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

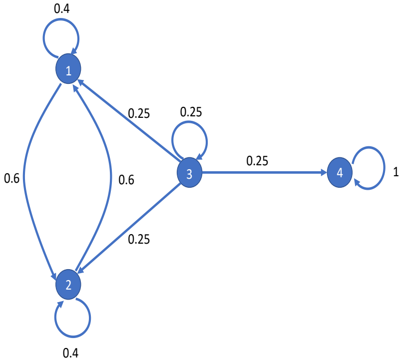
Outline

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 λ.



Time-Reversible VS Time-Irreversible

- A Markov chain with transition matrix P and stationary distribution π 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

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

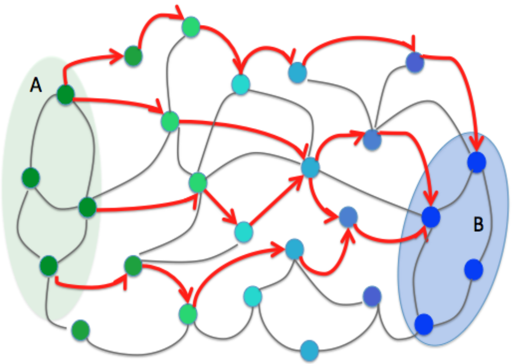
Objective

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

- Our goal is to develop efficient computational tools for the study of transition process in large and complex Markov chains.

Transition Path Theory: E & Vanden-Eijnden (2006)

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:

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

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:

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

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:

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

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

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

Key Concepts of Transition Path Theory

The probability current of reactive trajectories is given by

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

The effective current

$$f_{i,j}^+ = \max\{f_{i,j} - f_{j,i}, 0\} \tag{4}$$

Key Concepts of Transition Path Theory

The probability current of reactive trajectories is given by

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

The effective current

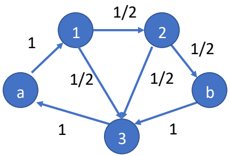
$$f_{i,j}^+ = \max\{f_{i,j} - f_{j,i}, 0\} \tag{4}$$

Transition rate

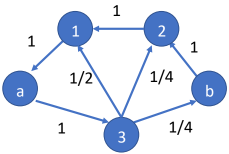
$$\nu_{AB} = \sum_{i \in A, j \in S} f_{ij} = \sum_{i \in S, j \in B} f_{ij} \tag{5}$$

Illustrative example

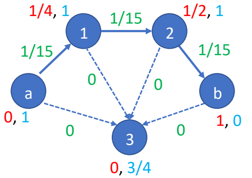
Original MC



Time-reversed MC



Forward committor, Backward committor and effective current



Transition rate:
ν_{ab}=1/15

Challenges with Time-Irreversible Markov Chains

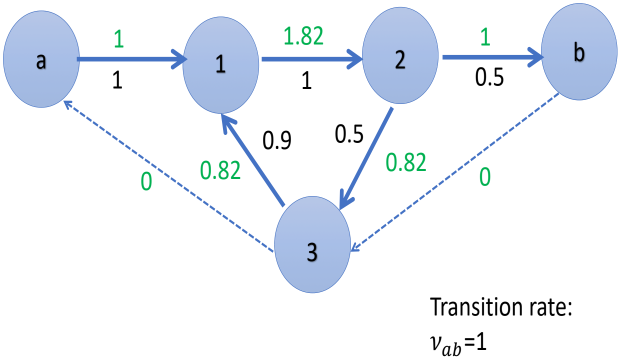


Figure: A cyclic effective current (green). Transition probabilities (black).

Cycle Removal Algorithm for Obtaining Acyclic Current

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
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

Chen, Csikasz-Nagy, Gyorffy, Val, Novak, & Tyson (2000)
Li, Long, Lu, Ouyang, & Tang (2003)

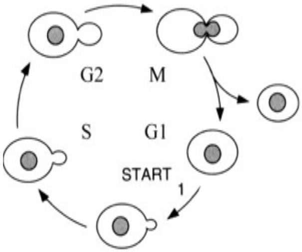


Figure: Budding Yeast Cell Cycle.

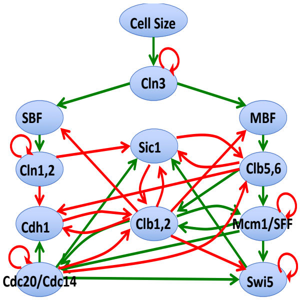


Figure: Gene-regulatory network of budding yeast.

Deterministic Model

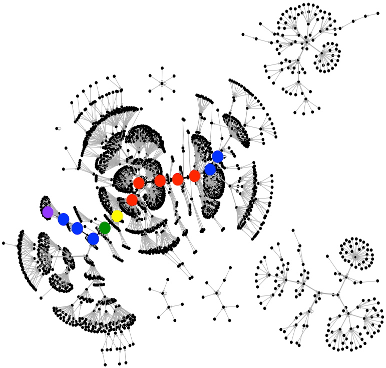


Figure: Dynamical trajectories

- $a_{ij} = 1$ 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 $S_i = 0$.

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

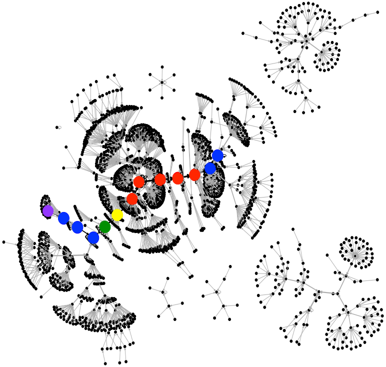


Figure: Dynamical trajectories

i	Cln3	SBF	MBF	Cln1,2	Sic1	Clb5,6	Cdh1	Clb1,2	Mcm1/SFF	Cdc20,14	Swi5	Phase
1	1	0	0	0	1	0	1	0	0	0	0	START
2	0	1	1	0	1	0	1	0	0	0	0	G1
3	0	1	1	1	1	0	1	0	0	0	0	G1
4	0	1	1	1	0	0	0	0	0	0	0	G1
5	0	1	1	1	0	1	0	0	0	0	0	S
6	0	1	1	1	0	1	0	1	1	0	0	G2
7	0	0	0	1	0	1	0	1	1	1	0	M
8	0	0	0	0	0	0	0	1	1	1	1	M
9	0	0	0	0	1	0	0	1	1	1	1	M
10	0	0	0	0	1	0	0	0	1	1	1	M
11	0	0	0	0	1	0	1	0	0	1	1	M
12	0	0	0	0	1	0	1	0	0	0	1	G1
13	0	0	0	0	1	0	1	0	0	0	0	G1*

Figure: Biological pathway of deterministic model

Stochastic Model: Zhang, Qian, Ouyang, Deng, Li, & Tang (2006)

let $v = As$

$$\mathbb{P}\{s_1(t+1), ..., s_{11}(t+1) | s_1(t), ..., s_{11}(t)\} = \Pi_{i=1}^{11} \mathbb{P}\{s_i(t+1) | s_1(t), ..., s_{11}(t)\} \quad (6)$$

where if $v_i \neq 0$

$$\mathbb{P}\{s_i(t+1) = 1 | s_1(t), ..., s_{11}(t)\} = \frac{e^{\beta v_i}}{e^{\beta v_i} + e^{-\beta v_i}} \quad (7)$$

$$\mathbb{P}\{s_i(t+1) = 0 | s_1(t), ..., s_{11}(t)\} = \frac{e^{-\beta v_i}}{e^{\beta v_i} + e^{-\beta v_i}} \quad (8)$$

if $v_i = 0$

$$\mathbb{P}\{s_i(t+1) = s_i(t) | s_1(t), ..., s_{11}(t)\} = \frac{1}{1 + e^{-\alpha}} \quad (9)$$

Results of Cycle Removal Algorithm to GRN

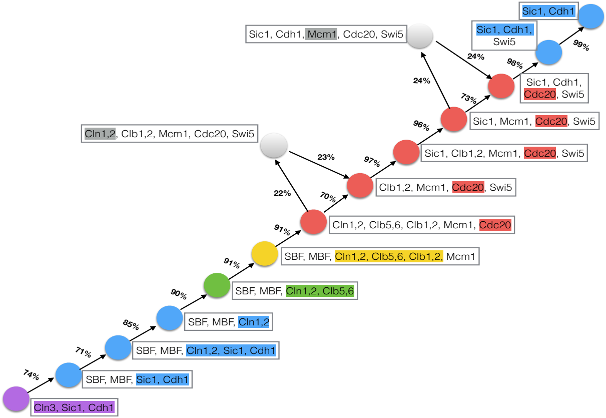


Figure: Acyclic current through cell cycle for $\alpha = 5$, $\beta = 6$.

Results of Cycle Removal Algorithm to Stochastic Model

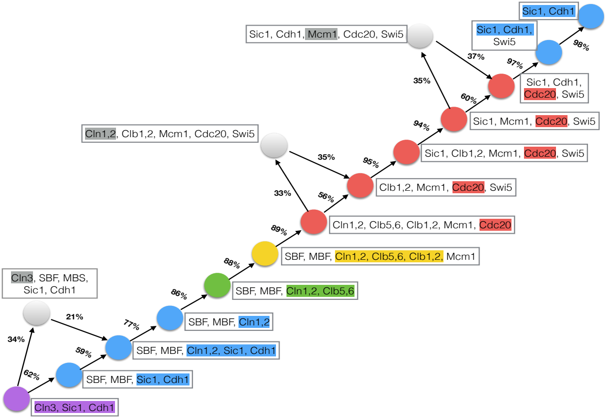


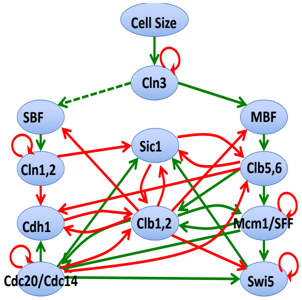
Figure: Acyclic current through cell cycle for $\alpha = 5$, $\beta = 3$.

Mutation Analysis (Cameron & Middlebrooks)

We use our cycle removal algorithm to identify essential edges in the GRN.

Mutation Analysis (Cameron & Middlebrooks)

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

	Non-Essential Edges
5	No effect: $Clb5,6 \rightarrow Sic1$ $Clb5,6 \rightarrow Cdh1$ $Cdh1 \rightarrow Clb1,2$ $Clb1,2 \rightarrow Sic1$ $Clb1,2 \rightarrow Cdc20/Cdc14$
6	Small effect: $Clb5,6 \rightarrow Mcm1/SFF$ $Clb1,2 \rightarrow Cdh1$ $Clb1,2 \rightarrow Mcm1/SFF$ $Clb1,2 \rightarrow Swi5$ $Mcm1/SFF \rightarrow Clb1,2$ $Mcm1/SFF \rightarrow Cdc20/Cdc14$
3	Lost of G2 phase: $Cln1,2 \rightarrow Cdh1$ $Sic1 \rightarrow Clb5,6$ $Clb5,6 \rightarrow Clb1,2$

Total: 14

Essential Edges
$Cln3 \rightarrow Cln3$
$Cln3 \rightarrow SBF$
$Cln3 \rightarrow MBF$
$SBF \rightarrow Cln1,2$
$MBF \rightarrow Cln5,6$
$Cln1,2 \rightarrow Cln1,2$
$Cln1,2 \rightarrow Sic1$
$Sic1 \rightarrow Clb1,2$
$Clb1,2 \rightarrow SBF$
$Clb1,2 \rightarrow MBF$
$Mcm1/SFF \rightarrow Mcm1/SFF$
$Mcm1/SFF \rightarrow Swi5$
$Cdc20/Cdc14 \rightarrow Sic1$
$Cdc20/Cdc14 \rightarrow Clb5,6$
$Cdc20/Cdc14 \rightarrow Cdh1$
$Cdc20/Cdc14 \rightarrow Clb1,2$
$Cdc20/Cdc14 \rightarrow Cdc20/Cdc14$
$Cdc20/Cdc14 \rightarrow Swi5$
$Swi5 \rightarrow Sic1$
$Swi5 \rightarrow Swi5$

Total: 20

Analysis of Stochastic Model

Non-Essential Edges			
5	No effect: $Clb5,6 \rightarrow Sic1$ $Clb5,6 \rightarrow Cdh1$ $Cdh1 \rightarrow Clb1,2$ $Clb1,2 \rightarrow Sic1$ $Clb1,2 \rightarrow Cdc20/Cdc14$	Small effect: $Cln3 \rightarrow Cln3$ $Cln3 \rightarrow MBF$ $Cln1,2 \rightarrow Cln1,2$ $Sic1 \rightarrow Clb1,2$ $Clb1,2 \rightarrow SBF$ $Clb1,2 \rightarrow MBF$ $Clb5,6 \rightarrow Mcm1/SFF$ $Clb1,2 \rightarrow Cdh1$ $Clb1,2 \rightarrow Mcm1/SFF$ $Clb1,2 \rightarrow Swi5$ $Mcm1/SFF \rightarrow Clb1,2$ $Mcm1/SFF \rightarrow Cdc20/Cdc14$ $Mcm1/SFF \rightarrow Mcm1/SFF$ $Mcm1/SFF \rightarrow Swi5$ $Cdc20/Cdc14 \rightarrow Sic1$ $Cdc20/Cdc14 \rightarrow Clb5,6$ $Cdc20/Cdc14 \rightarrow Cdh1$ $Cdc20/Cdc14 \rightarrow Clb1,2$ $Cdc20/Cdc14 \rightarrow Cdc20/Cdc14$ $Cdc20/Cdc14 \rightarrow Swi5$ $Swi5 \rightarrow Swi5$	21
4	Lost of G2 phase: $Cln1,2 \rightarrow Cdh1$ $Cln1,2 \rightarrow Sic1$ $Sic1 \rightarrow Clb5,6$ $Clb5,6 \rightarrow Clb1,2$		

Total: 30

Essential Edges	
Lost of S and G2 phases: $MBF \rightarrow Clb5,6$	1
Lost of S, G2 and M phases: $Swi5 \rightarrow Sic1$	1
No pathway found $Cln3 \rightarrow SBF$ $SBF \rightarrow Cln1,2$	2

Total:4

Comparison of Results

Deterministic

No effect:
$Clb5,6 \to Sic1$
$Clb5,6 \to Cdh1$
$Cdh1 \to Clb1,2$
$Clb1,2 \to Sic1$
$Clb1,2 \to Cdc20/Cdc14$

Small effect:
$Clb5,6 \to Mcm1/SFF$
$Clb1,2 \to Cdh1$
$Clb1,2 \to Mcm1/SFF$
$Clb1,2 \to Swi5$
$Mcm1/SFF \to Clb1,2$
$Mcm1/SFF \to Cdc20/Cdc14$

Lost of G2 phase:
$Cln1,2 \to Cdh1$
$Sic1 \to Clb5,6$
$Clb5,6 \to Clb1,2$

Stochastic

Small effect:
$Cln3 \to Cln3$
$Cln3 \to MBF$
$Cln1,2 \to Cln1,2$
$Sic1 \to Clb1,2$
$Clb1,2 \to SBF$
$Clb1,2 \to MBF$
$Mcm1/SFF \to Mcm1/SFF$
$Mcm1/SFF \to Swi5$
$Cdc20/Cdc14 \to Sic1$
$Cdc20/Cdc14 \to Clb5,6$
$Cdc20/Cdc14 \to Cdh1$
$Cdc20/Cdc14 \to Clb1,2$
$Cdc20/Cdc14 \to Cdc20/Cdc14$
$Cdc20/Cdc14 \to Swi5$
$Swi5 \to Swi5$

Lost of G2 phase:
$Cln1,2 \to Sic1$

Lost of S and G2 phases:
$MBF \to Clb5,6$

Lost of S, G2 and M phases:
$Swi5 \to Sic1$

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

1. Weinan, E., and Vanden-Eijnden, Eric. "Towards a Theory of Transition Paths." Journal of statistical physics 123.3 (2006).
2. Metzner, Philipp, Schütte, Christof, and Vanden-Eijnden, Eric. "Transition path theory for Markov jump processes." Multiscale Modeling & Simulation 7.3 (2009): 1192-1219
3. Maria Cameron and Eric Vanden-Eijnden. Flows in complex networks: theory, algorithms, and application to Lennard-Jones cluster rearrangement. Journal of Statistical Physics, 156(3):427–454, 2014.
4. Yuping Zhang, Minping Qian, Qi Ouyang, Minghua Deng, Fangting Li, and Chao Tang. Stochastic model of yeast cell-cycle network. Physica D: Nonlinear Phenomena, 219(1):35–39, 2006.
5. Katherine C Chen, Attila Csikasz-Nagy, Bela Gyorffy, John Val, Bela Novak, and John J Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. Molecular biology of the cell, 11(1):369–391, 2000.
6. Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. Introduction to algorithms. MIT press, 2009.
7. Robert Tarjan. Depth-first search and linear graph algorithms. SIAM journal on computing, 1(2):146–160, 1972.
8. Fangting Li, Tao Long, Ying Lu, Qi Ouyang, and Chao Tang. The yeast cell-cycle network is robustly designed. Proceedings of the National Academy of Sciences, 101(14):4781–4786, 2004.

Framework for Quantifying Transitions

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: $\sum_{i \in A} \sum_{j \in S} e_{ij} = \sum_{i \in S} \sum_{j \in B} e_{ij} = \nu_{AB}$

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*

$$\begin{cases} M_{ij} = \frac{e_{ij}}{\mu_i}, & i, j \in R \\ M_{is} = \sum_{j \in B} \frac{e_{ij}}{\mu_i}, & i \in R \\ M_{sj} = \frac{1}{1-\rho_R} \sum_{i \in A} e_{ij}, & j \in R \end{cases} \tag{10}$$

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

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*

$$\begin{cases} M_{ij} = \frac{e_{ij}}{\mu_i}, & i,j \in R \\ M_{is} = \sum_{j \in B} \frac{e_{ij}}{\mu_i}, & i \in R \\ M_{sj} = \frac{1}{1-\rho_R} \sum_{i \in A} e_{ij}, & j \in R \end{cases} \tag{10}$$

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.*

Framework for Quantifying Transitions

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 $E_{i,j} = e_{i,j} - e_{j,i}$.

$\square$