Partitioning a graph into minimum gap components

Maurizio Bruglieri [1]

Dipartimento di Design, Politecnico di Milano, Milano, Italy

Roberto Cordone [2]

Dipartimento di Informatica, Università degli Studi di Milano, Milano, Italy

Abstract

We study the computational complexity and approximability for the problem of partitioning a vertex-weighted undirected graph into p connected subgraphs with minimum gap between the largest and the smallest vertex weights.

Keywords: Graph partitioning, computational complexity, approximability

1 Introduction

Let $G = (V, E)$ be an undirected connected graph, w_v an integer *weight* coefficient defined on each vertex $v \in V$, and $p \leq |V|$ a positive integer number. Given a vertex subset $U \subseteq V$, we denote by $m_U = \min_{u \in U} w_u$ and $M_U = \max_{u \in U} w_u$ the minimum and maximum weight in U, respectively, and by *gap* their difference $\gamma_U = M_U - m_U$. The *Minimum Gap Graph Partitioning*

[1] Email: maurizio.bruglieri@polimi.it

[2] Email: roberto.cordone@unimi.it

Problem (*MGGPP*) requires to partition G into p vertex-disjoint connected subgraphs $G_r = (V_r, E_r)$, $(r = 1, \dots, p)$ with at least two vertices each. Its *min-max* and *min-sum* versions minimize, respectively, the maximum gap f^{MM} and the sum of the gaps f^{MS} over all subgraphs:

$$f^{MM} = \max_{r=1,\dots,p} \gamma_{V_r} \qquad f^{MS} = \sum_{r=1}^{p} \gamma_{V_r}$$

The *MGGPP* can find applications in agriculture (divide a land into parcels with limited difference in height [3]), in the location of gate houses along rivers, and in social network analysis (identify connected clusters of members with homogeneous features). It falls in the large field of graph partitioning problems [1,2], but, as far as we know, objective functions related to the differences between vertex weights in each subgraph have never been considered before.

2 Complexity

Theorem 2.1 *The* MGGPP *admits feasible solutions if and only if graph G contains a matching of cardinality at least p.*

Proof. Any maximum cardinality matching M induces on graph G a spanning forest of $|M|$ nondegenerate trees and $|V| - 2|M|$ isolated vertices. Each isolated vertex v has an incident edge e_v which is adjacent to an edge in M. Adding e_v to M for each isolated vertex v, we obtain a spanning forest of exactly $|M|$ trees. If $|M| > p$, we consider the edges connecting different trees, and we add them to M, stopping as soon as we obtain exactly p trees. This provides a feasible solution of the *MGGPP*. Vice versa, given a feasible solution, we can choose an edge from each subgraph (they all contain at least two vertices): these edges are nonadjacent, and yield a p-cardinality matching. □

Let $W_U = \{z \in \mathbb{Z} : \exists v \in U \text{ with } w_v = z\}$ be the set of values assumed by w on a subset of vertices $U \subseteq V$, and $\eta_U = |W_U|$ the number of such values.

Theorem 2.2 *The* MGGPP *with the* min-max *objective function is strongly $\mathcal{NP}$-hard even if $p = 2$ and $\eta_V = 3$.*

Proof. The decision version of the problem, obviously in $\mathcal{NP}$, amounts to verifying the existence of a solution such that the gap of all subgraphs is not larger than a given threshold. Given a generic instance of *SAT*, we build the following auxiliary graph. We introduce for each literal (x_i or $\bar{x}_i$) a vertex (v_i or $\bar{v}_i$) with $w_{v_i} = w_{\bar{v}_i} = 2$, and for each clause C_j a vertex c_j with weight $w_{c_j} = 1$;

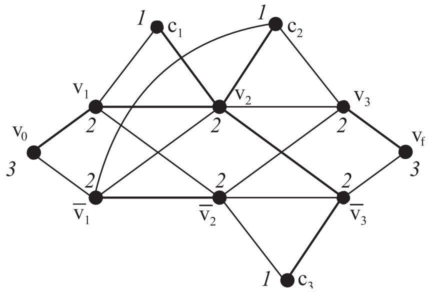

Fig. 1. Graph construction for the $\mathcal{NP}$-hardness proof of the min-max *MGGPP*

finally, we introduce two dummy vertices v_0 and v_f with weight $w_0 = w_f = 3$. Vertex v_0 is connected to v_1 and $\bar{v}_1$; vertex v_f is connected to v_n and $\bar{v}_n$; each vertex v_i (resp. $\bar{v}_i$) is connected to v_{i+1} and $\bar{v}_{i+1}$ $(i = 1, \ldots, n-1)$ and to all the clause vertices c_j such that literal x_i (resp. $\bar{x}_i$) occurs in clause C_j. We are looking for $p = 2$ connected subgraphs with gaps not larger than 1. Figure 1 shows the graph corresponding to $(x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$. If both subgraphs have gap ≤ 1, v_0 and v_f belong to the same subgraph, and this connects them through a path entirely made of vertices v_i or $\bar{v}_i$. By construction, this path contains at least one of v_i or $\bar{v}_i$ for each variable x_i. The other subgraph contains all the clause vertices c_j and connects them through adjacent vertices v_i or $\bar{v}_i$ which identify literals satisfying all clauses. Such a truth assignment is consistent because the subgraph includes at most one vertex for each variable x_i. Vice versa, any satisfying truth assignment identifies a partition of the graph into two subgraphs with gap ≤ 1. □

Theorem 2.3 *The* MGGPP *with the* min-sum *objective function is strongly $\mathcal{NP}$-hard even if $\eta_V = 2$.*

Proof (Sketch). The proof is by reduction from 3-*SAT*. □

3 Approximability

Theorem 3.1 *The* min-max MGGPP *cannot be approximated for any constant $\alpha < 2$ unless $\mathcal{P} = \mathcal{NP}$.*

Proof. Following Theorem 2.2, we can build an instance with optimum equal to 1 for any YES-instance of *SAT* and one with optimum equal to 2 for any NO-instance. By contradiction, a hypothetical α-approximated polynomial algorithm with $\alpha < 2$, would find on the former instances solutions with a value < 2 (by integrality, 1), and therefore solve *SAT* in polynomial time. □

Theorem 3.2 *The* MGGPP *is* 2*-approximable for $p = 2$.*

Proof. Let V_1^* and V_2^* be the unknown subsets of vertices of the optimal solution. The ranges of the weights in the two subgraphs, $\left[m_{V_1^*}; M_{V_1^*}\right]$ and $\left[m_{V_2^*}; M_{V_2^*}\right]$, are either separate or overlapping. In the former case, all the vertices in a subgraph have weights strictly smaller than those in the other. Then, the optimal solution can be found by exhaustively considering all pairs of intervals $[w_{\pi_1}, w_{\pi_k}]$ and $\left[w_{\pi_{k+1}}, w_{\pi_\eta}\right]$ $(k = 1, \ldots, \eta_V - 1)$, and building the subgraphs induced on G by the vertices whose weights fall in the two intervals. In the latter case, the two ranges overlap, and $f^{*MS} = \gamma_{V_1^*} + \gamma_{V_2^*} \geq \gamma_V$, which implies $f^{*MM} = \max(\gamma_1^*, \gamma_2^*) \geq \gamma_V/2$. Generating any feasible solution with Theorem 2.1, we obtain $f^{MS} \leq 2\gamma_V \leq 2f^{*MS}$ and $f^{MM} \leq \gamma_V \leq 2f^{*MM}$. □

4 Some special cases

The *MGGPP* admits some polynomially solvable special cases.

Proposition 4.1 *The* min-max MGGPP *is polynomially solvable if* $\eta_V = 2$.

Proof (Sketch). If there is a vertex whose weight is different from that of the adjacent vertices, the optimal solution is γ_V. Otherwise, we merge all the adjacent vertices of equal weight and consider the resulting vertex set V'. If $|V'| > p$, the optimum is γ_V; otherwise, a procedure similar to that of Theorem 2.1 provides an optimal solution with p subgraphs of zero gap. □

Proposition 4.2 *The* min-sum *and* min-max MGGPP *are polynomially solvable on line graphs.*

Proof (Sketch). The proof is based on the computation by dynamic programming of the minimum bottleneck path on a suitable graph. □

We are currently investigating the complexity of other special cases and working on the design of exact and heuristic algorithms.

References

[1] D. A. Bader, H. Meyerhenke, P. Sanders, and D. Wagner, eds. *Graph Partitioning and Graph Clustering*, v. 588 of *Contemporary Mathematics*. AMS, 2013.

[2] C.-E. Bichot and P. Siarry, editors. *Graph Partitioning*. Wiley-ISTE, 2013.

[3] Li Xiao, Li Hongpeng, Niu Dongling, Wang Yan, and Liu Gang. Optimization of GNSS-controlled land leveling system and related experiments. *Transactions of the Chinese Society of Agricultural Engineering*, 31(3):48–55, 2015.