Routledge
Taylor & Francis Group

International Journal of Social Research Methodology

ISSN: 1364-5579 (Print) 1464-5300 (Online) Journal homepage: www.tandfonline.com/journals/tsrm20

Case-based systems mapping: advancing a multimethod approach to social complexity

Brian Castellani, Corey Schimpf, Jonathan Wistow, Christopher Caden, Jutshi Agarwal & Peter Barbrook-Johnson

INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY
https://doi.org/10.1080/13645579.2025.2564177

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Case-based systems mapping: advancing a multimethod approach to social complexity

Brian Castellani[a], Corey Schimpf[b], Jonathan Wistow[a], Christopher Caden[a], Jutshi Agarwal[b] and Peter Barbrook-Johnson[c,d]

[a]Department of Sociology and Durham Research Methods Centre, Durham University, Durham, UK; [b]Department of Engineering Education, School of Engineering and Applied Sciences, University at Buffalo, State University of New York, New York, USA; [c]Institute for New Economic Thinking, University of Oxford, Oxford, UK; [d]Department of Arts & Sciences, University College London, London, UK

ABSTRACT
In this work, we introduce case-based systems mapping (CBSM), an integration of exploratory systems mapping and case-based complexity. CBSM explores different groups of cases and their distinct configurations and then generates submaps that display the causal flows for these different groups. Combining these allows users to advance on systems mapping by allowing them to (1) explore how the causal factors within the map may cluster based on groups of cases and (2) engage in greater data corroboration of exploratory systems maps to support, clarify, or further develop ideas. CBSM involves four steps: (1) generating an initial exploratory systems mapping; (2) identifying possible case-based clustering; (3) data corroboration via clustering, machine learning and network analysis; and (4) assessment. As a demonstration, we apply CBSM to a dataset on the social determinants of teenage pregnancy in 100 Local Authorities in England. We conclude with future directions for research.

ARTICLE HISTORY
Received 19 April 2025
Accepted 15 September 2025

KEYWORDS
Systems mapping; case-based complexity; multi-methods; configurational social science; methods platform

1. Introduction

This paper introduces readers to *case-based systems mapping*, a new component of the existing COMPLEX-IT platform[1] (Schimpf & Castellani, 2020). This new addition to COMPLEX-IT allows users to explore and understand the data they have for a complex system by integrating systems mapping techniques with case-based methods, cluster analysis, machine learning and social network analysis. Case-based systems mapping (CBSM) is part of the suite of methods called case-based complexity – a methodological field of study for modelling and exploring complex causality in systems by emphasising cases as the unit of analysis and by focussing on the configuration of characteristics or variables within a case as a whole, rather than on individual variables. Case-based complexity methods vary considerably from qualitative comparative analysis (QCA) to machine learning (Castellani & Gerrits, 2024).

CONTACT Brian Castellani brian.c.castellani@durham.ac.uk Department of Sociology and Durham Research Methods Centre, Durham University, Durham, UK

This article was originally published with errors, which have now been corrected in the online version. Please see the Correction (http://dx.doi.org/10.1080/13645579.2025.2577539).

As with any method, systems mapping has its limitations and potentials. For us, the main limitation is its inability to address the role of multiple cases and their places and their resulting differential outcomes in a map. Given this multi-method configurational approach, CBSM overcomes this limitation by allowing users to explore how the causal flows through a systems map impact an outcome, based on different underlying clusters. These case-based clusters of causal flows are called *case-based submaps*. Cases and their places, e.g. be they people, communities, organisations, are key to this approach, as they serve as the foundation for analysing and mapping the relationships between various components in a system map. The potential of systems mapping, which is gaining interest (e.g. Crielaard et al., 2024), is the ability to further explore or corroborate an initial exploratory systems map through additional data-driven mapping techniques. Through the use of cluster analysis, machine learning and network analysis, CBSM (particularly when employed through COMPLEX-IT) offers users an easy workflow to do this.

As Shown in Table 1, CBSM combines four core steps. Step 1 Exploratory systems mapping; Step 2 Case-based clustering; Steps 3, which focus on data corroboration, including Step 3A (cluster analysis) and Step 3B (network analysis); and Step 4, Assessment. As a demonstration, we will introduce readers to the CBSM tab in COMPLEX-IT. Our case study is a subsample of a dataset on multiple deprivation and health inequalities in England, UK, and their impact on teenage pregnancy rates. It is important to note that COMPLEX-IT's systems mapping tab is not the only way to employ CBSM. Readers can use a combination of exploratory systems mapping methods, clustering techniques and network methods, including a more qualitative approach to the latter two methods.

Our paper is organised as follows: We begin with an introduction to systems mapping. We then quickly review the limitations and potential of systems mapping that we seek to advance, turning from there to an in-depth review of CBSM. Having outlined our approach and its advance on systems mapping, we turn to our real-world case study and the COMPLEX-IT software package to walk readers through how CBMS actually works. We end with future directions for research.

2. Systems mapping: limits and potentials

Arriving at useful and valid models of social systems (e.g. organisations, urban planning and public policy) has become progressively more challenging and time-consuming for many applied researchers, civil servants, governments and public sector organisations. This challenge is due, in large measure, to the increasing complexity of the worlds in which we live and the difficulty of finding methods that can model such complexity (Castellani & Gerrits, 2024).

In response, over the last several years, systems mapping has received renewed interest. This renewed interest is due, in part, to the value of systems mapping for exploratory research, improving collective understanding of complex issues, visualising complex causality and helping to advance a multimethod approach grounded in the thinking of computational modeling, e.g. system dynamics models, network analysis and agent-based modelling (e.g. Crielaard et al., 2024; Hong et al., 2022).

Table 1. The four steps of case-based systems mapping.

Step	Description	Inputs	Outputs
1. Exploratory Systems Mapping	Create a visual map of the issue or outcome you want to understand. This map helps illustrate key components, relationships and factors relevant to the problem.	Expert and stakeholders' knowledge and opinions	*System map based on mental models of experts and stakeholders
2. Case-Based Clustering	Identify causal flows in the map by focusing on potential or known differences between cases and the groups they form in relation to the outcome of concern. These are called case-based configurational submaps, or case-based submaps for short.	Expert and stakeholders' knowledge and opinions	*Potential case-based submaps, based on ideas about how different clusters of cases representing different causal flows in the map.
3. Data Corroboration	**STEP 3A**: Apply data-driven cluster analysis and machine learning to explore and confirm the relationships and structures in the main system map and case-based submaps. **STEP 3B**: Use network analysis to further explore and validate the main system map and the identified case-based submaps.	Existing data for the factors and outcomes in the initial systems map	*Data-driven Case-based submaps. *Data-driven systems map to compare to initial map.
4. Assessment	Review the results, confirm or adjust the original system map and case-based submaps as needed, and consider what additional data or analysis methods might be required.	All of the outputs above	Iterated analysis New analysis

While variations in approach exist, systems mapping, at its core, involves individuals and groups creating a variety of depictions (mostly visual) of the systems in which they live, work, or study, including key factors such as subsystems, network linkages and feedback loops (Williams & Hummelbrunner, 2011). These linkages can be causal or not. Whatever the approach, a core function of systems mapping is to help people make sense of a system for the purposes of doing something, including identifying levers and barriers to change and potential power conflicts and tensions. For example, systems maps can unveil the dense, sparse, or unexpected ways in which components of a system or systems are interconnected and their possible causal paths. They may also reveal subnetworks or subsystems within the full map or ways in which feedback across components may happen (Barbrook-Johnson & Penn, 2022). Approaches range from Bayesian belief networks, causal loop diagrams to system dynamics and participatory systems mapping.

As stated in our introduction, systems mapping has its own limits and potentials. Our goal is to review the two that CBSM was created to address.

Limitation: multiple cases, places and differential outcomes

The role of multiple cases and their places are not, admittedly, a focus of systems mapping, given its emphasis on factors and their network-like causal linkages. While a system map can identify key agents and can be rather ethnographic in mapping a case, say an ecosystem, organisation or town, it does not tend to think of the map in terms of how it plays itself out for different clusters of people in particular places, such that these configurational variations of the map for each group can lead to differences in outcome (Castellani & Gerrits, 2024).

To understand why this is a limitation, we turn to three concepts from the case-based complexity literature – causal asymmetry, equifinality and multifinality (Rihoux & Ragin, 2009). All three are key to effectively illustrating the importance of considering these issues when building a system map. *Causal asymmetry* is the idea that, given differences in cases and context, the configuration of causal conditions that lead to some outcome may be very different from the configuration of conditions that leads to the absence of that outcome. *Equifinality* concerns those instances where different configurations of conditions lead to similar outcomes. *Multifinality* is the opposite of equifinality. It shows how similar configurations of causal conditions can lead to different outcomes. Applying these concepts to CBSM, *causal asymmetry* means that, for different clusters of cases and in a systems map, the factors that explain the outcomes of one group may differ from the explanatory factors of another. For example, in a city struggling with obesity amongst its adolescents, the configuration of factors or causal flows in a systems map contributing to high levels of obesity in a poor area can significantly differ from the factors in the systems map attributing to low levels of obesity in a wealthier area. *Equifinality* means that for different clusters of cases in a systems map, given different configurations of factors, can lead to the same outcome; or, in terms of *multifinality*, despite having similar configurations, the outcomes for different clusters of cases within a map may be different.

Potential: corroborating exploratory maps with additional data analyses

Exploratory systems mapping (as distinct from more formal approaches, such as system dynamics) is a powerful gateway method for helping stakeholders map a complex system in a relatively quick and straightforward manner. Inevitably, however, a key next step for many stakeholders is to determine the relative *utility* of the map, particularly when used for making policy recommendations. In such instances, corroboration with data is needed. We chose the word 'utility' over 'validity' for a specific reason. We believe that the best approach to corroboration of a systems map is *Bayesian*. As we detail below, we draw on Bayesian approaches to emphasise the distinct insights different actors or groups may have.

We align ourselves with those approaches to exploratory systems mapping that see it as qualitative and data exploratory, particularly the work of Peter Checkland on Soft Systems Methodology (Checkland & Poulter, 2020), as opposed to statistical and data confirming. From this perspective, exploratory systems mapping creates mental maps of a topic based on engagement with stakeholders around a complex social system of concern. These mental maps are designed to help achieve some degree of consensus or common ground on how best to conceptualise the key factors and causal flows (sub-maps) in a system. By definition, then, they are based on the group's knowledge and the data upon which it is based, not the collection of numerical data per se, be it of quantitative or qualitative origin.

Given our alignment, we see any form of corroboration through additional data analysis as an attempt to further engage a systems map to support, verify, clarify, modify or further develop its view of the world. Hence, our Bayesian take on corroboration.[2] Bayesians assign probabilities to our hypotheses about the world, while Frequentists assign probabilities to the data we collect about the world. When applied to exploratory systems mapping, this means that, instead of determining the empirical 'validity' of our map through additional data collection and analysis, we determine its conceptual 'utility' for making sense of a complex system. A Bayesian corroboration of a systems map tests the potential insights of a group's mental model and its related hypotheses about what is going on – with the idea that the potential insights of the model are made explicit in terms of for whom and how. If our additional data analyses suggest we need to change or modify the map, then we do so, and explore again. We also gain a sense of the certainty we can give to our map, or more accurately, its degree of uncertainty. Moreover, if system mapping is of any use, it is in identifying those areas of a complex system (i.e. configurations of factors or flows of causality) with the greatest degree of uncertainty. All of which takes us to the issue of multi-methods.

2. Case-based systems mapping

Case-based systems mapping (CBSM) is part of a suite of interdisciplinary methods making up the field of *case-based complexity* – which emerged out of Byrne's groundbreaking argument that cases meet the definitional criteria of a complex system (Byrne and Ragin, 2009). Case-based complexity is fundamentally configurational. This means that the focus is on how a set of factors (as opposed to any one variable) combine to produce an outcome. The combination of factors is not simply additive or acyclical. It is nonlinear, dynamic and comprised of feedback loops and, in the case of social systems, also comprises inequalities, conflicts, power relations, negotiations and social structures. Case-based complexity is also focused on the wider social contexts/systems in which cases and their configurations are situated. The particular 'place' for which a system map is made and the people of which a 'place' is composed means everything. It is within this wider platform that case-based systems mapping is situated.

As outlined in Table 1, CBSM is a multi-methods approach to exploring and understanding a system by integrating case-based reasoning with cluster analysis, machine learning and network analysis. CBSM explores different configurations of causal flows (case-based submaps) in a system map for different sets of cases and their places; how these different configurational flows cluster together into major and minor clusters; and the impact these different clusters have on an outcome.

CBSM does come with some requirements. The first is that some quantitative or numerical data be available or sought out, even if limited or partial. The second is that the exploratory systems mapping method must be able to model a system in complex terms, which means that it has to be able to map feedback loops and nonlinear causal flows. The third is that the map is focused on 'causality', however one defines it epistemologically – which excludes such systems mapping tools as mind mapping, concept mapping, Cynefin or giga-mapping.

Before turning to the steps outlined in Table 1, several of the ideas central to CBSM require unpacking, as they are potentially new concepts to readers. These include case-based submaps, systems as configurations, clustering cases in a system map, the challenge of data and the utility of data-driven corroboration.

The importance of case-based submapping

A common misconception of systems mapping is that it makes a topic more complex than it need be. This misconception is understandable, as systems mapping is often easy to start, but difficult to conduct well or communicate thoroughly. Submapping is also critical here. Even in the case of an extensive whole-systems mapping exercises, where one seeks to exhaustively identify all key factors and linkages, the ultimate goal is to focus more on various causal flows through the map, i.e. submapping. Submapping involves stakeholders directing their focus to parts of a map or causal flows through it to explore more 'specific questions or purposes, again in a highly exploratory and iterative manner' (Barbrook-Johnson & Penn, 2022, p. 63). The value of the larger map is to provide context, including identifying potential knock-on or unintended consequences or to see factors outside the submap that might be relevant or priorly unknown. As we outline later, CBSM takes a particular approach to submapping, based on how different causal flows for different cases cluster into identifiable trends.

Systems as configurations

The key insight of CBSM is that a systems map can be treated as configurations of factors that, given some set of cases, can be clustered into different systems map profiles, relative to some outcome of concern. To illustrate our point, consider a simple example. You are asked to run a workshop with stakeholders from across a dataset of Local Authorities (LAs) to evaluate policies for improving teenage pregnancy rates. Working together, you produce an exploratory system map, like Figure 1, and begin to identify the major factors impacting your outcome.

The ultimate question your stakeholders are keen to answer is, '*To what extent are these factors and their attempt to mitigate them impacting teenage pregnancy across England*?' It is this question that takes you to the cases in the system and their different configurations. Not every LA in England is similar. You have rural, suburban, urban LAs, which also vary in population density, deprivation, access to healthy foods, etc. It is at this point that your stakeholders turn to their administrative data, which shows that these LAs (cases) have different healthy weight outcomes.

This leads to the next key question, '*Are the causal flows in the system map the same for all these LAs or do they play out differently*?' In response, stakeholders need to move on to create a series of *case-based submaps*, based on how the LAs cluster, given their different factor-based configurations and causal flows. If data are available for the factors your stakeholders have identified, the next step is to explore the extent to which the overall map and its case-based submaps have value for evaluating the effectiveness of their interventions as a council. It is at this point that case-based systems mapping brings in other methods.

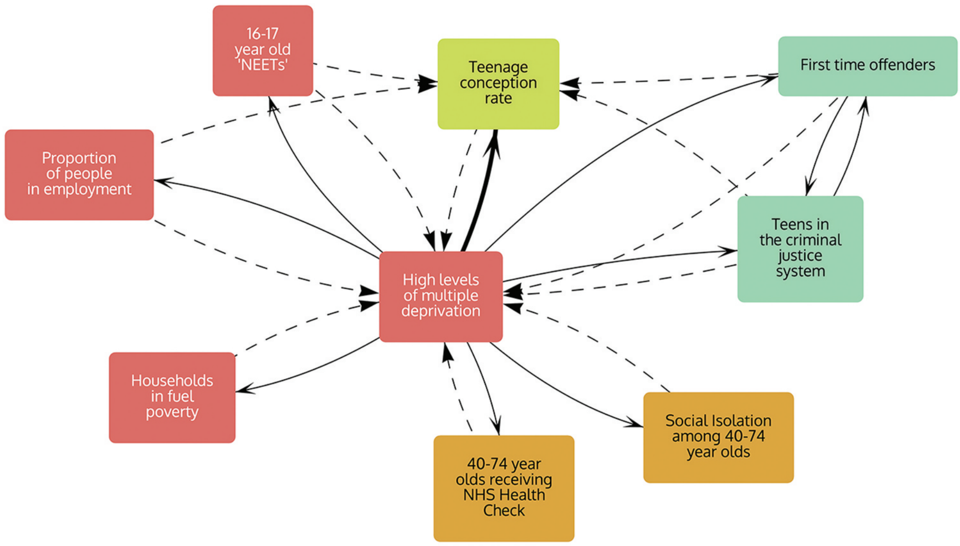

Figure 1. Initial exploratory systems map - teenage birth rate, deprivation and public health.

Data-driven clustering of cases in a systems map

To engage in data-driven collaboration of the initial exploratory map, CBSM involves the usage of data-driven cluster analysis, machine learning and network analysis. These techniques allow stakeholders to test their exploratory maps by, first, clustering cases according to their different configurations and second, creating empirically derived case-based submaps (which are the same as networks) using network analysis. With this step, we move into entirely new mapping territory. Certainly, some systems mapping methods, such as participatory systems mapping, consider how the outcome of a map can be different (or the same) for different cases, e.g. equifinality, multifinality and causal asymmetry. But this process is not done formally. It does not aim to map different cause-and-effect patterns (causal flows) or show how cases group based on differences in their system profiles. It also does not rely on data to do this. Hence, the purpose of Steps 3 and 4.

The challenge of data

With CBSM, a challenge immediately emerges. Having the right data! Through our engagement with stakeholders, we have found that, right from the get-go, case-based systems mapping challenges them to consider what sorts of data they collect, as it is typical for them not to be collecting data on many of the factors in their system map. For some stakeholders, this has led to a useful insight to change the data they collect; for others, it has meant going back to the first step and creating a system map based on the data they have, with an eye to what it reveals and what they might collect next. This gives CBSM a data mining feel, as it prompts questions, again, about the uncertainties in the map, but this time relative to the data needed to more fully understand these uncertainties.

Another issue is cross-sectional versus temporal data. The strength of CBSM and COMPLEX-IT is that it can work with either cross-sectional data, as used in the current paper, or on time-series or more dynamic and longitudinal data (Pagliarin & Gerrits, 2020). Because systems mapping tends to be more cross-sectional, looking at factors and their causal flows at a more meta-theoretical level, one can forget that these maps are dynamic. CBSM is built to conduct such analysis and develop such systems maps. But that is for another paper, as our goal here is to provide a first introduction to the approach. For more on the dynamic modelling capacity of COMPLEX-IT and case-based complexity, see (Schimpf & Castellani, 2020; Schimpf, Barbrook-Johnson & Castellani 2021).

Choosing the right clustering and network technique

There is no ultimate right answer as to which clustering or network analysis technique one should use for steps three and four of CBSM. In our work with COMPLEX-IT, which we will employ next to demonstrate how case-based systems mapping works, we use a combination of k-means cluster analysis, hierarchical cluster analysis and the self-organising map neural net (SOM). We use all three to corroborate the initial system map by comparing the k-means solution, which requires stakeholders to choose the number of clusters they think exist within their map, with the semi-supervised machine learning (where some of the data attributes are known or labelled beforehand and others are discovered through the analysis) solution of the SOM, which arrives at its own solution, and the hierarchical cluster solution the SOM additionally provides. In terms of network analysis, we use a standard program in R. The key is that the network is generated using zero-order correlations amongst the factors in the dataset, with the strength of the correlations determining the strength of the links. We do not stray much beyond basic descriptions, including indegree, outdegree and other basic network statistics. Whatever the program used, it is vital that, in addition to an overall network map, it generates case-based submaps for each cluster. This way stakeholders can not only compare their data-driven network map to their initial systems map, but also how the initial systems map differs for each cluster of cases.

But we are getting ahead of ourselves. If we are to truly understand how case-based systems mapping works and how it is, in its own way, an advance on systems mapping and, more specifically, exploratory systems mapping, then we need to turn to our example and the software package.

5. The four steps of case-based systems mapping

To demonstrate the four steps in our approach, we will overview COMPLEX-IT (Removed for review) and its case-based systems mapping tab, as applied to our UK case study on Local Authority (LA) level socio-economic and health factors and their links to teenage pregnancy rates in those LAs. Readers are welcome to use other combinations of exploratory systems mapping, cluster analysis, machine learning and network analysis to make use of CBSM. This section is organised as follows: we begin with a review of COMPLEX-IT and then our case study, followed by an overview of each of the four steps. Here is the link to COMPLEX-IT (https://www.complex-it-data.org/) and here is the case study dataset for readers to explore (https://art-sciencefactory.com/CBSMdata.html).

COMPLEX-IT

As shown in Figure 2, COMPLEX-IT is a case-based, multi-methods platform designed to increase non-expert access to the tools of computational social science to facilitate exploring and analysing complex data (Schimpf & Castellani, 2020). COMPLEX-IT consists of a bespoke suite of techniques: cluster analysis, machine learning, data visualisation, data forecasting, case-based scenario simulation and case-based systems mapping – the focus of the current paper. COMPLEX-IT supports applied social inquiry through a designed emphasis on learning through the analysis of the complex data/system under study – which makes it ideal for investigative methods, such as exploratory systems mapping. As an R-Shiny program, COMPLEX-IT can be run online or downloaded to run locally in R Studio.

On a technical level, COMPLEX-IT utilises many standard R packages such as the tidyverse (Wickham et al., 2019) collection including plyr, magrittr and ggplot2 (for data manipulation and representation) and the igraph (for statistical analysis of network maps) and visNetwork (for creating interactive network visualizations) packages for the systems mapping tab. The platform keeps an unaltered copy of the user's uploaded data and makes copies for various analyses. COMPLEX-IT, like all R software and packages, follows an open-source philosophy with its code publicly available for inspection and download1. All updates to COMPLEX-IT are reviewed by other team members before being shared publicly. For more technical information about COMPLEX-IT see (Schimpf & Castellani, 2020).

It is important to highlight that COMPLEX-IT has many routes through its tab structure, based on the type of exploratory analysis the user seeks to conduct

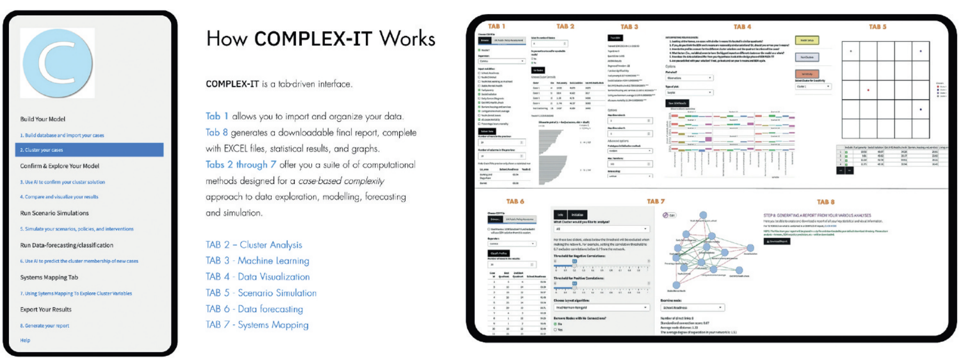

Figure 2. A visual overview of the tab structure of COMPLEX-IT. The lefthand side of Figure 2 shows the App on a phone; the middle of Figure 2 shows the seven methods tabs; tab 8 is for downloading all of the results; the righthand side of Figure 2 gives a quick overview of the tabs. To explore COMPLEX-IT, go to: (removed for review).

(Schimpf & Castellani, 2020). Some may use it entirely for running cluster analysis and machine learning, while others may use it for data forecasting or simulation. CBSM is a specific route through COMPLEX-IT, which can be augmented by or conducted as part of a wider research agenda in COMPLEX-IT. For example, some users may use CBSM in conjunction with the machine learning tab to corroborate further the clusters they find in their systems map, while others may use the results of the systems map to then engage in forecasting for future potential cases.

Case study: deprivation and public health in the UK

To illustrate CBSM, we have taken a theoretically informed selection of indicators from a larger health inequalities dataset covering all local authority areas in England. The exploratory case-study used here is based on the 100 English Local Authorities (LAs) and is largely derived from indicators published under the Public Health Outcomes Framework from the Department of Health and Social Care's profile of the Fingertips Public Health Data (Office for Health Improvement & Disparities) for England. Eight of the nine indicators are taken from this source, and they are: people in employment (proportion of population between 16–64 years); the proportion of households in fuel poverty; first-time offenders recorded (per 100K of population); teens (10–17 years of age) receiving their first reprimand, warning, or conviction per 100,000 population; proportion of 16–17 year olds not in education, employment or training (NEET) or whose activity is not known; the proportion of eligible population (40–74 years) receiving an NHS Health Check (a program to assess the risk of heart, stroke, kidney diseases and diabetes), and proportion of older adults in social isolation who self-identify as having as much social contact with people as they like. The remaining indicator is the Index of Multiple Deprivation (Ministry of Housing, Communities, & Local Government), which provides a weighted average of seven measures relating to deprivation.

Step 1: exploratory systems mapping

The first step in the CBSM process is to develop an exploratory system map of your topic of concern. Exploratory systems mapping helps stakeholders (1) *uncover* and *explore* their individual, conflicting and combined views and beliefs about a system; (2) *synthesise* these views (and the various data and mental models upon which they are based) into a more cohesive picture, including potential 'unknowns' or even 'unknown unknowns' in the system and its feedback loops; (3) *improve communication* around how people see a system; (4) *identify* actionable insights, including how best to address its problems or improve its functioning; and (5) *allow for* the development of a more formal 'systems' model using other methods such as computational modelling, simulation, complex network analysis, ethnography and statistics. To build an exploratory systems map, one follows standard procedures (Barbrook-Johnson & Penn, 2022). The map can be made by an individual or group of stakeholders, be they researchers, civil servants, etc. The only requirements for this first step, which we outlined earlier, are that the mapping approach be able to model a complex system, including feedback loops and nonlinear relationships; the second is that the map focuses on 'causality', however one defines it epistemologically; and the third is that it the system map is based on data, even if partially or with an eye to the data that will eventually be collected.

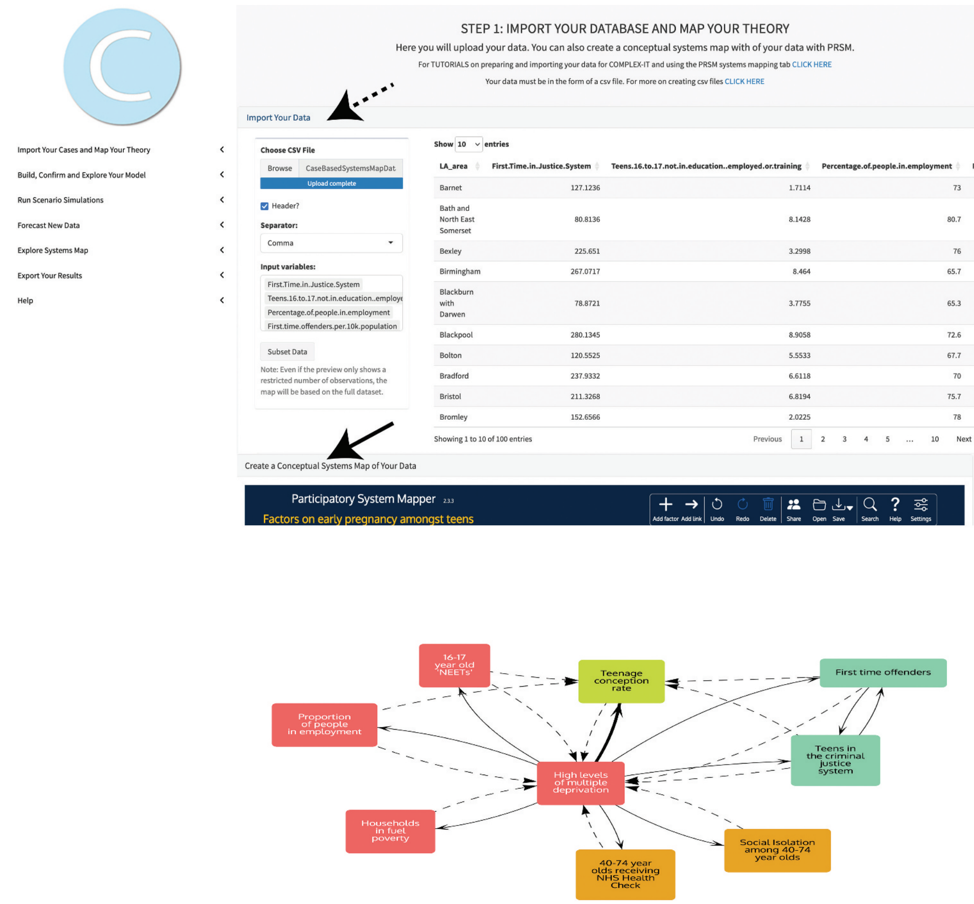

Figure 3. Tab 1 in COMPLEX-IT, with two subtabs for data upload and creating an exploratory systems map.

To accomplish this task in COMPLEX-IT (Figure 3), users are asked to do two things. First, they are asked to upload and visually explore their data to make sure it is all in working order. COMPLEX-IT. Second, they are asked to create an exploratory system map using the open-source platform, PRSM (participatory systems mapper) – https://prsm.uk/. Looking at Figure 3, we see the first three factors in our system map. Below that we see the systems mapping subtab, which is the same map shown in Figure 1.

Our exploratory systems map (Figure 1) was developed by one of the authors of this paper. This author was asked to consider the dataset we had and its factors and construct a systems map based on his expert knowledge of the area. Looking at Figure 1, the indicators have been grouped together and colour coded into three types. The red nodes are all related to socio-economic factors and deprivation, the green nodes to teenagers and crime and orange to adults and their wider health and wellbeing. There are three types of arrowed lines between the circles in the diagram, representing the anticipated direction, strength and certainty of causal inference between the various indicators and the outcome: dotted lines represent moderate causal flows; bold lines represent strong causal flows; and the thicker bold lines (between levels of multiple deprivation and teenage birth rates) represent very strong causal flows. In this respect, we can see that, for example, fuel poverty is hypothesised to contribute to deprivation, but with deprivation being a bigger driver of fuel poverty given the significance of deprivation within the overall environmental context of place. Given the importance of social and economic inequalities in relation to the social determinants of health (SDH) and inequalities of outcomes resulting from these (e.g. Marmot et al., 2020) multiple deprivation has a special status in Figure 1, which is represented by both its proximity to the outcome measure and the extent and complexity of causal flows associated with it.

Step 2: case-based clustering

Step 2 is where CBSM becomes truly case-based. Users hypothesise how causal pathways might differ across cases and their places, forming provisional case-based submaps. This exploratory, qualitative step also encourages users to think through how varying experiences or conditions may cluster, revealing distinct patterns of outcomes within the broader system map. The clusters users identify need not be precise; what matters is sketching them well enough to glimpse meaningful patterns. These provisional groupings lay the groundwork for Step 3, where initial insights evolve into more structured, data-informed understandings of how cases relate and diverge across the system's causal landscape. As an illustration, consider the systems map in Figure 1. After creating the map, the first and third authors discussed the potential causal flows through it based on how the 100 LAs in it might cluster together. They hypothesised there would be at least four or five clusters, with two to three of them being for poor local authorities, which would differ from one another insomuch as the specific factors related to deprivation and their causal influence on teenage pregnancy rates would differ.

Step 3: corroboration with cluster analysis and network analysis

The third step examines the potential or distinct insights of the user's exploratory system map and their case-based submaps (causal flows). To do so, data corroboration involves using *cluster analysis* and *network analysis*.

Step 3A: cluster analysis

In COMPLEX-IT (Figure 4), cluster analysis helps users identify major and (potentially) minor groups, based on different causal flows in a systems map. Each cluster identified is, in systems mapping terms, not just a cluster (in the conventional sense of the method) but also a case-based submap.

COMPLEX-IT uses k-means, which requires users to leverage their expertise and understanding of their system map, including the case-based submaps identified in Step 2, to select an appropriate number of clusters and to assess the validity and practicality of the results. One can even go on to further corroborate one's cluster analysis with a semi-supervised machine learning algorithm, but that is beyond the focus of the current paper. For more see (https://www.complex-it-data.org/) each cluster identified from the systems map is represented by a centroid, summarizing the average values for each case profile within that group. Ideally, distributions within clusters are tightly concentrated around their centroids and distinct from other clusters.

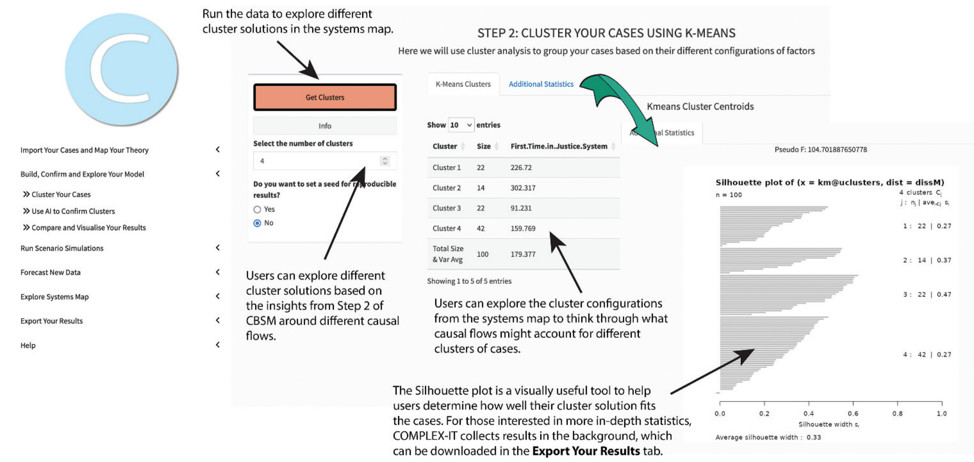

Figure 4. Table 2 (cluster analysis) in COMPLEX-IT showing the various tools for running a cluster solution on the factors and cases in the exploratory systems map.

COMPLEX-IT displays the resulting cluster profiles and their sizes through a series of graphics and quality measures (See Figure 4), Table 2 including pseudo – F and silhouette plots. The pseudo-F metric gauges how well cases are grouped within clusters and how distinct the clusters are, with higher values indicating stronger performance. The silhouette plot provides a visual and quantitative assessment of how well cases fit within their clusters (see Figure 6). Table 2, for example, was generated using the results from our study. As mentioned earlier, we identified four clusters based on different outcomes in Teenage Pregnancy and the case-based submaps that help to explain those differences. By reviewing these clusters alongside domain knowledge and these quality measures, users can determine the most effective arrangement for identifying major and minor clusters within their dataset.

Step 3B: network analysis (systems mapping tab)

While useful for identifying distinction profiles, cluster analysis does not tell us how the factors involved in a causal flow link to one another. Hence, the purpose of the Systems Mapping tab in COMPLEX-IT. As shown in Figure 5, this tab provides users with a series of tools to employ a variety of network analytics to help make sense of the data-driven map and its case-based submaps.

Warning! It is important to understand that the network links formed in an exploratory systems map are not relational. The model is causal, based on the zero-order correlations amongst the factors in the study dataset. This intentional ‘misuse’ of network analysis, while innovative and useful, is a violation of its assumptions. In a typical relational network one explores, for example, how Ruby knows Maggie, who knows Carol, which impacts the latter’s health outcomes. In a CBSM network, one uses the zero-order correlations amongst a set of factors to explore how, for example, a community’s lack of jobs impacts its healthcare system, which impacts teenage pregnancy rates. Case-based submaps are directed, correlational and cyclical networks. For an in-depth

Table 2. K-mean cluster solution for $N = 100$ cases in exploratory systems map.

Cluster ID and Size	Teens First Time in Justice (PER 100K POP)	Teens (16–17) Not in Edu/Employ/Training (%)	People in Employment (%16–64 Yrs)	First-Time Offenders (Per 100K POP)	Households in Fuel Poverty (%)	Older Adults in Social Isolation (%)	Adults Received NHS Health Check (% 40–74 Yrs)	Index of Multiple Deprivation	Under-18 Conception Rate (Per 1k POP)
1 (14)	302.3173	5.3918	74.8143	184.3753	11.3357	43.0286	39.4004	28.4577	16.7262
2 (42)	159.7687	4.9191	74.5071	158.8579	9.8881	45.2690	35.3320	22.1870	16.6868
3 (22)	226.7204	6.0182	73.8955	216.7806	11.6227	46.3000	29.2947	29.5585	19.2048
4 (22)	91.2310	5.3194	75.4545	123.8183	9.7636	47.4818	33.1542	17.5075	13.6461

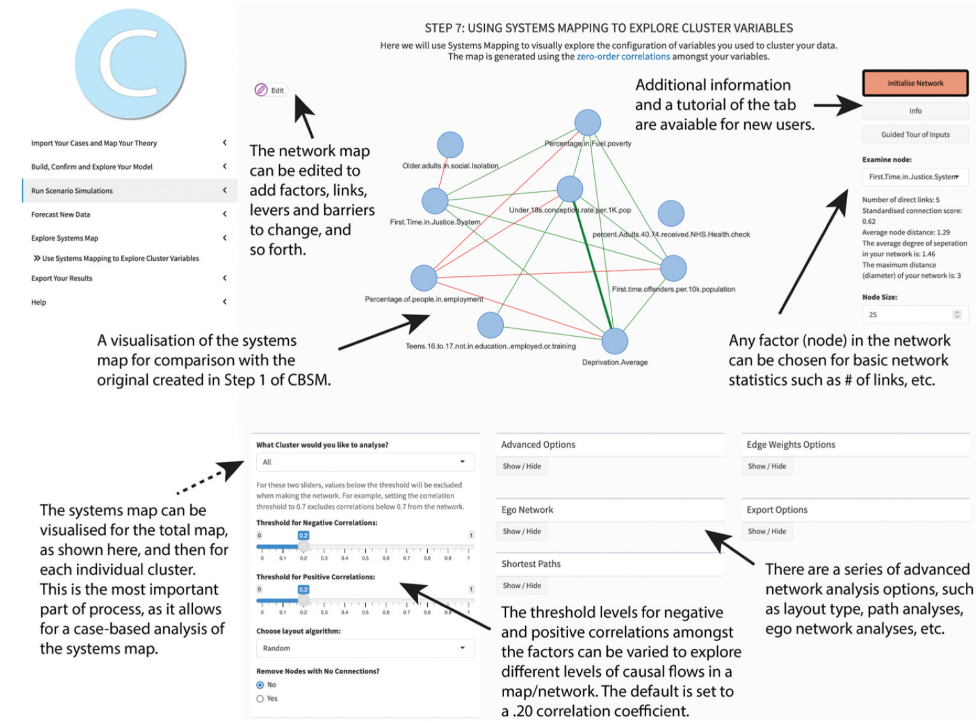

Figure 5. Tab 7 (systems mapping tab) in COMPLEX-IT, which uses the network analysis techniques, based on zero-order correlations amongst the factors in the user's dataset, to generate a data-driven systems map of a topic and case-based submaps, based on the clusters found in the data.

summary of the challenges and solutions to using network analysis for causal modelling, see the mental health (psychopathology) symptom network literature (Robinaugh et al., 2020). Given our intentional misuse of network analysis for exploratory systems mapping, we suggest refraining from using correlation coefficients (be they zero-order or otherwise) or any sort of numeric assignment to the links in a systems map.

Exploring the systems mapping Tab. The Systems Mapping Tab is comprised of a series of network analysis tools. Moving around Figure 5 clockwise and starting at the top, there is the INITIALISATION button, which creates a map. Once up, one can use the EXAMINE NODE dropdown to explore the network statistics for each factor in the network. One can then move on to the advanced options section at the bottom, such as looking at key network options as SHORTEST PATHS and EGO NETWORK, or adding user defined weights to the links between factors by changing the EDGE WEIGHT OPTIONS

When looking at the links in a map, it is necessary to consider the THRESHOLD for POSITIVE and NEGATIVE CORRELATION. Systems Mapping tab uses the zero-order correlation matrix amongst the factors in the dataset, with each correlation ranging from −1.00 to 1.00. The default setting is .20 correlation coefficient. Note that one can also request, as one raises the threshold, to have nodes (factors) removed from the network that are no longer linked to other factors in the network. There are several advantages to raising the threshold. A low-level setting of -/+.20 (or even lower) brings more factors

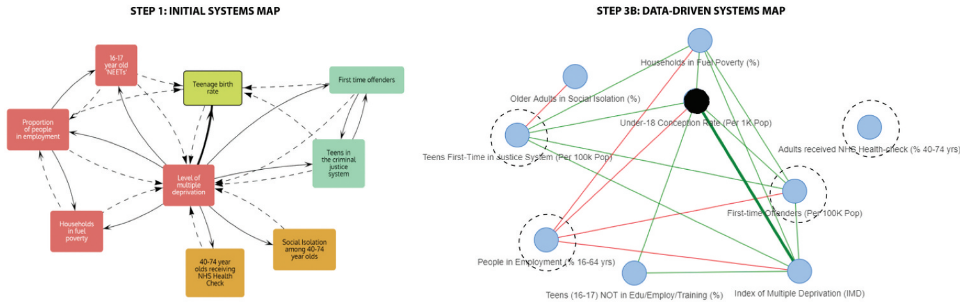

Figure 6. A comparison of step 1 initial systems map with the step 3B systems map.

into the network, thereby increasing the complexity of the map. This has value for revealing more complex potential causal flows through the systems map, including secondary or tertiary indirect causal links. Raising the threshold to -/+.50 or higher, even as high as -/+.70, reduces the complexity, revealing the most relevant factors and the most parsimonious potential causal flows. We caution such a reductionist approach, as the value of systems mapping is to understand the complexity of a topic. Relative to this point, even with the threshold set at -/+.20, one can visually identify the stronger links based on the thickness of the line. The thicker the line, the higher the correlation, allowing users to see which links are the strongest, be they negatively or positively correlated. Also, red lines are negative correlations; green lines are positive correlations.

Given the value of visualising a map, and of corroborating it with the initial systems map made in Step 1 of CBSM, there are several visualisation options in COMPLEX-IT. The CHOOSE LAYOUT ALGORITHM option provides several of the more popular.[3]

Corroborating the overall systems map. The first network created by the Systems Mapping tab is the overall map, which allows users to compare their initial map with this new data-driven map. For a quick illustration, see Figure 6. The first thing noticeable is that the data-driven map (right) supports the initial map (left) insomuch as they both link many of the deprivation factors similarly to teenage pregnancy rates. The only difference is that the initial map places deprivation more centrally, but that is due largely to a conceptual view. Still, the data-driven map used in the study can contradict the initial map – which would lead to its own set of insights. That is the purpose of corroboration.

Exploring case-based Submaps. The next step is to redraw the map for each of the clusters identified in Steps 2 and 3 of CBSM. Here is where issues of equifinality, multifinality and causal asymmetry emerge. For the purposes of illustration, we will explore the different case-based submaps for Cluster 3 and Cluster 4 from our study, which had the highest and the lowest teenage pregnancy outcomes in the dataset.

Looking at Figure 7, the first thing that stands out is how different the two maps are from the overall map shown in Figure 6, be it the initial map or the data-driven map; and also how different the two clusters are from each other. This makes our point hopefully in a visually compelling way: the causal flow through a systems map that does not consider the clustering of case-based differences can lead to an incorrect understanding of how

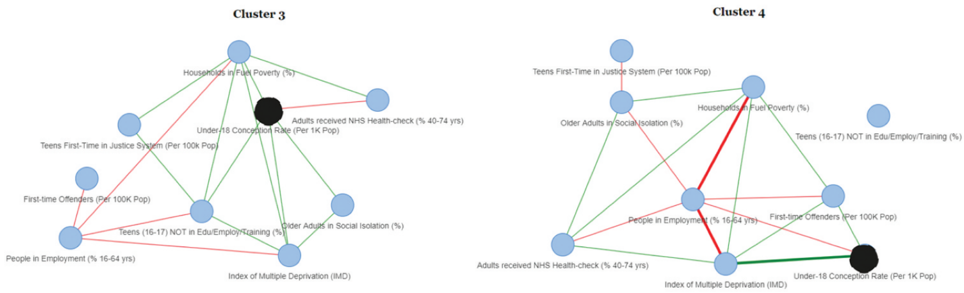

Figure 7. A comparison of cluster 3 and cluster 4 case-based submaps.

a set of factor impacts an outcome of concern, leading to potentially incorrect theoretical framings or policy and practice interventions.

For example, in highly deprived Cluster 3, teenage Pregnancy rates are more central to the network; while in Cluster 4, with the lowest rates, it is more peripheral to the network. And while teenagers not in employment are a key link in Cluster 3, given the role jobs play in mitigating economic deprivation, it is not even linked to the network in Cluster 4, at the .20 correlation coefficient threshold. Another thing that stands out is how different the first-degree links to Teenage Pregnancy are in Cluster 3 versus 4, suggesting a different causal flow in the two sub-maps. In Cluster 3, every single first-degree link is different from Cluster 4 except the measure of overall deprivation. This is a striking difference when one considers the relevance of these differences in terms of policy and intervention. The LAs in both clusters require variations in policy to improve Teenage Pregnancy rates.

Step 5: Re-assessment

In the spirit of data mining, the fifth step involves thinking over what has been learnt and deciding, from there, what to do next. This final step involves users considering their results to confirm or adjust their original map and its smaller case-based maps to more effectively address their outcome of concern, as well as consider what additional data or methods of analysis might be needed. This could involve, for instance, returning the data sources used and adding new factors that might explain differences in the case-based submaps and/or overall map or removing factors that are unrelated to other factors across the maps. Ragin (2009) called this process casing, or iteratively developing a more robust set of configurational factors to represent the cases under study. This assessment stage may also involve revisiting and revising prior qualitative outcomes, such as exploratory system maps or submaps. More generally, the process of building a map and case-based submaps qualitatively, followed by data-driven analysis forces the analyst to externalise their models, theories, or understanding of the system more concretely. Often these models may be implicit in the analyst's mind (e.g. see Kezar et al., 2015) and thus a re-assessment once an initial analysis is complete is often warranted to further refine and adapt one's model. Figure 8 provides a brief visual display of how analysts may (iteratively) proceed through the steps of CBSM and specifically where assessment may lead them to return to or explore next. The final step 'New line of inquiry' may include entirely new studies that emerge from the present analysis or extension of CBSM to

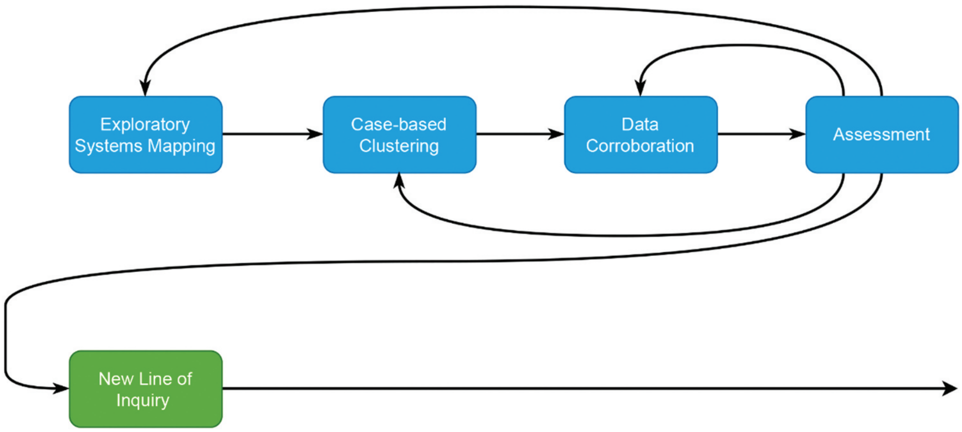

Figure 8. CBSM method process and possible assessment revisions.

include more methods or data, such as engaging documentary sources or additional modelling approaches. CBSM was developed in the spirit of methodological pluralism (Stern, 2015) and we believe exploring its integration with other methods represents a promising line of future research.

6. Conclusion

In this work, we introduce case-based systems mapping or CBSM, a multimethod approach that integrates exploratory systems mapping and case-based complexity to enable modelling of complex systems holistically and through distinct sets of cases or configurations that may exist within the system. CBSM is realised through four steps: (1) individual or group (participatory) visual mapping of the system of interest, (2) identification of potential case-based clusters within the larger map or system, (3) data corroboration through a data-driven cluster analysis and subsequent network analysis followed by (4) an assessment and potential iteration or refinement of prior analysis or new analysis directions. While we demonstrated CBSM step (3) with COMPLEX-IT (https://www.complex-it-data.org/), as it provides a compact and streamlined means for conducting this approach, analysts may use any combination of clustering and network analysis for this. The case study of multiple deprivation and health inequalities impact on teenage pregnancy rates in the UK illustrated how CBSM can unveil stark differences between a full system map and case-based submaps, as well as insights into the dynamics happening within and across the submaps. For instance, comparing the full map and submaps, or across submaps with different teenage pregnancy rates, factors shifted from being more central to more periphery; factors exhibited different connections or an absence of connections altogether; and factors showed notable shifts in the strength of connections.

In short, this approach helps analysts better capture the complexities of the social world and the dynamics and causal forces at play within and across distinct cases and places. CBSM thus holds promise for a broad array of fundamental and applied researchers, civil servants, government and public organisations seeking to better apprehend contemporary research challenges and promote evidence-based interventions. Following ongoing efforts to combine systems mapping with computational methods (e.g. see Barbrook-Johnson & Penn, 2022), future work with CBSM should explore different combinations of exploratory systems mapping and case-based complexity, including different cluster and network analysis methods, to better understand the strengths and weaknesses of each combination. Moreover, while CBSM is housed squarely within the growing calls for methodological pluralism (Stern, 2015), there remain opportunities to integrate CBSM with other methodological approaches be they computational, complexity-oriented, mixed methods or others to enhance the ongoing efforts to study social complexity.

Notes

1. For more on COMPLEX-IT, including freely accessing the software online or for download, as well as tutorials and examples, see (https://www.complex-it-data.org/)
2. Our approach is not Bayesian belief modeling, which simplify systems into static, acyclic, probabilistic structures.
3. The layout options, as well as other facets of the code underpinning the systems map itself, use functions from the igraph R package. The visualisation of the network is handled by the visNetwork R package, itself based on the vis.js javascript library.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Brian Castellani is Director of the Durham Research Methods Centre and Co-Director of the Wolfson Research Institute for Health and Wellbeing at Durham University, UK. He is also Visiting Professor at Nelson Mandela University (ZA), Adjunct Professor of Psychiatry at Northeastern Ohio Medical University (US), and Fellow of the UK National Academy of Social Sciences. Trained as a public health sociologist, clinical psychologist, methodologist and artist, he focuses on environmental health research and policy. His methodological focus is computational modelling, interdisciplinary methods, and visual complexity, including co-creating the COMPLEX-IT platform.

Corey Schimpf is Assistant Professor in the Department of Engineering Education at the University at Buffalo (US). Trained in engineering education and sociology, his research focuses on design thinking, human-centered design, student agency, and computational social science, with particular emphasis on participatory methods and educational technology. He teaches and studies methods in engineering education and undergraduate engineering, with particular interest in combining sociological insight and technical/data methods.

Jonathan Wistow is Associate Professor in the Department of Sociology at Durham University (UK). He is Co-Director of the Durham Research Methods Centre (DRMC), where he leads work on complexity, evaluation and health, and he also leads on health and place policy research in the Wolfson Research Institute for Health and Wellbeing. His research focuses on health inequalities, climate-change adaptation, political economy and place-based governance, applying complexity theory and qualitative comparative analysis to public health and policy.

Christopher Caden is Senior Programmer for the COMPLEX-IT platform and Research Assistant in the Durham Research Methods Centre at Durham University (UK). His work focuses on developing R-based multimethod platforms that make computational modelling and case-based complexity methods accessible to non-expert public stakeholders.

Jutshi Agarwal is Research Scientist in the Department of Engineering Education at the University at Buffalo, SUNY (US). Trained in aerospace engineering and engineering education, her research focuses on teaching preparation of graduate students, student agency in open-ended problems, teamwork in classrooms, and the professional development of K-12 teachers, using quantitative, qualitative, and mixed-methods approaches.

Peter Barbrook-Johnson is Lecturer in Social Economics in the Department of Arts & Sciences at University College London (UK) and Associate Fellow at the Institute for New Economic Thinking, University of Oxford. His research applies agent-based modelling, systems mapping, and complexity economics to environmental, energy and public health policy.

ORCID

Corey Schimpf http://orcid.org/0000-0003-2706-3282

References

Barbrook-Johnson, P., & Penn, A. S. (2022). *Systems mapping: How to build and use causal models of systems*. Springer Nature.

Castellani, B., & Gerrits, L. (2024). The Atlas of Social Complexity. Edward Elgar Publishing.

Checkland, P., & Poulter, J. (2020). Soft systems methodology. In M. Reynolds & S. Holwell (Retired) (Eds.), *Systems approaches to making change: A practical guide* (pp. 201–253). Springer London. https://doi.org/10.1007/978-1-4471-7472-1_5

Crielaard, L., Uleman, J. F., Châtel, B. D., Epskamp, S., Sloot, P., & Quax, R. (2024). *Refining the causal loop diagram: A tutorial for maximizing the contribution of domain expertise in computational system dynamics modeling*. Psychological Methods.

Hong, Q. N., Bangpan, M., Stansfield, C., Kneale, D., O'Mara-Eves, A., van Grootel, L., & Thomas, J. (2022). Using systems perspectives in evidence synthesis: A methodological mapping review. *Research Synthesis Methods*, *13*(6), 667–680. https://doi.org/10.1002/jrsm.1595

Kezar, A., Gehrke, S., & Elrod, S. (2015). Implicit theories of change as a barrier to change on college campuses: An examination of STEM reform. *The Review of Higher Education*, *38*(4), 479–506. https://doi.org/10.1353/rhe.2015.0026

Marmot, M., Allen, J., Boyce, T., Goldblatt, P., & Morrison, J. (2020). *Health equity in England: The Marmot review 10 years on*. Institute of Health Equity.

Pagliarin, S., & Gerrits, L. (2020). Trajectory-based qualitative comparative analysis: Accounting for case-based time dynamics. *Methodological Innovations*, *13*(3). https://doi.org/10.1177/2059799120959170

Ragin, C. (2009). Reflections on casing and case-oriented research. In D. Byrne & C. Ragin (Eds.), *The SAGE handbook of case-based methods* (pp. 522–534). Sage Publications, Inc.

Rihoux, B., & Ragin, C. C. (Eds.). (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques* (Vol. 51). Sage.

Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008-2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Schimpf, C., & Castellani, B. (2020). COMPLEX-IT: A case-based modelling and scenario simulation platform for social inquiry. *Journal of Open Research Software*, *8*(25). https://doi.org/10.5334/jors.298

Schimpf, C., Barbrook-Johnson, P., & Castellani, B. (2021). Cased-based modelling and scenario simulation for ex-post evaluation. Evaluation, *27*(1), 116–137. https://doi.org/10.1177/1356389020978490

Stern, E. (2015). *Impact evaluation: A guide for commissioners and managers*. Bond for International Development.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., & Spinu, V.Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686
Williams, B., & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioner's toolkit*. Stanford Business Books.