The Complex Interactions of the United Nations Sustainable Development Goals

Benjamin Kemp (ID: 52197910)
MSc Data Science, University of Aberdeen
2021/2022

Word Count: 4307

ABSTRACT Sustainable Development Goals (SDGs) have the aim of improving the planet and the lives of the people by 2030. SDGs cover a wide perspective of global issues facing many countries, allowing for a detailed plan to be developed for addressing these issues. We have focused on the interactions of the SDGs where a strong positive correlation between a pair of indicators is a synergy, and a strong negative correlation is a trade-off. Using network science to explore the interactions we use methods of community detection, eigenvector centrality and delayed correlation to explore the indicators further. Brazil and India both formed two distinct clusters of indicators based on correlation rather than their UN classification. Indicators with a large influence in synergies typically also had a strong influence in trade-offs and were found to connect the denser cluster together. Indicators with the greatest influence in trade-offs also had low influence in synergies and were trade-offs with most of the network such as total greenhouse gas emissions. Weak causal links were identified between funds for scholarships leading to more full-time researchers as well as a greater renewable energy capacity giving rise to more funds for infrastructure. Discovering the hidden structure of connections will allow for a quicker delivery of the sustainability goals.

1. Introduction

1.1. Sustainable Development Goals (SDGs)

The Sustainable Development Goals[1] are a group of 17 interlinked global goals designed to be a "blueprint to achieve a better and more sustainable future for all" set up in 2015 by the United Nations General Assembly (UN-GA) with the aim of being met by 2030[2]. The 17 SDGs (see Fig. 1) are (1) No Poverty, (2) Zero Hunger, (3) Good Health and Well-being, (4) Quality Education, (5) Gender Equality, (6) Clean Water and Sanitation, (7) Affordable and Clean Energy, (8) Decent Work and Economic Growth, (9) Industry, Innovation and Infrastructure, (10) Reduced Inequality, (11) Sustainable Cities and Communities, (12) Responsible Consumption and Production, (13) Climate Action, (14) Life Below Water, (15) Life on Land, (16) Peace, Justice and Strong Institutions, (17) Partnerships for the Goals.

Each goal then contains typically 8-12 targets which need to be delivered for the SDG to be achieved and each target has around 1-4 indicators which are used to measure progress towards that target. For example, SDG 1 is "End poverty in all its forms everywhere" which has, amongst others, Target 1.1 which is "By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day". One of the indicators to track progress towards this target, is Indicator 1.1.1 or the “Proportion of people living below the international poverty line”.

1.2. SDG Interactions

To analyse the interactions between SDGs it is vital to explore the synergies and trade-offs between indicator data. Links between indicators are classified as synergies when there is a strong positive correlation between them whilst strong negative correlations are classed as trade-offs. Pradhan[3] provides a quantification of synergies and trade-offs to be used in the study. The SusInfra (Sustainable Infrastructure) group carried out research *Towards Sustainable Sanitation in India and Brazil* (TOSSIB)[4] and *A systems approach towards sustainable sanitation challenges in urbanising China* (SASSI)[5]. With support from members of this group our research looked at a more general and network-based analysis to identify the key relationships among indicators which would allow policy makers to implement impactful changes to improve progress towards the 2030 agenda[2].

Figure 1. Sustainable Development Goals (source: United Nations)

2. Data and Methods

2.1. Data.

The United Nations Statistics Division provides data[6] on 247 indicators for monitoring the progress in achieving the SDGs (United Nations Economic and Social Council). Of these indicators there is time-series data for Brazil on 183 of them with 131 having at least 3 data points between 2000 and 2020. The time-series from these 131 indicators form the dataset considered for this work. I will however not use all of them, since I will only consider time-series with a statistically significant minimal number of points. These indicators are then split up, for example in terms of gender or age etc. leading to 967 disaggregated indicators and their respective time-series. All time-series used are kept at the highest level of disaggregation to avoid any bias or averaging of data.

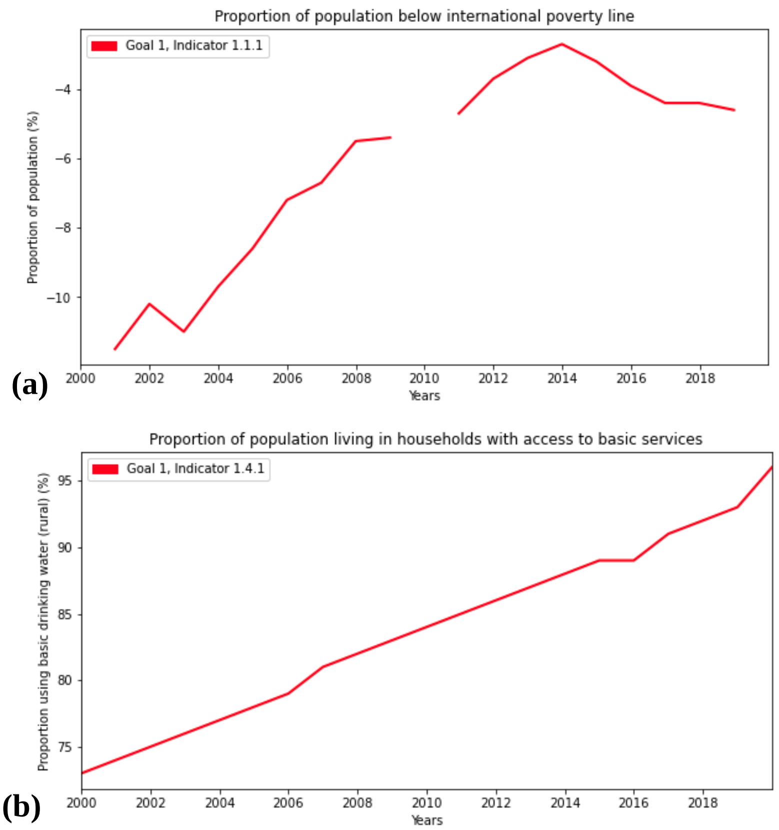


Figure 2. Two indicators from Brazil, with time series in (a) having been multiplied by -1 and in (b) original time-series.

2.2. Synergies and Trade-offs.

This analysis explored synergies and trade-offs in depth within Brazil and India. The initial correlation analysis is done on unique pairs of indicator time series using the country disaggregated data. Synergies occur when two indicators are positively correlated. However, for two indicators that are both getting worse they would be positively correlated and be classified as a synergy. We want to explore indicators improving so it is important to examine the data in depth to ensure that the synergy is correctly described.

Due to the nature of the time series (Fig. 2), for some indicators an increasing number signifies an improvement towards a target, such as indicator 1.4.1, "*Proportion of population living in households with access to basic services*". Whilst for others a decreasing number suggests an improvement in that target, such as indicator 1.1.1, "*Proportion of population living below the international poverty line*" both part of SDG 1 (End poverty in all its forms everywhere).

To be able to compare these directly by correlation the time series needs to be adjusted. This is done by applying a negative sign to any time series that requires a reduction in its value to signal an improvement, leading to any increase towards higher positive numbers showing an improvement in that indicator's performance. So now an increase in the proportion with access to basic services (Fig. 2b) and the negative of the proportion of population living below the international poverty line (Fig. 2a) will be classified as a positive correlation, since both time-series are now with a positive slope if it signifies an improvement.

2.3. Correlation.

Correlation describes the relationship between two variables, for different indicator pairs this relationship is visualised in Fig. 3. The angle of the line of best fit gives an idea of the direction of the relationship with the line pointing to the top right (positive slope) showing positive correlation while a line pointing to the bottom right (negative slope) will suggest negative correlation. How close the points are to this line of best fit then gives the strength of the correlation. For Pearson (linear) correlation, pairs that have all points directly on the straight line this will give a correlation of 1 or -1.

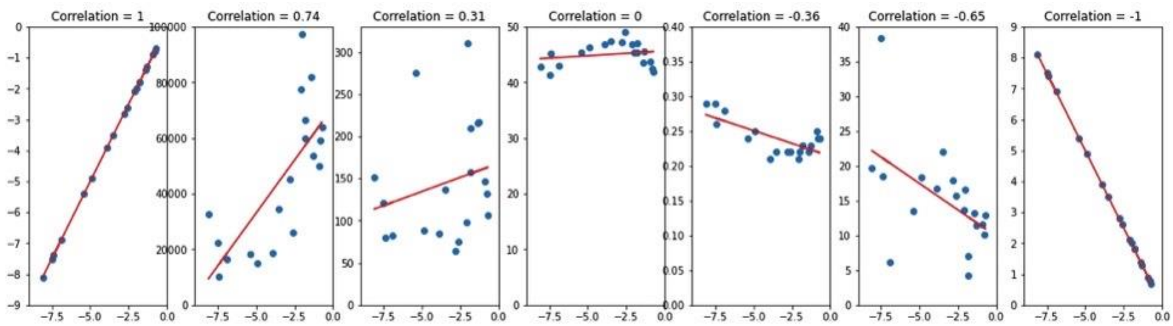


Figure 3. Blue filled points represent values from 2 time series. The red line is a linear fitting of the data. Panels show how the correlation between these two time series as a measure of linearity between them.

2.4. Spearman Correlation.

Spearman's rank-order correlation enables you to determine the strength and direction of a monotonic relationship between two time series. Spearman goes further than the Pearson product-moment correlation as it allows for measurements of non-linear relationships, as well as providing the correlation for data that is discrete. Pradhan[3] chose to perform correlation on only data pairs where there were more than 3 data points. While this may be fine for broad comparisons between country trends when wanting to investigate correlation further it is important to have as many data points as possible otherwise the result might be insignificant.

To determine the sample size requirements there are tables on the critical values of Spearman's Ranked Correlation Coefficient[7] to ensure significance. Typically, statistical tests classify significance to 5%, that is a measurement is significant if there is less than a 5% chance that the null hypothesis is correct (For a null hypothesis that there is no relationship between the variables). This equates to a p-value (α) of 0.05, while Pradhan[3] determined 0.6 (or -0.6) correlation to be the cut-off for classifying synergies (or trade-offs). Referring to the table[7] for a correlation of 0.6 or more and $\alpha = 0.05$, this requires a minimum 9 data points. To provide a safety net for correlation near 0.6 we will use 10 data points which ensures correlation over 0.564 can be determined as 5% significant.

For all analysis we will only consider time-series that contain a minimum number of 10 data points for correlation calculations. This represents a departure from the time series used in Pradhan[3], where all time series with at least 3 points were considered.

2.5. Network.

To further analyse beyond the scope that the TOSSIB[4] project has where the focus was on the pairwise analysis of synergies and trade-offs of pairs of indicators, we not only do similar pairwise analysis, but also evaluate the network global structure and examine the non-local connectivity within the network. Combining centrality measures with community detection can provide complementary insight into the impact important nodes have on their community and the whole network[8]. For that we focused on eigenvector centrality[9] and performing community detection[10]. To do this the network needs to be split into a synergy network and a trade-off network so that distinctions between the two can be made clear. We focused more on the synergy network as we wanted to identify the indicators that would provide the biggest improvement for a country. In the python Networkx[11] package the network is formed taking only connections between nodes that are above 0.6 correlation (for synergy) or below -0.6 correlation (for trade-off). The networks are displayed using a spring layout which places nodes closer together if they are more strongly correlated and further away when they are weakly correlated.

2.6. Eigenvector Centrality.

To determine the importance of a singular node in the network we can use eigenvector centrality[9,12]. The eigenvector centrality for the node v is obtained using the following approach that considers the eigenvalue equation:

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}, \quad (1)$$

where A is the adjacency matrix of the network G, whose elements take values between 0 and 1 corresponding to the weight of the correlation between nodes, x represents an eigenvector of A and λ is an eigenvalue associated to that eigenvector.

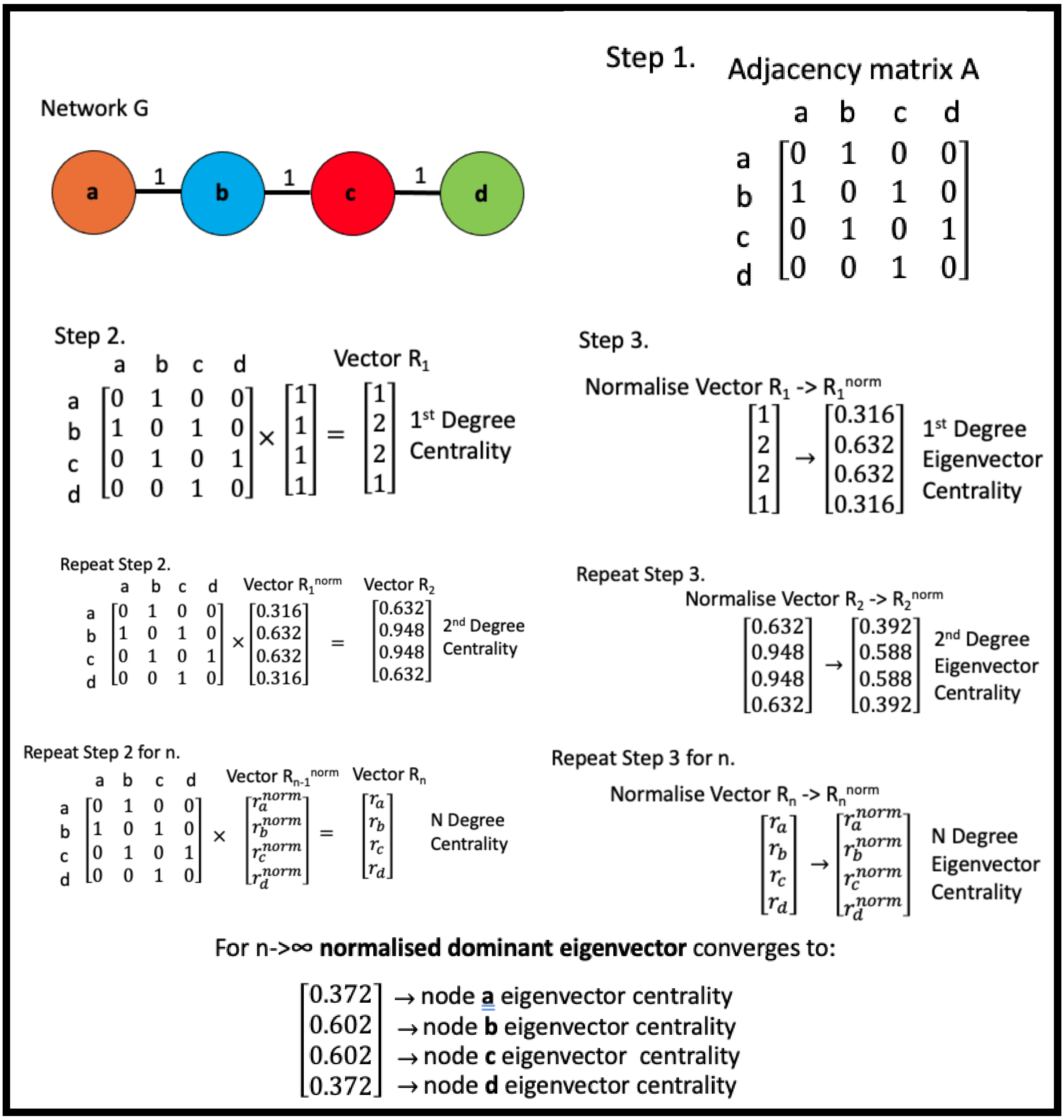


Figure 4. Eigenvector centrality of simple 4 node network where connections between nodes is either 1 for a connection of 0 for no connection

For a simple network (Fig. 4), step 1 gives its corresponding adjacency matrix A. It represents the graph of the network. Now, assume that there is a stochastic dynamical process on that graph. For example, a random walk process, where an initial random vector R would have elements that are related to the probability of finding a random walker on each one of the nodes in the graph. R_1 is the degree centrality of each node in the network found by multiplying A by a vector of 1's (step 2). Normalising R_1 gives the eigenvector centrality based on the value of the immediate neighbours (step 3). The vectors are normalised such that the square root of the sum of the power of 2 of all the elements inside the vector equal to 1. This is to say that after multiplying the vector R with the matrix A, the normalised vector R_1^{norm} has length 1.

This process can be repeated up to N times to include the values of the next N closest neighbours. After an infinite number of interactions Perron-Frobenius theorem[13], states that there is a unique and positive solution for R_n if λ is the largest eigenvalue associated with the eigenvector x of the adjacency matrix A, where R_n is equal to “dominant” eigenvector x of A. Basically, the power method is a computational efficient method to calculate the dominant eigenvector of A. The v^{th} component of R_n gives the relative centrality score of vertex v in the network. If you then normalise again R_n^{norm}(for $n \to \infty$), this has a physical meaning where the v^{th} component of R_n^{norm} represents the likelihood of finding the random walker in the node v of the network.

This takes the weight of the edges between nodes and calculates a normalised global centrality measure for each indicator in the graph. With a higher eigenvector score meaning a node is strongly connected to many other nodes in the network.

2.7. Community Detection.

The UN categorised the indicators within separate SDG’s based on social and economic links between them connecting them to the overall goal. I wanted to examine whether we could distinguish these links via a more statistical network analysis using a community detection algorithm. This is an unsupervised method of determining groups within a network based on the edges connecting nodes. Givan and Newman[14] said network nodes are tightly connected within communities and loosely connected between communities.

Yang[15] found that the multilevel (or Louvain[16]) algorithm produced the best results no matter the network size. We apply the Leiden[16] (an improved Louvain method) algorithm (through CDLIB[17]), to the synergy correlation network to evaluate its structure (Fig. 5).

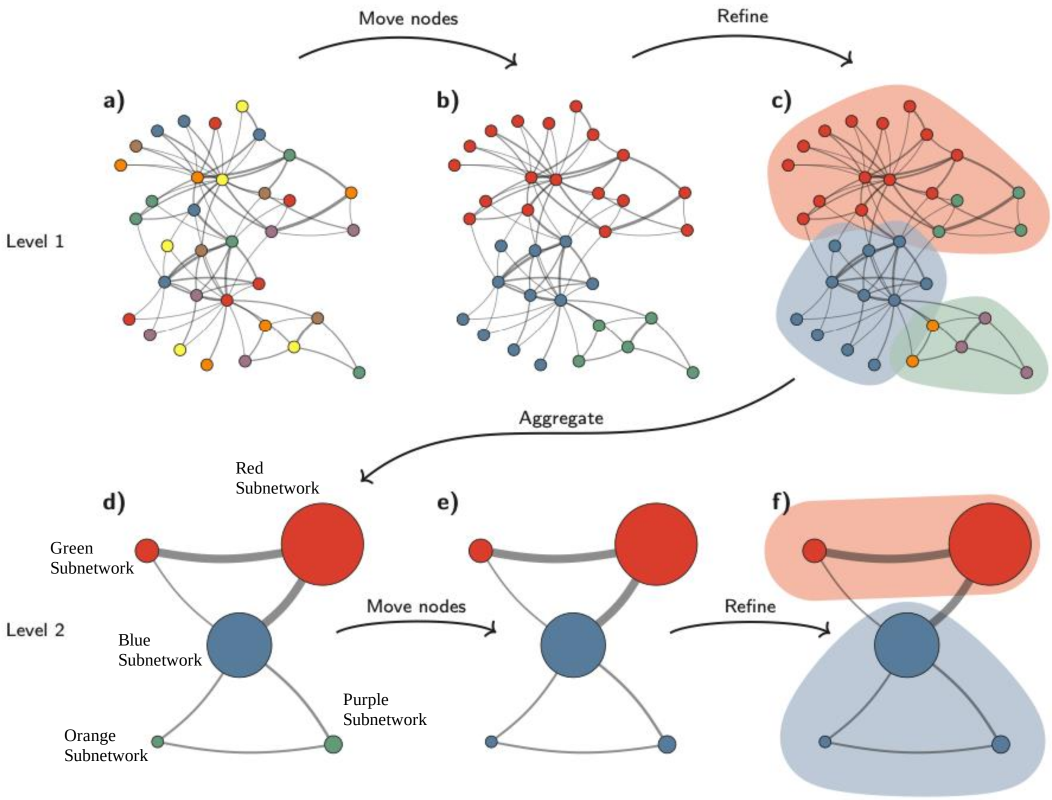


Figure 5. Leiden algorithm. All nodes separate (a). Local moving of nodes from one community to another to find a partition (b), Partition refined allowing for subcommunities to be found (c). Aggregate network created on new partition between subcommunities (d). Steps repeated on each newly aggregated network (e)-(f) until no move increases modularity. (**Graphic source:** Leiden[16])

Nodes are merged with a community only when there is an increase in the quality function; the method used is modularity[18]. This finds the difference between the actual number of edges within a community and the expected number if the edges were placed at random in the network. Modularity[19] Q is given by:

$$Q = \frac{1}{2m}\sum_{c}\left(m_c - \gamma\frac{K_c^2}{4m}\right) \quad (2)$$

Where m_c is the number of edges within community c. The expected number of edges is $\frac{K_c^2}{4m}$ where K_c is the sum of the degree of all nodes within community c and m is the total number of edges in the whole network. $\gamma > 0$ is the resolution parameter for which a larger value leads to more communities and a lower value to less (for the Leiden algorithm it is set to 1). When you have two communities (or singular nodes) if combining them would lead to an increase in the modularity then they are joined to form a single new community.

As the algorithms only group the nodes it does not provide a classification. Classification of the communities is done by determining the SDG that has the highest number of indicators in the group. Nodes that are classified as equally important are coloured with one node colour and the outline of the node is coloured as the other.

2.8. Delayed correlation.

To be able to action improvements after evaluating a node's influence it is useful to know the direction of impact an indicator has on its connections, or the direction of causality[20]. This allows us to determine the indicators that need to be improved to have a real difference on those directly linked. A method for estimating a weak notion of causality is to measure delayed correlation[21] between two highly correlated indicators which we apply to only positively correlated pairs (synergy) to determine what positive directional connections exist. For indicators A and B, we calculate the correlation between A and B (shifted by 3 years into the future) then the correlation between B and A (shifted by 3 years into the future). With the larger of the two correlations showing the indicator that has the greater effect on the future of the other, determining the direction of causality.

The idea behind this is that the causality measures how the future of one variable B is affected by the past and present of another variable A. This information can be used to predict the future of B. If the correlation for time shifted time-series is positive then the past of a variable can be used to predict the ranked relationship of the future of another variable. Moreover, as the delay is changed from a zero to a positive value, if the delayed correlation value is increased it means it will be easier to create linear models capable of predicting the future of B in terms of the present of A. This is by no means a rigorous measure of causality, we consider it to be a weak notion of causality allowing for predictability of similar monotonically ranked relationships of the data. However, it is an easy measure to make without needing any more elaborate computational effort.

3. Results.

3.1. UN Networks.

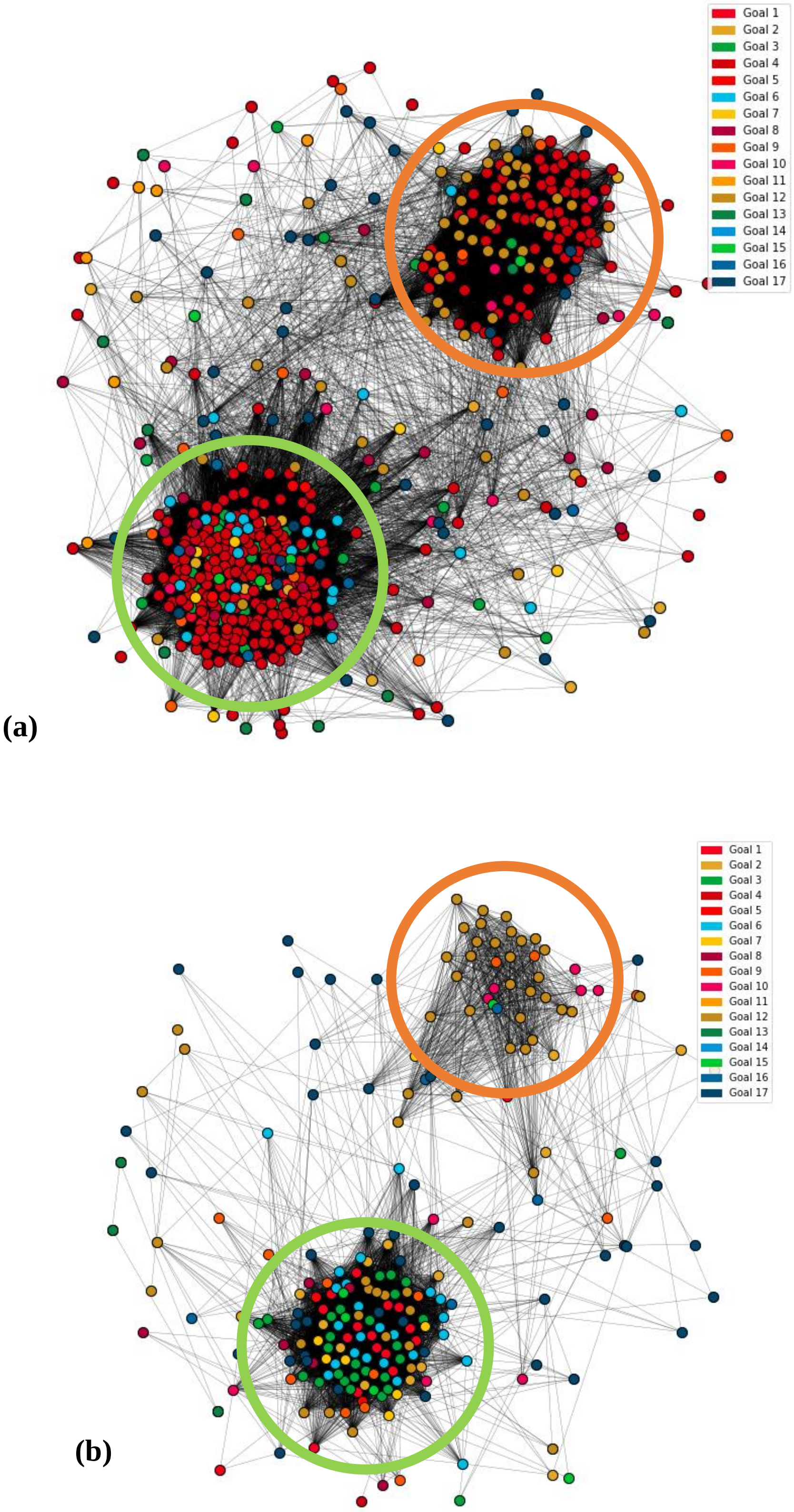


Figure 6. UN classification of all 17 SDGs Synergy Network of Brazil (a) and India (b) where nodes representing indicator time series are coloured by the UN SDGs classification.

When evaluating the network of synergies, between Brazil (Fig. 6a) and India (Fig. 6b) it is difficult to directly compare the two due to the limited data available. For Brazil there is enough data for 655 disaggregated indicators whilst for India there is less than half that at only 291 disaggregated indicators. Despite this it can still be useful to identify if any general trends appear. Both networks form two distinct groups (seen inside the circles) with one tightly grouped together (group 1 - green) and a less densely packed group (group 2 - orange), with the remaining nodes scattered about with connections to both groups. For Brazil both groups are dominated by SDG 4 (Quality Education), with group 1 containing nodes from SDG 3 and 6 while group 2 has nodes from SDG 12 and 17. For India group 1 is dominated by SDG 3 (Good Health and Well-Being) with nodes from SDG 1, 6 and 12. While group 2 mostly contains nodes from SDG 12 with some nodes from SDG 9 and 10.

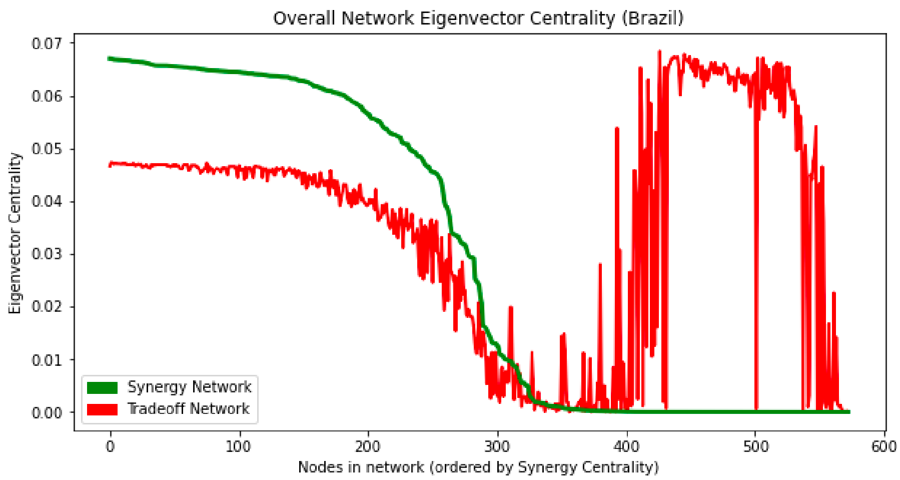
3.3. Eigenvector Centrality.

When examining the node eigenvector centrality, we first look at the individual node values. For both Brazil (Fig. 7a) and India (Fig. 7b) there is a clear trend that the nodes with the largest Eigenvector Centrality in the Synergy (ECS) network also have large Eigenvector Centrality in the Trade-off (ECT) network. This shows that nodes that are highly linked with synergies must also be highly linked in terms of their trade-offs. They contribute both positively and negatively for the development of the other goals. As the ECS value decreases so does the ECT value so as the nodes become less important in the synergy network their importance often also decreases in the trade-off network.

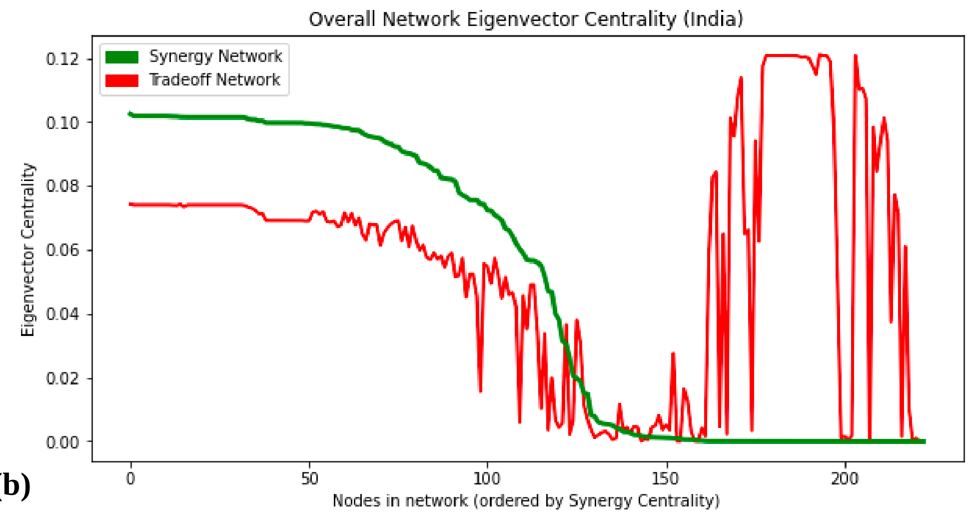
There is an interesting trend when the ECS value is extremely low, showing for both networks that there are many nodes that have an extremely high ECT value. For these nodes, it means that they are almost exclusively linked via trade-offs. Assuming that the general trend of the country is towards improvement, these indicators do not follow the country's trend and are most likely indicators that are getting worse at a significant rate.

For Brazil the largest trade-off eigenvector centrality node is indicator 13.2.2 (Total greenhouse gas emissions) from SDG 13 (Climate Action). Climate change is an increasingly important issue and looking more closely at the data it is seen that greenhouse gas emissions are getting worse, causing this indicator to be a trade-off with most of the other indicators available for Brazil[22].

For India there is a similar trend, and the highest trade-off node is indicator 9.4.1 (Total CO_2 emissions from fuel combustion) from SDG 9 (Industry, Innovation, and Infrastructure). Examining the data further this is also getting worse at an increased rate[23].



(a)



(b)

Figure 7. Eigenvector Centrality ordered by synergy for Brazil (a) and India (b)

Looking at the top eigenvector synergy network nodes, the number 1 node for Brazil (Table 1a) is number of full-time researchers (SDG 9), with number 2 related to Number of undernourished people (SDG 2) and the remaining 3-5 related to Education completion rates (SDG 4). However, there is not a significant difference between these ECS values and so it can be concluded that they are of similar importance to the overall synergy network.

For India (Table 1b) the important nodes come from different goals with the top node relating to installed renewable electricity-generating capacity (SDG 7/12) while the other top 4 are about sanitation services and open defecation (SDG 6). The rank 1 node has a slightly higher ECS value although it is very similar to the other top 5, so they all also have an equal impact on the network.

Rank	SDG Indicator	Description of Indicator	Eigenvector Centrality Synergy Value
1	9.5.2	Number of full-time researchers	0.0670
2	2.1.1	Number of undernourished people	0.0669
3	4.1.2	Lower Secondary Education completion rate (**for Urban, Bothsex and Wealth Quintile Q2**)	0.0668
4	4.1.2	Upper Secondary Education Completion Rate (**for Urban, Bothsex and Wealth Quintile Q3**)	0.0668
5	4.1.2	Lower Secondary Education completion rate (**for All Areas, Males and Wealth Quintile Q1**)	0.0668

Table 1a. Top 5 ranked Synergy eigenvector centrality nodes (Brazil)

Rank	SDG Indicator	Description of Indicator	Eigenvector Centrality Synergy Value
1	7.b.1/12.a.1	Installed renewable electricity-generating capacity	0.1026
2	6.2.1	Proportion of population using safely managed sanitation services (**for all areas**)	0.1019
3	6.2.1	Proportion of population practising open defecation (**for rural only**)	0.1019
4	6.2.1	Proportion of population practicing open defecation (**for urban only**)	0.1019
5	6.2.1	Proportion of population using safely managed sanitation services (**for rural only**)	0.1019

Table 1b. Top 5 ranked Synergy eigenvector centrality nodes (India)

Visualising the most important nodes in relation to the overall network can help to show why they are so key. For Brazil node 1 is enlarged and highlighted (Fig. 8a) so its position in the network is clear. As seen the node is part of the densely populated group and all its direct connections are coloured by their respective SDG. The node is almost exclusively linked to nodes within its group with it connected to the vast majority in the group.

For India (Fig. 8b) the same pattern is seen with the key node being part of the larger and denser group while also being directly connected to all the other nodes in the group.

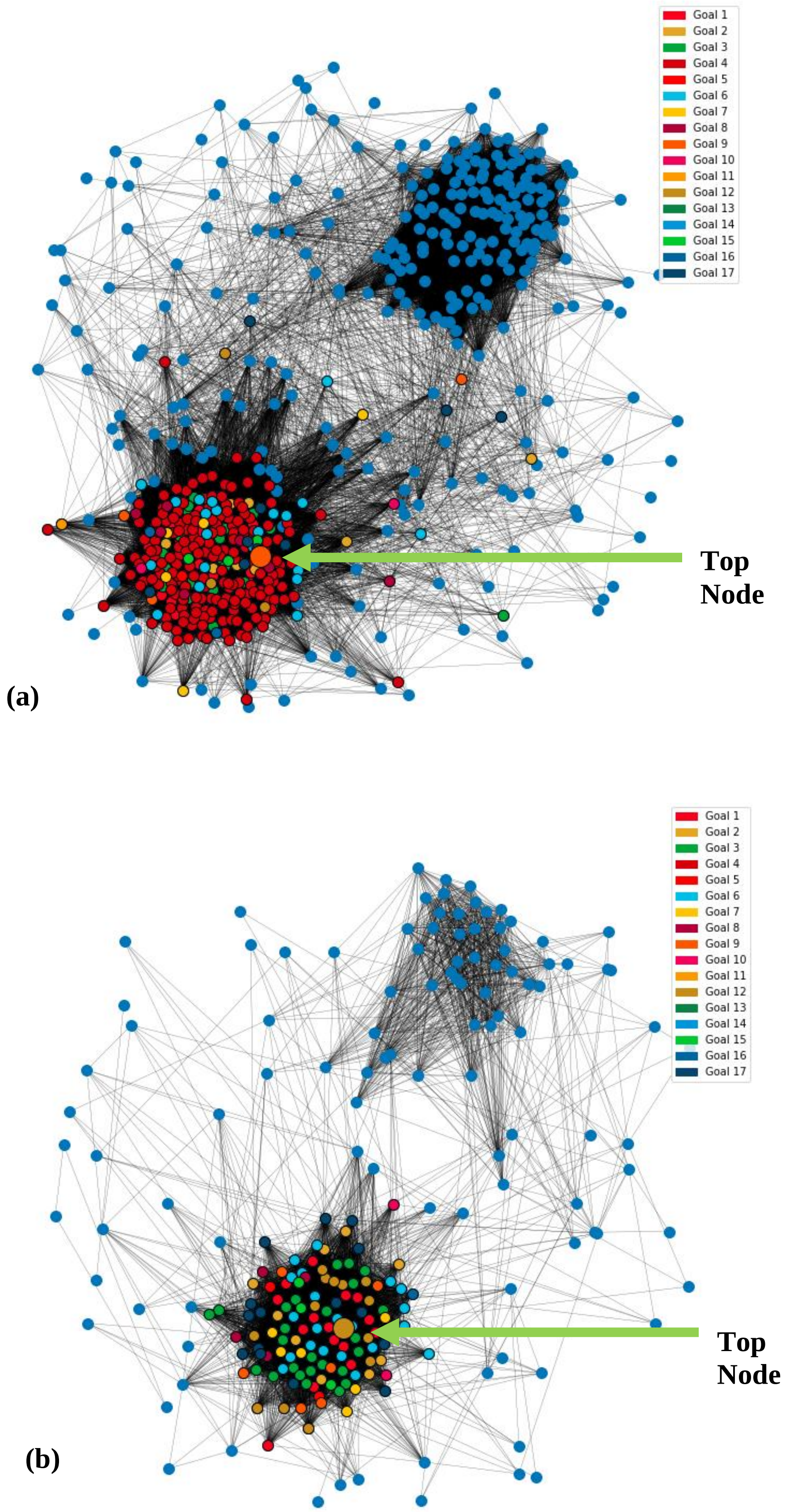


Figure 8. Top Synergy Eigenvector Node, for (a) Brazil (full-time researchers) and (b) India (installed renewable electricity-generating capacity), where nodes coloured are all direct connections to the top node.

3.3. Community detection.

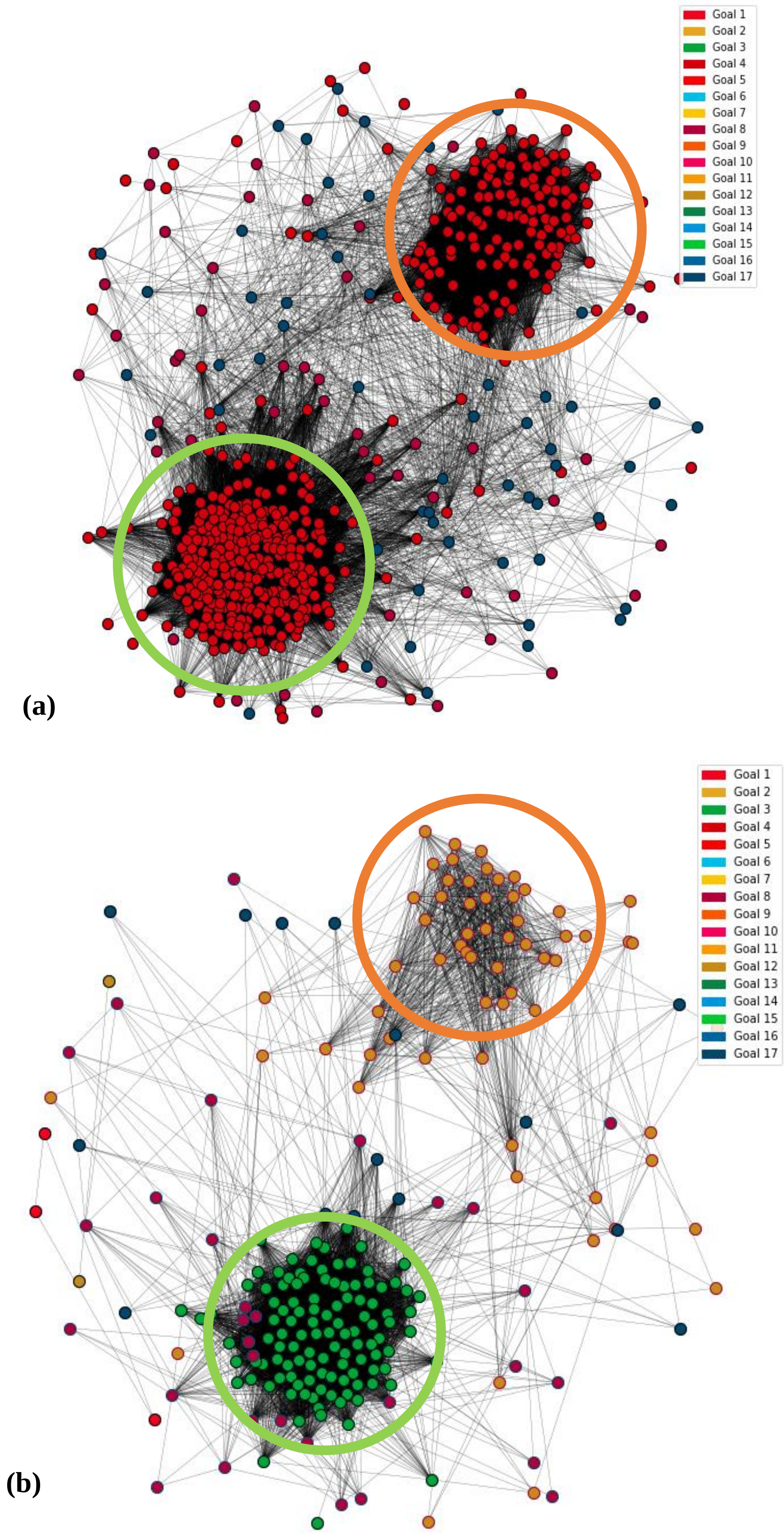


Figure 9. Leiden Community Detection on Synergy Network for Brazil (a) and India (b), where nodes are coloured based on the SDG with the largest number of nodes in that community.

For the Leiden community detection for Brazil (Fig. 9a) the network is plotted using the same spring layout. This time both group 1 and 2 although grouped separately, are classified as SDG 4. This leaves the remaining connector nodes between the 2 groups which were classified as being part of SDG 4, 8 or 17.

Applying Leiden community detection to India's network (Fig. 9b), group 1 is classified as SDG 3, with a few nodes classified as SDG 8/17. Group 2 is classified as SDG 8/12, leaving the remaining connector nodes classified as SDG 1, 8, 12 and 17.

3.4. Delayed Correlation.

For Brazil's complete network, the rank 1 node (Table 1a) indicator 9.5.2 (full-time researchers) feels a directional effect from indicator 4.b.1 (total official flows (funds/grants) for scholarships) (Table 2a). This suggest that increasing the total official flows has an influence on how many full-time researchers there will be. The overall direction is shown by the -0.57 with the minus sign meaning that 4.b.1 has causal effect on 9.5.2.

Scholarships are typically granted to those with high academic ability but who do not have the funds to pay. As the amount of money available for scholarships increases[24] the number of intellectual people going to university will increase. Many of these will likely continue to do further research leading to an increase in the numbers of full-time researchers in the country[25].

Eigenvector Central node (A)	Connected node (B)	Unshifted Correlation	Correlation A to B (shifted)	Correlation B to A (shifted)	Difference in Correlation (Cor A – Cor B)
Full Time Researchers	Total official flows for scholarships	0.87	0.37	0.94	-0.57
Number of undernourished people	Children moderately or severely overweight	0.85	0.16	0.96	-0.80

Table 2a. Delayed correlation results for top 2 indicators and a connected node (Brazil), a positive value in the final column suggests A weakly causes an effect on the future of B, while a negative value indicates B weakly causes an effect on the future of A

Examining the causal relationship for India rank 1 node (Table 1b) the correlation is large between the top indicator 7.b.1/12.a.1 (installed renewable electricity-generating capacity) and the future of indicator 9.a.1 (total official flows for infrastructure e.g. grants/funds). At a difference of 0.31 it shows that renewable capacity likely has a directional causal effect on infrastructure investment.

As renewable electricity-generating capacity that is installed increases[26] there will be greater demand for improved infrastructure that is cost-effective and sustainable[27] leading to more grants for this development. Although there is some directional relationship the other way, companies and governments are less likely to offer funds for improvement until after there has been the increase in capacity.

Eigenvector Central node (A)	Connected node (B)	Unshifted Correlation	Correlation A to B (shifted)	Correlation B to A (shifted)	Difference in Correlation (Cor A – Cor B)
Installed Renewable electricity-generating capacity	Total official flows for infrastructure	0.80	0.96	0.65	0.31
Population using safely managed sanitation services	Lakes and rivers permanent water area	0.81	0.61	0.90	-0.29

Table 2b. Delayed correlation results for top 2 indicators and a connected node (India), a positive value in the final column suggests A weakly causes an effect on the future of B, while a negative value indicates B weakly causes an effect on the future of A

An overview of the direction of causal relationships between the top 5 eigenvector centrality nodes and all their linked nodes that are a synergy (Table 3) shows that for both Brazil and India there are far more cases of other nodes that impact on the top 5 central nodes. For indicator correlations in Brazil, on average the top nodes impacted the future values of just 31.6% (28.9% for India) of their linked indicators while the top nodes were impacted by 68.3% (71.1% for India) of their links. This is a significant difference of over double suggesting that these key eigenvector nodes, if improved, would influence a target selected number of other indicators (85.8) positively. On the other hand, the top 5 nodes would be impacted by a broad range of other nodes. The top eigenvector centrality nodes therefore behave as a informational sink, whose future state can be better predicted by looking at the present of other nodes impacting on it. These sink central nodes can also potentially serve as useful in tracking the overall performance of a country, since they respond to what the other indicator nodes are performing. So, by following the changes in these small set of top central nodes would provide a good idea of how much a country is progressing towards the SDGs.

	Brazil	India
Impacts	85.8 (31.6%)	23.8 (28.9%)
Impacted by	185.6 (68.3%)	58.6 (71.1%)
Total Nodes	271.4	82.4

Table 3. Average number of nodes the top 5 eigenvector indicators likely impact or how many nodes they are impacted by

4. Discussion.

To fully understand the SDGs and their impacts on a country the full picture needs to be constructed. To understand how the network was linked this was examined using a network visualisation on the Spearman correlation values between indicator pairs.

Identifying key indicators allows for a deeper understanding of the network connectivity and eigenvector centrality gives a good measure of how connected a node is since it considers not only the number of connections a node has but also their strength. Across Brazil and India, the top nodes are not limited to a single SDG and the top 5 for the countries contain indicators from SDG 2, 4, 6, 7, 9 and 12. It is important to identify these individual key indicators as it is not possible to just analyse a single goal that can explain or improve the whole of the rest of the network.

Combining correlation with the Leiden community detection algorithm helped to determine if there was an underlying statistical significance for how indicators were grouped by the UN. Both networks formed two main communities with a larger denser one and a smaller less dense one. It was not clear why this was the case for both countries as each group did not contain the same SDGs. Perhaps due to how the Goals were constructed it led to particular pairs of SDGs or targets being highly correlated with each other forming the groups that we see. Due to some indicators being present in multiple goals these nodes would likely bridge two SDGs together as the indicators are likely to be highly correlated with indicators in both goals.

A simple weak causality method of delayed correlation provides a basic understanding of the directional influence that comes from the Spearman correlation between indicator pairs. From this analysis it can be determined that eigenvector centrality on an undirected network gives the key nodes that show the overall trend of a country. For Brazil this is through the number of full-time researchers (indicator 9.5.2) and monitoring this indicator would show the overall progress of Brazil is good if it improved. Similarly for India, tracking the installed renewable electricity-generating capacity (indicator 7.b.1/12.a.1) would give a good representation of the progress made towards a better and more sustainable future for all.

To improve understanding of the network structure, classifying the indicators in a single goal may allow for a clearer separation of the SDGs. Investigating the individual clusters more may allow us to understand why these groups are formed.

Understanding trade-offs is just as important as understanding synergies because it presents those nodes that are performing the worst in the network. Whilst improving them may not lead to the greatest positive improvement, allowing them to get worse would lead to significant negative implications for many nodes. Whilst it was not possible to evaluate trade-offs as closely here, they are still crucial in providing a full insight into the progress a country is making towards reaching the SDGs.

5. Acknowledgements.

The project has been partially financially supported by *A systems approach to sustainable sanitation challenges in urbanising China* (SASSI), funded by Natural Environmental Research Council (NERC).

I would like to thank my supervisor Dr Murilo S. Baptista, a reader at the University of Aberdeen for his support throughout the project. As well I would like to thank Ulysses Sengupta, a reader at the Manchester School of Architecture (MSA) for his help and assistance during the research.

References

1. Sustainable Development Goals - https://sdgs.un.org/goals
2. SDG 2030 Agenda - https://sdgs.un.org/2030agenda
3. Pradhan P., Costa L., Rybski D., Lucht W., Kropp JP. A systematic study of sustainable development goal (SDG) interactions. Earth's Future 5:1169–1179 (2017). https://doi.org/10.1002/2017EF000632
4. TOSSIB project website - https://susinfra.com/tossib
5. SASSI project website - https://susinfra.com/sassi
6. SDG data - https://unstats.un.org/sdgs/dataportal/database
7. Spearman's ranked correlation coefficient - http://webspace.ship.edu/pgmarr/geo441/tables/spearman%20ranked%20correlation%20table.pdf
8. Mester A., Pop A., Mursa B.E.M., Grebla H., Dios ̧an L., Chira C. Network Analysis Based on Important Node Selection and Community Detection. *Mathematics, 9*, 2294 (2019). https:// doi.org/10.3390/math9182294
9. Centrality comparisons - https://towardsdatascience.com/unveiling-important-nodes-in-a-network-4992a2ea1cca
10. Community detection algorithm comparisons - https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae
11. NetworkX documentation - https://networkx.org
12. Bihari A, & Pandia M. Eigenvector centrality and its application in research professionals' relationship network. 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE (2015). 10.1109/ABLAZE.2015.7154915.
13. Perron-Frobenius theorem - https://www.math.uchicago.edu/~may/VIGRE/VIGRE2007/REUPapers/FINALAPP/Khim.pdf
14. Girvan M., & Newman M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the*

United States of America, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

15. Yang, Z., Algesheimer, R. & Tessone, C. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Sci Rep* **6,** 30750 (2016). https://doi.org/10.1038/srep30750
16. Traag, V.A., Waltman, L. & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9,** 5233 (2019). https://doi.org/10.1038/s41598-019-41695-z
17. Rossetti, G., Milli, L & Cazabet, R. CDLIB: a python library to extract, compare and evaluate communities from complex networks. *Appl Netw Sci* **4**, 52 (2019). https://doi.org/10.1007/s41109-019-0165-9
18. Newman, M. E. J. Modularity and community structure in networks. PNAS, vol. 103, no. 23, 8577–8582 (2006). https://doi.org/10.1073/pnas.0601602103
19. Modularity method for Leiden algorithm used by CDLIB library - https://leidenalg.readthedocs.io/en/stable/reference.html#leidenalg.ModularityVertexPartition
20. Causality definition - https://www.newscientist.com/definition/causality/
21. Delayed correlation similar example – Chenhua, S. Analysis of detrended time-lagged cross-correlation between two nonstationary time series. Physics Letters A. Volume 379, Issue 7, 680-687, ISSN 0375-9601 (2015). https://doi.org/10.1016/j.physleta.2014.12.036.
22. Brazil CO_2 - https://ourworldindata.org/co2/country/brazil
23. India CO_2 - https://ourworldindata.org/co2/country/india
24. Brazil scholarships increasing - https://www.gov.br/en/government-of-brazil/latest-news/2022/the-brazilian-government-has-increased-by-5-the-number-of-scholarships-to-finance-masters-and-phd-students
25. Brazil number of full time researchers - https://www.ceicdata.com/en/brazil/enterprise-industry-no-of-employee-by-industry/no-of-employee-pt-scientific-research--development
26. India renewable energy capacity - https://www.investindia.gov.in/sector/renewable-energy
27. India Improving energy infrastructure - https://www.bloomberg.com/news/articles/2021-08-15/india-aims-to-end-energy-imports-boost-infrastructure-spending