INF421, Lecture 9
Drawing graphs

Leo Liberti

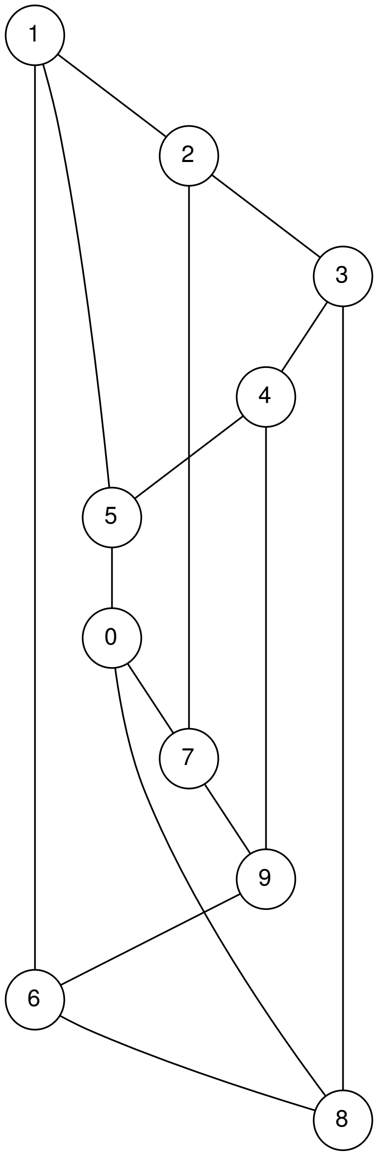
LIX, École Polytechnique, France

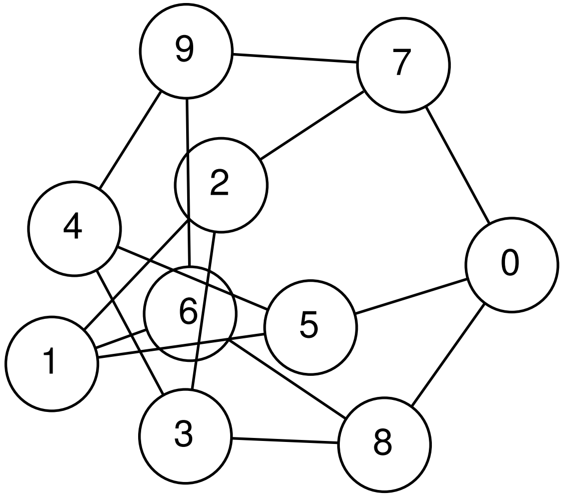
Course

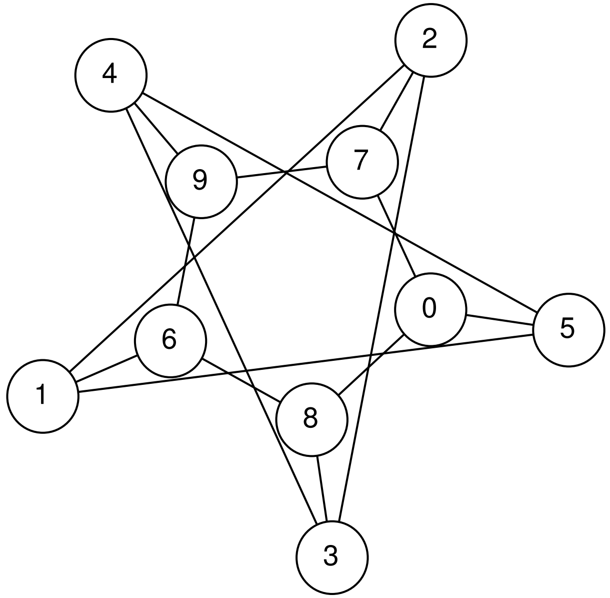
- **Objective**: teach notions AND develop intelligence
- **Evaluation**: TP noté en salle info, Contrôle à la fin. Note: $\max(CC, \frac{3}{4}CC + \frac{1}{4}TP)$
- **Organization**: fri 31/8, 7/9, 14/9, 21/9, 28/9, 5/10, 12/10, 19/10, 26/10, amphi 1030-12 (Arago), TD 1330-1530, 1545-1745 (SI:30-34)
- **Books**:
  1. K. Mehlhorn & P. Sanders, *Algorithms and Data Structures*, Springer, 2008
  2. D. Knuth, *The Art of Computer Programming*, Addison-Wesley, 1997
  3. G. Dowek, *Les principes des langages de programmation*, Editions de l'X, 2008
  4. Ph. Baptiste & L. Maranget, *Programmation et Algorithmique*, Ecole Polytechnique (Polycopié), 2006
- **Website**: `www.enseignement.polytechnique.fr/informatique/INF421`
- **Blog**: `inf421.wordpress.com`
- **Contact**: `liberti@lix.polytechnique.fr` (e-mail subject: INF421)

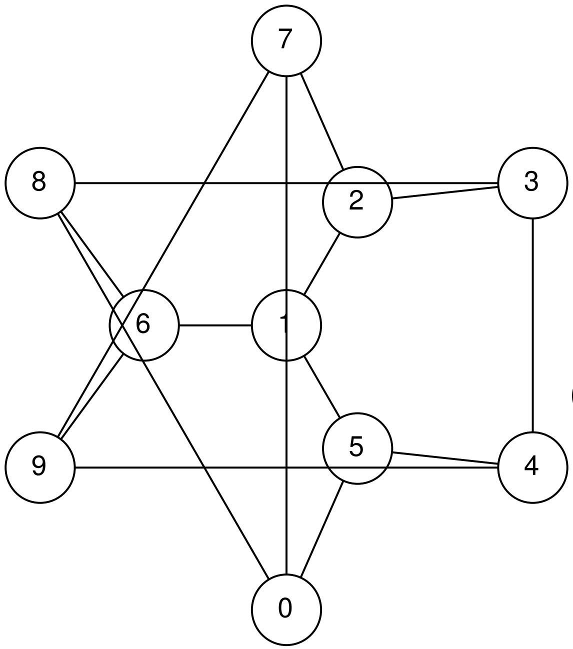
Today, a “research seminar”!

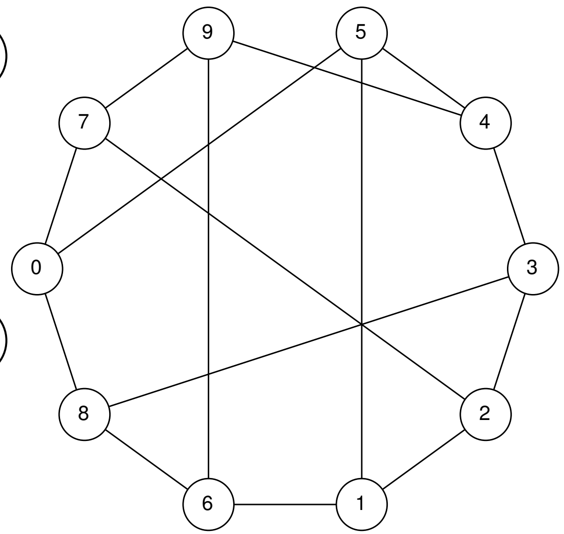
At a glance







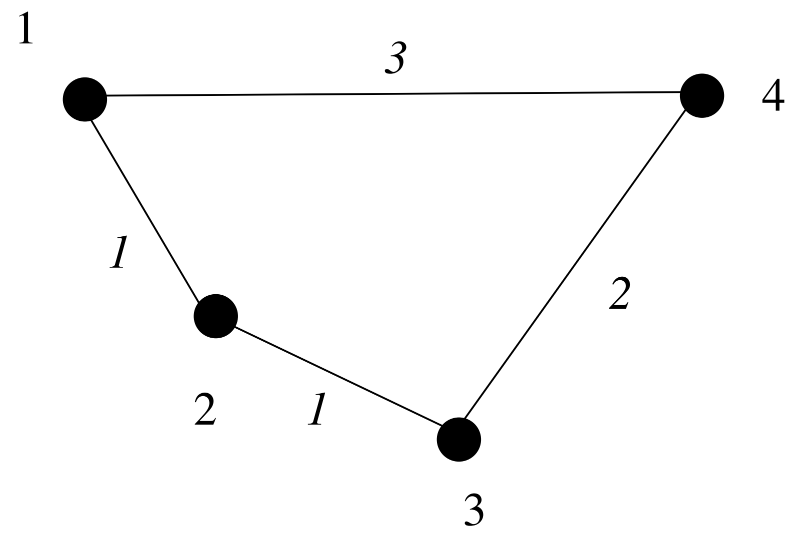




Which graph has most symmetries?

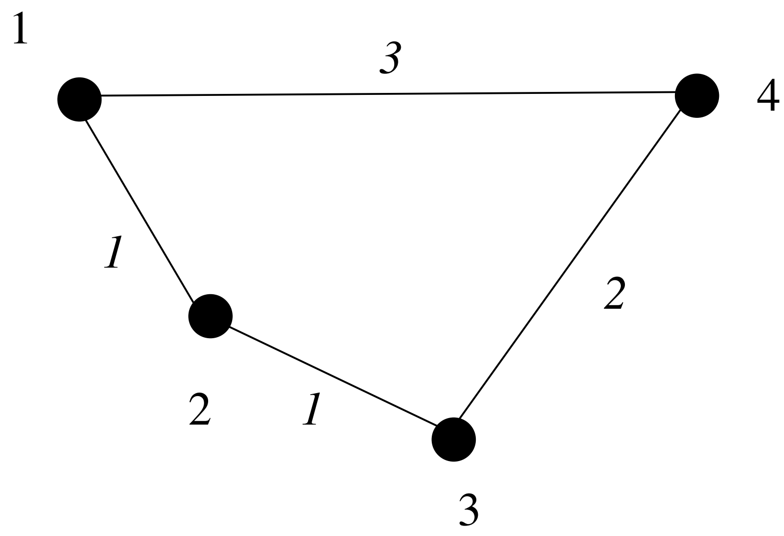
How does a weighted graph look?

- Like this?

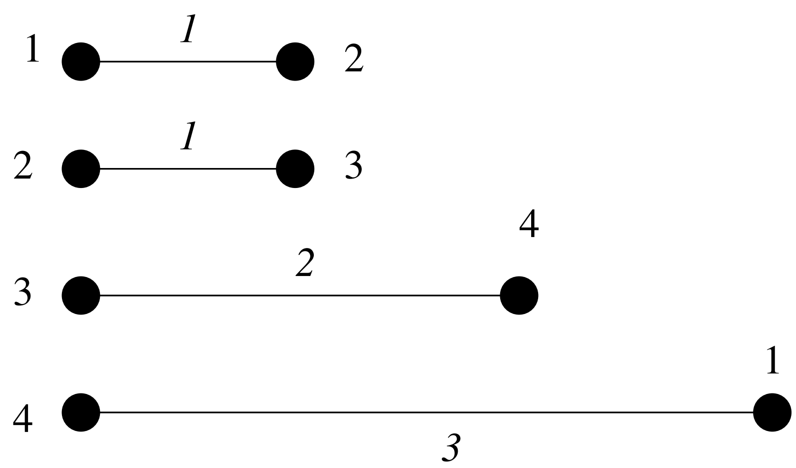


How does a weighted graph look?

- Like this?



- Perhaps like this?



Don't confuse a graph with its drawing

Clean energy

- Use hydrogen to produce chemical energy
- How to produce “pure hydrogen”?
- Photosystem II : *complex molecular conglomerate*
- Molecular function $\leftrightarrow$ 3D shape
- Molecule $=$ graph
  - Atoms $=$ vertices
  - Known inter-atomic distances $=$ edges

Draw a weighted graph in 3D

Other applications

Applications:

- Clock synchronization, phase retrieval (A. D'Aspremont, CMAP) — 1D
- Wireless sensor network localization — 2D
- Molecule conformation (me, LIX) / submarine localization — 3D
- Multidimensional scaling — (whatever)D

Drawing a graph

- Given a simple weighted undirected graph $G = (V, E)$ with a distance function $d : E \to \mathbb{R}_+$, solve the constraint system:

$$\forall \{u, v\} \in E \quad \|x_u - x_v\| = d_{uv} \tag{1}$$

- Obtain an embedding $x : V \to \mathbb{R}^2$

Global optimization

- Reformulate (1) to

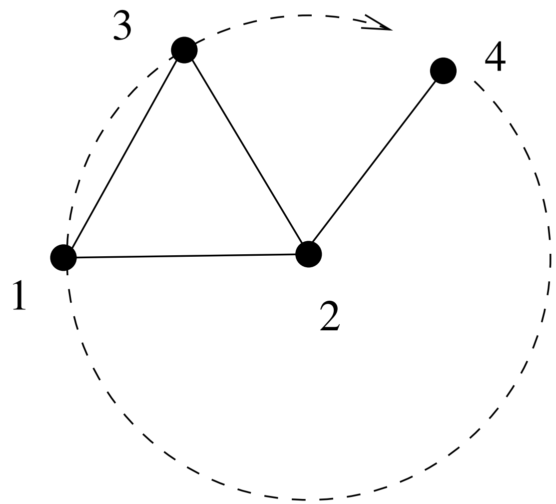
$$\min_x \sum_{\{u,v\}\in E} (\|x_u - x_v\|^2 - d_{uv}^2)^2 \qquad (2)$$

- G has an embedding $\Leftrightarrow$ optimum x^* of (2) has value 0.
- Eq (2) is nonconvex in x, many local optima

Try it on Matlab/Octave/Maple/whatever for simple data, you won't get very far (< 10 vertices)

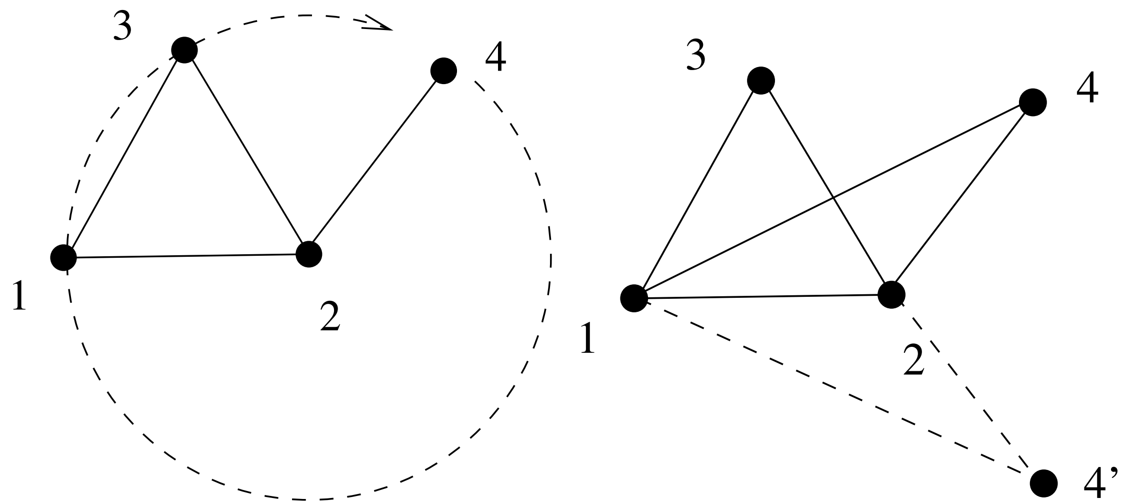
The number of embeddings

- Uncountably many (incongruent) embeddings



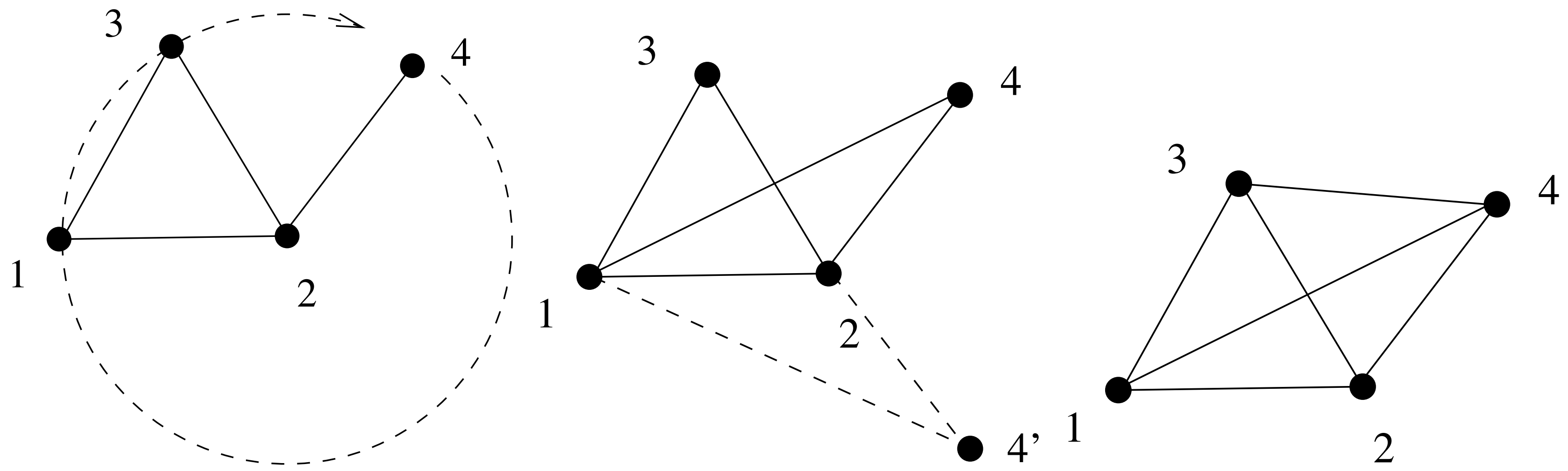
The number of embeddings

- Uncountably many (incongruent) embeddings
- Finitely many

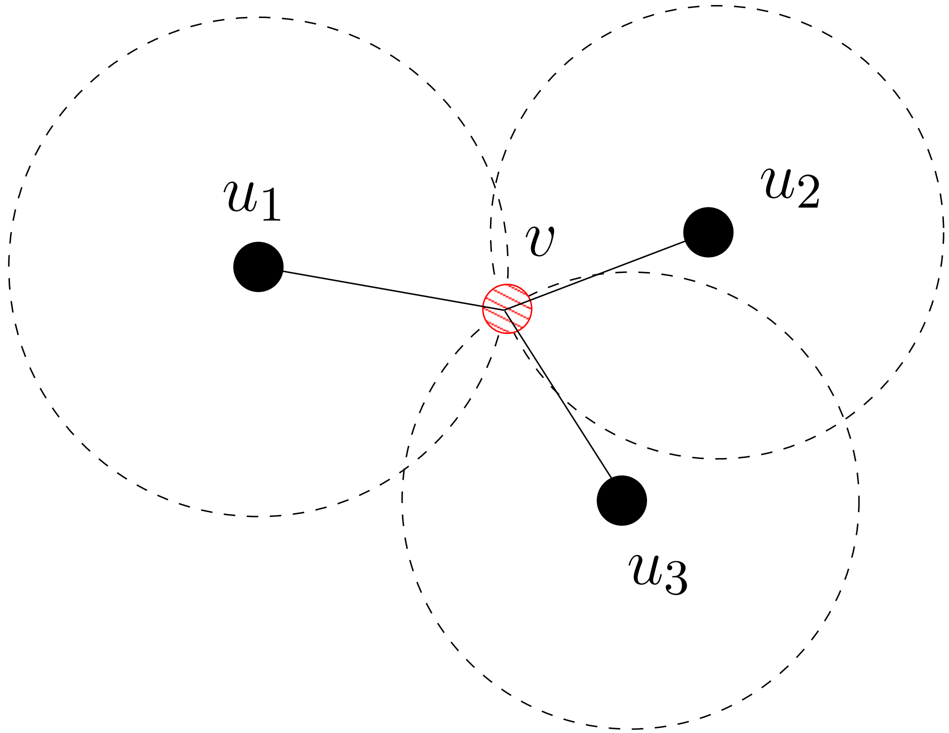


The number of embeddings

- Uncountably many (incongruent) embeddings
- Finitely many
- At most one



K-lateration



v has $\geq K + 1$ adjacencies with known general positions $\Rightarrow$

Find unique position for x_v in $\mathbb{R}^K$ in polytime

Example with $K = 3$

Given $U = \{1,2,3,4\} \subseteq V$ and a partial embedding $x_1, x_2, x_3, x_4 \in \mathbb{R}^3$

1. Consider v adjacent to all $u \in U$
2. Extend x to v by solving a linear system:

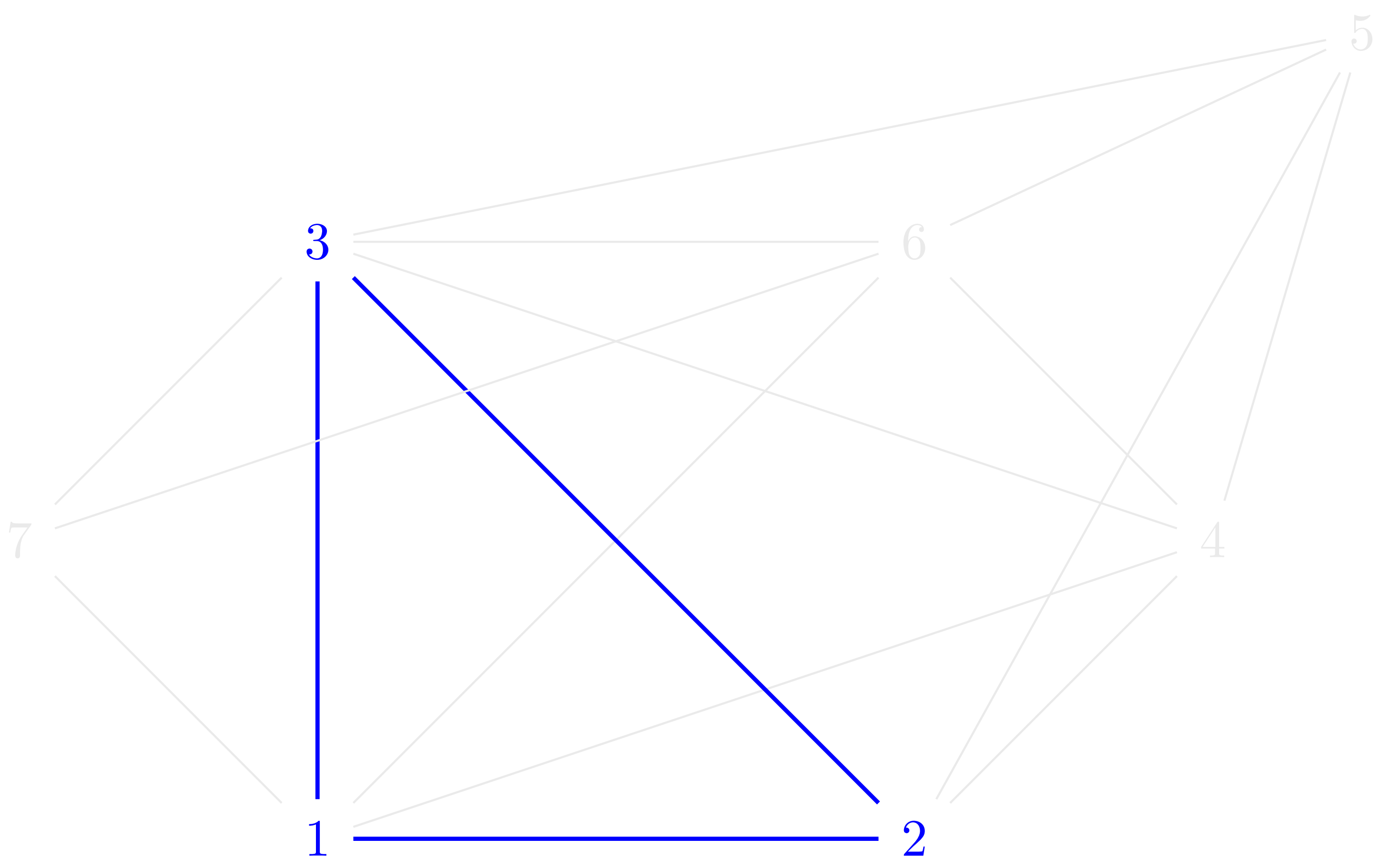
$$\boxed{\begin{array}{rcl} \|x_v - x_1\|^2 &=& d_{1v}^2 \\ \|x_v - x_2\|^2 &=& d_{2v}^2 \\ \|x_v - x_3\|^2 &=& d_{3v}^2 \\ \|x_v - x_4\|^2 &=& d_{3v}^2 \end{array}} \Rightarrow \boxed{\begin{array}{rcll} \|x_v\|^2 - 2x_v \cdot x_1 + \|x_1\|^2 &=& d_{1v}^2 & (3) \\ \|x_v\|^2 - 2x_v \cdot x_2 + \|x_2\|^2 &=& d_{1v}^2 & (4) \\ \|x_v\|^2 - 2x_v \cdot x_3 + \|x_3\|^2 &=& d_{1v}^2 & (5) \\ \|x_v\|^2 - 2x_v \cdot x_4 + \|x_4\|^2 &=& d_{1v}^2 & (6) \end{array}}$$

$$\boxed{\begin{array}{l} (6)\text{-}(7) \\ (6)\text{-}(8) \\ (6)\text{-}(9) \end{array}} \Rightarrow \boxed{\begin{pmatrix} 2(x_1 - x_4)^\mathsf{T} \\ 2(x_2 - x_4)^\mathsf{T} \\ 2(x_3 - x_4)^\mathsf{T} \end{pmatrix} x_v = \begin{pmatrix} (\|x_1\|^2 - \|x_4\|^2) - (d_{1v}^2 - d_{4v}^2) \\ (\|x_2\|^2 - \|x_4\|^2) - (d_{2v}^2 - d_{4v}^2) \\ (\|x_3\|^2 - \|x_4\|^2) - (d_{3v}^2 - d_{4v}^2) \end{pmatrix}}$$

Can do this in $O(K^3)$, if K is fixed, this is $O(1)$

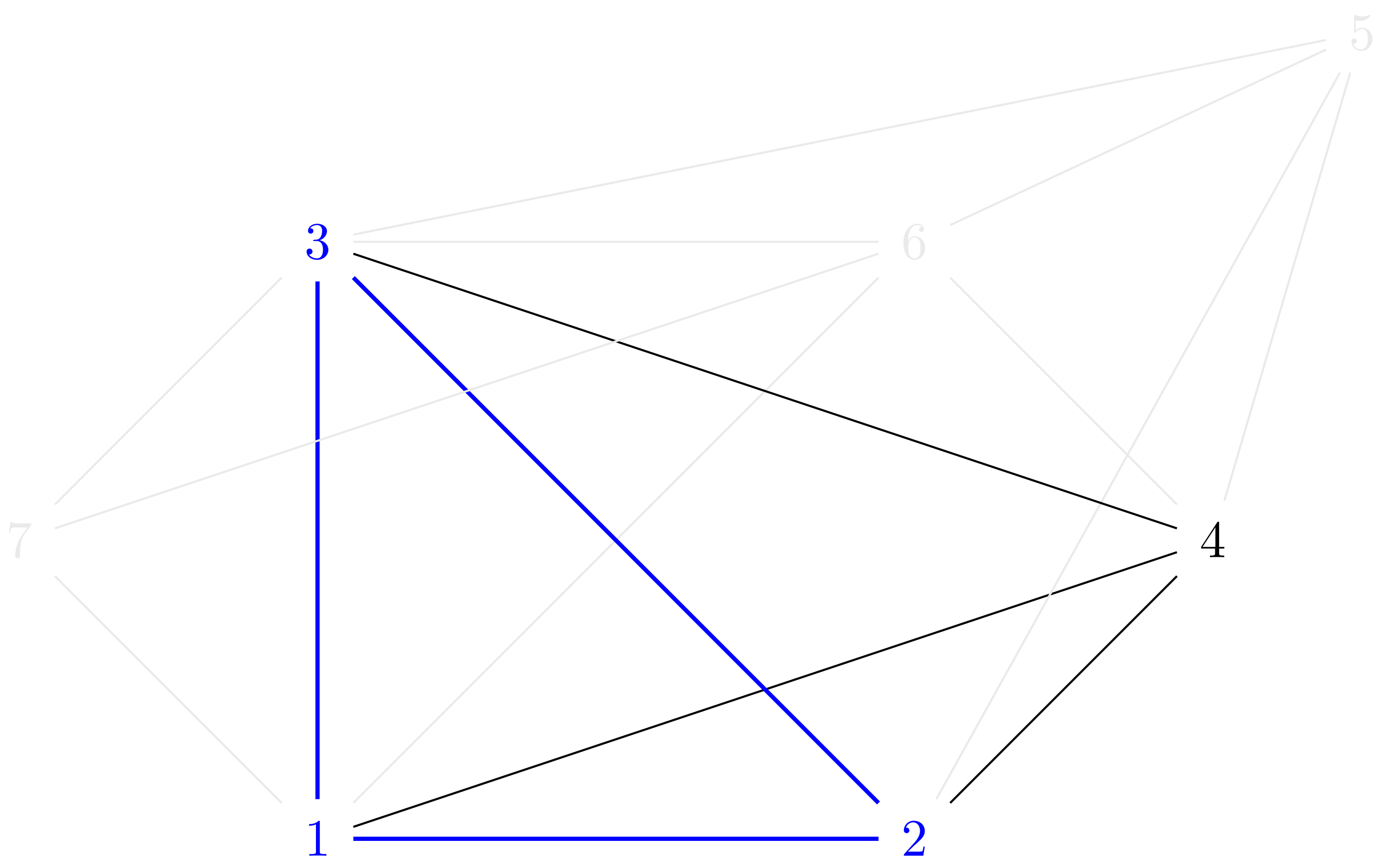
Combinatorial iterative approach

$K = 2$; *if* $\exists$ *vertex order s.t.* *next vertex* *has enough* **adjacent predecessors** *:*



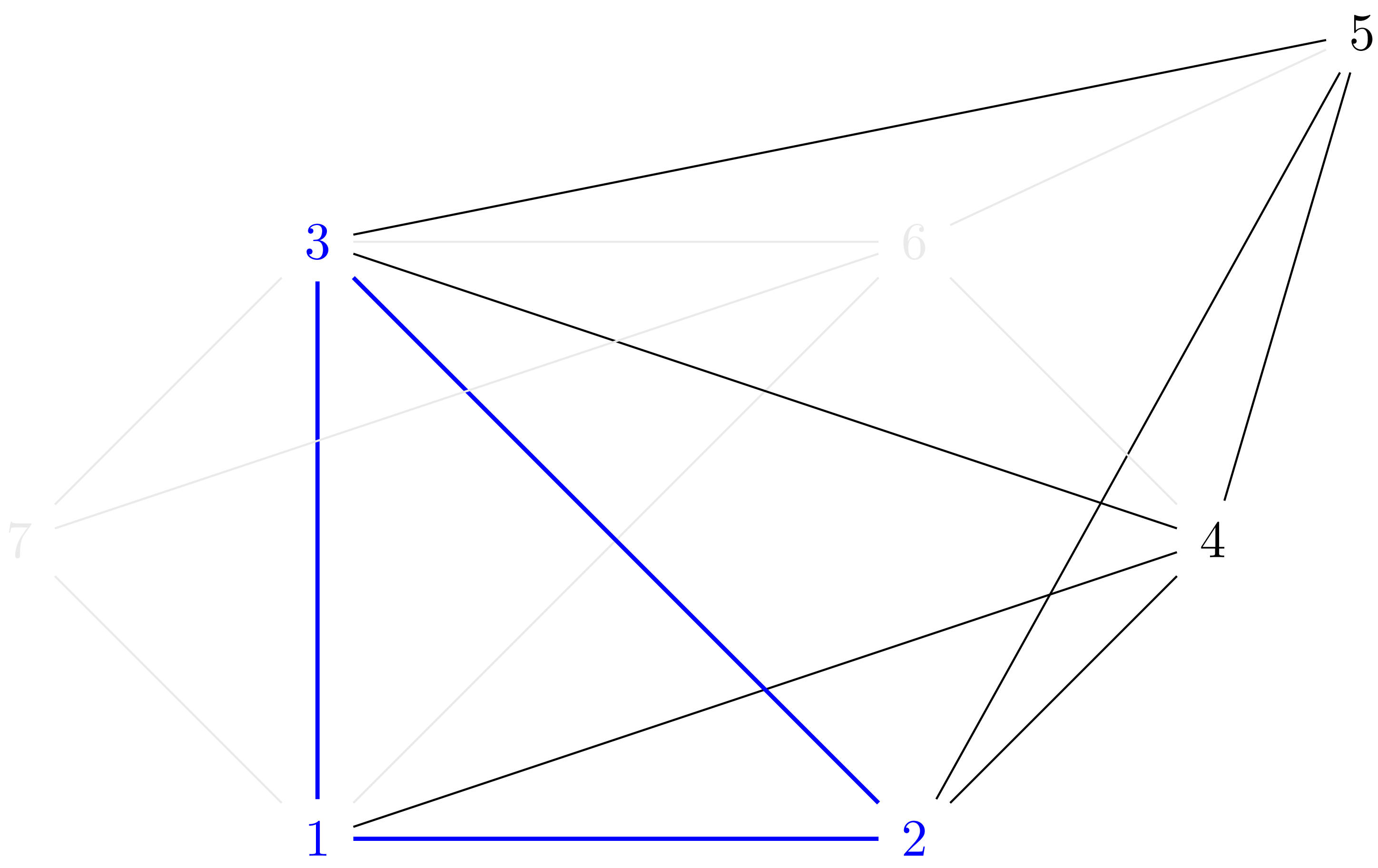
Combinatorial iterative approach

$K = 2$; *if* $\exists$ *vertex order s.t.* *next vertex* *has enough* **adjacent predecessors**:



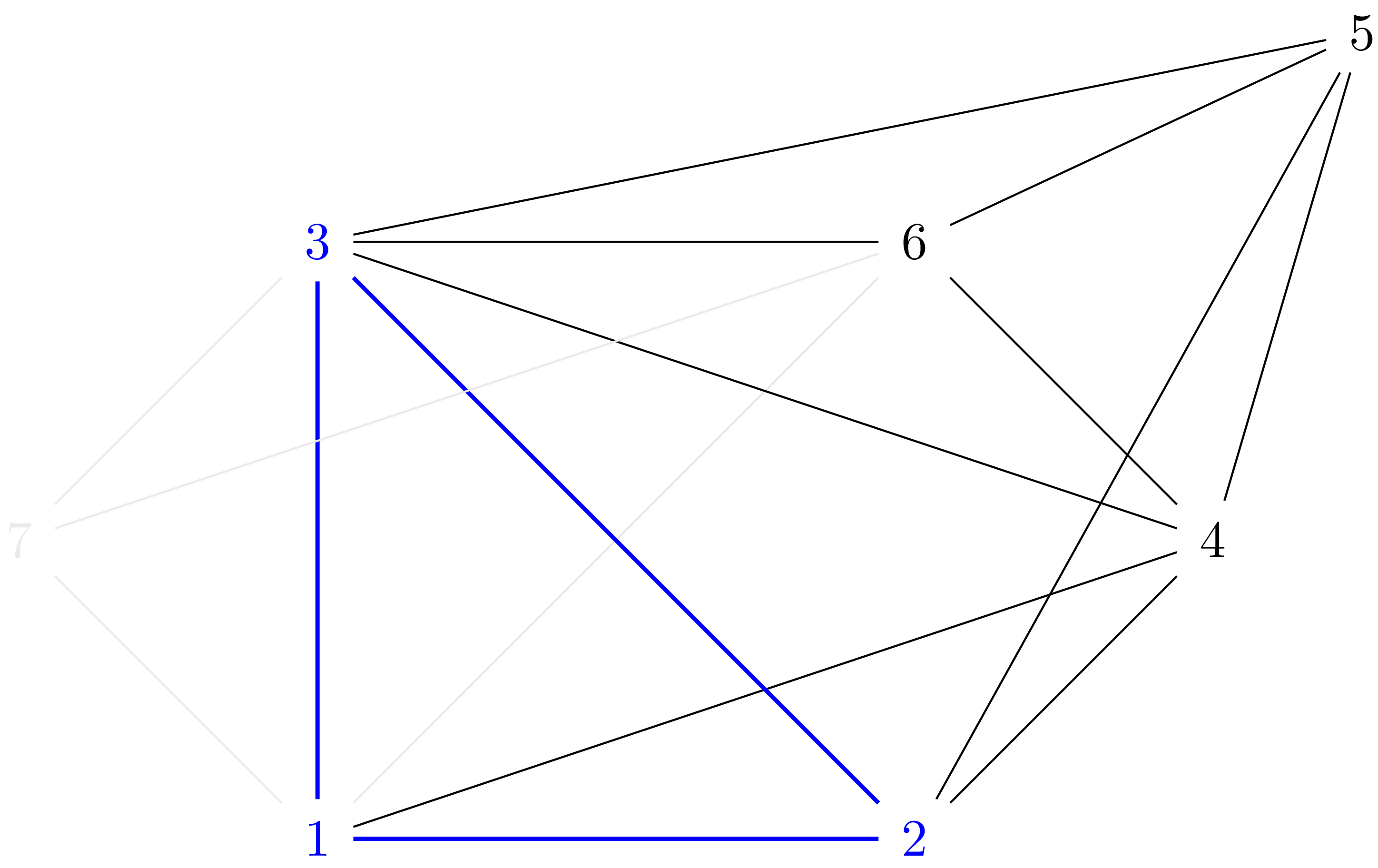
Combinatorial iterative approach

$K = 2$; *if* $\exists$ *vertex order s.t.* *next vertex* *has enough* **adjacent predecessors** *:*



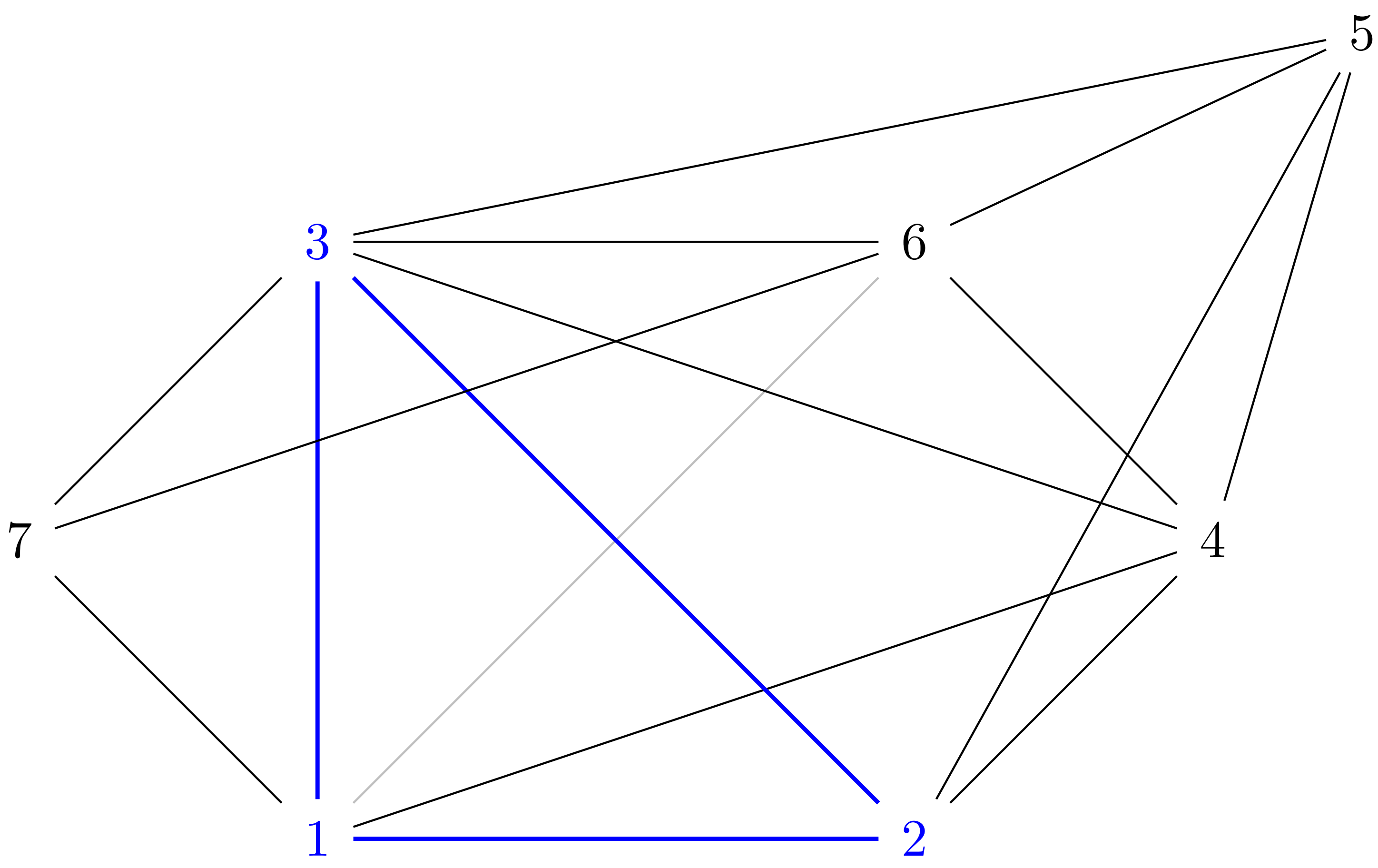
Combinatorial iterative approach

$K = 2$; *if* $\exists$ *vertex order s.t.* *next vertex* *has enough* **adjacent predecessors** *:*



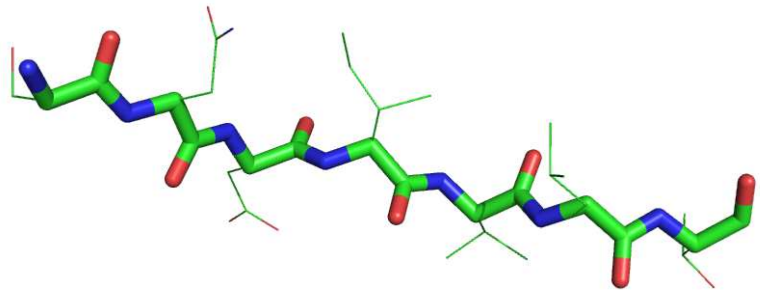
Combinatorial iterative approach

$K = 2$; *if* $\exists$ *vertex order s.t.* *next vertex* *has enough* **adjacent predecessors** *:*

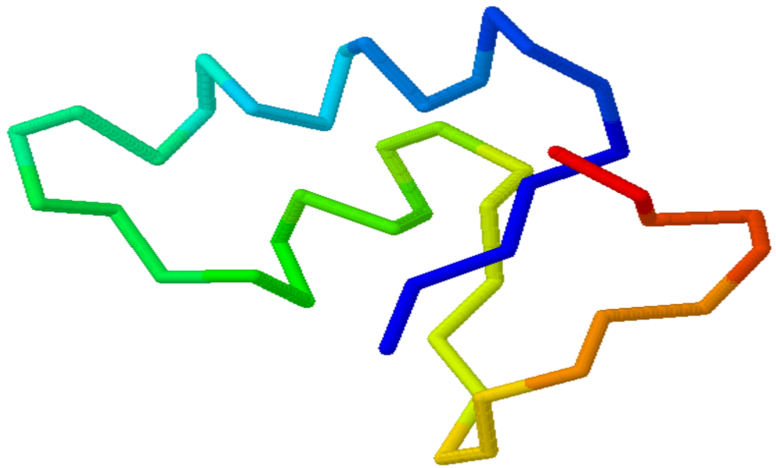


Proteins

- Proteins: *backbone* and *side chains*

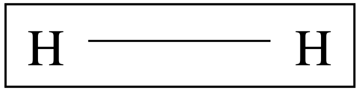


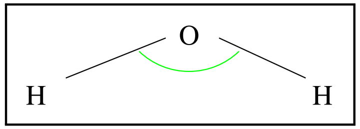
- Backbone: total order $<$ on a set V of atoms

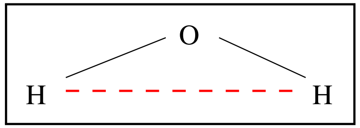


- Decompose the problem: embed the backbone, then plug the side chains in

Protein distances

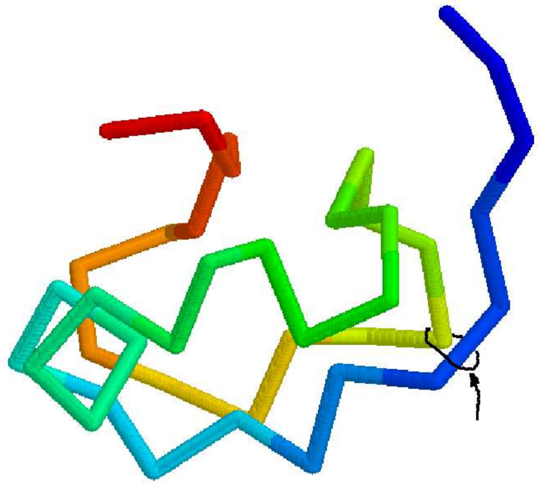
- Covalent bond distances $d_{v-1,v}$ are known 

- Angles between covalent bonds are known 

- $\Rightarrow d_{v-2,v}$ is known for all $v > 3$ 

- Distances $d_{v-3,v}$ are always < 6Å, so they can be measured using NMR techniques
- NMR might give other distances too

Atoms may be distant order-wise but closer than 6Å in space



Discretizable MDGP

- Protein backbones: 3 consecutive predecessors in 3D
- Weaken the condition $\geq K+1$ adjacent predecessors in $\mathbb{R}^K$ to:

$\geq K$ consecutive adjacent predecessors in $\mathbb{R}^K$

- DMDGP: given $x_1, \ldots, x_K \in \mathbb{R}^3$, and a vertex order as above, find $x_{K+1}, \ldots, x_n$ satisfying

$$\forall \{u, v\} \in E \ \|x_u - x_v\| = d_{uv}$$

- An **NP**-hard problem

Can we adapt the iterative method?

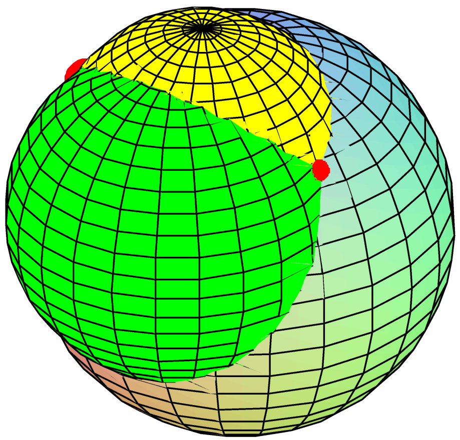
Sphere intersection

For given $v > 3$,

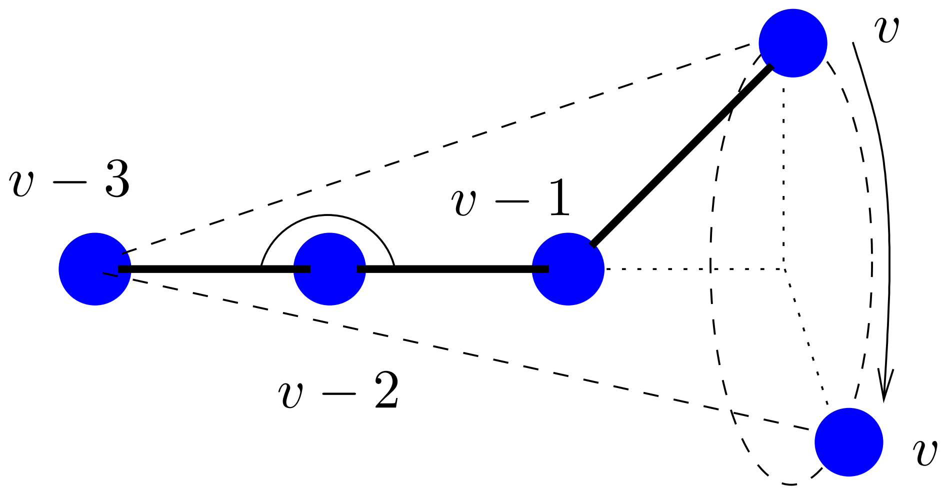
- $x_{v-3}, x_{v-2}, x_{v-1}$ are known
- $d_{v,v-1}, d_{v,v-2}, d_{v,v-3}$ are known

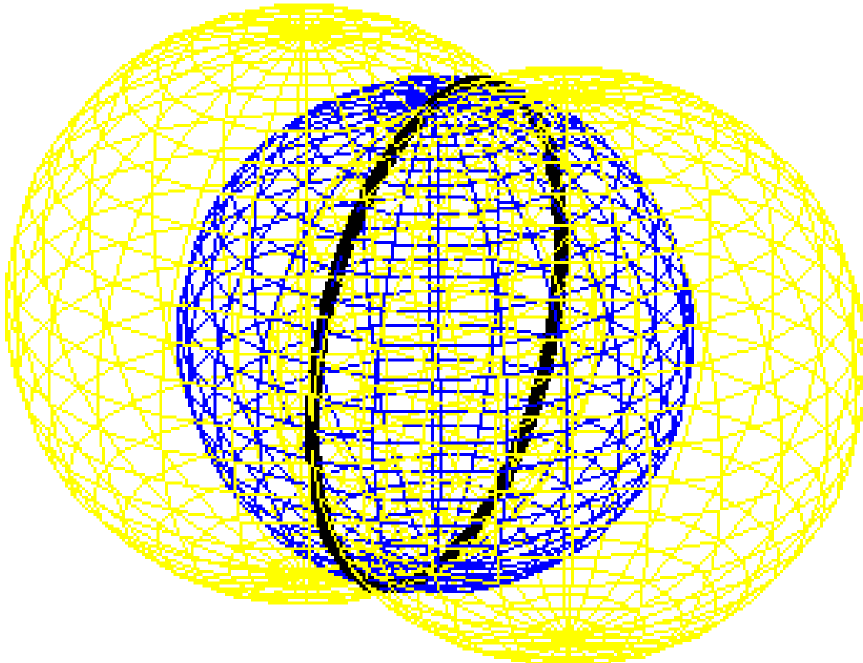
find x_v

Non-empty intersection of K spheres in $\mathbb{R}^K$ contains 2 points in general



Failure: collinearity





Probability 1

- We can develop a theory "modulo collinearity"
- Set of (configurations of n points in $\mathbb{R}^K$): all $\mathbb{R}^K$
- Collinearity in general: all points obey an equation $g(x) = 0$
- $\{x \mid g(x) = 0\}$: lower-dimensional manifold in $\mathbb{R}^K$, volume in $\mathbb{R}^K$ is 0
- Probability of sampling collinear embedding x: 0
- Results holding "with probability 1" $\equiv$ apart from a set of cases having volume 0 in the set of all possible cases

Finding the 2 points ($K = 3$)

Given $U = \{1, 2, 3\} \subseteq V$ and a partial embedding $x_1, x_2, x_3 \in \mathbb{R}^3$

1. Consider v adjacent to all $u \in U$
2. Extend x to v by solving a linear system:

$$\boxed{\begin{array}{rcl} \|x_v - x_1\|^2 & = & d_{1v}^2 \\ \|x_v - x_2\|^2 & = & d_{2v}^2 \\ \|x_v - x_3\|^2 & = & d_{3v}^2 \end{array}} \Rightarrow \boxed{\begin{array}{rcl} \|x_v\|^2 - 2x_v \cdot x_1 + \|x_1\|^2 & = & d_{1v}^2 \text{ (7)} \\ \|x_v\|^2 - 2x_v \cdot x_2 + \|x_2\|^2 & = & d_{1v}^2 \text{ (8)} \\ \|x_v\|^2 - 2x_v \cdot x_3 + \|x_3\|^2 & = & d_{1v}^2 \text{ (9)} \end{array}}$$

$$\boxed{\begin{array}{l} \text{(9)-(7)} \\ \text{(9)-(8)} \end{array}} \Rightarrow \boxed{\begin{pmatrix} 2(x_1 - x_3)^{\mathsf{T}} \\ 2(x_2 - x_3)^{\mathsf{T}} \end{pmatrix} x_v = \begin{pmatrix} (\|x_1\|^2 - \|x_3\|^2) - (d_{1v}^2 - d_{3v}^2) \\ (\|x_2\|^2 - \|x_3\|^2) - (d_{2v}^2 - d_{3v}^2) \end{pmatrix}}$$

3. Diagonalize the 2×3 linear system (one pivot)
4. Express x_{v1}, x_{v2} in function of x_{v3} linearly $\qquad (*)$
5. Replace x_{v3} in Eq. (9), solve quadratic in x_{v3}
6. Obtain two values for x_{v3}, use $(*)$ to find two points for x_v

Branch-and-Prune

v: rank of current atom $x_{<v}$: partial embedding to rank $v-1$
G: instance X: current pool of embeddings
$S(y,r)$: $\mathbb{R}^K$ sphere centered at y with radius r

BranchAndPrune(v, $x_{<v}$, G, X):

Let $\mathcal{S} \leftarrow \bigcap_{i\in\{1,\ldots,K\}} S(x_{v-i}, d_{v-i,v})$ =($\{s_1, s_2\}$ or $\varnothing$)

for $s \in \mathcal{S}$ **do**

Extend current embedding to $x = (x_{<v}, s)$

if $\forall u \in$ AdjPred(v) $\|x_u - x_v\| = d_{uv}$ **then**

if ($v = n$) **then**

Let $X \leftarrow X \cup \{x\}$

else

BranchAndPrune($v+1$, x, G, X)

end if

end if

end for

BP properties

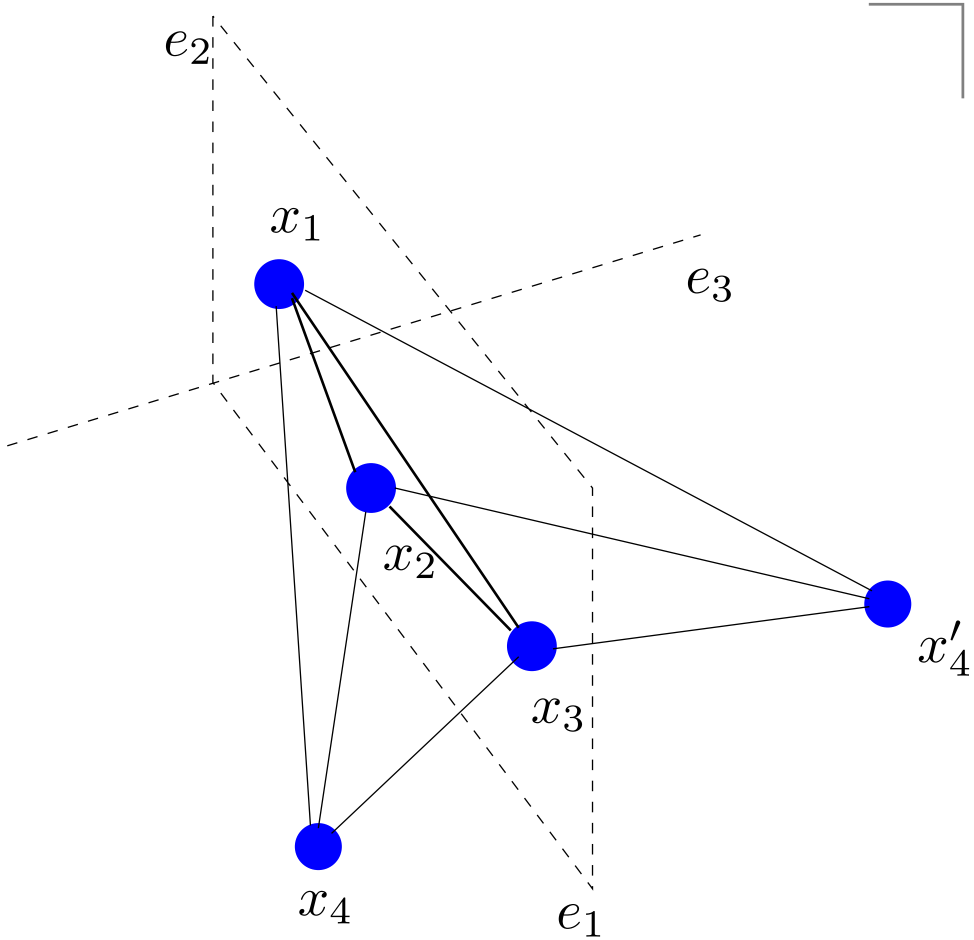
- BP: worst-case exponential time
- With probability 1, find all incongruent embeddings of G extending initial partial embedding
- Performs very efficiently (speed and accuracy)

  Embed 10,000 vertices in a 13 seconds of CPU time
- Two empirical observations:
  1. **the number of solutions it finds is always a power of two**
  2. $|V|$ **versus CPU time plots are always linear-like for PDB**

Symmetry

BP root node symmetry

- x_4' is a reflection of x_4 w.r.t. the plane defined by x_1, x_2, x_3
- $\Rightarrow$ BP tree symmetric below level 3
- Start branching from level 4, not 3



Number of solutions

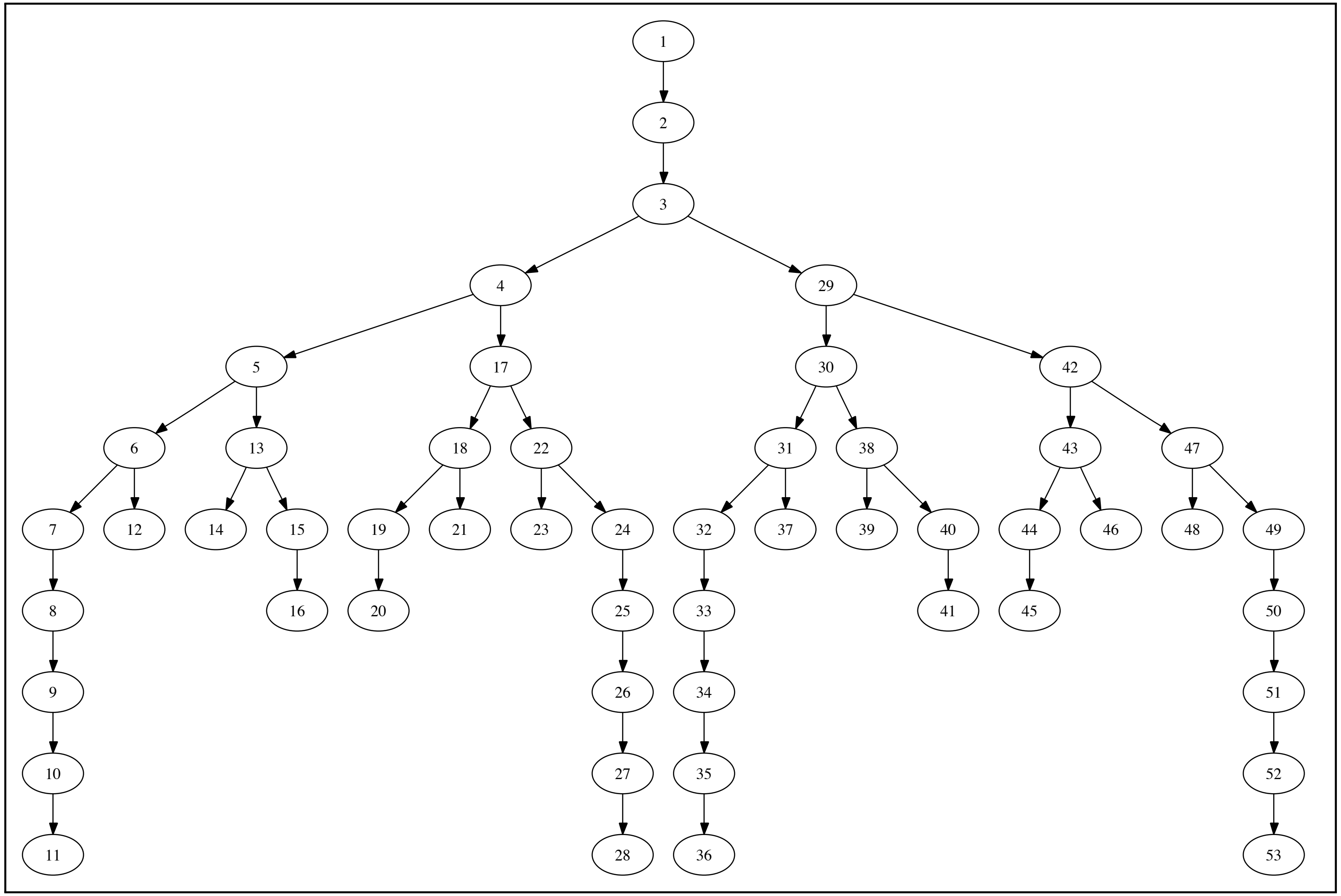
Instance	$\|X\|$
mmorewu-2	2
mmorewu-3	2
mmorewu-4	4
mmorewu-5	4
mmorewu-6	4
lavor10_0	4
lavor15_0	16
lavor20_0	8
lavor25_0	8
lavor30_0	2
lavor35_0	64
lavor40_0	2
lavor45_0	2
lavor50_0	4096
lavor55_0	64
lavor60_0	64

Instance	$\|X\|$
1brv	2
1aqr	4
2erl	2
1crn	2
1ahl	16
1ptq	2
1brz	4
1hoe	2
1lfb	2
1pht	2
1jk2	2
1f39a	2
1acz	8
1poa	2
1fs3	2
1mbn	2
1rgs	2
1m40	2
1bpm	2
1n4w	2
1mqq	2
1rwh	2
3b34	2
2e7z	2
1epw	2

For all tested DMDGP instances, $\exists \ell \in \mathbb{N}$ such that $|X| = 2^\ell$

A BP search tree example

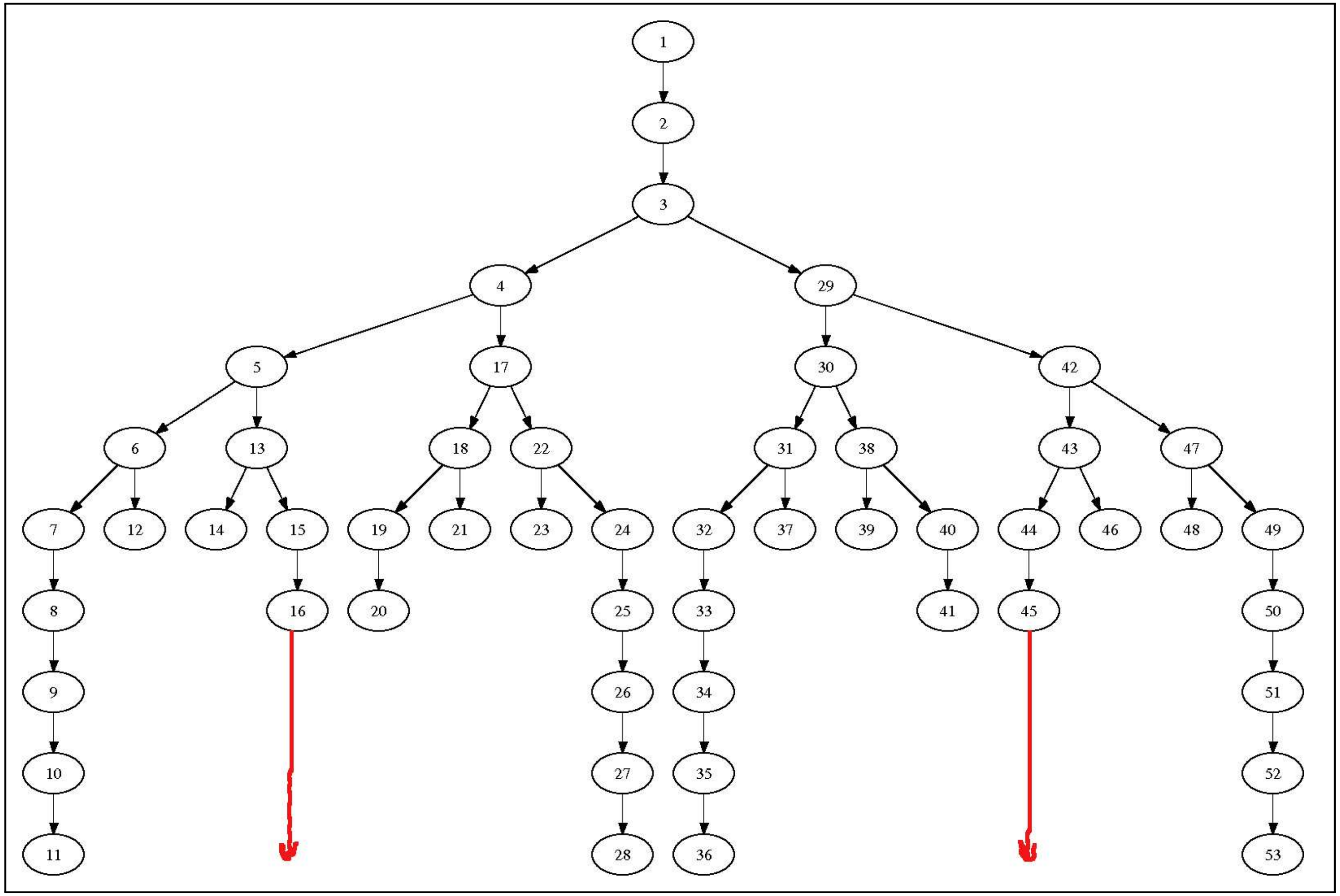
- Typical BP search tree (embeddings = paths root→leaves)



- Root node symmetry: $|X|$ is even
- No evident reason why $|X|$ should be a power of two

A BP search tree example

- Typical BP search tree (embeddings = paths root→leaves)

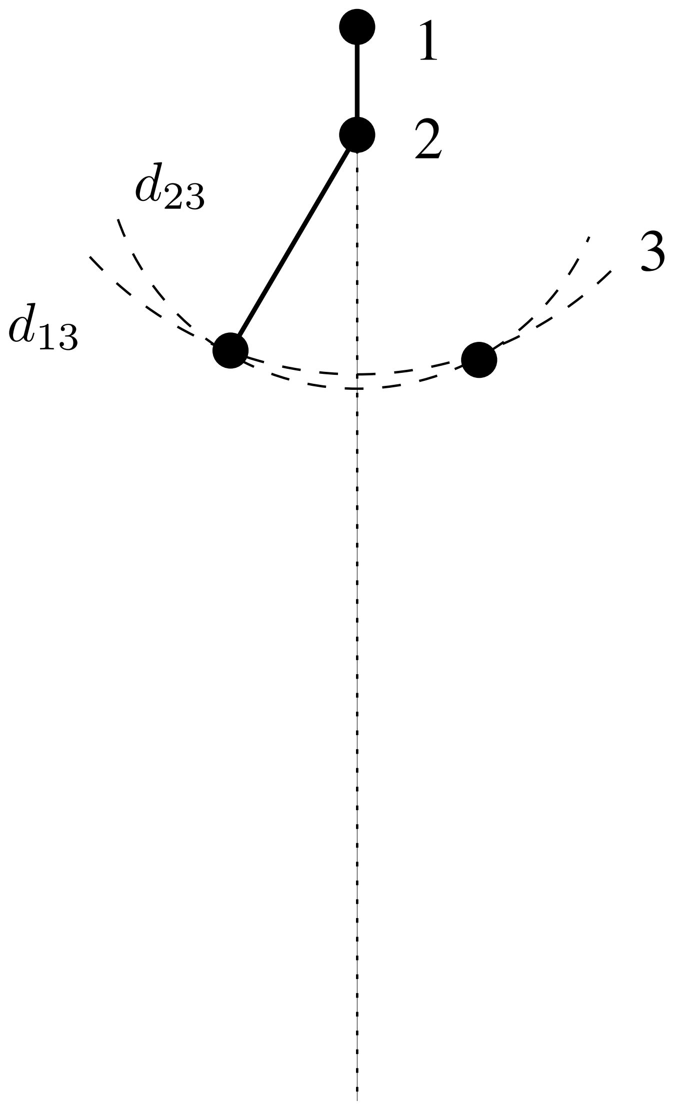


- Root node symmetry: $|X|$ is even
- No evident reason why $|X|$ should be a power of two (why not symmetric paths to level $|V|$ from nodes 16 and 45?)

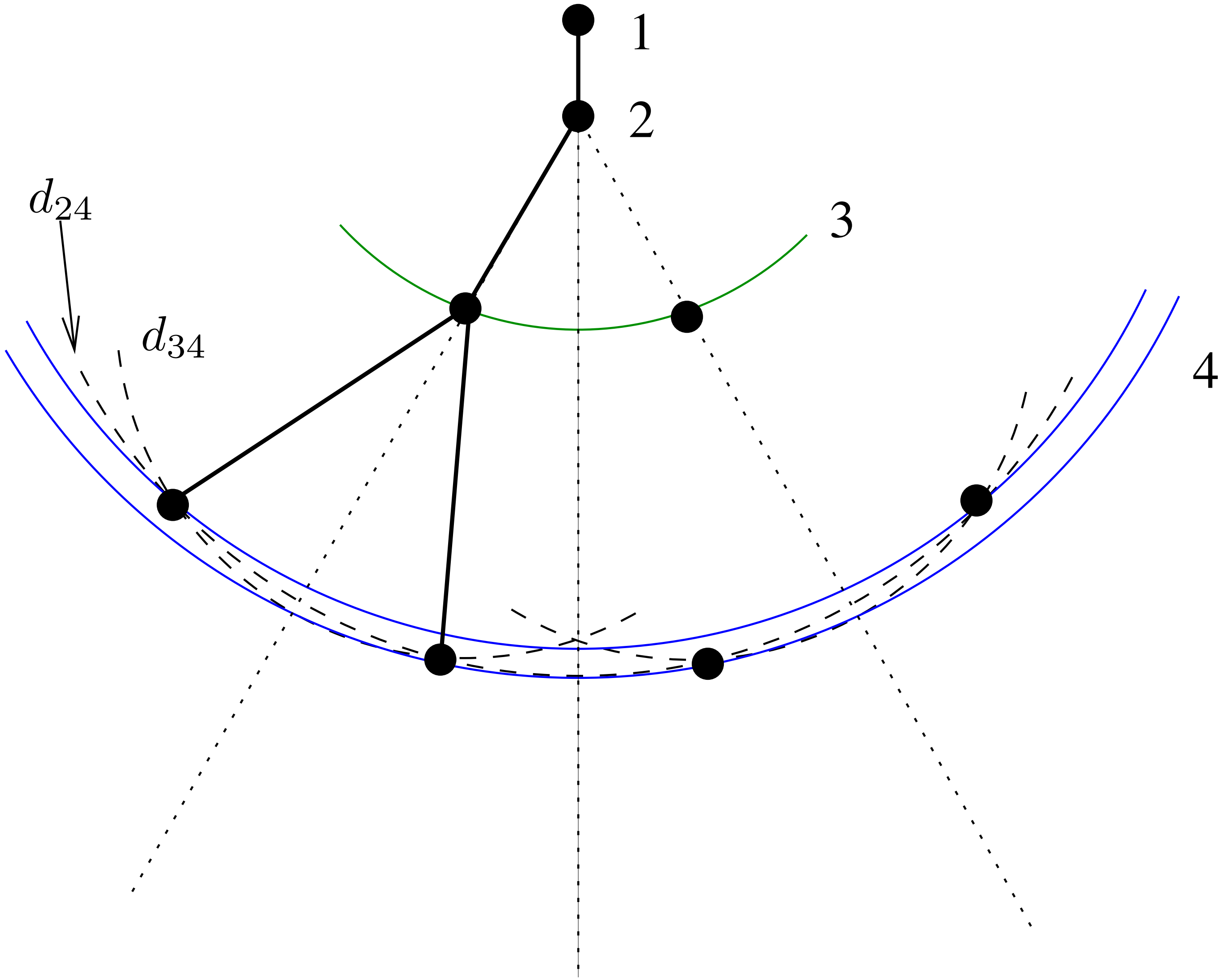
Discretization/pruning distances

- Let $E_D = \{\{u,v\} \mid |u-v| \le K\}$ and $E_P = E \smallsetminus E_D$
- E_D: **discretization distances**
  - they guarantee that the instance is a DMDGP
  - they allow the construction of the complete BP tree
  - this tree has $2^{|V|-3}$ leaves, $2^{|V|-4}$ if we consider root node symmetry
- E_P: **pruning distances**
  - they allow pruning of the BP tree
  - not clear why they should prune branches symmetrically

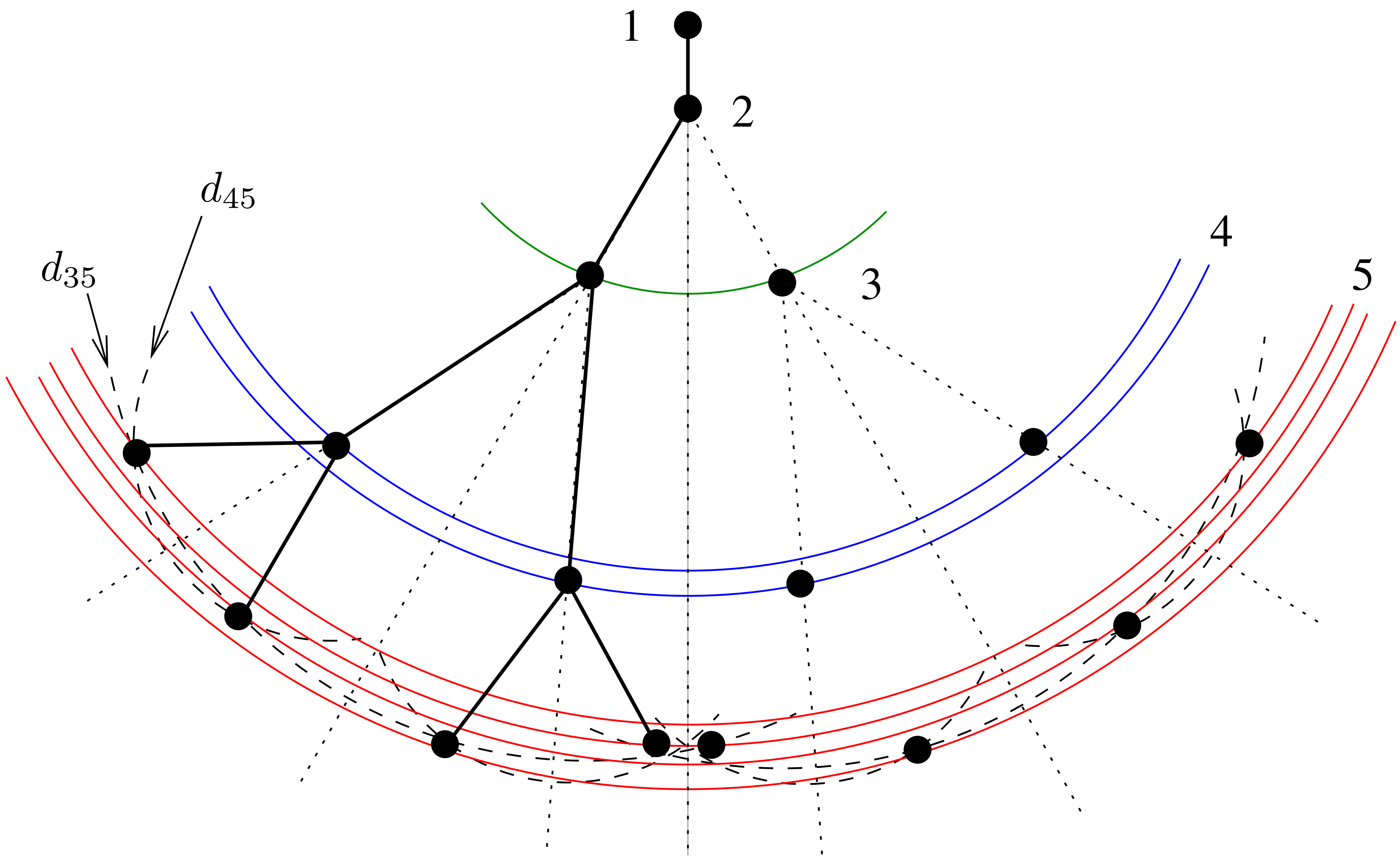
Structure of the BP tree ($\mathbb{R}^2$)



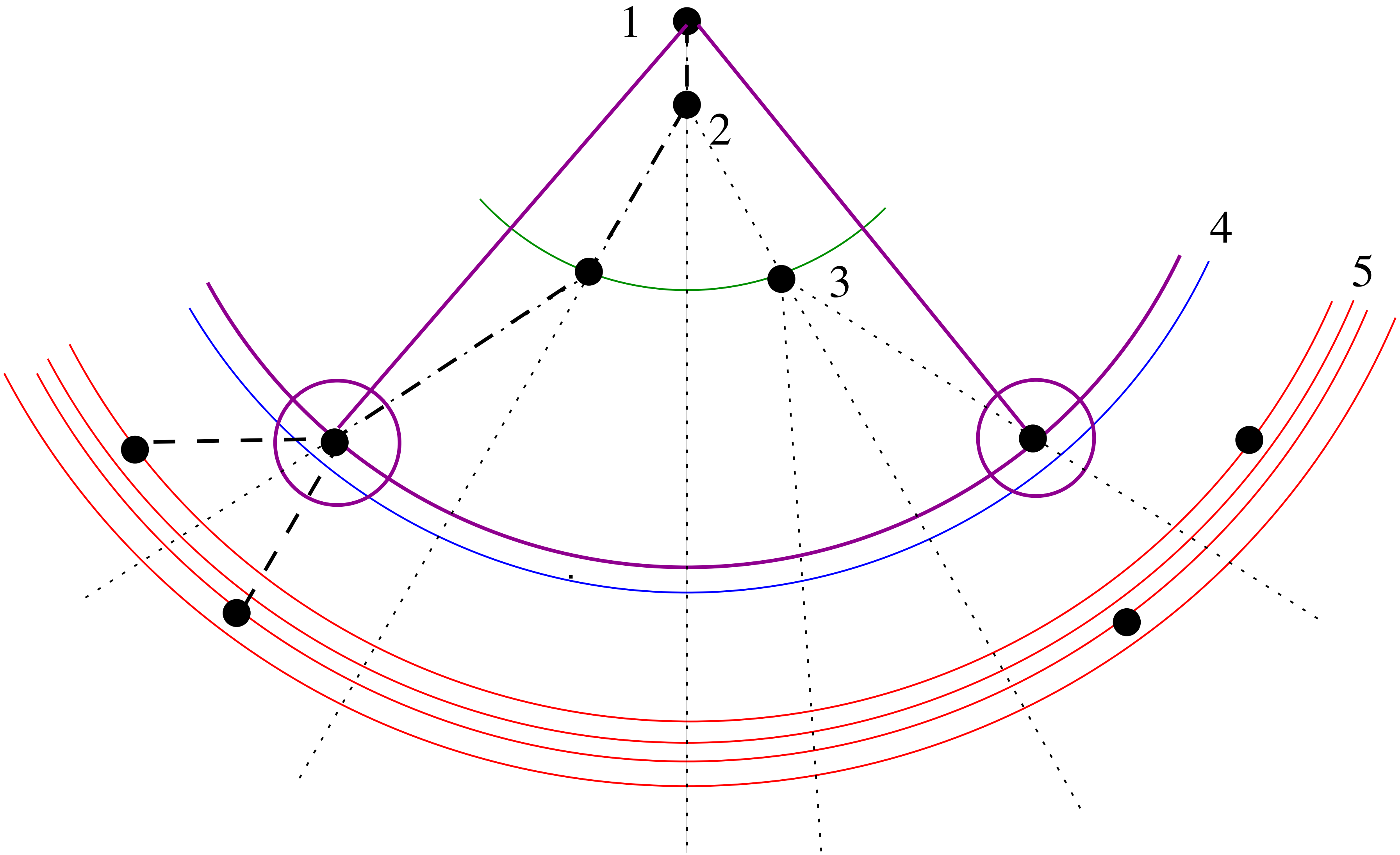
Structure of the BP tree ($\mathbb{R}^2$)



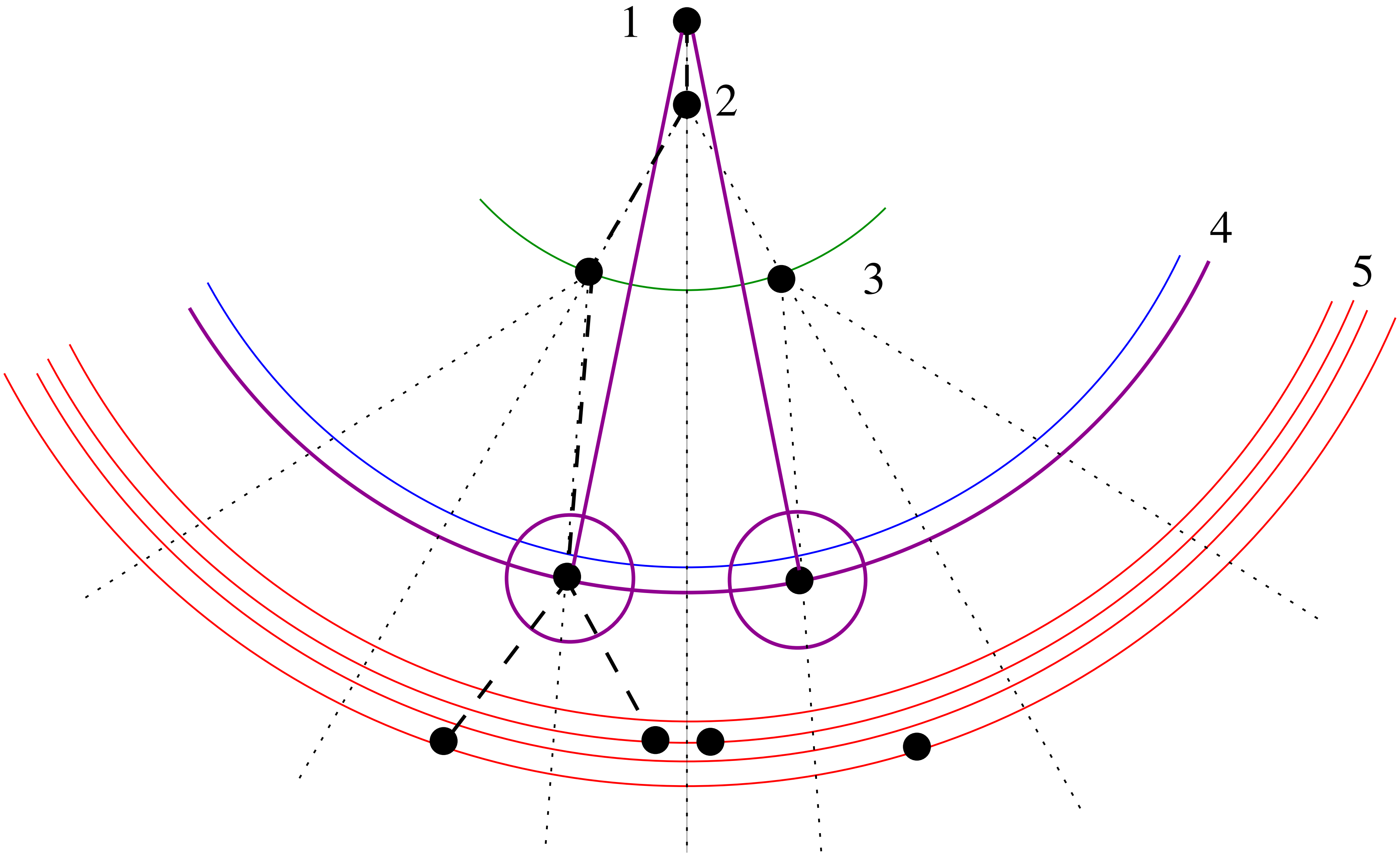
Structure of the BP tree ($\mathbb{R}^2$)



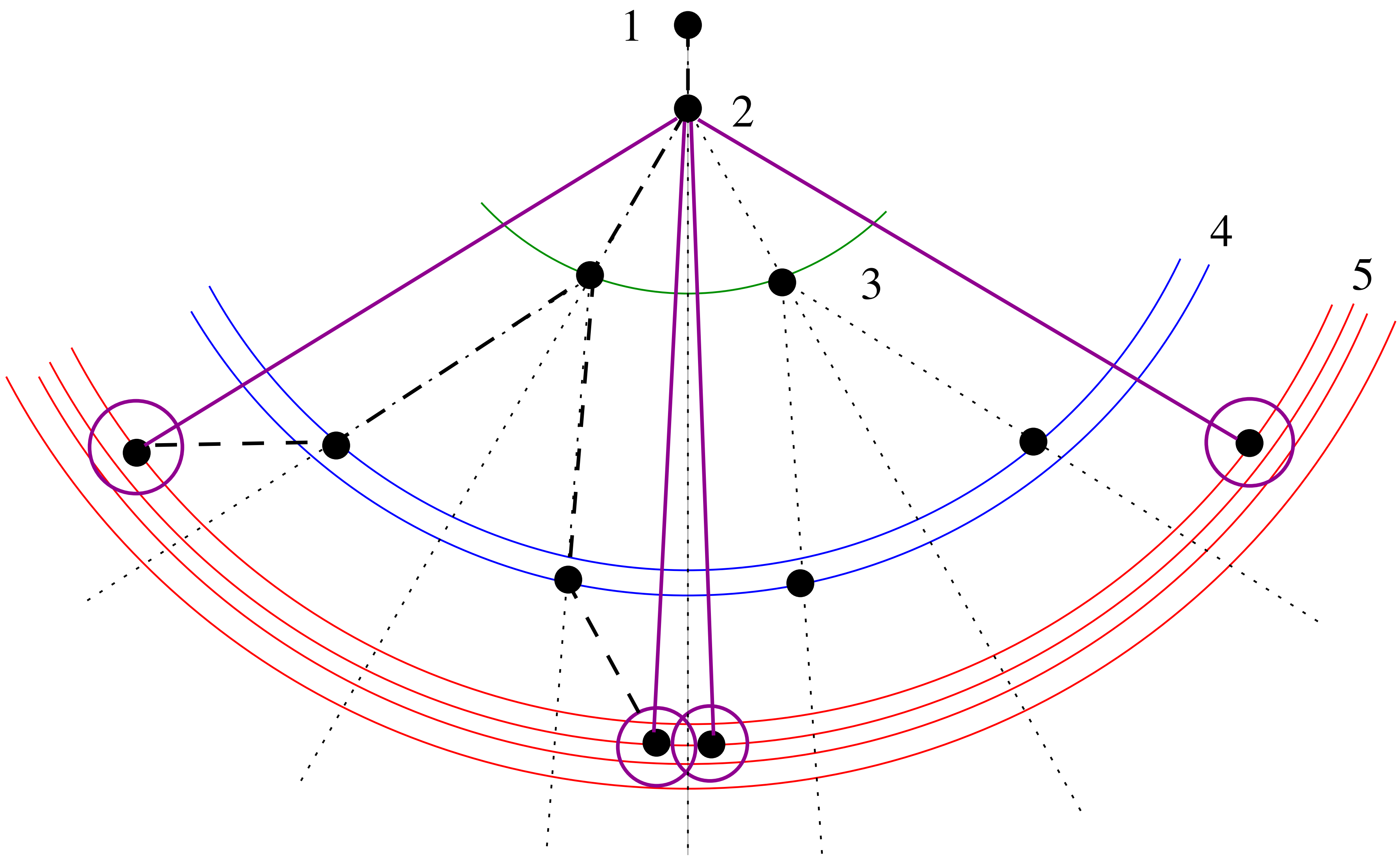
Effect of pruning distance d_{14}



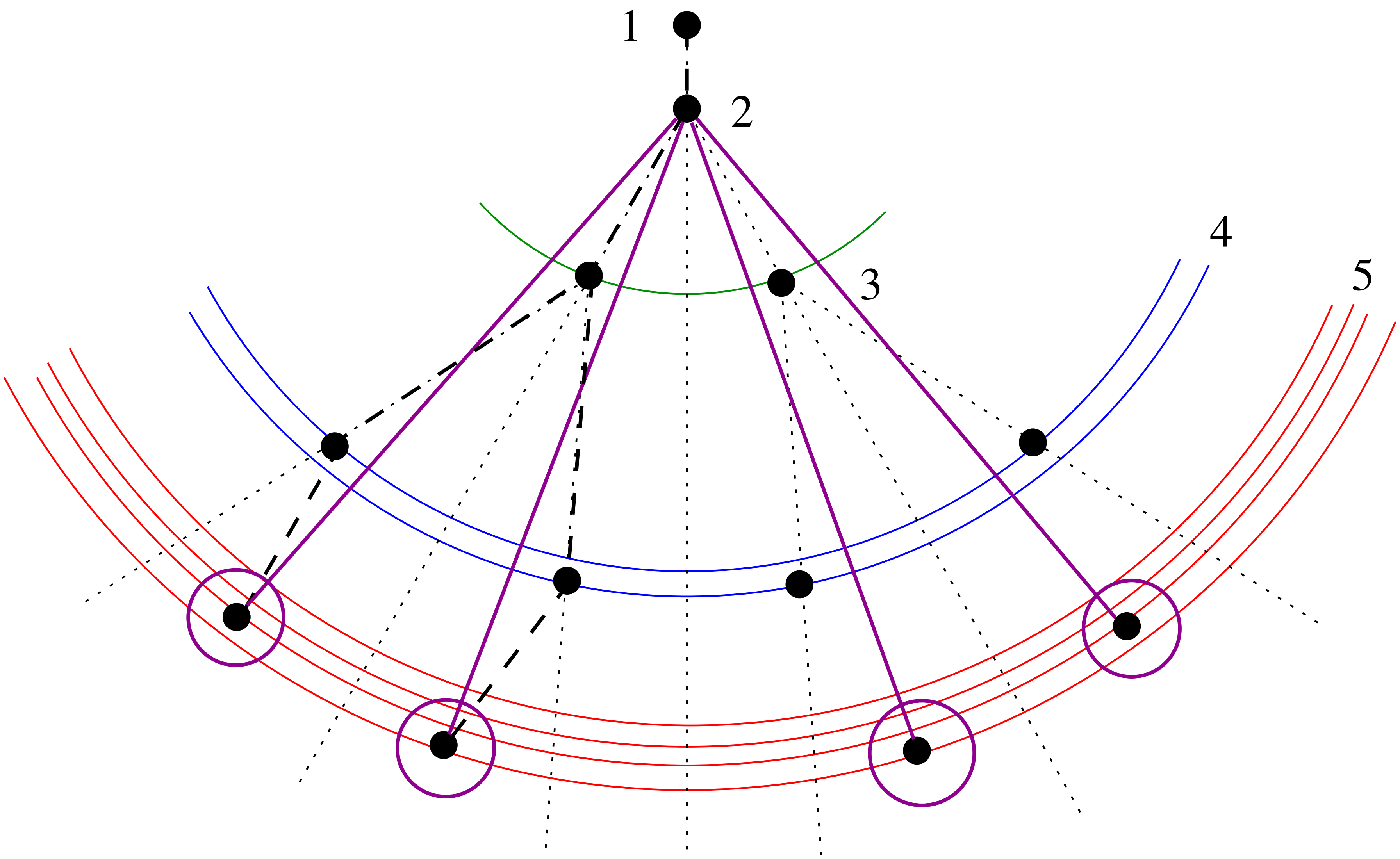
Effect of pruning distance d_{14}



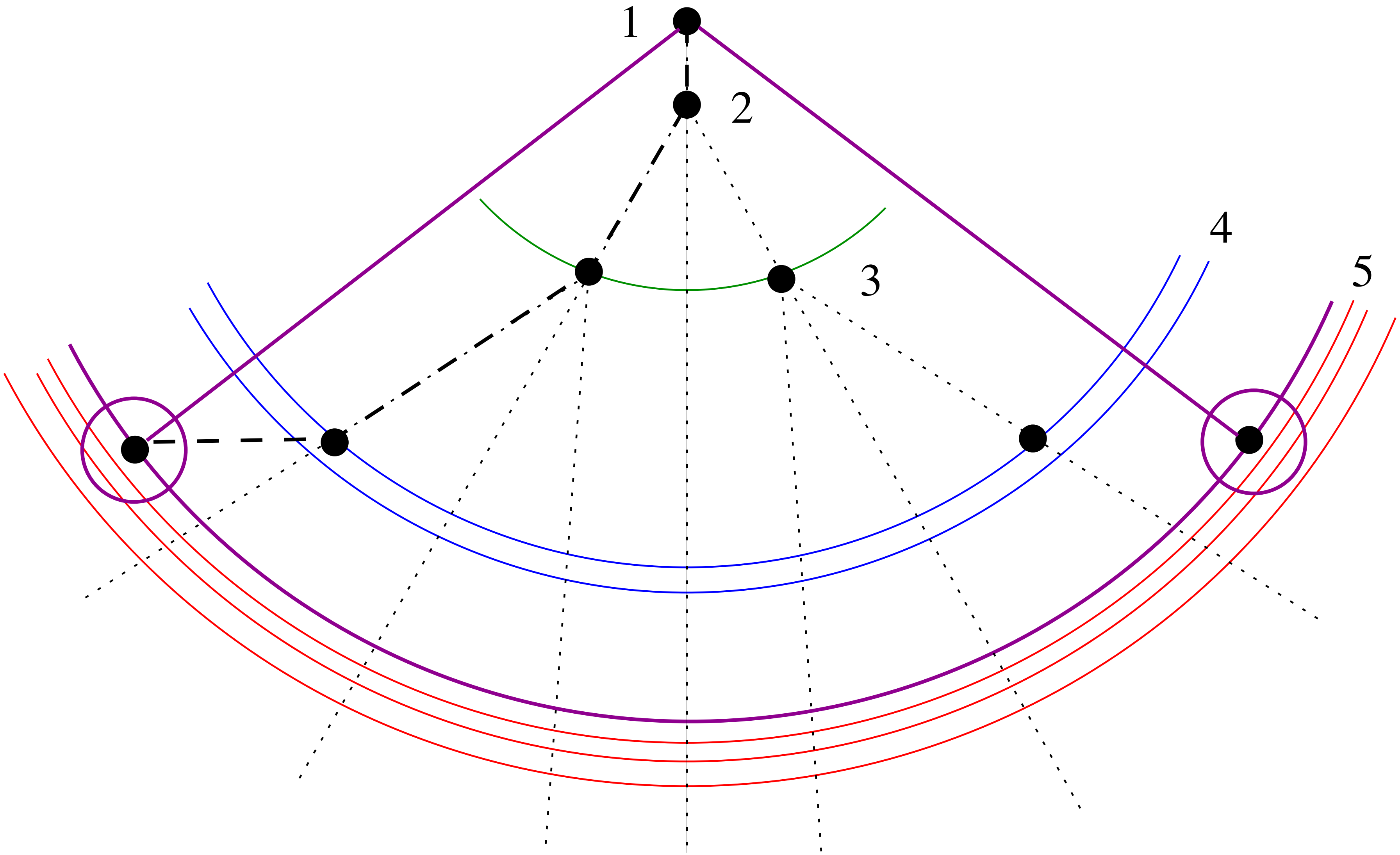
Effect of pruning distance d_{25}



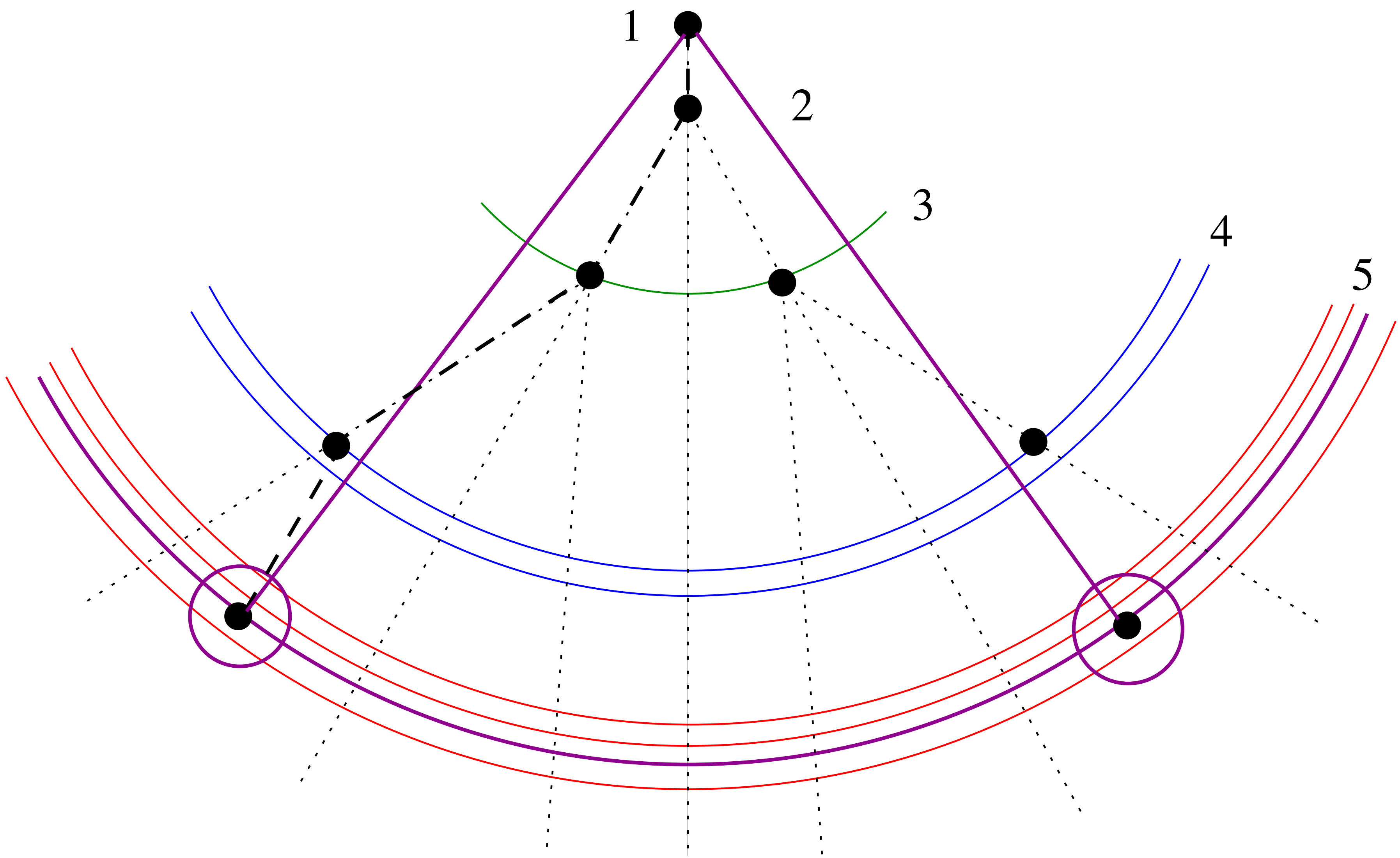
Effect of pruning distance d_{25}



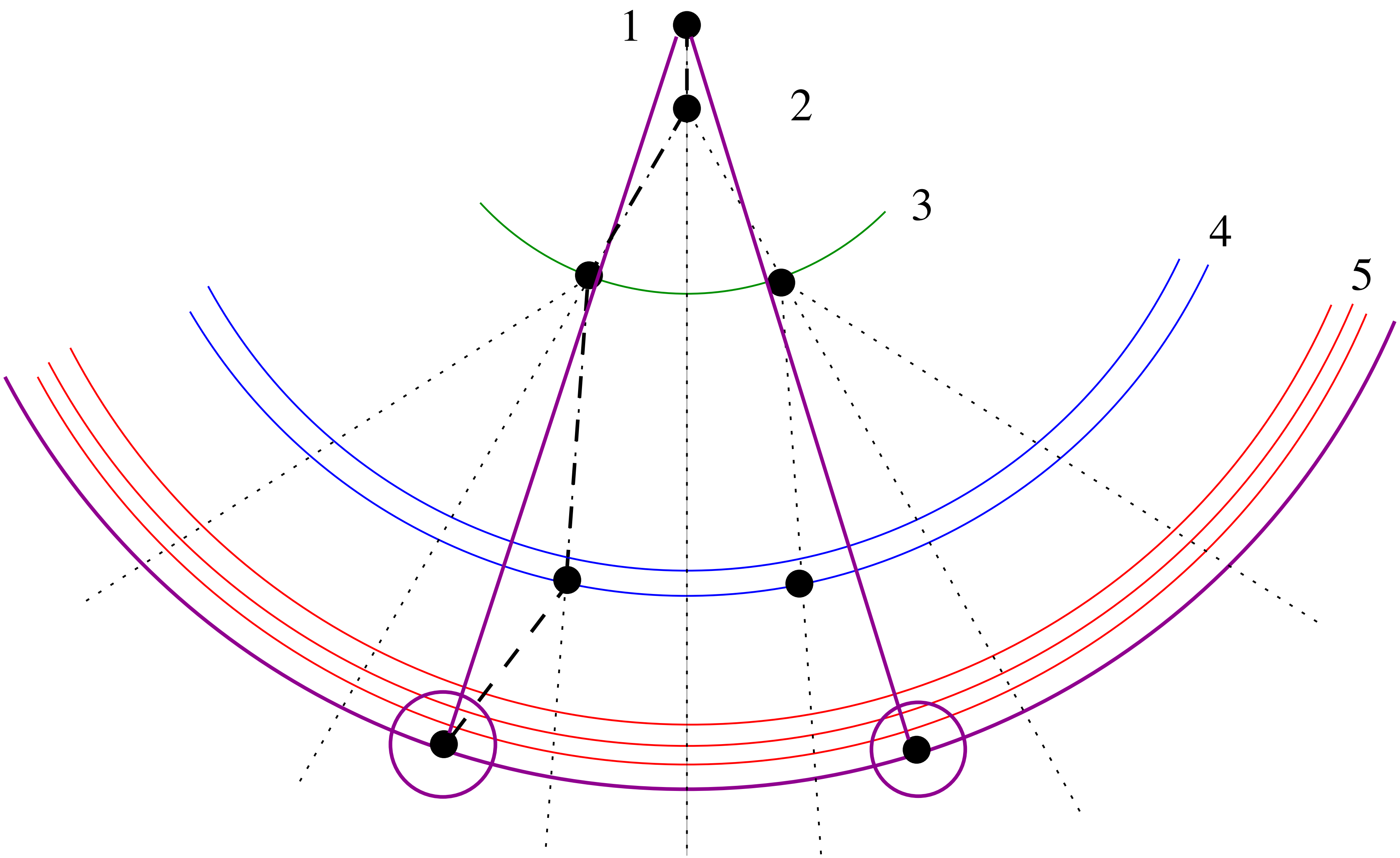
Effect of pruning distance d_{15}



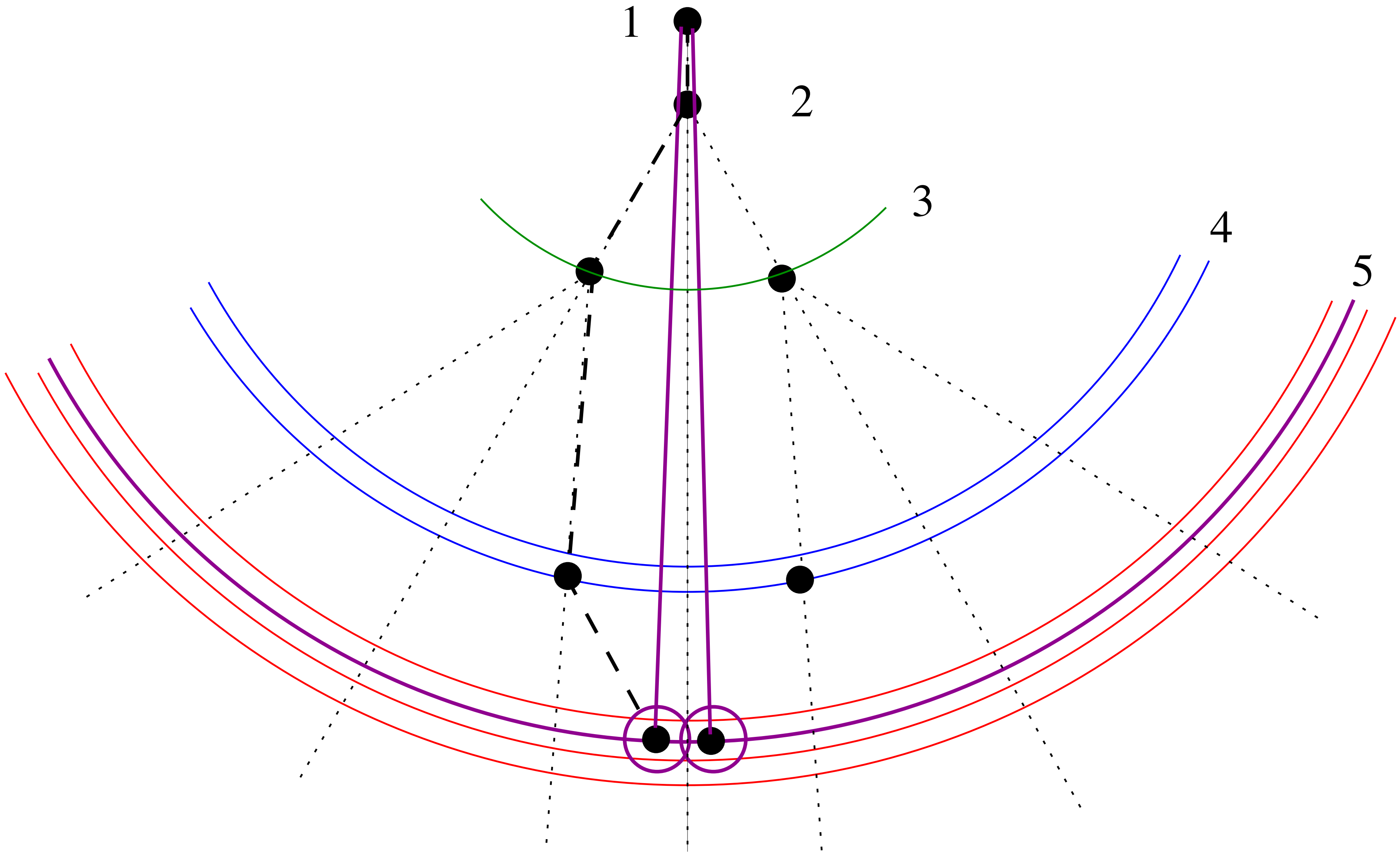
Effect of pruning distance d_{15}



Effect of pruning distance d_{15}

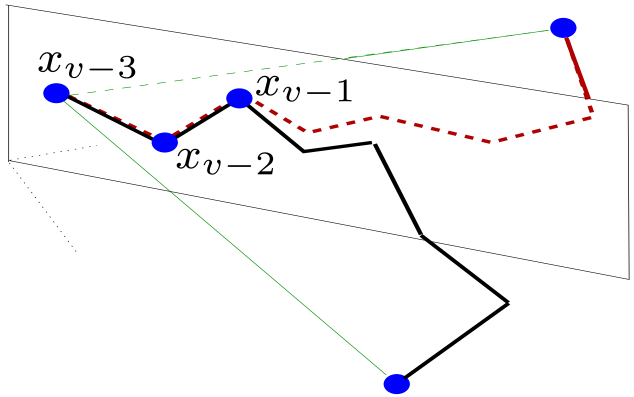


Effect of pruning distance d_{15}



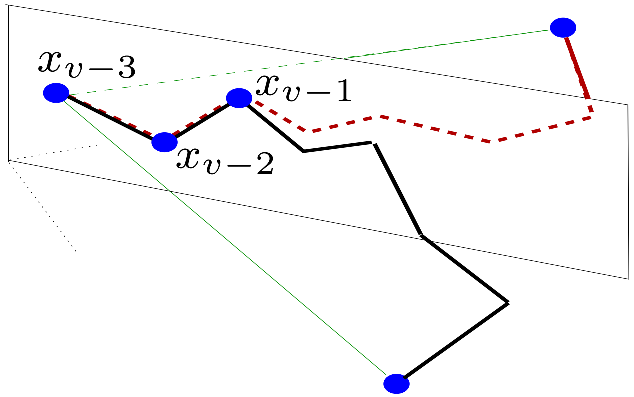
Symmetry by pruning distances

Given embedding x, $\quad R^v_x =$ reflection w.r.t. hyperplane $x_{v-K}, \ldots, x_{v-1}$



Symmetry by pruning distances

Given embedding x, R^v_x = reflection w.r.t. hyperplane $x_{v-K},\ldots,x_{v-1}$



Thm.

With prob. 1, for each $u, v \in V$ with $v > K, u < v - K$,

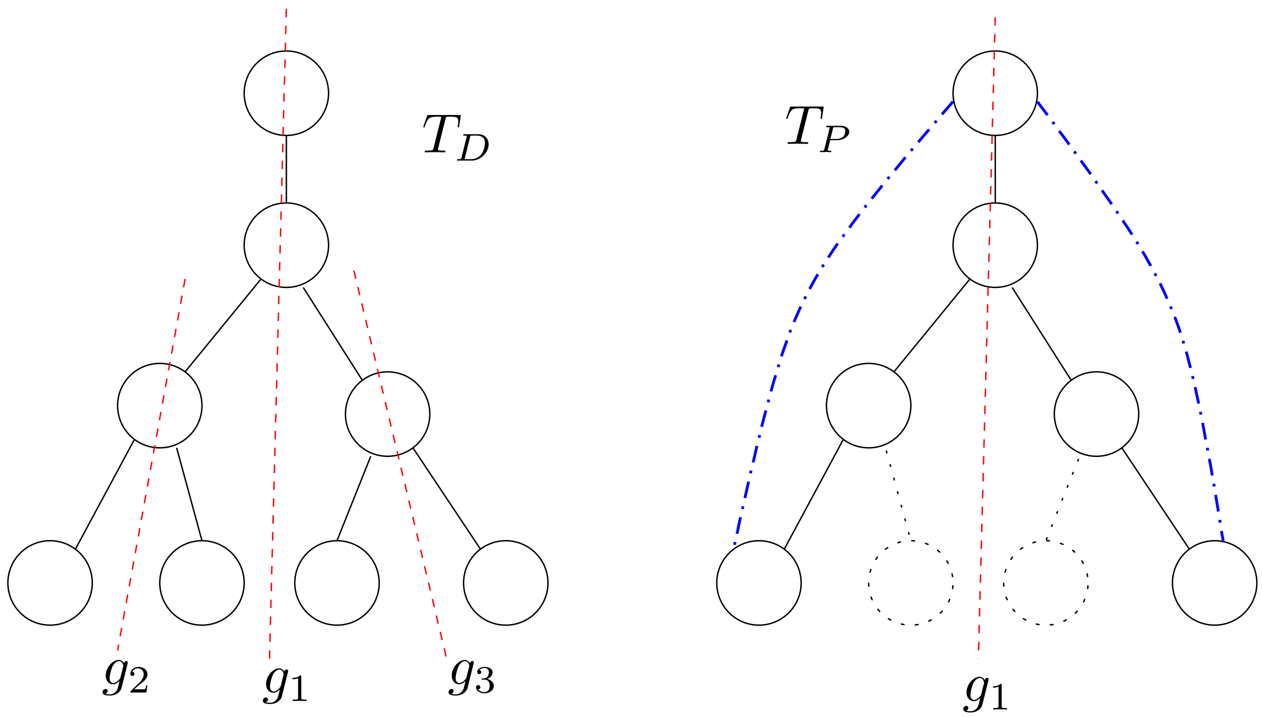
$$\forall x \neq x' \in X \quad \|x_u - x_v\| = \|x'_u - x'_v\| \Leftrightarrow x'_v = R^{u+K}_x(x_v)$$

Moreover, $\exists$ a finite set $H^{uv} \subseteq \mathbb{R}_+$ with $|H^{uv}| = 2^{v-u-K}$ s.t.

$$\forall x \in X \; (\underbrace{\|x_u - x_v\|}_{\text{plays the role of pruning dist.}} \in H^{uv})$$

Groups fixing the trees

- Let T_D be a full BP binary search tree
- Let T_P be the subtree of T_D representing only feasible branches
- Draw them so $T_P \subseteq T_P$
- Invariant group for T_D: all partial reflections (g_1, g_2, g_3)
- Invariant group for T_P: only some partial reflections (g_1)



Partial reflections

$$g_v(x) = (x_1, \ldots, x_{v-1}, R_x^v(x_v), \ldots, R_x^v(x_n))$$

Only reflect starting from vertex v

Discretization group

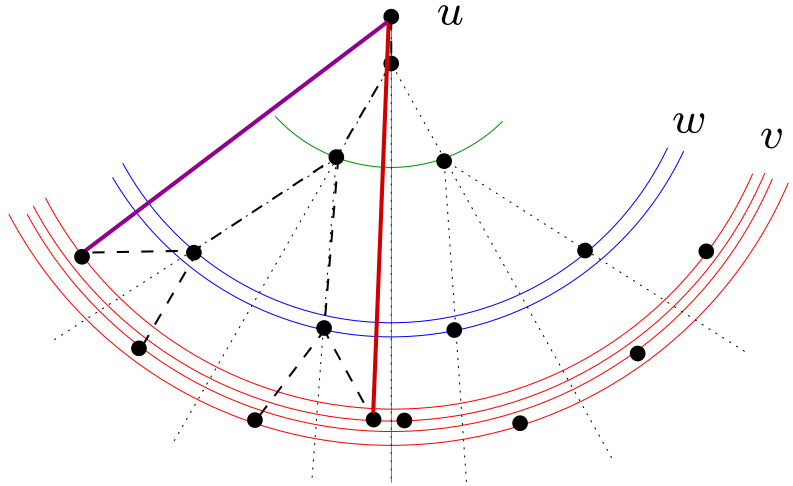
Group of partial reflections fixing the complete BP tree (no pruning distances)

- The following hold with probability 1 $\forall v > K$:
  1. g_v **is injective with probability 1 (by reflection)**
  2. $\boxed{g_v \text{ is idempotent}}$ **(by reflection)**
  3. $\forall u > K, u \neq v$, g_u **and** g_v **commute (nontrivial)**
- Thus, $\mathcal{G}_D = \langle g_v \mid v > K \rangle$ is an Abelian group under composition $\Rightarrow$ $\boxed{\text{isomorphic to } C_2^{n-K}}$)
- By previous thm, discretization distances are invariant under $\mathcal{G}_D$
- The action of $\mathcal{G}_D$ on X is transitive, i.e. $\forall x, x' \in X \exists g \in \mathcal{G}_D \ (x' = g(x))$
- This action has only one orbit, i.e. $X = \mathcal{G}_D x$

Pruning group

Group of partial reflections fixing the actual BP tree (with pruning distances)

- Assume DMDGP instance is YES, consider $\{u, v\} \in E_P$
- With probability 1, $d_{uv} \in H^{uv}$ (otherwise the instance would be NO)
- Notice $d_{uv} = \|x_v - x_u\| \neq \|g_w(x)_v - g_w(x)_u\|$ for all $w \in \{u + K + 1, \ldots, v\}$



- In order to keep invariance we remove such g_w's from the group
- **Pruning group**: $\mathcal{G}_P = \langle g_w \mid w > K \wedge \forall \{u, v\} \in E_P \; (w \notin \{u + K + 1, \ldots, v\}) \rangle$
- $\boxed{\mathcal{G}_P \leq \mathcal{G}_D}$ and all distances are invariant w.r.t. the pruning group
- Again, $\boxed{\text{action of } \mathcal{G}_P \text{ on } X \text{ is transitive}}$ (nontrivial proof)

Power of two

Thm.

$\exists \ell \in \mathbb{N} \; (|X| = 2^\ell)$

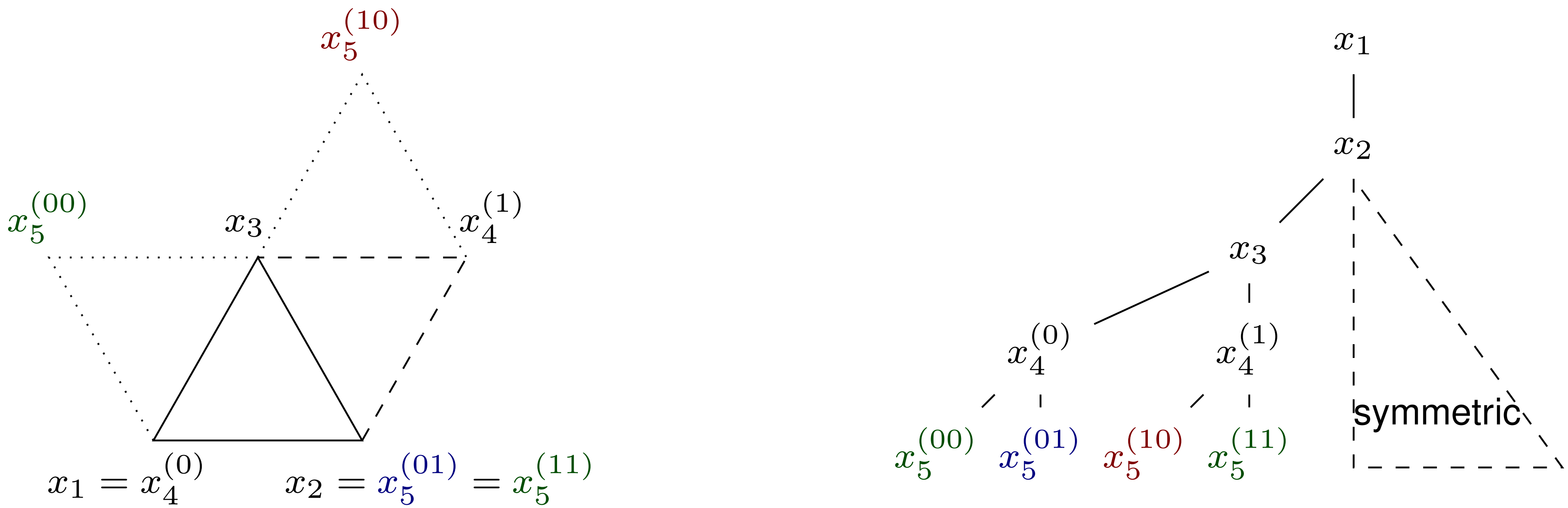
Proof

With probability 1:

- $\mathcal{G}_D \cong C_2^{n-K} \Rightarrow |\mathcal{G}_D| = 2^{n-K}$
- $\mathcal{G}_P \leq \mathcal{G}_D \Rightarrow |\mathcal{G}_P| \mid |\mathcal{G}_D| \Rightarrow \exists \ell \in \mathbb{N} \; |\mathcal{G}_P| = 2^\ell$
- Action of $\mathcal{G}_P$ on X is transitive $\Rightarrow \mathcal{G}_P x = X$
- Idempotency $\Rightarrow$ for $g, g' \in \mathcal{G}_P$, if $gx = g'x$ then $g = g' \Rightarrow |\mathcal{G}_P x| = |\mathcal{G}_P|$
- Thus, $|X| = |\mathcal{G}_P x| = |\mathcal{G}_P| = 2^\ell$

Why the “probability 1”?

- Not all “YES” DMDGP instances have $|X| = 2^\ell$
- But the set of such instances (with real data) has Lebesgue measure zero in the set of all DMDGP instances



Happens when > 1 vertices are embedded in the same position

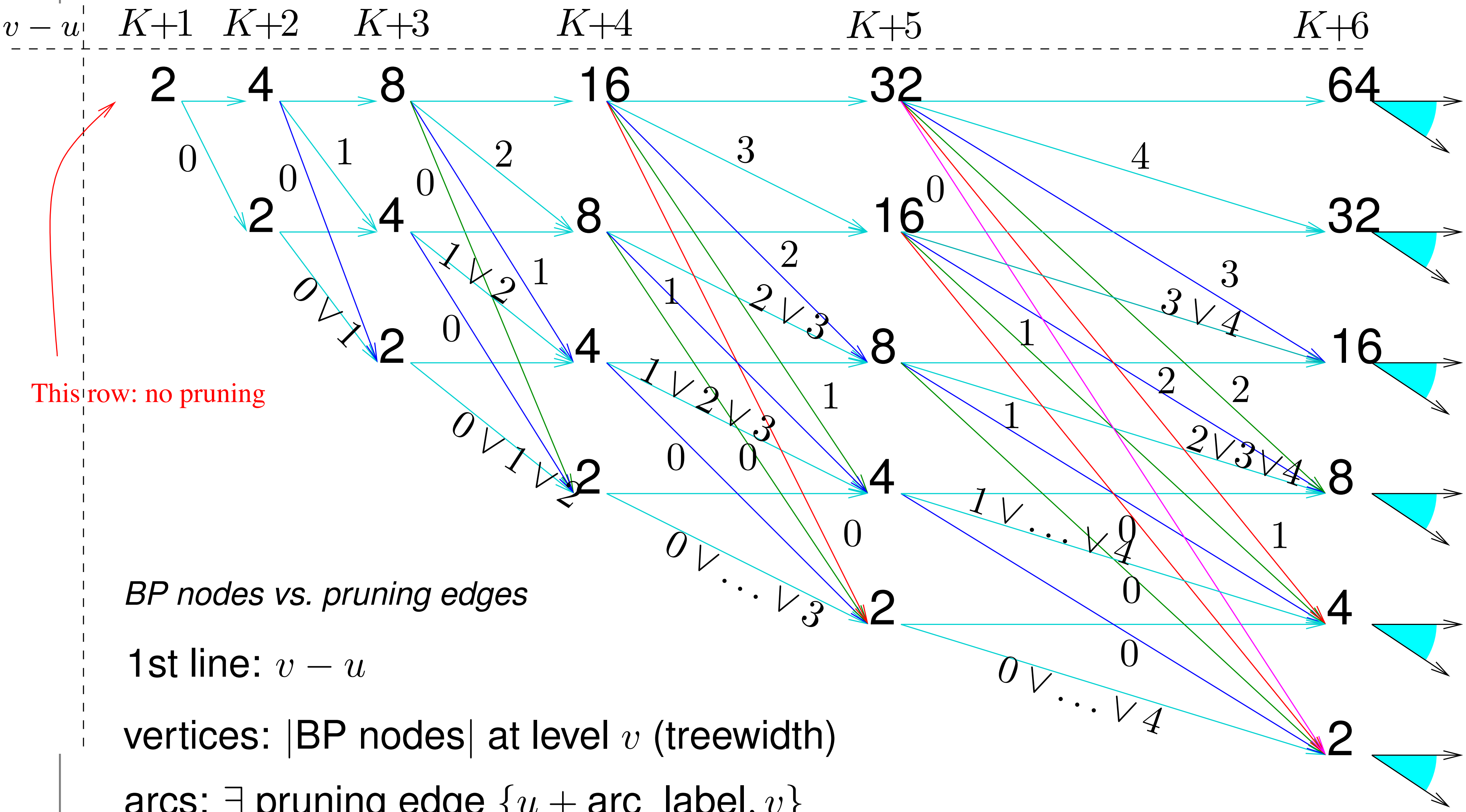
$x_5^{(01)}$ **should** be infeasible, but $x_5^{(01)} = x_5^{(11)}$ (event with prob. 0)

FPT behaviour

A polynomial BP?

- Empirically: never an exponential-time increase behaviour in our experiments (instances generated from PDB files)
- Embed 10000-atom protein backbones in 10-15s on one core
- Easy to show that BP has worst-case exponential complexity
- Are proteins a polynomial case of the DMDGP?
- Complexity depends on BP nodes; since height$\leq |V|$, only need to consider treewidth
- A pruning edge $\{u, v\}$ with $u < v - K$ reduces the number of nodes at level v from 2^{v-K} to $2^{v-K-(u-1)}$ (by symmetry)

BP subtree rooted at u



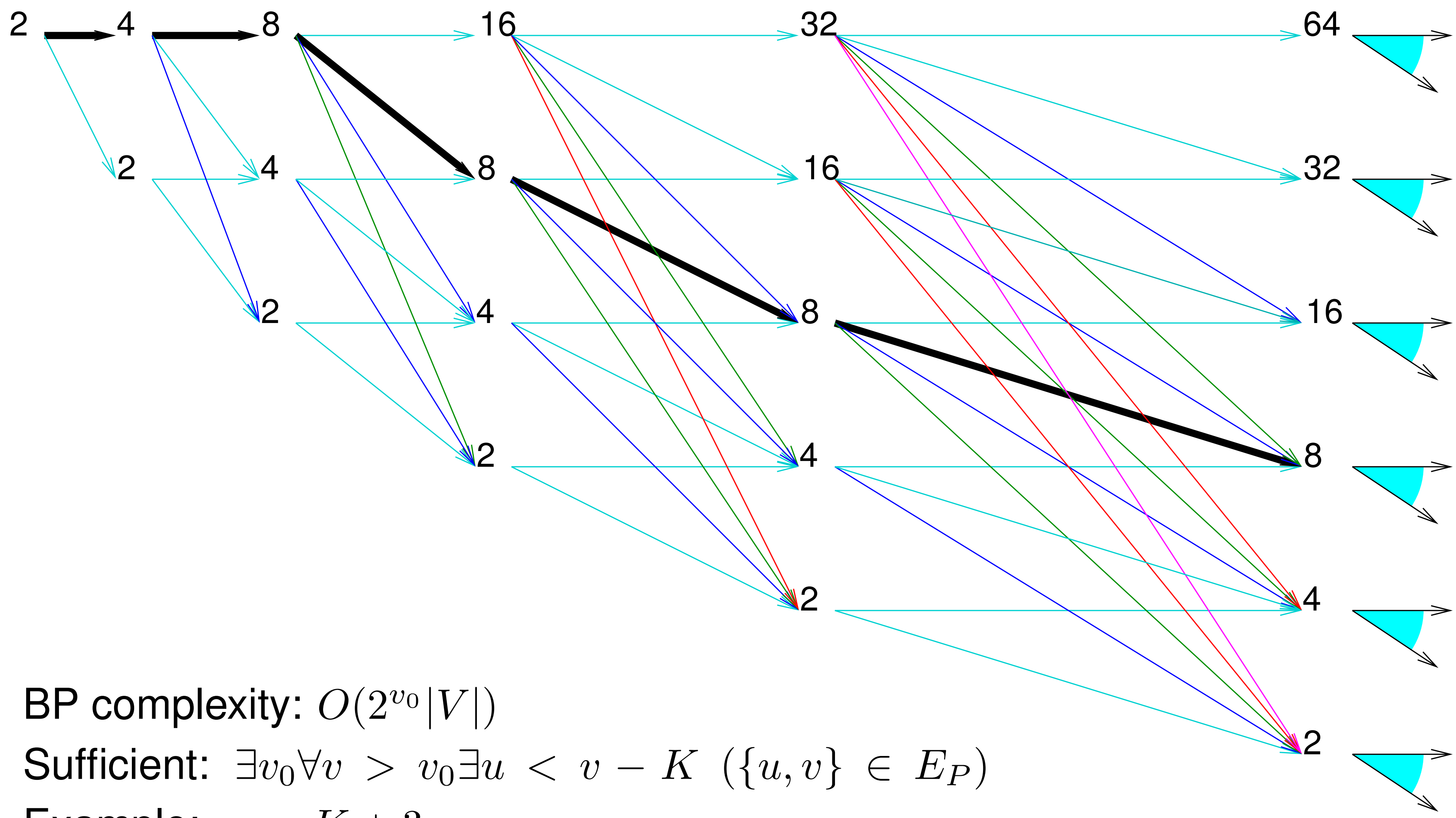
BP nodes vs. pruning edges

1st line: $v - u$

vertices: $|$BP nodes$|$ at level v (treewidth)

arcs: $\exists$ pruning edge $\{u + \text{arc\_label}, v\}$

Constant treewidth

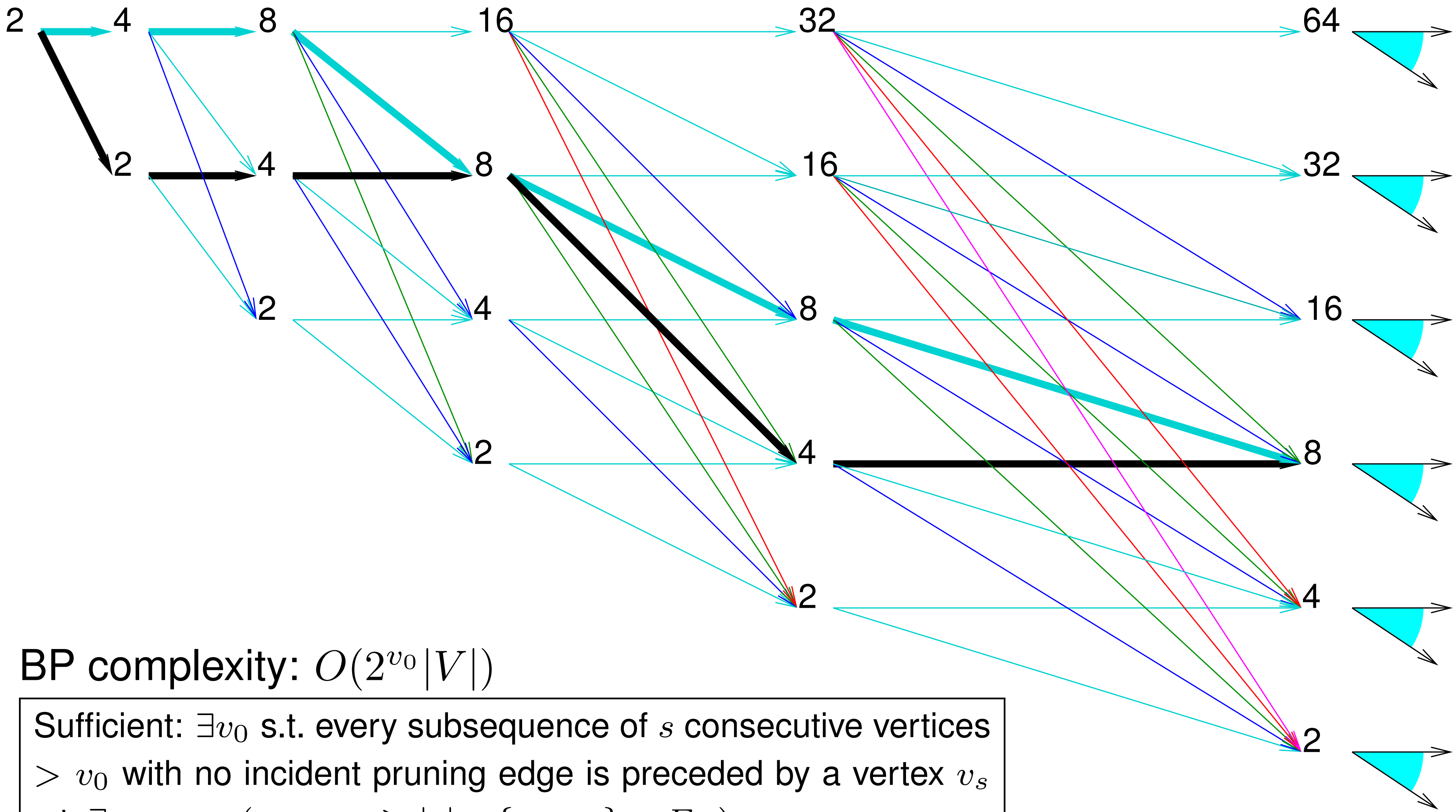


BP complexity: $O(2^{v_0}|V|)$

Sufficient: $\exists v_0 \forall v > v_0 \exists u < v - K \; (\{u,v\} \in E_P)$

Example: $v_0 = K + 3$

Constant-bounded treewidth



BP complexity: $O(2^{v_0}|V|)$

Sufficient: $\exists v_0$ s.t. every subsequence of s consecutive vertices $> v_0$ with no incident pruning edge is preceded by a vertex v_s s.t. $\exists u_s < v_s \ (v_s - u_s \geq \lvert s \rvert \wedge \{u_s, v_s\} \in E_P)$

"Any path under the constant path"

Fixed parameter tractability

- We can also allow treewidth growth as long as it's logarithmic in n
- This yields a fixed-parameter tractable behaviour for BP (w.r.t. v_0)

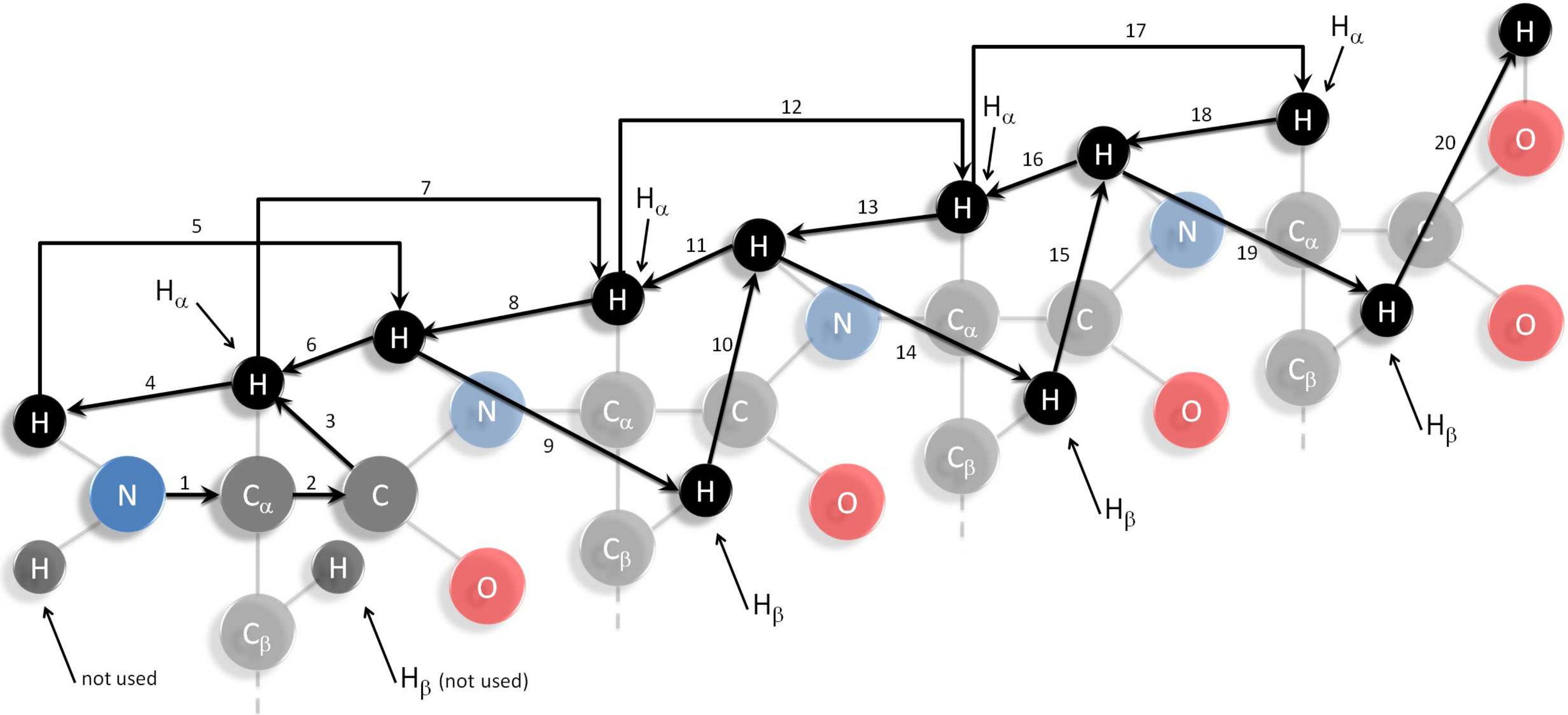
We tested all our protein instances: all display either constant or const-bounded treewidths **with very low** v_0 (i.e. $v_0 = 4$)

BP is polynomial on proteins (?)

Application to proteomics

Virtual hydrogen backbone

- The most accurate NMR distances are **between hydrogen atoms only**, but the actual backbone is a chain of N-C_α-C groups
- So find a *virtual* backbone composed of hydrogens only, and such that its order satisfies the DMDGP requirements



Certain hydrogens must be enumerated twice

Listing atoms twice

- If a hydrogen is listed twice, then there are $i \neq j \in V$ indexing the same atom
- Thus $x_i = x_j$ and $d_{ij} = 0$
- For all k such that $\{i,k\} \in E$, we have that $\{j,k\} \in E$ as $d_{jk} = d_{ik} + 0$, and

$$d_{ij} + d_{jk} = 0 + d_{jk} = d_{ik}$$

  so STRICT TRIANGULAR INEQUALITIES do not hold for all atom triplets
- However, it only fails on *nonconsecutive* triplets

  Hence, BP still applies
- Also, zero pruning distances help keeping floating point errors under control

Re-orders

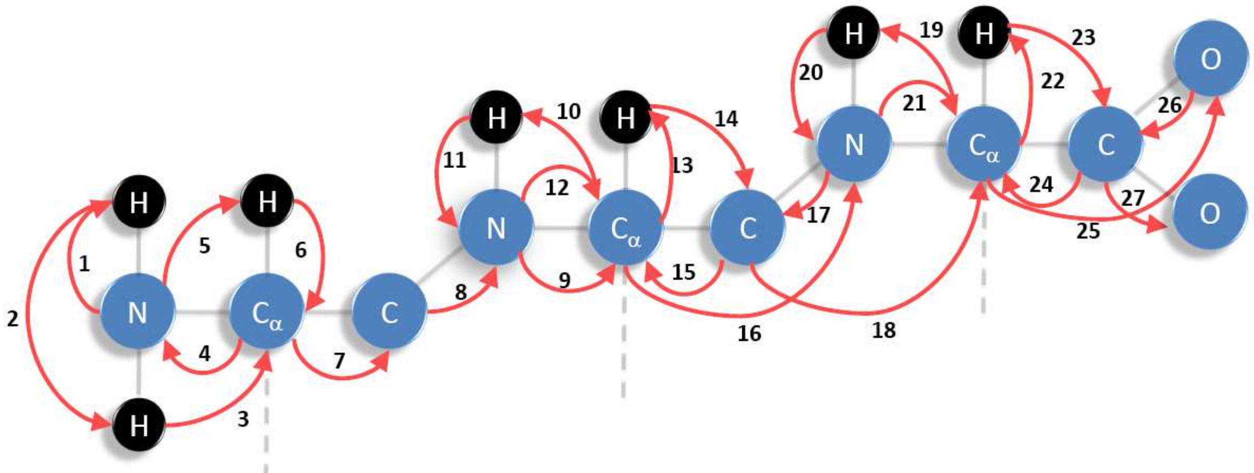
Defn.

A *repetition order* (re-order) is a finite sequence on V

- Re-orders generalize "counting vertices more than once"
- They add more flexibility to exploit certain distances as discretization distances
- Essentially, they provide a tool with which to hand-craft convenient vertex orders for interesting instance classes

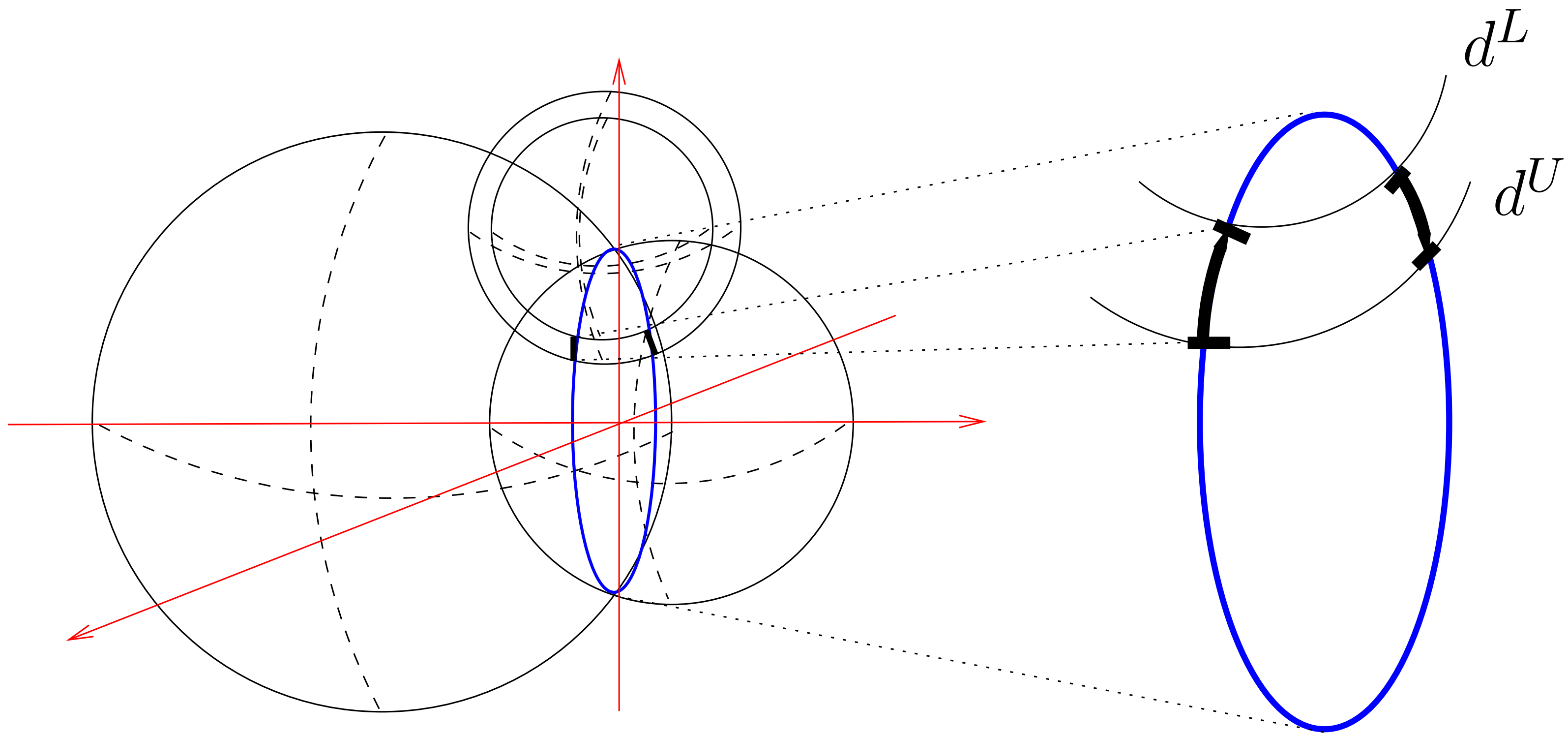
Not immediately evident how to best order proteins

Here's a re-order applying to all backbones



Uncertain distances

- Typically, NMR provides uncertain distances, modelled by intervals $[d^L_{uv}, d^U_{uv}]$
- Cannot be used for discretization



Two precise distances and an uncertain one

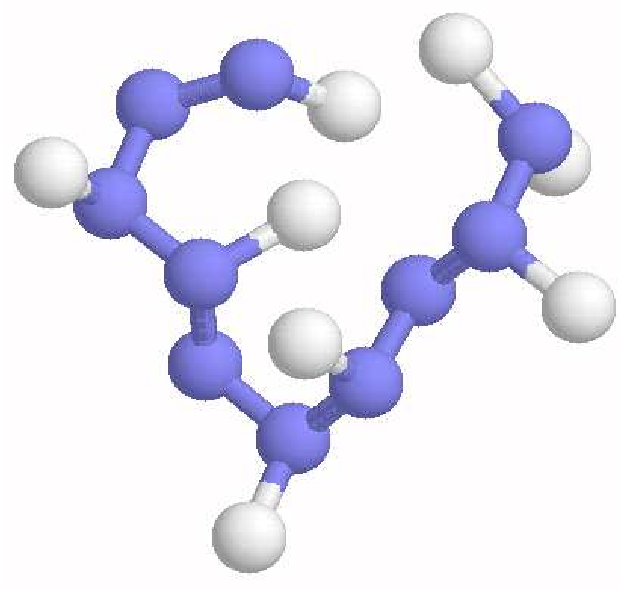
The actual situation

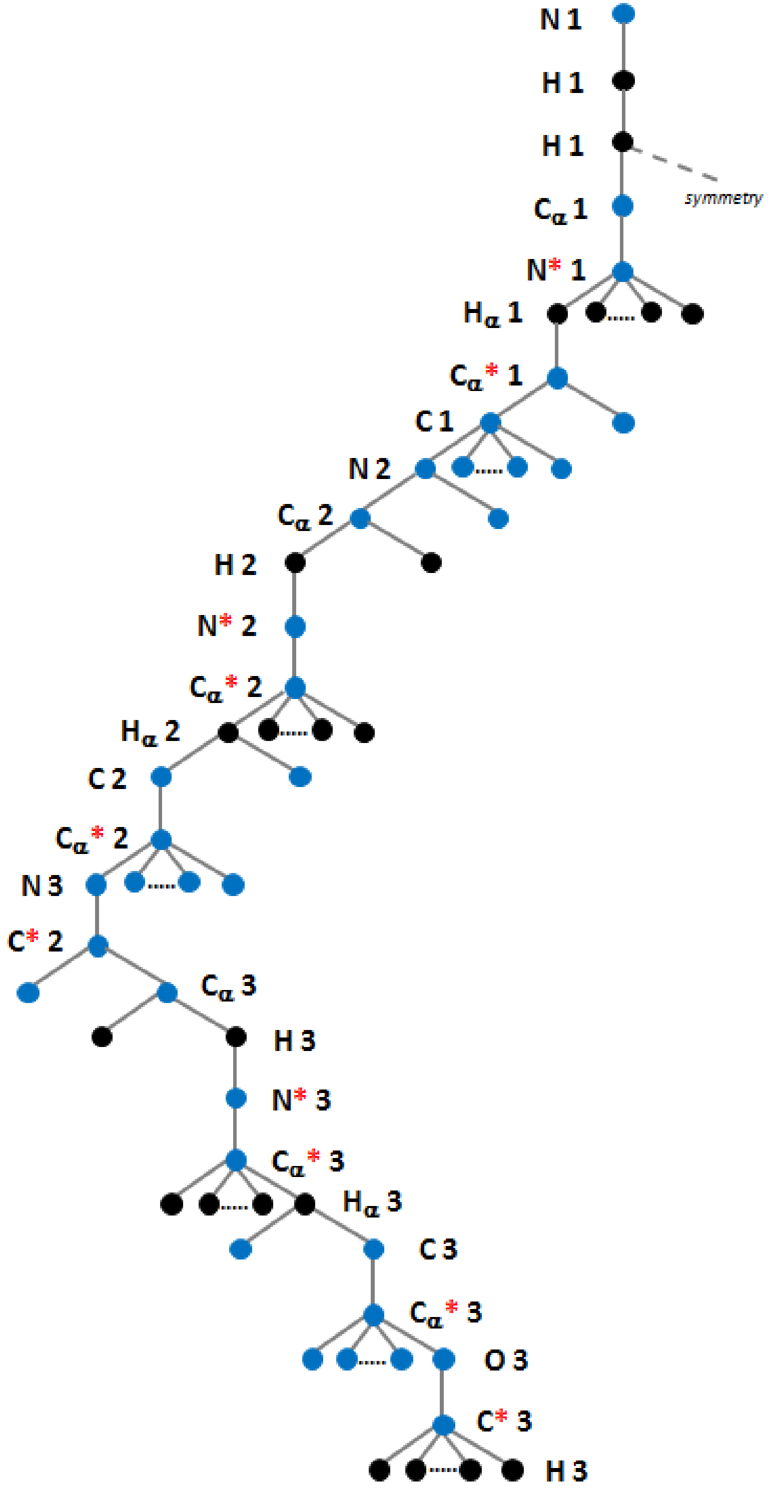
- We know several distances d_{uv} precisely because of chemical properties
- Some distances take values in a finite set D_{uv}
- The distribution of **precise/discrete/uncertain** distances on the protein backbone does not satisfy the DMDGP requirements

Re-orders provide a solution: use all **precise** distances for discretization, plus a few of the **discrete** whenever needed; **uncertain** distances are used for pruning

- Pruning with intervals is easy: if the current point x_v is s.t. $\|x_v - x_u\| \in [d^L_{uv}, d^U_{uv}]$ for all $u \in \alpha(v)$ accept it, otherwise prune it
- Discrete distances D_{uv} simply give rise to BP nodes at level $v-1$ with potentially $2|D_{uv}|$ subnodes

*i*BP




N 1
H 1
H 1
symmetry
C_α 1
N* 1
H_α 1
C_α* 1
C 1
N 2
C_α 2
H 2
N* 2
C_α* 2
H_α 2
C 2
C_α* 2
N 3
C* 2
C_α 3
H 3
N* 3
C_α* 3
H_α 3
C 3
C_α* 3
O 3
C* 3
H 3

Implementations

Sequential code

- The code is available in open source
- Download:
  `http://www.antoniomucherino.it/en/mdjeep.php`
- Any doubt, ask the MASTER (Antonio Mucherino)

Parallel code

Seconds of user CPU on Grid5000 (`www.grid5000.fr`)

	CPUs			
$\lvert V \rvert$	1	2	8	64
5000	3.21	1.30	0.54	0.36
7500	4.73	3.15	1.25	0.93
10000	13.38	5.49	2.49	1.57

Embed subgraphs then glue embeddings (rigidity $\Rightarrow$ exact)

A selection of current work

- Work with biochemists/bioinformaticians at Institut Pasteur to access and treat real NMR data
- Use $\mathcal{G}_P x = X$ result from symmetry to obtain all solutions from just one
- Extend complexity study to actual problem with discrete/uncertain distances
- Progress on "MDGP $\in$ **NP**?" question

See `http://www.lix.polytechnique.fr/~liberti/publications.html` for more papers

Surveys

- **Survey 1**: Liberti, Lavor, Mucherino, Maculan, *Molecular distance geometry methods: from continuous to discrete*, International Transactions in Operational Research, **18**:33-51, 2010
- **Survey 2**: Lavor, Liberti, Maculan, Mucherino, *Recent advances on the discretizable molecular distance geometry problem*, European Journal of Operational Research, **219**:698-706, 2012
- **Survey 3**: Liberti, Lavor, Maculan, Mucherino, *Euclidean distance geometry and applications*, SIAM Review, to appear (meanwhile: arXiv 1205.0349v1)

End of course

Appendix

Continuous formulation

- Solving the system

$$\forall \{i,j\} \in E \quad ||x_i - x_j|| = d_{ij}, \tag{10}$$

is numerically challenging

LHS involves $\sqrt{\text{arg}}$, floating point ops $\Rightarrow$ arg $< 0 \Rightarrow$ error and abort

$\Rightarrow$ square both sides

- Usually, cast as a penalty objective to be minimized

$$\min_x \sum_{\{i,j\} \in E} (||x_i - x_j||^2 - d_{ij}^2)^2. \tag{11}$$

- Unconstrained minimization of a polynomial of fourth degree

General-purpose methods

- sBB (exact): OK on small and medium-sized instances

because we know the optimal value of the objective (0), lower bound is tight at the initial tree levels

- VNS (heur): good for large(ish) instances
- MultiLevel Single Linkage (heur) [Kucherenko et al. '06]: so-so

		sBB		VNS		MLSL	
Atoms	Variables	OF Value	Time	OF Value	Time	OF Value	Time
cube8	24	0	0.22	0	1.21	0	13.56
cube27	81	0	30.39	0	34.01	0	300.285
cube64	192	0	2237.73	0	398.875	0	2765.13
lavor5	15	0	0.02	0	0.48	0	0.57
lavor10	30	0	1.12	0	7.06	0	69.71
lavor20	60	0	2.25	0	49.99	0	411.152
lavor30	90	0	488.87	0	352.06	0	1634.09
lavor40	120	-	-	0.09	1258.13	0.547	2376.01
lavor50	150	-	-	0	673.48	0	3002.88

MDGP-specific methods

- Smoothing-based:
  - Continuation method (heur) [Moré, Wu ’97]
  - Double VNS with smoothing (heur) [L. et al. ’09]
  - DC optimization with smoothing (heur) [An et al. ’03]
  - Hyperbolic smoothing (heur) [Xavier ’08]
- Alternating projections algorithm (heur) [Glunt et al. 90]:
  iterative updating of a dissimilarity matrix
- Geometric build-up (exact/heur) [Dong, Wu ’03 and ’07]: *triangulation*
- GNOMAD (heur) [Williams et al. ’01]
  iterative updating of atomic ordering minimizing error contribution
- Monotonic Basin Hopping (heur) [Grosso et al. ’09]
  funnel-based population heuristic
- Self-organization heuristic (heur) [Xu et al. ’03]
  pairwise atomic position modification heuristic
- SDP-based formulation *[Ye et al. ’09]*