Islands of Stability in Motif Distributions of Random Networks

M. V. Tamm, ^{1,4} A. B. Shkarin, ² V. A. Avetisov, ^{3,4} O. V. Valba, ^{4,5,6} and S. K. Nechaev ^{4,5,7}

¹Physics Department, Moscow State University, 119992 Moscow, Russia

²Department of Physics, Yale University, 217 Prospect Street, New Haven, Connecticut 06511, USA

³N.N. Semenov Institute of Chemical Physics of the Russian Academy of Sciences, 119991 Moscow, Russia

⁴Department of Applied Mathematics, National Research University Higher School of Economics, 101000 Moscow, Russia

⁵Université Paris–Sud/CNRS, LPTMS, UMR8626, Bâtiment 100, 91405 Orsay, France

⁶Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Russia

⁷P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991 Moscow, Russia

(Received 5 July 2013; revised manuscript received 26 March 2014; published 26 August 2014)

We consider random nondirected networks subject to dynamics conserving vertex degrees and study, analytically and numerically, equilibrium three-vertex motif distributions in the presence of an external field h coupled to one of the motifs. For small h, the numerics is well described by the "chemical kinetics" for the concentrations of motifs based on the law of mass action. For larger h, a transition into some trapped motif state occurs in Erdős-Rényi networks. We explain the existence of the transition by employing the notion of the entropy of the motif distribution and describe it in terms of a phenomenological Landau-type theory with a nonzero cubic term. A localization transition should always occur if the entropy function is nonconvex. We conjecture that this phenomenon is the origin of the motifs' pattern formation in real evolutionary networks.

DOI: 10.1103/PhysRevLett.113.095701 PACS numbers: 64.60.aq, 05.70.Fh, 89.75.Hc

Study of complex networks constitutes a rapidly developing interdisciplinary area [1,2] which unites investigation of various types of experimentally observed biological [3], social [4–6], and engineering [7] networks, as well as artificial random graphs constructed by various probabilistic techniques [8–12]. Many statistical properties of networks, such as vertex degree distribution, clustering coefficients, "small world" structure [13], and adjacency matrices spectra [14], have been studied.

One particular topological characteristic, seemingly very instructive in providing detailed information about local network topology, is the distribution of small subgraphs (motifs). The presence of motifs of a specific type is tightly linked to network function. For example, transcriptional regulatory networks are rich in autoregulation loops compared to random graphs with the same vertex degree distribution [15,16]. Some authors [17] connect the prevalence of specific motif types in the transcription factor networks of E.coili and S.cerevisiae with network evolution, which might lead to self-consistent optimal circuit design [18-20]. It is known that protein interaction networks have many short cycles and completely connected subgraphs [21] which may be necessary for signal transduction by feed-back loops [18]. These and many other examples clearly demonstrate that the prevalence of specific motifs strongly correlates with the network function.

The statistical analysis of motif distribution (MD) in various naturally observed directed networks demonstrates [22,23] that they can be split into four (or three, according to recent new evidence [24]) broad superfamilies with respect to their three-vertex motif (triad) distributions, and

the networks within the same superfamily tend to have similar functions. However, there is still no common opinion about the mechanism behind the separation of MDs into these superfamilies. In this Letter, we put forward a hypothesis which may give at least a partial answer to this question.

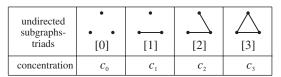
The basic idea of our approach is to notice that a MD gives a coarse-grained description of a network: one integrates out many internal degrees of freedom. Therefore, there exists an entropy corresponding to each particular motif distribution. The real observed state of a network is, then, as usual, a compromise of some sort of "energy" specific to a particular system (e.g., a selection pressure) and this generic entropic landscape. If some distribution is entropically advantageous, it will occur more often for different possible realizations of energy and will act as an effective trap for the network dynamics. Such entropically favorable distributions are the islands of stability in a sea of all MDs, as conjectured in [25].

We consider, here, statistical properties of the simplest motifs—the three-vertex subgraphs or triads—in N-vertex random networks. The microscopic configuration of the network is completely defined by N(N-1)/2 Boolean variables denoting presence or absence of edges, while the MD is described by a vector \mathbf{M} whose elements are the numbers of triads $\{M_0, M_1, ...\}$, or by vector \mathbf{c} of triad concentrations, $\{c_0, c_1, ...\}$, $c_i = M_i/M$, where M = N(N-1)(N-2)/6 is the total number of triads in the network [26]. For directed and nondirected graphs, there are 16 and 4 different triad types, respectively; however, not all of them are independent (see below).

To study the entropy of a network as a function of MD, we model it in an auxiliary external field, **H** coupled to the MD. That is, we consider an ensemble of networks with the partition function $W(\mathbf{H}) = \sum_{X}' e^{-\mathcal{H}(X)/T}$ with $\mathcal{H}(X) =$ $-\mathbf{HM}(X)$, where the sum runs over all microscopic configurations X and ' designates conditions imposed on the configurational space (e.g., the conservation of the network degree distribution). The unbiased case of equiprobable network configurations corresponds to H = 0. This model is athermic: in the absence of the external field, the partition function is purely a combinatorial object with no internal interactions and no temperature dependence (compare this, e.g., with athermic liquid mixtures or nonself-intersecting random walks [27]); therefore, it seems natural to introduce a normalized dimensionless field $\mathbf{h} = \mathbf{H}/T$, which is the only external variable governing the behavior of the model. This approach, reminiscent of the biased molecular dynamics[28,29], allows us, by varying **h**, to skew the motif distribution and, thus, sample the states of a network which are, otherwise, unaccessible. As a result, we obtain a full free energy landscape of a network as a function of motif distribution. To equilibrate a network, we permute links as shown in Fig. 1(b), preserving the node degree distribution [30], and use the Metropolis algorithm to accept or reject single steps.

For $\mathbf{h}=\mathbf{0}$, the system lives in the largest entropic basin corresponding to some equilibrium distribution of motifs. As $|\mathbf{h}|$ is increased, the motif distribution gets gradually more skewed. In the limit $|\mathbf{h}| \to \infty$, the entropic effects become irrelevant, and the network approaches the state with the largest possible value of the "energy," i.e., of the product $\mathbf{h}\mathbf{M}$. Depending on the particular shape of the entropy function, the motif vector \mathbf{M} can be either a smooth function of external field \mathbf{h} , or it can undergo abrupt jumps at some particular values of \mathbf{h} .

Here, we look at the simplest case of the triad distribution in undirected networks. The randomization procedure [30] consists of repeated permutations of randomly chosen pairs of links—see Fig. 1. Each permutation changes the number of triads of different types: four vertices encircled in Fig. 1 belong to 4 + 6(N - 4) +2(N-4)(N-5) different triads, of which (i) four consist of three vertices constructed from the set (1234), all of them are of type [1] and do not change in the elementary step, (ii) 6(N-4) triads including two vertices from (1234) and one external vertex, some of them change as a result of a permutation (we call these changes "elementary reactions" in what follows, not to be confused with "permutation steps" which, generally speaking, consist of many simultaneous elementary reactions), and (iii) 2(N-4)(N-5) triads including only one of the vertices from the set (1234), they do not change in a permutation. A direct check shows that there exists only one type of elementary reaction which changes the numbers of the triad types



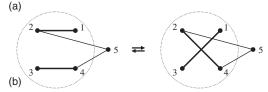


FIG. 1. (a) Possible triads in a nondirected network. (b) Single link permutation: links (12) and (34) are removed, and links (13) and (24) are created. Triad {135} goes from type [0] to type [1], triads {125, 345}—from type [2] to type [1], and triad {245}—from type [2] to type [3]: three new triads of type [1] and one triad of type [3] are created instead of three triads of type [2] and one of type [0], compare (1).

$$[0] + 3[2] \rightleftarrows [3] + 3[1].$$
 (1)

When a new [0] triplet is formed, it always coincides with the formation of three triplets of type [2], and elimination of one [3] triplet and three [1] triplets (see Fig. 1). Equation (1) sets a connection between the time derivatives of the triads concentrations

$$3\dot{M}_0(t) = \dot{M}_2(t) = -3\dot{M}_3(t) = -\dot{M}_1(t),\tag{2}$$

where $\dot{M}_i \equiv (dM_i/dt)$ (i=0...3). Since only one triad concentration is independent, the undirected network is effectively 1D in terms of triad distributions. This means that three independent combinations of $c_0, ..., c_3$ exist, which are conserved under (2), one can choose them as

$$I_1 = \sum_{i=0}^{3} M_i = M;$$
 $I_2 = \sum_{i=0}^{3} iM_i = 3pM;$ (3)
 $I_3 = \frac{1}{2}(M_0 + M_3).$

 I_1 and I_2 are, respectively, the total number of triads, M, and links, 3pM, and p is the fraction of links formed in the system. The invariance of I_3 follows from the conservation of the vertex degree distribution ([22], Supplemental Material). We parametrize the one-dimensional evolution of motifs by a variable $m = (1/2)(M_3 - M_0)$, or c = m/M. The external field h coupled to m is also one dimensional, and the dimensionless network energy in the external field reads

$$E = -hm = -\frac{1}{2}h(M_3 - M_0). \tag{4}$$

According to the detailed balance rule, in an equilibrium network subject to energy E, for the probabilities of

forward, p_+ , and backward, p_- , permutation steps, one has $p_+/p_-=e^{-\Delta E}$, where ΔE is the overall energy change due to all elementary reactions invoked by this permutation step. In the Metropolis algorithm, this is achieved by accepting the reaction with probability 1 if it decreases E, and with probability $\exp\{-\Delta E\}$, otherwise.

Since each permutation implies many simultaneous elementary reactions (1), the reactions are, generally speaking, not independent. A natural first approximation is, nevertheless, to neglect correlations between them and apply the law of mass actions (LMA) [31,32] to the chemical kinetics described by Eq. (1). It is straightforward to obtain

$$K \equiv \frac{p_{+}}{p_{-}} = e^{h} = \frac{c_{3}c_{1}^{3}}{c_{0}c_{2}^{3}} = \frac{(A+c)(2-3p-A+3c)^{3}}{(A-c)(3p-1-A-3c)^{3}}, \quad (5)$$

where $A = I_3/M$ [see (3)], and $K = e^h$ is the equilibrium reaction constant (K = 1 for h = 0).

The mean-field equation (5) is applicable to networks with any degree distribution. In what follows, we concentrate on a particular example of conventional Erdős-Rènyi (ER) networks [8] with bond formation probability p. The triad concentrations, in this case, are given by

$$\bar{c}_0 = (1-p)^3; \qquad \bar{c}_1 = 3(1-p)^2 p;
\bar{c}_2 = 3(1-p)p^2; \qquad \bar{c}_3 = p^3;
\bar{A} = \frac{1}{2}(\bar{c}_0 + \bar{c}_3) = \bar{I}_3/M = [p^3 + (1-p)^3]/2.$$
(6)

All concentrations, as expected, satisfy (5) with K=1. Starting with different p, we perform a randomization at nonzero fields $h=\ln K>0$. The typical resulting dependences $\phi(h)=c(h)-c(h=0)$ are shown in Figs. 2(a) and 2(b) for networks created at p=0.05 [2(a)] and p=0.35 [2(b)]. The dashed lines show the $\phi(h)$ dependence as given by solving (5) with $A=\bar{A}$ given by (6). At high h, the $\phi(h)$ curve saturates due to the depletion of triads of type 2

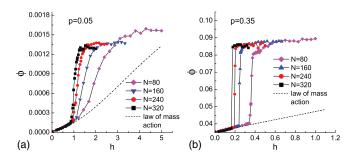


FIG. 2 (color online). The motif distribution $\phi(h) = c - c(h = 0)$ in ER networks with p = 0.05 (a) and p = 0.35 (b), reequilibrated at different h. Solid lines—numerical results for networks of sizes N = 80 (magenta diamonds), 160 (blue triangles), 240 (red circles), and 320 (black squares), dashed line—predictions of the LMA (5).

with growing c. In the vicinity of h = 0, the numerical results are in good agreement with the LMA (5), while for larger h, a sharp change in $\phi(h)$, not predicted by the LMA, is registered [33].

The discrepancy between theory and numerical experiment is due to correlations between elementary reactions constituting a single permutation step. To check this, we study [34] the distribution of the number of forward and backward elementary reactions constituting a single permutation step and show that while, for small h, this distribution is nearly Gaussian (as one expects for independent elementary reactions), in large fields, the distribution acquires a peculiar bimodal shape signaling strong correlations between elementary reactions.

For large p, the transition between two regimes is very narrow, reminiscent of the first-order phase transition; in [34], we show that hysteresis typical for such transitions is also observed. For smaller p, the transition is less narrow due to larger fluctuations, although its width decays with increasing N. For p = 0.05, the difference between the specific free energies of two competing phases is relatively small, and both phases could coexist within the whole transition region (compare to the coexistence of two phases in a liquid-liquid 1st order phase transition reported in [36]). To verify this, we have measured (see [34]) the distribution of links in the network with respect to the number of triangles (type [3] subgraphs) they belong to. The clear bimodality of the distribution shows that there is, indeed, a coexistence of triangle-poor and triangle-rich phases in the system.

We describe the observed transition via a phenomenological Landau-type theory [32] with ϕ as an order parameter. Since there is no $\phi \leftrightarrow -\phi$ symmetry, the Landau expansion of the free energy is expected to include odd and even powers of ϕ , giving up to the fourth order

$$H = H_0 - Mh\phi;$$

$$H_0/M = \frac{\chi}{2}\phi^2 - \frac{b(N,p)}{3}\phi^3 + \frac{g(N,p)}{4}\phi^4 + o(\phi^4).$$
 (7)

Here, H_0 is a purely combinatorial (i.e., temperature and field-independent) free energy of a network with given motif distribution in the absence of the external field; by definition, it has a minimum at $\phi=0$. The zero-field susceptibility χ , according both to LMA (5) and computer simulations at small h, is N independent and equals $\chi \equiv (\partial^2 H(\phi)/\partial(\phi)^2)|_{h=0} = (\partial h(c)/\partial c)|_{h=0}$ giving

$$\chi = \frac{1}{\bar{c}_0} + \frac{9}{\bar{c}_1} + \frac{9}{\bar{c}_2} + \frac{1}{\bar{c}_3} = \frac{1}{p^3 (1-p)^3},\tag{8}$$

where the last equality is true only for ER networks. Higher order coefficients b(N, p) and g(N, p) are expected to be N dependent.

The structure of the Hamiltonian (7) allows for a first order phase transition. Indeed, equilibrium value of ϕ is defined by minimizing $H(\phi)$ in (7), and is given implicitly by the equation

$$\chi \phi - b\phi^2 + g\phi^3 = h. \tag{9}$$

For $b^2 < 3g\chi$, this equation has a single solution for any h (the free energy H_0 is always convex), but for $b^2 > 3g\chi$, there exists a region with three solutions, which corresponds to two competing minima of the free energy and one unstable maximum in between. In the thermodynamic limit $N \to \infty$, the transition occurs when the values of H match in the minima, while in smaller systems, the fluctuations smear the transition.

The ansatz (9) allows us to fit numerical $\phi(h)$ dependences as shown in Fig. 3. While for p=0.35, the transition looks as a discontinuous jump of the order parameter at a point defined by the Maxwell rule, for p=0.05, the behavior in the transitional region is given by a linear combination of two competing phases.

Importantly, the transition point shifts to lower h with increasing size of the network seemingly as $N^{-\alpha}$ with $\alpha = 0.5 \pm 0.1$ [34]. In turn, the width of the transition decays more rapidly with N, approximately as N^{-1} .

To summarize, projection of microscopic (in terms of network nodes) onto macroscopic (in terms of triad concentration) description of a network gives rise to a nontrivial dependence of the entropy on a given motif distribution. In the case of a nondirected network, the macroscopic description is one dimensional. In the presence of an external field h, the equilibrium value of the motif concentration, $\phi = c(h) - c(h = 0)$, is determined by the balance of energy associated with h and entropy originating from mapping from microscopic to macroscopic description. If the entropic landscape is concave, a phase transition into a state with highly skewed motif distribution occurs at some critical h_c . This transition, observed numerically for ER networks, violates the law of

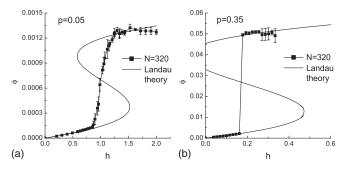


FIG. 3. Comparison of the numeric dependence $\phi(h)$ for N=320 (squares) and the best fitting Landau theory (9) (lines). (a) p=0.05, $\chi=p^{-3}(1-p)^{-3}=9330$, $b=1.72\times10^7$, $c=8.45\times10^9$; (b) p=0.35, $\chi=p^{-3}(1-p)^{-3}=85$, $b=4.46\times10^3$, $c=5.7\times10^4$.

mass action due to correlations between elementary reactions (1) in strong fields.

To check the generality of this result, we modified the elementary permutation rules allowing an edge connecting two arbitrary vertices (i, j) to be switched to some other pair (i, k) $(k \neq i, j)$. Under this dynamics, the nodes' degrees are not conserved, and the integral I_3 in (3) is absent. Accordingly, the dynamics in the motif space becomes effectively two dimensional with elementary reactions

$$2[2] \rightleftarrows [1] + [3];$$
 $2[1] \rightleftarrows [0] + [2].$

However, application of an external field \mathbf{h} (which in this case is a 2D vector) still leads to a transition to a localized motif distribution [37]. Another example of a similar transition, which corresponds to the dynamics with random creation and elimination of bonds with the external field coupled to triangles; $\mathbf{h} = 0, 0, h$ in the 3D space of motifs was first observed in [38] and discussed theoretically in [39]. Note, also, the conjugate effect of bimodality in the triangle parameter estimates in exponentially distributed networks reported in [40]. In [34], we provide preliminary data on networks with fixed scale-free distributions of edge degrees, showing that similar localization transition happens there as well.

The phenomenon of localization of the motif distribution under an external field into distinct entropic traps is apparently wide spread. We conjecture that stable motif profiles constituting superfamilies [22] may correspond to such stability islands inherent to the complicated underlying entropic landscape of a multidimensional motif space of directed networks. Notably, in the presence of well-defined stability islands, new information about the network structure may cause a drastic reclassification of a network from one family to another (akin that observed in [24]) rather than a gradual change in its motif profile.

The concept of entropically induced localization may be instrumental in various other fields. Compare it, for example, with the Eigen model of biological evolution in the space of heteropolymer sequences [41]. There, the localization-delocalization transition, known as the "error catastrophe," separates the state localized in the vicinity of a preferred pattern and one where it is completely random [42–44]. The transition occurs due to an interplay between the attraction to a pointlike potential well and the entropic repulsion from this well due to the exponential growth of the number of states with increasing Hamming distance from the well. In our case, a complimentary phenomenon occurs: the entropic landscape of a system acts as a source of effective attractive traps, while the uniform external field regulates the transitions between trapped states. It seems that trapping of a complex system in stability islands due to competition between selection and randomness provides a generic mechanism of localization in complex biological and social systems.

The authors are grateful to M. Kardar, P. Krapivsky, A. Mikhailov, and K. Sneppen for valuable discussions. This work was partially supported by the Grants No. ANR-2011-BS04-013-01 WALKMAT, No. FP7-PEOPLE-2010-IRSES 269139 DCP-PhysBio, as well as by a MIT-France Seed fund and the Higher School of Economics program for Basic Research.

- [1] R. Albert and A.-L. Barabsi, Rev. Mod. Phys. **74**, 47 (2002).
- [2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biologicals Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
- [3] A.-L. Barabasi and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).
- [4] J. P. Scott, Social Network Analysis: A Handbook (Sage Publications, Thousands Oaks, CA, 2000).
- [5] M. E. J. Newman and J. Park, Phys. Rev. E 68, 036122 (2003).
- [6] M. O. Jackson, Social and Economic Networks (Princeton University Press, Princeton, NJ, 2008).
- [7] B. Bollobas, Modern Graph Theory (Springer-Verlag, New York, 1998).
- [8] P. Erdős and A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. 5, 17 (1960).
- [9] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).
- [10] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. 85, 4629 (2000).
- [11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. 85, 4633 (2000).
- [12] G. Palla, L. Lovász, and T. Vicsek, Proc. Natl. Acad. Sci. U.S.A. 107, 7640 (2010).
- [13] D. J. Watts and S. H. Strogatz, Nature (London) 393, 440 (1998).
- [14] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. E **64**, 051903 (2001).
- [15] G. Balazsi, A. Barabasi, and Z. Oltvai, Proc. Natl. Acad. Sci. U.S.A. 102, 7841 (2005).
- [16] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. 31, 64 (2002).
- [17] G. Conant and A. Wagner, Nat. Genet. 34, 264 (2003).
- [18] A. Maayan, S. Jenkins, S. Neves, A. Hasseldine, E. Grace *et al.*, Science **309**, 1078 (2005).
- [19] M. Csete and J. Doyle, Science **295**, 1664 (2002).
- [20] S. Mangan and U. Alon, Proc. Natl. Acad. Sci. U.S.A. 100, 11980 (2003).
- [21] S. Wuchty, Z. Oltvai, and A. Barabasi, Nat. Genet. 35, 176 (2003).

- [22] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science 303, 1538 (2004).
- [23] U. Alon, Nat. Rev. Genet. 8, 450 (2007).
- [24] J. Sanz, E. Cozzo, J. Borge-Hothoefer, and Y. Moreno, BMC Syst. Biol. 6, 110 (2012).
- [25] V. A. Avetisov, S. K. Nechaev, and A. B. Shkarin, Physica (Amsterdam) 389A, 5895 (2010).
- [26] In [22], the authors use renormalized densities of the motifs. In our notation, the components of their motif vector \mathbf{Z} are equal to $Z_i = (c_i c_i^0)/\chi_i = \Delta c_i/\chi_i$, where $c_i^0 = c_i(h = 0)$, and χ_i is the susceptibility of ith concentration to an external field h_i coupled to it, taken at $h_i = 0$. This linear shift does not change the results presented in our work.
- [27] J. des Cloizeaux and G. Jannink, *Polymers in Solution: Their Modelling and Structure* (Oxford University Press: New York, 2010).
- [28] M. Müller and K. Daoulas, J. Chem. Phys. 128, 024903 (2008).
- [29] Y. Deng and B. Roux, J. Chem. Phys. **128**, 115103 (2008).
- [30] S. Maslov and K. Sneppen, Science 296, 910 (2002).
- [31] K. A. Connors, Chemical Kinetics: the Study of Reaction Rates in Solution, (VCH Press, New York, 1990).
- [32] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Pergamon, Oxford, 1981).
- [33] There is no $h \leftrightarrow -h$ symmetry in the problem (however, for ER networks, there is a symmetry with respect to simultaneous replacement $h \leftrightarrow -h$; $p \leftrightarrow 1-p$); here, we provide the results for h > 0 since it is where the sharp transition is observed.
- [34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.113.095701 for details, which includes Ref. [35].
- [35] E. A. Bender and E. R. Canfield, J. Comb. Theory Ser. A 24, 296 (1978).
- [36] S. Harrington, R. Zhang, P. H. Poole, F. Sciortino, and H. E. Stanley, Phys. Rev. Lett. 78, 2409 (1997).
- [37] V. A. Avetisov et al. (to be published).
- [38] D. Strauss, SIAM Rev. 28, 513 (1986).
- [39] J. Park and M. E. J. Newman, Phys. Rev. E **72**, 026136 (2005).
- [40] A. G. Chandrasekhar and M. O. Jackson, arXiv:1210.7375.
- [41] M. Eigen, Naturwissenschaften 58, 465 (1971).
- [42] S. Galluccio, R. Graber, and Y.-C. Zhang, J. Phys. A 29, L249 (1996).
- [43] L. Peliti, Europhys. Lett. 57, 745 (2002).
- [44] M. Kolár and F. Slanina, Physica (Amsterdam) **313A**, 549 (2002).