Computing invariants of knotted graphs given by sequences of points in 3-space

Vitaliy Kurlin, http://kurlin.org
Microsoft Research Cambridge
and Durham University, UK

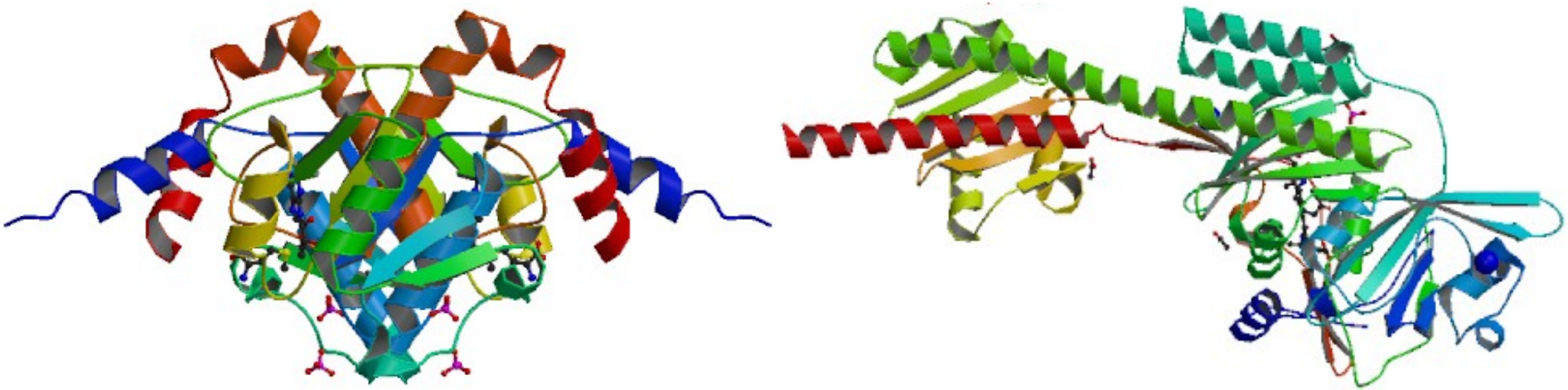
The secondment at Microsoft was supported by the EPSRC *Impact Acceleration Account*

Availability of real proteins

Protein Data Bank http://www.rcsb.org/pdb.

Each PDB file has 3D coordinates of atoms

linearly ordered along a protein backbone.

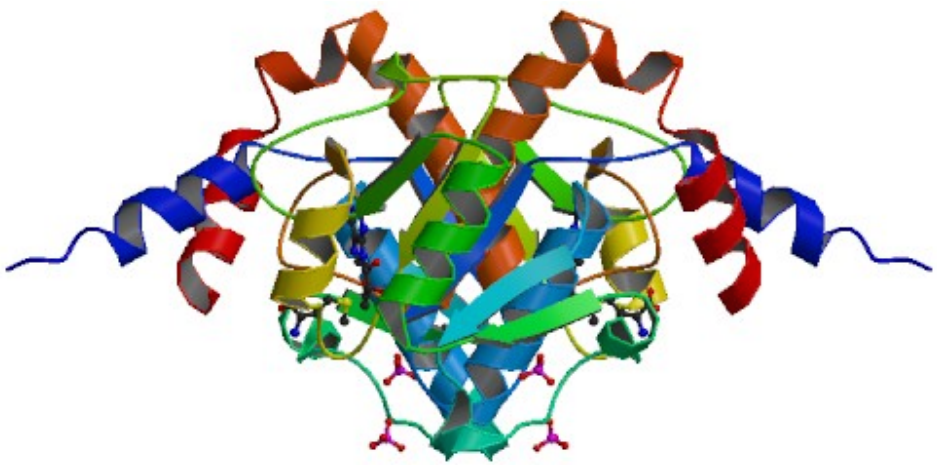


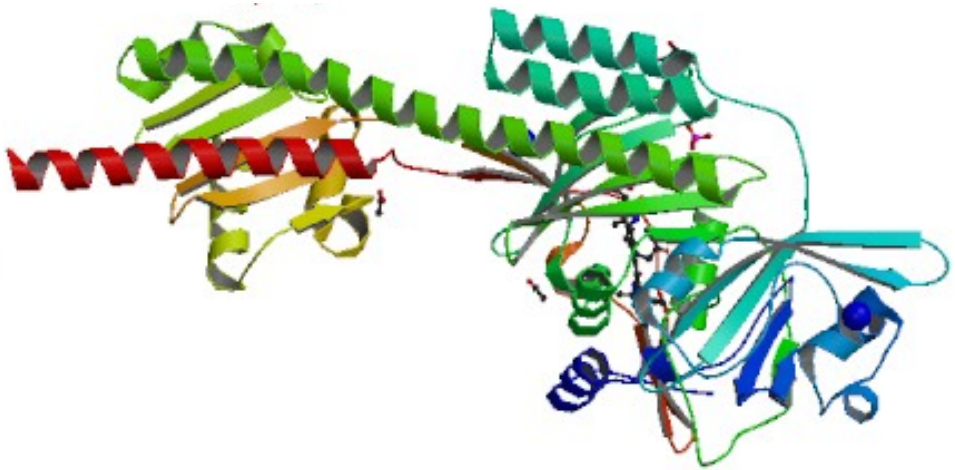
Left: 1v2x tRNA methyltransferase. Right: 3zq5.

Problems on knotted structures

Knotted graphs are good model for proteins,

harder to *visualize* and *recognize* than knots.

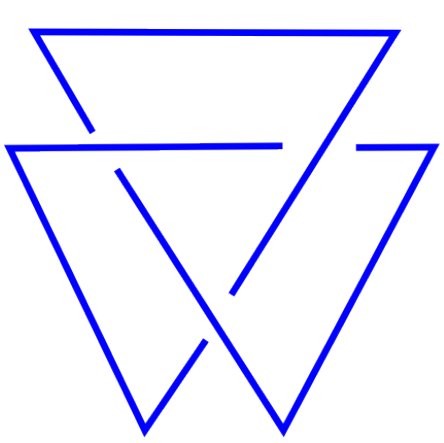


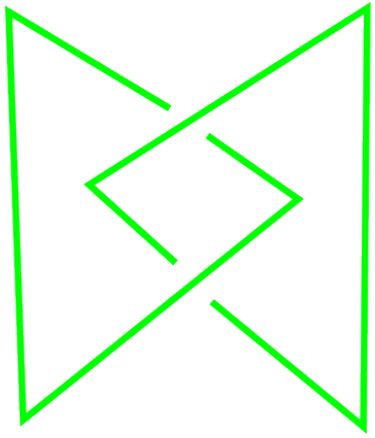


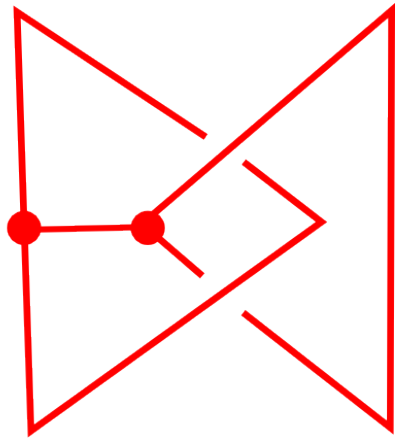
- *Encode* long knotted structures in a simple way.
- *Compare* knotted structures up to deformations.

Polygonal knotted graphs in $\mathbb{R}^3$

Def: a polygonal knotted graph is an embedding $f : G \to \mathbb{R}^3$ consisting of finitely many *straight segments*. So $f(G)$ has no self-intersections, but may have *double crossings* under $f(G) \to \mathbb{R}^2$.





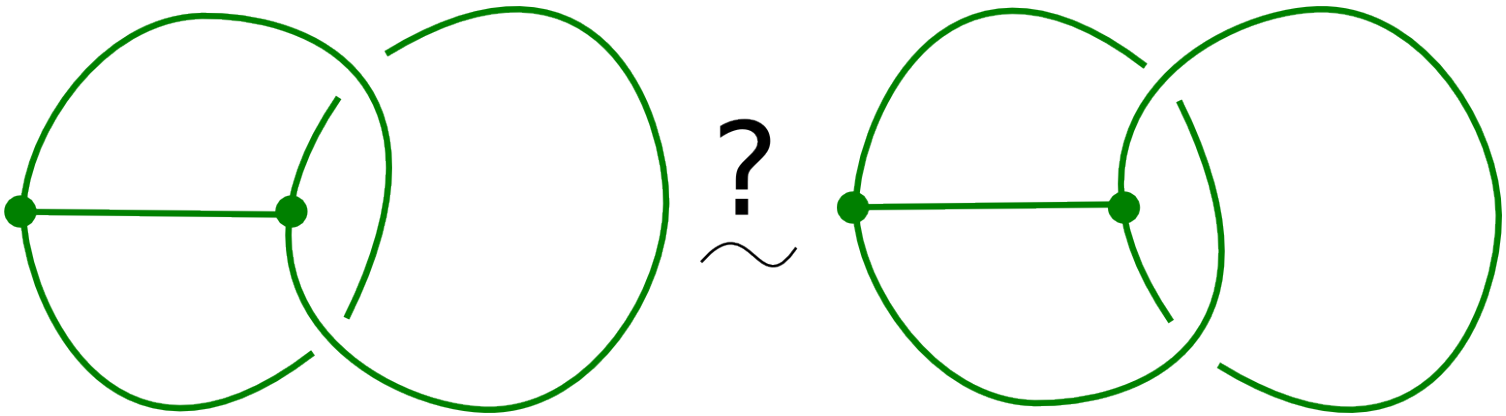


If $G \approx S^1$, the knotted graph $S^1 \subset \mathbb{R}^3$ is a knot.

If $G \approx \sqcup_{i=1}^{m} S_i^1$, the knotted graph is called a link.

Isotopy of knotted graphs

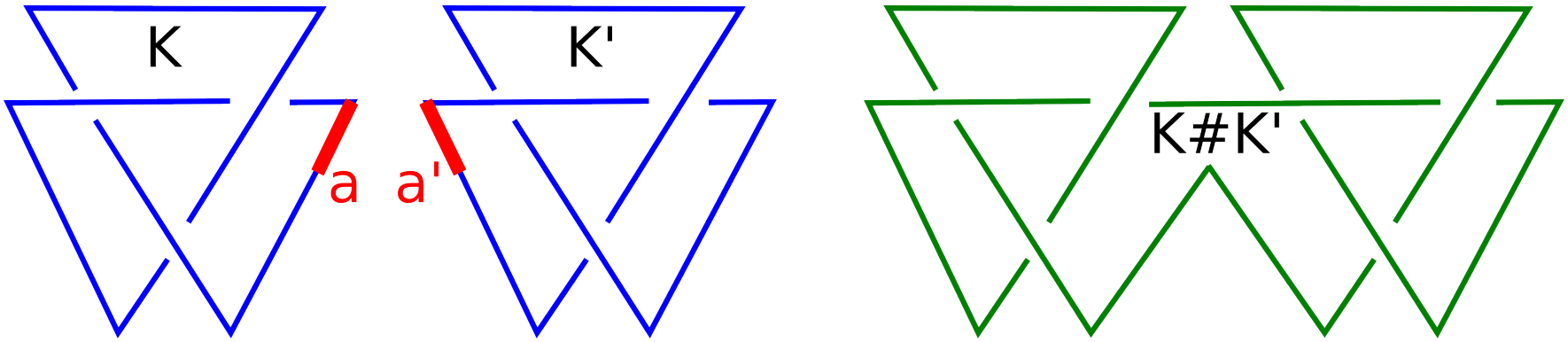
Def: an ambient isotopy between graphs $G, H \subset \mathbb{R}^3$ is a continuous family of ambient homeomorphisms $F_t : \mathbb{R}^3 \to \mathbb{R}^3$, $t \in [0, 1]$, where $F_0 = \mathbf{id}$ on $\mathbb{R}^3$ and $F_1(G) = H$.



Are these two graphs isotopic or not in $\mathbb{R}^3$?

Classify only prime knots

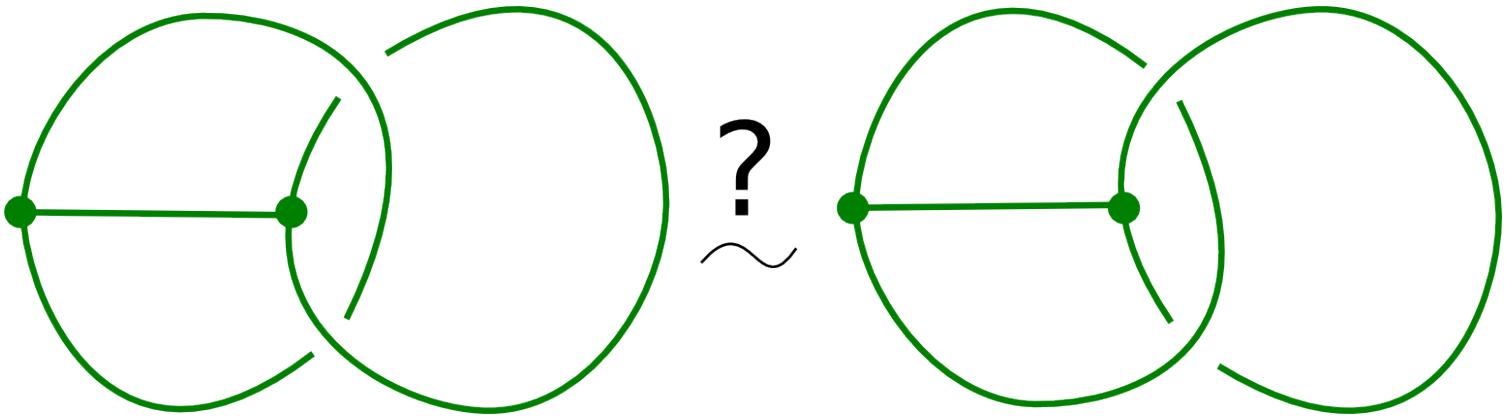
Def : a knot not isotopic to a *connected sum* of non-trivial knots is prime, similarly for graphs.



Any knot uniquely decomposes into prime knots (up to a permutation of summands as usual).

Isotopy invariants of graphs

An invariant is a function $\{\text{isotopy classes}\} \to \{\text{numbers}\}$ whose values are easy to compare.

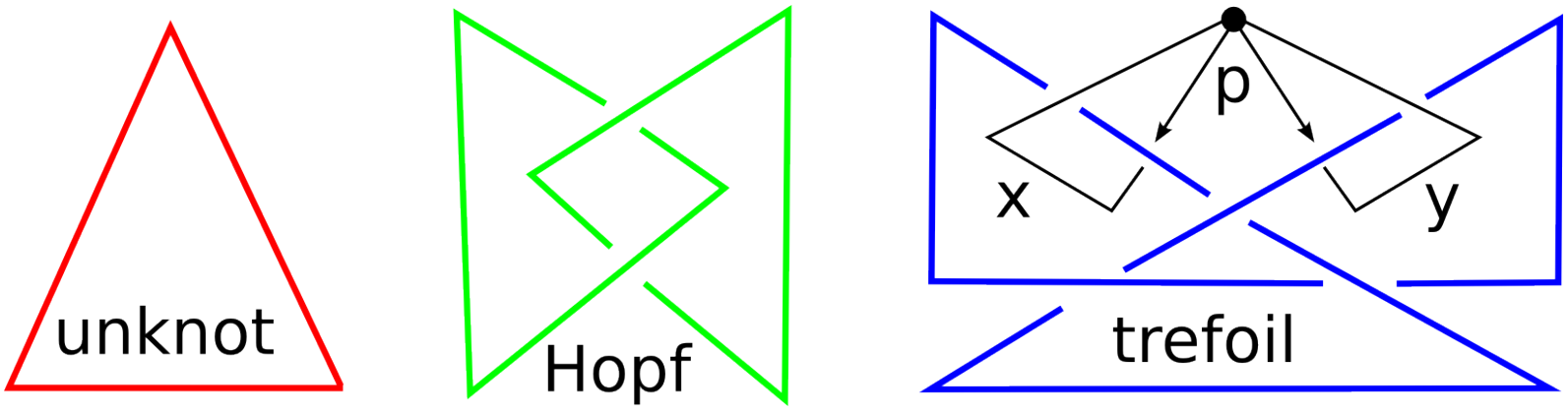


Simple: number of components of G is weak.

Powerful: 3D complement $\mathbb{R}^3 - G$ of the graph.

Knotted Graph Group $\pi_1(\mathbb{R}^3 - G)$

Def: KGG is the *group of closed loops* in $\mathbb{R}^3 - G$ at a base point p up to continuous movements.



Unknot : $\pi_1 = \mathbb{Z}$ (a loop can go around n times).

Hopf link : $\pi_1 = \mathbb{Z}^2$ ($\mathbb{R}^3 - G$ deforms to a torus).

Trefoil : $\pi_1 = \langle x, y \mid xyx = yxy \rangle$, see the picture.

π_1 : almost complete invariant

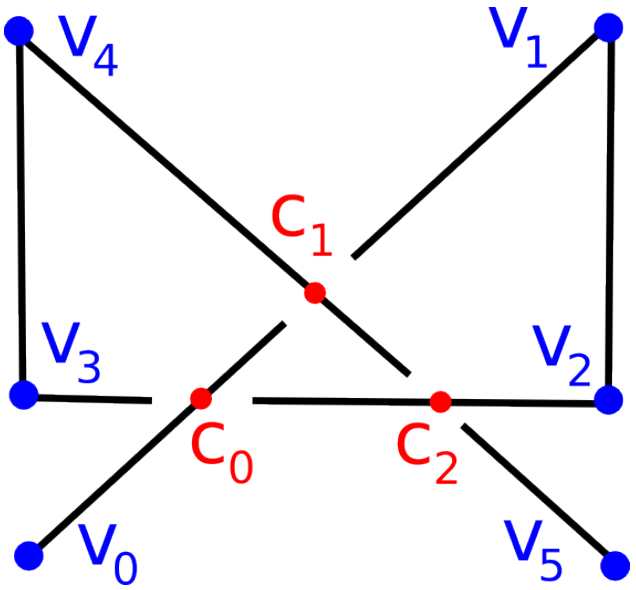
Th (Gordon, Luecke, 1989). Knots $K, K' \subset S^3$ are isotopic *if and only if* $S^3 - K \approx S^3 - K'$ (an orientation-preserving homeomoprhism).

Th (Whitten, 1987). For prime knots K, K', the complements $S^3 - K \approx S^3 - K'$ *if and only if* their groups $\pi_1(S^3 - K) \cong \pi_1(S^3 - K')$.

Input and output in practice

Each PDB file has 3D coordinates of all atoms.

Input: a *sequence of 3D points* along every edge-path of a polygonal graph $G \subset \mathbb{R}^3$, e.g.



$v_0, v_1, v_2, v_3, v_4, v_5 \in \mathbb{R}^3$.

Output: a presentation of the *Knotted Graph Group* $\pi_1(\mathbb{R}^3 - G)$, e.g. $\pi_1 = \langle x, y \mid xyx = yxy \rangle$.

Alexander polynomial is easier

$\pi_1(\mathbb{R}^3 - K)$ leads to *Alexander polynomial* of K.

Practically simple and powerful for small knots:

- computed in time $O(c^3)$ for c crossings of K,
- 550 values on 801 prime knots up to $c = 11$.
- used in *KnotProt* database of proteins.

For longer knots in larger real proteins, we now need more powerful invariants such as KGG.

Knotted Graph Group algorithm

$\{$Sequences of 3D points of $G\} \rightarrow \pi_1(\mathbb{R}^3 - G)$.

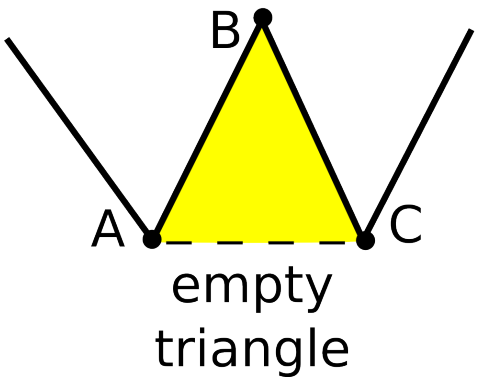
Stage 1. *Shorten a graph* $G \subset \mathbb{R}^3$ of a length n to a graph G' of a length $m \leq n$ in time $O(n^2)$.

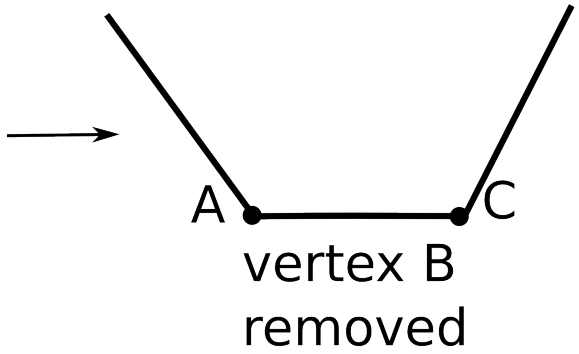
Stage 2. *Find a Gauss code* of a length $O(m^2)$ for the shortened graph G' in time $O(m^2)$.

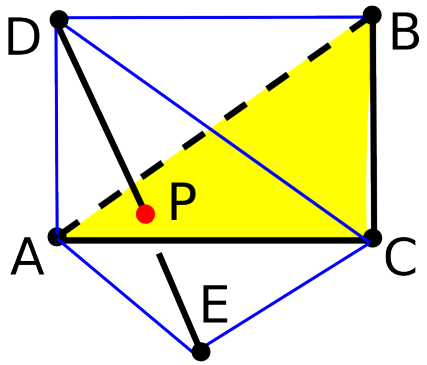
Stage 3. Turn a Gauss code into the *Knotted Graph Group* $\pi_1(\mathbb{R}^3 - G)$ in time $O(m^2)$.

Stage 1: shortening a graph G

KMT algorithm for shortening polygonal chains is used in *Rosetta* for 3D protein structures and detects if another edge DE meets $\triangle ABC$ by finding an intersection $P = DE \cap$ plane ABC and checking if $\angle APB + \angle BPC + \angle CPA = 2\pi$.







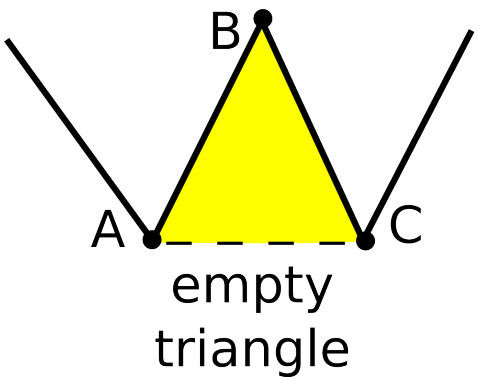
Arithmetic floating point errors are up to $3 \cdot 10^{-4}$.

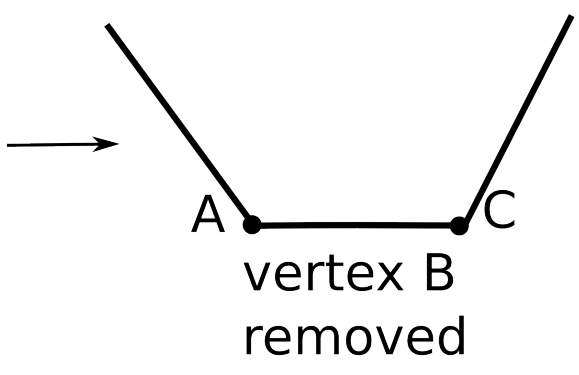
Improved KMT algorithm for KGG

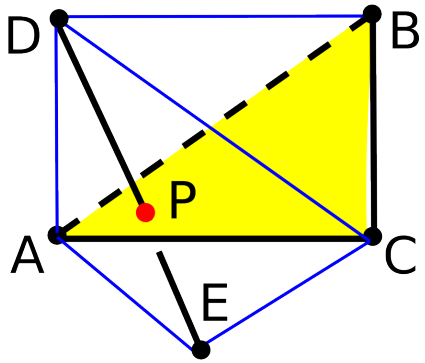
An edge DE meets $\triangle ABC$ if and only if the volumes of 5 tetrahedra satisfy the equation

$$|V_{ABCD}| + |V_{ABCE}| = |V_{ABDE}| + |V_{ACDE}| + |V_{BCDE}|.$$

Each signed volume is a 3×3 determinant.







This criterion has much smaller error $\approx 10^{-10}$ in comparison with $3 \cdot 10^{-4}$ in the standard KMT.

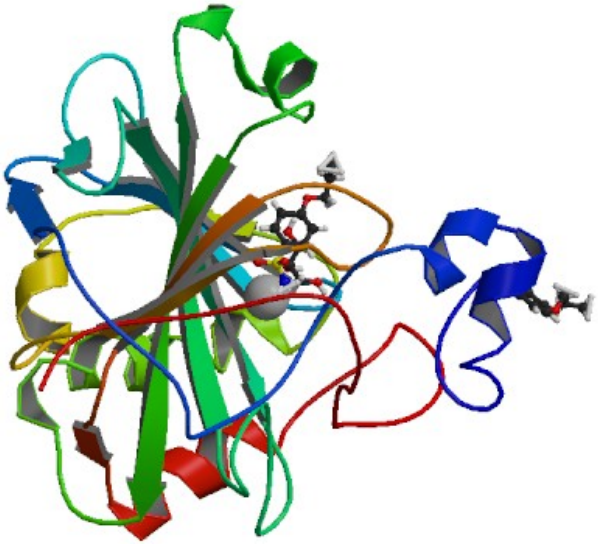
Experiments on knotted proteins

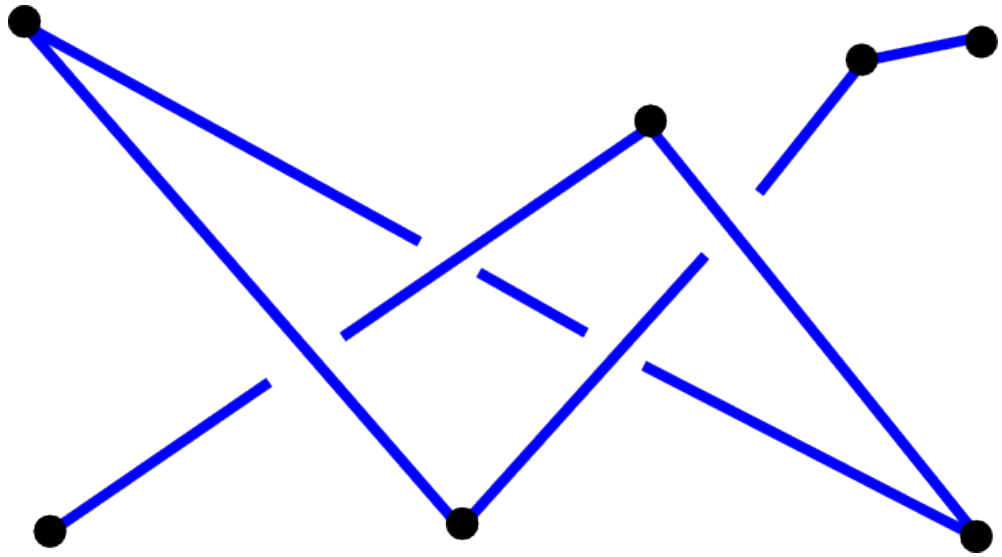
A polygonal chain $K \subset \mathbb{R}^3$ with $v(K)$ vertices and $c(K)$ crossings simplifies to K' whose Gauss code and knot type are easily found.

PDB	$v(K)$	$c(K)$	$v(K')$	$c(K')$	knot
1v2x	191	39	10	4	3_1
4ruy	263	101	7	4	3_1
3oil	267	102	7	3	3_1
2rh3	124	26	7	4	-3_1
3zq5	517	174	7	6	4_1

One simplified protein: 4ruy

The initial protein with 263 vertices and 101 crossings reduces to 7 vertices and 4 crossings.



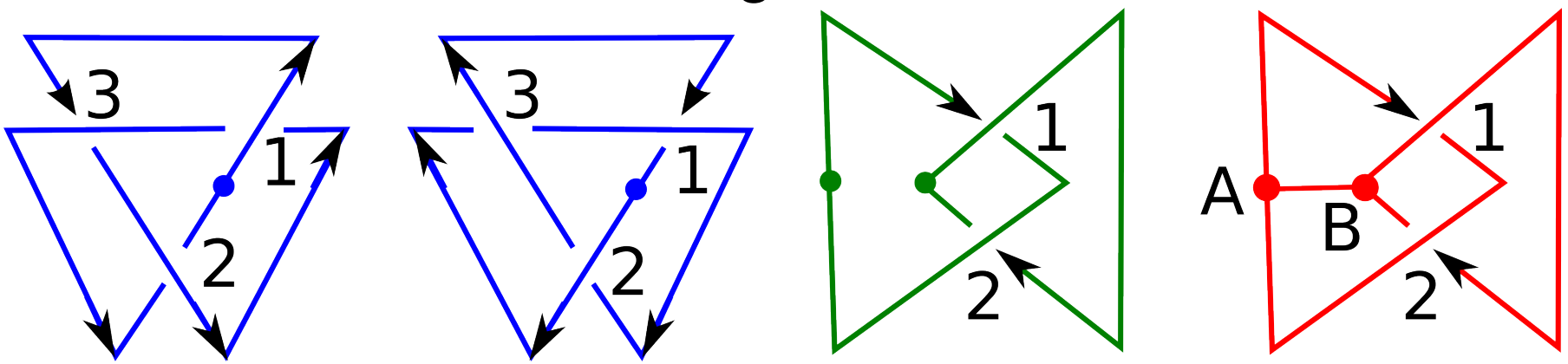


The diagram is *geometrically* (not topologically) *minimal* since no triangles can be removed.

The knot type is a *trefoil* confirmed by KGG.

Stage 2: Gauss code of a graph

Def: in a plane diagram of $G \subset \mathbb{R}^3$, *label vertices* and *crossings* along each directed edge-path of G. Each undercrossing has the extra minus.



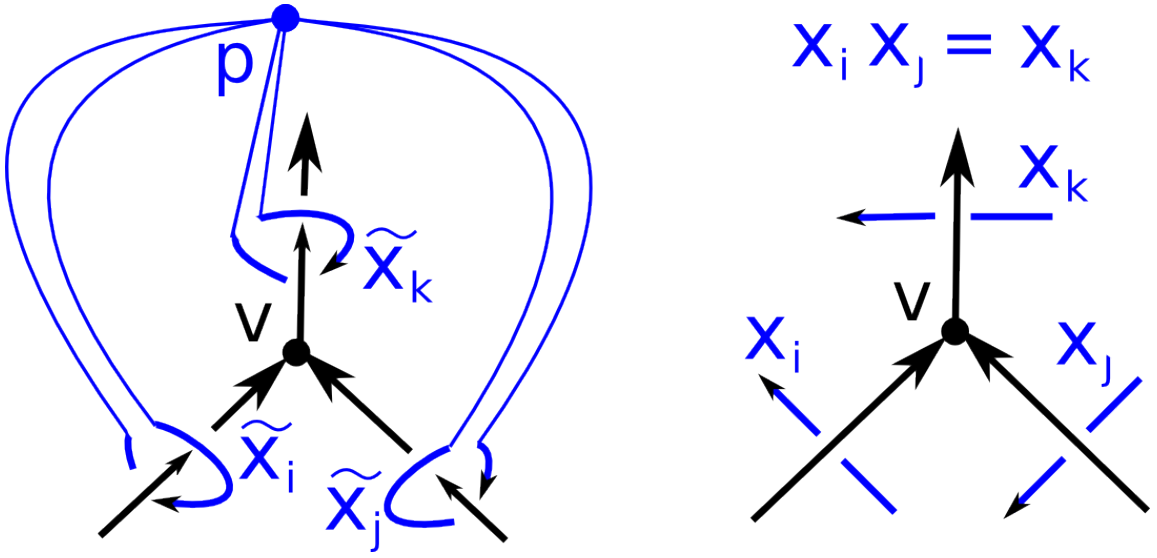
Trefoils have Gauss codes $(1,-3,2,-1,3,-2)$, $(2,-3,1,-2,3,-1)$. *Hopf link* : $(-1,2)$, $(1,-2)$.

Hopf graph : (A,B), $(A,-1,2,A)$, $(B,1,-2,B)$.

π_1 : relations for loops at vertices

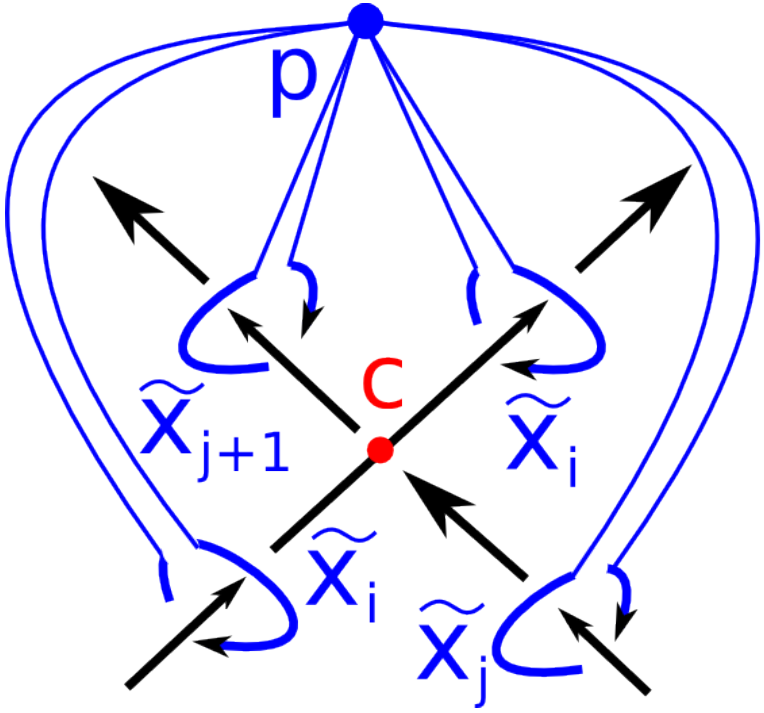
Fix a point p at ∞ above the plane diagram.

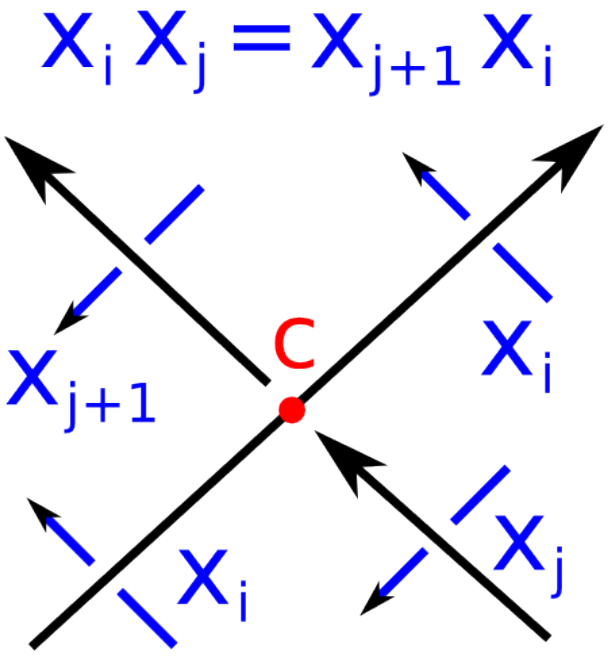
Loops $\tilde{x}_i$ around arcs represent generators x_i.



Relation for a vertex v is $x_i x_j x_k^{-1} = 1$, i.e. power +1 for incoming arcs, -1 for outgoing arcs.

π_1 : relations for crossings



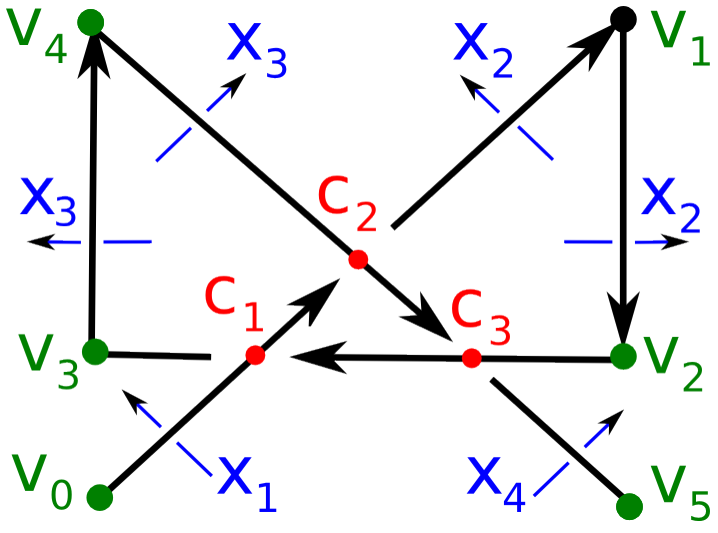


Relation for a crossing c is $x_i x_j x_i^{-1} = x_{j+1}$, i.e.

the next x_{j+1} is conjugate to the previous x_j.

Stage 3: presentation of $\pi_1(\mathbb{R}^3 - G')$

A diagram splits by crossings and vertices into



arcs: $a_1 = [v_0, c_1, c_2]$,

$a_2 = [c_2, v_1, v_2, c_3, c_1]$,

$a_3 = [c_1, v_3, v_4, c_2, c_3]$,

$a_4 = [c_3, v_5]$.

Use a Gauss code to write *Wirtinger relations*:

$$x_1 x_2 x_1^{-1} = x_3, \quad x_3^{-1} x_1 x_3 = x_2, \quad x_2 x_4 x_2^{-1} = x_3,$$

with $x_1 = x_4$ we get $\pi_1 = \langle x, y \mid xyx = yxy \rangle$.

Running time for stages 1, 2, 3

Stage 1: $O(n \log n)$ to order all deg 2 vertices by the increasing length $|AC|$ between neighbours.

$O(n)$ time to decide if ABC is replaced by AC.

Stage 2: go along the shortened graph G' and note $O(m^2)$ intersections of projected m edges.

Stage 3: $\pi_1(\mathbb{R}^3 - G')$ has $O(m^2)$ generators, convert each vertex/crossing into a relation.

Past work: π_1 from a big cubical 2-complex obtained from $\mathbb{R}^3 - K$ at a fixed resolution.

Abelian invariants of a group

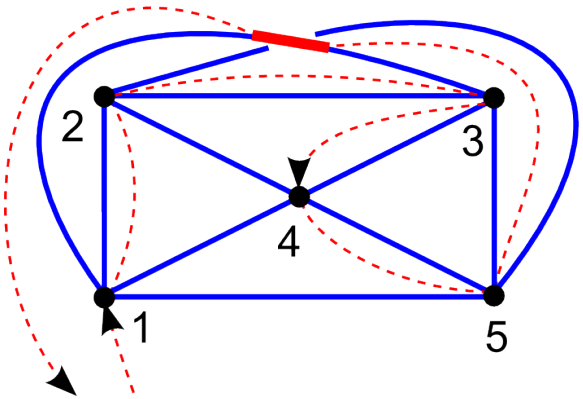
Th: any finitely generated abelian group is $\mathbb{Z}^r \oplus \mathbb{Z}_{q_1} \oplus \cdots \oplus \mathbb{Z}_{q_l}$, where abelian invariants $q_1, \ldots, q_l \geq 2$ are unique up to permutation.

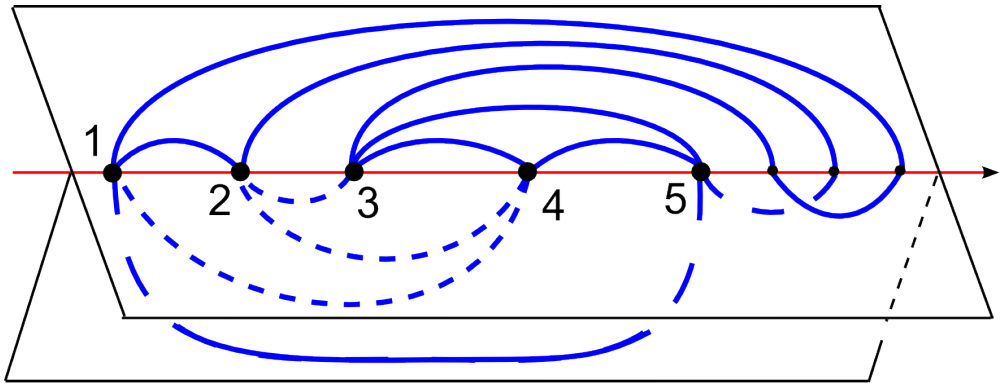
Def: the abelian invariants of any non-abelian group come from the abelian quotients $H/[H, H]$ for all subgroups H up to a certain index.

Brendel, Dlotko, Ellis, Juda, Mrozek: abelian invariants of $\pi_1(\mathbb{R}^3 - K)$ with indices up to 6 *distinguish all prime knots* up to 11 crossings.

3-page embedding in linear time

Th (VK, IVAPP'15): for any Gauss code W of a graph $G \subset \mathbb{R}^3$, in *time* $O(|W|)$ we can *embed* G into a *3-page book* (a union of 3 half-planes).





The *best recognition* algorithm for knots & links simplifies a 3-page embedding by local moves.

Next step : extend this simplification to graphs.

Summary and future work

- A knotted graph of a length n is shortened to a smaller length $m \leq n$ in time $O(n^2)$.
- A presentation of *Knotted Graph Group* $\pi_1(\mathbb{R}^3 - G)$ is computed in time $O(m^2)$
- *C++ code*, examples at http://kurlin.org.

- *Compute* abelian invariants of the Knotted Graph Group $\pi_1(\mathbb{R}^3 - G)$ using GAP.
- *Classify* periodic entanglements in 3-torus by comparing abelian invariants of KGG.