Sampling Random Chordal Graphs by MCMC (Student Abstract)

Wenbo Sun, Ivona Bezáková
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY, USA, 14623
{ws3109,ixbvcs}@rit.edu

Abstract

Chordal graphs are a widely studied graph class, with applications in several areas of computer science, including structural learning of Bayesian networks. Many problems that are hard on general graphs become solvable on chordal graphs. The random generation of instances of chordal graphs for testing these algorithms is often required. Nevertheless, there are only few known algorithms that generate random chordal graphs, and, as far as we know, none of them generate chordal graphs uniformly at random (where each chordal graph appears with equal probability). In this paper we propose a Markov chain Monte Carlo (MCMC) method to sample connected chordal graphs uniformly at random. Additionally, we propose a Markov chain that generates connected chordal graphs with a bounded treewidth uniformly at random. Bounding the treewidth parameter (which bounds the largest clique) has direct implications on the running time of various algorithms on chordal graphs. For each of the proposed Markov chains we prove that they are ergodic and therefore converge to the uniform distribution. Finally, as initial evidence that the Markov chains have the potential to mix rapidly, we prove that the chain on graphs with bounded treewidth mixes rapidly for trees (chordal graphs with treewidth bound of one).

Introduction

Chordal graphs arise in practical applications from a wide variety of fields, such as database management, computer vision, and Bayesian networks. Counting Markov equivalent DAGs in chordal graphs plays an important role in structural learning of Bayesian networks (Ghassami et al. 2019; Talvitie and Koivisto 2019). Generating chordal graphs uniformly at random is necessary for testing the performance of those algorithms. To the best of our knowledge, existing algorithms generating chordal graphs sample chordal graph instances far from the uniform distribution (Markenzon, Vernet, and Araujo 2008). This means that testing an algorithm's performance on such samples might give a false sense of security, since it is performed only on graph instances that are likely to be generated. In this paper, we address this weakness by proposing techniques that generate connected chordal graphs uniformly at random.

Chordal graphs are a very important concept in graphical models, and the running times of many learning and inference algorithms depend directly on the treewidth. Hence, instead of sampling all chordal graphs, one might want to sample only those with bounded treewidth. Therefore, we also focus on generating, uniformly at random, chordal graphs with a fixed treewidth bound. Previously known algorithms can be modified to generate chordal graphs with node and edge bounds, but not a treewidth bound. Of independent interest is also the relation of this problem to the well-studied problems of sampling spanning trees of a given graph (Broder 1989) and of sampling triangulations of a convex polygon (McShine and Tetali 1999) — these problems sample special chordal graphs of treewidth one and two, respectively, but use different Markov chains.

Brief Preliminaries

A graph is said to be *chordal* if each of its cycles of length four and greater has a chord, i.e., all induced cycles are of length three. Let Ω_n be the set of connected chordal graphs with n vertices. A vertex v is called *simplicial* if all its neighbors form a clique. The removal of a simplicial vertex will result in a graph in Ω_{n-1}. Our MCMC method aims to construct a Markov chain whose stationary distribution $\pi(G)$ is $|\Omega_n|^{-1}$ for all $G \in \Omega_n$. Achieving a desired distribution is typically done by the Metropolis-Hasting algorithm; in our case the distribution is uniform and, therefore, it suffices if our Markov chain is symmetric. The tricky part is to ensure that the Markov chain connects the state space, that is, that one can transform any state of the Markov chain into any other state using the steps of the chain. The mixing time of a Markov chain is defined as the number of steps the chain needs to take to get "close" to its stationary distribution.

Results

Efficient algorithms exist for deciding whether a graph is chordal. Therefore, a Markov chain for all chordal graphs can work as follow: Choose a random pair of vertices. If there is an edge between them, remove it if it keeps the graph chordal. If there is not an edge between them, add the edge

to the graph, if it is still chordal. This Markov chain connects the state space since one can use the theory of chordal graphs to find edges to remove, one by one, to get to a tree, and then design a mechanism how to convert one tree into another tree. This procedure is reminiscent to what we describe below for chordal graphs of bounded treewidth, hence we focus on the bounded treewidth case in this paper.

From now on, we will assume that we have a constant r bounding the treewidth, and, slightly abusing the notation, we denote by Ω_n all connected chordal graphs with treewidth bounded by r. We aim to design a Markov chain that samples graphs from Ω_n uniformly at random.

Algorithm 1 One step transition of the Markov chain

```
Pick a vertex u uniformly at random (u.a.r.).
if u is simplicial then
    Pick a set W of r vertices other than u u.a.r.
    Choose a random nonempty subset W' of W.
    if W' forms a clique then
        Deattach u from its current neighbors.
        Attach u to W'.
```

Every step of this chain happens with the same probability, hence the Markov chain is symmetric. It can convert the current chordal graph into a tree by always choosing W' of size 1. Hence, to prove the connectivity of the chain, it remains to show how it can convert one tree into another tree.

Therefore, we focus on $r = 1$ (sampling random (labeled) trees with n vertices uniformly at random), where we can not only formally prove the connectivity, but we can also give a polynomial mixing time bound.

Along the lines of the decomposition technique for bounding mixing times, we will decompose our state space into sets Ω_n^k containing chordal graphs on n vertices with exactly k simplicial vertices (and treewidth bound r).

We show that the above Markov chain, when restricted to only states in Ω_n^k (that is, any state not in Ω_n^k will be rejected), connects the state space and mixes rapidly, as long as $k \neq 2$ (for $k = 2$ every such chordal graph is a path) and $k \neq n - 1$ (star graphs).

We will use the following notation. For $G \in \Omega_n^k$, let S^G be the set of its simplicial vertices and let $S_0^G \subseteq S^G$ be the simplicial vertices whose neighbor is of degree 2. Whenever clear from the context, we will omit the superscript G.

Lemma 1. *The Markov chain over Ω_n^k for $3 \leq k \leq n - 2$ connects the state space, which has diameter $O(n^2)$.*

Proof. (Sketch.) Let the vertex set be $\{v_1, \ldots, v_n\}$. We will show that the chain can convert every graph in $G \in \Omega_n^k$ into the *canonical* graph, which has v_1 as its "root" and contains edges (v_1, v_j) for $2 \leq j \leq k + 1$ and (v_{j+1}, v_j) for $k + 1 \leq j \leq n$. This graph is a special case of a *comet* graph, which consists of a star graph with k vertices, and a path of $n - k$ vertices from the center of the star.

We will first show that any graph in Ω_n^k can be transformed, using steps of the Markov chain, to a comet graph with v_1 at the center of the star. We will first achieve that $v_1 \notin S$. If $v_1 \in S$, and there is a $u \in S_0$, we just reconnect u to v_1 (notice that the number of simplicial vertices did not change, since u was in S_0). If there is no such u, then suppose we root the tree at v_1 and u be the leaf furthest from v_1. This leaf has siblings (since it is not in S_0), we reconnect each of the siblings to an arbitrary vertex in $G - S$. Now we have $u \in S_0$ and we proceed as before.

Let t be a simplicial vertex. We will process the current simplicial vertices one by one. For a simplicial vertex $u \neq t$, if $u \in S_0$, reconnect u to t and let $t := u$; otherwise reconnect u to v_1. Both changes will keep the number of simplicial vertices unchanged. The final step is to transform the current "comet" graph into the canonical form. We omit the details due to the space constraints. □

Lemma 2. *The mixing time of the Markov chain on Ω_n^k for $3 \leq k \leq n - 2$ is $O(kn \log n)$.*

Proof. (Sketch.) We use the path coupling technique (Bubley and Dyer 1997). We say that two trees $X_t, Y_t \in \Omega_n^k$ are adjacent if they differ by exactly one Markov chain step. Let u be that different vertex, v_x be the parent of u in X_t, v_y be the parent of u in Y_t. The coupling chooses the same pair of vertices u', w' in both X_t and Y_t: u' will be deattached from its current neighbor and attached to w'.

There are two possible cases. The first is $u \in S_0$, then the removal of u will make v_x, v_y simplicial. So, in this case the good steps (when the distance between X_{t+1} and Y_{t+1} decreases) are when X_t removes (u, v_x), Y_t removes (u, v_y), and both add (u, w) and $w \in S$. Another possibility is $u \in S - S_0$, in which case the good steps are when X_t removes (u, v_x), Y_t removes (u, v_y), and both add (u, w) and $w \notin S$. It can be shown by induction that the number of good steps is strictly larger than the number of bad steps. □

Acknowledgments: This work was supported by the National Science Foundation, award DUE-1819546.

References

Broder, A. Z. 1989. Generating random spanning trees. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science (FOCS)*, 442–447.

Bubley, R., and Dyer, M. 1997. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS)*, 223–231.

Ghassami, A.; Salehkaleybar, S.; Kiyavash, N.; and Zhang, K. 2019. Counting and sampling from Markov equivalent DAGs using clique trees. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3664–3671.

Markenzon, L.; Vernet, O.; and Araujo, L. H. 2008. Two methods for the generation of chordal graphs. *Annals of Operations Research* 157(1):47–60.

McShine, L., and Tetali, P. 1999. On the mixing time of the triangulation walk and other Catalan structures. *Randomization methods in algorithm design* 43:147–160.

Talvitie, T., and Koivisto, M. 2019. Counting and sampling Markov equivalent directed acyclic graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 7984–7991.