LA-13

連結グラフの高速ランダム生成法

Fast Random Generation of Connected Graphs

河邊昌彦 † 二村良彦 †
Masahiko Kawabe Yoshihiko Futamura

1 初めに

従来のランダムグラフ生成法を用いては現実的には生成不可能であった，100,000 頂点 500,000 辺程度の大規模でかつランダムネスの保証されたラベル付き連結グラフでも，PC を用いて 4 分以内で生成可能にする $O(m \log m)$ アルゴリズムについて報告する．

最小全域木問題や最短経路問題などといった，グラフに関する問題を解くアルゴリズムが，数多く研究され開発されている．また，インターネットのような大規模ネットワークをグラフによって抽象化し，シミュレーションによって解析が行われることが多い．さらには，並列化コンパイラの研究においても，実行すべきタスクの関係がグラフによって表現されている．このように，グラフを扱うアルゴリズムは幅広い分野で応用されており，それらの性能を評価することは非常に重要である．

グラフアルゴリズムに対して評価を下すための一つの手段は，実際にグラフを入力して結果を解析することである．その際，精密な評価を得るためには，指定された性質を有しかつランダムネスの保証されたグラフを大量に生成する必要がある．また，評価にかかる時間を短くするためには，そのようなグラフの生成を高速に行わなければならない．

例えば，[3] では最小全域木問題を線形時間で解く確率的アルゴリズムが示されている．このアルゴリズムに対して，実際に統計的に線形時間で動くかということを検証する場合や，或いは他のアルゴリズムと性能を比較する場合を考える．すると，高精度に生成されたランダムグラフが大量に必要であることが容易に理解されるであろう．

指定された頂点数 n と辺数 m を有し頂点にラベルの付いた無向連結グラフに関して，現在までのところ，そのランダム生成には以下に挙げる方法が主に採られている．例えば，全域木に辺を追加していくことで連結グラフを得る方法 (全域木法) がある．また，n 個の頂点をランダムな m 本の辺で結び，それが連結グラフとなるまで繰り返す方法 (生成検査法) もある．或いは，構成比という指標を用いてランダムグラフを生成する方法 (構成比法) も提案されている [1,2]．しかし，これらの方法にはそれぞれの問題点が存在し，実用的に用いることを困難にしている．

本稿では，これらの問題点を実験データに基き明らかにする．その後で，構成比法における構成比を，頂点数に対する辺数の比 (辺密度) を用いて近似する，構成比近似法を提案する．そして，実際の生成時間と近似精度を示すことによって，実用的であることを示し，今後の応用についても触れる．

† 早稲田大学理工学部

2 従来の方法の問題点

全域木法は，簡単に実装できるランダムグラフ生成法として，用いられる場合がしばしばある．しかしこの方法は簡便法であり，各連結グラフの生成される確率が一様とはならない．実際に χ^2 検定を行った結果が表 1 である．ここでは頂点数が 6 で辺数が 7 から 14 までのグラフを，各 100 万個ずつ全域木法で生成し，χ^2 統計量を計算した．その値が，ランダムであると見なせる上限と下限の間から大きくはみ出しており，即ち不適正であることに注意されたい．

生成検査法も素朴なアルゴリズムであり，実装が容易であるため，用いられていることが多い．そのアルゴリズム上，生成された連結グラフのランダムネスは理論的に保証されている．従って，グラフアルゴリズムの高精度な評価を行える能力を備えている．

しかし，辺数が頂点数に近づくと，ランダムに作成したグラフが連結となる確率が急激に低下する．それにより，連結グラフが生成されるまでの繰り返し回数が増加し，高速に生成することが難しくなる (表 2)．よって，生成検査法は辺数が頂点数に近い時は実用的ではない．

構成比法は，連結グラフの総数に関する再帰方程式に基くアルゴリズムである．構成比，即ち連結グラフの総数に対する特定の下部構造を有する連結グラフの割合を求め，これに従ってグラフを生成する．この再帰方程式は [4–6] において示されている．また，計算機オーバーフローを回避して構成比を直接計算する方法は [1] で示されている．

このアルゴリズムは，理論的にランダムネスが保証されており，また，実際にもそれが確認されている [1]．そして，頂点数 n，辺数 m のランダムグラフを 1 個につき $O(mn)$ で生成することが可能である．しかし，そのためには $O(m^2n^2)$ 時間と $O(mn^2)$ 空間を要して構成比の表を予め作成しておく必要がある．従って，頂点数や辺数が大きくなるにつれて実用が難しくなる (表 2)．

頂点数	辺数	上限	下限	χ^2 統計量
6	7	5850.997	5549.003	55759.176
6	8	6322.035	6007.965	48968.240
6	9	5085.642	4804.358	36077.211
6	10	3106.490	2887.510	22049.935
6	11	1438.892	1291.108	11138.063
6	12	497.661	412.339	4283.022
6	13	125.494	84.506	1140.982
6	14	22.746	7.254	16.857

表 1: 全域木法の χ^2 検定 (サイズ 1,000,000)

頂点数	辺数	生成検査法	構成比法	構成比近似法
100	110	*	65.014	0.022
100	150	0.590	142.615	0.028
100	200	0.016	341.220	0.033
100	300	0.005	938.870	0.042
1,000	1,500	*	*	0.346
1,000	2,000			0.402
1,000	2,500	126.292		0.445
1,000	2,600	49.500		0.446
1,000	2,700	10.951		0.449
1,000	2,800	6.607		0.462
1,000	2,900	3.745		0.476
1,000	3,000	1.416		0.494
1,000	5,000	0.237		0.617
1,000	10,000	0.268		0.962
10,000	50,000	31.090	*	9.159
10,000	100,000	20.487		13.351
10,000	1,000,000	38.335		114.454
100,000	500,000	*	*	231.493
100,000	1,000,000			712.364

表 2: ランダムグラフの生成時間（単位は秒，PentiumIII 733MHz, Allegro Common Lisp 5.0 で測定，測定時間が大きくなりすぎた欄 (*) は空欄のまま残されている）

3 構成比近似法

構成比法のボトルネックは構成比を計算して表を作成する部分である．そこで我々は，構成比の単純な近似式を求めることによって，時間的コストおよび空間的コストの改善を目指すアプローチを取った．

まず，頂点数に対する辺数の比をグラフの辺密度と定義する．即ち，疎な連結グラフほど辺密度がほぼ 1 に近づき，密な連結グラフほど辺密度がほぼ $n/2$ に近づく．

すると，同じ辺密度を有する連結グラフは，構成比においても似通った性質を持つことが経験的に判明している．これは直感的には，例えば 50 頂点 100 辺の連結グラフと 100 頂点 200 辺の連結グラフの間には，何らかの線形関係がある，ということである．さらに，辺密度を固定して頂点数を十分に大きくした場合，構成比が辺密度の単純な関数で近似できることも判明している．即ち，予め構成比を計算して表を作らずとも，少ない計算量で辺密度から構成比を求めることが可能である．

また，構成比の別の特徴として，平均的な分布ではなく偏りの大きい分布であるということも判明している．これを利用することで，構成比に従う乱数を求める際の平均計算量を減少させることが可能である．

なお，これらは経験則に基くものであり，数学的証明はまだできていない．

以上の方法により，連結グラフの生成時間を約 $O(m \log m)$ にまで短縮することが可能となった (表 2)．特に重要な点は，生成検査法では実用が難しい辺密度の低いグラフに関しても，良い性能を示していることである．従って，グラフの疎密によって構成比近似法と生成検査法を使い分けることで，両アルゴリズムの長所を享受することができる．

辺数	真の構成比	構成比近似法 近似構成比	構成比近似法 誤差	全域木法 誤差
101	0.1727	0.1946	0.0219	0.8273
120	0.6667	0.6589	0.0078	0.3333
140	0.8210	0.8129	0.0081	0.1790
160	0.8957	0.8866	0.0091	0.1043
180	0.9366	0.9276	0.0090	0.0634
200	0.9606	0.9533	0.0073	0.0394

表 3: 構成比近似法と全域木法の誤差 (頂点数 100)

頂点数 100 の場合の，構成比近似法及び全域木法による構成比の近似誤差は表 3 となる．全域木法は構成比を陽には用いないが，この表における構成比を全て 1 と近似することに相当するので，それに基いて誤差を計算した．近似誤差がランダムネスに及ぼす数学的影響は不明であるが，誤差を比較する限り全域木法よりも精度が高いことが期待される．

4 終わりに

本稿では，まず，ランダムグラフ生成法の必要性について述べ，従来の生成法では不十分であることを示した．そして，辺密度を用いて構成比を近似する方法を報告し，それによって連結グラフをランダムに生成する構成比近似法が有効であることを示した．生成検査法の不得意とする疎なグラフに関しても高速に生成することが可能であり，また，全域木法に比べて近似誤差も小さい．構成比近似法と生成検査法を組み合わせることによって，幅広い頂点数，辺数のグラフに関して実用的な生成を行うことが可能である．

今後の課題としては，以上の成果をまとめ，研究および教育に利用できるようランダムグラフサーバとして実用化し，ネットワークを通じて広く公開することが挙げられる．それによって，テストデータとしてランダムグラフを必要とする研究者や学生が，容易に得ることができるようになる．また，頂点にラベルの付いた無向連結グラフ以外のグラフに関するランダムグラフサーバを実用化することも今後の課題である．

参考文献

[1] 矢農, 二村, “連結グラフのランダム生成法について,” アルゴリズム 60-1 (1998)

[2] 二村, 矢農, 他, “連結グラフの生成方法およびそのプログラム,” 特願平 9-259958 (1997)

[3] D.R.Karger, P.N.Klein, R.E.Tarjan, “A Randomized Linear-Time Algorithm to Find Minimum Spanning Trees,” Journal of the Association for Computing Machinery Vol 42, No 2 (1995), 321–328

[4] D.E.Knuth, “The Art of Computer Programming,” Vol.1, Addison-Wesley, Reading, Massachusetts (1973)

[5] A.Rényi, “On Connected Graphs I,” Publ. Math. Inst. Hungarian Acad. Sci. 4 (1959), 385–388

[6] N.C.Wormald, “Some Problems in the Enumeration of Labelled Graphs,” Ph.D. thesis, University of Newcastle (1978)