Computers & Operations Research 69 (2016) 68–78

Contents lists available at ScienceDirect

Computers & Operations Research

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

A heuristic based on negative chordless cycles for the maximum balanced induced subgraph problem

Fabrizio Marinelli, Angelo Parente

Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Via Brecce Bianche, I-60131 Ancona, Italy

ARTICLE INFO

Available online 21 December 2015

Keywords:
Network matrix
Signed graph
Independent set
Maximum balanced induced subgraph
Vertex frustration

ABSTRACT

A signed graph, i.e., an undirected graph whose edges have labels in $\{-1,+1\}$, is balanced if it has no negative cycles. Given a signed graph, we are interested in a balanced induced subgraph of maximum order (the MBIS problem). In the present work, we propose a greedy approach for the MBIS problem that is based on the progressive shortening of negative cycles, and that generalizes the well-known minimum-degree greedy heuristic for the maximum independent set problem. An extensive computational study on three classes of instances shows that the new algorithm outperforms the reference heuristics proposed in the literature.

1. Introduction

A signed graph $S=(G,\sigma)$ is an undirected graph $G=(V,E)$ whose edges are labeled either with $+1$ (*positive* edges) or with -1 (*negative* edges) by a *signature* function $\sigma: E \to \{-1,+1\}$. See [40,41] for a comprehensive survey. The sign of a subset C of edges, e.g., a path or a cycle, is given by the product of the signs of the edges in C. From a theoretical perspective, signed graphs and their incidence matrices are deeply connected with matroid theory [17,35].

Following the definition by Harary [22], a signed graph is said to be balanced if it has no negative cycles. Harary also proved that a signed graph is balanced if and only if the set of negative edges is a possibly empty cut. Given a signed graph, the Maximum Balanced Induced Subgraph problem (MBIS) asks for finding a balanced induced subgraph (BIS) of maximum order. The opposite numerical problem, i.e., finding the smallest number of vertices whose deletion makes the signed graph balanced, was introduced by Harary [23] and called *point index for structural balance*. Nowadays that number is known as the *vertex frustration number*; see [42].

Several problems originating in different domains, even far away from each other, can be modeled as MBIS. In social networking, for example, relationships between individuals can be represented by a signed graph where vertices are persons and positive (negative) edges express friendships (hostility). Based on the structural balance theory [10], positive cycles are supposed to indicate stable social situations, whereas negative cycles are supposed to be unstable. Hence the maximum balanced subgraph gives a measure of the cohesion of the social group.

Another application lies in the context of portfolio analysis [24]. In that case, the vertices of a signed graph denote stocks, and a positive (negative) edge represents the direct (inverse) correlation between its extremes. It is generally believed that the larger is the maximum balanced induced subgraph, the more predictable is the behavior of the portfolio.

MBIS is polynomial on series parallel graphs [4] but in general is NP-hard as it admits the odd cycle transversal (also known as maximum bipartite subgraph or vertex bipartization) and the maximum independent set (MIS) problems as special cases, the former with applications, e.g., in VLSI design [2,12], DNA sequencing [16] and computational biology [43], the latter arising as a subproblem or as a relaxation of many 0–1 integer problems (as a reference, think about the solution of the MIS problem on the so-called *conflict graph*, which is at the basis of many general preprocessing and probing techniques for integer linear programming problems [3,8]).

In particular, the odd cycle transversal problem on a graph G corresponds to the MBIS problem on the signed graph $S=(G,\sigma)$ obtained from G by signing all its edges as negative. In fact, odd cycles of G correspond to negative cycles of S, and a minimum odd cycle transversal on G corresponds to a smallest set of vertices to be removed from S to make it balanced. On the other hand, the MIS problem on a graph G corresponds to MBIS on the signed graph obtained from G by *signed expansion* [40, Section 7C], i.e., by doubling its edges and assigning opposite signs to each pair of parallel edges.

Our interest in the MBIS problem was originally motivated by the study of an equivalent problem known in the domain of integer

E-mail addresses: fabrizio.marinelli@univpm.it (F. Marinelli), angelo@inwind.it (A. Parente).

http://dx.doi.org/10.1016/j.cor.2015.12.004
0305-0548/

programming as Maximum Embedded Reflected Network (mern). Nowadays, the most successful methods for solving an integer linear program work by iteratively tightening (cutting planes) or recursively decomposing (branch-and-bound) a polyhedron which represents a continuous formulation of the problem. Such algorithms stop as soon as an integer extreme point is reached and its optimality is proven. When one of such methods is applied, the cases where the polyhedron is integral (or at least has an integer extreme point that is optimal for a given objective function) are of particular interest because the integer linear problem boils down to a continuous linear one. The integrality of the polyhedron depends on the structure of the coefficients of the corresponding integer linear program. A well-known family of such special structures is that of totally unimodular (TU) matrices: one of the famous theorems by Hoffman–Kruskal [27] states that the polyhedron $P = \{\mathbf{x} \in IR^n | \mathbf{b}' \le \mathbf{A}\mathbf{x} \le \mathbf{b}, \mathbf{c}' \le \mathbf{x} \le \mathbf{c}\}$, with $\mathbf{A}(m \times n)$, is integral for every integral $\mathbf{b}, \mathbf{b}' \in IR^m$ and $\mathbf{c}, \mathbf{c}' \in IR^n$ if and only if $\mathbf{A}$ is a TU matrix. Indeed, the recognition of special structures in the coefficient matrix of (integer) linear programs also helps in the solution of large-scale continuous models [6,7] and, on the other hand, can be exploited in the strategic choice of constraints to be either relaxed in a Lagrangian relaxation scheme [5] or convexified in a Dantzig–Wolfe decomposition.

Total unimodularity can be checked in polynomial time [36]; but most formulations of combinatorial optimization problems do not exhibit a TU matrix. We are therefore interested in finding a maximum embedded TU submatrix. A subclass of TU matrices is that of *reflected network* matrices; see Section 3. Since, for any given $\{0, \pm 1\}$-matrix $\mathbf{A}$, a signed graph S_A can be defined such that any reflected network submatrix of $\mathbf{A}$ obtained by row deletion corresponds to a balanced induced subgraph of S_A and vice versa [25,26], the mern problem, i.e., the task of finding a maximum reflected network submatrix by deleting rows, is, in combinatorial terms, the mbis problem. That explains our interest in mbis.

In this paper we propose a greedy heuristic for mbis that generalizes the minimum-degree greedy algorithm for mis. The heuristic is based on a graph transformation (the shortening of negative cycles) that preserves balance of any induced subgraph that is balanced in the original graph, and makes the mbis easier to solve. A broad computational experience shows that the heuristic largely outperforms the best previous algorithm for mbis.

The remainder of the paper is organized as follows. The graph terminology used throughout the paper and the main properties of signed graphs are introduced in Section 2. In Section 3 we describe the link between reflected networks and signed graphs and briefly survey the approaches for solving the mbis problem. In Section 4 we give the details of a heuristic for mbis and Section 5 illustrates our computational experience. Conclusions are sketched in Section 6.

2. Preliminaries

2.1. Graphs

Let $G=(V,E)$ be a finite undirected graph with vertex set $V=\{1,\dots,n\}$ and edge set E that is a set of unordered pairs of distinct vertices. The *density* of G is given by $2|E|/(|V|(|V|-1))$. The subgraph of G *induced* by the set of vertices $U \subseteq V$ is the graph $G[U]=(U,\{uv \in E | u,v \in U\})$. The subgraph of G *induced* by the set of edges $F \subseteq E$ is the graph $G[F]=(\{v | \exists\ uv \in F\}, F)$. The set $N(u)=\{v \in V | \exists uv \in E\}$ of vertices adjacent to $u \in V$ is the *neighborhood* of u. The cardinality of $N(u)$ is the *degree* $d(u)$ of u. A *path* is a non-empty graph $P=(U,F)$ of the form $U=\{u_0,u_1,\dots,u_k\}$ with $u_i \ne u_j$ for $0 \le i \ne j \le k$, and $F=\{u_0u_1,\dots,u_{k-1}u_k\}$. A *cycle* is a graph $C=(U,F)$ of the form $U=\{u_0,u_1,\dots,u_k\}$ with $u_i \ne u_j$ for $0 \le i \ne j \le k$, and $F=\{u_0u_1,\dots,u_{k-1}u_k,u_ku_0\}$. In the following, we denote by $u_0u_1\cdots u_k$ and $u_0u_1\cdots u_ku_0$ the path P and the cycle C, respectively. The length of a cycle $C=(U,F)$ is the number $|U|$ of its vertices (or edges). An edge which joins two vertices of a cycle but is not itself an edge of the cycle is a *chord*. A cycle without chords is said to be *chordless* and is usually called a *hole*. In the literature, chordless cycles commonly refer to cycles of length at least four; in this paper, we extend the definition also to cycles of length two (parallel pairs of edges) and three (triangles).

Given a bipartition $\{U,W\}$ of V, the (possibly empty) set $E(U,W)=\{uv \in E | u \in U, v \in W\}$ is called a *cut*. If $U=\{u\}$, the cut $E(U,V-u)$ is called the *star* of u and is denoted by $E(u)$. A *spanning tree* of $G=(V,E)$ is a connected subgraph $T=(V,E_T)$ such that T is acyclic. A set I of vertices is *independent* (or *stable*) if no two elements in I are adjacent. The *stability number* $\alpha(G)$ is the cardinality of the largest independent set of G.

In the following, we denote with $\mathcal{C}(G)$ the set of cycles of G and with $\mathcal{H}(G)$ the set of chordless cycles of G. With a slight extension of notation, $\mathcal{C}(uv)$ and $\mathcal{C}(u)$ ($\mathcal{H}(uv)$ and $\mathcal{H}(u)$) denote the sets of (chordless) cycles passing through edge uv and vertex u, respectively. Definitions not reported in this section can be found in Diestel's book [14].

2.2. Signed graphs

Let $S=(G,\sigma)$ be a signed graph. $G=(V,E)$ is called the *underlying* graph of S. A negative edge uv and the set of negative edges of S are denoted by $(uv,-)$ and E^- respectively. Analogously, $(uv,+)$ and E^+ denote (the set of) positive edges. A signed graph without negative edges and a vertex without incident negative edges are called *all positive*. In particular, we denote with S^+ the underlying subgraph $(V, E \backslash E^-)$, i.e., the subgraph of S obtained by removing all the negative edges. Similar notation and definitions are used for negative vertices and negative graphs. In all the figures of the paper, negative (positive) edges are depicted by dashed (solid) lines.

The set of negative (chordless) cycles of S is denoted by $\mathcal{C}^-(S)$ (by $\mathcal{H}^-(S)$). The positive counterparts are referred accordingly.

In the following, we only refer to signed graphs with no *loops*, *half-edges* and *loose edges*. Besides, we slightly generalize the definition of the underlying graph G by allowing *parallel* edges (of opposite sign in S): for the sake of conciseness we say that uv is a parallel pair of edges if both $(uv,-)$ and $(uv,+)$ are in E, and we indicate with E^P the set of parallel edges of S. Notice that any parallel edge pair uv counts two in both $d(u)$ and $d(v)$ and is a negative chordless cycle of length two. It follows that any cycle longer than two and containing parallel pairs of edges is not chordless.

Among the properties of signed graphs we are interested in *balance* and *switching equivalence*. A signed graph S is balanced if it has one of the following equivalent properties:

- S has no negative cycles;
- E^- is a (possibly empty) cut [22];
- all paths joining each pair of distinct vertices u, v have the same sign [22];
- there exists a vertex *signature* function $\gamma : V \to \{-1,+1\}$ such that $\sigma(uv)=\gamma(u)\gamma(v)$ for each edge uv.

Switching the set $U \subseteq V$ consists in negating the signs of the edges in the set $E(U, V \backslash U)$. In particular, switching a vertex u reverses the signs of the edges in the star $E(u)$. We denote by S^U the signed graph obtained by switching U in S. We say S and S^U are *switching equivalent*. Switching equivalence is an equivalence relation on signed graphs [40]. In particular, switching any $U \subseteq V$