

CORNELL UNIVERSITY MATHEMATICS DEPARTMENT SENIOR THESIS

*An Exploration of Transmission on Random
Graphs*

A THESIS PRESENTED IN PARTIAL FULFILLMENT
OF CRITERIA FOR HONORS IN MATHEMATICS

Divya Kirti

May 2010

BACHELOR OF ARTS, CORNELL UNIVERSITY

THESIS ADVISOR(S)

Richard Durrett
Department of Mathematics

This document is in the public domain.

ABSTRACT

Suppose that a new technology emerges in an industry, and the object of study is the adoption of this new technology. We view the comprising firms as nodes of a graph, and proceed by assuming a known random graph structure. To provide structure to analysis of the problem, we choose two different types of random graphs - the Erdős-Rényi, and Newman-Strogatz-Watts graphs, and assume that every vertex transmits to each of its neighbors at times that are independent and identically distributed as exponential waiting times. Related problems have been considered by Economists for many years – Blume [1], looked at interactions across lattices as a way of modeling game-theoretic interactions between entire populations where individual players only interact with a finite set of neighbors.

The structure of the model implies that expected percentage adoption depends only on the expected size of the *giant component* of the graph. Thus, in the Erdős-Rényi case, we model the process of transmission as an exploration of its connected component, which behaves like a branching process. Using the probability that this branching process becomes extinct we see that the two parameters of the model are identifiable sequentially.

We also use simulation times and extinction probabilities of branching processes to estimate the size of connected components given they are not the giant component as a function of the probability of connected in the Erdős-Rényi case. Finally, we consider the average distance between any two vertices of an Erdős-Rényi random graph as a way to further understand the transmission process and results.

BIOGRAPHICAL SKETCH

Divya Kirti was educated in New Delhi, India and Singapore, completing the International Baccalaureate at the United World College of South East Asia, Singapore. He graduated with the Class of 2010 from Cornell University, with majors in Economics and Mathematics.

Mathematics is in everything.

ACKNOWLEDGEMENTS

I would like to thank the faculty in the Mathematics department at Cornell University that have encouraged me. My thesis advisor Professor Durrett provided me excellent supervision, and Professors Ramakrishna, Muscalu, Hubbard and Brown all provided excellent advice over the years. I would also like to thank Professor Blume in the Economics department for suggesting the underlying problem and his subsequent comments.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Figures	vii
1 Introduction	1
2 Transmission along random graphs	6
2.1 Erdős-Rényi Graph	6
2.1.1 Algorithms	8
2.1.2 Simulation Results	10
2.2 Newman-Strogatz-Watts Graph	14
2.2.1 Algorithms	16
2.2.2 Simulation Results	17
3 Erdős-Rényi Adoption and the Branching Process	22
3.1 Diameter of Connected Component Given Extinction	23
3.2 Estimation	25
3.3 Expected Adoption	27
4 Distance Distributions	29
A Runtime Figures	31
B Miscellaneous Figures	33
References	35

LIST OF FIGURES

2.1	Erdős-Rényi Adoption times	10
2.2	Erdős-Rényi Adoption times with λ varying within $[1, 2]$	11
2.3	Erdős-Rényi final times with varying λ and superimposed scaled inverse logistic function	12
2.4	Erdős-Rényi final percentage adoption with $n = 300$	13
2.5	Erdős-Rényi Adoption as a function of λ and n	14
2.6	Newman-Strogatz-Watts Adoption times with $\alpha = 2.3$ and $\mu = 1$, 10 trials	18
2.7	Newman-Strogatz-Watts Adoption times with α varying within $[2, 3)$	19
2.8	Newman-Strogatz-Watts Adoption with $n = 500$	20
2.9	Newman-Strogatz-Watts Adoption	21
3.1	Erdős-Rényi Simulation Time	22
3.2	Erdős-Rényi Generating Function and Extinction probability, $\lambda =$ $1.5, n = 500$	25
3.3	Erdős-Rényi connected component given extinction, $n = 500$. . .	26
3.4	Erdős-Rényi simulated adoption (blue) and expected adoption (green)	27
3.5	λ corresponding to percentage adoption for $n = 250$ given Erdős- Rényi structure	28
4.1	Erdős-Rényi distance distribution, $n = 1000$	29
A.1	Erdős-Rényi Runtime	31
A.2	Newman-Strogatz-Watts Runtime	32
B.1	Newman-Strogatz-Watts Simulation Time	33
B.2	Power law distance distribution, $n = 1000$	34

CHAPTER 1

INTRODUCTION

The primary analysis is motivated by the following estimation problem: suppose that a new technology emerges in an industry, and the object of study is the process of adoption of this new technology. Related problems have been considered by Economists for many years – Blume [1], looked at interactions across lattices as a way of modeling game-theoretic interactions between entire populations where individual players only interact with a finite set of neighbors. The concept was subsequently generalized to more general types of graphs by Morris [2]. Young and Burke [3] focused on customs and accepted practices that lead to areas of local homogeneity, even though there may be global diversity. More recently, Young [4] has also considered diversity amongst agents as a way of explaining delay in the diffusion of objects of interest.

One way to model this industry is to view the comprising firms as nodes of a graph, where transmission is possible between nodes connected by edges. The idea, then, is to observe the pattern with which adoption occurs over time, and attempt to estimate parameters of the model¹.

It is clear that parameters relating to the following two properties are of interest:

1. The level of connectivity between firms.

¹Though this is a specific application we outline to motivate the analysis, precisely the same set-up can correspond to any form of transmission in a setting where agents can be connected to each other in a relevant manner. Epidemics are also commonly modeled as transmission across random graphs: in the model detailed below, a reinterpretation of transmission of technology as transmission of disease models the case where subjects are permanently infected. Transmission of information across social networks (e.g. the spread of rumors) can also be considered in a similar manner.

2. The speed of transmission given two firms are connected.

To provide structure to analysis of the problem, we choose two different types of random graphs - the Erdős-Rényi, and Newman-Strogatz-Watts graphs, and assume that every vertex transmits to each of its neighbors at times that are independent and identically distributed as exponential waiting times. An Erdős-Rényi graph consists of n vertices, where each possible edge is present independently with probability p . We will use $p = \lambda/n$ in order to keep the expected degree of every vertex independent of n .

A Newman-Strogatz-Watts graph assumes a fixed, known distribution for the degrees of vertices, drawing from this distribution to choose the vertex degrees. We interpret these degrees as ‘half-edges’ attached to their corresponding vertices, and then match up these half-edges randomly. Of particular interest are power law distributions for the degree, where the fraction of vertices of degree k , $p_k \sim Ck^{-\alpha}$ as $k \rightarrow \infty$.

Parameters relating to the likelihood of connection in a statistical sense for each type of graph (λ , the unscaled uniform probability of connection for the Erdős-Rényi case, and α , the exponent on the power law degree distribution for the Newman-Strogatz-Watts case) correspond to the level of connectivity, and the rate of the exponential distribution waiting times to the ‘speed’ of transmission.

Subsequently, we run simulate transmission scenarios on specific instances of generated random graphs and study the path of adoption. Transmission continues until either all vertices have been transmitted to, or all vertices connected to vertices that can transmit have already been transmitted to. We consider

changes in the path, and then the final percentage of adoption, as a result of changing basic parameters such as the size of the random graph, and parameters in the algorithms to generate the graphs.

In both cases, the rate of the exponential distribution, as a result of independence between waiting time in the model, primarily affects the time taken to reach the final level of adoption, and the structure of the adoption path to a much lower degree. Essentially then, in this model of transmission, the structural parameters of the random graph determine the distribution of the final percentage adoption achievable, and the rate of the exponential transmission determines the time taken. Therefore, we will use a rate of transmission $\mu = 1$ for all simulations without loss of generality and observe that time gets scaled by the rate.

Our simulations, due to computational time constraints², typically have fewer than 500 vertices. Though n is small, these sizes are interesting for two reasons: First, for the application outlined, technology diffusion across industry, this is actually a reasonable number of firms to consider. Second, random graph analysis is typically conducted asymptotically, that is as $n \rightarrow \infty$. One of the points we consider, then, is to what extent classical results apply to graphs of scales not so large.

As an example of the types of asymptotic results we have in mind, consider one of the basic results of random graph theory: For the Erdős-Rényi random graph, there is a key threshold for λ at 1: for $\lambda < 1$, the largest connected component of the graph remains small (roughly $O(\log n)$), and for $\lambda > 1$, a *giant component* emerges, and the largest component is $\theta(\lambda)n$, where $\theta(\lambda)$ is a constant

²See Appendix A for figures showing runtimes for simulations as functions of the parameters of the models.

depending on λ [5, pp. 4 – 7]. This change in behavior around a threshold is referred to as a *phase transition*. A similar phase transition occurs in the Newman-Strogatz-Watts case, depending on ν , a parameter defined in Section 2.2.

We use the Branching Process (initially a population growth model, where at each iteration each object from the previous iteration generates offspring) to model the process of exploring a connected component. Section 3.1 defines the process formally, but the intuitive connection between this process and random graphs is that starting with an individual vertex³ we can view its neighbors as its children. The neighbors' neighbors then become grandchildren. If the number of vertices at distance k is Z_k , then Z_k behaves like a branching process [5].

We find that the structure of the model implies that the expected percentage adoption depends largely on the expected size of the *giant component* of the graph. Thus, in the Erdős-Rényi case, we model the process of transmission as an exploration of its connected component, which behaves like a branching process. Using the probability that this branching process becomes extinct, we find a predicted level of adoption which closely matches simulation results. Subsequently, we see that it is possible to identify the two parameters of the model sequentially in this case. Section 3.3 explains the calculation and compares expected adoption with simulation results, concluding that the assumption of an Erdős-Rényi structure makes both λ and μ identifiable in that order.

As simulated final adoption times are effectively an average over scenarios where the connected component hit is the giant component and where the component is not the giant component, we also use simulation times and extinction probabilities of branching processes to estimate the size of connected compo-

³This is how our simulations work – we choose a vertex, and the process gets transmitted to precisely the set generated by a branching process.

nents given they are not the giant component as a function of the probability of connected in the Erdős-Rényi case. Finally, we also consider the average distance between any two vertices of an Erdős-Rényi random graph as a way to further understand the transmission process and results.

CHAPTER 2

TRANSMISSION ALONG RANDOM GRAPHS

We proceed by considering two different types of random graphs - the Erdős-Rényi, and Newman-Strogatz-Watts graphs - as models for generating sample graphs. Subsequently, by assuming a probability distribution for transmission between connected nodes, we can simulate the process of transmission. The Erdős-Rényi graph assumes any two nodes are connected independently with probability p , and the Newman-Strogatz-Watts graph assumes a known distribution for degrees of nodes, and pairs up half-edges after degrees are sampled.

2.1 Erdős-Rényi Graph

We begin by making basic assumptions regarding the structure of the industry to make the problem more specific. Suppose that an industry with n firms can be modeled by a *random* graph with n nodes, and that node a , if possessing the technology and connected to node b , transmits this technology at an exponential rate with parameter μ . Thus, the probability that the transmission takes less than t units of time is given by the cumulative distribution function

$$F(t, \mu) = 1 - e^{-\mu t}$$

We will use $\mu = 1$, and observe that the expected transition time just scales with μ everywhere, and nothing changes about the structure of results.

Consider the simplest possible structure for a random graph - each of the

2.1. Erdős-Rényi Graph

possible $n(n - 1)/2$ edges is present¹ independently with probability λ/n [5, pp. 4 – 7]. Scaling down by a factor of n implies that increasing the size of the graph does not affect the expected degree of each node. Consequently, we can consider the probability of connection λ/n as representing the level of connectivity of the chosen industry, and the rate of exponential transmission μ as capturing the speed of transmission.

Observe that the degree distribution of any vertex, given this structure, is Binomial($n - 1, \lambda/n$), and that as $n \rightarrow \infty$, this converges to a Poisson distribution with mean λ . One of the primary features of the Erdős-Rényi graph, as mentioned in Chapter 1, is the dramatic change in its connectivity properties around the threshold $\lambda = 1$. We state the following two formal theorems regarding the emergence of a giant component from Durrett [5, pp. 39–44].

Theorem 2.1.1. *Let C_i be the connected set containing vertex i . Suppose $\lambda < 1$ and let $\alpha = \lambda - 1 - \log(\lambda) > 0$. If $a > \frac{1}{\alpha}$, then*

$$P\left(\max_{1 \leq i \leq n} |C_i| \geq a \log n\right) \rightarrow 0. \quad (2.1)$$

Theorem 2.1.2. *Suppose $\lambda > 1$. Then \exists constant β so that with probability $\rightarrow 1$, there is only one component of the random graph with more than $\beta \log n$ vertices. The size of this component $\sim (1 - \rho(\lambda))n$ where $\rho(\lambda)$ is the extinction probability for the Poisson(λ) branching process².*

Thus, for $\lambda < 1$, we have an upper bound on the size of the largest connected component, and for $\lambda > 1$, a lower bound on the size of the largest component, as well as a formal statement about the uniqueness of the component.

¹Loops and multiple edges between nodes are excluded here, nor are edges directed.

²See Section 3.1 for a formal definition of the extinction probability of a branching process.

2.1.1 Algorithms

In the Erdős-Rényi case, generating the random graph is relatively simple:

1. Choose n , the number of vertices, λ , the unscaled probability of connection
2. Generate a sequence of i.i.d. binary random variables ζ_i , distributed 1 with probability p and 0 with probability $1 - p$, for $1 \leq i \leq n(n - 1)/2$.
3. Populate the upper right half of an $n \times n$ matrix G with this sequence of ζ_i , and reflect the sequence on the lower half to obtain a symmetric matrix corresponding with an undirected graph.

As noted above, we set $\mu = 1$ with no loss of generality, as becomes clear with the algorithm for simulating transmission; changing μ has a consistent effect on the expected time of individual jump. We use the following steps to simulate a transmission process given a specific random graph, corresponding to a matrix G generated as explained above:

1. Let H_j be the set of vertices possessing the technology at iteration j , with $H_0 = \{1\}$.
2. Let P_j be the set of vertices to which transmission is possible *after* iteration j . Let N_i be the set of neighbors of vertex i . Then $P_0 = N_1$, and in general

$$P_j = \left(\bigcup_{i \in H_j} N_i \right) - H_j$$

and the set of vertices to which transmission is possible is the set of vertices connected to vertices to which transmission has already occurred³.

³For two sets A and B , $A - B = A \cap B'$.

2.1. Erdős-Rényi Graph

3. For each vertex $i \in P_j$, calculate w_i , the number of vertices in H_j connected to vertex i .
4. Associate with each vertex $i \in P_j$ an exponentially distributed random variable $T_i \sim \exp(\mu \times w_i)$ (with these variables t_i mutually independent).
5. Sample $t_j \sim \min_{i \in P_j}(T_i)$ as the time between iterations $j - 1$ and j . Then t_j is distributed⁴ $\exp\left(\sum_{i \in P_j} \mu \times w_i\right)$. Choose the vertex transmission occurs to as vertex k with probability

$$\frac{w_k}{\sum_{i \in P_j} w_i}$$
 and note that $H_{j+1} = H_j \cup \{k\}$.
6. Return to step 2 and repeat until $P_j = \emptyset$.

Observe that the lack of memory property of the exponential distribution is implicitly used here in making the algorithm as described above consistent with one where an exponentially distributed variable is associated with every relevant edge, and actual transmission occurs at a time which is their minimum. Note, also, that the transmission process continues until the connected component containing the vertex labelled as 1 is reached. The proportion of vertices to which transmission occurs, then, depends largely on whether the initial vertex chosen is in the giant component. We use this observation to compare simulated percentages of adoption to expected adoption as λ varies in Section 3.3.

⁴Recall that for exponentially distributed random variables $x \sim \exp(\mu_1)$ and $y \sim \exp(\mu_2)$ $\min(x, y) \sim \exp(\mu_1 + \mu_2)$.

2.1.2 Simulation Results

Let H be the final set of vertices possessing the technology⁵. Define the fraction of eventual adoption as $|H|/n$. By keeping track of the simulation times for each of the vertices we can plot the adoption history of individual simulations. Figure 2.1 shows the adoption path for 10 independent runs of this simulation with $n = 500$, $\lambda = 2$, $\mu = 1$ and 10 trials.

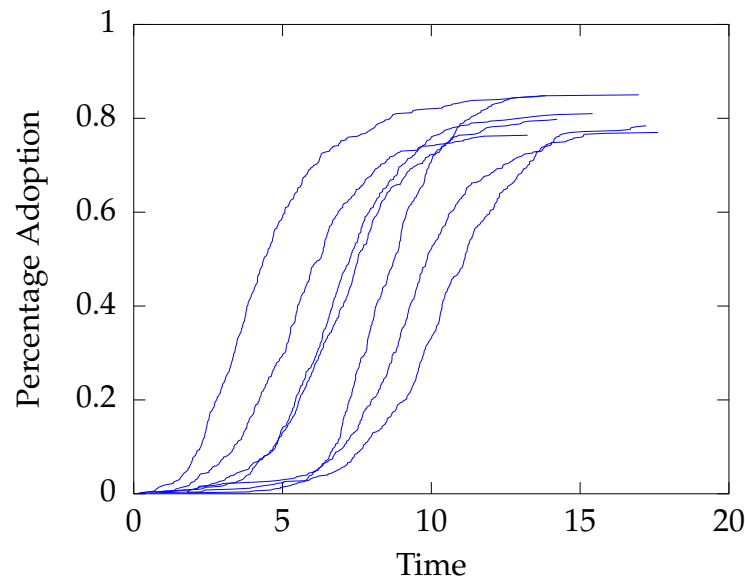


Figure 2.1: Erdős-Rényi Adoption times

Note that only for 7 of these trials does the technology reach a significant proportion of the vertices. For a specific λ , qualitatively it is clear that the variation in the clear S-curve is when adoption begins to accelerate, with a fairly similar final percentage of adoption given that node 1 is in the giant component.

The unscaled probability of connection, λ , which determines the expected

⁵Note that $H = C_1$ using the notation from Theorem 2.1.1.

2.1. Erdős-Rényi Graph

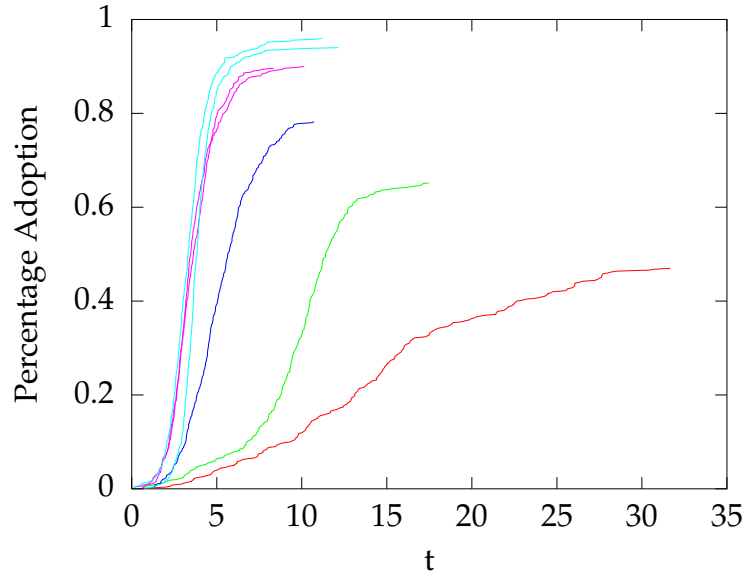


Figure 2.2: Erdős-Rényi Adoption times with λ varying within $[1, 2]$

proportion of vertices in the giant component, affects the final percentage adoption significantly. Figure 2.2 shows the adoption path from simulations with varying λ . As λ increases gradually from 1 to 2, we see an increase in the final percentage of adoption, as well as an increase in the overall rate of transmission. Thus, as λ increases, adoption paths shift up and to the left. An increase in λ also increases the probability that vertex 1 is in the giant component, and thus the likelihood that significant adoption occurs. It is also interesting to note that the exponential increase in the rate of transmission occurs much faster. Finally, we note that these are individual simulation paths, and some simulations start at a small connected component, and no significant adoption occurs.

To analyze the ‘S-curve’ behavior, we compare results to the classic S-curve – the solution to the logistic differential equation

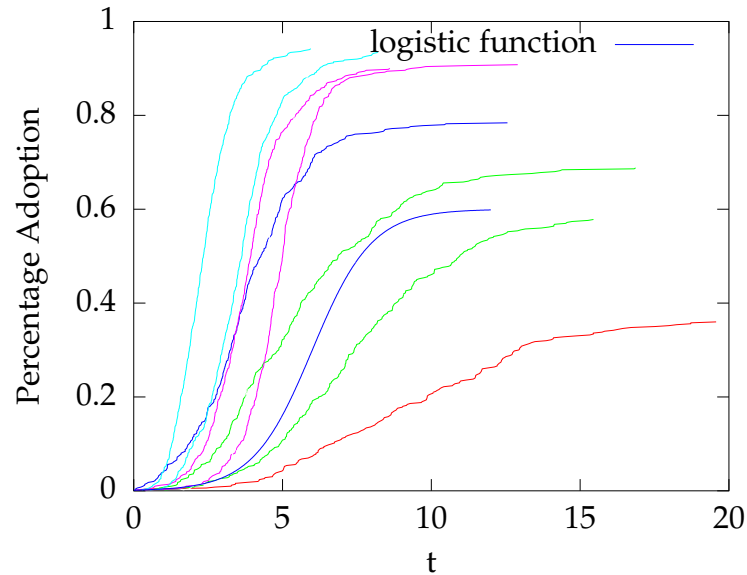


Figure 2.3: Erdős-Rényi final times with varying λ and superimposed scaled inverse logistic function

$$\frac{d}{dt}P(t) = P(t)(1 - P(t)) \quad (2.2)$$

which has solution

$$P(t) = \frac{e^t}{e^t + e^c} \quad (2.3)$$

where for convenience we set the constant of integration c so that $e^c = 1$.

Figure 2.3 shows the results of a simulation with identical parameters to those considered above, with $\mu = 1$ again, and includes a scaled solution to the logistic differential equation. It has been scaled to take time 12 and achieve 60% final adoption.

2.1. Erdős-Rényi Graph

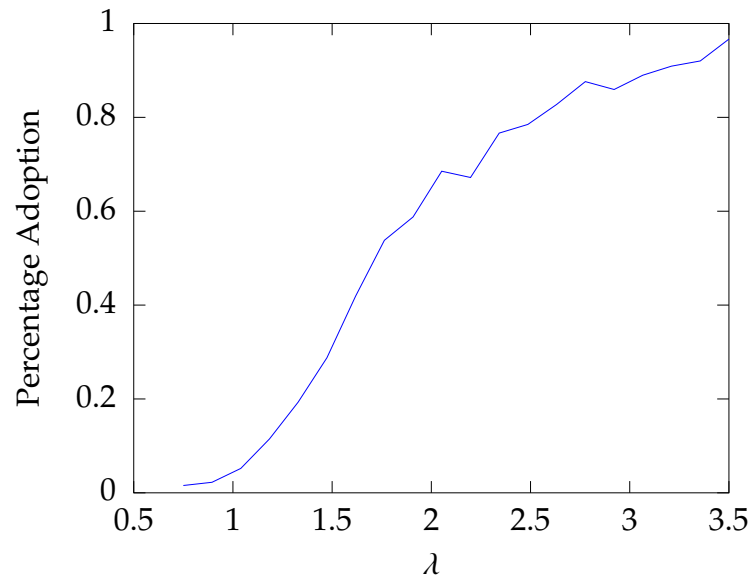
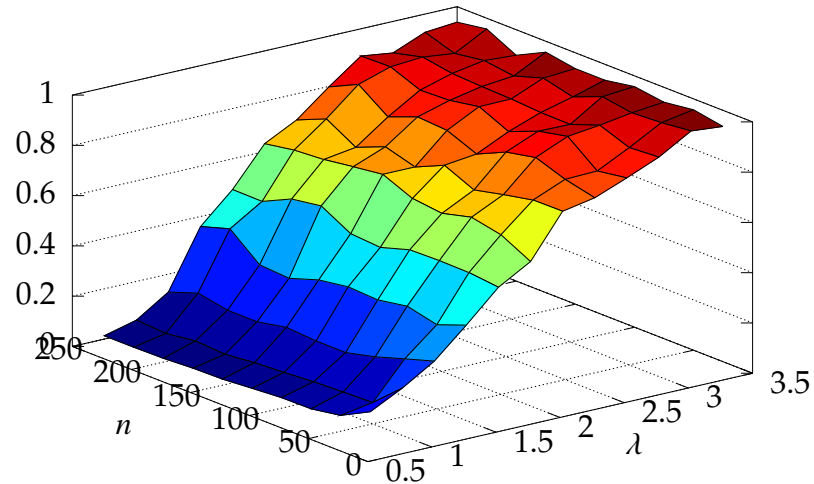


Figure 2.4: Erdős-Rényi final percentage adoption with $n = 300$

It is clear that the final percentage of adoption depends only on λ , and not on μ . Moreover, much of the variation is in how long transmission takes (and whether the giant component is ‘hit’). Consequently, we can summarize results by just considering the average final percentage of adoption over several trials as a function of λ . Figure 2.4 shows final percentage adoption as a function of λ , for $n = 300$. See Section 3.3 for a comparison of these simulation results with an analytical expression for average percentage adoption.

The final question to consider is the impact of n on the structure of results. Figure 2.5 shows the final percentage of adoption with varying λ and n , and it is clear that percentage adoption depends primarily on λ . Section 3.3 formulates the observations made here to compare simulated adoption as a function of λ for $n = 250$ with expected adoption.

Figure 2.5: Erdős-Rényi Adoption as a function of λ and n

2.2 Newman-Strogatz-Watts Graph

The simulations and results so far suggest that the primary data observations of interest are the final percentage of adoption, and the eventual time until transmission is completed, assuming this particular structure of the random graph. From these two points it should be possible to conclude the parameters for the random graph, though this effectively ignores the structure of the adoption path. Consequently, we next consider a different model for the underlying random graph, and study how this affects the results so far.

In an Erdős-Rényi random graph, vertices have degrees that have asymptot-

2.2. Newman-Strogatz-Watts Graph

ically a Poisson distribution. However, in social and communication networks, the distribution of degrees often has a power law form, and hence the fraction of vertices of degree k , $p_k \sim Ck^{-\alpha}$ as $k \rightarrow \infty$, for a constant C [5, pp. 71 – 76]. Consequently, a fixed degree degree distribution random graph provides, for large degrees, asymptotic behavior perhaps better matching empirical observations. For the particular problem being referred to here, though, it is unclear whether industrial connections mirror the behavior observed with social and broader communication networks⁶.

We use a slight modification of the Newman-Strogatz-Watts approach [6, 7]: Let d_1, \dots, d_n be independent and distributed $P(d_i = k) = p_k$. As d_i is the degree of vertex i , condition on the requirement $E_n = \{d_1 + \dots + d_n \text{ is even}\}$. Given this fixed degree distribution, draw degrees for vertices, and then match up the half pairs to produce a random graph. Our approach differs in that loops are explicitly not permitted.

As mentioned in Chapter 1, a phase transition occurs in the Newman-Strogatz-Watts case as well, leading to the existence of a giant component. The formal result we quote regards the original case in which loops are permitted, but the threshold presented remains a lower bound on the condition for the existence of a giant component. Suppose, then, that p_k is the degree distribution of the first vertex being analyzed. Given this information, it is no longer true that its neighbors have the same distribution, as vertices with degree k are k times as likely to be chosen to be connected. This leads to a size-biased degree distribution

⁶To the extent that the work done is also applicable to modeling other types of transmission, this particular distribution remains of interest. Particular firms are also likely to have a large impact on technology adoption; the rubric of an industry leader that is commonly followed is well known.

$$q_{k-1} = \frac{kp_p}{\sum_k kp_k} \quad (2.4)$$

where the index on the left hand side is $k - 1$ as it is already known that the ‘second’ generation of vertices is connected to the first one. We choose α here so that p_k has finite second moment, and thus q_k has a finite mean, ν . The condition on the existence of a giant component is that $\nu > 1$. We state the theorem from Durrett [5, pp. 71 – 76] below.

Theorem 2.2.1. *Let p_k have generating function*

$$G_0 = \sum_k p_k z^k$$

and q_k have generating function

$$G_1 = \sum_k q_k z^k.$$

The condition for the existence of a giant component is $\nu > 1$. In this case the fraction of vertices in the giant component is asymptotically $1 - G_0(\rho_1)$ where ρ_1 is the smallest fixed point of G_1 in $[0, 1]$.

2.2.1 Algorithms

In this case, we use the following algorithm to generate the graph:

1. Choose n , the number of vertices, α , the exponent of the degree distribution

2.2. Newman-Strogatz-Watts Graph

2. Generate a sequence of i.i.d. random variables d_i from the distribution⁷ $P(d_i = k) = Ck^{-\alpha}$ for $1 \leq i \leq n$. If E_n is not even, add let $d_1 = d_1 + 1$.
3. Consider each vertex i to have d_i half edges, and randomly pair these half edges. The approach differs from the strict Newman-Strogatz-Watts model in that loops are not permitted.
4. Populate the upper right half of an $n \times n$ matrix G with this sequence of edges, and reflect the sequence on the lower half to obtain a symmetric matrix corresponding with an undirected graph.

Multiple edges, though, are permitted, and can be interpreted as stronger connections between vertices, and this interpretation is consistent with simulation in that it enters in the parameter of the exponential waiting time for transmission between them. Once a random graph is generated, the simulation process is exactly analogous to that described in the case of the Erdős-Rényi random graph.

2.2.2 Simulation Results

We now present results corresponding to those shown in Section 2.1.2 for the Newman-Strogatz- Watts type random graph.

Figure 2.6 shows the adoption path for 10 independent runs of this simulation with $n = 500$, $\alpha = 2.3$, $\mu = 1$ and 10 trials. The difference between the two

⁷ C is chosen so that

$$C \sum_{n=1}^{\infty} n^{-\alpha} = 1$$

and the distribution is appropriately normalized.

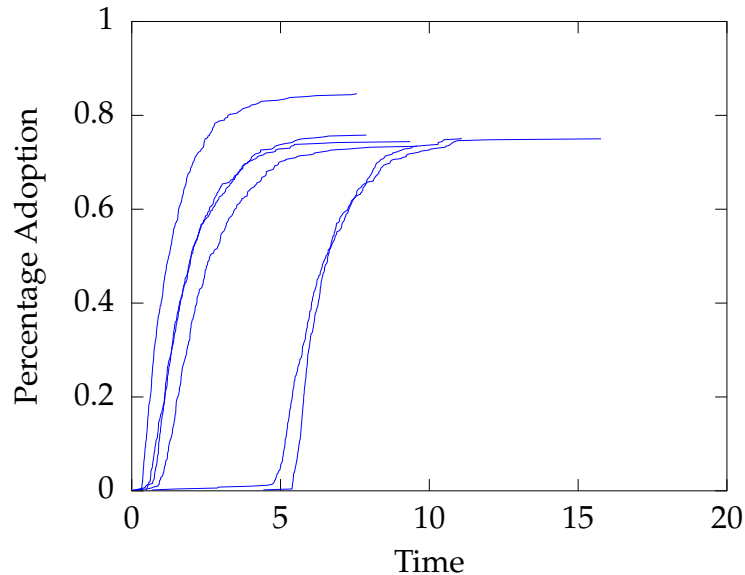


Figure 2.6: Newman-Strogatz-Watts Adoption times with $\alpha = 2.3$ and $\mu = 1$, 10 trials

assumed structures becomes clear with a comparison to Figure 2.1 – though both types of adoption paths can be described as S-curves, there is a distinct rapid jump from the initial, slow phase, which then slows into a much flatter final dissemination. On the other hand, the observations made above regarding the variation across simulations being largely in the time taken to reach this rapid jump, and in the final percentage of adoption to a much lesser degree, still seem valid.

As above, α is the primary parameter which determines the primary structure of the adoption path. Figure 2.7 shows the adoption path from simulation with varying α . As α decreases gradually from 3 to 2, we see an increase in the final percentage of adoption, as well as an increase in the overall rate of transmission. A comparison with Figure 2.2 further clarifies the distinction between

2.2. Newman-Strogatz-Watts Graph

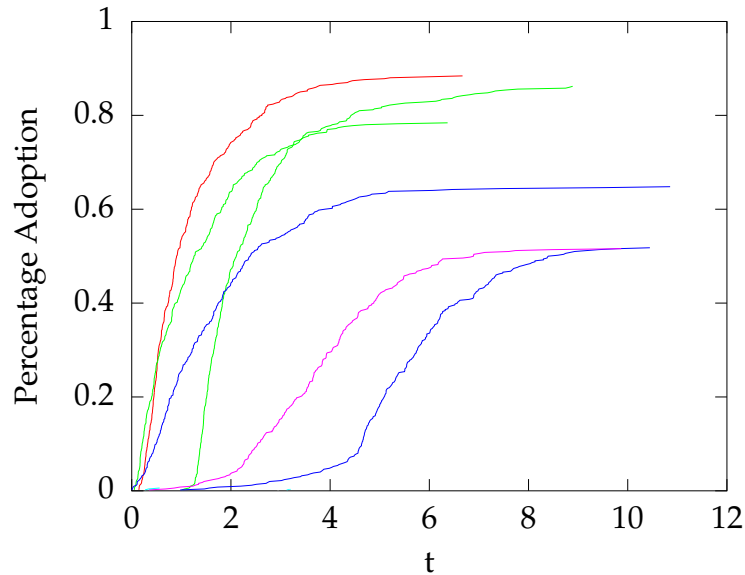


Figure 2.7: Newman-Strogatz-Watts Adoption times with α varying within $[2, 3)$

the two types of random graphs – in particular the final stage of transmission takes relatively longer, and as α approaches 2 (and the plot of the adoption path shifts to the left), the initial stage of the S-curve is no longer significant.

As above, we can summarize results by considering the average final percentage of adoption over several trials as a function of α . Figure 2.8 shows final percentage adoption as a function of α , for $n = 500$. It is, however, important to note that unlike the case for the Erdős-Rényi random graph structure, changing α changes the structure of the model to some extent. As the assumed degree distribution is $Ck^{-\alpha}$ with C changing with α as well, the expected degree of every vertex is changing⁸. In contrast, with the Erdős-Rényi case, the mean degree

⁸To some extent, this problem could be addressed by adding a constraint that the minimum degree be 3. In this case, most of the vertices will eventually be transmitted to. Another approach would be to use the re-wiring model considered by Strogatz, which begins with a connected ring and re-wires some edges randomly – unfortunately, this also means the graph is

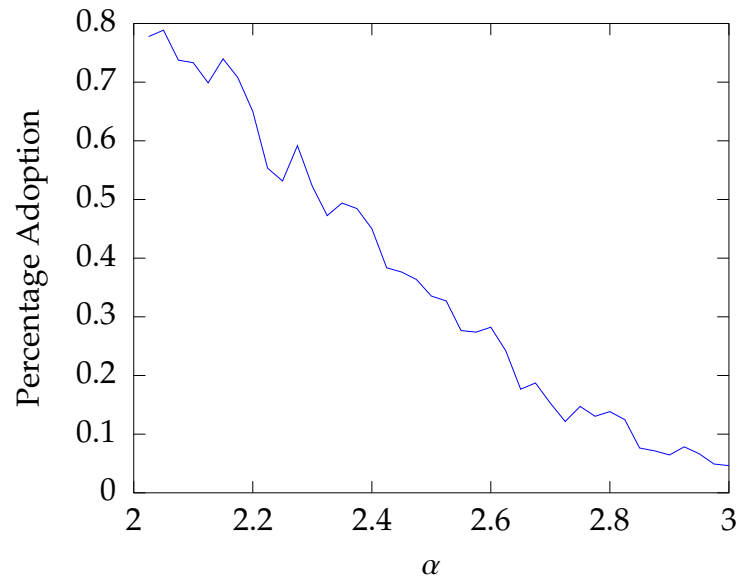


Figure 2.8: Newman-Strogatz-Watts Adoption with $n = 500$

does not change with λ as the probability is scaled.

Figure 2.9 shows the final percentage of adoption with varying α and n , and it is clear that percentage adoption depends primarily on α .

very likely to be one single component.

2.2. Newman-Strogatz-Watts Graph

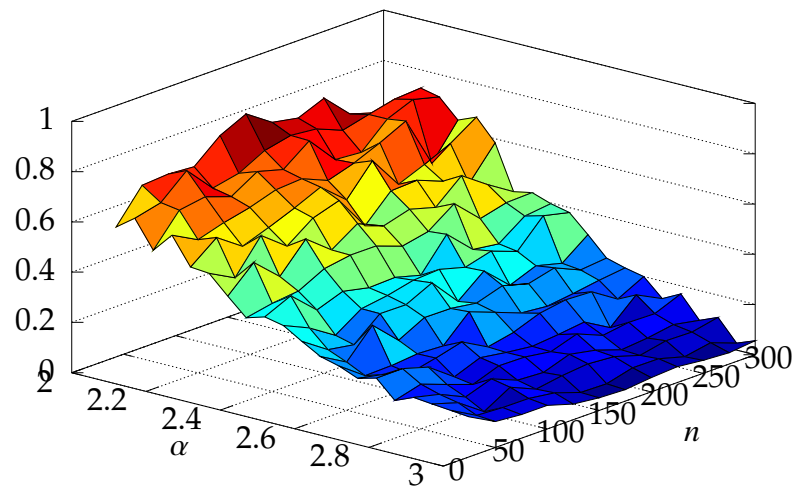


Figure 2.9: Newman-Strogatz-Watts Adoption

CHAPTER 3

ERDÖS-RÉNYI ADOPTION AND THE BRANCHING PROCESS

This chapter is motivated by observations stemming from the results shown in Figure 3.1, which shows the simulation time for transmission using the Erdős-Rényi random graph as a function of λ and n . As can be seen, for fixed n the simulation time increases initially as λ decreases, but then falls as λ continues to increase.

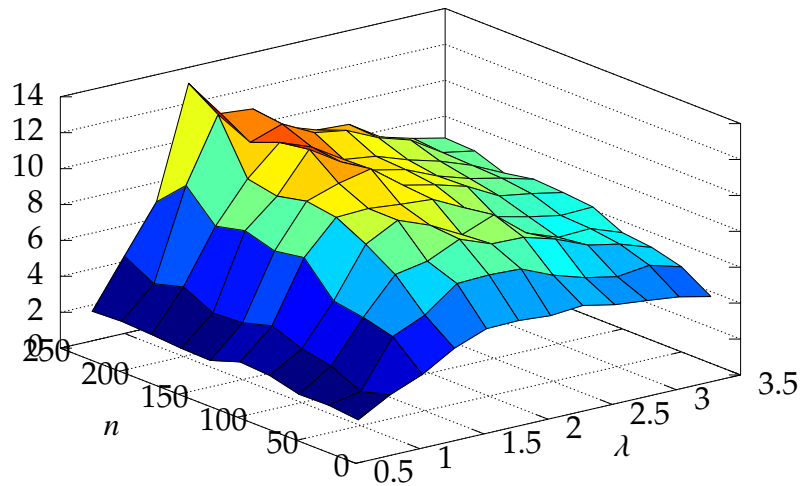


Figure 3.1: Erdős-Rényi Simulation Time

Intuitively, this follows from the fact that the simulation essentially consists of two different types of outcomes:

3.1. Diameter of Connected Component Given Extinction

- The chosen vertex is not in the giant component, and virtually none of the vertices are transmitted to. The process is relatively quick then.
- The chosen vertex is in the giant component, and virtually all vertices are transmitted to. The process takes relatively longer.

As the simulation time is averaged across these two scenarios, initially as λ increases, the probability the chosen vertex is in the giant component increases, increasing the average simulation time. Subsequently, as λ continues to rise, in most simulations the giant component is present, but as more edges are present transmission occurs much faster. By using the expected diameter of the giant component for large n , and the expected probability that a given edge is not in the giant component, we can estimate the size of the connected component of a vertex given it is not in the giant component as a function of λ .

3.1 Diameter of Connected Component Given Extinction

As Durrett [5, pp. 29 – 35] discusses, the evolution of the connected component of a given vertex of the Erdős-Rényi random graph can be approximated as a branching process, which we define formally below.

Definition 3.1.1 (Galton-Watson Process). *Let $X_i^t, i, t \geq 0$ be i.i.d random variables defined on \mathbb{Z}^+ . Define a sequence Z_t for $t \geq 0$ with $Z_0 = 1$, and subsequently*

$$Z_{t+1} = \begin{cases} X_1^{t+1} + \dots + X_{Z_t}^{t+1} & \text{if } Z_t > 0 \\ 0 & \text{if } Z_t = 0 \end{cases} \quad (3.1)$$

Z_t is then said to follow the Galton-Watson (branching) process.

As the branching process is defined, if $Z_t = 0$ for some t , the process is said to become extinct, whereas if $Z_t > 0 \forall t$, the process does not become extinct. Subsequently, asymptotic analysis of large graphs uses branching processes to analyze the growth of connected components of a chosen edge. Given an Erdős-Rényi structure, X_i^t is distributed Binomial with parameters $p = \frac{\lambda}{n}$ and n , where X_i^t represents the number of neighbors that the corresponding vertex has. Approximating the giant component as a branching process that survives forever, it is clear that if the chosen vertex is not in a giant component, the size of its connected component follows the size of a branching process that becomes extinct.

Now the probability that a branching process becomes extinct is given by the fixed point of its generating function on the interval $[0, 1)$ [8, pp. 29, 60 – 62]. Moreover, the generating function of a Binomial random variable with parameters p and n is

$$\sum_{k=0}^n P(X = k)z^k = \sum_{n=0}^k \binom{n}{k} (zp)^k (1-p)^{n-k}$$

and therefore

$$g_X(z) = (1 - p + pz)^n. \tag{3.2}$$

The probability that the branching process becomes extinct, then, is the fixed point of $g_X(z)$ for $z \in [0, 1)$.

Figure 3.2 shows the relevant graph for $\lambda = 1.5$ and $n = 500$. The extinction probability is approximately 0.42.

3.2. Estimation

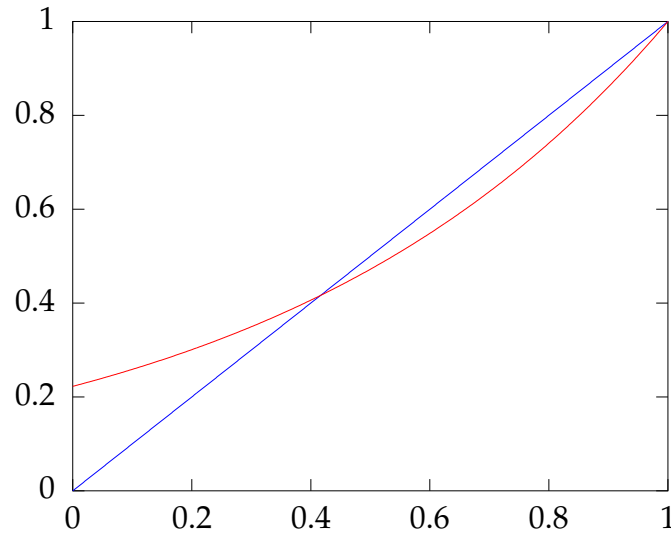


Figure 3.2: Erdős-Rényi Generating Function and Extinction probability, $\lambda = 1.5, n = 500$

3.2 Estimation

On order to get a point estimate of the overall population generated by the branching process (which corresponds to the size of the connected component), we need an estimate of the diameter of the giant component (to arrive at the expected simulation time when the chosen vertex *is* in the giant component). We quote the following result from [5, pp. 45 – 47].

Theorem 3.2.1. *Suppose $\lambda > 1$, and choose two vertices x and y randomly from the giant component of an Erdős-Rényi random graph. Then the following convergence in probability holds:*

$$\frac{d(x, y)}{\log n} \rightarrow \frac{1}{\log \lambda}.$$

Finally, suppose ρ , the fixed point of $g_X(z)$ for $z \in [0, 1)$, is the extinction probability of the branching process corresponding to an Erdős-Rényi graph with parameters λ and n . Then the expected simulation time is

$$\text{Simulation Time} = \rho c(\lambda) + (1 - \rho) \frac{\log n}{\log \lambda} \quad (3.3)$$

thus allowing a calculation of $c(\lambda)$ from the simulation time.

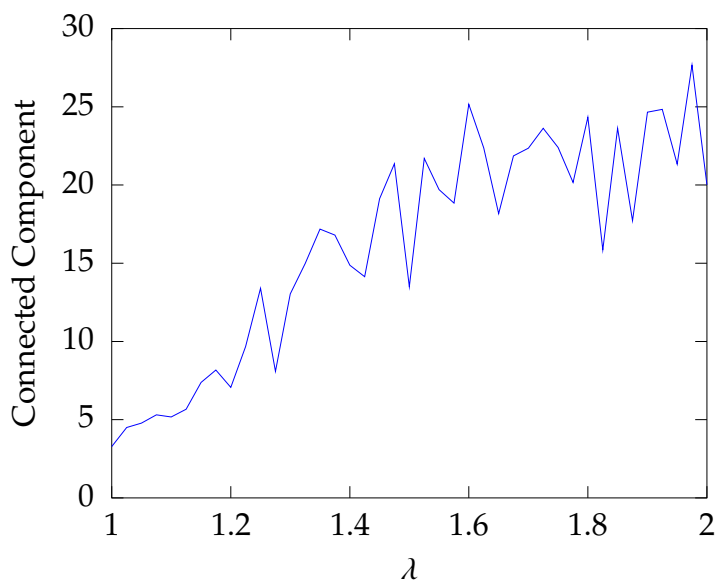


Figure 3.3: Erdős-Rényi connected component given extinction, $n = 500$

Figure 3.3 shows the diameter of the connected component as a function of λ , for $n = 300$.

3.3. Expected Adoption

3.3 Expected Adoption

A particularly interesting use of ρ , the probability that a branching process becomes extinct, is a simple expectation calculation that leads to a good analytical estimate of average percentage adoption. We proceed with a method similar to that used to formulate Equation 3.3. Essentially, as seen in the simulations seen in Section 2.1.2, if the initial vertex chosen happens to be a branching process which dies out, the percentage adoption is negligible, and there is minimal contribution to the average percentage adoption.

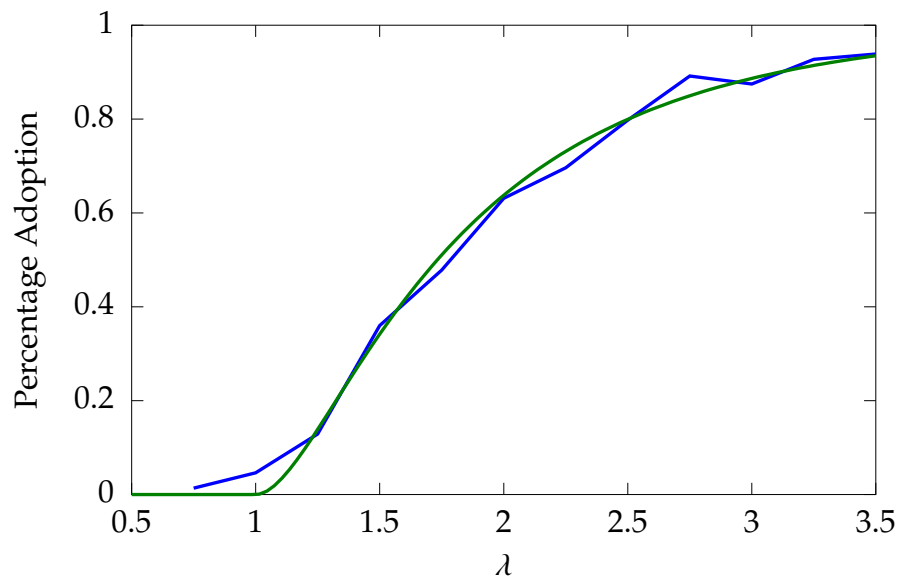


Figure 3.4: Erdős-Rényi simulated adoption (blue) and expected adoption (green)

Now the probability that the branching process does not die out is $1 - \rho(\lambda)$, which we will use twice to get the expected adoption to be

$$E\left(\frac{|H|}{n}\right) = [1 - \rho(\lambda)]^2 \quad (3.4)$$

as once the giant component is hit, one expects the entire giant component to be transmitted to. Figure 3.4 compares simulated adoptions for $n = 250$ as λ varies, and indeed this expression for the percentage adoption accurately describes the situation. Therefore, given empirical adoption paths and the size of the industry, if an Erdős-Rényi structure is assumed, we can find λ , as it corresponds to a unique probability of survival for the branching process. As there is a one-to-one correspondence between λ and the probability of extinction, it can be uniquely identified. Figure 3.5 shows λ as a function of percentage adoption.

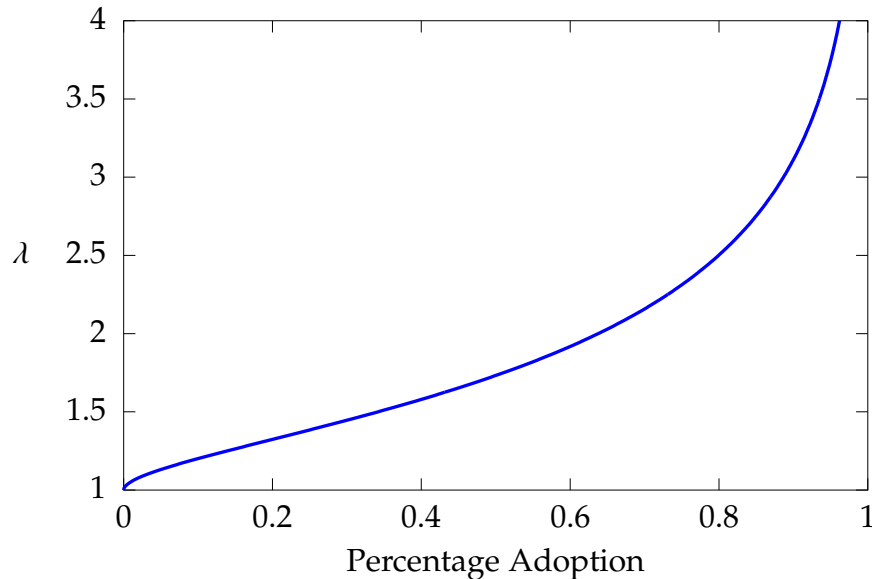


Figure 3.5: λ corresponding to percentage adoption for $n = 250$ given Erdős-Rényi structure

Subsequently, μ can be identified by the time taken for the process to complete.

CHAPTER 4

DISTANCE DISTRIBUTIONS

Another facet of the structure of the random graphs being generated is the distribution of $d(x, y)$ for any two vertices x, y . The nature of this distribution also provides insight into the results seen thus far. Figure 4.1 shows the distance distribution of Erdős-Rényi random graphs (averaged for 500 pairs of vertices each across 20 independently generated random graphs with $n = 1000$).

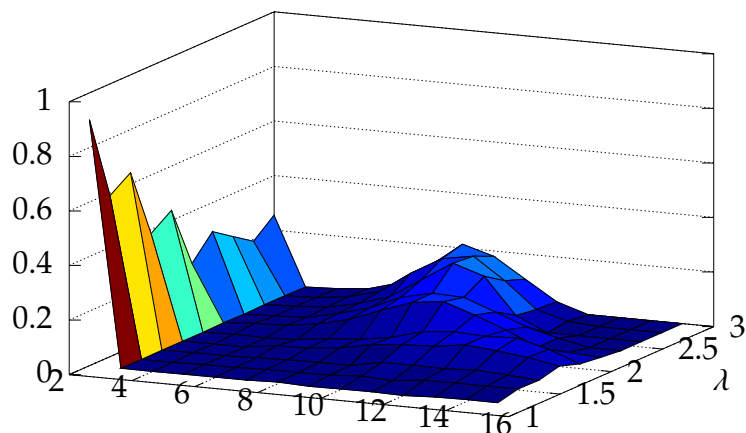


Figure 4.1: Erdős-Rényi distance distribution, $n = 1000$

Observe that level sets of Figure 4.1 (for fixed λ) correspond to probability ‘distributions’ which do not add up in probability to 1. This is because any two

vertices, randomly chosen, need not be connected, in which case the distance is ∞ . Moreover, observe that as λ increases, we begin to see a concentration of vertices in a middle bulge corresponding to the giant component.

The situation we analyze can be matched to that considered by van der Hofstad [9] – the problem of first passage percolation with weighted edges – by considering the realized exponential waiting times as weights on the edges they correspond to. Subsequently, given the manner of our analysis, vertices are reached precisely using the route of shortest cumulative weight, which here is minimal time. One of the results quoted in van der Hofstad [9] is that the expected graph distance between uniformly chosen pairs of connected vertices¹, for the Newman-Strogatz-Watts fixed distribution case with $\alpha \in (2, 3,)$ is uniformly bounded (see Figure B.2, which considers the case corresponding to this result).

¹Observe that this conditional assumption is not being made above, but rescaling so that the given distributions sum to 1 would correspond to this situation.

APPENDIX A
RUNTIME FIGURES

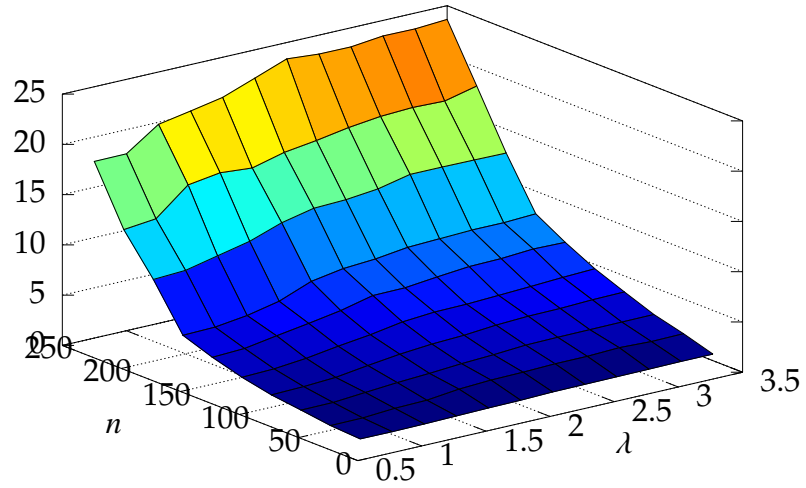


Figure A.1: Erdős-Rényi Runtime

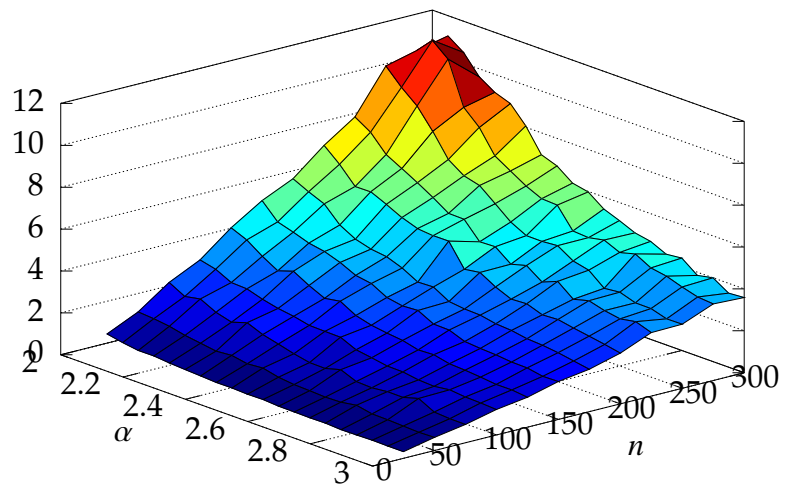


Figure A.2: Newman-Strogatz-Watts Runtime

APPENDIX B
MISCELLANEOUS FIGURES

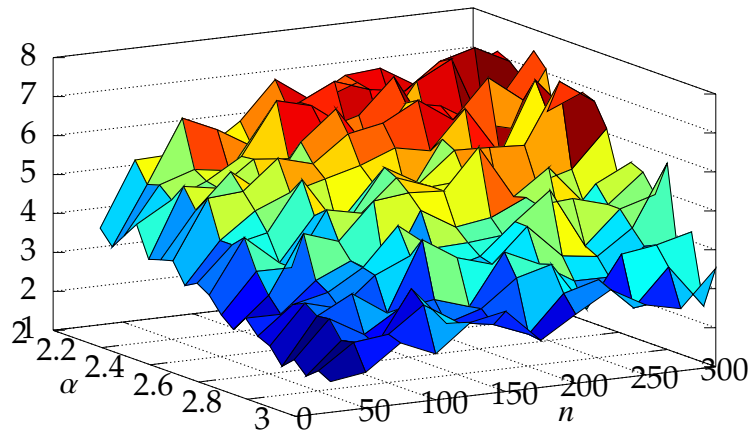


Figure B.1: Newman-Strogatz-Watts Simulation Time

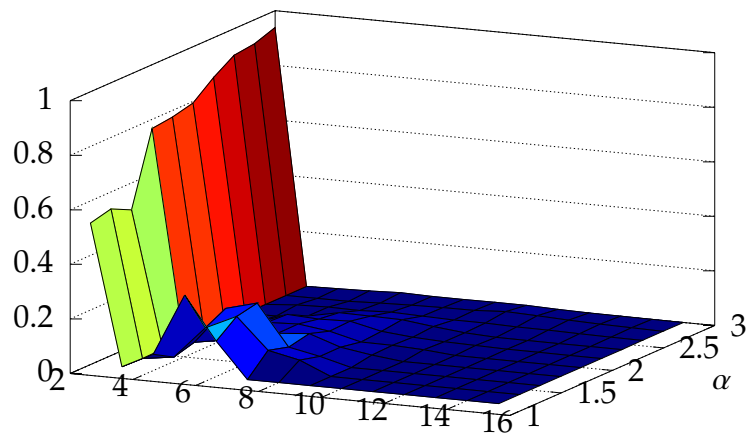


Figure B.2: Power law distance distribution, $n = 1000$

REFERENCES

- [1] Larry Blume, "The statistical mechanics of best-response strategy revision," *Games and Economic Behavior*, vol. 11, no. 2, 1995.
- [2] Stephen Morris, "Contagion," *The Review of Economic Studies*, vol. 67, no. 1, pp. 57–78, 2000.
- [3] H. Peyton Young and Mary A. Burke, "Competition and custom in economic contracts: A case study of illinois agriculture," *American Economic Review*, vol. 91, pp. 559–573, 2001.
- [4] H. Peyton Young, "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *American Economic Review*, vol. 99, pp. 1889–1924, 2010.
- [5] Rick Durrett, *Random Graph Dynamics*, Cambridge University Press, New York, 2007.
- [6] Mark Newman, Steven Strogatz, and Duncan Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 026118, 2001.
- [7] Mark Newman, Steven Strogatz, and Duncan Watts, "Random graphs models of social networks," *Proceedings of the National Academy of Science*, vol. 99, pp. 2566–2572, 2002.
- [8] Pierre Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, Springer, New York, 1999.
- [9] Remco van der Hofstad, "First passage percolation on random graphs," Transparencies of lecture in 'Order, disorder and double disorder' workshop, EURANDOM, August 22-26, 2009.