A practical fpt algorithm for Flow Decomposition and transcript assembly
Kyle Kloster, Philipp Kuinke, Michael P. O’Brien, Felix Reidl, Fernando Sánchez Villaamil, Blair D. Sullivan, Andrew van der Poel

The Motivation

Connecting overlapping segments and counting their frequencies yields a DAG and a flow. The problem is to split the flow into the least amount of s-t-paths, to recover the original DNA/RNA strands.

The Problem

k-Flow Decomposition (k-FD)
Input: (G, f, k) with G an s-t-DAG, f a flow on G, and k a positive integer.
Problem: Is there an integral flow decomposition of (G, f) using at most k paths?

The Algorithm

The routing g out of S3 (dashed lines) is an extension of the previous routings (solid paths). Each row in the constraint system L on the right corresponds to an arc; those shaded in gray are from arcs inside S3, and those in white come from g.

The Results

<table>
<thead>
<tr>
<th>dataset</th>
<th>instances</th>
<th>non-trivial</th>
<th>optimal</th>
<th>non-optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catfish</td>
<td>1,549,373</td>
<td>445,980</td>
<td>99.907%</td>
<td>0.053%</td>
</tr>
<tr>
<td>mouse</td>
<td>1,316,058</td>
<td>473,185</td>
<td>99.401%</td>
<td>0.074%</td>
</tr>
<tr>
<td>human</td>
<td>1,169,083</td>
<td>529,221</td>
<td>99.490%</td>
<td>0.043%</td>
</tr>
<tr>
<td>all</td>
<td>4,034,514</td>
<td>1,448,588</td>
<td>99.589%</td>
<td>0.056%</td>
</tr>
</tbody>
</table>

Since Toboggan finds optimal decompositions, we can investigate the Groundtruth for optimality.

Acknowledgments

This work is supported in part by the Gordon & Betty Moore Foundations Data-Driven Discovery Initiative through Grant GBMF4560 to Blair D. Sullivan.

Resources

The implemented solver is available on Github: /theoryinpractice/toboggan

Since Toboggan finds optimal decompositions we can investigate the Groundtruth for optimality.