Universality and Scaling in the Statistical Data of Literary Works

Chin-Kun Hu and Wei-Cheng Kuo
Universality and Scaling in the Statistical Data of Literary Works*

Chin-Kun Hu† Wei-Cheng Kuo

Academia Sinica The Affiliated High School of NTNU

In this paper we first review the concepts of universality and scaling for critical physical systems, including ordinary critical systems with tuning parameters and self-organized critical systems without any tuning parameter. We then plot the accumulated number of different words as a function of the total number of accumulated words in literary works in log-log scale and find a linear relationship between them. The slope of the linear curve can be used to represent writing style of the writer. We find that the slope for the English historical play Edward III (0.700) is different from the slopes of William Shakespeare’s historical plays (between 0.640 and 0.666) and the slope of Christopher Marlowe’s historical play Edward II (0.634), which suggests that Shakespeare and Marlowe did not write Edward III. The analysis of word-length distributions in these plays points to the same conclusion. We also calculate the Zipf’s distributions for some English and Chinese literary works and find that the distributions for all English (Chinese) literary works fall approximately on the same curve. However, the curves for English and Chinese literary works are quite different. We propose a model with two adjustable parameters to generate such curves. Finally, we discuss some problems for further research.

Key words: scaling, universality, authorship, Edward III, Zipf’s distributions

* Based on a talk presented by Hu at 2003 POLA (Project On Linguistic Analysis), 29 November-1 December 2003, at Aspire Park (Longtan, Taiwan) in honor of Professor William S-Y. Wang on the occasion of his 70th birthday.

We thank C.-K. Peng of Harvard Medical School, Ching-erh Chang of National Taiwan University, and Ting-Ting Chi of National Taiwan Normal University for useful discussions. We thank Ming-Chya Wu for help to prepare this paper. This work was supported by National Science Council (Taipei) under Grant No. NSC 93-2112-M 001-027 and the Ministry of Education (Taipei) under a high school student training program.

† Corresponding author with mailing address: Institute of Physics, Academia Sinica, Taipei 11529, Taiwan, electronic address: huck@phys.sinica.edu.tw, website: http://www.sinica.edu.tw/~statphys/ which contains PDF files of some cited papers by the present author.
1. Introduction

“Universality” means applicable in different content; “scaling” means functions of different quantities can collapse approximately on a single curve by a simple change of scale for each quantity. One of the objectives of scientific research is to use a small number of concepts, or functions, or laws to describe a large number of natural or laboratory phenomena. Thus it is not surprising that the concepts of universality and scaling have been used by researchers in different branches of sciences and it turns out that such concepts are quite useful.

Universality is the key idea in the search for fundamental interactions of matter. In the 17th century, Issac Newton (1642-1727 A.D.) discovered that the falling of apples to the earth, the motion of the moon around the earth, and the motion of the earth around the sun are governed by the same physical laws: universal law for gravitational forces of matter and Newton’s laws of motion. In the 19th century, James Clerk Maxwell (1831-1879 A.D.) proposed Maxwell’s equations to describe electric and magnetic interactions. In recent decades, S. Weiberg and A. Salam proposed a unified gauge field theory to describe the weak and electromagnetic interactions. The major objective of theoretical elementary particle physics is to formulate a unified theory which can describe electromagnetic, weak, strong, and gravitational interactions.

Besides basic interactions, the idea of universality is also useful for the understanding of many-body interacting systems, especially the critical systems. The quantitative and systematic researches on critical physical systems started in the later half of the 19th century in the study of liquid-gas critical systems, and made tremendous progress in the 20th century. The fluctuations in a critical systems contain a large number of different length scales and when one changes a length scale to examine the critical system, it appears to be similar to the original one. Thus the concept of scaling is useful for describing critical systems. In the following, we first review the concepts of universality and scaling in critical liquid-gas and other physical systems. We then point out that such concepts are also useful for describing statistical data of literary works.

2. Universality and scaling in critical physical systems

2.1 Critical liquid-gas systems

At the pressure $P$ of 76cm liquid Hg (1 atm), the liquid water has a phase transition to vapor at transition temperature $T_t = 100^\circ C$ and the density (mass per unit volume) of liquid water $\rho_l$ and the density of gas vapor $\rho_g$ has a finite difference, i.e. $\Delta \rho = \rho_l - \rho_g > 0$. In laboratory, one can put a sample of water in a container of volume $V$ and change the
pressure $P$ applied on the sample. As $P$ is increased, $T_t$ will increase, $\Delta \rho$ will decrease and the specific heat $C_p$, the compressibility (change of volume due to pressure change) $\alpha_p$, and correlation length (linear dimension over which fluctuations can correlate with each other) of the water will increase. As $P$ is increased to reach a critical pressure $P_c$, $T_t$ will increase to reach the corresponding critical temperature $T_c$ and $V$ becomes critical volume $V_c$, $\Delta \rho$ becomes zero and $C_p$, $\alpha_p$, and $\xi$ diverge. This is called critical phenomena (Stanley 1971). The critical exponents $\beta$, $\alpha$, $\gamma$, and $\nu$ are used to characterized critical behavior of $\Delta \rho$, $C_p$, $\alpha_p$, and $\xi$ near the critical point (Stanley 1971). Such critical exponents satisfy certain equality, called scaling relations, e.g. $\alpha + 2\beta + \gamma = 2$ (Stanley 1971).

In the later half of the 19th century, critical phenomena of many substances (e.g. CO$_2$ and H$_2$O) were discovered. In 1873, van der Vaals proposed that a mole ($\approx 6.02 \times 10^{23}$ molecules or atoms) of liquid-gas system can be described by the equation (Hu 1986, Stanley 1971)

$$P + \frac{a}{V^2} (V - b) = RT$$

where $a$ and $b$ are constants depending on the substance under study, $R$ is the ideal gas constant, and $T$ is the absolute temperature. We can use Eq. (1) to plot $P$ as a function of $V$ for fixed $T$ in the $P$-$V$ plane. For a small constant $T$ and a constant $P$, Eq. (1) could have three solutions for $V$; in this case Maxwell construction (an equal area rule for the parts below and above horizontal line) can be used to obtain a curve in the $P$-$V$ plane, which has a horizontal portion corresponding to liquid-gas coexistence region (Stanley 1971). When $T$ is increased, the size of this region decreases. When $T$ reaches the critical temperature $T_c$, the size of this region becomes 0. From this condition, one can derive that $V_c = 3b$, $P_c = a/27b^2$, $T_c = 8a/27bR$ from which one can determine $a$, $b$, and ideal gas constant $R$ from values of $V_c$, $P_c$, and $T_c$ measured by experiments. Two lines which connect boundaries of the horizontal portions for $T \leq T_c$ form phase boundaries, which separate liquid phase, liquid-gas coexistence phase, and gas phase. For large $T$, Equation (1) can be approximated by the Boyle’s law $PV = RT$; in this case for a constant pressure $P$ there is only one corresponding volume $V = RT/P$.

### 2.2 University of critical exponents

In 1881, van der Waals proposed “corresponding principle”. He defined reduced volume, reduced pressure, and reduced temperature as $\phi = V/V_c$, $\pi = P/P_c$, $\theta = T/T_c$, then he rewrote Eq. (1) as
which has been called reduced van der Waals’ equation. Equation (2) means that in terms of reduced variables data for different substance can collapse on the same curve. Equations (2) and (1) were confirmed by experiments. For this contribution, van der Vaals was awarded the Nobel Prize in physics in 1910.

In 1945, Guggenheim (Guggenheim 1945) reported that in the \( T/T_c \) versus \( \rho/\rho_c \) plane (\( \rho \) and \( \rho_c \) are density and critical density of the substance, respectively) the coexistence curves of Ne, Ar, Kr, Xe, N\(_2\), O\(_2\), CO, and CH\(_4\) are very consistent with each other. Thus van der Waals’s general idea for the universality of the coexistence curves was further confirmed. However, near the critical point it has been found that \( |\rho-\rho_c| \sim |1-(T/T_c)|^\beta \), where the critical exponent \( \beta \) is about 1/3, on the other hand van der Waals’s equation gives \( \beta = 0.5 \). The discovery of such deviation from van der Waals’s theory marks modern era of critical phenomena (Stanley 1971).

Besides liquid-gas critical systems, ferromagnet to paramagnet transitions and order-disorder transitions in alloys also show critical behavior. In early 1920s, Wilhelm Lenz (1888-1957) asked his student Ernst Ising (1900-1998) to study a simple model for ferromagnet to paramagnet transitions. Ising solved exactly the model on a one-dimensional lattice. Hereafter, the model has been called Lenz-Ising model or simply the Ising model (Brush 1967). For the model on a lattice \( G \) of \( N \) sites and \( E \) nearest neighbor bonds, the \( i \)-th site of the lattice is assigned an Ising spin \( \sigma_i \), \( 1 \leq i \leq N \), which can be +1 (spin up) or -1 (spin down). When two nearest-neighbor spins are in the same (different) direction, there is an interaction energy \(-J\) (\( J \)), where \( J \) is a coupling constant. The spin \( \sigma_i \) also couples with an external magnetic field \( B \) with an energy \(-B\sigma_i \). Ising found that his exact solution for the Ising model does not show the phase transition for finite temperatures. In 1944, Onsager solved exactly the square lattice Ising model in zero external field and found that the model has a phase transition with divergent specific heat at a finite critical temperature (Onsager 1944).

In 1952, Chen-Ning Yang (1952) derived the exact spontaneous magnetization \( M \) of the Ising model on a square lattice with isotropic interactions and found that the critical exponent \( \beta \) of \( M \) is 1/8. In the same year, Chang (1952) derived the exact spontaneous magnetization \( M \) of the Ising model on a square lattice with anisotropic interactions, i.e. the coupling constants in the horizontal direction \( J_1 \) and in the vertical direction \( J_2 \) are different. Chang (1952) found that, for \( 0 < J_1 / J_2 < \infty \), \( \beta \) is always equal to 1/8. Chang (1952) conjectured that for other planar lattices \( \beta \) is also equal to 1/8, which was confirmed by later calculations.

Now it is generally believed that for the Ising model on all two-dimensional lattices,
including the square (sq), the plane triangular (pt), the honeycomb (hc) lattices, etc, the specific heat exponent $\alpha$, the spontaneous magnetization exponent $\beta$, the magnetic susceptibility exponent $\gamma$, and the correlation length exponent $\nu$ are $0$ (logarithmic divergence), $1/8$, $7/4$, and $1$, respectively (Stanley 1971). It is also believed that for site and bond percolation models (to be defined below) on all two-dimensional lattices, the correlation length exponent $\nu$, the percolation probability exponent $\beta$, and the mean cluster size exponent $\gamma$ are $4/3$, $5/36$, and $43/18$, respectively (Stauffer and Aharony 1994).

Researches on critical physical systems till 1960s had converged to the widely accepted idea of universality: critical systems can be classified into different universality classes so that the systems in the same class have the same set of critical exponents. For example, the Ising (percolation) models on two-dimensional lattices are in the same universality class, but the Ising and percolation models are in different classes; three dimensional critical systems of liquid-gas, binary alloy, binary fluid, and the Ising model are in the same universality class. In early 1970s, K. G. Wilson proposed a renormalization group theory to understand the universal behaviors of critical systems. His theory can also provide a calculation scheme to obtain accurate critical quantities. For this contribution, K. G. Wilson was awarded the Nobel Prize in Physics in 1982.

### 2.3 Scaling and finite-size scaling

Another important concept in the theory of critical phenomena is scaling (Stanley 1971, Kadanoff 1990). For example, in a ferromagnetic system, e.g. CrBr$_3$, for temperatures $T$ near the critical temperature $T_c$ (also called the Curie temperature in ferromagnetic systems), if we plot $\sigma/|\epsilon|^\beta$ as a function of $h/|\epsilon|^\beta$, where $\sigma$ is the magnetization, $\epsilon = (T-T_c)/T_c$, and $h$ is the external magnetic field, then the experimental data for different temperatures collapse on a single curve, called the scaling function (Stanley 1971). In the following, we consider another kind of scaling, called finite-size scaling.

According to the theory of finite-size scaling (Fisher 1971, Privman and Fisher 1984, Cardy 1988, Privman 1990, Stauffer and Aharony 1994), if the dependence of a physical quantity $Q$ of a thermodynamic system (with the number of particles approaching $\infty$) on the parameter $\epsilon$, which vanishes at the critical point $\epsilon = 0$, is of the form $Q(\epsilon) \sim |\epsilon|^\alpha$ near the critical point, then for a finite system of linear dimension $L$, the corresponding quantity $Q(L, \epsilon)$ is of the form:

$$Q(L, \epsilon) \approx L^{-\gamma_y} F\left(\frac{\epsilon L^\nu}{L}\right)$$

where $y, (= \nu^{-1})$ is the thermal scaling power and $F(x)$ is the finite-size scaling function. It follows from (3) that the scaled data $Q(L, \epsilon) L^{\gamma_y}$ for different values of $L$ and $\epsilon$ can be...
described as a single function $F(x)$ of the scaling variable $x = \varepsilon L^\nu$. Thus it is important to know general features of the finite-size scaling function under various conditions. At the critical point $\varepsilon = 0$, we have

$$Q(L, \varepsilon = 0) \approx L^{-\nu} F(0)$$

Thus $Q$ and $L$ in log-log scale have a linear relationship whose slope is given by $-a/\nu$.

In 1984, Privman and Fisher (1984) proposed the idea of universal finite-size scaling functions (UFSSFs) and nonuniversal metric factors for static critical phenomena (Privman and Fisher 1984) for $T$ near $T_c$ and the external magnetic field $h$ near 0. Specifically, they proposed that, near $\varepsilon = 0$ and $h = 0$, the singular part of the free energy for a ferromagnetic system can be written as

$$f_s(\varepsilon, h, L) \approx L^{-d} Y \left( C_1 \varepsilon L^{1/\nu}, C_2 h L^{(\beta + \nu)/\nu} \right)$$

where $d$ is the spatial dimensionality of the lattice, $Y$ is a universal finite-size scaling function, and $C_1$ and $C_2$ are adjustable nonuniversal metric factors (Privman and Fisher 1984) which depend on the specific lattice structure. From Eq.(5) and the scaling relations $\nu d = 2 - \alpha$ and $\alpha + 2\beta + \gamma = 2$ (Stanley 1971), one obtains the scaling expression for the finite-size magnetization (Privman and Fisher 1984)

$$m = -\frac{\partial}{\partial h} f_s(\varepsilon, h, L) \approx C_2 L^{-\nu} Y( \left( C_1 \varepsilon L^{1/\nu}, C_2 h L^{(\beta + \nu)/\nu} \right)$$

which is the order parameter of the system. From 1984 to 1994, the progress in research on UFSSFs was very slow. The title of Privman and Fisher’s paper (Privman and Fisher 1984) is “Universal critical amplitudes in finite-size scaling” and most papers related to this paper only address the problem about the universality of critical amplitudes rather than universality of finite-size scaling functions. In the following two-dimensional percolation models will be used to show that critical systems can have very good finite-size scaling behaviors and universal finite-size scaling functions.

### 2.4 Universal finite-size scaling functions for percolation models

“Percolation” means spreading or penetration from one-side of the system to another. There are several kinds of percolation models. In the bond percolation model on a lattice $G$ of $N$ sites and $E$ bonds, each bond of the lattice is occupied with a bond probability $p$, $0 < p \leq 1$, and the sites connected by occupied bonds are defined to be in the same cluster.
For the site percolation model on $G$, each site of the lattice is occupied with a site probability $p_s$, $0 < p_s \leq 1$, and the nearest neighbor occupied sites are defined to be in the same cluster. When $p$ or $p_s$ increases, the average sizes of the clusters also increase. The cluster which extends from the top to the bottom of the lattice is called the percolating cluster. The average fraction of lattice sites in the percolation cluster is called “the percolation probability” $P$. The average size of non-percolating clusters is called the mean cluster size. The probability that at least one percolating cluster exists is called the the existence probability (Hu 1992a, b) (also called spanning probability, see Langlands et al. 1992) $E_p$. As the linear size of the lattice $L$ goes to $\infty$, $P > 0$ and $E_p > 0$ for $p > p_c$, where $p_c$ is called percolation threshold or critical point. For the bond percolation model on the square lattice, $p_c = 0.5$.

Hu (1992a, b) proposed a histogram Monte Carlo simulation method (HMCSM) to study percolation models. In 1995-1996, Hu, Lin and Chen (HLC) applied the HMCSM to calculate $E_p$, the percolation probability $P$, and the probability $W_n$ for the appearance of $n$ top to bottom percolating clusters on finite $L_1 \times L_2$ square (sq), honeycomb (hc), and planar triangular (pt) lattices, where $L_1$ and $L_2$ are linear dimensions of the lattice in the horizontal and vertical directions, respectively. They found that, by choosing appropriate aspect ratio $L_1/L_2$ for each lattice and non-universal metric factors for each model, $E_p$, $P$, and $W_n$ for six percolation models on planar lattices have UFSSFs (Hu et al. 1995a, b, Hu and Lin 1996). Hu and Wang (1997) used a random deposition Monte Carlo method to find that the continuum percolation of soft disks and hard disks have the same UFSSFs as percolation on planar lattices (Hu 1999). Hsu, Lin and Hu (2001) used Monte Carlo simulations to find that the bond percolation on random lattices has the same UFSSFs as the percolation model on regular lattices. We present examples of such studies in Figs. 1 and 2. Figure 1 shows that $W_n$ have very nice finite-size scaling behavior. Figure 2 shows that bond and site percolation models on sq, hc, and pt lattices have very nice UFSSFs. Figures 1 and 2 show that there are multiple percolating clusters at the critical point. Before Hu and Lin’s work (Hu and Lin 1996), it was widely believed that there is only one percolating cluster at the critical point of two and three dimensional percolation models (Stauffer 1997). Thus Hu and Lin (1996) inspired a series of works on critical multiple percolating clusters.

3. Self-organized critical systems

In the critical systems mentioned above, one should tune a parameter (e.g. temperature in liquid-gas systems or bond probability $p$ for the bond percolation model) so that the systems can reach its critical state. However, there are still many systems which show scaling behavior without turning parameter (Bak 1996). Such systems exhibit
self-organized criticality (SOC) and are called self-organized critical systems. There are many examples of self-organized critical systems (Bak 1996).

Fig. 1: (a) $W_n(L_1, L_2, p)$ for bond percolation on $128 \times 32$, $256 \times 64$, and $512 \times 128$ sq lattices, which are represented by dotted, dashed, and solid lines, respectively. (b) The data of (a) are plotted as a function of $z = (p - p_c)L^{1/\nu}$. The scaling function for $W_n(L_1, L_2, p)$ is denoted by $F_n(R, z)$, where $R = L_1/L_2$. The monotonic decreasing function is for $F_0(R, z)$. The $S$ shape curve is for $F_1(R, z)$. The bell shape curves from top to bottom are for $F_n(R, z)$ with $n$ being 2, 3, and 4, respectively. Taken from Hu and Lin (1996).
Fig. 2: (a) $W_n$ for bond and site percolations on 866×250 pt, 512×128 sq, and 866 ×125 hc lattices. (b) The data of (a) are plotted as a function of $x = D_1(p-p_c)L^{1/\nu}$, where $D_1$ is the non-universal metric factor. The universal scaling function for $W_n$ is denoted by $U_n(x)$. Taken from Hu and Lin (1996).
Fig. 3: Typical stable recurrent sandpile configuration on (a) honeycomb (hc), (b) square (sq), (c) triangle (ta), and (d) random lattices. For every lattice, sites are represented by solid black dots and links between neighboring sites are represented by solid lines. The dual of a given lattice is represented by dotted lines. Each site of a lattice has a corresponding cell of the dual lattice, which encloses the site. Thus the color of a cell is used to represent the height of the enclosed site. Red, orange, and yellow are used to represent the height \( z_i \) being \( z_i -1 \), \( z_i -2 \), and 0, respectively. In (b), (c), and (d), green represents \( z_i -3 \). In (c) and (d), purple represents \( z_i -4 \). In (c), blue represents 1. In (d), different depth of blue represents different values from 1 to \( z_i -5 \). When a particle is added at a red site, the site begins to topple and initiate an avalanche. Taken from Hu and Lin (2003).
3.1 Zipf’s Law

In 1949, George Zipf published an interesting result about statistical data in human writing (Zipf 1949). He found that in a given text corpus there is an approximate mathematical relation between the frequency of the occurrence of each word and its rank in the list of all the words in the text ordered by decreasing frequency (Zipf 1949, Montemurro 2001). Such frequency distribution will be called the Zipf’s distribution.

3.2 Bak-Tang-Wiesenfeld (BTW) Sandpile Model

In 1987, Bak, Tang, and Wiesenfeld (BTW) first proposed the idea of SOC and also proposed a lattice sandpile model to show these phenomena (Bak et al. 1987). Hu and Lin (2003) studied BTW models on two-dimensional systems, including hc, sq, triangle (ta), and random lattices (Fig. 3). The random lattice was constructed by randomly selecting points in the given domain and then connecting neighboring points by bonds; the method was described in Hsu et al. (2001). The coordinate numbers of hc, sq, and ta lattices are 3, 4, and 6, respectively; the coordinate numbers of the random lattice change from site to site with the most probable number being 6. BTW’s sandpile model on a general lattice $\mathcal{R}$ of $N$ sites is defined as follows. Each site of $\mathcal{R}$ is assigned a height integer; the $i$-th site is assigned $z_i$ for $1 \leq i \leq N$. In a stable configuration the height $z_i$ at any site $i \in \mathcal{R}$ takes values 0, 1, ..., or $z_i^{c_i}$, where the critical height $z_i^{c_i}$ is the coordination number (the number of the nearest neighbors) of the $i$-th site. A particle is added at a randomly chosen site and the addition of the particle increases the height at that site by one. If this height equals or exceeds the critical value $z_i^{c_i}$, then the site topples, and on toppling its height decreases by $z_i^{c_i}$ and the heights at all of its $z_i^{c_i}$ neighbors increase by 1. These neighboring sites may become unstable in their turn and the toppling process continues causing an avalanche. The open boundary conditions are used so that when a boundary site topples, the particle can leave the system. The dynamical process continues until $z_i < z_i^{c_i}$ for $1 \leq i \leq N$. In this way, a set of toppling sites with area $s$, occurs and forms an avalanche. Every avalanche may be represented as a sequence of more elementary events, waves of topplings, which can be organized as follows: if the site $i$ to which a grain was added becomes unstable, topple it once and then topple all other sites of the lattice that become unstable, keeping the initial site $i$ from a second toppling. The set of sites toppled thus far is called the first wave of topplings. After the first wave is completed, site $i$ is allowed to topple the second time, not permitting it to topple again until the second wave of topplings is finished. The process continues until site $i$ becomes stable and the avalanche stops, then we have the last wave of toppling. Since the BTW sandpile model is Abelian (Dhar 1990), the final stable configuration does not depend on how the waves of
toppling are organized.

Waves of topplings, being more elementary events than avalanches, also have much simpler properties. All the waves are individually compact, so we will characterize the waves of topplings only by their area, \( s \), and calculate total numbers and probability distributions of areas \( s \) for four categories of waves (Fig. 4a) (I) All waves. (II) Last waves. Each avalanche has exactly one last wave. When a particle is added to the \( i \)-th site with height \( z_i \), an avalanche is induced. Thus the total number of last waves is approximately equal to the product of the number of added particles and average fraction of lattice sites with height \( z_i \) (red sites in Fig. 3). (III) Dissipating waves which have particles leave the system from the boundary. (IV) Dissipating last waves which are both dissipating and last (Hu et al. 2000).

For a given probability distribution of waves, \( \mathcal{N}(s) \), the normalized condition \( \int_s^L \mathcal{N}(s) \, ds = 1 \) is imposed, where \( a \) is the lattice constant. In fact, Ivashkevich, Ktitarev, and Priezzhev (1994) proposed that for \( a^2 \ll s \ll L^2 \), the probability distribution of areas \( s \) for all waves of topplings is given by

\[
\mathcal{N}_{\text{last}}(s) \, ds \sim \frac{ds}{s}
\]

Fig. 4: (a) Examples of all wave (I), last wave (II), dissipating wave (III), and dissipating last wave (IV); (b). By a stereographic projection, the wave on a plane can be projected onto the waves on the surface of the sphere. The dissipating wave on the plane is corresponding to a last wave on the sphere, which touches the north pole. Taken from Hu et al. (2000).
Dhar and Manna (1994) proposed that the asymptotic probability distribution of the last wave is given by

\[ N_{\text{last}}(s) \sim \left( \frac{a^2}{s} \right)^{3/8} \frac{ds}{s} \]

or \( a^2 \ll s \ll L^2 \). Hu et al. (2000) observed that Eq. (7) is invariant under the inversion transformation (Fig. 4b) and argued that the asymptotic probability distribution of dissipating waves is:

\[ N_{\text{diss}}(s) \sim \left( \frac{s}{L} \right)^{3/8} \frac{ds}{s} \]

Hu et al. (2000) assumed that the events for last waves and dissipating waves are independent and obtained

\[ N_{\text{diss, last}}(s) \sim \left( \frac{a}{L} \right)^{3/4} \frac{ds}{s} \]

The critical exponents for Eqs. (7), (8), (9), and (10) are exactly 1, 11/8, 5/8, and 1, respectively.

Hu and Lin (HL) simulated the BTW sandpile model on hc, sq, ta, and random lattices with linear dimensions \( L = 32 \) - \( 1024 \), and \( N = L^2 \). For each \( L \) and lattice, they first generated a random stable configuration, and add \( 10^6 \) particles at randomly chosen sites to drive the system into critical recurrent configurations (Dhar 1990). They then began to take data by adding \( n = 4 \times 10^7 \) particles at randomly chosen sites. The total numbers of all waves \( A_{\text{all}} \), last waves \( A_{\text{last}} \), dissipating waves \( A_{\text{diss}} \), and dissipating last waves \( A_{\text{diss, last}} \) for these lattices as a function of \( L \) in log-log scale are plotted in Fig. 5(a). The data can be well represented by straight lines and slopes of the lines for four kinds of lattices are consistent with each other within numerical uncertainties. The average slopes for all, last, dissipating, and dissipating last waves are \( 0.19 \pm 0.01 \), \( 0.01 \pm 0.01 \), \( -0.18 \pm 0.02 \), and \( -0.71 \pm 0.01 \), respectively (Hu and Lin 2003).

To check whether waves of the BTW sandpile model on regular and random lattices have the same set of critical exponents predicted by Eqs. (7)-(10), HL used the data for \( 1024 \times 1024 \) hc, sq, ta, and random lattices to plot \( \log_2 \left[ s^q N'(s) L^r \right] \) as a function of \( \log_2 \left[ s/L^2 \right] \) in Fig. 5(b), where \( q \) is the predicted exact critical exponent. When the data are consistent with the predicted \( q \) values, the data should fall on horizontal lines. Figure
5(b) shows that the data are very consistent with the predictions of Eqs. (7)-(10) when \( 2^{-12} \leq s/L^2 \leq 2^{-3} \).

HL used ten independent set of data in the region \( 2^{-12} \leq s/L^2 \leq 2^{-3} \) to calculate critical exponents for four kinds of waves on four different lattices. All, last, dissipating, and dissipating last waves have exponents 1.00 ±0.01, 1.37 ±0.01, 0.61 ±0.02, and 1.02 ±0.02 for the hc lattice; exponents 1.01 ±0.01, 1.36 ±0.02, 0.61 ±0.02, and 1.02 ±0.02 for the sq lattice; exponents 1.01 ±0.01, 1.36 ±0.02, 0.62 ±0.01, and 1.01 ±0.02 for the ta lattice; exponents 1.01 ±0.01, 1.36 ±0.02, 0.62 ±0.02, and 1.02 ±0.02 for the random lattice. The exponents for four lattices are consistent with each other and also consistent with predicted exact values.

In summary, HL simulated the BTW sandpile model on hc, sq, ta, and random lattices and find that total numbers of waves and area distributions of waves on these lattices have nice finite-size scaling behavior and the same set of critical exponents. Thus the critical behavior of toppling waves is insensitive to the details of the lattices, which is similar to the critical behavior in ordinary critical systems.
Universality and Scaling in the Statistical Data of Literary Works

Fig. 5: (a) Number of waves $A$ as a function of $L$ for hc (solid line), sq (dotted line), ta (dashed line), and random (dotted-dashed line) lattices. Log-log plot of $A$ for all (○), last (□), dissipating (△), and dissipating last (▽) waves as a function of $L$. The shift parameter $Y$ is introduced so that we can show curves for four kinds of waves in the same figure. From top to bottom, the slopes for all waves are 0.19 ± 0.01, 0.19 ± 0.01, 0.19 ± 0.01, and 0.18 ± 0.01; the slopes for last waves are 0.01 ± 0.01, 0.01 ± 0.01, 0.01 ± 0.01, and 0.01 ± 0.01; the slopes for dissipating waves are 0.01 ± 0.01, -0.16 ± 0.02, -0.17 ± 0.01, -0.18 ± 0.01, and 0.19 ± 0.01; the slopes for dissipating last waves are -0.71 ± 0.02, -0.72 ± 0.01, and -0.70 ± 0.02, and 0.72 ± 0.02. (b) $\log_q [N(L^2)]$ as a function of $\log [L^2]$ for hc, sq, ta, and random lattices with linear dimension $L=1024$, where $q$ is the exact critical exponent and is 1, 11/8, 5/8, and 1 for all, last, dissipating, and dissipating last waves, respectively. The shift parameter $Y$ is introduced so that we can show curves for four kinds of waves in the same figure. Taken from Hu and Lin (2003).

3.3 Scaling behavior in family name distribution

Miyazima, Lee, and Nagamine (2000) plotted the total number of different family names $N$ as a function of the total number of population $S$ for five Japanese cities in log-log scale and find a linear relationship, i.e. $N \sim S^x$, with an exponent (slope) $x = 0.65 \pm 0.03$. In the following, we will present new evidences to show that human writing is a self-organized critical process and has interesting scaling behavior.
4. Writing as a self-organized critical process

4.1 Scaling in statistical data of literary works

In Fig. 6(a), we plot the accumulated number of different words as a function of the total number of accumulated words in log-log scale for 4 novels (The Legend of Condor Heros, Return of the Condor Heros, The Kung Fu Colt Master, and Swordsman) written by Jing Yong, 3 novels (Blue Blood, Diamond Flower, and Underground) written by Ni Kuang, and 3 novels written by J. K. Rowling about Harry Potter. Jing Yong and Ni Kuang are good friends and they have been living in Hong Kong during about the same period of time. Figure 6(a) shows that the novels written by the same author have very close slopes, but the novels written by different authors have quite different slopes. Such results suggest that the slope can be used to represent writing style of the author. The novels written by an author who prefers to use a larger (smaller) number of different words in the novel of given length will have larger (smaller), but quite consistent slopes. Thus there exists the possibility to use such slopes to answer the authorship problems of some literary works. In the following, we will try to use such slopes to answer the authorship problem of the English historical play: Edward III.

4.2 Scaling and the authorship of historical play: EDWARD III

The authorship of Edward III has been a mystery for a long time. Edward III was first printed anonymously in 1596 by the English publisher Cuthbert Burby; the play was reprinted also anonymously in 1599. In 1598, Meres Francis published Palladis Tamia: Wit’s Treasury, which listed all of William Shakespeare’s plays, but Edward III was not in the list (Francis 1973 reprinted). In 1760, Edward Capell included Edward III in the plays written by William Shakespeare, which initiated a series of debates about the authorship of Edward III (http://www.shakespeare.about.com). Based on statistical analysis of rare words, Slater (1988) proposed that Edward III was written by Shakespeare. However, Slater’s proposal was questioned by some researchers (http://drk.sd23.bc.ca/DeVere/) and supported by W. L. Godshalk (http://www.marshall.edu/engsr/SR1998.html).

A related and also interesting question is whether some of Shakespeare’s plays have been written by others, e.g. Christopher Marlowe (b. 1564). About this question, on a website (http://www2.localaccess.com/marlowe/pamphlet/pamphlet.htm), John C. Baker posted a long article with related references. Marlowe was born in the same year as Shakespeare (1564-1616) and was well known to have published the historical play: Edward II. Based on word rank order frequency statistics and phylogenetic tree construction, recently Yang et al. (2003) proposed that Edward III was written by...
Fig. 6: The accumulated number of different words as a function of the total number of accumulated words in log-log scale for novels and historical plays: (a) 4 novels by Jing Yong, 3 novels by Ni Kuang, and 3 novels by J. K. Rowling; (b) Edward III (blue) and historical plays by Shakespeare and Marlowe (Edward II, red).
Marlowe instead of Shakespeare. However, Yang et al. (2003) used only common words in two plays to calculate their distance and that might introduce some systematic errors. Thus it is of interest to use other statistical methods to answer the authorship problem of Edward III.

In Fig. 6(b), we plot the accumulated number of different words as a function of the total number of accumulated words in log-log scale for some historical plays and find that the slope for Edward III (0.700) is different from the slopes of William Shakespeare’s historical plays (between 0.640 and 0.666) and the slope of Christopher Marlowe’s historical play Edward II (0.634), which suggests that Shakespeare and Marlowe did not write Edward III. In Fig. 7, we plot word-length distributions in these plays and find that the curve for Edward III with a peak at word length 3 is different from the curves of William Shakespeare’s historical plays and Edward II which have a peak at the word length 4. Thus the data of Fig. 6(b) and Fig. 7 suggest that Edward III was not written by William Shakespeare and Christopher Marlowe.

![Fig. 7: Word length distributions for Edward III (blue), Edward II (red), and Shakespeare’s historical plays.](image)

Fig. 7: Word length distributions for Edward III (blue), Edward II (red), and Shakespeare’s historical plays.
Fig. 8: Zipf’s distributions for Chinese novels and English novels. Please note that the former and the latter fall on separate curves.

Fig. 9: Numerical simulations of Zipf’s distributions with exchange rate $g = 0.1$ and simulation steps: $10^3$, $10^4$, $10^5$, $10^6$, and $10^7$. 
Fig. 10: Numerical simulations of Zipf’s distributions: (a) number of words $N=4000$, exchange rate $g=0.085$ and simulation steps: $5 \times 10^3$, $10^4$, $5 \times 10^4$, $10^5$, and $10^6$. (b) number of words $N=10000$, exchange rate $g = 0.17$ and simulation steps $10^7$. 
5. Zipf’s distribution

5.1 Two classes of Zipf’s distributions

In Section 3.1, we have introduced Zipf’s distributions. Now we calculate Zipf’s distributions for five Chinese novels and five English novels, which are shown in Fig. 8. The Chinese novels were written in Ming Dynasty, Ching Dynasty, and recent decades. It is of interest to know that Zipf’s distributions for all Chinese novels fall approximately on the same curve and Zipf’s distributions for all English novels fall approximately on another curve, which is quite different from that of Chinese novels. Thus there are two classes of Zipf’s distributions: one for Chinese and another one for English.

It has been found that critical physical systems can be classified into different universality classes so that the systems in the same class have the same set of critical exponents (Stanley 1971, Wang and Hu 1997, Lin and Hu 1998), universal finite-size scaling functions (Privman and Fisher 1984, Hu et al. 1995a, b, Hu and Lin 1996), and amplitude ratios (Izmailian and Hu 2001). The criteria by which the critical systems can be classified into different universality classes is a problem of much academic interest. Figure 8 suggests Chinese literary works and English literary works could be considered as two separate “universality classes”. It is of interest to know what are differences between these two classes. In the following, we propose a model to generate such curves.

5.2 A model to generate Zipf’s distributions

A writer uses \( N \) different words \( w_i \), \( 1 \leq i \leq N \), in his writing and the probability that he uses \( w_i \) in his writing is \( x_i \), which satisfy the relation \( \sum_{i=1}^{N} x_i = 1 \). We can arrange \( x_i \) in a sequence from larger to smaller values and define an integer function \( I(x_i) \) of \( x_i \), which is the order number of \( x_i \) in the sequence. We assume that all \( x_i \) are different and we can define an inverse function \( V \) of \( I \) such that \( y_i \equiv x_i = V(I(x_i)) \).

We assume that a mature writer would have a steady \( x_1, x_2, \ldots, x_N \) and the corresponding \( y_1, y_2, \ldots, y_N \), which was developed through following process with an exchange parameter \( g \) and a range parameter \( r \):

1. In the beginning, every \( x_i \) takes a random number near \( 1/N \) such that \( \sum_{i=1}^{N} x_i = 1 \).
   From \( \{x_1, x_2, \ldots, x_N\} \), one has a corresponding \( \{y_1, y_2, \ldots, y_N\} \).
2. Randomly choose an integer \( j \) between 1 and \( N \), and subtract \( gy_j \) from \( y_j \):
   \[ y_j \to y_j - gy_j \]
3. Randomly choose an integer \( k \) between \( j-r \) and \( j+r \) and add \( g y_j \) to \( y_k \):
   \[ y_k \to y_k + gy_j \]
4. Update the order of \( y_i \) such that \( y_i > y_{i+1} \) for \( j \) between 1 and \( N+1 \).
5. Go to step 2 and iterate the process.

The simulation results for $y_I$ as a function of $I$ in log-log scale for $r=3$, $N = 1000$, $g = 0.1$, and iteration number $IN = 10^3$, $10^4$, $10^5$, $10^6$ and $10^7$ are presented in Fig. 9 which shows that as $IN$ increases, the Zipf’s distributions approach a steady curve similar to the curve for Chinese literary works in Fig. 8.

To understand the differences between curves for Chinese and English literary works in Fig. 8, we try to simulate curves in Fig. 8 and find that with $r=3$, $N = 4000$ and $g = 0.085$ the simulated curves will approach the curve for Chinese literary work (Fig. 10a) and with $r=3$, $N =10000$ and $g = 0.17$, the simulated curves will approach the curve for English literary work (Fig. 10b). Thus the English literary works have larger number of different words $N$ and exchange rate $g$ than Chinese literary works. This is not surprising because the former language is a alphabet spelling system and it is easier to generate new words to enrich the vocabulary or to use one word to replace another one during development of writing habit.

6. Conclusions

The discovery of universality and scaling in critical physical systems can be considered as one of the important scientific achievements of the 20th century, comparable to the discovery of quantum theory of matter proposed by Max Planck, Albert Einstein, Niels Bohr, E. Schrödinger, W. Heisenberg et al., and the relativity theory of space-time proposed by Albert Einstein.

In this paper, we first review the concepts of universality and scaling in critical physical systems and present some typical results for ordinary critical systems (Figs. 1 and 2) and self-organized critical systems (Fig. 5). We then show that the concepts of universality and scaling are also useful for analyzing statistical data of literary works (Figs. 6 and 8). We also propose a simple model with two parameters (number of different words $N$ and exchange rate $g$), which can reproduce the observed Zipf’s distributions (Fig. 10).

The results presented above suggest following problems for further studies:

1. Besides English, there are other languages with alphabet spelling systems (e.g., Russian, French, Germany, etc); besides Chinese, there are also other languages with patterns as words (e.g. historical Western Shia). It is of interest to calculate Zipf’s distributions of literary works in these languages and to know whether such distributions can collapse on curves of Fig. 8.
2. In the model of Sec. 5.2 the vocabulary pool of a writer and the number of different words $N$ is fixed, which is not realistic. In the next step, we would like to study a more realistic model in which new words can be added to the vocabulary pool and rarely used words can be deleted from the pool so that $N$ is a function of iteration steps.

3. The idea of universality classes and renormalization group theory are useful for understanding critical physical systems (Wilson 1983, Hu 1992a, b, Lin and Hu 2002). There are different language families in the world. It is of interest to know whether the idea of universality classes and renormalization group theory could be useful for understanding the language families.

References


