2010-12-15

Dynamical Grouping in Complex Systems

Zhenyuan Zhao
University of Miami, zhenyuanzhao@gmail.com

Recommended Citation

Zhao, Zhenyuan, "Dynamical Grouping in Complex Systems" (2010). *Open Access Dissertations*. 498.
https://scholarlyrepository.miami.edu/oa_dissertations/498

UNIVERSITY OF MIAMI

DYNAMICAL GROUPING IN COMPLEX SYSTEMS

By

Zhenyuan Zhao

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2010

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

DYNAMICAL GROUPING IN COMPLEX SYSTEMS

Zhenyuan Zhao

Approved:

Neil F. Johnson, Ph.D.
Professor of Physics

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Joshua Cohn, Ph.D.
Professor of Physics

Joseph Ashkenazi, Ph.D.
Associate Professor of Physics

Chris Cosner, Ph.D.
Professor of Mathematics

ZHAO, ZHENYUAN (Ph.D., Physics)

Dynamical Grouping in Complex Systems (December 2010)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Neil F. Johnson.
No. of pages in text. (183)

Quantifying the behavior of complex systems arguably presents the common "hard" problem across the physical, biological, social, economic sciences [1]. Individual-based or agent-based models have proved useful in a variety of different real world systems: from the physical, biological, medical domains through to social and even financial domains. There are many different models in each of these fields, each with their own particular assumptions, strengths and weaknesses for particular application areas. However, there is a lack of minimal model analysis in which both numerical and analytic results can be obtained, and hence allowing different application domains to be analyzed on a common footing.

This thesis focuses on a few simple, yet highly non-trivial, minimal models of a population of interacting objects (so-called agents) featuring internal dynamical grouping. In addition to analyzing these models, I apply them to a number of distinct real world systems. Both the numerical and analytical results suggest that these simple models could be key factors in explaining the overall collective behavior and emergent properties in a wide range of real world complex systems. In particular, I study variants of a particular model (called the EZ model) in order to explain the attrition time in modern conflicts, and the evolution of contagion phenomena in such a dynamically evolving population. I also study and explain the empirical data obtained for online guilds and offline gangs, leading to a team-based model which captures the common quantitative features of the data. I then move on to develop a resource competition model (i.e. the so-called El Farol model) and apply it to the carbon emissions market, mapping the different market factors into model parameters which enable me to explore the potential market behaviors under a variety of scenarios.

To my beloved parents for their endless love,

understanding, support and encouragement

Acknowledgements

I wish to express my sincerest gratitude to Professor Neil F. Johnson, who introduced me into the field of complex system and group-dynamics models. His continuous guidance and inspiration always helped me break through the difficulties in my research. I would also like to thank Professor Pak Ming Hui, Alex Dixon, Juan Camilo Bohorquez, J.P. Calderón, Chen Xu, Guannan Zhao, Jing Meng, Daniel J. Fenn, Riley Crane, B. Ruszczycki, B. Burnett, Choe Charley Sehyo, and all other colleagues, for their valuable discussions and productive collaboration.

I would like to extend my thanks to the faculty and staff members of the Physics Department at the University of Miami for their help with my education and for their efforts to make various resources available to me, and all my other friends for their help and collaboration during my Ph.D. studies.

Zhenyuan Zhao

University of Miami

December 2010

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Complex Systems - together with their dynamical behavior known as Complexity - are thought to pervade much of the natural, informational, sociological, and economic world [2]. A unique, all-encompassing definition of a Complex System is lacking - worse still, such a definition would probably end up being too vague. Instead, such Complex Systems are better thought of in terms of a list of common features which distinguish them from "simple" systems, and from systems which are just "complicated" as opposed to being complex. Although a unique list of Complex System properties does not exist, most people would agree that following would typically appear: feedback and adaptation at the macroscopic and/or microscopic level, many (but not too many) interacting parts, non-stationarity evolution, coupling with the environment, and observed dynamics which depend upon the particular realization of the system. Understanding the functionality of Complex Systems is of paramount importance, from both practical and theoretical viewpoints.

In a similar way to how many-body physics [3] developed to explain novel collective phenomena in solid state physics, physicists in the Complex Systems community have focused on explaining the appearance of novel non-Gaussian (e.g. power-law) statistical distributions which have been observed in many real-world, candidate complex systems (e.g. financial markets, conflict, traffic). In particular, the fact that such a distribution has not developed the typical Gaussian form of the Central Limit Theorem, means that it is natural to assume that rather non-trivial interactions and feedback must be at work on the microscopic level. Within mathematics, much headway into understanding the possible effects of clustering within a population (i.e. beyond mass action) have been obtained by adding functional response terms to differential equation descriptions of population models [4]. Within physics, the emphasis has instead been on looking for many-body type models inspired from condensed matter in which the output is a statistical distribution which is consistent with the observed empirical data. Two general types of model have been presented within the Physics community to deal with such strong internal correlations, or group dynamics: the first type, upon which most of this thesis focuses (Chaps. 2-5), is one in which agent interactions lead to a dynamically evolving set of internal groups. Members of a given group may then share opinions, viruses etc.. They may therefore imitate each other and thereby seem to act as a crowd. Such a crowding has been termed "herd formation" in the Physics literature. The second type is the EL Farol model (which is used in Chapter 6) in which it is assumed that the individual objects, or "agents", do not share any local information. The crowds are now global and unintentional in the sense that any agent may suddenly became a member of particular crowd via the strategy he is using, rather than because he decides to join that crowd per se - i.e. crowding results in strategy space. Irrespective of the type of model, our goal throughout this thesis is to build minimal yet justifiable models which are consistent with observed phenomena.

In a broader context, it is interesting that many different disciplines across the biological and social sciences face a “many-body problem” in one form or another, in which it is necessary to provide a framework for understanding the collective behavior of assemblies of interacting particles. As all fields would concur, the study of such a collection of particles is in general extremely complicated [2] with the emergent phenomena often bearing little resemblance to the underlying elementary laws. In Physics, many-body theory has been successful in synthesizing the behaviors which govern the macroscopic realm – in part, because the microscopic laws governing individual particles are so well known. Starting with Boltzmann [5], the idea of “emergent phenomena” has existed in physics for a long time, ranging from solid state physics [6] (e.g. spin glass [7], superconductivity [8]), quantum mechanics [9] through to high energy physics [10]. Many analytical and computational methods have been developed to attack the many-body problem [3].

Here in this thesis, the idea is to apply a similar philosophy to attack non-conventional real-world systems (see the following chapters). Instead of elementary particles, there could be proteins, polymers, human beings, groups, computers, websites, and so on. In addition, the microscopic interactions are in general much more versatile and complicated (e.g. history dependent). For example, a student in a discussion session, is interacting with the classmates and professor, while at the same time, the cell phone and internet enable him/her to interact with basically anybody in the university, country or even the whole world – and the way these interactions play out can depend on the specific point in time and the previous days’ events. Moreover, when the class is over, some of interactions stop while other interactions (e.g. personal links) get established. Such dynamics is a core theme of this thesis – in particular, the grouping dynamics. The novel feature of such a complex system for a physicist, is that it represents a many-body system of interacting particles (or so-called “agents”) which may or may not identical and whose interactions may now depend on time and space. The agents are typically alive and have some primitive ability to evolve, adapt, and hence think, learn, and decide, according to either the local and/or global information or interactions. These distinguishing features of agents will add a new level of heterogeneity into the system beyond that of a typical physical system.

One could argue that the behavior of complex systems presents the common "hard" problem across the physical, biological, social, economic sciences [1]. To add to the level of complexity, many, if not most, complex systems are influenced by some underlying network structure [11]. Early research on the networks made great strives into understanding the properties of static, equilibrium networks, such as grid lattice with constant or Gaussian distributed connectivity [12]. It is only recently that the study of "complex network" has become a major research area within the physics community. The explosion of research within the physics literature has led to the realization that similar network structures appears across a wide range of biological and social systems [11, 13–19]. The theoretical work to date has been rather abstract, and in the explicit case of complex networks, has focussed on the static, macroscopic structural properties of networks [20, 21]. On one hand, many works have been devoted to the studies of static (or semi-static with slow rewiring) lattice type networks, a multitude of measures have been developed, distinguishing and characterizing the various topological properties of complex networks: random [12, 22], small-world [15] or scale-free networks [13, 14]. On the other hand, much research has been performed using mean-field type mass action models, where the underlying population is assumed to be well mixed [23, 24].

Unfortunately, these previous studies offer relatively little insight into the dynamical properties of the networks associated with real-world complex systems. In particular, for networks associated with a population of interacting objects, it is how this network structure is affected by the dynamical behavior of the population, and how the population's behavior then feeds back on this network, which is of importance. For example, from a biological perspective, knowing that a given complex system network is scale-free is not, per se, that interesting. Instead it is the microscopic dynamics that lead to the macroscopic behavior, which is important. In short, it is how agents function with each other, and how the population evolves, that matters. So far, the precise nature of the interplay between dynamics and structure in complex network-based systems remains largely a mystery. These shortcomings in the current state of understanding of the dynamical properties of complex systems and networks provide the motivation for the research presented in this thesis.

1.2 Overview

Our main research interest lies in understanding the dynamics of groups in complex systems, and the effect on collective behavior. How do large systems evolve locally? How do different grouping dynamics change the overall shape of systems (e.g. cluster size destruction)? How do rumors, information, or viruses spread among the systems? Answers to these questions are vital to a range of application areas from forming efficient teams, through to helping prevent potential conflicts, recognizing the activities in financial market and detecting disease outbreaks. A basic premise behind the study of complex systems is that local and/or global interaction leads to complex collective behavior. We study, through several real-world cases from different fields, how group dynamics affect the evolution of large complex systems, and the emergent complex behavior. The following is an outline of the chapters in this thesis, which are all based on our published papers:

In Chapter 2, we start by studying the theoretical coalescence-fragmentation models, which are now of great interest across the physical, biological, and social sciences. They are typically studied from the perspective of rate equations, at the heart of which are the rules used for coalescence and fragmentation. The model is also sometimes referred as a local interaction model. Here we discuss how changes in these microscopic rules affect the macroscopic cluster-size distribution which emerges from the solution to the rate equation. Our analysis elucidates the crucial role that the fragmentation rule can play in such dynamical grouping models. We focus our discussion on two well-known models whose fragmentation rules lie at opposite extremes. In particular, we provide a range of generalizations and new analytic results for the well-known model of social group formation developed by Eguíluz and Zimmermann [V. M. Eguíluz and M. G. Zimmermann, Phys. Rev. Lett. **85**, 5659 (2000)]. We develop analytic perturbation treatments of this original model, and extend the analytic analysis to the treatment of growing and declining populations.

A remarkable recent application of the same coalescence - fragmentation has recently appeared in the physics community. In particular, it is able to explain a mysterious recent finding concerning quantum many-body effects in the high-temperature superconducting (SC) cuprate system $\mathrm{La_2CuO_{4+y}}$. [25, 26]

- "Relating the microscopic rules in coalescence-fragmentation models to the cluster-size distribution", *Eur. Phys. J. B* **72**, 289 (2009), B. Ruszczycki, B. Burnett, Z. Zhao, and N. F. Johnson.

- "Dynamical Clustering as a Generator of Complex System Dynamics", *Math-*

ematical Models and Methods in Applied Sciences, **19, Suppl.** 1539 (2009), Zhenyuan Zhao, Andy Kirou, B. Ruszczycki, Neil F. Johnson.

- "Modeling the self-assembly dynamics of macromolecular protein aggregates underlying neurodegenerative disorders", *Journal of Computational and Theoretical Nanoscience* **6**, 1 (2009), Zhenyuan Zhao, Rajiv Singh, Arghya Barman, Neil F. Johnson, and Rajeev Prabhakar.

In Chapter 3, we adapt the grouping dynamic model study to social systems – in particular, modern human conflicts, such as those ongoing in Iraq, Afghanistan and Colombia, which typically involve a large conventional force (e.g. a state army) fighting a relatively small insurgency having a loose internal structure. Here we adopt this qualitative picture in order to study the dynamics – and in particular the duration – of modern wars involving a loose insurgent force. We generalize a coalescence-fragmentation model from the statistical physics community in order to describe the insurgent population, and find that the resulting behavior is qualitatively different from conventional mass-action approaches. One of our main results is a counterintuitive relationship between an insurgent war's duration and the asymmetry between the two opposing forces, a prediction which is borne out by empirical observation.

- "Anomalously Slow Attrition Times for Asymmetric Populations with Internal Group Dynamics", *Phys. Rev. Lett.* **103**, 148701 (2009), Zhenyuan Zhao, Juan Camilo Bohorquez, Alex Dixon, and Neil, F. Johnson.
- "Statistical Physics and Modern Human Warfare", invited book chapter, in Mathematical Modeling of Collective Behavior in Socio-economic and Life-sciences, Alex Dixon, Zhenyuan Zhao, Juan Camilo Bohorquez, Russell Denney, and Neil Johnson.

In Chapter 4, we study the diffusion/spreading on the dynamic model in order to shed light on the role of network dynamics on the evolution of processes on a network. Despite the many works on contagion phenomena in both well-mixed systems and heterogeneous networks, there is still a lack of understanding of the intermediate regime where social group structures evolve on a similar timescale to individual level transmission. We address this question by considering the process of transmission through a model population comprising social groups which follow simple dynamical rules for growth and break-up. Despite the simplicity of our model, the profiles produced bear a striking resemblance to a wide variety of real-world examples – in particular, empirical data that we have obtained for social (i.e. YouTube), financial (i.e. currency markets), and biological (i.e. colds in schools) systems. The observation of multiple resurgent peaks and abnormal decay times is qualitatively reproduced within the model simply by varying the timescales for group coalescence and fragmentation. We provide an approximate analytic treatment of the system and highlight a novel transition which arises as a result of the social group dynamics.

- "Effect of social group dynamics on contagion", *Phys. Rev. E* **81**, 056107 (2010), Zhenyuan Zhao, J.P. Calderón, Chen Xu, Guannan Zhao, Dan Fenn, Didier Sornette, Riley Crane, Pak Ming Hui, and Neil F. Johnson.

- "Strong dependence of infection profiles on grouping dynamics during epidemiological spreading", Complex Sciences First International Conference, Complex 2009, Shanghai, China, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, ISSN 1867-8211 (2009), Zhenyuan Zhao, Guannan. Zhao, Cu Xu, Pak Ming Hui, and Neil F. Johnson.

In Chapter 5 we further generalize the multi-agent model to quantify the human group dynamics. Unlike animals and other biological systems, humans form groups in both real (offline) and virtual (online) spaces from potentially dangerous street gangs populated mostly by disaffected male youths, through to the massive global guilds in online role-playing games for which membership currently exceeds tens of millions of people from all possible backgrounds, age-groups and genders. We have compiled and analyzed data for these two seemingly unrelated offline and online human activities, and have uncovered an unexpected quantitative link between them. Although their overall dynamics differ visibly, we find that a common team-based model can accurately reproduce the quantitative features of each simply by adjusting the average tolerance level and attribute range for each population. By contrast, we find no evidence to support a version of the model based on like-seeking-like (i.e. kinship or "homophily").

- "Human group formation in online guilds and offline gangs driven by common team dynamic", *Phys. Rev. E* **79**, 066117 (2009), Neil F. Johnson, Chen Xu, Zhenyuan Zhao, Nicolas Ducheneaut, Nicholas Yee, George Tita, Pak Ming Hui.

In Chapter 6, we investigate a global interaction model, and apply it to the problem of controlling global carbon emissions. We explore quantitatively how different control schemes affect the collective emission dynamics of a population of emitting entities. We uncover a complex trade-off which arises between average emissions (affecting the global climate), peak pollution levels (affecting citizens' everyday health), industrial efficiency (affecting the nation's economy), frequency of institutional intervention (affecting governmental costs), common information (affecting trading behavior) and market volatility (affecting financial stability). Our findings predict that a self-organized free-market approach at the level of a sector, state, country or continent, can provide better control than a top-down regulated scheme in terms of market volatility and monthly pollution peaks.

- "Self-organized global control of carbon emissions", *Physica A* **389**, 17, 3546 (2010), Zhenyuan Zhao, Dan Fenn, Pak Ming Hui, and Neil F. Johnson.
- "Competitive carbon emission yields the possibility of global self-control", *J. Comput. Sci.* (2010 in press), Daniel J. Fenn, Zhenyuan Zhao, Pak Ming Hui, and Neil F. Johnson.

CHAPTER 2

Simple Model of An Interacting Dynamical Population

2.1 Background

The challenge to understand the dynamics of Complex Systems is attracting increasing attention, particularly in the socio-economic and biological domains [27, 28, 54–57,59–72]. For example, the recent turmoil in the financial markets has created significant public speculation as to the root cause of the observed fluctuations. At their heart, all Complex Systems share the common property of featuring many interacting objects from which the observed macroscopic features emerge. Exactly how this happens cannot yet be specified in a generic way – however, an important milestone in this endeavor is to develop a quantitative understanding of any internal clustering dynamics within the population. Coalescence-fragmentation processes have been studied widely in conventional chemistry and physics [73–94] – however, collective behavior in social systems is not limited by nearest neighbor interactions, nor are the details of social coalescence or fragmentation processes necessarily the same as in physical and biological systems. The challenge for a theorist is then twofold: (1) to provide a model which accounts correctly for the observed real-world behavior — e.g., in the case that power-laws are observed empirically, the model should be able to reproduce the power-law dependence itself, the value of the corresponding power-law exponent, and possibly also the form of the truncation; (2) the rules invoked in the model need to make sense in the context of the real-world system being discussed.

In this chapter, we discuss coalescence and fragmentation problems with a focus on social systems. In particular, we consider interactions which are essentially independent of spatial separation in order to mimic the effect of modern communications etc. Much of our discussion is focused around fragmentation processes in which an entire cluster breaks up into its individual pieces – thereby mimicking a social group disbanding – as opposed to the more typical case studied in physical and biological systems of binary splitting. We limit our discussion to the steady-state behavior corresponding to a constant population, or a steadily growing/declining population. In Sec. 2.1.2, we lay out a general formulation of such coalescence-fragmentation problems. In order to understand the quantitative effects of a particular choice of fragmentation rule, Sec. 2.2 then compares two well known coalescence-fragmentation models, with fragmentation rules which lie at opposite extremes of the spectrum. One of these is the well-known physics-inspired model of social group formation introduced by Eguíluz and Zimmermann [28] while the other is a standard model in mathematical ecology due to Gueron and Levin [54]. The explicit comparison between the two models allows us to elucidate the subtle differences in their microscopic rules that make their macroscopic distributions differ, and leads us to a better generic understanding of the crucial role that the fragmentation rule can play. We then proceed to focus on the physics-based model of Eguíluz and Zimmermann, generalizing it in several ways and providing new analytic results (Sec. 2.3). We analyze a perturbed version of the Eguíluz-Zimmerman model where spontaneous cluster formation is present (Sec. 2.3.1), as well as generalized versions in which there is a steadily growing (Sec. 2.3.3) or declining population (Sec. 2.3.3). Further realistic modifications of the Eguíluz-Zimmerman model are discussed in Sec. 2.3.4.

There is of course a huge volume of work in the mathematics, physics and chemistry literature on the topic of clustering within a many-body population of interacting particles [95]. The Smoluchowski coalescence equation is arguably the most famous and well-studied example [73–75]. Many of the existing studies of coalescence-fragmentation models, however, focus on distinct fragmentation mechanisms to the present EZ one studied in this chapter. This is because applications have focused on chemical kinetics, where the likely fragmentation of a polymer containing s units is simply to lose one monomer and hence produce an $s-1$ polymer – or where the polymer breaks into two identical pieces, or two pieces of general size (e.g., Gueron Levin Model discuessed in this chapter). Reference [95] provides an excellent recent review of coalescence-fragmentation models in physical and chemical systems from a mathematician's perspective – however we note that the socially-inspired models that we focus upon in this article are not discussed. Many previous studies have tended to focus on generic mathematical issues such as existence, uniqueness, mass conservation, gelation and finite size effects (see Refs. [76–80] and references therein). When it comes to Complex Systems – and in particular, social systems – the more pressing goal is to understand the emergent features of the population. In contrast to physical and chemical systems in which collision energetics play a crucial role in guiding the specification of microscopic coalescence and fragmentation rules, the precise microscopic rules in social systems are unknown – however, the overall macroscopic emergent phenomena such as cluster size distribution can be measured relatively easily. In financial markets, the collective dynamics of the population of traders is registered directly by means of the price. Indeed, as many prior works have shown, such collective behavior in social systems tends to produce near scale-free (i.e. power-law) networks and/or cluster sizes in a variety of real-world situations. For example, the distribution of transaction sizes follows a power-law with slope near 2.5 for each of the three major stock exchanges in New York, Paris and London [96]. In addition, it has been shown that the distribution of the severity of violent events inflicted in conflict by insurgent groups, and by terrorist groups, follow a power-law near 2.5 [56,57,59,97]. The model of Eguíluz and Zimmermann [28], which is the starting point of much of the paper's discussion, is therefore an attractive candidate model for such social systems. In addition to its intrinsic theoretical interest because of its non-binary fragmentation rule, which mimics the disbanding of social groups, it also happens to produce a robust power-law distribution of cluster sizes with slope 2.5 [28].

2.1.1 Modeling Social Systems

Many social systems seem to comprise a large number of dynamically evolving clusters. Over time, and in an apparently self-organized way, clusters either coalesce with each other to form even larger clusters, or fragment to form a collection of smaller ones. In addition to everyday social situations, these characteristics seem consistent with common sense notions of the dynamical connectivity within a community of financial traders [28], or even a loosely connected insurgent population or terrorist/criminal network [56,59]. Figure 2.1 illustrates the generic situation of interest in many recent works on coalescence-fragmentation models [28, 54, 60–72]. As a result of coalescence and fragmentation processes over time, the population of N objects undergoes dynamical partitioning into clusters $i, j, k, \ldots$ of size $s_i, s_j, s_k, \ldots$, where both the number of clusters and their membership are typically time-dependent. We have denoted the N objects in human form, but of course they could be animals,

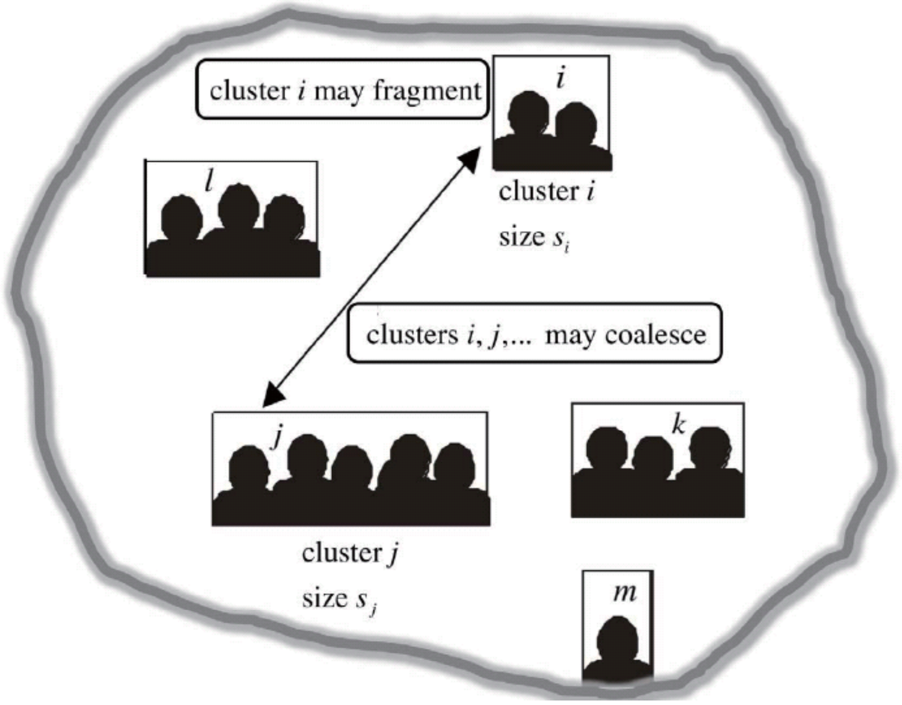


Figure 2.1: Schematic diagram indicating the presence of coalescence and fragmentation processes, for a population of $N = 15$ objects dynamically partitioned into clusters. The size of cluster i is $s_i = 2$, while the size of cluster j is $s_j = 6$ etc. The fragmentation process exhibits the richest range of possibilities, given the combinatorial number of ways in which a cluster can in principle be divided. There are many possible realizations of the objects themselves, e.g. humans, animals, macromolecules, though for simplicity we show them as humans.

macromolecules or other indivisible entities. Earlier studies tended to focus on situations in which the interactions between clusters might be expected to decay with physical separation – as in a simple solution of molecules interacting through Van der Waals interactions for example. However in modern-day social applications, where long-distance communication is as commonplace as communication with neighbors, it makes more sense to have interactions over all lengthscales, with the interaction probability effectively independent of physical separation. These are the type of interactions that we discuss here.

Of the two processes in Fig. 2.1, i.e. coalescence and fragmentation, the coalescence process is likely to be the simpler and more generic. Suppose we have a particular partition of a population of N objects into clusters as in Fig. 2.1, and that a cluster i of size $s_i = 2$ is to coalesce. It is unlikely to undergo three-body collisions and/or interactions, and hence its most likely coalescence event is to join with a single other cluster j. Given that the size of a cluster measures the number of objects in it, it is therefore reasonable to imagine that the coalescence probability should increase as the size of the clusters themselves increase. In a more human setting, the more objects that a cluster contains, the more likely it is that something will happen to one of its members in order to induce such an event, and hence the probability will increase with the size of the cluster. We therefore adopt size-dependent coalescence probabilities in this work. We note that although we are using the term "cluster" throughout this chapter for convenience, it can also be taken to mean a "community" in the language of network science [106] since it denotes a subset of the population who have very strong links between them, while the links between clusters are negligibly weak. We note also that the term "cluster" need not necessarily mean physical connection – instead it could represent a group of objects whose actions happen to be coordinated in some way. Hence the coalescing of two clusters, however distant in real space, can mean an instantaneous alignment of their coordinated activities, as one might expect in a financial market [55], organized crime or insurgent warfare [56, 59]. In such a situation, a common fragmentation event would then likely be a sudden disruption of this coordination – hence it is this type of fragmentation rule that forms the focus of our work. Although we do not explore the details of real-world applications such as financial markets or insurgent warfare here, it is useful to keep them in mind when we discuss the consequences of the different fragmentation rules later in the chapter.

As mentioned earlier, the distinct feature of many real-world systems is the existence of scale-free behavior in the time-averaged cluster size distribution [28, 55, 56, 59, 96, 98, 99], such that in the first instance these systems can be characterized by the exponent of their power law and by the range of its scale-free behavior. One may therefore ask: Which ingredients of the coalescence-fragmentation models, or combinations of ingredients, turn out to control the various observable aspects? It is this general question that motivates the present work.

2.1.2 General Formulation

Once the probabilities specifying the coalescence and fragmentation are given, the cluster size distribution may be computed either by a direct simulation of the model or in a mean-field theory approximation by solving an appropriate set of rate equations, often numerically. The rate equations are typically non-linear. The non-trivial question of existence and uniqueness of the time-independent solution therefore arises, and is addressed in seminal works such as Refs. [94, 102, 103]. For the social/economical models of current interest, the uniqueness and existence can be shown at the level of the rate equations, and verified by direct simulations. We consider mostly “steady-state” models, in which there is some form of robust long-time behavior.

The number of clusters of size s at time t is $n_s(t)$, and N is the total number of members (i.e. the population size). We will drop the explicit time-dependence for simplicity, since it will be clear from the context whether we are discussing $n_s(t)$ or its steady-state time-averaged value. In order to characterize a general system, we need to prescribe the following two functions, each of dimension $[\text{time}]^{-1}$:

- The *coalescence function* $\mathcal{C}(s, s')$ which is the rate describing the process by which two clusters of sizes s and s' merge. We only consider coalescence which depends on the details of a pair of clusters, and hence exclude the possibility that 3 (or more) clusters are involved in the merging process.

- The *fragmentation function* $\mathcal{FR}(s; m_1, m_2, \ldots, m_{s-1})$ which is the rate describing the process by which a cluster of size s fragments into a configuration which contains m_1 clusters of size 1, m_2 clusters of size 2, etc.

The functional form of the above two functions is taken to be time-independent. If we consider general fragmentation processes, we see that a large number of parameters are necessary to characterize the fragmentation. However in order to write down the rate equations and hence calculate the cluster size distribution, we do not need complete knowledge of the fragmentation function (i.e. we do not need knowledge about all possible partitions). It is sufficient to know the *reduced fragmentation function* $\mathcal{F}(s, s', m)$, defined as the rate at which a cluster of size s fragments into a configuration which contains m clusters of size s' plus any other clusters with sizes different to s'. In addition to $\mathcal{F}(s, s', m)$ we need to know the rate that the fragmentation of any given cluster of size s occurs, which we denote as $f(s)$. In principle we can calculate it by summing the complete fragmentation function over all partitions of the fragmentation products. By prescribing the deduced fragmentation function $\mathcal{F}(s, s', m)$ we do not characterize uniquely the fragmentation of the system and in general we may not be able to calculate $f(s)$ – yet it is possible in specific cases to do so once the assumption regarding the fragmentation products has been stated. Looking at the average number of clusters of size s that in unit time undergo the various processes (see Fig. 2.2), we may introduce the following notation:

- $L_F(s)$: *loss due to fragmentation*, the number of clusters of size s that fragment
- $L_C(s)$: *loss due to coalescence*, the number of clusters of size s that join with other clusters
- $G_C(s)$: *gain from coalescence*, the number of clusters of size s created from the merging of clusters of size smaller than s
- $G_F(s)$: *gain from fragmentation*, the number of clusters of size s created from fragmenting clusters of size larger than s

Symbolically the rate equations for any s are written as

$$\frac{\partial n_s}{\partial t} = -L_F(s) - L_C(s) + G_C(s) + G_F(s) \tag{2.1}$$

which explicitly reads as

$$\frac{\partial n_s}{\partial t} = -f(s)\, n_s - n_s \sum_{s'=1}^{N} n_{s'} \mathcal{C}(s, s') + \frac{1}{2} \sum_{s'=1}^{s-1} n_{s'} n_{s-s'} \mathcal{C}(s', s-s') + \sum_{s+1}^{N} n_{s'} \sum_{m=1}^{[N/s']} m\, \mathcal{F}(s', s, m)\,. \tag{2.2}$$

The last term represents the gain in the number of clusters of size s coming from fragmentation of other clusters of size $s' > s$, in such a way that among the fragmentation products we have m clusters of size s. We are summing over all possible values of m

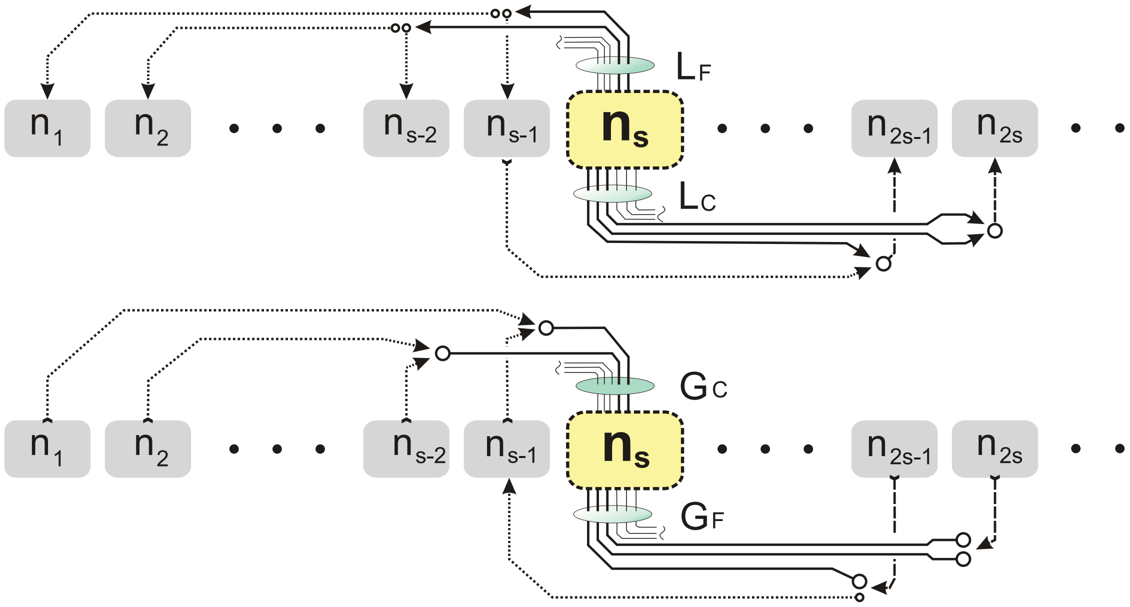


Figure 2.2: The various processes of cluster coalescence and fragmentation which give rise to L_F, L_C, G_F, G_C for any particular value of s. The bottom figure represents the appearance of new clusters of size s, the top one represents their loss. In the interests of simplicity, the fragmentation into two clusters has been depicted and only a few processes are shown.

and s'. Note that we sum over s' which is here the first (not the second) argument of $\mathcal{F}$. An explicit form for $f(s)$ is discussed above. It is convenient to formally define

$$\widetilde{\mathcal{F}}(s, s') = \sum_{m=1}^{[N/s']} m\, \mathcal{F}(s, s', m)\,. \tag{2.3}$$

We write therefore the last term of Eq. 2.2 as

$$\sum_{s'=s+1}^{N} n_{s'} \widetilde{\mathcal{F}}(s', s)\,.$$

2.2 Role of the fragmentation function

A logical first step in the quest to understand classes of models which differ in their cluster fragmentation process, is to look at extreme cases. One such case is the Eguíluz-Zimmermann (E-Z) model [28]. In the E-Z model, fragmentation of a cluster of size s always produces s clusters of size 1, i.e. the cluster breaks up into individual objects. At the other extreme, is the famous Gueron and Levin (G-L) model [54] in which fragmentation of a cluster yields two smaller pieces, i.e. the original cluster splits into two clusters. The original G-L model is formulated in terms of continuous distributions – however, since our aim is to analyze the effects of these rules on the same footing, we will focus on the discrete version of the G-L model, returning to the continuous formulation later on. The G-L model is in fact identical to Smoluchowski's coagulation-fragmentation model with binary fragmentation.

The common feature of the models that we discuss, is the presence of a separable coalescence function:

$$\mathcal{C}(s,s') = \alpha\, a(s)\, a(s') \ . \tag{2.4}$$

In principle, the multiplicative constant may be absorbed into $a(s)$, however we prefer to keep it explicitly and adopt a dimensionless $a(s)$. This class of model is further specified by introducing a coalescence mechanism on the microscopic scale, namely that two clusters merge when any member from one cluster connects to any member from the other cluster. In a macroscopic description, this is equivalent to assuming that $a(s) = s$. We note that Gueron and Levin [54], having the solution of the rate equations for $a(s) = s$, considered explicitly the other cases $a(s) = 1$ and $a(s) = 1/s$ by means of the substitution $n_s \to a(s)\, n_s$ – however, this substitution affects the form of the fragmentation function $\mathcal{F}(s, s', m)$.

2.2.1 Fragmentation function

Assuming that the cluster may only split into two pieces still does not uniquely specify the fragmentation, since we still need information about the probability distribution for the sizes of the fragments. In the G-L model, it is stated that the conditional distribution for fragments is uniform [54], i.e. the fragmentation of a cluster occurs with a probability which is independent of the way in which the cluster breaks. The reduced fragmentation function for $s > 1$ is therefore

$$\mathcal{F}_{GL}(s, s', m) = \beta\, b(s)\, [2\, \delta_{m,1}(1 - \delta_{2s',s}) + \delta_{m,2}\delta_{2s',s}] \tag{2.5}$$

where we have accounted for the fact that if $2s' = s$, the cluster breaks into two fragments of equal size. Using Eq. 2.3 one obtains immediately

$$\widetilde{\mathcal{F}}_{GL}(s, s') = 2\beta\, b(s) \quad . \tag{2.6}$$

The fragmentation probability is calculated as follows:

$$f_{GL}(s) = \frac{1}{2} \sum_{s'=1}^{s-1} \mathcal{F}_{GL}(s, s', m = 1) + \sum_{s'=1}^{s-1} \mathcal{F}_{GL}(s, s', m = 2) = \beta\, (s - 1)\, b(s)\,, \tag{2.7}$$

where the factor $1/2$ in the first term appears in order to avoid double-counting, and the second term represents splitting into two equal parts. In the E-Z fragmentation scheme, the cluster of size s can only break up into individual objects and there is only one mode of fragmentation, hence

$$\mathcal{F}_{EZ}(s, s', m) = \beta\, b(s)(1 - \delta_{s1})\delta_{s',1}\delta_{m,s}\,. \tag{2.8}$$

Using Eq. 2.3 we have

$$\widetilde{\mathcal{F}}_{EZ}(s, s') = \beta s\, b(s)\, (1 - \delta_{s1})\delta_{s',1}\,. \tag{2.9}$$

The fragmentation probability is

$$f_{EZ}(s) = \sum_{s'=1}^{s-1} \mathcal{F}_{EZ}(s, s', m = s) = \beta\, (1 - \delta_{s1}) b(s). \tag{2.10}$$

There is no double-counting problem here. A peculiar feature of the E-Z model is that the corresponding set of rate equations is semi-recursive, i.e. any k-th equation depends only on values of $n_{s'}$ for $s' \leq k$ and on a global constant depending on all

n_s. This is a feature by which it is easy to show the existence and uniqueness of the solution and also to solve the system numerically.

It is the common feature of both G-L and E-Z type models to assume that $a(s) = b(s)$. Mathematically, this acts to restrict the space of all possible solutions, otherwise the diversity of general solutions would be overwhelming. In physical terms, the justification for this assumption is that there is interest in the specific case $a(s) = b(s) = s$, since this describes the case of fragmentation of the cluster being triggered by a single member – hence the proportionality to s. Similarly the likelihood that two groups would become coordinated and hence act as a single unit (i.e. they coalesce) would be proportional to the size of each of the groups, if the underlying mechanism involved one member from each initiating the process by forming a link followed by all the other members.

With the assumptions made so far, it turns out that each system is described by three constants: α, β and the total population size N. For the time-independent system we need just two constants, and since α and β are of dimension $[\text{time}]^{-1}$ then only their ratio α/β should appear. The steady-state rate equations are as follows.

G-L system:

$$-\beta(s^2 - s)\, n_s - \alpha\, s\, n_s \sum_{s'=1}^{N} s' n_{s'} + \frac{\alpha}{2} \sum_{s'=1}^{s-1} s'\, n_{s'}\, (s - s') n_{s-s'} +\ 2\,\beta \sum_{s'=s+1}^{N} s'\, n_{s'} = 0\,. \tag{2.11}$$

E-Z system:

$$-\beta s\, (1 - \delta_{s1}) n_s - \alpha\, s\, n_s \sum_{s'=1}^{N} s' n_{s'} + \frac{\alpha}{2} \sum_{s'=1}^{s-1} s'\, n_{s'}\, (s - s') n_{s-s'} + \beta \delta_{s,1} \sum_{s'=s+1}^{N} s'^2\, n_{s'} = 0\,. \tag{2.12}$$

Eguíluz and Zimmermann [28] explicitly used the following constants:

$$\alpha = \frac{2(1-\nu)}{N^2}\,, \qquad \beta = \frac{\nu}{N}\,. \tag{2.13}$$

We see that both sets of equations 2.11 and 2.12 simplify if we express them in terms of $k_s = s\,n_s$, i.e. the number of agents contained in clusters of size s. Note that for general $a(s)$, we need to substitute $k_s = a(s)\,n_s$.

2.2.2 Equilibrium in Gueron-Levin model: Continuous formulation

Gueron and Levin's solution [54] to the G-L model, was obtained for the system with continuous cluster density which we denote as $n(s)$. In terms of $k(s) = s\,n(s)$, the integral rate equation corresponding to Eq. 2.11 with no limit on the maximum size of a cluster, is given by:

$$0 = -\beta\, s\, k(s) - \alpha\, k(s) \int_0^\infty ds'\, k(s') + \alpha \frac{1}{2} \int_0^s ds' k(s')\, k(s-s') + 2\beta \int_s^\infty ds'\, k(s')\,. \tag{2.14}$$

Looking at this equation, we guess that the solution is obtained by substituting an ansatz which satisfies $k(s+s') \propto k(s)k(s')$. The first form to try is $k(s) = A\, e^{-\mu\, s}$. With this ansatz we obtain

$$0 = -A\,\beta\, s\, e^{-\mu s} - A^2\alpha/\mu e^{-\mu s} + A^2\alpha/2\, s e^{-\mu s} + 2A\beta/\mu\, e^{-\mu s}\,. \tag{2.15}$$

There are two types of terms, of the form $\sim e^{-\mu s}$ or $\sim s\, e^{-\mu s}$. Eliminating the overall exponential factor we have

$$0 = s\left(-A\beta + A^2\frac{\alpha}{2}\right) + \frac{2}{\mu}\left(A\beta - A^2\frac{\alpha}{2}\right)\,. \tag{2.16}$$

Both terms in parentheses vanish if we choose

$$A = 2\frac{\beta}{\alpha}\,. \tag{2.17}$$

The scale factor μ in the exponent is determined to be $\mu = {}^{2\beta}/_{N\alpha}$ by normalization. The solution to Eq. 2.14 is just an exponential function which was obtained by Gueron and Levin by means of a Laplace transform.

We notice here a remarkable curiosity: If we take the actual solution of Eq. 2.14, then for any s the following equalities hold exactly:

$$L_F(s) = G_C(s), \qquad L_C(s) = G_F(s)\,. \tag{2.18}$$

This is in effect the detailed balancing. In other words, the following holds for the G-L model: *The average loss of clusters of size s due to the cluster fragmentation, is equal to the average gain obtained from the coalescence of clusters of sizes smaller than s. Also the average loss of clusters of size s due the coalescence with other clusters is equal to the average gain obtained from the fragmentation of clusters of sizes larger than s.*

In addition to its mathematical interest, this identity (which is not satisfied for the E-Z model as discussed below) shows up a fundamental feature of the G-L model, which arises in turn from the microscopic rules which characterize it. This symmetry is also revealed if we look at the behavior of the system with time flowing backwards.(In general, one does not obtain a stochastic system by time-reversing the recorded history of a second non-equilibrium stochastic system. Although this becomes an issue for discrete systems due to the presence of fluctuations, we may still discuss it from the perspective of the average quantities describing the equilibrium state). With the reversed time perspective, the coalescence of clusters is observed as fragmentation and vice-versa, but the average cluster size distribution remains unaltered in the equilibrium state. As far as this average quantity is concerned, the system is therefore invariant under an interchange of coalescence/fragmentation processes – and in the specific case of G-L model, the time-reversed processes are exactly the same as the original ones.

2.2.3 Cluster size distribution: The exponential cutoff

We now return to the discrete formulation. For the discrete version of the G-L system, it may be verified by direct computation that the steady-state value

$$n_s = 2\frac{\beta}{\alpha}\, s^{-1} \exp\left(-\mu\, s\right) \tag{2.19}$$

is also a solution of Eq. 2.11, once we make an approximation of extending the summation limits to infinity. Here the normalization condition is $N = \sum_{s'=1}^{\infty} s'\, n(s')$, from which we calculate

$$\mu = \ln\left(\frac{2\beta}{\alpha N} + 1\right) . \tag{2.20}$$

Thus we have

$$n_s = 2\frac{\beta}{\alpha}\, s^{-1} \left(\frac{2\beta}{\alpha N} + 1\right)^{-s} . \tag{2.21}$$

It is advantageous to consider $\beta/\alpha \propto N$, thus the exponent is independent of N and n_s is just proportional to N. If we use here the same constants (Eq. 2.13) as the original E-Z model, the solution is

$$\text{G-L}: \; n_s = N\frac{\nu}{1-\nu}\, s^{-1}\, (1-\nu)^s . \tag{2.22}$$

The solution to the E-Z model rate equations may be approximated as [61]

$$\text{E-Z}: \; n_s \sim N s^{-2.5} \left(\frac{4(1-\nu)}{(2-\nu)^2}\right)^s . \tag{2.23}$$

In order to compare the cluster size distribution for both models, we will for convenience characterize both using the same parameters N and ν. This means that they will have the same coalescence function, and their fragmentation functions will agree for the splitting of clusters of size $s = 2$. The difference between the two models then lies in the fragmentation of larger clusters. This allows them to be compared on a similar footing, focusing just on the effect of their respective fragmentation functions.

The cluster size distribution for both models is of the form $n_s \propto s^{-\kappa} e^{-\mu s}$. The scale of s at which the exponential cut-off becomes relevant, can be identified by looking at the ratio

$$\frac{n_{s+1}}{n_s} = e^{-\mu}\,\frac{(s+1)^{-\kappa}}{s^{\kappa}} = e^{-\mu}\left(1 - \frac{\kappa}{s} + O\left(\frac{1}{s^2}\right)\right) \,. \tag{2.24}$$

We assumed that $\mu \ll 1$ which is the regime in which such models exhibit power-law behavior. The exponential cutoff becomes dominant at the scale where $a \approx (1 - \frac{\kappa}{s})$, hence we may define

$$s_{\text{cutoff}} \equiv \frac{\kappa}{1 - e^{-\mu}} \,. \tag{2.25}$$

For the models of interest in this chapter with $\mu \ll 1$, and therefore $\nu \ll 1$, we have

$$\text{G-L}: \; s_{\text{cutoff}} = \nu^{-1} \,, \quad \text{E-Z}: \; s_{\text{cutoff}} = \frac{5}{2}\left(\frac{2-\nu}{\nu}\right)^2 \approx 10\nu^{-2} \,. \tag{2.26}$$

It is clear (see Fig. 2.3) that the range of cluster sizes for which one observes the power-law, is several orders of magnitude larger for the E-Z model than for the G-L model. We may also verify that the special equilibrium result mentioned earlier for the continuous G-L model (see statement in italics) is also a property of the corresponding discrete model, once the upper limits in the sums are extended to infinity. It also holds that

$$\text{G-L model}: \; L_F(s) \cong s\nu L_C(s) \,, \quad \text{E-Z model}: \; L_F(s) \cong \frac{\nu}{2} L_C(s) \,. \tag{2.27}$$

We see therefore that for the G-L model we can always find a value of s for which $L_F(s) \approx L_C(s)$ – in particular, it is the scale of the cluster size over which the exponential cutoff becomes apparent. By contrast, in the E-Z model for $\nu^{-1} \gg 1$ (i.e. for the wide range over which there is power-law behavior) we have $L_F(s) \ll$

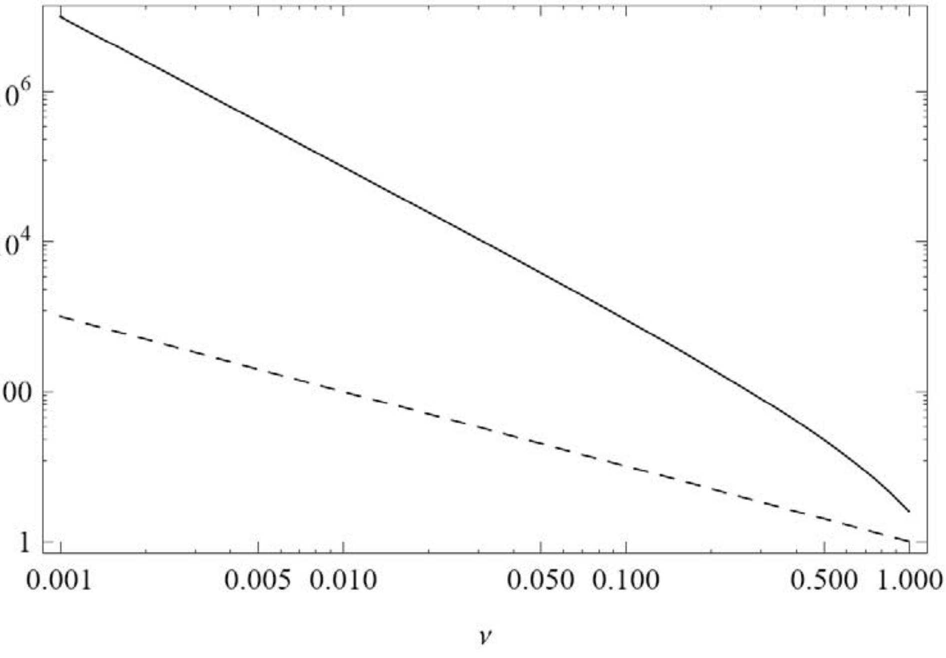


Figure 2.3: Scale of exponential cutoff for the E-Z model (solid curve) and for the G-L model (dashed curve) described by the same parameter ν. The range of cluster sizes for which one observes the power-law, is several orders of magnitude larger for the E-Z model than for the G-L model.

$L_C(s) \approx G_C(s)$. If we again compare both models, we find that the $L_F(s)$ function is usually much larger for the G-L model than for the E-Z model. Figure 2.4 illustrates this finding for a particular set of parameters.

2.2.4 Reservoir model: Fragmentation into fixed size clusters

So far we have looked at the cases in which the total population size N is treated as one of the parameters defining the model. Specifically, we considered a constant population such that the constraint $N = \sum_{s=1}^{N} sn_s(t)$ strictly holds at every instant in time. This highly idealized situation might not be realized in a particular real-world problem – however we can formulate a "reservoir model" version in which the total

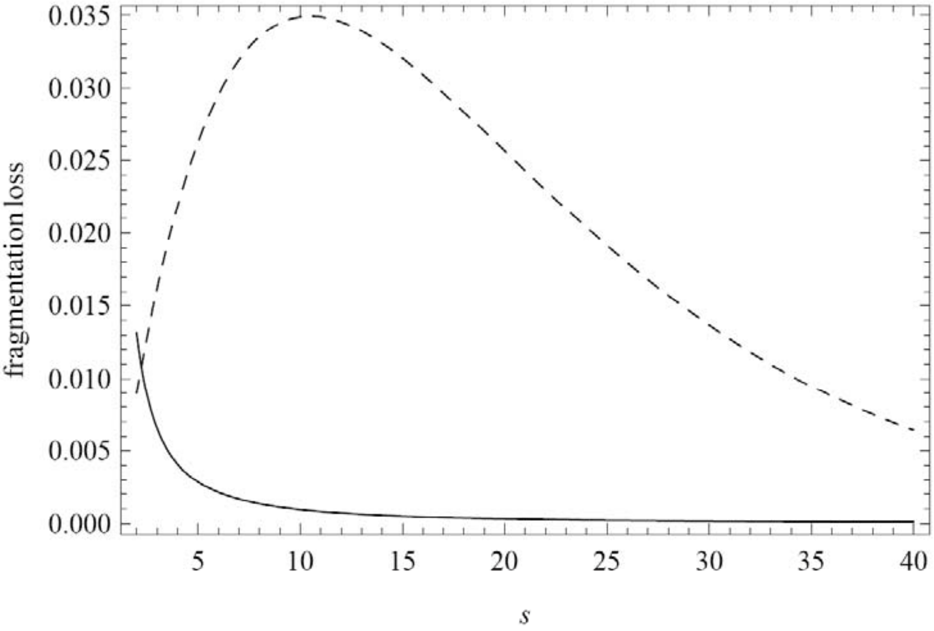


Figure 2.4: $L_F(s)$, *loss due to fragmentation* for the E-Z model (solid curve) and for the G-L model (dashed curve) with parameters $\nu = 0.1$ and $N = 1000$. The overall scale is determined up to a multiplicative constant (i.e. the scale of time). The graphs show that $L_F(s)$ for the G-L model is usually much larger than $L_F(s)$ for the E-Z model.

population size is no longer a parameter defining the model, but instead becomes a dynamical variable whose averaged equilibrium value is determined by the model itself: $N \equiv \sum_{s=1}^{\infty} s\langle n_s(t)\rangle$. We introduce a constant supply of individuals from a system reservoir, with γ denoting the rate at which single individuals are added. The products of the fragmenting cluster are then moved back to the reservoir. An equivalent interpretation is that a cluster stays in the system but ceases to interact (i.e. it does not merge with other clusters).

Here we discuss the case which in the remainder of the dynamics resembles the terms in the E-Z model, with βs being the rate of removing a cluster of size s and

$\alpha ss'$ being the coalescence rate. This particular reservoir model is therefore described by three parameters α, β, γ, with only two parameters required for the steady-state cluster size distribution. The master equations are

$$-\beta s\, n_s - \alpha\, s\, n_s \sum_{s'=1}^{\infty} s' n_{s'} + \frac{1}{2}\,\alpha \sum_{s'=1}^{s-1} s'\, n_{s'}\, (s-s')n_{s-s'} + \gamma \delta_{s1} = 0\,. \tag{2.28}$$

By summation of Eq. 2.28, the average number of participants is obtained as

$$\langle N \rangle = \frac{\sqrt{\beta^2 + 2\alpha\gamma} - \beta}{\alpha}\ . \tag{2.29}$$

The cluster size distribution has the same form as for the E-Z model Eq. 2.23, if expressed in terms of $\langle N \rangle$ and α/β. In this case there is no approximation made in extending the summation limit to infinity, and the solution in Eq. 2.23 is exact from the mean-field theory point of view. There is no limit on the maximum size of a cluster, which in principle may exceed $\langle N \rangle$ when the effect of fluctuations is non-negligible.

We note that the E-Z model changes very little if we consider the case where a cluster fragments into a set of smaller clusters, each of fixed size s_0. For the discrete system, there is naturally a divisibility problem regarding fragmentation of clusters of sizes which are not a multiple of s_0. Since we are interested in steady-state behavior, we may assume that such clusters do not fragment. Whatever the initial configuration is after a sufficiently long time, the system in equilibrium will consist almost entirely of clusters that are a multiple of s_0 in size. It turns out that the cluster size distribution has the same form as the E-Z model in Eq. 2.23, if we re-express it in terms of s_0 as the basic unit, i.e. if we substitute $s \rightarrow s/s_0$.

2.3 Generalization of the E-Z model

We now open up the above discussion to a broader class of coalescence-fragmentation models. The variety of coalescence-fragmentation-type processes which have been employed to describe physical, biological and social systems in the literature is enormous [27,28,54,60–63,65–72]. Here we focus on the E-Z model [28] given its potential relevance to understanding the empirical distributions observed in financial markets and insurgent behavior [56,57,59,96,97]. In particular, we will investigate the effect of variations in the rules, and perturbations, on the cluster size distribution.

2.3.1 Spontaneous cluster formation

Our first generalization mimics the situation in which a small number of clusters are allowed to spontaneously form from the population, as opposed to arising from the merger of two smaller clusters. In practice this is most simply viewed as the spontaneous formation of clusters from previously single agents/clusters of unit size. (The exact mechanism is unimportant). Let γ_s represent the rate of formation of clusters of size s by the non-hierarchical method. The value of γ_1 is implicitly defined by the requirement that the size N of the population remains constant, i.e., $\sum_{s=1}^{\infty} s\gamma_s = 0$, therefore $\gamma_1 < 0$. The rate equation is given by

$$\begin{aligned}
\frac{\partial n_s}{\partial t} &= -\beta s n_s \\
&\quad +\frac{1}{2}\,\alpha \sum_{r=1}^{s-1} r n_r \,(s-r)\, n_{s-r} \\
&\quad -\alpha s n_s \sum_{r=1}^{\infty} r n_r + \gamma_s
\end{aligned}$$

for $s \geqslant 2$, and

$$\frac{\partial n_1}{\partial t} = -\alpha n_1 \sum_{r=1}^{\infty} r n_r + \beta \sum_{r=1}^{\infty} r^2\, n_r + \gamma_1$$

for $s = 1$. In the steady state this may be written as

$$sn_s = A\left(\frac{1}{2}\,\alpha\sum_{r=1}^{s-1} rn_r\,(s-r)\,n_{s-r} + \gamma_s\right),$$

where A is defined by

$$A = \frac{1}{\beta + \alpha\sum_{r=1}^{\infty} rn_r}.$$

In deriving these results, we have extended the summation of appropriate low-order terms to infinity by introducing the approximation $\sum_{r=1}^{\infty} rn_r \approx N$. The generating function $g[y]$ is now introduced:

$$g[y] \equiv \sum_{r=2}^{\infty} rn_r y^r. \tag{2.30}$$

Taking the square of this function and using previous equation yields

$$\begin{aligned} 0 &= (g[y])^2 - 2\left(\frac{1}{A\alpha} - n_1 y\right) g[y] \\ &\quad + n_1^2 y^2 + \frac{2}{\alpha}\chi[y], \end{aligned} \tag{2.31}$$

where $\chi[y] \equiv \sum_{r=2}^{\infty} \gamma_r y^r$. Using the fact that $g[1] = \sum_{r=1}^{\infty} rn_r - n_1$, gives

$$n_1 = \frac{1 - A^2\left(\beta^2 + 2\alpha\chi[1]\right)}{2A\alpha}. \tag{2.32}$$

Solving Eq. (2.31) for general y and expanding the resulting radical using Taylor's theorem yields

$$\begin{aligned} g[y] &= A\chi[y] + \frac{1}{A\alpha}\sum_{k=2}^{\infty}\left(\frac{(2k-3)!!}{(2k)!!}\right. \\ &\quad \left.\times\left[2A\alpha\left(n_1 y + A\chi[y]\right)\right]^k\right). \end{aligned} \tag{2.33}$$

We will assume that the gamma term is small enough to be treated as a perturbation, i.e. $\frac{A\chi[y]}{n_1 y} \ll 1$ and hence a first-order binomial expansion of the exponential term in

Eq. (2.33) may be performed. In this case

$$\begin{aligned} g[y] \approx\ & A\chi[y] \\ & + \frac{1}{A\alpha}\sum_{k=2}^{\infty}\left(\frac{(2k-3)!!}{(2k)!!}\right. \\ & \times\left\{ [2A\alpha n_1 y]^k \right. \\ & \left.\left. + k\frac{A}{n_1}[2A\alpha n_1]^k \sum_{r=2}^{\infty}\gamma_r y^{r+k-1}\right\}\right). \end{aligned}$$

Comparing terms with Eq. (2.30) yields

$$n_2 = \frac{1}{2}A\gamma_2 + \frac{1}{4}A\alpha(n_1)^2$$

for $s = 2$. For large s, Stirling's approximation yields

$$\begin{aligned} n_s \approx\ & \left(\frac{e^2}{2\sqrt{\pi}A\alpha}\right)\left\{1 - A^2\left[\beta^2 + 2\alpha\chi\right]\right\}^s s^{-5/2} \\ & + A\left[\gamma_s + \frac{e^2}{2\sqrt{\pi}n_1}\sum_{r=2}^{s-1}\left(\left\{1 - A^2\left[\beta^2 + 2\alpha\chi\right]\right\}^r r^{-1/2}\gamma_{s-r+1}\right)\right]s^{-1}, \quad (2.34) \end{aligned}$$

where $\mathrm{X} \equiv \chi[1]$. Since A is constant for a given population, the general form of the above equation is

$$n_s \propto \kappa^s s^{-5/2} + Z[s]s^{-1}, \tag{2.35}$$

where $\kappa \equiv 1 - A^2(\beta^2 + 2\alpha\mathrm{X})$ and $Z[s]$ is a function whose form depends on the details of the perturbation induced by the γ_s terms.

2.3.2 Step perturbation

We now analyze a highly simplified example from the class of perturbations which die off as s increases. In particular, we consider a step function perturbation:

$$\gamma_s = \begin{cases} \frac{\Phi}{q-1}, & \text{for } 2 \leqslant s \leqslant q; \\ 0, & \text{for } s > q; \end{cases}$$

where q is an arbitrarily chosen cluster size and $\Phi > 0$. Using the original E-Z parametrization of Eq. 2.13 and Eq. 2.34, we obtain the cluster size distribution as

$$n_1 \approx N\frac{1-\Phi}{2-\nu},$$

$$n_2 \approx \frac{1}{2(2-\nu)}\left[\frac{\Phi}{q-1}+\frac{1-\nu}{(2-\nu)^2}(1-\Phi)^2\right],$$

$$\begin{aligned} n_s \approx\ & N\frac{(2-\nu)e^2}{4\sqrt{\pi}(1-\nu)}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^s s^{-5/2} \\ & +N\frac{1}{2-\nu}\frac{\Phi}{q-1}s^{-1} \\ & +\frac{e^2}{2\sqrt{\pi}}\frac{\Phi}{(1-\Phi)}\frac{1}{q-1} \\ & \times\left\{\sum_{r=2}^{s-1}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^r r^{-1/2}\right\}s^{-1}, \end{aligned}$$

for $3 \leqslant s \leqslant q$, and

$$\begin{aligned} n_s \approx\ & N\frac{(2-\nu)e^2}{4\sqrt{\pi}(1-\nu)}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^s s^{-5/2} \\ & +\frac{e^2}{2\sqrt{\pi}}\frac{\Phi}{(1-\Phi)}\frac{1}{q-1} \\ & \times\left\{\sum_{r=2}^{q}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^r r^{-1/2}\right\}s^{-1} \end{aligned}$$

for $s \geqslant q+1$. Examples of the resulting n_s distribution are plotted in Fig. 2.5.

Interestingly the greatest effect of the perturbation is found at high s, whereas the perturbation's definition means that it only directly affects the clustering at low s. This is because the perturbation creates small clusters by non-hierarchical means, which then serve as effective nucleation sites for the formation of larger clusters. The perturbation therefore greatly accelerates the formation of large clusters whereas, by contrast, the small clusters fragment sufficiently fast that their presence is hidden on the graph at low s. Figure 2.6 shows the predicted distribution of n_s for different

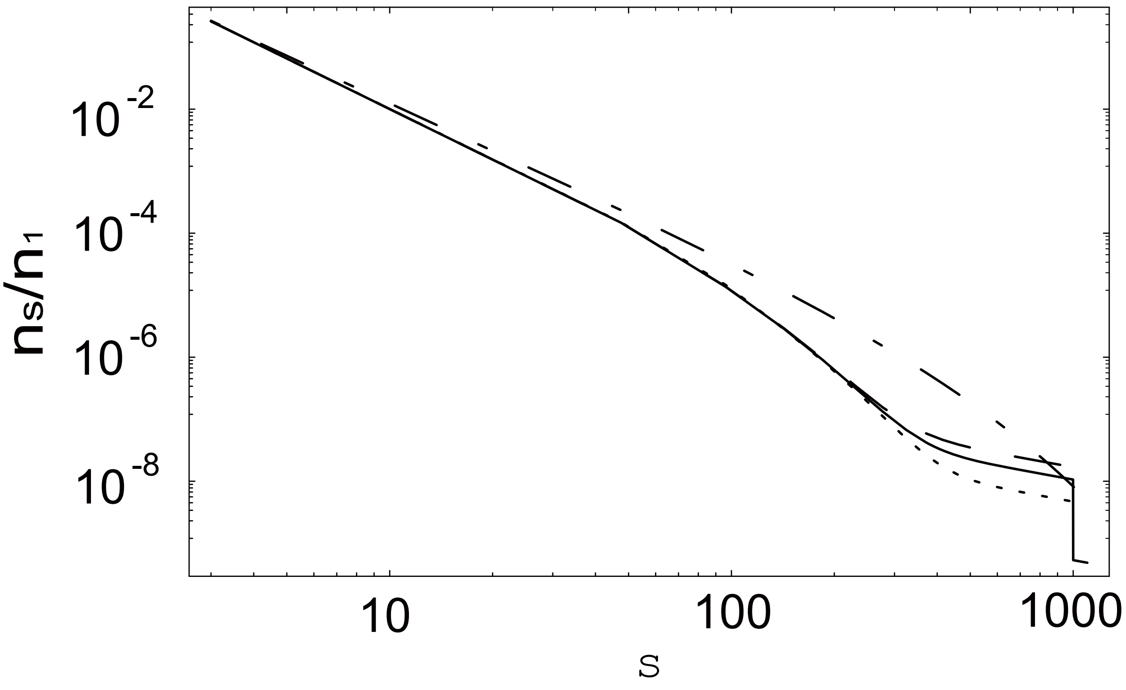


Figure 2.5: Predicted distribution of cluster sizes for the perturbed system described in Section 2.3.2, using $N = 1000$, $\nu = 0.1$ and $\Phi = 0.01$. The dot-dashed line shows the unperturbed population, while the dashed line shows $q = 10$, the dotted line shows $q = 100$, and the solid line shows $q = 1000$.

signs of the perturbation ($\pm\Phi$), together with the unperturbed result. Note that in the case of a negative sign, it is necessary that

$$\Phi < \frac{\nu^2}{4\,(1-\nu)} \tag{2.36}$$

in order that n_s remain finite as $s \to \infty$. The analytic predictions for the perturbed populations are quantitatively reliable for a wide range of s values. With primed quantities referring to the $-\Phi$ case, and using$N = 10000$, $\Phi = 0.001$, $\nu = 0.1$ and $q = 500$, we find that the effect of the perturbation is as follows:

$$\begin{aligned} \frac{n_1}{n_1'} &= 0.998, \\ \frac{n_{500}}{n_{500}'} &= 0.39, \\ \frac{n_{1000}}{n_{1000}'} &= 0.14. \end{aligned}$$

As claimed earlier, this small, low-s perturbation can be seen to have a very significant effect across a wide range of s, in particular at high s. We note that the interpretation of the perturbation is that statistically a cluster of size 500 or less spontaneously forms/fragments for $+/-\Phi$ cases respectively once in every 1000 timesteps, where a single timestep corresponds to any particular fragmentation or coalescence event in the system.

2.3.3 Variable population size

Only a small subset of real-world problems correspond to populations with a fixed size N, or with a fixed time-averaged size N. In this section, we develop an analytic treatment of a model which is analogous to the E-Z model, but which treats the case of a population whose size varies with time according to a simple law. As mentioned earlier, several real-world systems seem to have power-law behavior with

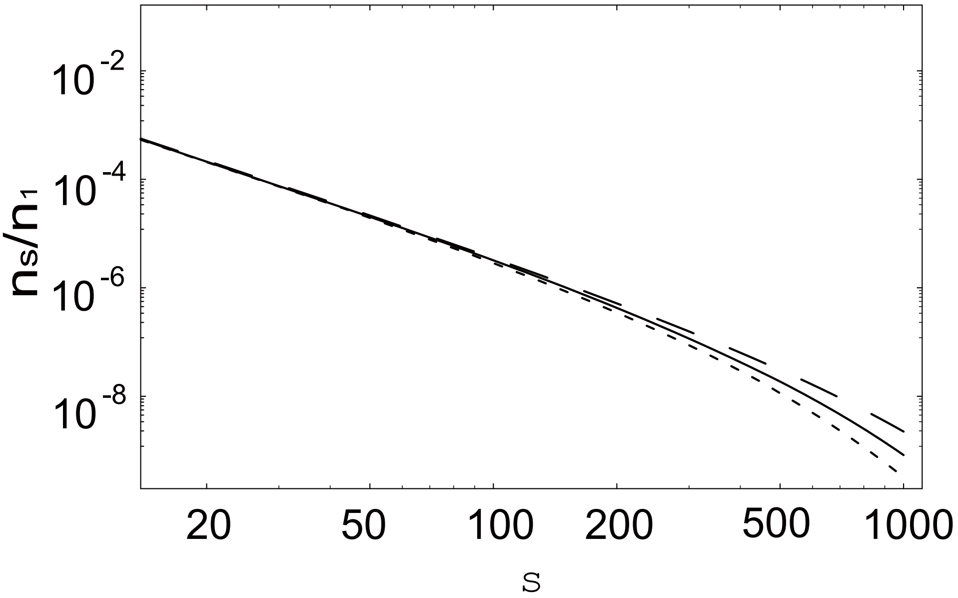


Figure 2.6: Predicted cluster size distribution for the $+\Phi$ case (dotted) and the $-\Phi$ case B (dashed) compared with the unperturbed model (solid). Parameter values: $N = 10000$; $\Phi = 0.001$; $\nu = 0.1$; $q = 500$.

exponent around 2.5. which is the same behavior as the unperturbed E-Z model – for example, the distributions of size of trades in markets, and the size of attacks in conflict and terrorism [55–57, 59, 97, 98]. Such real-world observations could therefore conceivably be attributed to the E-Z model – however this identification would be far more believable if the E-Z model's assumption of constant N did not have to be made. It is known that as the years pass in an active war, an insurgent population will generally increase in size as previously passive people become recruited. Likewise as a market grows, previously inactive individuals tend to join the trading. Hence a model with increasing N (or decreasing N for mature wars or markets that are dying off) is of interest. Real-world examples of declining populations are also known [100], [101].

We now look at a version which can be treated analytically under the assumption that the coalescence processes are negligible. Although this makes it arguably more restricted than our previous versions, the advantage is that the equation retains linear temporal dynamics and admits a novel solution. Including all coalescence terms would make it non-linear, and intractable.

Our model considers a population containing $N[t]$ agents instantaneously divided into $M[t]$ clusters, as in the E-Z model. First we focus on the number of agents increasing in time, and introduce the following E-Z-like rules:

1. In a single timestep, with probability $p[t]$, $L[t]$ new agents are added to a single cluster of size s, the cluster being selected with probability proportional to s.
2. Alternatively, with probability $q[t] = 1 - p[t]$, a randomly selected cluster fragments (selection of this cluster is independent of cluster size).

If the change in the number of agents is negative, then the model runs as follows:

1. In a single timestep, with probability $p[t]$, $L[t]$ agents are removed from a single cluster of size s, the cluster being selected with probability proportional to s. If the selected cluster has $s < |L[t]|$ then nothing occurs.
2. Alternatively, with probability $q[t] \equiv 1 - p[t]$, a randomly selected cluster fragments, with selection of this cluster independent of cluster size.

The rationale for adding or subtracting from a single cluster is that in many situations of interest, only a single cluster will likely be involved in an external event which changes the population size. As with all these generalizations, more realistic rules can of course be explored – but one runs the risk of obtaining increasingly complicated results.

Increasing population size: $L[t] > 0$

The model proposed above leads to the rate equations

$$\begin{aligned}\frac{\partial n_s}{\partial t} &= \frac{p[t]}{N[t]}\big((s-L[t])n_{s-L[t]} - sn_s\big) - \frac{q[t]}{M[t]}n_s \\ & \qquad\qquad\qquad\qquad\qquad\qquad \text{for } s > L[t], \\ \frac{\partial n_s}{\partial t} &= \qquad -\frac{p[t]}{N[t]}sn_s - \frac{q[t]}{M[t]}n_s \qquad \text{for } 2 \leqslant s \leqslant L[t], \\ \frac{\partial n_1}{\partial t} &= \quad -\frac{p[t]}{N[t]}n_1 + \frac{q[t]}{M[t]}\textstyle\sum_{r=2}^{\infty} rn_r \qquad \text{for } s = 1,\end{aligned}$$

with resulting totals

$$\frac{\mathrm{d}N}{\mathrm{d}t} = p[t]L[t], \tag{2.37}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = q[t]\left(\frac{N[t]}{M[t]} - 1\right). \tag{2.38}$$

The solution of the above equations clearly depends on the forms of $L[t]$ and $p[t]$. As a simple example, we take both to be constant: $L[t] \equiv L$ and $p[t] \equiv p$ for all t. In this case, it can be seen that for times $t \gg \frac{N[t=0]}{pL}$, Eq. (2.37) yields the linear solution

$$N[t] = pLt.$$

If we assume a similar asymptotically linear form for $M[t]$ at large t, $M[t] = \sigma t$, we can go on to deduce from Eq. (2.38) that

$$\sigma = \frac{q}{2}\left(\sqrt{4\frac{p}{q}L + 1} - 1\right).$$

We now assume a linear form for all n_s: $n_s[t] = c_s t$. In this case, one obtains the solution

$$c_1 = \frac{q}{\sigma}\frac{L}{L+1}\sum_{k=1}^{\infty}(1+kL)c_{1+kL} = \frac{q}{\sigma}\frac{L}{L+1}\left(pL - c_1\right).$$

Therefore

$$\begin{aligned} c_1 &= \frac{pq}{\sigma}\frac{L^2}{L(1+q/\sigma)+1}, \\ c_{1+kL} &= \frac{pq}{\sigma}\frac{L^2}{L(1+q/\sigma)+1}\frac{\rho!^{(L)}(1+(k-1)L)!^{(L)}}{(\rho+kL)!^{(L)}}, \end{aligned} \tag{2.39}$$

for $k = 1, 2, 3, \ldots$, where

$$\rho \equiv \left(\frac{q}{\sigma} + 1\right) L + 1,$$

and we have used the multifactorial function, defined recursively by

$$m!^{(n)} = \begin{cases} 1, & \text{if } 0 \leqslant m < n; \\ m(m-n)!^{(n)}, & \text{if } m \geqslant n. \end{cases}$$

Clearly $c_s = 0$ for $s \neq 1 + kL$. Via a generalization of Stirling's approximation,

$$\ln(n!^{(b)}) \sim \frac{1}{b}(n \ln n - n).$$

Applying this to Eq. (2.39), we obtain our solution:

$$c_{1+kL} \approx \frac{pq}{\sigma} \frac{L^2}{L(1+q/\sigma)+1} \frac{e^{(L-1)/L} \rho^{\rho/L} (1+(k-1)L)^{k-1+1/L}}{(\rho+kL)^{k+\rho/L}} \tag{2.40}$$

for integer $k \geqslant 1$. If we take a snapshot of this system at any given time, the observed cluster size distribution will be given by Eq. 2.40, modulo a multiplicative constant which grows linearly with time. The leading k-dependent behavior of Eq. (2.39) is

$$\frac{(kL + (1-L))^{1-L}}{(kL+\rho)^{\rho}} (kL)^{-(\rho+L-1)} \ . \tag{2.41}$$

Decreasing population size: $L[t] < 0$

For simplicity in the following analysis, we do not allow complete annihilation of clusters (i.e. we do not allow the removal of all of a cluster's members from the population). The rate equations for $L[t] < 0$ are as follows:

$$\begin{aligned} \frac{\partial n_s}{\partial t} &= \frac{p[t]}{N[t]} \big((s + |L[t]|) n_{s+|L[t]|} - s n_s\big) - \frac{q[t]}{M[t]} n_s \\ & \qquad\qquad\qquad\qquad \text{for } s > |L[t]|, \\ \frac{\partial n_s}{\partial t} &= \frac{p[t]}{N[t]} (s + |L[t]|) n_{s+|L[t]|} - \frac{q[t]}{M[t]} n_s \\ & \qquad\qquad\qquad\qquad \text{for } 2 \leqslant s \leqslant |L[t]|, \end{aligned}$$

and

$$\frac{\partial n_1}{\partial t} = \frac{p[t]}{N[t]}(1+|L[t]|)n_{1+|L[t]|} + \frac{q[t]}{M[t]}\sum_{r=2}^{\infty} r n_r$$

for $s = 1$, with resulting totals

$$\begin{aligned} \frac{\mathrm{d}N}{\mathrm{d}t} &= -\frac{p[t]|L[t]|}{N[t]}\sum_{r=1+|L|}^{\infty} r n_r, && (2.42) \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= q[t]\left(\frac{N[t]}{M[t]} - 1\right). && (2.43) \end{aligned}$$

As above, we can obtain a solution by assuming that p and L are both constant, and then introduce a linear trial solution of the form

$$\begin{aligned} N[t] &= N_0 - \gamma t, \\ M[t] &= M_0 + \sigma t, \\ n_s[t] &= C_s - c_s t. \end{aligned}$$

This approximation can only hold as long as the changes in each n_s are small compared to the size of the respective C_s. In this case (i.e., for t not too large) we obtain

$$\sigma \approx q\left(\frac{N_0}{M_0} - 1\right),$$
$$\gamma \approx \frac{p|L|}{N_0}\sum_{r=1+|L|}^{\infty} r C_r,$$

and for the c_s we obtain:

$$\begin{aligned} c_1 &\approx -\frac{p}{N_0}(1+|L|)C_{1+|L|} - \frac{q}{M_0}(N_0 - C_1) \\ & \qquad\qquad\qquad\qquad \text{for } s = 1, \\ c_s &\approx \frac{q}{M_0}C_s - \frac{p}{N_0}(s+|L|)\,C_{s+|L|} \qquad \text{for } 2 \leqslant s \leqslant |L|, \\ c_s &\approx \left(\frac{q}{M_0} + \frac{p}{N_0}s\right)C_s - \frac{p}{N_0}(s+|L|)\,C_{s+|L|} \\ & \qquad\qquad\qquad\qquad \text{for } s > |L|. \end{aligned}$$

With a suitable choice of initial conditions and a large population, one can therefore infer the small-t behavior of the system.

Decreasing population: proof of concept

As a simple example, we take $L < 0$ and a starting population of the form

$$n_s[t = 0] = \begin{cases} C_1 - \phi s, & \text{if } s < \frac{C_1}{\phi}; \\ 0, & \text{if } s \geqslant \frac{C_1}{\phi}. \end{cases}$$

In this case our equations from Section 2.3.3 yield

$$\begin{aligned} N_0 &= \frac{1}{6} C_1 \left[\left(\frac{C_1}{\phi} \right)^2 - 1 \right], \\ M_0 &= \frac{1}{2} C_1 \left(\frac{C_1}{\phi} - 1 \right), \\ \gamma &\approx p\,|L| \left\{ 1 - \frac{|L|\,(1 + |L|)\,(3C_1 - \phi - 2|L|\phi)}{C_1 \left[\left(\frac{C_1}{\phi} \right)^2 - 1 \right]} \right\}, \\ \sigma &\approx \frac{(1-p)}{3} \left(\frac{C_1}{\phi} - 2 \right). \end{aligned}$$

This leads to an expression for n_1 of the form

$$\begin{aligned} n_1[t] \approx\; & C_1 + \left[\frac{p}{N_0} (1 + |L|)\,(C_1 - \phi\,(1 + |L|)) \right. \\ & \left. + \frac{q}{M_0} (N_0 - C_1) \right] t, \end{aligned}$$

with corresponding n_s of the form

$$\begin{aligned} n_s[t] \approx\; & C_1 - \phi s - \left[\left(\frac{q}{M_0} C_1 - \frac{|L|p}{N_0} C_{|L|} \right) \right. \\ & \left. - \left(\frac{p}{N_0} C_{2|L|} + \frac{q}{M_0} \phi \right) s + \frac{p}{N_0} \phi s^2 \right] t \end{aligned}$$

for $2 \leqslant s \leqslant |L|$, and

$$\begin{aligned} n_s[t] \approx\; & C_1 - \phi s - \left[\left(\frac{q}{M_0} C_1 - \frac{|L|p}{N_0} C_{|L|} \right) \right. \\ & \left. + \left(\frac{2\,|L|\,p}{N_0} - \frac{q}{M_0} \right) \phi s \right] t \end{aligned}$$

for $s > |L|$. Figure 2.7 shows a plot of this model using illustrative parameter values.

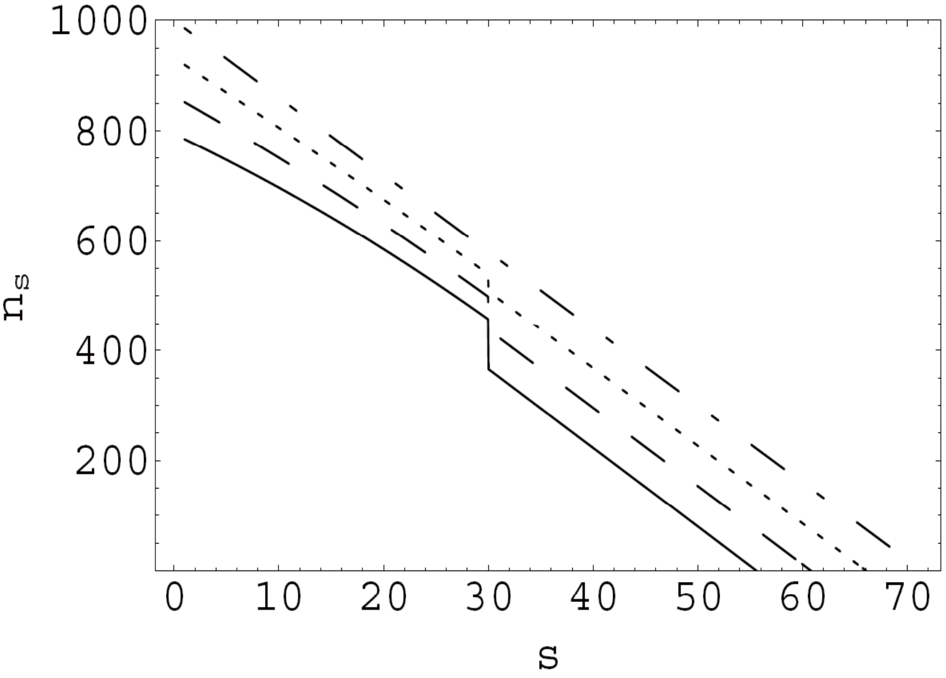


Figure 2.7: Predictions of the model of Section 2.3.3, using parameter values $C_1 = 1000$, $\phi = 14$, $p = 0.3$ and $L = -30$. This yields population size parameters of $N_0 = 850173$ and $M_0 = 35214$. Line styles reflect different values of the parameter t: dot-dashed ($t = 0$), dotted ($t = 5000$), dashed ($t = 10000$) and solid ($t = 15000$). Beyond $t = 15000$ it can be seen that the approximations made in the derivation of Section 2.3.3 become inaccurate.

2.3.4 Heterogeneity of members

In many real-world systems – in particular, biological or social systems – the population is heterogeneous. In addition to the basic question of whether approximating a heterogeneous system by a homogeneous model is justifiable, there is the deeper issue of how to formally introduce heterogeneity into coalescence-fragmentation systems. As we have seen in this chapter, small changes in coalescence-fragmentation rules can sometimes yield dramatic changes in the cluster size distribution, and vice versa. In other words, the "devil may be in the detail" in terms of the emergent phenomena that can be expected from a given set of microscopic rules. Our limited goal here is to explore some encouraging developments in this area, highlighting the circumstances in which the heterogeneity of the population allows an accurate description in terms of an effective homogenous model.

Reference [70] introduces a "character" to each object by means of an m-dimensional normalized vector which is formed from m-bit binary strings. The scalar product of any two such characters then becomes the argument of a function which controls the coalescence and fragmentation processes. The general case requires numerical simulation. Interestingly, however, this model produces a power-law over part of its range with a 2.5 slope which is identical to the homogenous E-Z model. Instead of the power-law exponent, it is the form of the exponential cut-off which turns out to depend on the heterogeneity of the population. We recently explored another type of heterogeneous E-Z-like model, showing that it can bridge the gap between the power-law slope of magnitude 2.5 for clusters in the E-Z model (and hence 1.5 for price returns) and the empirical value of financial market price returns which is typically closer to 4 [107]. A simple version of the vector model is provided via a fascinating recent variation proposed by Hui [104] in which the heterogeneity is represented by a character parameter $\rho_k \in [0, 1]$ which is assigned to each object in the entire population, where objects are numbered by $k = 1 \ldots N$. The probability that an agent i and another agent j form a link (and therefore for the inequivalent clusters to which these members belong to merge) depends on the value $|\rho_i - \rho_j|$. In principle it may be a general symmetric function $p(\rho_i - \rho_j)$. The fragmentation of a cluster may also depend on the characters of the members that form the particular cluster. One way of introducing this is by a mechanism in which fragmentation of the whole cluster is triggered by breaking any single link that belongs to it [104]. Since a weaker link is easier to break, it is assumed that the probability that the link breaks is proportional to $p(\rho_i - \rho_j)$ which may be interpreted as a measure of the strength of the link formed between members i and j. If $p(\rho_i - \rho_j)$ is a function which is sharply peaked at 0, we will have a situation where the newly formed clusters consist only of members of very similar character, and the whole system may be considered as a mixture of several homogeneous population subsystems which do not interact which each other. Each of these subsystems is described by the cluster size distribution of the form in Eq. 2.23 with constants determined by the distribution of characters across the population. The cluster size distribution for the whole system (regardless of the character) is then a sum of the distributions for the subsystems – therefore we still observe a scale-free behavior with variation in the form of the cut-off (i.e. diversity in the heterogeneity of the population induces diversity in the constants describing the subsystems, and hence lengthens the tail of the cluster size distribution). In the opposite limiting case, the function $p(\rho_i - \rho_j)$ does not vary sharply over its argument, e.g. $p(\rho_i - \rho_j) \propto 1 - |\rho_i - \rho_j|$, thereby yielding homogeneous mixing. The the distribution of characters across different clusters is uniform and the system can therefore be described as an effectively homogeneous one by Eqs. 2.12 and 2.23. The presence of the heterogeneity changes only the value of α/β in Eq. 2.12.

2.4 Conclusions and Implications

We have examined various coalescence-fragmentation systems, with the goal of elucidating how subtle changes in their underlying rules can affect the resulting distribution of cluster sizes. In the process, we have managed to connect the rules of coalescence and fragmentation with terms in the corresponding rate equations, and have identified the specific ways in which they affect the resulting distribution of cluster sizes. The connections are not always direct, but we have offered various insights which help establish a more direct link. In each case studied, the system senses the fragmentation function in two ways: the appearance of new clusters coming from the fragments of the fragmented cluster (represented by $G_F(s)$), and the disappearance of clusters that fragment (represented by $L_F(s)$).

As a result of our analysis, we can better understand what factors dictate when a power-law is likely to emerge, and what tends to control its exponent. We conclude that: (1) it is the substantial contribution of $L_F(s)$ in the equilibrium condition (Eq. 2.1) which may prevent the size distribution from showing a power-law behavior. (2) The presence or absence of $G_F(s)$ (i.e. the appearance of fragmentation products of new clusters) influences strongly the value of the power-law exponent itself, in cases where the power-law emerges. In the case where the parameter controlling the fragmentation is small but finite, it is hard to identify a common limiting case for the various systems studied – however, the form of the fragmentation function does influence the cluster size distribution regardless of the value of this parameter. Note that if the fragmentation rate tends to zero, the system cannot be clearly described using mean-field theory, since it performs quasi-oscillatory behavior associated with the build-up of one supercluster containing essentially the whole population, and this supercluster's eventual break-up. Whatever the mode of fragmentation, the exponent of the power-law may be controlled by altering the power of the cluster size s which is involved in the fragmentation and coalescence function. Specifying it realistically requires some detailed understanding of the system at the microscopic level. The most common mechanism of coalescence is created by building random links between the population members, yielding a coalescence function of the form $\sim ss'$.

If we adopt a point of view in which the system is considered as an evolving network, the clusters represent disconnected components. Depending on the particular rules, the fragmentation process now corresponds to breaking links. If the disconnected component in a network breaks predominantly into single members, it might be still interpreted in terms of the fragmentation being triggered by a single member, provided we allow some kind of link-breaking virus to spread rapidly throughout the entire disconnected component. Somewhat counter-intuitively, we have also seen that the behavior of the heterogeneous system does not substantially differ from the behavior of the homogeneous one. This results from two effects: the homogeneous mixing effect, and the coexistence of several non-interacting populations whose distinct "characters" lie hidden in the cluster size distribution.

Although we have mentioned various possible applications, we finish by noting a new one. Many of the neurodegenerative disorders associated with aging, for example Alzheimer's disease, are thought to be associated with the large-scale self-assembly of nanoscale protein aggregates in the brain [72]. Protein-aggregation has of course attracted much attention over the years in both the chemistry and physics literature – however, the problem of protein aggregates in neurodegenerative diseases is known to be much harder than traditional polymer problems, because of the complexity of the individual proteins themselves [72]. Given the wide range of possible heterogeneities *in vivo* within a cell, there is typically insufficient knowledge to specify either (i) a specific diffusion model and its geometry and boundary conditions, as a result of geometrical restrictions and crowding effects [105], or (ii) a specific reaction model for the binding rates, given the wide variety of conformational states in which molecules may meet. It therefore makes sense to assign some probabilities to the aggregation process – and in particular, coalescence and fragmentation probabilities to describe the joining of an n-mer with an n'-mer to give an n''-mer, where $\{n, n', n''\} \equiv 1, 2, 3, ...,$ and its possible breakup. The precise details of the coalescence and fragmentation rules now takes on a critical importance, since subtle changes in these rules can alter the resulting size distribution of the n-mer population. The practical question of how fatal a given realization of the disease will be in a particular patient, becomes intertwined with the question of whether the distribution of cluster sizes is a regular one in terms of its fluctuations – e.g. a Gaussian or Poisson distribution which both have a finite variance – or it is a power-law which may then have a formally infinite variance. Although in practice a cut-off always exists, a power-law with an exponent $\alpha < 2$ has (in principle) an infinite mean and infinite standard deviation; a power-law with $2 < \alpha < 3$ has (in principle) a finite mean but an infinite standard deviation; and a power-law with $\alpha > 3$ has a finite mean and finite standard deviation. The implication is that a coalescence-fragmentation process producing a power-law with $\alpha < 3$ as in E-Z-type models where $\alpha \sim 2.5$, has a significant probability of forming very large n-mers because of its (in principle) infinite standard deviation. Suppose for the moment that an n-mer of size $n \geq n_0$ can produce a neurodegenerative disorder, then the fraction of such dangerous n-mers in a soup of self-assembling polymer aggregates, will be non-negligible if $\alpha < 3$. In the highly crowded, heterogeneous n-mer population expected in the human body, the resulting value of any approximate power-law slope α could therefore be a crucial parameter to estimate. The possibility of engineering this α value such that large aggregates are unlikely, through subtle changes in the coalescence and fragmentation processes, then takes on a very real possibility. It also adds direct medical relevance which justifies further work on this topic in the future.

CHAPTER 3

Anomalous Slow Attrition Times in Interacting Dynamical Populations

3.1 Motivation

It is now more than fifty years since Richardson first uncovered empirical similarities in the total number of casualties for different wars [112]. Building on this, we presented some preliminary empirical analysis several years ago [113] which suggested a common power-law pattern in the frequency distribution of casualties within two individual wars, Iraq and Colombia. Specifically, our preliminary findings suggested that the probability that a violent intra-war event (i.e. a violent incident occurring within a war between the two opposing populations) produced x casualties, is well approximated by the power-law form $p(x) = Cx^{-\alpha}$ with C a positive constant and $\alpha \sim 2.5$. Intriguingly, this same approximate value $\alpha \sim 2.5$ was also identified by Clauset et al. for the dataset of global terrorist events [114]. Power-law behavior is known to be widespread across many physical and social systems [115, 116], however it was a surprise to find it describing individual events within a single war.

Here we explain how coalescence-fragmentation models from statistical physics can be adapted to understand this observation, and hence to develop promising quantitative descriptions of insurgent warfare. In the rest of Section 3.1, we present a simple model and discuss its attractiveness as a first-order approximation to a modern insurgency [28, 55]. Section 2 and beyond examines multiple population variants including different rules-of-engagement, with a focus on the war's duration. Comparisons to empirical data for duration of wars and casualties are encouraging, suggesting that this novel application of statistical physics to human conflict could have a very productive future.

3.1.1 Simple one-population model of insurgent dynamics

We refer to our basic one-population insurgent model as the "EZ" model [28, 55, 113]. Consider an insurgent force comprising many "agents", which are each casualty causing units. In the simplest case each agent is just a single fighter, but the definition also covers equipment such as explosives, or even information. We make the reasonable assumptions that (i) the insurgency does not have any external controller, (ii) groups of agents may spontaneously form and/or break up over time. These groups are neither fixed in size nor in number, and we will use the terms "group" and "cluster" interchangeably. A given group can merge with other groups to form larger groups, or it may fragment into single agents. We define the "attack strength" of a given group to be the average number of people who are killed or injured as a result of an attack involving that group. Each agent is taken to have an attack strength of 1, so that a cluster of size s has an attack strength of s. The number of clusters with a given attack strength is denoted by n_s, and the total attack strength of the population (which is equal to the total number of agents in it) is taken to be a constant, $N = \sum sn_s$. At each timestep in the model, a cluster (including those containing single agents) is selected with probability proportional to its size s. Equivalently, we could choose an agent at random from the population and then select the cluster to which it belongs. The probability of any cluster of size s being selected is therefore $P(s) = \frac{sn_s}{N}$. With a probability ν the group selected fragments into single agents. Otherwise, with a probability $(1 - \nu)$, a second group is selected with probability proportional to its size and the two clusters coalesce into a single group with size equal to the sum of the two constituent group sizes. The process then repeats at every subsequent timestep. For small fragmentation probability $\nu \ll 1$, this ongoing fragmentation and coalescence process results in a steady-state cluster size distribution [28, 118] featuring a power law with exponential cut-off:

$$n_s \approx N \left(\frac{4(1-\nu)}{(2-\nu)^2} \right)^s s^{-5/2} \ . \tag{3.1}$$

For small ν, the dominant s dependence is the power law with slope $5/2 = 2.5$ in agreement with the preliminary empirical observation in Ref. [113] for the wars in Iraq and Colombia, and the empirical observation of Clauset et al. for global terrorism events [114].

3.2 Two-population conflict model

We now add a second population comprising type B agents, with similar internal cluster dynamics. The second population comprises p_s clusters of size s and a total size $\sum sp_s = P$. A cluster is picked from this total population, $N+P$, with probability proportional to its size. In what follows, we will interchangeably refer to population A as having total size N_A or N, and population B as having total size N_B or P respectively – the specific choice will be clear from the context. The cluster then fragments with a fragmentation probability dependent on its population type, ν_A or ν_B. If the cluster does not fragment then a second cluster is selected from the total population, with probability proportional to its size. If the two clusters are of the

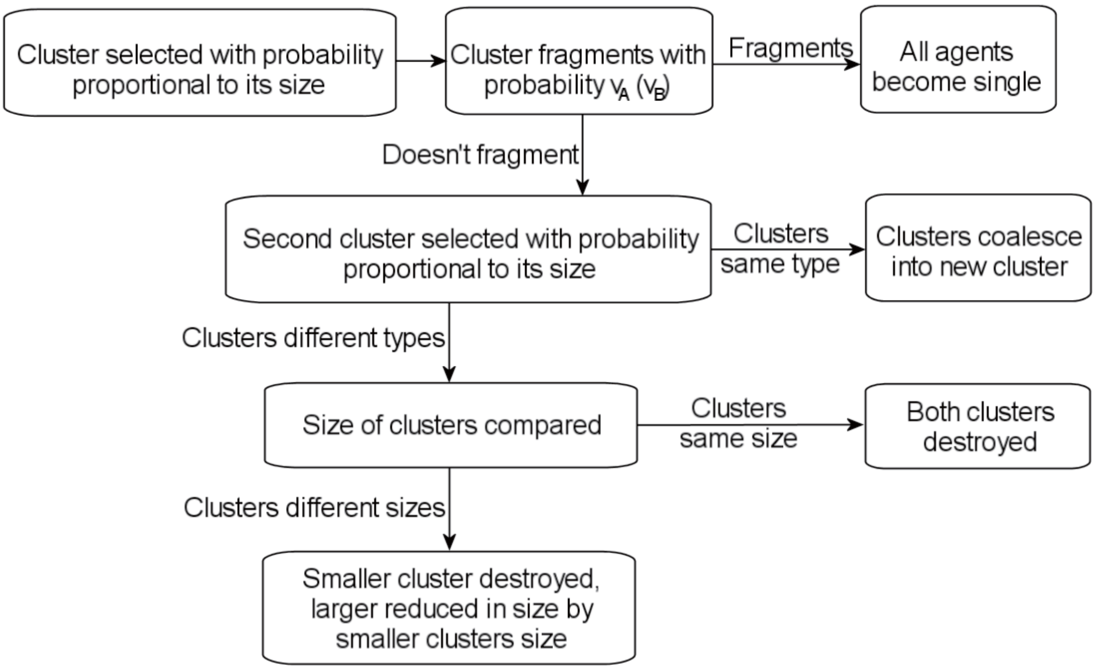


Figure 3.1: Flowchart of a two-population EZ model (TPEZ).

same type (A or B) they coalesce; if they are of different types then they interact. We start by employing very simple rules-of-enagagement for interactions, in order to illustrate the basic results. In an interaction, we assume that the smaller cluster is destroyed and the larger cluster is reduced in size by an amount equal to the smaller cluster's size. If they are both of the same size (but opposite type) then both clusters are destroyed. In this way, both populations lose the same amount of agents in any given interaction. These lost agents are then removed from the model. A flowchart of this model (referred to as TPEZ) is contained in Fig. 3.1. The initial A and B populations at timestep $t = 0$ are N_0 and P_0 respectively.

3.2.1 Numeric simulation results

A typical result from a numerical simulation is shown in Fig. 3.2. It shows the average cluster distribution (i.e. number of clusters of size s versus s) for both pop-

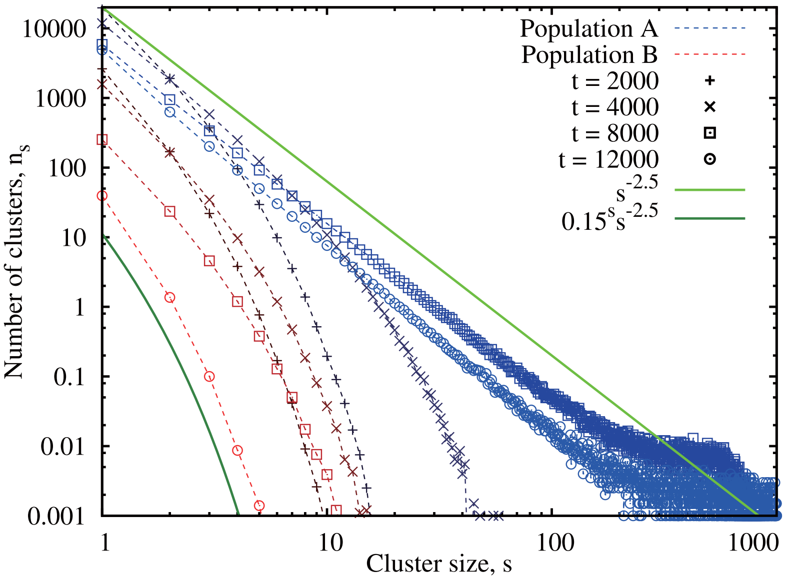


Figure 3.2: Cluster distribution for populations A and B at different timesteps in the TPEZ model. The A distribution has been rescaled by a factor of 5 for clarity. Initial conditions were $N_0 = 6000$, $P_0 = 4000$, $\nu_A = \nu_B = 0.01$, and all agents were initially single. The distribution is an average over 10^4 simulations.

ulations at various timesteps. The A distribution has been rescaled (displaced up the y-axis) to separate it from B for clarity. The larger A population develops into a power-law distribution, and this dependence remains as agents are destroyed. Both distributions continue to move towards the origin at higher timesteps, until no type B agents remain. From this point onward, the total A population stays constant as does its distribution. As both populations have their sizes reduced by the same amount in any interaction, the final A population is equal to the difference between the two initial populations. The power law exponent for both distributions is approximately 2.5, which is the same value as for the single population EZ model's steady-state distribution. The A population also exhibits the finite size effects observed in the EZ model; the power law becomes distorted as the cluster sizes reach the limit imposed by the total population size. This behavior is typical within the model, and is independent of the initial population sizes and fragmentation probabilities (ν_A, ν_B).

However, the the amount of time taken for the smaller population to be destroyed by the larger population (i.e. the war's duration) does depend on all of these variables. It is this duration that now becomes the focus of our discussion.

3.2.2 Analytic derivation of a war's duration

The time taken for one population to be destroyed (i.e. the time to extinction, or duration of the war) can be derived as follows. The probability Q_{AB} that any cluster of population A is selected and interacts with one of population B is the sum of the probabilities for an A cluster of size s to interact with any B cluster, $q_{AB}(s)$. The first factor in $q_{AB}(s)$ is the probability for a cluster of type A and size s to be selected, the second the probability for this cluster not to fragment, and the third factor is the probability for any cluster of type B to then be selected:

$$Q_{AB} = \sum_s q_{AB}(s) = \sum_s \frac{sn_s}{N+P}(1-\nu_A)\frac{\sum_r rp_r}{N+P} = \frac{NP}{(N+P)^2}(1-\nu_A) \tag{3.2}$$

using the fact that $\sum sn_s = N$, $\sum rp_r = P$. The probability Q_{BA} of selecting a B cluster and it interacting with an A is given by a similar expression, with ν_A replaced with ν_B. After an interaction, each population A and B is reduced by an amount equal to the size of the interaction (which is the size of the smallest cluster in the interaction). Introducing an average interaction size c, and starting from timestep $t = 0$, the populations then become $N = N_0 - c$, $P = P_0 - c$. After i interactions the populations will be $N = N_0 - ic$, $P = P_0 - ic$. Therefore the probability for an interaction between A and B after i previous interactions is

$$Q(i) = Q_{AB} + Q_{BA} = \frac{(N_0 - ic)(P_0 - ic)}{(N_0 + P_0 - 2ic)^2}(2 - \nu_A - \nu_B) \tag{3.3}$$

To reduce N_0 (and P_0) by c will take $1/Q$ timesteps on average. The total time to reduce one population to 0 is the sum of the timesteps required for each destructive interaction, until there are no agents left in the population. Taking P_0 to be the smaller population, it will require P_0/c interactions to destroy it, so the final interaction will happen after $P_0/c - 1$ previous interactions. Hence the war's duration is given by

$$
\begin{aligned}
T &= \sum_{i=0}^{\frac{P_0}{c}-1} \frac{1}{Q(i)} = \sum_{i=0}^{\frac{P_0}{c}-1} \frac{(N_0+P_0-2ic)^2}{(N_0-ic)(P_0-ic)(2-\nu_A-\nu_B)} \\
&= \frac{1}{(2-\nu_A-\nu_B)} \left(4\frac{P_0}{c} + \sum_{i=0}^{\frac{P_0}{c}-1} \left(\frac{N_0-P_0}{P_0-ic} + \frac{P_0-N_0}{N_0-ic} \right) \right) \\
&= \frac{N_0-P_0}{c(2-\nu_A-\nu_B)} \left(\frac{4P_0}{N_0-P_0} + \sum_{i=1}^{\frac{P_0}{c}} \frac{1}{i} - \sum_{\frac{N_0-P_0}{c}+1}^{\frac{N_0}{c}} \frac{1}{i} \right) .
\end{aligned}
$$

Using the standard result [119] that the harmonic series can be expressed as $\sum_1^n \frac{1}{i} = \gamma + \psi_0(n+1)$,where γ is the Euler-Mascheroni constant and ψ_0 the digamma function, and the fact $\sum_{a+1}^n = \sum_1^n - \sum_1^a$ allows us to express the duration T in this TPEZ model as

$$
\begin{aligned}
T = \quad & \frac{N_0-P_0}{c(2-\nu_A-\nu_B)} \\
& \left[\frac{4P_0}{N_0-P_0} + \left(\gamma + \psi_0(\frac{P_0}{c}+1) - (\psi_0(\frac{N_0}{c}+1) - \psi_0(\frac{N_0-P_0}{c}+1)) \right) \right]
\end{aligned}
\tag{3.4}
$$

This gives the war's duration in terms of the initial larger (N_0) and smaller (P_0) populations, their fragmentation probabilities (ν_A and ν_B) and the average size of a destructive interaction, c. Note that the interaction does not happen for each timestep, so c is not simply the average cluster/group size – however the variation from 1 is found from numerical simulations to be small and approximately linear, i.e. $c \approx 1 + 0.2(P_0/N_0)(1-\nu_B)(1-\nu_A)$.

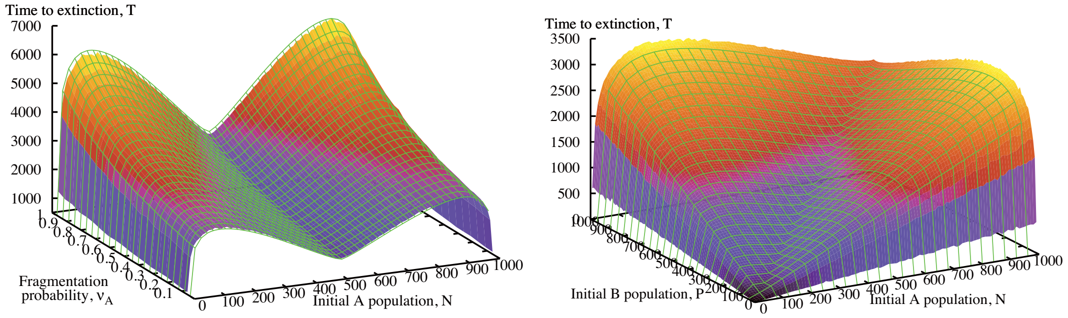


Figure 3.3: Time to extinction (T) in the TPEZ model as a function of initial populations (N, P) and fragmentation probability (ν_A), from numerical simulations and the analytic theory. Each data point is an average of 10^3 simulations.

3.2.3 Numeric simulation of a war's duration

Figure 3.3 shows the results from numerical simulations of the TPEZ model, together with the predicted analytic result from from Eq. 3.4. The agreement is good in all cases. T increases with fragmentation probability since interactions between the populations only occur when a cluster is selected and doesn't fragment, and interactions are required to destroy agents. T also increases with total population $N_0 + P_0$ since more agents require more time to be destroyed. The dependence of T on the ratio of N_0 to P_0 is in stark contrast to the expected behavior from mass-action theory [120], which would have instead suggested that a strong opponent would destroy a weaker one more quickly than if the two sides were of comparable strength. The numerical simulations and the analytic theory show that the opposite is actually true: The larger the relative imbalance in strengths, the longer the fight lasts. A population of 100 and 900 agents takes considerably longer to decay to extinction than two equal populations of 500. This surprising result can be understood by looking at the average number of "events" that occur in a given model war as a function of initial A population (keeping total population constant, so B population $P_0 = 1000 - N_0$),

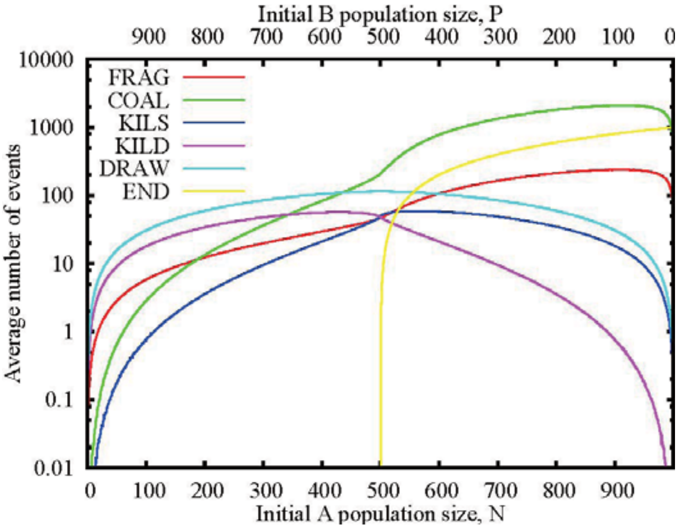


Figure 3.4: Average number of events as a function of relative initial populations within the TPEZ model, with $N_0 + P_0 = 1000$, $\nu_A = \nu_B = 0.1$.

see Fig. 3.4. The events are:

FRAG – Type A cluster selected and fragments

COAL – Type A cluster selected and coalesces

KILS – Opposite type clusters selected, A bigger

KILD – Opposite type clusters selected, B bigger

DRAW – Opposite type clusters selected, same size

and **END**, which is not an event but the final A population.

The number of type A coalescences and fragmentations (inter-population activity) increases rapidly as A become the majority population. (Note the logarithmic scale). At the same time the number of destructive interactions (KILD, KILS, DRAW) between the populations decreases. As the probability for any cluster being selected is proportional to its size, the probability for any population being selected is proportional to its total size. For unequal populations there is a greatly increased probabil-

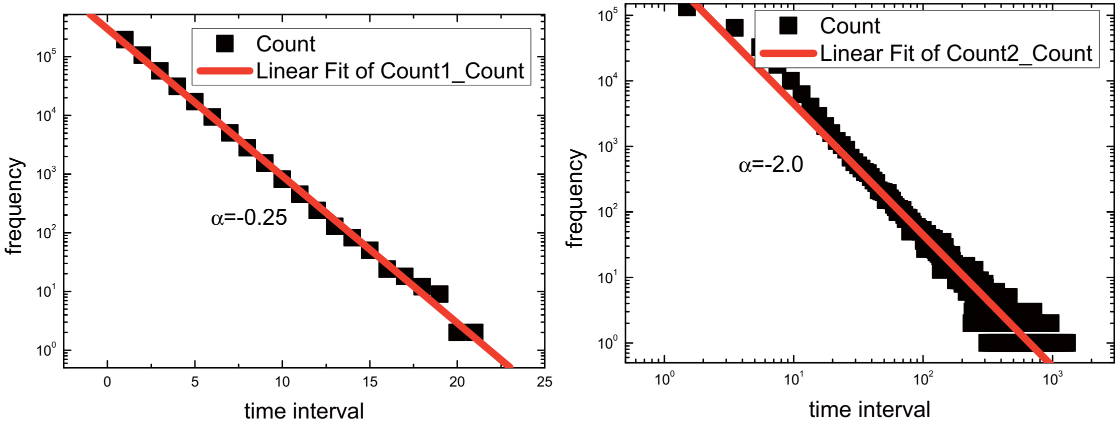


Figure 3.5: Frequency count of the time intervals between reactions. Left panel is a log-linear plot for $N_0 : P_0 = 5 : 5$, while the right panel is a log-log plot for $N_0 : P_0 = 7 : 3$.

ity for the larger population to self-interact (i.e. the same population being selected twice) as compared to equal populations where the probability for self and opposite interaction is the same. The larger population effectively gets in the way of its own search for the smaller population. As destructive events only occur with opposite interactions, this results in an increase in the time between agents being destroyed in asymmetric populations. This is not offset by the decrease in time due to fewer agents to destroy, leading to a net increase in the time required for extinction. Numerical simulation shows that the distribution of time intervals between interactions of A and B clusters shifts from exponential to power law, as the portion of $N_0 : P_0$ changes (see Fig. 3.5). While the result is unexpected, remarkably it reflects the empirical observation in Fig. 3.6 that asymmetric wars take longer to resolve than those in which the sides are of comparable strength [121].

3.2.4 Comparison of model and real war durations

In Fig. 3.6, the upper thick curve shows the theoretical T while the lower two curves show the mass-action predictions. The mass-action equations that we em-

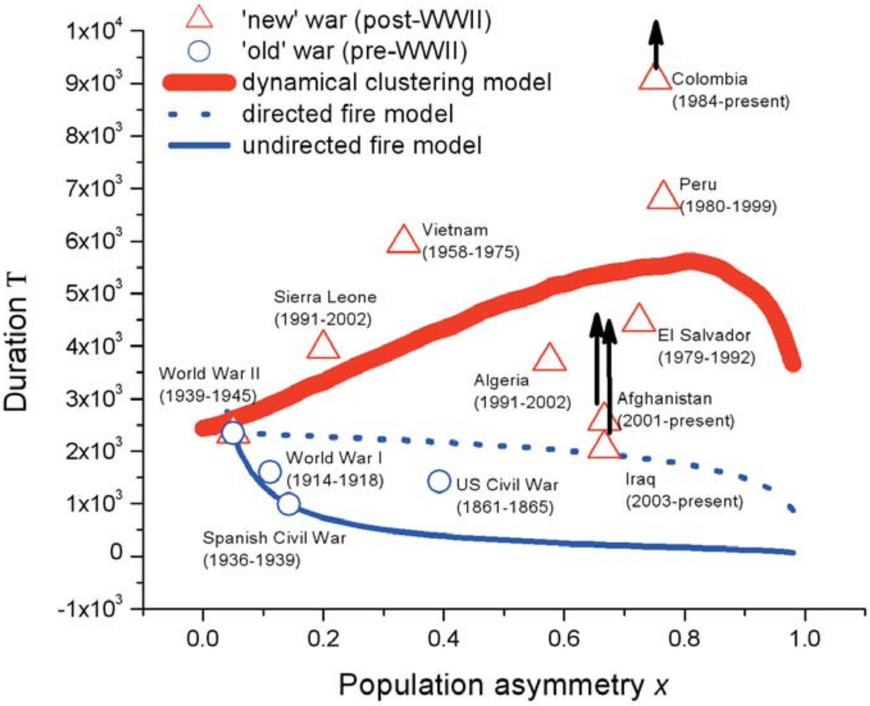
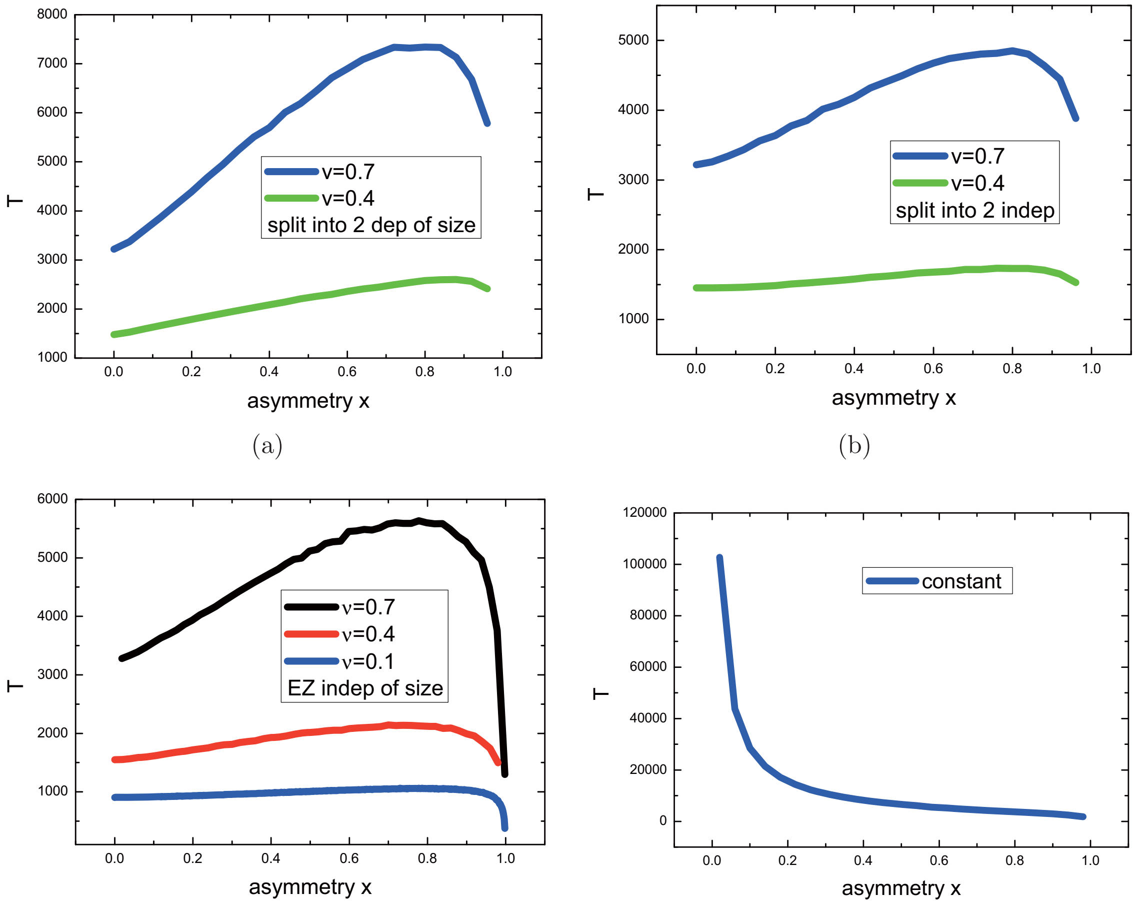


Figure 3.6: Adapted from Ref. [120]. Duration T of human conflicts as a function of asymmetry x between the two opposing military populations. $x = |N_0 - P_0|/(N_0 + P_0)$. Data are up to the end of 2008; hence, final data points for the three ongoing wars will lie above the positions shown, as indicated by arrows. The lower two lines are the mass-action results. The upper thick curve [i.e., Eq. 3.4] is generated using $\nu_A = \nu_B = 0.7$ and $N_0 + P_0 = 1000$ fixed. Changing ν_A and ν_B changes the height of the theoretical peak but leaves qualitative features unchanged.

ploy are those traditionally used for wars of attrition [122–128]: (1) $dN(t)/dt = -aN(t)P(t)$, and $dP(t)/dt = -aN(t)P(t)$ called Lanchester's undirected mass-action model, and (2) $dN(t)/dt = -bP(t)$, and $dP(t)/dt = -bN(t)$, called Lanchester's directed mass-action model, where a and b are constants. We take World War II as the dividing point between "old" wars and "new" wars. "Old" wars are well described by the mass-action models, while "new" wars are closer to our model prediction, implying an absence of grouping dynamics in "old wars" [129, 130].

3.2.5 Model variants



(a)

(b)

(c)

(d)

Figure 3.7: Duration T of model war. For (a) and (b), groups fragment into two randomly-sized groups. In (a), group is picked proportional to its size, while in (b) picking is independent of the group size. For (c) and (d), the rules follow the basic model, except that picking for (c) is independent of group size, while in (d) the destroyed agents are replaced by inactive members so that the total population in the system remains a constant.

We start with two explicit variations in the rules, before looking at more general model variants: (1) a group fragments into two random-sized groups [54]; (2) the group is picked independent of its size. Figure 3.7 shows the dependences. With the exception of Fig. 3.7(d), all variants (a), (b) and (c) retain the main features of the basic model, no matter how the group fragments or is picked. However, if we replenish the destroyed population by inactive members in order to keep the population constant, the curve reverts to the classic mass-action shape, i.e. maximum at symmetric point $x = 0$, as the lower two curves in Fig. 3.6. For the basic model as well as no-replenishment models, two opposing populations actively seek to fight with each other – by contrast, adding an inactive population to keep a constant total population, mimics the situation where the size of the space in which they are fighting is fixed. In short, this inactive sub-population now acts like a solvent that separates or delays the clashes and hence the conflict.

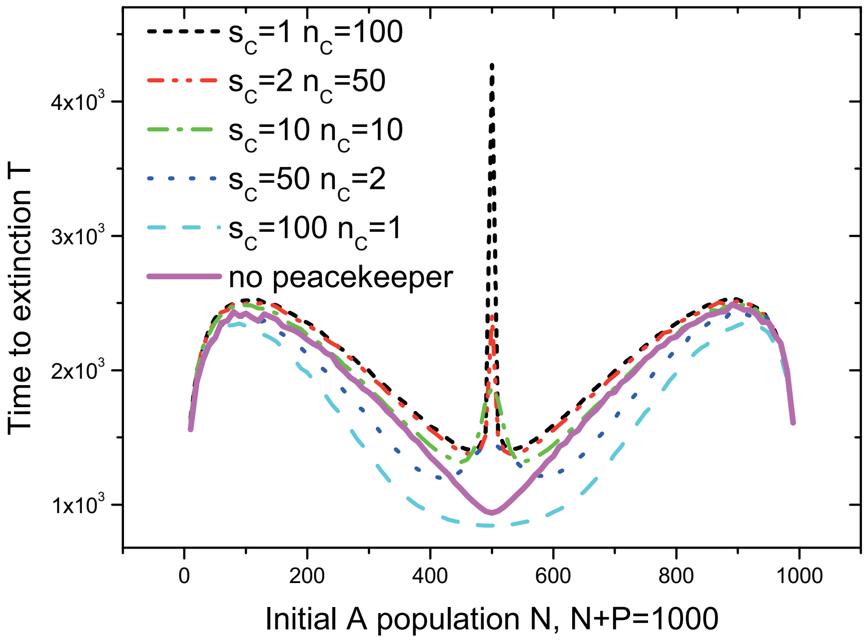


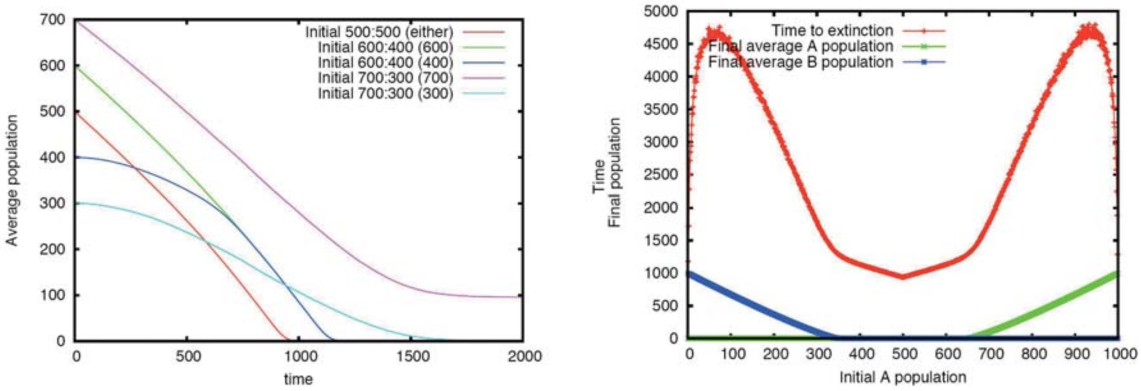
Figure 3.8: Effect of different allocations of 100 peacekeepers in the TPEZ model.

Peacekeeper variant

One topically pertinent model variant can be generated by adding a third species (population C). Like a peacekeeping force, C can block interactions. For simplicity, we assume the N_C members of C are permanently arranged into n_C groups, each with s_C permanent members. A group is selected as in the original model. If it is of type A or B, then it can fragment as usual with a probability ν_A or ν_B. If it doesn't fragment, then a second cluster is selected. If this cluster is A or B then coalescence/ fighting proceeds as before. If this second cluster is C, then a third cluster is selected. If this cluster is of the same type as the first, then they coalesce. If it is the opposite (A or B) type, then the C cluster is compared to the size of the selected A and B clusters. If it is greater than or equal in size, then there is no clash – otherwise, the A and B clusters fight as before. Figure 3.8 shows that if C comprises only a few large groups, then T decreases irrespective of the asymmetry. Having a few large C groups means that some sizable battles can be blocked; however, it also allows the buildup of sizeable groups of both A and B, which in turn makes the typical size of interactions bigger. By contrast, if C comprises many small groups, T can be much larger, showing a huge increase around the symmetric populations case (i.e. $x = 0$). If real-time management of the C population is possible, this duration profile T can be manipulated even further.

Minority advantage variant

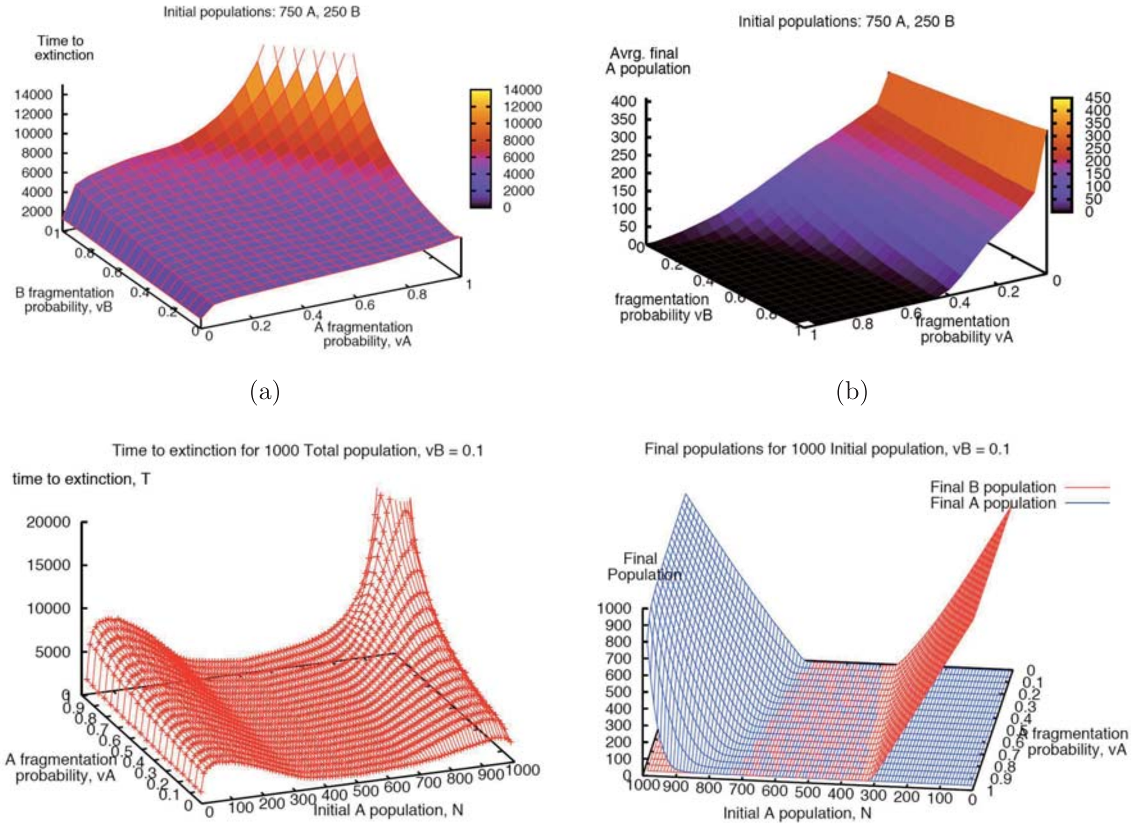
We next consider a class of model variant in which we change the behavior in the case of a draw (i.e. when two clusters of opposite species but the same size are selected). Previously both clusters were destroyed – but now, only the cluster belonging to the larger total population is destroyed. This could mimic a particular "home advantage" for a cluster from the minority population B, when faced with an equal-sized cluster of the invading army A. The first difference in the results from the original model is that the larger population does not always win any more. It has a larger probability of winning, but it isn't guaranteed. The larger population



(a) (b)

Figure 3.9: Left: Minority advantage variant, showing average populations as a function of time. Right: time to extinction (T) as a function of the initial A population (N_0) $N_0 + P_0 = 1000$.

loses members at a faster rate than the smaller, as shown in Fig. 3.9(a). The total initial population in the graph is constant at 1000, so the 600 and 400 groups are part of the same simulation as the 700 and 300. If the smaller population has enough members for it to last until the two populations are the same size (e.g. 600 : 400 case) then both populations decrease at the same rate until they are both destroyed. If the smaller population is destroyed before it reaches the same size as the larger (e.g. 700 : 300), then the larger population remains at the end on average. It is possible in the 700 : 300 for the two populations to reach the same size and both be destroyed. However, this is very unlikely – in contrast to the original model where it was simply not possible. Despite the difference in the rules as compared to the basic model, the graph of T (Fig. 3.9(b)) has a similar shape. This gives us further confidence that the results of the basic model are robust. The major difference is the sudden change in gradient which occurs at around $N_0 = 350$ and $N_0 = 650$. Between 0 and 350, A is too small to have any significant chance of becoming the same size as B, so B remains in the final state. Between $N_0 = 350$ and 650, A does have a high probability of catching up with B (or vice versa above 500), with both populations being destroyed. Above 650 then it is the same as below 350 but reversed; A cannot be caught up by B so it remains at the end. These effects determine T.

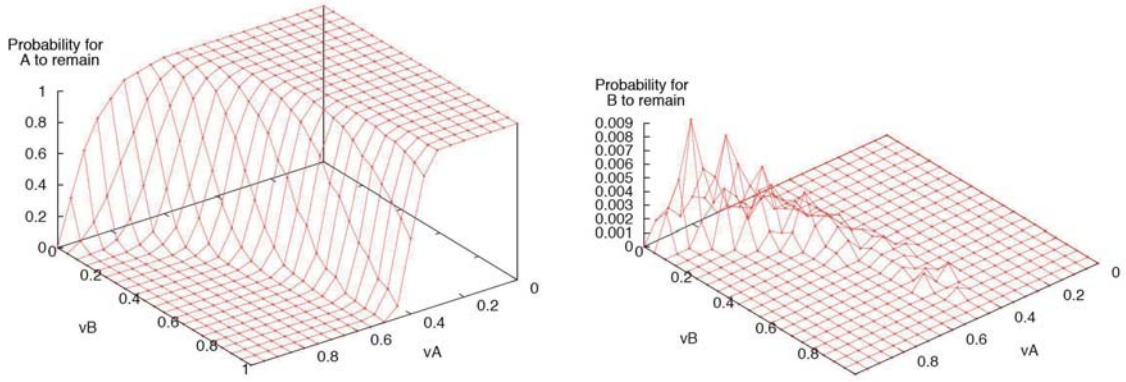


(a) (b)

(c) (d)

Figure 3.10: With minority advantage (a) duration dependence on fragmentation probability, $N_0 = 750$, $P_0 = 250$; (b) final state population dependence on fragmentation probability, $N_0 = 750$, $P_0 = 250$; (c) duration dependence on initial population and fragmentation probability, $N_0 + P_0 = 1000$, $\nu_B = 0.1$; and (d) final state population dependence on fragmentation probability, $N_0 + P_0 = 1000$, $\nu_B = 0.1$.

The duration T depends on fragmentation probability in a similar way to before; if either ν is higher then there are less fatal interactions and so the war lasts a longer time, leading to a large peak at ν_A, $\nu_B \approx 1$. The dependence isn't quite symmetric though since the time actually depends more strongly on the minority population's fragmentation probability. This is because if ν_A is high then A is mostly single and



(a) (b)

Figure 3.11: Both correspond to the minority advantage variant: (a) Probability for majority population to remain and (b) probability for minority population to remain.

many draws occur, reducing A until it is the same size as B. Both populations then reduce to 0 quite quickly. If ν_B is high though then B will be mostly single and can be killed off by A. Killing off a smaller population takes longer than two-equally sized populations, and even longer in this model because draws don't reduce the smaller population. The final state population (Fig. 3.10(b)) varies strongly with ν_A (majority rate) and slightly with ν_B. This is for the same reasons as T's variation. If ν_A is high then there are lots of single A agents, so lots of draws and both populations go to 0. If ν_A is low, then A kills off B effectively so it remains at the end. The ν_B dependence is similar but less important: If there are lots of single B (high ν_B) then there are more draws, and A and B become of the same size and both effectively disappear. The graph of initial population and fragmentation probability versus T (Fig. 3.10(c)) is similar to the previous model, with two differences. First, while the center section ($N_0 \approx 300$ to 700) is the same, the "wings" are much higher. This is because the dominating process for removing agents is draws. If the minority population is too small to be able to reach the majority (which happens at around $N = 300$) then there are still lots of draws which reduce the majority population but not the minority. This leads to an increase in time to destroy the minority: It can't win but all the draws slow down the process. The other major difference is the huge peak as $N_0 \approx 900$ and $\nu_A \approx 0.9$. The high fragmentation rate means that A is mostly single, while the lack of B agents means they don't get a chance to coalesce. In this case in the basic model, draws would occur and kill off B. In this model however, draws just reduce the much larger A population. In order for B to be destroyed there needs to be a "big A meets small B" event, which is very rare with these distributions, so the time required is very large. The final state population of the A is also lower than for lower ν_A values, as all the draws reduce A. The final state populations reflect these differences (Fig. 3.10(d)). When A is the minority the final B population is a straight line, it does not depend on ν_A. When A is the majority population though, the amount of fragmentation in A has an effect: The final population is slanted, with higher ν_A leading to lower final populations. A majority population having a low fragmentation probability (forming large clusters) leads to the quickest removal of B, with the most agents left. For the minority population, having a larger ν (forming single agents) is the best plan, since this has the highest chance of destroying the other population along with yours, and also takes the longest for you to be destroyed. There are three possible outcomes to the model; either the majority A population only, the minority B population only or neither population remains at the end. The probability for the majority win outcome is shown in Fig. 3.11(a) as a function of fragmentation probability, for a fixed initial concentration of $N_0 = 750$, $P_0 = 250$. The probability at the same conditions for the minority win outcome is shown in Fig. 3.11(b). As the probability for the minority win is very small, the probability for no population to remain is the inverse of Fig. 3.11(a). The position and shape of the boundary depends on initial population concentrations.

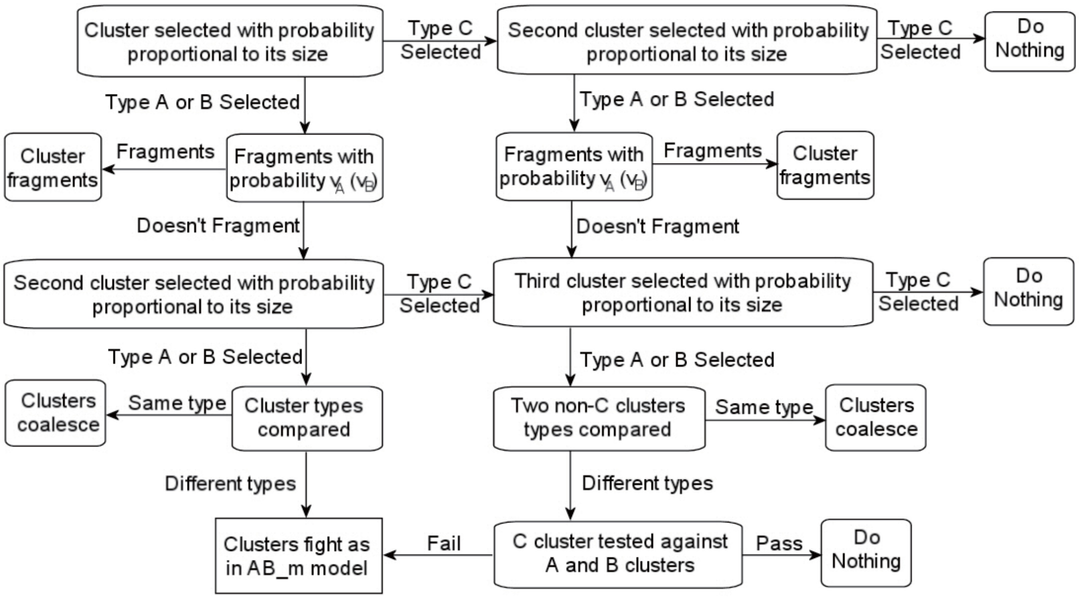
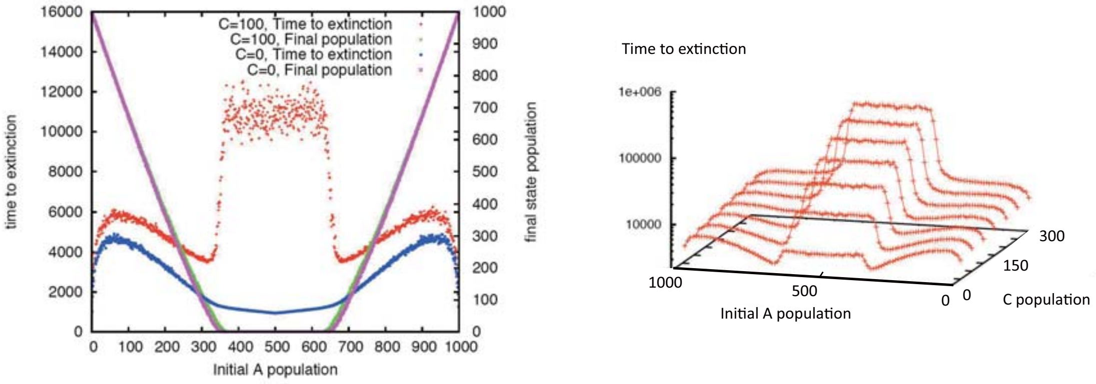


Figure 3.12: Flowchart for peacekeeper with minority advantage variation. "AB_m" means minority advantage.

Peacekeepers with minority advantage

This variation adds a third species (population C) to the minority advantage model as described in flowchart 3.12. The evolution of the system to a steady state with one, or both, populations becoming extinct proceeds as before. The presence of the third population C does increase the time required for a population to be destroyed however. This time increase is most dramatic in the region where N and P are close enough together for both to be destroyed, as shown in Fig. 3.13(a). This graph is for $N_C = 100$ or 0, $N_0 + P_0 = 1000$. The increase in duration time T with N_C is displayed in Fig. 3.13(b). All C agents were single, and there was no "test" for C, i.e. it always blocked an interaction if it was present. The large increase in time with C when both populations become extinct, can be explained. In these cases both N and P become very small, while N_C remains constant. As the probability for a particular population being chosen is proportional to its total size, the most likely event is then



(a) (b)

Figure 3.13: (a) Time to extinction (i.e. duration T) and (b) final population for $N_0 + P_0 = 1000$, and $\nu_A = \nu_B = 0.1$ with peacekeeper and minority advantage. However, there is no "test", i.e. whenever C is picked, there is no fight.

two C's being selected, leading to nothing happening for that time step. The next most likely process involves a C and any other two clusters – since C blocks any fatal interaction with which it is involved, nothing then happens. By contrast when one population remains at the end, this population still has a probability comparable to C of being selected, so the probability of agents being removed is not reduced by the same magnitude.

3.3 Encounter fragmentation model

The EF model also involves two populations of agents, the same as TPEZ, which can group together to form clusters within their populations. On each timestep a cluster is selected from the total population (A+B) with uniform probability, so each cluster has the same chance of being selected. A cluster is then selected out of the total population with probability proportional to its size, resulting in each agent having the same probability of being selected. The two clusters are then compared; if

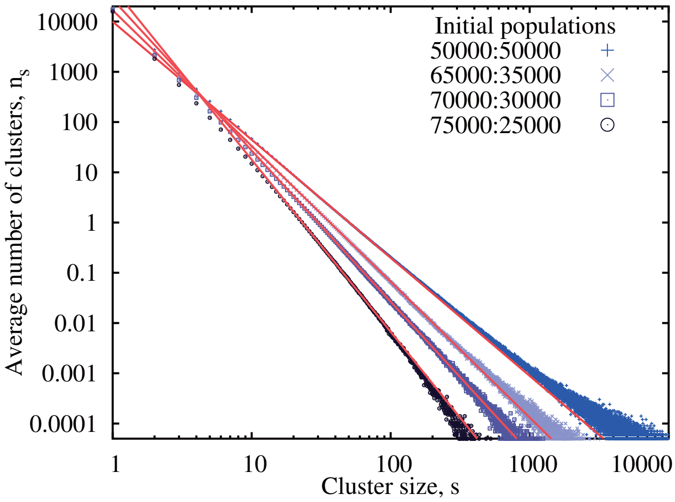


Figure 3.14: Steady state distribution of minority population cluster sizes in encounter fragmentation (EF) model. Total population $N + P = 10^5$ is constant, and each distribution is an average of 10^5 simulations. The lines are LSR fits to the data, see Table 3.1.

they are of the same type (A or B) they coalesce. If they are of different population type then the smaller cluster fragments and the larger is unaffected, or if they are the same size both clusters fragment.

3.3.1 Numerical simulation results

Numerical simulations of the model show that the steady state involves the larger population coalescing into a single massive cluster, with size equal to its total population. The smaller population in contrast is distributed into a range of cluster sizes in the steady state, with the distribution given by a power law with an exponent depending on the relative initial population sizes. We find that the steady state is independent of the initial conditions – for example all agents could start single or both populations could start in a single cluster. Figure 3.14 shows the cluster distributions obtained for various initial populations, along with power law fits to the data. Note the presence of finite size effects distorting the power law for cluster sizes nearing the population limit, similar to the case in the EZ model. This limit can be raised by using larger populations, albeit at the expense of computation time. Using well established methods [113, 114] the power law coefficients can be robustly determined. As power laws diverge as $x \rightarrow 0$ then distributions of the type $p(x) = Cx^{-\alpha}$ only exist above some minimum value x_{min} [116]. Above x_{min}, α can be found using a maximum-likelihood function: $\alpha = 1 + n\left[\sum_{i=1}^{n} ln\left(\frac{x_i}{x_{min}}\right)\right]^{-1}$ where x_i is a single measurement (i.e. in our case a single cluster of size s) and n the total number of measurements above x_{min} (the total number of clusters). The Kolmogorov- Smirnov goodness-of-fit test is used to compare the distribution $Cx^{-\alpha}$ with the simulation data, and by minimizing the D-statistic from this test, x_{min} (and hence α) is found. As a check α is also estimated using least squares regression. Table 3.1 summarizes the coefficients determined for different initial population ratios (as power laws are scale independent, α is independent of the total population). The confidence intervals of α determined from the maximum likelihood method were calculated using bootstrap re-sampling [132] of the distribution with 1000 replications. The limits stated are at the 0.95 confidence level. The uncertainties given for the least-squares regression α are the fit standard errors. The estimated and calculated values for α agree in all cases to at least two significant figures (except $5:15$, where all three values differ by approximately 0.1, suggesting that the low upper cut off in cluster size at $s \approx 300$ is adversely affecting the statistics). Least squares regression is known to be unreliable in this application [133], and the analytic value is technically only valid for $N \rightarrow \infty$, accounting for the minor discrepancies between the results. For comparison, the values obtained for the standard single population EZ model (analytic value $-5/2$) are $\alpha_1 = 2.66$, $\alpha_2 = 2.60$ [28].

N:P	α_1	α_-	α_+	α_2	α_e	x_{min}
10:10	2.3261	2.3250	2.3267	2.3498(1)	2.3333	21
9:11	2.3919	2.3907	2.3929	2.4007(1)	2.4337	23
8:12	2.5384	2.5374	2.5398	2.5197(1)	2.5625	24
7:13	2.7463	2.7444	2.7479	2.7020(2)	2.7316	27
6:14	3.0360	3.0328	3.0396	3.0075(7)	2.9608	25
5:15	3.5311	3.5240	3.5393	3.423(3)	3.2857	30

Table 3.1: Encounter fragmentation model cluster distribution coefficient, as determined from numerical simulations, α_1 using maximum likelihood (ML) and Kolmogorov-Smirnov test and α_2 using least squares regression and analytic solution α_e, see Sec. 3.3.2. The value in parenthesis is the error in the last digit. Also shown is the minimum x value above which the power law holds (determined by K-S test). α_+ and α_- are the 95% confidence limits for the ML estimate, determined by bootstrap resampling.

3.3.2 Analytic solution

We present an outline derivation of the steady state cluster distribution of the smaller population, with full details in the Appendix A. Defining n_s and p_s as the number of clusters with size s from population A and B respectively, we can construct equations describing their change between timesteps. For $s \geq 2$:

$$\begin{aligned}
\frac{dn_s}{dt} = \quad & -\frac{n_s}{\sum n_{s'} + \sum p_{s'}} \frac{\sum_{s' \geq s} s' p_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} - \frac{\sum_{s' \geq s} p_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{s' n_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} \\
& -\frac{n_s}{\sum n_{s'} + \sum p_{s'}} \frac{s' n_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} - \frac{\sum_{s' \geq s} n_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{s' n_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} \\
& + \frac{\sum_{s'}^{s-1} n_{s'} (s - s') n_{s-s'}}{(\sum n_{s'} + \sum p_{s'})(\sum s' n_{s'} + \sum s' p_{s'})}
\end{aligned} \tag{3.5}$$

and for $s = 1$:

$$\begin{aligned}
\frac{dn_1}{dt} = \quad & -\frac{n_1}{\sum n_{s'} + \sum p_{s'}} \frac{\sum s' n_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} - \frac{\sum n_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{n_1}{\sum s' n_{s'} + \sum s' p_{s'}} \\
+ \quad & \frac{\sum_{s'=2} s' n_s}{\sum n_{s'} + \sum p_{s'}} \frac{\sum_{r' \geq s'} r' p_{r'}}{\sum s' n_{s'} + \sum s' p_{s'}} + \frac{\sum_{s'=2} p_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{\sum_{r'=2}^{r' \leq s'} r'^2 n_{r'}}{\sum s' n_{s'} + \sum s' p_{s'}}
\end{aligned}$$

Similar equations hold for population B. The first two terms on the RHS of Eq. 3.5 are due to a cluster selecting (being selected by) a larger cluster of the opposite population, which causes it to fragment. The next two terms are due to a cluster of size s selecting any other cluster from the same population and coalescing with it. The final term is due to two clusters joining together to form a new cluster of size s. In Eq. 3.6 the first two terms correspond to a single agent being selected to coalesce, and the last two terms with a larger cluster fragmenting into single agents. Since we know that the steady state involves the larger population forming a single cluster, then taking B to be the larger population ($P > N$, where N and P are A and B's total populations) we can simplify the above equations by using the fact that in the steady state $\sum p_{s'} = 1$, $\sum s' p_{s'} = P$, $\sum s' n_{s'} = N$. We also make the approximation that $\sum n_{s'} \gg 1$, so that the probability of a B cluster being picked first is negligible, which is valid for large N. As derived in detail in the Appendix A, the steady-state solution to these equations is:

$$n_i = n_1 \frac{\left(\frac{N}{P} + \frac{P}{N} + 3\right)!}{\left(\frac{(N+P)^3}{N^2 P + 2NP^2}\right)!} e^{\left(\frac{P}{N}\frac{N}{P+2}+2\right)} i^{-\left(\frac{P}{N}\frac{P}{N+2P}+2\right)} \tag{3.6}$$

which is valid for $i \gg 1$ and $P > N$. This gives a power law cluster distribution of the form $n_s = Cs^{-\alpha}$ with C an irrelevant constant and α dependent only on the initial population sizes, N and P.

3.3.3 Variants and modifications

We now describe several ways in which the encounter fragmentation (EF) model can be extended.

Casualty variation, EFF

The EF model is altered to include a fighting element in interactions between opposite populations. This results in agents being destroyed (removed from the system)

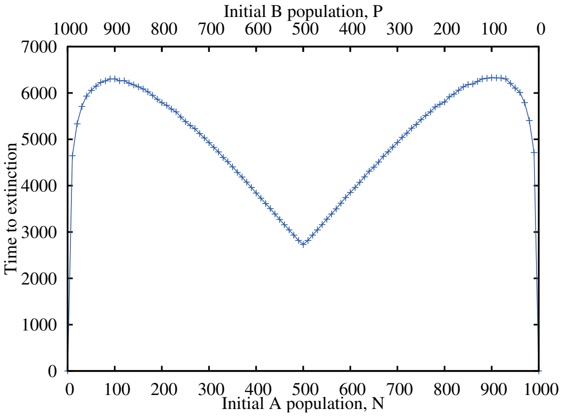


Figure 3.15: War duration dependence T on initial populations in EF model (with casualties). Initial total population $N + P = 1000$.

when an A and B cluster are selected. Previously the smaller cluster would have fragmented, but now the smaller cluster loses half its agents before fragmentation occurs. The larger cluster loses an amount of agents equal to the number lost by the smaller cluster, but does not fragment.

Numerical simulations show that the end state of the system involves the initially smaller population being completely destroyed, while the other population remains in a single cluster with size equal to the difference between the initial populations. This is reminiscent of the result of the two population EZ model discussed in Sec. 3.2, and indeed the time required for one population to be destroyed follows a similar dependence on the initial populations as that model. As shown in Fig. 3.15, for a constant total population, the duration T is much greater for an asymmetric population (e.g. $900 : 100$) than for a more even population (e.g. $600 : 400$).

Civilian population variation, EFC

A third "civilian" population is introduced into the EF model. This population (type C) can only be selected in the second selection step, and is selected in this step with probability proportional to its size, the same as for the A or B population. Note that as the C population can only be selected second it cannot interact with itself (coalesce), only with the A or B population. If an interaction with the C population occurs, C takes damage equal to the size of the A or B cluster selected first, whereas this cluster is unaffected. The size of the damage, and the population which inflicted it, is recorded. The C population has no effect on the behavior of the model. The distribution of damage event sizes reflects, in the steady state, the distribution of the population causing the damage. The size of the C population scales this event distribution, but otherwise has no effect. Therefore provided that the C population is large enough so it cannot be wiped out, then it makes no difference if C agents are actually destroyed in interactions.

Recruitment variation, EFR

In this variant agents are added to one or both populations at the start of each timestep. The mechanism used to do this can be one of several; new agents can be created singly, or can be attached to an existing cluster selected at random or with probability proportional to its size. New agents can also either be added at a constant rate, or created whenever the population is below its initial level. Additionally, all the above can be done with agents removed instead of added. Without any mechanism for destroying agents then this system cannot achieve a steady state. However, provided the rate of adding new agents is not too fast (less than 1 per timestep) the effect is simply to change the overall population size, regardless of which actual mechanism is used to attach the new agents. The minority population remains distributed in a power law, with an exponent which depends on the ratio of populations. As one (or both) population sizes are changing, then this power law also changes with time. The same is also true if agents are removed, except in this case one population will

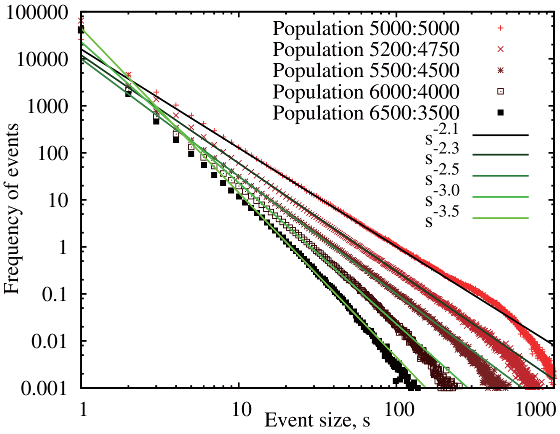


Figure 3.16: Distribution of B population damage to C in EF model with casualties, for different initial A and B populations. The graph is an average of 10^4 simulations.

eventually become extinct.

Army variation, EFA

As it is observed that the larger initial population invariably forms a single maximum sized cluster, we may take this population (conventionally the A population) as being constantly distributed in one or more large clusters, rather like a conventional army. The model operates as previously, except that now A no longer fragments or coalesces, and as such starts and ends in large groups. The variation produces the same power law distribution of the B population as before, while the A population remains in its initial distribution. The duration T is insensitive to whether the majority population exhibits internal grouping or not [120].

3.3.4 Combination of variants

Casualties and civilian population

This is a combination of the EFF and EFC variants. While the cluster distribution of the two populations does not become stable when both populations are losing

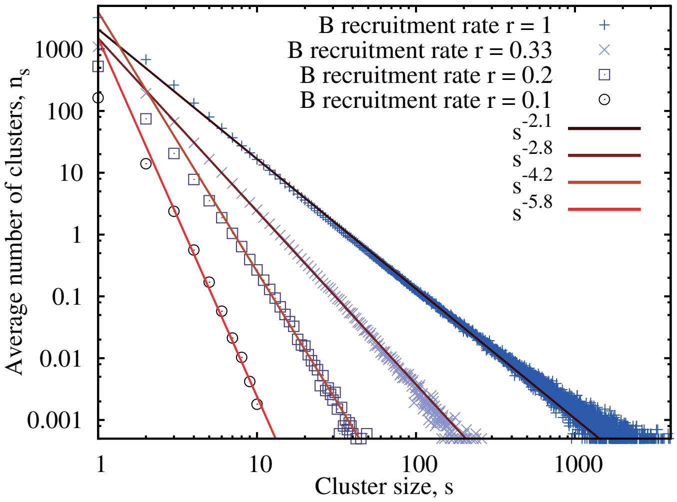


Figure 3.17: Minority (B) cluster distribution in steady state for encounter fragmentation model, with army, casualties and insurgent reinforcements. The AB interaction distribution is identical to the distributions shown. Also shown are approximate power laws. The type A population was 10^4, changing this would only affect the scale. The distributions are an average of 10^4 simulations.

agents, the distribution of the size of damage inflicted upon the C population in an interaction does. This event distribution, the number of events which cause a given amount of damage to C, is shown in Fig. 3.16 for several different initial A and B populations. Again, it is the ratio between the population sizes which is important, any change in total population just alters the scale. The graph shows the distribution of events between the minority (B) population and C; the distribution of events between the majority population (A) and C is a power law which varies only from -2.1 (the value for equal populations, shown on the graph) to -2.0 as A becomes larger than B. Interestingly, the distribution of the number of agents killed in an interaction between A and B (the AB event distribution) is identical to the BC event distribution, except for its scale.

3.3.5 Army, reinforcement and casualties

This is a combination of the EFA, EFR and EFF variants. The A population is now explicitly a conventional army, as such its distribution does not change. In fighting interactions then the A population does not lose agents, however B (the insurgent population) does as normal. This can be interpreted as the army reinforcing its units to full strength after any clash. The B population in contrast is replenished as in the reinforcement variation; with a probability r each timestep it receives a new agent, which is added to an existing cluster selected with probability proportional to its size. The effect of agents being both introduced and destroyed is that the system achieves a steady state. The distribution of the B population in the steady state is shown in Fig. 3.17. As can be seen it is a power law with an exponent which depends only on the rate of new agents being recruited. The A population distribution is fixed, and the distribution of fatal interaction events between A and B is identical to the B cluster distributions shown in the figure (except for the scale). In the steady state, the size of the A population has no effect at all on the AB event distribution, and no effect on the B cluster distribution except for determining the total population size at which B will be stable, so that the ratio of N to P is fixed for a given value of r.

3.3.6 Comparison with Conflict Data

The final model developed has many features in common with guerrilla warfare. Guerrilla armies are known to organize from the ground up [134], with small groups forming and then joining up with other groups. In encounters with a larger force they also tend to fragment and withdraw [134]. Recruitment can also be erratic, with new members almost always recruited into existing groups (new guerrillas are unlikely to just appear and start attacking). Larger groups are also likely to be more successful at recruiting, or simply be easier for new members to find or hear about. Recalling that agents can represent not only people but also equipment, larger groups are again more likely to be able to acquire more. This is exactly the same in the model, where a new agent is added to a group with probability proportional to its size. A conventional army on the other hand tends to have rigidly organized units. These units operate independently and do not tend to just join together upon meeting, nor do they generally fragment unless they take heavy losses [135]. The conventional army is also assumed to have sufficient resources to be able to reinforce its units immediately in the case of losses: This assumption is generally true as armies engaged in ongoing active war have reserves to draw upon. If the army is starting to lose substantial numbers that cannot be replaced to the insurgents, then the conflict is likely to either develop into conventional warfare or end, either of which are not covered by this model. The first cluster selection process also gives each cluster an equal chance of action, which means each group acts as a single unit, and also gives the initiative to the population with more groups; this is almost always the insurgents. The second cluster selection picks a cluster with probability proportional to its size: This means that the first cluster selected is more likely to encounter/interact with larger clusters. This makes sense, as apart from the fact that larger groups occupy more area, they are also easier to detect. In the following, the final EF model variant (EFA+EFR+EFF, with A explicitly an army, B insurgents, and casualties and recruitment occurring) is compared with casualty data for two conflicts using data from CERAC [136]. The model's casualty event distribution from AB interactions is compared, but the distribution would be exactly the same if we used interactions involving a civilian population. In order to minimize inaccuracies and statistical fluctuations in the data and computer simulation, both are plotted as

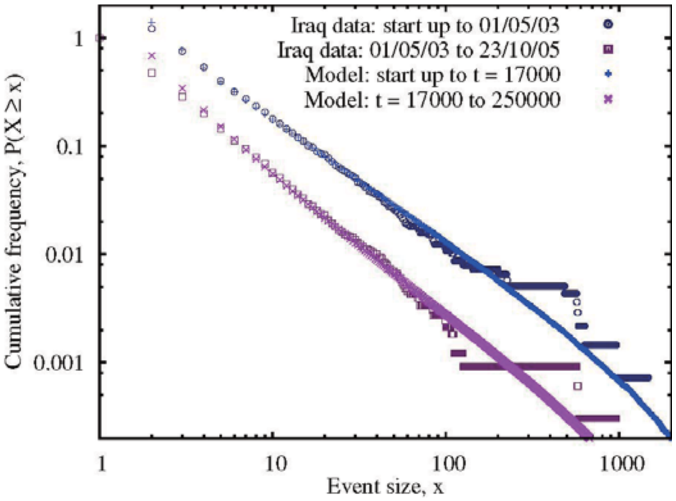


Figure 3.18: Comparison of model with Iraq data at two different time points. The first set of data and model have both been rescaled for clarity.

cumulative distribution functions, $P(X \geq x)$. This is defined as the probability that an event is greater than or equal to a given size, x.

Iraq: 14/03/03 - 23/10/05

The data available is divided into two periods, one from the start of the war to 01/05/03, and the second from then up to 23/10/05. Shortly after the war started there were substantial numbers of insurgents (remnants of the army, Ba'athists etc.) already in place, and subsequently more insurgents have been recruited (foreign fighters, nationalists etc.) [137]. The model was therefore started with a large initial B (insurgent) population, a relatively large A (coalition army) population, and a constant recruitment rate for insurgents. The numbers used are: $N_0 = 4500$, $P_0 = 8500$, $r = 0.45$. The distribution of events (casualty sizes between A and B) during two different time periods are plotted, along side the corresponding data from Iraq, in Fig. 3.18.

As can be seen the correlation in both cases is good for all values except for events

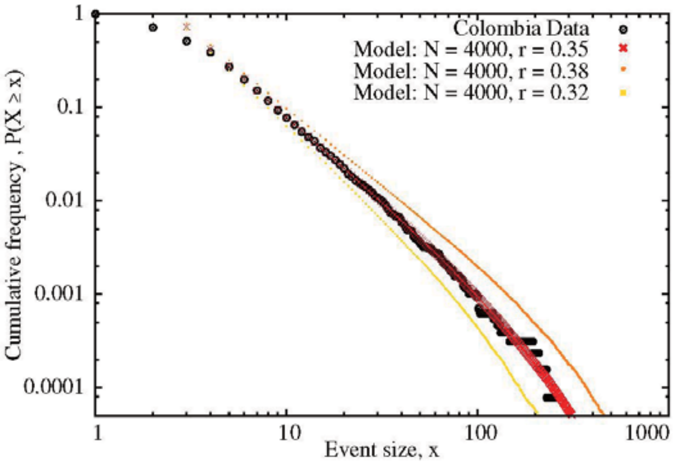


Figure 3.19: Comparison of model with Colombia data

of size 1 and 2. The model has almost reached a steady state by $t = 250000$, and while we cannot know if the data has reached a steady-state with only two time points, it has been suggested elsewhere [113] that α (the slope) has varied smoothly between these two time points and is tending to a constant value. This is identical to the model, further validating it and the choice of parameters.

Colombia: 01/01/88 - 31/12/04

For the case of Colombia the guerrilla war has been ongoing for a long period of time, so a steady state model distribution is appropriate. The initial B population in this case does not matter, the recruitment rate was set at a lower rate than Iraq ($r = 0.35$), as was the army population ($N_0 = 4000$). The result of this distribution in the steady state is shown in Fig. 3.19. The match is again good except at very low s values. The power law coefficient (α) is known to fluctuate with time in this conflict [113], for this reason the effect of altering the recruitment rate in the steady state to two different values is also shown in the figure. The result of this is to vary α as observed.

3.4 Outlook

We hope that the present chapter has given an indication of what statistical physics might have to offer to the daunting challenge of quantifying human conflict. This field is new, in the sense that new insurgent warfare is not within the standard military modeling approach using mass-action partial differential equation approaches (e.g. Lanchester). There may also be other areas in which these models could potentially be useful, e.g. the battles of the immune system fighting invading pathogens. Much remains to be explored in these fields, and many exciting new results undoubtedly await discovery.

CHAPTER 4

Transmission Theory of Dynamical Population

4.1 Overview

The world recently witnessed a baffling variety of global outbreak phenomena: the huge fluctuations across the world's financial markets, driven in part by the rapid global spread of rumors [139]; an unexpected global outbreak of swine flu [140], driven in part by rapid social mixing (e.g. within schools [141]); and even the sudden rise to global fame of an unknown Scottish singer, driven in part by word-of-mouth sharing [142–144]. To understand how such phenomena might arise, consider the following: The number and identity of the people with whom we are each in instantaneous electronic or physical contact – and with whom we can therefore instantaneously exchange information, rumors or viruses – can change slowly or rapidly within any given day, according to the activities which we undertake and hence the instantaneous social groups within which we happen to find ourselves. Even on the shortest plane journeys, for example, passengers find themselves momentarily confined in an enclosed space with complete strangers for up to an hour or more, enabling the exchange of respiratory pathogens. On the blogosphere and on the Web, ephemeral groups form around topics or content and exchange information, opinions and social contacts before flickering out of existence. The transient transnational nature of online discussion groups and chat-rooms, as frequented by financial traders or YouTube users [139, 143, 144] provides a vivid illustration. A full description of such specific transmission processes would likely require rather sophisticated epidemiological models which incorporate system-specific details and considerations (e.g. spatial topology, differential susceptibility). There are indeed many sophisticated epidemiological models already under construction and study in the literature [145–154]. Some of these focus on the well-mixed (i.e. mass-action) limit, some of these focus on the limit of heterogeneous networks [148, 151–154] – and some attempt to move between the two by adding patch-like structure to mass-action models, or dynamical link rewirings to network models.

In this paper, we focus on the less well understood dynamical regime where the group-level dynamics and individual-level transmission processes can evolve on the same timescale, and hence the number and identity of a given individual's contacts can change abruptly at any given moment in time (see, for example, Fig 4.1(a) and (b)). In Sec. 4.2, we introduce and analyze a simple model which mimics the dynamical processes of social group formation/break-up and person-to-person transmission of a virus or information, allowing them to co-exist on comparable timescales. By varying the probabilities of group coalescence (ν_{coal}) and fragmentation (ν_{frag}) relative to the standard SIR (Susceptible→Infected→Recovered) probabilities [145, 150] for person-to-person transmission (p) and individual recovery (q), the entire range of relative timescales can be easily explored – from a very slowly changing social network structure (i.e. essentially a static network with infrequent rewirings) through to a rapidly changing social network structure (i.e. essentially a well-mixed population). Most importantly, this includes the complicated intermediate regime where both processes co-exist on the same timescale. Figure 4.1 illustrates this intermediate regime, while Fig. 4.2 shows how an associated infection profile $I(t)$ is qualitatively very different from the two limiting cases of the static (or quasi-static) network, and the well-mixed population. Instead, the interplay of the group dynamics and individual-level transmission generates epidemic profiles which exhibit a rich structure (e.g. multiple resurgences and abnormal decay times, see Fig. 4.2). It turns out that such profiles are strikingly similar to real-world outbreaks across the social, financial and biological domains (see Fig. 4.3, and Sec. 4.3). While it is conceivable that infection profiles similar to Fig. 4.3 can also be obtained using alternative, more sophisticated epidemiological models (e.g. by adding spatial topology or differential susceptibility), such models will typically have more parameters and be more system-specific. By contrast, our model only has four stochastic parameters for the probabilities (and hence timescales) of the individual level transmission and group dynamics, i.e. p, q for the SIR process, and ν_{coal} and ν_{frag} which describe the probability of social groups coalescing or fragmenting. We find it intriguing that the qualitative shapes of such a wide range of empirical profiles (Fig. 4.3) can be reproduced simply by varying these relative timescales. While we cannot prove that the empirically observed profiles in Fig. 4.3 are indeed generated by such a simple model as ours, it seems that more complex models are not required in order to reproduce their main features. In Sec. 4.4, we offer an approximate analytic analysis of the properties of our model. Although a detailed theoretical description of the infection profiles $I(t)$ remains an open future challenge, we find that the overall properties can be captured by making a mean-field approximation of the behavior of connected pairs within the population. In Sec. 4.5, we comment on how our results also suggest a minimally-invasive dynamical method for controlling outbreaks (see Fig. 4.5). Section 4.6 extends the discussion to other types of viral transmission models which are widely studied in the literature (e.g. SIS). Section 4.7 provides a summary and outlook.

4.2 Our model

There are many plausible rules for generating human group dynamics [155]. The grouping process that we choose involves simple cluster dynamics at each timestep. In terms of the viral model, our focus will be on individual-level transmission via SIR because of its dual relevance to real virus transmission and to the spread of rumors and information in a social system. Here SIR means the viral process Susceptible → Infected → Recovered, the person-to-person transmission probability at each timestep is p and the individual recovery probability at each timestep is q, as described in the previous section. Provided that the choice of social dynamics permits similar intrinsic self-amplification and suppression processes by sporadically injecting infected individuals into susceptible groups (see Fig. 4.1(b) and bottom of Fig. 4.3), the resulting epidemic profiles from alternative choices of cluster mechanism should exhibit similar characteristics – in particular, multiple resurgences and abnormal decay times (see Figs. 4.2 and Fig. 4.3). In Sec. 4.4 we extend this SIR focus by presenting quantitative results for other commonly studied viral processes (e.g. SIS).

Our choice of cluster mechanism features the coalescence and fragmentation of groups as described below, and illustrated in Fig. 4.1(b). There is a huge volume of work in the mathematics, physics and chemistry literature on cluster models within a many-body population of interacting particles [156]. For modern social systems, one is typically interested in mechanisms which mimic the long-range interactions that people can have (either through transport in the case of transmission of viruses, or

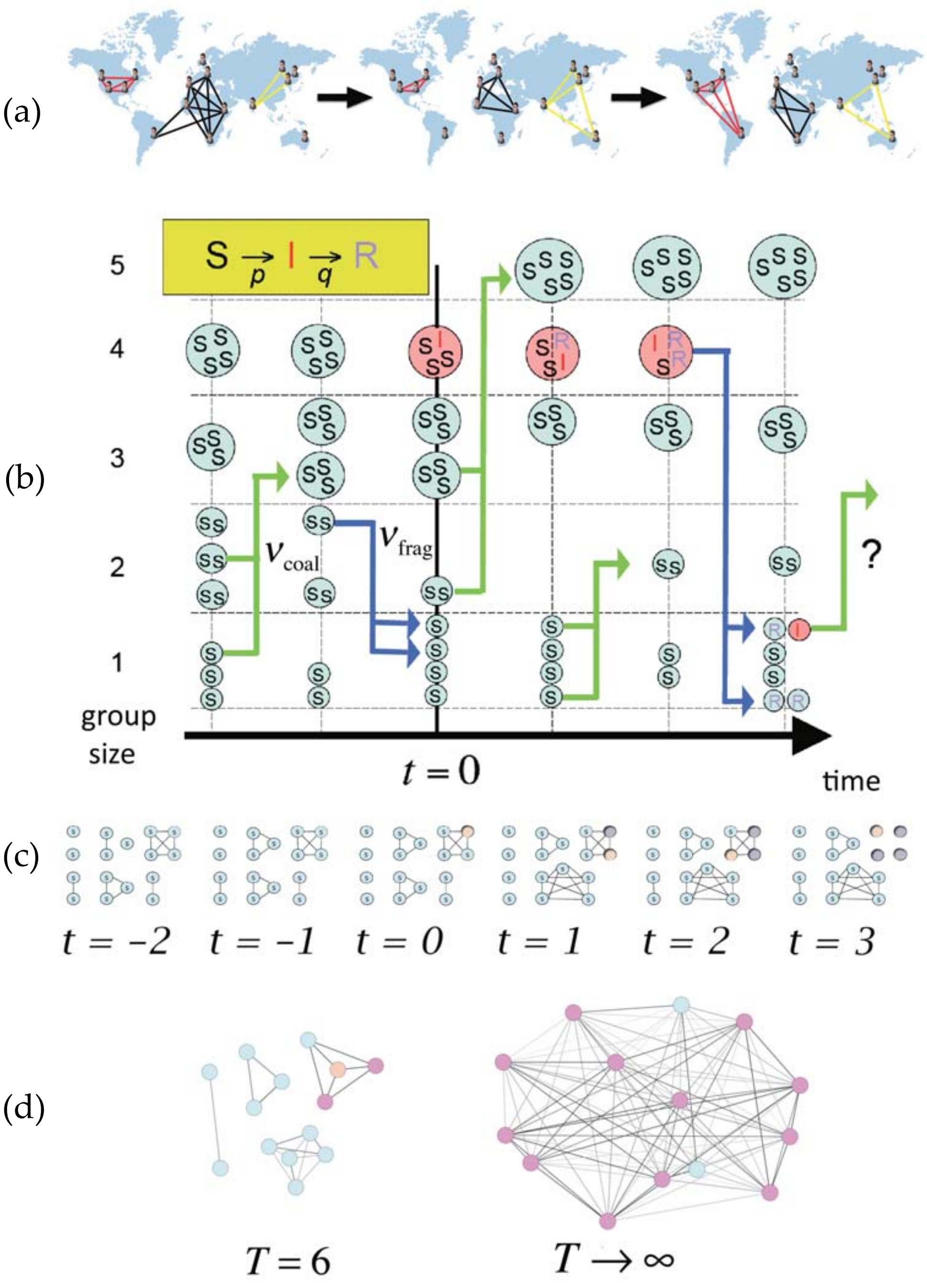


Figure 4.1: (Color online) a: Schematic of dynamical grouping of traders or YouTube users on the Internet b: Schematic of our model, featuring spreading in the presence of dynamical grouping via coalescence and fragmentation. Vertical axis shows number of groups of a given size at time t. c: Instantaneous network from Fig. 1b at each timestep. d: Weighted network obtained by aggregating links over time-window T.

communications in the case of transmission of a rumor or information). We choose the rate of coalescence of two groups of size n_1 and n_2 respectively to be proportional to the combinatorial number of pairwise encounters between individuals, one from each group, i.e., the rate of coalescence is equal to $\nu_{\text{coal}} \cdot (n_1/N) \times (n_2/N)$, where $\nu_{\text{coal}} \leq 1$ is a coalescence probability. Similarly, a given group of n individuals may break up (i.e. completely fragment) with a total rate equal to $\nu_{\text{frag}} \cdot (n/N)$, where $\nu_{\text{frag}} \leq 1$ is a fragmentation probability with $\nu_{\text{frag}} + \nu_{\text{coal}} \leq 1$, reflecting the increasing fragility of large groups (i.e. standard size effect). The implementation of this social dynamics is essentially the same as Ref. [28], with the generalization that the coalescence and fragmentation probabilities are general (i.e. $\nu_{\text{frag}} + \nu_{\text{coal}} \leq 1$ but ν_{frag} is otherwise unrelated to ν_{coal}, in contrast to Ref. [28]). This specific cluster process has real-world relevance for several reasons. First, it embodies the rare but dramatic changes of contact networks that can occur, as mentioned in the introduction. Second, it produces a distribution of group sizes which is power-law with exponent 2.5 for $\nu_{\text{coal}} \gg \nu_{\text{frag}}$ when time-averaged, as shown in the related model for financial markets [28, 55]. This theoretical model is therefore consistent with the observation of Gabaix et al. [96] who found that the distribution of transaction sizes follows a power-law with slope near 2.5 for each of the three major stock exchanges in New York, Paris and London. Third, this power-law exponent 2.5 is also consistent with the distribution of group sizes inferred for terrorists and insurgent groups based on an analysis of casualty figures [113]. Fourth, the model is structurally robust in that the group dynamic rules can be generalized to different positive power exponents $\alpha \neq 1, \beta \neq 1, \gamma \neq 1$, with coalescence and fragmentation rates given by $\nu_{\text{coal}} \cdot n_1^{\alpha} \times n_2^{\beta}$ and $\nu_{\text{frag}} \cdot n^{\gamma}$, respectively, without losing the main qualitative features of the dynamics of the number $I(t)$ of infected individuals.

Individual connectivities within our model may change significantly on the same timescale as the SIR process, thereby mimicking individuals participating in YouTube viewing, financial systems, and schools, who sometimes exhibit rapid moves among peer groups either online or in real space, while simultaneously picking up and spreading rumors or pathogens. A discrete illustration of our model over six time steps is shown in Fig. 4.1b and c. Fig. 4.1d contrasts the short-time group structure between individuals with the long-term linkage between them: as time increases without bound, by ergodicity, all individuals will have eventually been part of some common group. While the latter long-time network structure is the one usually emphasized in models of epidemic processes on complex networks, the short-time limited linkage is essential to understand the competition between individual isolation (which tends to stop an epidemic) and group coalescence which amplifies its spreading. In short, this model provides us with a simple framework within which to explore and quantify the interaction of these group dynamical processes and conventional SIR dynamics.

In the numerical implementation of the model, we run the above coalescence-fragmentation dynamics until the time-averaged distribution of group sizes has become stationary. Then, at some instant taken at the origin of time $t = 0$, one group is selected, and an arbitrary individual in this group becomes infected, and hence the infection profile unfolds according to the SIR process within each group, with all the groups undergoing at the same time the coalescence-fragmentation dynamics according to the two rates ν_{coal} and ν_{frag}. A typical simulation is shown in Figure 4.2, and is compared with the popular approach that models spreading on static networks: (i) an instantaneous network ($T = 1$) and (ii) a global network formed by time-aggregating instantaneous contacts over long times ($T \rightarrow \infty$). SIR spreading dynamics on fixed networks obtained at different intermediate T gives curves that lie in the shaded area

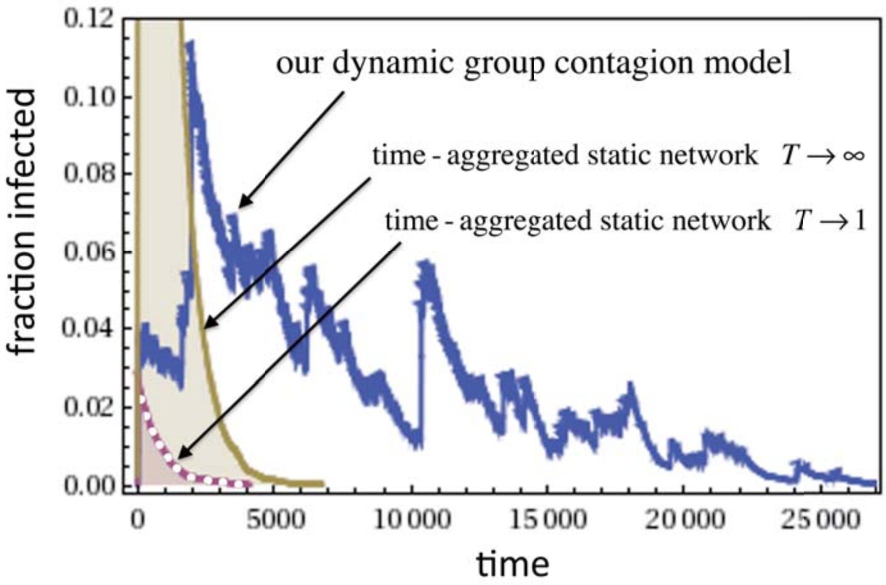


Figure 4.2: (Color online) Theoretical profile $I(t)$. Thick (blue) curve shows our dynamical group contagion model, with $\nu_{\text{frag}} = 0.05$, $\nu_{\text{coal}} = 0.95$, $p = 0.001$ and $q = 0.001$. Using same p and q values, dotted (purple) curve corresponds to stochastic SIR model on a static network with $T \to 1$, i.e. the $t = 0$ network in Fig. 1. Thin solid (green) curve corresponds to stochastic SIR on a $T \to \infty$ network.

of Figure 4.2. Our model can generate not only this type of dynamics, but also qualitatively new regimes that arise from adjusting the coalescence-fragmentation rates: the large fluctuations, resurgences, and abnormally long decay time which are observed in our model (as illustrated in Figure 4.2 (thick curve)) are generated by self-amplification and suppression processes due to the coalescence-fragmentation group dynamics at all group-size scales.

4.3 Theoretical and empirical results

Figure 4.3 provides a comparison between two of the empirical profiles that we have collected from each of three distinct real-world domains, and the theoretical infection profiles $I(t)$ produced by our model. By providing two empirical profiles for each real-world system, our aim is to give confidence that the empirical features observed

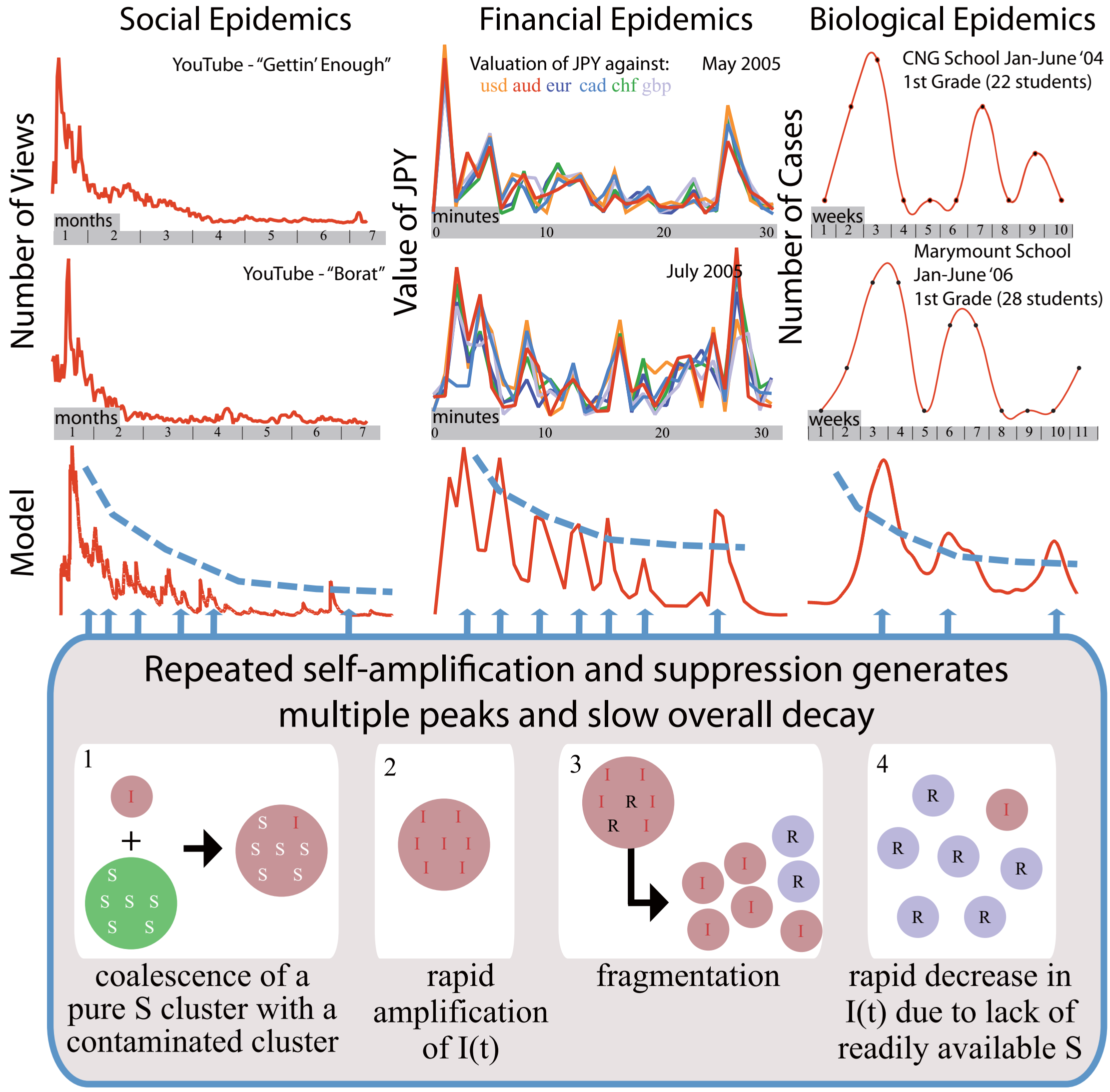


Figure 4.3: (Color online) Top two rows: Empirical activity profile $I(t)$ in three distinct real-world systems. Third row: Results from our model. Dashed line is a guide to the eye. Left: YouTube download activity. Middle: Currency trading activity (i.e. absolute value of price-change, hence the excess demand to buy or sell at each timestep). Different shades correspond to different currencies. See main text. Right: Fraction of children with colds within a school. Lower panel: Simple example of the repeated self-amplification and suppression processes which spontaneously arise within our the model. When replicated at all scales of group size, these processes generate a unified quantitative description of the empirical $I(t)$ profiles.

are not simply the result of some irreproducible external noise. We could generate a number of metrics for comparing the empirical and theoretical profiles (e.g. number of peaks, time interval between peaks, peak-to-trough ratio) as is our intention for future studies when additional data is available – but for now, the message that we wish to convey is a visual one, i.e. that our simple model manages to capture the main qualitative features for each of these three distinct empirical systems.

The sociological example (left column of Fig. 4.3) shows downloads for two similar YouTube clips [143, 144]. Such downloads are typically driven by YouTube users absorbing and spreading opinions as they share information in their social groups [144]. The two downloads appealed to similar age-groups, and were measured close together in time, implying that a similar pool of users accessed them both, in line with our model's assumptions. Our model accounts well for the long memory and aftershock-like decay. The top-left panel corresponds to "Gettin' Enough" from `http://www.youtube.com/watch?v=AiXxMrkeklg`. This video was uploaded to the site on 8 November, 2006 at 13:33:12 GMT. The first record of a download is 9 November, 2006 at 17:21:35 GMT, with a view count of 5708. The last record is 24 May, 2007 at 21:34:30 GMT, with a view count of 257759. The video's length is 225 seconds. The middle-left panel shows downloads of music video "Borat" from `http://www.youtube.com/watch?v=b1xXERFt_Zg`. The video was uploaded to the site on 3 November, 2006 at 11:04:15 GMT. The first record of a download was 7 November, 2006 at 10:35:57 GMT with a view count of 20745. The last record is 24 May, 2007 at 22:29:25 GMT, with a view count of 254918. The video's length is 154 seconds. These two music downloads are similar in terms of appeal, age-group, total number of downloads, and lack of any public/global announcement, news or advertisement, and are hence consistent with spreading through contagion. The bottom-left panel shows our model's output with $\nu_{\mathrm{frag}} = 0.05$, $\nu_{\mathrm{coal}} = 0.81$, $p = 0.001$, $q = 0.001$.

The financial example (middle column of Fig. 4.3) shows foreign exchange movements as a result of a specific rumor spreading among traders concerning revaluation of the Chinese Yuan currency. This same rumor circulated twice in the space of a few months. The fact that the currency pairs follow a similar dynamical pattern in each case, suggests that the same underlying group dynamics developed, in line with our model. Note that this financial epidemic is characterized by the largest coalescence rate ν_{coal} and much larger infectivity parameter p among the three examples, reflecting the efficiency of the information cascade among currency traders. The top-centre panel corresponds to the CNY (Chinese currency) revaluation rumor as detected from trader chat-rooms by HSBC bank (courtesy of S. Williams). Specifically, we show the absolute returns on the timescale of 1-minute intervals for the JPY (Japanese currency) exchange rates from 08:22 to 08:53 GMT, on 11 May 2005. The middle-centre panel corresponds to the CNY actual revaluation. Absolute returns on the timescale of 1-minute intervals are shown for JPY (Japanese currency) exchange rates from 11:03 to 11:34 GMT, on 21 July 2005. Since the CNY was not one of the directly traded currencies, its effects on the JPY-X rates (where X is another currency) are indirect in both cases, suggesting influence through contagion of the rumor/information. There was no public announcement or global news to trigger this activity, which also supports spreading through contagion. The bottom-centre panel shows our model output with $\nu_{\mathrm{frag}} = 0.05$, $\nu_{\mathrm{coal}} = 0.95$, $p = 0.009$, $q = 0.002$. Since data are on 1-minute scale, but prices can change on 1-second scale, we show an averaged output by providing value at regular equispaced intervals, mimicking 1-minute.

The biological example (right column of Fig. 4.3) shows incidences of a cold among

1st grade students in two schools in Bogota, Colombia. These data come from an ongoing monitoring experiment carried out by members of our team, in conjunction with the Universidad de Los Andes, Bogota, based on weekly surveys of all students and staff at Colegio Nueva Granada (CNG) and Marymount School. The schools' locations at the top of the Andes guarantee that seasonal temperature variations are minimal. In addition, the student population of each approximates to a closed system due to local issues of security and social segregation. Given the unchanging climatic conditions of this part of Colombia, we argue that the microbes responsible for colds remain present in the environment of children, hence successive bursts can be expected to follow similar dynamics. Within our approach, the school cold dynamics are found to be best described by the lowest fragmentation rate ν_{frag} and highest recovery parameter q, mirroring the more rigid structure of inter-children contacts and the crucial role of multiple recuperations. The bottom-right panel shows our model output with $\nu_{\text{frag}} = 0.001$, $\nu_{\text{coal}} = 0.5$, $p = 0.001$, $q = 0.004$. The model output is smoothed, to mimic fact that data is recorded on the 1-week time interval.

The sociological example corresponds to the regime of ν_{coal} values where the group dynamics dominate, hence the strong observed clustering effect in $I(t)$. Large values of ν_{coal} do indeed make sense for social online systems, given the ease and rapidity with which people can now be exposed to new information and rumors through virtual communities. The financial example corresponds to the regime of large transmission probability per timestep p, thereby promoting both the spread and survival of the information "virus" and hence leading to persistently noisy (but remarkably reproducible) fluctuations. The biological example seems to lie in between, since it has more peaks as compared with sociological example and larger fluctuations as compared with the financial example. We stress that more complex explanations of each empirical profile in Fig. 4.3 are undoubtedly possible – and may even be deemed as more realistic by specialists within each field. However, our purpose here is to focus on a minimal description of common dynamical features, and to highlight the fact that a common minimal description is indeed possible. Although we have only presented three examples of our model's output, each of which has a quite distinct visual form, it turns out that this subset of three broadly encompasses the full range of $I(t)$ profile behaviors observed across the entire parameter space of our model. One might speculate whether other empirical examples chosen from social, financial and biological domains, might also fall crudely into one of these three classes of infection profile behavior shown in Fig. 4.3. A casual observation of similar types of empirical data available on the Web suggests that this might indeed be the case [144, 158].

4.4 Approximate analytic analysis

4.4.1 Limiting case relationship with mass-action epidemiological theory

A full analytic description of the $I(t)$ profiles generated by our model represents a fascinating open challenge for the community. However some important features can be captured straight away, by suitably generalizing existing epidemiological machinery. A key quantity is the probability $\eta_{\mathrm{SI}}(t)$ that a particular link instantaneously exists *and* that it connects a susceptible and an infected. Thus, out of a potential totality of $N(N-1)/2$ links among N individuals, only $\eta_{\mathrm{SI}}(t) \cdot N(N-1)/2$ are typically present. This provides an accurate equation for the number of susceptibles $S(t)$ in a given epidemic sequence: $\dot{S}(t) = -p.\eta_{\mathrm{SI}}(t).N(N-1)/2$, with p is the infectivity parameter quantifying the probability that a particular S is infected by a particu-

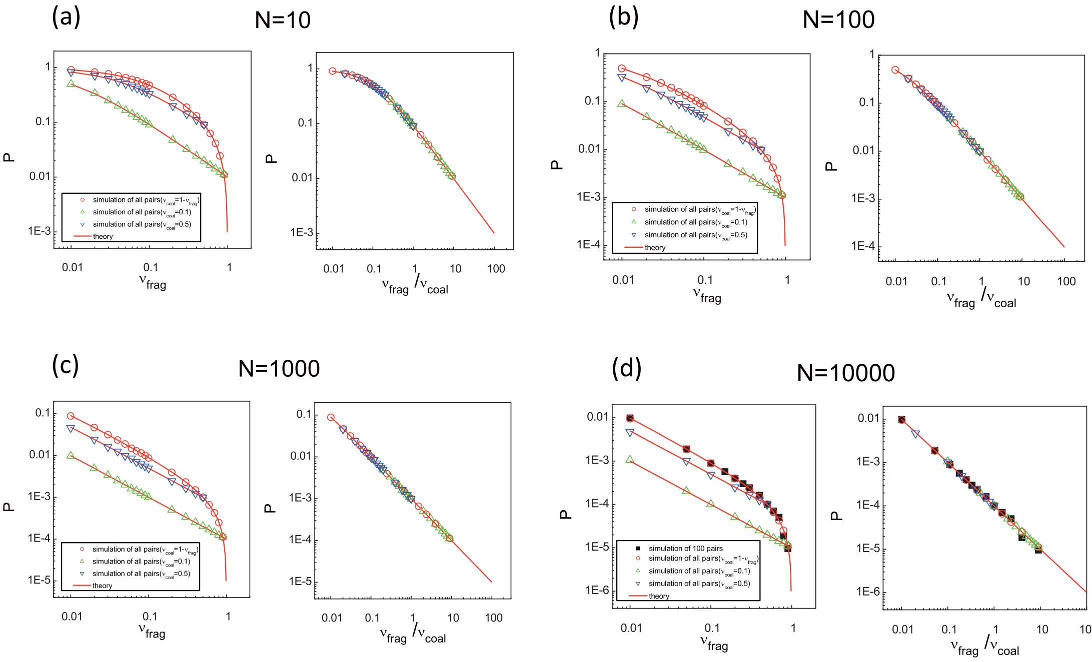


Figure 4.4: (Color online) Demonstration of accuracy of Eq. (4.1) across a range of parameter space. Solid line is theoretical expression from Eq. (1), while discrete points are numerical simulation results. Population size (a) $N = 10$, (b) $N = 100$, (c) $N = 1000$ and (d) $N = 10,000$. Left panel for each N value: P value obtained by tracking all pairs over time for the illustrative case $(\nu_{\text{coal}} + \nu_{\text{frag}}) = 1$, and also for specific values $\nu_{\text{coal}} = 0.1$ and 0.5. Left panel of (d) additionally shows result of tracking just 100 pairs over time, demonstrating that the same results are obtained as long as the time window is sufficiently long. Right panel for each N value: P as a function of $\nu_{\text{frag}}/\nu_{\text{coal}}$ on a log-log scale. Data corresponding to different systems fall onto the same curve as predicted by Eq. (4.1).

lar I to which it is linked. We rewrite this in the conventional mass-action form $\dot{S}(t) = -pP_{\mathrm{SI}}(t)S(t)I(t)$, where $P_{\mathrm{SI}}(t)$ is equivalent to $\eta_{\mathrm{SI}}(t)N(N-1)/[2S(t)I(t)]$ and hence now incorporates the complex dynamics which are so hard to capture analytically. We now approximate $P_{\mathrm{SI}}(t)$ by a constant term P, which is the time-averaged probability that any two arbitrarily chosen nodes belong to the same cluster independent of SI-infection status. This approximation throws out the dynamical details of $I(t)$, but can provide useful insights, as shown below on the non-spreading to spreading transition, provided that several coalescence-fragmentation processes occur over the timescale of the entire outbreak. In order to develop a more detailed analytic analysis, we start with the identification of P as the average value of $P_{i,j}$, where $P_{i,j}$ means the probability that specific nodes i and j are connected (i.e. i and j are in the same cluster). Our numerical simulations enable us to track individual pairs and hence deduce numerical values for P as a function of ν_{coal}, ν_{frag} and N.

Figure 4.4 presents P values taken from the numerical simulation, which suggest that P has a remarkably simple functional form. Results are shown for populations of size $N = 10$ (Fig. 4.4(a)), $N = 100$ (Fig. 4.4(b)), $N = 1,000$ (Fig. 4.4(c)) and $N = 10,000$ (Fig. 4.4(d)), and for different sets of parameters ν_{coal} and ν_{frag}. There are several ways that one can obtain the quantity P numerically. One could follow a certain pair of nodes over a very long simulation run of the model, counting how many timesteps these two nodes happen to be in the same cluster. Alternatively, to reduce run times, one may follow a certain number of pairs (e.g. 50 or 100) over a particular time-window. As confirmed by the left-hand panel of Fig. 4.4(d), the results do not depend on the precise method used. The numerical results in Fig. 4.4 suggest that the variation in P can be well described by the following empirical relation:

$$P = \frac{\nu_{\text{coal}}}{\nu_{\text{coal}} + N\nu_{\text{frag}}} = \frac{1}{1 + N\frac{\nu_{\text{frag}}}{\nu_{\text{coal}}}} \quad . \tag{4.1}$$

This expression contains an explicit dependence on N, the number of nodes in the system, and a dependence on the ratio $\nu_{\text{frag}}/\nu_{\text{coal}}$. For each N value in Fig. 4.4, the right-hand panel shows that different simulations with different parameters $(\nu_{\text{coal}}, \nu_{\text{frag}})$, yield dependences of P as a function of $\nu_{\text{frag}}/\nu_{\text{coal}}$ which all collapse on the same curve. For $N\nu_{\text{frag}} \gg \nu_{\text{coal}}$, Eq. 4.1 reduces to the approximate form

$$P = \frac{\nu_{\text{coal}}}{N\nu_{\text{frag}}} \quad . \tag{4.2}$$

4.4.2 Analytic derivation of P using Master Equation

Having demonstrated the quantitative accuracy of Eq. 4.1 using quantitative results from the model simulation, we now turn to the task of trying to understand this result analytically. Starting with the master equation approach, the dynamics of $P_{i,j}$ follow:

$$\frac{dP_{i,j}}{dt} = -P_{i,j}\frac{1}{N}\sum_{k\in\{..i.\}} P_{k,i}\nu_{\text{frag}} + (1 - P_{i,j})\frac{1}{N}\sum_{m\in\{..i.\}} P_{m,i}\frac{1}{N}\sum_{n\notin\{..i.\}} P_{n,j}\nu_{\text{coal}} \quad . \tag{4.3}$$

As indicated, the sums are restricted to either the nodes within the cluster containing node i, or to the nodes outside that cluster. We note that the first term on the right-hand side can alternatively be written in a completely symmetric way in terms of i and j, simply by splitting the term into two equal pieces (one in i and one in j) and adding a factor of one half to avoid double counting. Two particular situations will significantly affect the value of $P_{i,j}$: one situation is where the nodes i and j are together but are going to break-up at the next timestep. The second situation is where the node i and j are not currently together, but are going to join together at the next timestep. The two terms on the right-hand side of Eq. 4.3 correspond to these two cases. The first term on the right-hand side of Eq. 4.3 describes the case where nodes i and j are together ($P_{i,j}$) and then one node k is picked ($\frac{1}{N}$) which is in the same cluster as i and j ($P_{k,i}$) and so the cluster fragments (ν_{frag}). The second term of Eq. 4.3 describes the case where nodes i and j are not together ($1 - P_{i,j}$), and then one node m is picked ($\frac{1}{N}$) which is in the same cluster as i ($P_{m,i}$), as well as one node n being picked ($\frac{1}{N}$) in the same cluster as j ($P_{n,j}$). They then coalesce (ν_{coal}). We stress that the third summation $\sum_{n \notin \{..i.\}}$ sums over all the nodes outside the cluster containing node i, and hence $\sum_{n \notin \{..i.\}} P_{n,j} = 1$. In the steady-state, we can write $P = P_{i,j}$ for all i and j. The remaining two summations over nodes within the cluster containing node i, then become equal to the average cluster size multiplied by P. Setting the left-hand side of Eq. 4.3 equal to zero and solving for P, yields

$$P = \frac{\nu_{\text{coal}}}{\nu_{\text{coal}} + N\nu_{\text{frag}}} = \frac{1}{1 + N\frac{\nu_{\text{frag}}}{\nu_{\text{coal}}}}$$

which is exactly Eq. 4.1.

4.4.3 Analytic derivation of P using coupled cluster equations

The limiting case Eq. 4.2 can also be obtained in the limit $N\nu_{\text{frag}} \gg \nu_{\text{coal}}$, by considering the master equation for the number n_s of clusters with size s in the model:

$$\frac{\partial n_s}{\partial t} = -\frac{\nu_{\text{frag}} s n_s}{N} + \frac{\nu_{\text{coal}}}{N^2} \sum_{s'=1}^{s-1} s' n_{s'} (s - s') n_{s-s'} - \frac{2\nu_{\text{coal}} s n_s}{N^2} \sum_{s'=1}^{\infty} s' n_{s'} \qquad (4.4)$$

for $s \geq 2$, with a similar but truncated form for $s = 1$:

$$\frac{\partial n_1}{\partial t} = \frac{\nu_{\text{frag}}}{N} \sum_{s'=2}^{\infty} (s')^2 n_{s'} - \frac{2\nu_{\text{coal}} n_1}{N^2} \sum_{s'=1}^{\infty} s' n_{s'} \quad . \qquad (4.5)$$

For a steady-state cluster distribution, we have

$$sn_s = \frac{\nu_{\text{coal}}}{2\nu_{\text{coal}} + \nu_{\text{frag}}} \frac{1}{N} \sum_{s'=1}^{s-1} s' n_{s'} (s - s') n_{s-s'} \tag{4.6}$$

for $s \geq 2$, while for $s = 1$ we have

$$n_1 = \frac{\nu_{\text{frag}}}{2\nu_{\text{coal}}} \sum_{s'=2}^{\infty} (s')^2 n_{s'} \quad . \tag{4.7}$$

Therefore on average, we obtain:

$$P = \sum_{s=2}^{\infty} \frac{sn_s}{N} \frac{s-1}{N} = \frac{1}{N^2} \sum_{s=2}^{\infty} (s^2 n_s - sn_s) = \frac{1}{N^2} \frac{2\nu_{\text{coal}}}{\nu_{\text{frag}}} n_1 - \frac{N - n_1}{N^2} \tag{4.8}$$

where the only unknown quantity is n_1. We now define a generating function:

$$G[y] = \sum_{k=0}^{\infty} k n_k y^k = n_1 y + \sum_{k=2}^{\infty} k n_k y^k = n_1 y + g[y] \quad . \tag{4.9}$$

Taking Eq. 4.6 multiplied by y^s and then summing from $s = 2$ to ∞, yields

$$g[y] = \frac{\nu_{\text{coal}}}{2\nu_{\text{coal}} + \nu_{\text{frag}}} \frac{1}{N} G[y]^2 \tag{4.10}$$

i.e.

$$g[y]^2 - (\frac{2\nu_{\text{coal}} + \nu_{\text{frag}}}{\nu_{\text{coal}}} N - 2n_1 y) g[y] + n_1^2 y^2 = 0 \tag{4.11}$$

where $g[y] = \sum_{s=2}^{\infty} sn_s y^s$ and $g[1] = N - n_1$. Solving this quadratic equation Eq. 11 gives

$$n_1 = \frac{\nu_{\text{frag}} + \nu_{\text{coal}}}{\nu_{\text{frag}} + 2\nu_{\text{coal}}} N \ . \tag{4.12}$$

Substituting into Eq. 4.8, we obtain Eq. 4.2.

4.5 Dynamical control of outbreaks

We now use our numerical model and the approximate analytic analysis of Sec. 4.4, in order to address the following highly topical question: *Will there be epidemic*

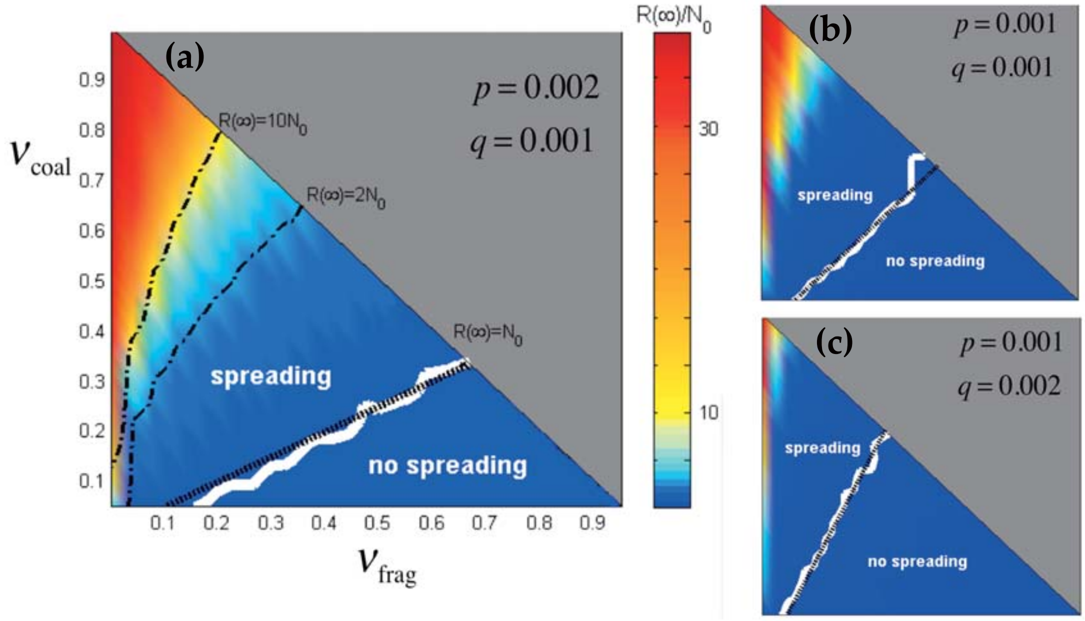


Figure 4.5: (Color online) Phase diagrams show theoretically obtained transition (i.e. $\frac{p\nu_{\rm coal}}{q\nu_{\rm frag}} = 1$, black dashed line) and the numerical result (white solid line) separating regimes of spreading (i.e. overall number of infecteds exceeds initial group size, hence $R(\infty) > N_0$) and no-spreading (i.e. $R(\infty) < N_0$). Population reacts to news of the initial infection at $t = 0$ by changing its dynamical grouping from $\nu_{\rm frag} = 0.001$ and $\nu_{\rm coal} = 0.99$, to the new values shown on the axes. Shading shows the population (in units of N_0) who become infected, and hence recovered, over the lifetime of the outbreak. Solid triangular shaded region is unphysical since $\nu_{\rm frag} + \nu_{\rm coal} > 1$.

spreading in a population in which it is publically known that N_0 persons have been infected with a given pathogen or rumor, but where the precise identity of infected persons cannot be disclosed? At $t = 0$, $N_0 \ll N$ individuals of the instantaneously largest group are infected and news of an infection is announced without disclosing the infected's identities. The population reacts by adjusting its group dynamics, i.e. it adopts a new $\nu_{\rm coal}$ and $\nu_{\rm frag}$. Although many further features could be added to mimic the population's subsequent adjustment to knowledge of an outbreak, we focus here on a simple case in order to better understand the effect of the initial reaction.

Numerical results are presented in Fig. 4.5, together with analytic curves for the transition threshold, as a function of the new ν_{coal} and ν_{frag}. Our analytic analysis exploits the generalized SIR equations developed in Sec. 4.4, and builds upon the theoretical framework discussed in detail in Ref. [150]. The number of susceptibles in the long-time limit $S(\infty)$ with $N \gg 1$ is then given by the solution $\overline{z}$ to the following generalized form, $z = \exp[-\kappa(1-z)]$ where $z \equiv S(\infty)/N$ and $\kappa \equiv p\nu_{\text{coal}}/q\nu_{\text{frag}}$ where κ plays the role of the basic reproductive rate. For $\kappa \leq 1$, the only solution is $\overline{z} = 1$, corresponding to a vanishingly small fraction of infected individuals (i.e. total number of infected $R(\infty)$ does not exceed $N_0 \ll N$). This solution bifurcates at $\kappa = 1$ into the following stable solution $\overline{z} = -(1/\kappa) \cdot W\left(-\kappa \cdot e^{-\kappa}\right)$ valid for $\kappa > 1$, where $W(z)$ is the Lambert function. For $\kappa > 1.5$, $\overline{z}$ is very well-approximated by $\overline{z} \approx e^{-\kappa}/\kappa$. This shows a rather abrupt transition from non-spreading epidemics for $\kappa < 1$ to global infection of a finite fraction of the population for $\kappa > 1$. The form of the epidemic control parameter $\kappa \equiv p\nu_{\text{coal}}/q\nu_{\text{frag}}$ exemplifies that infectivity and coalescence play together against recovery and fragmentation in controlling the propagation of the epidemics: Infectivity and coalescence promote the infection propagation, while recovery and fragmentation hinders its spread.

Not only is our theory for the spreading threshold (dashed black line in Fig. 4.5) in good agreement with the numerical results (white solid line), its simple analytic form suggests an epidemic control scheme based on manipulation of the group coalescence and fragmentation timescales (i.e. ν_{coal}^{-1} and ν_{frag}^{-1}). An imminent epidemic can be *suppressed* (i.e. $R(\infty) < N_0$) by increasing the timescale for group coalescence with respect to the timescale for group fragmentation (i.e. decrease ν_{coal} with respect to ν_{frag}), but it will get *amplified* if we decrease the coalescence timescale with respect to the fragmentation timescale (i.e. increase ν_{coal} with respect to ν_{frag}). Not only would such modest intervention allow the overall system to continue functioning, it does *not* require knowledge of the infected's identities. There is also no assumption that the N_0 members of the group which carries the initial infected case at $t = 0$, remain in that group. In the school setting, schedules could be adjusted to slow down or speed up classroom use and recess, without the need for disruptive school closures [141] or the need to test, label or isolate infected children. In particular, the coalescence rate in a school could be reduced by simply staggering the lunchtimes for separate classes, as opposed to isolating them entirely. Having individual classes in the cafeteria at different times does not count as a coalescence event, while a common lunchtime period would. Hence the coalescence rate can been reduced without removing any of the daily routine activity. In financial markets, if one wanted to prevent highly noisy fluctuations of the type observed in Fig. 4.3, similar control might be achieved by basing the joining and leaving rules of the online chatrooms frequented by financial traders, on present occupancy. In viral marketing, the attractiveness of the message or product quantified by the infectivity p can be completely subjugated by suitable management of the group dynamics (ν_{coal} versus ν_{frag}), as firms using e-commerce and e-advertisement are now realizing [157]. These findings are potentially applicable to many other scenarios, given that many real-world activity/infection curves resemble those in Figure 4.2.

4.6 Results for general infection models

Having focused exclusively on the SIR viral process, we now compare the infection profiles generated using other commonly studied viral processes. To ease comparison, the population maintains the *same* coalescence-fragmentation group dynamics as featured throughout this paper. We find that a similarly large range of $I(t)$ profiles arise

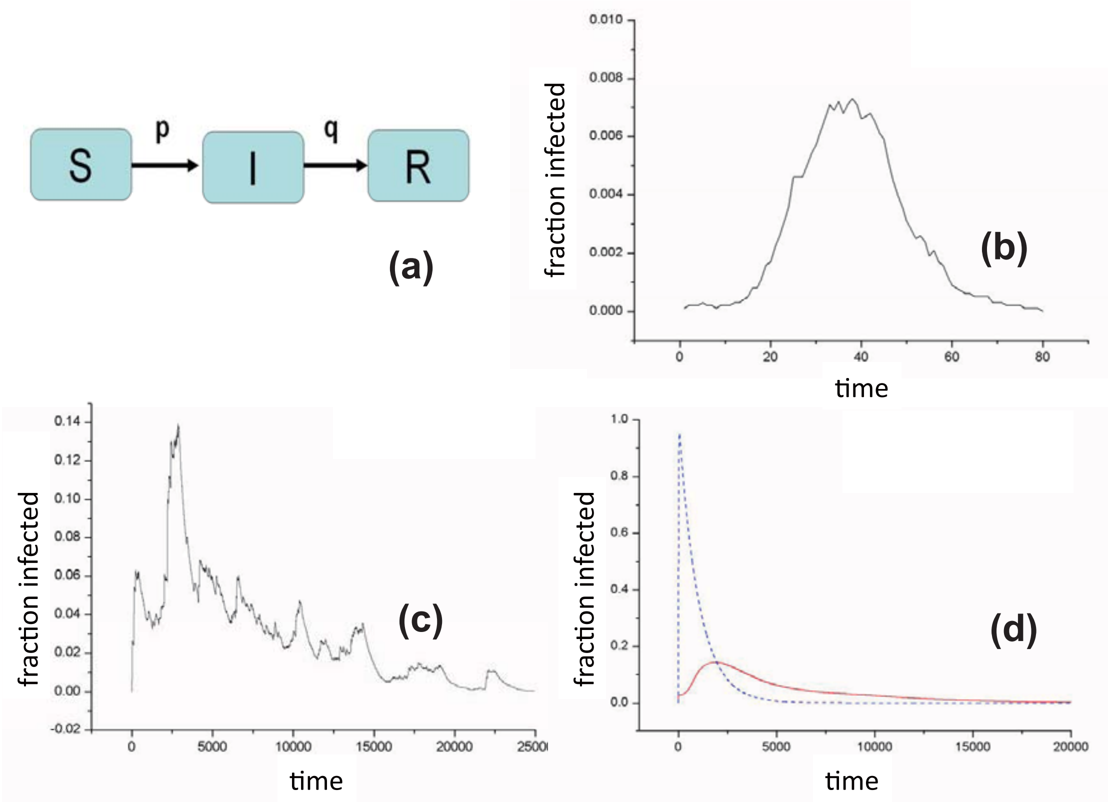


Figure 4.6: (Color online) SIR process in presence of the coalescence-fragmentation group dynamics. (a) Schematic SIR process. (b) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.001$, $\nu_{\text{coal}} = 0.99$, $p = 0.001$ and $q = 0.1$. (c) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.01$, $\nu_{\text{coal}} = 0.9$, $p = 0.001$ and $q = 0.001$. (d) Comparison between run-averaged $I(t)$ (solid curve, red) using the same parameters as (c), and $I(t)$ for a weighted network (blue dotted curve) in which all nodes are connected with strength P.

for these other viral processes as for SIR, which is understandable given that they are all being driven by the same background group dynamics. In addition to p and q, some of the viral processes will use μ which is a birth and death probability rate, γ which is a probability rate of transition from I to S, and ω is a probability rate of transition from R to S.

As suggested in the earlier parts of this paper, we might expect the numerical simulations to exhibit three main classes of behavior for each type of viral process: (1) group dynamics is much slower than epidemic spreading, hence the virus tends to exist only within the initial group; (2) group dynamics is comparable with the epidemic spreading, hence grouping plays a significant role in suppressing or amplifying the spreading leading to generic spiky behavior in the infection profile $I(t)$; (3) group dynamics is much faster than epidemic spreading, in which case analytic theory can be developed as discussed earlier in Sec. 4.4 for SIR. Figure 4.6 summarizes some typical results from the SIR process in the presence of group dynamics, as discussed throughout this paper. As in Fig. 4.3, a broad range of $I(t)$ profiles can be obtained – in particular those with large resurgences and a long-tailed decay – simply by varying ν_{coal}, ν_{frag}, p, and q. As shown in Fig. 4.6(d), individual runs cannot be explained by considering the SIR process on a static weighted network. This is because any static network model neglects the dynamically changing nature of the transmission pathways, i.e. the dynamically changing contact structure.

Figure 4.7 presents the SIS viral process which is appropriate for diseases without immunity. Recovery from infection is followed by an instant return to the susceptible pool. The slow growth and low stable state (endemic equilibrium) are both due to suppression by the group dynamics, which only allows a part of the population to come in contact with infecteds, while others are isolated in different groups. Before an infected node can contact the whole population and infect them, it may recover and become susceptible again. Figures 4.7(b) and (c) show how the restoration rate γ, combined with ν_{frag} and ν_{coal}, affect the stability of $I(t)$ after reaching the endemic equilibrium. Large γ values yield large jumps, and vice versa, because it leads to quickly replenishment of the pool of susceptibles. Figure 4.8 presents the SIRS viral process which is appropriate for diseases in which an infected has a recovery time and then returns to susceptible status. It appears similar to SIS, but produces bigger fluctuations than SIS since there is the possibility that infecteds maintain an inert

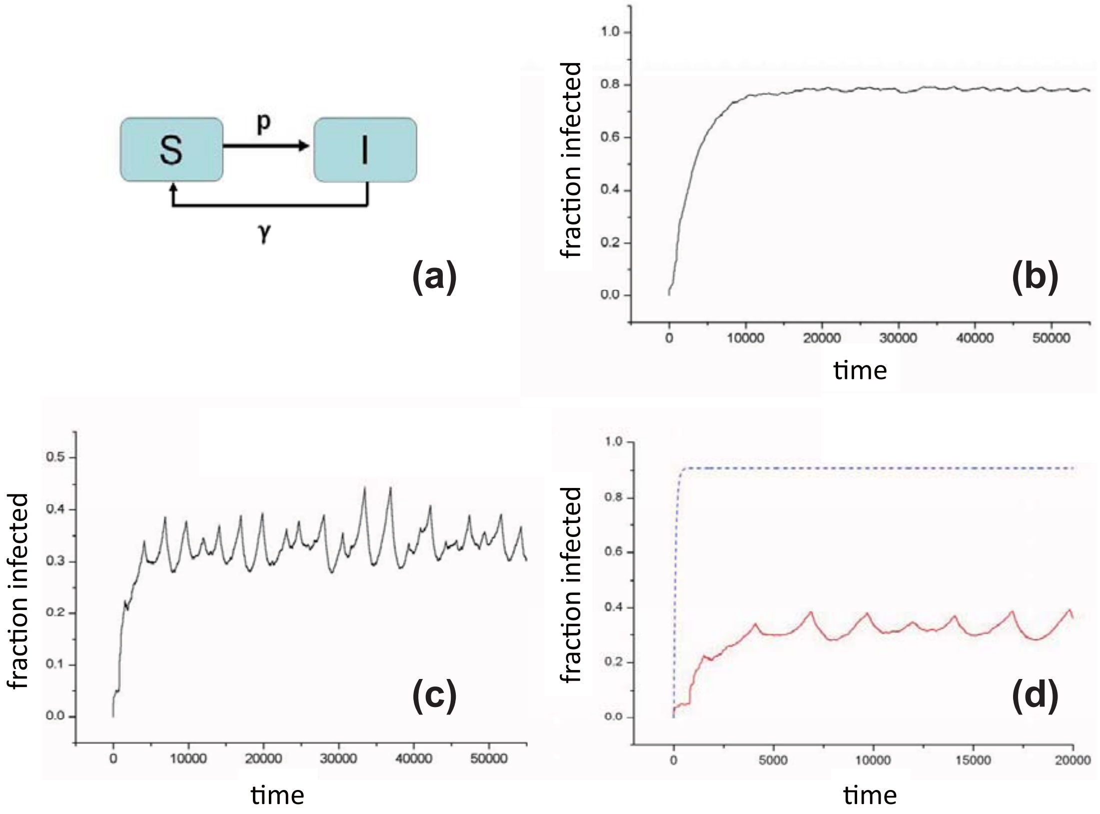


Figure 4.7: (Color online) SIS process in presence of the coalescence-fragmentation group dynamics. (a) Schematic SIS process. (b) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.01$, $\nu_{\text{coal}} = 0.99$, $p = 0.01$ and $\gamma = 0.0001$. (c) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.001$, $\nu_{\text{coal}} = 0.99$, $p = 0.01$ and $\gamma = 0.001$. (d) Comparison between $I(t)$ (solid curve, red) using the same parameters as (c), and $I(t)$ for a weighted network (blue dotted curve) in which all nodes are connected with strength P.

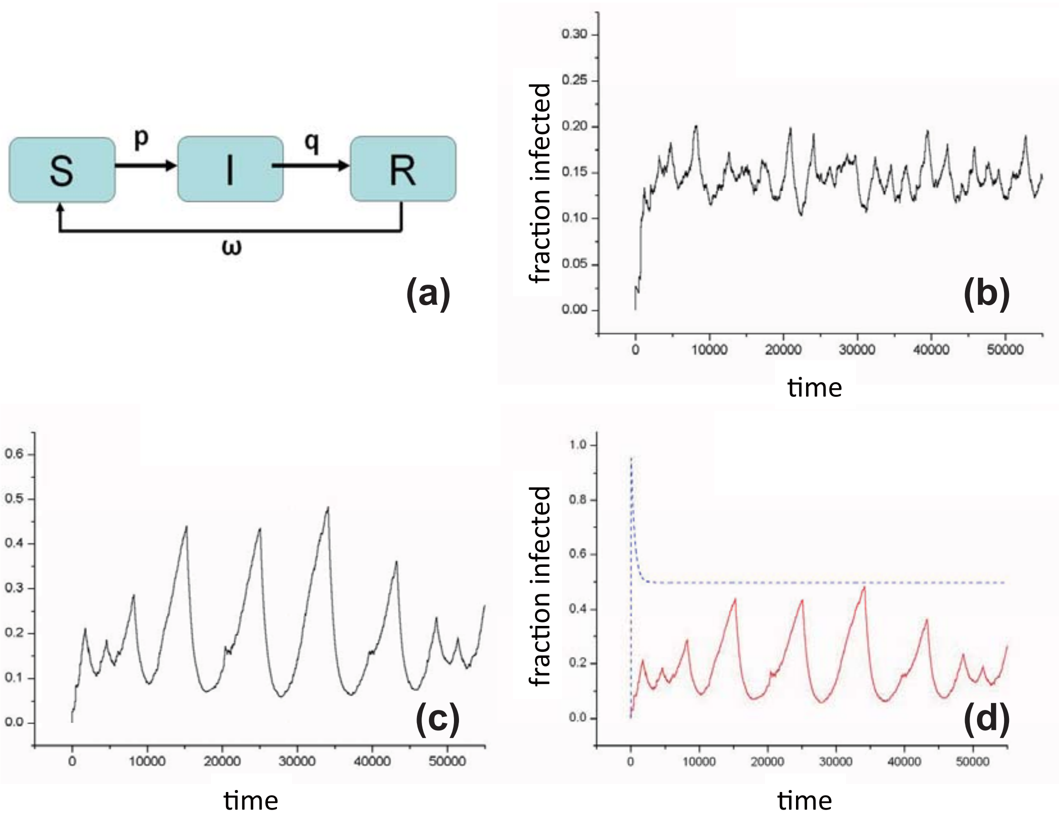


Figure 4.8: (Color online) SIRS process in presence of the coalescence-fragmentation group dynamics. (a) Schematic SIR process. (b) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.01$, $\nu_{\text{coal}} = 0.99$, $p = 0.01$,$q = 0.001$ and $\omega = 0.001$. (c) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.001$, $\nu_{\text{coal}} = 0.99$, $p = 0.01$,$q = 0.001$ and $\omega = 0.001$. (d) Comparison between $I(t)$ (solid curve, red) using the same parameters as (c), and $I(t)$ for a weighted network (dotted curve, blue) in which all nodes are connected with strength P.

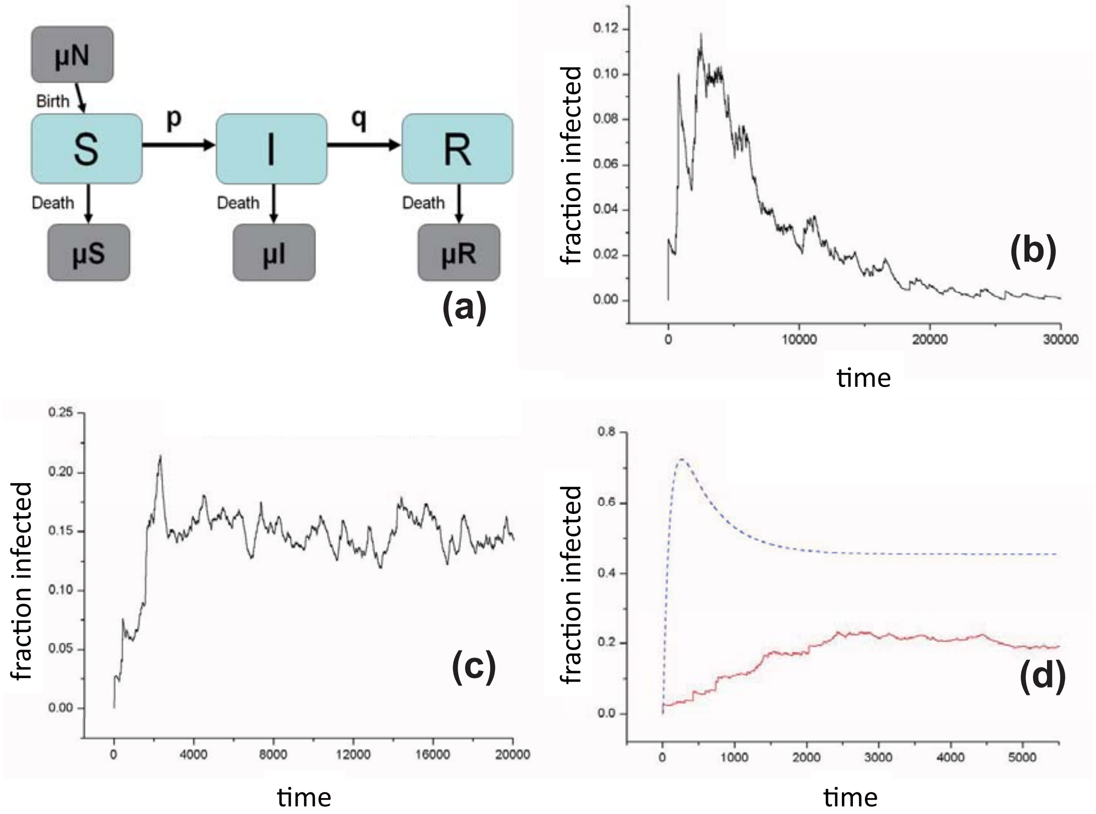


Figure 4.9: (Color online) SIRD process (i.e. SIR with demography) in presence of the coalescence-fragmentation group dynamics. (a) Schematic SIR process. (b) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.01$, $\nu_{\text{coal}} = 0.99$, $p = 0.01$, $q = 0.001$ and $\mu = 0.000001$. (c) Typical individual simulation run showing $I(t)$ for $N = 10000$, $\nu_{\text{frag}} = 0.01$, $\nu_{\text{coal}} = 0.99$, $p = 0.01$, $q = 0.001$ and $\mu = 0.001$. (d) Comparison between $I(t)$ (solid curve, red) using the same parameters as (b) but with $\mu = 0.001$, and $I(t)$ for a weighted network (blue dotted curve) in which all nodes are connected with strength P.

state (R) for a finite time, which hence allows $I(t)$ to become very small. The net effect is that the $I(t)$ peak-trough ratio is enhanced as compared to SIS.

Figure 4.9 presents the SIRD viral process which corresponds to the addition of demography (i.e. births and deaths) to SIR. It assumes that there is a natural mortality μ (i.e. each individual has a lifespan $1/\mu$). In order to keep the total population constant $S + I + R = N$, μ also represents the birth rate of the population. Figure 4.8(b) shows that when the birth-death processes are much slower than infection and recovery processes, the infection profile $I(t)$ is essentially the same as SIR model including group dynamics. The fact that the endemic equilibrium is not stable can be solved by increasing μ. However we stress that as in all other cases, the resultant $I(t)$ profile is different from the corresponding result for a static weighted network (Fig. 4.9(d)).

4.7 Summary and outlook

We have presented and analyzed a simple model of contagion within a population featuring dynamically evolving connectivity, allowing group-level dynamics and the individual-level transmission process to co-exist on similar timescales. In spite of the simplicity of our model, we find that the profiles produced bear a striking resemblance to a wide variety of real-world examples from social, financial and biological domains. The common features of multiple resurgent peaks and abnormal decay times are observed both theoretically and empirically. To demonstrate further the generic nature of the empirical profiles that we show in Fig. 4.3, and hence the relevance of our theoretical model, we refer interested readers to examine the explicit YouTube profiles in Ref. [144], and the wider range of examples in Ref. [158]. In terms of further justifying the underlying group dynamics that we imposed, we stress that the recent Ref. [113] shows explicitly that these coalescence-fragmentation group dynamics are consistent with that observed in insurgencies. Adding in the feature of transmission, as we do in the present paper, elevates this particular real-world application to the hotly debated issue of understanding how information and know-how about IEDs (Improvised Explosive Devices) – in particular innovations in design – spreads through an insurgency. This latter topic is a very important practical one given the recent insurgent preference for IED use in the current wars in Iraq and Afghanistan.

Although it is of course possible that such empirically observed $I(t)$ profiles can be generated by other more sophisticated models – e.g. more detailed social dynamic mechanisms, the introduction of spatial heterogeneity, or more elaborate generalizations of the SIR transmission process – we find it intriguing that our simple analysis suffices. We hope that these findings stimulate future work on the potential effects of different social group dynamics (see, for example, Palla et al. [155]) and on detailed analytic descriptions of the resulting infection profiles.

CHAPTER 5

Adding Degree of Freedom to Individual Particles

5.1 Background

Quantifying the group dynamics of living objects is a fundamental challenge across the sciences [54, 159–169]. Humans represent a particularly difficult case to analyze, since their groups can be formed in both real (offline) and virtual (online) spaces. Such fascinating sociological challenges have attracted the attention of many physicists in recent years under the heading of Econophysics and Sociophysics [165–170]. Indeed, the Econophysics website provides a rapidly increasing repertoire of such investigations [171].

Massively multiplayer online games typically allow individuals to spontaneously form, join or leave a formal group called a guild [172, 173]. The design of the game encourages players to form such groups by making the most rewarding quests (i.e., missions) too difficult to accomplish alone. Millions of people worldwide log on to the world's largest online game (World of Warcraft (WoW)) for the equivalent of several days every week. Indeed, online games are one of the largest collective human activities on the planet and hence of interest from the perspectives of global commerce [174], security [175] and even epidemiology [176]. A seemingly unrelated social phenomenon which is also of great concern, is urban gangs. Urban gangs have been gaining in popularity among young people both nationally and internationally [177,179,193]. There are obvious differences in the settings and history of online guilds and offline gangs, however the empirical datasets that we have compiled enable us to perform a unique comparative study of their respective grouping dynamics [172,173,179].

Studies of the formation and evolution of groups have long occupied a central position within the sociological and organizational theory literatures, particularly in terms of understanding how individual level characteristics (e.g., demographics, skill sets) shape group dynamics [54,159–163,166–169,180–184]. Proponents of homophily tend to argue that individuals choose to participate in groups that minimize within-group heterogeniety, since sameness facilitates communication and reduces potential conflict [161,185–188]. With respect to stability, previous research has suggested that members of groups that are most unlike the other members of the group are also more likely to exit the group [189]. By contrast, some researchers suggested that rather than minimize diversity among members, members might instead join groups that maximize the diversity of skills in the group (team) [190, 191] since a wider skill set might be more efficient in meeting particular goals [168, 190, 191].

In this paper, we analyze data obtained from street gangs in the offline, real world [177, 179, 193] and Internet guilds in virtual online worlds within massively multiplayer online role-playing games [172–174]. We develop and employ a physically motivated model to analyze these two high profile, yet seemingly unrelated, human activities. The underlying datasets were obtained from online WoW guilds [172, 173] and urban street gangs in Long Beach, California [179]. They have been separately compiled by members of our team over the past few years through a combination of field-work and data compilation, and are believed to be state-of-the-art datasets for each respective system. As a result of our analysis, we uncover evidence of a quantitative link between the collective dynamics in these two systems. Although the observable group-size distributions are very different, we find that a common microscopic mechanism can reproduce the observed grouping data for each, simply by adjusting the populations' average attribute property. In particular, we find that the evolution of gang-like groups in the real and virtual world can be explained using the same team-based group formation mechanism. In contrast to the quantitative success of our team-based model, we find that a homophilic version of the model fails. Our findings thus provide quantitative evidence that online guilds and offline gangs are both driven by team-formation considerations, rather than like-seeking-like. Interestingly, each server's Internet Protocol (IP) address seems to play an equivalent role to a gang ethnicity. Given the current public concern regarding the social consequences of intensive Internet game-playing, and separately the current rise in street gangs [177, 179, 193], we hope that the present findings help contribute to the debate by setting these systems on a common footing.

The plan of the paper is as follows: Sec. 5.2 gives the main empirical results that are to be modelled. Section 5.3 gives the key ideas and a detailed description of our self-organized team formation model. The main results comparing the cumulative group size distributions from our model and real data for both WoW guilds and LA gangs are also presented, so as to establish the validity of our model. Section 5.4 gives further analysis on the non-cumulative WoW guild size distributions for separate servers, as well as the group size distributions for LA gangs of different ethnic groups. These results represent a more stringent test on our model. Section 5.5 defines the kinship model and demonstrates its inadequacy. Finally, Sec. 5.6 VI provides the conclusions and discusses the implications.

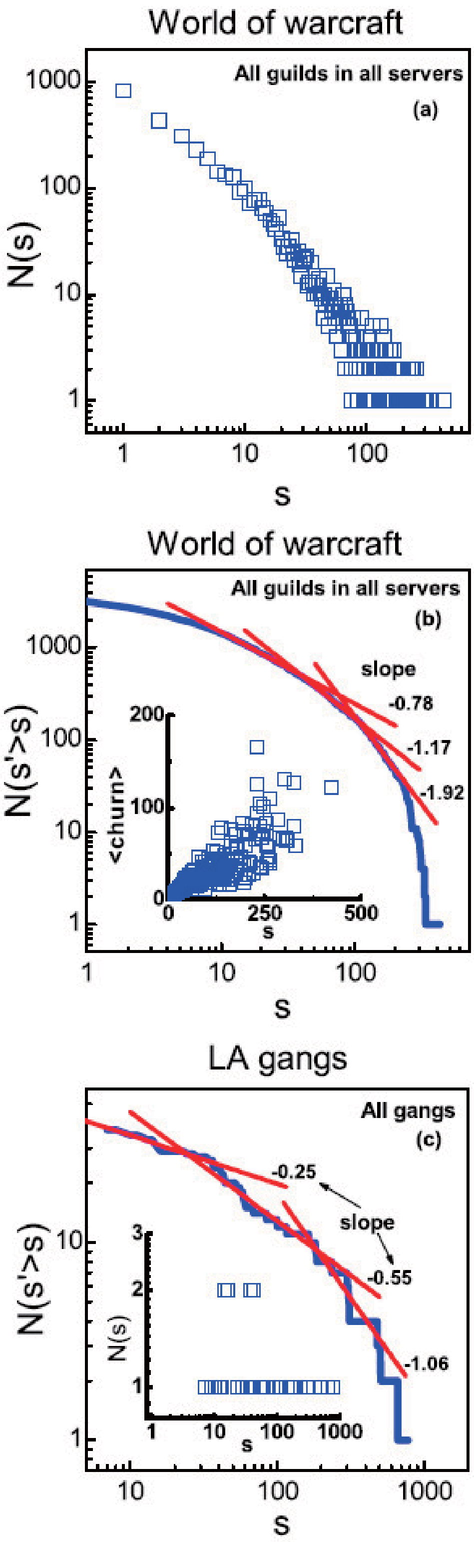


Figure 5.1: (Color Online) Internet guilds and street gangs. (a) Empirical data from World of Warcraft on all servers. (b) Cumulative distribution differs significantly from a power-law. Inset shows the averaged churn of the guilds. (c) Cumulative distribution for Long Beach (i.e. “LA”) gangs. Inset shows the underlying discrete distribution.

5.2 Main empirical results

Figures 5.1(a) and 1(b) give the guild size distribution $N(s)$ and cumulative guild size distribution $N(s' > s)$ for the WoW dataset, using October 2005 as a representative month. Overall, data were collected from three different servers, each representing a different game environment, between June 2005 and December 2005. There are 76686 agents involved in a total of 3992 guilds spread across three servers: S1, S2, and S3. The cumulative distributions for the separate servers S1, S2 and S3 will be shown later (see Fig. 5.3(a)). All three servers are based in the US and were selected at random, with the servers' identities anonymized to preserve players' privacy. The vertical axis $N(s)$ is the number of guilds of size s. Data is shown using October 2005 as a representative month, however other months show similar behavior as demonstrated in a later section. Interestingly, the distribution is neither a Gaussian nor a power-law. Figure 5.1B confirms that if we were to insist on power-law behavior, the supposedly constant slope in $N(s' > s)$ would vary unacceptably. The inset in Fig. 5.1(b) shows the averaged churn $\langle \text{churn} \rangle$ versus the guild size s, where $\langle \text{churn} \rangle$ describes the monthly guild dynamics as follows: The membership of a guild is recorded at the beginning and end of each month, with the churn being the number of players who were members at the beginning of the month but who then left during that month. For guilds which have the same size at the beginning of the month, we then average over the churn values and call this averaged quantity $\langle \text{churn} \rangle$. We have checked across different months, and have also looked at different measures, in order to convince ourselves that the data in Fig. 5.1 are typical of the WoW data. Figure 5.1(c) shows our empirical data for the 5214 members of street gangs in Long Beach, California just outside of Los Angeles. The data are shown for June 2005, but again other months show similar behavior. For convenience, we label these as “LA gangs”.

All gangs are included irrespective of their ethnicity (e.g. Latino). The number of real gangs is much smaller than the number of guilds in WoW. $N(s' > s)$ for gangs is not smooth – nor is it a power-law with a well-defined slope, as shown explicitly in Fig. 5.1(c).

5.2.1 WoW: Monthly Guild Size Distributions and Churn

To demonstrate that the form of the distribution in October 2005 is typical of the WoW data, we also analyzed the data for all the remaining months. For each month, we repeat the same exercise of counting the guilds and their sizes for each server. Here, we show the data for several additional months (i.e. June, August and December 2005) as well as October 2005. The empirical data shows that the number of players in each month was 80183 (June), 93127 (August), 76686 (October), and 93322 (December). Figure 5.2 shows $N(s)$ and $N(s' > s)$ for these four months. The distributions for different months behave in a similar way. The results indicate that the guild size distribution measured at any time during the data collection process, represents a general property of the game during the entire data collection window.

In the inset of Fig. 5.1(b), we showed the values of $\langle \text{churn} \rangle$ for all the guilds in the three servers (S1, S2, S3) for October 2005. Here, the data of $\langle \text{churn} \rangle$ for June, August, and December 2005 are shown in Fig. 5.3. The data indicate that the behavior of $\langle \text{churn} \rangle$ versus guild size is almost the same for every month. Thus, the behavior $\langle \text{churn} \rangle \sim s$ is a general feature of the WoW data. We have also analyzed the data for separate servers, and the behavior is again nearly the same. Note that there are necessarily fewer data points for a single server, hence it is more convenient to show the results corresponding to all servers bundled together. Later, we will compare

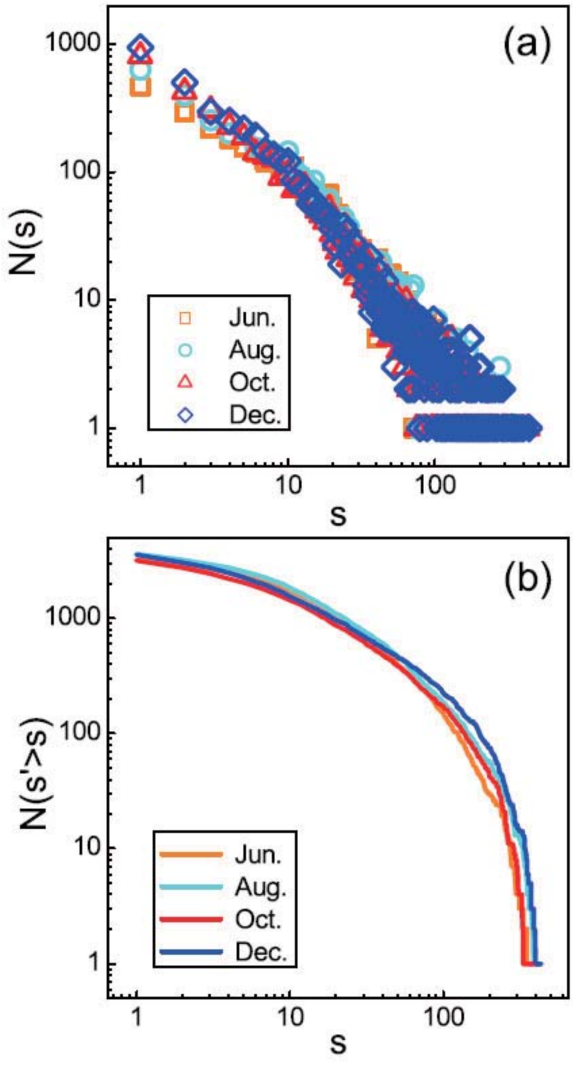


Figure 5.2: (Color Online) (a) WoW guild size distributions $N(s)$ for the months June, August, October, and December 2005. The total numbers of players in these months are 80183, 93127, 76686, and 93322, respectively. (b) The cumulative guild size distributions $N(s' > s)$ for each of the four months.

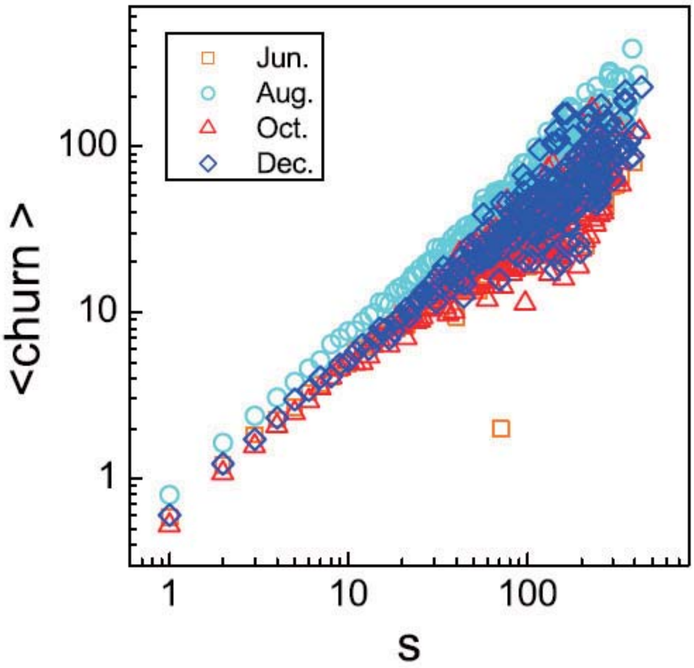


Figure 5.3: (Color Online) The average churn ⟨churn⟩ (as defined in text) as a function of guild size in the WoW dataset in a log-log plot, treating the data in all three servers collectively. Data for the months June, August, October, and December 2005 are shown.

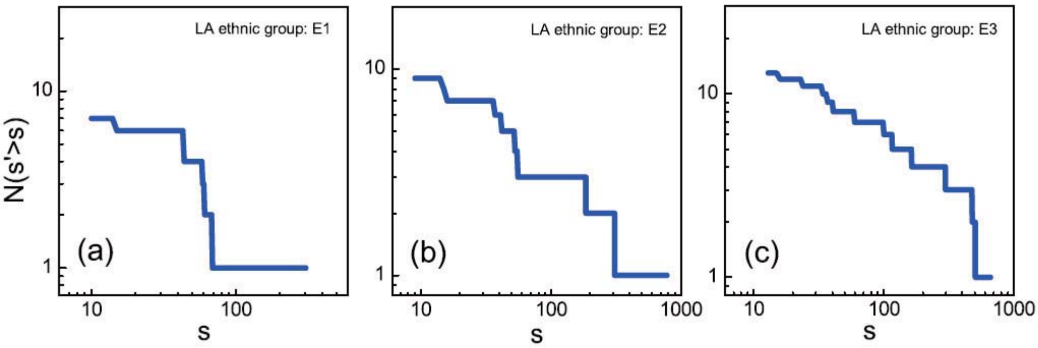


Figure 5.4: (Color Online) The cumulative gang size distribution $N(s' > s)$ for LA gangs of three main ethnic groups. (a) Cumulative gang size distribution for gangs with ethnicity E1. The total membership is $N = 608$. (b) Ethnicity E2 with total membership $N = 1504$. (c) Ethnicity E3 with total membership $N = 2552$.

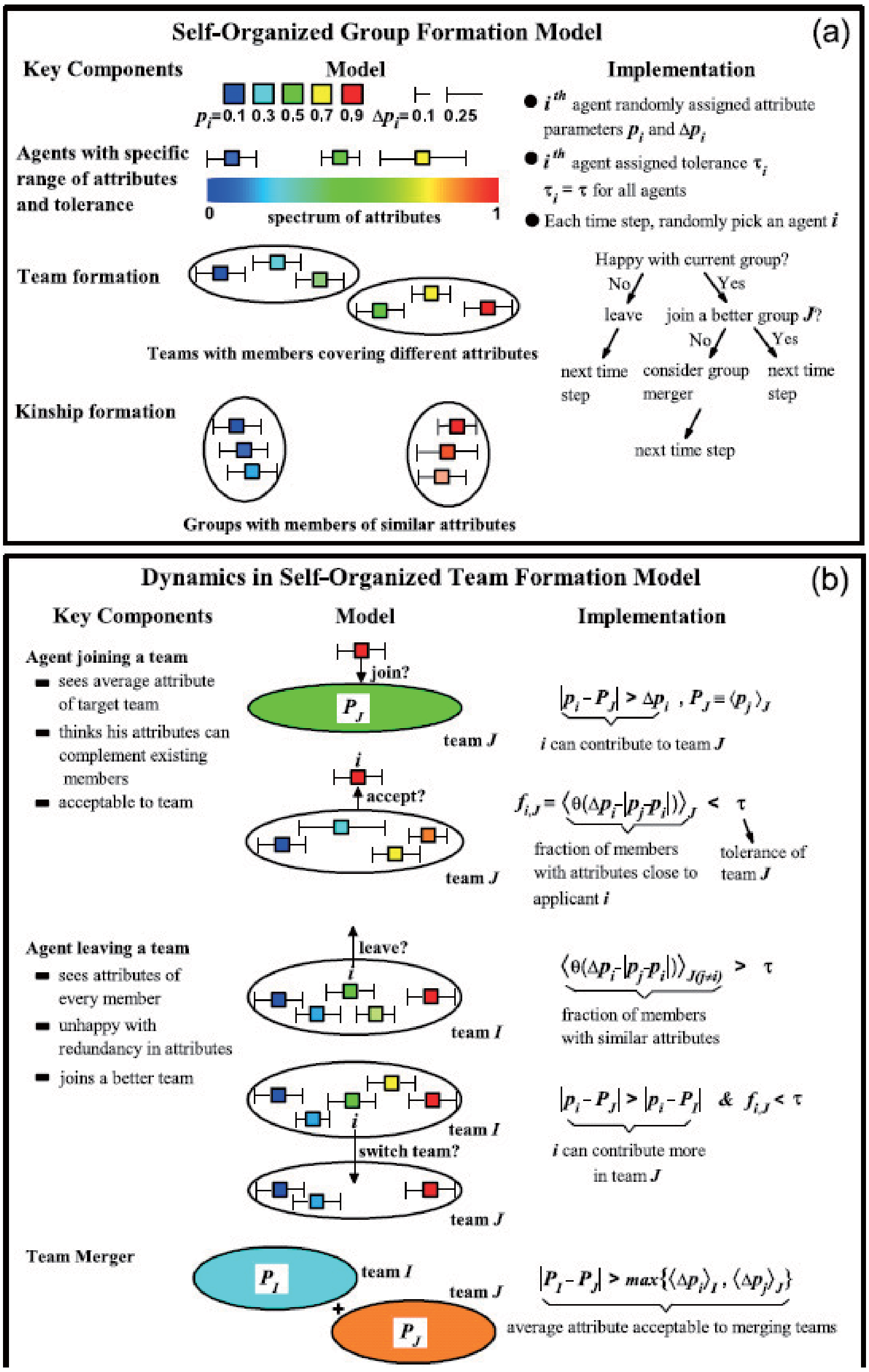


Figure 5.5: (Color Online) Our generic model of group dynamics. (a) The basic model setup, without yet specifying the criterion that an agent uses when seeking to join or leave a group. Two possible extremes are the team-formation model shown in Fig. 5.2.1(b), where an agent seeks a group with a suitable niche in p-space, and the kinship model (not shown) where an agent seeks a group having members with a similar p-value. Details of the implementation and specific rule-sets are discussed in Sec. 5.3.

results of $N(s)$ as obtained by our team-formation model with data of separate servers (see Fig. 5.7).

5.2.2 LA gangs: Different ethnic groups

Our dataset on LA gangs collected in June 2005 consists of the sizes and the ethnicity of the gangs. Putting all the data together, there are a total of 5214 members. The cumulative distribution $N(s' > s)$ was shown in Fig. 5.1(c). The distribution shows a similar shape as for the WoW cumulative guild size distribution. From the information on the ethnicity of the gangs, there are three main ethnic groups that one can identify. For privacy reasons, we label these groups as E1, E2, and E3, with membership 608, 1504, 2552, respectively. Figure 5.4 shows the cumulative gang size distributions $N(s' > s)$ for the three major ethnic groups. For each of these ethnic groups, the number of gangs is very small (around 10). For this reason, $N(s' > s)$ shows step-like behavior. Comparing with WoW data, the total number of gangs and the number of members in the LA gang data are both much smaller than the corresponding numbers in WoW. In a later section, we will compare the results for our team-formation model with these data of different ethnicity groups (see Fig. 5.8).

5.3 Self-organized team formation model

In this section, we introduce the key ideas and describe in detail our model and show that our simply model reproduces accurately the quantitative features in empirical data. As an overview, Fig. 5.2.1(a) shows our generic model of self-organized group formation which acts as the core setup for implementing specific rule-sets for joining and leaving a group – for example, team-formation (see Fig. 5.2.1(b)) or homophilic kinship. Our generic model (Fig. 5.2.1(a)) creates a heterogeneous population by assigning an attribute p_i to each person (i.e. agent) i. Since people may have a range of attributes, we assign each agent a spread Δp_i around p_i. With the goal of building a minimal model, we choose each p_i to be a single number chosen randomly from a uniform distribution between zero and one. More complicated models can be built by assigning an array of numbers to describe the attributes of a person. The values of Δp_i's are random numbers drawn from a single-peaked distribution with mean $\langle \Delta p_i \rangle$ and spread (i.e. standard deviation) $\sigma_{\Delta p}$. The Δp_i values are shown in Fig. 5.2.1 as horizontal bars around the corresponding color-coded p_i value. We then assign a tolerance to every agent – for simplicity, we choose the same value τ for each agent. The tree on the right-hand side of Fig. 5.2.1(a) applies to both team-formation and kinship versions. In the team-formation version, the group contains agents with complementary attributes (i.e. a team) while in the kinship version a group contains agents with similar attributes (i.e. like with like).

The model can be constructed without considering a particular context. It could represent players in WoW, members in gangs, employees in companies, etc. Figure 5.2.1(b) describes what happens in one timestep in the team-formation implementation of Fig. 5.2.1(a), both schematically and mathematically. The kinship model, described later, essentially corresponds to an opposite set of add-on rules to the team-formation model. The team-formation model, as we shall see, works better for the empirical data and we will focus on it in this section. We will use the words "team" and "group" interchangeably in the following discussion. However we emphasize that for the portions of the following discussion concerning Fig. 5.2.1(a), the word "team" can be replaced by "group" since the statements apply equally to the team-formation model and the kinship model.

Parameters – Consider a population of N agents or players. The attributes of an agent i are described by a set of numbers $(p_i, \Delta p_i, \tau_i)$, where p_i describes the ith-agent's mean attribute. Δp_i describes the i-th player's range of attributes around p_i, or equivalently a breadth of skills around the mean skill. The value of Δp_i is independent of the value of p_i. Here, τ_i is a parameter that describes the tolerance of an agent in deciding whether to leave a group, after he compares how close his attributes are to the members of the group. In the present model, we have not included the possible evolution of attributes, although this is an interesting problem for future studies.

Initialization – Initially, each agent is randomly assigned his attribute parameter p_i, the value of which is chosen randomly from a uniform distribution between 0 and 1. The agents' Δp_i's are assumed to follow a Gaussian distribution characterized by a mean $\langle \Delta p_i \rangle$ and standard deviation $\sigma_{\Delta p}$. Each agent is then assigned a value of Δp_i from this Gaussian distribution. With p_i and Δp_i, the agent i covers the attributes $p_i \pm \Delta p_i$, for attributes characterized by the range between 0 and 1. The coverage of attributes is not allowed to go below 0 or above 1, i.e., when $p_i + \Delta p_i > 1$, the upper bound is set at 1 and when $p_i - \Delta p_i < 0$, the lower bound is set at 0. For simplicity, the values of τ_i are taken to be the same for all agents, i.e., $\tau_i = \tau$ for all agents. The total number of agents in the system N can be easily taken from the real data. Thus, the model is completely characterized by four physically meaningful parameters: N, $\langle \Delta p_i \rangle$, $\sigma_{\Delta p}$ and τ.

Key Ideas and Model Implementation – In each timestep, an agent i is randomly picked. The attachment of the agent i to a group then follows the rules below.

(a) **For a single agent joining a team** – This step is imposed when the agent i being picked is an isolated agent. In this case, another agent j is randomly picked. The agent j belongs to a team labelled J with n_J members. Note that $n_J = 1$ if j is an isolated agent. The key idea is that it is a two-way consideration when an agent i wants to join a team J: the agent must find a team to which his attributes could contribute, and that team must in turn find the agent's attributes acceptable. Moreover, the agent can only see the average attributes of the team to which he is applying. In other words, when joining a team, an agent will be guided by general information about the team (i.e. the average attribute of the team) rather than detailed information about all its members. This mimics the fact that an outsider cannot be expected to be aware of all the details of a team's members before joining, since such knowledge can generally only be gained after being a member of that team. Once inside the team, this information can then be gained either through direct access to insider knowledge, or simply through osmosis.

An agent i therefore assesses a team J of which he considers to join, by looking at the average attribute P_J of that team:

$$P_J = \frac{1}{n_J} \sum_{k \in J} p_k, \tag{5.1}$$

where the sum is over all members of team J. The agent i will find the team suitable if his attributes complement those of the existing members. Therefore, if his attributes are too close to that of the existing members of the team, he feels that he could not contribute much and he will not join the team. The condition that the agent i finds the team J acceptable can thus be modelled by $|p_i - P_J| > \Delta p_i$.

For the team J, it will consider whether to enroll agent i as a new member. As an applicant to the team J, the team will know the range of attributes that agent i could cover and then assess the potential contribution of agent i to the team. This can be measured by counting the number of existing members with attributes in the range of agent i, normalized by the team size n_J. Thus, we define $f_{i,J}$ as

$$f_{i,J} = \frac{1}{n_J} \sum_{j \in J} \theta(\Delta p_i - |p_j - p_i|), \tag{5.2}$$

where the sum is over all members in team J and $\theta(x)$ is the Heaviside function, i.e., $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ otherwise. In deciding whether to accept a new member, we define a team's tolerance by averaging the individual tolerance of its members, i.e.,

$$\tau_J = \frac{1}{n_J} \sum_{j \in J} \tau_j. \tag{5.3}$$

For $\tau_j = \tau$ for all agents, $\tau_J = \tau$. Note that $f_{i,J}$ is a quantity less than unity. If $f_{i,J}$ is large, many existing members in team J have attributes that are close to that of agent i and thus the team tends not to accept agent i as a new member due to redundance in attributes. Thus, the condition that the team J will accept agent i as a new member is $f_{i,J} < \tau_J$.

Considering joining a team requires two-way consideration, the criteria for an agent i joining a team J are: $|p_i - P_J| > \Delta p_i$ and $f_{i,J} < \tau_J$.

(b) **For an agent leaving a group, finding a better group, or for groups merging** – This step is imposed when the agent i being picked belongs to a group labelled I with n_I ($n_I > 1$) members. The following attempts are implemented **in sequence**.

(i) **Agent i decides whether he can tolerate the team** – After being a member of team I for a while, the agent i has the chance to explore the microscopic details (individual attributes) of the team members. The key idea is that if he finds that there are many members with similar attributes to his, he will leave. To decide whether he can tolerate the team, he looks at the fraction f_i of members in the team with attributes within his range of coverage, i.e.,

$$f_i = \frac{1}{n_I - 1} \sum_{k \in I (k \neq i)} \theta(\Delta p_i - |p_k - p_i|), \tag{5.4}$$

where the sum is over all the agents in the team I except the agent i himself. Note that $0 \leq f_i \leq 1$. If f_i is close to 1, then there are too many members with similar attributes and the agent i will have a higher tendency to leave. If $f_i > \tau_i$, the i-th agent cannot tolerate the team any more and he *leaves* the group to become an isolated agent. If this happens, the timestep ends.

(ii) Another key idea is team switching. If the agent i finds that he can tolerate the team, it does not necessarily mean that he is very happy with the team. He will try to find a better (more suitable) team to join. An agent j, who belongs to a group J, is randomly picked. The agent i will then compare whether the current team I or the team J is more suitable for him. He intends to join team J if $|p_i - P_J| > |p_i - P_I|$. This criterion implies that the agent i finds that he can contribute more in team J than in team I. Whether team J would accept agent i as a new member is again determined by the criterion $f_{i,J} < \tau_J$, as in step (a). Thus, the criteria for agent i to switch from team I to team J successfully are $|p_i - P_J| > |p_i - P_I|$ and $f_{i,J} < \tau_J$. If there is group switching, the timestep ends. We remark that the steps (b)(i) and (ii) are similar to job hunting. If the job is too bad, then one will simply quit without finding a new job. This is reflected in (b)(i). However, even if the job is acceptable, one tries to look for a better job. In job hunting, it is a two-way process: The employer is looking for someone who can cover the weaker aspects or services in a company, and the employee is looking for a better place. This is reflected in (b)(ii).

(iii) The next key idea is to allow for team mergers. If nothing actually happened in (i) and (ii), i.e., the i-th agent does not leave the team I, either because he is happy or because team switching does not work, we consider the possibility of allowing two teams to merge. Team I to which agent i belongs, merges with another team J under the criterion $|P_I - P_J| > \Delta P_I$, where $\Delta P_I = (1/n_I)\sum_{i \in I} \Delta p_i$. Similarly, team J considers merging with team I under the criterion $|P_J - P_I| > \Delta P_J$. That is to say, if $|P_I - P_J| > max(\Delta P_I, \Delta P_J)$, then teams I and J merge to form a bigger team. Note that there are two ways to implement mergers. The team J could be the same team that the j-th agent belonged to in procedure (ii) above, or a new agent j can be picked randomly when mergers are considered. Results are nearly identical for the two ways.

To summarize, the key ingredients in our team-formation model are: (i) Teams tend to recruit members to cover a spectrum of attributes; (ii) agent joins a team by assessing his potential contribution to the team; (iii) agent joining a team only sees an average of the attributes of a team; (iv) team accepts new member by assessing his potential contribution; (v) agent leaves a team when there are many members with similar attributes; (vi) agent always looks for better teams where he could contribute more; and (vii) team tends to expand by mergers when its membership becomes stable. Each of these ingredients seems reasonable based on our common knowledge of how people behave in team situations. We remark that this set of rules allowed us to produce results that are similar to the real data in both the WoW and gangs data. If more data are available to put further constraints on the modelling or if we only want to model a selected behavior in the data, then more rules may be needed or some rules may be simplified. For example, the ingredient (vi) is needed to model the ⟨churn⟩ in WoW data and the ingredient (vii) is needed to get at the proper sizes of the bigger groups in real data.

Figure 6 highlights the main modelling results. Figure 5.3(a) shows that excellent agreement is obtained across the entire range of observed group sizes for $N(s' > s)$, between the empirical WoW guild data from Fig. 5.1(b) (dark blue) and the team-formation model (red) of Fig. 5.2.1. Throughout this paper, the model(s) is implemented with the observed number of agents as an input. Here $\tau = 0.69$, $\langle \Delta p_i \rangle = 0.16$ and $\sigma_{\Delta p} = 0.022$, but we stress that good agreement can be obtained across a reasonably wide range of parameter choices. The remaining panels (Fig. 5.3(b)-(d)) show the data separated by server. The parameter values used are within 10% of those quoted above. To calculate ⟨churn⟩ in the model, we record the membership for each guild in a run during 0.7 Monte Carlo timesteps (after a transient of 1000 Monte Carlo steps). A Monte Carlo time step is the duration over which each agent has, on average, been chosen once for carrying out the dynamics in the model, i.e. each agent has been given a chance to join or leave a group. We have tried different time windows so as to obtain the ⟨churn⟩ in WoW data and found that 0.7 Monte Carlo timesteps give good agreement. However, we remark that it is different to relate the time scale in simulations with real time exactly. A similar process is followed for the LA gangs in Fig. 5.3(e). In an analogous way to the breakdown by computer server in Fig. 5.3(a)-(d), one can breakdown the LA gang data by ethnicity. The fit by gang ethnicity (see Sec. 5.4 later) is good even though the numbers are much smaller than WoW and hence more prone to noise. This surprising connection between ethnicity and server is consistent with the fact that it is essentially impossible to change one's real-world ethnicity or virtual-world server (unless a large fee is paid to WoW in the latter case,

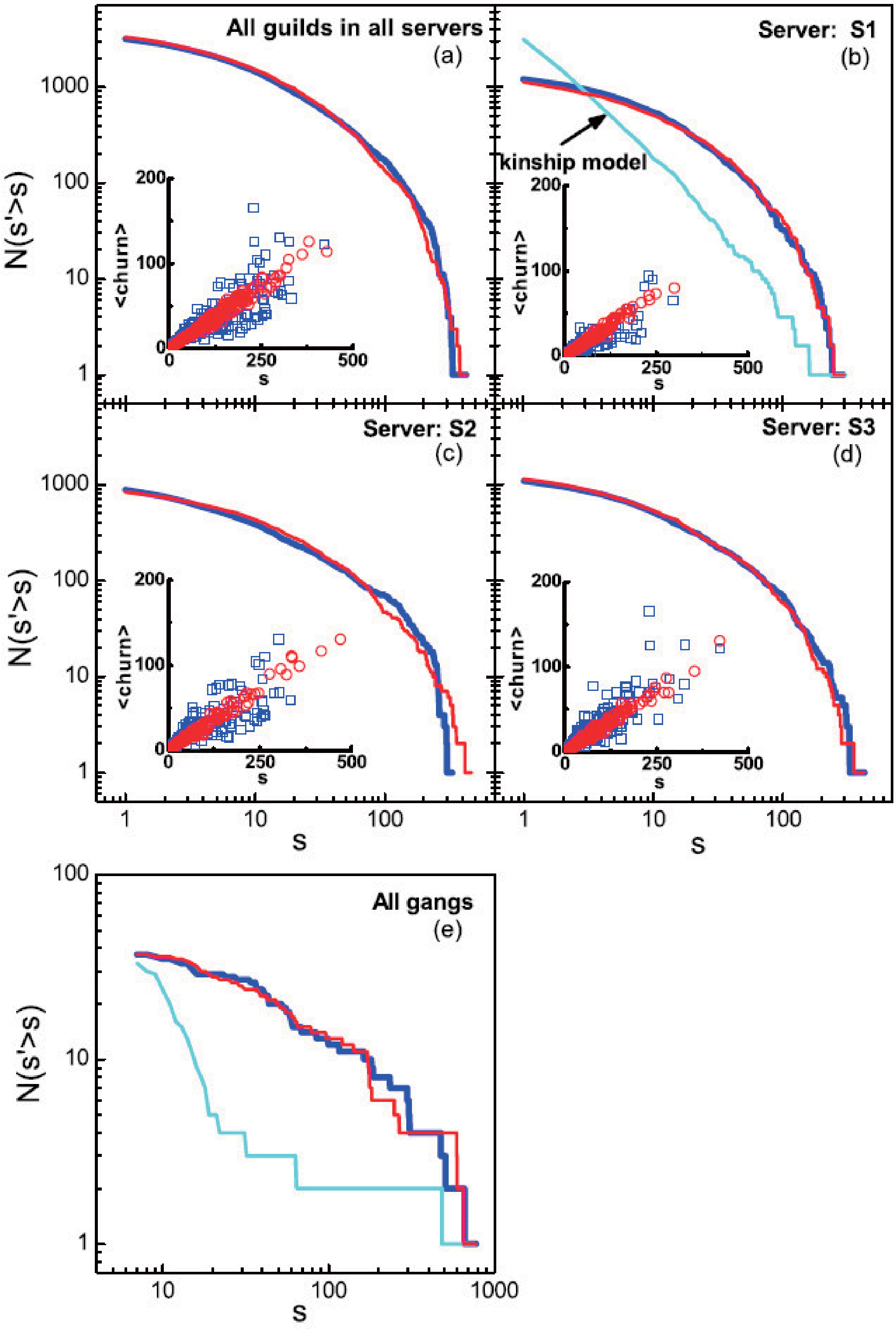


Figure 5.6: (Color Online) Empirical data and model comparison for (a)-(d) World of Warcraft and (e) LA gangs. Empirical data are dark blue, and the team-formation model from Fig. 5.2.1(b) is in red. The kinship model (light blue) produces a poor fit in both cases.

and even then it is an irreversible process). It is also intriguing that the best-fit model parameter values are so similar across WoW servers, and across gang ethnicities. This suggests a quasi-universal behavior in terms of the *way* in which people form gang-like groups online and offline. The small observed server-dependences (and ethnicity-dependences) can be explained by players on different servers (and gang members of different ethnicities) perceiving their environments differently, and hence adopting slightly different tolerances. Our team-formation model thus manages to capture all the features of the empirical gang and guild dynamics, including the approximately linear increase of the averaged churn with guild size in WoW. By contrast, the kinship (i.e. homophilic) version (see Sec. 5.5 later) of the model does *not* reproduce the empirical results of either WoW or the gangs, even qualitatively, as demonstrated by the light blue curves in Fig. 5.3.

5.4 Further analysis

The agreement of our model with WoW data in $N(s' > s)$ for all and individual servers (Fig. 5.3(a)-(d)) can be further illustrated by comparing the underlying, i.e., non-cumulative, distribution for the guild size distribution $N(s)$. Since $N(s)$ is less smooth and thus more noisy than the cumulative distribution $N(s' > s)$, we are actually executing a more stringent test of the model by carrying out the team-formation model comparison based on $N(s)$ instead of $N(s' > s)$.

From the WoW dataset, we count the number of players in all the guilds in each server, and also the total number of players in all servers. In each case, we take the number of players as input for N and run our team-formation model. By adjusting the parameters $\langle \Delta p_i \rangle$, $\sigma_{\Delta p}$ and τ in the model, we obtained the guild size distributions

$N(s)$ for each of the three separate servers and for the three servers collectively. Figure 5.7 shows the $N(s)$ for our model obtained from one run in each of these cases, together with the distribution obtained from the data. The parameters are given in the figure caption. The results from the team-formation model capture the essential features in the WoW guild size distributions.

From the parameters for each of the servers, it can be seen that they are very similar but not identical. This indicates that while the behavior of the players in different servers are not too different, there are slight differences indicating some kind of special characteristic of a server or game environment. We will see that similar features also appear in the LA gang data, when treating ethnicity separately. To the extent to which the server identity mimics an ethnicity, this seems to open up some deeper sociological questions which can be explored in future research on guilds and gangs.

From our attempts in modelling the real data, we now make a few comments on the model as related to the key features in real data: Step (a) (see Sec. 5.3) that sets the criteria for an agent to join a team and a team to accept a new member is the essence of the team-formation model. This is essential in getting the *shape* of $N(s' > s)$. We observed that the shape of $N(s' > s)$, and thus $N(s)$, is more sensitive to the parameter τ. In the WoW data, there is a quantity called $\langle$churn$\rangle$. In order to get reasonable values for $\langle$churn$\rangle$, a mechanism is required for agents to leave a team or to switch teams readily. Steps (b)(i) and (b)(ii) serve to provide such a mechanism. In order to get at the largest size of the guilds in real data, we need a mechanism for guilds to merge. Step (b)(iii) serves this purpose.

If we were to focus *only* on fitting the guild-size or gang-size distributions, and hence decided not to care about simultaneously fitting the churn $\langle$churn$\rangle$ in the WoW

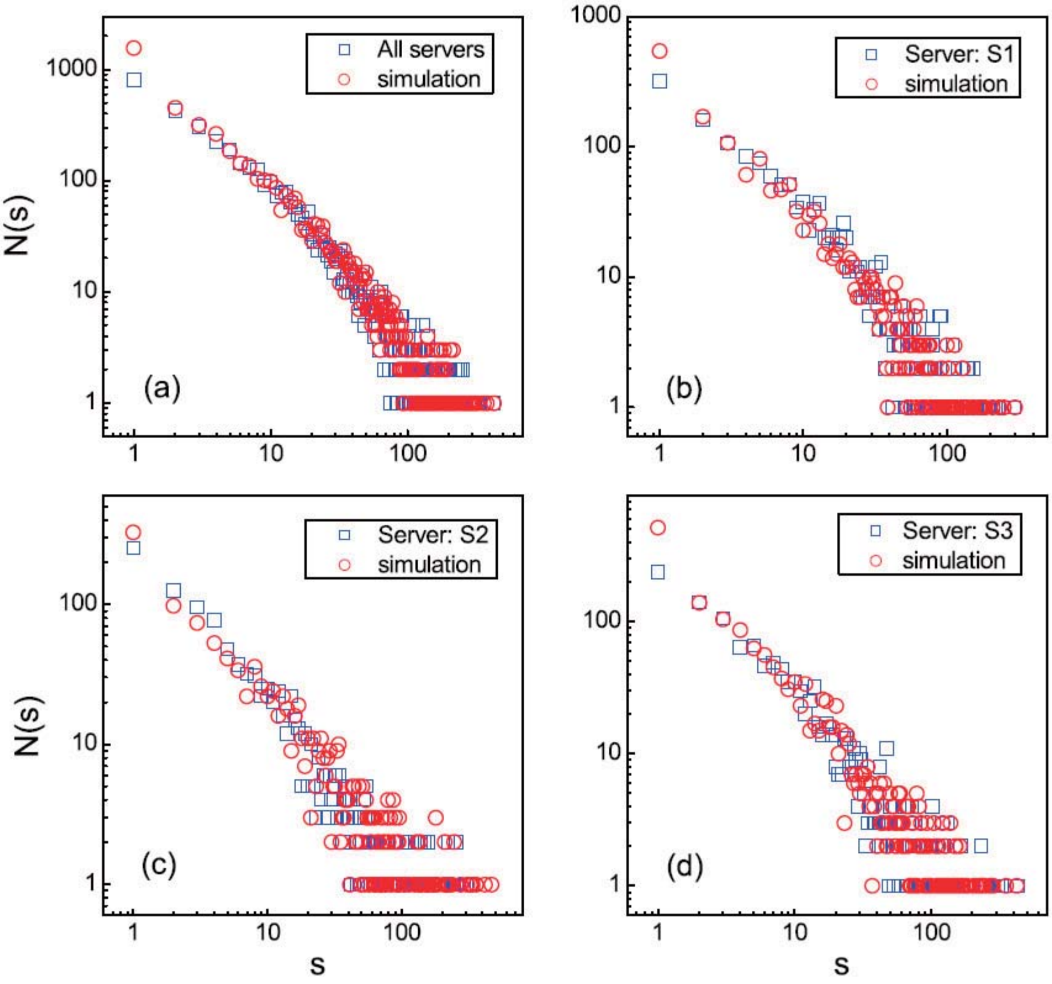


Figure 5.7: (Color Online) The WoW guild size distribution $N(s)$ in October 2005. (a) Guild size distribution treating all servers collectively. The parameters used for team formation are $N = 76686$, $\langle \Delta p_i \rangle = 0.160$, $\sigma_{\Delta p} = 0.022$, and $\tau = 0.69$. (b) Guild size distribution of server S1. The parameters used for team formation simulation are $N = 24033$, $\langle \Delta p_i \rangle = 0.160$, $\sigma_{\Delta p} = 0.020$, and $\tau = 0.67$. (c) Guild size distribution of server S2. The parameters used for team formation simulation are $N = 24477$, $\langle \Delta p_i \rangle = 0.160$, $\sigma_{\Delta p} = 0.025$, and $\tau = 0.75$. (d) Guild size distribution of server S3. The parameters used for team formation simulation are $N = 28176$, $\langle \Delta p_i \rangle = 0.161$, $\sigma_{\Delta p} = 0.020$, and $\tau = 0.70$. Each simulation result is obtained from one particular run of the team-formation model. Note that the parameters for different servers are very similar.

data, we could construct even simpler versions of our model and yet still obtain group-size distributions similar to the real data. For example, a model with slower team switching and more static groups can be used to get at $N(s' > s)$ similar to real data. However, with our present team-formation model we have managed to fit these size distributions *and* account for the churn. One implication of our work is therefore that previous grouping models which have been proposed to explain time-averaged group sizes in real data *without* churn, should be re-examined once such churn data becomes available. Fitting churn as well as the group-size distribution presents a stringent challenge which relatively few candidate models will survive. Performing studies analogous to our present one, would therefore be a very useful way of reducing the number of competing models. Likewise our own extensive experimentation indicates that it would be very hard to identify an alternative model to our team-formation one, in which equally high quantitative accuracy was obtained and yet the structure and/or set of microscopic rules were fundamentally different. This gives us confidence that our analysis has indeed identified a realistic group formation mechanism.

We have tested our model against the empirical data of $N(s' > s)$ for LA gangs data, treating all the gangs collectively (Fig. 5.3(e)). Here, we further test our model for the three major ethnic groups as shown in Fig. 5.4. Treating the ethnicity as the analogy of servers in WoW, we counted the number of members in the gangs in each of the ethnic groups. For the three major ethnic groups E1, E2, and E3, there are 608, 1504, and 2552 members, respectively. These numbers are used as inputs to our model. We then adjust the parameter values in the model to give a distribution $N(s' > s)$ that resembles the empirical data, for each of the three cases. Figure 5.8 shows the results for the ethnic groups E1, E2, E3. It is very encouraging that our model manages to capture the main features of the empirical data for the LA gangs,

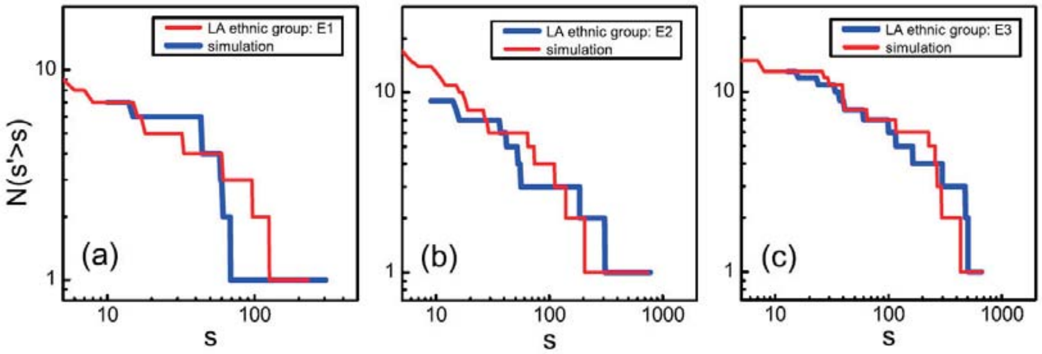


Figure 5.8: (Color Online) The cumulative gang size distribution $N(s' > s)$ for LA gangs of different ethnicity. (a) $N(s' > s)$ of membership of LA gangs of ethnicity E1. The parameters used for the team formation model are $N = 608$, $\langle \Delta p_i \rangle = 0.150$, $\sigma_{\Delta p} = 0.016$, and $\tau = 0.73$. (b) $N(s' > s)$ of membership of LA gangs of ethnicity E2. The parameters used for the team formation model are $N = 1504$, $\langle \Delta p_i \rangle = 0.142$, $\sigma_{\Delta p} = 0.014$, and $\tau = 0.72$. (c) $N(s' > s)$ of membership of LA gangs of ethnicity E3. The parameters used for the team-formation model are $N = 2552$, $\langle \Delta p_i \rangle = 0.141$, $\sigma_{\Delta p} = 0.016$, and $\tau = 0.72$. Each model result corresponds to one run of the team-formation model simulation. Note that the parameters for different ethnic groups are very similar, as was the case for different servers in WoW.

even though the individual gang sizes and number of gangs in each ethnic group are much smaller than for the case of WoW. Note that the parameters (given in the caption) are quite similar for different ethnic groups. Interestingly, using the value of N from the empirical data for each of the ethnic groups, the resulting number of groups in our team-formation model turns out to be similar to that for the empirical data. From the results of WoW guilds and street gangs, we can see that the role of server in WoW has a direct analogy with the role of ethnicity in street gangs.

In summary, our team-formation model reproduces the main quantitative features of the empirical WoW guild size distribution and the cumulative distribution (Fig. 5.1(a), Fig. 5.1(b), Fig. 5.2), in the case when the servers are considered collectively (Fig. 5.3(a)-(d)) *and* in the case when the servers are considered individually (Fig. 5.7). The model *also* reproduces the main feature in the group dynamics

(Fig. 5.3(a)-(d)) observed in the empirical data on churn (Fig.5.1(b) and Fig. 5.3). Furthermore, the agreement between model and empirical data extents to results in different time windows (i.e. months). Our team-formation model *also* reproduces the main quantitative features of cumulative gang size distributions in empirical data (Fig. 5.1(c)), taking the ethnicity collectively (Fig. 5.3(e)) and separately (Fig. 5.8). Thus, our self-organized team-formation model captures quantitatively the features of the group dynamics resulting from cyber-world interactions, as in the case of WoW guilds, and real-world interactions as in the case of street gangs.

5.5 Inadequacy of the alternative kinship model

There are lines (in light blue) in Fig. 5.3(b) for WoW server S1 and in Fig. 5.3(e) for street gangs that show the results for a kinship model. The kinship model is in many ways the "opposite" of the team-formation model, and was introduced to explore homophily as a possible alternative group-formation mechanism. In the team-formation model, the teams tend to recruit members with attributes that spread over the whole spectrum of attributes, i.e., the attributes of the agents complement each other. By contrast in the kinship model, groups form around agents with similar attributes. In short, agents tend to dislike being in a group with agents having very different attributes. Here, we briefly discuss the mechanisms in this kinship model.

We can readily modify our team-formation model in order to create a kinship formation model, as follows. The framework in Fig. 5.2.1(a) remains the same, and so does Fig. 5.2.1(b) in terms of its structure – however we flip the inequalities in Fig. 5.2.1(b) for the criteria for an agent joining a group and for a group accepting a member. A kinship model can hence be defined which is diametrically opposite to our team-formation model, and yet can be discussed on the same footing. In step (a) (see Sec. 5.3), the criteria for an agent i joining a group J are: $|p_i - P_J| \leq \Delta p_i$ and $f_{i,J} \geq \tau_J$. These imply that an agent wants to join a group with an average attribute close to his own, and a group wants to accept new members having attributes close to its existing members. In step (b)(i), an agent i cannot tolerate a group I when he finds the members are too different from him. Thus the agent leaves if $f_i < \tau_i$. In step (b)(ii), each agent is continually looking for a better group which has a more similar average attribute to him. So group switching happens if $|p_i - P_J| < |p_i - P_I|$ and $f_{i,J} \geq \tau_J$. Finally, when membership becomes stable, a group tends to expand by merging with groups having similar average attributes. Thus two groups I and J merge if $|P_I - P_J| \leq min(\Delta P_I, \Delta P_J)$.

In fact, for every team formation model that incorporates the idea of agents with different attributes tending to form a team, a corresponding kinship model can be identified, built around the opposite idea of agents with similar attributes forming groups. However the cumulative distribution function obtained from the kinship model cannot capture even the basic qualitative shape of the empirical data. The detailed reason is that the kinship model tends to produce too many groups of small sizes.

5.6 Conclusions and implications

Our results show that irrespective of their respective settings and origins, the observed dynamics in two very distinct forms of human activity – one offline activity which is widely considered as a public threat, and one online activity which is by contrast considered as relatively harmless – can be reproduced using the same, simple model of individuals seeking groups with complementary attributes, i.e., they want to form a team, as opposed to seeking groups with similar attributes, i.e., homophilic kinship. Just as different ethnicities may have different types of gangs in the same city in terms of their number, size, and stability, the same holds for the different computer servers on which online players play a given game. Our quantitative results provide a novel addition to the group formation debate by being (i) able to reproduce the quantitative features of both the dynamical and time-averaged behavior observed in the empirical datasets, (ii) plausible in terms of the individual-based rules that are used to describe group membership, (iii) robust in terms of its insensitivity to small perturbations in the model's specification and parameter values, (iv) the number of free parameters in the model is kept to a minimum given the available sets of data to be modelled, and (v) able to shed light on what mechanistic rules drive people to join and leave such groups in offline and online situations and provide the basis for further investigations.

This close relationship between gangs and guilds might be less surprising if it were true that both are populated by a similar sector of society. However this is not the case. Online games are played equally by men and women across all age groups, locations and backgrounds [172–174], while gangs are mostly populated by teenage urban males from particular backgrounds [177]. Instead, we believe that our results demonstrate a commonality in the *way* in which humans form such offline and online groups. Interestingly this echoes recent claims by international law enforcement agencies concerning the hybrid nature of transnational gangs ("maras"), crime organizations, insurgencies and terrorist groups, whose interactions and activities are now beginning to blur the boundaries between real and virtual spaces [177].

Finally we note that this work throws up the interesting challenge of providing analytic solutions to accompany the empirical findings and numerical simulations. Work is proceeding in this direction, though the difficulty of including internal degrees of freedom, i.e., the model attributes, in general coalescence-fragmentation problems is daunting.

CHAPTER 6

Effect of Global Interactions

6.1 Motivation

A CO_2 emissions level of $\leq$ 500 ppm [192] would set the probability of a potentially catastrophic 5°C warming at 3% [192]. At a recent G8 summit, leaders agreed to "strongly consider" at least halving global emissions by 2050 [193]. However, there is still no national or international consensus on how these reductions can be systematically achieved and maintained [194], nor is there any deep quantitative understanding of the trade-offs which could arise at the local and global level. Given the recent instabilities in global financial markets and apparent inevitability of human irrationality [195], it is also unclear whether a free-market approach can ever be trusted [196].

Here we analyze a simple, yet realistic dynamical model of a competitive emissions market which allows us to investigate the simultaneous interplay between myriad competing real-world factors. Our model is a non-trivial generalization of the El Farol bar problem [197] which has attracted much attention among physicists [55,198,199]. In addition to offering the physics community a novel generalization and application of the El Farol model, we believe that our work provides the first unified, quantitative discussion of the underlying trade-offs between average emissions, instantaneous peak

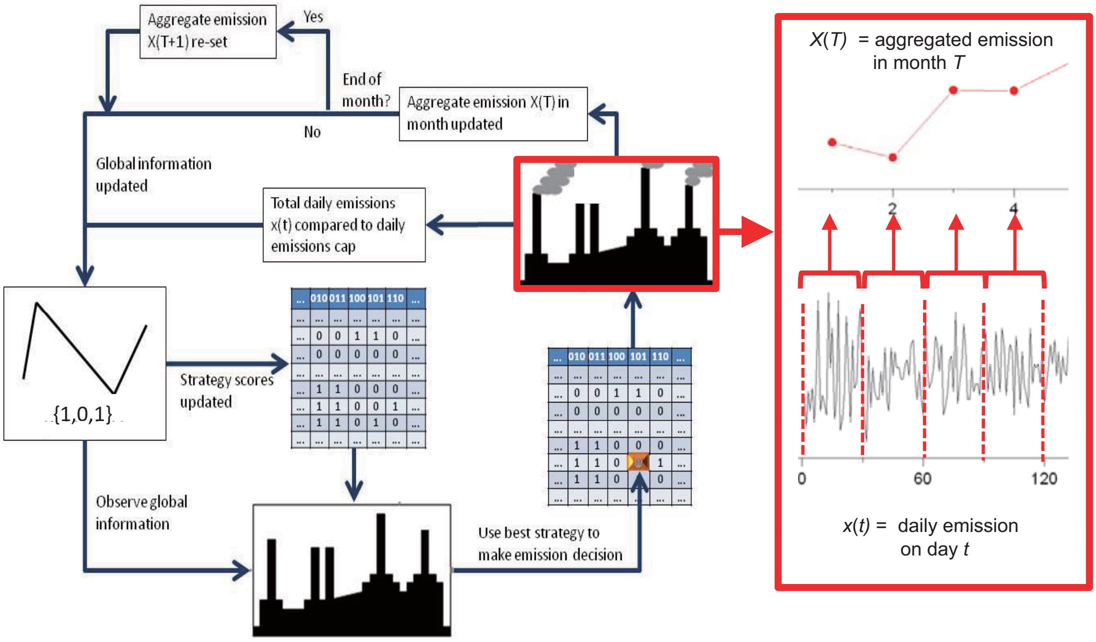


Figure 6.1: (Color) Schematic diagram of the carbon market model and the resulting daily time series $x(t)$ for day $t = 1, 2, \ldots,$ together with the aggregated monthly time-series $X(T)$ for month $T = 1, 2, \ldots.$

pollution levels, market stability, efficiency of production, and common information. Our model predicts that a completely self-organized emissions market with collective competition and no top-down management, can offer distinct advantages over a managed system in terms of peak emission values. Although helpful with respect to the mean monthly emission, top-down monthly management can by contrast induce a far bigger volatility and hence aggravate the uncertainty in emissions.

6.2 Our emission model

Figure 6.2 shows a schematic describing our generic emissions scenario, comprising N emitters (e.g. companies) who each decide whether to emit or not during a particular timestep t (e.g. day). All companies are assumed to have the same emission capabilities (i.e., one unit of carbon each timestep). The system's (e.g. national) safe emission level is L over some period Δt (e.g. month $\Delta t = 30$), with successive periods (e.g. months) labelled $T = 1, 2, \ldots$. Hence the average emission cap per timestep (e.g. day) is $\bar{L} = L/\Delta t$. Depending on the top-down management infrastructure of interest, the emitters could equally well be industries within a sector, companies within a state, states within a country, countries within a continent, or countries or continents within some global organization – likewise, the relevant timescales t and T need not be days and months respectively. From a governmental perspective, the ideal outcome would be that the total emission each month $X(T)$ is exactly equal to L units of carbon pollutants: If $X(T) > L$ then too much carbon dioxide is emitted into the atmosphere, while $X(T) < L$ means that the nation has wasted some of its allowed production capacity [194]. Companies are rewarded in some generic way (e.g. favorable public opinion, or a monetary compensation) for choosing to emit on low-pollution days ($x(t') \leq \bar{L}$) or abstaining from emitting on high-pollution days ($x(t') > \bar{L}$), and receive punishments otherwise. Each day's outcome is represented in terms of its collective emission: 1 if $x(t') \leq \bar{L}$ for a given t' and 0 if $x(t') > \bar{L}$. Companies rely on common, publicly disclosed information when deciding whether or not to emit at a given timestep. We take this common information to be dominated by the previous m days' outcomes, a bit-string of length m compromised by 0 or 1, but in principle it could include other information from government, public or other competitors. The fact that all participants have access to, and use, the same information, can generate correlations between their actions. A strategy is a specific prediction 0 or 1 (and hence action, emit or don't emit) for each of the 2^m possible information bit-strings, hence there are 2^{2^m} strategies. Companies randomly select s strategies from the strategy space with repetitions allowed during the assignment. Each company uses its best performing strategy at a given timestep, with an individual strategy's score updated by $+1$ (-1) at a given timestep, if it would have made the correct (incorrect) decision. The correct decisions are emitting (not emitting) when the cap is not exceeded (exceeded), and vice versa for incorrect decisions. Tied best-performing strategies are broken by random choices. Our setup therefore incorporates the generic complex system features of Arthur's El Farol problem and Challet and Zhang's binary version [55, 197–199]. Most importantly, companies do not communicate directly among themselves, nor do they need to know the number of competitors around, nor are they managed by some governmental entity. Instead, by competing to emit, they interact through the common information that their collective actions create. There is recent independent evidence that groups of human do indeed employ such general decision-based mechanisms as in Fig. 6.2 [200]. More generally, our model mimics a simple cap-and-trade scenario in which emitters who decide to emit on a given day immediately purchase a permit to do so. The less emitters per day, the lower the demand for permits, and hence the lower that day's permit price, and vice versa. Hence the time-series of emissions $x(t)$ mimics the time-series of permit prices.

Figure 6.2 shows that this simple model is capable of reproducing the highly irregular, non-Gaussian distribution of the 2009 EU carbon market to date. The model parameters and their values have a direct and reasonable interpretation: $N = 100$ suggests that the actions of approximately one hundred entities (e.g. large companies) is moving the market in 2009, and hence visibly impacting the overall emissions; $m = 4$ suggests that just less than one week of prior outcomes is considered relevant for making a decision; $s = 6$ suggests that individual entities are using approximately six strategies to make a decision about whether to buy a permit and hence emit. Choosing to emit is equivalent to buying a permit and using it on that day – if less

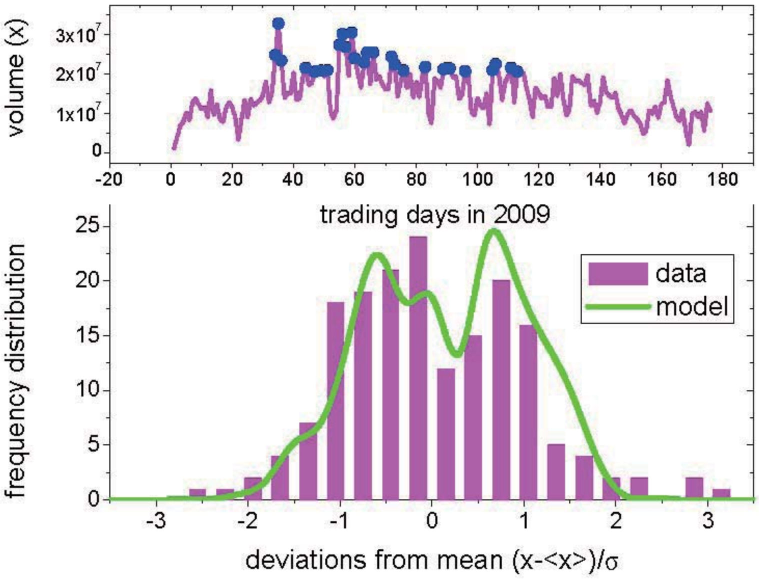


Figure 6.2: (Color) The upper panel is the daily time series of the first 176 trading days' total volume in 2009 from ECX EUA futures contract, in tonnes of CO_2 (EU allowances). The lower panel shows the empirical frequency distribution of $(x - \langle x \rangle)/\sigma)$, as compared to our model $m = 4$, averaged over 100 runs. Our model is a good representation of the empirical data, but has the added advantage that it eliminates the extreme peaks observed in the EU market (indicated by circles in the top panel). This suggests that our unmanaged free market mechanism would provide tighter control than the existing EU market.

people apply on a given day, the permit price is low which means that the time-series of the number of emitters and the price mimic each other. Hence as a surrogate of the actual daily emissions, we have taken the daily carbon price to represent the daily demand for permission to emit, and hence the resulting volume of emissions. The quantity displayed, $(x - \langle x \rangle)/\sigma$, is independent of the number of participants N, for large N. We note that our model shows a smaller occurrence of extreme events than the empirical data, suggesting that our competitive, self-organized setup might provide better control of large fluctuations than the present EU scheme which is operating. If the distribution were more Gaussian-like as in regular financial markets, this would suggest that the market should contain many noisy speculators – however, the multi-modal form in Fig. 6.2 implies that this is not the case.

Figure 6.2 compares the predictions of our model for monthly emissions between an unmanaged (red curve) and managed (blue curve) system, as a function of the amount of common information about previous outcomes (i.e. m). The average daily emissions cap is $\bar{L} = 60$. In the managed system, at the end of month T, the government will reduce or increase the emissions capacity $L(T+1)$ for month $T+1$ by the amount that the aggregated emissions $X(T)$ was above or below $L(T)$. In the unmanaged system, there is no such external control and hence L is constant. The overall system performance can be assessed through the time-series for monthly emissions $X(T)$ (Fig. 6.2): In particular (top to bottom in Fig. 6.2) the mean $\langle X \rangle$, the maximum $\max(X)$ (where $\max(X)$ is the largest monthly emission value during the time-window of the numerical simulation) over some fixed period (e.g. a year), and the standard deviation (i.e. volatility $\sigma(X)$) about the mean. An interesting comparison system is obtained by considering the "random" case of an unmanaged system in which companies decide to emit by tossing a coin each day. In the absence

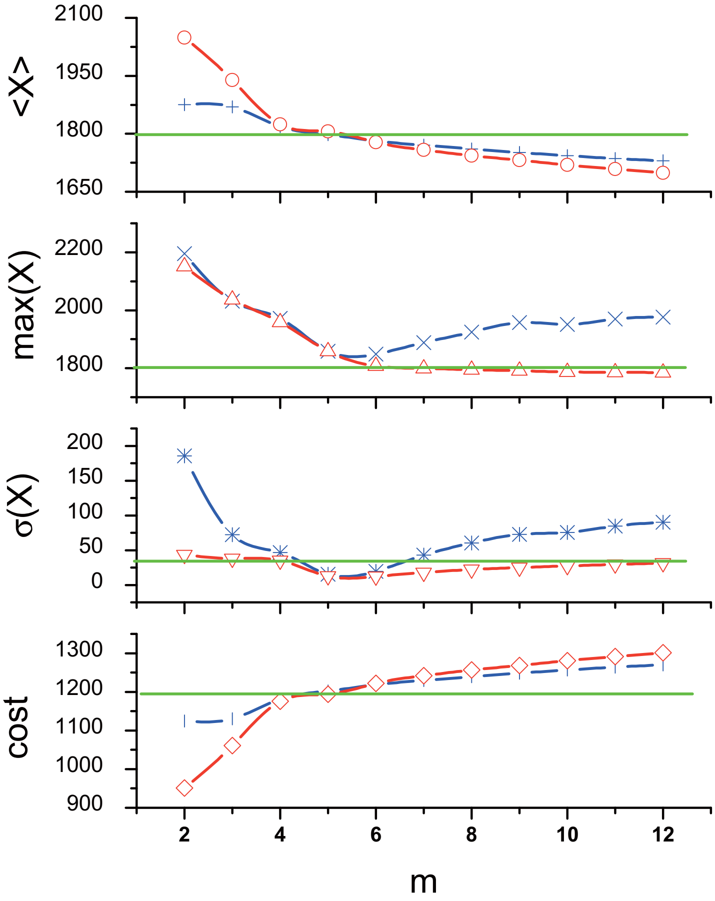


Figure 6.3: (Color) Monthly emissions for our model, for $N = 100$, $s = 6$. Top to bottom: the mean monthly emission $\langle X \rangle$, the maximum (i.e. peak) monthly emission, the monthly volatility $\sigma(X)$, and the governmental cost of compensation to companies for not emitting. Red: unmanaged system. Blue: managed system. Green: random result for learning $p = \bar{L}/N$.

of any learning (i.e. the system is non-adaptive) every decision is an independent coin toss and hence the a priori probability to emit would be $p = 0.5$. However if the entities are gradually able to learn from the feedback of the previous experience and adapt to the ideal ratio $\bar{L}/N$ (at least, at the collective level) then $p = \bar{L}/N$ yielding the green curves shown in Fig. 6.2.

The mean monthly emission $\langle X \rangle$ decreases monotonically as m increases for both systems. The monthly control exerted in the managed system pulls the value closer to the capacity limit of 1800 than for the unmanaged system. However, this improved performance due to top-down management is accompanied by a significantly higher volatility for $m > 6$ as well as a significantly higher peak pollution level. Indeed, the managed system does *worse* than both the unmanaged system *and* the random system with learning. This is because the month-by-month adjustment to $\bar{L}$ induces a delayed oscillatory effect which in turn generates significant volatility. A transition occurs around $m \sim 4$ where all the curves seem to cross the green (i.e. random learning) curve. This m value coincides with the system's dynamical de Bruijn path (which has duration 2.2^m [201]) becoming equal to the finite duration of the emission interval (i.e. 30 days, hence $2.2^m \sim 30$ which yields $m \sim 4$). This precedes a minimum in the volatility around $m \sim 5$ for both managed *and* unmanaged systems, which is smaller than for random learning. As for the El Farol problem [55, 197–199], this unintentional collective cooperation emerges as a result of cancellation between the actions of crowds of emitters using one strategy, and anticrowds using the exact opposite strategy. The cost result (bottom panel) reflects a simple one-unit payout given to any company not emitting on a given day. The choice of emitting or not-emitting becomes essentially cost-neutral to a given company – however for public relations reasons, and because they want to stay active in business, each company still continues to compete. A higher $\langle X\rangle$ hence incurs a lower cost.

Figure 6.2 shows the model's daily emission (Fig. 6.2(a), crosses) and volatility (Fig. 6.2(b), crosses) as a function of the daily emissions cap $\bar{L}$. The red shaded area in Figs. 6.2(a) and (b) is the "learning zone" bounded by the two analytically obtained limits of no learning ($p = 0.5$, the probability that a company is going to emit at a timestep, horizontal red line) and learning ($p = \bar{L}/N$, red diagonal line in Fig. 6.2(a) and convex curve in Fig. 6.2(b)). The standard deviation for daily emissions in the random case, is given analytically by the usual binomial form, i.e. $[Np(1-p)]^{\frac{1}{2}}$. Using the lower bound value $p = \bar{L}/N$ yields the convex curve $[\bar{L}(1-\bar{L}/N)]^{\frac{1}{2}}$, while using the upper bound value $p = 0.5$ yields the horizontal line $0.5N^{\frac{1}{2}} = 5$. The model's mean emission values (crosses) lie within the shaded area in Fig. 6.2(a) but are closest to the limit $p = \bar{L}/N$, thereby demonstrating that the unmanaged, self-organized market collectively learns. For intermediate $\bar{L}$ values (Fig. 6.2(b)) the corresponding volatility tends to be smaller than the random value, however it moves above it for very large or small $\bar{L}$. For small m values, which corresponds to the crowded regime of the strategy space, numerical runs can show significantly large volatilities (green circle).

Figure 6.2(c) explores the implications of our results for the *derivative* emissions markets. If emission markets follow the path of the mature non-emission financial markets, it is likely that such derivatives (e.g. options) markets will become as large, or even larger, than the primary emissions market itself [55]. In this respect, our findings serve as a warning of the dangers of simply applying standard financial theory for such derivative instruments [55]. Standard option pricing theory uses the volatility over a given time increment as the input to the Black-Scholes pricing formula [202].

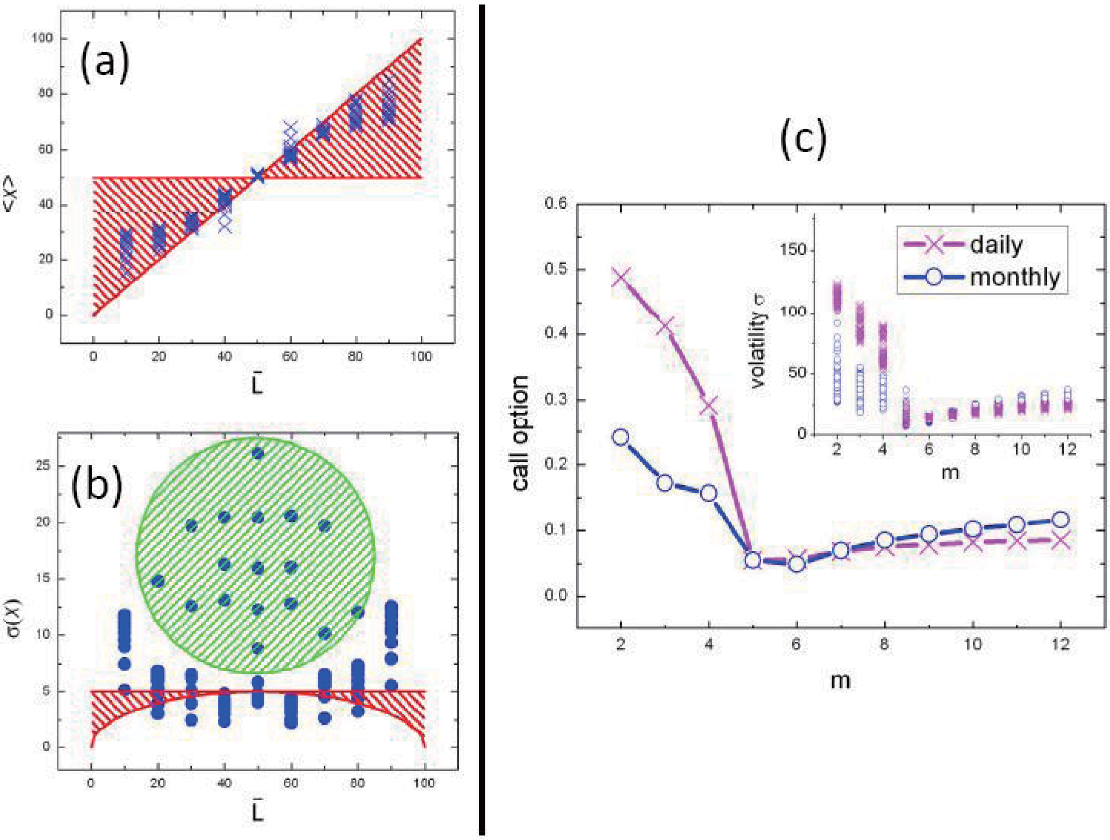


Figure 6.4: (Color) Daily emissions for our model with m varying from 2 to 12 ($s = 6$). Red shaded region is the analytically obtained zone of learning (see text). (a) Mean emission. Horizontal boundary line corresponds to no learning, diagonal boundary line (slope of unity) is for learning. (b) Volatility. Horizontal boundary line corresponds to no learning, convex curve is for learning. Green shaded circle shows low m, high volatility region due to crowding in strategy space. (c) European call option prices for different measurements of volatility according to the standard derivative pricing theory (i.e. Black-Scholes equation). Risk-free interest rate $r = 0$, current value x is set to the individual mean divided by 100, and strike price X_s equals $\bar{L}/100$. The volatility σ is scaled by $1/\sqrt{100}$. The time is one day before expiration. Blue curve uses the results from monthly measurement, while purple one is from daily measurement. The insert shows the anomalous scaling of volatility of emissions, from daily to monthly scales. There are 64 runs for each m. The purple crosses are the standard deviation calculated by taking the daily volatility and multiplying by $\sqrt{30}$, which would be *exactly* equal to the monthly volatility if the time-series followed a random walk.

This assumes that the market approximates to a random walk and hence that the monthly volatility over Δt timesteps is simply the volatility over one timestep multiplied by $\sqrt{\Delta t}$. Figure 6.2(c) shows not only that this is incorrect (see inset), but also that the discrepancy depends on the amount of common information m – and that as a consequence, the price of a call option (Fig. 6.2(c)) can be mispriced according to whether daily or monthly volatility estimates are used. This opens up an intriguing but dangerous situation in the event of any abnormal periods in the market: Following some external news event (e.g. collapse of an oil company), it may happen that the number of previous days' outcomes that are thought relevant, becomes very small (i.e. $m \to 0$). As shown, the corresponding mispricing then becomes huge, leading to possible financial instabilities.

6.3 Formal analysis of our emissions model

It remains a fascinating and important practical challenge to produce analytic expressions in close quantitative agreement with the preceeding plots, especially about the volatility. However, we provide in this section a formal platform on which such analytic expressions can be built. In particular, this section serves to (1) formalize the underlying model structure, connecting it to well-known ideas in computational science such as Hamming distances and binary look-up tables, and (2) develop analytic expressions for the special case where the resource level is exactly one half the number of competing agents N. In this case, the problem maps onto the so-called Minority Game [198]. However, far from being a very isolated special limit, we note that our numerical simulations show that similar qualitative results are obtained for *any* value of resource level L which is not too extreme (i.e. $L \sim N/2$). Hence the analytic analysis carried out in this paper serves to take us tantalizingly close to a full analytic expression for the emissions problem.

In order to develop neat closed-form expressions, we will make the notation symmetric, by mapping action $0 \to -1$ and $1 \to +1$. The emissions cap L is also set equal to $(N-1)/2$ (assuming N is odd, as is the case in the present simulations) in order to enable analytic progress, although we stress that our simulations show similar qualitative behaviors for general L values as long as L is not too close to 0 or N. We will focus on calculating the daily volatility, since arguably one of the most striking results in our earlier figures is the dependence of the fluctuations on the individual agent memory size m. At timestep t, each agent (i.e. emitting company) decides whether to emit (1 under the previous notation, which now becomes $+1$) or not (0 under the previous notation, which now becomes -1). Recall that the global information available to the agents is a common memory of the most recent m outcomes, which are represented as either 0 (e.g. emissions below capacity L) or 1 (e.g. emissions above capacity L). Hence this outcome history is represented by a binary bit-string of length m. For general m, the system will therefore move within the space of $P = 2^m$ possible history bit-strings. These history bit-strings can alternatively be represented in decimal form: $\mu = \{0, 1, ..., P-1\}$. For $m = 2$, for example, $\mu = 0$ corresponds to 00, $\mu = 1$ corresponds to 01 etc. A strategy consists of a predicted action, -1 or $+1$, for each possible history bit-string. Hence there are $2^{P=2^m}$ possible strategies.

Figure 6.5 shows the history space from Figure 6.2, for the case of individual agent memory $m = 2$, while Fig. 6.6 shows the corresponding strategy space. Figure 6.5 is familiar to computer scientists as De Bruijn graphs, while Fig. 6.6 resembles a look-up table (Fig. 6.6, left panel) and a Boolean hypercube which is typically used to discuss

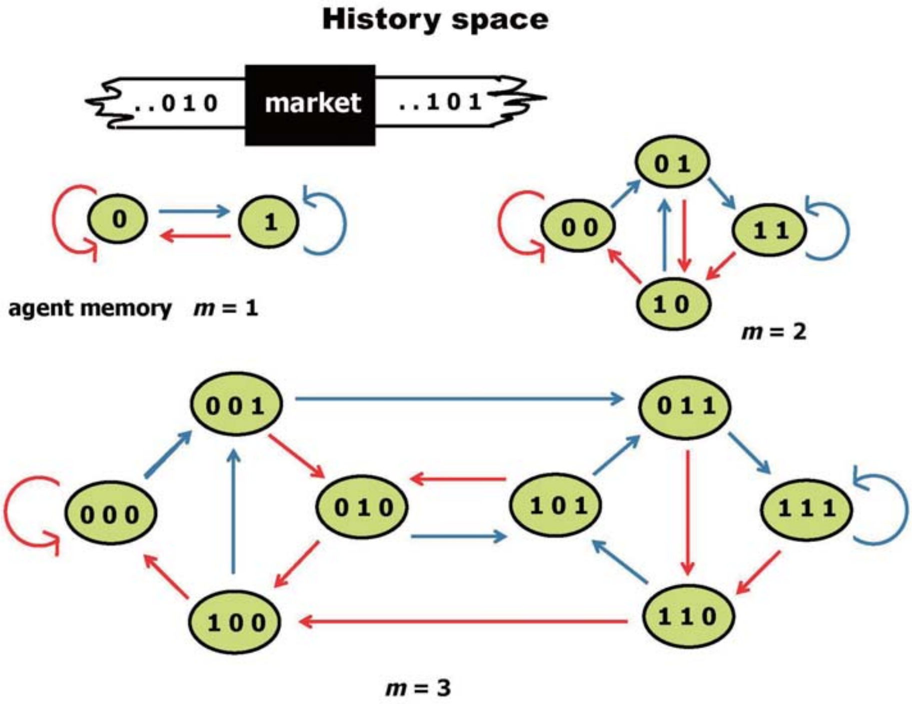


Figure 6.5: History Space for agent memory $m = 2$. Examples shown for $m = 1$, $m = 2$ and $m = 3$. The history space in each case takes the form of a De Bruijn graph. Only certain transitions are allowed because of the finite number of ways in which the global history can change from timestep t to timestep $t + 1$.

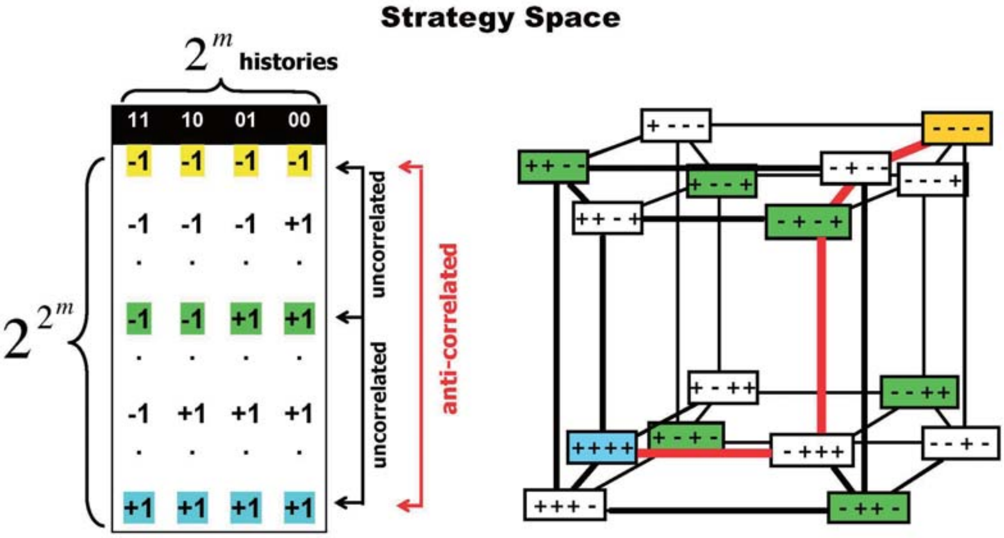


Figure 6.6: Left: the strategy space shown for $m = 2$, together with example strategies. This strategy space shown is called the Full Strategy Space (FSS) and it contains all possible permutations of the actions -1 and $+1$ for each history. There are 2^{2^m} strategies in the FSS. Right: the 2^m dimensional hypercube shows all 2^{2^m} FSS strategies at its vertices. The shaded strategies span a Reduced Strategy Space RSS, which has $2.2^m = 2P$ strategies. The thick line connects two strategies whose Hamming distance separation is 4.

Hamming distances (Fig. 6.6, right panel). Each strategy is a set of instructions to describe what action an agent should take, given any particular history μ. The strategy space is the set of strategies from which agents are allocated their strategies. The strategy space shown is known as the Full Strategy Space (FSS) and contains all possible permutations of the actions -1 and $+1$ for each history. As such there are 2^{2^m} strategies in this space. One can choose a subset of 2.2^m strategies, called a Reduced Strategy Space (RSS), such that any pair within this subset has one of the following two characteristics: (i) *Anti-correlated.* For example, any two agents using the $(m = 2)$ strategies $(-1, -1, +1, +1)$ and $(+1, +1, -1, -1)$ respectively, would take the opposite action irrespective of the sequence of previous outcomes and hence the history. If we consider the excess demand $D[t] = n_{+1}[t] - n_{-1}[t]$ at timestep t, where $n_{+1}[t](n_{-1}[t])$ is the total number of agents choosing action $+1(-1)$ at timestep t, then their net effect on the excess demand cancels out at each timestep, regardless of the history. Hence they will not contribute to fluctuations in $D[t]$. We note that alternatively we could define the excess demand as the deviation from L, however since we are focusing on $L = (N - 1)/2$ the size of the fluctuations are equivalent in the two cases, apart from an unimportant constant factor. (ii) *Uncorrelated.* For example, any two agents using the strategies $(-1, -1, -1, -1)$ and $(-1, -1, +1, +1)$ respectively, would take the opposite action for two of the four histories, and the same action for the remaining two histories. If the histories occur equally often, the actions of the two agents will be uncorrelated on average. Note that the strategies in the RSS can be labeled by $R = \{1, 2, ..., 2P = 2.2^m\}$. A tensor Ω can be used to describe the strategy allocation among the N individual agents. This strategy allocation acts as a quenched disorder in the system since it is typically fixed from the beginning of the game. Its rank is given by the number of strategies S that each agent holds.

Many possible “microstates” describing the specific partitions of strategies among the agents, correspond to a single Ω “macrostate”. For this reason, the formalism that we are now presenting, if retained at the level of a given Ω, can describe the set of all games which belong to that same Ω macrostate. We also note that Ω can be made symmetric since the agents do not distinguish between the order in which the two strategies are picked. For this reason, we henceforth focus on $S = 2$ and employ a symmetrized version of the strategy allocation matrix: $\Psi = \frac{1}{2}(\Omega + \Omega^T)$.

We now show that the the variation in the size of the fluctuations of the emissions, as shown in the numerical results of this paper, can vary significantly as the memory m changes – and that this inter-relation can be understand as a competition between a crowd and its anticrowd. Consider an arbitrary timestep t during a run of the game. The excess demand $D[t] \equiv D\left[\underline{S}[t], \mu[t]\right] = n_{+1}[t] - n_{-1}[t]$, where $\underline{S}[t]$ is the $2P$-dimensional score-vector whose R'th component is the virtual point score for strategy R. We note that strategies gain/lose one virtual point at each timestep, according to whether their predicted action would have been a winning/losing action. At timestep t, the current history is $\mu[t]$. Hence for this given run, the standard deviation of $D[t]$ is given by a time-average for a given set of initial conditions and a given realization of Ψ. We next introduce the quantity $a_R^{\mu[t]} = \pm 1$, which is the action predicted by strategy R in response to the history bit-string μ at time t. Carrying out a sum over the RSS yields: $D\left[\underline{S}[t], \mu[t]\right] = \sum_{R=1}^{2P} a_R^{\mu[t]} n_R^{\underline{S}[t]}$. Here the quantity $n_R^{\underline{S}[t]}$ counts the number of agents at time t using strategy R. We use the notation $\langle X[t] \rangle_t$ to denote a time-average over the variable $X[t]$ for a given Ψ. Therefore

$$\left\langle D\left[\underline{S}[t], \mu[t]\right]\right\rangle_t = \sum_{R=1}^{2P} \left\langle a_R^{\mu[t]} n_R^{\underline{S}[t]} \right\rangle_t = \sum_{R=1}^{2P} \left\langle a_R^{\mu[t]} \right\rangle_t \left\langle n_R^{\underline{S}[t]} \right\rangle_t$$

where we have employed the property that $a_R^{\mu[t]}$ and $n_R^{\underline{S}[t]}$ are uncorrelated. We now focus on the common situation where all histories are visited equally on average, noting that even if this situation does not hold for a specific Ψ, it may indeed hold once the averaging over Ψ has also been taken. For example, in the present case with $L = (N-1)/2$, all histories are visited equally at small m and a given Ψ. If we take the additional average over all Ψ, then the same is also true for large m. Using this property of equal histories, we obtain:

$$\begin{aligned}
\langle D\left[\underline{S}[t], \mu[t]\right]\rangle_t &= \sum_{R=1}^{2P}\left(\frac{1}{P}\sum_{\mu=0}^{P-1} a_R^{\mu[t]}\right)\left\langle n_R^{\underline{S}[t]}\right\rangle_t \\
&= \sum_{R=1}^{P}\left(\frac{1}{P}\sum_{\mu=0}^{P-1} a_R^{\mu[t]} + a_{\overline{R}}^{\mu[t]}\right)\left\langle n_R^{\underline{S}[t]}\right\rangle_t = \sum_{R=1}^{P} 0.\left\langle n_R^{\underline{S}[t]}\right\rangle_t \\
&= 0
\end{aligned} \tag{6.1}$$

where we have used the exact result $a_R^{\mu[t]} = -a_{\overline{R}}^{\mu[t]}$, together with the approximate result $\left\langle n_R^{\underline{S}[t]}\right\rangle_t = \left\langle n_{\overline{R}}^{\underline{S}[t]}\right\rangle_t$. This is reasonable since there is no a priori best strategy typically. When the strategies are distributed reasonably evenly over the agent population, the average number of agents playing each strategy becomes approximately equal and hence $\left\langle n_R^{\underline{S}[t]}\right\rangle_t = \left\langle n_{\overline{R}}^{\underline{S}[t]}\right\rangle_t$. In the event that all histories are not equally visited over time, even after averaging over all Ψ, it may still happen that the system's dynamics is restricted to equal visits to some *subset* of histories. In this case one can then carry out the averaging in Equation (6.1) over this subspace of histories. We note that more generally, the averagings which arise in this formalism can be carried out with appropriate frequency weightings for each history – indeed, we can incorporate any non-ergodic dynamics provided we know the appropriate history path.

The excess demand $D[t]$ has a variance given by

$$\sigma_\Psi^2 = \left\langle D\left[\underline{S}[t], \mu[t]\right]^2\right\rangle_t - \langle D\left[\underline{S}[t], \mu[t]\right]\rangle_t^2 \ . \tag{6.2}$$

We make the reasonable assumption that the game output is unbiased and therefore $\langle D\left[\underline{S}[t],\mu[t]\right]\rangle_t = 0$. Therefore

$$\sigma_\Psi^2 = \left\langle D\left[\underline{S}[t],\mu[t]\right]^2\right\rangle_t = \sum_{R,R'=1}^{2P} \left\langle a_R^{\mu[t]} n_R^{\underline{S}[t]} a_{R'}^{\mu[t]} n_{R'}^{\underline{S}[t]} \right\rangle_t .$$

We now employ the exact identities $\underline{a_R}.\underline{a_{R'}} = P$ (fully correlated), $\underline{a_R}.\underline{a_{R'}} = -P$ (fully anti-correlated), and $\underline{a_R}.\underline{a_{R'}} = 0$ (fully uncorrelated) where $\underline{a_R}$ is a vector of dimension P with components $a_R^{\mu[t]}$ for $\mu[t] = 1, 2, ..., P$. This yields

$$\begin{aligned} \sigma_\Psi^2 &= \sum_{R=1}^{2P} \left\langle \left(n_R^{\underline{S}[t]}\right)^2 - n_R^{\underline{S}[t]} n_{\overline{R}}^{\underline{S}[t]} \right\rangle_t + \sum_{R \neq R' \neq \overline{R}}^{2P} \left\langle a_R^{\mu[t]} a_{R'}^{\mu[t]} \right\rangle_t \left\langle n_R^{\underline{S}[t]} n_{R'}^{\underline{S}[t]} \right\rangle_t \\ &= \sum_{R=1}^{2P} \left\langle \left(n_R^{\underline{S}[t]}\right)^2 - n_R^{\underline{S}[t]} n_{\overline{R}}^{\underline{S}[t]} \right\rangle_t . \end{aligned} \tag{6.3}$$

We can equivalently write the sum over $2P$ terms as a sum over P terms,

$$\begin{aligned} \sigma_\Psi^2 &= \sum_{R=1}^{P} \left\langle \left(n_R^{\underline{S}[t]}\right)^2 - n_R^{\underline{S}[t]} n_{\overline{R}}^{\underline{S}[t]} + \left(n_{\overline{R}}^{\underline{S}[t]}\right)^2 - n_{\overline{R}}^{\underline{S}[t]} n_R^{\underline{S}[t]} \right\rangle_t \\ &= \sum_{R=1}^{P} \left\langle \left(n_R^{\underline{S}[t]} - n_{\overline{R}}^{\underline{S}[t]}\right)^2 \right\rangle_t \equiv \left\langle \sum_{R=1}^{P} \left(n_R^{\underline{S}[t]} - n_{\overline{R}}^{\underline{S}[t]}\right)^2 \right\rangle_t . \end{aligned}$$

For each R, the actual values of $n_R^{\underline{S}[t]}$ and $n_{\overline{R}}^{\underline{S}[t]}$ depend on the precise form of Ψ. We denote the ensemble-average over all possible realizations of the strategy allocation matrix Ψ, as $\langle ... \rangle_\Psi$. Employing the notation $\langle \sigma_\Psi^2 \rangle_\Psi = \sigma^2$, gives

$$\sigma^2 = \left\langle \left\langle \sum_{R=1}^{P} \left(n_R^{\underline{S}[t]} - n_{\overline{R}}^{\underline{S}[t]}\right)^2 \right\rangle_t \right\rangle_\Psi . \tag{6.4}$$

Analogous expressions for the variances in $n_{+1}[t]$ and $n_{-1}[t]$ can be obtained in a parallel way.

The time-averaged fluctuations in the excess demand are given by Eq. (6.4). In order to reduce Eq. (6.4) to some explicit analytic expressions, approximations must be made. Depending on the level of approximation that one is prepared to make, Eq. (6.4) can be manipulated in a variety of ways, with the precise form of any resulting analytic expression depending on the approximation's details. Here we adopt one particular approach which is well-suited to the low m regime. We start by relabelling the strategies, rewriting the sum in Eq. (6.4) to be over a *virtual-point ranking* K as opposed to R. The variation in points for a given strategy, as a function of time for a given realization of Ψ, will have a ranking (i.e. label) in terms of virtual-points score which will typically change rather rapidly in time – particularly in the low m regime of interest where strategies are continually being overturned. The specific identity of the "K"th highest-scoring strategy' therefore changes frequently in time, and hence $n_R^{S[t]}$ varies considerably in time. In order to proceed, we shift the focus onto the time-evolution of the highest-scoring strategy, second highest-scoring strategy etc., since this should have a much smoother time-evolution than the time-evolution for a given strategy. With all strategies starting off at zero points, the anticorrelated strategies appear as the mirror-image, i.e. $S_K[t] = -S_{\overline{K}}[t]$. We can denote the rank in terms of strategy score by the label K, i.e. $K = 1$ is the highest scoring strategy position, $K = 2$ is the second highest-scoring strategy position etc. with

$$S_{K=1} > S_{K=2} > S_{K=3} > S_{K=4} > \ldots \tag{6.5}$$

assuming no strategy-ties. Since $S_R = -S_{\overline{R}}$ (i.e. all strategy scores start off at zero), it follows that $S_K = -S_{\overline{K}}$. We can therefore rewrite Eq. (6.4) exactly as

$$\sigma^2 = \left\langle \left\langle \sum_{K=1}^{P} \left(n_K^{S[t]} - n_{\overline{K}}^{S[t]} \right)^2 \right\rangle_t \right\rangle_\Psi . \tag{6.6}$$

In the systems of interest the agents are typically playing their highest-scoring strategies, hence knowledge of a strategy's relative ranking is the relevant quantity for determining how many agents will instantanously play a given strategy, not the actual value of its virtual points score. The quantities $n_K^{S[t]}$ and $n_{\overline{K}}^{S[t]}$ will therefore fluctuate relatively little in time, enabling us to develop the problem in terms of time-averaged values. The number of agents playing the strategy in position K at any timestep t, can be rewritten as some constant value n_K plus a fluctuating term $n_K^{S[t]} = n_K + \varepsilon_K[t]$. Simulations confirm that there is indeed a suitable constant n_K such that the fluctuation $\varepsilon_K[t]$ represents a small noise term. Therefore

$$\begin{aligned}\sigma^2 &= \left\langle \sum_{K=1}^{P} \left\langle [n_K + \varepsilon_K[t] - n_{\overline{K}} - \varepsilon_{\overline{K}}[t]]^2 \right\rangle_t \right\rangle_\Psi \qquad (6.7)\\ &\approx \left\langle \sum_{K=1}^{P} \left\langle [n_K - n_{\overline{K}}]^2 \right\rangle_t \right\rangle_\Psi = \left\langle \sum_{K=1}^{P} [n_K - n_{\overline{K}}]^2 \right\rangle_\Psi .\end{aligned}$$

We have assumed the noise terms average out to be small – as with all our assumptions, we have checked that it is well supported by the numerical simulation. Taking Ψ inside the sum, each term can be rewritten exactly using the joint probability distribution for n_K and $n_{\overline{K}}$, which we shall refer to as $P(n_K, n_{\overline{K}})$. Therefore

$$\sigma^2 = \sum_{K=1}^{P} \left\langle [n_K - n_{\overline{K}}]^2 \right\rangle_\Psi = \sum_{K=1}^{P} \sum_{n_K=0}^{N} \sum_{n_{\overline{K}}=0}^{N} [n_K - n_{\overline{K}}]^2 \, P(n_K, n_{\overline{K}}).$$

Consider Eq. (6.8) in the limiting case where the averaging over the quenched disorder matrix is dominated by matrices Ψ which are nearly flat. This is a good approximation in the "crowded" limit of small m in which there are many more agents than available strategies, since in this regime the standard deviation of an element in Ψ (i.e. the bin-size standard deviation) will be much smaller than the mean bin-size. As a result, the probability distribution $P(n_K, n_{\overline{K}})$ is sharply peaked around the n_K and $n_{\overline{K}}$ values given by the mean values for a flat quenched-disorder matrix Ψ. Defining these mean values as $\overline{n_K}$ and $\overline{n_{\overline{K}}}$, yields $P(n_K, n_{\overline{K}}) = \delta_{n_K, \overline{n_K}} \delta_{n_{\overline{K}}, \overline{n_{\overline{K}}}}$ and therefore

$$\sigma^2 = \sum_{K=1}^{P} \left[\overline{n_K} - \overline{n_{\overline{K}}}\right]^2 . \qquad (6.8)$$

Equation (6.8) has a very simple interpretation, as the sum of the variances for each Crowd-Anticrowd pair. There is an anticorrelated strategy $\overline{K}$ for any given strategy

K. Hence there are $\overline{n_K}$ agents using strategy K who are doing the opposite of the $\overline{n_{\overline{K}}}$ agents using strategy $\overline{K}$, regardless of the history bit-string. Therefore the effective group-size for each Crowd-Anticrowd pair is $n_K^{eff} = \overline{n_K} - \overline{n_{\overline{K}}}$, which represents the net step-size d of the Crowd-Anticrowd pair in a random-walk contribution to the total variance. The net contribution of this Crowd-Anticrowd pair to the variance is therefore given by

$$[\sigma^2]_{K\overline{K}} = 4pqd^2 = 4.\frac{1}{2}.\frac{1}{2}[n_K^{eff}]^2 = [\overline{n_K} - \overline{n_{\overline{K}}}]^2 \tag{6.9}$$

where for a random walk we know that $p = q = 1/2$. All the strong correlations have been included (i.e. anti-correlations), therefore we can assume that the separate Crowd-Anticrowd pairs execute random walks which are uncorrelated with respect to each other. This follows since the properties of the RSS guarantee that all the remaining strategies are uncorrelated. The total variance is given by the sum of the individual variances,

$$\sigma^2 = \sum_{K=1}^{P} [\sigma^2]_{K\overline{K}} = \sum_{K=1}^{P} [\overline{n_K} - \overline{n_{\overline{K}}}]^2 , \tag{6.10}$$

which therefore corresponds exactly to Equation (6.8). In the event that there are frequent strategy-ties, we have to be more careful about evaluating $\overline{n_K}$ because the tie-breaking rule may affect its value. This effect will become quite important in the case of very small m in the presence of network connections, since very small m naturally leads to crowding in strategy space and hence mean-reverting virtual scores for strategies. The mean-reversion will be amplified further by the presence of network connections which increase the crowding, thereby increasing the chance of strategy ties.

For the important limiting case of small m, we now evaluate the Crowd-Anticrowd expressions. Crowding effects will be important since there are many more agents

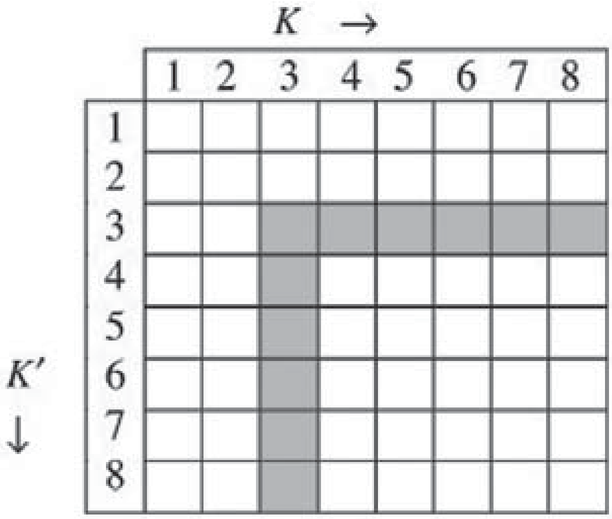


Figure 6.7: Strategy allocation matrix Ψ with $m = 2$ and $S = 2$, in the RSS, shown schematically. The strategies are ranked according to strategy score, and are labelled by the rank K. When Ψ is essentially flat, the number of agents playing the K'th highest-scoring strategy will just be proportional to the number of shaded bins at that K.

than available strategies. Each element of Ψ has a mean of $N/(2P)^S$ agents per "bin". The fluctuations in the number of agents per bin will be small compared to this mean value in the case of small m and hence densely-filled Ψ – therefore Ψ will look uniform or "flat" in terms of the occupation numbers in each bin. Figure 6.7 provides a schematic representation of Ψ in the RSS with $m = 2$, $S = 2$. When Ψ is flat, any re-ordering due to changes in the strategy ranking has no effect on its form, hence the number of agents playing the K'th highest-scoring strategy will always be proportional to the number of shaded bins at that K (see Fig. 6.7 for $K = 3$). We find for general m and S:

$$\begin{aligned}\overline{n_K} &= \frac{N}{(2P)^S}[S(2P-K)^{S-1}+\frac{S(S-1)}{2}(2P-K)^{S-2}+...+1] \qquad (6.11)\\ &= \frac{N}{(2P)^S}\sum_{r=0}^{S-1}\frac{S!}{(S-r)!r!}[2P-K]^r\\ &= \frac{N}{(2P)^S}([2P-K+1]^S-[2P-K]^S)\\ &= N.\left(\left[1-\frac{(K-1)}{2P}\right]^S-\left[1-\frac{K}{2P}\right]^S\right),\end{aligned}$$

with $P \equiv 2^m$. When each agent holds two strategies, i.e. $S = 2$, $\overline{n_K}$ can be rewritten as

$$\overline{n_K} = N.\left(\left[1-\frac{(K-1)}{2P}\right]^2-\left[1-\frac{K}{2P}\right]^2\right) = \frac{(2^{m+2}-2K+1)}{2^{2(m+1)}}N\ . \qquad (6.12)$$

Therefore

$$\begin{aligned}\sigma^2 &= \sum_{K=1}^{P}\left[\frac{(2^{m+2}-2K+1)}{2^{2(m+1)}}N-\frac{(2K-1)}{2^{2(m+1)}}N\right]^2 \qquad (6.13)\\ &= \frac{N^2}{2^{2(2m+1)}}\sum_{K=1}^{P}[2^{m+1}-2K+1]^2 = \frac{N^2}{3.2^m}(1-2^{-2(m+1)})\ .\end{aligned}$$

We have assumed that there are no strategy ties – in particular, we have assumed that the simulation rules governing strategy ties do not affect the identical forms of the rankings in terms of popularity and highest virtual points. As a result, we have underestimated the size of the Anticrowds using low-ranking strategies and overestimated the size of the Crowds using high-ranking strategies. Hence the analytic form for σ overestimates the numerical value, as shown in Figure 6.8. Despite this, there is remarkably good agreement between our analytic theory and the numerical results (assuming $L = (N-1)/2$ and L is time-independent, as in the Minority Game). In exactly the same way, this formal approach can be extended to deal with the important complementary regimes of a non-flat quenched disorder matrix Ψ, at small m, and a non-flat quenched disorder matrix Ψ, at large m. The resulting agreement for these regimes is also excellent, as shown in Figure 6.8. Most importantly, the analytic expressions in Fig. 6.8 capture some of the general qualitative trends seen in Figs. 6.2 and 6.2 for the full emissions simulation.

6.4 Summary

Despite recent skepticism surrounding the stability of free markets, our analysis predicts that an *unmanaged* carbon emissions market can provide significant advantages over a managed one. For a given sector, state, country or continent, our model helps identify the appropriate degree of governmental management such that annual *global* emissions targets are achieved, while simultaneously allowing for individual choice regarding the trade-off between *local* social issues as listed in the abstract. Finally, we have checked that our main conclusions are reasonably robust to different sets of parameter values.

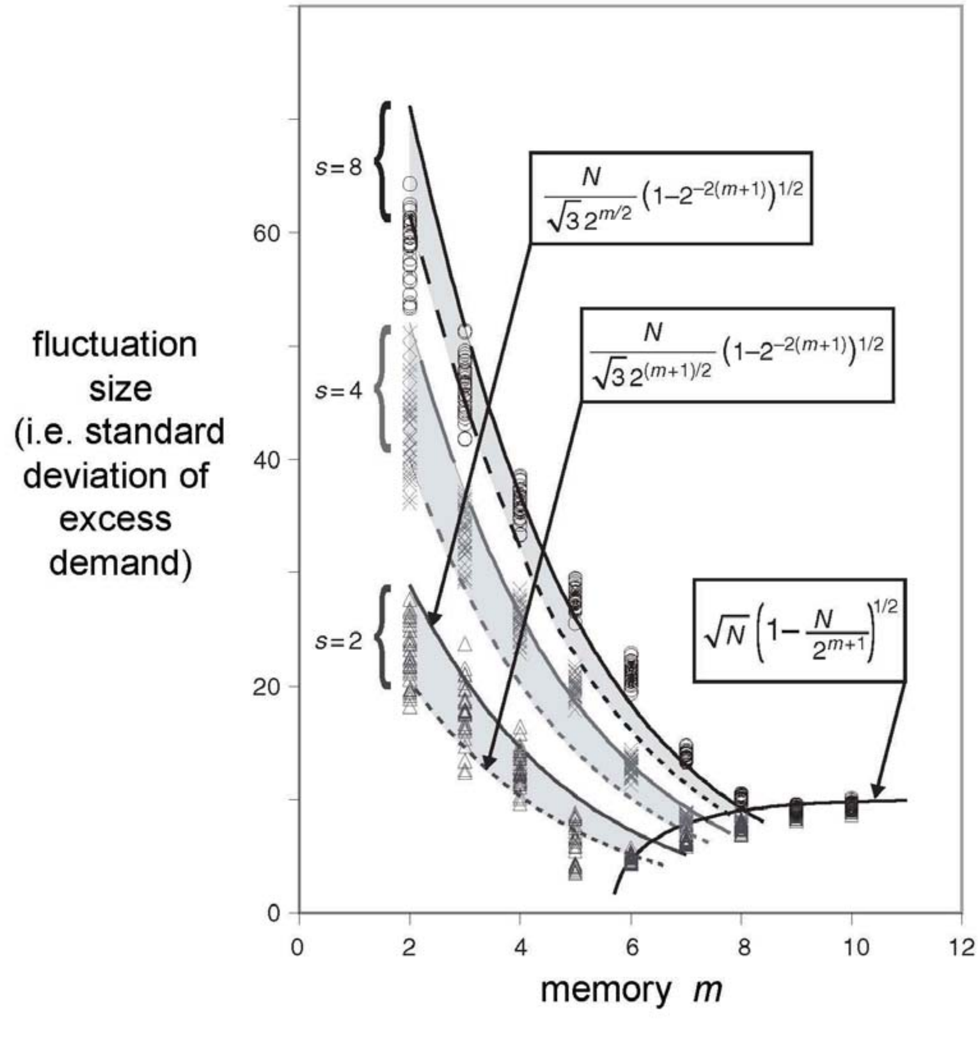


Figure 6.8: Our formal theory vs. numerical simulation results for the fluctuations σ in the emissions model with $L = (N-1)/2$ as a function of memory size m, for $N = 101$ agents, at $S = 2$, 4 and 8. Analytic forms of σ (i.e. standard deviation in excess demand) are shown at each S value. The numerical values were obtained from different simulation runs (triangles, crosses and circles). Figure adapted from Ref. [55].

CHAPTER 7

Summary and Future Work

7.1 Summary

The theoretical research presented in this thesis has made several new contributions to the topic of grouping dynamics in Complex Systems. These are summarized below:

In the first part of the thesis, we investigated a classic group-dynamics model – EZ model [28], and how changes of microscopic rules affect the macroscopic cluster-size distribution which emerges from the solution to the rate equation. We then compared this model to another social group model (Gueron and Levin's Model [54]). Our analysis pointed out the crucial role of fragmentation among these dynamical grouping models. We also extended our analytic treatment to the perturbation, and varying populations of the basic EZ model.

Based on these theoretical studies, we then applied the dynamic models to different fields, providing new aspects to the understanding and explanation of complex system phenomena in different fields. We adapted the generalized coalescence-fragmentation model to describe an insurgent population, which is qualitatively different from conventional mass-action approaches, and we were able to explain the attrition time of modern conflict. Our work suggested some new insights in the study of modern conflict/competition.

Following this line of thought, we also studied diffusion and spreading behavior in such a dynamical clustering model. In particular, we analyzed the empirical data from school colds, through to foreign exchange markets and Youtube downloads. We found that our agent model is capable of reproducing the main features in the data: long decay, and multiple resurgent peaks. We also looked at the epidemic threshold. We were able to provide an analytical expression for a novel transition that we uncovered, which in turn suggested a new strategy for controlling the spreading of a virus. We then generalized our multi-agent system to a more complicated model, by adding internal character (i.e. hidden variables) to the particles, and hence could successfully capture the quantitative features of both online game and on street gangs. This revealed several counter-intuitive common behaviors.

Apart from the coalescence-fragmentation models (i.e. local interaction models), we also studied a global interaction model based on a generalization of the El Farol bar problem [55, 197–199]. Applying the agent model, we proposed a quantitative framework for studying the emerging carbon emissions market. Our model reveals some novel underlying relationships between average emissions, instantaneous peak pollution levels, market stability, efficiency of production, and common information – and moreover highlights the potential mispricing of the options for carbon markets.

7.2 A vision for the future

As an extension of the work in this thesis, a long-term research goal would be to harness other similarly large data sets which have not been available until recently, in order to understand, predict, and ultimately enhance social and technological systems across all relevant real-world domains. In turn, this will involve crucially important collaborations with expert colleagues from different fields, as well as further testing of the core models presented here.

In the shorter term, directly related to this thesis. I think there are several promising avenues along which this thesis may be extended:

- Relate our temporal grouping dynamic models to complex networks, i.e. reproduce the characteristic parameters (properties) found in the theoretical network models [18–21] – or better, reproduce the numbers found in the real world systems.
- Combine the local grouping dynamic models and global interaction models into a multi-level and more comprehensive model, which would be more realistic and more capable of attacking the above point.
- More tools need to be developed to analyze the data and compare to the features emerging from the models. In particular, suitable measures related to burstiness and memory would be useful.
- Most importantly, I think more work could be done in each of the domains discussed in this thesis. For example, if one could do a deeper investigation of the WOW data (i.e. with more complete data), it would be possible to pin down the optimal structure for a team in the game, which might then provide a generic method for determining the optimal social group in a variety of real-world situations.

Bibliography

[1] J. L. Casti, *Would-be Worlds*, Wiley, New York (1997).

[2] N. F. Johnson, *Two's Company, Three is Complexity*, Oneworld, New York (2007).

[3] G. D. Mahan, *Many-Particle Physics*, Kluwer Academic/plenum Publishers Oct (2000).

[4] C. Cosner, D. L. Deangelis, J. S. Ault, and D. B. Olson, Theoretical Population Biology **56**, 65-75 (1999).

[5] E Broda, and L Gray. '*Ludwig Boltzmann : man, physicist,philosopher*, Woodbridge (1983).

[6] Neil W. Ashcroft, and N. David Mermin, *Solid State Physics*, Harcourt: Orlando (1976).

[7] D. Sherrington, S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).

[8] J. Ashkenazi, S.E. Barnes, F. Zuo, G.C. Vezzoli, and B.M. Klein (editors), *High Temperature Superconductivity: Physical properties, Microscopic Theory, and Mechanisms*, Plenum Press, NY (1992).

[9] J. Mehra, and H. Rechenberg, *The historical development of quantum theory*, Springer-Verlag (1982).

[10] K. Nakamura et al., J. Phys. G **37**, 075021 (2010).

[11] Strogatz, S. Nature **410**, 268 (2001).

[12] P. Erdös, and A. Rényi, Publ. Math. Debrecen, **6**, 290 (1959).

[13] R. Albert, H. Jeong, and A.-L. Barabasi, Nature, **401**, 130 (1999).

[14] R. Albert, H. Jeong, and A.-L. Barabasi, Nature, **406**, 378 (2000).

[15] D. J. Watts, and S. H. Strogatz, Nature **393** 440 (1998).

[16] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts Phys. Rev. Lett. **85**, 5468 (2000).

[17] R. Albert, and A.-L. Barabasi, Phys. Rev. Lett. **85**, 5234 (2000).

[18] M. E. J. Newman, C. Moore, and D. J. Watts Phys. Rev. Lett. **84**, 3201 (2000).

[19] R. Albert, and A.-L. Barabasi, Reviews of Modern Physics, **74**(1) 47 (2002).

[20] S. N. Dorogovtsev, and J. F. F. Mendes, *Evolution of Networks From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford (2003).

[21] M. E. J. Newman, SIAM Review, **45** 167 (2003).

[22] P. Erdös, and A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci., **5** 17 (1960).

[23] A. J. Lotka, *The Elements of Physical Biology*, Williams & Wilkins, Baltimore (1925).

[24] V. Volterra, Mem. Acad. Lincei. (6) **2**, 31-113 (1926).

[25] M. Fratini, N. Poccia, A. Ricci,G. Campi, M. Burghammer, G. Aeppli, and A. Bianconi, Nature **466**, 841-844 (2010).

[26] http://arxiv.org/abs/1011.6398.

[27] J. Ferkinghoff-Borg, M. H. Jensen, J. Mathiesen, and P. Olesen, Europhys. Lett. **73**, 422 (2006).

[28] V. M. Eguíluz, and M. G. Zimmermann, Phys. Rev. Lett. **85**, 5659 (2000).

[29] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Nat. Acad. Sci. USA, 101(9):2658C2663 (2004).

[30] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002)

[31] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E, **70**, 066111 (2004).

[32] L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas, J. Stat. Mech., **29**(09) 09008 (2005).

[33] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, Nature textbf435(7043), 814 (2005).

[34] A. Clauset, C. Moore, and M. E. J. Newman, Nature **453**(7191), 98 (2008).

[35] U. Alon, Nature Reviews Genetics **8**(6) 450 (2007).

[36] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science, **298**(5594) 824 (2002).

[37] S. S. S. Shai, R. Milo, S. Mangan, and U. Alon, Nature Genetics, **31** 64 (2002).

[38] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Phys. Rep., **424**(4-5) 175 (2006).

[39] B. Bollobas, and O. Riordan. in S. Bornholdt and H. Schuster, editors, *Handbook of Graphs and Networks*, pages 1C37 (2003).

[40] F. R. K. Chung, and L. Lu., Adv. Appl. Math., **26**(4):257C279 (2001).

[41] H. A. Simon, Biometrika, **42** 425, (1955).

[42] D. J. de Solla Price, Science, **149**(3683) 510 (July 1965).

[43] L. Lu, In SODA'01: Proceedings of the 12th annual ACM-SIAM Symposium on Discrete algorithms, pages 912C921 (2001).

[44] G. Bianconi, and A. L. Barabasi, Europhys. Lett., **54** 436C442 (2001).

[45] A. D. Flaxman, A. M. Frieze, and J. Vera, In WAW'04: Proceedings of the 2nd Workshop On Algorithms And Models For The Web-Graph, pages 44C55 (2004).

[46] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, In COCOON 99: Proceedings of the International Conference on Combinatorics and Computing (1999).

[47] E. A. Bender, and E. R. Canfield, J. Comb. Theory A **24**, 296-307 (1978).

[48] J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E **66**, 055101 (2002).

[49] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler,. A.C.S. **5**, 43-54 (2002).

[50] D. Chakrabarti and C. Faloutsos, ACM Computing Survey, **38**(1) 2 (2006).

[51] B. Bollobas and O. Riordan, in S. Bornholdt and H. Schuster, editors, *Handbook of Graphs and Networks*, pages 1C37 (2003).

[52] R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **86**, 3200-3203 (2001).

[53] R. Pastor-Satorras, and A. Vespignani, in S. Bornholdt and H. G. Schuster (eds.), *Handbook of Graphs and Networks*, Wiley-VCH, Berlin (2003).

[54] S. Gueron, and S.A. Levin, Mathematical Biosciences **128**, 243 (1995).

[55] N. F. Johnson, P. Jefferies, and P.M. Hui, *Financial Market Complexity* (Oxford University Press, 2003).

[56] N.F. Johnson, "Complexity in Human Conflict", in *Managing Complexity: Insights, Concepts, Applications* ed. D. Helbing (Springer, Berlin, 2008) p. 303-320; *Back Page*, APS News, (November, 2006); Policing (20 June, 2008) doi:10.1093/police/pan018.

[57] A. Clauset, and F.W. Wiegel, e-print arXiv:0902.0724v1 [physics.soc-ph].

[58] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Review **51**(4), 661C703 (2009).

[59] N. Johnson, M. Spagat, J. Restrepo, J. Bohorquez, N. Suarez, E. Restrepo and R. Zarama, e-print http://lanl.arxiv.org/abs/physics/0605035.

[60] Y.B. Xie, B.H. Wang, H.J. Quan, W.S. Yang, and P.M. Hui, Phys. Rev. E **65**, 046130 (2002).

[61] R. D'hulst, and G.J. Rodgers, Int. Jour. of Theor. and Appl. Fin. **3**, 609 (2000).

[62] J. O. Matthews, K. I. Hopcraft, and E. Jakeman, J. Phys. A: Math. Gen. **36**, 11583 (2003).

[63] E. Jakeman, K. I. Hopcraft, and J .O. Matthews, J. Phys. A: Math. Gen. **38**, 6447 (2005).

[64] Y.B. Xie, B.H. Wang, B. Hu, and T. Zhou, Phys. Rev. E **71**, 046135 (2005).

[65] D.F. Zhang, G.J. Rodgers, P.M. Hui, Physica A **310**, 480 (2002).

[66] D.F. Zhang, P.M. Hui, K.F. Yip, and N.F. Johnson, Eur. Phys. J. B **27**, 213 (2002).

[67] G.J. Rodgers and D.Zheng, Physica A **308**, 375 (2002).

[68] G.J. Rodgers and Y.J. Yap, Eur. Phys. J. B **28**, 129 (2002).

[69] S. Rawal, G.J. Rodgers, Physica A **344**, 50 (2004).

[70] A. Wyld, and G.J. Rodgers, Physica A **374**, 491 (2007).

[71] E. Bonabeau, L. Dagorn and P. Freon, Proc. Natl Acad. Sci. USA **96**, 4472 (1999).

[72] A.J. Modler, K. Gast, G. Lutsch, and G. Damaschun, J. Mol. Biol. **325**, 135 (2003).

[73] M. Smoluchowski, Physik. Zeitschr., **17**, 557 (1916).

[74] M. Smoluchowski, *Zeitschrift f. Physik. Chemie*, **92**, 129 (1917)

[75] F. Leyvraz, Phys. Rev. A **29**, 854 (1984).

[76] F. Leyvraz, Phys. Rep. **383**, 95 (2003).

[77] Ph. Laurençot, J. Differential Equations **167**, 145 (2000).

[78] Ph. Laurençot and S. Mischler, Rev. Mat. Iberoamericana **18**, 731 (2002).

[79] R. L. Drake, *Topics in Current Aerosol Research (Part 2)* , Volume 3 of *International Reviews in Aerosol Physics and Chemistry*, (Pergamon, London, 1972) p. 201.

[80] D. J. Aldous, Bernouilli, **5**, 3 (1999).

[81] V.M. Kolybasova and Yu. N. Sokolskikha, Phys. Lett. B **276**, 409 (1992).

[82] A. Menchaca-Rocha, F. Huidobro, A. Martinez-Davaloz, K. Michaelian, A. Perez, V. Rodriguez and N. Carjan J. of Fluid Mech. **346**, 291 (1997).

[83] R. Fries, V. Greco, P. Sorensen Annual Review of Nuclear and Particle Science **58**, 177 (2008).

[84] V. Greco, C.M. Ko and R. Rapp, Phys. Lett. B **595**, 202 (2004).

[85] J. S. Goldstein, A. J. Mazzella, Il Nuovo Cimento B **21**, 142 (1974).

[86] F. Family, P. Meakin, J. M. Deutch, Phys. Rev. Lett. **57**, 727 (1986).

[87] P. Carruthers, M. Duong-Van, Phys. Lett. B **131**, 116 (1983).

[88] Y. Yoshii, H. Saio, Astrophys. Jour. **295**, 521 (1985).

[89] A.M. Mazzone, J. of Computer-Aided Materials Design **7**, 133 (2000).

[90] S. Scherer, H. Stocker, Proceedings of the International Symposium on Atomic Cluster Collisions, St. Petersburg, Russia, July 18-21, 2003, (2004), p. 169.

[91] S. Lovejoy, H. Gaonac'h, and D. Schertzer, J. Geophys. Res. **109**, 11203 (2004).

[92] L. Zaichik, V. Alipchenkov, High Temperature **46**, 666 (2008).

[93] D. Han, X. Sheng Zhang, W. An Zheng, Acta Mathematica Sinica **24**, 121 (2008).

[94] J. M. Ball, J. Carr, J. Stat. Phys. **61**, 203 (1990).

[95] J.A.D. Wattis, Physica D **222**, 1 (2006).

[96] X. Gabaix, P. Gopakrishnan, V. Plerou and H.E. Stanley, Quarterly Journal of Economics, May Issue, **121**, 461 (2006).

[97] A. Clauset, M. Young, and K. S. Gleditsch, Journal of Conflict Resolution **51**, 58 (2007).

[98] A. Clauset, C. R. Shalizi, M. E. J. Newman, e-print arXiv:0706.1062v2 [physics.data-an].

[99] See for example, the wide range of publications and conferences on computational modeling of socio-economic systems at *www.unifr.ch/econophysics*.

[100] S. Saavedra, F. Reed-Tsochas, and B. Uzzi, Proc. Nat. Acad. Sci. **105**, 16466 (2008).

[101] M. Kaivan, M. Rosenzwieg, Econ. Trans. **13**, 427 (2005).

[102] J. Carr, Proc. Roy. Soc. Edinburgh A, **121**, 231 (1992).

[103] J. Carr, F. P. da Costa, J. Stat. Phys. **77**, 89 (1994).

[104] P. M. Hui, private communication.

[105] S. Schnell and R. Hancock, *The intranuclear environment* in Methods in Molecular Biology - The Nucleus, to appear (2009).

[106] See articles in *Large Scale Structures and Dynamics of Complex Networks*, ed. G. Caldarelli and A. Vespignani (World Scientific, Singapore, 2007).

[107] A. Kirou, B. Ruszczycki, M. Walser and N.F. Johnson, in *Proceedings of the International Conference on Computational Science 2008* (Springer, Berlin, 2008) p. 33 of Part 1.

[108] J.-P. Bouchaud, M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, Cambridge University Press, Cambridge (2004).

[109] R. N. Mantegna, H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge (1999).

[110] R. Cont, J.-P. Bouchaud, Macroeconomic Dynamics **4**, 170 (2000).

[111] D. Challet and Y.C. Zhang, Physica A **246**, 407 (1997).

[112] L.F. Richardson, J. Amer. Statistical Assn, **43**, 523 (1948).

[113] J.C. Bohorquez, S. Gourley, A.R. Dixon, M. Spagat, and N.F. Johnson, Nature **462**, 911 (2009).

[114] A. Clauset, M. Young, *http://xxx.lanl.gov/abs/physics/0502014* (2005).

[115] L. Cederman, Am. Pol. Sci. Rev., **97**, 135 (2003).

[116] M.E.J. Newman, Contemporary Phys., **46**, 323 (2005).

[117] T.X. Hammes, *The Sling and The Stone, On War in the 21st Century*, Zenith, St. Paul, MN (2004).

[118] R. D'Hulst, G.J. Rodgers, it Eur. Phys. J. B, **20**, 619 (2001).

[119] M. Abramowitz, I.A. Stegun,*Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York (1964).

[120] Z. Zhao, J.C. Bohorquez,A. Dixon, N.F. Johnson, Phys. Rev. Lett., **103**, 148701 (2009).

[121] D. Cunningham, K. Gleditsch, I. Salehyan, 46th Annual Convention of the International Studies Association (2005).

[122] N. Plowes, E. Adams, Proc. R. Soc. B, **272**, 1809 (2005).

[123] S. Tanachaiwiwat, A. Helmy, Technical Report CS 05-859, University of Southern California (2005).

[124] S. Yano, New Lanchester strategy. Lanchester, California (1996).

[125] J. Epstein, *Nonlinear Dynamics, Mathematical Biology and Social Sciences*, Addison-Wesley, Reading (1997).

[126] N. MacKay, Math. Today, **42**, 170 (2006).

[127] G. Wu, S. Yan, Amer. Jour. of Infect. Diseases, **1**, 156 (2005); Comp. Clin. Pathol, **11**, 178 (2005).

[128] A. Radford, A. Du Plessis, Animal Behaviour, **68**, 803 (2004).

[129] M. Kaldor, *New and Old Wars*, Stanford University Press (1999).

[130] J. Robb,*Brave New War. Wiley*, New York (2007).

[131] B. Ruszczycki, B. Burnett, Z. Zhao, and N.F. Johnson, Eur. Phys. J. B, in press (2009) DOI: 10.1140/epjb/e2009-00354-5.

[132] B. Efron, SIAM Reviews, **21**, 460-480 (1979).

[133] M.L. Goldstein, S.A. Morris, G.G. Yen, Eur. Phys. J. B, **41**, 255 (2004).

[134] U.S. Army Insurgency Doctrine *www.usma.edu/dmi/IWmsgs/doctrinal_template_1of3.pdf* (2005).

[135] V.M. Creveld,*The Art of War: War and Military Thought*, Cassell, Wellington House (2000).

[136] Conflict Analysis Resources Center, *www.cerac.org.co*

[137] A.H. Cordesman,*The Developing Iraqi Insurgency*, Center for Strategic and International Studies (CSIS.org) (2004).

[138] T.X. Hammes, presentation on the 7th Nov at the Carr Center, Harvard University (2005).

[139] F.E. Harmon, Emergency Order by U.S. Securities and Exchange Commission: *False rumors can lead to a loss of confidence in our markets*, Release Number 58166, July 15 (2008.

[140] D.G. McNeil, *Containing flu is not feasible, specialists say*, New York Times, Front Page, Thursday, April 30 (2009).

[141] Report by The American Academy of Pediatrics (AAP), *Pandemic Influenza: Warning, Children At-Risk* (2007).

[142] S. Lyall, *Susan Boyle, Unlikely Singer Is YouTube Sensation*, New York Times, Front Page, Friday, April 17 (2009).

[143] D. Sornette, F. Deschâtres, T. Gilbert, and Y. Ageon, Phys. Rev. Lett. **93**, 22 (2004).

[144] R. Crane, and D. Sornette, Proc. Nat. Acad. Sci. **105**, 15649 (2008).

[145] M.J. Keeling, and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press (2007).

[146] R.M. May, and A.L. Lloyd, Phys. Rev. E **64**, 066112 (2001).

[147] F. Ball, D. Mollison, and G. Scalia-Tomba, Ann. Appl. Probab. **7** 46 (1997).

[148] J. Shao, S. Havlin, and H.E. Stanley, Phys. Rev. Lett. **103**, 018701 (2009).

[149] J.S. Koopman, in *Biological Networks* ed. F. Kepes, World Scientific, London (2007).

[150] J.D. Murray, *Mathematical Biology: I. An Introduction* 3rd Edition, Springer, New York (2007) Ch. 10.

[151] V. Colizza, and A. Vespignani, Phys. Rev. Lett. **99**, 148701 (2007).

[152] D.J. Watts, R. Muhamad, D.C. Medina, and P.S. Dodds, Proc. Nat. Acad. Sci. **102**, 11157 (2005); P.S. Dodds and D.J. Watts, Phys. Rev. Lett. **92**, 218701 (2004).

[153] T. Gross, C. Dommar, and B. Blasius, Phys. Rev. Lett. **96**, 20 (2006).

[154] L.B. Shaw, and I.B. Schwartz, Phys. Rev. E, **77**, 066101 (2008).

[155] G. Palla, A.L. Barabasi, and T. Vicsek, Nature **446**, 664 (2007).

[156] M. Smoluchowski, Physik. Zeitschr., **17**, 557 (1916).

[157] R. Colbaugh and K. Glass, e-print http://arxiv.org/abs/0912.5242; *Social Marketing* at http : //knowledge.wharton.upenn.edu/article.cfm?articleid = 1864; J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance and M. Hurst, e-print http : //arxiv.org/abs/0704.2803v1; R. Dye, Harvard Business Review, November-December 2000, p. 139.

[158] Additional candidate "infection" profiles where the contagion activity level $I(t)$ is measured as webhits, downloads etc, and where $I(t) \sim 0$ before the initial event, as expected for an outbreak-like contagion phenomenon: (i) Swine-flu at the level of a county, state or country, following initial cases – for example, Florida, U.S. at $\mathrm{http://www.doh.state.fl.us/Disease_ctrl/epi/swineflu/Reports/reports.htm}$; (ii) Code-Red outbreak of Internet worm, see D. Moore, C. Shannon and K. Claffy at $\mathrm{http://www.icir.org/vern/imw-2002/}$; (iii) SARS virus, see p. 18 at $\mathrm{http://www.who.int/csr/sars/WHOconsensus.pdf}$; (iv) website hits for movie *Brokeback Mountain* following death of actor Heath Ledger. Prior to his death, the web traffic had been essentially zero (i.e. $I(t) \sim 0$ as required) for an extended period. See $\mathrm{http://www.alexa.com/}$ for BrokebackMountain; (v) website hits for a new album by band Radiohead in October 2007. See $\mathrm{http://www.alexa.com/}$ for inrainbows; (vi) outbreak of number of Improvised Explosive Devices (IED) in Afghanistan following their first reported use. See $\mathrm{http://www.icasualties.org/OEF/Index.aspx}$

[159] M. Mesterton-Gibbons and E.S. Adams, Science **298**, 2146 (2002).

[160] I.D. Couzin, J. Krause, N.R. Franks, and S.A. Levin, Nature **433**, 513 (2005).

[161] R. Durrett and S. Levin, Journal of Economic Behavior and Organization **57**, 267 (2005).

[162] J.M. Epstein and R. Axtell, *Growing artificial societies: social science from the bottom up*, The Brookings Institute, Washington (1996).

[163] L.E. Cederman, American Political Science Review **97**, 135 (2003).

[164] M.C. Gonzalez, C.A. Hidalgo, and A.L. Barabasi, Nature **453**, 779 (2008).

[165] J.M. Kumpula, J-P. Onnela, J. Saramaki, K. Kimmo, J. Kertesz, Phys. Rev. Lett. **99**, 228701 (2007).

[166] R. Guimera, B. Uzzi, S. Jarrett, and L.A. Nunes Amaral, Science **308**, 697 (2005).

[167] A.L. Barabasi, Nature **435**, 207 (2005).

[168] S. Wuchty, B. Jones, B. and B. Uzzi, Science **316**, 1036 (2007).

[169] G. Palla, A.L. Barabasi, and T. Vicsek, Nature **446**, 664 (2007).

[170] S. Galam, Int. J. Mod. Phys. C **19**, 409 (2008) and references therein.

[171] See www.unifr.ch/econophysics for details.

[172] N. Ducheneaut, N. Yee, E. Nickell, and R.J. Moore, Games and Culture **1**, 281 (2006).

[173] D. Williams, N. Ducheneaut, X. Li, Y. Zhang, N. Yee, and E. Nickell, Games and Culture **1**, 338 (2006).

[174] International Business Machines study: *Gaming as a Metaphor For 21st Century Leadership*, available at www.ibm.com/gio (2007).

[175] D. Their, Wired Magazine, 20 March (2008).

[176] R.D. Balicer, Epidemiology **18**, 260 (2007).

[177] US Justice Department Report: *Hybrid and Other Modern Gangs*, Juvenile Justice Bulletin, December 2001.

[178] http://news.bbc.co.uk/1/hi/world/americas/7328967.stm

[179] G. Tita, in *The Atlas of Crime*, eds. Hendrix, E.H., Dent, B. and Turnbull, L.S., Oryx Press, Phoenix (2000).

[180] N. Katz, and D. Lazer, Small Group Research **35**, 307 (2004).

[181] P. Gongla, and C.R. Rizzuto, IBM Systems Journal **40**, 842 (2001).

[182] D. Kreps, and R. Wilson, Journal of Economic Theory **27**, 253 (1982).

[183] B. Mayhew, J. McPherson, T. Rotolo, Social Forces **74**, 15 (1995).

[184] M. Ruef, H.E. Aldrich, and N.M. Carter, American Sociological Review, **68**, 195 (2003).

[185] P. Lazarsfeld, and R.K. Merton, Friendship as a Social Process: A Substantive and Methodological Analysis, in *Freedom and Control in Modern Society*, eds. Berger, M., Abel, T. and Page, C.H. (New York, Van Nostrand, 1954) p. 18.

[186] P.M. Blau, American Journal of Sociology, **83** 26 (1977).

[187] J.M. McPherson, and L. Smith-Lovin, American Sociological Review, **52**, 370 (1987).

[188] J.M. McPherson, L. Smith-Lovin, and J. Cook, J. Annual Review of Sociology, **27**, 415 (2001).

[189] P. Popielarz, and J.M. McPherson, American Journal of Sociology, **101**, 698 (1995).

[190] J.W. Thibaut, and H.H. Kelley, The Social Psychology of Groups, New York, Wiley (1959)

[191] P.J. Hinds, K.M. Carley, D. Krackhardt, and D. Wholey, Organizational Behavior and Human Decision Processes, **81**, 226 (2000).

[192] N. Stern, *The Economics of Climate Change: The Stern Review*, Cambridge Univ. Press, Cambridge (2007); *A Blueprint for a Safer Planet*, Random House, London (2009), pp. 39; A.D. Ellerman, *et al. Markets for Clean Air: The U.S. Acid Rain Program*, Cambridge Univ. Press, Cambridge (2000).

[193] See news.bbc.co.uk/1/hi/world/europe/6732787.stm.

[194] *Emission Trading: Environmental Policy's New Approach*, Eds. R.F. Kosobud, D.L. Schreder, and H.M. Biggs, John Wiley & Sons, Inc., New York (2000); *Voluntary Carbon Markets: An International Business Guide to What They Are and How They Work* Eds. R. Bayon, A. Hawn, and K. Hamilton, Earthscan Publications Ltd., London (2009); *Emission Trading: institutional Design, Decision Making and Corporate Strategies*, Eds. R. Antes, B. Hansjurgens, and P. Letmathe, Springer, New York (2008).

[195] See news.bbc.co.uk/2/hi/business/8244600.stm.

[196] J. O'Brien, *Engineering a Financial Bloodbath*, World Scientific, Singapore (2009).

[197] W.B. Arthur, Amer. Econ. Assoc. Papers. Proc., **84**, 405 (1994); Science **284**, pp. 107-109 (1999).

[198] D. Challet, and Y.C. Zhang, Physica A, **246**, 407 (1997); D. Challet, M. Marsili, and Y.C. Zhang, *Minority Games*, Oxford University Press (2005); A.C.C. Collen, *The mathematical theory of Minority Games*, Oxford University Press (2005); T.Galla, D. Sherrington, J. Stat. Mech., P 10009 (2005); D. Sherrington, E. Moro, and J.P. Garrahan, Physica A, **311**, 527 (2002); T. Galla, and A. De Martino, J. Phys. A: Math. and Theor., **41** 324003 (2008).

[199] N.F. Johnson, et al., Physica A, **258**, 230 (1998).

[200] W. Wang, Y. Chen, and J. Huang, Proc. Natl. Acad. Sciences. U.S.A., **106**, 8423 (2009).

[201] P. Jefferies, M.L. Hart, and N.F. Johnson, Phys. Rev. E, **65**, 016105 (2001).

[202] Price of the European call option, using Black-Scholes equation, is [55]: $V(x,t) = x\Phi[d_1] - X_s e^{-r(t_0-t)}\Phi[d_2]$, where $\Phi[z] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-0.5y^2}$, $d_1 = \frac{ln(x/X_s)+(r+0.5\sigma^2)(t_0-t)}{\sigma\sqrt{t_0-t}}$, $d_1 = \frac{ln(x/X_s)+(r-0.5\sigma^2)(t_0-t)}{\sigma\sqrt{t_0-t}}$, where σ is the volatility.

APPENDIX A

Derivation of Eq. 3.6

There are two populations A and B. Population A comprises an average of n_s clusters of size s and has a total population N, while population B has p_s clusters of size s and total population P. Each timestep, a cluster is selected at random from the total population $N + P$, so the probability of a cluster of size s being selected is proportional to n_s (p_s for B). A second cluster is then selected with probability proportional to its size, sn_s (sp_s). The two cluster types (A or B) are compared, if they are the same the two clusters coalesce, and if they are different then the smaller of the two clusters selected fragments (both fragment if they are the same size). The time evolution of the number of clusters of size s, n_s , is given for $s \geq 2$ by

$$\begin{aligned} n_s[t+1] \quad & -n_s[t] = -\frac{n_s}{\sum n_{s'} + \sum p_{s'}} \frac{\sum_{s' \geq s} s' p_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} \\ & -\frac{\sum_{s' \geq s} p_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{s n_s}{\sum s' n_{s'} + \sum s' p_{s'}} - \frac{n_s}{\sum n_{s'} + \sum p_{s'}} \frac{\sum s' n_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} \\ & -\frac{\sum n_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{s n_s}{\sum s' n_{s'} + \sum s' p_{s'}} + \frac{\sum_{s'}^{s-1} n_{s'} (s-s') n_{s-s'}}{(\sum n_{s'} + \sum p_{s'})(\sum s' n_{s'} + \sum s' p_{s'})} \end{aligned} \tag{A.1}$$

and for $s = 1$ by

$$\begin{aligned} n_1[t+1] - n_1[t] = & -\frac{n_1}{\sum n_{s'} + \sum p_{s'}} \frac{\sum s' n_{s'}}{\sum s' n_{s'} + \sum s' p_{s'}} \\ -\frac{\sum n_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{n_1}{\sum s' n_{s'} + \sum s' p_{s'}} & + \frac{\sum_{s'=2} s' n_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{\sum_{r' \geq s'} r' p_{r'}}{\sum s' n_{s'} + \sum s' p_{s'}} \\ +\frac{\sum_{s'=2} p_{s'}}{\sum n_{s'} + \sum p_{s'}} \frac{\sum_{r'=s'}^{r' \leq s'} r'^2 n_{r'}}{\sum s' n_{s'} + \sum s' p_{s'}} & \end{aligned} \tag{A.2}$$

Similar equations hold for population B. The first two terms on the RHS are due to a cluster selecting (being selected by) a larger cluster of the opposite population, which causes it to fragment. The next two terms are due to a cluster of size s selecting any other cluster from the same population and coalescing with it. The final term is due to to two clusters joining together to form a new cluster of size s. The $s = 1$ equation's first two terms correspond to a single agent being selected to coalesce, and the last two terms with a larger cluster fragmenting into single agents. The steady state of this system is found from computer simulation to have the larger of the two initial populations forming a single cluster of maximum size. Taking B to be the larger population, $(P > N)$ then we can simplify the above equations by using the fact that

$$\sum p_{s'} = 1 \qquad \sum s' p_{s'} = P \qquad \sum s' n_{s'} = N$$

giving

$$n_s[t+1] - n_s[t] = \frac{-n_s P - s n_s - n_s N - s n_s \sum n_{s'} + \sum_{s'}^{s-1} (s-s') n_{s'} n_{s-s'}}{(\sum n_{s'} + 1)(N+P)} \quad s > 1 \tag{A.3}$$

$$n_1[t+1] - n_1[t] = \frac{-n_1 N - n_1 \sum n_{s'} + P \sum_{s'=2} s' n_{s'} + \sum_{s'=2} s'^2 n_{s'}}{(\sum n_{s'} + 1)(N+P)} \quad s = 1 \tag{A.4}$$

We now make the approximation that $\sum n_{s'} \gg 1$, so that the probability of a B cluster being picked first is negligible. This is applicable for large N. With this approximation, in the steady state these equations become

$$0 = \frac{-n_s(P + N + s\sum n_{s'}) + \sum_{s'}^{s-1} n'_s n_{s-s'}}{(\sum n_{s'})(N+P)} \qquad 0 = \frac{-n_1 N - n_1 \sum n_{s'} + P(N - n_1)}{(\sum n_{s'})(N+P)}$$

$$n_s = \frac{1}{(P + N + s\sum n_{s'})} \sum_{s'}^{s-1} s' n_{s'} n_{s-s'} \qquad n_1 = \frac{NP}{(N + \sum n_{s'} + P)} \tag{A.5}$$

Introducing the generating function $h(\omega)$, and its derivative

$$h(\omega) = \sum_{r=2}^{\infty} n_r e^{-\omega r} \qquad \frac{dh(\omega)}{d\omega} = -\sum_{r=2}^{\infty} r n_r e_{-\omega r} \tag{A.6}$$

We note that

$$
\begin{aligned}
-h(\omega)\frac{dh(\omega)}{d\omega} &= e^{-4\omega}(2n_2n_2) + e^{-5\omega}(2n_2n_3 + 3n_3n_2) + e_{6\omega}(2n_2n_4 + 3n_3n_3 + 4n_4n_2) + ... \\
= \quad & \sum_{i=4} e^{-i\omega} \sum_{s'=2}^{i-2} s' n_{s'} n_{i-s'} = \sum_{i=4} e^{-i\omega} \left(\sum_{s'=1}^{i-1} s' n_{s'} n_{i-s'} - n_1 n_{i-1} - (i-1) n_{i-1} n_1 \right)
\end{aligned}
$$

Using Eq. A.5 to replace the inner sum,

$$
\begin{aligned}
-h(\omega)\frac{dh(\omega)}{d\omega} &= \sum_{i=4} e^{-i\omega} \left(n_i \left(P + N + i \sum n_{s'} \right) - i n_1 n_{i-1} \right) \\
= \quad & \sum e^{-i\omega} \left(n_i \left(P + N + i \sum n_{s'} \right) - i n_1 n_{i-1} \right) - e^{-2\omega} \left(n_2 \left(P + N + 2 \sum n_{s'} \right) - 2 n_1 n_1 \right) \\
& - e^{-3\omega} \left(n_3 \left(P + N + 3 \sum n_{s'} \right) - 3 n_1 n_2 \right)
\end{aligned}
$$

Comparing to the $s = 2$ and $s = 3$ terms of Eq. A.5,

$$
\begin{aligned}
-h(\omega)\frac{dh(\omega)}{d\omega} &= \sum_{i=2} e^{-i\omega} n_i \left(P + N + i \sum n_{s'} \right) - n_e^{-\omega} \sum_{i=2} e^{-(i-1)\omega} i n_{i-1} + n_1 n_1 e^{-2\omega} \\
= \quad & (P+N) \sum_{i=2} n_i e^{-i\omega} + \sum n_{s'} \sum_{i=2} i n_i e^{-\omega} - n_1 e^{-\omega} \sum_{i=1} e^{-i\omega} (i+1) n_i + n_1 n_1 e^{-2\omega} \\
= \quad & (P+N) h(\omega) - \sum n_{s'} \frac{dh(\omega)}{d\omega} - n_1 e^{-\omega} \sum_{i=2} e^{-i\omega} (i+1) n_i - 2 n_1 n_1 e^{-2\omega} + n_1 n_1 e^{-2\omega}
\end{aligned}
$$

$$
-h(\omega)\frac{dh(\omega)}{d\omega} = (P + N - n_1 e^{-\omega}) h(\omega) - \left(\sum n_{s'} - n_1 e^{-\omega} \right) \frac{h(\omega)}{d\omega} - (n_1 e^{-\omega})^2 \quad \text{(A.7)}
$$

Setting $\omega = 0$, we obtain

$$
h(0) = \sum_{r=2}^{\infty} n_r = \sum n_s - n_1 \qquad \frac{dh(0)}{d\omega} = -\sum_{r=2}^{\infty} r n_r = -N + n_1
$$

$$
\begin{aligned}
(N - n_1) \left(\sum n_s - n_1 \right) \quad &= \quad (P + N + n_1) \left(\sum n_s - n_1 \right) (-N + n_1) - n_1^2 \\
0 = P \left(\sum n_s - n_1 \right) \quad &+ \quad (N - n_1) \sum n_{s'} - N n_1 \qquad n_1 = \frac{(N+P) \sum n_s}{N + P + \sum n_s}
\end{aligned}
$$

Comparing with Eq. A.5 gives

$$
\sum n_s = \frac{NP}{N+P} \qquad n_1 = \frac{N+P}{\frac{N}{P} + \frac{P}{N} + 2} \quad \text{(A.8)}
$$

Comparing powers of $e^{-\omega}$ in Eq. A.7 Making the assumption that the number of groups of size 2 is likely to be much smaller than the number of size 1, we can parameterize the n_2 and higher terms by a small value, β. From computer simulations the group size distribution is expected to be a power law, and so as $1^x(i-1)^x > j^x(i-j)^x$ for $j < i-1$ and any x, the higher order terms are always smaller than the n_1. With this it can be seen that Iterating this back to n_1

$$\begin{aligned}
n_i &= n_{i-2}n_1\frac{i-1+\beta}{(N+P+(i+1)\sum n_s)}n_1\frac{i+\beta}{(N+P+i\sum n_s)} \\
n_1 &= n_1\prod_{j=2}^{j=i} n_i\frac{j+\beta}{(N+P+j\sum n_s)} \\
n_1 &= n_1^i\frac{(i+\beta)!}{(1+\beta)!}\frac{1}{\left(\sum n_s\right)^{i-1}}\frac{\left(\frac{N+P}{n_s}+1\right)}{\left(\frac{N+P}{\sum n_s}+i\right)!}
\end{aligned}$$

Taking logarithms and applying Stirling's formula

$$\ln\, n_i = \ln\left(n_1\left(\frac{n_1}{\sum n_s}\right)^{i-1}\frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{(1+\beta)!}\right) + \ln(i+\beta)! - \ln\left(\frac{N+P}{\sum n_s}+i\right)!$$

$$\begin{aligned}
= \quad & \ln\left(n_1\left(\frac{n_1}{\sum n_s}\right)^{i-1}\frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{(1+\beta)!}\right) + \frac{1}{2}\ln(2\pi) + (i+\beta+\frac{1}{2})\ln(i+\beta) \\
& -(i+\beta) - \frac{1}{2}\ln(\pi) - \left(\frac{N+P}{\sum n_s}+i+\frac{1}{2}\right)\ln\left(\frac{N+P}{\sum n_s}+i\right) + \left(\frac{N+P}{\sum n_s}+i\right) + ...O\left(i^{-1}\right) \\
= \quad & \ln\left(n_1\left(\frac{n_1}{\sum n_s}\right)^{i-1}\frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{(1+\beta)!}\right) + (i+\beta+\frac{1}{2})\ln(i+\beta) \\
& -\left(\frac{N+P}{\sum n_s}+i+\frac{1}{2}\right)\ln\left(\frac{N+P}{\sum n_s}+i\right) + \frac{N+P}{\sum n_s} - \beta + ...O\left(i^{-1}\right)
\end{aligned}$$

As i becomes large we can neglect the terms of order i^{-1}. Also as $\sum n_s = \frac{NP}{N+P}$ (Eq. A.8) then $\frac{N+P}{\sum n_s} = \frac{N}{P} + \frac{P}{N} + 2$ and therefore for N and P of the same order of magnitude, this is a term of order 1. β is defined as a small quantity, so in the i large limit we can take the two rightmost logarithms as equal to i.

$$\begin{aligned}
\ln n_i \approx \quad & \ln\left(n_1\left(\frac{n_1}{\sum n_s}\right)^{i-1}\frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{(1+\beta)!}exp\left(\frac{N+P}{\sum n_s}-\beta\right)\right)\\
& +\left(i+\beta+\frac{1}{2}-\frac{N+P}{\sum n_s}-i-\frac{1}{2}\right)\ln i\\
= \quad & \ln\left(n_1\left(\frac{n_1}{\sum n_s}\right)^{i-1}\frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{(1+\beta)!}exp\left(\frac{N+P}{\sum n_s}-\beta\right)i^{\left(\beta-\frac{N+P}{\sum n_s}\right)}\right)\\
n_i = \quad & n_1\left(\frac{n_1}{\sum n_s}\right)^{i-1}\frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{(1+\beta)!}exp\left(\frac{N+P}{\sum n_s}-\beta\right)i^{\left(\beta-\frac{N+P}{\sum n_s}\right)}
\end{aligned}$$

Using the value of n_1 and $\sum n_s$ from Eq. A.8 this becomes

$$n_i = n_1\left(1+\frac{NP}{(N+P)^2}\right)^{1-i}\frac{\left(\frac{N}{P}+\frac{P}{N}+3\right)!}{(1+\beta)!}e^{\left(\frac{N}{P}+\frac{P}{N}+2-\beta\right)}i^{\left(\beta-\frac{N}{P}-\frac{P}{N}-2\right)} \tag{A.9}$$

The term to the power $1-i$ can be seen to approach 1 as P becomes larger than N, which was one of our initial assumptions. The dominant i dependence is then the power law with exponent $(\beta - 2 - \frac{N}{P} - \frac{P}{N})$. Comparing with the power law coefficients determined from numerical simulations it can be seen that $\beta \approx 1.7$, confirming our assumption of it as small compared to i. To remove the β parameter, we can use the form of Eq. A.9 to substitute into Eq. A.7. Comparing powers of $e^{-\omega}$ as before, we find for the general case

$$n_i = \frac{1}{N+P+i\sum n_s}\sum_{j=1} n_j(i-j)n_{i-j} \qquad n_i = \frac{n_{i-1}}{N+P+i\sum n_s}\sum_{j=1}(i-j)\frac{n_j n_{i-j}}{n_{i-1}}$$

Substituting in the form of Eq. A.9, $n_j = AB^j j^c$ with A, B, c constants

$$\begin{aligned}
n_i &= \frac{n_{i-1}}{N+P+i\sum n_s}\sum_{j=1}(i-j)\frac{AB^j j^c AB^{i-j}(i-j)^c}{AB^{i-1}(i-1)^c}\\
n_i &= \frac{n_{i-1}}{N+P+i\sum n_s}\sum_{j=1} ABj^c\frac{(i-j)^{c+1}}{(i-1)^c}\\
n_i &= \frac{n_{i-1}}{N+P+i\sum n_s}\sum_{j=1} ABj^c(i-1)\left(\frac{i}{i-1}-\frac{i}{j-1}\right)^{c+1}
\end{aligned}$$

As i is large we can take $\frac{i}{i-1} \approx 1$ and expand the bracket using the binomial series,

$$n_i = \frac{n_{i-1}}{N+P+i\sum n_s} \sum_{j=1} ABj^c(i-1)\left(1-(c+1)\frac{j}{i-1}+\frac{(c+1)c}{2!}\left(\frac{-j}{i-1}\right)^2+...\right)$$

As i becomes large we neglect the terms of order i^{-1} and lower,

$$\begin{aligned} n_i &= \frac{n_{i-1}}{N+P+i\sum n_s} \sum_{j=1} ABj^c(i-1-(c+1)j) \\ n_i &= \frac{n_{i-1}}{N+P+i\sum n_s} \left(i\sum_{j=1} ABJ^c - \sum_{j=1} ABj^c - c\sum_{j=1} ABjj^c - \sum_{j=1} ABjj^c\right) \\ n_i &= \frac{n_{i-1}}{N+P+i\sum n_s} \left(i\sum_{j=1} n_j - \sum_{j=1} n_j - c\sum_{j=1} jn_j - \sum_{j=1} jn_j\right) \end{aligned}$$

which uses the result from Eq. A.9 that $B \approx 1$. Then as $\sum jn_j = N$,

$$n_i = \frac{n_{i-1}}{N+P+i\sum n_s}\left(i\sum n_s - \sum n_s - cN - N\right)$$

This is similar to Eq. **??**, and it can be iterated in an analogous way to yield

$$n_i = n_1 \frac{\left(\frac{N+P}{\sum n_s}+1\right)!\left(i-1-\frac{cN}{\sum n_s}-\frac{N}{\sum n_s}\right)!}{\left(\frac{N+P}{\sum n_s}+i\right)!\left(1-1-\frac{cN}{\sum n_s}-\frac{N}{\sum n_s}\right)!}$$

Applying Stirlings formula in the i large limit as perviously results in

$$n_i = n_1 \frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{\left(-\frac{cN}{\sum n_s}-\frac{N}{\sum n_s}\right)!} e^{\left(\frac{N+P}{\sum n_s}+\frac{cN}{\sum n_s}+\frac{N}{\sum n_s}+1\right)} i^{-\left(-\frac{N+P}{\sum n_s}-\frac{cN}{\sum n_s}-\frac{N}{\sum n_s}-1\right)} \tag{A.10}$$

where c was defined as the power law coefficient of n_i, so for self consistency

$$\begin{aligned} c &= -\frac{N+P}{\sum n_s} - \frac{cN}{\sum n_s} - \frac{N}{\sum n_s} - 1 \\ c\left(1+\frac{N}{\sum n_s}\right) &= -\frac{N+P}{\sum n_s} - \frac{N}{\sum n_s} - 1 \\ c &= -\frac{N+P}{\sum n_s}\frac{\sum n_s}{N+\sum n_s} - 1 \end{aligned}$$

With c determined, we can write Eq. A.10 as

$$n_i = n_1 \frac{\left(\frac{N+P}{\sum n_s}+1\right)!}{\left(\frac{N}{\sum n_s}\frac{N+P}{N+\sum n_s}\right)!} e^{\left(\frac{N+P}{N+\sum n_s}+1\right)} i^{-\left(-\frac{N+P}{N+\sum n_s}-1\right)}$$

using the value of $\sum n_s$ from Eq. A.8

$$n_i = n_1 \frac{\left(\frac{N}{P} + \frac{P}{N} + 3\right)!}{\left(\frac{(N+P)!}{N^2P+2NP^2}\right)!} e^{\left(\frac{P}{N}\frac{N}{P+2}+2\right)} i^{-\left(\frac{P}{N}\frac{P}{N+2P}+2\right)} \tag{A.11}$$

$$n_i = n_1 C i^{-\left(\frac{P}{N}\frac{P}{N+2P}+2\right)} \tag{A.12}$$

where C is a constant. This is valid for $i \gg 1$ and $P > N$.