Facebook – A smaller world

研究生：高瑋琳

指導教授：傅恆霖 教授

Department of Applied Mathematics

National Chiao Tung University

Aug. 12, 2012

Outline

Section 1

Introduction

1. Introduction
   - Small world
   - Facebook – An introduction

Small world

- What is “small world” ?

Small world

- What is “small world” ?
- General properties of small world networks:

Small world

- What is “small world” ?
- General properties of small world networks:
  1. Short average distance.

Small world

- What is “small world” ?
- General properties of small world networks:
  1. Short average distance.
  2. High clustering coefficient.

Small world

- What is “small world” ?
- General properties of small world networks:
  1. Short average distance.
  2. High clustering coefficient.
  3. The degree distribution follows a power law – the fraction of vertices with degree k decay as $k^{-\lambda}$ for some exponent λ.

Facebook

- Social networking service website established by Mark Zuckerberg et. al in 2004.

Facebook

- Social networking service website established by Mark Zuckerberg et. al in 2004.
- On 2011, a new study from Facebook and the University of Milan shows that any two people on the site are on average separated by just 4.74 intermediate connections, [2, 3].

Section 2

Our main work

2 Our main work

- Facebook model
- Propositions
- Degree distribution of the Facebook network

Definition

- $G_t = (V(G_t), E(G_t))$: A graph with vertices set $V(G_t)$ and edges set $E(G_t)$.
- v_t: Number of vertices in G_t.
- e_t: Number of edges in G_t.
- $d_t(u)$: Degree of vertex u in G_t.
- $\bar{d}_t$: Average degree of G_t.

Facebook model

We let each user of Facebook be a vertex and the friendship between two users be an edge.

Facebook model

We let each user of Facebook be a vertex and the friendship between two users be an edge.

Let the graph G_1 consists 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:

Facebook model

We let each user of Facebook be a vertex and the friendship between two users be an edge.

Let the graph G_1 consists 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

Facebook model

We let each user of Facebook be a vertex and the friendship between two users be an edge.

Let the graph G_1 consists 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
2. Birth of edges

Facebook model

We let each user of Facebook be a vertex and the friendship between two users be an edge.

Let the graph G_1 consists 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
2. Birth of edges
3. Death of vertices

Facebook model

We let each user of Facebook be a vertex and the friendship between two users be an edge.

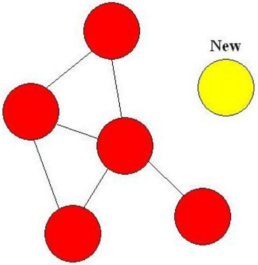
Let the graph G_1 consists of two vertices connected by an edge, and in each discrete time-step $t+1$, $t>0$, the graph G_{t+1} is constructed from G_t in which one of the following four steps is carried out:

1. Birth of vertices
2. Birth of edges
3. Death of vertices
4. Death of edges

Facebook model

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

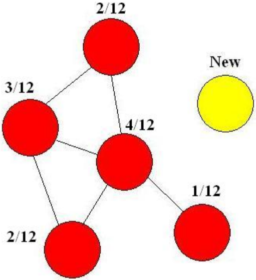
$$\mathbb{P}_{t+1}(z = u) = \frac{d_t(u)}{\sum_{w \in V(G_t)} d_t(w)} = \frac{d_t(u)}{2e_t} \quad \text{for } u \in V(G_t). \tag{1}$$



Facebook model

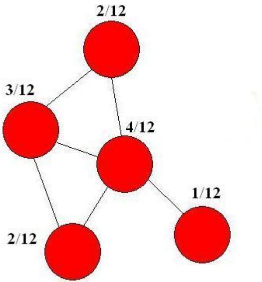
- Birth of vertices: With probability $p_1 > 0$, a new vertex with one edge is added to the graph. The edge is connected to an existing vertex z chosen according to the following probability distribution:

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



Facebook model

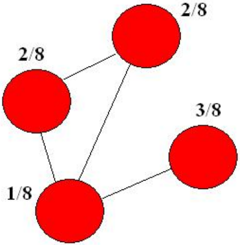
- Birth of edges: With probability $p_2 > 0$, a new edge is added between a vertex chosen by (1) and another vertex chosen randomly among $V(G_t)$.



Facebook model

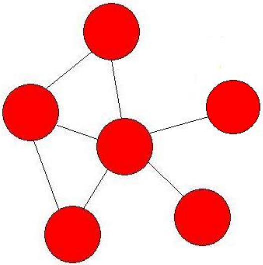
- Death of vertices: With probability $p_3 > 0$, an existing vertex z is chosen for deletion along with all the edges incident to z in G_t. The vertex z is chosen by the probability distribution:

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



Facebook model

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



Number of vertices of Facebook network

- The probability of birth of vertices is p_1.
- The probability of birth of edges is p_2.
- The probability of death of vertices is p_3.
- The probability of death of edges is p_4.

Proposition 1

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

Number of vertices of Facebook network

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

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

Number of vertices of Facebook network

Proof. For $t > 0$, $v_{t+1} = v_t + X_{t+1}$, 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 conditional expectation of v_{t+1} with G_t fixed is

$$\mathbb{E}[v_{t+1} \mid G_t] = v_t + \mathbb{E}[X_{t+1}]. \tag{2}$$

Number of vertices of Facebook network

By taking the expectations of both sides of (2), we get

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

Then we solve the recursive equation (3) with initial condition $\mathbb{E}[v_1] = 2$, yields:

$$\mathbb{E}[v_t] = (p_1 - p_3)(t - 1) + 2,$$

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

Number of edges of Facebook network

Proposition 2

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

Definition

Definition 1

$N^{(1)}_{k,t}$ is the number of neighbors of degree k of the vertex chosen for deletion during step t.

Definition 2

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

Degree distribution in the neighborhood of the deleted vertex

Proposition 3

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

The power-law of degree distribution

Theorem 1

In our model for Facebook network, if $p_1 > p_3 + 2p_4$, then it satisfies the small world phenomenon: the degree distribution obeys the power-law.

Degree distribution of the Facebook network

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

$$\begin{aligned} &\mathbb{E}[N_k(t+1) - N_k(t) \mid G_t] \\ &= p_1 C_k^{(1)}(t) + p_2 C_k^{(2)}(t) + p_3 C_k^{(3)}(t) + p_4 C_k^{(4)}(t) + p_1 \delta_{k1}, \end{aligned} \tag{4}$$

Degree distribution of the Facebook network

$$\mathbb{E}[N_k(t+1) - N_k(t) \mid G_t] = p_1 C_k^{(1)}(t) + p_2 C_k^{(2)}(t) + p_3 C_k^{(3)}(t) + p_4 C_k^{(4)}(t) + p_1 \delta_{k1},$$

where

Degree distribution of the Facebook network

$$\mathbb{E}[N_k(t+1) - N_k(t) \mid G_t] = p_1 C_k^{(1)}(t) + p_2 C_k^{(2)}(t) + p_3 C_k^{(3)}(t) + p_4 C_k^{(4)}(t) + p_1 \delta_{k1},$$

where

- $C_k^{(1)}(t) = N_{k-1}(t)\dfrac{(k-1)}{2e_t} - N_k(t)\dfrac{k}{2e_t},$

Degree distribution of the Facebook network

$$\mathbb{E}[N_k(t+1) - N_k(t) \mid G_t] = p_1 C_k^{(1)}(t) + p_2 C_k^{(2)}(t) + p_3 C_k^{(3)}(t) + p_4 C_k^{(4)}(t) + p_1 \delta_{k1},$$

where

- $C_k^{(1)}(t) = N_{k-1}(t)\dfrac{(k-1)}{2e_t} - N_k(t)\dfrac{k}{2e_t},$
- $C_k^{(2)}(t) = N_{k-1}(t)\dfrac{(k-1)}{2e_t} + \dfrac{N_{k-1}(t)}{v_t - 1} - N_k(t)\dfrac{k}{2e_t} - \dfrac{N_k(t)}{v_t - 1},$
- $C_k^{(3)}(t) = (k+1)N_{k+1}(t)\dfrac{v_t - 2\bar{d}_t}{v_t^2 - 2e_t} - N_k(t)\dfrac{(v_t - k)}{v_t^2 - 2e_t} - kN_k(t)\dfrac{v_t - 2\bar{d}_t}{v_t^2 - 2e_t},$
- $C_k^{(4)}(t) = N_{k+1}(t)\dfrac{(k+1)}{e_t} - N_k(t)\dfrac{k}{e_t}.$
- $p_1\delta_{k1}$ reveals the fact that the degree of a new vertex is always one.

Degree distribution of the Facebook network

Assume that $\mathbb{E}[N_k(t)]/t$ converges to a_k as $t \to \infty$.

Degree distribution of the Facebook network

Assume that $\mathbb{E}[N_k(t)]/t$ converges to a_k as $t \to \infty$.

To obtain a recursion for a_k, we take the expectation of (4) and find the limit as $t \to \infty$. By above Proposition 1, 2, 3, this yields

$$
\begin{aligned}
&[\alpha_2(k+2)+\beta_2]a_{k+2} + [\alpha_1(k+1)+\beta_1]a_{k+1} + [\alpha_0 k+\beta_0]a_k \\
&= 2p_1(p_1-p_3)(p_1+p_2-p_4)\delta_{k1}. \qquad (5)
\end{aligned}
$$

Degree distribution of the Facebook network

To solve (5), we will use Laplace's method as described in [4] to solve the homogeneous equation:

$$[\alpha_2(k+2)+\beta_2]a_{k+2}+[\alpha_1(k+1)+\beta_1]a_{k+1}+[\alpha_0 k+\beta_0]a_k=0 \quad \text{for } k\geq 1. \quad (6)$$

Degree distribution of the Facebook network

To solve (5), we will use Laplace's method as described in [4] to solve the homogeneous equation:

$$[\alpha_2(k+2)+\beta_2]a_{k+2}+[\alpha_1(k+1)+\beta_1]a_{k+1}+[\alpha_0 k+\beta_0]a_k=0 \quad \text{for } k\geq 1. \qquad (6)$$

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

$$a_k=\int_a^b t^{k-1}h(t)\,dt \qquad (7)$$

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

Degree distribution of the Facebook network

Then $[\alpha_2(k+2)+\beta_2]a_{k+2}+[\alpha_1(k+1)+\beta_1]a_{k+1}+[\alpha_0 k+\beta_0]a_k=0$ is satisfied if a, b and $h(t)$ are chosen such that

$$[t^k h(t)(1-t)(\alpha_0-\alpha_2 t)]_a^b=0 \tag{8}$$

and

$$\frac{h'(t)}{h(t)}=\frac{\beta_1 t+\beta_0}{t(1-t)(\alpha_0-\alpha_2 t)}. \tag{9}$$

Degree distribution of the Facebook network

By integrating both sides of equation (9), we get

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

Degree distribution of the Facebook network

By integrating both sides of equation (9), we get

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

where

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

The power-law of degree distribution

Assume that $p_1 > p_3 + 2p_4$, then

$$\frac{\alpha_0}{\alpha_2} = \frac{(p_1+p_2)(p_1+p_3)}{2p_1(p_3+p_4)+2p_2p_3} > 1, \quad \lambda_2 = \frac{2p_1(p_1+p_2-p_4)}{(p_1+p_2)(p_1-p_3)-2p_1p_4} > -1$$

and $[\alpha_2 t^{k+\lambda_1}(1-t)^{\lambda_2+1}(\frac{\alpha_0}{\alpha_2}-t)^{\lambda_3+1}]_a^b = 0$ is satisfied with $a=0$ and $b=1$.

The power-law of degree distribution

Assume that $p_1 > p_3 + 2p_4$, then

$$\frac{\alpha_0}{\alpha_2} = \frac{(p_1+p_2)(p_1+p_3)}{2p_1(p_3+p_4)+2p_2p_3} > 1, \quad \lambda_2 = \frac{2p_1(p_1+p_2-p_4)}{(p_1+p_2)(p_1-p_3)-2p_1p_4} > -1$$

and $[\alpha_2 t^{k+\lambda_1}(1-t)^{\lambda_2+1}(\frac{\alpha_0}{\alpha_2}-t)^{\lambda_3+1}]_a^b = 0$ is satisfied with $a=0$ and $b=1$.

Hence we obtain

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

The power-law of degree distribution

$$\begin{aligned}
a_k &= \int_0^1 t^{k-1} t^{\lambda_1} (1-t)^{\lambda 2} (\frac{\alpha_0}{\alpha_2} - t)^{\lambda 3}\, dt \\
&\asymp \int_0^1 t^{k+\lambda_1-1} (1-t)^{\lambda_2}\, dt \\
&= \frac{\Gamma(k+\lambda_1)\Gamma(1+\lambda_2)}{\Gamma(k+1+\lambda_1+\lambda_2)} \qquad (10) \\
&\sim k^{-(1+\lambda_2)}, \quad k \geq 1.
\end{aligned}$$

Here (10) is obtained by a formula in [6] (Table 1 (5) p.27) ■

The power-law of degree distribution

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

$$a_k \sim k^{-5}, \quad \text{for } k \geq 1. \qquad \blacksquare$$

Section 3

Conclusion

Conclusion

- Our model for Facebook generates graphs with asymptotically power-law degree distribution.

Conclusion

- Our model for Facebook generates graphs with asymptotically power-law degree distribution.
- Future work:

Conclusion

- Our model for Facebook generates graphs with asymptotically power-law degree distribution.
- Future work:
  - Finding the average distance of the random graph model theoretically.

Conclusion

- Our model for Facebook generates graphs with asymptotically power-law degree distribution.
- Future work:
  - Finding the average distance of the random graph model theoretically.
  - Constructing a general model for various networks.

Reference

- A. L, Barabási and R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509-512.
- L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. CoRR, abs/1111.4570, 2011.
- J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the Facebook social graph. CoRR, abs/1111.4503, 2011.
- C. Jordan, Calculus of Finite Differences, Rottig and Romwalter, 1939.
- N. Deo and A. Cami, Preferential deletion in dynamic models of web-like networks, Inf. Proc. Lett. 102 (2007) 156-162.
- B. David, Nouvelles tables d'intégrales définies, Leide : P. Engels, 1867.

Reference

- A. Cami, Techniques for analyzing dynamic random graph models of web-like networks: An overview, Networks 51 (2008) 211-255.
- M. Deijfen and M. Lindholm, Growing networks with preferential deletion and addition of edges, Physica A (2009) 4297-4303.
- R. Durrett, Probability: theory and examples, Wadsworth and Brooks/Cole, 2003.
- S. Dorogovtsev and J. Mendes, Scaling behaviour of developing and decaying networks, Europhys. Lett. 52 (1) (2000) 33-39.
- P. Erdös and A. Rényi, On random graphs, Publ. Math. 6 (1959) 290-297.
- C. Cooper, A. Frieze and J. Vera, Random deletion in a scale-free random graph process, Internet Math. 1 (4) (2004) 463-483.

Reference

- S. Milgram, The small world problem, Psychology Today 2 (1967) 60-67.
- F. Chung and L. Lu, Coupling online and offline analyses for random power law graphs, Internet Math. 1 (4) (2004) 409-461.
- D. J. Watts and S. H. Strogatz, Collective dynamics of ’small-world’ networks, Nature 393 (1998) 440-442.
- D. B. West, Introduction to graph theory, Pearson Education Taiwan Ltd. 2008.

Thank you for your attention!

祝 張鎮華 老師 60大壽 生日快樂!