Scaling of the distribution of shortest paths in percolation

Nikolay V. Dokholyan[1*] , Youngki Lee,[1] Sergey V. Buldyrev,[1]
Shlomo Havlin,[2] Peter R. King[3,4], and H. Eugene Stanley[1]
[1] *Center for Polymer Studies and Physics Dept., Boston University, Boston, MA 02215 USA*
[2] *Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan, Israel*
[3] *BP Exploration Operating Company Ltd., Sunbury-on-Thames, Middx., TW16 7LN, UK*
[4] *Department of Engineering, Cambridge University, Cambridge, UK.*
(*J. Stat. Phys.*, **93:** 603–613 (1998).)

We present a scaling hypothesis for the distribution function of the shortest paths connecting any two points on a percolating cluster which accounts for *(i)* the effect of the finite size of the system, and *(ii)* the dependence of this distribution on the site occupancy probability p. We test the hypothesis for the case of two-dimensional percolation.

Key words: percolation, minimal path, chemical distance, finite size effect, off-criticality.

Dedicated to Professor Leo Kadanoff on the occasion of his 60$^{\rm th}$ birthday.

The *chemical distance or minimal path*, ℓ, between two sites is defined as the shortest path on a percolating cluster connecting the two sites. The quantity of interest here is the conditional probability, $P(\ell|r)$, that two sites taken from the same cluster, separated by geometrical distance r, are ℓ chemical distance away. For example, in oil recovery the first passage time from the injection well to a production well a distance r away is related to $P(\ell|r)$. In many realistic problems where the disordered media controls a transport process, dynamic properties such as conductivity and diffusion can also be expressed in terms of chemical distance, in the case of loopless aggregates or aggregates for which loops can be neglected [1]. It is known that the average chemical distance $\bar{\ell}$ scales as $r^{d_{\rm min}}$, where various estimates of $d_{\rm min}$ are $d_{\rm min} \approx 1.130 \pm 0.005$ [2] and $d_{\rm min} \approx 1.1307 \pm 0.0004$ [3]. There has been an extensive theoretical and computer work done on studying the scaling of $P(\ell|r)$ [4–6], but the complete scaling form of $P(\ell|r)$ which accounts for the finite size effect, and off-critical behavior has not been studied in detail.

We start by recalling the scaling form for a different but related function, the conditional probability $P(r|\ell)$, the probability that two sites, separated by chemical distance ℓ, are a geometric distance r away. For isotropic media this probability distribution was studied extensively (see [1,4–8]). In analogy with the theory of self-avoiding random walks (SAWs) [9–11], it was proposed [1] that

$$P(r|\ell) = A_l \left(\frac{r}{\ell^{\tilde{\nu}}}\right)^{g_r} f_0\left(\frac{r}{\ell^{\tilde{\nu}}}\right), \tag{1}$$

where $A_\ell \sim 1/\ell^{\tilde{\nu}}$, $\tilde{\nu} = 1/d_{\rm min} = 0.88 \pm 0.02$, $g_r = 2.2 \pm 0.3$ [12] and the scaling function $f_0(x) = \exp(-ax^{\tilde{\delta}})$ with $\tilde{\delta} = (1-\tilde{\nu})^{-1}$. The relation between exponent g_r and other known exponents was proposed by Havlin [13], who used arguments similar to those of de Gennes [9] for SAWs. The distribution $P(r|\ell)_{r=1}$ determines the probability of the minimal path returns to a nearest neighbour site of the origin, which is from Eq. (1)

$$\frac{1}{\ell^{\tilde{\nu}}}\left(\frac{1}{\ell^{\tilde{\nu}}}\right)^{g_r} = \frac{1}{\ell^{\tilde{\nu}+\tilde{\nu}g_r}}. \tag{2}$$

On the other hand, the total number of clusters $N(\ell)$ of chemical size ℓ scales as $M^{-\tau+1}(dM/d\ell)$, where M is the total mass of the cluster, which scales as $\ell^{\tilde{\nu}d_f}$, where d_f is the fractal dimension of the percolation cluster, and $\tau = 1 + d/d_f$ (see e.g. [1]). Hence

$$N(\ell) \sim \ell^{\tilde{\nu}d_f(-\tau+1)+\tilde{\nu}d_f-1} \sim \ell^{\tilde{\nu}d_f(-\tau+2)-1}. \tag{3}$$

Due to the fact that in the mean field regime the effect of the excluded volume effect is absent, we conjecture[1] that the number of clusters $N_0(\ell)$ in which the minimum path ℓ returns to a nearest neighbour site of the origin obeys the

* *Correspondence to:* N. V. Dokholyan. E-mail: dokh@miranda.bu.edu
[1]This conjecture is made in analogy with the number of self-avoiding random walks coming to the origin, which scaling lacks prefactor $N^{\gamma-1}$, i. e. $\gamma = 1$ as in mean field regime (see [9]).

mean field values for which $\tilde{\nu} = 1/2$, $\tau = 5/2$, and $d_f = 4$. The probability of a cluster with the path returning to a nearest neighbour site of the origin is proportional to the fraction of paths $N_0(\ell)/N(\ell)$ which return to a nearest neighbour site of the origin and the probability of an end of the path in a volume V to be at the specific site $1/V$. Thus,

$$P(r|\ell)_{r=1} \sim \frac{N_0(\ell)}{N(\ell)} \frac{1}{V}, \tag{4}$$

where V is the volume of the cluster $V \sim \ell^{d\tilde{\nu}}$. The expressions (2) and (4) should be equal and thus

$$\frac{1}{\ell^{\tilde{\nu}+\tilde{\nu}g_r}} \sim \frac{1}{\ell^{\tilde{\nu}d_f(-\tau+2)+1+d\tilde{\nu}}}. \tag{5}$$

Finally we have

$$g_r = d_f(-\tau+2) + d_{\min} + d - 1 = d_f + d_{\min} - 1. \tag{6}$$

This value is within the error bars found numerically for g_r in $d = 2$ and $d = 3$ [14]. Numerical simulations for percolation clusters in $d = 2$ are in good agreement with the above scaling form [1,4,15].

We start by recalling a scaling Ansatz for $P(\ell|r)$, which has been developed in [4–6]. Exactly at $p = p_c$ in the infinite system ($L = \infty$), in analogy to (1),

$$P(\ell|r) = A_r \left(\frac{\ell}{r^{d_{\min}}}\right)^{-g_\ell} f_1\left(\frac{\ell}{r^{d_{\min}}}\right) \tag{7}$$

where $f_1(x) \sim \exp(-ax^{-\phi_1})$ is the scaling function corresponding to f_0 and $A_r \sim \frac{1}{r^{d_{\min}}}$ is the normalization factor. The conditional probabilities $P(r|\ell)$ and $P(\ell|r)$ are related by

$$P(r|\ell)P(\ell) = P(\ell|r)P(r) = P(r,\ell)\,, \tag{8}$$

where $P(r) = \int P(r,\ell)d\ell$ and $P(\ell) = \int P(r,\ell)dr$. For the case of the infinite percolation cluster, i. e. when we restrict ourself only to the points on the infinite cluster, $P(r)$ and $P(\ell)$ scale as [15]

$$P(\ell) \sim \ell^{d_\ell - 1}\,,\; P(r) \sim r^{d_f - 1}\,, \tag{9}$$

where $d_\ell = d_f/d_{\min}$. Thus,

$$P(\ell|r) \sim P(r|\ell)\frac{\ell^{d_\ell - 1}}{r^{d_f - 1}}\,. \tag{10}$$

Substituting Eqs. (1) and (7) into Eq. (10), we find that

$$f_1(x) \sim \exp(-ax^{-\phi_1}) \tag{11}$$

where

$$\phi_1 = \tilde{\delta}\tilde{\nu} = \tilde{\nu}(1-\tilde{\nu})^{-1} = 1/(d_{\min} - 1), \tag{12}$$

and [15]

$$g_\ell - 1 = (g_r - 1)\tilde{\nu} + (2 - d_f)\tilde{\nu}. \tag{13}$$

Using Eq. (6) we find $g_\ell = 2$ for all $d \geq 2$. Note, that the numerical value for $g_\ell \approx 2.04$ found here is very close to this prediction.

Let us now consider the scaling of $P'(\ell|r)$, which is defined in the same way as $P(\ell|r)$, but for any two randomly chosen points, separated by geometrical distance r and not necessarily belonging to an infinite cluster. In this case,

$$P'(r) \sim rG(r)\,, \tag{14}$$

where $G(r)$ is a pair connectedness function and the factor r comes from the summation over all points which are equidistant from the origin. The scaling of the $G(r)$ is described in [16]. At $p = p_c$:

$$G(r) \sim r^{2-d-\eta}\,, \tag{15}$$

where $\eta = 4 - 2d_f$. The exact value of $\eta = 5/24 \approx 0.21$ is known for $d = 2$. Substitution of Eqs. (7) and (14) into Eq. (8), and normalization, yields

$$P'(r,\ell) \sim rG(r)\frac{1}{r^{d_{\min}}}\left(\frac{\ell}{r^{d_{\min}}}\right)^{-g'_l} f_1\left(\frac{\ell}{r^{d_{\min}}}\right) \sim r^{1-d_{\min}-\eta}\left(\frac{\ell}{r^{d_{\min}}}\right)^{-g'_l} f_1\left(\frac{\ell}{r^{d_{\min}}}\right). \tag{16}$$

The integration of $P'(r,\ell)$ over r gives

$$P'(\ell) \sim \ell^{2\tilde{\nu}-1-\tilde{\nu}\eta}\,. \tag{17}$$

Next, we substitute Eqs. (16) and (17) to Eq. (8) and compare with (1). We obtain

$$P'(\ell|r) = A_r\left(\frac{\ell}{r^{d_{\min}}}\right)^{-g'_\ell} f_1\left(\frac{\ell}{r^{d_{\min}}}\right), \tag{18}$$

where

$$g'_\ell - 1 = (g_r - 1)\tilde{\nu} + 2(2 - d_f)\tilde{\nu}. \tag{19}$$

Thus from (13) and (19),

$$g'_\ell - g_\ell = (2 - d_f)\tilde{\nu}. \tag{20}$$

The latter relation between exponents g'_ℓ and g_ℓ has simple probabilistic meaning, since a pair of two randomly selected points separated by a chemical distance ℓ should belong to the cluster of chemical size $\ell_0 > \ell$. The probability of this event scales as $\ell^{-\eta\tilde{\nu}/2}$. Once two points belong to such a cluster, the probability that their chemical distance is equal to ℓ scales the same way as on an infinite cluster and is proportional to ℓ^{-g_ℓ}. Hence the probability that two randomly selected points are separated by a chemical distance ℓ is proportional to the product of these two probabilities $\ell^{-g_\ell-\eta\tilde{\nu}/2}$, which, by definition, is $\ell^{-g'_\ell}$. Hence $g'_\ell = g_\ell + \eta\tilde{\nu}/2$.[2] We summarize the values of the above exponents for $d = 2$ in Table I.

Next we study the finite size effect and off-critical behavior using Eq. (7) as starting point. We propose that for $\xi > r$, $P'(\ell|r)$ has the following scaling form

$$P'(\ell|r) \sim \frac{1}{r^{d_{\min}}}\left(\frac{\ell}{r^{d_{\min}}}\right)^{-g'_\ell} f_1\left(\frac{\ell}{r^{d_{\min}}}\right) f_2\left(\frac{\ell}{L^{d_{\min}}}\right) f_3\left(\frac{\ell}{\xi^{d_{\min}}}\right), \tag{21}$$

where the scaling functions are $f_1(x) = \exp(-ax^{-\phi_1})$, $f_2(x) = \exp(-bx^{\phi_2})$ and $f_3(x) = \exp(-cx)$. Here ξ is a characteristic length of pair connectedness function and has a power-law dependence of occupancy p as

$$\xi \sim |p - p_c|^{-\nu}. \tag{22}$$

The first function f_1 accounts for the lower cut-off due to the constraint $\ell > r$, while f_2 and f_3 account for the upper cut-off due to the finite size effect and due to the finite correlation length respectively. Either f_2 and f_3 becomes irrelevant, depending on which of the two values L or ξ is greater. For $L < \xi$, f_2 dominates the upper cut-off, otherwise f_3 dominates. We assume the independence of the finite size effect and the effect of the concentration of the vacant sites, so the Eq. (21) can be represented as a product of the terms which are responsible for the finite size effect (f_2) and the effect of the concentration (f_3). Our simulations confirm this assumption numerically.

There is no particular reason for the choice of the simple exponential decay of the function f_3, but we find that this function is a straight line on a semilogarithmic plot (see discussion below). However, due to the fact that the pair connectedness function $G(r) \sim r^{2-d-\eta}\exp(-r/\xi)$ has simple exponential form [16] it is reasonable to assume that

---

[2] Indeed, $\eta\tilde{\nu}/2 = \beta\tilde{\nu}/\nu$, where β is defined as the exponent of the probability to belong to the infinite cluster, $P_\infty \sim |p - p_c|^\beta$. Analogous relation is known, for example, for the exponents γ and $1/\sigma$, which define respectively the scaling of the average and the typical size of the cluster: $1/\sigma = \gamma + \beta$.

function f_3 has a form of a stretched exponential $f_3 \sim \exp(-cx^{\tilde{\nu}})$. However, since the exponent $\tilde{\nu} \approx 0.88$ is close to 1, we are not able to resolve this question in our simulations.

Next, using the Leath method [17–19], which corresponds to calculation of $P'(\ell, r)$ [3], we numerically test the scaling conjecture Eq. (21) exactly at the percolation threshold $p = p_c$. In this case, $\xi = \infty$ so $f_3 = \text{const}$. Hence Eq. (21) reduces to

$$P'(\ell|r) \sim \frac{1}{r^{d_{\min}}} \left(\frac{\ell}{r^{d_{\min}}}\right)^{-g'_\ell} f_1\left(\frac{\ell}{r^{d_{\min}}}\right) f_2\left(\frac{\ell}{L^{d_{\min}}}\right), \quad (p = p_c). \tag{23}$$

Indeed, Fig. 1a shows that $P'(\ell|r)$ has power law behavior for $r^{d_{\min}} < \ell < L^{d_{\min}}$ and rapidly vanishes for $\ell < r^{d_{\min}}$ and for $\ell > L^{d_{\min}}$. In order to test Eq. (23) we compute rescaled probability distribution

$$\Phi\left(\frac{\ell}{r^{d_{\min}}}\right) \equiv P'(\ell|r)\ell^{g'_\ell} r^{-d_{\min}(g'_\ell - 1)} = f_1 \cdot f_2, \tag{24}$$

and plot it against scaling variable $x = \ell/r^{d_{\min}}$ (see Fig. 1b). According to Eq. (23),

$$\Phi(x) = A f_1(x) f_2\left[x\left(\frac{r}{L}\right)^{d_{\min}}\right]. \tag{25}$$

Therefore, $\Phi(x)$ should depend only on x and the ratio L/r. Indeed, Fig. 1b shows excellent data collapse for $L/r = 16$, with sharp cutoffs governed for small $x < 1$ by $f_1(x)$ and for large $x > (L/r)^{d_{\min}}$ by $f_2[x(r/L)^{d_{\min}}]$.

In order to test the assumption that the functions f_1 and f_2 are stretched exponentials with exponents ϕ_1 and ϕ_2 we plot $\Pi(x) = \log[A/\Phi(x)]$ versus x in double logarithmic scale for various values of normalization constant A. (See Fig. 2.) If the stretched exponential conjecture is correct, $\Pi(x)$ should have two straight line asymptotes for $\log x \to \infty$ with the slope ϕ_2 and for $\log x \to -\infty$ with the slope $-\phi_1$. The slopes ϕ_1 and ϕ_2 of the straight line fits depend weakly on the value of A. Using $A = 1.65$, we obtain the longest regimes of straight line behavior. For this A we obtained $\phi_1 \approx 7.3$ and $\phi_2 \approx 4.0$. The value of ϕ_1 is in good agreement with Eq.(12), while ϕ_2 seems to be a new exponent.

Finally, in order to test the dependence of $P'(\ell|r)$ on p we obtain data for very large system size L and for several values of $p \neq p_c$. In this case, the upper cutoff of the distribution Eq. (21) is governed by f_3 and the functional form of the rescaled probability Φ is given by

$$\Phi(\ell/r^{d_{\min}}) \sim f_1\left(\frac{\ell}{r^{d_{\min}}}\right) f_3\left(\frac{\ell}{\xi^{d_{\min}}}\right). \tag{26}$$

For large ℓ, we suggest an exponential decay of Φ:

$$\Phi(\ell/r^{d_{\min}}) \sim \exp\left(-c\frac{\ell}{\xi^{d_{\min}}}\right). \tag{27}$$

Indeed, semi-logarithmic plots of $\log \Phi(\ell/r^{d_{\min}})$ versus ℓ shown in Fig. 3a can be approximated by the straight lines with slopes which approach to zero as $p \to p_c$. According to Eq. (27), these slopes $k(p)$ should be proportional to $\xi^{-d_{\min}} = |p - p_c|^{d_{\min}\nu} \approx |p - p_c|^{1.51}$. Fig. 3b shows double logarithmic plots of $|k(p)|$ versus $|p - p_c|$ for $p < p_c$ and for $p > p_c$, which can be well approximated by straight lines with slopes 1.55 and 1.57 in good agreement with scaling conjecture, Eq. (21).

The scaling form, Eq.(21), is limited to the case when $\xi > r$. For $\xi < r$, the finite size effects can be neglected, the power law regime vanishes, and the minimal path can be divided into r/ξ independent blobs each of length $\xi^{d_{\min}}$, so that the distribution $P'(\ell|r)$ approaches Gaussian form with mean $r\xi^{d_{\min}-1}$ and variance $r\xi^{2d_{\min}-1}$.

---

[3] In the Leath method, starting a cluster from the origin corresponds to selection of a random point in the disordered media. Construction of the Leath cluster corresponds to the determination of all the points which are connected to the starting one. In our procedure, we compute $P'(\ell, r)$ by dividing the number of the pair of points $N_{\ell,r}$ which are separated by r and ℓ by the total number of points N in all constructed clusters. The quantity of interest $N_{\ell,r}/N$ is equivalent to the quantity $P'(\ell, r)$, which by definition is the probability that two randomly chosen points, belonging to the same cluster, are separated by r and ℓ.

In many realistic problems an interesting quantity is the first passage time distribution. If we assume that the first passage time t between two points is linearly proportional to shortest chemical distance ℓ between them then we can identify the distribution $P'(\ell|r)$ as a first passage time distribution. This is only true when the spreading velocity through the media remains constant along the path. In the oil recovery problem it is the injection rate, the volume of the fluid pumped into the media per unit time, which is maintained constant, not the velocity itself. We can generalize the relation between first passage time and chemical distance by $t \sim \ell^z$. For the given fractal dimension d_B of the backbone of the percolating cluster we know that $M(r) \sim r^{d_B}$ where $M(r)$ is the total mass of the backbone of the percolation cluster of radius r. Constant pumping rate gives $dt \sim dM \sim r^{d_B-1}dr$. Combining with the relation $\ell \sim r^{d_{\min}}$ and their derivative $d\ell \sim r^{d_{\min}-1}dr$ we can get another scaling relation $z = d_B/d_{\min} = 1.625/1.14 \approx 1.43$. If we substitute $\ell \sim t^{1/z}$ in Eq. (21) it gives the correct scaling form for the first passage time distribution in case of constant injection rate.

Another problem comes from the presence of a second well, which might establish some pressure potential field through the media and bring some anisotropic spreading front toward the production well when we apply some external driving pressure. This can be corrected by calculating the potential field which is a solution of Laplace's equation with two point boundaries and calculating the local velocity on the percolating cluster including the quenched disorder of the percolation cluster. Practically the problem of multiple wells scattered along the oil fields with equal spacing is of particular interest in oil recovery.

In summary, we have studied the scaling properties of the shortest paths distribution for fixed two points on the percolating cluster which accounts for the finite size effect, also off criticality. We propose a plausible scaling hypothesis for the distribution, which is supported by theoretical argument and tested by numerical simulation. The lower and upper cut-offs of the distribution has been numerically observed and fitted successfully by stretched exponential function. Off the critical point the upper cutoff, due to the finite correlation length, becomes a pure exponential form. Further work is underway to test the conjectures and hypothesis developed here.

We thank Prof. D. Stauffer for fruitful discussion, the referee for the helpful suggestions, and BP for the financial support.

---

[1] S. Havlin and D. Ben-Avraham, Adv. Phys. **36**, 695 (1987).
[2] H. J. Herrmann and H. E. Stanley, J. Phys. A: Math. Gen **21**, L829 (1988).
[3] P. Grassberger, J. Phys. A: Math. Gen. **25**, 5475 (1992).
[4] A. U. Neumann and S. Havlin, J. Stat. Phys. **52**, 203 (1988).
[5] A. Bunde, H. E. Roman, S. Russ, A. Aharony, and A. B. Harris, Phys. Rev. Lett. **69**, 3189 (1992).
[6] J.-P. Hovi and A. Aharony, Phys. Rev. E **56**, 172 (1997).
[7] P. Ray, J. Phys. A **18**, L657 (1985).
[8] M. Barma and P. Ray, Phys. Rev. B **34**, 3403 (1986).
[9] P. G. De Gennes, *Scaling Concepts in Polymer Physics* (Ithaca: Cornell University Press) (1979).
[10] C. Domb, Adv. Chem, Phys. **15**, 229 (1969).
[11] M. E. Fisher, J. Chem. Phys. **44**, 616 (1966).
[12] M. Porto, A. Bunde, S. Havlin and H. E. Roman, Phys. Rev. E. **56**, 1667 (1997).
[13] S. Havlin, (*unpublished*).
[14] S. Schwarzer, S. Havlin, and A. Bunde, Phys. Rev. E (*in press*) (1998).
[15] A. Bunde and S. Havlin (eds.), *Fractals and Disordered Systems*, **2nd** Ed., Springer, New York (1996).
[16] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, New York (1971).
[17] P. L. Leath, Phys. Rev. B **14**, 5046 (1976).
[18] Z. Alexandrowicz, Phys. Lett. A **80**, 284 (1980).
[19] R. Pike and H. E. Stanley, J. Phys. A **14**, L169 (1981).

β	ν	η	$d_{\min}$	$\tilde{\nu}$	g_r	g_ℓ	g'_ℓ	ϕ_1	ϕ_2
0.14	1.33	0.20	1.13	0.88	2.2	2.04	2.14	7.3	4.0

Table I: Numerical estimates for the various exponents.

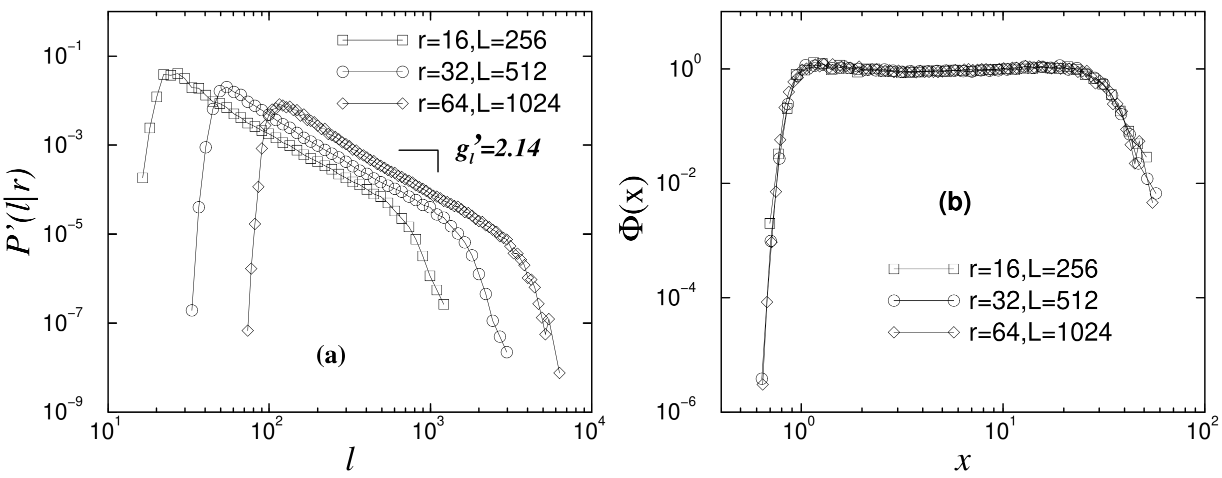

FIG. 1. (a) Log-log plot of $P'(\ell|r)$ for $p = p_c = 0.593$ and for different sets of parameters, (r, L). The straight line regime has slope $g'_\ell = 2.14$. (b) Log-log plot of rescaled probability $\Phi(x) \equiv P'(\ell|r)\ell^{g'_\ell} r^{-d_{\min}(g'_\ell - 1)}$ against rescaled length $x = \ell / r^{d_{\min}}$ using the values, $g'_\ell = 2.14$ and $d_{\min} = 1.14$.

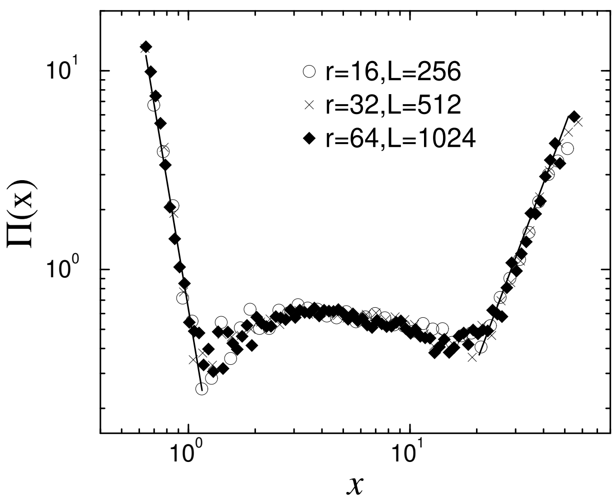

FIG. 2. Log-log plot of transformed probability $\Pi(x) = log[A/\Phi(x)]$ versus $x = \ell / r^{d_{\min}}$. The slopes of the solid lines give the power of the stretched exponential function f_1 and f_2 in Eq. (25). Using the parameter $A = 1.65$, the slopes give $\phi_1 \approx 7.3$ for the lower cut-off and $\phi_2 \approx 4.0$ for the upper cut-off.

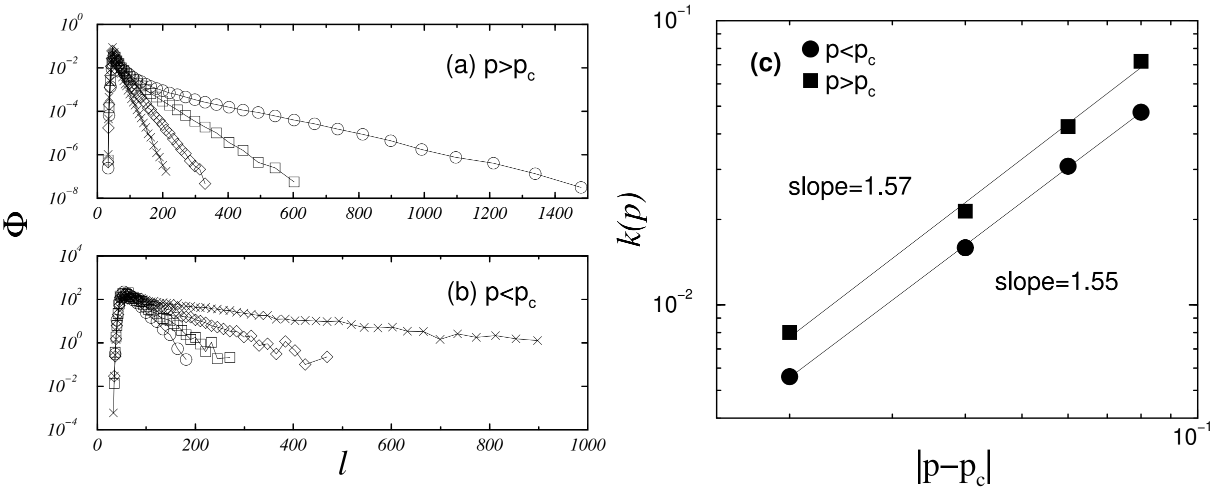

FIG. 3. Semi-logarithmic plot of transformed probability Φ versus ℓ ($r = 32$) shows pure exponential behavior of f_3. (a) (○) denotes $p = 0.613$, (□) $p = 0.633$, (◇) $p = 0.655$ and (×) $p = 0.673$. (b) (○) denotes $p = 0.533$, (□) $p = 0.553$, (◇) $p = 0.573$ and (×) $p = 0.583$. (c) The slope of the log-log plot of the coefficient in exponential function f_3 as a function of $|p - p_c|$ gives the value $\nu d_{\rm min} \approx 1.55$ for $p > p_c$ and 1.57 for $p < p_c$.