Medical Sociology and Case-Based Complexity Science: A User's Guide

31

Brian Castellani, Corey Schimpf, and Frederic Hafferty

31.1 Introduction

In this chapter, we provide readers with the first formal overview of case-based complexity science and its related methodology, case-based modeling. **Case-based modeling**, championed largely by [1], constitutes a fifth major method for modeling complex systems, offering itself as an alternative to (and also integration of) agent (rule-based) modeling, dynamical (equation-based) modeling, qualitative (idiographic) modeling, and statistical (aggregate-based) modeling. For us, as medical sociologists, case-based modeling makes sense because, fundamentally, medicine is about the case. Case-based modeling also resonates with our particular practice of a **case-based complexity science**, which can be defined as a generalist approach, grounded in the epistemological perspectives of Byrne's complex realism—which we explain later.

B. Castellani, Ph.D. (✉)
Kent State University, 3300 Lake Road West, Ashtabula, OH, USA
e-mail: bcastel3@kent.edu

C. Schimpf, M.A.
Purdue University, School of Engineering Education, Neil Armstrong Hall of Engineering, Room 1300,701 W. Stadium Avenue, West Lafayette, IN, USA
e-mail: cschimpf@purdue.edu

F. Hafferty, Ph.D.
Program in Professionalism and Ethics, Mayo Clinic, Minneapolis, MN, USA
e-mail: fredhafferty@mac.com

In terms of case-based modeling, we employ the **SACS Toolkit**. The SACS Toolkit is a new, computationally grounded, case-based method we created for modeling complex social systems as a set of cases [2, 3]. In terms of health, we use the SACS Toolkit to study communities, school systems and stress and coping issues as different types of complex systems [4, 5]; and, in terms of health care, we use it to study the complexities of medical professionalism and medical education [6, 7]. By the end of our review, interested readers should have enough knowledge of case-based modeling and the SACS Toolkit to determine its viability for their own research.

Our chapter is organized as follows. We begin with an overview of case-based method, providing a quick history of how case-based complexity science and, more specifically, case-based modeling emerged and how we position our approach relative to other ways of modeling complex systems. From here we turn to the SACS Toolkit, providing a quick overview of how it models complex systems. Detailed reviews of the SACS Toolkit currently exist—one qualitative in focus [2] and the other mathematical [3]. However, a few advances are made in the current chapter, as it is (1) our first attempt to clarify the SACS Toolkit's explicit links to case-based complexity science and (2) our first effort to integrate our two previous versions into a new, updated version. To help readers grasp a basic understanding of the SACS Toolkit, our review will draw on some examples from our research in medical sociology.

J.P. Sturmberg and C.M. Martin (eds.), *Handbook of Systems and Complexity in Health*,
DOI 10.1007/978-1-4614-4998-0_31,

31.2 Case-Based Complexity Science: An Overview

Over the last several years, Professor David Byrne of *Durham University, UK* has emerged as a leading international figure in what most scholars see as two highly promising but distinct fields of study (1) case-based method and (2) the sociological study of complex systems. An example of the former is Byrne's *Sage Handbook of Case-Based Methods* [1]—which he co-edited with Charles Ragin, the most prominent figure in case-based method. An example of the latter is his widely read *Complexity Theory and the Social Sciences* [8]. What scholars (including the current authors) are only beginning to grasp, however, is the provocative premise upon which Byrne's work in these two fields is based. His premise, while simple enough, is ground-breaking:

> Cases are the methodological equivalent of complex systems; or, alternatively, complex systems are cases and therefore should be studied as such.

With this premise, Byrne adds to the complexity science literature an entirely new approach to modeling complex systems, alongside the current repertoire of agent (rule-based) modeling, dynamical (equation-based) modeling, statistical (variable-based) modeling, network (relational) modeling, and qualitative (meaning-based) method.

Working independently of and yet in tandem with Byrne, we have used his premise to develop a case-based, computationally grounded, mixed-methods technique called the SACS Toolkit [2, 3]. However, because it is designed for studying both small-database and large-database complex systems, the SACS Toolkit makes a slight variation on Byrne's premise: it models a complex system as a *set* of cases, ranging from, at minimum, 1 case to any large number of cases. In the language of matrix algebra, these cases are k dimensional vectors (See 3 for complete mathematical overview). The goal of the SACS Toolkit is to compare and contrast and then condense and cluster databases comprised of a large number of cases to create a low-dimensional model of a complex system's structure and dynamics across time/space. To create these models, the SACS Toolkit employs a variety of computational techniques—including cluster analysis, network analysis, agent-based modeling, and artificial neural nets—as well as statistics, historiography and qualitative method.

Before we can overview the SACS Toolkit, however, it is necessary to situate it within the larger fields of case-based method and complexity science. We begin with case-based method.

31.2.1 Case-Based Method

Case-based method is an umbrella term for a somewhat varied set of techniques that have a long history in the social sciences and others fields such as biology, history, archaeology, and medicine [1].

Case-based methods, whatever the type, can be explanatory or descriptive. They can be static or longitudinal, retrospective or prospective. Despite differences, the goal of these methods is to study a case or set of cases more holistically, systematically, and ideographically. The simplest example of a case-based method is the **case study**, which is an in-depth investigation of a single case. Most approaches, however, tend to study a set of cases, engaging in what is called **case-comparative method**.

The most popular version of case-comparative method is Ragin's **qualitative comparative analysis** (QCA) [9]. Over the last decade, QCA has developed into a set of comparative techniques that allow case comparative methods to move beyond the limitations of traditional qualitative method. While case comparison is implicitly the purpose of such statistical techniques such as cluster analysis and discriminatory analysis, it is rarely couched in such terms. As such, most case-comparative methods are grounded in a qualitative tradition, focusing on a small number of cases. QCA pushes case-comparative method into a mixed-methods frame, allowing researchers to capitalize on the strengths of both qualitative and quantitative analysis, insomuch as it uses

Boolean algebra and its matrices to search for patterns and to make generalizations with larger datasets [1]. As a side note, Boolean algebra is a variant of algebra that works only with 1 s and 0 s, as truth values: cases either do or do not belong to a dominant profile identified—for example, sick or not sick patients. Ragin has also developed a fuzzy-set version of QCA which overcomes the limitations of Boolean algebra and its crisp sets to allow cases to have "degrees of membership" in the main profiles identified by a study—think, for example, of a study that allows people to be, in varying degrees, both healthy and sick [9, 10].

Regardless of the case-based method used, a **case** can be a person, event, place, concept, etc. Whatever studied, the case is the focus of the study, not the individual variables or attributes of which it is comprised. Case-based researchers would not, for example, study the impact gender has on the professional behavior of physicians. Instead, they would study how the different profiles of physicians explain their dissimilar professional behaviors, with gender being one of the key attributes examined. Case-based methods also treat the cases they study as **composites**, viewing them as comprised of an interdependent, interconnected set of variables, factors, or attributes that form some type of emergent configuration, such that the whole is more than the sum of its part. Each variable, therefore, is not an isolated factor impacting the case of study; instead, it is part of a larger, context-specific set of factors which collectively define the case of study, usually in rather nonlinear ways. For example, in Ragin's QCA, variables are treated as sets [10]. A case profile, therefore, tells us which sets (a.k.a variables) a case belongs to and in what manner or degree.

Case-based methods do, however, differ from one another in the degree to which they seek to **generalize** their findings. Byrne [11, 12] for example—the leading figure in case-based complexity science—advocates contextualized and limited forms of case-based generalization. Despite these differences, all case-based methods treat cases as particular instances, examples, occurrences, or types of some larger population.

It is its **configuration approach** to variables, however, that ultimately makes case-based method a radical departure from normative, variable-based inquiry, as defined by the majority of statistical methods used in the social sciences and, more specifically medical sociology—think here of **conventional method**. Variable-based statistics has no interest in cases or any in-depth understanding of how a set of variables collectively define or impact these cases. Instead, variable-based inquiry seeks to understand the relationship variables have with each other, and usually in the most parsimonious, reductionist, nomothetic, linear, unidirectional manner possible. To illustrate, let us go back to our example of physician professionalism. A variable-based study might examine which factor (amongst some set of supposedly independent variables) best explains the different professional behaviors of physicians. For example, which variable is more important for later misconduct? Is it the networks students hang out in or the number of times they were cited for unethical behavior? In contrast, a case-based approach would examine how the different variable-based configurations of some set of physicians (cases) account for differences in professional misconduct. For example, one may find that male students, specializing in surgery, who attended schools that failed to really punish their misconduct, and who socialized in student networks that approved of their "bad" behavior went on to practice in similar ways later in life: that is, they engaged in misconduct while working at hospitals that did little about their behavior, and they worked in physician networks that approved of their behavior.

31.2.2 Case-Based Complexity Science

As Byrne recognized in his research [1, 8, 11, 12], not only is case-based method a radical departure from variable-based inquiry but also it has strong affinity with complexity science. Going even further, it also, in some very useful ways, advances the study of complex systems.

Based on Byrne, we wish to introduce in this chapter two new terms: case-based complexity science and case-based modeling. **Case-based**

complexity science is defined as scholarly activity that seeks to actively integrate case-based method with complexity science for the purpose of modeling complex systems as cases. **Case-based modeling** is defined as the set of techniques scholars use to conduct case-based complexity science.

In addition to Byrne, scholars involved in the development of case-based complexity science and case-based modeling include 2,9,10,13, and 14. The argument is simple enough. Cases are the methodological equivalent of complex systems. If one thinks about it, complexity scientists and case-based researchers make a similar argument (1) variable-based inquiry is insufficient for modeling complex systems; (2) needed instead are methods that employ an idiographic approach to modeling, one grounded in the techniques of constant comparison; (3) the whole of a case or system is more than the sum of its part; (4) and yet, the study of parts and their complex interactions, from the ground-up, including the interactions these parts have with the case or system as a whole, is the basis to modeling. We can go on. Bottom line: cases are complex systems; complex systems are cases.

The above argument, however, is as far as the similarities go. Fact is, Byrne (as well as ourselves) set case-based complexity science as its own particular approach, distinct from the approach *en vogue* within complexity science today. To clarify this distinction, several comments are in order.

31.2.2.1 Situating Case-Based Complexity Science

In the last thirty years, Academia has witnessed the emergence of what many scholars—including Stephen Hawking—call a "new kind of science." The name of this new, massively interdisciplinary science is **complexity**. While young, complexity science (like many new scientific innovations of late) has captured part of the academic and public imagination—in this case with discussions of six-degrees of separation, swarm behavior, computational intelligence, and simulated societies. This popularity, however, has come with a price: confusion over the field's core terminology and the disciplinary divisions within it. As Mitchell explains in her popular work, *Complexity: A Guided Tour* [15], while it is popular to refer to complexity science in the singular, "*neither a single science of complexity nor a single complexity theory exists yet, in spite of the many articles and books that have used these terms*" (2009, p. 14).

If one follows Castellani and Hafferty [2], however, complexity science's confusion over terminology has less to do with its age, and more to do with its interdisciplinary and therefore interstitial (between things) character. Interstitial areas of thinking, no matter how novel, replicate the dominant intellectual divisions of academia, such as science versus theory or qualitative method versus statistics. Complexity science, given that it situates itself within the full range of academic inquiry—from the humanities and the social sciences to mathematics and the natural sciences—is replete with such divisions. As such, while oriented toward the study of complex systems in general, scholars in complexity science find themselves struggling with significant divisions regarding the complexity theories they use, the methods they employ, the epistemologies upon which they rely, and the definitions of a complex system they embrace. Given these divisions, a few clarifications are in order—all of which help us to understand better the goal of case-based complexity science.

1. The first clarification concerns the goals of science. As mentioned by Mitchell [15], complexity science is really the complexity sciences. To date, complexity science can be organized into several competing types, based on different combinations of the dominant distinctions in academia [16].

For Byrne (and for us), one of the most important distinctions is between what Morin [16] calls restricted versus general complexity science. **Restricted complexity science** is popular in economics and the natural sciences. It is defined as the empirical study of complex systems via the methods of rule-based, computational modeling. Its goal is quasi-reductionist, as it seeks to identify and explore the set of rules out of which complex systems emerge, so it can generate quasi-general laws about complex systems. In contrast is **general complexity science**, which is defined as the empirical study of complex systems via the broader methods of the humanities and the sciences. Its goal is more qualitative and holistic, seeking to model complex systems to create context-specific, grounded theoretical understandings of complex systems. **Case-based complexity science** situates itself in the latter approach.

As Klüver and Klüver make clear in their book *Social Understanding: On Hermeneutics, Geometrical Models and Artificial Intelligence* [17], most sociological phenomena are simply too complex to be reduced to the emergent consequence of rule-following. A more general approach, as Byrne explains [12], is one that that acknowledges this point: context and messiness and the mutual influence of macroscopic and microscopic structures and dynamics are crucial to understanding social systems.

2. The second clarification concerns computational modeling. A defining feature of the complexity sciences (restricted and general) is their reliance upon the latest developments in computational modeling. As Mitchell [15] explains, while the complexity sciences offer scholars a handful of new concepts (autopoiesis, self-organized criticality), their major advancement is **method**. Case in point: one can go back to the 1800s to Weber, Marx, Pareto, or Spencer to find reasonably articulate theories of society as a complex system; or, one can go back to the 1950s to systems science and cybernetics (or, more recently, social network analysis in sociology) to find many of the concepts complexity scientists use today. Despite their theoretical utility—which, albeit critically received, is widespread—all the aforementioned theories ultimately stalled in terms of the study of complex systems because (amongst other reasons) they lacked a successful methodological foundation.

Computational modeling is the usage of computer-based algorithms to construct reasonably simplified models of complex systems. There are three main types of computational models used in complexity science: agent (rule-based) modeling, network (relational) modeling, and dynamical (equation-based) modeling. Different methods yield different results. Situating itself within the latest advances in computational modeling, case-based complexity science seeks to use these tools. Byrne [12] and Uprichard [14], for example, use cluster analysis; and our own work employs agent-based modeling, cluster analysis, neural nets, and network analysis [3]. But, the focus is on comparing cases and searching for common case-based profiles, as concerns a particular health outcome. The consequence of this focus is the causal model built—not the techniques used. Focusing on cases is a search for profiles: context dependent assemblages of factors (k dimensional vectors) that seem to explain well for example different types of health outcomes. For example, one could use computational modeling to examine a set of health factors (e.g., income level, education, gender, age, and residential location) to see which case-based assemblage of these factors relate to differences in mortality rates.

3. The third clarification concerns the distinction between complexity science and complexity theory. Like complexity science, there are multiple complexity theories, which form a loosely organized set of arguments, concepts, theories, and schools of thought from across the humanities and the social sciences that various scholars use in a variety of ways to address different topics.

In terms of intellectual lineage, these theories are strongly grounded in two intersecting epistemological and theoretical traditions: the one stems from systems theory, Gestalt psychology, biological systems theory, second-order cybernetics, and ecological systems theory; while the other stems from semiotics, post-structuralism, feminism, postmodernism, constructivism, constructionism, and critical realism [2].

Complexity theories and their related epistemologies are also tied up in the substantive systems theories of sociology, anthropology, political science, economics, psychology, and managerial studies. As such, complexity theories can differ dramatically from one another. For example, Niklas Luhmann uses complexity theory to articulate a new, *metaphorical* theory of global society (a grand theory with no agents, only a communicating society); while John Holland uses complexity theory to build a bottom-up, agent-based *computational theory* of complex emergent systems.

Perhaps the sharpest distinction between complexity theory and complexity science, however, is that neither necessarily has affinity for the other. In fact, complexity theories need not be data driven, empirically grounded, computational, or scientific. They can even be anti-data, anti-empirical, anti-computation, and anti-scientific. For example, Francois Lyotard uses early *empirical* research in complexity science (mainly chaos theory) to end grand narrative and place a limit on the conditions of science, which he called post-modernity. Meanwhile, most scholars in the managerial sciences use complexity theory in a *prescriptive* manner, with almost no empirical backing whatsoever [16]. In contrast, the complexity sciences, while reliant upon key concepts from complexity theory, such as self-organization or emergence, tend to ignore theory (Mitchell 2009). For example, most rule-based complexity science is theoretically vacuous.

Given the above distinctions, the generalist approach of case-based complexity science is grounded in a post-positivistic epistemology, albeit one that has learned from the errors and shortcomings of much of postmodernism and post-structuralism. This seasoned viewpoint is best described as complex realism, which combines Bhaskar's critical realism with Cillier's understanding that knowledge and the world are complex interdependent processes. Together, these two ideas form what Byrne calls *complex realism*. Here is an all-too-short overview of its main point. For an in-depth review, see Byrne [12]. Complex realism seeks to overcome two key problems.

The first is epistemological. Why is reality so hard to comprehend? Is it because our minds cannot know reality? No, it is not because we are immured within a solipsist (simulated) mind-constructed view of the world. Complex realism explains that much of the contingency in knowing (causal modeling) is not because reality cannot be apprehended. Reality escapes us because it is fundamentally complex, both in terms of the real and the actual.

Second, in relation to this complexity, we have a methodological problem. **Quantitative modeling** (statistics) fails us because it does not know how to model complexity and is lost in a reductionist world of variables and parsimony. In turn, **qualitative modeling** limits itself because it cannot deal with generalization and often falls prey to problematic post-positivist ideas, such as post-modernism and radical post-structuralism. **Restrictive complexity** limits itself because it fails to actually address complexity, primarily in the form of context and contingency—that is, the manner in which things practiced are done so uniquely and done so in contextual frames, larger complex systems, etc. Complex systems are more than just rules. **Conventional case-comparative method** has all the methodological tools and the epistemological basis, but it does not have yet an explicit theory of complexity and complex systems. Finally, **equation-based modeling** cannot get beyond the dynamics of simple systems. So, what is the solution? Complex realism coupled with a generalist complexity theory coupled with case-comparative method—that is the solution. The link pin to this 'trifecta coupling' is the idea that cases are the methodological equivalent of complex systems. If reality and our knowledge of it is complex, then complexity is the issue to address. If complex systems are cases, then complex systems cannot be reduced to some set of rules or variables, and context has to be explicitly modeled. If cases are complex systems, then case-based researchers need a wider explicit vocabulary grounded in a wider set of methods, including computational modeling.

So, how do these clarifications help us contextualize the SACS Toolkit? The SACS Toolkit is part of the case-based complexity science agenda. It was designed to be the first explicit case-based modeling method designed for modeling complex social systems. Epistemologically speaking, it embraces a generalist complexity science and complex realism perspective, tempering this approach with an equal embrace of Michel Foucault's post-structuralism (which it uses to develop its theoretical framework) and Richard Rorty's neo-pragmatist understanding of the tool value of modeling; that is, scientific models of complex systems are true insomuch as they work, not because they gain a direct understanding of a complex system in its entirety. With these clarifications established, we turn now to a review of how the SACS Toolkit works.

31.3 The SACS Toolkit

The SACS Toolkit is a case-based, mixed-method, system-clustering, data-compressing, theoretically-driven toolkit for modeling complex social systems. It is comprised of three main components:

1. First, it is comprised of a theoretical blueprint for studying complex systems called **social complexity theory**. Social complexity theory is not a substantive theory; instead, it is a theoretical framework comprised of a series of key concepts necessary for modeling complex systems. These concepts include field of relations, network of attracting clusters, environmental forces, negotiated ordering, social practices, and so forth. Together, these concepts provide the vocabulary necessary for modeling a complex system.
2. Second, it is comprised of a set of case-based instructions for modeling complex systems from the ground up called **assemblage**. Regardless of the methods or techniques used, assemblage guides researchers through a seven-step process of model building—which we review below—starting with how to frame one's topic in complex systems terms, moving on to building the initial model, then on to assembling the working model and its various maps to finally ending with the completed model.
3. Third, it is comprised of a recommend list of case-friendly modeling techniques called the **case-based toolset**. The case-based toolset capitalizes on the strengths of a wide list of techniques, using them in service of modeling complex systems as a set of cases. Our own repertoire of techniques include k-means cluster analysis, the self-organizing map neural net, Ragin's QCA, network analysis, agent-based modeling, hierarchical regression, factor analysis, grounded theory method, and historical analysis.

As stated earlier, the SACS Toolkit is a variation on Byrne's [1, 11, 12] general premise regarding the link between cases and complex systems. For the SACS Toolkit, case-based modeling is the study of a complex system as a set of n-dimensional vectors (cases), which researchers compare and contrast, and then condense and cluster to create a low-dimensional model (map) of a complex system's structure and dynamics over time/space.

Because the SACS Toolkit is, in part, a data-compression technique that preserves the most important aspects of a complex system's structure and dynamics over time, it works very well with databases comprised of a large number of complex, multi-dimensional, multi-level (and ultimately, longitudinal) variables. Compression can be done using a variety of techniques, from qualitative to computational.

It is important to note, however, before proceeding, that the act of **data compression** is different from reduction or simplification. Data compression maintains complexity, creating low-dimensional maps that can be "dimensionally inflated" as needed; reduction or simplification, in contrast, is a nomothetic technique, seeking the simplest explanation possible. This distinction is crucial. At no point during the model building process is the full complexity of a system lost. Searching for the most common case-based configurations and patterns amongst the data is a way of generating a causal model, upon which the full complexity of a topic can be arranged, managed, and further data-mined. For example, while cluster analysis identifies the most common profiles in a database, one still knows which cases belong to which profiles and the degree to which they belong. Consider Fig. 31.1 as a demonstration, which comes from a study we conducted on community-level health disparities in a county

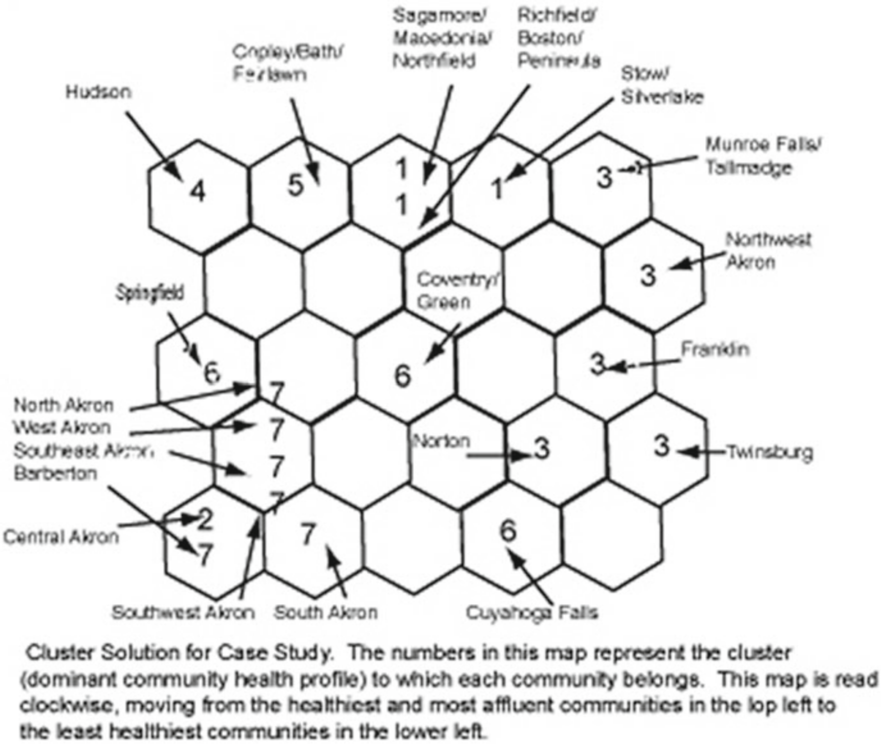

Fig. 31.1 Cluster solution for 20 communities in summit county

of 20 communities [4, 5]. In this database, we found seven clusters—each cluster represents one of the main profiles in the complex system (County) of study. Each community in Fig. 31.1 is identified by the profile (cluster) to which it belongs. One can see, however, that compression still allows us to examine every case in our database; in fact, we could (and do) go on to further cluster and differentiate any one cluster into further profile gradations, such as different types of poor communities. It all depends upon the level of granularity sought.

The SACS Toolkit is also versatile and consolidating. The strength, utility, and flexibility of the SACS Toolkit come from the manner in which it is, mathematically speaking, put together. The SACS Toolkit emerges out of the assemblage of a set of existing theoretical, mathematical, and methodological techniques, and fields of inquiry. The "assembled" quality of the SACS Toolkit, however, is its strength. While it is grounded in a highly organized and well defined mathematical framework, with key theoretical concepts and their relations, it is simultaneously open-ended and therefore adaptable and amenable, allowing researchers to integrate into it many of their own computational, mathematical, and statistical methods. Researchers can even develop and modify the SACS Toolkit for their own purposes.

31.3.1 The SACS Toolkit Updated

To date, we have written two pieces that address the SACS toolkit as a method. First, there is our book, *Sociology and Complexity Science: A New Field of Inquiry* [2], which provides a theoretical and qualitative overview, including a historically grounded case study. The second is our article *Case Based Modeling and the SACS Toolkit: A Mathematical Outline* [3], which provides a mathematical overview, including a quantitatively grounded case study. Moving from the book to the article, in addition to providing a mathematical foundation to our method, we made several major advances, most of which had to do with assembling what is called the network of attracting clusters—which we explain below. In the current chapter, we make two more minor advances. First, as we did in the previous section, we clarified the explicit links the SACS Toolkit shares with case-based complexity science and case-based modeling. In this section we make a second minor advance: we integrate the qualitative and mathematical outlines of the assemblage algorithm into a new updated version, representing current practice. Let us explain.

Given the limited space of a book chapter we felt the most useful rendition of the SACS Toolkit would be to lay out the updated assemblage algorithm, threaded with key conceptual and methodological points found in our general reviews of the SACS Toolkit. In this way we aim to provide a holistic, albeit brief, sketching of the toolkit that will hopefully give the reader enough grounding to tackle past treatments and future developments of the SACS Toolkit—interested readers could turn to our book or article for more information. And so, beginning with recasting a topic in complexity terms, through model construction and to the final leveraging of the model to answer the researcher's question(s), we now conduct a brief excursion through the Assemblage process. As a side note, we will use the following examples from our recent research: the first on medical professionalism [6, 7] and community health [4, 5]. We cite them here so we do not have to repeat them below.

31.3.2 Assembling a Complex System: A Basic Overview

The assemblage algorithm involved a series of seven basic steps. They are outlined as follows:

31.3.2.1 Converting One's Topic into a Complex Systems Framework

This first step comes in two parts. To begin, researchers need to conceptualize their topic in complex systems terms. For those relatively new to the study of complex systems, we recommend a starter text that defines the basic concepts for a complex system, such as Mitchell [15]. Conceptualizing a topic in complex systems terms means looking at your topic as a complex system/network and asking yourself such questions as (1) What will be gained by studying my topic as a complex system? (2) Do I really know what it means that my topic is emergent or self-organizing? (3) Can I thinking about my topic as a system or a network, evolving over time? Or (4) can I think of my topic in terms of the interactions amongst variables and parts, rather than the parts themselves? For example, in our study of medical professionalism we turned the concept on its head, realizing that professionalism could be thought of as a complex system, instead of a single entity. Furthermore, as a system, we saw medical professionalism comprised of several competing types (dominant profiles around which cases cluster) of professionalism, a few of which were vying for control over its future trajectory; namely, nostalgic, entrepreneurial, and lifestyle professionalism. This was an extremely novel way of thinking about professionalism. As another example, in our study of community health we examined a Midwestern county in the USA, treating its 20 major communities (cases) as a complex system—Figure 31.1, which we discussed earlier, is the cluster map for these 20 communities. Thinking about this county as a complex system was novel insomuch as it forced us to think of its 20 communities as a network, interrelated, and interdependent.

Once a topic is recast in complexity terms researchers are ready to pose a complex systems research question. This second part is just as challenging as the first, as researchers really need to make sure that the research question is a complexity science one, and not a conventional question wrapped in the new language of complexity science. For example, in our study of medical professionalism we examined how medical professionalism, as a complex system, has evolved over the last decade, including which professional types were the most dominant. This proved to be a very novel question, which only thinking in terms of complex systems could have allowed, as previous research treated

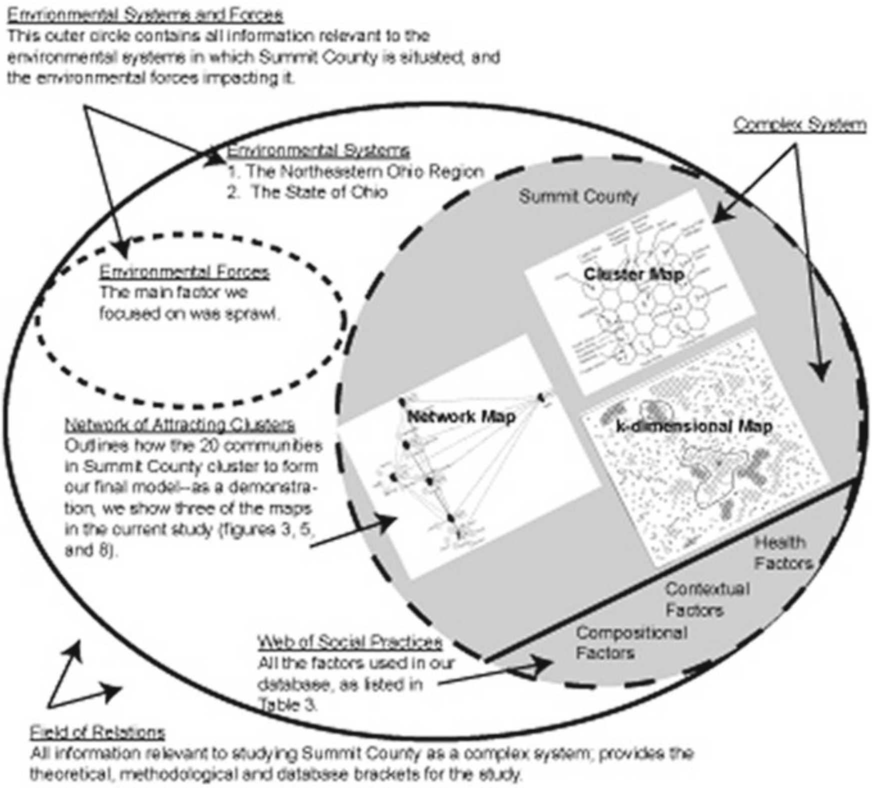

Fig. 31.2 Example of the initial model created for our community health Study

professionalism as a singular profile, and anything else as a deviation from it. As another example, in our community health study, thinking about the county as a network allowed us to ask how the health and wellbeing of the more affluent communities, across time/space, have kept the poor communities caught in a poverty trap, despite all efforts to get out of this poverty—again, a very novel question that required a complex systems viewpoint.

31.3.2.2 Building an Initial Model

Perhaps slightly counter-intuitive, the second step in the assemblage algorithm is to construct the initial model of the complex system of study. This initial model is essential because it forces researchers right from the beginning to see their topic as a complex system and to begin, albeit very fuzzily, to employ the vocabulary of social complexity theory in the building of their model. It is also crucial because it helps to define how the database for the study should be built and developed across time and what sorts of techniques to use to assemble and data-mine the model.

In our study of medical professionalism, for example, we built an early model depicting the fight in the 1990s and 2000s between three major professional types (nostalgic, entrepreneurial, and lifestyle) as they sought control over the future of medical professionalism—see our article for a picture of the model. This was very helpful because it pushed us to see how professionalism could be a system of competing types and how these different types could and were influencing the ethical behavior of physicians in the USA. As another example, as shown in Fig. 31.2, in our study of community health, our initial model was a conceptual map of our county, onto which we projected the various issues we saw contributing to the health disparities between the affluent and poor communities. This was very useful because it got us thinking about our topic as a system, helped us to identify the cases and factors we wanted to study, and helped us build our database.

31.3.2.3 Constructing the Database

Initial model in hand, the third step in the assemblage algorithm is to construct the database. The database—qualitative, numerical, or otherwise—can be usefully thought of (and, in the case of numerical data, actually assembled) as a table. In such a table, the rows are the cases. Each case is composed of a series of factors (aka variables, sets, etc). Each factor is a column. Together, these columns form the k-dimensional profile of the cases. Profiles are comprised of two types of factors: social practices and environmental forces. Social practices refer to all the factors that make up a complex system of study; environmental forces are those impacting the complex system of study; they are all put together to make the profile for the cases [3]. Construction of the database can either proceed from social practices and environmental forces to cases or from cases to social practices. Key to building the database is adopting a "data-mining" perspective: that is, researchers need to realize that modeling complex system is an iterative, evolving, and dynamic process that pushes the database to grow and change as the model is fleshed out and the framework is applied to the topic.

For example, in our study of medical professionalism, the social practices used to construct our case profiles had to do with ten key aspects of medical work. These practices included values such as altruism; skills such as interpersonal communication; economic practices such as entrepreneurialism; and personal beliefs such as an emphasis on lifestyle. In turn, environmental forces included commercialism and government regulation. In our community health study, the practices included individual factors such as household income and community level factors such as quality of school system. The main environmental force we examined was suburban sprawl.

31.3.2.4 Constructing the Field of Relations

Here, in Step 4, is where the actual model building process really gets going: creating the **field of relations**. As the term implies, the SACS Toolkit is all about relationships: the similarities and dissimilarities amongst the cases based on differences in profile; the relationships (ties, connections links) amongst the cases as a network; and the relationships amongst the factors making up the profiles. These are the three main relationships the SACS Toolkit studies, and we will visit them all again in Step 5 as we assemble the network of attracting clusters. The only difference in the next step is that we will seek to condense these relationships from full matrices to simplified maps.

The first set of relationships is referred to as the **proximity matrix**; the second as the **adjacency matrix**; and the third as the **correlation matrix**—all three matrices are algebraic terms to describe what these relationships look like mathematically. However, we use the same terms for qualitative and historical inquiry, primarily for the purposes of consistency, noting however that the construction of such matrices in these latter forms will be more loosely defined and assembled and, perhaps, even metaphorical.

Proximity Matrix: The proximity matrix is the most important of the matrices, as will be seen when map generation is discussed in the next step. Within the proximity matrix each case has a profile which shows how similar or dissimilar the case is from all the other cases. Similarity or dissimilarity is determined holistically, by looking at all of the dimensions (the complete profile) of a case and comparing them to the dimensions of other cases. Qualitative researchers can build such a matrix in any qualitative software package; quantitative studies can build such a matrix in any statistical software package. For example, in our study of medical professionalism, we used all historical and qualitative data and therefore built our profiles and sorted (compared and contrasted) them by considering a variety of cases we found in the academic medicine literature, newspapers, and through first-person qualitative interviews. In our community health study, we used an actual statistical database matrix, so the initial sorting of cases was done statistically and computationally—See Castellani and Rajaram [3] for a complete explanation.

Adjacency Matrix: Secondly there is the adjacency matrix which displays relationships or links amongst cases in the database. Think here of the new science of networks and hubs, strong ties, and small worlds, etc. The focus in this second matrix is identifying what the key relationships amongst the cases are relevant to some question of interest. In other words, whereas the proximity matrix compares all of the dimensions of one case to another to determine their similarity or dissimilarity, the adjacency matrix looks for ties or connections between two cases depending on their values on one of their shared dimensions. For example, in our medical professionalism study, we thought of the seven major professional types as a network of cases, with each case labeled according to its type and the relationships amongst these cases having to do with professional relationships of one type or another.

As our example of medical professionalism suggests, because each case contains many dimensions, it is possible to build multiple adjacency matrices. The key point in building these matrices is to look for meaningful or informative relationships among the cases that will help the researcher model their topic. For example, in our medical professionalism study, we were very interested in, from a network perspective, how students learn their professionalism, through their interactions and relationships with peers and clinical faculty. So, this has been one of the networked relationships we have been exploring.

Correlation Matrix: Finally there is the correlation matrix which consists of relationships among the social practices and environmental forces themselves. In full matrix form this means conceptualizing statistically all pairwise correlations for the factors in a study. In more practical terms, the relationships between factors could be approached with a variety of techniques or perspectives. For example, in our community health study, we were very interested in the link between micro-level residential mobility behaviors (where people moved over time) and community-level health outcomes (how residential migration patterns lead to the segregation of rich and poor communities). Whereas in our medical professionalism study we were interested in the correlation between altruism and commercialism—two of the ten factors on medical work we used to construct our profiles.

31.3.2.5 Constructing the Network of Attracting Clusters

On their own, the three matrices comprising the field of relations provide a fundamental mapping of the relationships in the database but, particularly when there are a large number of cases or dimensions (as the SAC Toolkit was designed for), are less useful for modeling the topic. Their usefulness is limited because of the overwhelming amount of data generated when analyzing all possible relationships in the database, which make identifying key relationships and patterns among the cases difficult. To overcome this limitation the next step in assemblage is to condense and compress the field of relations into a series of maps: the cluster, network, and dimensional maps, collectively known as the **network of attracting clusters**.

Compression involves taking the field of relation's different matrices and condensing all the relationships therein to a smaller set of salient and common patterns of relationships. Each matrix has its own type of map: the proximity matrix is turned into cluster maps; the adjacency matrix is turned into network maps; and the correlation matrix is turned into dimensional maps.

Together, these maps make up the network of attracting clusters, which, as the name implies, is a model that combines maps of the main profiles (clusters around which the cases cluster) in a system and the relationships these clusters and their cases share as a network, as well as the relationships amongst the dimensional factors of which they are comprised.

Cluster Maps: As we have discussed repeatedly in this chapter, the cluster maps seek to identify the most common profiles among cases in the system by grouping or clustering cases with common profiles. As shown in Fig. 31.1, cases are grouped around an attractor point, or centroid around which similar cases cluster. All cases are clustered in this way, compressing the collection of cases to a smaller collection of clusters that each contains a set of cases. A series of clusters thus enables comparison across groups of similar cases, instead of requiring that every case be compared with every other case—which is often impossible in large databases. Clustering can be done using traditional case-comparative method, Ragin's QCA or computationally, as in our usage of k-means cluster analysis and the self-organizing map neural net—see 1 and 3 in references for more information.

It is possible to generate multiple cluster maps by starting off with different types of cases or by clustering at different levels (e.g., a few big clusters or many smaller clusters). The proximity matrix can be compressed using a variety of techniques, depending upon the data used. In the case of numeric data, compression can be done using cluster analysis and the self-organizing map algorithm; for qualitative data basic sorting techniques can be used.

Note: in small databases, clustering may be very simple, amounting to little more than sorting five or seven cases. In such instances, compression may not be necessary; and so the proximity matrix and the cluster map are similar. Most databases, however, require some type of compression.

Network Maps: Network maps are directly informed by the cluster maps. Various techniques commonly used in network analysis and the new science of networks (e.g., hubs, degrees of separation, cliques, etc) are applicable for generating this kind of map. The goal is to go beyond just finding common profiles, as in Ragin's QCA, to understand how these profiles are influenced by the relationships and interactions amongst the cases. For example, in our medical professionalism study, we discussed how clinical faculties transmit their professional type to students through the hidden curriculum. Again, in a small database, the network map and the proximity matrix may be similar.

Dimensional Maps: Dimensional maps explore the relationships either between dimensions within a case or between dimensions in different cases. While the SACS Toolkit is primarily concerned with cases and relationships amongst them, dimensional maps are still useful for understanding dynamics within a case and in some instances between them, which both inform the larger focus of studying cases. Further, studying dimensional relationships across cases can inform network maps, as they suggest interactions or links between cases. Many techniques can be used to study these relationships from qualitative, statistical, agent based, to equation based. For example, in our medical professionalism study, we examined how commercialism (as an environmental force) impacted the importance of altruism in the various professional types.

Stitching the Model Together: Once the above maps have been generated, it is time to bring them together to create an integrated model. Looking over the maps, researchers at this stage should ask themselves: (1) What do these maps tell us about the model as a whole (for one point in time/space)? And (2), what do the maps tell us about the cases? Using these questions as prompts the researcher takes the three maps to create a complete network(s) of attracting clusters or single map of the given time point.

The aim for this stage is to make a map of the **negotiated ordering** of the system for this point in time/space. Negotiated ordering, as the name implies, involves the spatial and conceptual arrangement of the clusters in relation to each other, the dynamics and relationships between the clusters, how their ordering relates to the larger system in which they are part and eventually (as time points are added) negotiated ordering also incorporates the trajectory of the clusters over time. For example, Fig. 31.2 is a rough sketch of how all the maps in our model came together to create a picture of the county we studied and its 20 communities.

Furthermore, if multiple cases or cluster levels are examined or different network linkages are examined, it is also possible to make different network of attracting clusters for each time point. For example, in our medical professionalism study, we examined medical professionalism at three different levels: micro, meso, and macro, examining how our three maps came together at each level and across levels.

Time/Space: After a network(s) of attracting clusters has been completed for one time period this is repeated for other time periods, from the database to the matrices, to the initial maps and a new final network(s) of attracting clusters. This is an important point. Complexity science is ultimately about modeling complex systems across time/space.

The SACS Toolkit follows the same logic (1) Cases are not static they are dynamic and change over time. (2) Therefore, as a case based methodology, the best way to study a system is to study how the cases (or clusters) develop across time/space.

Deciding how many time-points a topic should examine should be informed by the questions the researcher is attempting to answer as well as the topic that is being studied (e.g., some systems are more stable than others, and different levels of stability may call for more time periods). To model the system across time, the network of attracting clusters is assembled on a timeline as a series of discrete time points. These discrete time points can be connected longitudinally using qualitative, computationally, statistically, or other techniques. In particular the researcher should examine the negotiated ordering of the cases/clusters present in the network(s) of attracting clusters to understand how the system is arranged and changes over time. For example, in our community health study, we examined how the 20 communities in our county evolved over a ten-year period (specifically in terms of sprawl) and the impact this had on community-level health outcomes. In our medical professionalism study, we conceptualized it as a social movement that has been evolving over the last two decades through the negotiated conflict amongst these major professional types, and in response to environmental forces and their interplay with these types.

Validity Checking: Given that just about anything in the health sciences can be seen as complex, researchers need to be careful that their study is nothing more than the same old ideas restated in the fancy language of complexity science. As such, throughout the modeling process it is necessary to do a series of validity checks. Key questions researchers should ask themselves are (1) Does modeling this topic as a system offer any substantively, theoretically, or methodologically meaningful new insights beyond conventional modes of study for this topic?; and (2) Am I forcing my topic to fit the SACS Toolkit framework or does it naturally connect to and develop with the framework? If the model passes these questions, then it is time to transition to conclude the study.

31.3.2.6 Concluding One's Study

At some point it is necessary to end model construction. Like Step 1, this last step is in two parts: drawing a study to close and answering the initial research question.

A strong signal that it is time to advance to the next part of the model or to end model construction overall is reaching the saturation point. The saturation point is the time in which adding new parts to the complex model or generating new/different maps yield marginal insights or differences from past attempts. Another sign that model construction or the present step should come to a close is when the researcher violates the validity check on forcing the model to fit the topic. If new additions or maps require forced or drawn out explanations, it is likely approaching time to end the given stage of model construction. One thing the researcher should be wary of is iterative looping disorder—the need to keep iterating on the model for fear of missing some details. Saturation or violation of the validity check is an indication that iterations should come to an end.

Finally the last step in assemblage is one that has already begun throughout model construction: answering the research question. The model settled upon as iterations on the working model come to a close become the final model, a series of network(s) of attracting clusters, used to study the topic of interest. One final point is that not all parts of the model may be necessary to answer the research question that motivated the study or some parts of the model may be more salient for the particular question.

For example, in our medical professionalism study, we realized that there is almost no end to the detail we could explore, given our model is conceptualized simultaneously at the macro, meso, and micro level. So, we have built our general model and have realized that we can go back to this model, repeatedly, to data mine it (and, also, add or develop new data) to address specific questions we want to explore, such as how, in medical schools, new students learn their professional type through theirs social networks and the larger informal and hidden curriculum in which they are situated. In our community health study, we only really addressed one key environmental force, sprawl, and its impact on community-level disparities in health. But, there are so many other forces that can be addressed, such as the evolution of the health care systems that care for the people living in our county of study.

References

1. Byrne D, Ragin C, editors. The sage handbook of case-based methods. CA: Sage; 2009.
2. Castellani B, Hafferty F. Sociology and complexity science: a new field of inquiry. Germany: Springer; 2009.
3. Castellani B, Rajaram R. Case-Based Modeling and the SACS Toolkit: A Mathematical Outline. Computational and Organizational Theory; forthcoming.
4. Rajaram R, Castellani B. Modeling complex systems macroscopically: case/agent-basedm modeling, synergetics and the continuity equation. Complexity, 2012. Online version: http://onlinelibrary.wiley.com/doi/10.1002/cplx.21412/abstract.
5. Castellani B, Rajaram R, Buckwalter G, Ball M, Hafferty F. Are communities complex systems? A case study and empirical test. Proceedings of the Center for Complexity in Health; 2011.
6. Hafferty F, Castellani B. The increasing complexities of professionalism. Acad Med, Flexner Centenary Issues. 2010;85(2):288–301.
7. Castellani B, Hafferty F. The complexities of professionalism: a preliminary investigation. In: Delese Wear, Julie Aultman's (Eds) Professionalism in medicine: critical perspectives. New York: Kluwer Academic Publishers, 2006
8. Byrne D. Complexity theory and the social sciences. London: Routledge; 1998.
9. Rihoux B, Ragin C, editors. "Configurational Comparative methods: qualitative comparative ANalsysi (QCA) and related techniques". Applied social research methods Series, vol. 51. CA: Sage; 2009.
10. Ragin C. Redesigning social inquiry: fuzzy sets and beyond. Chicago: University of Chicago Press; 2008.
11. Byrne D. Complexity, configurations and cases. Theory, Culture and Society. 2005;22(5):95–111.
12. Byrne D. Using cluster analysis, qualitative comparative analysis and invivo in relation to the establishment of causal configurations with pre-existing large-N datasets: machining hermeneutics. In: Byrne D, Ragin C, editors. The Sage handbook of case-based methods. CA: Sage; 2009. p. 260–8.
13. Blackman T. Placing health: Neighborhood renewal, health improvement and complexity. UK: The Policy Press; 2006.
14. Uprichard E. Introducing cluster analysis: what it can teach us about the case. In: Byrne and Ragin's Sage handbook of case based methods, Sage Press, CA. 2009; pp 132–147
15. Mitchell M. Complexity: a guided tour. New York, NY: Oxford University Press; 2009.
16. Richardson K, Cillers P. Special editors note: what is complexity science? A view from different directions. Emergence. 2001;3(1):5–22.
17. Klüver J, Klüver K. Social Understanding: On Hermeneutics, geometrical models and artificial intelligence. Springer 2009