

DISTRIBUTION DYNAMICS OF COMPLEX SYSTEMS

A Dissertation

Submitted to Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Gordon A. and Mary Cain Department of Chemical Engineering

by

Young-Pyo Jeon

B.S., Kangwon National University, Korea, 1997

M.S., Kangwon National University, Korea, 1999

December 2006

*To Jaewook and Yeounoak,
for their love and support*

ACKNOWLEDGMENTS

First, I thank my father Jaewook Jeon and mother Yeounoak Kim, for their unconditional love, patience, encouragement, and support. Without your support and guidance through the years of my life, I would not be where I am today.

There are many people in the chemical engineering department at the Louisiana State University whom I must acknowledge. First, I thank my advisors, Martin A. Hjortso and Benjamin J. McCoy, for giving me great latitude to exercise my creativity and to study interesting problems. I am deeply indebted for you to make time to answer my questions always, no matter how trivial. I am always amazed by your ability to identify the important problems in a field. It has been a great honor to work with you and learn from you. I also thank Professors Kalliat T. Valsaraj, Karsten E. Thompson, and Kenneth A. Rose for your constructive advises and suggestions, when they served as my graduate committee member. I am also grateful to Professors Elizabeth J. Podlaha-Murphy and Srinath V. Ekkad for you to become members of my final examination committee. I also thank members of my research group, Jiao Yang, Lisa Brenskelle-Elmer, and Rujun Li. I could always count on either reasoning out research problems or taking a break for humor with you.

I am also grateful to Professor Yong Jung Kwon, who advised me during my Master course, for your conscientious advises and suggestions. I thank Kwangbok Yi and Sungho Lee, who helped me to settle down my life in Baton Rouge. I also wish to thank my dear friends: Jonghoon Kim, Keyyoung Park, and Hana Kim, who have loved and supported, although perhaps I would have graduated sooner if you had not come to Baton Rouge in fall 2005. I also thank my friend Kyuhwan who have cheered and supported me from Korea, far enough away from Baton Rouge. Finally, I have had the great pleasure of

getting to know Heejoung An. I thank you for your love, patience, and support over the past year, and the years in Baton Rouge would not have been as special without you.

Young-Pyo Jeon

Louisiana State University

December 2006

TABLE OF CONTENTS

| | |
|--|-------------|
| DEDICATION | ii |
| ACKNOWLEDGMENTS | iii |
| LIST OF TABLES | vii |
| LIST OF FIGURES | viii |
| LIST OF NOMENCLATURE | x |
| ABSTRACT | xii |
| | |
| CHAPTER 1. INTRODUCTION | 1 |
| 1-1 Complex Systems | 1 |
| 1-2 Networks | 11 |
| 1-3 Power Law Distributions | 13 |
| 1-4 Accelerating Networks | 14 |
| 1-5 Human Dynamics | 15 |
| | |
| CHAPTER 2. DISTRIBUTION DYNAMICS OF EVOLVING NETWORKS | 18 |
| 2-1 Population Balance Dynamics | 18 |
| 2-2 Networks | 20 |
| 2-3 Distribution Kinetics | 23 |
| 2-4 Random Networks | 26 |
| 2-5 Power Law Networks | 33 |
| 2-6 Conclusion | 40 |
| | |
| CHAPTER 3. EVOLUTION OF POWER LAW DISTRIBUTIONS IN SCIENCE AND SOCIETY | 42 |
| 3-1 Introduction | 42 |
| 3-2 Cluster Distribution Dynamics | 45 |
| 3-3 Power Laws | 49 |
| 3-4 Moment Expressions | 55 |
| 3-5 Conclusion | 57 |
| | |
| CHAPTER 4. ACCELERATING NETWORKS WITH AND WITHOUT PREFERENTIAL ATTACHMENT | 59 |
| 4-1 Introduction | 59 |
| 4-2 Model | 61 |
| 4-3 Distribution Kinetics | 63 |
| 4-4 Exponential Networks: Absence of Preferential Attachment | 65 |
| 4-5 Power Law Networks: Effect of Preferential Attachment | 70 |

| | |
|---|------------|
| 4-6 Conclusion | 78 |
| CHAPTER 5. DISTRIBUTION KINETICS OF HUMAN DYNAMICS | 81 |
| 5-1 Introduction | 81 |
| 5-2 Distribution Kinetics of Human Activities | 84 |
| 5-3 Results and Discussions | 89 |
| 5-4 Conclusion | 93 |
| CHAPTER 6. | |
| SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS | 95 |
| 6-1 Summary | 95 |
| 6-2 Conclusions | 96 |
| 6-3 Model Applications and Limitations | 98 |
| 6-4 Recommendations | 100 |
| REFERENCES | 103 |
| APPENDIX: LETTERS OF PERMISSION | 111 |
| VITA | 113 |

LIST OF TABLES

| | |
|---|----|
| Table 2-4.1 The asymptotic behavior and long-time limits of number average, polydispersity, and variance for $I_i = 0$ (no node generation) | 31 |
| Table 2-4.2 The asymptotic behavior and long-time limits of number average, polydispersity, and variance for $I_i = \alpha$ (constant node generation), and $I_i = \alpha t$ (time-dependent node generation) | 32 |
| Table 2-5.1 Time dependence of the linkage distribution $p(\xi, \theta)$ for network growth | 35 |
| Table 3-1.1 Power law systems and distributed properties | 44 |
| Table 3-1.2 Power of the frequency distribution, $\xi^{-\lambda}$, for different systems | 44 |
| Table 3-3.1 Parameters for comparison of corporation size data with our model ... | 53 |
| Table 3-4.1 Moment results and their asymptotes | 56 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1-2.1 Two types of networks: (A) random networks, (B) power-law networks | 12 |
| Figure 1-3.1 Schematic representation of birth and growth of a power law networks | 14 |
| Figure 2-4.1 Evolution of the Gaussian distribution for (A) network growth ($\gamma = 1.5$ and $\kappa = 1.0$) and (B) breakage ($\gamma = 1.0$ and $\kappa = 1.5$) | 28 |
| Figure 2-4.2 The discrete Poisson distribution for irreversible network growth ($\kappa = 0$, $\gamma = 0.5$) at times $\theta = 1, 5, 10, 50, 100$ | 29 |
| Figure 2-5.1 The general power law distribution for case (a) of Table 2-5.1 | 36 |
| Figure 2-5.2 Evolution of the distribution for network growth (b) of Table 2-5.1, and the scaled time θ increases from top to bottom in steps of 4 from $\theta = 1$ to 29 | 36 |
| Figure 2-5.3 Evolution of the distribution for network growth, (c) of Table 2-5.1, where the scaled time θ increases from left to right in steps of 4 from $\theta = 1$ to 29 | 37 |
| Figure 2-5.4 Evolution of the distribution for network growth, (d) of Table 2-5.1, where the scaled time θ increases from left to right in steps of 1 from $\theta = 1$ to 4 (A) and in steps of 10 from $\theta = 1$ to 61 (B) | 37 |
| Figure 2-5.5 Comparison to statistical data of Oregon Internet growth in different years with the model distribution | 38 |
| Figure 3-3.1 Evolution of the size distribution for cluster growth cases Eqs. 3-3.5 and 3-3.8 | 52 |
| Figure 3-3.2 Evolution of the size distribution for cluster growth cases Eqs. 3-3.10 | 53 |
| Figure 3-3.3 Comparison of the model and statistical data of U.S. company size-distribution growth in different years | 54 |
| Figure 4-4.1 Evolution of the Gaussian (A) and Poisson (B) distributed network growth based on the moment results in Eqs. 4-4.4 – 4-4.6 (A) and Eqs. 4-4.10 – 4.4.12 (B) | 67 |
| Figure 4-4.2 Evolution of the Gaussian (A) and Poisson (B) distributed network growth based on the moment results in Eqs. 4-4.4 – 4-4.6 (A) and | |

| | |
|---|----|
| Eqs. 4-4.10 – 4-4.12 (B) | 69 |
| Figure 4-5.1 Evolution of moments of the power law distributed network (Eqs. 4-5.4 – 4-5.6) with initial condition, $p(\xi, t=0) = 0$ and boundary condition, $p(\xi=1, t) = p_o^{(0)} e^{\theta/\tau}$ | 72 |
| Figure 4-5.2 The scaled degree versus total number of nodes for the model (Eqs. 4-5.4 – 4-5.6 with $\lambda = 1$, $p_o^{(0)} = 100$, $\xi_m = 1000$, and $\tau = 20$) | 74 |
| Figure 4-5.3 The Scaled degree versus total number of nodes for the model (Eqs. 4-5.5 – 4-5.7 with $\lambda = 1$, $p_o^{(0)} = 100$, $\xi_m = 1000$, and $\tau = 20$) | 77 |
| Figure 5-3.1 The waiting-time distributions by (A) the random-order protocol, $\lambda = 0$, with the initial condition, $p(\xi, t=0) = 0$, and boundary condition, $p(\xi=1, t) = p_o e^{t/\tau}$; and (B) the priority-relevant protocol, $\lambda = 1$, with the initial condition, $p(\xi, t=0) = 0$, and boundary condition, $p(\xi=1, t) = p_o(1 - e^{-t/\tau})$ | 90 |
| Figure 5-3.2 Comparison of the model and statistical data for email and printing server systems on log-log coordinates | 92 |

LIST OF NOMENCLATURE

- b power exponent for Zipf's frequency distribution
 c_i constants
 $C^{(0)}$ zeroth moment for clusters (total number of cluster)
 $C^{(1)}$ first moment for clusters (mass of cluster)
 $C^{(2)}$ second moment for clusters
 $C(\xi)$ cluster with mass ξ .
 $C(\xi,t)$ cluster distribution (total number of clusters at time t in the differential properties range ξ to $\xi+d\xi$)
 C_o initial value for cluster distribution
 C^{avg} average moment for cluster
 C^{pd} polydispersity for cluster
 C^{var} variance for cluster
 D hypothetical diffusivity for convective diffusion equation.
 $f(\xi)$ initial condition for power law distribution
 $g(\theta)$ boundary condition for power law distribution
 I_i node insertion or removal terms for networks and nucleation for cluster systems
 k network or cluster growing intensity
 $m_o^{(0)}$ initial zeroth moment of monomer
 $m(\xi,t)$ monomer distribution (total number of monomer at time t in the differential properties range ξ to $\xi+d\xi$)
 $M(\xi')$ monomer with mass ξ'
 $p(\xi,t)$ node distribution (total number of nodes at time t in the connection range from ξ to $\xi+d\xi$)
 $P(\xi)$ node with ξ connections.
 $P^C(\xi,t)$ cumulative distribution defined as, $P^c(\xi,\theta) = \int_{\xi}^{\infty} p(\xi,\theta) d\xi$.
 p_o initial number of cluster
 $p_o^{(0)}$ initial zeroth moment
 $p_o^{(1)}$ initial first moment

- p_o^{avg} initial average moment
 p_o^{pd} initial polydispersity
 $p^{(0)}(t)$ zeroth moment, which represents the total number of nodes for networks
 $p^{(1)}(t)$ first moment, which represents the total number of connections
 $p^{(2)}(t)$ second moment, which provides further information for the distribution
 $p^{(n)}(t)$ general n^{th} moment defined as $p^{(n)}(t) = \int p(\xi, t) \xi^n d\xi$.
 $p^{\text{avg}}(t)$ average distribution defined as a ratio between first and zeroth moments
 $(p^{\text{avg}}(t) = p^{(1)}(t)/p^{(0)}(t))$
 $p^{\text{pd}}(t)$ polydispersity defined as $p^{\text{pd}} = p^{(2)} p^{(0)}/p^{(1) 2}$
 $p^{\text{var}}(t)$ variance defined as $p^{\text{var}} = p^{(2)}/p^{(0)} - p^{\text{avg} 2}$
 r rank for Zipf's frequency distribution
 s Laplace transform variable
 t time
 u unit step function where $u(x) = 1$ if $x \geq 0$ and $u(x) = 0$ if $x < 0$
 V hypothetical velocity for convective diffusion equation

 α time-dependent node addition parameter.
 $\delta(\xi-\xi_i)$ Direct Delta function
 δ_{nm} Kronecker delta, 0 if $n \neq m$ and 1 if $n = m$.
 γ prefactor of the growth rate coefficient.
 κ prefactor of the dissociation rate coefficient.
 λ power of the growth rate coefficient.
 ν power of the dissociation rate coefficient.
 θ hypothetical time defined by using time t and $p^{(0)}_o$ or ξ_m .
 τ node or monomer addition controlling parameter
 ξ number of links or mass of clusters
 ξ_o value for Direct Delta functions in Gaussian distribution.
 ξ^* critical size of clusters
 $\xi_m(t)$ unit mass of monomer
 $\xi_{\text{max}}(\theta)$ maximum value of cluster mass

ABSTRACT

A complex system is defined as a system with many interdependent parts having emergent self-organization; analyzing and designing such complex systems is a new challenge. A common observable structure of many complex systems is the network, which is connections among nodes, and thus inherently difficult to describe. The goal of this research is to introduce an effective methodology to describe complex systems, and thus we will construct a population balance (distribution kinetics) model based on the association-dissociation process to describe the evolution of complex systems.

Networks are commonly observed structures in complex systems with interdependent parts that self-organize. How networks come into existence and how they change with time are fundamental issues in numerous networked systems. Based on the nodal-linkage distribution, we propose a unified population dynamics approach for the network evolution. Size-independent rate coefficients yield an exponential network without preferential attachment, and size-dependent rate coefficients produce a power law network with preferential attachment.

For nonlinearly growing networks, when the total number of connections increases faster than the total number of nodes, the network is said to accelerate. We propose a systematic model, a population dynamics model, for the dynamics of growing networks represented by distribution kinetics equations, and perform the moment calculations to describe the dynamics of such networks.

Power law distributions have been observed in numerous physical and social systems; for example, the size distributions of particles and cities are often power laws. Each system is an *ensemble* of clusters, comprising units that combine with or dissociate from the cluster. To describe the growth of clusters, we hypothesize that a distribution

obeys a governing population dynamics equation based on reversible association-dissociation processes. The rate coefficients considered to depend on the cluster size as power expressions provide an explanation for the asymptotic evolution of power law distributions.

To mathematically represent human-initiated phenomena, which recently recognized as power law distributions, we apply the framework of cluster kinetics to the study of waiting-time distributions of human activities. The model yields both exponential and power law distributed systems, depending on the expressions for the rate coefficients in a Fokker-Planck equation.

CHAPTER 1. INTRODUCTION

1-1. Complex Systems

Due to the rapid addition of new information and innovations in science and technology that occur daily, an engineer must continuously expand his or her perspective, and indeed technological developments force engineers to apply their skills to a wide range of topics. Engineers work in fields that are, for example, biological, genetic, environmental, medical, physical and social. In many of these fields, analyzing and designing complex systems is a new challenge.

A complex system is a system with many interdependent parts having collective complex characteristics: self-organization and adaptability. Complex system study in a unified framework has become a recent scientific interest and is recognized as a new discipline, notable for interdisciplinary research. Many systems surrounding us are complex. In spite of the complexity and variety of the systems, an essential aspect of a complex system study is universal law. Many scientific attempts are based, to a greater or lesser degree, on the existence of universality, which manifests itself in diverse ways. Extracting the universality, as a part of complex system studies, can enhance our ability to understand complexity.

The systems in nature ranging from atomic, cellular, to biological, social, physical and chemical systems consist of many parts depending on the degree of complexity in the systems. The primary questions are why and how parts of a complex system are intrinsically related to the nature of the system. Simple systems may also consist of many parts but have a smaller number of interactions without showing collective complex behaviors. To qualitatively understand complex system behavior, we should understand

and analyze how system parts act together to produce desired functions as a whole, not just the behavior of the parts. To describe the whole complex system, it is, however, necessary to describe each part and the interactions with other parts, and these are what make it difficult to understand and analyze complex systems.

In the study of complex systems, how is it possible that well separated fields such as biology and physics can become unified in a single discipline? We may answer the question by study of universal properties (principles), which is particularly important for the complex system study. Though universal principles are usually considered intuitive and not explicitly specified, careful consideration of such principles can help us approach complex systems systematically. We have applied this systematic approach throughout our studies on complex systems, from networks to general human dynamics.

The purpose of this chapter is to introduce complex systems and their general concepts without detailed method and mechanisms biased as to conclusions. To initiate the complex system study, we will consider examples, quantities, and mechanisms that are relevant to the study as a part of the dissertation. Let us first look at some examples of complex systems. To help understand their properties, consider actual systems ranging from physical and chemical to ecological, biological, and social, for example, human brain and body, individuals, families, computers, the Internet, weather, the Earth, and the universe. These systems are categorized focusing on functions, structures, and diverse expressions. In addition, time also plays an important role in complex systems. Thus, the properties of complex systems are birth, change, growth, and death, which are possible components of a life cycle. We propose to apply a well-defined theory to analyze and model such systems with the form of a life cycle: population balance dynamics.

Adaptability, an important characteristic of complex systems, can be extracted as a result

of the cycle changes with time and the interactions between complex system and its environment. Population balance models, formulated for chemical engineering purposes, are very powerful methods to describe systems with underlying structure; these models are widely used to describe and control a wide range of particulate processes including crystallization, combustion, polymerization, etc. In general, these models refer to distributed systems in which the distributed particles that form an interaction with each other, such as addition, breakage, aggregation, and de-aggregation. Moreover, these models describe birth and death processes, which take place, for example, when monomers are added to or removed from the polymer. Population balance models, generally governed by integrodifferential equations for the dynamics of the population, can be used to address a range of problems of interest.

How is the complexity of the whole related to the complexity of the parts? Relationships between parts and the whole are essential to organize complex system properties. Though some systems show simplicity if considered in a macroscopic viewpoint, the systems may also display complexity if considered from a microscopic viewpoint. For example, the Earth orbiting the Sun can be considered as a simple system in a macroscopic viewpoint. However, the Earth can also be considered as a cluster of numerous components that are somehow related from a microscopic viewpoint. We can, therefore, appreciate that a system can be both simple and complex according to the viewpoints. In this regard, the possibility where a system composed of simple parts shows complex collective behavior is called emergent complexity, and any complex system formed out of atoms can be an example. Also, the possibility where a system composed of complex parts displays simple collective behavior is called emergent simplicity, and a planet orbiting around a star is a useful example; the orbiting behavior

of the planet is quite simple, even if the planet is the Earth, a planet with many complex systems upon it. This illustrates that the collective system has a behavior at a different scale than its parts. On the smaller scale the system may behave in a complex way, but on the larger scale all the complex details may not be relevant.

We begin to describe complex systems beginning with identification of properties important to the systems. The components of complex systems are numbers of parts, interactions among them, formation and operation with time, diversity, relationship to environment, and activities and objectives. Complex systems can be described by several ways using words, illustrations, audio or video recordings, and we define the system itself as a separate part of universe distinguished from environment, the rest by an imaginary boundary. A simple kind of emergent property of a system appears as patterns and interdependence, the tendency systems and environments to interact. Thus, parts of a complex system are working together to produce desired functions. One of the useful ways to probe complex system behavior is examine how a system responds to an applied force or change, for example, a node or link addition or removal from a network. Self-organization properties make a complex system stable by evolution through interactions among many interdependent parts. Robustness is what makes the complex system strong against damages or failures of its parts, and appears as a result of such interdependent interactions. Based on these complex system properties, adaptability arises, the ability of complex systems to adapt to external stimuli or failures of parts. We will model such complex systems with the described properties based on population balance dynamics.

As described, complex systems consist of many parts and interactions among them, and each part may have complex or simple characteristics. We may question how the complexity of the system can be related to their parts, why we should analyze the

systems as the whole, and how and where the complexity emerges. The emergent complexity is the idea that many simple parts interact in such a way to make the behavior of the system complex. Consider the two types of complex systems, where the characteristic components are either complex or simple. For the system where the components are already complex, it is easy to say the system is complex. However, if a system composed of simple parts shows collective complex behavior, how can the emergent complexity of the systems be related to the complexity of the parts? We will explain the answer through the dissertation.

Before we go to the next step, we will define some characterization of complex systems such as space, time, self-organization, and complexity, and consider some of questions rising from complex system studies. Many complex systems have their own responses to the stimuli from the environment that require their internal structure change. We question how the structure of the system responds, and when will dynamic processes reach an equilibrium state. Most basically, how does the complex system come into existence? What kind of dynamic processes give rise to complex systems? How do such processes develop to self-organize the systems? We are going to answer these questions through other chapters. The dissertation consists of research on networks, power law distributions, and human dynamics, and the aim is to discuss the complex system properties in the context of specific examples. Therefore we do not attempt to cover the entire fields of networks, (power law) distribution systems, and human dynamics, but we do provide a mathematical framework for their study.

The concepts of emergence and complexity, once understood, reveal the context in which universal properties of complex systems arise. Specific universal phenomena, such as the evolution of networked or biological systems, can then be better grasped.

What make systems complex and what is complexity? The primary issue is how we define complexity quantitatively. Researchers have used statistics, dynamics and computer simulations to quantify complexity. There have been recent book-length efforts to define complex systems. As a part of recent efforts, Ottino [Ottino, 2003] defines a complex system as a system composed of many parts and the interactions among them, whose behavior cannot be simply understood from the behavior of its parts. Another effort [Backlund, 2002; Waldrop 1992] states that the complexity of a system can be measured by the amount of information necessary to describe the collective behavior of a system.

To understand complex systems, it is necessary to recognize that simple parts should somehow, in large numbers, give rise to collective complex behaviors. The most simple and basic question that the complex system study faces is how and when this occurs. To approach the problem, consider the term, “emergence.” When collective behavior appears in a small part of the system, the concept of emergence arises because the collective behavior is not easily understood from the behavior of the parts. It also arises when collective behavior pertains to the system as a whole, and this is particularly relevant to the study of complex systems. An example of emergent property is system pressure or temperature, which becomes relevant only when the system contains many particles together. Another example of emergent property is the formation of water from hydrogen and oxygen atoms. The properties of oxygen and hydrogen molecules are not apparent as properties of water molecules. But the properties of water are not independent of the properties of components. In the study of complex system, we are mainly interested in more complicated types of emergent properties, though careful mathematical treatments are required to appreciate and understand them.

To help understand the concept of emergence in complex systems, we consider a network as an example. If a network consists of small number of nodes with simple connections, simple emergent behavior is the outcome. However, if a network consists of large number of nodes with many types of connections, complex emergent behavior will arise if a network is sufficiently rich in nodes and connections. In this example, if a hub, a node with many connections, is removed, the network may lose the ability to function properly. This kind of behavior is what characterizes emergent properties. Complex emergent properties can be studied by looking at each of the parts in the context of the system as a whole, not by taking a system apart and examining the parts. If the behavior of the small part, where it is a part of the larger system, is different in isolation, a complex emergent property will arise. If we think about the system as a whole, rather than the small part of the system, we can identify the system that has a complex emergent property as being formed out of interdependent parts. The term interdependent should be distinguished from the term interconnected, because the term interconnected does not pertain directly to the influence one part has on another. It is also distinct from the term interacting, because even strong interactions do not necessarily imply interdependency of behavior. Therefore, we can characterize complex systems through the effect of removal of part of the system, though it is not easy to describe for systems such as networks. The possibility most appealing as a model of complex system is that its properties are also affected by the removal of a part. Such a system has a collective behavior depending on the behavior of all of its parts, and this concept will become more precise if we quantitatively measure the complexity. As mentioned, the amount of information needed to describe a system is the complexity of the system, and provides how complex a system

is. The complexity of the whole system must involve a description of the parts, if the behavior of the system depends on the behavior of the parts.

Complex systems are not far from the traditional concerns of chemical engineers. Almost all engineering systems are composed of many different interdependent parts. Because of this interdependency, most systems inevitably have complex characteristics, and we therefore call them “complex systems.” The interactions among elements may occur with immediate neighbors or distant ones. According to Ottino [Ottino, 2003], a common characteristic of all complex systems is that they show organization without conforming to any external rule, and adaptability and robustness are often byproducts of their organization. Because of these characteristics, if a part of the system works improperly, the system may still function properly. A key characteristic of complex systems, by this argument, is adaptability, so complex systems spontaneously respond to external stimuli, for example, species survival in changing ecosystems. These complex systems can be broadly categorized as physical and chemical systems, biological systems, and social systems and organizations. It seems obvious that chemical engineers, who are exposed to a wide range of time and space scales and are trained to think in terms of systems, can grasp the opportunity to take a leadership position in the area of complex system research.

Complex systems can be specified by what they do and how they can be analyzed. Metabolic pathways, ecosystems, the Internet, the World-Wide-Web, highways, the US power grid and the propagation of infections are examples of complex systems that already have a great impact on our lives. Before continuing, we should distinguish between complex and complicated systems. For a complicated system, every single part, no matter how many or how elaborate, can be understood by knowing how the single

parts of the system work together to produce desired functions. For the most elaborate mechanical machine, a failure of any single part of this complicated system can cause a serious malfunction. In other words, complicated systems do not adapt or self organize against external or internal variations. As explained, complex systems cannot be well understood in isolation. Interactions between parts and the overall functions that emerge from the interactions are the intrinsic nature of the complex systems. Therefore, complex systems have to be analyzed as a whole and with respect to adaptation.

We consider that engineering is about optimum design and consistency of operation, assembling pieces that produce desired functions. As engineers, our question is how to analyze and design complex systems. Based on knowledge and experience, engineers need to build complicated systems having characteristics such as adaptation and self-organization, so called, complex systems. In designing a process, engineers always balance between performance and risk. These two criteria, high efficiency and low risk, are mostly in conflict with each other. It is difficult to keep high efficiency without risk, because usually an efficient state is a high-risk state. From this perspective, if we can design complicated systems having adaptation and self-organization characteristics, in other words, if we can design complex systems, then we can operate systems at optimum conditions – high performance and safety.

These complex systems can often be expressed as networks that are inherently difficult to describe. Networks are composed of nodes and links, such that properties of complex systems evolve with their basic components. First, nodes are not identical, there are many different kinds of nodes, and each node can vary in time. Second, the links among nodes could have different length, weight, direction and sign, and they can also vary in time. For example, synapses in the nervous system can be strong or weak,

inhibitory or excitatory. Third, network wiring diagrams could be changed in numerous ways: nodes can be inserted or removed from the networks, and links can be lost or created among existing or introduced nodes. Unfortunately, many real complex systems are beyond present mathematical analysis. For such cases, we need to begin with a structural or topological approach.

Most traditional engineering process designs have multiple configurations but, once finalized, the process does not adapt or self-organize. Nevertheless, engineers need to have insight into complex systems because of their growing importance. From this perspective, the most substantial theory that can be applied to design complex systems is network theory.

Our purpose in studying complex systems is to extract general principles. General principles can be many forms. However, most of them are expressed as relationships between properties, and will be quantitatively expressed as equations. Therefore, mathematical modeling based on dynamic theory is required to come up with such equations. To model complex systems, there are some rules and simplifications we should follow. The first, complex systems should be analyzed as the whole, since interactions between parts of a complex system are essential to understand its behavior. The second, much of the quantitative study of complex system cannot be described by a uniform model, different non-linear static and dynamic models may be used. The third, the study of complex system behavior should be focused on many independent parameters at the same time, not focusing on only one or two parameters.

Among many approaches, two types are frequently used for studying complex systems. The first approach is a method that identifies and describes parts as well as interactions among them for a specific system. The objective is to show how the behavior

of the system emerges from them. Another type of approach considers how the essential properties of such systems are described. Statistical analysis can be used to obtain properties and describe behavior of the systems. The first type of approach is used for the work. In the text, we introduce population balance dynamics and the approaches they are based on.

1-2. Networks

The objective of the present network research is to apply population balance dynamics to complex evolving systems. A common characteristic of many complex systems is that they have a network structure. For the description, analysis, and understanding of these complex systems, network theory has appeared recently as a unifying concept with great potential for applications to a wide range of phenomena.

Many systems can be seen as networks. For example, in a polymerization reaction, each monomer and molecular bond can be referred to as a node (edge) and a connection (link), respectively. The World Wide Web, food webs, metabolic pathway, and protein networks within cells are examples of networks. Species are connected by predator-prey relationships in food webs, and molecules are connected by reactions in chemical networks. Metabolic pathways and eco-systems are biological networks, whereas the Internet is an example of a human-created network. Propagation of viruses, including HIV infection, exemplifies a biological and sociological network. The connections among nodes make up the observable or underlying structure for numerous physical and social systems, and structure always affects function, for instance, the structure of social networks affects the spread of information or disease. Systems of metabolic reaction pathways, food webs, and pipelines are physical examples; acquaintanceships, viral contacts, commodity trade, and scientific collaborations are social examples of networks.

Before it was realized that they share similar architectures, these systems seemed not to have anything in common.

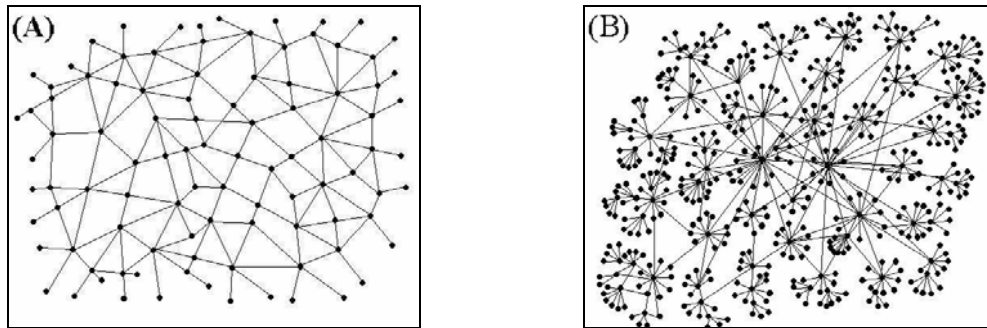


Figure 1-2.1 Two types of networks: (A) random (exponential) and (B) scale-free (power law) networks: points represent nodes and lines represent connections between them.

Networks can be specified into two main categories; (1) random or exponential networks (single scale) – the number of links per node follows a Gaussian, Poisson, or exponential distribution, and (2) power law networks (scale-free) – the number of links per node follows a power law. Figure 1-2.1 shows two representative network structures schematically: (a) exponential networks – nodes are connected exponentially, (b) power law networks – connections per node follow a power law distribution. It may be helpful to think of the analogy to road maps and airline connections.

In general, networks are not static but evolve with time. How networks come into existence and how they change with time are fundamental issues in many applications. Networks are usually growing, but also sometimes disintegrate and possibly vanish due to random breakage or intentional attacks. Polymers, likewise, have large numbers of repeating units (monomers) making up their chain-length distributions, and change with time. Crystals undergoing growth or dissolution also are composed of many molecular

units. Polymer reaction kinetics and crystallization dynamics are typically formulated as population balance equations governing the statistical properties of the molecular-weight distribution. We propose that networks analogously have statistical properties that can be computed by population dynamics (distribution kinetics) modeling.

1-3. Power Law Distributions

Distributions in nature, economy, and society that consist of a small number of rare events and a large number of common events often present a regular power law form. Likewise, many human created and naturally occurring phenomena are distributed following a power law distribution. The popular event can have hundreds, thousands or even millions of relationships among the events. For instance, scale-free networks, which present no characteristic length, contain hubs, nodes with many links, and the distribution of node linkages follows a power law.

Power law distributions have been observed and investigated recently and characterize numerous systems such as city sizes, personal incomes, word frequencies, earthquake magnitudes, aerosol masses, and many others in the areas of biology, chemistry, linguistics, economics, and computer science. A power law distribution appears as a straight line on a log-log plot.

Power law networks are composed of many nodes with a few connections and a few nodes with many connections, usually called hubs. Hubs are an essential feature of power law networks, such as Yahoo or Google in the World Wide Web and ATP (Adenosine Tri-Phosphate) in metabolic networks.

A power law in complex networks can be established based on a mechanism of growth with preferential attachment. Growth means that the network emerges through the addition of new links and nodes. Preferential attachment means that nodes prefer to link

to more connected nodes allowing highly connected nodes to acquire new links faster than less connected nodes. These two mechanisms are essential for network evolution and generation of hubs through a “rich get richer” phenomenon, producing power law networks. Figure 1-3.1 schematically express the birth and growth processes with preferential attachment of a power law network.

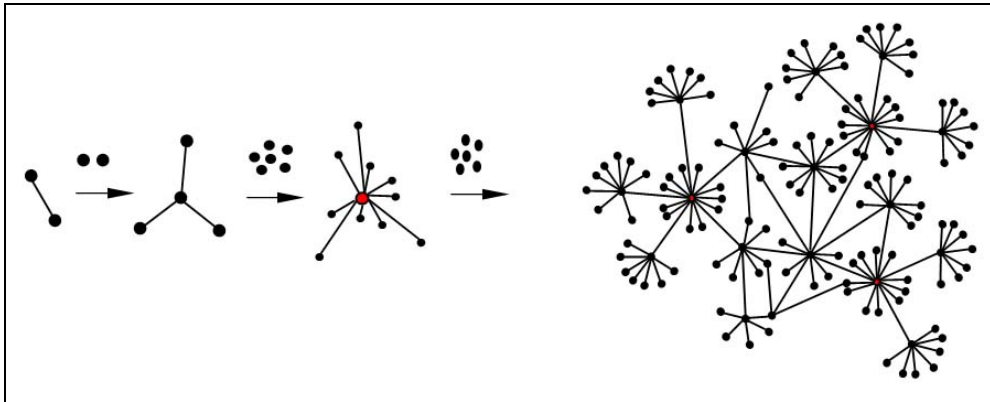


Figure 1-3.1 Schematic representation of birth and growth of a power law network

1-4. Accelerating Networks

For complex systems with interacting and interdependent parts that self-organize, a common observable structure is a network composed of many connections among many nodes. Most network studies have focused on relatively simple connected systems such as phone exchange server or the Internet. These networks are scale-free in that their structures in terms of the average number and the degree distribution of their connections per node show little change as they grow. For functionally well-organized systems such as stock exchanges and protein network controlling gene expression, operation of such systems depends on the activity of the connected nodes. The number of connections per node should increase with the size of network. In such networks, the total number of

connections between nodes has to be increased faster than the total number of nodes, in other words, it has to be accelerated. Many natural or man-made networks under the category usually grow with time. A majority of them show non-linear growth where the total number of connections increases faster than the total number nodes; such networks are called “accelerating networks.” The goal of this study is to describe the dynamics of such accelerating network growth.

The moments correspond to the properties of non-directional networks with finite number of nodes and connections; if the general n^{th} moment is expressed as $p^{(n)}(t)$, the total number of nodes is $p^{(0)}(t)$, the total number of connections is $\frac{1}{2} p^{(1)}(t)$, and degree distribution, the average number of connections per node, represented by the average moment is $\frac{1}{2} p^{\text{avg}}(t)$, where $p^{\text{avg}}(t) = p^{(1)}(t)/p^{(0)}(t)$.

In this study, we will propose the comprehensive and systematic model for the dynamics of growing networks, either exponential or power law networks, in the context of their kinetics represented by distribution dynamics equations. We will study the accelerating networks by following steps such as; define the nodal-linkage distribution, $p(\xi, t)d\xi$, construct a population dynamics equation based on the association-dissociation process with the proposed rate coefficients, $k_g(\xi) = \gamma\xi^\lambda$ and $k_d(\xi) = \kappa\xi^\lambda$, and perform the moment calculations to describe the dynamics of such networks. Depending on the power in the coefficients, the model with the rate coefficients will describe both exponential network in the absence of preferential attachment and accelerating power law network with preferential attachment accounting for the accelerated growth.

1-5. Human Dynamics

As an example of complex systems, we study human dynamics based on a deterministic distribution kinetics approach. Human activities are somehow connected

and perhaps get together creating social or characteristic clusters to produce desired functions. We study the dynamics of collective human activities based on a deterministic distribution kinetics approach, and find that it develops power law structures similar to those appearing in many nonlinear dynamic systems.

Understanding human activity patterns is essential for some problems of practical interest such as cell phone or the internet server design, products and inventory control strategies, etc. In human dynamics, however, extracting regularities is very difficult except the obvious daily behavior and seasonal periodicities. Unlike physical or chemical sciences, which would be commonly described by accurate calculation tools, predicting patterns of human actions and social behavior is often trivial.

By distinguishing characteristics, the timing of human activities can be classified as two categories; types of activities executed independently and dependently of each other. The patterns of human activities such as sending emails or making phone calls are commonly modeled by the Poisson process showing exponential distribution. Increasing empirical evidence reveals that such human actions are well characterized by a power law distribution providing a better quantitative description. Most human initiated activities are not independent of others. For instance, in task executions, since the selection of one task also implies the exclusion of others, some tasks with low priority should wait to be executed, and therefore, the distribution of waiting times in job performing processes can be well described not by Poisson processes but by power law distributions.

To mathematically represent such human behavior, we apply the framework of cluster kinetics to the study of waiting-time distributions of human activities. The model yields both exponential and power law distributed systems, depending on the expressions for the rate coefficients in a Fokker-Planck equation. A derived truncation power law

quantitatively describes the observed waiting-time distribution data for email and printing server systems.

CHAPTER 2.

DISTRIBUTION DYNAMICS OF EVOLVING NETWORKS

2-1. Population Balance Dynamics

Many systems of engineering interest are composed of entities that are distributed with respect to a property, and continuous distribution kinetics can be applied when there are many entities. We propose that such systems can be described by a distribution whose temporal and spatial variance is governed by a population balance dynamics. The population balance dynamics can describe and apply to time evolution of multivariate distribution reactions such as branched macromolecules, complex polymer mixture systems, and so on. For these systems, entities like molecules can combine randomly and break simultaneously to smaller sizes that may be distributed randomly or nonrandomly.

Population balance models can describe a broad range of dynamic behaviors and are suited for processes undertaken in groups of entities that have individual properties. Regarding particulate systems, two important variables of population balance dynamics are time and any property or constituent to which a conservation law is applicable such as mass, volume, etc.

Kinetics and dynamics of many complex systems can be expressed as population balance dynamics applied to networks. Based on the concept of a nodal linkage distribution, we propose a unified population dynamics approach for the evolution of networks to random or power law conformations. The functional form of the rate coefficients for addition or removal of links usually governs the asymptotic forms, which are independent of initial states. Based on the population balance equations, we propose

kinetic relationships, moment and distribution solutions, and continuity equations to represent the network structure and dynamics. We focus on exponential and power law networks.

The population balance equation, cast either as an integrodifferential equation, a difference-differential equation, or a partial differential equation, can be solved by standard methods, including moment techniques. The large-scale properties of the network can be formulated as moments of the distribution, such as total number of nodes, total number of connections, and average number of connections per node. The moments are solutions of ordinary differential equations in time, and are particularly useful for random networks. Power law networks have an intrinsic nonlinear character and require an approach different from random networks. The addition or removal of connections can be written as a reaction-like reversible process. The growth and dissolution rate coefficients that are used with a power of the linkage number yield an asymptotic power law distribution. Their asymptotes depend on the power form of rate coefficients under appropriate boundary conditions for the first order partial differential equation. Rate coefficients, independent of linkage number, yield exponential networks, the Poisson or the Gaussian distribution networks. The first order partial differential equation from the population balance equation yields the power law networks, which display a temporal evolution that depends on their initial and boundary conditions.

We are guided by experience in distribution kinetics developed through population dynamics equations, which has proven a productive approach to polymerization and depolymerization [McCoy and Madras, 2001; Sterling and McCoy, 2001], particulate fragmentation and aggregation [Madras and McCoy, 2002a], and crystal growth and dissolution [Madras and McCoy, 2002b].

2-2. Networks

Networks are connections among nodes that make up the observable or underlying structure for numerous complex physical and social systems [Albert and Barabasi, 2002; Strogatz, 2001]. Systems of metabolic reaction pathways, food webs, pipelines, telephone lines, highways, and railroads are physical examples; acquaintanceships, viral contacts, commodity trades, and scientific collaborations are social examples of networks. Network theory, which has emerged recently as a unifying concept for complex systems, has great potential for applications to a wide range of phenomena [Ottino, 2003]. The two classes of networks are random (exponential) and scale-free (power law) networks, which were schematically illustrated in Fig. 1-2.1. In general these networks are not static, but evolve with time [Barabasi and Albert, 1999; Strogatz, 2001; Albert and Barabasi, 2002; Barabasi et al., 2002], often growing, but also sometimes dissipating and possibly vanishing due to accidental or intentional breakage of links. A quantitative understanding of how networks come into existence and how they change with time is desirable for recognizing cause and effect in these strategic systems.

The aim of the present work is to discuss the dynamics of networks in the context of their kinetics represented by distribution dynamics equations (also called master equations or population balances). Significant work has appeared recently on this issue. For example, in a study [Barabasi et al., 2002] of the temporal evolution of networks of scientific collaborations, extensive data were analyzed and a master equation was proposed to represent the network structure and dynamics. The approach showed how either discrete or continuous mathematics can describe network dynamics. Quantities such as node separation and clustering coefficients could not be described by the model, and were simulated by Monte Carlo calculations. An essential feature of power law

network growth is the preferential linkage to nodes already well connected [Barabasi and Albert, 1999]. Random network evolution, on the other hand, has been described by probabilistic arguments [Erdos and Renyi, 1960; Strogatz, 2001; Albert and Barabasi, 2002]. For a random network (graph) with a constant number of nodes, if the number of links among nodes is small the network is composed of separate clusters of nodes. As the number of links increases, the clusters grow by linking, eventually coalescing into a single interconnected cluster [Erdos and Renyi, 1960; Strogatz, 2001]. Although similarities are apparent in the different approaches for power law and random networks, a generic theory has not yet emerged.

Degradation or disintegration of networks is of current interest [Albert et al., 2000; Dorogovtsev and Mendes, 2001a], with examples in collapse of electrical power networks, cybernetic attacks to the Internet, and environmental and ecological deterioration. The approach outlined here provides some insights into such network breakage, but complete models are much more difficult than for network growth. Similar difficulties are encountered for particle fragmentation and polymer degradation, where the representation of breakage kernels is quite distinct from growth or aggregation kernels [Kodera and McCoy, 2002]. Thus, although we are unable to solve completely the problem for power law networks, we offer useful solutions for random network disintegration.

The current objective is to describe the time evolution of a general network, either random or power law, in which nodes are being added or removed and connections between nodes are being established or eliminated at given rates. To illustrate fundamental ideas, we will see that even elementary models yield a rich variety of behaviors. Thus for now, we consider connections (links or edges) as binary interactions

(nondirectional and of indeterminate length) between nodes. We define the nodal-linkage distribution so that $p(\xi,t)d\xi$ is the number of nodes at time t with number of connections in the interval ξ to $\xi+d\xi$. Even though $\xi \geq 0$ are integers, for a large number of connections we can treat the distribution as a continuous function of ξ (a discrete distribution would replace integrals with summations). The distribution of the number of links is given by $\frac{1}{2}\xi p(\xi,t)d\xi$, which is the number of connections in the interval ξ to $\xi+d\xi$. (The analogy with polymer molecular weight distributions [Erdos and Renyi, 1960] is useful in explaining this concept. If $p(x)dx$ is the number of macromolecules having mass in the interval $(x, x+dx)$, then $xp(x)dx$ is the mass of macromolecules in the same interval.) Because we consider non-directional networks, each connection is associated with two nodes, hence the factor of $1/2$. The moments of the nodal-linkage distribution are defined as

$$p^{(n)}(t) = \int p(\xi, t) \xi^n d\xi \quad (2-2.1)$$

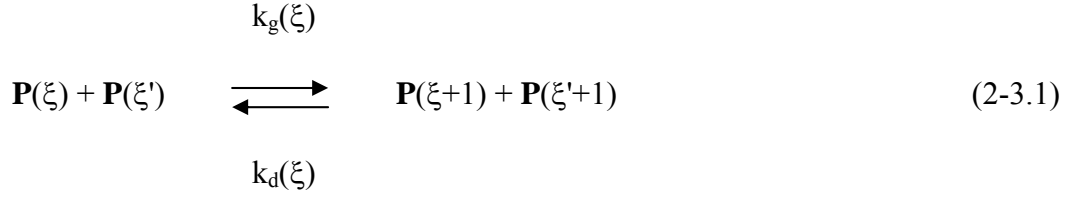
where the integration limits are determined by the domain of $p(\xi,t)$. The total number of nodes is $p^{(0)}(t)$, the total number of connections is $\frac{1}{2} p^{(1)}(t)$, and thus the average number of connections per node is $\frac{1}{2} p^{\text{avg}}(t)$, where $p^{\text{avg}}(t) = p^{(1)}(t)/p^{(0)}(t)$. Higher moments provide further information about the character and shape of the distribution. The variance is $p^{\text{var}} = p^{(2)}/p^{(0)} - p^{\text{avg}^2}$, and $p^{\text{var}}/p^{\text{avg}^2} = p^{\text{pd}} - 1$, where the polydispersity index is $p^{\text{pd}} = p^{(2)}p^{(0)}/p^{(1)^2}$.

We will focus on the two classifications: random (single-scale) and power law (scale-free) networks (Fig. 1-2.1). For random networks, the distribution $p(\xi)$ is unimodal (peaked) with well-defined moments so that statistical properties such as mean and

variance can be defined and measured, e.g., Gaussian, binomial, or Poisson distributions. Scale-free networks have a power law form, $p(\xi) \sim \xi^{-\lambda}$, where λ is a constant. Such networks lack an inherent scaling factor because moments are not defined on the interval $(0, \infty)$. As we will show, however, when the evolving power law distribution has a finite domain, the integral, Eq. 2-2.1, can be defined. The aim is to develop a framework that determines the evolution of the two network types by a systematic and consistent approach. We are guided by experience in distribution kinetics developed through population dynamics equations, which has proven a productive approach to polymerization and depolymerization [Sterling and McCoy, 2001; McCoy and Madras 2001], particulate fragmentation and aggregation [Madras and McCoy, 2002a], and crystal growth and dissolution [Madras and McCoy, 2002b]. This approach follows a tradition of chemical engineering science; fundamental relationships are defined, general principles are explained, and governing differential equations are written for the hypothesized model. An attribute of this method is that analytical solutions are possible for numerous interesting cases. These solutions show clearly the effects of parameters that govern the network evolution rate. The algebraic computations for these solutions would be extremely tedious and difficult, however, if attempted by hand. Therefore, all work described here was done using a computer algebra software (Mathematica)

2-3. Distribution Kinetics

We consider links added one at a time to available nodes, allowing for the possibility that connected nodes, or indeed entire networks, might coalesce by such linking processes. The addition or removal of connections can be written as a reversible rate process,



where a node with ξ connections is schematically represented by $\mathbf{P}(\xi)$. The formation of a connection between two nodes adds a single link between them. The rate coefficients for addition (growth) and removal (dissociation) are $k_g(\xi)$ and $k_d(\xi)$, respectively, considered in general to depend on the number of connections. In the present work we propose to use power expressions for the rate coefficients,

$$k_g(\xi) = \gamma \xi^\lambda \quad \text{and} \quad k_d(\xi) = \kappa \xi^\nu \tag{2-3.2}$$

where the constants γ , κ , λ , and ν are positive definite. Equation 2-3.1 suggests that either addition or removal of a connection must involve two nodes, and will increase with the distributions (or densities) of these two available nodes, thus, second-order kinetics will apply. The process of Eq. 2-3.1 is unchanged if ξ is replaced with $\xi-1$ or ξ' is replaced with $\xi'+1$. To construct the population dynamics equation we need expressions for rates of generation or loss of nodes as connections are made or broken. Formulating the governing equations for networks, polymers [Sterling and McCoy, 2001], and crystallization dynamics [Madras and McCoy, 2002b] have points of similarity.

Formation of a link between two nodes is more probable if $\mathbf{P}(\xi)$ and $\mathbf{P}(\xi')$ are in greater abundance. Similar to mass-action reaction kinetics formulation of bimolecular rate expressions, or to aggregation kinetics, this leads to a second-order rate of linkage growth for networks. For example, the loss of $\mathbf{P}(\xi)$ on the left-hand side of Eq. 2-3.1 is the product of $p(\xi,t)$ and $p(\xi',t)$, with all possible partners having connections ξ' being accounted to give the zeroth moment,

$$-k_g(\xi)p(\xi,t) \int p(\xi', t) d\xi' = -k_g(\xi)p(\xi,t)p^{(0)}(t) \quad (2-3.3)$$

Likewise, the removal of a link between $\mathbf{P}(\xi)$ and $\mathbf{P}(\xi')$ is proportional to $p(\xi,t)$ and $p(\xi',t)$, and hence second-order. The rate of insertion (nucleation) or removal of nodes with ξ_i connections into the network is $I_i(t) \delta(\xi-\xi_i)$. If unconnected nodes are introduced, i.e., $\xi_{i=1} = 0$, then the Dirac delta ensures that these inserted nodes have no connections. The response to loss of nodes, including major hubs (nodes with many connections, $\xi_i \gg 1$), would be modeled with a negative rate, $I_i(t) < 0$. We sum over a finite number of different functions, $I_i(t)\delta(\xi-\xi_i)$, that can possibly affect the network evolution.

With these preliminary concepts we can write the population balance (or distribution dynamics) equation for Eq. 2-3.1 with generation and loss terms,

$$\begin{aligned} \partial p(\xi,t)/\partial t &= -k_g(\xi)p(\xi,t)p^{(0)}(t) + k_g(\xi-1)p(\xi-1,t)p^{(0)}(t) \\ &\quad -k_d(\xi)p(\xi,t)p^{(0)}(t) + k_d(\xi+1)p(\xi+1,t)p^{(0)}(t) + \sum_{i=0} I_i(t) \delta(\xi-\xi_i) \\ &= \gamma p^{(0)}(t)[(\xi-1)^\lambda p(\xi-1,t) - \xi^\lambda p(\xi,t)] \\ &\quad + \kappa p^{(0)}(t)[(\xi+1)^\nu p(\xi+1,t) - \xi^\nu p(\xi,t)] + \sum_{i=0} I_i(t) \delta(\xi-\xi_i) \end{aligned} \quad (2-3.4)$$

where in the second equality we have substituted Eq. 2-3.2. Equation 2-3.4 has a form similar to a master equation, except that it displays second-order kinetics whereas master equations usually have first-order kinetics [Kampen, 1992]. Related population dynamics equations for crystallization [Madras and McCoy, 2002b] or polymerization [Madras and McCoy, 2002a] describe growth by monomer addition for clusters or polymers, and serve as examples of how distribution kinetics can be applied to physical and chemical processes. As in other applications of continuous distribution kinetics [Madras and

McCoy, 2002a; Madras and McCoy, 2002b], the distribution in $\xi_{\pm 1}$ can be expanded in a series around ξ so that Eq. 2-3.4 is replaced by a Fokker-Planck (continuity) equation,

$$\begin{aligned} \partial p(\xi, t) / \partial t = p^{(0)}(t) \partial [(k_d(\xi) - k_g(\xi)) p(\xi, t)] / \partial \xi \\ + \frac{1}{2} p^{(0)}(t) \partial^2 [(k_d(\xi) + k_g(\xi)) p(\xi, t)] / \partial \xi^2 + \dots + \sum_{i=0} I_i(t) \delta(\xi - \xi_i) \end{aligned} \quad (2-3.5)$$

where the ellipsis (...) represents omitted third- and higher-order terms.

2-4. Random Networks

We can substitute a new time variable, $d\theta = p^{(0)}(t) dt$, such that

$$\theta = \int_0^t p^{(0)}(t) dt \quad (2-4.1)$$

If we keep terms up to second-order and set source terms to zero, $I_i = 0$, Eq. 2-3.5

becomes

$$\partial p(\xi, \theta) / \partial \theta = \partial [(k_d(\xi) - k_g(\xi)) p(\xi, \theta)] / \partial \xi + \frac{1}{2} \partial^2 [(k_d(\xi) + k_g(\xi)) p(\xi, \theta)] / \partial \xi^2 \quad (2-4.2)$$

Because the number of nodes is constant in the absence of source or sink terms, $p^{(0)}$ is constant and $\theta = p^{(0)} t$. The resemblance of Eq. 2-3.4 to a one-dimensional random walk and its reduction to a convective diffusion equation, Eq. 2-4.2, suggests how a Gaussian distribution for a random network is obtained when the rate coefficients are constants [Chandrasekhar, 1943; Feller, 1957], $k_g(\xi) = \gamma$ and $k_d(\xi) = \kappa$. The convective diffusion equation can be expressed by substituting a "velocity," $v = (\gamma - \kappa)$, and a "diffusivity," $D = (\gamma + \kappa)/2$, into Eq. 2-4.2, as follows,

$$\partial p(\xi, \theta) / \partial \theta = -v \partial [p(\xi, \theta)] / \partial \xi + D \partial^2 [p(\xi, \theta)] / \partial \xi^2 \quad (2-4.3)$$

The exact solution can be obtained by Fourier transformation of Eq. 2-4.3 and the initial condition, $p(\xi, \theta=0) = p_0^{(0)} \delta(\xi - \xi_0)$, in terms of a Dirac delta such that initially each of

the $p_o^{(0)}$ nodes has ξ_o links. One boundary condition is $p(\xi \rightarrow \infty, \theta) = 0$, which means no node can have an unlimited number of links. The boundary condition $p(\xi \rightarrow -\infty, \theta) = 0$ is not realistic for $\xi \geq 0$. However, if the solution peak is far enough away from $\xi = 0$, then the peak is not influenced by the boundary. In convective diffusion theory [Chandrasekhar, 1943], the Peclet number is defined as $N_{Pe} = v \xi_o / D = 2(\gamma - \kappa)\xi_o / (\gamma + \kappa)$, and for $N_{Pe} \gg 1$, the boundary condition [Levenspiel and Smith, 1957] can be reasonably approximated by $p(\xi \rightarrow 0, \theta) = 0$, which means every node in the network has at least one connection. The Peclet number is large if ξ_o is large, where the initial distribution is positioned.

The solution [Levenspiel and Smith, 1957] for the convective diffusion equation, Eq. 2-4.2, is,

$$p(\xi, \theta) = p_o^{(0)} / (4\pi D \theta)^{1/2} \exp[-(\xi_o + \xi - v\theta)^2 / (4D\theta)] \quad (2-4.4)$$

Equation 2-4.4 is approximated by a Gaussian distribution for ξ and θ when $N_{Pe} \gg 1$,

$$p(\xi, \theta) = p_o^{(0)} / (2\pi(\gamma + \kappa)\theta)^{1/2} \exp[-(\xi_o + \xi - (\gamma - \kappa)\theta)^2 / (2(\gamma + \kappa)\theta)] \quad (2-4.5)$$

expressed in terms of rate coefficients. The moments of the both Eqs. 2-4.4 and 2-4.4 are readily found by integration (Eq. 2-2.1) with the results,

$$p^{avg}(t) = \xi_o + (\gamma - \kappa) p_o^{(0)} \theta \quad (2-4.6)$$

and

$$p^{var}(t) = (\gamma + \kappa) p_o^{(0)} \theta \quad (2-4.7)$$

Clearly, if $\gamma > \kappa$, nodes are being connected, the average number of links increases, and the network grows. If $\gamma < \kappa$, breakage of links occurs, the average number of links decreases, and the network deteriorates. For either growth or breakage according to Eq.

2-4.7 the network variance increases. This behavior is illustrated in Figs. 2-4.1A and 2-4.1B showing the growth and breakage, respectively, of random networks when $\xi_0 = 100$.

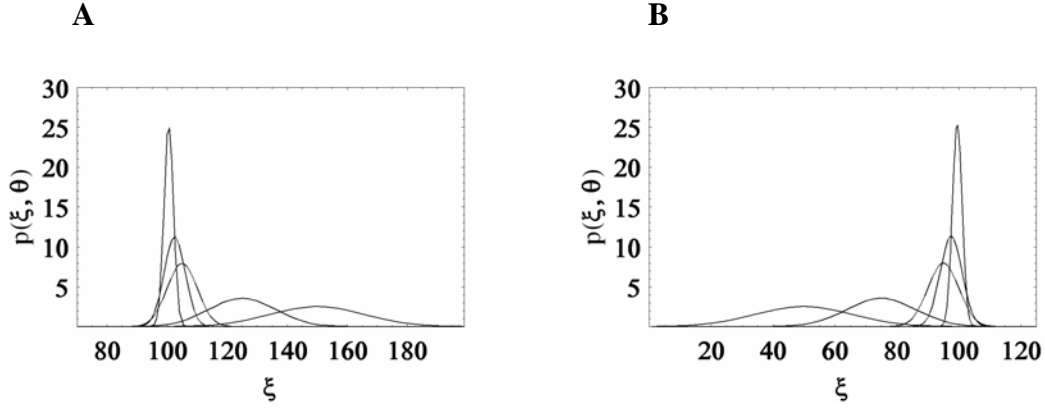


Figure 2-4.1 Evolution of the Gaussian distribution for (A) network growth ($\gamma = 1.5$ and $\kappa = 1.0$) and (B) breakage ($\gamma = 1.0$ and $\kappa = 1.5$). The initial distribution is a Dirac delta at $\xi_0 = 100$ with $p_0^{(0)} = 100$. Values of time shown are $\theta = 1, 5, 10, 50, 100$.

The discrete Poisson distribution for the random network derives from more restricted conditions. If network growth is irreversible ($k_d = 0$), source terms vanish, $I_i = 0$, and $k_g(\xi) = \gamma$, then Eq. 2-3.4 can be written as,

$$\partial p(\xi, \theta) / \partial(\gamma \theta) = -p(\xi, \theta) + p(\xi - 1, \theta) \quad (2-4.8)$$

where ξ takes only positive integer values. This is a first-order difference-differential equation similar (but not identical) to basic equations in chain polymerization [McCoy and Madras, 2001] and stirred-tank cascade modeling [Dotson et al., 1996]. The boundary and initial conditions are $p(\xi < 0, \theta) = 0$ and $p(\xi, \theta = 0) = p_0^{(0)} \delta_{0\xi}$, here expressed in terms of the Kronecker delta for the integer variable ξ . The solution, found by Laplace transformation, is closely related to a Poisson distribution [McCoy and Madras, 2001],

$$p(\xi, \theta) = p_0^{(0)} (\gamma \theta)^{\xi+1} e^{-\gamma \theta} / (\xi+1)! \quad (2-4.9)$$

The average is

$$p^{\text{avg}}(\theta) = \gamma\theta / (1 - e^{-\gamma\theta}) - 1 \quad (2-4.10)$$

which for long time ($\gamma\theta \gg 1$) is $\gamma\theta/2$. The variance is

$$p^{\text{var}}(\theta) = \gamma\theta e^{\gamma\theta} (e^{\gamma\theta} - \gamma\theta - 1) / (e^{\gamma\theta} - 1)^2 \quad (2-4.11)$$

which for large values of time is also $\gamma\theta/2$. Thus, like the Poisson distribution, for Eq. 2-4.9 the average and variance asymptotically reach the same expressions at long time.

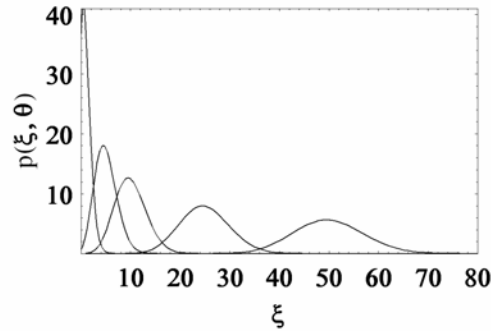


Figure 2-4.2 The discrete Poisson distribution for irreversibile network growth ($\kappa = 0$, $\gamma = 0.5$) at scaled time $\theta = 1, 5, 10, 25, 50$. The initial condition is $p_0^{(0)} = 100$ unlinked nodes ($\xi = 0$).

Figure 2-4.2 illustrates network growth as a distribution moving with time to larger values of ξ , with average and variance in accord with Eqs. 2-4.10 and 2-4.11. The size-independent rate coefficients thus allow either Gaussian or Poisson distribution solutions and explain the evolution of random networks.

Another way to obtain information from the population balance, Eq. 2-3.4, is to solve directly for moments. For integer values of λ and ν , a general moment equation can be derived by the operation of the moment definition, Eq. 2-2.1, on Eq. 2-3.4. The integrals are evaluated by substituting new integration variables for $\xi+1$ and $\xi-1$, and applying the binomial expansion before defining the moments. One obtains,

$$\begin{aligned}
dp^{(n)}/dt = & \gamma p^{(0)}[-p^{(n+\lambda)} + \sum_{j=0}^n \binom{n}{j} p^{(j+\lambda)}] \\
& + \kappa p^{(0)}[-p^{(n+v)} + \sum_{j=0}^n \binom{n}{j} (-1)^{n-j} p^{(j+v)}] + \sum_{i=0}^n I_i(t) \xi_i^n \quad (2-4.12)
\end{aligned}$$

where the binomial coefficient is defined as $\binom{n}{j} = n!/j!(n-j)!$. Consider the case of unconnected nodes being introduced or eliminated. Then $\xi_i = 0$ and ξ_i^n is replaced with δ_{n0} , which is 0 if $n > 0$, and 1 if $n = 0$. This indicates that only the zeroth moment is affected by insertion or removal of such nucleation nodes at the rate(s), $I_i(t)$. Values of λ and v that are combinations of 0 and 1 are of most interest. For $n = 0$ we have

$$dp^{(0)}/dt = \sum_{i=0}^n I_i(t) \quad (2-4.13)$$

independent of λ and v . The increase or decrease in number of nodes is therefore governed by the net generation rate. Considering the possible functions $I_i(t)$, we have a variety of scenarios to evaluate. The zeroth moment influences all higher moments, e.g., for $n = 1$ we have

$$dp^{(1)}/dt = \gamma p^{(0)} p^{(\lambda)} - \kappa p^{(0)} p^{(v)} \quad (2-4.14)$$

For $n = 2$ we have

$$dp^{(2)}/dt = \gamma p^{(0)} [p^{(\lambda)} + 2 p^{(\lambda+1)}] + \kappa p^{(0)} [p^{(v)} - 2 p^{(v+1)}] \quad (2-4.15)$$

Network dynamics represented by moment expressions for different generation expressions provide practical results. To assess the dynamics, let us first consider constant generation rate, $I_i = 0$ or α , following Barabasi et al. [Barabasi et al. 2002], and also a time dependent rate of node generation, $I_i(t) = \alpha t$. Tables 2-4.1 and 2-4.2 display the derived expressions for long-time limits and asymptotic behavior (after the initial

transient has passed) for number average p^{avg} , variance p^{var} , and polydispersity p^{pd} . For network breakage and growth, the limits for p^{avg} are 0 and ∞ , respectively, representing total network dissolution and complete connection. When $k_g(\xi) = \gamma \xi^\lambda$ and $k_d(\xi) = \kappa \xi^\nu$ with powers $\lambda = \nu$ equal to either 0 or 1, the direction of network change depends only on the relative magnitudes of γ and κ (Tables 2-4.1 and 2-4.2).

Table 2-4.1 The asymptotic behavior and long-time limits of number average, polydispersity, and variance for $I_i = 0$ (no node generation). Constants are denoted by c , c_1 , or c_2 .

| | p^{avg} | | p^{pd} | | p^{var} |
|---|---|--------------------|---|----------|------------------|
| | Asymptote | Limit | Asymptote | Limit | Limit |
| $k_g = \gamma, k_d = \kappa, I_i = 0$ | | | | | |
| $\gamma = \kappa$ | $p^{\text{avg}} = p_o^{\text{avg}}$ | p_o^{avg} | $p^{\text{pd}} = p_o^{\text{pd}} + 4\gamma p_o^{(0)3} t / p_o^{(1)2}$ | ∞ | ∞ |
| $\gamma > \kappa$ | $p^{\text{avg}} = p_o^{\text{avg}} + (\gamma - \kappa) p_o^{(0)} t$ | ∞ | $p^{\text{pd}} \sim 1 + c / t$ | 1 | ∞ |
| $\gamma < \kappa$ | $p^{\text{avg}} = p_o^{\text{avg}} + (\gamma - \kappa) p_o^{(0)} t$ | 0 | $p^{\text{pd}} \sim 1 + c / t$ | 1 | 0 |
| $k_g = \gamma \xi, k_d = \kappa \xi, I_i = 0$ | | | | | |
| $\gamma = \kappa$ | $p^{\text{avg}} = p_o^{\text{avg}}$ | p_o^{avg} | $p^{\text{pd}} = p_o^{\text{pd}} + 4\gamma p_o^{(0)2} t / p_o^{(1)}$ | ∞ | ∞ |
| $\gamma > \kappa$ | $p^{\text{avg}} = p_o^{\text{avg}} \exp[(\gamma - \kappa) p_o^{(0)} t]$ | 0 | $p^{\text{pd}} \sim p_o^{\text{pd}} + (\gamma + \kappa) / p_o^{\text{avg}} (\gamma - \kappa)$ | constant | ∞ |
| $\gamma < \kappa$ | $p^{\text{avg}} = p_o^{\text{avg}} \exp[(\gamma - \kappa) p_o^{(0)} t]$ | 0 | $p^{\text{pd}} \sim c_1 \exp[c_2 t]$ | ∞ | 0 |

For the case when $k_g(\xi) = k_d(\xi)$ with no node generation, $I_i = 0$, the network stays at the dynamic equilibrium state. With the non-zero node generation term, the network grows continuously by node addition followed by establishment of connections. With continuous addition of nodes, however, if the link removal process is dominant, the network will disintegrate unless the link removal process ceases. For the non-zero node

generation case, therefore, even if rate coefficients are equal, the average number of connections per node decreases as time increases.

Table 2-4.2 The asymptotic behavior and long-time limits of number average, polydispersity, and variance for $I_i = \alpha$ (constant node generation), and $I_i = \alpha t$ (time-dependent node generation). Constants are denoted by c , c_1 , or c_2 .

| | | p^{avg} | | p^{pd} | | p^{var} |
|--|--|---|----------|--|----------|------------------|
| | | Asymptote | Limit | Asymptote | Limit | Limit |
| $k_g = \gamma, k_d = \kappa, I_i = \alpha$ | | | | | | |
| $\gamma = \kappa$ | | $p^{\text{avg}} = p_o^{(1)} / (p_o^{(0)} + \alpha t)$ | 0 | $p^{\text{pd}} \sim c t^4$ | ∞ | ∞ |
| $\gamma > \kappa$ | | $p^{\text{avg}} \sim c t^2$ | ∞ | $p^{\text{pd}} \sim 1.2$ | 1.2 | ∞ |
| $\gamma < \kappa$ | | $p^{\text{avg}} \sim c_1 - c_2 t$ | 0 | $p^{\text{pd}} \sim 1.2$ | 1.2 | ∞ |
| $k_g = \gamma\xi, k_d = \kappa\xi, I_i = \alpha$ | | | | | | |
| $\gamma = \kappa$ | | $p^{\text{avg}} = p_o^{(1)} / (p_o^{(0)} + \alpha t)$ | 0 | $p^{\text{pd}} \sim c t$ | ∞ | ∞ |
| $\gamma > \kappa$ | | $p^{\text{avg}} \sim c_1 / (t \exp[c_2 t])$ | 0 | $p^{\text{pd}} \sim c_1 + c_2 t$ | ∞ | 0 |
| $\gamma < \kappa$ | | $p^{\text{avg}} \sim c / (t \exp[c_2 t])$ | 0 | $p^{\text{pd}} \sim c_1 / (t \exp[c_2 t])$ | 0 | 0 |
| $k_g = \gamma, k_d = \kappa, I_i = \alpha t$ | | | | | | |
| $\gamma = \kappa$ | | $p^{\text{avg}} = 2p_o^{(1)} / (2p_o^{(0)} + \alpha t^2)$ | 0 | $p^{\text{pd}} \sim c t^7$ | ∞ | ∞ |
| $\gamma > \kappa$ | | $p^{\text{avg}} \sim c t^3$ | ∞ | $p^{\text{pd}} \sim 1.25$ | 1.25 | ∞ |
| $\gamma < \kappa$ | | $p^{\text{avg}} \sim c_1 - c_2 t^3$ | 0 | $p^{\text{pd}} \sim 1.25$ | 1.25 | ∞ |
| $k_g = \gamma\xi, k_d = \kappa\xi, I_i = \alpha t$ | | | | | | |
| $\gamma = \kappa$ | | $p^{\text{avg}} = 2p_o^{(1)} / (2p_o^{(0)} + \alpha t^2)$ | 0 | $p^{\text{pd}} \sim c t^7$ | ∞ | ∞ |
| $\gamma > \kappa$ | | $p^{\text{avg}} \sim c_1 / (t^2 \exp[c_2 t])$ | 0 | $p^{\text{pd}} \sim c_1 t^7 \exp[c_2 t]$ | ∞ | ∞ |
| $\gamma < \kappa$ | | $p^{\text{avg}} \sim c_1 \exp[c_2 t] / t^2$ | ∞ | $p^{\text{pd}} \sim c_1 t^7 / \exp[c_2 t]$ | 0 | 0 |

For the case when $k_g(\xi) = k_d(\xi)$ with no node generation, $I_i = 0$, the network stays at the dynamic equilibrium state. With the non-zero node generation term, the network grows continuously by node addition followed by establishment of connections. With

continuous addition of nodes, however, if the link removal process is dominant, the network will disintegrate unless the link removal process ceases. For the non-zero node generation case, therefore, even if rate coefficients are equal, the average number of connections per node decreases as time increases.

When the powers λ and ν are different, however, an interesting behavior is revealed for the case $\lambda = 0$ and $\nu = 1$. For $I_i = 0$, one can show

$$p^{\text{avg}}(t) = \kappa^{-1} [(\kappa - \gamma)\exp(-\kappa p_o^{(0)}t) + \gamma] \quad (2-4.16)$$

and

$$p^{\text{var}}(t) = (\gamma/\kappa)[1 - \exp(-\kappa p_o^{(0)}t)] + \exp(-\kappa p_o^{(0)}t) + [p_o^{(2)}/p_o^{(0)} - 1] \exp(-2\kappa p_o^{(0)}t) \quad (2-4.17)$$

The limit as $t \rightarrow \infty$ is a stationary state with both average and variance approaching γ/κ .

The time dependence of the moments is much more complicated for $I_i = \alpha$ and αt , but the limits are, remarkably, the same ratio of γ to κ . A proof for any $I_i(t)$ can be fashioned algebraically by setting expressions for dp^{avg}/dt and dp^{var}/dt to zero and taking the limit as $t \rightarrow \infty$. This suggests that networks with constant growth ($\lambda = 0$) and size-dependent breakage ($\nu = 1$) are stable in the sense of reaching a constant limiting condition.

2-5. Power Law Networks

The evolution of power law distributed networks can be understood by examining cases when the rate coefficients themselves have the power law expression, Eq. 2-3.2, with $\lambda = \nu$. We assume the source terms are zero, $I_i = 0$, and truncate Eq. 2-3.5 to first order. By substituting the new time variable, Eq. 2-4.1, we write a first-order partial differential equation for the growth of the distribution,

$$\partial p(\xi, \theta) / \partial \theta + \partial [G p(\xi, \theta)] / \partial \xi = 0 \quad (2-5.1)$$

where $G = (\gamma - \kappa)\xi^\lambda$ is the growth (or dissolution) rate. This partial differential equation, having the common form of a continuity equation, is fundamental to population balance modeling [Randolph and Larson, 1986]. An exact solution can be obtained by Laplace transformation for the initial condition, $p(\xi, \theta=0) = 0$ (initially no nodes exist). We consider a boundary condition, $p(\xi=1, \theta) = p_o^{(0)}(1 - e^{-\theta/\tau})$, which means that the number of nodes with one connection increases with time to the constant $p_o^{(0)}$. The Laplace-transformed solution,

$$p(s, \xi) = (p_o^{(0)}/s(1+s\tau)) \xi^{-\lambda} \exp[-s(\xi^{1-\lambda} - 1)/k(1 - \lambda)] \quad (2-5.2)$$

inverts to

$$p(\xi, \theta) = p_o^{(0)} \xi^{-\lambda} (1 - \exp[-\theta/\tau + (\xi^{1-\lambda} - 1)/k\tau(1 - \lambda)]) u[\theta - (\xi^{1-\lambda} - 1)/k(1 - \lambda)] \quad (2-5.3)$$

where $u(x)$ is the unit step function defined as $u(x < 0) = 0$ and $u(x > 0) = 1$. The method of characteristics [Goldenfeld, 1992] can also be used to solve Eq. 2-5.1. As the time variable, θ , becomes sufficiently large, the step function equals unity. Therefore, the power law, $\xi^{-\lambda}$, in Eq. 2-5.3 dominates the asymptotic behavior. For the special case $\lambda = v = 1$, the result is

$$p(\xi, \theta) = (p_o^{(0)}/\xi)(1 - \xi^{1/k\tau} e^{-\theta/\tau}) u[\theta - \ln(\xi)/k] \quad (2-5.4)$$

such that the asymptote is ξ^{-1} . The domain of Eq. 2-5.3 extends from $\xi = 1$ to a value that increases with time. The moments, Eq. 2-2.1, thus exist for all but $t \rightarrow \infty$. The different boundary conditions required for the power law solution mean that a comparison with the moment solution ($v = \lambda = 1$, $I_1 = 0$ in Table 2-4.1) is not appropriate.

We illustrate these ideas by listing exact solutions of Eq. 2-5.1 for $G = k \xi^\lambda$ with $k = \gamma - \kappa$ for several initial and boundary conditions (Table 2-5.1). Part (a) lists the solution

for initial conditions, $p(\xi, \theta=0) = p_o^{(0)} \xi^{-\lambda}$ (initial number of nodes is related to the number of links as a power law) and the constant boundary condition, $p(\xi, \theta=0) = p_o^{(0)}$ (a constant number of nodes with one connection are always present and available for growth). Part (b) lists the solution with constant initial condition, $p(\xi, \theta=0) = p_o^{(0)}$ and constant boundary condition, $p(\xi=1, \theta) = p_o^{(0)}$. Part (c) lists the solution for the zero initial condition and time-dependent boundary condition, $p(\xi=1, \theta) = p_o^{(0)} (1 - e^{-\theta/\tau})$ (the number of nodes with one connection becomes constant with time). Part (d) in Table 2-5.1 gives solutions for the initial condition, $p(\xi, \theta=0) = p_o^{(0)} \xi^{-\lambda}$ and the boundary condition, $p(\xi=1, \theta) = p_o^{(0)} e^{\theta/\tau}$. Even though each pair of initial and boundary condition produces different solutions, the dominant term is always $\xi^{-\lambda}$. As we see in Table 2-5.1, the asymptotic solutions are all $p_o^{(0)} \xi^{-\lambda}$, and therefore the asymptotic solutions are independent of initial conditions, as expected.

Table 2-5.1 Time dependence of the linkage distribution $p(\xi, \theta)$ for network growth: the exact solutions of Eq. 2-5.1 with $G = k \xi^\lambda$ and $k = \gamma - \kappa$ for given initial and boundary conditions. The function $u(\cdot)$ is the unit step function.

| | | |
|-----|---|--|
| (a) | $p(\xi, \theta=0) = p_o^{(0)} \xi^{-\lambda}$ $p(\xi=1, \theta) = p_o^{(0)}$ | $p(\xi, \theta) = p_o^{(0)} \xi^{-\lambda}$ |
| (b) | $p(\xi, \theta=0) = p_o^{(0)}$ $p(\xi=1, \theta) = p_o^{(0)}$ | $\lambda = 1$ $p(\xi, \theta) = p_o^{(0)} \xi^{-1} \exp[-k\theta] \{1 - (1 - \exp[k\theta]) u[\theta - \ln[\xi]/k]\}$ |
| (c) | $p(\xi, \theta=0) = 0$ $p(\xi=1, \theta) = p_o^{(0)} (1 - e^{-\theta/\tau})$ | $p(\xi, \theta) = p_o^{(0)} \xi^{-\lambda} \{1 - \exp[-\theta/\tau + (\xi^{1-\lambda} - 1)/k\tau (1-\lambda)] u[\theta - (\xi^{1-\lambda} - 1)/k(1-\lambda)]\}$ |
| (d) | $p(\xi, \theta=0) = p_o^{(0)} \xi^{-\lambda}$ $p(\xi=1, \theta) = p_o^{(0)} e^{\theta/\tau}$ | $p(\xi, \theta) = p_o^{(0)} \xi^{-\lambda} \{1 - (1 - \exp[\theta/\tau - (\xi^{1-\lambda} - 1)/k\tau (1-\lambda)]) u[\theta - (\xi^{1-\lambda} - 1)/k(1-\lambda)]\}$ |

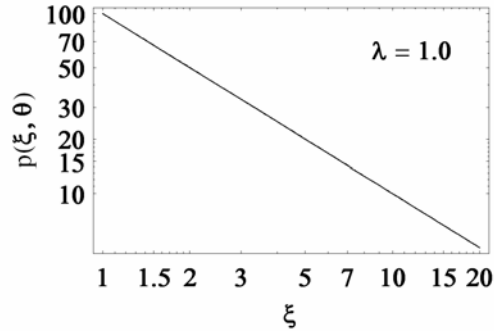


Figure 2-5.1 The general power law distribution for case (a) in Table 2-5.1.

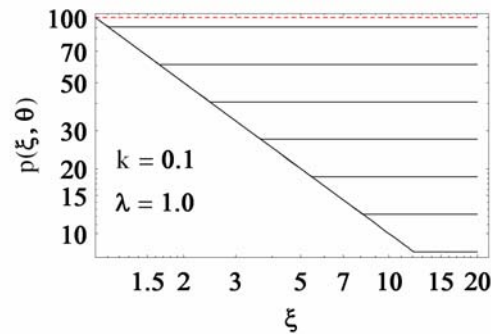


Figure 2-5.2 Evolution of the distribution for network growth (b) of Table 2-5.1, and the scaled time θ increases from top to bottom in steps of 4 from $\theta = 1$ to 29. The dotted line is the initial condition.

The solutions in Table 2-5.1 are plotted in Figs. 2-5.1 – 2-5.4. The parameter values chosen for the plots are $p_0^{(0)} = 100$, $\lambda = 1.0$, $\tau = 2-20$, and $k = 0.03-0.2$ for growth. The boundary condition is fixed at the point $\xi = 1$, and the dotted lines represent initial conditions. Figure 2-5.1 shows the linkage distribution for part (a) in Table 2-5.1. Figure 2-5.2 shows the time evolution of a node distribution for part (b) in Table 2-5.1. The horizontal line moves down with its left-side value following the asymptotic solution, $\xi^{-\lambda}$. The dynamic behavior is easily understood by viewing time-step animations for the solutions (see the website, <http://www.che.lsu.edu/faculty/mccoy/Networks/Networks.htm>). Figures 2-5.3 – 2-5.4, the plots of part (c) and (d) of Table 2-5.1, show how the

evolution of the distribution is converted into the power law distribution as time progresses. These plots show that the solutions in Table 2-5.1 approach a common asymptote, $p_0^{(0)} \xi^{-\lambda}$, plotted in Fig. 2-5.1.

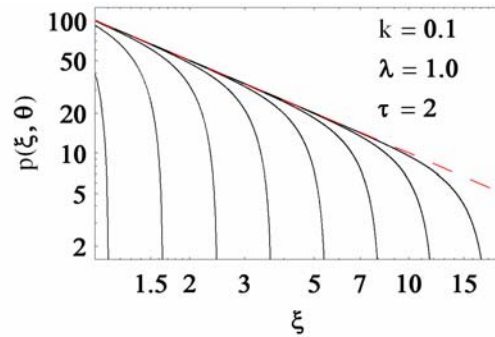


Figure 2-5.3 Evolution of the distribution for network growth, (c) of Table 2-5.1, where the scaled time θ increases from left to right in steps of 4 from $\theta = 1$ to 29. The dashed line is the common asymptote.

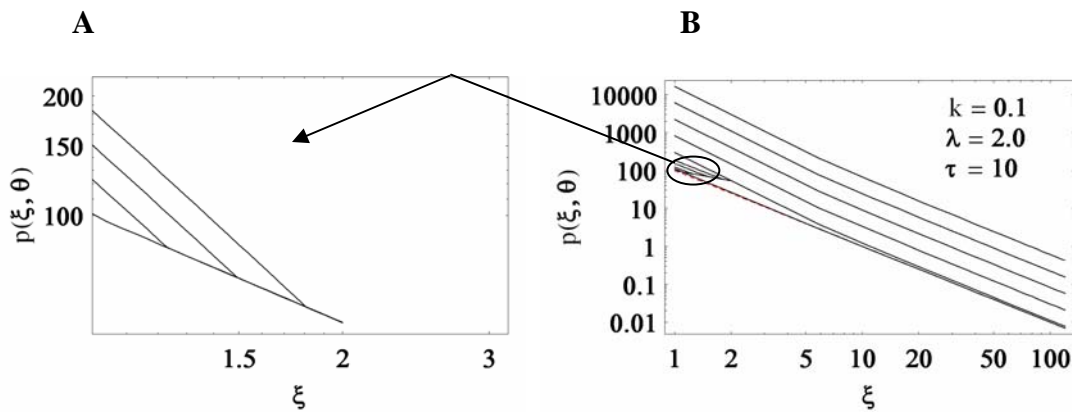


Figure 2-5.4 Evolution of the distribution for network growth, (d) of Table 2-5.1, where the scaled time θ increases from left to right in steps of 1 from $\theta = 1$ to 4 (A) and in steps of 10 from $\theta = 1$ to 61 (B). The dotted line is an initial condition.

We now examine the capability of the model, in particular part (c) of Table 2-5.1, to describe real-world network dynamics. The data analyzed consist of cumulative distributions for the Oregon system of Internet routers for the years 1997, 1999, and 2001, reported by Pastor-Satorras and Vespignani [Pastor-Satorras and Vespignani,

2004]. In Figs. 2-5.5A, 2-5.5B, and 2-5.5C, the points represent data showing how the linkage increased with time with the same slope for different years. The cumulative distribution is defined as,

$$P^c(\xi, \theta) = \int_{\xi}^{\infty} p(\xi, \theta) d\xi \quad (2-5.5)$$

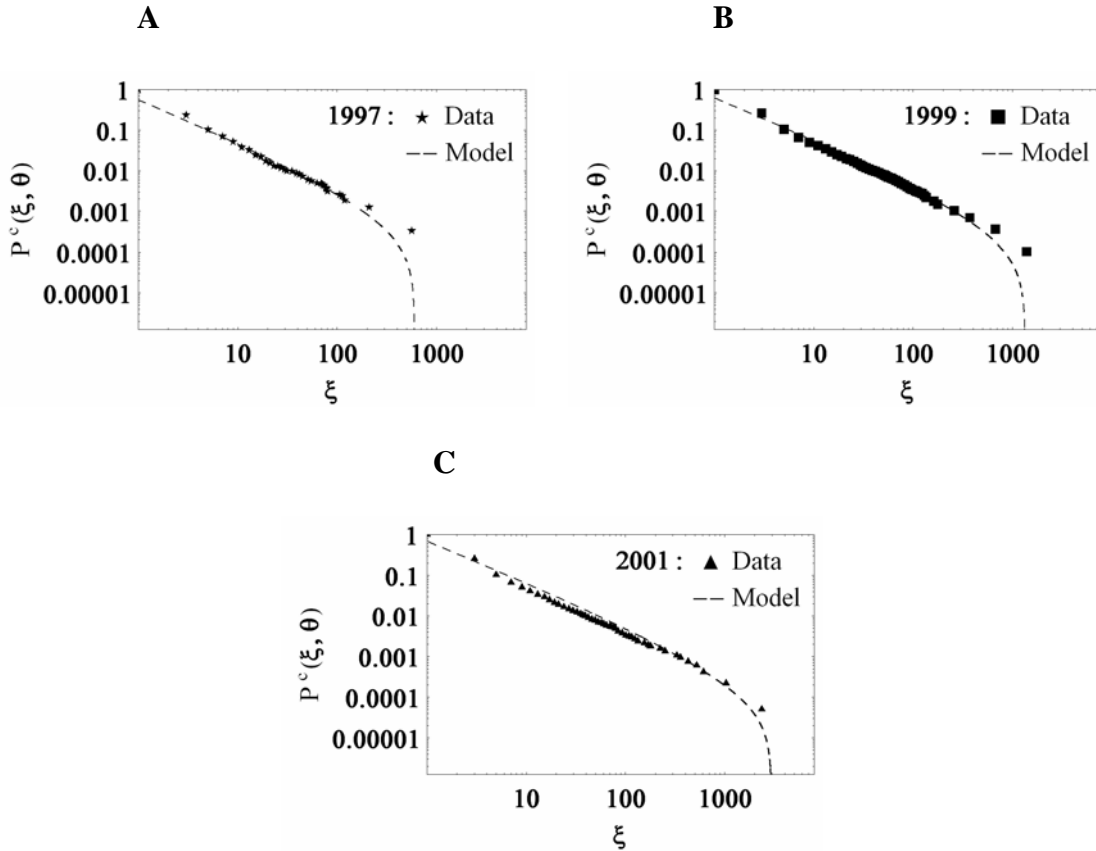


Figure 2-5.5 Comparison to statistical data of Oregon Internet growth in different years with the model distribution. Symbols ((A) star, 1997; (B) box, 1999; (C) triangle, 2001) represent statistical data and the lines are the distributions predicted by the model solution of part (c) of Table 2-5.1.

The data show a cumulative power law with slope -1.1 , which yields the power $\lambda = 2.1$ for the distribution. The model parameters are $p_0^{(0)} = 9$, $k = 0.1$, $\tau = 1000$, and $\theta = 64$ (1997), 72 (1999), and 80 (2001), respectively. From Eq. 2-4.1, scaled time is defined as $\theta = p_0^{(0)} t$ when the source term I_i is zero. Assuming that the cumulative distribution of

the Internet growth began in 1991, (i.e., $t = 0$ and scaled time $\theta = 0$ in 1991), t would then be 8 in 1999, giving a scaled time $\theta = 72$ for that year. From the comparison plotted in Fig. 2-5.5, we confirm that the network dynamics model gives a good estimate of the Internet growth. In accord with the data, the power law distributions are truncated at the value of ξ for the node with the largest number of links: 600 in 1997, 1300 in 1999, and 3000 in 2001.

The goal of parameter estimation is to determine the set of parameters that best reconciles the data with model predictions. For the sake of model prediction, we obtain optimal parameter estimates using a nonlinear parameter estimation method. In general, there is no algebraic expression for the best-fitting parameters, and thus numerical optimization algorithms incorporated in the spreadsheet program are applied to determine the best-fitting parameters. To obtain a best-fitting curve with a minimal deviation from a given set of the data, we applied the method of polynomial least squares. In the model comparison, the most significant deviations of the model prediction with the real system data arise at the end of distributions. Figure 2-5.5 demonstrates that the model predictions may have comparable deviations to predict truncated-tail behaviors of real power law structures. Since the parameter estimations are based on the nonlinear regression method, deviations in the model comparison with the data arise due to the lack of sufficient real system data and the errors associated with the integration of the first order partial differential equation for the fast extents. Thus it is plausible that the model may show deviation in the prediction of truncated-tail behavior where the system data are not sufficient, since the inherent sampling error may be greater than the error introduced by the model prediction.

2-6. Conclusion

Networks having perhaps millions of nodes and many more connections seem complex because of their size and intricacy. Many physical systems, however, are also composed of unimaginably large numbers (e.g., Avogadro's number of molecules per mole of gas), yet their statistical properties can be calculated and observed (moments of the velocity distribution are governed by the Boltzmann equation). Polymers, likewise, have large numbers of repeat units (monomers) making up their chain-length distributions. Crystals undergoing growth or dissolution also are composed of many molecular units. Polymer reaction kinetics and crystallization dynamics are typically formulated as population balance equations governing the statistical properties of the molecular-weight distribution. Networks analogously have statistical properties that can be computed from a population dynamics equation, as we have demonstrated. Mathematical modeling can assist in understanding functional forms and relationships between parameters and variables, and in recognizing which of these forms, relationships, parameters, and variables might be important or dominant, in contrast to minor or negligible.

The basis of the current theory is the (continuous) distribution of nodes as a function of time and number of connections. Obviously this continuous distribution approach that we have applied becomes a better description if the network is large. The nodes are considered independent of spatial position, implying that links are also spatially indeterminate. Nodal position would be an issue for fixed lattices restricted to near-neighbor interactions. For nondirectional connections between two nodes, we have shown how an integrodifferential PBE can be formulated based on formation and removal of connections, which naturally leads to second-order kinetics. The large-scale properties of

the network can be formulated as moments of the distribution, such as total number of nodes, total number of connections, and average number of connections per node. The moments are solutions of ordinary differential equations in time, and are particularly useful for random networks. Power law networks, as developed here, have an intrinsic nonlinear character and require an approach different from random networks.

The current theory based on population dynamics shows how network distributions are governed by integrodifferential equations that reduce to difference or differential equations under appropriate conditions. Distribution kinetics as mathematically presented through population balance equations allows the simultaneous analysis of many length and time scales. The time-dependent solutions reveal how the distributions, either random or power law, are maintained under nonstationary state conditions. This self-organizing tendency or universality of networks is thus a consequence of their distributions following well-known mathematics of physics and engineering science.

CHAPTER 3.

EVOLUTION OF POWER LAW DISTRIBUTIONS IN SCIENCE AND SOCIETY

3-1. Introduction

Power law distributions are obvious features of many complex systems and go by different names, e.g., fat tails in economics and Zipf's law in demographics and linguistics [Newmann, 2000]. Several processes have been proposed to explain power laws, for example, self-organized criticality [Bak et al., 1989] has been suggested as the origin of power laws in complex systems, highly optimized tolerance [Carson and Doyle, 1999] is a mechanism that relates evolving structure to power laws in interconnected systems, and random walk models describe the movement of particles influenced by a stochastic mechanism [Newmann, 2000; Marsili and Zhang, 1998]. Based on previous studies of kinetics, we propose a reversible association-dissociation mechanism that can produce power distributions.

The tendency for many physical and social systems to appear as power law distributions is well known, but how they evolve is not well understood. Table 3-1.2 suggests how a range of systems, including particles, aerosols, corporations, and cities are often distributed in frequency as power laws, here written as proportional to $\xi^{-\lambda}$. A frequency distribution is constructed for a system by a binning operation, which divides the total size range into intervals (bins) and then counts the number of items in each bin. The frequency can be plotted versus size on log-log coordinates, yielding a straight line with slope $-\lambda$ if a power law is obeyed. A frequency distribution can be transformed by summation or integration into a cumulative distribution, such that all items larger than (or

smaller than) the given size are plotted. Integrating $\xi^{-\lambda}$ yields $\xi^{1-\lambda}$, so that on log-log coordinates, a cumulative distribution has the slope $1 - \lambda$.

When $\lambda = 2$ the cumulative distribution has the slope -1 and is known as Zipf's distribution. According to Table 3-1.2, city [Marsili and Zhang, 1998] and corporation size distributions [Axtell, 2001] have the Zipf form. Generally, Zipf's law refers to the frequency y of an event relative to its size rank r , $y \sim r^{-b}$, where the power b is close to 1. This can be converted into a cumulative distribution expressed as $r \sim y^{-1/b}$, or $P(\xi \geq y) \sim \xi^{-1/b}$. The derivative of the cumulative distribution gives $\lambda = 1 + 1/b$ in a power law distribution $\xi^{-\lambda}$, so that $b = 1$ in Zipf's law will give a power law with $\lambda = 2$.

Networks in the physical and social spheres often display power law form. A number of large distributed systems, ranging from social to communication to biological networks, have power law distributions in their node degree (number of links). Such distributions reflect the existence of a few nodes with very high degree and many with low degree, a feature not found in standard random graphs [Adamic et al., 2001], which are frequently normal, or Gaussian, distributions [Chandrasekhar, 1943; Feller, 1957].

Our approach for investigating the formation and evolution of power law distributions is based on previous studies of polymer and particulate systems that add or remove monomers, represented by the property value ξ_m , to clusters according to kinetic rate expressions [McCoy, 2002; Sterling and McCoy, 2001]. Such a growth or dissolution process is visualized as analogous for individuals arriving or leaving a city, and for dollars received or paid out by a corporation, for example. We use the same terminology here, so that a monomer is any unit adding to a cluster. Table 3-1.1 explains the

Table 3-1.1 Power law systems and distributed properties.

| System Property | Particle cluster made up of monomers | City population | Corporate size by receipts in dollars |
|--------------------|--|----------------------------------|--|
| ξ | ξ = cluster mass/ monomer mass | ξ = number of individuals | ξ = dollars of receipts |
| ξ_m | ξ_m = 1, each monomer is one unit | ξ_m = one person | ξ_m = one dollar, the unit of a transaction |

Table 3-1.2 Power of the frequency distribution, $\xi^{-\lambda}$, for different systems.

| System | λ |
|---|-----------|
| The degree distribution of co-authorship network (physics) [Newman, 2001]. | 0.91-1.3 |
| Cluster size distribution of phase ordering system at steady state [Das et al., 2001]. | 1.25 |
| Distribution of financial stock market price changes [Equiluz and Zimmerman, 2000]. | 1.5 |
| Distribution of terrestrial animal species as a function of their length [Schroeder, 1991] | 2 |
| English word frequency (Zipf distribution) [Zipf, 1932; Adamic and Huberman, 2002] | 2 |
| Mass distribution of atmospheric aerosols for coagulation [Camacho, 2001]. | 2 |
| Size distribution of cities (population larger than 10^5) in the U.S. and India [Marsili and Zhang, 1998]. | 2 |
| Size distribution of U.S. firms based on receipts [Adamic et al., 2001]. | 2 |
| Outlink degree distribution for telephone calls between individuals [Adamic et al., 2001]. | 2.1 |
| Web connectivity [Barabasi and Albert, 1999]. | 2.1 |
| Internet backbone [Adamic et al., 2001]. | 2.15-2.2 |
| Size distribution of businesses in a price driven market [D'Hulst and Rodgers, 201]. | 2.2 |
| Collaborations of film actors [Barabasi and Albert, 1999; Adamic et al., 2001]. | 2.3 |
| Distribution of wealth for the 400 richest people in U.S. [Malcai et al., 1999]. | 2.36 |
| Distribution of total market values of companies in the stock market [Malcai et al., 1999]. | 2.4 |
| Probability that a certain web document contains k outgoing links [Barabasi et al., 1999]. | 2.45 |
| Size distribution of businesses and customers [Zheng et al., 2002]. | 2.5 |
| The degree distribution of co-authorship network (biomedicine) [Newman, 2001] | 2.5 |
| Size distribution of ion clusters in particle fragmentation [Lejeune et al., 2003]. | 2.64 |
| Citation patterns of scientific publications [Barabasi and Albert, 1999]. | 3 |
| Electric power grid of the western U.S. [Barabasi and Albert, 1999] | 4 |

relationships and symbolism for ξ and ξ_m ; Cluster size is the property ξ (e.g., dollars of receipts) and its unit, or monomer, value is ξ_m (e.g., one dollar). Just as in crystal growth,

where many monomers may deposit on the cluster at the same time, we nevertheless consider they deposit independently and separately. We have also applied the general ideas of the method, including the formulation and moment solution of population balance equations [McCoy, 2002], to investigate the growth and disassembly of networks.

3-2. Cluster Distribution Dynamics

For clusters of particles, cities, businesses, and other systems, the size distribution is defined by $C(\xi, t)d\xi$, representing the number of clusters at time t in the differential property range $(\xi, \xi+d\xi)$. The size (or other property) ξ is defined for particles, cities, and corporate clusters in Table 3-1.2. Moments are defined as integrals over the property ξ ,

$$C^{(n)}(t) = \int C(\xi, t) \xi^n d\xi \quad (3-2.1)$$

where the limits of integration are minimum and maximum values of ξ . The system property ξ is a positive integer, and for such discrete distributions, moments are defined by summations. For large ξ , however, the difference between discrete and continuous distributions is negligible, and a summation from $\xi = 1$ can be replaced by the integral in Eq. 3-2.1. In general the mathematical moments do not exist for power distributions unless the largest size is limited. The zeroth moment, $C^{(0)}(t)$, is the time-dependent number of clusters, and the average cluster property value is $C^{avg} = C^{(1)}/C^{(0)}$, defined in terms of the first moment, $C^{(1)}$. The variance,

$$C^{var} = C^{(2)}/C^{(0)} - [C^{avg}]^2 = [C^{avg}]^2 [C^{pd} - 1] \quad (3-2.2)$$

and the polydispersity, $C^{pd} = C^{(2)}C^{(0)}/C^{(1)2}$, are measures of the size-distribution breadth.

For particle growth by monomer addition, the number concentration, $m^{(0)}(t)$, of a

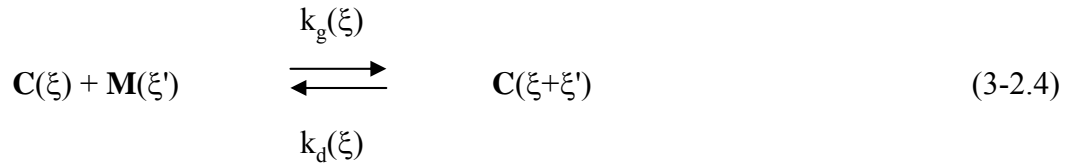
monomer having the property value, ξ_m , is the zeroth moment of the monomer distribution, $m(\xi, t) = m^{(0)}(t)\delta(\xi - \xi_m)$.

The cumulative distribution is defined as

$$C^{\text{cum}}(\xi) = \int_1^{\xi} C(\xi') d\xi' \quad (3-2.3)$$

so that $C^{\text{cum}}(\xi)$ is the number of entities having values less than ξ , and $C^{\text{cum}}(\xi)$ becomes $C^{(0)}$ as ξ approaches its maximum value. A power frequency distribution $C(\xi) \sim \xi^{-\lambda}$ obviously gives $C^{\text{cum}}(\xi) \sim \xi^{-\lambda+1}$.

Following methods previously reported [Madras and McCoy, 2001], we hypothesize that power law distributions obey a governing population balance equation. The growth or shrinkage process by which units having the property value, $\xi' = \xi_m$, are reversibly added to or dissociated from a cluster of mass ξ can be written as association-dissociation process,



where $C(\xi)$ is the cluster composed of number of units ξ and $M(\xi' = \xi_m)$ is the monomer.

This process intrinsically conserves the properties designated by ξ , and is naturally represented by balance equations in terms of ξ . The balance equations governing the cluster distribution, $C(\xi, t)$, and the monomer distribution, $m(\xi, t)$, are

$$\begin{aligned} \partial C(\xi, t) / \partial t = & -k_g(\xi)C(\xi, t) \int_0^{\infty} m(\xi', t) d\xi' + \int_0^{\xi} k_g(\xi - \xi')C(\xi - \xi', t)m(\xi', t) d\xi' \\ & - k_d(\xi)C(\xi, t) + \int_{\xi}^{\infty} k_d(\xi')C(\xi', t)\delta(\xi - (\xi' - \xi_m)) d\xi' - I \delta(\xi' - \xi^*) \end{aligned} \quad (3-2.5)$$

According to Eq. 3-2.4, unit additions are second-order in $C(\xi,t)$ and $m(\xi,t)$, whereas dissociation is first-order in $C(\xi,t)$. Expressed differently, the probability of combination is proportional to the product of their relative abundance of monomers and clusters. Nucleation of clusters of mass ξ^* at rate I are source terms, or sink terms if clusters are lost (denucleation). In growth and coarsening of a crystal size distribution, clusters may shrink to their critical size, ξ^* , and then spontaneously vanish [McCoy, 2001]. For growth or dissolution without such source or sink terms, we set $I = 0$. The initial condition for Eq. 3-2.5 is $C(\xi,t=0) = f(\xi)$.

Our purpose here is to show how simplified forms of Eq. 3-2.5 yield power distributions. As the main simplification, we will neglect the source term. The size distribution changes according to Eq. 3-2.5, which becomes, when the integrations over the Dirac distributions are performed, the finite-difference differential equation,

$$\begin{aligned} \partial C(\xi,t)/\partial t = & -k_g(\xi)C(\xi,t)m^{(0)}(t) + k_g(\xi-\xi_m)C(\xi-\xi_m,t)m^{(0)}(t) \\ & - k_d(\xi)C(\xi,t) + k_d(\xi+\xi_m)C(\xi+\xi_m,t) \end{aligned} \quad (3-2.6)$$

where $m^{(0)}(t)$ is monomer concentration, here considered constant. Equation 3-2.6 shows that $C(\xi,t)$ increases by addition of unit ξ_m to the cluster $(\xi-\xi_m)$ and decreases by the loss of ξ_m . The first two terms on the right-hand-side account for cluster growth by addition of monomer by second-order kinetics. If monomers are abundant and are not limiting in cluster growth, first-order kinetics holds. The remaining terms account for cluster breakage by the loss of one monomer by first-order kinetics. The modification of the equation allowing for a size distribution of monomers or including source terms is straightforward. A formal expansion Eq. 3-2.6 for $\xi_m \ll \xi$ yields a Fokker-Planck equation,

$$\begin{aligned} \partial C(\xi,t)/\partial t = \xi_m \partial[(k_d(\xi) - k_g(\xi)m^{(0)}(t)) C(\xi,t)]/\partial \xi \\ + \frac{1}{2} \xi_m^2 \partial^2[(k_d(\xi) + k_g(\xi)m^{(0)}(t)) C(\xi,t)]/\partial \xi^2 + \dots \end{aligned} \quad (3-2.7)$$

The rate coefficients for addition (growth) and removal (dissociation) are $k_g(\xi)$ and $k_d(\xi)$, respectively, considered in general to depend on ξ , the size of the cluster. As proposed in our previous work [Madras and McCoy, 2003], we use power expressions for the rate coefficients,

$$k_g(\xi) = \gamma \xi^\lambda \text{ and } k_d(\xi) = \kappa \xi^\nu \quad (3-2.8)$$

If monomers are not a limiting factor in cluster growth, then $m^{(0)}(t)$ can be considered as constant, $m_o^{(0)}$, and we can define the dimensionless time variable θ , and a rate coefficient ratio k ,

$$\theta = t \gamma m_o^{(0)}, \quad \alpha = \kappa / (\gamma m_o^{(0)}) \quad (3-2.9)$$

For growing systems, k has a value between zero and one ($0 \leq k \leq 1$). If distribution growth is controlled by limited monomer, then $m^{(0)}(t)$ decreases as individuals form clusters, influencing the evolution, as in crystallization from a saturated solution [McCoy, 2000; Madras and McCoy, 2001; Madras and McCoy, 2002b].

When the exponents of the rate coefficients are equal, $\lambda = \nu$, Eq. 3-2.6 yields the dimensionless difference-differential equation,

$$\begin{aligned} \partial C(\xi,\theta)/\partial \theta = (\xi-1)^\lambda C(\xi-1,\theta) - \xi^\lambda C(\xi,\theta) \\ + \alpha [(\xi+1)^\lambda C(\xi+1,\theta) - \xi^\lambda C(\xi,\theta)] \end{aligned} \quad (3-2.10)$$

where we have set $\xi_m = 1$. Similarly, Eq. 3-2.7 yields the partial differential equation,

$$\begin{aligned} \partial C(\xi,\theta)/\partial \theta = -\partial[(1 - \alpha) \xi^\lambda C(\xi,\theta)]/\partial \xi \\ + \frac{1}{2} \partial^2[(1 + \alpha) \xi^\lambda C(\xi,\theta)]/\partial \xi^2 + \dots \end{aligned} \quad (3-2.11)$$

Equation 3-2.11 with second-order derivative terms is a convective diffusion equation, which for the special case $\lambda = 0$ has a well-known exponential solution [Chandrasekhar, 1943; Feller, 1957]. For the growing system ($k > 0$) with the characteristic cluster size ($\xi \sim L$), Eq. 3-2.11 becomes,

$$\partial C(\xi, \theta) / \partial \theta = -\partial [k \xi^\lambda C(\xi, \theta)] / \partial \xi + O(1/L^2) \quad (3-2.12)$$

Compared to the first-order term ($\sim 1/L$), the second order term ($\sim 1/L^2$) is negligible if L is large. As we will demonstrate, the first-order solution is sufficient to derive power distributions. The first-order Fokker-Planck equation (Eq. 3-2.12) can be satisfied even for the case when $k = 0$, because we are describing cluster kinetics. In this case the rate of growth and dissociation are same, $k_g(\xi)m_o^{(0)} = k_d(\xi)$. The time derivative is zero, the system becomes an equilibrium state, and the first-order Fokker-Planck equation is satisfied.

The difference-differential equation, Eq. 3-2.6, is similar to stochastic equations for the transition probability with birth and death rate power expressions [Cox and Miller, 1965; Roehner and Valent, 1982]. Whereas birth and death rates in transition probability equations usually are restricted to linear or quadratic dependence [Roehner and Valent, 1982], the proposed model can be applied with any λ (usually between 0 and 5).

3-3. Power Laws

We now illustrate how power law distributions evolve according to the dimensionless population balance, Eq. 3-2.10 or 3-2.11, representing growth or dissolution of a distribution. The evolution can be understood by considering the rate coefficients with a power expression in the first-order partial differential equation for distribution growth. For the case when $\lambda = \nu$, we truncate Eq. 3-2.11 to first-order:

$$\partial C(\xi, \theta) / \partial \theta + \partial [k \xi^\lambda C(\xi, \theta)] / \partial \xi = 0 \quad (3-3.1)$$

where $k = (1 - \alpha)$ is the growth (or dissolution) rate. This partial differential equation, Eq. 3-3.1, having the common form of a continuity equation, is fundamental to population balance modeling. The ξ -derivative growth term [Himmelblau and Bischoff, 1968; Randolph and Larson, 1986] conventionally appears in crystal growth. A solution can be obtained by Laplace transformation for the general initial condition, $C(\xi, \theta=0) = f(\xi)$. For a distribution to grow, it is necessary either that new clusters nucleate or existing clusters are ready to grow. We apply the general boundary condition, $C(\xi=1, \theta) = g(\theta)$, which means that a certain number, $g(\theta)$, of emergent clusters of size $\xi = 1$ are present and available for cluster growth.

The Laplace-transformed solution of Eq. 3-3.1 for the general initial and boundary conditions ($f(\xi)$ and $g(\theta)$) is,

$$C(\xi, s) = (\xi^{-\lambda} / k) \exp[-s \xi^{1-\lambda} / k(1-\lambda)] \int_1^\xi \exp[sy^{1-\lambda} / k(1-\lambda)] f(y) dy + g(s) \xi^{-\lambda} \exp[s(1-\xi^{1-\lambda}) / k(1-\lambda)] \quad (3-3.2)$$

Dominating the result, the term $\xi^{-\lambda}$ represents a distribution with slope $-\lambda$ on log-log coordinates. The long-time asymptotic dominance of the power term $\xi^{-\lambda}$ is readily understood by recognizing that early-time transients will dissipate. Then in Eq. 3-3.1 the time derivative becomes negligible relative to the ξ derivative; thus $\partial [k \xi^\lambda C(\xi, \theta)] / \partial \xi \sim 0$, which integrates to $C \sim \xi^{-\lambda}$. This reveals the underlying mathematical reason for evolution to the power distribution.

For the special case when $\lambda = \nu = 1$, Eq. 3-3.1 becomes

$$\partial C(\xi, \theta) / \partial \theta + \partial [k \xi C(\xi, \theta)] / \partial \xi = 0 \quad (3-3.3)$$

and the Laplace-transformed solution, Eq. 3-3.2, for the general initial and boundary conditions is simplified as

$$C(\xi, s) = (\xi^{-(1+s/k)} / k) \int_0^\xi y^{s/k} f(y) dy + g(s) \xi^{-(1+s/k)} \quad (3-3.4)$$

A quite simple case that illustrates this behavior is the initial condition $f(\xi) = 0$ and boundary condition $g(\theta > 0) = C_0$. This means that at an instant after $\theta = 0$, a constant number of emergent clusters, or nuclei, become available for growth of the distribution. As in chain polymerization [McCoy and Madras, 2001], the chain-reaction nature of monomer addition ensures that a distribution of cluster sizes will be obtained. In terms of the unit step function, defined as $u(\xi < 0) = 0$ and $u(\xi \geq 0) = 1$, the solution for Eq. 3-3.1 is

$$C(\xi, \theta) = C_0 \xi^{-\lambda} u[\theta - (\xi^{1-\lambda} - 1)/k(1-\lambda)] \quad (3-3.5)$$

Other initial and boundary conditions also yield the power distribution. For example, the initial condition,

$$f(\xi) = C_0 [1 - u(\xi - \xi_0)] \quad (3-3.6)$$

is a rectangular distribution with a step down to zero at $\xi = \xi_0$. An exponentially increasing boundary condition from 0 up to C_0 is

$$g(\theta) = C_0 (1 - \exp[-\theta/\tau]) \quad (3-3.7)$$

which means that at long time the number of emergent clusters of size $\xi = 1$ approaches C_0 , constant with time. For Eq. 3-3.7 and the initial condition Eq. 3-3.6, when $\lambda = 1$, the distribution is

$$C(\xi, \theta) = C_0 \exp[-k\theta] \{1 + (\exp[k\theta]/\xi - \exp[\theta(k-1/\tau)]\xi^{-1+1/k\tau} - 1)u[\theta - \ln(\xi)/k] - u[\xi - \xi_0] (1 - u[\theta - \ln(\xi/\xi_0)/k])\} \quad (3-3.8)$$

Eqs. 3-3.5 and 3-3.8 are plotted in Figs. 3-3.1A and 3-3.1B, respectively.

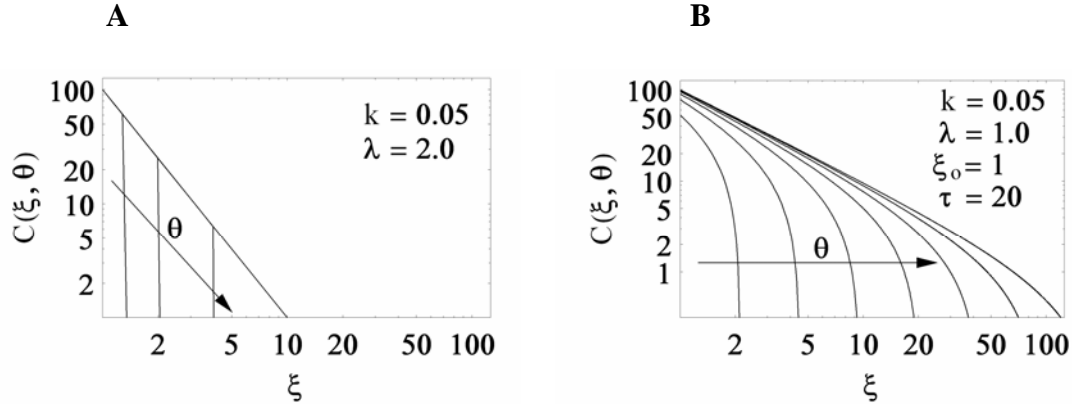


Figure 3-3.1 Evolution of the size distribution for cluster growth case Eqs. 3-3.5 and 3-3.8: $C_0 = 100$, $\xi_0 = 1$, and the scaled time θ increases in steps of 5 (A) from $\theta = 5$ up to 20 with $\lambda = 2$, and in steps of 15 (B) from $\theta = 15$ up to 105.

Consider the case when the initial condition is a power law from $\xi = 1$ to $\xi = \xi_0$,

$$C(\xi, \theta=0) = C_0 \xi^{-\lambda} (1 - u[\xi - \xi_0]) \quad (3-3.9)$$

For the exponentially increasing boundary condition Eq. 3-3.7 and the initial condition Eq. 3-3.9, the distribution is

$$C(\xi, s) = C_0 \xi^{-\lambda} \{1 - \exp[-\theta/\tau - (1 - \xi^{1-\lambda})/k\tau(1-\lambda)] u[\theta - (\xi^{1-\lambda} - 1)/k(1-\lambda)] - u[\xi - \xi_0](1 - u[\theta - (\xi^{1-\lambda} - \xi_0^{1-\lambda})/k(1-\lambda)])\} \quad (3-3.10)$$

Figure 3-3.2 is the plot of Eq. 3-3.10 for the cluster growth case when the initial distribution is a power law. We note that cluster size distributions become truncated power laws as ξ_0 approaches unity, as depicted in Figs. 3-3.1B and 3-3.2A. Although the above solutions differ for each initial and boundary condition, the dominant term is always $\xi^{-\lambda}$. This shows that a power law distribution evolves from an arbitrary initial distribution, subject to the conditions that the rate coefficient has the power form. Transients in the boundary conditions die out as time increases, leading to the asymptotic power behavior.

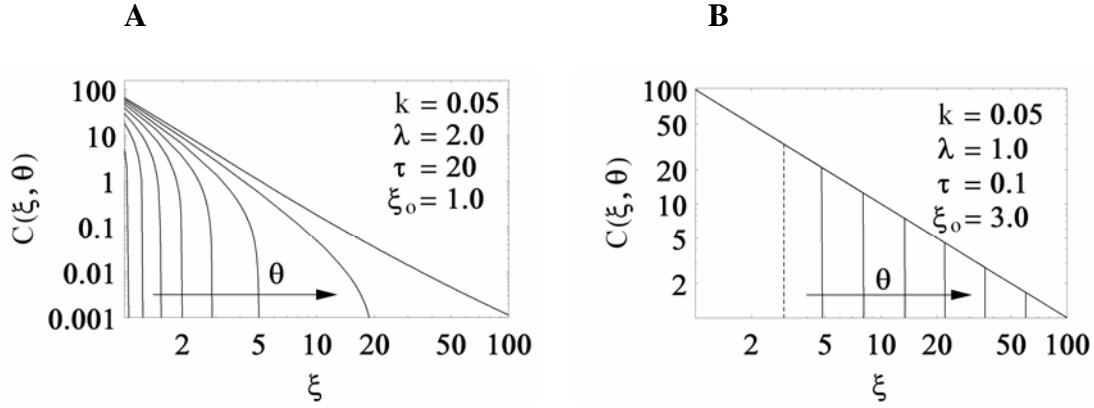


Figure 3-3.2 Evolution of the size distribution for cluster growth case Eq. 3-3.10: $C_0 = 100$, $\xi_0 = 3$, and the scaled time θ increases in steps of 3 (A) from $\theta = 4$ up to 22 with $\lambda = 2$, and in steps of 10 (b) from $\theta = 10$ to 70 with $\lambda = 2$. The dotted line in (B) is the initial condition.

As a consequence of examining cluster size distribution dynamics, we conclude that our population balance model can describe cluster growth systems. Many physical and social systems intrinsically grow and thus have an historical character, so our approach is reasonable for such accumulative systems.

Table 3-3.1 Parameters for comparison of corporation size data with our model. ξ_{\max}^* is the truncation size uncorrected for inflation.

| | | | | |
|-------------------|-------------------|-------------------|--------------------|--------------------|
| year | 1967 | 1977 | 1987 | 1997 |
| θ | 142 | 158 | 174 | 190 |
| t | 89 | 99 | 109 | 119 |
| ξ_{\max}^* | 0.7×10^9 | 2.5×10^9 | 8.0×10^9 | 25.0×10^9 |
| CPI based on 1997 | 4.998 | 2.749 | 1.467 | 1.0 |
| ξ_{\max} | 3.5×10^9 | 6.9×10^9 | 11.7×10^9 | 25.0×10^9 |

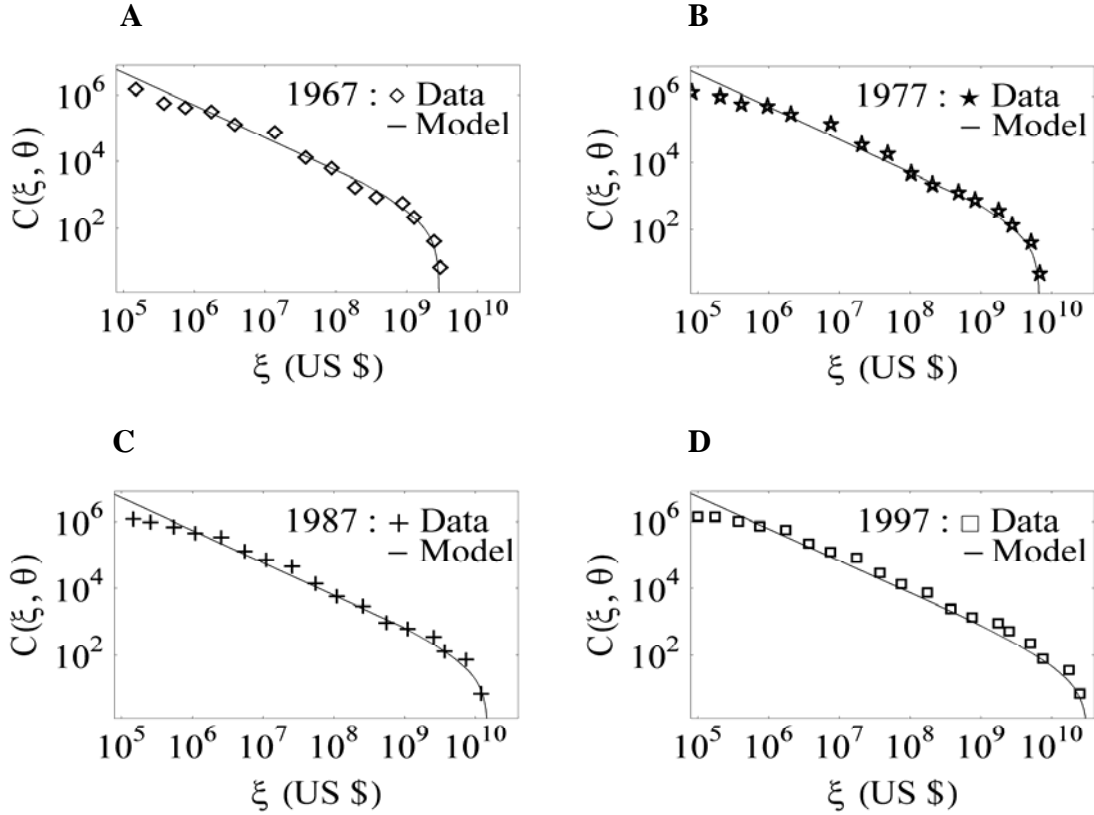


Figure 3-3.3 Comparison of the model and statistical data of U.S. company size-distribution growth in different years. The lines are the model predictions, Eq. 3-3.10, and symbols represent statistical data: (A) diamond (\diamond), 1967; (B) star (\star), 1977; (C) cross (+), 1987; (D) box (\square), 1997.

We now investigate the capability of the model, in particular Eq. 3-3.10, to describe the power law evolution of corporation size-distribution data. Based on the reversible, reaction-like process described in Eq. 3-2.4, the model excludes cluster-cluster interactions such as aggregation. Although including these interactions is possible [Madras and McCoy, 2003], here we assume corporate mergers are negligible. The data examined are cumulative U.S. firm size distributions classified by receipts size for the years from 1967 to 1997 in steps of 10 years, reported by the U.S. Bureau of the Census [U.S. Bureau of the Census, 1967-1997]. In Figs. 3-3.3A to 3-3.3D, the symbols represent data showing how the number and sizes of enterprises increased with time for

different years. The data shown in Fig. 3-3.3 have a cumulative power law with slope -0.94 for all years, and thus the power $\lambda = 1.94$ for the frequency distribution. The model parameters are $k = 0.1$, $\tau = 50$, and $C_o = 6 \times 10^6$. Values of θ (Table 3-3.1) are scaled time defined as $\theta = t \gamma m_o^{(0)}$ (Eq. 3-2.9), so that for the base year 1878, where $t = \theta = 0$, time t has the values given in Table 3-3.1. From the comparison plotted in Fig. 3-3.3, we confirm that the power law distribution model gives a good estimation of the U.S. firm growth. In accord with the data, the power law distributions are truncated by the exponential part of the equation at the value of ξ (ξ_{\max}^* in Table 3-3.1) for the cluster with the largest number of monomers (unit US dollars). To compensate for inflation, we applied the Consumer Price Index (CPI) to get ξ_{\max} in Table 3-3.1.

3-4. Moment Expressions

Moment results also show the time dependence of the size distribution, $C(\xi, \theta)$, for growth. Power law expressions with unlimited ξ do not have proper integrals, and thus have no moments. But as we have demonstrated, power laws evolve by extending to increasingly larger values of ξ , and thus moment integrals can be defined for finite time. For integer values of λ and ν , a general moment equation can be derived by the operation of the moment definition, Eq. 3-2.1. We examine the case when both rate coefficients have the same power, $\lambda = \nu$.

The results for the evolving distributions we have derived are discontinuous, and can be used to derive moments if the integral in Eq. 3-2.1 is sectioned into parts according to the step behavior. One part is defined from $\xi = 1$ or ξ_o up to $\xi = \xi_{\max}(\theta)$, which is a function of θ . Another part is defined from $\xi = 1$ or ξ_o up to $\xi = g(\theta)$, which can be determined based on the step function in the distribution. This time dependent

function can have different forms such as $e^{\theta k}$ or $\xi_0 e^{\theta k}$ when $\lambda = 1$ and $(k\theta(1-\lambda)+1)^{1/(1-\lambda)}$ or $(k\theta(1-\lambda)+\xi_0)^{1/(1-\lambda)}$ when we keep a general power $\lambda \neq 1$. Moments defined in Eq. 3-2.1 are conveniently evaluated, therefore, as the sum,

$$C^{(n)}(\theta) = \int_1^{\xi_{\max}(\theta)} (C_1(\xi, \theta) + C_2(\xi, \theta)u[\xi-\xi_0]) \xi^n d\xi + \int_1^{g(\theta)} (C_3(\xi, \theta) + C_4(\xi, \theta)u[\xi-\xi_0]) \xi^n d\xi \quad (3-4.1)$$

Table 3-4.1 Moment results and their asymptotes.

| IC \ BC | A and C | B |
|---------|---|---|
| | $C(\xi=1, \theta) = C_o$ $C(\xi=1, \theta) = C_o(1-\exp[\theta/\tau])$ | $C^{(0)}(\theta) \sim \theta$ $C^{\text{avg}}(\theta) \sim e^{k\theta}/\theta$ $C^{\text{var}}(\theta) \sim e^{2k\theta}/\theta$ $C^{\text{pd}}(\theta) \sim \theta$ |

The moments, $C^{(n)}(\theta)$, C^{avg} , C^{var} , and C^{pd} , have complicated expressions that can be derived by computer algebra, so here we list only their long time limits and asymptotic behaviors, which are determined based on their boundary condition, $C(\xi=1, \theta) = g(\theta)$, and the three different initial conditions, Eqs. 3-3.5, 3-3.9, and 3-3.13. In Table 3-4.1, we organize each case based on boundary conditions. For case A and C in Table 3-4.1, the asymptotes are identical, characteristic of a constant boundary condition. The zeroth moment, representing the number of clusters, has a long-time asymptote linear in time for cases A and C and exponential for B in Table 3-4.1. The number average and variance

have exponential asymptotes, but increase more rapidly for case B. The polydispersity index increases proportionally with time for cases A and C and exponentially for case B in Table 3-4.1.

3-5. Conclusion

Size distribution power laws have been observed in numerous science and social systems composed of clusters. To understand how power laws evolve for these systems, we have constructed a model based on reaction-like processes of reversible monomer addition to clusters. A population dynamics equation similar to those used for distribution kinetics of crystal growth and chain polymerization describes cluster growth and/or dissolution with cluster size-dependent rate coefficients. The hypothesis that power law distributions are governed by a population balance equation realistically describes cluster-growth systems.

Mathematical solution to the population balance equation provides relationships among parameters and variables for the distributions, and yields the functional form of the dominant power law term, $\xi^{-\lambda}$. The cluster size distributions have statistical properties, such as moments, that can be calculated from the population dynamics equations. Derived cluster size distributions show the development of the asymptotic power law, $\xi^{-\lambda}$, at long time. A central feature of the evolving distribution is that the initial distribution is transformed into a power law at points increasing with time. Thus as time progresses, the power law overtakes the initial distribution and initial transients dissipate.

As a consequence of examining cluster size distribution dynamics, we conclude that our population balance model can describe cluster growth systems. However, for the cluster shrinkage cases, we encounter discontinuous regions between time steps. This

suggests that the distribution decomposition process is different fundamentally from its growth, and that our model as presently formulated is not suitable to describe cluster reduction systems. A similar situation applies for modeling polymerization (growth) and polymer degradation (decomposition), where different reaction orders, different stoichiometric kernels, and different rate coefficients must be considered for growth and its reverse. Many physical and social systems, however, intrinsically grow and thus have an historical character, so our approach is reasonable for such accumulative systems.

CHAPTER 4.

ACCELERATING NETWORKS WITH AND WITHOUT PREFERENTIAL ATTACHMENT

4-1. Introduction

As stated, a complex system can be defined as a system with many interacting and interdependent parts having emergent self-organization [Ottino, 2003]. New technologies and rapidly changing societies, as well as biological evolution, increase the need for a better understanding of these complex systems and their structure. Data for complex systems [Dorogovtsev and Mendes, 2001b] often reveal network structure, consisting of many connections among many nodes [Albert and Barabasi, 2002; Ottino, 2003; Strogatz, 2001]. Among the models for network simulation, the Erdos and Renyi model [Bollobas, 1985; Erdos and Renyi, 1960] generates undirected random connection networks. The small-world network model [Watts and Strogatz, 1998] is an interpolation between regular lattice models and random graphs [Vazquez, 2000]. Compared to the random graph model the small world network has a much larger clustering coefficient [Watts and Strogatz, 2000], which is the probability that two randomly chosen nodes have a connection with each other. The Barabasi-Albert network model [Barabasi and Albert, 1999], a preferential growth model producing a power law structure, clarified the time dependence of power law networks. Generally, networks evolve with time [Albert and Barabasi, 2002; Barabasi et al., 2002; Ottino, 2003; Strogatz, 2001], typically growing unless they undergo breakage of connections or removal of nodes. For growing networks, the number of nodes and number of connections are typically increasing with time [Albert and Barabasi, 2002; Barabasi and Albert 1999; Barabasi et al., 2002;

Dorogovtsev and Mendes, 2001b]. As few networks show linear growth with their size [Jeong et al., 2000], the majority of networks grow nonlinearly, with the total number of connections increasing faster than the total number of nodes. Such networks are called accelerating networks [Albert and Barabasi, 2002; Dorogovtsev and Mendes, 2002; Mattick and Gagen, 2005].

The present aim is to propose a different approach for the dynamics of linear and nonlinear growing networks based on distribution kinetics [Jeon and McCoy, 2005a; Jeon and McCoy, 2005b]. For a class of growing network models [Dorogovtsev and Mendes, 2000; Dorogovtsev and Mendes, 2001b; Mattick and Gagen, 2005], where the addition of new nodes leads to power law structure [Barabasi and Albert, 1999; Barabasi et al., 1999; Huberman and Adamic, 1999], the degree distribution follows a power law, $p(\xi) \sim \xi^{-\lambda}$. Examples of power law networks [Dorogovtsev and Mendes 2001c] are communication networks, such as the World Wide Web and the Internet [Baran, 1964; Huberman et al., 1998], citation networks in the scientific literature [Lahererre and Sornette, 1998; Render, 1998], collaboration networks [Amaral et al., 2000; Newman, 2001], and metabolic reaction networks [Jeong et al., 2000].

The current model we are going to develop for the accelerating networks, allowing multiple connections between any two nodes, is more general than the connection-limited networks, allowing only one connection between any two nodes. An example of the model network is an airline network where several flying routes exist from one airport to another. For connection-limited networks, the most effective structure is the saturated state, where all nodes are connected with all other nodes [Mattick and Gagen, 2005; Jeong et al., 2000]. Such saturated structures may not be realistic even for the connection-limited networks, because real-world complex systems are trade-offs

between cost and efficiency, and unlinked node pairs may be present. For instance, if a computer is directly connected to the main server, surfing the Internet will be more than 100 times faster than the normal broadband high-speed Internet connection. For billions of Internet users and given the cost of direct connections, this would be unrealistic; thus the Internet, a representative communication network, follows the second best effective structure, the power law network.

4-2. Model

We begin with the general concept of frequency distribution of nodes and connections, in which nodes and connections are added and established for network growth, and removed and eliminated for breakage [Jeon and McCoy, 2005a]. To exemplify fundamental ideas, we will show even simple models can produce diverse behaviors. Connections (links or edges) are binary interactions between nodes that are nondirectional and of indeterminate length. For such connections, the nodal-linkage distribution $p(\xi, t)d\xi$ is defined as the number of nodes at time t with number of connections in the interval ξ to $\xi+d\xi$. Because ξ is a positive integer, the distribution is discrete, but for a large number of connections one can substitute the discrete distribution with a continuous distribution (replacing summations for the discrete distribution with integrals for a continuous distribution). The distribution of the number of connections in the interval ξ to $\xi+d\xi$ can be expressed as $\frac{1}{2}\xi p(\xi, t)d\xi$. Since each connection is associated with two nodes in non-directional networks, the factor $\frac{1}{2}$ appears. This is similar to polymer molecular weight distributions [Madras and McCoy, 2002a], where the number of macromolecules having mass in the range $(x, x+dx)$ is $p(x)dx$ and the mass of macromolecules in the same interval is $xp(x)dx$.

The moments of the nodal-linkage distribution are defined as,

$$p^{(n)}(t) = \int p(\xi, t) \xi^n d\xi \quad (4-2.1)$$

where the integration limits are determined by the domain of $p(\xi, t)$. From the definition, the total number of nodes and total number of connections are $p^{(0)}(t)$ and $\frac{1}{2} p^{(1)}(t)$, respectively, and the average number of connections per node, $\frac{1}{2} p^{\text{avg}}(t)$, where $p^{\text{avg}}(t) = p^{(1)}(t)/p^{(0)}(t)$. The variance and polydispersity index in terms of the second moment are defined as $p^{\text{var}} = p^{(2)}/p^{(0)} - p^{\text{avg}^2}$ and $p^{\text{pd}} = p^{(2)}/p^{(0)}/p^{(1)2}$, providing further information for the character and shape of the distribution.

To estimate the maximum number of connections for a connection-limited network, consider a simple saturated network with four nodes (three connections each where any two nodes have only one connection between them). Three connections for the first node, two connections for the second node, and one connection for the third node can be counted without repeating, and the summation of these gives the total number of connections. Therefore, the maximum possible number of connections, $\frac{1}{2} p^{(1)}(t)$, for a connection-limited network expressed as an arithmetic progression from 1 up to $(p^{(0)}(t) - 1)$ is $\frac{1}{2} p^{(0)}(t)(p^{(0)}(t) - 1)$ [Albert and Barabasi, 2002; Dorogovtsev and Mendes, 2002].

We will focus on the network growth for two classifications: exponential networks (single-scale), randomly connected in the absence of preferential attachment, and power law (scale-free) networks constructed by preferential attachment (Fig. 1-2.1). The distribution of exponential networks, for instance, Gaussian, binomial, or Poisson distributed networks, is unimodal (peaked) with well-defined moments. Therefore, statistical properties such as mean and variance can be easily defined and measured. The distribution of power law networks has a power law expression, $p(\xi) \sim \xi^{-\lambda}$, where λ is usually a positive constant. The moments of such networks are not defined on the interval

$(0, \infty)$, because they do not have an inherent scaling factor. However, if the evolving power law network has an expanding finite domain, the moment integral in Eq. 4-2.1 can be defined within the domain. For that reason, moments of the power law networks within a finite domain are defined as,

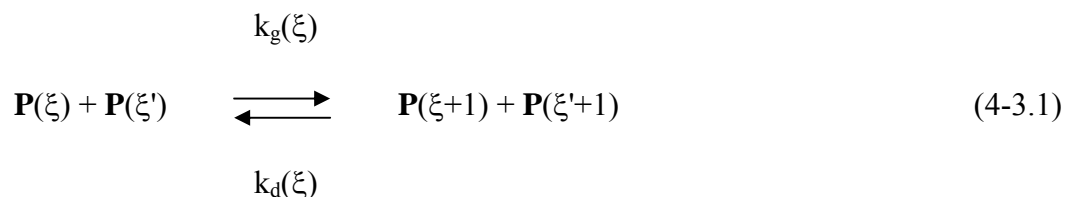
$$p^{(n)}(t) = \int_0^{\xi_m(t)} p(\xi, t) \xi^n d\xi \quad (4-2.2)$$

where $\xi_m(t)$ represents the maximum number of connections, generally a function of time.

Our aim is to develop a framework that determines the evolution of the two types of network. We will approach the problem with knowledge from distribution kinetics based on population dynamics, which has proven a productive approach to polymerization and depolymerization [McCoy and Madras, 2001], fragmentation and aggregation [Madras and McCoy, 2002a] of particulate systems, and growth and dissolution of crystal systems [Madras and McCoy, 2002b]. By this method, we can obtain solutions for numerous interesting systems, and show the effect of the parameters that govern the network evolution.

4-3. Distribution Kinetics

We describe network structure based on the association-disassociation process, written below as a reaction-like expression. We consider connections added one at a time to available nodes, having the possibility that connected nodes, or indeed entire networks, might coalesce by such connecting processes. The addition or removal of connections can be written as a reversible rate process,



where $\mathbf{P}(\xi)$ schematically represents a node with ξ connections. A connection can be formulated by the interaction between two nodes. We propose the rate coefficients $k_g(\xi)$ and $k_d(\xi)$ for addition (growth) and removal (dissociation), respectively, by adopting power expressions, which are generally considered to depend on the number of connections,

$$k_g(\xi) = \gamma \xi^\lambda \quad \text{and} \quad k_d(\xi) = \kappa \xi^\nu \quad (4-3.2)$$

where the constants γ , κ , λ , and ν are positive definite. The process of Eq. 4-3.1 is unchanged by replacing $\xi-1$ with ξ or $\xi'+1$ with ξ' . In addition to expressions for connection formulation or removal, expressions for rates of node generation or loss are also required to construct the population dynamics equation. Constructing the governing equations for networks is similar to polymer [McCoy and Madras, 2001], and crystallization [Madras and McCoy, 2002b] kinetics. For example, the loss of $\mathbf{P}(\xi)$ on the left-hand side of Eq. 4-3.1 is the product between $p(\xi, t)$ and $p(\xi', t)$, however, if $\mathbf{P}(\xi')$ is abundant, we can assume that $\mathbf{P}(\xi)$ is the limiting reactant. Therefore, the loss of $\mathbf{P}(\xi)$ is expressed as,

$$-k_g(\xi) p(\xi, t) \quad (4-3.3)$$

Similarly, the removal of a connection between $\mathbf{P}(\xi)$ and $\mathbf{P}(\xi')$ is proportional to $p(\xi, t)$. The node addition or node removal rate with ξ_i connections can be expressed as $I_i(t) \delta(\xi - \xi_i)$, or can be incorporated into boundary conditions. Based on these preliminary concepts we write the population dynamics equation for Eq. 4-3.1 with generation and loss terms,

$$\begin{aligned} \partial p(\xi, t) / \partial t = & -k_g(\xi) p(\xi, t) + k_g(\xi-1) p(\xi-1, t) \\ & - k_d(\xi) p(\xi, t) + k_d(\xi+1) p(\xi+1, t) + \sum_{i=0} I_i(t) \delta(\xi - \xi_i) \end{aligned}$$

$$\begin{aligned}
&= \gamma [(\xi-1)^\lambda p(\xi-1,t) - \xi^\lambda p(\xi,t)] \\
&\quad + \kappa [(\xi+1)^\nu p(\xi+1,t) - \xi^\nu p(\xi,t)] + \sum_{i=0} I_i(t) \delta(\xi-\xi_i) \quad (4-3.4)
\end{aligned}$$

By substituting rate coefficients in Eq. 4-3.2, we obtain Eq. 4-3.4, similar to a master equation with first-order kinetics [Kampen, 1992]. The population dynamics equation for network growing process is similar to how crystallization [Madras and McCoy, 2002b] or polymerization [Madras and McCoy, 2002a] affords growth by monomer addition for clusters or polymers. Because we will describe node insertion using the boundary conditions, we set the source terms to zero, $I_i = 0$. We expand the distribution in $\xi_{\pm 1}$ in a series around ξ , as in other distribution kinetics applications [Madras and McCoy, 2002a; Madras and McCoy, 2002b], and obtain a Fokker-Planck equation from Eq. 4-3.4,

$$\partial p(\xi,t)/\partial t = \partial \{ [k_d(\xi) - k_g(\xi)] p(\xi,t) \} / \partial \xi + \frac{1}{2} \partial^2 \{ [k_d(\xi) + k_g(\xi)] p(\xi,t) \} / \partial \xi^2 + \dots \quad (4-3.5)$$

where we have omitted third- and higher-order terms.

4-4. Exponential Networks: Absence of Preferential Attachment

We keep up to the second-order terms in Eq. 4-3.5,

$$\partial p(\xi,t)/\partial t = \partial [(k_d(\xi) - k_g(\xi)) p(\xi,t)] / \partial \xi + \frac{1}{2} \partial^2 [(k_d(\xi) + k_g(\xi)) p(\xi,t)] / \partial \xi^2 \quad (4-4.1)$$

A one-dimensional random walk and its expression as a convective diffusion equation are similar to Eqs. 4-3.4 – 4-4.1, suggesting how a Gaussian distribution for the exponential network is obtained when the rate coefficients are constants [Chandrasekhar, 1943; Feller, 1957], $k_g(\xi) = \gamma$ and $k_d(\xi) = \kappa$. For this case, connections are randomly established and yield exponential networks. By substituting a "velocity," $v = (\gamma - \kappa)$, and a "diffusivity," $D = (\gamma + \kappa)/2$, into Eq. 4-4.1, the convective diffusion equation can be expressed as,

$$\partial p(\xi,t)/\partial t = -v \partial [p(\xi,t)] / \partial \xi + D \partial^2 [p(\xi,t)] / \partial \xi^2 \quad (4-4.2)$$

In stochastic theory, Eq. 4-4.2 is also called the Fokker-Planck or Kolmogorov forward equation, and diffusivity and velocity correspond to the constant infinitesimal mean and variance [Cox and Miller, 1965]. By a Fourier transformation of Eq. 4-4.2 and the initial condition, $p(\xi, \theta=0) = p_o^{(0)} \delta(\xi - \xi_o)$, which means $p_o^{(0)}$ nodes with ξ_o connections exist initially, the exact solution can be obtained. Two additional boundary conditions are required to solve the forward equation. The first condition is $p(\xi \rightarrow \infty, t) = 0$, which means no node can have an unlimited number of connections (there should be a maximum number of connections per node). Because ξ is a positive integer ($\xi \geq 0$) for the network systems, a typical second condition, $p(\xi \rightarrow -\infty, t) = 0$, is not realistic. If the solution peak, located at ξ_o initially, is far enough away from $\xi = 0$, the boundary condition does not affect the peak. In convective diffusion theory [Chandrasekhar, 1943], if the Peclet number for this system defined as $N_{Pe} = v\xi_o/D = 2(\gamma - \kappa)\xi_o / (\gamma + \kappa)$ is much greater than 1, the boundary condition [Levenspiel and Smith, 1957] can be approximated by $p(\xi \rightarrow 0, t) = 0$, which ensures that the network excludes nodes without connection. As ξ_o , where the initial distribution is positioned, becomes larger, the Peclet number is larger

The solution for the convective diffusion equation, Eq. 4-4.2, can be approximated by a Gaussian distribution for ξ and t , if $N_{Pe} \gg 1$,

$$p(\xi, t) = [p_o^{(0)} / (4\pi / N_{Pe})^{1/2}] \exp[-(\xi/\xi_o)^2 N_{Pe}/4] \quad (4-4.3)$$

By applying integration in Eq. 4-2.2, the moments of Eq. 4-4.3 are readily found as

$$p^{(0)}(t) = p_o^{(0)} \quad (4-4.4)$$

$$p^{(1)}(t) = p_o^{(0)} [(\gamma - \kappa) t + \xi_o] \quad (4-4.5)$$

The average moment and the variance are

$$p^{avg}(t) = (\gamma - \kappa) t + \xi_o \quad (4-4.6)$$

$$p^{\text{var}}(t) = (\gamma + \kappa) t \quad (4-4.7)$$

If the growth rate coefficient, $k_g(\xi) = \gamma$, is greater than the dissolution rate coefficient, $k_d(\xi) = \kappa$, nodes are being connected, the average number of connections, corresponding to the degree distribution, $\frac{1}{2} p^{\text{avg}}(t)$, increases, and the network grows. If $\gamma < \kappa$, the average moment of connections decreases and the network deteriorates by connection removal. For either growth or breakage, the network variance increases according to Eq. 4-4.7. As described in Eqs. 4-4.4 and 4-4.6, the degree distribution for this network with fixed number of nodes does not change with network size because the number of nodes is constant. Thus, the network exemplifies how the total number of connections among a constant number of nodes increases with time in the absence of preferential attachment.

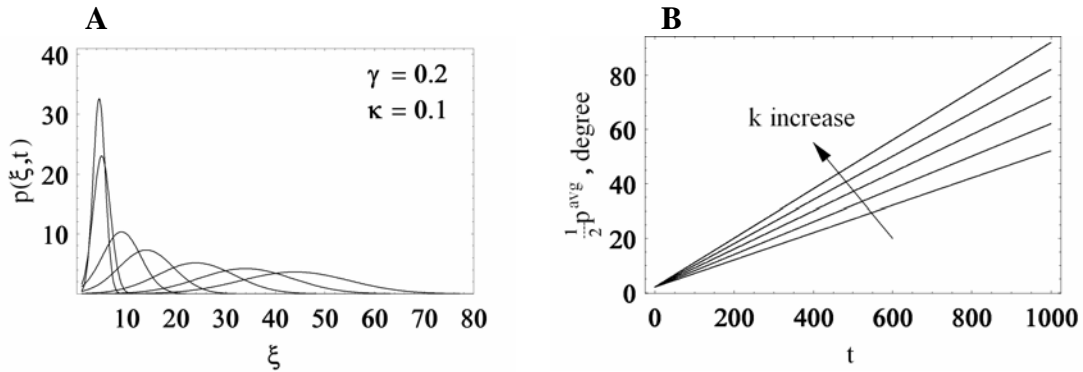


Figure 4-4.1 Evolution of the Gaussian distributed network for growth with $p_o^{(0)} = 100$ and $\xi_o = 4$ based on the moment results in Eqs. 4-4.4 – 4-4.6. The scaled times in (A) are $t = 5, 10, 50, 100, 300, 500, 700$ and the growth rate $k = \gamma - \kappa$, in (B) increases in steps of 0.02 from 0.1 to 0.2.

Figure 4-4.1 shows how the degree distribution of Gaussian distributed exponential networks grows with time. The growing behavior is illustrated in Fig. 4-4.1 showing the growth of Gaussian distributed exponential networks. Figure 4-4.1B shows time dependence of degree distribution increasing with time. As plotted in Fig. 4-4.1B, if

the growth rate, $k = \gamma - \kappa$, increases, the degree distribution with a constant number of nodes increases.

The current model, allowing multiple connections between any two nodes, is more general than the connection-limited networks (only one connection between any two nodes). For the connection-limited networks, the most effective structure is the saturated state, where all nodes are connected with all other nodes. But due to the network growth constraints, such as connecting or organization costs [Mattick and Gagen, 2005], the saturated state may not occur even for the connection-limited networks.

The discrete Poisson distribution for the exponential network derives from more constrained conditions: irreversible network growth ($k_d = 0$), lack of source terms ($I_i = 0$), and constant rate coefficient ($k_g = \gamma$). With these restrictions, Eq. 4-3.4 can be written as,

$$\partial p(\xi, t) / \partial (\gamma t) = -p(\xi, t) + p(\xi - 1, t) \quad (4-4.8)$$

where only positive integers ($\xi > 0$) are considered. Equation 4-4.8 is a first-order difference-differential equation similar (but not identical) to governing equations in chain polymerization [McCoy and Madras, 2001] and stirred-tank cascade modeling [Dotson et al., 1996]. The initial and boundary conditions are $p(\xi, t=0) = p_o^{(0)} \delta_{0\xi}$ and $p(\xi < 0, t) = 0$; the initial condition is expressed in terms of the Kronecker delta, representing unconnected node insertion, and the boundary condition ensures the variable ξ is positive. Equation 4-4.8 can be solved by Laplace transformation and the solution is similar to a Poisson distribution [McCoy and Madras, 2001],

$$p(\xi, t) = p_o^{(0)} e^{-\gamma t} (\gamma t)^{\xi+1} / (\xi+1)! \quad (4-4.9)$$

The moments for the distribution by the calculation in Eq. 4-2.2 are,

$$p^{(0)}(t) = p_o^{(0)} (1 - e^{-\gamma t}) \quad (4-4.10)$$

$$p^{(1)}(t) = p_o^{(0)} (\gamma t - 1 + e^{-\gamma t}) \quad (4-4.11)$$

The average moment, which has a long time limit, γt , and the variance are,

$$p^{\text{avg}}(t) = (\gamma t - 1 + e^{-\gamma t}) / (1 - e^{-\gamma t}) \quad (4-4.12)$$

$$p^{\text{var}}(t) = \gamma t e^{\gamma t} (e^{\gamma t} - \gamma t - 1) / (1 - e^{\gamma t})^2 \quad (4-4.13)$$

The variance also shows a linear time behavior (degree $\sim \gamma t$) for large values of time.

Thus, similar to the Poisson distribution, the average and variance of Eq. 4-4.9

asymptotically reach the same expressions at long time.

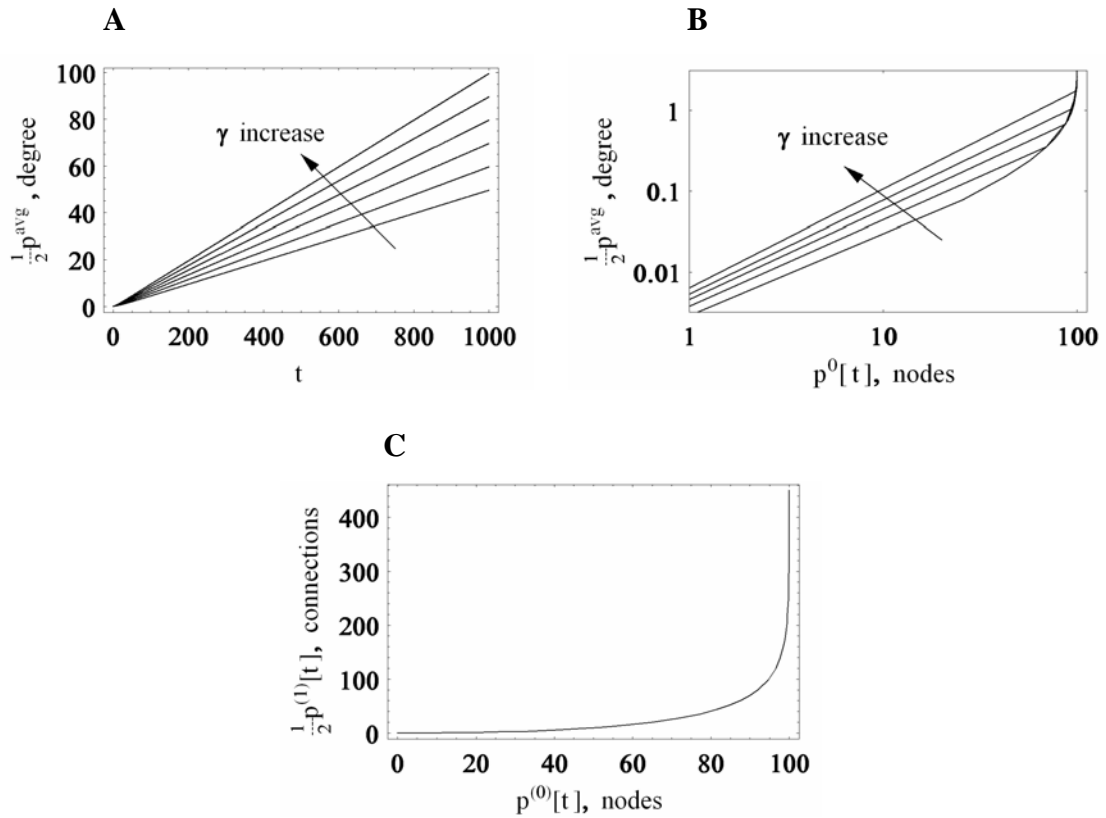


Figure 4-4.2 Evolution of the Gaussian (A) and Poisson (B) distributed network growth based on the moment results in Eqs. 4-4.4 – 4-4.6 (A) and Eqs. 4-4.10 – 4.4.12 (B): $p_o^{(0)} = 100$, $\xi_o = 4$, and the growth rate $k = \gamma - \kappa$, increases in steps of 0.02 from 0.1 to 0.2 for (A), and $p_o^{(0)} = 100$ and γ increases in steps of 3 from 1 to 13 for (B).

Figure 4-4.2 shows how the degree distribution of Poisson distributed exponential network grows with time (Fig. 4-4.2A) and the number of nodes (Fig. 4-4.2B) and the number of connections with the number of nodes (Fig. 4-4.2C). With the absence of preferential attachment, the Poisson distributed network in Fig. 4-4.2A exemplifies how the total number of connections among a constant number of nodes increases with time, the Poisson distributed network in Fig 4-4.2B (log-log coordinates) demonstrates nonlinear accelerating growth of degree distribution with network size (the total number of connections grows faster than the total number of nodes), and the total number of connections in Fig. 4-4.2C shows nonlinear increase with the network size expressed by the total number nodes.

As plotted in Fig. 4-4.2B, if the growth rate, γ , increases, the degree distribution with a constant number of nodes increases. The diverging behavior of degree distribution explains that the model allows multiple connections between nodes. The linear behavior on log-log coordinates (Fig. 4-4.2B) implies that the degree distribution is proportional to a power of network size.

4-5. Power Law Networks: Effect of Preferential Attachment

Many complex systems have power law size distributions. The well-known mechanism to produce power law networks is preferential attachment [Barabasi and Albert, 1999] where a new node introduced into networks preferentially connects with highly connected ones. To describe the evolution of power law distributed networks, we truncate Eq. 4-3.5 to first-order with the power expression of the rate coefficients (Eq. 4-3.2, with $\lambda = \nu \neq 0$).

$$\partial p(\xi, t) / \partial t + \partial [G p(\xi, t)] / \partial \xi = 0 \quad (4-5.1)$$

where $G = (\gamma - \kappa)\xi^\lambda$ is the growth rate. This partial differential equation, similar to a continuity equation, is common to population balance modeling [Randolph and Larson, 1986], and can be solved by Laplace transformation. We examine two different cases with different initial conditions. We first consider the initial condition $p(\xi, t=0) = 0$, which means initially no nodes exist, and the boundary condition, $p(\xi=1, t) = p_o^{(0)} e^{t/\tau}$, which represents the number of nodes with one connection increasing exponentially. The distribution for this set of initial and boundary conditions can be obtained by Laplace transformation as well as the method of characteristics,

$$p(\xi, t) = p_o^{(0)} \xi^{-\lambda} \exp[t/\tau + (1 - \xi^{1-\lambda})/k\tau(1 - \lambda)] u[t - (\xi^{1-\lambda} - 1)/k(1-\lambda)] \quad (4-5.2)$$

where $u(x)$ is the unit step function defined as $u(x < 0) = 0$ and $u(x \geq 0) = 1$. The moments of Eq. 4-5.2 can be obtained by integration (Eq. 4-2.2).

The analytical solution for general power λ is complicated, and we will show the evolution and network size dependence graphically. Here we list the simplest moment results where node connection probability is linearly proportional to node degree, $\lambda = 1$.

Eq. 4-5.2 for $\lambda = 1$ can be written as,

$$p(\xi, t) = p_o^{(0)} \xi^{-(1 + k\tau)/k\tau} \exp(t/\tau) u[t - \ln(\xi)/k] \quad (4-5.3)$$

For this linearly proportional connection probability, the n^{th} moment obtained by integration within a finite domain is,

$$p^{(n)}(t) = p_o^{(0)} k\tau [\exp(t/\tau) - \exp(nkt)] / (1 - nkt) \quad (4-5.4)$$

Thus the zeroth, first, and average moments are,

$$p^{(0)}(t) = p_o^{(0)} k\tau [\exp(t/\tau) - 1] \quad (4-5.5)$$

$$p^{(1)}(t) = p_o^{(0)} k\tau [\exp(kt) - \exp(t/\tau)] / (kt - 1) \quad (4-5.6)$$

and

$$p^{\text{avg}}(t) = [\exp(kt) - \exp(t/\tau)] / \{(k\tau - 1)[\exp(t/\tau) - 1]\} \quad (4-5.7)$$

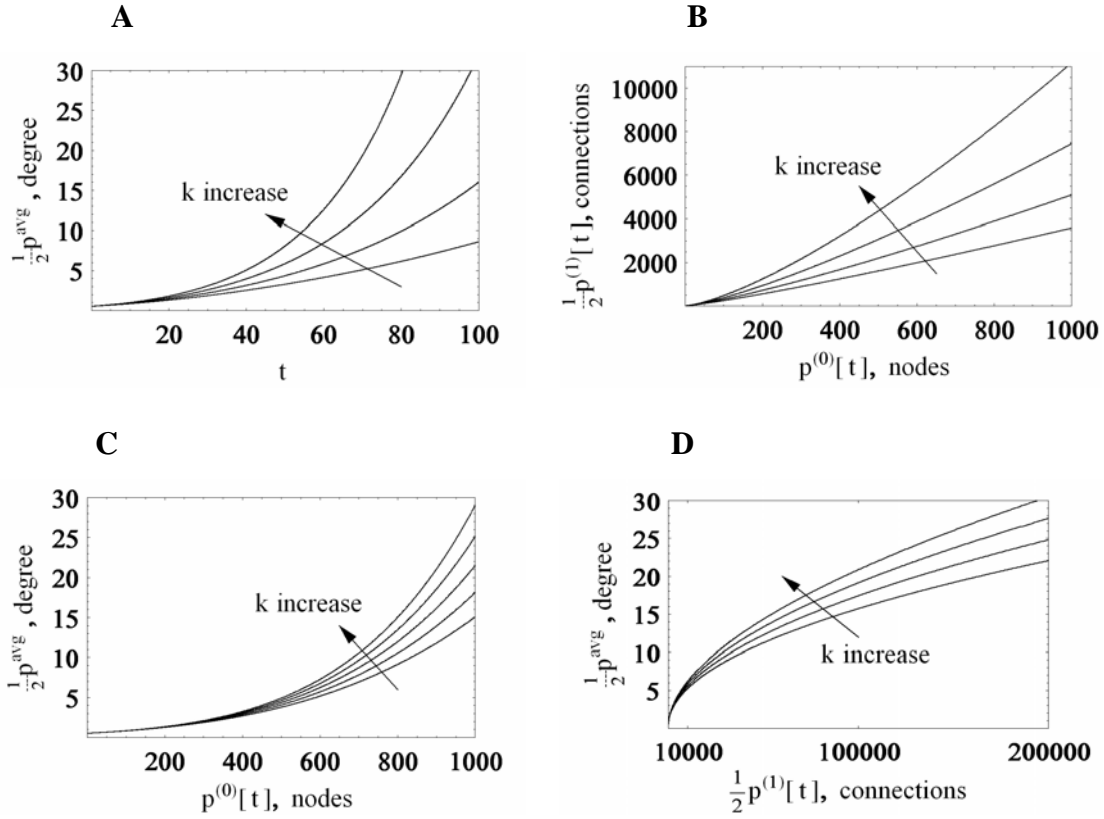


Figure 4-5.1 Evolution of moments of the power law distributed network (Eqs. 4-5.5 – 4-5.7) with initial condition, $p(\xi, t=0) = 0$ and boundary condition, $p(\xi=1, t) = p_o^{(0)} e^{\theta/\tau}$. The growth rate k increases in steps of 0.01 from 0.11 to 0.14 for parameters $\lambda = 1$, $p_o^{(0)} = 100$, $\xi_m = 1000$, $\tau = 10$.

As explained, the zeroth moment, $p^{(0)}(t)$, represents the total number of nodes, and the total number of connections can be represented by $\frac{1}{2} p^{(1)}(t)$. Only a particular network [Jeong et al., 2000] shows linear growth, where the number of connections is linearly proportional to network size. However, many networks display accelerating nonlinear growth, where the total number of connections grows faster than the total number of nodes, for example, the communication networks (the WWW and Internet), citation networks, and collaboration networks [Amaral et al., 2000; Baran, 1964; Huberman et al.,

1998; Lahererre and Sornette, 1998; Newman, 2001; Render, 1998] We now demonstrate such nonlinear, accelerating network growth showing evolution of a power law distribution. Figure 4-5.1A shows the time dependence of power law networks with proposed initial and boundary conditions, Fig. 4-5.1B demonstrates nonlinear accelerating growth of the degree distribution with network size, and Figs. 4-5.1C and 4-5.1D illustrate accelerating growth of the degree distributions with the number of nodes and connections.

We next consider the initial condition $p(\xi, t=0) = p_o^{(0)} \xi^{-\lambda}$, where connections are initially distributed as a power law, and the exponentially increasing boundary condition, $p(\xi=1, t) = p_o^{(0)} e^{t/\tau}$. The network distribution for this set of initial and boundary conditions is,

$$p(\xi, t) = p_o^{(0)} \xi^{-\lambda} \{1 - \{1 - \exp[t/\tau + (1 - \xi^{1-\lambda})/k\tau(1-\lambda)]\} u[t - (\xi^{1-\lambda} - 1)/k(1-\lambda)]\} \quad (4-5.8)$$

As explained, the general moment of Eq. 4-5.8 is also complicated. Therefore, we will show its simplest expression and explain its behavior graphically. When $\lambda = 1$, Eq. 4-5.8 can be simplified as,

$$p(\xi, t) = (p_o^{(0)}/\xi) \{1 - [1 - \exp(t/\tau) - \xi^{-1/k\tau}] u[t - (\xi^{1-\lambda} - 1)/k(1-\lambda)]\} \quad (4-5.9)$$

and its general n^{th} moment obtained by integration is,

$$p^{(n)}(\tau) = p_o^{(0)} [\exp(nkt) - nk\tau \exp(t/\tau) + \xi_m^n (nk\tau - 1)] / [n(nk\tau - 1)] \quad (4-5.10)$$

and the zeroth, first, and average moments are,

$$p^{(0)}(t) = p_o^{(0)} \{ \ln(\xi_m) - k[t + \tau(1 - \exp(t/\tau))] \} \quad (4-5.11)$$

$$p^{(1)}(t) = p_o^{(0)} [\exp(kt) - k\tau \exp(t/\tau) + \xi_m(k\tau - 1)] / (k\tau - 1) \quad (4-5.12)$$

and

$$p^{\text{avg}}(t) = [\exp(kt) - k\tau \exp(t/\tau) + \xi_m(k\tau - 1)] / \{(k\tau - 1)[\ln(\xi_m) - k\{t + \tau[1 - \exp(t/\tau)]]\} \quad (4-5.13)$$

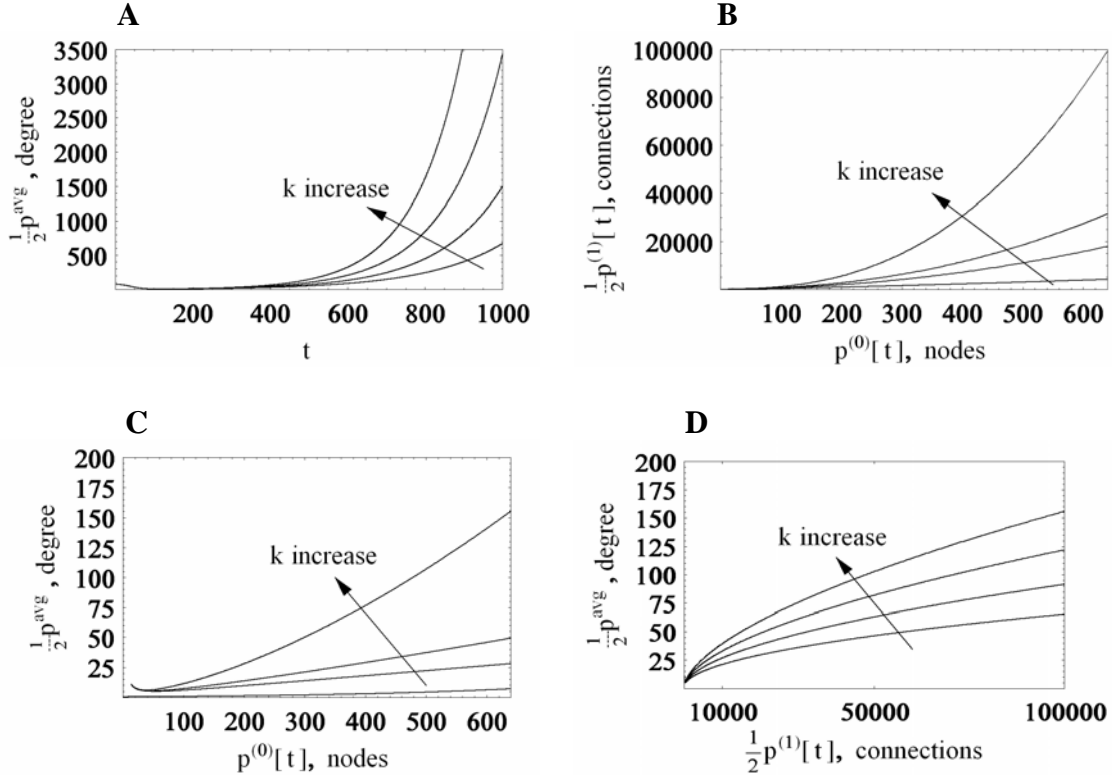


Figure 4-5.2 Evolution of the power law distributed network with initial condition, $p(\xi, t=0) = p_o^{(0)}\xi^{-\lambda}$ and boundary condition, $p(\xi=1, t) = p_o^{(0)} e^{\theta/\tau}$, using moment expressions in Eqs. 4-5.11 – 4-5.13. The growth rate k increases in steps of 0.001(A) from 0.055 to 0.058, 0.01 (B-D) from 0.1 to 0.13: $\lambda = 1$, $p_o^{(0)} = 100$, $\xi_m = 1000$, $\tau = 20$.

If k and τ are large, Eqs. 4-5.10 – 4-5.13 reduce to Eqs. 4-5.4 – 4-5.7, and the degree distribution in Fig. 4-5.2 shows nonlinear accelerating behavior similar to Fig. 4-5.1.

Figure 4-5.2A shows time dependence of power law networks with the proposed initial and boundary conditions, Fig. 4-5.2B displays the non-linearly growing total number of connections with the total number of nodes, and Figs. 4-5.2C and 4-5.2D demonstrate accelerating growth of the degree distribution with network size.

As expressed in Eqs. 4-5.4 – 4-5.13 and Fig. 4-5.1, the model for power law networks contains two parameters, k and τ , which allow quantitative description of many nonlinearly growing systems: As k (growth rate) increases, the network size and nonlinearity, the ratio between the total number of connections and nodes, increases. As τ (node addition intensity) increases the total number of connections exceeds the maximum number of connections of connection-limited networks, $p^{(0)}(t)(p^{(0)}(t)-1)/2$, indicating multiple connections between nodes.

It is also interesting that when $\tau k/2 = 1$, the number of nodes represented by the zeroth moments in Eqs. 4-5.5 and 4-5.11 is approximated as $p^{(0)}(t) \sim \exp[kt/2]$. The number of connections expressed by the first moments in Eqs. 4-5.6 and 4-5.12 is written as $\frac{1}{2} p^{(1)}(t) \sim (\exp[kt/2])^2$, and therefore the number of connections increases quadratically with network size, $\frac{1}{2} p^{(1)}(t) \sim (p^{(0)}(t))^2$. Because the degree distribution is defined by the average moment (degree = $\frac{1}{2} p^{avg}(t)$ and $p^{avg}(t) = p^{(1)}(t)/p^{(0)}(t)$), this shows linear behavior with the total number of nodes, thus, degree $\sim p^{(0)}(t)$. The quadratic increase of number of connections with number of nodes was reported to apply for supercomputers and regulatory gene networks [Mattick and Gagen, 2005].

A vector-parallel high-performance computer developed by ESRDC (the Earth Simulator Research and Development Center)/NEC, the Earth Simulator is registered as the world's fastest supercomputer with 35.61 TFlops (trillion operations per second) according to Linpack benchmark test results. The Earth Simulator consists of 640 supercomputers, one at each node, with 8 vector processors for a total of 5120 processors connected by a high-speed network with 12.3 GBytes data transfer speed. Recent

research [Mattick and Gagen, 2005; Sato, 2004] indicates that more than 400,000 connections are required to connect 640 nodes by 83,000 wires.

For the gene regulatory network of single-celled prokaryotic organisms, general arguments on the network control indicate that regulatory gene number grows relatively fast with genome size increasing, and the number of transcriptional regulators scales quadratically with the total number of genes. Recent studies [Gagen and Mattick, 2005; Croft et al., 2003] on the gene regulatory network reveals that the percentage of regulatory gene increase from 2.5% to 9% as the network size grows from 4,000 to 8,000 bacterial genes. The model predictions are compared with the accelerating behaviors of the Earth Simulator and the gene regulatory network [Mattick and Gagen, 2005] in Fig. 4-5.3.

Real-world complex systems such as the Earth Simulator, bacteria, or business organizations show nonlinear accelerating growth behavior [Mattick and Gagen, 2005], because of the connection and organization costs. As a test we compare model results with data for a supercomputer (Earth Simulator) and for regulatory gene networks [Mattick and Gagen, 2005]. As explained for the connection-limited networks, the most connected network is the most effective one, for example, when a new node introduced into networks has connections with all other nodes. This will appear as a straight line on a scaled plot of degree with network size. Predicting network growth path by using the model solution is straightforward and convenient as pictured in the scaled plots, Figs. 4-5.3A and 4-5.3B with different k . We normalized the axes, the degree and the number of nodes in Fig. 4-5.3 by using values at $t = 100$ (which is sufficient time to see nonlinear accelerating behavior). When the parameter k is 0.1, the straight line, which corresponds to the most effective network for the connection-limited network, was obtained, i.e., $\tau k/2$

= 1.0. The model prediction for data in Fig. 4-5.3 shows nonlinear growth of degree distribution, $\frac{1}{2} p^{\text{avg}}$, with network size, $p^{(0)}(t)$, through quadratic increases in the total number of connections, $\frac{1}{2} p^{(1)}(t)$. With $k = 0.15$, the model adequately describes the data for accelerating network growth.

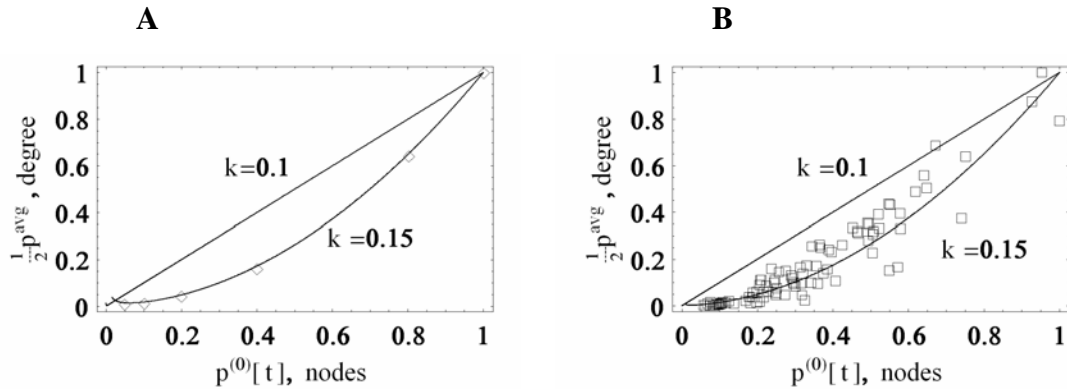


Figure 4-5.3 The scaled degree versus total number of nodes for the model (Eqs. 4-5.5 – 4-5.7 with $\lambda = 1$, $p_o^{(0)} = 100$, $\xi_m = 1000$, and $\tau = 20$). The lines are model predictions and symbols indicate data [Mattick and Gagen, 2005] for (A) Earth simulator and (B) regulatory gene networks. The predictions with $k = 0.15$ demonstrate that the model can describe real-world accelerating network growth. The predictions with $k = 0.1$ ($k\tau/2 = 1$) show the most effective network growth path for connection-limited networks.

The network study can be classified as several categories depending on the behavior they analyze, e.g., degree distribution, diffusion, percolation, clustering, or evolution. The model cannot be applied to evaluate all of these quantities. However, as we mentioned in the introduction, the aim of the work is to present a general model for accelerating network evolution. Quantities such as node separation or clustering coefficients were previously calculated [Dorogovtsev and Mendes, 2003; Park and Newman, 2005] by Monte Carlo simulations, and thus we do not cover the calculation of such network properties. We, therefore, focus on the temporal evolution of accelerating

networks, which gave quantitatively correct predictions for the real-world accelerating network growth.

4-6. Conclusion

What kind of mechanism do growing networks follow? How do they construct and maintain their exponential or power law structures? To answer these questions, we have suggested a model based on population balance dynamics (distribution kinetics). The approach shows how continuous distributions can describe network dynamics and how either exponential or power law networks can be constructed. Our aim has been to introduce a generalized model for growing networks.

For the distributions in population balance kinetics, we have proposed a growing network model with and without preferential attachment, i.e., power law and exponential networks, respectively. The model with size-independent rate coefficients, $k_g(\xi) = \gamma$ and $k_d(\xi) = \kappa$, yields Gaussian or Poisson distributed exponential networks. The model with the size-dependent rate coefficients, $k_g(\xi) = \gamma\xi^\lambda$ and $k_d(\xi) = \kappa\xi^\nu$, produces power law networks, and the nonlinear network size dependence of the number of connections describes how such power law networks evolve. We explored the nonlinear growth of power law degree distribution with time and network size. Our model for power law network evolution has two parameters: τ , which controls node addition, and the key parameter, k , which is the difference between pre-factors of the rate coefficient ($\gamma - \kappa$) and manages network-growing intensity. For real-world complex systems, k reflects changes of the internal or external conditions of networks, such as the accumulation of connection load [Holme and Kim, 2002; Moreno et al., 2002], which increases

connecting and organization costs, or technological innovations, which diminish the costs.

Moment results show the time dependence of the degree distribution, $p(\xi, t)$, for either linearly or nonlinearly growing networks. Power law expressions with unlimited ξ do not have proper integrals, and thus their moments are indeterminate. But as we have demonstrated, power laws evolve to increasingly larger values of ξ , and thus moment integrals can be defined for finite time. A general moment equation was derived by the integration of the moment definition in Eq. 4-2.2 when integer values of λ and ν are identical, $\lambda = \nu$. The moments, $p^{(n)}(t)$, $p^{\text{avg}}(t)$, and $p^{\text{var}}(t)$, have complicated expressions that can be derived by computer algebra. We also derived time and network size dependent behaviors of the model using moment solutions, determined based on the initial and boundary conditions.

The growing network model based on distribution dynamics can describe growing network systems and represent data. The degree distributions of the network model for exponential and power law networks increase with time and network size through nonlinear accelerating growth (Figs. 4-4.1 – 4-5.2). The proposed model, allowing multiple connections between any two nodes, is general compared to the connection-limited networks, which allow only one connection between two nodes. We have demonstrated that the model is also able to describe accelerating nonlinear growth of networks by plotting degree distributions and the total number of connections with network size.

The aim of this paper was to present a model for accelerating network evolution by developing a framework for the dynamics of linear and nonlinear growing networks.

We focused on accelerating networks [Mattick and Gagen, 2005] and how the number of nodes and connections and the degree distribution evolve in time. The model in its present state treats only those quantities that can be represented as moments of the distribution, $p(\xi, t)$, and thus not quantities such as node separation or clustering coefficient, which have been computed via Monte Carlo simulations [Dorogovtsev and Mendes, 2003]. At present, no single model quantitatively describes all possible phenomena associated with networks. Population balance (nonequilibrium distribution kinetics) modeling can incorporate additional qualities as added parameters and variables in the distribution function, and this is the subject of continuing investigation.

CHAPTER 5. DISTRIBUTION KINETICS OF HUMAN DYNAMICS

5-1. Introduction

A wide range of human activities produces the many social, technological, and economic phenomena that illustrate diverse human dynamics. Individuals in modern society perform a large number of tasks daily, ranging from personal to social to work activities. The ubiquitous appearance of coherent macroscopic patterns in human activities has led to the introduction of several theoretical models aimed at understanding how human activities distribute and evolve in time. More recently, approaches based on continuous deterministic [Alien et al., 1977] or stochastic diffusion [Cox and Miller, 1965] models, or kinetic particle diffusion [Schweitzer and Steinbrink, 1997] have appeared.

Among many human activities, problems of practical interest require us to understand patterns of such actions. Typical examples are the design of telephone systems or web servers, where it is critical to know how many will use the service simultaneously. When individuals perform tasks based on job performing priority, the timing properties of the tasks are heavy tailed and thus well modeled by power law distributions. Most tasks have short waiting times whereas a few have very long waiting times [Barabasi, 2005], a prominent feature of power law distributions. The power law or Pareto distribution (a cumulative version of the power law) also has been reported in a number of physical and human initiated phenomena: distributions of incomes (or incomes exceeding a minimum value), sizes distribution of asteroids, islands, and cities, and extinction events [Kauffman, 1993; Mandelbrot, 1963]. In communications, power law distributions have been used to model telephone call holding times [Duffy et al, 1994],

and frame sizes for variable-bit-rate video [Garrett and Willinger, 1994]. The model of exponential arrival times has been shown to be inadequate for describing wide area network traffic [Faxon and Floyd, 1995], including local-area network (LAN) and wide-area network (WAN) traffic and the distribution of packet inter-arrivals. These are clearly different from random-order protocols, which assume uniform distribution of events. Some examples of heavy tail distributions in computer systems include: computer networks both in terms of their connectivity [Willinger et al., 2002] and their traffic patterns [Willinger et al., 1996], file systems [Gribble et al., 1998], video traffic [Beran et al., 1995], and software caches [Voldman et al., 1981], and the job size distributions on a single processor [Harchol-Balter, 1999] as well as on supercomputers [Feitelson, 2000]. These power law distributions also have important implications for the phenomena, indicating a significant possibility of huge deviations from average, which can be interpreted as overload of machine capacity for the case of supercomputers, a critical and damaging earthquake for the earthquake size distributions, and quite long waiting-times for event completion in the timing of human activities.

In executing human activities, each individual arranges and orders the tasks based on criteria such as relative importance, personal preference, and amount of time required for completing each task. The timing of human activities obviously varies in a complex way, and inevitable characteristics of human dynamics--diversity and irregularity--complicate prediction. Since individuals are performing one task at a time, all tasks should inevitably wait to be performed, and thus waiting-time of tasks will vary with their own criteria, whether random or prioritized. Previous models to describe human-initiated phenomena assume that waiting time, the time interval between two consecutive actions performed by the same individual, is randomly distributed, and can be described

by exponential distributions for Poisson processes [Haight, 1967; Reynolds, 2003]. If activity patterns are assumed homogeneous, human behavior can be modeled by a Poisson process, that is, time intervals will distribute following an exponential distribution. Although such processes have been commonly postulated, empirical evidence [Barabasi, 2005; Faxon and Floyd, 1995] indicates that human activity patterns are rather heterogeneous with short waiting-times of many activities and relatively long waiting-times of a few activities. The incompatibility of the Poisson model has been recognized for human behaviors as well as for the electronic information networks such as telephone circuits [Berger and Mandelbrot, 1963] and network traffic [Leland et al., 1994]. For the distribution of human activities, the power law or Pareto distribution provides a better approximation for waiting-time distributions. These phenomena are also found in the email networks [Eckmann et al., 2004] and human printing behavior [Harder and Paczuski, 2005]; time intervals for sending and responding to emails are well-described by the power law distribution based on heterogeneous statistics, not by the Poisson model based on uniform inter-event statistics. For human printing behavior, the distribution of file sizes and timing of printing jobs can be described by a truncated power law, and thus, waiting-times between individual requests are broadly distributed from seconds to weeks. To forecast human activities, Barabasi [Barabasi, 2005] suggested that most human dynamics should be represented by the power law and not the Poisson model. The tendency can be found in many physical and social systems with underlying structures, e.g., the size distribution of clusters [Lejeune et al., 2003], corporations [Adamic et al., 2001], and cities [Marsili and Zhang, 1998].

Recently an application [Vazquez, 2005] of the Barabasi model was proposed to provide exact results for statistics of stochastic human activities. The model explained the

appearance of a power law and self-scaling keeping all moments finite. It reproduced waiting-time distributions quite accurately, particularly waiting-times for email servers. Here we attempt to address the problem from a different point of view, assuming that the evolution of the waiting-time distribution can be modeled by an approach based on cluster distribution kinetics. This allows us to calculate the waiting-time distribution of human activities that are either uniform or heterogeneous. The aim of the present work is to introduce by a deterministic approach an alternative explanation of the power law representation of human dynamics.

5-2. Distribution Kinetics of Human Activities

Compared to computers that execute several tasks at the same time by parallel computing, most human activities cannot be performed simultaneously. Whenever an individual has a series of tasks to perform, ordering the tasks based on job specific criteria is necessary, and new tasks are also added to the list based on these criteria. The majority of tasks await a moment to be executed, and such time intervals are distributed as a waiting-time distribution. From this perspective, human-initiated working patterns can be classified by three selection protocols [Cohen, 1969]: the first-in-first-out protocol, the random-order protocol, which disregards job priority, and the priority-relevant protocol, which arranges human jobs according to their priorities. The first protocol is straightforward and does not need a mathematical model for its explanation. Separate models, however, are required for the other two protocols: a model for randomly distributed human activities by the random-order protocol, and a model for power law distributed human activities by the priority-relevant protocol. These selection protocols and mechanisms to produce networks [Albert and Barabasi, 2002] have points of similarity; for examples, the random-order protocol can be related to the node connection

mechanism for exponential networks, and the priority-relevant protocol corresponds to the preferential attachment for power law networks. As described, in contrast to the random-order protocol of human activities, which assumes a uniform distribution, the priority-relevant protocol is based on the assumption that the time intervals between human activities are not uniform but rather heterogeneous. The model we present here can describe human dynamics for the second and third selection protocols by changing rate coefficient expressions. We define the distribution of human activities such that $p(\xi, t)d\xi$ represents the number of events at time t in the differential waiting-time range $(\xi, \xi+d\xi)$. The goal behind the method is to explore how a general and simple model might describe complex patterns of human activity distribution.

First we consider the differences among the power law distribution and its other forms, the Pareto and Zipf laws, which can be applied to systems consisting of a small number of rare events and large number of common events, depending on the quantity used in ordering the events. In a clustered system, for example, if y is the size of cluster and r is its rank (the r^{th} largest cluster), the Zipf law cluster distribution with power exponent b is expressed as $y \sim r^{-b}$, showing how the cluster size decreases with its rank. On the other hand, Pareto's law, $r \sim y^{-1/b}$, is the inversion of Zipf's law, showing how many clusters have a number of monomers larger than y , $p(\xi \geq y) \sim \xi^{-1/b}$. The rank r , the r^{th} largest cluster, can also be interpreted as the number of clusters with y or more monomers. Because Pareto's law is cumulative, the power law distribution $p(\xi)$ is obtained by a derivative, $\xi^{-(1/b+1)}$, and we will define the power as $\lambda = (1/b + 1)$ so that $p(\xi) \sim \xi^{-\lambda}$ where $\lambda \geq 0$. Thus any system expressed as either Zipf's or Pareto's law can be transformed into a power law distribution. The cluster kinetics model we present here is

based on the first order Fokker-Planck equation, which also can be derived from a population balance equation [Jeon and McCoy, 2005a; Jeon and McCoy, 2005b]. The model has a large number of applications to the modeling of many physical and social phenomena, particularly in the field of non-linear particle or event-distributed systems [Jeon and McCoy 2005b]. The cluster distribution kinetics has been applied in many different fields, such as distribution of US firms, incomes, and cities. In this paper, we apply the cluster distribution kinetics to the waiting-time distribution of human activities, a phenomenon more related to statistical or social sciences than to physical sciences.

To motivate our model for human activity patterns in the distribution of human actions, we now discuss analogies to dynamical and statistical models developed to understand complex distributing processes. While not necessarily exact, the models proposed below are meant to capture the essential dynamics that occur in the job submission and completion processes.

Many dynamic, distributed systems are based on the general (second-order) Fokker-Planck equation [Cox and Miller, 1965; Feller, 1957],

$$\partial p(\xi,t)/\partial t = -\partial[v(\xi)p(\xi,t)]/\partial \xi + 1/2 \partial^2[d(\xi)p(\xi,t)]/\partial \xi^2 \quad (5-2.1)$$

where $v(\xi)$ and $d(\xi)$ are convection velocity and diffusion coefficient, respectively, defined as $v(\xi) = k\xi^\lambda$ and $d(\xi) = (k+c)\xi^\lambda$ [Jeon and McCoy, 2005b]. The constants k and λ are positive and c can be either positive or negative depending on the systems. Equation 5-2.1 is also known as the Kolmogorov forward equation in stochastic theory [Cox and Miller, 1965] or the convective diffusion equation in diffusion theory [Crank, 1975; Levenspiel and Smith, 1957]. The Poisson distribution, corresponding to the Eq. 5-2.1 with $\lambda = 0$, can be obtained from the difference-differential equation [Carson and Doyle,

1999], which is derived from the size independent Fokker-Planck equation. The difference-differential equation is similar to a fundamental equation in chain polymerization [McCoy and Madras, 2001] and stirred-tank cascade modeling [Dotson et al., 1996], or stochastic equations for the transition probability [Feller, 1957; Roehner and Valent, 1982]. Two examples are when monomers are added to the clusters one at a time regardless of the cluster size in the cluster systems, and when connections are made between nodes independent of their connectivity in a network [Jeon and McCoy, 2005a; Jeon and McCoy, 2005b].

To describe a waiting-time distribution of human activities, we write a first-order Fokker-Planck equation, which as mentioned can also be derived starting from a population balance equation [Jeon and McCoy, 2005b]. Since the effect of the second-order derivative term is not significant unless a diffusion-like process is in effect, we write the first-order differential equation,

$$\partial p(\xi, t) / \partial t + \partial [k \xi^\lambda p(\xi, t)] / \partial \xi = 0 \quad (5-2.2)$$

Equation 5-2.2 is satisfied even for $k = 0$ [Jeon and McCoy, 2005b], since the time derivative should be zero, and the system becomes either a dynamic or static equilibrium state.

The main advantage of the proposed model is the application flexibility. If $\lambda = 0$, the model yields an exponential distribution for the human activities driven by the random-order protocol corresponding to the Poisson process and showing exponential decay. If $\lambda > 0$, usually between 0 and 5 [Jeon and McCoy, 2005b], the model produces a power law distribution for the human behavior induced by the priority-relevant protocol, and corresponds to the Pareto distribution. We will see how such a parameter change affects to the model description through its analytical and graphical solutions.

To explain how the model describes the systems, we will derive results for $\lambda = 0$ and $\lambda > 0$. The first-order partial differential equation, Eq. 5-2.2, can be solved by Laplace transformation as well as by the method of characteristics. The Laplace transform general solution for the initial condition, $p(\xi, t=0) = f(\xi)$, and the boundary condition, $p(\xi=1, t) = g(t)$, is

$$p(\xi, s) = (\xi^{-\lambda}/k) \exp[-s\xi^{1-\lambda}/k(1-\lambda)] \int_1^\xi \exp[sy^{1-\lambda}/k(1-\lambda)] f(y) dy + g(s) \xi^{-\lambda} \exp[s(1-\xi^{1-\lambda})/k(1-\lambda)] \quad (5-2.3)$$

For the randomly distributed events with $\lambda = 0$, Eq. 5-2.3 reduces

$$p_{\lambda=0}(\xi, s) = \exp(-s\xi/k)/k \int_1^\xi \exp(s\xi/k) f(y) dy + g(s) \exp[s(1-\xi)/k] \quad (5-2.4)$$

and for the power law distributed events with $\lambda = 1$, Eq. 5-2.3 reduces

$$p_{\lambda=1}(\xi, s) = (\xi^{-(1+s/k)}/k) \int_1^\xi y^{s/k} f(y) dy + g(s) \xi^{-(1+s/k)} \quad (5-2.5)$$

To show the behaviors of exponential distributions, we consider $\lambda = 0$ with the zero initial condition ($p(\xi, t=0) = 0$), which means there is no event initially. The exponentially increasing boundary condition ($p(\xi=1, t) = p_0 e^{t/\tau}$) means the number of tasks with unit waiting-time is exponentially increasing with time. With these initial and boundary conditions, the waiting-time distribution for the random-order protocol is

$$p_{\lambda=0}(\xi, t) = p_0 \exp[t/\tau + (1-\xi)/k\tau] u[t - (\xi-1)/k] \quad (5-2.6)$$

where the function $u[x]$ is the unit step function, defined as $u(x < 0) = 0$ and $u(x \geq 0) = 1$.

The exponential distribution, Eq. 5-2.6, is expressed as straight lines on semilog coordinates showing exponential decay in the tail of the distribution. This behavior is common for distributions of random variables, for example, lifetime of light bulbs,

waiting time in the line at the post office, and when students are called on randomly in the classroom [Easton and McColl, 1997]. The distributions of such variables are also exponential and show a similar behavior with the exponential waiting-time distribution in Eq. 5-2.6.

A typical feature of many power law distributed systems is that they are truncated at large ξ . To describe such behavior, we next consider the case when $\lambda > 0$ with the same initial condition and the exponentially increasing boundary condition up to a finite value, $p(\xi=1,t) = p_0 (1 - e^{-t/\tau})$. Derived by Laplace transformation, the power law distribution for the initial and boundary conditions is

$$p_{\lambda>0}(\xi,t) = (p_0 \xi^{-\lambda}) \{1 - \exp[(\xi^{-\lambda}-1)/[k\tau(1-\lambda)] - t/\tau]\} u\{t - (\xi^{-\lambda}-1)/[k(1-\lambda)]\} \quad (5-2.7)$$

If $\lambda = 1$, Eq. 5-2.7 can be simplified as

$$p_{\lambda=1}(\xi,t) = (p_0 / \xi) [1 - \exp(-t/\tau) \xi^{1/k\tau}] u[t - \ln(\xi)/k] \quad (5-2.8)$$

The model parameters are k , τ , and λ , where the parameter k determines the number of jobs, τ controls the intensity of job addition with unit waiting-time, and λ determines whether the model describes the exponential distribution ($\lambda = 0$) or the power law distribution ($\lambda > 0$).

5-3. Results and Discussions

Figure 5-3.1 shows how the model describes waiting-time distributions of human actions by random-order and priority-relevant protocols. Figure 5-3.1A exhibits the exponential waiting-time distribution, Eq. 5-2.6, graphically as straight lines on the semilog coordinates when $\lambda = 0$. As shown in Fig. 5-3.1B with log-log coordinates, the power law waiting-time distribution, Eq. 5-2.8, with $\lambda = 1$ is expressed as straight lines

with truncation of the tails. It is interesting to note that this behavior implies that even the jobs with lower priority will be executed, though they have to wait longer than higher priority jobs. Figures 5-3.1A and 5-3.1B explain that the model description mainly depends on the power λ : the model with $\lambda = 0$ and $\lambda > 0$ yields the exponential and power law waiting-time distributions, respectively. Below we will compare the model prediction with statistical data.

The model based on cluster kinetics gives quantitatively correct descriptions for both exponential and power law distributions of human activities. Adding to the model capability, the model can also be applied for describing systems with underlying networked structure. For example, jobs and waiting-time in the model correspond to nodes and number of connections in network theories.

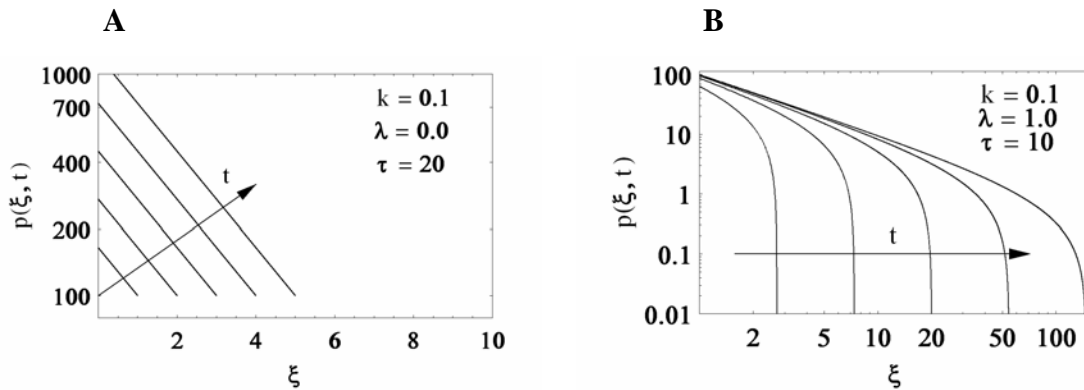


Figure 5-3.1 The waiting-time distributions by (A) the random-order protocol, $\lambda = 0$, with the initial condition, $p(\xi, t=0) = 0$, and boundary condition, $p(\xi=1, t) = p_0 e^{t/\tau}$; and (B) the priority-relevant protocol, $\lambda = 1$, with the initial condition, $p(\xi, t=0) = 0$, and boundary condition, $p(\xi=1, t) = p_0 (1 - e^{-t/\tau})$. A waiting-time distribution, $p(\xi, t)$, represents the number of jobs with the waiting-time ξ , where ξ represents waiting-time in seconds.

The model describes two kinds of human dynamics: human activities distributed by random-order and priority-relevant protocols. But even for jobs distributed by the

power law according to priority, there is always a probability that the lower priority items are executed before all higher priority items are done. A common example is a task with a deadline. To provide model details, we may need to specify such item priorities, for example, by defining a priority for the job with a deadline using a Dirac-delta function at the deadline. By including this feature, many generalizations are possible and perhaps necessary to apply the model to specific systems.

As evidence of model applicability we have compared the model predictions with the timing of printing jobs [Harder and Paczuski, 2005] and waiting-time distribution for sending and responding to emails [Barabasi, 2005; Faxon and Floyd, 1995] induced by the priority-relevant protocol. We represent these processes by a cluster distribution model, which is presented as both analytical and graphical solutions. The waiting-time distribution data and the cumulative form of model predictions are plotted in Fig. 5-3.2. Based on Eq. 5-2.8, the cumulative distribution is evaluated by integrating the distribution [Jeon and McCoy, 2005a; Jeon and McCoy, 2005b],

$$\begin{aligned}
 P^C(\xi, t) &= \int_{\xi}^{\infty} p(\xi, t) d\xi \\
 &= p_0 \left\{ \xi^{-(\lambda-1)} / (\lambda-1) + k\tau \exp[(\xi^{1-\lambda} - 1) / k\tau(1-\lambda)] - t/\tau \right\} \\
 &\quad (1 - \exp[-\xi^{1-\lambda} / k\tau(1-\lambda)]) \} u[t - (\xi^{1-\lambda} - 1) / k(1-\lambda)] \quad (5-3.1)
 \end{aligned}$$

which is plotted in Fig. 5-3.2 as model predictions for the real waiting-time distribution data. The data in Fig. 5-3.2A are the time intervals between consecutive emails sent by the same user over a three-month time interval examined by Barabasi [Barabasi, 2005]; the parameters used for the model prediction, Eq. 5-3.1, are $k = 0.01$, $\lambda = 1.94$ and $\tau = 0.1$. The time intervals taken by the user to reply to the received emails [Barabasi, 2005] are plotted as data points in Fig. 5-3.2B with the line determined by parameters $k = 0.02$,

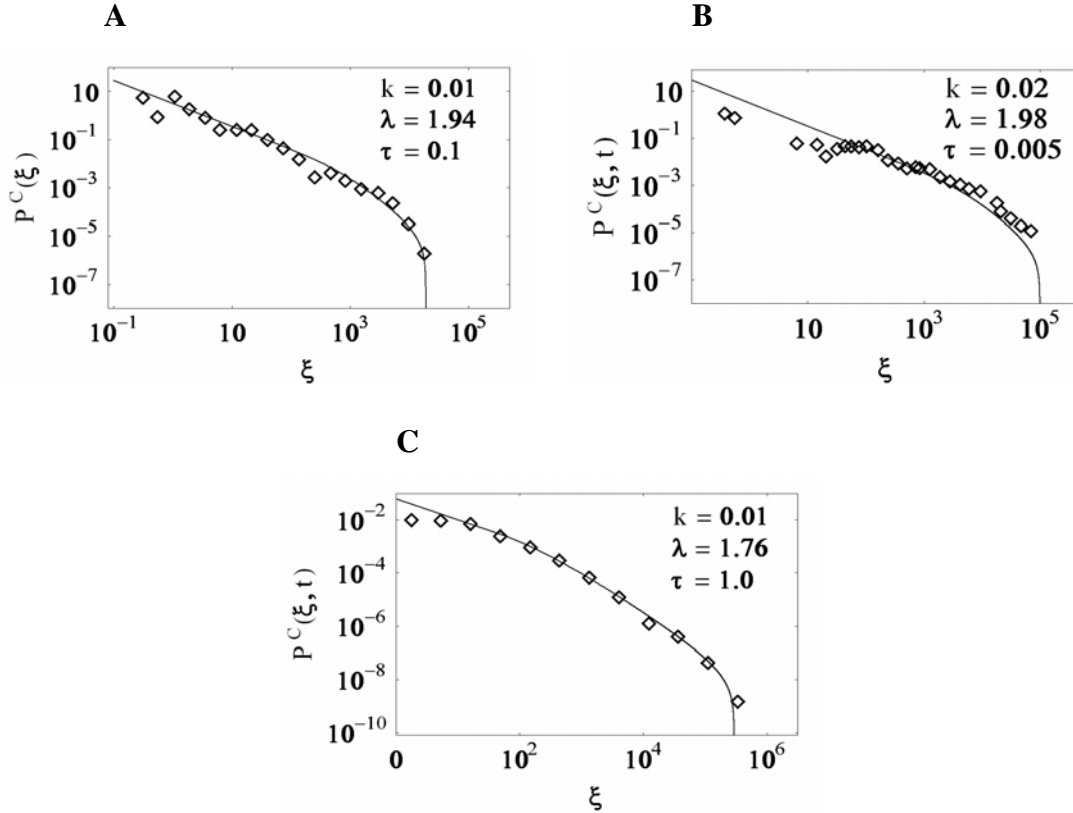


Figure 5-3.2 Comparison of the model and statistical data for email and printing server systems [Barabasi, 2005] on log-log coordinates. A cumulative distribution, $P^C(\xi, t)$, represents number of jobs with the waiting-time longer than given value ξ , and ξ represents waiting-time in seconds. The lines are the model predictions, Eq. 5-3.1, and symbols indicate statistical data collected by Barabasi [Barabasi, 2005] for (A) and (B) and Harder and Paczuski [Harder and Paczuski, 2005] for (C).

$\lambda = 1.98$, and $\tau = 0.005$ in Eq. 5-3.1. Figure 5-3.2C is the comparison of the model prediction with the plot of the inter-arrival times between subsequent requests submitted by users to the printer ‘Chrome’ in 2003 [Harder and Paczuski, 2005]. The data points are obtained by a binning operation (count the number of time differences in growing bins and normalize the count by the bin size) and the parameters for the prediction are $k = 0.01$, $\lambda = 1.76$ and $\tau = 1$ in Eq. 5-3.1. According to the dominating (first) term in Eq. 5-3.1, the cumulative distributions of Fig. 5-3.2 are proportional to $\xi^{-(\lambda-1)}$, and therefore the slopes of the model predictions are -0.94 (A), -0.98 (B), and -0.76 (C) respectively. As

shown in Fig. 5-3.2, the model predictions, Eq. 5-3.1, expressed as solid lines describe reasonably well the waiting-time distribution of email exchange systems by the cumulative distribution derived from the power law distribution.

5-4. Conclusion

Let us discuss some of the implications of the value of the parameters. As explained, a larger λ yielding a steeper slope in the distribution implies a heavier tailed distribution with the existence of jobs with very long waiting-time, and is therefore subject to larger deviations from average than those with lighter tails. A smaller τ (job addition intensity controlling parameter) implies a smaller effect of terms other than $\xi^{-(\lambda-1)}$ in Eq. 5-3.1, and the distribution becomes straighter. As τ increases to extremely large values, truncation behavior becomes minimized and therefore the model may not be able to describe truncation behavior. In addition, the system parameters also show the characteristics of a complex system. The parameters, k , λ , and τ , are interdependent, and thus, have to be adjusted together for the systems: for instance, the parameter k , which determines the number of jobs, power λ , and job addition intensity τ function together to determine the shape of the distribution and the longest waiting-time at cutoff. These results have important implications for understanding and designing complex systems, particularly for the systems of human-initiated behaviors. Typical examples requiring such understanding are the design of phone exchange and printing servers, or web and email servers, which require knowledge of the maximum possible number using the service concurrently.

We have studied the dynamics of collective human activities and found that it develops power law structures similar to those appearing in many nonlinear dynamic

systems. As we have shown, the model from cluster kinetics can describe waiting-time distributions of human actions driven both by the random-order and the priority-relevant protocols. We also have demonstrated that the waiting-times of jobs to the machine are not simply exponential or pure power law tails, but truncated power laws. As mentioned, previous models to describe human activity patterns use uniform inter-event statistics, however, the current model uses heterogeneous statistics of individual actions. The graphs in this paper are generated by the analytical solutions we derived based on the first order partial differential equation, Eq. 5-2.2. Because of the non-Poisson nature of human dynamics, most human activities are distributed as power laws and the model faithfully illustrates such distributions. Understanding the nature of human actions is essential, for instance, in designing web pages, email server systems, and hubs for Internet connection, constructing an optimum broadcasting station for phone connections, and even establishing strategies for stock market investigation.

CHAPTER 6.

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6-1. Summary

In the dissertation, we have introduced several complex systems in chapters 2 to 5: network evolutions, power law distributed systems, non-linearly growing accelerating networks, and distribution of human dynamics. We initiated the dissertation by presenting a network evolution model to show how structure of complex system is changing over time. Based on the nodal linkage distribution we define the distribution of nodes as a function of time and number of links, and provide a model based on the population balance equation yielding both exponential and power law confirmations. By presenting asymptotic expressions of the model, we show how they evolve in time in the form of analytical solutions as well as graphical solutions.

Because of ubiquity of power laws in science and social (natural and manmade) systems with cluster structures, we next introduced power law distributions to emphasize the growing importance of self-organized structures in complex system studies. To understand power law distributed systems, we constructed a model assuming that the power law distributions in cluster-growth systems are governed by the population balance equation. The model presents asymptotic power law behaviors at long time by going beyond the initial transient behavior, and we show that the population balance model governing the power law distribution realistically describes cluster-growth systems.

Based on distribution kinetics, we next examined non-linearly growing accelerating networks. We show that how continuous distributions describe network

dynamics and structures in a framework for the dynamics of linear and nonlinear growing networks. The model with size dependent and independent rate coefficients provides descriptions on the effect of preferential attachment according to the power λ . Nonlinear network size dependency of the number of connections describes the evolution of such power law networks. The moment technique is applied to describe the network properties such as the total number of nodes and links and the degree distribution. Time and network size dependency of the model are also examined by moment solutions based on the initial and boundary conditions. The model well describes the time evolution of the number of nodes and links and the degree distribution of nonlinearly growing accelerating networks.

As a part of complex system studies, we have studied human dynamics, and found that the waiting time distribution of human activities also develops power law structures, because of the heterogeneity and non-Poisson nature of human dynamics. We have demonstrated that statistics of human activity patterns are not homogeneous rather heterogeneous. The model solution, a truncated power law derived based on the first order partial differential equation, well illustrates the waiting-time distribution of human activities driven by the priority-relevant protocols.

6-2. Conclusions

The definitions of the complexity of systems are based on the theories of information and computation. We discuss relevant concepts, structures, and several example systems that may be used to understand complexity and complex systems. As scientists, whenever we encounter something new, our very first objective is to understand it. This understanding enables us to use, modify, control, or appreciate it. We achieve understanding in a number of ways, through classification, description, and

ultimately through the ability to predict behavior. In this regard, complexity is a measure of the inherent difficulty to achieve the desired understanding, and we shortly characterize the complexity of a system as the amount of information necessary to describe it.

For dynamic systems, the description of complexity should include the changes of the system over time. The complexity in the current models is expressed as analytical and graphical forms of model solutions, and since we are primarily dealing with dynamic systems, we provide the time evolution of such solutions in time.

Our objectives are to understand the complexity of systems composed of physical entities such as nodes-links and clusters-connections, to develop models to describe complex system structures, and to predict such complex system changes over time. In our current work, models are built based on the distribution kinetics approach in the context of population balance technique, which is a powerful method to describe complex systems and their common network structure. Following a traditional chemical engineering approach, we first define fundamental relationships and explain general principles. We next hypothesize the systems can be described by the distribution of quantities, make a theoretical bridge between system properties and model parameters, and construct a model based on the hypothesis and theories. Then, we solve model equations by standard methods including moment technique providing solutions, which describe status and changes of system properties in both analytical and graphical ways, and finally prove that the model can actually predict the system changes over time. By this method, we can obtain solutions for numerous science and engineering interests, and these solutions clearly illustrate the effects of parameters that govern the changes of model solutions over time.

The overall conclusion is that the distribution dynamics approach based on the population balance method faithfully describes several complex systems and the time evolution of their network structures. Further investigations are needed for more practical and specific complex systems as well as for other systems in engineering interests.

6-3. Model Applications and Limitations

There is a drive to extend the capability of population dynamics to simulate diverse complex systems, including situations in which network structure growth or particle formation and growth processes are important. These processes require efficient population balance modeling algorithms including moment methods to describe nonlinear processes. For example of the cluster distributed systems, well-defined distribution dynamics such as the method of moments with interpolative closure and the quadrature method of moments are already applied to the problems of the simultaneous cluster aggregation and de-aggregation. These population dynamics models have been applied to the problems of multicomponent coagulation, collision and coalescence of particles, and step, chain, and cross-linking polymerization. For instance to design and build a process for Nitroxide mediated polymerization (NMP), we need to model the process to obtain information such as monomer and polymer concentrations, molecular weight distributions, variance, polydispersity of products, optimum temperature and pressure, etc. To obtain such information, we need to consider many different reaction mechanisms and factors: initiation process, reversible addition and fragmentation chain transfer (RAFT) radical mechanism, termination process, temperature, pressure, volume, activation energies of reactions, etc. We initiate the modeling by assuming that the process is governing by the distribution kinetics, and write population balance equations (PBEs) to describe steps ranging from initiation, radical propagation, to termination: an initiation

step of radical reaction by adding heat into Nitroxide monomer, a reversible addition and fragmentation chain transfer (RAFT) step, and a termination step. We next propose a complex form of rate coefficients for each step, accounting for the temperature and pressure effects, e.g., $k_g = k_{g0} \exp(-E_g/k_B T) \exp(-PV_g/k_B T)$ where k_{g0} is prefactor, E_g is activation energy, k_B is the Boltzmann's constant, and T , P , and V_g is temperature, pressure, and activation volume, respectively. In the rate coefficient, the first exponential accounts for the temperature effect by using the Arrhenius formula and the second exponential displays the pressure effect by using activation volume. We next solve the PBEs simultaneously using the method of moment, and thus obtain general moment equations for monomer and polymer. Based on the moment solutions, we can obtain essential information on the process: concentration of molecules (0^{th} moment), weight of molecules (1^{st} moment), number average molecular weight (average moment, a ratio of 1^{st} and 0^{th} moments), weight average molecular weight (ratio of 2^{nd} and 1^{st} moments), and variance and polydispersity index. In addition, temperature and pressure effects will also be explained by two exponential terms in the rate coefficient we proposed. We, therefore, are able to design the Nitroxide mediated polymerization process based on the information obtained from the modeling.

In the thesis, we demonstrated how population balance modeling can be used in description and prediction of linear and nonlinear complex systems as they appear in network dynamics. The model developed here mainly describes the dynamics of growth of complex networks, and agrees with statistical data of many real world complex systems. However, the model we proposed in the thesis is not capable of describing the disintegration of the power law structures, since the dissociation process does not follow general breakage kernels, and is thus inherently difficult to describe when intentional

attacks are imposed on hubs of power law structures. In addition, the model developed here is not capable of capturing some real system behaviors; initial power shifting behavior of power law distributions and truncation-tail into zero behavior at the maximum value of the distributions. This is due to the size effects as there are no entities in real life which correspond to nodes with an infinite number of links. For instance, no company exists with infinite amount of income in the firm size distribution. To capture these real system behaviors, different methodologies or other factors relevant to those behaviors should be considered; numerical calculation of the partial differential equation including higher-order derivative terms, introducing new parameters or modification of current parameters accounting for power shifting and truncated-tail behaviors.

6-4. Recommendations

Introduced complex system studies are only part of complex system disciplines, and there are many stems in which complex system researches are extending. Among complex system researches, the modeling of disintegration of structures such as network breakdown, particle fragmentation, cluster de-aggregation, and polymer degradation is of current interest, and much progress has been made. Describing network evolution, evaluating their statistical properties, identifying self-organized and self-optimized mechanisms, and understanding real world complex systems will be crucial for all branches of engineering and science. In order to apply our understanding of complexity of mathematical constructs to physical and social systems, we should first develop a fundamental understanding of system representations.

Network disintegration is a quite interesting subject with many applications [Albert et al., 2000; Dorogovtsev and Mendes, 2001]. Robustness of power law networks is based on their structural characteristics. Unlike random networks, power law networks

are stable with more than 80% loss of nodes, unless we have intentional attacks on hubs [Barabasi and Oltvai, 2004]. Power law networks do not have a critical threshold for breakdown against random attacks. In other words, random attack on power law networks is not an efficient way to disrupt the network. However, the selective removal of a few hubs can cause a total disintegration [Dorogovtsev and Mendes, 2003].

Based on population balance dynamics, our next task is to describe network breakdown with random and intentional attacks. As mentioned in Ch. 2-1., similar difficulties are encountered for particle fragmentation and polymer degradation, and establishing models that completely describe network breakage is difficult. However, the approach we used here can provide some vision into such network disintegration.

If the above model for network disintegration is established, the following effort to apply our understanding would be modeling of species extinction in Ecosystems. Species are networked together by a multitude of interactions. The network that allows closest relationships and interactions are power laws. Whether random or power law, the dismantling of the network by loss of nodes can be modeled by a population balance method. With the help of population balance dynamics, we will be able to analyze distribution patterns of the species that have power relationships in space and time. When we describe a system, we, however, are not generally interested in a microscopic description of the positions and moving velocities of all of species. This is indeed the reason we use only the total number of nodes and links and the degree distribution when we model networks structures. The same strategy will be applied for this application, and therefore, we will develop an understanding of complexity of species interactions that is not tied to the microscopic description, but relevant to observations at a particular length and time scale.

The distribution of a species in an ecosystem reflects its ability to persist over time. Species that are widespread and flourish over a long time period show great ecological tolerance, whereas species with less resilience persist for a shorter time period. The analysis of the distribution patterns of the species over time can provide useful information. However, studies for the distribution of a species should include the interplay of many independent factors such as: the structure of food webs, patterns in the relative abundance of species, patterns in the number of species or number of individual physical size, and observations about the commonness or rarity of organisms. Fortunately, it is the nature of the population balance dynamics to incorporate these several factors. Therefore, based on the population balance dynamics established in current dissertation, we would predict how species respond to future global change and the loss of biodiversity.

REFERENCES

- L.A. Adamic and B.A. Huberman, 2002, "Zipf's law and the Internet" *Glottometrics* **3**, 143.
- L.A. Adamic, R. M. Lukose, A. R. Puniyani and B. A. Huberman, 2001, "Search in power-law networks" *Phys. Rev. E* **64**, 046135-1.
- R. Albert and A.L. Barabasi, 2002, "Statistical mechanics of complex networks" *Rev. Mod. Phys.* **74**, 47.
- R. Albert, H. Jeong, and A.L. Barabasi, 2000, "Error and attack tolerance of complex networks" *Nature* **406**, 378.
- P.M. Alien, J.L. Denenbourg, M. Sanglier, F. Boon, A. de Palma, 1977, "Dynamic urban models" Reports to the Department of Transportation, TSC-1185.
- L.A.N. Amaral, A. Scala, M. Barthelemy, and H.E. Stanley, 2000, "Classes of small-world networks" *Proc. Natl. Acad. Sci. USA* **97**, 11149.
- L.A.N. Amaral and J.M. Ottino, 2004, "Complex systems and networks: challenges and opportunities for chemical and biological engineers" *Chem. Eng. Sci.* **59**, 1653.
- H.R. Anderson, 2003, *Fixed Broadband Wireless System Design*, Wiley, New York.
- R. L. Axtell, 2001, "Zipf distribution of U.S. firm sizes" *Science* **7**, 1818.
- P. Bak, K. Chen, and M. Creutz, 1989, "Self-organized criticality in the 'Game of Life'" *Nature* **342**, 780.
- A. Backlund, 2002, "The concept of complexity in organizations and information systems" *Kybernetes*, **31**, 30.
- A.L. Barabasi, 2005, "The Origin of bursts and heavy tails in human dynamics" *Nature* **435**, 207.
- A.L. Barabasi and R. Albert, 1999, "Emergence of Scaling in Random Networks" *Science* **286**, 509.
- A.L. Barabasi, R. Albert, and H. Jeong, 1999, "Mean-field theory for scale-free random networks" *Physica A* **272**, 173.
- A.L. Barabasi, R. Albert, H. Jeong and G. Bianconi, 2000, "Power-Law Distribution of the World Wide Web" *Science* **287**, 2115.

- A.L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, 2002, "Evolution of the social network of scientific collaborations" *Physica A* **311**, 590.
- A.L. Barabasi and Z.N. Oltvai, 2004, "Network Biology: Understanding the Cell's Functional Organization" *Nature Reviews* **5**, 101.
- P. Baran, 1964, "Introduction to Distributed Communications Networks" RM-3420-PR. (<http://www.rand.org/publications/RM/RM3420/>).
- J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, 1995, "Long-Range Dependence in Variable-Bit-Rate Video Traffic" *IEEE Trans. Commun.* **43**, 1566.
- J.M. Berger and B. Mandelbrot, 1963, "A new model for the clustering of errors on telephone circuits" *IBM J. RES. DEV.* **7**, 224.
- B. Bollobas, 1985, *Random Graphs*, Academic Press, London.
- J. Camacho, 2001, "Scaling in steady-state aggregation with injection" *Phys. Rev. E* **63**, 046112.
- J.M. Carson and J. Doyle, 1999, "Highly optimized tolerance: A mechanism for power laws in designed systems" *Phys. Rev. E* **60**, 1412.
- S. Chandrasekhar, 1943, "Stochastic Problems in Physics and Astronomy" *Revs. Mod. Phys.* **15**, 1.
- J.W. Cohen, 1969, *The Single Server Queue*, North Holland, Amsterdam.
- D.R. Cox and H.D. Miller, 1965, *The Theory of Stochastic Processes*, Wiley, New York, P. 161.
- J. Crank, 1975, 2nd Ed., *The Mathematics of Diffusion*, Oxford University Press, New York.
- L.J. Croft, M.J. Lercher, M.J. Gagen, and J.S. Mattick, 2003, "Is prokaryotic complexity limited by accelerated growth in regulatory overhead?" <http://arxiv.org/abs/q-bio.mn/0311021>.
- D. Das, M. Barma, and S.N. Majumdar, 2001, "Fluctuation-Dominated Phase Ordering Driven by Stochastically Evolving Surfaces" *Phys. Rev. E* **64**, 046126.
- R. D'Hulst and G.J. Rodgers, 2001, "Models for the size distribution of businesses in a price driven market" *Eur. Phys. J. B* **21**, 447.
- S.N. Dorogovtsev and J.F.F. Mendes, 2000, "Evolution of networks with aging of sites" *Phys. Rev. E* **62**, 1842.

S.N. Dorogovtsev and J.F.F. Mendes, 2001a, "Comment on "Breakdown of the Internet under Intentional Attack" *Phys. Rev. Lett.* **87**, 219801-1

S.N. Dorogovtsev and J.F.F. Mendes, 2001b, "Effect of the accelerating growth of communications networks on their structure" *Phys. Rev. E* **63**, 025101.

S.N. Dorogovtsev and J.F.F. Mendes, 2001c, "Heterogeneous condensation in dense media" *Phys. Rev. E* **63**, 056125.

S.N. Dorogovtsev and J.F.F. Mendes, 2002, "Evolution of networks" *Adv. Phys.* **51**, 1079.

S.N. Dorogovtsev and J.F.F. Mendes, 2003, *Evolution of Networks*, Oxford.

N.A. Dotson, R. Galvan, R.L. Laurence, and M. Tirrell, 1996, *Polymerization Process Modeling*, New York, VCH Publishers Inc.

D. Duffy, A. McIntosh, M. Rosenstein, and W. Willinger, 1994, "Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks" *IEEE J. Sel. Areas Commun.* **12**, 544.

V.J. Easton and J.H. McColl, 1997, "Random variables and probability distributions" (http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html)

J.-P. Eckmann, E. Moses, and D. Sergi, 2004, "Entropy of dialogues creates coherent structures in e-mail traffic" *Proc. Natl. Acad. Sci. USA* **101**, 14333.

V. M. Equiluz and M.G. Zimmerman, 2000, "Transmission of Information and Herd Behavior: An Application to Financial Markets" *Phys. Rev. Lett.* **85**, 5659.

P. Erdos and A. Renyi, 1960, "On the evolution of random graphs" *Publ. Math Inst. Hung. Acad. Sci.* **5**, 17.

V. Faxon and S. Floyd, 1995, "Wide-Area Traffic: The Failure of Poisson Modeling" *IEEE/ACM Trans. Netw.* **3**, 226.

D. Feitelson, 2000, "Sensitivity of Parallel Job Scheduling to Fat-Tailed Distributions" Technical Report 2000-44, Inst. Computer Science, The Hebrew University of Jerusalem (<http://www.cs.huji.ac.il/feit/pub.html>).

W. Feller, 1957, *An Introduction to Probability Theory and Its Applications*, Vol I, 2nd Ed., Wiley, New York.

M.J. Gagen and J.S. Mattick, 2005, "Inherent size constraints on prokaryote gene networks due to "accelerating" growth" *Theory in Bioscience* **123**, 381.

M. Garrett and W. Willinger, 1994, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic" *Proc. ACM SIGCOMM*, 269.

J. Gleick, 1987, *Chaos: Making a New Science*, Penguin, New York.

N. Goldenfeld, 1992, *Lectures on Phase Transitions and the Renormalization Group*, Perseus Books, Reading, Mass.

J.H. Greene, 1997, *Production and Inventory Control Handbook*, McGraw-Hill, New York.

S.D. Gribble, G.S. Manku, D.S. Roselli, E. A. Brewer, T.J. Gibson, and E.L. Miller, 1998, "Self-Similarity in File Systems" *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and Modeling of Computer Systems*, Madison, Wisconsin, US, June 1998, pp. 141-150.

J.L. Guillaume and M. Latapy, 2004, "Bipartite Graphs as Models of Complex Networks" *Int. Proc. Lett.* **90**, 215.

F.A. Haight, 1967, *Handbook of the Poisson Distribution*, Wiley, New York.

M. Harchol-Balter, 1999, "The Effect of Heavy-Tailed Job Size Distributions on Computer System Design" *Proceedings of ASA-IMS Conference on Applications of Heavy Tailed Distributions in Economics*.

U. Harder and M. Paczuski, 2005, "Correlated dynamics in human printing behavior" *Physica A* **361**, 329.

E.A. Hardly and B.A. Maurer, 2001, "Spatial and temporal patterns of species diversity in montane mammal communities of western North America" *Evol. Ecol. Res.* **3**, 477.

D.M. Himmelblau, K.B. Bischoff. 1968, *Process Analysis and Simulation-Deterministic Systems*, Wiley, New York, P.192.

P. Holme and B.J Kim, 2002, "Vertex overload breakdown in evolving networks" *Phys. Rev. E* **65**, 066109.

J.H. Holland, 1999, *Emergence: From Chaos to Order*, Perseus Books Group.

B.A. Huberman and L.A. Adamic, 1999, "Growth Dynamics of the World Wide Web" *Nature (London)* **401**, 131.

B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, and R.J. Lukose, 1998, "Strong Regularities in World Wide Web Surfing" *Science* **280**, 95.

Y.-P. Jeon and B.J. McCoy, 2005a, "Distribution dynamics of evolving networks" *Comput. Chem. Eng.* **29**, 2429.

Y.-P. Jeon and B.J. McCoy, 2005b, “Evolution of power law distributions in science and society” *Phys. Rev. E* **72**, 037104.

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, 2000, “The large-scale organization of metabolic networks” *Nature (London)* **407**, 651.

S. Kauffman, 1993, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press.

Y. Kodera and B.J. McCoy, 2002, “Distribution Kinetics of Polymer Thermogravimetric Analysis: A Model for Chain-End and Random Scission” *Energy & Fuels* **16**, 119.

J. Lahererre and D. Sornette, 1998, “Stretched exponential distributions in nature and economy: “fat tails” with characteristic scales” *Eur. Phys. J. B* **2**, 525.

A. Lejeune, J. Perdang, and J. Richert, 2003, “Dynamics of many-particle fragmentation in a cellular automaton model” *Phys. Rev. E* **67**, 046214.

W.E. Leland, M. Taqqu, W. Willinger, and D. Wilson, 1994, “On the self-similar nature of Ethernet traffic” *IEEE/ACM Trans. Netw.* **2**, 1.

O. Levenspiel and W.K. Smith, 1957, “Notes on the diffusion-type model for the longitudinal mixing of fluids in flow” *Chem. Eng. Sci.* **6**, 227.

R. Lewin, 1992, *Complexity: Life at the Edge of Chaos*, Macmillian, New York.

K. Li, Y. Pan, H. Shen, G.H. Young, and S.Q. Zheng, 1998, “Lower Bounds for Dynamic Tree Embedding in Bipartite Networks” *J. Paral. and Distr. Comp.* **53**, 119.

G. Madras and B. J. McCoy, 2001, “Distribution Kinetics Theory of Ostwald Ripening” *J. Chem. Phys.* **115**, 6699.

G. Madras and B.J. McCoy, 2002a, “Numerical and Similarity Solutions for Reversible Population Balance Equations with Size-Dependent Rates” *J. Coll. Interf. Sci.* **246**, 356.

G. Madras and B.J. McCoy, 2002b, “Ostwald ripening with size-dependent rates: Similarity and power-law solutions” *J. Chem. Phys.* **117**, 8042.

G. Madras and B.J. McCoy, 2003, “Ostwald ripening in two dimensions: Time dependence of size distributions for thin-film islands” *Phys. Chem. Chem. Phys.* **5**, 5459.

O. Malcai, O. Biham, and S. Solomon, 1999, “Power-law distributions and Levy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements” *Phys. Rev. E* **60**, 1299.

- B. Mandelbrot, 1963, "New Methods in Statistical Economics" *J. Polit. Economy* **71**, 421.
- M. Marsili and Yi-Cheng Zhang, 1998, "Interacting Individuals Leading to Zipf's Law" *Phys. Rev. Lett.* **80**, 2741.
- J.S. Mattick and M.J. Gagen, 2005, "Accelerating Networks" *Science* **307**, 856.
- R.M. May, 1974, *Stability and Complexity in Model Ecosystems*, 2nd Ed., Princeton.
- R.M. May, 1988, "How many species are there on Earth" *Science* **241**, 1441.
- B.J. McCoy, 2000, "Vapor Nucleation and Droplet Growth: Cluster Distribution Kinetics for Open and Closed Systems" *J. Coll. Int. Sci.* **228**, 64.
- B. J. McCoy, 2001, "Distribution Kinetics Modeling of Nucleation, Growth, and Aggregation Processes" *Ind. Eng. Chem. Res.* **40**, 5147.
- B.J. McCoy, 2002, "Cluster kinetics for glassforming materials" *J. Phys. Chem. Solids* **63**, 1967.
- B.J. McCoy and G. Madras, 2001, "Discrete and Continuous Kinetic Models for Polymerization and Depolymerization" *Chem. Eng. Sci.* **56**, 2831.
- Y. Moreno, R. Pastor-Satorras, A. Vazquez, and A. Vespignani, 2003, "A Critical load and congestion instabilities in scale-free networks" *Europhys. Lett.* **62**, 292298.
- M.E.J. Newman, 2000, "Applied mathematics: The power of design" *Nature* **405**, 412.
- M.E.J. Newman, 2001, "The structure of scientific collaboration networks" *Proc. Natl. Acad. Sci. USA* **98**, 404.
- M.E.J. Newman, D.J. Watts, and S.H. Strogatz, 2002, "Random graph models of social networks" *Proc. Natl. Acad. Sci. USA* **99**, 2566.
- J.M. Ottino, 2003, "Complex Systems" *AIChE J.* **49**, 292.
- J.M. Ottino, 2004, "Engineering complex systems" *Nature* **427**, 399.
- J. Park and M.E.J. Newman, 2005, "Solution for the properties of a clustered network" *Phys. Rev. E* **72**, 026136.
- R. Pastor-Satorras and A. Vespignani, 2004, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, New York.
- V. Paxson and S. Floyd, 1996, "Wide area traffic: the failure of Poisson modeling" *IEEE/ACM Trans. Netw.* **2**, 226.

- A.D. Randolph and M.A. Larson, 1986, *Theory of Particulate Processes*, 2nd Ed., Academic Press, New York.
- S. Render, 1998, "How popular is your paper? An empirical study of the citation distribution" *Eur. Phys. J. B* **4**, 131.
- P. Reynolds, 2003, *Call Center Staffing*, The Call Center School Press, Lebanon, Tennessee.
- B. Roehner and G. Valent, 1982, "Solving the birth and death process with quadratic asymptotically symmetric transition rates" *SIAM J. Appl. Math.* **42**, 1020.
- T. Sato, 2004, "The earth simulator: Roles and impacts" *Nuclear Physics B (Proc. Suppl.)* **129&130**, 102.
- M.R. Schroeder, 1991, *Fractals, Chaos, Power Laws*, W.H. Freeman and Company, New York, P.103.
- F. Schweitzer and J. Steinbrink, 1997, *Urban cluster growth: Analysis and computer simulations of urban aggregations*, in: *Self-organization of complex structures*, Gordon and Breach science publishers, Amsterdam.
- N. Shiode and M. Batty, 2000, "Power Law Distributions in Real and Virtual worlds" *CSCA* working paper 19.
- H.A. Simon, 1996, *The Sciences of the Artificial*, 3rd Ed., MIT press, Cambridge.
- K. Sneppen, M. Rosvall, A. Trusina, and P. Minnhagen, 2004, "A simple model for self organization of bipartite networks" (<http://arxiv.org/abs/nlin.AO/0403005>).
- W.J. Sterling and B.J. McCoy, 2001, "Distribution Kinetics of Thermolytic Macromolecular Reactions" *AIChE J.* **47**, 2289.
- S.H. Strogatz, 2001, "Exploring complex networks" *Nature* **410**, 268.
- S.H. Strogatz, 2003, *Sync: the emerging science of spontaneous order*, 1st Ed., Hyperion, New York.
- A. Thomas, R. Cannings, N.A.M. Monk, and C. Cannings, 2003, "On the structure of protein-protein interaction networks" *Biochem. Soc. Trans.* **31**, 1491.
- U.S. Bureau of the Census, *Enterprise Statistics: Company Summary* (U.S. Dept. of Commerce, Bureau of the Census, Washington, D.C., 1967-1997).
- N.G. van Kampen, 1992, *Stochastic Processes in Physics and Chemistry*, North-Holland Publishing Co.

A. Vazquez, 2000, “Knowing a network by walking on it: emergence of scaling” *eprint arXiv:cond-mat/0006132*.

A. Vazquez, 2005, “Exact Results for the Barabasi Model of Human Dynamics” *Phys. Rev. Lett.* **95**, 248701.

J. Voldman, B.B. Mandelbrot, L.W. Hoewel, J. Knight, and P. Rosenfeld, 1981, “Fractal Nature of Software-Cache Interaction” *IBM J. RES. DEV.* **27**, 164.

M.M. Waldrop, 1992, *Complexity: The Emerging Science at the Edge of Order and Chaos*, Simon&Schuster, New York.

D.J. Watts and S.H. Strogatz, 1998, “Collective dynamics of ‘small-world’ networks” *Nature* **393**, 440.

W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, 2002, “Scaling phenomena in the Internet: Critically examining criticality” *Proc. Natl. Acad. Sci. USA* **99**, 2573.

W. Willinger, M. Taqqu, and A. Erramilli, 1996, “A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks” in *Stochastic Networks: Theory and Applications*, F.P. Kelly, S. Zachary, and I. Ziedins Eds., Oxford University Press, Oxford, P. 339-366.

D.H. Zanette and S.C. Manrubia, 1997, “Role of Intermittency in Urban Development: A Model of Large-Scale City Formation” *Phys. Rev. Lett.* **79**, 523.

D. Zheng, G.J. Rodgers, and P.M. Hui, 2002, “A model for the size distribution of customer groups and businesses” *Physica A* **310**, 480.

G.K. Zipf, 1932, *Selective Studies and the Principle of Relative Frequency in Language*, Harvard University Press, Cambridge, MA.

APPENDIX: LETTERS OF PERMISSION

1. A permission letter from Elsevier for the article ‘Y.-P. Jeon and B.J. McCoy, Distribution dynamics of evolving networks, *Comput. Chem. Eng.* Vol. **29**, Issue 11-12, 2429-2437 (2005)’.



Dear Young-Pyo Jeon

We hereby grant you permission to reproduce the material detailed below at no charge **in your thesis** subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.
2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

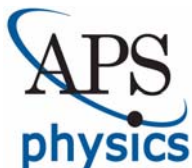
“Reprinted from Publication title, Vol number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier”.
3. Reproduction of this material is confined to the purpose for which permission is hereby given.
4. This permission is granted for non-exclusive world **English** rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form. Should you have a specific electronic project in mind please reapply for permission.
5. This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Yours sincerely



Jennifer Jones
Rights Assistant

2. A permission letter from APS physics for the article ‘Y.-P. Jeon and B.J. McCoy, Evolution of power law distributions in science and society, *Phys. Rev. E* **72**, 037104 (2005)’.



SOCIETY

AMERICAN PHYSICAL

One Physics Ellipse, College Park, MD 20740 · <http://www.aps.org>

October 2, 2006

Young-Pyo Jeon
Department of Chemical Engineering
Louisiana State University

Ref #4210

Thank you for your permission request dated September 29, 2006, we are pleased to grant you non-exclusive, non-transferable permission, English rights, but limited to **print format** only, provided you meet the criteria outlined below. Permission is for one-time use and does not include permission for future editions, additional printings, updates, any electronic forms, databases, translations, or any other matters. Permission must be sought for each additional use. This permission does not include the right to modify APS material.

Please print the required copyright credit line on the first page that the material appears: “Reprinted (abstract/excerpt/figure) with permission from [FULL REFERENCE CITATION] as follows: authors names, journal title, volume number, page number and year of publication. Copyright (YEAR) by the American Physical Society.

You must also obtain permission from at least one of the authors for each separate work, if you haven’t done so already. The author’s name and address can be found on the first page of the published Article.

Use of the APS material must not imply any endorsement by the American Physical Society.

Permission is granted for use of the following APS material only:

- Phys. Rev. E 72, 037104 (2005) by Y.-P. Jeon and B.J. McCoy

Permission is limited to the single title specified or single edition of the publication as follows:

- Ph.D. dissertation by Young-Pyo Jeon

If you have any questions, please refer to the Copyright FAQ at: <http://forms.aps.org/author/copyfaq.html> or contact me at assocpub@aps.org.

Sincerely,

A handwritten signature in black ink that reads 'Eileen LaManca'.

Eileen LaManca
Marketing Assistant

VITA

Young-Pyo Jeon was born in Jeongson, South Korea, to Jaweok Jeon and Yeounoak Kim. In February 1991, he graduated as valedictorian of his class from Jeongson High School in Jeongson, Korea. In February 1997, he graduated with departmental honors from Kangwon National University in Korea with a Bachelor of Science degree in chemical engineering. During his college years, he entered Korean Army as mandatory service of a Korean citizen in September 1992, and he was honorably discharged from the service in December 1994. In February 1999, he graduated from the Graduate School of Kangwon National University in Korea under the direction of Yong Jung Kwon with a Master of Science degree in chemical engineering. During the following four years, he was employed as a researcher in KOESF (Korean Organization of Engineering and Science Foundation) at Chunchon, Korea. In the fall 2002, he flew to the United States and began his doctoral studies under the direction of Benjamin J. McCoy and Martin A. Hjortso in the Department of Chemical Engineering at Louisiana State University. After surviving four Baton Rouge summers and hurricane seasons, he will be heading for a cooler climate as a post-doctoral fellow at Clemson University in Clemson, South Carolina.

Permanent Address: 47-42 Bongyang, Jeongson
Kangwon, 233800, South Korea