Snapshots of modern mathematics
from Oberwolfach

№ 6/2019

Counting self-avoiding walks on the hexagonal lattice

Hugo Duminil-Copin[1]

In how many ways can you go for a walk along a lattice grid in such a way that you never meet your own trail? In this snapshot, we describe some combinatorial and statistical aspects of these so-called *self-avoiding walks*. In particular, we discuss a recent result concerning the number of self-avoiding walks on the hexagonal ("honeycomb") lattice. In the last part, we briefly hint at the connection to the geometry of long random self-avoiding walks.

1 Self-avoiding walks on a lattice

Around the middle of the 20th century, Paul J. Flory[2] and W. J. C. Orr introduced *self-avoiding walks* (SAWs) as a mathematical model for the shape of ideal polymers, that are, very long chain-like molecules composed from many identical small links called monomers [5, 11].

Let us consider a lattice such as the square lattice, the triangular lattice, or the hexagonal lattice, which you can see in Figure 1. A self-avoiding walk on the lattice is a sequence of neighboring vertices that avoids returning to any previously visited vertex. The first question that pops to mind is the following:

How many different SAWs of length n start at a given point?

[1] Hugo Duminil-Copin is supported by the NCCR SwissMap and an IDEX chair of Paris Saclay.

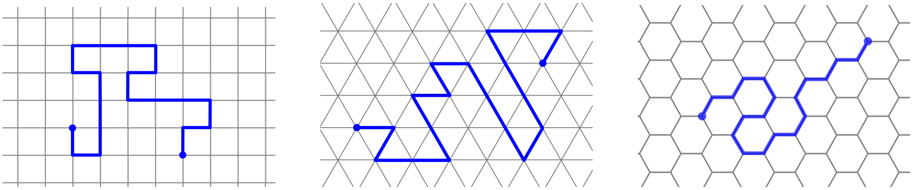

Figure 1: Examples of self-avoiding walks of length 19 on the square, triangular, and hexagonal lattices. The unit of length is chosen to be the length of an edge in the lattice.

In order to address this question, we will first have a more formal look at the definitions. A *graph* is a set of *vertices* together with a set of *edges*, where an edge links pairs of vertices: when two vertices are connected by an edge, they are said to be *neighbors* of one another.[3] A graph is *transitive* if every vertex can be mapped to every other vertex by a "symmetry" of the graph, which means the graph looks the same from every vertex. In the polymer interpretation, vertices correspond to monomers, which are linked via edges of the graph.

A graph is called *infinite* if it has infinitely many vertices, and *locally-finite* if every vertex has only finitely many neighbors. We define a *lattice* to be a transitive locally-finite infinite graph. In addition to the already encountered square, triangular, and hexagonal lattices, which are all lying in the plane $\mathbb{R}^2$, a further example of a lattice is the *cubic lattice* in $\mathbb{R}^3$, which has vertex set $\mathbb{Z}^3$, and two vertices are neighbors whenever their distance is 1.

Given a graph $\mathbb{G}$, a *walk* of length $n \in \mathbb{N}$ on $\mathbb{G}$ is a map $\gamma : \{0, \ldots, n\} \to \mathbb{G}$ such that $\gamma(i)$ and $\gamma(i+1)$ are neighbors for each $i \in \{0, \ldots, n-1\}$. A walk is called *self-avoiding* if it is injective, that is, $\gamma(i) \neq \gamma(j)$ for $i \neq j$. From now on, we will only consider walks on graphs that are lattices.

Returning to our previous question of counting self-avoiding walks, choose a lattice $\mathbb{L}$ and let c_n denote the number of self-avoiding walks on $\mathbb{L}$ starting from a fixed point. Since the lattice is transitive, this number does not depend on the choice of that point. For small values of n, the number c_n can be computed by hand and it is fun. For instance, one can find by hand that $c_6 = 16\,926$ on $\mathbb{L} = \mathbb{Z}^3$. However, as n grows, such computations quickly become impossible to perform. This is due to the fact that, as we will see below, c_n grows exponentially fast. With today's technology and efficient algorithms, one may count walks

[2] Paul John Flory (1910–1985) eventually received the Nobel Prize in Chemistry in 1974, giving a Nobel Lecture on *Spatial Configuration of Macromolecular Chains*.

[3] Note to the expert reader: we only consider *simple* graphs, that is, graphs without loops or double edges.

up to length 36 on $\mathbb{Z}^3$, see [12], where a new algorithm is used together with 50 000 hours of computing time to get $c_{36} = 2\,941\,370\,856\,334\,701\,726\,560\,670$. On the square lattice $\mathbb{Z}^2$, the largest known enumeration corresponds to walks of length 71, see [2]. No exact formula is expected to hold for general values of n, yet it is still possible to study the *asymptotic*, or limiting, behavior of c_n as n becomes very large. John Hammersley observed in [6] that the sequence c_n has the property that there is a certain positive real number $\mu_c(\mathbb{L})$, called the *connective constant* of the lattice, with

$$\lim_{n\to\infty} c_n^{1/n} = \mu_c(\mathbb{L}). \tag{1}$$

Thus c_n is "roughly" $\mu_c(\mathbb{L})^n$ for large n. Hammersley's elegant argument has become classical in statistical physics and probability. It runs as follows.

Since a SAW of $n+m$ steps can be cut into an n-step SAW and a parallel translation of an m-step SAW, we infer that

$$c_{n+m} \leq c_n c_m,$$

hence the $(c_n)_{n\in\mathbb{N}}$ form what is called a *sub-multiplicative* sequence. From here, it follows by an exercise on such sequences that there indeed is a constant $\mu_c(\mathbb{L})$ such that Equation (1) holds. Note that $\mu_c(\mathbb{L})$ is larger than or equal to 1 and smaller than or equal to the number of neighbors of a point[4] minus 1.

As an example, consider the *tree of degree* $d+1$. This is the lattice $\mathbb{T}_d$ that is uniquely defined by the following two properties: every vertex has degree $d+1$, and for any two vertices, there is exactly one self-avoiding walk starting at one vertex and ending at the other. Figure 2 shows an illustration of a part of the tree $\mathbb{T}_2$.[5] One can directly check that $c_n(\mathbb{T}_d) = d^n$, and hence $\mu_c(\mathbb{T}_d) = d$.

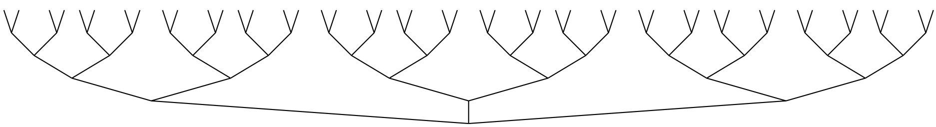

Figure 2: A finite part of $\mathbb{T}_2$, the tree of degree 3. Unlike the image might suggest, we consider all edges, and thus all the steps in a walk, to have the same length.

[4] Note that this number, called the *degree* of the lattice, does not depend on the point. For instance, the hexagonal, square, triangular, and cubic lattices have degrees 3, 4, 6, and 8, respectively.

[5] Note that this is only a combinatorial image – we consider all edges to have the same length.

Unfortunately, explicit formulæ for $\mu_c(\mathbb{L})$ are not expected to be frequent, and mathematicians and physicists only possess numerical[6] predictions for the most common lattices. For instance, the papers [3, 2] give the following approximations for the connective constants of $\mathbb{Z}^2$ and $\mathbb{Z}^3$ – the parentheses indicate a possible error in these digits:

$$\mu_c(\mathbb{Z}^2) = 2.63815853035(2),$$
$$\mu_c(\mathbb{Z}^3) = 4.684039931(27).$$

2 The connective constant of the hexagonal lattice

In 1980, Bernard Nienhuis [10] suggested that the hexagonal lattice $\mathbb{H}$ is an exception among lattices, in the sense that its connective constant can be very explicitly stated. More precisely, Nienhuis conjectured that

$$\mu_c(\mathbb{H}) = \sqrt{2+\sqrt{2}}.$$

This beautiful prediction relied on a correspondence between different models of statistical physics and remained mathematically elusive.

Recently, Stanislas Smirnov and I provided a rigorous proof of this result [4]. The argument is quite instructive. While we cannot describe it in such a short number of pages, let us still highlight one important aspect of the proof, and refer the avid reader to the original paper.

From now on, we view the hexagonal lattice $\mathbb{H}$ as lying in the complex plane $\mathbb{C}$. Consider a finite *domain* $\mathcal{D}$, that is, a finite subset of edges of $\mathbb{H}$, and pick a mid-edge a on the boundary of $\mathcal{D}$, as in Figure 3. For every mid-edge z in $\mathcal{D}$, consider the self-avoiding walks in $\mathcal{D}$ that start from a and end at z. For such a walk γ, let $W_\gamma(a,z)$ be equal to the number of left turns minus the number of right turns made by γ when going from a to z, times $\frac{\pi}{3}$. Introduce a complex-valued function F defined on every mid-edge z, and depending on parameters σ and x, by the summation formula[7]

$$F(z) := \sum_{\substack{\gamma \text{ SAW in } \mathcal{D} \\ \text{from } a \text{ to } z}} e^{-\mathrm{i}\sigma W_\gamma(a,z)} x^{\#\text{ vertices in }\gamma}. \tag{2}$$

In our example in Figure 3, the blue walk from a to z passes through 24 vertices, making 9 left and 15 right turns, and thus contributes a term $e^{2\pi\mathrm{i}\sigma}x^{24}$ to the sum in Formula (2).

---

[6] that is, approximatively computed

[7] the symbol # stands for 'number of'

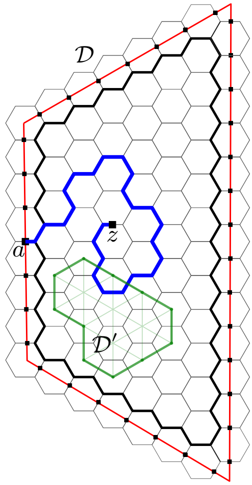

Figure 3: A domain $\mathcal{D}$ in the hexagonal lattice, with its boundary contour (in red) and little boxes at the boundary mid-edges. Blue: a self-avoiding walk in $\mathcal{D}$ from a boundary mid-edge a to an interior mid-edge z. Green: a discrete contour bounding a subdomain $\mathcal{D}'$ of $\mathcal{D}$, together with its decomposition into elementary triangular contours.

The advantage of the function F, which is called the *parafermionic observable*, is that when $\sigma = 5/8$ and $x = 1/\sqrt{2+\sqrt{2}}$, it enjoys a very special property: around every vertex v of $\mathcal{D}$, the values of F at the adjacent mid-edges p, q, and r, indexed in counter-clockwise order, satisfy the relation

$$F(p) \;+\; \mathrm{e}^{2\mathrm{i}\pi/3}\, F(q) \;+\; \mathrm{e}^{4\mathrm{i}\pi/3}\, F(r) \;\;=\;\; 0. \tag{3}$$

The coefficients in Relation (3) are such that the left-hand side can be seen as a sum along an “elementary contour on the dual lattice”, up to a multiplicative factor. While this sentence may appear as an adjunction of strange words, the interpretation is quite simple:

A *contour* in $\mathbb{H}$ is a path $c = (z_i)_{i \le n}$ of neighboring faces of $\mathbb{H}$, where z_i is the complex number at the center of the corresponding face. The contour is *closed* when c starts and ends on the same face, that is, z_n and z_0 are equal. For an example, see the green paths in Figure 3 and Figure 4. Define the *discrete integral* of a function F on mid-edges along such a contour c by

$$\oint_c F(z)\, dz := \sum_{i=0}^{n-1} F(p_i)(z_{i+1} - z_i), \tag{4}$$

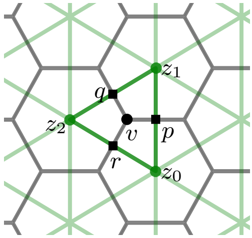

Figure 4: Zooming in on a vertex v of the hexagonal lattice, we see the adjacent mid-edges p, q, and r, and the centers z_0, z_1, and z_2 of the adjacent faces forming the corresponding elementary contour. All elementary contours together form the triangular "dual" lattice.

where p_i is the center of the edge bordered by the faces corresponding to z_i and z_{i+1}. Note that $-\frac{\mathrm{i}}{\sqrt{3}}(z_{i+1} - z_i) \in \{1, \mathrm{e}^{2\mathrm{i}\pi/3}, \mathrm{e}^{4\mathrm{i}\pi/3}\}$ for every $0 \leq i < n$.

Equation (3) at vertex v states that the discrete integral of F along the discrete "triangular" contour going through the three faces around v vanishes, see Figure 4. Now, assume furthermore that the domain $\mathcal{D}$ has no "hole". Then one may decompose any closed discrete contour in $\mathcal{D}$ into a sum of elementary triangular contours. Since the discrete integral of the observable F along each of these elementary contours is zero, we deduce that the discrete integral of F vanishes for any discrete contour in $\mathcal{D}$.

This property seems to be a discretization of a property characterizing *holomorphic functions*, which are complex-valued functions of a complex variable that are "complex differentiable": the contour integrals of a holomorphic function vanish, as long as the contour does not go around a "hole" of the domain.[8] Thus, one may interpret Relation (3) as a discrete version of holomorphicity. In our context, *discrete*, as opposed to "continuous", describes the concept that the variable "jumps" from mid-edge to mid-edge of the lattice $\mathbb{H}$, instead of moving continuously in the complex plane $\mathbb{C}$.

Nevertheless, a word of caution is necessary. One of the key properties of holomorphic maps is that so-called *boundary value problems* have unique solutions – when two holomorphic functions have the same values on the boundary of a domain, they must agree on its interior, too. Let us try to do the same at the discrete level. Imagine for a moment that we wish to determine a discrete function F using only its boundary values and Relation (3) around every vertex. We have one unknown variable per mid-edge z, namely the value

[8] Note to the more advanced reader: this is known as *Cauchy's integral theorem*. Taken together with *Morera's theorem*, which states the converse, it asserts that this property can actually be taken as definition of holomorphicity.

of $F(z)$, and one relation per vertex. For example, a discrete boundary value problem for the green domain $\mathcal{D}'$ in Figure 3 would have 11 boundary values, 20 unknowns, and 17 relations. Note that

$$3\,\#\,\text{interior vertices} = \#\,\text{boundary mid-edges} + 2\,\#\,\text{interior mid-edges},$$

because every vertex is adjacent to 3 edges, and every mid-edge is adjacent to 1 or 2 interior vertices, depending on whether it is on the boundary or on the interior of the domain. Therefore, there are typically more unknown variables than relations, and therefore many solutions to this system of linear equations.

We are seemingly facing a dead end: in the discrete case, the fact that the discrete contour sums vanish seems to provide little information about the function F. In conclusion, a function satisfying Relation (3) around every vertex can be seen as some kind of "weakly discrete holomorphic" function, but the relations do not allow us to do as much as the standard notion of holomorphicity does.

Fortunately, the property of vanishing contour sums is not meaningless. A careful analysis of contour sums along the boundary of well chosen domains $\mathcal{D}$ implies that the value $\sqrt{2+\sqrt{2}}$ mentioned above has to be the connective constant of the hexagonal lattice. This part is not straightforward, and may even appear as a miracle in the light of the previous paragraph. Again, we refer to the original paper [4] for further details about this strategy.

3 Geometry of the SAW on the hexagonal lattice

Computing the connective constant should be considered as a stepping stone towards a bigger goal: physicists and mathematicians are ultimately interested in the geometry of large *random* SAWs. Let us thus depart from our combinatorial question of counting SAWs and focus on a geometrical one:

How does a large SAW typically look like?

Let us start by mentioning a related stochastic model, called the simple random walk (SRW), which is obtained as follows: Consider all – possibly self-intersecting – walks on the hexagonal lattice that start at the origin and take n steps from there. Combinatorially, there are exactly 3^n such walks. The *n-step simple random walk* is the uniformly random choice of a sample from the set of n-step walks starting at the origin.[9] The probability of picking any particular walk is 3^{-n}. Further, we allow the *meshsize* δ of the lattice, that is, the length of an edge – and hence of a step in the walk – to vary as n increases.

[9] A method of taking a random sample is called *uniform* if all possible choices are equally likely.

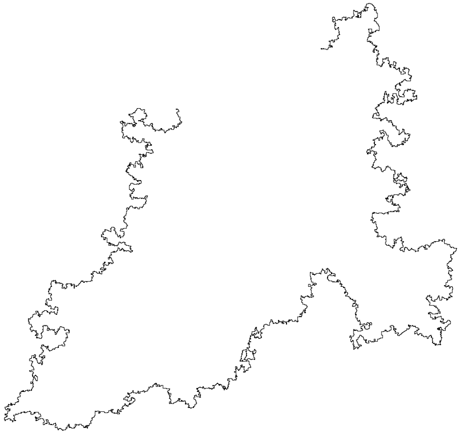

Figure 5: A 10 000-step self-avoiding walk on a lattice $\delta\mathbb{Z}^2$. The distance between start and end point is about $273\,\delta$.

Thus, let Φ_n denote the n-step simple random walk on a rescaled hexagonal lattice $\delta_n\mathbb{H}$. We are interested in the asymptotic behavior of Φ_n as n grows large.

If $\delta_n = 1$ for all n, we effectively do not rescale the lattice and therefore obtain larger and larger walks. On the contrary, if δ_n decays too fast, the walks collapse to the origin as n tends to infinity. Now, if δ_n is taken to be $n^{-1/2}$, then as n tends to infinity, the walk Φ_n converges, as a random object, to a continuous random curve Φ called the *Brownian Motion*, see [9].[10] As a corollary, we deduce that asymptotically, a SRW of length n is typically of "size" $\sqrt{n}$.[11]

Physicists have studied the corresponding question for the SAW.[12] Let Γ_n denote the random choice, with uniform probability c_n^{-1}, of a SAW among all n-step SAWs on $\delta_n\mathbb{H}$ starting at the origin. Again, let n grow larger and larger.

In the paper where he introduced SAWs, Flory predicted that if we choose $\delta_n = n^{-3/4}$, then the random object Γ_n, and hence a typical n-step SAW,

---

[10] In physics, the Brownian Motion was first studied as a mathematical model for the movement of a small particle in a fluid at the microscopic scale.

[11] The *size* of a walk can be measured by the distance between its starting and ending points.

[12] In comparison to the SAW, the SRW is simpler already from a combinatorial viewpoint: while there are 3^n simple walks of length n, the number c_n of SAWs is not known exactly for larger n. Moreover, in contrast to the SAW, it can be shown that the SRW can also be sampled step-by-step, by choosing at each step independently and uniformly choosing a random next step from the 3 neighbors of a given vertex.

remains of constant size. Again, this implies that as n increases, a uniformly chosen SAW on $\mathbb{H}$ should typically end up at distance $n^{3/4}$ from the origin. For $n = 10\,000$, we would therefore expect the distance between its endpoints to be close to $1\,000\,\delta$. An example of a 10 000-step self-avoiding walk on the square lattice is shown in Figure 5.

While Flory's original argument is known today to be erroneous, incredibly, his prediction seems to be correct. We now have more compelling arguments suggesting that δ_n should indeed be taken as $n^{-3/4}$.

For this choice of δ_n, the random variables Γ_n should even converge to a continuous random curve Γ, which plays a role for the random SAW similar to the role of the Brownian Motion for the SRW. This random curve is a random "fractal", called *Schramm–Loewner Evolution*. This object appeared in recent years in the study of two-dimensional models of statistical physics "at criticality". We refer to [7] for more details and references on this more advanced subject.

In conclusion, for large n, a simple random walk of length n on average ends at a point at distance $\sqrt{n}$, while a self-avoiding walk of length n should on average end at a point at distance $n^{3/4}$. The first claim is very well understood mathematically, but the second one represents one of the main conjectures in statistical physics and remains very mysterious. While this discussion seems to have been unrelated to the discussion of the previous section, they are in fact intimately connected: one possible way to prove Flory's prediction is to show that properly "renormalized" parafermionic observables, when defined on graphs $\mathcal{D}_\delta := \delta\mathbb{H} \cap \mathcal{D}$, converge as δ tends to zero.

For the reader looking for more on the self-avoiding walk, we refer to [8] and [1].

Further reading

For further snapshots

- related to another mathematical model of statistical mechanics, the dimer model, see Snapshot 2/2016 *Random sampling of domino and lozenge tilings* by Éric Fusy, and Snapshot 16/2015 *Domino tilings of the Aztec diamond* by Juanjo Rué.
- containing a summary of basic concepts in probability theory, as well as a discussion of the simple random walk on $\mathbb{Z}$, see Snapshot 14/2015 *Quantum diffusion* by Antti Knowles, especially Chapter 2.
- related to a different discretization of holomorphic functions, see Snapshot 1/2017 *Winkeltreue zahlt sich aus* (in German) by Felix Günther.

Image credits

Figure 5 "A self-avoiding walk of length 10 000". Courtesy of Vincent Beffara, http://vbeffara.perso.math.cnrs.fr/pictures.html, visited on April 3, 2019.

References

[1] R. Bauerschmidt, H. Duminil-Copin, J. Goodman, and G. Slade, *Lectures on self-avoiding walks*, Probability and statistical physics in two and more dimensions, Clay Mathematics Proceedings, vol. 15, American Mathematical Society, 2012, pp. 395–467.

[2] N. Clisby, *Calculation of the connective constant for self-avoiding walks via the pivot algorithm*, Journal of Physics. A. Mathematical and Theoretical **46** (2013), no. 24, 245001.

[3] N. Clisby and I. Jensen, *A new transfer-matrix algorithm for exact enumerations: self-avoiding polygons on the square lattice*, Journal of Physics. A. Mathematical and Theoretical **45** (2012), no. 11, 115202.

[4] H. Duminil-Copin and S. Smirnov, *The connective constant of the honeycomb lattice equals* $\sqrt{2+\sqrt{2}}$, Annals of Mathematics. Second Series **175** (2012), no. 3, 1653–1665.

[5] P. Flory, *Principles of polymer chemistry*, Cornell University Press, 1953.

[6] J. M. Hammersley, *Percolation processes: II. The connective constant*, Mathematical Proceedings of the Cambridge Philosophical Society **53** (1957), 642–645.

[7] G. F. Lawler, O. Schramm, and W. Werner, *On the Scaling Limit of Planar Self-Avoiding Walk*, Fractal Geometry and Applications: a Jubilee of Benoît Mandelbrot, Part 2, Proceedings of Symposia in Pure Mathematics, vol. 72, American Mathematical Society, 2004, pp. 339–364.

[8] N. Madras and G. Slade, *The Self-Avoiding Walk*, Probability and its Applications, Birkhäuser, 1993.

[9] P. Mörters and Y. Peres, *Brownian Motion*, vol. 30, Cambridge University Press, 2010.

[10] B. Nienhuis, *Exact Critical Point and Critical Exponents of* $\mathrm{O}(n)$ *Models in Two Dimensions*, Physical Review Letters **49** (1982), no. 15, 1062–1065.

[11] W. J. C. Orr, *Statistical treatment of polymer solutions at infinite dilution*, Transactions of the Faraday Society **43** (1947), 12–27.

[12] R. D. Schram, G. T. Barkema, and R. H. Bisseling, *Exact enumeration of self-avoiding walks*, Journal of Statistical Mechanics: Theory and Experiment **2011** (2011), P06019.

Hugo Duminil-Copin *is a permanent professor at the Institut des Hautes Études Scientifiques as well as a full professor at the University of Geneva.*

Mathematical subjects
Probability Theory and Statistics

DOI
10.14760/SNAP-2019-006-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

ISSN 2626-1995

Mathematisches Forschungsinstitut
Oberwolfach gGmbH
Schwarzwaldstr. 9 – 11
77709 Oberwolfach
Germany

Director
Gerhard Huisken

Mathematisches Forschungsinstitut Oberwolfach

Member of Leibniz Association

IMAGINARY open mathematics