A Scientific Network Model Around the 17 UN SDGs

Yipin Sun[1a,*], Jichi Li[2b], Rui Feng[3c]

{ Corresponding author[*]:syp211100@163.com[a], ljc1851503492@163.com[b], Fengrui7256@163.com[c]}

School of Mechanical Engineering, Hefei University of Technology, Hefei,230009, Anhui,China[1,2], School of Electrical and Automation Engineering, Hefei University of Technology, Hefei, 230009, Anhui, China[3]

Abstract.The United Nations (UN) has set 17 Sustainable Development Goals (SDGs). Achieving these goals will ultimately improve the lives of many people and promote economic development around the world. In order to find the relationship between these goals and explore the ways to achieve them, this paper constructs a scientific network model around the 17 UN Sustainable Development Goals (SDGs),providing reasonable solutions for achieving sustainable development goals.Our team innovatively construct a suitable network model and can iteratively calculate the accurate optimal solution, resulting in relatively accurate results.Meanwhile,an index system is constructed through entropy weight method to objectively evaluate the importance of the index.

Keywords:SDGs,network model,index sysdem, average clustering coefficient

1 Introduction

1.1 Restatements of Problems

The Sustainable Development Goals (SDGs) formulated by the United Nations (UN) have made great efforts to promote global sustainable development. The goal of sustainable development is not only a common problem and challenge faced by the whole world for a long time, but also a common market with global demand, and a list of actions aimed at promoting the well-being of all mankind. The goals of sustainable development include poverty eradication, zero hunger, clean water and sanitation, affordable clean energy, climate change and inequality, and many important problems faced by human social development. Achieving the goal of sustainable development marks a better sustainable future, but the priority of sustainable development plays an important role in the development process.

Sustainable development goals, as the focus of global attention, are the consensus reached by Community of Shared Future for Mankind. In this process, the realization of some goals will have positive or negative effects on the other goals. In order to promote the improvement of human life in the process of realizing these goals, we need to establish a mathematical model to solve the problems: what kind of influence exists among various sustainable development goals and how technological progress, global pandemics, climate change, regional wars and refugee flows hinder the above sustainable development goals.[1-5]

MSEA 2023, May 26-28, Nanjing, People's Republic of China

DOI 10.4108/eai.26-5-2023.2334469

1.2 Model assumptions

Due to the differences in network models, it is necessary to make the following assumptions to simplify the problem:

It is assumed that the relationship between goals is static and will not change over time. This assumption allows us to model the relationship between goals as a fixed network structure and measure it uniformly when analyzing past indicators.

Assume that the indicator data of each sustainable development goal is reliable and can be compared in different countries and regions. We also assume that every country uses the same standards and methods when collecting and reporting data to ensure the reliability and consistency of data.

It is assumed that each country's investment in sustainable development goals can be simplified to global investment. In this way, we can express the amount of capital investment as a numerical value, so that it is more convenient to compare and analyze the capital demand and distribution of different goals.

It is assumed that the priority between sustainable development goals is linear, that is, after one goal is completed, the next goal will be considered until all the goals are completed. In this way, we can clearly determine and arrange the priorities among different goals to guide policies and decisions.

It is assumed that a goal is completed when it reaches a certain threshold. In this way, we can define the completion criteria of each goal in more detail, so as to better track and evaluate the progress achieved and determine whether it is necessary to adjust policies and resource allocation.

1.3 Data sources and collection

Seventeen goals of the United Nations can be refined into more sub-indicators. At present, the United Nations has determined a total of 169 specific target to quantify the score of each country's seventeen goals. The 169 indicators of the United Nations Sustainable Development Goals are a huge and complex data set, which need to be collected and collated from multiple sources. We collected the response data of 169 indicators from 30 random countries in recent five years from official website (https://SDGs.UN.org/goals), the United Nations sustainable development goal. Some of the data were missing or abnormal, so we cleaned up the data and finally got three sub-indicators corresponding to each goal, totaling 50 sub-indicators.

For example, for poverty eradication, the following three complete indicators are selected(see Table 1):

Table 1 Three sub-indicators of table zero hungry

Zero hungry		
prevalence of stunting in children under 5 years of age (%, 2019)	prevalence of wasting in children under 5 years of age (%, 2019)	prevalence of obesity, bmi ≥ 30 (% of adult population, 2016)

And here are 30 countries randomly selected(see Table 2):

Table 2 randomly selected national samples

Germany	South Korea	Brazil	Indonesia	Serbia	Luxembourg
France	Canada	Argentina	Saudi Arabia	Gabon	Nauru
Britain	Australia	China	South Africa	Congo	Tonga
Japan	USA	Turkey	India	Qatar	New Zealand
Italy	Russian Federation	Mexico	Myanmar	Laos	SLibya

2 Network-based influence relationship

2.1 network construction

The following are the steps to build a relational network through the correlation coefficient matrix:

Step1:Prepare data

Collect 169 groups of variable data and calculate the correlation coefficient matrix between them.

Step2:Determine correlation threshold

Set a correlation threshold, and determine whether there is a strong relationship between nodes according to the value of correlation coefficient matrix. Generally, the absolute value of correlation coefficient greater than 0.5 is considered as a strong relationship.

Step3:Building a network

Variables are taken as nodes, correlation coefficients are taken as the weights of edges, and threshold filtering is used to determine whether the connections between nodes exist.

In order to ensure the accuracy and reliability of the relationship network we obtained, we need to carefully consider when selecting the correlation coefficient method. In this case, we choose to use Pearson correlation coefficient.Pearson correlation coefficient is calculated as follows(see equation (1)):

$$r = \frac{\sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y)}{\sqrt{\sum_{i=1}^{n}(X_i - \mu_X)^2 (Y_i - \mu_Y)^2}} \tag{1}$$

Where x and y are the values of two variables, μ_x and μ_y are the average values of two variables respectively.

The calculated correlation coefficient matrix is as follows(see Fig 1):

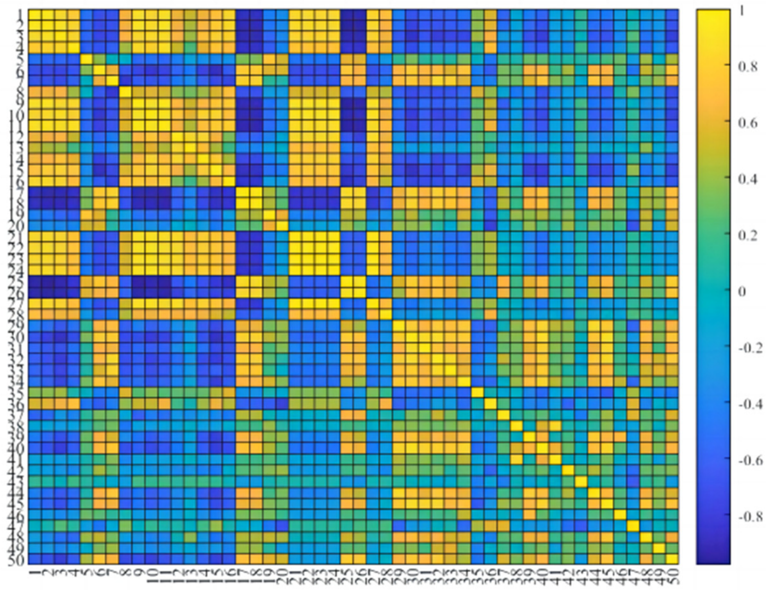

Fig 1 Thermal diagram of correlation coefficient of graph sub-index

Next, the correlation threshold is set, with 0.5 as the threshold, and the correlation coefficient r satisfies(see equation(2)):

$$|r| < 0.5 \tag{2}$$

In statistics, the absolute value of correlation coefficient r less than 0.5 is usually regarded as weak correlation or irrelevant, because this value shows that the linear relationship between variables is weak. Therefore, when calculating the correlation coefficient, if it is found that its absolute value is less than 0.5, it can be set to 0 to show that there is almost no linear relationship between the two variables(see Fig 2).

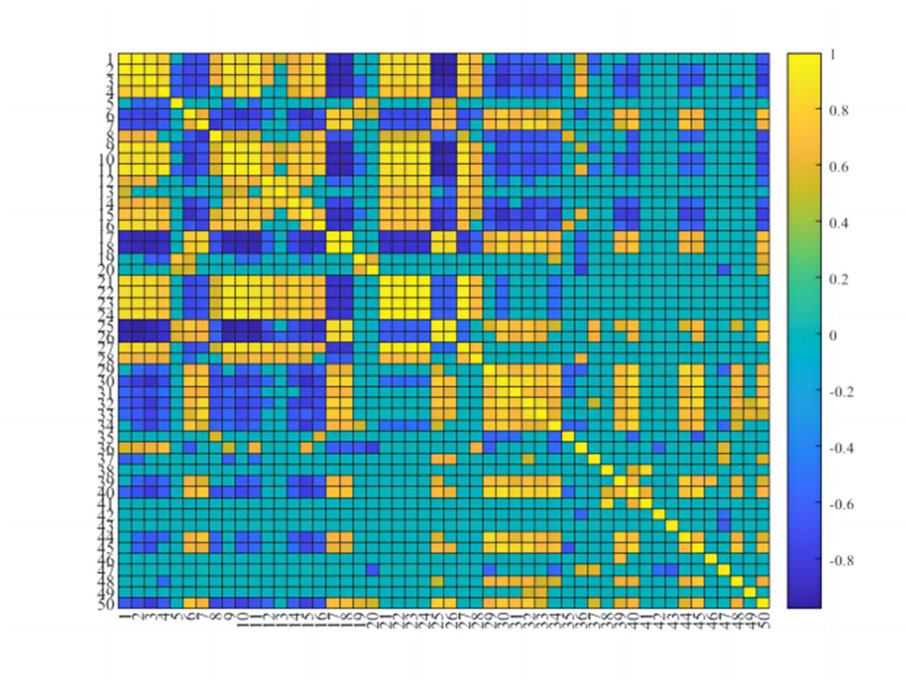

Fig 2 Thermal map of correlation coefficient after setting threshold

After getting the corresponding correlation coefficient matrix, we can start to build the sub-index relationship network of sustainable development goals. In this relational network, each node represents a sub-index of the sustainable development goal, and the size of the node represents the value of the index. Because the units of these indicators are different, we need to standardize them for comparison. There are edges between indicators, and each edge has a weight, which is equal to the correlation coefficient between indicators. If the weight is 0, it means that there is no edge connection between the two indicators. In this way, we can build a network of edges and nodes, which can help us understand the relationship between indicators and their contribution to the sustainable development goals(see Fig 3).

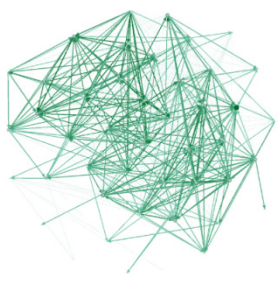

Fig 3 Relationship network between graph sub-indexes

2.2 network analysis

Eigenvector centrality is an index to evaluate the influence of a particular node in the network based on its connection with other nodes. However, this centrality does not simply calculate the number of connections of nodes, but reflects the relative importance of nodes in the network by considering the connection weight with other nodes. However, not every connection with other nodes has a fixed contribution to the centrality of the node, and the contribution of each connection to the centrality of the feature vector is proportional to the centrality of the connected node. This means that establishing connections with those nodes with higher feature vector centrality ranking has greater influence on the centrality ranking of the nodes than establishing connections with those nodes with lower feature vector centrality ranking. Therefore, when analyzing and comparing the influence of nodes in the network, it is necessary to comprehensively consider the centrality ranking of different nodes.

If we denote a graph with N nodes by G and $A = (a_{n,t})$ is an adjacency matrix, we have $(a_{n,t}) = 1$ when node n is connected to node t by an edge, otherwise we have $a_{n,t} = 0$. If we define$x_a$as the value of the eigenvector centrality of node A, then the value of the eigenvector centrality of node N is(see equation(3)):

$$X_n = \frac{1}{\lambda} \sum_{t \in M(n)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{n,t}\, x_t \tag{3}$$

Where λ is a constant, and m(n) represents the set of neighboring nodes of node n.

Table 3 Sub-indicators with the largest centrality of the first six feature vectors in the table

Subindex	Eigenvector Centrality
Subindex 11:Neonatal mortality rate (per 1,000 live births, 2020)	1
Subindex 3: Prevalence of stunting in children under 5 years of age	0.991779
Subindex 2: Poverty headcount ratio at $3.20/day	0.99071
Subindex 6: Human Trophic Level	0.987374
Subindex 15: Traffic deaths (per 100,000 population, 2019)	0.98389
Subindex 10: Neonatal mortality rate (per 1,000 live births, 2020)	0.980648

It can be seen that the mortality rate has a significant impact on health indicators because it reflects a region's or country's ability to protect citizens' health, and the mortality rate has a direct impact on people's quality of life and longevity. The mortality rate is strongly related to the level of medical care and the economic situation. A high mortality rate may indicate a lack of adequate medical resources and safeguard measures in these regions or countries, or it may indicate a low level of social and economic development. As a result, improving mortality rates should be a key goal of public health and socioeconomic development in any country or region(see Table 3).

The average clustering coefficient is a measure of the degree of node aggregation in social networks and other complex networks. It refers to the ratio of a node's actual connections to the possible connections, or to measuring the closeness of the social circle in which a node is located.

The average clustering coefficient is significant because it reflects the degree of aggregation of the social circle in which a node is located. Because it is easier for a node's neighbors to form close ties in a highly clustered network, its average clustering coefficient will be higher. On the other hand, if a network is widely dispersed, such a connection is unlikely to form, resulting in a lower average clustering coefficient.

The average clustering coefficient can also be used to examine the overall network characteristics. The average clustering coefficient of different networks, for example, can be compared to better understand their social structure and connectivity.

Table 4 The first six sub-indicators with the largest clustering coefficient in the table.

Subindex	Clustering coefficient
Subindex 35: Gini coefficient (2019)	1
Subindex 49: Corporate Tax Haven Score	1
Subindex 38: Production-based nitrogen emissions (kg/capita, 2015)	1
Subindex 41: CO_2 emissions embodied in fossil fuel exports (kg/capita, 2021)	1
Subindex 42: Mean area that is protected in terrestrial sites important to biodiversity (%, 2020)	1
Subindex 13: New HIV infections (per 1,000 uninfected population, 2020)	0.984848

From the standpoint of sustainable development, the Gini coefficient, tax haven score of enterprises, nitrogen emissions based on production, CO emissions reflected in fossil fuel exports, and the protected area of biodiversity-rich land sites are important indicators that constitute the center of clustering degree. These indicators correspond to reduced inequalities, goal-oriented partnerships, responsible consumption and production, climate action, and life below water, in that order. These core indicators reflect many aspects of sustainable development, including reduced inequalities, which focuses on closing the wealth gap and promoting social justice. Partnerships for the Goals focuses on multi-party cooperation to achieve shared sustainable development goals; Responsible consumption and production focuses on the sustainability of consumption and production methods; Climate action focuses on mitigating climate change; and Life Below Water focuses on protecting underwater biodiversity. As a result, the indicators of these cluster centers are critical factors in assessing a country's or region's level of sustainable development, as well as key indicators in achieving the goal of sustainable development(see Table 4).

3 Conclusions

In order to find the relationship between 17 SDGs, for this purpose we collected data of 17 sets of subindicators from the official website of the United Nations for 30 countries,found the importance of the subindicators by entropy weighting method, and constructed a relationship network among the sub-indicators by building a correlation coefficient matrix.

References

[1] Joleen T, Le B P , Hisham Z , et al. A policy nexus approach to forests and the SDGs: tradeoffs and synergies[J]. Current Opinion in Environmental Sustainability, 2018, 34:7-12.
[2] Sachs, J., Schmidt-Traub, G., Kroll, C., Durand-Delacre, D., and Teksoz, K. (2016). SDG Index and Dashboards – Global Report. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN). Available at: http://www.sdgindex.org/
[3] Organization W H . World health statistics 2016: monitoring health for the SDGs, sustainable development goals[M]. 2016.
[4] Organization W H . Health in 2015: from MDGs to SDGs[J]. 2015.
[5] WHO. Health in 2015, from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals[M]. 2015.