



Deterministic tourist walk

Rodrigo Silva González¹, Juliana Militão Berbert¹, César Augusto Sangaletti Terçariol²,
Wilnice Tavares Reis Oliveira¹ and Alexandre Souto Martinez³

Manuscript received on Month ??, 200? / accepted on Month ??, 200?

ABSTRACT

A set of N points is randomly spread in a d -dimensional hypercube of unitary edges. The neighborhood statistics among any pair of points is known as the “random point problem” (RPP). A walker can move over these points following the deterministic rule of going, at each time step, to the nearest site not visited in the previous μ steps. This partially self-avoiding walk is known as “deterministic tourist walk” (DTW). Here we present some results on the RPP and on dynamic aspects of the DTW.

Keywords: tourist walks, disordered media, partially self-avoiding walks.

Correspondence to: Alexandre Souto Martinez

¹Departamento de Física e Matemática (DFM), Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes, 3900, 14040-901 Ribeirão Preto, SP, Brazil. E-mails: rsgonzalez@pg.ffclrp.usp.br / bjuliana@usp.br / wilnice@pg.ffclrp.usp.br

²Centro Universitário Barão de Mauá, rua Ramos de Azevedo, 423, 14090-180 Ribeirão Preto, SP Brazil. E-mail: cesar@pg.ffclrp.usp.br

³Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes, 3900, 14040-901, Ribeirão Preto, SP, Brazil; and National Institute of Science and Technology for Complex Systems, Avenida Bandeirantes, 3900, 14040-901 Ribeirão Preto, SP, Brazil. E-mail: asmartinez@usp.br

1 INTRODUCTION

Random walks over regular lattices and random media have been often studied and have wide range of applications [1, 2, 3]. Although not so thoroughly investigated, deterministic walks in regular [4, 5, 6] and disordered media [7] also present very interesting results. Here, we are interested in a partially self-avoiding deterministic walk algorithm, which we call the deterministic tourist walk (DTW) [8, 9].

A standard way to construct a disordered medium consists in randomly spreading N points (with coordinates uniformly distributed along the edges) on a d -dimensional hypercube. The determination of the distance and neighborhood statistics between any pair of points is known as the *random point problem* (RPP). One can consider these points as sites in a random landscape and a walker can move on them. Following the deterministic rule of, at each discrete time step, going to the nearest site not visited in the previous μ steps, a walker performs a partially self-avoiding walk, where self-avoidance is limited to the memory window $\tau = \mu - 1$. This quantity represents a characteristic time to the site to become attractive to the walker again (refractory time) and prohibits a trajectory from intersecting itself inside this memory window.

The tourist behavior depends strictly on the data set configuration. The tourist movements are entirely performed based on a neighborhood table. This table represents the *tourist graph*, i.e., nodes with $\mu - 1$ fixed directed and weighted outgoings each and with a variable number of incomings. Notice that the distances among the sites is simply a way to rank their neighborhood.

The paper is organized as follows. In Section 2, we present a study of the static characteristics of the underlying random medium on which the DTW is performed. Some geometrical properties are shown. In Section 3, the dynamics of the DTW is presented, showing its behavior when the memory μ and the dimension d are combined. Conclusions are presented in Section 4.

2 RANK NEIGHBORHOOD STATISTICS

The probability $P_{m,n}^{(d,N)}$ that an arbitrary point is the m th nearest neighbor of its own n th nearest neighbor in the RPP has been initially studied by Clark and Evans [10] and Clark [11] on some aspects of spatial pattern in biological populations. Figure 1 shows a geometrical representation for $d = 2$. They devised the term reflexive neighbors for the case $m = n$. Their work has been corrected by Dacey [12] ($m, n > 1$) in the context of geographical analysis and then generalized (for $m \neq n$) by Cox [13]. We call $P_{m,n}^{(d,N)}$ the *Cox probabilities*. We have calculated [14] these

probabilities using only Poisson distribution instead of the various distinct distributions used in the original paper [13]. Unlike Cox, we have written the resulting expression in terms of known functions (rather than in terms of an integral) and known distributions (multinomial, binomial and hypergeometric).

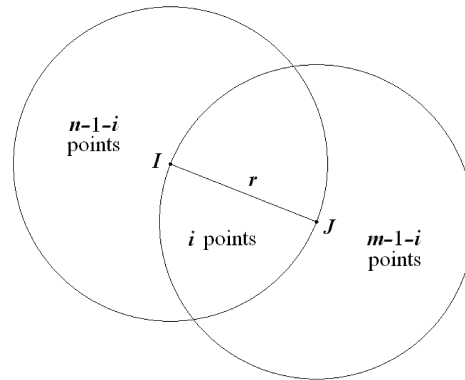


Figure 1 – Two-dimensional Poissonian process. There are i points in the intersection of the surfaces and in the I and J crescents there are $n - 1 - i$ and $m - 1 - i$ points, respectively.

In the following, we show the Cox probabilities for a finite number of dimensions. Then, we compute its high dimensional limit, that leads to the *random link model* (RLM). Also the *random map model* (RMM) is presented.

2.1 Cox probabilities in finite dimensionality

The original form of Cox probabilities is written as:

$$P_{m,n}^{(d)} = \sum_{i=0}^{\min(m-1,n-1)} \frac{(m+n-2-i)!}{i!(m-1-i)!(n-1-i)!} \times \frac{(1-I_d)^i I_d^{m+n-2-2i}}{(1+I_d)^{m+n-1-i}} \tag{1}$$

with $m = 1, 2, \dots, \infty$, $n = 1, 2, \dots, \infty$ and

$$I_d = I_{1/4} \left(\frac{1}{2}, \frac{d+1}{2} \right) \tag{2}$$

where

$$I_z(a, b) = \frac{\int_0^z dt t^{a-1} (1-t)^{b-1}}{B(a, b)}$$

with $\text{Re}(a) > 0$, $\text{Re}(b) > 0$ is the normalized incomplete beta function [15]. Geometrically, I_d represents the relative volume of the crescent compared to the hypersphere in a poissonian process, as shown in Figure 1 for $d = 2$ [14]. Notice that I_d depends exclusively on the dimensionality d of the RPP. Letting i vary from 1 to $\min(m, n)$ and rearranging the terms, one is able

to rewrite Eq. 1 as:

$$\frac{P_{m,n}^{(d)}}{P_{1,1}^{(d)}} = \sum_{i=1}^{\min(m,n)} \text{Mult} \left(i-1, m-i, n-i; \frac{1-I_d}{1+I_d}, \frac{I_d}{1+I_d}, \frac{I_d}{1+I_d} \right) \quad (3)$$

$$P_{1,1}^{(d)} = \frac{1}{1+I_d}, \quad (4)$$

where $P_{1,1}^{(d)}$ is the couple density (mutually nearest neighbors) and $\text{Mult}(n_a, n_b, n_c; \pi_a, \pi_b, \pi_c) = (n_a + n_b + n_c)! \pi_a^{n_a} \pi_b^{n_b} \pi_c^{n_c} / (n_a! n_b! n_c!)$ is the multinomial distribution.

2.2 High dimensionality probabilities and random link model

The distances between any pair of points are not all independent random variables, due to boundary effects and triangular restrictions. For fixed N in the RPP, as d increases, the boundary effects become more pronounced and the distances between pair of points become less correlated. Boundary effects can be minimized considering periodic boundary condition and, in the limit $d \rightarrow \infty$, all the two-point distances are independent and identically distributed (i.i.d.) random variables. This is the *random link model* (RLM) [16, 17], which is a mean field description of the RPP. We stress that in the RLM, there exist two Euclidean constraints: (i) the distance from a point to itself is always null ($D_{ii} = 0$, for all i) and (ii) the forward and backward distances are equal ($D_{ij} = D_{ji}$, for all i, j).

The high dimensionality limit for the Cox probabilities can be obtained directly from Eq. 3. In the thermodynamic limit ($N \gg 1$), the multinomial distribution becomes:

$$\frac{P_{m,n}^{(r,l)}}{P_{1,1}^{(r,l)}} = \text{Bin} \left(m-1, n-1, \frac{1}{2}, \frac{1}{2} \right) \quad (5)$$

$$P_{1,1}^{(r,l)} = \frac{1}{2}, \quad (6)$$

where $\text{Bin}(n_a, n_b; \pi_a, \pi_b) = (n_a + n_b)! \pi_a^{n_a} \pi_b^{n_b} / (n_a! n_b!)$ is the binomial distribution. Simple expressions can be obtained such as: $P_{1,n}^{(r,l)} = 1/2^n$, $P_{2,n}^{(r,l)} = n/2^{n+1}$.

Since here, Euclidean distances are only a means to obtain the ranking neighborhood probabilities, it is independent of particular choice for the distance probability distribution function (pdf) [18]. For simplicity, we will consider uniform deviates in the interval $[0, 1]$ for the distances among the N points. In the

random link case, one is able to obtain the Cox probability for a finite system. To take into account finite size effects, just replace the binomial distribution of Eq. 5 by:

$$\frac{P_{m,n}^{(r,l,N)}}{P_{1,1}^{(r,l,N)}} = \text{Hypg}(N-2, N-2; m-1, n-1) \quad (7)$$

$$P_{1,1}^{(r,l,N)} = \frac{N-1}{2N-3}, \quad (8)$$

with $m = 1, 2, 3, \dots, N-1$ and $n = 1, 2, 3, \dots, N-1$, where

$$\text{Hypg}(N_a, N_b; n_a, n_b) = \frac{\binom{N_a}{n_a} \binom{N_b}{n_b}}{\binom{N_a + N_b}{n_a + n_b}}$$

is the hypergeometric distribution. Obviously, these equations (Eqs. 7 and 8) reduce to Eqs. 5 and 6 as $N \gg 1$.

We observe that one can use the determinism of the algorithm of pseudo random number generator to obtain very large random link model data sets [19]. Up to date, we have not being able to consider finite size effects for Eq. 3 and 4. One must generalize the hypergeometrical distribution so that for $N \gg 1$ one obtains the multinomial one.

2.3 Random map model

Breaking the distance symmetry constraint in the RLM the *random map model* (RMM) [20] is obtained. The RMM is the mean field approximation to the Kauffman's model [21] and analytical results may be obtained. Also, Cox probabilities can be obtained for the RMM.

In the case in which the constraint $D_{ii} = 0, \forall i$ is preserved, if an arbitrary point I is chosen, its m th neighbor J is automatically set, but the n th neighbor of J is equally probable to be anyone of the other $N-1$ points, since the distances are totally independent. Thus, the probability $P_{m,n}^{(r,m)}$ that the point I is the n th neighbor of its m th neighbor is simply:

$$P_{m,n}^{(r,m)} = \frac{1}{N-1}, \quad (9)$$

where $m = 1, 2, \dots, N-1$ and $n = 1, 2, \dots, N-1$.

On the other hand, in the case which $D_{ii} \neq 0$ is allowed, the probability $P_{m,n}^{(r,m)}$ is twice as large for reflexive neighbors than for non-reflexive ones, because now one must consider that every point is always its own m th nearest neighbor, for some m . Therefore:

$$P_{m,n}^{(r,m)} = \frac{1 + \delta_{m,n}}{N+1}, \quad (10)$$

where $\delta_{m,n}$ is the Kronecker's delta, $m = 1, 2, \dots, N$ and $n = 1, 2, \dots, N$. Notice that in the thermodynamic limit $N \gg 1$, these cases are still distinguishable due to the presence of the factor 2 for the reflexive neighbors.

3 SOME RESULTS ON THE DETERMINISTIC TOURIST WALK

All DTW have an initial transient part of length t and end in cycles. Each cycle has a period p . Both transient time and cycle period can be combined in the joint distribution $S_{\mu,d}^{(N)}(t, p)$. In the following we consider some special cases obtained combining μ and d .

3.1 Lazy tourist

The simplest case to deal with the DTW is to consider $\mu = 0$. In the following we present the result for an arbitrary dimensionality and for the random map model.

3.1.1 Arbitrary dimensionality Euclidean space

This case is trivial, since the walker has a null-size memory. The walker remains at the same site and the trajectory has a zero-length transient and a cycle of period $p = 1$. The transient and period joint distribution is simply given by:

$$S_{0,d}^{(N)}(t, p) = \delta_{t,0} \delta_{p,1}, \quad (11)$$

where $\delta_{i,j}$ is the Kronecker's delta.

Despite its triviality, this becomes interesting because it is the simplest situation of the stochastic tourist walk [22]. Also, in the deterministic situation, the $\mu = 0$ case can be analytically solved and presents non-trivial solutions for the RMM, as we see below.

3.1.2 Random map

In this model, the lazy tourist may not exist, since the distance, which may be interpreted as the cost, can be greater to remain in a site than going away, in this case the walker may not stay at the same site. Thus, even for $\mu = 0$, non-trivial transient time and cycles distributions may exist. Indeed, the transient time and period cycles joint distribution for a random map with N points is [18]:

$$S_{0,rm}^{(N)}(t, p) = \frac{\Gamma[N]}{\Gamma[N+1-(t+p)]N^{t+p}}. \quad (12)$$

We call attention to the fact that the relevant variable is $n_e = t + p$, which is the number of explored sites along the

trajectories. The period distribution is simply:

$$S_{0,rm}^{(N)}(p) = \sum_t S_{0,rm}^{(N)}(t, p),$$

and for $N \gg 1$, it can be approximated by:

$$S_{0,rm}^{(N)}(p) = \sqrt{\frac{\pi}{2N}} \operatorname{erfc}\left(\frac{p}{\sqrt{2N}}\right) \approx \frac{e^{-p^2/(2N)}}{p}, \quad (13)$$

where it is interesting to point out that for $p \ll \sqrt{2N}$ the period distribution decays as a power law: $S_{0,rm}^{(N)}(p) \approx p^{-1}$, independently of the size N of the system.

3.2 Memoryless tourist

For $\mu = 1$, the walker knows only the nearest neighbor of the current site. Thus, at each time step, he must leave the current site and go to the nearest one. The name "memoryless tourist" is devised because the walker knows the site he is, but does not remember any of the previously visited ones. This rule does not lead to exploration of the random medium, since after a very short transient time the walker gets trapped by a couple of mutually nearest neighbors. These couples are the attractors of the system, and the period distribution is simply

$$S_{1,d}^{(N)}(p) = \sum_t S_{1,d}^{(N)}(t, p) = \delta_{p,2}.$$

The transient time and period joint distribution has been analytically obtained for $N \gg 1$ [18]:

$$S_{1,d}^{(\infty)}(t, p) = \frac{\Gamma(1 + I_d^{-1})(t + I_d^{-1})}{\Gamma(t + p + I_d^{-1})} \delta_{p,2}, \quad (14)$$

where $\Gamma(z)$ is the gamma function and I_d is given by Eq. 2.

3.3 Tourist with memory

Interesting phenomena occur when greater values of μ are considered. In this case, the cycle distribution is no longer peaked at $p_{\min} = \mu + 1$, but presents a whole spectrum of cycles with period $p \geq p_{\min}$, with possible power-law decay [8, 9, 23].

3.4 One-dimensional systems

It is interesting to point out that, for 1D systems, determinism imposes serious restrictions, as we discuss below.

3.4.1 Very short memory

For $\mu \ll N$, cycles of period $p \in [2\mu + 1, 2\mu + 3]$ are forbidden. Additionally, for $\mu = 2$ all odd periods but $p_{\min} = 3$

are forbidden. The number of site per attractor is [8]:

$$n(p) = \frac{p}{2} + \tau + 1. \quad (15)$$

3.4.2 Crossover between localized and extended regimes

As one can see in Figure 3, the heavy tail of the period marginal distribution $S_{\mu,d}^{(N)}(p)$ allows the walker to explore the whole system even for small memory values ($\mu \ll N$). If a walker starts to walk in one extremity of a 1D map, he can reach the other extremity (using open boundary conditions) if he takes N steps without changing its walking direction. So, he needs a memory μ large enough to avoid being trapped in cycles. For a given memory μ , this occurs with probability [24, 25]:

$$P(N|\mu) = (1 - 2^{-\mu})^{N-\mu-1}. \quad (16)$$

Figure 2 shows the graphics of Eq. 16 for three values of N . One can easily see the existence of an abrupt change in the walker's exploratory behavior in 1D at a narrow memory range. Considering $N \gg \mu + 1$ and computing the inflection point of Eq. 16, one can determine the value of memory at which the transition takes place:

$$\mu_1 = \log_2 N. \quad (17)$$

The width of the transition range is independent of N : $\varepsilon = e/\ln 2$.

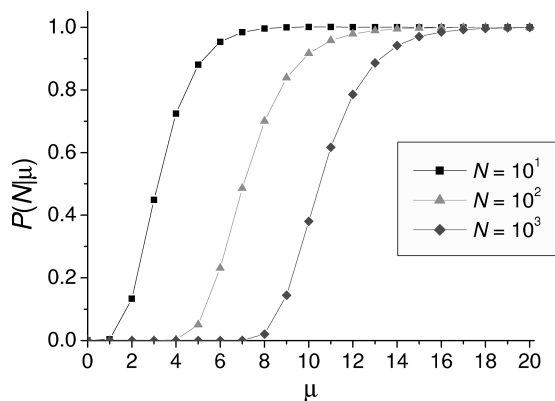


Figure 2 – Graphics of Eq. 16 for $N = 10, 100$ and 1000 . It is possible to see the existence of an abrupt transition in the value of $P(N|\mu)$ at $\mu_1 = \log_2 N$ in a narrow memory range $\varepsilon = e/\ln 2$.

On one hand, for $\mu < \mu_1$, the walker is trapped in cycles inside the system. The walker cannot reach the other extremity and the exploration remains local. On the other hand, if $\mu > \mu_1$, the number of visited points increases, allowing the walker to reach the other extremity and the exploration becomes global.

3.5 Higher dimensionalities

Figure 3 shows the t and p joint distribution histograms for $\mu = 9$ and $\mu = 25$ with $N = 1000$ and $d = 2$. As application of these deterministic walks, the cycles have been used as a thesaurus characterization [23], a clusterization method [26] and in texture analysis in images [27, 28].

3.6 Mean field approximation

The mean field approximation consists on making each point to interact with the same intensity with all the others in the system. In the following, we analyze two systems which have this feature: the *random link* and the *random map* models. The first is the natural high dimensional limit for the tourist walk problem. The second is the only analytically solved case that has a non-trivial distribution of cycles even for the lazy and memoryless tourist cases. Both models have been studied for the tourist walk in Ref. [18].

3.6.1 Random link

As the number of dimensions of the euclidean space increases, the correlations between distances weaken down. When $d \rightarrow \infty$, these correlations can be neglected and the distances between points considered as independent random variables. In this model, the transient time distribution, in the thermodynamical limit, to the memoryless tourist walk can be obtained taking the limit of $d \rightarrow \infty$ in Eq. 14:

$$S_{1,rl}^{(\infty)}(t) = \frac{t+1}{(t+2)!}. \quad (18)$$

For $\mu = 2$, the transient time and period cycle joint distribution is [29]:

$$S_{2,rl}^{(N)}(t, p) = \frac{e^{-\frac{(t+p-2)(t+p-3)}{3N}}}{(3 - \delta_{t,0})N}. \quad (19)$$

The marginal distributions of transient time and period are, respectively:

$$S_{2,rl}^{(N)}(t) = \left(1 + \frac{\delta_{t,0}}{2}\right) \sqrt{\frac{\pi}{6N}} \operatorname{erfc}\left(\frac{t}{\sqrt{6N}}\right), \quad (20)$$

$$S_{2,rl}^{(N)}(p) = \sqrt{\frac{\pi}{6N}} \operatorname{erfc}\left(\frac{p-7/2}{\sqrt{6N}}\right) \approx \frac{e^{-p/(6N)}}{p}. \quad (21)$$

We call attention to the fact that the decay of the transient time distribution for $\mu = 1$ is much faster than for $\mu \geq 2$. This is due to the main mechanism, which is driven by extreme value statistics in the $\mu = 1$ case and by combinatorial statistics for

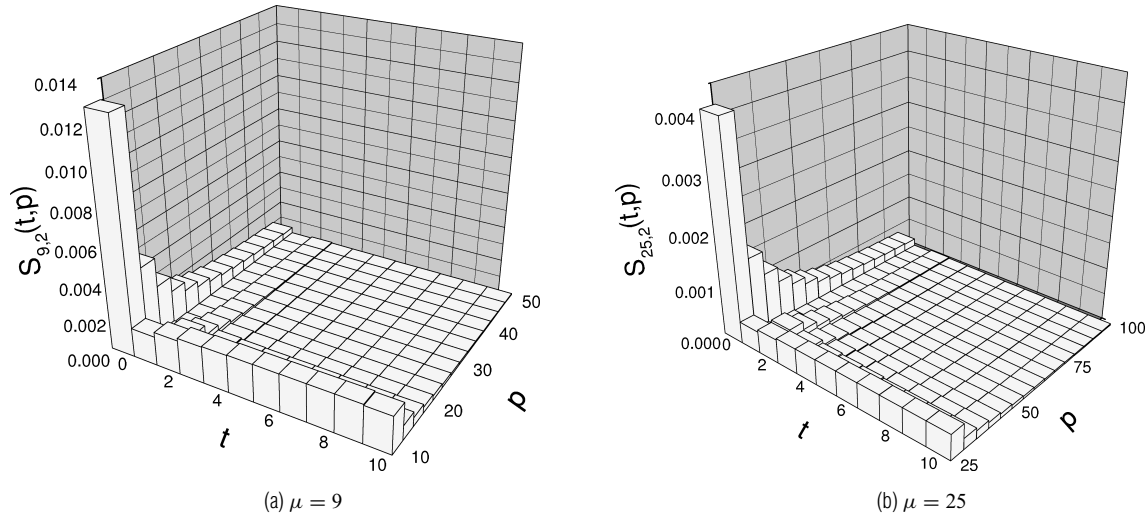


Figure 3 – Joint distribution of transient time (t) and period (p) with $N = 1000$ and $d = 2$ for two values of μ . One observes that the general aspects of both histograms are similar.

$\mu = 2$. This change, from extreme value statistics to combinatorial one, converts the $\delta_{p,2}$ distribution for $\mu = 1$ to a wide period distribution power-law decay for $\mu \geq 2$ (Eq. 21). From these considerations, one concludes that even a memory window of a single step drastically favors the walker to explore the random medium.

3.6.2 Random map

In this model, all symmetry restrictions are relaxed, and the distance from one point to itself is not necessarily null. So, for a N -point map, there are N^2 independent random distances. The transient time and period joint distribution and the period marginal distribution for this case, with $\mu = 0$, are respectively given by Eq. 12 and Eq. 13. For the quantity $n_e = t + p$, Eq. 12 becomes:

$$S_{0,rm}^{(N)}(n_e) = \frac{n_e \Gamma[N]}{\Gamma[N - n_e - 1] N^{n_e}}. \quad (22)$$

One can observe that this distribution diverges at the thermodynamical limit.

3.6.3 Equivalence between the random link and random map models

The random link and random map models are the mean field approximations of the random point problem and Kaufmann's model, respectively. These models are very different in essence, nevertheless their mean field approximations are similar. If one

compares Eq. 21 to Eq. 13, one sees a non-trivial equivalence between RMM with $3N$ points and $\mu = 0$ and RLM with N points and $\mu = 2$ [29].

4 CONCLUSION

Although less investigated than random walks, deterministic walks also present very interesting results. In this paper we revised the random point problems and the deterministic tourist walks. For the random point problem, we have shown some geometrical characteristics and the static properties of the medium in which the DTW takes place. We have presented the behavior of the DTW depending on the system dimensionality d and the walker's memory μ , pointing out the transient time and period joint distributions and the existence of a crossover transition in 1D systems.

ACKNOWLEDGMENTS

R.S. González (140420/2007-0), J.M. Berbert (134461/2007-0) and A.S. Martinez (303990/2007-4 and 476862/2007-8) acknowledges the Brazilian agency CNPq for support. W.T.R. Oliveira acknowledges the Brazilian agency FAPEMAT for support.

REFERENCES

- [1] FISHER DS. 1984. Phys. Rev. A, 30: 960.
- [2] METZLER R & KLAFTER J. 2000. Phys. Rep., 339: 1.
- [3] DERRIDA B. 1997. Physica D 107: 186.
- [4] FREUND H & GRASSBERGER P. 1992. Physica A, 190: 218.

- [5] BUNIMOVICH LA & TROUBETZKOY SE. 1992. *J. Stat. Phys.*, 67: 289.
- [6] GALE D, PROPP J, SUTHERLAND S & TROUBETZKOY S. 1995. *Math. Intelligencer*, 17: 48.
- [7] BUNIMOVICH LA. 2004. *Physica D*, 187: 20.
- [8] LIMA GF, MARTINEZ AS & KINOUCI O. 2001. *Phys. Rev. Lett.*, 87: 010603.
- [9] STANLEY HE & BULDYREV SV. 2001. *Nature (London)*, 413: 373.
- [10] CLARK PJ & EVANS FC. 1955. *Science*, 121: 397.
- [11] CLARK PJ. 1956. *Science*, 123: 373.
- [12] DACEY MF. 1969. *Geographical Analysis*, 1: 385.
- [13] COX TF. 1981. *Biometrics*, 37: 367.
- [14] TERÇARIOL CAS, MOURA KIIPPER F & MARTINEZ AS. 2007. *J. Phys. A*, 40: 1981.
- [15] ABRAMOWITZ M & STEGUN IA. (eds.) 1972. *Handbook of Mathematical Functions* (Dover Publications, New York).
- [16] MÉZARD M & PARISI G. 1986. *Europhys. Lett.*, 2: 913.
- [17] PERCUS AG & MARTIN OC. 1996. *Phys. Rev. Lett.*, 76: 1188.
- [18] TERÇARIOL CAS & MARTINEZ AS. 2005. *Phys. Rev. E*, 72: 021103.
- [19] TERÇARIOL CAS & MARTINEZ AS. 2006. *Braz. J. Phys.*, 36: 232.
- [20] DERRIDA B & FLYVBJERG H. 1987. *J. Phys. (Paris)*, 48: 971.
- [21] KAUFFMAN SA. 1969. *J. Theor. Biol.*, 22: 437.
- [22] RISAU-GUSMAN S, MARTINEZ AS & KINOUCI O. 2003. *Phys. Rev. E*, 68: 016104.
- [23] KINOUCI O, MARTINEZ AS, LIMA GF, LOURENO GM & RISAU-GUSMAN S. 2002. *Physica A*, 315: 665.
- [24] TERÇARIOL CAS, GONZALEZ RS & MARTINEZ AS. 2007. *Phys. Rev. E*, 75: 061117.
- [25] TERÇARIOL CAS, GONZALEZ RS, OLIVEIRA WTR & MARTINEZ AS. 2007. *Physica A*.
- [26] CAMPITELI MG, KINOUCI PDO & MARTINEZ AS. 2006. *Phys. Rev. E*, 74: 026703.
- [27] BACKES AR, BRUNO OM, CAMPITELI MG & MARTINEZ AS. 2006. *Lect. Note Comput. Sci.*, 4225: 784.
- [28] CAMPITELI MG, MARTINEZ AS & BRUNO OM. 2006. *Lect. Note Comput. Sci.*, 4140: 159.
- [29] TERÇARIOL CAS & MARTINEZ AS. 2008. *Physical Review E*, 78: 070603.