BUILDING IMPROVED GENDER EQUALITY COMPOSITE INDICATORS BY OBJECT-ORIENTED BAYESIAN NETWORKS

Lorenzo Giammei [1], Flaminia Musella[2], Fulvia Mecatti[1] and Paola Vicard[3]

[1] Department of Sociology and Social Research, University of Milan-Bicocca, (e-mail: lorenzo.giammei@unimb.it, fulvia.mecatti@unimib.it)

[2] Department of Life Sciences, Health and Health Professions, Link Campus Unversity, (e-mail: f.musella@unilink.it)

[3] Department of Economics, Roma Tre University, (e-mail: paola.vicard@uniroma3.it)

ABSTRACT: This work proposes a novel methodology for constructing gender equality indicators using an Object-Oriented Bayesian Network (OOBN). The methodology is illustrated by focusing on the composite indicator known as Gender Equality Index, annually released by the European Institute of Gender Equality (EIGE). By using province-level ISTAT data, the index is re-constructed in a modern AI environment, able to enhance its information capacity and, at the same time, to preserve its original architecture. The modularity of the OOBN ensures a computational logic that is consistent with composite indicators, while also providing additional information about the relational structure of variables.

KEYWORDS: Object-Oriented Bayesian Networks, gender equality, composite indicator, regional indicator, sustainable development goals.

1 Introduction

Gender based inequalities represent a threat to socio-economic well-being on an individual level as well as for the society as a whole. Gender equality is one of the objectives pursued by the Sustainable Development Goals (SDG) of the UN Agenda 2030, as stated in the ambitious Goal 5: to achieve gender equality and empower all women and girls. In order to reach Goal 5, gender equality measurement plays a key role. The most widely used approach for measuring national gender equality is through a composite indicator. Composite indicators are useful in monitoring complex multidimensional phenomena, including gender equality, by providing a single value information, e.g. ranking of countries in their progresses toward SDG 5. However, composite indicators do not show the process of how a country has reached its own level of national gender equality nor they allow for monitoring or predicting the effect of policies and interventions. In this work we aim at empowering the role of gender inequality data analysis, by proposing a new method to measure the gender gap that can be used alongside composite indicators to obtain a richer set of information. In particular we employ an OOBN that follows the idea behind the computation of the European Union's Gender Equality Index (EU-GEI) but takes into account the multivariate dependence structure among all the variables generating a certain level of gender equality.

2 Object Oriented Bayesian Networks

A Bayesian network (BN) (Cowell, 1998; Pearl, 1998) is a probabilistic statistical model representing the joint distribution of a set of variables by means of a directed acyclic graph (DAG). In a DAG, nodes represent variables and edges denote the influence of one variable to another one. Bayesian networks possess a relevant and crucial property, named *modularity*, by which a possibly complex multivariate relation structure can be decomposed into smaller modules encoding conditional independencies. In a sense, BN can serve as basic building blocks for an extended tool called Object-Oriented Bayesian Networks (Koller & Pfeffer, 2013). An OOBN is a multi-instances network made of objects and special nodes. Objects are also called instance nodes representing simpler networks; the flow of information between networks is allowed by two kinds of interface nodes: input nodes, that import information from the OOBN into the instance; output nodes, that broadcast information from the instance to the OOBN. Since an instance node is a BN encapsulated in the OOBN, these models take advantage of the statistical properties of Bayesian networks. The inference process of OOBNs is efficient due to conditional independence between standard nodes in the instance and the OOBN, given the interface nodes. The architecture of OOBNs is particularly useful for managing large and complex domains, particularly when hierarchical structures are present. In the literature, there are many applications of OOBNs; the most relevant contribution for our purpose arises from the managerial framework (Musella & Vicard, 2015) where different quality aspects have been combined to provide a synthetic global quality indicator.

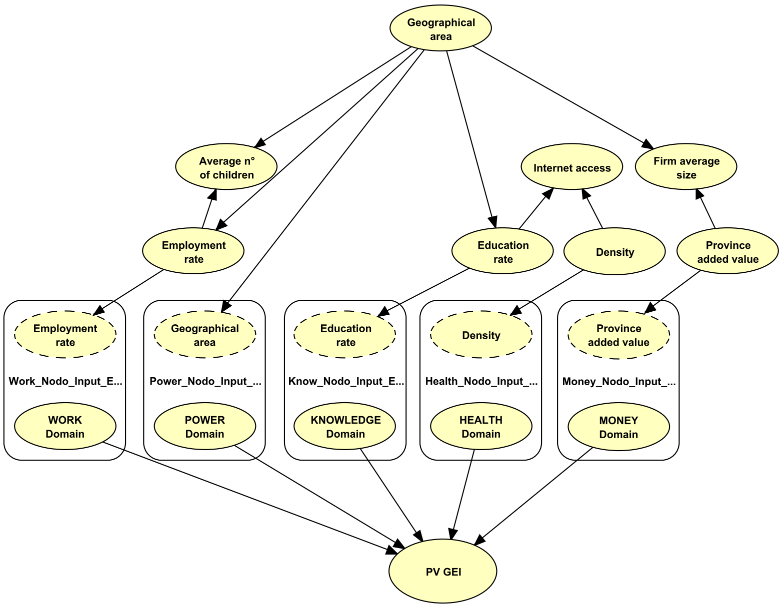

Figure 1. *Object-Oriented Bayesian Network model for the Italian province-level GEI*

3 Application to Real Data

In this work, we employ an OOBN to develop a gender equality indicator for Italian provinces based on the architecture of the EU-GEI. The EU-GEI, annualy released by the EIGE, is based on 6 domains: work, money, knowledge, time, power and health. We use province-level data from ISTAT to obtain a set of variables that is consistent with the one employed to compute the EU-GEI. Due to limited data availability, it is not possible to perfectly replicate the national GEI at the province level. To overcome the scarcity of gender-sensitive data at a fine granularity, proxy variables have been included if available. As a result all the EU-GEI domains are represented in the province-level GEI (PV-GEI), except for the Time domain. In addition, some extra socio-economic variables, such as *province added value* and *firm average size* are included to investigate their relationship with GEI ingredients. The resulting OOBN, obtained employing the statistical software Hugin, is depicted in Figure 1. Each box (rounded rectangle) in the OOBN represents an instance, which is a simpler network representing a specific PV-GEI domain that is linked to the whole OOBN. In each instance, the input node (represented as a node with a dashed outline in the figure) is selected from among the extra variables, while the output node is the summary value of the domain. According to the EU-GEI methodology, this value is the geometric mean of the sub-domain measure, which is in turn the arithmetic mean of the ingredient variables. PV-GEI ingredients are not visible in the OOBN at this level of representation because they are part of the sub-networks given by the domain instances. The different domains are then aggregated to compute the PV-GEI as a weighted geometric mean of domain measures. This architecture allows information to flow from extra socio-economic variables to PV-GEI ingredients, which in turn generates a certain level of the domain nodes and of the PV-GEI. The resulting model is consistent with the EU-GEI architecture and, at the same time, constitutes a powerful tool to simulate scenarios of how the PV-GEI changes when ingredient or socio-economic variables take different values.

4 Discussion and future research

The proposed methodology enriches composite indicators and provides a new perspective on the analysis of the gender gap. By employing an OOBN, we not only obtain a measure of gender equality that is consistent with that of composite indicators but also gain insight into the multivariate relationships between ingredients and other variables of interest. In addition, a refined granularity can be reached depending on the available data. These findings can support policy decision-making by shedding additional light on the complex net of factors that affect the gender (in)equalities. Finally, the estimated OOBN provides a simulation engine to predict the effect of policies and intervention aiming at reducing gender-based inequalities in the Country.

References

COWELL, R. 1998. Introduction to inference for Bayesian networks. *Learning in graphical models*, 9–26.

KOLLER, D, & PFEFFER, A. 2013. Object-oriented Bayesian networks. *arXiv preprint arXiv:1302.1554*.

MUSELLA, F, & VICARD, P. 2015. Object-oriented Bayesian networks for complex quality management problems. *Quality & Quantity*, **49**, 115–133.

PEARL, J. 1998. *Probabilistic reasoning in intelligence systems: networks of plausible inference*. Los Altos, CA: Morgan Kaufmann.