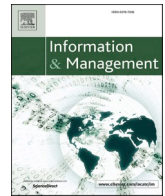


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Information & Management

journal homepage: www.elsevier.com/locate/im

Estimating networks of sustainable development goals

Luis Ospina-Forero^a, Gonzalo Castañeda^b, Omar A. Guerrero^{c,d,*}^a Alliance Manchester Business School, The University of Manchester, Manchester, United Kingdom^b Centro de Investigación y Docencia Económica (CIDE), Mexico City, Mexico^c Department of Economics, University College London, London, United Kingdom^d The Alan Turing Institute, London, United Kingdom

ARTICLE INFO

Keywords:

Sustainable development
Complexity
Estimation
SDGs
Networks
Indicators

ABSTRACT

An increasing number of researchers and practitioners advocate for a systemic understanding of the Sustainable Development Goals (SDGs) through interdependency networks. Ironically, the burgeoning network-estimation literature seems neglected by this community. We provide an introduction to the most suitable estimation methods for SDG networks. Building a dataset with 87 development indicators in four countries over 20 years, we perform a comparative study of these methods. We find important differences in the estimated network structures as well as in synergies and trade-offs between SDGs. Finally, we provide some guidelines on the potentials and limitations of estimating SDG networks for policy advice.

1. Introduction

In recent years, the concepts of inclusion and sustainability have become central to socioeconomic development. Perhaps, the best example is the Sustainable Development Goals (SDGs), the leading international agenda for national and regional development strategies. As the 2030 Agenda has progressed, it has been recognized that, to truly achieve sustainable development, it is necessary to understand how its multiple dimensions interact with each other [1]. In other words, a multidimensional view of development requires well-defined procedures to quantify and operationalize networks of interdependencies between different goals (or their indicators). Following this idea, several studies have attempted to measure such networks through different methods, for example, subjective criteria from expert advice, text mining applied to official documents and names of development indicators, and proximity measures between indicators that are relevant for a given country.

Overall, these efforts are commendable as they shift the discussion on socioeconomic development from topical silos to a systemic view [2]. However, scholars from development studies have not provided a –much needed– thorough reflection and analysis on the virtues and shortcomings of the different methods available for the empirical estimation of SDG networks. Furthermore, the number of alternatives considered in these studies has been rather small. This stands in stark contrast with the

growing literature on network estimation and causal inference models produced by data and network scientists.¹

This paper analyses different network-estimation methods to identify the most suitable procedures to build networks of development indicators. For this reason, we examine their underlying assumptions, strengths, and limitations. Broadly speaking, we classify the available methods into five families: correlation thresholding, Granger causality, chordal information filtering, statistical structure learning, and physics-inspired methods. Interestingly, only a handful of methods can be employed with currently available empirical development-indicator data. Thus, we perform a comparative study using the methods that show the potential to work with this kind of observational data.

Before introducing the reader to network-estimation methods and developing the empirical analysis, we provide a critical account of the current attempts to estimate SDG networks, as well as a brief but necessary discussion on the differences between association, structural dependence, and causation. The paper is structured in the following way: Section 2 reviews the methods that have been used to estimate SDG networks and discusses their contributions and drawbacks. Section 3 provides a general overview of five methodological families that are frequently employed by network and data scientists. Section 4 explains, in more detail, these four methods selected from these families that are best suited to construct SDG networks. Section 5 presents a comparative study applying these four selected methods on development-indicator

* Corresponding author at: The Alan Turing Institute, United Kingdom.

E-mail address: oguerrero@turing.ac.uk (O.A. Guerrero).

¹ The literature on Quantitative Sustainability Assessments has already made a call for the need to embrace causal methods in SDG [73].

<https://doi.org/10.1016/j.im.2020.103342>

Received 1 March 2019; Received in revised form 26 June 2020; Accepted 27 June 2020

Available online 8 July 2020

0378-7206/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

data. Finally, the paper closes with the conclusions in Section 6, where we synthesize empirical results, discuss some of the limitations of causal inference on SDG networks, and propose some guidelines.

2. On networks of sustainable development goals

As discussions on socioeconomic development have shifted toward a multidimensional view [3], international organizations and development analysts have concluded that a network framework is indispensable. At the same time, it has become evident that complexity arises when trying to understand how these multiple dimensions interact with each other. For example, reductions in infant malnutrition may cause children to perform better at the Programme for International Student Assessment (PISA) tests. Therefore, a policy originally aimed at improving the dimension of public health may also have implications in the dimension of education. In this example, causality may need to be considered both ways, as improvements in education may also create consciousness about the risks of unhealthy diets. Either way, it is clear that development does not take place within isolated topics, but it is rather a coevolutionary process across different dimensions.

Owing to the complexity arising from interdependencies between development indicators, it has been argued that policy interventions can only be properly designed when the synergies and trade-offs between different goals have been identified. This is especially important when social, economic, and sustainability indicators are jointly considered in a government's development strategy. To frame development goals in a setting of complex networks, it is important to take into account the following issues: (1) the way of interpreting an interdependency network (does a link represent causation, structural-dependence or correlation?), and (2) the assumptions underlying the estimation methods (which methods are more suitable given the nature of the available data?). In this section, we discuss issue 1, while leaving 2 for the coming sections. Before entering into these discussions, however, it is important to provide a critical account of the studies that have previously attempted to estimate SDG networks.

2.1. The first generation of SDG networks

The first studies that tried to understand SDGs through the lens of networks appeared in the development literature a decade ago. According to their estimation procedures, we can classify them into two groups: (1) subjective studies that rely on qualitative information (e.g., the conceptual description of the variables), and (2) statistical ones that make use of panel data (countries through time). In group 1, subjectivity comes from determining links based on opinions. Thus, in this group, we can find studies that take either a brainstorming (e.g., expertise and stakeholder's knowledge) or a heuristic approach (e.g., informal text mining). In contrast, group 2 applies quantitative techniques to development-indicator data. In the networks estimated in group 2, each node corresponds to a specific indicator, while a link or edge denotes the strength of a relationship between nodes. Usually, these graphs have weighted edges and, in fewer cases, these edges have directions.

Among the subjective studies, we can find Pedercini and Barney [4]; Blanc [5]; Collste et al. [6]; Weitz et al. [7]; Allen et al. [8]. Examples of statistical studies are Czyzewska and Mroczek [9]; Ceriani and Gagliarano [10]; Castañeda et al. [11]; Cinicioglu et al. [12]; Pradhan et al. [13]; El-Maghrabi et al. [14].² Finally, the work of Zhou and Moinuddin [15] seems to be the only one standing in both camps.

For clarity, and to align our exposition with the literature semantics, we use the terms *policy issues*, *development indicators*, and *development*

goals (or just goals) interchangeably to refer to a specific topic where a policy can be directed (e.g., education, public health, poverty, etc.). Then, to refer to the development aspirations of a government or society, we use the term *objectives*. Likewise, an SDG is a development pillar that encompasses a set of policy issues that, *ex-ante*, are thought to be closely associated. Thus, we can think of policy issues as the topics of interest and of the objectives as the final states where the policymaker wants to arrive (i.e., specific values for a set of development indicators belonging to different SDGs).

In most of the first-generation studies, the aim has been to identify policy issues with synergistic effects to other goals (positive spillovers) in order to promote them. Similarly, estimating goals that face trade-offs (negative spillovers) is also important to discourage policies that improve certain indicators but obstruct other objectives. This type of analysis prevails in subjective studies, in which first- and second-order spillovers are inferred (e.g., Blanc [5]; Pradhan et al. [13]; Weitz et al. [7]). In other studies, different centrality measures are calculated with the purpose of identifying influential policy issues (e.g. [8,15]).

There are several ways in which subjective studies build SDG networks. First, edges can be constructed through information derived from previous studies. Second, knowledge from stakeholders and experts in specific fields can be leveraged to propose specific graphs. Third, the text of official documents (e.g., from the United Nations) can be mined to build networks of development goals connected by commonly mentioned policy issues. Regarding statistical studies, we identify three dominant approaches: inference through plain correlations [13,15,11], Bayesian techniques [9,10,12], and a co-occurrence methodology [14].³ The co-occurrence methodology establishes, first, if the performance of a country in a particular indicator is above the average country from nations with similar per capita income. Then, the method infers the likely co-occurrence between two indicators (i.e., if they have proximate mechanisms for delivering an above-average performance).

2.2. Shortcomings of the first generation

In the context of SDGs, where the ultimate goal is to facilitate the prescription and evaluation of public policies (besides inferring the interdependencies of the development process), it is important to specify broad desirable qualities of network-estimation methods. In our view, such attributes are scalability, replicability, specificity, directionality, and validity. Next, we elaborate on each of these concepts and suggest that these desirable properties are rarely fulfilled jointly in the first generation of SDG networks.

2.2.1. Scalability

A network-estimation procedure for SDGs should be easily scalable to incorporate as many socioeconomic indicators as possible (appealing to the multidimensional nature of development). This has become self-evident as we passed from the 48 indicators considered by the Millennium Development Project to the 232 (and growing) of the SDGs. Such high dimensionality calls for representing the policy space as a complex network of goals. This, however, is not trivial when the number of observations for each indicator (usually less than 10) is lower than the number of dimensions typically observed in development-indicator data. Another aspect of scalability is the capacity to estimate the network with relative ease. This is hardly achieved by methods relying on expert advice. For instance, a government may be unable to gather enough experts in so many different issues due to budgetary and time constraints, or simply because such experts are unavailable (something common among poor countries and, at the subnational level, even in

² Czyzewska and Mroczek [9], Ceriani and Gagliarano [10], Castañeda et al. [11], and Cinicioglu et al. [12] do not make a direct reference to SDGs; nonetheless, we consider them relevant because they are early contributions that use development-indicator data.

³ Note that, although Czyzewska and Mroczek [9] briefly discuss the challenge of inferring causality, none of these studies formally attempt to establish causal relationships, yet their results are usually interpreted as if this were the case.

middle and high-income nations). Furthermore, as the number of indicators increases, the number of experts needed to understand all the possible direct and indirect relations becomes impossible to fulfill.⁴

2.2.2. Replicability

Methods that facilitate the replication of empirical studies are desirable in all scientific endeavors. Hence, the techniques through which information is obtained and processed should be accessible to third parties who wish to verify or refine an estimated network. In this view, studies based on expert advice are rarely replicable because human capital is highly scarce in numerous developing countries. Furthermore, even if such experts exist, it does not mean that they are accessible. Moreover, the process through which they arrived at their original estimations might have changed because it may not be systematic or transparent, but rather the result of their individual experiences and perceptions. Finally, brainstorming methods are often biased through “echo chamber” effects, or the current conventional wisdom (e.g., the relationship between economic openness and job creation). Hence, dealing with such biases requires additional protocols on how to collect expert advice.

2.2.3. Specificity

Extensive evidence from real-world development policies has shown that context matters [16] and, hence, policies have to be adjusted to the social, economic, and political environment where they are supposed to be applied. For this reason, networks that are estimated by pooling data from different countries may fail to be specific to the country of interest (as it is done in Czychowska and Mroczek [9]; Ceriani and Gigliarano [10]; Cincioğlu et al. [12]; El-Maghrabi et al. [14]). Similarly, tweaking a “master” network –previously built for analyzing other countries– also runs the risk of neglecting a country’s context (as in Pedercini and Barney [4]). When time series are too short, a second-best alternative consists of pooling data from a reduced group of countries with structural similarity (for example, demonstrated through statistical cluster analysis).

2.2.4. Directionality

As the purpose of building SDG networks is to conduct evidence-based policymaking, one should be cautious when interpreting the outcomes of these estimates. For instance, it is not enough to identify the high centrality of an indicator to argue that the associated policy issue should be prioritized (e.g., when recognizing that the country’s electrification is critical for its development because it connects to many other topics). Among other things, it is crucial to identify the direction of the edges. Hence, estimations built on standard text mining, co-occurrence methods, or plain contemporary correlations fail to meet the property of directionality.

In principle, “common sense” can be used to discern the potential direction of an edge, as there may be indicators that cannot precede others. However, as we move to larger datasets, more complex topologies, and higher topical specificity, establishing directions in this fashion takes us back to the scalability problems found when using expert advice. Therefore, detecting directionality in a data-driven way is desirable.

2.2.5. Validity

As with any estimation procedure, network inference requires some form of validation to be considered scientifically reliable. In the context of SDGs, external validation through out-of-sample prediction is unfeasible because of the short span of the SDG time series. A popular alternative among network scientists is to create data-generating models and test how well the estimation methods recover the networks specified in such models.

⁴ This does not mean, however, that this kind of data is not useful to estimate SDGs. In fact, in section 5.2, we argue that expert advice and anecdotal experience can be used to construct complementary ex post validation tests.

This, however, is not possible with SDGs because there is no comprehensive understanding of how development indicators emerge from the interactions of socioeconomic agents at a much more disaggregated level. Therefore, an admissible alternative is appealing to validated methods, which were developed in other disciplines, but which are designed for data with a similar structure.⁵ Accordingly, none of the methods used in the first generation have passed any formal validation test.

It is important to note that we consider all the attributes indicated in Section 2.2 essential to network-estimation methods in the context of SDGs. However, their joint fulfillment represents a major challenge. Besides, it should be pointed out that all of the first-generation studies lack most of these requirements. Although, from our description, it is also clear that some of them have certain strengths when evaluating particular aspects.

2.3. What can be inferred from first-generation networks?

The first-generation of SDG networks has provided a first approximation of the policy space that is relevant to the 2030 Agenda. Among other useful estimates, we find associations between policy issues that share links, clusters of nodes obtained through community detection algorithms, signed edges indicating a positive or negative co-movement between pairs of variables, node centrality (e.g., degree, eigenvector, betweenness, or closeness) indicating influence on the network or ability to connect communities, and network sparsity showing if indicators tend to connect more with issues from their same SDG than with others. Besides these descriptive insights, one of the main virtues of the first-generation networks is that they have helped analysts and policy-makers grasp the complex and systemic nature of development.

First-generation networks might also be helpful to improve policy heuristics. For instance, in the co-occurrence approach [14], it is argued that policy priorities can be determined through two criteria: (1) the feasibility of improving an underperforming indicator (for similar countries) given its proximity to other nodes that perform relatively well (as indicated in density analyses), and (2) their potential for improving other indicators, measured through degree centrality of the intervened policy issue. Nevertheless, one should still be careful when interpreting these criteria. For example, a highly connected node might be the result of incoming links that are not observed when estimating undirected networks and, thus, its promotion could result in poor outcomes.

Finally, an important way in which first-generation networks can be used is not for direct policy interpretation, but as structural information for more comprehensive models. For example, together with other structural information such as input-output matrices, the tradition of “system dynamics” incorporates these networks as model components. This approach highlights the fact that policy intervention analysis cannot be solely based on an exercise of synergies and trade-offs of development indicator data alone. It suggests that the causal chain from resource allocation to changes in indicators is complex, limiting our ability to provide policy prescriptions out of network estimates of a partially observed system (i.e., raw sustainable indicator data).

2.4. On causality and policy prescriptions

The recurrent “correlation does not imply causation” phrase should have already come into the reader’s mind. Thus, it is pertinent to draw

⁵ The methods reviewed in the latter sections of this paper fall under this category of methods, that is estimation methods validated for their respective disciplines by their corresponding authors. In other words, by employing these methods, we are assuming that development-indicator data come from the data-generating processes used in such validation. It is important to mention that, while this validation strategy may not be ideal in the context of SDGs, it is still an improvement over first-generation studies, where most of the methods have no validation at all.

distinctions between the different types of inferences that can be achieved through existing network-estimation methods, as well as their underlying assumptions and limitations. This is especially important in the context of SDGs, where much of the analysis tends to be used to inform policymakers (unavoidably implying some sort of causation).

First, let us differentiate three types of inference relations: association, structural-dependence, and causation. An association is a co-movement between two variables without distinguishing their origin or direction. That is, when X (Y) changes, Y (X) is observed to change too.⁶ Structural-dependencies between variables have explicit directions; for example, when X varies, we can also observe changes in Y , but when the latter varies we do not need to observe changes in the former.⁷ Finally, causation indicates that there exists “cause and effect” relationships between variables. That is, X is a cause of another variable Y , when manipulations of the former systematically affect the outcomes of the latter, after controlling for a set of Z variables, possibly related to Y .⁸ From this explanation, it becomes evident that, for causal inference to be possible, a directional link between two variables is a necessary but not sufficient condition.

All first-generation SDG networks measure either associations or structural-dependencies, but not causation. However, causality may be the most relevant relationship to be inferred from an SDG network because, ultimately, one would like to provide advice on which nodes to intervene through public policies. Consequently, any policy advice derived solely from association and structural-dependence networks should be taken with a pinch of salt. This makes evident that, besides the recurrent correlation \neq causation argument, other important aspects are generally overlooked in the estimation of SDG networks, for example, how feasible is it to exercise a policy intervention in an indicator? do interventions take place at the same level of aggregation as the outcome variables? is it possible to identify all the confounding factors affecting causal relationships and/or the indicators? etc. While these questions are out of the scope of this paper, we offer some thoughts in Section 6.

It must be noted that most of the definitions of causation used by the methods reviewed in this paper fall into a particular type: the “dependence account” [17]. Here, causal factors are those whose presence makes possible the existence of one or several outcomes (effects).⁹ Because development indicators are aggregate variables (see Section 2.5), by estimating causal SDG networks under the dependence account and considering only indicator information, one implicitly assumes, among other things, that outcomes and interventions take place at the indicator level; something that may not hold in the context of SDGs. This is the case, for instance, when analyzing the impact of financial development on economic growth. Here, both the “causal variable” and the outcome are measured at the country level, while policy interventions

⁶ An association can be described in different forms: co-occurrence (*i.e.*, a pair of variables that tend to appear in tandem) and contemporaneous correlation (*i.e.*, variables that tend to exhibit a monotonic relationship).

⁷ In Bayesian networks, the link’s presence is given through a conditional probability $Pr(Y | X)$, so that X is the “parent” node and Y is the direct “child” (X precedes Y). In networks of non-contemporaneous correlations, the direction indicates a time sequence.

⁸ In mathematical terms, following the notation of Pearl [76], causation relates to probabilities of the type $Pr(Y | do(X_0))$, where $do(X_0) := do(X = X_0)$ means that variable X has been intervened. Thus, classical statistical inference is not the same as causal statistical inference.

⁹ According to Casini and Manzo [66], acyclic causal graphs make use of a dependence account of causation and a horizontal view of the causal mechanisms. Such a view consists of a set of stable relationships in a network of variables. They can be represented as algebraic equations or as network topologies, such as cascades ($X \rightarrow Y \rightarrow Z$) and branches ($Y \leftarrow X \rightarrow Z$). An alternative account is, for example, the “production” one, in which causal factors are those that help to *generate* or bring about specific outcomes. This account, however, requires a different set of techniques that lie beyond the statistical methods here covered.

(*e.g.*, rules of supervision and competition) are implemented at the microlevel (*e.g.*, banks and other financial institutions). Hence, recommendations derived from the dependence account need to be carefully considered.

2.5. Challenges of SDG data

Now that we established the desirable qualities for SDG-network estimations and differentiated the types of inferences, it is important to provide some clarifications on the nature of the data and on the challenges related to building an “ideal” method. First, it should be clear by now that estimating SDG networks through statistical methods require quantitative data. The relevant data available for this type of analysis consist of development indicators. Second, these indicators have a time-series structure (*i.e.*, today’s values tend to depend on previous ones). Therefore, an ideal method should take temporal dependence into account. Third, these time series are short (*e.g.*, usually less than 20 annual observations, depending on how many countries and indicators are collected for the study). Thus, the ideal method should be able to handle the statistical problems associated with small samples. Fourth, SDG data are high dimensional; hence an ideal method should also work in this setting. Fifth, development indicators are aggregated variables at the country (or sub-national) level, for example, the gross domestic product, yet some of them are the result of lower-level processes or conformed by other unobserved variables.¹⁰

The fact that development indicators have temporal dependencies does not imply that time is their only or main determinant. Likely, many of the associated policy issues are continuously influenced by policy interventions to the system. Because of this, a common concern in the time-series analysis is the presence of “structural breaks” within the time interval under study, especially if the sampling period involves different government administrations. Therefore, while time series based methods account for the time dependence; the fact that available data has few observations, mostly collected with low frequency (yearly), means that there may be something to be learned from simpler methodologies that assume independence, but which still incorporate intervention based causal learning.

We should add that an “ideal” method for estimating SDG networks with development-indicator data should contain directed and weighted edges. Few of the first-generation studies present these two topological features jointly. However, their inclusion is critical in so far as arrows are indispensable to account for a correct dependence or causal structure (either of intervention impacts or flows between endogenous variables), and weights are essential to measuring the “strength” or influence of the relationship between goals.

These challenges serve as guidelines to select the methods that are best suited to estimate SDG networks. Furthermore, in selecting estimation procedures, we privilege those techniques that are transparent

¹⁰ Another challenge that may be a concern to some readers is the imbalance in the number of indicators between SDGs. The relevance of this problem depends on whether the inferences are made at such an aggregate level. Nevertheless, it is important to mention that this issue is endemic to all studies that try to cover the high-dimensional nature of development. In part, such imbalance is a historical legacy of how different development agendas have evolved over the years, giving priority to the measurement of some policy issues before others. For example, it is not surprising to see more development indicators about poverty and economic growth than environmental ones. This is so because the Millennium Development Project prioritized the eradication of poverty, while environmental topics recently became important.

and replicable through the provision of publicly available code.¹¹

3. Broad families of network-estimation methods

Network estimation methods are abundant in the literature of complex systems, in statistics, and machine learning (e.g., Han and Zhu [18]; Smith et al. [19]; Linderman and Adams [20]; Aragam et al. [21]). In this paper, we focus on a subset of methodological families that have explicit underlying concepts of causation. Nevertheless, and as a comparative benchmark, we also present results from computing rank correlations (Spearman). In total, we classify the methods into five families (although we don't consider our classification to be definite or exhaustive).

Intending to make this paper accessible to a broad audience, we try to keep technicalities and jargon at a minimum. However, using some mathematical expressions might be unavoidable. Therefore, before explaining each family, it is useful to introduce some notation. Let us consider the development indicators as variables $\{X_1, X_2, \dots, X_p\}$, where p is the total number of indicators. Let G denote a network whose nodes represent individual development indicators. Each link or edge in G is informative about the presence, direction, and magnitude of an association, structural-dependence, or causal strength between two nodes (i.e., policy issues or goals). Finally, we denote the weight of edge $i \rightarrow j$ as G_{ij} .

3.1. Correlation thresholding

One of the most commonly used techniques to estimate networks is thresholding correlation matrices (e.g., [18,22–24]). Its popularity owes, to a considerable extent, to its simplicity, straightforward interpretability, and lack of restrictive assumptions about the variables' interdependencies (i.e., a particular type of structure that the estimated network should have). Broadly speaking, the approach consists of, first, obtaining a complete correlation matrix. Secondly, given a minimum acceptable correlation magnitude—referred to as threshold—an initially empty network G is populated by edges between nodes i and j . An edge in G exists whenever the magnitude of the correlation between X_i and X_j is larger than the threshold.

The existence of different correlation measures and the possibility of considering lagged values give place to numerous methodological variations in this family. To mention a few, we can find the Pearson correlation, Spearman, Kendall, partial correlation, etc. Besides, there are various ways in which one could select a threshold, for example, by arbitrarily defining it, by choosing a significance level for the p -value of each correlation, or by selecting a threshold under which the resulting network structure best fits the desired property (e.g., number of edges, the distribution of connections, clustering patterns, etc.). Finally, depending on the temporal nature of the correlation, one could generate directed (with lagged time series) or undirected networks (with contemporary correlations).

3.2. Granger causality

The most visible representative of time-series causal inference methods was proposed by Granger [25]. Granger causality (also known as G-causality) can be defined as follows: for a given time point t and time series X and Y , “We say that X_t is causing Y_t if we are better able to predict Y_t using all available information than if the information apart

¹¹ This means that more sophisticated methods that, potentially, could jointly tackle different challenges may not be applied in this study. However, we provide further details about some of these methodologies in the next sections. Finally, we should add that it is particularly important for development studies, a multi-disciplinary field where researchers and consultants are not experts in the science of networks, to have transparent access to novel methods through publicly available code.

from X_t had been used” [25]. This view of causality has spawned diverse methods to estimate networks.

Most methods in this family stem from time series analysis and, more specifically, from the vector autoregressive family of models (e.g., [26–28]). Nevertheless, we can also find G-causality in methodologies developed using concepts from physics, for example, mixtures of G-causality and transfer entropy [29]. A common characteristic among these approaches is the use of hypothesis testing to determine the existence of an edge between a pair of nodes ($i \rightarrow j$), representing indicator variables X_i and X_j . Such tests evaluate whether the prediction of X_j at time t —when only using its previous history—improves by including X_i at time t , and other potential lags, in the set of predictors (e.g. [27]).

Overall, Granger-causality network-estimation methods generate directed networks and weighted edges as well. However, they have two major drawbacks when it comes to the challenges enlisted in Section 2.5. First, evaluating the presence or absence of an edge is often tested without considering all the other variables in the system. Second and most importantly, given that these tests are built on autoregressive models, they perform poorly when applied to short time series. Strictly speaking, G-causality methods are more in line with structural-dependence than causal inference. This is so because structural-dependence is mainly concerned with discovering the links' directions between two nodes (e.g., its co-movement in specific directions), and not in terms of the effect of exogenous interventions.

One of the first studies to explicitly address causation in SDG networks was developed by Dörge et al. [30], who took a Granger-causality approach. To overcome the limitations of G-causality, they pool cross-national data, which in their case, precluded obtaining country-specific networks.

3.3. Chordal information filtering graphs

Filtering graph methods were initially proposed as a mechanistic way to remove edges in highly dense and complex networks, for example, in full correlation matrices. The original purpose of this approach was to eliminate spurious links that are irrelevant in terms of weight and structural integrity of the network [31]. Building on these ideas, more recent developments provide network estimation procedures that perform in a fast and scalable fashion (e.g. [32]). Among these proposals, LoGo [33] stands out as a suitable framework to estimate causation when the lengths of the time series are shorter than the number of variables.

In broad terms, one of the main characteristics of the filtered graphs family is how they account for the network's structural integrity. For example, methods such as Triangulated Maximally Filtered Graphs (TMFG) [32] and LoGo [33] propose to begin the estimation process with a network that satisfies a mathematical property called the “planarity constraint”. In simple terms, the planarity constraint is a problem in which, given a complex and dense network, one must redraw it on a surface (a plane) in such a way that as many edges as possible are preserved, but that none of them intersect each other. Although there is no clear intuition or a causality-driven motivation for using the planarity constraint, it has been shown that the resulting networks can be used in forecasting.¹²

One of the main strengths of this family is its ability for inferring networks from data with more variables than observations. In particular, LoGo can be used to infer the links' weights and to generate a potentially causal interpretation of their direction. Note, however, that chordal filtered networks might be biased toward substructures called triangles (3-cliques) and/or other k -cliques (k nodes, all connected to the other k

¹² Another common characteristic is that the estimated graphs are “chordal”, a topological property for which there is no clear motivation or interpretation in terms of causality (“every cycle of length > 4 has a chord, an edge not belonging to the cycle that joins two non-adjacent vertices of the cycle” [32]).

– 1 nodes). An important drawback of LoGo is that, when a dataset has numerous indicators that are proportionally interrelated (often the case in SDG data), the estimation may result unfeasible. This is so because LoGo performs matrix inversions requiring specific properties (being nonsingular) that are unlikely to hold in high-dimensional and largely correlated datasets. Hence, one may need to resort to strategies to ameliorate such correlations, for example, discriminating indicators in somehow arbitrary ways, or introducing noise.

3.4. Statistical structure learning

The last three decades have seen numerous studies of the multivariate causal structure of systems (e.g., [34–42, 77]). Currently, there exists a wide variety of statistical methods for this purpose. Many of these techniques were developed from different theoretical standpoints. Therefore, there is a considerable variation in terminology, often referring to similar or even the same ideas. In trying to overcome this lack of clarity, we classify the family of statistical learning methods into causal graphical models and structural causal models. Here, we provide a brief description for each one.

3.4.1. Causal graphical models

Graphical models are sometimes referred to as Bayesian networks. They consist of a network of conditional dependencies where the nodes represent the variables and the links their dependencies. Altogether, these networks are modeled as directed acyclic graphs (DAGs).¹³ In statistical terms, a DAG imposes a conditional-dependence structure that induces a probability distribution over a set of random variables [34, 41]. Thus, causal graphical models are designed to estimate the DAG of conditional dependencies.

A typical estimation procedure to construct G consists of performing conditional independence tests across all the observed variables, to discard edges from an initial G . This procedure is performed until all the topological properties of a DAG have been met.¹⁴ A problem with this approach, however, is that there may be multiple possible DAGs that can, for example, meet the results of the conditional independence checks, thus making them all candidate estimates of G .¹⁵ Thus, these methods usually generate multiple estimates of G , which can then be further discriminated by using auxiliary methods. Examples of this family are the PC algorithm (pcalg) [43] and the *Intervention calculus when DAG is absent* (IDA) framework [44].

3.4.2. Structural causal models

Structural causal models are also referred to as structural equation models (and should not be confused with econometric structural models [45]). In their most general form [37,41], they can be characterized in terms of “parent-child” relations between variables and noise terms. More specifically, for a given variable of interest X_i (a development indicator), we say that its parents PA_i are a subset of other nodes in the system that are structural determinants of the value of X_i . Together, PA_i and a noise component N_i determine X_i through a function f_i , i.e., $X_i := f_i(PA_i, N_i)$. Therefore, depending on the way that one chooses to model the parent-child relationship, it is possible to define a variety of structural equation models as shown in Table 1.

This approach provides ample flexibility to model structural relationships and noise terms. Aragam and Zhou [46], for example, propose a scalable network estimation method based on a linear Gaussian structural equation model. This approach can handle cases where the sample size is much smaller than the number of variables in the network. Peters et al. [47], in contrast, provide an alternative approach for

Table 1

Different specifications of structural causal models.

Structural causal model	$X_i = f_i(X_{PA_i}, N_i)$
Additive noise model	$X_i = f_i(X_{PA_i}) + N_i$
Causal additive model	$X_i = \sum_{k \in PA_i} f_{ik}(X_k) + N_i$
Linear Gaussian model	$X_i = \sum_{k \in PA_i} \beta_{ik} X_k + N_i$ and $N_i \sim N(0, \sigma_i)$

Based on Table 7.1 from Peters et al. [41]. Other popular specifications are nonlinear Gaussian additive noise models and linear non-Gaussian acyclic models, among others.

multivariate time series (TiMINo) that does not require a linear Gaussianity assumption. Sometimes, structural causal models are also referred to as Bayesian networks. This confusion in terminology arises because the set of structural equations implies a graphical representation of the dependence structure of the variables X_1, \dots, X_p . For example, if the parents of X_1 and X_2 are $PA_1 = X_5$ and $PA_2 = X_1$ respectively, then, the functional assignments $\{X_1 = f_1(X_5, N_1), X_2 = f_2(X_1, N_2), X_5 = f_5(N_5)\}$ imply a network $X_5 \rightarrow X_1 \rightarrow X_2$. Note that, in contrast to graphical models, structural equation models provide a data-generating mechanism that can be easily implemented by following the structural assignments. More specifically, following the previous example, values of (X_1, X_2, X_5) can be obtained by drawing a value n_5 from the noise distribution of N_5 . Then, we can compute the value $x_5 = f_5(n_5)$, from which $x_1 = f_1(x_5, n_1)$ and, subsequently $x_2 = f_2(x_1, n_2)$ can be obtained as well (where n_1 and n_2 are noise values drawn from the distributions of N_1 and N_2 , respectively). On the contrary, in the graphical models of the previous subsection, such a data-generating mechanism is not proposed.

In general, both causal graphical models and structural equation models introduce some common assumptions that may be considered as important limitations. First, the estimated network G is often required to be acyclical. This implies that reinforcing cycles (virtuous and vicious) are not allowed, for example, as with the bidirectional effects between improvements in public health and education. A second important assumption is that the data points are assumed to be independent draws of the distribution that generates the development indicator. This means that the temporal dependence of the data needs to be added to the initial formulation of the models, as done, for example, by Peters et al. [47].¹⁶

3.5. Physics-inspired approaches

In addition to statistically inspired approaches, there are techniques with a fundamentally different basis; for example inspired on physical and biological phenomena. For instance, Hu et al. [29], and Servadio and Convertino [48] employ transfer entropy (i.e., the amount of directed information flowing from one variable to another) to infer the association network in multivariate time series. Furthermore, Hempel et al. [49] developed the *inner composition alignment* (iota) method to infer directed networks from short time series.¹⁷

Among the methods inspired by the natural sciences, there is one class that has a strong mathematical backbone: *state-space reconstruction methods*. These techniques spawned from the literature on nonlinear dynamical systems and, in recent years, they have become relatively popular to estimate causal relationships in physical, biological, and

¹⁶ In this family, there exist more sophisticated methods that can deal with the possibility of cycles (e.g., [39]), with nonlinearity (e.g., [39, 47]), with time dependence, and there are different theoretical results that tackle the presence of latent variables as well [41].

¹⁷ It is important to note that, although some methods belong to a single-family, several others share characteristics with different families. For example, Servadio and Convertino [48] propose a transfer-entropy approach comparable to the one used in correlation thresholding. Similarly, the chordal-filtering- graphs approach of Aste and Di Matteo [33] also employs transfer entropy to quantify the edges that carry causal interpretations.

¹³ A DAG is a network in which no path can return to its starting node.

¹⁴ In addition, likelihood-based approaches can also be employed to estimate G .

¹⁵ This is known as the “identifiability problem”.

ecological systems (e.g., [50–54]). The basic idea behind state space reconstruction is that any variable in the system can be used individually, through its time series, to recover the attractor to which the system tends to evolve (the so-called Takens' theorem).¹⁸

Inspired in this theory, Sugihara et al. [55] developed the method of *cross convergent mapping* (ccm). This technique allows estimating directed networks by testing causation between a pair of variables, even if they are weakly coupled (note that bidirectional causality is possible). The test is built around the idea that X causes Y when the dynamics of the former can be recovered (predicted) by the dynamics of the latter. In other words, the information concerning X is already reflected in the evolution of Y , and cannot be removed from the universe of all possible causative factors. Accordingly, through Y 's attractor manifold (a mathematical object that represents a geometric space in which the variable moves, usually denoted as M_y), it is possible to make local-neighborhood predictions on X 's manifold M_x . As local-neighborhood tests can be data demanding, several methods have been proposed to deal with short time series for noisy and high-dimensional systems (e.g., [56–58]).

Overall, state-space reconstruction methods have an appealing theoretical basis and allow the estimation of directed networks without imposing topological constraints. Nevertheless, they also have drawbacks that should be mentioned. On the theoretical side, these models assume that the system is in a “steady-state” fluctuating around the attractor. Thus, for a system that is transitioning to a different attractor, these methods would provide the wrong causal relations. Furthermore, the assumption of remaining near a specific attractor may conflict with the principles of socioeconomic development; a process with technological innovations, emerging social norms, and considerable political turbulence that may push nations toward different attractors. In addition, the weights of the inferred networks do not have a clear interpretation other than a score about how well one variable predicts the other.¹⁹

4. Eligible estimation methods for SDG networks

As we have argued in Section 2.5, SDG data convey several challenges that turn most network-estimation methods ill-suited. For example, most methods in the classical Granger-causality family are inadequate because of the short length of development-indicators time series.²⁰ Therefore, after a comprehensive review of numerous approaches, we have selected four methods to perform empirical estimations given the data restrictions. These are *intervention calculus when the DAG is absent* through the PC algorithm (pcalg), *concave penalized estimation through sparse Gaussian Bayesian networks* (sparsebn), *inner composition alignment* (iota), and *cross convergent mapping* (ccm). That is, the first two belong to the family of statistical structure learning, while the last two correspond to the physics-inspired family.²¹

Overall, no single method “ticks all boxes” when it comes to the challenges of SDG networks. However, the chosen ones, while assuming that all variables relevant to the causal system are being considered, have the common ability to cope with a small number of observations and a large number of variables. Therefore, while the eligible methods are not a definitive statement about the correct way to estimate SDG

networks, they do provide an important guide on a variety of ways to exploit SDG data. Finally, and as mentioned above, all of the chosen methods have publicly available software. In this section, we explain further details on each of these four methods to perform the empirical analysis in Section 5.

4.1. Intervention calculus through the PC algorithm

The *intervention calculus when DAG is absent* (IDA), proposed by Maathuis et al. [44], is a causal graphical model designed to estimate causation directly from observational data. The method consists of three steps. First, the basic structure of the dependence network –called the skeleton– is inferred (i.e., a network displaying only undirected edges and no weights). The skeleton is particularly useful as it highlights all the possible causal relations between all variables in the network. In the second stage, edge directions are inferred, so the “completed partially directed acyclic graph” (CPDAG) is created (see Fig. 1.b).

The CPDAG is called partial because the inference of some directions is theoretically infeasible, hence, CPDAG displays those edges as undirected. The reason for this infeasibility lies in the fact that multiple DAGs can be obtained from the CPDAG. These multiple DAGs encode the same distributional and conditional dependence constraints inferred from the data, making them indistinguishable through purely observational information. Nevertheless, the CPDAG can be thought of as a summary network displaying all the possible directed dependence networks that fit the data constraints. Finally, the third stage estimates the size of the causal effects (edge's weights) for each of the possible relations in the CPDAG, generating a weighted network.

Following Maathuis et al. [44], the first two stages can be performed through a method called the PC algorithm [43]. The method starts with an undirected graph that includes all possible edges. Then, it iteratively deletes edges that pass (do not reject) a conditional independence test. In the first iteration, a simple independence test is performed between two connected nodes. If the nodes are found to be independent, that edge is deleted from the network. Next, once all pairwise tests are performed, each remaining pair of connected nodes (i, j) is tested for conditional independence given any single node connected to i , or any single node connected to j . Whenever a pair of nodes is found to be conditionally independent, its edge is removed. Then, in a new iteration of conditional independence tests, the size of the conditioning set is increased by one node. This edge deletion process leads to the skeleton of the inferred dependence network.²²

Once the network skeleton is obtained, edge directions, directions are inferred by considering each triplet of connected nodes ($i - k - j$). If in the first step node k was never part of a “passed” (not rejected) conditional independence test between i and j , then the triplet takes the form $i \rightarrow k \leftarrow j$, (see Spirtes et al. [43] for full details on the PC algorithm). Finally, IDA is used to estimate the upper and lower bounds of the weights (of the causal effects). The idea is to estimate the effect that interventions in variable X_i can have on variable X_j . This effect can be interpreted as the change in the mean of X_j , when X_i is forced to take a given value (see Maathuis et al. [44] and Lauritzen [34] for full details). In this paper, we use the minimum reported bound as the causal effect (Fig. 1.a).²³

¹⁸ For a review of this literature see Ma et al. [42].

¹⁹ See Cobey and Baskerville [72] for more technical criticisms on ccm and the conceptual challenges of state space reconstruction.

²⁰ This may explain the limited results obtained by Dörge et al. [30].

²¹ Although chordal information filtering methods such as LoGo provide accessible code, the required matrix inversions are unfeasible because of the highly correlated nature of some development indicators (which could also be caused or heightened by the short length of the time series). Similarly, and despite its code availability, TIMIno was unable to provide estimates for our data. For this reason, we do not use them in our empirical application.

²² An alternative method designed to deal with more variables than observations has been developed by Meinshausen et al. [75]. The authors also provide tools to compare network estimation methods of this family.

²³ The R package *pcalg* [79] provides all the required tools to estimate these networks on development-indicator data. In this paper, we use the code provided for Gaussian variables.

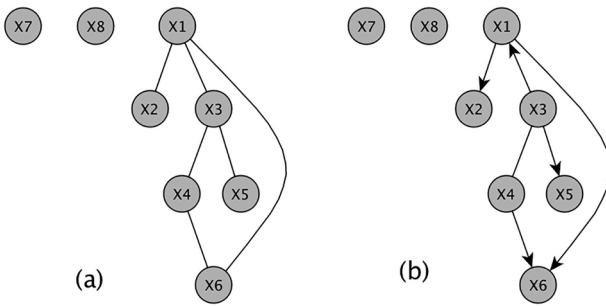


Fig. 1. Skeleton and CPDAG of the PC algorithm. (a) Example of the skeleton of a graphical model with 8 variables. The undirected edges represent the presence of a dependence relation between nodes. However, the directions are absent. (b) Example of a completed partially directed acyclic graph, providing directions for some of the edges in the skeleton. Some edges, however, remain undirected as the method cannot decide which is the correct direction.

4.2. Concave penalized estimation through sparse Gaussian Bayesian networks

Aragam and Zhou [46] propose a structural causal model for systems composed of a large number of variables that follow a multivariate Gaussian distribution. They estimate the coefficients of a linear Gaussian structural equation model (SEM) (see Table 1) by solving a convex optimization problem.²⁴ With the estimated parameters in hand, it is possible to construct a directed weighted network through the system of equations defined in the SEM. There are two key innovations in this approach. The first is a parameter transformation that allows linking the optimization problem to the estimation of high-dimensional Gaussian distributions (*i.e.*, allows estimating large networks). The second is a penalty term in the optimization procedure that allows the estimation of sparser networks (*i.e.*, eliminates edges to address overfitting).²⁵ Note, however, that the estimated coefficients (*i.e.*, the weights) cannot be directly interpreted as causal relations without further analysis of interventional data. Instead, the method “focuses on finding the most parsimonious representation of the true distribution [of the variables] as a set of structural equations” [46].²⁶

4.3. Inner composition alignment

Iota [49] is a method inspired in physical dynamics where a constantly changing system can be analyzed by describing its different states and possible ways to transition between them. For example, in a biological system such as gene regulatory networks, one way to infer when genes are coupled is through pairwise tests between time series. More specifically, *iota* identifies the coupling of two time-series when the ranking of the values of one of them is similar to the ranking of the other. In detail, it checks if a non-decreasing ordering applied to time series X_i can also be applied (approximately) to X_j , to achieve a non-decreasing ordering of X_j .

Hempel et al. [49] build on these ideas to relate two time-series by how monotonic (increasing or decreasing) X_j is when it is sorted based on the ordering of the other time series (X_i). More precisely, they propose a statistic that captures the degree of monotonicity of the reordered

variable X_j to test the presence and direction of the edge $i \rightarrow j$. This statistic can be understood as an index that depends on the number of intersections between the reordered time series X_j and the imaginary horizontal lines fixed at the values of X_j , when the ordering is given by X_i (see Fig. 2). Through pairwise computations and a permutation test of the aforementioned statistic, *iota* aims at estimating a dependence network G . The method works for noticeably short time series and allows inferring the direction of associations between variables.²⁷ Furthermore, given that it checks for monotonicity on reordered variables, this method can also detect nonlinear relations between the time series.

Note that *iota* has been constructed for truly short time-series and does not impose assumptions about the structure of the dependence network. However, rather than being based on a generative theory of the data, it is based on an observed pattern. Therefore, it is not clear if it can infer causal or structural-dependence networks.²⁸

4.4. Cross convergence mapping

The method known as *ccm* was proposed by Sugihara et al. [55], based on the idea that complex systems exhibit nonlinearities and, thus, they tend to evolve toward strange attractors. Therefore, when two variables (X, Y) are coupled, their corresponding manifolds (M_x, M_y)—obtained through the coordinates of their lagged values—should approximately describe the attractor. Hence, if the information embedded in M_y can predict the information embedded in M_x , it is argued that X causes Y . Reconstructing the state space, however, requires a large number of observations when a system is high-dimensional.

In its supplemental material, [55] claim that *ccm* works well when studying data of fisheries with approximately 70 species and only 35–40 observations. Moreover, recent extensions attempt to improve the resolution of the manifolds for small samples. This can be done, for instance, by increasing the number of observations through replications of the same unit of analysis but in a different physical space [57]. An alternative is to use a neural network to estimate the cross-map between the manifolds [56].

According to the *ccm* method, causation between two nodes exists when there is convergence in fitting; that is, when the correlation coefficient between the predictor and the observed variable increases rapidly with the series length. In the seminal paper of Sugihara and coauthors, a nearest-neighbor algorithm is used to establish the weights to be used when calculating the predicted values.

Despite the time series of our data having less than 20 observations, we use the original *ccm* methodology. We do not apply the multispatial *ccm* [57] because we only have one observation per country/year.

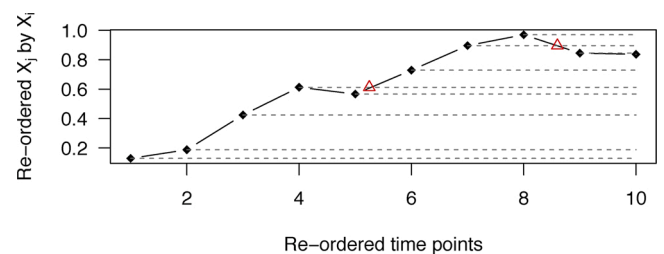


Fig. 2. Intuition behind the *iota* statistic. An illustration of the number of crossings (marked as red triangles) between the reordered time series X_j and the imaginary horizontal dashed lines fixed at the values of X_j , according to the ordering obtained from X_i .

²⁴ This is an approximation of the non-convex optimization problem that arises when directly considering the optimization of the log-likelihood of the data through the SEM [46].

²⁵ Additionally, Aragam and Zhou [46] develop another parameter transformation to convert the non-convex optimization problem into a convex one.

²⁶ The R package *sparsebn* [21] provides all the required tools to estimate these networks on development-indicator data. This tool assumes that the data are an independent and identically distributed (i.i.d.) sample from a multivariate Gaussian distribution.

²⁷ Note that, although *iota* is inspired in time series observations, it does not explicitly consider the temporal dependence of variables.

²⁸ The R package *iota* (tocsy.pik-potsdam.de/iota.php) provides all the required tools to estimate these networks on development indicator data.

However, as the software associated with this procedure is publicly available, we use it for our estimations (with only one replica).²⁹ The *ccm* approach does not impose topological requirements on G . However, more nodes imply higher dimensionality of the underlying attractor. Finally, the method does consider the time-dependence structure of the indicators.

5. Data and results

5.1. Data

In spite of worldwide efforts to build comprehensive datasets tracking the 17 SDGs of the 2030 Agenda [3], it is still challenging to assemble information with a large coverage of countries, indicators, and observations at the same time. Because, in this paper, we perform a comparative analysis between methods, the estimation of SDG networks demands the following data requirements: (i) all countries should contain the same indicators, (ii) all observations must be contemporaneous (they should cover the same sampling period), (iii) no indicator should be a linear renormalization of another (*i.e.*, no two indicators should have a perfect linear correlation), and (iv) all indicators should exhibit temporal variation (we assume that a constant indicator is independent of all others).³⁰

Taking all previous considerations into account, we build a dataset where we reach a compromise between maximizing the length of the sampling period, the number of indicators, the SDGs covered, and the number of countries in the sample. In particular, we prioritize the number of indicators and years because our motivation for estimating SDGs is to map a complex web of structural-dependencies between numerous policy issues. Our sample for this study consists of four countries from different continents: Egypt (EGY), Indonesia (IDN), Mexico (MEX), and Turkey (TUR). We analyze the relationships between 87 development indicators (covering 16 of the 17 SDGs) during the 1995–2014 period for each country in an independent fashion. Each indicator has been normalized between 0 and 1.³¹ The indicators have been adjusted so that larger values denote better outcomes. The main source of the dataset is the United Nations Global SDG Database [59]. In addition, we complement it with the World Bank Sustainable Development Goals Database [60] and the poverty indices from the World Bank Poverty and Equity Data. [61]. Fig. 3 presents the summary statistics for the indicators aggregated in terms of the corresponding SDG and country.³²

Fig. 3 shows the differences between the four countries. For example, Egypt is substantially less developed in SDG 10 (reduced inequalities), while Indonesia is in SDG 1 (no poverty). It is also interesting to observe that all the four countries are poorly developed in SDG 6: clean water and sanitation and SDG 14: life below water (from the four indicators that were available for the four nations).

With this general picture of the data, we proceed to present the estimated networks. We do this in a twofold fashion. First, we show aggregate results at the level of the SDGs. That is, we study the total number of incoming and outgoing edges (and their weights) in each SDG. This allows gaining insights into the structure of synergies and

trade-offs between SDGs at the level of each country. Second, we introduce and compute formal metrics to compare the topologies of the estimated SDG networks, with a special emphasis on comparing the networks produced for the same country through different methods. This provides a more rigorous understanding of the implications of using different frameworks.

5.2. SDG networks

As discussed in Section 4, we present results for each country on four specific methods:³³ *iota*, *ccm*, *pccalg*, and *sparsebn*. In addition we also provide calculations obtained from correlation networks (*lcorr*). The latter networks are built by computing pairwise Spearman correlations, where the explanatory variable is lagged by one period. In this way, the resulting networks are not necessarily symmetrical. We only keep those edges where the correlation has a p -value lower than 10%.³⁴

Table 2 shows the summary statistics about the connectivity in the estimated SDG networks. When talking about the number of links, connectivity is usually referred to as degree. When edges have weights, however, the correct term is strength (the sum of the weights connected to a node). We have split each network into edges with positive and negative weights. In other words, we analyze synergies and trade-offs separately. Notice that the topological structures of the estimated networks vary considerably, as reflected by their synergies and trade-offs across each country's estimates. However, the estimated weights have slightly different interpretations because each method has a different underlying concept of causality (if any at all); hence, a comparison of the raw magnitudes of synergies/trade-offs between the different methodologies may not be relevant. In the light of these variations, one method can be preferred over another depending on the type of the analysis to be pursued, or, perhaps more importantly, on the consistency of the estimates with prior information (see below).

Notice that the five methods analyzed here do not identify the same country as the one with maximum synergistic strength. Indonesia is identified twice (*lcorr*, *pccalg*), Mexico twice (*ccm*, *sparsebn*), and Turkey once (*iota*). A similar assessment emerges when comparing the identity of the nodes with maximum strength for each country across estimation methods, except for Turkey, where *lcorr* and *sparsebn* present the same identifier for trade-offs [46]. While in the case of synergies, none of the methods identify the same node for a specific country. These results point out the method-dependent sensitivity of interpretations about the strength of complementary (or conflicting) links between policy issues. Next, we present additional analyses comparing the estimated network.

As a substantial part of the discussions in the sustainable development literature relates to encouraging complementary policies, let us take a closer look at synergistic networks. Fig. 4 shows the structure of synergies as the links that are directed from one indicator to a group of nodes (we provide a similar illustration for trade-offs in Fig. 6 of Appendix B). The first feature that stands out in Fig. 4 is the striking

³³ The estimation method was applied independently for each country and considering its whole multivariate time-series.

³⁴ Network topology is a very important property in the analysis of complex systems, and because the focus of this paper is to highlight the multiple estimation methodologies and some of their practical differences; we opt for using a p -value of 10%, instead of the more standard 5%, as this choice of p -value would better illustrate the global structural differences between the estimated networks from the different methodologies. This is the case as the 5% p -value for some methods leads to network estimations that are reduced to only a few edges, where no sense of global structure can be appreciated. However, it is important to note that the choice of larger p -values tends to increase the presence of falsely estimated interactions. Thus, when utilizing these network estimation methodologies in practice, a robust selection of the p -value should be considered (see Table 7 in Appendix B).

²⁹ The R package multi-spatial CCM provides all the required tools to estimate these networks on development-indicator data.

³⁰ Hence, we focus on nonduplicated time-varying indicators.

³¹ For indicator i of country c in period t , we normalize according to a standard procedure: $[I_{i,t,c} - \min(I_{i,\dots})] / [\max(I_{i,\dots}) - \min(I_{i,\dots})]$, where I is a matrix with development indicators of all countries, not only those in the sample. The reason for using all countries in the normalization is that cross-national data provide an adequate benchmark to normalize indicators (*i.e.*, zero values should be assigned to those indicators that equal the lowest levels across all countries and years).

³² See Table 5 in Appendix A for a complete list of the 87 indicators and their respective SDGs, and Table 6 for the numeric values of the summary statistics.

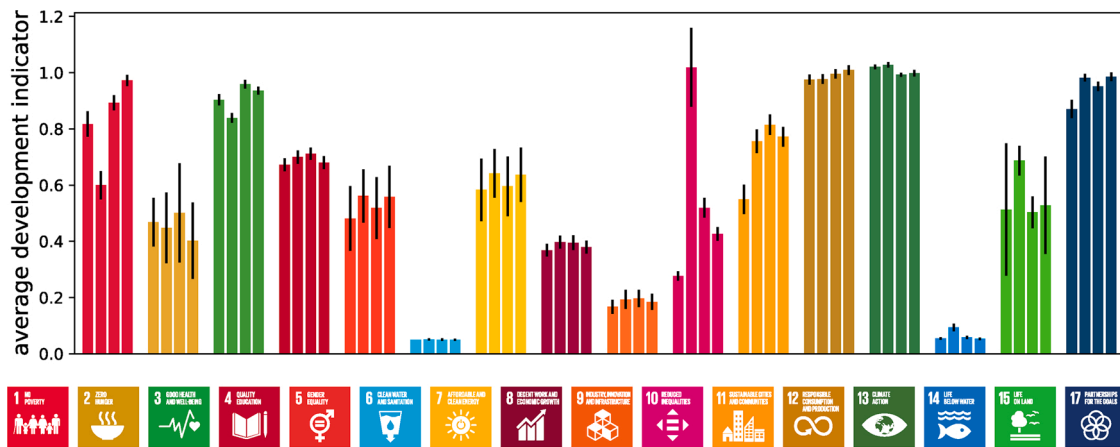


Fig. 3. Summary statistics of development indicators by country and SDG. Each color corresponds to an SDG. All indicators in each SDG have been aggregated by averaging, first, their values through time and, then, through all the indicators in the SDG. Within each color, each bar corresponds to a country in the following order: EGY, IDN, MEX, and TUR. The vertical black lines denote the standard error. All indicators in SDG 6 have nearly zero values. Therefore, to better illustrate their comparison against all others, all indicators in this plot have been added an amount of 0.05.

Table 2
SDG-network strength statistics.

Method	EGY		IDN		MEX		TUR	
	Synergy	Trade-off	Synergy	Trade-off	Synergy	Trade-off	Synergy	Trade-off
<i>lcorr</i>	30.99	25.1	43.06	24.48	37.82	29.86	41.43	30.62
	(12.11)	(11.3)	(19.21)	(16.51)	(15.83)	(14.0)	(18.66)	(16.03)
	[31]	[51]	[40]	[43]	[5]	[32]	[45]	[46]
<i>iota</i>	12.83	11.1	13.01	11.01	11.66	10.44	13.38	11.4
	(6.53)	(4.79)	(5.9)	(5.0)	(3.69)	(4.78)	(6.2)	(5.71)
	[5]	[43]	[4]	[7]	[27]	[13]	[36]	[81]
<i>ccm</i>	21.1	14.48	22.23	11.17	27.24	19.06	26.25	13.36
	(15.55)	(10.9)	(16.86)	(10.45)	(17.11)	(11.62)	(19.14)	(11.94)
	[56]	[68]	[41]	[36]	[23]	[36]	[24]	[36]
<i>pcalg</i>	1.29	0.64	1.33	0.53	1.26	0.44	1.15	0.55
	(0.82)	(0.8)	(0.97)	(0.66)	(0.79)	(0.64)	(0.86)	(0.67)
	[74]	[36]	[16]	[35]	[3]	[19]	[11]	[29]
<i>sparsebn</i>	8.9	7.03	9.63	7.66	10.62	8.0	6.3	3.36
	(4.89)	(4.31)	(5.02)	(4.33)	(5.81)	(5.3)	(3.8)	(2.63)
	[85]	[8]	[15]	[42]	[1]	[70]	[3]	[46]

Reported connectivity is the average strength (the sum of the weights) of a node's connections, separated into positive (synergies) and negative (trade-offs) links. Standard deviations are inside round brackets. Nodes with the maximum strength are inside squared brackets.

differences between physics-inspired methods (*iota* and *ccm*) and statistical structure learning methods (*pcalg* and *sparsebn*).³⁵

On the one hand, *iota* and *ccm* generate considerably denser SDG networks than *pcalg* and *sparsebn*. In addition, there are also important differences between *pcalg* and *sparsebn*, with the former estimating considerably fewer synergies than the latter. These differences vary from country to country, highlighting the importance of not pooling arbitrarily cross-national data to estimate SDG networks. On the other hand, when making within-country comparisons, it is also evident that, depending on the method of choice, the distribution of weights across SDGs is significantly different. For instance, in the case of Egypt, *iota* suggests the presence of many synergies coming in and out from SDG 1 (no poverty). While synergies entering to this node are drastically reduced in *ccm*, they vanish in the other two methods.

To provide more detail on how differently the inferences can be derived from these methods, let us rank each SDG in each country and method. We produce this ranking according to how synergistic an SDG is to the others. That is, for a given SDG in a specific network, we take its outgoing edges with destinations in different SDGs. By averaging their weights, we obtain an estimate of how synergistic that SDG is to other

goals. Table 3 shows the top five SDGs in each ranking. No country presents the same list of indicators for all methods, and only Egypt and Indonesia exhibit the same list for two methods: *lcorr* and *ccm* (although in a different ordering). However, 5 indicators from Egypt are repeated in at least three lists of the top five, 4 indicators from Indonesia, 4 from Mexico, and 3 from Turkey. These results indicate that, given the sensitivity of the different methods, a potential approach could be finding a consensus among networks. Due to space limitations, this concept will be developed more formally in Appendix D.

5.3. Prior-informed tests

As we have explained in Section 2.2, a desirable quality in an SDG network estimation method should be external validity. However, this is hardly achievable in the context of SDGs because the data-generating process is rather complex and takes place at different aggregation levels and time scales. Therefore, one has to appeal to some form of internal validity, such as a qualitative judgment on the consistency between the properties of the data and the method requirements (e.g., no temporal dependence). Furthermore, there are ways in which prior information may be used to make a judgment on how valid an estimated SDG network can be. For instance, the subjective information employed in some of the first-generation methods can be exploited to –ex-post–

³⁵ Remember that *pcalg* and *sparsebn* assume no temporal dependence in the data.

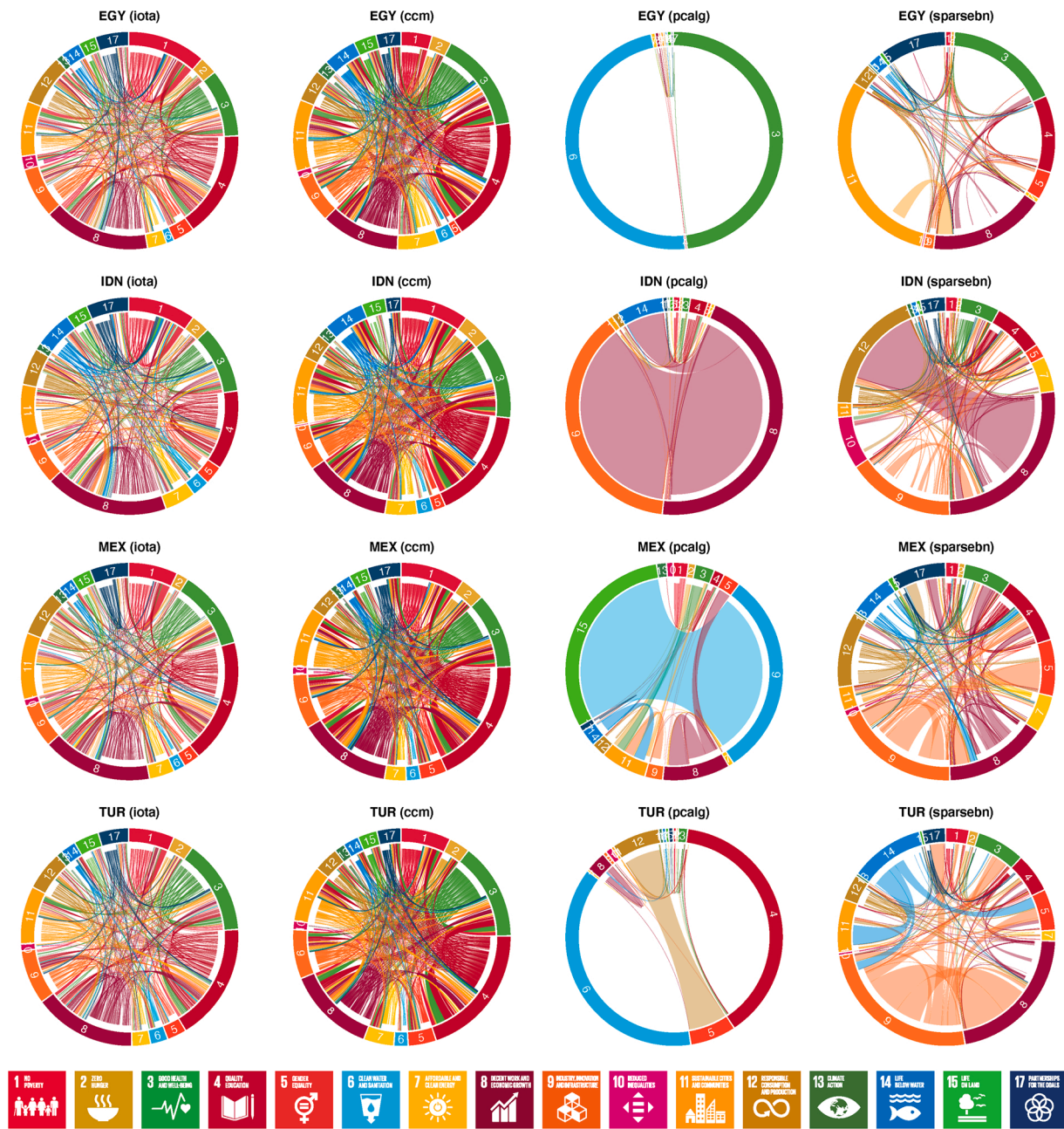


Fig. 4. Estimated synergies between SDGs by country and method. Weights have been normalized so they are comparable in terms of proportion to all other weights in the network. Edge colors indicate the SDG from which they originate. The length of each colored segment of each circle should not be interpreted, as they are the result of a visualization algorithm.

Table 3
Top five SDGs by outgoing synergies.

Country	<i>lcorr</i>	<i>iota</i>	<i>ccm</i>	<i>pcalg</i>	<i>sparsebn</i>
EGY	15, 14, 2, 11, 3	1, 2, 3, 5, 10	2, 3, 11, 14, 15	3, 7, 10, 14, 15	2, 8, 9, 11, 14
IDN	2, 14, 1, 3, 11	1, 6, 14, 15, 17	1, 2, 3, 11, 14	3, 4, 5, 7, 8	3, 8, 9, 12, 13
MEX	2, 1, 5, 11, 14	3, 4, 6, 11, 15	1, 2, 3, 4, 11	2, 3, 6, 9, 14	7, 8, 9, 12, 14
TUR	2, 3, 1, 5, 17	2, 3, 6, 11, 17	1, 2, 3, 4, 11	3, 7, 12, 13, 14	1, 2, 9, 12, 14

The outgoing synergy of an SDG is computed by averaging the weights of positive outgoing links with a destination in nodes (indicators) belonging to a different SDG. In other words, we do not count within-SDG synergies. Each set of five SDGs has been sorted from most to least synergistic (left to right).

partially validate an estimation. That is, rather than –ex-ante– relying on subjective data to estimate the network, this information can be leveraged to evaluate whether an estimated set of edges constitutes false positives (FP), or whether absent edges translate into false negatives (FN). Here, we show how this approach sheds light on the performance of the estimation methods.

To evaluate FP, we identify 10 pairs of development indicators for which one would not expect any causal or structural-dependence relation in any direction and sign. Therefore, a false positive means that a method has estimated a synergy or a trade-off that, a priori, is known to have no logic in the context under study (goes against common sense). Thus, there can be up to 40 false positive links, by considering both

directions and positive or negative weights.³⁶ For the FN, we have identified 10 pairs of indicators that almost certainly have some type of relationship, whether this is causal, a structural-interdependency, synergistic, a trade-off, or it goes in any direction. A false negative is the complete absence of any type of edge between a pair of nodes in the estimated network when, in reality, that edge exists.³⁷ Thus, there can be up to 10 FN at most. Appendix C provides the two lists of indicators that have been identified for this analysis.

Table 4 presents the number of FP and FN for each estimated network. Concerning the FN, *pcalg* shows the best performance. However, the difference with the other methods is not significant. With respect to FP (i.e., links that should not be there), *lcorr* presents the highest number, and thus worst raw performance. However, the large number of links established (T) in this method increases the probability of false positives. Thus, if we normalize by T, *lcorr* and *ccm* present a relatively large number of FP (>0.002) for the four country cases, but this threshold is not crossed in two methods: *iota* (Egypt and Mexico) and *sparsebn* (Egypt and Turkey). Not surprisingly, the extremely sparse networks estimated with *pcalg* receive the best possible scores. Thus, in our judgment, *iota*, *sparsebn*, and *pcalg* are potential candidates, although a more extensive and rigorous test should be conducted.

5.4. Comparing SDG networks

Although the data allow comparisons between countries, we would not want our results to be interpreted as having potential policy implications because the aim of the paper is not to estimate the definite networks for our country examples. Instead, our purpose is to understand how the five methodologies differ in their results. For this reason, we emphasize methodological comparisons rather than country ones.

To provide a more rigorous analysis of the structural differences between alternative methods, it is necessary to briefly elaborate on some relevant metrics. A common practice in network science is to measure topological differences between networks by focusing on specific traits. For example, some studies try to quantify discrepancies on how network connectivity is distributed across nodes, while others concentrate on internal communities and clusters. Broadly speaking, we can characterize these different emphases in a spectrum between *connectivity* and *structure*. Different metrics have been developed to compare networks in this spectrum. It should be noted that, as of today, there is no “gold standard” when it comes to these comparisons. Furthermore, because networks may be directed or undirected, weighted or unweighted, single or multilayered, etc., a metric that is popular for a certain type of network might not be well defined for others. For these reasons, we employ three metrics that are well suited for comparing SDG networks and try to cover the spectrum between connectivity and structure. In the three metrics, higher values imply larger dissimilarity between networks.

5.4.1. ψ Distance

Xu et al. [62] proposed a comparative measure for the connectivity of two networks called the ψ distance. This measure is general enough to consider weights and directions, besides, it can handle networks with different sets of nodes. In the context of this study, the ψ distance becomes quite simple because we compare networks with the same development indicators (the same sets of nodes). More specifically, this metric is the sum of the magnitudes of the differences between the weights of the same edge in the two different networks. Formally, for

³⁶ We can obtain up to 40 FP for 10 pairs of nodes because, for each pair, there can be 4 possible edges: one positive and another negative from X to Y, and one positive and another negative from Y to X.

³⁷ For the analysis of FN, we have identified indicators whose relationship could even be considered tautological. That is, the two indicators in a pair belong to the same SDG and are conceptually closely related.

two networks G and Q , the ψ distance is defined as

$$d_{\psi} = \sum_{i,j \in V_{G,Q}} |G_{ij} - Q_{ij}|, \quad (1)$$

where $V_{G,Q}$ is the set of nodes in G and Q , and G_{ij} denotes the weight of edge $i \rightarrow j$ in network G , Q_{ij} is defined equivalently. Note that for the raw weights of the methods considered, this distance will focus on relating methods with closer weight scales.

5.4.2. Hamming distance

Somewhere in the spectrum between connectivity-based and structure-based metrics we can find the popular Hamming distance [63]. Interestingly, Richard Hamming developed this metric to detect coding errors, but he had a general theory in mind for any object that could be studied through geometry. For instance, the metric has become extremely popular to study distances between words. In network science, this metric is also known as the edit distance, and it is very intuitive. Given a network G , its Hamming distance from another network Q is the minimum number of edits required to transform it into Q . In our context, since G and Q have the same nodes, edits translate into edge removals and additions. Computing the Hamming distance is usually performed through an algorithm provided in most programming languages.³⁸ It is important to clarify that this metric does not consider edge weights.

5.4.3. NetEmd

NetEmd [64],³⁹ is a more recent network-comparison method initially developed to directly compare the structure of complex systems. Similar to previous structural metrics [65], NetEmd measures the frequency of occurrence of different subgraphs. For example for nodes i, j, k, l, m , the occurrence of triangles is given by the presence of edges $i-j, j-k, k-i$; 4-stars exist if we observe edges $j-i, j-k, j-l, j-m$; square subgraphs are given by edges $i-j, j-k, k-l, l-i$ and so forth. In total, NetEmd uses 30 types of subgraphs.⁴⁰

Given G , NetEmd builds the probability distribution of observing a node that is attached to x subgraphs of a particular configuration (i.e., to one of the 30 configurations).⁴¹ Then, it compares the shape of the distribution to the one obtained from another network Q . The comparison involves some normalization techniques, after which, the final product is a score indicating the average distance between G and Q in terms of the similarity of the shape of their subgraph distributions.⁴²

5.4.4. Results

Once the distance metrics have been defined, we can compute them for each estimated SDG network. Fig. 5 presents the different comparisons of the five networks per country and per metric. We can see that statistical structure learning methods (*pcalg* and *sparsebn*) are very similar in terms of the Hamming and ψ distances, but not under NetEmd. In contrast, the *lcorr* and *iota* methods are quite different under the three metrics, while *pcalg* and *sparsebn* are also very different in comparison to *lcorr*. We can also say that *iota* exhibits moderate levels of resemblance with respect to *pcalg* and *sparsebn* irrespective of the metric considered. Moreover, the topological discrepancies between *ccm* and the other methods are salient, yet they vary notoriously depending on the distance

³⁸ The R package *pcalg* provides the function *shd* to compute the Hamming distance between two networks.

³⁹ The R package for NetEmd can be downloaded from github.com/alan-turing-institute/network-comparison.

⁴⁰ These sub-networks can further be decomposed into 73 configurations when the node position is considered. By default, NetEmd uses these 73 configurations.

⁴¹ More precisely, to a specific location in one of the 30 sub-graphs.

⁴² More specifically, NetEmd uses the minimum Wasserstein distance [78] – also known as the earth mover’s distance – and a few normalization procedures.

Table 4
Prior-informed false positives and false negatives.

Method	EGY			IDN			MEX			TUR		
	FP	FN	T	FP	FN	T	FP	FN	T	FP	FN	T
lcorr	12	2	4288	11	2	5130	14	0	5406	18	0	5932
iota	1	0	1195	3	2	1207	1	1	1112	5	2	1279
ccm	8	2	2799	8	2	2825	11	3	3818	10	0	3357
pcalg	0	0	84	0	0	81	0	0	74	0	0	74
sparsebn	0	0	693	4	3	752	2	1	810	0	0	420

FP: false positives. FN: false negatives. T: total number of directed edges estimated in the network.

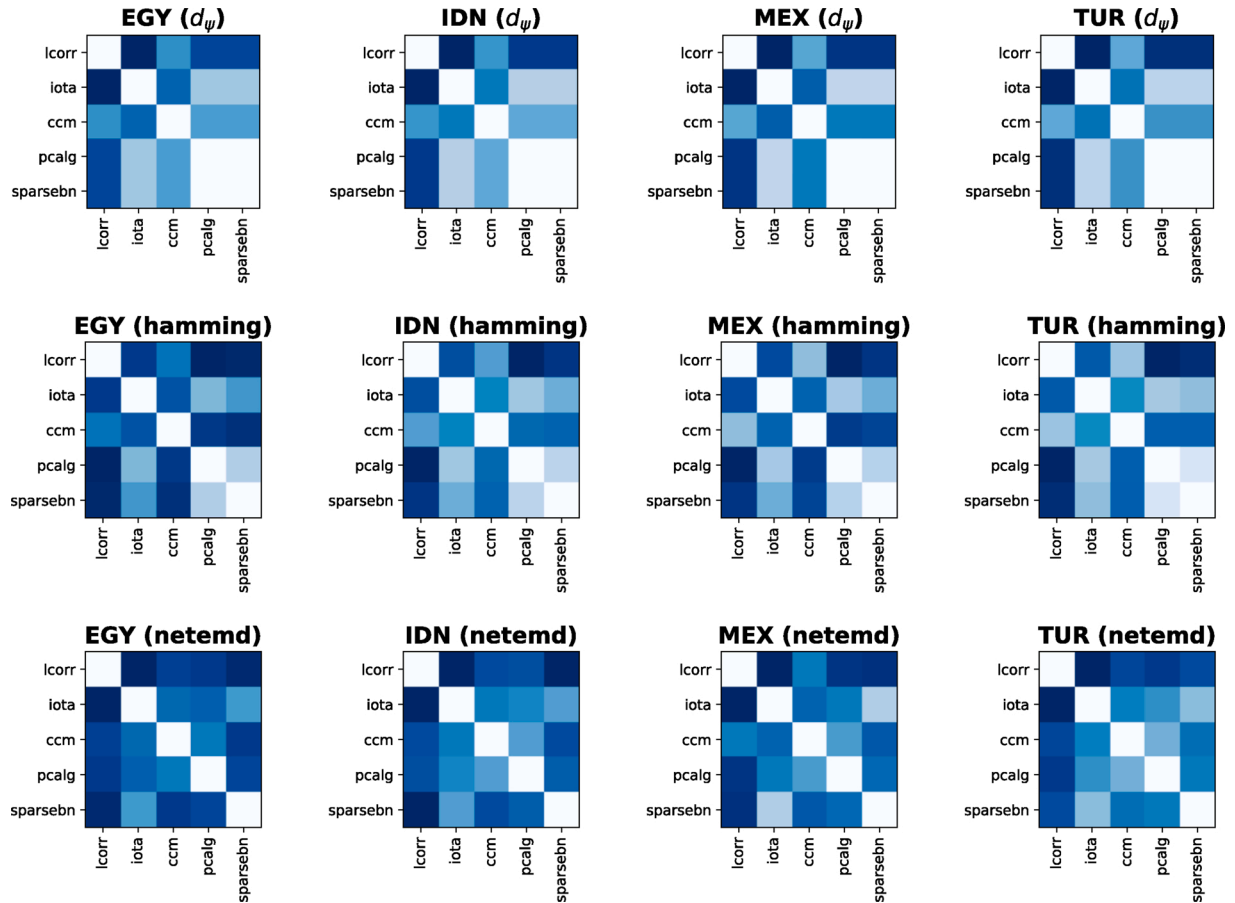


Fig. 5. Differences between estimated SDG networks by country and metric. Darker tones denote more difference between networks.

metric used. All in all, we conclude that, except for the pair (*sparsebn* and *pcalg*), the methods here considered produce different topologies. Presumably, the reason for *sparsebn* and *pcalg* to show closer results is that they belong to the same family and both tend to produce sparse networks.⁴³

6. Discussion and conclusions

This paper provides an overview of the different network-estimation families that are better suited for inferring synergies and trade-offs between SDGs. However, it warns that policy recommendations directly obtained from these networks should be taken with a pinch of salt, even when using causal inference methods. This is so because network-estimation methods only consider the dependence account of causation, not the production one [17]. In the production account, causal factors are those that help to generate or bring about, specific outcomes.

In other words, it calls for modeling the specification of vertical causation which, in the context of SDGs, allows the generation of data from mechanisms that are known to be part of the SDG system. Among these mechanisms, we would need to include the policy-making process, agent-level interactions, adaptation, among others, which are currently not part of the SDG indicator datasets.

A production account of causation is better-suited for advising development policies because policy interventions take place at a much lower level of aggregation than the one provided by the currently available data for SDGs (*i.e.*, the development indicators). Hence, the need for considering lower-level system dynamics of high resolution that account for vertical causal mechanisms seen at the global level. Scenarios of policy interventions, which take place at an agent level perspective (high resolution), and emergent properties at the macro level is the “bread and butter” of complex adaptive systems such as markets and ecosystems. Casini and Manzo [66] argue that agent-based models are a more adequate tool to establish vertical mechanisms and study the causal impact of policy interventions. These models are particularly useful in the absence of empirical information at multiple

⁴³ Appendix D presents a consensus-network approach to compare topologies.

levels of aggregation.

A way to conciliate a more policy-oriented account of causation in the context of the 2030 Agenda is by estimating SDG networks and then using them as inputs for an agent-based model of the policymaking process (with vertical causal mechanisms that stem from the policy-making process). Some progress has been made in this vein of research through policy priority inference [67,68]. Starting with a set of exogenous objectives, indicators' initial values, and an estimated network of structural-dependencies, the model generates endogenous policy priorities that can be used to make causal inference in a variety of topics such as policy resilience [69], ex-ante policy evaluation [70], policy coherence [71], and corruption [74].⁴⁴ In this framework, the purpose of the network is only to specify how the SDG indicators co-evolve once certain variations are induced by the adaptation of the budgetary allocations.

6.1. Concluding remarks

This paper contributes to the literature of the SDGs by reviewing first-generation network studies and by comparing networks estimated with state-of-the-art methodologies coming from statistics, complexity science, and machine learning. In contrast to the first-generation tools, estimations based on the newer methods have the potential of meeting more desirable properties for SDG networks (scalability, replicability, specificity, directionality, and validity). An endemic property of SDG indicators, however, is that they have a limited number of observations.

By selecting five specific methods, we estimate networks using data from four countries (Egypt, Indonesia, Mexico, and Turkey), covering 20 years and 16 out of the 17 SDGs. The chosen methods for these estimations were: Spearman correlations (*lcorr*), intervention calculus through *pclag*, concave penalized estimation through *sparsebn*, *ccm*, and *iota*. The second and third belong to the statistical structure learning family, while the last two are physics-inspired methods. Although this list is not exhaustive, it offers important insights into the key problems of SDG network estimations.

Our main conclusion is that applied methods –suitable for small sample sizes– produce topologically different networks. Even when these differences are not very sharp, they still produce differentiated implications (e.g., differences in the top-5 synergistic SDGs). Sensitivity to the method of choice and the impossibility of comprehensive validation tests (due to the opaqueness of the data-generating process of SDGs) demands filters for selecting a suitable method. We consider two criteria: discarding those methods whose theoretical assumptions are less consistent with the nature of SDG data and discarding those methods that produce relatively more false positives.

According to our results, *pclag*, *iota*, and *sparsebn* are the best methods when analyzing the number of false positives. However, in the first one, the number of estimated links is extremely low. Then, on the theoretical front, *pclag* and *sparsebn* have shortcomings in terms of assuming temporal independence and acyclical graphs. Likewise, the edge weight estimation in *iota* is not theoretically very sound and, thus, it needs to be combined with other methods. Despite that these five procedures generate very different topologies, the performance of the networks at the country-level still requires a more nuanced evaluation in further studies. Besides, one could consider a consensus network built from a collection of graphs. Therefore, consensus methods are a topic that deserves to be explored in future research.

This paper provides the first view into “second-generation”

⁴⁴ In this strategy, causation is an emergent feature of the policy-making dynamics provided by the model, not by network. Therefore, the network should rather be interpreted as the set of structural dependencies (a stylized fact) that the data-generating process should take into account; not as causal linkages. This ameliorates concerns of finding causal relations in data that may be the result of numerous unobservable confounders.

methodologies for the inference of synergies and trade-offs between SDGs. Naturally, new questions will arise as these methods are disseminated in the sustainability community, so there should be a concerted effort to answer them. Some examples are: How accurate are the different network estimation methodologies? How robust are the methodologies for small samples or the presence of structural breaks? Are some methods better suited for different types of countries? Which of the methodologies is better able to capture true causal links (or structural-dependencies)? How robust are the methods to violations of their assumptions, e.g., temporal dependence, presence of cycles, linearity, and Gaussianity?

Declaration of Competing Interest

The authors confirm that there is no potential conflict of interest in this paper.

Acknowledgement

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.im.2020.103342>.

References

- [1] M. Nilsson, D. Griggs, M. Visbeck, Policy: Map the interactions between Sustainable Development Goals, *Nature News* 534 (7607) (2016) 320.
- [2] OECD, Policy Coherence for Sustainable Development 2018: Towards Sustainable and Resilient Societies, OECD, Paris. OCLC, 2018, 1040196886.
- [3] General Assembly, (25-Sep-2015). Resolution Adopted by the General Assembly on 25 September 2015. Technical Report A/70/L.1, United Nations, 2020.
- [4] M. Pedercini, G. Barney, Dynamic Analysis of Interventions Designed to Achieve Millennium Development Goals (MDG): The Case of Ghana, *Socioecon. Plann. Sci.* 44 (2) (2010) 89–99.
- [5] D. Blanc, Towards Integration at Last? The Sustainable Development Goals as a Network of Targets, *Sustain. Dev.* 23 (3) (2015) 176–187.
- [6] D. Collste, M. Pedercini, S. Cornell, Policy Coherence to Achieve the SDGs: Using Integrated Simulation Models to Assess Effective Policies, *Sustain. Sci.* 12 (6) (2017) 921–931.
- [7] N. Weitz, H. Carlsen, M. Nilsson, K. Skanberg, Towards systemic and contextual priority setting for implementing the 2030 agenda, *Sustain. Sci.* 13 (2) (2018) 531–548.
- [8] C. Allen, G. Metternicht, T. Wiedmann, Prioritising SDG Targets: Assessing Baselines, Gaps and Interlinkages, *Sustain. Sci.* (2018).
- [9] M. Czyżewska, T. Mroczek, Bayesian Approach to the Process of Identification of the Determinants of Innovativeness, *Finansowy Kwartalnik Internetowy e-Finanse* 10 (2) (2014) 44–56.
- [10] L. Ceriani, C. Gagliarano, Multidimensional Well-Being: A Bayesian Networks Approach. Technical Report 399, ECINEQ, Society for the Study of Economic Inequality, 2016.
- [11] G. Castañeda, G. Íñiguez, F. Chávez-Juárez, The Complex Network of Public Policies. An Empirical Framework for Identifying Their Relevance in Economic Development. Background Paper, Governance and the Law, The World Bank, 2017.
- [12] E. Cinioglu, G. Ulusoy, S. Ekici, F. Ülengin, B. Ülengin, Exploring the interaction between competitiveness of a country and innovation using bayesian networks, *Innov. Dev.* 0 (0) (2017) 1–36.
- [13] P. Pradhan, L. Costa, D. Rybski, W. Lucht, J. Kropp, A Systematic Study of Sustainable Development Goal (SDG) Interactions, *Earths Future* 5 (11) (2017) 1169–1179.
- [14] M. El-Maghrabi, S. Gable, I. Osorio-Rodarte, J. Verbeek, Sustainable Development Goals Diagnostics: An Application of Network Theory and Complexity Measures to Set Country Priorities, *World Bank Working Paper* (2018) 1–22. WPS8481.
- [15] X. Zhou, M. Moinuddin, Review of the SDG index and dashboards: an example of Japan's global ranking results, *IGES Working Paper*. (2016).
- [16] D. Rodrik, *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*, Princeton University Press, Princeton, 2009.
- [17] N. Hall, Two concepts of causation, in: J. Collins, N. Hall, L. Paul (Eds.), *Causation and Counterfactuals*, MIT Press, Cambridge MA, 2004, pp. 225–276.
- [18] L. Han, J. Zhu, Using Matrix of Thresholding Partial Correlation Coefficients to Infer Regulatory Network, *Biosystems* 91 (1) (2008) 158–165.
- [19] S. Smith, K. Miller, G. Salimi-Khorshidi, M. Webster, C. Beckmann, T. Nichols, J. Ram-sey, M. Woolrich, Network modelling methods for fMRI, *NeuroImage* 54 (2) (2011) 875–891.

- [20] S. Linderman, R. Adams, Discovering latent network structure in point process data. *Proceedings of The 31st International Conference on Machine Learning*, 2014, pp. 1413–1421, pages.
- [21] B. Aragam, J. Gu, Q. Zhou, Learning Large-scale Bayesian Networks With the Sparsebn Package, arXiv preprint arXiv:1703.04025, 2017.
- [22] V. Boginski, S. Butenko, P. Pardalos, Statistical Analysis of Financial Networks, *Comput. Stat. Data Anal.* 48 (2) (2005) 431–443.
- [23] W.-Q. Huang, X. Zhuang, S. Yao, A Network Analysis of the Chinese Stock Market, *Phys. A Stat. Mech. Its Appl.* 388 (14) (2009) 2956–2964.
- [24] K. Dimitrios, O. Vasileios, A network analysis of the greek stock market, *Procedia Econ. Financ.* 33 (2015) 340–349.
- [25] C. Granger, Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica* 37 (3) (1969) 424–438.
- [26] G. Castagneto-Gissey, M. Chavez, F.D.V. Fallani, Dynamic Granger-Causal Networks of Electricity Spot Prices: A Novel Approach to Market Integration, *Energy Econ.* 44 (2014) 422–432.
- [27] M. Barigozzi, C.T. Brownlee, NETS: network estimation for time series, SSRN preprint (2017), <https://doi.org/10.2139/ssrn.2249909>.
- [28] X. Gao, S. Huang, X. Sun, X. Hao, F. An, Modelling Cointegration and Granger Causality Network to Detect Long-Term Equilibrium and Diffusion Paths in the Financial System, *R. Soc. Open Sci.* 5 (3) (2018), 172092.
- [29] Y. Hu, H. Zhao, X. Ai, Inferring Weighted Directed Association Network from Multivariate Time Series with a Synthetic Method of Partial Symbolic Transfer Entropy Spectrum and Granger Causality, *PLoS One* 11 (11) (2016), e0166084.
- [30] G. Dörge, V. Sebestyén, J. Abonyi, Evaluating the Interconnectedness of the Sustainable Development Goals Based on the Causality Analysis of Sustainability Indicators, *Sustainability* 10 (10) (2018) 3766.
- [31] M. Tumminello, T. Aste, T. Matteo, R. Mantegna, A Tool for Filtering Information in Complex Systems, *Proc. Natl. Acad. Sci.* 102 (30) (2005) 10421–10426.
- [32] G.P. Massara, T. Matteo, T. Aste, Network filtering for big data: triangulated maximally filtered graph, *J. Complex Netw.* 5 (2) (2017) 161–178.
- [33] T. Aste, T. Di Matteo, Sparse Causality Network Retrieval from Short Time Series, *Complexity* (2017).
- [34] S. Lauritzen, *Graphical Models*, Oxford Science Publications. Clarendon Press, 1996.
- [35] M. Kalisch, P. Bühlmann, Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm, *J. Mach. Learn. Res.* 8 (2007) 613–636. March.
- [36] G. Lacerda, P. Spirtes, J. Ramsey, P. Hoyer, Discovering cyclic causal models by independent components analysis. *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, Virginia, United States. AUAI Press, Arlington, 2008, pp. 366–374. UAI'08, pages.
- [37] J. Pearl, Causal inference in statistics: an overview, *Stat. Surv.* 3 (2009) 96–146.
- [38] P. Bühlmann, Causal statistical inference in high dimensions, *Mathematical Methods of Operations Research* 77 (3) (2013) 357–370.
- [39] A. Hyvärinen, S. Smith, Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models, *J. Mach. Learn. Res.* 14 (2013) 111–152. January.
- [40] S. Shimizu, LiNGAM: non-gaussian methods for estimating causal structures, *Behaviormetrika* 41 (1) (2014) 65–98.
- [41] J. Peters, D. Janzing, B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*, MIT Press, Cambridge, MA, USA, 2017.
- [42] H. Ma, S. Leng, C. Tao, X. Ying, J. Kurths, Y.-C. Lai, W. Lin, Detection of time delays and directional interactions based on time series from complex dynamical systems, *Phys. Rev. E* 96 (1) (2017), 012221.
- [43] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction, and Search*, vol. 81, Springer, 1993.
- [44] M. Maathuis, M. Kalisch, P. Bühlmann, Estimating high-dimensional intervention effects from observational data, *Ann. Stat.* 37 (6A) (2009) 3133–3164.
- [45] J. Heckman, E. Vytlacil, Chapter 70 econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation, in: J. Heckman, E. Leamer (Eds.), *Handbook of Econometrics*, vol. 6, Elsevier, 2007, pp. 4779–4874.
- [46] B. Aragam, Q. Zhou, Concave Penalized Estimation of Sparse Gaussian Bayesian Networks, *J. Mach. Learn. Res.* 16 (1) (2015) 2273–2328.
- [47] J. Peters, D. Janzing, B. Schölkopf, Causal inference on time series using restricted structural equation models, in: C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2013, pp. 154–162, 26.
- [48] J.L. Servadio, M. Convertino, Optimal Information Networks: Application for Data-Driven Integrated Health in Populations, *Sci. Adv.* 4 (2) (2018), e1701088.
- [49] S. Hempel, A. Koseska, J. Kurths, Z. Nikoloski, Inner composition alignment for inferring directed networks from short time series, *Phys. Rev. Lett.* 107 (5) (2011), 054101.
- [50] L. Heskamp, A. Abeelen, J. Lagro, J. Claassen, Convergent Cross Mapping: A Promising Technique for Cerebral Autoregulation Estimation, *Int. J. Clin. Neurosci. Ment. Health* 1 (Suppl. 1) (2014) S20.
- [51] H. Ye, E. Deyle, L. Gilarranz, G. Sugihara, Distinguishing Time-Delayed Causal Interactions Using Convergent Cross Mapping, *Sci. Rep.* 5 (2015) 14750.
- [52] A. Tsonis, E. Deyle, R. May, G. Sugihara, K. Swanson, J. Verbeten, G. Wang, Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature, *Proc. Natl. Acad. Sci.* 112 (11) (2015) 3253–3256.
- [53] J. McBride, X. Zhao, N. Munro, G. Jicha, F. Schmitt, R. Kryscio, C. Smith, Y. Jiang, Sugihara Causality Analysis of Scalp EEG for Detection of Early Alzheimer's Disease, *Neuroimage Clin.* 7 (2015) 258–265.
- [54] Y. Wang, J. Yang, Y. Chen, P. Maeyer, Z. Li, W. Duan, Detecting the Causal Effect of Soil Moisture on Precipitation Using Convergent Cross Mapping, *Sci. Rep.* 8 (1) (2018) 12171.
- [55] G. Sugihara, R. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, S. Munch, Detecting causality in complex ecosystems, *Science* 338 (6106) (2012) 496–500.
- [56] H. Ma, K. Aihara, L. Chen, Detecting Causality from Nonlinear Dynamics with Short-term Time Series, *Sci. Rep.* 4 (2014) 7464.
- [57] A. Clark, H. Ye, F. Isbell, E. Deyle, J. Cowles, G. Tilman, G. Sugihara, Spatial convergent cross mapping to detect causal relationships from short time series, *Ecology* 96 (5) (2015) 1174–1181.
- [58] B. Zhang, W. Li, Y. Shi, X. Liu, L. Chen, Detecting Causality from Short Time-Series Data Based on Prediction of Topologically Equivalent Attractors, *BMC Syst. Biol.* 11 (7) (2017) 128.
- [59] United Nations Statistics Division, United Nations Global SDG Database, 2016. <https://unstats.un.org/sdgs/indicators/database>.
- [60] The World Bank Group, Sustainable Development Goals Database, 2016. <https://datacatalog.worldbank.org/dataset/sustainable-development-goals>.
- [61] The World Bank Group, World Development Indicators (WDI) – Data Catalog, 2010 world-development-indicators, <https://datacatalog.worldbank.org/dataset/>.
- [62] Y. Xu, S. Salapaka, C. Beck, A distance metric between directed weighted graphs, 52nd IEEE Conference on Decision and Control (2013) 6359–6364.
- [63] R. Hamming, Error detecting and error correcting codes, *Bell Syst. Tech. J.* 29 (2) (1950) 147–160.
- [64] A. Wegner, L. Ospina-Forero, R. Gaunt, C. Deane, G. Reinert, Identifying Networks with Common Organizational Principles, *J. Complex Netw.* (2017).
- [65] W. Ali, T. Rito, G. Reinert, F. Sun, C. Deane, Alignment-Free Protein Interaction Network Comparison, *Bioinformatics* 30 (2014) i430–i437.
- [66] L. Casini, G. Manzo, Agent-based Models and Causality: A Methodological Appraisal, *The IAS Working Paper Series* 2016 (7) (2016).
- [67] G. Castañeda, F. Chávez-Juárez, O.A. Guerrero, How do governments determine policy priorities? Studying development strategies through networked spillovers, *J. Econ. Behav. Organ.* 154 (2018) 335–361.
- [68] O. Guerrero, G. Castañeda, Policy Priority Inference: a Computational Method for the Analysis of Sustainable Development. SSRN Preprint, [dx.doi.org/10.2139/ssrn.3604041](https://doi.org/10.2139/ssrn.3604041), 2020.
- [69] G. Castañeda, O. Guerrero, The Resilience of Public Policies in Economic Development, *Complexity* 2018 (2018).
- [70] G. Castañeda, O. Guerrero, The Importance of Social and Government Learning in Ex Ante Policy Evaluation, *J. Policy Model.* (2019).
- [71] O. Guerrero, G. Castañeda, Quantifying the Coherence of Development Policy Priorities, *Dev. Policy Rev.* (2020). Forthcoming.
- [72] S. Cobey, E. Baskerville, Limits to Causal Inference with State-Space Reconstruction for Infectious Disease, *PLoS One* 11 (12) (2016), e0169050.
- [73] S. Cucurachi, S. Suh, Cause-effect analysis for sustainable development policy, *Environ. Rev.* 25 (3) (2017) 358–379.
- [74] O. Guerrero, G. Castañeda, Does Better Governance Guarantee Less Corruption? Evidence of Loss in Effectiveness of the Rule of Law (2019) arXiv preprint arXiv:1902.00428.
- [75] N. Meinshausen, A. Hauser, J. Mooij, J. Peters, P. Versteeg, P. Bühlmann, Methods for causal inference from gene perturbation experiments and validation, *Proc. Natl. Acad. Sci.* 113 (27) (2016) 7361–7368.
- [76] J. Pearl, Causal inference without counterfactuals: comment, *J. Am. Stat. Assoc.* 95 (450) (2000) 428–431.
- [77] J. Ramsey, M. Glymour, R. Sanchez-Romero, C. Glymour, A Million Variables and More: The Fast Greedy Equivalence Search Algorithm for Learning High-Dimensional Graphical Causal Models, with an Application to Functional Magnetic Resonance Images, *Int. J. Data Sci. Anal.* 3 (2) (2017) 121–129.
- [78] Y. Runber, C. Tomasi, L. Guibas, A metric for distributions with applications to image databases, *IEEE International Conference Computer Vision* (1998) 59–66.
- [79] M. Kalisch, M. Mächler, D. Colombo, M. Maathuis, P. Bühlmann, Causal Inference Using Graphical Models with the R Package pcalg, *Journal of Statistical Software* (2012).

Luis Ospina-Forero is an academic in Data Science in the Alliance Manchester Business school. Before his position at Manchester University, he was a data scientist at The Alan Turing Institute (London). His research interests relate to Network Inference, Network Comparison, and the application of Network Analysis methodologies to practical and meaningful problems in finance and society. Luis holds a Ph.D. in Statistics from the University of Oxford and is a statistician by training from the National University of Colombia.

Gonzalo Castañeda is currently a professor-researcher in the Economic Division at Centro de Investigación y Docencia Económica (CIDE), Mexico City. He has a Ph.D. in Economics from Cornell University. He is also a member of the National Research System, level III (i.e., the top rank granted by the National Council of Science and Technology, CONACYT). Prof. Castañeda works on Complex Adaptive System with issues related to economic development and just wrote a new book with the title “The Paradigm of Social Complexity: An Alternative Way of Understanding Societies and their economies.”

Omar A. Guerrero is an ESRC-Turing Fellow at the UCL Department of Economics and the Alan Turing Institute. He has a Ph.D. in Computational Social Science from George Mason University and studies problems related to economics and public policy. Previously, he was a fellow at the Oxford Martin School, the Saïd Business School and the Institute of New Economic Thinking at the University of Oxford. His work spans development economics, political economy, sustainable development, housing economics, firm and labor dynamics, and the improvement of policymaking through computational analysis.