CHEMICAL ENGINEERING TRANSACTIONS

VOL. 70, 2018

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Timothy G. Walmsley, Petar S. Varbanov, Rongxin Su, Jiří J. Klemeš

ISBN 978-88-95608-67-9; **ISSN** 2283-9216

DOI: 10.3303/CET1870131

Automated Analysis of the Interactions Between Sustainable Development Goals Extracted from Models and Texts of Sustainability Science

Gyula Dörgő, Gergely Honti, János Abonyi*

MTA-PE "Lendület" Complex Systems Monitoring Research Group, University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary
janos@abonyilab.com

The design and monitoring of sustainable policies should rely on models that can handle complex and interconnected variables and subsystems of sustainability issues. Structuring knowledge has been identified as an essential first step in building models of sustainability science. Although it is known that all models yield a reduced view of the examined topic and no models can include all the variables that would make the representation closed and comprehensive, in the case of sustainability issues it is critical to synthesize as many critical aspects as possible that could have an impact on the studied problem. The key idea of our research is that strategic plans, sustainability reports and scientific studies reflect these variables, therefore, with the tools of text mining, the most important focus points and interactions can be determined. These key aspects and their connections can be represented by a network structure and compared to the subsystems of the dynamic models of sustainability to explore the deficiencies of the models or the lack of focus of the related policies and documentations. In the present work, the proposed methodology through the analysis of five strategical documents is demonstrated and the determined aspects with the structure of the famous World3 system dynamics model compared. The comparison highlighted the incomplete view of the original World3 model since certain topics were not critical issues whilst the World3 model was in development.

1. Introduction

With today's accelerating global economic, environmental and social change, the understanding of the undergoing processes and the forming of knowledge systems are of upmost importance in meeting the needs of current and future generations and therefore maintaining sustainable development. However, there is a clear consensus that scientific knowledge plays a key role in determining the aspects of unsustainability, and therefore in forming the related policies for promoting sustainable development, the development of such a knowledge base faces significant difficulty (Cash et al., 2003). One of the key problems is that, although nowadays several policy documents are being made (Chofreh et al., 2017), the formation of these policies has fallen behind the new emerging challenges and the needs of urgent knowledge are not sufficiently met (Cornell et al., 2013). The other main conflict that scientists must confront is that the models and theories that are used to generate their results relate to a number of their own assumptions, beliefs and values (Tàbara and Chabay, 2013).
Based on Forrester's revolutionary work in both system dynamics and sustainability (Forrester, 1971), the trends in sustainability have been interpreted as the results of dynamic processes for an extended period of time. One of the most well-known works is "The Limits to Growth" (Meadows et al., 1972), which not only focuses on the issues of sustainability but describes the deep conceptual understanding of the modelled system of complex interactions before being modelled quantitatively using stock-and-flow simulations. This work has undergone several improvements by scientists all over the world, e.g. the WorldWater model was introduced by Simonovic (2002) and the calibration of the original model was updated by Pasqualino et al. (2015) which was also improved by the authors themselves first in the "Beyond the Limits: Confronting Global Collapse" (Meadows et al., 1992) and later in the "Limits to Growth: The 30-Year Update" (Meadows et al., 2004). Even though dynamic simulations of aspects of sustainability can assist the work of policymakers and engineers, the principle that all

Please cite this article as: Dorgo G., Honti G., Abonyi J., 2018, Automated analysis of the interactions between sustainable development goals extracted from models and texts of sustainability science , Chemical Engineering Transactions, 70, 781-786 DOI:10.3303/CET1870131

models are as good as their underlying assumptions and simplifications must not be forgotten. Formerly Nabavi et al. (2017) discussed the opportunities and bad practices of the application of system dynamics in support of policymaking. In the present article by following this concept, it is assumed that system dynamics and policymaking should go hand in hand, meaning that policymakers should test their assumptions using the tools of system dynamics, while the key variables and subsystems of dynamic simulations should include the latest aspects of sustainability. A novel algorithm for the extraction of interactions from texts and dynamic models is presented and represented in the form of a network structure. This network representation can help to unfold the missing aspects of policies and the oversimplified or neglected subsystems of dynamic models.

Our key idea is that sustainability reports and scientific studies reflect the significant variables of sustainability, so with the tools of text mining the most important focus points and interactions can be determined. The proposed methodology is based on the analysis of multiplex co-word networks. The clusters of words are studied, and the viewpoints of documents compared based on the rank-based correlation of the centralities (PageRank) of the keywords and on node similarity analysis (Jaccard and SimRank-based cluster map and MDS). A similar approach can be applied to the analysis of system dynamics models. The variables of these models can also be investigated in terms of a network-based approach to unfold hidden interactions that are not directly presented in the causal loop diagram of the model. To realise this idea, a Python program was developed to extract networks from system dynamics models (Insight Maker XML files and Vensim SDM-Doc HTML extracts (Martinez-Moyano, 2012)). With the determination of the central nodes, the detection of the key subsystems and the comparison of the resultant structure with the network extracted from the related documents are desired.

The applicability of the developed tools and methodology is demonstrated by comparing the sustainability studies over previous decades to obtain an insight into how the key aspects change over time and what the persistent (problematic) fields of sustainability are. Five documents of sustainability are studied as case studies: Sustainable Development Goals of the United Nations: The 2030 Agenda for Sustainable Development (UN_SDG) (UN General Assembly, 2015), A Guide to SDG Interactions: from Science to Implementation (International Council for Science, 2017), The Limits to Growth (Meadows, 1972), Limits to Growth: The 30-Year Update (Meadows, 2004), and Pope Francis' encyclical on the environment and human ecology (Laudato Si') (Francis and McDonagh, 2016). Extracted key aspects of sustainability were compared with the structure of the World3 model (Meadows, 1972). The methodology exposes the incomplete view of the World3 model compared to the analysed documents as the topics of water, energy and climate were not critical issues whilst the World3 model was in development.

2. Methodology

A novel methodology to extract a network of key variables from system dynamics models and documents of sustainability is presented. The extracted networks can be compared to detect missing or misinterpreted areas of both modelling and policymaking.

2.1 From system dynamics models to a network of subsystems

The modelling of complex processes usually starts with causal loop diagrams to visualise how the variables in a system are interrelated. These diagrams are then transformed into stock-and-flow diagrams, which are the foundations of system dynamics modelling, consisting of stocks, flows, parameters and the relations between them represented by functions and/or converters. The main difference between stocks and flows is that stocks accumulate over time from flows, therefore, indicating conditions or states of the process, while flows represent activities and subsequently fill or drain stocks. According to another suggestive rule of thumb, stocks represent the nouns, while flows represent the verbs of the modelled process. A didactic illustration of the structure of a stock-and-flow diagram can be seen in Figure 1, together with Eq(1), which describes how this diagram represents a differential equation through the example of stock x_i:

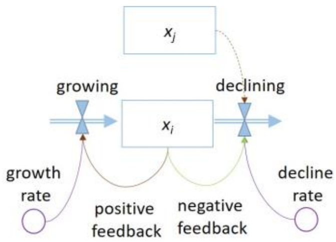
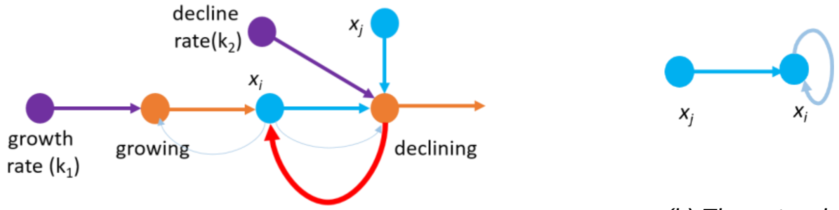


Figure 1: A simple example to illustrate how a stock-and-flow diagram represents differential equations.

$$x_i(t) = \int_0^t k_1 x_i(t) - k_2\big(x_i(t) - x_j(t)\big)dt \tag{1}$$

The system represented by a stock-and-flow diagram can be directly mapped into a network, where the nodes represent the parameters, flows and stocks, as illustrated by Figure 2(a). An important contribution of our work is drawing attention to the influence of the outlet flow on the stock itself, as the network has to be extended according to the internal logic of the dynamical system. This implemented extra connection is highlighted in red in Figure 2(a). The network of state variables (stocks) can be extracted from the result of the reachability analysis. This simplified network provides a good opportunity to evaluate the controllability and observability of the system. The network of state variables (i.e. stocks) can be seen in Figure 2(b). The colours indicate the type of the node: purple represents the parameters, orange stands for the flows, while blue illustrates the stocks of the process.



(a) Interactions between the elements of the stock-and-flow diagram can be directly mapped into a network.

(b) The network of the state variables.

Figure 2: An illustration of the transition from stock-and-flow diagrams to a network of stocks, flows and variables is presented in part (a). The causal relationship of the state variables is highlighted by the reachability analysis, as presented in part (b).

This network-based representation provides the opportunity to calculate the modularity of the nodes of the network. Modularity is the fraction of the edges that fall within the given groups of nodes minus the expected fraction if the edges follow a random distribution. The detected modules can be evaluated by the means of possible subsystems, which should reflect the key aspects of the modelled process and can be used to determine the closeness of each module, which is an important measure of controllability and observability. The determined subsystems represent the cognitive map of the modelled system.

2.2 From texts to the network of the concepts of sustainability

With regard to the processing of the original text documents, the KNIME Analytics Platform software was applied due to its easy-to-use graphical user interface and the high number of implemented machine learning and data mining algorithms. After the parsing of each document, only the numbers, nouns and verbs were kept for further analysis based on the POS tag of each word and all words longer than three characters were converted to lower case. By applying the bag-of-words model, the frequency of occurrence of each word and word pair was determined. The frequency of occurrence of each word reflects the key topics of the given document, while the co-occurrence frequency of these words provides the opportunity for a network-type analysis of the documents. In the case of smaller networks, even the multilayer visualisation of the networks can result in interesting recognitions, but in the case of longer documents, other, more illustrative measures should be defined to identify the connections between the different contents. In the present work, the similarity of these networks was measured based on the rank correlation of the centralities of the keywords.

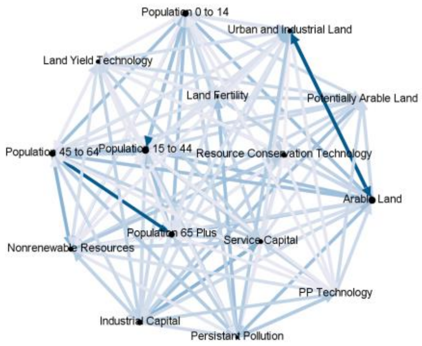
3. Results

To test the proposed network-based analysis method, a Python program to extract networks from system dynamics models (Insight Maker XML files and Vensim SDM-Doc HTML extracts (Martinez-Moyano, 2012)) and documents of sustainability were developed. Using the developed algorithms, the applicability of this methodology through the analysis of the famous World3 model and five important strategic plans from previous decades was demonstrated.

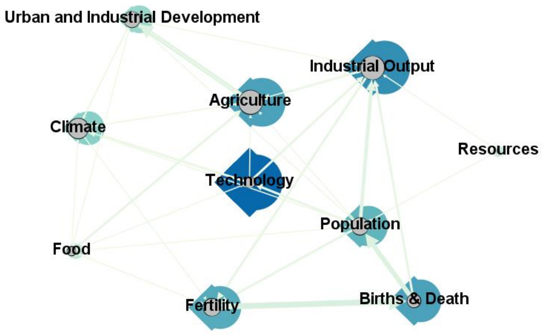
3.1 Analysis of the World3 model

The World3 model is a system dynamics model originally documented and applied in the book ‘The Limits to Growth’ (Meadows, 1972) for the simulation of the global population, industrial growth, food production and

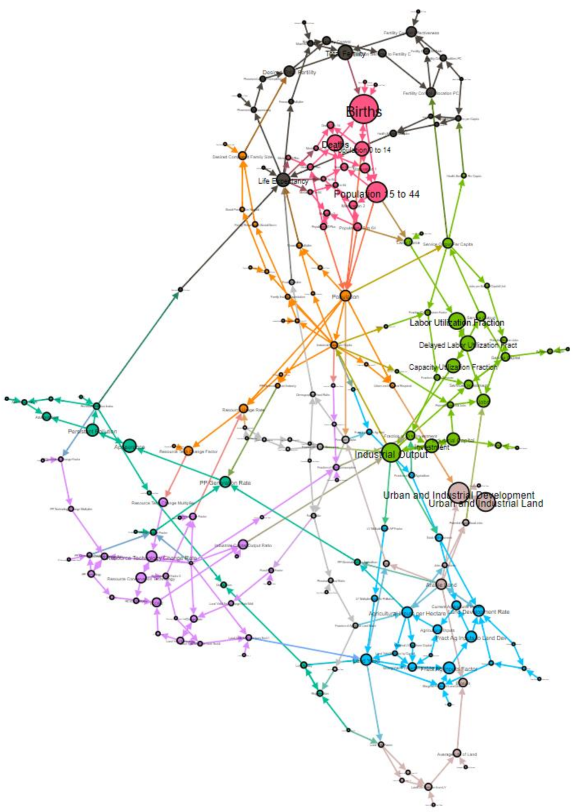
limits in the ecosystem of the earth. The main systems in the original model were that of food (dealing with agriculture and food production), industry, population, non-renewable resources and pollution.
The state variables, i.e. the stocks of the original stock-and-flow diagram, can be characterised based on their closeness to each other, which is an important measure of the ease of controlling or observing these nodes. In Figure 3(a) the colour of the edge represents the closeness of the given variables, therefore, the darker the edge, the closer the given nodes are to each other.
The key areas of the modelled processes are reflected in the communities of the network of nodes extracted from the World3 model, as is presented in Figure 3(b). The colours represent the modules of the model, reflecting the key areas of the Club of Rome study published in the book, The Limits to Growth (1972). By illustrating the connections between these subsystems in a similar way to that shown in Figure 3(c), the cognitive map of the system dynamics model can be seen:



(a) The colour of the edges illustrates the closeness of the state variables (the darker the edge, the closer the nodes are to each other).



(c) The cognitive map of the World3 model based on the detected communities.

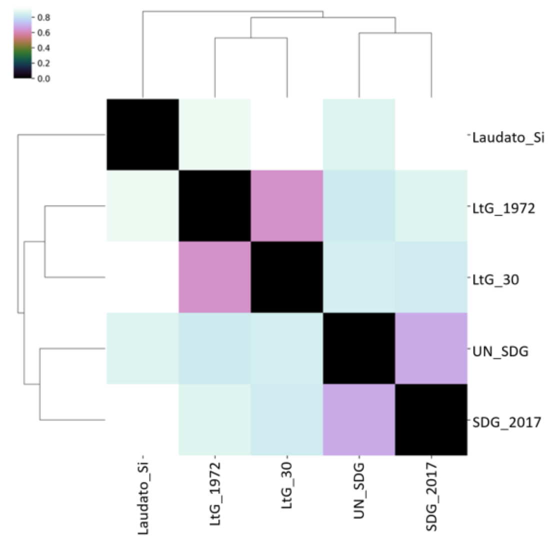


(b) The network extracted from the World3 model.

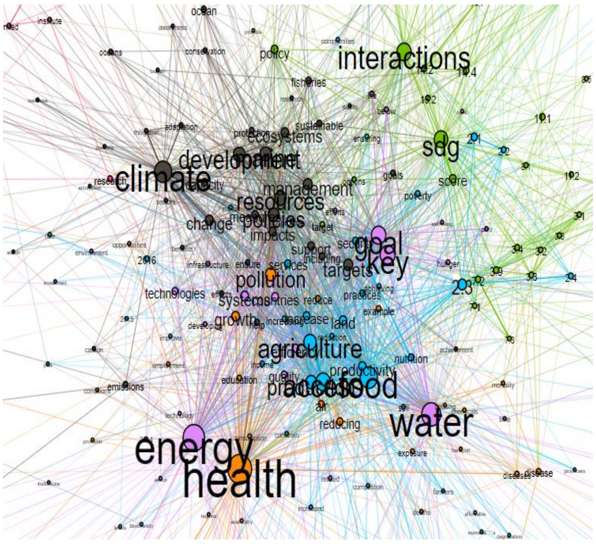
Figure 3: (a) The closeness of state variables and (b) the results of community detection in the World3 model. (c) The key areas of the subject process of modelling can be determined and illustrated with a cognitive map.

3.2 From texts to the key concepts of sustainability

By comparing the sustainability studies of previous decades, an insight can be made into how the key aspects of sustainability change over time and what the persistent (problematic) fields are. The similarities of the documents were evaluated based on the rank correlation of the centralities of the keywords. According to the results presented in Figure 4(a), it can be stated that Pope Francis holds a unique view of the problem of sustainability. To find what the abbreviations stand for in full see the appropriate paragraph in the Introduction section. The most important words in the Sustainable Development Goals of the United Nations (UN_SDG) (UN General Assembly, 2015) and Pope Francis' encyclical on the environment and human ecology (Laudato Si) (Francis and McDonagh, 2016) were compared. According to the most frequent words presented in Table 1, it can be observed that the topics reflected by these words are fundamentally different. It is believed that the omission of important social variables highlighted in the Laudato Si' (Francis and McDonagh, 2016) is a serious defect of the other studied documents.



(a) The similarities of the documents based on the rank correlation of the centralities of the keywords. The darker the colour, the stronger the similarity is.



(b) The network extracted from SDG_2017 illustrates how the goals (represented by numbers) are interconnected and related to the extracted problems of sustainability.

Figure 4: (a) The heatmap of similarity measures of documents and (b) the interconnection of sustainability goals in "A guide to SDG interactions: from science to implementation" (SDG_2017).

Table 1: The most important words in the UN_SDG and Laudato Si' documents.

Laudato Si	UN_SDG
Human	Development
World	Sustainable
Development	Country
Life	Nature
Social	Global
Nature	Financial

The network-based representation of long and complex documents with multiple topics or aspects can also assist the visualisation of interactions between topics. Figure 4(b) shows the network extracted from "A guide to SDG interactions: from science to implementation" (SDG_2017), which illustrates how the goals (represented by the same numbers in the document and Figure 4(b)) are interconnected and related to the extracted problems of sustainability.

4. Conclusions

The design and monitoring of sustainable policies and the analysis of system dynamics models of sustainability should go hand in hand to be able to handle the constantly changing system of complex and interconnected variables. In the present article, a novel network theory-based methodology was proposed, which helps the structuring of knowledge for policymakers and engineers. A python program to extract networks from system dynamics models and documents and compare these networks to determine deficiencies and misinterpretations was developed. Node centralities were determined and the communities in the networks extracted from the World3 model identified. The detected communities can be applied to define thematic subsystems and form cognitive maps. The similarities of the documents (and models) were evaluated based on the rank correlation of the centralities of the keywords. The proposed methodology is fast and effective in the comparison of large number of documents and system dynamics models giving an overview of the discussed topics. The comparison of the documents highlights that Pope Francis holds a unique view of the problem.

The network extracted from “A Guide to SDG Interactions: from Science to Implementation” illustrates how the goals are interconnected and related to the extracted problems of sustainability. Pope Francis' Laudato Si highlights the omission of important social variables from the UN_SDG and the studied models, which is a serious flaw of these studies. The comparison of the network of documents with the extracted network of the studied World3 model exposes the incomplete view of the problem as the topics of water, energy and climate were not critical issues whilst the model was in development. The results illustrate that a network-based representation can help in the development of new system dynamics models and can also be useful in the implementation of new and so far, ignored topics and effects into the existing models. Moreover, this representation also provides the opportunity to compare the layer of texts and models, highlighting the similarities and differences between the documents and simulators.

Acknowledgments

This research has been supported by the National Research, Development and Innovation Office NKFIH, through the project OTKA-116674 (Process mining and deep learning in the natural sciences and process development) and the EFOP-3.6.1- 16-2016- 00015 Smart Specialization Strategy (S3) Comprehensive Institutional Development Program. Gyula Dörgő was supported by the ÚNKP-17-3 New National Excellence Program of the Ministry of Human Capacities.

References

Cash D.W., Clark C.C., Alcock F., Dickson N.M., Eckley N., Guston D.H., Jäger J., Mitchell R.B., 2003, Knowledge systems for sustainable development, PNAS, 100, 8086–8091.

Chofreh A.G., Goni F.A., Klemeš J.J., 2017, Development of a framework for the implementation of sustainable enterprise resource planning, Chemical Engineering Transactions, 61, 1543-1548.

Cornell S., Berkhout F., Tuinstra W., Tàbara J. D., Jäger J., Chabay I., de Wit B., Langlais R., Mills D., Moll P., Otto I.M., Petersen A., Pohl C., van Kerkhoff L., 2013, Opening up knowledge systems for better responses to global environmental change, Environmental Science & Policy, 28, 60–70.

Forrester J.W., 1971, World dynamics. Wright-Allen Press, Cambridge, Mass., USA.

Francis P., McDonagh S., 2016, On care for our common home: laudato si’ - the encyclical of pope francis on the environment. Orbis Books (Ecology and Justice), Maryknoll, New York, USA.

International Council for Science, 2017, A guide to sdg interactions: from science to implementation. international council for science, Paris, France, DOI: 10.24948/2017.01.

Martinez-Moyano I.J., 2012, Documentation for model transparency, System Dynamics Review, 28, 199–208,.

Meadows D.H., Meadows D.L., Randers J., Behrens W., 1972, The Limits to growth: A report for the Club of Rome's project on the predicament of mankind. Universe Books, New York, New York.

Meadows D.H., Randers J., Meadows D.L., 2004, Limits to growth: the 30-year update. Chelsea Green Publishing, London, UK.

Meadows D.H., Meadows D.L., Randers J., 1992, Beyond the limits: confronting global collapse, envisioning a sustainable future. Chelsea Green Publishing, London, UK.

Nabavi E., Daniell K.A., Najafi H., 2017, Boundary matters: the potential of system dynamics to support sustainability? Journal of Cleaner Production, 140, 312–323, DOI: 10.1016/J.JCLEPRO.2016.03.032.

Pasqualino R., Jones A. W., Monasterolo I., Phillips A., 2015, Understanding global systems today—a calibration of the world3-03 model between 1995 and 2012, Sustainability, 7, 9864-9889.

Simonovic S.P., 2002, World water dynamics: global modeling of water resources, Journal of Environmental Management, 66, 249–267.

Tàbara J.D., Chabay I., 2013, Coupling human information and knowledge systems with social–ecological systems change: reframing research, education, and policy for sustainability, Environmental Science & Policy, 28, 71–81.

UN General Assembly, 2015, Transforming our world: the 2030 Agenda for Sustainable Development <www.refworld.org/docid/57b6e3e44.html> accessed 27.11.2017

Zwillinger D., 1997, Handbook of differential equations. Academic Press, Cambridge, Mass., USA.