数理モデル化と問題解決 38－5
（2002． 3． 4）

メモリ領域が小さい
確率的なグラフ連結性判定アルゴリズムについて

池田 諭　品野 勇治　中森 眞理雄
〒184-8588 東京都小金井市中町 2 － 24 － 16 東京農工大学工学部
e-mail : {bisu, yshinano, nakamori}@cc.tuat.ac.jp

概要

与えられたグラフ $G=(V,E)$ が連結か否かを判定することを考える．$O(|V|+|E|)$ のメモリ領域を用いることが出来るならば，$O(|V|+|E|)$ の手間で連結性が判定できることが R. E. Tarjan によって示されている．本論文の目的は，分散システム上で動作する連結性判定分散アルゴリズムを求めることである．

本論文では，R.Aleliunas, R.M.Karp, R.J.Lipton, L.Lovaász および C.Rackoff によるランダムウォークを用いる手法を取り上げている．そして，簡単な前処理をすることが可能な状況ならば，$O(|V|^2 \log |V|)$ の手間でほぼ確実に無向グラフの連結性が判定出来ることを示した．また，有向グラフの場合もグラフの直径と次数の上界が分かっているならば同様にして多項式オーダーで判定できることを示した．

Probabilistic Algorithm for the Reachability Problem with small memory area

Satoshi IKEDA　Yuji SHINANO　Mario NAKAMORI
Tokyo University of Agriculture and Technology
Nakacho 2-24-16, Koganei, Tokyo 184-8588
e-mail : {bisu, yshinano, nakamori}@cc.tuat.ac.jp

abstract

We consider the reachability problem for directed and undirected graphs. Let $G=(V,E)$ be a directed graph. In [1], R. E. Tarjan has showed the algorithm of $O(|V|+|E|)$ sapce and time complexity. We show a probablistic distributed algorithm for this problem in space $O(1)$ and time $O(|V|^2 \log |V|)$.

In this paper, we discuss a variant of the algorithm which based on random walk, due to R.Aleliunas, R.M.Karp, R.J.Lipton, L.Lovaász, and C.Rackoff. We show an improvement of their algorithm from $O(|V|^3 \log |V|)$ to $O(|V|^2 \log |V|)$ in time helped with a simple preprocessing. In a directed graph, we also show their original algorithm run in time $O(|V|^2 \log |V|)$, if the diameter and the degree of the graph is bounded.

1 はじめに

与えられたグラフ $G=(V,E)$ が連結か否かを判定することを考える．良く知られているように，$O(|V|+|E|)$ のメモリ領域を用いることが出来るならば，$O(|V|+|E|)$ の手間で連結性が判定できる (R. E. Tarjan[1])．しかし，この手法ではインターネットのサイトの接続関係をモデル化したような巨大なグラフを対象とする場合には適さない．そこで，分散システム上で動作する連結性判定アルゴリズムが必要となる．本論文の目的は，分散システム上で動作する連結性判定分散アルゴリズムを求めることである．

論文 [2] の中で，R.Aleliunas, R.M.Karp, R.J.Lipton, L.Lovaász および C.Rackoff は，$O(|V||E| \log |V|)$ の長さのランダムウォークによって，ほぼ確実に無向グラフの連結性が判定出来ることを主張している．彼らの手法は，つぎのようなものである．無向グラフの頂点 u, v が連結かどうかを判定するものとする．u から出発して，隣接する頂点を等確率で選んで移動するという操作を繰り返す．そして，十分な操作を繰り返しても v に到達出来ないならば u と v は連結でないと判定するのである．もちろん，誤った結論を導く可能性もあるがその確率が十分小さいならば実用上は問題がないと考えられる．

本論文では，簡単な前処理をすることが可能な状況ならば，$O(|V|^2 \log |V|)$ の手間で彼らと同様な主張が出来ることを示す．また，モデルが有向グラフの場合は，グラフの直径と次数の上界が分かっているならば，論文 [2] の手法で多項式オーダーで判定できることを示す．現在，インターネット上の強連結成分の直径は，20 以下であると言われているので [7]，この仮定は非現実的なものではない．

2　記号の定義

特に断りがない限り，本論文を通して，グラフ $G=(V,E)$ は無向で連結な単純グラフとする．ただし，$|V|=n, |E|=m$ であるものとしておく．頂点 $u\in V$ に対して，隣接する頂点の集合を $N(u)=\Big\{v:\{u,v\}\in E\Big\}$ と記す．グラフ G は無向であるから $v\in N(u)\iff u\in N(v)$ が成り立つ．頂点 u の次数 $|N(u)|$ を $\deg(u)$ と表す．また，頂点の次数の最大値をグラフ G の次数と呼び，$\deg G$ と記すことにする．すなわち，$\deg G=\max_{v\in V}\deg(v)$ である．グラフ G の直径を $D(G)$ と記し，

$$D(G)=\max_{u,v\in V}D(u,v)$$

で定義する．ただし，$D(u,v)$ は，$u\neq v$ であるときは

$$D(u,v)=\inf\Big\{\ell\in\mathbf{N}\mid u=v_0,v_1,\cdots,v_\ell=v,\ \{v_i,v_{i+1}\}\in E\text{ for }i=0,1,\cdots,\ell-1\Big\}$$

とし，$u=v$ であれば $D(u,v)=0$ と定義する．

頂点の片側無限列の集合を $\Omega=V^{\mathbf{N}\cup\{0\}}$ とする（ただし，$\mathbf{N}$ は自然数の集合とする）．集合 Ω の元を $\omega=(\omega_0,\omega_1,\cdots)$ と表記することにする．ω の $i+1$ 番目の要素 ω_i のことを $X_i(\omega)$ と書くことにする．すなわち，$\omega=(\omega_0,\omega_1,\cdots)\in\Omega$, $n\in\mathbf{N}\cup\{0\}$ に対して，

$$X_n:\Omega\to V,\quad X_n(\omega)=\omega_n$$

とする．

集合 Ω 上の Markov 測度全体の集合を $\mathcal{M}(\Omega)$ とする．Markov 測度 $\mu\in\mathcal{M}(\Omega)$ は，初期分布 $\mathbf{q}^{(0)}=(q_v^{(0)})\in[0,1]^V$ と遷移確率行列 $Q=(q_{uv})\in[0,1]^{|V|\times V}$ を持つものとする．集合 Ω は，$G=(V,E)$ 上のランダムウォークの通った道を表すものだから，一般性を失うことなく以後は扱う測度のクラスを

$$\mathcal{M}^+(\Omega)=\left\{\mu\in\mathcal{M}(\Omega)\ \middle|\ q_{uv}>0\text{ if }v\in N(u)\ \text{かつ}\ q_{uv}=0\text{ if }v\notin N(u)\cup\{u\}\right\}$$

に制限して議論することにする．$\omega=(\omega_0,\omega_1,\cdots)\in\Omega$ と $v\in V$ に対して

$$t(\omega)=\inf\Big\{i\in\mathbf{N}\cup\{0\}\mid\{X_0(\omega),X_1(\omega),\cdots,X_i(\omega)\}=V\Big\},\sigma_v(\omega)=\inf\Big\{i\in\mathbf{N}\mid X_i(\omega)=v\Big\}$$

とする．$t(\omega)$ は，グラフ $G=(V,E)$ 上の隣接した頂点を $\omega_0,\omega_1,\cdots$ の順番にランダムウォークして全ての頂点をカバーするまでのステップ数を意味している．与えられた自明でないグラフ $G=(V,E)$ と $\mu\in\mathcal{M}^+$ に対して $v\in V$ の mean hitting time $H_\mu(u,v)$ と cover time $C_\mu(G)$ を次のように定義する．

$$H_\mu(u,v)=E_\mu[\sigma_v(\omega)\mid X_0(\omega)=u],\quad C_\mu(G)=\max_{u\in V}C_\mu(G,u).$$

ただし，$u\in V$ に対して $C_\mu(G,u)=\mathbf{E}_\mu\Big[t(\omega)|X_0(\omega)=u\Big]$ であるものとする．$H_\mu(u,v)<\infty$ かつ $C_\mu(G)<\infty$ である．また，一般に $H_\mu(u,v)\neq H_\mu(v,u)$ である．

本論文では，遷移確率 $P=(p_{uv})\in[0,1]^{V\times V}$, $\hat{P}=(\hat{p}_{uv})\in[0,1]^{V\times V}$ がそれぞれ

$$p_{uv}=\begin{cases}1/\deg(u) & \text{if }v\in N(u),\\ 0 & \text{otherwise,}\end{cases}\tag{2.1}$$

$$\hat{p}_{uv}=\begin{cases}0 & \text{if }u\neq v\text{ and }v\notin N(u),\\ \min\{1/\deg(u),1/\deg(v)\} & \text{if }v\in N(u),\\ 1-\sum_{w\in N(u)}\hat{p}_{uw} & \text{otherwise if }u=v.\end{cases}\tag{2.2}$$

で与えられる2種類のランダムウォークを扱う．ただし，初期分布はそれぞれ $\mathbf{p}^{(0)}=(p_v^{(0)})\in[0,1]^{|V|}$, $\hat{\mathbf{p}}^{(0)}=(\hat{p}_v^{(0)})\in[0,1]^{|V|}$ としておく．以後，遷移確率が (2.1),(2.2) で与えられるマルコフ測度をそれぞれ $\tilde{\mu}$ および $\hat{\mu}$ と記す．明らかに $\tilde{\mu},\hat{\mu}\in\mathcal{M}^+(\Omega)$ である．本質的でない議論を避けるため，以降の議論では遷移確率行列が非周期的な場合のみを扱うことにする．

3 主要な結果

論文 [2] の中で，R.Aleliunas, R.M.Karp, R.J.Lipton, L.Lovaász および C. Rackoff は，グラフの連結性を確率的に判定する手法を提案している．彼らの手法は，与えられたグラフ上のランダムウォークを用いるもので現在の状態だけを保持するだけの小さな記憶領域しか必要としない．彼らの主張は，次のものである．

定理 1 任意の無向連結グラフ $G=(V,E)$ と $r \geq 1$ に対して，

$$\tilde{\mu}\left(\left\{\omega \in \Omega \,\middle|\, t(\omega) \geq r|V||E|\right\}\right) \leq 1/r$$

が成り立つ．

これは，$|V||E|$ の r 倍回のランダムウォークによって到達できない 2 頂点が連結である確率は，$1/r$ 以下であることを主張するものである．よって，$O(|V|^3 \log |V|)$ の試行で，2 頂点の連結性をほぼ確実に判定できることがわかる．つぎに我々の主要な結果を示す．

定理 2 与えられた無向連結グラフ $G=(V,E)$ の任意の $u,v \in V$ および $r>0$ に対して，

$$\hat{\mu}|_u\left(\left\{\omega \in \Omega \,\middle|\, \sigma_v(\omega) \geq rD(G)|V|,\ X_0(\omega)=u\right\}\right) < \min\left\{6/r, 1\right\}$$

が成り立つ．ただし，$u \in V$ に対して $\hat{\mu}|_u$ は，$\hat{\mu}$ を $\left\{\omega \mid X_0(\omega)=u\right\}$ へ制限した条件つき確率測度とする．

定理 2 は，実際は連結である 2 頂点の一方から出発して $D(G)|V|$ の r 倍の推移を繰り返したときに，もう一方に到達出来ない確率は $6/r$ 以下であることを意味している．常に $D(G) \leq |V|-1$ が成り立つことから，この結果は，$O(|V|^2 \log |V|)$ の試行で，ほぼ確実に 2 頂点の連結性を判定出来ることを意味している．この計算量の意味で，Aleliunas らが提案する遷移規則より我々の提案する手法の方が有利である．

一方，良く知られているように，[1] で Tarjan は，$O(|V|^2)$ の手間で任意の有向グラフの連結性が判定できることを示した．彼の手法は，決定論的なアルゴリズムなため判定は確実である．また，計算量の観点からも有利であると考えられる．しかし，記憶領域に関しては，Tarjan の手法は，$O(|V|+|E|)$ の領域を必要とするのに対し，我々の手法では $O(1)$ の小さな領域で十分である．確率的な手法は，とてつもなく大きなグラフの連結性を判定するのに適していると考えられる．

定理 1 と定理 2 の主張は少し違っている．前者の評価は，cover time を使っているのに対して，後者は，mean hitting time を用いているためである．2 頂点の連結性の評価をするだけなので，定理 1 は，cover time の代わりに mean hitting time を用いることで，より良い評価を得られる可能性がある．しかし，実際は定理 1 の証明のために mean hitting time を用いても，本質的な改良にならない．なぜならば，

$$H_{\hat{\mu}}(u,v) = O(n^2)$$

であるのに対して，

$$H_{\tilde{\mu}}(u,v) = \Theta(n^3)$$

となることがあるからである．このため，cover time の代わりに mean hitting time を用いても，定理 1 の評価を改善することは出来ない．定理 1 および定理 2 は，それぞれオーダーとしては最良の評価であり，オーダーの意味で等号が成り立つことが実際にある．すなわち，定理 1 と定理 2 の違いは評価するときのテクニカルな問題ではなく実際に定理 2 の設定の方が早いのである．

4 有向グラフの場合

これまでは，無向グラフについて議論してきた．しかし，Web のグラフ構造を考えると有向グラフで考える方が現実的である．この章では，有向グラフの場合を議論する．以後は，$G=(V,E)$ を有向グラフの意味で単純グラフであり，かつ強連結であるものとする．ランダムウォークが，有向辺に沿って移動すること以外は，3 章までの設定と記号を踏襲して使用する．

頂点 u から v へ向かう有向辺を $(u,v)\in E$ と記す．また，$u\in V$ に対して，隣接する頂点の集合は，

$$N^-(u)=\Big\{v:(u,v)\in E\Big\},\ N^+(u)=\Big\{v:(v,u)\in E\Big\}$$

と記す．また，以下同様にして

$$\deg^-(u)=|N^-(u)|,\ \deg^-(G)=\max_{u\in V}\deg^-(u)$$

および

$$\deg^+(u)=|N^+(u)|,\ \deg^+(G)=\max_{u\in V}\deg^+(u)$$

とする．直径の定義は，無向辺を有向辺に置き換えるだけで良い．また，(2.2) は，つぎのようになる．

$$p_{uv}=\begin{cases}1/\deg^-(u) & \text{if } v\in N^-(u),\\ 0 & \text{otherwise.}\end{cases} \tag{2.2$'$}$$

有向グラフに関して，一般につぎの評価が得られる．

定理 3 与えられた有向連結グラフ $G=(V,E)$ の任意の $u,v\in V$ および $r>0$ に対して，

$$\tilde{\mu}|_u\left(\Big\{\omega\in\Omega\ \Big|\ \sigma_v(\omega)\geq r|V|^2,\ X_0(\omega)=u\Big\}\right)<\deg^-(G)\Big(\deg^-(G)\deg^+(G)\Big)^{D(G)}r^{-1}$$

が成り立つ．

5 まとめ

本論文では，インターネットのサイトの接続関係をモデル化したような巨大なグラフの連結性を判定するための，確率的な分散アルゴリズムについて考察した．

その結果，前処理が可能な状況ならば従来知られていた判定方法より少ない手間 $O(n^2\log n)$ で連結性を判定できるアルゴリズムを提案した．ただし，n は頂点の数を意味するものとする．

また，有向グラフの場合，従来の手法での手間の評価を行った．そして，グラフの半径と次数が有界ならばやはり $O(n^2\log n)$ で判定できることを示した．

References

[1] R. E. Tarjan, "Depth first search and linear graph algorithms," *SIAM J. on Computing,* Vol.1, pp.146-160, 1972.

[2] R. Aleliunas, R.M Karp, R.J. Lipton, L. Lovaász, and C. Rackoff, "Random Walks, universal traversal sequences, and the complexity of maze problems," *Proc. 20th Ann. Symposium on Foundations of Computer Science,* pp.218-223, 1979.

[3] G. Brightwell and P. Winkler, "Maximum hitting time for random walks on graphs," *J. Random Structures and Algorithms,* Vol.3, pp.263-276, 1990.

[4] D. Coppersmith, P. Tetali, and P. Winkler, "Collisions among random walks on a graph," *SIAM J. on Discrete Mathematics,* Vol.6, No.3, pp.363-374, 1993.

[5] R. Motowani and P. Raghavan, *Randomized Algorithms,* Cambridge University Press, New York, 1995.

[6] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," 1998, http://www-db.stanford.edu/ backrub/pageranksub.ps

[7] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi, " Internet: Diameter of the World-Wide Web," *Nature,* Vol.401, pp.130-131, 1999.

[8] S. Ikeda, I. Kubo, N. Okumoto, and M. Yamashita, "Fair circulation of a token," *IEEE Trans. Parallel and Distributed Systems,* Vol.13, No.4, 2002 (to appear).