

The Impact of Heterogeneity and Geometry on the Proof Complexity of Random Satisfiability

Thomas Bläsius ()^{KIT} Karlsruhe Institute of Technology), Tobias Friedrich ()^{HPI} Hasso Plattner Institut Digital Engineering · Universität Potsdam), Andreas Göbel ()^{HPI} Hasso Plattner Institut Digital Engineering · Universität Potsdam),
Jordi Levy ()^{CSIC} Consejo Superior de Investigaciones Científicas), Ralf Rothenberger ()^{HPI} Hasso Plattner Institut Digital Engineering · Universität Potsdam)



Satisfiability

k-SAT

- input: CNF-formula with k literals per clause

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

Satisfiability

k-SAT

- input: CNF-formula with k literals per clause

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- goal: find satisfying assignment

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

Satisfiability

k-SAT

- input: CNF-formula with k literals per clause

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- goal: find satisfying assignment

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- ...or prove that no such assignment exists

Satisfiability

k-SAT

- input: CNF-formula with k literals per clause

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- goal: find satisfying assignment

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- ...or prove that no such assignment exists

$$\frac{\begin{array}{c} \vdots \\ x_1 \vee x_3, \quad x_2 \vee \neg x_3 \end{array}}{x_1 \vee x_2}$$

Satisfiability

k-SAT

- input: CNF-formula with k literals per clause

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- goal: find satisfying assignment

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- ...or prove that no such assignment exists

$$\begin{array}{c} \vdots \\ \frac{x_1 \vee x_3, \quad x_2 \vee \neg x_3}{x_1 \vee x_2} \\ \\ \frac{x_1 \vee x_2, \quad \neg x_1 \vee x_2}{x_2} \end{array}$$

Satisfiability

k -SAT

- input: CNF-formula with k literals per clause

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- goal: find satisfying assignment

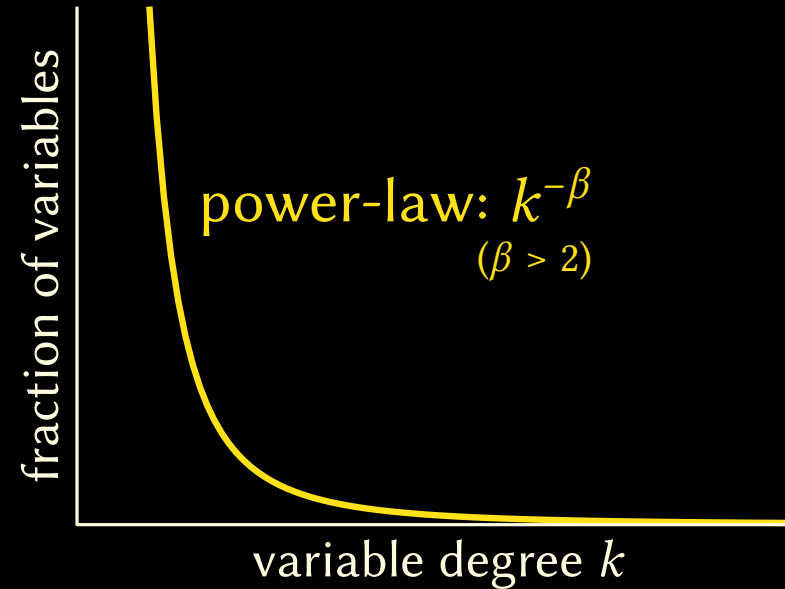
$$(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee x_4)$$

- ...or prove that no such assignment exists

$$\begin{array}{c}
 \vdots \\
 \frac{x_1 \vee x_3, \quad x_2 \vee \neg x_3}{x_1 \vee x_2} \\
 \\
 \frac{x_1 \vee x_2, \quad \neg x_1 \vee x_2}{x_2} \qquad \qquad \qquad \frac{\neg x_2 \vee \neg x_3, \quad \neg x_2 \vee x_3}{\neg x_2} \\
 \\
 \frac{x_2, \quad \neg x_2}{\emptyset}
 \end{array}$$

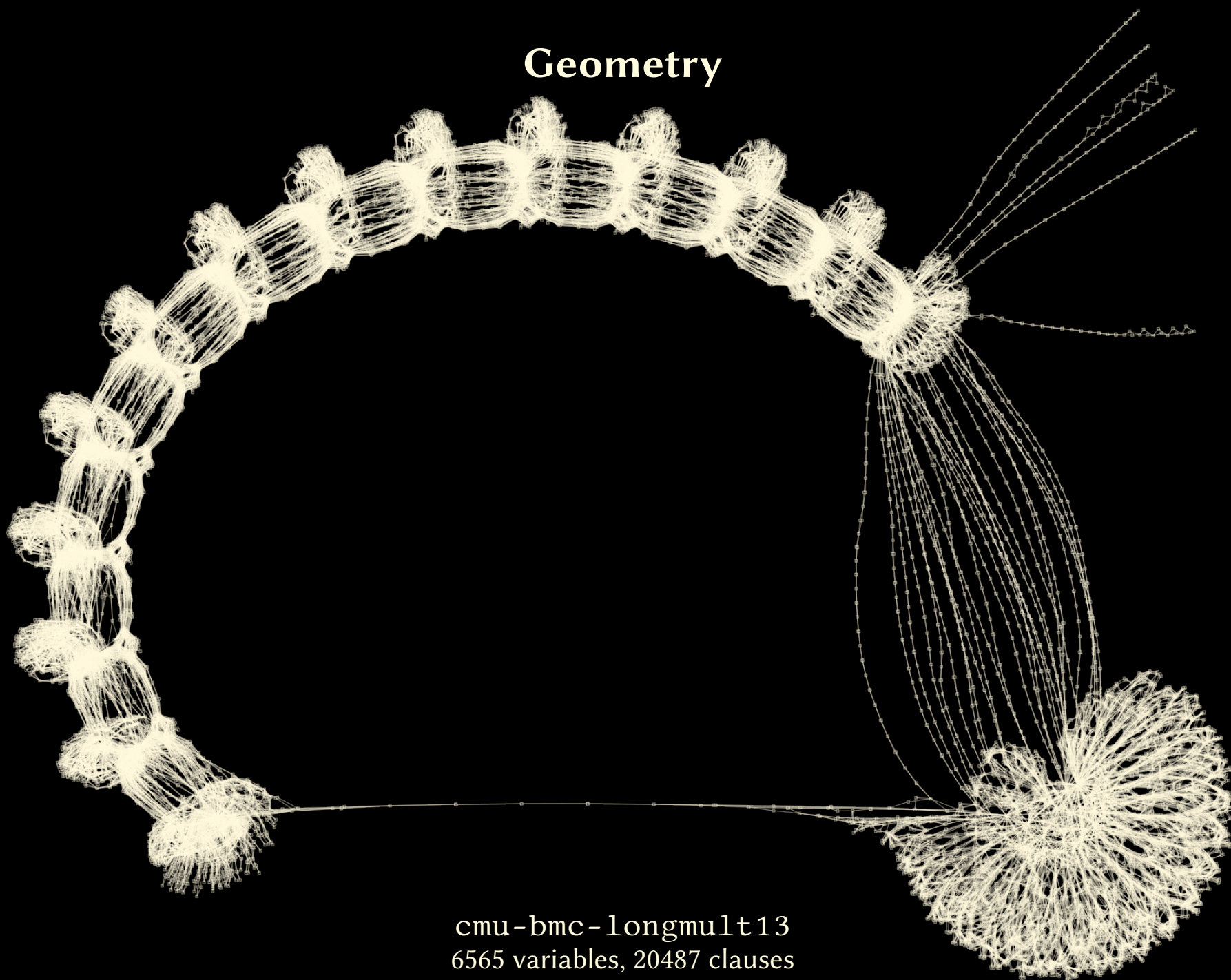
Typical Properties of Industrial SAT Instances

Heterogeneity



Typical Properties of Industrial SAT Instances

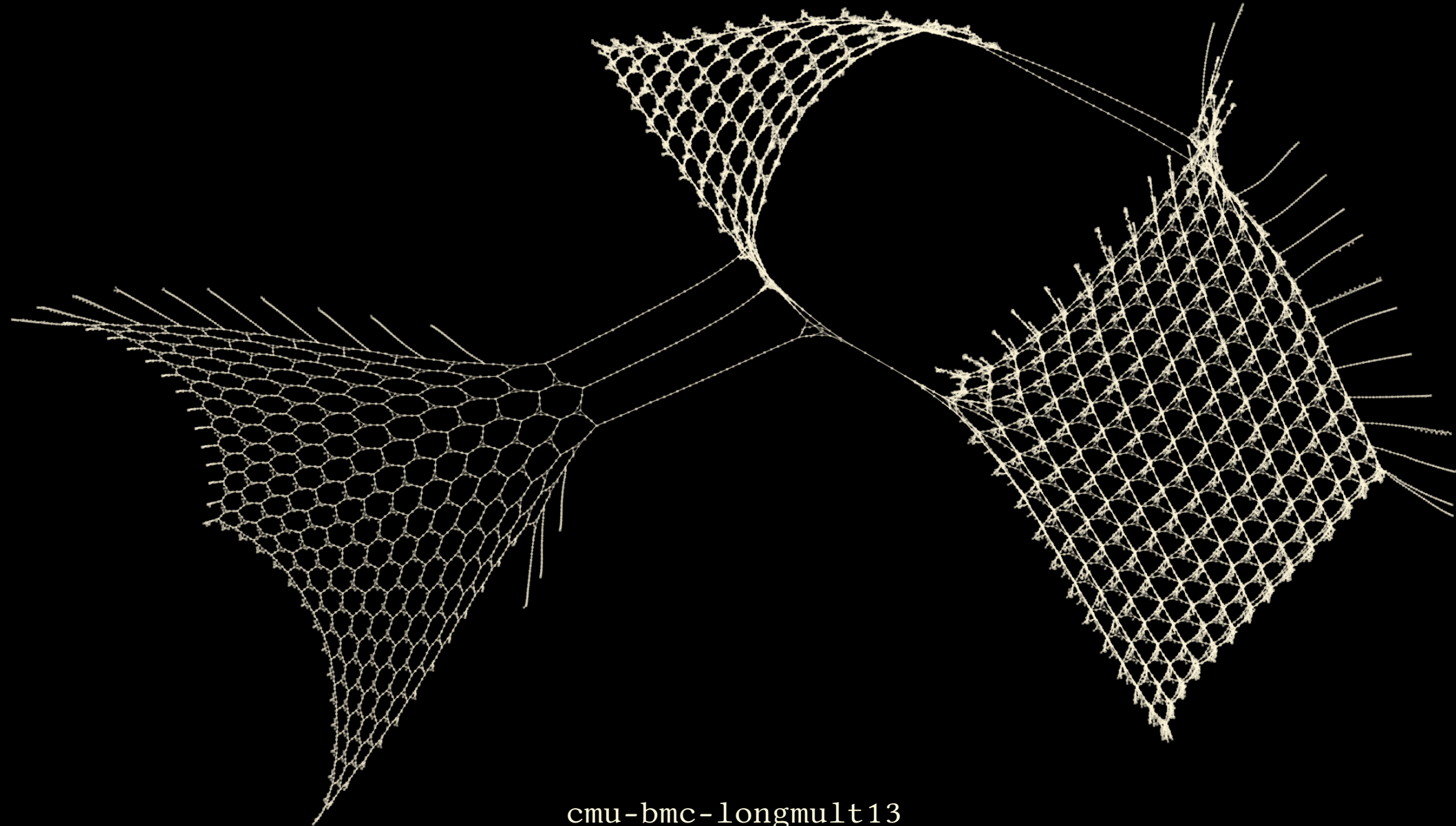
Geometry



cmu-bmc-longmult13
6565 variables, 20487 clauses

Typical Properties of Industrial SAT Instances

Geometry



cmu-bmc-longmult13
6565 variables, 20487 clauses

Random SAT with Heterogeneity and Geometry

Heterogeneity: Power-Law Random k -SAT [Ansótegui, Bonet, Levy 2009]

- for each clause: independently draw k variables without repetition
- variable $v \in \{1, \dots, n\}$ is chosen with probability proportional to $w_v = v^{-\frac{1}{\beta-1}}$
- independently negate each variable with probability $\frac{1}{2}$

Random SAT with Heterogeneity and Geometry

Heterogeneity: Power-Law Random k -SAT [Ansótegui, Bonet, Levy 2009]

- for each clause: independently draw k variables without repetition
- variable $v \in \{1, \dots, n\}$ is chosen with probability proportional to $w_v = v^{-\frac{1}{\beta-1}}$
- independently negate each variable with probability $\frac{1}{2}$

Geometry: Geometric Random k -SAT

- related to:
 - popularity-similarity SAT [Giráldez-Cru, Levy 2017]
 - hyperbolic random graphs [Krioukov et al. 2010]
 - GIRGs [Bringmann, Keusch, Lengler 2017]

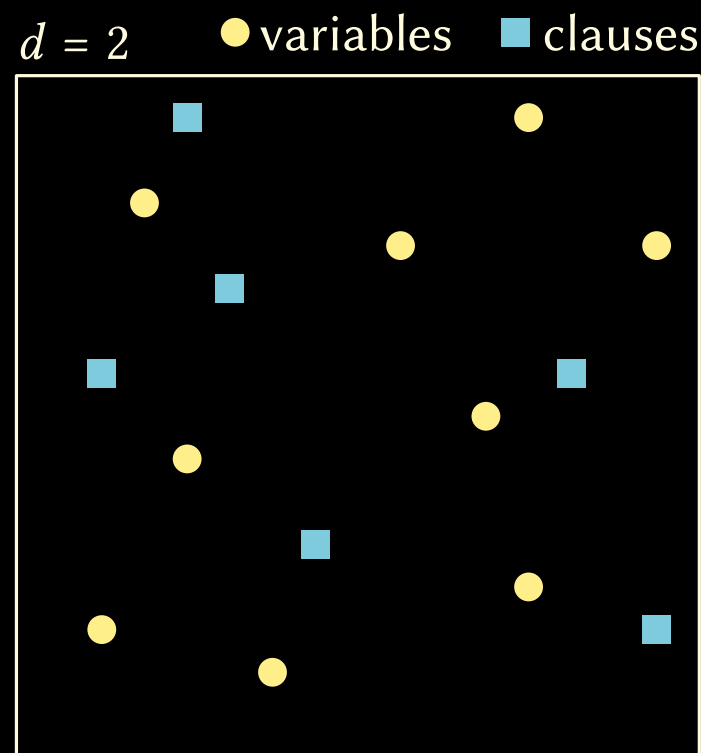
Random SAT with Heterogeneity and Geometry

Heterogeneity: Power-Law Random k -SAT [Ansótegui, Bonet, Levy 2009]

- for each clause: independently draw k variables without repetition
- variable $v \in \{1, \dots, n\}$ is chosen with probability proportional to $w_v = v^{-\frac{1}{\beta-1}}$
- independently negate each variable with probability $\frac{1}{2}$

Geometry: Geometric Random k -SAT

- related to:
 - popularity-similarity SAT [Giráldez-Cru, Levy 2017]
 - hyperbolic random graphs [Krioukov et al. 2010]
 - GIRGs [Bringmann, Keusch, Lengler 2017]
- sample d -dimensional positions



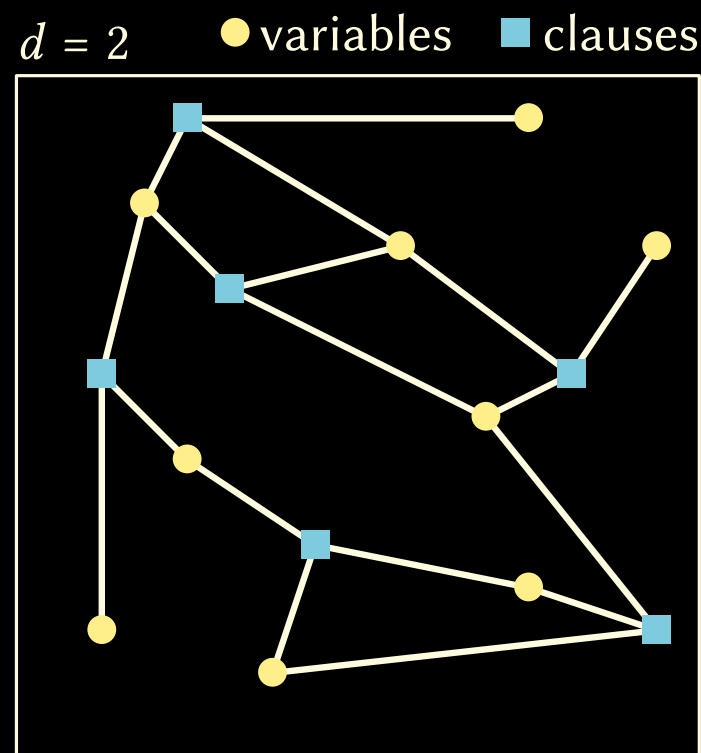
Random SAT with Heterogeneity and Geometry

Heterogeneity: Power-Law Random k -SAT [Ansótegui, Bonet, Levy 2009]

- for each clause: independently draw k variables without repetition
- variable $v \in \{1, \dots, n\}$ is chosen with probability proportional to $w_v = v^{-\frac{1}{\beta-1}}$
- independently negate each variable with probability $\frac{1}{2}$

Geometry: Geometric Random k -SAT

- related to:
 - popularity-similarity SAT [Giráldez-Cru, Levy 2017]
 - hyperbolic random graphs [Krioukov et al. 2010]
 - GIRGs [Bringmann, Keusch, Lengler 2017]
- sample d -dimensional positions



Random SAT with Heterogeneity and Geometry

Heterogeneity: Power-Law Random k -SAT [Ansótegui, Bonet, Levy 2009]

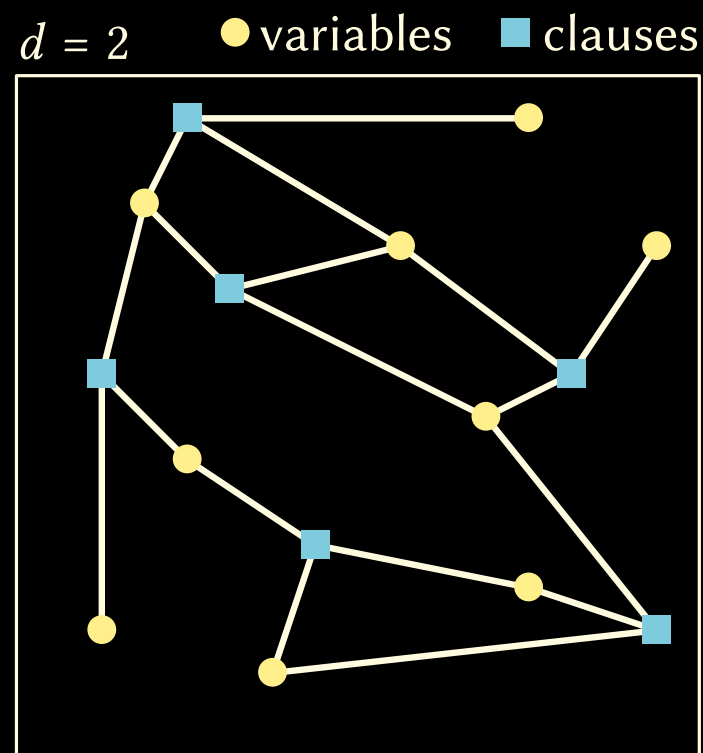
- for each clause: independently draw k variables without repetition
- variable $v \in \{1, \dots, n\}$ is chosen with probability proportional to $w_v = v^{-\frac{1}{\beta-1}}$
- independently negate each variable with probability $\frac{1}{2}$

Geometry: Geometric Random k -SAT

- related to:
 - popularity-similarity SAT [Giráldez-Cru, Levy 2017]
 - hyperbolic random graphs [Krioukov et al. 2010]
 - GIRGs [Bringmann, Keusch, Lengler 2017]
- sample d -dimensional positions
- connection weight for clause c and variable v :

$$X(c, v) = \left(\frac{w_v}{\|c - v\|^d} \right)^{1/T}$$

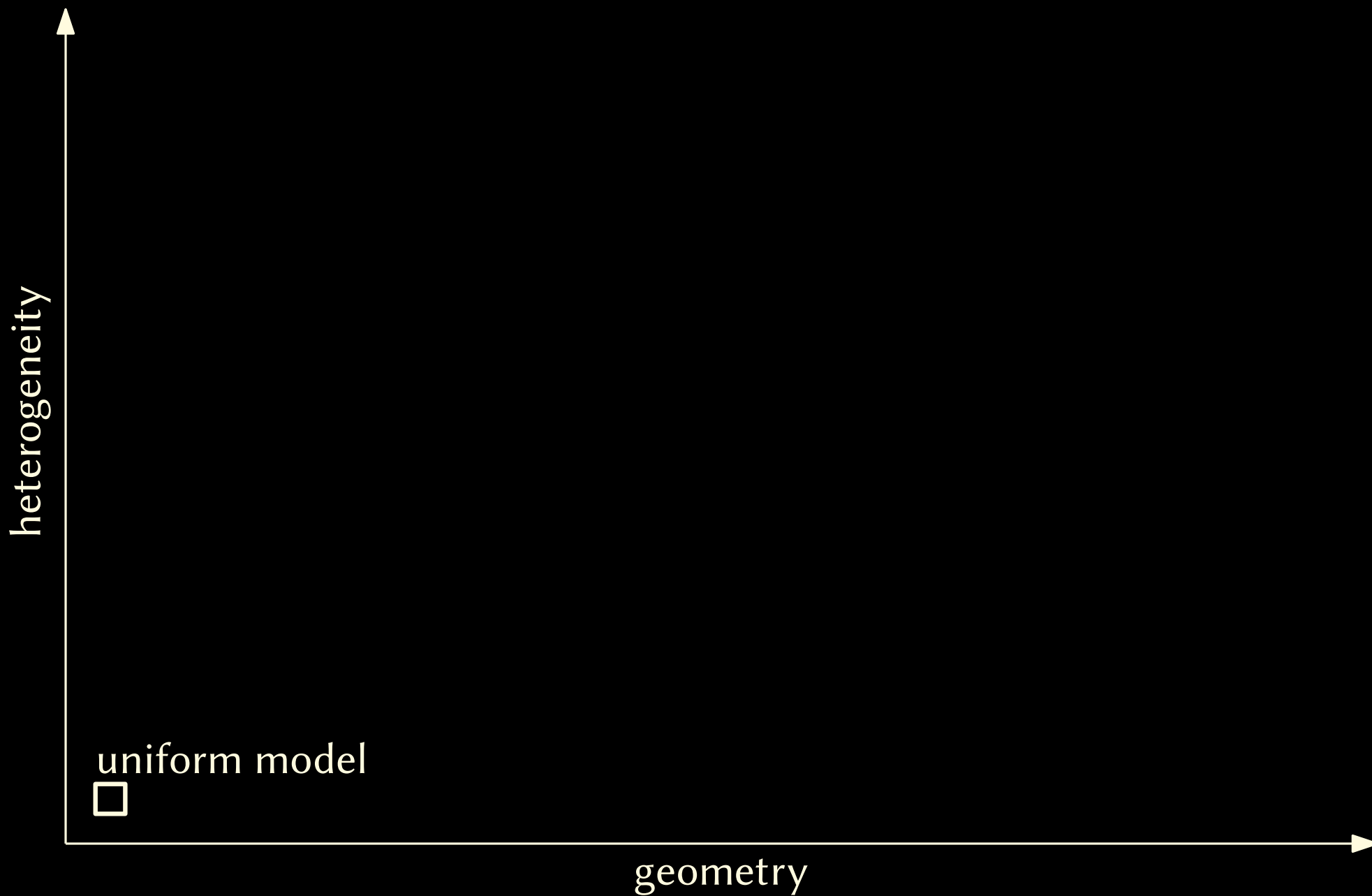
- for clause c : draw k different variables with probabilities proportional to $X(c, v)$



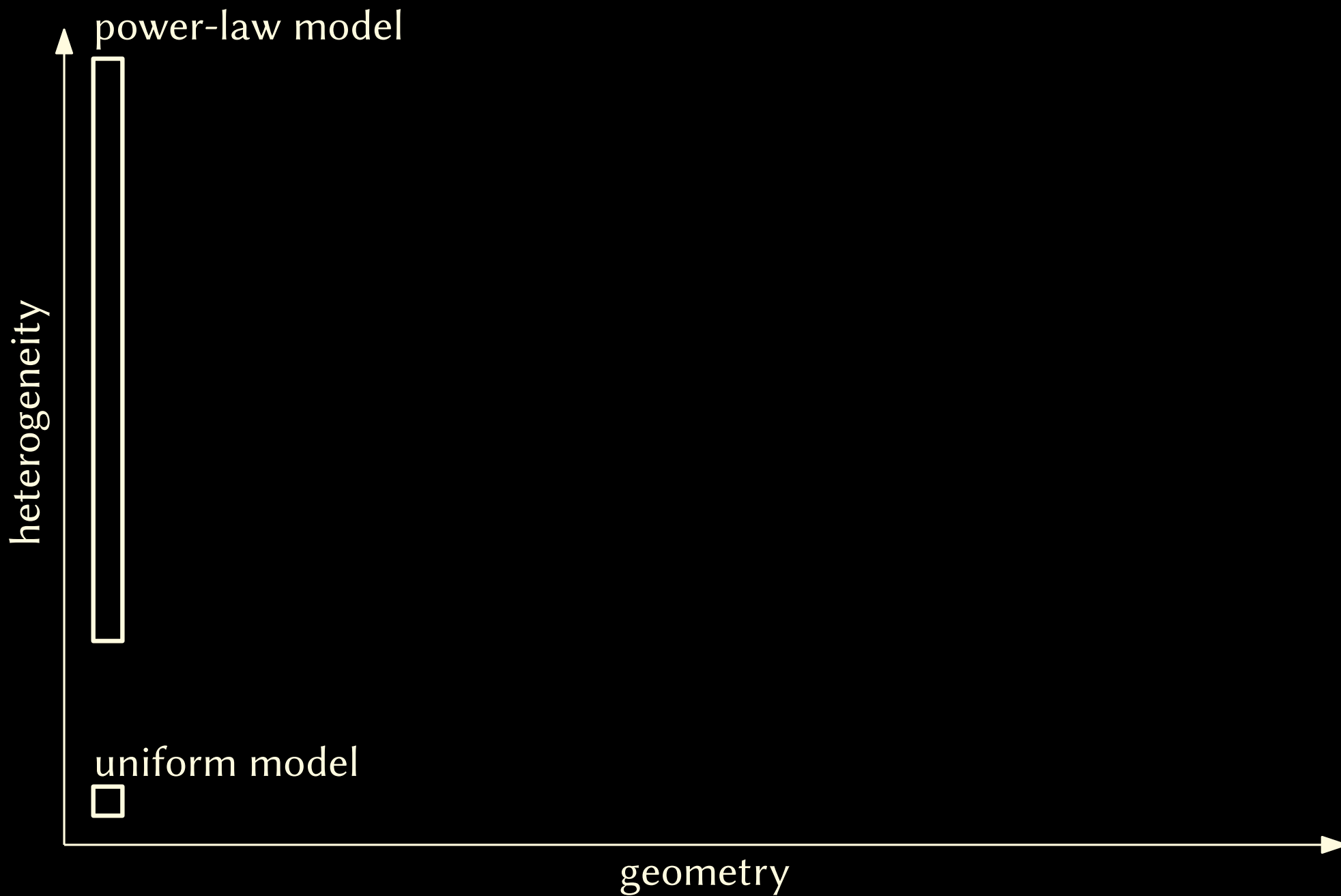
The Big Picture



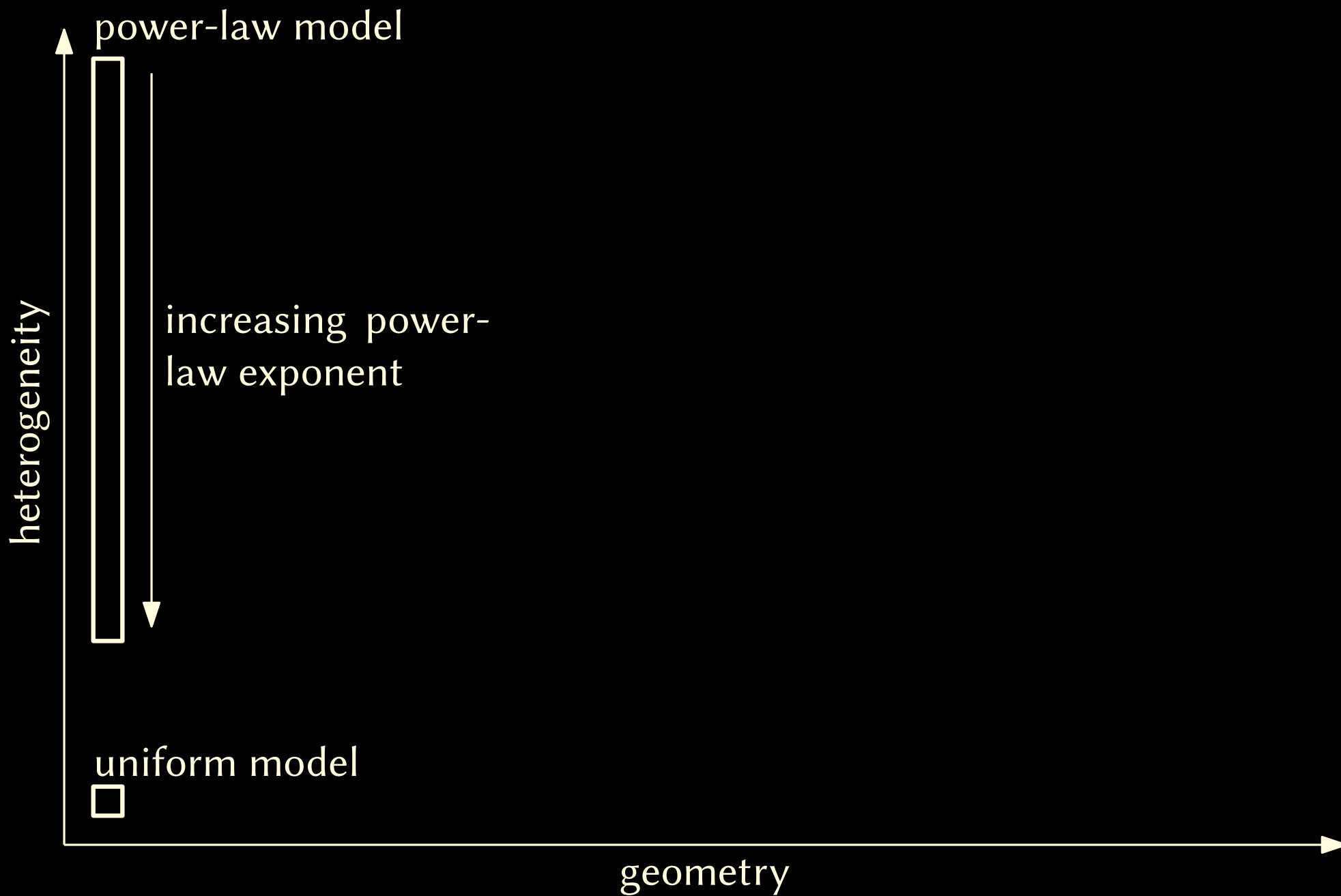
The Big Picture



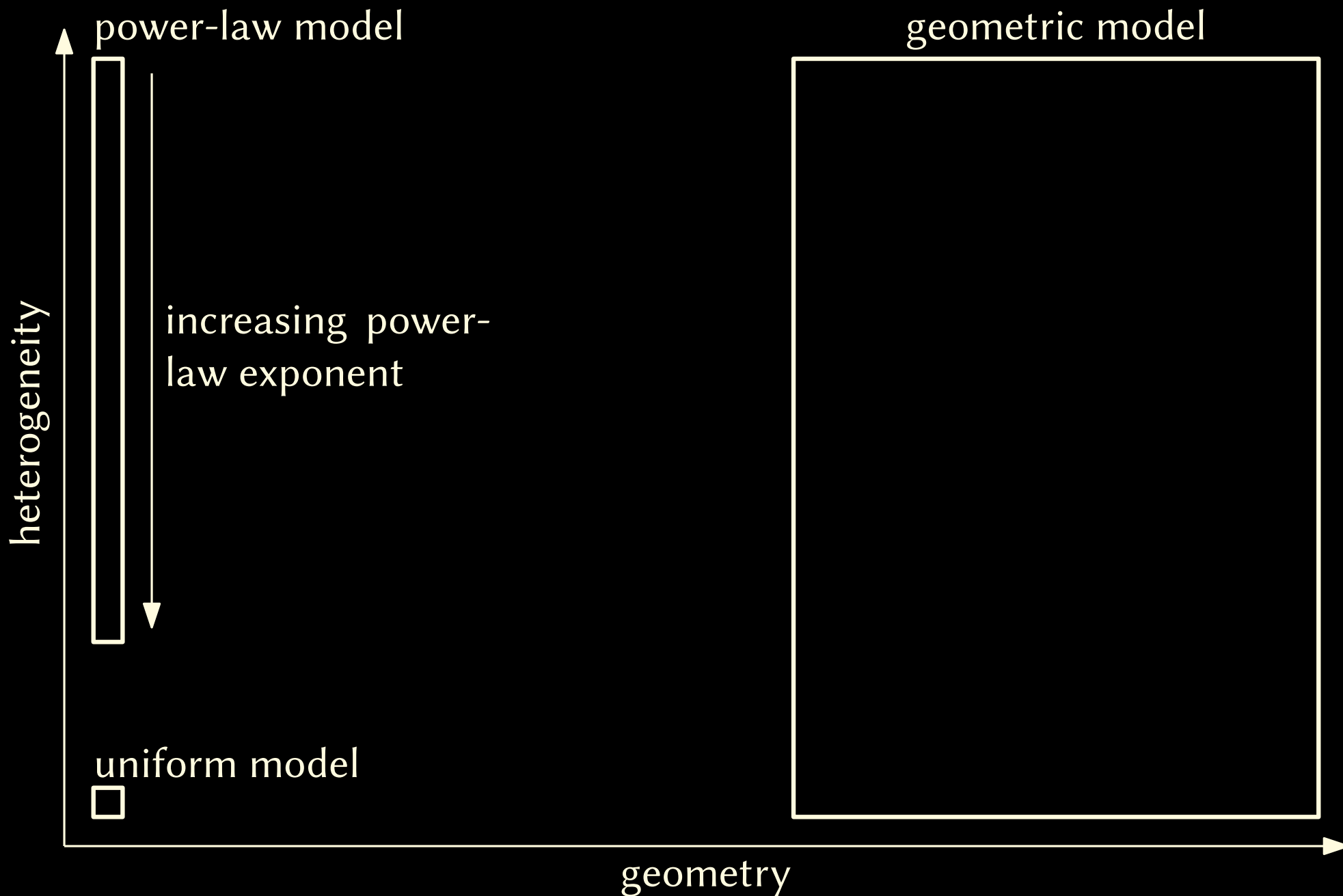
The Big Picture



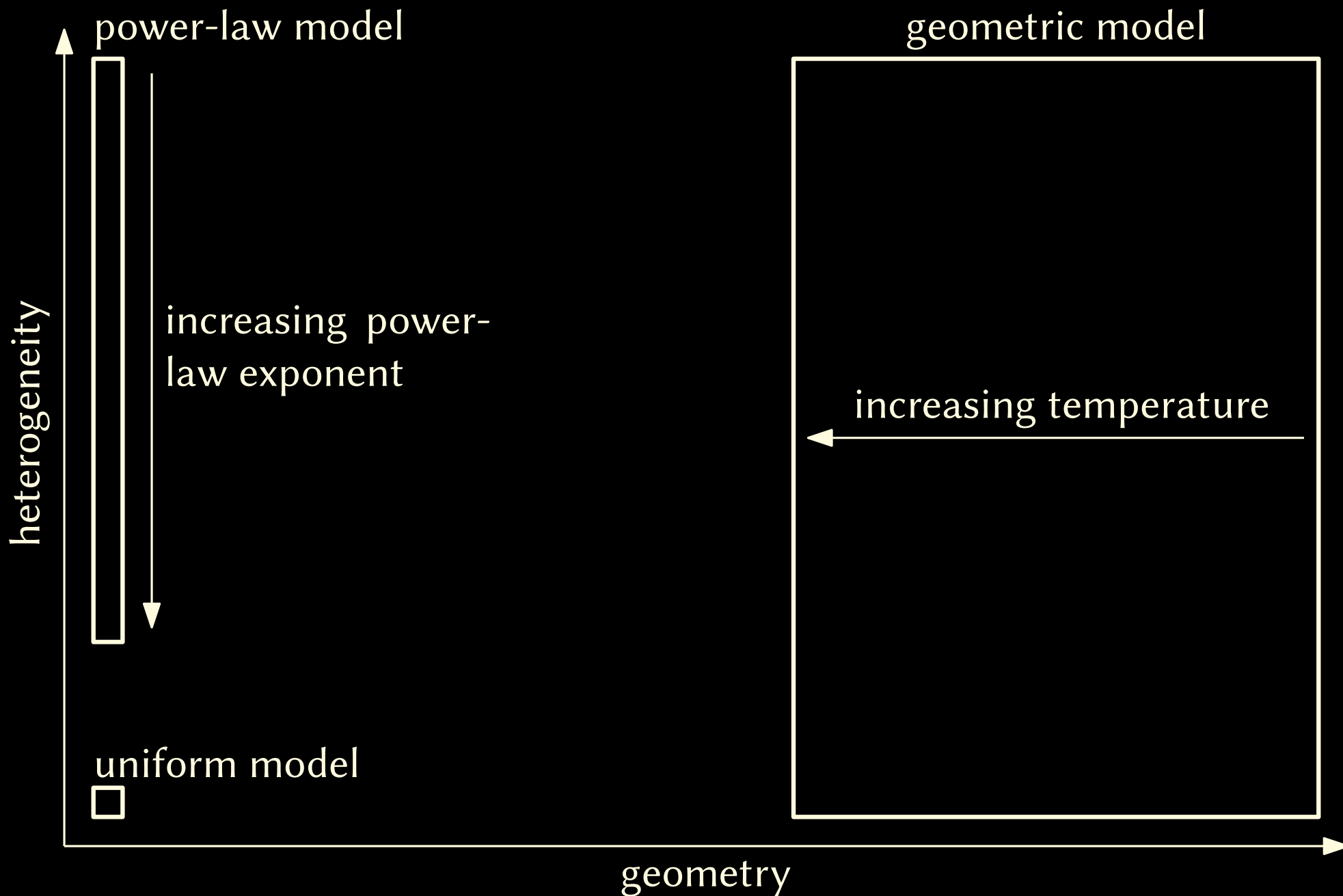
The Big Picture



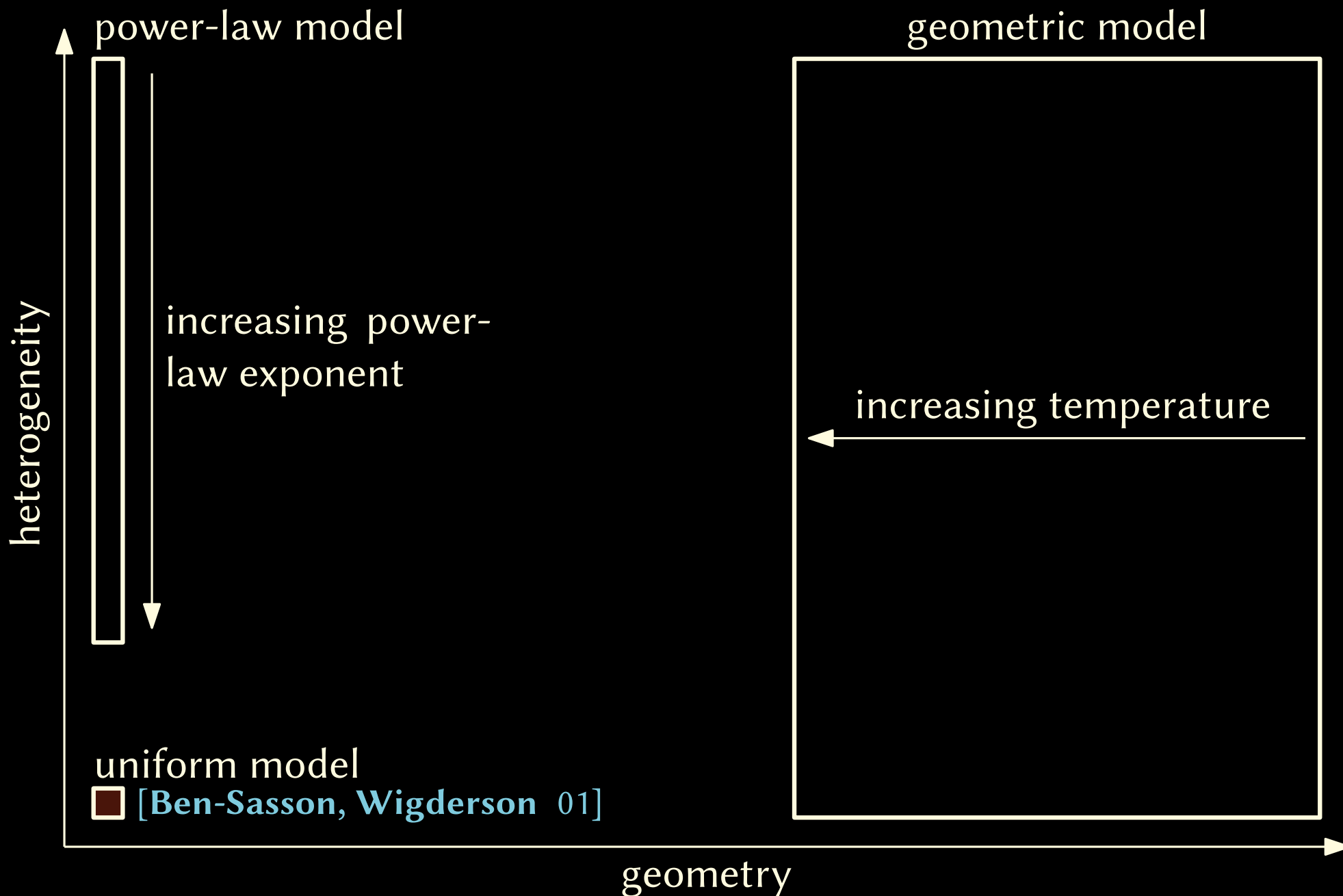
The Big Picture



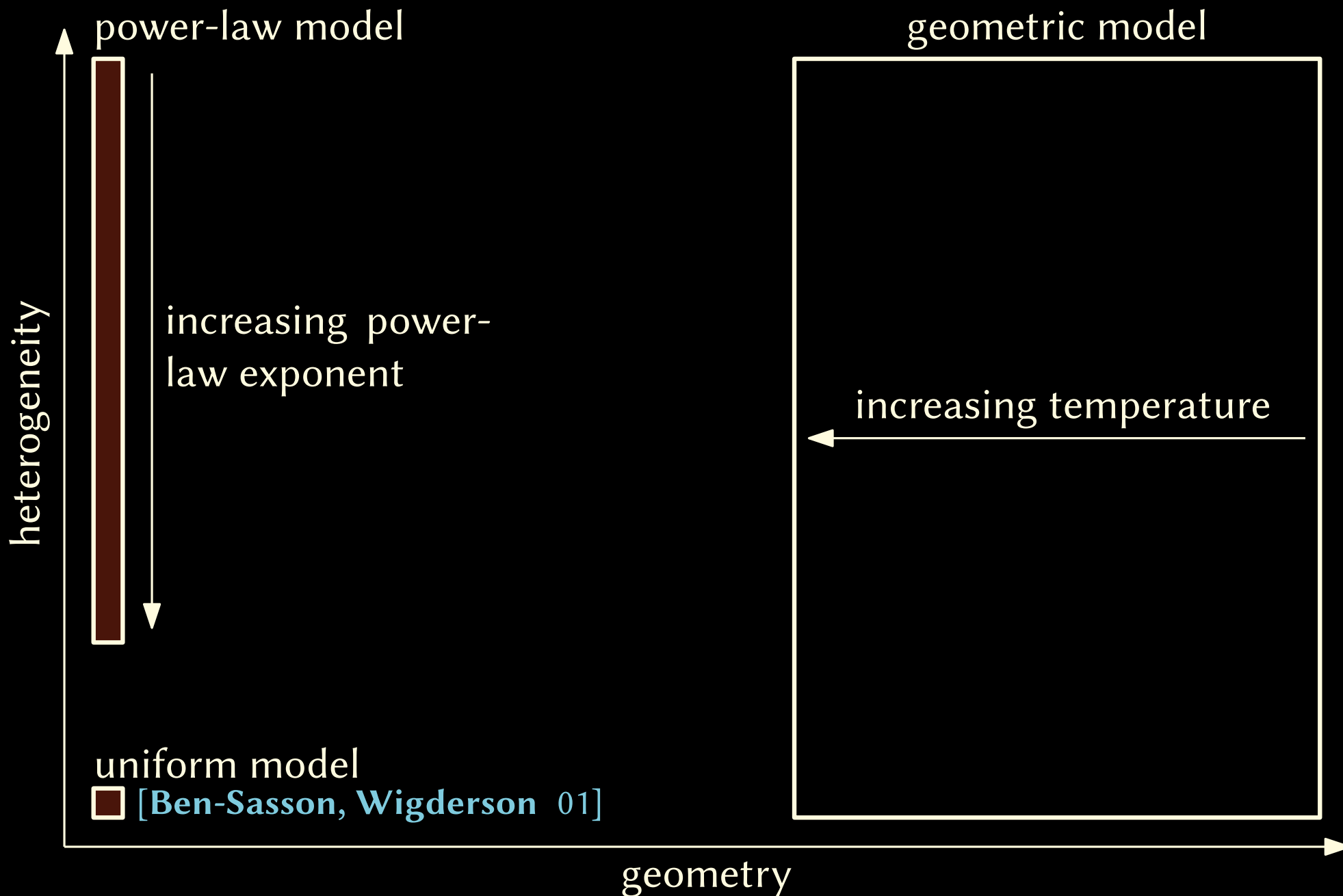
The Big Picture



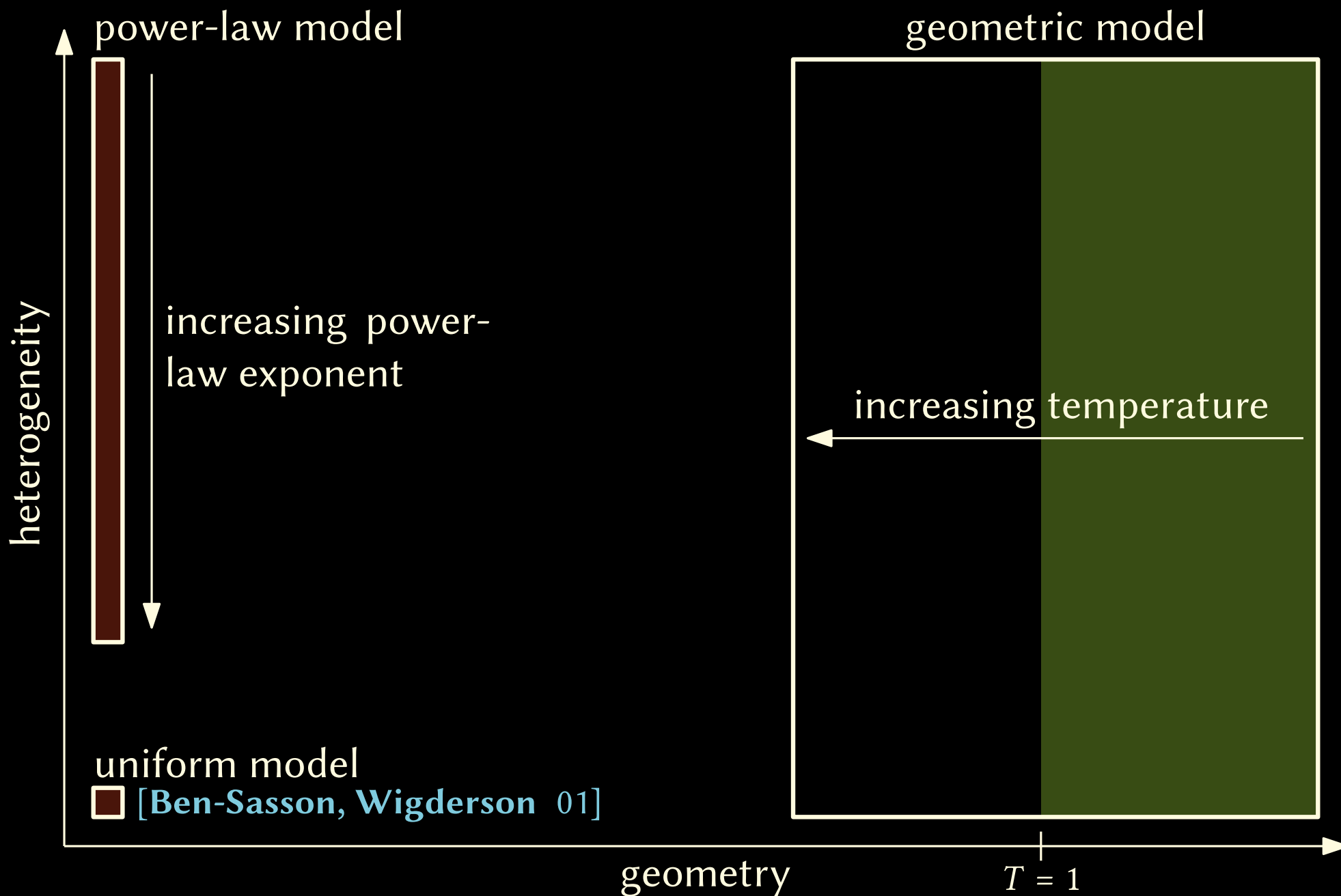
The Big Picture



The Big Picture



The Big Picture



Power-Law Random k -SAT

Proof Complexity Lower Bounds

[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size
 - $\exp(\Omega(w^2/n))$
 - $\exp(\Omega(w))$ for tree-like resolution

Power-Law Random k -SAT

Proof Complexity Lower Bounds

[Ben-Sasson, Wigderson 01]

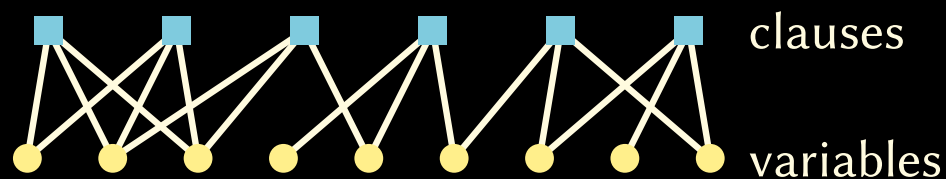
- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - $\exp(\Omega(w^2/n))$ \rightarrow CDCL
 - $\exp(\Omega(w))$ for tree-like resolution \rightarrow DPLL

Power-Law Random k -SAT

Proof Complexity Lower Bounds

[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - $\exp(\Omega(w^2/n))$ \rightarrow CDCL
 - $\exp(\Omega(w))$ for tree-like resolution \rightarrow DPLL
- high bipartite expansion \Rightarrow high width

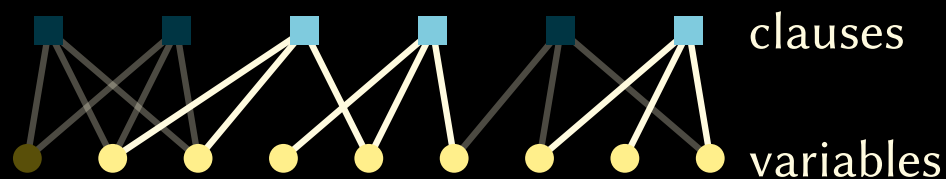


Power-Law Random k -SAT

Proof Complexity Lower Bounds

[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - $\exp(\Omega(w^2/n))$ \rightarrow CDCL
 - $\exp(\Omega(w))$ for tree-like resolution \rightarrow DPLL
- high bipartite expansion \Rightarrow high width

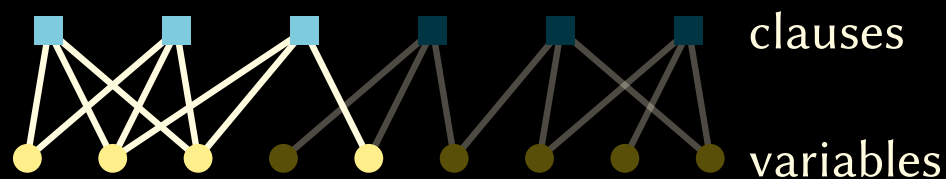


Power-Law Random k -SAT

Proof Complexity Lower Bounds

[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - $\exp(\Omega(w^2/n))$ \rightarrow CDCL
 - $\exp(\Omega(w))$ for tree-like resolution \rightarrow DPLL
- high bipartite expansion \Rightarrow high width

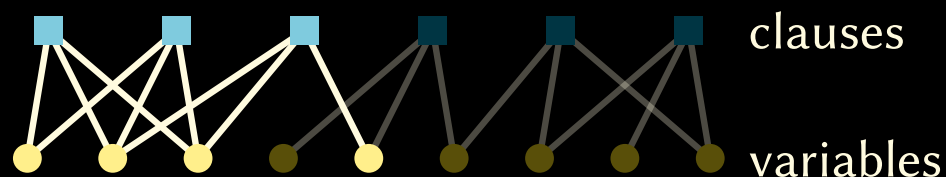


Power-Law Random k -SAT

Proof Complexity Lower Bounds

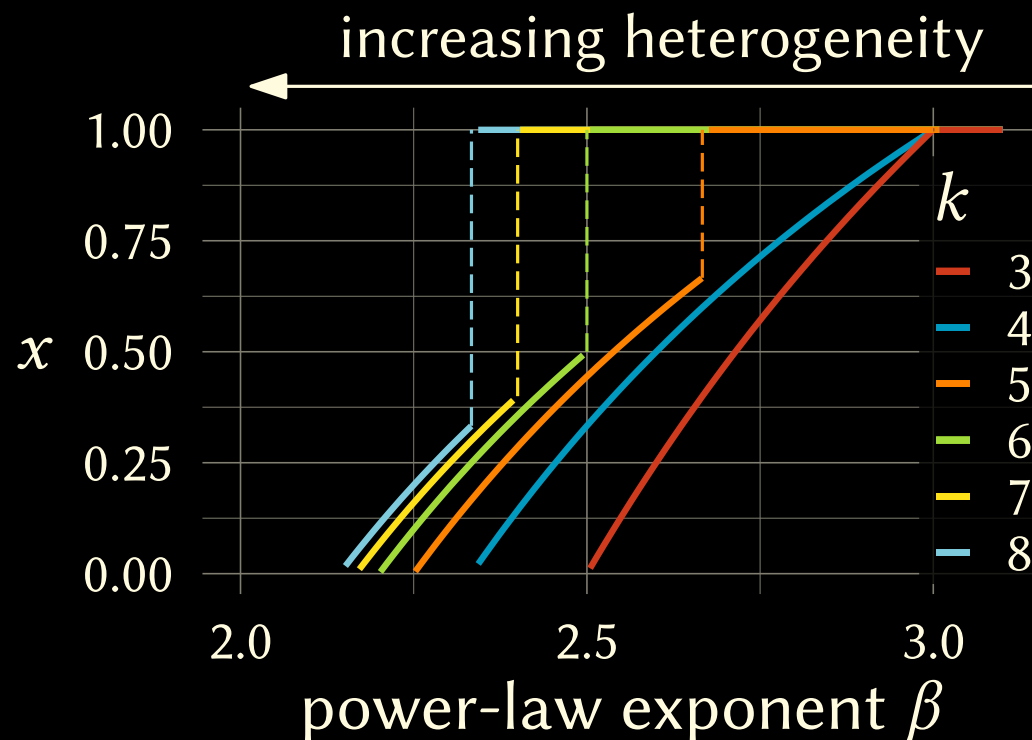
[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - \rightarrow CDCL
 - \rightarrow DPLL
- high bipartite expansion \Rightarrow high width



Our Lower Bounds (simplified)

- resolution width $w \in \tilde{\Omega}(n^x)$

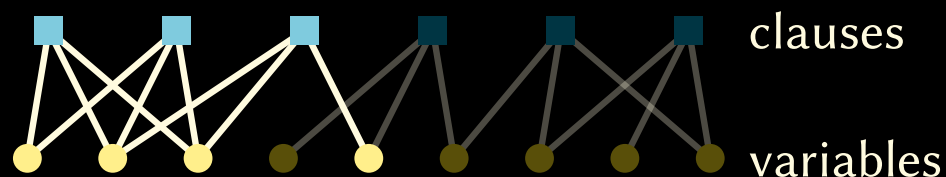


Power-Law Random k -SAT

Proof Complexity Lower Bounds

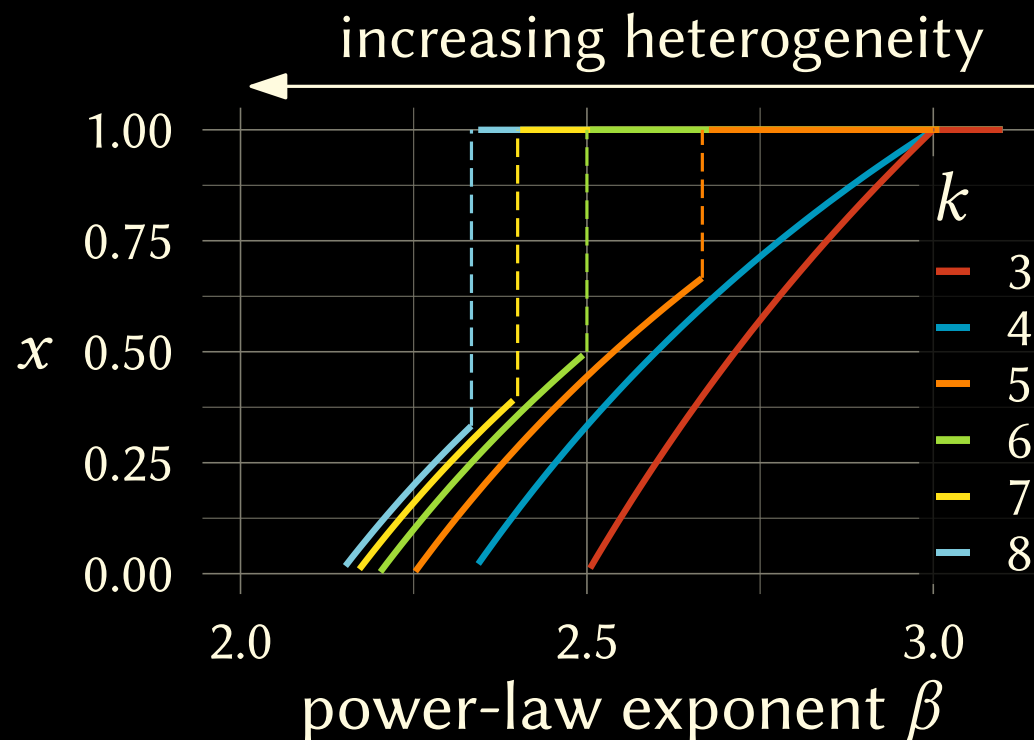
[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - \rightarrow CDCL
 - \rightarrow DPLL
- high bipartite expansion \Rightarrow high width



Our Lower Bounds (simplified)

- resolution width $w \in \tilde{\Omega}(n^x)$
- holds for arbitrarily large constant clause–variable ratio

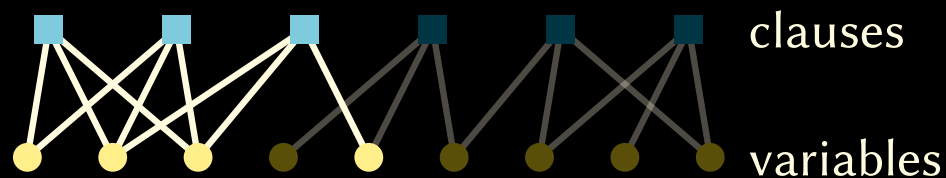


Power-Law Random k -SAT

Proof Complexity Lower Bounds

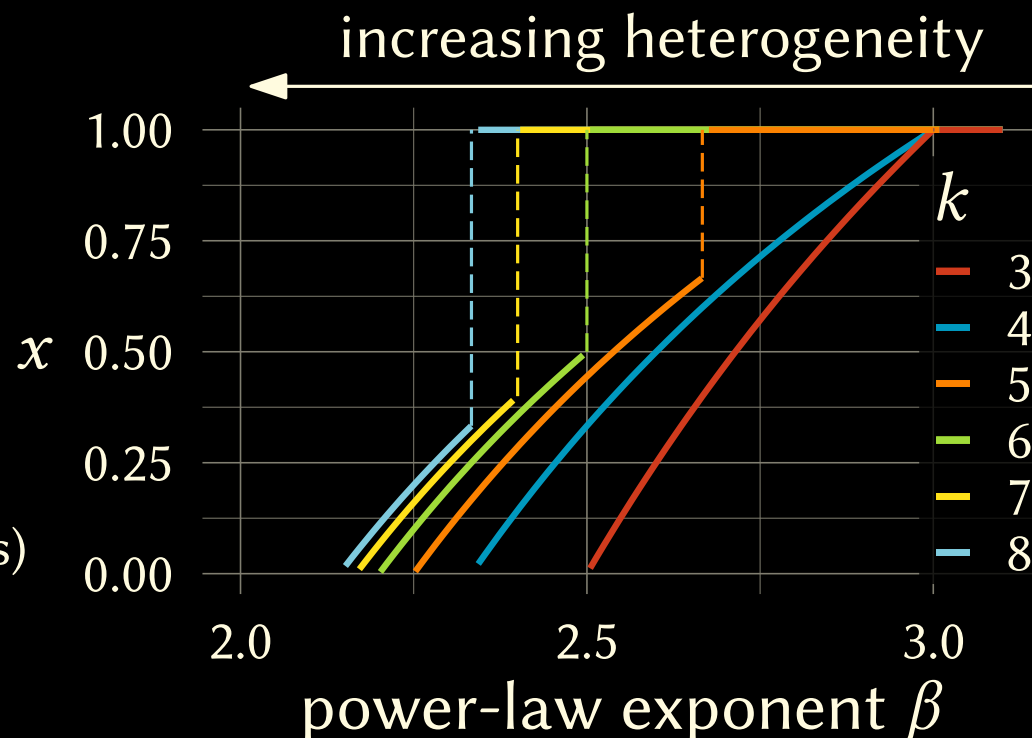
[Ben-Sasson, Wigderson 01]

- *resolution width* w = largest clause in resolution proof (min over all proofs)
- lower bounds on the resolution proof size \rightarrow lower bounds for solvers
 - \rightarrow CDCL
 - \rightarrow DPLL
- high bipartite expansion \Rightarrow high width



Our Lower Bounds (simplified)

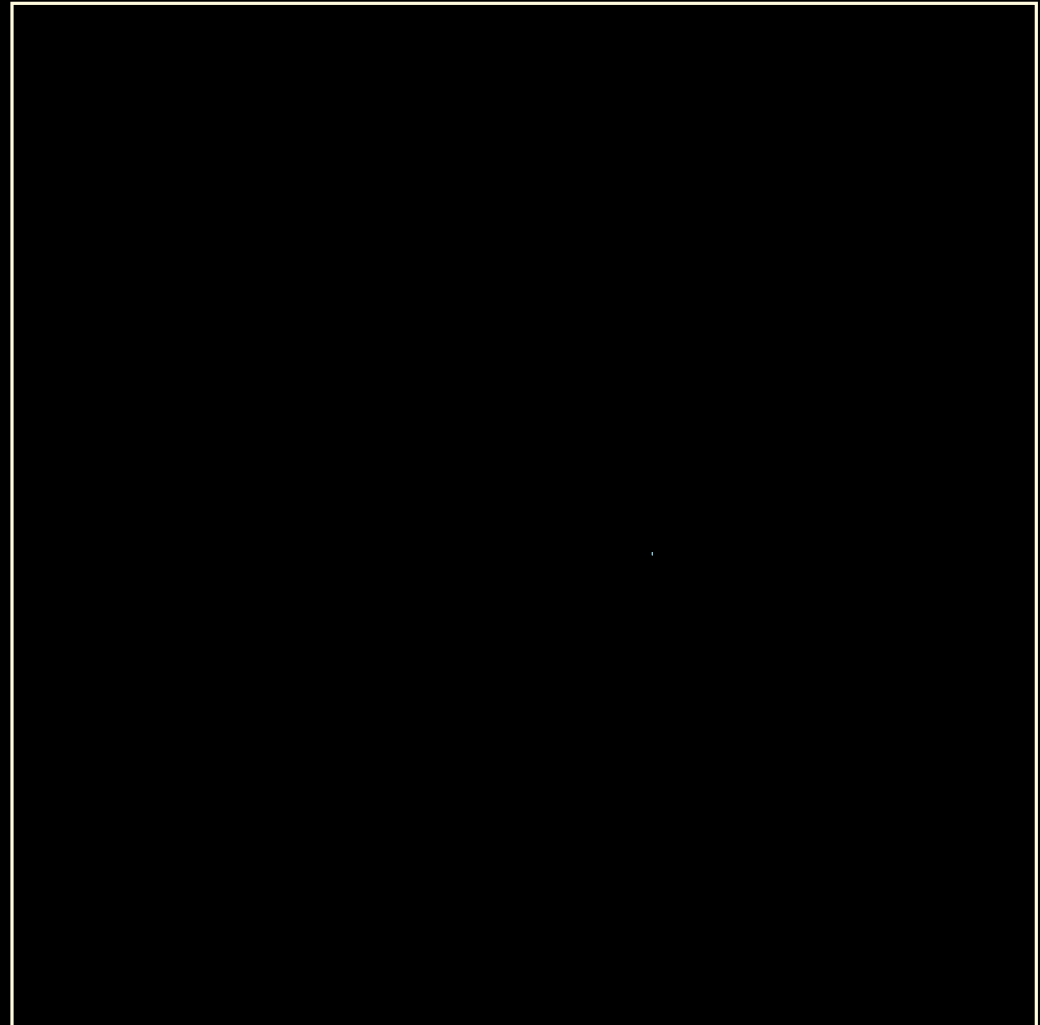
- resolution width $w \in \tilde{\Omega}(n^x)$
- holds for arbitrarily large constant clause–variable ratio
(super-constant ratio: slightly weaker bounds)



Geometric Random k -SAT

Simplifying Assumptions

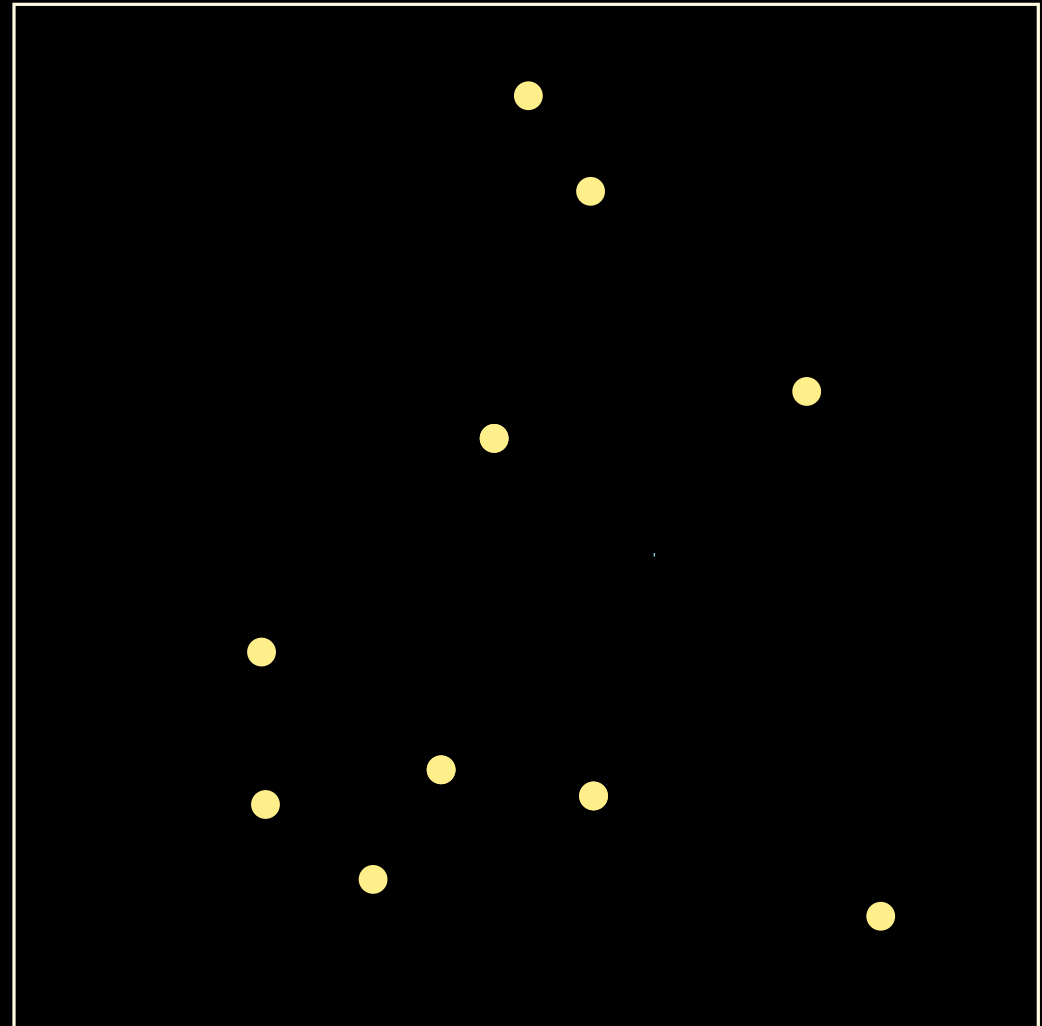
- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights



Geometric Random k -SAT

Simplifying Assumptions

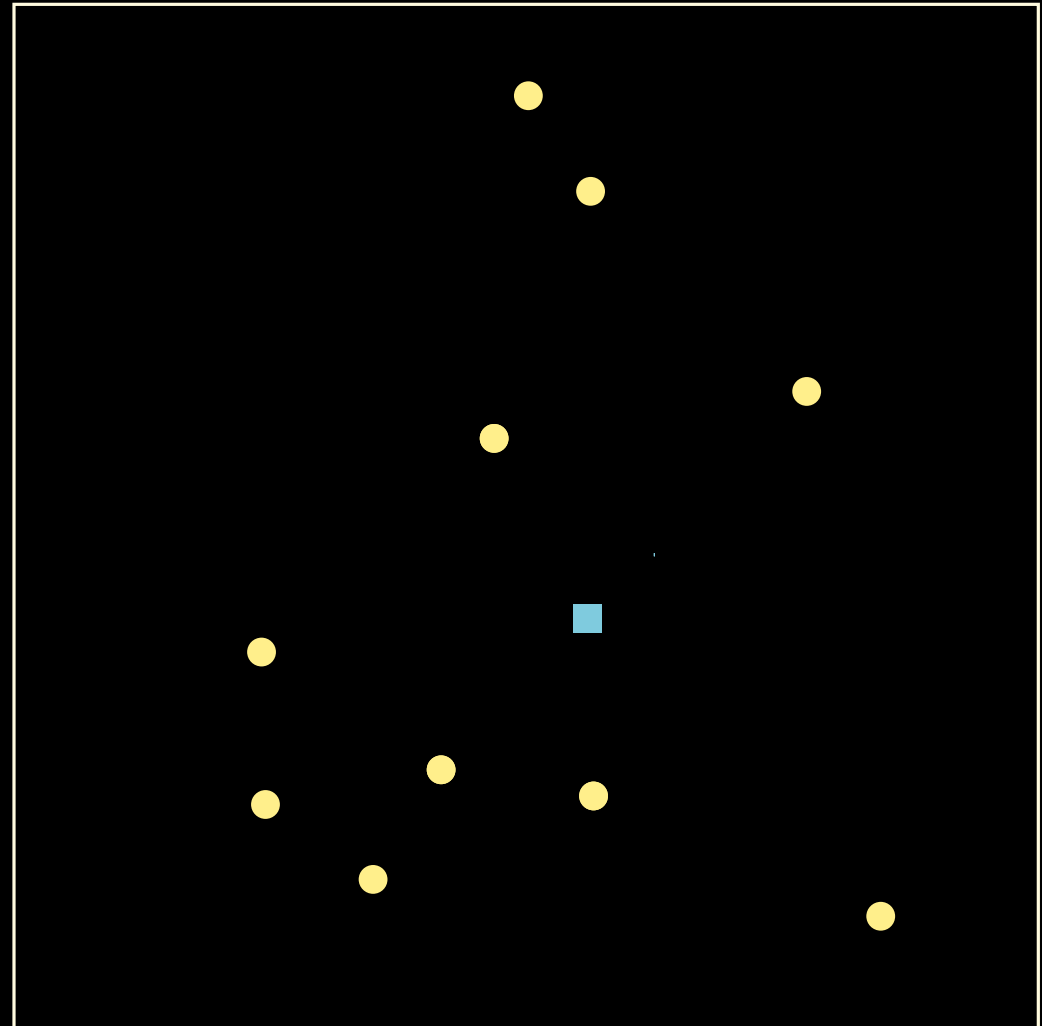
- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights



Geometric Random k -SAT

Simplifying Assumptions

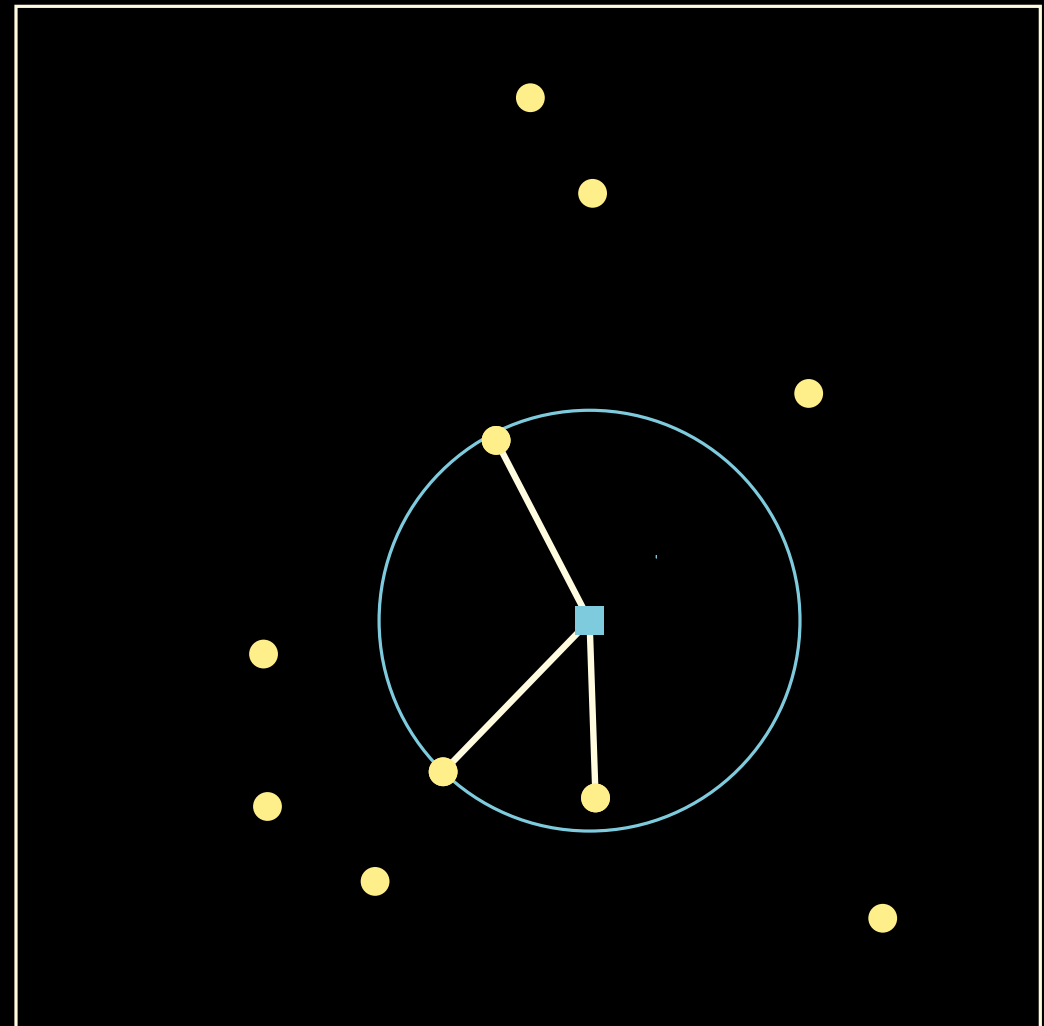
- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights



Geometric Random k -SAT

Simplifying Assumptions

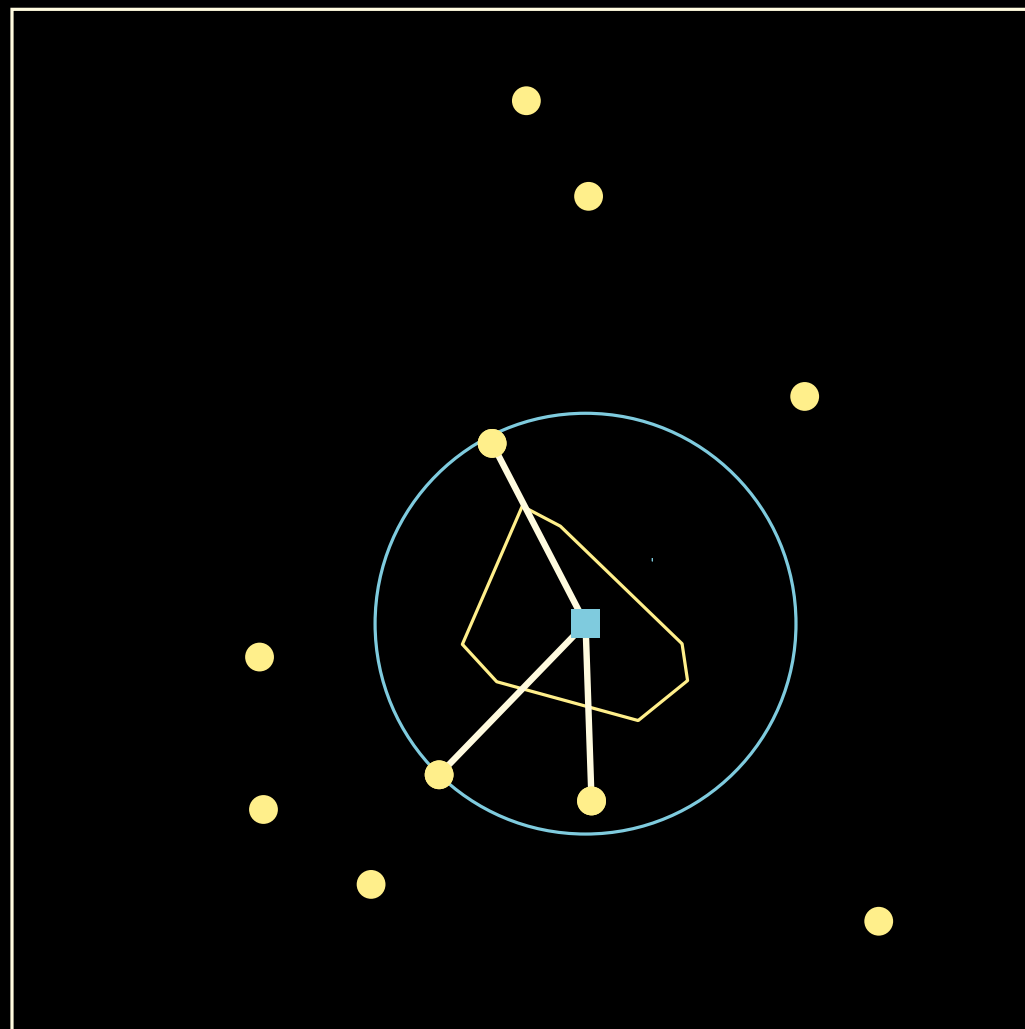
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

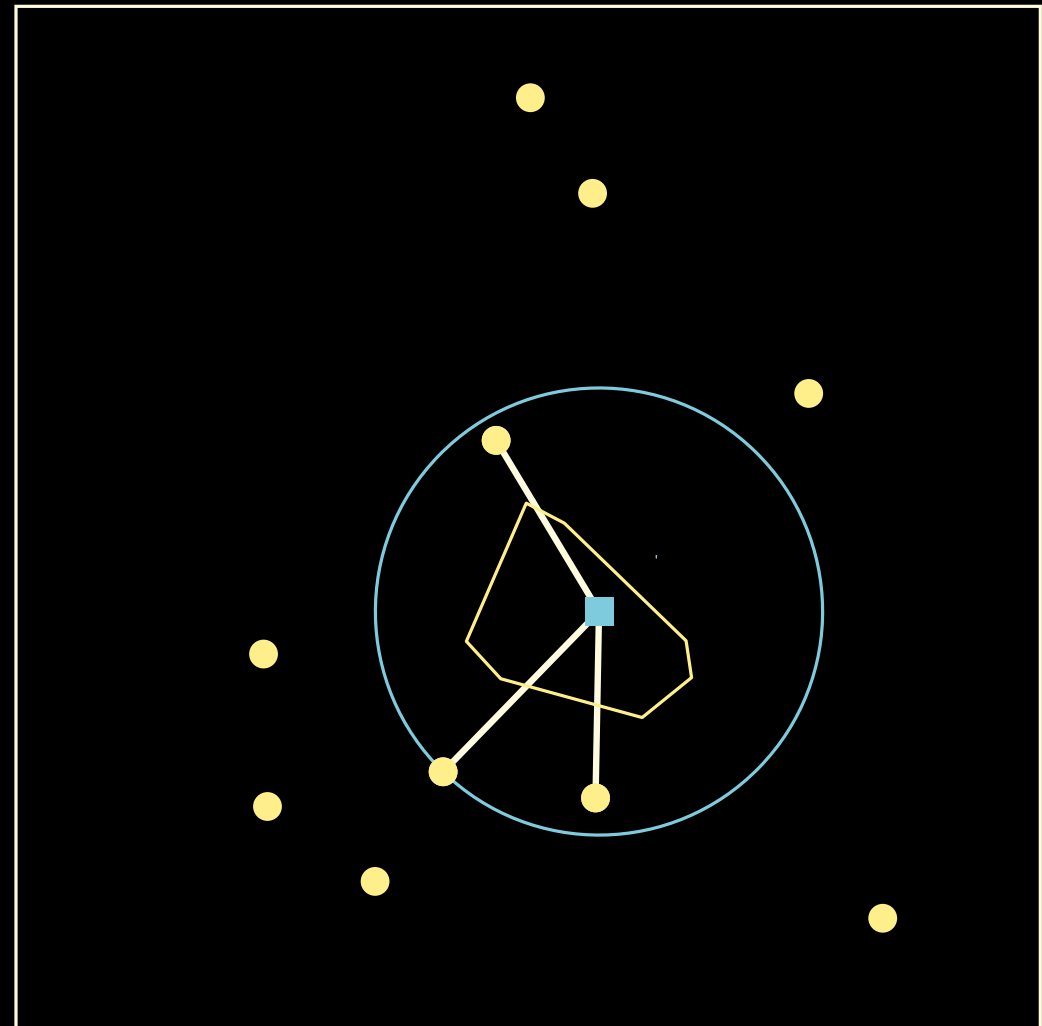
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

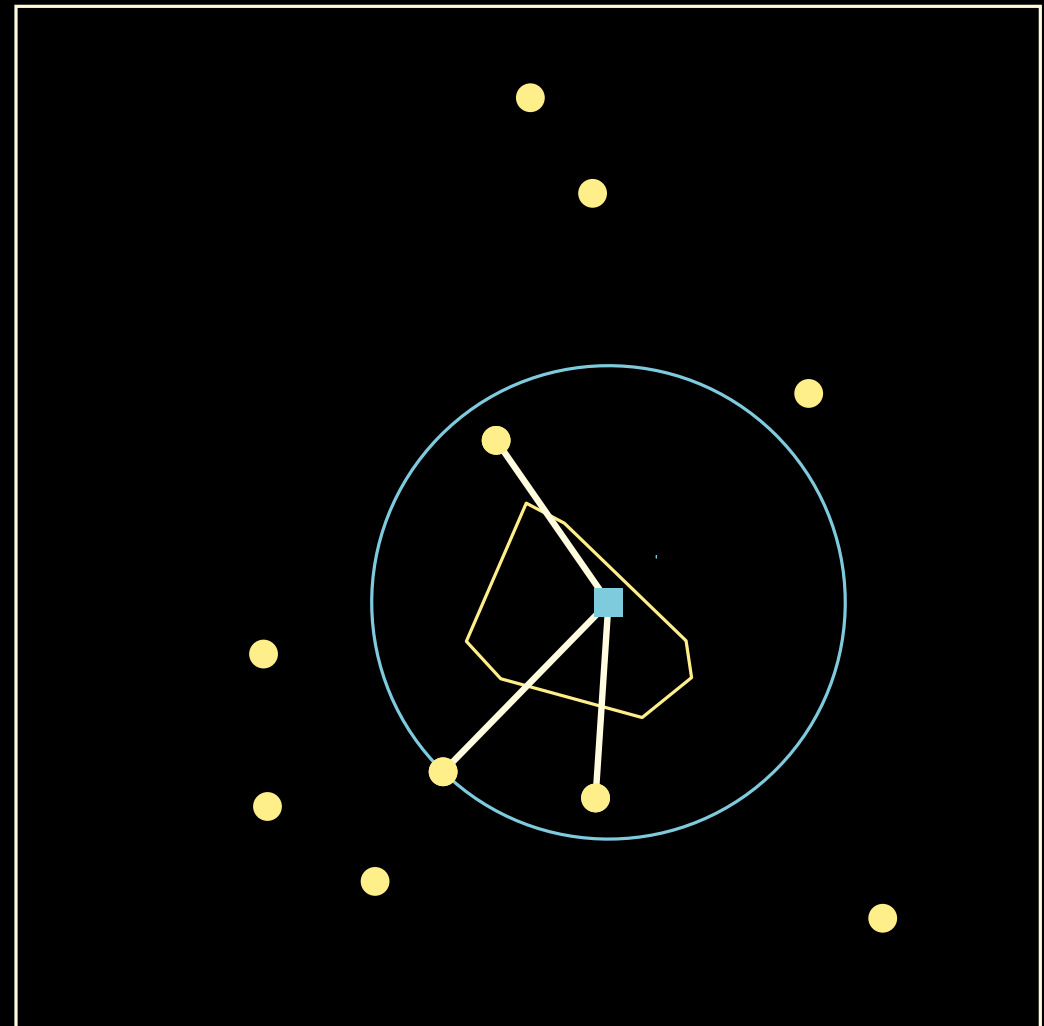
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

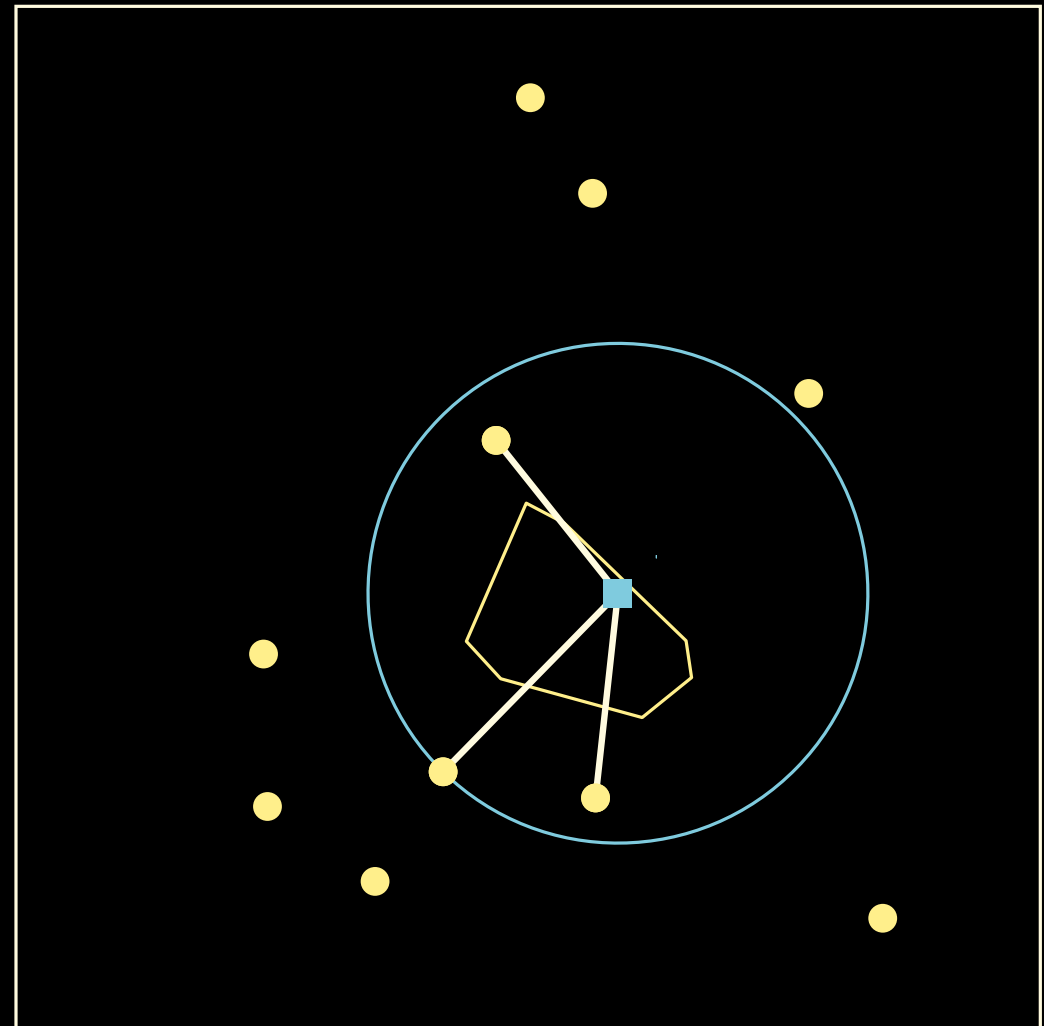
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

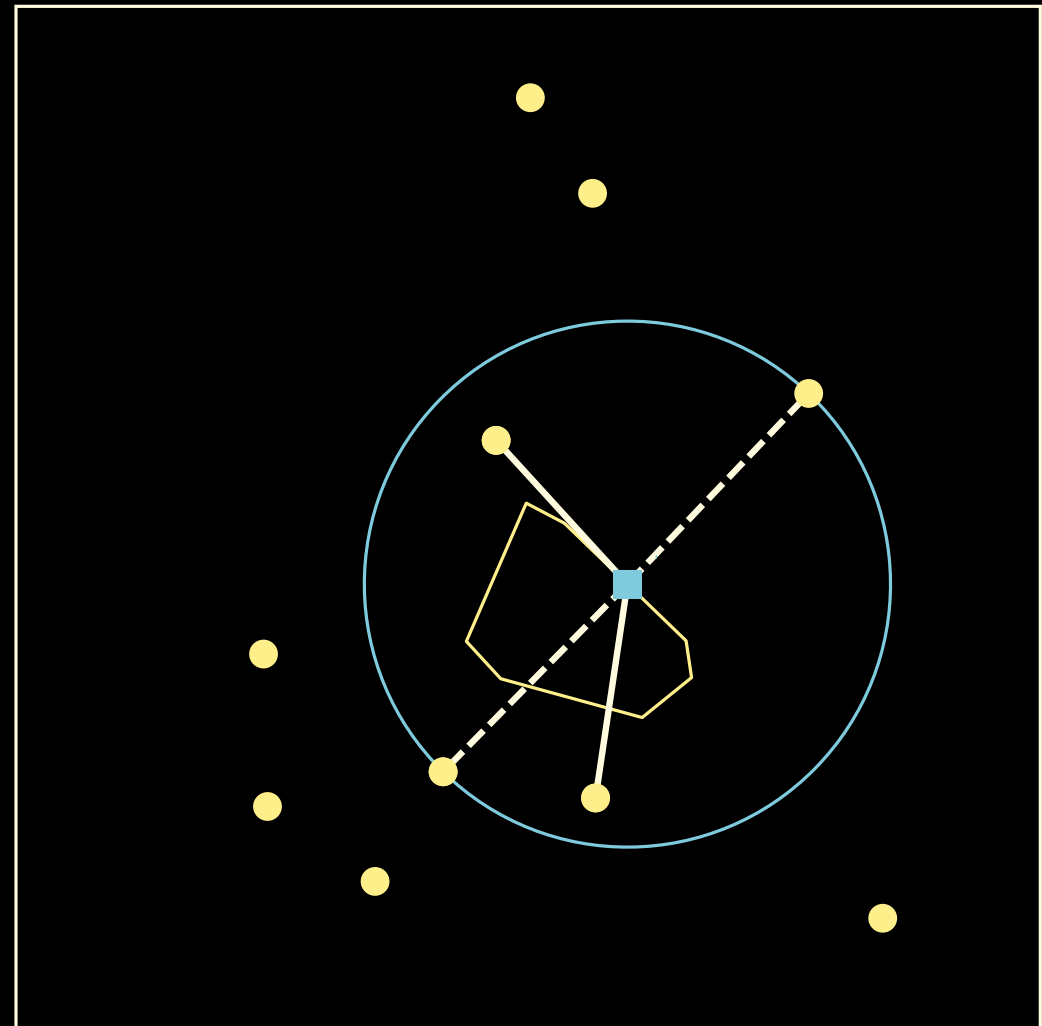
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

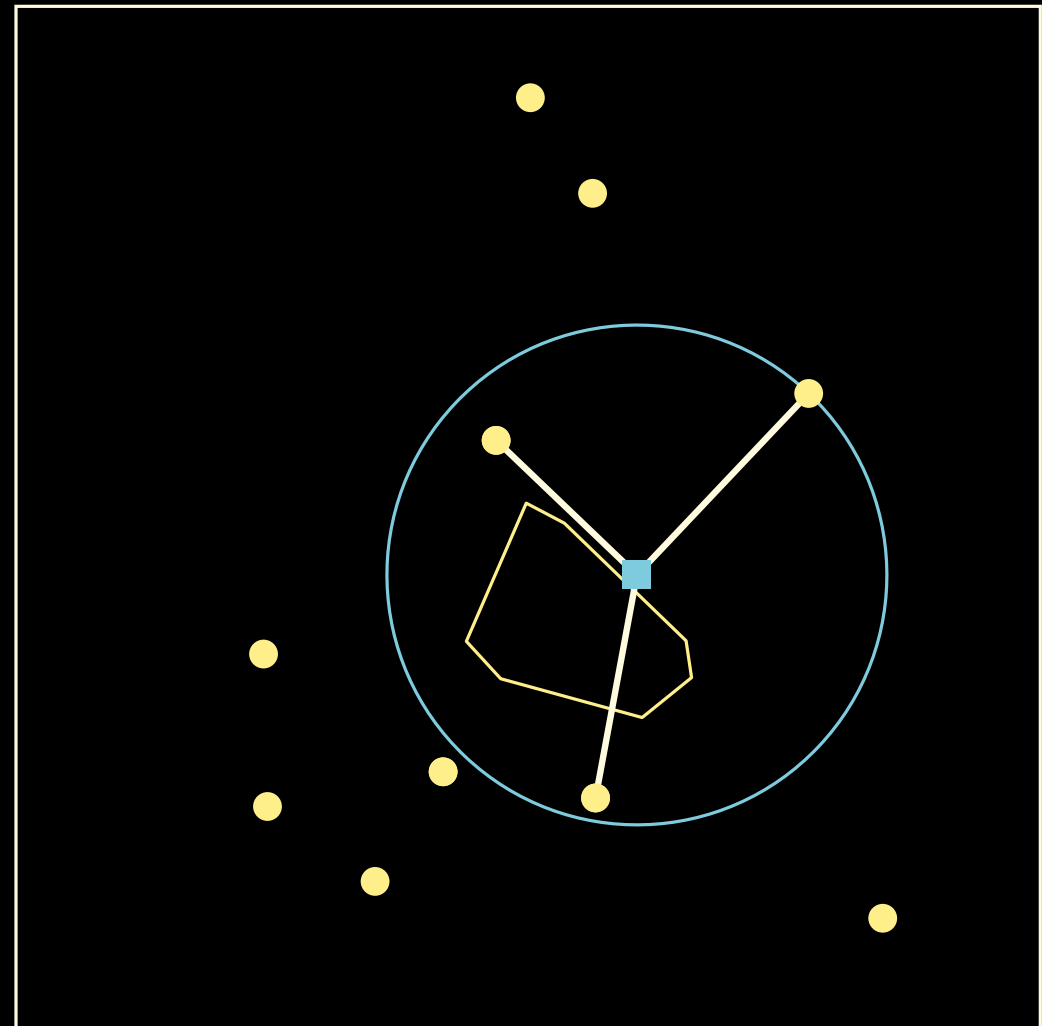
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

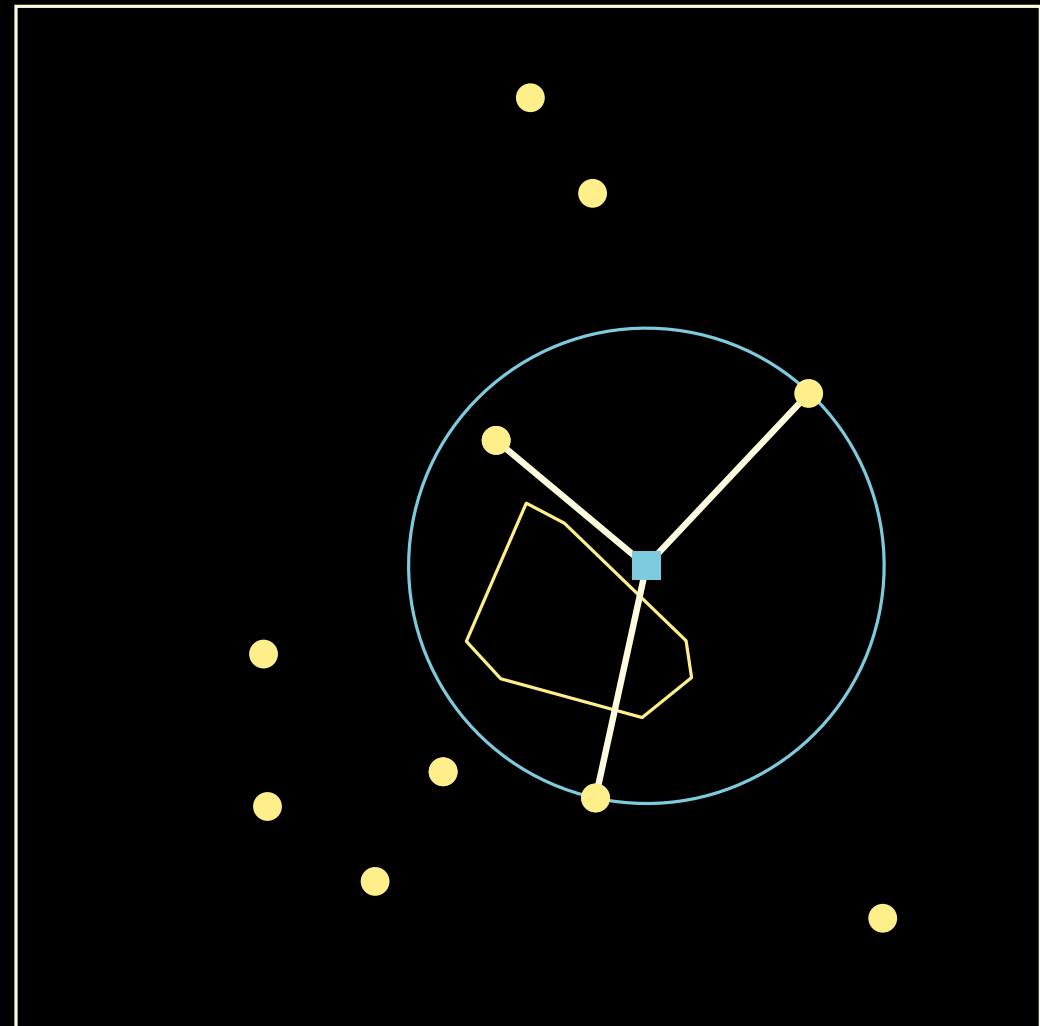
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

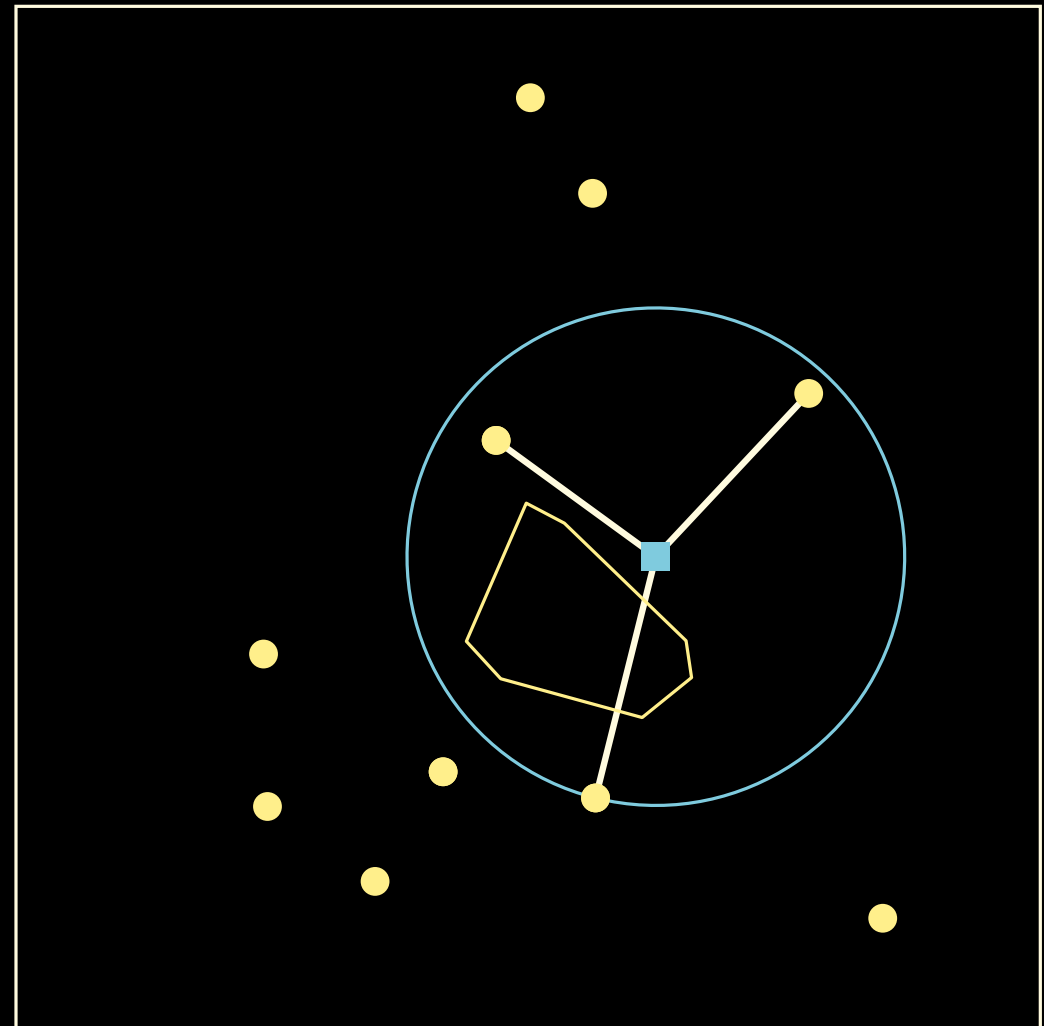
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

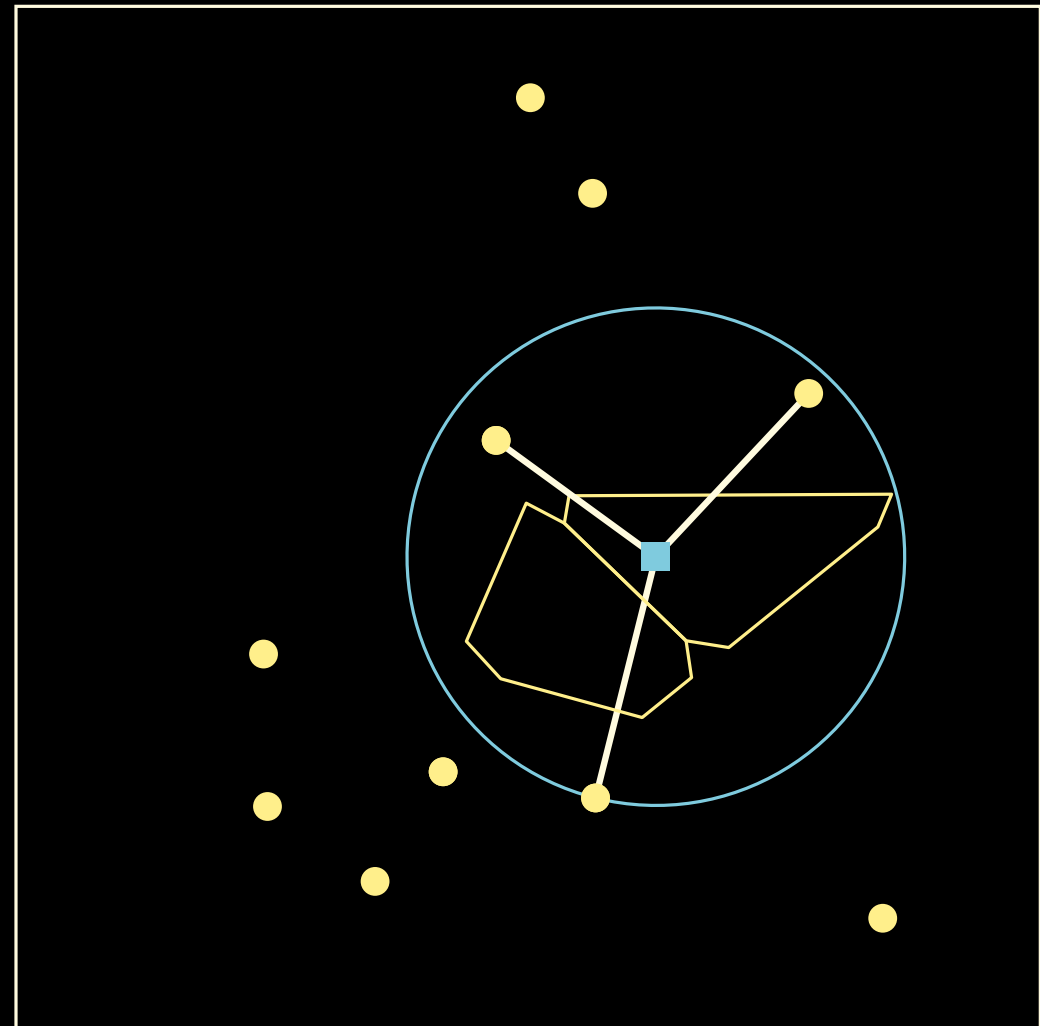
- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses



Geometric Random k -SAT

Simplifying Assumptions

- each clause contains the $k = 3$ closest variables ($T = 0$)
 - 2-dimensional ground space (square)
 - uniform weights
- variables ■ clauses

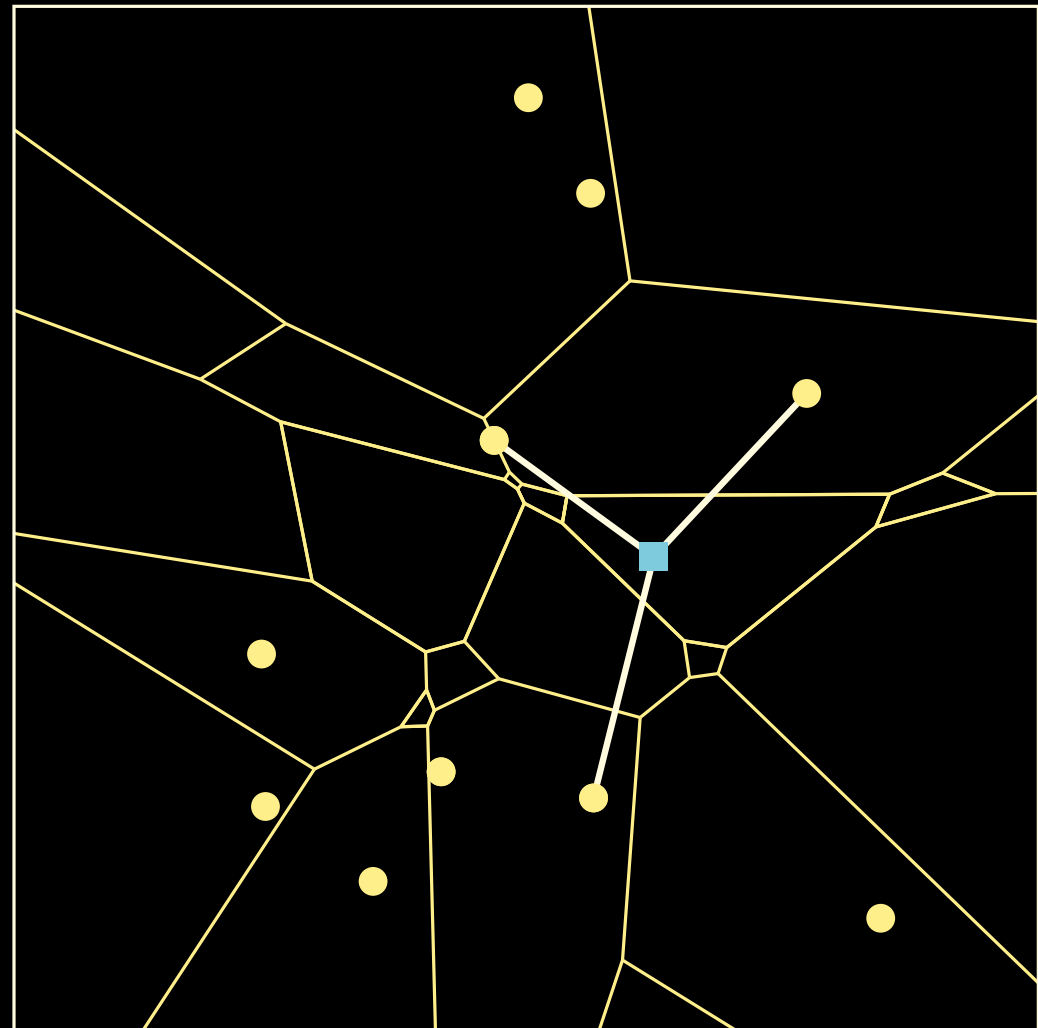


Geometric Random k -SAT

Simplifying Assumptions

- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights

Order- k Voronoi Diagram



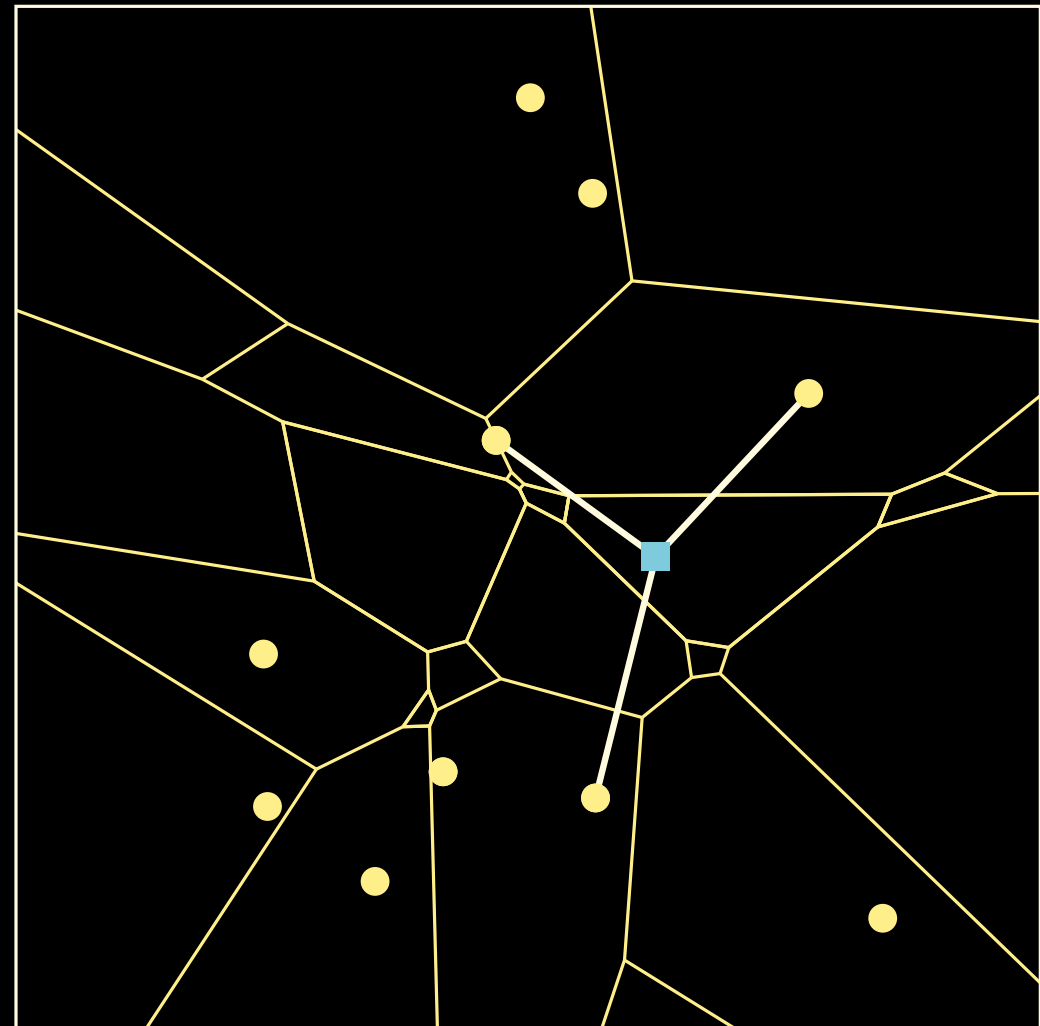
Geometric Random k -SAT

Simplifying Assumptions

- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights

Order- k Voronoi Diagram

- $O(n)$ regions [Bohler et al. 15]



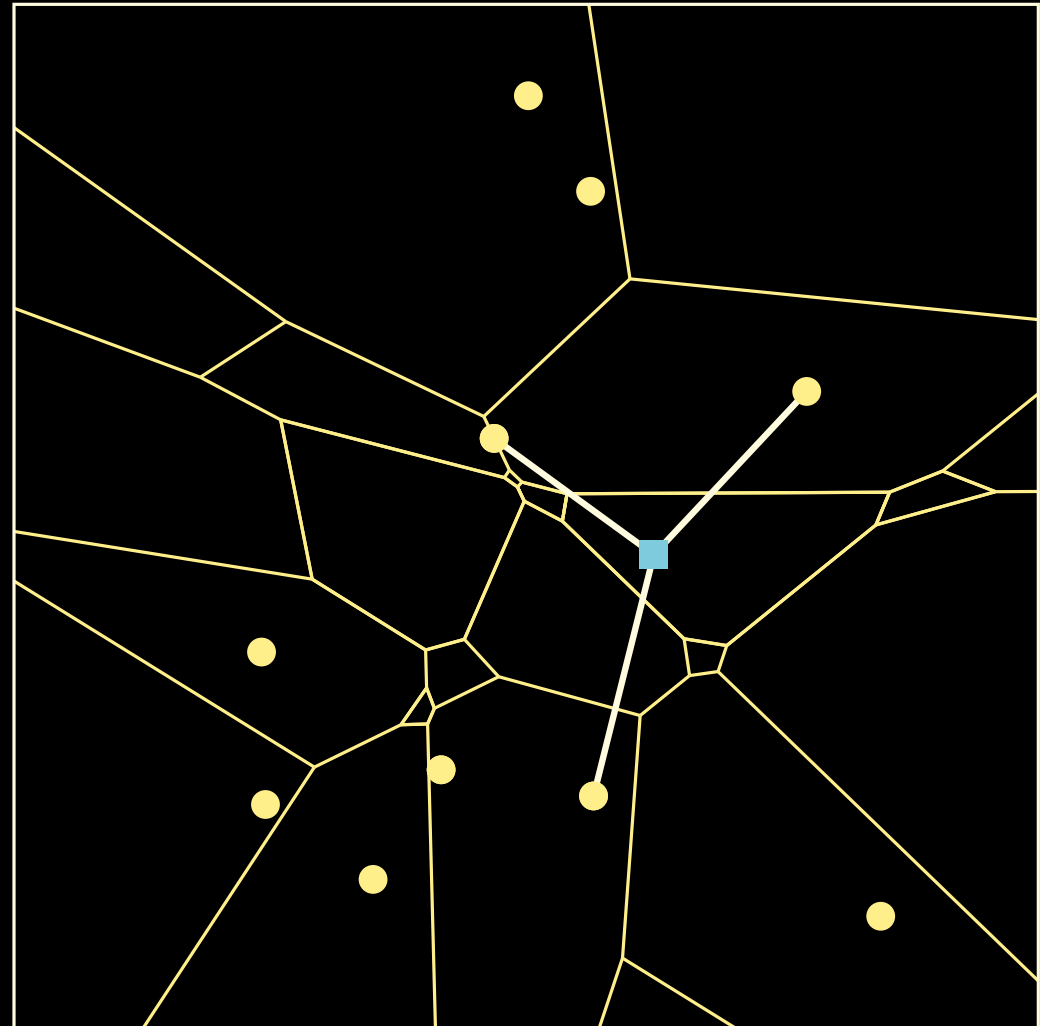
Geometric Random k -SAT

Simplifying Assumptions

- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights

Order- k Voronoi Diagram

- $O(n)$ regions [Bohler et al. 15]
- $\Omega\left(\frac{n}{\text{polylog } n}\right)$ clauses [Raab, Steger 98]
- (a.a.s.) \Rightarrow region with $\Omega\left(\frac{\log n}{\log \log n}\right)$ clauses



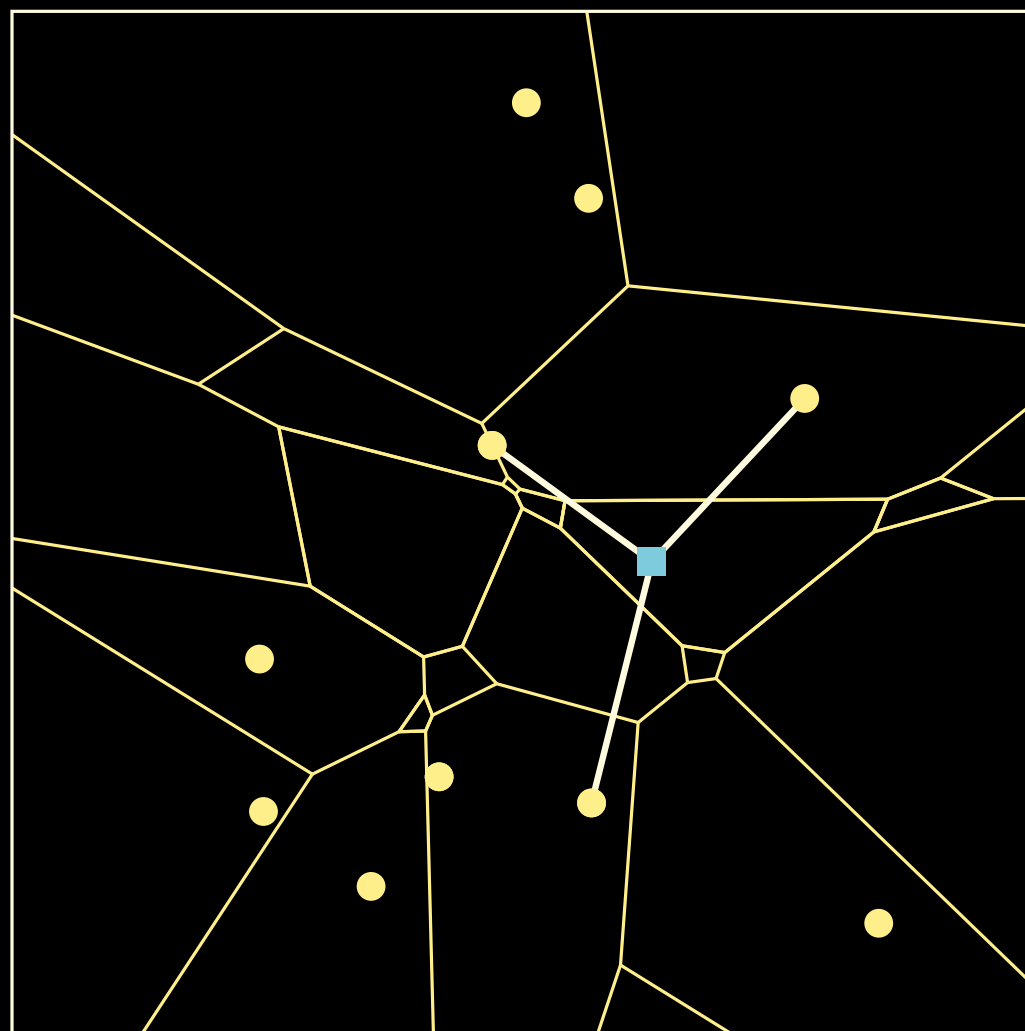
Geometric Random k -SAT

Simplifying Assumptions

- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights

Order- k Voronoi Diagram

- $O(n)$ regions [Bohler et al. 15]
- $\Omega\left(\frac{n}{\text{polylog } n}\right)$ clauses [Raab, Steger 98]
(a.a.s.)
 \implies region with $\Omega\left(\frac{\log n}{\log \log n}\right)$ clauses
- k -tuple of vars with $\omega(1)$ clauses



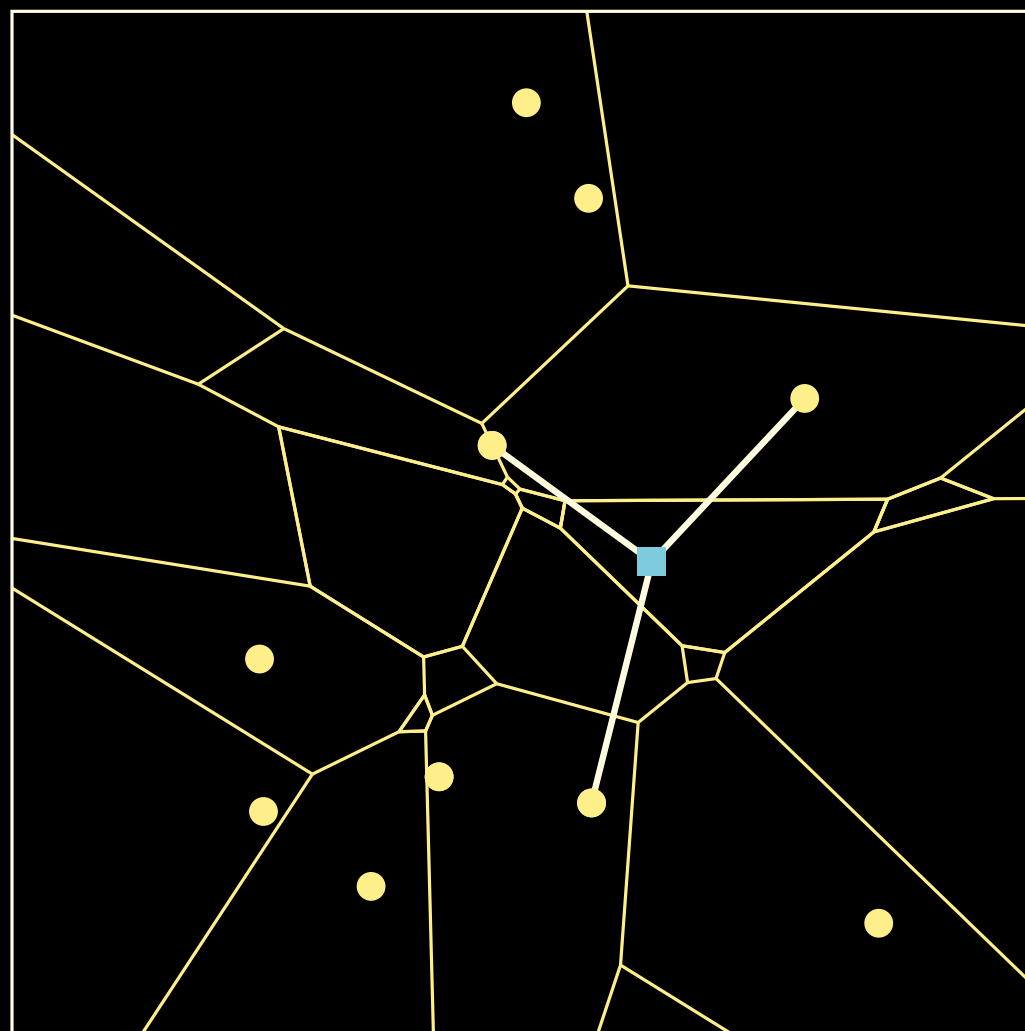
Geometric Random k -SAT

Simplifying Assumptions

- each clause contains the $k = 3$ closest variables ($T = 0$)
- 2-dimensional ground space (square) ● variables ■ clauses
- uniform weights

Order- k Voronoi Diagram

- $O(n)$ regions [Bohler et al. 15]
- $\Omega\left(\frac{n}{\text{polylog } n}\right)$ clauses [Raab, Steger 98]
- (a.a.s.) \Rightarrow region with $\Omega\left(\frac{\log n}{\log \log n}\right)$ clauses
- k -tuple of vars with $\omega(1)$ clauses
- unsatisfiable subformula of constant size



Geometric Random k -SAT – General Setting

Weighted Variables & Higher Dimensions

■ $\Omega(n^2)$ in each of these settings:

□ weighted, 2D, $k = 3$

[Aurenhammer, Edelsbrunner 84]

□ unweighted, 3D, $k = 4$

[Klee 80][Seidel 87]

Geometric Random k -SAT – General Setting

Weighted Variables & Higher Dimensions

- $\Omega(n^2)$ in each of these settings:
 - weighted, 2D, $k = 3$ [Aurenhammer, Edelsbrunner 84]
 - unweighted, 3D, $k = 4$ [Klee 80][Seidel 87]
- *our result*: random positions \Rightarrow only $O(W)$ regions for ($W =$ total weight)
 - arbitrary constant k
 - arbitrary constant dimension
 - p -norm for $p \in \mathbb{N}^+ \cup \{\infty\}$

Geometric Random k -SAT – General Setting

Weighted Variables & Higher Dimensions

- $\Omega(n^2)$ in each of these settings:
 - weighted, 2D, $k = 3$ [Aurenhammer, Edelsbrunner 84]
 - unweighted, 3D, $k = 4$ [Klee 80][Seidel 87]
 - *our result*: random positions \Rightarrow only $O(W)$ regions for ($W =$ total weight)
 - arbitrary constant k
 - arbitrary constant dimension
 - p -norm for $p \in \mathbb{N}^+ \cup \{\infty\}$
- \Rightarrow previous argument applies to weighted variables and higher dimensions

Geometric Random k -SAT – General Setting

Weighted Variables & Higher Dimensions

- $\Omega(n^2)$ in each of these settings:
 - weighted, 2D, $k = 3$ [Aurenhammer, Edelsbrunner 84]
 - unweighted, 3D, $k = 4$ [Klee 80][Seidel 87]
 - *our result*: random positions \Rightarrow only $O(W)$ regions for ($W =$ total weight)
 - arbitrary constant k
 - arbitrary constant dimension
 - p -norm for $p \in \mathbb{N}^+ \cup \{\infty\}$
- \Rightarrow previous argument applies to weighted variables and higher dimensions

Higher Temperatures

- relax requirement that each clause contains the k closest variables

Geometric Random k -SAT – General Setting

Weighted Variables & Higher Dimensions

- $\Omega(n^2)$ in each of these settings:
 - weighted, 2D, $k = 3$ [Aurenhammer, Edelsbrunner 84]
 - unweighted, 3D, $k = 4$ [Klee 80][Seidel 87]
 - *our result*: random positions \Rightarrow only $O(W)$ regions for ($W =$ total weight)
 - arbitrary constant k
 - arbitrary constant dimension
 - p -norm for $p \in \mathbb{N}^+ \cup \{\infty\}$
- \Rightarrow previous argument applies to weighted variables and higher dimensions

Higher Temperatures

- relax requirement that each clause contains the k closest variables
- *our result*: enough clauses contain the k closest variables if $T < 1$
(given some reasonable technical assumptions)

Geometric Random k -SAT – General Setting

Weighted Variables & Higher Dimensions

- $\Omega(n^2)$ in each of these settings:
 - weighted, 2D, $k = 3$ [Aurenhammer, Edelsbrunner 84]
 - unweighted, 3D, $k = 4$ [Klee 80][Seidel 87]
 - *our result*: random positions \Rightarrow only $O(W)$ regions for ($W =$ total weight)
 - arbitrary constant k
 - arbitrary constant dimension
 - p -norm for $p \in \mathbb{N}^+ \cup \{\infty\}$
- \Rightarrow previous argument applies to weighted variables and higher dimensions

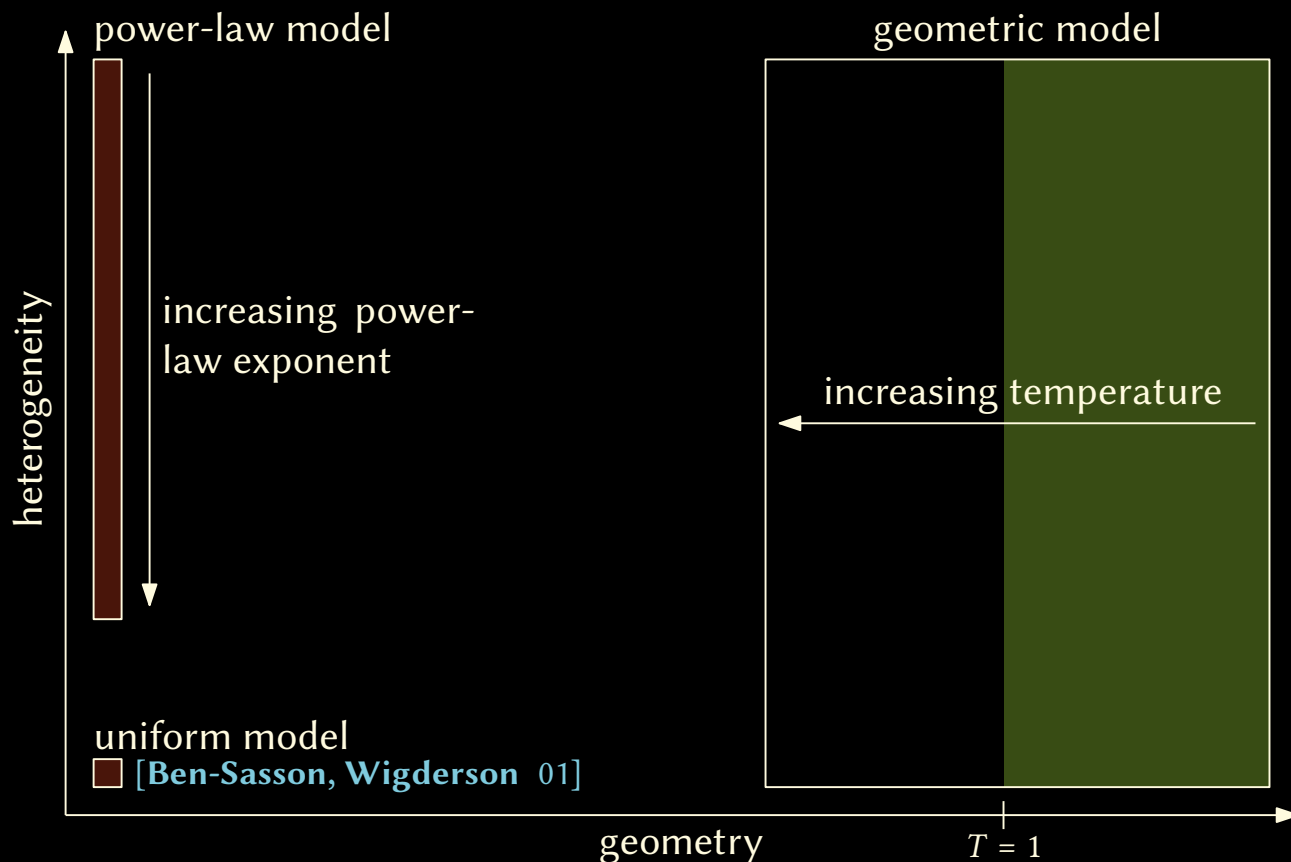
Higher Temperatures

- relax requirement that each clause contains the k closest variables
 - *our result*: enough clauses contain the k closest variables if $T < 1$
(given some reasonable technical assumptions)
- \Rightarrow previous argument applies to higher-temperature regime

Conclusion: Why are SAT-Solvers so Fast?

The Big Picture

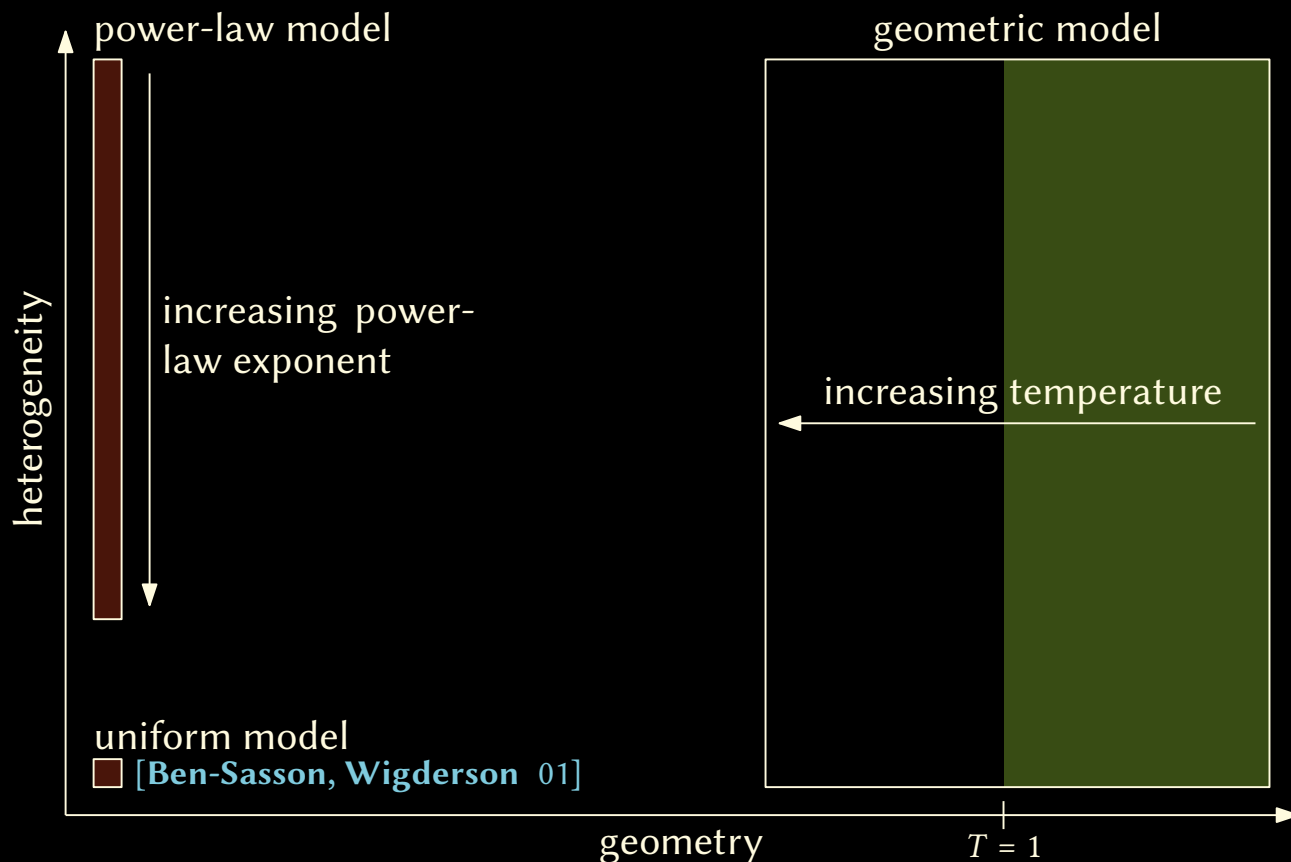
- heterogeneous variables:
not very helpful
- underlying geometry:
very helpful



Conclusion: Why are SAT-Solvers so Fast?

The Big Picture

- heterogeneous variables:
not very helpful
- underlying geometry:
very helpful

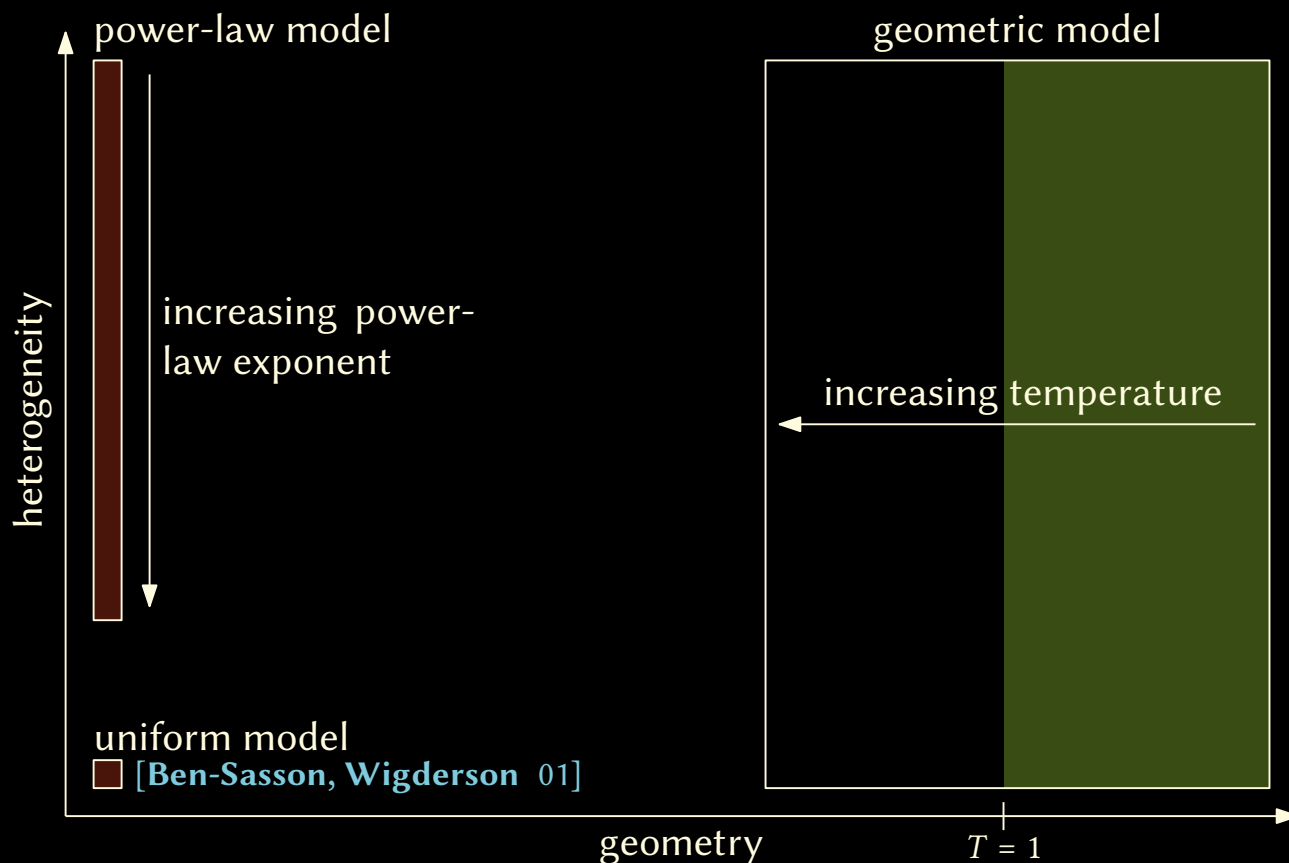


Are We Done Now?

Conclusion: Why are SAT-Solvers so Fast?

The Big Picture

- heterogeneous variables:
not very helpful
- underlying geometry:
very helpful



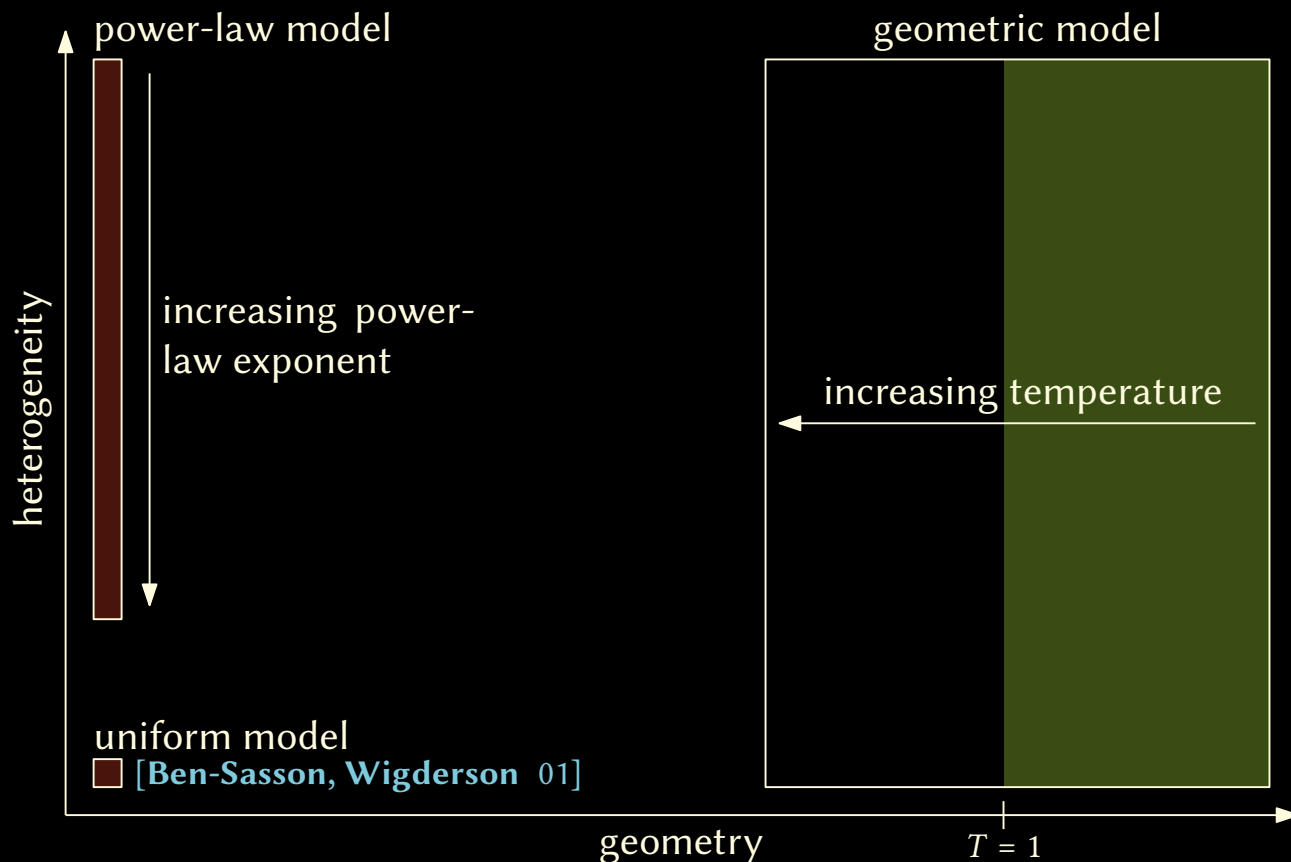
Are We Done Now?

- the geometric instances are unrealistically easy (at least in the limit)

Conclusion: Why are SAT-Solvers so Fast?

The Big Picture

- heterogeneous variables:
not very helpful
- underlying geometry:
very helpful



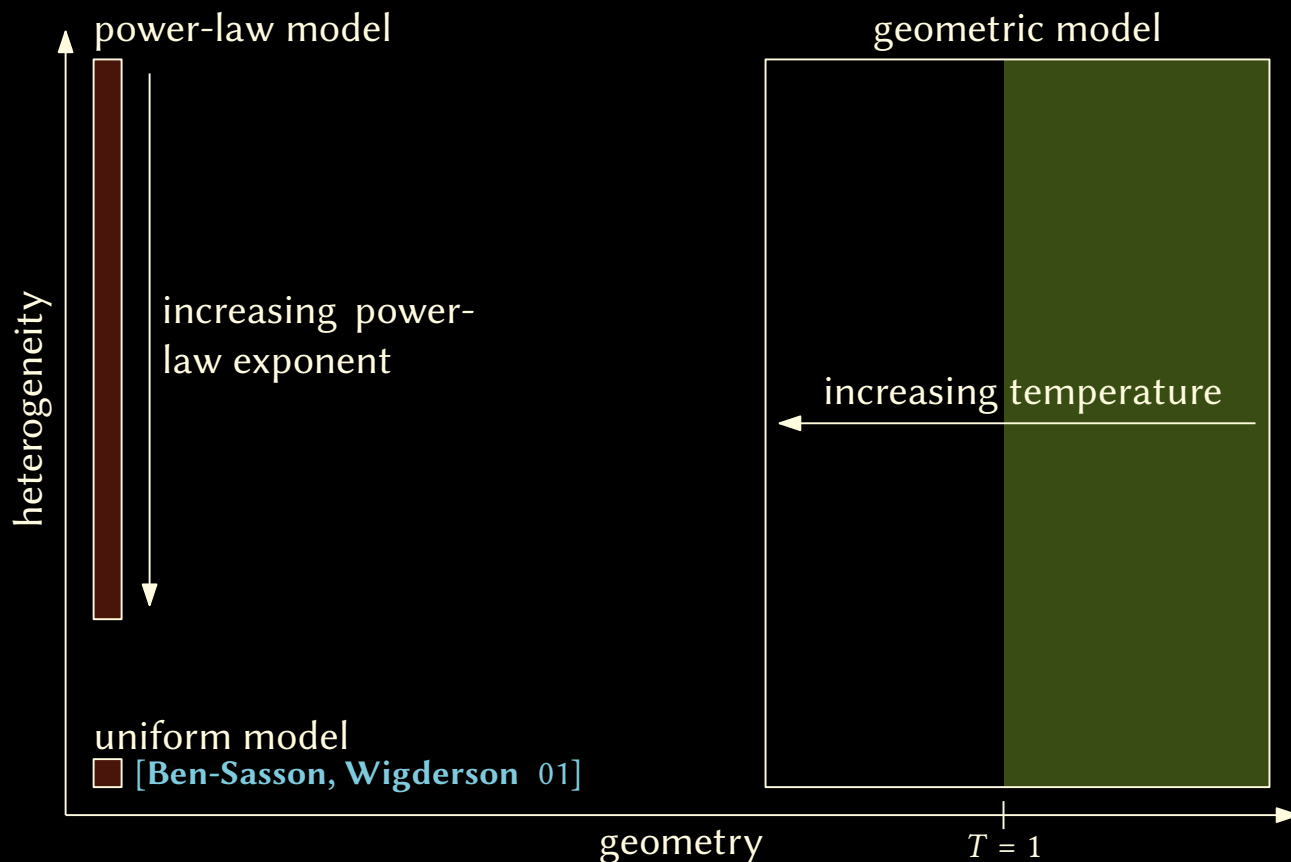
Are We Done Now?

- the geometric instances are unrealistically easy (at least in the limit)
- first theoretical evidence that underlying geometry helps
→ further investigations could help close the theory–practice gap

Conclusion: Why are SAT-Solvers so Fast?

The Big Picture

- heterogeneous variables:
not very helpful
- underlying geometry:
very helpful



Are We Done Now?

- the geometric instances are unrealistically easy (at least in the limit)
- first theoretical evidence that underlying geometry helps
→ further investigations could help close the theory–practice gap
- geometric model provides an easy average case
→ the truth might lie between average and worst case