Controllability of complex networks for sustainable system dynamics

Pahola T. Benavides$^{1,2}$ and Urmila Diwekar$^{1,2}$

$^1$Department of Bio Engineering, University of Illinois, Chicago, IL 60607, USA
$^2$Center for Uncertain Systems: Tools for Optimization & Management (CUSTOM), Vishwamitra Research Institute, Crystal Lake, IL 60012, USA

Corresponding author. Email: urmila@vri-custom.org

AND

Heriberto Cabezas

Department of Computer Science and Systems Technology, University of Pannonia, Veszprém, Hungary

Edited by: Allesina Stefano

[Received on 8 July 2014; accepted on 18 December 2014]

Successful implementation of sustainability ideas in ecosystem management requires a basic understanding of the often non-linear and non-intuitive relationships among different dimensions of sustainability, particularly the system-wide implications of human actions. This basic understanding further includes a sense of the time scale of possible future events and the limits of what is and is not likely to be possible. With this understanding, systematic approaches based on control theory can then be used to develop policy guidelines for the system. Therefore, controllability of the system is very important to determining long term sustainability of the system. An article Liu et al. (2011, Nature, 473, 167) presents a new analytical approach to study the controllability of complex systems. We apply this approach to three dynamic systems developed to study the sustainability of our planet. These three systems consist of an ecosystem based on wild and domesticated compartments: a simple economic model, an ecosystem model with an industrial system, and an integrated model involving the ecosystem, industrial systems, and energy producers. The goal is not to develop detailed predictions, but to explore the feasibility of general strategies for sustainability, and to establish scientific criteria which would indicate their likelihood of being successful. We argue that controllability of this system is necessary for its long-term sustainability, and we present our arguments in the light of previous studies of these systems.

Keywords: sustainability, controllability, complex networks, ecological systems.

Nomenclature

- $\beta$: Inverse of temperature
- $C$: Carnivores
- EP: Energy producer
- ES: Energy source
- $E_G(\beta)$: Internal energy function
- $H_i$: Herbivores
- HH: Humans households

© The authors 2015. Published by Oxford University Press. All rights reserved.
1. Introduction

The intensification of human activities resulting from continuous technological, economic and social evolution has severely depleted and deteriorated much of the earth’s natural resources. This situation has challenged the scientific, political and social communities to explore alternatives that allow them to understand and manage the effects of development. As a result, sustainability has emerged as a new subject of analysis and evaluation. In the Brundtland Commission report [2], sustainability was defined as development that meets the needs of the present without compromising the ability of future generations to meet their own needs. The sustainability of a system is determined by the interactions between the various dimensions of a system represented such as ecology, human society, economics, technology and many other disciplines. Very often, these interactions are non-linear and intertwined, so that their effects are not easily perceived [3]. The goal of sustainability, then, is to promote policies and solutions that improve the ecology, human, economics component of the system, such that long-term human existence on the Earth becomes possible. In other words, sustainability aims to enhance human well-being while protecting the planet. However, the successful implementation of sustainability strategies (e.g. in ecosystem management) requires a basic understanding of the natural, social and technological systems and the relationships of their related components. Only through such understanding we can manage those complex systems. Therefore, with sustainability emerging as one of the critical themes in policy and decision-making, the analysis of complex systems is essential.

In the literature, it can be found that real-world complex systems have been a subject of network studies [4–8]. These complex systems typically have a large number of components which may act
according to dynamic rules that change over time. When the complex systems are represented as networks, the nodes are the different elements of systems and the links are the set connections that represent the interaction among these elements. Examples of real-work systems are social networks [6,9] where nodes are persons and their relationships (i.e. friendship, family or work-related) are represented by links. In nature, ecological systems are also a good example of complex systems. In this case, nodes are represented by animals, plants, humans etc., and links represent the interactions such as predator–prey. These models are commonly found as a food web networks or ecological networks [8,10–13]. The study of complex system as networks is important because: (i) there are interesting features of complex systems that can be captured by simpler networks and (ii) there are families of networks that share certain properties that can be quantified with simple statistics. Therefore, network analysis can help us to understand how elements of a particular network interact with each other and how connections are made between them.

Mathematical models have been used to understand the behaviour and dynamics of complex networks such as ecological systems. For instance, (4) ecologists have used mathematical models to determine the stability of a community based on the size and connectivity of the food web in the face of fluctuations of the system (i.e. invasion of new species, population dynamics etc.). Lotka–Volterra models used to describe the predator–prey interactions (9) represent one classical theory used to describe the dynamics of these biological systems. These equations represent interactions between the predator and prey by first order, non-linear, differential equations. In addition, there have been ecological models which have been incorporated varying degrees of socio-economic features. For example, Acutt and Mason [10] provided important issues in the field of environmental economics like the valuation of paying for environmental preservation or application of environmental taxes, and Ludwig et al. [15] optimized the phosphorous loading for a potentially eutrophic lake. They found a strong interaction between economic and ecological parameters in determining the optimal policies and proposed their results to illustrate economic ideas such as sustainable development. Brock and Xepapadeas [16] analysed the equilibrium ecosystem resulted from Nature’s equilibrium and two management problems (i.e. privately and socially), characterized them as ecological/economic model with linear structures and provided policy rules so that the privately optimal state could be driven through the socially optimal or natural equilibrium. Therefore, mathematical models of real ecosystems are very important tools and they can also aid in the study of sustainability. In this paper, we study three ecosystem models—considered as complex systems—that mimic some of the dynamic behaviour of the earth using networks. The first model is a simple version of the earth ecosystem that presents very rudimentary social and ecological interactions [17]. The two remaining models are enhanced versions of the first model since they integrate social and ecological interactions with economic interactions, and biofuels as an energy source [18,19]. With these three different socio-ecological-economical systems, we study their controllability which is closely linked to sustainability.

According to control theory, controllability is the ability to guide a system’s behaviour from any initial state to any desired final state through the appropriate manipulation of a suitable choice of inputs [1]. Therefore, if we lead a system to a desired state under certain conditions, then, the system is controllable. Moreover, if we lead the system through a controllable path, where the successions of states favour human existence over the long term, we are leading the system through a sustainable path. Two important characteristics are involved in this ability: the system’s architecture which is described by the network representation and the dynamic rules that capture the time-dependent interactions. Using these two characteristics of the system, we can study controllability using an approach proposed by Liu et al. [1] and determine whether the system can in principle be controlled to make it sustainable. This is important because in principle any environmental system can be represented as a
network, and any management strategy or policy for sustainability is at its core an effort to control the system.

This paper is arranged in the following order: Section 1 introduces the three dynamic models for studying sustainability which are the simple, intermediate and integrated model. Section 2 presents the network representation for these models, while Section 3 describes the methodology of controllability and the application to sustainability. Section 4 discusses the results. Finally, Section 5 summarizes the paper.

2. Dynamic models for sustainability

In this paper, three dynamic models used for sustainability are presented. These models describe the earth as a complex dynamic network that mimic a general ecosystem with very elementary social, ecological and economic interactions. The interactions are regulated through flow of mass of biological resources such as biomass, nutrients, water etc. The first model is a simple version of the ecological system studied in [17]. The second model has modifications of the first model that involve economic aspects through a price-setting model [18], and the third model further includes energy sources [19]. These models are close to mass, meaning that the accumulative sum of mass in all system compartments is constant. This useful assumption is important for the study of sustainability because the system must function with finite material resources [18]. On the other hand, the mass balance is formulated through non-linear differential equations that describe the rate of change in the mass of various components over time based on Lotka–Volterra-type expressions. These equations represent the biological growth and death due to interactions between the compartments. The model structure is divided into two characteristic branches: domestic and non-domestic. The domestic branch represents the agricultural and livestock activities which are private property and have economic value. The non-domestic branch represents species that are hunted or gathered, and it has no property rights and no economic value. In other words, the domestic branch represents species that humans can own, buy, and sell to be consumed or used, and the non-domestic branch represents species that are owned, bought, or sold by humans.

2.1 Model 1: The simple model

Figure 1 depicts the first ecosystem model. As it can be seen, there are a total of 12 compartments that compromise: four plants \( P_1, P_2, P_3 \) and \( P_4 \) two of which can represent cultivated crops such as corn and soybean and the others any species of plants; three groups of herbivore animals \( H_1, H_2 \) and \( H_3 \) from mammals, birds and reptiles one of which is domesticated; two groups of carnivore animals \( C_1 \) and \( C_2 \); a human household \( \text{HH} \); a resource pool \( \text{RP} \) and inaccessible resource pool \( \text{IRP} \). The resource pool represents all biological resources needed by plants (e.g. water, nitrogen, carbon dioxide etc.). Primary producers or plants \( P_1 \ldots P_4 \) make available mass and energy from the accessible resource pool to the rest of the food web. The inaccessible resource pool corresponds to biologically unavailable resources that result from human activity (e.g. polluted water which is biologically unusable unless it gets purified by slow biological process). The arrows represent the mass flows from one compartment (i.e. origination) to another compartment (i.e. termination), and all living compartments have an implied mass flow back to the resource pool that represents death. In this figure, there are three types of arrows: the solid arrows describe transfer of mass due to biological or geological drivers (e.g. species of \( P_1 \) are consumed by species of \( H_1 \)), the dashed arrows represent the transfer of mass from the resource pool to the inaccessible pool that occurs as a byproduct of human activities necessary to
support agriculture and product distribution (e.g. water pollution, paving etc.), and the dotted arrows are the human influenced mass flows. The dashed arrow connecting \( RP \) to \( IRP \) represents mass which becomes biologically unavailable as a by-product of activities associated with \( P_1, P_2 \) and \( H_1 \). The dashed arrow connecting \( HH \) to \( IRP \) represents mass made biologically unavailable by human consumption. The primary producers feed on \( RP \) and use solar energy to make this mass available to the rest of the system. A small amount of mass from \( IRP \) is recycled back into the system by \( P_3 \) and \( P_4 \) which symbolizes degradation by the actions of microorganisms. All nine biological compartments recycle mass back to \( RP \) through death. The details of the mathematical model can be found in [17].

2.2 Model 2: The intermediate model

The intermediate model is an extended version of the first model since it includes the industrial sector (IS). This new sector represents—at a very elementary level—generic human industrial activity that offers a benefit to the human population. As it is shown in Fig. 2, this model also contains the same 12 compartment shown in Fig. 1 that track the flows of mass. However, \( IS \) is not considered as a compartment since no mass is resident. The industrial sector simply takes mass from compartments: \( P_1, RP \) and \( H_3 \) and combines it to form a product. This plays an important role in the economic model of this system. Therefore, the intermediate model consists of the ecosystem dynamic model and a very simple economic model. For details, please refer to [18]. In this model, the solid arrows represent transfers of mass due solely to biological or geological drivers, with no interference from humans. The dashed arrows represents a transfer of mass from the resource pool to the inaccessible resource pool.
that occurs as a by-product of human activities as before, including consumption ($HH \rightarrow IRP$) and production ($RP \rightarrow IRP$ and $IS \rightarrow IRP$). The mass flows represented by the dashed arrows are under some level of human control.

2.3 Model 3: The integrated ecological and economic model

The third model is a further enhanced version of the two previous models minus one plant species. In this case, the model incorporates some crucial and representative elements of the real world social, economic, technological and biological system. This models is based on the model presented by Whitmore et al. [20], however, despite of the unique features of Whitmore’s model, it has a very limiting assumption as it presumed that an infinitive amount of energy was available without any cost to the industrial sector and various components of the ecological system. This assumption does not reflect the real world since factors related to energy not only have a cost but geopolitical ramification. They also causes enormous stress on certain components of the ecosystem. Therefore, this new version of the model, although still a simplified version of the any real system, considers one or more aspects related to the production and utilization of energy. The production of energy considers various types of energy sources. As shown on Fig. 3, this modified model contains three plants ($P1, P2$ and $P3$), three herbivores ($H1, H2$ and $H3$), two carnivores ($C1$ and $C2$), a human households ($HH$), an industrial sector ($IS$), energy source ($ES$), energy producer ($EP$), the resource pool ($RP$) and inaccessible resource pool ($IRP$) resulting in 14 compartments. The compartments $HH, IS, EP, P1$ and $H1$ describe the economic perspective of this model. The price-setting macroeconomic model determines flows of economic goods.
and labour that govern the dynamic (decisions) of these five compartments. This price-setting model is based on four firms (IS, EP, P1 and H1) and the fact that HH attempts to maximize humans well-being. For instance, the energy source represents a finite non-renewable resource. The energy producer is a firm that uses labour to transform mass (ES) into a usable form of energy. This energy is supplied to HH and IS. EP is also capable of producing energy using P1, and this would represent the production of energy using biomass. For example, sugar cane is used to produce bio-ethanol or soybean oil is used to produce biodiesel. The IS produces products valuable to HH using P1 and RP. The use of the IS products does not increase the mass of HH, but it instead passes through and is discarded to increase the mass of IRP. Similarly, the use of mass by the EP to produce energy results in waste that increases the mass of IRP, and a corresponding decrease in the mass of ES. For details, please refer to [19].

As shown in Fig. 3, the dashed lines indicate mass flows that occur under anthropogenic influence. The dotted lines indicate the flow of energy from EP to HH and IS. The square dotted lines between P2, P3 and IRP indicate slow transfers of mass as a result of microbial activity. The formulation was presented in [19,21]. For this model, it is assumed that 30% of the total energy demand by the integrated system is being provided by the biomass. If a sufficient amount of biomass is not available, the maximum available biomass is used for the production of energy and the remaining energy is produced from the non-renewable energy source. Other important parameters involved in the formulation of this model are the demands for goods and labour: the wages rate is set by the industrial sector, and demands of various products are set by human households. Note that there is a cost to IS and EP for the generation of biologically unavailable mass which is discharged into IRP.

Table 1 shows a summary of the state variables (i.e. \( y_i \)) involved in the mathematical model of the three socio-ecological-economical systems presented before. Each of these variables represents a...
Table 1  Summary of states variables for each model

<table>
<thead>
<tr>
<th>Description</th>
<th>Model 1: simple model</th>
<th>Model 2: intermediate model</th>
<th>Model 3: integrated economic-ecological model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary producers</td>
<td>$y_{P1}$, $y_{P2}$, $y_{P3}$, $y_{P4}$</td>
<td>$y_{P1}$, $y_{P2}$, $y_{P3}$, $y_{P4}$</td>
<td>$y_{P1}$, $y_{P2}$, $y_{P3}$</td>
</tr>
<tr>
<td>Herbivorous Animals</td>
<td>$y_{H1}$, $y_{H2}$, $y_{H3}$</td>
<td>$y_{H1}$, $y_{H2}$, $y_{H3}$</td>
<td>$y_{H1}$, $y_{H2}$, $y_{H3}$</td>
</tr>
<tr>
<td>Carnivorous Animals</td>
<td>$y_{C1}$, $y_{C2}$</td>
<td>$y_{C1}$, $y_{C2}$</td>
<td>$y_{C1}$, $y_{C2}$</td>
</tr>
<tr>
<td>Humans Households</td>
<td>$y_{HH}$</td>
<td>$y_{HH}$</td>
<td>$y_{HH}$</td>
</tr>
<tr>
<td>Industries</td>
<td>-</td>
<td>-</td>
<td>$y_{IS}$, $y_{EP}$</td>
</tr>
<tr>
<td>Natural resources</td>
<td>$y_{RP}$</td>
<td>$y_{RP}$</td>
<td>$y_{RP}$</td>
</tr>
<tr>
<td>Inaccessible resources</td>
<td>$y_{IRP}$</td>
<td>$y_{IRP}$</td>
<td>$y_{IRP}$</td>
</tr>
<tr>
<td>Deficits</td>
<td>-</td>
<td>-</td>
<td>$y_{P1H1d}$, $y_{P1ISd}$, $y_{P1HHd}$, $y_{H1HH}$, $y_{ISHHd}$, $y_{NHH}$</td>
</tr>
<tr>
<td>Human population</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

different node of the network. For instance, $y_{P1}$ represents the state variable for plant 1 or node $P1$ while $y_{H1}$ is the node for compartment $H1$. As it can be observed, Model 3 has eight more nodes as a result of the integration of the price setting economic model: $y_{P1H1d}$, $y_{P1ISd}$, $y_{P1HHd}$, $y_{H1HH}$ and $y_{ISHHd}$. These represent the states variables involved with the mass flows [19,21].

3. Network representations

The increase of available data of biological, socio-ecological and economical systems, and the advance in computational tools giving us wide access to these data has placed a new focus on the complex networks (4). This situation has extended the study of network theory to a wide range of systems including the internet, power-grid and ecological systems. As mentioned before, a network is a collection of points joined together in pairs by lines (6) which uses a graph as an abstract mathematical representation [22]. The conception of network theory started with the solution of the Knigsberg bridge puzzle presented by [23]. This work was of remarkable importance since it identified the topology (i.e. the arrangement of links and nodes) as the key issue of the problem [24]. Therefore, we can say that the relationship between the number of links, nodes and their collective behaviour play an important role in understanding networks. Some interesting features of networks can be used for this purpose are, for instance, degree of a node and degree distributions. The degree of a node ($k$) is the number of edges that are connected to that node. On the other hand, the degree distribution is a statistical property that expresses the probability $P(k)$ of nodes in the network having degree $k$. This property depends on the type of network, that is, whether the network is directed or undirected. In a directed network, links go in only one direction while in an undirected network links go in both directions (1). Analytical tools [1] developed to study the controllability of an arbitrary complex network were developed for directed networks using the degree distribution for calculation of controllability.

3.1 Model visualization

Model visualization is a crucial part of network analysis. It allows us to have a general idea of the structure of the model and its architecture represented through a graph. For this purpose, we used the discrete network dynamic tool found in the Network workbench NWB software [25]. The NWB is a large-scale network analysis, modelling and visualization toolkit used in several areas of scientific
research. Using this software, we created the state space graph of the three models based on the dynamic interactions of each element (i.e., states variable) in the network. The representation scheme used here was the radial tree visualization which is shown in Figs. 4–6 for each model, respectively. These figures show that the three networks are directed networks. In the NWB tool, we used the Network Analysis Toolkit (NAT) to obtain basic attributes of the networks such as the number of nodes, the number edges, average mean degree etc. These attributes are used to compute the minimum number of nodes or driver nodes that will allow us to control the system. Table 2 summarizes these attributes for the three models.

3.2 Degree distribution

This property describes the distribution of node connections in the network. The degree distribution is the result of a histogram of degree of nodes generated by calculating the degree \( k \) at each node. Since the networks studied in this paper are directed networks, the degree distribution is classified as in-degree and out-degree distributions. This information is given by Network workbench NWB software. The in-degree and out-degree distributions for Model 1–3 are presented in the following figures.
The degree distribution for model 1 is shown in Fig. 7. As it can be seen, the in-degree distribution shows that \( \sim 30-50\% \) of the nodes, respectively, have 5 and 4 ingoing degrees to the target nodes. On the other hand, the out-degree plot looks more distributed, that is, 42\% have 5 outgoing links to the targets nodes while 33\% are disperse between 4 and 6, and only one node has 9 outgoing links to the target nodes.

Figure 8 presents the degree distribution for the intermediate model. In this case, the in-degree distribution shows that most of the nodes (i.e. 58 and 25\%) have degree of 4 and 5 while only 8\% of the nodes have higher degrees (i.e. 12). On the other hand, the out-degree distribution shows that 66\% of nodes have degree of 4 and 5 while only 8\% of the nodes higher degree (i.e. 10).

For model 3, both in-degree and out-degree distributions look sparser compared with Models 1 and 2 which means that the degrees of nodes have more variability. This is also described by degree of heterogeneity \((H)\) which is a measure of statistical dispersion that represents the spread between the less connected and the more connected nodes [1]. As shown in Fig. 9, in the in-degree distribution, the nodes with few links (low \(k\)) are less common than the nodes with large numbers of links, that is, 32\% of the nodes have 10 incoming connections to the target nodes. In the case of out-degree, the heterogeneity is more appreciable since 26, 16 and 5\% of the nodes have 4, 5 and 6 outgoing links to target nodes, respectively, while 5, 32, 5 and 8\% have 10, 11, 12 and 13 outgoing links, respectively.
Fig. 6. Network visualization for Model 3.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Model 1: simple model</th>
<th>Model 2: intermediate model</th>
<th>Model 3: integrated economic-ecological model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes ($N$)</td>
<td>12</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>Number of links ($L$)</td>
<td>59</td>
<td>61</td>
<td>155</td>
</tr>
<tr>
<td>Mean degree</td>
<td>9.83</td>
<td>10.16</td>
<td>16.32</td>
</tr>
<tr>
<td>Average in-degree ($k_{in}$)</td>
<td>4.92</td>
<td>5.08</td>
<td>8.16</td>
</tr>
<tr>
<td>Average out-degree ($k_{out}$)</td>
<td>4.92</td>
<td>5.08</td>
<td>8.16</td>
</tr>
</tbody>
</table>

3.3 Degree of heterogeneity

The measure and analysis of heterogeneity is important in order to understand the behaviour of complex networks. In fact, complex systems are characterized by having large values of the degree of heterogeneity ($H$). The degree of heterogeneity can be calculated by Equation (1) [1]:

$$H = \frac{\Delta}{\langle k \rangle}$$

(1)
Fig. 7. Degree distributions of the simple ecological Model 1.

Fig. 8. Degree distributions of the intermediate Model 2.

Fig. 9. Degree distributions of the integrated economic and ecological Model 3.
where \( \langle k \rangle \) is the mean degree (Table 2) and \( \Delta \) is the average absolute degree difference of all pairs of nodes \( (i \) and \( j) \) drawn from the respective probability degree distribution \( P(k) \). \( \Delta \) is calculated by the following equation:

\[
\Delta = \sum_i \sum_j |k_i - k_j| P(k_i) P(k_j)
\]  

\( \Delta \) is calculated by the following equation:

\[
\Delta = \sum_i \sum_j |k_i - k_j| P(k_i) P(k_j)
\]  

\( H \) can take values from 0 to 2. Table 3 presents the degree of heterogeneity for the three models studied here. Since we have in-degree and out-degree distribution, we computed the degree of heterogeneity for each distribution. As it can be seen, the degree of heterogeneity is higher in Model 3 since their degree distributions look sparser.

### 4. Controllability of systems with cavity method

The structural controllability criteria defined by Liu et al. [1] determines the driver nodes for the dynamic system—also defined as the minimum number of nodes \( N_D \)—that are needed to guide the dynamics of the complete system. Thus, these nodes provide a lower bound on the number of inputs that must be manipulated in order to control the system; however, they will not tell us how to design a control law for the system. These driver nodes are obtained through the network’s properties such as the degree distribution. The methodology consists of obtaining the average size and number of maximum matching using the cavity method. The minimum number of nodes is related to the size of the maximum matching in a corresponding digraph (i.e. directed graph). In a directed graph, \( M \) is the set of input nodes also called origin, and the state nodes connected to the origins are called controlled nodes. Thus, a link subset \( M \) is a matching or independent link set if it is a set of links without common nodes. A node is matched if it is an ending node of a link in the matching. Otherwise, it is unmatched. A maximum matching also known as maximum cardinality matching [26] is a matching that contains the largest possible number of nodes. The matching number is the size of a maximum matching. For more detailed information regarding maximum matching and the cavity method please reference to the supplementary information linked to the online version of the paper by Liu et al. [1].

The cavity method is a powerful method employed to compute the properties of the ground state in condensed matter physics and optimization problems [27]. This method was developed in statistical physics [28]. This method involves a cost function (e.g. internal energy function \( E_G(\beta) \)) which gives, for each matching \( M \), twice the number of unmatched nodes. Moreover, the cavity method involves a \( \beta \) parameter which is the inverse of the temperature. In general, when \( \beta \to \infty \) (zero temperature limit), the internal energy function gives the ground states properties such as the average number or density of unmatched nodes (i.e. \( n_D \)). Although the cavity method was originally developed at finite temperature, it is easier to apply when it is directed at zero temperature [27]. The cavity method is important for counting matching in a graph or network. Therefore, based on these ideas, Liu et al. [1] developed a set of self-consistent equations whose inputs are degree distributions and whose outcome is the average number \( n_D \). In other words, the minimum number of driver nodes \( (N_D) \) is determined by the number of

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Degree of heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1: simple model</td>
</tr>
<tr>
<td>( H ) (in-degree)</td>
<td>0.17</td>
</tr>
<tr>
<td>( H ) (out-degree )</td>
<td>0.17</td>
</tr>
</tbody>
</table>
incoming and outgoing links of each node which are described through the in-degree and out-degree distributions $P(k_{in})$ and $P(k_{out})$, respectively. The set found as driver nodes is not usually unique, i.e. it can be possible to obtain different potential control configurations with the same number of driver nodes. However, identifying those driver nodes is not a trivial task. Particularly, when the in-degree and the out-degree distributions are not similar and not defined by continuous simple distributions like a Poisson distribution. Without going into excessive detail, we provide the set of equations which were used in our analysis to find the minimum number of driver nodes.

To compute the number of drivers nodes ($ND$) of each model, we use the following equation derived from the cavity method for zero temperature (1):

$$ND = N \ast n_D$$

where $N$ is the number of nodes (i.e. states variables) involved in the network and $n_D$ is the minimum density of unmatched nodes or equivalently the minimum density of driver nodes computed from the following equation:

$$n_D = \frac{1}{2} \left[ (G(\hat{w}2) + G(1 - \hat{w}1) - 1) + (\hat{G}(w2) + \hat{G}(1 - w1) - 1) + \frac{\langle k \rangle}{2} (\hat{w}1(1 - w2) + w1(1 - \hat{w}2)) \right]$$

where each expression of $G$ and $W$ is functions which are defined by:

$$G(\hat{w}2) = \sum_{k_{out}=0}^{\infty} P(k_{out}) (\hat{w}2)^{k_{out}}$$

$$G(1 - \hat{w}1) = \sum_{k_{out}=0}^{\infty} P(k_{out}) (1 - \hat{w}1)^{k_{out}}$$

$$\hat{G}(w2) = \sum_{k_{in}=0}^{\infty} P(k_{in}) (w2)^{k_{in}}$$

$$\hat{G}(1 - w1) = \sum_{k_{in}=0}^{\infty} P(k_{in}) (1 - w1)^{k_{in}}$$

$$\hat{w}1(1 - w2) = \sum_{k_{in}=0}^{\infty} \frac{(k_{in} + 1)P(k_{in} + 1)}{\langle k \rangle} (1 - w2)^{k_{in}}$$

$$w1(1 - \hat{w}2) = \sum_{k_{out}=0}^{\infty} \frac{(k_{out} + 1)P(k_{out} + 1)}{\langle k \rangle} (1 - \hat{w}2)^{k_{out}}$$
Table 4  Characteristic of the network

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Nodes (N)</th>
<th>Number of Links (L)</th>
<th>minimum density of driver nodes (n_D)</th>
<th>Number of nodes to be controlled (N_D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple model</td>
<td>12</td>
<td>59</td>
<td>0.12</td>
<td>2</td>
</tr>
<tr>
<td>Intermediate model</td>
<td>12</td>
<td>61</td>
<td>0.12</td>
<td>2</td>
</tr>
<tr>
<td>Integrated economic-ecological model</td>
<td>19</td>
<td>155</td>
<td>0.59</td>
<td>12</td>
</tr>
</tbody>
</table>

However, the values of \(w_1, w_2, \hat{w}_1\) and \(\hat{w}_2\) are the weights (probabilities) of cavity fields computed by:

\[
w_1 = \sum_{k_{\text{out}}=0}^{\infty} \frac{(k_{\text{out}} + 1)P(k_{\text{out}} + 1)}{\langle k \rangle} (\hat{w}_2)^{k_{\text{out}}}
\]

\[
w_2 = 1 - \sum_{k_{\text{out}}=0}^{\infty} \frac{(k_{\text{out}} + 1)P(k_{\text{out}} + 1)}{\langle k \rangle} (1 - \hat{w}_1)^{k_{\text{out}}}
\]

\[
\hat{w}_1 = \sum_{k_{\text{in}}=0}^{\infty} \frac{(k_{\text{in}} + 1)P(k_{\text{in}} + 1)}{\langle k \rangle} \hat{w}_2^{k_{\text{in}}}
\]

\[
\hat{w}_2 = 1 - \sum_{k_{\text{in}}=0}^{\infty} \frac{(k_{\text{in}} + 1)P(k_{\text{in}} + 1)}{\langle k \rangle} (1 - \hat{w}_1)^{k_{\text{in}}}
\]

\[
0 \leq n_D \leq 1
\]

where the expressions of \(P(k_{\text{in}})\) and \(P(k_{\text{out}})\) represent the in-degree and out-degree probability, respectively, taken from Figs. 5–7, depending which model we were studying.

It can be seen that Equations (11–14) can be solved simultaneously using a non-linear programming optimization method. Table 4 presents the results of for the three models. In the first two models, two nodes are sufficient to control the system, while the third model needs 12 nodes out of 19 to control the system.

5. Discussion of results

We have used ideas and concepts from network analysis to characterize the structure of three prototypical complex networks and to evaluate their controllability. Among the properties analysed, the most relevant characteristic for controllability of a complex network was the degree distribution. As mentioned before, the degree distribution is used to determine the minimum numbers of nodes \((N_D)\) that will provide the number of inputs needed to drive or control the dynamics of the system. It was found that for Models 1 and 2, one pair of nodes was needed to control the system. This is also independently confirmed by Shastri et al. [29]. They studied optimal control of these two systems and proposed that the intermediate system is fully controllable if we control two variables. Table 5 presents four parameters involved in the dynamic of the systems, these parameters also represent the possible control policies that can be manipulated. For example, the mass transfer coefficients from \(RP\) to various plants (i.e. \(g_{RP1}\) and \(g_{RP2}\)) govern the consumption patterns of natural resources. Since \(P1\) and \(P2\) are domesticated...
(i.e. agriculture), manipulation of $gRPP1$ and $gRPP2$ reflect changes in the intensity of agricultural activities [29].

On the other hand, the integrated model needed 12 nodes to control its system. Although, the increase in the number of nodes (i.e. states variables) made this model more complicated, this increase was not necessarily the main factor that made the third model more difficult to control. In fact, the problem is the dispersion of the model which represents the less connected and the more connected nodes as it is described by the degree of heterogeneity. When we calculated the degree of heterogeneity, we observed that Model 3 had higher values compared with Models 1 and 2. This outcome is also confirmed by Liu’s results [1]. He stated that the larger the differences between the node degrees, the more driver nodes that are needed to control the system making the system more difficult to control. As mentioned before, Model 3 was studied by Kotecha et al. [19]. In this paper, the system was subjected to two different scenarios: population explosion and per capita consumption increase. It was found that this system, in the current condition and including both scenarios, is not sustainable leading to economic instability and biological extinctions. Therefore, the study of controllability becomes very useful tool for this model as we determined that we need to find 12 accessible nodes to control this system to make it as sustainable as possible. These 12 nodes can be described through policies and strategies that can be impose by government or private entities such us pollution fee (e.g. discharge fee to the industry sector due to waste disposal) or economic price (e.g. regulation of the plant consumption P1 by herbivorous animals H1). In Doshi et al. [30], we presented a further study of the controllability of Model 3. The objective there is to derive socio-economic policies and maximize the sustainability of this system using optimal control techniques based on the controllability study presented here.

These results are important because most real environmental systems can be represented as networks, and most environmental management strategies and policies are essentially efforts to control the dynamics of the network. Then, according to our results, these strategies and policies are likely to be successful only if they are designed to access the minimum necessary number of nodes. This is not always possible in real systems, and this may be one reason why efforts at environmental management for sustainability are not always successful. Our results, however, provide guidance on how to approach the problem in a general manner using the three aforementioned models as illustrations and, possibly, other more elaborate models.

6. Summary

In this paper, we studied the controllability of three illustrative complex systems whose structures represent in very simplified form the Earth’s ecology coupled with an economy, industrial production, energy generation, and elements of a social system. A recently proposed theory shows how the controllability
of complex systems can be derived from their network properties; therefore, we studied network characteristics such as the degree of nodes, and the degree distributions for these systems. Based on these characteristics, we found the minimum numbers of nodes that provide the number of inputs needed to control and drive the system towards sustainability. As a result, it was found that the more complex Model 3 is more difficult to control because the minimum number of driver nodes represents 60% of the total nodes present, meaning that we need 12 nodes out of 19 to fully control this system. In contrast, the simpler Models 1 and 2 needed only 11% of its nodes to control it system. The fact that control of a more complex system requires access to more nodes than is the case for a simpler one may seem simply logical, but the fact that exactly 12 nodes are required is not obvious. Note also that while the progression from Model 1 to Model 2 to Model 3 may seem relatively straightforward, the progression in the number of nodes needed for control is not a simple progression. The reason why this important is that we are attempting to demonstrate that policies derived for simple model systems cannot in generally be expected to perform as well when applied to more complex real life problems. One reason is that the structure of the networks is simply not the same, and significant modifications are, therefore, needed to make the control or management strategies perform well. The work presented here provides some illustrations on how this transfer could be successfully done. This is important as humanity tries to manage the very complex problems associated with maintaining the sustainability of the planet. In future work, we will design optimal control profiles for these systems to increase the life span of sustainability of this third system. Further using the insights from the theory and three prototype models presented here, we hope to be able to analyse the management of real systems for sustainability.

Acknowledgement

The authors will like to thank Dr. Yang Yu Liu and the NWB Team for their valuable comments and clarifications.

Funding

This work is financially supported by Center for Uncertain Systems: Tools for Optimization & Management (CUSTOM), Vishwamitra Research Institute, USA.

REFERENCES