Zipf's law in percolation

Makoto S. Watanabe
Faculty of Liberal Arts, Hosei University, Fujimi, Chiyoda-ku, Tokyo 102, Japan
(Received 24 October 1995; revised manuscript received 26 December 1995)

Cluster-size distributions are examined for percolation processes on square and triangular lattices by means of the Monte Carlo simulation. Zipf's law is found to exist in the relation between cluster sizes and their ranks in the size order on finite lattices. The law is predicted to appear just at the thresholds p_c for infinite lattices. The value of critical exponent τ is derived analytically from the sizes with the law. The value is found to be $\tau=2$, showing the validity of the appearance of the law at p_c for the lattices. [S1063-651X(96)11104-1]

PACS number(s): 05.70.Jk, 05.90.+m

In the field of linguistics, Zipf's law [1,2] has been known as a statistical phenomenon concerning the relation between English words and their frequency in literature. The law states that, when we list the words in the order of decreasing population, the frequency of a word is inversely proportional to a rank of the word in a frequency list [1–3]. This relation was found not only in linguistics but also in other fields of science. For instance, the law appeared in distributions of populations in cities, and distributions of areas of lakes [3]. One of the research themes has been to examine what kind of phenomena accompany the law.

Percolation theory, as is well known, has been widely studied in analyzing many problems in various fields of investigations [4–7]. In the theory the values of thresholds and critical exponents have been examined for various lattices [4–16]. Cluster sizes, their numbers, and the structure of perimeters have been analyzed in the theory [17–28]. In these examinations, the Monte Carlo (MC) simulation has played an important role in analyzing the problems.

This paper clarifies that the law exists in the distributions of cluster sizes in percolation processes on square and on triangular lattices with the MC simulation. The law states that the relation between the sizes and their ranks is described by $s_n=An^{-1}$ $(n=1,2,3,\ldots)$, where A is a constant and s_n is the size of rank n in a size list when we arrange the clusters in the order of decreasing size. For instance the size s_2 of the second largest cluster with rank $n=2$ is one-half of the size n_1 of the largest cluster, the size s_3 of the third largest cluster with rank $n=3$ is one-third of the size n_1 of the largest cluster, and so on.

First we constructed an array of vacant sites on the lattice points of the square lattice. Their total number N of the sites was $N=L\times L$, where L is the number of sites on a side of the system. Then we determined occupied sites, the number of which is N_p, at random on the vacant sites. The number N_p was defined as $N_p=N\times p$, where p is the occupation probability of the sites. From the occupied sites determined at random, we examined the structure of clusters constructed in the system. Here the cluster was defined as a group of the occupied sites connected by nearest-neighbor bonds.

For a sample with a random array of clusters, we examined sizes which are defined as the numbers of the occupied sites belonging to each cluster. Then we arranged the sizes in order of decreasing size. For a specified value of p, we repeated this process with varying simulation samples, i.e., varying random number seed, in order to obtain statistically reliable values of the sizes. Then we averaged the sizes for each rank in size lists of the samples: we averaged the sizes for the largest clusters in each sample, averaged them for the second largest clusters, averaged them for the third largest clusters, and so on. From the sizes averaged, we examined the relation between the sizes s_n and their ranks n.

Here let us examine the relation for the system with $L=200$. Filled squares in Fig. 1 show the relation for $p=0.5000$. The top 20 clusters with ranks $1\leq n\leq 20$ were plotted logarithmically in this figure. The linear relation appears in it. The straight line in the figure shows the least square fitting of the 20 clusters. The slope S in it has been found to be $S=-0.35$. Then we increased the value of p, and examined the same relation between them. Open squares in Fig. 1 show the relation for $p=0.5500$. The linear relation

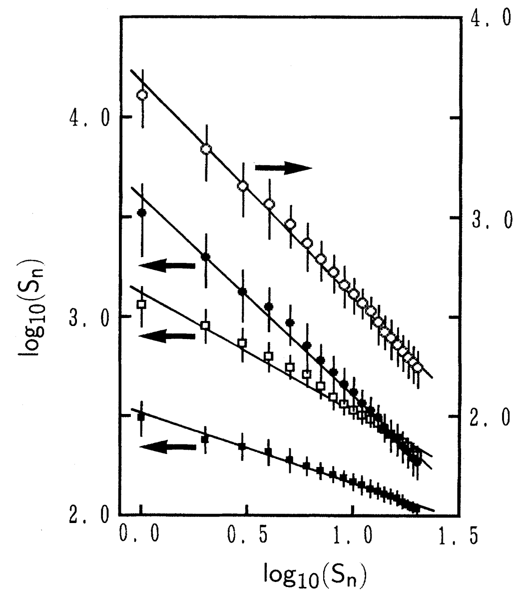

FIG. 1. A log-log plot of the relation between cluster sizes s_n and their ranks n in the size order on the square lattice with $L=200$. Filled squares are for $p=0.5000$, open squares for $p=0.5500$, filled circles for $p=0.5720$, and open circles for $p=0.5760$. Straight lines show the least-square fitting of the relation. Error bars are the standard deviations of the sizes.

 © 1996 The American Physical Society

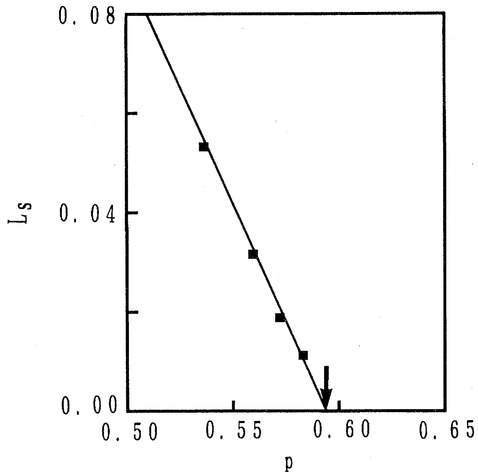

FIG. 2. Occupation probabilities $p^{\mathrm{Zipf}}(L)$ at which Zipf's law appears on the square lattice are plotted against the scaled size $L_s = L^{-1/\nu}$, with $\nu = \frac{4}{3}$. The arrow shows the threshold $p_c = 0.5\,927\,460$ of this lattice.

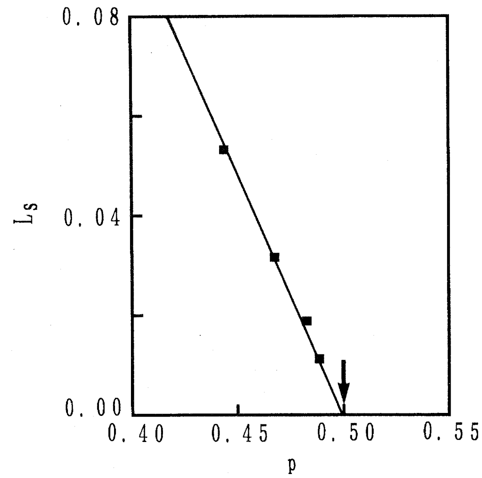

FIG. 3. Occupation probabilities $p^{\mathrm{Zipf}}(L)$ at which Zipf's law appears on the triangular lattice are plotted against the scaled size $L_s = L^{-1/\nu}$, with $\nu = \frac{4}{3}$. The arrow shows the threshold $p_c = \frac{1}{2}$ of this lattice.

appears in it. The slope of the line was $S = -0.61$. Zipf's law does not exist for lines with these values of p, because the law requires the slope $S = -1$.

We increased the value of p in the system, and examined the same relation as above. The filled circles in Fig. 1 show the result for $p = 0.5720$. The linear relation appears in it. The slope S in it has been found to be $S = -1.00$. This shows that the law exists in the distribution on the square lattice with the MC simulation. We continued to increase the value of p in the system. Open circles in Fig. 1 show the relation for $p = 0.5760$. The linear relation appears in it. The slope steepens with increasing p. The slope was $S = -1.08$, which shows that the law disappears with increasing p.

In the above examinations, the system size was finite. Here let us predict the value of p at which the law appears for infinite system size. When we examine the values p_c of thresholds of percolation, we usually extrapolate the effective thresholds $p(L)$, which are the thresholds for finite system size L, as $L \to \infty$ [13,29,30]. This is based on the extrapolation rule $|p_c - p(L)| \propto L^{-1/\nu}$, where p_c is the threshold for infinite lattices and ν the critical exponent for correlation length. The exponent was found to be $\nu = \frac{4}{3}$ [4–6,31–33] for all lattice structures in two dimensions showing the universality. Here let us examine whether or not the rule is valid even for the probability p at which the law appears in the system.

We simulated the processes for the other sizes with $L = 50$, 100, and 400. Then we searched for the values p at which the slope becomes $S = -1$ for each L. Hereafter the probability p with the slope $S = -1$ is denoted by $p^{\mathrm{Zipf}}(L)$ for each L: $p^{\mathrm{Zipf}}(L)$ shows the ratio of the number of the occupied sites to the total number N when the system accompanies the law. Figure 2 shows $p^{\mathrm{Zipf}}(L)$ as a function of the scaled size $L_s = L^{-1/\nu}$ for $L = 50$, 100, 200, and 400. A linear relation appears in it. The values of $p^{\mathrm{Zipf}}(L)$ increase linearly with decreasing L_s. This shows the rule, i.e., $|p^{\mathrm{Zipf}}(L) - p_c| \propto L^{-1/\nu}$ with $\nu = \frac{4}{3}$, is valid even for the probability $p^{\mathrm{Zipf}}(L)$ for the appearance of the law. The line intersects the abscissa just at the arrow which shows the value of threshold $p_c = 0.592\,746\,0$ [16] for this lattice as $L \to \infty$. This shows that the law exists just at the threshold p_c on the square lattice with infinite system size.

Here we simulated the process on a triangular lattice with $L = 200$. A linear relation was observed between cluster sizes s_n and their ranks n for various values of p. The law was found to exist at $p = 0.4824$ for $L = 200$. Then we searched for the values p at which the law exists for the other system sizes with $L = 50$, 100, and 400. Figure 3 shows the relation between $p^{\mathrm{Zipf}}(L)$ and scaled size $L_s = L^{-1/\nu}$ for this lattice. A linear relation is found in this figure. The line intersects the abscissa just on the arrow, which shows the threshold $p_c = 1/2$ for this lattice. Thus the law appears just at p_c for the triangular lattice as $L \to \infty$.

From the statistical point of view, Zipf's law is related to the critical exponent τ which is the Fisher exponent [4,34]. When we appreciate the nature of cluster-size distributions in percolation processes, we have often referred to the value of the exponent. The exponent was obtained from a slope S in a log-log plot of the relation between cluster sizes s and their numbers n_s in systems constructed with the MC simulation [4–6]: the exponent is defined as $\tau = -S$ in the relation. Stauffer examined the value of τ on the triangular lattice just at the transition point $p = 1/2$ of percolation [4,5]. The value was found to be nearly $\tau = 2$ when the system became fully large [4]. The value from series expansions was predicted to be $\tau = 2.055$ [4,5,35], which almost coincides with the above value. Yonezawa, Sakamoto, and Hori examined the values for various lattices in two dimensions such as square, kagome, dice, Penrose, and dual of Penrose lattices [14]. The ratios τ/τ_0 with MC simulations have been between 0.94 and 0.97 at the transition points for these lattices, where $\tau_0 = 2.055$ [14]. This gives the values from $\tau = 1.932$ to 1.993 for the lattices. Thus the exponent τ has been known to be nearly $\tau = 2$ when the system percolates. Here we examine the relation between cluster sizes with Zipf's law and the value of the exponent τ.

Let us consider the distribution of cluster sizes as $k(s)$. The valuable s in $k(s)$ is the size of clusters with the relation $s=An^{-1}$. Here $k(s)ds$, which is equal to $-dn$, is considered to be the number of clusters within size range between s and $s+ds$. This gives the relation $k(s)ds=-dn$ from which the size distribution $k(s)=-dn/ds \propto s^{-2}$ is derived. This shows the distribution with the law is proportional to s^{-2}, indicating the value of the Fisher exponent is $\tau=2$. Since the value $\tau=2$ has given the transition of percolation, Zipf's law with $\tau=2$ exists validly just at the percolation thresholds p_c on these lattices.

Some discussions are made in the last part of this paper. We found that Zipf's law appeared for infinite lattices just at the thresholds p_c. This has been clarified from the extrapolation of $p^{\mathrm{Zipf}}(L)$ which is the probability of the occupied sites when the law appears in finite lattices. Here let us examine whether or not the values of $p^{\mathrm{Zipf}}(L)$ coincide with the effective thresholds $p(L)$ for finite lattices. The thresholds $p(L)$, which are the thresholds for finite system size L, were examined in Ref. [36] for the square lattice between $L=40$ and 100. Open squares in Fig. 4 show the thresholds $p(L)$ for this lattice, which have been defined as the average probability of the sites when the system has first percolated for each L [36]. Error bars are for standard deviations around the average thresholds $p(L)$. The probabilities $p^{\mathrm{Zipf}}(L)$, which are shown as filled squares, are found to be smaller than the open squares, and then the left limit of their fluctuation areas. This indicates that $p^{\mathrm{Zipf}}(L)$ are not always coincident with the thresholds $p(L)$ for finite lattices.

Figure 1 shows the logarithm plot of cluster sizes s_n against the ranks n in a size list of clusters for the square lattice with $L=200$. In the figures, the top 20 clusters with $1 \leqslant n \leqslant 20$ have been plotted. This indicates the clusters for $n>20$, whose sizes are less than $0.05=\frac{1}{20}$ of the largest size, have not been depicted. This number 20 is found to be large. We examined the total number of clusters that appear in the system for the square lattice with $L=200$. We counted the numbers, and found 1470.9 per one sample for $p=0.5720$. Although the number 20 is only about 1.36% of the total number of clusters, the sum of the sizes of the clusters belonging to the 20 clusters has been counted to be as much as 14017.9, which is 61.27% of the total number $N_p=N\times p=22\,880$ of occupied sites scattered in the simulation.

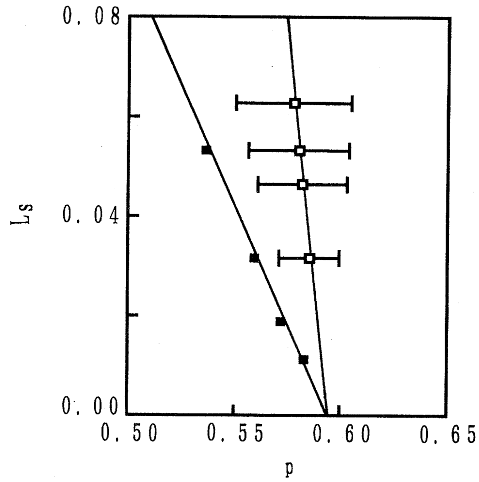

FIG. 4. Effective thresholds $p(L)$ and occupation probabilities $p^{\mathrm{Zipf}}(L)$ are plotted against the scaled size $L_s=L^{-1/\nu}$, with $\nu=\frac{4}{3}$. Open squares are for $p(L)$ and filled ones for $p^{\mathrm{Zipf}}(L)$. Error bars show the standard deviations around $p(L)$.

We found that the law appeared in cluster-size distributions on square and triangular lattices. This appearance is for two dimensional lattices. Here let us clarify whether or not the law exists in the distributions for other dimensions. We simulated the process on a simple cubic lattice with three dimensions. The system size N has been chosen to be $N=20\times20\times20$. The same relation as Fig. 1 has been observed for the lattice. The law has been found to exist at $p=0.3120$ for the lattice. Thus the law appears not only on two dimensional lattices but also on three dimensional lattices in percolation processes with the MC simulations.

[1] D. Crystal, *The Cambridge Encyclopedia of Language* (Cambridge University Press, Cambridge, 1987), p. 86.
[2] H. Takayasu, *Fractals in the Physical Sciences* (Manchester University Press, Manchester, 1990), p. 55.
[3] T. Musha, *Blue Backs B442* (Koudansha, Tokyo, 1980), p. 60 (in Japanese).
[4] D. Stauffer, *Introduction to Percolation Theory* (Taylor & Francis, London, 1985).
[5] D. Stauffer, Phys. Rep. **54**, 1 (1979).
[6] V. K. S. Shante and S. Kirkpatrick, Adv. Phys. **20**, 325 (1971).
[7] G. Deutscher, R. Zallen, and J. Adler, *Percolation Structures and Processes* (Hilger, Bristol, 1983).
[8] H. Scher and R. Zallen, J. Chem. Phys. **53**, 3759 (1970).
[9] M. F. Sykes, D. S. Gaunt, and M. Glen, J. Phys. A **9**, 97 (1976).
[10] D. Wilkinson and M. Barsony, J. Phys. A **17**, L129 (1984).
[11] M. Rosso, J. F. Gouyet, and B. Sapoval, Phys. Rev. B **32**, 6053 (1985).
[12] R. M. Ziff and B. Sapoval, J. Phys. A **19**, L1169 (1986).
[13] F. Yonezawa, S. Sakamoto, and M. Hori, Phys. Rev. B **40**, 636 (1989).
[14] F. Yonezawa, S. Sakamoto, and M. Hori, Phys. Rev. B **40**, 650 (1989).
[15] R. P. Langlands, C. Pichet, Ph. Pouliot, and Y. Saint-Aubin, J. Stat. Phys. **67**, 553 (1992).
[16] R. M. Ziff, Phys. Rev. Lett. **69**, 2670 (1992).
[17] J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).
[18] M. F. Sykes and M. Glen, J. Phys. A **9**, 87 (1976).
[19] C. Domb and C. J. Pearce, J. Phys. A **9**, L137 (1976).
[20] A. Frammang, Z. Phys. B **28**, 47 (1977).
[21] P. L. Leath and G. R. Reich, J. Phys. C **11**, 4017 (1978).
[22] J. Hoshen, R. Kopelman, and E. M. Monberg, J. Stat. Phys. **19**, 219 (1978).

[23] J. Hoshen, D. Stauffer, G. H. Bishop, R. J. Harrison, and G. D. Quinn, J. Phys. A **12**, 1285 (1979).

[24] H. Nakanishi and H. E. Stanley, Phys. Rev. B **22**, 2466 (1980).

[25] T. C. Lubensky and A. J. McKane, J. Phys. A **14**, L157 (1981).

[26] A. Margolina, Z. V. Djordjevic, D. Stauffer, and H. E. Stanley, Phys. Rev. B **28**, 1652 (1983).

[27] A. Margolina, H. Nakanishi, D. Stauffer, and H. E. Stanley, J. Phys. A **17**, 1683 (1984).

[28] D. C. Rapaport, J. Phys. A **18**, L175 (1985).

[29] P. J. Raynolds, H. E. Stanley, and W. Klein, J. Phys. A **11**, L199 (1978).

[30] P. J. Raynolds, H. E. Stanley, and W. Klein, Phys. Rev. B **21**, 1223 (1980).

[31] D. W. Heermann and D. Stauffer, Z. Phys. B **40**, 133 (1980).

[32] P. D. Eschbach, D. Stauffer, and H. J. Herrmann, Phys. Rev. B **23**, 422 (1981).

[33] J. Adler, Y. Meir, A. Aharony, and A. B. Harris, Phys. Rev. B **41**, 9183 (1990).

[34] M. E. Fisher, Physics **3**, 255 (1967).

[35] D. S. Gaunt and M. F. Sykes, J. Phys. A **9**, 1109 (1976).

[36] M. S. Watanabe, Phys. Rev. E **51**, 3945 (1995).