

Generating Connected Acyclic Digraphs Uniformly at Random

Fabrice Philippe, Guy Melançon

► To cite this version:

Fabrice Philippe, Guy Melançon. Generating Connected Acyclic Digraphs Uniformly at Random. Information Processing Letters, Elsevier, 2004, 90 (4), pp.209-213. lirmm-00108549

HAL Id: lirmm-00108549

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00108549>

Submitted on 23 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Generating connected acyclic digraphs uniformly at random

Guy Melançon, Fabrice Philippe

LIRMM CNRS UMR 5506

Montpellier, France

`{Guy.Melancon, Fabrice.Philippe}@lirmm.fr`

June 10, 2003

Abstract

We describe a simple algorithm based on a Markov chain process to generate simply connected acyclic directed graphs over a fixed set of vertices. This algorithm is an extension of a previous one, designed to generate acyclic digraphs, non necessarily connected.

Keywords graph algorithms, random generation, simply connected acyclic directed graphs.

1 Introduction

Combinatorial objects can sometimes be investigated through random generation in order to discover various properties, estimate statistics or test conjectures. Also, objects might be generated randomly to test the behaviour or performance of algorithms, with the hope that all possible or interesting cases will be obtained from the generation process. For example, graph drawing algorithms can be tested or studied by running them over a large set of examples. This approach is widely used and many authors provide test sets on the web (see [5, 8] for instance).

Graphs are widely used by non-mathematicians as a modelling tool in various fields (social sciences, computer science and biology to name just a few). A common approach is to develop a theoretical but specific framework based on graphs and to prove its relevancy and usefulness by experimentally testing it on many examples. However, the lack of "real-world" example sets motivate the construction of artificial ones. Randomly generating graphs appears as a solution, provided that the generating process can be somewhat controlled or kept within a given class of graphs.

Acyclic digraphs (directed graphs containing no circuits) prove to be appropriate in many situations. The absence of circuits can indeed have different interpretations (logical ordering of tasks or processes, ordering induced from time, etc.). Many authors have studied their properties providing insight on the global and average shape of acyclic digraphs [9, 10, 2, 3, 6]. An algorithm generating acyclic digraphs uniformly at random has only been published recently [7]. However, the authors did not provide any specific conditions showing how to use their algorithms to generate *simply connected* acyclic digraphs. This condition is often seen as being central in

many applications and motivated our work. As we shall see, a careful investigation of the Markov chain introduced in [7] reveals how and why the same algorithm can be slightly modified and used to generate simply connected acyclic digraphs uniformly at random.

This note is organized as follows. We first recall basic results and notations and describe the Markov chain introduced in [7]. We then discuss the conditions under which it can be adapted as to provide an algorithm for generating simply connected acyclic digraphs. The last part of the note is devoted to the proof of our result.

2 A Markov chain algorithm

Let $N \geq 2$ be a fixed integer and let $V = \{1, \dots, N\}$ denote a finite set of *vertices*. We consider the set \mathcal{A} of all acyclic directed graphs over V , that is, graphs containing no circuits¹. Next, we define the Markov chain M over the set \mathcal{A} . Each state of M is an acyclic digraph in \mathcal{A} . Because the set V of vertices is fixed, we will not distinguish between a digraph in \mathcal{A} and the set of its arcs. The transition between any two states in M is given as follows. Let X_t denote the state of the Markov chain at time t . Suppose a couple of integers (i, j) has been drawn uniformly at random from the set $V \times V$.

(T_1) If (i, j) is an arc in X_t , it is deleted from X_t . That is, $X_{t+1} = X_t \setminus (i, j)$.

(T_2) If (i, j) is not an arc in X_t , then

- (i) it is added to X_t , provided that the resulting graph is acyclic. That is, $X_{t+1} = X_t \cup (i, j)$.
- (ii) otherwise, nothing is done. That is, $X_{t+1} = X_t$.

The probability of a transition going from a state X to a state $Y \neq X$ is either 0 or $1/N^2$. Note that the transition rules make the transition matrix A symmetric. This is a key property in proving that the Markov chain converges towards the uniform distribution. More precisely, we have:

$$\lim_{n \rightarrow \infty} A^n \cdot d = \bar{1}, \quad (1)$$

where d is any initial distribution over \mathcal{A} and $\bar{1}$ stands for the normalized vector $(\frac{1}{|\mathcal{A}|}, \dots, \frac{1}{|\mathcal{A}|})$. The other key element in the proof is that the state space of M is *irreducible*. That is, given any two acyclic digraphs G_1 and G_2 , there exists a sequence of transitions going from G_1 to G_2 . This is straightforward since, starting from G_1 , one can travel by rule (T_1) to the graph with no arc (which is obviously acyclic) then by rule (T_2) to G_2 .

This result can be converted into an algorithm for randomly generating acyclic digraphs, each having the same probability of being drawn. Indeed, condition (T_2) relies on the ability to test whether the addition of (i, j) would create a circuit, which can be easily done [1]. Hence, starting from a graph with an empty set of arcs one can iteratively apply the rules (T_1) and (T_2) to build an acyclic digraph. Applying those rules a large (but finite) number of times leads to an algorithm, which randomly outputs acyclic digraphs (with an *almost* uniform distribution).

¹Recall that a circuit in a directed graph G is a sequence of vertices $v_0, v_1, \dots, v_k = v_0$ where each (v_i, v_{i+1}) , $i = 0 \dots k-1$, is an arc in G .

3 Customizing the algorithm

In [7, Theorem 1], the authors prove that random acyclic digraphs have $N^2/4$ arcs on average. However, in many applications, the total number of edges in graphs is proportional to their number of vertices². Hopefully, the chain can be adapted to these situations by imposing further conditions on the transition (T_2) . For instance, one can allow the insertion of a new arc provided that, additionally to condition (T_2) , the total number of arcs in the graph does not exceed a given upper bound. The condition can also be refined in order to impose an upper bound on each vertex individually.

Two observations must be made here. First, it is important to mention that additional conditions imposed on transitions (T_1) or (T_2) actually induce a restriction on the state space on which they act. The corresponding transition matrix may not be symmetric any more, so that other sufficient properties have to be established. Also, the irreducibility of this space has to be established before we can conclude that Eq. 1 holds for the restricted transition matrix. However, the situations considered in [7] are rather simple and in each case the irreducibility of the restricted space is straightforward.

The contribution of this short note is to show how the Markov chain can be adapted to produce simply connected acyclic digraphs³. In the sequel, we will refer to simply connected graph as connected graph, for sake of brevity.

In many applications, this appears to be an essential property in the modelling phase. Indeed, social networks are rarely made of several disconnected components, since it is exactly the interactions between the actors of the network that motivates their study. Also, disconnected acyclic digraphs are irrelevant in modelling task synchronization problems, since tasks belonging to distinct components can be performed independently. As a further example, chronology forces acyclicity in Bayesian networks.

An obvious approach one can adopt in order to adapt the Markov chain M to generate connected acyclic digraphs is to impose on the rule (T_1) an additional condition: the deletion of an arc should only be allowed if the arc is not *disconnecting*, that is, if the resulting graph is still connected. Furthermore, a disconnecting arc belongs to no (undirected) cycle, so that it may be reversed without creating a circuit. This seemingly superfluous rule has two advantages at least: it enables further transitions, slightly shortening the diameter of the transition graph, and it makes the chain irreducible for $N = 2$. Observe that the modified conditions preserve the symmetry of the transition matrix, admitted that the state space underlying this restricted Markov chain is now the set of all *connected* acyclic digraphs. So, for instance, the graph with no arc does not belong to the state space anymore. The minimal number of arcs is now $N - 1$. Each of these connected digraphs with $N - 1$ arcs can actually be built by constructing a (non directed) tree and by assigning appropriate orientations to its edges.

For sake of clarity, we denote by \mathcal{A}' the set of all connected acyclic digraphs and modify the transition rules as follows. Take uniformly at random (i, j) in V^2 .

(T'_1) If (i, j) is an arc in X_t then

²When generating graphs to test drawing algorithms, for instance, it makes no sense to generate graphs with too many edges. Graphs having $4N$ edges are actually considered as *dense* in the Graph Drawing community.

³Recall that a directed graph G is *simply connected* if its underlying non directed graph is connected. That is, if for any two vertices u, v there exists a sequence of vertices $u = v_0, v_1, \dots, v_p = v$ such that either (v_i, v_{i+1}) or (v_{i+1}, v_i) is an arc in G ($i = 0 \dots p - 1$).

- (i) if it is not disconnecting, it is deleted from X_t . That is, $X_{t+1} = X_t \setminus (i, j)$.
 - (ii) if it is disconnecting, it is reversed. That is, $X_{t+1} = X_t \setminus (i, j) \cup (j, i)$.
- (T_2) If (i, j) is not an arc in X_t , then
- (i) it is added to X_t , provided that the resulting graph is acyclic. That is, $X_{t+1} = X_t \cup (i, j)$.
 - (ii) otherwise, nothing is done. That is, $X_{t+1} = X_t$.

Write M' for this Markov chain running under the rules (T'_1) and (T_2). Since the transition matrix is symmetric, the problem we face is to show that the state space \mathcal{A}' underlying M' is irreducible. Observe that, contrarily to what was done in [7], this is not straightforward. Indeed, it is not immediate to see that given two connected acyclic digraphs G_1 and G_2 there exists a sequence of transitions from G_1 to G_2 and going through connected acyclic digraphs only. The proof is given in the next section.

In [7] the authors mention an algorithm by Alon and Rodeh [1] to test for circuits in directed graphs which can be used for conditions (T_2). Any algorithm testing a directed graph for simple connectedness can be used for condition (T'_1). This can be performed in a time proportional to the number of arcs in the graph (using a depth first search for instance, see, e.g., the textbook by Jungnickel [4]).

The complexity of the algorithm is dominated by the total number of iterations. The problem of estimating a good upper bound of this number is a difficult one and is at the centre of the theory on Markov chains [11]. As we will see in the next section, the transition graph is compact, with diameter in the order of N^2 .

4 Irreducibility of the Markov chain

We shall now give the complete proof of the irreducibility of the Markov chain M' . It is worthwhile to point out that the result holds, with the same proof, if \mathcal{A}' is replaced by any subclass containing all digraphs on V with $N - 1$ and N arcs.

Theorem 1 *Let M' be the Markov chain defined over the space \mathcal{A}' of all connected acyclic digraphs, together with the transitions defined by the rules (T'_1) and (T_2). The chain M' is irreducible, that is, given two connected acyclic digraphs G and H , there exists in M' a sequence of transitions $G = G_0 \rightarrow G_1, G_1 \rightarrow G_2, \dots, G_{p-1} \rightarrow G_p = H$ ($p \geq 1$) with all $G_i \in \mathcal{A}'$ ($i = 0 \dots p$). Moreover, such a sequence exists with length at most $(N + 7)(N - 3/2)$.*

Proof of Theorem 1

Since transitions in M' are reversible, it suffices to show that there exists a sequence of transitions from any given connected acyclic digraph G to a fixed element in \mathcal{A}' . We proceed in three steps. Let H be any acyclic digraph over V , denote by \bar{H} the undirected graph obtained from H by ignoring the orientations of its arcs.

Step 1 There exists in M' a sequence of transitions from G to a directed tree T (i.e., an acyclic digraph such that \bar{T} contains no cycle). Indeed, \bar{T} can be obtained by computing a spanning tree for \bar{G} . The transition is then defined by a sequence of arc deletions by rule (T'_1), leaving only the arcs present in the spanning tree.

Step 2 There exists in M' a sequence of transitions from T to a directed chain C (i.e., a directed tree over V such that \bar{C} is a chain; for short, a *dichain*). We proceed by diminishing the number $\ell(T)$ of the leaves in \bar{T} . Suppose $\ell(T) > 2$. Starting from a leaf u , one can find in \bar{T} a simple path $u = u_1, u_2, \dots, u_p = w$ such that u_i has degree 2 for all $i = 2, \dots, p - 1$ and w has degree at least 3. Let $v \neq u_2$ be

another neighbour of w (see the left-hand part of Figure 1, where the dashed lines stand for one or more edges and the dotted ones for possible edges). By rule (T_2) ,



Figure 1: Turning the leaf u into a vertex of degree 2

either the arc (v, u) or (u, v) can be added to T without creating a circuit, after what the arc between w and v cannot be disconnecting. Deleting it by rule (T'_1) , we get a directed tree T' such that $\ell(T') = \ell(T) - 1$, because w has still degree at least 2 and u is not a leaf anymore. The result follows by iterating until $\ell(T) = 2$.

Step 3 There exists in M' a sequence of transitions from C to any Hamiltonian directed chain, say $\{(i, i+1), 1 \leq i < N\}$. By rule (T'_1) indeed, this holds for $N = 2$ trivially. Set $N \geq 3$. By rule (T'_1) , it suffices to show that there is in M' a sequence of transitions from C to one of the dichains associated with the chain $\{(i, i+1), 1 \leq i < N\}$. Let $\bar{C} = \{\{v_i, v_{i+1}\}, 1 \leq i < N\}$, we first bring C into a dichain C' such that $\bar{C}' = \{\{v'_i, v'_{i+1}\}, 1 \leq i < N\}$ and $v'_1 = 1$. If there is $j > 1$ such that $v_j = 1$, an arc between v_N and v_1 may be added to C by rule (T_2) , and the arc in C relative to the edge $\{v_{j-1}, v_j\}$ may then be deleted by rule (T'_1) (see Figure 2, where orientations of the arcs have been omitted for sake of simplicity, and vertices joined by dotted paths are possibly identical). Now $v'_1 = 1$ in C' , let k be the least

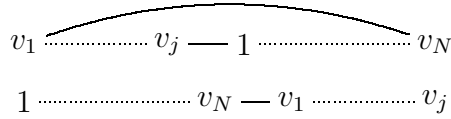


Figure 2: Turning the vertex 1 into a leaf

integer such that $v'_k \neq k$. Suppose $k < N$, then the (simple) path in \bar{C}' from 1 to k is longer than the one from 1 to v'_k . We next bring C' to a dichain C'' such that $\bar{C}'' = \{\{v''_i, v''_{i+1}\}, 1 \leq i < N\}$ with $v''_i = i$ for $1 \leq i \leq k$. Since its justification is quite similar to the previous one, we only describe the construction in Figure 3. Iterating it until $k = N$ completes the proof of the third step.

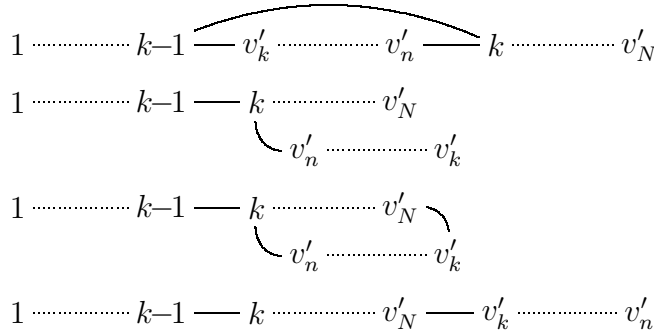


Figure 3: Placing the vertex k (orientations are omitted).

Finally, let G_1 and G_2 be in \mathcal{A}' . Then G_i ($i = 1, 2$) may be turned into a directed tree T_i by $N(N-1)/2 - (N-1)$ transitions at most, the worst case obtaining when

G_i is a tournament. Then T_i may be turned into a dichain C_i by $2((N - 1) - 2)$ transitions at most, the worst case obtaining when T_i is a star. Finally, C_1 may be made a Hamiltonian dichain by $(N - 1)/2$ arc reversions at most, and C_2 may be turned into the latter Hamiltonian dichain by $2 + 4(N - 2)$ transitions at most.

Remark For $N \geq 3$, the reversion rule $T'_1(ii)$ is not indispensable for Theorem 1 to hold. It is not clear whether it fastens convergence of the process or not.

Acknowledgements We wish to thank our colleague Alain Jean-Marie for stimulating discussions. This work originally started from email exchanges with Fabio G. Cozman.⁴

References

- [1] I. Alon and M. Rodeh. Finding a minimal circuit in a graph. *SIAM Journal of Computing*, 7(4):413–423, 1978.
- [2] E. A. Bender, L. B. Richmond, R. W. Robinson, and N. C. Wormald. The asymptotic number of acyclic digraphs. *Combinatorica*, 6(1):15–22, 1986.
- [3] Ira M. Gessel. Counting acyclic digraphs by sources and sinks. *Discrete Mathematics*, 160:253–258, 1996.
- [4] D. Jungnickel. *Graphs, Networks and Algorithms*. Springer Verlag, 1999.
- [5] Thorsten Koch. Steinlib testsets, 2001 (see the URL elib.zib.de/steinlib/).
- [6] Brendan McKay. On the shape of a random acyclic digraph. *Mathematical Proceedings of the Cambridge Philosophical Society*, 106:459–465, 1989.
- [7] G. Melançon, I. Dutour, and M. Bousquet-Mélou. Random generation of directed acyclic graphs. In Jaroslav Nešetřil, Marc Noy, and Oriol Serra, editors, *Comb01, Euroconference on Combinatorics, Graph Theory and Applications*, volume 10 of *Electronic Notes in Discrete Mathematics*, 2001.
- [8] Maurizio Patrignani. Graph drawing test suite, (see the URL www.dia.uniroma3.it/~patrigna/).
- [9] R. W. Robinson. Counting labeled acyclic digraphs. In F. Harary, editor, *New Directions in the Theory of Graphs*, pages 28–43, Ann Arbor, Michigan, 1973. Academic Press.
- [10] R. W. Robinson. Counting unlabeled acyclic digraphs. In C. H. C. Little, editor, *Combinatorial Mathematics V*, volume 622 of *Lecture Notes in Mathematics*, pages 239–273, Australia, 1977. Springer-Verlag.
- [11] A. Sinclair. *Algorithms for Random Generation & Counting: A Markov Chain Approach*. Progress in Theoretical Computer Science. Birkhauser, Boston, 1993.

⁴<http://www-2.cs.cmu.edu/~fgcozman/>