Minimum flow decompositions from multiassembly problems to line planning

Fernando Dias[1] and Philine Schiewe[2]

[1]Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland, ✉ fernando.dias@aalto.fi
[2]Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland, ✉ philine.schiewe@aalto.fi

1 Background and motivation

In this work, we introduce a novel variant of the minimum flow decomposition problem that handles undirected edges, aiming to find a minimum set of paths that covers every edge within a given range. We show how this problem can be used to model multiassembly problems in genome reconstruction [5] as well as line planning problems in public transport [3].

Although finding a feasible flow decomposition in directed flow networks is a trivial task, determining a minimum set is NP-hard even for directed acyclic graphs [4]. As this translates to undirected graphs, we propose a mixed-integer linear program to find minimum flow decompositions for undirected graphs. The formulation is tested on biological data as well as transport data.

2 Problem formulation

In this work, we consider *undirected flow networks* $(G = (V, E), \underline{f}, \overline{f})$ with undirected graph G, vertex set V, edge set $E \subseteq \binom{V}{2}$ and lower and upper bounds on the flow values $\underline{f}, \overline{f} \colon E \to \mathbb{N}$. A *path* in this network is an edge progression connecting two different vertices without any vertex repetition.

Definition 1 (Undirected minimum path flow decomposition problem ($\min k-$UFP))**.** *A* path decomposition $(\mathcal{P}, w)$ *of an undirected flow network corresponds to a set of undirected paths* $\mathcal{P} = (P_1, \dots, P_k)$ *and their associated weight* $w = (w_1, \dots, w_k) \in \mathbb{N}^k$ *that satisfy*

$$\underline{f}_{\{u,v\}} \leq \sum_{j\colon \{u,v\} \in P_j} w_j \leq \overline{f}_{\{u,v\}}$$

for all edges $\{u, v\} \in E$*. The* undirected minimum path flow decomposition problem ($\min k-$UFP) *is to find a path decomposition* $(\mathcal{P}, w)$ *of minimal cardinality* $|\mathcal{P}|$*.*

We modify undirected flow networks as follows to simplify the corresponding mixed-integer linear program.

Definition 2 (Oriented flow network)**.** *A tuple* $N'' = (G'' = (V'', E''), \underline{f}, \overline{f})$ *is said to be an* oriented flow network *based on an undirected flow network* $N = (G = (V, E), \underline{f}, \overline{f})$ *if* $V'' = V \cup \{s, t\}$ *(for artificial source and sink* $s, t \notin V$*) and* $E'' = E' \cup \{(s, u), (u, t) \colon u \in V\}$ *where* $E' = \{(u, v), (v, u) \colon \{u, v\} \in E\}$ *is the set of oriented edges corresponding to* E*.*

It is easy to show that oriented flow networks can be used to model $\min k-$UFP.

Lemma 1. *There is a undirected path decomposition* $(\mathcal{P}, w)$ *of* N *with cardinality* k *if and only if there is decomposition of* N'' *into* k *directed* s-t *paths* $\mathcal{P}' = (P_1, \dots, P_k)$ *with associated weights* $w'_1, \dots, w'_k$ *which satisfies*

$$\underline{f}_{\{u,v\}} \leq \sum_{j\colon (u,v) \in P_j} w'_j + \sum_{j\colon (v,u) \in P_j} w'_j \leq \overline{f}_{\{u,v\}}.$$

Based on Lemma 1, we adapt the arc-based mixed-integer program from [1] to $\min k-$UFP. As the underlying graph is no longer directed and acyclic, we need to explicitly model vertex-repetition-free paths by introducing auxiliary variables and constraints analogously to vehicle routing formulations.

3 Experimental evalution

We apply the undirected minimum path flow decomposition problems in two contexts. In *multiassembly problems*, multiple genomic sequences in different abundances must be reconstructed from a set of *reads* sequenced from them. A *weighted graph* is constructed from the reads where each read spanning two exons leads to an edge with the direction arbitrary based on the reference sequence. Hence, different reference sequences might yield different orientations. By using undirected graphs, we provide a decomposition that accounts for this uncertainty and simultaneously allows optimal solutions regardless of edge direction. For our experiments, we utilise data from three benchmarking datasets from biological data (*BIO*). Table 1 shows that our model can find optimal solutions for all instances within the time limit of 600s.

For *line planning in public transport*, the undirected flow network represents the infrastructure network with bounds on the minimum and maximum service frequency between neighbouring stops. A path naturally represents a *line* while the weight corresponds to the *frequency* of operation. It is easy to extend $\min k-$UFP by further practical constraints such as terminals and line costs. Note that $\min k-$UFP models the integrated problem of line generation, line selection and frequency setting, see [3]. Figures 1 and 2 show line plans for the benchmarking data set *Mandl* determined by our approach and the costs-minimizing standard procedure in the open source software package LinTim [2]. Note that our model reduces the costs of the line plan by 40.9% due to the integration of line generation.

	$\min k$	Amount	Avg.	Σ	Solved
SRR020730 Salmon	4-10	5691	5.13	29194	100
	11-15	95	12.40	1178	100
	16-20	16	17.30	276	100
	21-*max*	8	23.50	188	100
Reference Sim	4-10	6512	7.56	49230	100
	11-15	260	8.65	2249	100
	16-20	78	10.40	811	100
	21-*max*	40	34.60	1384	100
SRR30790 StringTie	4-10	864	18.70	16156	100
	11-15	104	43.60	4534	100
	16-20	70	80.50	5635	100
	21-*max*	27	200.00	5400	100

Table 1: Data set *BIO* with minimum path decomposition ($\min k$), number of instance (Amount), average and total runtime (Avg., $\sum$) and percentages of instances solved to optimality (Solved).

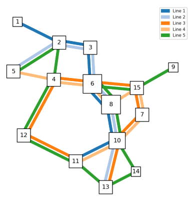

Figure 1: *Mandl* solved by $\min k-$UFP.
5 lines, costs 360.93.

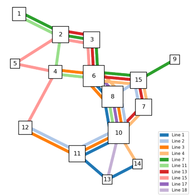

Figure 2: *Mandl* solved by LinTim.
10 lines, costs 610.71.

4 Conclusion and future work

Undirected minimum path flow decompositions show promise for modeling real-world problems related to both multiassembly problems in bioinformatics and line planning problems in public transport. In future work we aim to strengthen the developed mixed-integer programming formulation by introducing tailored cutting planes. In an experimental evaluation, we aim to quantify the influence of different objective functions as well as additional problem-specific constraints.

References

[1] F.H.C. Dias, L.W., B. Mumey, and A.I. Tomescu. Fast, Flexible, and Exact Minimum Flow Decompositions via ILP. In *RECOMB 2022 - 26th Annual International Conference on Research in Computational Molecular Biology*, volume 13278 of *Lecture Notes in Computer Science*, pages 230–245, -, 2022. Springer.

[2] P. Schiewe, A. Schöbel, O. Herrala, M. Rihlmann, S. Roth, S. Albert, C. Biedinger, T. Dahlheimer, L. Dittrich, K. Hoffmann, S. Jäger, P. Pattanaik, A. Schiewe, M. Stinzendörfer, and R. Urban. Documentation for lintim 2025.11, 2025.

[3] M. Schmidt and A. Schöbel. Planning and optimizing transit lines. *arXiv preprint arXiv:2405.10074*, 2024.

[4] B. Vatinlen, F. Chauvet, P. Chrétienne, and P. Mahey. Simple bounds and greedy algorithms for decomposing a flow into a minimal set of paths. *European Journal of Operational Research*, 185(3):1390–1401, 2008.

[5] Y. Xing, A. Resch, and C. Lee. The multiassembly problem: reconstructing multiple transcript isoforms from est fragment mixtures. *Genome research*, 14(3):426–441, 2004.