Coupling On-line and Off-line Random Graphs

Woojin Kim

March 1st

Goal for this presentation

We are going to explore

- Several **random graph models**

Goal for this presentation

We are going to explore

- Several **random graph models**
- The method to analyze them

Goal for this presentation

We are going to explore

- Several **random graph models**
- The method to analyze them
  (Especially, by relating one random graph to another random graph)

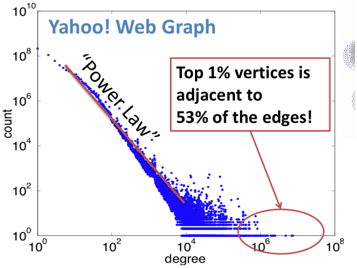
Reminder

- A graph is called a **power law graph** if the fraction of vertices with degree k is proportional to $\frac{1}{k^{\beta}}$ for some $\beta > 0$

Reminder

- A graph is called a **power law graph** if the fraction of vertices with degree k is proportional to $\frac{1}{k^{\beta}}$ for some $\beta > 0$

  - Natural Graphs:



Reminder

- A **random graph** means a probability space $(\Omega, \mathcal{F}, P)$ where the set Ω consists of graphs

Reminder

- A **random graph** means a probability space $(\Omega, \mathcal{F}, P)$ where the set Ω consists of graphs

e.g. Erdos-Renyi model $G(n, p)$

Reminder

- A **random graph** means a probability space $(\Omega, \mathcal{F}, P)$ where the set Ω consists of graphs

e.g. Erdos-Renyi model $G(n, p)$

e.g. $F(n, m)$

Reminder

- A random graph G almost surely **satisfies a property** P if

$$Pr(G \text{ satisfies } P) = 1 - o_n(1)$$

Reminder

- A random graph G almost surely **satisfies a property** P if

$$Pr(G \text{ satisfies } P) = 1 - o_n(1)$$

e.g. $G(n, n^{-1.1})$ is almost surely triangle free.

Reminder

- A random graph G almost surely **satisfies a property** P if

$$Pr(G \text{ satisfies } P) = 1 - o_n(1)$$

e.g. $G(n, n^{-1.1})$ is almost surely triangle free.

e.g. $G(n, n^{-0.9})$ almost surely contains triangle.

Off-line vs On-line

All random graph models for power law graphs belong to the following two categories; the off-line model and the on-line model

Off-line vs On-line

All random graph models for power law graphs belong to the following two categories; the off-line model and the on-line model

For the off-line model, the graph under consideration has a fixed number of vertices, say n vertices.

Off-line vs On-line

All random graph models for power law graphs belong to the following two categories; the off-line model and the on-line model

For the off-line model, the graph under consideration has a fixed number of vertices, say n vertices.

e.g. The uniform distribution on the set of all graphs on n vertices

Erdos-Renyi model $G(n, p)$

Off-line vs On-line

All random graph models for power law graphs belong to the following two categories; the off-line model and the on-line model

For the off-line model, the graph under consideration has a fixed number of vertices, say n vertices.

e.g. The uniform distribution on the set of all graphs on n vertices

Erdos-Renyi model $G(n, p)$

The probability distribution of the random graph depends upon the choice of the model.

Off-line vs On-line

The on-line model is often called the generative model.

Off-line vs On-line

The on-line model is often called the generative model.

At each tick of the clock, a decision is made for adding or deleting vertices or edges.

Off-line vs On-line

The on-line model is often called the generative model.

At each tick of the clock, a decision is made for adding or deleting vertices or edges.

The on-line model can be viewed as an infinite sequence of off-line models where the random graph model at time t may depend on all the earlier decisions.

Off-line vs On-line

The on-line models are **harder to analyze** than the off-line models, but **closer** to the way that realistic networks are generated.

Off-line vs On-line

The on-line models are **harder to analyze** than the off-line models, but **closer** to the way that realistic networks are generated.

We analyze the on-line models using the knowledge that we have about the off-line models.

Off-line vs On-line

The on-line models are **harder to analyze** than the off-line models, but **closer** to the way that realistic networks are generated.

We analyze the on-line models using the knowledge that we have about the off-line models.

Our goal is **to couple the on-line model with the off-line model of random graphs with a similar power law degree distribution** so that we can apply the techniques from the off-line model to the on-line model.

Graph property

A *graph property* P can be viewed as a set of graphs.

Graph property

A *graph property* P can be viewed as a set of graphs.

We say a graph G satisfies property P if $G \in P$.

Graph property

A *graph property* P can be viewed as a set of graphs.

We say a graph G satisfies property P if $G \in P$.

A graph property is said *monotone* if whenever a graph H satisfies property A, then any graph containing H must also satisfy property A.

Graph property

A *graph property* P can be viewed as a set of graphs.

We say a graph G satisfies property P if $G \in P$.

A graph property is said *monotone* if whenever a graph H satisfies property A, then any graph containing H must also satisfy property A.

Examples

- The property of containing the complete graph K_3

Graph property

A *graph property* P can be viewed as a set of graphs.

We say a graph G satisfies property P if $G \in P$.

A graph property is said *monotone* if whenever a graph H satisfies property A, then any graph containing H must also satisfy property A.

Examples

- The property of containing the complete graph K_3
- The property of being connected

Graph property

A *graph property* P can be viewed as a set of graphs.

We say a graph G satisfies property P if $G \in P$.

A graph property is said *monotone* if whenever a graph H satisfies property A, then any graph containing H must also satisfy property A.

Examples

- The property of containing the complete graph K_3
- The property of being connected (Non-example)

Dominance

Definition

Given two random graphs G_1 and G_2 on n vertices.

Dominance

Definition

Given two random graphs G_1 and G_2 on n vertices. We say G_1 *dominates* G_2, if

Dominance

Definition

Given two random graphs G_1 and G_2 on n vertices. We say G_1 *dominates* G_2, if
For any monotone graph property A,

$$Pr(G_1 \text{ satisfies } A) \geq Pr(G_2 \text{ satisfies } A).$$

Dominance

Definition

Given two random graphs G_1 and G_2 on n vertices. We say G_1 *dominates* G_2, if
For any monotone graph property A,

$$Pr(G_1 \text{ satisfies } A) \geq Pr(G_2 \text{ satisfies } A).$$

In this case, we write

$$G_1 \geq G_2.$$

Dominance

e.g. For any $p_1 \leq p_2$, $G(n, p_1) \leq G(n, p_2)$

Dominance

Definition

For any $\epsilon > 0$, we say G_1 dominates G_2 with an error estimate ϵ if

$$Pr(G_1 \text{ satisfies } A) + \epsilon \geq Pr(G_2 \text{ satisfies } A)$$

Dominance

Definition(Almost surely dominate)

If G_1 dominates G_2 with an error estimate $\epsilon = \epsilon_n$, which goes to zero as n approaches infinity,

Dominance

Definition(Almost surely dominate)

If G_1 dominates G_2 with an error estimate $\epsilon = \epsilon_n$, which goes to zero as n approaches infinity, We say G_1 almost surely dominates G_2. In this case, we write

$$\text{Almost surely } G_1 \succeq G_2$$

Dominance

Definition(Almost surely dominate)

If G_1 dominates G_2 with an error estimate $\epsilon = \epsilon_n$, which goes to zero as n approaches infinity, We say G_1 almost surely dominates G_2. In this case, we write

$$\text{Almost surely } G_1 \succeq G_2$$

e.g. For any $\delta > 0$, we have almost surely

$$G(n, (1-\delta)\frac{m}{\binom{n}{2}}) \preceq F(n, m) \preceq G(n, (1+\delta)\frac{m}{\binom{n}{2}})$$

Edge-independent

Definition

A random graphs G is called **edge-independent** if there is an edge-weighted function $p : E(K_n) \to [0, 1]$ satisfying

$$Pr(G = H) = \prod_{e \in H} p_e \times \prod_{e \notin H} (1 - p_e)$$

Edge-independent

Definition

A random graphs G is called **edge-independent** if there is an edge-weighted function $p : E(K_n) \to [0,1]$ satisfying

$$Pr(G = H) = \prod_{e \in H} p_e \times \prod_{e \notin H} (1 - p_e)$$

For any given random graph model, **it would be advantageous if we can establish some comparisons with edge-independent random graph**

A Growth-Deletion Model for Random Power Law Graphs

Here we consider a general on-line model that combines deletion steps with the preferential attachment model.

A Growth-Deletion Model for Random Power Law Graphs

Here we consider a general on-line model that combines deletion steps with the preferential attachment model.

Vertex-growth step: Add a new vertex v and form a new edge from v to an existing vertex u chosen with probability proportional to an existing vertex u chosen with probability proportional to d_u

A Growth-Deletion Model for Random Power Law Graphs

Here we consider a general on-line model that combines deletion steps with the preferential attachment model.

Vertex-growth step: Add a new vertex v and form a new edge from v to an existing vertex u chosen with probability proportional to an existing vertex u chosen with probability proportional to d_u

Edge-growth step: Add a new edge with endpoints to be chosen among existing vertices with probability proportional to the degrees. If it already exists in the current graph, the generated edge is discarded. The edge-growth step is repeated until a new edge is successfully added.

A Growth-Deletion Model for Random Power Law Graphs

Vertex-deletion step: Delete a vertex and all incident edges randomly.

A Growth-Deletion Model for Random Power Law Graphs

Vertex-deletion step: Delete a vertex and all incident edges randomly.

Edge-deletion step: Delete an edge randomly.

A Growth-Deletion Model for Random Power Law Graphs

For non-negative values p_1, p_2, p_3, p_4 summing to 1, we consider the following growth-deletion model $G(p_1, p_2, p_3, p_4)$:

A Growth-Deletion Model for Random Power Law Graphs

For non-negative values p_1, p_2, p_3, p_4 summing to 1, we consider the following growth-deletion model $G(p_1, p_2, p_3, p_4)$:

At each step, with probability p_1, take a vertex-growth step;

A Growth-Deletion Model for Random Power Law Graphs

For non-negative values p_1, p_2, p_3, p_4 summing to 1, we consider the following growth-deletion model $G(p_1, p_2, p_3, p_4)$:

At each step, with probability p_1, take a vertex-growth step;

With probability p_2, take an edge-growth step;

A Growth-Deletion Model for Random Power Law Graphs

For non-negative values p_1, p_2, p_3, p_4 summing to 1, we consider the following growth-deletion model $G(p_1, p_2, p_3, p_4)$:

At each step, with probability p_1, take a vertex-growth step;

With probability p_2, take an edge-growth step;

With probability p_3, take a vertex-deletion step;

A Growth-Deletion Model for Random Power Law Graphs

For non-negative values p_1, p_2, p_3, p_4 summing to 1, we consider the following growth-deletion model $G(p_1, p_2, p_3, p_4)$:

At each step, with probability p_1, take a vertex-growth step;

With probability p_2, take an edge-growth step;

With probability p_3, take a vertex-deletion step;

Otherwise, with probability $p_4 = 1 - p_1 - p_2 - p_3$, take an edge-deletion step.

A Growth-Deletion Model for Random Power Law Graphs

$G(p_1, p_2, p_3, p_4)$

A Growth-Deletion Model for Random Power Law Graphs

$G(p_1, p_2, p_3, p_4)$

- We assume $p_3 < p_1$ and $p_4 < p_2$ so that the number of vertices and edge grows as t goes to infinity.

A Growth-Deletion Model for Random Power Law Graphs

$G(p_1, p_2, p_3, p_4)$

- We assume $p_3 < p_1$ and $p_4 < p_2$ so that the number of vertices and edge grows as t goes to infinity.
- If $p_3 = p_4 = 0$, the model is the usual preferential attachment model which generates power law graphs with exponent $\beta = 2 + \frac{p_1}{p_1 + 2p_2}$.

A Growth-Deletion Model for Random Power Law Graphs

Facts:

$G(p_1, p_2, p_3, p_4)$

A Growth-Deletion Model for Random Power Law Graphs

Facts:

$G(p_1, p_2, p_3, p_4)$

Suppose $p_3 < p_1$ and $p_4 < p_2$.

A Growth-Deletion Model for Random Power Law Graphs

Facts:

$G(p_1, p_2, p_3, p_4)$

Suppose $p_3 < p_1$ and $p_4 < p_2$. Then almost surely the degree sequence of the growth-deletion model $G(p_1, p_2, p_3, p_4)$ follows the power law distribution with the exponent

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

A Growth-Deletion Model for Random Power Law Graphs

Facts:

$G(p_1, p_2, p_3, p_4)$

Suppose $p_3 < p_1$ and $p_4 < p_2$. Then almost surely the degree sequence of the growth-deletion model $G(p_1, p_2, p_3, p_4)$ follows the power law distribution with the exponent

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

A Growth-Deletion Model for Random Power Law Graphs

- A random graph in $G(p_1, p_2, p_3, p_4)$ almost surely has expected average degree $\frac{p_1 + p_2 - p_4}{p_1 + p_3}$.

A Growth-Deletion Model for Random Power Law Graphs

- A random graph in $G(p_1, p_2, p_3, p_4)$ almost surely has expected average degree $\frac{p_1 + p_2 - p_4}{p_1 + p_3}$.
- For p_i's in certain ranges, this value can be below 1 and the random graph is not connected.

A Growth-Deletion Model for Random Power Law Graphs

- A random graph in $G(p_1, p_2, p_3, p_4)$ almost surely has expected average degree $\frac{p_1 + p_2 - p_4}{p_1 + p_3}$.
- For p_i's in certain ranges, this value can be below 1 and the random graph is not connected.

$\Longrightarrow$ We consider the modified model $G(p_1, p_2, p_3, p_4, m)$ for some integer m which will generate random graphs which have expected degree $m\frac{(p_1 + p_2 - p_4)}{(p_1 + p_3)}$.

A Modified Growth-Deletion Model

$G(p_1, p_2, p_3, p_4, m)$:

A Modified Growth-Deletion Model

$G(p_1, p_2, p_3, p_4, m)$:

At each step, with probability p_1, add a new vertex and form m new edges from v to existing u chosen with probability proportional to d_u

A Modified Growth-Deletion Model

$G(p_1, p_2, p_3, p_4, m)$:

At each step, with probability p_1, add a new vertex and form m new edges from v to existing u chosen with probability proportional to d_u

With probability p_2, take m edge-growth steps;

A Modified Growth-Deletion Model

$G(p_1, p_2, p_3, p_4, m)$:

At each step, with probability p_1, add a new vertex and form m new edges from v to existing u chosen with probability proportional to d_u

With probability p_2, take m edge-growth steps;

With probability p_3, take a vertex-deletion step;

A Modified Growth-Deletion Model

$G(p_1, p_2, p_3, p_4, m)$:

At each step, with probability p_1, add a new vertex and form m new edges from v to existing u chosen with probability proportional to d_u

With probability p_2, take m edge-growth steps;

With probability p_3, take a vertex-deletion step;

Otherwise, with probability $p_4 = 1 - p_1 - p_2 - p_3$, take m edge-deletion step.

A Modified Growth-Deletion Model

Suppose $p_3 < p_1$ and $p_4 < p_2$.

A Modified Growth-Deletion Model

Suppose $p_3 < p_1$ and $p_4 < p_2$.

- Then almost surely the degree sequence of the growth-deletion model $G(p_1, p_2, p_3, p_4, m)$ follows the power law distribution with the exponent β being the same as the exponent for the model $G(p_1, p_2, p_3, p_4)$.
- Many results for $G(p_1, p_2, p_3, p_4, m)$ can be derived in the same fashion as for $G(p_1, p_2, p_3, p_4)$

Definition: Almost surely edge-independent

A random graph G is "almost surely edge-independent"

Definition: Almost surely edge-independent

A random graph G is "almost surely edge-independent" if there are two edge-independent random graphs G_1 and G_2 on the same vertex set satisfying:

$$G_1 \leq G \leq G_2$$

and

Definition: Almost surely edge-independent

A random graph G is "almost surely edge-independent" if there are two edge-independent random graphs G_1 and G_2 on the same vertex set satisfying:

$$G_1 \leq G \leq G_2$$

and

For any two vertices u and v, let $p_{uv}^{(i)}$ be the probability of edge uv in G_i for $i = 1, 2$. We have

$$p_{uv}^{(1)} = (1 - o(1))p_{uv}^{(2)}$$

The main theorem 1: Fan Chung and Linyuan Lu, 2004

Suppose $p_3 < p_1, \quad p_4 < p_2 \quad and \quad \log n \ll m < t^{\frac{p_1}{2(p_1+p_2)}}$. Then,

The main theorem 1: Fan Chung and Linyuan Lu, 2004

Suppose $p_3 < p_1, \quad p_4 < p_2 \quad and \quad \log n \ll m < t^{\frac{p_1}{2(p_1+p_2)}}$. Then,

(1) Almost surely the degree sequence of the growth-deletion model $G(p_1, p_2, p_3, p_4, m)$ follows the power law distribution with the exponent

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

The Main theorem 1: Fan Chung and Linyuan Lu, 2004

Suppose $p_3 < p_1, \quad p_4 < p_2 \quad and \quad \log n \ll m < t^{\frac{p_1}{2(p_1+p_2)}}$. Then,

(2) $G(p_1, p_2, p_3, p_4, m)$ is almost surely edge-independent. It dominates and is dominated by an edge-independent graph with probability $p_{ij}^{(t)}$ of having an edge between vertices i and j, $i < j$, at time t, satisfying:

The Main theorem 1: Fan Chung and Linyuan Lu, 2004

Suppose $p_3 < p_1, \quad p_4 < p_2 \quad and \quad \log n \ll m < t^{\frac{p_1}{2(p_1+p_2)}}$. Then,

(2) $G(p_1, p_2, p_3, p_4, m)$ is almost surely edge-independent. It dominates and is dominated by an edge-independent graph with probability $p_{ij}^{(t)}$ of having an edge between vertices i and j, $i < j$, at time t, satisfying:

$$p_{ij}^{(t)} \approx \begin{cases} \frac{p_2 m}{2p_4\tau(2p_2-p_4)} \frac{t^{2\alpha-1}}{i^\alpha j^\alpha}\left(1+\left(1-\frac{p_4}{p_2}\right)\left(\frac{j}{t}\right)^{\frac{1}{2r}+2\alpha-1}\right), & \text{if } \; i^\alpha j^\alpha \gg \frac{p_2 m t^{2\alpha-1}}{4\tau^2 p_4} \\ 1-(1+o(1))\frac{2p_4\tau}{p_2 m} i^\alpha j^\alpha t^{1-2\alpha}, & \text{if } \; i^\alpha j^\alpha \ll \frac{p_2 m t^{2\alpha-1}}{4\tau^2 p_4} \end{cases}$$

where $\alpha = \frac{p_1(p_1+2p_2-p_3-2p_4)}{2(p_1+p_2-p_4)(p_1-p_3)}$ and $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3}$

The Main theorem 2: Fan Chung and Linyuan Lu, 2004

Without the assumption on m, we have the following general but weaker result. We say the index of a vertex u is i if u is generated at time i.

The Main theorem 2: Fan Chung and Linyuan Lu, 2004

Without the assumption on m, we have the following general but weaker result. We say the index of a vertex u is i if u is generated at time i.

The Main theorem 2: Fan Chung and Linyuan Lu, 2004

In $G(p_1, p_2, p_3, p_4, m)$ with $p_3 < p_1, \quad p_4 < p_2$, let S be the set of vertices with index i satisfying

$$i \gg m^{\frac{1}{\alpha}} t^{1-\frac{1}{2\alpha}}.$$

Let G_S be the induced subgraph of $G(p_1, p_2, p_3, p_4, m)$ on S. Then we have

(1) G_S dominates a random power law graph G_1, in which the expected degrees are given by

$$w_i \approx \frac{p_2 m}{2p_4\tau(2p_2 - p_4)(\frac{p_1}{p_1-p_3} - \alpha)} \frac{t^\alpha}{i^\alpha}$$

The Main theorem 2: Fan Chung and Linyuan Lu, 2004

(2) G_S is dominated by a random power law graph G_2, in which the expected degrees are given by

$$w_i \approx \frac{m}{2p_4\tau(\frac{p_1}{p_1-p_3}-\alpha)}\frac{t^\alpha}{i^\alpha}.$$

The Main theorem 3: Fan Chung and Linyuan Lu, 2004

In $G(p_1, p_2, p_3, p_4, m)$ with $p_3 < p_1$, $p_4 < p_1$, let T be the set of vertices with index i satisfying

$$i \ll m^{\frac{1}{\alpha}} t^{1-\frac{1}{2\alpha}}.$$

Then the induced subgraph G_T of $G(p_1, p_2, p_3, p_4, m)$ is almost a complete graph. Namely, G_T dominates an edge-independent an edge-independent graph with $p_{ij} = 1 - o(1)$

Ingredient of Proof for the Main Theorems

- The basic idea : the martingale method
- But with substantial difference

- A martingale involves a sequence of functions with consecutive functions having small bounded differences, each function is defined on a fixed probability space Ω.

- For the on-line model, the probability space for the random graph generated at each time instance is different in general. (We have a sequence of probability spaces where two consecutive ones have "small" difference.)

Bibliography

Complex Graphs and Networks; Fan Chung and Linyuan Lu (2004)

The Probabilistic method 3rd ed; Noga Alon and Joel H. Spencer (2008)