Durham University

Synthesis

Computational Social Science

Lecture 10

Overview

Social science deals with complex problems

- Revisit ideas from Lecture 1

Computational social science provides a set of methods that are suited to complex problems

- Combine data and complexity-aware theory

Compare how methods incorporate complexity

- Complexity methods: to statistical models
- Four focus methods: to each other

One thing…

At the end of lecture, I will ask you

- Something(s) you learned in this module that you think is important
- Something(s) you learned in this module that you found interesting

It would be great if you can think of more than one!

Giving you some warning so you can think about it

Particularly important to not read ahead this week

Complex Realism

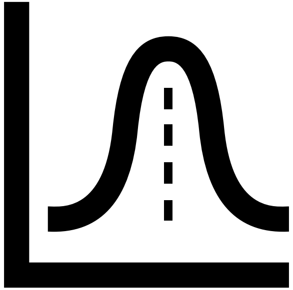
It is a nuisance, but God has chosen to give the easy problems to the physicists.

Lave & March (1993). *An Introduction to Models in the Social Sciences.* University Press of America. Lanham, MD

Weaver's organisation of scientific problems

Phase 1: to 1900

- Physical sciences
  - Problems of simplicity
  - Experimental methods isolate and compare
  - Technological advancement
- Life sciences
  - Classify and describe



Phase 2: to 1948

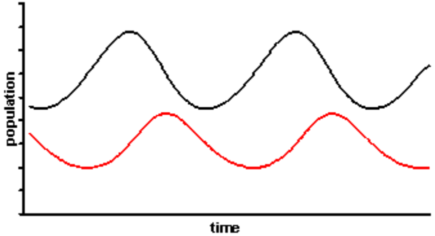
- Problems of disorganised complexity
- Statistical methods effective as average meaningful

Phase 3: after 1948

- Organised complexity: *sizeable number of factors which are interrelated into an organic whole*
- Common in social and life sciences

Require

- Computing power
- Multidisciplinary teams



Durham University

Weaver (1948). "Science and Complexity". *American Scientist*, 36(4): 536-44

Current social problems less amenable to traditional methods

Global population overload

Global warming and climate change

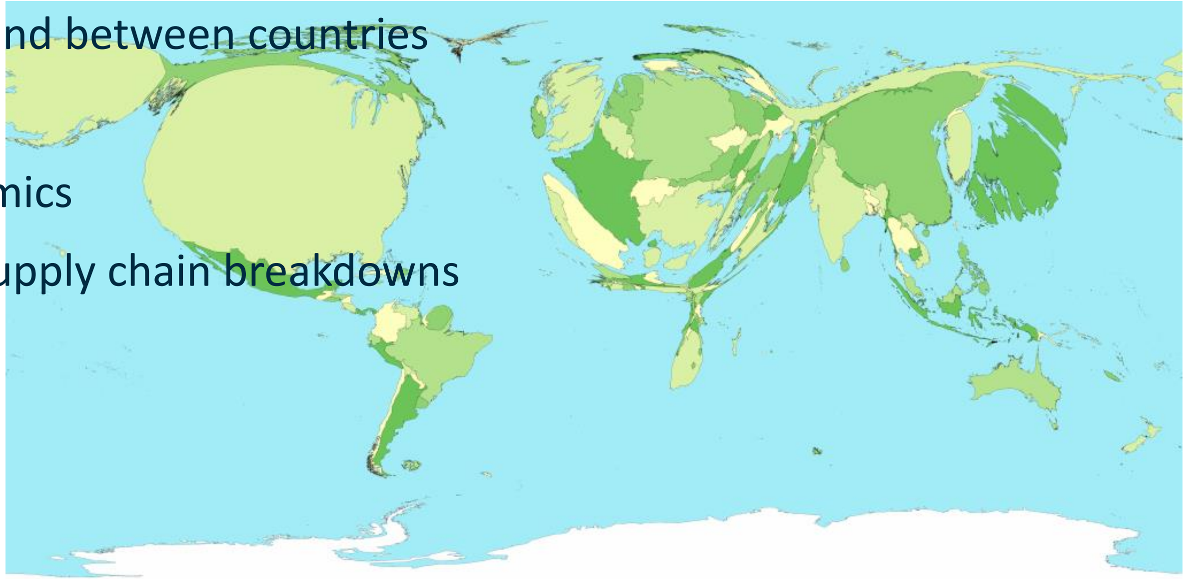
Increasing ecological challenges – access to water, food

Entrenched inequality, within and between countries

Cultural conflict and terrorism

Fast moving epidemics, pandemics

Fragile infrastructure such as supply chain breakdowns



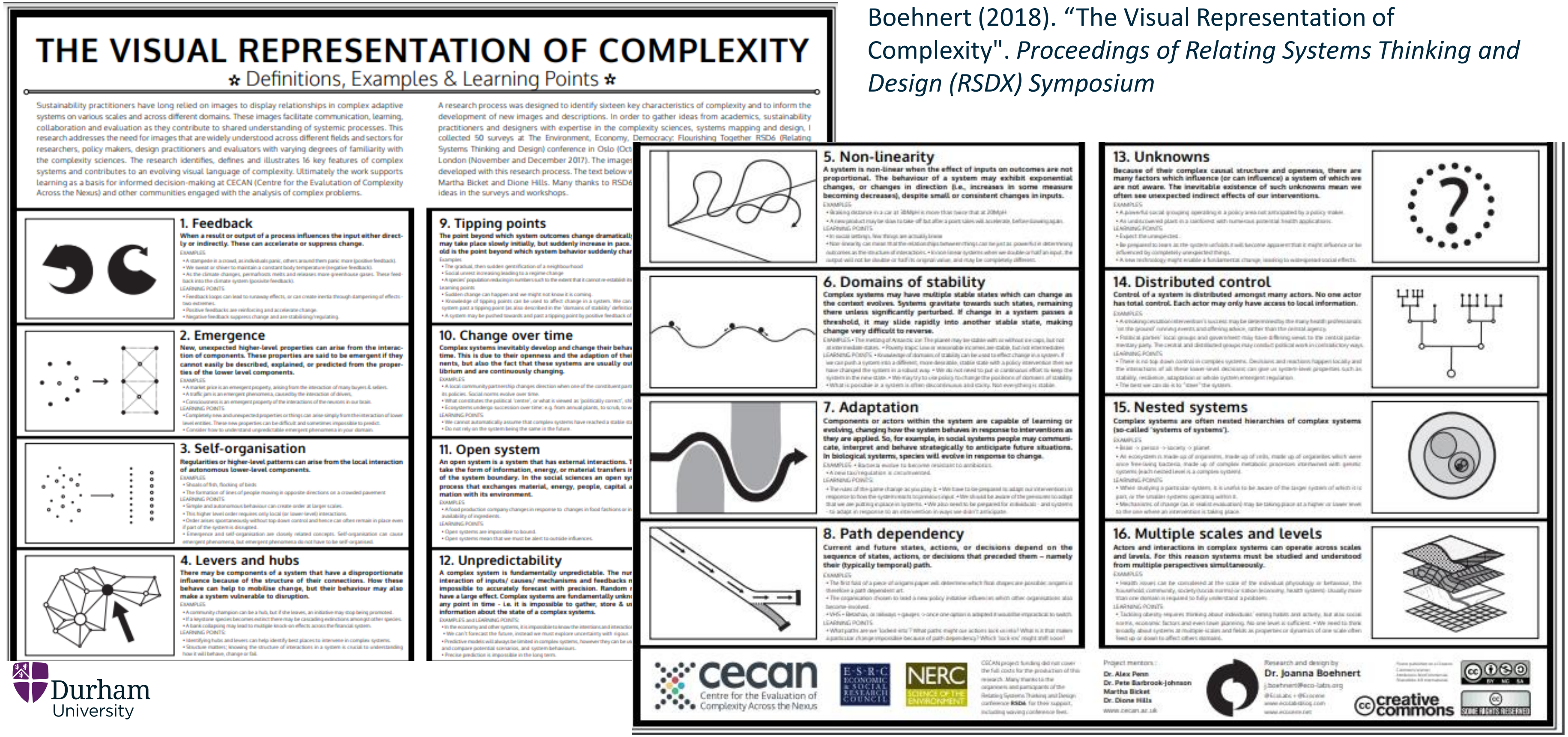
Durham University

Social problems are qualitatively different: complex

A system is complex if the ***system's behaviour of interest is driven by the interactions*** between the individual parts, not simply the behaviour of the parts independently.

Something ‘more’ than having many parts that are linked in complicated ways

Interaction at the core of complexity gives rise to characteristic properties



Boehnert (2018). “The Visual Representation of Complexity". *Proceedings of Relating Systems Thinking and Design (RSDX) Symposium*

Durham University

Complex systems involve Collective Behaviour

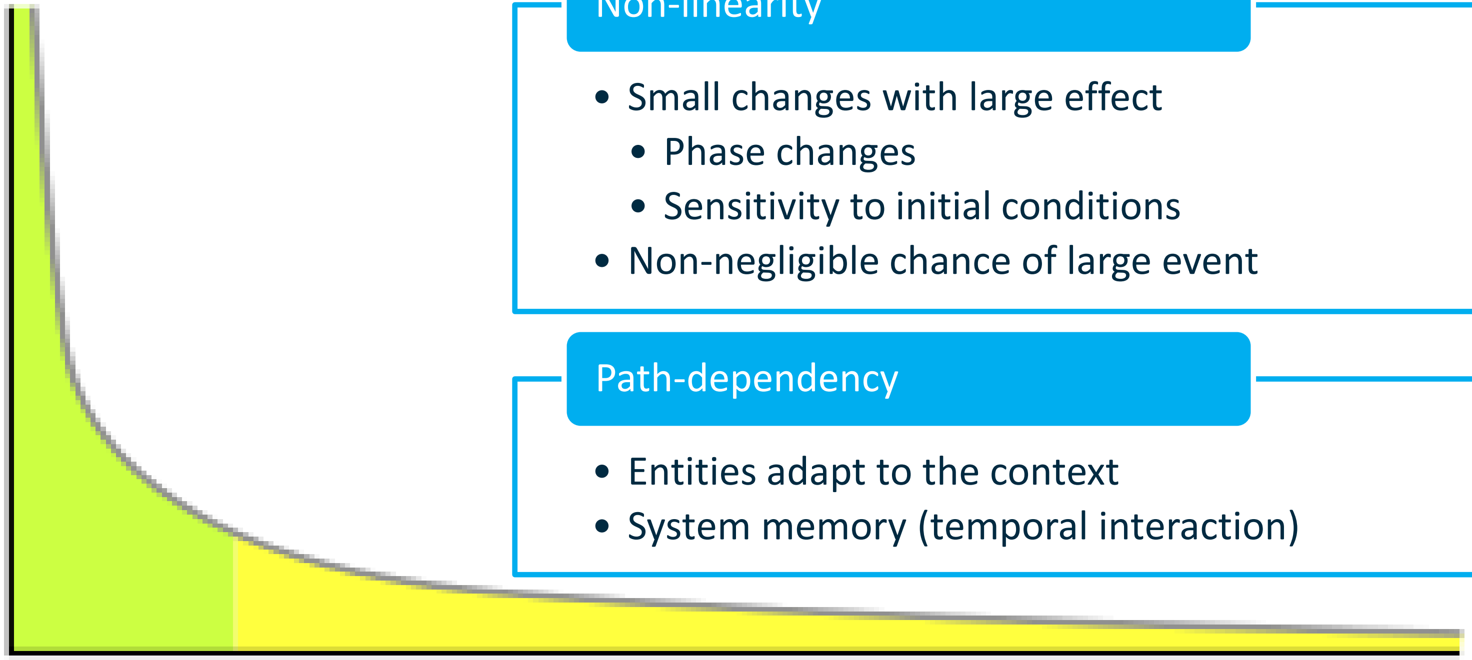
Emergence

- Formal term for the system having properties that the parts do not have

Self-organisation

- The system exhibits order
- Parts (entities) are autonomous, without central control

Complex Systems involve Feedback



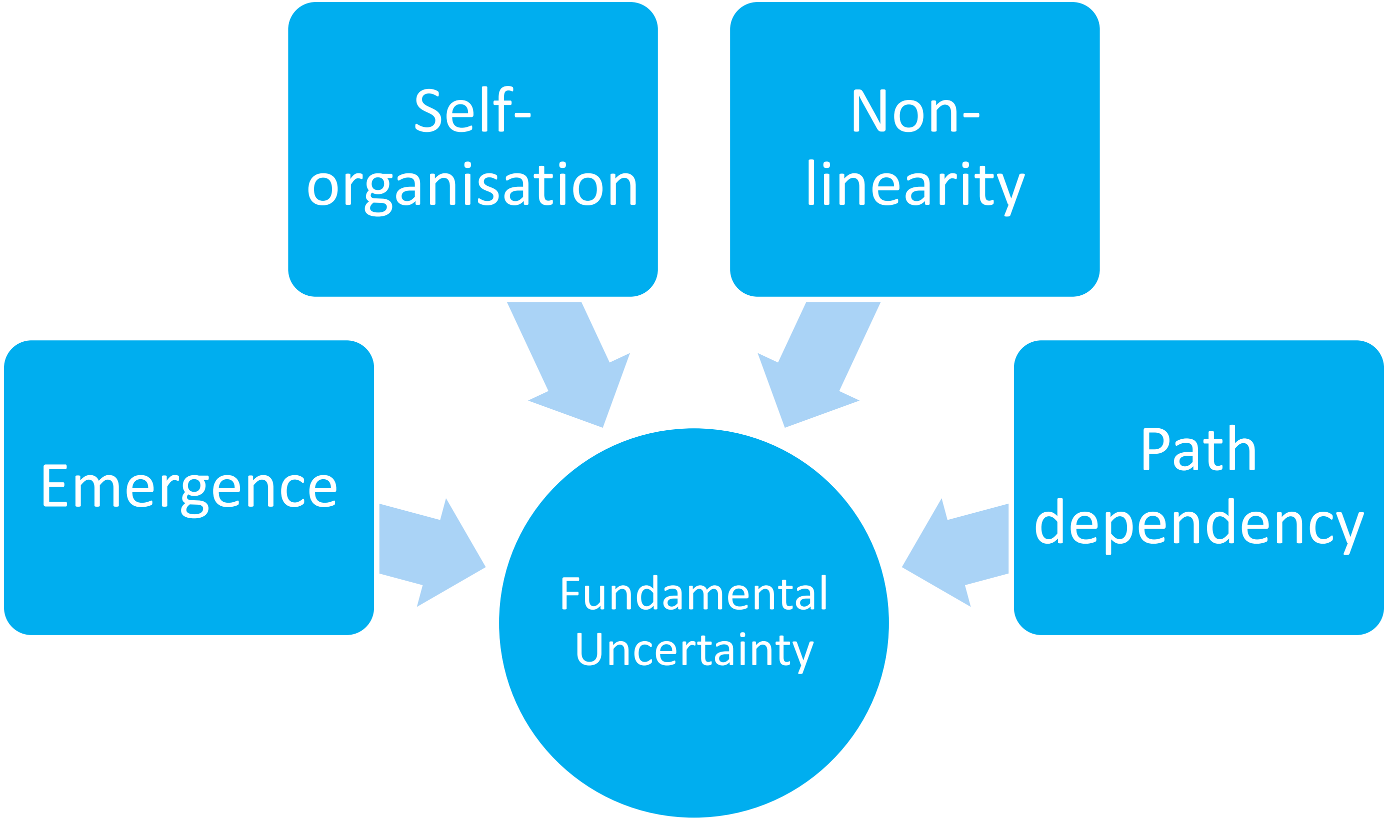
Non-linearity

- Small changes with large effect
  - Phase changes
  - Sensitivity to initial conditions
- Non-negligible chance of large event

Path-dependency

- Entities adapt to the context
- System memory (temporal interaction)

Complex systems difficult to understand and predict



Durham University

Positivism and Empiricism

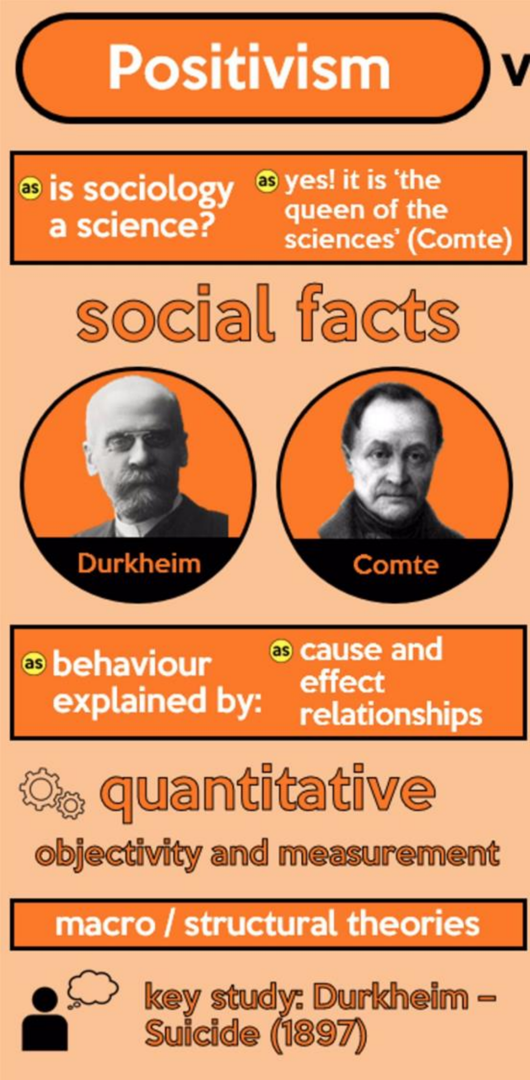
The role of social science is to uncover the rules that govern the social world

- Emulates natural sciences with proof, methods
- Experience, observation are only source of knowledge

These rules describe the way in which the macro-structure and norms of society lead to the behaviour of people

- Attributes of individuals (eg gender) expose them to different norms

Positivists (unlike strict empiricists) allow logic to supplement observation



Durham University

Twitter @allsociology (11 Feb 2020)
https://twitter.com/allsociology/status/1227133621984792576

Interpretivism and Constructivism

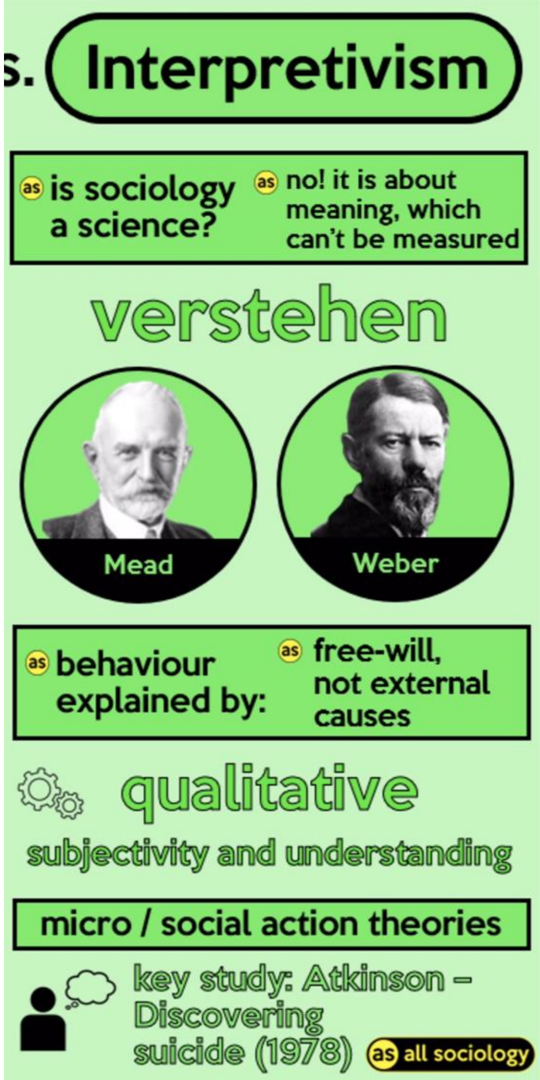
Multiple realities

- Individuals experience and understand the same event differently

Interpretivism: How are these realities experienced and perceived?

- People are experts in their own reality

Constructivism: How are realities constructed through interaction with others?



Durham University

Twitter @allsociology (11 Feb 2020)
https://twitter.com/allsociology/status/1227133621984792576

Complex Realism

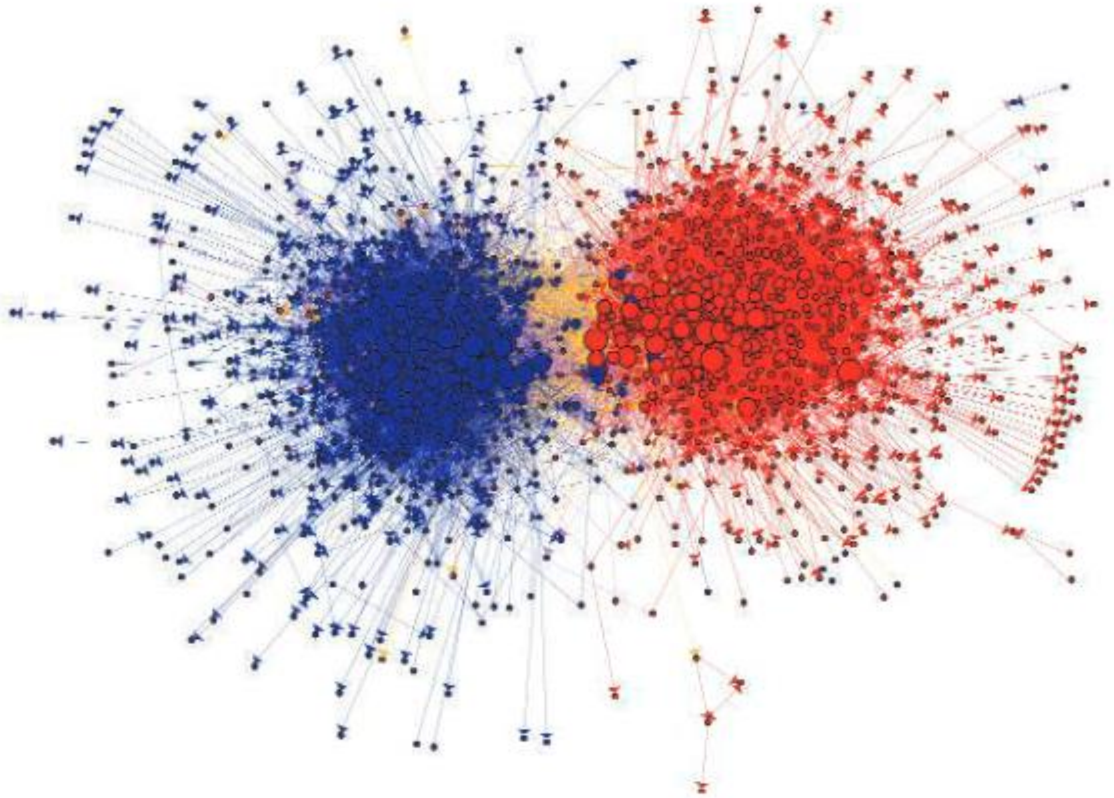
Realism lies between positivism and interpretivism

- There is a real social world to measure and understand that exists independently of our minds (Positivism)
  - Otherwise could not get things wrong (Sayer 2000)
- Actions have an effect on society (Constructivism)
- Combines quantitative and qualitative methods

That real social world is complex

- Social structures emergent property of social relations
- Social systems are open and evolve

What is computational social science?



Lazer et al (2009), Computational Social Science
doi: 10.1126/science.1167742

Competing definitions of CSS: Lazer et al

Lazer et al (2009), Computational Social Science
doi: 10.1126/science.1167742

Substantial social data available

- Communication networks
- Social media
- Contents of searches
- Virtual worlds for experiments

Why not (yet) transformed social science

- Privacy concerns
- Limited resources
- Interdisciplinary: training, academic careers, journals

SOCIAL SCIENCE

Computational Social Science

A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit

ment agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of aca-

critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environ-

Competing definitions of CSS: Conte et al

Conte et al (2012), Manifesto of computational social science. doi: 10.1140/epjst/e2012-01697-8

Formidable social problems

Digital and data are changing society

Generative explanations combine theoretically based models with data

- Primarily about ABM
- Many scientific challenges
  - How much detail in agents?

Manifesto of computational social science

R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J. -P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel & D. Helbing

The European Physical Journal Special Topics **214**, 325–346 (2012) | Cite this article

Abstract

The increasing integration of technology into our lives has created unprecedented volumes of data on society's everyday behaviour. Such data opens up exciting new opportunities to work towards a quantitative understanding of our complex social systems, within the realms of a new discipline known as Computational Social Science. Against a background of financial crises, riots and international epidemics, the urgent need for a greater comprehension of the complexity of our interconnected global society and an ability to apply such insights in policy decisions is clear. This manifesto outlines the objectives of this new scientific direction, considering the challenges involved in it, and the extensive impact on science, technology and society that the success of this endeavour is likely to bring about.

Much in common…

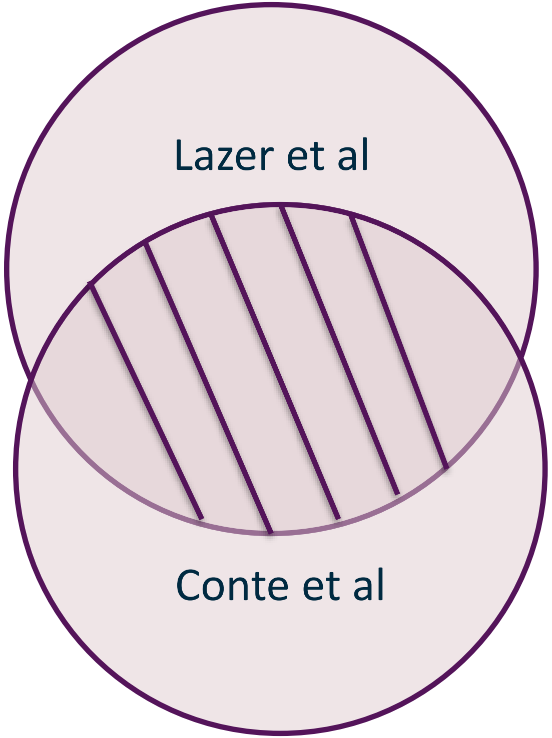
Potential of big data for social science

- Digital traces of individual and group behaviour
- Enormous scale of the available data
- Similar data sources

Opportunity to apply quantitative methods to understand society

- Data science provides methods

Interdisciplinary: social science and data science



Durham University

… but fundamentally different

Lazer et al: Applying data science methods to social data

- Data driven analysis of patterns of behaviour

… leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors

What I think of as social analytics

Conte et al: using computational methods to do social science

- Explicit integration of analysis with theory and complexity

… unprecedented volumes of data … opens up exciting new opportunities to work towards a quantitative understanding of our complex social systems

This module

Durham University

Is someone analysing Facebook friend data doing social science?

If yes, why?

About social interactions and behaviour

Can research phenomena such as diffusion, conformity, opinion change, characteristics of groups of interest

If no, why not?

No meaning defined by friendship

Social science connects to some theory

Analysis is primarily descriptive

Ultimately, depends on the research question: what is the purpose of the analysis?

CSS is quantitative, but different from statistics

2010 ESRC review of UK sociology found

- Weak training in quantitative methods compared to US tradition
- Statistics core to social science

Byrne's commentary

- Agree more quantitative needed
- Review equated that to statistics
- Methods must deal with complexity
  - Statistical methods do not

Sociology
46(1) 13–24

DOI: 10.1177/0038038511419178
soc.sagepub.com
SAGE

UK Sociology and Quantitative Methods: Are We as Weak as They Think? Or Are They Barking up the Wrong Tree?

David Byrne
Durham University, UK

Abstract

This piece responds to the *Benchmarking Review of UK Sociology*'s assertion that the discipline has a deficit in quantitative methods and that the solution involves a recognition that: '… statistical methods form the core of social science.' It argues that whilst a quantitative programme is essential and we can agree that there are problems in relation to the quantitative competencies of sociologists at all levels in the UK, a turn to conventional statistical methods is not the way to go. The argument is developed first in relation to epistemic critiques of those methods by Pawson and Goldthorpe and then by the outlining of an alternative founded in a synthesis of complexity and systematic comparison. The key issue is that we need a quantitative programme which actually corresponds to social reality and that is not to be found in statistical methods which reify variables and consider causality in linear terms.

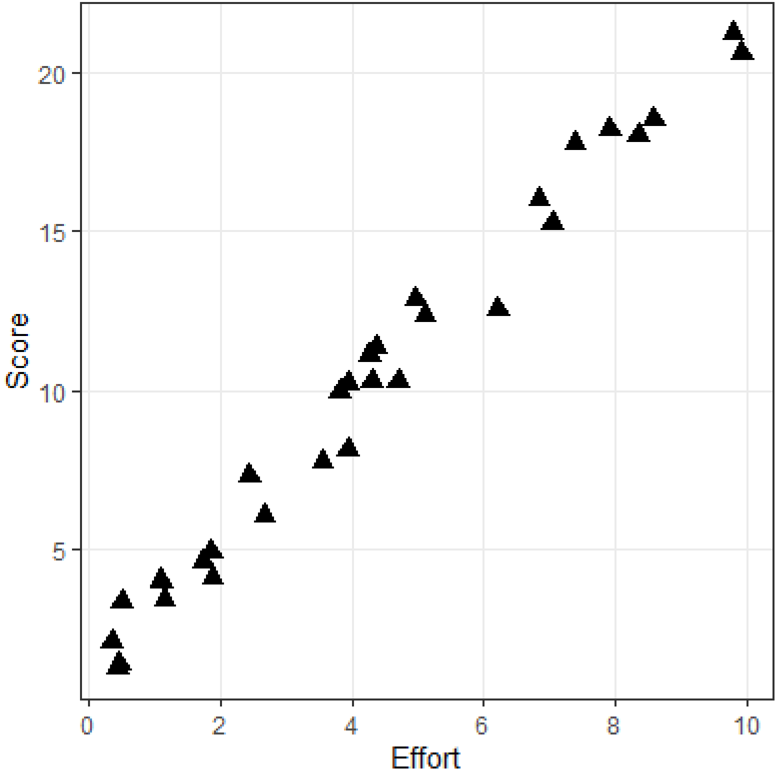
Durham University

Explanation with linear statistical modelling

How would you summarise the attributes of these cases?

- People who put in only a small amount of effort achieve a low score, and those who put in a large effort achieve a high score
- Each additional amount of effort is associated with the same increase in score

Linear: constant increase regardless of effort



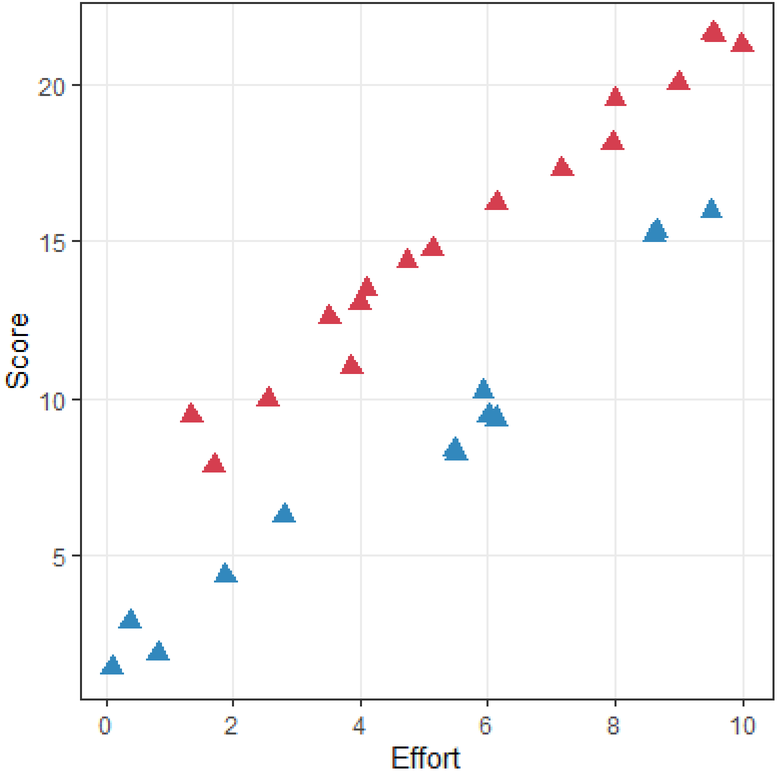
Durham University

Explanation with multiple relevant attributes

How would you summarise the attributes of these cases?

- As before, score related to effort
- Additional factor indicated by colour
  - Those marked as red have consistently higher scores for same effort

Additive and independent: additional score for red is constant regardless of effort

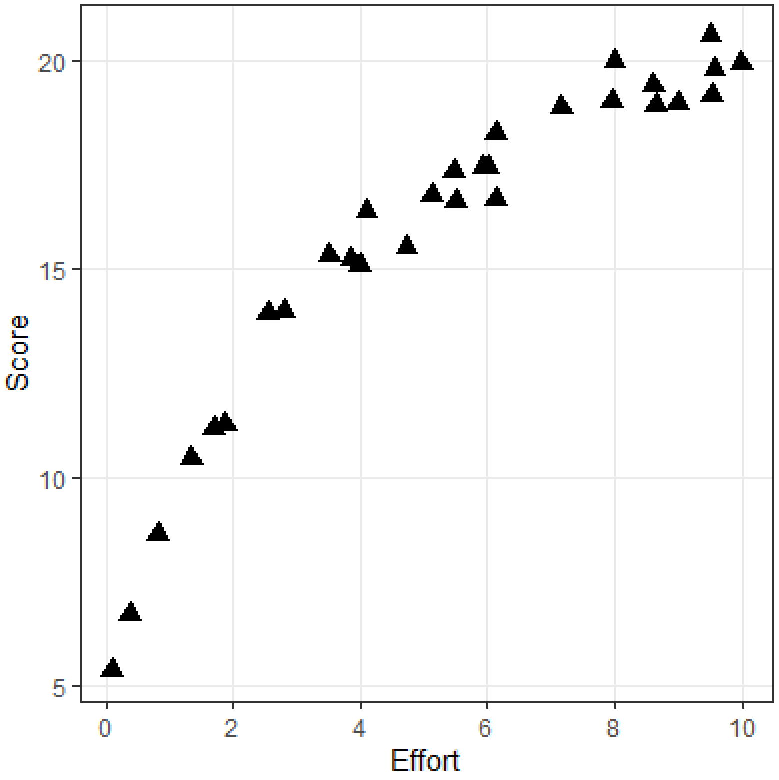


Explanation with nonlinear statistical modelling

How would you summarise the attributes of these cases?

- Score increases with effort
- After a certain amount of effort, each additional amount of effort has smaller impact

More complicated mathematical relationships are possible, but all have a single equation to set out how the value of one attribute is estimated by some combination of the values of all others



Durham University

Difference in type of explanation most evident for QCA

Cases that are irrelevant in the logical structure of QCA would reduce the explanatory power of statistical model

Example: (from lecture) of hypothetical austerity protests

- Subset of cases with severe austerity
- Rapid price rise sufficient to protest
- Protests without rapid price rise undermine statistical explanation

Rapid price rises	Protests	
	Yes	No
T	23	0
F	6	7

Durham University

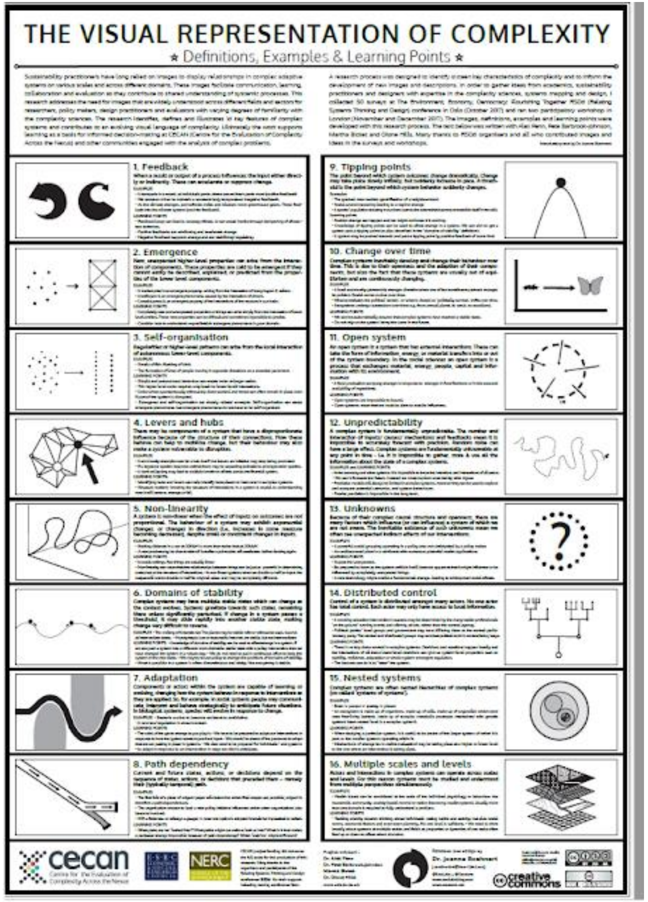
Explanation in complex realism

In Complex Realism effects are

- Complex: involve interactions between multiple factors
- Emergent: cannot be reduced to individual factors or elements
- Contingent: function differently depending on context

Cannot be described by some mathematical relationship between values of attributes

Dealing with complexity



Boehnert (2018). “The Visual Representation of Complexity". *Proceedings of Relating Systems Thinking and Design (RSDX) Symposium*
https://rsdsymposium.org/the-visual-representation-of-complexity-definitions-examples-learning-points/

How do the four methods deal with complexity?

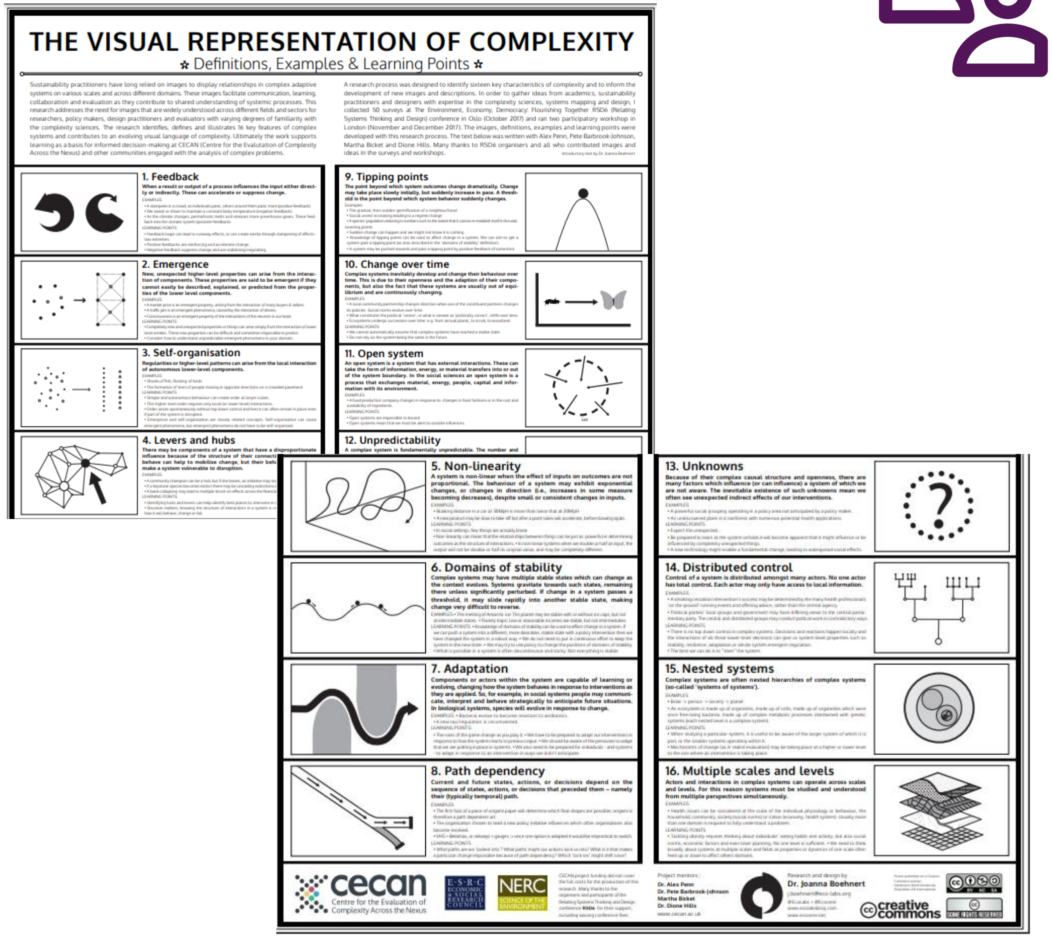
Identify 2 or 3 complex realism aspects for each method

- Clustering (case based)
- QCA (configurational)
- Network analysis
- Simulation

Note

- Also consider interactions
- Characteristics are dynamic, but methods also static

Small groups



Clustering: Key Idea

Find patterns in cases based on their whole set of attributes

Also referred to as case-based complexity

Workshop case study: Criminology

- Patterns of arrests in USA States

Case →

Area Code	Income	Employment	Health						
	People in income deprivation (%)	Working-age people in employment deprivation (%)	GP-recorded chronic condition (rate per 100)	Limiting long-term illness (rate per 100)	Premature death (rate per 100,000)	GP-recorded mental health condition (rate per 100)	Cancer incidence (rate per 100,000)	Low birth weight (live single births less than 2.5kg) (%)	Children aged 4-5 who are obese (%)
	16	10	14.3	22.7	382.4	23.2	611.9	5.5	11.8
Isle of Anglesey	15	10	13.4	20.8	359.6	23.0	601.1	5.5	12.7
Gwynedd	13	8	12.9	19.5	348.9	20.3	604.0	5.0	13.0
Conwy	15	10	12.9	20.6	375.6	23.7	593.6	5.1	11.4
Denbighshire	17	11	14.7	21.8	397.4	28.4	639.3	6.1	12.2
Flintshire	12	8	14.1	19.7	358.1	23.1	647.8	5.4	11.2
Wrexham	15	9	14.3	21.5	393.7	24.3	637.3	6.4	12.4
Powys	11	7	12.8	18.8	309.1	19.0	579.8	4.7	10.5
Ceredigion	12	8	12.7	20.0	322.4	19.9	545.5	4.8	10.5
Pembrokeshire	15	10	13.1	20.5	345.8	22.1	606.1	5.2	12.5
Carmarthenshire	15	11	13.9	23.7	365.5	20.0	602.6	5.4	12.8

Clustering: Steps

Normalise data

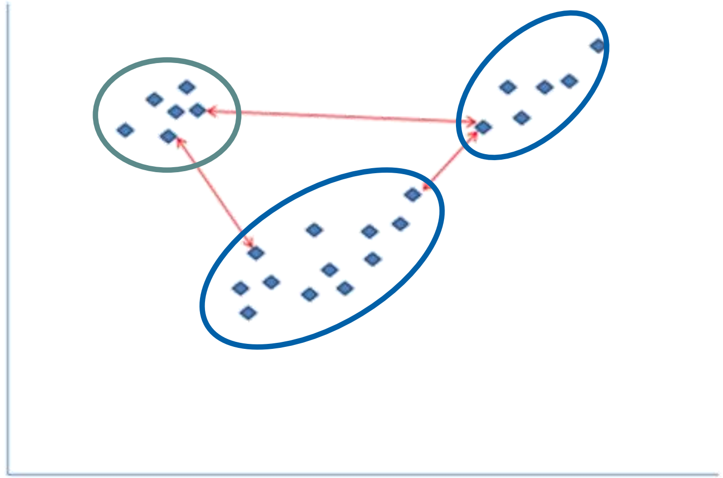
Choose appropriate distance measure

Use a clustering algorithm

- Group close cases into a cluster
- Maximise separation between clusters

Assess whether the clusters are meaningful

- Visualise
- Interpret



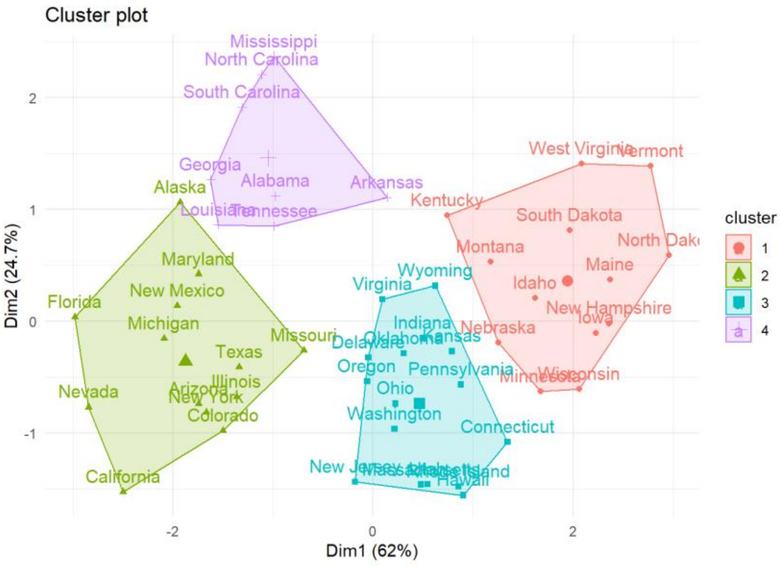
Clustering: Challenges

Only relevant for scale data attributes

- Relies on mapping attributes to coordinates in spatial dimensions

Easy to obtain spurious results

- Algorithms work whether or not structure exists



Clustering: Complexity

Emergence: cases treated as complete entities

- Close on all dimensions, not just some

Self-organisation

- Structure is revealed by the method

Multiple scales and levels

- Small clusters may occur within broader clusters

QCA: Key Idea

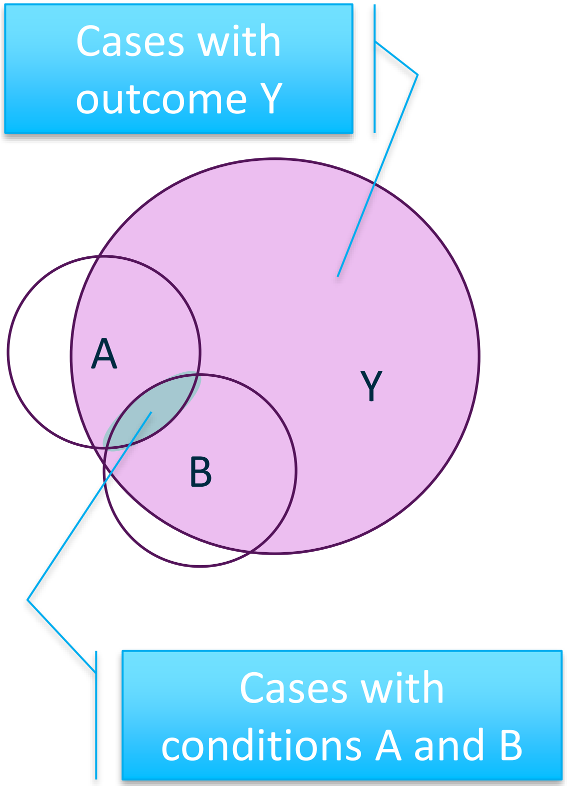
Structured method for case based comparative research

- Select attributes that appear to be relevant
- Compare attribute patterns where the phenomenon occurs and where it doesn't

Also referred to as configurational complexity

Workshop case study: Political Science

- Explain survival/breakdown of democracy



QCA: Steps

Identify social phenomenon and potential contributing conditions

Select moderate number of cases

Calibrate attribute values

Construct truth table

- Resolve contradictions

Simply truth table

- May include configurations without cases

Assess quality of solution

Conditions		Cases by outcome	
A	B	Y	Y
T	T	≥1	0
T	F	any	≥1
F	T	any	≥1
F	F	any	any

QCA: Challenges

Diverse cases are essential

- The logic of comparing sets of truth values relies on many combinations available from the dataset
- Too many remainders limit the opportunity to identify recipes

Operationalising attributes (crisp or fuzzy) relies on judgement and iteration

QCA: Complexity

Emergence: cases treated as composite of their attributes

Non-linearity: Recipes use logical operators

Levers and hubs: Causal recipe is lever for outcome

Recipes respect diversity and heterogeneity of the cases and their contexts

- Adaptation?

Recipes have a close relationship with theory

- Realist perspective, something exists to influence observations

Networks: Key Idea

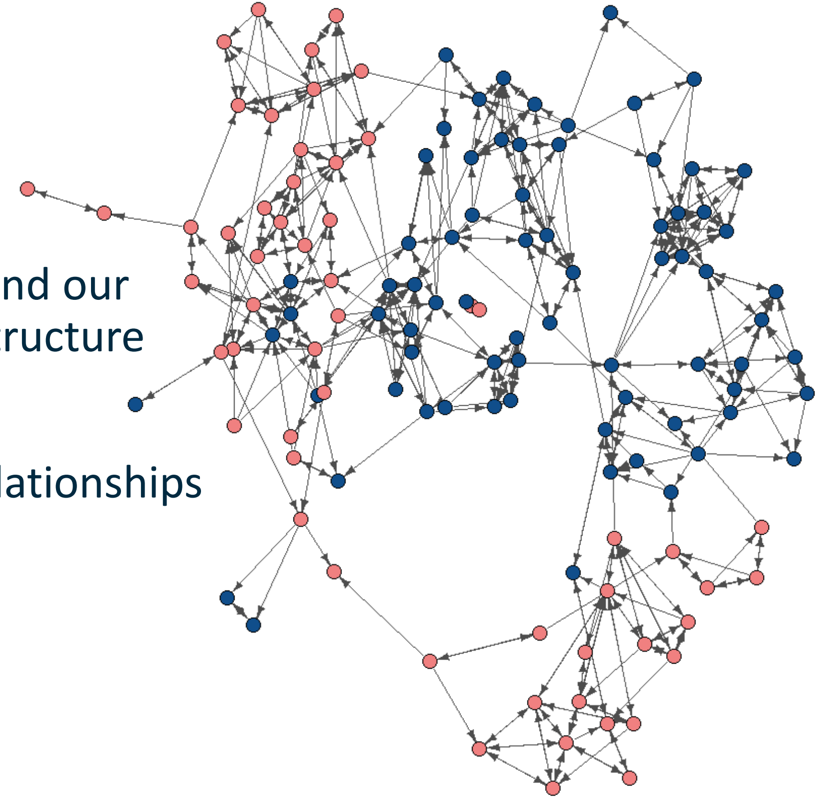
Relationship between cases is focus of analysis

Network analysis provides a set of tools to extend our intuition of the patterns that construct social structure

- Visualise to highlight patterns and structure
- Develop theories about the way in which relationships influence social phenomena and processes
- Empirically test such theories

Workshop case study: Sociology

- Diffusion of family planning practices



Durham University

Networks: Steps

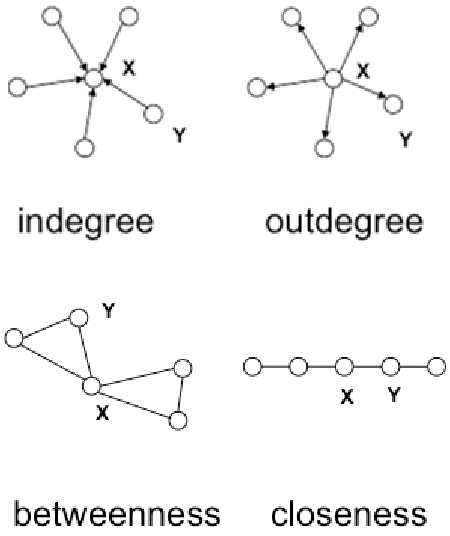
Many different methods with different steps

Collect information about cases and the relationships

- Construct network

Understand the network

- Visualise with attributes and layout
- Calculate centrality and other properties
- Consider communities



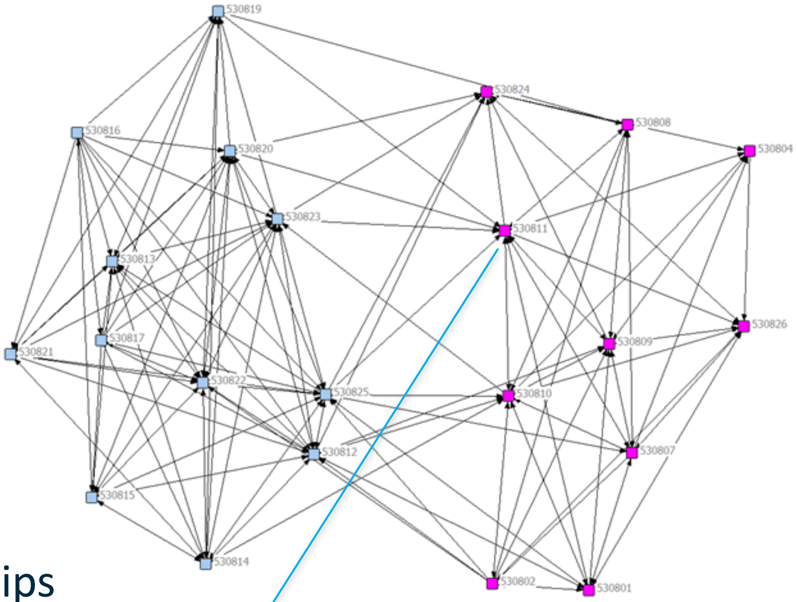
Networks: Challenges

Terminology

- Two (or three) traditions
- Entire discipline, not just one method

Sensitive to data quality issues

- Missing cases/attribute data
- Misattribution of information and relationships
- Retrospection error: recalling interactions



Imagine if this node (and its edges) is missing

Durham University

Network: Complexity

Specifically maps interactions as relationships

Levers and hubs: Influence arises from network position

Non-linearity and tipping points

- Dynamics over networks (eg epidemics)

Path dependency: Which edges are involved in a relationship, not just how many

Multiple scales and levels: Analyse at node, community, whole of network

Simulation (ABM): Key Idea

Modelling some social process

- Justified stories are plausible coherent explanations of a dynamic phenomenon

Agent-centric thinking

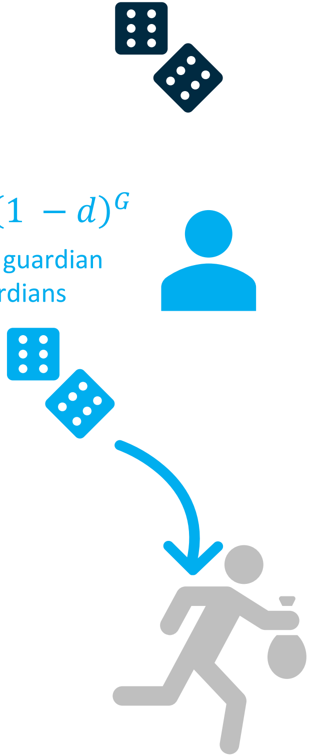
- I, the agent, have certain characteristics and beliefs of my own as well as information about the world around me, and therefore will decide on some action

Workshop case study: Social Psychology

- Conformity and segregation

$$P(deterrence) = 1 - (1 - d)^G$$

d is deterrence per guardian
G is number of guardians



Simulation: Steps

Relevant entities

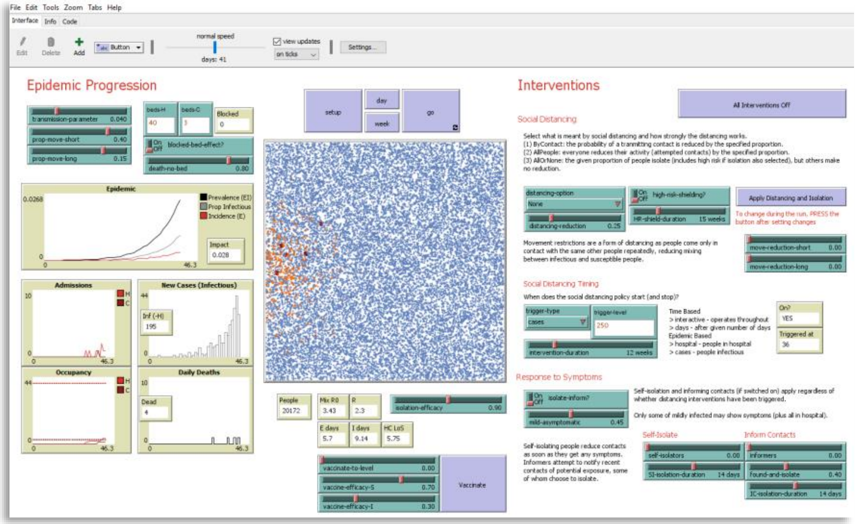
- Agents, Environment, Attributes

Develop rules

- How do agents behave?
- Influences on that behaviour and effect of decision

Build model, testing iteratively

Run scenarios (experiments)



Durham University

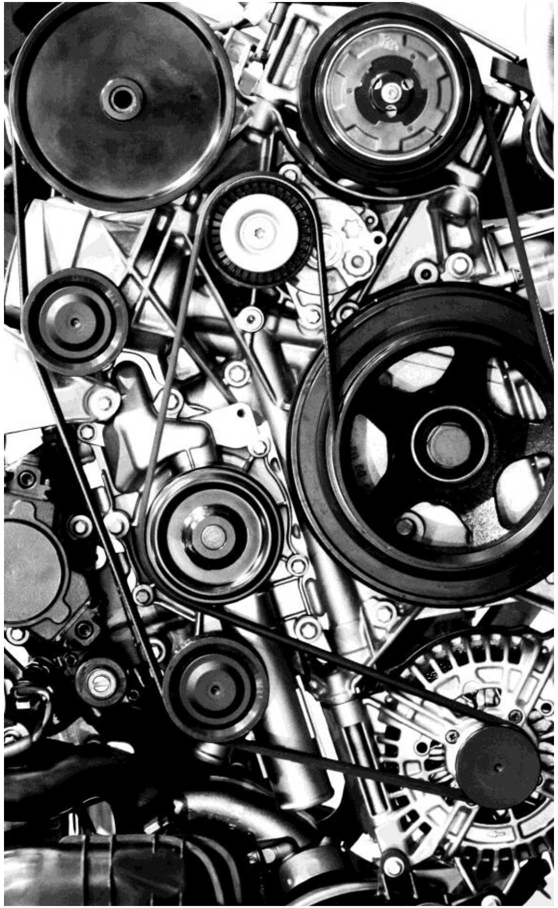
Simulation: Challenges

Process or mechanism orientation

- Different type of thinking about theory and data
- Data not available about mechanism, only outcomes

Balance between simplicity and detail

Programming skills



Simulation: Complexity

Agents interact with each other (social environment) and with their physical environment

- Choice of action affected by environment and selected action has impact on the environment

Dynamic interactions

- Feedback / Tipping Points / Path Dependency
- Change over time / Non-linearity

Distributed control: definition of agent

Theory and method work together

Computational Social Science is truly interdisciplinary

- Mathematical/computer scientists have analytical skills and training to develop methods and work with data
- Social scientists bring ideas, applications and interpretation
- Theory organises the ideas and interpretation

As much as it bears similarities to previous work, however, what is new about the current generation of network-related research is a rapidly emerging and highly interdisciplinary synthesis of new analytical techniques, enormously greater computing power, and an unprecedented volume of empirical data. Sociologists have much to gain from this progress, and also much to contribute. Most of the work discussed in this review is taking place in the mathematical sciences, particularly in physics. That is hardly surprising, as physics and mathematics typically lead the way methodologically, whereas biological and social sciences follow with applications. But in this particular case, the flow of ideas has been bidirectional, with many of the core ideas—not just applications—having come from sociology. Physicists may be marvelous technicians, but they are mediocre sociologists. Thus, if the science of networks is to live up to its early promise, then the other disciplines—sociology in particular—must offer guidance in, for example, the interpretation of empirical and theoretical findings, particularly in the context of policy applications, and also in suggesting measures and models that are increasingly relevant to the important problems at hand.

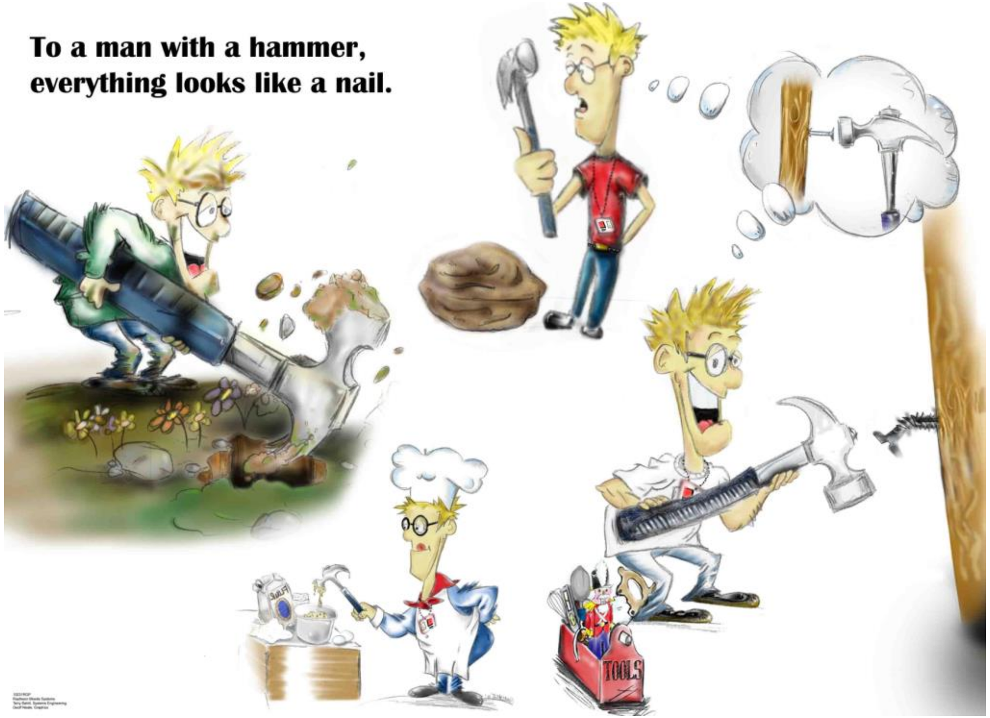
Durham University

Watts (2004), The "New" Science of Networks, *Annual Review of Sociology*. 30:1, 243-270

Question first, then select the method

Each computational method responds to different aspects of complex systems

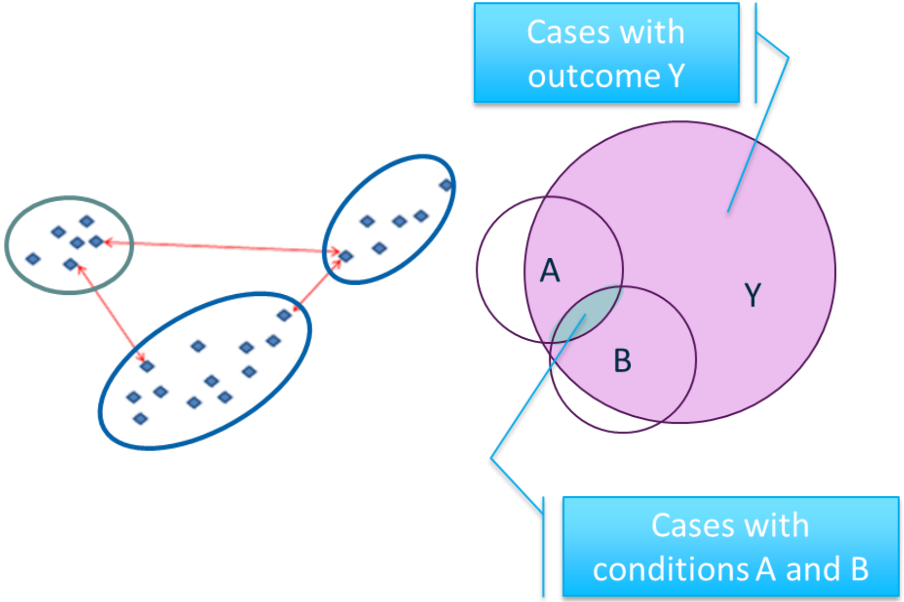
Choosing a method asserts something about the structure of the system being examined and the validity of specific types of questions



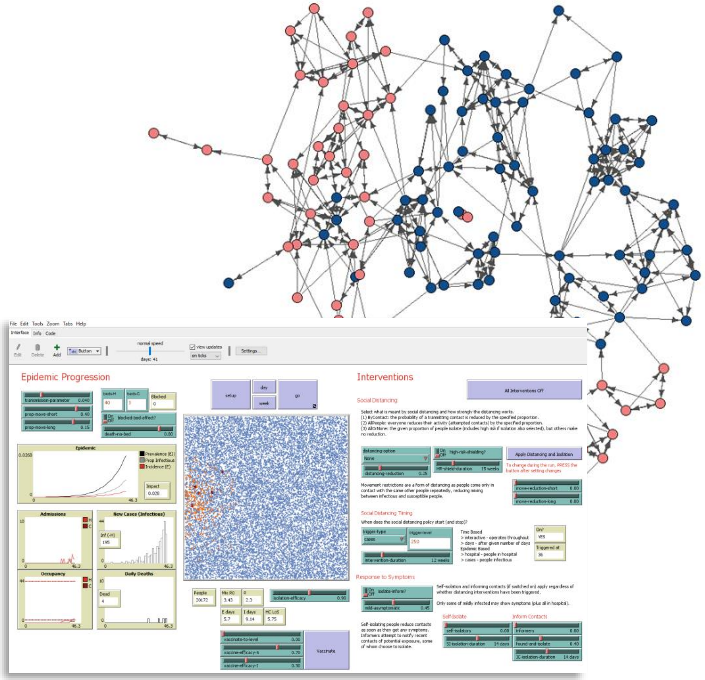
Durham University

Where does the complexity lie (roughly)?

Internal: within cases



External: processes, relationships



Beyond the module

Assessment due 22 April at 13:00

Make sure research question and proposed method(s) match

The method(s) MUST deal with relevant aspects of complexity that you identify

1/ Compare two methods

- What methods could complement each other?

2/ Use one of the methods

- Choose question with relevant dataset / model
- Be guided by what is possible with your dataset or model

Remember the reader

- Explain how the method(s) you choose could be useful, not about answering a research question

Extensions and adjustments have forms

Come and talk to me if unsure

Your reflections

Something(s) you learned in this module that you think is important

Something(s) you learned in this module that you found interesting

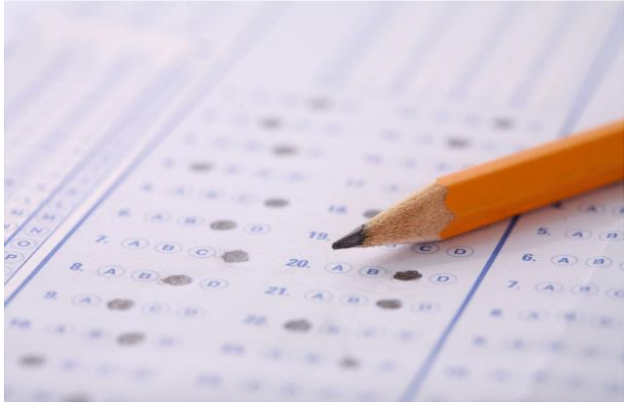
The End

Thank you

- For being a great group to teach
- For taking on something a little different

Please complete the Module Evaluation Questionnaire when you are asked

- I don’t find out what you think of the module unless lots of you fill it in



Durham University