THE AUSTRALIAN NATIONAL UNIVERSITY

INTEGRATION INSIGHTS

Number 12

A COMPENDIUM OF MODELLING TECHNIQUES

Jennifer Badham

ISSN 1834-304X

May 2010

Australian Government
Australian Research Council

CEPS
ARC Centre of Excellence in Policing and Security

DPMP
Drug Policy Modelling Program

ANU COLLEGE OF MEDICINE, BIOLOGY & ENVIRONENT

INTEGRATION INSIGHTS

Number12, May 2010

A COMPENDIUM OF MODELLING TECHNIQUES

Jennifer Badham

This Integration Insight provides a brief overview of the most popular modelling techniques used to analyse complex real-world problems, as well as some less popular but highly relevant techniques. The modelling methods are divided into three categories, with each encompassing a number of methods, as follows: 1) Qualitative Aggregate Models (Soft Systems Methodology, Concept Maps and Mind Mapping, Scenario Planning, Causal (Loop) Diagrams), 2) Quantitative Aggregate Models (Function fitting and Regression, Bayesian Nets, System of differential equations / Dynamical systems, System Dynamics, Evolutionary Algorithms) and 3) Individual Oriented Models (Cellular Automata, Microsimulation, Agent Based Models, Discrete Event Simulation, Social Network Analysis). Each technique is broadly described with example uses, key attributes and reference material.

PREFACE

by Gabriele Bammer

Modelling is one key approach to better understanding and responding to complex real world problems, be they in national and international security, population health, education, environment or the myriad other dimensions of human endeavour. Integration and Implementation Sciences (I2S) is a discipline which supports research on such problems by bringing together disciplinary and stakeholder knowledge, understanding and managing unknowns, and providing integrated research support for policy and practice change. Modelling is a central element of I2S, as it can be used to address each of these three domains. In other words, models can be used to synthesise knowledge, to better comprehend unknowns and their consequences, and to provide decision support.

Models can only ever be partial representations of the real world, but different modelling approaches provide handles on different facets of a problem's complexity. Both the process of model building and the model itself provide a systematic way of developing a more comprehensive understanding of key aspects of the problem. Nevertheless, it is difficult to find accounts comparing modelling methods, the elements of complexity that they tackle, and their strengths and weaknesses. This seems to be a fundamental gap, one which this compendium starts to address.

The aim of this document is to provide a solid basis for discussion and additional work in refining and expanding the compendium. Everyone will not agree with the categorisations presented here, and there are other modelling methods which could be included. Comments, discussion and suggestions for additions are therefore very welcome.

The stimulus for producing this compendium comes from the ARC Centre of Excellence in Policing and Security, which recognises that it can learn from techniques developed in environment, population health and elsewhere. Similarly, environment researchers can learn from approaches to security and population health, and so on. The compendium therefore has relevance across the board.

This Integration Insight is a departure from those produced previously. While it is still an overview of techniques, this is a substantial compilation rather than a brief digest. The importance of the topic warrants this expansion in the role of the Integration Insights series.

INTRODUCTION

Terrorism, drug use and climate change are all examples of real world problems that are particularly difficult to analyse and resolve, because they arise from the behaviour of complex systems. They have several characteristic features, including:

- They arise from the behaviour of people each making individual decisions in response to their beliefs and objectives and the information they have about their situation;
- There are multiple perspectives about what the problem is and why it exists;
- There are interactions between aspects of the problem, with the result that a change to address one part of the problem has unintended immediate effects on another facet and potentially counterintuitive future consequences.

There are many different systems thinking techniques which can be used to understand, communicate and forecast particular aspects of a problem. Each technique highlights specific characteristics and hides others. An example demonstrates how different modelling approaches are associated with very different ways of thinking about an issue (as reported in Richmond 1993, pg 128):

> A popular economic journal published the research of a noted economist who had developed a very sophisticated econometric model designed to predict milk production in the United States. The model contained a raft of macroeconomic variables woven together in a set of complex equations. But nowhere in that model did cows appear. If one asks how milk is actually generated, one discovers that cows are absolutely essential to the process. Thinking operationally about milk production, one would focus first on cows, then on the rhythms associated with farmers' decisions to increase and decrease herd size, the relations governing milk productivity per cow, and so on.

Before choosing a technique, you need to know both what you want to achieve by modelling and which aspects of the system you want to highlight. For many problems, it may be most useful to apply several different techniques to build a richer understanding.

This guide provides a brief overview of the most popular modelling techniques used to analyse complex real-world problems, as well as some less popular but highly relevant techniques. It does not, however, deal with all the modelling methods that could be useful. Each technique is broadly described with example uses and key attributes. Where appropriate, the guide also identifies reference material and software to help you implement the chosen technique(s).

The modelling methods are divided into three categories, with each encompassing a number of methods, as follows:

Qualitative Aggregate Models

> Soft Systems Methodology
> Concept Maps and Mind Mapping
> Scenario Planning
> Causal (Loop) Diagrams

Quantitative Aggregate Models

> Function fitting and Regression
> Bayesian Nets
> System of differential equations / Dynamical systems
> System Dynamics
> Evolutionary Algorithms

Individual Oriented Models

- Cellular Automata
- Microsimulation
- Agent Based Models
- Discrete Event Simulation
- Social Network Analysis

Before describing each of these, additional background information is provided about systems and models.

What is a System?

A system is any group of entities that acts together. In many systems, the group acts together for a purpose. For example, control systems are used in engineering to regulate extreme behaviour of some group of physical components

In complex systems, the behaviour of the system arises naturally from the interactions between the components. This is referred to as emergence because, in some sense, the characteristics of the system are distinct from the characteristics of the individuals and 'emerge' from the interactions. Examples include the price set by an economic market (arising from individual buying and selling decisions), oscillations in the populations of predators, mob rioting, and the average throughput of an assembly line.

What is a Model?

A model is any representation of relevant features of the entity under consideration. The most effective model will depend on the characteristics of the system to be modelled and the question to be answered. Sufficient detail is required to preserve those aspects relevant to the question, but other details must be excluded or the model will be as complicated as the original entity.

For the purposes of this guide, a model is primarily a description of the relationship between system components. For complex systems, the model is expected to also demonstrate the behaviours of the system being modelled.

A model generally refers to the system as it currently exists. That is, models can be used to consider ways in which the system may change as the strengths of relationships between components change, but cannot be used to examine systems with a different set of components or interactions. In these cases, a different model would be required. For example, a model can assist with options that change the size of an incentive, but a new incentive requires a change to the model structure.

Why Model?

There are many reasons why an analyst or decision maker may wish to model a system (see Epstein 2008 for another perspective). These reasons also impact on the choice of modelling technique, as different approaches are more effective in achieving specific purposes.

A formal modelling task can be useful in extracting *information* about the system from the various people with knowledge of the system, each with their own perspective. That is, development of a model provides a methodical approach to identify perceptions about the key aspects of the system.

More generally, a model can stimulate two-way or group *communication* about how a system works. For example, a scale physical model of a proposed building is an efficient way for an architect to convey their vision to prospective clients and facilitate discussion about the features of the building. For a complex system, this communication may be achieved with the design of a model, rather than the model itself. That is, the structure, interaction rules and other aspects that define the model provide a shared understanding of the key relationships and other features of the system. This communication may be achieved regardless of whether the model is actually built, simply by formally participating in the model design.

Models can also be used to investigate and falsify competing *explanations* for certain behaviour or features in the system. To achieve this, models that represent each proposed explanation are built to determine whether the proposal does, in fact, produce the behaviour in question. Of course, generating the required behaviour does not imply the explanation is 'true', but inability may suggest the explanation is false.

For those systems where there is behaviour data for the system and its components, a model can be used for *data compression*. In a sense, the model summarises all of this data. This compression results in a loss of information, but the model can be supplemented with information about its accuracy at various tolerance levels and/or retained complete data for more extreme situations that are poorly represented by the model.

Projections and 'what if?' analysis can only be conducted where a model has been validated against detailed data sets. Projections use the model tuned to current system parameters to estimate future system behaviour. Input parameters can also be varied to identify possible consequences of changing the system slightly. Such estimates assume that the model accurately includes all relevant structures and relationships, and that parameters are correct. The complex behaviour of a system makes extrapolations particularly risky if more than one parameter is varied or any parameter is varied outside of the values available in the validation dataset. Due to the wide use and reporting of economic projections, this type of analysis is familiar and may be the most common 'front of mind' use of modelling, despite its limited value for many problems.

For policy and research integration, the discipline of creating an explicit model may provide the greatest benefit. This is particularly true where the modelling software or documentation provides a clear description of the model structure. While most of the modelling techniques described in this guide can be implemented efficiently in JAVA, C++ or some other programming language, the guide focuses on purpose built software with visual construction of model structure and built in reporting as these features facilitate communication between modellers and other participants.

BUILDING A MODEL

For policy and research integration, understanding and communication may be the major objectives and much of the modelling benefit arises from the development process rather than directly from the model itself. This process has several steps. The implementation of each step differs substantially depending on the purpose of the model and the modelling technique to be used, but each step must be explicitly considered.

Initialisation

The first step is to define the problem. Without a clear problem or question to be answered, it is difficult to assess the relevance of various system components and relationships, so the model is likely to be confused and too detailed.

The participants in the modelling project must also be identified early in the project, together with their roles. Participants are likely to include decision makers responsible for implementing any recommendations from the project, subject matter experts and modelling specialists. Subject matter experts include people affected by the system being modelled (such as consumers), operational managers who make administrative decisions within the system and data managers.

As for any other project, many of the participants will have limited time and may be independent of the organisation developing the model. Their involvement may be constrained, and obtaining their interest and commitment is necessary for the model to be useful. Typically, the modeller and a coordinator with some subject matter knowledge would devote substantial time to the project with other participants involved at key points only.

Design

The second phase is to design the model. The modeller and subject matter experts must agree on the factors to be included in the model and the assumptions concerning the factors excluded from the model. Other scope decisions may include the time frame for dynamic models and whether the model should be qualitative or quantitative.

The objective of the design phase is to teach the modeller how the system operates. The design should include a description of the pattern of influences between system components. One useful approach for many systems is to trace possible paths of people (in social systems) or objects through the system and identifying branching points in these paths and the factors that determine which branch occurs.

This knowledge can be captured with some of the qualitative modelling techniques described below. The design, or business rules, should be agreed by the participants before the model is constructed.

Build

If the objective of the model is simply to understand and communicate the key features of a system, the design document may constitute the model and no separate 'build' phase is necessary. Generally, however, the modeller will construct a model that implements the design to the extent possible.

Depending on the modelling technique, the build phase can include:

> Translate the design into the perspective of the modelling technique;
> Discover the exact relationship between the cause and effect for each identified relationship from research and/or data;
> Identify missing information or data;
> Construct the user interface to receive input parameters and generate output data and charts;
> Basic testing, including handling of extreme cases;
> Calibrate the model against historical or snapshot data;
> Preparation of technical and user documentation.

It is unlikely that the model will be built exactly as designed. Wherever the implementation differs from the design, the change and the reason for that change should be documented.

Confirmation

In addition to basic testing, a formal validation and verification phase is required with all participants to confirm that the model is reasonable. This includes checking that:

> Change to each input individually results in an overall change of the expected direction and magnitude;
> Real world results that were not used in the calibration can be replicated in the model;
> The user interface is easy to interpret;
> Each business rule agreed in the design phase has been implemented or, if relevant, included in the change documentation;
> Design changes made during the build phase are appropriate;
> The question originally asked is illuminated by the model.

As a result of the confirmation phase, further building and confirmation may be required.

Application

The model is then used. For models that must be 'run', this may involve the modeller running a set of scenarios and preparing a report, or distributing the model for participants to run directly. For conceptual models, use may be circulating the documentation for discussion and comment.

It is likely that some relationships could not be completely determined from available research and data. In such cases the application phase may also include sensitivity analysis to assess whether the missing information has a significant effect on the model outcome. This analysis tests various possible values for the missing information and examines how the outcome varies. If the model is sensitive to some of the missing information, the analysis can also assist in identifying which information is the most important to obtain.

QUALITATIVE AGGREGATE MODELS

Qualitative aggregate modelling techniques are primarily used for understanding an issue, problem structuring, integrating different perspectives and communicating the system structure. While they cannot be used for projection, since they do not contain quantitative information, some further analysis may be able to identify relatively important features or characteristic behaviour.

Qualitative methods can also be used in the design phase for a quantitative technique to identify the key relationships for which data is required.

Soft Systems Methodology

Soft Systems Methodology (SSM) was developed as a systematic approach to examine human problems in organisations and agree policy responses. The method is a straightforward process of understanding the problem, developing options and implementing the decision. However, there are two innovative methods used to elicit different perspectives on the problem from the group, Rich Text Pictures and CATWOE.

A Rich Text Picture is relatively unstructured but there are some content requirements. It includes a boundary that limits the scope of the problem to be described. External constraints are listed outside the boundary. Within the boundary, each stakeholder is pictured with their concerns and the relationship with other stakeholders.

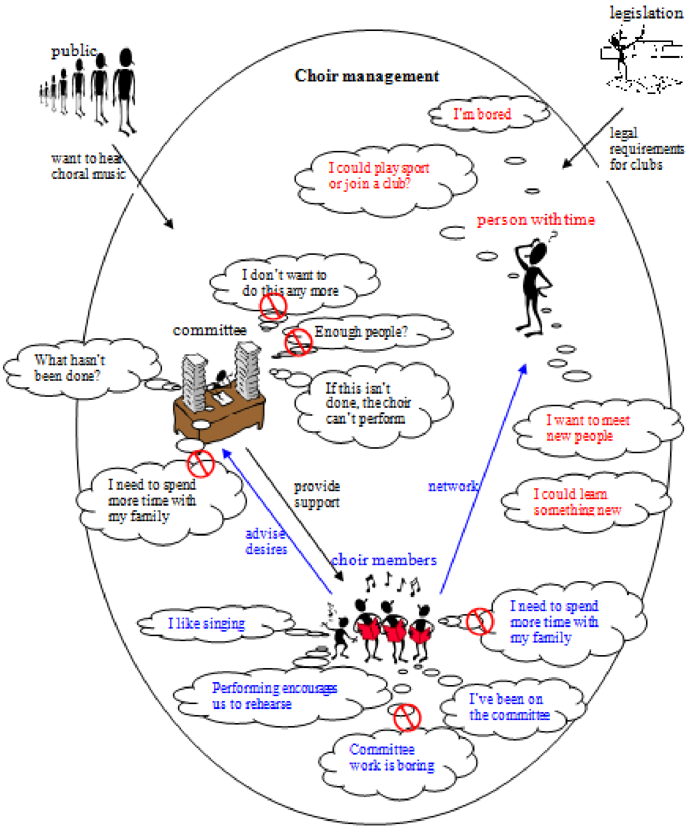

Figure 1: A possible Rich Text Picture describing management of a community choir

An example is shown at Figure 1. It describes a conflict within a community choir where certain management functions must be undertaken for legal, publicity or performance reasons but there are insufficient volunteers.

CATWOE is an acronym for customers, actors, transformation, Weltanschauung (or worldview), owner and environment.

For the same example, one CATWOE response could be: customers are choir members and audience, actors are choir members and their friends, the transformation desired is to ensure essential functions are undertaken, worldview is that performance is the reason for the choir's existence, owner of the system is the choir management committee, and the environment is the relevant laws.

Key features	What it answers:	Assists groups to elicit various perspectives about an issue and formalises statement of problem.
	What is highlighted:	Competing perspectives.
	Inputs:	Opinions from group members.
	Outputs:	Richer understanding of the issue and potential conflicts.
	Relationships:	Stakeholder perspectives of the problem and potential solutions are explicitly elicited.
	Traps:	Relies entirely on the expertise and opinions of group members.
	Handling uncertainty:	Nil
	Data needs:	Supporting information for claims made by stakeholders.
Resources	References:	Checkland and Poulter 2006, Learning For Action: A Short Definitive Account of Soft Systems Methodology, and its use for Practitioners, Teachers and Students
	Software:	None generally used.

Concept Maps and Mind Mapping

Concept Maps are used to structure knowledge about an issue. They can be conceived as a formal approach to mind mapping and knowledge extraction. Each concept or aspect of the issue (generally nouns) is drawn in a box, and links are drawn between those boxes where there is a relationship between the concepts. The link is annotated with a description of the relationship (often verbs).

An example is shown at Figure 2, a section of a concept map about Sickle Cell Disease (Rendas 2006). This section distinguishes between the presentation of some acute symptoms ('can be') and the underlying causes ('due to').

Concept maps may be maintained in electronic form so that more detailed information can be hyperlinked to the relevant aspect of the knowledge structure.

Mind maps also link related concepts and ideas but use a radial hierarchy, generally without link annotation. A person or group brainstorms some central issue and the mind map diagram documents the brainstorming, so that the first ring expands the central issue and the second ring expands each of the issues in the first ring and so on.

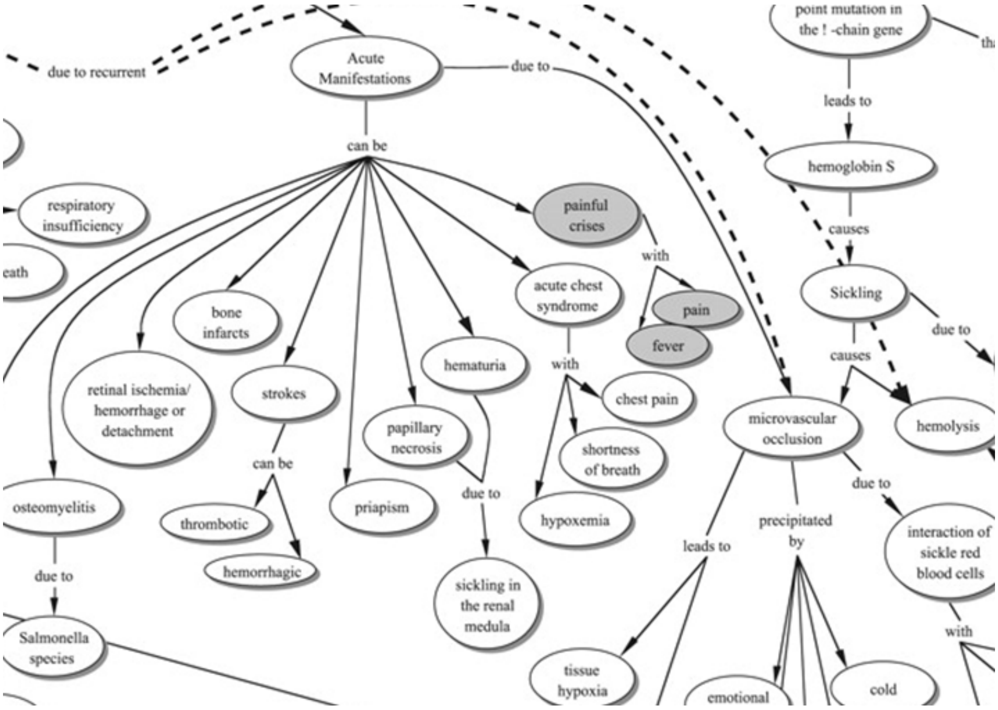

Figure 2: Section of a concept map concerning Sickle Cell Disease (Rendas 2006)

Key features	What it answers:	What concepts are related to some central issue and the relationships involved.
	What is highlighted:	Concepts, ideas and other aspects of knowledge.
	Inputs:	The perspectives of those involved in the brainstorming about what is relevant.
	Outputs:	Documentation of the structure of knowledge or ideas.
	Relationships:	Relationships between ideas are explicitly documented in some mapping approaches.
	Traps:	Relies entirely on the expertise and opinions of group members.
	Handling uncertainty:	Generally nil, though ideas about what is unknown can be included.
	Data needs:	Nil
Resources	Software:	CMapTools, concept maps, by the Institute of Human Cognition (cmap.ihmc.us) PersonalBrain, mind mapping, but without the radial hierarchy, so links can connect any concepts, by TheBrain Technologies (www.thebrain.com) FreeMind, an open source mind mapping project in JAVA (http://freemind.sourceforge.net/wiki/index.php/Main_Page) MindManager, a commercial mind mapping application (www.mindjet.com)

Scenario Planning

Scenario Planning is a structured approach to identify plausible future situations to which an organisation may need to respond. That is, it concerns the external environment of an organisation rather than internal issues over which it has some control. Once these situations are identified, they can be planned for, usually including some way of monitoring so that the likelihood of the scenario occurring can be regularly reassessed. Most modelling techniques describe the current system and can incorporate only small variations. In contrast, scenario planning is used to describe possible extreme situation.

There are several essential steps in the process. The first is to identify major trends and key factors that shape the environment. Each of these trends and factors is assessed for importance (such as strength of influence or potential impact) and for uncertainty about future character. Scenarios are then defined by combining extremes of the various trends and factors (providing they are not incompatible), particularly focusing on those of relatively high probability.

Each scenario is then fleshed out, developing a description of the situation associated with the combination of extreme outcomes. These descriptions then provide the basis of a standard planning process: developing a research program, monitoring techniques and responses.

Key features

What it answers:	Plausible future situations that impact on an organisation but are outside the organisation's control.
What is highlighted:	Competing perspectives.
Inputs:	Trends in factors that shape the organisation's environment, such as economic and political features.
Outputs:	Set of plausible scenarios that describe the potential extreme situations in which the organisation may find itself in the future.
Relationships:	Relationships between factors are considered when combining factors to generate scenarios.
Traps:	Does not generate solutions.
Handling uncertainty:	Extreme situations are specifically sought.
Data needs:	Historical data for correlations in changes.

Resources

Software:	None generally used.

Causal (Loop) Diagrams

Causal Loop Diagrams (CLD, also Causal Diagrams or Influence Diagrams) focus on the causal links within a system that relate to change over time. Two components of a system are linked if a change in one causes a change in the other. The link is depicted as an arrow from the cause to the effect. Each arrow also has a '+' or '-' to indicate whether an increase in the cause leads to an increase or decrease in the effect respectively.

An example Causal Loop Diagram is at Figure 4, showing the predator-prey relationship between foxes and rabbits. Births and natural deaths increase as population increases and, in turn, increase and decrease population respectively. Rabbits Eaten is of particular interest is; as Foxes increase, they eat more rabbits, which increases rabbit deaths and decreases fox deaths.

This example also displays a key feature of CLDs, feedback loops. A cycle of arrows from one system element to another and eventually to the original is referred to as a feedback loop (such as RabbitsEaten to FoxDeaths to FoxPopn to RabbitsEaten). If the number of negative arrows is even, the loop is a positive feedback or reinforcing loop. Any change is exaggerated, leading to exponential growth (or decay). On the other hand, if the number of negative arrows is odd, the loop is a negative feedback or balancing loop. Any change is damped within the system.

Identifying and classifying the strongest feedback loops can provide insight into the system's behaviour without quantitative input. CLDs are often used as part of the model design process for System Dynamics Models.

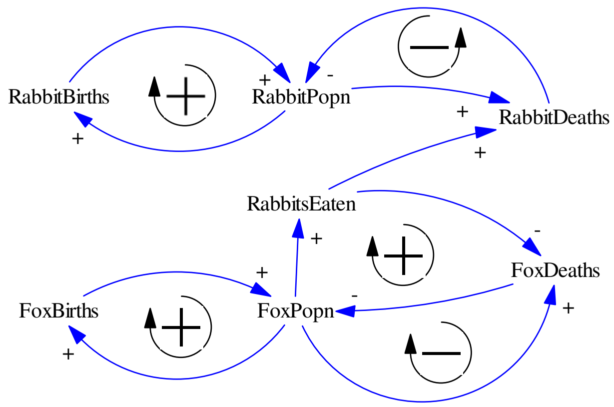

Figure 3: Causal Loop Diagram for the predator prey system (Adapted from Vensim sample models www.vensim.com)

Key features	What it answers:	How changes in one part of the system flow through to other parts and the system as a whole.
	What is highlighted:	System response to changes and feedback loops.
	Inputs:	Components of a system that change over time. Causal relationships between changes in system parts.
	Outputs:	Documentation of system structure.
	Relationships:	Direction of change in one system element caused by another system element.
	Traps:	Difficult to document indirect influences such as constraints or limits.
	Handling uncertainty:	Nil
	Data needs:	Supporting information for claims of influence.
Resources	References:	Roberts et al 1983, Introduction to Computer Simulation: The System Dynamics Modeling Approach
	Software:	Vensim, by Ventana Systems (www.vensim.com).

QUANTITATIVE AGGREGATE MODELS

Quantitative aggregate models use equations to describe relationships between the averages of pairs (or larger groups) of system components. The model is simply the set of equations. All models in this category have two limitations. The first is that confining the relationships to averages can be misleading when the average is not a meaningful description of the behaviour of the system part. The other limitation arises from the focus on parts of the system that can be quantified. Any system aspect that has substantial qualitative aspects cannot be modelled with these methods. This may also contribute to missing connections between quantitative measures, where a qualitative component connects them. For example, a link between duration of a treatment program and success is clearer with an intermediate link through quality of treatment.

Function Fitting and Regression

Regression assumes that one system measure of interest is dependent only on the values of some set of explanatory system measures. The purpose of regression is to use empirical data to fit a specific functional form that relates the value of interest to the values of the explanatory measures. Regression is a specific example of fitting a function that links input and output data.

For example, a dataset that contains the height, weight and age of many children could be used to estimate weight from a combination of age and height. It could also be used to estimate height from age. Both of these would be valid uses of regression.

The functional form must be provided in advance and cannot be derived from the regression. The same example dataset could be used to estimate age from a combination of height and weight. However, some thought suggests that weight depends on age and height, but age does not depend on height and weight. Regression cannot identify the correct interpretation; a statistically significant model could equally be developed for the invalid functional form.

Regression is particularly used for economic models. The related statistical modelling technique of time series analysis is used where time is one of the variables. Factor analysis or principal component analysis may be used prior to regression to identify a smaller set of system aspects on which to build the model.

Key features

What it answers:	Identifies the best fit version of a predefined function that relates one system aspect to a set of other aspects.
What is highlighted:	The explanatory power of a particular function.
Inputs:	Dataset that contains multiple instances of sets of system measures. Expected functional form relating one aspect to all other aspects.
Outputs:	Coefficients for the predefined function. Estimate of how well the model fits.
Relationships:	One aspect assumed to depend on the other system aspects.
Traps:	Identifies associations, not causal relationships. Different definitions of 'best fit' will select different model parameters. Generally used is the sum of the square of the errors (difference between estimate and actual), which strongly weights data outliers.

Incorrect interpretation of large or highly significant coefficients as the most important associations.

Limited capacity to assess whether the predetermined functional form is valid.

Handling uncertainty:	Difference between predicted and actual result is minimised. Specific use of goodness of fit and residuals tests to assess whether derived model fits the data.
Data needs:	Dataset with multiple records that each contains values for each of the system aspects of interest.

Resources

References:	Any statistics textbook
Software:	Spreadsheets (eg Excel) for small applications. Specialist statistical packages such as R (open source), SAS, Stata and SPSS.

Bayesian Nets

Bayesian Nets (BN, also Bayes Nets, Bayesian Networks and Belief Networks) depict conditional probability relationships. The nodes in the network represent system components and the links represent probabilistic dependencies between the states of those components. The network summarises the joint probability distribution of the component states. The network can be used to estimate the probability of specific states given information about some other states.

An example network is shown at Figure 4. The 'Explosion' node depends on the 'Oxygen' and 'Hydrogen' nodes. That is, the probability of an explosion depends on whether oxygen and hydrogen are both present (see Table 1).

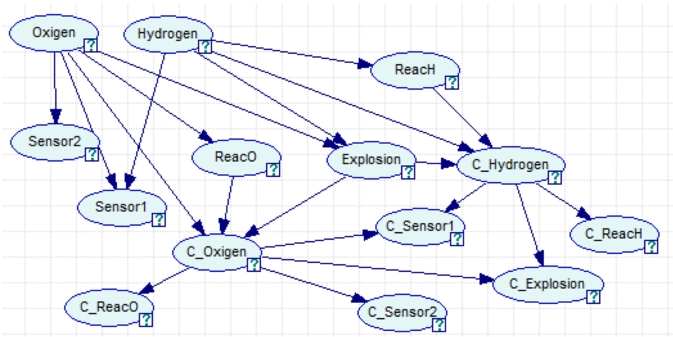

Figure 4: Bayesian Net screenshot from GeNIe (Decision Software Laboratory at http://genie.sis.pitt.edu) depicting the probability of an explosion and the ability of sensors to detect relevant factors.

Table 1: Probability structure for 'Explosion' node.

Present?	Yes	No
O yes / H yes	0.9	0.1
O yes / H no	0	1
O no / H yes	0	1
O no / H no	0	1

Key features

What it answers:	Probability that specific system component states are true, given knowledge of other (relevantly influential) system states.

	What is highlighted:	Probability dependencies between system component states.
	Inputs:	Set of probabilistic dependencies between states of system entities.
	Outputs:	Probability of unobserved states.
	Relationships:	Conditional probabilities.
	Traps:	Confidence in the belief is not represented. That is, an estimate of probability based on research or data has no greater influence on the model than a 'guess'.
	Handling uncertainty:	The probability distributions can be subjective beliefs rather than based on relative frequencies.
	Data needs:	Sufficient to justify belief structure.
Resources	Software:	GeNIe, by University of Pittsburg (genie.sis.pitt.edu)

System of Differential Equations/ Dynamical systems

Differential equations (DE) describe the change in a system as some function of the current values within the system. Unlike simulation techniques where time is updated in discrete steps, DEs deal with continuous time. The system is described by a set of DEs, which must be solved simultaneously. In many cases, the mathematics of a system of DEs is intractable. That is, the function describing the system at any time cannot be derived through mathematical manipulation. However, certain features of the system may be tractable, such as equilibrium points, using the techniques of dynamical system analysis.

For example, the predator prey model first shown in Figure 4 can be quantified with the Lotka-Volterra equations:

$$\frac{dF}{dt} = -\alpha F + \beta FR \qquad \text{where } F \text{ is foxes, } R \text{ is rabbits}$$

$$\frac{dR}{dt} = \gamma R - \delta FR \qquad \alpha, \beta, \gamma, \delta \text{ are constants}$$

In this formulation, the constants α and γ represent the difference between birth and death rates for foxes and rabbits respectively in the absence of the other subpopulation. The number of rabbits eaten by foxes is proportional to the probability of meeting, with β and δ converting this probability to the impact of these meetings on the rabbit and fox populations. Analytical techniques can be used to identify the fox to rabbit ratio at which the system is stable. At other ratios, the system oscillates: rabbits increase until there is sufficient food supply for foxes to increase, but then the high number of foxes leads to a shortage in the rabbit supply and foxes starve, which then allows the rabbit population to increase again.

Key features	What it answers:	Gives a complete description of the system at any point in time.
	What is highlighted:	A complete description of the system.
	Inputs:	Equations that relate the change in a system over time to the state of the system.

	Outputs:	State of the system at any time (if tractable).
	Relationships:	Equations relate change in the system to state of the system.
	Traps:	Only tractable for systems with few variables. Only includes quantitative information.
	Handling uncertainty:	Nil
	Data needs:	Nil
Resources	Software:	Specialist mathematical software, such as Mathematica, Matlab or SciLab.

System Dynamics

System Dynamics (SD) enables numerical analysis of a system of differential equations. The model consists of entities and an equation for each entity that includes other relevant entities and values. Arrows indicate those entities that contribute to an equation. SD uses the analogy of stocks and flows to represent amounts and changes in amounts respectively. Other structures in the system provide auxiliary information needed to create the equations.

Time is incremented in discrete steps, so flows are defined per unit time. At each time step the equations are calculated based on current values, including integration over time. All values are then updated synchronously in preparation for the next timestep.

An example System Dynamics model diagram is at Figure 5 and builds on the causal loops example provided in Figure 4. The fox and rabbit populations are stocks, which increase with the inflow of births and decrease with the outflow of deaths. The flexibility of SD models enables complex equations to be included, in this case concerning the impact of population ratios on the number of rabbits eaten by foxes and consequently on fox mortality.

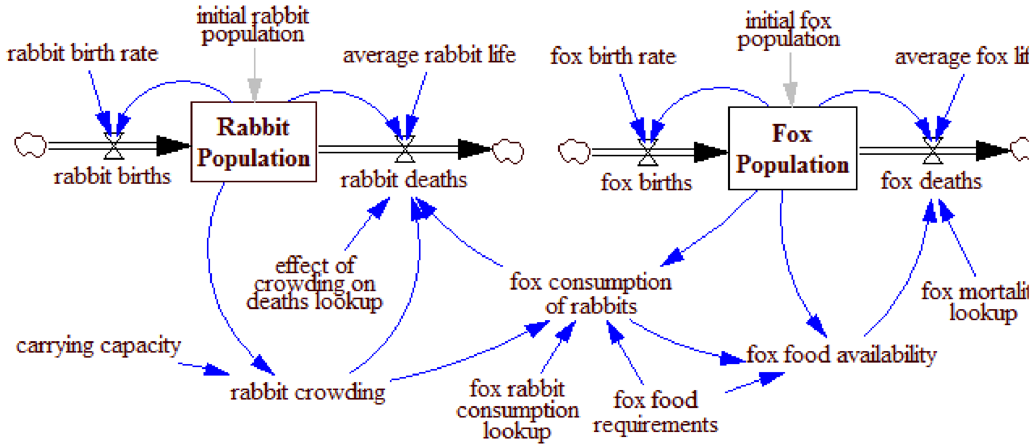

Figure 5: System Dynamics diagram for the predator prey system (Adapted from Vensim sample models www.vensim.com). The equations must also be specified to complete the model.

Key features	What it answers:	How the whole system behaves over time.
	What is highlighted:	Connections between aggregate aspects of the system, with each aspect represented by an average or typical value.

	Inputs:	Model structure based on stocks, flows and information. Quantitative estimates of the relationship between a set of causes and the combined effect for each aspect of the system. Values for each system aspect at the initial time.
	Outputs:	Aggregate value for each system element at each point in time.
	Relationships:	Equations are necessary for each system element that describes how it responds to the state of input influences.
	Traps:	Only includes quantitative information. Underlying time step must be small compared to natural times within system to avoid artificial behaviour.
	Handling uncertainty:	Sensitivity analysis can be used, constructing multiple simulations with information values drawn from distributions.
	Data needs:	Data required for calibration of the model.
Resources	References:	Sterman 2000, Business Dynamics: Systems Thinking and Modeling for a Complex World.
	Software:	Vensim, by Ventana Systems (www.vensim.com). Powersim Studio (www.powersim.com) STELLA / iThink, by isee systems (www.iseesystems.com)

Evolutionary Algorithms

Evolutionary Algorithms (EAs) are used to identify optimal solutions to a quantitative problem where the optimum cannot be found with analytical methods. An initial population of trial solutions is generated randomly. The value of the system for each of these solutions is calculated. The next population of trial solutions is generated by combining the better solutions.

For example, the 'hill climb' approach can be used to search for the maximum of a complex function. The function is calculated for several input values. The next generation would then be sampled from values near the initial values that had higher values of the function.

Another type of application is training a computer system to classify. For this application, the number of errors in classification is to be minimised, without specifying the classification rules in advance.

Key features	What it answers:	Where are the near optimal points (that is, maximum or minimum) for some complex function?
	What is highlighted:	Extreme values for a specific mathematical function.
	Inputs:	Function to be optimised. Process to select next generation of trial solutions.
	Outputs:	Sets of variable values with 'good' results.
	Relationships:	Function that converts input sets of values to a result.

Traps:	Difficult to find solutions where the function is discontinuous or very steep, as geometrically close points do not necessarily have similar function values.
Handling uncertainty:	Uses probability to generate each succeeding population of test solutions so that some diversity in spaces being search is maintained.
Data needs:	Depends on application. Generally nil for optimising a function and a training dataset for a classifier.

INDIVIDUAL ORIENTED MODELS

Individual oriented modelling techniques track individuals and calculate results by counting the relevant individuals. These are often referred to as 'bottom up' approaches as the model works at the lowest level within the system, explicitly representing only local interactions. Depending on the technique and the application, the individuals may be of different types and, within a type, may display different characteristics.

Cellular Automata

Cellular automata (CA) models are particularly effective for processes where physical location and geographical neighbourhoods are the most important cause of system change.

The archetypal model is Conway's game of life. The game starts with a 2D grid of cells that may be 'alive' (dark blue in Figure 6) or 'dead'. Whether a cell is alive in the next timestep is determined by its current status (dead or alive) and how many of the eight cells it is touching are alive. All cells are updated simultaneously and characteristic patterns emerge quickly.

More generally, a CA model has two components, a grid and a set of rules. The grid is typically made up of square cells set out in two dimensions and wrapped top to bottom (that is, the neighbours of the cells in the top row include cells in the bottom row) and right to left. However, grids can use identical cells of any shape in any number of dimensions. The rules define neighbourhoods and how the features of the relevant neighbourhood update the state of each cell. For some applications, the rule set may include a small probability of a random state update.

Typical applications of a relatively sophisticated CA model include the spread of disease in a crop, the spread of a bushfire or changes in opinion. However, the uniform nature of the structure limits the capacity to develop useful CA models. Refer to Agent Based Models and Social Network Analysis techniques for related but more realistic models.

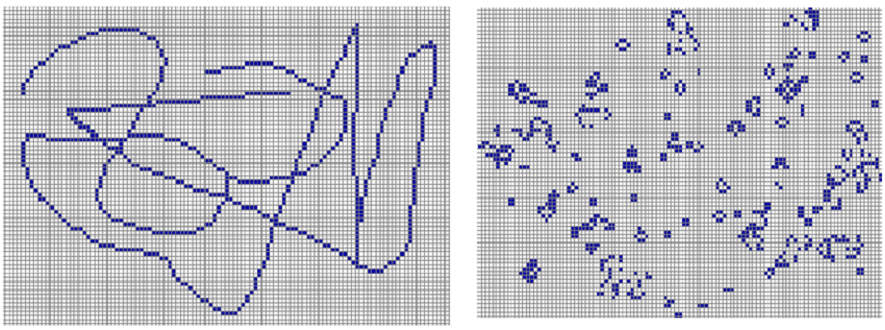

Figure 6: Screenshots from JAVA implementation of Conway's Game of Life (Hensel 2001): initialisation (left) and after several timesteps (right).

Key features	What it answers:	How do signals move through a geographically fixed population?
	What is highlighted:	Influence of geographical neighbours.
	Inputs:	Grid representing a physical space with uniform characteristics. Rules defining how a cell's state is affected by the states of neighbouring cells.
	Outputs:	Stable repeating patterns of state changes, or completely fixed cell states.
	Relationships:	Cells represent people, businesses, vegetation or any entity that is physically fixed but able to change state. Influential cells are entirely defined by neighbourhood.
	Traps:	Cells and rules are identical so there is no variation in cell responses or behaviour. No variation in susceptibility to influence, strength of influence on neighbours, and number of neighbours.
	Handling uncertainty:	Different initial states will lead to different stable patterns. Sets of simulations can be used to estimate the probability of particular patterns emerging.
	Data needs:	Calibration data for transmission speed.
Resources	References:	Gilbert and Troitzsch 1999, Simulation for the Social Scientist
	Software:	Any Agent Based Model software can be used to develop CA models (eg NetLogo, RePast). MatLab can be adapted naturally, with a matrix to represent the grid and matrix elements to represent cell states.

Microsimulation

Microsimulation models are effective where individuals do not interact with each other and have many characteristics that affect some issue of interest. A typical use would be to calculate and compare the effect of a proposed taxation change on household income for different population groups (such as single parents of two children with an income of about $50,000 per year).

The general process starts with a detailed unit record dataset, such as a census file. A single record represents multiple people with the same set of characteristics and the record includes a weight to identify the number of people represented. Other detailed datasets are merged for additional characteristics. This merging uses fine cross-tabulations based on data items common to both datasets. The unit records are also 'aged' as necessary to update for changes in the population. For example, detailed death rates are used to reduce weights in accordance with expected deaths, and incomes are altered in accordance with changes in income distribution over time. Finally, the relevant policy is applied to each record and its effect calculated. The average effect is then calculated from the record specific effect and the record's weight.

Key features	What it answers:	The effect of a policy on specific subpopulations.
	What is highlighted:	Winners and losers for each policy option.
	Inputs:	Detailed datasets describing population characteristics. Proposed policy rules, including application and effect.
	Outputs:	Impact of policy by population group.
	Relationships:	None; the individuals modelled do not interact with each other.
	Traps:	Ignores interactions between people and responses of people to a changing environment.
	Handling uncertainty:	Statistical methods are used for dataset matching and ageing of the datasets.
	Data needs:	Large detailed datasets from major national collections.
Resources	References:	Gilbert and Troitzsch 1999, Simulation for the Social Scientist
	Software:	Specialist statistical packages that efficiently process large datasets, such as SAS, Stata and SPSS.

Agent Based Models

Agent Based Modelling (ABM) is a very flexible bottom up modelling technique. It is generally used to explore possible ways in which simple individual behaviour can result in recognisable group patterns, and how those patterns are affected by differences in the individual behaviour. Very detailed (high fidelity) ABMs have also been developed for forecasting in social or economic systems. Multiple simulations are used to estimate the relative frequency of different patterns.

In ABMs, each agent represents a single person, firm or other entity that can change state and display behaviour. The behaviour of agents responds to current state, personal characteristics, environment and rules. The environment includes any features of the agent's location and the states of agents that are geographical or network neighbours. Behaviour rules can be different for each agent and can change over time.

The archetypal model is SugarScape (Epstein & Axtell 1996). This is a simple ABM with a regular grid of cells representing the landscape. Some landscape cells have resources (sugar or spice). Agents deplete resources by moving around the landscape, but can add resources when they land on appropriate cells (see Figure 7).

The implementation of the predator-prey model as an ABM is very different to a model using differential equations (as shown in Figure 6). Two types of agents would be established: predator (fox) and prey (rabbit). Each fox or rabbit would have some probability of giving birth or dying in each timestep. In addition, the foxes and rabbits would move around and, if they landed in the same place, the fox would eat the rabbit. Foxes would also die after some number of timesteps without eating a rabbit. Average results over many simulations would be expected to replicate the dynamic behaviour arising in SD models.

More sophisticated ABMs are used to develop a broad range of artificial societies. They can be used to simulate economic markets, where goods are traded based on the individual preferences of buyers and sellers. They can model the transmission of disease, with individual people only becoming infected where a neighbour is infectious, each person has a susceptibility that affects probability of transmission and the infectivity of the disease varies over the period of a person's infectious state. For belief models, a person's opinion can be affected by how many of their friends have specific opinions and how influential those friends are and their own tendency to be influenced by others. Other applications include transport systems and crowd behaviour.

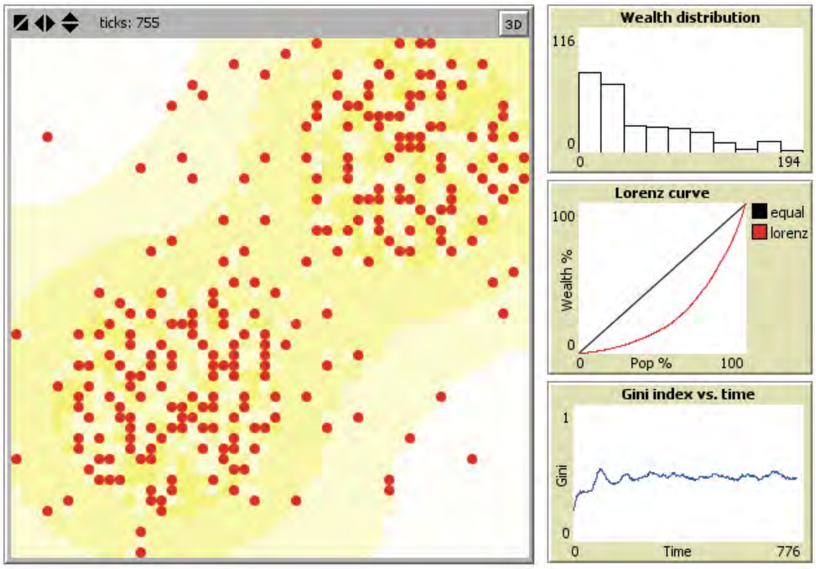

Figure 7: Screenshot of NetLogo SugarScape implementation (Wilensky 1999; Li & Wilensky 2009). Dots are agents and grid squares are coloured by resource amounts.

Key features

What it answers:	What social patterns emerge from simple behaviour of individuals?
What is highlighted:	The relationship between individual behaviour and group behaviour.
Inputs:	Personal characteristics for each agent. Behaviour rules (may vary by agent, or may be based on characteristics that vary by agent), which can include movement. Pattern of connections between agents.
Outputs:	Aggregate measures of system behaviour over time. Measures of variation within the system, such as distribution of resources.
Relationships:	Agents represent people, businesses or any other entity that can display behaviour (changing state). Physical location represents geographical environment, which may have features such as resource availability. Network links represent relationships between influential agents. Rules describe the relationship between agent behaviour and the influence of other agents and location.
Traps:	It can be difficult to interpret results as the connection between individual behaviour and system behaviour is visible only through repeated simulation.

	Handling uncertainty:	Individual simulations generate different results. Large numbers of simulations are run to estimate relative frequency of patterns of social behaviour.
	Data needs:	Minimal for models that are intended to explore how certain behaviour can emerge as they are not calibrated to real world systems. For real world problems: • Detailed data concerning the distribution of relevant behaviours and contacts, and the process of transmission or influence. • System aggregate data for calibration.
Resources	References:	Gilbert 2008, Agent-Based Models.
	Software:	CORMAS (http://cormas.cirad.fr/en/outil/outil.htm): Oriented to environmental research. NetLogo (http://ccl.northwestern.edu/netlogo): Easy to learn and good for relatively simple models. MASON (http://cs.gmu.edu/~eclab/projects/mason): Specialised for large number of agents. RePast (http://repast.sourceforge.net) Swarm (www.swarm.org)

Discrete Event Simulation

Discrete Event Simulation is used to model systems where there are processes of known duration and queues for resources. Typical applications include assembly lines, queuing systems (such as customer service) and supply chains (see Figure 8).

The general approach is to program a sequence of events, including arrivals, conveyors and assembly points. Each event can trigger a future event. For example, an arrival will also generate the time for the next arrival, and completing a process may move a waiting person or item from the queue and restart the same process. The simulation moves from event to event (in sequential order) rather than updating in regular time.

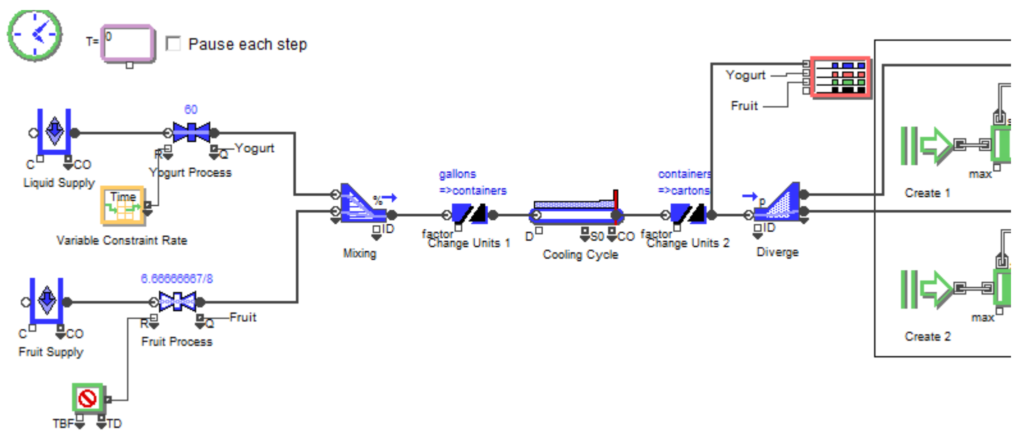

Figure 8: Screenshot of section of ExtendSim demonstration model of a yoghurt manufacturing process (ExtendSim model copyright © 1987-2010 Imagine That Inc. All rights reserved).

Key features	What it answers:	Identifies bottlenecks and allows trialling of options to improve efficiency.
	What is highlighted:	How constraints in the system affect overall system performance.
	Inputs:	Distribution of entry to the system. Processes within the system and duration information. Paths that persons or items follow through the system.
	Outputs:	Statistical information concerning any object in the system, such as average queue length and waiting time.
	Relationships:	Sequence of processes within the system.
	Traps:	For processes that involve people, any changes over time in the behaviour of people are ignored. For example, if bank queues are always long, arrivals to the bank may reduce over time as people decide to bank elsewhere.
	Handling uncertainty:	System reports can include various statistical measures that include averages and variation.
	Data needs:	Relatively straightforward information about processing time.
Resources	Software:	Simul8, by Simul8 Corporation (www.simul8.com) ExtendSim, Imagine That (www.extendsim.com)

Social Network Analysis

Social Network Analysis (SNA) focuses on the structure of the relationships between entities (generally people, households or organisations). Each entity is represented by a node (or vertex or actor) and two nodes are connected by an edge (or link or arc) if they have the specified relationship. The same set of nodes can have different sets of edges and hence networks, each representing a different relationship. Typical relationships of interest include friendship or capacity to influence.

More sophisticated models can set weights for the edges to represent, for example, strong and weak friendship, but generally the edge exists or does not exist. Also, relationships can be directional (that is, A is a friend of B does not mean that B is a friend of A) or undirected (such as 'work together'). Directed relationships are denoted with an arrow.

While the network model is qualitative in the sense that it describes only the existence of a particular relationship, many of the analysis techniques are quantitative and enable comparisons between individuals or groups within the network. Typical applications identify clusters within the network, highly influential individuals, or locations within the network where transmission could be stopped.

A typical application is shown in Figure 9, which depicts a high school friendship network, with nodes coloured by race and year groups indicated by shape of the node (Moody, 2001). Race segregation is clearly visible and, in contrast, different age groups mix well.

Figure 9: US high school friendship network demonstrating same race preference but mixing of year groups (Moody, 2001, Figure 1).

Key features

What it answers:	What is the role of social structure in the system?
What is highlighted:	Connections between individuals that make up a system.
Inputs:	The pairs of individuals with the relationship of interest.
Outputs:	Documentation of relationship structure.
Relationships:	Specified in the definition of the network of interest.
Traps:	Most relationships are strong/weak rather than yes/no but the analysis techniques are less able to deal with weighted edges.
Handling uncertainty:	Generally minimal. Can construct sets of randomly generated networks with a small number of specific features to assess the statistical significance of some features of the specific network.
Data needs:	Whether the relationship exists for every pair of nodes.

Resources

References:	Newman 2003, The structure and function of complex networks
Software:	Pajek (http://pajek.imfm.si/doku.php?id=pajek) UCINet by Analytical Technologies (http://www.analytictech.com/ucinet)

REFERENCES

Epstein, J. M. & Axtell, R. 1996, Growing Artificial Societies, Brookings Institution Press, Washington, DC.

Epstein, J. M. 2008, 'Why model?', *Journal of Artificial Societies and Social Simulation*, vol. 11, no. 4, p. 12.

Gilbert, N. & Troitzsch, K. G. 1999, Simulation for the social scientist, Open University Press, Buckingham.

Gilbert, N. 2008, Agent-Based Models, Sage Publications, Los Angeles.

Hensel 2001, Conway's Game of Life Applet, available at www.ibiblio.org/lifepatterns.

Li, J. & Wilenksy, U. 2009, NetLogo Sugarscape 3 Wealth Distribution Model, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, available at http://ccl.northwestern.edu/netlogo/models/Sugarscape3WealthDistribution.

Moody, J. 2001, 'Race, school integration, and friendship segregation in America', *The American Journal of Sociology*, vol. 107, no. 3, pp 679-716.

Newman, M. E. J. 2003, 'The structure and function of complex networks', *SIAM Review*, vol. 45, no. 2, pp. 167-256.

Rendas, A. B., Fonseca, M. & Pinto, P. R. 2006, 'Toward meaningful learning in undergraduate medical education using concept maps in a PBL pathophysiology course', *Advances in Physiology Education*, vol. 30, no. 1, pp 23-39.

Roberts, N. H., Andersen, D. F., Deal, R. M., Garet, M. & Shaffer, W. A. 1983, Introduction to Computer Simulation: The System Dynamics Modeling Approach, Addison-Wesley, Reading, MA.

Sterman, J. D. 2000, Business Dynamics: Systems Thinking and Modeling for a Complex World, Irwin/McGraw-Hill, Boston.

Wilenksy, U. 1999, NetLogo, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, available at http://ccl.northwestern.edu/netlogo.

CONTACT

Professor Gabriele Bammer, National Centre for Epidemiology and Population Health, ANU College of Medicine, Biology and Environment, The Australian National University, Canberra ACT 0200, Australia.

P: +61 2 6125 0716

E: Gabriele.Bammer@anu.edu.au

CITATION

Badham, J. (2010) A compendium of modelling techniques. Integration Insights #12, May. Available at http://i2s.anu.edu.au.

COPYRIGHT

PUBLISHER

The National Centre for Epidemiology and Population Health, ANU College of Medicine, Biology and Environment, The Australian National University, Canberra ACT 0200, Australia.

THANKS

The production of this Integration Insight was funded by the ARC Centre of Excellence in Policing and Security and the Colonial Foundation Trust through the Drug Policy Modelling Program (DPMP).

Ros Hales provided valuable assistance with formatting.

PREVIOUS ISSUES

Bammer, G. (2006) A systematic approach to integration in research. Integration Insights #1, September. Available at http://i2s.anu.edu.au.

Bammer, G. (2006) Illustrating a systematic approach to explain integration in research – the case of the World Commission on Dams. Integration Insights #2, October. Available at http://i2s.anu.edu.au.

Bammer, G. (2006) Principled negotiation – a method for integrating interests. Integration Insights #3, November. Available at http://i2s.anu.edu.au.

Bammer, G., McDonald, D., Deane, P. (2007) Dialogue methods for research integration. Integration Insights #4, May. Available at http://i2s.anu.edu.au.

Bammer, G. (2007) Key concepts underpinning research integration. Integration Insights #5, May. Available at http://i2s.anu.edu.au.

Bammer, G. (2008) The case for a new discipline of Integration and Implementation Sciences (I2S). Integration Insights #6, May. Available at http://i2s.anu.edu.au.

Bammer, G. and Smithson, M. (2008) Understanding uncertainty. Integration Insights #7, May. Available at http://i2s.anu.edu.au.

Smithson, M and Bammer, G. (2008) Uncertainty: Metaphor, Motives and Morals. Integration Insights #8, June. Available at http://i2s.anu.edu.au.

Kasperson, R.E. (2008) Coping with deep uncertainty. Integration Insights #9, June. Available at http://i2s.anu.edu.au.

Bammer, G. (2008) Enhancing Research Collaborations. Integration Insights #10, June. Available at http://i2s.anu.edu.au.

Bammer, G. (Compiler, 2008) Checklists for assessing research-policy interactions. Integration Insights #11, July. Available at http://i2s.anu.edu.au.