Multigraph realizations of degree sequences: Maximization is easy, minimization is hard

Heather Hulett, Todd G. Will, Gerhard J. Woeginger

PII: S0167-6377(08)00055-2
DOI: 10.1016/j.orl.2008.05.004
Reference: OPERES 5160

To appear in: *Operations Research Letters*

Received date: 3 March 2008
Accepted date: 31 May 2008

Please cite this article as: H. Hulett, T.G. Will, G.J. Woeginger, Multigraph realizations of degree sequences: Maximization is easy, minimization is hard, *Operations Research Letters* (2008), doi:10.1016/j.orl.2008.05.004

Multigraph realizations of degree sequences: Maximization is easy, minimization is hard

Heather Hulett [*] Todd G. Will [†] Gerhard J. Woeginger [‡]

Abstract

The following minimization problem is shown to be NP-hard: Given a graphic degree sequence, find a realization of this degree sequence as loopless multigraph that minimizes the number of edges in the underlying support graph. The corresponding maximization problem is known to be solvable in polynomial time.

Keywords: Computational complexity; combinatorial optimization; graph theory.

AMS subject classification: 05C12.

1 Introduction

A sequence $d = \langle d_1, \ldots, d_n \rangle$ of non-negative integers is called *graphic* if it is the degree sequence of some loopless multigraph G. Such a multigraph G then contains $\frac{1}{2}\sum_{k=1}^{n} d_k$ edges, and is called a *realization* of sequence d. For a multigraph $G = (V, E)$, the underlying *support graph* is a simple graph on the same vertex set V, that contains an edge between two vertices u and v in V, if and only if the multigraph contains at least one edge between u and v in E.

Degree sequences of *simple* graphs are well-understood. They have nice combinatorial characterizations (Hakimi [5]), and they can be recognized in polynomial time. There is a close connection between degree sequences of simple graphs and general graphic degree sequences (of loopless multigraphs), which is based on the following procedure for transforming a multigraph into a simple graph: "As long as there exist two vertices u and v with at least two parallel edges between them, subdivide one of these edges by creating a new vertex of degree 2."

Proposition 1 *(Owens & Trent [8])*
Let t be an integer, let $d = \langle d_1, \ldots, d_n \rangle$ be a sequence of non-negative integers, and let the sequence d' result from d by appending t copies of the integer 2. Then the following two statements are equivalent:

[*] hulett.heat@uwlax.edu. Department of Mathematics, University of Wisconsin-La Crosse, La Crosse, WI 54601, USA.

[†] will.todd@uwlax.edu. Department of Mathematics, University of Wisconsin-La Crosse, La Crosse, WI 54601, USA.

[‡] gwoegi@win.tue.nl. Department of Mathematics and Computer Science, TU Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

(i) The sequence d is the degree sequence of a loopless multigraph, whose support graph contains at least $\frac{1}{2}\sum_{k=1}^{n} d_k - t$ edges.

(ii) The sequence d' is the degree sequence of a simple graph. $\square$

Next, let us formulate two natural optimization problems around degree sequences and support graphs.

> Max-Realization:
> For a given graphic sequence d, find a loopless multigraph realization that *maximizes* the number of edges in the underlying support graph.

> Min-Realization:
> For a given graphic sequence d, find a loopless multigraph realization that *minimizes* the number of edges in the underlying support graph.

The maximization problem is solvable in polynomial time. This easy fact is an immediate consequence of Proposition 1. Kleitman [6] discusses other, perhaps faster algorithms for Max-Realization.

The minimization problem is more challenging. Will & Hulett [9] study the combinatorial structure of support graphs with the minimum number of edges, and they show that there are only two possible types of connected components for them: A connected component is either a tree, or a tree plus one edge where the unique cycle has odd length. In this note we will show that the minimization problem is NP-hard.

Theorem 2 *It is strongly NP-hard to decide for a given graphic sequence d and an integer bound B, whether there exists a loopless multigraph realization with at most B edges in the underlying support graph.*

Theorem 2 settles an open question of Will & Hulett [9]. It also adds another item to the list of optimization problems for which the minimization version and the maximization version behave very differently. Other items on this list are, for instance, cuts in graphs (min is easy, max is hard), paths in graphs (min is easy, max is hard), jumps and bumps in linear extensions of partial orders (min is hard, max is easy; see [4]), or the travelling salesman problem in the plane with the Manhattan metric (min is hard, max is easy; see [1]).

2 The hardness proof

This section contains the proof of Theorem 2. Our reduction is done from the partial Latin square completion problem. A *Latin square* of order p is a $p \times p$ matrix with entries from the color set $\{1, 2, \ldots, p\}$, such that each row contains each color exactly once, and each column contains each color exactly once. A *partial Latin square* is a $p \times p$ matrix where each entry is either empty or contains a color from $\{1, 2, \ldots, p\}$, such that each row (column) contains each color at most once. Colbourn [2] established NP-hardness of the following problem.

Partial Latin square completion (PLSC):

Instance: A partial $p \times p$ Latin square L.

Question: Can the empty entries in L be filled with colors from $\{1, 2, \ldots, p\}$, such that the resulting matrix is a Latin square?

Now let us describe the reduction. We consider an arbitrary partial Latin square L with m empty entries as an instance of PLSC, and we will construct an instance of Min-Realization from it. Let $q = 2p$. Then, for the kth row, the ℓth column, and for color c we do the following (where k, ℓ, c run through all values in $\{1, \ldots, p\}$).

- If color c does not occur in the kth row, then we put the so-called *x-number* $x(k, c) = 2q^6 + kq - c$ into the degree sequence d.
- If color c does not occur in the ℓth column, then we put the so-called *y-number* $y(\ell, c) = 7q^6 + \ell q^2 + c$ into the degree sequence d.
- If the entry $L(k, \ell)$ at the crossing of kth row and ℓth column is empty, then we put the so-called *z-number* $z(k, \ell) = 9q^6 + kq + \ell q^2$ into the degree sequence d.

The bound on the number of edges in the support graph is defined as $B = 2m$. This completes the construction of the instance d and B of Min-Realization.

The degree sequence d contains altogether $3m$ numbers, of which m are x-numbers, m are y-numbers, and m are z-numbers. The following two Lemmas 3 and 4 state crucial properties of our construction, and then Lemmas 5 and 6 establish the correctness of our reduction.

Lemma 3 *The $3m$ numbers in the degree sequence d are pairwise distinct.*

Proof. The x-numbers are from the range $2q^6$ to $2q^6 + q^2$, and the value $2q^6 + kq - c$ uniquely determines k and c. The y-numbers are from the range $7q^6$ to $7q^6 + q^3$, and the value $7q^6 + \ell q^2 + c$ uniquely determines ℓ and c. The z-numbers are from the range $9q^6$ to $9q^6 + q^4$, and the value $9q^6 + kq + \ell q^2$ uniquely determines k and ℓ. Since the three ranges are disjoint, one easily sees that the $3m$ numbers are pairwise distinct. $\square$

Lemma 4 *Assume that two entries d_i and d_j with $d_i < d_j$ in the degree sequence d add up to a third entry d_k. Then d_i must be an x-number $x(k, c)$, d_j must be a y-number $y(\ell, c)$, and d_k must be a z-number $z(k, \ell)$ for three appropriate values k, ℓ, c.*

Proof. Straightforward case distinctions on the size of the involved numbers imply that $d_i = x(k_1, c_1)$, $d_j = y(\ell_2, c_2)$, and $d_k = z(k_3, \ell_3)$ must hold for appropriate values of k_1, k_3, ℓ_2, ℓ_3, c_1, and c_2. This yields

$$(2q^6 + k_1 q - c_1) \ + \ (7q^6 + \ell_2 q^2 + c_2) \ = \ (9q^6 + k_3 q + \ell_3 q^2).$$

By considering this equation modulo q, we get $c_1 = c_2$. Then considering the equation modulo q^2 yields $k_1 = k_3$. Finally, $c_1 = c_2$ and $k_1 = k_3$ also imply $\ell_2 = \ell_3$. $\square$

Lemma 5 *If the instance L of PLSC can be completed to a Latin square, then the constructed instance of* Min-Realization *has answer YES.*

Proof. Whenever an empty entry $L(k,\ell)$ in the partial Latin square receives color c in the completed Latin square, we create three (new) corresponding vertices u, v, w in the multigraph together with $x(k,c)$ edges between u and v, and together with $y(\ell,c)$ edges between u and w. No other edges are incident to u, v, w. Then v is of degree $x(k,c)$, and w is of degree $y(\ell,c)$, and u is of degree $x(k,c) + y(\ell,c) = z(k,\ell)$.

In this fashion, for every empty entry in L we introduce three corresponding vertices and two corresponding edges in the underlying support graph. Altogether, this yields a support graph with $2m = B$ edges. $\square$

Lemma 6 *If the constructed instance of* Min-Realization *has answer YES, then the instance L of PLSC can be completed to a Latin square.*

Proof. Consider a loopless multigraph realization $G = (V,E)$ of the degree sequence d with at most $B = 2m$ edges in the underlying support graph $G' = (V,E')$. Since the sequence d does not contain any zero entries, a connected component in G cannot consist of a single vertex. Since by Lemma 3 the entries in d are pairwise distinct, a connected component in G cannot consist of exactly two vertices. Consequently every connected component in G and G' contains at least three vertices, and there are at most $m = |V|/3$ connected components. In any graph, the number of edges is greater or equal to the number of vertices minus the number of connected components. For G', this yields $|E'| \geq |V| - m = 2m = B$. This implies that $|E'| = 2m$, and that every connected component in G' is a path on three vertices.

Now consider a connected component in the multigraph G. Since the degree of the middle-vertex equals the sum of the degrees of the two outer vertices, Lemma 4 implies that these three degrees are $x(k,c)$, $y(\ell,c)$, and $z(k,\ell)$ for three appropriate values k, ℓ, c. We fill the empty entry $L(k,\ell)$ with color c, and repeat this for all other connected components.

Every empty entry in $L(k,\ell)$ is filled (since the corresponding z-number $z(k,\ell)$ is the degree of a middle-vertex in one of the components). In the kth row every missing color c shows up exactly once (since the corresponding x-number $x(k,c)$ is the degree of an outer vertex in one of the components). In the ℓth column every missing color c shows up exactly once (since the corresponding y-number $y(\ell,c)$ is the degree of an outer vertex in one of the components). $\square$

By Lemmas 5 and 6, our reduction is correct. Note furthermore that the numbers in the degree sequence d are polynomially bounded in p. This establishes the strong NP-hardness of Min-Realization, and completes the proof of Theorem 2.

3 Appendix: Three-partitioning with distinct integers

In this appendix, we extract two corollaries from the NP-hardness proof in the preceding section. We feel that these corollaries are of independent interest, and that they may prove useful in other lines of investigation; for instance Li [7] fixes a proof by previous authors who, as he points out, had prematurely assumed that NMTS with distinct integers (see below for a definition of this problem) was NP-hard.

The well-known book [3] by Garey & Johnson lists two NP-hard integer packing problems that since then have been used in literally hundreds of NP-hardness proofs: THREE-PARTITIONING and NUMERICAL MATCHING WITH TARGET SUMS.

> THREE-PARTITIONING:
>
> INSTANCE: A sequence $a_1, \ldots, a_{3n}$ of $3n$ positive integers; an integer B with $\sum_{k=1}^{3n} a_k = nB$.
>
> QUESTION: Is it possible to partition these $3n$ integers into n disjoint triples, such that in every triple the three elements add up to B?

> NUMERICAL MATCHING WITH TARGET SUMS (NMTS):
>
> INSTANCE: Three sequences $a_1, \ldots, a_n$, and $b_1, \ldots, b_n$, and $c_1, \ldots, c_n$ of positive integers, such that $\sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k = \sum_{k=1}^{n} c_k$.
>
> QUESTION: Is it possible to partition the $3n$ given numbers into n triples, such that each triple contains one a_i, one b_j, and one c_k, with $a_i + b_j = c_k$?

In the standard NP-hardness proof for THREE-PARTITIONING (as presented in [3]), the $3n$ numbers $a_1, \ldots, a_{3n}$ are not pairwise distinct. Quite to the contrary, the proof introduces repeated integers at many places, and this seems to be an inherent feature of this proof.

The standard NP-hardness arguments for NUMERICAL MATCHING WITH TARGET SUMS also introduce repeated integers. Yu, Hoogeveen & Lenstra [10] provide a very sophisticated proof that the special case of NMTS with $a_k \equiv b_k \equiv k$ is NP-hard. However, their construction yields numerous repeated integers among $c_1, \ldots, c_n$.

Corollary 7 *The special case of* THREE-PARTITIONING *where the* $3n$ *integers* $a_1, \ldots, a_{3n}$ *are all distinct is strongly NP-hard.*

Corollary 8 *The special case of* NUMERICAL MATCHING WITH TARGET SUMS *where the* $3n$ *integers* $a_1, \ldots, a_n$*,* $b_1, \ldots, b_n$*,* $c_1, \ldots, c_n$ *are all distinct is strongly NP-hard.*

The proofs of both corollaries follow the construction in the preceding section. For Corollary 7 we use all x-numbers, all y-numbers, and for every z-number $z(k, \ell)$ we use the number $B - z(k, \ell)$ with $B = 19q^6$. For Corollary 8 we simply use the x-numbers as $a_1, \ldots, a_n$, the y-numbers as $b_1, \ldots, b_n$, and the z-numbers as $c_1, \ldots, c_n$. Lemma 3 yields that these numbers are all distinct. Lemma 4 yields the correctness of these reductions from PLSC to THREE-PARTITIONING and NMTS, respectively.

Acknowledgement. This research has been supported by the Netherlands Organisation for Scientific Research (NWO), grant 639.033.403, and by BSIK grant 03018 (BRICKS: Basic Research in Informatics for Creating the Knowledge Society).

References

[1] A. Barvinok, S.P. Fekete, D.S. Johnson, A. Tamir, G.J. Woeginger, R. Woodroofe (2003). The geometric maximum traveling salesman problem. *Journal of the ACM 50*, 641–664.

[2] C.J. Colbourn (1984). The complexity of completing partial latin squares. *Discrete Applied Mathematics 8*, 25–30.

[3] M.R. Garey and D.S. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, San Francisco.

[4] M. Habib, R.H. Möhring, and G. Steiner (1988). Computing the bump number is easy. *Order 5*, 107–129.

[5] S. Hakimi (1962). On realizability of a set of integers as degrees of the vertices of a linear graph. *SIAM Journal on Applied Mathematics 10*, 496–506.

[6] D.J. Kleitman (1970). Minimal number of multiple edges in realization of an incidence sequence without loops. *SIAM Journal on Applied Mathematics 18*, 25–28.

[7] W.N. Li (1995). The complexity of segmented channel routing. *IEEE Trans. Computer-Aided Design 14*, 518–523.

[8] B. Owens, H.M. Trent (1967). On determining minimal singularities for the realization of an incidence sequence. *SIAM Journal on Applied Mathematics 15*, 406–418.

[9] T.G. Will and H. Hulett (2004). Parsimonious multigraphs. *SIAM Journal on Discrete Mathematics 18*, 241–245.

[10] W. Yu, H. Hoogeveen, J.K. Lenstra (2004). Minimizing makespan in a two-machine flow shop with delays and unit-time operations is NP-hard. *Journal of Scheduling 7*, 333–348.