

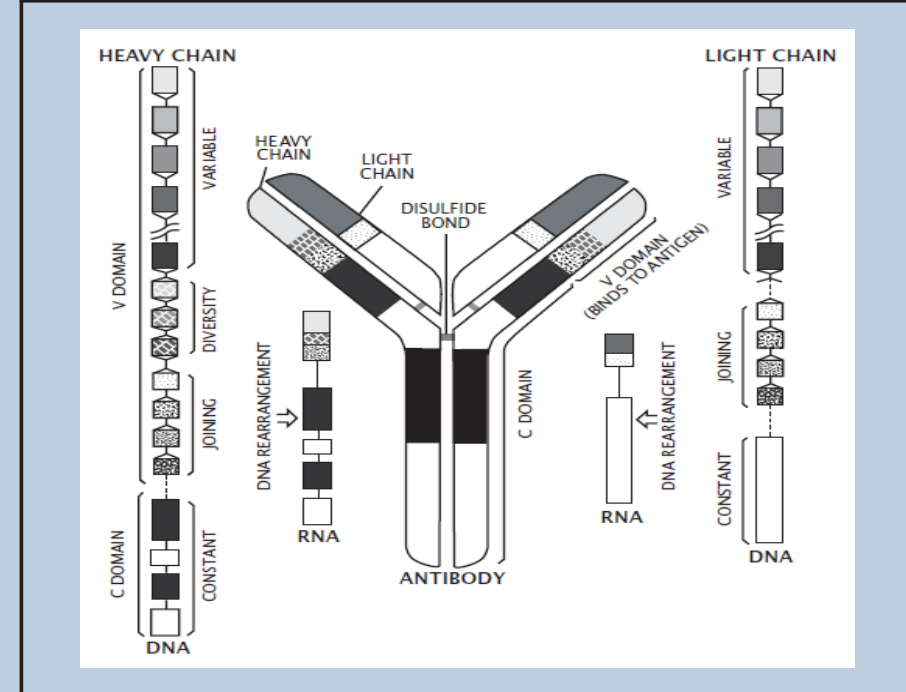
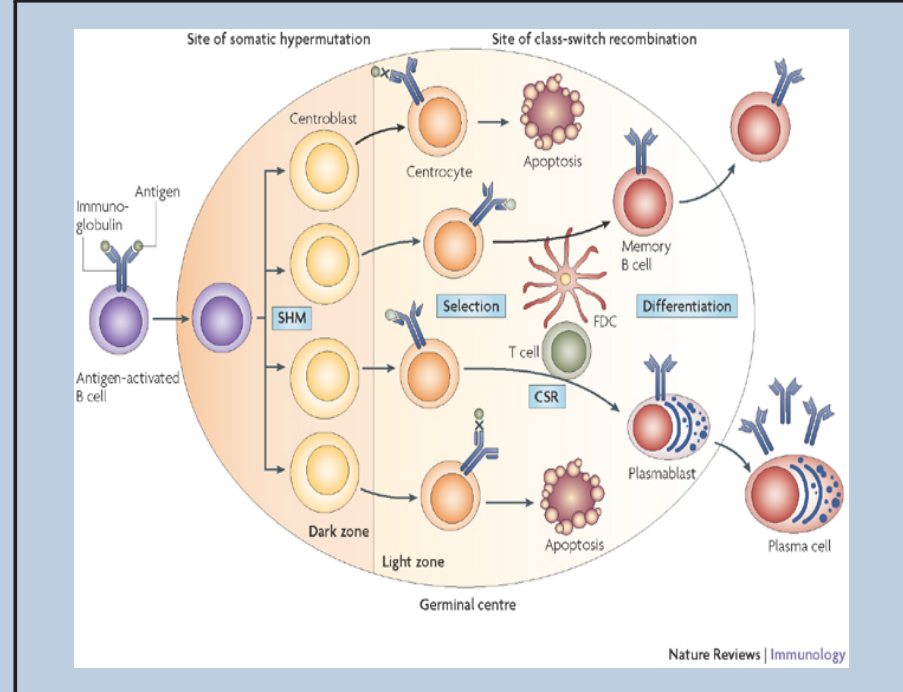
A mathematical model for Somatic Hypermutation



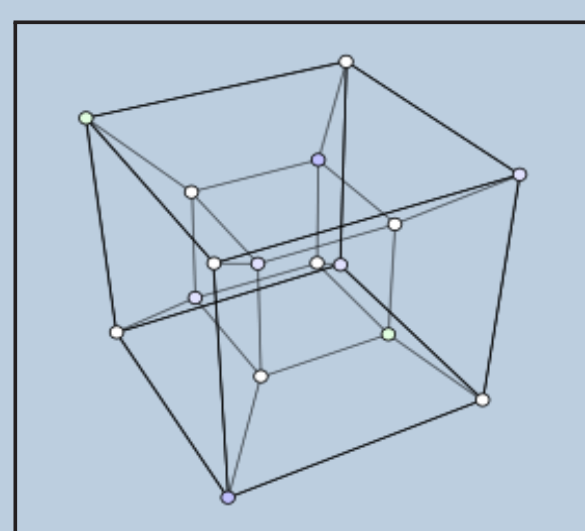
Irene Balelli (LAGA - University Paris 13)
balelli@math.univ-paris13.fr

Biological background

In the **adaptive immune system**, immunity is conferred by **antigen-specific antibodies**. Their production is assured by **B-cells**, that undertake an **evolutionary mutation-selection process** in order to improve their ability to recognize a particular antigen. **Somatic hypermutation** is the basis for the affinity maturation of B-cells. It involves a programmed process of mutation affecting B-cells receptors (BCR).



Modeling SHM



We suppose we are allowed to **classify the amino acids** which determine the chemical properties of both BCR and antigen into **two classes**.

We represent BCR and antigen as two **binary strings** with the same **fixed length N** and we estimate the **affinity** in terms of the **hamming distance**. To define a **mutation rule** means to define a **random walk on the hypercube**.

The Basic Model

Mutation rule: class switch of a randomly chosen amino acid.

$(X_n)_{n \in \mathcal{H}_N}$ is a simple random walk on the hypercube

$$\mathbb{P}(X_n = x_j | X_{n-1} = x_i) = \begin{cases} \frac{1}{N} & \text{if } x_j \sim x_i \\ 0 & \text{otherwise} \end{cases} \rightarrow \mathcal{P}$$

Stationary distribution: $\mu(x) = 2^{-N}, \forall x \in \mathcal{H}_N$

Convergence $\mathcal{P} \xrightarrow{\mathcal{L}} \mu$: if we add N loops at each vertex to break the periodicity. Indeed, \mathcal{H}_N is bipartite.

Eigenvalues: $-1, \frac{-N+2}{N}, \frac{-N+4}{N}, \dots, \frac{N-4}{N}, \frac{N-2}{N}, 1$

The distances process

$(D_n) = H(X_n, \bar{x})$ where H is the Hamming distance. (D_n) is a random walk on $\{0, \dots, N\}$.

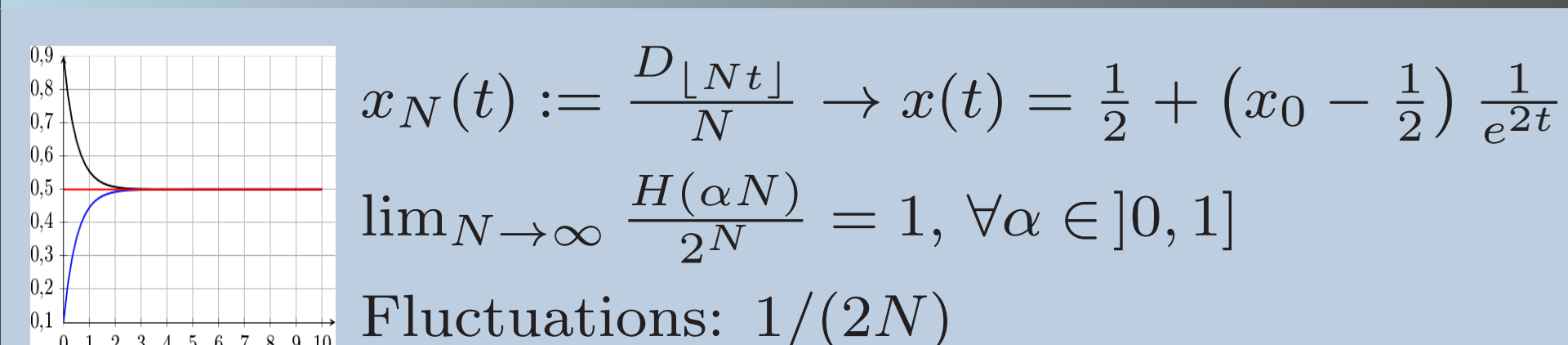
$$\mathbb{P}(D_n = d' | D_{n-1} = d) = \begin{cases} \frac{d}{N} & d' = d - 1 \\ \frac{N-d}{N} & d' = d + 1 \\ 0 & |d' - d| \neq 1 \end{cases} \rightarrow \mathcal{Q}$$

Stationary distribution: $\mathbf{B}(N, \frac{1}{2}) = \binom{N}{d} \frac{1}{2^N}$

Convergence $\mathcal{Q} \xrightarrow{\mathcal{L}} \mathbf{B}(N, \frac{1}{2})$: iff we add loops.

Eigenvalues: the same as \mathcal{P} , with multiplicity 1.

$N \rightarrow \infty$



References

- [1] Chinmoy Dutta, Gopal Pandurangan, Rajmohan Rajaraman & Scott Roche. *Coalescing-branching random walk on graphs*, 2013.
- [2] László Lovász. *Random walks on graphs: a survey*. Combinatorics, Paul erdos is eighty, 2(1):1-46, 1993.

The hitting time

Definition: $H(i, j)$ = average number of steps needed to reach j , starting from i .

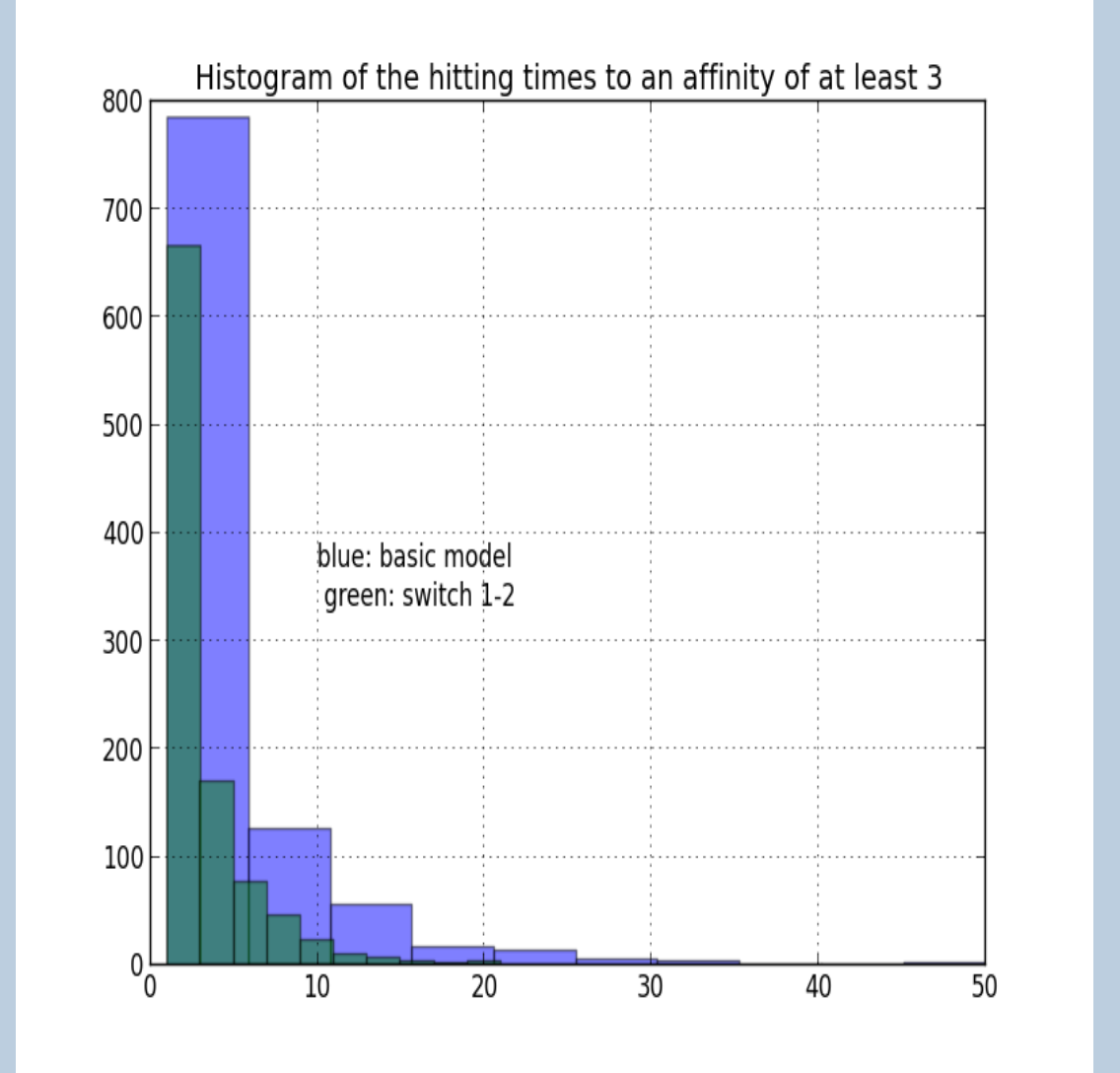
$H(\bar{d})$ = average number of steps to reach \bar{x} , with $H(X_0, \bar{x}) = \bar{d}$.

Biological interpretation: the hitting time represents the expected number of mutations we need to obtain an optimal BCR, given a particular antigen.

Computation (Basic model): $H(i, j) = 2^N \sum_{k=2}^{2^N} \frac{1}{1-\lambda_k} (v_{kj}^2 - v_{ki}v_{kj})$

$$H(\bar{d}) = \sum_{d=0}^{\bar{d}-1} \frac{\sum_{j=1}^{N-1-d} \binom{N}{d+j} + 1}{\binom{N-1}{d}}$$

Switch 1-2, a new mutation rule: we're allowed to switch the class of 1 or 2 length strings, depending on the Hamming distance. The hitting time corresponding to this case is reduced by a factor 2, compared to the basic model.



2-Branching random walk on the N -dimensional hypercube

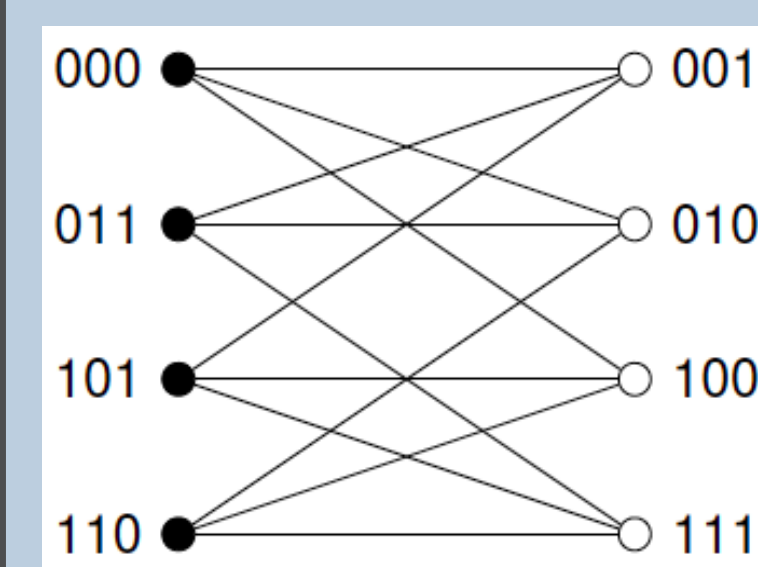
Biological motivation: we want to couple the mutational process with the cellular proliferation.

We give two different definitions of the 2-Branching random walk on \mathcal{H}_N :

Simple: the process starts at an arbitrary node, labelled as active. Then if at time t a node is active, it chooses two of its neighbors, independently and with replacement, to become active at time $t+1$ and he becomes inactive. In this case we are not interested in how many times a node is chosen to become active.

With multiplicity: the process starts with a clone lying on a random node of the hypercube, which represents its configuration. At each time step each clone divides and moves, independently from all the others, in a neighbor vertex. As an immediate consequence, at time t we will have exactly 2^t clones lying on \mathcal{H}_N .

Branching random walk and bipartiteness



Notation: $S_t = \{\text{active nodes at time } t\}$; $N(S_t) = \{\text{neighbors of } S_t\}$

Proposition: Let $G_b(V_1 \sqcup V_2, E)$ be a bipartite graph. If the initial distribution \mathbf{p} is concentrated on V_1 or on V_2 , then: $|S_t| \leq \max_{i=1,2} (|V_i|)$ for all $t \geq 0$

Theorem: Given a k -branching random walk on a finite non-bipartite connected graph $G = (V, E)$, independently from \mathbf{p} we have that: $\mathbb{P}[\exists t > 0, t < \infty \text{ s.t. } S_t = V] > 0$

Partial cover time for the Simple BRW

Preliminary result: For any $N \geq 1$, \mathcal{H}_N is a N -regular $(r, 2^{-r})$ -expander graph, i.e.:

$$\forall r \in \{1, \dots, N\}, \forall S \subset \{0, 1\}^N \text{ s.t. } |S| \leq 2^{N-r} \Rightarrow |N(S)| \geq r|S|$$

Theorem [1]: Given a simple 2-branching random walk on \mathcal{H}_N , there exists a time T such that:

$$T = \mathcal{O}(N) \text{ and } |S_T| \geq 2^{N-r}, \text{ for } r > \frac{N^2 e^{-2} + N - 2}{N e^{-2} + N - 2}$$

BRW with multiplicity

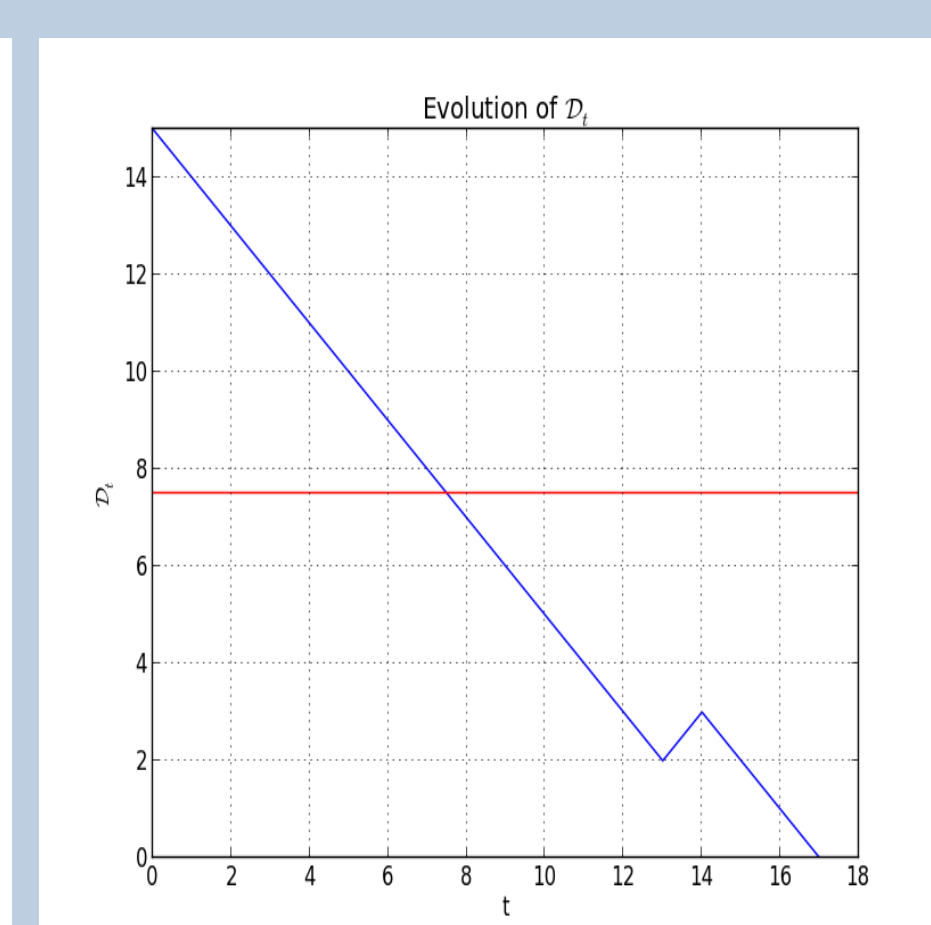
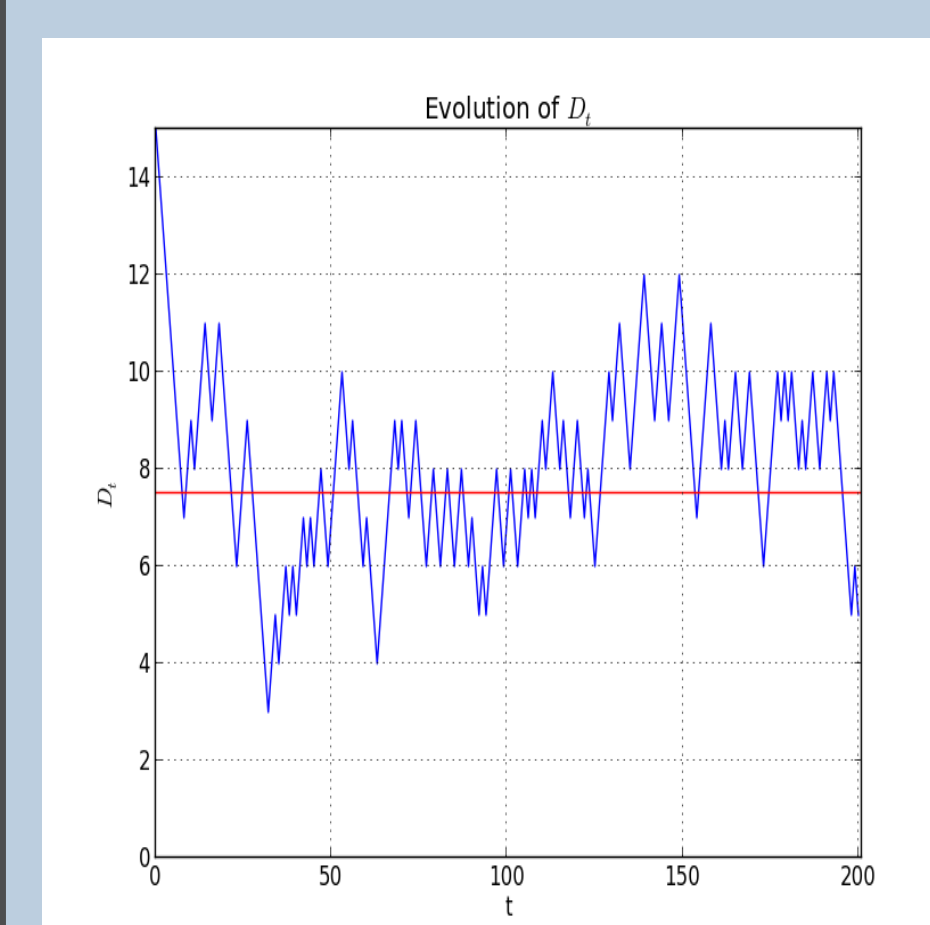
X_t^i = number of particles lying on vertex i at time t . Then we have the following result:

$$\mathbb{P}[X_t^i = s | X_{t-1}^j \forall j \sim i] = \begin{cases} \binom{2n_{t-1}^i}{s} \frac{(N-1)^{2n_{t-1}^i - s}}{N^{2n_{t-1}^i}} & \text{if } s \leq 2n_{t-1}^i \\ 0 & \text{otherwise} \end{cases}, \text{ where } n_{t-1}^i := \sum_{j \sim i} X_{t-1}^j$$

k -BRW on the complete graph on N vertices \mathcal{K}_N :

$$\mathbb{P}[X_t^i = s | X_{t-1}^i = s'] = \begin{cases} \binom{2^t - ks'}{s} \left(\frac{1}{N-1}\right)^s \left(1 - \frac{1}{N-1}\right)^{2^t - ks' - s} & \text{if } s \leq 2^t - ks' \\ 0 & \text{otherwise} \end{cases}$$

Comparing the distances processes: basic model and BRW



Definition: Let \bar{x} be a fixed node in \mathcal{H}_N .

Let $\mathbf{X}_t := (x_1^t, \dots, x_{2^t}^t) \in (\mathcal{H}_N)^{2^t}$ be the 2-branching process.

We define the distance \mathcal{D}_t of \mathbf{X}_t to \bar{x} as:

$$\mathcal{D}_t = \min_{x_i^t \in \mathbf{X}_t} (H(\bar{x}, x_i^t))$$

where $H(\bar{x}, x_i^t)$ is the Hamming distance between \bar{x} and x_i^t .