

Comparison of the innovation process on several networks

メタデータ	言語: jpn 出版者: 公開日: 2021-04-05 キーワード (Ja): キーワード (En): 作成者: 片山, 明, 木下, 修一 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/1498

いくつかのネットワーク上でのイノベーションプロセスの比較

Comparison of the innovation process on several networks

片山 明¹

Akira Katayama

木下 修一²

Shu-ichi Kinoshita

1 はじめに

我々の社会では、日々新しい概念が提案されそれが実現化されており、このような実現化された新しい概念は、しばしば“イノベーション”と呼ばれる。革新的なイノベーションは新しいイノベーションを誘発し、イノベーションが連鎖的に発生する。このイノベーションが連鎖的に発生するプロセスこそが、我々の社会に文化的多様性や経済的な繁栄をもたらしており、イノベーションプロセスを理解することは、我々の社会の理解を大きく進めることに繋がる。これまで、一般的にイノベーションの痕跡が残ることは少なく、このようなイノベーションプロセスを理解することは困難であった。しかし、現在、イノベーションの痕跡は、インターネット・デジタル技術の発展により主にデジタルデータとして蓄積されており、そのデータの種類は科学論文 [1]、映画 [2]、言語 [3]、商品 [4]、知識や情報 [5] など、多岐にわたる。このようなデータを解析することで、イノベーションプロセスに関するいくつかの統計的特徴が明らかになっている。代表的な特徴の1つは、時系列データの中に発生する累積の新規データ数 $D(t)$ には相関が存在し、Heaps' 則と呼ばれるべき的な振る舞い ($D(t) \sim t^\beta$ ($0.0 < \beta < 1.0$)) を共通して示すことであり、もう一つは異なるイベントの出現頻度の順位 R の順位分布 $f(R)$ は Zipf's 則 ($f(R) \sim R^{-\alpha}$) に従うことである [6, 7]。そこで、このような統計則を再現するイノベーションプロセスに関する数理モデルの提案がいくつか行われており、代表的なモデルとしてポリヤの壺に基づくモデルがあげられる [8]。一方、このようなイノベーションプロセスにおいて“adjacent possible”と呼ばれる重要な概念が S. Kauffman により提案された [9]。この概念は、彼が生物および分子進化過程を研究する中で提案した概念である。進化過程において生物や分子に突然変異が起こるとき、必ずしも莫大な数の状態の中からランダムに選ばれた状態が実現されるのではなく、あらかじめ制限された可能なより小さい状態からランダムに選ばれ実現されるという概念である。例えば、高分子の進化を考える場合、元の高分子 A がより大きな高分子に進化する場合は結合可能な分子は化学的な組成や性質により制限される。この場合、元の高分子 A に結合可能な分子集団が adjacent possible に相当する。より

¹ 武蔵野大学大学院工学研究科数理工学専攻 / 修士2年

² 武蔵野大学数理工学センター員 / 武蔵野大学工学部数理工学学科准教授

簡潔に述べるならば, adjacent possible とは元々実現されているモノ・コトから行くことが可能な一歩先にあるモノ・コトの集合である. Tria らはポリヤの壺モデルを拡張し adjacent possible の概念を取り入れることで, Heaps' 則, Zipf's 則, その他の統計則の再現に成功するだけでなく, さらにそのモデルが 3 つのパラメータを持つ成長するネットワーク上におけるランダムウォークモデルと対応することを示した [10]. この結果を踏まえ, Iacopini らは 1 つのパラメータのみを持ち, 重みが増加するネットワーク上におけるランダムウォークモデルである edge-reinforced random walk モデル (ERRW) を提案し, Tria らと同様に Heaps' 則, Zipf's 則, その他の統計則の再現に成功した [11]. ERRW は 1 つのパラメータのみで統計則を再現できるだけでなく, adjacent possible をネットワーク構造として初めから表現しており, adjacent possible とイノベーションプロセスの関係を調べる上で非常に有用なモデルである. Iacopini らの結果はイノベーションプロセス過程を理解する上で非常に興味深い結果であるが, 細かい点でいくつか気になる点がある. 1 つ目としては, スモールワールド・ネットワーク (SWN) 作成に関して確率 p でエッジを付加する手法を用いている点である. この手法はスモールワールド性を表す指標 (クラスター係数, 平均経路長) を変化させながら, ネットワークの強連結性を維持できるという面で便利ではあるが, p に応じて次数が大きくなるという欠点がある. もう, 1 点は Heaps' 則以外の統計則に関して, SWN 以外のネットワーク上での結果を示していないことである. この点を踏まえ, 本研究では Iacopini らの結果をより明確化するため, 前者に関しては次数一定かつ強連結性を維持するアルゴリズムを用い SWN の作成を行った. また, 後者に対応するため Heaps' 則以外の統計則を円環ネットワーク (CN), ランダム・ネットワーク (RN) 上でも数値計算により調べ, SWN 上の統計則との比較を行った.

2 edge-reinforced random walk (ERRW)

本章では edge-reinforced random walk モデル (ERRW) について述べる. ERRW では, ウォーカーがネットワーク上をランダムウォークする過程が, イノベーションを生み出すまでの探索過程に対応し, 新しいノードに到達することが, イノベーションの発生に対応する. まず, ノード数 N , 平均次数 $\langle k \rangle$ さらに各エッジに重み $w_{i,j}^t$ が割り当てられた重み付きネットワークを考える. ここで, i, j はノード番号を表している ($i, j = 1, 2, \dots, N$). ウォーカーはこのネットワーク上を $w_{i,j}^t$ に応じてランダムウォークし, ウォーカーの動きによって $w_{i,j}^t$ は時間変化する. 初期条件時刻 $t = 0$ において, 全てのエッジの重みは同じ値をとり $w_{i,j}^0 = 1$ とする. 時刻 t におけるノード i からノード j へのウォーカーの遷移確率は次式で表される

$$Prob^t(i \rightarrow j) = \pi_{j,i}^t = \frac{w_{i,j}^t}{\sum_l w_{i,l}^t}. \quad (1)$$

遷移確率は正規化されており, $\sum_j \pi_{j,i}^t = 1$ を満たす. また, ネットワークは自己ループを持たないため, ウォーカーは各時間ステップで必ずノードを移動する. ウォーカーが時刻 t で移

動する時、選択されたエッジ (i, j) の重みは次式のように強化される

$$w_{i,j}^{t+1} = w_{i,j}^t + \delta w. \quad (2)$$

ここでの δw は強化値と呼ばれ、このモデルにおいて唯一のコントロールパラメータであり、このパラメータを調節することで、ウォーカーが到達したノードの時系列パターンに相関が生ずる。本研究では、次のようにネットワークを作成する。まず、円環上にノードを並べ全てのノードについて両隣のノード（最近接ノード）および最近接のさらに隣のノード（第二隣接ノード）にエッジを張る。この操作によって次数 $k = 4$ の CN が作成される。続いて、それらのすべてのエッジから確率 p でエッジを選び、エッジの端のノードを全てのノードからランダムに選びなおしエッジのつなぎ変えを行う。この時、 p の値によって、ネットワークのクラスター係数および平均経路長をコントロールし様々なネットワークを作成できる [12]。ただし、この手法ではネットワークの強連結性（強連結性とはネットワーク上の任意の 2 ノード間に経路が存在する事である）が失われる可能性があるため、作成したネットワークの強連結性を調べ強連結性を維持しているネットワークサンプルのみを用いた。また、本研究では新規ノードの統計則を表す時系列の長さとして $t = 10^4$ 程度を想定しているため、それと比較し十分に大きいネットワークを用いる必要がある。本研究では、時系列の 10 倍程度の大きさ $N = 10^5$ 、平均次数は $\langle k \rangle = 4$ とし、CN ($p = 0.0$)、SWN ($p = 0.02$)、RN ($p = 1.00$) の 3 種類のネットワーク上で EPRW の数値計算を実行し、その結果の比較を行った。

3 数値計算結果

本モデルでは、ウォーカーが到達したノードの時系列パターンの統計的特徴に着目する。図 1 はウォーカーが、ある時刻 t までに到達した新しいノードの累積回数 $D(t)$ の時間発展の両対数プロットである。(a)CN, (b)SWN, (c)RN の全ての場合で、累積回数は $D(t) \propto t^\beta$ という関係 (Heaps' 則) が成り立っている。この結果は、Heaps' 則がネットワーク構造に依らない頑健な法則であることを示唆している。ただし、ネットワーク構造の違いや δw の違いによりベキ指数 β の値は異なっている。そこで、各ネットワーク毎に δw に対するベキ指数 β の依存性を調べた結果が図 2 である。CN, SWN, RN のそれぞれのネットワーク上の ERRW から得ることが可能な β の範囲はそれぞれ $\beta < 0.50$, $\beta < 0.87$, $\beta < 0.98$ となる。先行研究 [10] から現実のデータ (Gutenberg, Last.fm lyrics, Last.fm.artist, Wikipedia, del.icio.us) のベキ指数はそれぞれ $\beta = 0.45, 0.68, 0.56, 0.77, 0.78$ となっている。RN および SW では、 δw をコントロールすることで、この全ての β の値を再現することが可能であるのに対し、CN では $\beta > 0.5$ となる β を再現できないことが分かる。

続いて、ウォーカーが到達したノードの時系列の相関関係に焦点を当てる。ネットワーク上の ERRW から得られる時系列は、ノードの出現頻度自体が初めて出現する時間に依存する非定常な時系列となる。このような非定常な時系列において相関性を論ずる決定的な統計量は存在せず、データの特徴にあった統計量を考える必要がある。本研究では、あるノードから同じ

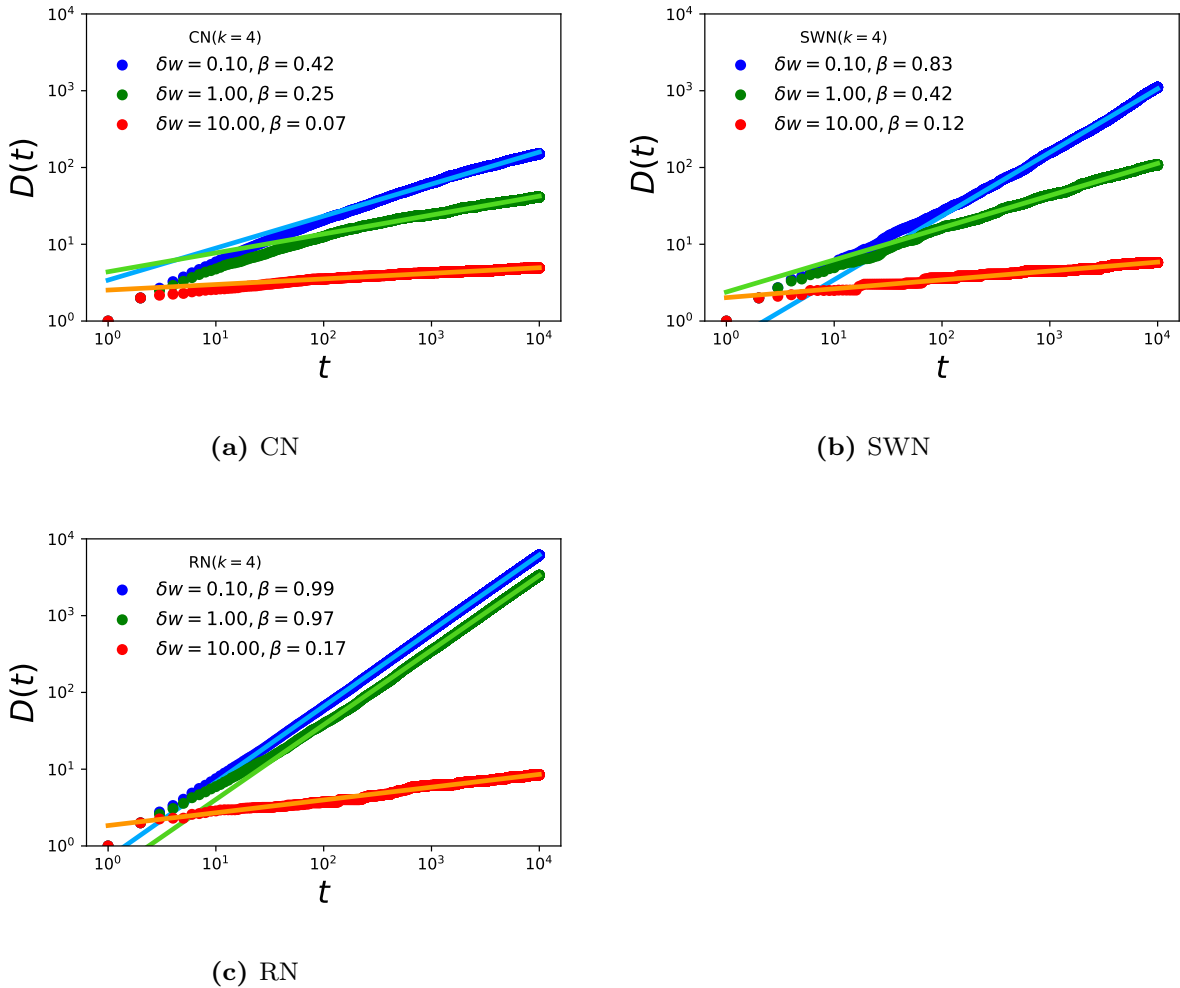


図 1: ウォーカーが新しいノードに到達した累積回数 $D(t)$ の時系列 (両対数). (a)CN, (b)SWN, (c)RG の $D(t)$ 表し, 各点の色の違いはエッジの強化値の違いを表している ($\delta w = 0.10$ (青), 1.00 (緑), 10.0 (赤)). 実線はフィッティング関数 $f(t)$ を表し, $f(t) \propto t^\beta$ である.

ノードへ次に到達するまでの時間間隔分布と出現頻度に対する時系列のエントロピーの 2 つの統計量を用い, 相関のない時系列から得られる結果と比較することで時系列の相関について論ずる. 比較のための相関のない時系列は元のデータから Fisher-Yates shuffle 法を用い作成した [13].

図 3 はウォーカーがあるノードから次の新しいノードへ到達するまでの時間間隔分布 $f(\delta t)$ である. (a)CN, (b)SWN, (c)RG の時系列から得られた $f(\delta t)$ である. δt の小さい領域では, 3 つ全てのネットワークにおいて $f(\delta t)$ に相関があり, δt のすべての領域で, (a) と (b) では $f(\delta t)$ (青) がベキ的にふるまう. このベキ的なふるまいは現実のデータで現れた特徴と一致する [10]. この結果は (a), (b) では一部のノードをウォーカーが繰り返し訪問していることを示している. 一方, (c) においては $f(\delta t)$ (青) はベキ的なふるまいを示さないことが分かった.

続いて, 時系列のエントロピーについて比較する. 時系列内に出現するノード A のエント

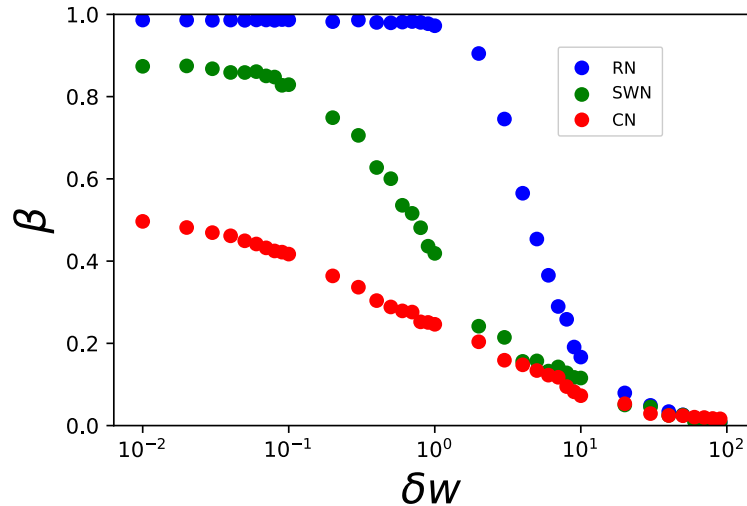


図 2: Heaps law のベキ指数 β の δw 依存性. ネットワーク毎に色を変えて表現している (CN(赤), SWN(緑), RN(青)). 各点はサンプル数 100 個の平均である.

ロピーは先行研究 [10] に従い以下のように定義される

$$H_A(n) = - \sum_{i=1}^n \frac{f_i}{n} \log \frac{f_i}{n}. \quad (3)$$

ここで注意すべき点は、用いられる時系列はノード A が初めて出現して以降の時系列であるという事である. n は時系列内に出現したノード A の数, f_i は時系列を n 等分した際の, その区間内にあるノード A の数である. n 回現れたノードのエントロピーの平均を $H(n)$ とし, それを $\log n$ で規格化したものが図 4 である. (a), (b), (c) の全ての場合において, ランダムシャッフルした時系列データよりエントロピーが小さくなっているのが分かる. また, $n < 10$ の範囲では $H(n)/\log(n) < 0.1$ となり, 時系列内の一部にノードが偏って出現していることが分かる. 一方, $10 < n < 100$ の範囲では (a), (b) と (c) では傾向が大きく異なる. (a), (b) ではエントロピーは単調に増加し, ランダムシャッフルされた時系列のエントロピーに近づく. 一方, (c) では $10 < n < 100$ の範囲内に要素の最大出現回数 n_{max} が存在する. これは, CN, SWN と異なり RN のクラスター係数も小さいという構造上の特徴が影響していると考えられる. つまり, クラスター (ネットワークに存在する三角形のループ構造) が少ない場合, 例えエッジが強化されても特定のノード間を確率的ではあるが再帰的に巡るモードが現れ難くなっていることがこの結果から示唆される. 先行研究において, 現実のデータから同様のエントロピーが計算されており定性的に (a), (b) と同じ振る舞いを示す [10].

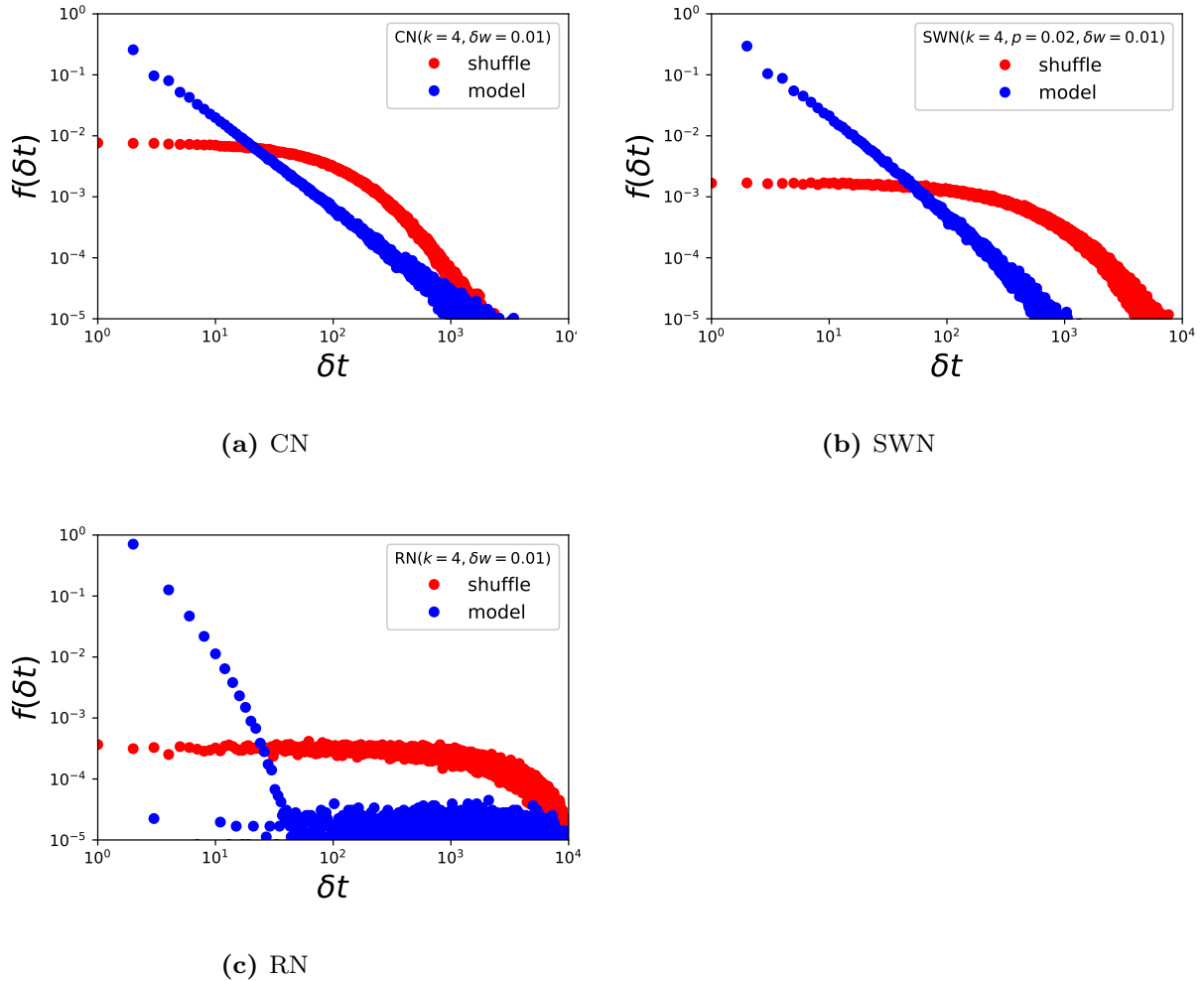


図 3: 同じノードに到達した時間 δt の時間間隔分布 $f(\delta t)$ (両対数). (a)CN, (b)SWN, (c)RN に対応しており, モデルから得られた $f(\delta t)$ は青, 比較のためランダムシャッフルした時系列より得られた $f(\delta t)$ は赤で表している. 各点はサンプル数 100 個の平均である.

4 まとめと議論

本研究では, 現実のイノベーションプロセスのデータに現れる統計則とネットワーク構造の関係を明らかにするための第一歩として, 平均次数 $\langle k \rangle$ を一定としたときの, CN, SWN, RN 上における ERRW の振る舞いを数値計算を用いて調べた. 全てのネットワーク上でウォーカーが新しいノードに到達する累積累積回数は Heaps' 則 $D(t) \sim t^\beta$ となり (図 1), 現実のデータが示す β の値 ($\beta > 0.5$) を実現できるネットワークは SWN および RN であることが分かった (図 2). 続いて, ウォーカーが到達するノードの時系列の相関を時間間隔分布 $f(\delta t)$ およびエントロピー $H(n)$ を用い調べた. CN と SWN 上の ERRW から得られた時間間隔分布 $f(\delta t)$ はべき的に減衰することが分かった (図 3). これは, 現実のデータでも現れる特徴である. 最後に, 出現頻度のエントロピー $H(n)$ を計算し, CN と SWN 上の ERRW から得ら

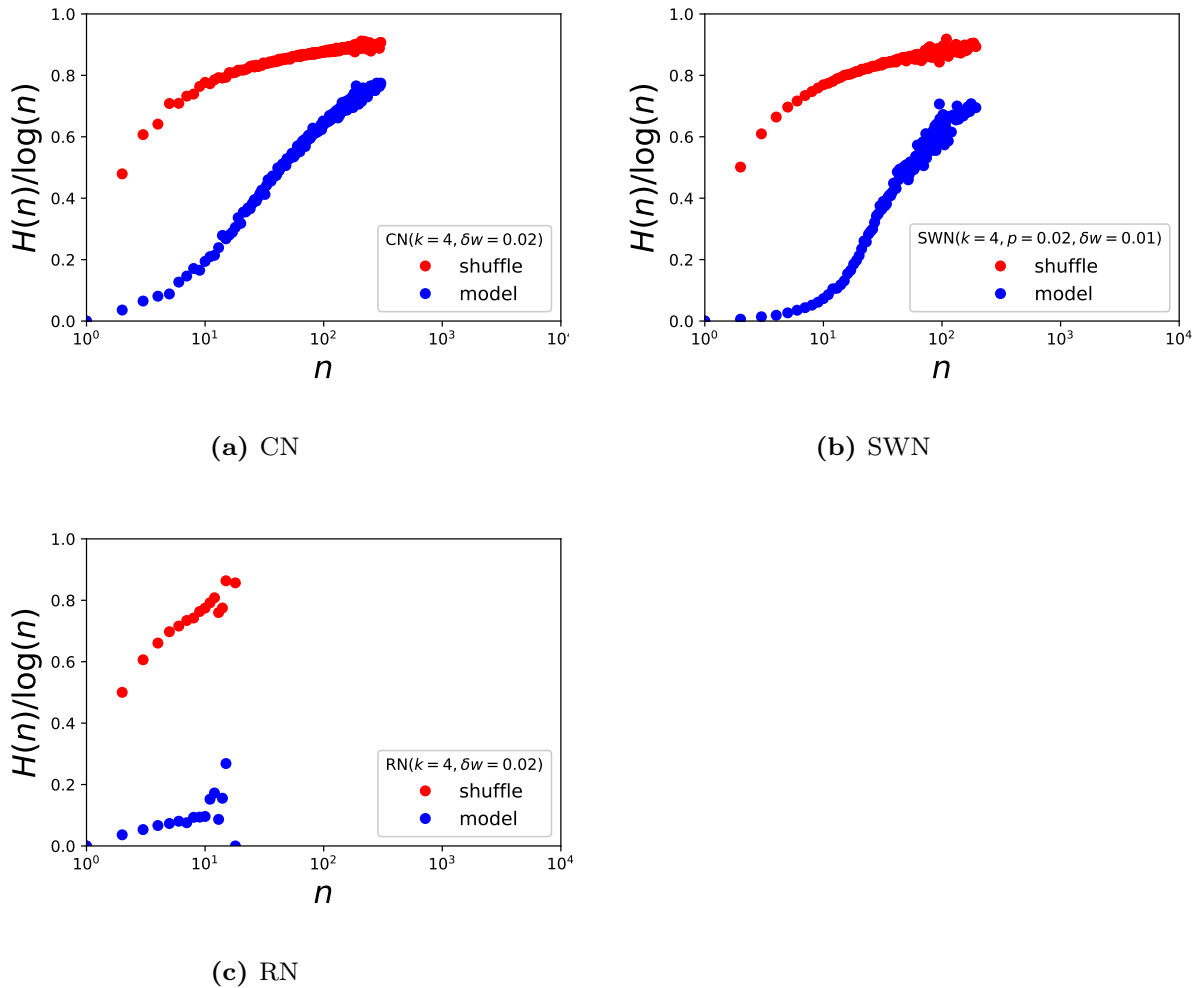


図 4: 出現頻度 n に対する規格化されたエントロピー $H(n)$ (x 軸は対数表示). (a)CN, (b)SWN, (c)RN に対応しており, モデルから得られた $H(n)/\log(n)$ は青, 比較のためランダムシャッフルした時系列より得られた $H(n)/\log(n)$ は赤で表している. 各点はサンプル数 100 個の平均である.

れた $H(n)$ は現実のデータが示す関数形と類似しており, 単調増加することが分かった. 以上より, 3つの統計量を比較した結果から SWN 上の ERRW のみ, 現実のイノベーションプロセスが示す全ての統計的性質をみとることが明らかになった. これは, イノベーションプロセスにおける adjacent possible を表現するネットワーク構造としてはスモールワールド性が重要であることを示唆している. 本研究の内容は片山による 2020 年度武蔵野大学大学院修士論文に基づき一部の結果を示したものである [14].

参考文献

- [1] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabasi, *Science* **354** aaf5239 (2016).

- [2] S. Sreenivasan, *Sci.Rep.* **3**, 2758 (2013).
- [3] A. Puglisi, A. Baronchelli, and V. Loreto, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7936 (2008).
- [4] F. Saracco, R. Di Clemente, A. Gabrielli, and L. Pietronero, *PLoS One* **10**, e0140420 (2015).
- [5] M. Andjelković, B. Tadić, M.M. Dankulov, M. Rajković, and R. Melnik, *PLoS One* **11**, e0154655 (2016).
- [6] H. S. Heaps, *Information Retrieval: Computational and Theoretical Aspects* (Academic, New York, 1978).
- [7] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Ravenio, Cambridge, England, 2016).
- [8] G. Pólya, *Annales de l' I.H.P.* **1**, **117** (1930).
- [9] S. A. Kauffman, *Investigations: The Nature of Autonomous Agents and the Worlds They Mutually Create. SFI working papers* (Santa Fe Institute, 1996).
- [10] F. Tria, V. Loreto, V. D. P. Servedio, and S. H. Strogatz, *Sci. Rep.* **4**, 5890 (2014).
- [11] I. Iacopini, S. Milojevic, and V. Latora, *Phys. Rev. Lett.* **120**, 048301 (2018).
- [12] M. Newman, *Networks*, Oxford Univ. press (2018).
- [13] R. A. Fisher, *Statistical tables for biological, agricultural and medical research*, (Oliver & Boyd, London 1957).
- [14] 片山 明, 2020 年度武蔵野大学大学院修士論文 (2021).

(原稿提出: 2021 年 1 月 16 日; 修正稿提出: 2021 年 2 月 7 日)