Branching Random Walks Applied to Antibody Affinity Maturation

IRENE BALELLI

SUPERVISORS: VUK MILIŠIĆ, GILLES WAINRIB

EXTERNAL PARTNERS: NADINE VARIN-BLANK'S TEAM (INSERM U978)

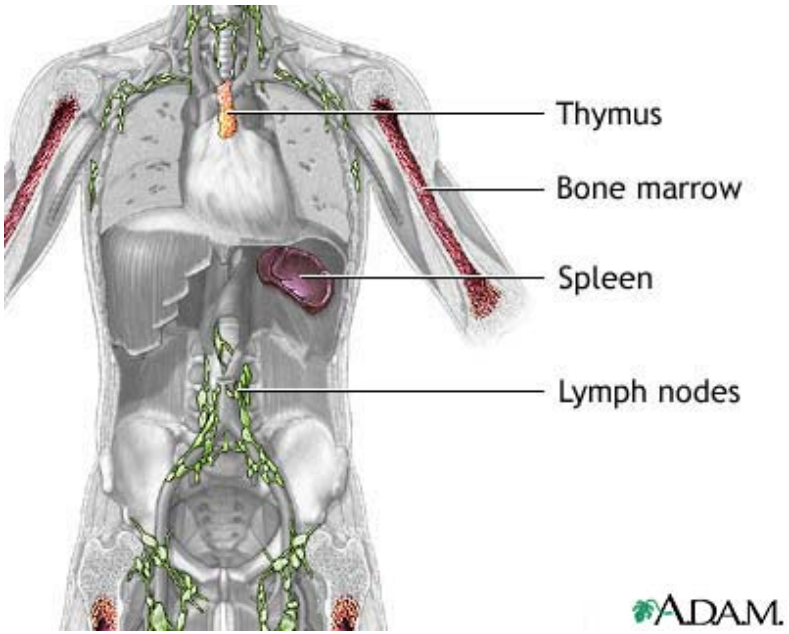
LAGA - University Paris 13

IHÉS - Bures-sur-Yvette - May 17, 2016

Contents

Plan

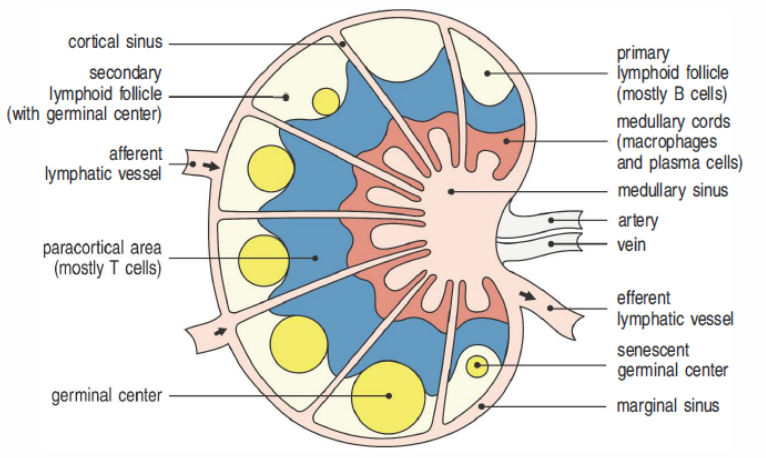
The immune system



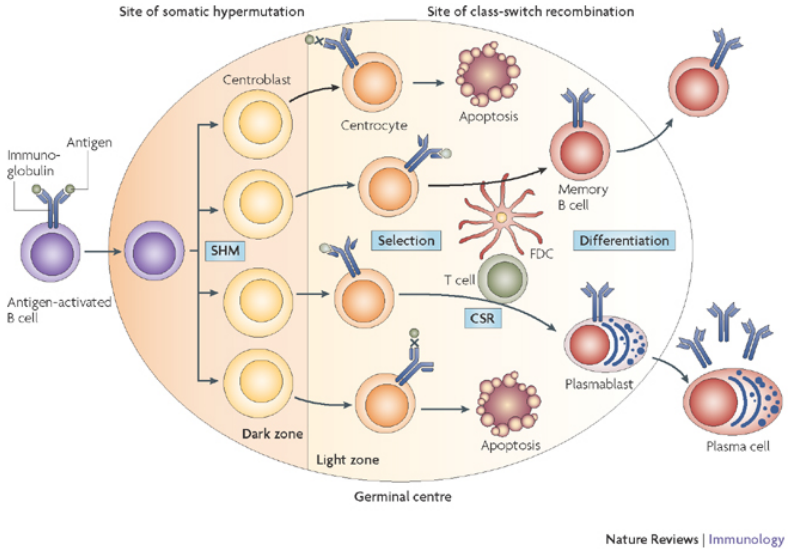
The structure of the immunity system (Encyclopedia of University of Maryland Medical Center)

- The immune system:
  - innate
  - adaptative
- Production of antigen-specific antibodies:
  - assured by B-cells
  - evolutionary mutation - seclection process

The germinal center reaction

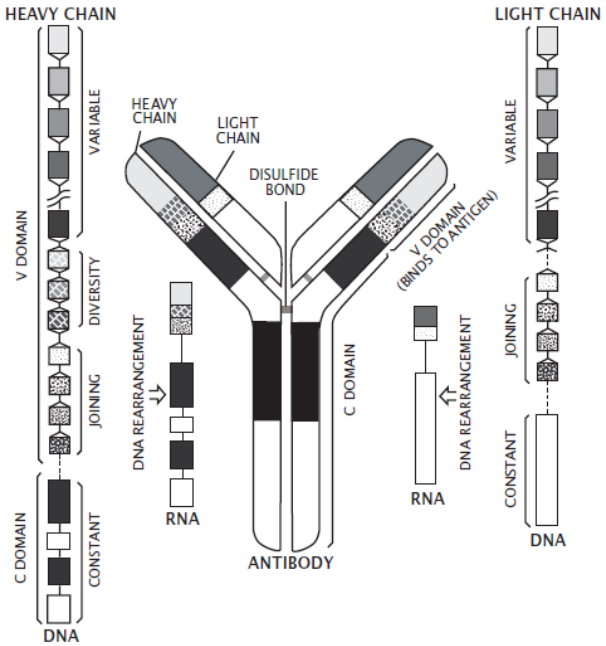


Organization of a lymph node (Janeway's immunobiology, 2012)



The germinal center microenvironment (Germinal centres: role in B-cell physiology and malignancy, *Nature Reviews Immunology* **8**, 2008)

Somatic hypermutation



The coding and assembly of BCR molecules during somatic hypermutation (Immunology and evolution of infectious disease, 2002)

- Genetic mutations on the variable region of the BCR, the antigen binding site
- Extremely high rate of mutation ($+10^5$-10^6)
- Random mutations

Aim and Model

Aim :
To build a mathematical framework to investigate the interactions between division, mutation and selection

Model :

- 2 amino acid classes: 0 or 1
- BCR and antigen = N-length binary strings ($\mathcal{H}_N := \{0,1\}^N$)
- Affinity = $N-$ Hamming distance between the strings
- To define a mutation rule = to define a random walk on $\mathcal{H}_N$

Plan

Definitions (1/2)

(a) Simple point mutations: class switch of a randomly chosen amino acid. **Mathematically:** Simple Random Walk on $\mathcal{H}_N$.

$\mathcal{P}$:= transition probability matrix

ex. $N = 5$: 1 1 0 0 1 $\longrightarrow$ 1 0 0 0 1

(b) Class switch of 1 or 2-length strings depending on affinity: class switch of 1 or 2 randomly chosen amino acids depending on the affinity between BCR and antigen.
Mathematically: graph divided into 2 components. The one containing the antigen is accessible from the other, not conversely.

ex. $N = 5$: $\bar{\mathbf{x}} =$ 1 1 0 0 1 ; $X_0 =$ 1 0 0 1 0

1 0 0 1 0 $\rightarrow$ 1 0 0 0 1 $\rightarrow$ 1 0 1 0 1

Definitions (2/2)

(c) Multiple point mutations: with probability a_i, i independent simple point mutations, $1 \leq i \leq k$, $k \leq N$ fixed.
Mathematically: two models proposed

- $\boxed{\mathcal{P}^{(k)} := \frac{1}{k} \sum_{i=1}^{k} \mathcal{P}^i}$

  With probability $1/k$ at each time step between 1 and k independent simple point mutations

- $\boxed{\mathcal{P}^{k^*},\ k^* = 2\lfloor (k+1)/2 \rfloor - 1}$

  At each time step exactly k^* independent simple-point mutations

The hitting time (1/2)

Definition : The expected number of steps to reach a specific node in $\mathcal{H}_N$, given the departure node.

$\mathbb{E}_{\mathbf{x}_i}[\tau_{\{\mathbf{x}_j\}}]$, where $\tau_{\{\mathbf{x}_j\}} := \inf\{n \geq 0 \mid \mathbf{X}_n = \mathbf{x}_j\}$.

Interpretation : The expected time we need to wait until the optimal BCR is obtained, given a particular antigen.

Computation : For the mutational models introduced, we determine explicit formulas to evaluate this quantity (or at least estimations for N big enough).

The hitting time (2/2)

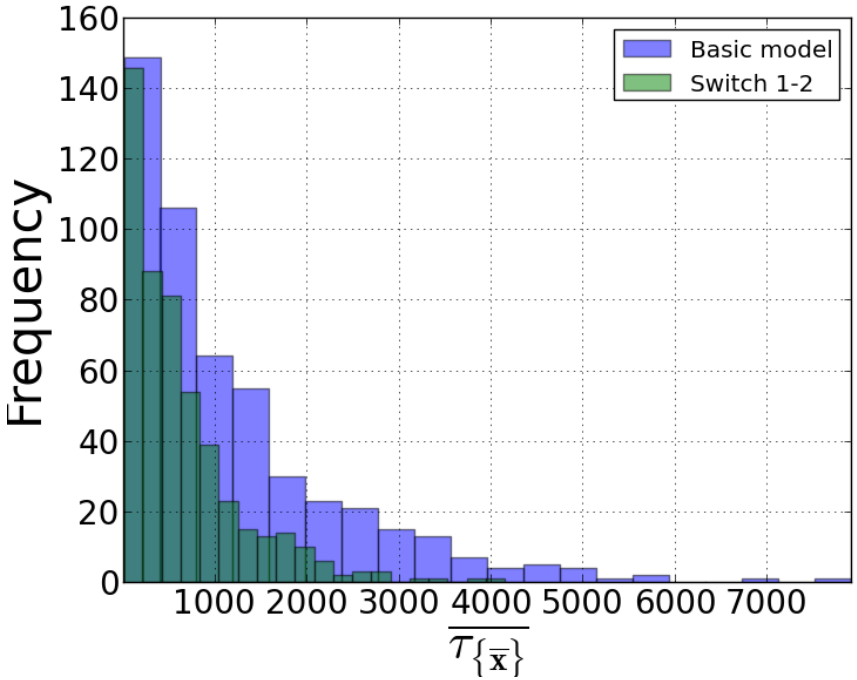
Let $\overline{d}$ be the initial Hamming distance between BCR and antigen.

Rule (a) $H(\overline{d}) = \sum_{d=0}^{\overline{d}-1} \frac{\sum_{j=1}^{N-1-d} C_N^{d+j} + 1}{C_{N-1}^{d}} \sim 2^N$, for N big enough

Rule (b) $\sim 2^{N-1}$, for N big enough

Rule (c) $\overline{T}_N^{(k)}(\overline{d}) = \sum_{l=2}^{2^N} \mu_l^{(k)} - \frac{1}{2^N C_N^{\overline{d}}} \sum_{l=2}^{2^N} \mu_l^{(k)} R_N(l, \overline{d})$

Class switch of 1 or 2 length strings, depending on the Hamming distance to $\overline{\mathbf{x}}$



Histogram of the hitting times. $N = 10$

Theoretical result:

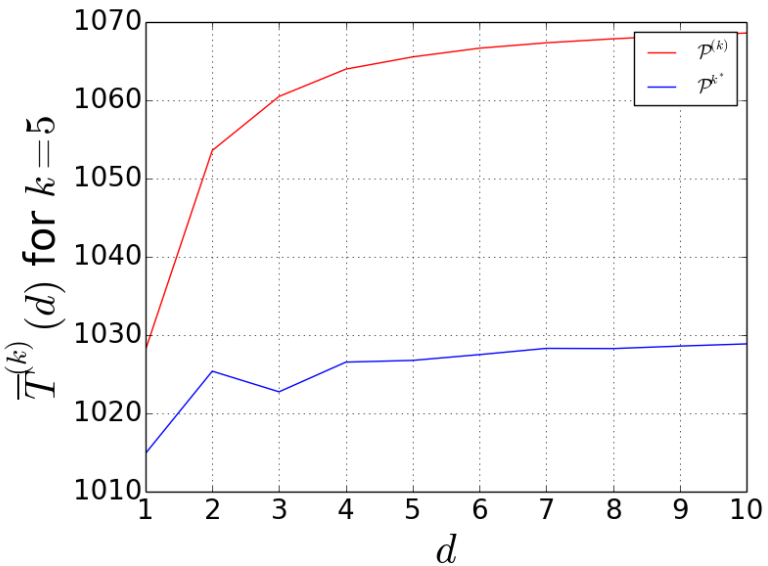
hitting time $= \frac{1}{2}$(hitting time for the basic model)

Experimental results for $N = 10$ (over 5000 simulations):

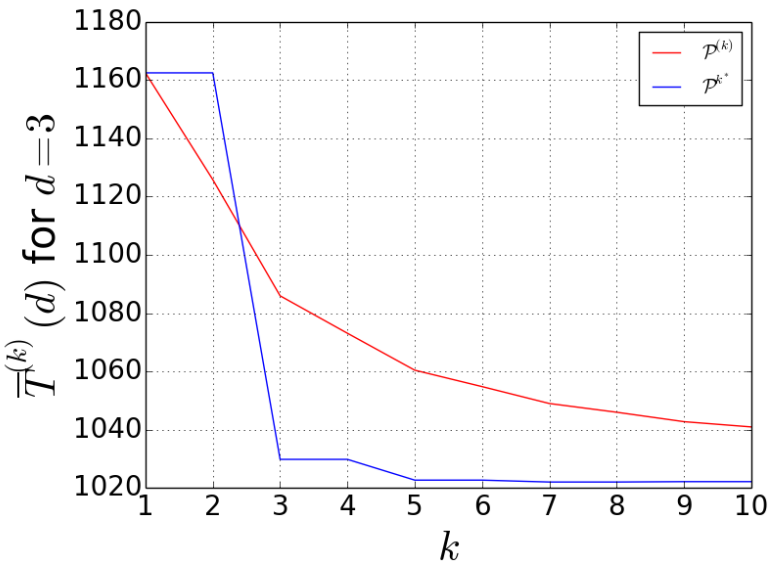
Basic model: 1188.8 ± 16.3
Switch 1-2: 602.8 ± 8.5

1 to k mutations

The spectral analysis let us conclude that $\mathcal{P}^{k^*}$ optimizes the mean hitting time to cover a given distance d, if $k > 2$.



Dependence of d on $\overline{T}^{(5)}_{10}(d)$.



Dependence of k on $\overline{T}^{(k)}_{10}(3)$.

$\overline{T}^{(k)}_{N}(d)$ = mean hitting time from a distance d allowing 1 to k mutations.

Plan

Motivation and definitions

Purpose : Introduction and analysis of the division process

Definition : Simple 2-Branching Random Walk

- $t = 0$: a randomly chosen node is labelled as active
- $t \to t + 1$: each active node chooses 2 neighbors to become active (independently and with replacement)
- possible states: active or non-active (never mind if a node is chosen more than once)

Portion of $\mathcal{H}_N$ covered in $\mathcal{O}(N)$

Notation 1 $S_t = \{\text{active nodes at } t\} \quad \Rightarrow \quad |S_t| = \#S_t$

Notation 2 2-BRW-$\mathcal{M}$ = a simple 2-BRW on a graph whose transition probability matrix is $\mathcal{M}$

Theorem

Given a simple 2-BRW-$\mathcal{P}$ on $\mathcal{H}_N$, in a time $T = \mathcal{O}(N)$ w.g.p. $|S_T| \geq 2^{N-r}$, *for* $r > \frac{N^2 e^{-2} + N - 2}{N e^{-2} + N - 2}$.

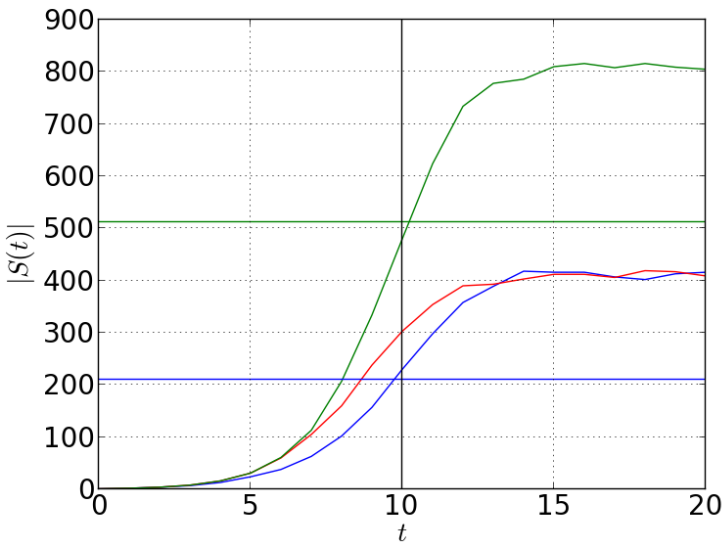
Theorem

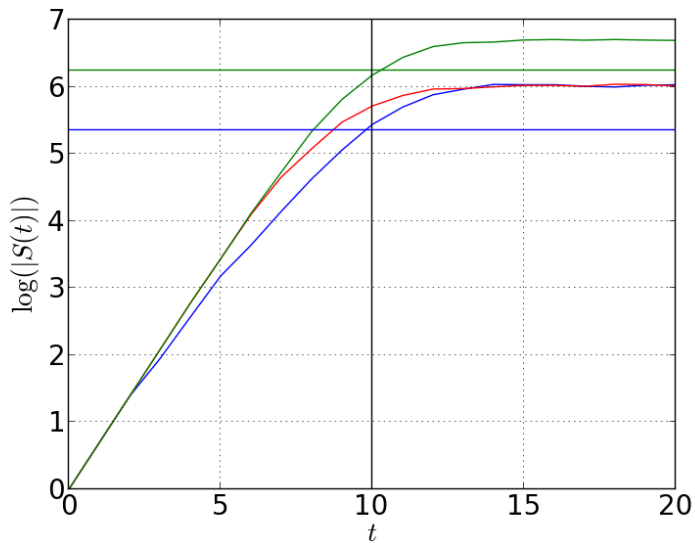
Given a simple 2-BRW-$\mathcal{P}^{(k)}$ on $\mathcal{H}_N$, in a time $T = \mathcal{O}(N)$ w.g.p. $|S_T| \geq \delta 2^N$, *for* $\delta \leq 1/2$.

[Dutta, C., Pandurangan, G., Rajaraman, R., Roche, S. 2013]

BRW with respect to $\mathcal{P}$, $\mathcal{P}^{k^*}$, $\mathcal{P}^{(k)}$

Using as transition probability matrix $\mathcal{P}$ or $\mathcal{P}^{k^*}$ the graph is bipartite: we can not have more than a half part of $\mathcal{H}_N$ active. With $\mathcal{P}^{(k)}$ we do not have this problem anymore: we can invade all the state space.





$|S(t)|$, comparing the 2-branching random walk for $\mathcal{P}$ (blue), $\mathcal{P}^7$ (red) and $\mathcal{P}^{(7)}$ (green).

$N = 10$

Plan

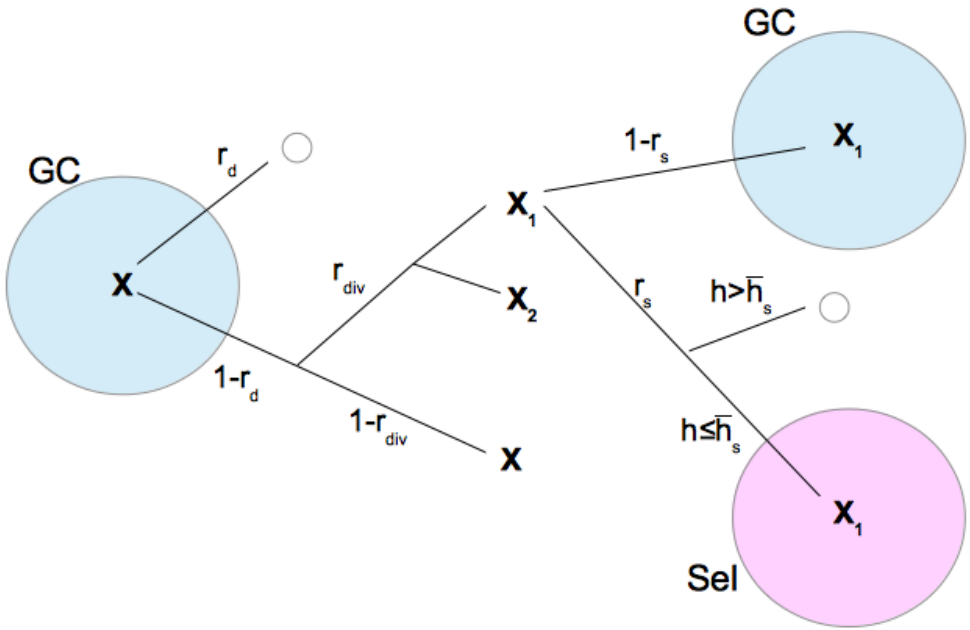
Definitions

$t = 0$: A B-cell enters GC with initial Hamming distance h_0

$t \to t+1$: Death rate: r_d; Division rate: r_{div} ; Selection rate: r_s

(a) If $h > \overline{h}_s \Rightarrow$ death ; if $h \leq \overline{h}_s \Rightarrow$ selected pool

(b) If $h > \overline{h}_s \Rightarrow$ nothing ; if $h \leq \overline{h}_s \Rightarrow$ selected pool



Evolution of the selected pool

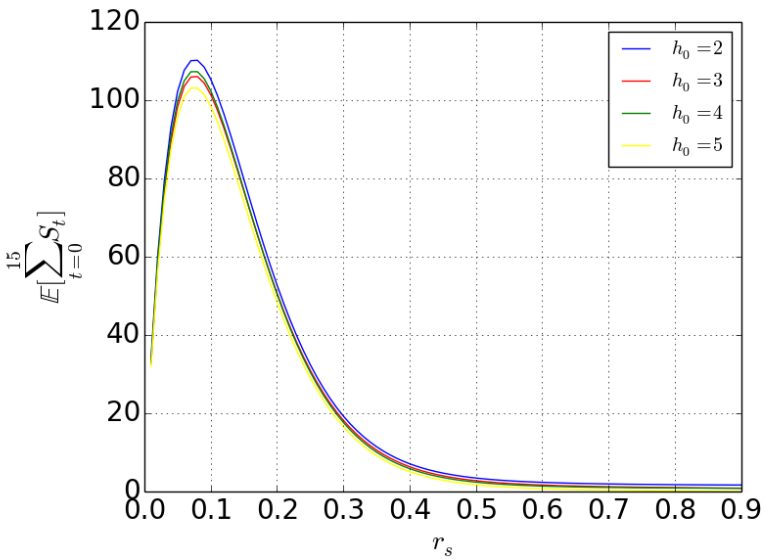
$(N+3)$-type Galton Watson process:

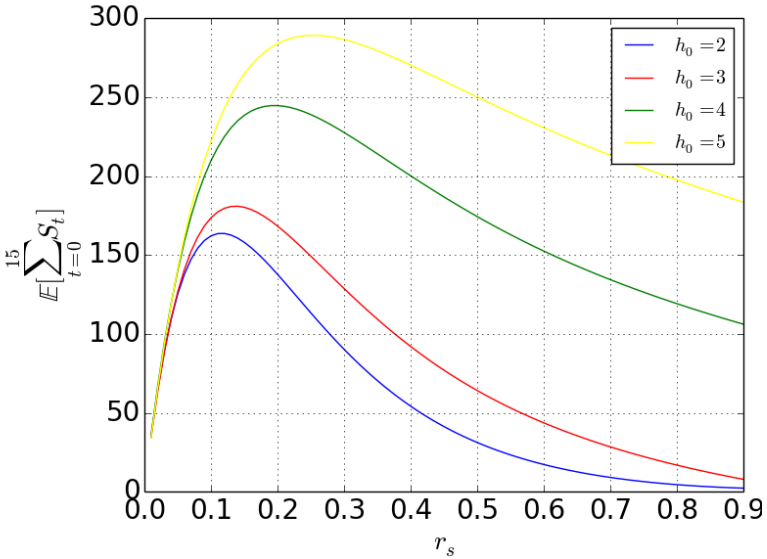
$$\boxed{\mathbf{Z}_t^{(\mathbf{i})} = (Z_{t,0}^{(\mathbf{i})}, \ldots, Z_{t,N+2}^{(\mathbf{i})})}$$

- $0 \leq j \leq N$: $Z_{t,j}^{(\mathbf{i})} = \#$ GC B-cells having Hamming distance j
- $Z_{t,N+1}^{(\mathbf{i})} = \#$ selected B-cells
- $Z_{t,N+2}^{(\mathbf{i})} = \#$ death B-cells

at time t, when the process is initiated in state $\mathbf{i} = (i_0, \ldots, i_N, 0, 0)$

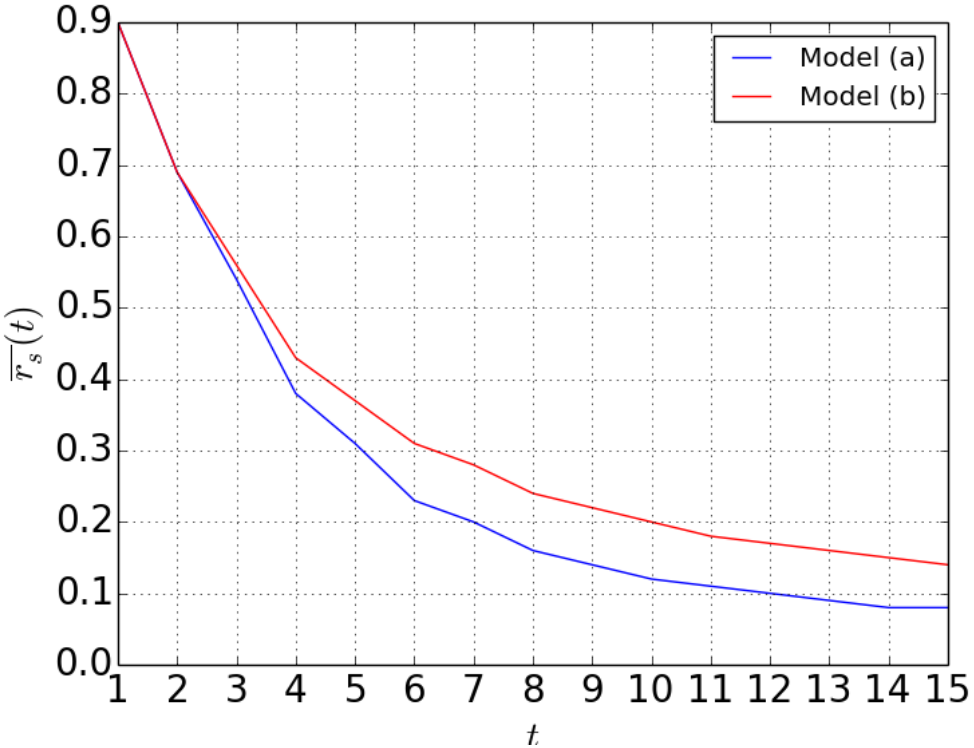
Expected number of selected cells depending on r_s





Expected number of selected B-cells after 15 time steps with mutational model corresponding to matrix $\mathcal{P}$ for model (a) and (b) respectively. $N = 7$, $r_{div} = 0.9$, $r_d = 0.1$, $\overline{h}_s = 3$.

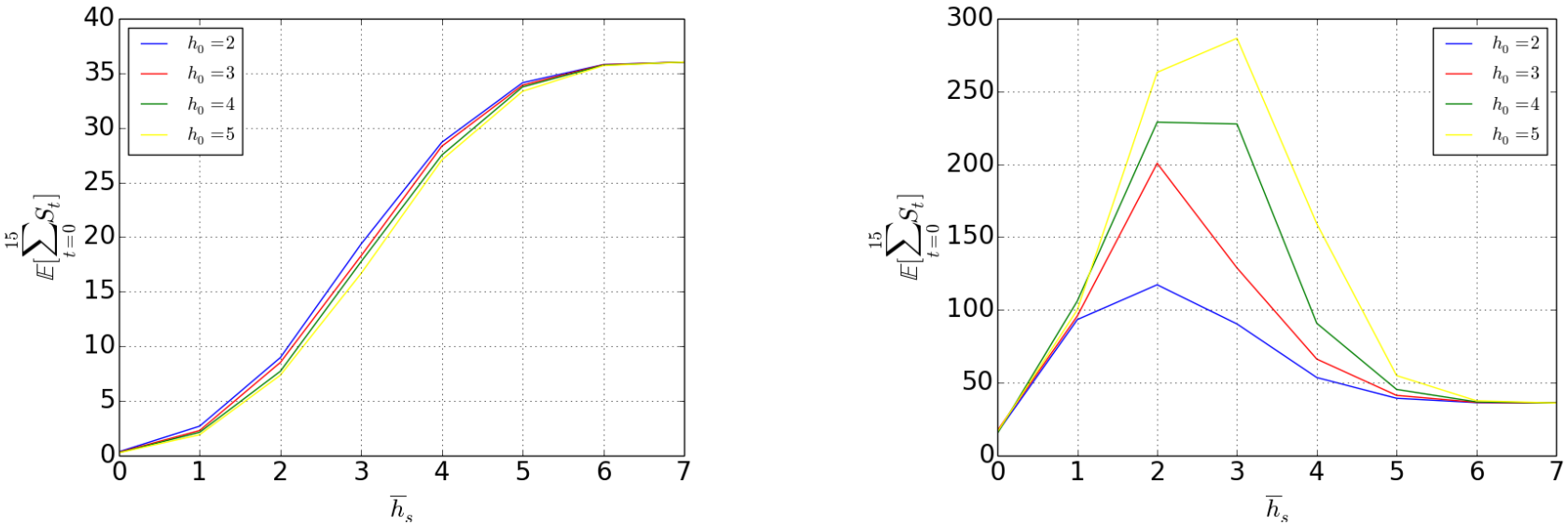
Estimation of the best $r_s(t)$



Estimation of the best choice of r_s depending on t for model (a) and (b) respectively, with mutational model corresponding to matrix $\mathcal{P}$.

$N = 7$, $r_{div} = 0.9$, $r_d = 0.1$, $\overline{h}_s = 3$, $h_0 = 3$.

Expected number of selected cells depending on $\overline{h}_s$



Expected number of selected B-cells after 15 time steps with mutational model corresponding to matrix $\mathcal{P}$ for model (a) and (b) respectively. $N = 7$, $r_{div} = 0.9$, $r_d = 0.1$, $h_0 = 3$, $r_s = 0.3$.

Plan

Conclusions

- Results:
  - Mathematical environment allowing us to introduce and study mutations characteristic of somatic hypermutation
  - It allows to fix our point of view: genetic mutations on DNA or effective mutations on amino acids
  - Introduction of a new kind of branching random walks on graphs
  - Galton-Watson processes with affinity dependent selection
- Future objectives:
  - Mathematical analysis of the model including division, mutation and selection
  - Evaluation of other characteristics of the process (mutation rate, final population size, quality of the final clones, mutational lineage trees, etc.)