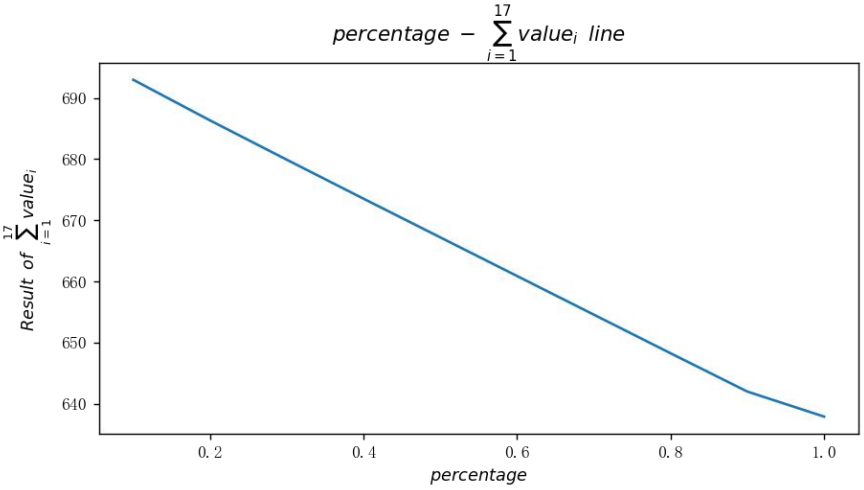

Figure17. sensitivity analysis of the initial financial support

It is obvious that increase of the financial support will reduce the urgency level. Moreover, it can be concluded that sum of the urgency index is almost linear with the percentage of urgency that initial financial support can reduce.

V. Strengths and Weaknesses

5.1 Strengths

The Bidirectionality of correlation between SDGs. We consider bilateral influence in case of network system building.

The authority and diversity of data. Instead of subjective results from questionnaire, We use data released officially which is very accurate and reliable and reflect objective international attitude. Also, for one SDGs, we mine multidimensional indicators to reach more reasonable conclusions.

The urgency of the SDG itself matters. We consider both the importance of SDG and correlation between SDGs when build up the model of the flow of the

5.2 Weaknesses

The limitation of time span. In this case, we only take year 2016-2020 into consideration due to limitation of data. If more data is available, more accurate and reliable correlations can be calculated.

VI. References

[1] Pradhan, P. et al. (2017) “A systematic study of Sustainable Development Goal (SDG) interactions,” Earth's Future, 5(11), pp. 1169–1179. Available at: https://doi.org/10.1002/2017ef000632.

[2] Allen, C., Metternicht, G. and Wiedmann, T. (2018) “Prioritising SDG targets: Assessing baselines, gaps and Interlinkages,” Sustainability Science, 14(2), pp. 421–438. Available at: https://doi.org/10.1007/s11625-018-0596-8.

[3] Yang, S. et al. (2020) “Prioritizing Sustainable Development Goals and linking them to ecosystem services: A global expert's knowledge evaluation,” Geography and Sustainability, 1(4), pp. 321–330. Available at: https://doi.org/10.1016/j.geosus.2020.09.004.

[4] *Topsis (2022) Wikipedia.* Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/TOPSIS (Accessed: February 21, 2023).

[5] Sentence-transformers/paraphrase-minilm-L6-V2 · hugging face (no date) sentence-transformers/paraphrase-MiniLM-L6-v2 · Hugging Face. Available at: https://huggingface.co/sentence-transformers/paraphrase-MiniLM-L6-v2 (Accessed: February 21, 2023).

[6] *K-means clustering* (2023) *Wikipedia*. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/K-means_clustering (Accessed: February 21, 2023).