國 立 交 通 大 學

應用數學系

碩　士　論　文

臉書—小小世界

Facebook — A smaller world

研　究　生：高瑋琳

指導老師：傅恆霖　教授

中 華 民 國 一 〇 一 年 六 月

臉書 — 小小世界

Facebook – A smaller world

研究生：高瑋琳 Student: Wei-Lin Kao
指導教授：傅恆霖 教授 Advisor: Hung-Lin Fu

國 立 交 通 大 學

應 用 數 學 系

碩 士 論 文

A Thesis
Submitted to Department of Applied Mathematics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Applied Mathematics
June 2012
Hsinchu, Taiwan, Republic of China

中 華 民 國 一 ○ 一 年 六 月

臉書 — 小小世界

研究生：高瑋琳　　　　　　指導教授：傅恆霖 教授

國 立 交 通 大 學

應 用 數 學 系

摘要

“六度分隔理論”告訴我們：任兩個陌生人之間，平均最多只要透過六個人就可以認識。這世界人口如此地多，人際網路卻是個小世界。隨著網際網路的盛行，社群網路臉書的崛起，人和人的距離似乎又拉近了許多。實驗發現，在臉書上，任兩個陌生人的平均距離最多只需要五步，世界似乎更小了。在這篇論文中，我們提出了一個動態隨機圖模型來模擬臉書，將每一個用户看成點，好友關係看成邊，試著去刻畫隨機圖在時間很大的時候的樣貌。在模型的建構過程中，我們用不同的機率分佈來加入新的點和邊，和刪去舊有的點和邊，引入優先附加和相對弱者易被淘汰的概念，以符合臉書上的實際狀況。我們發現，這個模型的度分佈(degree distribution)也滿足冪次律(power-law)—小世界網路(small world network)的明顯特徵。因此，我們可以推斷，臉書也是一個小世界。

中 華 民 國 一○一 年 六 月

Facebook – A smaller world

Student: Wei-Lin Kao Advisor: Hung-Lin Fu

Department of Applied Mathematics
National Chiao Tung University

June, 2012

“Six degree of Separation” told us: any two individuals, selected randomly from almost anywhere on the planet, can know each other via a chain of average no more than six intermediate acquaintances. There are more tens of millions of people around the world, but the social network is a small world. With the dramatic growth of the World Wide Web and the Internet, even the rise of the social network-Facebook, the distance between two people seems much shorter than before. Through the experiment result, on Facebook, any two individuals are connected in five steps or fewer, on average. The world seems smaller. In this thesis, we construct a dynamic random graph model to simulate Facebook. We regard each user of Facebook as a vertex and the friendship between two users as an edge, and try to depict the pattern of the random graph as time being approximately infinity. In the process of the construction, we applied different probability distributions to adding new vertices and edges, and deleting existing vertices and edges. Based on the preferential attachment and the idea of the weaker tends to be weeded out, the model seems to conform with Facebook. Furthermore, we prove that the degree distribution satisfies the power-law, a common feature of the small world networks. Therefore, we conclude that Facebook is also a small world.

謝誌

碩一剛入學，有點迷失自己的方向，要特別謝謝小泡學姊和敏筠學姊的引薦，讓我有機會可以加入「傅家子弟」這個大家族，和傅老師以及學長姊們一起學習受益良多。

首先要感謝我的指導教授—傅恆霖老師，謝謝傅老師願意收我這分析組的學生，且不論是在研究、待人處事上，老師總是給予我許多指點和鼓勵，讓我有動力和方向可以繼續完成論文。同時也要謝謝傅家子弟的每一位學長姊，特別是敏筠學姊和施智懷學長，在作研究的過程中，總是很有耐心的指導我、給予我意見，提供我非常多的幫助。也要謝謝惠蘭、明輝、惠娟、貓頭、軒軒等學長姊，在學習的過程中給予我非常受用的建議和鼓勵。

再來就是要謝謝我的同學們，謝謝同研究室的惠雯、吉米、DK、連以及佳穎學姊，如果沒有你們互相扶持，或許我會畢不了業，當然你們也是我研究所生活最好的玩伴。謝謝組合組的姿慧、憶妏和沙發，熱心的對我伸出援手，讓我轉組卻不孤單。還有許多應數所的同學和學長姊，謝謝大家這段時間的照顧，讓我的研究所生活增添了許多美好回憶。

還要謝謝中山的學姊十大和大樹，因為有你們在，讓我在新竹還可以繼續打球，和願意聽我發牢騷以及一起懷念看海的日子。也要謝謝我以前的好朋友們，雖然大家分散在各地，但還是不會忘記給彼此鼓勵，因為擁有你們讓我覺得自己很幸福。

最後要謝謝的我的家人，謝謝你們對我的支持與鼓勵，因為距離相離有點遙遠，總是無法花很多的時間陪伴你們，也謝謝你們的體諒。現在，我可以很驕傲的說，我沒有讓你們失望!

Contents

1 Introduction

Throughout of this thesis, all notations, terms and graph properties on graph theory, we refer to the textbook written by D. B. West, see [16]. And for the facts on probability we refer to [9] written by R. Durrett.

1.1 Small world

Since the name "small world" was mentioned in [13] by S. Milgram, people start to study more carefully about the properties of the "small world." Among them "six degree of separation" attracts more attention. That is to say, any two individuals, selected randomly from almost anywhere on the planet, are "connected" via a chain of on average no more than six intermediate acquaintances. The idea can be traced back to the 1960s with Stanley Milgram's small-world experiments [13]. Milgram called it the "lost-letter technique." He selected 296 volunteers and asked them to dispatch a message to a specific individual, a stockholder living in the Boston suburb of Sharon, Massachusetts. The volunteers were told that they couldn't send the message directly to the target person (unless the sender knew him personally), but that they should route the message to a personal acquaintance that was more likely than the sender to know the target person. Milgram found that the average number of intermediate persons in these chains was 5.2 (representing about 6 hops). Milgram's observation became famous and passed into popular folklore in the phrase "six degree of separation." This is known as the small world network in recent years.

Empirical observations on not only social networks but also Internet graphs and biological networks have revealed similar properties. Generally, among the small world networks there are three common characteristics: (1) short average distance; (2) high clustering coefficient; and (3) the degree distribution generally follows a power law, that is, the number of vertices with degree k decay as $k^{-\lambda}$ for some exponent λ. Here average distance is the average length of the shortest path between any two vertices in the graph. And the clustering coefficient of vertex i is

$$C_i = \frac{2e_i}{k_i(k_i - 1)},$$

where e_i is the number of edges in the subgraph induced by vertex i and its k_i neighbors. The clustering coefficient C of a graph $G = (V, E)$ is defined to be the average of the clustering coefficients of all vertices,

$$C = \frac{\sum_{i=1}^{|V|} C_i}{| V |}.$$

Earlier study of network structure has focused on random graph. The classical model of random graph was introduced by Erd$\ddot{o}$s and Renyi in 1960 [11], of which vertices have equal probability of connecting each other. But it is not suitable for modeling these real-life networks since it does not have power-law degree sequences. This has driven the development of various alternative models for random graphs.

Beginning with the small-world model by Watts and Strogatz in 1998 [15] and the preferential attachment model by Barab$\acute{a}$si and Albert in 1999 [1], a lot of new random graph models have been defined and studied in recent years; see [5] for a survey. All the models can be classified into two groups: static (also known as explicit or off-line) and dynamic (also known as recursive or on-line). The difference between these two groups can be conveniently explained in the context of the algorithmic method for defining a random graph. In a static model, the set of vertices is fixed and the algorithm operates on the set of edges. However, in a dynamic model, the set of vertices and edges may change during the course of the defining algorithm.

One of the most widely-studied models of static models, after the random graph, is a model proposed by Watts and Strogatz in 1998 which shows both the short global separation and high clustering coefficient.

In addition, the most studied dynamic models are the birth-only ones (where only the addition of vertices and edges takes place). On 1999, Barab*á*si and Albert have proposed a simple model for internet growth which generates power-law through a random multiplication process–a kind of "rich get richer" phenomenon in which the vertices with most edges are the ones that gain new edges at the fastest rate.

In contrast, models that conclude birth and death (addition and deletion of vertices and edges) have been studied much lesser. Dorogovtsev and Mendes [10] studied a model which interleaves the addition of nodes and edges with a uniform deletion of edges. Later, Chung and Lu [3] and Cooper et al. [4], independently, have studied a dynamic model that combines the addition of nodes and edges with a uniform deletion of both vertices and edges. These birth-death models have also been found to generate graphs with power-law degree sequences with exponents that depend on the addition/deletion probabilities.

1.2 Facebook – An introduction

Facebook is a social networking service website originated in the United States and launched in 2004. It was established by Mark Zuckerberg with his college roommates and fellow students Eduardo Saverin, Dustin Moskovitz and Chris Hughes. As of now, Facebook has over 900 million active users around the world. Users must register before using the site, after which they may create a personal profile, add other users as friends, and exchange messages, including automatic notifications when they update their profile.

Furthermore, on 2011, a new study from Facebook and the University of Milan has shown that people in the world are better connected than before, with users of the social network now connected by less than five contacts. The study, which used data taken from Facebook's 721 million active users (more than 10% of the global population), with 69 billion friendships among them, shows that any two people on the site are on average separated by just 4.74 intermediate connections, see [2, 14]. That shows indeed Facebook is a smaller world.

In this thesis, we will consider a dynamic model where interleaves addition and deletion of vertices and edges for Facebook. As a consequence, we show that the model has the power-law degree distribution as well. By doing so, we have created a new mathematical model for Facebook network which shows that it is a "smaller" world.

2 A smaller world

We start with the design of our model.

2.1 The idea of design

We regard each user of Facebook as a vertex and the friendship between two users as an edge. Note that the probability with which a new user makes friends to the existing users is not uniform. But there is a higher probability that the new comer makes friends to someone who already has a large number of friends. This is the idea known as preferential attachment. Similarly, the older user more likely establishes friendship to the user who has more friends. Conversely, if a user has few friends on Facebook, he may withdraw from Facebook due to lack of interest. This implies that the vertices with lower degree have the higher probability of deletion. Furthermore, the friendship between two users may break with some unknown reasons. Hence, the probability of edge-deletion is the same among all edges of the graph.

By the above observation, the mathematical model for Facebook network can be designed as following.

2.2 Facebook model

As mentioned above, we let each user of Facebook be a vertex and the friendship between two users be an edge. Now, we can build up our Facebook model starting with two vertices which are connected by an edge.

Let the graph G_1 consist of two vertices connected by an edge, and in each discrete time-step $t+1$, $t > 0$, the graph G_{t+1} is constructed from G_t

in which one of the following four steps is carried out:

1. Birth of vertices: With probability $p_1 > 0$, a new vertex with one edge is added to the graph. To incorporate preferential attachment, the edge is connected to an existing vertex z chosen according to the following probability distribution:

$$\mathbb{P}_{t+1}(z = u) = \frac{d_t(u)}{\sum_{w \in V(G_t)} d_t(w)} = \frac{d_t(u)}{2e_t} \quad \text{for } u \in V(G_t), \qquad (1)$$

where $d_t(u)$ is the degree of the vertex u of G_t and $e_t = \mid E(G_t) \mid$.

2. Birth of edges: With probability $p_2 > 0$, a new edge is added between a vertex chosen by (1) and another vertex chosen randomly among $V(G_t)$.

3. Death of vertices: With probability $p_3 > 0$, an existing vertex z is chosen for deletion along with all the edges incident to z in G_t. To make small-degree vertices with larger possibility of deletion than the higher-degree ones, the vertex z is chosen by the probability distribution:

$$\mathbb{P}_{t+1}(z = u) = \frac{v_t - d_t(u)}{v_t^2 - 2e_t} \quad \text{for } u \in V(G_t), \qquad (2)$$

where $v_t = \mid V(G_t) \mid$.

4. Death of edges: With probability $p_4 = 1 - p_1 - p_2 - p_3$, one randomly chosen edge is deleted from $E(G_t)$.

In following analysis, we are concerned with the behavior in approximately infinity time. Hence, we ignore the influence of multiple edges and allow the isolate vertices to exist. And we can set the probability $p_1 > p_3$ and $p_2 > p_4$ so that an empty graph G_t occurs very rarely.

In order to obtain the degree distribution of the Facebook model, we need more informations. First, we are concerned with the number of vertices in the Facebook network.

2.3 Number of vertices in the Facebook network

Proposition 1. *The expectation of the number of vertices in the Facebook network at time t is $\Theta[(p_1 - p_3)t]$.*

Proof. We assume that $p_1 > p_3$ so that the number of vertices in the graph is indeed growing (on average). Hence it is assumed that $v_t > 0$ for all $t > 0$.

For all $t > 0$, $v_{t+1} = v_t + X_{t+1}$, where X_{t+1} is a discrete random variable and

$$X_{t+1} = \begin{cases} 1, & \text{with probability } p_1; \\ -1, & \text{with probability } p_3; \text{ and} \\ 0, & \text{with probability } p_2 + p_4. \end{cases}$$

Thus, the expectation of v_{t+1} is

$$\mathbb{E}[v_{t+1}] = \mathbb{E}[v_t] + (p_1 - p_3), \text{ for } t > 1. \tag{3}$$

Which implies that $\mathbb{E}[v_t] = \Theta[(p_1 - p_3)t]$. ■

2.4 Number of edges in the Facebook network

In this section, we consider the number of edges in the Facebook network.

Proposition 2. *The expectation of number of edges in the Facebook network at time t is $\Theta\left[\frac{(p_1 - p_3)(p_1 + p_2 - p_4)}{p_1 + p_3} t\right]$.*

Proof. For convenience, we assume $p_2 > p_4$ to ensure that the number of edges is growing(on average).

Let $N_k(t)$ be the number of vertices of degree k in G_t, and define $N_{-1}(t) = 0$ for all t.

For all $t > 0$, $e_{t+1} = e_t + Y_{t+1}$, where Y_{t+1} is a discrete random variable specified by

$$Y_{t+1} = \begin{cases} 1, & \text{with probability } p_1; \\ 1, & \text{with probability } p_2; \\ -k, & \text{with probability } p_3 N_k(t)\dfrac{(v_t - k)}{v_t^2 - 2e_t},\ k \in [0, \Delta(G_t)]; \\ -1, & \text{with probability } p_4; \end{cases}$$

where $\Delta(G_t)$ is the maximum degree of G_t.

Note that $\dfrac{(v_t - k)}{v_t^2 - 2e_t}$ is the probability of the event that a vertex with degree k is chosen to delete.

Hence, with G_t fixed, the number of edges after the $(t+1)$th step is

$$\mathbb{E}[Y_{t+1} \mid G_t] = p_1 + p_2 - p_3 \sum_{k \geq 0} k N_k(t) \frac{v_t - k}{v_t^2 - 2e_t} - p_4.$$

This implies that

$$\begin{aligned} \mathbb{E}[e_{t+1}] &= \mathbb{E}[e_t] + p_1 + p_2 - p_3 \mathbb{E}\left[\sum_{k \geq 0} k N_k(t) \frac{v_t - k}{v_t^2 - 2e_t}\right] - p_4 \\ &= \mathbb{E}[e_t] + p_1 + p_2 - p_4 - p_3 \mathbb{E}\left[\sum_{k \geq 0} \frac{k N_k(t)}{v_t - \bar{d}_t}\right] + p_3 \mathbb{E}\left[\sum_{k \geq 0} \frac{k^2 N_k(t)}{v_t^2 - 2e_t}\right], \end{aligned} \tag{4}$$

where $\bar{d}_t$ is the average degree of G_t.

Then we evaluate the two expectations multiplied by p_3 in (4).

Firstly, since

$$\sum_{k \geq 0} \frac{k N_k(t)}{v_t} = \bar{d}_t,$$

we obtain

$$\mathbb{E}\left[\sum_{k\geq 0}\frac{kN_k(t)}{v_t-\bar{d}_t}\right]=\mathbb{E}\left[\frac{v_t}{v_t-\bar{d}_t}\sum_{k\geq 0}\frac{kN_k(t)}{v_t}\right]=\mathbb{E}\left[\frac{2e_t}{v_t-\bar{d}_t}\right].$$

Secondly, using the approximation as in [7],

$$\sum_{k\geq 0}\frac{k^2N_k(t)}{v_t}\approx 2(\bar{d}_t)^2=\frac{8e_t^2}{v_t^2}.$$

Hence,

$$\mathbb{E}\left[\sum_{k\geq 0}\frac{k^2N_k(t)}{v_t^2-2e_t}\right]=\mathbb{E}\left[\frac{1}{v_t-\bar{d}_t}\sum_{k\geq 0}\frac{k^2N_k(t)}{v_t}\right]\approx\mathbb{E}\left[\frac{8e_t^2}{v_t^2(v_t-\bar{d}_t)}\right].$$

Then, by substituting them into (4), we have

$$\mathbb{E}[e_{t+1}]\approx\mathbb{E}[e_t]+p_1+p_2-p_4-p_3\mathbb{E}\left[\frac{2e_t}{v_t-\bar{d}_t}\right]+p_3\mathbb{E}\left[\frac{8e_t^2}{v_t^2(v_t-\bar{d}_t)}\right].$$

We consider the equation as $t\to\infty$, so the differences between the number of edges at time t and mean are relatively small. We may ignore some terms, yield

$$\mathbb{E}[e_{t+1}]-(1-\frac{2p_3}{\mathbb{E}[v_t-\bar{d}_t]})\mathbb{E}[e_t]-(\frac{8p_3}{\mathbb{E}[v_t^2(v_t-\bar{d}_t)]})\mathbb{E}[e_t]^2\approx p_1+p_2-p_4, \quad (5)$$

which is a non-linear difference equation.

Methods for solving such equations are known only for a few special cases. Because we add or delete at most k edges at one time, we may assume that $\mathbb{E}[e_t]=\varepsilon t$, where ε is a constant that does not depend on t. Substituting $\mathbb{E}[e_t]$ with εt and $\mathbb{E}[v_t]$ with $\Theta[(p_1-p_3)t]$ by Proposition 1 in (5), and then taking the limits as $t\to\infty$, we get

$$\varepsilon+\frac{2p_3}{p_1-p_3}\varepsilon=p_1+p_2-p_4.$$

Therefore,

$$\varepsilon = \frac{(p_1 - p_3)(p_1 + p_2 - p_4)}{p_1 + p_3}.$$

$$\text{i.e., } \mathbb{E}[e_t] = \Theta\left[\frac{(p_1 - p_3)(p_1 + p_2 - p_4)}{p_1 + p_3} t\right]. \qquad \blacksquare$$

2.5 Degree distribution in the neighborhood of the deleted vertex

Before turning our attention to the degree distribution in G_t, we need to evaluate one more quantity, namely the expectation of $N_{k,t}^{(1)}$ - the number of neighbors of degree k of the vertex chosen for deletion during step t. And $N_{k,t}^{(1)}(u)$ is the number of neighbors with degree k of the vertex u at time t.

Proposition 3. *The expectation of $N_{k,t}^{(1)}$ in the Facebook network at time t is approximate $k\mathbb{E}[N_k(t)]\mathbb{E}\left[\frac{v_t - 2\bar{d}_t}{v_t(v_t - \bar{d}_t)}\right]$.*

Proof. As G_t is fixed,

$$\begin{aligned}
\mathbb{E}[N_{k,t}^{(1)} \mid G_t] &= \sum_{u \in V(G_t)} N_{k,t}^{(1)}(u) \frac{v_t - d_t(u)}{v_t^2 - 2e_t} \\
&= \frac{1}{v_t - \bar{d}_t} \sum_{u \in V(G_t)} N_{k,t}^{(1)}(u) - \frac{1}{v_t(v_t - \bar{d}_t)} \sum_{u \in V(G_t)} N_{k,t}^{(1)}(u) d_t(u).
\end{aligned} \tag{6}$$

Then, note that the two summations of (6) are

$$\sum_{u \in V(G_t)} N_{k,t}^{(1)}(u) = kN_k(t)$$

and

$$\sum_{u \in V(G_t)} N_{k,t}^{(1)}(u) d_t(u) = \sum_{i=1}^{N_k(t)} \sum_{j=1}^{k} d_{k,i,j,t}.$$

Here $d_{k,i,j,t}$ denotes the degree of the jth neighbor of the ith vertex of degree k after step t. It may be approximated by the average degree $\bar{d}_t^{(1)}$ of a random neighbor of a random vertex. This quantity $\bar{d}_t^{(1)} \approx 2\bar{d}_t$. Hence we get

$$\begin{aligned}\mathbb{E}[N_{k,t}^{(1)} \mid G_t] &\approx \frac{kN_k(t)}{v_t - \bar{d}_t} - \frac{2kN_k(t)\bar{d}_t}{v_t(v_t - \bar{d}_t)} \\ &= \frac{kN_k(t)}{v_t - \bar{d}_t}(1 - \frac{2\bar{d}_t}{v_t}).\end{aligned}$$

By taking the expectations of both sides, we obtain

$$\mathbb{E}[N_{k,t}^{(1)}] \approx k\mathbb{E}[N_k(t)]\mathbb{E}\left[\frac{v_t - 2\bar{d}_t}{v_t(v_t - \bar{d}_t)}\right]. \qquad \blacksquare$$

2.6 Degree distribution of the Facebook network

Finally, after finishing the above analytic results, we turn our attention to the degree distribution of the graph G_t.

Theorem 2.6.1. *If* $\frac{(p_1 + p_2)(p_1 + p_3)}{2p_1(p_3 + p_4) + 2p_2p_3} > 1$ *and* $\frac{2p_1(p_1 + p_2 - p_4)}{p_2(p_1 - p_3) + p_1(p_1 - p_3 - 2p_4)} > -1$, *the mathematical model we construct for Facebook network satisfies the small world phenomenon: the degree distribution obeys the power-law.*

Proof. By analyzing the change in $N_k(t)$ between the t th and the $(t+1)$ th step, we have

$$\begin{aligned}&\mathbb{E}[N_k(t+1) - N_k(t) \mid G_t] \\ &= p_1C_k^{(1)}(t) + p_2C_k^{(2)}(t) + p_3C_k^{(3)}(t) + p_4C_k^{(4)}(t) + p_1\delta_{k1},\end{aligned}$$

where

$$
\begin{aligned}
C_k^{(1)}(t) &= \frac{(k-1)N_{k-1}(t)}{2e_t} - \frac{kN_k(t)}{2e_t},\\
C_k^{(2)}(t) &= \frac{(k-1)N_{k-1}(t)}{2e_t} + \frac{N_{k-1}(t)}{v_t} - \frac{kN_k(t)}{2e_t} - \frac{N_k(t)}{v_t},\\
C_k^{(3)}(t) &= (k+1)N_{k+1}(t)\frac{v_t - 2\bar{d}_t}{v_t(v_t - \bar{d}_t)} - \frac{N_k(t)(v_t - k)}{v_t^2 - 2e_t} - kN_k(t)\frac{v_t - 2\bar{d}_t}{v_t(v_t - \bar{d}_t)},\\
C_k^{(4)}(t) &= \frac{(k+1)N_{k+1}(t)}{e_t} - \frac{kN_k(t)}{e_t}.
\end{aligned} \tag{7}
$$

Term $C_k^{(1)}(t)$ in (7) reflects the expected change in $N_k(t)$ due to the birth of vertices. When a new vertex with one edge is added to the graph G_t, if the other end of the edge connects to an existing vertex of degree $k-1$, then the number of vertices of degree k in G_{t+1} will increase by one. On the other hand, if the end vertex connects to a vertex of degree k, then the number of vertices of degree k in G_{t+1} will decrease.

Term $C_k^{(2)}(t)$ in (7) exhibits the expected change in $N_k(t)$ due to the birth of edges. As a new edge is added to the graph G_t, if a vertex of degree $k-1$ is chosen to being an end vertex, the number of vertices of degree k in G_{t+1} will increase. On the other hand, a vertex of degree k is chosen, then $N_k(t+1)$ will decrease.

Term $C_k^{(3)}(t)$ expresses the expected change in $N_k(t)$ due to the death of vertices. There are two different ways of deleting a vertex which can cause $N_k(t)$ to decrease: (a) a vertex of degree k is deleted; and (b) the deleted vertex is adjacent to one or more vertices of degree k. The expected fall due to deletion of a vertex of degree k is $\mathbb{E}[N_k(t)]\mathbb{E}[(v_t - k)/(v_t^2 - 2e_t)]$. In addition, Proposition 3 implies that the expected drop due to deletion of a vertex which has one or more neighbors of degree k is $k\mathbb{E}[N_k(t)]\mathbb{E}[(v_t -$

$2\bar{d}_t)/(v_t^2 - 2e_t)]$. In a similar manner, one may also lead to the increase of $N_k(t)$ due to deleting vertices.

Term $C_k^{(4)}(t)$ reflects the expected change in $N_k(t)$ due to the death of edges. If the deleted edge has an end vertex of degree $k+1$, then $N_k(t)$ increases. On the other hand, if an end vertex is of degree k, then $N_k(t)$ decreases.

The last term in (7) comes from the fact that the degree of a new vertex is always one.

Assume that $\mathbb{E}[N_k(t)]/t$ converges to a_k as $t \to \infty$. Notice that if $\mathbb{E}[N_k(t)]/t$ is not converging to any number, then there is no solution for a_k. To obtain a recursion for a_k, we take the expectation of (7) and find the limit as $t \to \infty$. By Proposition 1, 2 and 3, this yields

$$\begin{aligned}&[\alpha_2(k+2)+\beta_2]a_{k+2} + [\alpha_1(k+1)+\beta_1]a_{k+1} + [\alpha_0 k + \beta_0]a_k \\ &= 2p_1(p_1-p_3)(p_1+p_2-p_4)\delta_{k1},\end{aligned}$$

where

$$\begin{aligned}\alpha_2 &= -2p_1(p_3+p_4) - 2p_2p_3, \\ \beta_2 &= 0, \\ \alpha_1 &= (p_1+p_3)(p_1+p_2+2p_4) + 2p_3(p_1+p_2-p_4), \\ \beta_1 &= 2(p_1+p_2)(p_1+p_2-p_4), \\ \alpha_0 &= -(p_1+p_2)(p_1+p_3), \\ \beta_0 &= -2p_2(p_1+p_2-p_4). \qquad (8)\end{aligned}$$

To solve (8), we will use Laplace's method as described in [12] to solve the nonlinear homogeneous equation:

$$[\alpha_2(k+2)+\beta_2]a_{k+2}+[\alpha_1(k+1)+\beta_1]a_{k+1}+[\alpha_0 k+\beta_0]a_k=0, \quad \text{for } k\geq 1. \tag{9}$$

Assume that the solution of the homogeneous equation is of the form:

$$a_k=\int_a^b t^{k-1}h(t)\,dt, \tag{10}$$

where the function $h(t)$ and the limits of integration a, b are yet to be determined.

Note that integration by parts of (10) yields

$$ka_k=[t^k h(t)]_a^b-\int_a^b t^k h^{'}(t)\,dt. \tag{11}$$

Now, define

$$\phi_\alpha(t):=\alpha_2 t^2+\alpha_1 t+\alpha_0=\alpha_2(1-t)(\frac{\alpha_0}{\alpha_2}-t) \text{ and}$$
$$\phi_\beta(t):=\beta_2 t^2+\beta_1 t+\beta_0=\beta_1 t+\beta_0.$$

By substituting (10) and (11) into (9), we obtain

$$[t^k\phi_\alpha(t)h(t)]_a^b-\int_a^b t^k\phi_\alpha(t)h^{'}(t)\,dt+\int_a^b t^{k-1}\phi_\beta(t)h(t)\,dt=0.$$

Then (9) is satisfied if a, b and $h(t)$ are chosen such that

$$[t^k h(t)\phi_\alpha(t)]_a^b=0 \tag{12}$$

and

$$\frac{h^{'}(t)}{h(t)}=\frac{\phi_\beta(t)}{t\phi_\alpha(t)}. \tag{13}$$

By integrating both sides of equation (13), we get

$$h(t) = t^{\lambda_1}(1-t)^{\lambda_2}(\frac{\alpha_0}{\alpha_2} - t)^{\lambda_3},$$

where

$$\begin{aligned}
\lambda_1 &= \frac{\beta_0}{\alpha_0} = \frac{2p_2(p_1+p_2-p_4)}{(p_1+p_2)(p_1+p_3)}, \\
\lambda_2 &= \frac{\beta_0+\beta_1}{\alpha_2-\alpha_0} = \frac{2p_1(p_1+p_2-p_4)}{p_2(p_1-p_3)+p_1(p_1-p_3-2p_4)}, \\
\lambda_3 &= -\frac{\beta_0\alpha_2+\beta_1\alpha_0}{\alpha_0(\alpha_2-\alpha_0)} \\
&= \frac{2(p_1+p_2-p_4)[2p_2(p_1p_3+p_2p_3+p_1p_4)-(p_1+p_2)^2(p_1+p_3)]}{(p_1+p_2)(p_1+p_3)[(p_1+p_2)(p_1-p_3)-2p_1p_4]}.
\end{aligned}$$

Now, by our assumption

$$\frac{(p_1+p_2)(p_1+p_3)}{2p_1(p_3+p_4)+2p_2p_3} = \frac{\alpha_0}{\alpha_2} > 1$$

and

$$\frac{2p_1(p_1+p_2-p_4)}{p_2(p_1-p_3)+p_1(p_1-p_3-2p_4)} = \lambda_2 > -1,$$

then (12) is satisfied with $a = 0$ and $b = 1$. Using the notation $\asymp$ to denote that the left hand side is bounded from above and below by constants times the right hand side, and the notation $\sim$ to denote that the quotient of the right hand side and the left hand side converges to a constant. Hence we obtain

$$\begin{aligned}
a_k &= \int_0^1 t^{k-1}t^{\lambda_1}(1-t)^{\lambda 2}(\frac{\alpha_0}{\alpha_2} - t)^{\lambda 3}\, dt \\
&\asymp \int_0^1 t^{k+\lambda_1-1}(1-t)^{\lambda_2}\, dt \\
&= \frac{\Gamma(k+\lambda_1)\Gamma(1+\lambda_2)}{\Gamma(k+1+\lambda_1+\lambda_2)} \\
&\sim k^{-(1+\lambda_2)}, \quad k \geq 1.
\end{aligned} \tag{14}$$

Here, (14) is obtained by a formula in [6] (Table 1 (5) p.27).

Hence, asymptotically the degree distribution of G_t follows a power-law with exponent $1 + \lambda_2$. ∎

Moreover, we have the following corollary.

Corollary 2.6.2. *If $p_1 > p_3 + 2p_4$, the mathematical model for Facebook network satisfies the small world phenomenon: the degree distribution obeys the power-law.*

Proof. By Theorem 2.6.1, the degree distribution of the facebook model obeys the power-law, if the followings hold:

$$\frac{\alpha_0}{\alpha_2} = \frac{(p_1 + p_2)(p_1 + p_3)}{2p_1(p_3 + p_4) + 2p_2p_3} > 1$$

and

$$\lambda_2 = \frac{2p_1(p_1 + p_2 - p_4)}{p_2(p_1 - p_3) + p_1(p_1 - p_3 - 2p_4)} > -1.$$

Firstly, since p_1, p_2, p_3, $p_4 > 0$, $2p_1(p_3 + p_4) + 2p_2p_3 > 0$.

Hence we consider

$$\begin{aligned} &(p_1 + p_2)(p_1 + p_3) - (2p_1(p_3 + p_4) + 2p_2p_3) \\ &= p_1^2 - p_1p_3 + p_1p_2 - p_2p_3 - 2p_1p_4 \\ &= p_2(p_1 - p_3) + p_1(p_1 - p_3 - 2p_4). \end{aligned} \tag{15}$$

If $p_1 > p_3 + 2p_4$, then (15)> 0, i.e., $\dfrac{\alpha_0}{\alpha_2} > 1$.

Secondly, if $p_1 > p_3 + 2p_4$, then $p_2(p_1 - p_3) + p_1(p_1 - p_3 - 2p_4) > 0$ and $2p_1(p_1 + p_2 - p_4) > 0$. Which implies

$$\lambda_2 = \frac{2p_1(p_1 + p_2 - p_4)}{p_2(p_1 - p_3) + p_1(p_1 - p_3 - 2p_4)} > 0 > -1.$$

Hence, if $p_1 > p_3 + 2p_4$, then the two conditions of Theorem 2.6.1 hold. That is to say, if $p_1 > p_3 + 2p_4$, then the degree distribution of the facebook model obeys the power-law. ■

The condition $p_1 > p_3 + 2p_4$ of Corollary 2.6.2 is close to the situation of the real-network of Facebook, which is that the probability of birth of vertices is much higher than death of vertices and edges. This ensures that the graph model is growing.

Example 1. *If we pick* $p_1 = 0.5$, $p_2 = 0.25$, $p_3 = 0.125$, *then* $p_4 = 0.125$ *and* $p_1 > p_3 + 2p_4$. *Hence we obtain*

$$a_k \sim k^{-5}, \quad \textit{for } k \geq 1.$$ ■

3 Conclusion

In this thesis, we consider a "random" graph model that combines the addition and deletion of vertices and edges in order to set up a model which fits the Facebook network. As a consequence, we found that our model for Facebook generates graphs with asymptotically power-law degree distribution, the common feature of the small world networks. That is to say, we have more confidence to believe Facebook is also a small world. Indeed, it should be smaller.

As for future work, the model in this thesis is only in accordance with Facebook. It would be interesting to construct a general model for various networks. On the other direction of research, we need to find the average distance of the random graph model theoretically. So far, only experimental results are obtained for small worlds including Facebook network.

References

[1] A. L, Barabási and R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509-512.

[2] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. CoRR, abs/1111.4570, 2011.

[3] F. Chung and L. Lu, Coupling online and offline analyses for random power law graphs, Internet Math. 1 (4) (2004) 409-461.

[4] C. Cooper, A. Frieze and J. Vera, Random deletion in a scale-free random graph process, Internet Math. 1 (4) (2004) 463-483.

[5] A. Cami, Techniques for analyzing dynamic random graph models of web-like networks: An overview, Networks 51 (2008) 211-255.

[6] B. David, Nouvelles tables d'intégrales définies, Leide : P. Engels, 1867.

[7] N. Deo and A. Cami, Preferential deletion in dynamic models of web-like networks, Inf. Proc. Lett. 102 (2007) 156-162.

[8] M. Deijfen and M. Lindholm, Growing networks with preferential deletion and addition of edges, Physica A (2009) 4297-4303.

[9] R. Durrett, Probability: theory and examples, Wadsworth and Brooks/Cole, 2003.

[10] S. Dorogovtsev and J. Mendes, Scaling behaviour of developing and decaying networks, Europhys. Lett. 52 (1) (2000) 33-39.

[11] P. Erdős and A. Rényi, On random graphs, Publ. Math. 6 (1959) 290-297.

[12] C. Jordan, Calculus of Finite Differences, Rottig and Romwalter, 1939.

[13] S. Milgram, The small world problem, Psychology Today 2 (1967) 60-67.

[14] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the Facebook social graph. CoRR, abs/1111.4503, 2011.

[15] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440-442.

[16] D. B. West, Introduction to graph theory, Pearson Education Taiwan Ltd. 2008.