University of Trento
Master's Degree in Physics

Percolation - Theory and Applications

Student: *Guglielmo Grillo*

Course: *Multi-scale methods in soft matter physics [145889]* – Teacher: *Prof. R. Potestio*
Date: *February 1, 2023*

1 Abstract

The Ising model is usually the archetypical model of phase transition for physicists due to its close link with magnetism. On the other hand, mathematicians are typically introduced to the concept of phase transition via a different model called percolation. My aim with this review is to present an introduction to the Bernoulli percolation model and prove that a phase transition does indeed exist in any dimension $d \geq 2$. In the final part of the review, I'll also present attempts to apply Bernoulli percolation to physical problems.

2 Introduction

2.1 Wetting of a Porous City

Naples' underground presents a huge network of tunnels dug in the tuff. Part of those tunnels date back to the Roman era and were used to supply running water to the whole city by filling wells beneath each building. As the tunnels were dug into the tuff, collapses and clogs were not infrequent. This was not a big deal as long as a path from the water source to the sea was present. But if enough tunnels got closed the city's water supplies would eventually run off or stagnate leading to epidemics. Conceptually, this intricate network can be viewed as a graph where the nodes are the buildings' wells and the edges are the tunnels connecting them. Add on top of this Naples' grid plan and the sewers can be thought of as a planar square lattice. The closing of a tunnel due to collapse/clog, which can be assumed to be independent from the state of the others, can be modeled as removing the edge between two lattice points. The question now is: what's the critical clog probability at which the sewers need constant maintenance and occasional cleaning is not enough? The answer to this question can be found using a model called Bernoulli percolation.

2.2 Bernoulli percolation

Let $\mathcal{Z}^2$ be the plane square lattice and let p be a number satisfying $0 \leq p \leq 1$. Each vertex is declared **open** with probability p or **closed** with probability $1-p$ independently of all the other edges[1]. Once a vertex is declared open (closed) it can't change its state.

[1] In contrast to Naples' example, Percolation theory is formulated in terms of open edges instead of closed (clogged) edges. This is not a big deal as we can always require $p_{clog} = 1 - p_{open}$. Furthermore, this mapping will come back later when we introduce the dual lattice.

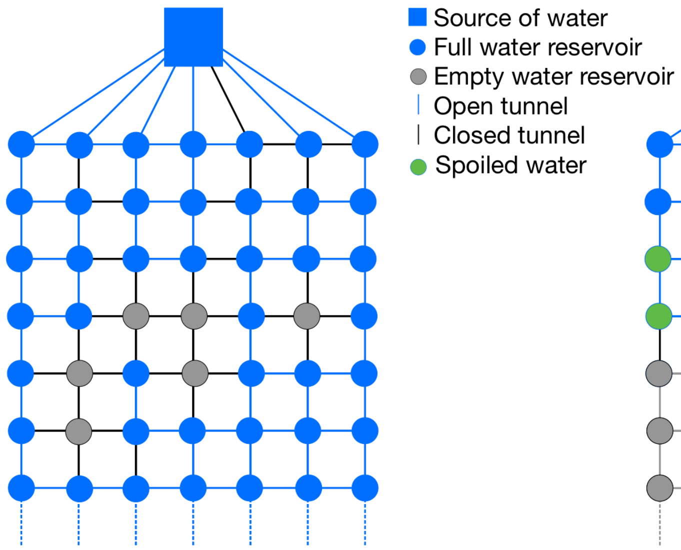

Figure 1: Schematic representation of Naples' underground. If all the tunnels around a reservoir are closed the well runs out of water quickly.

This process will form **clusters**, regions of space whose points are all reachable from one to the other using only open edges. In the extreme case of $p = 1$ the graph itself will be the only cluster available. On the other hand if $p = 0$ each point is a cluster on its own. The real question is: what happens for intermediate values of the probability? Intuitively,

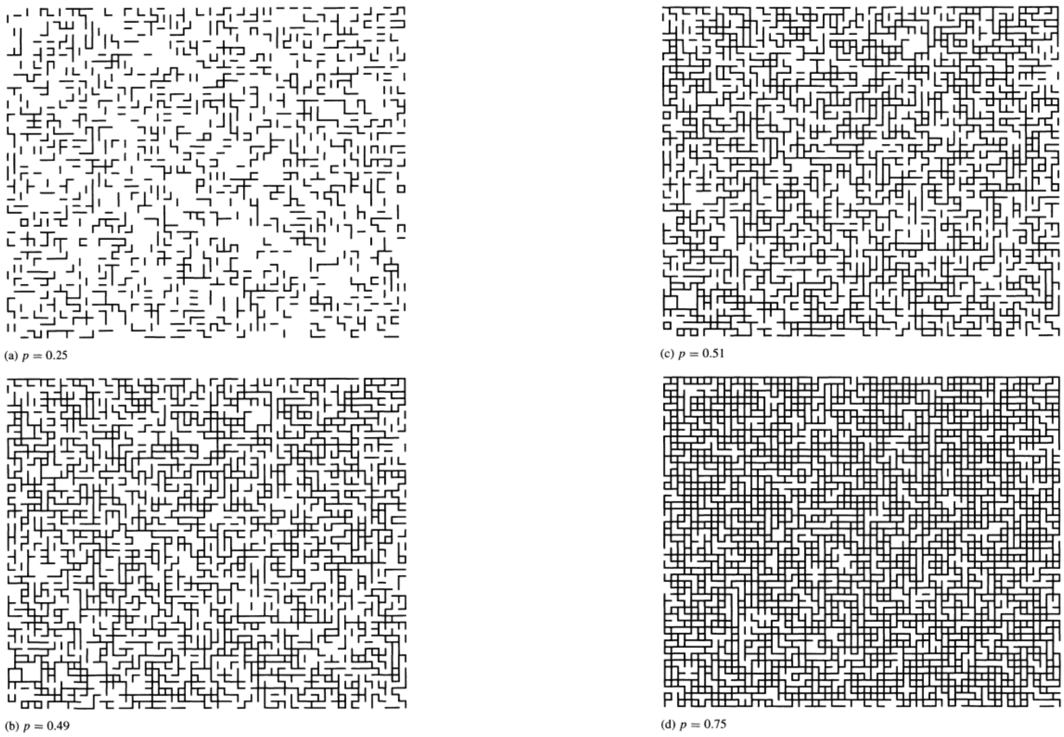

Figure 2: Samples of percolation for various values of p. Image borrowed from [1].

there are two different regimes (phases) separated by a critical point. In the supercritical phase there exists at least one open path from top to bottom while in the subcritical regime it doesn't. In the case of an infinite grid, this problem can be reformulated by asking if there exists a path starting from a point called **origin** and reaching infinite. As one can imagine, a finite grid does introduce finite-size effects, but these are negligible for large enough grids.
Once the existence of two phases is proved, a few more questions arise: what is the exact value of the critical point? Do we really need the infinite path to start from the origin? What happens around the critical point? The answer to these questions will be provided in the next chapter.

3 Percolation and the existence of the critical point

This section will present the mathematics behind Percolation and prove the existence of a critical point. The main reference for this part will be *Percolation* by *G. Grimmett* [1], which is also the reference for the notation used.

3.1 Bond Percolation

3.1.1 Mathematical fundaments

Let's take a lattice $\mathcal{Z}^d$ where d is an integer number representing the number of dimensions. Given a particular value of d, the points of this space are represented as $\vec{x} = (x_1, ..., x_d)$ and their distance is computed using a taxicab metric:

$$\delta(\vec{x}, \vec{y}) = \sum_{i=1}^{d} |x_i - y_i| \qquad \vec{x}, \vec{y} \in \mathcal{Z}^d. \tag{1}$$

Any particular point of the lattice can be taken as the origin o. This allows defining a pseudodistance from the origin:

$$||\vec{x}|| = max\{|x_i| : 1 \leq i \leq d\}\,. \tag{2}$$

This second definition allows to easily quantify the distance of a point from the origin but loses the notion of order relation.
Consider now a graph whose vertices are the lattice point of $\mathcal{Z}^d$ and place an edge between any couple of points with distance $\delta(x_i, x_j) = 1$. As already stated, each one of these edges can either be open or closed with uniform probability p independent from the state of all the other vertices.

3.1.2 Clusters

If we remove all the closed edges from the graph we are left with a partially connected graph. We define as **cluster** C a set of points such that any point inside the cluster can be reached by any other point using only open edges[2]. A single point is a cluster of its own. An **infinite cluster** is a cluster that contains both the origin o and at least one point x^* such that $||x^*|| = \infty$ in the general sense. We define the length of a cluster using the maximum distance between two points: $|C| = max\{||\vec{x} - \vec{y}|| : \vec{x}, \vec{y} \in C\}$.
The notion of cluster is closely linked to the notion of path. As all the points inside a cluster are reachable from any other point of the same cluster, it is always possible to find a path that connects any two points of a cluster. At the same time, a path represents a cluster because any point of the path can reach any other point of the path by using the path itself. An infinite path is a path that starts from the origin and reaches the point $x_* : ||x^*|| = \infty$. An infinite cluster contains at least one infinite path. We will use the term cluster and path interchangeably when possible.

3.1.3 Excursus: Uniform coupling

In order to simplify the problem at hand [2] suggests to introduce the uniform coupling. This coupling changes the role of the parameter p. Instead of considering p as the probability that an edge is open, each edge is assigned a uniform random number $u \in [0, 1]$. If

[2]Every cluster is a connected component (connected sub-graph) of the original graph.

$u \le p$ the edge is considered open, otherwise it is closed. The difference between the two definitions becomes apparent once we change the value of the parameter p on the same graph. In the first case, it is necessary to "reroll the dice" from scratch and generate a new set of vertices. In the second case, it is just a matter of checking which vertices were changed by changing the parameter. Other than speeding up numerical simulations, this coupling assures that, given a graph $\mathcal{G}$, if $p < p'$ then the number of open edges satisfies $Edges(\mathcal{G}_p) \le Edges(\mathcal{G}_{p'})$. This might not always be the case in the uncoupled Bernoulli percolation due to statistical fluctuations.
Remark: I didn't find anything related to the equivalence between the Bernoulli percolation and the uniform coupling. My guess is that the uniform coupling is a way to express the average / expected behavior of Bernoulli percolation.

3.1.4 Percolation probability

The *percolation probability* $\theta(p)$ is defined as the probability that the cluster of the origin is an infinite cluster. Formally, this quantity is expressed using the complementary event, that is the probability that the origin does not belong to a finite cluster:

$$\theta(p) = P_p(|C| = \infty) = 1 - \sum_{n=1}^{\infty} P_p(|C| = n) \tag{3}$$

Intuitively, this probability will be zero up to a value and then start rising to one. Notice that we don't know if this "rising" is continue or abrupt. It might be the case that there exists a jump from zero to any other finite value greater than zero. The critical point p_c is the point that distinguishes these two phases and can be defined as:

$$p_c = sup\,\{p : \theta(p) = 0\} \tag{4}$$

We will now prove the existence of the critical point for a cubic lattice of any dimension $d \ge 2$.

3.2 Critical Point - lower bound

This part of the proof is taken from [2] and is called *Peierls argument. Grimmett* [1] uses a similar approach but expresses the result in terms of the connective constant $\lambda(d) = \lim_{n\to\infty} [\sigma(n)]^{\frac{1}{n}}$ where $\sigma(n)$ is the number of self-avoiding paths of a lattice $\mathcal{L}^d$ whose length is n. This choice allows to maintain the result independent from the specific lattice choice but makes the argument less clear.

As an infinite path contains a path of length l, the probability that an infinite path exists is smaller than the probability that any path of length l exists.

$$\theta(p) = P_p(|C| = \infty) \le P_p(|C| = l) \tag{5}$$

In turn, this probability is bounded by the probability that l consecutive edges are open times the number of paths of length l. If crossing is allowed but backtracking is forbidden, at each vertex after the first one there are $2d-1$ possible choices for the direction to take[3]. This provides a further upper bound.

[3]Given a point, there are d direction to choose from and for each direction we can either move forward or backward. If this is not the first point, one of the directions is the one where we come from and must be excluded.

$$\begin{aligned}\theta(p) &\leq P_p(|C| = l) &(6)\\ &\leq p^l \cdot \#\{\text{of paths of length l}\} &(7)\\ &\leq p^l \; 2d \; (2d-1)^{l-1} = \frac{2d}{2d-1}\,[p(2d-1)]^l &(8)\end{aligned}$$

If we take the limit as $l \to \infty$, the term on the right is zero for every value of $p < \frac{1}{2d-1}$. This proves that, given a finite number of dimensions d, there exists a non-zero lower bound for the critical probability:

$$p_c \geq \frac{1}{2d-1}. \tag{9}$$

3.3 Critical Point - upper bound

First of all, let's notice that a d-dimensional lattice $\mathcal{L}^d$ can be embedded in a $d+1$-dimensional lattice $\mathcal{L}^{d+1}$. This means that if $\mathcal{L}^d$ percolates at p_c^d so must $\mathcal{L}^{d+1}$ implying $p_c(d+1) \leq p_c(d)$. With this in mind, the upper bound for $d = 2$ must be an upper bound for all dimensions $d \geq 2$.

3.3.1 Dual graph

Secondly, we have to introduce the notion of dual graphs. A way to do so is to give a definition by construction:

1. Take a planar graph $\mathcal{G}$ and place a vertex inside each face of the graph.
2. If the graph is finite, place a vertex outside the graph. This corresponds to the infinite face of the graph.
3. If two faces are adjacent, i.e. they are separated by an edge of $\mathcal{G}$, draw an edge connecting them.
4. The edges and the vertices just drawn constitute the dual graph of the original graph.

In *Figure* 3 it is possible to see the construction of the dual graph for the two-dimensional cubic lattice. Notice how $\mathcal{Z}^2$ is self-dual, that is, its dual is just $\mathcal{Z}^2$ itself translated.
To finally start proving that an upper bound exists, we need to couple the primal grid (i.e. the original graph) with its dual. To do so we state that an edge of the open graph is open if and only if the edge of the primal grid that it would intersect is closed and vice-versa. This is the same as requiring that the edges of the dual are open with probability $p_{dual} = 1 - p_{primal}$. Due to this coupling, any finite cluster starting from the origin of the primal grid will be encircled by a closed path in the dual grid[4].

[4]The converse is also true. Finite paths in the dual graph are surrounded by closed paths of the primal grid. This is irrelevant for the proof but will come in handy in finding the correct value of the critical probability.

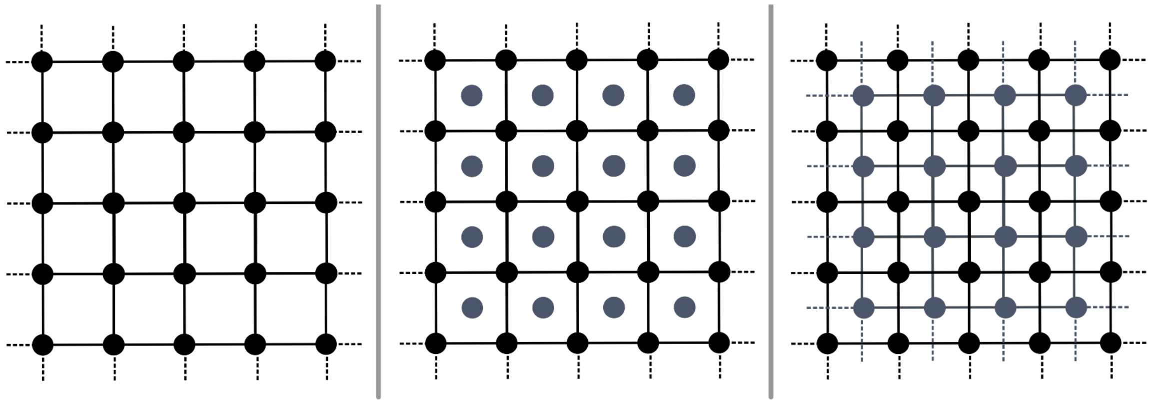

Figure 3: (left) Two dimensional cubic lattice $\mathcal{Z}^2$, also called primal graph. (center) Place a vertex inside each face of $\mathcal{Z}^2$. (right) Connect vertices whose corresponding face share an edge. The dual lattice is made of the vertices and the edges shaded in blue-grey.

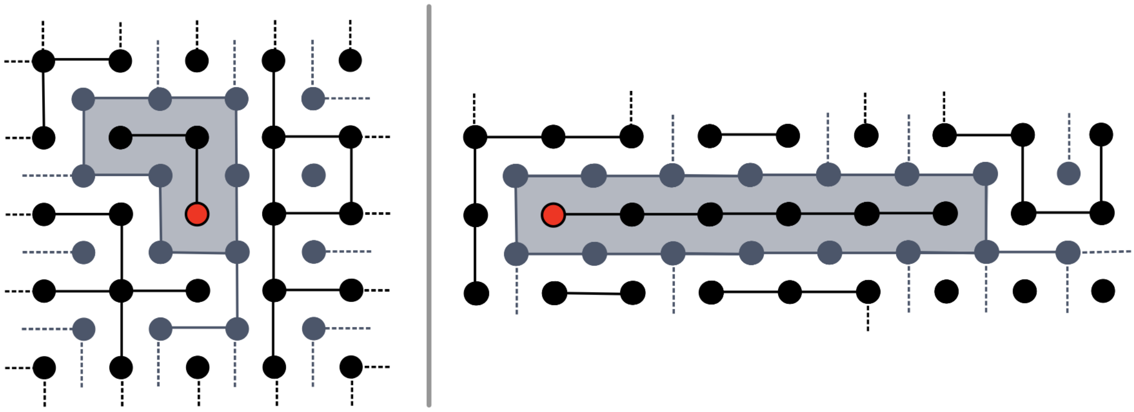

Figure 4: (left) Coupling between the primal graph in black and the dual graph shaded blue-grey. The red point represents a possible choice for the origin. Notice how, with this coupling, any finite path of the primal graph is surrounded by a closed path in the dual graph. An example of this property is highlighted in grey. (right) Highlighted in grey: longest possible path in the primal graph for a closed path of length $l = 14$ in the dual graph.

3.3.2 Upper bound

Let's consider a closed path of length l in the dual grid that encircles the origin. In the worst case scenario, the open cluster can have length at maximum $\frac{l}{2} - 2$. The probability that such closed path exists must be smaller than the probability that any closed path of length l (not necessarily encircling the origin) exists.

$$P\left(\{\text{path of length l encircling the origin}\}\right) \leq P\left(\{\text{any path of length l}\}\right). \tag{10}$$

As before, this is bounded by the probability that l edges are open in the dual graph times the number of possible paths of length l.

$$P\left(\ \{\text{any path of length l}\}\right) \leq p_{dual}^{l} \cdot \#\{\text{of paths of length l}\}\ . \tag{11}$$

As you can guess, the final bound is given releasing once again the constraint and counting all paths of length l that do not backtrack

$$p^l_{dual} \cdot \#\{\text{of paths of length l}\} \ \leq p^l_{dual} \ 4 \cdot 3^{l-1} = \frac{4}{3} \left(3p_{dual}\right)^l . \tag{12}$$

To enclose an infinite path we impose $\lim l \to \infty$. In this limit, the probability that a circle surrounding the origin exists is zero if $p_{dual} < \frac{1}{3}$ and non-zero otherwise. Recalling the coupling $1-p_{primal} = p_{dual}$ we get that an infinite closed path in the dual grid surrounding the origin can not exist if $p_{primal} > \frac{2}{3}$.
What about finite-length closed paths that enclose the origin? Those are not a problem: take the closed path that surrounds the longest cluster. As no longer closed path can exist (by definition), the surface of the outermost closed path must be connected to infinity. If that wasn't the case, an even longer closed path would exist. Finally, inside a finite closed path only a finite number of vertices can exist. This implies that however small, there must exist a finite probability that a path starting from the origin crosses the aforementioned perimeter.
Both the events

- The perimeter of the outermost closed path is connected to infinite and
- The origin is connected to the perimeter

involve different sets of edges, thus the two events are independent. As both events are independent and have finite probability, the probability that both events happen at the same time is the product of the two. This implies that the probability that an infinite cluster exists is finite as it is the product of two different finite quantities. We have thus proven that $p_c(d) \leq p_{primal} = \frac{2}{3}$. Combining this result with the previous one we get:

$$\frac{1}{2d-1} \leq p_c(d) \leq \frac{2}{3} \tag{13}$$

implying that **a phase transition must exist for every** $d \geq 2$.

3.4 Existence of the infinite cluster

Before talking about the actual critical value for percolation probability let's pause for a moment and think about the infinite clusters. What does it mean that there is a finite probability that an infinite cluster exists in an infinite graph? Does it actually exists or not? Both [1] and [2] report the same proof based on *Kolmogorov's Zero-One law*[5] to show that if we relax the requirement by allowing the infinite path to start from any point and not only from the origin the infinite cluster will certainly exist above the critical threshold.

3.4.1 Existence above the critical threshold

Above the critical threshold, there exists a finite probability that the cluster of the origin is infinite. Of course, the probability that there exists an infinite cluster starting from

[5]Kolmogorov's zero–one law states that if some assumptions are satisfied, a tail event will either almost surely happen or almost surely not happen; that is, the probability of such an event occurring is zero or one.

any point must be greater or equal to this:

$$0 < P(\{\text{cluster of the origin is infinite}\}) \quad (14)$$
$$\leq P(\{\exists \text{ infinite cluster on any point}\}) \quad (15)$$

Furthermore, the existence of the infinite cluster does not depend on any set of finite edges in the graph therefore it is a tail event. By Kolmogorov's zero-one law, its probability can only be either 0 or 1. As the possibility that$P(\{\exists \text{ infinite cluster on any point}\})$ is zero was ruled out, only $P(\{\exists \text{ infinite cluster on any point}\}) = 1$ is left.

3.4.2 Nonexistence below the critical threshold

On the other hand, below the critical point the probability that the cluster of the origin is infinite is zero. As the choice of the origin is arbitrary, the infinite cluster can not contain any point and, therefore, it can not exists $P(\{\exists \text{ infinite cluster on any point}\}) = 0$.

3.4.3 How many infinite clusters?

One last thing before talking about the critical threshold. How many infinite clusters are there? An in-depth treatment would be out of scope, but *Grimmett* states:
"*the infinite cluster is (almost surely) unique whenever it exists.*"
It is easy to grasp why this is true, at least in $d = 2$ dimension. Given two infinite clusters, there are only two possible cases:

- The two clusters intersect, thus merging into a single infinite cluster;
- The two clusters are parallel. In this case, they must be separated by a finite number of edges. As their distance is almost constant, the probability that a finite orthogonal path crossing both clusters exists is non-zero. If that path exists, the two clusters are merged into one.

3.5 Critical Point - actual value

We have proven that $\frac{1}{2d-1} \leq p_c < \frac{2}{3}$. But, what about the actual values? The real answer is an open problem. It turns out that the value of the critical probability depends on both the lattice structure and the number of dimensions chosen. Furthermore, it is conjectured that no explicit expression for the critical probability exists for $d \geq 3$ and numerical simulations are the only option available. The best that it is possible to do in a small review like this is to treat the trivial case of $d = 1$ and give a hint on why the critical probability for the square lattice in $d = 2$ must be $p_c = \frac{1}{2}$.

3.5.1 Case d=1

$\mathcal{Z}^1$ is just a linear chain. This implies that if any closed edge exists both on the left and on the right of the origin an infinite cluster can not exist. As the probability that n subsequent edges are closed scales as p^n, the only way to have an infinite cluster is to require $p = 1$. The critical value is thus $p_c = 1$.

3.5.2 Case d=2

To gather some intuition on why the critical probability must be $\frac{1}{2}$ we need to think about the dual lattice. By looking at *Figure* 3 *(right)* it is clear that the dual lattice is just the lattice itself translated by $(\frac{1}{2}, \frac{1}{2})$ (relative units). The coupling between the two graphs $p_{dual} = 1 - p_{primal}$ implies that if one system is above critical probability the other must be below critical probability. Furthermore, we have $p_c^{dual} = 1 - p_c^{primal}$. The only value of p_c that preserve the symmetry of the system is $p_c = \frac{1}{2}$.

3.6 The two phases

Before concluding the mathematical treatment I'd like to report some of the results presented in *Grimmett* [1].

3.6.1 Subcritical phase

In the subcritical phase all clusters have finite length and their probability exponentially decay to zero with the length:

$$\exists\alpha(p) > 0 \quad : \quad P_p(|C| = n) \sim e^{-n\alpha(p)} \qquad \textbf{as } n \to \infty \tag{16}$$

3.6.2 Supercritical phase

In the supercritical phase all the finite-size clusters start merging with the infinite cluster. The longer the cluster, the faster it will merge with the infinite cluster. It is known that there exists two parameters $\beta_1(p), \beta_2(p) : 0 < \beta_2(p) \leq \beta_1(p) < \infty$ such that:

$$e^{-\beta_1(p)} e^{n^{(d-1)/d}} \leq P_p(|C| = n) \leq e^{-\beta_2(p)} e^{n^{(d-1)/d}} \qquad \forall n \tag{17}$$

Furthermore, it is believed that:

$$\delta(p) = \lim_{n\to\infty} \left\{-n^{-(d-1)/d} \log\left(P_p(|C| = n)\right)\right\} \tag{18}$$

exists and is strictly positive when $p > p_c$.

3.6.3 At the critical point

What happens at the critical point is still an open question. Up to now, it was proven that no infinite cluster exists for $d = 2$ and $d \geq 19$, but there are no proofs for intermediate values, although it is conjecture that no infinite cluster exists even there.
Finally, it is believed that the length of the clusters at the critical point decays as

$$P_{p_c}(|C| > n) \sim n^{-1/\delta} \quad \text{as } n \to \infty \tag{19}$$

implying that δ is an example of a critical exponent.

3.6.4 Near the critical point

Near the critical point some power laws are hypothesised[1]:

$$\gamma = -\lim_{p \to p_c^-} \frac{\log \chi(p)}{log|p - p_c|} \tag{20}$$

$$\beta = \lim_{p \to p_c^+} \frac{\log \theta(p)}{log(p - p_c)} \tag{21}$$

$$\delta^{-1} = -\lim_{n \to \infty} \frac{\log P_{p_c}(|C| \geq n)}{log(n)} \tag{22}$$

Here $\chi(p)$ is the expected number of vertices in the cluster of the origin.

$$\chi(p) = E_p|C| \tag{23}$$

$$= \infty \; P_p(|C| = \infty) + \sum_{n=1}^{\infty} nP_n(|C| = n) \tag{24}$$

$$= \infty \; \theta(p) + \sum_{n=1}^{\infty} nP_n(|C| = n) \tag{25}$$

$$= \begin{cases} \sum_{n=1}^{\infty} nP_n(|C| = n) & \text{if } p \leq p_c \\ \infty \quad \text{if } p > p_c \end{cases} \tag{26}$$

Notice that *Grimmett* didn't specify whether δ is related to $\delta(p)$ (*eq.* 18). It shouldn't be the case as *eq.* 18 was defined requiring $p > p_c$ but the two expressions are close enough that a connection might be possible when $\lim p \to p_c^+$.

4 Applications

Generally speaking, Bernoulli percolation is too simple of a model to quantitatively predict physical phenomena due to its total absence of coupling between parts. Despite that, its formulation allows for a qualitative description of numerous phenomena. In this chapter, I'm first going to list some problems where Bernoulli's percolation theory can provide some qualitative insight and then cover the results presented by *Brunk et all* about the collapse of virus capsids [3].

4.1 Disordered Electrical System

Let's assume we are given two materials, a perfect insulator A and a perfect conductor B. What's the proportion of the two materials that allows the mixture to still conduct electricity? If we consider the mixture as made of infinitesimal blocks, the problem can be mapped to the problem of percolation on sites. If the infinitesimal block is made of a conductive material the corresponding site can be taken as open, while if the block is made of insulating material, the site is closed. With this simple model, we can set an upper bound for the fraction of the conductor in the mixture equal to the critical probability[6].

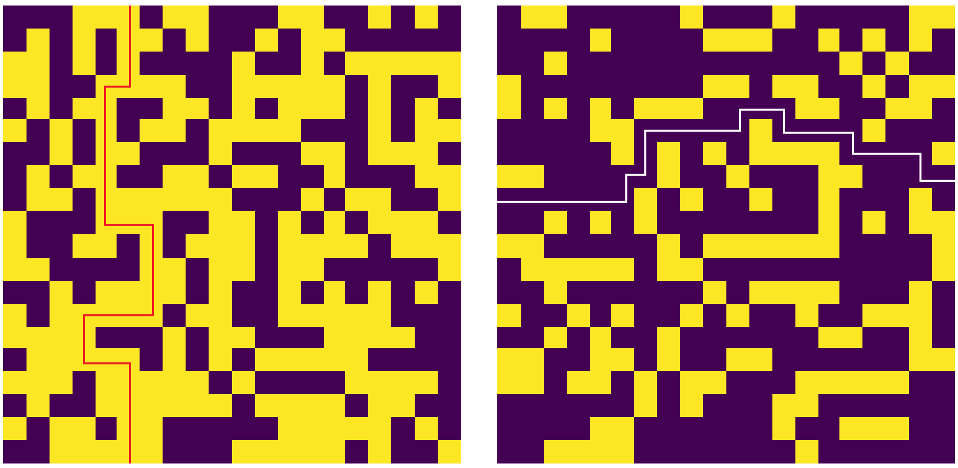

Figure 5: Simulation of a $2D$ mixture of an insulator (dark purple) with a conductor (yellow) on a 20×20 grid. Current is assumed to flow from top to bottom. The simulation on the left is performed at $p_{conductor} = 0.51 > p_c$ and shows conductivity (percolation for the conductor, red path) while the one on the right does not ($p_{conductor} = 0.49 < p_c$) implying that the mixture is an insulator. Notice how the non-percolation of the conductor can be highlighted by looking for the percolation of the insulator in the orthogonal direction (white path).

4.2 Epidemics and fires in orchards

Orchards are usually planned with a lattice structure depending on plant type and harvest easiness arguments. This allows considering the plants as the sites of a $2D$ lattice. If a

[6]If the structure of the mixture is known it is possible to improve the guess by performing appropriate simulations or consulting a list of critical probabilities according to the lattice structure; e.g. https://en.wikipedia.org/wiki/Percolation_threshold

particular plant catches a new disease, this will be transmitted with stochastic probability p to neighbor plants. As above critical probability, the infinite open cluster is unique, if the plants are so closely packed that the infection is transmitted with a probability greater than the critical one, the whole orchards will be blighted.

4.3 Virus capsids collapse

This part will refer to the article *Molecular jenga: the percolation phase transition (collapse) in virus capsids*[3].

Virus capsids are polymeric shells whose function is to protect viral genetic material. Their study is a key component of both antiviral drug design and general drug design. In the first case, the idea is to study capsid assembly/disassembly in order to stop the diffusion of the pathogen from one cell to the other. In the latter case, capsids are taken as biological reference for the design of structures capable of protecting drugs and delivering their content only to specific locations. Brunk et al. tried to apply percolation theory to study the collapse (i.e. percolation) of hepatitis B virus (HBV) capsid.

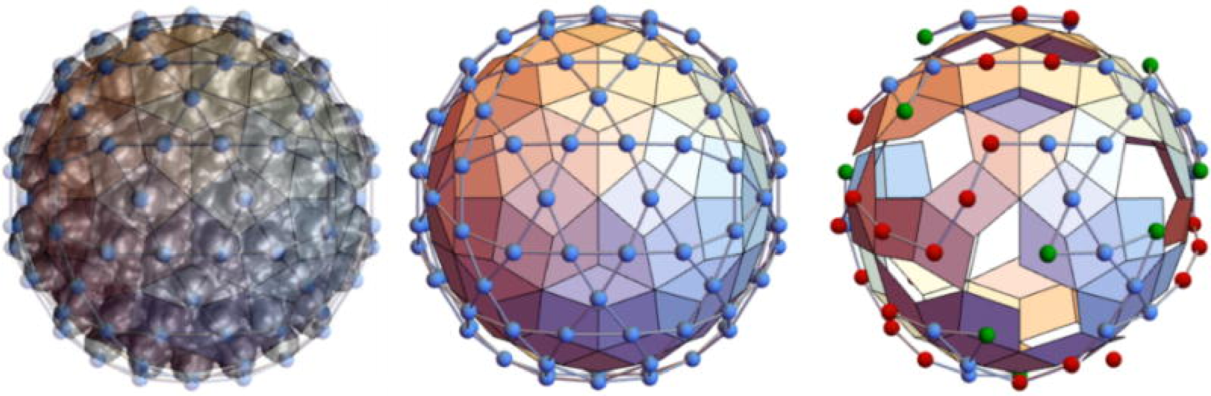

Figure 6: (left) The $T = 4$ HBV capsid overlaid with its dual graph. (center) Polyhedral model of such capsid was obtained assuming that all the sub-units are identical. (right) Polyhedral model after some sub-units were removed. Blue dots in the dual lattice indicate "almost untouched" sub-units, green dots trimeric clusters, and red dots linear chains. Image borrowed from[3].

HBV's capsid is made of 120 equivalent subunits (dimers) arranged with a $T = 4$ icosahedral symmetry (*Figure* 6). Dimers can be removed by gentle treatment with a denaturant and the product of its fragmentation can be studied with single-particle mass spectrometers. A remarkable result of this kind of experiments is that almost no fragmented capsid with less than 90 dimers can be found. This suggests that the removal of $\simeq 31$ sub-units makes the whole structure unstable and prone to fragmentation. To simulate this behavior, the authors assumed all the 120 dimers to be completely equivalent and described the capsid as a wrapped $2D$ regular lattice. Then, they proceeded to randomly remove sub-units (vertices in the dual graph) one at a time and examine fragmentation as a function of the number of removed sub-units.

Assumptions Before presenting their result, it is necessary to clearly state all the assumption made and their consequences:

- The capsid was considered as a $2D$ surface. Although reasonable, effects due to the $3D$ nature of proteins are not to be excluded.
- HBV capsid is made of 120 sub-units while percolation theory assumes an infinite lattice. This might lead to finite-size effects and a more gradual phase transition.
- The vertices of the dual graph represent complex proteins that may change conformation depending on the presence or absence of their neighbors introducing a sort of coupling between vertices.

4.3.1 Results: dimer-dimer interaction energy

First, the authors examined the effects of dimer-dimer interaction energy by studying edge counting as a function of the number of sub-units removed (*Figure* 7).

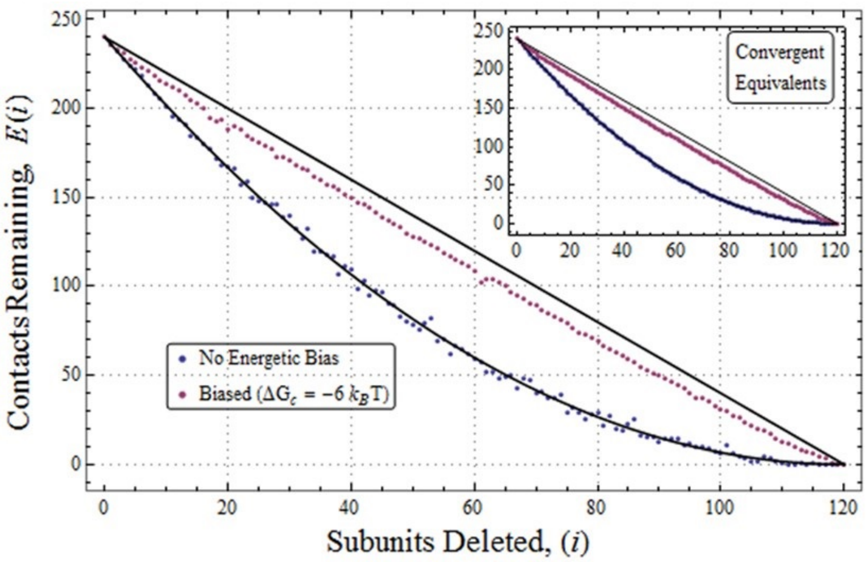

Figure 7: Graph edge count as a function of the sub-units deleted. When accounting for dimer-dimer interaction (purple points) the points have a linear behavior. If dimer-dimer interaction is neglected (blue points) the counting follows a quadratic decay; (Insert) the same plot but with several replicas of the numerical experiment. Image borrowed from[3].

In the *unbiased energy case* where dimer-dimer interaction was neglected, vertices were removed randomly from anywhere in the graph. This led to a quadratic decay which has an explicit formula:

$$E(i) \simeq \bar{E}(i) = E_0 - \bar{k}_o i + \frac{\bar{k}_0}{2V_0} i^2. \tag{27}$$

$E(i)$ is the number of edges left after removing i vertices, $\bar{E}(i)$ its mean after many Monte Carlo replicas, E_0 the initial number of edges, $\bar{k}_0$ the mean number of edges per vertex of the graph (prior to disassembly) and V_0 the initial number of vertices. The linear term takes into account edge removal when totally independent vertices are removed: as each vertex is connected with $\bar{k}_0$ neighbors, its removal causes the loss of $\bar{k}_0$ edges. The quadratic term, instead, handles the case of removing adjacent sub-unit and, thus, removing fewer edges. By recasting the quadratic term as $\frac{\bar{k}_0}{2}\frac{i}{V_0}i$ one can notice how, for a large value of the hole-density $\frac{i}{V_0}$ only half edges ($\frac{\bar{k}_0}{2}$) are taken away for each vertex.

The dimer-dimer interaction was introduced by assuming that the removal can happen only in the proximity of a preexisting defect thus generating a single growing hole. The corresponding edge count follows a linear relation (*Figure* 7), as expected by the $1D$ surface available to the edge removal process.
These results show that the dimer-dimer interaction[7] can not be a key quantity in the fragmentation process, otherwise, they would actually suppress fragmentation. Notice how both curves show a continuous decay without any sign of abrupt transition from an intact capsid to a series of small fragments. Although they highlighted this property, the author did not give any explanation on why this happened. My hypothesis is that this simulation was focused on the dimer-dimer interaction and non-removed vertices were held fixed in the structure, even if all the surrounding vertices were removed.

4.3.2 Percolation Probability

By defining the percolation probability $\bar{P}_{con}(i)$ as the probability that a partially disassembled graph is still connected[8], the authors showed that the expected percolation threshold is identical to the one measured in experiments, namely, the critical value for HBV occurs at $i = 31$ missing sub-units.

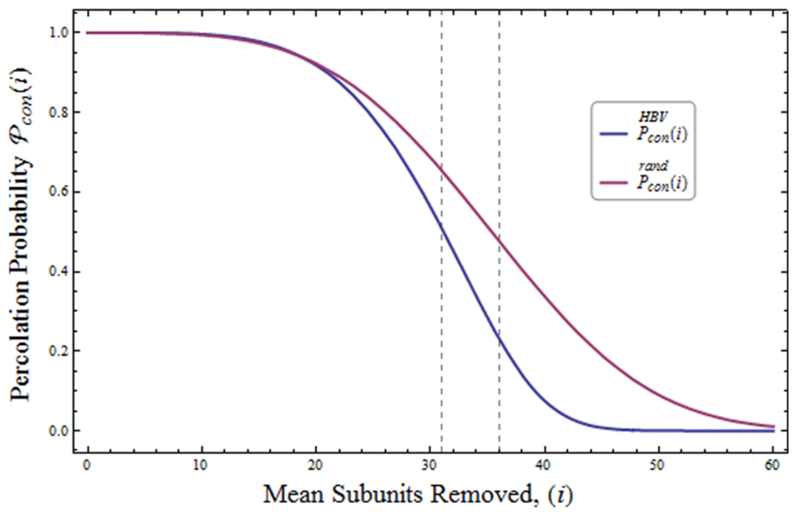

Figure 8: Percolation probability as a function of the number of sub-units removed. The percolation threshold (dashed vertical bar) is obtained by looking at the maximum of the derivative. P_{con}^{HBV} refers to the HBV virus while P_{con}^{rand} to simulation performed on multiple (unspecified) non-planar random 4-regular graphs. Image borrowed from[3].

[7] At least an attractive dimer-dimer interaction. Recall that sub-units were taken away from an already existing defect.

[8] In this particular setting *connected* means that no patch of vertices has detached from the capsid.

5 Conclusions

5.1 About the model

The most attentive reader will have noticed a lot of similarities between Bernoulli percolation and the Ising Model. This is a coincidence. As *Duminil-Copin* [4] reports, both Bernoulli percolation and the Ising model are particular cases of the **random cluster model**. The reader interested in a bird's-eye view review can refer to the article by *Duminil-Copin* while a reader who wants to have an in-depth look into the topic may refer to *The Random-Cluster Model*[5] by *Grimmett*. I haven't read that article, but basing myself on his approach in [1] I think the treatment will be enlightening, even though it may be a bit too mathematical in some points.

5.2 About applications

Due to its lack of coupling between elements, Bernoulli percolation seems too simple of a model to obtain quantitative esteems. Any attempt to do so like [3] appears as forced and unnatural. Despite that, Bernoulli percolation can explain the presence of a phase transition even when all the elements constituting a system are independent. This is a marvelous result that can give qualitative insight into some problems and help develop intuition on more complicated phenomena.

References

1. Grimmett, G. in *Percolation* (ed Grimmett, G.) 1–31 (Springer, Berlin, Heidelberg, 1999). ISBN: 978-3-662-03981-6. `https://doi.org/10.1007/978-3-662-03981-6_1` (2023).
2. Spectral Collective. *Percolation: a Mathematical Phase Transition* `https://www.youtube.com/watch?v=a-767WnbaCQ` (2023).
3. Brunk, N. E., Lee, L. S., Glazier, J. A., Butske, W. & Zlotnick, A. Molecular Jenga: The Percolation Phase Transition (Collapse) in Virus Capsids. *Physical biology* **15,** 056005. ISSN: 1478-3967. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6004236/` (2023) (June 6, 2018).
4. Duminil-Copin, H. *Sixty years of percolation* Dec. 13, 2017. arXiv: `1712.04651[math-ph]`. `http://arxiv.org/abs/1712.04651` (2023).
5. Grimmett, G. *The Random-Cluster Model* 2002. `https://arxiv.org/abs/math/0205237`.