## 10. Project: Percolation

10.1. **Choice of the Random Number Generator.** The permuted congruential generator (PCG) is a modern, simple, high quality pseudorandom number generator (RNG) made by O'Neill in 2014. It is essentially just a linear congruential generator (LCG), except the output of the LCG is not used directly, but instead is used to rotate some bits.[8][O'N14] Surprisingly, if done carefully, this simple change turns the unusable LCG into an excellent generator.

All code in this essay has used the PCG generator, since it is the default for NumPy, and because it is what we used in C++ also. But we used the XSH-RR variant in C++ because it's the most standard version, whereas NumPy uses the XSL-RR variant.

Testing a RNG means testing for correlations. This is highly nontrivial, there are many types of correlation failures a RNG may have. It's best to use standardized test suites like NIST SP 800-22. Others have already tested the XSH-RR variant of the PCG generator in this test suite with good results, see e.g. Cook's results.[9] O'Neill's paper also includes such tests.[O'N14]

10.2. **Percolation Theory and Bond Percolation.** Percolation theory studies links in a network, and how disconnected clusters of links become infinite connected 'spanning clusters' after a phase transition, which is marked by a finite probability of obtaining such a spanning cluster.

The example we will consider is 'bond percolation' in $d = 2$ dimensions. Here we consider a lattice of bonds (links), i.e. a lattice graph. We have some parameter $p$, which is the (i.i.d.) probability any particular connection becomes an open bond (traversible). We are interested in the the connectedness of the graph, e.g. the sizes of the clusters, and whether there is a finite probability to obtain a spanning cluster, the existence of which means the lattice has 'percolation'.

Intuitively, as $p$ increases, the subgraphs become more connected, giving rise to fewer overall connected subgraphs. There exists a critical probability $p_c$ at which the probability for *a* cluster subgraph to be connected to $\infty$ becomes 1. However, the lattice is mathematically infinite. So the probability this subgraph is visible in the finite part of the lattice that we can sample is $\leq 1$. For large $p$, a single cluster will dominate, and then we can be more confident that the dominant subgraph is the one connected to $\infty$, i.e. the spanning cluster resulting in percolation.

Clearly $p_c$ is analogous to $T_c^{-1}$ in the Ising model; this single cluster dominating and reaching $\infty$ can be seen as a ferromagnetic phase. But whereas Ising model clusters have only 2 possible identities: spin up and spin down, there are infinitely many bond percolation cluster identities.

$p_c$ was solved analytically (in $d = 2$ dimensions) in two steps, by Harris in 1960 and Kesten in 1980. Harris showed that subgraphs stay disconnected from $\infty$ for $p = \frac{1}{2}$, i.e. $p_c \geq \frac{1}{2}$. Kesten showed that at least one connects for $p > \frac{1}{2}$, i.e. $p_c \leq \frac{1}{2}$.[10] (By definition, $p_c$ itself has no chance of percolation, but $p > p_c$ does.) Unfortunately Kesten's proof is difficult to find, and there are more elegant ways to do his proof. So better to see [BR04]

[8] `https://www.pcg-random.org/index.html`

[9] `https://www.johndcook.com/blog/2017/07/07/testing-the-pcg-random-number-generator/`, `https://www.johndcook.com/PCG_NIST_output.txt`.

[10] It has also been proven that this one is unique, so it is also correct to say exactly one connects for $p > p_c$, but this was only shown later.

than the original works, it fully proves the $p_c = \frac{1}{2}$ result. In $d = 3$, just like the Ising model, the problem is unsolved.

10.3. **Cluster Visualization.** We can visualize the clusters by decomposing the lattice graph into a set of connected subgraphs, and assigning a random color to each cluster. We will show the nonexistent bonds as black. Figures 29 to 32 show some visualizations for $p = 0.45$, $p = 0.5$, $p = 0.55$, and $p = 0.6$. For $p = 0.45$ (subcritical), the clusters stay small. For $p = 0.5$ (critical), the clusters vary more in size, but it's not particularly dramatic. For $p = 0.55$ we see large clusters. At this point the probability for any particular bond to be connected to $\infty$ is finite, but still clearly small. For $p = 0.6$ we have a large cluster, so that there is a significant probability that this large connected subgraph extends to $\infty$, and that it is this unique subgraph that is percolating. (Of course the simulation is finite, but mathematically we deal with infinite lattices. We are seeing only a small, arbitrary part of the infinite lattice.)

10.4. **Proof that $p_c \in \left[\frac{1}{3}, \frac{2}{3}\right]$.** It is interesting to prove (to a mathematical physics level of rigor) that $p_c$ is nontrivial, i.e. $p_c \in (0, 1)$. We can do better even and show $p_c \in \left[\frac{1}{3}, \frac{2}{3}\right]$ quite easily. We're not sure who to credit for this proof, but the $p_c < 1$ part is due to Peierls. We will follow mostly the notes from the Math 541 class at the University of Arizona by Kennedy, and also a review by Duminil-Copin.[11][DC17]

Instead of searching for this one particular subgraph connected to $\infty$, we will consider the subgraph $C$ containing a particular arbitrary lattice site, called the 'origin' $o$. The point is that, because the lattice is infinite, and because the origin is arbitrary, proving that the probability $\theta(p)$ for $C$ to be connected to $\infty$ is finite, is sufficient to prove that there must exist a subgraph connected to $\infty$, i.e. we have percolation through the lattice graph. So that is what we will do instead: show where $\theta(p) = 0$ and where $\theta(p) > 0$.

Let's be more explicit and write $\theta_\infty(p) \equiv \theta(p)$ and $\theta_\ell(p)$ for the probability for $o$ to be connected to a Manhattan distance $\ell$. Consider the set of all possible paths of distance $\ell$ from the origin. For the 1st step we can choose 4 directions. For the $\ell - 1$ remaining steps we may only choose $\leq 3$. (It is $\leq 3$ and not 3 because loops can't exist, since then the path no longer has length $\ell$, as there would be a shortcut.) An upper limit for the number of these paths is $3^{\ell-1}4$. And each obviously has a probability $p^\ell$ of actually existing; each step must be a bond. So $\theta_\ell(p) \leq 3^{\ell-1}4p^\ell = 4p(3p)^{\ell-1}$ for any finite $\ell$. Now if $p < \frac{1}{3}$, this goes to 0 as $\ell \to \infty$, so there is no chance of $C$ being connected to $\infty$, and thus $p_c \geq \frac{1}{3}$.

Each lattice graph has an associated dual. In this context they're called the 'primal' and 'dual' grids respectively. This dual grid is defined so that its vertices lie at the centers of the primal grid's vertices, and so that a bond in the primal grid at some location becomes closed in the dual grid, and vice versa. A physical interpretation is that if the primal grid represents a maze (which is by definition the *paths*), the dual grid represents the *walls*. Anyway, this is an interesting construction because the parameter of the dual grid is $1 - p$. By understanding the relation between the primal grid and dual grid a bit better, we can make a statement about $p = 1 - \frac{1}{3} = \frac{2}{3}$.

Consider $C$. Let's say it is not connected to $\infty$, then this is equivalent to the statement that the dual grid has a sector that fully encloses $C$, i.e. there is no way out of the maze. Write $C_\ell(p)$ for the probability that the 'circuit' enclosing $C$ in the dual grid is of length $\ell$. Now draw a line of length $\frac{\ell}{2}$ that is part of $C$. Then for the circuit to enclose it, it must

[11] `https://www.math.arizona.edu/~tgk/541/`

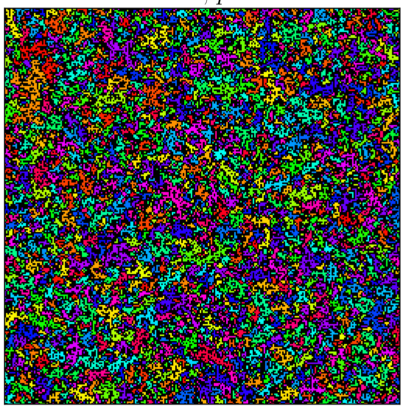


FIGURE 29. Subcritical bond percolation.

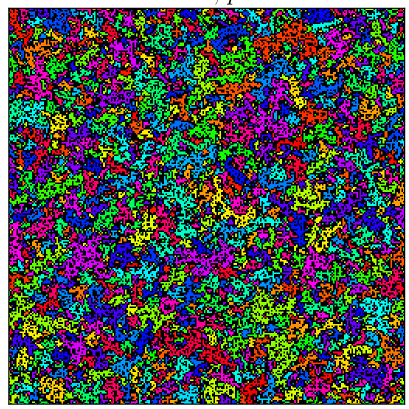


FIGURE 30. Critical bond percolation.

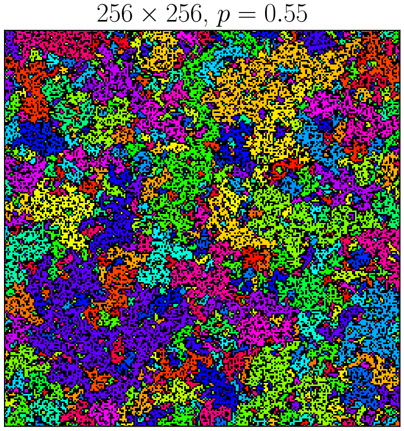


FIGURE 31. Slightly supercritical bond percolation.

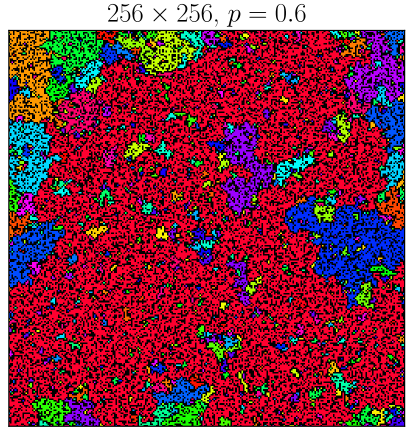


FIGURE 32. Significantly supercritical bond percolation.

have a length $\ell + 4 > \ell$, which is not possible since its length is defined as $\ell$. So the circuit must cross this line of length $\frac{\ell}{2}$, i.e. we can start constructing the circuit on this line. There are $\frac{\ell}{2}$ choices for where to start, and at most $3^{\ell-1}$ other choices, for the same reason as before. A particular circuit of length $\ell$ in the dual grid has a probability $(1-p)^\ell$ to exist. So the probability for the circuit enclosing $C$ (that is not connected to $\infty$) to be of length $\ell$ is $C_\ell(p) \leq (1-p)^\ell \frac{\ell}{2} 3^{\ell-1} \leq (1-p)^\ell \ell 3^\ell = \ell(3(1-p))^\ell$. Now the probability for any such circuit to exist that can 'contain' $C$ is at most $\sum_\ell \ell(3(1-p))^\ell$. The probability for such a circuit to be at least of length $L$ is $\sum_{\ell=L} \ell(3(1-p))^\ell$. $C$ can be arbitrarily large because the lattice is infinite. If we have $p > \frac{2}{3}$, $\sum_{\ell=L} \ell(3(1-p))^\ell \to 0$ as $L \to \infty$. But we can't make $C$ infinite. It is sufficient however to notice that there will be some finite $L$ so that $\sum_{\ell=L} \ell(3(1-p))^\ell < 1$. This means the probability that $C$ cannot be contained becomes $1 - \sum_{\ell=L} \ell(3(1-p))^\ell > 0$ for some finite size of $C$. $C$ may always have some large finite size since the lattice is infinite and $o$ arbitrary. So there is percolation for $p > \frac{2}{3}$, and thus $p_c \le \frac{2}{3}$.

10.5. **Intuition for the $p_c = \frac{1}{2}$ Proof.** In general, the primal and dual grids are opposite. If the clusters in one are small, the clusters in the other are large. If we had $p_c < \frac{1}{2}$ or $p_c > \frac{1}{2}$, it would be possible for the primal and dual grids to simultaneously be subcritical or supercritical. If both were subcritical, both would have probability 0 for any cluster to be connected to $\infty$, which makes little sense. Similarly both being supercritical makes little sense. This outlines how Kesten showed that $p_c = \frac{1}{2}$. (But this does not extend to $d = 3$, there $p_c \approx 0.249$.)

10.6. **Subcritical and Supercritical Behavior.** Some more specific behavior is known for $d = 2$ bond percolation.

- In the subcritical $p < p_c$ phase, $\theta_\ell(p)$ decreases exponentially in $\ell$. (This is the specific way in which it goes to 0 at $\infty$.)[DC17] Instead of the probability of being connected up to some Manhattan distance $\ell$, we can also consider the probability of the size of $C$ being $|C|$, it is also exponentially decreasing. This gives the cluster size distribution $S_{|C|}(p) \sim e^{-c|C|}$.
- $S_{|C|}(p_c) \sim |C|^{-\tau}$ as $|C| \to \infty$, where $\tau$ is called the Fisher exponent, which should be $\tau = \frac{187}{91} \approx 2.05$.
- In the supercritical $p > p_c$ phase near $p_c$, $\theta(p) \sim (p - p_c)^\beta$.[12]
- etc.

An efficient way to test the aformentioned subcritical behavior is to sample a single huge lattice, because $o$ and thus $|C|$ are arbitrary; it should be equivalent to sampling many smaller ones where we only consider the origin $o$. Figure 33 shows clearly that this is indeed exponentially decreasing at a constant rate. The very small sizes are slight outliers, which we believe is due to the clusters no longer being able to shrink due to the discretization, and the very large sizes are also outliers due to very few clusters existing there, so low S/N ratio. But overall it still looks linear, even for the largest of sizes, i.e. constant spatial decay rate. We see in Figure 34 that this spatial decay rate does not depend on the lattice size, as expected; the distribution is essentially identical.[13]

We can't fit the Fisher exponent very well because of this $|C| \to \infty$ requirement. We can't generate enough samples for large $|C|$ to get a good fit. We don't know how large $|C|$ should be to begin with, but clearly more than we practically can do here since we consistently find $\tau > 2.05$. Anyway, Figure 35 shows some results along with a fit. Here we find $\tau \approx 2.85$, which doesn't seem too bad. At least we can see a clear deviation from the subcritical exponential decay for large $|C|$. This reflects that larger clusters are starting to become more likely, and at some point so much so that a spanning cluster forms, resulting in percolation.

[12] `https://www.math.arizona.edu/~tgk/541/`

[13] Usually $S_{|C|}$ is normalized by the lattice size, which we didn't do here, but that should not affect the power $c$.

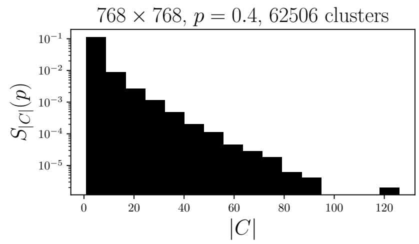


FIGURE 33. Subcritical exponentially decreasing $S_{|C|}(p)$.

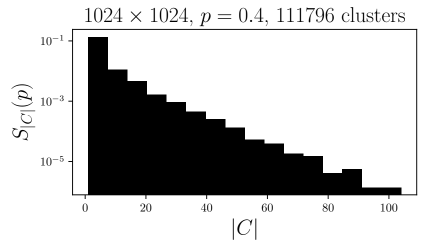


FIGURE 34. Subcritical exponentially decreasing $S_{|C|}(p)$.

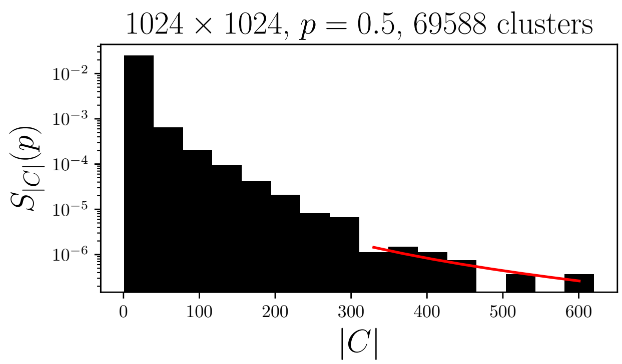


FIGURE 35. Critical power law behavior, where we estimate the Fisher exponent to be $\tau \approx 2.85$.

The supercritical behavior we test by generating many lattices and checking if $o$ is connected to $\infty$ (i.e. whether there is percolation), this estimates $\theta(p)$ directly. We must make many samples at each $p$ to estimate $\theta(p)$. Our main interest is a narrow range of $p > p_c$ values, but it's also interesting to consider a wider range because it should display the phase transition at $p_c = \frac{1}{2}$, which we haven't numerically estimated yet. Figure 36 shows the overall behavior of the parameter $p \in [0.2, 0.8]$, where we can for the 1st time nicely see that $p = \frac{1}{2}$ does look like a phase transition, where the probability for percolation becomes finite. Here we consider 20 discrete values of $p$, and take $n = 200$ samples to estimate $\theta(p)$. Figure 37 shows the behavior near $p_c$, so $p \in [0.5, 0.55]$. We consider 10 discrete values of $p$ and take $n = 1000$, so that each point should be quite accurate. We sample here a power law, i.e. $\theta(p) \sim (p - p_c)^\beta$. We find $\beta \approx 4.8$. Figure 38 shows a fit in $p \in [0.53, 0.58]$, which should have better S/N due to larger $\theta(p)$, but it might no longer be sufficiently near $p = p_c$ for the same power law to be valid. Here we find $\beta \approx 4.1$. Either way, it's clear that the power law fit is excellent, with the main limitation being the low S/N ratio due to low $\theta(p)$.

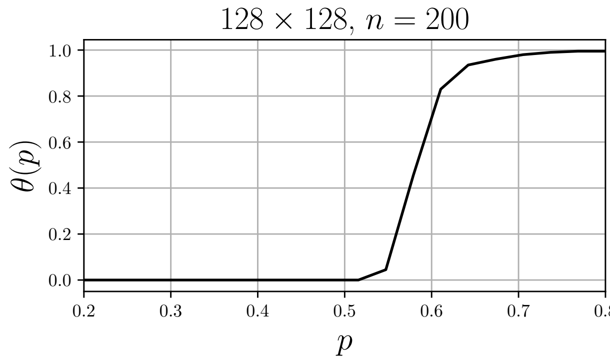


FIGURE 36. Overall percolation probability $\theta(p)$. $n = 200$ samples for 20 discrete values of $p$.

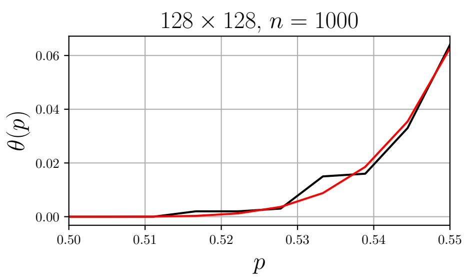


FIGURE 37. Supercritical power law behavior of $\theta(p)$ near $p_c$. $n = 1000$ samples for 10 discrete values of $p$. Fits $\beta \approx 4.8$

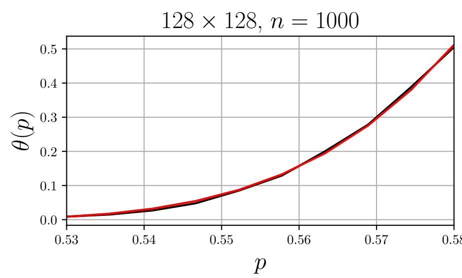


FIGURE 38. Supercritical power law behavior of $\theta(p)$ near $p_c$. $n = 1000$ samples for 10 discrete values of $p$. Fits $\beta \approx 4.1$.

## References

[BR04] B. Bollobas and O. Riordan. "A short proof of the Harris-Kesten Theorem". In: (2004). URL: https://arxiv.org/abs/math/0410359.

[DC17] Hugo Duminil-Copin. Sixty years of percolation. 2017. arXiv: 1712.04651 [math.PR].

[NB99] M. E. J. Newman and G. T. Barkema. Monte Carlo methods in statistical physics. Oxford: Clarendon Press, 1999.

[O'N14] M.E. O'Neill. "PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms for Random Number Generation". In: (2014).

KU Leuven
*Email address*: olaf.willocx@student.kuleuven.be