Diameter of Random Clusters

Don Blair and Jon Machta

Department of Physics, University of Massachusetts, Amherst, MA 01003-3720

(Dated: 16 January 2010)

Abstract

We define a new qeometric quantity, the diameter, on Potts Model clusters, and determine the scaling of the diameter with system size for $q = 1, 2, 3, 4$ in 2D, and $q = 2$ in 4D, using a particularly efficient algorithm for $q = 1$. We find that that scaling exponents for the diameter and the chemical distance are the same within error for these models.

I. INTRODUCTION

Potts Model. The Potts model – a generalization of the Ising Model to q spin states – has seen applications in a wide variety of fields, including conformal field theory, percolation theory, knot theory, mathematical biology, and complex networks. It consists of a lattice of Potts spin variables σ_i, each of which can have integer values $\sigma_i = 1...q$. Any two neighboring spins σ_i and σ_j contribute an amount $-K$ to the Hamiltonian if they have the same value, or zero otherwise; the Hamiltonian can thus be written as $H = -K \sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j}$, with K a dimensionless coupling constant [11]. Despite its simplicity, the Potts Model has a rich phase diagram, and can be mapped onto the Random Cluster Model for all $q \geq 0$, with $q \to 1$ corresponding to the Percolation Model, and $q \to 2$ corresponding to the Ising Model. For $q \leq 4$, the model exhibits a second-order phase transition at the critical point; for $q > 4$, the transition is first order [1].

The Percolation Model (Potts, $q = 1$) exhibits a particularly simple and intuitive phase transition: after the probability for occupying any site (in site percolation) or bond (in bond percolation) reaches the critical value (p_c), "infinite clusters", or clusters that span the size of the system, first begin to appear. The intuitive geometric nature of this model, as well as its varied applications in the fields of, e.g., graph theory and oil recovery, have ensured that geometric properties of percolation clusters (their shape, the extent of their perimeter, etc.) continue to be of great interest. One such geometric quantity that until recently [2] had been studied only for the $q = 1$, or Percolation model, is the *chemical distance*, or the average shortest path between two sites on a cluster. This quantity is of particular interest in applications related to transport in random media, e.g. packets on a communication network or oil through porous rock. The scaling of the chemical distance with the Euclidian distance r between the sites is known to be

$$\langle l \rangle \propto r^{d_{min}} \tag{1}$$

for critical Percolation clusters. In this work, we extend the study of the shortest path exponent d_{min} to the $q = 1, 2, 3, 4$ 2D Potts Model. For Percolation (Potts $q = 1$), the notion of a "path" between cluster sites is part of the definition of the model; for Potts $q > 1$, paths are generated via the mapping of the Potts Model onto the Random Cluster Model. In our simulations of the Potts model we employ the Swendsen-Wang algorithm [REF], which exploits this mapping explicitly in

2

order to provide the Potts model with a dynamics; the bonds that are created between cluster sites in this algorithm constitute the paths we study here.

The diameter. We also introduce a new geometric property of random clusters, the cluster *diameter*, w, which we define as the longest of all the shortest paths between sites on a cluster. The diameter of a cluster defined in this way is closely related to the graph theoretic notion of diameter, and thus makes contact with an extensive literature on the diameter of graphs of various types. In the case of the Potts model, the diameter of a cluster has implications for the maximum transport time from any site to any other site on the cluster, and, in particular applications, to correlation times associated with processes occurring on satisfied cluster bonds [REFS]. The similarity of this quantity to the chemical distance suggests that it will scale as:

$$\langle w \rangle \propto r^{w_{min}} \tag{2}$$

at criticality. Through numerical simulations, we show that this is indeed the case, and in fact that d_{min} is equal to w_{min} for $q = 1, 2, 3, 4$ 2D Potts Models within the error bounds of our analysis. Since the naive $O(N^2)$ algorithm for determining the diameter of a cluster (by finding shortest paths between all possible pairings of the N sites of the cluster, and choosing the largest of these paths) renders the larger system sizes desirable for a finite-size scaling analysis computationally inaccessible; we employ a novel algorithm for measuring the diameter for $q = 1$, described below.

At and above the upper critical dimension it is expected that mean field theory should apply [REF]; this means that the underlying graph of connected sites that form the critical cluster should be well approximated by a complete graph [REF Bollobas 96, Wu 92) on n vertices – i.e., a simple graph in which every pair of vertices is connected by an edge. It has been shown [7] that the diameter of the complete graph at criticality scales as

$$w(n) \propto n^{1/3} \tag{3}$$

where n is the number of vertices in the graph. In order to determine whether this scaling applies to the diameter of Potts model clusters at and above the upper critical dimension, we have also determined numerically the value of d_{min} in the $q = 2$ (Ising) Potts Model, for which the upper critical dimension is 4. Since the mapping of the complete (linear) graph to the Potts random graph in 4D is $L^4 = n$, $w(L) \propto L^{4/3}$; thus, we may expect that w_{min} should equal 4/3 for $q = 2$ in 4D.

3

II. METHODS

Swendsen-Wang Algorithm. We performed Monte Carlo simulations of the Potts Model using the Swendsen-Wang algorithm [10]. This algorithm, which is itself based on the work of Fortuin and Kasteleyn [3], works by first introducing bonds between neighboring spins, with probability

$$p(\sigma_i, \sigma_j) = \delta_{\sigma_i, \sigma_j}(1 - e^{-K}), \tag{4}$$

thus creating clusters of bonded spins. All clusters thus formed are then, with probability $1/2$, flipped by choosing a random spin value from the q possible values, and assigning this value to all sites in the cluster. Such cluster-flipping algorithms dramatically reduce critical slowing down in computer simulations of spin models, as compared with algorithms that flip each spin individually [8] (e.g. the Metropolis algorithm [6]). In our case, the bonds that are formed by the Swendsen-Wang algorithm also provide the cluster bonds that will be analyzed as paths in our chemical distance and diameter analysis.

Previous Methods for Determining the Chemical Distance. Various methods have been used to determine the average chemical distance on critical percolation clusters. [REFS]. One commonly-used method is the method introduced by Leath. Recall that in the percolation model, sites or bonds on a lattice are populated with probability p; the value of p above which infinite clusters (clusters that span the system, as $L \to \infty$) are first formed is known the critical probability p_c. To analyze the statistics of clusters in a percolation model with occupation probability p, it is therefore clear that one need not simulate the entire lattice: it is sufficient generate a cluster by "growing" the cluster from a seed, using the method first proposed by Leath [5]: using a random number generator, one assigns all the bonds associated with the seed site the status "occupied" or "unoccupied" with probability p. If a bond is assigned "occupied" status, the site to which this bond connects is deemed a "growth site", and is added to cluster. All the sites thus added to the cluster in this round form a "chemical shell" of distance l from the seed site. This process is then continued for subsequent generations of growth trials, each associated with a larger chemical shell; the growth process stops naturally when one of the growth rounds generates no new growth sites. (Note: sites not added to the cluster in a particular round get another chance to be added to the cluster in subsequent rounds; but, once added, are no longer considered as possible growth sites.) Various improvements on this basic scheme have been proposed; that of Paul et al [9] provided significant reductions in the memory required to store the occupancy status of cluster

4

sites; Grassberger[4] employed a two-seed cluster growth method, in which two different clusters are grown simultaneously; if at any point the clusters then share a common growth site, it is a simple matter to determine the chemical distance between the seed sites, this path can be considered to have been measured from two random points on the typically much larger cluster that would have resulted from allowing the growth process to continue.

Procedure for $q > 1$. For $q > 1$, we cannot rely these cluster-growing techniques, as we must allow the entire system to evolve. For $q = 1$, we cannot employ the memory-saving features of Paul et al, as we must retain information about all sites and bonds in the cluster in order to later determine the cluster diameter; and the two-seed method of Grassberger does not save us computational work, since we must grow the entire cluster in order to measure the diameter. We can, however, increase the efficiency of our diameter measurement by recognizing that the diameter of a cluster must have its endpoints on the external perimeter of the cluster. Our procedure for $q > 1$ is then:

1. Generate a new cluster configuration using the Swendsen-Wang algorithm (see above) with periodic boundary conditions. The identification of connected clusters in this steps allows us to determine the largest cluster in the system.

2. Choose a random site s on this cluster as the seed site.

3. Beginning with the seed site s, determine all sites in the largest cluster by "growing" along satisfied cluster bonds (this process does not change the bonds that were determined in step 1).

4. The chemical shell reached in the final step of this growth process, $shell_{final}$, is considered to be the randomly-chosen chemical distance on the largest critical cluster, and is added to our statistics for the chemical distance.

5. All the i sites at the end of this growth process whose nearest neighbors are all occupied are deemed to be perimeter sites, p_i. This set includes all of the external perimeter sites of the cluster.

6. A similar Leath growth process is preformed using each of the perimeter sites as seeds, and $shell_{final_i}$ from each of these growth processes is stored.

7. The diameter for the largest cluster is then $max\{shell_{final_i}\}$.

This method for finding the diameter is an improvement over the naive N^2 algorithm for solving the all-pairs maximum shortest path problem on the paths formed along cluster bonds. It is expected to scale as $O(pN)$, where p is the number of perimeter sites on the largest critical cluster.

Procedure for $q = 1$. For $q = 1$, it is possible to grow a cluster from a seed site. We can still not employ the efficiencies introduced by Paul et al and Grassberger; but we save computational effort in the following ways. First, we recognize, as before, that the diameter must have its endpoints on perimeter sites. Second, we note that any "pins", or singly-connected paths on the external perimeter of the cluster, contain sites that can be eliminated as possible diameter endpoints; and it is straightforward to show that the existence of such a "pin" also allows us to eliminate as candidate diameter endpoints that lie within the "body" of the cluster as well.

Let us call the set of all sites on the pin P, and call p_{tip} the site that is the outermost tip of a given pin (i.e., the site with only one nearest neighbor) and p_{attach} the site that attaches this pin to the body of the cluster (i.e., a site with more than 2 nearest neighbors). Imagine that we were to include as a candidate site in S some site from P that was not p_{tip}, resulting in a candidate diameter D'; it would be immediately clear that rejecting this site in favor of p_{tip} would result in a new candidate diameter $D'' > D'$. We can therefore exclude all sites in in P that are closer than p_{tip} to S. Similar considerations [PROVE THIS?] allow us to additionally exclude from S all sites in N that have a chemical distance from p_{attach} less than or equal to the chemical distance between p_{tip} p_{attach} (i.e., the length of the pin). Our method is then to begin, for every site is in S, a "Leath growth" search that examines the chemical distance between along the cluster between s and every other site on the cluster, terminating when all cluster sites have been examined. The maximum chemical distance found across all such searches is then D.

We thus need only consider a relatively small proportion [quantify this proportion, on average] of cluster sites as possible diameter endpoints, greatly reducing the number of "Leath scans" required in order to determine the diameter exactly. Note that this method does not work for periodic boundary conditions, however; we must therefore grow clusters from a seed site, retaining only those clusters that do not grow to touch the boundaries of the lattice. Our procedure for $q = 1$, then, is:

1. Choose a growth seed site in the center of the lattice.

2. Perform a Leath growth from this site until the cluster dies, or reaches the boundaries of the maximum lattice size of L_{max}. If any cluster site borders L_{max}, begin again at step 1.

3. Identify all the perimeter sites in the cluster by choosing all sites in the final growth step that are perimeter sites (i.e., those that have less than the maximum number of allowed nearest neighbors). In this geometry, all the sites in the final chemical shell will be external perimeter sites.

4. Identify all the "pins" among these perimeter sites by performing a Leath growth from each pin site until one finds a site that is not singly-connected. All of the sites in the "neck" of the pin are eliminated from consideration as diameter endpoints.

5. Beginning from the point of attachment of the pin to the body of the cluster, continue the Leath scan until one has achieved a chemical shell equal to the distance (along sites) between the point of attachment and the end of the pin. All of sites thus scanned are also eliminated from consideration as diameter endpoints.

6. Perform Leath growths from all of the remaining perimeter sites p_i, collecting the maximum chemical shells reached in each instance; the largest of these chemical shells is then the diameter of the cluster.

We compared this method to the method described for $q > 1$, and found that the fraction of perimeter sites eliminated as candidates for diameter endpoints was approximately $X\%$ in our runs with $L_{max} = XX$.

A.   Simulation Methods

Overview. We used the Swendsen-Wang algorithm to simulate Potts Models 2D at criticality for values of L between 8 and 2048 for our measurements of l, and 4 and 128 for our measurements of w. For $q = 2$ in 4D, L ranged between 4 and 64. All simulations began in a random configuration.

Values of p_{add} used. For $q = 1$ in 2D, p_{add} is known exactly [REF]. For $q = 2, 3, 4$ in 2D, p_{add} = X [REF], X [REF], and X [REF], respectively. For $q = 2$ in 4D, $p_{add} = X$ [REF].

Thermalization. For $q > 1$, the simulations require some time to achieve an equilibrium state, and should therefore be thermalized. Accordingly, each simulation for system size L was run for at least $X\tau_{int,m}$ before measurements were taken, where $\tau_{int,m}$ was the estimated integrated autocorrelation time for the mass of the largest cluster for that value of L. (A table of integrated autocorrelation times for the largest system sizes measured is provided below at [TABLE]).

Run times. In 2D, our simulations were run for a length of $X\tau_{int,m}$; for measurements of w, and for $X\tau_{int,m}$ for measurements of l. For our 4D, $q = 2$ measurements, simulations were run for a length of $X\tau_{int,m}$ for our measurements of l. Some of our simulations consisted of a single, long run; others were the result of combining data from several runs begun from different initial random number generator seeds.

Random Number Generator. Random numbers for the simluations were generated using the Mersenne Twister method [REF: Matsumoto + Nishimura 1998], with parameters chosen to provide a period of at least X [determine this].

Tests. As a check on our simulation methods, we also measured the mass of the largest cluster for each lattice size L in order to determine the fractal dimension. The agreement betwen our values and the latest from the literature was good [Summary of comparisons below?].

CPU time. The CPU time for simulations measuring the diameter w was approximately $XL^2\mu s/$ iteration; for l it was approximately $XL^2\mu s/$ iteration, when run on the [details of cluster].

B. Data Analysis

Blocking Method. We used the 'blocking' method [8] to extract the proper standard deviation for chemical distance and diameter from our measurements. This method works by clustering the measurements of the quantity O into blocks of size s; the average of O is then found for each block independently; the standard deviation in O is then taken to be the standard deviation in these block averages, or

$$\sigma = \sqrt{\frac{\langle m^2 \rangle - \langle m \rangle^2}{n-1}} \tag{5}$$

where n is the number of blocks.

Fitting Methods. For $q = 1, 2, 3$, we attempted fits using the Ansätze $y = aL^b$ and $y = aL^b + L/c$, including in the fit data points down to L value of L_{min}, where L_{min} was the smallest value of L that still yielded a reasonable goodness-of-fit value, Q; the fitting form $y = aL^b$ provided the best fits for all values of q. For $q = 4$, we also attempted a fit of the form $y = A + B \log L$; the fit was not as good as the Ansätz $y = aL^b$.

III. RESULTS AND DISCUSSION

Comparison of chemical distance and diameter. [Wait on analysis]

Comparison of our results with those of Deng et al. Our numerical results appear to match the conjecture of Deng et al. [2] within error for $q = 1$ and $q = 2$; for $q = 3$, we find [wait until results of new blocking analysis are in]. For $q = 4$, we were unable to find a fit of high quality; but our results seem to support Deng et. al's conjecture...

Discussion of systematic errors.

q	1	2	3	4
d_{min} (conj.)	1.13021	1.09375	1.06667	1.03125
d_{min} (num.)	1.1304(2)	1.0933(2)	1.0641(1)	1.0335(9)
w_{min}	XXX	XXX	XXX	XXX

TABLE I: **Results for 2D Potts Model.** Scaling exponent for the chemical distance (d_{min}), and for the diameter (w_{min}), for the 2D Potts model with various values of q, with system size L=16,32,64,128,256,512. Included for comparison is the exact conjecture for d_{min} of Deng et. al [REF]. For $q = 1$, the values of w_{win} were determined using the special algorithm described in the text, and included system sizes up to L=XXX.

IV.   FIGURES

FIG. 1: A typical percolation cluster is shown. The chemical distance, or shortest path, between two arbitrary sites on the cluster is colored red; the diameter is colored green.
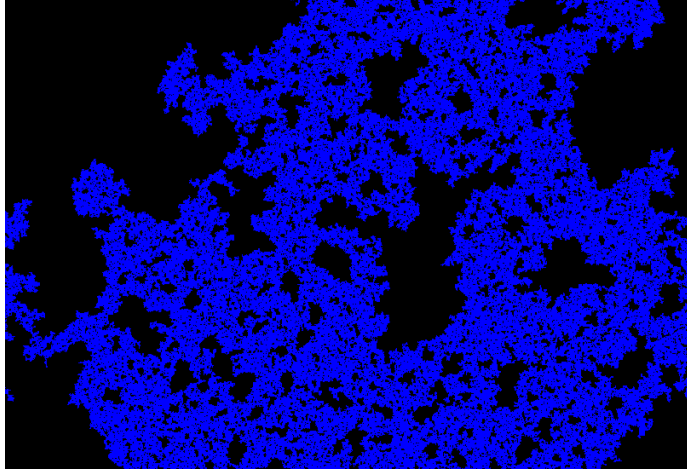
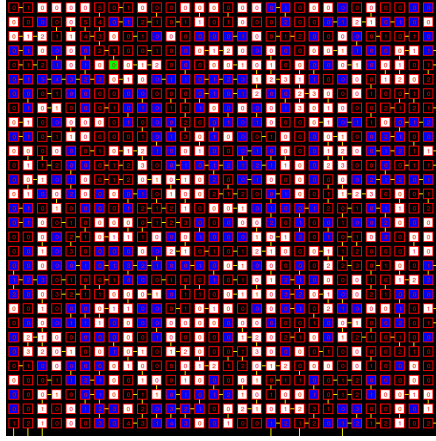FIG. 2: Example Potts Cluster, 2D, $q = 1$ (Percolation).
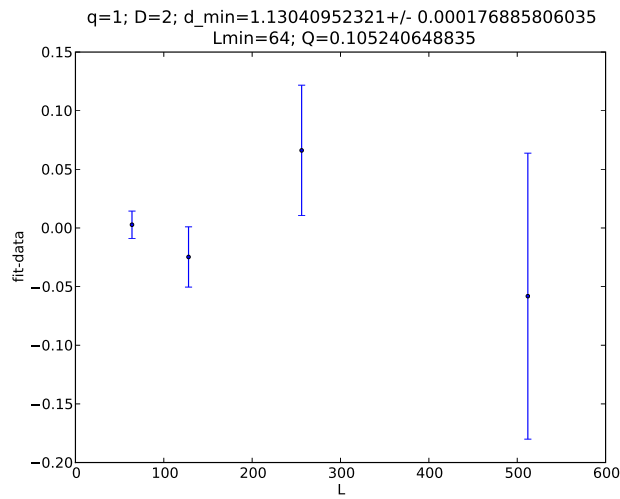

FIG. 3: Leath Growth Example, 2D, $q = 1$.

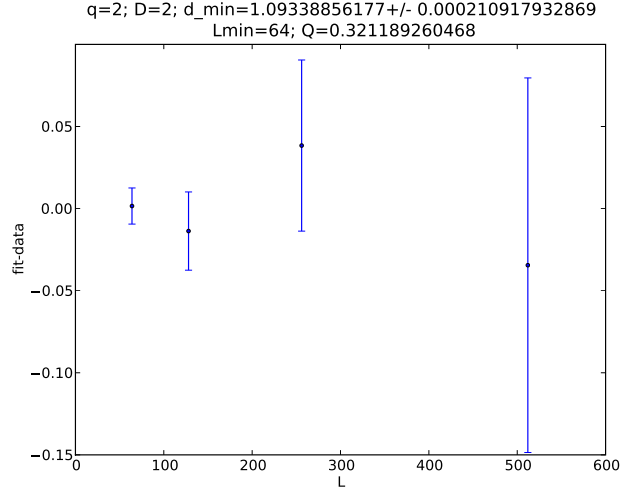FIG. 4: Residuals for l and w, 2D Potts, $q = 1$, fit of form $y = AL^B$.

10

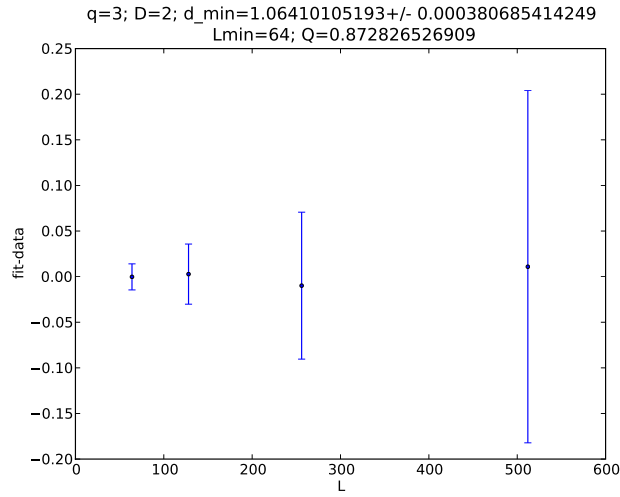FIG. 5: Residuals for l and w, 2D Potts, $q = 2$, fit of form $y = AL^B$.

FIG. 6: Residuals for l and w, 2D Potts, $q = 3$, fit of form $y = AL^B$.

V. BIBLIOGRAPHY

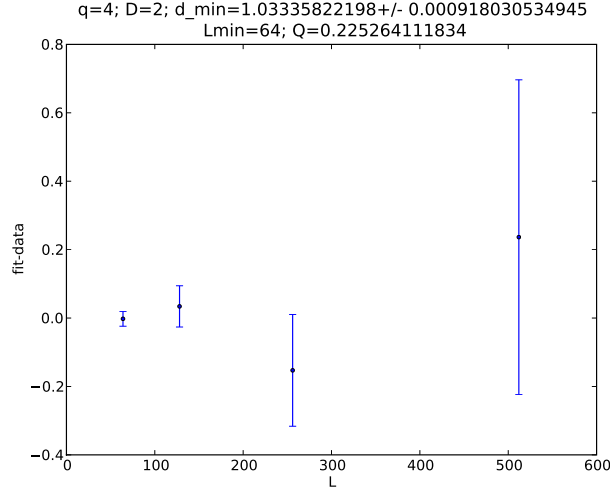[1] R. J. Baxter. Potts model at the critical temperature. *J. Phys. C*, 6:L445, 1973.

FIG. 7: Residuals for l and w, 2D Potts, $q = 4$, fit of form $y = AL^B$.

[2] Youjin Deng, Wei Zhang, Timothy M Garoni, Alan D Sokal, and Andrea Sportiello. Some geometric critical exponents for percolation and the random-cluster model. pages 16–19, 2010.

[3] C. M. Fortuin and P. M. Kasteleyn. On the random-cluster model. *Physica*, 57:536, 1972.

[4] P. Grassberger. Pair connectedness and shortest-path scaling in critical percolation. *J. Phys. A: Math. Gen.*, 32:6233, 1995.

[5] P. Leath. Cluster size and boundary distribution near percolation threshold. *Phys. Rev. B*, 14:5046, 1976.

[6] N. Metropolis, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J.Chem.Phys.*, 21:1087, 1953.

[7] Asaf Nachmias and Yuval Peres. Critical random graphs: diameter and mixing time. *Time*, pages 1–18.

[8] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Oxford, Oxford, 1999.

[9] Gerald Paul, Robert M. Ziff, and H. Eugene Stanley. Percolation threshold, Fisher exponent, and shortest path exponent for four and five dimensions. *Physical Review E*, 64(2):1–8, 2001.

[10] R. H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86, 1987.

[11] F. Y. Wu. The Potts model. *Rev. Mod. Phys.*, 54:235, 1982.