Sci. Prog., Oxf. (1977) **64**, 65–83.

Percolation and related topics

D. J. A. Welsh

1. Introduction

As its name suggests, percolation theory is concerned with flow in random media. Its origin, in 1957 in the work of Broadbent & Hammersley,[2] was as a model for molecules penetrating a porous solid, electrons migrating over an atomic lattice, a solute diffusing through a solvent or disease infecting a community. In this paper we shall attempt to survey the main results in classical percolation theory and also to relate it with other areas of applied mathematics in which the underlying structure is a random medium and which can be perhaps regarded as 'percolation in a wider sense'. These include such topics as the Ising model of ferromagnetism, the shortest route problem in random networks, models of growth processes and random clumping. For excellent surveys of this area we refer to the review papers of Essam[6] and Shante & Kirkpatrick.[29] This article is obviously dependent on Essam and Shante & Kirkpatrick but will aim at covering a different area of the subject and from a different viewpoint.

2. Random graphs

As an example of percolation in the wider sense consider the following problem in communication theory.

Let N be the network shown in Fig. 1. Suppose each directed edge has probability p of being reliable, that is allowing a message to pass. Suppose further that the reliability of each edge is independent of the reliability of any other edge. What is the probability that there is a path from A to B consisting only of reliable edges? Simple calculation shows that it is just the probability that not all the routes from A to B are unreliable. Since the routes have no edge in common we are dealing with independent random variables and we have

$$\begin{aligned} P[\mathrm{A} \rightsquigarrow \mathrm{B}] &= 1-P[\text{ACB closed, ADB closed, AEB closed}] \\ &= 1-P[\text{ACB closed}]\,P[\text{ADB closed}]\,P[\text{AEB closed}] \\ &= 1-(1-p^2)^3 \end{aligned}$$

Dr Welsh is a Fellow of Merton College, Oxford.

However, if we try the same problem for the network N′ (Fig. 2) the problem is much more complicated. This is due solely to the *dependence* in N′ of the events 'the route ACDB is reliable' and 'the route ACB is reliable'.

This problem illustrates the intrinsic difficulty of percolation problems–stochastic dependence occurs in most cases and makes computation very difficult. Indeed, even with the speed of modern computing machines it is still difficult to determine the reliability of networks–see for example the paper of Wing & Demetriou[36] or the book of Frank & Frisch.[11]

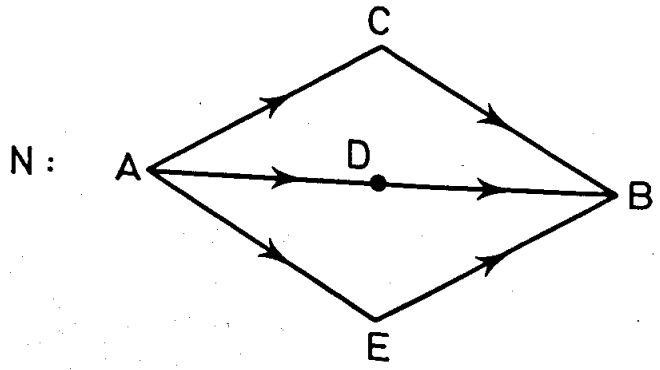


Fig. 1

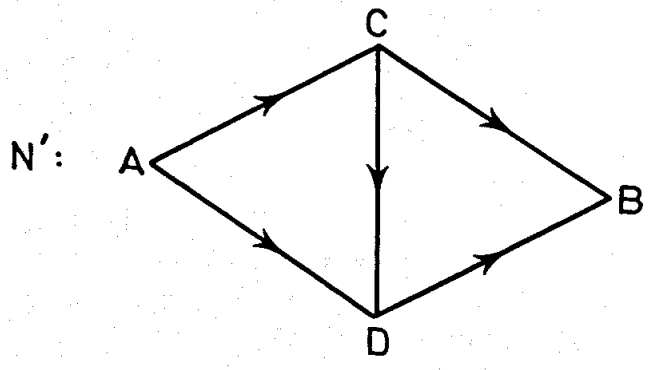


Fig. 2

Throughout this paper, unless otherwise specified, a *graph* consists of a set V of vertices and a set E of edges joining certain pairs of these vertices. Formally E is a subset of V × V. A graph is *finite* if V and E are finite sets, otherwise it is *infinite*. A *digraph* or *directed graph* or *network* is a graph in which each edge is assigned a direction or orientation. The networks of Figs. 1 and 2 are examples. A *random graph* in the sense of this paper will mean a structure consisting of a set of vertices and a probability mechanism in which each pair of vertices is joined by an edge with probability p or not joined with probability q $(= 1-p)$, independently for each pair of vertices.

Thus the number of edges in a random graph on n vertices is a random variable X with a distribution given by

$$P[X=k]=\binom{n}{k}p^k q^{\binom{n}{2}-k},\ 0\leqslant k\leqslant\binom{n}{2}.$$

This is a slightly different model from that studied by Erdös & Renyi[8] as they considered graphs on n vertices having $N(n)$ edges where $N(n)$ is a prescribed function and each N-subset of the $\binom{n}{2}$ possible edges occurs with equal probability. However in the broad sense the two theories are very similar.

Suppose we let $C_n(p)$ be the probability that the random graph on n vertices is *connected*, that is that there exists a path between any pair of vertices. Clearly

$$C_n(0)=0, \qquad C_n(1)=1$$

and $C_n(p)$ is an increasing function of p. Gilbert[12] used this model for routing calls through central telephone offices. Imagine the n vertices to be telephone offices and suppose that each pair of offices has the same probability p that there is an idle direct line between them. Suppose further that a new call between two offices can be routed via other offices if necessary. Then $C_n(p)$ is the probability that each office can call every other office. Exact expressions for $C_n(p)$ are given by Gilbert[12] but for large n they are unwieldy. Bounds on $C_n(p)$ show that as $n\to\infty$

$$C_n(p)\sim 1-n(1-p)^{n-1}$$

To illustrate the dependence of this on n and p consider the following table taken from Gilbert:[12]

p	0.9	0.5	0.1
$C_2(p)$	0.90000	0.50000	0.10000
$C_4(p)$	0.99581	0.59375	0.01293
$C_6(p)$	0.99994	0.81569	0.00624

At first sight the result that for any fixed $p>0$

$$\lim_{n\to\infty} C_n(p)=1$$

is slightly surprising since for $p=0.1$ the early terms of the sequence do not indicate such limiting behaviour. However, on an intuitive level we note that the average number of edges in a random graph is essentially $pn^2/2$ and since the smallest connected graphs on n-vertices (spanning trees) have only $n-1$ edges it is to be expected that for large n a random graph is almost surely connected. Indeed more can be said, for if $\delta_n(p)$ denotes the *diameter* of the random graph, defined in the obvious way as the maximum of the shortest distances between any pair of vertices, it can be shown that for any fixed $p>0$, as $n\to\infty$, with probability one

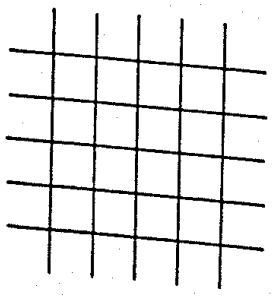
$$\lim_{n\to\infty} \delta_n(p)=2$$

Thus loosely speaking, in the terminology of telephone offices, for large n it is highly unlikely that any new call will have to be rerouted through more than one station. A further result of Grimmett & McDiarmid[14] implies that as $n \to \infty$ between any two vertices there are approximately pn disjoint paths. For further details and references see Grimmett & McDiarmid.[14]

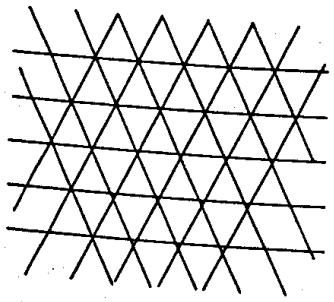
3. Classical percolation theory

By classical percolation theory we mean that branch of random graph theory which originated in the paper by Broadbent & Hammersley[2] and which is concerned with the probability of infinite clusters in a 'regular crystal lattice'. The definition of what exactly is a ' regular crystal lattice' is rather difficult to formulate precisely– indeed it varies from author to author. For the purposes of this paper it can be regarded as typified by the regular lattices shown in Fig. 3.

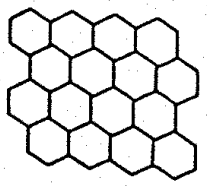
In the main we shall be concerned with the two dimensional square lattice L shown in Fig. 3(a). L is an infinite graph with vertices situated at all integer coordinated points of the plane.



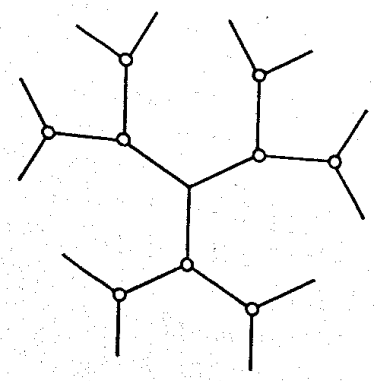
(a) Square Lattice



(b) Triangular Lattice



(c) Hexagonal or Honeycomb



(d) Bethe Lattice

Fig. 3

Bond percolation

Suppose that there is a supply of fluid at the origin and that each edge of L allows fluid to pass along it with probability p, independently for each edge. Let $P_n(p)$ be the probability that at least n vertices of L get wet by the fluid. Thus

$$P_1(p) = 1$$
$$P_2(p) = 1-(1-p)^4$$

and *in theory* $P_N(p)$ can be calculated for any integer N. However, the reader will rapidly find it prohibitively time consuming ($N = 7$ is a fair piece of work!). Obviously

$$P_N(p) \geqslant P_{N+1}(p)$$

and hence we know that $P(p)$ $(= P_\infty(p))$ exists where

$$P(p) = \lim_{N\to\infty} P_N(p)$$

and it represents the probability that fluid spreads an infinite distance from the origin.

Theoretically, little has been proved about $P(p)$. Even though each $P_N(p)$ is a polynomial in p and hence we would expect $P(p)$ to be a continuous function of p this has not yet been proved. The main result of Broadbent & Hammersley[2] is that there exists a *critical probability* p_H such that

$$p < p_H \Rightarrow P(p) = 0$$
$$p > p_H \Rightarrow P(p) > 0$$

and Monte Carlo simulations suggest that for all the well-known lattices the behaviour of $P(p)$ is roughly the same in the qualitative sense. A typical picture of the form of $P(p)$ is shown in Fig. 4.

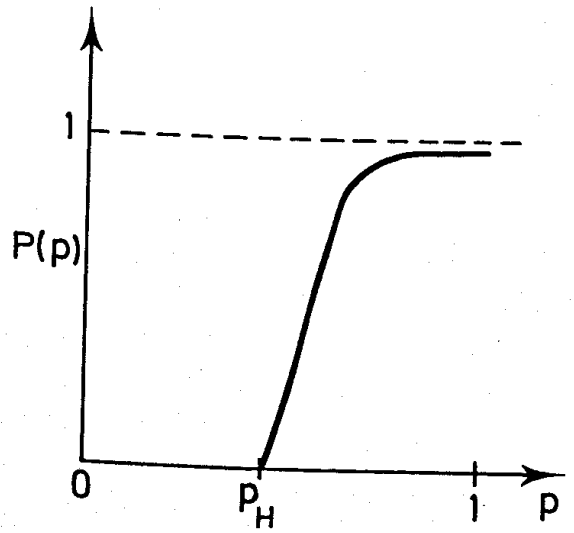


Fig. 4

The existence of this critical probability was proved by Broadbent & Hammersley[2] for a wide class of lattices; we discuss its evaluation in Section 5.

Atom or site percolation

The reader may have noticed that though the last section was headed 'bond percolation', no mention of 'bond' occurred subsequently. Historically, the subject of percolation had statistical mechanics overtones, and in this area 'bond' is usually used to denote an 'edge' of a graph, similarly 'site' or 'atom' denotes a 'vertex'. We shall depart from custom and use the mathematically more common vertex–edge notation.

In atom percolation on L instead of each edge of L being randomly blocked with probability $1-p$ or open with probability p the vertices of L are blocked with probability p or open with probability $q = 1-p$. Again we are interested in the probability of fluid spreading locally or an infinite distance.

Exactly analogous results hold for atom percolation as for bond percolation, though of course the numerical values of the critical probabilities p_H and percolation probabilities $P(p)$ differ.

In fact it can be argued that atom percolation is the more important to study on the grounds that any bond percolation problem on a lattice L can be turned into an

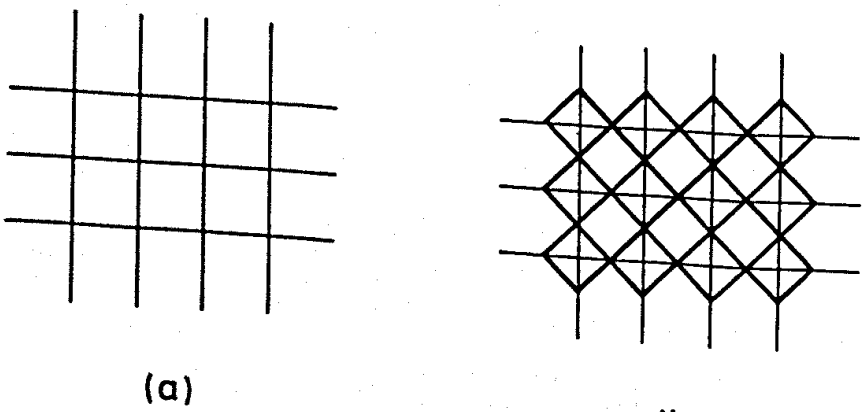


Fig. 5

atom percolation problem on a related lattice $\tilde{L}$. This lattice $\tilde{L}$ is known in graph theory as the *line graph* of L and is got by letting each edge of L be a vertex in $\tilde{L}$ and joining two vertices of $\tilde{L}$ if and only if the corresponding edges of L are incident. For example, Fig. 5(b) shows the line graph of the square lattice L.

It has been shown that for any regular lattice, if $P^A(p)$, $P^B(p)$ represent respectively the atom and bond percolation probabilities on the lattice then in general

$$P^A(p) < P^B(p) \quad 0 < p < 1$$

for details see Fisher[9] and Hammersley.[17]

Asymmetric percolation

The idea of a critical probability below which fluid (or infection) will only spread locally can be extended to the situation where the vertical edges of the square lattice L have probability p_1 of being open, and the horizontal edges have a probability p_2 of being open. This time we find that instead of a critical probability there is a *critical curve* passing through the points (0, 1) and (1, 0) such that if the point (p_1, p_2) lies on the same side of the curve as the origin then with probability 1 only a finite number of vertices will be wet whereas if (p_1, p_2) lies outside the curve there is a non-zero probability that an infinity of vertices will be wet. The critical curve obviously passes through the point (p_H, p_H) and Mauldon[24] gives estimates for the values of r where the line $p_1 = r\cos\theta, p_2 = r\sin\theta$ meets the curve for a range of values of θ between 0 and $\pi/2$.

4. The cluster problem

Intimately related with percolation theory is the study of the distribution of white and black clusters when the edges (or vertices) of a graph are painted white with probability p and black with probability $q = 1-p$.

Again we shall concentrate on the edge problem for the square lattice. A white *cluster* is a maximal connected subset of black edges of the lattice. For example, in Fig. 6 where the heavy lines indicate the edge is painted white there are ten black clusters containing at least one edge and four isolated vertices at A, B, C, D, giving a total of fourteen clusters. We note that some writers do not count these isolated vertices as clusters. The two main quantities of physical interest are: (a) the average number of white clusters; (b) the average number of vertices in a white cluster.

To be more precise let L_m denote a square section of the square lattice containing m^2 vertices and hence $2(m-1)^2$ edges. If ω denotes a particular black/white painting of L_m then let $c_m(\omega)$ denote the number of white clusters and let its average value over all paintings ω be denoted by $K_m(p)$.

Similarly if we let the distinct clusters under ω be labelled $A_1, \ldots, A_{c(\omega)}$, we define

$$S_m(p) = \mathscr{E}\left[\frac{|V(A_1)|+\ldots+|V(A_{c(\omega)})|}{c_m(\omega)}\right]$$

where $|V(A_i)|$ denotes the number of vertices in A_i, and $\mathscr{E}$ denotes the expectation or averaging operator over all black-and-white paintings. Thus $S_m(p)$ is the average number of vertices in a white cluster.

Note that if isolated points are not counted as clusters then the expected number of clusters in this sense is given by $K_m(p)-m^2q^4$ where $q = 1-p$. This is because the probability that a particular vertex forms an isolated cluster is just the

probability that the four edges incident with it are painted black, that is q^4 where $q = 1-p$. Thus the average number of isolated points amongst the white clusters is m^2q^4.

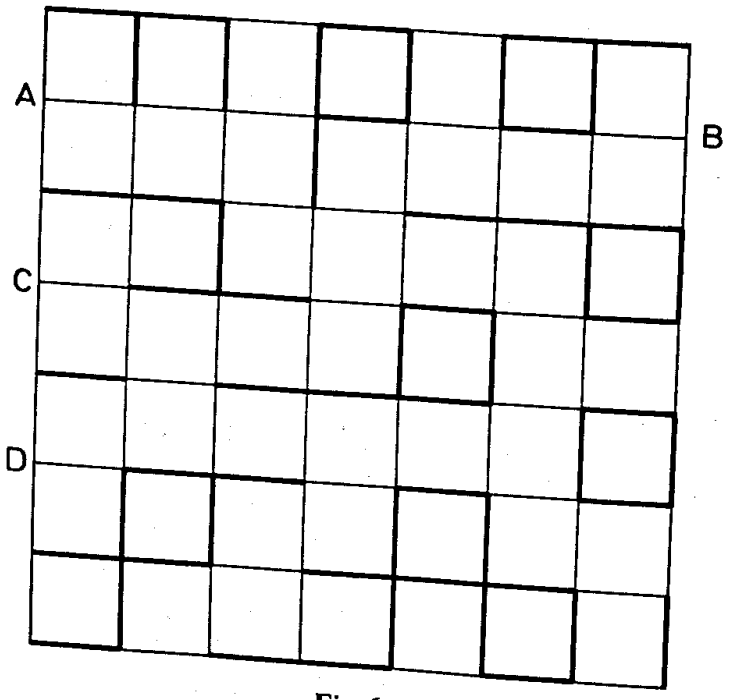


Fig. 6

The average number of black clusters is obviously $K_m(1-p)$ and the average number of vertices in a black cluster is $S_m(1-p)$. Very little is known theoretically about either of these functions. Methods of Grimmett[13] show that

$$K_m(p) \sim m^2\lambda(p) \qquad \text{as } m \to \infty$$

where λ is an undetermined function of p, moreover as $m \to \infty$

$$\frac{C_m(p)}{m^2} \to \lambda(p)$$

with probability one.

Temperley & Lieb[33] have related the problem of enumerating clusters (of at least one white edge) with classical graph colouring problems, and have shown that

$$\lim_{m\to\infty} K_m(\tfrac{1}{2}) \sim 0.098\, m^2$$

It is rather curious that their methods do not seem to extend to $p \neq \frac{1}{2}$.

Roughly speaking the quantities $K_m(p)$ and $S_m(p)$ are reciprocal, though theoretically all that can be proved is that

$$S_m(p) \geqslant m^2/K_m(p)$$

For p greater than the critical probability p_H we have positive probability of an infinite white cluster in L_∞. Hence *a fortiori* as $p \to p_H$ the average size of a cluster tends to ∞. Numerical evidence of Sykes, Gaunt & Glen[32] suggests that as p approaches p_H from below there exists constants C and γ such as $m \to \infty$ $S_m(p) \to S(p)$ where

$$S(p) \sim C(p_H - p)^{-\gamma}$$

where moreover γ is an invariant depending only on the dimensionality of the lattice. A conjectured form of $S(p)$ is therefore shown in Fig. 7.

Probably one of the most interesting results on the cluster problem has been rigorously proved by Harris.[15] It says:

Theorem

Consider the edge cluster problem on the infinite square lattice. If p is strictly greater than the critical probability p_H then with probability one the set of white edges contains only one infinite component.

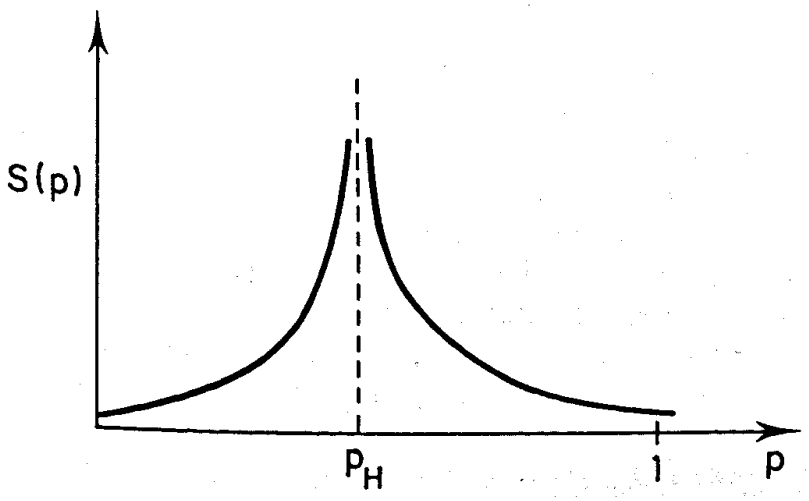


Fig. 7

5. The critical probability or probabilities

As stated earlier, p_H, the critical probability, proved to exist by Broadbent & Hammersley,[2] is defined to be the critical value below which there is zero probability that fluid from a source at the origin spreads to infinitely many points. At least two other 'critical probabilities' occur in the literature and there is some considerable confusion about the relationship between them. The first, p_T, is defined to be the critical value of p above which the *average number* of points wet by fluid from the origin becomes infinite. Now it is easy to see that if there is a positive probability that infintely many points are wet then *a fortiori* the average number of points wet is infinite. Thus for any lattice.

$$p_T \leqslant p_H \qquad (1)$$

Proving equality in (1) seems to be very difficult. Sykes & Essam[31] in a very ingenious paper obtained some very elegant results about a quantity p_E which they call the critical probability but which is defined in terms of singularities of functions giving the mean number of clusters on the lattice. For bond percolation on the square lattice L for example they prove

$$p_E(L) = \tfrac{1}{2}$$

and for the triangular lattice T and hexagonal lattice H they show that

$$p_E(T) = 2 \sin(\pi/18) = 1 - p_E(H).$$

It seems to be extremely difficult to relate p_E with either of the other two 'critical probabilities' p_H and p_T, and physically it does not appear (from its definition at least) to be as natural an object as p_H or p_T. Exact rigorous bounds for p_H and p_T seem difficult to obtain. For example, for the bond percolation problem on the square lattice the best theoretical results are those of Harris[15] and Hammersely[16] which show that

$$0.5 \leqslant p_H(L) \leqslant 0.646790$$

Much closer numerical bounds have been obtained by Domb.[3] For a good survey of the different estimates of the 'critical probability' for the different lattices see the review papers of Essam[6] and Shante & Kirkpatrick,[29] though we should point out that in many cases it is not clear 'which critical probability' has been numerically calculated. Indeed throughout the literature little attention has been paid to the difference between these three critical probabilities and there is a 'folk lore myth' that the estimation of critical probabilities for bond percolation on the square, triangular and hexagonal lattices has been settled by the Sykes–Essam result for p_E. It would be most interesting mathematically and perhaps even physically if the three critical probabilities p_H, p_E, p_T turned out to be all different.

6. First passage percolation

Consider a population $P_1, P_2, \ldots, P_i, \ldots$ of individuals set along the integer points of the real line. Suppose that an individual P_i is infected and that in any time $(t, t+\delta t)$ it has probability $\lambda \delta t + o(\delta t)$ of infecting P_{i+1}. If P_0 is infected at time $t = 0$, the number N_t of individuals infected by time t is the well-known Poisson process, whose distribution is given by

$$P[N_t = k] = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \qquad (k = 0, 1, \ldots)$$

A more general situation is where individuals such as trees are situated at the vertices of a graph G in which neighbours are joined by an edge and each tree, once infected has probability $\lambda \delta t + o(\delta t)$ of infecting any of its neighbour, independent-

ly for each neighbour. The problem is completely equivalent to the following problem. Assign to each edge e_i of the graph G a random variable t_i drawn from the exponential distribution with density function

$$f(x) = \begin{cases} e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases} \tag{1}$$

We call t_i the *time coordinate* associated with the edge e_i joining u to v and regard t_i as the time taken for u to infect v. The problem is to see how far disease spreads in a given time.

First passage percolation is the more general problem when to each edge of the directed graph G we associate a random variable drawn from a fixed distribution and then try to determine the probability distribution of the fastest time between pairs of points of the graph. For example, let G be the graph shown in Fig. 8.

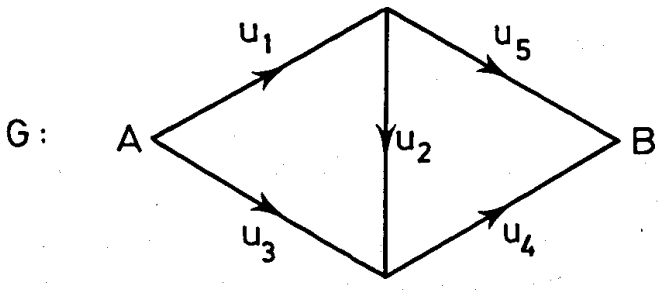


Fig. 8

The *first passage time* t_{AB} between A and B is given by

$$t_{AB} = \min \begin{pmatrix} u_1 + u_5 \\ u_1 + u_2 + u_4 \\ u_3 + u_4 \end{pmatrix}$$

Even when G is such a small and relatively simple graph, finding the distribution of t_{AB} is non-trivial. This is because if two paths P_1, P_2 have a common edge the times to travel along them are dependent random variables and there is no easy way of calculating their distribution.

A very crude technique, but one which is used quite extensively, is to reduce the random problem to a deterministic problem by replacing each random variable by its expected value and then to solve the associated deterministic problem. Thus if the u_i above have expected value $\bar{u}$, we would have

$$\mathscr{E} t_{AB} \leqslant \mathscr{E} \min \begin{pmatrix} u_1 + u_5 \\ u_1 + u_2 + u_4 \\ u_3 + u_4 \end{pmatrix}$$

$$\leqslant \min \mathscr{E}\begin{pmatrix} u_1 + u_5 \\ u_1 + u_2 + u_4 \\ u_3 + u_4 \end{pmatrix}$$

$$= 2\,\bar{u}$$

So far first passage percolation theory has received far less attention than classical percolation theory. We describe briefly the main results of the theory obtained so far.

Consider the infinite square lattice with vertices at all integer points (m,n) of the plane. Suppose that to each edge e_i of this lattice we independently assign a non-negative random 'length' or *time coordinate* u_i drawn from a given distribution $F(x)$.

Let t_n be the random first passage (shortest) time from the origin to $(n,0)$ in this lattice and let $\tau(n)$ be its expected value over all possible distribution of time coordinates u_i.

The basic result of Hammersley & Welsh[20] is that there exists a *time constant* μ such that as $n \to \infty$

$$\frac{\tau(n)}{n} \to \mu$$

The constant μ depends only on the distribution F. It is clearly not greater than $\bar{u}$ the mean of the distribution F since it is obvious that for any n,

$$\tau(n) \leqslant n\bar{u}$$

If we denote by s_n the fastest time from the origin to the line $x = n$, it is clear that $s_n \leqslant t_n$. However, a second result of Hammersley & Welsh[20] is that if $\psi(n)$ is the expected value of s_n

$$\frac{\psi(n)}{n} \to \mu$$

as $n \to \infty$. Later work of Kingman[21] proves the stronger result that as $n \to \infty$, with probability one, both t_n/n and s_n/n converge to this time constant μ.

The constant μ can be interpreted as the reciprocal of the *limiting velocity* of motion in the square lattice. Unfortunately it seems to be very difficult to obtain an exact analytic expression for μ, and indeed the best known theoretical bounds for μ seem to be weak. For example when the u_i are uniformly distributed between 0 and 1 the best theoretical bounds to date are

$$0 \leqslant \mu \leqslant 0.425$$

whereas Monte Carlo simulation carried out by Welsh[34] suggests

$$\mu \simeq 0.323$$

A similar problem has been studied by Morgan & Welsh[26] and Hammersley.[18] this is a study of first passage percolation on the non-negative quadrant of the square lattice oriented in the positive x and y directions and with exponential time coordinates. Loosely speaking this is a model for the spread of infection through an orchard in the presence of a very strong south-west wind! For example if time coordinates are as shown in Fig. 9, the set of trees infected by time 17 is marked with

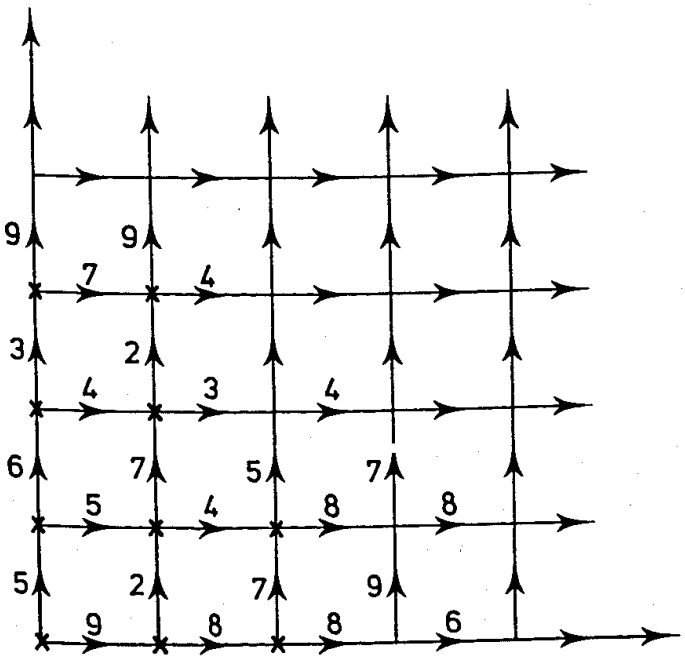


Fig. 9

a cross. Well-known probability theory shows that if X_t denotes the coordinate of the furthest point on the x-axis infected by time t then $X_t \simeq \lambda t$ where λ is a constant of the exponential distribution (1). More generally if I_t denotes the set of points infected by time t, we let N_t be a point of I_t of maximum distance away from the origin and let the *frontier* R_t be the subset of infected points which have a non-infected neighbour. If $M(t)$ denotes the expected value of N_t it is proved by Morgan & Welsh[26] and Hammersley[18] that there exists a constant C, the *velocity* of the process such that

$$\lim_{t\to\infty} \frac{M(t)}{t} = C\lambda$$

and that $M(t)$ is differentiable with a derivative satisfying

$$\frac{dM(t)}{dt} = 2\lambda R(t) \qquad (2)$$

where again λ is the parameter of the underlying exponential distribution. The constant C is shown to lie between 2.18 and 4.31 though numerical evidence suggests that its true value is a little more than 3.4. What is interesting about (2) is that it is a first analytic result expressing the intuitively appealing idea that the rate of growth of a spatial epidemic is proportional to the size of its frontier. This idea has been extended by Mollison[25] to other epidemics. An interesting open problem is to obtain some result of the same nature as (2) when the underlying distribution is not exponential.

For a more recent work on the ideas of this section see Hammersley,[19] Kingman[22] or Smythe.[30]

7. Growth processes

Closely related to the percolation processes described above is a class of growth processes studied by Richardson.[27] His family of processes includes first passage percolation and the model studied by Williams & Bjerknes,[35] for the spread of an abnormal clone in the basal layer of the epithelium. It also includes an interesting growth process studied by Eden[4] and various spatial epidemic models.

Let S be two-dimensional Euclidean space and let T be a division of S into cells, such that T has only finitely many cells in any bounded region of S. A *configuration* of T is a painting of the cells with the colours black and white. The *black boundary* of the configuration C is the set of black cells which have at least one white neighbour. A *growth process* in its widest sense starts with a given configuration at time $t = 0$ and a random mechanism which governs the change of configurations as time progresses. This is best illustrated by the following examples:

The growth process Gp

Let T be a regular tessellation or honeycomb of two-dimensional space, say the triangular lattice of Fig. 3. Time progresses in discrete jumps $t = 0, 1, 2 \ldots$ At time $t = 0$, the cell at the origin is black and all other cells are white. If a cell is black at time t it is also black at time $t + 1$. If a cell is white at time t and has one or more black neighbours then it becomes black at time $t + 1$ with probability p. Thus if we draw the dual lattice got by inserting a point in each face and assign to each edge of this dual lattice a time coordinate u where u is a random variable taking the values

$$P[u = k] = q^{k-1}p \qquad k = 1, 2, \ldots \tag{1}$$

we see that the time for a cell to become black in this Gp process is exactly the first passage time from the origin to this cell in their dual lattice with the geometric time distribution given by (1).

The growth process Ga(p)

This process is exactly like Gp except that a white cell with n black neighbours at time t becomes black at time $t + 1$ with probability $a(n,p)$ where $a(n,p)$ is a func-

tion of n and p which is greater than 0 and less than or equal to 1 for all n and all p which are greater than zero. A simple argument shows that this is almost exactly the same as first passage percolation on the dual tessellation with the restriction that the time coordinates u_i have to take integer values.

A growth process with deaths

Again this is a growth process of the same type as Gp except that black cells do not remain black permanently. Instead if a cell is black at time t and has a white neighbour it becomes white at time $t+1$ with probability q.

An interesting problem in connection with this process is finding conditions which guarantee the survival of the black cells. A fairly obvious necessary condition is that $q < p$. As far as I know it is still not known whether this is sufficient.

An epidemic model

Divide the plane into squares and start the process at time $t=0$ with one particle in each square. Suppose that the particle at the origin is unwell and all the others are healthy. As time passes the particles follow independent Brownian motion trajectories. A square becomes black (= unhealthy) at the first instant it contains a sick particle. A healthy particle becomes unwell as soon as it shares a square with an unwell particle.

The Eden growth process

This is probably the simplest growth process and was first studied by Eden.[4] The plane is divided into squares. At time $t=0$ the one cell is black and all other cells are white. At stage $t=n+1$ we pick at random an edge from the boundary of the black configuration at time n, and make the white cell sharing this edge black. Thus at time $t=n$ we have a connected configuration consisting of exactly n black cells.

Thus after say three stages of time the possible configurations (with origin

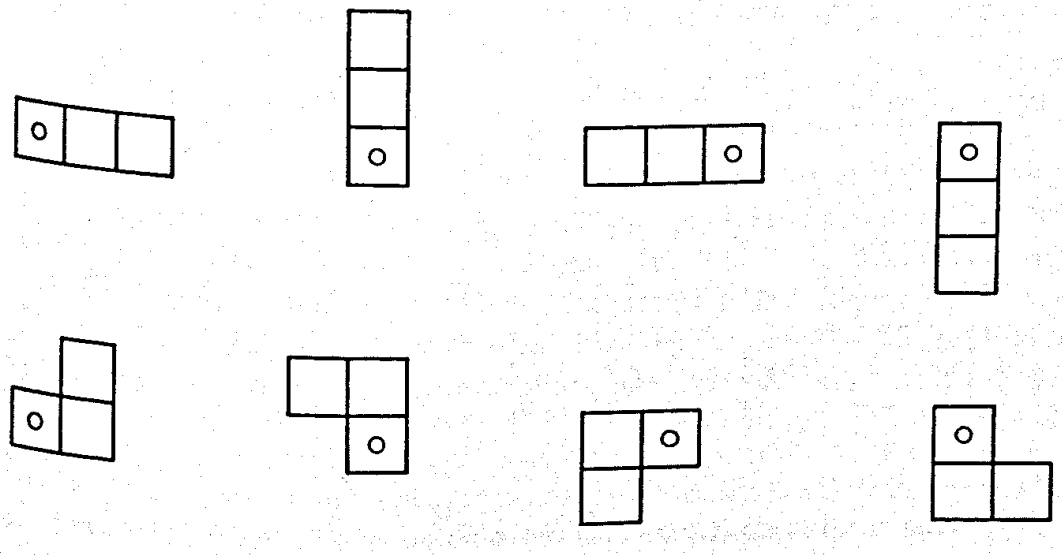


Fig. 10

marked) are as shown. Let C_n be the set of all possible n-cell configurations. With each $c \in C_n$ associate a probability $P(c)$ that at stage n of the Eden process the configuration is in fact c.

Problem. Let t_n = the time for the point $x = n$ to become black. Obviously $t_n \geqslant n$ and by symmetry arguments $t_n \geqslant 2$.

An interesting and still unsolved combinatorial problem in connection with this is enumerating how many different configurations there are in C_n. For details of computer counts and recent results see Lunnon.[23]

Richardson[27] studies the 'shape' of the black configuration in these processes as time progresses. His formal results are not easily stated and in some ways pose more problems than they solve! For example in the process Gp computer simulations indicate that as p varies from 1 to 0 the 'shape' of the black configuration varies from a diamond to a circle. Roughly speaking as p decreases a typical boundary of the black region tends to become more jagged. At the moment there is no theoretical reason for this.

8. Related problems

We conclude with a brief description of some related problems.

The Ising model of ferromagnetism

Consider a finite graph G. With each vertex v of G associate a *spin* σ_v which can take the values ± 1. If there is a coupling constant J_e associated with each edge $e = (u,v)$ of G the Hamiltonian of the system ω is defined to be

$$H(\omega) = -\sum_{e \in E(G)} J_e(\sigma_e - 1)$$

where $E(G)$ is the edge set of G and σ_e the spin of the edge e under ω is given by

$$\sigma_e = \sigma_u \sigma_v$$

where σ_u, σ_v are the spins assigned by ω to its endpoints u,v. Thus if $J_e = J$ for all e,

$$H(\omega) = 2J(|E(G)| - N(\omega))$$

where $N(\omega)$ is the number of edges whose endpoints have like spins.

There is an obvious similarity between this and the cluster problem on the vertices of a graph. For a rigorous account of the interrelation between this, classical percolation and other combinatorial-type problems of mathematical physics such as the Ashkin-Teller model, and Potts problem see Fortuin & Kasteleyn.[10]

Random clumping

Many practical problems reduce to the simple model in which laminae of negligible thickness are scattered randomly and independently on a plane surface and the

problem is to determine the average number of 'clumps' or the average size of a clump. Typical examples arise in the study of quantitative problems connected with dust particles in the air, bacterial counting, or even in one dimension traffic flow. For a detailed account of such phenomena and the various approaches to settling them we refer to the monograph of Roach.[28] Basically the situation is that except in very special cases no exact analytic solution has been found and various approximations or simplifications have to be made. We show here how one such approximation given by Roach (Chapter 5) is essentially a percolation problem.

Suppose that the laminae are placed at random points on a lattice and that the assumption is made that only adjacent laminae overlap. That is, clumps are formed by linking only those laminae whose centres are on adjacent points of the square lattice. If now we take p as the probability that a laminae falls on a particular vertex of the square lattice L we have exactly the cluster problem for the vertex percolation model, in which clusters correspond exactly to clumps.

As a measure of the difficulty of the percolation model Roach in Chapter 6 considers the clumping problem (= vertex percolation) model for the $2 \times n$ lattice.

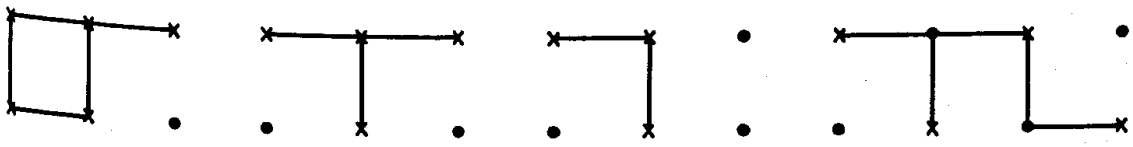


Fig. 11

This model has been solved, for details see Roach.[28]

PERT networks

The concept of a *PERT* or *critical path* network was first used in 1959 for measuring and controlling development progress for the Polaris missile programme. Since then it has been widely used in operations research—for a bibliography see Bigelow.[1]

A *PERT* network consists of a network which has no directed cycles. The edges of the network represent jobs and the orientation and arrangement of the edges depends on the precedence and priorities of the jobs. For example if a project consists of four jobs J_1, J_2, J_3, J_4 in which jobs J_1, J_2 must be completed before J_3 is started and J_3 must be completed before J_4 is started the corresponding network is shown in Fig. 12. Now suppose that the jobs are assigned duration times; this is equivalent to assigning lengths to the edges of the network. The time to complete the whole project is then the length of the longest path in the network. In many applications the time needed to complete any job is uncertain—that is it is taken to be a random variable from a known distribution. The problem then is to find a longest route in a random network. Since the network has no directed cycles this is easily seen to be equivalent to a random shortest route problem—namely first passage percolation.

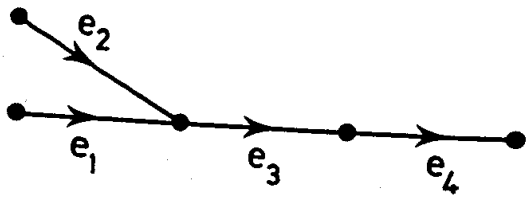


Fig. 12

Colouring problems

By a *k-colouring* of the vertices of a graph G we mean an assignment of one of k-colours $C_1 \ldots . C_k$ to each vertex of G in such a way that if two vertices are joined by an edge they have different colours. Probably the best known unsolved problem in combinatorial mathematics is deciding whether or not every planar graph has a four-colouring.

It is well known that for any graph G the number of distinct k-colourings is a polynomial in k, $P(G; k)$ called the *chromatic polynomial* of the graph G. Various authors, in particular Temperley & Lieb,[33] Fortuin & Kasteleyn[10] and Essam,[5] have noticed that mathematically the problem of determining the chromatic polynomial of a graph G is intimately related to the cluster problem for G. In particular outstanding open mathematical problems whose solutions would surely throw considerable light on the percolation problem are: (a) determine the chromatic polynomial $P_n(\lambda)$ of the $n \times n$ square lattice; (b) it is known that as $n \to \infty$ $[P_n(\lambda)]^{1/n}$ converges to a limiting function $P(\lambda)$ in the sense of a thermodynamic limit of statistical mechanics, find $P(\lambda)$.

References

1. Bigelow G.G. (1962) *Operations Res.* 10, 728.
2. Broadbent S.R. & Hammersley J.M. (1957) *Proc. Camb. phil. Soc.* 53, 629.
3. Domb C. (1964) *J. Roy. Statist. Soc.* B 26, 367.
4. Eden M. (1961) *Proc. Fourth Berkeley Symposium* 4, 223.
5. Essam J.W. (1971) *Discrete Math.* 1, 83.
6. Essam J.W. (1972) *Phase Transition and Critical Phenomena* (ed. by C. Domb and M. S. Green), Vol. 2, p. 197. Academic Press, London.
7. Essam J.W. & Sykes M.F. (1966) *J. Math. Phys.* 7, 1573.
8. Erdös P. & Renyi A. (1960) *Publ. Math. Inst. Hung. Acad. Sci.* 5A, 17.
9. Fisher M. (1961) *J. Math. Phys.* 2, 620.
10. Fortuin C.M. & Kasteleyn P.W. (1972) *Physica* 57, 536.
11. Frank H. & Frisch I.T. (1971) *Communication, Transmission and Transportation Networks.* Addison-Wesley, New York.
12. Gilbert E.N. (1959) *Ann. Math. Statist.* 30, 1141.
13. Grimmett G. (1976) *J. London Math. Soc.* (to appear).
14. Grimmett G. & McDiarmid C.J.H. (1975) *Proc. Camb. phil. Soc.* 77, 313.
15. Harris T. (1960) *Proc. Camb. phil. Soc.* 56, 13.
16. Hammersley J.M. (1959) *Colloques Internationaux du Centre National de la Recherche Scientifique* 87, 17.
17. Hammersley J.M. (1961) *J. Math. Phys.* 2, 728.

18. Hammersley J.M. (1966) *J. Roy. statist. Soc.* **28**, 491.
19. Hammersley J.M. (1973) *Annals Probability* **1**, 910.
20. Hammersley J.M. & Welsh D.J.A. (1965) *Bernoulli, Bayes, Laplace Anniv. Volume*, p. 61. Springer, Berlin.
21. Kingman J.F.C. (1968) *J. Roy. statist. Soc.* **30**, 499.
22. Kingman J.F.C. (1973) *Annals Probability* **1**, 883.
23. Lunnon W.F. (1972) *Graph Theory and Computing* (ed. by R.C. Read), p. 87. Academic Press, London.
24. Mauldon J.G. (1960) *Proc. Fourth Berkeley Symposium*, p. 337.
25. Mollison D. (1972) *Proc. Sixth Berkeley Symposium* **3**, 579.
26. Morgan R.W. & Welsh D.J.A. (1965) *J. Roy. statist. Soc.* B **27**, 497.
27. Richardson D. (1973) *Proc. Camb. phil. Soc.* **74**, 515.
28. Roach S.A. (1968) *The Theory of Random Clumping.* Methuen, London.
29. Shante V.K.S. & Kirkpatrick S. (1971) *Adv. Phys.* **20**, 325.
30. Smythe R.T. (1976) *J. appl. Prob.* (to appear).
31. Sykes M.F. & Essam J.W. (1964) *J. Math. Phys.* **5**, 1117.
32. Sykes M.F., Gaunt D.S. & Glen M. (1976) *J. Phys.* A **9**, 97.
33. Temperley H.N.V. & Lieb E.H. (1971) *Proc. Roy. Soc. Lond.* A **322**, 251.
34. Welsh D.J.A. (1965) *J. appl. Math. Phys.* **16**, 520.
35. Williams T. & Bjerknes R. (1972) *Nature* **236**, 3.
36. Wing O. & Demetriou P. (1964) *IEEE Trans. Communications Technology*, p. 38.