University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations | Dissertations and Theses

July 2018

Power laws in complex graphs: parsimonious generative models, similarity testing algorithms, and the origins

Shan Lu
University of Massachusetts, Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

 Part of the Other Computer Engineering Commons

Recommended Citation

Lu, Shan, "Power laws in complex graphs: parsimonious generative models, similarity testing algorithms, and the origins" (2018). *Doctoral Dissertations*. 1257.
https://scholarworks.umass.edu/dissertations_2/1257

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

POWER LAWS IN COMPLEX GRAPHS: PARSIMONIOUS GENERATIVE MODELS, SIMILARITY TESTING ALGORITHMS, AND THE ORIGINS

A Dissertation Presented

by

SHAN LU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2018

Electrical and Computer Engineering

© Copyright by Shan Lu 2018

All Rights Reserved

POWER LAWS IN COMPLEX GRAPHS: PARSIMONIOUS GENERATIVE MODELS, SIMILARITY TESTING ALGORITHMS, AND THE ORIGINS

A Dissertation Presented

by

SHAN LU

Approved as to style and content by:

Weibo Gong, Co-chair

Don Towsley, Co-chair

Lixin Gao, Member

Patrick A. Kelly, Member

Christopher V. Hollot, Department Chair
Electrical and Computer Engineering

DEDICATION

To my family!

ACKNOWLEDGMENTS

I would like to thank so many people whose names may not all be enumerated here.

First of all, I would like to express my sincere gratitude and thanks to my advisor, Dr. Weibo Gong for imparting his knowledge and expertise in my Ph.D. study. His enthusiasm and passion in scientific research inspired me deeply. His understanding and caring of my personal life also provided a warm environment for my graduate study.

I place on record, my gratitude and appreciation to my co-advisor Dr. Don Towsley. I am extremely grateful to him for his expert and valuable guidance. My great gratitude is also extended to the remaining members of my dissertation committee, Dr. Patrick A. Kelly and Dr. Lixin Gao for their helpful suggestions and comments.

I am highly indebted to Dr. Gennady Samorodnitsky from Cornell University for his effective collaborations and supports. He provided a huge amount of his precious time and effort for my projects. The completion of my dissertation could not have been possible without his help.

I also take this opportunity to thank all who have lent their helping hand in this venture. I am especially thankful to a group of people: Bo Jiang, Fabricio Murai, Kun Tu, Chang Liu, James Atwood, from Computer Networks Research Group, and my colleagues: Yan Cai, Sheng Xiao, Jieqi Kang, from Complex Systems Modeling and Control Lab, for their friendship and contributive discussions. At the same time, I would like to acknowledge my gratefulness to a group of close friends inside UMass,

Shuo Li, Xiaolin Xu, Ze He, Zhongyun Huang, Shikang Xu for their friendship and support in the past seven years.

Finally, my thanks and appreciations also go to my family. I thank my husband Jieqi Kang, who is also my colleague, for his consistent love and support. I thank my lovable son Yihan Kang for delaying my graduation for two years. In addition, I would like to thank my parents and my twin sister who are fully supportive for all my choices and decisions. I really appreciate their help during the hard time.

ABSTRACT

POWER LAWS IN COMPLEX GRAPHS: PARSIMONIOUS GENERATIVE MODELS, SIMILARITY TESTING ALGORITHMS, AND THE ORIGINS

MAY 2018

SHAN LU

B.Sc., SICHUAN UNIVERSITY, CHINA

M.Sc., UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Weibo Gong and Professor Don Towsley

This dissertation mainly discussed topics related to power law graphs, including graph similarity testing algorithms and power law generative models.

For graph similarity testing, we proposed a method based on the mathematical theory of diffusion over manifolds using random walks over graphs. We show that our method not only distinguishes between graphs with different degree distributions, but also graphs with the same degree distributions. We compare the undirected power law graphs generated by Barabási-Albert model and directed power law graphs generated by Krapivsky's model to the random graphs generated by Erdös-Rényi model. We also compare power law graphs generated by four different generative models with the same degree distribution. The comparison results show that, our

method performs better compared to the traditional features like eigenvalue spectrum and degree distributions.

To study the generative mechanism of bivariate power law data in social networks, we use Poisson Counter Driven Stochastic Differential Equation (PCSDE) models as mathematical tool. We propose three types of bivariate PCSDE models. We study the tail dependence of the models and compare the models to real data in social networks. Type 1 model with Markov on-off modulation generates tail dependence coefficient (TDC) with values either zero or one; while the Type 2 model with coupled growth has the values between zero and one. The first two types of models do not fit the real data in distribution. Type 3 model keeps the shared Poisson counter in Type 1, but uses independent Brownian motion components instead of independent Poisson counters. We show that second Type 3 model with $0 < \gamma < 1$ has fractional TDC and synthetic data fits the real data in distribution.

We study the applications of our proposed bivariate models. At first, the connection between Type 3 model to the existing network growing models is discussed. By connecting the two, our model explains why correlated bivariate power law in directed growing networks. The idea of exponential growth and random stopping can also be used to explain the existence of power law in many other natural or man-made phenomenons. We show that bivariate power law data also exists in natural images. We propose a new generative model for self-similar images based on our second Type 3 model.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Background

Power law distributions have been observed in many natural phenomena (sizes of earthquakes [44], sizes of neural avalanches clusters [76], etc) and man-made phenomena (frequency of use of words, population of cities, etc) [19, 70]. The degree distribution of many complex networks follow power law distributions, such as networks of film actors, peer-to-peer networks, etc. In [69], the authors gave basic statistics of some published networks. Power laws also appear in some critical phenomena, such as power law cluster size distribution appears at the phase transition point in percolation problems [11, 1, 70]. Natural images are also related to power law distributions, such as the the power spectra of natural images [89, 90], cluster size distribution in the k-bilevels of natural images [5, 39].

We begin with the definition of power law distribution. A random variable follows a power law distribution if its Complementary Cumulative Distribution Function (CCDF) satisfies:

$$\bar{F}_X(x) = \mathbb{P}(X > x) \sim x^{-\alpha}, \tag{1.1}$$

for some $\alpha > 0$. In (1.1), $f(x) \sim g(x)$ denotes

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = c,$$

where c is some positive constant.

Power laws are famous for the 80/20 Rule, which is also called Pareto Principle. In 1971, Pareto pointed out that 80% of the wealth in Italy belonged to 20% of the population [74]. The 80/20 Rule can also be applied to many other domains. For example, we normally spend 80% of our time to complete some hard part of a project, which takes up only 20% of the project.

As we mentioned before, complex networks are also related to power-law phenomenon. Complex networks normally relate to real-life networks, such as social networks, biological networks, etc. Complex networks are typically very large, and exhibit some features not found in simple graphs. For example, connections between the nodes in complex networks are usually random; the degree distributions follows power law.

A network with power law distribution is often called a scale-free network. In this dissertation, we focus on topics related to scale-free networks. Power law degree distributions have been observed in many real networks with the creation of World Wide Web (WWW) at the beginning of the 90s [4]. In [3], the authors presented the degree distribution exponents of several scale-free networks, such as Internet [28, 14], email networks [24], citation networks [77], etc. Power laws are also observed in financial economics [31, 32]. Social networks [25] have attracted millions of users since their birth. Normally, social networks contain users and relationship between the users. In [54, 61], power law distributions had been found in some online social networks, like Twitter, Youtube, LiveJournal etc. In [84], Ribeiro *et al.* studied the distribution of number of friends in MySpace and found that it followed a double-Pareto distribution. In [52], the authors presented several measurements of two social networks, Flickr and Yahoo!, and proposed a biased preferential attachment model to study the evolution of social networks. Social networks exhibit some features that differ from the traditional Internet, Web and citation networks. We will talk more

about the differences in the later parts. These differences motivate us to study the origin of power law data in social networks.

1.2 Related work

1.2.1 Research on graph comparison

Our first interest is to check similarity or dissimilarity of complex networks. This problem is practical and useful in data analysis.

Based on research in [49, 93], the traditional notion of graph similarity includes graph isomorphism, edit distance, and maximum common subgraph/minimum common subgraph. Two graphs are isomorphic when they have the same number of nodes and the nodes are connected in the same way. Edit distance refers to the minimum set of operations required to transform one graph to another. It is a generalization of isomorphism. Another generalization is the minimum/maximum common subgraph/supergraph, which means one graph is isomorphic to a subgraph of the other. This notion is very useful in comparing simple graphs, such as chemical structures. Checking for graph isomorphism is believed to be NP-Complete, thus may not be useful for complex graph comparison.

In [73], the author proposed several algorithms for Web graph similarity based on existing graph similarity measures, such as vertex/edge overlap (VEO), Vertex/edge vector similarity, etc. These algorithms are designed to detect anomalies in Web graphs, such as missing random vertices or connectivity changes. These algorithms need the two graphs to have similar sizes to compare and need to know the nodes mapping information of the two compared graphs. This is a not suitable for many complex networks without no detailed vertex information.

When the nodes' mapping information is unknown, algorithms to measure the similarities between nodes are needed. Based on the principle that two nodes are similar if their neighborhoods are also similar, some iterative methods were proposed.

The iterative methods need to calculate the pairwise similarity scores between graphs elements. Some examples of iterative algorithms include SimRank algorithm [41], Similarity Flooding algorithm [59], etc. Such iterative algorithms do not scale well.

Feature extraction methods extract features such as degree distribution, diameter, and eigenvalues to compare between graphs. Using these extracted features, a similarity measure is applied to compute the similarity scores between graphs, like similarity measures between probability density functions in [15]. Euclidean distance is one of the most popular measures. Feature extraction methods significantly reduce the scale of data needed for comparing and do not need the nodes' mapping information; so are more suitable for complex networks. However, some features may fail in some special cases. In Chapter 2, we will discuss some failure cases of existing features, such as the degree distribution and Laplacian spectrum.

1.2.2 Research on power law generative models

The ubiquity of power law distributions has motivated researchers to search for mechanisms to explain their origins. In [63, 31], the authors summarized several generative mechanisms for power law distributions, including preferential attachment, optimization, multiplicative models, and so forth. Generative mechanisms for double-Pareto distribution were discussed in [64] and [79]. In [79], the model was used to fit income and other size distributions. In [37], the authors argued that some of the proposed mechanisms are not robust and existing data is usually not sufficient for classifying the tail of a distribution.

Network growth models are designed to generate scale-free graphs to explain the origin of power law degree distributions in complex graphs. The famous undirected generative models like Barabási-Albert Model (B-A) model [8] was proposed by Barabási and Albert in 1999. B-A model rediscovered the world for undirected graphs. This model is simple and elegant; but the power law exponent the model

generated is always 3 no matter how you change the parameters. Some extensions and modifications of the B-A model are proposed by followers. Albert and Barabási summarized those models in Table III in [3]. Directed generative models are also proposed to generate directed power law graphs. In 1965, the famous Price' model was proposed to simulate the growing process in citation networks [22]. Many more directed models were proposed after that for more general cases, like Krapivsky's models [51, 50], Aiello's models [2], Bollobás' model [9], etc. In [69], the pros and cons of existing undirected and directed network growing models were summarized and discussed in Chapter 7.

In [78, 80], authors presented a simple mechanism by showing that power law emerges when an exponentially growing process is stopped at exponentially distributed random time. They used a Geometric Brownian Motion (GBM) as the mathematical model. In their work [81] in 2003, they applied the same mechanism in [80] to explain the occurrence of heavy-tailed distributions in gene family size distribution and the evolution networks. That means that their mechanism can be applied to explain many specific problems. In [43], the authors showed the power of this idea for generating different kinds of power law, log-normal and double-Pareto distributions with some modifications to the original model. This idea is so powerful and is not constrained to explain power laws in complex graphs. We use models in [43] as our mathematical basis in our works.

1.3 Problem outlines and main results

As has mentioned in the previous section, we focus on developing algorithms and models for problems related to complex networks. The problems include:

1. Design an algorithm to check the similarity between two complex graphs. The algorithm needs to be scale in the size of the graphs. And the features we extract

should be better representatives of the graph compared to existing features, not too loose or too tight.

2. Propose generative models for bivariate power law data. Existing power law explanations are normally for one dimensional power law phenomenon, except for some directed network growing models. The network growing models were designed to model the evolution of directed scale-free graphs. We will design bivariate models not only aiming at directed graphs but also bivariate power law data in many other fields.

The main results of our work include:

1. We designed a graph comparison algorithm based on existing heat content theory. This algorithm is computationally efficient for classification. It is robust against minor changes in graphs. Our algorithm also works well in cases where more traditional algorithms fail.

2. We propose several bivariate power law generative models based on single variate models in [81] and [43]. Our Type 3 models fit real bivariate power law data in scale-free networks. This type of models can be connected to the existing network growing models. They are also useful in explaining bivariate power law data in other fields, like natural images.

1.4 Dissertation structure

The thesis contains four chapters. In Chapter 2, we propose a graph similarity testing algorithm and apply the algorithm to classify graphs with different degree distributions and graphs with the same degree distribution. In Chapters 3 and 4, we propose three different types of bivariate Poisson Counter Driven Stochastic Differential Equation (PCSDE) models to generate correlated bivariate power law data. The synthetic data generated by our models are compared to real data in social network.

In Chapter 5, we discuss possible applications of the model proposed in Chapter 4. We draw conclusions and discuss future works in Chapter 6.

CHAPTER 2

COMPLEX NETWORK COMPARISON USING RANDOM WALKS

2.1 Introduction

2.1.1 Background

Graph similarity checking among complex networks is a challenging task since graph sizes increase extremely fast in diverse areas. Traditional methods for simple graphs may not apply to complex networks. The complexity and large size of graphs require scalability in the algorithms. How do we know whether two graphs are structurally similar if the two graphs are large? In Figure 2.1, we show three graphs from [40]. Can one tell how similar or dissimilar these three graphs are? Does one believes the three graphs actually have the same degree distribution?

As mentioned in Section 1.2.1, isomorphism is effective in comparing simple graphs and iterative algorithms are not scalable. Feature extraction is more suitable to compare large scale complex graphs. However, some previously proposed features may

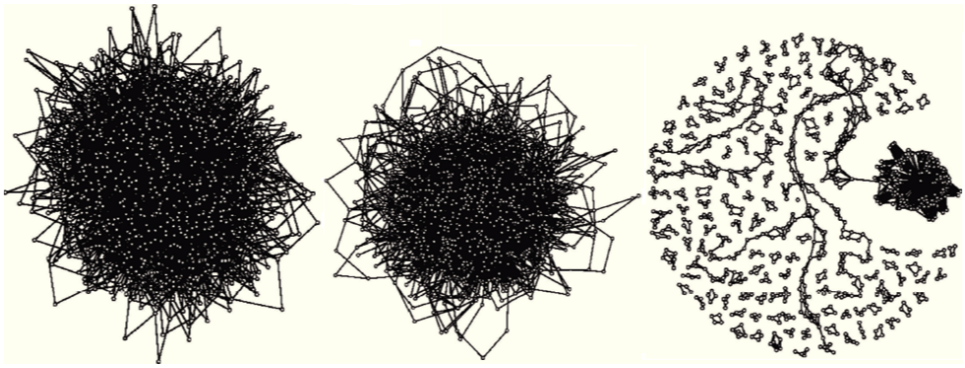


Figure 2.1. How to check similarity/dissimilarity of complex networks?

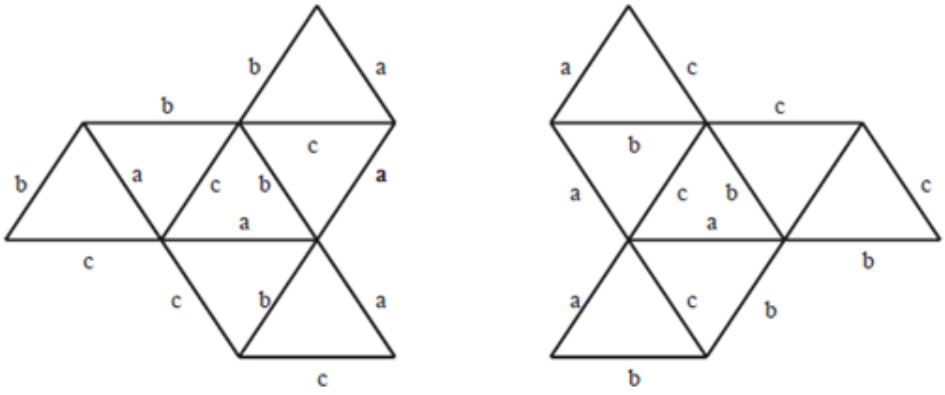


Figure 2.2. An example of isospectral planar graphs in [58]

not reflect the network connectivity structure very well, such as degree distribution and eigenvalues. For example, In [40], the authors analyzed the structural properties of graphs with the same degree distribution and found that different networks with the same degree distribution can have distinct structural properties (as shown in Figure 2.1). In [58], the authors gave an example where two iso-spectral non-isometric planar graphs could be distinguished by the heat content, despite the fact they shared the same set of eigenvalues (as shown in Figure 2.2). In our work, we apply the heat content method in [58] to graph similarity checking.

2.1.2 Mathematical tool

The asymptotic behavior of the heat content has been used as a tool to understand the geometry of a manifold domain [91, 75], or the connectivity structure of a graph [57, 58]. The **Heat equation** describes the distribution of heat in a given region over time,

$$\partial_t u = -\Delta u,$$

where Δ is the Laplacian, a differential operator. $\Delta = \partial_x^2 + \partial_y^2$ in $\mathbb{R}^2$. The heat equation is normally used to study the geometry of manifolds. Likewise, the heat equation can be defined on graphs too. Accordingly, we replace the continuous Laplacian opera-

tor by the normalized graph Laplacian. We will introduce the formal definition later. Similarly, heat equation on graphs can also be used to study the graph structure. **Heat content**, as the solution of the heat equation associated with the Laplacian operator, summarizes the heat diffusion in the manifold domain or on the graph as a function of time,

$$Q(t) = \int_D u(\mathbf{x}, t) d\mathbf{x}.$$

Heat content measures the heat stored in a given domain. One property is that its asymptotic behavior as $t \to 0$ separates the heat content curves of different structures. This enables one to develop fast algorithms for comparing complex graphs. In [35, 36] it was pointed out that Monte-Carlo simulations of diffusions on graphs are effective in testing the similarity of complex graphs and that such simulations provide plausible mechanisms for many brain activities.

Using random walks to compare graphs is not a new idea. In [65], graphs were compared based on their mixing times. Mixing time is the time needed for a random walk on the graph to approach its stationary distribution. If the expected variation distance between the distribution after random walk and the stationary distribution is too small, this method may expect long walk distance for the computation. Our method, on the other hand, focuses on the first few random walk steps to compare. And the random walk step is fixed for all graphs. In [29], the author developed an efficient algorithm named RWT using random walk to check the structural similarities of sub-graphs between knowledge networks. Our method is similar to that in [29], but based on more rigorous mathematical background.

Our algorithm exhibits the following features. First, our method summarizes graph structure into a single time function so as to facilitate similarity testing. Second, the behavior of this function around time $t = 0$ forms the basis for the comparison, so that we can greatly reduce the computation time. Third, we use a lazy random walk to estimate the heat content function, thereby avoiding the need to compute eigenvalues

and eigenvectors of the graph Laplacian while retaining the spectral information. Fourth, like most of the feature extraction methods, we do not need to know the mapping information between nodes of two graphs to compare. Finally we note that our method is robust to minor changes in large graphs according to the interlacing theorem [13]. With these features, our algorithm is capable of handling very large complex networks.

The whole chapter is organized as follows. In Section 2.2, we give notations and review the concept of heat equation and heat content for graphs. In Section 2.3, we use the lazy random walk simulation method to estimate the heat content. In Section 2.4, the graph generative models used in experiment part are introduced. Experiment settings and results are presented in Section 2.5. Section 2.6 summarizes the main results and discusses future work of this part.

2.2 Heat equation and heat content for graphs

In this part, we introduce the definition of heat equation and heat content for graphs. We begin with basic notations for graphs.

2.2.1 Notations

Let $G = (V, E)$ denote a graph with vertex set V and edge set $E \subseteq V \times V$ with adjacency matrix

$$A = [a_{uv}],$$

where $a_{uv} = 1$ if there is an edge from u to v; otherwise $a_{uv} = 0$.

The out-degree matrix

$$D = diag[d_u],$$

with $d_u = \sum_v a_{uv}$.

The graph Laplacian L is defined as follows

$$L = D - A,$$

and the normalized Laplacian $\mathcal{L}$ [17] is defined as

$$\mathcal{L} = D^{-1/2} L D^{-1/2}.$$

We can also write $\mathcal{L} = (\mathcal{L}_{uv})$, with

$$\mathcal{L}_{uv} = \begin{cases} 1 & \text{if } u = v, \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

The random walk Laplacian L_r differs from the normalized Laplacian $\mathcal{L}$. L_r is related to the random walk on graph,

$$L_r = D^{-1} L.$$

Thus,

$$L_r = D^{-1/2} \mathcal{L} D^{1/2}.$$

Without loss of generality, we assume that the Laplacian L is diagonalizable and hence $\mathcal{L}$ is diagonalizable. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ the eigenvalues of $\mathcal{L}$ and $\phi_i, i = 1, \cdots, n$ the corresponding eigenvectors. With

$$\Lambda = diag[\lambda_i]$$

and

$$\Phi = [\phi_1, \cdots, \phi_n],$$

$\mathcal{L}$ is diagonalized to be

$$\mathcal{L} = \Phi\Lambda\Psi,$$

where $\Psi = \Phi^{-1} = [\psi_1; \psi_2; \cdots; \psi_n]$. Meanwhile

$$L_r = (D^{-1/2}\Phi)\Lambda(D^{-1/2}\Phi)^{-1}. \tag{2.1}$$

L_r and $\mathcal{L}$ share the same set of eigenvalues. $\mathcal{L}$ is the normalized graph Laplacian used in the heat equation on a graph. We use the relationship between $\mathcal{L}$ and L_r to develop a random walk simulation method in Section 2.3.

2.2.2 Heat equation and heat content

Define the heat equation on graph associated with the normalized graph Laplacian as follows

$$\frac{\partial H_t}{\partial t} = -\mathcal{L}H_t, \tag{2.2}$$

with initial condition $H_0(u, u) = 1$. $H_t(u, v)$ measures the amount of heat that initiates from vertex u and ends up at vertex v at time t. In the heat equation (2.2), we consider using the normalized Laplacian instead of the regular one for the following reasons.

1. As in [17], the spectra of the normalized Laplacian relate well to other graph invariants; while, the other two definitions adjacency matrix and standard Laplacian fail to do. For example, for bipartite graphs, complete graphs and regular graphs, spectra of normalized laplacian satisfy certain rules; for random graphs, the spectra satisfy a semi-circle law.

2. The spectra of normalized Laplacian is also consistent with the eigenvalues in spectral geometry and in stochastic processes [17].

3. Eigenvalues of normalized Laplacian are in "normalized" forms with $0 \leq \lambda \leq 2$ and the eigenvalues are bounded according the interlacing theorem for the normalized Laplacian in [13].

Vertex set V is partitioned into two subsets, the set of all boundary nodes B and the set of all interior nodes B^c; thus $V = B \cup B^c$. Heat flows are from $B^c \to B$ but not $B \to B^c$. Let $N = |V|$ and $n = |B^c|$. Label the interior vertices as the first n nodes; the normalized Laplacian $\mathcal{L}$ can be partitioned into four parts. The part related to the interior domain is denoted as $\mathcal{L}_{B^c}$,

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{B^c} & \mathcal{L}_{B,B^c} \\ \mathcal{L}_{B^c,B} & \mathcal{L}_B \end{bmatrix}.$$

Since we are interested in the heat remaining in the interior domain, define interior domain as an $n \times n$ matrix h_t with $h_t(u,v) = H_t(u,v)$ (for $u, v \in B^c$). The heat equation in the interior domain is

$$\begin{cases} \frac{\partial h_t}{\partial t} = -\mathcal{L}_{B^c} h_t, \\ h_0(u,u) = 1. \end{cases} \tag{2.3}$$

The solution to the heat equation is $h_t = e^{-\mathcal{L}_{B^c} t}$. Heat content $Q(t)$ is defined as

$$Q(t) = \mathbf{1}^T h_t \mathbf{1} \tag{2.4}$$

For convenience, we slightly abuse notation and use Λ and Φ as the eigenvalue matrix and eigenvector matrix of $\mathcal{L}_{B^c}$. After performing the eigen-decomposition, we have $\mathcal{L}_{B^c} = \sum_i \lambda_i \phi_i \psi_i$. Letting $\big(\alpha_i = (\mathbf{1}^T \phi_i)(\psi_i \mathbf{1})\big)$ yields

$$Q(t) = \sum_{i=1}^{n} \alpha_i e^{-\lambda_i t}. \tag{2.5}$$

For symmetric graphs, $\Phi^{-1} = \Phi'$, which yields $\psi_i = \phi_i,\ i = 1, \ldots, n$. Thus, we have the heat content for symmetric graphs:

$$Q(t) = \sum_{i=1}^{n} e^{-\lambda_i t} \sum_{uv} \phi_i(u)\phi_i(v). \tag{2.6}$$

2.2.3 Heat content for asymmetric graph

The eigenvalues and eigenvectors of the normalized graph Laplacian $\mathcal{L}$ of the asymmetric graph can be complex valued, and exist as complex conjugates. For any complex conjugates pairs of eigenvalues $\lambda = a + bi$ and $\overline{\lambda} = a - bi$ with corresponding eigenvectors ϕ, ψ, $\overline{\phi}$ and $\overline{\psi}$, suppose that

$$w = \sum_{uv} \phi(u)\psi(v) = \alpha + \beta i.$$

Then we have

$$\begin{aligned} \overline{w} = \alpha - \beta i &= \overline{\sum_{uv} \phi(u)\psi(v)}, \\ &= \overline{(\sum \phi)(\sum \psi)}, \\ &= (\sum \overline{\phi})(\sum \overline{\psi}), \\ &= \sum_{uv} \overline{\phi}(u)\overline{\psi}(v). \end{aligned} \tag{2.7}$$

The summation of the heat contents of the pair of complex conjugates becomes

$$e^{-\lambda t} w + e^{-\overline{\lambda} t}\overline{w} = e^{-(a+bi)t}(\alpha + \beta i) + e^{-(a-bi)t}(\alpha - \beta i) = 2e^{-at}(\alpha \cos(bt)) + \beta \sin(bt)). \tag{2.8}$$

Consequently the summation is still a real-valued function. Therefore, the total summation is still real valued and can be written as

$$Q(t) = \sum_{i=1}^{n} e^{-a_i t}(\sqrt{\alpha_i^2 + \beta_i^2} \sin(b_i t + \arctan \frac{\beta_i}{\alpha_i})). \tag{2.9}$$

Compared to the heat content for symmetric graphs in Equation (2.6), the new oscillatory heat content (OHC) is no longer a sum of purely exponentially decaying functions. It becomes an oscillatory function that contains components of different frequencies, amplitudes and phases. In the total heat content, there is one component $Q_1(t)$ corresponding to λ_1, the smallest eigenvalue, which is a real number. For large t, this low frequency component dominates the total heat content curve. We exhibit this feature through the following example.

Consider a fully connected graph $G = (V, E)$ with $|V| = 30$ with different weights (as shown in Figure 2.3). The edge set $E = \{(u, v), u, v = 1, \ldots, |V|\}$ is separated into two parts: $E = \{E_1, E_2\}$. Define the weight matrix $W = [w_{i,j}]$ with each element as below:

$$w_{i,j} = \begin{cases} \frac{100}{d_{i,j}} & (i,j) \in E_1, \\ \frac{1}{d_{i,j}} & (i,j) \in E_2, \end{cases}$$

where $d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the Euclidean distance between vertex i and j. In Figure 2.3, the edges in E_1 are draw in the graph using arrows. Edges in E_2 are not shown.

Boundary nodes are selected as the first six nodes in graph: nodes 1, 2, 3, 4, 5 and 6. There are 14 frequency components in the whole heat content, including four real eigenvalues components and ten complex conjugates pair components. The eigenvalues (λ_k) and weights of eigenvectors (w_k) of the graph's Laplacian is shown in Table 2.2.3. For real eigenvalues, the heat content component is simple. Take λ_1 as an example,

$$Q_1 = 40.64e^{-0.02t}.$$

Two conjugated eigenvalues can be seen as one frequency component of the heat content function. For example, the heat content component of λ_2 and λ_3 is

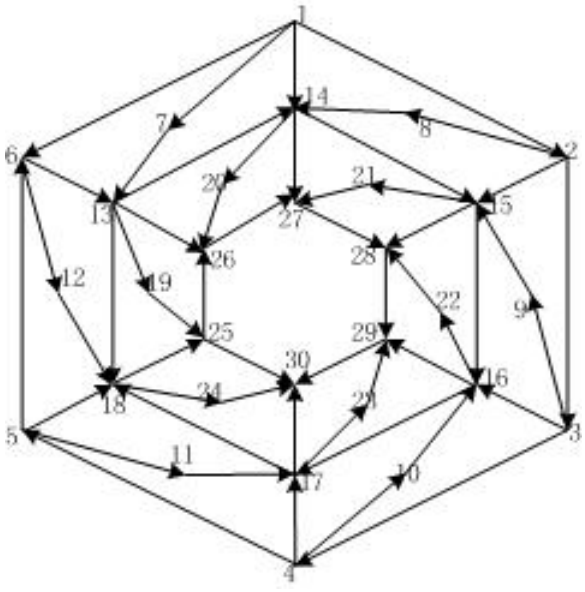


Figure 2.3. The asymmetric graph as an example; edges in E_2 are not shown

$$\begin{aligned} Q_{2,3} =& (-2.14 - 1.10i)e^{-(1.33+0.44i)t} + (-2.14 + 1.10i)e^{-(1.33-0.44i)t} \\ =& 2e^{-1.33t}(-2.14\cos(0.44t) - 1.10\sin(0.44t)) \end{aligned}$$

k	1	2	3	4	5	6
λ_k	0.02	1.33+0.44i	1.33-0.44i	1.49	0.73+0.35i	0.73-0.35i
w_k	40.64	-2.14-1.10i	-2.14+1.10i	1.14	8.41-13.05i	8.41+13.05i
k	7	8	9	10	11	12
λ_k	1.36	1.02+0.31i	1.02-0.31i	0.73+0.07i	0.73-0.07i	1.12+0.19i
w_k	-15.22	-12.37-10.69i	-12.37+10.69i	-2.40+2.92i	-2.40-2.92i	8.01+0.22i
k	13	14	15	16	17	18
λ_k	1.12-0.19i	0.90+0.07i	0.90-0.07i	0.995+0.13i	0.995-0.13i	0.97+0.09i
w_k	8.01-0.22i	0.95+0.22i	0.95-0.22i	-0.07-0.86i	-0.07+0.86i	-0.89-0.74i
k	19	20	21	22	23	24
λ_k	0.97-0.09i	1.08+0.09i	1.08-0.09i	1.15+0.01i	1.15-0.01i	1.11
w_k	-0.89+0.74i	-0.12+0.23i	-0.12-0.23i	0.01+0.32i	0.01-0.32i	-1.33

Table 2.1. Eigenvalues and eigenvectors of the graph's Laplacian

We draw the fourteen components separately and compare them with total heat content in Figure 2.4. As shown in figure, at the beginning part of the heat content curve, the high frequency components pull down the curve; and at last, the low

frequency component $Q_1(t)$ dominates, which makes it almost overlap with the total heat content curve.

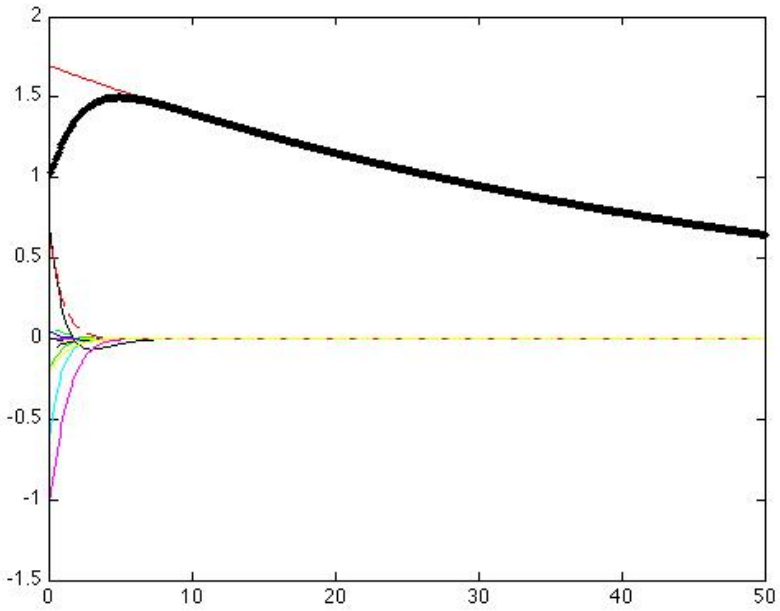


Figure 2.4. Heat content and heat content components for asymmetric graph in Figure 2.3 (**dark black line**: the total heat content $Q(t)$; solid red line : $Q_1(t)$, corresponding to λ_1).

2.2.3.1 Where is the oscillation?

As shown in Figure 2.4, the oscillatory part at the beginning of the total heat content comes from the complex conjugates pair components. However, it seems there is no oscillation after that. In Figure 2.5, we show that without $Q_1(t)$, the rest part of the heat content contains oscillations all the way to infinity time. However, since $Q_1(t)$ decreases slowly to 0 and is much larger than the rest part of the total heat content when t is large, we only observe the first oscillation in the total heat content.

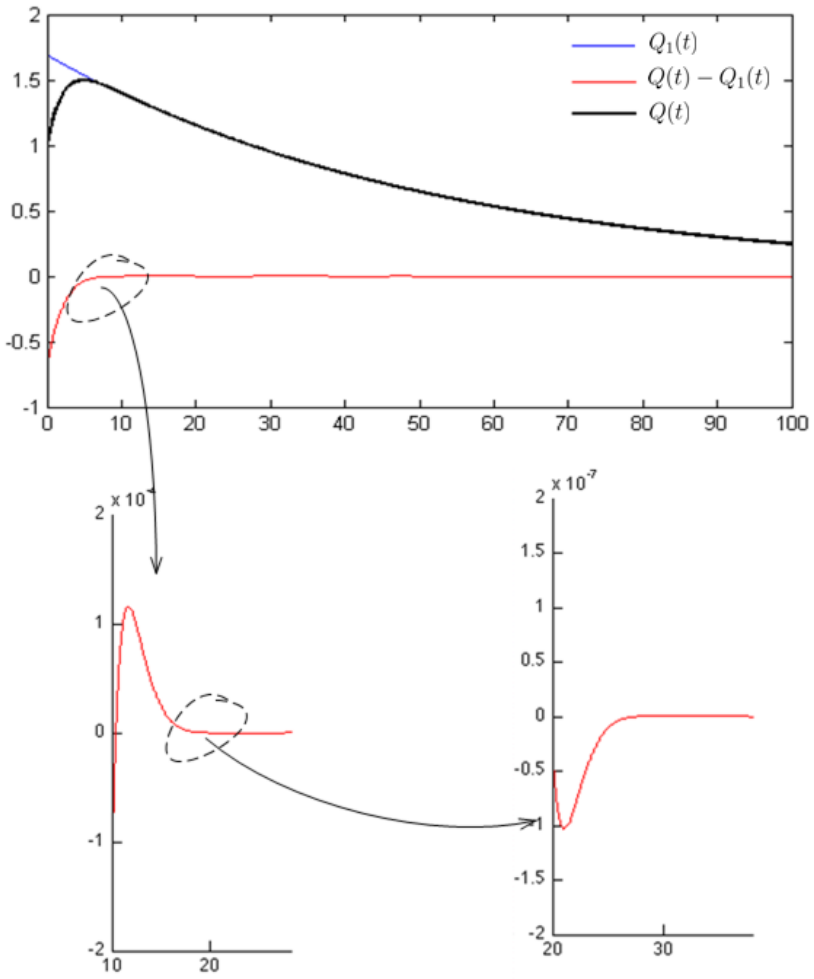


Figure 2.5. Where is the oscillations in heat content for asymmetric graphs?

2.3 Random walk methods for heat content estimation

Computing eigenvalues and eigenvectors of the Laplacian matrix needed for evaluating the heat content is very time consuming for large complex networks. We consider a random walk where the walker moves from vertex u to a neighboring vertex v with probability a_{uv}/d_u. Define the transition matrix $M = D^{-1}A$ and the lazy random walk transition matrix as

$$M_L = (1-\delta)I + \delta M \ (0 < \delta < 1).$$

For any given time $t = k\delta$, we have

$$P_t = M_L^k P_0 = [I - \frac{t}{k} L_r]^k P_0 \ \rightarrow \ e^{-L_r t} P_0. \tag{2.10}$$

Here the arrow ($\rightarrow$) implies taking the limit as $k \rightarrow \infty$ (at the same time $\delta \rightarrow 0$ while keeping $k\delta = t$). P_0 is the initial distribution of random walkers. We have $M_L^k \rightarrow e^{-L_r t}$. $M_L^k(u,v)$ measures the probability that a random walker starting at vertex u ends up at vertex v in k steps in the lazy random walk.

From equations (2.1), we obtain the following approximation for $Q(t)$:

$$\hat{Q}(t) = \sum_{u \in B^c} \sum_{v \in B^c} M_L^k(u,v) \sqrt{\frac{d_u}{d_v}}. \tag{2.11}$$

With the lazy random walk approximation, our algorithm avoids computation of the eigenvalues and weights of the eigenvectors. When the graph is not too big, we can simply use matrix multiplication method in Equation 2.11 to estimate the heat content for precision purpose. Matrix multiplication is the method we use in our experiment part since the networks we generate have only 2000 nodes.

Instead of computing $M_L^k(u,v)$ using matrix multiplication, we can also use the Monte Carlo method to estimate $M_L^k(u,v)$ based on the definition of $M_L^k(u,v)$. The

variance of the estimated value is inversely proportional to the amount of random walkers. Therefore, random walk simulation provides a trade off between precision and computation time. For each random walker in the graph, we only need to record its starting node and its current node. Each random walker walks independently and the next node the random walker moves to is calculated based on its current node's local information. So, we can compute and estimate the heat content in parallel. This property is important since the algorithm needs to be scale in the size of the graphs.

2.4 Generative models

We consider the following generative models to generate graphs to compare, including models for undirected graphs and directed graphs.

2.4.1 Undirected graph models

To check the ability of our method to distinguish symmetric graphs with different degree distributions, we introduce the following two models to generate undirected graphs with different degree distributions.

2.4.1.1 Erdös-Renyi (E-R) model

This model was first introduced in 1959 by Paul Erdös and Alfred Renyi [27]. They introduced two models. In our work, we consider the simpler one $G(n, p)$. The $G(n, p)$ graph is constructed by connecting nodes randomly and independently. An edge is added to each pair of vertices with a given probability p. For a graph with n vertices, the graph is connected with probability one when the edge adding probability is larger than $2\ln(n)/n$. the degree distribution follows a binomial distribution.

2.4.1.2 Barabási-Albert (B-A) model

This model was first introduced in [8]. The model starts with m_0 initial nodes. Each new node is connected to $m(m \leq m_0)$ existing nodes with a probability pro-

portional to the number of links that the existing nodes already have. The degree distribution follows $P(D = d) \sim d^{-3}$.

2.4.2 Directed power law graph models

To check the ability of our method to distinguish asymmetric graphs with different degree distributions, we introduce the following two types of directed graph growing models.

2.4.2.1 Directed E-R model

The directed E-R model is similar to the undirected model. A directed edge is added with a given probability p.

2.4.2.2 Directed power law model

In [51], a graph generative model is proposed to describe growing processes in the Web Graphs (WG).

1. With probability p, a new node is introduced, which immediately attaches to an existing node u with probability proportional to $d_u^{\text{in}} + \lambda_{\text{in}}$, where d_u^{in} is the in-degree of node u.

2. With probability q, a new edge from existing node v to node u is created with probability proportional to $(d_u^{\text{in}} + \lambda_{\text{in}})(d_v^{\text{out}} + \lambda_{\text{out}})$, where d_v^{out} is the out-degree of node v.

This model produces directed graphs with marginal in-degree and out-degree distributions that are both heavy tailed. The mean in/out-degree is p^{-1}. Let $P(d^{\text{in}} = i) \sim i^{-v_{\text{in}}}$ and $P(d^{\text{out}} = j) \sim j^{-v_{\text{out}}}$. We have

$$v_{\text{in}} = 2 + p\lambda_{\text{in}}$$

and

$$v_{\text{out}} = 1 + q^{-1} + p\lambda_{\text{out}}/q.$$

2.4.3 Generative models for the same degree distribution

In [40], the authors used the following four models to generate graphs with the same degree distribution. The four models are described as follows. We will use the four models to check the ability of our algorithm in distinguishing graphs with the same degree distribution.

2.4.3.1 Molloy-Reed (MR) model

M-R model is proposed by Molly and Reed in 1998 [66]. The algorithm is as follows:

1. Give each node a degree from the given distribution.
2. Connect a pair of vertices each time randomly, vertices are selected with probability proportional to the nodes open connections.
3. Repeat step 2 until there is no vertex with open connections.

2.4.3.2 Kalisky model

Kalisky's model was first proposed by Kalisky *et al.* in 2004 [45]. The algorithm of this model is as follows:

1. Give each node a degree from the given distribution.
2. Start with the maximal degree (K) vertex as the first layer. Randomly choose K open connections. The neighbors of a node in the first layer form the second layer.
3. Repeat the procedure for nodes in the second layer with open connections.

2.4.3.3 Model A (MA)

Model A and the following Model B are proposed by Grisi-Filho *et al.* in [40]. The algorithm associated with Model A is:

1. Give each node a degree from the given distribution.
2. Start with the maximal degree (K) vertex;
3. Connect the vertex with K other randomly selected vertices whose open connections is nonzero.
4. Repeat step 2 and 3 until there is no vertex with an open connection.

2.4.3.4 Model B (MB)

Model B differs the most from the other ones. At first, as with previous models, it assigns each node a degree from the given distribution. It also maintains a vector of vertices list and open connections. Then

1. Start with the maximal degree (K) vertex.
2. Connect the vertex with the first K other vertices in the vector.
3. Repeat step 1 and 2 until there is no vertex with open connections.

Since Model B selects target nodes with a given sequence, not randomly, the graphs generated by this model can easily contain some subgraphs with bipartite structure. In Figure 2.1, the three graphs are from B-A, MA and MB one by one. We observe that the graph from MB obviously differs from the other two. B-A and MA graphs have some minor differences compared to MB. Is our method enough to tell the differences between those graphs? In next part, we will design experiments to check.

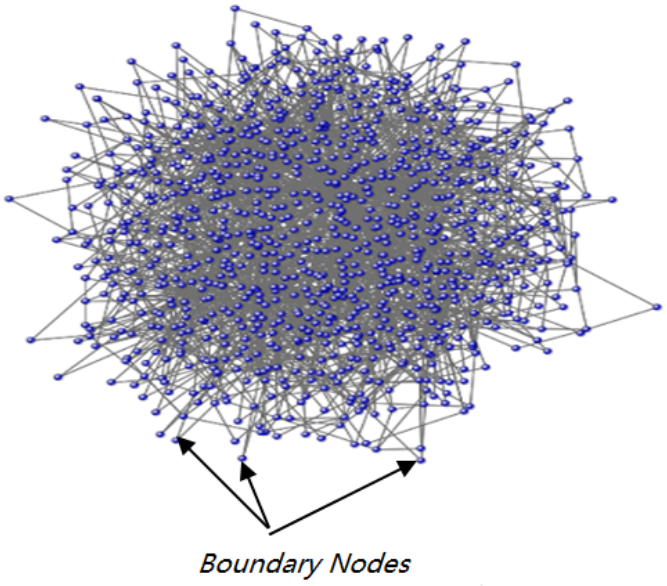


Figure 2.6. Boundary nodes selection

2.5 Experimental results

In the experiment part, we generate several groups of graphs to compare based on the generative models introduced in previous section.

For boundary nodes, we randomly select a fixed percentage of nodes in the graph with the smallest degree (as shown in Figure 2.6). The selection is based on the following reasons: (1) the selection of nodes should not change the graph structure dramatically. (2) the heat content should not go to zero too fast. For directed graphs, we selected nodes with the smallest in- and out-degree product. However, nodes with no in-degree are not included.

2.5.1 Undirected graphs with different degree distributions: B-A vs. E-R

Two groups of graphs are generated using the B-A and E-R models in Section 2.4.1, respectively. We generate random number in each step to determine whether to keep an edge or determine which node to connect. We use Matlab since it is efficient in matrix multiplication computing. The total number of nodes is 2000 and the number of boundary nodes is 40. Each group includes four graphs with average

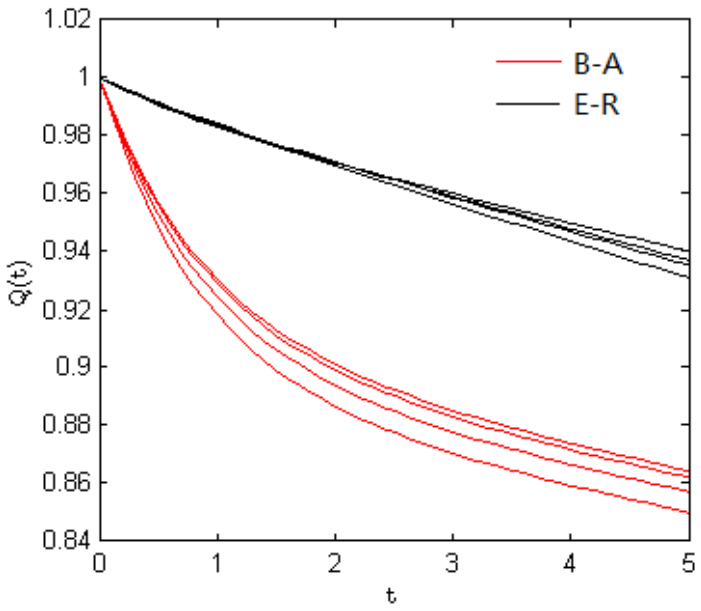


Figure 2.7. Heat content of B-A graphs and E-R graphs with different mean degrees

degree varying from 20 to 50. As shown in Figure 2.7, the heat contents of the two groups of graphs follow different patterns. When t is close to zero, the heat content for a power law graph drops faster than an E-R random graph, but the decrease speed slows down after the quick drop. In the Figure, we can see the graphs of the same type with different mean degrees are also differentiated.

For the spectra of these two kinds of graphs, Chung *et.al.* [18] proved that eigenvalues of the normalized Laplacian for both E-R random graphs and power law graphs satisfy the semicircle law. The circle radius is almost the same for graphs with the same mean degree (as shown in Figure 2.8). Using only the Laplacian spectrum we can hardly distinguish the two types of graphs. However, according to Equation (2.5), the values of α_i also play an important role in the heat contents. In Figure 2.9, we compare the weights for the two types of graphs except for the weight α_1 for the smallest eigenvalue λ_1 (since its too large to be shown in the same figure with the other weights). As shown in the figure, the weights (α) for the power law graph are much larger than those for the E-R random graph, which explains the different heat

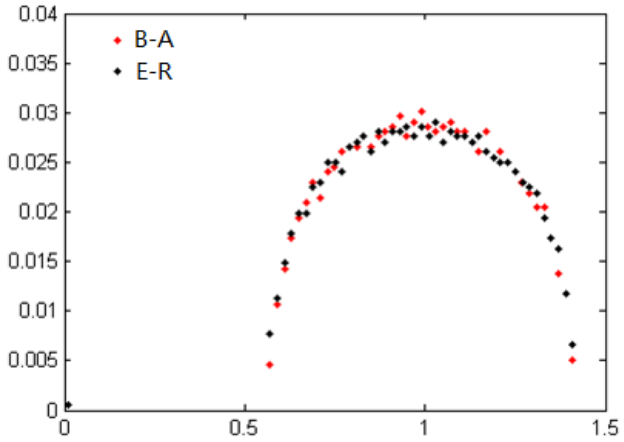


Figure 2.8. Spectra of a B-A graph and a E-R graph with the same mean degree

content behaviors for the two kinds of graphs. For the B-A graphs, the weights for the larger eigenvalues are much larger than that for E-R graphs, which pull down the heat content curves at the beginning part.

2.5.2 Directed graphs with different degree distributions: Krapivsky's Model vs. directed E-R Model

For directed graph comparison, two groups of graphs are generated using the 'WG' model and the E-R model in Section 2.4.2. Each group contains four graphs with different average degrees by setting p in 'WG' model to be 0.1, 0.15, 0.2 and 0.25, respectively. The total number of nodes is 2000 with 10% of those assigned to be boundary. As shown in Figure 2.10, directed power law graphs and E-R random graphs exhibit similar behavior to undirected graphs.

2.5.3 Graphs with the same degree distribution

We first generate a 2000 node power law graph using B-A model with mean degree 10. The number of boundary nodes is 40. Next using each one of the four generative models MR, Kalisky, MA and MB, we independently generate one graph with the same degree distribution as the graph generated by the B-A model. We use Matlab

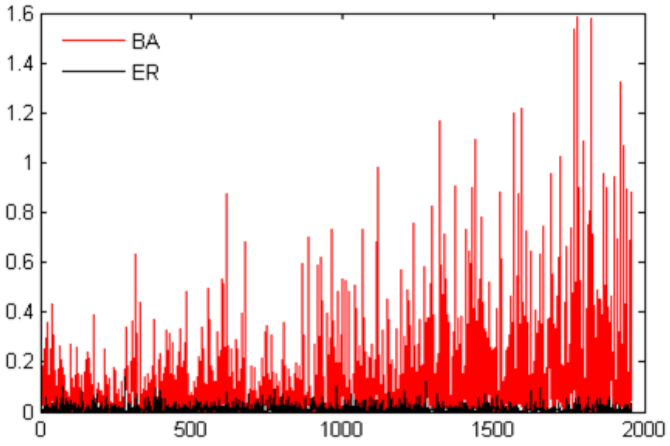


Figure 2.9. Weights α of a B-A graph and a E-R graph with the same mean degree (without showing weight α_1 for the smallest eigenvalue λ_1)

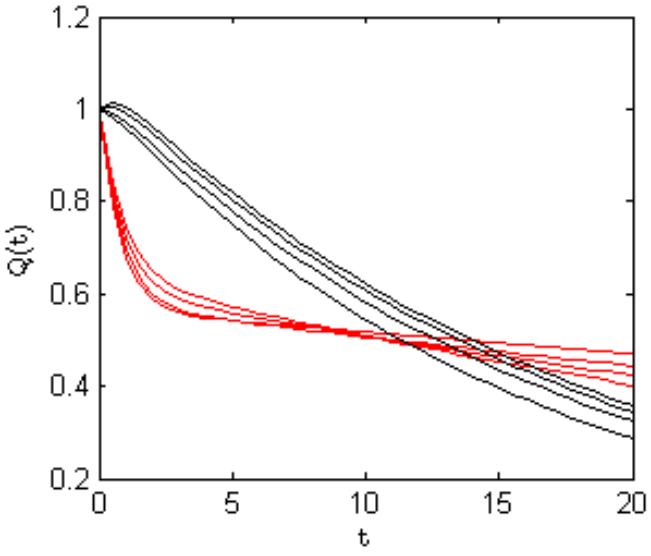


Figure 2.10. Heat content of Krapivsky's graphs and directed E-R graphs with different mean degrees

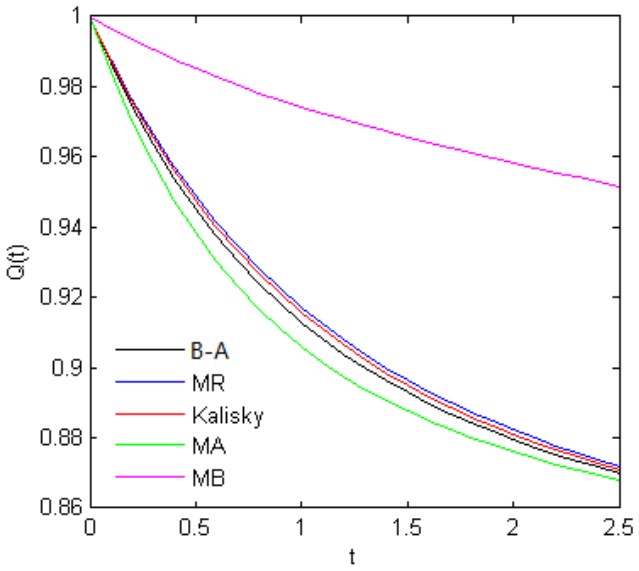


Figure 2.11. Heat Content of a B-A graph and graphs generated by M-R model, Kalisky, Model A and Model B with the same degree distribution

and follow the steps introduced in Section 2.4.3. The heat contents comparison results are shown in Figures 2.11.

We observe that graphs with the same degree distribution can be distinguished according to their heat content behaviors. Even with the same degree distribution, the differences of the heat contents between the five generative models are still noticeable. We also notice that the heat contents for model B (the curves in color magenta) perform differently from the other four models (B-A, MR, Kalisky, and MA). This result is consistent with the conclusions in [40] that, although with the same degree distribution, model B gives the most different network compared to the other four models.

Next, we compare the spectra of the five models (including B-A Model) in Figure 2.12 and the eigenvector weights in Figure 2.13. We observe that, the spectra of the graphs generated by B-A, MR and Kalisky all follow semi-circle. For MA and MB model, the spectra do not follow a perfect semi-circle as the other three graphs.

For the weights of eigenvectors, the graphs using the same model with different parameters have similar patterns. The weights for MA and MB are different in an apparent way. For MB, the weights are larger for smaller eigenvalues, so it drops slower compared to the other four. We can tell the difference between BA and the other two graphs (MR and Kalisky) at the very beginning part. The heat content of MR and Kalisky graphs are quite closer and the weights of the two graphs look similar. In [40], the authors also pointed out that MR and Kalisky networks showed similar number of components and giant component size. In their paper, they plot the average degree of the nearest neighbors as a function of the degree of a given vertex for the graphs generated by each model. The results for MR and Kalisky looked similar to each other. So, our results are consistent to the results in [40]. Reader is referred to [40] for more details.

We try to explain the similar between MR and Kalisky as follows. For Kalisky model, although it uses a layered structure, the nodes with the largest degrees will be connected to a huge number of the other nodes in the graph. The selected nodes are placed in the second layer. And then, for the third layer, it will exhaust almost all the remaining nodes in the graph. Thus the layer structure in the Kalisky model is not quite obvious for a graph with power law distribution. Second, the two models both randomly select open connections, which makes the probability of a node being selected proportional to the nodes' open connections. The two reasons make the power-law graphs generated by the two models similar to each other.

2.6 Conclusions and future work

In this chapter, we proposed a random walk method to estimate the heat content on graphs for the purpose of determining if two graphs are similar or not. We first applied the method to compare graphs with different degree distributions. Graphs with heavy tail degree distributions have different heat content curves compared to

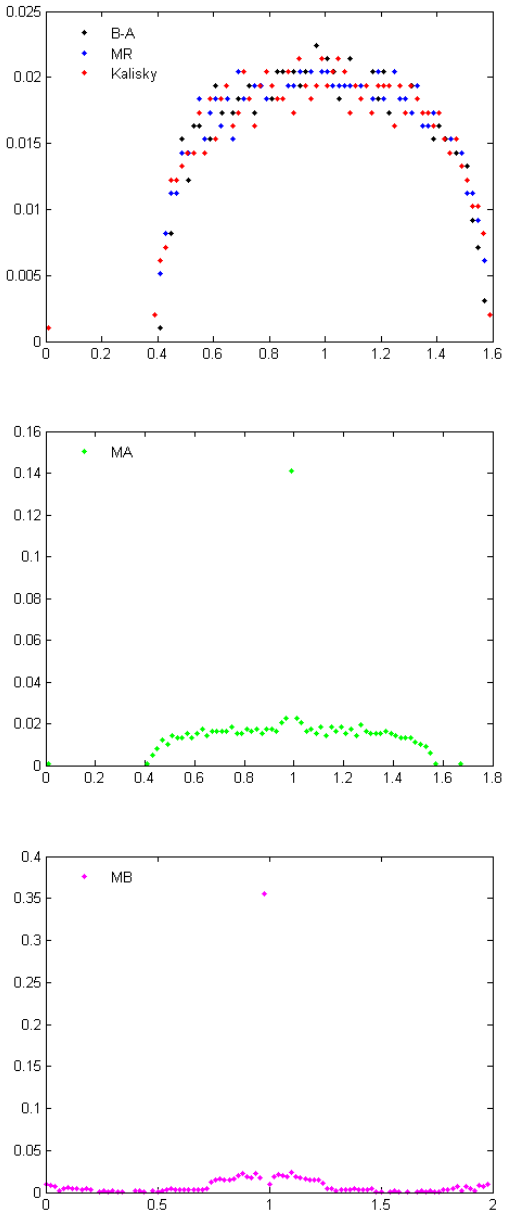


Figure 2.12. Spectra of the graphs generated by B-A model and the other four generative models with the same degree distribution

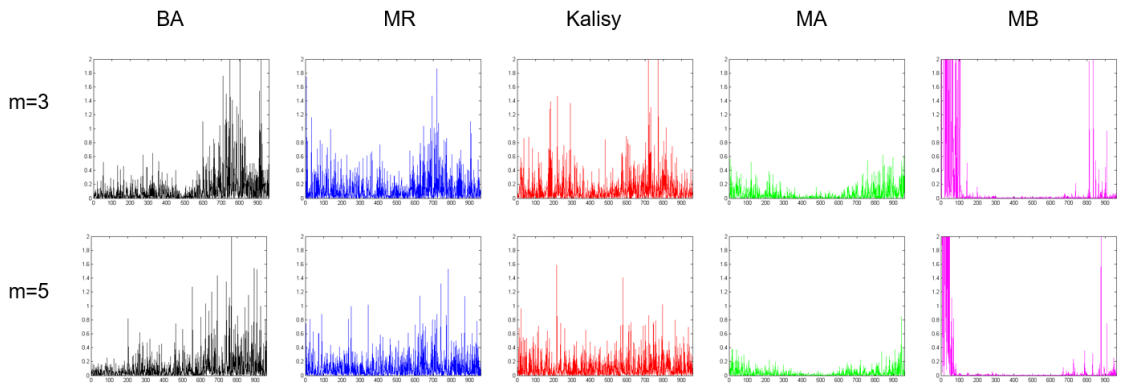


Figure 2.13. Weights α of the graphs generated by B-A model and the other four generative models with the same degree distribution, $m = 3$ and 5 in B-A model

random graphs generated by the E-R model: the decrease rate for power law graphs is much larger than E-R graphs at the very beginning part. Our method can also distinguish graphs with the same degree distribution but different structural properties. Experiments show that our algorithm performs better in graph comparison than some other feature extraction methods using eigenvalues and degree distributions. We have published a paper in [55] for this part of work. The reader are encouraged to read this paper as a reference.

Our algorithm can also be applied in other classification problems, such as image retrieval as in [46, 47]. Readers are referred to the two papers for details. We will discuss the pros and cons of this application in Chapter 6. In our future work, we consider modifying our algorithm to amplify the oscillation and applying our method to other classification problems like audio classification problems.

CHAPTER 3

PCSDE MODELS FOR BIVARIATE HEAVY TAILED DISTRIBUTIONS-PART I: MODELS WITH A SHARED POISSON COUNTER AND COUPLED GROWTH

3.1 Introduction

In Section 2.4.2, we introduced a directed power law network growth model developed by Krapivsky. There are many other directed network growth models, such as Price's model for citation networks [22]; Bollobás' model [9] for World Wide Web. Those models are useful in modeling the generative process of some bivariate data in real world networks. Studying the origin of high dimensional power law behavior is an important task. In our work, we propose mathematical models to explain the origin of bivariate power law data.

3.1.1 Background

Bivariate power law distributions have been found in directed real world networks, such as citation networks (arXiv, CiteSeer, US patent), social networks (Facebook, Youtube, Flickr, Livejournal), and Web networks (Google) [53]. We are particularly interested in studying the social network datasets, since we all use social networks everyday. By comparison, we will also study a Google hyperlink network. Some basic information about the datasets we will study in this thesis are provided in Table 3.1. The Facebook wall posts dataset is a small subset of posts to other user's wall on Facebook. The nodes of the network are Facebook users and each directed edge represents one post, linking the users writing a post to the users whose wall the post is written on. The other three datasets are the same datasets as in [61]. They are

Graph	Size (vertices)	Volume (edges)	Type
Facebook wall posts	46,952	876,993	Directed, Social
Youtube links	1,138,499	4,942,297	Directed, Social
Flickr	2,302,925	33,140,017	Directed, Social
LiveJournal	4,847,571	68,475,391	Directed, Social
Google Hyperlink	875,713	5,105,039	Directed, Web

Table 3.1. Datasets statistics

social networks with users and friendship connections. Google hyperlink network is a network of web pages connected by hyperlinks. It is a webgraph from the Google programming contest in 2002.

Take Youtube and Google hyperlinks as examples (as shown in Figures 3.1 and 3.2). In these networks, both in-degree and out-degree exhibit power law distributions at the tail. Strong dependence correlations are shown between node in-degrees and out-degrees in Youtube. As shown in Figure 3.3, in-degrees and out-degrees of the four social network datasets are gathered along the line where the two variables are equal. We use the Pearson correlation coefficient, which is a measure of the linear correlation between two variables, to quantify the dependency between in-degree and out-degree of the datasets. The correlation coefficients of the four social network datasets Youtube, Facebook, Flickr and LiveJournal, are 0.9492, 0.8470, 0.7558 and 0.6478 separately. On the other hand, for Web Google, the correlation coefficient between in-degree and out-degree is only 0.1365. In [61], the authors plot overlap percentage as a function of fraction of users ranked by in-degree and out-degree in descending order. We redo the experiments on the four social network datasets (Facebook, Youtube, Flickr, and LiveJournal) and one Web dataset (Google). As shown in Figure 3.4, the overlap percentage is larger for the four social network datasets, which is different from the Web dataset. Experiments shows that, for social network datasets, the top 1% of nodes ranked by out-degree has a more than 64% overlap percentage with the top 1% of nodes ranked by out-degree (Youtube: 90.62%;

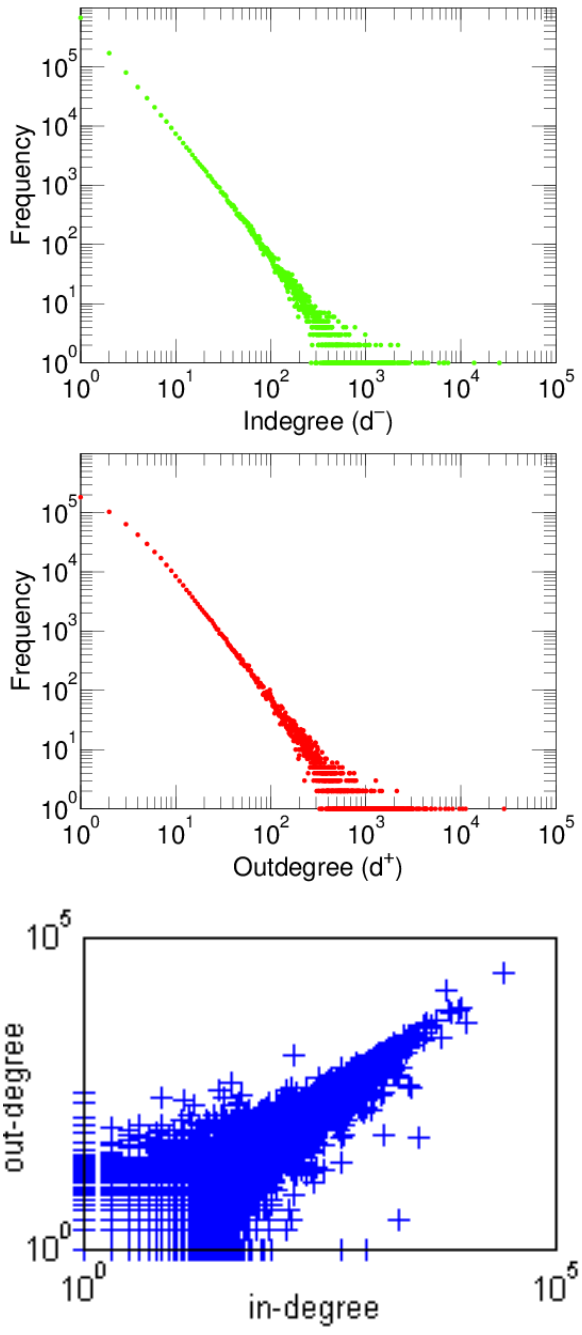


Figure 3.1. Youtube dataset

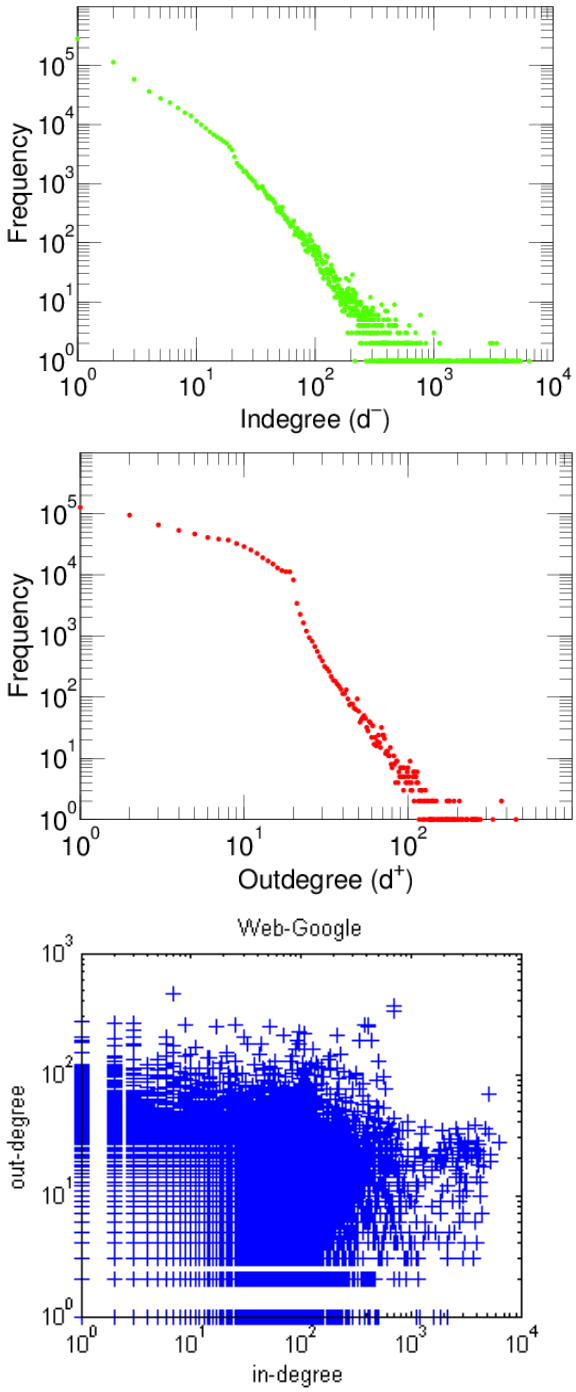


Figure 3.2. Web Google

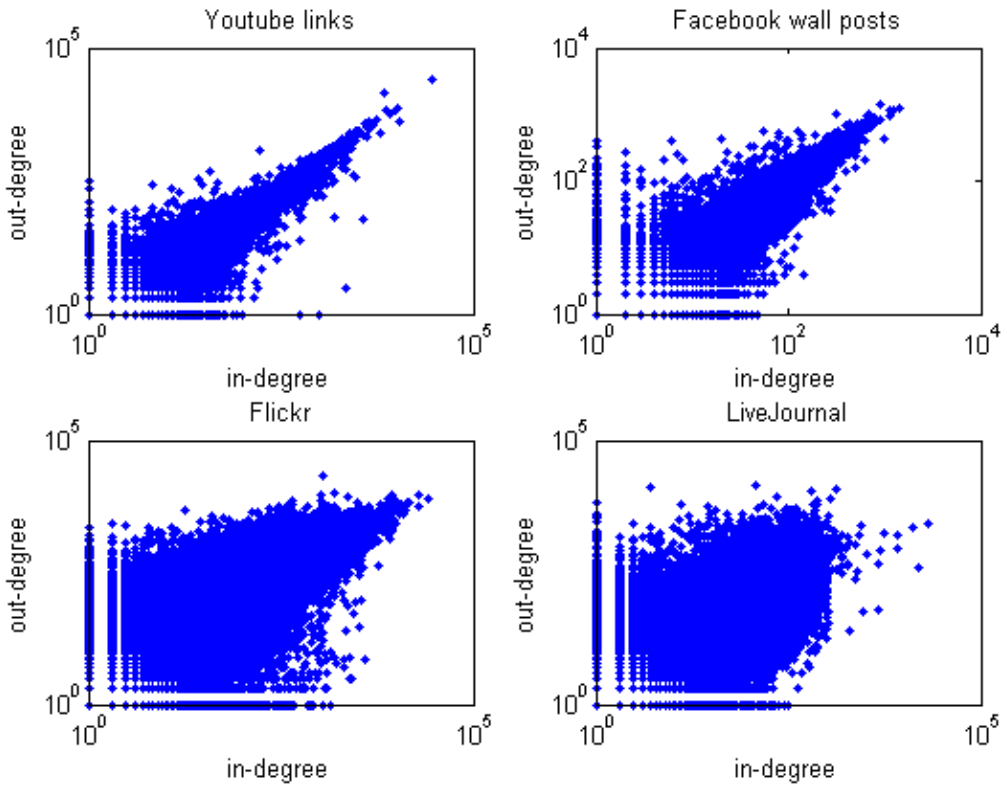


Figure 3.3. Scatter plot of four social network datasets

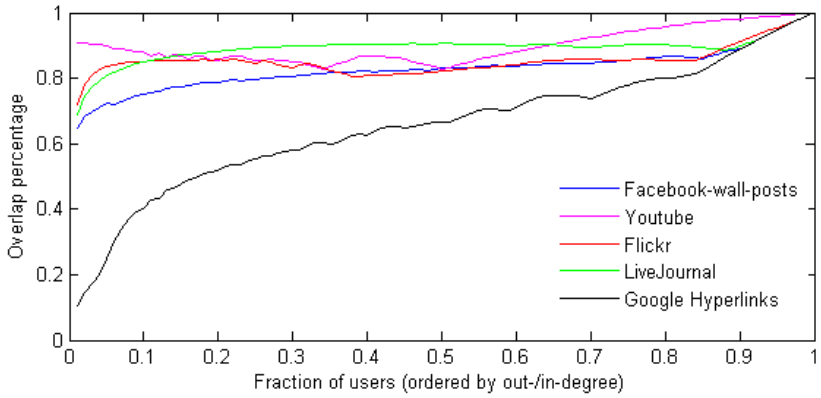


Figure 3.4. Overlap between top $x\%$ of nodes ranked by out-degree and in-degree

Flickr: 72.09%; LiveJournal: 68.94%; Facebook: 64.82%). The corresponding overlap percentage in the Google Web is only 10.72%.

For the tail behavior, if we find a node in social network with a high in-degree, it is very possible that the node also has a large out-degree. The tendency for a large in-degree nodes to have a large out-degree in a Web dataset is much less. This feature can be measured by the following notion: Tail Dependence Coefficient (TDC), which is defined as follows:

$$TDC = \lim_{t \to 1^-} \mathrm{P}(d^+ > F_{d^+}^{-1}(t) | d^- > F_{d^-}^{-1}(t)), \tag{3.1}$$

where d^+ denotes out-degree, d^- denotes in-degree and $F_X(x) = P(X \leq x)$, the cumulative distribution function. The tail dependence coefficient in our work is an upper tail dependence coefficient [26]. The tail dependence of a pair of random variables is a measure of their co-movements in the tails of the distributions. If the two variables share the same marginal distribution, then Equation (3.1) can also be written into the following equation:

$$TDC = \lim_{x \to \infty} \mathrm{P}(d^+ > x | d^- > x). \tag{3.2}$$

For finite x, we define dependence coefficient $DC(x) = \mathrm{P}(d^+ > x | d^- > x)$ as a function of x.

In [61], the authors explained the large correlation in social network as a result of the tendency of users to reciprocate links from other users who point to them. In our work, we try to design a mathematical model to explain the behavior observed in social networks.

3.1.2 Mathematical tools

Stochastic differential equations (SDEs) are widely used to model various phenomena such as stock prices and thermal fluctuations. We are interested in the following models.

- Geometric Brownian motion (GBM) [71]

$$dX = \beta X dt + \sigma X dW \tag{3.3}$$

- Constant Elasticity of Variance (CEV) model [20]

$$dX = \beta X dt + \sigma X^{\gamma} dW, \ (\gamma \geq 0, \ \sigma \geq 0) \tag{3.4}$$

- Cox-Ingersoll-Ross (CIR) model [21]

$$dX = a(b - X)dt + \sigma\sqrt{X}dW \tag{3.5}$$

Those models normally consist of an ordinary differential equation and an additional random white noise term (W is standard Brownian motion). In [78] and [82], the authors explain how to use the GBM model to generate a double Pareto distribution and how to estimate the parameters to fit the model to empirical distributions. By adding Poisson counters to the above SDE models, we arrive at a new set of models, the PCSDE models. In [43], Jiang *et al.* first presented different PCSDE models with no Brownian motion part to generate power law distributions at different tail parts, such as at lower tail, upper tail, or near a critical point. Then, by adding a Brownian motion component, this paper proved that the PCSDE model based on GBM could be used to produce double Pareto distributions, which confirms results in [78] and [82]. We will use the univariate PCSDE models in [43] as the basis of our bivariate model.

3.1.3 Chapter outline

Our work starts by reviewing the univariate PCSDE model for lower tail power law distribution in [12], upper tail power law distribution and double Pareto distribution in [43]. Then we explore a bivariate extensions to explain correlated bivariate power law behavior in social networks.

In [7], Asimit *et al.* proposed a new type of multivariate Pareto distribution. This distribution has arbitrarily parameterized margins compared to the traditional multivariate Pareto distribution of the second kind [6]. The first type of bivariate PCSDE model with a shared Poisson counter [42] is formulated in the spirit of this paper and this model can generate a correlated bivariate power law distribution just like the one proposed in [7]. We call the PCSDE model with a shared Poisson counter **Type 1** model. In Type 1 model, the two growth processes are uncoupled within each session. But the life time (the time between occurrence of the last jump of the Poisson counter and the observation point) of the two processes are the same from time to time.

The second model takes a complementary approach: the two growth processes are coupled, but their life times are independent. We call the model with coupled growth **Type 2** model.

Our study of each model divides into three parts. First, we check whether the marginal density of the proposed bivariate model follows a power law distribution at the upper tail. Then, we check whether the model can generate synthetic data that is consistent with the real data in joint distribution. Last, we check the tail dependence of the model. Reviewing the definition TDC in equation (3.1), TDC relates to the dependence among the extreme values. We are especially interested in models with fractional TDCs in order to match to the tail behavior of real datasets in social networks.

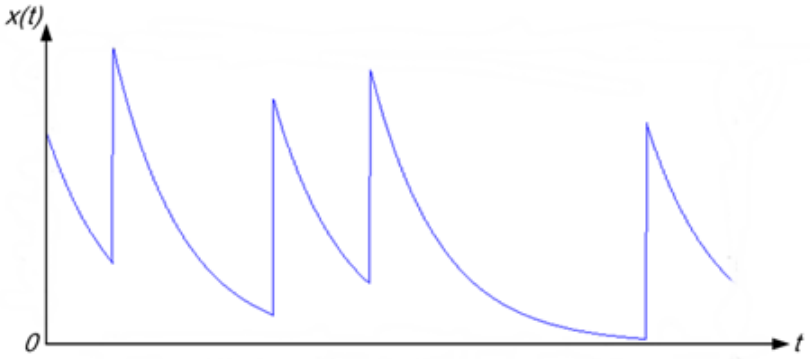


Figure 3.5. Sample path of lower tail power law model in Equation (3.6)

3.2 Univariate PCSDE models

In this section, we review several univariate PCSDE Models for: lower tail power law, upper tail power law and double Pareto distribution. These univariate models form the basis for our bivariate extensions.

3.2.1 Univariate PCSDE model for lower tail power law

As discussed in [12], the following PCSDE model generates a lower tail power law distribution:

$$dX(t) = -\beta X(t)dt + \sigma dN(t), \tag{3.6}$$

where $\beta, \sigma > 0$ and N is a Poisson process with rate λ. Figure 3.5 illustrates a sample path of this model.

As discussed in [34], the Fokker-Planck equation of this model is

$$\frac{\partial f_X(x,t)}{\partial t} = \beta \frac{\partial}{\partial x}[x f_X(x,t)] + \lambda f_X(x-\sigma,t) - \lambda f_X(x,t). \tag{3.7}$$

Since $f_X(x,t) = 0$ for $x \leq 0$, $f(x-\sigma,t) = 0$ when $x \in (0,\sigma]$. Let $t \to \infty$, for the steady state, we have

$$\beta \frac{\partial}{\partial x}[x f_X(x)] - \lambda f_X(x,t) = 0, x \in (0,\sigma], \tag{3.8}$$

which yields

$$f_X(x) = Cx^{\alpha-1}, x \in (0, \sigma], \tag{3.9}$$

where $\alpha = \frac{\lambda}{\beta}$.

This model produces power law distribution at the lower tail. This model can be converted to an upper tail power law generator.

3.2.2 Univariate PCSDE model for upper tail power law

This subsection presents two types of PCSDE models for upper tail power law. We need the following Ito's rule for PC SDEs [34].

Ito's rule for PC SDE

Let

$$dX = f(X)dt + \sum_{1}^{n} g_i(X)dN_i$$

where N_i is an independent Poisson counter. Then

$$d\psi(X) = \langle \frac{\partial \psi}{\partial X}, f(X) \rangle dt + \sum_{i=1}^{n} [\psi(X + g_i(X)) - \psi(X)]dN_i$$

where$\langle \cdot, \cdot \rangle$.

3.2.2.1 Convert lower tail to upper tail

As in the previous subsection, PCSDE model $dX(t) = -\beta X(t)dt + \sigma dN(t)$ produces a lower tail power law distribution. In [43], it is shown that this model can be converted to an upper tail generator by letting $Y(t) = X(t)^{-1}$. With $\epsilon \triangleq \sigma^{-1}$, we have

$$dY(t) = \beta Y(t)dt - \frac{Y(t-)^2}{\epsilon + Y(t-)}dN(t) \tag{3.10}$$

Figure 3.6 presents a sample path of this model. As shown in the figure, Y starts at a small positive value. It grows exponentially and reverts to a value that relates

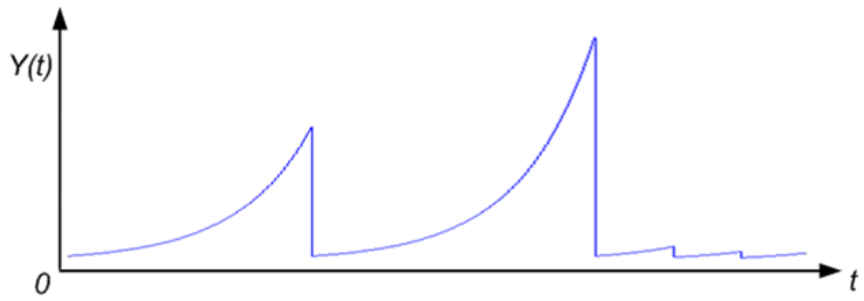


Figure 3.6. Sample path of upper tail power law model in Equation (3.10)

to its current value. For a smooth function $\psi(Y)$, by using Ito's rule for PC SDE, we have

$$\begin{aligned} d\psi(Y) =& \frac{\partial \psi}{\partial Y}(\beta Y)dt + [\psi(Y - \frac{Y^2}{\epsilon + Y}) - \psi(Y)]dN \\ =& \frac{\partial \psi}{\partial Y}(\beta Y)dt + [\psi(\frac{\epsilon Y}{\epsilon + Y}) - \psi(Y)]dN. \end{aligned} \tag{3.11}$$

Taking expectation of both sides of (3.11),

$$\frac{dE\psi(Y)}{dt} = \beta E[\frac{\partial \psi(Y)}{\partial Y}Y] + \lambda E\psi(\frac{\epsilon Y}{\epsilon + Y}) - \lambda E\psi(Y).$$

Assuming that the density $f_Y(y)$ vanishes at $y = \pm\infty$,

$$\int \psi(y)\frac{df_Y(y)}{dt} = -\beta \int \psi(y)\frac{\partial}{\partial y}[yf_Y(y)] + \lambda \int \psi(y)f_Y(h(y))\frac{\partial h(y)}{\partial y} - \lambda \int \psi(y)f_Y(y), \tag{3.12}$$

or,

$$\frac{df_Y(y)}{dt} = -\beta\frac{\partial}{\partial y}[f_Y(y)y] + \lambda f_Y(h(y))\frac{\partial h(y)}{\partial y} - \lambda f_Y(y), \tag{3.13}$$

where $h(y) = \frac{y\epsilon}{\epsilon - y}$.

Assuming initial condition $Y(0) > 0$, we have $Y(t) > 0$. So, $f_Y(y) = 0$ when $y \leq 0$. When $y > \epsilon$, $h(y) < 0$, the Fokker-Planck equation of this model becomes

$$\frac{df_Y(y)}{dt} = -\beta\frac{\partial}{\partial y}[f_Y(y)y] - \lambda f_Y(y), \ y \geq \epsilon. \tag{3.14}$$

When $t \to \infty$, for steady state, we have

$$\beta \frac{\partial}{\partial y}[f_Y(y)y] = -\lambda f_Y(y),\ y \geq \epsilon, \tag{3.15}$$

which gives

$$f_Y(y) = Cy^{-(1+\frac{\lambda}{\beta})}, y \geq \epsilon. \tag{3.16}$$

Thus, this model produces a power law distribution at the upper tail.

3.2.2.2 A simpler model

By letting $Y(t)$ in Equation (3.10) revert to a fixed point, we produce a simpler PCSDE model

$$dX(t) = \beta X(t)dt + (\epsilon - X(t-))dN(t), \tag{3.17}$$

where $\beta, x_0 > 0$. N is a Poisson process with rate λ. In this model, X grows exponentially with rate β, and reverts to a fixed value ϵ after an exponential distributed life time with rate λ. The whole process then repeats.

We give the corresponding characteristic function $\Phi_X(k,t) = \mathrm{E}[e^{jkX(t)}]$ by Ito's rule

$$(\frac{\partial}{\partial t} - \beta k \frac{\partial}{\partial k})\Phi_X(k,t) = -\lambda \Phi_X(k,t) + \lambda e^{jk\epsilon}. \tag{3.18}$$

Solving Equation (3.18) as in [43] yields

$$\Phi_X(k,t) = e^{-\lambda t}\Phi_X(ke^{\beta t}, 0) + \lambda \int_0^t e^{-\lambda(t-s)} e^{j\epsilon k e^{\beta(t-s)}} ds.$$

A change of variable $x = \epsilon e^{\beta(t-s)}$ along with letting $t \to \infty$ yields

$$\Phi_X(k,\infty) = \frac{\lambda}{\beta\epsilon} \int_\epsilon^\infty \left(\frac{x}{\epsilon}\right)^{-\frac{\lambda}{\beta}-1} e^{jkx} dx. \tag{3.19}$$

The steady-state density of X is given by taking the inverse Fourier transform

$$f_X(x) = \frac{\lambda}{\beta\epsilon} \left(\frac{x}{\epsilon}\right)^{-\frac{\lambda}{\beta}-1},\ x \geq \epsilon, \tag{3.20}$$

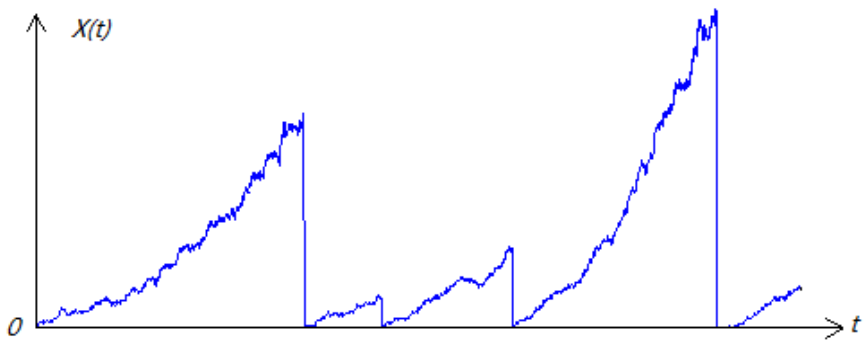


Figure 3.7. A sample path of the model in Equation (3.22) ($\lambda = 1$, $\beta = 1$, $\sigma = 0.2$ and $\epsilon = 1$)

and the Complementary Cumulative Distribution Function (CCDF) is

$$\bar{F}_X(x) = \left(\frac{x}{\epsilon}\right)^{-\frac{\lambda}{\beta}}, \quad x \geq \epsilon. \tag{3.21}$$

From this computation, we see that this model yields similar result to the model in the previous sub-subsection. In our following bivariate extension, we use this model as a basis because of its conciseness.

3.2.3 Univariate PCSDE model for double-Pareto distribution

A PCSDE model with Browian motion component based on Gemetric Brownian motion (GBM) is

$$dX(t) = \beta X(t)dt + \sigma X(t)dW(t) + (\epsilon - X(t))dN(t), \tag{3.22}$$

where $\beta, \sigma, \epsilon > 0$, W is standard Brownian motion and N is a Poisson process with rate λ, independent of W. A sample path of this model is shown in Figure (3.7).

This model can be used to produce a Pareto distribution. This model was first described in [78]. Reed further discussed the implications of this model in modeling size distributions like income in [82] and [79]. In [30], many more Pareto distributions

are discussed, like friendship in social networks, number of downloads on the Internet, oil field reserves, etc.

Since this mode is not used in the first two types of bivariate PCSDE models, we will leave it here and discuss more details about this model in Chapter 4.

In the following, we will propose different types of bivariate extensions based on univariate PCSDE models that we discussed above, and talk about their pros and cons in matching real datasets.

3.3 Type 1: bivariate PCSDE model with a shared Poisson counter

3.3.1 Basic model

To develop a bivariate model, we start with two independent univariate upper tail power law generators

$$dX_i = \beta_i X_i + (\epsilon_i - X_i)dN_i,\ i = 1, 2.$$

We omit the t in parentheses in this and the following equations. We observe, X_1 and X_2 grow separately and their life time are controlled by two independent Poisson counters. So the above model generates independent bivariate power law data. To introduce dependence between the two variables, a shared Poisson counter N_0 is needed. This model was first presented in [42].

3.3.1.1 Model formulation

We simplify the above model by setting growth rates $\beta_1 = \beta_2 = 1$ and initial values $\epsilon_1 = \epsilon_2 = 1$,

$$dX_i = X_i dt + (1 - X_i)(dN_0 + dN_i),\ i =, 1, 2. \tag{3.23}$$

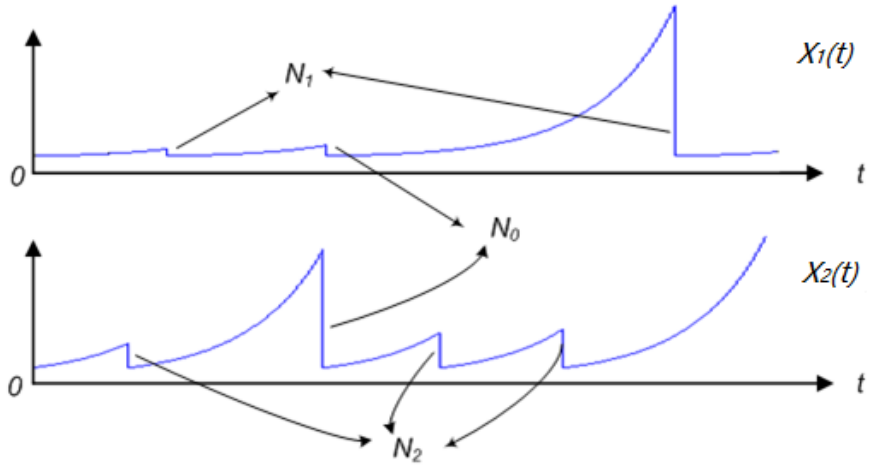


Figure 3.8. Sample path of basic bivariate model with a shared Poisson counter

Here N_0, N_1, and N_2 are independent Poisson counters with rates λ_0, λ_1, and λ_2. Figure 3.8 illustrates a sample path of this model.

3.3.1.2 Marginal and joint density

The marginal steady-state density is

$$f_{X_i}(x_i) = (\lambda_0 + \lambda_i) x_i^{-(\lambda_0+\lambda_i+1)}, \; x_i \geq 1, \tag{3.24}$$

and the marginal CCDF is

$$\bar{F}_{X_i}(x_i) = x_i^{-(\lambda_0+\lambda_i)}, x_i \geq 1, \; i = 1, 2. \tag{3.25}$$

We use the characteristic function to compute the joint density. Let

$$\Phi(k_1, k_2, t) = \mathrm{E}[e^{j \sum_i k_i X_i(t)}],$$
$$\Phi_i(k_i, t) = \mathrm{E}[e^{j k_i X_i(t)}].$$

Applying Ito's rule yields

$$\left(\frac{\partial}{\partial t} - k_i \frac{\partial}{\partial k_i}\right) \Phi - \lambda_+ \Phi + \lambda_0 e^{j \sum_i k_i} + \lambda_1 e^{j k_1} \Phi_2 + \lambda_2 e^{j k_2} \Phi_1, \tag{3.26}$$

where $\lambda_+ = \lambda_0 + \lambda_1 + \lambda_2$. Solving equation (3.26) and letting $t \to \infty$ yields

$$\begin{aligned} \Phi(k_1, k_2, \infty) = & \int_1^\infty x^{-\lambda_+ - 1} \lambda_0 e^{j \sum_i k_i x} dx \\ & + \int_1^\infty x_1^{-\lambda_+ - 1} \lambda_1 e^{j k_1 x_1} \Phi_2(k_2 x_1, \infty) dx_1 \\ & + \int_1^\infty x_2^{-\lambda_+ - 1} \lambda_2 e^{j k_2 x_2} \Phi_1(k_1 x_2, \infty) dx_2, \end{aligned} \tag{3.27}$$

and the inverse Fourier transform gives

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) = & \lambda_0 x_1^{-\lambda_+ - 1} u(x_1 - 1) \delta(x_1 - x_2) \\ & + \lambda_1 x_1^{-\lambda_+ - 1} f_{X_2}(x_2 x_1^{-1}) x_1^{-1} u(x_1 - 1) \\ & + \lambda_2 x_2^{-\lambda_+ - 1} f_{X_1}(x_1 x_2^{-1}) x_2^{-1} u(x_2 - 1), \end{aligned} \tag{3.28}$$

where $u(x) = 1$ when $x \geq 0$; otherwise, $u(x) = 0$; $\delta(x)$ is the Dirac delta function. The two variables in this model are not independent since $f_{X_1,X_2}(x_1, x_2) \neq f_{X_1}(x_1) f_{X_2}(x_2)$.

The joint CCDF of the model is computed from (3.28)

$$\begin{aligned} & \bar{F}_{X_1,X_2}(x, x) \\ = & \int_x^\infty dx_1 \int_x^\infty dx_2 f_{X_1,X_2}(x_1, x_2), \\ = & \lambda_0 \int_x^\infty dx_1 x_1^{-\lambda_+ - 1} \\ & + \lambda_1 \int_x^\infty dx_1 x_1^{-\lambda_+ - 1} \int_{x_1}^\infty dx_2 f_{X_2}(x_2 x_1^{-1}) x_1^{-1} \\ & + \lambda_2 \int_x^\infty dx_2 x_2^{-\lambda_+ - 1} \int_{x_2}^\infty dx_1 f_{X_1}(x_1 x_2^{-1}) x_2^{-1}, \\ = & x^{-\lambda_+}. \end{aligned} \tag{3.29}$$

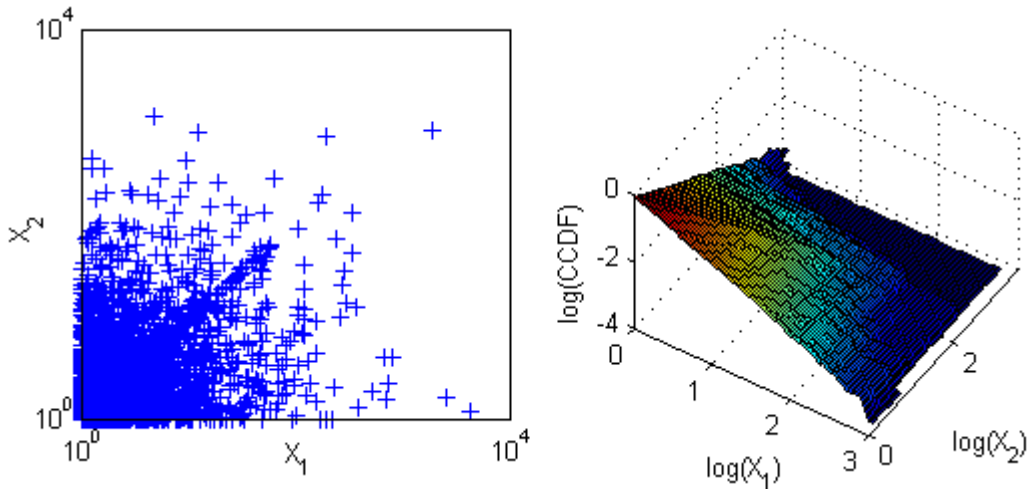


Figure 3.9. Data sample from Type 1 model and joint CCDF of the sample data

With the marginal CCDF in (3.25), we obtain the tail dependence coefficient (TDC) of this model

$$P(X_2 > x | X_1 > x) = \frac{\bar{F}_{X_1,X_2}(x,x)}{\bar{F}_{X_1}(x)} = x^{-\lambda_2} \xrightarrow{x\to\infty} 0. \tag{3.30}$$

3.3.1.3 Experimental results

We present a scatter plot and CCDF of a group of sample data to compare this model to the real datasets in social networks, as shown in Figures 3.9 and 3.10. The scatter plot shows that the density of this model does not fit the real dataset. The data from the Type 1 model are formed by two types of data: (1) X_1 and X_2 are independent; (2) X_1 and X_2 equals. This is not the case in real datasets.

On the other hand, as indicated in (3.30), although the model in (3.23) is useful for generating correlated bivariate power law data, the TDC of this model is 0. We try to explain why this model is asymptotically independent as follows. Since the observed state is an exponential function of the life time, we require the life time to be larger than $\ln x$ to observe a value larger than x. Given $X_1 > x$, this means no Poisson event happens for N_0 and N_1 during that long period. Since N_2 is independent of both of

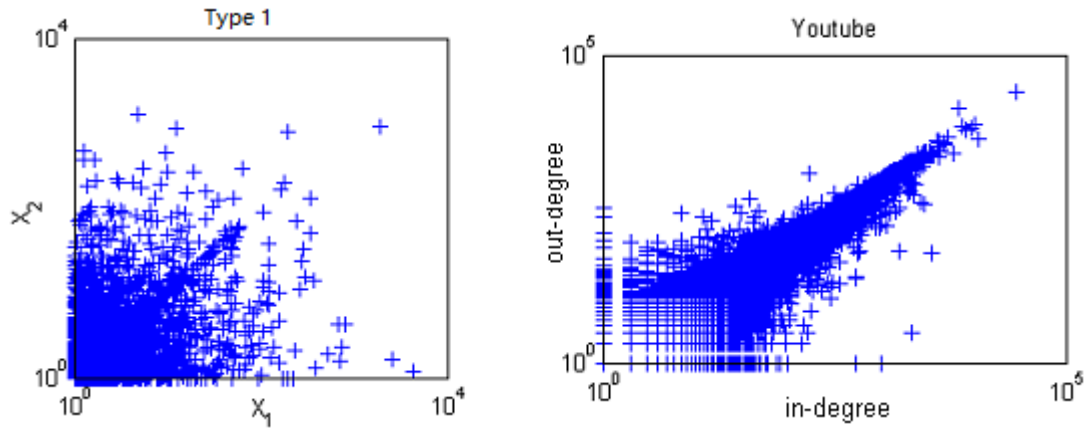


Figure 3.10. Comparing Type 1 model to Youtube dataset

N_0 and N_1, it requires N_2 never to occur during at least $\ln x$ life time to achieve $X_2 > x$. Recall that, the life time follows exponential distribution. As $x \to \infty$, the probability of N_2 never to occur goes to 0.

In the next two sections, we modify the model in (3.23) to produce a nonzero TDC. Even though it is not likely that we could change the shape of the scatter plot of Type 1 model with the following modifications, we are still interested in producing non-zero TDC for this model.

3.3.2 Modulated model with Markov on-off modulation

Define a Markov on-off process Y ($Y \in \{0, 1\}$)

$$dY = (1 - Y)dM_1 - YdM_2 \tag{3.31}$$

where M_1 and M_2 are independent Poisson counters with rate μ_1 and μ_2.

Our idea is to use this Markov on-off process to separate the active times of the independent Poisson counters (N_1 and N_2) and shared Poisson counter N_0. Our modified bivariate PCSDE model is

$$dX_i = X_i dt + (1 - X_i)\left((1 - Y)dN_0 + YdN_i\right), \ i = 1, 2$$

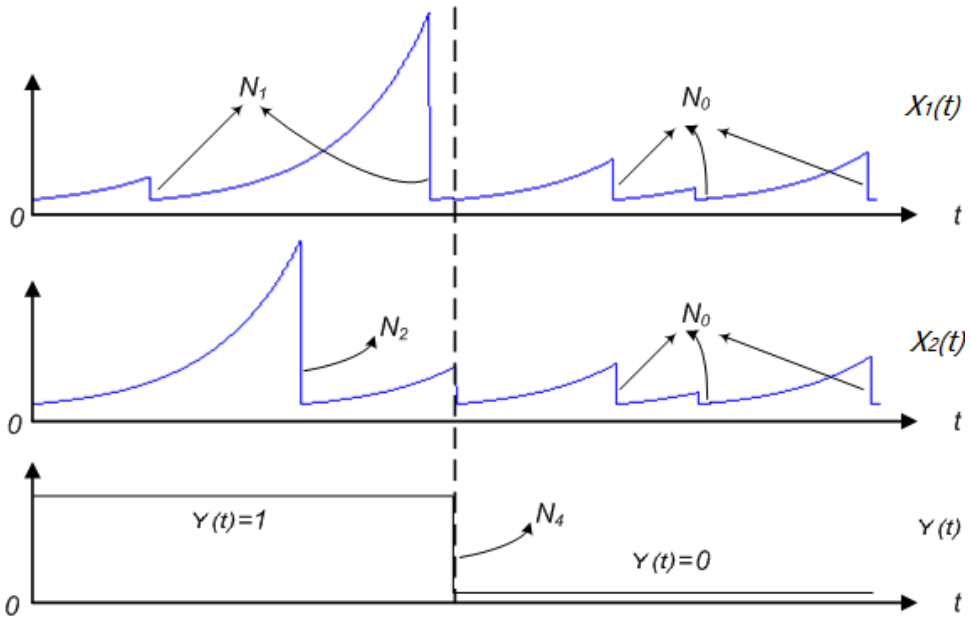


Figure 3.11. A sample path of the bivariate model with a shared Poisson counter and Markov on-off modulation

Thus, the two independent Poisson counters are active when the Markov on-off process is "on" and the shared Poisson counter N_0 is active when the Markov on-off process is "off". A sample path explains how this model works (as shown in Figure 3.11).

We solve this model using characteristic function as we did previous models. Define

$$\Phi(k_1, k_2, t) = \mathrm{E}[e^{j\sum_i k_i X_i(t)}], \;\; \Phi_i(k_i, t) = \mathrm{E}[e^{jk_i X_i(t)}],$$
$$\Psi(k_1, k_2, t) = \mathrm{E}[Y(t)e^{j\sum_i k_i X_i(t)}], \;\; \Psi_i(k_i, t) = \mathrm{E}[Y(t)e^{jk_i X_i(t)}],$$

and let $m(t) = \mathrm{E}[Y(t)]$. For the marginal, Ito's rule yields

$$\left(\frac{\partial}{\partial t} - k_i \frac{\partial}{\partial k_i}\right) H_i = -A_i H_i + b_i e^{jk_i}, \tag{3.32}$$

where

$$H_i = \begin{pmatrix} \Phi_i \\ \Psi_i \end{pmatrix}, \quad A_i = \begin{pmatrix} \lambda_0 & \lambda_i - \lambda_0 \\ -\mu_1 & \lambda_i + \mu_1 + \mu_2 \end{pmatrix},$$

and

$$b_i = \begin{pmatrix} \lambda_0(1 - m(\infty)) + \lambda_i m(t) \\ \lambda_i m(t) \end{pmatrix}.$$

Equation (3.32) can be solved as

$$H_i(k_i, t) = e^{-A_i t} H_i(k_i e^t, 0) + \int_0^t e^{-A_i(t-s)} b_i(s) e^{jk_i e^{t-s}} ds. \tag{3.33}$$

Execute a change in variables by letting $x_i = e^{t-s}$,

$$H_i(k_i, t) = e^{-A_i t} H_i(k_i e^t, 0) + \int_1^{e^t} e^{-A_i \log x_i} b_i(t - \log x_i) e^{jk_i x_i} x_i^{-1} dx_i. \tag{3.34}$$

Letting $t \to \infty$,

$$H_i(k_i, \infty) = \int_1^{\infty} e^{-A_i \log x_i} b_i(\infty) e^{jk_i x_i} x_i^{-1} dx_i. \tag{3.35}$$

Taking the inverse Fourier transform, the marginal steady-state density can be computed as

$$f_{X_i}(x_i) = a x_i^{-A_i} b_i(\infty) x_i^{-1}, \ x_i > 1 \tag{3.36}$$

where $a = (1, 0)$. Let $\gamma = (1, m(\infty))^T = A_i^{-1} b_i(\infty)$.

The marginal CCDF can be computed as

$$\begin{aligned}\bar{F}_{X_i}(x) &= \int_x^\infty a x_i^{-A_i} b_i(\infty) x_i^{-1} dx_i \\ &= a x^{-A_i} A_i^{-1} b_i(\infty), \\ &= a x^{-A_i} \gamma. \end{aligned} \tag{3.37}$$

For the joint case, Ito's rule yields

$$\left(\frac{\partial}{\partial t} - \sum_i k_i \frac{\partial}{\partial k_i} \right) H = -AH + b e^{j \sum_i k_i} + c\lambda_1 e^{jk_1} \Psi_2 + c\lambda_2 e^{jk_2} \Psi_1, \tag{3.38}$$

where

$$H = \begin{pmatrix} \Phi \\ \Psi \end{pmatrix}, \quad A = \begin{pmatrix} \lambda_0 & \sum_{i=1,2} \lambda_i - \lambda_0 \\ -\mu_1 & \sum_{i=1,2} (\lambda_i + \mu_i) \end{pmatrix},$$

and

$$b = \begin{pmatrix} 1 - m(t) \\ 0 \end{pmatrix} \lambda_0, \quad c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Equation (3.38) can be solved as

$$\begin{aligned} H(k_1, k_2, t) = e^{-At} H(k_1 e^t, k_2 e^t, 0) \\ + \int_0^t e^{-A(t-s)} B(s) e^{j \sum_i k_i e^{t-s}} ds \\ + \int_0^t e^{-A(t-s)} C\lambda_1 e^{jk_1 e^{t-s}} \Psi_2(k_2 e^{t-s}, s) ds \\ + \int_0^t e^{-A(t-s)} C\lambda_2 e^{jk_2 e^{t-s}} \Psi_1(k_1 e^{t-s}, s) ds. \end{aligned}$$

The solution to equation (3.38) at $t \to \infty$ is

$$\begin{aligned} & H(k_1, k_2, \infty) \\ &= \int_1^\infty dx e^{-A \log x} b(\infty) e^{j \sum_i k_i x} x^{-1} \\ &+ \int_1^\infty dx_1 e^{-A \log x_1} c\lambda_1 e^{jk_1 x_1} \Psi_2(k_2 x_1, \infty) x_1^{-1} \\ &+ \int_1^\infty dx_2 e^{-A \log x_2} c\lambda_2 e^{jk_2 x_2} \Psi_1(k_1 x_2, \infty) x_2^{-1}. \end{aligned} \tag{3.39}$$

Taking inverse Fourier transform, we have the joint density

$$\begin{aligned}
&f_{X_1,X_2}(x_1,x_2)\\
=&ax_1^{-A}b(\infty)x_1^{-1}u(x_1-1)\delta(x_1-x_2)\\
+&ax_1^{-A}c\lambda_1 f_{X_2}(x_2x_1^{-1})m(\infty)x_1^{-2}u(x_1-1)\\
+&ax_2^{-A}c\lambda_2 f_{X_1}(x_1x_2^{-1})m(\infty)x_2^{-2}u(x_2-1).
\end{aligned} \tag{3.40}$$

Let $\gamma=(1,m(\infty))^T=A^{-1}[b(\infty)+\lambda_1 cm(\infty)+\lambda_2 cm(\infty)]$, we have

$$\begin{aligned}
&\bar{F}_{X_1X_2}(x,x)\\
=&\int_x^\infty dx_1 ax_1^{-A}b(\infty)x_1^{-1}\\
+&\int_x^\infty dx_1 ax_1^{-A}c\lambda_1 x_1^{-1}\int_{x_1}^\infty dx_2 f_{X_2}(x_2x_1^{-1})m(\infty)x_1^{-1}\\
+&\int_x^\infty dx_2 ax_2^{-A}c\lambda_2 x_2^{-1}\int_{x_2}^\infty dx_1 f_{X_1}(x_1x_2^{-1})m(\infty)x_2^{-1},\\
=&ax^{-A}A^{-1}[b(\infty)+\lambda_1 cm(\infty)+\lambda_2 cm(\infty)],\\
=&ax^{-A}\gamma.
\end{aligned} \tag{3.41}$$

Let $\xi_\pm^i$ be the eigenvalues of A_i and $\xi_\pm$ be the eigenvalues of A, we have

$$\begin{aligned}
\xi_\pm^{(1)}=&\frac{\lambda_0+\lambda_1+\mu_1+\mu_2}{2}\\
&\pm\frac{\sqrt{(\lambda_1-\lambda_0+\mu_2-\mu_1)^2+4\mu_1\mu_2}}{2},
\end{aligned} \tag{3.42}$$

and

$$\begin{aligned}
\xi_\pm=&\frac{\lambda_0+\lambda_1+\lambda_2+\mu_1+\mu_2}{2}\\
&\pm\frac{\sqrt{(\lambda_1+\lambda_2-\lambda_0+\mu_2-\mu_1)^2+4\mu_1\mu_2}}{2}.
\end{aligned} \tag{3.43}$$

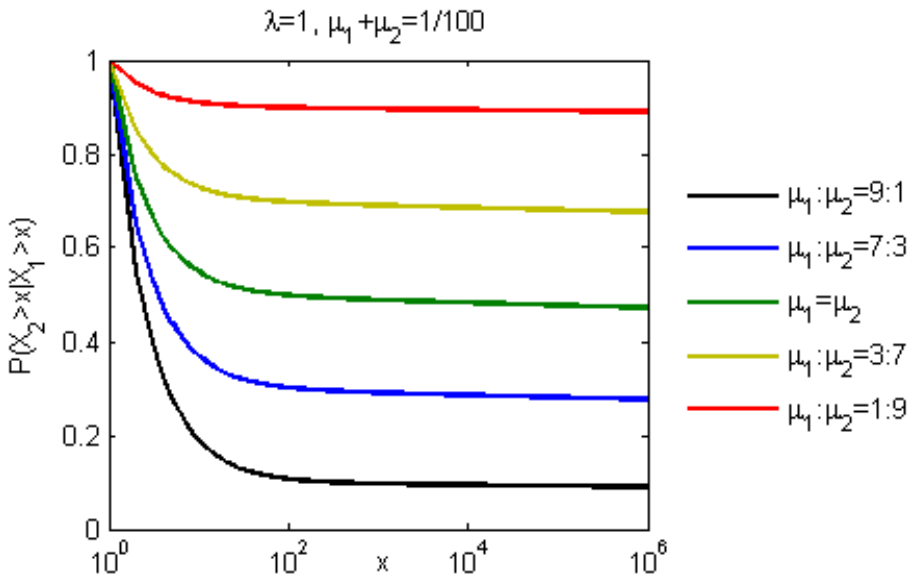


Figure 3.12. Theoretical TDC for a special case of the first modulated model

It is easy to check that $\xi_- - \xi_-^{(1)} > 0$, which implies

$$P(X_2 > x | X_1 > x) \sim C x^{-(\xi_- - \xi_-^{(1)})} \xrightarrow{x \to \infty} 0. \tag{3.44}$$

As indicated in (3.44), this model is still asymptotically independent. Consider the following special case. Let the arrival rates of Poisson counters N_0, N_1 and N_2 be the same, denoted as $\lambda_1 = \lambda_2 = \lambda_0 \triangleq \lambda$. When $\mu_1, \mu_2 \ll \lambda$, we have

$$P(X_2 > x | X_1 > x) \sim \frac{\mu_2}{\mu_1 + \mu_2} x^{-(\xi_- - \xi_-^{(1)})} \tag{3.45}$$

and $\xi_- - \xi_-^{(1)} < \mu_1$, where $\frac{\mu_2}{\mu_1 + \mu_2} = 1 - m(\infty)$. When μ_1 is small enough, the model can produce non-zero dependence coefficient over long decades under this special condition (as shown in Figure 3.12).

This model can also be solved using another method, as presented in Appendix A. It yields the same results as using characteristic function above.

3.3.3 Modulated model with Markov on-off modulation and synchronized reverting

In this subsection, we consider manually reverting the variables to their initial values as soon as the Markov on-off process changes its state. Thus, for any individual growth process between two successive reverting, the Markov on-off process is either in "on" or "off" state during the whole period. The new model is

$$dX_i = X_i dt + (1 - X_i)\left((1 - Y)(dN_0 + dM_1) + Y(dN_i + dM_2)\right), i = 1, 2. \tag{3.46}$$

Use the same method in Section 3.3.2, we have

$$\bar{F}_{X_i}(x) = ax^{-A_i}\gamma \quad \bar{F}_{X_1,X_2}(x, x) = ax^{-A}\gamma, \tag{3.47}$$

where in this model,

$$A_i = \begin{pmatrix} \lambda_0 + \mu_1 & \lambda_i - \lambda_0 + \mu_2 - \mu_1 \\ 0 & \lambda_i + \mu_2 \end{pmatrix},$$

and

$$A = \begin{pmatrix} \lambda_0 + \mu_1 & \sum_{i=1,2} \lambda_i - \lambda_0 + \mu_2 - \mu_1 \\ 0 & \sum_{i=1,2} \lambda_i + \mu_2 \end{pmatrix},$$

and a and γ are the same as in Section 3.3.2.

Let $\lambda_1 = \lambda_2 \triangleq \lambda$ and perform eigen-decompositions on A_i and A. The marginal and joint CCDF of this model become

$$\bar{F}_{X_i}(x) = x^{-(\lambda+\mu_2)}m(\infty) + x^{-(\lambda_0+\mu_1)}(1 - m(\infty)), \tag{3.48}$$

and

$$\bar{F}_{X_1,X_2}(x, x) = x^{-(2\lambda+\mu_2)}m(\infty) + x^{-(\lambda_0+\mu_1)}(1 - m(\infty)). \tag{3.49}$$

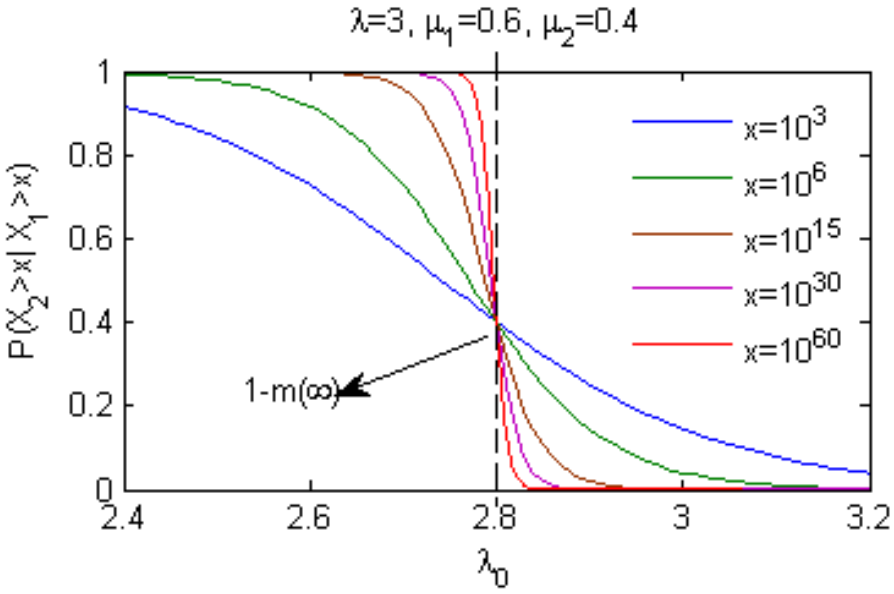


Figure 3.13. The dependence coefficient $P(X_1 > x|X_1 > x)$ of the second modulated model as a function of λ_0 with different x values

Denote $\Delta\mu = \mu_1 - \mu_2$. The tail dependence coefficient of this model is

$$\lim_{x\to\infty} P(X_2 > x|X_1 > x) = \begin{cases} 1 & \lambda > \lambda_0 + \Delta\mu, \\ \frac{\mu_2}{\mu_1+\mu_2} & \lambda = \lambda_0 + \Delta\mu, \\ 0 & \lambda < \lambda_0 + \Delta\mu. \end{cases} \tag{3.50}$$

This modulated model successfully generates nonzero TDCs. However, it seems that the case when fractional TDC appears is unstable. When a parameter is perturbed, the coefficient goes to one or zero. As shown in Figure 3.13, when x is finite, the dependence coefficient $\mathbb{P}(X_2 > X|X_1 > x)$ is continuous with the increasing of λ_0 (with the other parameters fixed). As x goes to infinity, the TDC becomes discontinuous at a critical point.

3.3.4 Summary of Type 1 models

In Section 3.3, we proposed a PCSDE model with a shared Poisson counter and two modulaitons are made to the original model to produce fractional TDC. As we observe, even though the second modulation generates non-zero TDC, it is not the way real datasets do. The synthetic data of this type of models look different to real datasets too.

We try to propose a model which can generate fractional TDC in a more natural way. This thought leads to a new type of model in the following.

3.4 Type 2: bivariate PCSDE model with coupled growth

In the previous section, we presented the Type 1 model with a shared Poisson counter and its modulations. This type of model does not succeed in either producing non-zero TDC, or fitting real datasets. Reconsidering how to introduce dependence among variables X_1 and X_2, we propos a new model, where dependence is introduced by using **Coupled Growth**, instead of a shared Poisson counter. We call this the **Type 2** model.

3.4.1 Model formulation

The model is formulated as

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} 1 - X_1(t) \\ 0 \end{pmatrix} dN_1(t) + \begin{pmatrix} 0 \\ 1 - X_2(t) \end{pmatrix} dN_2(t). \tag{3.51}$$

In this model, the life time of the two variables are controlled by two independent Poisson counters N_1 and N_2; however, the growth part of the two variables are coupled. The increment of X_1 (X_2) is also related to the current value of X_2 (X_1).

In the following computations, we set the Poisson rates of the independent Poisson counters N_1 and N_2 equal to make the computation easier.

3.4.2 Marginal tail

3.4.2.1 Tail exponent

In this part, we will prove the marginal tail of this model follows power law distribution. We write the matrix

$$\beta_M = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix},$$

$$\lambda_1 = \lambda_2 = \lambda.$$

Note that the DE

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \beta_M \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt \tag{3.52}$$

has the solution

$$\begin{aligned} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} &= e^{t\beta_M} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix}, \\ &= \frac{1}{2} e^{t(1+\beta)} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix} \\ &\quad + \frac{1}{2} e^{t(1-\beta)} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix}. \end{aligned} \tag{3.53}$$

Let X_n be the value of $X_1(t)$ at the n^{th} arrival of the Poisson process N_2. By the PASTA property [92], the stationary distribution of (X_n) is the stationary distribution of $(X_1(t))$. We will prove that (X_n) satisfies a stochastic recursion

$$X_{n+1} = A_{n+1} X_n + B_{n+1},\ n = 1, 2, \ldots \tag{3.54}$$

where $(A_n, B_n),\ n = 2, 3, \ldots$ are i.i.d. and (A_{n+1}, B_{n+1}) is independent of X_n in (3.54). Based on the results in [48, 33, 88], once (3.54) has been proved, we know that for a stationary random variable X such that

$$X \stackrel{d}{=} AX + B \tag{3.55}$$

(the stationary distribution of (X_n)) we have

$$\mathbb{P}(X > x) \sim x^{-\alpha}, \tag{3.56}$$

where $\alpha > 0$ is such that

$$EA^{\alpha} = 1. \tag{3.57}$$

We now prove (3.54). Let $K = 0, 1, 2, \ldots$ be the number of arrivals of N_1 in the interval $(0, E)$, where E is the first arrival of N_2,

$$E \sim \exp(\lambda).$$

Let $0 < T_1 < T_2 < \ldots < T_K < E$ be the arrival times of N_1. Let $Y_{n,j}$ be the state of $X_2(t)$ at $t = T_j, j = 1, \ldots, K$. Note that, at time 0,

$$X_1(0) = X_n,\ X_2(0) = 1.$$

An illustration of the above state update process is shown in Figure 3.14.

We conclude by (3.53),

$$\begin{aligned} Y_{n,1} &= \frac{1}{2}e^{T_1(1+\beta)}(X_n + 1) + \frac{1}{2}e^{T_1(1-\beta)}(-X_n + 1) \\ &= X_n\left(\frac{1}{2}e^{T_1(1+\beta)} - \frac{1}{2}e^{T_1(1-\beta)}\right) + \frac{1}{2}e^{T_1(1+\beta)} + \frac{1}{2}e^{T_1(1-\beta)}. \end{aligned} \tag{3.58}$$

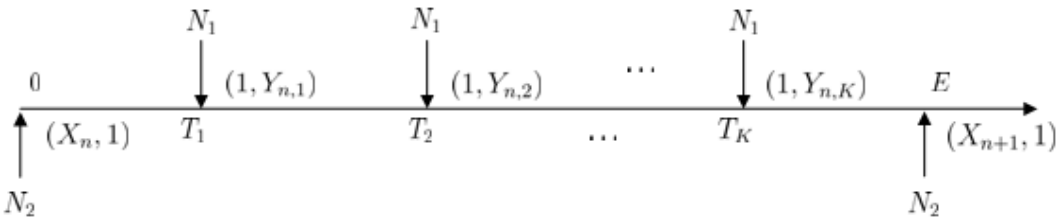


Figure 3.14. The state update process of Type 2 model

Similarly, for $j = 2, \ldots, K$,

$$Y_{n,j} = Y_{n,j-1}\left(\frac{1}{2}e^{(T_j - T_{j-1})(1+\beta)} + \frac{1}{2}e^{(T_j - T_{j-1})(1-\beta)}\right) + \left(\frac{1}{2}e^{(T_j - T_{j-1})(1+\beta)} - \frac{1}{2}e^{(T_j - T_{j-1})(1-\beta)}\right). \tag{3.59}$$

We conclude that

$$\begin{aligned} Y_{n,K} = X_n &\frac{e^{T_1(1+\beta)} - e^{T_1(1-\beta)}}{2} \prod_{j=2}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2} \\ &+ \prod_{j=1}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2} \\ &+ \sum_{i=2}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} - e^{(T_j - T_{j-1})(1-\beta)}}{2} \prod_{j=i+1}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2}. \end{aligned} \tag{3.60}$$

Finally, we use (3.53) once again:

$$\begin{aligned} X_{n+1} &= \frac{1}{2}e^{(E-T_K)(1+\beta)}(1 + Y_{n,K}) + \frac{1}{2}e^{(E-T_K)(1-\beta)}(1 - Y_{n,K}), \\ &= Y_{n,K}\left(\frac{e^{(E-T_K)(1+\beta)} - e^{(E-T_K)(1-\beta)}}{2}\right) + \frac{e^{(E-T_K)(1+\beta)} + e^{(E-T_K)(1-\beta)}}{2}. \end{aligned} \tag{3.61}$$

Combining (3.60) and (3.61) yields (3.54) with

$$A = \frac{e^{T_1(1+\beta)} - e^{T_1(1-\beta)}}{2} \cdot \prod_{j=2}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2} \tag{3.62}$$

$$\cdot \frac{e^{(E-T_K)(1+\beta)} - e^{(E-T_K)(1-\beta)}}{2}. \tag{3.63}$$

There is a more convenient way of representing A in (3.62). Let (F_j) be i.i.d $\exp(2\lambda)$ random variables, independent of a $Geom(1/2)$ random variable N. The proof of the distribution of F_j and N is given in Appendix B. Then

if $N = 0$,

$$A = \frac{e^{F_1(1+\beta)} + e^{F_1(1-\beta)}}{2}; \tag{3.64}$$

if $N \geq 1$, then

$$A = \frac{e^{F_1(1+\beta)} - e^{F_1(1-\beta)}}{2} \cdot \frac{e^{F_2(1+\beta)} - e^{F_2(1-\beta)}}{2} \cdot \prod_{j=3}^{N+1} \frac{e^{F_j(1+\beta)} + e^{F_j(1-\beta)}}{2}. \tag{3.65}$$

Now we need to find $\alpha > 0$ that (3.57) holds.

We write

$$EA^\alpha = \frac{1}{2} E\left(\frac{e^{F_1(1+\beta)} + e^{F_1(1-\beta)}}{2}\right)^\alpha + \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} \left(E\left(\frac{e^{F_1(1+\beta)} - e^{F_1(1-\beta)}}{2}\right)^\alpha\right)^2 \left(E\left(\frac{e^{F_1(1+\beta)} + e^{F_1(1-\beta)}}{2}\right)^\alpha\right)^{n-1}. \tag{3.66}$$

Let $E\left(\frac{e^{F_1(1+\beta)} + e^{F_1(1-\beta)}}{2}\right)^\alpha = I_1$ and let $E\left(\frac{e^{F_1(1+\beta)} - e^{F_1(1-\beta)}}{2}\right)^\alpha = I_2$, we have

$$\begin{aligned} EA^\alpha =& \frac{1}{2} I_1 + I_2^2 \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} I_1^{n-1}, \\ =& \frac{1}{2} I_1 + I_2^2 \frac{\frac{1}{4}}{1 - \frac{1}{2} I_1}, \\ =& \frac{1}{2} I_1 + I_2^2 \frac{1}{4 - 2I_1}. \end{aligned} \tag{3.67}$$

where

$$I_1 = 2^{-\alpha} \int_0^\infty 2\lambda e^{-2\lambda x} \left(e^{x(1+\beta)} + e^{x(1-\beta)}\right)^\alpha dx, \tag{3.68}$$

and

$$I_2 = 2^{-\alpha} \int_0^\infty 2\lambda e^{-2\lambda x} \left(e^{x(1+\beta)} - e^{x(1-\beta)}\right)^\alpha dx. \tag{3.69}$$

Let $y = e^{-x}$, we have

$$I_1 = 2^{-\alpha} \int_0^1 2\lambda y^{2\lambda-1} \left(y^{-(1+\beta)} + y^{-(1-\beta)}\right)^\alpha dy, \tag{3.70}$$

and

$$I_2 = 2^{-\alpha} \int_0^1 2\lambda y^{2\lambda-1} \left(y^{-(1+\beta)} - y^{-(1-\beta)}\right)^\alpha dy. \tag{3.71}$$

First compute I_2,

$$\begin{aligned} I_2 &= 2\lambda \cdot 2^{-\alpha} \int_0^1 y^{2\lambda-1} y^{-\alpha(1+\beta)} (1-y^{2\beta})^\alpha dy, \\ &= 2\lambda \cdot 2^{-\alpha} \int_0^1 y^{2\lambda-\alpha(1+\beta)-1} (1-y^{2\beta})^\alpha dy. \end{aligned}$$

Let $z = y^{2\beta}$, so $y = z^{1/(2\beta)}$, so

$$\begin{aligned} I_2 &= 2\lambda \cdot 2^{-\alpha} \int_0^1 z^{\frac{2\lambda-\alpha(1+\beta)-1}{2\beta}} (1-z)^\alpha \cdot \frac{1}{2\beta} z^{1/(2\beta)-1} dz, \\ &= \frac{\lambda 2^{-\alpha}}{\beta} \int_0^1 z^{\frac{2\lambda-\alpha(1+\beta)}{2\beta}-1} (1-z)^\alpha dz, \\ &= \frac{\lambda 2^{-\alpha}}{\beta} B\left(\frac{2\lambda-\alpha(1+\beta)}{2\beta}, \alpha+1\right), \end{aligned} \tag{3.72}$$

where $B(x, y)$ is the beta function.

Using the same method as in computing I_2, we have

$$I_1 = \frac{\lambda 2^{-\alpha}}{\beta} \int_0^1 z^{\frac{2\lambda-\alpha(1+\beta)}{2\beta}-1}(1+z)^{\alpha} dz. \tag{3.73}$$

We cannot get a closed form solution for I_1 similar to that of I_2. So we estimate I_1 using numerical integration. Consider the following 'rectangle rule'. Let $N > 0$ be a sufficient large integer,

$$\int_0^1 f(x)dx \approx \frac{1}{N} \sum_{n=0}^{N-1} f(\frac{n}{N}). \tag{3.74}$$

Then

$$I_1 \approx \frac{\lambda 2^{-\alpha}}{\beta} \frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{n}{N}\right)^{\frac{2\lambda-\alpha(1+\beta)}{2\beta}-1} \left(1+\frac{n}{N}\right)^{\alpha}. \tag{3.75}$$

3.4.2.2 Necessary condition

Equation $E(A^{\alpha}) = 1$ in (3.57) does not always have a solution for different values of β. Let $h(\alpha) = EA^{\alpha} - 1$, it is easy to prove by the definition of A that

$$h(0) = 0, \quad h''(\alpha) > 0.$$

A necessary condition for function $h(\alpha) = 0$ having a solution other than 0 is $h'(0) < 0$, which gives $E(\log A) < 0$.

Figure 3.15 gives us an illustration. As shown in Figure 3.15, $h(0) = 0$ and $h''(\alpha) > 0$. Only when $h'(0) < 0$, does the function $h(\alpha)$ cross the "0 line" again, which gives $h(\alpha) = 0$ a positive solution. If $h'(0) > 0$, the function will start at 0, increase with α and never return to 0 again.

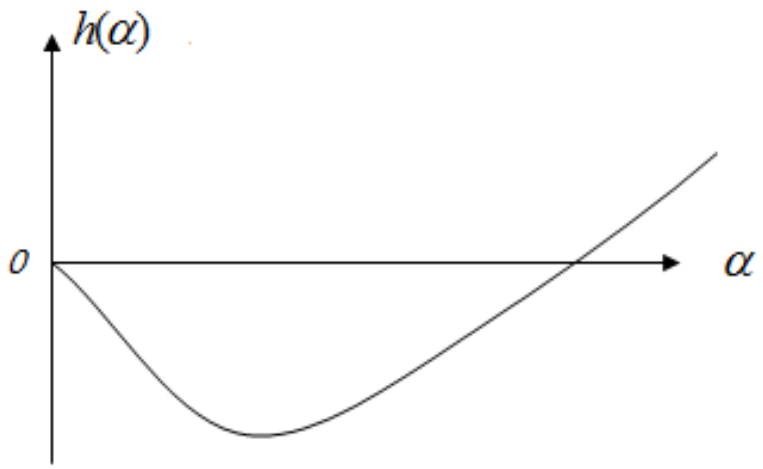


Figure 3.15. $h(\alpha) - 1$ as a function of α

Write

$$
\begin{aligned}
&E(\log A)\\
=&\frac{1}{2}E\left(\log\frac{e^{F_1(1+\beta)}+e^{F_1(1-\beta)}}{2}\right)\\
+&\sum_{n=1}^{\infty}\frac{1}{2^{n+1}}\left(2E\left(\log\frac{e^{F_1(1+\beta)}-e^{F_1(1-\beta)}}{2}\right)+(n-1)E\left(\log\frac{e^{F_1(1+\beta)}+e^{F_1(1-\beta)}}{2}\right)\right).
\end{aligned}
\tag{3.76}
$$

Let $J_1 = E\left(\log\frac{e^{F_1(1+\beta)}+e^{F_1(1-\beta)}}{2}\right)$ and $J_2 = E\left(\log\frac{e^{F_1(1+\beta)}-e^{F_1(1-\beta)}}{2}\right)$. Then we have

$$
\begin{aligned}
E(\log A) =&\frac{1}{2}J_1+\sum_{n=1}^{\infty}\frac{1}{2^{n+1}}(2J_2+(n-1)J_1),\\
=&\frac{1}{2}J_1+\sum_{n=1}^{\infty}\frac{1}{2^n}J_2+\sum_{n=1}^{\infty}\frac{n-1}{2^{n+1}}J_1,\\
=&\frac{1}{2}J_1+\frac{\frac{1}{2}}{1-\frac{1}{2}}J_2+\frac{\frac{1}{2^3}(\frac{1}{1-\frac{1}{2}})}{1-\frac{1}{2}}J_1,\\
=&J_1+J_2,
\end{aligned}
\tag{3.77}
$$

where

$$J_1 = \int_0^\infty 2\lambda e^{-2\lambda x} \log \frac{e^{x(1+\beta)} + e^{x(1-\beta)}}{2} dx,$$

and

$$J_2 = \int_0^\infty 2\lambda e^{-2\lambda x} \log \frac{e^{x(1+\beta)} - e^{x(1-\beta)}}{2} dx.$$

Let $y = e^{-x}$, we have

$$J_1 = \int_0^1 2\lambda y^{2\lambda-1} \log \frac{y^{-(1+\beta)} + y^{-(1-\beta)}}{2} dy, \tag{3.78}$$

and

$$J_2 = \int_0^1 2\lambda y^{2\lambda-1} \log \frac{y^{-(1+\beta)} - y^{-(1-\beta)}}{2} dy. \tag{3.79}$$

With equations (3.78) and (3.79), we can also use the rectangle rule to perform the numerical integration. Then using Equation (3.77), we can estimate $E(\log A)$.

We will give the numerical results for the exponent α and $E(\log A)$ as a function of β at the experiment part.

3.4.3 Joint tail

Consider our system in steady state. Let Y be a random variable with the stationary distribution of the value of $X_1(t)$ at the moment when the counter N_2 has an arrival. Consider the combined counter $N_1 \cup N_2$. Its points are $W_1, W_2, \ldots$, with $(W_{n+1} - W_n)$ i.i.d, $\exp(2\lambda)$.

The state of the system at these points is a Markov chain, with state space

$$(\{1\} \times [1, \infty) \vee ([1, \infty) \times \{1\}).$$

The stationary distribution has the form

$$\begin{pmatrix} 1 \\ Y \end{pmatrix} \quad w.p. \quad \frac{1}{2},$$
$$\begin{pmatrix} Y \\ 1 \end{pmatrix} \quad w.p. \quad \frac{1}{2}. \tag{3.80}$$

The stationary distribution of the continuous-time process can be computed by

$$\begin{aligned}
&\mathbb{P}_{st}((X_1, X_2) \in A) \\
=&\frac{E\int_{W_n}^{W_{n+1}} \mathbf{1}\left((X_1(t), X_2(t)) \in A\right) dt}{E(W_{n+1} - W_n)}, \\
=&\frac{E\int_0^{\infty} 2\lambda e^{-2\lambda u} du \int_0^u \mathbf{1}\left((X_1(t), X_2(t)) \in A\right) dt}{\frac{1}{2\lambda}}, \\
=&4\lambda^2 \int_1^{\infty} e^{-2\lambda u} du \times \frac{1}{2} \int_1^{\infty} F_Y(dy) \\
&\times \int_0^u [\mathbf{1}((y\frac{e^{t(1+\beta)} - e^{t(1-\beta)}}{2} + \frac{e^{t(1+\beta)} + e^{t(1-\beta)}}{2}, y\frac{e^{t(1+\beta)} + e^{t(1-\beta)}}{2} + \frac{e^{t(1+\beta)} - e^{t(1-\beta)}}{2}) \in A) \\
&+\mathbf{1}((y\frac{e^{t(1+\beta)} + e^{t(1-\beta)}}{2} + \frac{e^{t(1+\beta)} - e^{t(1-\beta)}}{2}, y\frac{e^{t(1+\beta)} - e^{t(1-\beta)}}{2} + \frac{e^{t(1+\beta)} + e^{t(1-\beta)}}{2}) \in A)]dt.
\end{aligned} \tag{3.81}$$

Therefore, in the stationary regime

$$(X_1, X_2) \stackrel{d}{=} \begin{cases} (YV + W, YW + V), & w.p. \quad \frac{1}{2}, \\ (YW + V, YV + W), & w.p. \quad \frac{1}{2}, \end{cases} \tag{3.82}$$

where (V, W) is a random vector independent of Y, with the following distribution. Let $T \sim \exp(2\lambda)$. Given $T = t$, let $u \sim U(0, t)$. Then

$$\begin{aligned}
V &= \frac{e^{u(1+\beta)} - e^{u(1-\beta)}}{2}, \\
W &= \frac{e^{u(1+\beta)} + e^{u(1-\beta)}}{2}.
\end{aligned} \tag{3.83}$$

Before continuing, we state Breiman's lemma [10, 23]:

Breiman's Lemma

Suppose that X and Y are two independent nonnegative random variables such that $\mathbb{P}\{X > x\}$ is regularly varying of index $-\alpha$, $\alpha \geq 0$, and $\mathbb{E}\{Y^{\alpha+\epsilon}\} < \infty$ for some $\epsilon > 0$. Then

$$\mathbb{P}\{XY > x\} \sim \mathbb{E}\{Y^{\alpha}\}\mathbb{P}\{X > x\} \tag{3.84}$$

as $x \to \infty$.

With the lemma, we have

$$\begin{aligned}
&\frac{\mathbb{P}(\|(X_1, X_2)\| > x, \frac{(X_1,X_2)}{\|(X_1,X_2)\|} \in A)}{\mathbb{P}(\|(X_1, X_2)\| > x)} \\
=&\frac{1}{2}\frac{\mathbb{P}(Y\|V, W\| > x, \frac{(V,W)}{\|(V,W)\|} \in A)}{\mathbb{P}(Y > x)}\frac{\mathbb{P}(Y > x)}{\mathbb{P}(Y\|(V, W)\| > x)} \\
+&\frac{1}{2}\frac{\mathbb{P}(Y\|W, V\| > x, \frac{(W,V)}{\|(W,V)\|} \in A)}{\mathbb{P}(Y > x)}\frac{\mathbb{P}(Y > x)}{\mathbb{P}(Y\|(W, V)\| > x)}, \\
\xrightarrow{t\to\infty}&\frac{1}{2}E\left[\|(V, W)\|^{\alpha}\mathbf{1}\left(\frac{(V, W)}{\|(V, W)\|} \in A\right)\right]\frac{1}{E\|(V, W)\|^{\alpha}} \\
+&\frac{1}{2}E\left[\|(W, V)\|^{\alpha}\mathbf{1}\left(\frac{(W, V)}{\|(W, V)\|} \in A\right)\right]\frac{1}{E\|(W, V)\|^{\alpha}}.
\end{aligned} \tag{3.85}$$

The right hand side of (3.85) is the spectral measure.

In particular, for $x_1, x_2 > 0$,

$$
\begin{aligned}
&\frac{\mathbb{P}(X_1 > tx_1, X_2 > tx_2)}{\mathbb{P}(Y > t)} \\
=&\frac{1}{2}\frac{\mathbb{P}(YV > tx_1, YW > tx_2)}{\mathbb{P}(Y > t)} + \frac{1}{2}\frac{\mathbb{P}(YW > tx_1, YV > tx_2)}{\mathbb{P}(Y > t)}, \\
=&\frac{1}{2}\frac{\mathbb{P}(Y > t\frac{x_1}{V}, Y > t\frac{x_2}{W})}{\mathbb{P}(Y > t)} + \frac{1}{2}\frac{\mathbb{P}(Y > t\frac{x_1}{W}, Y > t\frac{x_2}{V})}{\mathbb{P}(Y > t)}, \\
=&\frac{1}{2}\frac{\mathbb{P}(Y > \max(\frac{x_1}{V}, \frac{x_2}{W})t)}{\mathbb{P}(Y > t)} + \frac{1}{2}\frac{\mathbb{P}(Y > \max(\frac{x_1}{W}, \frac{x_2}{V})t)}{\mathbb{P}(Y > t)}, \\
=&\frac{1}{2}\frac{\mathbb{P}(Y \min(\frac{V}{x_1}, \frac{W}{x_2}) > t)}{\mathbb{P}(Y > t)} + \frac{1}{2}\frac{\mathbb{P}(Y \min(\frac{W}{x_1}, \frac{V}{x_2}) > t)}{\mathbb{P}(Y > t)}, \\
\xrightarrow{t\to\infty}&\frac{1}{2}E\min\left(\frac{V}{x_1}, \frac{W}{x_2}\right)^{\alpha} + \frac{1}{2}E\min\left(\frac{W}{x_1}, \frac{V}{x_2}\right)^{\alpha}.
\end{aligned} \tag{3.86}
$$

Then the conditional probability becomes

$$
\begin{aligned}
\mathbb{P}(X_2 > tx_2 | X_1 > tx_1) =& \frac{\mathbb{P}(X_1 > tx_1, X_2 > tx_2)}{\mathbb{P}(X_1 > tx_1)}, \\
=& \frac{\mathbb{P}(X_1 > tx_1, X_2 > tx_2)}{\mathbb{P}(Y > t)} \frac{\mathbb{P}(Y > t)}{\mathbb{P}(X_1 > tx_1)}, \\
\xrightarrow{t\to\infty}& \frac{E\min\left(\frac{V}{x_1}, \frac{W}{x_2}\right)^{\alpha} + E\min\left(\frac{W}{x_1}, \frac{V}{x_2}\right)^{\alpha}}{E\left(\frac{V}{x_1}\right)^{\alpha} + E\left(\frac{W}{x_1}\right)^{\alpha}}.
\end{aligned} \tag{3.87}
$$

Let $x_1 = x_2 = 1$, we have

$$
\mathbb{P}(X_2 > t | X_1 > t) \xrightarrow{t\to\infty} \frac{2E\min(V, W)^{\alpha}}{EV^{\alpha} + EW^{\alpha}} = \frac{2EV^{\alpha}}{EV^{\alpha} + EW^{\alpha}}, \tag{3.88}
$$

where

$$
EV^{\alpha} = \int_0^{\infty} 2\lambda e^{-2\lambda t} dt \int_0^t \frac{1}{t} \left(\frac{e^{u(1+\beta)} - e^{u(1-\beta)}}{2}\right)^{\alpha} du, \tag{3.89}
$$

and

$$
EW^{\alpha} = \int_0^{\infty} 2\lambda e^{-2\lambda t} dt \int_0^t \frac{1}{t} \left(\frac{e^{u(1+\beta)} + e^{u(1-\beta)}}{2}\right)^{\alpha} du. \tag{3.90}
$$

Let $s = e^{-t}$ and $v = e^{-u}$, we have

$$
EV^{\alpha} = 2^{-\alpha} \cdot 2\lambda \int_0^1 s^{2\lambda - 1} \frac{1}{-\ln s} ds \int_s^1 \frac{1}{v} \left(v^{-(1+\beta)} - v^{-(1-\beta)}\right)^{\alpha} dv, \tag{3.91}
$$

and

$$EW^{\alpha} = 2^{-\alpha} \cdot 2\lambda \int_0^1 s^{2\lambda-1} \frac{1}{-\ln s} ds \int_s^1 \frac{1}{v} \left(v^{-(1+\beta)} + v^{-(1-\beta)}\right)^{\alpha} dv. \tag{3.92}$$

Based on the definition of TDC, the TDC of this model can be computed using Equation (3.88), where EV^{α} and EW^{α} are both computed using numerical method introduced before with Equations (3.91) and (3.92).

Another method to compute EV^{α} and EW^{α} is using Equation (3.83). Given β, we can generate samples $(V_1, V_2, \ldots, V_N)$ and $(W_1, W_2, \ldots, W_N)$ by generating sample data of $(u_1, u_2, \ldots, u_N)$ in Equation (3.83). Then EV^{α} and EW^{α} can be computed by using the sample mean method:

$$EV^{\alpha}_{est} = \frac{1}{N} \sum_{n=1}^{N} V_n^{\alpha}, \quad EW^{\alpha}_{est} = \frac{1}{N} \sum_{n=1}^{N} W_n^{\alpha}. \tag{3.93}$$

In the experiment part, we use the sample mean method to compute TDC, since this method is more straightforward and runs fast.

3.4.4 Experimental results

Here we first present some numerical results of our Type 2 model using Matlab. Then we compare synthetic data generated by Type 2 model to real social network datasets.

3.4.4.1 Numerical results

In this part, we present our experimental results based on the theoretical analysis of the previous subsections. First let $\lambda = 1/4$, $1/2$, 1, and 2 separately. By increasing β, we compute the corresponding exponent value α and TDC. $E(\log A)$ is also computed to verify the necessary condition in subsection 3.4.2.2.

As shown in Figure 3.16, when $\beta = 0$, the two processes are independent, which yields $\alpha = \lambda$. When β increases, the exponent α decreases, which means the tail

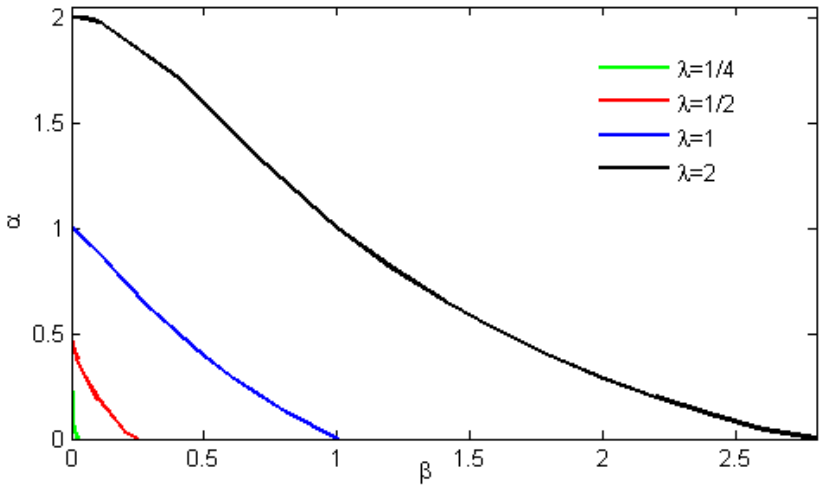


Figure 3.16. Exponent α as a function of β with different λ values

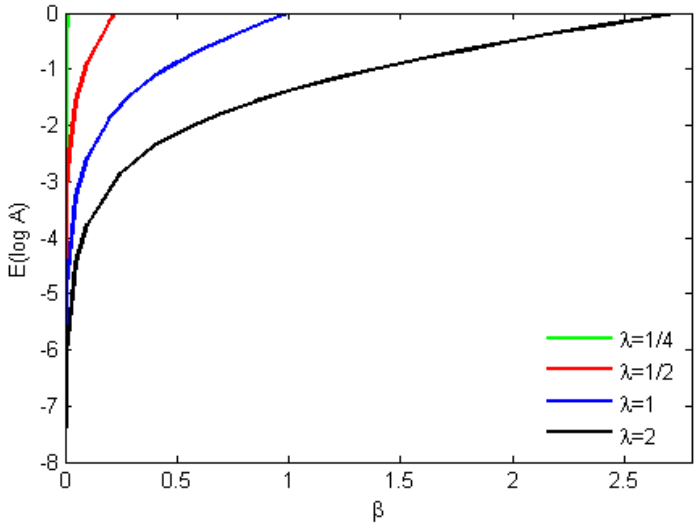


Figure 3.17. $E(\log A)$ as a function of β with different λ values

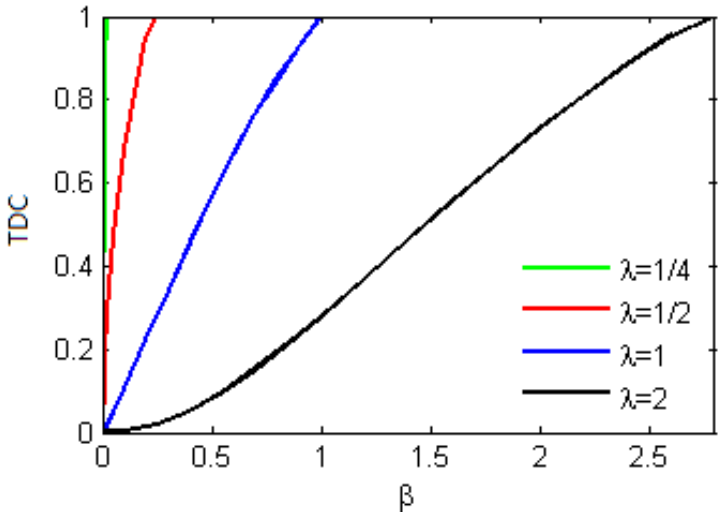


Figure 3.18. TDC as function of β with different λ values

becomes heavier. Exponent α reaches 0 at some point, and this point varies for different values of λ.

As discussed in Subsection 3.4.2.2, we obtain a positive solution to Equation (3.57) only when $E(\log A) < 0$. Based on the results in Fig 3.17, we observe that $E(\log A)$ is negative when β is close to 0 (negative infinity when $\beta = 0$). $E(\log A)$ increases with β and eventually reaches 0. The point varies for different values of λ and is the same as the point when exponent α reaches 0 in Figure 3.16.

When $\beta = 0$, the two processes are independent, which yields $TDC = 0$. As β increases, TDC increases. TDC reaches one at some point which corresponds to the point when α reaches zero and $E(\log A)$ reaches zero.

3.4.4.2 Comparing Type 2 to real datasets

To obtain a better picture of the joint distribution of Type 2 model, we generate samples pairs of (X_1, X_2) using Equation (3.82), and draw the scatter plot and CCDF of the samples. With $\lambda = 2$, $\beta = 0.2$, we have $\alpha = 1.9203$. $100,000$ samples are generated and the scatter plot and CCDF are shown in Figure 3.19. We compare the scatter plot of Type 2 datasets to the real data in Youtube in Figure 3.20. We

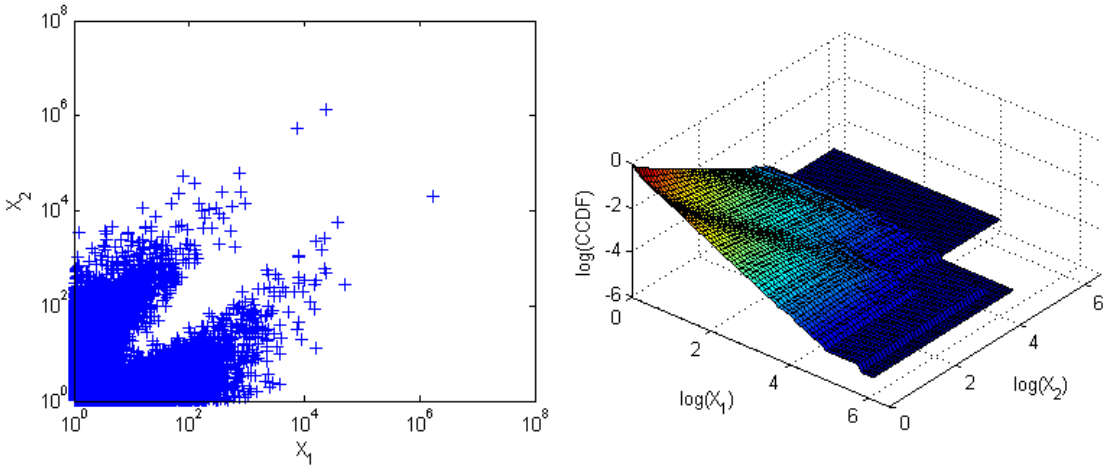


Figure 3.19. Synthetic data from Type 2 model and CCDF ($\lambda = 2, \beta = 0.2$)

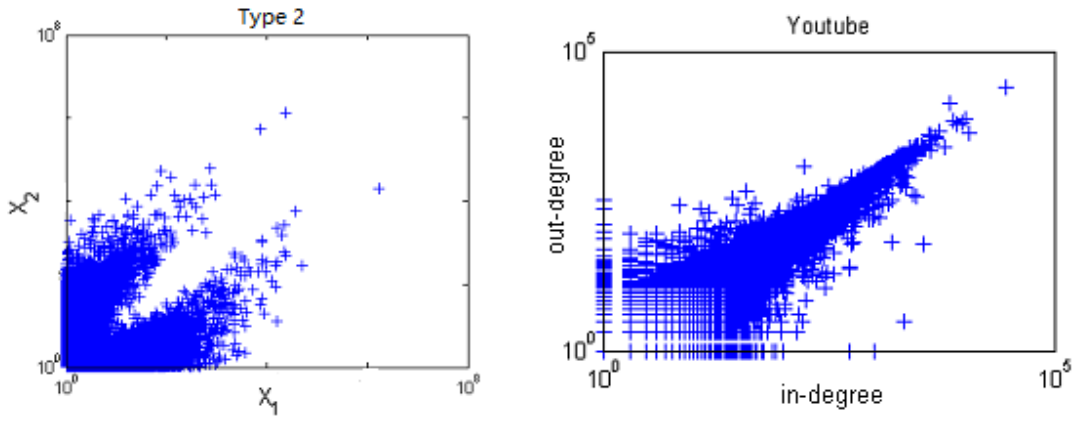


Figure 3.20. Comparing Type 2 model to Youtube dataset

observe from the figure that the scatter plot and CCDF of the synthetic data has a weird shape which actually not present in bivariate power law data. The data is not present along the equal line, but totally on the opposite way.

3.4.5 A generalized model

3.4.5.1 Model formulation

Consider a more general form of the model where the two processes are not symmetric

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} 1 - X_1(t) \\ 0 \end{pmatrix} dN_1(t) + \begin{pmatrix} 0 \\ 1 - X_2(t) \end{pmatrix} dN_2(t), \tag{3.94}$$

with $\beta_1, \beta_2 > 0$.

3.4.5.2 Theoretical results

As we did in our previous section, we write the matrix

$$\beta_M = \begin{pmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{pmatrix},$$

$$\lambda_1 = \lambda_2 = \lambda.$$

For the DE

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \beta_M \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt, \tag{3.95}$$

we have the solution

$$\begin{aligned} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} &= e^{t\beta_M} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix}, \\ &= \frac{1}{2} e^{t(1+\sqrt{\beta_1\beta_2})} \begin{pmatrix} 1 & \sqrt{\frac{\beta_1}{\beta_2}} \\ \sqrt{\frac{\beta_2}{\beta_1}} & 1 \end{pmatrix} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix} \\ &+ \frac{1}{2} e^{t(1-\sqrt{\beta_1\beta_2})} \begin{pmatrix} 1 & -\sqrt{\frac{\beta_1}{\beta_2}} \\ -\sqrt{\frac{\beta_2}{\beta_1}} & 1 \end{pmatrix} \begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix}. \end{aligned} \tag{3.96}$$

Using the same method as in the case when $\beta_1 = \beta_2 = \beta$, we prove that X_1 and X_2 both satisfy the stochastic recursion in (3.54), where A is the same as in Equation

(3.64) and Equation (3.65). The only difference is that we need to let $\beta = \sqrt{\beta_1\beta_2}$. The proof of this is in Appendix C. The marginal exponent α is computed by solving the Equation $EA^\alpha = 1$, as we did for the symmetric model.

For the joint distribution, given W and V the same as in Equation (3.83), let $V_1 = \sqrt{\frac{\beta_1}{\beta_2}}V$ and $V_2 = \sqrt{\frac{\beta_2}{\beta_1}}V$, we have

$$(X_1, X_2) \stackrel{d}{=} \begin{cases} (XV_1 + W, XW + V_2) & w.p.\ \frac{1}{2}, \\ (XW + V_1, XV_2 + W) & w.p.\ \frac{1}{2}, \end{cases} \tag{3.97}$$

and we can compute the conditional probability of this model as follows,

$$\begin{aligned} &\lim_{x\to\infty} \frac{P(X_1 > t, X_2 > t)}{P(X_i > t)} \\ =&\frac{\mathrm{E}\left[\min\left(V_1, W\right)^\alpha\right] + \mathrm{E}\left[\min\left(W, V_2\right)^\alpha\right]}{\mathrm{E}[V_i^\alpha] + \mathrm{E}[W^\alpha]},\ i = 1, 2. \end{aligned} \tag{3.98}$$

Note that the above conditional probability is not the TDC when $\beta_1 \neq \beta_2$.This due to the fact that the marginal distribution of the two variables are not the same when the two processes are not symmetric, even though the exponents are the same. The conditional probability with either X_1 or X_2 given are different.

3.4.5.3 Experimental results

We did experiments for the general case. Let $\lambda = 1/4$, fix the value $\beta_1 = 0.001$ and increase β_2 value. First we compute the exponent α. As shown in Figure 3.21, the exponent α decreases as β_2 increases. $\beta = \sqrt{\beta_1\beta_2}$ also increases in β_2.

Next we compute the conditional probability at infinity $\lim_{x\to\infty} P(X_2 > x|X_1 > x)$ and $\lim_{x\to\infty} P(X_1 > x|X_2 > x)$ when $\beta_1 > \beta_2$. As shown in Figure 3.22, the conditional probability at infinity increases with β_2. Meanwhile, $\lim_{x\to\infty} P(X_2 > x|X_1 > x) < \lim_{x\to\infty} P(X_1 > x|X_2 > x)$ when $\beta_1 > \beta_2$ and vice versa. Note that, when $\beta_1 = \beta_2$, we know the conditional probability at infinity is also the TDC.

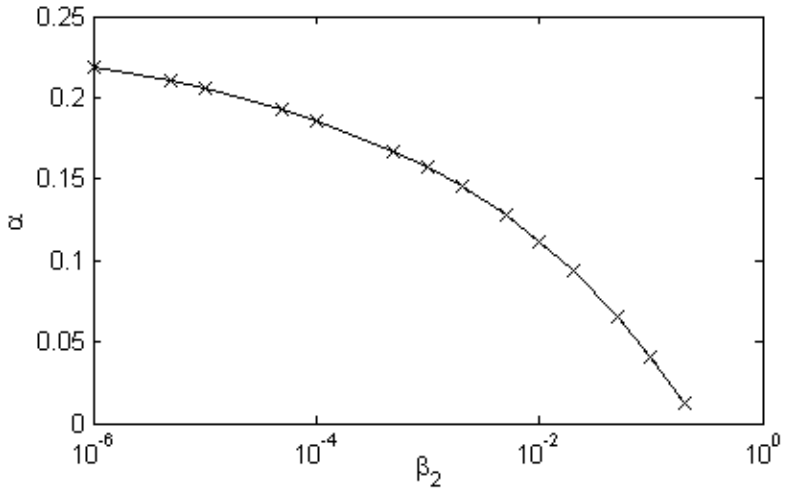


Figure 3.21. α as a function of β_2 ($\lambda = 0.25, \beta_1 = 0.001$)

Synthetic data generated from the generalized model is shown in Figure 3.23 with $\lambda = 2$, $\beta_1 = 0.04$ and $\beta_2 = 1$. With $\beta = \sqrt{\beta_1 \beta_2} = 0.2$, we have the same marginal exponents as the synthetic data in symmetric model, $\alpha = 1.9203$. We observe from the figure, when $\beta_1 \neq \beta_2$, the synthetic data from this model becomes asymmetric. This can be explained by analyzing the joint distribution of this model. When $\beta_1 = \beta_2$, we can get the density at point (x_1, x_2) is the same as the density at (x_2, x_1) from Equation (3.82); however this is not true when $\beta_1 \neq \beta_2$ from Equation (3.97).

3.5 Summary of Type 1 and Type 2 models

For the Type 1 and Type 2 models, we have published a paper in [56]. The reader could read that paper for a reference. For the joint density, we compare the synthetic data of the two types of models to the real dataset Youtube (as shown in Figure 3.24). We observe from the figure, that neither of the two types of models fits the real dataset in distribution. However, Type 1 model is better since it has some data concentrate on the 'equal line' as we discussed before. Type 2 model is totally on the opposite way. So, we believe the dependence between the two variables are

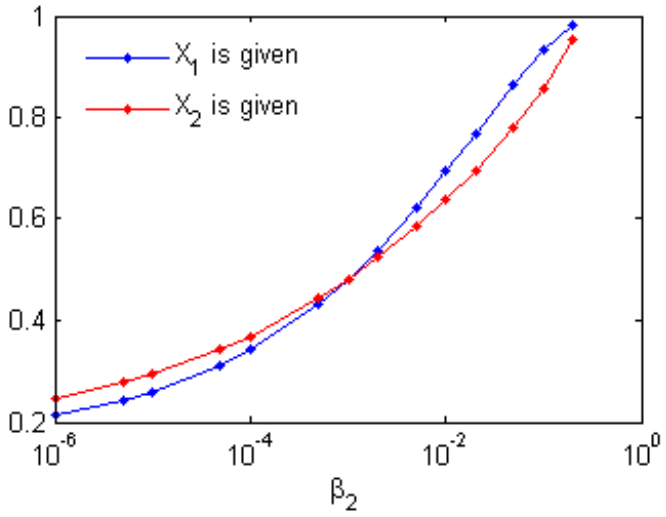


Figure 3.22. Comparing conditional probability $\lim_{x\to\infty} P(X_2 > x|X_1 > x)$ and $\lim_{x\to\infty} P(X_1 > x|X_2 > x)$ as a function of β_2 ($\lambda = 0.25, \beta_1 = 0.001$)

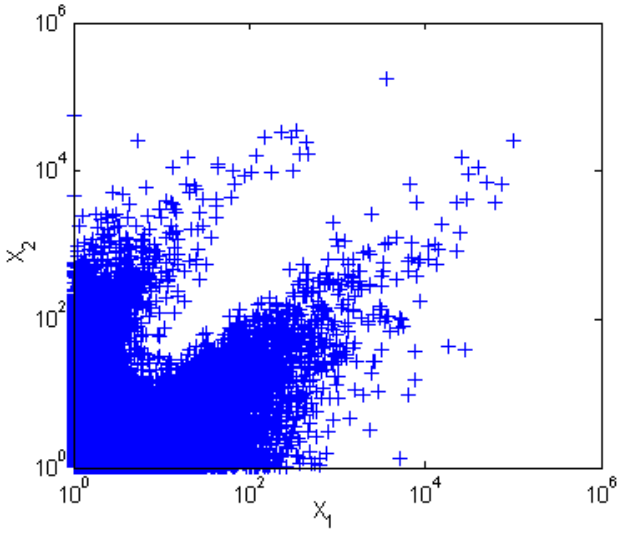


Figure 3.23. Synthetic data from generalized Type 2 model ($\lambda = 2, \beta_1 = 0.04, \beta_2 = 1$)

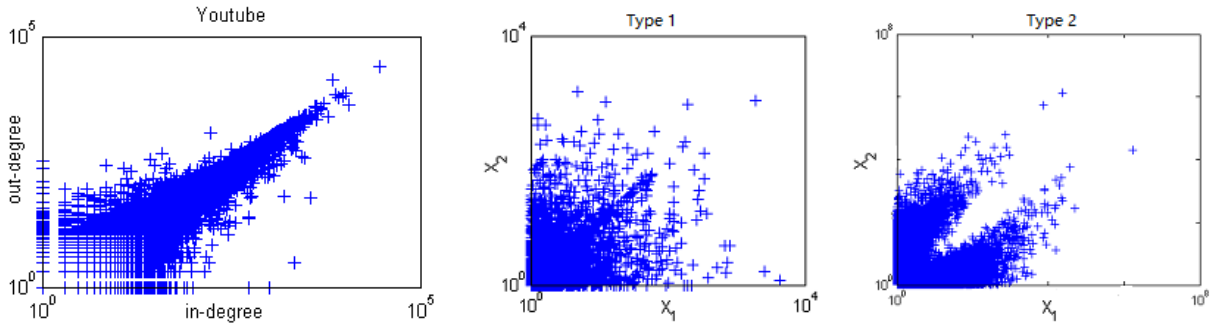


Figure 3.24. Comparing Type 1 and Type 2 models to real dataset

more likely to be introduced by the **shared Poisson counters**, but not the coupled growth.

The problem with the Type 1 model is the way to introduce in the independence. As we observe from Figure 3.24, Type 1 model has some data where X_1 and X_2 are totally independent. This is not the case in real datasets.

In the next Chapter, we will introduce a new type model. In the new model, the independence of the two variables are introduced in by the **independent Brownian motion components** instead of the independent Poisson counters in Type 1 model. We name this a **Type 3** model. The Type 3 model is important since it (1) fits real datasets; (2) connects to network growing models. We will discuss more in the following chapters.

CHAPTER 4

PCSDE MODELS FOR BIVARIATE HEAVY TAILED DISTRIBUTIONS-PART II: MODELS WITH BROWNIAN MOTION COMPONENTS

4.1 Mathematical background

In the previous chapter, we discussed two types of bivariate PCSDE models. The two types models do not fit real data observed in social networks. The Type 1 model is more promising but it has a problem generating independence. In this section, we introduce a new type of model, which keeps the shared Poisson counter in Type 1 model, but uses independent Brownian motion components to replace the two independent Poisson counters.

Before we continue to the PCSDE models with Brownian motion component, let's first review some properties of Brownian motion and some models related to Brownian motion.

4.1.1 Brownian motion

Brownian motion describes the random movement of a small particle suspend in water or in the air. The movement is caused by collisions between the small particle and fast moving molecules around it. The Brownian motion we focus on in this chapter is **Wiener process**.

The Wiener process $W(t)$ is also called standard Brownian motion. It has the following properties:

1. $W(0) = 0$;

2. $W(t)$ is continuous in t almost surely;

3. $W(t)$ has independent increments and $W(t) - W(s) \sim \mathcal{N}(0, t-s)$.

$W(t)$ has distribution

$$f_{W_t}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}},$$

with expectation

$$E(W(t)) = 0,$$

and variance

$$Var(W(t)) = t.$$

We can use the following SDE model to model the Brownian motion as in [34]

$$dX(t) = \frac{1}{\sqrt{\lambda}}[dN_1(t) - dN_2(t)]. \tag{4.1}$$

The two independent Poisson counters N_1 and N_2 have rates $\frac{\lambda}{2}$. The SDE model in Equation (4.1) describes the following process: a particle is at the origin at time 0, i.e. $X(0) = 0$, where $X(t)$ is the position of the particle at time t. When the particle is hit from the left by a water molecule, X increases by $\frac{1}{\sqrt{\lambda}}$. When it is hit from the right, X decreases by $\frac{1}{\sqrt{\lambda}}$. The arrivals of hits from left and right. When $\lambda \to \infty$, X converges to Brownian motion.

In Figure 4.1, 25 sample paths of Wiener process are plotted. Although the mean of samples keeps 0, the samples diverge with an increase in t.

4.1.2 Ito's rule for Brownian motion

Consider the following SDE driven by a Wiener process

$$dX = f(X)dt + g(X)dW. \tag{4.2}$$

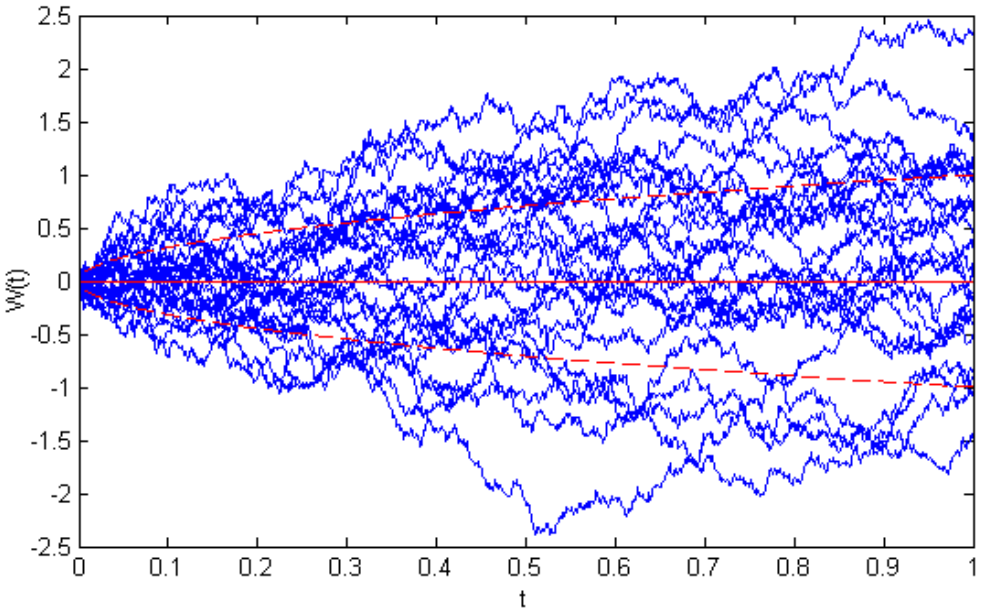


Figure 4.1. Brownian motion: 25 sample paths (blue lines), mean path (red solid line), one standard deviation from the mean (red dash line)

For a smooth function $\psi(X)$, Ito's rule yields

$$d\psi = \frac{\partial \psi}{\partial X}[f(X)dt + g(X)dW] + \frac{1}{2}\frac{\partial^2 \psi}{\partial X^2}g^2(X)dt. \tag{4.3}$$

Ito's rule for Brownian motion is useful in our calculations. Next we discuss some processes related to Wiener process, such as Brownian motion with drift and geometric Brownian motion [72][85].

4.1.3 Brownian motion with drift

Brownian motion with drift $X(t)$ is the solution to the following SDE

$$dX(t) = \beta dt + \sigma dW(t). \tag{4.4}$$

With initial condition $X(0) = \epsilon$, the solution to the SDE is

$$X(t) = \epsilon + \beta t + \sigma W(t).$$

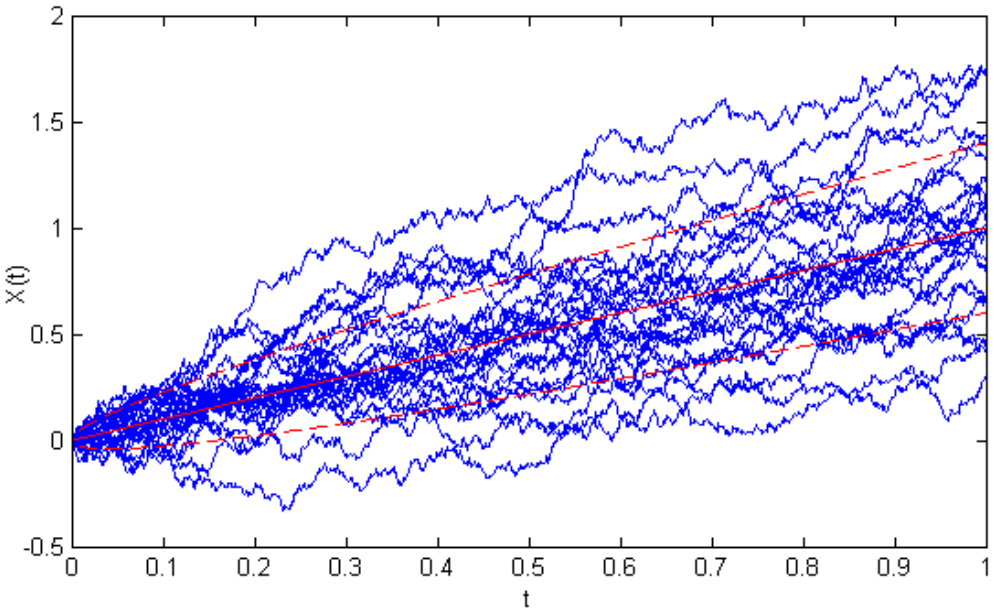


Figure 4.2. Brownian motion with drift: 25 sample paths (blue lines), mean path (red solid line), one standard deviation from the mean (red dash line)

Based on the distribution of $W(t)$, we have the distribution of $X(t)$:

$$f_X(x,t) = \frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{(x-\epsilon-\beta t)^2}{2\sigma^2 t}},$$

with expectation:

$$E(X(t)) = \epsilon + \beta t,$$

and variance:

$$Var(X(t)) = \sigma^2 t.$$

In Figure 4.2, 25 sample paths of Brownian motion with drift are plotted ($\beta = 1$, $\sigma = 0.4$, and $\epsilon = 0$). From the figure, we observe that the mean and standard deviation of the samples increase in t. Based on the computation results, the standard deviation increases more slowly than the mean.

4.1.4 Geometric Brownian motion (GBM)

Geometric Brownian motion is a more complicated model. As we stated in Section 3.1.2, this model has been widely used in finance modeling. The model is

$$dX(t) = \beta X(t)dt + \sigma X(t)dW(t). \tag{4.5}$$

Let $Y(t) = \log X(t)$, using Ito's rule for Brownian motion, we have

$$dY(t) = (\beta - \frac{1}{2}\sigma^2)dt + \sigma dW(t),$$

which is a Brownian motion with drift. With initial condition $X(0) = \epsilon$, we have $Y(0) = \log \epsilon$. Hence,

$$Y(t) = \log \epsilon + (\beta - \frac{1}{2}\sigma^2)t + \sigma W(t),$$

and

$$X(t) = \epsilon e^{(\beta - \frac{1}{2}\sigma^2)t + \sigma W(t)}.$$

Thus, $Y(t)$ follows normal distribution and $X(t)$ follows log-normal distribution. The density of $X(t)$ is

$$f_X(x,t) = \frac{1}{\sigma x\sqrt{2\pi t}} e^{-\frac{(\log x - \log \epsilon - (\beta - \frac{1}{2}\sigma^2)t)^2}{2\sigma t}},$$

with expectation

$$E(X(t)) = \epsilon e^{\beta t},$$

and variance

$$Var(X(t)) = \epsilon^2 e^{2\beta t}(e^{\sigma^2 t} - 1).$$

The expectation and variance can be computed as follows. First, take the expectations of both sides of the SDE in Equation (4.5). Setting $E(dW(t)) = 0$ yields

$$dE(X) = \beta E(X)dt,\ X(0) = \epsilon. \tag{4.6}$$

And the solution to the above DE is

$$E(X) = \epsilon e^{\beta t}. \tag{4.7}$$

Next we compute the second order moments. First, let $Y = X^2$. By applying Ito's rule for Brownian motion, we have

$$\begin{aligned} dX^2 =& 2X(\beta X dt + \sigma X dW) + \sigma^2 X^2 dt, \\ =& (2\beta + \sigma^2)X^2 dt + 2\sigma X^2 dW. \end{aligned} \tag{4.8}$$

Taking expectation

$$dE(X^2) = (2\beta + \sigma^2)E(X^2)dt,\ X^2(0) = \epsilon^2. \tag{4.9}$$

And the solution of the above DE is

$$E(X^2) = \epsilon^2 e^{(2\beta+\sigma^2)t}. \tag{4.10}$$

Hence, the variance of GBM is

$$Var(X) = E(X^2) - (E(X))^2 = \epsilon^2 e^{(2\beta+\sigma^2)t} - \epsilon^2 e^{2\beta t} = \epsilon^2 e^{2\beta t}(e^{\sigma^2 t} - 1). \tag{4.11}$$

The standard deviation of GBM is $\sigma_{GBM} = \epsilon e^{\beta t}\sqrt{e^{\sigma^2 t} - 1}$. We see that the standard deviation grows even faster compared to the mean. In Figure 4.3, we plot 25

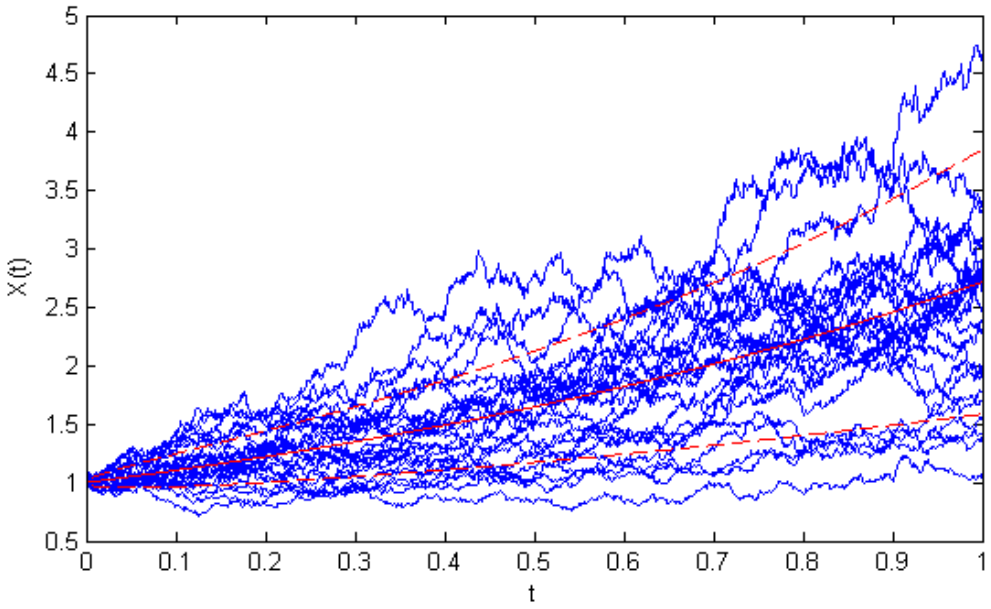


Figure 4.3. GBM: 25 sample paths (blue lines), mean path (red solid line), one standard deviation from the mean (red dash line)

sample paths ($\beta = 1$, $\sigma = 0.4$, and $\epsilon = 1$), corresponding mean path and 1 standard deviation lines to either side of the mean.

4.1.5 Constant Elasticity of Variance (CEV) model

Constant Elasticity of Variance (CEV) Model is first proposed by John Cox in 1975 [20]. The model was already introduced in Section 3.1.2. CEV is a more general model compared to GBM. When $\gamma = 1$, CEV becomes GBM. In our work, we consider the case when $\gamma < 1$:

$$dX(t) = \beta X(t)dt + \sigma X^{\gamma}(t)dW(t), \tag{4.12}$$

where $\beta, \sigma > 0, 0 < \gamma < 1$.

We compute the expectation and variance as we did for GBM. This is not easy for all $\gamma < 1$, we compute the simplest case when $\gamma = 1/2$,

$$dX(t) = \beta X(t)dt + \sigma X^{1/2}(t)dW(t). \tag{4.13}$$

First, take expectation of the SDE in Equation (4.13). $E(dW(t)) = 0$ yields the same DE as GBM,

$$dE(X) = \beta E(X)dt,\ X(0) = \epsilon. \tag{4.14}$$

And the solution to the above DE is

$$E(X) = \epsilon e^{\beta t}. \tag{4.15}$$

Now we compute the second order moments. First, let $Y = X^2$. By applying Ito's rule for Brownian motion we have

$$\begin{aligned} dX^2 =& 2X(\beta X dt + \sigma X^{1/2} dW) + \sigma^2 X dt, \\ =& (2\beta X^2 + \sigma^2 X)dt + 2\sigma X^{3/2} dW. \end{aligned} \tag{4.16}$$

Taking expectation and with $E(X) = \epsilon e^{\beta t}$, we have

$$dE(X^2) = 2\beta E(X^2)dt + \sigma^2 \epsilon e^{\beta t},\ X^2(0) = \epsilon^2. \tag{4.17}$$

The solution to the above DE is

$$\begin{aligned} E(X^2) =& \epsilon^2 e^{2\beta t} + \sigma^2 \int_0^t \epsilon e^{2\beta(t-s)} e^{\beta s} ds, \\ =& \epsilon^2 e^{2\beta t} + \frac{\sigma^2 \epsilon}{\beta}(e^{2\beta t} - e^{\beta t}). \end{aligned} \tag{4.18}$$

The variance of CEV is

$$Var(X) = E(X^2) - (E(X))^2 = \frac{\sigma^2 \epsilon}{\beta}(e^{2\beta t} - e^{\beta t}). \tag{4.19}$$

The standard deviation of CEV ($\gamma = 1/2$) is $\sigma_{CEV} = \sigma\sqrt{\frac{\epsilon}{\beta}}\sqrt{(e^{2\beta t} - e^{\beta t})}$. Compared to GBM, the standard deviation grows at the same speed as the mean. In

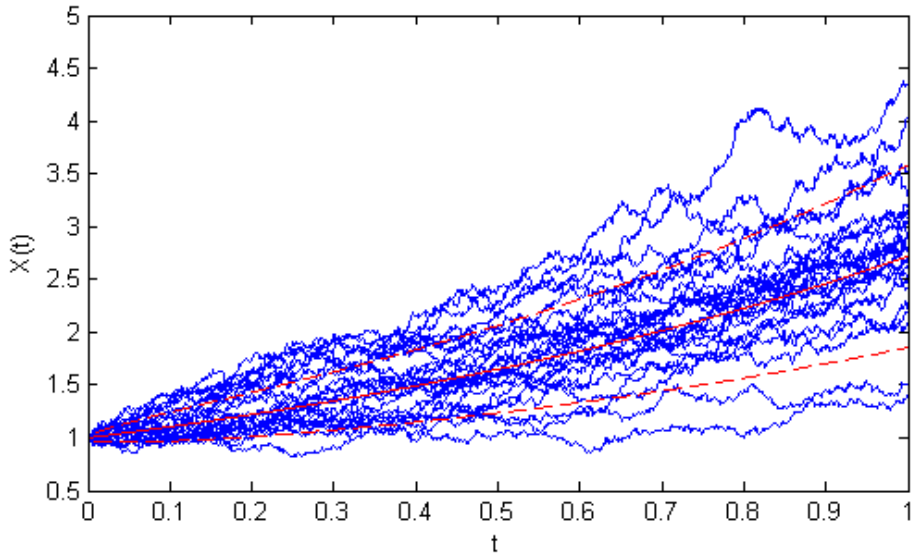


Figure 4.4. CEV ($\gamma = 1/2$): 25 sample paths (blue lines), mean path (red solid line), one standard deviation from the mean (red dash line)

	$E(X(t))$	Var(X(t))	$STD(X(t))$
GBM	$\epsilon e^{\beta t}$	$\epsilon^2 e^{2\beta t}(e^{\sigma^2 t} - 1)$	$\epsilon e^{\beta t}\sqrt{e^{\sigma^2 t} - 1}$
CEV($\gamma = \frac{1}{2}$)	$\epsilon e^{\beta t}$	$\frac{\sigma^2}{\beta}\epsilon(e^{2\beta t} - e^{\beta t})$	$\frac{\sigma\sqrt{\epsilon}}{\sqrt{\beta}}\sqrt{e^{2\beta t} - e^{\beta t}}$

Table 4.1. Statistics comparison of GBM and CEV ($\gamma = \frac{1}{2}$).

Figure 4.4, we plot 25 sample paths ($\beta = 1$, $\sigma = 0.4$, and $\epsilon = 1$), corresponding mean path and one standard deviation lines on both sides of the mean as we did for GBM. We compare the statistics of GBM and CEV ($\gamma = 1/2$) in Table 4.1. With the same parameters, CEV performs quite differently from GBM. It is reasonable that we expect different behaviors for the PCSDE models based on GBM and based on CEV. We will discuss the two kinds of PCSDE models in the following.

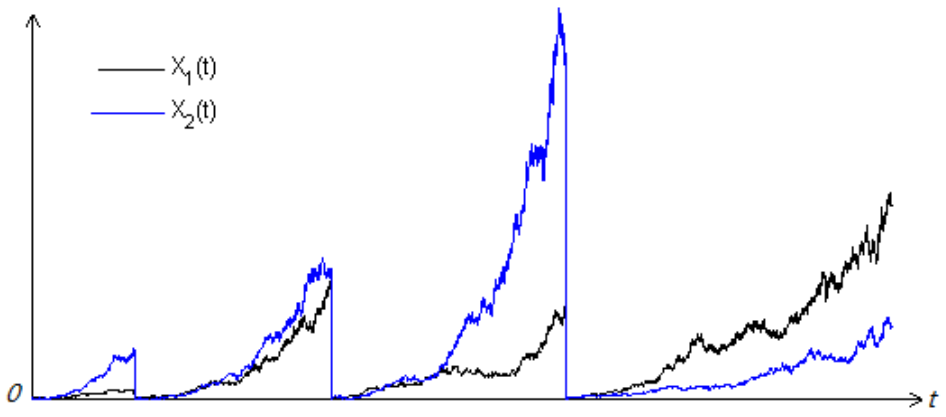


Figure 4.5. Sample path of the first Type 3 model in Equation (4.21) ($\lambda = 0.5, \mu = 1, \sigma = 0.4$)

4.2 PCSDE model based on Geometric Brownian Motion (GBM) (the first Type 3 model)

4.2.1 Model formulation

By adding a Poisson counter to Geometric Brownian motion, we obtain the following PCSDE model

$$dX(t) = \beta X(t)dt + \sigma X(t)dW(t) + (\epsilon - X(t))dN(t). \tag{4.20}$$

This model has been introduced in Section 3.2.3 to produce double-Pareto distribution.

A bivariate extension with a shared Poisson counter is

$$dX_i = \beta_i X_i dt + \sigma_i X_i dW_i + (\epsilon_i - X_i)dN, i = 1, 2, \tag{4.21}$$

where N is a shared Poisson counter with rate λ. We call the bivariate PCSDE model based on GBM **the first Type 3 model**. A sample path of this model is shown in Figure (4.5).

4.2.2 Marginal tail

In [43], the authors used a characteristic function method to compute the distribution of this model. Please refer to the paper for more details. Unfortunately, we cannot apply this method to get the result for the bivariate extension of this model. Instead we introduce another method in the following based on previous results in [64] and [38].

Let $Y = \ln X$, we have

$$dY(t) = \left(\beta - \frac{\sigma^2}{2}\right) dt + \sigma dW(t) + (\ln \epsilon - Y(t))dN(t).$$

The solution to the differential equation $dY(t) = (\beta - \frac{\sigma^2}{2})dt + \sigma dW(t)$ is

$$Y(t) = \ln \epsilon + \left(\beta - \frac{\sigma^2}{2}\right) t + \sigma W(t),$$

and

$$X(t) = \epsilon e^{\left(\beta - \frac{\sigma^2}{2}\right)t + \sigma W(t)}.$$

We are interested in the case when $\beta - \frac{\sigma^2}{2} > 0$.

The Poisson counter N_t has rate λ. Define the life time T: the time between the last active Poisson counter occurring and the observation time. By PASTA property [92], we can prove that $T \sim \exp(\lambda)$. The proof is as follows:

Let $U \sim \exp(\lambda)$. Given $U = u$, $T \sim \mathbb{U}(0, u)$. We have

$$\begin{aligned}
\mathbb{P}(T \geq x) &= \frac{E\int_0^U \mathbf{1}(t \geq x)dt}{EU} \\
&= \frac{\int_0^\infty \lambda e^{-\lambda u}du \int_0^u \mathbf{1}(t \geq x)dt}{\int_0^\infty \lambda e^{-\lambda u}du} \\
&= \frac{\int_x^\infty \lambda e^{-\lambda u}(u-x)du}{\frac{1}{\lambda}} \\
&= \frac{\lambda e^{-\lambda x}\int_x^\infty \lambda e^{-\lambda(u-x)}(u-x)d(u-x)}{\frac{1}{\lambda}} \\
&= \frac{\lambda e^{-\lambda x}\frac{1}{\lambda^2}}{\frac{1}{\lambda}} = e^{-\lambda x}.
\end{aligned}$$

Thus $T \sim \exp(\lambda)$.

Given $W(t) \sim \mathcal{N}(0,t)$, X follows a lognormal distribution

$$\ln(X) \sim \mathcal{N}(\ln\epsilon + (\beta - \frac{\sigma^2}{2})T, \sigma^2 T), \tag{4.22}$$

which gives

$$f_X(x,T) = \frac{1}{\sqrt{2\pi T}\sigma x}e^{-\frac{\left(\ln x - \ln\epsilon - (\beta-\frac{\sigma^2}{2})T\right)^2}{2\sigma^2 T}},$$

and thus

$$f_X(x) = \int_0^\infty \lambda e^{-\lambda t}\frac{1}{\sqrt{2\pi t}\sigma x}e^{-\frac{\left(\ln(x/\epsilon) - (\beta-\frac{\sigma^2}{2})t\right)^2}{2\sigma^2 t}}dt.$$

Let $t = s^2$,

$$f_X(x) = \frac{2\lambda}{\sqrt{2\pi}\sigma x}e^{\frac{\ln(x/\epsilon)(\beta-\frac{\sigma^2}{2})}{\sigma^2}}\int_0^\infty e^{-\frac{\lambda+\left(\beta-\frac{\sigma^2}{2}\right)^2}{2\sigma^2}s^2}e^{-\frac{(\ln(x/\epsilon))^2}{2\sigma^2 s^2}}ds.$$

Given the identity

$$\int_{z=0}^\infty e^{-az^2 - b/z^2}dz = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-2\sqrt{ab}},$$

we have

$$f_X(x) = \begin{cases} \theta_1 \epsilon^{-1} \left(\frac{x}{\epsilon}\right)^{\frac{\beta-\sigma^2/2+\theta_2}{\sigma^2}-1}, & x < \epsilon, \\ \theta_1 \epsilon^{-1} \left(\frac{x}{\epsilon}\right)^{\frac{\beta-\sigma^2/2-\theta_2}{\sigma^2}-1}, & x \geq \epsilon, \end{cases} \tag{4.23}$$

where $\theta_1 = \frac{\lambda}{\sqrt{(\beta-\sigma^2/2)^2+2\lambda\sigma^2}}$ and $\theta_2 = \sqrt{(\beta-\sigma^2/2)^2+2\lambda\sigma^2}$. The result is the same as using the characteristic function method in [43].

Consider the case $\epsilon = 1$, $\beta = 1$ and $\sigma = 1$. Based on the result in Equation (4.23), the marginal density at the upper tail

$$f_X(x) = \frac{\lambda}{\sqrt{2\lambda+\frac{1}{4}}} x^{-(\sqrt{2\lambda+\frac{1}{4}}+\frac{1}{2})}, \ x \to \infty. \tag{4.24}$$

4.2.3 Joint CCDF and TDC

Given $T \sim \exp(\lambda)$, $X_1 \sim \mathcal{N}(\ln\epsilon_1 + (\beta_1 - \frac{\sigma_1^2}{2})T, \ \sigma_1^2 T)$ and $X_2 \sim \mathcal{N}(\ln\epsilon_2 + (\beta_2 - \frac{\sigma_2^2}{2})T, \ \sigma_2^2 T)$. X_1 and X_2 are independent. We have

$$f_{X_1,X_2}(x_1,x_2,T) = \frac{1}{2\pi T\sigma_1\sigma_2 x_1 x_2} e^{-\left(\frac{\left(\ln(\frac{x_1}{\epsilon_1})-(\beta_1-\frac{\sigma_1^2}{2})T\right)^2}{2\sigma_1^2 T}+\frac{\left(\ln(\frac{x_2}{\epsilon_2})-(\beta_2-\frac{\sigma_2^2}{2})T\right)^2}{2\sigma_2^2 T}\right)}. \tag{4.25}$$

We give the joint distribution

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2 x_1 x_2}\int_0^\infty \lambda e^{-\lambda t}\frac{1}{t} e^{-\left(\frac{\left(\ln(\frac{x_1}{\epsilon_1})-(\beta_1-\frac{\sigma_1^2}{2})t\right)^2}{2\sigma_1^2 t}+\frac{\left(\ln(\frac{x_2}{\epsilon_2})-(\beta_2-\frac{\sigma_2^2}{2})t\right)^2}{2\sigma_2^2 t}\right)} dt. \tag{4.26}$$

We prove the joint density of this model at the tail

$$f_{X_1,X_2}(x_1,x_2) \sim x_1^{-\frac{3}{2}+\frac{\beta_1}{\sigma_1^2}} x_2^{-\frac{3}{2}+\frac{\beta_2}{\sigma_2^2}} H^{-1/4} e^{-\sqrt{2B}H^{1/2}}, \tag{4.27}$$

where

$$H = \left(\frac{(\ln x_1)^2}{\sigma_1^2} + \frac{(\ln x_2)^2}{\sigma_2^2} \right),$$

and

$$B = \lambda + \left(\frac{(\beta_1 - \frac{\sigma_1^2}{2})^2}{2\sigma_1^2} + \frac{(\beta_2 - \frac{\sigma_2^2}{2})^2}{2\sigma_2^2} \right).$$

The proof of conclusion in Equation (4.27) is in Appendix D.1.

To analyze the tail dependence, we need to compute the joint CCDF of this model. To simplify the computation, let $\beta_1 = \beta_2 = 1, \sigma_1 = \sigma_2 = 1$. We begin with the definition of CCDF

$$\begin{aligned} P(X_1 > x, X_2 > x) &= \int_x^\infty \int_x^\infty f_{X_1,X_2}(x_1, x_2) dx_1 dx_2, \\ &\sim \frac{(2B)^{-1/4}}{\sqrt{2\pi}\sigma_1\sigma_2} \int_x^\infty \int_x^\infty x_1^{-1/2} x_2^{-1/2} \left((\ln x_1)^2 + (\ln x_2)^2 \right)^{-1/4} \\ &\quad e^{-\sqrt{2B}\left((\ln x_1)^2 + (\ln x_2)^2 \right)^{1/2}} dx_1 dx_2. \end{aligned} \tag{4.28}$$

We prove the joint CCDF of this model at the tail

$$P(X_1 > x, X_2 > x) \sim x^{-(\sqrt{4\lambda+1}-1)}. \tag{4.29}$$

The detailed proof is in Appendix D.2.

The marginal CCDF of this model can be computed from the marginal density in Equation (4.24), that

$$P(X_1 > x) = \int_x^\infty f_{X_1}(x_1) dx_1 \sim x^{-(\sqrt{2\lambda+\frac{1}{4}}-\frac{1}{2})}.$$

With the joint CCDF in Equation (4.29), we prove that

$$P(X_2 > x | X_1 > x) \sim x^{-(\sqrt{4\lambda+1}-\sqrt{2\lambda+\frac{1}{4}}-\frac{1}{2})}. \tag{4.30}$$

It is easy to show that $\sqrt{4\lambda+1} - \sqrt{2\lambda+\frac{1}{4}} - \frac{1}{2} > 0$ for all λ, which means $TDC = \lim_{x\to\infty} P(X_2 > x | X_1 > x) = 0$.

Thus, the first Type 3 model is an asymptotically independent model. Before we proceed to the new model, we perform experiments using data generated from the first Type 3 model.

4.2.4 Experimental results

We generate two sets of synthetic data from the first Type 3 model with different σ values ($\lambda = 1, \beta = 0.9$). As shown in Figure 4.6, the synthetic data from the first Type 3 model based on GBM does not fit the real data. And as the values increase, the dependence between X_1 and X_2 decreases. This corresponds to our theoretical results in Section 4.2.3, that the model is asymptotic independent. Although the theoretical result is for special case when $\beta_1 = \beta_2 = 1, \sigma_1 = \sigma_2 = 1$, it should work on other cases too. We skip the proof for the other cases.

The marginal CCDF ($P(X_1 > x)$ and $P(X_2 > x)$) and dependence coefficients ($P(X_1 > x | X_2 > x)$ and $P(X_2 > x | X_1 > x)$) as a function of x of the above two groups of data are shown in Figure 4.7. We observe from the figure, the dependence coefficients decrease to 0 with x increasing and the speed to 0 is faster when σ is larger.

In the next section, we consider a more generalized model based on CEV ($0 < \gamma < 1$) instead of GBM. As we said in Section 4.1.5, we expect this model to perform differently from the model based on GBM.

4.3 PCSDE Model based on Constant Elasticity of Variance (CEV) Model with $0 < \gamma < 1$ (the second Type 3 model)

We restate the CEV model ($0 < \gamma < 1$) from Section 4.1.5

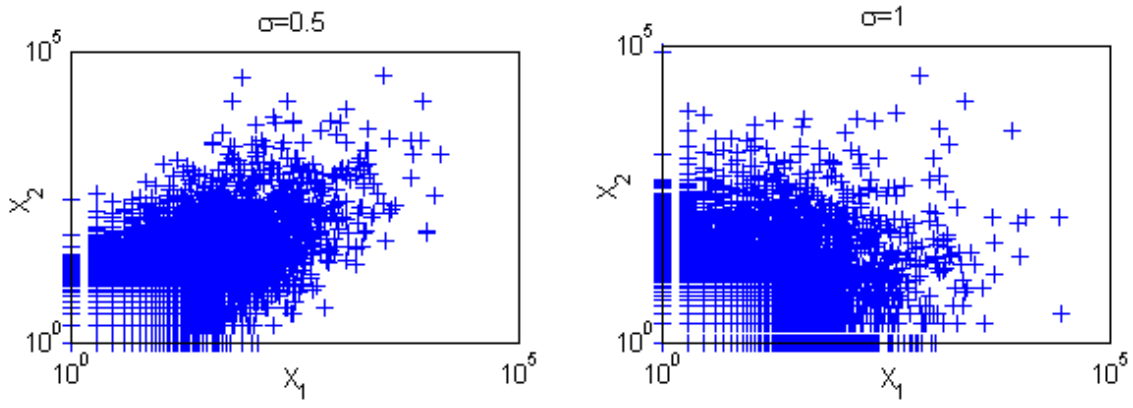
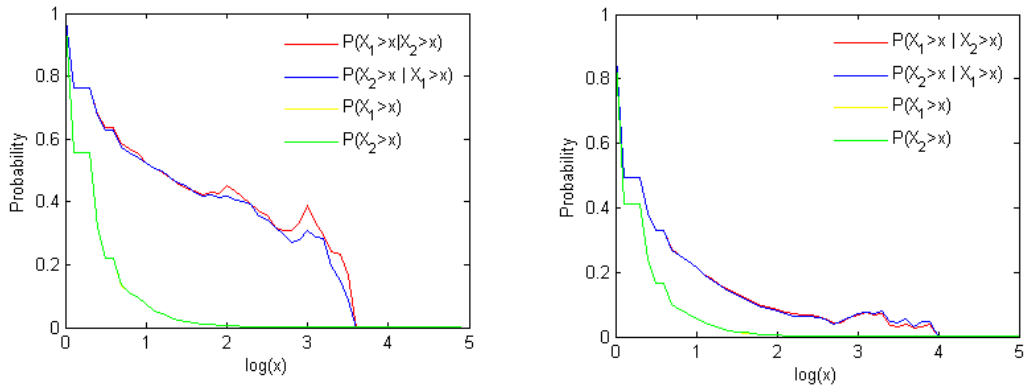


Figure 4.6. Synthetic data of the first Type 3 model



(a) $\sigma = 0.5$ (b) $\sigma = 1$

Figure 4.7. Marginal CCDF and dependence coefficients of data in Figure 4.6

$$dX(t) = \beta X(t)dt + \sigma X^{\gamma}(t)dW(t), \tag{4.31}$$

where $\beta, \sigma > 0, 0 < \gamma < 1$.

Theorem 2 in [16] gives the density of this model

$$f_{X_t|X_0}(x,t) = 2(1-\gamma)\tilde{k}^{\frac{1}{2(1-\gamma)}}(\tilde{x}\tilde{z}^{1-4\gamma})^{\frac{1}{4(1-\gamma)}}\exp\{-\tilde{x}-\tilde{z}\}I_{\frac{1}{2(1-\gamma)}}(2(\tilde{x}\tilde{z})^{\frac{1}{2}}), \tag{4.32}$$

where

$$\tilde{k} = \frac{\beta}{\sigma^2(1-\gamma)(e^{2\beta(1-\gamma)t}-1)},$$
$$\tilde{x} = \tilde{k}X_0^{2(1-\gamma)}e^{2\beta(1-\gamma)t},$$
$$\tilde{z} = \tilde{k}x^{2(1-\gamma)},$$

and $I_v(z)$ is the modified Bessel function of the first kind of order v

$$I_v(z) = (\frac{1}{2}z)^v \sum_{n=0}^{\infty} \frac{(\frac{1}{4}z^2)^n}{n!\Gamma(v+n+1)}. \tag{4.33}$$

The density for this model is more complicated than that of the original model GBM. Since it is not easy to compute the density of PCSDE model based on this model, we consider a specific condition $\gamma = 1/2$ in the following.

4.3.1 Model formulation

By adding a Poisson counter component to the CEV model, we obtain a new type of PCSDE model

$$dX(t) = \beta X(t)dt + \sigma X^{\gamma}(t)dW(t) + (\epsilon - X(t))dN(t), \tag{4.34}$$

where W is standard Brownian motion, N is a Poisson counter with rate λ and $0 < \gamma < 1$.

A bivariate extension of the above single variate model is

$$dX_i = \beta_i X_i dt + \sigma_i X_i{}^{\gamma} dW_i + (\epsilon - X_i) dN, i = 1, 2, \tag{4.35}$$

where W_1 and W_2 are two independent Wiener processes. We call the bivariate PCSDE model based on CEV ($\gamma = 1/2$) **the second Type 3 model**.

4.3.2 Marginal tail

To compute the marginal tail, we further simplify this model by setting $\epsilon = 1$ and $\gamma = 1/2$. In Section 5.1, we will show that the case when $\gamma = 1/2$ also relates to network growing models.

We have the density function for the DE in (4.32) becomes

$$f_{X_t}(x,t) = \frac{2\beta}{\sigma} x^{-1/2} \frac{e^{\beta t/2}}{e^{\beta t} - 1} \exp\left\{ -\frac{2\beta}{\sigma} \frac{e^{\beta t} + x}{e^{\beta t} - 1} \right\} I_1\left(\frac{4\beta}{\sigma} \frac{e^{\beta t/2}}{e^{\beta t} - 1} x^{1/2} \right), \tag{4.36}$$

where $I_1(z) = (\frac{1}{2}z) \sum_{n=0}^{\infty} \frac{(\frac{1}{4}z^2)^n}{n!\Gamma(n+2)}$. With $t \sim \exp(\lambda)$, we have

$$f_X(x) = \lambda \frac{2\beta}{\sigma} x^{-1/2} \int_0^{\infty} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t} - 1} \exp\left\{ -\frac{2\beta}{\sigma} \frac{e^{\beta t} + x}{e^{\beta t} - 1} \right\} I_1\left(\frac{4\beta}{\sigma} \frac{e^{\beta t/2}}{e^{\beta t} - 1} x^{1/2} \right) dt. \tag{4.37}$$

It is not easy to obtain a closed-form solution to Equation (4.37). We try to analyze the marginal density at the tail as $x \to \infty$.

We obtain the following conclusion

$$\lim_{x \to \infty} x^{1+\lambda/\beta} f_X(x) = \lambda \frac{2}{\sigma} \int_0^{\infty} y^{-3/2-\lambda/\beta} \exp\left\{ -\frac{2\beta}{\sigma} y^{-1} \right\} I_1\left(\frac{4\beta}{\sigma} y^{-1/2} \right) dy \in (0, \infty), \tag{4.38}$$

which means $f_X(x) \sim x^{-(1+\lambda/\beta)}$. The reader is referred to Appendix E.1 for a detailed proof.

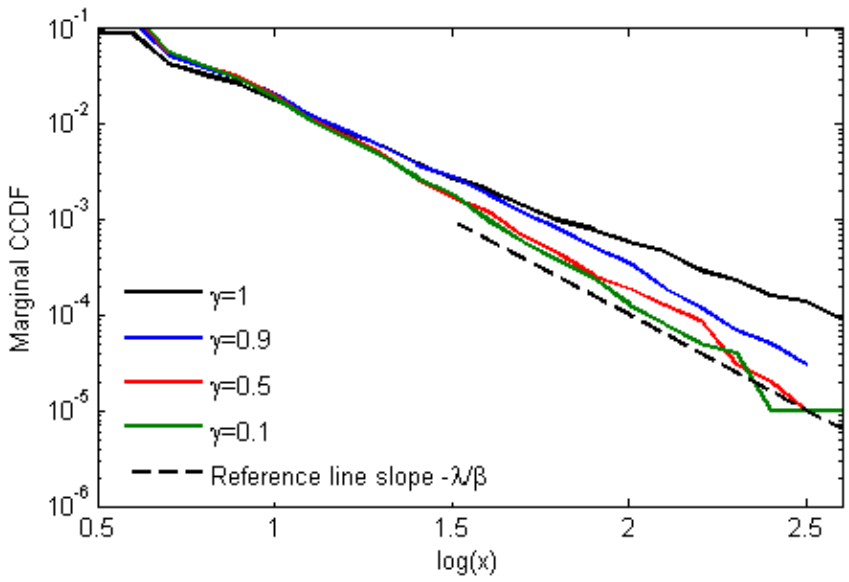


Figure 4.8. The marginal CCDF of synthetic data for the second Type 3 model with γ varies

The marginal tail is the same as the original PCSDE model without a Brownian motion component in Equation (3.17). This is quite different from the first Type 3 model. In Figure 4.8, we plot the marginal CCDF of the synthetic data generated by the second Type 3 model with different γ values. For comparison, we also plot the CCDF of synthetic data of the first Type 3 model with $\gamma = 1$. The black dashed line is a reference line with slope $-\lambda/\beta$, which is the exponent of the original model with no Brownian motion component. We observe that the tail exponent is the same as the original model even with $\gamma = 0.9$, which is quite close to $\gamma = 1$. Actually we can prove this in an easier way and the proof is applied to all the cases when $0 < \gamma < 1$ with no need to specify $\gamma = 1/2$.

Let $Y = \log X$. By applying Ito's rule for Brownian motion, Equation (4.34) becomes

$$dY = \left(\beta - \frac{\sigma^2}{2} e^{-2Y(1-\gamma)}\right) dt + \sigma e^{-Y(1-\gamma)} dW - Y dN.$$

With $0 < \gamma < 1$, we have $0 < (1-\gamma) < 1$. When $Y \to \infty$,

$$dY = \beta dt - Y dN.$$

With $X = e^Y$, we have

$$dX = \beta X dt + (1 - X) dN,\ X \to \infty,$$

which gives the marginal tail

$$f_X(x) \sim x^{-\lambda/\beta - 1},\ x \to \infty.$$

4.3.3 Joint CCDF and TDC

In Appendix E.2, we give a rigorous analysis of the joint tail for the case when $\gamma = 1/2$. The conclusion is that

$$P(X_1 > x, X_2 > x) \sim x^{-\lambda/\beta}.$$

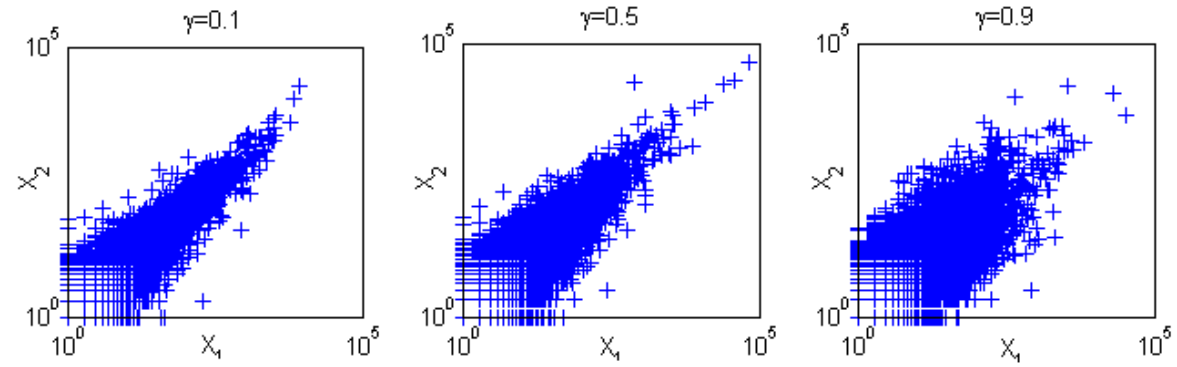
Based on the marginal density in Equation 4.38, we have the marginal CCDF

$$P(X > x) \sim x^{-\lambda/\beta}. \tag{4.39}$$

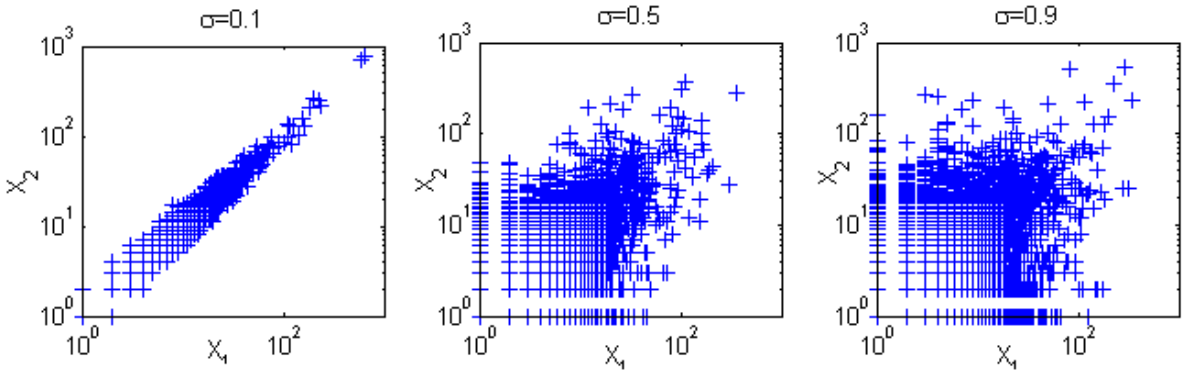
With the same tail exponent of the marginal and joint CCDF, we conclude the tail dependence of the second Type 3 model is non-zero given $\gamma = 1/2$. The reader is referred to Appendix E.2 for a detailed proof.

4.3.4 Experimental results

We generate two groups of synthetic data from the second Type 3 model with different γ and different σ values. In each group, there are three sets of data. As shown in Figure 4.9, with the increasing of γ and σ values, the dataset becomes less dependent between X_1 and X_2. Meanwhile, with bigger γ value, the tail part of



(a) γ vary ($\beta = 0.9$, $\sigma = 0.5$)



(b) σ vary ($\beta = 0.5$, $\gamma = 0.5$)

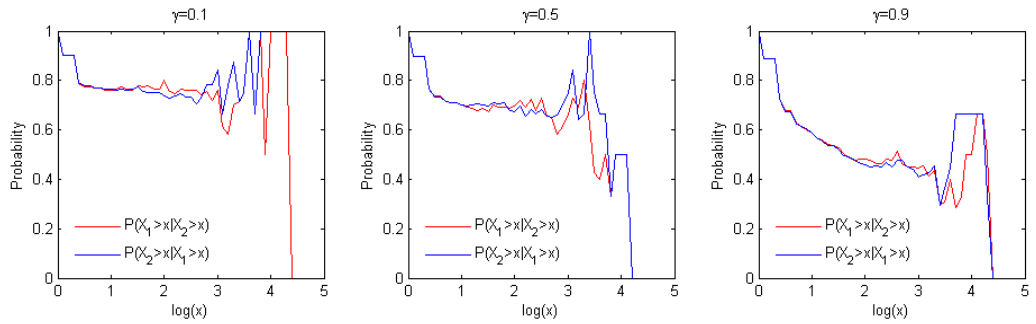
Figure 4.9. Synthetic data from the second Type 3 model with γ and σ varying

the data becomes fatter; while the body part of the dataset becomes fatter when σ becomes bigger.

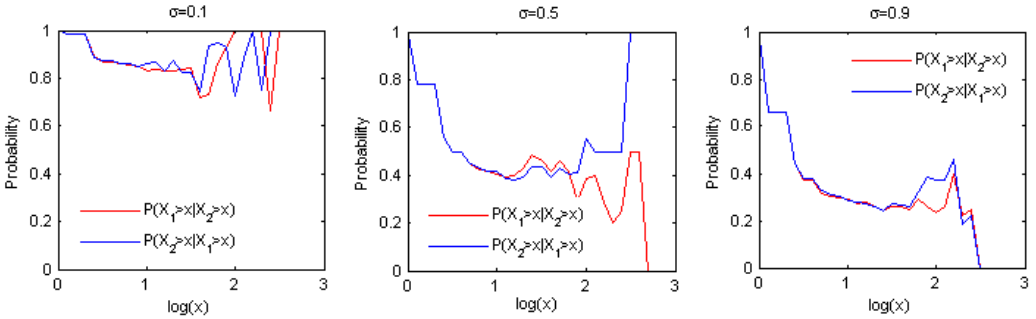
For different values of σ and γ, we plot the dependence coefficients of the datasets we generated in Figures 4.10. We observe that the dependence coefficients tend to stabilize around a fractional value when x goes to large and the TDC is smaller when σ and γ is larger.

4.4 Pearson correlation coefficients

In Table 4.1, we give the first and second moments for GBM and CEV ($\gamma = 1/2$). In this section, we will use the moments to calculate the moments for PCSDE models. From Table 4.1, we observe that the expectations for GBM and CEV ($\gamma = 1/2$) are



(a) γ vary ($\beta = 0.9$, $\sigma = 0.5$)



(b) σ vary ($\beta = 0.5$, $\gamma = 0.5$)

Figure 4.10. Dependence coefficients of datasets in Figure 4.9

the same

$$E(X(t)) = e^{\beta t}.$$

But for the second moments, we have

$$E(X^2(t)) = e^{(2\beta+\sigma^2)t},$$

for GBM; and

$$E(X^2(t)) = e^{2\beta t} + \frac{\sigma^2}{\beta}(e^{2\beta t} - e^{\beta t}),$$

for CEV ($\gamma = 1/2$).

Given X_1 and X_2 having the same life time but generated with independent random seeds,

$$E(X_1(t)X_2(t)) = E(X_1(t))E(X_2(t)) = e^{2\beta t}.$$

With Poisson resetting in the PCSDE model, life time $t \sim \exp(\lambda)$. Thus we have the moments for the Type 3 PCSDE model:

$$E(X^m) = \int_0^\infty \lambda e^{-\lambda t} E(X^m(t)) dt.$$

It is easy to get that, when $\beta < \lambda < 2\beta$, $E(X)$ exists; $E(X^2)$ for both GBM and CEV ($\gamma = 1/2$) are infinite. For a finite dataset, it normally has an upperbound for the life time (the observation time) T. Assume life time follow truncated exponential distribution with an upperbound T. Let $\beta_1 = \beta_2 = \beta$, $\sigma_1 = \sigma_2 = \sigma$, $\lambda_1 = \lambda_2 = \lambda$. We have

$$E(X_1) = E(X_2) = \int_0^\infty \lambda e^{-\lambda t} e^{\beta t} dt = \frac{\lambda}{\lambda - \beta} \left(1 - e^{-(\lambda-\beta)T}\right),$$

and

$$E(X_1 X_2) = \int_0^\infty \lambda e^{-\lambda t} e^{2\beta t} dt = \frac{\lambda}{2\beta - \lambda} \left(e^{(2\beta-\lambda)T} - 1\right) = O(e^{(2\beta-\lambda)T}).$$

For GBM-based PCSDE,

$$\begin{aligned}
\sigma_{X_i}^2 =& E(X_i^2) - (E(X_i))^2, \\
=& \int_0^T \lambda e^{-\lambda t} e^{(2\beta+\sigma^2)t} dt - \left(\frac{\lambda}{\lambda - \beta} \left(1 - e^{-(\lambda-\beta)T}\right)\right)^2, \\
=& \frac{\lambda}{2\beta + \sigma^2 - \lambda} \left(e^{(2\beta+\sigma^2-\lambda)T} - 1\right) - \left(\frac{\lambda}{\lambda - \beta} \left(1 - e^{-(\lambda-\beta)T}\right)\right)^2, \\
=& O(e^{(2\beta-\lambda+\sigma^2)T}).
\end{aligned}$$

For CEV-based PCSDE ($\gamma = 1/2$),

$$\begin{aligned}
\sigma_{X_i}^2 =& \int_0^T \lambda e^{-\lambda t} \left((\frac{\sigma^2}{\beta} + 1) e^{2\beta t} - \frac{\sigma^2}{\beta} e^{\beta t}\right) dt - \left(\frac{\lambda}{\lambda - \beta} \left(1 - e^{-(\lambda-\beta)T}\right)\right)^2, \\
=& \frac{\lambda}{(2\beta - \lambda)} (1 + \frac{\sigma^2}{\beta}) \left(e^{(2\beta-\lambda)T} - 1\right) - \frac{\lambda\sigma^2}{\beta(\lambda - \beta)} \left(1 - e^{-(\lambda-\beta)T}\right) - \left(\frac{\lambda}{\lambda - \beta} \left(1 - e^{-(\lambda-\beta)T}\right)\right)^2, \\
=& O(e^{(2\beta-\lambda)T}).
\end{aligned}$$

We observe that the variance $\sigma^2_{X_i}$ grows faster with T than $E(X_1X_2)$ for GBM-based PCSDE; while $\sigma^2_{X_i}$ and $E(X_1X_2)$ grow at the same speed for CEV-based PCSDE ($\gamma = 1/2$).

We calculate the Pearson correlation coefficient between X_1 and X_2 to be

$$Corrcoef(X_1, X_2) = \frac{E(X_1X_2) - E(X_1)E(X_2)}{\sigma_{X_1}\sigma_{X_2}}.$$

When $T \to \infty$, we have

$$Corrcoef(X_1, X_2) \sim \frac{2\beta + \sigma^2 - \lambda}{2\beta - \lambda} e^{-\sigma^2 T} \to 0,$$

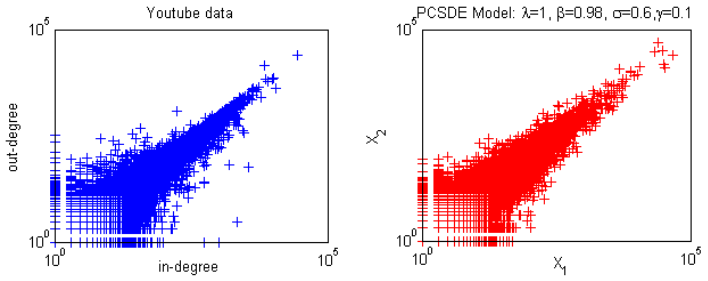
for the first Type 3 model; and

$$Corrcoef(X_1, X_2) \to \frac{1}{1 + \frac{\sigma^2}{\beta}} \in (0, 1),$$

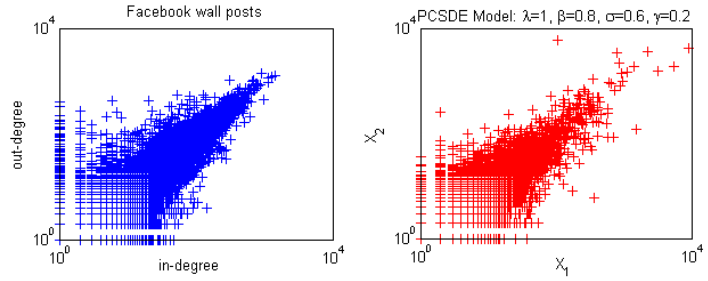
for the second Type 3 model ($\gamma = 1/2$).

Although the above results for the second Type 3 model is under a specific condition $\gamma = 1/2$, we have shown in our previous sections that, the results for the other parameters $0 < \gamma < 1$ should be similar. The statistics of the four datasets are given in Table 4.2. Compared to the statistics in Table 4.2, the second Type 3 model based on CEV fits the data better. The reason is that $E[d_{in}d_{out}]$ and $\sigma^2_{d_{in}}$ ($\sigma^2_{d_{out}}$) of the real datasets are in the same order of magnitude. Meanwhile, the correlation coefficients are much larger than 0.

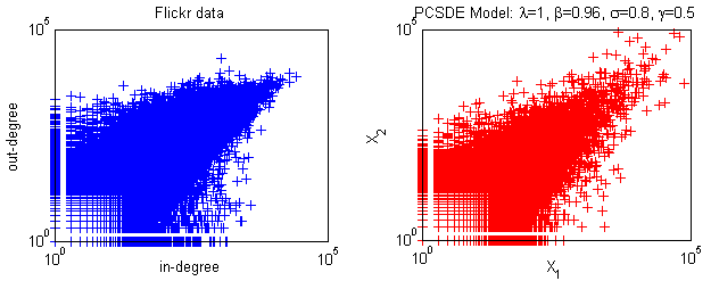
To fit real datasets, we tune the parameters λ, β, σ and γ of the second Type 3 model in Equation (4.35) and generate synthetic data to compare. The comparisons between scatter plot of the real datasets and synthetic data generated by the PCSDE model are shown in Figure 4.11. The results show that our synthetic data with different parameters fit different social network datasets well.



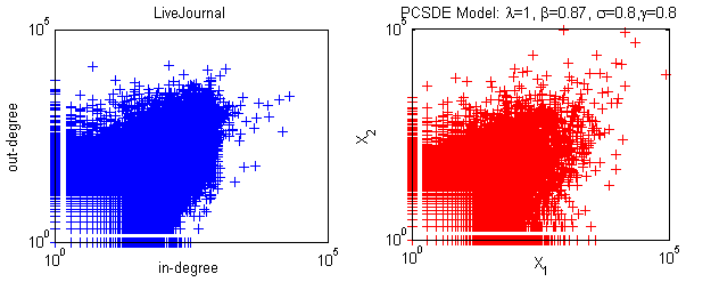
(a) Youtube



(b) Facebook



(c) Flickr



(d) Livejournal

Figure 4.11. Synthetic data from the second Type 3 model compared to real datasets

Graph	$E[d_{in}]$	$\sigma^2_{d_{in}}(\sigma^2_{d_{out}})$	$E[d_{in}d_{out}]$	Corrcoef
Youtube	4.34	2.37e+003 (1.61e+003)	1.88e+003	0.95
Facebook	18.7	1.91e+003 (2.17e+003)	2.08e+003	0.85
Flickr	14.4	1.25e+004 (1.00e+004)	8.66e+003	0.76
Livejournal	14.1	1.30e+003 (1.88e+003)	1.21e+003	0.65

Table 4.2. Statistics in social network datasets

4.5 Summary

In this chapter, we proposed two types of models. Type 3 models are based on the idea that two variables are correlated if they have the same life time but independent random terms. By compared to the real data, the second Type 3 model based on CEV with $0 < \gamma < 1$ fits the real data in social networks.

We also find that the synthetic data from the fist Type 3 model in Figure 4.6 with $\sigma = 1$ looks similar to the Web Google dataset in Figure 3.2. We compare the two datasets in Figure 4.12. Thus, the first Type 3 model might be useful in explaining some bivariate power law data in citation or Web networks. To note that, all the fitting work for the Type 3 model is not a rigorous fitting. The reason is that it is pretty hard to get the theoretical joint distribution for Type 3 model. And it is even not possible to compute the second and higher order moments of the second Type 3 model with $\gamma \neq 1/2$.

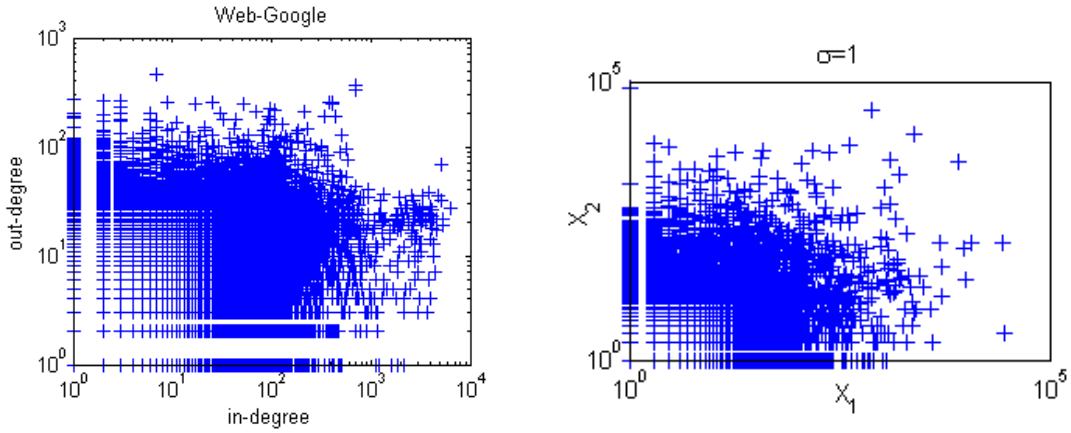


Figure 4.12. Synthetic data from the first Type 3 model v.s. Web Google dataset

CHAPTER 5

APPLICATIONS: NETWORK GROWING MODELS, NATURAL IMAGES

The second Type 3 model based on CEV ($0 < \gamma < 1$) is an interesting model since it fits the real data in social network in distribution and it generates fractional TDC. In this chapter, we first connect a special case ($\gamma = 1/2$) of this model to the network growing models. Since the network growing models are only designed to explain the origin of power law in complex networks, our model serves as a more general explanation to all the power law observations. For example, bivariate power law distributions in natural images.

5.1 Network growing models and PCSDE models

We start with a generalized network growing model and then take Bollobás' model as an example.

5.1.1 A generalized network growing model

Consider the following generalized model. In each step

- the total number of new nodes added to the existing graph is n;
- the total number of new degree (in-degree/ out-degree for directed graph) added is m with m_1 the degree attached to the new nodes and m_2 the degree attached to nodes in existing graph;

- when selecting the node in existing graph to attach to, we use preferential attachment mechanism, say the node is selected with probability proportional to the node's current degree.

Our target is to use our PCSDE model to simulate the degree growth process of a randomly selected node in the graph. First, the PCSDE model assumes the life time follows an exponential distribution. However, the nodes normally grow linearly in network growing models since a constant number of nodes and edges are added in each step. In real life, the size of online social networks normally grows exponentially at the rapid growing period [25, 67] (as shown in Figure 5.1). Thus, it is reasonable to assume the nodes grows exponentially, which means the new nodes added in each "*time step*" is proportional to the current number of nodes in the existing graph. Note that *time step* is different from *step* in growing models.

Given initial number of nodes in the network n_0, denote $N(t)$ the number of nodes in the network at time t

$$N(t) = n_0 e^{\lambda t}, \tag{5.1}$$

which yields

$$dN(t) = \lambda n_0 e^{\lambda t} dt = \lambda N(t) dt. \tag{5.2}$$

At a very large T, we observe the nodes' life time. Denote the life time as a random variable L. With the assumption in Equation (5.1), we have

$$\begin{aligned} \mathbb{P}(L > t) &= \frac{\text{Number of nodes appear before time } T-t}{\text{Total number of nodes at time } T}, \\ &= \frac{N(T-t)}{N(T)}, \\ &= \frac{n_0 e^{\lambda(T-t)}}{n_0 e^{\lambda T}}, \\ &= e^{-\lambda t}. \end{aligned} \tag{5.3}$$

Thus, $L \sim \exp(\lambda)$.

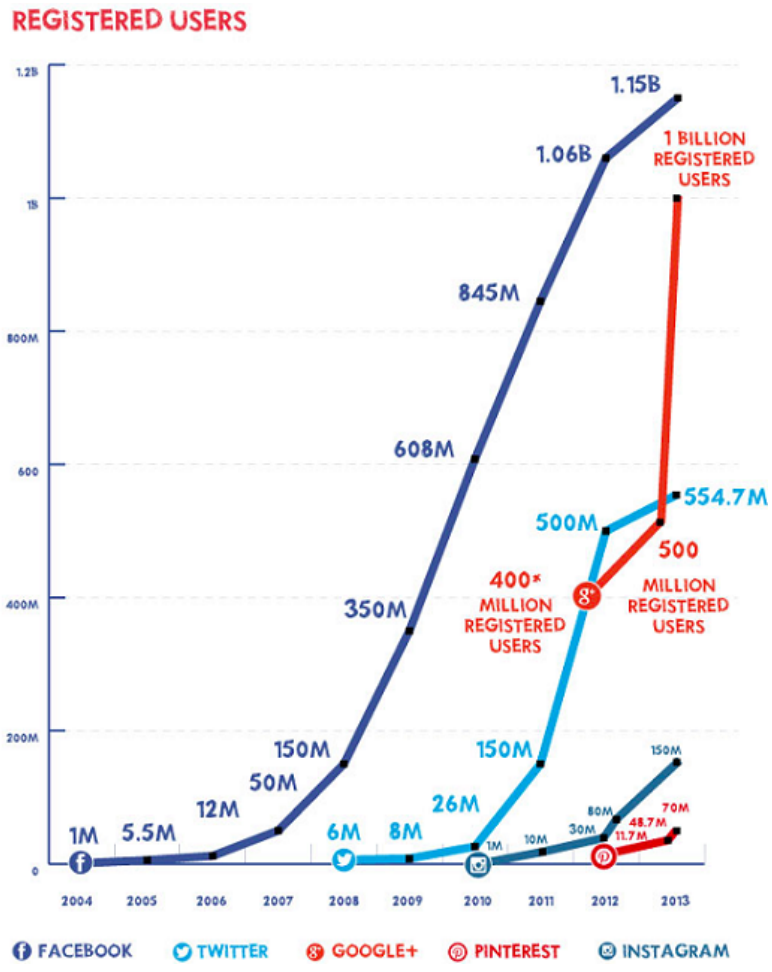


Figure 5.1. Registered users in social networks from 2004 to 2013 (Figure is copied from [67])

Next, denote $M(t)$ as the total degree in the graph at time t. Assume the initial number of degree in the graph is m_0 and $m_0 = \frac{m}{n} n_0$, then we have

$$M(t) = \frac{m}{n} N(t) = m_0 e^{\lambda t}, \tag{5.4}$$

and

$$dM(t) = \lambda m_0 e^{\lambda t} dt = \lambda M(t) dt. \tag{5.5}$$

Now, consider a randomly selected node in the existing graph at time t with degree $D(t)$, we denote the new degree added to this node in dt as $dD(t)$. For each one new degree, this node is selected with probability $p = \frac{D(t)}{M(t)}$. And the total degree added to the nodes in existing graph is $\frac{m_2}{m} dM(t)$. Thus $dD(t)$ follows binomial distribution

$$dD(t) \sim B\left(\frac{m_2}{m} dM(t), \frac{D(t)}{M(t)}\right), \tag{5.6}$$

with expectation (denote $\beta \triangleq \frac{m_2}{m}\lambda$)

$$E(dD(t)) = \frac{m_2}{m} dM(t) \frac{D(t)}{M(t)} = \frac{m_2}{m} \lambda D(t) dt = \beta D(t) dt. \tag{5.7}$$

When $p = \frac{D(t)}{M(t)} \to 0$, we have $p(1-p) \to p$. Since $M(t)$ grows faster than $D(t)$, we have $\frac{D(t)}{M(t)} \to 0$ as $t \to \infty$. Thus

$$Var(dD) = \frac{m_2}{m} dM(t) \frac{D(t)}{M(t)} \left(1 - \frac{D(t)}{M(t)}\right) \to \frac{m_2}{m} dM(t) \frac{D(t)}{M(t)} = \beta D(t) dt. \tag{5.8}$$

If we let $D(0) = 0.1M(0)$, which is the case in our experiments, the variance can be approximated by the above equation even when t is very small.

We use the following normal distribution to approximate the above binomal distribution

$$dD(t) \sim B\left(\frac{m_2}{m} dM(t), \frac{D(t)}{M(t)}\right) \approx \mathcal{N}\left(\beta D(t) dt, \beta D(t) dt\right). \tag{5.9}$$

Given standard Brownian motion $dW(t) \sim \mathcal{N}(0, dt)$, we have the following SDE

$$dD(t) = \beta D(t)dt + \sqrt{\beta D(t)}dW(t). \tag{5.10}$$

Combing with the conclusion drawn from Equation (5.3), we use the following PCSDE to simulate the degree evolution in the graph.

Let X denote the undirected degree in symmetric graph (in-degree or out-degree for directed graph). We have

$$dX(t) = \beta X(t)dt + \sqrt{\beta X(t)}dW(t) + (x_0 - X(t))dN(t), \tag{5.11}$$

where $N(t)$ is a Poisson counter with rate λ and

$$\lambda = \frac{m}{m_2}\beta. \tag{5.12}$$

Equation (5.11) is a special case of PCSDE model of the second Type 3 in Equation (4.34) with $\gamma = 1/2$ and $\sigma = \sqrt{\beta}$. With the conclusion in Equation (4.38), the marginal density is $f_X(x) \sim x^{-(\frac{\lambda}{\beta}+1)}$. With the relationship between λ and β in Equation (5.12), we have

$$f_X(x) \sim x^{-(\frac{m}{m_2}+1)}. \tag{5.13}$$

For B-A model in [8], it attaches half of the undirected edges to the new node and half to the nodes in the existing network in each step, as shown in Figure 5.2. With $m_2 = \frac{1}{2}m$, we have the tail exponent $1 + \frac{m}{m_2} = 3$ for B-A model. The result is consistent with the result in [8].

5.1.2 A directed example: Bollobás' model

Consider the following directed network growing model in [9]. In [87], the authors proved that the joint distribution of this model has jointly regularly varying tails. In this model, the network grows by adding a new node or a new directed edge in each step.

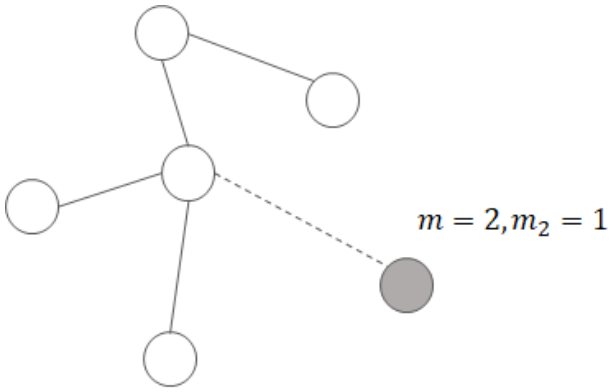


Figure 5.2. Illustration of growing process in B-A model (white circle: existing nodes; gray circle: new nodes; solid line: existing connections; dashed line: new connections)

- With probability p_1, append a new node to the graph with an edge from an existing node in the graph to the new node.
- With probability p_2, append a new node to the graph with an edge from the new node to an existing node in the graph.
- With probability $q = 1 - p_1 - p_2$, append to the existing graph a directed edge from v to w.
- The initiating node v is chosen with probability depending on its out-degree;.

$$p(v \text{ is chosen}) = \frac{D_{out}(v) + \epsilon_{out}}{\sum_u (D_{out}(u) + \epsilon_{out})},$$

The targeting node w is chosen with probability depending on its in-degree

$$p(w \text{ is chosen}) = \frac{D_{in}(w) + \epsilon_{in}}{\sum_u (D_{in}(u) + \epsilon_{in})}.$$

In the directed model, in-degree and out-degree grow separately. However, the life time of the two is the same. The in-degree and out-degree bias ϵ_{in} and ϵ_{out} can be

seen as extra initial in-degree and out-degree given to each node. When introducing a new node with an in-going link, the new node start with $1+\epsilon_{in}$ in-degree and ϵ_{out} out-degree; while with an out-going link, the new node start with ϵ_{in} in-degree and $1+\epsilon_{out}$ out-degree. The expectation of in-degree added at each step is $m^{in} = 1 + \epsilon_{in}(p_1 + p_2)$ with $m_2^{in} = p_1 + q$ attached to nodes in existing graph; while the expectation of out-degree added at each step is $m^{out} = 1 + \epsilon_{out}(p_1 + p_2)$ with $m_2^{out} = p_2 + q$ attached to nodes in existing graph. With the result in Section 5.1.1, we obtain the following bivariate PCSDE model:

$$\begin{aligned} dX_1 &= \beta_1 X_1 dt + \sqrt{\beta_1}\sqrt{X_1} dW_1 + (1 + \epsilon_{in} - X_1) dN_1 + (\epsilon_{in} - X_1) dN_2; \\ dX_2 &= \beta_2 X_2 dt + \sqrt{\beta_2}\sqrt{X_2} dW_2 + (\epsilon_{out} - X_2) dN_1 + (1 + \epsilon_{out} - X_2) dN_2, \end{aligned} \tag{5.14}$$

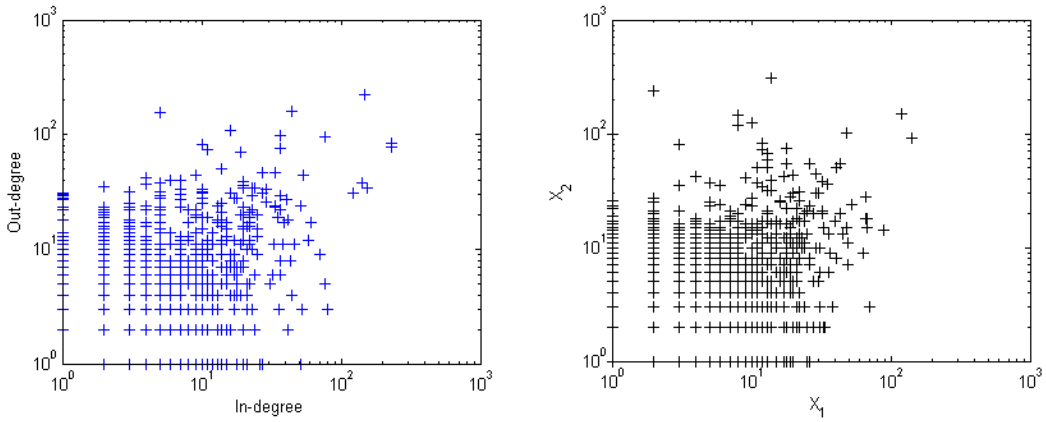
where N_1 and N_2 are independent Poisson counters with rates λ_1 and λ_2, with the relationship:

$$\begin{aligned} \lambda &= \lambda_1 + \lambda_2, \\ \lambda_1/\lambda_2 &= p_1/p_2. \end{aligned}$$

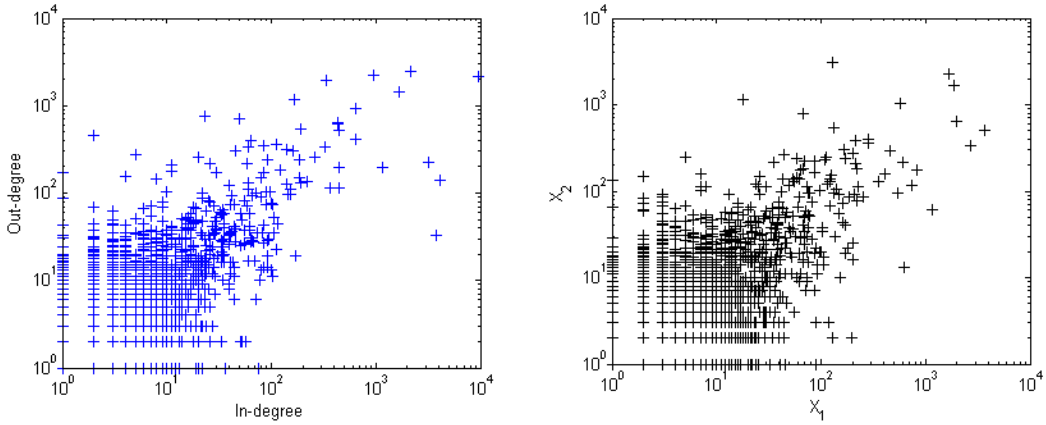
and

$$\begin{aligned} \beta_1 &= \frac{p_1 + q}{1 + \epsilon_{in}(p_1 + p_2)}\lambda, \\ \beta_2 &= \frac{p_2 + q}{1 + \epsilon_{out}(p_1 + p_2)}\lambda. \end{aligned}$$

The marginal tail exponents can be computed with Equation (4.39), which are the same with the results in [9] and [87]. We also prove the asymptotic dependence between in-degree and out-degree with the conclusion in Section 4.3.3.



(a) $N = 5000, p_1 = p_2 = 0.2, \epsilon_{in} = \epsilon_{out} = 1$



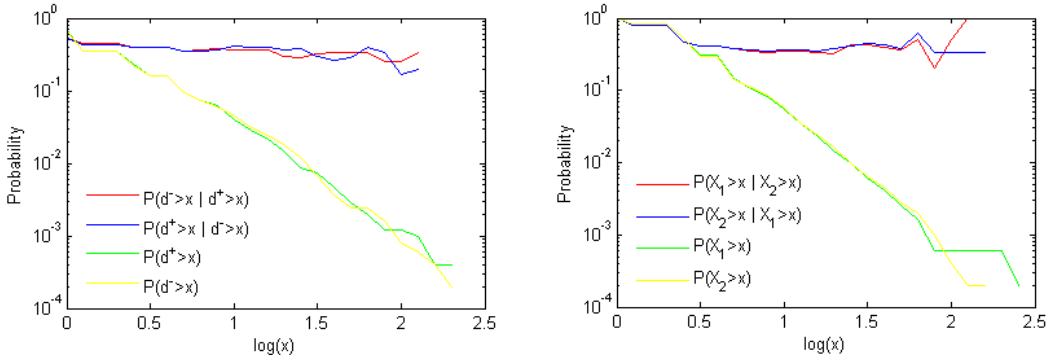
(b) $N = 5000, p_1 = p_2 = 0.05, \epsilon_{in} = \epsilon_{out} = 1$

Figure 5.3. Comparison between the data generated from Bollobás' model and the data from PCSDE model in (5.14) (**Blue**: Bollobás model; **Black**: PCSDE model).

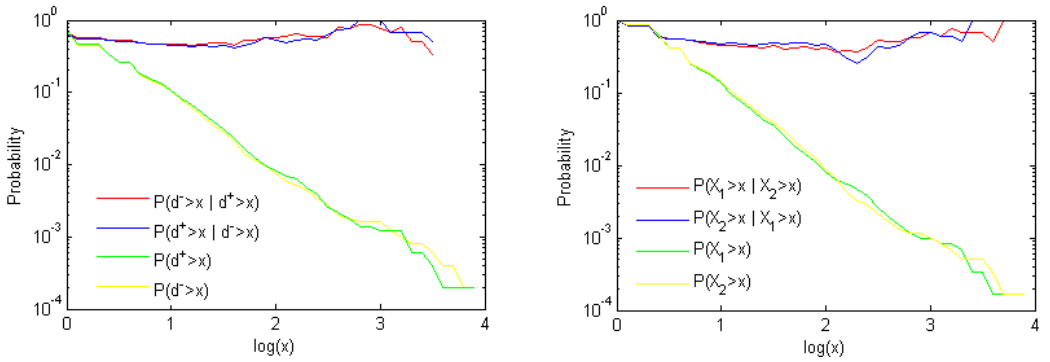
5.1.3 Experimental results

We generate two groups of data using the Bollobás' model in Section 5.1.2 with $p_1 = p_2 = 0.2$ in the first group and $p_1 = p_2 = 0.05$ in the second group. Let $\lambda = 1$ in bivariate PCSDE model in Equation (5.14) and we compute the other corresponding parameters. We have $\beta_1 = \beta_2 = 0.5714$ for the first and $\beta_1 = \beta_2 = 0.8636$ for the second. The total number of nodes N in each group is 5000.

The scatter plot of the synthetic data from the growing model and from the PCSDE model are compared in Figure 5.3. We observe that the data from the PCSDE model looks quite similar to the data from growing model. To illustrate



(a) $N = 5000, p_1 = p_2 = 0.2, \epsilon_{in} = \epsilon_{out} = 1$



(b) $N = 5000, p_1 = p_2 = 0.05, \epsilon_{in} = \epsilon_{out} = 1$

Figure 5.4. Comparison of the CCDF and dependence coefficients between the data generated from Bollobás' model and the data from PCSDE model in (5.14) (Left column: Bollobás model; Right column: PCSDE model).

the two sets of data following the same distribution, we further compare the CCDFs and dependence coefficients between the synthetic data from the Bollobás' model and from the PCSDE model. As shown in Figure 5.4, the two sets of data have the same marginal tail exponents and the dependence coefficients stay at the same value when the value of x is large.

Thus, our PCSDE model can be used to explain the degree growth of a randomly selected node in the graph. Since the simulation process of generating new nodes and new connecting edges is very time consuming, we could use our PCSDE model to generate synthetic data instead of using the original growing model to build the whole graph.

5.1.4 Some thoughts about new network growing models based on Type 3 model

We have shown that a special case ($\gamma = 1/2$) of our second Type 3 model connects to network growing models. Could we come up with new network growing models, which connect to the other cases of our PCSDE model?

The problem in the existing network growing model is that each new degree is added randomly and independently. The accumulated variance is fixed and is just the case in PCSDE model when $\gamma = 1/2$. What about the other cases? Let's see the growing part of our PCSDE model, the CEV model

$$\frac{dX}{X} = \beta dt + \frac{\sigma}{X^{1-\gamma}} dW.$$

In CEV model, the ratio between the increment value and original value $\frac{dX}{X}$ follows a normal distribution with a constant mean. The instantaneous variance is also constant when $\gamma = 1$. When $\gamma < 1$, the instantaneous variance decreases when X increases.

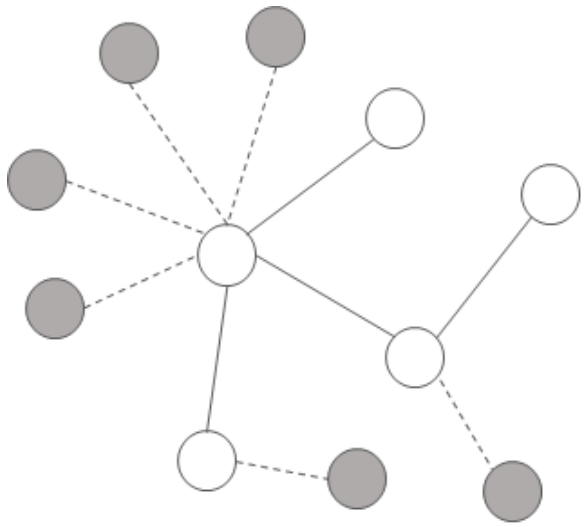


Figure 5.5. Illustration of new network growing model based on Type 3 model (white circle: existing nodes; gray circle: new nodes; solid line: existing connections; dashed line: new connections)

So, we let the existing nodes attract a number of new nodes in each step in our new network growing model, which gives a real exponential growth. As shown in Figure 5.5, the number of new nodes a given node attracted depends on its current degree. Instantaneous variance decreases with degree actually makes sense in real social networks. For a node with large degree, new connections are normally coming for a reason (a fan, interesting content); so it has smaller instantaneous variance compared to nodes with smaller degree. γ value reflects the property of the network. Smaller γ means faster drop with degree, which further indicates that famous nodes have bigger impact in such kind of networks.

To summarize, here is only some brief thoughts about how to interpret our Type 3 model in real network growing scenario. And we only talk about new growing models for undirected networks. For directed network growing models, it will be more complicated. In the next section, we will present another interesting application of our Type 3 model in natural image.

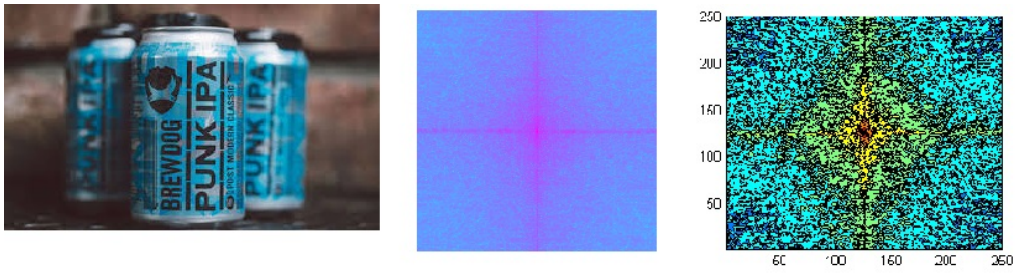


Figure 5.6. Power spectrum of a natural image (image, power spectrum, contour of power spectrum)

5.2 Power law in natural image

5.2.1 Background

Natural images are highly related to power law phenomenon. Average power spectra of natural images normally follow a power law [89, 90, 68] (as shown in Figure 5.6). In [90], authors present spectral signatures of different image categories (as shown in Figure 5.7). We observe that images with natural and man-made objects are different in power spectra shapes.

Power law spectrum is also observed in internet traffic [62]. In [62], an aggregation of Markovian Hierarchical On-Off Processes is proposed to model the internet traffic. However, how to extend this model to two dimension to apply it to two dimensional images is unclear. In our work, we focus on a more straight forward explanation: occluding objects with power law distributed sizes.

As explanations in [62] for power law power spectrum in internet traffic, a provocative interpretation of the power law spectra in natural images is self-similarity [68, 86, 60]. A self-similar object is similar to a part of itself. There is evidence that the distribution of the object sizes is self-similar in natural image and explanation models have been delivered in [86, 60]. In their models, the images are composed of independently occluding objects with constant intensities, whose sizes follow power

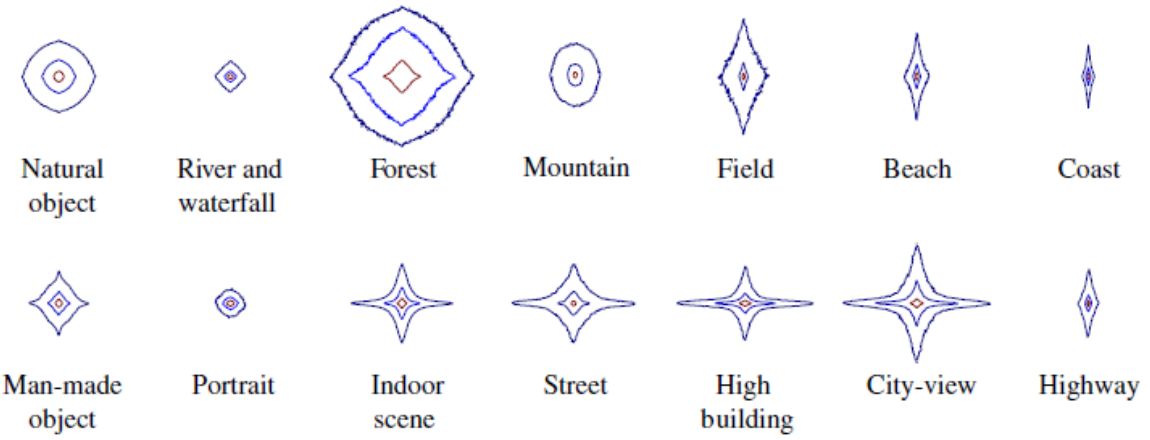


Figure 5.7. Spectral signature of different image categories (image comes from [90])

law distribution. To model self-similar images, power law size distributed circles and squares are generated [86, 60] (as shown in Figure 5.8).

Power law sized cluster distribution, on the other hand, has been found in the k-bilevels of natural images [5, 39]. In the works, the authors equally divided pixel levels $[0, 255]$ into k regions. The images could be represented by k-bilevels. A k-bilevel of a binary images ($l = 1, \ldots, k$) is as follows:

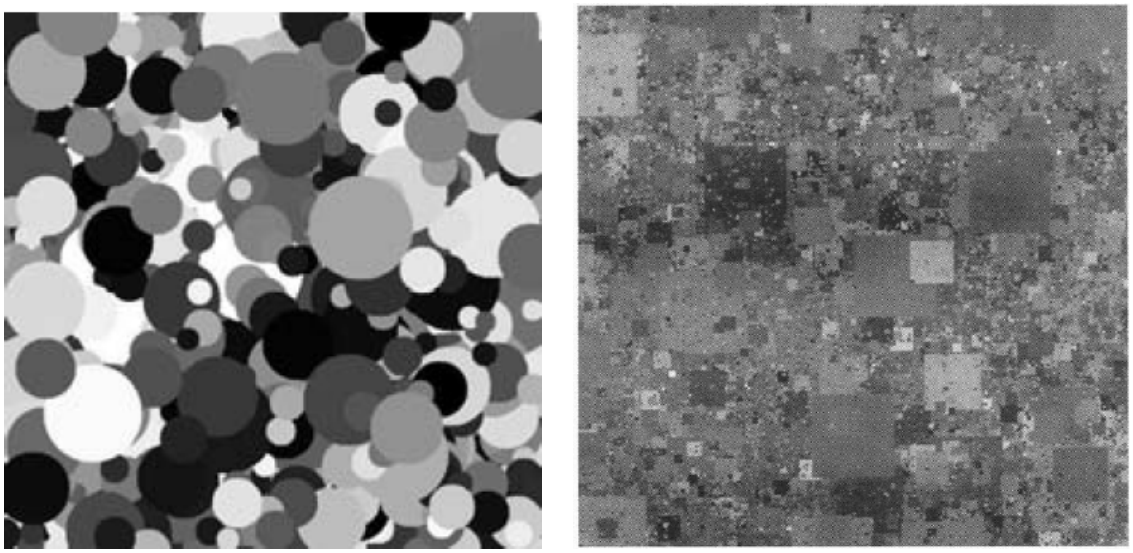
$$I_l(i,j) = \begin{cases} 1 & I(i,j) \in [(l-1)\frac{N}{k}, l\frac{N}{k}], \\ 0 & \text{otherwise.} \end{cases}$$

From experiments, the size distribution of connected components in all the k-bilevels of an image can be approximated by a power law distribution (as shown in Figure 5.9).

The works in [5, 39] are evidences for the power law sized objects in [86, 60].

5.2.2 Bivariate power law in natural images

In [86, 60], the objects are represented by squares and circles. In fact, the shape of the connected components in natural image is not always regular. We use different



(a) (b)

Figure 5.8. Model self-similar images ((a) from "Scaling and Power Spectra of Natural Images" (2003); (b) from "Origins of Scaling in Natural Images" (1997))

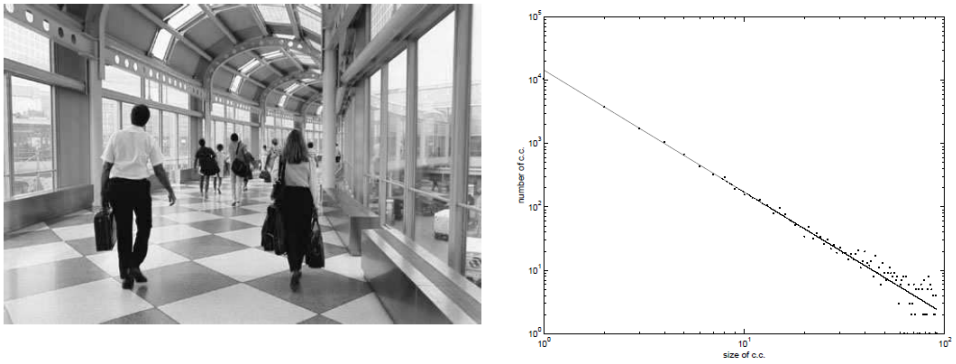


Figure 5.9. The distribution of sizes of connected components of 10-bilevels (from Gousseau, Y. and Morel, J.M.(2001) "Are natural images of bounded variation?")

colors to mark the connected components of natural images using k-bilevel method with $k = 8$ (as shown in Figure 5.10).

We observe that the shapes of the objects in real images are quite arbitrary. The squares and circles with fix diameters are not good enough in capturing the shapes of the objects in natural images. Since images have two dimensions, we are wondering whether there exists bivariate power law data in natural images?

We measure the height (from the north-est pixel to the south-est pixel) and the width (from the west-est pixel to the east-est pixel) of each connected component and plot a scatter plot of $(height, width)$ jointly (as shown in Figure 5.11). Meanwhile, we plot the marginal CCDF and the dependence coefficients between the two variables in Figure 5.12. As shown in Figures 5.11 and 5.12, $(height, width)$ pair of the connected components in most of the natural images follow power law marginally and their joint densities and dependence behaviors look similar to the synthetic data generated by the second Type 3 model.

5.2.3 Model self-similar in natural images

The observation in Section 5.2.2 motivates us to propose a new model to generate self-similar images. We believe that using rectangles and ellipses instead of squares and circles should be better choices. We use rectangles to simulate man-made objects, like building, futurities, etc. The width and height of a rectangle follows joint distribution in the second Type 3 model. We use ellipses to simulate natural objects, like leaves, flowers, etc. And the major axis and minor axis of the ellipse follows joint distribution in the second Type 3 model.

Two self-similar images generated using rectangles and ellipses are shown in Figures 5.13 and 5.14. We also plot the power spectra and the contour of the spectra. Compared to the results in Figure 5.7, the power spectrum contour of the image with rectangles has similar shape to the spectral signature of image with natural object;

(a) Image with natural objects

(b) Clusters of image with natural objects

(c) Image with man-made objects

(d) Clusters of image with man-made objects

Figure 5.10. Connected components ($k = 8$) in images with man-made and natural objects

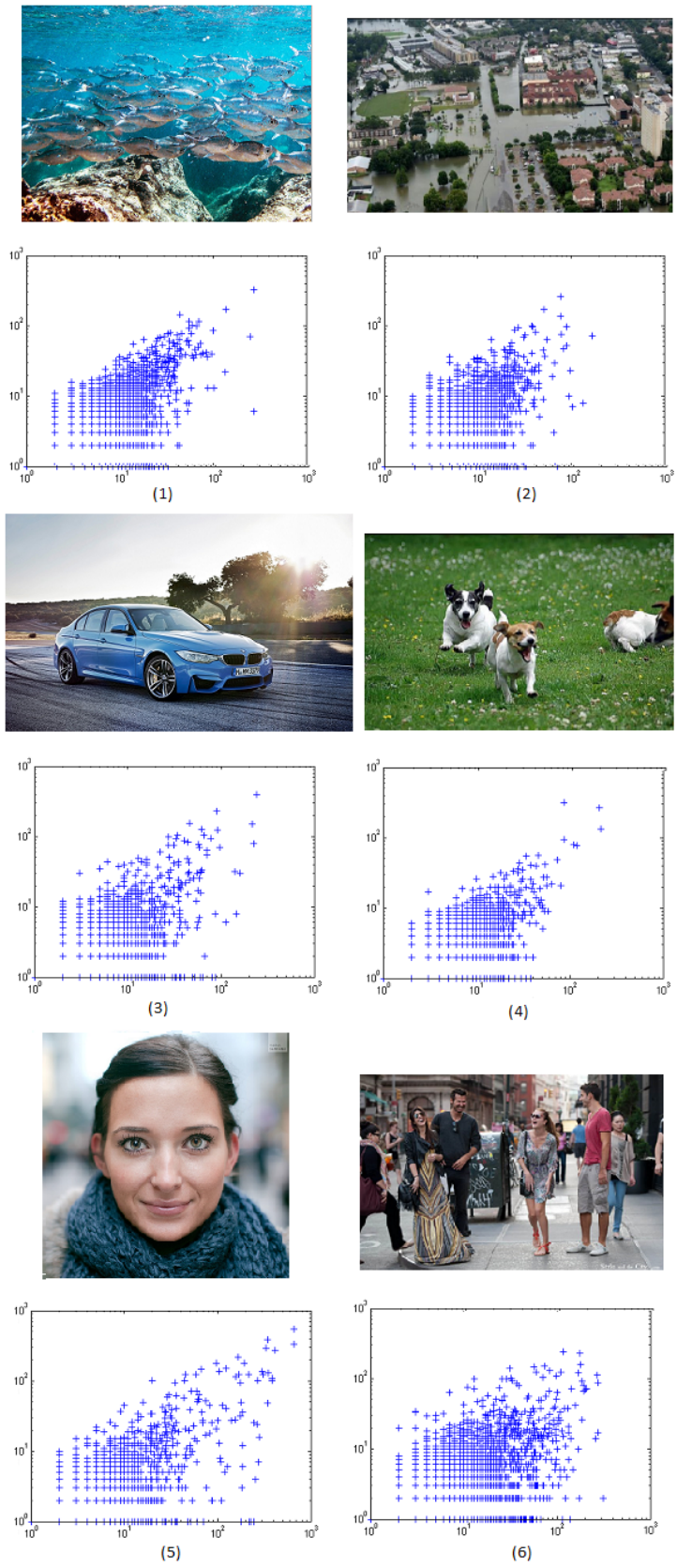
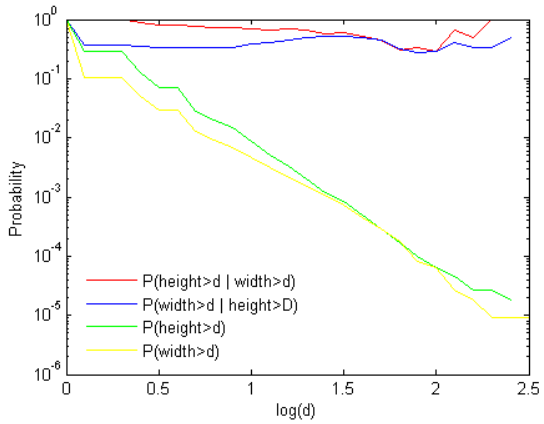
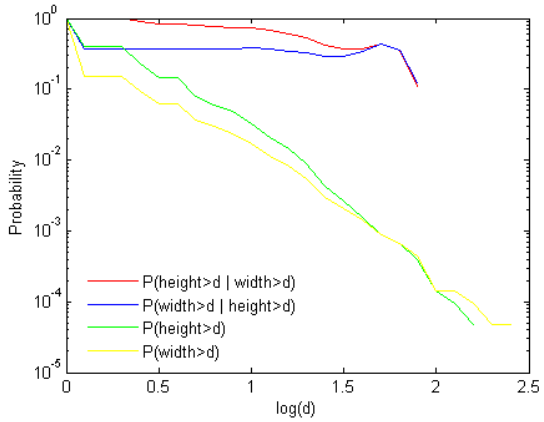


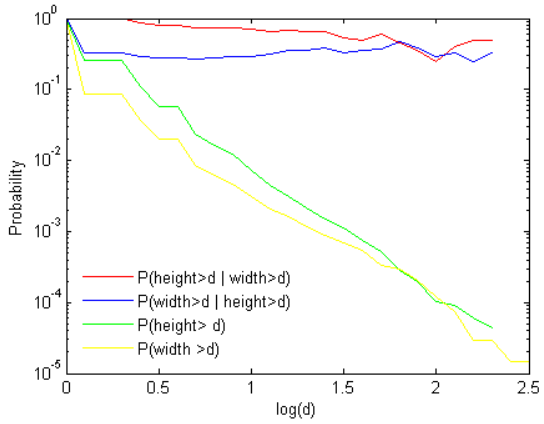
Figure 5.11. Scatter plot of $(height, width)$ of connected components $(k = 8)$ in natural images



(a) Image (1)



(b) Image (2)



(c) Image (3)

Figure 5.12. Images: marginal CCDF $P(height > d)$ $(P(width > d))$ and dependence coefficients $P(height > d|width > d)$ and $P(width > d|height > d)$ as a function of value d

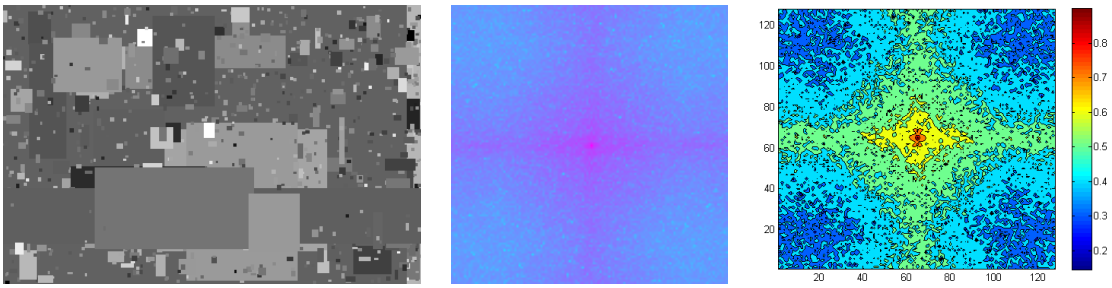


Figure 5.13. Model self-similar images with rectangles

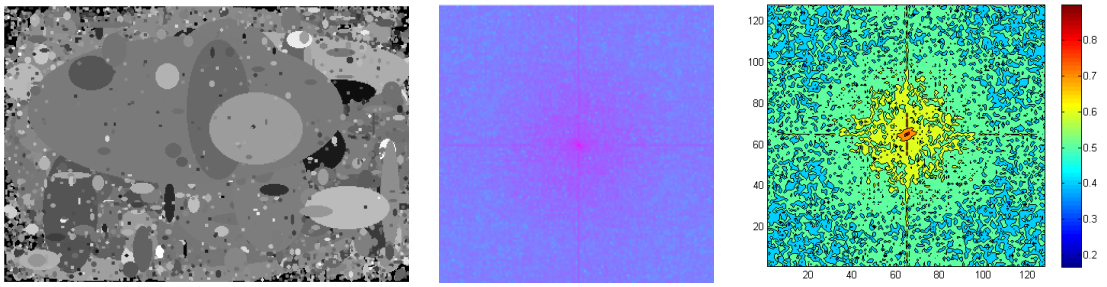


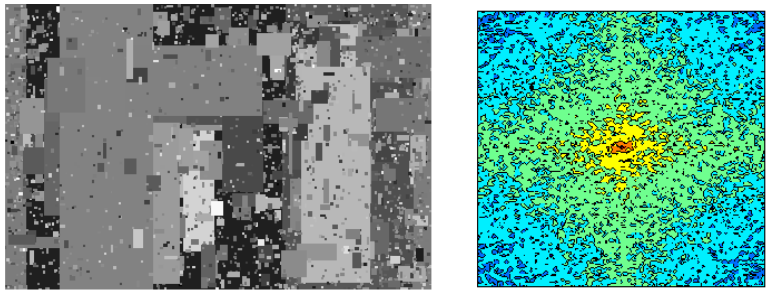
Figure 5.14. Model self-similar images with ellipses

while the power spectrum contour of the image with ellipses has similar shape to the spectral signature of image with man-made object.

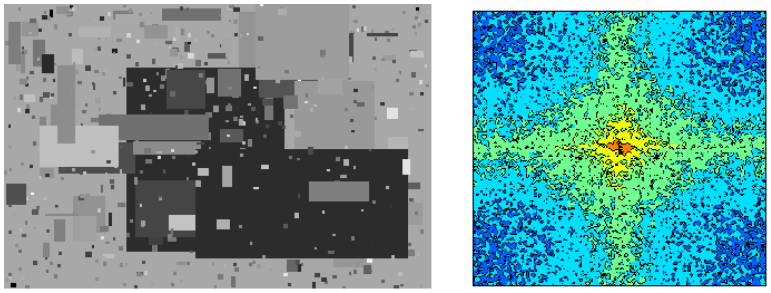
Meanwhile, by changing the parameters in the PCSDE model, we can also simulate different scene scale in real images (as shown in Figure 5.15). As indicated in [90], close-up views on man-made objects tend to produce images with flat and smooth surfaces. As distance between the observer and the scene background increases, it is likely to encompass more objects in the image. Thus, the power spectra for close-up views is concentrated mainly in low spatial frequencies. While, as the distance increases, the spectral energy corresponding to high spatial frequencies increases. In our experiments, larger ϵ corresponds to close-up scene.

5.2.4 Summary

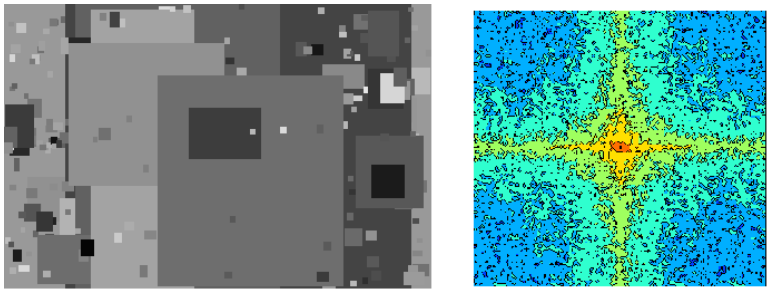
In this part, we show two major applications of our second Type 3 model.



(a) $\epsilon = 0.6$



(b) $\epsilon = 1$



(c) $\epsilon = 3$

Figure 5.15. Model self-similar images with different scene scales

In the first application, our model can be used to generate synthetic data to save the time in generating a whole connected graph. To note that, we can only get the in-degree and out-degree pairs of the nodes in the network. However, the information of the connections between nodes are lost.

In the second application, we observe the width and height of the connected components in an image follow bivariate power law distribution similar to the synthetic data generated by the second Type 3 model. Based on this observation, we propose a modified model to generate self-similar images. Our new model with rectangle and ellipse performs better in capturing the shapes of connected components in natural images.

Since power laws are observed in many other fields, we believe our PCSDE models have other potential applications. And we are curious to learn more in our future works.

CHAPTER 6

DISCUSSIONS

In this part, we conclude our work and discuss the pros and cons in the work. This dissertation focused on solving problems related to complex graphs. The first problem is to propose similarity testing algorithms for complex graphs. The second problem is to study generative models for bivariate power law distributions observed in social networks. In this chapter, we first summarize the major contributions in this dissertation. We discuss the limitations in our work in the second section. At last, we draw conclusions and present future works.

6.1 Contributions

- We proposed a fast and effective graph comparison algorithm based on heat content. Heat content transfers graph connectivity information into a one dimensional feature. We compute a fixed and short length of heat content feature to compare for all graphs with different structures. Evidence proves that heat content feature is more effective in graph comparison compared to some existing features like degree distribution and eigenvalues. Like many other feature extraction methods, heat content allows us to compare graphs of different sizes. And we do not need the nodes' mapping information between two graphs. Meanwhile, random walk estimation method provides possibility of parallel computing and scaling. With all these features, we could apply our similarity testing algorithm to other comparison tasks, such as image retrieval and classification. We will discuss this more when we discuss the limitations in our algorithm.

- We proposed several different bivariate PCSDE models to generate correlated bivariate power law distributions. We analyzed and discussed the property of each model. We put emphasis on Type 3 model, since this type of bivariate PCSDE models are useful in explaining the existence of bivariate power law data in social networks, web networks and natural images. By connecting Type 3 model to existing network growing models, our PCSDE model provides an quick way to study the joint density and tail dependence behavior of a network growing model. Although the other two types of proposed PCSDE models fail in generating power law distributions similar to real network datasets, they are interesting theoretically and may have potential usages in other fields.

6.2 Limitations

6.2.1 Limitations in our graph similarity testing algorithm

To do more complicated classification tasks like image classification problems, we need to build a graph based on image pixel levels at first. Then, we compute the heat content of the generated graph.

The first method is to classify images using only heat content information. We say two images are similar if their heat contents are close to each other. A problem of this method is that our heat content method is fixed when the graph is given, which means our model lacks of the flexibility to apply a training process. A training process is common in the most popular image classification algorithms, like logistic regression, neural networks and deep learning. Without a training process, the precision of our classification result can not be improved by tuning the parameters.

The second method is to use heat content as a one dimensional feature of the image. Combining with other features, we apply classification algorithms like k-nearest-neighbor, logistic regression to classify images. Some image classification

works were done based on this approach. The reader is referred to [46, 47] for more details.

The problem is that, if we use heat content as a feature, generating this feature is time consuming compared to using raw pixel levels as an input. The graph generation process also has a scaling problem. The size of the graph is determined by the number of pixels in the image; and the number of edges in the graph is the square of the number of pixels. To save the time in building graphs, we have resized all the images to a fixed small size, which reduced the precision of our algorithm.

6.2.2 Limitations in our bivariate PCSDE generative models

Type 1 and Type 2 models are interesting theoretically. However, we have not found any data show the same distributions as the two models. Type 2 model is a special model. This model gives perfect theoretical results including power law marginal tail and non-zero tail dependence coefficients. However, this model generates weird synthetic data. When the value becomes large, there is no data with two variables sharing the same value. This does not obey our common sense.

Type 3 model gained great success in generating data similar to real data in Web and social networks. The problem is that the Brownian motion component in the PCSDE model makes it difficult to compute the theoretical marginal and joint density. Meanwhile, we can only compute the second moments of the model under some special cases ($\gamma = 1/2$). Lacking of theoretical results make it difficult to fit real data sets in social networks. We do experiments by tuning the parameters manually to make the datasets look similar to the real datasets in shapes.

For the applications of Type 3 model in natural images, we have shown our new model to generate self-similar images using rectangles and ellipses. Actually, for real images the shape might be very irregular. We did not come up with a solution to generate random shaped object with a given width and height in this thesis.

6.3 Conclusion and future works

To summarize, this thesis discussed two major topics related to power law graphs: similarity testing and the origin. The work included both theoretical results and experimental results. The importance of our work is presented by discussing possible applications of our work in image retrieval, social network data fitting, network growing model explanation, self-similar image generation, etc.

We conclude the whole thesis by discussing our future works. The future works are mainly based on the limitations we discussed in the previous section.

- We seek improvement to our heat content method. For example, we can amplify the asymmetric part of the graph Laplacian to create a new kind of oscillatory heat content. The weight for the asymmetric part can be used as a training parameter. In [47], two new feature extraction methods were introduced.

- We look for other applications of our graph similarity test algorithm. For example, we could use spectrogram of an audio sample as an image and do audio retrieval by using the same technique in our image retrieval. We did some small experiments in [46] and the results showed the potential of our method to succeed in more complicated tasks.

- For models to generate self-similar images, we could propose a solution to generate random shaped object with a given width and height to mimic real objects in the nature.

- We will explore applications of the Type 3 models in many other fields; and we are also curious about the applications of our first two types of models.

- We could extend our PCSDE generative models to higher dimensions, such as three dimensions. In a color image, the three rgb channels make it possible to produce three dimensional power law data.

APPENDIX A

ANOTHER CALCULATION METHOD FOR THE MODIFIED TYPE 1 MODEL

Consider X_1 first. Let us consider the sequence of times when X_1 is revert to ϵ. Let $P_{on}^{(1)}$, $P_{off}^{(1)} = P_{on}^{(1)}$ be the probabilities that, at a given one of these times, the ON-OFF system is ON (correspondingly, OFF).

For $t > 0$ let $h_{on}(t)$ be the probability that Y_1 is not revert to ϵ again, given that it starts in an ON period, due to N_1. The meaning of $P_{off}(t)$ is analogous.

We start with the asymptotic analysis of $P_{on}(t)$. We claim that

$$P_{on}(t) \sim a_{on} e^{-\beta t}, t \to \infty, \tag{A.1}$$

where

$$\beta = \frac{(\lambda_1 + \lambda_3 + \lambda_4) - \sqrt{(\lambda_1 + \lambda_3 + \lambda_4)^2 - 4\lambda_1\lambda_3}}{2}. \tag{A.2}$$

To this end, we will write a renewal equation for $P_{on}(t)$. We have

$$P_{on}(t) = e^{-(\lambda_1+\lambda_4)t} + \int_0^t (\lambda_1+\lambda_4) e^{-(\lambda_1+\lambda_4)x} dx\theta \left[e^{-\lambda_3(t-x)} + \int_0^{t-x} \lambda_3 e^{-\lambda_3 y} dy P_{on}(t-x-y) \right], \tag{A.3}$$

where

$$\theta = \frac{\lambda_4}{\lambda_1 + \lambda_4}. \tag{A.4}$$

We rewrite A.3 in the form

$$P_{on}(t) = z(t) + \int_0^t f(x) P_{on}(t-x) dx, \tag{A.5}$$

where

$$z(t) = e^{-(\lambda_1+\lambda_4)t} + \theta \int_0^t (\lambda_1+\lambda_4) e^{-(\lambda_1+\lambda_4)x} e^{-\lambda_3(t-x)} dx$$
$$= e^{-(\lambda_1+\lambda_4)t} + \theta \frac{\lambda_1+\lambda_4}{\lambda_3-\lambda_1-\lambda_4}(e^{-(\lambda_1+\lambda_4)t} - e^{-\lambda_3 t}) \quad (A.6)$$

if $\lambda_3 \neq \lambda_1 + \lambda_4$, and

$$z(t) = e^{-\lambda_3 t} + \theta\lambda_3 t e^{-\lambda_3 t} \quad (A.7)$$

if $\lambda_3 = \lambda_1 + \lambda_4$.

Further,

$$f(x) = \theta\lambda_3(\lambda_1+\lambda_4) \int_0^x e^{-(\lambda_1+\lambda_4)y} e^{-\lambda_3(x-y)} dy,$$
$$= \theta\lambda_3(\lambda_1+\lambda_4) e^{-\lambda_3 x} \int_0^x e^{(\lambda_3-\lambda_1-\lambda_4)y} dy,$$
$$= \theta \frac{\lambda_3(\lambda_1+\lambda_4)}{\lambda_3-\lambda_1-\lambda_4}(e^{-(\lambda_1+\lambda_4)x} - e^{-\lambda_3 x}), \quad (A.8)$$

if $\lambda_3 \neq \lambda_1 + \lambda_4$, and

$$f(x) = \theta\lambda_3^2 x e^{-\lambda_3 x}, \quad (A.9)$$

if $\lambda_3 = \lambda_1 + \lambda_4$.

In any case,

$$\int_0^\infty f(x)dx = \theta < 1. \quad (A.10)$$

We now use Prop. 3.11.1 in Reznick "Adventures in Stochastic Process" [83]. We are looking for $\beta \in R$ such that

$$\int_0^\infty e^{\beta x} f(x)dx = 1. \quad (A.11)$$

Suppose first that $\lambda_3 \neq \lambda_1 + \lambda_4$. Then (A.11) becomes, for

$$\beta < \min(\lambda_3, \lambda_1 + \lambda_4), \tag{A.12}$$

$$\begin{aligned} 1 =& \theta \frac{\lambda_3(\lambda_1+\lambda_4)}{\lambda_3-\lambda_1-\lambda_4}\left[\int_0^\infty e^{-(\lambda_1+\lambda_4-\beta)x}dx - \int_0^\infty e^{-(\lambda_3-\beta)x}dx\right], \\ =& \theta \frac{\lambda_3(\lambda_1+\lambda_4)}{\lambda_3-\lambda_1-\lambda_4}\left[\frac{1}{\lambda_1+\lambda_4-\beta} - \frac{1}{\lambda_3-\beta}\right], \\ =& \theta \frac{\lambda_3(\lambda_1+\lambda_4)}{(\lambda_1+\lambda_4-\beta)(\lambda_3-\beta)} = \frac{\lambda_3\lambda_4}{(\lambda_1+\lambda_4-\beta)(\lambda_3-\beta)}. \end{aligned} \tag{A.13}$$

That is, we obtain a quadratic equation for β,

$$\beta^2 - \beta(\lambda_1+\lambda_3+\lambda_4) + (\lambda_1+\lambda_4)\lambda_3(1-\theta) = 0, \tag{A.14}$$

or,

$$\beta = \frac{(\lambda_1+\lambda_3+\lambda_4) \pm \sqrt{(\lambda_1+\lambda_3+\lambda_4)^2 - 4\lambda_3(\lambda_1+\lambda_4)(1-\theta)}}{2}. \tag{A.15}$$

It is easy to check that only the smaller root satisfies (A.12).

This root is given by (A.2), and it is positive. The property (A.12) guarantees that the function

$$z^*(t) = e^{\beta t} z(t), t \geq 0. \tag{A.16}$$

So by Prop. 3.11.1 in Resnick we conclude that (A.1) holds.

The case $\lambda_3 = \lambda_1 + \lambda_4$ leads to the same quadratic equation and the same solution. Hence, (A.1) still holds.

APPENDIX B

PROVE THE DISTRIBUTION OF N AND F_j

B.1 Prove $N \sim Ge(1/2)$

$$\begin{aligned}
\mathbb{P}(N=k) &= \int_0^\infty \lambda e^{-\lambda x}(e^{-\lambda x}\frac{(\lambda x)^k}{k!})dx, \\
&= \int_0^\infty e^{-2\lambda x}\frac{(\lambda x)^k}{k!}d(\lambda x), \\
&= \int_0^\infty e^{-2x}\frac{x^k}{k!}dx, \\
&= \frac{1}{k!2^{k+1}}\int_0^\infty y^k e^{-y}dy, \\
&= \frac{1}{k!2^{k+1}}\Gamma(k+1) = \frac{1}{2^{k+1}}.
\end{aligned}$$

B.2 Prove $F_j \sim \exp(2\lambda)$

$$\begin{aligned}
\mathbb{P}(T_1 > t) &= \int_{t<x_1<x_2}^\infty \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}dx_1dx_2, \\
&= \int_{x_1>t}^\infty \lambda e^{-\lambda x_1}e^{-\lambda x_1}dx_1, \\
&= \int_{x_1>t}^\infty \lambda e^{-2\lambda}dx_1, \\
&= \frac{1}{2}e^{-2\lambda t}.
\end{aligned}$$

Given $\mathbb{P}(N>0) = 1 - \mathbb{P}(N=0) = \frac{1}{2}$, we have

$$\mathbb{P}(T_1 > t | N > 0) = \frac{\mathbb{P}(T_1 > t, N > 0)}{\mathbb{P}(N>0)} = e^{-2\lambda t}.$$

APPENDIX C

PROVE THE RESULTS FOR THE GENERALIZED TYPE 2 MODEL

Let $K = 0, 1, 2, \ldots$ be the number of arrival of N_1 in the interval $(0, E)$, where E is the first arrival of N_2,

$$E \sim \exp(\lambda).$$

Let $0 < T_1 < T_2 < \ldots < T_K < E$ be the arrival time of N_1. Let $Y_{n,j}$ be the state of $X_2(t)$ at $t = T_j, j = 1, \ldots, K$. Note that, at time 0,

$$X_1(0) = X_n,\ X_2(0) = 1.$$

Let $\beta = \sqrt{\beta_1\beta_2}$, we conclude by (3.96),

$$\begin{aligned} Y_{n,1} &= \frac{1}{2}e^{T_1(1+\beta)}(\sqrt{\frac{\beta_2}{\beta_1}}X_n + 1) + \frac{1}{2}e^{T_1(1-\beta)}(-\sqrt{\frac{\beta_1}{\beta_2}}X_n + 1), \\ &= X_n\sqrt{\frac{\beta_2}{\beta_1}}\left(\frac{1}{2}e^{T_1(1+\beta)} - \frac{1}{2}e^{T_1(1-\beta)}\right) + \frac{1}{2}e^{T_1(1+\beta)} + \frac{1}{2}e^{T_1(1-\beta)}. \end{aligned} \tag{C.1}$$

Similarly, by (3.96), for $j = 2, \ldots, K$,

$$\begin{aligned} Y_{n,j} = Y_{n,j-1}&\left(\frac{1}{2}e^{(T_j - T_{j-1})(1+\beta)} + \frac{1}{2}e^{(T_j - T_{j-1})(1-\beta)}\right) \\ &+ \left(\frac{1}{2}\sqrt{\frac{\beta_2}{\beta_1}}e^{(T_j - T_{j-1})(1+\beta)} - \frac{1}{2}e^{(T_j - T_{j-1})(1-\beta)}\right). \end{aligned} \tag{C.2}$$

We conclude that

$$
\begin{aligned}
Y_{n,K} = X_n \sqrt{\frac{\beta_2}{\beta_1}} \frac{e^{T_1(1+\beta)} - e^{T_1(1-\beta)}}{2} \prod_{j=2}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2} \\
+ \prod_{j=1}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2} \\
+ \sqrt{\frac{\beta_2}{\beta_1}} \sum_{i=2}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} - e^{(T_j - T_{j-1})(1-\beta)}}{2} \prod_{j=i+1}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2}.
\end{aligned}
\tag{C.3}
$$

Finally, we use (3.96) once again,

$$
\begin{aligned}
X_{n+1} &= \frac{1}{2} e^{(E-T_K)(1+\beta)} (1 + \sqrt{\frac{\beta_1}{\beta_2}} Y_{n,K}) + \frac{1}{2} e^{(E-T_K)(1-\beta)} (1 - \sqrt{\frac{\beta_1}{\beta_2}} Y_{n,K}), \\
&= Y_{n,K} \sqrt{\frac{\beta_1}{\beta_2}} \left(\frac{e^{(E-T_K)(1+\beta)} - e^{(E-T_K)(1-\beta)}}{2} \right) + \frac{e^{(E-T_K)(1+\beta)} + e^{(E-T_K)(1-\beta)}}{2}.
\end{aligned}
\tag{C.4}
$$

Combing (C.3) and (C.4), we obtain

$$
\begin{aligned}
A = & \frac{e^{T_1(1+\beta)} - e^{T_1(1-\beta)}}{2} \\
& \cdot \prod_{j=2}^{K} \frac{e^{(T_j - T_{j-1})(1+\beta)} + e^{(T_j - T_{j-1})(1-\beta)}}{2} \\
& \cdot \frac{e^{(E-T_K)(1+\beta)} - e^{(E-T_K)(1-\beta)}}{2}.
\end{aligned}
\tag{C.5}
$$

If we let $0 < T_1 < T_2 < \ldots < T_K < E$ be the arrival time of N_2 and let $Y_{n,j}$ be the state of $X_1(t)$ at $t = T_j, j = 1, \ldots, K$, we can compute A analogously.

APPENDIX D

PROVE THE JOINT DENSITY AND CCDF OF THE FIRST TYPE 3 MODEL

D.1 Prove the joint density

Let $\epsilon_1 = \epsilon_2 = 1$ in Equation (4.26), and let

$$
\begin{aligned}
I &= \int_0^\infty e^{-\lambda t} \frac{1}{t} e^{-\left(\frac{\left(\ln x_1 - (\beta_1 - \frac{\sigma_1^2}{2})t\right)^2}{2\sigma_1^2 t} + \frac{\left(\ln x_2 - (\beta_2 - \frac{\sigma_2^2}{2})t\right)^2}{2\sigma_2^2 t} \right)} dt, \\
&= \int_0^\infty e^{-\lambda t} \frac{1}{t} e^{-\frac{1}{2t}\left(\frac{(\ln x_1)^2}{\sigma_1^2} + \frac{(\ln x_2)^2}{\sigma_2^2}\right) - \left(\frac{(\beta_1 - \frac{\sigma_1^2}{2})^2}{2\sigma_1^2} + \frac{(\beta_2 - \frac{\sigma_2^2}{2})^2}{2\sigma_2^2}\right)t + \left(\frac{\ln x_1 (\beta_1 - \frac{\sigma_1^2}{2})}{\sigma_1^2} + \frac{\ln x_2 (\beta_2 - \frac{\sigma_2^2}{2})}{\sigma_2^2}\right)} dt, \\
&= x_1^{-\frac{1}{2} + \frac{\beta_1}{\sigma_1^2}} x_2^{-\frac{1}{2} + \frac{\beta_2}{\sigma_2^2}} \int_0^\infty e^{-\lambda t} \frac{1}{t} e^{-\frac{1}{2t}\left(\frac{(\ln x_1)^2}{\sigma_1^2} + \frac{(\ln x_2)^2}{\sigma_2^2}\right) - \left(\frac{(\beta_1 - \frac{\sigma_1^2}{2})^2}{2\sigma_1^2} + \frac{(\beta_2 - \frac{\sigma_2^2}{2})^2}{2\sigma_2^2}\right)t} dt. \qquad \text{(D.1)}
\end{aligned}
$$

Let

$$
J = \int_0^\infty e^{-\lambda t} \frac{1}{t} e^{-\frac{1}{2t}\left(\frac{(\ln x_1)^2}{\sigma_1^2} + \frac{(\ln x_2)^2}{\sigma_2^2}\right) - \left(\frac{(\beta_1 - \frac{\sigma_1^2}{2})^2}{2\sigma_1^2} + \frac{(\beta_2 - \frac{\sigma_2^2}{2})^2}{2\sigma_2^2}\right)t} dt, \qquad \text{(D.2)}
$$

with

$$
A = \left(\frac{(\beta_1 - \frac{\sigma_1^2}{2})^2}{2\sigma_1^2} + \frac{(\beta_2 - \frac{\sigma_2^2}{2})^2}{2\sigma_2^2} \right),
$$

and

$$
H(x_1, x_2) = \left(\frac{(\ln x_1)^2}{\sigma_1^2} + \frac{(\ln x_2)^2}{\sigma_2^2} \right),
$$

we have

$$\begin{aligned} J &= \int_0^\infty e^{-\lambda t} \frac{1}{t} e^{-\frac{H}{2t} - At} dt, \\ &= \int_0^\infty \frac{e^{-(\lambda + A)t}}{t} e^{-\frac{H}{2t}} dt. \end{aligned} \tag{D.3}$$

Let $t = HW$, we have

$$J = \int_0^\infty \frac{e^{-(\lambda + A)HW}}{W} e^{-\frac{1}{2W}} dW. \tag{D.4}$$

Consider very large x_1 and x_2, we integra W from 0 to a small number δ. Let $B = \lambda + A$,

$$\begin{aligned} J &\sim \int_0^\delta \frac{e^{-(\lambda + A)HW}}{W} e^{-\frac{1}{2W}} dW, \\ &= \int_0^\delta \frac{dW}{W} e^{-(BHW + \frac{1}{2W})}. \end{aligned} \tag{D.5}$$

We find $W^* = \frac{1}{(2BH)^{1/2}}$ that minimize $BHW + \frac{1}{2W}$, which gives

$$J \sim (2B)^{1/2} H^{1/2} \int_0^\delta e^{-(BHW + \frac{1}{2W})} dW. \tag{D.6}$$

Let

$$K = H^{1/2} \int_0^\delta e^{-(BHW + \frac{1}{2W})} dW. \tag{D.7}$$

Set $W = zH^{-1/2}$, we have

$$K = \int_0^{\delta H^{1/2}} e^{-H^{1/2}(Bz + \frac{1}{2z})} dz. \tag{D.8}$$

Let $f(z) = Bz + \frac{1}{2z}$, this function is minimized at the point when $z^* = \frac{1}{\sqrt{2B}}$. Do the Taylor expansion, we have

$$f(z) = f(z^*) + \frac{1}{2}(z - z^*)^2 f''(z^*) + O((z - z^*)^3).$$

Then let

$$\begin{aligned} K &= \int_{z^*-\epsilon}^{z^*+\epsilon} e^{-H^{1/2} f(z)} dz, \\ &\sim \int_{z^*-\epsilon}^{z^*+\epsilon} e^{-H^{1/2}\left(f(z^*)+\frac{1}{2}f''(z^*)(z-z^*)^2\right)} dz, \\ &= e^{-H^{1/2} f(z^*)} 2 \int_0^{\epsilon} e^{-H^{1/2}\frac{1}{2}f''(z^*)z^2} dz, \end{aligned} \tag{D.9}$$

and let $z' = \sqrt{H^{1/2} f''(z^*)}z$, we have

$$\begin{aligned} K &\sim e^{-H^{1/2} f(z^*)} \frac{2\sqrt{2\pi}}{\sqrt{f''(z^*)} H^{1/4}} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{z'^2}{2}} dz', \\ &= H^{-1/4} e^{-H^{1/2} f(z^*)} \frac{\sqrt{2\pi}}{\sqrt{f''(z^*)}}, \end{aligned} \tag{D.10}$$

where $f(z^*) = \sqrt{2B}$ and $f''(z^*) = (2B)^{\frac{3}{2}}$.

Back to the joint distribution in Equation (4.26), we have

$$f_{X_1,X_2}(x_1, x_2) \sim \frac{(2B)^{-1/4}}{\sqrt{2\pi}\sigma_1\sigma_2} x_1^{-\frac{3}{2}+\frac{\beta_1}{\sigma_1^2}} x_2^{-\frac{3}{2}+\frac{\beta_2}{\sigma_2^2}} H^{-1/4} e^{-\sqrt{2B}H^{1/2}}. \tag{D.11}$$

Thus, conclusion in Equation (4.27) is proved.

D.2 Prove the joint CCDF

Let $t_i = \ln x_i$, $x_i = e^{t_i}$. Then let

$$\begin{aligned} P &= \int_x^\infty \int_x^\infty x_1^{-1/2} x_2^{-1/2} \left((\ln x_1)^2 + (\ln x_2)^2\right)^{-1/4} e^{-\sqrt{2B}\left((\ln x_1)^2+(\ln x_2)^2\right)^{1/2}} dx_1 dx_2, \\ &= \int_h^\infty \int_h^\infty e^{\frac{t_1+t_2}{2}} (t_1^2+t_2^2)^{-1/4} e^{-\sqrt{2B}(t_1^2+t_2^2)^{1/2}} dt_1 dt_2, \end{aligned} \tag{D.12}$$

where $h = \ln x$.

Let $t_1 = r\cos\psi$ and $t_2 = r\sin\psi$, then

$$P = \int_0^{\pi/2} d\psi \int_0^\infty 1(r > \frac{h}{\cos\psi} \vee \frac{h}{\sin\psi}) r^{1/2} e^{\frac{r}{2}(\cos\psi+\sin\psi)} e^{-\sqrt{2B}r} dr. \tag{D.13}$$

Let $\delta = \sqrt{2B} - \frac{\cos\psi+\sin\psi}{2}$, we compute the following integral

$$\begin{aligned} \int_a^\infty r^{\frac{1}{2}} e^{-\delta r} dr &= (\delta)^{-2/3} \int_{a\delta}^\infty u^{\frac{1}{2}} e^{-u} du, \\ &\sim \frac{a^{\frac{1}{2}}}{\delta} e^{-a\delta}, a \to \infty. \end{aligned} \tag{D.14}$$

We have

$$\begin{aligned} P &\sim \int_0^{\frac{\pi}{2}} \frac{(\frac{h}{\cos\psi} \vee \frac{h}{\sin\psi})^{1/2}}{\delta} e^{-(\frac{h}{\cos\psi} \vee \frac{h}{\sin\psi})\delta} d\psi, \\ &= h^{\frac{1}{2}} \int_0^{\frac{\pi}{2}} \frac{(\frac{1}{\cos\psi} \vee \frac{1}{\sin\psi})^{1/2}}{\sqrt{2B} - \frac{\cos\psi+\sin\psi}{2}} e^{-h(\frac{1}{\cos\psi} \vee \frac{1}{\sin\psi})(\sqrt{2B} - \frac{\cos\psi+\sin\psi}{2})} d\psi. \end{aligned} \tag{D.15}$$

Let $g(\psi) = (\frac{1}{\cos\psi} \vee \frac{1}{\sin\psi})(\sqrt{2B} - \frac{\cos\psi+\sin\psi}{2})$. $g(\psi)$ achieves minimum at $\psi^* = \frac{\pi}{4}$, and

$$g(\psi^*) = \sqrt{2}(\sqrt{2B} - \frac{1}{\sqrt{2}}) = 2\sqrt{B} - 1.$$

Then

$$\begin{aligned}
P &\sim h^{\frac{1}{2}} \frac{2^{1/4}}{\sqrt{2B} - \frac{1}{\sqrt{2}}} \int_{\frac{\pi}{4}-\epsilon}^{\frac{\pi}{4}+\epsilon} e^{-hg(\psi)} d\psi, \\
&\sim h^{\frac{1}{2}} \frac{2^{1/4}}{\sqrt{2B} - \frac{1}{\sqrt{2}}} \int_{\frac{\pi}{4}-\epsilon}^{\frac{\pi}{4}+\epsilon} e^{-h\left(g(\psi^*) + \frac{1}{2}g''(\psi^*)(\psi-\psi^*)^2\right)} d\psi, \\
&= h^{\frac{1}{2}} \frac{2^{1/4}}{\sqrt{2B} - \frac{1}{\sqrt{2}}} e^{-hg(\psi^*)} \int_{\frac{\pi}{4}-\epsilon}^{\frac{\pi}{4}+\epsilon} e^{-\frac{h}{2}g''(\psi^*)(\psi-\psi^*)^2} d\psi, \\
&= h^{\frac{1}{2}} \frac{2^{1/4}}{\sqrt{2B} - \frac{1}{\sqrt{2}}} e^{-hg(\psi^*)} \int_{-\epsilon}^{\epsilon} e^{-\frac{h}{2}g''(\psi^*)\psi^2} d\psi, \\
&= \frac{2^{1/4}}{\sqrt{2B} - \frac{1}{\sqrt{2}}} e^{-hg(\psi^*)} \int_{-\epsilon(hg''(\psi^*))^{1/2}}^{\epsilon(hg''(\psi^*))^{1/2}} e^{-\frac{\psi'^2}{2}} d\psi', \\
&\sim \frac{2^{3/4}\sqrt{\pi}}{(\sqrt{2B} - \frac{1}{\sqrt{2}})(g''(\psi^*))^{1/2}} e^{-\ln x(2\sqrt{B}-1)}, \\
&\sim \frac{2^{3/4}\sqrt{\pi}}{(\sqrt{2B} - \frac{1}{\sqrt{2}})(g''(\psi^*))^{1/2}} x^{-(\sqrt{4\lambda+1}-1)}.
\end{aligned} \tag{D.16}$$

So we have

$$P(X_1 > x, X_2 > x) \sim \frac{(B)^{-1/4}}{\sigma_1\sigma_2} \frac{1}{(\sqrt{2B} - \frac{1}{\sqrt{2}})(g''(\psi^*))^{1/2}} x^{-(\sqrt{4\lambda+1}-1)}. \tag{D.17}$$

Thus we prove Equation (4.29).

APPENDIX E

THEORETICAL RESULTS FOR THE SECOND TYPE 3 MODEL

E.1 Marginal tail

We need to analyze the asymptotic behaviour of the density of the second Type 3 model

$$f_X(x) = \lambda \frac{2\beta}{\sigma} x^{-1/2} \int_0^\infty e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t} - 1} \exp\left\{ -\frac{2\beta}{\sigma} \frac{e^{\beta t} + x}{e^{\beta t} - 1} \right\} I_1\left(\frac{4\beta}{\sigma} \frac{e^{\beta t/2}}{e^{\beta t} - 1} x^{1/2} \right) dt \tag{E.1}$$

as $x \to \infty$. We will prove that

$$\lim_{x\to\infty} x^{1+\lambda/\beta} f_X(x) = \lambda \frac{2}{\sigma} e^{-2\beta/\sigma} \int_0^\infty y^{-3/2-\lambda/\beta} \exp\left\{ -\frac{2\beta}{\sigma} y^{-1} \right\} I_1\left(\frac{4\beta}{\sigma} y^{-1/2} \right) dy \in (0, \infty) . \tag{E.2}$$

We will use several facts about the modified Bessel function. First of all, I_1 is bounded on compact intervals. Furthermore,

$$I_1(x) \sim \sqrt{\frac{1}{2\pi}} x^{-1/2} e^x \quad \text{as } x \to \infty. \tag{E.3}$$

Let $M > 0$ be a large number. Denote

$$g_1(x) = \int_0^{\beta^{-1}\log(x/M)} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t} - 1} \exp\left\{ -\frac{2\beta}{\sigma} \frac{e^{\beta t} + x}{e^{\beta t} - 1} \right\} I_1\left(\frac{4\beta}{\sigma} \frac{e^{\beta t/2}}{e^{\beta t} - 1} x^{1/2} \right) dt ,$$

$$g_2(x) = \int_{\beta^{-1}\log(x/M)}^{\beta^{-1}\log(xM)} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t} - 1} \exp\left\{ -\frac{2\beta}{\sigma} \frac{e^{\beta t} + x}{e^{\beta t} - 1} \right\} I_1\left(\frac{4\beta}{\sigma} \frac{e^{\beta t/2}}{e^{\beta t} - 1} x^{1/2} \right) dt ,$$

$$g_3(x) = \int_{\beta^{-1}\log(xM)}^{\infty} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t}-1} \exp\left\{-\frac{2\beta}{\sigma}\frac{e^{\beta t}+x}{e^{\beta t}-1}\right\} I_1\left(\frac{4\beta}{\sigma}\frac{e^{\beta t/2}}{e^{\beta t}-1}x^{1/2}\right)\, dt\,.$$

The claim (E.2) will follow from the following three statements.

$$\lim_{M\to\infty} \limsup_{x\to\infty} x^{\lambda/\beta+1/2} g_1(x) = 0\,, \tag{E.4}$$

$$\lim_{M\to\infty} \limsup_{x\to\infty} x^{\lambda/\beta+1/2} g_3(x) = 0\,, \tag{E.5}$$

$$\begin{aligned}\lim_{M\to\infty} \liminf_{x\to\infty} x^{\lambda/\beta+1/2} g_2(x) &= \lim_{M\to\infty} \limsup_{x\to\infty} x^{\lambda/\beta+1/2} g_2(x) \\ =& e^{-2\beta/\sigma}\frac{1}{\beta}\int_0^{\infty} y^{-3/2-\lambda/\beta} \exp\left\{-\frac{2\beta}{\sigma}y^{-1}\right\} I_1\left(\frac{4\beta}{\sigma}y^{-1/2}\right)\, dy\,.\end{aligned} \tag{E.6}$$

We start with proving (E.4). Note that for every $\theta > 0$ there are θ-dependent finite positive constants $c_1, c_2, \ldots$ such that

$$\begin{aligned}&\int_0^{\theta} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t}-1} \exp\left\{-\frac{2\beta}{\sigma}\frac{e^{\beta t}+x}{e^{\beta t}-1}\right\} I_1\left(\frac{4\beta}{\sigma}\frac{e^{\beta t/2}}{e^{\beta t}-1}x^{1/2}\right)\, dt, \\ &\le c_1 \int_0^{\theta} t^{-1} e^{-c_2 t^{-1}} e^{-c_3 t^{-1} x} I_1\big(c_4 t^{-1} x^{1/2}\big)\, dt, \\ &\le c_5 \int_0^{\theta} t^{-1} e^{-c_2 t^{-1}} e^{-c_3 t^{-1} x} e^{c_6 t^{-1} x^{1/2}}\, dt, \\ &\le c_7 e^{-c_8 x} \int_0^{\theta} t^{-1} e^{-c_2 t^{-1}}\, dt\,,\end{aligned}$$

an exponentially fast decaying function of x. Next, with the same notation for constants,

$$\begin{aligned}&\int_{\theta}^{\beta^{-1}\log(x/M)} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t}-1} \exp\left\{-\frac{2\beta}{\sigma}\frac{e^{\beta t}+x}{e^{\beta t}-1}\right\} I_1\left(\frac{4\beta}{\sigma}\frac{e^{\beta t/2}}{e^{\beta t}-1}x^{1/2}\right)\, dt, \\ &\le c_1 \int_{\theta}^{\beta^{-1}\log(x/M)} e^{-\lambda t} e^{-\beta t/2} \exp\big\{-c_2 x e^{-\beta t}\big\} I_1\big(c_3 x^{1/2} e^{-\beta t/2}\big)\, dt, \\ &= c_3 \int_{e^{\beta\theta}}^{x/M} w^{-\lambda/\beta-3/2} e^{-c_2 x/w} I_1\big(c_3 x^{1/2} w^{-1/2}\big)\, dw, \\ &\le c_3 x^{-\lambda/\beta-1/2} \int_0^{1/M} y^{-\lambda/\beta-3/2} e^{-c_2 y^{-1}} I_1\big(c_3 y^{-1/2}\big)\, dy\,,\end{aligned}$$

and so (E.4) follows.

We proceed with proving (E.5). The argument is similar to the above. We have for $M \geq 1$ and large x,

$$\begin{aligned}
&\int_{\beta^{-1}\log(xM)}^{\infty} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t}-1} \exp\left\{-\frac{2\beta}{\sigma}\frac{e^{\beta t}+x}{e^{\beta t}-1}\right\} I_1\left(\frac{4\beta}{\sigma}\frac{e^{\beta t/2}}{e^{\beta t}-1}x^{1/2}\right)\, dt, \\
&\leq c_3 x^{-\lambda/\beta-1/2} \int_M^{\infty} y^{-\lambda/\beta-3/2} e^{-c_2 y^{-1}} I_1\big(c_3 y^{-1/2}\big)\, dy\,,
\end{aligned}$$

and so (E.5) follows.

Finally, we prove (E.6). Let $\varepsilon > 0$. For a fixed $M \geq 1$, and for all large enough x,

$$\begin{aligned}
&e^{-2(1+\varepsilon)\beta/\sigma} \int_{\beta^{-1}\log(x/M)}^{\beta^{-1}\log(xM)} e^{-\lambda t} e^{-\beta t/2} \exp\left\{-\frac{2(1+\varepsilon)\beta}{\sigma} e^{-\beta t} x\right\} I_1\left(\frac{4\beta}{\sigma} e^{-\beta t/2} x^{1/2}\right)\, dt, \\
&\leq \int_{\beta^{-1}\log(x/M)}^{\beta^{-1}\log(xM)} e^{-\lambda t} \frac{e^{\beta t/2}}{e^{\beta t}-1} \exp\left\{-\frac{2\beta}{\sigma}\frac{e^{\beta t}+x}{e^{\beta t}-1}\right\} I_1\left(\frac{4\beta}{\sigma}\frac{e^{\beta t/2}}{e^{\beta t}-1}x^{1/2}\right)\, dt, \\
&\leq (1+\varepsilon) e^{-2\beta/\sigma} \int_{\beta^{-1}\log(x/M)}^{\beta^{-1}\log(xM)} e^{-\lambda t} e^{-\beta t/2} \exp\left\{-\frac{2\beta}{\sigma} e^{-\beta t} x\right\} I_1\left(\frac{4(1+\varepsilon)\beta}{\sigma} e^{-\beta t/2} x^{1/2}\right)\, dt.
\end{aligned}$$

Since

$$\begin{aligned}
&\int_{\beta^{-1}\log(x/M)}^{\beta^{-1}\log(xM)} e^{-\lambda t} e^{-\beta t/2} \exp\left\{-\frac{2(1+\varepsilon)\beta}{\sigma} e^{-\beta t} x\right\} I_1\left(\frac{4\beta}{\sigma} e^{-\beta t/2} x^{1/2}\right)\, dt, \\
&= x^{-\lambda/\beta-1/2} \frac{1}{\beta} \int_{1/M}^{M} y^{-3/2-\lambda/\beta} \exp\left\{-\frac{2(1+\varepsilon)\beta}{\sigma} y^{-1}\right\} I_1\left(\frac{4\beta}{\sigma} y^{-1/2}\right)\, dy\,,
\end{aligned}$$

we conclude that for any $\varepsilon > 0$,

$$\liminf_{x\to\infty} x^{\lambda/\beta+1/2} g_2(x) \geq e^{-2(1+\varepsilon)\beta/\sigma} \frac{1}{\beta} \int_{1/M}^{M} y^{-3/2-\lambda/\beta} \exp\left\{-\frac{2(1+\varepsilon)\beta}{\sigma} y^{-1}\right\} I_1\left(\frac{4\beta}{\sigma} y^{-1/2}\right)\, dy\,.$$

Letting $\varepsilon \to 0$ we obtain

$$\liminf_{x\to\infty} x^{\lambda/\beta+1/2} g_2(x) \geq e^{-2\beta/\sigma} \frac{1}{\beta} \int_{1/M}^{M} y^{-3/2-\lambda/\beta} \exp\left\{-\frac{2\beta}{\sigma} y^{-1}\right\} I_1\left(\frac{4\beta}{\sigma} y^{-1/2}\right)\, dy\,.$$

Hence

$$\lim_{M\to\infty}\liminf_{x\to\infty} x^{\lambda/\beta+1/2}g_2(x) \geq e^{-2\beta/\sigma}\frac{1}{\beta}\int_0^\infty y^{-3/2-\lambda/\beta}\exp\left\{-\frac{2\beta}{\sigma}y^{-1}\right\}I_1\left(\frac{4\beta}{\sigma}y^{-1/2}\right)\,dy\,.$$

This proves one part of (E.6), and the second part is completely analogous.

E.2 Joint tail

Instead of the density, let us investigate the joint tail

$$\int_{y_1}^\infty\int_{y_2}^\infty f_{X_1,X_2}(x_1,x_2)dx_1dx_2$$

as $y_1, y_2 \to \infty$. As in the single variate case, the main part is to investigate the function

$$P(X_1 > y_1, X_2 > y_2) = \int_{y_1}^\infty\int_{y_2}^\infty x_1^{-1/2}x_2^{-1/2}dx_1dx_2$$
$$\int_0^\infty e^{-(\lambda+\beta)t}\exp\{-\frac{2\beta}{\sigma}e^{-\beta t}(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}e^{-\beta t/2}x_1^{1/2})I_1(\frac{4\beta}{\sigma}e^{-\beta t/2}x_2^{1/2})dt$$

as $y_1, y_2 \to \infty$. We have

$$P(X_1 > y_1, X_2 > y_2) = \frac{1}{\beta}\int_{y_1}^\infty\int_{y_2}^\infty x_1^{-1/2}x_2^{-1/2}dx_1dx_2$$

$$\int_0^1 z^{\lambda/\beta}\exp\{-\frac{2\beta}{\sigma}z(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}z^{1/2}x_1^{1/2})I_1(\frac{4\beta}{\sigma}z^{1/2}x_2^{1/2})dz,$$
$$\sim \frac{1}{\beta}\int_{y_1}^\infty\int_{y_2}^\infty x_1^{-1/2}x_2^{-1/2}dx_1dx_2$$
$$\int_0^\infty z^{\lambda/\beta}\exp\{-\frac{2\beta}{\sigma}z(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}z^{1/2}x_1^{1/2})I_1(\frac{4\beta}{\sigma}z^{1/2}x_2^{1/2})dz,$$
$$= \frac{1}{\beta}\int_0^\infty z^{\lambda/\beta}dz$$
$$\int_{y_1}^\infty\int_{y_2}^\infty x_1^{-1/2}x_2^{-1/2}\exp\{-\frac{2\beta}{\sigma}z(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}z^{1/2}x_1^{1/2})I_1(\frac{4\beta}{\sigma}z^{1/2}x_2^{1/2})dx_1dx_2,$$

$$= \frac{1}{\beta}\int_0^\infty z^{\lambda/\beta-1}dz$$

$$\int_{zy_1}^\infty\int_{zy_2}^\infty x_1^{-1/2}x_2^{-1/2}\exp\{-\frac{2\beta}{\sigma}(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}x_1^{1/2})I_1(\frac{4\beta}{\sigma}x_2^{1/2})dx_1dx_2,$$

$$= \frac{1}{\beta}\int_0^\infty\int_0^\infty x_1^{-1/2}x_2^{-1/2}\exp\{-\frac{2\beta}{\sigma}(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}x_1^{1/2})I_1(\frac{4\beta}{\sigma}x_2^{1/2})dx_1dx_2$$

$$\int_0^\infty z^{\lambda/\beta-1}\mathbf{1}(z<\min(x_1/y_1,x_2/y_2))dz,$$

$$= \frac{1}{\lambda}\int_0^\infty\int_0^\infty x_1^{-1/2}x_2^{-1/2}(\min(x_1/y_1,x_2/y_2))^{\lambda/\beta}\exp\{-\frac{2\beta}{\sigma}(x_1+x_2)\}$$

$$I_1(\frac{4\beta}{\sigma}x_1^{1/2})I_1(\frac{4\beta}{\sigma}x_2^{1/2})dx_1dx_2.$$

Set $y_1 = y_2 = x$, we have

$$\lim_{x\to\infty} P(X_1>x, X_2>x)x^{\lambda/\beta}$$

$$\sim \int_0^\infty\int_0^\infty x_1^{-1/2}x_2^{-1/2}(\min(x_1,x_2))^{\lambda/\beta}\exp\{-\frac{2\beta}{\sigma}(x_1+x_2)\}I_1(\frac{4\beta}{\sigma}x_1^{1/2})I_1(\frac{4\beta}{\sigma}x_2^{1/2})dx_1dx_2 \in (0,\infty).$$

BIBLIOGRAPHY

[1] Aharony, A., and Stauffer, D. *Introduction to percolation theory.* Taylor and Francis, 2003.

[2] Aiello, W., Chuang, F., and Lu, L. Random evolution in massive graph. *Insurance: Mathematics and Economics 4* (2002), 97–122.

[3] Albert, R., and Barabsi, A.L. Statistical mechanics of complex networks. *Reviews of modern physics 74(1)* (2002), 47.

[4] Albert, R., Jeong, H., and Barabasi, A.L. Diameter of the world-wide web. *Nature 401(6749)* (1999), 130–131.

[5] Alvarez, L., Gousseau, Y., and Morel, J.M. The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics 111* (1999), 167–242.

[6] Arnold, B.C. *Pareto Distributions.* International Co-operative Publishing House, 1983. Chapter 6.

[7] Asimit, A.V., Furman, E., and Vernic, R. On a multivariate pareto distribution. *Insurance: Mathematics and Economics 46(2)* (2010), 308–316.

[8] Barabási, A.L, and Albert, R. Emergence of scaling in random networks. *Science 286* (1999), 509–512.

[9] Bollobás, B., Borgs, C., Chayes, J., and Riordan, O. Directed scale-free graphs. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms* (Jan. 2003), Society for Industrial and Applied Mathematics, pp. 132–139.

[10] Breiman, L. On some limit theorems similar to the arc-sin law. *Theory of probability and its applications X(2)* (1965), 323–331.

[11] Broadbent, S.R., and Hammersley, J.M. Percolation processes. *Mathematical Proceedings of the Cambridge Philosophical Society 53(3)* (1957), 629–641.

[12] Brockett, R.W. Talk at kyoto university. Tech. rep., Kyoto University, March 2007.

[13] Butler, S. Interlacing for weighted graphs using the normalized laplacian. *Electronic Journal of Linear Algebra 16* (2007), 90–98.

[14] Caldarelli, G., Marchetti, R., and Pietronero, L. The fractal properties of internet. *EUROPHYSICS LETTERS 52(4)* (2000), 386–391.

[15] Cha, S.H. Comprehensive survey on distance/similarity measures between probability density functions. *Internatioanl Journal of Mathematical models and methods in applied sciences 1(4)* (2007), 300–307.

[16] Chen, R. R., and Lee, C. F. A constant elasticity of variance (cev) family of stock price distributions in option pricing, review, and integration. In *Handbook of Quantitative Finance and Risk Management*. Springer US, 2010, pp. 1615–1625.

[17] Chung, F. *Spectral Graph Theory*. American Mathematical Society, 1997. Chapter 1.

[18] Chung, F., Lu L., and Vu, V. Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences 100(11)* (2003).

[19] Clauset, A., Shalizi, C.R., and M.E.J., Newman. Power-law distributions in empirical data. *SIAM REVIEW 51(4)* (2009), 661–703.

[20] Cox, J. Notes on option pricing i: Constant elasticity of diffusions. Stanford University, 1975.

[21] Cox, J.C., Ingersoll, J.E., and Ross, S.A. A theory of the term structure of interest rates. *Econometrica 53* (1985), 385407.

[22] de Solla Price, D.J. Networks of scientific papers. *Science 149(3683)* (1965), 510–515.

[23] Denisov, D., and Zwart, B. On a theorem of breiman and a class of random difference equations. *J. Appl. Probab. 44(4)* (2007), 1031–1046.

[24] Ebel, H., Mielsch, L.I., and Bornholdt, S. Scale-free topology of e-mail networks. *Physical review E 66(3)* (2002), 035103.

[25] Ellison, N.B. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication 13(1)* (2007), 210–230.

[26] Embrechts, P., Lindskog, F., and McNeil, A. Modelling dependence with copulas. *Rapport technique, Dpartement de mathmatiques, Institut Fdral de Technologie de Zurich* (Sep 2001).

[27] Erdös, P., and Rényi, A. On random graphs i. *Publications Mathematicae 6* (1959), 290–297.

[28] Faloutsos, M., Faloutsos, P., and Faloutsos, C. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review 29(4)* (1999), 251–262.

[29] Fang, Z. *Finding statistics characteristics and similarities of substructures between knowledge networks*. PhD thesis, University of Massachusetts Lowell, 2012.

[30] Fang, Z., Wang, J., Liu, B., and Gong, W. Double pareto lognormal distributions in complex networks. In *Handbook of Optimization in Complex Networks*. Springer US, 2012, pp. 55–80.

[31] Farmer, J.D., and Geanakoplos, J. Power laws in economics and elsewhere. *Santa Fe Institute* (May, 2008).

[32] Gabaix, X. Power laws in economics and finance. *National Bureau of Economic Research* (2008), w14299.

[33] Goldie, C.M. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability* (1991), 126–166.

[34] Gong, W. Stochastic dynamic systems. UMASS ECE735 HANDBOOK, 2007.

[35] Gong, W. Can one hear the shape of a concept? In *Proceedings of the 31st Chinese Control Conference (Plenary Lecture)* (Hefei, China, July 2012), pp. 22–26.

[36] Gong, W. Transient response functions for graph structure addressable memory. In *Proceedings of the 52th IEEE Conference on Decision and Control* (Florence, Italy, December 2013).

[37] Gong, W., Liu, Y., and Misra, V. On the tails of web file size distributions. *Proceedings of the annual allerton conference on communication control and computing 39(1)* (2001), 192–201.

[38] Gong, W., Liu, Y., Misra, V., and Towsley, D. Self-similarity and long range dependence on the internet: a second look at the evidence, origins and implications. *Computer Networks 48* (2005), 377–399.

[39] Gousseau, Y., and Morel, J.M. Are natural images of bounded variation? *SIAM Journal on Mathematical Analysis 33(3)* (2001), 634–648.

[40] Grisi-Filho, J.H., Ossada, R., Ferreira, F., and Amaku, M. Scale-free networks with the same degree distribution: Different structural properties. *Physics Research International 2013* (2013), 234180.

[41] Jeh, G., and Widom, J. Simrank: a measure of structural-context similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (2002), pp. 538–543.

[42] Jiang, B. Multivariate power law. Tech. rep., University of Massachusetts, Amherst, August 21, 2012.

[43] Jiang, B., Brockett, R., Gong, W., and Towsley, D. Stochastic differential equations for power law behaviors. In *51st IEEE Conference on Decision and Control* (December 10-13, 2012).

[44] Kagan, Y. Y. Earthquake size distribution: Power-law with exponent beta = 1/2? *Tectonophysics 490(1)* (2010), 103–114.

[45] Kalisky, T., Cohen, R., , Ben-Avraham, D., and Havlin, S. Tomography and stability of complex networks. In *Complex Networks*. Springer Berlin Heidelberg, 2004, pp. 3–34.

[46] Kang, J., Lu, S., Gong, W., and Kelly, P.A. A complex network based feature extraction for image retrieval. In *IEEE International conference on image processing (ICIP)* (2014), pp. 2051–2055.

[47] Kang, J., Lu, S., Gong, W., and Kelly, P.A. Image feature extraction based on spectral graph information. In *IEEE International Conference on image processing (ICIP)* (Sep 25-28, 2016), pp. 46–50.

[48] Kesten, H. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica 13(1)* (1973), 207–248.

[49] Koutra, D., Parikh, A., Ramdas, A., and Xiang, J. Algorithms for graph similarity and subgraph matching. Available at `https://www.cs.cmu.edu/~jingx/docs/DBreport.pdf`.

[50] Krapivsky, P.L., and Redner, S. A statistical physics perspective on web growth. *Computer Networks 39(3))* (2002), 261–276.

[51] Krapivsky, P.L., Rodgers, G.J., and Redner, S. Degree distributions of growing networks. *Physical Review Letters 86(23)* (2001), 5401.

[52] Kumar, R., Novak, J., and Tomkins, A. Structure and evolution of online social networks. In *Link mining: models, algorithms, and applications*. Springer New York, 2010, pp. 337–357.

[53] Kunegis, J. Konect: the koblenz network collection. In *WWW 2013 Companion* (Rio de Janeiro, Brazil, May 13-17, 2013).

[54] Kwak, H., Lee, C., Park, H., and Moon, S. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (April 26, 2010), pp. 591–600.

[55] Lu, S., Kang, J., Gong, W., and Towsley, D. Complex network comparison using random walks. In *Proceedings of the 23rd International Conference on World Wide Web* (2014), pp. 727–730.

[56] Lu, S., Samorodnitsky, G., Gong, W., Jiang, B., Kang, J., and Towsley, D. Poisson process driven stochastic differential equations for bivariate heavy tailed distributions. In *American Control Conference (ACC)* (July 6, 2016), pp. 2977–2982.

[57] Mcdonald, P., and Meyers, R. Diffusions on graphs, poisson problems and spectral geometry. *TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY 354(12))* (2002), 5111–5136.

[58] Mcdonald, P., and Meyers, R. Isospectral polygons, planar graphs and heat content. *Proceedings of the American Mathematical Society 131(11)* (2003), 3589–3599.

[59] Melnik, S., Garcia-Molina, H., and Rahm, E. Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings. of 18th International Conference on Data Engineering* (2002), pp. 117–128.

[60] Millane, R.P., Alzaidi, S., and Hsiao, W.H. Scaling and power spectra of natural images. *Proc. Image and Vision Computing New Zealand* (Nov. 2003), 148–153.

[61] Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., and Bhattacharjee, B. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (Oct. 2007, 2012), pp. 29–42.

[62] Misra, V., and Gong, W.B. A hierarchical model for teletraffic. *Proceedings of the 37th IEEE Conference on Decision and Control 2* (1998), 1674–1679.

[63] Mitzenmacher, M. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics 1(2)* (2004), 226–251.

[64] Mitzenmacher, M. Dynamic models for file sizes and double pareto distributions. *Internet Mathematics 1(3)* (2004), 305–333.

[65] Mohaisen, A., Yun, A., and Kim, Y. Measuring the mixing time of social graphs. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (Nov 1, 2010), pp. 383–389.

[66] Molloy, M., and Reed, B. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing 7(03))* (1998), 295–305.

[67] Morrison, K. The growth of social media: From passing trend to international obsession. Available at `http://www.adweek.com/socialtimes/the-growth-of-social-media-from-trend-to-obsession-infographic/142323`.

[68] Mumford, D. Pattern theory: the mathematics of perception. Available at `arXivpreprintmath/0212400`.

[69] Newman, M.E. The structure and function of complex networks. *ACM SIGCOMM computer communication review 45(2)* (2003), 167–256.

[70] Newman, M.E. Power laws, pareto distributions and zipf's law. *Contemporary physics 46(5)* (2005), 323–351.

[71] Oksendal, B. *Stochastic differential equations: an introduction with applications.* Springer Science and Business Media, 2013. Chapter 5.

[72] Pacati, C. Brownian motion and geometric brownian motion. Graphical representaions, 2010-2011.

[73] Papadimitriou, P., Dasdan, A., and Carcia-Molina, H. Web graph similarity for anomaly detection. *Journal of International Service and Applications 1(1)* (2010), 19–30.

[74] Pareto, V., and Page, A. N. Translation of manuale di economia politica. *Manual of political economy* (1971).

[75] Park, J., and Kim, K. The heat energy content of a riemannian manifold. *Trends in Mathematics, Information Center for Mathematical Sciences 5(2))* (2002), 125–129.

[76] Petermann, T., Thiagarajan, T.C., Lebedev, M.A., Nicolelis, M.A., Chialvo, D.R., and Plenz, D. Spontaneous cortical activity in awake monkeys composed of neuronal avalanches. *Proceedings of the National Academy of Sciences 106(37)* (2009), 15921–15926.

[77] Redner, S. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems 4(2)* (1998).

[78] Reed, W.J. The pareto, zipf and other power laws. *Economics Letters 74* (2001), 15–19.

[79] Reed, W.J. A parametric model for income and other size distributions and some extensions. *International Journal of Statistics 64(1)* (2006).

[80] Reed, W.J., and Hughes, B.D. From gene families and genera to incomes and internet file sizes: why power laws are so common in nature. *Physical Review E 66(6)* (2002).

[81] Reed, W.J., and Hughes, B.D. Power-law distributions from exponential processes: An explanation for the occurrence of long-tailed distributions in biology and elsewhere. *Scientiae Mathematicae Japonicae Online 8* (2003), 329–339.

[82] Reed, W.J., and Jorgensen, M. The double pareto-lognormal distributiona new parametric model for size distributions. *Communications in Statistics-Theory and Methods 33(8)* (2004), 1733–1753.

[83] Resnick, S. *Adventures in Stochastic Process.* Birkhauser Basel, 2002. p254.

[84] Ribeiro, B., Gauvin, W., Liu, B., and Towsley, D. On myspace account spans and double pareto-like distribution of friends. In *INFOCOM IEEE Conference on Computer Communications Workshops* (March 15-19, 2010).

[85] Ross, S.M. Variations on brownian motion. In *Introduction to Probability Models (11th ed.)*. Amsterdam: Elsevier, 2014, pp. 612–614.

[86] Ruderman, D.L. Origins of scaling in natural images. *Vision research 37(23)* (1997), 3385–3398.

[87] Samorodnitsky, G., Resnick, S., Towsley, D., Davis, R., Willis, A., and Wan, P. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *Journal of Applied Probability 53(1)* (2016), 146–161.

[88] Siegmund, D. Note on a stochastic recursion. *Lecture Notes-Monograph Series* (2001), 547–554.

[89] Simoncelli, E.P., and Olshausen, B.A. Natural image statistics and neural representation. *Annual review of neuroscience 24(1)* (2001), 1193–1216.

[90] Torralba, A., and Oliva, A. Statistics of natural image categories. *Network: computation in neural systems 14(3)* (2014), 391–412.

[91] Van Den Berg, M., and Gilkey, P.B. Heat content asymptotics of a riemannian manifold with boundary. *Journal of Function Analysis 120* (1994), 48–71.

[92] Wolff, R.W. Poisson arrivals see time averages. *Operations Research 30(2)* (1982), 223–231.

[93] Zager, L. *Graph similarity and matching*. PhD thesis, Massachusetts Institute of Technology, 2005.