修 士 論 文

時系列データに基づいた Scale Free Graph モデルに関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

森本 真一

2009 年 3 月

修 士 論 文

時系列データに基づいた Scale Free Graph モデルに関する研究

指導教官　上原 隆平 准教授

審査委員主査　上原 隆平 准教授
審査委員　浅野 哲夫 教授
審査委員　宮地 充子 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

0710069 森本 真一

提出年月: 2009 年 2 月

概 要

近年，WWW やインターネットをモデル化できるものとして Scale Free Graph が注目されている．中でも，より解析しやすい Scale Free Graph モデルとして，時系列データに基づく Scale Free Graph (Scale Free Interval Graph) が提案されている．本論文では，Blog データを用いて，より実ネットワークに即した時系列データに基づく Scale Free Graph モデルの提案をおこなった．そして提案したネットワークモデルの妥当性を実験的に解析した．

目 次

第1章　はじめに

1.1　背景と目的

近年，WWW やインターネットをモデル化できるものとして，Scale Free Graph や Small World が注目を集めている [1][2]．これは従来の Erdös-Renyi による一様な構造を持つ Random Graph とは違い，非一様な構造を持っており，様々な現実の社会ネットワークをモデル化していると考えられている．

Scale Free Graph ではベキ法則と呼ばれる法則が成立しており，このベキ法則を実現できるいくつかの Scale Free Graph モデルがすでに知られている [3]．しかしそれらのモデルは次数分布の評価に微分方程式が使われており，モデルから得られたグラフの解析は複雑でグラフの結合構造も簡単には見る事ができない．解析を容易にすることは，Scale Free Graph 上で動くアルゴリズムをデザインする際にも非常に重要であると考えられる．そのため，比較的単純な確率や組合わせによって解析することができるモデルとして，時系列データに基づいた Scale Free Graph である Scale Free Interval Graph が提案されている [4].

Scale Free Interval Graph は Interval Graph を用いてモデル化されている．Interval Graph は Intersection Graph の一種である．Interval Graph では一つの頂点を数直線上の区間で表現し，二つの区間に重なりがある場合，それらの間に辺を持つ．

Scale Free Interval Graph はこの区間を時系列データとみなし，生存確率に偏りを持たせた結果として得られる Scale Free Graph である．しかしこのモデルは理論的な確率モデルであって，実際のネットワークに沿わない特徴を持つ．Scale Free Interval Graph では同じ時間を共有する頂点がある場合，その頂点は全て隣接し，クリークになってしまう特徴が存在する．これは実際のネットワークでは考えにくい仮定である．

本論文では Excite 社から提供された Blog データを用いて，より実ネットワークに即した Scale Free Interval Graph モデルの提案を目的としている．

Max-Tolerance Graph は Interval Graph を一般化したものと考えられる．Max-Tolerance Graph は Interval Graph と同様に頂点を区間で表現するが，それぞれの区間は重みを持つ．二つの区間の重なりの長さが，それぞれの重みの大きい方よりも勝れば，それらの間に辺を持つ．本論文では Interval Graph の代わりに Max-Tolerance Graph を用いてモデル化を行う．Max-Tolerance Graph を用いることで，同じ時間を共有する頂点同士がクリークになってしまう特徴を回避し，より実ネットワークに即した Scale Free Interval Graph の提案が可能であると考えられる．

1.2 本論文の構成

本論文では，2 章を本論文で扱う基本的な用語の定義，3 章を提供された Blog データとネットワークの対応についての方法，4 章を新しいモデルの提案と結果とし，5 章をまとめとする．

第2章　準備

2.1　グラフ

本章では，グラフの基本的な定義，性質について説明する．

2.1.1　無向グラフ

無向グラフ$G = (V, E)$は，頂点集合Vと辺集合Eからなる．辺$e \in E$は，頂点$v_i, v_j \in V$の非順序対で，$\{v_i, v_j\}$と表される．頂点v_iとv_jの間に辺があるとき，v_iとv_jは隣接していると言い，v_iとの間に辺を持つ全ての頂点をv_iの隣接点と言う．図2.1において，頂点集合Vは$\{a, b, c, d, e, f\}$，辺集合Eは$\{\{a, b\}, \{b, a\}, \{b, c\}, \{b, d\}, \{b, e\}, \{c, b\}, \{c, d\}, \{d, b\}, \{d, c\}, \{d, e\}, \{d, f\}, \{e, b\}, \{e, d\}, \{e, f\}, \{f, d\}, \{f, e\}\}$となる．

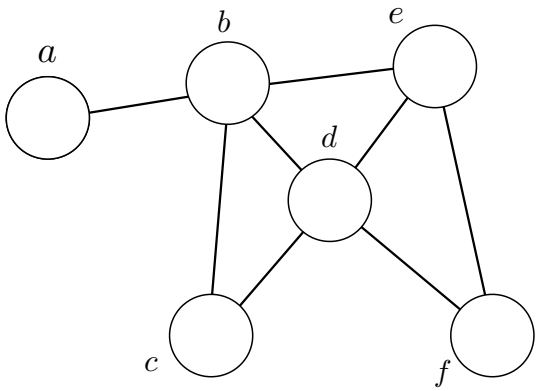


図 2.1: 無向グラフの例

2.1.2　有向グラフ

有向グラフ$G = (V, E)$は，頂点集合Vと辺集合Eからなる．辺$e \in E$は，頂点$v_i, v_j \in V$の順序対で，(v_i, v_j)と表される．頂点v_iからv_jへの辺があるとき，v_jはv_iに隣接していると言い，v_iからの辺を持つ全ての頂点をv_iの隣接点と言う．図2.2において，頂点集

合 V は $\{a, b, c, d, e, f\}$，辺集合 E は $\{(a, b), (b, c), (b, d), (b, e), (c, d), (d, e), (d, f), (e, f)\}$ となる．

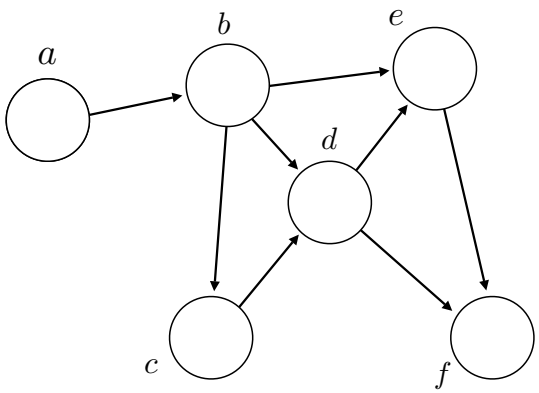


図 2.2: 有向グラフの例

2.1.3 次数

無向グラフにおいて，頂点 v と接続する辺の数を頂点 v の次数と呼び, $\deg(v)$ と表す．

有向グラフにおいて，頂点 v に入ってくる辺の数を頂点 v の入次数, 頂点 v から出て行く辺の数を頂点 v の出次数と呼び，それぞれ $in\_\deg(v), out\_\deg(v)$ と表す．また，有向グラフにおける次数は，$\deg(v) = in\_\deg(v) + out\_\deg(v)$ と定義する．図 2.3 において，頂点 v の次数は 4 となる．

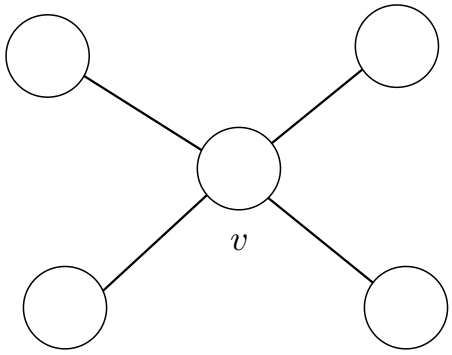


図 2.3: 次数の例

2.1.4 セルフループ

同じ頂点 v を結ぶ辺 $\{v, v\}$ や (v, v) をセルフループと呼ぶ.

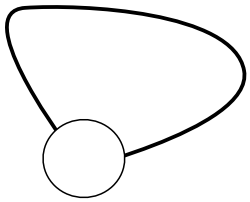


図 2.4: セルフループの例

2.1.5 多重辺

頂点 u から頂点 v への辺が複数存在するとき，それらを多重辺と呼び，それらの辺は平行であると言う.

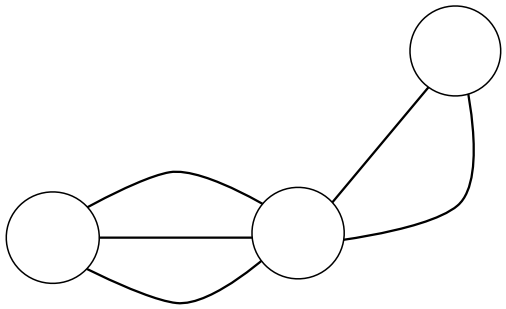


図 2.5: 多重辺の例

2.1.6 パス

無向グラフ $G = (V, E)$ において，$i = 1, 2, \ldots, l$ に対して，$\{v_{i-1}, v_i\} \in E$ であるとき，頂点の列 $[v_0, v_1, \ldots, v_l]$ は v_0 から v_l への長さ l のパスであるという．グラフが有向グラフで $i = 1, 2, \ldots, l$ に対して，$(v_{i-1}, v_i) \in E$ であるとき，この順列を有向パスと呼ぶ．

2.1.7 連結グラフと連結成分

無向グラフ $G = (V, E)$ に対して，V の任意の 2 つの頂点を結ぶパスが存在するとき，グラフ G は連結であるという．また，グラフ G の極大な連結部分グラフを G の連結成分と呼ぶ．

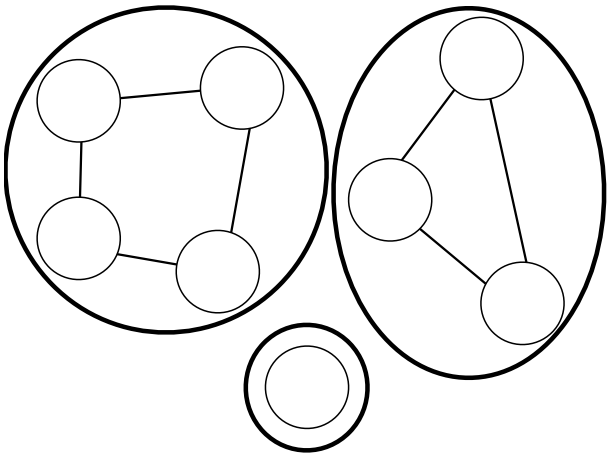


図 2.6: 連結成分の例

2.1.8 Interval Graph と区間表現

グラフ $G = (V, E)$ の区間表現とは，数直線上の区間の集合 I であり，以下の条件を満たすものである．

- V の各頂点は I の区間と 1 対 1 対応する
- G において頂点が隣接するための必要十分条件は，対応する区間同士が重なりを持つことである

区間表現を持つグラフを Interval Graph という．図 2.7 において，(a) は Interval Graph を表し，(b) は (a) の Interval Graph の区間表現を表している．

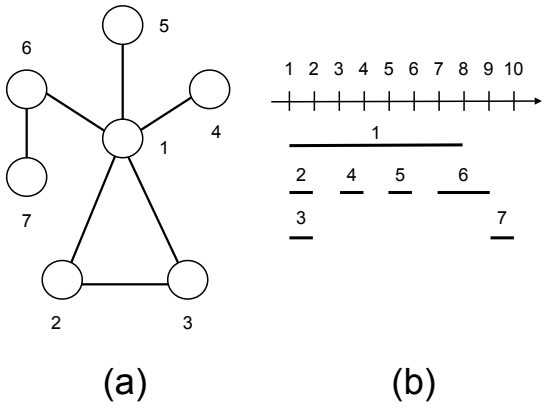


図 2.7: Interval Graph と対応する区間表現

2.1.9 Max-Tolerance Graph

Max-Tolerance Graph は Interval Graph を一般化したものと考えられる [5]．このグラフクラスは Interval Graph と同様に区間表現を持ち，またそれぞれの区間 I_i は重み t_i を持つ．Max-Tolerance Graph $G = (V, E)$ において，頂点が隣接するための必要十分条件は以下の通りである．

$$\{v_1, v_2\} \in E \Leftrightarrow |I_1 \cap I_2| \geq max(t_1, t_2) \tag{2.1}$$

ここで $t_i \leq |I_i|$ である．すべての重みを 0 とすれば，Interval Graph の定義と一致するので，Interval Graph は Max-Tolerance Graph である．図 2.8 において，(a) は Max-Tolerance Graph を表し，(b) は (a) の Max-Tolerance Graph の区間表現と重みを表している．なお，図 2.8(a) は Interval Graph ではない．

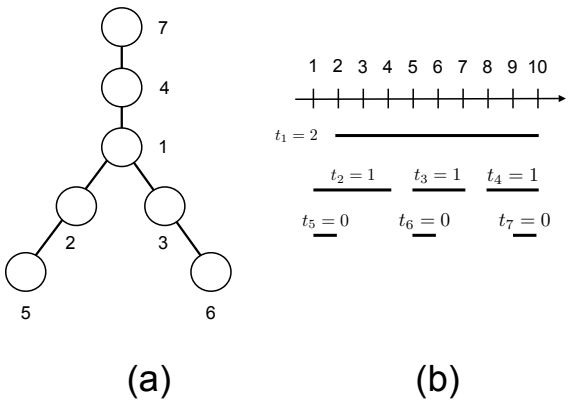


図 2.8: Max-Tolearance Graph と対応する区間表現

2.2 Scale Free Graph

2.2.1 Scale Free Graph

World Wide Web を始めとする多くの実在のネットワークは，その頂点の次数分布 $P(k)$ が，ある定数 γ に対する，べき法則 (power law) 分布

$$P(k) \sim k^{-\gamma} \tag{2.2}$$

に従うという共通の性質を持つ．次数 k はいくらでも大きなサイズをとりうることから，この性質を “scale free” と呼ぶ．実際に，WWW，インターネット，航空路線などのネットワークは，Scale Free Graph であることが報告されている．図 2.9 に Scale Free Graph の次数分布の例を示す．

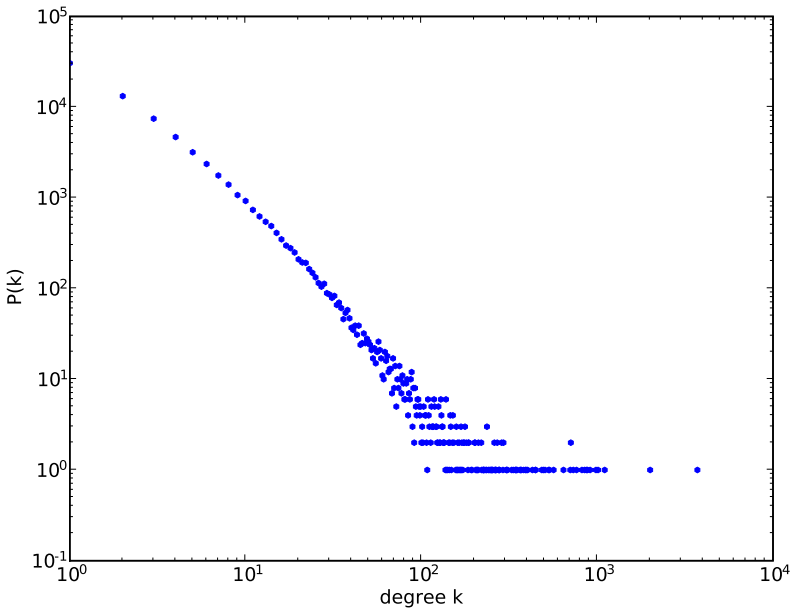


図 2.9: Scale Free Graph の次数分布

従来の Erdös-Renyi の Random Graph モデルではこれらの Scale Free Graph を再現することが難しかった．近年，この scale free を実現する Random Graph モデルが注目されるようになり，いくつかの異なる Scale Free Graph のモデルが提案されている．

2.2.2 Barabási-Albert の Preferential attachment モデル

Barabási と Albert による preferential attachment モデルは，いくつかの異なる Scale Free Graph モデルの中で，最も広く受け入れられている．

このグラフモデルは，グラフに新しい頂点を加えていき，新しい頂点は次数の大きい頂点との間に優先的に辺を張るということを繰り返し，グラフを生成していく．

Preferential attachment モデルは，Bollobás らによって次のように一般化されている．

- 各時刻 t で，頂点 v_t を加え，v_t とある頂点 u の間に 1 つの辺を張る．u は次の確率分布に従ってランダムに選択される．

$$Pr(u = v_i) = \begin{cases} \frac{d_{t-1}(v_i)}{2t-1} & v_i \neq v_t \text{のとき} \\ \frac{1}{2t-1} & v_i = v_t \text{のとき} \end{cases} \tag{2.3}$$

  ここで，$d_{t-1}(v)$ は時刻 $t-1$ における頂点 v の次数を表す．

- m は任意の定数で，時刻 $t \equiv 0 \pmod m$ のとき，$t, t-1, t-2, \cdots, t-m+1$ で加えられた m 頂点を 1 つの頂点にまとめる．

このグラフの次数分布は

$$P(k) = \frac{2m^3}{k^3} \tag{2.4}$$

となることが示されている [6]．また，このモデルは非常に簡潔なモデルではあるが，セルフループや多重辺が存在する可能性があることに注意しなければならない．

2.2.3 時系列データに基づいた Scale Free Graph (Scale Free Interval Graph) モデル

従来の Scale Free Graph モデルは次数分布の評価に微分方程式が使われており，得られたグラフの結合構造も簡単には見ることができない．そのため，比較的単純な確率や組合わせによって解析することができるモデルとして，時系列データに基づいた Scale Free Graph である Scale Free Interval Graph が提案されている [4]．

Scale Free Intreval Graph は次のような特徴を持つ．

- それぞれの頂点は時間軸上の閉区間
- 区間が重なりを持つ時は，頂点は辺で結ばれる
- それまでの生存区間が長い頂点は次の世代でも生存している確率が高い

また区間の長さと個数の分布は次のベキ法則に従う．

$$P(k) = \frac{1}{\zeta(\alpha)}(k+1)^{-\alpha} \tag{2.5}$$

ここで k は区間の長さ，$P(k)$ は k の長さを持つ区間の個数，$\zeta(\alpha) = \sum_{i=1}^{\infty} i^{-\alpha}$ は *Riemann* のゼータ関数である．図 2.10 は Scale Free Interval Graph の区間の長さと個数を両対数

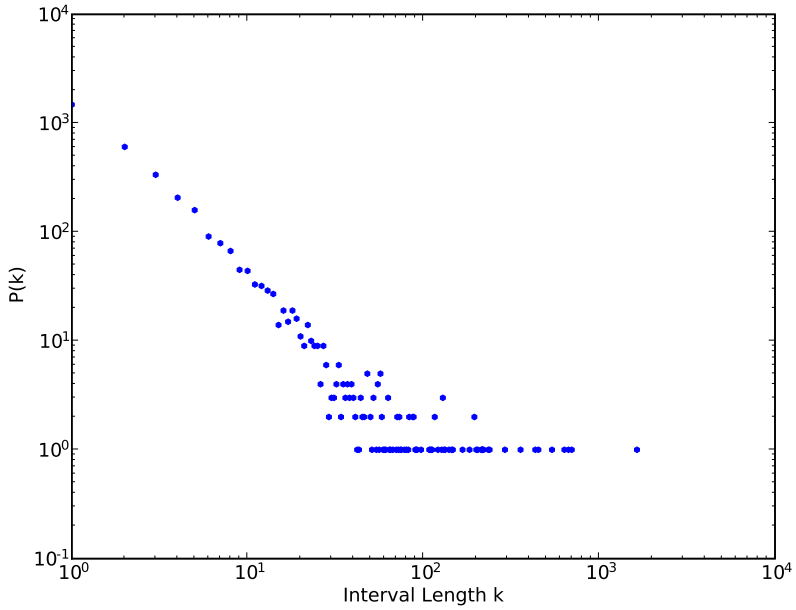


図 2.10: 区間の長さと個数

でプロットしたグラフである．図 2.10 において，区間の長さと個数の分布は直線になっており，ベキ法則に従っていることが分かる．

Scale Free Interval Graph におけるある区間 I_i の次数はその区間 I_i が誕生した時点で既に存在している区間の数 $N[I_i]$ とその区間の始点から終点までに誕生した区間の数 $N(I_i)$ とを足し合わせたものである (図 2.11)．

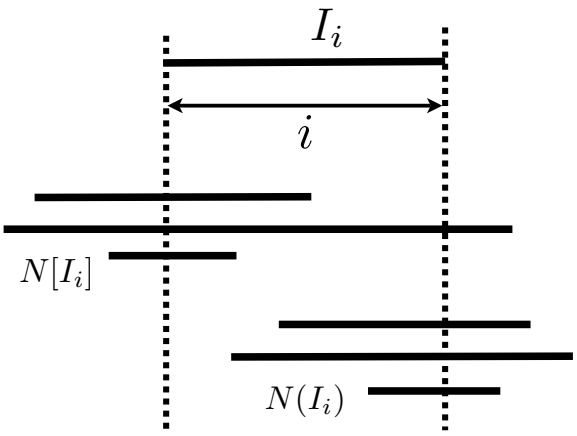


図 2.11: Scale Free Interval Graph における次数

$N[I_i]$ はポアソン分布に従い，$N(I_i)$ はベキ法則に従う．図 2.12 に $N[I_i]$ の分布を，図 2.13 に $N(I_i)$ の分布を示す

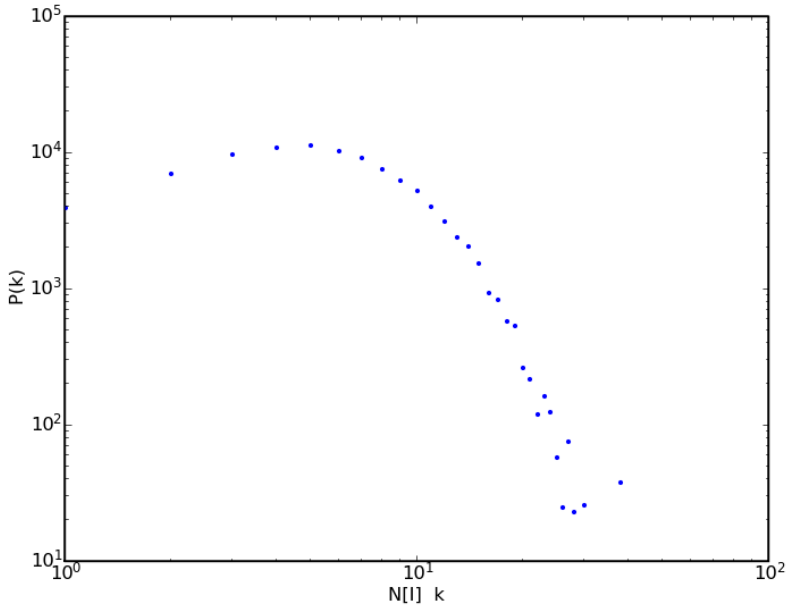


図 2.12: $N[I_i]$ の分布

.

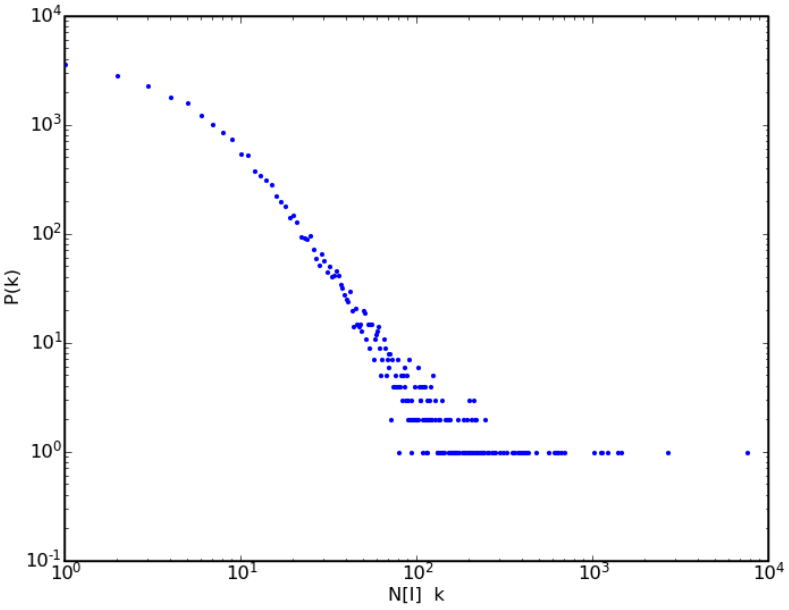


図 2.13: $N(I_i)$ の分布

Scale Free Interval Graph の次数については一般に以下のことがいえる．

- 区間が長いと次数が大きくなり，またその逆も成立する
- 区間が長いと $N(I_i)$ が支配項になる

したがって次数が大きくなるにつれ，$N(I_i)$ が $N[I_i]$ を支配するので，Scale Free Interval Graph の次数分布は次数が小さいところではポアソン分布，次数が大きくなるとベキ法則に従う．図 2.14 に Scale Free Interval Graph の次数分布を示す．

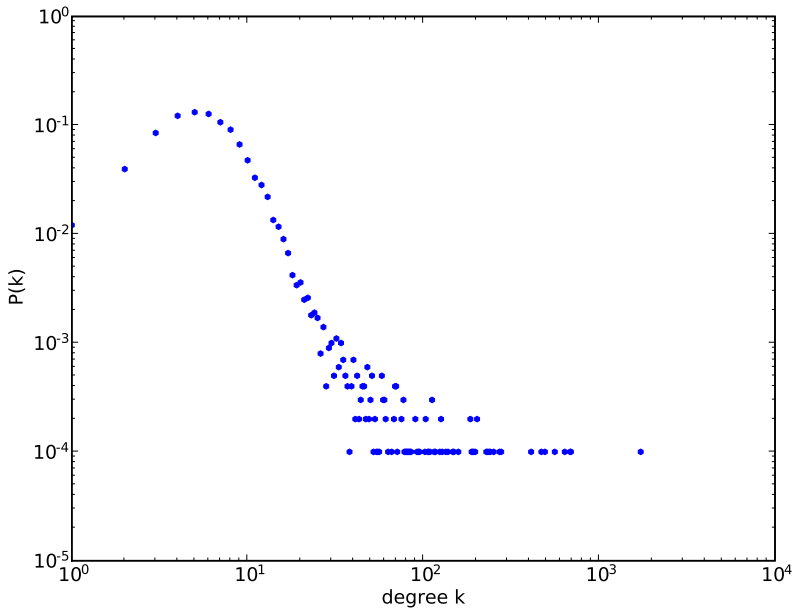


図 2.14: Scale Free Graph の次数分布

Scale Free Interval Graph が実ネットワークに沿わない特徴として，同じ時間を共有する頂点同士は全て隣接し，クリークになってしまう点があげられる．

第3章　Blog データ

3.1　Blogデータの利用目的

Scale Free Interval Graph モデルでは同じ時間を共有する頂点がある場合，その頂点は全て隣接し，クリークになってしまう特徴が存在する．これは実際のネットワークでは考えにくい仮定といえる．そのため本論文では Excite 社から提供されたブログデータ (2004 年 2 月 1 日〜2007 年 10 月 30 日) を解析し，実ネットワークにおいて隣接する傾向を検証することを目的としている．本章では，利用する Blog データのネットワークの設定と区間の対応について説明する．また Blog データの区間の分布，Scale Free 性について，Scale Free Interval Graph のそれと比較し，考察する．

3.2　Blog データの取り扱い

3.2.1　Blog ネットワークの設定

Excite Blog では通常，１つの Blog につきユニークな ID が与えられており．また通常のリンクのほかに他人の Blog の記事に自分の Blog へのリンクを作成する TrackBack 機能も実装されている (図 3.1).

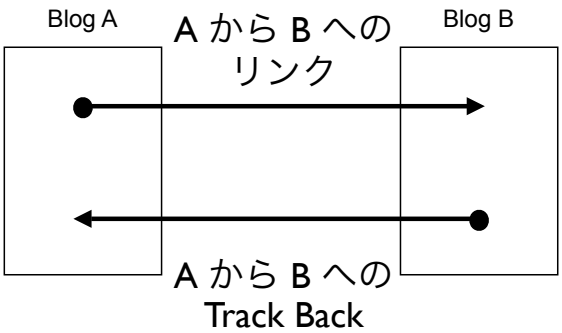


図 3.1: リンクと TrackBack の違い

本論文では Blog データを以下のようにしてネットワークとして定義する (図 3.2).

- Blog におけるユニークな ID を頂点とする
- Excite Blog 内におけるリンク及び TrackBack だけを有効とし，Excite Blog 外へのリンク及び TrackBack は扱わない
- リンク及び TrackBack を無向辺とみなす

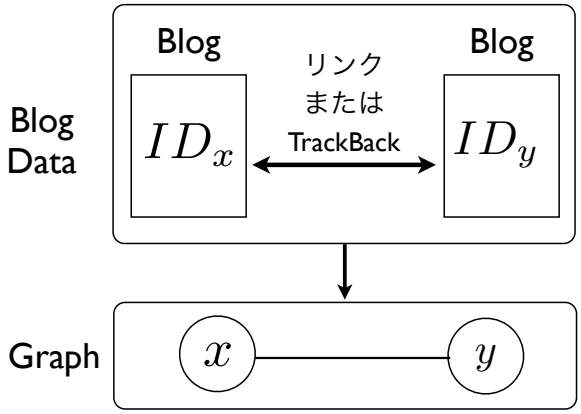


図 3.2: Blog データのネットワーク設定

以上のようにして構成した Blog ネットワークは多数の連結成分を含んでいるため，本論文では最大の連結成分のみを扱う．以下に連結成分数と全体の頂点数，辺数，今回扱うネットワークの頂点数と辺数を示す．

- 連結成分数　：5348
- 全体の頂点数：87779
- 全体の辺数　：416222
- 今回扱うネットワークの頂点数：73863　(全体の 84%)
- 今回扱うネットワークの辺数　：198567 (全体の 48%)

最大の連結成分における頂点数が全体の 84%, 辺数が全体の 48% と大きな差が見られる．これは多数の小さな連結成分においては，リンクファームすなわち人工的なクリークが形成されているためだと考えられる．

3.2.2　Blog と区間の対応

Blog データと区間の対応については，区間の始点はその Blog の最初の更新日，終点はその Blog の最後の更新日とする．また Blog の更新回数を対応する区間の重みとする．

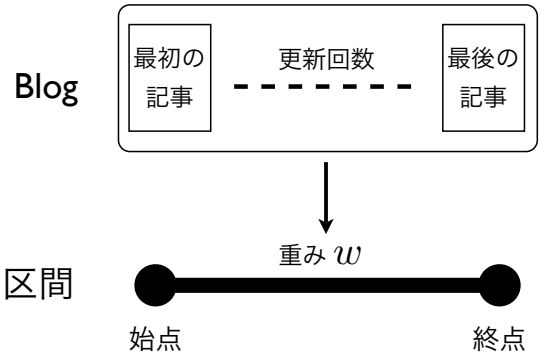
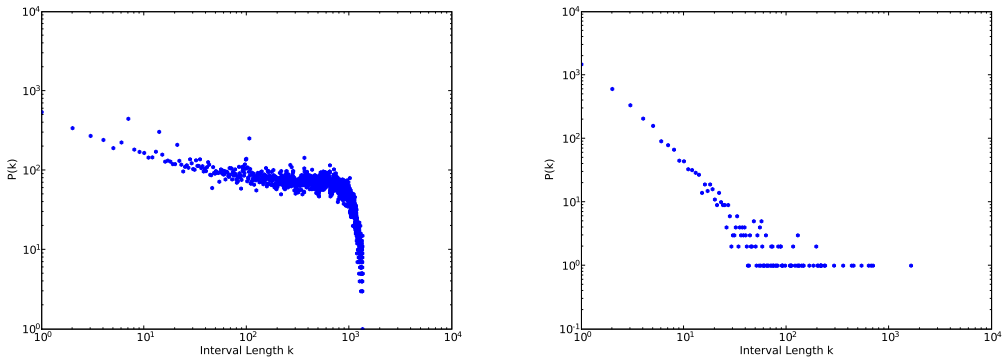


図 3.3: Blog データと区間の対応

3.3 Blog ネットワークの解析

3.3.1 区間解析

Scale Free Interval Graph では区間の分布（区間の長さとその長さを持つ頂点の個数の分布）はベキ法則に従った．本節では，Blog ネットワークにおいても区間の分布がベキ法則に従うか解析した．図 3.4(a) は x 軸に区間の長さ，y 軸にその長さを持つ頂点の個数をプロットしたグラフである．



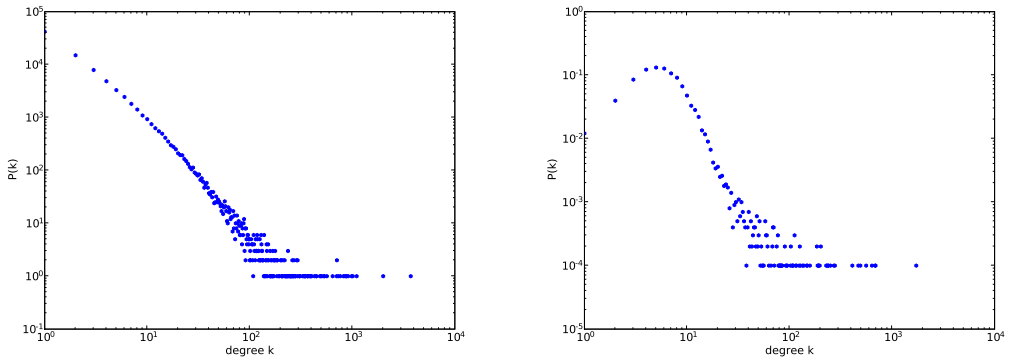
(a) Blog ネットワークの区間分布　　(b) Scale Free Interval Graph の区間分布

図 3.4: 区間分布

図 3.4(a) を見ると，Scale Free Interval Graph の区間分布をプロットした図 3.4(b) と比べると傾きが小さいものの，ベキ法則に従っていることが見てとれる．右端を見ると頂点の個数が急速に落ちているが，これは元のデータの区間が途中で切れているためだと考えられる．

3.3.2 Scale Free 性の解析

一般に WWW の次数分布はベキ法則に従うことが知られている．そのため，今回利用する Blog ネットワークの次数分布がベキ法則に従うかどうかを解析する．図 3.5(a) は Blog ネットワークの次数分布をプロットしたグラフである．



(a) Blog ネットワークの次数分布　　(b) Scale Free Interval Graph の次数分布

図 3.5: 次数分布

図 3.5(a) のグラフを見ると，両対数グラフで直線に乗っているので，ベキ法則に従っていることが見てとれる．また Scale Free Interval Graph の次数分布をプロットした図 3.5(b) のグラフとは違い，次数が小さい場所でもベキ法則に従っていることが分かる．

第4章　Blog データの解析と新しいモデルの提案

4.1　決定的モデル

4.1.1　手法

Scale Free Interval Graph は Interval Graph を用いてモデル化されている．そのため，同じ時間を共有する頂点同士がクリークになってしまう特徴が存在する．本節では，この特徴を回避するために Interval Graph の代わりに Max-Tolerance Graph を用いてモデル化する．今回のモデル化は区間の長さと更新回数の間の相関を用いる．

Max-Tolerance Graph を用いたモデル化は以下の手順で行う．

- 区間の長さと更新回数の間の相関を実データから求める．
- 区間の長さに応じて，上で求めた相関から更新回数を与える (図 4.1)．
- 頂点が隣接するための必要十分条件は以下のとおりである (図 4.2)．

$$\{v_1, v_2\} \in E \Leftrightarrow (c \times |I_1 \cap I_2|) \geq max(w_1, w_2) \tag{4.1}$$

ここで，w_i は I_i の重み（更新回数），c は調整のための係数である．

3 章にて設定した Blog ネットワークの頂点と対応する区間に対して，このモデルを適用し，得られたグラフの次数分布を調べ，このモデルの有効性を考察する．

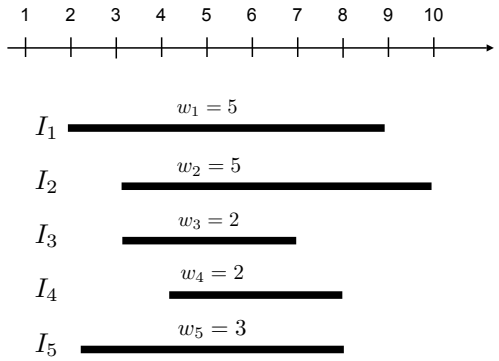


図 4.1: 実データから得られた相関から区間の長さに応じて，更新回数を与える．したがって区間の長さが同じであれば，更新回数も同じになる．

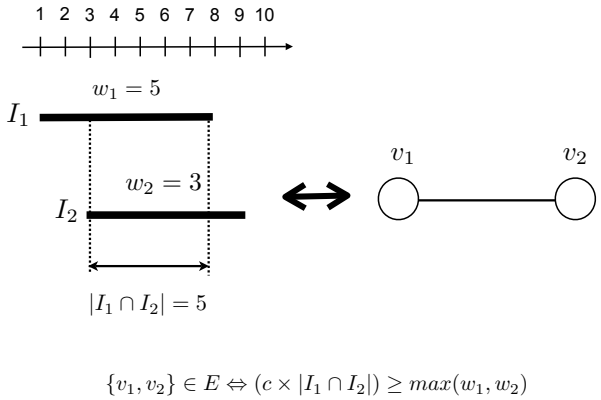


図 4.2: 隣接する条件

4.1.2 結果

図 4.3 は x 軸に Blog ネットワークにおける区間の長さ，y 軸に更新回数をプロットしたものである (図中の赤色の直線は最小 2 乗法によって得られた回帰直線).

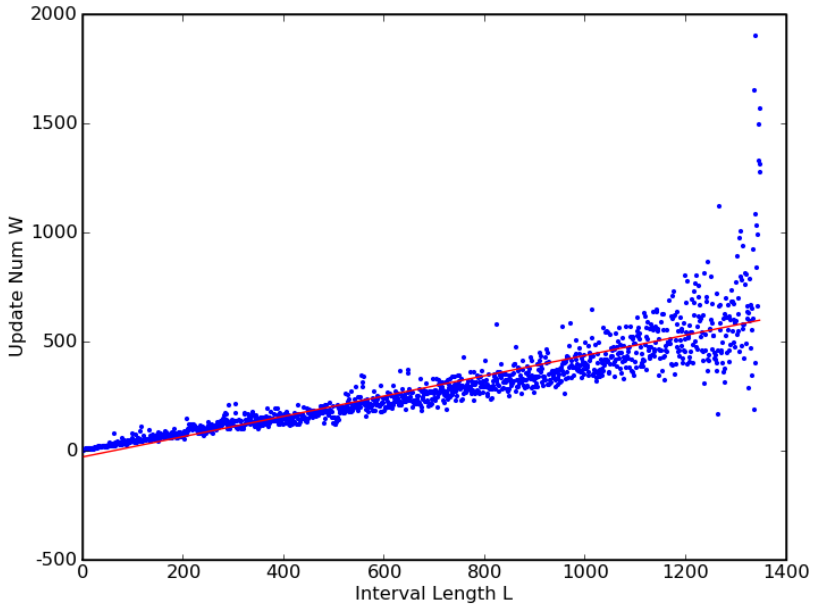


図 4.3: 隣接する条件

最小 2 乗法によって得られたブログネットワークにおける区間の長さと更新回数の相関は以下の通りである．ただし以下の式で得られる更新回数が 1 未満の時，更新回数は 1 とする (どの区間も少なくとも 1 回は更新されているため).

$$\text{更新回数} = 0.465088147012 \times \text{区間の長さ} - 28.7058825467 \tag{4.2}$$

上記の式を用いて，Blog ネットワークの区間に対して重みを与える．以下にモデルより得られたグラフの次数分布を示す．図 4.4 は $c = 0.5$，図 4.5 は $c = 1$，図 4.6 は $c = 1.5$，図 4.7 は $c = 2$ の係数を与えた結果である．

上記の式より更新回数は区間の長さの 1/2 程度になる．そのため式 (4.1) より，係数を 0.5 よりも小さくするとモデルより得られるネットワークは過度に疎になってしまう．本節では，Scale Free Interval Graph における同じ時間を共有する頂点同士がクリークになってしまう特徴を回避するために Max-Tolerance Graph をモデル化に用いている．これは，隣接する条件が区間に重なりを持つだけではなく，その区間の重なりがある程度大きくなければならない事を意味している．そのため，本節では区間の重なりの長さが長い方の区間の長さの少なくとも 1/4 以上ある時に，頂点間に辺が張られるように係数を調整する．長い方の区間の長さを l とした時，更新回数 w は区間の長さの 1/2 程度なので $l/2$

となる．また，区間の重なりの長さが長い方の区間の長さの 1/4 とすると．区間の重なりの長さは $l/4$ と表せる，この時，式 (4.1) の右辺は以下のように表せる．

$$(c \times l/4) \geq l/2 \tag{4.3}$$

上記の不等式は $c \geq 2$ で常に成り立つ．よって，本節では係数の最大値として 2 を与える．

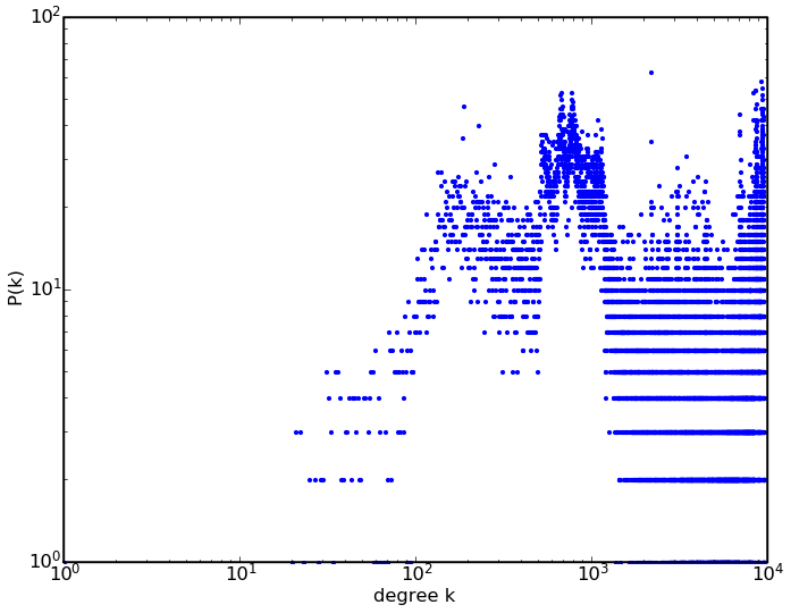


図 4.4: 決定的モデルによって得られたグラフの次数分布 ($c = 0.5$)

ネットワークの次数分布がベキ法則に従うならば，その次数分布は両対数プロットで直線になる．それぞれの係数のグラフを見ると，どのグラフの次数分布も直線にはなっていないことが分かる，そのため，どのグラフにも Scale Free 性は見られないといえる．また，もし決定的モデルが Blog ネットワークを再現しているのならば，モデルによって得られるネットワークの次数分布のグラフは，本来の Blog ネットワークの次数分布を示している図 3.5(a) のグラフと一致するはずである．しかしどの係数の次数分布のグラフを見ても，図 3.5(a) のグラフと一致しているとはいえない．そのため，決定的モデルは本来の Blog ネットワークを再現していないと結論づけられる．

それぞれのグラフについて，小さい次数を持つ頂点が極端に少なく，また大きい次数を持つ頂点が多いことが見てとれる．これは Blog ネットワークの区間の分布がベキ法則に従ってはいるものの，傾きが小さく，その結果長い区間を持つ頂点が多数存在したためだと考えられる (図 3.4(a))．

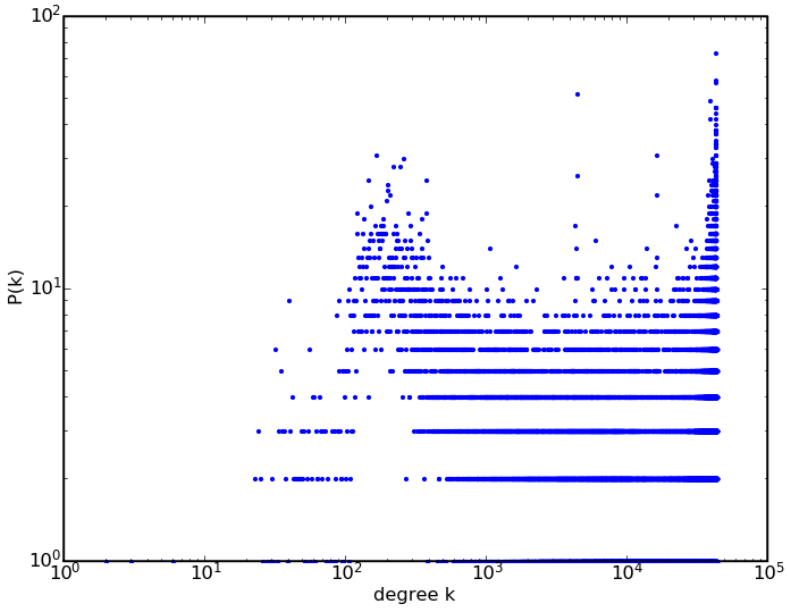


図 4.5: 決定的モデルによって得られたグラフの次数分布 ($c = 1$)

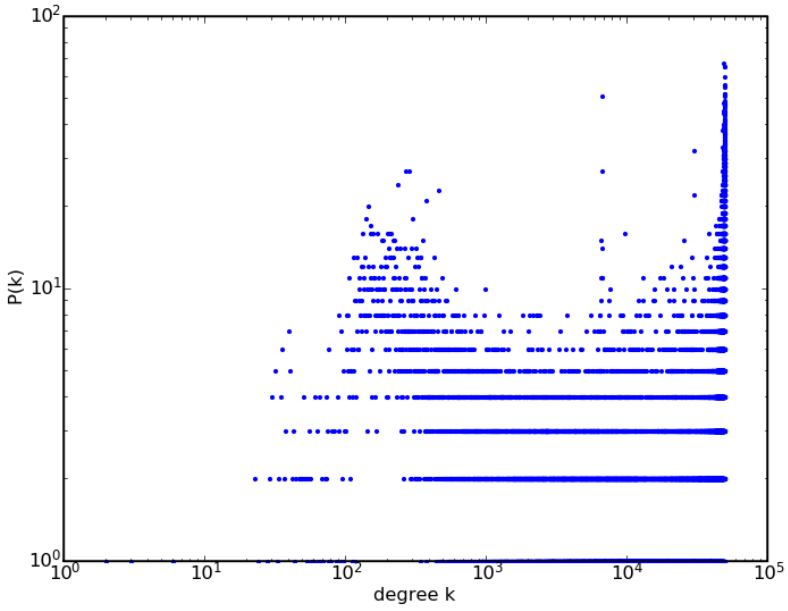


図 4.6: 決定的モデルによって得られたグラフの次数分布 ($c = 1.5$)

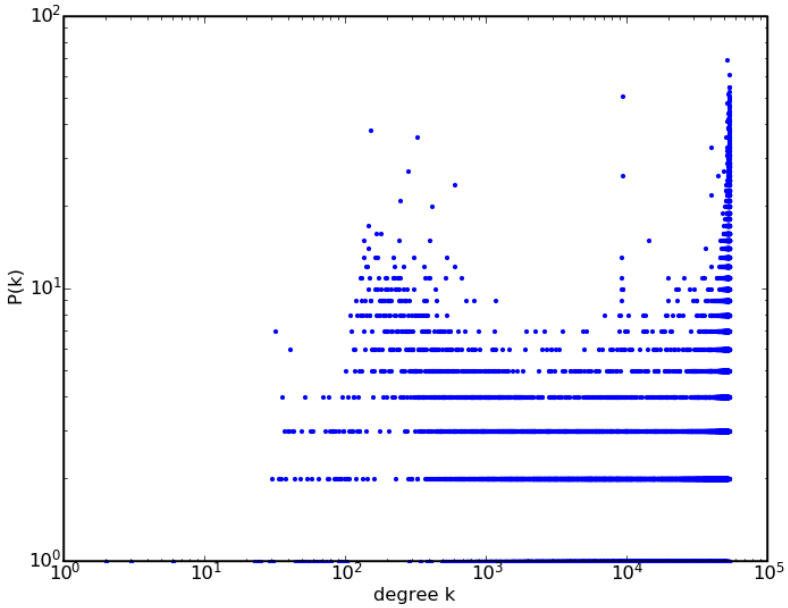


図 4.7: 決定的モデルによって得られたグラフの次数分布 ($c = 2$)

4.2 確率的モデル

4.2.1 依存モデル

決定的モデルでは，Scale Free 性を再現することはできなかった．本節では，Blog ネットワークにおける辺が張られる傾向を解析し，その傾向によって確率的に辺を張るモデルを検証する．

Blog ネットワークにおける辺が張られる傾向については，Max-Tolerance Graph を参考に以下の 5 つを考える．

(1) 区間長モデル

区間に重なりがある場合，区間の長い方の長さに依存して，辺が張られる傾向にあるのではないかと予想 (図 4.8).

(2) 更新回数モデル

区間に重なりがある場合，更新回数の多い方の更新回数に依存して辺が張られる傾向にあるのではないかと予想 (図 4.9).

(3) 更新頻度モデル

区間に重なりがある場合，更新頻度 (更新回数を区間の長さで割ったもの) の高い方の更新頻度に依存して辺が張られる傾向にあるのではないかと予想.

(4) 生存区間長モデル

2 つの区間において，どちらか片方が更新を終えた時，それ以降の生存している区間の

長さは更新を終了した区間には影響を与えないと考えられる．つまり，2つの区間において，どちらか片方の更新が終えたときまでの区間の長い方の長さに依存して，辺が張られる傾向にあるのではないかと予想 (図 4.10).

(5) 共通区間長モデル

区間の重なりの長さに依存して辺が張られる傾向にあるのではないかと予想 (図 4.11).

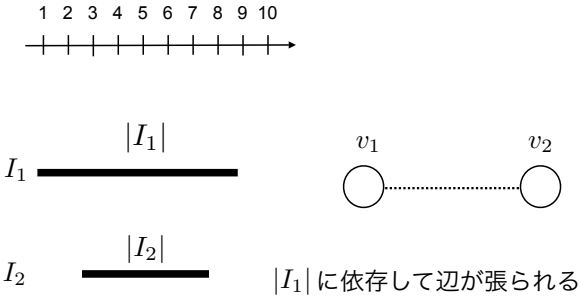


図 4.8: 区間長モデル（区間の長い方の長さに依存して辺が張られる）

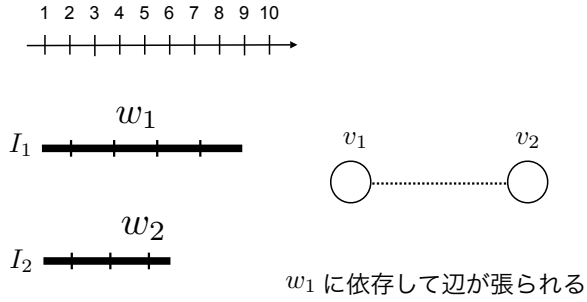


図 4.9: 更新回数モデル（更新回数の多い方の更新回数に依存して辺が張られる）

4.2.2 辺が張られる割合

以下に辺が張られる割合をプロットしたグラフを示す．図 4.12 は区間長モデル，図 4.13 は更新回数モデル，図 4.14 は更新頻度モデル，図 4.15 は生存区間長モデル，図 4.16 は共通区間長モデルである．

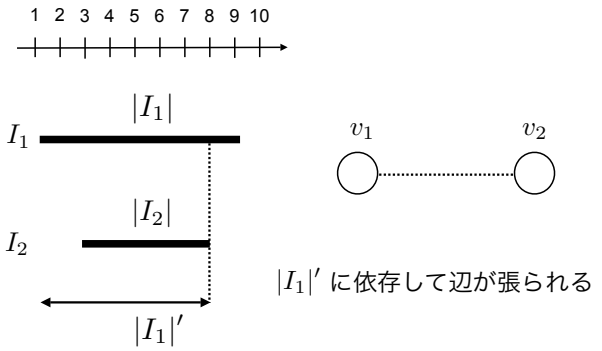


図 4.10: 生存区間長モデル（片方の更新が終えた時までの区間の長い方の長さに依存して辺が張られる）

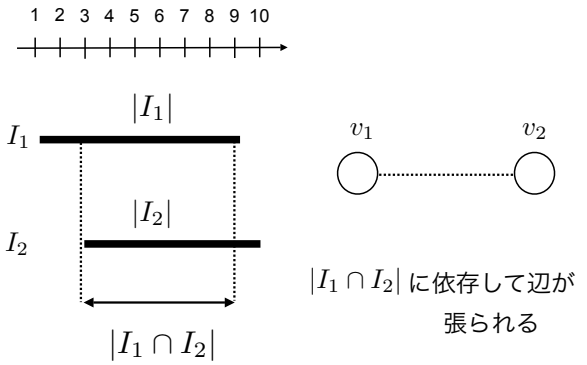


図 4.11: 共通区間長モデル（区間の重なりの長さに依存して辺が張られる）

それぞれについて，依存する要素が大きくなるにつれて，辺が張られる割合が高くなる傾向が見られる．しかし共通区間長モデル以外のモデルについて，依存する要素が小さいところでも辺が張られる割合が高くなっている．このような現象が起こる原因については，宣伝目的のために無差別にリンクを張る spam blog の影響ではないかと考えられる．

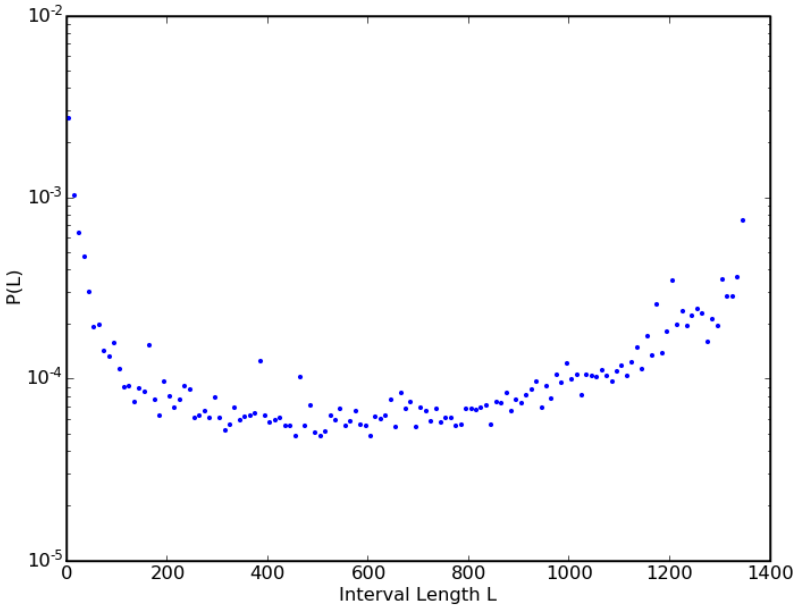


図 4.12: 区間長モデルの辺が張られる割合

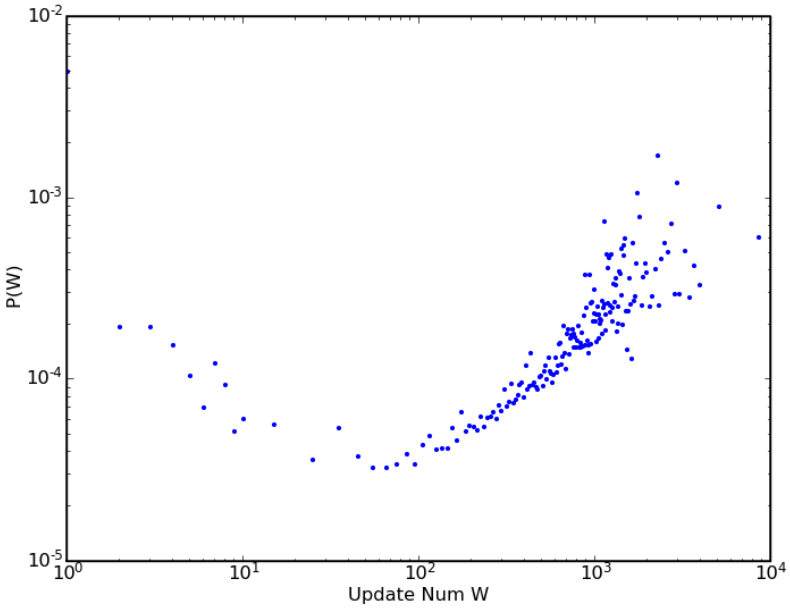


図 4.13: 更新回数モデルの辺が張られる割合

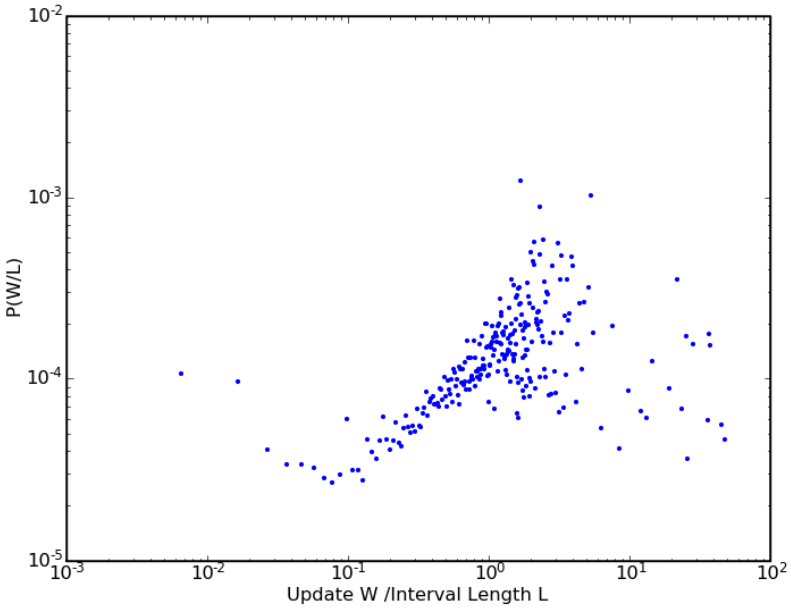


図 4.14: 更新頻度モデルの辺が張られる割合

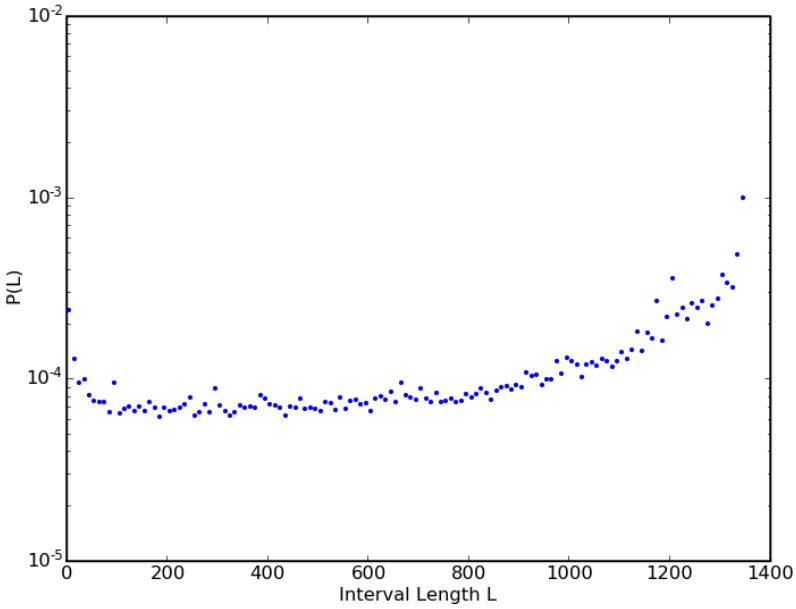


図 4.15: 生存区間長モデルの辺が張られる割合

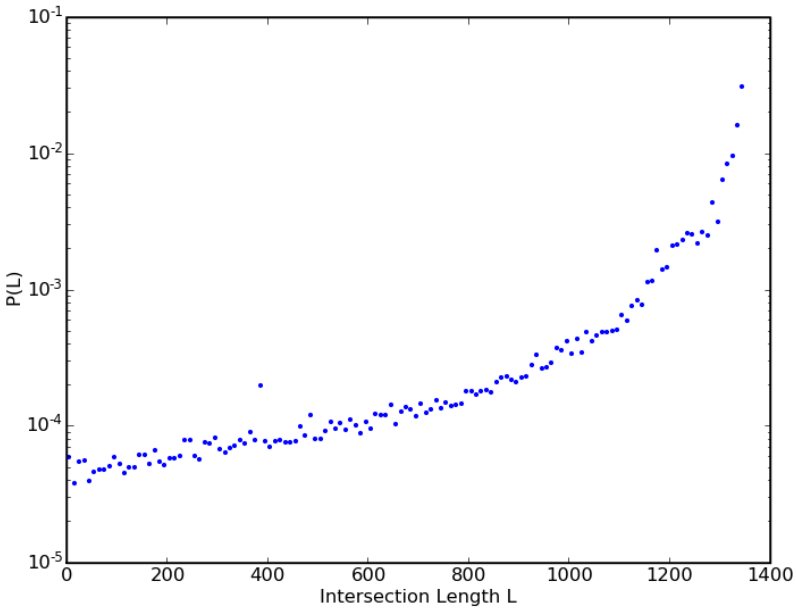


図 4.16: 共通区間長モデルの辺が張られる割合

4.2.3 辺が張られる傾向の検証

4.2.1 節で求めた辺が張られる傾向が，本当に Blog ネットワークにおける辺が張られる傾向を示しているのか次の方法で検証した (図 4.17).

- Blog ネットワークから辺を取り除く
- 残った区間に対して，4.2.1 節で求めた確率で辺を張り直す
- 辺を張り直したネットワークの次数分布を求め，元の Blog ネットワークの次数分布と比較する

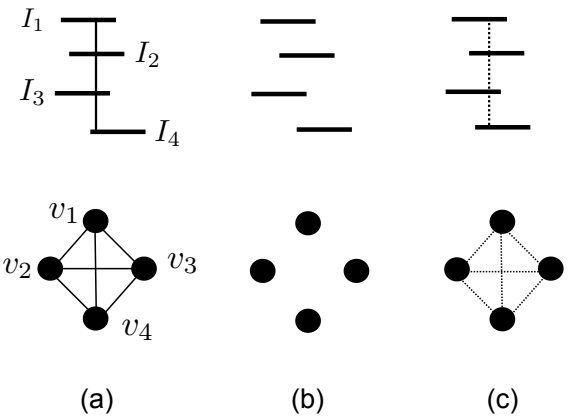


図 4.17: (a)Blog ネットワークから辺を取り除く．(b) モデルに応じた確率で辺を張り直す．(c) 辺を張り直したネットワークの次数分布を求める．

4.2.4 結果

以下に辺を張り直した Blog ネットワークの次数分布を示す．図 4.18 は区間長モデル，図 4.19 は更新回数モデル，図 4.20 は更新頻度モデル，図 4.21 は生存区間長モデル，図 4.22 は共通区間長モデルである．なお赤色でプロットされている点は本来の Blog ネットワークの次数分布である．

それぞれのモデルの次数分布について，次数が大きいところではベキ法則に従っているかに見える．しかし次数が小さいところでは全てのモデルにおいて，ベキ法則に従わないという結果が得られた．

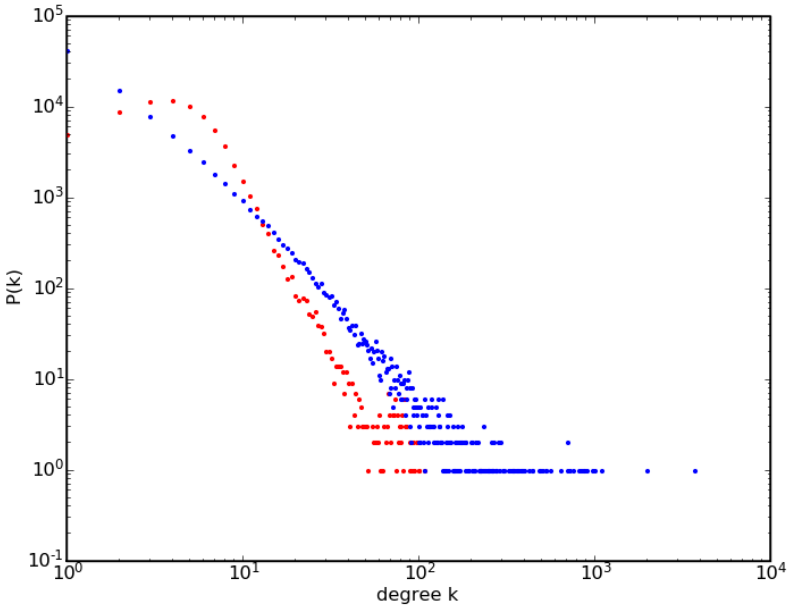


図 4.18: 区間長モデルの辺を張り直したネットワークの次数分布

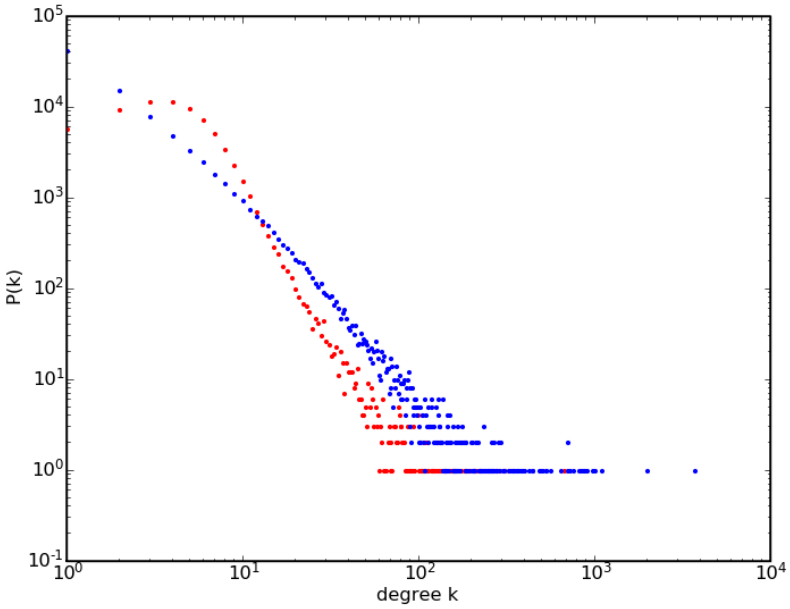


図 4.19: 更新回数モデルの辺を張り直したネットワークの次数分布

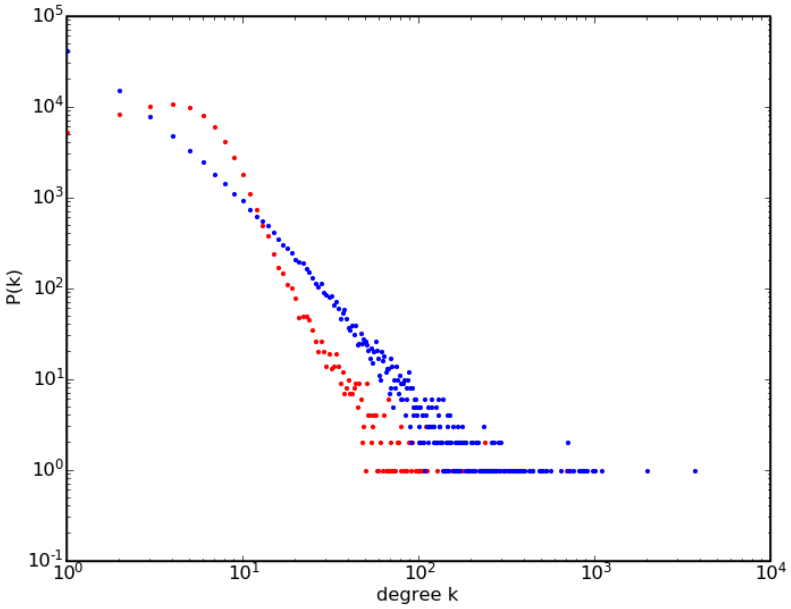


図 4.20: 更新頻度モデルの辺を張り直したネットワークの次数分布

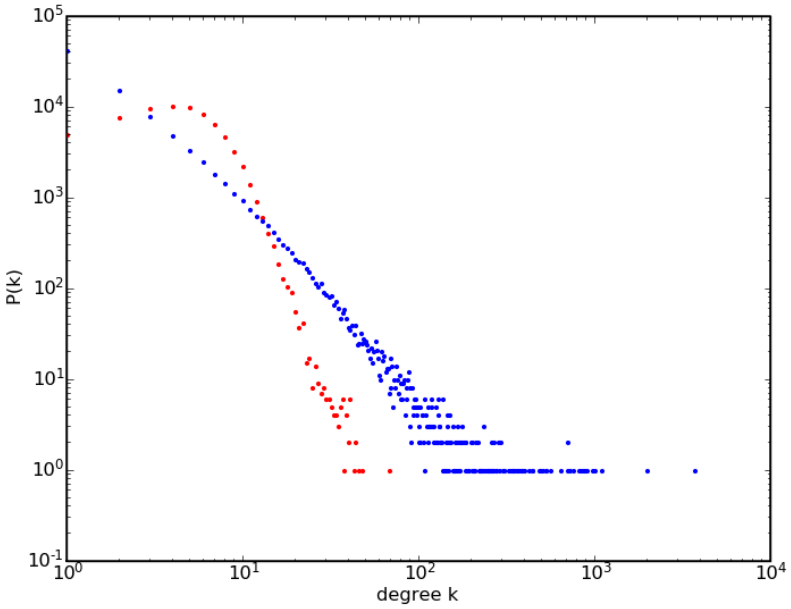


図 4.21: 生存区間長モデルの辺を張り直したネットワークの次数分布

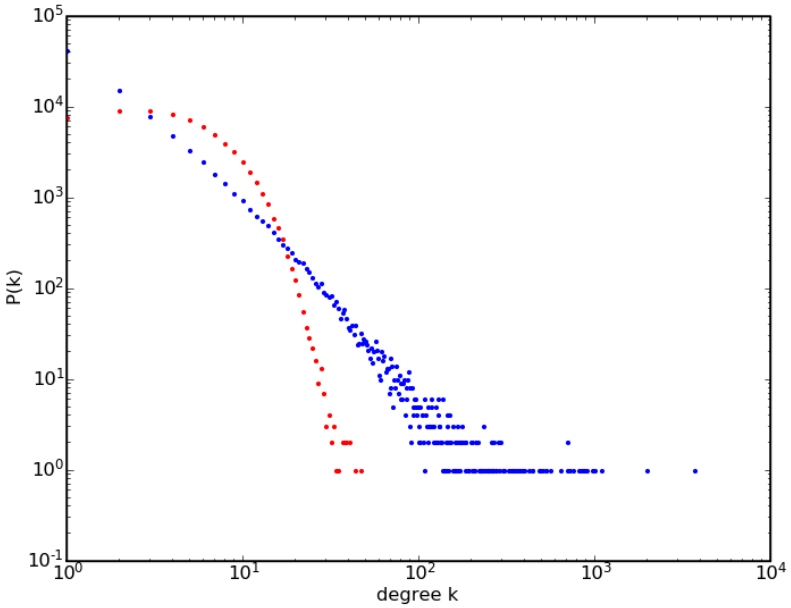


図 4.22: 共通区間長モデルの辺を張り直したネットワークの次数分布

4.3 確率的モデル + Preferential attachment モデル

4.3.1 手法

本節では前節の確率的モデルと Barabási と Albert による preferential attachment モデルを組み合わせたモデルについて検証する．モデル化は以下の手順で行う．

1. Blog ネットワークを解析して
   (a) 区間の長さと出次数の相関を求める．
   (b) 更新回数と出次数の相関を求める．
2. それぞれの頂点は，1 で得られた相関より
   (a) 区間の長さに応じた出次数を持つ (図 4.23)．
   (b) 更新回数に応じた出次数を持つ．
3. それぞれの頂点は，区間に重なりを持つ頂点に対して，自分の出次数が 2 で決定された出次数に至るまで辺を張る．その時，区間に重なりを持つ頂点は
   (a) 区間の長さが長ければ長い程，辺を張られやすい (図 4.24)．
   (b) 更新回数が多ければ多い程，辺を張られやすい．

4. 以上の操作により得られたネットワークは無向グラフとする

上記の操作の (a) を区間長モデル，(b) を更新回数モデルとする．

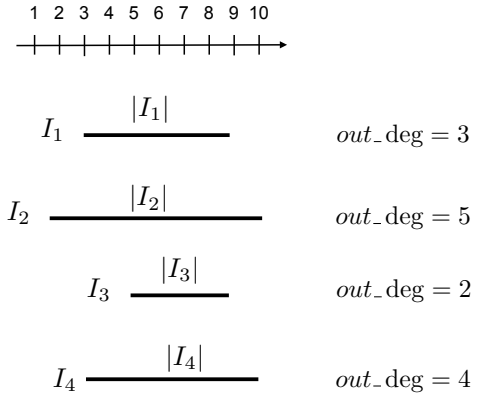


図 4.23: 区間の長さと出次数の相関より区間の長さに応じた出次数が決定される．

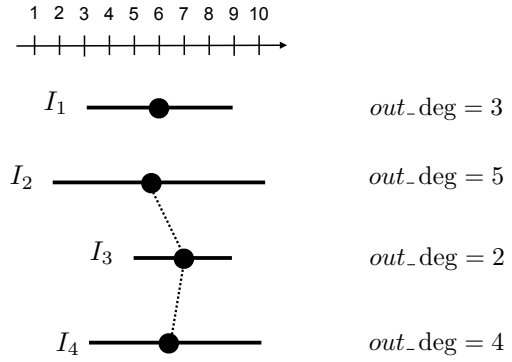


図 4.24: 区間の長さが長ければ長い程，辺が張られやすい

4.3.2 結果

以下に確率的モデルと preferential attachment モデルを組み合わせたモデルにより得られたネットワークの次数分布を示す．図 4.25 は区間の長さに依存するモデル，図 4.26 は更新回数に依存するモデルである．赤色でプロットされた点は本来の Blog ネットワークにおける次数分布である．

図 4.25，図 4.26 ともに，確率的モデルで見られた次数が小さい場所での次数分布の落込みが抑えられている．特に更新回数モデルの結果である図 4.26 では，傾きこそ本来の

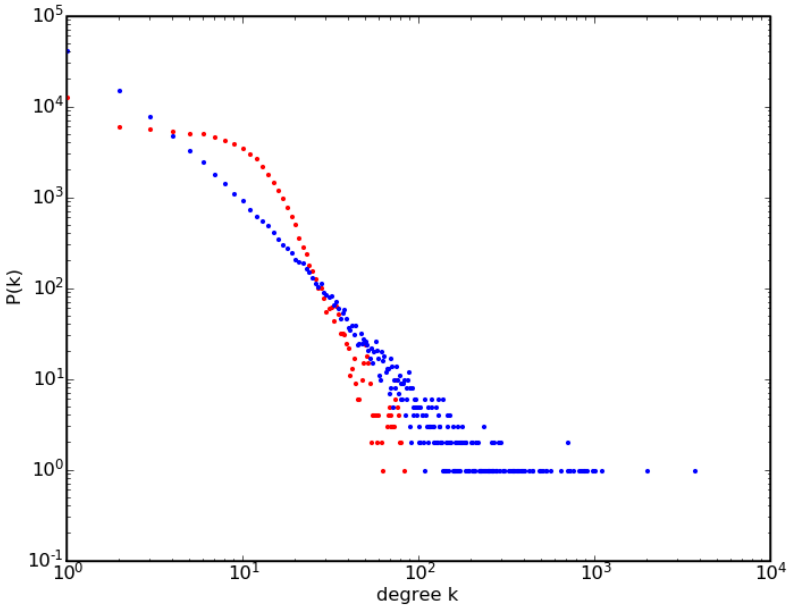


図 4.25: 区間長モデルの次数分布

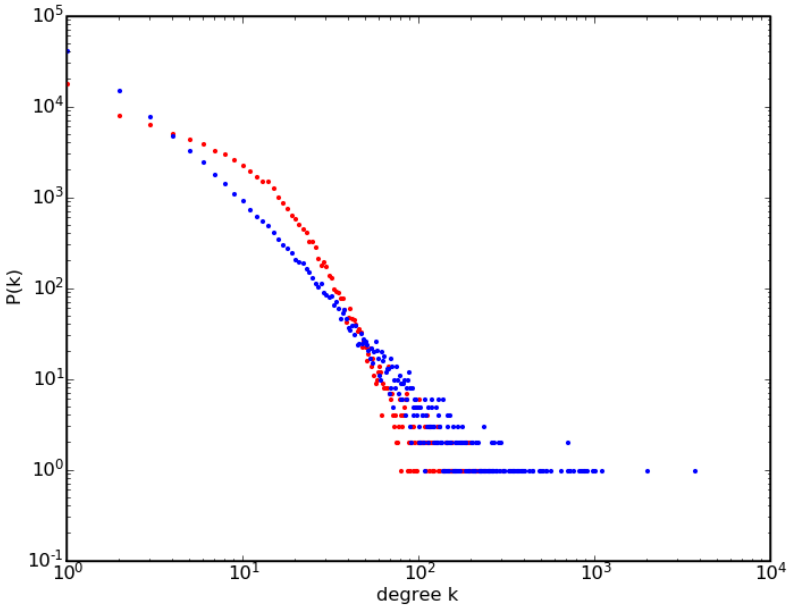


図 4.26: 更新回数モデルの次数分布

Blog ネットワークと異なるものの，次数が小さいところ，大きいところの両方でベキ則に従っていることが見てとれる．

第5章　まとめ

本論文では，時系列データに基づいた Scale Free Interval Graph について新しいモデルの提案をおこなった．

決定的モデル，確率的モデル，確率的モデルと preferential attachment モデルを組み合わせたモデルの 3 つのモデルを検証し，その中で確率的モデルと preferential attachment モデルを組み合わせたモデルから得られるネットワークが 本来の Blog ネットワークに一番近い結果となった．

本論文では実験的手法のみに留まっており，これらのモデルの理論的な解析が今後の課題として上げられる．

謝辞

本研究の遂行にあたり，日頃より懇切丁寧なご指導を賜りました上原隆平准教授に，心から感謝いたします．浅野哲夫教授，金沢工業高等専門学校元木光雄准教授，清見礼助教，斎藤寿樹氏をはじめとする浅野研究室，上原研究室の学生の皆様には，数多くの有益な助言やご支援をいただき，厚くお礼申し上げます．また，有益かつ貴重なデータを提供して頂いた エキサイト株式会社の方々に深く感謝いたします．最後に，大学院での研究を支えてくれた家族に感謝します．

参考文献

[1] Barabási A., Albert R., Emergence of Scaling in Random Networks, Science 286(5439), p.509–512, 1999

[2] Watts D.J., Strogatz D.H., Collective Dynamics of 'Small-World' Networks, Nature 393, p.440-442, 1998

[3] 増井直樹，今野紀雄，複雑ネットワークの科学，産業図書株式会社， 2005

[4] Miyoshi N., Shigezumi T., Uehara R., Watanabe O., Scale Free Interval Graphs, International Conference on Algorithmic Aspects in Information and Management(AAIM 2008), Lecture Notes in Computer Science Vol. 5034, p.292–303, 2008

[5] GOLUMBIC M.C., TRENK A. N., Tolerance Graphs, CAMBRIDGE UNIVERSITY PRESS, 2004

[6] Bollobás B., Rordan O., Spencer J., Tusnady G., The degree sequence of a scale-free random graph process, Random Structures and Algorithms 18, p.279–290, 2001