Quantifying Flows in Time-Irreversible Markov Chains: Application to Gene Regulatory Network

Danielle Middlebrooks

Advisor: Dr. Maria Cameron

Applied Mathematics, Statistics and Scientific Computation
University of Maryland- College Park

January 15, 2020

This work is partially supported by NSF COMBINE-NRT Program under Grant No. DGE-1632976 and NSF Career grant DMS 1554907
1 Background
   ■ Markov Chains
   ■ Transition Path Theory

2 Methodology for Quantifying Transitions

3 Application to Gene Regulatory Networks
   ■ Exploring the Dynamical Network
   ■ Mutation Analysis

4 Conclusion
Discrete-Time Markov Chains

A discrete-time Markov chain is defined by:

- A sequence of random variables $(X_n)_{n \geq 0} \in a countable set $S$ characterized by the Markov property,
- Transition matrix $P$,
- Initial distribution $\lambda$. 

A $2x2$ transition matrix is:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.6</td>
</tr>
</tbody>
</table>

A $3x3$ transition matrix is:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>0.6</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.6</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.25</td>
<td>1.0</td>
</tr>
</tbody>
</table>
A Markov chain with transition matrix $P$ and stationary distribution $\pi$ satisfying

$$\pi P = \pi, \quad \sum_{i \in S} \pi = 1$$

is called **time-reversible** if it satisfies the detailed balance condition, i.e

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

A Markov chain is called **time-irreversible** if the detailed balance condition is not satisfied. Hence, the transition probabilities for the reversed process is given by

$${\hat{P}}_{i,j} = \frac{\pi_j}{\pi_i} P_{j,i}$$
Objective

- Time-irreversible Markov chains can arise in applications in Economics, Physics, Social sciences, Biology, and Etc.

- Our goal is to develop efficient computational tools for the study of transition process in large and complex Markov chains.
Transition Path Theory (TPT) is a framework to analyze the statistical properties of reactive trajectories i.e. those going from $A$ to $B$ without returning to $A$ in between.
Key Concepts of Transition Path Theory

The forward committor function $q^+ = (q^+_i)_{i \in S}$ is the probability that starting at a state $i$, the trajectory will reach set $B$ prior to set $A$ and satisfies:

$$
\begin{cases}
q^+_i = \sum_{j \in S} P_{i,j} q^+_j, & i \in S \setminus (A \cup B) \\
q^+_i = 0, & i \in A \\
q^+_i = 1, & i \in B
\end{cases}
$$

(1)
Key Concepts of Transition Path Theory

The forward committor function $q^+ = (q^+_i)_{i \in S}$ is the probability that starting at a state $i$, the trajectory will reach set $B$ prior to set $A$ and satisfies:

$$\begin{cases} 
q_i^+ = \sum_{j \in S} P_{i,j} q_j^+, & i \in S \setminus (A \cup B) \\
q_i^+ = 0, & i \in A \\
q_i^+ = 1, & i \in B 
\end{cases}$$ (1)

The backward committor $q^- = (q^-_i)_{i \in S}$ is the probability that the process arriving at state $i$ last came from $A$ rather than $B$ and satisfies:

$$\begin{cases} 
q_i^- = \sum_{j \in S} \hat{P}_{i,j} q_j^-, & i \in S \setminus (A \cup B) \\
q_i^- = 1, & i \in A \\
q_i^- = 0, & i \in B 
\end{cases}$$ (2)

with $\hat{P}_{i,j}$ being the transition matrix for the time-reversed process.
Key Concepts of Transition Path Theory

The probability current of reactive trajectories is given by

$$f_{i,j} = \begin{cases} 
\pi_i q_i^- P_{i,j} q_j^+ , & \text{if } i \neq j \\
0 , & \text{otherwise}
\end{cases}$$

(3)
Key Concepts of Transition Path Theory

The probability current of reactive trajectories is given by

\[ f_{i,j} = \begin{cases} \pi_i q_i^- P_{i,j} q_j^+, & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases} \] (3)

The effective current

\[ f_{i,j}^+ = \max\{f_{i,j} - f_{j,i}, 0\} \] (4)
Key Concepts of Transition Path Theory

The probability current of reactive trajectories is given by

$$f_{i,j} = \begin{cases} \pi_i q_i^- P_{i,j} q_j^+, & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The effective current

$$f_{i,j}^+ = \max\{f_{i,j} - f_{j,i}, 0\} \quad (4)$$

Transition rate

$$\nu_{AB} = \sum_{i \in A, j \in S} f_{ij} = \sum_{i \in S, j \in B} f_{ij} \quad (5)$$
Quantifying Flows in Time-Irreversible Markov Chains: Application to Gene Regulatory Network

Background

Transition Path Theory

Illustrative example

Original MC

Time-reversed MC

Forward committor, Backward committor and effective current

Transition rate: $\nu_{ab} = 1/15$
Challenges with Time-Irreversible Markov Chains

Figure: A cyclic effective current (green). Transition probabilities (black).
We develop an algorithm for generating a weighted directed acyclic graph $G(S, \{F^+\})$

**Input:** Weighted directed graph $G(S, \{f^+\})$

**Output:** Weighted directed acyclic graph $G(S, \{F^+\})$

**The main body**

```plaintext
while flag == 0 do
    Find cycle in $G(S, \{f^+\})$ using DFS algorithm
    if cycle is found then
        Find minimum current in the cycle $f_{min}^+$
        for each edge in the cycle
            subtract minimum current from edge
        end for
    else
        flag == 1
    end
end
```
Application to Gene Regulatory Network
GRN for Budding Yeast Cell Cycle


Figure: Budding Yeast Cell Cycle.

Figure: Gene-regulatory network of budding yeast.
Quantifying Flows in Time-Irreversible Markov Chains: Application to Gene Regulatory Network

Application to Gene Regulatory Network

Background

Deterministic Model

- \( a_{ij} = 1 \) for protein \( j \) activating protein \( i \).
- \( a_{ij} = -1 \) for protein \( j \) repressing protein \( i \).
- Each node \( i \) has only two states, \( S_i = 1 \) and \( S_i = 0 \).

\[
S_i(t+1) = \begin{cases} 
1, & \sum_i a_{ij} S_j(t) > 0 \\
0, & \sum_i a_{ij} S_j(t) < 0 \\
S_i(t), & \sum_i a_{ij} S_j(t) = 0 
\end{cases}
\]

Figure: Dynamical trajectories
Quantifying Flows in Time-Irreversible Markov Chains: Application to Gene Regulatory Network

Application to Gene Regulatory Network

Background

Figure: Dynamical trajectories

Figure: Biological pathway of deterministic model
let \( v = As \)

\[
P\{s_1(t+1), \ldots, s_{11}(t+1) | s_1(t), \ldots, s_{11}(t)\} = \prod_{i=1}^{11} P\{s_i(t+1) | s_1(t), \ldots, s_{11}(t)\} \tag{6}
\]

where if \( v_i \neq 0 \)

\[
P\{s_i(t + 1) = 1 | s_1(t), \ldots, s_{11}(t)\} = \frac{e^{\beta v_i}}{e^{\beta v_i} + e^{-\beta v_i}} \tag{7}
\]

\[
P\{s_i(t + 1) = 0 | s_1(t), \ldots, s_{11}(t)\} = \frac{e^{-\beta v_i}}{e^{\beta v_i} + e^{-\beta v_i}} \tag{8}
\]

if \( v_i = 0 \)

\[
P\{s_i(t + 1) = s_i(t) | s_1(t), \ldots, s_{11}(t)\} = \frac{1}{1 + e^{-\alpha}} \tag{9}
\]
Results of Cycle Removal Algorithm to GRN

Figure: Acyclic current through cell cycle for $\alpha = 5$, $\beta = 6$.  

Cln3, Sic1, Cdh1
SBF, MBF, Sic1, Cdh1
Cln1,2, Clb1,2, Mcm1, Cdc20, Swi5
SBF, MBF, Cln1,2, Clb1,2, Mcm1, Cdc20
Cln1,2, Clb1,2, Mcm1, Cdc20
Sic1, Cdh1, Mcm1, Cdc20, Swi5
Cln1,2, Clb1,2, Mcm1, Cdc20
Sic1, Cdh1

$74\%$
$24\%$
$96\%$
$71\%$
$97\%$
$91\%$
$90\%$
$85\%$
$70\%$
$91\%$
$99\%$
$98\%$
$73\%$
$24\%$
$23\%$
$22\%$

$Sbf$, $MBF$, $Sic1$, $Cdh1$
$SBF$, $MBF$, $Cln1,2$, $Clb5,6$
$SBF$, $MBF$, $Cln1,2$, $Clb1,2$, $Cln1,2$
$SBF$, $MBF$, $Clb1,2$, $Mcm1$, $Cdc20$, $Swi5$
$Sbf$, $MBF$, $Cln1,2$, $Clb5,6$
$Sbf$, $MBF$, $Cln1,2$, $Sic1$, $Cdh1$

Quantifying Flows in Time-Irreversible Markov Chains: Application to Gene Regulatory Network

Application to Gene Regulatory Network

Background
Results of Cycle Removal Algorithm to Stochastic Model

Figure: Acyclic current through cell cycle for $\alpha = 5$, $\beta = 3$. 
We use our cycle removal algorithm to identify essential edges in the GRN.
We use our cycle removal algorithm to identify essential edges in the GRN.

- Recompute transition matrix
- Run cycle removal algorithm on $G(S, \{f^+\})$ to obtain acyclic current
- Run DFS algorithm to obtain pathways
### Analysis of Deterministic Model

<table>
<thead>
<tr>
<th>Non-Essential Edges</th>
<th>Essential Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No effect:</strong></td>
<td><em>Cln3</em> → <em>Cln3</em></td>
</tr>
<tr>
<td><em>Clb5,6</em> → <em>Sic1</em></td>
<td><em>Cln3</em> → <em>SBF</em></td>
</tr>
<tr>
<td><em>Clb5,6</em> → <em>Cdhl</em></td>
<td><em>Cln3</em> → <em>MBF</em></td>
</tr>
<tr>
<td><em>Cdhl</em> → <em>Clb1,2</em></td>
<td><em>SBF</em> → <em>Cln1,2</em></td>
</tr>
<tr>
<td><em>Clb1,2</em> → <em>Sic1</em></td>
<td><em>MBF</em> → <em>Cln5,6</em></td>
</tr>
<tr>
<td><em>Clb1,2</em> → <em>Cdc20/Cdc14</em></td>
<td><em>Cln1,2</em> → <em>Cln1,2</em></td>
</tr>
<tr>
<td></td>
<td><em>Cln1,2</em> → <em>Sic1</em></td>
</tr>
<tr>
<td></td>
<td><em>Sic1</em> → <em>Clb1,2</em></td>
</tr>
<tr>
<td></td>
<td><em>Clb1,2</em> → <em>SBF</em></td>
</tr>
<tr>
<td></td>
<td><em>Clb1,2</em> → <em>MBF</em></td>
</tr>
<tr>
<td></td>
<td><em>Mcm1/SFF</em> → <em>Mcm1/SFF</em></td>
</tr>
<tr>
<td></td>
<td><em>Mcm1/SFF</em> → <em>Swi5</em></td>
</tr>
<tr>
<td></td>
<td><em>Cdc20/Cdc14</em> → <em>Sic1</em></td>
</tr>
<tr>
<td></td>
<td><em>Cdc20/Cdc14</em> → <em>Clb5,6</em></td>
</tr>
<tr>
<td></td>
<td><em>Cdc20/Cdc14</em> → <em>Cdhl</em></td>
</tr>
<tr>
<td></td>
<td><em>Cdc20/Cdc14</em> → <em>Clb1,2</em></td>
</tr>
<tr>
<td></td>
<td><em>Cdc20/Cdc14</em> → <em>Cdc20/Cdc14</em></td>
</tr>
<tr>
<td></td>
<td><em>Cdc20/Cdc14</em> → <em>Swi5</em></td>
</tr>
<tr>
<td></td>
<td><em>Swi5</em> → <em>Sic1</em></td>
</tr>
<tr>
<td></td>
<td><em>Swi5</em> → <em>Swi5</em></td>
</tr>
</tbody>
</table>

| Total: 14                                                                          | Total: 20                              |
|                                                                                   |                                        |
| **Small effect:**                                                                 |                                        |
| *Clb5,6* → *Mcm1/SFF*                                                             |                                        |
| *Clb1,2* → *Cdhl*                                                                  |                                        |
| *Clb1,2* → *Mcm1/SFF*                                                             |                                        |
| *Clb1,2* → *Swi5*                                                                 |                                        |
| *Mcm1/SFF* → *Clb1,2*                                                             |                                        |
| *Mcm1/SFF* → *Cdc20/Cdc14*                                                        |                                        |
| **Lost of G2 phase:**                                                              |                                        |
| *Cln1,2* → *Cdhl*                                                                  |                                        |
| *Sic1* → *Clb5,6*                                                                  |                                        |
| *Clb5,6* → *Clb1,2*                                                                |                                        |
|                                                                                   |                                        |
| **Total: 20**                                                                      |                                        |
## Analysis of Stochastic Model

<table>
<thead>
<tr>
<th>Non-Essential Edges</th>
<th>Essential Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>4</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>Total: 30</strong></td>
<td><strong>Total: 4</strong></td>
</tr>
</tbody>
</table>

### Non-Essential Edges:
- **No effect:**
  - $Clb_{5,6} \rightarrow Sic1$
  - $Clb_{5,6} \rightarrow Cdh1$
  - $Cdh1 \rightarrow Clb_{1,2}$
  - $Clb_{1,2} \rightarrow Sic1$
  - $Clb_{1,2} \rightarrow Cdc_{20}/Cdc_{14}$

### Essential Edges:
- **Lost of G2 phase:**
  - $Cln_{1,2} \rightarrow Cdh1$
  - $Cln_{1,2} \rightarrow Sic1$
  - $Sic1 \rightarrow Clb_{5,6}$
  - $Clb_{5,6} \rightarrow Clb_{1,2}$

- **Small effect:**
  - $Cln3 \rightarrow Cln3$
  - $Cln3 \rightarrow MBF$
  - $Cln_{1,2} \rightarrow Cln_{1,2}$
  - $Sic1 \rightarrow Clb_{1,2}$
  - $Clb_{1,2} \rightarrow SBF$
  - $Clb_{1,2} \rightarrow MBF$
  - $Clb_{5,6} \rightarrow Mcm1/SFF$
  - $Clb_{1,2} \rightarrow Cdh1$
  - $Clb_{1,2} \rightarrow Mcm1/SFF$
  - $Clb_{1,2} \rightarrow Swi5$
  - $Mcm1/SFF \rightarrow Clb_{1,2}$
  - $Mcm1/SFF \rightarrow Cdc_{20}/Cdc_{14}$
  - $Mcm1/SFF \rightarrow Mcm1/SFF$
  - $Mcm1/SFF \rightarrow Swi5$
  - $Cdc_{20}/Cdc_{14} \rightarrow Sic1$
  - $Cdc_{20}/Cdc_{14} \rightarrow Clb_{5,6}$
  - $Cdc_{20}/Cdc_{14} \rightarrow Cdh1$
  - $Cdc_{20}/Cdc_{14} \rightarrow Clb_{1,2}$
  - $Cdc_{20}/Cdc_{14} \rightarrow Cdc_{20}/Cdc_{14}$
  - $Cdc_{20}/Cdc_{14} \rightarrow Swi5$
  - $Swi5 \rightarrow Swi5$

- **Lost of S and G2 phases:**
  - $MBF \rightarrow Clb_{5,6}$

- **Lost of S, G2 and M phases:**
  - $Swi5 \rightarrow Sic1$

- **No pathway found:**
  - $Cln3 \rightarrow SBF$
  - $SBF \rightarrow Cln_{1,2}$

- **Non-Essential Edges**

- **Essential Edges**
Comparison of Results

**Deterministic**
- No effect:
  - $Clb_{5,6} \rightarrow Sic1$
  - $Clb_{5,6} \rightarrow Cdhl$
  - $Cdhl \rightarrow Clb_{1,2}$
  - $Clb_{1,2} \rightarrow Sic1$
  - $Clb_{1,2} \rightarrow Cdc20/Cdc14$

- Small effect:
  - $Clb_{5,6} \rightarrow Mcm1/SFF$
  - $Clb_{1,2} \rightarrow Cdhl$
  - $Clb_{1,2} \rightarrow Mcm1/SFF$
  - $Clb_{1,2} \rightarrow Swi5$
  - $Mcm1/SFF \rightarrow Clb_{1,2}$
  - $Mcm1/SFF \rightarrow Cdc20/Cdc14$

- Lost of G2 phase:
  - $Cln_{1,2} \rightarrow Cdhl$
  - $Sic1 \rightarrow Clb_{5,6}$
  - $Clb_{5,6} \rightarrow Clb_{1,2}$

**Stochastic**
- Small effect:
  - $Cln3 \rightarrow Cln3$
  - $Cln3 \rightarrow MBF$
  - $Cln_{1,2} \rightarrow Cln_{1,2}$
  - $Sic1 \rightarrow Clb_{1,2}$
  - $Clb_{1,2} \rightarrow SBF$
  - $Clb_{1,2} \rightarrow MBF$
  - $Mcm1/SFF \rightarrow Mcm1/SFF$
  - $Mcm1/SFF \rightarrow Swi5$
  - $Cdc20/Cdc14 \rightarrow Sic1$
  - $Cdc20/Cdc14 \rightarrow Clb_{5,6}$
  - $Cdc20/Cdc14 \rightarrow Cdc20/Cdc14$
  - $Cdc20/Cdc14 \rightarrow Clb_{1,2}$
  - $Cdc20/Cdc14 \rightarrow Cdc20/Cdc14$
  - $Cdc20/Cdc14 \rightarrow Swi5$
  - $Swi5 \rightarrow Swi5$

- Lost of G2 phase:
  - $Cln_{1,2} \rightarrow Sic1$

- Lost of S and G2 phases:
  - $MBF \rightarrow Clb_{5,6}$

- Lost of S, G2 and M phases:
  - $Swi5 \rightarrow Sic1$
Conclusion

- We developed a methodology supported by theoretical results for quantifying transition processes in time-irreversible Markov chains.

- This technique is applied to the Budding yeast GRN.

- Stochastic GRN is much more robust to mutation analysis compared to deterministic GRN.

Future research:

- Develop strategy for selecting key subset of nodes to make applicable to larger and more complex networks.
References


Cameron and Vanden-Eijnden (2013)

Two modified Markov jump processes were designed for the original time-reversible irreducible Markov chain.

- The stationary probability current coincided with the probability current of reactive trajectories.
- The stationary probability current was equal to the reactive current.

Cameron and Middlebrooks

- Combined these two propositions to a generalized version for time-irreversible Markov chains.
Theorem (Transition Path Process: Cameron & Middlebrooks)

Suppose we have defined a current $e$ satisfying the following properties:

1. Non-negativity: $e_{ij} \geq 0$
2. The conservation of current: $\forall i \in S_R$, $\sum_{j \in S} (e_{ij} - e_{ji}) = 0$
3. Transition rate: $\sum_{i \in A} \sum_{j \in S} e_{ij} = \sum_{i \in S} \sum_{j \in B} e_{ij} = \nu_{AB}$
Conclusion

Theorem (Transition Path Process: Cameron & Middlebrooks)

Suppose we have defined a current $e$ satisfying the following properties:

1. Non-negativity: $e_{ij} \geq 0$
2. The conservation of current: $\forall i \in S_R, \sum_{j \in S} (e_{ij} - e_{ji}) = 0$
3. Transition rate: $\sum_{i \in A} \sum_{j \in S} e_{ij} = \sum_{i \in S} \sum_{j \in B} e_{ij} = \nu_{AB}$

Let $S_R = S \setminus (A \cup B)$ and $R := \{i \in S_R | \exists j \in S : e_{ij} > 0\}$. Consider the process on the state space $\tilde{S} = R \cup \{s\}$ defined by the generator $M$ given by

$$
\begin{align*}
M_{ij} &= \frac{e_{ij}}{\mu_i}, & i, j \in R \\
M_{is} &= \sum_{j \in B} \frac{e_{ij}}{\mu_i}, & i \in R \\
M_{sj} &= \frac{1}{1 - \rho_R} \sum_{i \in A} e_{ij}, & j \in R
\end{align*}
$$

where $\rho_R = \sum_{i \in R} \mu_i$. 

(10)
Theorem (Transition Path Process: Cameron & Middlebrooks)

Suppose we have defined a current \( e \) satisfying the following properties:

1. Non-negativity: \( e_{ij} \geq 0 \)
2. The conservation of current: \( \forall i \in S_R, \sum_{j \in S} (e_{ij} - e_{ji}) = 0 \)
3. Transition rate: \( \sum_{i \in A} \sum_{j \in S} e_{ij} = \sum_{i \in S} \sum_{j \in B} e_{ij} = \nu_{AB} \)

Let \( S_R = S \setminus (A \cup B) \) and \( R := \{ i \in S_R | \exists j \in S : e_{ij} > 0 \} \). Consider the process on the state space \( \tilde{S} = R \cup \{s\} \) defined by the generator \( M \) given by

\[
\begin{align*}
M_{ij} & = \frac{e_{ij}}{\mu_i}, \quad i, j \in R \\
M_{is} & = \sum_{j \in B} \frac{e_{ij}}{\mu_i}, \quad i \in R \\
M_{sj} & = \frac{1}{1 - \rho_R} \sum_{i \in A} e_{ij}, \quad j \in R
\end{align*}
\]

(10)

where \( \rho_R = \sum_{i \in R} \mu_i \). Then the desired invariant probability distribution of the transition path process is given by \( \tilde{\mu}_i = \begin{cases} 
\mu_i, & i \in R \\
1 - \rho_R, & i = s \end{cases} \) and the stationary current in the network with state space \( \tilde{S} \) and the generator matrix \( M \) coincides with the current \( e \) in the original network.
Outline of proof

Proof.

To show the stationary current in the network with state space $\tilde{S}$ and the generator matrix $M$ coincides with the current $e$ in the original network we must show the following:

- The invariant distribution in the modified MJP is $\tilde{\mu}$, i.e show
  \[ \sum_{i \in R \cup \{s\}} \tilde{\mu}_i M_{ij} = 0. \]

- The stationary current in the MJP with generator matrix $M$ coincides with stationary current $E_{i,j} = e_{i,j} - e_{j,i}$. 