Vol.2015-AL-153 No.2
2015/6/12

階層的スケールフリーネットワークにおける万能定数時間アルゴリズム

伊藤 大雄[1,2,a)]

概要：ビッグデータへの劣線形時間アルゴリズムによるアプローチが最近有望視されている。代表的なビッグデータであるスケールフリーネットワークには階層的構造があることが指摘されており、我々はそれを利用した自然な多重グラフのクラス HSF (Hierarchical Scale-Free multigraphs) を定義する。そして、このクラスにおいて、超有限性 (hyperfiniteness) が成立することを証明する。この結果から「本クラスの多重グラフにおいては、すべての性質が定数時間検査可能である」という驚くべき結論を導くことができる。これは疎でかつ次数上限の無いグラフモデルに対する初めての万能検査アルゴリズムであり、現実の複雑ネットワーク上の超高速アルゴリズムの構築を可能とする基礎理論である。

1. はじめに

ビッグデータは理論計算機科学の分野においても重要な問題となってきている。各国で様々なプロジェクトが起こっているが、我が国でも JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」が始まり、その課題の一つとして「ビッグデータ時代に向けた革新的アルゴリズム基盤」Foundations of Innovative Algorithms for Big Data (ABD14) が 2014 年より発足している [*1]。

そこでの中心概念は「劣線形 (sublinear)」であり、ビッグデータに対して、データのごく一部（データ量を N とすると $o(N)$ 個のデータ）しか読まない劣線形時間アルゴリズムが重要な道具として期待されている。

劣線形時間アルゴリズムの中でも「性質検査」(property testing) はもっとも研究されている分野である。任意の固定された「性質」に対する検査アルゴリズム（検査機）とは、与えられた入力のうちの定数個だけを見ることで、その入力がその性質を満たせば高い確率（例えば 2/3 以上）で受理し、その性質からほど遠ければ高い確率で拒否する乱択アルゴリズムである。検査機の存在する性質のことを「検査可能」であるという [9]。

グラフに対する性質検査は特に良く研究されている [2], [3], [9], [10], [11], [12], [18], [20]。グラフにおける性質検査には大きく分けて以下の三つのモデルがある。密グラフ（隣接行列）モデル、次数制限モデル、一般モデルである。このうち密グラフモデルが最も良く解明されており、検査可能な性質の特徴づけ（必要十分条件）が明らかにされている [2]。しかし現実のものをグラフ化すると、ほとんどの場合、グラフは疎になることが知られているので、密グラフモデルの適用は難しい。次数制限グラフモデルに関しては、マイナーに関して閉じている性質はすべて検査可能であることが分かっている [3]。さらにこの結果は、超有限（定義は後述）なグラフのクラスに関しては任意の性質が検査可能である [18] という驚くべき結果に結びついている。超有限は次数制限グラフに関しては、マイナーについて閉じている性質を含む広いクラスである。しかし、現実のネットワークをモデル化した場合、次数制限は相性が悪い。実際、ウェブグラフなどのスケールフリーネットワークには次数が大変大きいハブと呼ばれる頂点が存在することが知られており [1], [16]、このようなグラフに対しては、これらのアルゴリズムは適用できない。

グラフにおける典型的なビッグデータは「スケールフリーネットワーク」と呼ばれ、様々なモデルが研究、提案されている [1], [4], [5], [6], [8], [16], [17], [19], [21], [22], [23]。最近、スケールフリーネットワークにおける孤立クリークの分布における階層的構造に着目した有力なモデルが繁住・宇野・渡辺 [21] によって提案された。孤立クリークとは伊藤・岩間 [14], [15] によって提案されたもので、非負定数 $c \geq 0$ に対し c **孤立クリーク**とは、k 頂点クリークで、そのクリークとクリーク外を結ぶ辺の数が ck 未満であるものである。1 孤立クリークのことを単に**孤立クリーク**と

1 電気通信大学
The University of Electro-Communications (UEC)
2 科学技術振興機構, CREST
CREST, JST
a) itohiro@uec.ac.jp

*1 http://www.alg.cei.uec.ac.jp/itohiro/e-science/crest2014.html

呼ぶこともあり、本稿では「孤立クリーク」はこの意味で用いる。

このモデル [21] に基づいて我々は階層的スケールフリー多重グラフ (Hierarchical Scale-Free Multigraphs; HSF; 定義 3.2) を導入する。これはスケールフリーネットワークの自然なモデル化であるが、これに対して、我々は以下の結果 (詳細は定理 3.4) を得た。

指数法則の指数が 2 より大きい HFS においては、
任意の性質が検査可能である。

本結果から、ウェブグラフなど多くのグラフ形状のビッグデータ上の問題が理論上は定数時間検査可能であることになる。このアルゴリズムは Newman と Sohler の STOC ' 11 のアルゴリズム [18] を利用しているが、彼らの結果は次数制限モデルに対するものであり、HSF は次数制限を持たない。本結果は一般グラフモデルに対する初めての万能検査アルゴリズムである。

2. 準備

本稿で扱うグラフは自己ループを持たない無向多重グラフである。本稿ではそれらを単に「グラフ」といい、$G = (V, E)$ で表す。V は頂点集合、E は辺集合である。グラフ $G = (V, E)$ とその頂点集合 $X, Y \subseteq V$ に対し、X と Y の間の辺集合を $E_G(X, Y) = \{(x, y) \in E \mid x \in X, y \in Y\}$ で表す。$E_G(X)$ は $E_G(X, V - X)$ のことである。$|E_G(X)|$ を $d_G(X)$ で表す。頂点 $v \in V$ に対し、v に接続する辺数をその頂点の**次数 (degree)** と呼ぶ。ただ一つの要素からなる集合、例えば $\{x\}$ は誤解の恐れない場合には、集合を表す中括弧を省略して x のように表現することもある。従って v の次数は $d_G(v)$ と表現できる。これらの添字の G はそれらが明らかな場合は $E(*)$, $d(*)$ のように省略できる。

グラフ $G = (V, E)$ と頂点集合 $X \subseteq V$ に対し、X **による誘導部分グラフ (subgraph induced by X)** とは $G(X) = (X, \{(u, v) \in E \mid u, v \in X\})$ である。頂点部分集合 $X \subseteq V$ に対し、X を新しい頂点 v_X で置き換え、$E_G(X)$ 内の辺 (x, u) $(x \in X)$ を (v_X, u) で置き換える操作を X **の縮約 (contract)** と呼ぶ。すなわち、$X \subseteq V$ の縮約によってグラフ $G = (V, E)$ は

$$\begin{aligned} V' &= V - X \cup \{v_X\}, \text{ and} \\ E' &= E - \{(v, u) \mid v \in X, u \in V - X\} \\ &\qquad \cup \{(v_X, u) \mid u \in V - X\}. \end{aligned}$$

であるグラフ $G' = (V', E')$ に置き換わる。

なお、上の操作における (v_X, u) と (v, u) を同一視する。すなわち、辺 (v, u) は G' において (v_X, u) として残る。ここでは多重グラフを扱っているので、もし異なる二つの $v, v' \in X$ に対して $(v, u), (v', u) \in E(X)$ である場合には、並列辺（どちらも (v_X, u) と表現される）が残り、その片方が (v, u) に、もう片方が (v', u) に対応している。

二つのグラフ $G_1 = (V_1, E_1)$ と $G_2 = (V_2, E_2)$ は頂点間に一対一対応 $\Phi : V_1 \to V_2$ が存在し、任意の $u, v \in V_1$ に対して $|E_{G_1}(u, v)| = |E_{G_2}(\Phi(u), \Phi(v))|$ が成立するときに**同型 (isomorphic)** であるという。グラフの**性質 (property)** とは、同型性について閉じている、グラフの部分集合である（無限集合でも良い）。例えば、「平面性」とは、すべての平面グラフの集合のことであり、これは（もちろん）同型性について閉じている。

定義 2.1 (ϵ-far) $G = (V, E)$ と $G' = (V', E')$ を同じ頂点数 $|V| = |V'| = n$ である二つのグラフとする。G を G' と同型にするために必要な削除あるいは付加する辺の数の最小値を $m(G, G')$ とする。G と G' の**距離 (distance)** を $\mathrm{dist}(G, G') = m(G, G')/n$ で定める。*2 G と G' が ϵ **遠隔 (ϵ-far)** であるとは、$\mathrm{dist}(G, G) > \epsilon$ であることをいい、そうでない場合には ϵ **近接 (ϵ-close)** であるという。P を空でない性質とする。グラフ G と性質 P との**距離**を $\mathrm{dist}(G, P) = \min_{G'' \in P} \mathrm{dist}(G, G'')$ で定義する。$\mathrm{dist}(G, P) > \epsilon$ であるとき、G は P より ϵ **遠隔**であるといい、そうでないとき、ϵ **近接**であるという。 □

性質 P に対する**検査アルゴリズム (testing algorithm)** もしくは**検査機 (tester)** とは、グラフ $G = (V, E)$ に対し、与えられた質問をすることで G の情報を得て、G が性質 P を満たすならば確率 2/3 以上で受理し、G が性質 P から ϵ 遠隔ならば確率 2/3 以上で拒否するものであり、質問回数が G によらない（ただし ϵ には依存して良い）定数回で抑えられるようなもののことをいう。一般グラフモデルに対する質問は、任意の頂点 $v \in V$ に対して、一回の質問で、次数 $d(v)$ を訪ねるか、$1 \leq i \leq d(v)$ である整数 i について、v の i 番目の隣接頂点が何かを聞くことができる。*3

グラフ $G = (V, E)$ に対して、その孤立クリークをすべて縮約して得られたグラフを $\mathcal{E}(G)$ と表す。なお、異なる二つの孤立クリークはその定義から、重なることはないので、上記の操作は一意的に定まる。

定義 2.2 (超有限 [7]) 二つの正実数 $t > 0$ と $\epsilon > 0$ に対し、n 頂点よりなるグラフ $G = (V, E)$ が (t, ϵ) **超有限 ((t, ϵ)-hyperfinite)** であるとは、G より高々 ϵn 本の辺を削除することによって、各連結成分の大きさ（頂点数）を t 以下にできることをいう。関数 $\rho : \boldsymbol{R}^+ \to \boldsymbol{R}^+$ に対し、G が ρ **超有限 (ρ-hyperfinite)** であるとは、任意の $\epsilon > 0$ に対し $(\rho(\epsilon), \epsilon)$ 超有限であることである。グラフの集合 $\mathcal{G}$ が ρ **超有限**であるとは、任意の $G \in \mathcal{G}$ が ρ 超有限であること

*2 $m(G, G') > n$ となりえることから、この定義では距離は 1 を超える場合がある。しかし我々は疎なグラフのみを扱うので、グラフ集合には平均次数の上界値（定数）d があり、従って距離の上限も $d/2$ で抑えられる。

*3 一般グラフモデルの場合、2 頂点 $u, v \in V$ に対して、辺 (u, v) が存在するか否かを聞くこともできるとすることが一般的であるが、本稿で扱うアルゴリズムはこの質問は使用しない。

である。グラフの集合 $\mathcal{G}$ が**超有限 (hyperfinite)** であるとは、$\mathcal{G}$ が ρ 超有限　となるような関数 ρ が存在することである。 □

超有限性は、次数制限グラフにおいては、マイナーについて閉じているすべての性質を含む、たいへん広いクラスである。そして超有限性は次の定理の存在によって、性質検査において非常に重要なクラスとなっている。

定理 2.3 ([18]) 次数制限グラフモデルにおいて、超有限なグラフクラスに対し、任意の性質は検査可能である。

3. 本稿の成果

3.1 クラス $\mathcal{SF}$ と $\mathcal{HSF}$

どんな性質でも検査可能であるとする定理 2.3 はたいへん強力である。しかし一方で、現実のスケールフリーネットワークにおいては次数の非常に大きなハブと呼ばれる頂点が存在すると言われており、次数制限グラフモデルは適さない、という問題点がある。我々はこの強力な結果を現実のスケールフリーネットワークに適用できないか考察した。そして、スケールフリーネットワークの自然なモデル化である $\mathcal{SF}$ と $\mathcal{HSF}$ というクラス（後者は前者の部分集合）を以下のように定義する。

定義 3.1 三つの正実数 $c_1, c_2 > 1$ と $\gamma_1 > 1$ に対し、クラス **Scale-Free Graphs (SF)** $\mathcal{SF} = \mathcal{SF}(c_1, c_2, \gamma_1)$ を以下の条件を満たす（多重）グラフ $G = (V, E)$ の集合と定義する。

(i) $d(v) = i$ である頂点 $v \in V$ の数を ν_i とすると、次の式が成立する。

$$\lfloor c_1 i^{-\gamma_1} n \rfloor \le \nu_i \le c_2 i^{-\gamma_1} n, \quad \forall i \in \{2, 3, \ldots, \}. \qquad (1)$$

□

上記の性質 (i) は一般に「指数法則 (power-low)」と呼ばれ、スケールフリーグラフの代表的な性質とされる。さらに多くの場合、$2 < \gamma_1 < 3$ を満たすとされている [1]。

定義 3.2 (Hierarchical Scale-Free Graphs) 六つの正実数 $c_1, c_2, c_3, c_4 > 1$ と $\gamma_1, \gamma_2 > 1$ に対し、クラス **Hierarchical Scale-Free Graphs (HSF)** $\mathcal{HSF} = \mathcal{HSF}(c_1, c_2, c_3, c_4, \gamma_1, \gamma_2)$ を以下の条件を満たす（多重）グラフ $G = (V, E)$ の集合と定義する。

(i) $G \in \mathcal{SF}(c_1, c_2, \gamma_1)$

(ii) $|Q| = i$ である孤立クリーク $Q \subseteq V$ の個数を μ_i とすると、次の式が成立する：

$$\lfloor c_3 i^{-\gamma_2} n \rfloor \le \mu_i \le c_4 i^{-\gamma_2} n, \quad \forall i \in \{2, 3, \ldots, \}, \qquad (2)$$

(iii) $\mathcal{E}(G) \in \mathcal{HSF}$. □

上記の (ii) と (iii) は [21] によって指摘されたものであり、多くのスケールフリーネットワークにおいて $\gamma_2 \approx \gamma_1 + 1$ であることが同文献内で指摘されている。

3.2 定理

我々は以下の成果を得た。

定理 3.3 $\gamma_1, \gamma_2 > 2$ を満たす任意の $\mathcal{HSF} = \mathcal{HSF}(c_1, c_2, c_3, c_4, \gamma_1, \gamma_2)$ と任意の正実数 $\epsilon > 0$ に対し、$\mathcal{HSF}$ が $(t_{3.3}, \epsilon)$ 超有限となるような正実数 $t_{3.3} = t_{3.3}(\mathcal{HSF}, \epsilon)$ が存在する。 □

そして、本定理で示された超有限性によって定義される頂点集合の分割を得る大域的アルゴリズムを与える。この分割アルゴリズムは乱数を用いないので、決定的である。すなわち、グラフ $G = (V, E)$ とパラメータ ϵ が固定されれば、分割も一通りに定まる。そして、このアルゴリズムは局所的なものに容易に変換できる。すなわち、我々は「決定性の」分割神託を得る。なお、これまでに知られている分割神託はすべて乱択アルゴリズムである。この分割神託を用いると、[18] とほぼ同じ議論によって、次の定理を得る。

定理 3.4 任意の性質は $\gamma_1, \gamma_2 > 2$ である任意の $\mathcal{HSF}(c_1, c_2, c_3, c_4, \gamma_1, \gamma_2)$ 上で検査可能である。 □

証明については [13] を参照のこと。

4. まとめと今後の課題

我々はスケールフリーネットワークの自然なモデル化である多重グラフのクラス $\mathcal{HSF}$ を定義し、そのクラス内の広い部分クラス（指数法則の指数が 2 より大きいもの）が超有限であることを証明した（定理 3.3）。この結果を用いて、このクラスにおいては任意の性質が検査可能であることを証明した（定理 3.4）。

$\mathcal{HSF}$ は孤立クリークの分布の階層構造を利用している。ここでの「孤立クリーク」は「孤立した密な部分集合 [14]」に拡張できる可能性が有る。そのようなさらに広いクラスにおいてもやはり超有限性が成立する可能性があるので、それを調べることは今後の重要な研究課題である。

謝辞

日頃から様々な情報をいただき、本研究に際しても非常に貴重なご意見を賜った、国立情報学研究所の吉田悠一准教授に心より感謝いたします。文献 [21] の著者であり、スケールフリーネットワークに関して色々ご教授いただいた東京工業大学の渡辺治教授および大阪府立大学の宇野裕之准教授に深く感謝の意を表します。本研究は科学技術振興機構クレスト「ビッグデータ時代に向けた革新的アルゴリズム基盤 (ABD14)」および、文部科学省 科研費 新学術領域研究「多面的アプローチの統合による計算限界の解明 (ELC) (24106003)」、日本学術振興会 科研費 挑戦的萌芽研究「ゲーム解析の新パラダイム (24650006)」から補助を受けて行われました。ここに感謝いたします。

参考文献

[1] R. Albert and A. -L. Barabási: Statistical mechanics of complex networks, Review of Modern Physics, Vol. 74, 2002, pp. 47–97.

[2] N. Alon, E. Fischer, I. Newman, and A. Shapira: A combinatorial characterization of the testable graph properties: it's all about regularity, SIAM J. Comput., Vol. 39, No. 1, 2009, pp. 143–167.

[3] I. Benjamini, O. Schramm, and A. Shapira: Every minor-closed property of sparse graphs is testable, Proc. STOC 2008, ACM, 2008, pp. 393–402.

[4] A. Bottreau, Y. Métivier: Some remarks on the Kronecker product of graphs, Information Processing Letters, Vol. 68, Elsevier, 1998, pp. 55–61.

[5] A.Z. Broder, S.R. Kumar, F. Maghoul, R. Raghavan, S. Rajagoplalan, R. Stata, A. Tomkins and J.L. Wiener: Graph structure in the Web, Computer Networks, Vol. 33, 2000, pp.309–320.

[6] C. Cooper and R. Uehara: Scale free properties of random k-trees; Mathematics in Computer Science, Vol. 3, 2010, pp. 489–496.

[7] G. Elek: L^2-spectral invariants and convergent sequence of finite graphs, Journal of Functional Analysis, Vol. 254, No. 10, 2008, pp. 2667–2689.

[8] Y. Gao: The degree distribution of random k-trees, Theoretical Computer Science, Vol. 410, 2009, pp. 688–695.

[9] O. Goldreich (Ed.): Property Testing — Current Research and Surveys, LNSC 6390, 2010.

[10] O. Goldreich and D. Ron: Property testing in bounded degree graphs: Proc. STOC 1997, 1997, pp. 406–415.

[11] O. Goldreich, S. Goldwasser, and D. Ron: Property testing and its connection to learning and approximation: Journal of the ACM, Vol. 45, No. 4, July, 1998, pp. 653–750.

[12] A. Hassidim, J. A. Kelner, H. N. Nguyen, and K. Onak: Local graph partitions for approximation and testing, Proc. FOCS 2009, IEEE, pp. 22–31.

[13] H. Ito, Every property is testable on a natural class of scale-free multigraphs, arXiv: 1504.00766, Cornell University, April 6, 2015.

[14] H. Ito and K. Iwama, Enumeration of isolated cliques and pseudo-cliques, ACM Transactions on Algorithms, Vol. 5, Issue 4, Oct. 2009, Article 40.

[15] H. Ito and K. Iwama, and T. Osumi: Linear-time enumeration of isolated cliques, Proc. ESA2005, LNCS, 3669, Springer, 2005, pp. 119–130.

[16] J. Kleinberg and S. Lawrence: The structure of the Web, Science, Vol. 294, 2001, pp. 1894–1895.

[17] M. Mahdian and Y. Xu: Stochastic Kronecker graphs, Proc. WAW 2007, LNCS, 4863, Springer, 2007, pp. 179–186.

[18] I. Newman and C. Sohler: Every property of hyperfinite graphs is testable, Proc. STOC 2011, ACM, 2011, pp. 675–784.

[19] M. E. J. Newman: The structure and function of complex networks, SIAM Review, Vol. 45, 2003, pp. 167–256.

[20] R. Levi and D. Ron: A quasi-polynomial time partition oracle for graphs with an excluded minor, Proc. ICALP 2013 (1), LNCS, 7965, Springer, 2013, pp. 709–720.

[21] T. Shigezumi, Y. Uno, and O. Watanabe: A new model for a scale-free hierarchical structure of isolated cliques, Journal of Graph Algorithms and Applications, Vol. 15, No. 5, 2011, pp. 661–682.

[22] D.J. Watts and S.H. Strogatz: Collective dynamics of 'small-world' networks, Nature, Vol. 393, 1998, pp. 440–442.

[23] Z. Zhang, L. Rong, and F. Comellas: High-dimensional random apollonian networks, Physica A: Statistical Mechanics and its Applications, Vol. 364, 2006, 610–618.