Simulating aggregates of bivalents in 2n = 40 mouse meiotic spermatocytes through inhomogeneous site percolation processes

Soledad Berríos, Julio Lopez Fenner, Aude Maignan

Simulating aggregates of bivalents in 2n=40 mouse meiotic spermatocytes through inhomogeneous percolation processes

Soledad Berríos and Julio López Fenner [*] and Aude Maignan [†]

S. Berríos
Programa Genética Humana
ICBM, Facultad de Medicina
Universidad de Chile, Santiago, Chile
e-mail: sberrios@med.uchile.cl

J. López Fenner (✉)
Ingeniería Matemática
Facultad de Ingeniería y Ciencias
Universidad de La Frontera, Temuco, Chile
e-mail: julio.lopez@ufrontera.cl

A. Maignan (✉)
*Univ. Grenoble Alpes, CNRS, Grenoble INP**
LJK, 38000 Grenoble, France
** Institute of Engineering Univ. Grenoble Alpes*
e-mail: aude.maignan@univ-grenoble-alpes.fr

Abstract: We show that an inhomogeneous Bernoulli percolation process running upon a dual of a fullerene C_{1200} can be used for representing bivalents attached to the nuclear envelope in mouse *Mus M. Domesticus* 2n=40 meiotic spermatocytes during pachytene. It is shown that the induced clustering generated by overlapping percolation domains correctly reproduces the probability distribution observed in the experiments (data) after fine tuning the parameters.

MSC 2010 subject classifications: Primary 60K35; secondary 60C05, 62P10, 92B05, 92-08..
Keywords and phrases: Inhomogeneous percolation, Fullerene, *Mus M. Domesticus* 2n=40 meiotic spermatocytes..

1. Introduction

The question about randomness in biological processes constitutes a source of many interesting cross-disciplinary discussions among biologists and mathematicians. More often than not, uniform probability distributions seem to constitute the very paradigm of randomness, while the appearance of other probability distributions are often considered to be the result of dynamical rules applied to uniform random variables.

[*]Partially supported by Chilean MINEDUC Grant MECE-SUP 2016-2017.
[†]Partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program *Investissement d'avenir*.
[‡]"This is a pre-print of an article published in Journal Math. Biology. The final authenticated version is available online at: https://doi.org/10.1007/s00285-018-1254-6".

On other side, it has been frequently discussed whether biological models at the scales considered in this work, in the range of 1 to 10 μm, should be better approached from the continuous rather than the discrete side of modeling.

The challenge of producing a mathematical model of randomness for explaining the (apparent) bivalent's association clusters observed in the Meiosis's Pachytene for the house mouse $2n = 40$ - *Mus m. domesticus* spermatocytes was initiated in [3] and continued in [8], in an eminently discrete approach, by considering first surface hexagonal tilings as approximation of the nuclear envelope and next by a discrete regular graph with a finite number of vertices. The reason of this sort of approach is simple: Objects being modeled are distinct entities and the sought probability distributions referred to the (discrete) partitions of a number. Hence, more than absolute positions of points in a sphere, the mathematical/biological constraints for the stochastic distribution of a finite number of indistinguishable objects, related to each other by a neighborhood-dependent condition, lead us naturally to a description using graphs, with satisfactory results [8].

In this work we propose an enhanced discrete model along the lines of the earlier works, but in which the nuclear envelope is now modelled by the dual C'_{1200} of a fullerene, [13], which is a particular planar graph embedded in the sphere. On this graph the objects of interest are percolation domains obtained by a suitable 'Inhomogeneous Percolation Process', which we call P-percolation (see [6]), that are put into correspondence with the units of biological interest (bivalents, or, more precisely, CTC's, see below). Induced clustering generated by overlapping domains can be now explored and analyzed using the model.

Hence, the article is structured as follows: Section 2 contains the biology context, in which a minimal set of notions and working definitions (necessary for understanding the underlying biological problem) are presented. It includes some new derived statistics from the original data set, that were omitted in the previous works, which we use as test frames for our model. The interested reader and the biological audience will find the complete theory with details in the references.

In section 3 we argue against using the theory of random graphs $G(n, p)$, as introduced by Erdös [4], for explaining the clustering probabilities obtained in the dataset. We show that there can be no value for the probability p that adequately fit or reproduces the data.

Section 4 deals with the actual construction of the graph used to model the nuclear envelope. It will turn out to be the dual of fullerene C_{1200} (subsection 4.1). This graph appears naturally by imposing size and connectivity constraints upon the number of vertices and planarity of the graph, that correlates well with both, the size and form of the biological objects being modeled. The construction of random domains of vertices via P-percolation, (called '*ranches*') that represent the CTC's (subsections 4.2 and 4.3), coupled with a simple Laplace dynamics for the parallel evolution of the positions of vertices with attributes (subsection 4.4), produce a structure with which the probability distribution of the association clusters is correctly reproduced. Finally, subsection 4.5 presents the results from the simulation process.

Section 5 hands the word back to the biologist by presenting a biological discussion and interpretation of the mathematical and simulation results obtained.

2. The Biology Behind

Meiosis is an extraordinary process that produces haploids and genetically diverse gametes in organisms of sexual reproduction and differs significantly to the more widely known process of Mitosis. It is characterized by a plethora of stages, of which one of the conspicuous, remarkable ones is called the *pachytene in early prophase I*, which we are interested in. At pachytene, the homologous chromosomes synapse along (by or through) a proteinacious structure called *synaptonemal complex* (SC), which enables recombination between them, a process that is known to produce genetic variation. Chromosomes in synapse are called *bivalents* and describe arcs 'floating' inside the nucleous, bound to the nuclear envelope by both their extremes, called *telomeres*.

Unions or associations among bivalents, mediated by overlapping domains of their own constitutive heterochromatin, are frequently observed during this stage.

Enter the house mouse: $2n = 40$ *Mus m. domesticus*. This species, which is widely represented in Europe, exhibits all their chromosomes (save for the X and Y), to be of similar morphology and size, hence making them particularly well suited for establishing a reference biological subject with which it should be possible to contrast mathematical models. In our case, we will only consider the 19 bivalents of $2n = 40$ *Mus m. domesticus* (hence leaving the X and Y out) and will consider them to be indistinguishable units.

The *centromere - telomere - complex* (CTC) is a structure of the short arm of each one of the 19 autosomal bivalents and - under appropriate staining techniques, as immunocytochemical staining - several structures of them can be brought to light. So, for example the *Synaptonemal Complexes* (SC's) and also the *constitutive pericentromeric heterochromatin* domains (CPCH's).

Spreads are obtained as the result of removing the nuclear envelope and projecting the spermatocyte nucleus's to a flat 2D surface. A typical spread is presented in Figure 1. a), in which the pericentromeric heterochromatin domains (CPCH's) are stained red, while the Synaptonemal Complexes (SC's) are stained green.

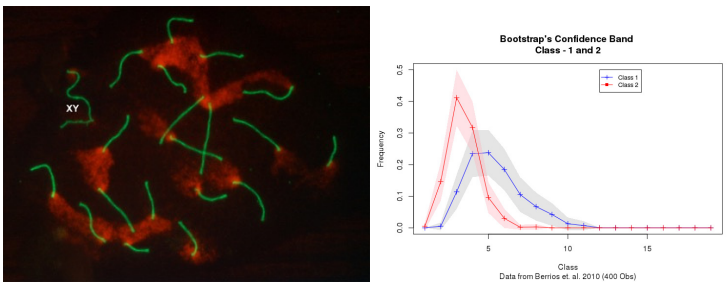

FIG 1. *a) A representative pachytene spermatocyte spread of* $2n = 40$ *Mus domesticus mice treated by immunocytochemical techniques. In red the pericentromeric heterochromatin domains, in green the synaptonemal complexes. The (sex) bivalent XY is indicated. b) Observed* 1^{st}*-Class and* 2^{nd}*-Class frequencies with bootstrap variation*

Notice the apparent clustering induced by the overlapping domains of CPCH. These clusters will be interpreted as domains in which the corresponding CTC's are in asso-

ciation and, in order to describe the frequencies of their appearance in the data, they will be associated 1:1 to a partition of the number 19 (see [3]):

The 1^{st}-*Class* (or simply '*Class*') of a spread is a characterization of the induced partition given by the number of bivalents present in the biggest association cluster observed in it. The 2^{nd}-*Class* is the size of the second biggest subcluster observed, and so forth.

So, for example, if a given spread exhibits one association cluster of size 4, three of size 2, and all others in singletons, then the associated partition of 19 is written as $19 = 4 + 2 + 2 + 2 + 1 + \cdots + 1$ and we say that the spread belongs to 1^{st}-Class 4, 2^{nd}-*Class* 2, etc.

In order to make this article as self-contained as possible, we reprocessed the original data from [3] and reproduced the empiric 1^{st}-*Class* and 2^{nd}-*Class* distribution, which we completed by adding the intrinsic variation in each class, computed by simple bootstrap.

The data is given in Table 1, results are reported in Figure 1, b), where the shaded areas represent two times (the square root of) the corresponding bootstrap variances.

1^{st} - and 2^{nd} - Class Elementary Statistics				
Size	1^{st}-Class *Freq* (%)	$\sqrt{Var_1}$	2^{nd}- Class *Freq* (%)	$\sqrt{Var_2}$
1	–	–	0.48	0.006
2	0.50	0.006	14.61	0.030
3	11.25	0.027	41.18	0.044
4	23.25	0.037	31.74	0.0413
5	23.50	0.036	9.55	0.0241
6	18.50	0.033	2.99	0.0153
7	10.25	0.028	0.19	0.004
8	6.50	0.022	0.27	0.004
9	4.25	0.018	0	0
10	1.25	0.01	0	0
11	0.75	0.01	0	0
12	0	0	0	0
⋮	⋮	⋮		
19	0	0	0	0

TABLE 1

1^{st} - *and* 2^{nd} - *Class Elementary Statistics (see Fig. 1).*

For each spread in the data set, the corresponding partition was determined and the frequencies of their appearance computed. Figure 2. a) depicts the full set of frequencies, where the x-axis denotes the set of all partitions ordered canonically, i.e., the first partition is $19 = 1+1+1+\cdots+1$, the second one is $19 = 2+1+1+\cdots+1$, until partition number 489, which is $19 = 18 + 1$ and the last one, $19 = 19$, the 490th partition. One cannot help but notice a suggestive fractal-like form for the distribution showing up. Figure 2, b) shows the observed distribution for 1st-Class 4.

Remark: The reader will notice a small abuse in notation while using the letter n: Each time we refer to the mouse under study, we shall write $2n = 40$, since this is standard notation in biology: $2n = 40$ *Mus M. Domesticus*. In contrast, whenever we talk about placing the 19 autosomal bivalents (not considering the X and Y) upon the

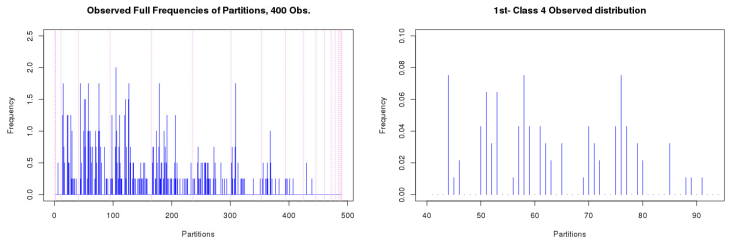

FIG 2. *a) The full set of 400 observations, with their frequencies. b) The subset of 1st-Class 4*

nuclear envelope and determining the corresponding association clusters, we shall put $n = 19$, since it is the partitions of 19 we will be referring to.

3. Random Graphs

In this section we show that the observed partitions do not follow the distribution obtained when applying the theory of random graphs $G(n, p)$.

Indeed, the exact probability of $G(n, p)$ being connected, which we denote by $f(n)$, can be computed recursively by means of [5]

$$\begin{aligned} f(1) &= 1 \\ f(n) &= 1 - \sum_{i=1}^{n-1} f(i) \left(\begin{array}{c} n-1 \\ i-1 \end{array} \right) (1-p)^{i(n-i)},\ n > 1. \end{aligned} \tag{3.1}$$

The probability $P(r)$ for a random graph composed of n nodes to be split into k components of size $r = (r_1, r_2, ...r_k)$ is known to be [1]:

$$P(r) = \frac{n!}{\Pi_m d_m! \Pi_i r_i!} \Pi_l f(r_l) \Pi_{i<j} (1-p)^{r_i r_j},$$

in which d_m is the number of containers of size m, which essentially is a combinatoric result obtained in the determination of the so called 'Bell's numbers' (see [2], [9]).

So we ask ourselves whether there is a value p, for which the observed partition frequencies corresponds to the theoretical frequencies of the cluster distributions of $G(19, p)$.

To answer this we take, for example, 1st-Class 4, of which there are 93 observations in the dataset, corresponding to 28 different partitions.

Under this model, the probability of partition $r = (4, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ in $G(19, p)$ is given by

$$\begin{aligned} P(r) = \Big[& a - a\,(1-p)^3 - b\,p(1-p)^4 \\ & - b\left(1 - (1-p)^2 - 2\,p(1-p)^2\right)(1-p)^3 \Big] p^3\,(1-p)^{162}, \end{aligned}$$

in which the coefficients are, respectively, $a = 290990700$, $b = 872972100$.

Solving for p such that $\sum_{r\in Cl(4)}(P(r)) = 93/400$, the observed frequency for 1st Class 4, yields three possible values: $p = -.121, p = -0.0125, p = 1.807$. Hence we conclude that no p will ever furnish a correct approximation for the observed distribution using $G(n = 19, p)$.

For this reason, we now take the nuclear architecture of the spermatocytes into consideration and propose a random partition model that suits better the experimental data.

4. Model Construction

We take first into account the scales of the biological units considered in our study (as reported in [3]): It is known that approximately $1\mu m^2$ portion of the CTC's surface adheres to the internal face of the nuclear envelope, whose size - in turn - range between $10\mu m - 15\mu m$ in diameter. This means that, for example, in a nucleus of $14\mu m$ diameter, the total surface of the nuclear envelope measures approximately $615\mu m^2$, hence accepting a tiling containing 615 cells of $1\mu m^2$ each, while in a nucleus of $13\mu m$ diameter, a similar tiling would contain 530 cells of $1\mu m^2$.

We propose a discrete structure (a connected undirected graph $G = (V, E)$, with V a set of vertices (or nodes) and E a set of edges, $E \subseteq V \times V$) for modeling the nuclear envelope. The difference with our previous model (reported in [8]) is that now the bivalent's structures attached to the envelope are spatially extended structures which will be represented as attributes to the vertices, and we replace the assumption of a rigid CTC by a random structure representing two of the highlighted components in the spreads: the Constitutive Pericentromeric Heterochromatine (CPCH) and the Synaptonemal Complex (SC).

To capture the biological constraints posed above, we search for a graph G satisfying (at least) the conditions:

C1. Regular and maximally connected.
C2. Embedded on a sphere.
C3. $|V| \sim 600$.

Condition C1 represents a biological constraint, since the SC's are spacially extended objects placed upon a surface, which imposes naturally a six-degree of regularity upon the nuclear architecture, considered as a graph. To respect the observed sizes of the nuclei, we require C3 and, while a six-regular embedding of a surface to a sphere is certainly not possible, the constraint of being maximally connected yields a candidate: The dual fullerene.

4.1. Dual Fullerene

Embedding a graph in the plane is equivalent to embedding it on the sphere, hence we use planar graph theory [12]. We need the following results, which we state without proof:

Theorem 4.1. *If G is a planar graph of $N \geq 3$ vertices and e edges, then $e \leq 3N - 6$.*

Corollary 4.1. *Every planar graph of $N \geq 3$ vertices contains a vertex of degree at most 5.*

A graph satisfying conditions C1 and C2 above with exactly $N - 12$ vertices of degree 6 and 12 vertices of degree 5 can be constructed in a straightforward manner:

Indeed, from the previous theorem and corollary, $e \leq 3N - 6$. If we consider $N - 12$ vertices of degree 6 and 12 vertices of degree 5, then $e = \frac{1}{2}(6(N-12) + 5 \times 12)$, which is exactly $3N - 6$.

A *Fullerene* is a cubic planar graph having all faces 5- or 6-cycles, i.e., a 3-regular planar graph such that all the face's sizes are pentagons or hexagons [13]. Hence the graph we consider is a fullerene dual and is constructed as follows: Start with a node, attach to it a ring made of 6 connected vertices, then a second ring of 12 vertices, etc ..., until attaching the i^{th} ring with $6i$ vertices, see Figure 3.

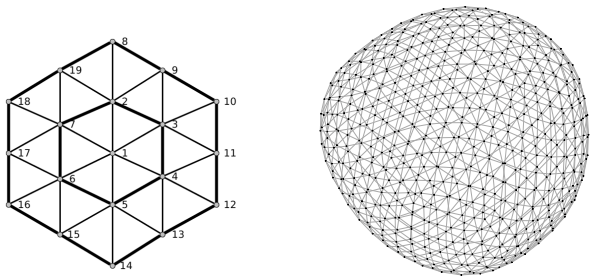

FIG 3. *A local planar view of the (almost) 6-regular sphere and the resulting graph C'_{1200}.*

After having built this structure with i rings, which we call a semi-sphere, the graph has exactly $N_i = 1+3i(i+1)$ vertices, all of them of degree 6, save those that are located at the exterior ring, which are of degree 3 or 4.

Construct now a second half-sphere with $i - 1$ rings and connect with the first one as to leave only 12 vertices of degree 5, all other being of degree 6.

The total number of vertices is therefore $N_i + N_{i-1} = 1+3i(i+1)+1+3i(i-1) = 2+6i^2$ and all vertices of degree 5 are located on the i^{th} ring of the first semi-sphere.

For $i = 10$ we finally obtain a graph of 602 vertices, which matches our expectations for the size of the nucleus (see Figure 3). Of them, 590 vertices are of degree 6 and the rest (twelve) are of degree 5. It is C'_{1200}, the *Dual of the Fullerene* C_{1200} of 1200 vertices. The presence of these twelve degree 5 vertices will prove to be non significant for the computations/simulations, in the sense that without loss of accuracy, the planar graph can be assumed to be plain 6-regular.

4.2. Random Chromatin Neighborhoods

Now we turn to the CTC, which will be characterized by two of their main constituents: The SC and the CPCH. It is a 3-d structure that attaches to the nuclear envelope. Hence the contact zone will be described by the position of the SC, i.e., a vertex in C'_{1200}, and

the positions of the surrounding CPCH, described by vertices in the neighborhood of the SC.

Let $G = (V, E)$ be a graph, v an arbitrary but otherwise fixed vertex. Let $\mathcal{N}_G(v)$ be the set of neighbors of v in G, i.e., the set of vertices that are connected to v by an edge in E. For $v, w \in V$, let $d_G(v, w)$ the distance between them to be defined as the length of the shortest path connecting them.

For each vertex v we assign a state: $state(v)$, which belongs to the set $\{\neg Ex, S, \neg S\}$, in which $\neg Ex$ stands for "non explored", S for "selected" and $\neg S$ for "not selected".

Let $P = (P_1, P_2, \ldots, P_\kappa) \in [0, 1]^\kappa$ be a finite length vector of probabilities $0 \leq P_i \leq 1$, $1 \leq i \leq \kappa$.

In analogy to [6], we construct an *Inhomogeneous Bernoulli Percolation Process* upon C'_{1200}, also called *P- percolation process*, as a site selection process according to the following algorithm:

P-percolation

- Input: $G \leftarrow C'_{1200}$, $v \in V$, $P = (P_1, P_2, \ldots P_\kappa)$.
- Put vertices in $V \setminus \{v\}$ to be in state $\neg Ex$, assign $state(v) \leftarrow S$ and set $V_0^P(v) = \{v\}$.
- for j from 1 to κ do:

  $\forall k \in V_{j-1}^P(v)$, $\forall l \in \mathcal{N}_G(k)$:

  - if $state(l) == \neg Ex$ then:
    - Put $X \sim B(1, P_j)$, the Bernoulli random variable with probability P_j;
    - If $X == 1$ then add l to $V_j^P(v)$ and assign $state(l) \leftarrow S$.
  - else $state(l) \leftarrow \neg S$.
- Ouput: $\cup_{j=0}^{\kappa} V_j^P(v)$.

Definition 4.1. *Let $P = (P_1, P_2, \ldots, P_\kappa) \in [0, 1]^\kappa$. Let $v \in V$ be a given (fixed) vertex of G. The* **random chromatin neighborhood** *of v, denoted by $ranch^P(v) \subseteq V$, is the set of vertices in state S resulting from a P-percolation processes around v.*

As a result, $ranch^P(v) \leftarrow \cup_{j=0}^{\kappa} V_j^P(v) \subseteq V$ is obtained layer by layer by adding successive rings $V_j^P(v)$ of vertices centered at v in which the selection follows independent Bernoulli trials, hence building a random neighborhood that represents a natural probabilistic structure with which bivalents (CTC's) will be modeled. Clearly, $ranch^P(v) \equiv V(v)$ iff all the values of P are 1, otherwise it is a subset of nodes of $V(v)$.

We assume that the CTC will be attached to the nuclear envelope at the points given by the vertices in $ranch^P(v)$, $v \in V$. v denotes the position of the SC, and the set $ranch^P(v) \setminus \{v\}$ the CPCH, see Figure 4. ● represents the position of a SC and • corresponds to the intersection between the chromatine and the nucleus surface discretization.

From the fact that – by construction – the sets $V_j^P(v)$, $j = 1, 2; \ldots$ are mutually

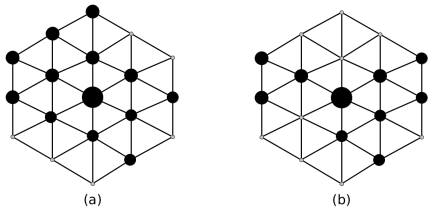

FIG 4. *(a) Chromatin neigborhood created with* $P = (1, 0.5)$. *(b) Chromatin neigborhood created with* $P = (0.8, 0.6)$.

disjoint, the number of elements of the ranch is given by

$$|ranch^P(v)| = \sum_{j=0}^{\kappa} |V_j^P(v)|.$$

We restrict ourselves to the graph $G = (V, E)$ with $|V| = N = 602$ nodes, as introduced in section 4.1, see also Figure 3, and the vector P a given κ-dimensional vector of probabilities.

Let $P = (P_1, P_2, \dots P_\kappa)$. We shall derive an expression for the first two moments of the probability distribution of the size of the $ranch^P(v)$ of a given vertex $v \in V$: $\mathbb{E}(|ranch^P(v)|)$ and $\sigma^2(|ranch^P(v)|)$. For this, we assign to the central position of the local planar graph in which v is contained the (local) coordinate (0,0). In this coordinate system, vertices located at distance 1 will be given the positions $(1, 0), (1, 1), \dots (1, 5)$ counting clockwise, as in figure 5, and so forth for vertices located at greater distances: for $i \geq 1$, $j \in \{0, 1, \dots 6i - 1\}$, the vertex (i, j) will be the j-th vertex counting clockwise from the central vertex located at distance i from the center.

Denote by $Z(i, j)$ the Bernoulli random variable $Z(i, j) \sim B(1, p_{i,j})$, $0 \leq p_{i,j} \leq 1$ that takes the value 1 with probability $p_{i,j}$ if the vertex (i, j) belongs to the $ranch^P$, 0 otherwise. This means that (i, j) is reached from the central vertex by the P-percolation described above, i.e., by a percolation path of length i.

Notice that at each level $i = 1, 2, \dots, \kappa$, once the preceding level has been determined by the P percolation algorithm, the set of reachable vertices depends strongly upon the previous level. In terms of the random variables $Z(i, j)$, this means that $Z(i, j)$ depends always upon some $Z(i - 1, l)$ that lies below, for some l. The dependency/independency of the Z variables is illustrated in Table 2. Due to the particular construction and symmetry of the local architecture of the graph, being 6-regular, variables $Z(i, 0)$ and $Z(i, i \cdot k)$, for $k \in \{1, 2, \dots 5\}$, are independent, identically distributed Bernoulli random variables with parameter $\pi_i := \prod_{j=1}^{i} P_j$.

As for the probability $p_{i,j}$ of $Z(i, j) \sim B(1, p_{i,j})$ for $j \neq i \cdot k$, $k \in \{1, 2, \dots 5\}$, it can be readily seen that it satisfies the following finite difference equation

$$p_{i,j} = P_i \left[p_{i-1,j-1} + p_{i-1,j} - p_{i-1,j-1} \cdot p_{i-1,j} \right], \qquad 1 \leq i \leq \kappa. \tag{4.1}$$

This equation follows from the fact that, by the graph architecture, nodes at distance i of the central node v can be reached either by a unique path of length i, which is the

case for nodes at $(i,0), (i,i), \ldots, (i,5i)$ or by the two sole adjacent paths reaching (i,j): one that pass through the node situated at $(i-1,j)$, the other one through $(i-1,j-1)$, so that Eqn. (4.1) is nothing else than the probability of the union.

Theorem 4.2. *Let $v = (0,0)$ be a given node of the graph G, $P = (P_1, P_2)$ Then the expected size and variance of the size of $ranch^{P_1,P_2}(v)$ is given by*

$$\mu(|ranch^P(v)|) = 1 + 6P_1 + 6P_1P_2(3 - P_1) \tag{4.2}$$

$$\begin{aligned}\sigma^2(|ranch^P(v)|) = 6P_1(1-P_1) + 6P_1P_2\left(9 - 11P_1 + 4P_1^2\right) \\ + 6P_1P_2^2(6 - 19P_1 + 14P_1^2 - 3P_1^3).\end{aligned} \tag{4.3}$$

Proof. The mean size and variance of $|ranch^P(v)|$ will be given by the number of vertices in the neighborhood of $v = (0,0)$ which are selected by the procedure.

With our notation, (see figure 5)

$$|ranch^P(v)| = 1 + \sum_{i=1}^{2}\sum_{j=0}^{6i-1} Z(i,j),$$

from which

$$\begin{aligned}\mathbb{E}(|ranch^P(v)|) &= 1 + \sum_{i=1}^{2}\sum_{j=0}^{6i-1} p_{i,j} \\ &= 1 + 6P_1 + 6P_1P_2(3 - P_1)\end{aligned}$$

because all of them are Bernoulli random variables, which is the easy part and uses equation (4.1).

The variance, though, is more involved, as the random variables are not mutually independent, save those of the first ring $Z(1,0), \ldots Z(1,5)$.

Variables $Z(1,i) \sim B(1,P_1)$, $i = 0,1,\ldots 5$, are independent identically distributed Bernoulli random variables with probability P_1. It is immediate to recognize that $Z(2,0)$, $Z(2,2), \ldots Z(2,10)$ are all i.i.d $\sim B(1,P_1P_2)$. From equation (4.1), variables $Z(2,1)$, $Z(2,3), \ldots Z(2,11)$ are all $B(1,P_1P_2(2-P_1))$. But they are not all independent among them, nor among the variables directly below following any path of length 2 to the central vertex $(0,0)$:

	Depends on	Independent of
$Z(2,0)$	$Z(2,11), Z(2,1)$	$Z(2,2) - -Z(2,10)$
$Z(2,1)$	$Z(2,11), Z(2,0), Z(2,2), Z(2,3)$	$Z(2,4) - -Z(2,10)$
$\vdots$	$\vdots$	$\vdots$

TABLE 2
Z dependencies at level 2

This is relevant for the computation of the variance, since then the joint probability distribution must be determined. Since

$$\sigma^2(ranch^{P_1,P_2}(v)) = \mathbb{E}\left(|ranch^{P_1,P_2}(v)|^2\right) - \mathbb{E}^2(|ranch^{P_1,P_2}(v)|),$$

we expand the terms involved and get, for example

$$\left(1 + \sum_{i=1}^{2}\sum_{j=0}^{6i-1} Z(i,j)\right)^2 = 1 + 2\sum_{j=0}^{5} Z(1,j) + 2\sum_{j=0}^{11} Z(2,j) + 2\sum_{k=0}^{5}\sum_{l=0}^{11} Z(1,k)\cdot Z(2,l) + \sum_{k=0}^{5}\sum_{l=0}^{5} Z(1,k)\cdot Z(1,l) + \sum_{k=0}^{11}\sum_{l=0}^{11} Z(2,k)\cdot Z(2,l).$$

Also,

$$\mathbb{E}^2(|ranch^{P_1,P_2}(v)|) = \mathbb{E}^2\left(1 + \sum_{j=0}^{5} Z(1,j) + \sum_{j=0}^{11} Z(2,j)\right) = (1 + 6P_1 + 6P_1P_2(3 - P_1))^2,$$

so that

$$\sigma^2(|ranch^{P_1,P_2}|) = 1 + 2\cdot 6P_1 + 2(6P_1P_2 + 6P_1P_2(2 - P_1)) + (6P_1 + 30P_1^2) + 2\sum_{k=0}^{5}\sum_{l=0}^{11}\mathbb{E}\left(Z(1,k)\cdot Z(2,l)\right) + \sum_{k=0}^{11}\sum_{l=0}^{11}\mathbb{E}\left(Z(2,k)\cdot Z(2,l)\right) - (1 + 6P_1 + 6P_1P_2(3 - P_1))^2.$$

It remains only to compute the terms $\sum_{k=0}^{5}\sum_{l=0}^{11}\mathbb{E}\left(Z(1,k)\cdot Z(2,l)\right)$ for the interactions between level 1 and 2, and $\sum_{k=0}^{11}\sum_{l=0}^{11}\mathbb{E}\left(Z(2,k)\cdot Z(2,l)\right)$ for level 2, which is done by using the symmetry of the local graph structure (6-regularity) and the fact that the product of Bernoulli random variables is again Bernoulli.

Thus, for example, by conditioning upon the random variables at the level below, we determine that $Z(1,0)\cdot Z(2,0) \sim B(1, P_1P_2)$ and it has the same distribution as $Z(1,0)\cdot Z(2,1)$ or $Z(1,0)\cdot Z(2,11)$ and that $Z(1,0)$ is independent of $Z(2,2), Z(2,3), \ldots Z(2,10)$.

Hence the products yield either a $B(1, P_1^2P_2)$ or $B(1, P_1^2P_2(2 - P_1))$ distribution, so that

$$\sum_{j=0}^{11}\mathbb{E}\left(Z(1,0)\cdot Z(2,j)\right) = 3P_1P_2 + 5P_1^2P_2 + 4P_1^2P_2(2 - P_1).$$

Since there are six terms of this kind, we conclude that

$$\sum_{k=0}^{5}\sum_{l=0}^{11}\mathbb{E}\left(Z(1,k)\cdot Z(2,l)\right) = 6(3P_1P_2 + 5P_1^2P_2 + 4P_1^2P_2(2 - P_1)) = 6P_1P_2(3 + 13P_1 - 4P_1^2).$$

In the similar lines of reasoning, the products $Z(2,i) \cdot Z(2,j)$ are determined to be Bernoulli, with probabilities $\pi_2 := P_1P_2$, $\tilde{\pi_2} := \pi_2(2-P_1)$, $a := P_1P_2^2 = \pi_2 \cdot P_2$, $b := P_1^2P_2^2 = \pi_2^2$, $c := P_1^2P_2^2(2-P_1) = \pi_2 \cdot \tilde{\pi_2}$, $d := P_1P_2^2(1+P_1-P_1^2) = a \cdot (1+P_1-P_1^2)$, and $e := P_1^2P_2^2(2-P_1)^2 = \tilde{\pi_2}^2$. We explain a few of these expressions, the rest follows from symmetry considerations:

- $Z(2,0) \sim B(1, P_1P_2) \Rightarrow Z(2,0)^2 \sim B(1, P_1P_2) = B(1, \pi_2)$.
- $Z(2,1) \sim B(1, P_1P_2(2-P_1)) \Rightarrow Z(2,1)^2 \sim B(1, P_1P_2(2-P_1)) = B(1, \tilde{\pi_2})$,
- $Z(2,0) \cdot Z(2,1) \sim Z(2,0) \cdot Z(2,11) \sim B(1, P_1P_2^2) = B(1,a)$, which is deduced by conditioning upon $Z(1,0)$ and $Z(1,1)$, see below.
- $Z(2,0) \cdot Z(2,2) \sim B(1, P_1^2P_2^2) = B(1,b)$ since they are independent,
- $Z(2,0) \cdot Z(2,3) \sim B(1, P_1^2P_2^2(2-P_1)) = B(1,c)$ by independence,
- $Z(2,1) \cdot Z(2,2) \sim B(1, P_1P_2^2) = B(1,a)$ by symmetry,
- $Z(2,1) \cdot Z(2,3) \sim B(1, P_1P_2^2(1+P_1-P_1^2)) = B(1,d)$, which is obtained by conditioning upon $Z(1,0)$, $Z(1,1)$ and $Z(1,2)$,
- $Z(2,1) \cdot Z(2,4) \sim Z(2,0) \cdot Z(2,3) \sim B(1, P_1^2P_2^2(2-P_1)) = B(1,c)$, by rotating the figure,
- $Z(2,1) \cdot Z(2,5) \sim B(1, P_1^2P_2^2(2-P_1)^2) = B(1,d)$ since they are independent, etc,
- The other products follows by symmetry, i.e., by rotating the figure.

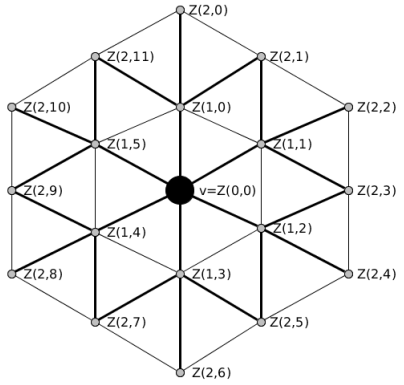

FIG 5. *Reachable points from v with* $P = (P_1, P_2)$

In order to see why – for example – $Z(2,0) \cdot Z(2,1) \sim B(1, P_1P_2^2)$, it suffices to recall that once a vertex, say w, at a given level i has been selected by the percolation process, all vertices at level $i+1$ that can be reached from it are available for the next Bernoulli selection $B(1, P_{i+1})$.

Hence, the joint probability $P(Z(2,0) = 1 \wedge Z(2,1) = 1)$ is determined upon conditioning with respect to variables $Z(1,0)$ and $Z(1,1)$: Let $Y_0 = Z(2,0)$, $Y_1 = Z(2,1)$,

$X_0 = Z(1,0)$ and $X_1 = Z(1,1)$. Then

$$\begin{aligned} P(Y_0 = 1 \wedge Y_1 = 1) &= \sum_{i,j\in\{0,1\}} P(Y_0 = 1 \wedge Y_1 = 1 | X_0 = i \wedge X_1 = j) \cdot P(X_0 = i \wedge X_1 = j) \\ &= P_2^2 \cdot P(X_0 = 1 \wedge X_1 = 1) + P_2^2 \cdot P(X_0 = 1 \wedge X_1 = 0) \\ &+ 0 \cdot P(X_0 = 0 \wedge X_1 = 1) + 0 \cdot P(X_0 = 0 \wedge X_1 = 0) \\ &= P_2^2 \left(P_1^2 + P_1(1 - P_1)\right) = P_1 P_2^2. \end{aligned}$$

If we denote by $M = (M_{i,j})$ the matrix with entries $m_{i,j} = \mathbb{E}(Z(2, i-1) \cdot Z(2, j-1))$, we obtain

$$M = \begin{bmatrix} \pi_2 & a & b & c & b & c & b & c & b & c & b & a \\ & \tilde{\pi_2} & a & d & c & e & c & e & c & e & c & d \\ & & \pi_2 & a & b & c & b & c & b & c & b & c \\ & & & \tilde{\pi_2} & a & d & c & e & c & e & c & e \\ & & & & \pi_2 & a & b & c & b & c & b & c \\ & & & & & \tilde{\pi_2} & a & d & c & e & c & e \\ & & & & & & \pi_2 & a & b & c & b & c \\ & & & & & & & \tilde{\pi_2} & a & d & c & e \\ & & & & & & & & \pi_2 & a & b & c \\ & & & & & & & & & \tilde{\pi_2} & a & d \\ & & & & & & & & & & \pi_2 & a \\ & & & & & & & & & & & \tilde{\pi_2} \end{bmatrix}$$

in which we only wrote the upper diagonal part of the matrix, since it is symmetric.

Hence the second sum is found to be

$$\begin{aligned} \mathbb{E}\left(\sum_{k=0}^{11}\sum_{l=0}^{11} Z(2,k) \cdot Z(2,l)\right) &= 6\pi_2 + 6\tilde{\pi_2} + 24a + 30b + 48c + 12d + 18e \\ &= 6P_1P_2(3 - P_1) + 6P_1P_2^2(4 + 23P_1 - 16P_1^2 + 3P_1^3). \end{aligned} \tag{4.4}$$

Adding up all terms, we conclude that

$$\begin{aligned} \sigma^2(|ranch^{P_1,P_2}(v)|) = 6P_1(1 - P_1) &+ 6P_1P_2(9 - 11P_1 + 4P_1^2) \\ &+ 6P_1P_2^2(6 - 19P_1 + 14P_1^2 - 3P_1^3). \end{aligned}$$

□

As a small check of consistency, we can point out that if $P_1 = P_2 = 1$, the percolation process reach every vertex of the neighborhood at distance 2 with probability one, hence the variance is zero: $0 + 6(9 - 11 + 4) + 6(6 - 19 + 14 - 3) = 0$; if $P_1 = 1$, all vertices at distance 1 are reached by the percolation process, rendering the variance – at level 2 – of the sum of 12 independent Bernoulli random variables $B(1, P_2)$, which yields $6P_2(9 - 11 + 4) + 6P_2^2(6 - 19 + 14 - 3) = 12P_2(1 - P_2)$, or, if $P_2 = 0$, $P_1 \neq 0, 1$, only the first layer of vertices is reached, which produces the variance of the sum of 6 independent Bernoulli's $B(1, P_1)$, i.e., $6P_1(1 - P_1)$.

Corollary 4.2. *The homogeneous (P_1, P_1) percolation process produces ranches with average size $1 + 6P_1 + 12P_1^2 + 6P_1^2(1 - P_1)$ and variance $6P_1 + 48P_1^2 - 30P_1^3 - 90P_1^4 + 84P_1^5 - 18P_1^6$.*

The maximum variance is obtained with $P_1 = 0.59511$, at which the expected ranch size is 9.68 and variance 8.43. In contrast, the nonhomogeneous Bernoulli percolation process with $P = (P_1, P_2) = (1, 0.47)$ produce ranches with average size 12.64 and variance 2.98, respectively.

Theorem 4.3. *Let $v = (0, 0)$ be a given node of the graph G, $P = (P_1, P_2, \cdots, P_\kappa)$. Let $\pi_i := \prod_{j=1}^{i} P_j$, $i = 1, 2, \cdots \kappa$, and*

$$p_{i,j} = \begin{cases} \pi_i & \text{if } j = k \cdot i, k \in \{0, 1, \ldots 5\} \\ P_i\left[p_{i-1,j-1} + p_{i-1,j} - p_{i-1,j-1} \cdot p_{i-1,j}\right] & j \neq k \cdot i, k \in \{0, 1, \ldots 5\}, \end{cases} \tag{4.5}$$

Define $\partial_\kappa := \sum_{j=0}^{6\kappa-1} Z(\kappa, j)$, so that

$$|ranch^{P_1,\cdots,P_\kappa}(v)| = |ranch^{P_1,\cdots,P_{\kappa-1}}(v)| + \partial_\kappa. \tag{4.6}$$

Then the expected size and variance of $|ranch^{P_1,P_2,\cdots,P_\kappa}(v)|$ is given by

$$\begin{aligned} \mu(|ranch^{P_1,\cdots,P_\kappa}(v)|) &= 1 + \sum_{i=1}^{\kappa} \sum_{j=0}^{6i-1} p_{i,j} \\ &= \mu(|ranch^{P_1,\cdots,P_{\kappa-1}}(v)|) + \mu(\partial_\kappa), \end{aligned} \tag{4.7}$$

$$\begin{aligned} \sigma^2(|ranch^{P_1,\cdots,P_\kappa}(v)|) = \sigma^2(|ranch^{P_1,\cdots,P_{\kappa-1}}(v)|) + \sigma^2(\partial_\kappa) \\ + 2COV\left(|ranch^{P_1,\cdots,P_{\kappa-1}}(v)|, \partial_\kappa\right). \end{aligned} \tag{4.8}$$

In light of the computations made so far, the first result (expected value) is by now straightforward. The second result is just the variance of the sum of two random variables, the detailed computation requires computing the joint distribution of all variables that are dependent with each other, so for example $Z(i, j)$ with $Z(l, j)$ whenever $1 \leq l \leq i$, $j = k \cdot i$ or $Z(i, j)$ and $Z(i, l)$ for $0 \leq j \leq l \leq i - 1$, etc. We omit this computation.

4.3. Farms

Once ranches have been appropriately defined into our graph structure, we turn over to the associations between them, which will be given by overlapping clouds of CPCH and can hence undertake the problem of determining the association distribution:

Definition 4.2. *Let $P = (P_1, P_2, \ldots, P_\kappa)$, $G = (V, E)$ as before.*

1. *Two ranches r_1 and r_2 are said to* overlap *if and only if $r_1 \cap r_2 \neq \emptyset$.*

2. *Two ranches r_1 and r_2 are said to be* connected *or* in association *if and only if there exists $r_{i_1}, r_{i_2}, \ldots r_{i_k}$ such that r_1 overlaps r_{i_1}, r_{i_j} overlaps $r_{i_{j+1}}$, $1 \leq j \leq k-1$ and r_{i_k} overlaps r_2.*
3. *A* Farm *is a partition induced upon any set of ranches by the equivalence relation of "being connected" defined between ranches.*

Remark: From the biological point of view, an association cluster according to this definition consists of a set of vertices whose overlapping domains represent actual sites in which CPCH from different bivalents come into contact, hence providing a scaffold in which genetic material can be paired, compared, copied, etc., (we do not fix ourselves or pursue this topic further). The observed partitions are the sizes of the equivalence classes determined upon the constructed farm. From first principles, any random distribution of ranches will provide a different partition distribution, it remains to check whether we can reproduce the observed distribution with sufficient accuracy. □

For example, in Figure 6 we depict the results obtained for the simulation of 19 SC's distributed randomly upon the set of 602 vertices, along with their corresponding ranches and for the probability vectors a) $P = (2/3, 1/2, 2/3)$ and b) $P = (1, 0.5)$. Boundaries of chromatin have been drawn in order to emphasize the similarity with the observations.

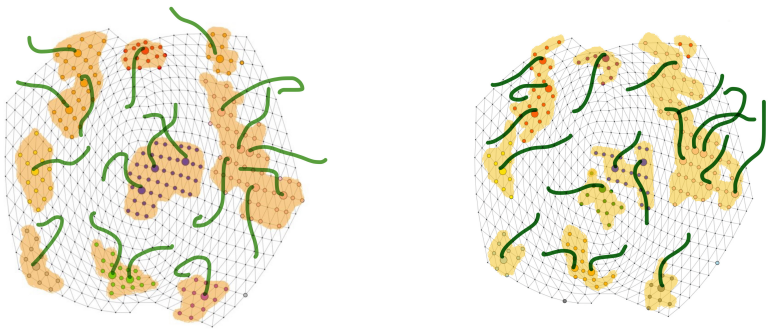

FIG 6. *Artist's rendering of simulated spermatocytes (a) $P = (1, 1/2)$, partition* $19 = 6+4+3+2+1+1+1+1+1$. *(b) $P = (2/3, 1/3, 7/24)$, partition* $19 = 7+3+2+2+1+1+1+1+1$. *(Lengths are not to scale).*

4.4. SC Dynamics

Notice that in the field of dynamics upon a fixed regular net, cellular automata and agent based models have been widely used, see for example [7], and also to a lesser extent some dynamics have been defined upon planar tilings made of hexagonal and/or pentagonal cells, as in [14]. From this point of view, we can consider that the SC's can be thought of as being agents interacting upon the surface of the nucleus (a finite connected field constituted by 590 hexagonal cells and 12 pentagonal cells).

It is known that in an earlier stage of the prophase, previous to pachytene, the size of the nuclear envelope is a fraction of the one used in our setting, and that at this

stage bivalents appear to be grouped into a single "bouquet" (Figure 7), see [10], it is during the pachytene stage that the bivalents (and more specifically the SC's) seem to be randomly (i.e;, uniformly) distributed.

While the available data does not provide information about the wandering processes of the SC's gliding upon the internal surface of the nucleus, we nevertheless propose and explore a simple dynamics for the displacement of the SC's upon the internal surface of the fullerene dual. It will turn out that it actually converges to the uniform distribution:

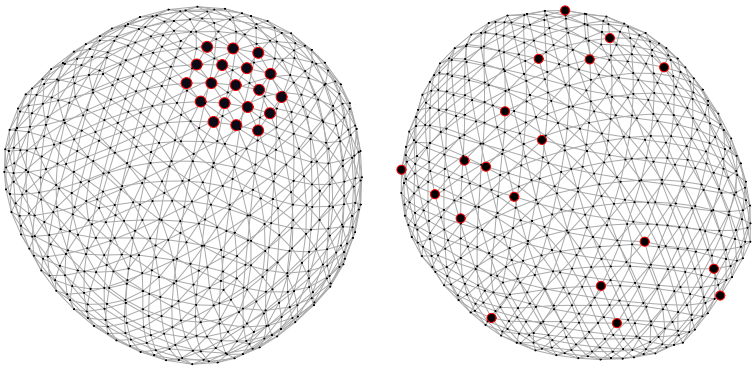

FIG 7. *Initial state (left) and result of the dynamics (right) after approximately* 10^4 *time steps*

Let $G = (V, E)$ be the graph representing the nuclear envelope of a spermatocyte. A node $v \in V$ corresponds to a position upon the nuclear envelope that can accommodate either the SC or the CPCH of a given bivalent.

To decide whether at time t this position is occupied by an SC or not, we add an attribute to each vertex:

Definition 4.3. *The* attribute *at time t of a vertex $v \in V$, denoted by $\alpha_v^t \in \{0, 1\}$ is defined by: $\alpha_v^t = 1$ if and only if a SC is located at (the position of) vertex $v \in V$ at time t.*

At time $t = 0$, we assume that all the SC are grouped and occupy the 19 first nodes, see Figure 7. (The assumption is not restrictive, as any other compact set of vertices forming a disc can also be chosen.) Thus the vector $\alpha^0 = (\alpha_1^0, \ldots, \alpha_{602}^0)$, which describes the "bouquet" configuration, is defined as $\alpha_i^0 = 1$, $i \in \{1, 2, \ldots, 19\}$, and $\alpha_i^0 = 1$ for all $i \in \{20, 21, \ldots, 602\}$. The only constraint is that two SC cannot share the same position at the same time. Thus the random trajectory of an SC is only affected by the position of the other SC's.

At time $t > 0$, with vector α^t already been computed, the positions of the SC's will be given by the following algorithm:

SC-Dynamics

- Input $G \leftarrow C'_{1200}$, $n.IT \leftarrow$ number of iterations.
- Initialize: $\alpha_i^0 \leftarrow 1$, for $i \in \{1, 2, \ldots, 19\}$, otherwise $\alpha_i^0 \leftarrow 0$, $i \in \{20, \ldots, 602\}$.
- Iterate for $t \in \{0, \ldots, n.IT - 1\}$:
  - Step 1 : for all $1 \leq i \leq 602$ do:
    - If $\alpha_i^t == 1$ then choose $v \in \mathcal{N}_G(i)$ such that $\alpha_v^t == 0$ if any; put (i, v) in a non ordered list L.
  - Step 2 : $\alpha^{t+1} \leftarrow \alpha^t$; while $L \neq \emptyset$ do:
    - extract randomly from L an element (i, v)
    - If $(\alpha_v^{t+1} == 0)$ then $\alpha_i^{t+1} \leftarrow 0$ and $\alpha_v^{t+1} \leftarrow 1$
- Output: $MC \leftarrow \{i \in V : \alpha_i^{n.IT} == 1\}$

At each time unit the complete set of SC's is allowed to simultaneously choose a free neighbor and move to it, or to remain at the present position if either there are no free neighbors or the chosen one was already selected marked for movement. In this dynamics, the sum $\sum_{i=1}^{602} \alpha_i^t = 19$ is a conserved quantity. It is a conflict free parallel discrete time upgrading dynamics.

Figure 8 illustrates one possible step from time t to time $t+1$ for a parallel evolution of three nodes in a subgraph of G.

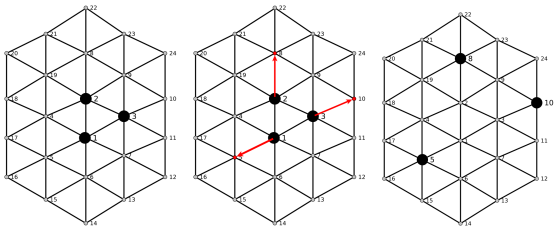

FIG 8. *Description of 1 step of (parallel) evolution in a subgraph of G. Red lines indicate direction of displacement.*

Simulations produce convergence to a uniform distribution for the positions of the SC's upon the nuclear envelope, already after $\sim 10^4$ iterations. Hence, we can implement now an algorithm for assessing the percolation process after a uniform sample of nodes has been reached.

4.5. Simulation results

We simulate the random bivalent's associations via the following *Farm-Partition* algorithm:

Farm-Partition

- Input: $G \leftarrow C'_{1200}$, $P = (P_1, P_2, \ldots P_\kappa)$.
  - MC <- SC-Dynamics $(G, n.IT \leftarrow 10^4)$.
  - For $v \in MC$: $ranch^P(v)$ ←P-Percolation(G, v, P).
  - $Farm \leftarrow \{F_1, \ldots, F_r\}$ vertex sets of the connected components of $\cup_{v \in MC} ranch^P(v)$ in G.
- Output: $Part \leftarrow$ sort($|F_j|$, $j = 1, \ldots, r$, decreasing = TRUE).

This algorithm was implemented using R software [11] and iterated *ad libitum* in order to produce a simulated set of nuclei with which available data could be contrasted. Notice that in our notation, the biggest subcluster found in the partition corresponds to $Part[1]$, i.e., 1^{st}-Class, the second one in $Part[2]$, i.e., 2^{nd}-Class, etc. Best approximation to the observed distribution for the 1^{st}- and 2^{nd}- Class was obtained for different values of $P = (P_1, P_2)$, where a bootstrapping technique was used for assessing the intrinsic variation in the data/simulations in both cases, the homogeneous and the non homogeneous one.

Figure 9 shows, for example, the results obtained for the homogeneous percolation with $P = (0.68, 0.68)$ after simulating 10^4 spermatocytes (i.e., by running Farm-Partition 10^4 times with $G \leftarrow C'_{1200}$, $P = (0.68, 0.68)$).

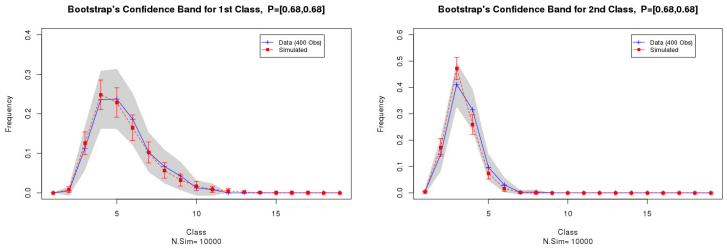

FIG 9. *Best approximation for 1^{st}-Class and 2^{nd}-Class distributions for the homogeneous case $P = (0.68, 0.68)$. Similar approximations are obtained in the non homogeneous case for various values of the P-vector, see Table 3 anf Figure 10.*

P_1	P_2	μ	σ^2	$\hat{\mu}$	$\hat{\sigma}^2$
1.0	0.47	12.64	2.99	12.55	3.00
0.8	0.60	12.14	5.84	12.05	5.97
0.68	0.68	11.52	8.03	11.40	8.03
0.6	0.78	11.34	10.01	11.22	10.06
0.52	1.0	11.86	13.50	11.82	13.86

TABLE 3

A sample of (P_1, P_2) values for which the associated P percolation process gives the best approximation for the 1^{st}- and 2^{nd}-Class probability distribution, together with theoretical and simulated ranch sizes and their variances.

Other values with good agreement with respect to the data are the non homogeneous case $P = (1, 0.47)$, or $P = (0.8, 0.6)$, etc., see Table 3. Included are the theoretical values for the mean value and variance of the ranch's sizes in the 6-regular approximation (equations (4.2) and (4.3) from Theorem 4.2), as well as those obtained by simulation in the dual fullerene C'_{1200}. The homogeneous percolation with minimal variance $P = (0.595, 0.595)$ does not furnish a satisfactory 1^{st}-Class probability distribution, so it is not included in the table.

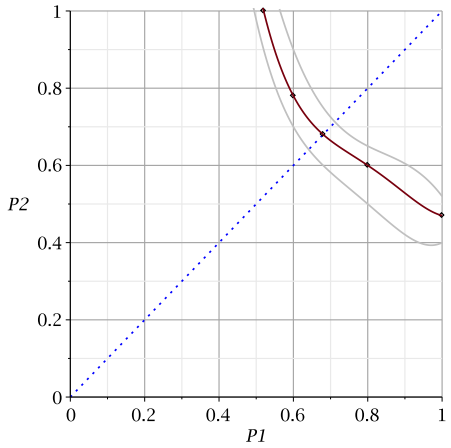

FIG 10. *Best approximation regions in $P = (P_1, P_2)$- space for 1^{st}-Class and 2^{nd}-Class distributions obtained with P-percolation.*

Figure 10 depicts the zone in (P_1, P_2)-space with the best fitting to the observed frequencies in both, 1^{st}-Class and 2^{nd}-Class. The criteria used for determining the lower and upper curves was that the entire simulated mean distribution does not fall outside the variation zone around the observed mean values (frequencies) in the data set. The red curve depicts the best fit in which both, the simulated and the observed mean distribution lie close together. Depicted here is also the homogeneous percolation region $P = (P_1, P_1)$, which - intersected with the region in between the lower and upper curves - yield the best approximation range for the distribution of 1^{st}-Class and 2^{nd}-Class in the homogeneous case.

5. Comments and Discussion

Chromosomal bivalent's associations through intersecting CPCH create rich dynamic and diverse scenarios via the participating elements. They are triggered by the corresponding associations of CPCH located at the short arms of the bivalents, but also by the resulting convergence of the rest of the constituent chromatin along them. This supra chromosomal bonds allows, for example, joint expression of genes coming from different bivalents.

Undoubtedly, the better understanding of the general principles behind bivalent's associations in prophase meiotic nuclei, as well as precising the type of randomness being

at play at this stage could bring us also a step closer to a better understanding of the different chromosome combinations present in the gametes. Since these associations and combinations persist until the meiotic divisions, the chromosomal associations as described here necessarily leaves some imprint in the chromosomal sets passed on to gametes and hence their importance to evolution. While available data provides no distinctive insights upon prophase progression from the bouquet state to the observed associations, whose distributions are studied here, it is by simulating different scenarios that we obtain a simple conflict free parallel dynamics for randomly selecting free neighbors, with which these distributions, as measured by 1^{st}- and 2^{nd}-Classes, can be correctly reproduced. For this, a suitable choice of vector P has to be made, which is a parameter needed for the ranch determination and therefore is an architectural parameter worth considering.

The approach pursued in this work possess clear advantages over earlier modeling attempts, since now we have a tool that not only reproduces available data, but can also drive further biological research in topics that have not been explored yet, for example expressing or assessing individual properties attached to the individual chromosomes via different choices of their ranches, i.e., via assigning individual vector P's to each CTC. To our knowledge this has not been reported in the literature and will be undertaken next. We conjecture a correlation between our ranch-sizes, expressed in terms of the number of neighboring vertices with chromatin attributes, and the actual sizes of the chromosomes (respectively their chromatin content). These conditions should allow us to propose new experiments for assessing individual chromatin content (for example as a ratio volume/surface) or to characterize it by means of appropriate choices of vectors P. This approach would allow us to effectively relax our previous hypothesis of indistinguishable CTC's, or at least, to put it to a test.

From the biological point of view, the introduction of chromatin neighborhoods with (random) variable sizes reflects more properly the true nature of the objects being modeled and, perhaps more significantly, it is the very intersection of chromatin domains, which appears here naturally for describing the interactions of CTC's via its induced clustering relation, that represents better the biological importance of the described process, because, while in hetero-chromatin there is no actual genic expression, it certainly contributes to bringing together genes located at other chromosomal domains, so that the associations themselves could be conceived as a form of dynamical organization that contributes to the functionality of the joint genic expression, which – ultimately – may be looking at us hidden behind the available data.

References

[1] Allard, A., Hébert-Dufresne, L., Noël, P. A., Marceau, V. and Dubé, L. J. (2012). Exact solution of bond percolation on small arbitrary graphs. *EPL (Europhysics Letters)* **98** 16001.

[2] Berend, D. and Tassa, T. (2010). Improved bounds on Bell numbers and on moments of sums of random variables. *Probability and Mathematical Statistics* **30** 185–205.

[3] Berríos, S., Manterola, M., Prieto, Z., López-Fenner, J., Page, J. and Fernández-Donoso, R. (2010). Model of chromosome associations in Mus domesticus spermatocytes. *Biological research* **43** 275–285.
[4] Erdös, P. and Rényi, A. (1959). On random graphs. I. *Publicationes Mathematicae* **6** 290–297.
[5] Gilbert, E. N. (1959). Random Graphs. *Ann. Math. Statist.* **30** 1141–1144.
[6] Grimmett, G. R., Manolescu, I. et al. (2013). Inhomogeneous bond percolation on square, triangular and hexagonal lattices. *The Annals of Probability* **41** 2990–3025.
[7] Leng, B., Wang, J., Zhao, W. and Xiong, Z. (2014). An extended floor field model based on regular hexagonal cells for pedestrian simulation. *Physica A: Statistical Mechanics and its Applications* **402** 119–133.
[8] López-Fenner, J., Berríos, S., Manieu, C., Page, J. and Fernández-Donoso, R. (2014). Bivalent associations in Mus domesticus 2n= 40 spermatocytes. Are they random? *Bulletin of mathematical biology* **76** 1941–1952.
[9] Rota, G.-C. (1964). The number of partitions of a set. *Am Math Mon* **71** 498–504.
[10] Scherthan, H. (2001). A bouquet makes ends meet. *Nature reviews Molecular cell biology* **2** 621–627.
[11] R Core Team (2013). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
[12] Trudeau, R. J. (1993). Introduction to Graph Theory (Corrected, enlarged republication. ed.).
[13] Weisstein, E. W. Fullerene MathWorld–A Wolfram Web Resource.
[14] Wąs, J., Porzycki, J., Lubaś, R., Miller, J. and Bazior, G. (2016). Agent based approach and cellular automata: A promising perspective in crowd dynamics modeling? *Acta Physica Polonica B Proceedings Supplement* **9** 133–144.