European Journal of Operational Research 245 (2015) 402–414

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

Mathematical programming approaches for classes of random network problems

Jordi Castro[a,*], Stefano Nasini[b]

[a] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Jordi Girona 1–3, 08034 Barcelona, Spain
[b] IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 19 March 2014
Accepted 12 March 2015
Available online 16 March 2015

Keywords:
Interior-point methods
Integer programming
Complex networks
Total unimodularity
Central path

ABSTRACT

Random simulations from complicated combinatorial sets are often needed in many classes of stochastic problems. This is particularly true in the analysis of complex networks, where researchers are usually interested in assessing whether an observed network feature is expected to be found within families of networks under some hypothesis (named conditional random networks, i.e., networks satisfying some linear constraints). This work presents procedures to generate networks with specified structural properties which rely on the solution of classes of integer optimization problems. We show that, for many of them, the constraints matrices are totally unimodular, allowing the efficient generation of conditional random networks by specialized interior-point methods. The computational results suggest that the proposed methods can represent a general framework for the efficient generation of random networks even beyond the models analyzed in this paper. This work also opens the possibility for other applications of mathematical programming in the analysis of complex networks.

1. Introduction

The use of random simulation is quite common when statistically studying properties of highly combinatorial sets. In many of those cases, closed-form expressions are hard to be found and the availability of *efficient* and *correct* simulation procedures might be of remarkable importance.

This is particularly true in the analysis of complex networks, an interdisciplinary field which brings together tools and methods from discrete mathematics and computer science with a great concern toward empirical applications, among others, in business, marketing, epidemiology, engineering, etc. Researchers are often interested in assessing the hypothesis of whether a particular network property is likely to appear under a uniform distribution of all networks verifying given constraints, named *conditional random networks* (Bollobas, 1985). In the absence of closed-form expressions (as it is often the case for most of random network models), large random samples of networks satisfying particular properties are required to test these hypotheses. This work introduces novel procedures to generate this sample, based on linear and integer optimization. They result in a general approach for random network simulation, which outperforms in versatility some currently available methods.

Following the standard notation (Ahuja, Magnanti, & Orlin, 1991), a graph $G = (V, E)$ is defined by a finite set V of n nodes, and a set of m pairs of them $E \subseteq V \times V$, named edges or arcs. A graph can be represented by a $n \times n$ binary matrix X, called *adjacency matrix* (AM from now on), whose (i, j)-entry, x_{ij}, is equal to 1 if there is a link between nodes i and j, and 0 otherwise. We will assume the graph has no loops, so that the diagonal of X is null. A network is a graph whose arcs or nodes have associated numerical values (arc costs, arc capacities, node supplies, etc.). In this work we will make no distinction and the two terms "graph" and "network" will be used as synonyms.

The study of random graphs begins with the seminal work of Erdös and Rainyi (1959), who considered a fixed set of nodes and an independent and equal probability of observing edges among them. There are two closely related variants of the Erdös–Rainyi model:

- *the $G(n, p)$ model*, where a network is constructed by connecting nodes randomly with independent probability p;
- *the $G(n, m)$ model*, where a network is chosen uniformly at random from the collection of all graphs with n nodes and m edges.

Both models possess the considerable advantage of being exactly solvable for many of their average properties: clustering coefficient, average path length, giant component, etc. (For more details about network properties, see Bollobas (1985), and Wasserman and Faust (1994).) In other words, the expectation of many structural properties

* Corresponding author. Tel.: +34 934015854.
E-mail addresses: jordi.castro@upc.edu (J. Castro), SNasini@iese.edu (S. Nasini).

http://dx.doi.org/10.1016/j.ejor.2015.03.021
0377-2217/

of networks generated by the Erdös–Rainyi processes is analytically obtainable. Conditional uniform models can be seen as a generalization of the $G(n, m)$ model, when the conditioning information is not necessarily the number of edges but whatever other arbitrary network property. Unfortunately, in this case we have very few analytical results and simulation is required to obtain empirical distributions of their average properties.

Although other Operations Research tools have been used in the context of social networks (Berghammer, Rusinowska, & de Swart, 2010; Gómez, Figueira, & Eusébio, 2013), as far as we know, this work is the first attempt to use linear and integer optimization for the generation of several classes of conditional random graphs. Previous approaches, developed within the fields of mathematical and computational sociology, were ad hoc procedures for some particular types of networks, in general difficult to generalize and not very efficient. For instance, the distribution of all networks conditioned to the nodes in- and out-degree has difficult combinatorial properties, as its analytical study involves binary matrices with fixed marginal rows and columns. In this respect, some combinatorial results have been obtained by Ryser (1957), who derived necessary and sufficient conditions for two vectors of non-negative integers to constitute the row sums and column sums of some zero-one matrix. On the other hand, ways to generate uniform random networks with given degree distribution were developed in Snijders (1991), Rao, Jana, and Bandyopadhyay (1996), Charon, Germa, and Hudry (1996), Roberts (2000) and Verhelst (2008), although they were computationally expensive and prohibitive for very large AMs.

In practice one would like to go even further in conditioning, which however leads to self-defeating attempts because of combinatorial complexity. This work provides a general methodological framework to generate networks with constraints, representing structural features we wish to control for.

Let x_{ij} be entries of the AM of either a directed or undirected graph with no loops or multiples edges. The AM is an element of the set of binary matrices

$$\chi = \{x_{ij} \in \{0, 1\}, (i, j) \in H\},$$

$$\begin{aligned} \text{where } & H = \{(i, j) : 1 \le i \le n-1, i < j \le n\} \text{ for undirected graphs} \\ \text{or } & H = \{(i, j) : 1 \le i \le n, 1 \le j \le n, i \ne j\} \text{ for directed graphs.} \end{aligned} \tag{1}$$

The continuous relaxation of χ, name it $CR(\chi)$, is obtained by replacing $x_{ij} \in \{0, 1\}$ by $x_{ij} \in [0, 1]$, in (1). Clearly, all extreme points of $CR(\chi)$ are integer. If we consider a conditional graph by adding extra linear constraints to χ, then $CR(\chi)$ may contain fractional extreme points, unless its constraints matrix is totally unimodular (TU, from now on). As shown in Heller and Tompkins (1956), the next theorem provides sufficient conditions for a matrix to be TU:

Theorem 1. *Let $A \in \{-1, 0, 1\}^{m \times n}$ be a matrix obtained by elementary operations of $B \in \mathbf{Z}^{m \times n}$ and consider a partition of the rows of A in two disjoint sets $\mathcal{J}_1$ and $\mathcal{J}_2$. The following three conditions together are sufficient forB to be TU:*

1. *Every column of A contains at most two non-zero entries, which are either 1 or −1.*
2. *If two non-zero entries in a column of A have the same sign, then the row of one is in $\mathcal{J}_1$, and the other in $\mathcal{J}_2$.*
3. *If two non-zero entries in a column of A have opposite signs, then the rows of both are either in $\mathcal{J}_1$ or $\mathcal{J}_2$.*

The above theorem will be extensively used in next section. More details on unimodularity in integer programming can be found in Schrijver (1998). If the constraints matrix of $CR(\chi)$ is TU, each extreme point of $CR(\chi)$ represents a graph. Therefore, it is possible to generate a bunch of graphs by merely solving linear programs (LP) with random gradients in the objective function, or by non-degenerated simplex pivoting, starting from a given initial extreme point (Padberg, 1999). Moreover, they can be generated in polynomial time if interior-point methods are used (Wright, 1996).

The paper is organized as follows. Section 2 is devoted to the characterization of the convex hull of polytopes associated to some families of conditional random networks. We will differentiate between families whose constraints are TU, and those which may give rise to fractional AMs. Supported by these results, Section 3 presents two particular procedures for the generation of conditional random networks, and analyzes the probability distribution of the LP solutions. Section 4 illustrates these techniques using some real-world data sets.

Throughout the paper we denote the vector of variables associated to the components of the AM as either $\mathbf{x}^T = [x_{12}, \dots, x_{1n}, x_{23}, \dots, x_{(n-1)n}, x_{21}, \dots, x_{n(n-1)}]$ (i.e., the rowwise upper triangle of AM followed by its columnwise lower triangle) for directed graphs, or $\mathbf{x}^T = [x_{12}, \dots, x_{1n}, x_{23}, \dots, x_{(n-1)n}]$ (only the rowwise upper triangle of AM) for undirected graphs.

2. Total unimodularity of constraints from some conditional random networks

Let χ be the set of AMs of a family of either directed or undirected networks with n nodes, and let $CR(\chi)$ be its continuous relaxation. For about twenty families of networks the extreme points of $CR(\chi)$ can be seen to be integer. Although making an extensive list of all these families is out of the scope of this work, some of the most relevant ones will be discussed in the following sections.

Next Proposition 1, which provides a sufficient condition for the existence of a bijection between extreme points of $CR(\chi)$ and the set of feasible networks, will be useful to show that some constraints matrices are TU.

Proposition 1. *For a given family of either directed or undirected networks with n nodes, let $F \in \mathbb{R}^{l \times m}$, be a matrix of $l \le m$ linear constraints characterizing the family of networks under consideration, where $m = n(n-1)$or $m = n(n-1)/2$ for, respectively, directed and undirected networks. Let $CR(\chi) = \{\mathbf{x} \in [0, 1]^m : F\mathbf{x} = \mathbf{b}\}$be the continuous relaxation of the constraints. If $\mathbf{b}$ is integer and F can be reduced by elementary row operations to a matrix, call itF', with a unique unitary element (either +1 or −1) per column and all the elements of the same row with the same sign, then there is a bijection between the extreme points of $CR(\chi)$ and the set of networks under consideration.*

Proof. In standard form, the system of linear constraints associated to $CR(\chi)$ is

$$\left[\begin{array}{c|c} I & I \\ \hline F' & \end{array}\right] \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{s} \end{array}\right] = \left[\begin{array}{c} \mathbf{e} \\ \hline \mathbf{b} \end{array}\right], \qquad \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{s} \end{array}\right] \ge 0. \tag{2}$$

From Theorem 1, the constraints matrix of (2) is TU by considering the following partition of rows: set the first m rows (associated to the identities) in $\mathcal{J}_1$; if elements of row i of F' are negative, then set this row in $\mathcal{J}_1$; otherwise, if they are positive, set the row in $\mathcal{J}_2$. Therefore, all extreme points of $CR(\chi)$ are integer and they correspond to the AM of a network. In addition, no integer point may be located in the interior of $CR(\chi)$ since it is a subset of the unit hypercube, completing the proof. □

In some cases there is no bijective relation between a family of conditional random networks and the extreme points of its polyhedron, since some basic solutions may be fractional. However, if we can ensure that no integer solution is in the interior of the polyhedron, this injective relation (i.e., any random network is associated to an extreme point, but not the opposite) is still useful, whenever some kind of acceptance–rejection technique is considered for fractional solutions. This is the case, for instance, of networks conditioned