1

Facets of Complex Systems

What is complex about a complex system? This seemingly simple question has neither a simple answer nor a unique one (Ladyman and Wiesner, 2020; Gell-Mann, 2002). Over the past few decades, the science of complex systems has been applied across many disciplines ranging from the physical and biological sciences to the social sciences. Given the widespread usage of the term, an all-encompassing definition is hard to come by. One such definition comes from Rosser (1999) who in turn borrowed the idea from Day (1994): "a dynamical system is complex if it endogenously does not tend asymptotically to a fixed point, a limit cycle, or an explosion." This description squarely relies on the dynamical behavior of a system to categorize it as a complex system. For our purpose, we will take a complementary approach which is considerably broader in scope and phenomenological in its essence (Ladyman et al., 2013). The idea is to not define the system *a priori*, but to describe the characteristics or different facets of complex systems, which may help us to identify a system as complex and to decipher its properties. Loosely speaking, a complex system is one that comprises a *large* number of interacting simple entities, which leads to emergent behavior at the macro level. This emergent behavior of the system cannot be reduced to the behavior of individual entities at the micro level. Complex systems, however different they may be at the micro level, have three broad classes of characteristics associated with their behavior: *heterogeneity*, *interactions*, and *emergence*.

Let us start with a simple example to elucidate the idea of a complex system. Viewed from the angle of literature on complex systems, a wheel is a "simple" entity. A car is a "complicated" object, which represents a collection of many simple entities. And traffic jam is a "complex" phenomenon – literally, not figuratively. The explanation is as follows. A wheel's functionality can be understood by analyzing simple linear dynamics. The functioning of a car, which consists of many parts each of which is simple, can be understood by taking it apart and studying the properties of each constituent part separately. However, the dynamics of a traffic jam cannot be reduced to the motion of each individual car – it is a *complex* system. Instead of looking into how cars behave in isolation, we have to look at how cars interact with each other resulting in a new emergent property called “traffic jam.” Such emergent properties are ubiquitous in nature, and they go beyond the realm of human interactions. For example, scientists noted a long time back that how an ant colony functions cannot be understood by observing a single *representative* ant (Kirman, 1993)!

However, the mere existence of a certain type of system may not necessarily imply that it is useful to build a paradigm to study its properties. We also need to know how generalizable the ideas are. It turns out that many phenomena taking place around us do seem to possess the facets of complex systems, at both a theoretical and a practical level. Therefore, a natural question arises: how do we see real-world systems through the lens of complex systems? Real-world systems are dynamic, evolving, and intertwined. Examples of such systems can be found in a wide variety of contexts ranging from the economy to living organisms to the environment to financial markets. The ways in which public opinion swings from one extreme to another, flocks of birds fly in sync, financial markets crash, and languages compete with each other like species with the emergence of dominant languages are all examples of emergent behavior at the scale of underlying systems as a result of interactions among constituent parts.

The next question, therefore, is: how do we make sense of these widely different systems and their behavior? Deciphering complexity is a daunting task as the world around us steadily assumes a more complex and interrelated form. One approach could be from a purely theoretical perspective: for example, by using toy models to establish the so-called “universal” behavior exhibited by some intricate real-world systems. In this respect, sand-pile models exhibiting self-organized criticality have been widely studied. Interacting particle models exhibiting scaling behavior in size distributions are yet another class of models that has become very popular. However, the wide variety of quantitative and qualitative behavior that is demonstrated by complex systems make it seemingly impossible to unify all possible features under one grand theory. Therefore, we follow a different approach. Instead of defining precise universal behavior, which are rare to begin with, we aim to quantify the behavioral traits that describe and/or indicate complexity and build a toolkit to achieve that objective. Fundamentally, the necessary framework to conceptualize a complex system encompasses analytical, computational, and empirical apparatus for studying dynamics that display a wide spectrum of behavior ranging from critical phenomena and complex interactions to the emergence of patterns. Therefore, the challenge appears in the form of approaching the problem with an integrative perspective, in contrast to a reductionist approach.

We pursue an empirical approach toward complex socio-economic systems. The ongoing digital transformation has brought forth new challenges to our current understanding and a new era in studying social and economic systems. A huge amount of data are being generated every moment across different spheres that can be used to understand the complex relationships between the underlying variables. And though the social and economic systems that we study may appear very different from one another, often they have similar characteristics in their dynamics. A crucial aspect of our empirical approach is the explicit attention to phenomenology in the form of a top-down zooming-in approach to deciphering complexity, rather than the prevalent bottom-up approach of describing assumptions and theoretically working out the resulting aggregate implications. In the context of complexity-driven economics, especially macroeconomics, a similar point was made by Di Guilmi et al. (2017), although they differ from us in terms of their emphasis on theory. In the empirical portion of this book, we describe the top-down approach to characterizing macroscopic behavior. Once having built insights into the characterization of macroscopic behavior, we provide a bottom-up approach to arrive at such behavior from microscopic interaction rules similarly to the reductionist approach that connects the behavior of a system with its constituent parts. However, these two approaches retain their differences which arise in the form of emergence. As opposed to the sum of all the smallest component-level behavior representing the macroscopic behavior, we emphasize systems where the macroscopic behavior is more than the sum of its parts.

The nature and features of complexity differ across socio-economic, physical, and biological systems. For quite some time, scientists have gathered sizable amounts of data on physical and biological systems. In recent times, the economy and human society have gone through a revolution due to the explosive growth of information science. This, in turn, is facilitated by rapid technological advances, yielding a high volume and a tremendous variety of data in multiple domains at any given point of time. The big challenge of the present day is in extracting meaningful insights from the data in ways that go beyond data-crunching and data-engineering. More clearly pinned-down rules of interactions governing the dynamics of social and economic agents will lead to a clearer understanding of the corresponding emergent properties. This conjunction is our main thesis.

To explore the emergence of macroscopic behavior from microscopic interactions, we focus on some of the most pressing problems of the present-day world, ranging from social segregation and economic inequality to the extinction of languages, among others. We argue that the paradigm of complex systems can shed light on these phenomena. Clearly, a lot more data are available now to study such systems than was available earlier. We argue that the features of heterogeneity, interactions, and emergence are embedded in each of the systems we examine here, which make the systems complex. With this understanding, we can make sense of the connections between microscopic and macroscopic properties. While this description of complexity is less stringent than other definitions (e.g., see the discussion in Ladyman et al. [2013]), it allows us to simplify our exposition considerably. With that in our view, we will first explain the defining features of complex systems in terms of measurable empirical properties.

1.1 Features of Complex Systems

Some of the early attempts to understand complex systems came from the physics literature and there is a specific reason for that. A natural feature of a complex system is its dynamics, and the study of the mathematics of dynamical systems in the natural world found a home in the physics literature. Such dynamics can be either stochastic or deterministic (including chaotic), often displaying nonlinear and emergent phenomena. To capture system-level dynamics by aggregating the dynamics of all of a system's constituent entities is a non-trivial and often intractable problem. Moreover, the constituents of complex systems are heterogeneous and there can be time-varying interactions between them. Most of the formalisms of statistical physics that are used to understand macroscopic or collective behavior from the dynamics of the microscopic constituents have insufficiencies in this regard. One fascinating example where the application of such theories has been very successful is the type of systems that exhibit self-organization and critical behavior. Notably, the famous "sand-pile" models in physics have been very useful for understanding the universal behavior arising out of such systems. However, as we will discuss later, a direct mapping of such systems on to real-world phenomena has had very limited success. To summarize, the standard approach used in statistical physics has its limitations, although there has been enormous progress in terms of the theoretical understanding of the dynamical properties of complex systems. In order to develop these ideas further, we first need to describe the features of complex systems beyond their dynamics.

The first feature is the stability of an interacting dynamical system. As we have noted above, complex systems can be large and the constituent entities may display heterogeneity. This, in turn, influences the stability of the system. It is probably easiest to grasp this idea with an intuition from ecology. A dominant paradigm in the ecology literature has been centered around the idea that diverse systems would be more stable than their less diverse counterparts. Intuitively, a more heterogeneous system will be able to adapt more to adverse shocks and therefore should display more stability. However, Robert May's work famously turned this idea upside down when he showed that a larger and more diverse system can be actually prone to more instability, rather than stability (May, 1972). The stability of a system also relates to the linkages between its constituent entities. Are more linkages good or bad for stability of a system? May's work showed that more linkages can be bad for stability. A similar insight also carries through in the context of financial markets. More connected financial markets provide more opportunities to diversify and thereby, potentially, create more stability. However, the recent literature recognizes that too many interlinkages may lead to hidden feedback loops, which can destabilize the system rather than imparting stability to it (May et al., 2008; Haldane and May, 2011).

What is the role of interlinkages beyond influencing the stability of a system? To understand this, we need the help of the network view of the system, a useful paradigm for capturing the topological characteristics of a system with non-trivial interlinkages (Newman, 2010). The main understanding comes from the idea that for many (if not most) socio-economic systems, the connections between the constituent entities do not display symmetry and regularity. Connections are heterogeneous and so are their corresponding influences on the system-level properties (Page, 2010). For example, consider a growing online social network. It is dynamic and heterogeneous in its connections across all the users and the diffusion of a piece of news on the network may heavily depend on its topology. More importantly, different social media may actually exhibit non-trivial differences in topology, resulting in differences in macroscopic behaviors such as the diffusion of a piece of news or a rumor. Thus the nature of interlinkages is a crucial component of the complexity of the system.

Competition and the emergence of dominant traits represent another key feature of complex systems (Chakrabarti and Sinha, 2016). One can conceptualize this mechanism via heterogeneity in the macro-behavior of the system. It has long been recognized that even when there is little heterogeneity in core characteristics among the constituent entities of a given system, the outcomes for each of them may diverge widely. An oft-cited example of such a system is income inequality. All characteristics of human beings are possibly well approximated by bell curves – normal distributions, which have a small variance. However, the corresponding income or wealth distribution among the people interacting through the economy has a variance many times larger than the underlying variances in characteristics of the people. The famous Pareto law was first found in income distribution itself, which says that income m is distributed as a power law $p(m) \sim m^{-(1+\gamma)}$ where γ is the Pareto coefficient (typically with a magnitude close to 1, although there are notable exceptions). We will explore this law in much more detail later on when we study size distributions in income and cities. For the time being, one can see the implication of the law in the following way: going by this law with the coefficient being equal to 1, 20% of the entities are responsible for 80% of all events. In this case, that translates into the idea that 20% of the people acquire 80% of the total income. In the context of the present-day world, this degree of heterogeneity and inequality is fairly accurate in an empirical sense although the exact magnitude may differ. Such heterogeneity is seen not only in economic contexts but also in scientific paradigms where citations are power-law distributed and in many other systems that we will describe later. Chakrabarti and Sinha (2016) observed a unique case with movie income distributions, which actually possess bimodal characteristics where success and failure literally translate into two different modes of the outcome distribution. In the extreme, an asymptotic case of inequality in outcomes is seen in the case of language dynamics: some languages may actually die completely, which leads to the emergence of extreme dominance by the remaining languages.

1.2 A Data-Driven View of Complexity

In what follows, we adopt a data-driven view of complex systems. Our focus will be less on the mathematical modeling of complex systems, and much more on the empirical analysis of the resulting behavior of such systems. Recent advances in computation for solving empirical problems have pushed the frontier of science by leaps and bounds. Google's AlphaFold is possibly one of the most impressive achievements in discovering the structures of proteins. This builds on a series of inventions in computer hardware and software over the last century – from the breaking of the Enigma code during the Second World War to IBM's Deep Blue defeating the famous chess grandmaster Garry Kasparov, from IBM's Watson to Google's AlphaZero to the latest technology ChatGPT from OpenAI which can mimic human conversation. This has been possible due to massive advances not only in computational ability but also in data management. At the outset, this may not look like a challenging problem. But often data management turns out to be the biggest practical challenge in computation. One of the great resources has come in the form of shared and distributed computing and data annotation. As we will see in Chapter 4, many learning algorithms require labeled data. Many software packages and websites, among them GitHub open codes, Google Colab, and Amazon Mechanical Turk, have helped the transition from single-user to multi-user crowdsourced computation and data annotation. In what follows, we will not data management any further and will focus almost exclusively on data analysis and modeling.

We consider data collected at different frequencies: low-frequency data (e.g. societal changes spanning centuries), mid-frequency data (e.g. economic data collected at business cycle frequencies in the order of a few years), and high-frequency data (e.g. financial data collected daily or even more frequency). We will refer to such data as time series data. The other dimension of the data would be its heterogeneity. A set of observations at any given point of time would allow us to conceptualize and quantify heterogeneity. We will refer to such data as cross-sectional data.

In order to develop the empirical apparatus, we first provide an introduction to probability and statistics, covering both classical and Bayesian statistics (Chapter 2). We go through a discussion of the classical approach to probability and develop the concept of statistical estimation and hypothesis testing, leading to a discussion of Bayesian models. Then we discuss time series models to analyze evolving systems (Chapter 3). In particular, we develop ideas to model stationary and non-stationary systems. Additionally, we review some ideas from financial econometrics that have proved to be very useful for modeling time-varying conditional second moment: that is, volatility. In the next part of the book, we review machine learning techniques emphasizing numerical, spectral, and statistical approaches to machine learning (Chapter 4). Then we discuss network theory as a useful way to think about interconnected systems (Chapter 5). These four components constitute the building blocks of the data science approaches to complex systems.

1.3 A World of Simulations

There is no single unique framework for studying emergent behavior, and as mentioned above, we often have to borrow tools from several domains such as economics, mathematics, statistics, computer science, and physics. In some sense, the abstract and non-unique nature of this mixed bag of tools is quite useful as one can expand or compress the scope of analysis as necessary.

Due to the property of emergence, for many of the complex systems and their dynamical patterns, characterization of local rules may not directly capture systemic behavior. Whether we are studying self-organization in economic markets, the dynamics of ethnic conflicts and cooperations, migration networks, or the spread of epidemics, the modeling approach has to emphasize the varying scales and nature of interactions. However, there is another dimension to this problem. Along with making predictions for the expected behavior of the system, one may also want to see what are the possible extreme cases. More generally, one may want to first figure out what are the boundaries of possible behavior of a given system. This is very important for systems that are prone to sudden collapse. Markets, in particular financial markets, provide a great example of such behavior. For policy-makers or even for market participants, a better understanding of large fluctuations is typically more useful, for hedging against extreme outcomes, than the prediction of average behavior.

How can we delineate the boundaries of all possible behaviors of a system consisting of a large number of interacting entities, each with potentially non-trivial

https://doi.org/10.1017/9781108953597.002 Published online by Cambridge University Press

dynamics? This poses a big challenge. Analytically enumerating all possible behaviors is a daunting task and often impossible for all practical purposes. The complex systems literature has taken an alternative path to solving this problem. Rather than trying to do brute-force calculations by ourselves, it is easier to ask the computer to do so given the cheapness of computational power. The idea is to simulate possible behaviors with simple iterative rules governing interactions between individual agents, which can be tuned to mimic the dynamics of emergent phenomena. Additionally, this approach is useful for finding the average behavior of systems which are analytically intractable.

In order to develop insights into real-world socio-economic systems, in the later part of the book we describe six models that illuminate different facets of complex systems. We start with a model of social segregation which introduces a spatial model of interaction between agents following the seminal work by Schelling (1971). This is an elementary model that delivers the astounding insight that even a minor preference for a homogeneous neighborhood in a heterogeneous population may quickly result in strong segregation of the population. At the same time, this model also started the literature on interaction models of grids or lattices, connecting it to the idea of cellular automata. Famous examples of this kind of model include Conway's Game of Life model and cellular automata model. Interested readers may consult Wolfram (1984, 2002) for an in-depth view of such systems in the context of discrete mathematics. Next, we describe a more intricate scaling behavior on a square lattice in the form of the Bak–Tang–Wiesenfeld model (Bak et al., 1987; Bak, 2013) that exhibits self-organized criticality through the famous simulated sand-piles. A follow-up description of the Bak–Sneppen model sheds light on extinction through competition in a multi-agent setup. In particular, this model generates the scaling behavior of avalanches which models extinctions. A simple time series analysis shows periods of bursts and calm, capturing *punctuated equilibrium* behavior – reminiscent of the theory proposed by Gould and Eldredge (1977). Building on these abstract spatial models, next we focus on city size distribution which possesses a well-known defining feature in the form of Zipf's law, a special case of the Pareto law described above. This law provides a natural restriction on the spatial organic growth of cities. We study two different mechanisms that explain the appearance of Zipf's law. Then we introduce bilateral interaction between agents to explain the dispersion of agent-level attributes. As a modeling paradigm, we then take up models of asset inequality. Empirically, such inequality seems to exhibit robust behavior across economies and time especially in the right-hand tails of income and wealth distributions, in the form of Zipf's law. We show that heterogeneity in general and Zipf's law in particular arise out of interactive systems. Next, we consider competition across groups of agents. As a modeling paradigm, we take up the case of linguistic competition where the emergence of dominant and recessive languages occur as an outcome of competition. Until this point, all of the models we have discussed can be thought of as representing interactions between zero-intelligence agents. In the last model, we consider agents with limited intelligence or rationality who can strategically compete against each other. We analyze a resource competition game to study the evolution of coordination and anti-coordination. We close the discussion with a brief detour through the trade-off between the realism and the generalizability of these models and corresponding insights.