DOMINANT-EDGE PATHWAY:
A Weighted Graph Algorithm for Identifying Dominant Metabolic Pathways

Ehsan Ullah & Soha Hassoun
Department of Computer Science

Kyongbum Lee
Department of Chemical and Biological Engineering

Tufts University
Ethanol Synthesis in *E. Coli*

(Wlaschin, et al., 2006)
Graph Representation of Metabolic Networks

\[ R_1 : A \rightarrow B \]
\[ R_2 : B \rightarrow C \]
\[ R_3 : C \rightarrow D \]
\[ R_4 : B \rightarrow E + F \]
\[ R_5 : B \rightarrow 2G \]
\[ R_6 : G \rightarrow H \]
\[ R_7 : F \rightarrow H \]
\[ R_8 : H \rightarrow I \]
Elementary Flux Mode

• An elementary flux mode is a minimal set of reactions that can operate at steady state (Schuster & Hilgetag, 1994)

• Metabolic flux distributions in living cells can be represented as non-negative linear combinations of elementary modes
Elementary Flux Mode Analysis: An Example

A → D
A → E + I
A → 2I
Elementary Modes Can be Combined to Produce Several Possible Metabolic Flux Distributions
Our Approach for Pathway Analysis

• Weighted graph search for a path containing a “dominant edge” that provides the most restrictive bottleneck

• Dominance defined using edge weights that represent Gibbs free energy change
Chemical Potential

Electric Circuits

Chemical Reactions

A

High Chemical Potential

∆G < 0

Energy

Spontaneous

B

Low Chemical Potential

∆G > 0

Low Chemical Potential

High Chemical Potential
Why Gibbs Free Energy?

• $\Delta G$ gives estimate of reaction likelihood
• For most metabolic reactions $\Delta G \approx 0$
• Flux through an entire pathway will be heavily influenced by a smaller subset of reactions that have very negative or positive $\Delta G$
Gibbs Free Energy

• Gibbs Free Energy (G)
  – Maximum amount of work that can be extracted from a closed system
  – Calculated using Group Contribution Theory (Mavrovouniotis, 1990)

• Gibbs Free Energy Change (ΔG)
  – Difference of Gibbs Free Energy of products and substrates
Dominant-Edge Pathway

• Dominant reactions
  ▪ for a series of reactions, it has the most positive $\Delta G$
  ▪ for parallel reactions, it has the most negative $\Delta G$

• Stoichiometrically balanced pathway
Series Reactions

• Weights
  \(WR_4 = -0.5\)
  \(WR_7 = -2.5\)
• \(R_4\) dominates \(R_7\)
Parallel Reactions

• Weights
  \( WR_4 = -0.5 \)
  \( WR_5 = -2.5 \)
• \( R_5 \) dominates \( R_4 \)
Problem Statement

Given a metabolic network graph $G_m = (V, E)$, and starting and ending vertices $s$ and $t$, find a dominant-edge pathway from $s$ to $t$. 
Algorithm

1. Identify dominant-edge path from $s$ to $t$
   - Use modified Dijsktra’s algorithm to identify dominant-edge pathways

2. Augment the found path to realize a stoichiometrically balanced pathway
   - Use modified Dijsktra’s algorithm starting with the intermediate byproducts
Intermediate Byproducts

source S and target T

- S
- B
- G
- H
- F
- E
- C
- D
- T

Arrows indicate the reaction paths: R₁, R₂, R₃, R₄, R₅, R₆, R₇, R₈, R₉.
Intermediate Byproducts

Dominant-Edge Path from S to T with E as a byproduct.

The path is not stoichiometrically balanced.
Modified Dijkstra’s Algorithm

• Dijkstra’s Shortest Path Algorithm
  ▪ Finds shortest paths from a given source vertex \( v \in V \) to all other vertices in weighted graph \( G=(V, E) \), such that all edge weights are nonnegative

• Modified Dijkstra’s Algorithm (Bottleneck Problem)
  (D.R. Fulkerson, 1966)
  ▪ Selects reaction with most negative \( \Delta G \) in parallel reactions
  ▪ Select reaction with most positive \( \Delta G \) in series reaction
Example Problem

Find Dominant-Edge Pathway from S to T
Example Problem

\[ S = \{ \} \]
\[ Q = \{ S, A, B, C, D, E, F, T \} \]
Example Problem

$S = \{ \}$

$Q = \{ S, A, B, C, D, E, F, T \}$
Example Problem

\[ S = \{ S \} \]
\[ Q = \{ A, B, C, D, E, F, T \} \]
Example Problem

\[ S = \{ S \} \]
\[ Q = \{ A, B, C, D, E, F, T \} \]
Example Problem

$S = \{ S, F \}$

$Q = \{ A, B, C, D, E, T \}$
Example Problem

\[ S = \{ S, T, A, B, C, D, E, F \} \]
\[ Q = \{ \} \]
Dominant-Edge Path

Node D is an intermediate byproduct
Augmenting Path

Path from byproducts to final product
Runtime

• Dominant Path: $O(|V|^2 + |E|)$

• Augmenting Path: $O(|V|^2 + |E|)$
Test Cases

1. Pentose Fermentation in *Zymomonas mobilis* (bacterium)
2. *Escherichia coli*
3. Liver Cell

<table>
<thead>
<tr>
<th></th>
<th>Compounds</th>
<th>Reactions</th>
<th>Total EFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test case 1</td>
<td>21</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Test case 2</td>
<td>47</td>
<td>60</td>
<td>33,000</td>
</tr>
<tr>
<td>Test case 3</td>
<td>38</td>
<td>60</td>
<td>188</td>
</tr>
</tbody>
</table>
Test Case 1

(Altintas, et al., 2006)
Test Case 2

(Wlaschin, et al., 2006)
Test Case 2

Total elementary flux modes : 33,000

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Compounds</th>
<th>Reactions</th>
<th>Super EFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fructose</td>
<td>Ethanol</td>
<td>12</td>
<td>11</td>
<td>169</td>
</tr>
<tr>
<td>Glucose</td>
<td>Ethanol</td>
<td>13</td>
<td>12</td>
<td>156</td>
</tr>
<tr>
<td>Xylose</td>
<td>Ethanol</td>
<td>21</td>
<td>19</td>
<td>725</td>
</tr>
</tbody>
</table>

Each DOMINANT-EDGE pathway between an input sugar and ethanol exactly corresponds to the maximal conversion pathway engineered through gene knockouts by Wlaschin et al., 2006
### Test Case 3

Total elementary flux modes : 188

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Compounds</th>
<th>Reactions</th>
<th>Overlapping EFMss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Glucose</td>
<td>10</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Alanine</td>
<td>Urea</td>
<td>8</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Glucose</td>
<td>10</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Urea</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Glycine</td>
<td>Alanine</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Glycine</td>
<td>Glucose</td>
<td>11</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
</tbody>
</table>
### Runtime

- Windows Machine
- C++ Implementation
- 4GB Memory

<table>
<thead>
<tr>
<th>Test Case</th>
<th>EFM Analysis</th>
<th>Our Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt; 1s</td>
<td>&lt; 1s</td>
</tr>
<tr>
<td>2</td>
<td>10s</td>
<td>&lt; 1s</td>
</tr>
<tr>
<td>3</td>
<td>&lt; 1s</td>
<td>&lt; 1s</td>
</tr>
</tbody>
</table>
ΔG Used in Recent Studies

• ΔG analysis of a *E. Coli* resulted in identifying four reactions representing thermodynamic bottlenecks in the production of growth (Henry et al., 2006)

• Using thermodynamics-based constraints to enforce the exclusion of thermodynamically infeasibilities when calculating flux distribution (e.g. a flux cannot be positive unless ΔG is negative) (Henry et al., 2007)
Conclusion

• An Efficient way of finding a maximum conversion pathway with runtime advantages over enumeration-based approaches

• The Dominant-Edge Pathway found is identical, proper subset, or partially overlaps with EFM results
Applications & Future Work

• Pathway Analysis has promising applications
  ▪ Understanding living systems
  ▪ Efficient *in silico* experimentation
  ▪ Engineering microorganisms to perform specific synthesis tasks

• Future Work
  ▪ Find an energetically favored pathway (most negative overall Gibbs energy change)
  ▪ Find alternative pathways with (nearly) equivalent bottleneck or pathway weights
Acknowledgements

• Gautham Sridharan (BCE Tufts University)
Questions
Thank You
Questions for Ehsan?

• Why use EFM instead of Extreme Pathways? How would your results be different?
• How did you calculate the delta_Gs?
• How do you ensure that your paths are stoichiometrically balanced in terms of conversion?