UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

CEU Department of Network and Data Science | CENTRAL EUROPEAN UNIVERSITY

Exploring the relationship between Sustainable Development Goals and cancer incidence by means of complex networks and machine learning

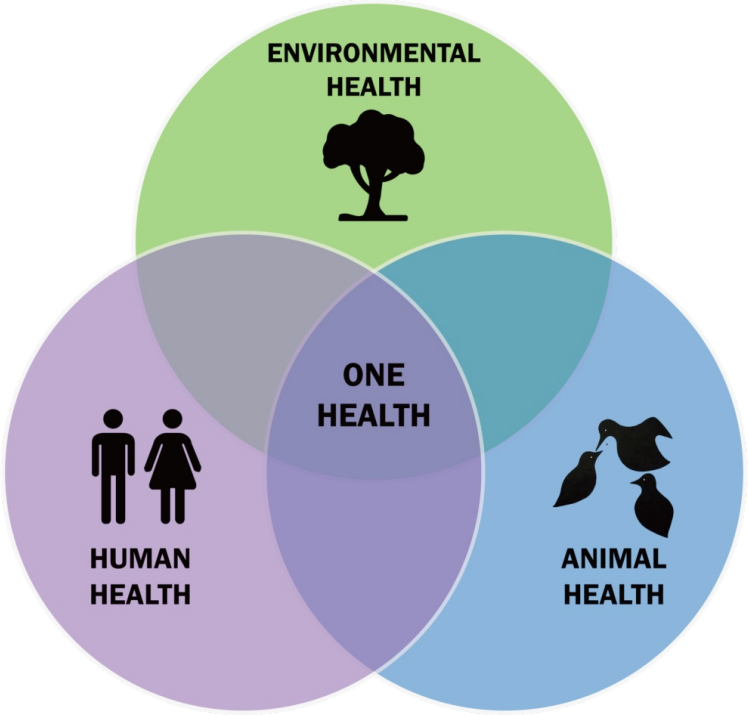
A. Lo Sasso, L. Bellantuono and E. Omodei

15 Sept. 2025

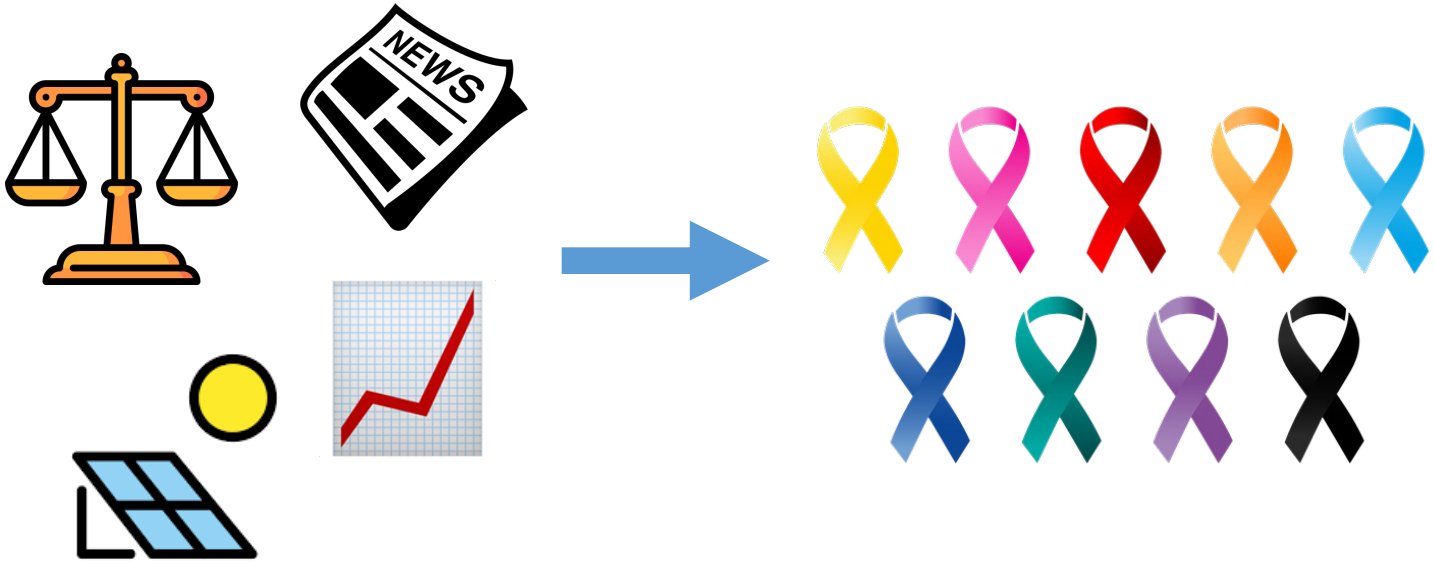
Complex Systems Society – Bari

COMPLEX SYSTEMS SOCIETY

In recent years, awareness of the impact of environmental, economic, and social factors, such as lifestyle and human-nature relationships, on citizens' well-being and the prevention of chronic diseases has grown.



Nevertheless, in the study of disease onset, potential socioeconomic causes remain not fully understood.

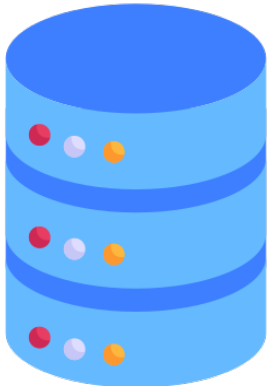


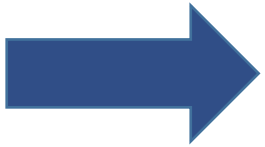
In recent years, studies correlating diseases with socioeconomic causes in United Nations Member States (UNMS) have sought to identify causes for specific pathologies [1, 2]. However, there is a lack of a comprehensive set of indicators that can collectively be associated with the onset of diseases such as cancer.

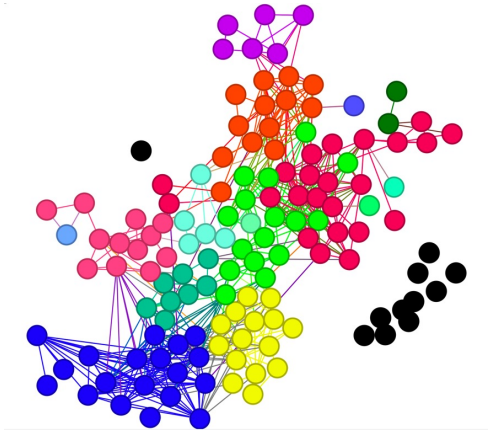
Among the indicators used to study the behavior of UNMS, the Sustainable Development Goals (SDGs) exhibit the greatest variability and abundance.

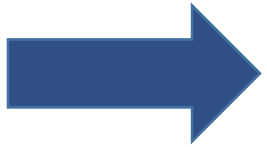
[1] Marais, B. J., et al. "Poverty and tuberculosis: is it truly a simple inverse linear correlation?." *European Respiratory Journal* 33.4 (2009): 943-944.
[2] Velimirovic, B., et al. "Socio-economic and environmental factors and human health example of cholera El Tor in Manila." *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Erste Abteilung Originale. Reihe B: Hygiene, Praventive Medizin* 160.1 (1975): 1-27.

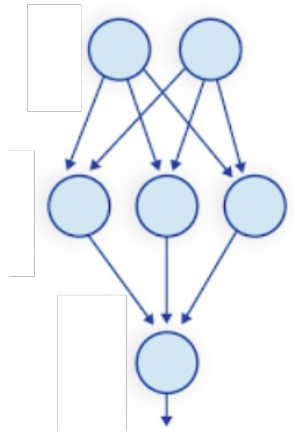
Outline









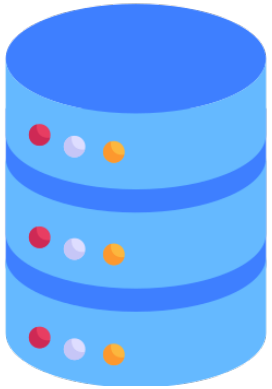


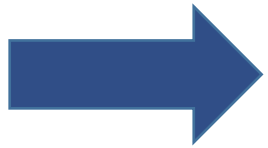
Data

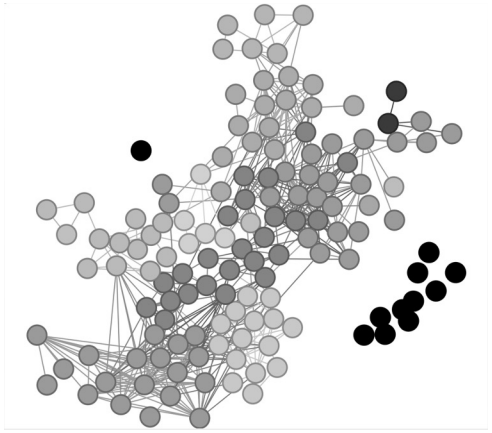
Network Filter

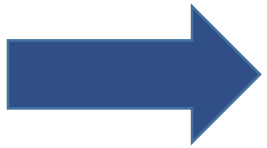
AI analysis

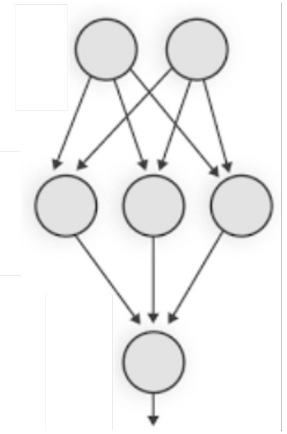
Outline











Data

Network Filter

AI analysis

Pre-processing – SDGs Data

The dataset available on the United Nations Sustainable Development Goals (SDG) Data Portal serves as a comprehensive global repository of statistical data related to the 17 SDGs established by the 2030 Agenda [3] .

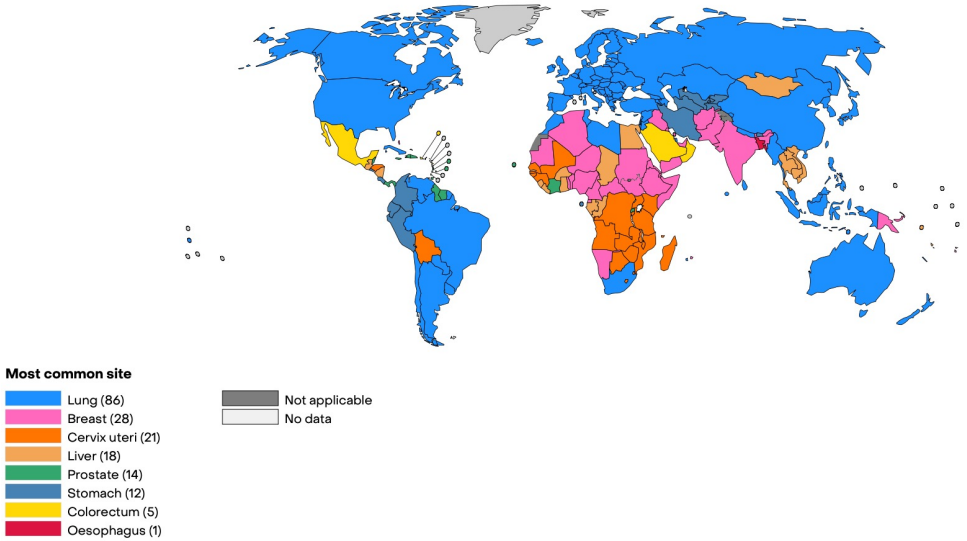
For analysis, we selected SDG indicators from 2022[4], ensuring higher data availability. The dataset contained 509 indicators, and we apply a selection criterion of at least 70% data completeness, to ensure reliability.

[3] United Nations Sustainable Development Goals (SDG) Data Portal : https://unstats.un.org/sdgs/dataportal/database
[4] UN DESA - Indicators of Sustainable Development: Guidelines and Methodologies (2007)

Pre-processing – WHO Data

The Cancer Today dataset is available on the Global Cancer Observatory (GCO), managed by the International Agency for Research on Cancer (IARC). It provides estimates of cancer incidence, mortality, and prevalence [5].

It tracks the incidence and mortality rates in UNMS, offering a standardized approach to analysing cancer trends across regions for research, prevention, and policy-making.

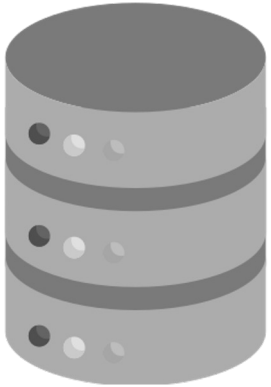


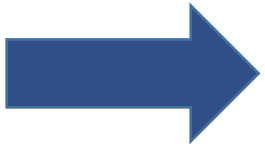
International Agency for Research on Cancer

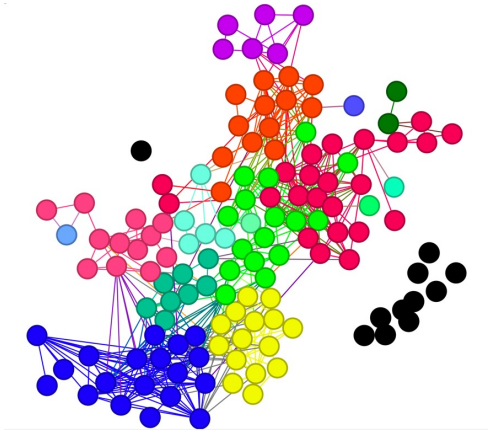
World Health Organization

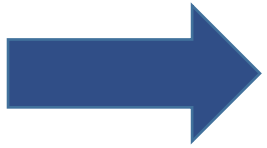
[5] Global Cancer Observatory: https://gco.iarc.fr/today/en/dataviz/maps-most-common-sites?mode=cancer&key=total&types=1&cancers=15

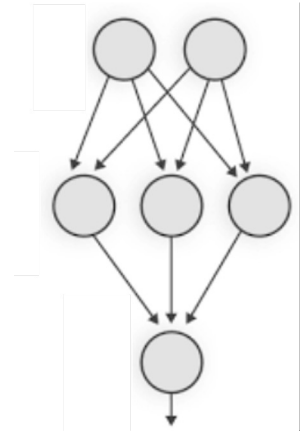
Outline











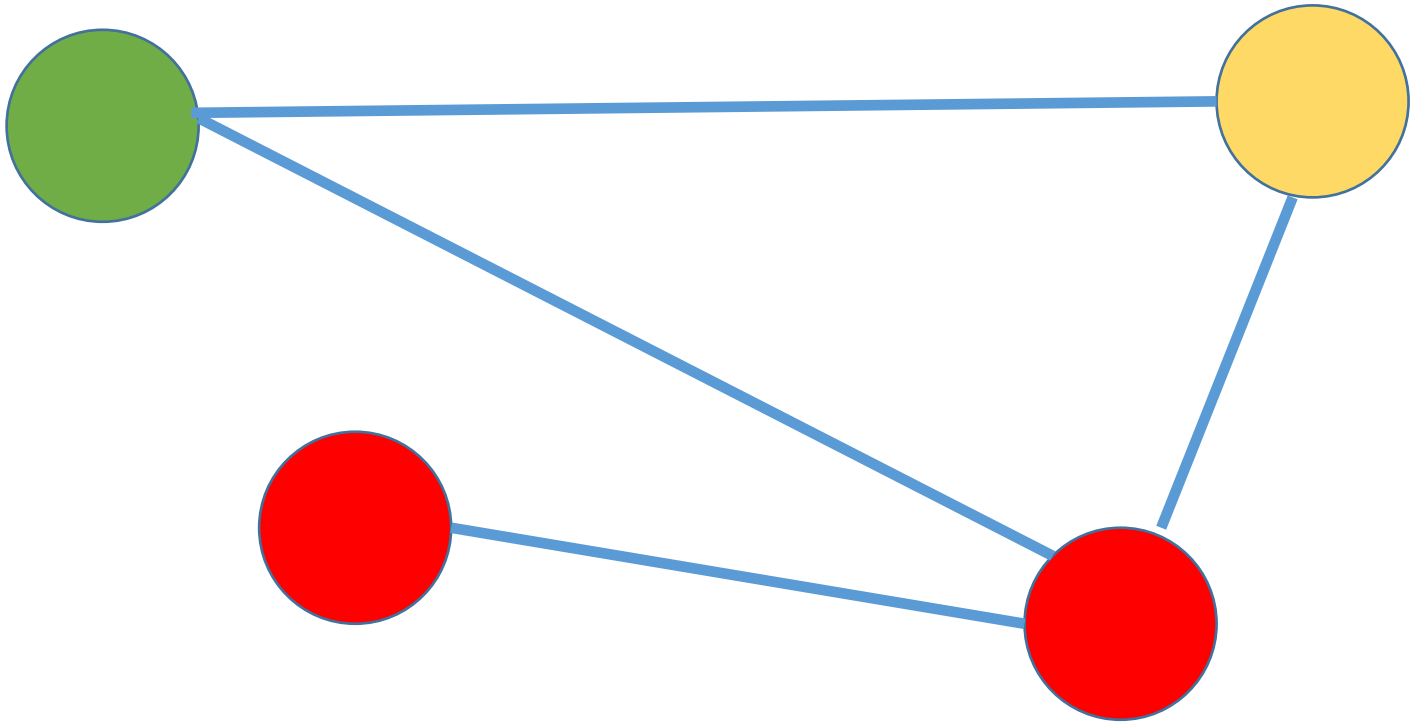
Data

Network Filter

AI analysis

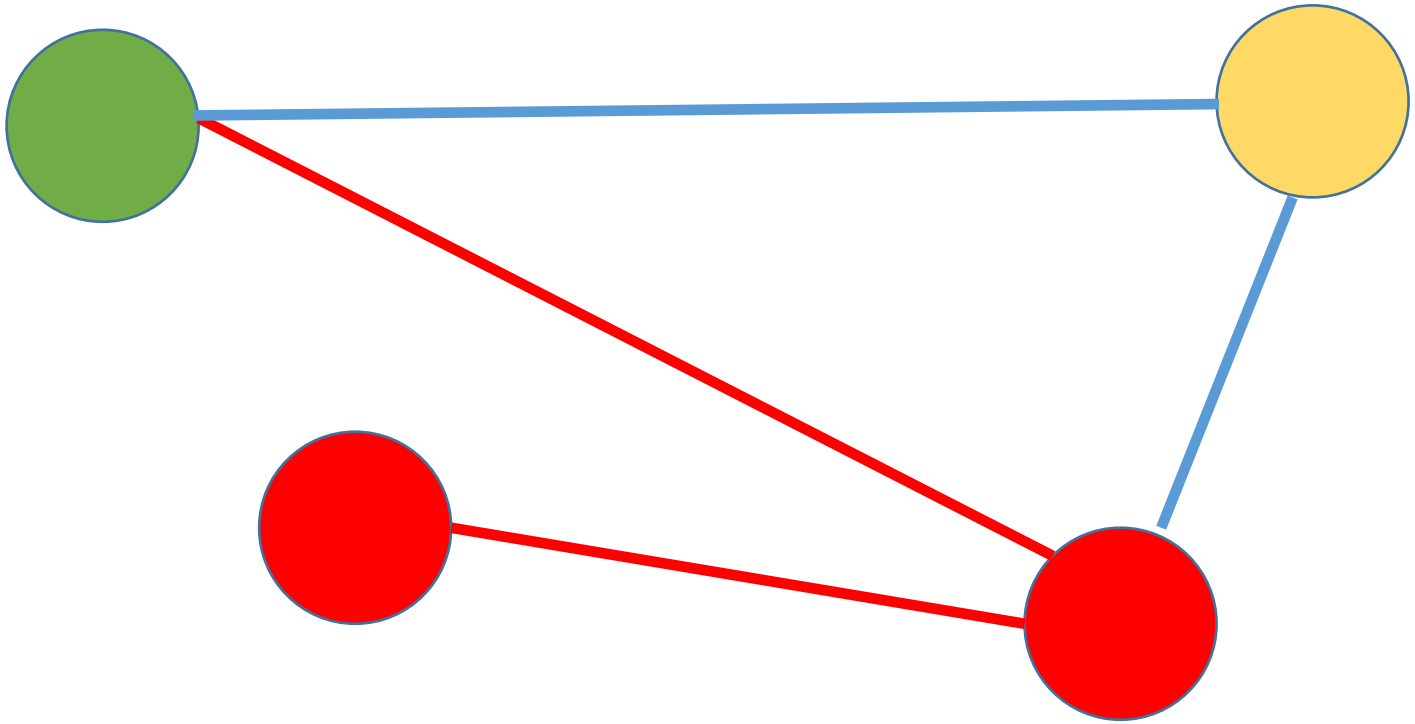
Network filter

We construct a weighted complex network, where nodes are SDG indicators, and links correspond to statistically significant correlations between node calculated by SDG data recorded across UNMS ($p < 10^{-6}$).



Network filter

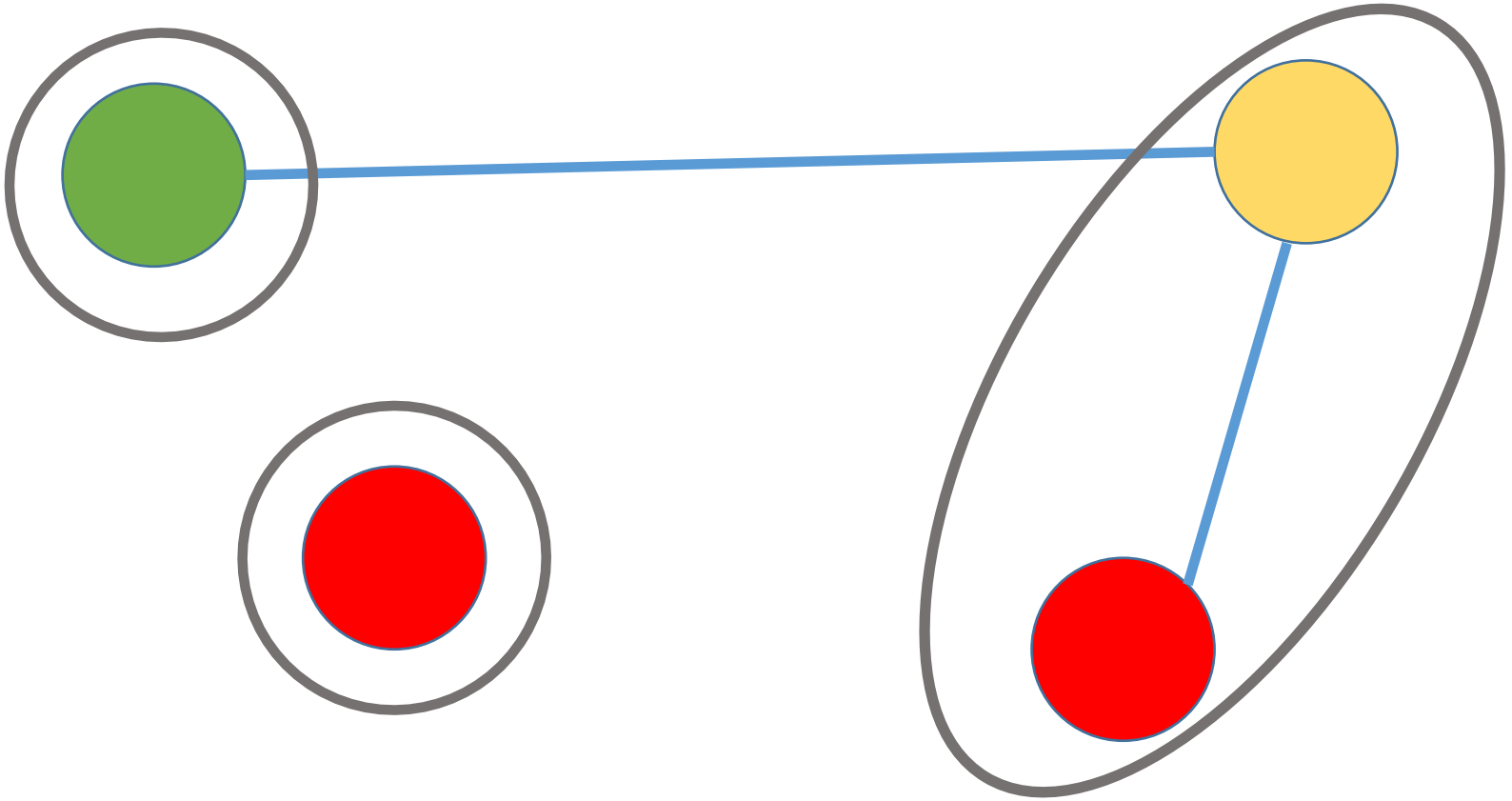
Applying the disparity filter [6], we reduce network complexity by preserving significant connections while eliminating statistically weak links.



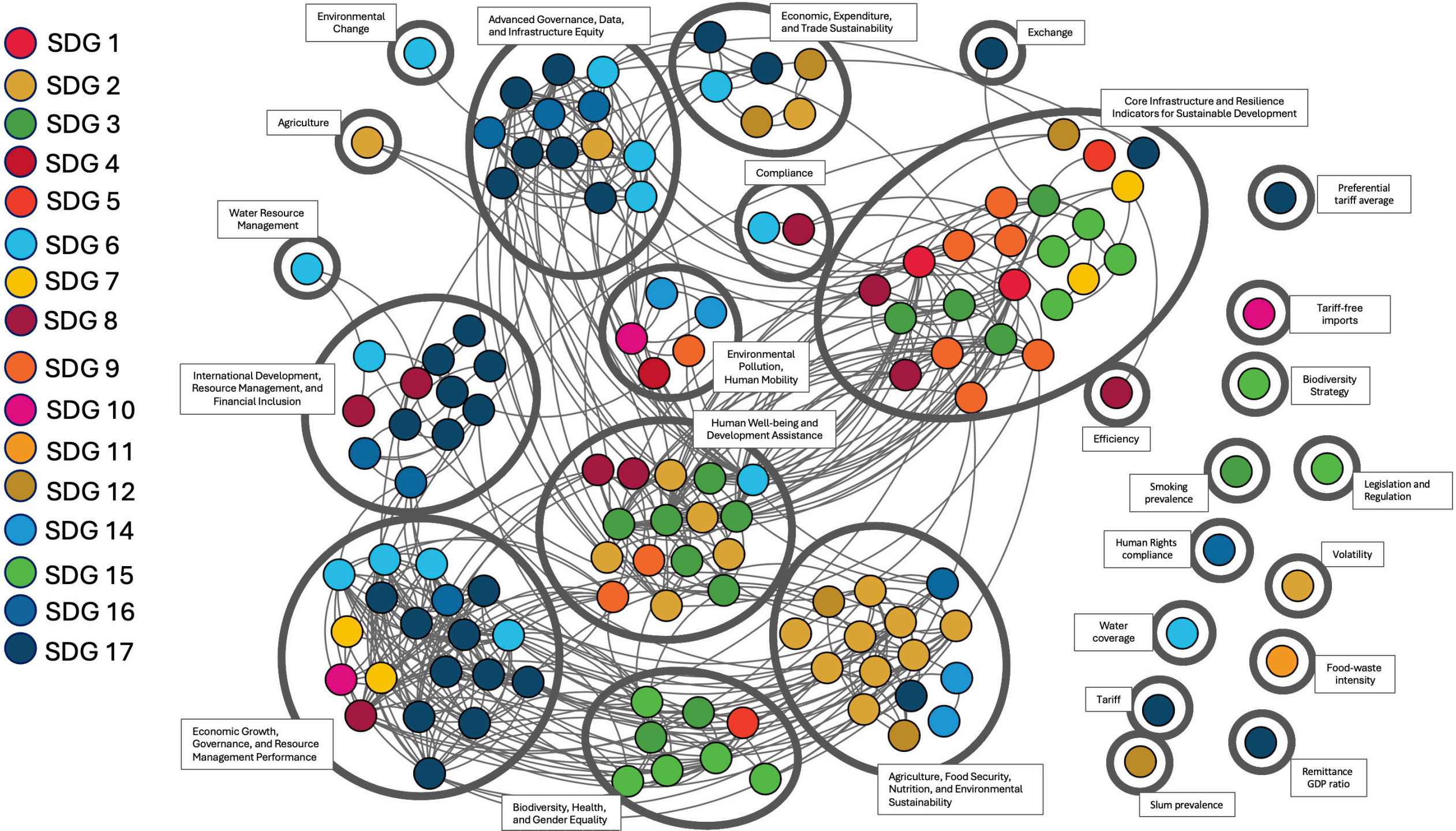
[6] Serrano, M. Ángeles, Marián Boguná, and Alessandro Vespignani. "Extracting the multiscale backbone of complex weighted networks." *Proceedings of the national academy of sciences* 106.16 (2009): 6483-6488.

Network filter

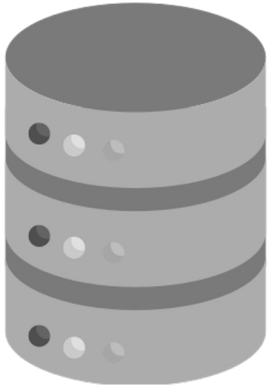
After performing community detection through modularity optimization [7], we select the most representative node in each community.

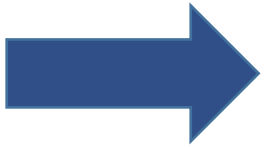


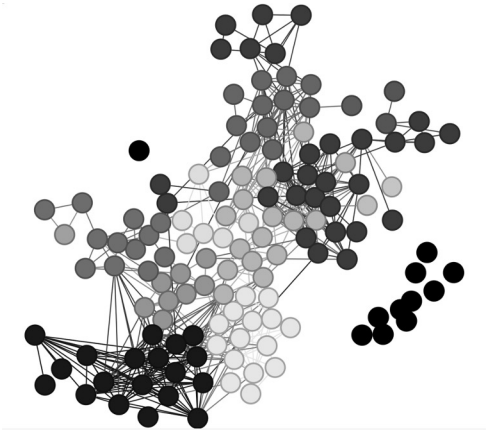
[7] Traag, Vincent A., Ludo Waltman, and Nees Jan Van Eck. "From Louvain to Leiden: guaranteeing well-connected communities." *Scientific reports* 9.1 (2019): 1-12.

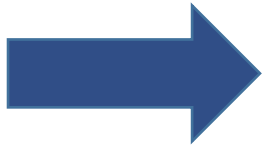

SDG 1
SDG 2
SDG 3
SDG 4
SDG 5
SDG 6
SDG 7
SDG 8
SDG 9
SDG 10
SDG 11
SDG 12
SDG 14
SDG 15
SDG 16
SDG 17
Environmental Change
Advanced Governance, Data, and Infrastructure Equity
Economic, Expenditure, and Trade Sustainability
Exchange
Agriculture
Core Infrastructure and Resilience Indicators for Sustainable Development
Compliance
Preferential tariff average
Water Resource Management
Tariff-free imports
Biodiversity Strategy
International Development, Resource Management, and Financial Inclusion
Environmental Pollution, Human Mobility
Efficiency
Human Well-being and Development Assistance
Smoking prevalence
Legislation and Regulation
Human Rights compliance
Volatility
Water coverage
Food-waste intensity
Tariff
Remittance GDP ratio
Economic Growth, Governance, and Resource Management Performance
Agriculture, Food Security, Nutrition, and Environmental Sustainability
Slum prevalence
Biodiversity, Health, and Gender Equality

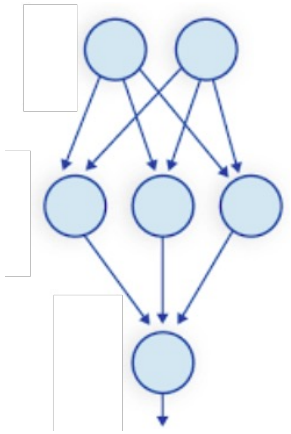
Outline









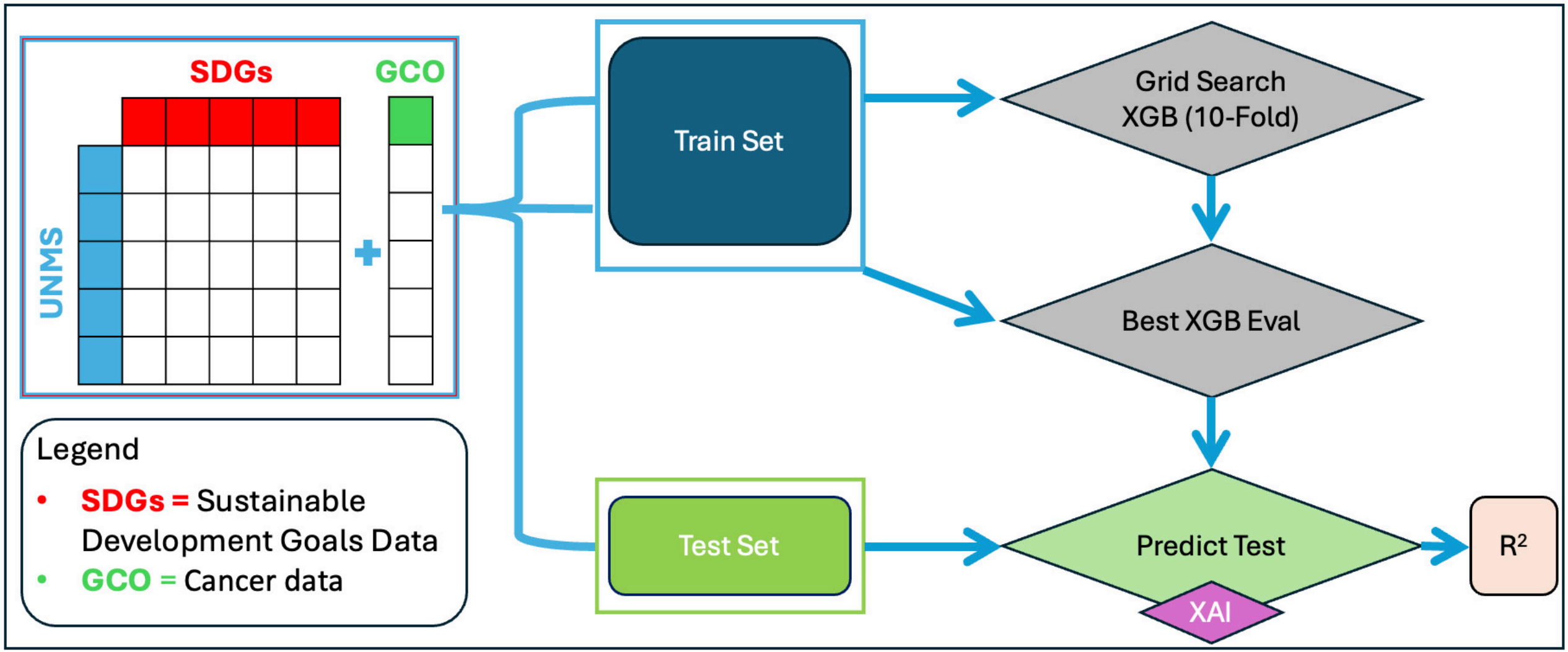


Data

Network Filter

AI analysis

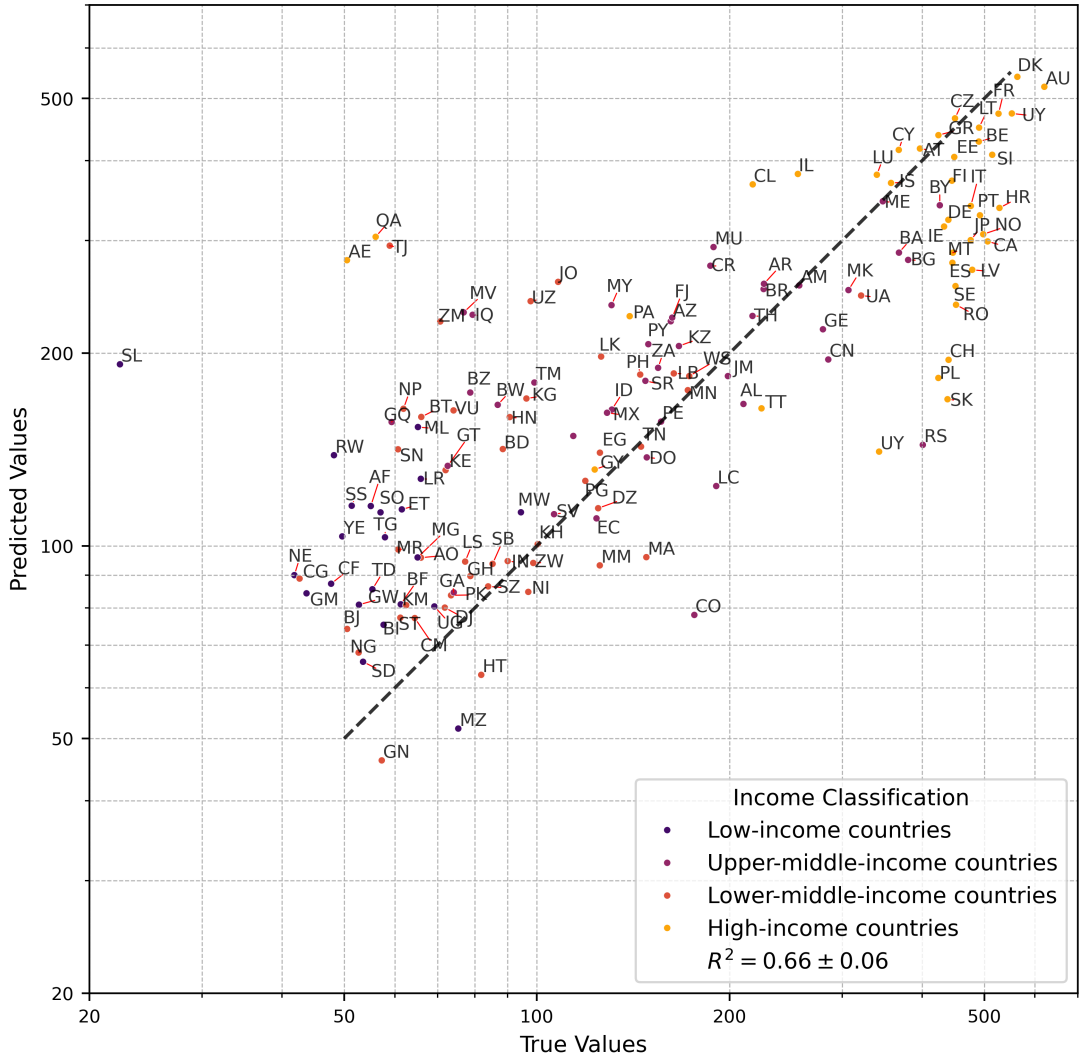
Machine Learning workflow



Comparing GCO data with ML predictions

Colouring the ML predictions of UNMS according to the income labels assigned to them by the World Bank, our findings indicate that countries with the highest predicted cancer incidence levels are also those classified as high-income level.

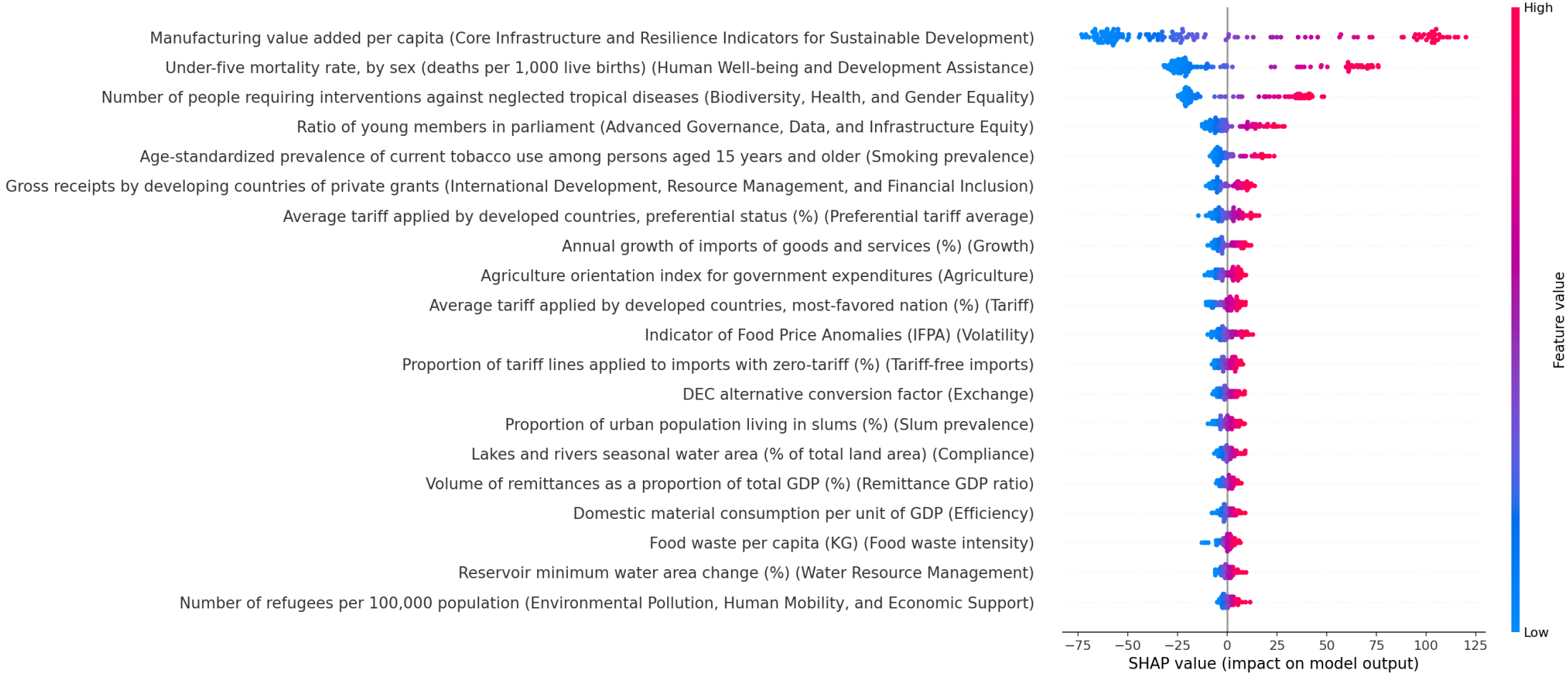
This is in line with previous studies [8] that correlate GDP and cancer incidence.



[8] Chahoud, J. et al. "Wealth, health expenditure, and cancer: A national perspective." *Journal of the National Comprehensive Cancer Network* 14.8 (2016): 972-978.

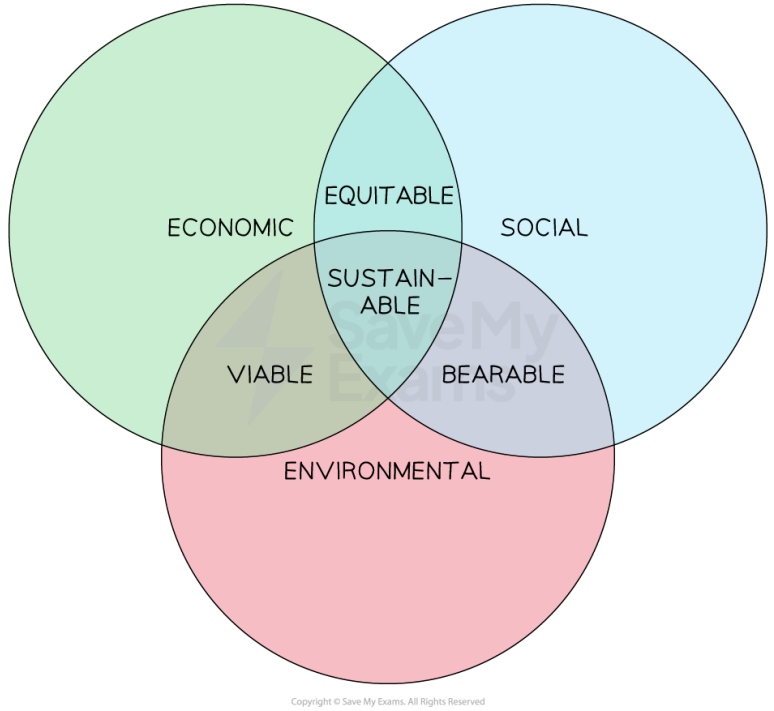
XAI analysis

The XAI analysis highlights the key role indicators for predicting cancer incidence do not include only SDG 3 (Good Health and well-being). We find also the socio-economic indicators referring to agricultural economy, and the management and cleanliness of water resources (linked to SDG 6) emerge as influential factors.



Conclusions

The XAI analysis highlights the significant role of socio-economic indicators, such as agricultural economy and water resource management in predicting cancer incidence.



Thank you for attention!

Andrea Lo Sasso
University of Bari
INFN Bari

Loredana Bellantuono
University of Bari
INFN Bari

Elisa Omodei
Central European University

Andrea Lo Sasso  andrea.losasso@uniba.it  @a_l_s