Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

Electronic Notes in Discrete Mathematics 53 (2016) 43–55

www.elsevier.com/locate/endm

Uniform Sampling of Directed and Undirected Graphs Conditional on Vertex Connectivity

Salem A. Alyami [a,b,1], A. K. M. Azad [b] and Jonathan M. Keith [b]

[a] *Department of Mathematics and Statistics*
Al Imam Mohammad Ibn Saud Islamic University (IMSIU)
Riyadh, Saudi Arabia

[b] *School of Mathematical Sciences*
Monash University
Melbourne, Australia

Abstract

Many applications in graph analysis require a space of graphs or networks to be sampled uniformly at random. For example, one may need to efficiently draw a small representative sample of graphs from a particular large target space. We assume that a uniform distribution $f(N, E) = 1/|\mathcal{X}|$ has been defined, where N is a set of nodes, E is a set of edges, (N, E) is a graph in the target space $\mathcal{X}$ and $|\mathcal{X}|$ is the (finite) total number of graphs in the target space. We propose a new approach to sample graphs at random from such a distribution. The new approach uses a Markov chain Monte Carlo method called the Neighbourhood Sampler. We validate the new sampling technique by simulating from feasible spaces of directed or undirected graphs, and compare its computational efficiency with the conventional Metropolis-Hastings Sampler. The simulation results indicate efficient uniform sampling of the target spaces, and more rapid rate of convergence than Metropolis-Hastings Sampler.

Keywords: Sampling graph space, Markov chain Monte Carlo, Bayesian networks.

[1] Supported by Monash University, Australia. Email: salem.alyami@monash.edu

http://dx.doi.org/10.1016/j.endm.2016.05.005

1 Introduction

Random graphs (RGs) are the subject of a broad area of research originating in the early works of [7] and [4], drawing insights from both graph theory and probability. One particular area of interest is the generation of random graphs according to some criteria. Two commonly used approaches to generate random graphs are by using a probability distribution, or by using a random process [1]. An example of using probability distributions is the Erdös-Reńyi model [4], which generates random graphs with a uniform probability distribution over the space of all graphs with a given number of nodes and edges. The Erdös-Reńyi model can be implemented by starting with a given number of nodes with no edges and then iteratively adding one new edge at a time, sampled uniformly over the set of missing edges, until the required number of edges is obtained [1]. An example of using random processes is the simple algorithm used by [12] to simulate Bayesian graphs (directed acyclic graphs) in which the number of nodes and their average degree are taken as an input. The algorithm then computes a threshold value t, and for each pair of nodes, a random number $r \in [0, 1]$ is generated. If $r \leqslant t$, then the pair is linked by an edge. In this paper, we present an efficient sampler that is capable of sampling graphs that are uniformly distributed over certain spaces. Specifically, we consider spaces consisting of directed or undirected graphs in which all vertices are connected. Such graphs arise in a variety of applications, in particular in the study of Bayesian networks. A graph $\mathcal{G}$ is expressed as a pair (N, E), where N is a set of nodes and E is a set of edges. In the context of Bayesian networks, nodes represent random variables $V_1, V_2, \ldots, V_N$. A directed edge between any pair of variables $V_i \to V_j$ indicates conditional dependency of V_j upon V_i. Typically, edges represent cause and effect relationships and thus such graphs are acyclic. This paper considers two types of graph structures: a Connected Directed Acyclic Graph (CDAG) and a Connected Un-Directed Graph (CUDG). Figure 1 illustrates some types of graph structures.

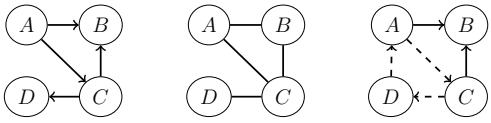

Fig. 1. From left to right: CDAG, CUDG and directed cyclic graph. Dashed lines indicate the cycle in the rightmost graph.

In CDAGs, as the number of nodes increases, the size of the space of possible graphs grows exponentially. For example, the sizes of spaces of CDAGs for three, four and five nodes are 18, 446 and 26430, respectively. Like CDAGs,