

ループ強化に基づく逆優先選択とメッセージ 伝搬法の情報拡散や攻撃耐性に対する強固な 効果

Liao Fuxuan

主任指導教員 林 幸雄
副主任指導教員 小谷 一孔
副主任指導教員 坂本 二郎

北陸先端科学技術大学院大学
先端科学技術研究科
融合科学共同専攻

A common topological structure, known as scale-free (SF), is observed in many social, technological, and biological networks. SF networks have efficient short paths between nodes but are highly vulnerable to intentional attacks targeting hubs with high degrees. This combination of efficient paths and extreme vulnerability creates a double-edged sword. Additionally, SF networks can be generated using a preferential attachment rule, such as the Barabasi and Albert (BA) model, where new nodes are linked to existing nodes with a probability proportional to their degree.

Due to the presence of hub nodes, many researchers believe that these hub nodes possess stronger dissemination capabilities, often referred to as influential nodes. Identifying influential nodes as seeds in real networks is an important problem with wide-ranging applications. However, conventional heuristic methods do not account for the overlap phenomenon. To address this issue and prevent overlapping of spreading, we propose a new method that combines statistical physics approaches and multi-hop coverage. We also introduce a faster epidemic model that eliminates the need for stochastic behavior averaging. Computer simulations demonstrate that our method outperforms conventional methods in terms of stronger spreading power per seed.

Furthermore, Self-organization of robust and efficient networks is important for the future designs of communication or transportation systems, because both characteristics are not coexisting in many real networks. As one of the candidates for the coexisting, the optimal robustness of onion-like structure with positive degree-degree correlations has recently been found, and it can be generated by incrementally growing methods based on a pair of random and intermediation attachments with the minimum degree selection. We introduce a continuous interpolation by a parameter $\beta \leq 0$ between random and the minimum degree attachments to investigate the reason why the minimum degree selection is important. However, we find that the special case of the minimum degree attachment can generate highly robust networks but with low efficiency as a chain structure. Furthermore, we consider two intermediation models modified with the inverse preferential attachment for investigating the effect of distance on the emergence of robust onion-like structure. The inverse preferential attachments in a class of mixed attachment and two intermediation models are effective for the emergence of robust onion-like structure. However, a small amount of random attachment is necessary for the network efficiency, when β is large enough. Such attachment models indicate a prospective direction to the future growth of our network infrastructures.

Keywords:

Network science, Self-organization, Minimum degree attachment, Influence maximum problem, Statistical physics approach

目次

第1章 緒言	1
1.1 研究の背景と目的	1
1.2 論文の構成	6
第2章 ネットワークの基本的な用語や既存の解析指標及び近似解法	7
2.1 基本的な用語と概念	7
2.1.1 ネットワーク科学の基本的な用語	7
2.2 主な解析指標	9
2.2.1 次数 (Degree)	9
2.2.2 三角形数とクラスタリング係数	9
2.2.3 ネットワーク上のパーコレーション	12
2.2.4 近接中心性 (closeness centrality)	12
2.2.5 媒介中心性 (betweenness centrality)	13
2.2.6 次数相関 (Degree-degree correlations)	15
2.2.7 頑健性 (Robustness)	15
2.2.8 通信効率 (Efficiency)	17
2.3 Feedback Vertex Set (FVS)	18
第3章 最小頂点被覆問題 (Minimum Vertex Cover)	20
3.1 最小頂点被覆問題とは	20
3.2 コンピューター科学における2-近似アルゴリズム	23
3.3 統計物理のアプローチ Survey Propagation 法	25
3.4 初期値やラウンド数の非依存性	27
3.5 Survey Propagation(SP) による最小頂点被覆と2-近似法による最小頂点被覆の近似解の比較	30
第4章 影響力最大化問題における複数の種の探索	32
4.1 IMP とは	32

4.1.1	独立カスケード (IC) モデル	32
4.1.2	線形閾値 (LT) モデル	36
4.1.3	LT モデルと IC モデルの違い	40
4.2	拡散処理と高速化計算	41
4.2.1	従来の SIR モデル	41
4.2.2	高速化 SIR モデル	42
4.2.3	従来の SIR モデル <i>vs</i> 高速 SIR モデル	43
4.3	従来の IMP に対するヒューリスティック手法	46
第 5 章	IMP への近似手法の提案	51
5.1	提案する組み合わせ法	51
5.2	提案手法と従来の手法の比較結果	54
5.3	本章のまとめ	61
第 6 章	頑健性かつ効率的なネットワークを生成するアタッチメントモデルの族	62
6.1	Scale-Free 性と優先的選択 BA モデル	62
6.1.1	Scale-Free(SF) 性	62
6.1.2	優先的選択 (Preferential Attachment) と BA モデル	63
6.2	逆優先的選択 (Inverse Preferential Attachment) と玉ねぎ状ネットワーク (Onion-like Network)	67
6.2.1	玉ねぎ状ネットワーク (Onion-like Network)	68
6.2.2	玉ねぎ状ネットワークにおける最小次数結合の要因	69
6.2.3	少量のランダムアタッチメントの必要性和効率的な経路	80
6.3	$k^{-\beta}$ アタッチメントを用いた改良された p-モデル	84
6.4	IPA モデル上の IMP のシミュレーション結果	94
6.5	本章のまとめ	101
第 7 章	結言	102
7.1	研究成果の要約	102
7.2	今後の課題	103

目 次

1.1	三点の関連性。青色は既知の部分を示し、緑色は私の提案部分を示す。	4
2.1	三つのノード（黒丸）と二本のリンク（黒線）で構成された単純無向グラフの例。	8
2.2	多重リンク、自己ループ、孤立ノードの例。 これらを含んでいないネットワークのみを考える。	8
2.3	クラスタリング係数の例	11
2.4	近接中心性の例。	13
2.5	次数順攻撃に対する最大連結成分サイズの減少。一つのネットワークに対して Hub 攻撃を適用した結果を示している。 R 値は各線の下側領域面積に相当する。	17
3.1	VC の例	21
3.2	貪欲アルゴリズム VC の例	23
3.3	SP 法で3つの異なる状態を確率的に決定する方法。0、*、1 は、白色、灰色の、黒色の丸で表される。下部のノード i の状態は、上部の隣接ノード ∂i の状態によって決定される。(a) ∂i に白色の丸がない場合、下部のノード i は被覆する必要がなく、白色（状態0）になる。(b) ∂i に白色の丸が1つある場合、下部のノード i は灰色（ジョーカー状態）になりる。(c) ∂i に白色の丸が2つ以上ある場合、下部のノード i は常に被覆される状態（状態1）で黒色になる。	25
3.4	木の中の VC を見つけるための割り当て法 [1]。	31
4.1	Step 0、ノード 8 とノード 9 が最初の活性ノードである。	33
4.2	Step 1、ノード 1 と 6 が成功して活性化されたと仮定する。	34
4.3	Step 2、ノード 1 と 6 が活性化された。	35

4.4	Step 3、ノード 4 の隣接ノードはどれも活性化されていないと仮定する。	35
4.5	Step 0、ノード 8 とノード 9 が最初の活性化ノードである。	36
4.6	Step 1、ノード 5 が活性化する。	37
4.7	Step 2、ノード 6 が活性化する。	38
4.8	Step 3、ノード 7 と 1 が活性化する。	38
4.9	最終段階	39
4.10	SIR モデルのフローチャート	42
4.11	従来の SIR モデルと高速 SIR モデルにおける、三つの状態 S、I、R の割合の時間変化を示す。	45
4.12	k-core 分解	47
4.13	k-shell 分解	47
4.14	A set of vertexes in 2-ホップ from a vertex w in the process of LC.	49
5.1	(a) 重なり合う状況と、(b) 重なりを避ける状況。	51
5.2	l -ホップ被覆の概念を説明する図解。	52
5.3	LastFM の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 1, 2, 4$ の場合を比較している。	56
5.4	各種のペア (i, j) 間の距離 $d_{i,j}$ の分布を 2-または 4-ホップ被覆 ($l = 2, 4$)。	57
5.5	LastFM の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。	59
6.1	航空網データ AirTraffic を (A) 可視化したものと (B) 次数分布. ノードは次数が高いほど大きく表示されている。また次数分布は両対数スケールになっている。	64
6.2	BA モデルを生成の見取り図	65
6.3	玉ねぎ状ネットワークの生成	68
6.4	SF ネットワークと玉ねぎ状ネットワーク次数分布の比較	69
6.5	分布の裾野を緑の線に $e^{-\beta k \log k}$ で推定。紫の点と緑の線は、 $k^{-\beta}$ アタッチメントによって成長するネットワークの次数分布とその推定値を表している。時間ステップごとの追加リンク数は $m = 2$ である。	73

- 6.6 分布の裾野を緑の線に $e^{-\beta k \log k}$ で推定。紫の点と緑の線は、 $k^{-\beta}$ アタッチメントによって成長するネットワークの次数分布とその推定値を表している。時間ステップごとの追加リンク数は $m = 2$ である。 74
- 6.7 $\beta \rightarrow \infty$ の最小次数アタッチメントによるリンクの結合先。黒い線は、新しいノードから既存の m ノードへの最小次数アタッチメントによるリンクの結合を示している。右から左に3つの部分がある。右側の最初の部分では、 $m - 1$ 個の既存のノードのうち次数が最小の $m, m + 1, \dots, 2m - 2$ に新しいノードが結合される。最小次数アタッチメントのためである。2番目の部分（中央）では、残りの1つのリンクが次数 $2m - 1$ のノードに結合される。 Δt 個の次数が $2m$ 未満のノードが存在するため、次数が $2m - 1$ の最も古いノードの追加時間は $t - \Delta t$ となる。 $t - 1 - \Delta t$ 個のノードのうち、第3部分（左側）には、次数が $2m$ のノードが $t - 1 - \Delta t$ 個存在する。これらのノードの追加順序と、時間 t におけるノードの次数の関係は、下から2行目と3行目に示されている。 77
- 6.8 $k^{-\beta}$ アタッチメントによって生成されたネットワークにおけるノードサイズ N に対する指標： r 、 R 、 E 。 $m=2$ の場合で、リンク/時間ステップ。紫色（十字印）、緑色（正方形印）、シアン（丸印）、黒（三角形印）、オレンジ（逆三角形印）の線は、それぞれ $\beta = 0, 1, 10, 50$ 、および 100 に対応する。 78
- 6.9 $k^{-\beta}$ アタッチメントによって生成されたネットワークにおけるノードサイズ N に対する指標： r 、 R 、 E 。 $m=2$ の場合で、リンク/時間ステップ。紫色（十字印）、緑色（正方形印）、シアン（丸印）、黒（三角形印）、オレンジ（逆三角形印）の線は、それぞれ $\beta = 0, 1, 10, 50$ 、および 100 に対応する。 79

- 6.10 ランダムと $k^{-\beta}$ アタッチメントの混合アタッチメントによるネットワークにおける、ノードサイズ $N = 5000$ で $m = 2$ (左) と $m = 4$ (右) の場合の、混合確率 p に対する効率の変化。ここで、 p はランダムアタッチメントの割合を示している。紫色の線 (十字印)、緑色の線 (正方形印)、茶色の線 (五角形印)、水色の線 (丸印)、黒色の線 (三角形印)、オレンジ色の線 (逆三角形印)、赤色の線 (ダイヤモンド印) は、それぞれ $\beta = 0, 1, 4, 10, 50, 100$ 、および $\beta \rightarrow \infty$ を表している。 81
- 6.11 $k^{-\infty}$ アタッチメントにおける平均経路長 $\langle L \rangle$ とノード数 N の関係を示したグラフ。縦軸のスケールは (a)~(d) で異なることに注意。ここで、 p はランダムアタッチメントの割合を示し、 $m = 2$ (右) および $m = 4$ (左) での結果を示している。 82
- 6.12 成長過程を示す図として、(a) ~ (c) に、異なるランダム結合の数の持つ修正された p-model [2] を示す。図 (d) に示されているように、M-MED モデルは、MED モデル [3, 4] の最小次数選択ではなく、 $k^{-\beta}$ attachment で修正されている。距離 d と $\mu + 1$ は、p-models と MED モデルの場合、ランダムに選択されたノードからのホップ数で測定される。 85
- 6.13 p-model0 における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。 89
- 6.14 p-model1 における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。 90
- 6.15 p-model2 における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。 91
- 6.16 M-MED における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。 92

6.17	成長するネットワークにおける4つの修正モデルの次数分布は、 $m = 4$ で似たような分布を持っている。各々の (a) (b) (c) (d) において、p-model0、p-model1、p-model2、M-MED モデルで似たような分布がある。分布の裾は、半対数プロットで指数分布として線形である。 $\beta \rightarrow \infty$ の緑の線（三角形印）とオレンジの線（逆三角形印）の傾きは、 $\beta = 1$ の紫の線（十字印）とシアンの線（アスタリスク印）の傾きよりも急である。したがって、大きな次数は β が増加するにつれて制限される。つまり、ハブは存在しない。	93
6.18	SF ネットワークと IPA ネットワークの比較	95
6.19	β は1の時、 $l = 2, 3, 4$ で、SF ネットワーク上各 IMP 解法のシミュレーション結果。	96
6.20	β は-10の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各 IMP 解法のシミュレーション結果。	97
6.21	β は-20の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各 IMP 解法のシミュレーション結果。	98
6.22	β は-50の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各 IMP 解法のシミュレーション結果。	99
6.23	β は-100の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各 IMP 解法のシミュレーション結果。	100
7.1	Polblogs の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	106
7.2	Polblogs の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を2から10まで変化させた場合の図。	107
7.3	Airtraffic の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	108
7.4	Airtraffic の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を2から10まで変化させた場合の図。	109
7.5	Hamster の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	110
7.6	Hamster の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を2から10まで変化させた場合の図。	111

7.7	Gnutella の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	112
7.8	Gnutella の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。	113
7.9	GRQC の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	114
7.10	GRQC の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。	115
7.11	HeptH の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	116
7.12	HeptH の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。	117
7.13	PGP の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。	118
7.14	PGP の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。	119

表 目 次

3.1	Maaya 上の VC 初期値変更。	27
3.2	Hamster 上の VC 初期値変更。	27
3.3	OpenFlight 上の VC 初期値変更。	28
3.4	Polblogs 上の VC 初期値変更。	28
3.5	Maaya 上の VC ランダム処理。	28
3.6	Hamster 上の VC ランダム処理。	28
3.7	OpenFlight 上の VC ランダム処理。	29
3.8	Polblogs 上の VC ランダム処理。	29
3.9	逆温度パラメータ y が異なる場合について、最小 VC に関する SP 法、BP 法と 2-近似アルゴリズムによる推定値の $ VC $ を比較する。 太字の数値は最小 VC における最良の結果を表している。	30
3.10	LastFM における異なるラウンド数 T での SP 法による $ VC $ 。ノード数 $N = 7624$ 。	31
4.1	従来の SIR モデルにおいて、拡散時間が 2 から 7 のサンプルにおける、最も低い累積感染数と最も高い累積感染数のギャップを示す。	44
4.2	高速と従来の SIR モデルにおける収束までの差と計算時間について (CPU : i7-11800H、メモリ : 16GB)。	44
5.1	8 つのソーシャルネットワークの基本的なトポロジー特徴。	54
5.2	提案手法と従来の 6 つの手法による従来の SIR モデルと高速 SIR モデルで選択された各種ノードの感染力。 $N \times R(t_c)/N_s$ は、種ノードあたりの感染力を示し、 $R(t_c)$ は収束時間 t_c での累積感染を示す。	58
6.1	$N = 5000$ 個のノードを持つ初期完全グラフから、 $m = 2, 4$ の場合に生成された、 $\beta \rightarrow \infty$ の最小次数アタッチメントによる直径の推定値 $N/\Delta t$ と実測値。	75

6.2	修正された p-モデルと MED モデルにおける β 値の影響により、玉 ねぎ状ネットワークが現れる。MED-rand と MED-kmin は、それぞ れ $\beta = 0$ と $\beta = \infty$ の特殊な例である。RA と D-D は、ランダムア タッチメントと複写モデルを表す。	88
7.1	パラメータの組合せ	105

第1章 緒言

1.1 研究の背景と目的

2000 年前後、現実の複雑なネットワークの繋がりをデータ解析することで、多くの現実のネットワークに共通の繋がり構造を持つことが明らかになった [5, 6]。このような現象を研究する分野は、「複雑ネットワーク」または「ネットワーク科学」と呼ばれている。歴史的には 20 世紀に、Erdős-Rényi によって、単純な点と線として抽象化した 2 つのノードが一定の確率で結合されるランダムグラフ [7] が研究された。しかしながら、多くの現実のネットワークは、このようなランダムグラフにおける次数分布が Poisson 分布に従う特徴は持っていない。そこで、コンピューターによる大規模なデータ解析を用いて、現実のネットワークに共通する繋がり構造として Scale-Free(SF) 性が発見されてから、様々な分野で 21 世紀初頭から精力的に研究が行われるようになった。ここで、次数 (Degree) とは、あるノード i の次数 k_i はノード i に繋がるリンクの数で、ネットワークにおいて次数 k を持つノードの個数の割合を示すものを次数分布 $P(k)$ という。SF 性とは、ネットワーク内のノードの次数分布が以下のべき分布に従うことを指す。

$$P(k) \propto k^{-\gamma}. \quad (1.1)$$

現実のネットワークでは、べき指数が $2 < \gamma < 3$ となっている場合が多い [7]。べき乗則に従う分布を分かりやすく言い換えると、ごく少数のハブノードが非常に多いリンクを持ち、大多数のノードはごく少数のリンクしか持っていない繋がり構造を指す。このネットワーク構造は、上記のランダムグラフと比較して、一様ランダムノード故障に対しての連結性により強い頑健性がある。なぜなら、ランダムにいくつかのノードやリンクを削除しても、ハブは残る確率が高く、このネットワークはその連結性と伝達機能を維持することができる。この SF ネットワーク構造の存在性は、ソーシャルネットワーク [8]、インターネット [9]、タンパク質相互作用ネットワーク [10] など、多くの実際の応用例で広く確認されている。

ハブノードの存在により、多くの研究者はこれらのハブノードがより強力な拡散能力を持てと考える。ところで、ネットワーク上の伝達機能を考えた影響最大化 (IMP) は、情報拡散をモデル化したときに、多くのノードに情報が拡散するように、拡散の起点となる拡散の種のノード集合を求める問題となる [11]。例として、バイラルマーケティングが挙げられ、少数の人に商品の無料サンプルを渡すことで、その商品の噂が口コミを通して多数の人に伝わることを期待できる [12]。拡散力を持つ局所的なネットワークの特性（ハブ）を利用して、単一または複数の種を見つける方法を設計しているが、これらの手法には、二つの種間の距離を考慮してなく、拡散の重なり現象が発生する。そこで、我々は、拡散の重なり問題を考慮して、IMP 問題を解決するための新しい方法を見つけることを望んでいる。しかしながら、IMP は NP 困難な問題であり、そのため正確な解を迅速に見つけることは困難である。IMP 問題は頂点被覆 (VC) 問題と相互変換することができる。これは、IMP 問題を解決するために VC 問題のアルゴリズムを使用することができることを意味する [13]。VC 問題には、コンピュータ科学の 2-近似的アルゴリズムや統計物理学の Survey Propagation (SP) 法など、多くの近似解法が存在する。特に、統計物理学の SP 法は、VC 問題の解を見つける際に高い効率と優れた精度を持つことが証明されていた [14]。つまり、一つ目の課題として、ループと拡散の重なりを着目して、情報科学のマルチホップ被覆と、統計物理学的なメッセージ伝搬による近似解法を融合して、新たな方法を提案する [Liao2]。種間の距離を長くすることで、拡散の重なり問題を回避させる。更に、実験の過程で、私たちは従来の感染モデルでは、平均化するため多くの時間を要することを発見した。そこで、計算時間を短縮するために、ネットワーク科学における集団的影響力 (Collective influence) から着想を得た高速計算法を提案する [Liao2]。

更に、SF ネットワークは二つノード間にハブを経由するなどして効率的な最短経路が存在するが、巨大な次数を持つハブノードへの意図的な攻撃に対しては極めて脆弱で、経路の効率と極端な脆弱性の組み合わせは、諸刃の剣と言える。また、SF ネットワークは Barabasi and Albert (BA) モデル [6] のように、新しいノードからリンクされたノードがその次数に比例した確率で選択されるような、優先的選択の利己的ルールによって生成される。もし、ハブがなくなった場合、ネットワークはどのように変化するのかについて、我々は、成長するネットワークにおいて、強い頑健性と高い通信効率という重要な特性の両方を備える、利己的でな

い逆優先的選択（IPA）による、より良いアタッチメント法を考察する [Liao1]。

一方、次数相関ではなく、ループと頑健性のより本質的な関係が明らかにされつつある。特に、ノード除去により最大連結成分サイズがある一定値以下になる最小ノード集合 Dismantling Set を見つける問題は、除去によりループ無となる最小ノード集合 Feedback Vertex Set (FVS) を見つける問題と漸近的に等価になることが証明されている [15]。直感的にも、ループの無い木構造は攻撃に対して脆弱であることが分かる。さらに、ループ除去に必要なノード数 (FVS サイズ) が多いネットワークほど、攻撃に対して頑健であると考えられる [16]。また、ノード数が増加して成長する玉葱状ネットワークにおいて、FVS サイズと頑健性指標に強い相関関係があることが数値的に示されている [16]。玉ねぎ状ネットワークは、次数の大きい順に中心から周リンクに同心円上にノードを配置すると、正の次数相関を強くする、同程度の次数を持つノードが結合することで玉葱状に可視化されることからそう呼ばれる。次数分布を固定した時、次数相関が正の値を持つ玉葱状ネットワークが最適な頑健性を持つことが明らかになっているが、ランダム、仲介、距離、逆優先選択が玉葱状ネットワークの出現に支配的であるかは、明らかにされていない。そのため、私たちはパラメータの変化に応じてネットワークの成長方法が変化するモデル ($k^{-\beta}$ attachment) を提案した [Liao1]。最終的に、逆優先選択が玉ねぎ状ネットワーク形成の主要な要因であることが判明した。また、IPA によって生成される Regular-like ネットワークの拡散能力も研究する価値がある。そのため、私たちは SF ネットワークと IPA ネットワークを IMP 上で比較する実験を行い、以下の結論に達した：ハブを持つ SF ネットワークは、ノードの次数分布がより集中し、ネットワークはより脆弱である。一方、IPA ネットワークはより均等な次数分布を持ち、より頑健性がある。

上記のような背景に基づき、本研究では次の三点を具体的な目的としている。また、それら三点の関連性を図 1.1 に示す。

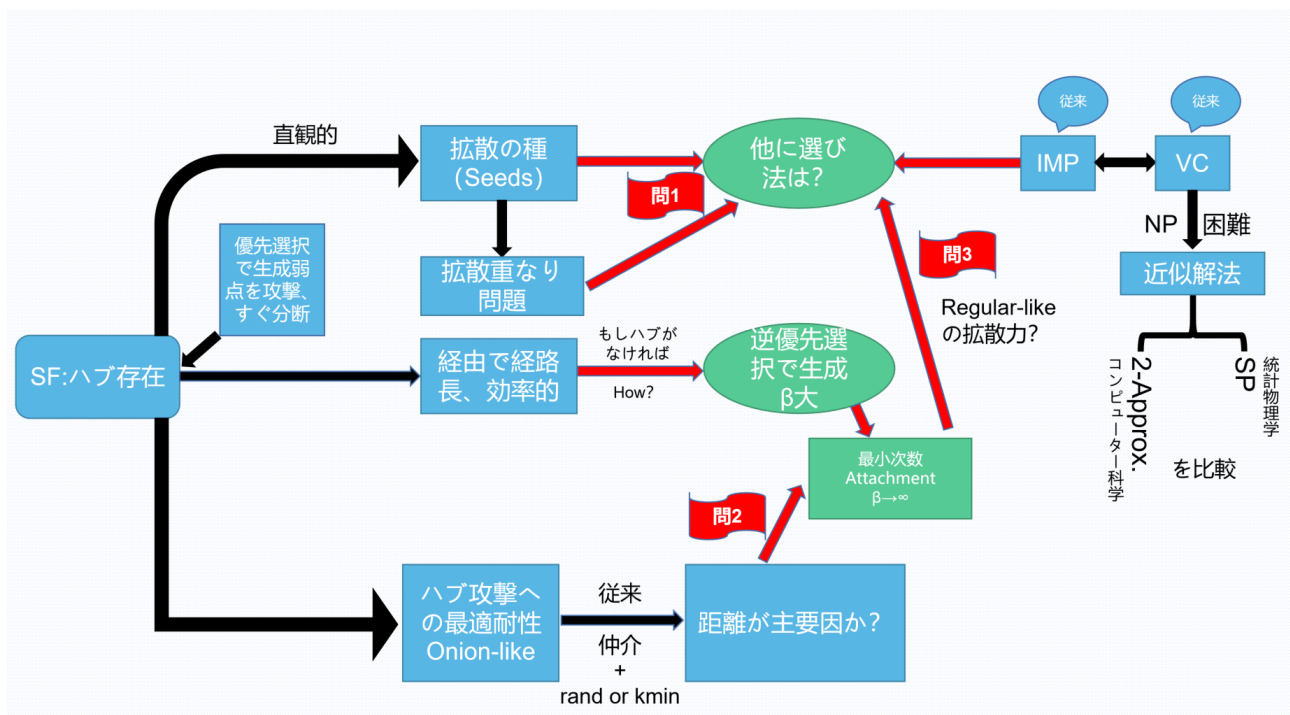


図 1.1: 三点の関連性。青色は既知の部分を示し、緑色は私の提案部分を示す。

問 1: IMP 問題に対して、拡散の種として影響力のあるノードを見つけることは、広い応用範囲に関わる重要な問題である。従来のヒューリスティックな手法では、ループと拡散の重なりを考慮しない。そこで、ループと拡散の重なりに着目して、統計物理学的アプローチ（ループ強化）とマルチホップ被覆（重なり解消）を組み合わせた新しい手法を提案する。より強力な拡散力を持つ種ノードを見つけるためには、拡散の重なりを回避することが重要であることが明らかになる。更に、従来の拡散処理では、多くの試行（平均化）が必要であるが、計算時間を短縮するために、統計物理のメッセージ伝搬に基づいて、新たな高速化計算で平均化不要として、数倍の速度向上が実現される。この成果は IMP の問題解決に新たなアプローチと手法を提供し、多くの応用分野が応用できる。例えば、グローバルな流行病の拡大を防止するための問題 [17] など。

問 2: 次数分布の幅が狭いほど頑健性が高くなると考えられ、分布を底上げする逆優先選択は重要な役割を果たす。頑健性の向上への逆優先選択の有効性を示すために、ランダムと最小次数のアタッチメントの間に連続補間を導入し、頑健性、次数相関、通信効率、複数の種攻撃をそれぞれ比較する。逆優先選択が頑健性と通信効率の向上に重要であることを示す。また、ランダムと IPA モデルの混合アタッチメントを提案して、ランダムアタッチメントが通信効率と平均経路長に与える影響を調べる。これらの結果は、高い頑健性と通信効率を持つネットワークの発見において、新たな視点を提供する。

問 3: 玉葱状ネットワークを生成するために、色々な手法が多く提案されてきた (MED-Rand/Kmin)。玉葱状構造の生成に重要な要因を見い出すために、従来の玉葱状ネットワーク生成法を部分変更して、次数相関及び頑健性と通信効率に関する指標を比較する。玉葱状構造の生成に対して、結合距離より逆優先選択は主な要因であることが分かる。これらは、玉ねぎ状ネットワークの生成に対する理解を深め、利己的な優先的選択の Rich get Richer 原理を得るというルールではなく、少量のランダムな結合がノードに等しく出会いの機会を提供することを意味する。

1.2 論文の構成

本論文の構成を以下に示す。

本論文では、全体を六章で構成している。

第一章の緒言に続いて、第二章では、複雑ネットワークの基本的な概念としてネットワークの定義について述べる。さらに、主な解析指標として、次数、クラスティング係数、パーコレーション、近接中心性、媒介中心性、次数相関、頑健性、通信効率について説明する。その後、Feedback Vertex Set を概説する。

第三章では、最小頂点被覆問題の定義を説明する。そして、コンピューター科学のアプローチからの 2-近似アルゴリズムと統計物理のアプローチからの Survey Propagation (SP) 法を説明する。さらに、SP 法の同時計算を提案した後、SP 法と 2-近似アルゴリズムで提案した最小頂点被覆の大きさを予備実験として比較する。

第四章では、影響力最大化問題 (IMP) に関して代表的な、独立カスケード (IC) モデルや線形閾値 (LT) モデルについて説明する。従来のサンプル数を必要とする SIR モデルと比較するため、メッセージ伝搬方程式から着想した高速 SIR モデルを提案する。また、IMP 問題のヒューリスティック手法について説明する。

第五章では、IMP を解決するための提案手法と従来の手法の比較結果を示す。

第六章では、広く知られた Scale-Free トポロジー構造及び逆優先的選択 (Inverse Preferential Attachment) と玉ねぎ状ネットワーク (Onion-like Network) の関係を明らかにする。さらに、 $k^{-\beta}$ アタッチメントを用いた改良された p-モデルを提案し、玉ねぎ状ネットワークの主要因を明らかにする。

第七章では、結論として以上の成果をまとめる。

第2章 ネットワークの基本的な用語 や既存の解析指標及び近似 解法

まず、2.1 節において本研究で扱うネットワークの基本的な用語や解析指標を紹介する。2.2 節では既存の主な解析指標を説明する。2.3 節で 最小 Feedback Vertex Set と二種類の近似解法を説明する。

2.1 基本的な用語と概念

2.1.1 ネットワーク科学の基本的な用語

本研究で議論するネットワークは、図 2.1 に示すように、グラフ $G = (V, E)$ における、ノード（ノード）の集合 $V(G) = \{1, 2, 3\}$ と、ノードを繋ぐリンク（リンク）の集合 $E(G) = \{e_{ij}\}$ から構成される。ここで、ネットワーク内のノード i とノード j を結ぶリンクを (i, j) と表記する。

但し、リンクの両端が 同じノードに繋がる自己ループ (self loop)、2つのノード間に複数のリンクが存在する多重リンク (multiple links) は認めない単純グラフ (simple graph) とする。また、リンクの向きを考慮する場合は有向グラフ (digraph: directed graph)、考慮しない場合は無向グラフ (undirected graph) と呼ばれるが、本論文で扱うネットワークは、全て無向グラフである。さらに、ネットワークには、どのリンクとも繋がっていない孤立ノードは無いものとする。図 2.2 に多重リンクと自己ループ、孤立ノードの例を示す。これらを持たない単純ネットワークを対象としている。本論文ではグラフ理論におけるノードとリンクを、ノードとリンクと呼ぶことで統一的に表記する。

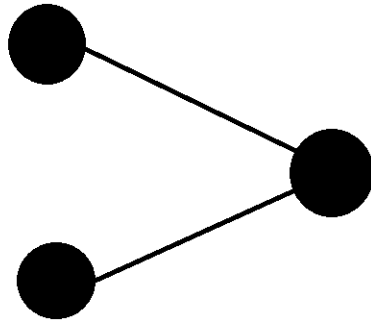
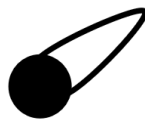


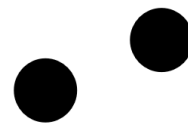
図 2.1: 三つのノード（黒丸）と二本のリンク（黒線）で構成された単純無向グラフの例。



多重リンク



自己ループ



孤立ノード

図 2.2: 多重リンク、自己ループ、孤立ノードの例。これらを含んでいないネットワークのみを考える。

2.2 主な解析指標

一般に、ネットワークを分析する際、様々な指標を用いる。以下、本論文で扱う解析指標として、次数、次数相関、頑健性、通信効率指標について述べる。

2.2.1 次数 (Degree)

N 個のノードと M 本のリンクからなるネットワークにおいて、あるノード i の次数 k_i は、ノード i に繋げるリンクの数 k_i として定義される。

$$k_i = \sum_{j=1}^N A_{ij}. \quad (2.1)$$

ここで、 A_{ij} は隣接行列で、ノード i と j の間に、リンクが存在する場合は $A_{ij} = 1$ 、存在しないの場合は $A_{ij} = 0$ と定める。リンクの両端には二つノードがあるため、次数を全てのノードで合計すると、 $\sum_{j=1}^N k_j = 2M$ となる。ネットワークを分析する上では、ある次数 k を持つノードの存在確率 $P(k)$ として次数分布 (degree distribution) [6] を議論することが多い。なぜなら、後述するように、その分布がどのような形をしているかによって、ネットワークの性質が大きく異なるからである。

2.2.2 三角形数とクラスタリング係数

三角形数 (Triangle Count) は、グラフ中の各ノードが 1 ホップのノードと直接的に三角形を形成できる数で定義される。三角形は、3 つのノードで構成される集合において、それぞれが他の 2 つのノードと直接結合している関係にある。全てのノードに対して三角形カウントアルゴリズムを実行することで、データセット全体のある種の集約状態を評価できる。

クラスタリング係数 (Clustering Coefficient) は、グラフのクラスタリングの程度を測定するために使用される。クラスタリング係数には、局所クラスタリング係数 (Local Clustering Coefficient) とグローバルクラスタリング係数 (Global

Clustering Coefficient) の2種類がある。あるノードの局所クラスタリング係数は、そのノードの近傍ノード同士が相互に結合されている度合いを示す。あるノードのクラスタリング係数が高いほど、そのノード周囲のノード同士の関係性が高いことを示し、グループの集約度が高いことを示す。その最大値は1であり、あるノードの周囲のすべての点同士が関係性を持っている場合である。

三角形数の計算式は、1つのノードから1ホップの隣接ノード同士の間でリンクが存在することで定義される。

$$Triangles = r_v. \quad (2.2)$$

ここで、 r_v は、ノード v の隣接の任意の2つのノードのペア数を表す。

局所クラスタリング係数の計算式は三角形のカウントに依存して、以下で定義される。

$$LCC_v = \frac{2r_v}{d_v(d_v - 1)}, \quad (2.3)$$

d_v はノード v の次数を表す。グローバルクラスタリング係数の計算式は以下で定義される、

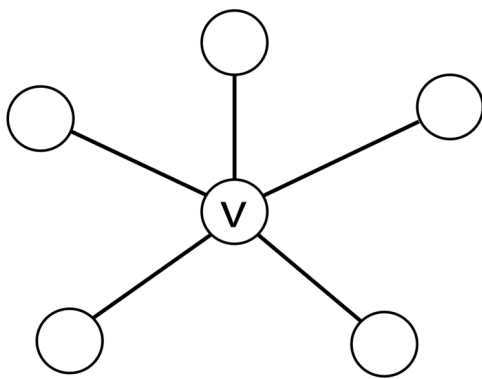
$$GCC = \frac{ClosedTriplets}{TotalTriplets}, \quad (2.4)$$

ここで、

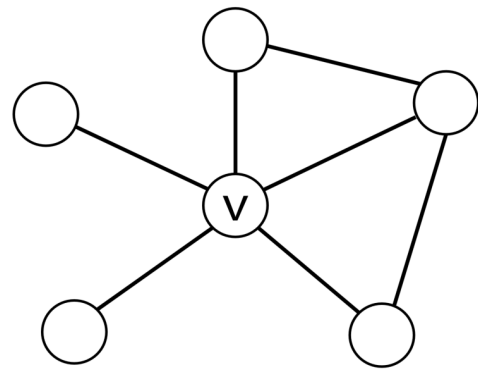
$$ClosedTriplets = 3 * \sum_{v \in G} r_v, \quad (2.5)$$

$$TotalTriplets = \sum_{v \in G} \frac{d(d_v - 1)}{2}. \quad (2.6)$$

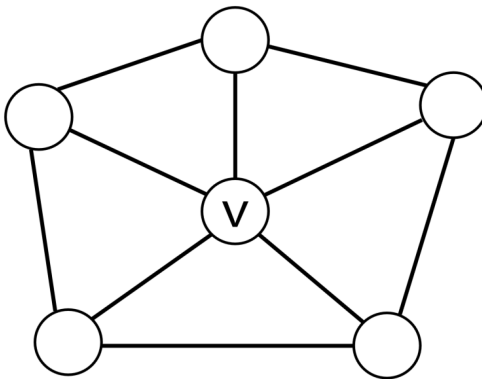
例えば、図 2.3(a) では $Triangle = 0$ 、 $LCC_v = 0$ 、(b) では $Triangle = 3$ 、 $LCC_v = 0.3$ 、(c) では $Triangle = 5$ 、 $LCC_v = 0.5$ 、(d) では $Triangle = 10$ 、 $LCC_v = 1$ となる。



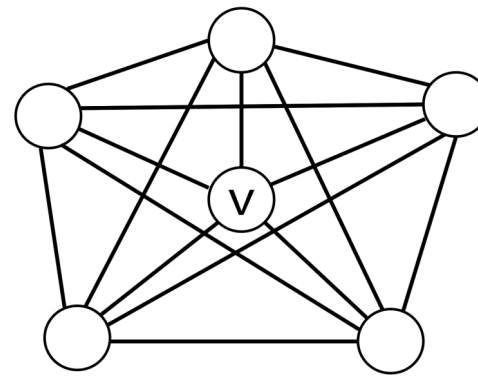
(a)



(b)



(c)



(d)

図 2.3: クラスタリング係数の例

2.2.3 ネットワーク上のパーコレーション

ランダムな木状のネットワークについて、次数相関がない場合、そのようなネットワークには巨大な連結成分が存在することが示され、全体が繋がるランダムな占有ノードの割合としてのパーコレーション閾値は次の式で与えられる [18]。

$$p_c = \frac{1}{g'_1(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \quad (2.7)$$

ここで、 $g'_1(1)$ は余剰次数分布に対応する母関数であり、 $\langle k \rangle$ はネットワークの平均次数、 $\langle k^2 \rangle$ は $\sum_{k=kmin}^{kmax} k^2 P(k)$ 、 $kmin$ と $kmax$ はそれぞれ最小と最大の次数である。

クラスタリングが低い、すなわち $0 < C \ll 1$ のネットワークの場合は、上記の臨界点は $(1 - C)^{-1}$ によってスケーリングされるため、以下に近似される [18]：

$$p_c = \frac{1}{1 - C} \frac{1}{g'_1(1)}. \quad (2.8)$$

上式は、ある次数分布に対して、クラスタリングがより大きなパーコレーション閾値をもたらすことを示している。言い換えると、その大きな閾値は、固定されたリンク数に対して、クラスタリング構造がネットワークの連結性を弱体化させることに相当する。但し、高いクラスタリングを持つネットワークの場合は、中心核と周リンクの core-periphery 構造を引き起こす可能性があるため、core と periphery が異なる臨界点でパーコレートする可能性があり、上記の近似的な処理は適用されない [19] ことに注意されたい。

2.2.4 近接中心性 (closeness centrality)

連結されたネットワークにおいて、あるノードの近接中心性は、中心性の尺度の一つであり、そのノードと他のノードとのホップ数で測った最短経路の長さの合計の逆数として計算される。したがって、あるノードがより中心的であるほど、他のすべてのノードに近くなる。

近接中心性は、Bavelas (1950) によって距離の逆数として定義される [20, 21]。

$$C_c(x) = \frac{1}{\sum_{y \in V} d(y, x)}. \quad (2.9)$$

ここで、 $d(y, x)$ はノード x と y の最小ホップ数で測った距離（最短経路の長さ）を表す。上記は標準化されていない近接中心性である。通常、その標準化された形式に言及し、最短経路の合計ではなくその平均値を表すような前の式に $N - 1$ をかけた以下の形式が広く用いられている。

$$C(x) = \frac{N - 1}{\sum_{y \in V} d(y, x)}. \quad (2.10)$$

例として図 2.4 では、各ノードの横に書かれた数字は、そのノードと赤い四角いノードとの間の最短経路の長さに基づいており、そのノードからの距離を表す。緑色のリンクは、赤い四角形のノードと赤い円形のノードとの間の最短経路のうちの 1 つを示す。したがって、赤い四角形のノードの近接中心性は $5/(1+1+1+2+2) = 5/7$ となる。

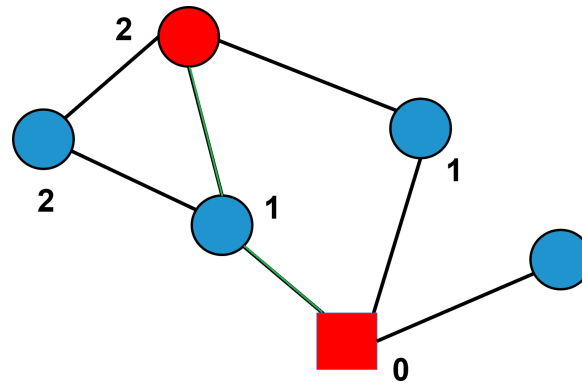


図 2.4: 近接中心性の例。

2.2.5 媒介中心性 (betweenness centrality)

まず説明前に、媒介数 (Betweenness) とは何かを知る必要がある。媒介数は、リンク介数とノード介数の 2 種類に分類される。

リンク介数は、ネットワーク内のあるリンクが、すべての最短経路の中で現れる回数を指す。つまり、このリンクがどの程度の最短経路の一部であるかを表す。通常、リンク介数が高いリンクほど重要であると考えられている。なぜなら、こ

これらのリンクは異なるノード間の情報伝搬においてより重要な役割を果たすからである。

ノード介数とは、ネットワーク内のあるノードが全ての最短経路で出現する頻度のことであり、最短経路においてそのノードが中継点としてどのくらい機能するかを表す。通常、ノード介数が高いノードは重要度が高いと考えられている。なぜなら、これらのノードは異なるノード間の情報伝搬においてより重要な役割を果たしているからである。

媒介数は、対応するノードまたはリンクがネットワーク全体でどのような役割や影響力を持っているかを反映する重要なグローバル指標であり、現実的な意味を持っている。社会関係ネットワークや技術ネットワークでは、媒介数の分布特性が、例えば、対応する生産関係における異なる人員、リソース、および技術の地位を反映する [22, 23]。これは、重要なリソース、技術、および人材を発見し、保護するために重要である。

媒介中心性は、1977年に Freeman によって提案され [22]、情報フローに対するグラフ上の影響力が大きいノードを見つけることができ、値が大きいほどそのノードが情報フローに与える影響も大きいことを示す。あるノードの媒介中心度は、ネットワーク内の各ノード対の最短経路数を分母とし、その中で、ノードを通過する最短経路数を分子として合計する。

$$C_b(v_i) = \sum_{i \neq j \neq v_i, i < j} \frac{g_{ij}(v_i)}{g_{ij}}. \quad (2.11)$$

ここで、 g_{ij} はノード i からノード j への最短経路の数を表し、 $g_{ij}(v_i)$ はノード i からノード j へのノード v_i を通る最短経路の数を表す。

次数中心性に似た性質を持つ、媒介中心性の値はネットワークのサイズに依存する。しかし、ネットワークのサイズに関係なく、相対値を計算するためにはネットワークサイズの影響を排除する必要があり、最大媒介中心性を計算することが一般的である。この時、ノード v_i が任意のノード対 (v_i, v_j) の全ての最短経路に含まれる場合、そのノードの媒介中心性は最大値 1 になる。したがって、最大媒

介中心性は、

$$C_b(v_i) = \sum_{i \neq j \neq v_i, i < j} \frac{g_{ij}(v_i)}{g_{ij}} = \sum_{i \neq j \neq v_i, i < j} 1 = \frac{2 \binom{n-1}{2}}{2} = \frac{(n-1)(n-2)}{2}. \quad (2.12)$$

となるので、標準化された媒介中心性は以下で定義される。

$$C'_b(v_i) = \frac{2C_b(v_i)}{(n-1)(n-2)}. \quad (2.13)$$

2.2.6 次数相関 (Degree-degree correlations)

次数相関とは、隣接するノード間の次数の相関のことであり、Pearson の次数相関をリンク両端のノードの次数に対して適用することで、以下の式で定義する [24]。

$$r = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2M) k_i k_j}{\sum_{ij} (k_i \delta_{i,j} - k_i k_j / 2M) k_i k_j} \quad (2.14)$$

ここで、 $\delta_{i,j}$ は Kronecker のデルタであり、 $i = j$ のときに 1、 $i \neq j$ のとき 0 になる。また、式変形することで次式 (2.4) が得られる [24]。

$$r = \frac{S_1 S_e - S_2^2}{S_1 S_3 - S_2^2} \quad (2.15)$$

ここで、

$$S_1 = \sum_i k_i, S_2 = \sum_i k_i^2, S_3 = \sum_i k_i^3, S_e = \sum_{ij} A_{ij} k_i k_j \quad (2.16)$$

である。

上式で計算される次数相関 r の範囲は $[-1, 1]$ となる、 r が負値のときは、次数が異なるノード同士が隣接しやすいことを示し、 r が正值のときは、次数が同じノード同士が隣接しやすいことを示す。

2.2.7 頑健性 (Robustness)

ネットワークへの攻撃に対する結合耐性を評価する指標として、以下のような頑健性指標 R が広く用いられている [25]。頑健性指標 R は、ノードを順に除去し

たとき、ネットワークの連結性がどの程度まで保存されているかを評価する。具体的には、除去率 q % のノードを除去した場合、最大連結成分比 $S(q)/N$ の総和を頑健性指標 R として評価する (q % のノードを削除して、ネットワークが様々な島に分断されたとしても、その中で一番大きい島が最大連結成分である)。以下に頑健性指標 R の定義式 (2.6) を示す。

$$R = \left(\sum_{q=1/N}^1 S(q)/N \right) / N \quad (2.17)$$

ここで、 N はネットワークの全体のノード数で、 $S(q)$ はノード除去率 q の時の最大連結成分 (GC:Ginat Component) の大きさを、 $\sum_{q=1/N}^1$ は除去率の割合 $1/N$ から $2/N, 3/N, \dots, N/N = 1$ までの和を表す。 $S(q)/N$ が最大連結成分比になり、除去率 q の範囲は $[0,1]$ である。つまり、ノードを一つずつ順番に取り除いて、その $S(q)/N$ を平均したものである。除去ノードの攻撃法として、次数が高いノード (ハブ) から順に除去する、ハブ攻撃 [25] がよく使われている。図 2.5 に、Hub 攻撃による最大連結成分サイズの崩壊の例を示す。

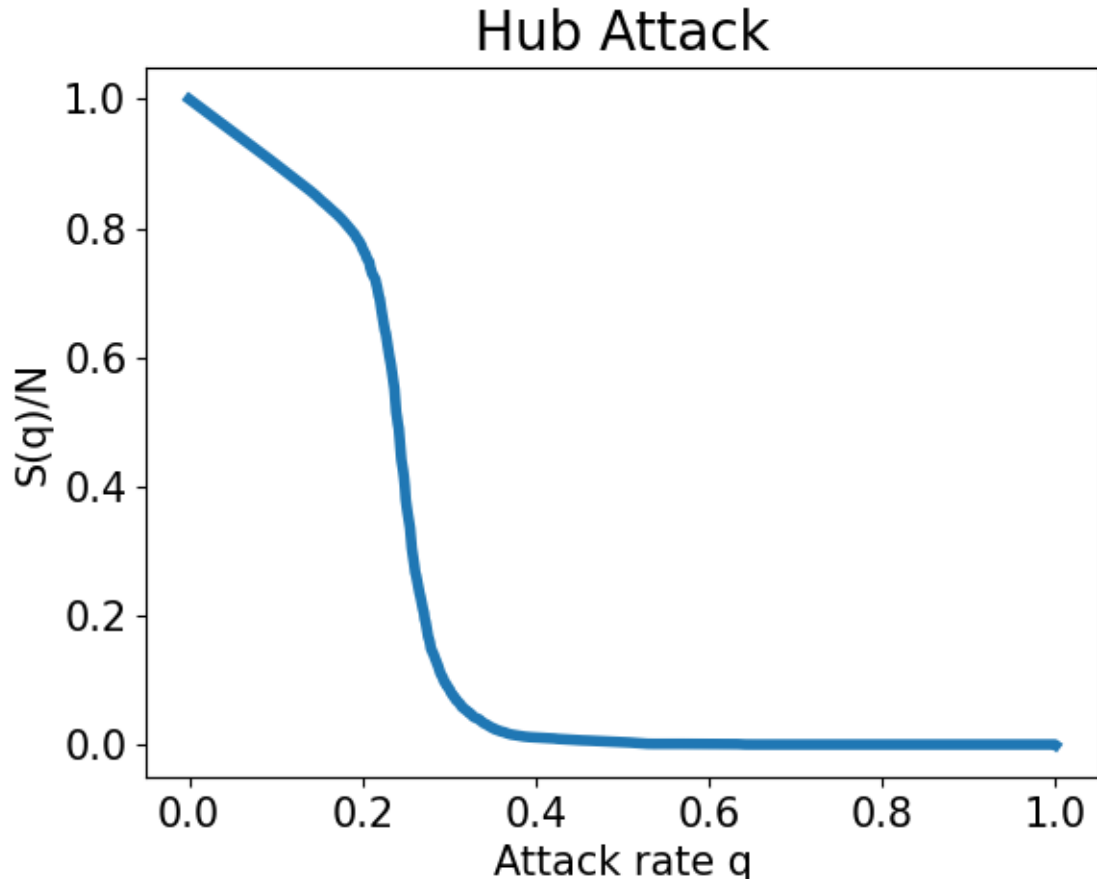


図 2.5: 次数順攻撃に対する最大連結成分サイズの減少。一つのネットワークに対して Hub 攻撃を適用した結果を示している。 R 値は各線の下側領域面積に相当する。

2.2.8 通信効率 (Efficiency)

通信効率 (Efficiency) は、ネットワークにおいて2つのノード間の最短経路長の逆数の和で定義される [26]。通信効率が高いほど、ネットワーク内のノード同士が効率的に情報を伝達できることを示す。具体的には、通信効率は以下の式で定義される。

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{L_{ij}}, \quad (2.18)$$

ここで、 N はノード数、 L_{ij} はノード i と j の間の最短経路長を表す。通信効率は、ネットワークの構造をより深く理解するために広く用いられる指標の1つである。

2.3 Feedback Vertex Set (FVS)

あるグラフ $G = (V, E)$ から、一部のノードを削除した、残りのグラフ G' において、ループが存在しない状態まで、削除したその最小除去ノード集合は Feedback Vertex Set (FVS) と呼ばれる。しかしながら、FVS を探すのは NP 困難な組み合わせ問題 [13] であり、多項式時間で解くアルゴリズムが存在しない。そこで、厳密に最適解を求めるのが非常に困難であるため、近似アルゴリズムが適用される。例えば、**コンピューター科学における 2-近似アルゴリズム**がある。これは、Vijay V. Vazirani が 2001 年に書籍「近似アルゴリズム」に紹介された方法で、最適解の高々 2 倍の近似値が求められる [27]。

下記の Algorithm 1 は、「近似アルゴリズム」[27] から転載した手順を示す。ここで、入力はグラフ G と重み関数 ω であり、出力は FVS に相当した F_0 である。

Algorithm 1 Approximation Algorithm of Feedback Vertex Set

Input: Given an undirected graph $G=(V,E)$ and a function ω assigning nonnegative weights to its vertices.

Output: F_0

```

1:  $H \leftarrow G, \omega_{prime} \leftarrow \omega, i \leftarrow 0$ 
2: for all  $H$  is not acyclic do
3:    $c \leftarrow \min_{u \in H} \left\{ \frac{\omega'(u)}{\delta_H(u)} \right\}$ 
4:    $G_i \leftarrow H, t_i \leftarrow c \delta_{G_i}, \omega' \leftarrow \omega' - t_i$ 
5:    $H \leftarrow$  the subgraph of  $G_i$  induced by vertices  $u$  with  $\omega'(u) > 0$ 
6:    $i \leftarrow i + 1$ 
7: end for
8:  $F_k \leftarrow \emptyset$ 
9:  $i = k, \dots, 1$  extend  $F_i$  to a feedback vertex set  $F_{i-1}$  of  $G_{i-1}$  by adding a minimal
   set of vertices from  $V_{i-1} - V_i$ 
10: return  $F_0$ .
```

一方、Algorithm 2 は、**統計物理 Cavity 法に基づく Belief Propagation (BP) 法**である [28]。

Algorithm 2 Belief Propagation Algorithm of Feedback Vertex Set

Input: Given an undirected graph $G=(V,E)$

Output: F_0

- 1: $H \leftarrow G$
 - 2: $qi \leftarrow$ the probability of node becoming FVS node.
 - 3: **for all** H is not acyclic **do**
 - 4: candidate $\leftarrow$ a node who has the biggest qi
 - 5: $G' \leftarrow$ delete the candidate from H
 - 6: $H \leftarrow G'$
 - 7: $i \leftarrow i + 1$
 - 8: **end for**
 - 9: **return** F_0 .
-

第3章 最小頂点被覆問題 (Minimum Vertex Cover)

本章の3.1節は、最小頂点被覆問題を定義する。3.2節は、それに対する2-近似アルゴリズムについて、3.3節は、統計物理のアプローチからの Survey Propagation (SP) 法を説明する。3.4節は、SP法における初期値やラウンド数の非依存性を数値例を用いて示す。3.5節は、SP法と2-近似法で推定した最小頂点被覆の大きさを予備実験として比較する。以下、3.1節と3.2節ではグラフ理論の流儀にならい、ノードとリンクを頂点と辺と呼ぶ。

3.1 最小頂点被覆問題とは

最小頂点被覆問題 (最小 VC) は、グラフ理論において重要な概念の1つであり、あるグラフ $G = (V, E)$ において全てのリンクのうち少なくとも片方のノードが被覆されるノードの集合を指す。つまり、その集合に含まれるノードが全てのリンクを被覆しているということである。この概念は、最適化問題の1つとしても扱われ、最小 VC は NP 困難問題として知られている [14]。最小 VC は、様々な分野で応用されており、例えば、バイオインフォマティクス [29]、ソーシャルネットワーク分析 [30] などに利用されている。

無向グラフ $G = (V, E)$ の頂点被覆 V' は、 $(u, v) \in E$ ならば $u \in V'$ 又は $v \in V'$ を満たす V の部分集合である。つまり、頂点被覆 V' は、すべてのリンクの少なくとも一方の端点が V' に含まれ、このような集合は G のリンクを被覆すると言われる。図 3.1 は、いくつかの頂点被覆 V' が赤で印された2つの頂点被覆の例を示している。

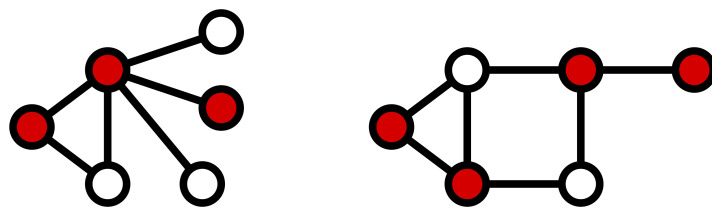


図 3.1: VC の例

最小 VC

最小 VC とは、与えられたグラフにおいて、最小の $|V'|$ となる頂点被覆 V' を見つける最適化問題である。

入力 グラフ G

出力 G が頂点被覆 k を持つ場合、最小の数 k

もし問題が決定問題の場合、それは頂点被覆問題 (Vertex Cover Problem) と呼ばれる。

入力 グラフ G と正の整数 k

出力 G がサイズが k 以下の頂点被覆を持つか？

上記の頂点被覆の決定問題は NP 完全問題であり、Karp の 21 の NP 完全問題の 1 つである [13]。厳密に最適解を求めるのが非常に困難であるため、近似アルゴリズムが適用される。以下、3.2 節と 3.3 節で述べる二つの近似法を試し、予備実験として 3.4 節にて比較してみた。

3.2 コンピューター科学における 2-近似アルゴリズム

リンクの両端点を順に頂点被覆に追加し、グラフから削除することを繰り返すことで、近似比 2 の解を得ることができる。

まず、貪欲アルゴリズムによって最大マッチング M を見つけ、 M に含まれるリンクの両端点から構成される頂点被覆 C を作成する。ここで、マッチングとは、お互いに端点を共有しないリンクの集合で、その最大リンク数である。以下の図 3.2 では、最大マッチング M は赤で、頂点被覆 C は青で示されている。

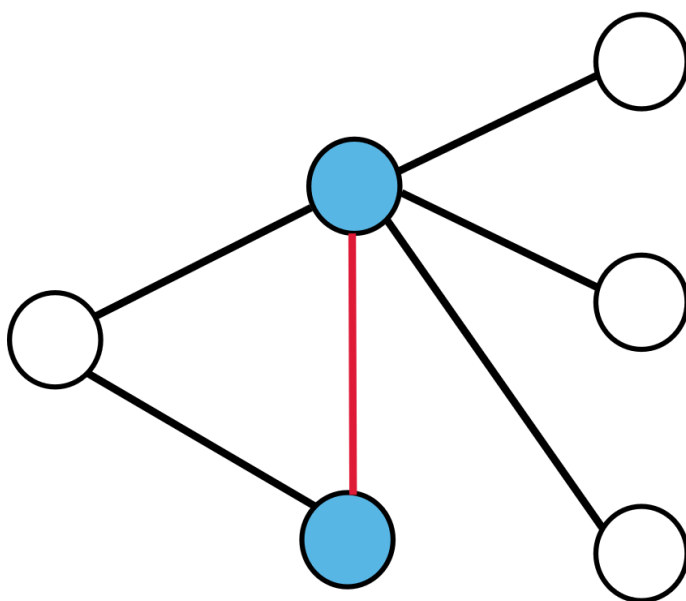


図 3.2: 貪欲アルゴリズム VC の例

この方法で構築された集合 C は頂点被覆であり、リンク e が C によって被覆されていないとすると、 $M \cup \{e\}$ はマッチングであり、 $e \notin M$ であり、これは M が極大であるという仮定に矛盾する。さらに、 $e = \{u, v\} \in M$ である場合、最適な頂点被覆を含む任意の頂点被覆は u または v (または両方) を含まなければならず、そうでない場合、リンク e は被覆されない。つまり、最適な頂点被覆は M 内の各リンクの少なくとも 1 つの端点を含み、全て片方の端点の場合を考えると、集合 C は最適な頂点被覆の 2 倍以下になることが分かる。

上記は、Vijay V. Vazirani が 2001 年に書籍「近似アルゴリズム」で紹介した方法で、最適解の高々 2 倍の近似値が求められる [27] のである。

Algorithm 3 は、「近似アルゴリズム」[27] から転載した手順を示す。入力グラフ G であり、出力は VC に相当した C である。

Algorithm 3 Approximation Algorithm of Vertex Cover Problem

Input: Given an undirected graph $G=(V,E)$.

Output: C

- 1: $C = \emptyset, E' = E$
 - 2: **for all** $E' \neq \emptyset$ **do**
 - 3: E' の任意のリンク (u, v) を考える。
 - 4: $C = C \cup \{u, v\}$
 - 5: E' から u または v に隣接する全てのリンクを取り除く。
 - 6: **end for**
 - 7: **return** C .
-

3.3 統計物理のアプローチ Survey Propagation 法

このアルゴリズムは、統計物理学の Cavity 法に基づく Survey Propagation (SP) という方法 [14] で、各ノード i は 3つの状態を持つ。1は被覆された状態、0は被覆されなかった状態、そして、*は時々被覆されたりされなかったりする状態（ジョーカー状態）である。ジョーカー状態*は状態0と1の間に位置することに注意されたい。ジョーカー状態を導入することによって拡張された探索空間を調整できる。これが SP と呼ばれる理由である。図 3.3 に示すように、これらの確率は、それぞれ状態 1 を表す $\hat{\pi}_{j \rightarrow i}^{(1)}$ 、状態 0 を表す $\hat{\pi}_{j \rightarrow i}^{(0)}$ 、ジョーカー状態 * を表す $\hat{\pi}_{j \rightarrow i}^{(*)}$ で表記される。

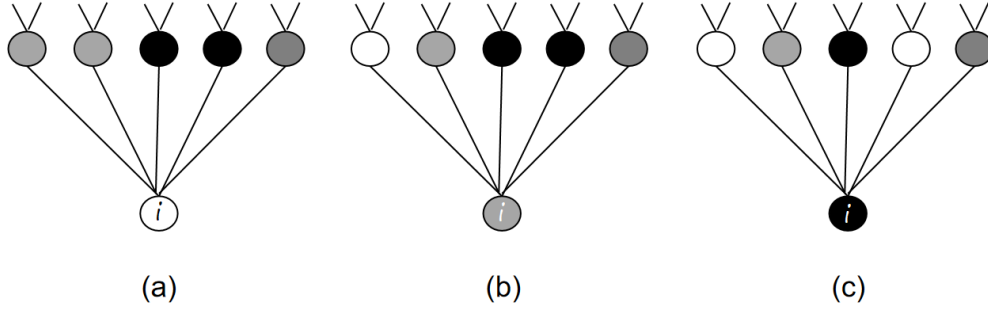


図 3.3: SP 法で 3つの異なる状態を確率的に決定する方法。0、*、1 は、白色、灰色の、黒色の丸で表される。下部のノード i の状態は、上部の隣接ノード ∂i の状態によって決定される。(a) ∂i に白色の丸がない場合、下部のノード i は被覆する必要がなく、白色（状態 0）になる。(b) ∂i に白色の丸が 1つある場合、下部のノード i は灰色（ジョーカー状態）になる。(c) ∂i に白色の丸が 2つ以上ある場合、下部のノード i は常に被覆される状態（状態 1）で黒色になる。

最小 VC を推定するメッセージ伝搬について説明するにあたり、4.2 節にて後述する情報伝搬における SIR モデル [31] とは異なることに注意する必要がある。各ノード i に対して、メッセージ伝搬の方程式 [14] は以下のように与えられる。

$$\hat{\pi}_i^{(0)} = C_i^{-1} \prod_{j \in \partial i} (1 - \hat{\pi}_{j \rightarrow i}^{(0)}), \quad (3.1)$$

$$\hat{\pi}_i^{(*)} = C_i^{-1} e^{-y} \sum_{j \in \partial i} \hat{\pi}_{j \rightarrow i}^{(0)} \prod_{j' \in \partial i \setminus j} (1 - \hat{\pi}_{j' \rightarrow i}^{(0)}), \quad (3.2)$$

$$\hat{\pi}_i^{(1)} = C_i^{-1} e^{-y} \left[1 - \prod_{j \in \partial i} (1 - \hat{\pi}_{j \rightarrow i}^{(0)}) - \sum_{j \in \partial i} \hat{\pi}_{j \rightarrow i}^{(0)} \prod_{j' \in \partial i \setminus j} (1 - \hat{\pi}_{j' \rightarrow i}^{(0)}) \right], \quad (3.3)$$

ここで、 $\partial i \setminus j$ はノード i の最近傍ノード集合で、ノード j を含めないものを表す。また、 e^{-y} は VC のサイズを最小化するためのペナルティ因子であり、 y は逆温度パラメータである。規格化定数は、全てのメッセージの積の逆数として与えられる。

$$C_i = e^{-y} \left[1 - (1 - e^y) \prod_{j \in \partial i} (1 - \hat{\pi}_{j \rightarrow i}^{(0)}) \right]. \quad (3.4)$$

また、各リンク $i \rightarrow k$ に対して、確率は以下の式で与えられる。

$$\hat{\pi}_{i \rightarrow k}^{(0)} = C_{i \rightarrow k}^{-1} \prod_{j \in \partial i \setminus k} (1 - \hat{\pi}_{j \rightarrow i}^{(0)}), \quad (3.5)$$

$$C_{i \rightarrow k} = e^{-y} \left[1 - (1 - e^y) \prod_{j \in \partial i \setminus k} (1 - \hat{\pi}_{j \rightarrow i}^{(0)}) \right]. \quad (3.6)$$

式 (3.3) から式 (3.6) は、 N 個のノードのランダム置換順に左リンクから右リンクへの T 回のラウンドの繰り返しによって計算される。ある程度の収束後、Decimation として [14]、 $\hat{\pi}_i^{(1)}$ が最大のノード i が VC として選択される。その後、ノード i が除かれ、上記を再計算する、以下の処理が行われる。

ステップ 1 式 (3.3)～(3.6) により、ノード i の確率 $\hat{\pi}_i^{(1)}$ を計算し、最小 VC を推定する。

ステップ 2 Decimation として、 $\hat{\pi}_j^{(1)}$ が最も高いノード j が VC ノードとして選択される。選択されたノード j と孤立したノード (Degree は 0 のノード) がネットワークから削除される。頂点被覆集合の数は $|VC| \leftarrow |VC| + 1$ と更新される (最初は $|VC| = 0$ と設定されている)。

ステップ 3 すべてのノードがネットワークから削除されるまで、ステップ 1 とステップ 2 を繰り返す。最終的に、複数の VC ノードのサイズは $|VC|$ として得られる。

3.4 初期値やラウンド数の非依存性

上記のアルゴリズムでは、各ラウンド数 T で、全てのノードが更新されるまでに、ランダムな順番で一つずつ更新される。以下では、ランダムな初期値とラウンド数 T を変更して、推定された $|VC|$ にどのような影響を与えるかを数値的に調べた結果を示す。表中 $b-a-c$ に対して、 a は違う Rand 及びラウンド数 T から得られた結果の同じ部分のノード個数、 b と c は違う部分のノード個数である。全ての表から VC の個数は大体同じである。表 3.1 から表 3.4 までに、初期値 Rand を変更して、 $|VC|$ はほとんど同じである。また、各初期値から探す VC ノードは大体同じである。表 3.5 から表 3.8 までに、ラウンド数 T を変更する。しかし、 $|VC|$ はほとんど同じである。また、各ラウンド数 T から探す VC ノードも大体同じである。すなわち、この統計物理学の VC アプローチはうまく収束した、ラウンド及び初期値は最終的な結果に影響を与えない。

	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	VC
Rand1		15—147—15	6—156—6	15—147—15	5—157—5	15—147—15	162
Rand2			16—146—16	5—157—5	15—147—15	5—157—5	162
Rand3				14—148—14	5—157—5	16—146—16	162
Rand4					15—147—15	6—156—6	162
Rand5						13—149—13	162
Rand6							162

表 3.1: Maaya 上の VC 初期値変更。

	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	VC
Rand1		46—419—47	54—411—55	67—398—69	43—422—42	45—420—44	465
Rand2			45—421—45	64—402—65	48—418—46	47—419—45	466
Rand3				57—409—58	55—411—53	53—413—51	466
Rand4					62—405—59	63—404—60	467
Rand5						31—433—31	464
Rand6							464

表 3.2: Hamster 上の VC 初期値変更。

	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	VC
Rand1		60—472—60	50—482—49	61—471—62	61—471—62	60—472—60	532
Rand2			59—473—58	58—474—59	58—474—59	56—476—56	532
Rand3				55—476—57	57—474—59	62—469—63	531
Rand4					56—477—56	63—470—62	533
Rand5						58—475—57	533
Rand6							532

表 3.3: OpenFlight 上の VC 初期値変更。

	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	VC
Rand1		3—354—3	3—354—3	3—354—3	4—353—4	5—352—5	357
Rand2			5—352—5	3—354—3	3—354—3	4—353—4	357
Rand3				3—354—3	5—352—5	5—352—5	357
Rand4					4—353—4	4—353—4	357
Rand5						1—356—1	357
Rand6							357

表 3.4: Polblogs 上の VC 初期値変更。

T	5	10	20	50	100	200	VC
5		6—156—6	7—155—7	6—156—6	6—156—6	6—156—6	162
10			4—158—4	5—157—5	6—156—6	7—155—7	162
20				6—156—6	6—156—6	6—156—6	162
50					6—156—6	5—157—5	162
100						6—156—6	162
200							162

表 3.5: Maaya 上の VC ランダム処理。

T	5	10	20	50	100	200	VC
5		46—419—47	54—411—55	67—398—69	43—422—42	45—420—44	465
10			45—421—45	64—402—65	48—418—46	47—419—45	466
20				57—409—58	55—411—53	53—413—51	466
50					62—405—59	63—404—60	467
100						31—433—31	464
200							464

表 3.6: Hamster 上の VC ランダム処理。

T	5	10	20	50	100	200	VC
5		61—471—61	57—475—56	63—469—63	58—474—57	58—474—58	532
10			67—465—66	62—470—62	51—481—50	61—471—61	532
20				61—470—62	61—470—61	64—467—65	531
50					55—477—54	61—471—61	532
100						56—475—57	531
200							532

表 3.7: OpenFlight 上の VC ランダム処理。

T	5	10	20	50	100	200	VC
5		5—352—5	6—351—6	4—353—4	5—352—5	5—352—5	357
10			4—353—4	5—352—5	5—352—5	5—352—5	357
20				5—352—5	5—352—5	5—352—5	357
50					3—354—3	3—354—3	357
100						5—352—5	357
200							357

表 3.8: Polblogs 上の VC ランダム処理。

3.5 Survey Propagation(SP) による最小頂点被覆と 2-近似法による最小頂点被覆の近似解の比較

予備実験として、現実のネットワークである LastFM を用いた数値シミュレーションを通じて、3.2 節の 2-近似法と 3.3 節の SP 法による最小 VC の近似解を比較し、SP 法が最小 VC をより正確に推定できることを示す。ここで、LastFM は 2020 年 3 月に公開 API から収集された、アジア諸国の LastFM ユーザーのソーシャルネットワークである。ノードはユーザーであり、リンクは彼らの間の相互フォロー関係である [32]。

表 3.9 に示されるように、SP 法による解はほぼ最適であり、一方で 2 近似法 [33] による解は最適解のほぼ 2 倍の大きさと考えられる。SP 法は統計物理学的アプローチであり、2 近似法は最大 2 倍の大きさを保証するコンピュータ科学的アプローチであることを思い出されたい。さらに、SP 法と比較するために、Belief Propagation(BP) アルゴリズム [28] を適用して、Feedback Vertex Set から VC に還元してみた。但し、FVS を除去した後、ネットワークの残りは部分木になるので、図 3.4 に示すように、よく知られた方法 [1] を適用して、木を奇数層と偶数層に分割し、サイズが小さい偶数層または奇数層のどちらかを VC として選択する。表 3.9 に示されるように、SP 法によって推定された $|VC|$ は BP 法によって推定された $|VC|$ よりもわずかに優れている。逆温度パラメータを $y = 7$ とした場合、最小 VC の結果は最小サイズのものである。さらに、表 3.10 に示されるように、 $T = 50$ ラウンドが最小 VC の最良の結果を得ている。したがって、以降では、SP 法を $y = 7$ および $T = 50$ で適用する。

ITP y	0	0.5	1	2	3	5	7	10	11	12
$ VC $ by SP	3520	3510	3517	3510	3508	3511	3507	3511	3514	3508
$ VC / N$	0.462	0.460	0.461	0.460	0.460	0.461	0.460	0.461	0.461	0.460
$ VC $ by BP	3520	3514	3516	3519	3515	3523	3524	3522	3519	3517
$ VC / N$	0.462	0.461	0.461	0.461	0.461	0.462	0.462	0.462	0.462	0.641
$ VC $ by 2-approx	5498									
$ VC / N$	0.721									

表 3.9: 逆温度パラメータ y が異なる場合について、最小 VC に関する SP 法、BP 法と 2-近似アルゴリズムによる推定値の $|VC|$ を比較する。太字の数値は最小 VC における最良の結果を表している。

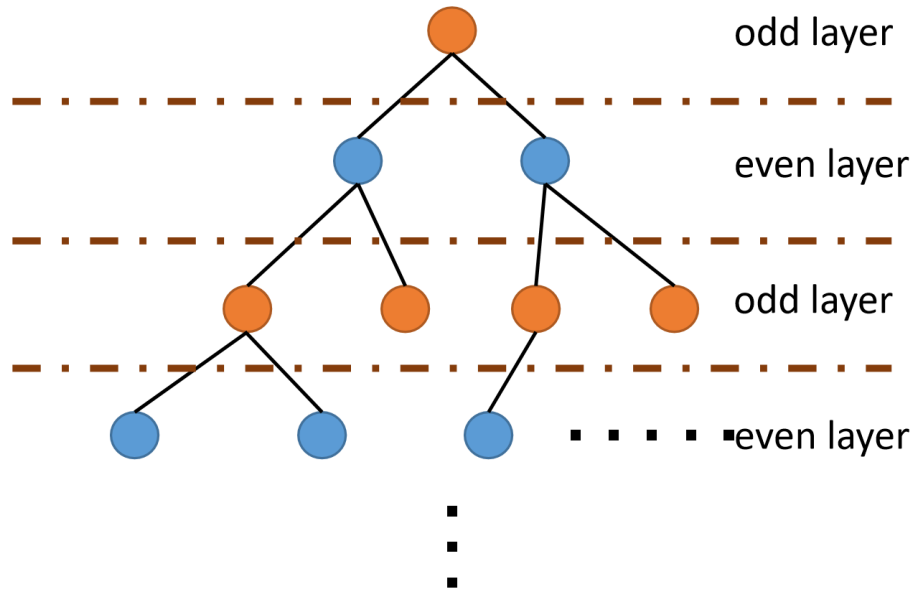


図 3.4: 木の中の VC を見つけるための割り当て法 [1]。

Round T	5	10	20	50	100	200
$ VC $ by the SP	3511	3512	3519	3508	3516	3516
$ VC / N$	0.461	0.461	0.462	0.460	0.461	0.461

表 3.10: LastFM における異なるラウンド数 T での SP 法による $|VC|$ 。ノード数 $N = 7624$ 。

第4章 影響力最大化問題における複数の種の探索

本章の4.1節では、影響力最大化問題（IMP）に関する独立カスケード（IC）モデルや線形閾値（LT）モデルについて紹介する。これらを既知であれば、読み飛ばしても差し支えない。4.2節では、多くのサンプル数を必要とする従来のSIRモデルに対して、高速SIRモデルを提案する。4.3節では、従来のIMP問題のヒューリスティック手法について説明する。ここでSIRとは、Susceptible-Infected-Recoveredモデルを指す[31]。

4.1 IMPとは

IMPは、ソーシャルネットワーク上で口コミ情報などの拡散の影響力が最大化されるように、ノードの部分集合である N_s 個の種ノードを見つける最適化問題であり、 $N_s \geq 1$ は定数である。この問題には、ウィルス性マーケティング[34, 35]、脳活性化[36]、コミュニティの情報伝播[37]、噂のブロック[38]、そして接触ネットワークにおけるグローバルな流行病の拡大を防止するための問題[17]など、多くの応用分野がある。IMPでは、ある種ノード個人から情報伝播をシミュレートするために拡散モデルが用いられる[11]。拡散モデルには以下の独立カスケード（IC）モデルや線形閾値（LT）モデルがある[39]。

4.1.1 独立カスケード（IC）モデル

ICモデルは、情報拡散モデルの一つであり、ノード u, v 間に対して伝染確率 $P_{u,v}$ を、相互作用の頻度、地理的近接性、または過去の感染履歴などに基づいて割り当てられる。各ノード u は感染すると、次の時間ステップでそのリンク (u, v) に関連付けられた確率 $P_{u,v}$ に基づいてその隣接ノード v を感染させる能力を持つもの

とする。独立カスケードモデルのダイナミクスでは、時間ステップ t ごとに、 $t-1$ 時刻で新たに感染したノードの集合 S^{t-1} に対して、各 $v \in S^{t-1}$ が確率 $P_{u,v}$ で未活性の隣接ノード ∂u を感染させる [39]。Kempe らは、IC モデルの下で IMP を、NP 完全の Set Cover 問題の一般的なケースと見なすことができることを示している [11]。また、特に、全てのリンク (u, v) で一定確率 $P_{u,v} = \beta$ の時、4.2 節の SIR は IC モデルの特別例である。

図 4.1、ネットワーク中のすべてのノードに対して、 $P_{w,v}=0.5$ であると仮定する。つまり、ノードが活性化すると、50%の確率でその隣接ノードが活性化する。まず、ノード 8 とノード 9 が最初の活性化ノードであると仮定する。ノード 8 と 9 から始まり、IC モデルは $P_{w,v}=0.5$ の値に基づいて、それらの隣接ノードを活性化する。図 4.2 から 4.4 は、IC モデルの条件下で情報拡散がどのように進展していくか示している。以下、より詳しくその動作を説明する。

図 4.1 の Step1 では、ノード 8 とノード 9 は、隣接するノード 5、6、7 を活性化しよ

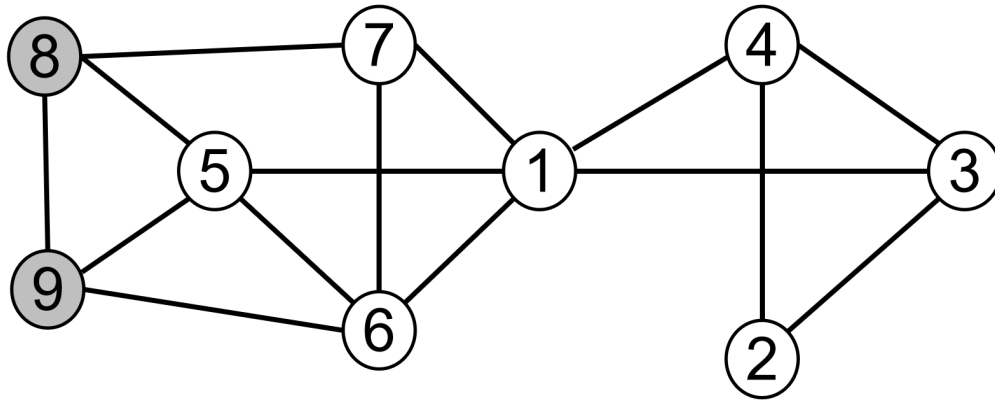


図 4.1: Step 0、ノード 8 とノード 9 が最初の活性ノードである。

うとする。ノード 5 とノード 7 のみが成功して活性化され、ノード 6 は活性化に失敗したとする。ノード 5 と 7 が活性化された場合、それらの隣接ノードを $P_{w,v}=0.5$ の確率で続けて活性化を行い、ノード 1 と 6 が活性化に成功したと仮定する。

次に、ノード 6 のすべての隣接ノードが既に活性化しているため、図 4.3 の Step2

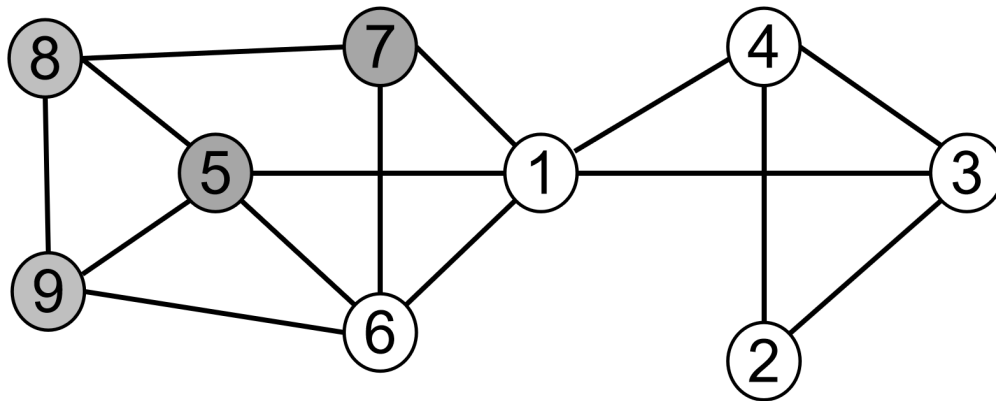


図 4.2: Step 1、ノード 1 と 6 が成功して活性化されたと仮定する。

では、ノード 1 の隣接ノードのみを考慮する。IC モデルに従い、ノード 3 とノード 4 は 50% の確率で活性化して、ノード 4 のみが成功的に活性化されたと仮定する。

図 4.4 の Step3 では、最近活性化されたノード 4 の隣接ノードが考慮される。ノード 4 の隣接ノードはどれも活性化されていないと仮定する。このように、ノード 1、4、5、6、7、8、9 が活性化されたときに、ネットワーク内の情報拡散プロセスは停止する。

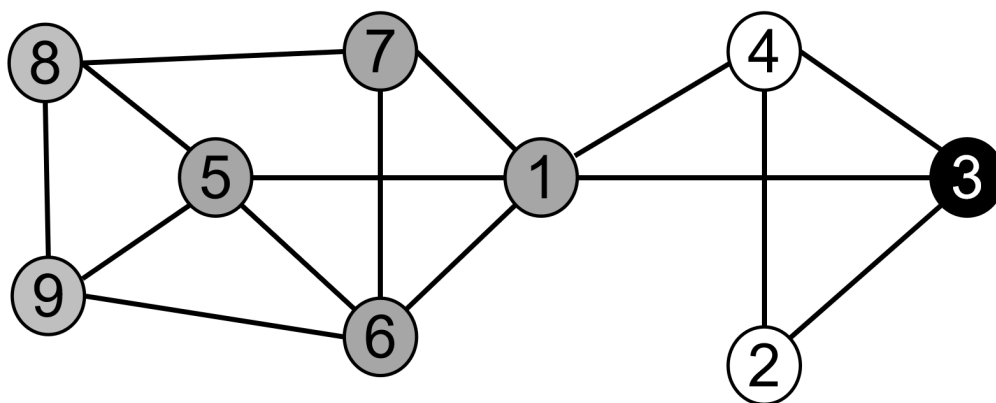


図 4.3: Step 2、ノード 1 と 6 が活性化された。

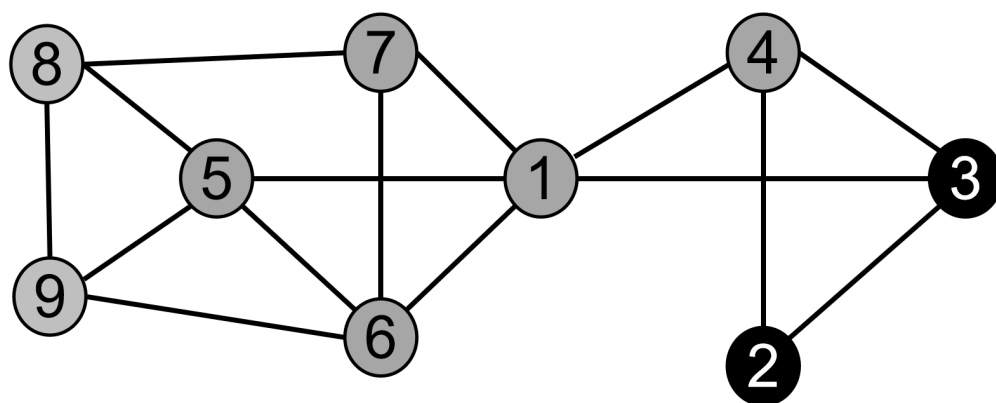


図 4.4: Step 3、ノード 4 の隣接ノードはどれも活性化されていないと仮定する。

4.1.2 線形閾値 (LT) モデル

LT モデルでは、任意のノード $u \in V$ に対して、受信リンク (u, v) の合計重みが 1 以下になるように、 $\sum_{u \in \partial v} b(u, v) \leq 1$ とする。モデルのダイナミクスは以下のように定義される。各ノード v は、 $[0, 1]$ の区間で一様ランダムに閾値 θ_v を選択する。次に、 H_{t-1} は時刻 $t-1$ またはそれ以前に活性化されたノードの集合、各非活性ノード v は、 $\sum_{u \in \partial v \cap H_{t-1}} b(u, v) \geq \theta_v$ であれば活性化する [39]。ここで、 ∂v は v の隣接ノードの集合である。LT モデルにおける IMP は NP 困難である [11]。ネットワークが有向グラフであると仮定し、 $b_{u,v} \neq b_{v,u}$ は異なる役割を持つ。簡略化するため、 $(b_{u,v}) = 1/k_v$, $(b_{v,u}) = 1/k_u$ と仮定し、各ノードのしきい値を 0.5 とする。ノード 8 とノード 9 が最初の活性化ノードであると仮定する。図 4.5 は、LT モデルの条件下で情報拡散がどのように進展していくかを示している。

このプロセスの最初のステップでは、ノード 5 は 2 つの隣接ノードが活性化して

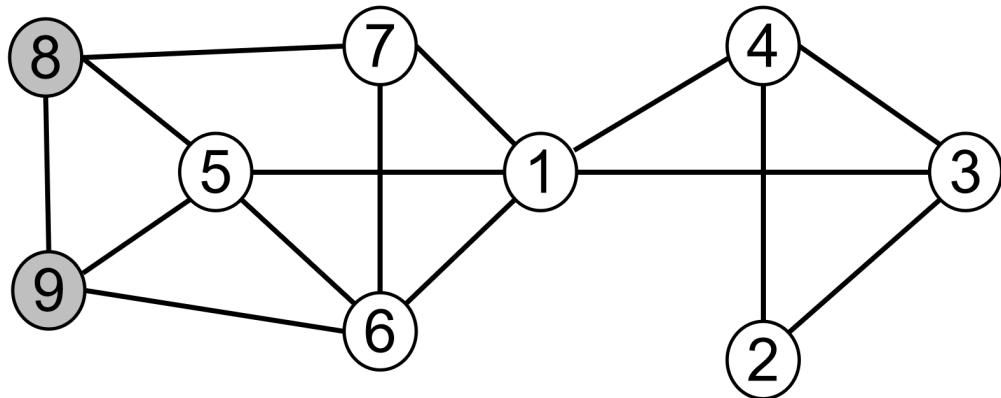


図 4.5: Step 0、ノード 8 とノード 9 が最初の活性化ノードである。

いるため、ノード 5 に向かう重み $b_{8,5} + b_{9,5} = 1/3 + 1/3 = 2/3$ である。2/3 は閾値 0.5 より大きいため、ノード 5 が活性化になる。

図 4.7 の Step2 では、ノード 6 の 2 つの隣接ノードであるノード 5 とノード 9 が活性化状態であるため、ノード 6 が活性化される。

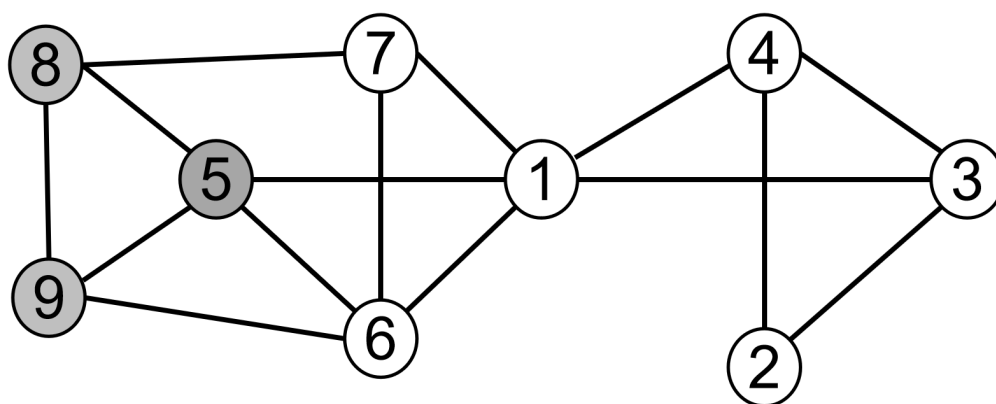


図 4.6: Step 1、ノード 5 が活性化する。

図 4.8 の Step3 では、隣接ノードの重みの合計が閾値 0.5 より高くなるため、ノード 7 とノード 1 が活性化される。ノード 1、5、6、7、8、9 が活性化されると、残りのネットワークのノードはすべて活性化の条件を満たせなくなるため、情報拡散プロセスは停止する。

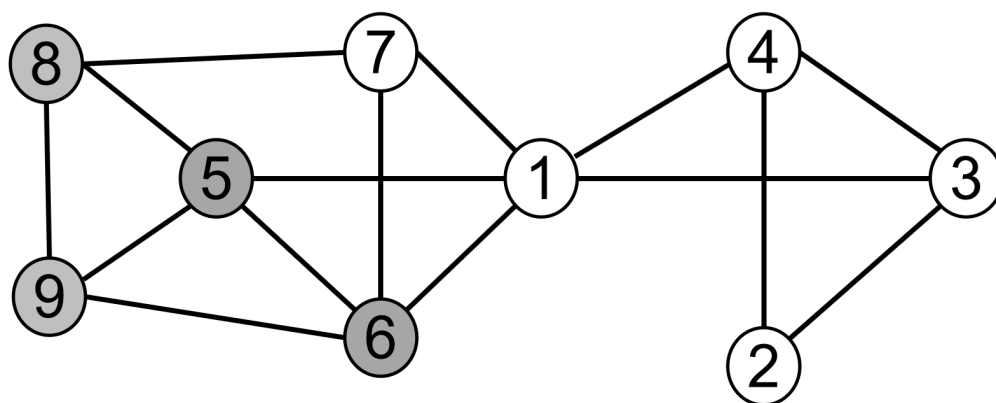


図 4.7: Step 2、ノード 6 が活性化する。

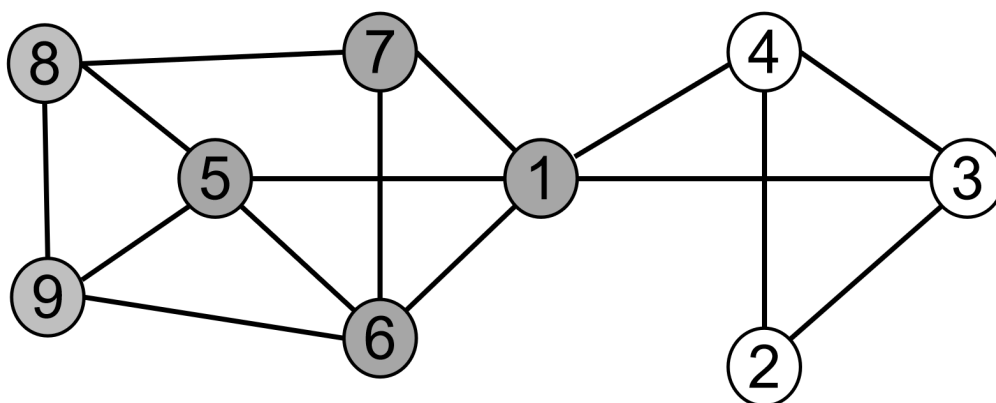


図 4.8: Step 3、ノード 7 と 1 が活性化する。

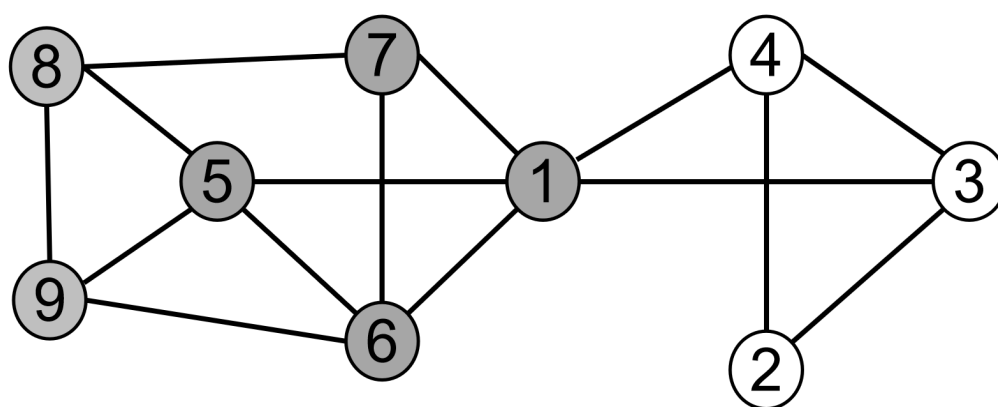


図 4.9: 最終段階

4.1.3 LT モデルと IC モデルの違い

IC モデルと LT モデルは、拡散能力や拡散方式などの特徴を一定程度反映している [39]。両モデルにはランダム性が導入されている。LT モデルアルゴリズムは、情報拡散を開始する前に、ノードにランダムに閾値を選択する。一方、IC モデルでは、ノードが隣接するノードを活性化する確率 $P_{w,v}$ である。しかしながら、両者には以下のように明らかな違いがある。

線形閾値 (LT) モデル

- 受信者主導 (receiver-centered) で、あるノードのすべての隣接ノードを観察し、そのノードの次数と総次数の割合がある閾値以上ならば活性化できるかどうかを決定することを意味する。
- ノードの活性化は、ノードのすべての隣接ノードに依存してる。
- LT モデルで閾値が一旦与えられると、ネットワーク内の情報拡散プロセスも確定する。

独立カスケード (IC) モデル

- 送信者主導 (sender-centered approach) で、ノードが活性化された後、そのノードは隣接するすべてのノードを活性化しようとする。
- 1つのノードが独立してそのすべての隣接ノードを活性化する（必ずしもすべてが活性化されるわけではない）。
- 情報伝播のプロセスは、確率的に変化するため、ネットワーク内の情報伝播プロセスは不確定になる。

4.2 拡散処理と高速化計算

一般的に SIR 感染モデル [31] は複数のランダム状態の初期値からの伝搬を平均化するために多くの試行が必要である。そこで、計算時間を短縮するために、本節は CI [40] のメッセージ伝搬方程式から着想して、新たな高速 SIR モデルを提案する。

4.2.1 従来の SIR モデル

一般的に、SIR 感染モデル [31] は、複数の種からの拡散処理を行うために適用され、感染症の流行予測 [41] や、予防・制御策の評価 [42] などに役立つ。このモデルは、Susceptible（感受性保持者）、Infected（感染者）、Recovered（回復者）の3つの状態を考慮している。感受性保持者は、まだ感染していない人々を指し、感染者はすでに感染している人々を指す。回復者は、病気から回復した人々を指し、彼らは再度感染することができない。SIR モデルは、感染者が感染を広げる速度（感染率 β ）と回復者が感染拡大を止める速度（回復率 γ ）を考慮している。

SIR モデルでは、各ノードが感受性保持者（S）、感染者（I）、および回復者（R）の3つ状態に分類される。各時間ステップにおいて、感染拡大プロセスでは、感染しているノードは確率 β で隣接ノードの状態を S から I に変更し、回復確率 γ で自身の状態を I から R に変更する。感受性保持 S は感受性保持 S と感染 I の積に比例して確率 β で感染 I になる。一方、回復率 γ で回復者になる。回復者は再度同じ病気に感染しない免疫を持つものとする。図 4.10 の時間発展を非線形常微分方程式で記述すれば、

$$\frac{dS(t)}{dt} = -\beta S(t)I(t), \quad (4.1)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t), \quad (4.2)$$

$$\frac{dR(t)}{dt} = \gamma I(t), \quad (4.3)$$

となる。

ここで、 $S(t)$ は時間 t における感受性保持者の数、 $I(t)$ は感染者の数、 $R(t)$ は回復者の数、一定数 $N = S(t) + I(t) + R(t)$ は総人口の数、 β は感染率、 γ は回復

率を表す。式 (4.1)、(4.2)、(4.3) の和は

$$\frac{d}{dt}(S(t) + I(t) + R(t)) = 0, \quad (4.4)$$

となり、この式は全人口の変化率がゼロであることを示しており、感染症が広がる期間において、人口全体の総和が変化しないことを意味する。感染者が増える一方で、感受性保持者と回復者が減少するため、これらの変化率の和がゼロになる。この式は、SIR モデルの基本的な性質を表している。

ここでは、確率的 SIR モデルの平均的な振る舞いについて考える [31]。拡散プロセスは、各時間ステップで、感染したノードは、例えば確率 $\beta = \lambda \frac{\langle k \rangle}{\langle k^2 \rangle}$ [43] で、その近隣ノードの状態を S から I に変更し、自身の状態を I から R に回復確率 $\mu = 1$ で変更する。ここで、 $\frac{\langle k \rangle}{\langle k^2 \rangle}$ は 2.2.3 節で説明したのパーコレーション値である。通常、複数の種からの拡散を行うために従来の SIR モデルが適用されるが、例え同じ種集合からでも、確率 β の確率的振る舞いのサンプリングの平均化には多数の試行が必要である。つまり、ネットワークのサイズが非常に大きい場合、従来の SIR モデルでは平均化に多くの時間が必要になる。

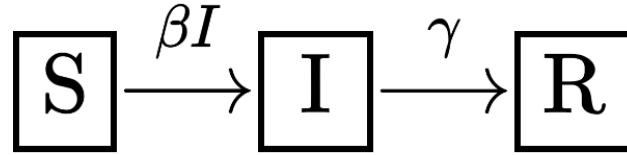


図 4.10: SIR モデルのフローチャート

4.2.2 高速化 SIR モデル

上記の計算時間を短縮するために、CI [40] のメッセージ伝搬方程式から着想して、以下の方程式を考慮する。

$$P_i^I(t+1) = P_i^S(t) \left[1 - \prod_{j \in \partial i} (1 - \beta P_j^I(t)) \right], \quad (4.5)$$

$$P_i^R(t+1) = P_i^R(t) + P_i^I(t), \quad (4.6)$$

$$P_i^S(t+1) = 1 - P_i^I(t+1) - P_i^R(t+1). \quad (4.7)$$

ここで、 $P_i^I(t+1)$ 、 $P_i^R(t+1)$ 、 $P_i^S(t+1)$ は、それぞれ時刻 $t+1$ におけるノード i の状態 I、R、S の確率を表す。すでに平均化された確率値 $P_i^I(t)$ 、 $P_i^R(t)$ 、 $P_i^S(t)$ は、時間ステップによって更新される。これを高速 SIR モデルと呼ぶ。これらのメッセージ伝搬方程式 (4.5)(4.6)(4.7) はすでに平均化された確率変数に基づいているため、確率的な振る舞いを平均化するために多数のサンプルが不必要となる。

4.2.3 従来の SIR モデル *vs* 高速 SIR モデル

この項は、従来の SIR モデルと高速 SIR モデルによる計算時間を予備実験として比較する。高速 SIR モデルと従来の SIR モデルは似たような拡散挙動を示す。図 4.11 においては、パーコレーション閾値 $\lambda \frac{\langle k \rangle}{\langle k^2 \rangle}$ ($\lambda = 1$) で、感染率 $\beta = 0.12$ の条件下で、種のサイズ N_s が 885 ($l = 2$)、516 ($l = 3$)、407 ($l = 4$) の 3 つのそれぞれに対して調べている。これらの l の値は、異なるサイズ N_s の種を与え、割合は、それぞれ $N_s/N = 0.12$ ($l = 2$)、0.07 ($l = 3$)、0.05 ($l = 4$) である。ここで、 $S(t) = \sum_{i=1}^N P_i^S(t)/N$ 、 $I(t) = \sum_{i=1}^N P_i^I(t)/N$ 、 $R(t) = 1 - S(t) - I(t)$ で定義する。

図 4.11(a)(b)(c) では、 $I(t)$ は単調増加してから、単調減少、最終的には 0 に収束する。 $R(t)$ は単調増加し、0.4 に収束する。 $S(t)$ は単調減少し、0.6 に収束する。円、四角、三角形印の黒い線は従来の SIR モデルの状態 S、I、R の確率を表す。円、四角、三角形印の赤い線は、高速 SIR モデルの状態 S、I、R の確率を表す。種のサイズが異なるにもかかわらず、高速 SIR と従来の SIR の各状態 S、I、R の赤と黒の線はほぼ一致している。さらに、 $R(t)$ は (a) で 0.4、(b) で 0.37、(c) で 0.35 に収束し、 $t > t_c$ (t_c : 収束時間で、すべての感染したノードが回復したときに定義される) である。さらに、 l が増加するにつれて、 t_c も徐々に増加する ($t_c = 5$ (a)、 $t_c = 7$ (b)、 $t_c = 8$ (c))。 $I(t) = 0$ のため、 $S(t) + R(t) = 1$ である。各状態 S、I、R の高速 SIR と従来の SIR の赤と黒の線がほぼ一致しているにもかかわらず、高速 SIR は (平均化が必要ないため) サンプル数倍速くなる。

また、収束時間 t_c の前 (初期の感染拡大期) には、高速 SIR と従来の SIR の赤色と黒色の線にはわずかに違いがある。図 4.11 の $t = 2$ から $t = 7$ までの赤色線と黒色線の差は、従来の SIR モデルのサンプルの中で最も高い $R(t)$ と最も低い $R(t)$ の間に隙間があるためである (表 4.1 参照)。言い換えると、差が現れる理由は、

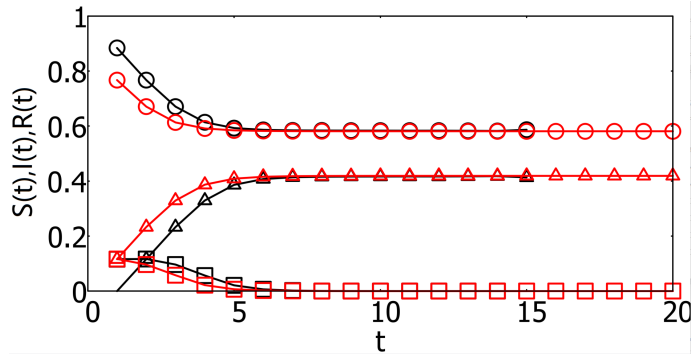
最低値のため従来の SIR モデルで $I(t)$ と $S(t)$ が過小評価されていることである。最低値と最高値の $R(t)$ の間の差は大きくないが、ネットワークサイズ $N = 7624$ のため、感染ノードの累積数 $N \times R(t)$ は十分大きくなる。さらに、表 4.2 には、高速 SIR モデルが従来の SIR モデルの 30 倍速いことが示されており、1000 サンプルは約 $30 \times T\text{-rounds}$ ($T = 50$) に相当する。速度向上率 (AVG-SIR の計算時間/高速 SIR の計算時間) は 20 から 30 までである。

t	2	3	4	5	6	7
lowest $R(t)$	0.11608	0.22258	0.31269	0.36791	0.39178	0.39821
highest $R(t)$	0.11608	0.24278	0.34693	0.40424	0.42195	0.42799
average $R(t)$	0.11608	0.23283	0.32923	0.38586	0.40707	0.41360
variance of $R(t)$	0	1.5401e-06	4.1998e-06	4.5865e-06	4.2386e-06	4.2224e-06

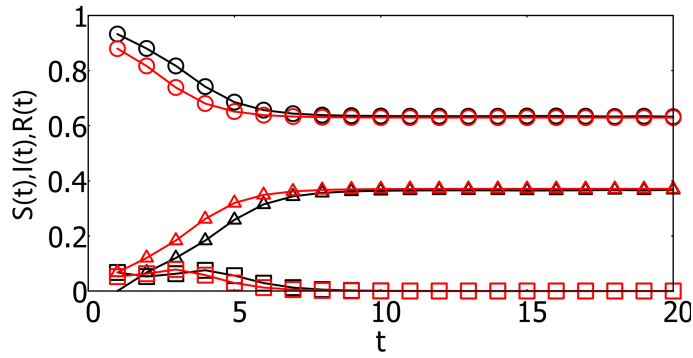
表 4.1: 従来の SIR モデルにおいて、拡散時間が 2 から 7 のサンプルにおける、最も低い累積感染数と最も高い累積感染数のギャップを示す。

	AVG	MP
Difference	Need the averaging of stochastic behavior	Equations for already averaged
Calculation time (Sec)	602.28630	20.06478

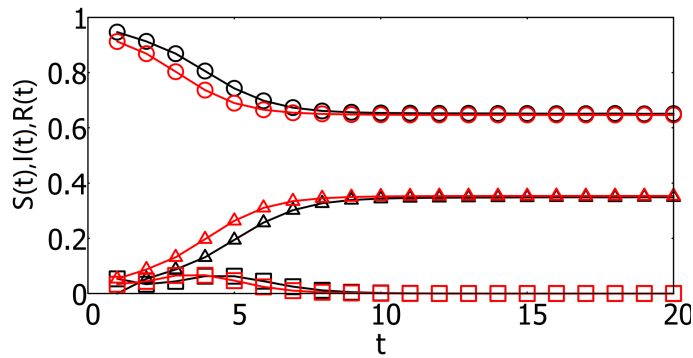
表 4.2: 高速と従来の SIR モデルにおける収束までの差と計算時間について (CPU: i7-11800H、メモリ: 16GB)。



(a) size of seed=885 at $l=2$



(b) size of seed=516 at $l=3$



(c) size of seed=407 at $l=4$

図 4.11: 従来の SIR モデルと高速 SIR モデルにおける、三つの状態 S 、 I 、 R の割合の時間変化を示す。

4.3 従来のIMPに対するヒューリスティック手法

IMP に対する、従来のヒューリスティック手法について説明する。

High Degree

High degree (HD) 法は、次数が高い順に k 個のノードを影響力のある種として選択する [44, 45]。この方法は、最も近い隣接ノード数に関する局所的なトポロジー情報のみを必要として、シンプルで効率的に種を見つけることができる。

k-core

ネットワーク内で密に連結された核として、k-core は少なくとも互いに K 値以上のノードに結合されている部分グラフを指す。

k-core に含めるには、その部分グラフ内で少なくとも自身のノード以外の K 個の他ノードにリンクしている必要があるが、グループ外で他のいくつかのノードに結合しているかは関係ない。

例えば、その部分グラフが、その部分グラフ内の少なくとも他の 2 個のノードに結合しているすべての対象を含んでいる場合、その部分グラフは 2 コアになる。同様に、その部分グラフがその部分グラフ内の少なくとも 3 個の他のノードにリンクしているすべてのノードを含んでいる場合、その部分グラフは 3 コアになる。3 コア内のすべてのノードは、3 コアの他のすべてのメンバーへのリンクを少なくとも 3 個持っている必要がある。そのような対象は必然的に、3 コアの他のすべてのメンバーへのリンクを少なくとも 2 個持っている必要がある。言い換えれば、3 コアは 2 コアのサブセット (2 コアに含まれる) になる。より一般的には、k-core は、階層的に生成され、0 コアには 1 コアが含まれ、1 コアには 2 コアが、2 コアには 3 コアが含まれている (以下同様)。K の値が増加すると、コアサイズは小さくなるが、核内の連結度は高まる [46]。

以下の図 4.12 は、単純なネットワークの k-core の分解図を示しており、より濃い灰色で示された 0 コア、1 コア、2 コア、および 3 コアの階層化された集合を示す。図 4.12 の場合、3 コアは、実際には 2 つの別個のグループで構成されていることに注意されたい。

一方、k-shell は k-core の分解方法と同様に描出できる。但し、k-shell は k-core -

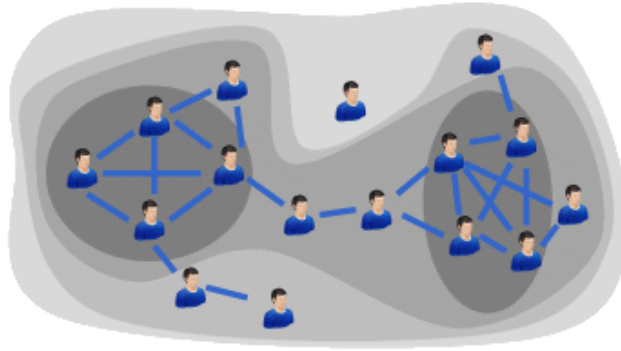


図 4.12: k-core 分解

($K-1$)-core として求められる。

以下は k-shell 分解の具体的な手順である。

Step0: 時刻 $t=0$ に次数が $K = 1$ の葉ノードを再帰的に削除して、削除した部分は 1-shell になる。

Step1: 毎時刻 $t=1,2,3,4,\dots$ に次数が $K = t + 1$ より小さいノードを再帰的に削除して、削除した部分は k-shell になる。

Step2: 全部ノードを削除するまで Step1 を繰り返す。

下の図 4.13 では色分けされた 3 つの shell に分解される。

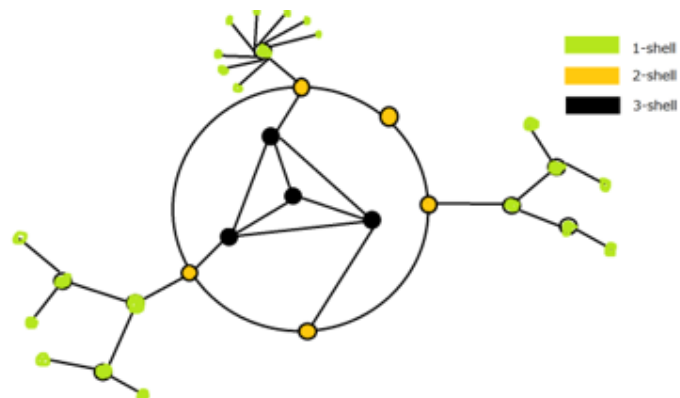


図 4.13: k-shell 分解

k-core 法を種ノードの選択に適用する際、種ノードはその Shell 数 k_s の値によってランク付けされる。その値は k-shell 分解を通じて計算され、葉ノードがなくなるまで繰り返す。実際、k-shell 分解では、すべてのノードが 1 の ID をもつどれかの shell に分割される。一般に、周リンクノードと比較して、コアノードはより大きな拡散を伴う傾向がある。したがって、最大の k_s を持つコア内のノードを種ノードとして定義する。

Local Centrality and Local Structure Centrality

HD 法はシンプルで効率的であるが、大域的なネットワークの性質を無視している。例えば、ハブの近傍が葉ノードである場合、ハブは弱い拡散力となり、効率が低い。一方、媒介中心性（Betweenness）と近接中心性（Closeness）は大域的な情報を考慮するが、計算がやや複雑である。そこで、Local Centrality(LC) は、局所性と計算にかかる時間とのトレードオフを考慮している [47]。LC は以下のように定義される。

$$Q_{LC}(u) = \sum_{w \in \partial u} \sum_{k=0}^2 |\partial Ball(w, R)|, \quad (4.8)$$

$$C_{LC}(v) = \sum_{u \in \partial v} Q(u), \quad (4.9)$$

ここで、 ∂u はノード u の近傍ノード集合を示し、 $\partial Ball(w, R)$ は図 4.14 に示すようにノード w から R ホップ離れたノードの集合を示す。 $||$ は集合の大きさを表す。種として、 $C_L(v)$ の値が大きい順にノード v が選択される。なお、 LC は近接中心性と同様に広がり強さを与えることが知られている [47]。

また、Local Structure Centrality(LSC) は、 LC の拡張である [48]。LSC は、調整

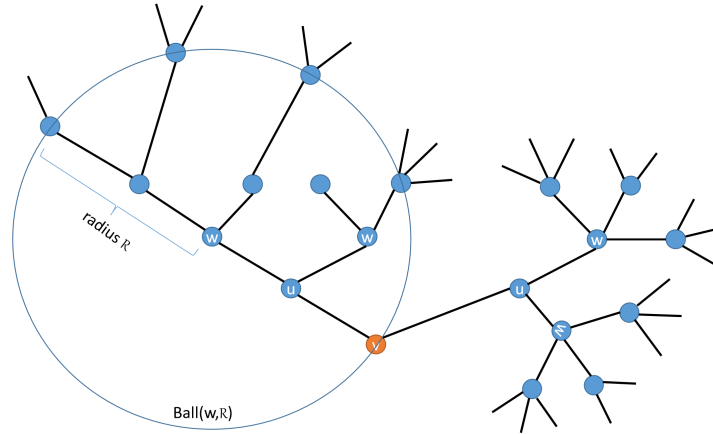


図 4.14: A set of vertexes in 2-ホップ from a vertex w in the process of LC.

可能なバランスパラメータ $0 \leq \alpha \leq 1$ で 2.2.2 節に述べた局所クラスタリング係数

C_w [5] と LC の線形補間によって以下のように定義される。

$$Q_{LSC}(u) = \alpha \sum_{k=0}^2 |\partial Ball(u, k)| + (1 - \alpha) \sum_{w' \in \partial Ball(u, 2)} C_{w'}, \quad (4.10)$$

$$C_{LSC}(v) = \sum_{u \in \partial v} Q(u), \quad (4.11)$$

ここで、 $\partial Ball(u, 2)$ はノード u の 2 番目に近い近傍のノードの集合を表す。また、[48] で述べられている、 $\alpha = 0.7$ と設定する。

Collective Influence

Collective influence (CI) は、以下のようにして IMP の最小集合を見つけることを目的としている [40]。まず、非線形メッセージパッシング方程式を考える、

$$v_{i \rightarrow j} = n_i \left[1 - \prod_{k \in \partial i \setminus j} (1 - v_{k \rightarrow i}) \right], \quad (4.12)$$

ここで、 $\partial i \setminus j$ は i の隣接ノード集合 ∂i のうち j を除いたものである。 $v_{i \rightarrow j}$ が全て 0 ではない時、 $v_{i \rightarrow j} = 1$ となり i から j に伝搬することを表す。次に、除去ノードをどのように選ぶかに関する $n = (n_1, \dots, n_N)$ と除去率 q に依存した上記の式の反複写像における原点の安定性の条件は、式の左リンクの線形近似における Jacobian 行列 $\left[\frac{\partial v_{i \rightarrow j}}{\partial v_{k \rightarrow i}} \right]$ の最大固有値によって決定される。つまり、最大固有値が 1 未満の場合、影響を与えるノード集合 $i : n_i = 0$ を削除することによって、拡散が停止する。すなわち、最も少ない除去ノード集合は、ネットワークを分断させ、言え変えると、それらは情報拡散に必要不可欠なものとして CI を定める。そのため、貪欲アルゴリズムを使用して固有値を最小化することで、各ノード i に対して CI を導出する [49]。

$$CI_{\mathcal{R}}(i) = (k_i - 1) \sum_{j \in \partial Ball(i, \mathcal{R})} (k_j - 1), \quad (4.13)$$

ここで、 $\mathcal{R}$ はノード i からホップ数 $\mathcal{R}$ までの半径を表している。最も高い $CI_{\mathcal{R}}(i)$ が種として選択される。ノード i を削除した後、ネットワーク内の残りのノード $i' \in V$ について、 $CI_{\mathcal{R}}(i')$ が再計算される。この手法は、ネットワーク全体ではなく、半径 $\mathcal{R}$ のボール内の局所的なトポロジー構造のみを必要とする。

第5章 IMP への近似手法の提案

本章 5.1 節では、IMP を解決するための提案手法について説明する。5.2 節では、提案手法と従来手法 [44, 45, 46, 47, 48, 40] の比較結果を示す。5.3 節では、本章のまとめを述べる。

5.1 提案する組み合わせ法

従来のヒューリスティック手法 ([44, 45, 46, 47, 48, 40]) などは複数の種を見つけるために適用できるが、拡散の重なり現象を考慮していない。図 5.1 (a) に示されているように、中心的なノードが近くにある場合、高次数法 (HD) [44, 45] は複数の種を見つけるのに適していない。なぜなら、その拡散領域が重複するからである。この重複を避けるために、統計物理学的アプローチと l -ホップ被覆に着想を得た近似を考える。

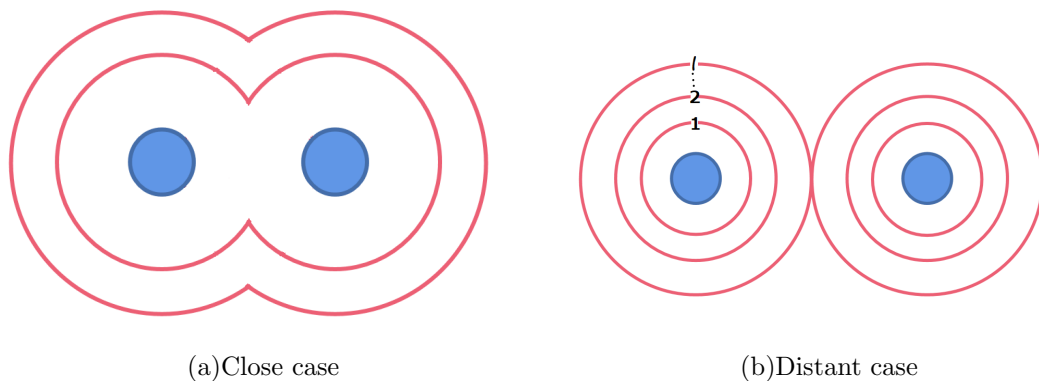


図 5.1: (a) 重なり合う状況と、(b) 重なりを避ける状況。

重なりがなく拡散力が大きい複数の種を見つけることは、 l -ホップ被覆 (l が 1 より大きい) を使用した最小集合被覆拡張をすることができる。また、集合被覆問題は頂点被覆 (VC) 問題に帰着することができる [13]。本研究におけるアイデアと

して、コンピュータ科学の l -ホップ被覆と最小 VC の統計物理アプローチ [14] を、情報拡散のより高速なシミュレーションを通じて組み合わせる。ここで、 l -ホップ被覆とは、種が l ホップ先のノードに感染することを想定する。 $l = 1$ の特殊な場合として、セット被覆、支配集合、VC 問題は 1-ホップ被覆に相当する。

その組み合わせ方法の概要は以下の通りである。まず、種数 N_s は、各種間の距離が l ホップ以上である l -ホップ被覆によって決定される。比較のため、 N_s は $l=1, 2, 3, 4$ に対応する固定値とする。次に、 l -ホップ被覆を説明する。図 5.2 に示されているように、ノード i を VC の最初の候補として選択する。ノード i とその l -ホップ近傍を削除した後、残りのネットワークから VC の 2 番目の候補として、ノード j が選択される。その後、上記の手順を、ネットワークにノードが存在しなくなるまで繰り返す。

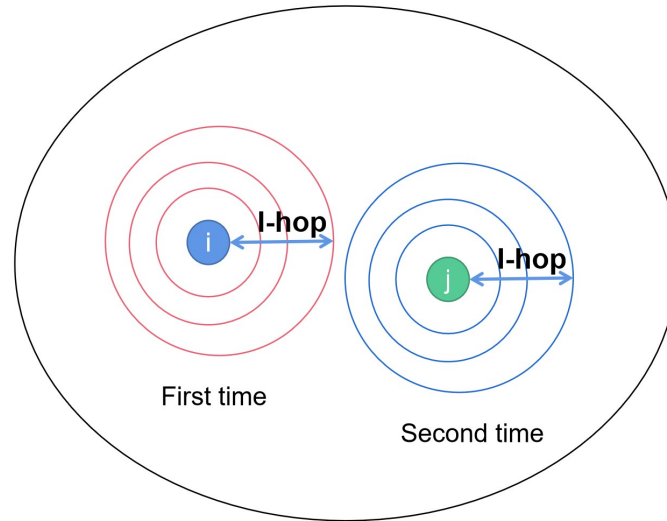


図 5.2: l -ホップ被覆の概念を説明する図解。

拡張された l -ホップ被覆における最小 VC の詳細は以下の通りである。

ステップ 1 式 (3.3)～(3.6) を用いて N 個のノードのランダム置換順の T ラウンドの更新反復後、各ノード i の確率 $\hat{\pi}_i^{(1)}$ を求める。

ステップ 2 Dismation では、最も高い $\hat{\pi}_j^{(1)}$ を持つノード j が種として選択される。 l -ホップ被覆に含まれる $\partial Ball(j, l-1)$ を考え、選択されたノード j とその

$\partial Ball(j, l-1)$ がネットワークから削除される。種のは数は、 $N_s \leftarrow N_s + 1$ (最初に $N_s = 0$ と設定) と更新される。

ステップ 3 ステップ 1 と 2 をすべてのノードがネットワークから取り除かれるまで繰り返す。最終的に、複数の種のは数は N_s として得られる。

5.2 提案手法と従来の手法の比較結果

提案手法で選択された複数の種と、従来のHD、k-core、LC、LSC、CI法によって選択された複数の種の情報拡散力を表5.1中の8つのソーシャルネットワークで比較した。なお、提案手法と従来手法は同じ種サイズ N_s の場合で情報拡散力を比較する。CI3が他の従来手法を上回るため、それをベースラインとして考える。

Data set	$ V $	$ E $	Diameter D	$\frac{\langle k \rangle}{\langle k^2 \rangle}$	N_s at $l=2$	$l=3$	$l=4$
LastFM	7624	278060	15	0.06	885	516	407
AirTraffic	1226	2408	17	0.07	192	129	78
Hamster	1788	12476	14	0.02	116	53	12
Polblogs	1222	16714	8	0.01	58	17	3
Gnutella	6299	20776	9	0.06	534	233	147
GRQC	4158	13422	17	0.06	544	378	306
HepTh	8638	24806	18	0.06	1115	761	625
PGP	10680	24316	24	0.06	1508	1055	795

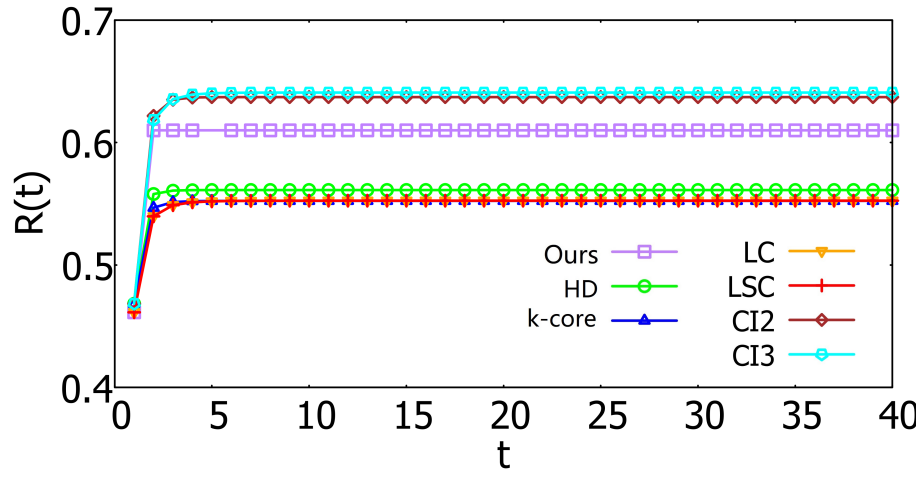
表 5.1: 8つのソーシャルネットワークの基本的なトポロジー特徴。

ソーシャルネットワークの1つであるLastFMの典型的な結果を述べる。図5.3は、4.2.3項で定義した累積感染者数の時間発展 $R(t)$ を示す。図5.3(a)に示されるように、複数の種ノードが l -ホップ被覆抜きの通常の最小VC法によって種ノードを選択した場合の紫の線（正方形印）は、茶色（ダイヤモンド印）の線（ CI_2 ）とシアン（五角形印）の線（ CI_3 ）よりも低い値を示している。これは、複数のホップの被覆を考慮せず、重複を回避できない最小VC法は、いくつかの従来法に比べて、複数の種の選択において拡散効果が低いことを示唆している。

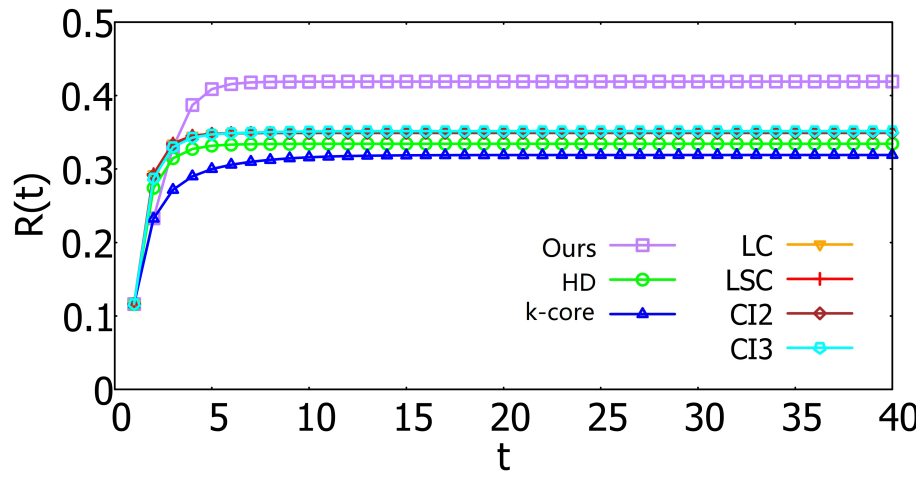
一方、図5.3(b)(c)に示されるように、茶色（ダイヤモンド印）の線（ CI_2 ）、シアン色（五角形印）の線（ CI_3 ）、オレンジ色（逆三角形印）の線（LC）、赤色（十字印）の線（LSC）は、緑色（円形印）の線（HD）および青色（三角形印）の線（k-core）よりも高い。CI、LC、およびLSCは、種の最近傍だけでなく、次近傍やその次の近傍なども考慮するため、HDやk-coreより拡散力が高いことが指摘されているが、これらの結果はそれと矛盾しない。 N_s はそれぞれ、3517、885、407であり、それらに対応する N_s/N は、それぞれ0.46、0.12、0.05である。

特に、高速 SIR モデルの場合、紫色 (四角形の印) の線 (提案手法) は、緑 (円形の印) の線 (HD)、青 (三角形の印) の線 (k-core)、オレンジ色 (逆三角形の印) の線 (LC)、赤い (十字印) の線 (LSC)、茶色 (ダイヤモンド印) の線 (CI_2)、およびシアン (五角形の印) の線 (CI_3) より最も高いことがわかる。図 5.1(b) に示されているように、提案手法で選ばれた種は互いに離れているため、種同士の影響範囲が重複しないことによって、高い拡散力を持つという結果を得たと考えられる。

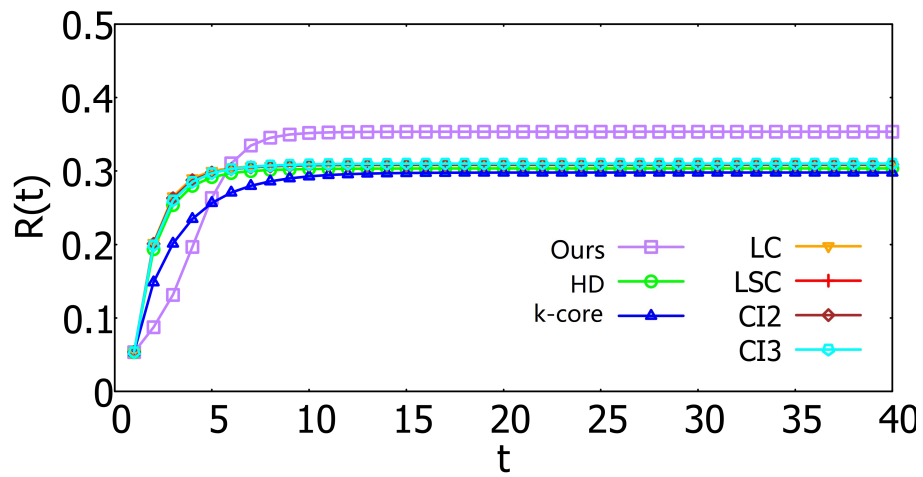
さらに、収束時間 t_c 以降、図 5.3(b) の紫色 (正方形の印) の線と他の線の差は、大きくなっている。一方、種数が少なくなるほど、1つの種あたりの拡散力は大きくなる。ここで、 t_c の値は 10 以上である。 l が増加するにつれて、 t_c も増加するが、種数は減少する。これらの結果に対する具体的な原因は下の図 5.4 を用いて議論する。



(a) size of seed=3517 at $l=1$



(b) size of seed=885 at $l=2$



(c) size of seed=407 at $l=4$

図 5.3: LastFM の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda=2$ で、 $l=1, 2, 4$ の場合を比較している。

図 5.4 は、2-ホップまたは4-ホップ被覆で選んだ2つの種 i, j のホップ数距離 $d_{i,j}$ の分布を示している。2本の紫色の線のピークが他色の線のピークよりも右にあるため、提案手法で選択された種は、従来手法で選択された種よりもお互いにより遠くに位置していることを意味する。2つの種間のより大きな距離は拡散の重なりを減らすため、当研究で採用した手法が従来の手法より効果的であると言える。また、塗りつぶされた正方形印で示される紫色の線のピーク（距離 $d = 6$ ）は、正方形印で示される紫色の線のピーク（距離 $d = 5$ ）よりも右にある。これは、 l （ホップ数）が増加するにつれて、種間の距離が増加することを示している。但し、表 5.2 でも後述するように、大きな l には限界があることに注意する必要がある。

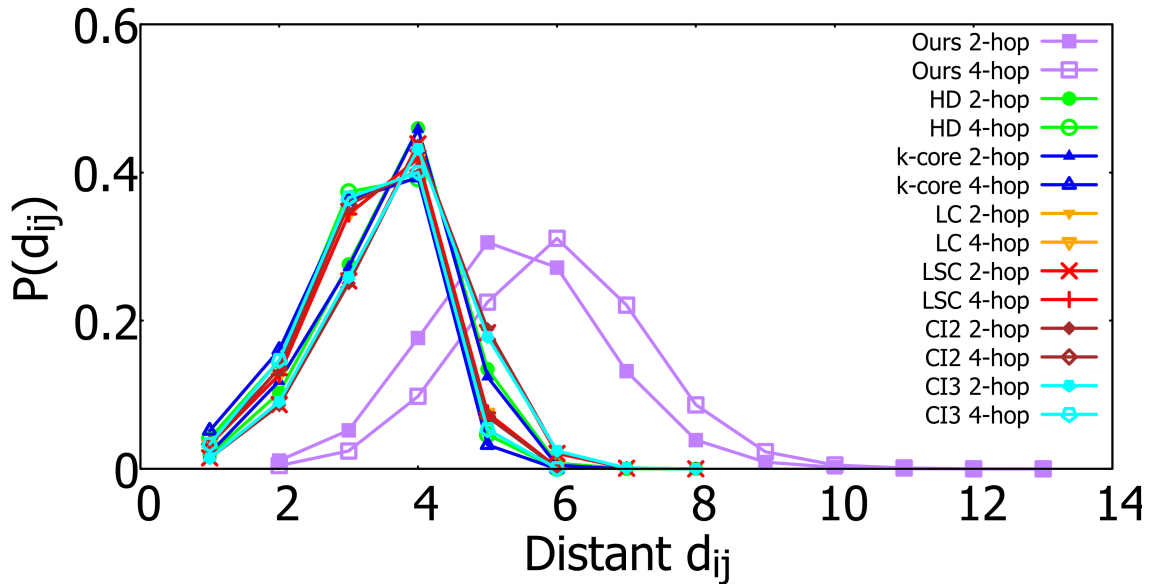


図 5.4: 各種のペア (i, j) 間の距離 $d_{i,j}$ の分布を 2-または 4-ホップ被覆 ($l = 2, 4$)。

また、異なる拡散率 $\beta = \lambda \frac{\langle k \rangle}{\langle k^2 \rangle}$ ($\lambda = 2, 4, 6, 8$) について、提案手法によるが複数の種の拡散力を調べた。なお、感染率がパーコレーション閾値 $\beta = \frac{\langle k \rangle}{\langle k^2 \rangle}$ [43] を上回ることは、現実的に妥当である [50]。なぜなら、感染率がパーコレーション閾値 $\beta = \frac{\langle k \rangle}{\langle k^2 \rangle}$ [43] を上回る時のみ、感染は持続的に行われるが、そうでないと、すぐに停止してしまう。図 5.5 に示されているように、横軸は感染パラメータ λ を示しており、2 から 10 の範囲 (β は 0.12 から 0.6 の範囲) になっている。なお、 $\lambda = 2$ の場合は図 5.3 に対応する。縦軸の $R(t_c)$ は、収束時間 t_c での累積感染数を示している。

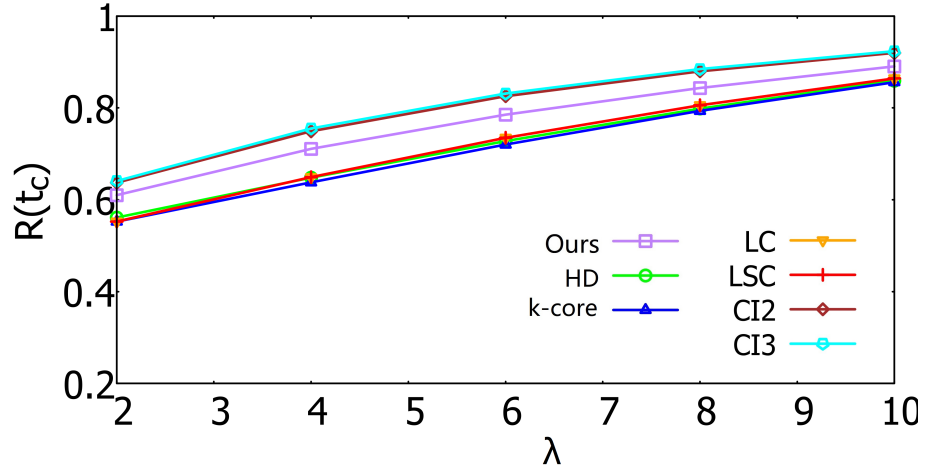
図 5.5(a) においては、拡散の重なり現象のために ($l=1$ は多ホップ被覆を考慮して

いない)、紫色の四角い印の線が最高ではない。 $l > 1$ の場合、紫の四角い印の線 (提案手法) は他の手法より常に高くなる。しかし、図 5.5(b)(c) において、 λ の値が増加するにつれて、提案手法 (紫色の正方形の線) と他の方法との差が徐々に小さくなっていくことが分かる。というのは、感染率が増加すると、種がより多くのノードを感染させるためである。

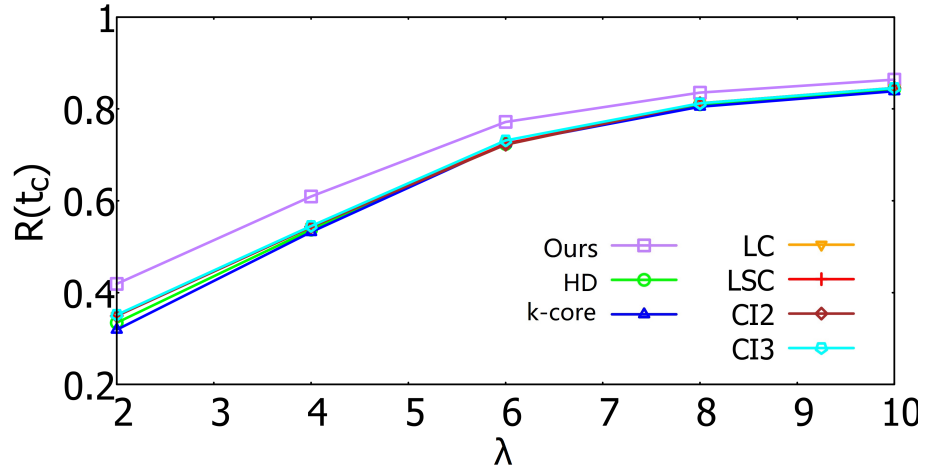
さらに、表 5.2 から、提案手法によって選ばれた 1 つの種あたりの情報拡散力 ($N \times R(t_c)/N_s$) が、従来手法によって選ばれたものよりも大きいことが分かる (l 値ごとに各最良の性能は、太字で強調されている)。特に、被覆距離 l が増加するにつれて、提案手法で選択された種ごとの拡散力が大きくなる。したがって、より大きな被覆距離 l で選択された種は、 l が増加するにつれてより良い拡散力を持つが、 l はネットワークの直径 $D - 1$ よりも小さい範囲内に制限されている。ここで、 D は、ノード間の最短経路の最大距離と定義されるネットワークの直径である。 l が $D - 1$ を超えると、最初の種が選択された後にすべてのノードが削除されるので、複数の種ではなくなる。

l-ホップ # of seed	$N \times R(t_c)/N_s$ in AVG-SIR			$N \times R(t_c)/N_s$ in 高速 SIR		
	1	2	4	1	2	4
Our method	1.322	3.562	6.579	1.341	3.601	6.622
HD	1.216	2.853	5.484	1.242	2.879	5.685
k-core	1.197	2.671	5.213	1.217	2.748	5.580
LC	1.197	2.999	5.702	1.199	3.008	5.770
LSC	1.210	3.000	5.713	1.221	3.010	5.785
CI_2	1.382	3.011	5.559	1.411	3.007	5.782
CI_3	1.389	2.989	5.683	1.419	3.028	5.808

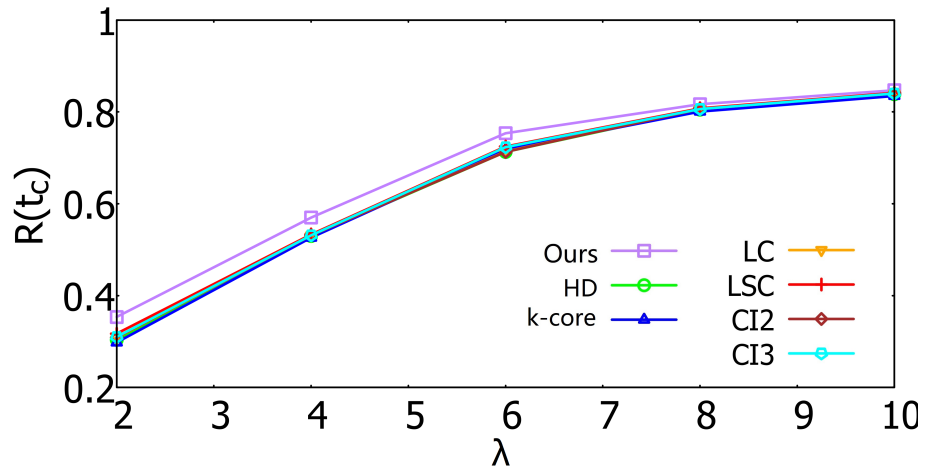
表 5.2: 提案手法と従来の 6 つの手法による従来の SIR モデルと高速 SIR モデルで選択された各種ノードの感染力。 $N \times R(t_c)/N_s$ は、種ノードあたりの感染力を示し、 $R(t_c)$ は収束時間 t_c での累積感染を示す。



(a) size of seed=3517 at $l = 1$



(b) size of seed=885 at $l = 2$



(c) size of seed=407 at $l = 4$

図 5.5: LastFM の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。

残りの7つの現実的なネットワークの結果は付録に表示されている。

5.3 本章のまとめ

重なりがなく拡散力のある複数の種ノードを効果的に見つけるために、私たちはIMP問題を近似的に解決する新しい方法を提案した。その鍵となるアイデアは、コンピューター科学の最小VCと l -ホップ被覆を統計物理学的アプローチに組み合わせ、拡散の重なりを避けることである。また、従来のSIRモデルよりも多くのサンプルを必要としない高速SIRモデルを提案している。そのため、サンプル数/ T ラウンド数 T に比例して通常のSIRモデルよりも高速となる。よって、このより速いSIRモデルを使用して、拡散プロセスを迅速にシミュレーションすることを可能にしている。蓄積された感染の時間発展の結果から、提案手法は、サイズが異なる社会ネットワークに対して、他の従来手法よりも優れた性能を発揮できると言える。

第6章 頑健性かつ効率的なネットワークを生成するアタッチメントモデルの族

本章 6.1 節では、広く知られた Scale-Free トポロジー構造を説明する。6.2 節では、提案する逆優先的選択 (Inverse Preferential Attachment) と従来の玉ねぎ状ネットワーク (Onion-like Network) の関係を説明する。6.3 節では、 $k^{-\beta}$ アタッチメントを用いた改良された p-モデルを説明する。6.4 節は、IPA モデル上の IMP のシミュレーション結果を示す。6.5 節では、本章のまとめである。

ところで、以下、リンクの結合によって成長するネットワーク生成に共通している。すなわち、各時間ステップで、新しいノードが一つ追加され、自己ループと多重リンクを禁止しながら、BA モデル [51] と同様に既存の m 個ノードに結合する。

6.1 Scale-Free 性と優先的選択 BA モデル

6.1.1 Scale-Free(SF) 性

SF ネットワークの特徴の 1 つは、平均次数よりもはるかに高い次数を持つノードが存在することである。最も高い次数を持つノードは通常、「ハブ (hub)」と呼ばれ、ネットワーク内で特別な役割を果たすと考えられている。SF 性は、ネットワークの攻撃に対処する能力と密接に関係している。もしノード攻撃がランダムに発生し、多数のノードが小さな次数を持っている場合、Hub ノードはほとんど影響を受けない。選択的に主要な Hub ノードを取り除くと、ネットワークは孤立した複数の部分グラフに分かれる。そのため、Hub ノードは SF ネットワークの利点であると同時に、欠点である。Cohen[52, 53] らや Callaway[54] らは、パーコレー

ション理論を用いてこれらの特性を解析している。Cohen[52] は、 $N \rightarrow \infty$ の SF ネットワークの一様ランダム故障に相当する臨界浸透閾値が $p_c \rightarrow 0$ であることを証明した。これは、平均次数が $\bar{k}$ の Erdős-Rényi グラフと対照的で、Erdős-Rényi グラフでは、臨界値は $p_c = 1/\bar{k}$ となる。

SF ネットワークのもう一つの重要な特徴は、ノードの次数が増加するにつれて減少するクラスタ係数分布である。この分布もべき乗則に従う [55]。これは、次数が低いノードが密な部分グラフに属しており、これらの部分グラフが Hub ノードを介して相互に結合されていることを示している。

一方で、異なる生成メカニズムに基づく SF ネットワークの詳細な特性は異なる。例えば、優先的選択によって生成されたネットワークでは大部分のノードをランダムに削除しても、ネットワーク全体の連結性にはあまり影響を与えない。これは、この種のトポロジーがセキュリティの向上に役立つということを示しているが、いくつかのターゲット攻撃はすぐにその連結性を破壊することができる。

一方、ほとんどの SF ネットワークにおいて、ノード間の平均距離（小世界ネットワークなど）は、規則的な格子グラフに比べて非常に小さい。Cohen と Havlin[56] は、 $2 < \gamma < 3$ のべき指数の SF ネットワークでは、非常に小さな半径 $d \sim \ln(\ln(N))$ となることを示した。実際には、 N が大きい SF ネットワークの半径は、ほぼ定数と見なすことができる。多くの現実世界のネットワークは、SF ネットワークであると考えられている。例えば、インターネットや World Wide Web、金融ネットワーク [57, 58]、Semantic networks[59] などである。

図 6.1 に SF ネットワークを可視化したものと、次数分布を示す。図 6.1A の可視化から SF ネットワークは、リンクが集中しているハブと低次数ノードがあることが分かる。また図 6.1B から、次数分布が両対数プロットで直線的になっていることから、それがべき乗則に従うことが分かる。

6.1.2 優先的選択 (Preferential Attachment) と BA モデル

最も知られている SF ネットワークの生成モデルは、Barabási と Albert (1999) [51] による利己的な Rich get richer 原理によるモデルで、入次数に比例する非均一

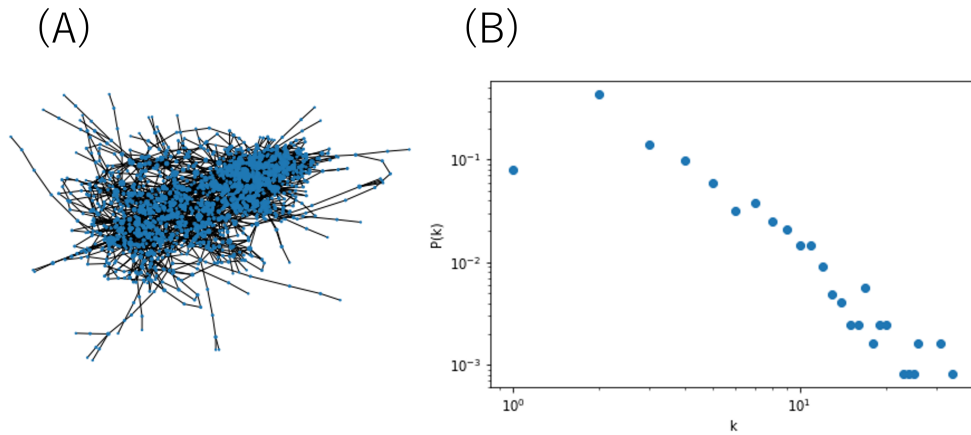


図 6.1: 航空網データ AirTraffic を (A) 可視化したものと (B) 次数分布. ノードは次数が高いほど大きく表示されている。また次数分布は両対数スケールになっている。

な確率に基づいて生成される。このモデルは、1965 年に Price が累積優位と名付けた [60] が、現在の名前 (BA モデル) は、Barabási によって再発見された後に注目を集めたことによる。例えば、より多くの入次数を持つウェブページは、一般的なウェブページよりも多くのリンクを集める傾向がある。

BA モデルにおいて、SF 性をもつ重要な要因は、優先的選択とネットワーク成長の 2 つである。優先的選択により新ノードから次数が大きい既存ノードを繋げると、時間が経つにつれて古株の次数がより大きくなる。成長と優先的選択による、ノード数の増大に伴って、次数が大きいノードはより大きくなり、巨大なリンク数を持って、Hub と呼ばれる結合が集中したノードが発生する。以下は BA モデルによるネットワーク構築のアルゴリズムである。

図 6.2 は、BA モデルにおいて、緑ノードは追加するノード、青ノードの次数は大きく、結合されやすい。ここで、ノード i の選択確率は以下で定義される [51]。

$$p_i = \frac{k_i}{\sum_{n \in N} k_n}$$

Step0: 時刻 $t=0$ に初期ネットワークを構築する。Step1 で追加されるリンク数 m

から、初期ネットワークは少なくとも m 以上のノードを持つ。そこで、本研究では $m+1$ の完全グラフを初期ネットワークとする。

Step1: 毎時刻 $t=1,2,3,4\dots$ 新しいノードから既存ノードに PA に従って m 本リンクを追加する。既存ノード i の確率選択は式 (2.6) に従う。但し、毎時刻で各ノードの多重リンクは禁止され、式 (2.6) の確率選択が再試行される。

Step2: 構築したいノード数 N に達するまで Step1 を繰り返す。

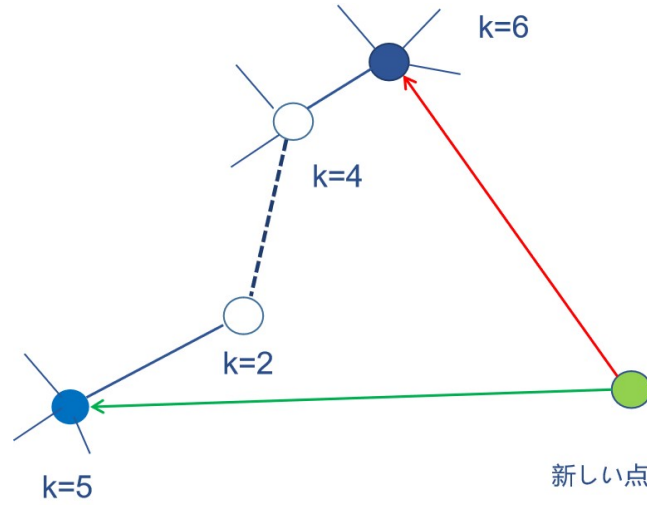


図 6.2: BA モデルを生成の見取り図

BA モデルによって生成されるネットワークの顕著な特徴は、べき乗次数分布を持つことである。以下、その理論解析 [51] を手短かに述べる。 $p(k, t_i, t)$ を、時刻 t_i にノード i が追加され、時刻 t で次数がちょうど k になる確率と定義する。BA ネットワークでは、新しいノードがネットワークに追加されるとき、ノード i の次数が 1 増加する確率は $k/2t$ であり、そうでなければノードの次数は変化しない。このことから、以下のマスター方程式と呼ばれる再帰式が得られる：

$$p(k, t_i, t+1) = \frac{k-1}{2t} p(k-1, t_i, t) + \left(1 - \frac{k}{2t}\right) p(k, t_i, t). \quad (6.1)$$

BA ネットワークの次数分布は次のようになる：

$$p(k) = \lim_{t \rightarrow \infty} \left(\frac{1}{t} \sum_{t_i} p(k, t_i, t) \right) \quad (6.2)$$

以下の再帰式：

$$p(k) = \begin{cases} \frac{k-1}{k+2}p(k-1), k \geq m+1, \\ \frac{2}{m+2}, k = m, \end{cases}$$

により、BA の SF ネットワークの次数分布が得られる。

$$p(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \sim k^{-3} \quad (6.3)$$

これは、BA ネットワークが3のべき指数を持つことを示している。

6.2 逆優先的選択 (Inverse Preferential Attachment)

と玉ねぎ状ネットワーク (Onion-like Network)

残念ながら、SF ネットワークは、巨大な次数を持つハブのノードへの意図的な攻撃に対して非常に脆弱である [61]。一方で、ハブを経由するなどして、任意の2つのノード間には効率的な短い経路が存在しやすい。効率的な経路と極端な脆弱性の組み合わせは、二重の刃の剣となる。したがって、私たちは、成長するネットワークにおいて、従来のものよりも強い頑健性と高いネットワーク通信効率の両方の重要な特性を備えた、より良いアタッチメント法を見つけることを目指す [6, 61, 51]。

過去数十年間、成長するネットワークにおいて多様なアタッチメント法が提案されてきた。例えば、ランダムアタッチメントと優先的アタッチメントを混合したアタッチメント法が議論されてきた [62, 63]。他の混合アタッチメントとしては、優先的アタッチメントと調整可能なクラスタリングの組み合わせも研究されている [64]。成長ネットワーク (GN: Growing Network) モデルにおいて [65, 66, 67]、次数が k であるノードが、 k^v に比例する確率でリンク先として選択される。その際、 v の値によって、 $v = 0$ 、 $0 < v < 1$ 、 $v = 1$ 、 $v > 1$ のそれぞれに対して、指数関数、指数カットオフを持つべき乗則、SF となるべき乗則、排他的な星状ネットワークなど、次数分布が連続的に変化する。また、複写モデルでは、ランダムに選択されたノードの最近傍に結合する複製またはコピーによって、近似解析で導かれたべき乗則分布が生成される [68, 69]。しかしながら、次数 k を持つノードが選択される確率が $k^{-\beta}$ に比例する逆優先結合は、 $\beta = 1$ の特殊な場合を除いてはほとんど考慮されていない [70, 71]。特に、次数分布の解析を超えて、逆優先アタッチメントについて、ネットワーク通信効率と頑健性は、ほとんど議論されていない。

一方、正の次数相関を持つ玉ねぎ状構造が、与えられた次数分布下で次数順の意図的な攻撃に対して最適な頑健性を持つことがわかっている [72, 25]。このような構造は、同じ次数のノードが、中心から周リンクに向かって次数の減少する同心円状に配置される場合に視覚的に表現される。次数相関を大きくするいくつかのリワイヤリング法が提案されている [72, 73] が、これらの方法は既存のリンクを破

棄するため、現実のネットワークへの適用が困難である。一方、ショートカットの追加と協調的部分コピーを適用する逐次的な成長法 [74] や仲介を利用する方法も提案されている [3, 4]。それらで生成された玉ねぎ状ネットワークにおいては、頑健性と通信効率の両方が共存することが数値的に示されている [3]。しかしながら、ランダム、仲介、逆優先のどれが玉ねぎ状ネットワークの出現にとってに主要因になるかは十分理解されていない。

6.2.1 玉ねぎ状ネットワーク (Onion-like Network)

近年、数値シミュレーションと理論解析により、玉ねぎ状ネットワークが発見された [4]。このようなネットワークは次数の大きい順に中心から周リンクに同心円上にノードを配置すると、正の次数相関を強くする同程度の次数を持つノードが結合することで玉葱状に可視化されることから玉ねぎ状と呼ばれる。以下、玉ねぎ状ネットワークによるネットワーク構築のアルゴリズムを述べ [4]、図 6.3 はその Step1, Step2 を説明するものである。

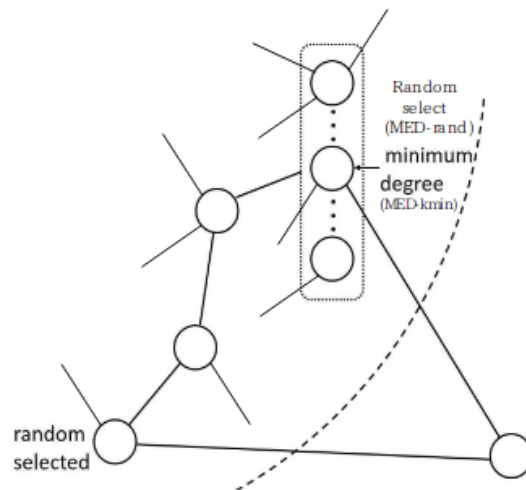


図 6.3: 玉ねぎ状ネットワークの生成

Step0: 時刻 $t=0$ に初期ネットワークを構築する。Step1 で追加されるリンクの数 m から、初期ネットワークは少なくとも m 以上のノードを持つ。そこで、本研究では $m+1$ の完全グラフを初期ネットワークとする。

Step1: 毎時刻 $t=1,2,3,4\dots$ に m 本リンクをもつ新しいノード 1 個を既存ネットワークに追加する。

Step2: 追加する時、既存ノードをランダムに一個選んで、新しいノードと選んだノードを結合する。更に、ランダム選択ノードから一定な仲介数 μ を経た $(\mu + 1)$ ホップ先の最小のノードと繋げる。毎時刻で各ノードの多重リンクは禁止される。

Step3: ノード数 N に達するまで Step1, Step2 を $\frac{m}{2}$ 回繰り返す。

図 6.4(a)(b) において、BA モデルによる SF ネットワークと上記の Step 1,2,3 による玉ねぎ状ネットワークの次数分布を比較する。図 6.4(a) の両対数の直線部分から、スケールフリーネットワークはべき乗次数分布に従う。一方、図 6.4(b) の縦軸のみ片対数の直線部分から、上記で生成した玉ねぎ状ネットワークは指数分布に従うことがわかる。

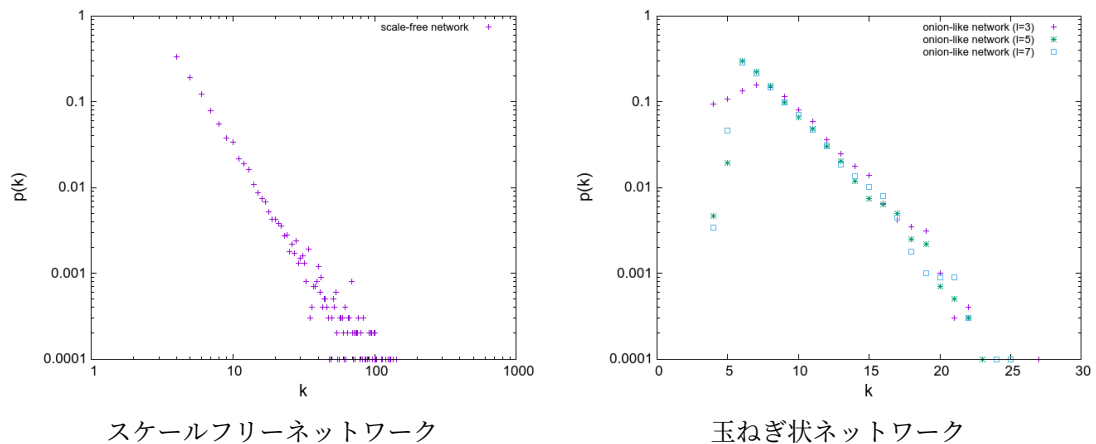


図 6.4: SF ネットワークと玉ねぎ状ネットワーク次数分布の比較

6.2.2 玉ねぎ状ネットワークにおける最小次数結合の要因

我々は、以下のように強固な玉ねぎ状ネットワークを生成するために、ノードのランダム選択と最小次数選択に焦点を当てる。一方、正の相関として知られる次数相関を持つ玉ねぎ状ネットワークを生成するために、ランダム選択と intermediation (MED) 選択のペアに基づく逐次的な成長法が提案されている [3, 4]。そこで、最小次数選択とランダム選択のいずれかで、MED モデルのタイプを MED-kmin ま

たは MED-rand と呼ぶ [4]。

MED-kmin

各 MED-kmin ペアにおいて、片方のリンクの結合先は一様ランダムに選択され、もう一方のリンクの結合先は、ランダムに選ばれたペアのノードから、仲介ホップ数 μ で最小次数のノードが選択される。 μ ホップの仲介とは、 $\mu + 1$ 番目の隣に結合することを指す。また、数値シミュレーションにより、 $\mu \geq 3, 4$ ホップのノードへの反復的な結合により、比較的長いループが形成され、頑健性が向上することが示されている [2]。もしネットワークにループが少ない場合、ノードを取り除くことで木構造になってしまう可能性があるため、ノードに対する攻撃が行われると、ネットワークは容易に分断されてしまう。この現象は、Dismatling と Decycling 問題の同等性によって支持されている [4]。

MED-rand

MED-rand の場合は、最小次数のノードの代わりに、ランダムに選択されたペアのノードから $\mu + 1$ 番目の近傍にあるノードが一様ランダムに選択される。古いノードは大きな次数を持つ傾向があるから、ランダムアタッチメントで、正の次数相関を持つ大きな次数のノード同士を結びつける可能性が高い。一方、最小次数のノードへのアタッチメントは、小さな次数のノード間での正の次数相関の向上に貢献する。なぜなら、その第二のアタッチメントによって、 $\mu + 1$ 近傍の最小次数ノードとネットワーク全体で最小次数 m の新しいノードとの結合が確立され、小さな次数のノード間の正の次数相関関係を向上させることになる。但し、第二のアタッチメントの効果が弱いため、MED-rand では、MED-kmin のような玉ねぎ状のネットワークは形成されない [4]。

本研究では、 $k^{-\beta}$ アタッチメント ($0 < \beta < \infty$) によって、 $\beta = 0$ でランダムアタッチメント、 $\beta \rightarrow \infty$ で最小次数アタッチメントの補間として、玉ねぎ状ネットワークが生じることを示す [Liao1]。ここで、関連する議論が第 6 章の MED に至るまで、距離の影響はアタッチメントに含まれていない (MED-kmin の距離効果を調整するパラメーター μ を含む修正については、第 6 章で議論されることに注意されたい)。

まず、 $k^{-\beta}$ -アタッチメントによる次数分布を導出する。後述するが、 β が大きくなるにつれて最大次数が制限され、 $\beta \rightarrow \infty$ での最小次数アタッチメントでは後述する特殊なチェーン構造が得られることがわかる。この構造は、長距離経路が存

在するため、ネットワークの効率が低下する原因となる。次に、次数相関 r 、連結の頑健性指標 R 、ネットワークの通信効率 E について数値的に調査し、どの β の値が $k^{-\beta}$ アタッチメントによって効率的で堅牢な玉ねぎ状ネットワークを生成するのに貢献するかを明らかにする。

$k^{-\beta}$ アタッチメントとして、ランダムなアタッチメントと最小次数アタッチメント ($\beta = 0, \infty$ で) の間のパラメータ的な補間として、既存ノード i を確率 $\frac{k_i^{-\beta}}{\sum_j k_j^{-\beta}}$ で選び、 $0 < \beta < \infty$ で $k^{-\beta}$ アタッチメントを考慮する。各時刻で新しいノードから m 本のリンクを追加する。一般的に、次数 k_i の確率関数 $f(k_i)$ によって選択されたノード i に対して非線形なアタッチメントを行った場合、次数分布 $P(k)$ は [75] の式 (9) から次のように導かれる。

$$P(k) = \frac{\langle f \rangle}{\langle f \rangle + m \times f(K_{\min})} \prod_{\kappa=K_{\min}+1}^k \frac{m \times f(\kappa-1)}{\langle f \rangle + m \times f(\kappa)}, \quad (6.4)$$

$$\langle f \rangle \stackrel{\text{def}}{=} \sum_{k=K_{\min}}^{K_{\max}} f(k) \times P(k), \quad (6.5)$$

ここで、 $K_{\min}$ と $K_{\max}$ は最小と最大次数である。 $f(k) = k^{-\beta}$ の場合、 $m = K_{\min} \leq k \leq K_{\max} < \infty$ として、以下のように書き換えられる。

$$P(k) = \exp \left[\log \left\{ \frac{\langle f \rangle}{\langle f \rangle + m \times f(K_{\min})} \prod_{\kappa=K_{\min}+1}^k \frac{m \times f(\kappa-1)}{\langle f \rangle + m \times f(\kappa)} \right\} \right], \quad (6.6)$$

$$= \exp \left[\sum_{\kappa=K_{\min}+1}^k \{ \log(m) + \log((\kappa-1)^{-\beta}) - \log(\langle f \rangle + m \times \kappa^{-\beta}) \} + C \right], \quad (6.7)$$

ここで、 $C = \log \frac{\langle f \rangle}{\langle f \rangle + m \times f(K_{\min})}$ は定数である。式 (6.7) の右リンクでは、右リンクの定数項 C や $\log(m)$ と $\log(\langle f \rangle + m \times K_{\min}^{-\beta}) > \log(\langle f \rangle + m \times \kappa^{-\beta})$ 、 $K_{\min} + 1 \leq \kappa \leq k$ は支配的ではないため、

$$\exp \left[\int -\beta \log(\kappa-1) d\kappa \right] = \exp [-\beta(k-1) \log(k-1) - 1] \sim e^{-\beta k \log k} \quad (6.8)$$

のように、定数因子を無視して近似される。

図 6.5 と 6.6 は、 $\beta = 1, 4, 10$ および $m = 2, 4$ の場合について、 $\log(N \times P(k))$ と $k \log k$ の最小二乗法による良好なフィッティングを示している。図 6.5、6.6 では、より大きな β に対して最大次数が減少し、 $\beta \rightarrow \infty$ の場合には $2m$ によって上限が設定される。また、 $k \geq 1$ において、 k と $k \log k$ の間に単調増加の関係がある。

次に、 $\beta \rightarrow \infty$ の場合、最小次数アタッチメントとして特別なチェーン構造が生じることを明らかにする [Liao1]。時刻 t のアタッチメントでは、新しいノードは時刻 $t-1, t-2, \dots, t-(m-1)$ に追加された既存のノードに結合し、その次数は最小の $m, \dots, 2m-2$ である。各時刻 t において、新たなノードは $t-1, t-2, \dots, t-(m-1)$ に追加された既存ノードに最小次数優先で繋げるため、次数は最小値 $m, \dots, 2m-2$ になる。多重リンクが禁止されているため、次数は各時刻で一つずつしか増加せず、合計 m 個のリンクのうち、 $2m-2$ より小さい次数で $2m-1$ になる。図 6.7 は、議論を簡単にするために、次数が $2m-1$ の最も古いノードを選択するという特殊なケースを示している。次数が $2m$ であるノードは、 $2m-1$ の次数を持つノードが複数存在するため、もうリンクを取得しない。上記の議論と図 6.7 から、これを確認する。時刻 t における次数の合計 M_t について、以下の式が成り立つ [Liao1]。

$$M_t = 2M_0 + 2mt, \quad (6.9)$$

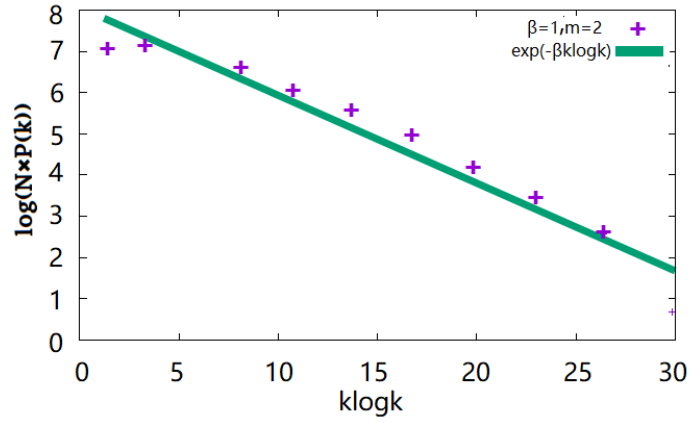
$$= 2m(N_0 + t - \Delta t) + (2m-1)(\Delta t - m) \quad (6.10)$$

$$+ \frac{((m+1) + (2m-2+1))(m-1)}{2} + m, \quad (6.11)$$

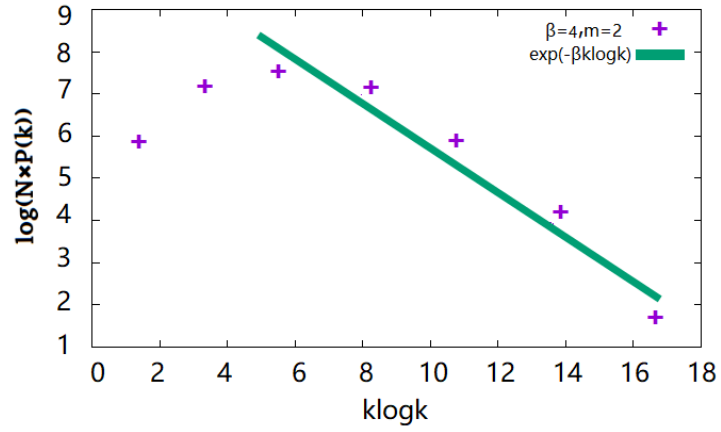
ここで、 M_0 と N_0 は初期 ($t=0$) ネットワークにおけるリンク数とノード数を示す。 Δt は、現在の時点と次数が $2m$ 未満の最も古いノードの挿入時刻の差を表している。したがって、 Δt は時間の長さではなく、ノードの数を表す。式 6.11 の右リンクにおいて、 $2m(N_0 + t - \Delta t)$ と $(2m-1)(\Delta t - m)$ はそれぞれ、次数が $2m$ と $2m-1$ のノードの総和を表している。3 番目の項 $\frac{((m+1)+(2m-2+1))(m-1)}{2}$ は、 $m+1$ から $2m-2+1$ までの等差数列の和を表している。最後の項は新しいノードの次数 m である。式 (6.11) を解くと、

$$\Delta t = 2mN_0 - 2M_0 - \frac{m(m-1)}{2}. \quad (6.12)$$

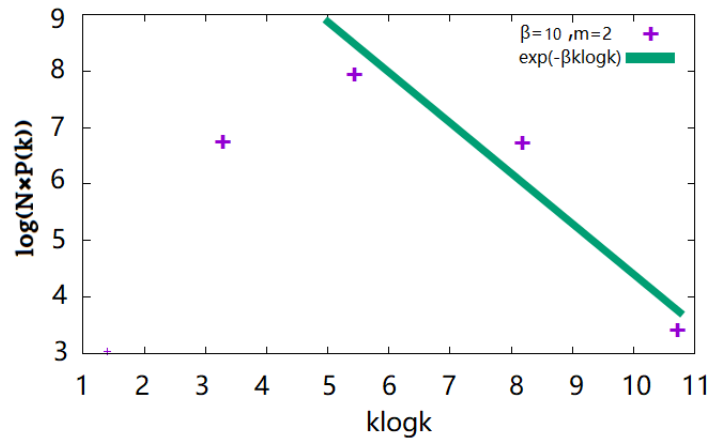
初期完全グラフの場合、 $N_0 = m+1$ 、 $M_0 = \frac{m(m+1)}{2}$ を式 6.12 に代入すると、以下の定数項が得られる。



(a) $m = 2, \beta = 1$

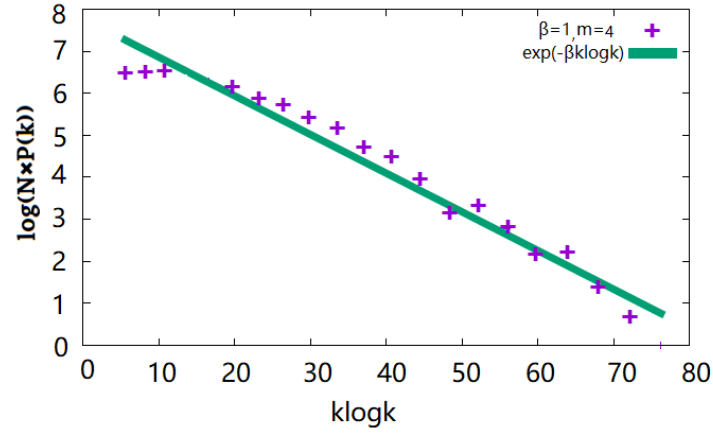


(b) $m = 2, \beta = 4$

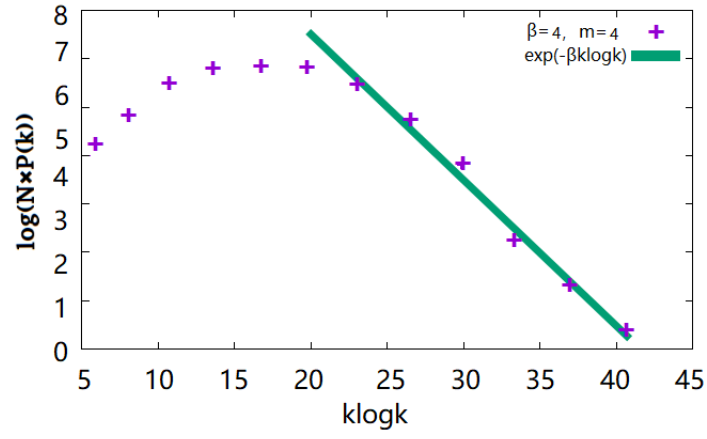


(c) $m = 2, \beta = 10$

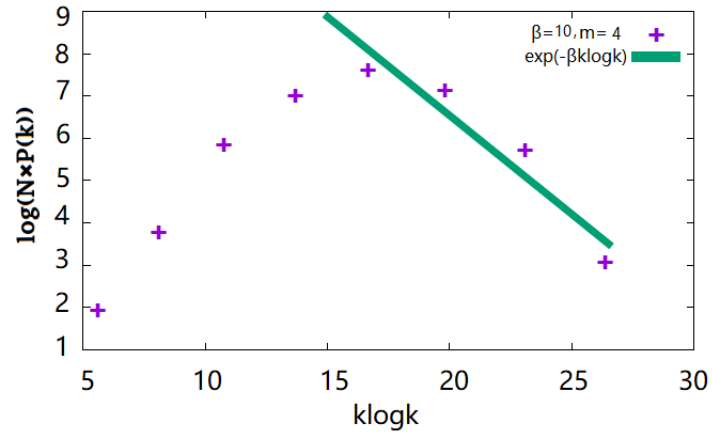
図 6.5: 分布の裾野を緑の線に $e^{-\beta k \log k}$ で推定。紫の点と緑の線は、 $k^{-\beta}$ アタッチメントによって成長するネットワークの次数分布とその推定値を表している。時間ステップごとの追加リンク数は $m = 2$ である。



(a) $m = 4, \beta = 1$



(b) $m = 4, \beta = 4$



(c) $m = 4, \beta = 10$

図 6.6: 分布の裾野を緑の線に $e^{-\beta k \log k}$ で推定。紫の点と緑の線は、 $k^{-\beta}$ アタッチメントによって成長するネットワークの次数分布とその推定値を表している。時間ステップごとの追加リンク数は $m = 2$ である。

$$\Delta t = \frac{m(m+3)}{2} > m. \quad (6.13)$$

式 (6.13) からわかるように、 $2m-1$ の次数を持つノードの数は $\Delta t - (m-1) > 1$ となる。したがって、 $2m-1$ の次数を持つ複数のノードが存在することがわかる。最小次数アタッチメントによって生成されるネットワークは、長時間経過後に多くの $t-1-\Delta t$ ノードが次数 $2m$ を持つため、ほとんど正則グラフでと言える。さらに、このネットワークが、 Δt の間隔を持つチェーン構造であることを強調する。チェーン構造の場合、最長の最短経路の長さ（ネットワークの直径）は、 $N/\Delta t$ と推定される。表 6.1 は、実測値が推定値 $N/\Delta t$ に非常に近いことを示している。わずかな違いは、図 6.7 の特別な場合の推定において最も古いノードではなく、測定においては次数 $2m-1$ のノードがランダムに選択されることに起因すると考えられる。 $0 < \beta < \infty$ に対して、 β が大きくなるほど長さが長くなる。 $\beta=0$ の場合はランダムアタッチメントの場合である。

m	Practical	Estimated $N/\Delta t$	Δt
2	1001	1000	5
4	365	357	14

表 6.1: $N = 5000$ 個のノードを持つ初期完全グラフから、 $m = 2, 4$ の場合に生成された、 $\beta \rightarrow \infty$ の最小次数アタッチメントによる直径の推定値 $N/\Delta t$ と実測値。

$k^{-\beta}$ アタッチメントによって、効率的かつ頑健な玉ねぎ状ネットワークを生成するために、どのような β の値が貢献しているか、次数相関 r [76, 77]、頑健性指標 R [72]、および通信効率 E [26] を調べる。

図 6.8 (a) と図 6.9 (a) は、 $k^{-\beta}$ アタッチメントによる成長ネットワークにおけるサイズ N に対する $\beta = 0, 1, 10, 50, 100$ 、および $\beta \rightarrow \infty$ の次数相関 r を示している。赤色のダイヤモンド印の線 ($\beta \rightarrow \infty$) は、紫の十字印の線 ($\beta = 0$)、緑の四角印の線 ($\beta = 1$)、シアン丸印の線 ($\beta = 10$)、黒の三角印の線 ($\beta = 50$)、オレンジの逆三角印の線 ($\beta = 100$) よりもより強い次数相関 r を示している。したがって、より大きな β の場合、より高い次数相関 r が得られる。同じ色の線や印を比較すると、 $m = 4$ の図 6.9 (a) の次数相関 r は、 $m = 2$ の図 6.8 (a) よりも高くなって

いる。ここでは、図 6.8(b) および図 6.9(b) に示されているように、同じ色の線または印で比較すると、 $m = 2$ の図 6.8(b) よりも、 $m = 4$ の図 6.9(b) の方が頑健性指標 R が高いことがわかる。さらに、図 6.8(a) においては、次数相関 r の線は β を 0 から ∞ に変化させるにつれて単調に増加する。一方、図 6.8(b) においては、次数相関 r の線は β を 0 から 10 に変化させるにつれて増加し、10 から 50 に変化させるにつれて減少し、50 から ∞ に変化させるにつれて再び増加する。また、図 6.8 (b) と 図 6.9 (b) において、頑健性指標 R は、 β が 0 から 10 に変化するにつれて増加し、10 から 50 に変化するにつれて減少し、50 から 100 に変化するにつれて再び増加する。 β が 100 を超えると、頑健性指標 R は一定値に近づく。図 6.8(c) と 図 6.9(c) において、通信効率 E は β を 0 から ∞ に変化させるにつれて単調に減少している。次数相関 r と頑健性指標 R がこのように変化する理由は、正確にはわかっていない。但し、以前に議論されたように、チェーン構造がそれらに影響を与える可能性がある。特に、 $m = 4$ の場合の図 6.9(a)(b) における $\beta \rightarrow \infty$ の赤い線は、比較的高い次数相関 ($r > 0.2$) と頑健性 ($R > 0.3$) を持つ、玉ねぎ状のネットワークとして振る舞うことを示している。強すぎる次数相関は、 R の値を減少させ頑健性を弱める傾向がある [25]。 r と R が玉ねぎ状のネットワークであるための正確な条件は存在しないため、 $r > 0.2$ および $R > 0.3$ として考えている。なぜなら、同じ N および m に対して BA モデルは玉ねぎ状ネットワークではなく、 $r \approx 0$ および $R < 0.23$ であるからである [3, 4]。図 6.8 と 6.9 (c) から、 β が ∞ に近づくにつれて、効率 E の値（赤い線）が、 $\beta = 0, 1, 10, 50, 100$ （紫、緑、シアン、黒、オレンジの線）よりも低くなっている。低い効率の原因は、図 6.7 や表 6.1 で述べたように、チェーン構造に起因すると考えられる。

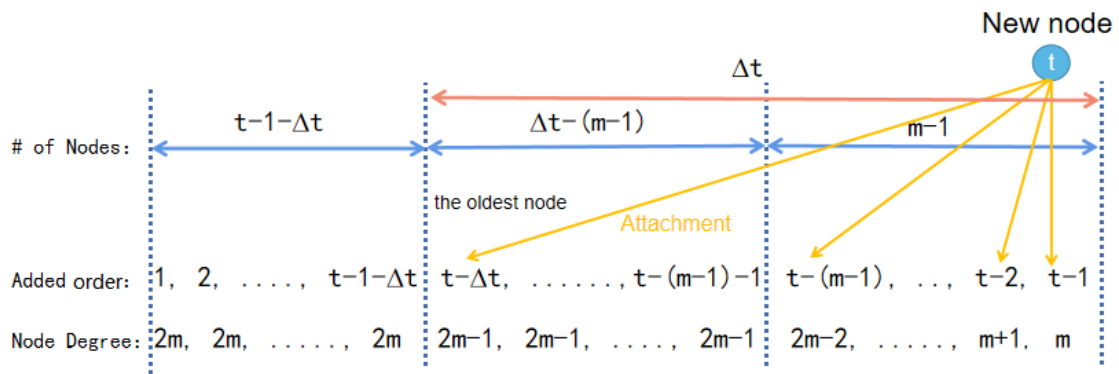


図 6.7: $\beta \rightarrow \infty$ の最小次数アタッチメントによるリンクの結合先。黒い線は、新しいノードから既存の m ノードへの最小次数アタッチメントによるリンクの結合を示している。右から左に 3 つの部分がある。右側の最初の部分では、 $m-1$ 個の既存のノードのうち次数が最小の $m, m+1, \dots, 2m-2$ に新しいノードが結合される。最小次数アタッチメントのためである。2 番目の部分（中央）では、残りの 1 つのリンクが次数 $2m-1$ のノードに結合される。 Δt 個の次数が $2m$ 未満のノードが存在するため、次数が $2m-1$ の最も古いノードの追加時間は $t - \Delta t$ となる。 $t-1-\Delta t$ 個のノードのうち、第 3 部分（左側）には、次数が $2m$ のノードが $t-1-\Delta t$ 個存在する。これらのノードの追加順序と、時間 t におけるノードの次数の関係は、下から 2 行目と 3 行目に示されている。

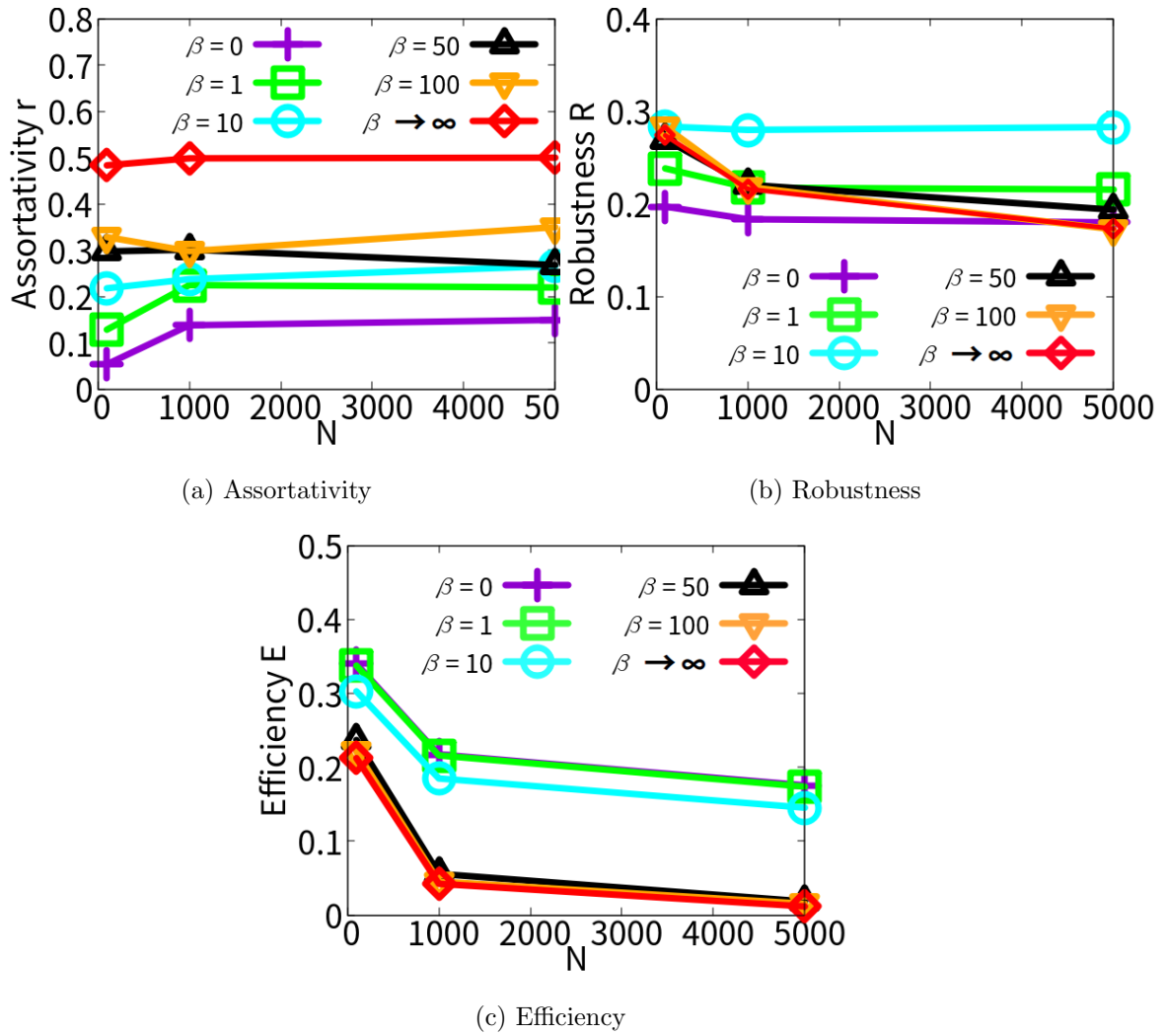
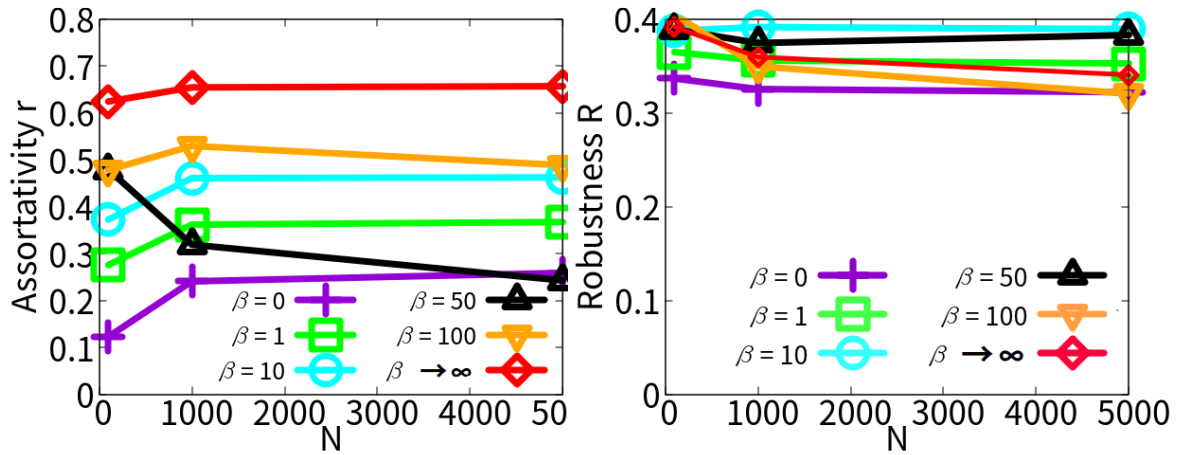
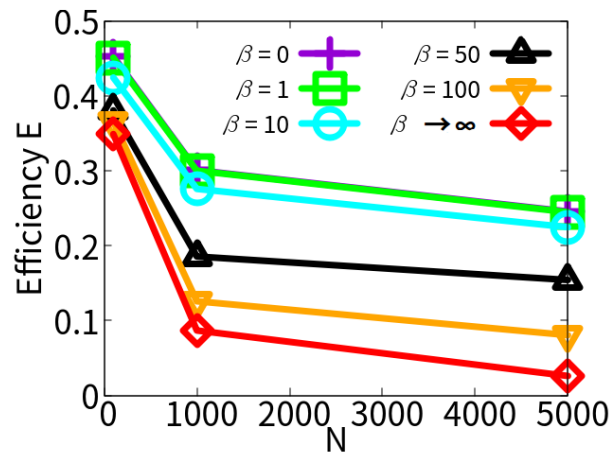


図 6.8: $k^{-\beta}$ アタッチメントによって生成されたネットワークにおけるノードサイズ N に対する指標: r 、 R 、 E 。 $m=2$ の場合で、リンク/時間ステップ。紫色 (十字印)、緑色 (正方形印)、シアン (丸印)、黒 (三角形印)、オレンジ (逆三角形印) の線は、それぞれ $\beta = 0$ 、1、10、50、および 100 に対応する。



(a) Assortativity

(b) Robustness



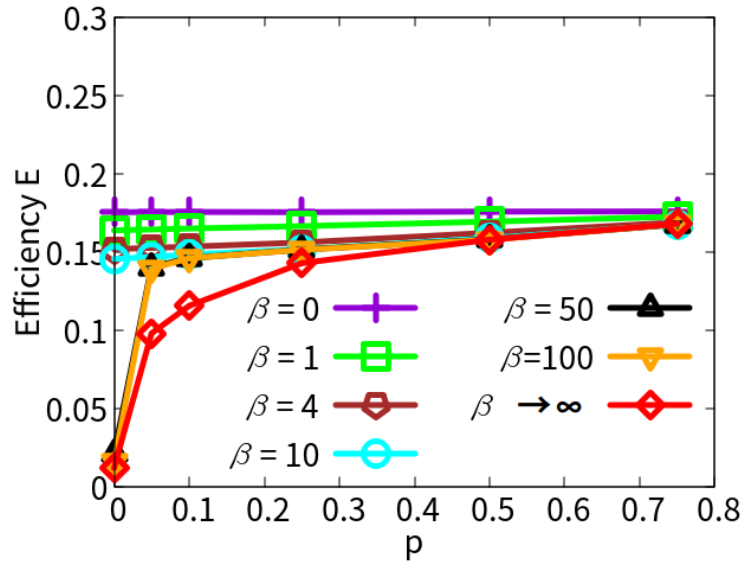
(c) Efficiency

図 6.9: $k^{-\beta}$ アタッチメントによって生成されたネットワークにおけるノードサイズ N に対する指標: r 、 R 、 E 。 $m=2$ の場合で、リンク/時間ステップ。紫色 (十字印)、緑色 (正方形印)、シアン (丸印)、黒 (三角形印)、オレンジ (逆三角形印) の線は、それぞれ $\beta = 0$ 、1、10、50、および 100 に対応する。

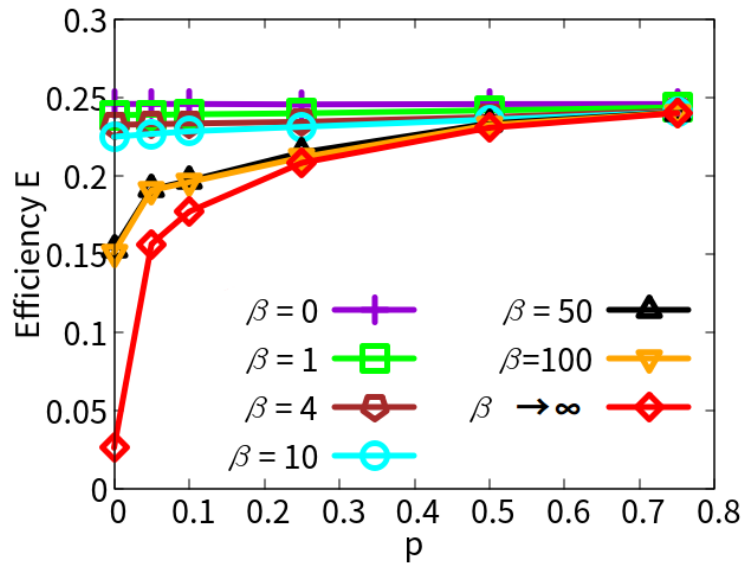
6.2.3 少量のランダムアタッチメントの必要性と効率的な経路

ランダムと $k^{-\beta}$ アタッチメントの混合アタッチメントにおいて、平均経路長が $O(\log N)$ となり、ネットワークの通信効率 E が向上すること調べる。6.2.2 項で述べた、 $\beta = 100$ の実測と、 $m = 2$ ($m = 4$) に対する推定によって、725.5 と 1000 ($m=4$ では 365 と 103) と、似たような値が得られ、 β が十分大きい場合、チェーン構造に近づくため、直径は $O(N)$ になる。以下の混合アタッチメントでは、新しいノードが各時刻で追加されるが、リンクの結合先は、確率 p でランダムアタッチメントで選択され、確率 $1 - p$ で $k^{-\beta}$ アタッチメントで選択される。但し、 $\beta = 0$ の場合は純粋なランダムアタッチメントである。

図 6.10 において、緑色（四角形印）、茶色（五角形印）、青緑色（丸印）の線（ $\beta = 1, 4, 10$ ）は、比率 p が大きくなるにつれて効率が向上する傾向があることを示している。黒（三角形印）、オレンジ（逆三角形印）、赤（ダイヤモンド印）の線（ $\beta = 50, 100$ 、と $\beta \rightarrow \infty$ ）は、横軸に対して劇的に効率が向上することを示していることに注目する。この現象は、鎖状構造が現れることに対応している。同じ色の線や印を比較すると、図 6.10(b) の $m = 4$ の場合、図 6.10(a) の $m = 2$ の場合よりも、玉ねぎ状ネットワークの形成によりより多くのリンクを使用しているため、効率が高くなっている。最短経路の平均長は算術平均と調和平均の関係で $\langle L \rangle \approx 1/E$ で与えられる。例えば、効率 $E=0.1, 0.2, 0.25$ は、それぞれ平均経路長が 10、5、4 ホップに対応する。通信効率 E が低くなるほど、平均経路長 $\langle L \rangle$ が増加する傾向がある。図 6.11(a)、(c) または (b)、(d) において、赤色（星印）、紫色（十字印）、緑色（正方形印）、シアン色（丸印）、オレンジ色（三角印）、黒色（逆三角印）、青色（菱形印）の各線が、 $p = 0$ （上部）から $p=0.001, 0.0025, 0.005, 0.05, 0.25$ 、および 0.5（下部）に増加するにつれて平均経路長 $\langle L \rangle$ が減少することが示されている。特に、純粋な $k^{-\infty}$ アタッチメントの場合、 $p = 0$ に対して急速に増加する（赤色の）曲線が得られるため、6.2.2 項で言及したようなチェーン構造のため、これらの結果は $\log N$ に比例しない。しかしながら、平均経路長が $O(N)$ と $O(\log N)$ の間で連続的に遷移するため、正確な臨界値 p_c は不明である。図 6.11 (c) (d) で示されるように、 $O(\log N)$ の直線状のものよりも長い平均経路長を示すものがあるので、 p_c は 0.0025 から 0.005 の間にあると推定される。また、同じ色の線や印と比較して、図 6.11 (d) の $m = 4$ の場合の平均経路長は、より多くの追加リンクを使用するため、図 6.11 (c) の $m = 2$ の場合よりも短くなっている。 $\beta = \infty$ だけでなく、他の値（ $\beta = 100$ ）でも、図 6.11 の類似した結果が得ら

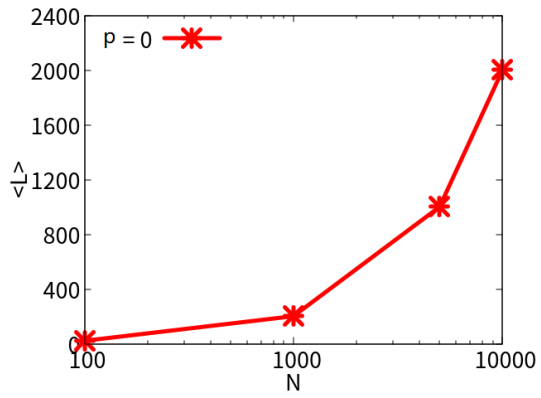


(a) $m = 2$

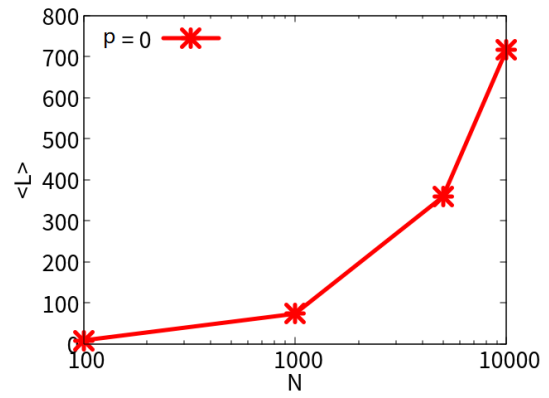


(b) $m = 4$

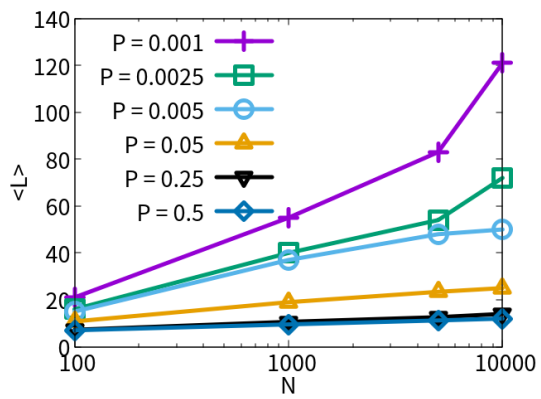
図 6.10: ランダムと $k^{-\beta}$ アタッチメントの混合アタッチメントによるネットワークにおける、ノードサイズ $N = 5000$ で $m = 2$ (左) と $m = 4$ (右) の場合の、混合確率 p に対する効率の変化。ここで、 p はランダムアタッチメントの割合を示している。紫色の線 (十字印)、緑色の線 (正方形印)、茶色の線 (五角形印)、水色の線 (丸印)、黒色の線 (三角形印)、オレンジ色の線 (逆三角形印)、赤色の線 (ダイヤモンド印) は、それぞれ $\beta = 0, 1, 4, 10, 50, 100$ 、および $\beta \rightarrow \infty$ を表している。



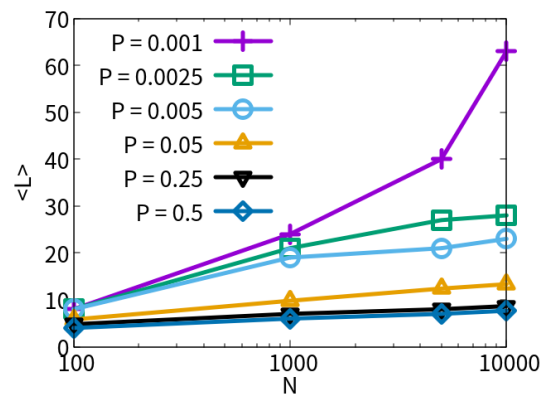
(a) $p = 0, m = 2$



(b) $p = 0, m = 4$



(c) $p > 0, m = 2$



(d) $p > 0, m = 4$

図 6.11: $k^{-\infty}$ アタッチメントにおける平均経路長 $\langle L \rangle$ とノード数 N の関係を示したグラフ。縦軸のスケールは (a)~(d) で異なることに注意。ここで、 p はランダムアタッチメントの割合を示し、 $m = 2$ (右) および $m = 4$ (左) での結果を示している。

れ、臨界値は $0.0025 \leq p_c \leq 0.005$ となる。そのため、ランダムアタッチメントが混合アタッチメントにおける効率的な経路の出現に必要であることがわかる。

6.3 $k^{-\beta}$ アタッチメントを用いた改良された p-モデル

本節、 $k^{-\beta}$ アタッチメントによる p-モデルの改良版を提案する。元々の p-モデル [2] は、MED モデル [3, 4] に似ているため、両方のモデルで、ランダムに選択されたノードから距離 d または $\mu + 1$ ホップ離れたノードがリンクの行き先として選択される。これを距離アタッチメントと呼ぶ。MED モデルと p-モデルの本質的な違いは、MED モデルでは距離アタッチメントが決定論的であるが、p-モデルでは確率的である。既存のノードから d または $\mu + 1$ ホップ離れたノードの候補集合にノードがいくつかある場合、元の p-モデルと MED-rand では、その中から一つが一樣ランダムに選択される。p-モデルが玉ねぎ状のネットワークを生成できるかどうかは明確ではない。そのため、候補ノードのランダム選択を、確率 $k^{-\beta}$ での選択に変更して、玉ねぎ状ネットワークの出現を調べることにする。

p-モデルの改良版では、結合ノード i はその次数 k_i と距離 d の二倍に応じたの確率で選択される。以前に示したように、 $m = 4$ の場合には玉ねぎ状ネットワークが生成されるため、今回は $m = 4$ の場合のみを考慮する。図 6.12 は、p-model0、p-model1、p-model2、および M-MED モデルにおける成長過程を示している。ここで、接尾辞の数字 0、1、2 は、ランダムなアタッチメントは 0 本の仮想的な選択を行う場合、1 本のランダムなアタッチメントを行う場合、および 2 本のランダムおよび距離（または仲介）のアタッチメントのペアを行う場合をそれぞれ意味する。M-MED モデルは、 $k^{-\beta}$ アタッチメントを持つ MED モデルを改良したものであり、p-model2 と比較するために導入されている。 $m/2$ のペアを基にした p-model2 と同様の構成に基づき、 $\alpha > 0$ 、 $\beta > 0$ 、 $d = \mu + 1 \geq 1$ は与えられたパラメータである。

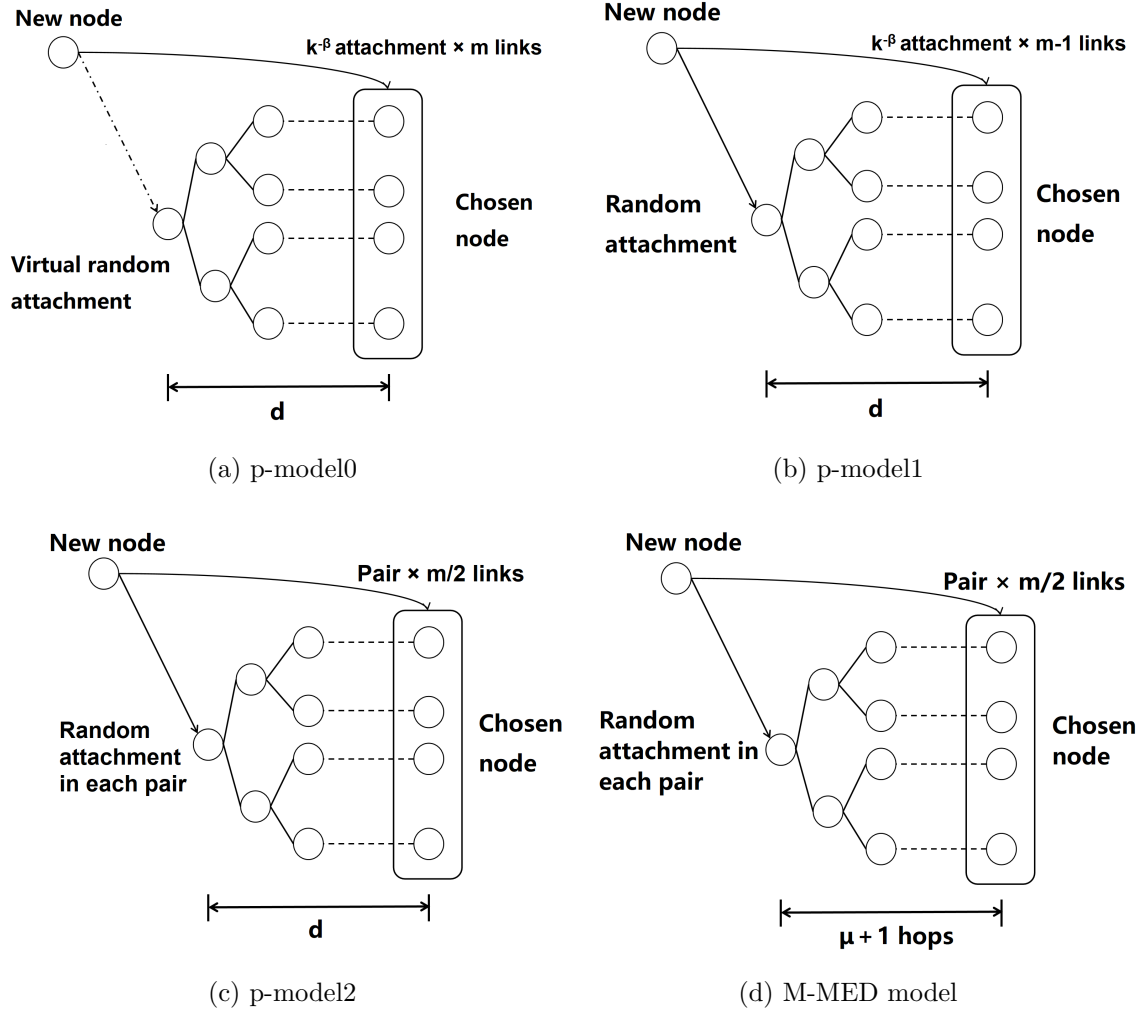


図 6.12: 成長過程を示す図として、(a) ~ (c) に、異なるランダム結合の数を持つ修正された p-model [2] を示す。図 (d) に示されているように、M-MED モデルは、MED モデル [3, 4] の最小次数選択ではなく、 $k^{-\beta}$ attachment で修正されている。距離 d と $\mu + 1$ は、p-models と MED モデルの場合、ランダムに選択されたノードからのホップ数で測定される。

p-model0 新しいノードからの m 本のリンクについて、既存のノードの中から一様ランダムに選択される。各リンクについて、以下の2つのステップの選択がある。まず、仮想的に選択されたノードから距離 d の候補ノード集合を $d^{-\alpha}$ に比例する確率で選択する。次に、候補ノード集合から $k_i^{-\beta}$ に比例する確率でノード i が選択される。候補集合にノードが1つしかない場合は、そのノードが選択される。

p-model1 最初にアタッチされるノードは既存のノードの中から一様ランダムに選択される。その後、p-model0 においては仮想選択は行わず、元の p-model [2] と同様に、 $k^{-\beta}$ アタッチメントによる修正を行いながら、他の $m - 1$ 本のリンクについても同様のプロセスを繰り返す。

p-model2 新しいノードから出る $m/2$ ペアのランダムと仲介アタッチメントのリンクのうち、1つのリンク先は既存ノードの中から一様ランダムに選択される。その後、残りのリンク先について、ランダムに選択した仲介ノードから距離 d にあるノードの候補集合が、 $d^{-\alpha}$ に比例する確率で選ばれる。そして、それぞれのリンクのために、候補集合内のノード i が $k_i^{-\beta}$ に比例する確率で選ばれる。

M-MED model p-model2 と同様に、 $m/2$ のペアノードが選択されるが、ランダムに選択する代わりに、それぞれのランダムに選択されたノードから定数距離 $\mu + 1$ の候補ノードの集合が決定される。この集合に対して、M-という修飾子が付け、ランダム選択の代わりに $k^{-\beta}$ アタッチメントが適用される。

以下の議論では、異なるランダムアタッチメントの数が、玉ねぎ状ネットワークの発生に与える影響を調べるために、p-model0 と p-model1 と比較する。同様に、距離 $d = \mu + 1$ が玉ねぎ状ネットワークの発生に与えられる影響を調べるために、p-model2 と M-MED と比較する。図 6.13 (a) は、 $\alpha = 0.1$ と 1 の場合（十字と三角形の印を持つ紫色と緑色の線）、 β が 0 から 10 に増加するにつれて、p-model0 での次数相関が急速に増加し、 $\beta > 10$ 以降は安定することを示している。 $\alpha = 3$ においては（逆三角形のシアンの線）、ランダムに選ばれたノードの最も近い隣接ノードに結合する、という弱い次数相関が現れる。この場合は複写モデル [68] に対応する。複写モデルは、常に仮想的なランダム選択ノードの最も近い隣接ノードに、リンクの数を（一定の m ではなく）変化させながら結合し、SF ネットワークを生成する [78] 方法である。また、図 6.13 と 6.14 (a) に示されているように、p-model1 の $\alpha=0.1$ と 1 の場合（十字と三角形の印を持つ紫色と緑色の線）の分類係数は、p-model0 の場合よりも低くなる。また、図 6.14 (a) において $\alpha = 3$ （逆三角形の印を持つシアンの線）と $\alpha=0.1$ と 1（十字と三角形の印を持つ紫色と緑色の線）の間にはほとんど差がない。したがって、距離 d の効果は、次数相関の増加において主要な影響を与えない。さらに、図 6.13 (b) に示すように、 $\alpha = 0.1$ および 1（紫色および緑色の線で、十字と三角形の印が付いている）の場合、 $\beta < 10$ では頑健性指標 R が急速に増加するが、 $\alpha = 3$ （逆三角形の印が付いているシアンの線）の場合はわずかに頑健性指標 R が弱くなる。これらのネットワークは、玉ねぎ状の構造として、高い次数相関と頑健性を持っている。

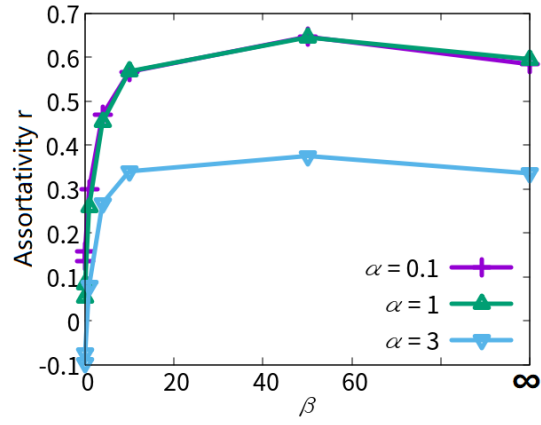
次に、p-model2 と M-MED モデルの次数相関 r 、頑健性指標 R 、効率 E について議論する。ここで、 $\alpha = 3$ または $\mu = 0$ 、 $\alpha = 1$ または $\mu = 2$ 、 $\alpha = 0.1$ または $\mu = 5$ は、それぞれ近、中、遠の仲介に対応する。p-model2 と M-MED モデルの図 6.15 と 6.16 (a) は、同じ色の線や印と比較して、近、中、遠の仲介に対する次数相関が同様に高いことを示している。さらに、p-model2 と M-MED モデルにはほとんど差がない。 $\mu + 1$ の距離が次数相関、頑健性、効率にあまり影響を与えないことに注意する必要がある。表 7.1 は、修正された p-model と MED モデルにおける β 値の出現をまとめたものである。 $\beta > 0$ の場合、すべての修正モデルが玉ねぎ状の構造になる。

さらに、図 6.17 (a)、(b)、(c)、(d) は、p-model0、p-model1、p-model2、および M-MED モデルにおける次数分布を示している。半対数プロットで指数関数分

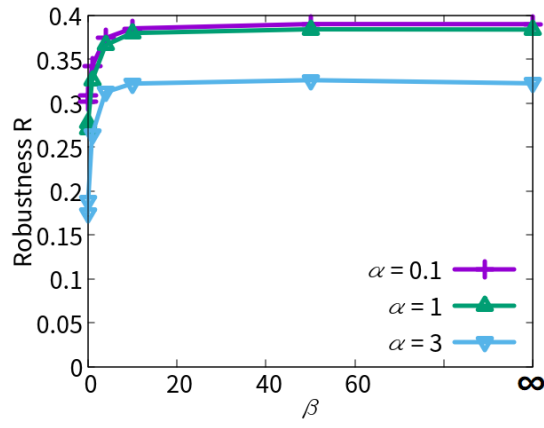
布として表される分布の分布の裾野は直線的である。 $\beta \rightarrow \infty$ に対して、緑色の三角形印とオレンジ色の逆三角形印の傾きが、 $\beta = 1$ に対する紫色の十字印と青緑色の星印の傾きよりも急である。したがって、大きな次数のノードが存在しないため、 β が増加するにつれて大きな次数が制限され、強い頑健性となる。

	RA: $\beta = 0$	$0 < \beta < \infty$	kmin: $\beta \rightarrow \infty$	Ratio of RA
p-model0	Not Onion-like (D-D: SF at $\alpha \rightarrow \infty$)	Onion-like	Onion-like	0
p-model1	Not Onion-like	Onion-like	Onion-like	$1/m$
p-model2	Not Onion-like	Onion-like	Onion-like	0.5
M-MED model	Not Onion-like (MED-rand)	Onion-like	Onion-like (MED-kmin)	0.5

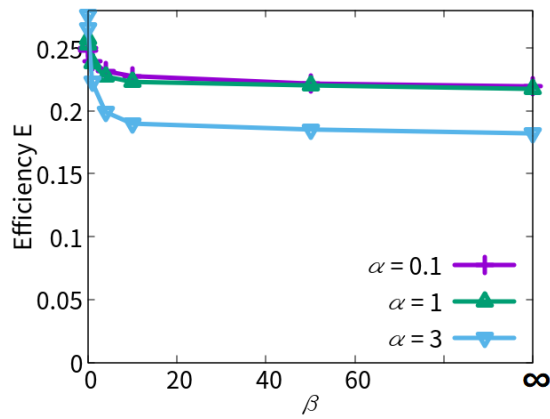
表 6.2: 修正された p-モデルと MED モデルにおける β 値の影響により、玉ねぎ状ネットワークが現れる。MED-rand と MED-kmin は、それぞれ $\beta = 0$ と $\beta = \infty$ の特殊な例である。RA と D-D は、ランダムアタッチメントと複写モデルを表す。



(a) Assortativity, p-model0

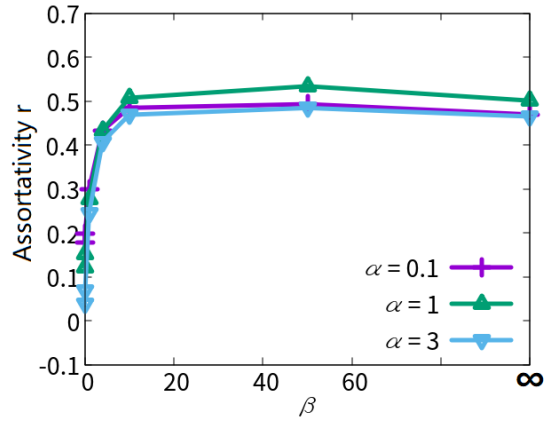


(b) Robustness, p-model0

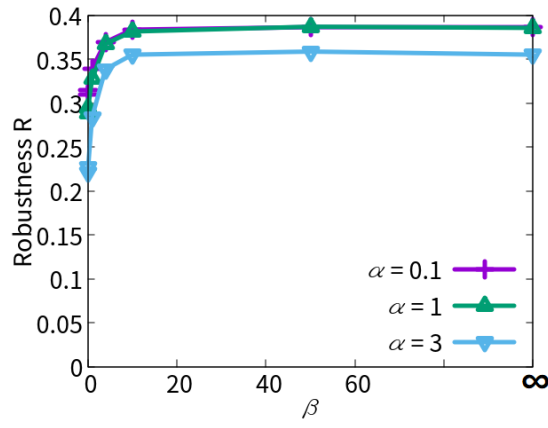


(c) Efficiency, p-model0

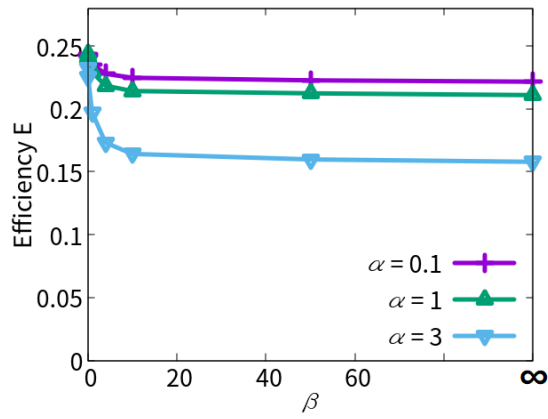
図 6.13: p-model0 における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。



(a) Assortativity, p-model1

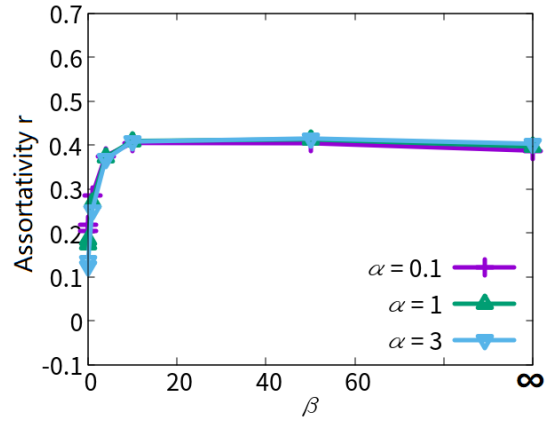


(b) Robustness, p-model1

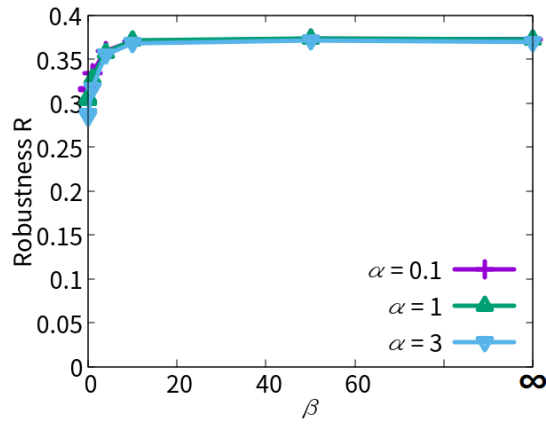


(c) Efficiency, p-model1

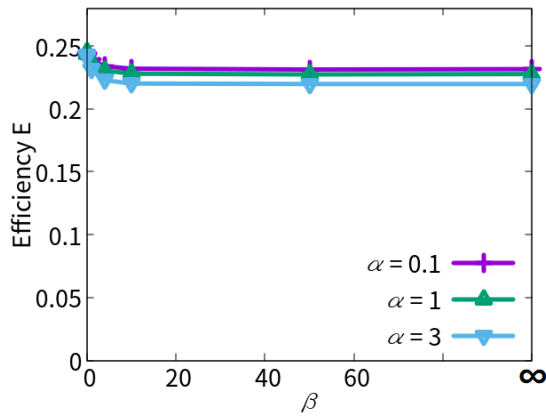
図 6.14: p-model1 における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。



(a) Assortativity, p-model2

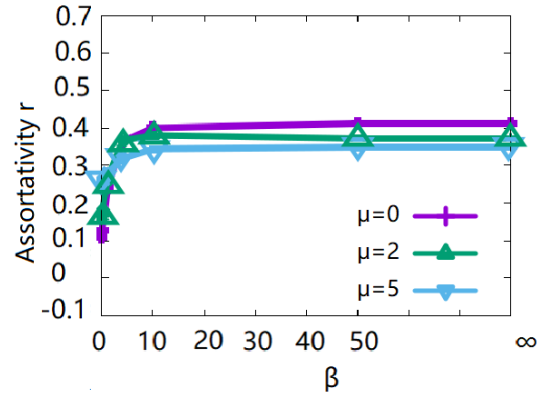


(b) Robustness, p-model2

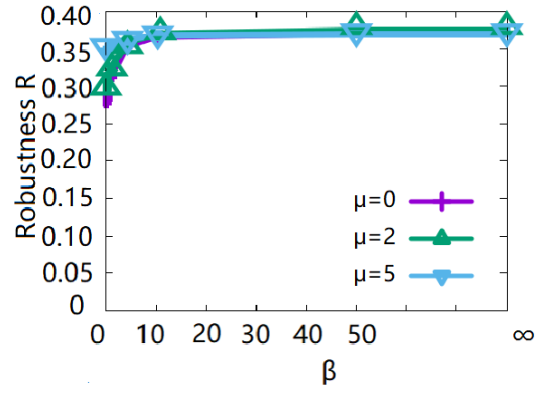


(c) Efficiency, p-model2

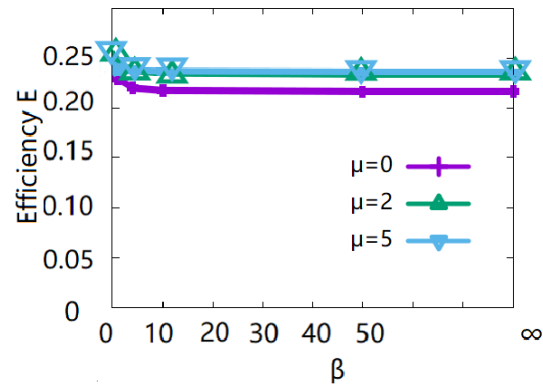
図 6.15: p-model2 における 3 つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。



(a) Assortativity, M-MED



(b) Robustness, M-MED



(c) Efficiency, M-MED

図 6.16: M-MED における 3つの基本的な指標 r (上)、 R (中)、 E (下) をパラメータ β に対してプロットしたグラフ。サイズ $N = 5000$ で、時間ステップあたりのリンク数 m は 4 本。

図 6.13 と 6.14(b) において、同じ色の線や印と比較して、 $m = 2$ の (a) よりも、より多くのリンクを追加して玉ねぎ状のネットワークを形成するため、効率が高くなっている。

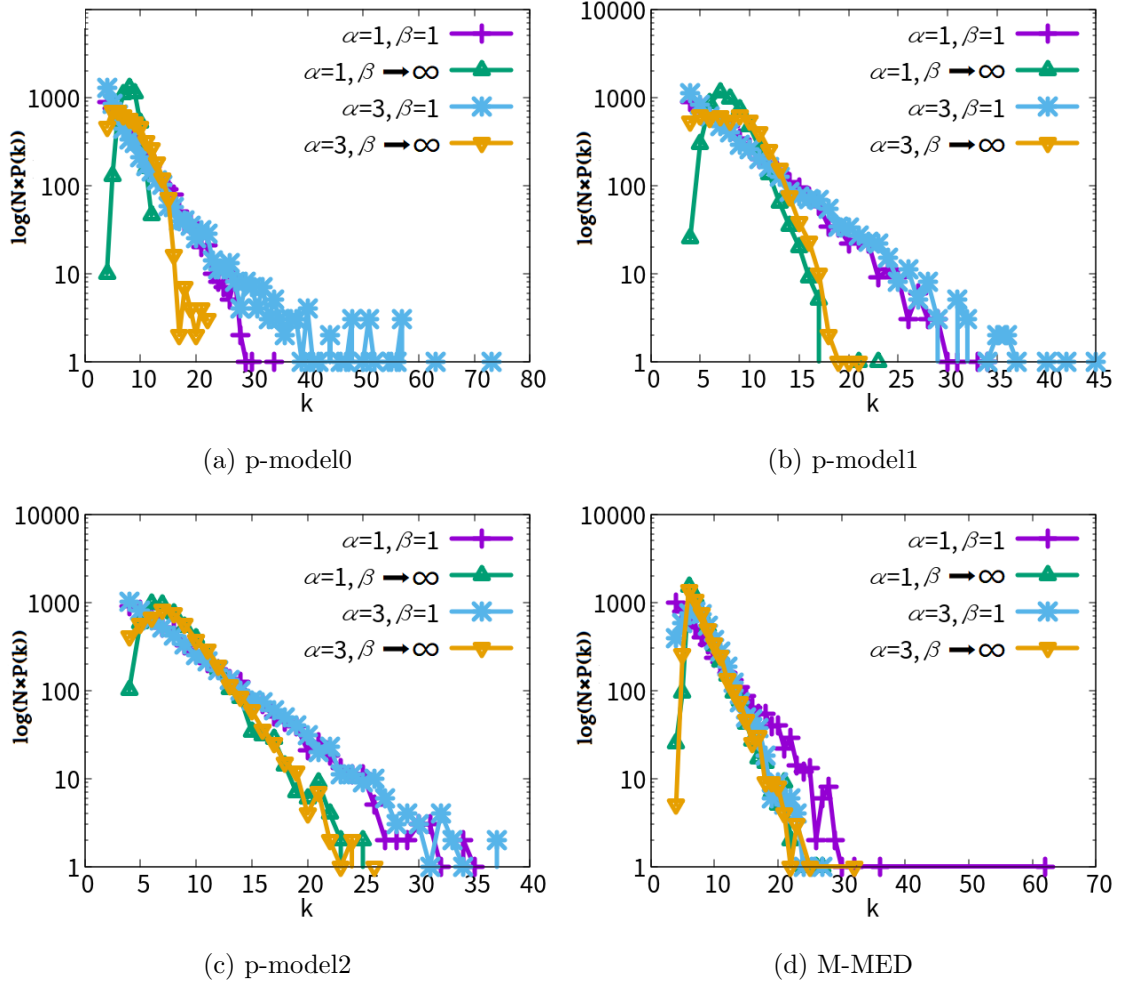


図 6.17: 成長するネットワークにおける 4 つの修正モデルの次数分布は、 $m = 4$ で似たような分布を持っている。各々の (a) (b) (c) (d) において、p-model0、p-model1、p-model2、M-MED モデルで似たような分布がある。分布の裾は、半対数プロットで指数分布として線形である。 $\beta \rightarrow \infty$ の緑の線（三角形印）とオレンジの線（逆三角形印）の傾きは、 $\beta = 1$ の紫の線（十字印）とシアン（アスタリスク印）の傾きよりも急である。したがって、大きな次数は β が増加するにつれて制限される。つまり、ハブは存在しない。

6.4 IPA モデル上のIMP のシミュレーション結果

Regular-like ネットワークの拡散力を調べるために、SF ネットワークとなる BA モデルと $\beta < 0$ の IPA モデルで生成した [Liao3] ネットワークを使って拡散処理のシミュレーションを行った。図 6.19 と図 6.20、6.21、6.22、6.23 を比較することで、どの IMP 方法を使っても、図 6.19 の $R(\infty)^{SF}$ 値は他の図よりも高い。従って、下の式が成立。

$$R(\infty)^{SF} > R(\infty)^{\beta=-10} \approx R(\infty)^{\beta=-20} \approx R(\infty)^{\beta=-50} \approx R(\infty)^{\beta=-100} \quad (6.14)$$

$$N_S^{SF} < N_S^{\beta=-10} \approx N_S^{\beta=-20} \approx N_S^{\beta=-50} \approx N_S^{\beta=-100} \quad (6.15)$$

ここの $R(\infty)^{SF}$ は、SF ネットワークの累感染が収束する十分大きい時間 ($t \rightarrow \infty$) までの累積感染数である。そこで、

$$\frac{R(\infty)^{SF}}{N_S^{SF}} > \frac{R(\infty)^{\beta=-10}}{N_S^{\beta=-10}} \approx \frac{R(\infty)^{\beta=-20}}{N_S^{\beta=-20}} \approx \frac{R(\infty)^{\beta=-50}}{N_S^{\beta=-50}} \approx \frac{R(\infty)^{\beta=-100}}{N_S^{\beta=-100}} \quad (6.16)$$

式 (6.16) から、SF ネットワークに対して、ある種の平均感染能力は IPA モデルよりも強力であると言える。更に、図 6.19 の緑色 (Degree 法)、青色 (k-shell 法)、黄色 (LC 法) の線が他の図よりも明らかに上にあることが分かる。図 6.18 のような、SF ネットワークが非均等な次数分布と集中的な拡散力を持ち、結合耐性は極端に脆弱である。IPA ネットワークが均等な分布と分散的な拡散力を持ち、結合耐性は非常に強い。

一方、従来の中心性で種ノードを決める IMP 解法 [44, 45, 46, 47, 48, 40] との比較において、提案手法を用いて（四角形印を付けた線）得られた種の感染能力が一番高いことを図 6.19 から図 6.23 は示している。その理由として、提案手法は拡散の重なり問題を回避させ、従来法を凌駕する結果が得られたと考えられる。

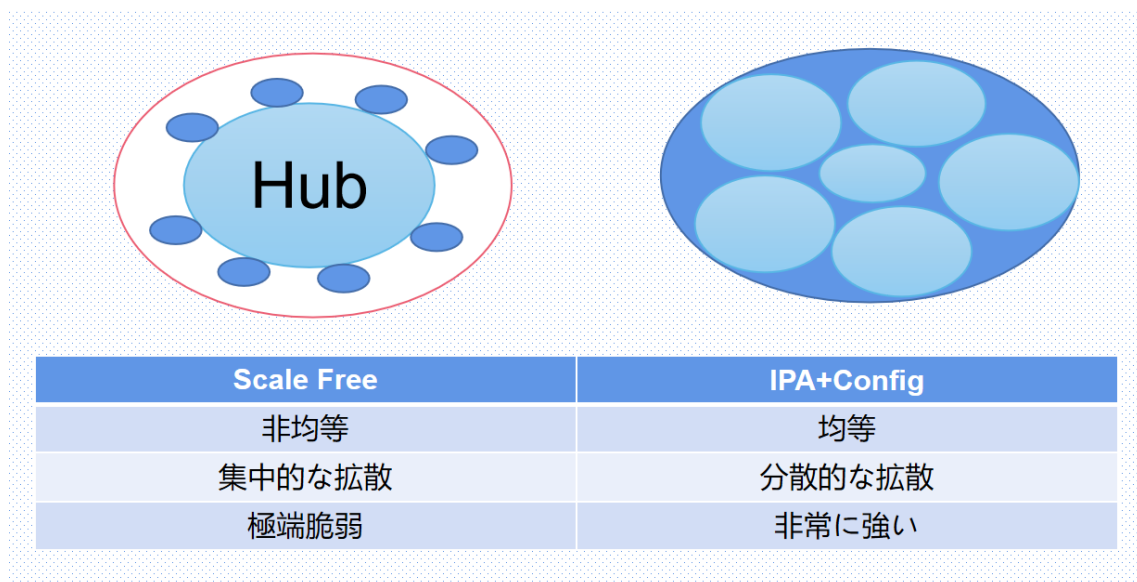
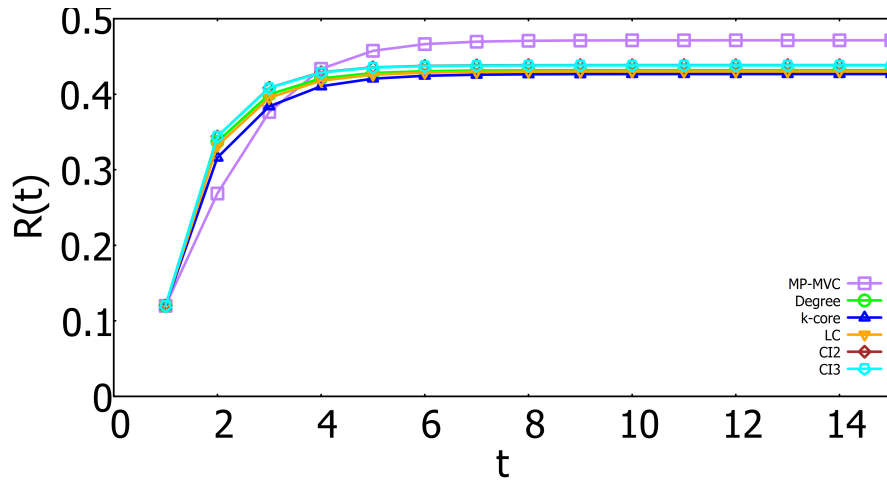
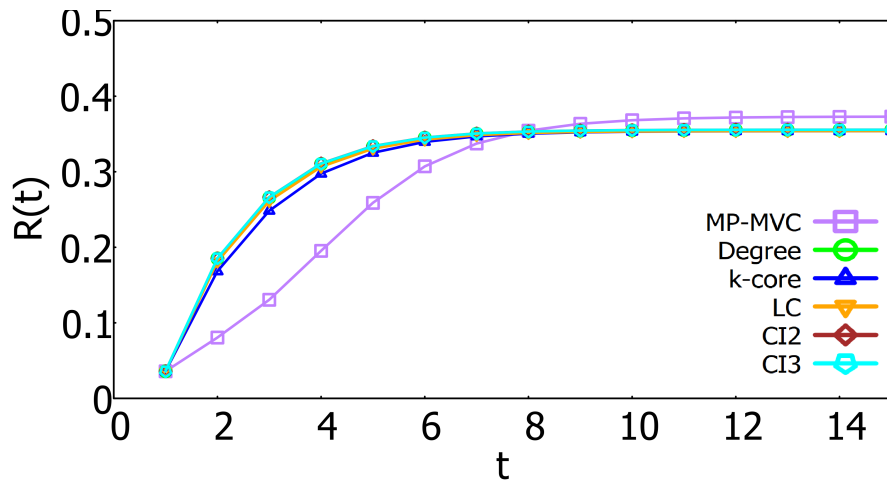


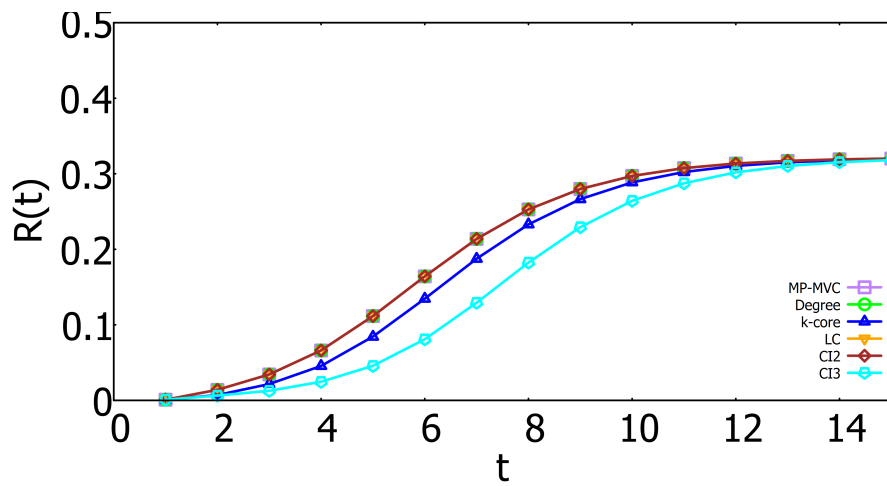
図 6.18: SF ネットワークと IPA ネットワークの比較



(a) $l = 2$

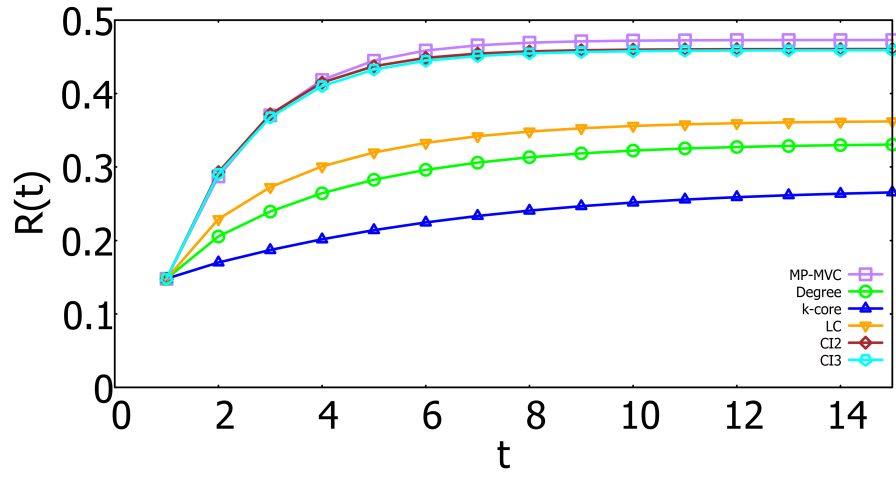


(b) $l = 3$

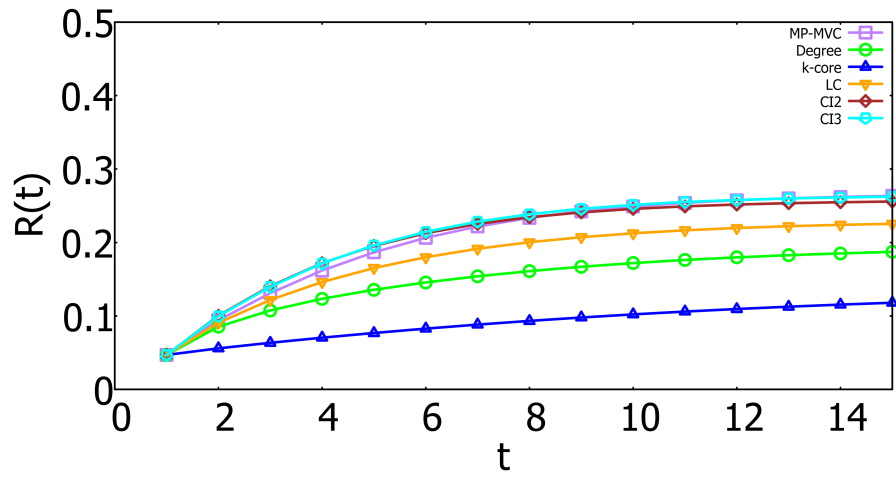


(c) $l = 4$

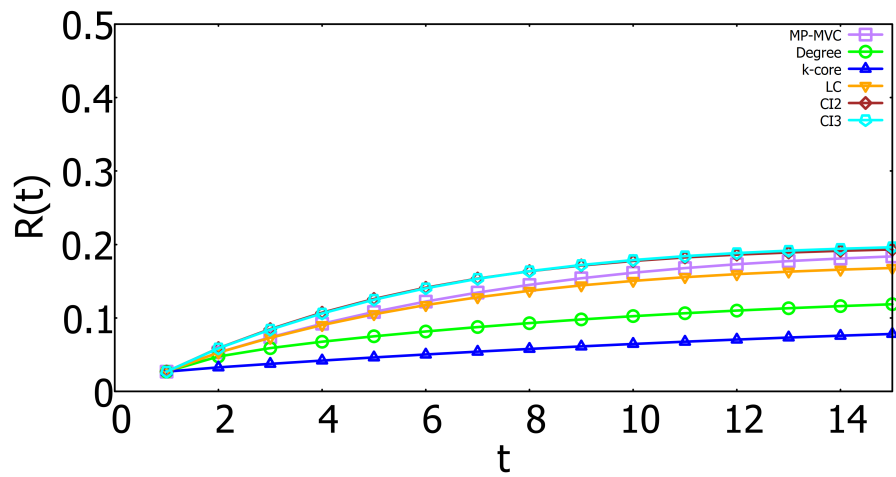
図 6.19: β は 1 の時、 $l = 2, 3, 4$ で、SF ネットワーク上各 IMP 解法のシミュレーション結果。



(a) $l = 2$

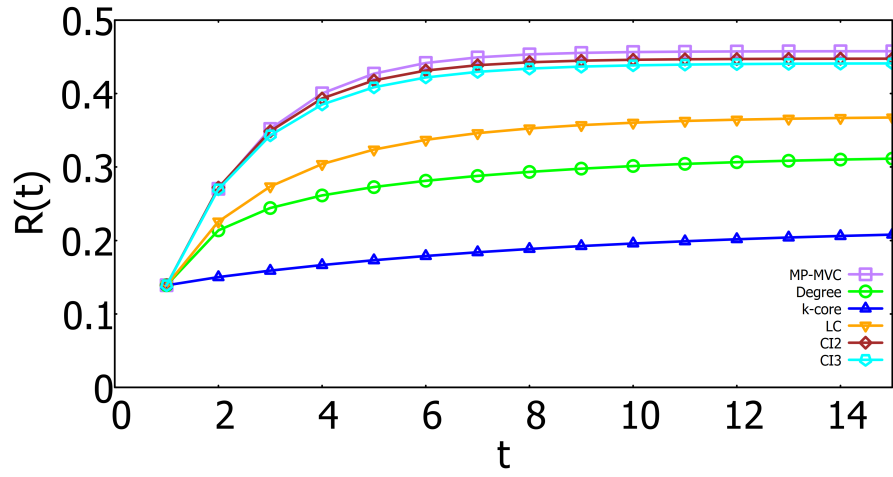


(b) $l = 3$

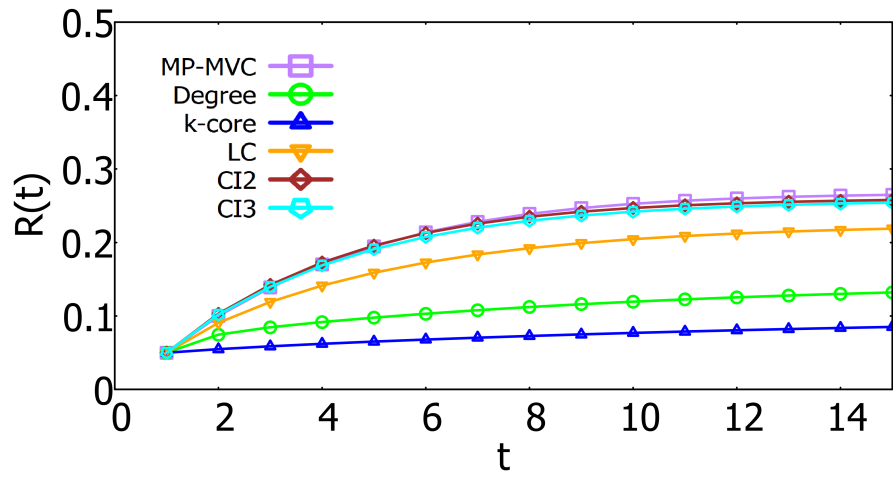


(c) $l = 4$

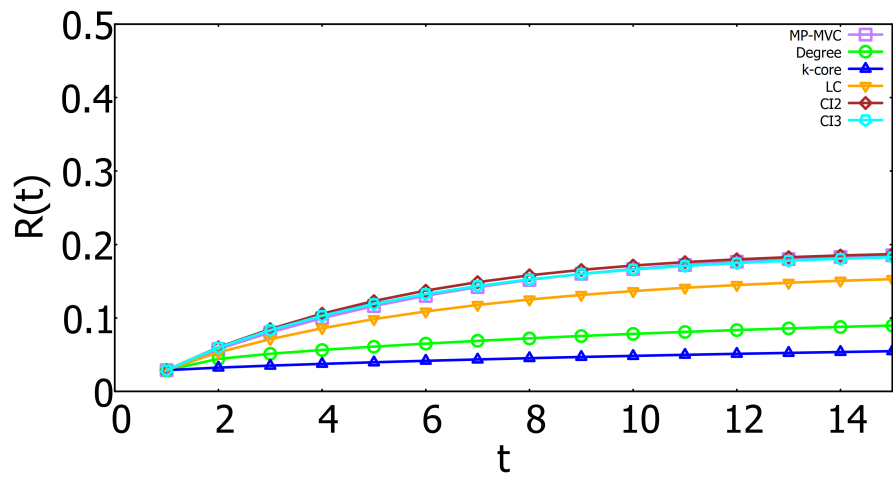
図 6.20: β は-10 の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各IMP 解法のシミュレーション結果。



(a) $l = 2$

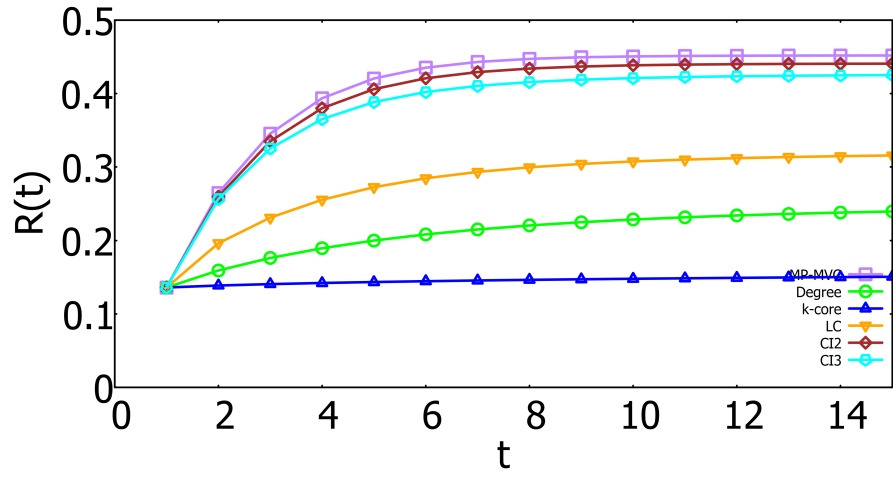


(b) $l = 3$

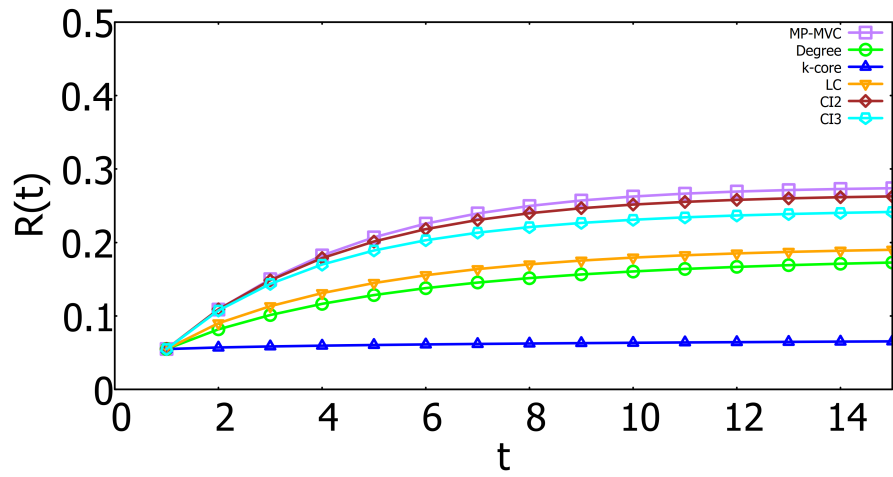


(c) $l = 4$

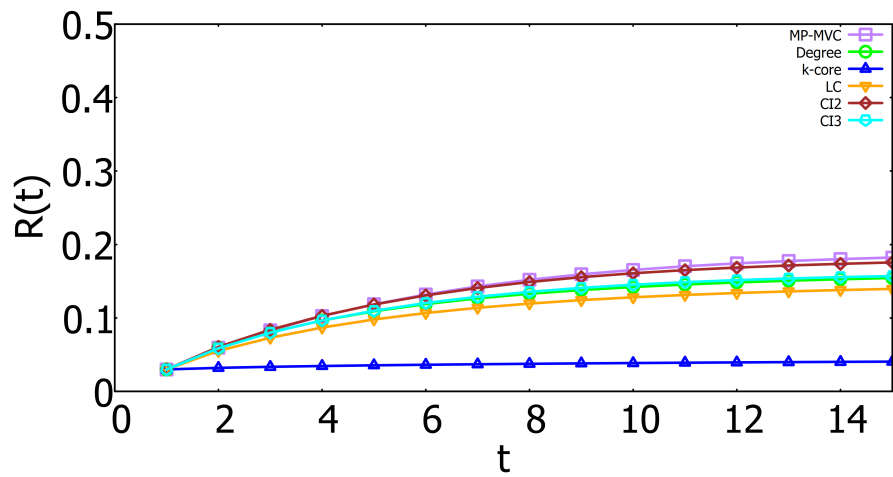
図 6.21: β は-20 の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各IMP 解法のシミュレーション結果。



(a) $l = 2$

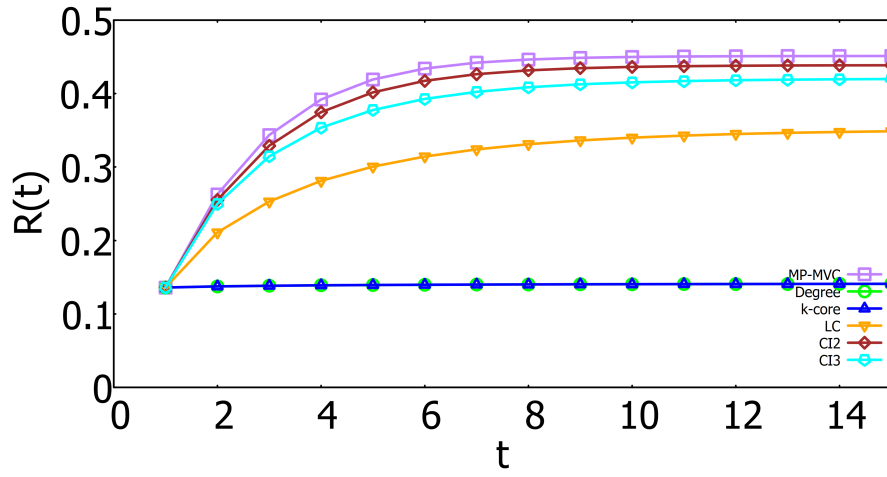


(b) $l = 3$

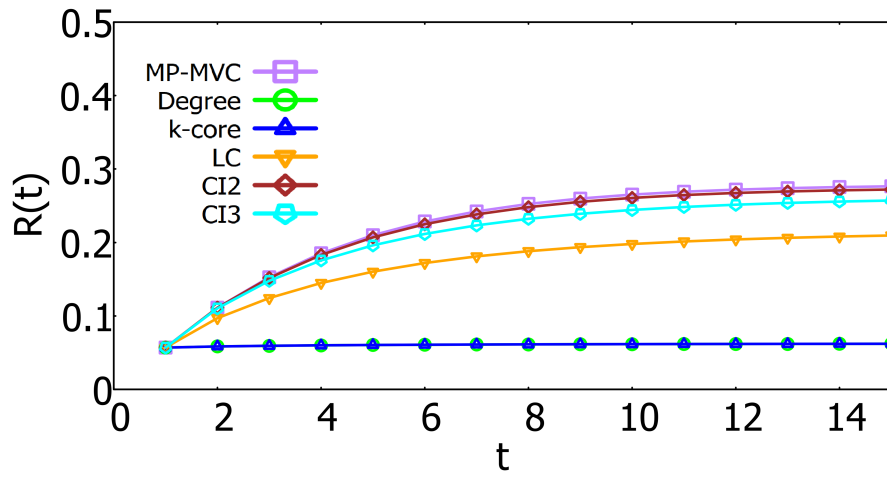


(c) $l = 4$

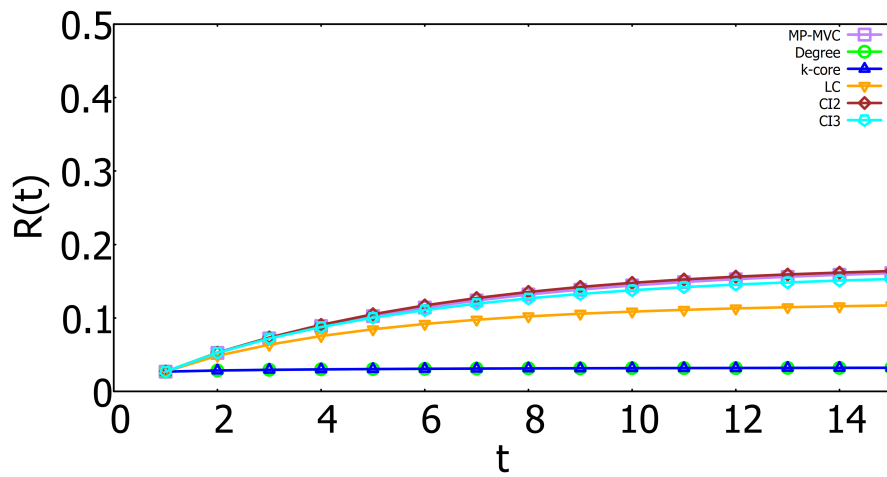
図 6.22: β は-50 の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各IMP 解法のシミュレーション結果。



(a) $l = 2$



(b) $l = 3$



(c) $l = 4$

図 6.23: β は-100 の時、 $l = 2, 3, 4$ で、IPA ネットワーク上各 IMP 解法のシミュレーション結果。

6.5 本章のまとめ

パラメータ $\beta \geq 0$ による逆優先選択として、ランダムアタッチメントと最小次数アタッチメントを補間する $k^{-\beta}$ アタッチメントを用いて、強く頑健な玉ねぎ状ネットワークを生成できることを示した。特に、 $\beta > 0$ が大きい場合、 $k^{-\beta}$ アタッチメントが高い次数相関と頑健性を持つ玉ねぎ状ネットワークを生成する。しかしながら、 $\beta \rightarrow \infty$ では効率的でない $O(N)$ の最長経路（直径として）を持つチェーン構造が得られる。そのため、ランダムアタッチメントと $k^{-\beta}$ アタッチメントの混合を考え、平均経路長が $O(\log N)$ になるよう、効率的な経路の出現にはわずかなランダムアタッチメントが必要であることが分かった。また、 $k^{-\beta}$ アタッチメントを用いた玉ねぎ状ネットワークの性質について、p-model [2] および MED model [3, 4] を改良して、 $k^{-\beta}$ アタッチメントと距離 $d = \mu + 1$ の効果を調べた。得られた結果から、広範囲の $\beta > 0$ で $k^{-\beta}$ アタッチメントを用いて玉ねぎ状ネットワークを生成でき、ランダムに選ばれたノードからの距離はあまり影響を与えないことも分かった。これらの結果は、玉ねぎ状ネットワークの生成に関する理解を深めることに貢献する。また、6.4 節についても、IPA モデル上の IMP のシミュレーション結果を示した。IMP に対する、提案手法は従来法よりより良い結果が得られた。

第7章 結言

7.1 研究成果の要約

本研究は、拡散の重なりを回避できるの IMP 問題に対する近似解法の提案と、これまで最適耐性を持つと考えられてきた玉ねぎ状ネットワーク形成する要因の解明を目的として、筆者のおこなってきた研究成果をまとめたものである。以下、各章を要約する。

第一章では、現代社会において欠かせない大規模かつ複雑なネットワークの繋がり構造が攻撃に対して脆弱な問題について指摘した。そこで、頑健なネットワーク構造の必要性と IMP 問題について説明し、本研究の目的を述べ、各章の構造を示した。

第二章では、複雑ネットワークの基本的な概念を述べた後、主な解析指標である、次数、三角形数とクラスタリング係数、ネットワーク上のパーコレーション、近接中心性と媒介中心性、次数相関、頑健性、通信効率について述べた。

第三章では、最小頂点被覆問題の定義を説明した後、コンピューター科学アプローチからの 2-近似アルゴリズムとより良い統計物理のアプローチからの SP 法を説明し、SP 法の同時計算を提案し、SP と 2-近似法による最小頂点被覆を予備実験として比較した。

第四章では、影響力最大化問題 (IMP) に関する、独立カスケード (IC) モデルや線形閾値 (LT) モデルについて紹介した。そして、従来拡散処理では、多くの試行 (平均化) が必要であるが、計算時間を短縮するために、統計物理のメッセージ伝搬に基づいて、ネットワーク科学における集団的影響力 (Collective influence) から着想を得た高速計算法を提案した。この高速計算法は平均化が不要で、従来法に比べて数倍速くなることを示した。

第五章では、次数相関ではなく、ループと頑健性のより本質的な関係が明らかにされつつある。しかしながら、従来のヒューリスティックな手法は、ループと拡散の重なりを考慮しない。そこで、ループと拡散の重なりを着目して、情報科学

のマルチホップ被覆と、統計物理学的なメッセージ伝搬による近似解法を融合して、新たな方法を提案した。種間の距離を長くすることで、拡散の重なり問題を回避させた。提案手法と従来手法の比較結果を示し、提案手法は拡散の重なり問題を回避させ、従来法を凌駕する結果が得られたと考えられる。

第六章では、次数分布を固定した時、リンク両端のノードの次数の相関係数が正の値を持つ玉葱状ネットワークが最適な頑健性を持つことが明らかになっているが、玉葱状ネットワークを生成するために、色々な手法が多く提案されてきた (MED-Rand/Kmin) ことに着目する。特に、ループ強化を促す、次数分布の幅が狭いほど頑健性が高くなると考えられ、分布を底上げする最小次数（逆優先）選択は重要な役割を果たすと考え、逆優先選択の有効性を明らかにした。そして、ランダムアタッチメントと最小次数のアタッチメントの間に連続補間を導入し、頑健性、次数相関、効率性、複数の種攻撃を比較することで、逆優先選択が頑健性と効率性の向上に重要であることを示した。玉葱状構造の生成に対して、結合距離より逆優先選択は重要な要因であることが分かった。一方、我々は、分断されやすいチェーン構造を避けるために、ランダムアタッチメントと $k^{-\beta}$ アタッチメントの混合を考え、平均経路長が $O(\log N)$ になるよう、効率的な経路の出現にはわずかなランダムアタッチメントが必要であることが分かった。また、 $k^{-\beta}$ アタッチメントを用いた玉ねぎ状ネットワークの性質について、p-model および MED model を改良して、 $k^{-\beta}$ アタッチメントと距離 $d = \mu + 1$ の効果を調べた。得られた結果から、広範囲の $\beta > 0$ で $k^{-\beta}$ アタッチメントを用いて玉ねぎ状ネットワークを生成でき、ランダムに選ばれたノードからの距離はあまり影響を与えないことも分かった。

7.2 今後の課題

本研究の今後の課題について述べる。IMP に対する提案手法は、マルチホップ被覆において、重なりを避けるために最適なホップ数はまだ解決されていない。今後の研究では、この問題を考慮し、IMP に最も効果的なホップ数を示す最適な数値を探ることが考えられる。優先的選択における Rich get Richer 原理の代わりに、少量のランダムなアタッチメントがノードに等しく結合の機会を提供する。一方、最小次数アタッチメントは、小さな次数を持つ貧弱で有用性の低いノードに結合するための役立つ機会を提供して、ネットワークに対する頑健性を高めるこ

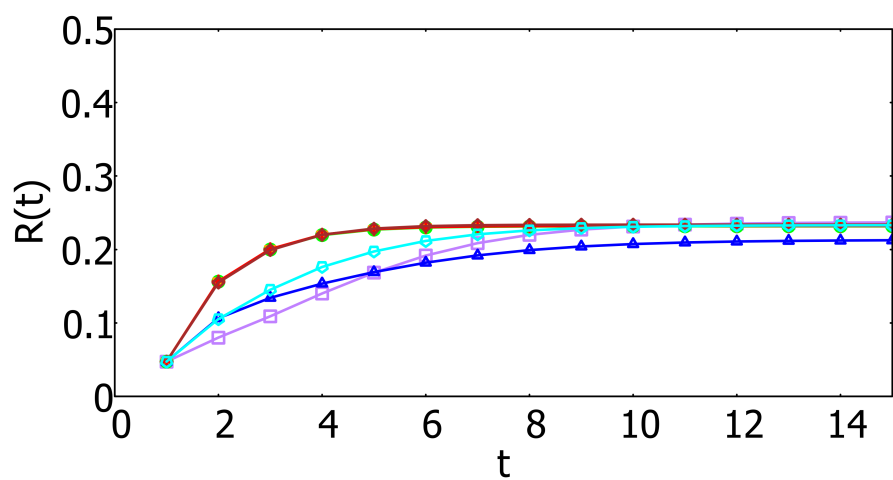
とに繋がる。このような結論は、将来のシステムで強い頑健性と高い通信効率を兼ね備えたより良いネットワークの実現に役立つ可能性がある。

付録

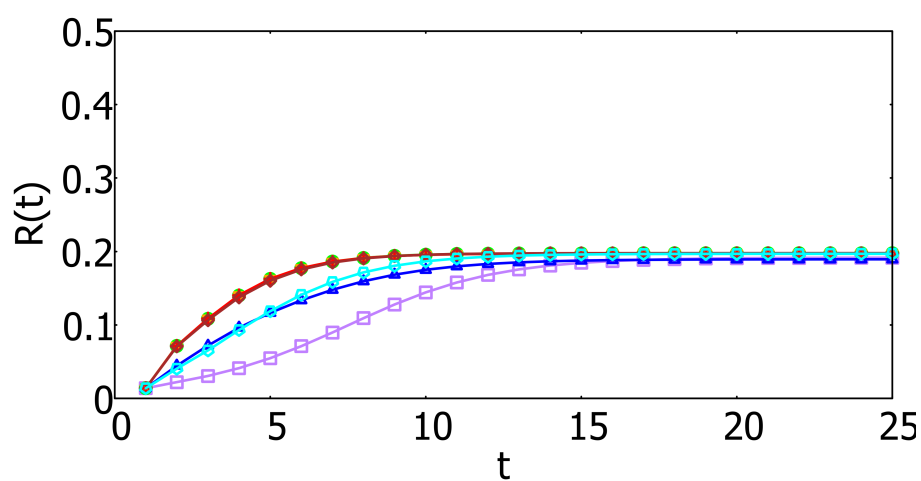
図 7.1 から 7.14 から分かるように、紫色の線は常に最も高い位置にある。これは、私たちが提案した手法が、ネットワークでも累積感染数 $R(t)$ 値が高く優れた伝搬効果を持っていることを意味している。下の表はパラメータの組合せである。

図	ネット名	λ 値	β 値	横軸	縦軸
図 7.1	Polblogs	2	0.12	t	$R(t)$
図 7.2	Polblogs	2~10	0.12~0.6	λ	$R(t_c)$
図 7.3	Airtraffic	2	0.12	t	$R(t)$
図 7.4	Airtraffic	2~10	0.12~0.6	λ	$R(t_c)$
図 7.5	Hamster	2	0.12	t	$R(t)$
図 7.6	Hamster	2~10	0.12~0.6	λ	$R(t_c)$
図 7.7	Gnutella	2	0.12	t	$R(t)$
図 7.8	Gnutella	2~10	0.12~0.6	λ	$R(t_c)$
図 7.9	GRQC	2	0.12	t	$R(t)$
図 7.10	GRQC	2~10	0.12~0.6	λ	$R(t_c)$
図 7.11	Hepth	2	0.12	t	$R(t)$
図 7.12	Hepth	2~10	0.12~0.6	λ	$R(t_c)$
図 7.13	PGP	2	0.12	t	$R(t)$
図 7.14	PGP	2~10	0.12~0.6	λ	$R(t_c)$

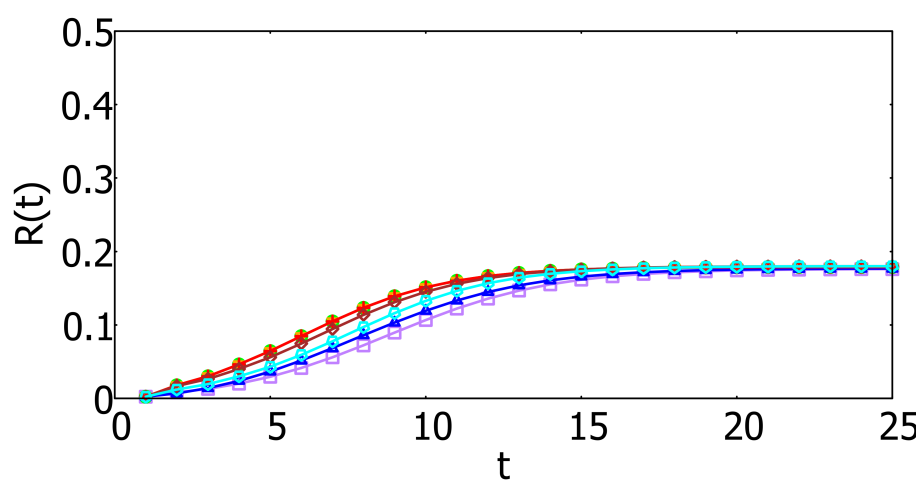
表 7.1: パラメータの組合せ



(a) $l = 2$

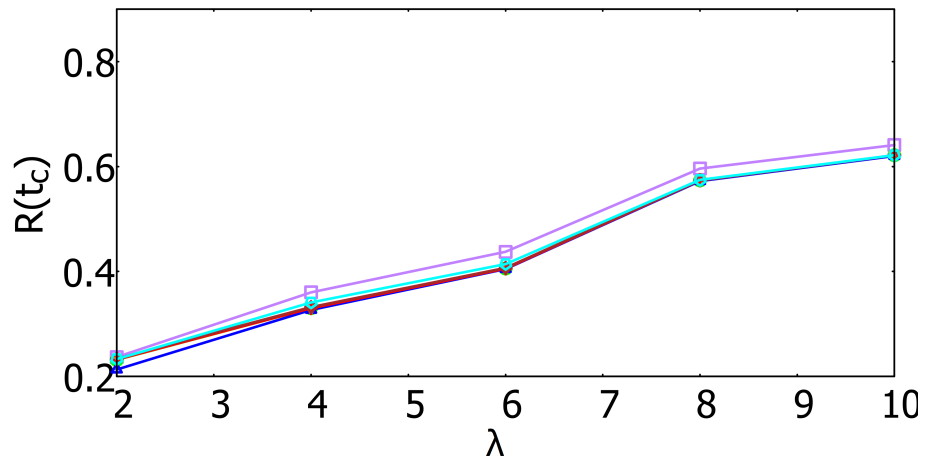


(b) $l = 3$

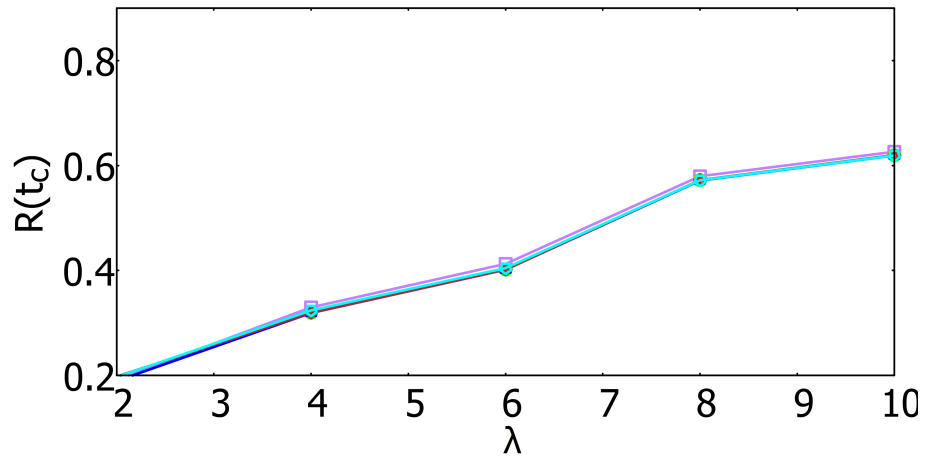


(c) $l = 4$

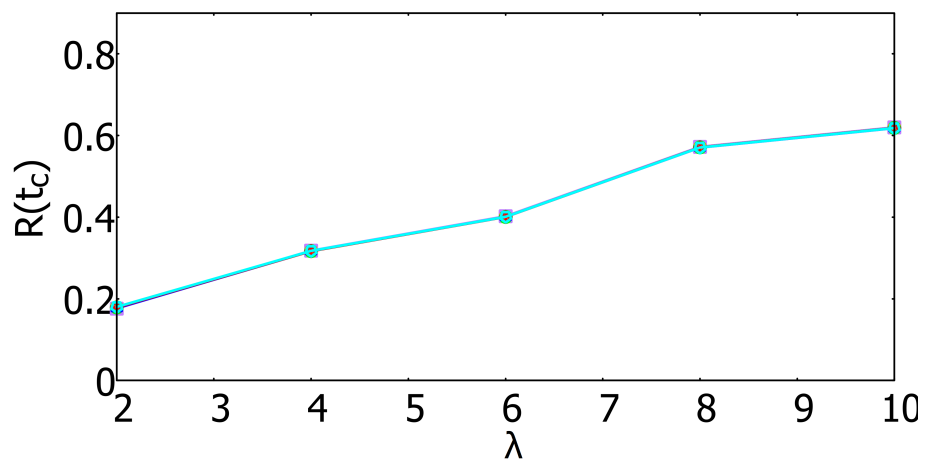
図 7.1: Polblogs の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$

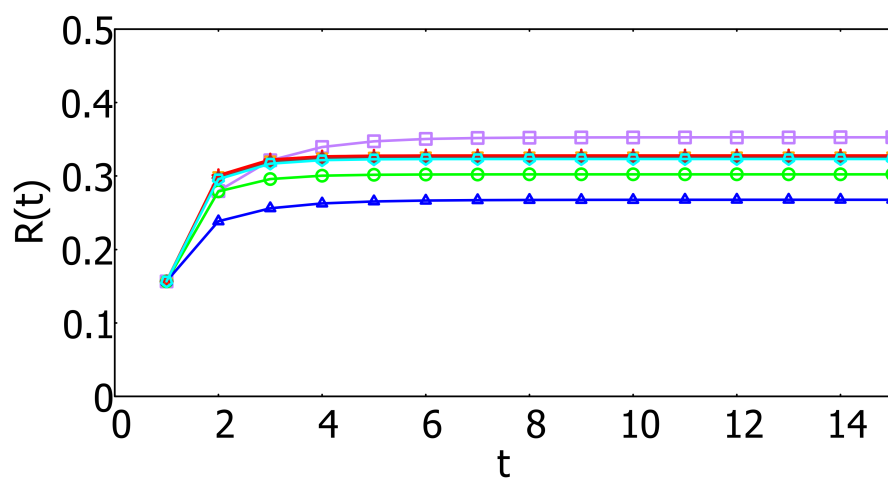


(b) $l = 3$

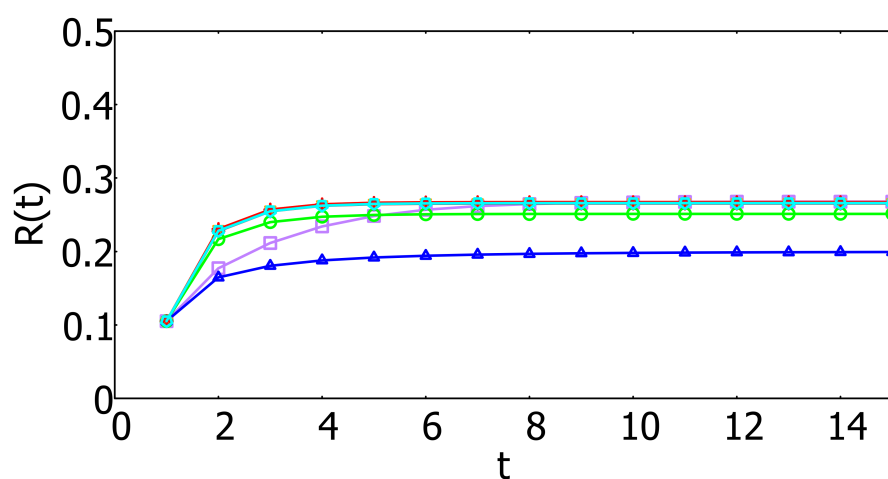


(c) $l = 4$

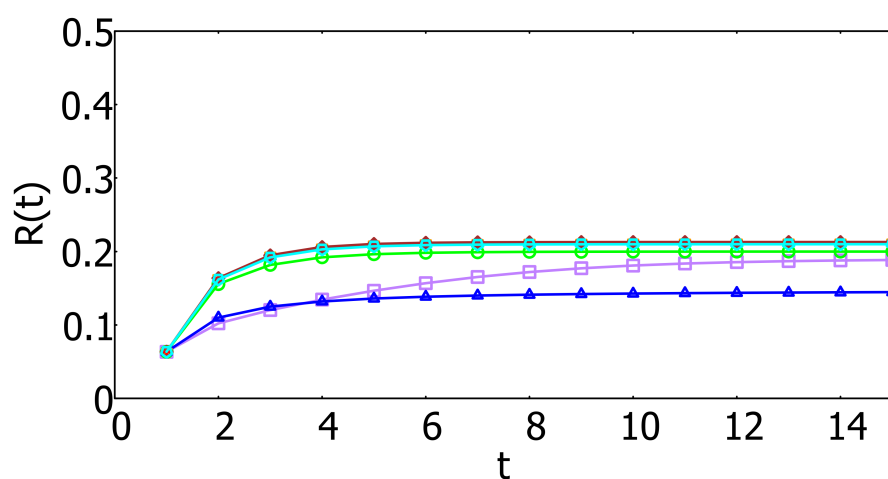
図 7.2: Polblogs の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。



(a) $l = 2$

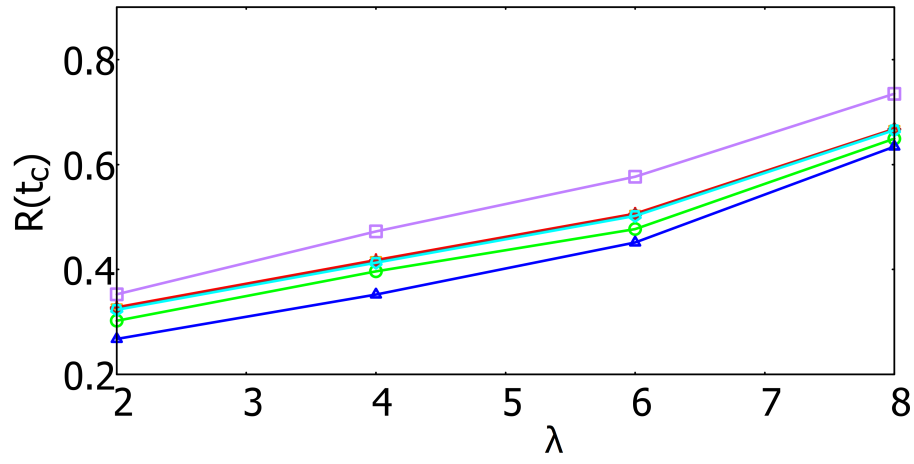


(b) $l = 3$

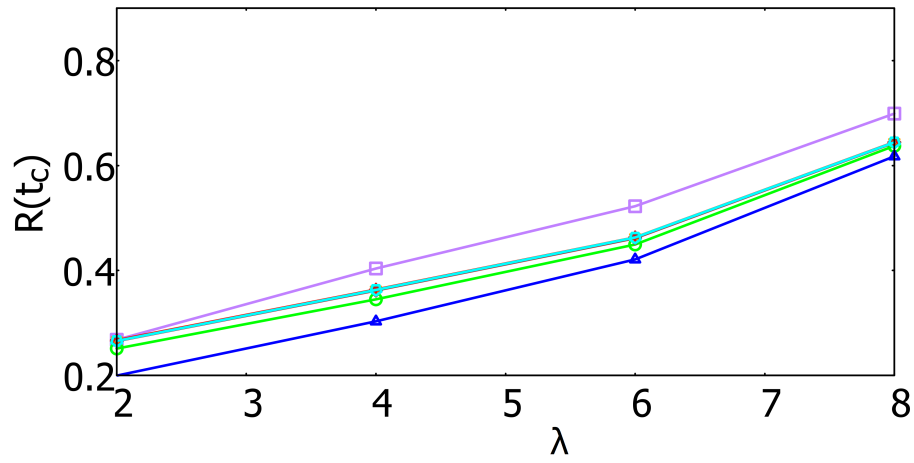


(c) $l = 4$

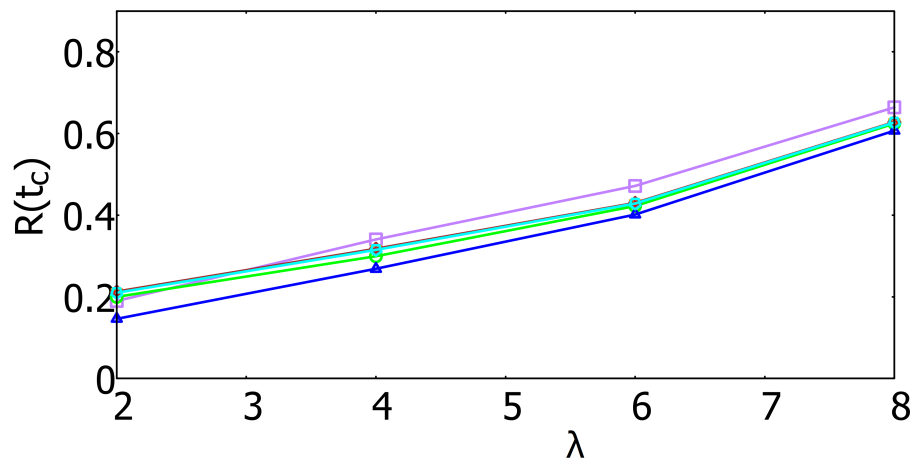
図 7.3: Airtraffic の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$

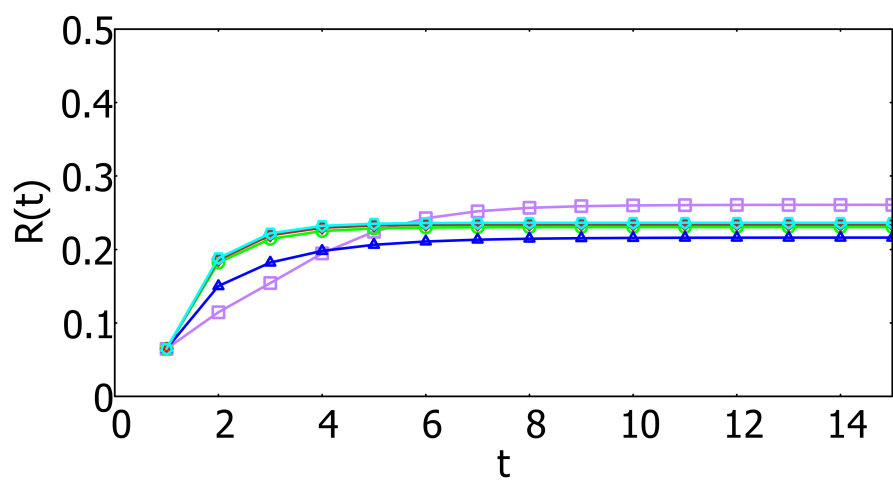


(b) $l = 3$

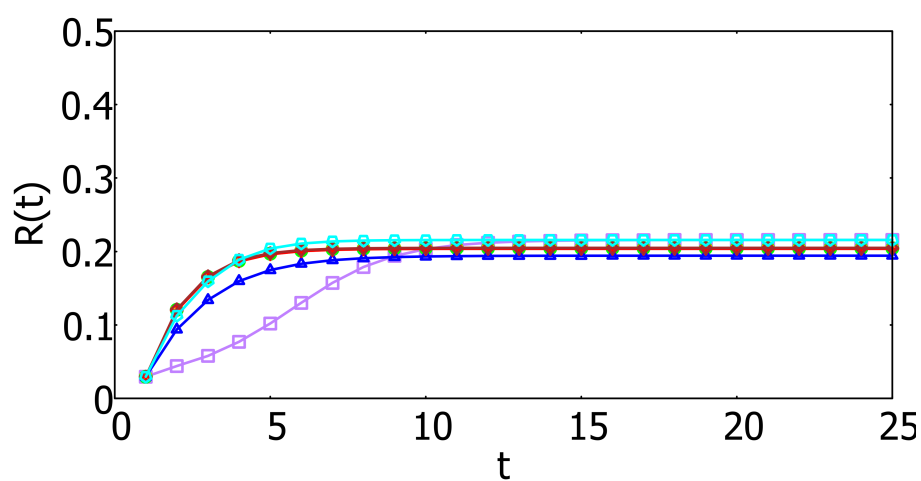


(c) $l = 4$

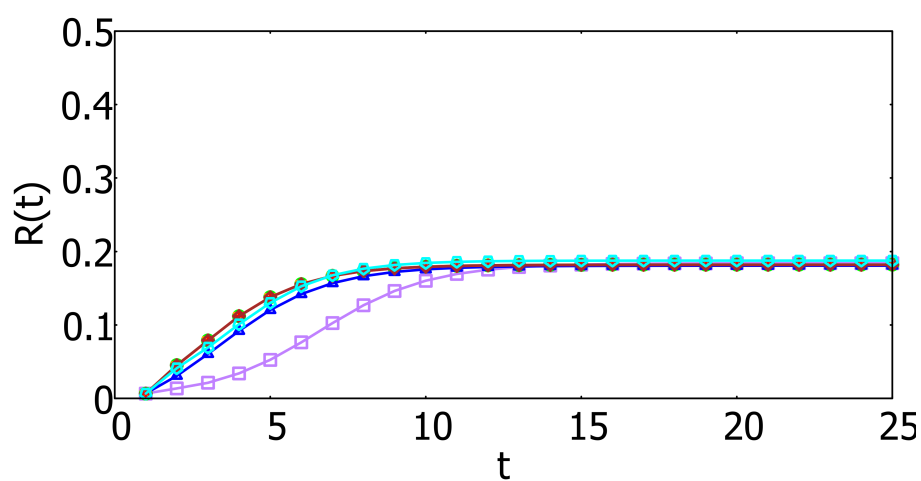
図 7.4: Airtraffic の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。



(a) $l = 2$

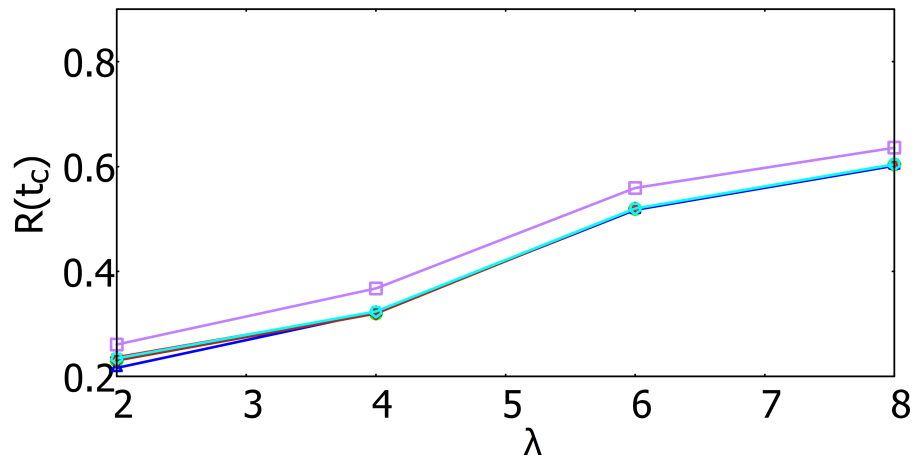


(b) $l = 3$

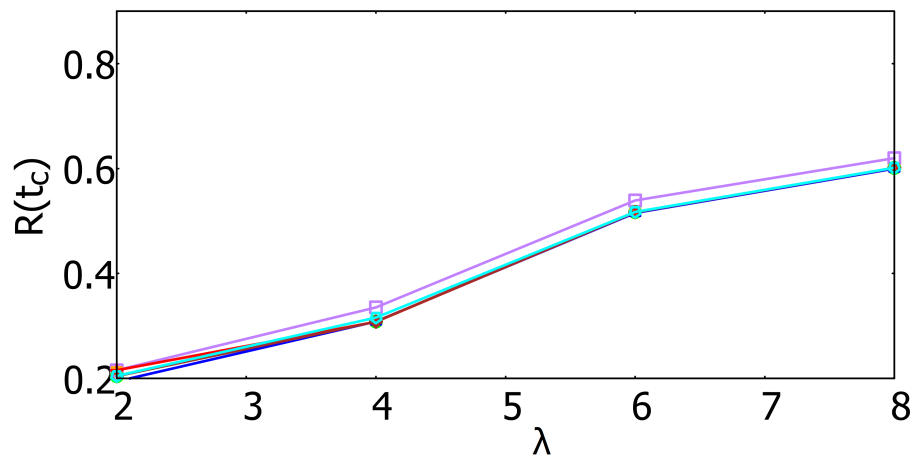


(c) $l = 4$

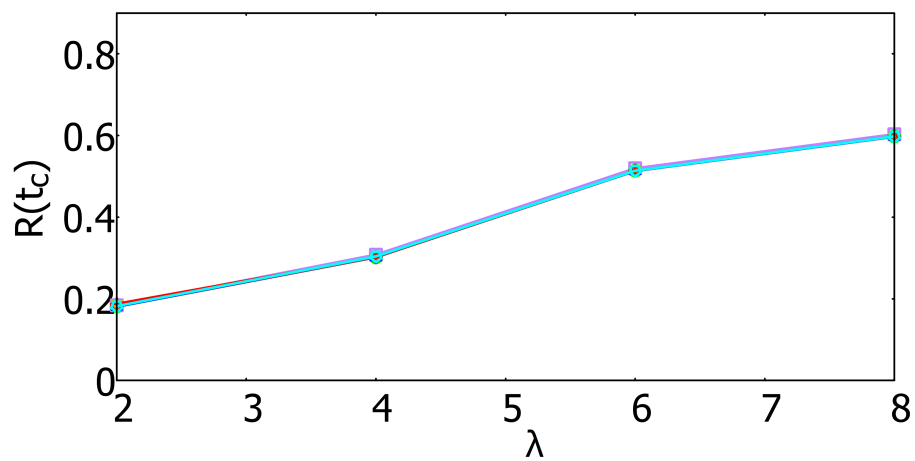
図 7.5: Hamster の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$

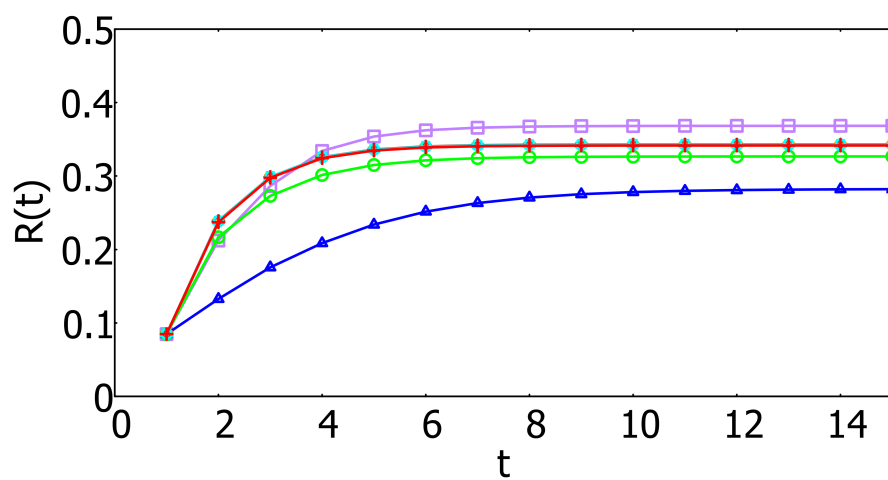


(b) $l = 3$

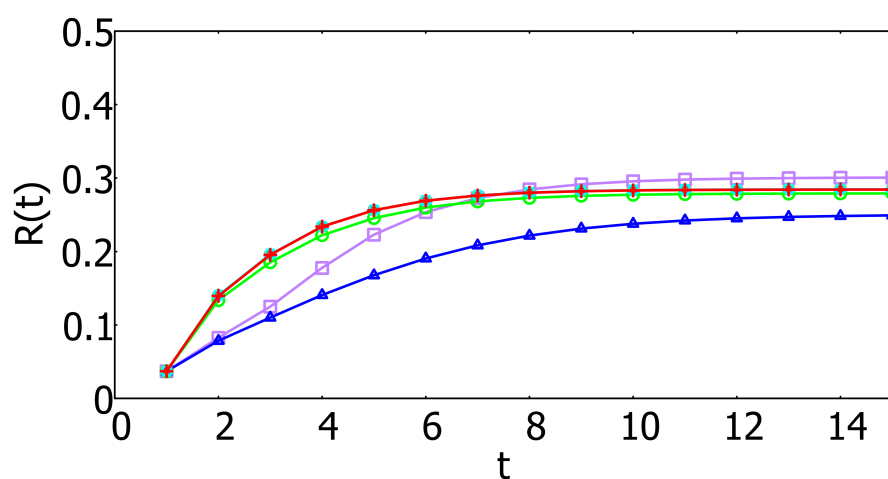


(c) $l = 4$

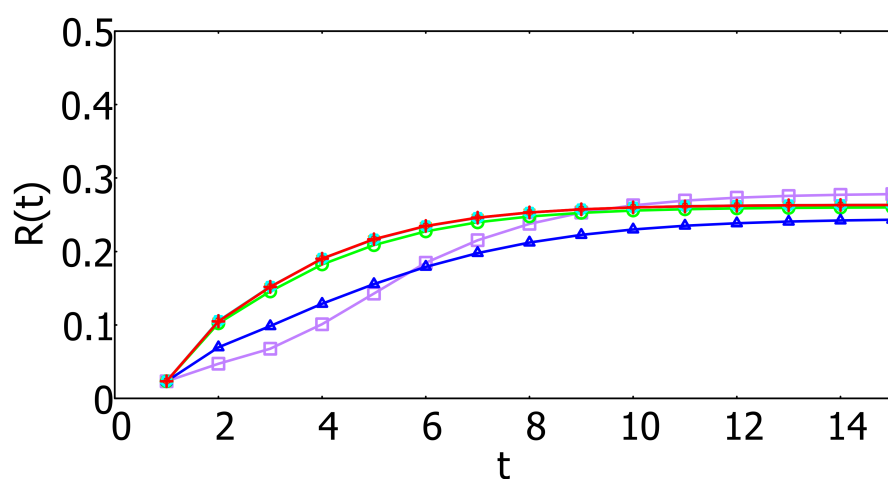
図 7.6: Hamster の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。



(a) $l = 2$

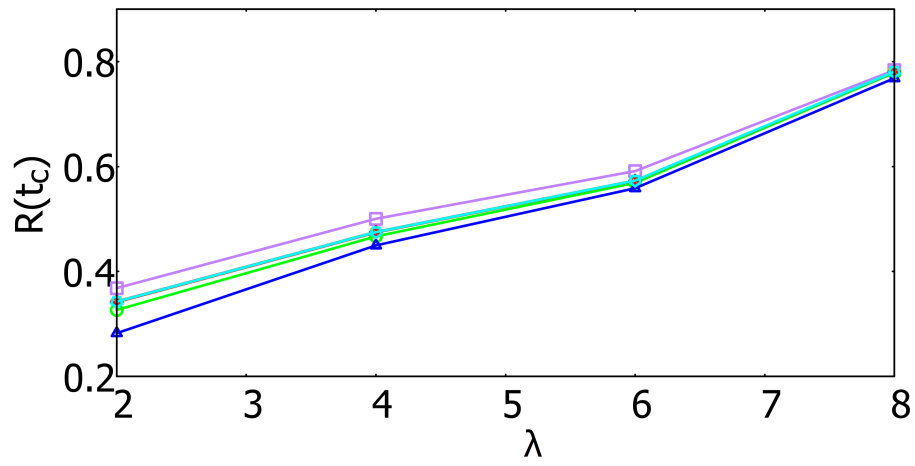


(b) $l = 3$

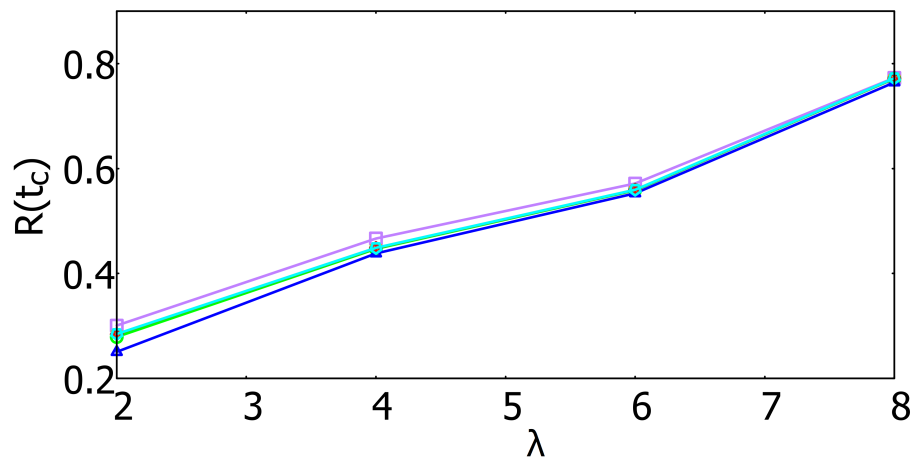


(c) $l = 4$

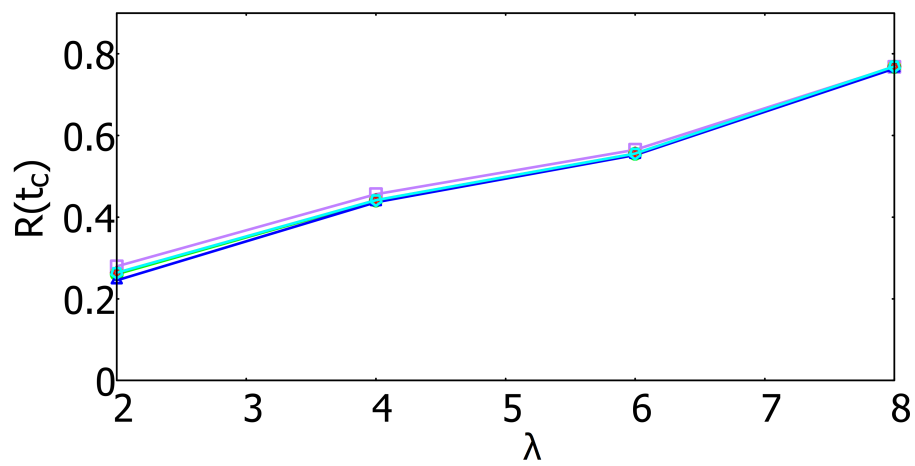
図 7.7: Gnutella の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$

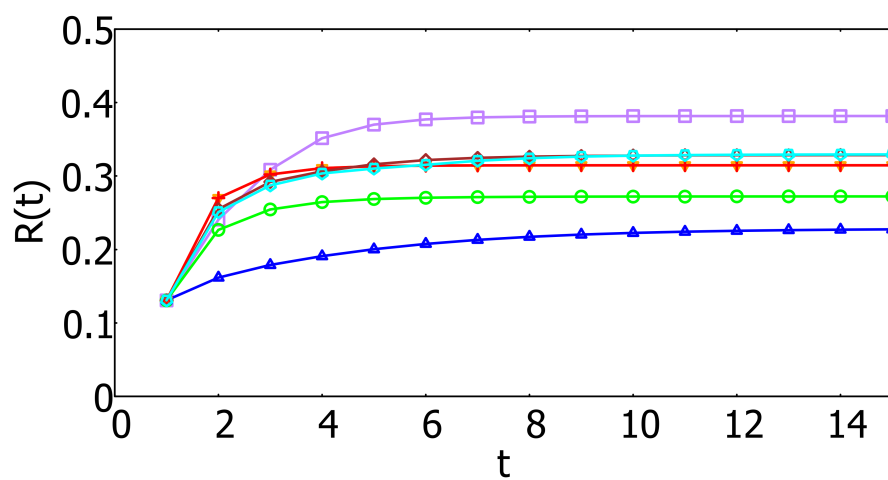


(b) $l = 3$

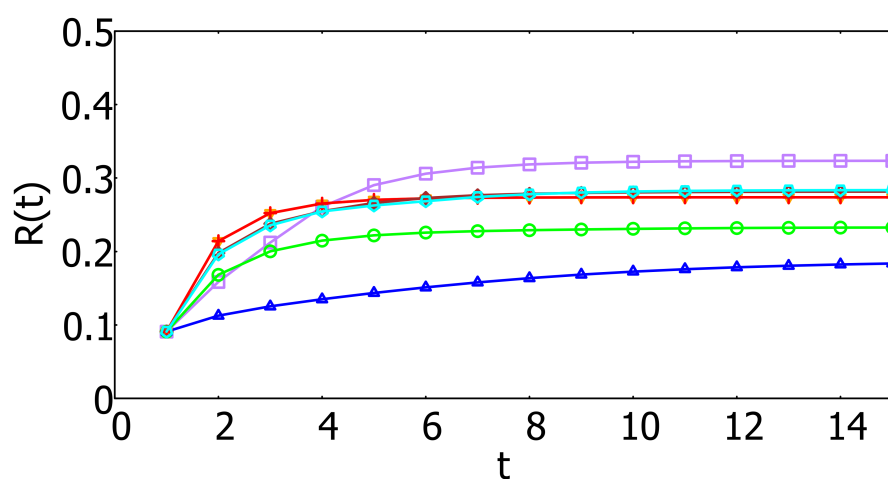


(c) $l = 4$

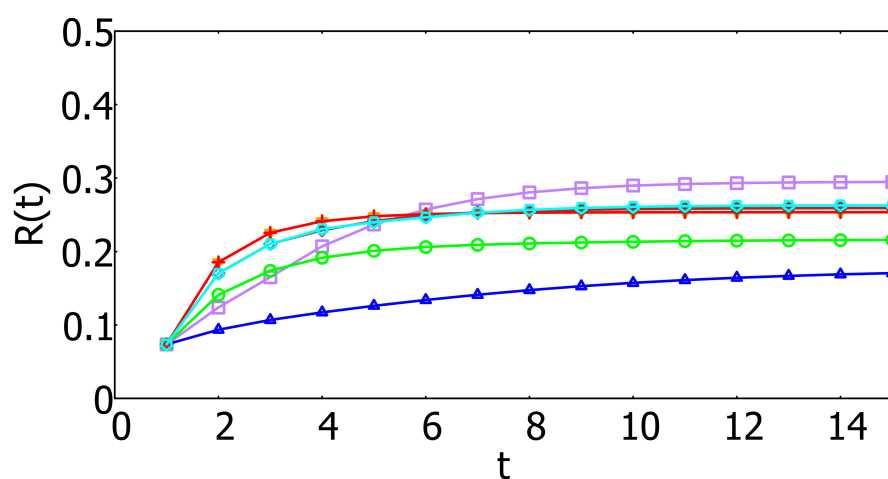
図 7.8: Gnutella の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。



(a) $l = 2$

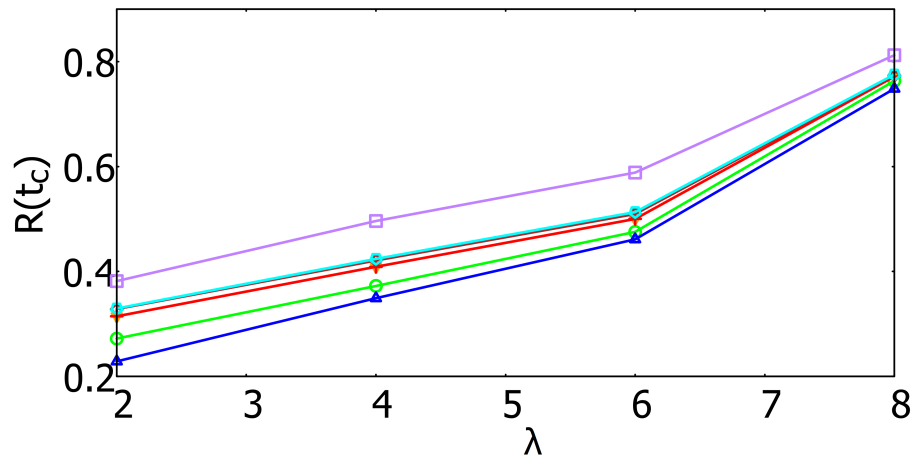


(b) $l = 3$

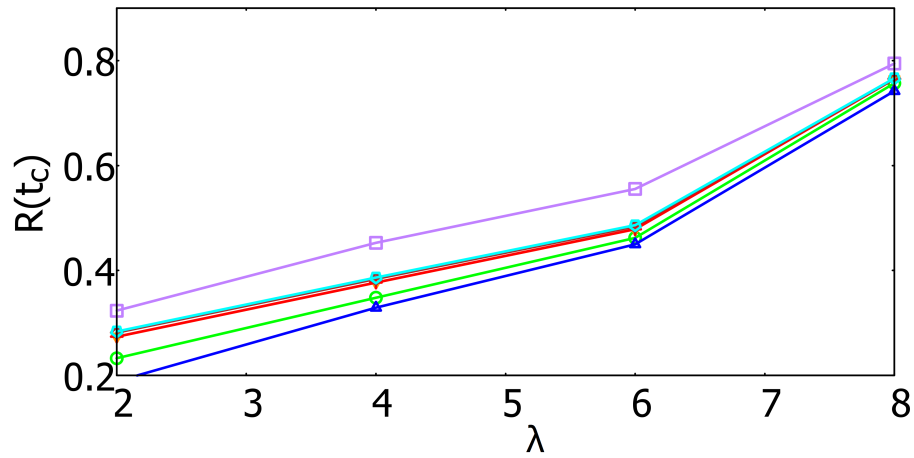


(c) $l = 4$

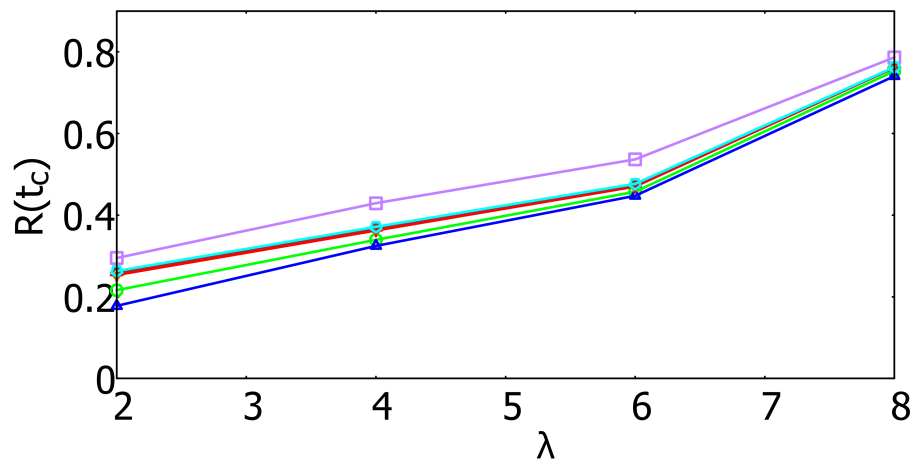
図 7.9: GRQC の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$

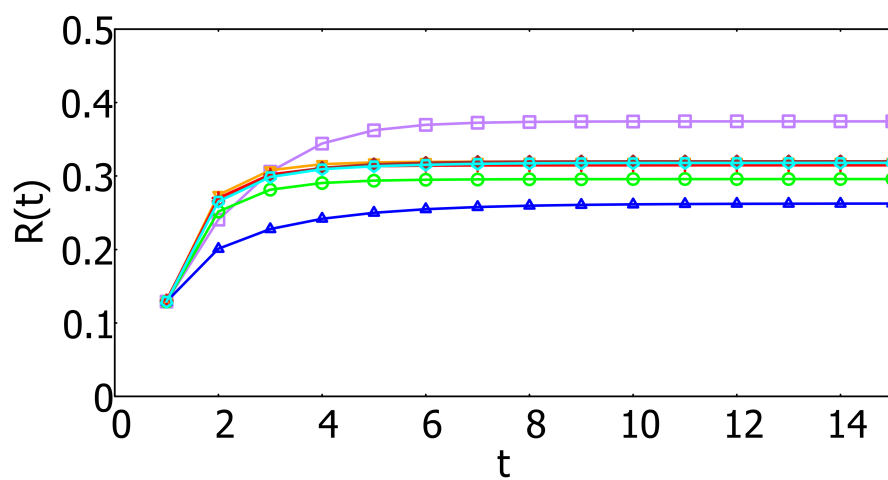


(b) $l = 3$

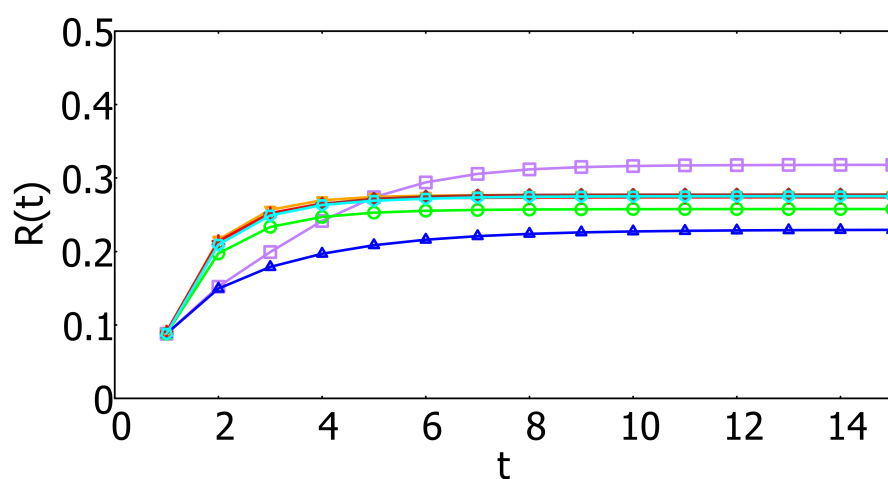


(c) $l = 4$

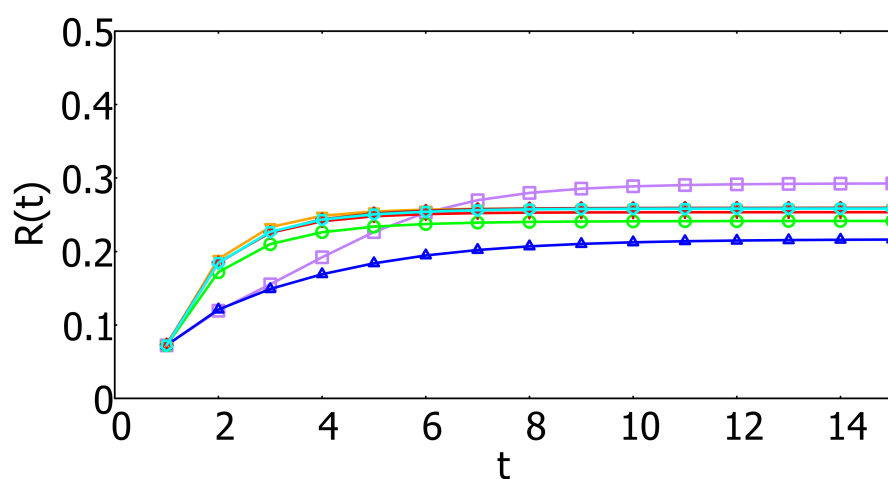
図 7.10: GRQC の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。



(a) $l = 2$

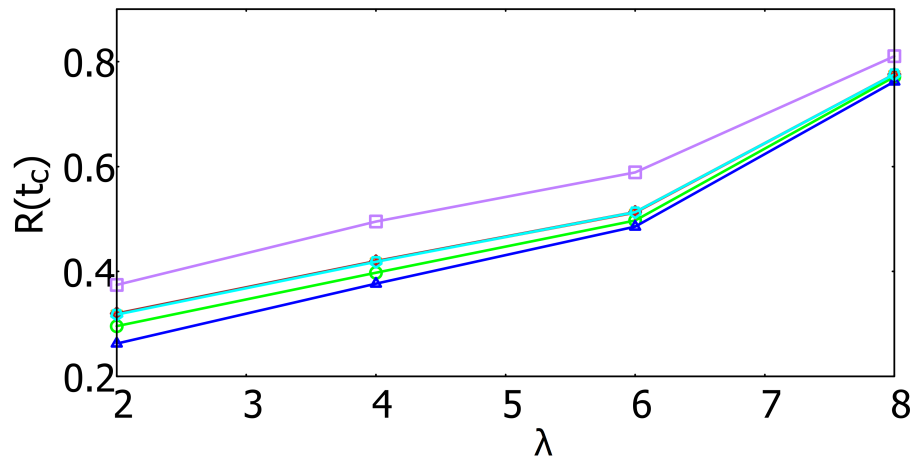


(b) $l = 3$

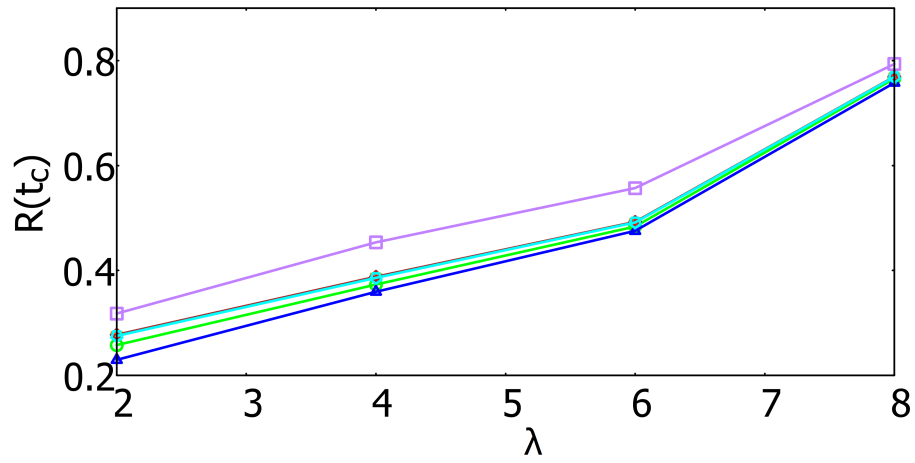


(c) $l = 4$

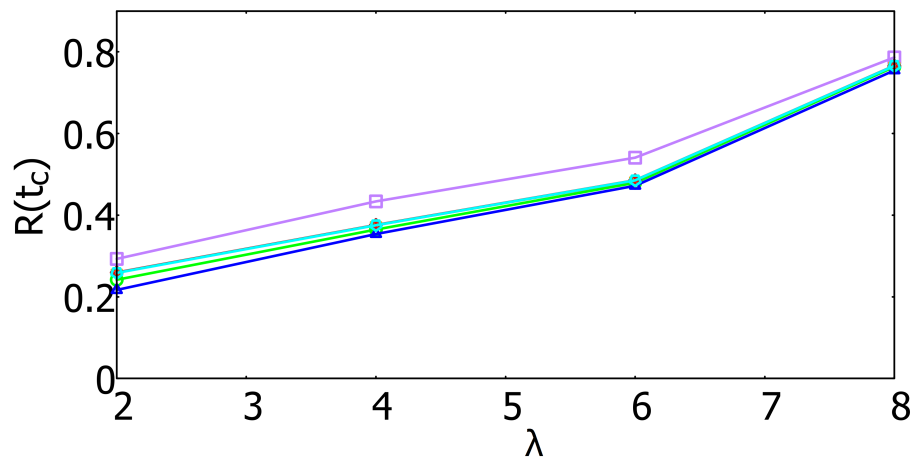
図 7.11: Hepth の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$

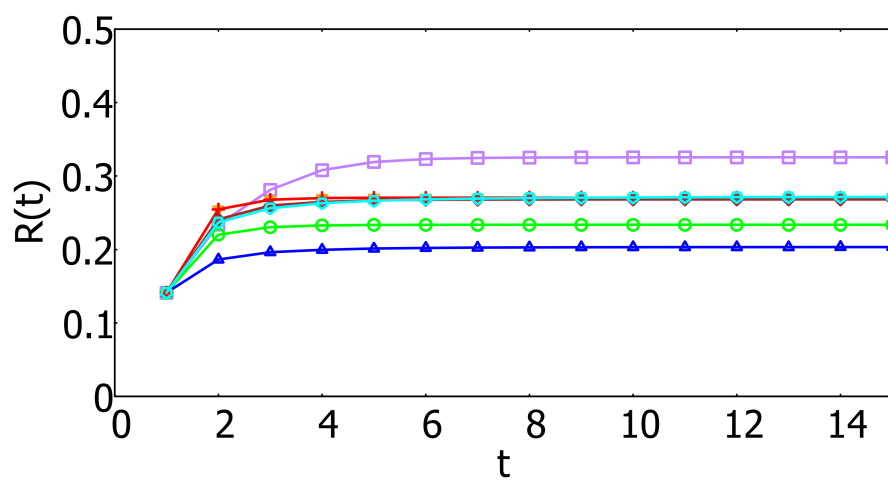


(b) $l = 3$

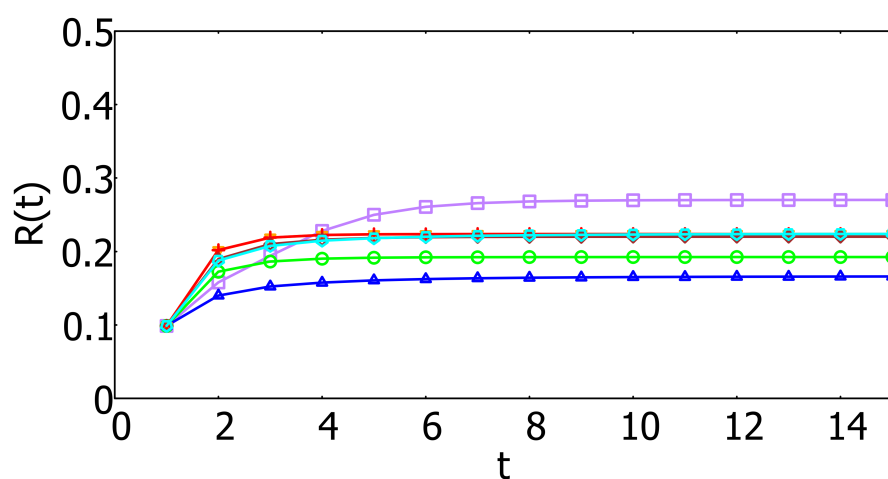


(c) $l = 4$

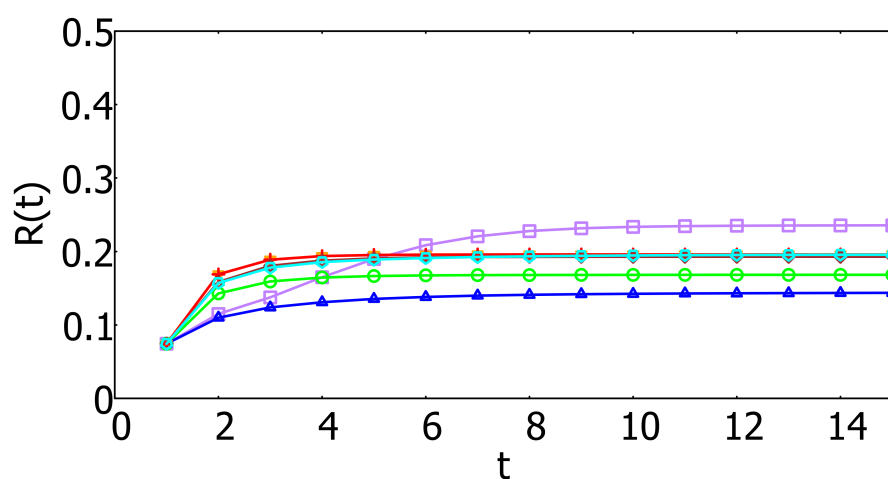
図 7.12: Hepth の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。



(a) $l = 2$

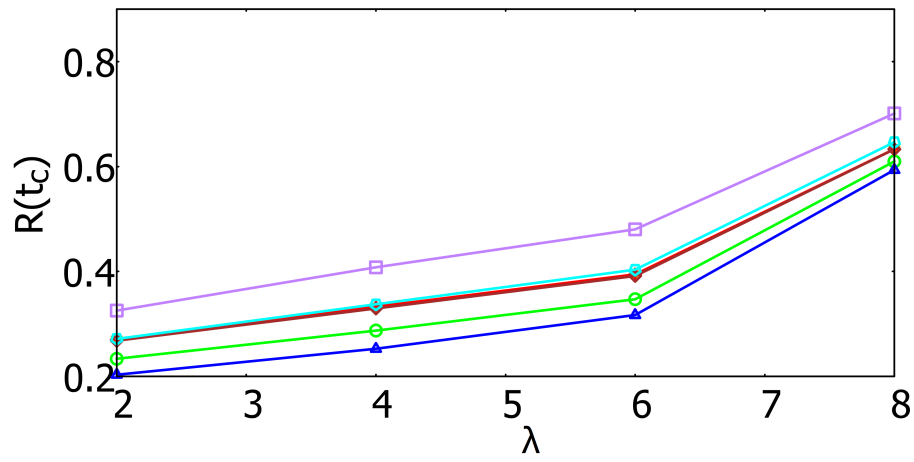


(b) $l = 3$

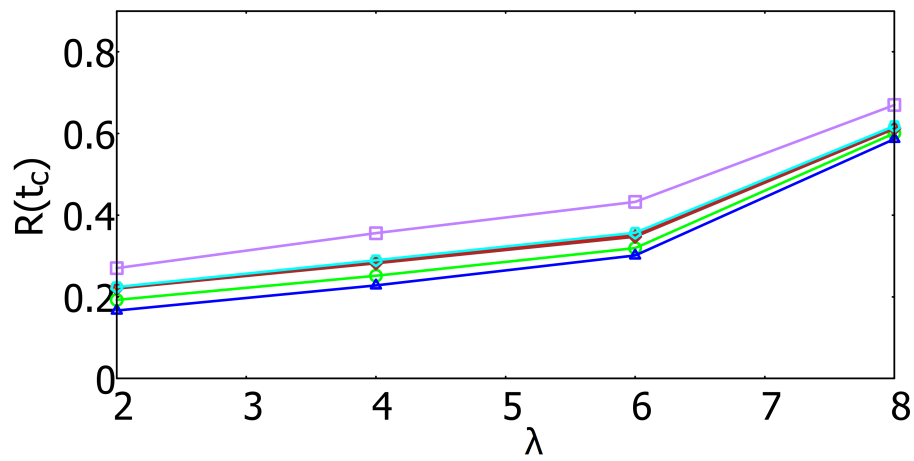


(c) $l = 4$

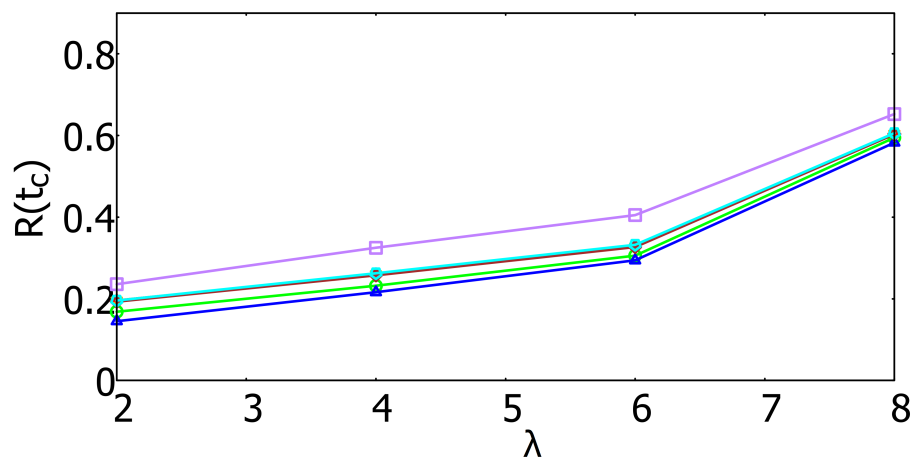
図 7.13: PGP の高速 SIR モデルにおける累積感染数 $R(t)$ の時間発展を示す図。
 $\lambda = 2$ で、 $l = 2, 3, 4$ の場合を比較している。



(a) $l = 2$



(b) $l = 3$



(c) $l = 4$

図 7.14: PGP の高速 SIR モデルにおいて、収束時間 t_c における累積感染数を、感染パラメータ λ を 2 から 10 まで変化させた場合の図。

発表論文・口頭発表

発表論文

[Liao1] Liao Fuxuan, Yukio Hayashi, Emergence of robust and efficient networks in a family of attachment models, *Physica A: Statistical Mechanics and its Applications* 599,(127427) 2022 .

[Liao2] Liao Fuxuan, Yukio Hayashi. Identify multiple seeds for influence maximization by statistical physics approach and multi-hop coverage. *Applied Network Science* 7, (52) 2022.

国際会議

Liao Fuxuan, Yukio Hayashi, Collective computation to find optimal multiple seeds for influence maximization in multi-hop coverage, *The 7th Combinatorics and Graph Theory Conference(Online)*, May. 27-29.2022.

Liao Fuxuan, and Yukio Hayashi, Strongly Robust Onion-like Networks generated in a Family of Attachment Models, *Networks 2021: A Joint Conference of Sunbelt and NetSci(Online)*, July. 5-10, 2021.

Liao Fuxuan, and Yukio Hayashi, A new relation of k-shell and feedback vertex set, *International School and Conference on Network Science*, Tokyo Japan, Jan. 11-13, 2020.

Liao Fuxuan, and Yukio Hayashi, Identifying a crucial role for robustness and spreading in complex network”, *The 8th International Conference on Complex*

Networks and their Applications, Book of abstracts pp.569-571, Lisbon Portugal, Dec. 21-23, 2019.

国内会議

Liao Fuxuan, 林 幸雄, “頑健性と拡散性に影響を与えるネットワーク指標の分析”, 電子情報通信学会 2020 年総合大会, N-2-3, 広島, 2020 年 3 月, 口頭発表.

[Liao3] Liao Fuxuan, 林 幸雄, “影響力最大化問題に対する種の感染力に影響する要因”, 電子情報通信学会 2023 年総合大会, N-2-1, 東京, 2023 年 3 月, 口頭発表.

参考文献

- [1] Hao Chen and Jürgen Jost. Minimum vertex covers and the spectrum of the normalized laplacian on trees. *Linear Algebra and Its Applications*, 437(4):1089–1101, 2012.
- [2] Lazaros K Gallos, Shlomo Havlin, H Eugene Stanley, and Nina H Fefferman. Proximity drives the emergence of network structure and density. *Proceedings of the National Academy of Sciences USA*, 116(41):20360–20365, 2019.
- [3] Yukio Hayashi. A new design principle of robust onion-like networks self-organized in growth. *Network Science*, 6(1):54–70, 2018.
- [4] Yukio Hayashi and Naoya Uchiyama. Onion-like networks are both robust and resilient. *Scientific Reports*, 8(1):1–13, 2018.
- [5] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [7] Paul Erdős and Alfréd Rényi. On random graphs i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.
- [8] Piotr Fronczak. *Scale-Free Nature of Social Networks*, pages 2300–2309. Springer New York, New York, NY, 2018.
- [9] Albert-László Barabási. Scale-free networks: a decade and beyond. *science*, 325(5939):412–413, 2009.
- [10] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell’s functional organization. *Nature reviews genetics*, 5(2):101–113, 2004.

- [11] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.
- [12] Robert V Kozinets, Kristine De Valck, Andrea C Wojnicki, and Sarah JS Wilner. Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of marketing*, 74(2):71–89, 2010.
- [13] Richard Manning Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Springer, 1972.
- [14] Martin Weigt and Haijun Zhou. Message passing for vertex covers. *Physical Review E*, 74(4):046110, 2006.
- [15] Alfredo Braunstein, Luca Dall’Asta, Guilhem Semerjian, and Lenka Zdeborová. Network dismantling. *Proceedings of the National Academy of Sciences*, 113(44):12368–12373, 2016.
- [16] Masaki Chuijo and Yukio Hayashi. Adding links on minimum degree and longest distance strategies for improving network robustness and efficiency. *Plos one*, 17(10):e0276733, 2022.
- [17] Zhiou Xu, Xiaobin Rui, Jing He, Zhixiao Wang, and Tarik Hadzibeganovic. Superspreaders and superblockers based community evolution tracking in dynamic social networks. *Knowledge-Based Systems*, 192:105377, 2020.
- [18] Yakir Berchenko, Yael Artzy-Randrup, Mina Teicher, and Lewi Stone. Emergence and size of the giant component in clustered random graphs with a given degree distribution. *Physical Review Letters*, 102:138701, Mar 2009.
- [19] Ming Li, Run-Ran Liu, Linyuan Lü, Mao-Bin Hu, Shuqi Xu, and Yi-Cheng Zhang. Percolation on complex networks: Theory and application. *Physics Reports*, 907:1–68, 2021. Percolation on complex networks: Theory and application.
- [20] Alex Bavelas. Communication Patterns in Task - Oriented Groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 06 2005.

- [21] G Sabidussi. The centrality index of a graph. *Psychometrika* 31, page 581–603, 1966.
- [22] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.
- [23] Paramita Dey, Subhayan Bhattacharya, and Sarbani Roy. A survey on the role of centrality as seed nodes for information propagation in large scale network. *ACM/IMS Transactions on Data ScienceArchive* Vol, 2(3), aug 2021.
- [24] Mark Newman. *Networks: An Introduction*. Oxford University Press, 03 2010.
- [25] Toshihiro Tanizawa, Shlomo Havlin, and H. Eugene Stanley. Robustness of onionlike correlated networks against targeted attacks. *Physical Review E*, 85(4):046109, Apr 2012.
- [26] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87(19):198701, 2001.
- [27] Vijay V Vazirani. *Approximation algorithms*, volume 1. Springer, 2001.
- [28] Hai-Jun Zhou. Spin glass approach to the feedback vertex set problem. *The European Physical Journal B*, 86(11):1–9, 2013.
- [29] Rodney G Downey, Michael R Fellows, and Ulrike Stege. Parameterized complexity: A framework for systematically confronting computational intractability. *Contemporary Trends in Discrete Mathematics*, 49:49–99, 1997.
- [30] Ning Chen. On the approximability of influence in social networks. *SIAM Journal on Discrete Mathematics*, 23(3):1400–1415, 2009.
- [31] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, Aug 2015.
- [32] Benedek Rozemberczki and Rik Sarkar. Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, page 1325–1334. ACM, 2020.

- [33] Reuven Bar-Yehuda and Shimon Even. A local-ratio theorem for approximating the weighted vertex cover problem. In *Analysis and Design of Algorithms for Combinatorial Problems*, volume 109 of *North-Holland Mathematics Studies*, pages 27–45. North-Holland, 1985.
- [34] Thomas Valente and Rebecca Davis. Accelerating the diffusion of innovations using opinion leaders. *The Annals of the American Academy of Political and Social Science*, 566(1):55–67, 1999.
- [35] Jianxiong Guo, Yapu Zhang, and Weili Wu. An overall evaluation on benefits of competitive influence diffusion. *IEEE Transactions on Big Data*, pages 1–1, 2021.
- [36] Flaviano Morone, Kevin Roth, Byungjoon Min, H. Eugene Stanley, and Hernán A. Makse. Model of brain activation predicts the neural collective influence map of the brain. *Proceedings of the National Academy of Sciences*, 114(15):3849–3854, 2017.
- [37] Linyuan Lü, Duan-Bing Chen, and Tao Zhou. The small world yields the most effective information spreading. *New Journal of Physics*, 13(12):123005, 2011.
- [38] Jianxiong Guo, Tiantian Chen, and Weili Wu. A multi-feature diffusion model: Rumor blocking in social networks. *IEEE/ACM Transactions on Networking*, 29(1):386–397, 2021.
- [39] Paulo Shakarian, Abhinav Bhatnagar, Ashkan Aleali, Elham Shaabani, and Ruocheng Guo. *The Independent Cascade and Linear Threshold Models*, pages 35–48. Springer International Publishing, Cham, 2015.
- [40] Xian Teng, Sen Pei, Flaviano Morone, and Hernán A Makse. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Scientific Reports*, 6(1):1–11, 2016.
- [41] 西浦博 and 稲葉寿. 感染症流行の予測: 感染症数理モデルにおける定量的課題. 2006.

- [42] 荒牧英治, 増川佐知子, 森田瑞樹, et al. Twitter catches the flu: 事実性判定を用いたインフルエンザ流行予測. *研究報告自然言語処理 (NL)*, 2011(1):1–8, 2011.
- [43] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Physical Review E*, 65:036104, Feb 2002.
- [44] Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. Locating privileged spreaders on an online social network. *Physical Review E*, 85(6):066123, 2012.
- [45] Gouhei Tanaka, Kai Morino, and Kazuyuki Aihara. Dynamical robustness in complex networks: the crucial role of low-degree nodes. *Scientific Reports*, 2(1):1–6, 2012.
- [46] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.
- [47] Duanbing Chen, Linyuan Lü, Ming-Sheng Shang, Yi-Cheng Zhang, and Tao Zhou. Identifying influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*, 391(4):1777–1787, 2012.
- [48] Shuai Gao, Jun Ma, Zhumin Chen, Guanghui Wang, and Changming Xing. Ranking the spreading ability of nodes in complex networks based on local structure. *Physica A: Statistical Mechanics and its Applications*, 403:130–147, 2014.
- [49] Nam Parshad Bhatia and Giorgio P Szegő. *Stability theory of dynamical systems*. Springer Science & Business Media, 2002.
- [50] Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4):521–529, 2002.
- [51] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1):173–187, 1999.

- [52] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Physical Review Letters*, 85:4626–4628, Nov 2000.
- [53] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Breakdown of the internet under intentional attack. *Physical Review Letters*, 86:3682–3685, Apr 2001.
- [54] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468, 2000.
- [55] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific American*, 288(5):60–69, 2003.
- [56] Reuven Cohen and Shlomo Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5):058701, 2003.
- [57] Giulia De Masi, Giulia Iori, and Guido Caldarelli. Fitness model for the italian interbank money market. *Physical Review E*, 74(6):066112, 2006.
- [58] Kimmo Soramäki, Morten L Bech, Jeffrey Arnold, Robert J Glass, and Walter E Beyeler. The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1):317–333, 2007.
- [59] Mark Steyvers and Joshua B Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1):41–78, 2005.
- [60] Derek J. de Solla Price. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.
- [61] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.
- [62] Zhi-Gang Shao, Xian-Wu Zou, Zhi-Jie Tan, and Zhun-Zhi Jin. Growing networks with mixed attachment mechanisms. *Journal of Physics A: Mathematical and General*, 39(9):2035–2042, feb 2006.

- [63] Lei Min and Zhao Qinggui. Degree distribution of a mixed attachments model for evolving networks. In Yanwen Wu, editor, *Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009)*, pages 815–822, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [64] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65(2):026107, 2002.
- [65] Paul L Krapivsky, Sidney Redner, and Francois Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85(21):4629, 2000.
- [66] Paul L Krapivsky and Sidney Redner. Organization of growing random networks. *Physical Review E*, 63(6):066123, 2001.
- [67] Paul L Krapivsky and Sidney Redner. A statistical physics perspective on web growth. *Computer Networks*, 39(3):261–276, 2002.
- [68] Romualdo Pastor-Satorras, Eric Smith, and Ricard V Solé. Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, 222(2):199–210, 2003.
- [69] P. L. Krapivsky and S. Redner. Network growth by copying. *Physical Review E*, 71:036118, Mar 2005.
- [70] Cynthia S. Q. Siew and Michael S. Vitevitch. Investigating the influence of inverse preferential attachment on network development. *Entropy*, 22(9), 2020.
- [71] Cynthia S. Q. Siew and Michael S. Vitevitch. An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General*, 149(12), 2020.
- [72] Christian M. Schneider, André A. Moreira, José S. Andrade, Shlomo Havlin, and Hans J. Herrmann. Mitigation of malicious attacks on networks. *Proceedings of the National Academy of Sciences*, 108(10):3838–3841, 2011.

- [73] Zhi-Xi Wu and Petter Holme. Onion structure and network robustness. *Physical Review E*, 84:026106, Aug 2011.
- [74] Yukio Hayashi. Growing self-organized design of efficient and robust complex networks. In *2014 IEEE Eighth International Conference on Self-Adaptive and Self-Organizing Systems*. IEEE,Xplore, 2014.
- [75] VN Zadorozhnyi and EB Yudin. Growing network: models following non-linear preferential attachment rule. *Physica A: Statistical Mechanics and its Applications*, 428:111–132, 2015.
- [76] Mark EJ Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.
- [77] Mark Newman. 8.7 Assortative mixing. In *Networks*, pages 266–268. Oxford University Press, 2010.
- [78] Xu-Hua Yang, Shun-Li Lou, Guang Chen, Sheng-Yong Chen, and Wei Huang. Scale-free networks via attaching to random neighbors. *Physica A: Statistical Mechanics and its Applications*, 392(17):3531–3536, 2013.