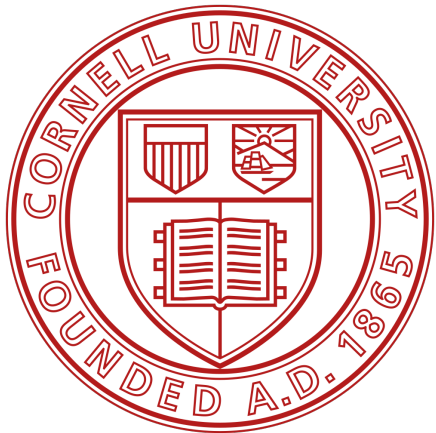# Combinatorial Generation Problems

Guillaume Perez

Motivations

- **Combinatorial problems are everywhere**

Knapsack problems

Dams placement in amazon

Pickup and deliveries

Music and text generation

Combinatorial Optimization Problems

- Finding an **optimal** object from a **finite set** of objects
  - Example: Finding the shortest path in a graph

- Decision variables are **discrete**
  - Different from continuous optimization

Knapsack problems

$$\max \sum_{i \in S} x_i v_i$$
$$\text{subject to: } \sum_{i \in S} x_i w_i \leq W$$

Easy with continuous x_i variables

Hard with discrete x_i variables

The end?

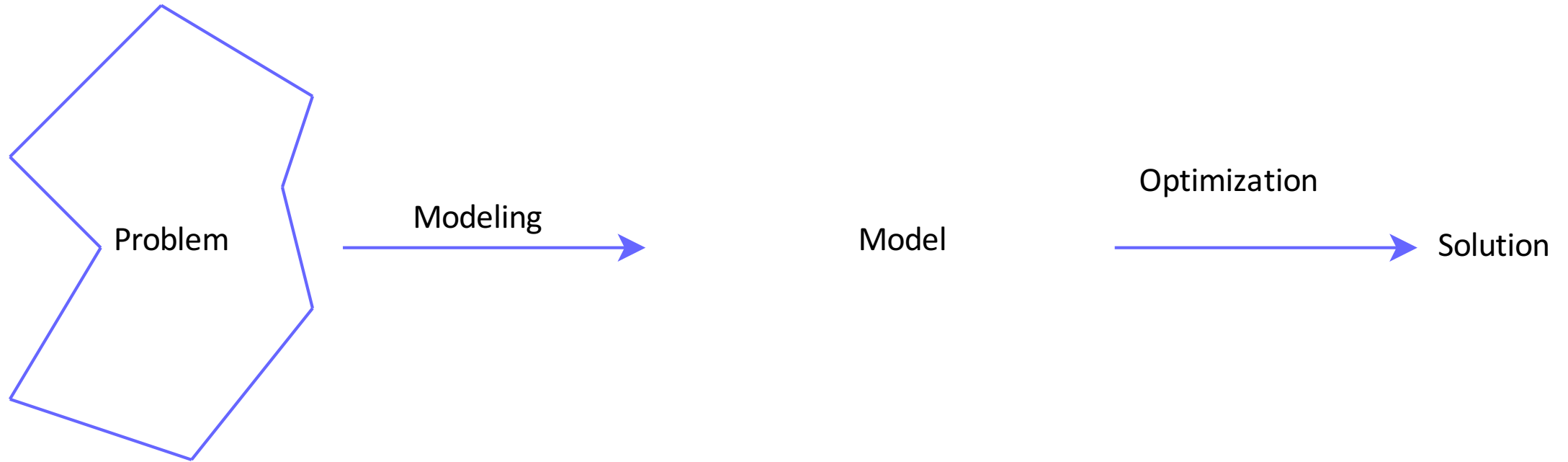# Thank You

- Questions?
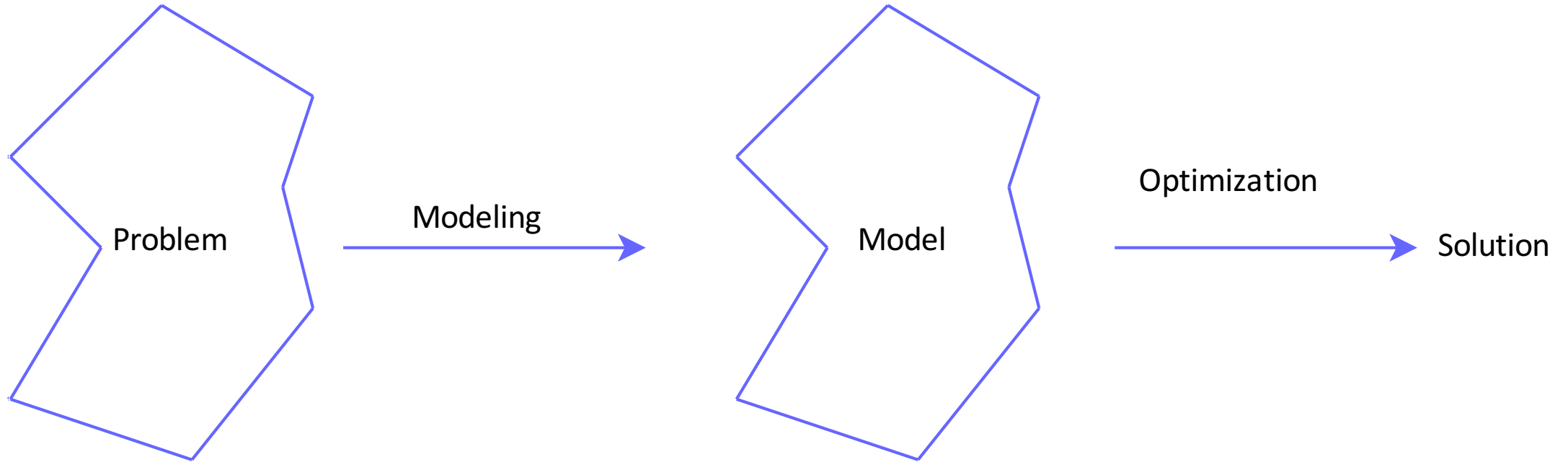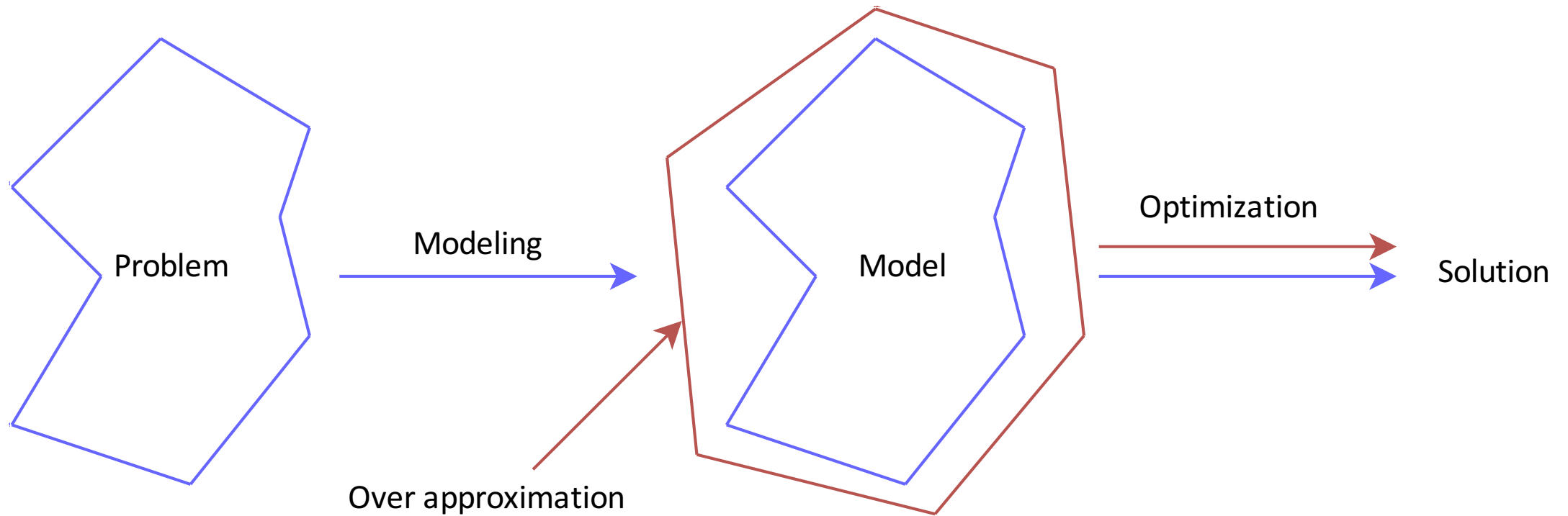
On Solving Combinatorial Problems
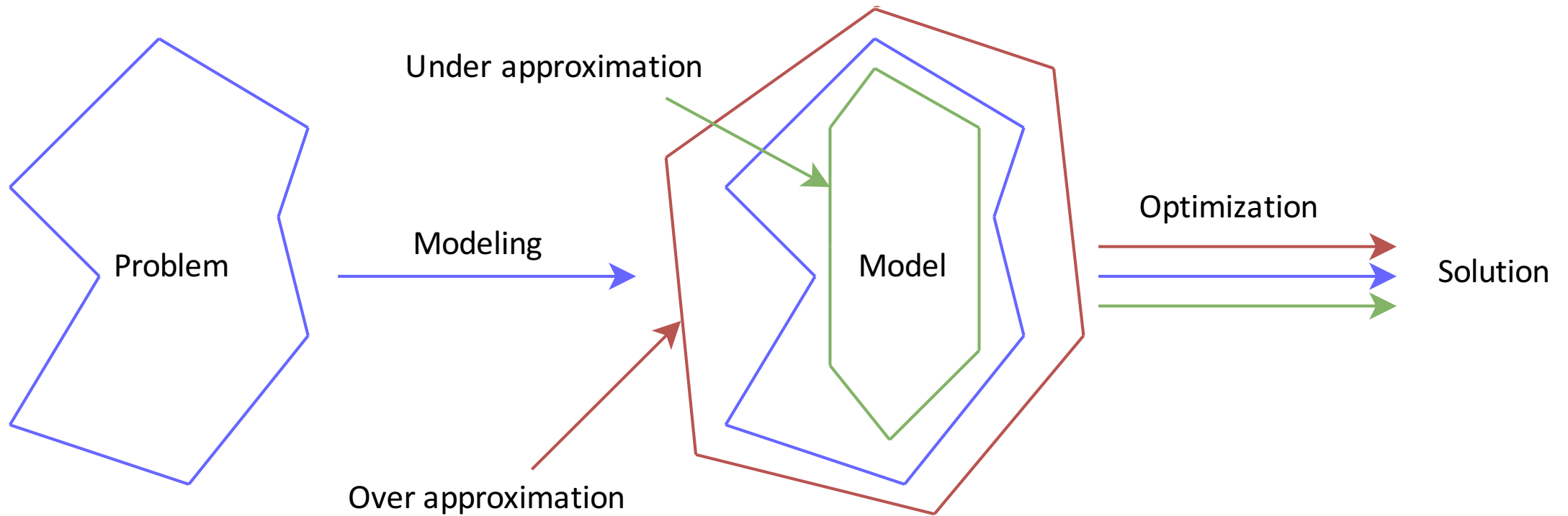
Problem $\xrightarrow{\text{Modeling}}$ Model $\xrightarrow{\text{Optimization}}$ Solution

On Solving Combinatorial Problems

Problem

Modeling →

Model

Optimization →

Solution

On Solving Combinatorial Problems

Problem

Modeling

Model

Optimization

Solution

Over approximation

On Solving Combinatorial Problems

Under approximation

Over approximation

Problem

Modeling

Model

Optimization

Solution

Modeling is an Art, but optimizing is a science

Sequence Generation

- Generate a **sequence**

x_1 x_2 x_3 x_4 x_5 x_6 x_7
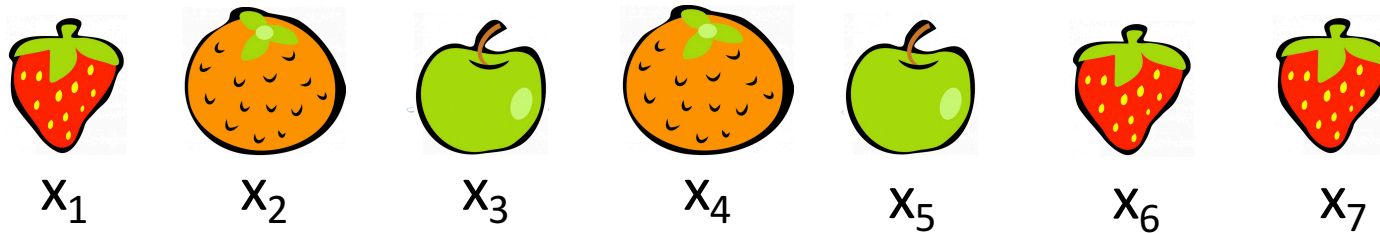
- But with **constraints**
  - #orange < 3
  - #apple > 1
  - Finish by strawberry

Product line scheduling
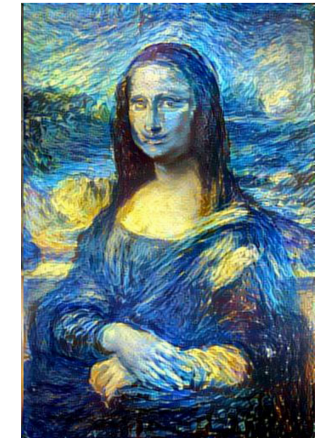
Generation under constraints

- What if I want to generate a text with rhymes ?

Find a rhyme with combinator**ial**
That doesn't sound artific**ial**

- Text and music using someone's "style" ?

- Music without plagiarism ?

Let it go, let it go
Can't hold it back anymore

Example with image

- Poetry with a given number of syllables ?

In the month of the long decline of roses
I, beholding the summer dead before me,
Set my face to the sea and journeyed silent,

Algernon Charles Swinburne

Generation under constraints

- **What if I want to generate a text with rhymes ?**

Find a rhyme with combinator**ial**
That doesn't sound artific**ial**

- Text and music using someone's "style" ?

- Music without plagiarism ?

Let it go, let it go
Can't hold it back anymore

Example with image

- Poetry with a given number of syllables ?

In the month of the long decline of roses
I, beholding the summer dead before me,
Set my face to the sea and journeyed silent,
Algernon Charles Swinburne

Sequence with Rhymes

- Example, rhymes between each two lines

I be Puerto Rican day parade floa**tin'**
That Benz Marina Del Rey coas**tin'**
She in school to be a real estate ag**ent**
Last month I helped her with the car paym**ent**

Famous – Kanye West   https://youtu.be/1wYXSxCvN68?t=73

- How to **express** that?
  - 1 variable per syllable?
  - **1 variable per word?**

$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad \mathbf{x_5}$$

$$x_6 \quad x_7 \quad x_8 \quad x_9 \quad \mathbf{x_{10}}$$

Constraint: $\mathbf{x_5}$ and $\mathbf{x_{10}}$ must rhyme.

Sequence with Rhymes

- Given a **corpus**

$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad \mathbf{x_5}$$

$$x_6 \quad x_7 \quad x_8 \quad x_9 \quad \mathbf{x_{10}}$$

- Generate sentences that **rhyme**

Constraint: $\mathbf{x_5}$ and $\mathbf{x_{10}}$ must rhyme.

- What values for the X_i with $i \neq \{5,10\}$?

Section 2 in exercises

- What values for the X_i with $i = \{5,10\}$?
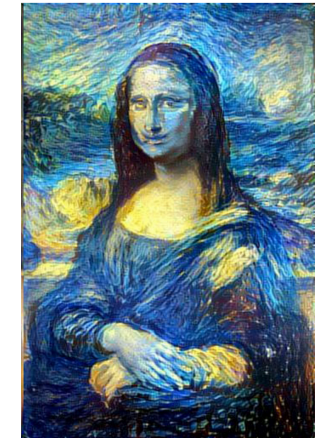
Generation under constraints

- What if I want to generate a text with rhymes ?

Find a rhyme with combinator**ial**
That doesn't sound artific**ial**

- **Text and music using someone's "style" ?**

- Music without plagiarism ?

Let it go, let it go
Can't hold it back anymore

Example with image

- Poetry with a given number of syllables ?

In the month of the long decline of roses
I, beholding the summer dead before me,
Set my face to the sea and journeyed silent,
Algernon Charles Swinburne

Sequence with Style

- What's someone **style**?
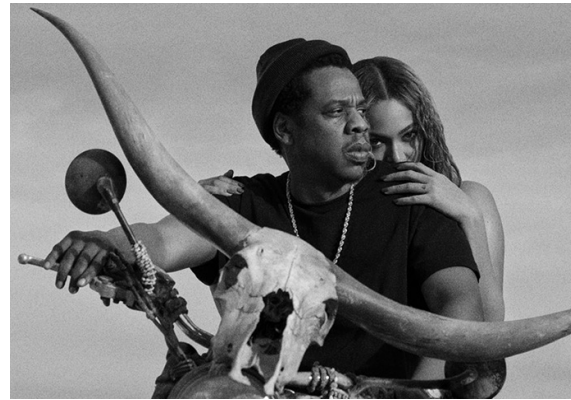 - It's own way of **doing** or **presenting things**.

- How to **extract** it?

- How to use it to **generate** new contents?

Blind we are, if creation of this clone army we could not see.
Yoda

generate

extract   Style

Generate

Sequence with Style

- Example
  - **Extract** the words transition process

This is a state ☺

Section 3 in exercises
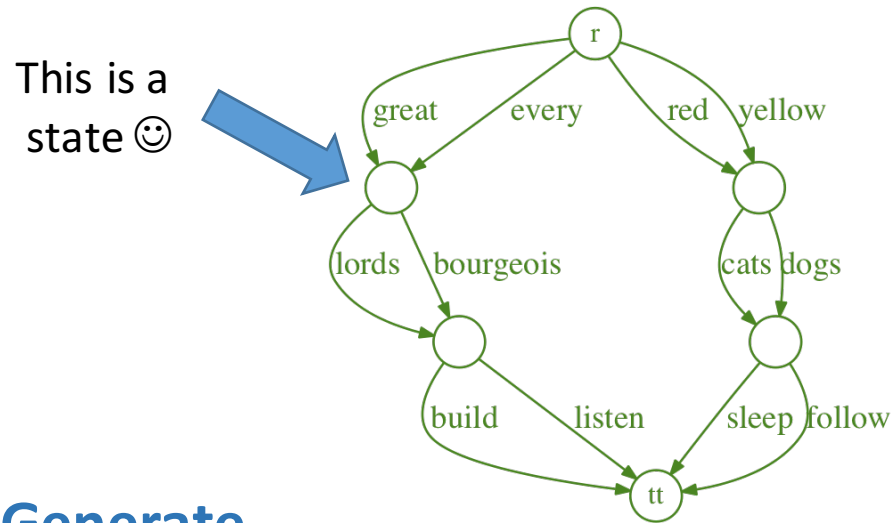
  - **Generate**
    - **Great lords** build

- Toy corpus:
  Great lords listen *but*
  every bourgeois listen while
  great bourgeois build when
  every lords build

  red cats follow but
  yellow dogs sleep while
  red dogs follow when
  yellow cats sleep

Sequence with Style and Rhymes

- **Complexity**?
  - **NP-complete** (reduction from 3-SAT) to prove that it exists a solution
  - **#P-complete** (reduction from counting 3-SAT) for exact sampling

The end?

Thank You

- Questions?

Modeling With Constraints

- Many frameworks exist
  - Sat solvers
  - MIP solvers
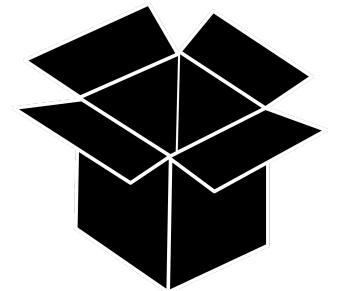 - **Constraint Programming (CP) solvers**

- Modeling
  - **Variables** with domain

  - **Constraints**

Black box Solvers

$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad \mathbf{x_5}$$
$$x_6 \quad x_7 \quad x_8 \quad x_9 \quad \mathbf{x_{10}}$$

Constraint: $\mathbf{x_5}$ and $\mathbf{x_{10}}$ must rhyme.
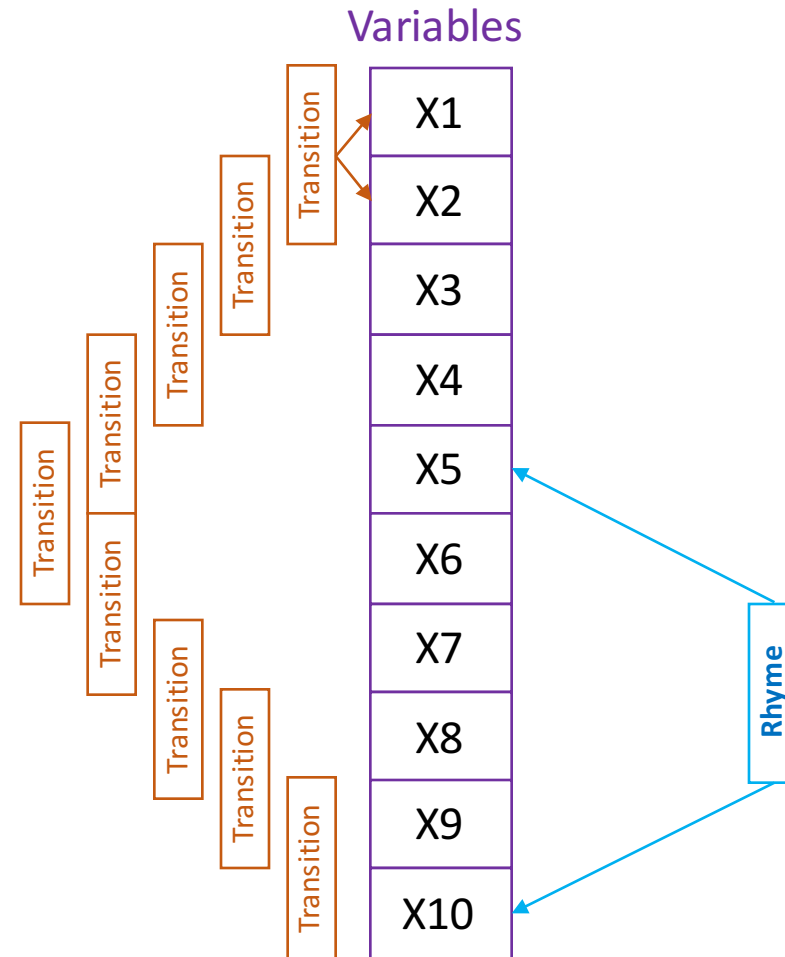
Sequence with Style and Rhymes
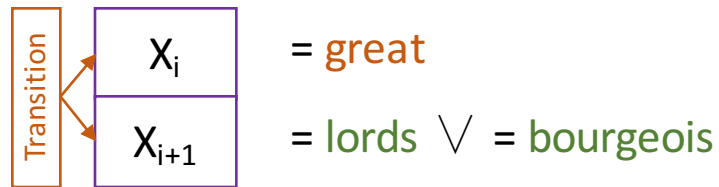
- Our first model
  - 10 Variables
    - 5 by sentence
    - Domain: all possible words from corpus

  - Transition constraint
    - Between consecutive variables

  - Rhyme constraint
    - Between the two last words

Variables

Transition	X1
Transition	X2
Transition	X3
Transition	X4
Transition	X5
Transition	X6
Transition	X7
Transition	X8
Transition	X9
Transition	X10

Rhyme

Transition constraint?

- Standalone
  - Just pick one word in the possible next

Transition	X_i	= great
	X_{i+1}	= lords $\vee$ = bourgeois

- Constraint
  - Boolean formula:

  X_i = great $\wedge$ X_{i+1} = lords $\vee$
  X_i = great $\wedge$ X_{i+1} = bourgeois

  - Table:

X_i	X_{i+1}
great	lords
great	bourgeois
…	…

- Toy corpus:

  Great lords listen but
  every bourgeois listen while
  great bourgeois build when
  every lords build

  red cats follow but
  yellow dogs sleep while
  red dogs follow when
  yellow cats sleep

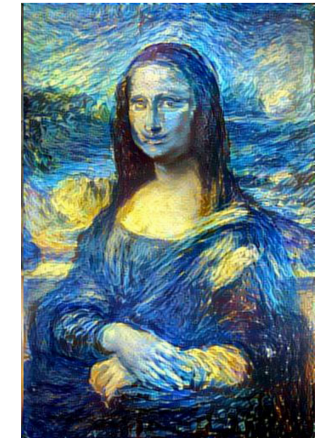**Section 5 in exercises**

21

Generation under constraints

- What if I want to generate a text with rhymes ?

Find a rhyme with combinator**ial**
That doesn't sound artific**ial**

- Text and music using someone's "style" ?

Example with image

- **Music without plagiarism ?**

Let it go, let it go
Can't hold it back anymore

- Poetry with a given number of syllables ?

In the month of the long decline of roses
I, beholding the summer dead before me,
Set my face to the sea and journeyed silent,
Algernon Charles Swinburne

Sequence Without plagiarism

- Problem:
  - **MaxOrder** (Papadopoulos et al., AAAI 2014)
  - Avoiding **plagiarism** in Markov sequence generation

- Example of goal:
  - Based on a **corpus**

  - **Generate** sequences

  - **All** the subsequences of **size 2 belong to the corpus**
    - "*Great lords sleep*" Fail (unknown transition)

  - **None** of the subsequences of **size 3 belongs to the corpus**
    - "*Great lords listen*" Fail (plagiarism)

  - "*Red cats sleep*" **Success**

- Toy corpus:

  Great lords listen but

  every bourgeois listen while

  great bourgeois build when

  every lords build

  red cats follow but

  yellow dogs sleep while

  red dogs follow when

  yellow cats sleep

Sequence with Style and Rhymes

- Our first model
  - 10 Variables
    - 5 by sentence
    - Domain: all possible words from corpus

  - Transition constraint
    - Between consecutive 2 variables

  - Rhyme constraint
    - Between the two last words

  - Plagiarism Constraint
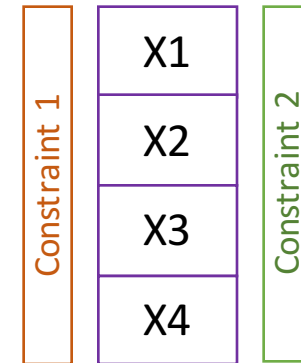   - Between consecutive 3 variables
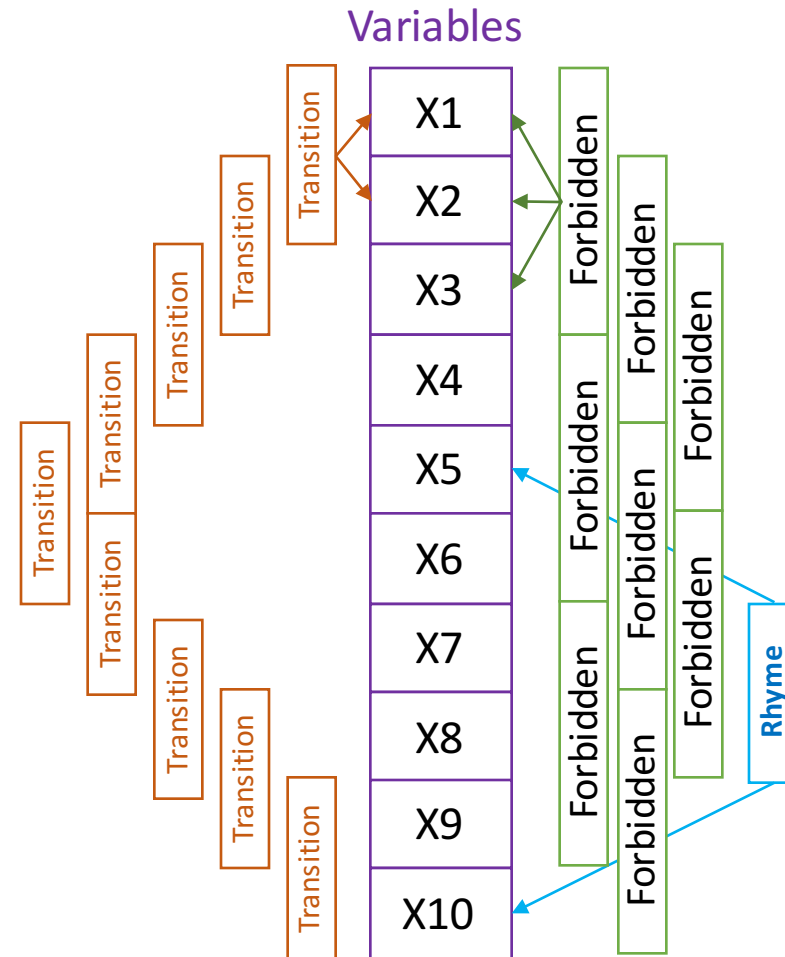
Section 6 in exercises

Inconsistency

- Example

X_1	X_2	X_3	X_4
a	b	a	b
b	a	b	a

Constraint 1

X_1	X_2	X_3	X_4
a	a	a	a
b	b	b	b

Constraint 2

Constraint 1 | X1 | Constraint 2
X2
X3
X4

No solution !

- Solution:
  - Use **modeling tools**

Combinatorial problem solving

- Need for efficient **modeling tools**
  - Hard to express
    - Let's use the solutions

- Efficient data structure:
  1. Efficiently represents discrete problem solutions
     1. **Compression**

  2. Efficiently combinable
     2. **Combination**

  3. Integrated into solvers via fast and incremental algorithms.
     3. **Integration**

Multi-valued Decision Diagram (MDD)

- BDD generalization

$$f : \{0...d-1\}^r \to \{true, false\}$$

- Each **Layer** represents a **variable**
  - Each **arc** is labeled by a **value**

- Each **path** from ***root*** node to ***tt*** node represents a **valid assignment**

- **Exponential** compression power
  - Exponential number or tuples
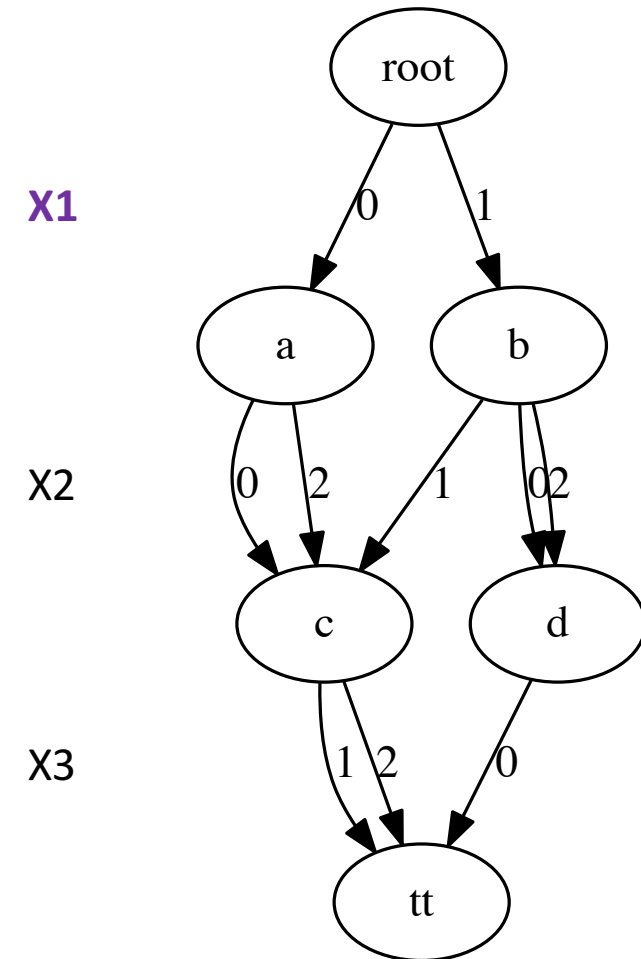
Multi-valued Decision Diagram (MDD)

X1	X2	X3
0	0	1
0	0	2
0	2	1
0	**2**	**2**
1	1	1
1	1	2
1	0	0
1	2	0

24 values   **VS**   10 arcs + 6 nodes

10⁹⁰ tuples   VS   14,000 nodes
**100 variables   600,000 arcs**
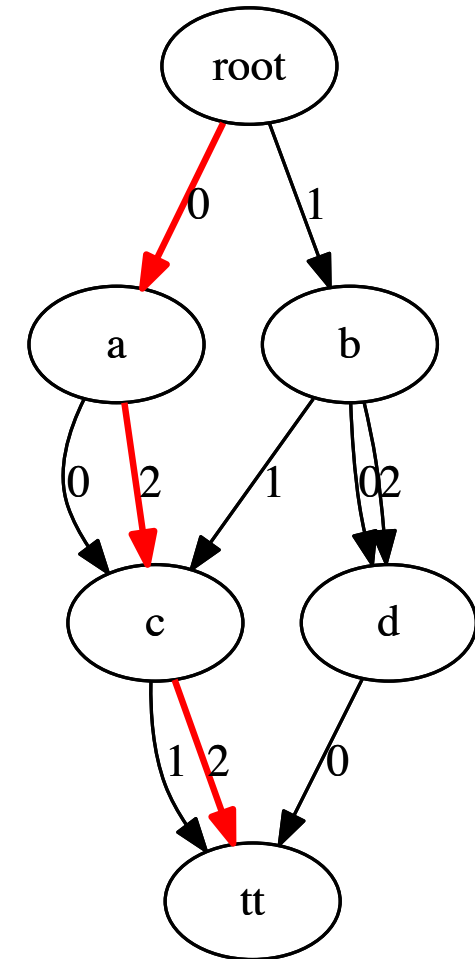
Reduction

- Operation which **merges *equivalent* nodes**

- Two nodes are equivalent if they have

  the same outgoing arcs

- For a given variable ordering
  - *Canonical*
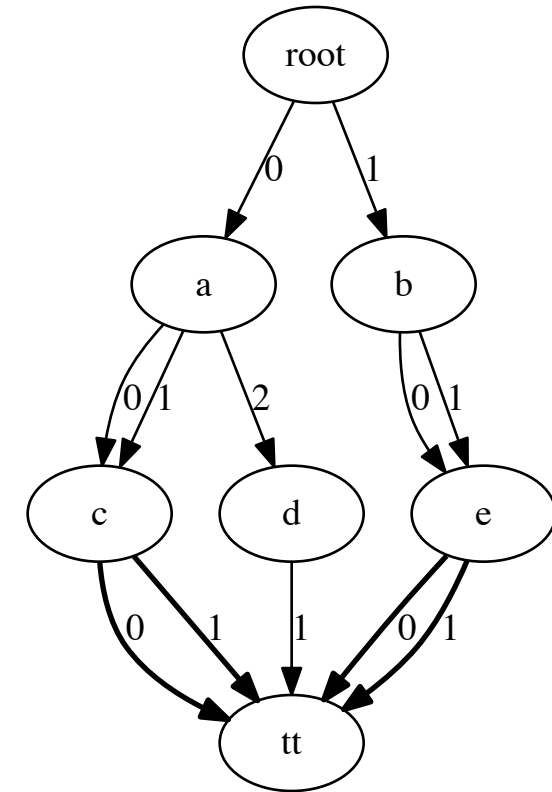
Reduction

- Operation which **merges *equivalent* nodes**

- Two nodes are equivalent if they have
  the same outgoing arcs

- For a given variable ordering
  - *Canonical*

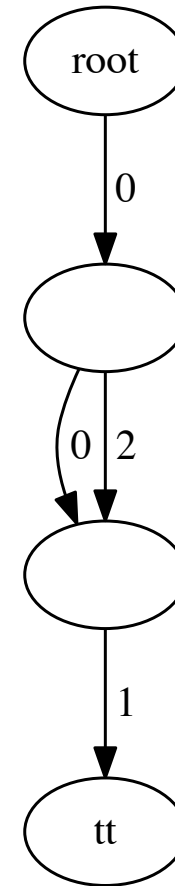- **Compression is made by reduction**
  - Sub-graph sharing
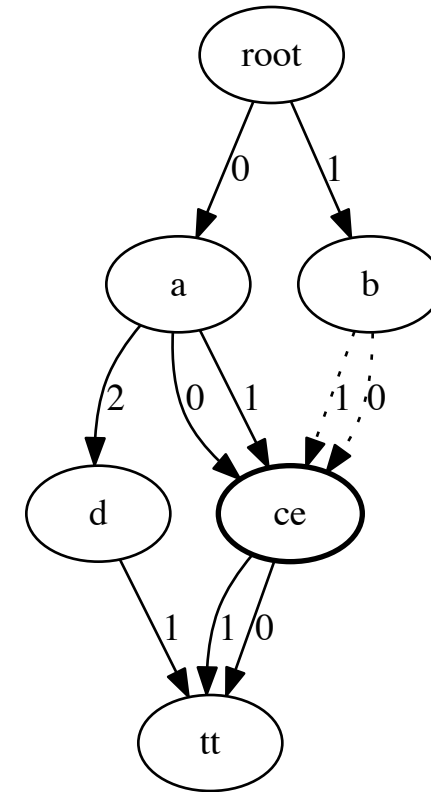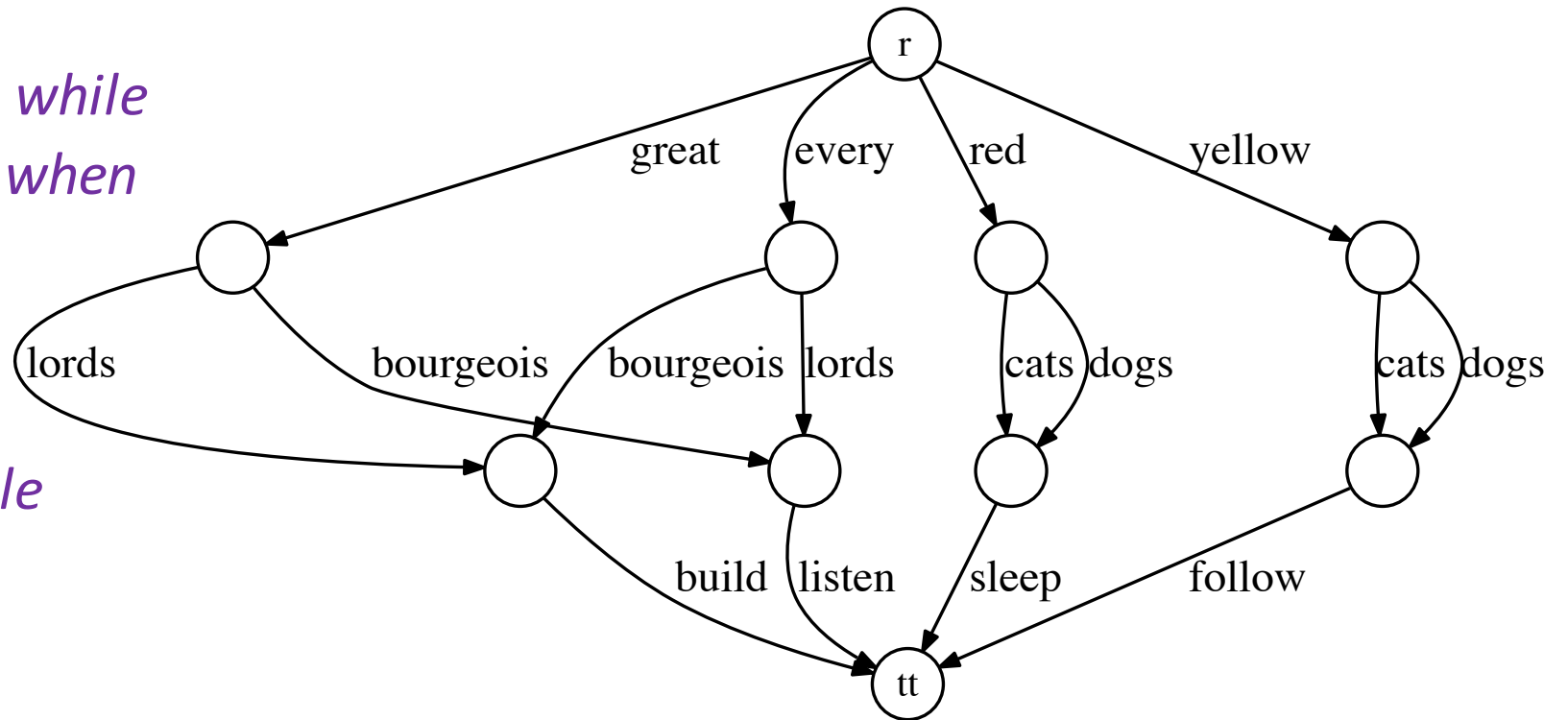
Combination

All set operations

$\bigcap$

$=$
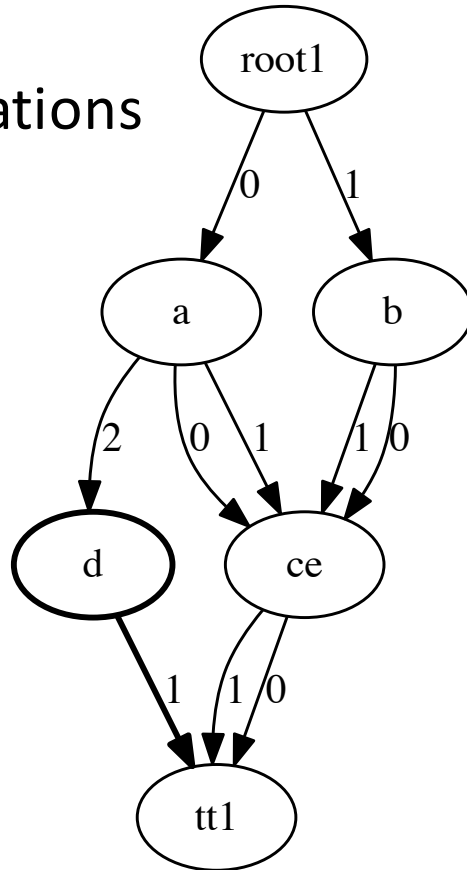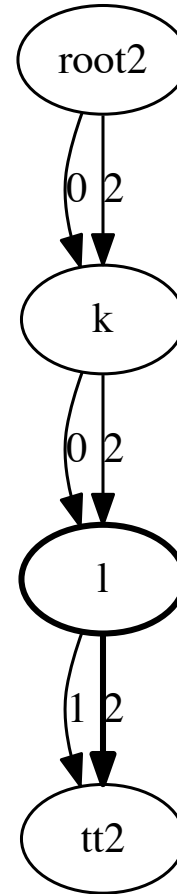
Goal MDD Transition ∩ ¬Plagiarism

Red cats sleep
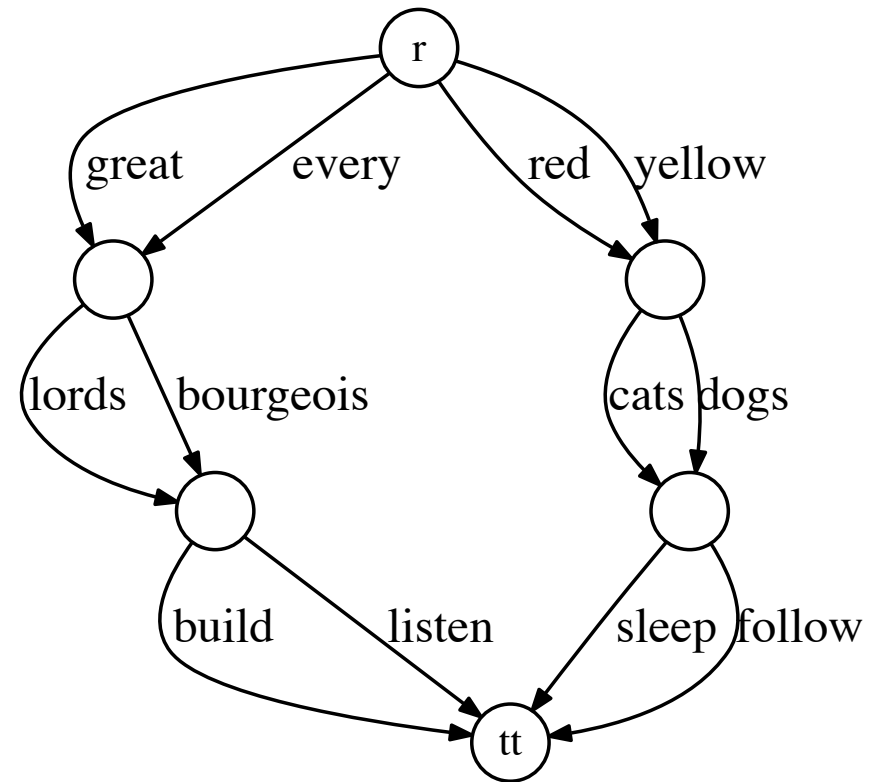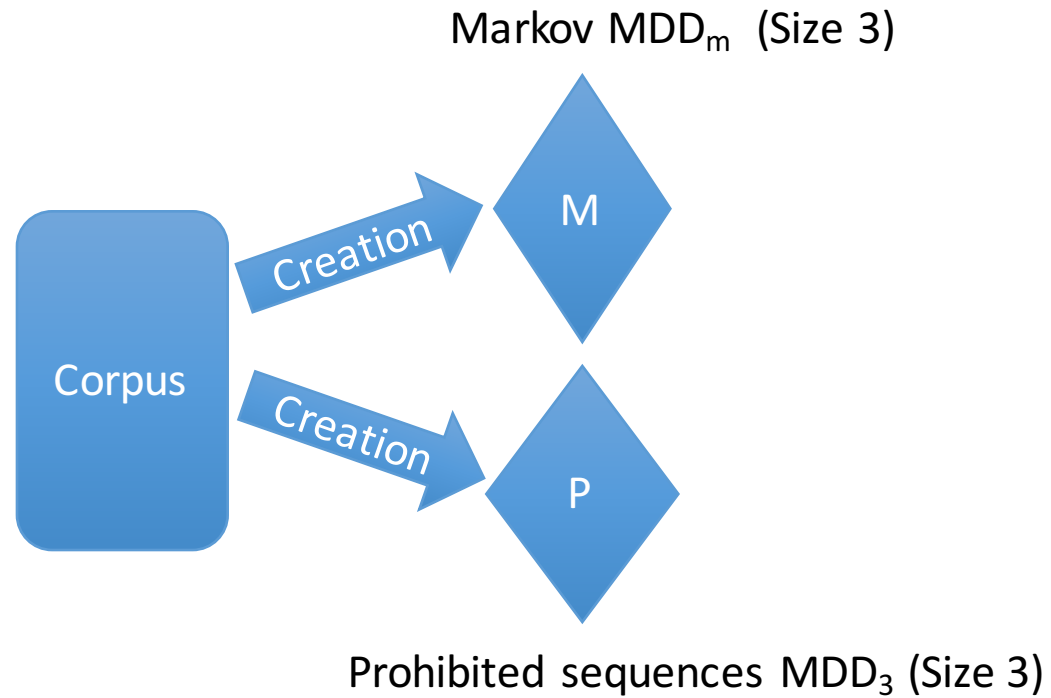
- Toy corpus:

  Great lords listen but
  every bourgeois listen while
  great bourgeois build when
  every lords build

  red cats follow but
  yellow dogs sleep while
  red dogs follow when
  yellow cats sleep

MaxOrder MDD model

Markov MDD_m (Size 3)

Corpus

Creation

M

Markov MDD

- Toy corpus:

  <u>*Great lords* listen</u> *but*

  every bourgeois listen while

  great bourgeois build when

  every lords build

  red cats follow but

  yellow dogs sleep while

  red dogs follow when

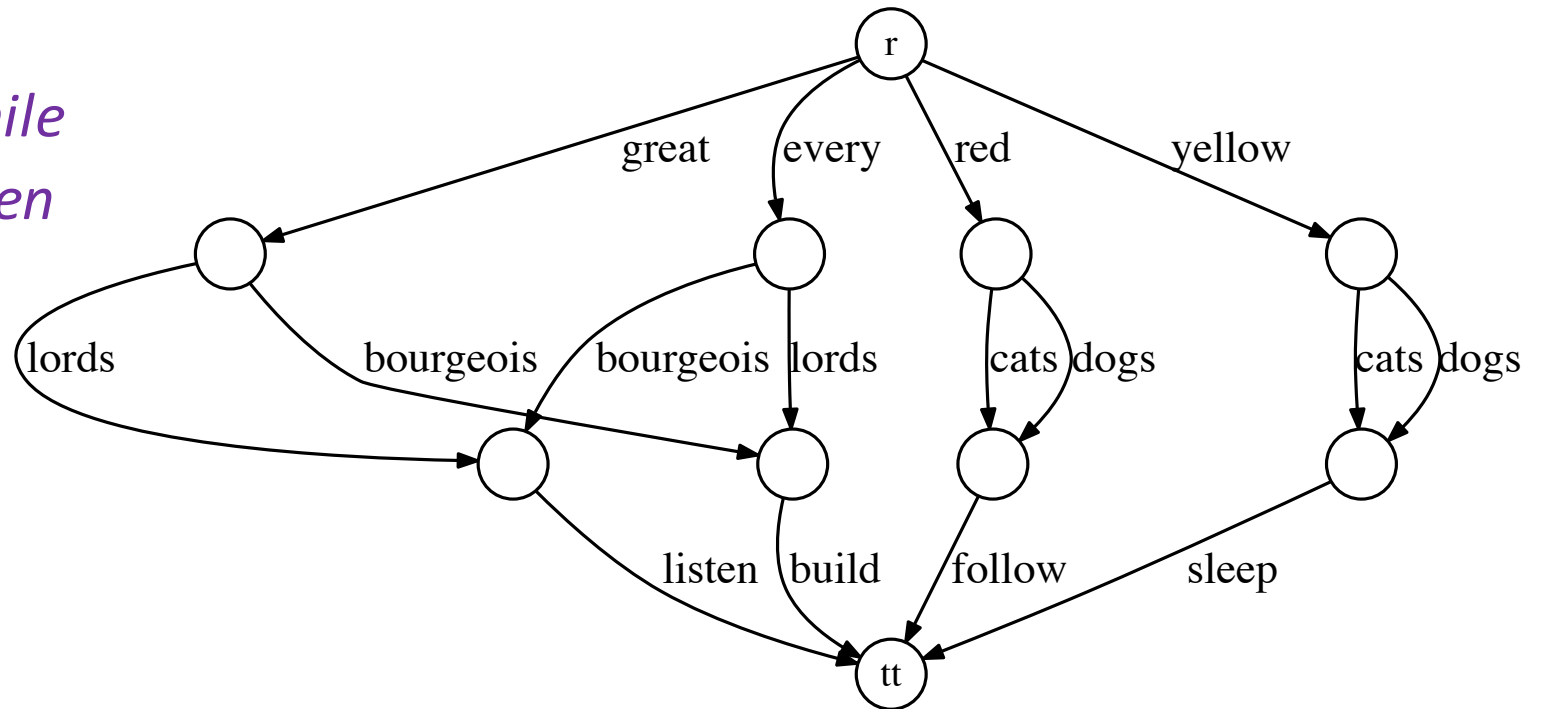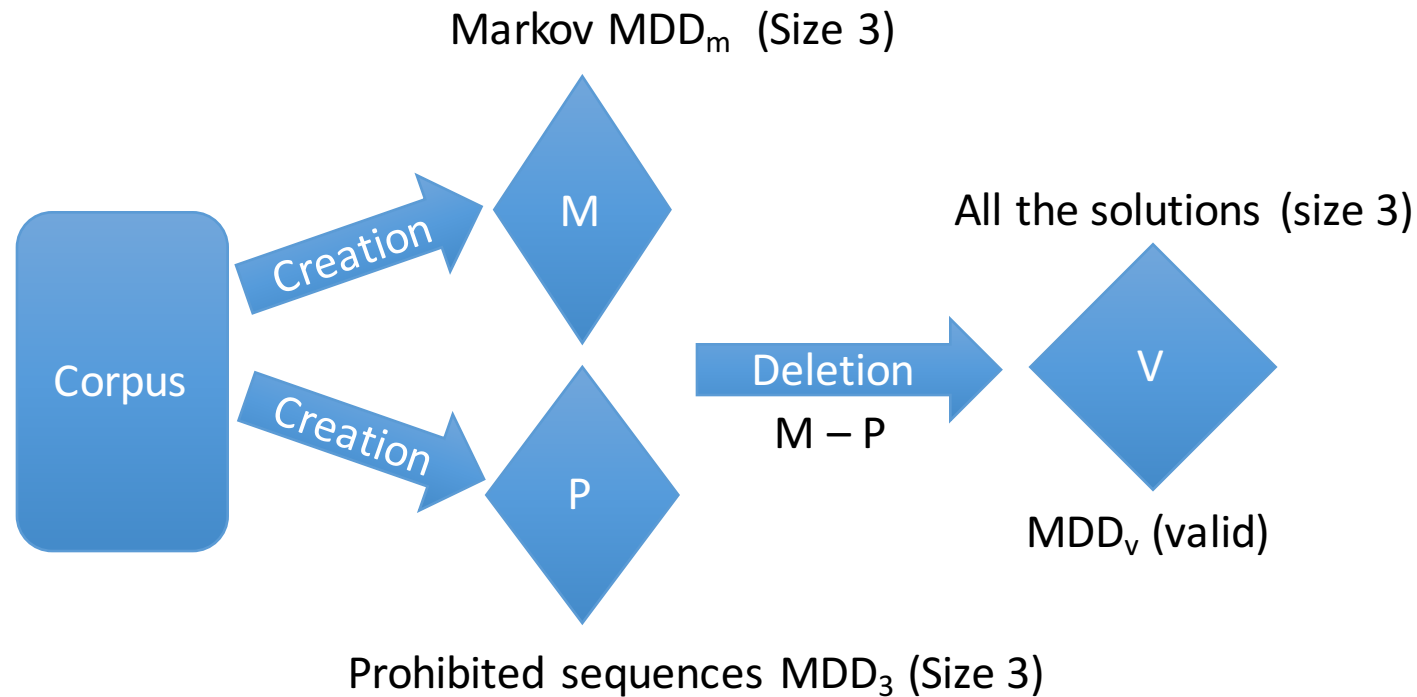 *yellow cats sleep*

Red cats sleep

MaxOrder MDD model

Markov MDD$_m$ (Size 3)

Corpus

Creation

M

Creation

P

Prohibited sequences MDD$_3$ (Size 3)

Prohibited MDD

- Toy corpus:

  Great lords listen but

  every bourgeois listen while

  great bourgeois build when

  every lords build

  red cats follow but

  yellow dogs sleep while
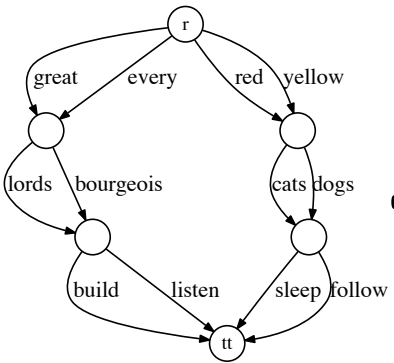
  red dogs follow when

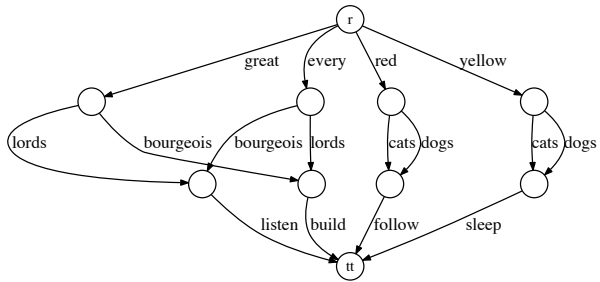 _yellow cat sleep_

MaxOrder MDD model

Markov MDD_m (Size 3)

Corpus

Creation

M

All the solutions (size 3)

Creation

Deletion

$M - P$

V

P

MDD_v (valid)

Prohibited sequences MDD_3 (Size 3)

Remark: $P \subseteq M$
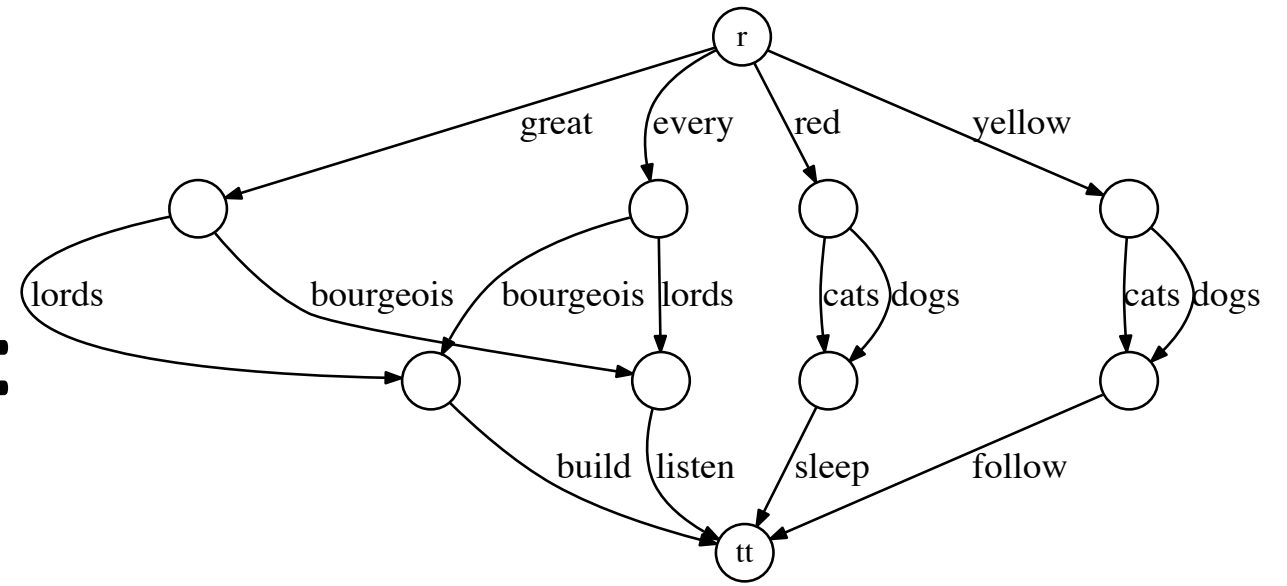
Removing

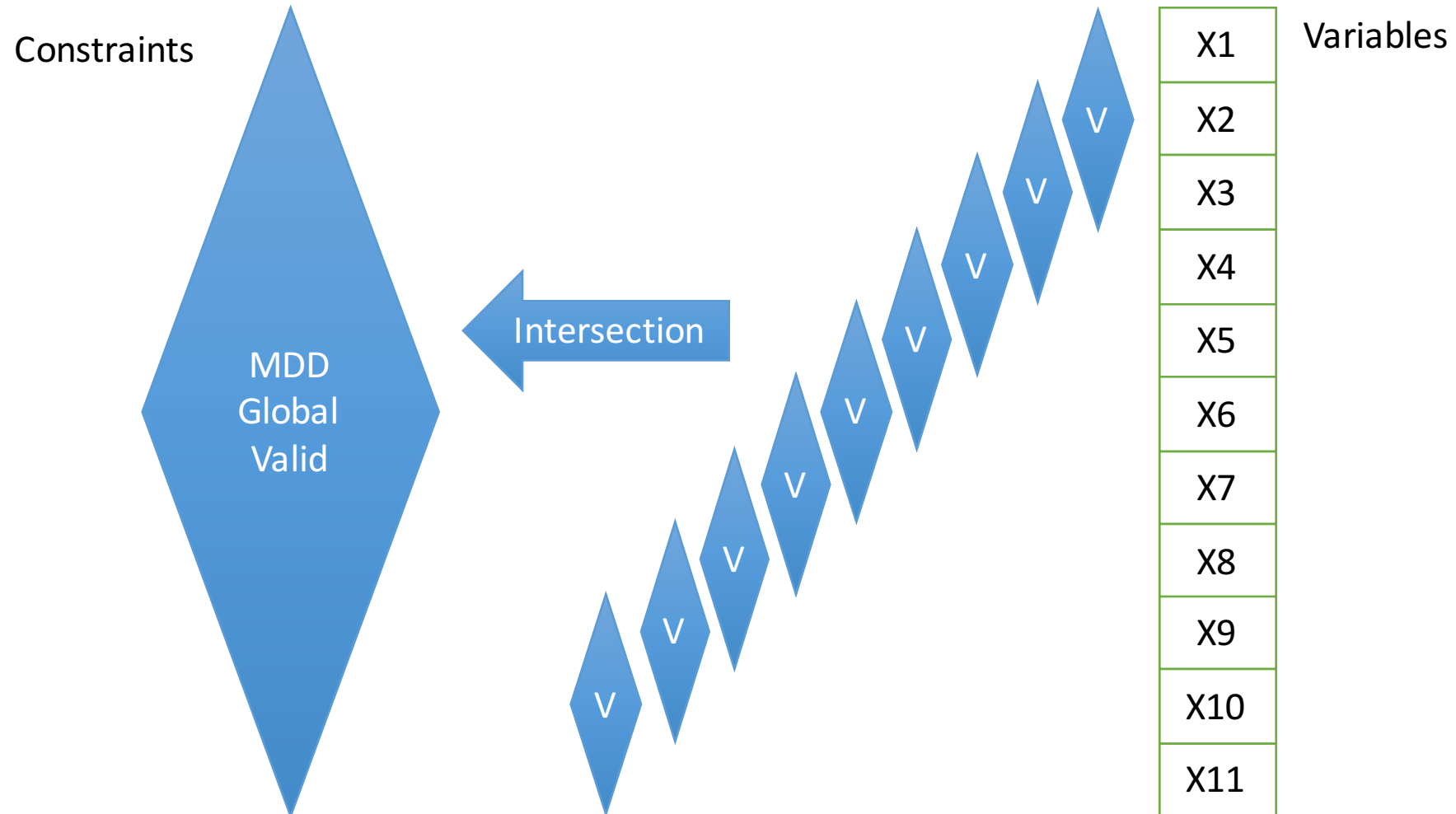**Markov MDD** **Prohibited MDD**

All solutions

MaxOrder MDD model

MaxOrder Results

- MDD contains all the solutions
  - **No longer need** a solver ☺
  - Can be embedded with other **constraints**

- Benchmark "Proust – À la recherche du temps perdu"
  - 11,000 different words (domain)
  - Max Plagiarism: 4
  - Generate sequences of length 20
  - 2 creations & 1 deletion & 16 intersections

  - MDD Size
    - 1 M nodes, 190 M arcs
    - **$2.2*10^{35}$** tuples

  - Results
    - First: ~400 seconds, equivalent to the ad hoc algorithm
    - Ordering effect: **143 s**

Compression ratio: **10^{27}**
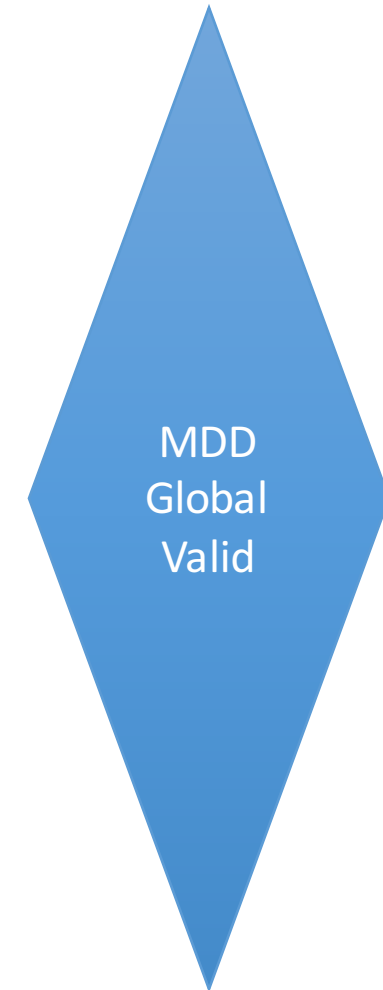
MDD
Global
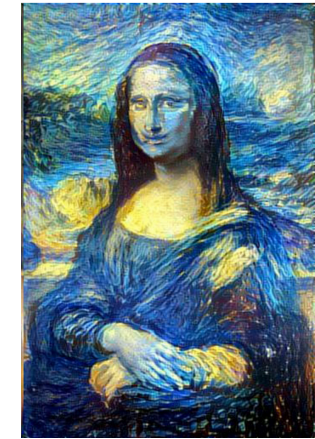Valid

Generation under constraints

- What if I want to generate a text with rhymes ?

Find a rhyme with combinator**ial**
That doesn't sound artific**ial**

- Text and music using someone's "style" ?

Example with image

- Music without plagiarism ?

Let it go, let it go
Can't hold it back anymore

- **Poetry with a given number of syllables ?**

In the month of the long decline of roses
I, beholding the summer dead before me,
Set my face to the sea and journeyed silent,
Algernon Charles Swinburne

Sequence with a given number of syllables

- Syllabic poetic meter
  - 12 syllables
- French poets from 17th to 19th century

- Automatic generation from a corpus
- Example
  - *Grands seigneurs, tout bourgeois veut bâtir comme un Bœuf*
    - Great lords, every bourgeois wants to build like a Beef
  - *Grands – sei – gneurs - tout – bour – geois – veut – bâ- tir – comme – un – Bœuf*
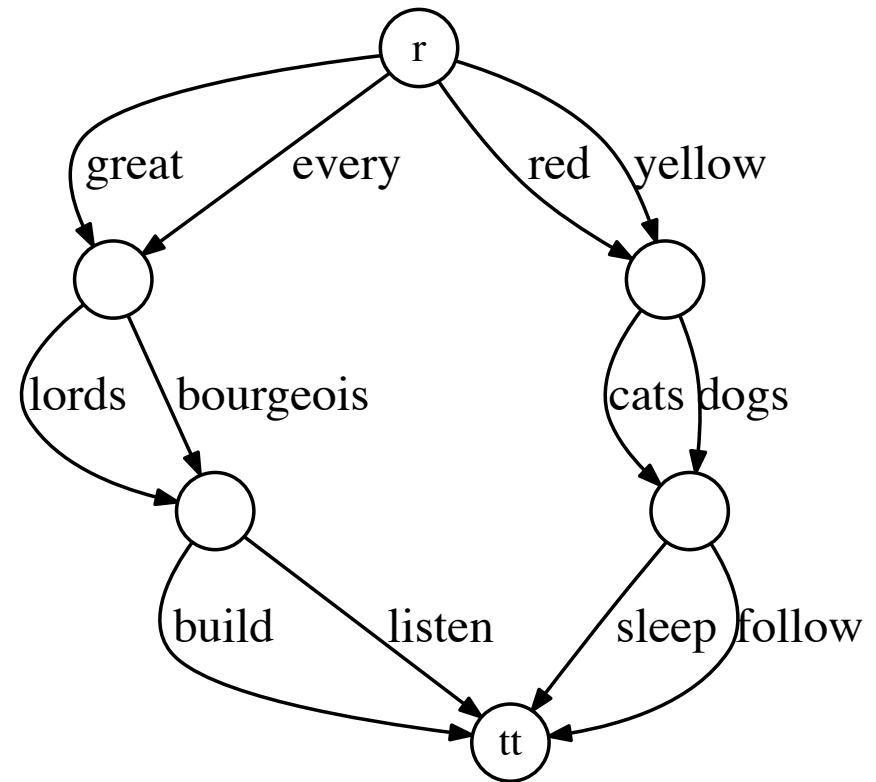 - *   1   2   3   4   5   6   7   8   9   10   11   12*

Sampling   Model
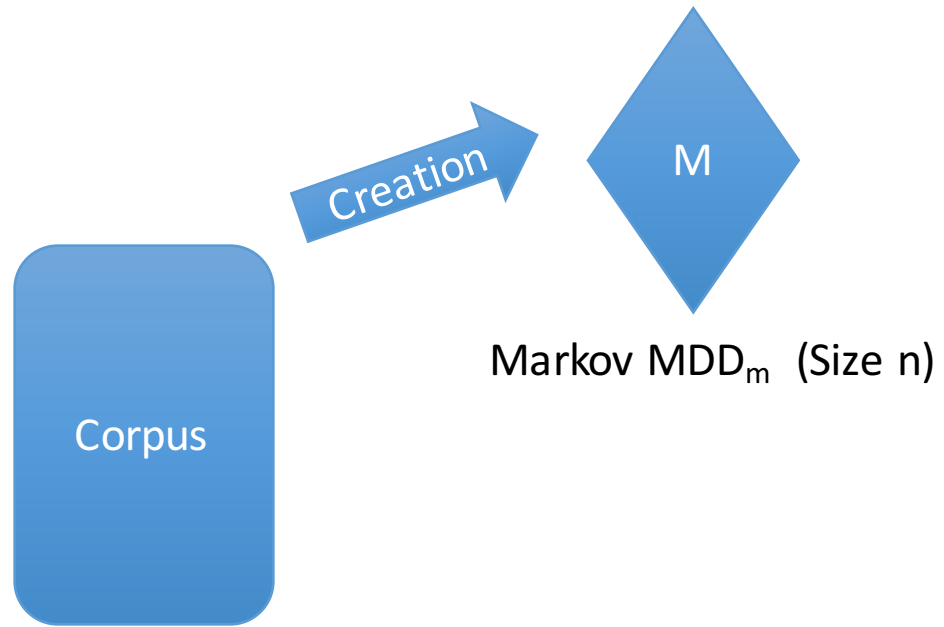
Markov MDD$_m$ (Size n)

Markov MDD

- Toy corpus:

  <u>*Great lords* </u>listen *but*

  every bourgeois listen while

  great bourgeois build when

  every lords build

  red cats follow but

  yellow dogs sleep while
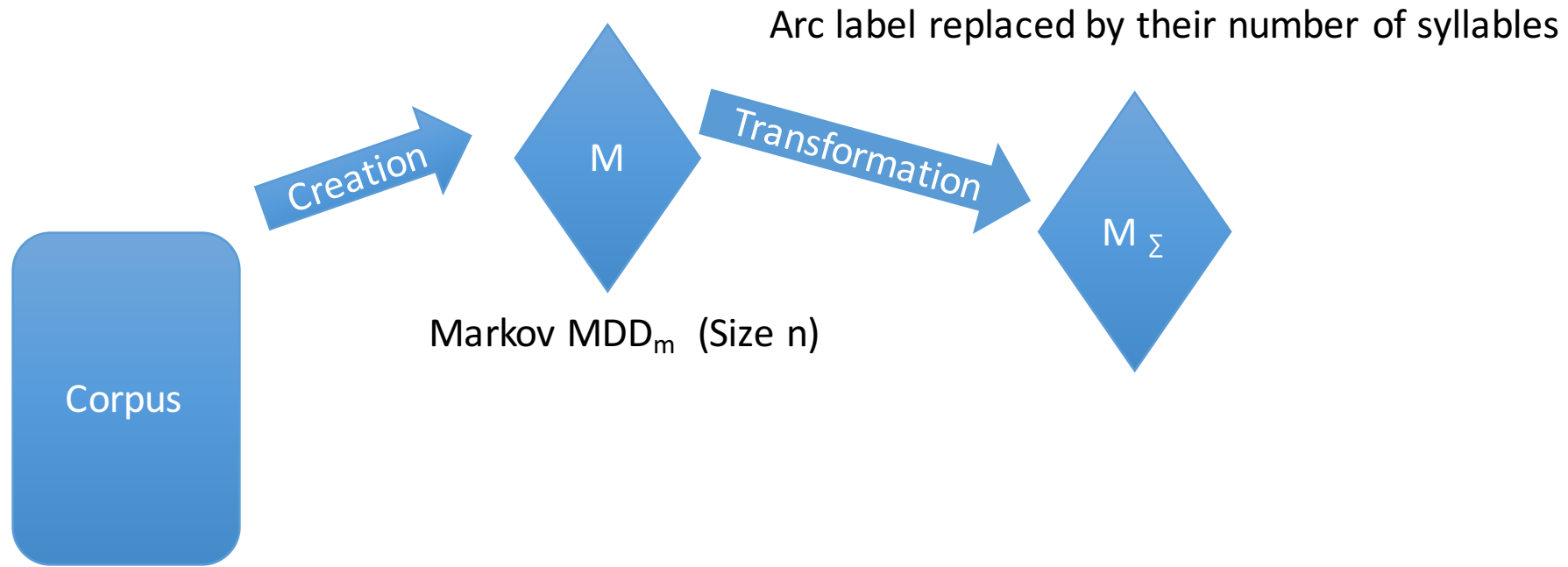
  red dogs follow when

  yellow cat sleep

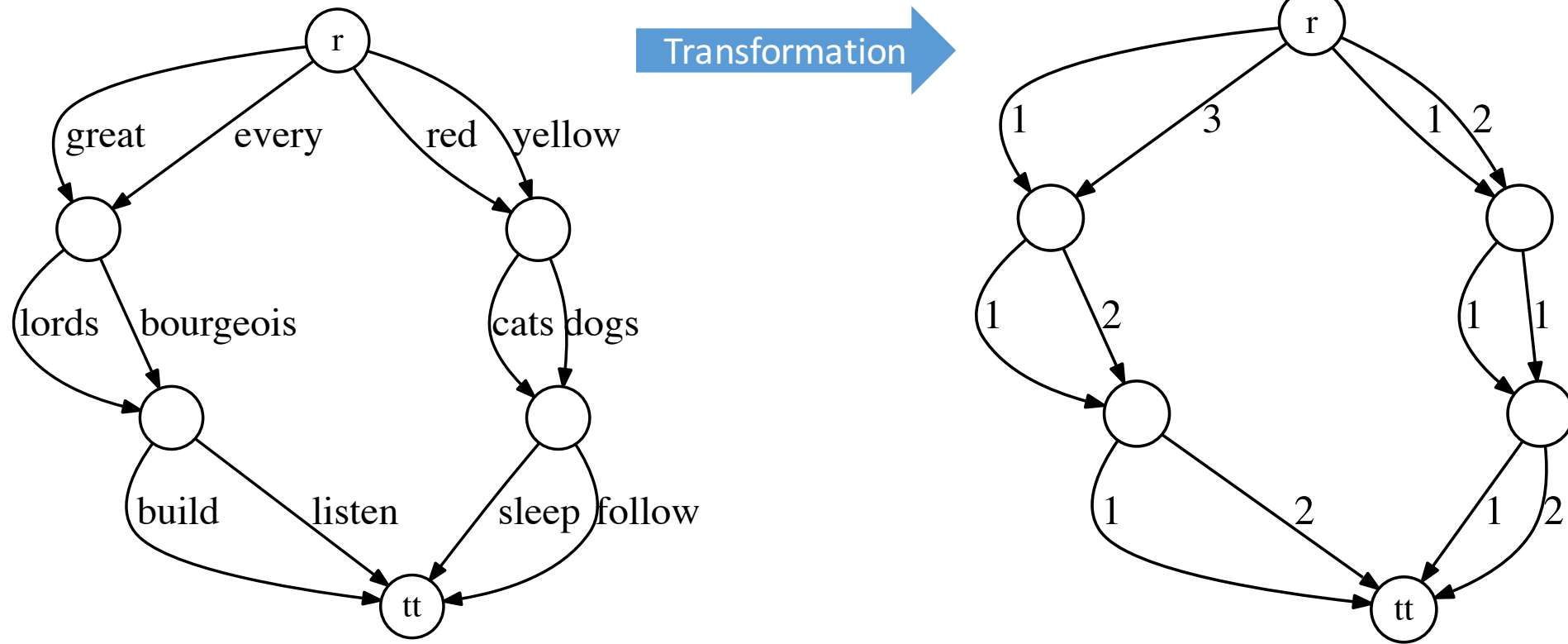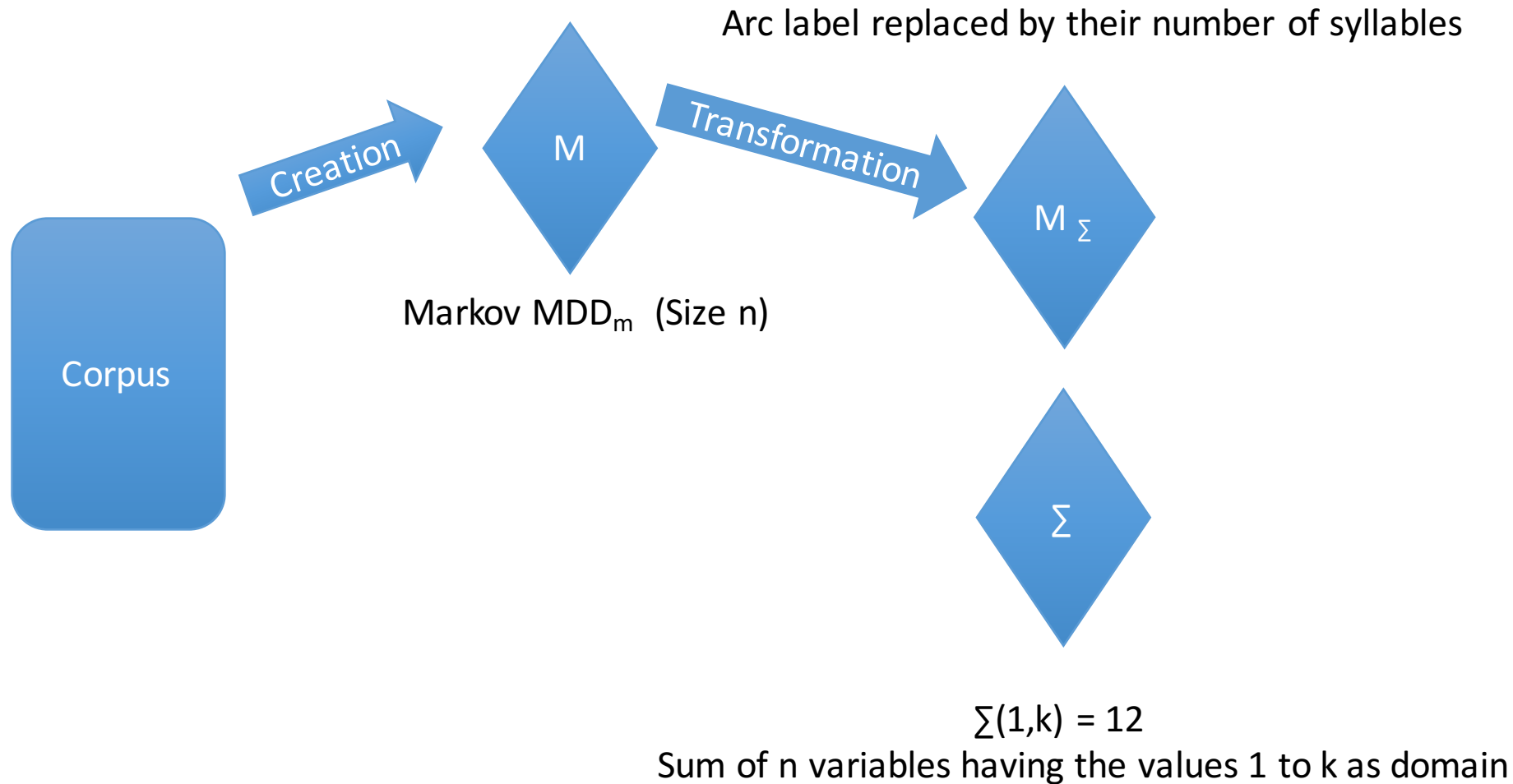Red cats sleep

Sampling   Model

Arc label replaced by their number of syllables

Creation → M → Transformation → M_{Σ}

Markov MDD_m (Size n)

Corpus

Value Transformation

Sampling   Model

Arc label replaced by their number of syllables

Creation

M

Transformation

M_{Σ}

Markov MDD_m (Size n)

Σ

Corpus

$\Sigma(1,k) = 12$
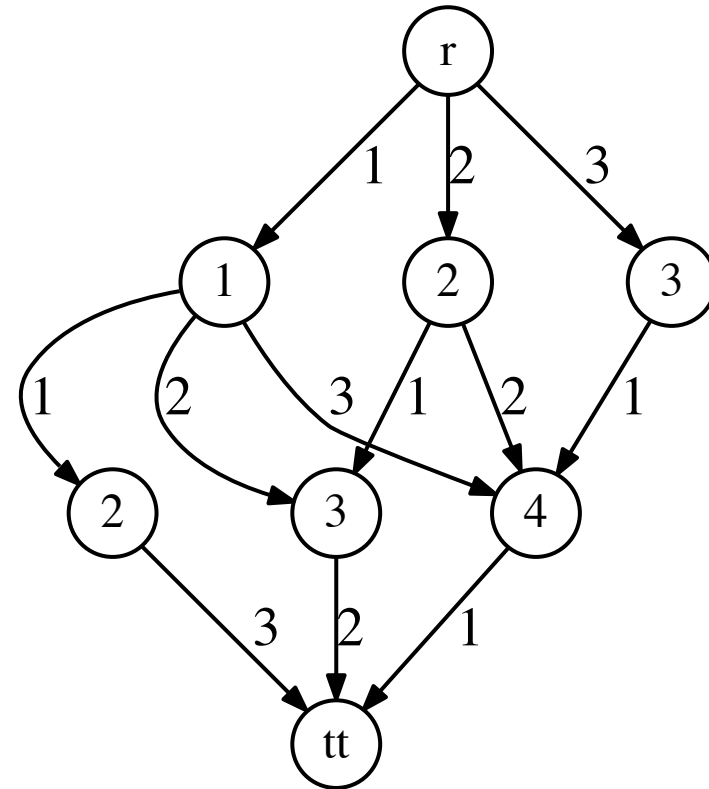Sum of n variables having the values 1 to k as domain

Sum MDD

- ∑(1,3) = 5
  - All the paths are composed of values whose sum is 5.

Sampling   Model

Arc label replaced by their number of syllables

Creation

M

Markov MDD$_m$ (Size n)

Transformation

M $_\Sigma$

Intersection

A

All the solutions

Σ

$\sum(1,k) = 12$
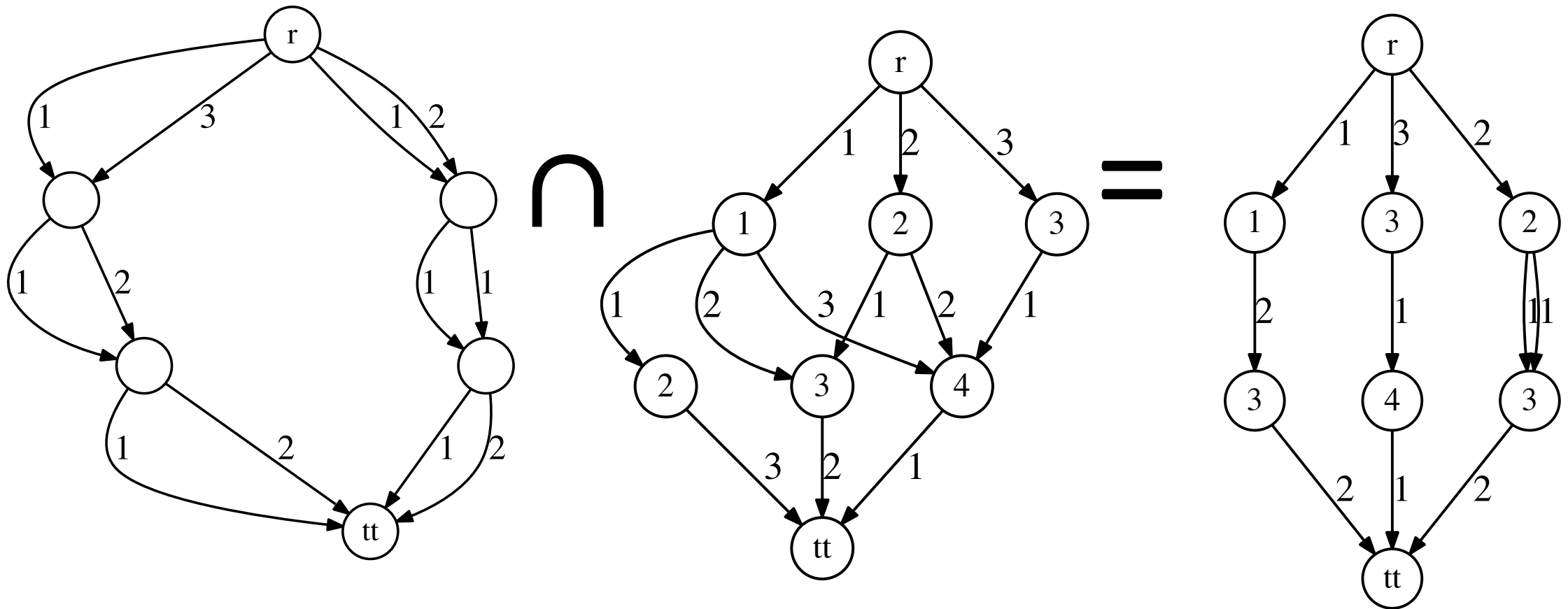Sum of n variables having the values 1 to k as domain
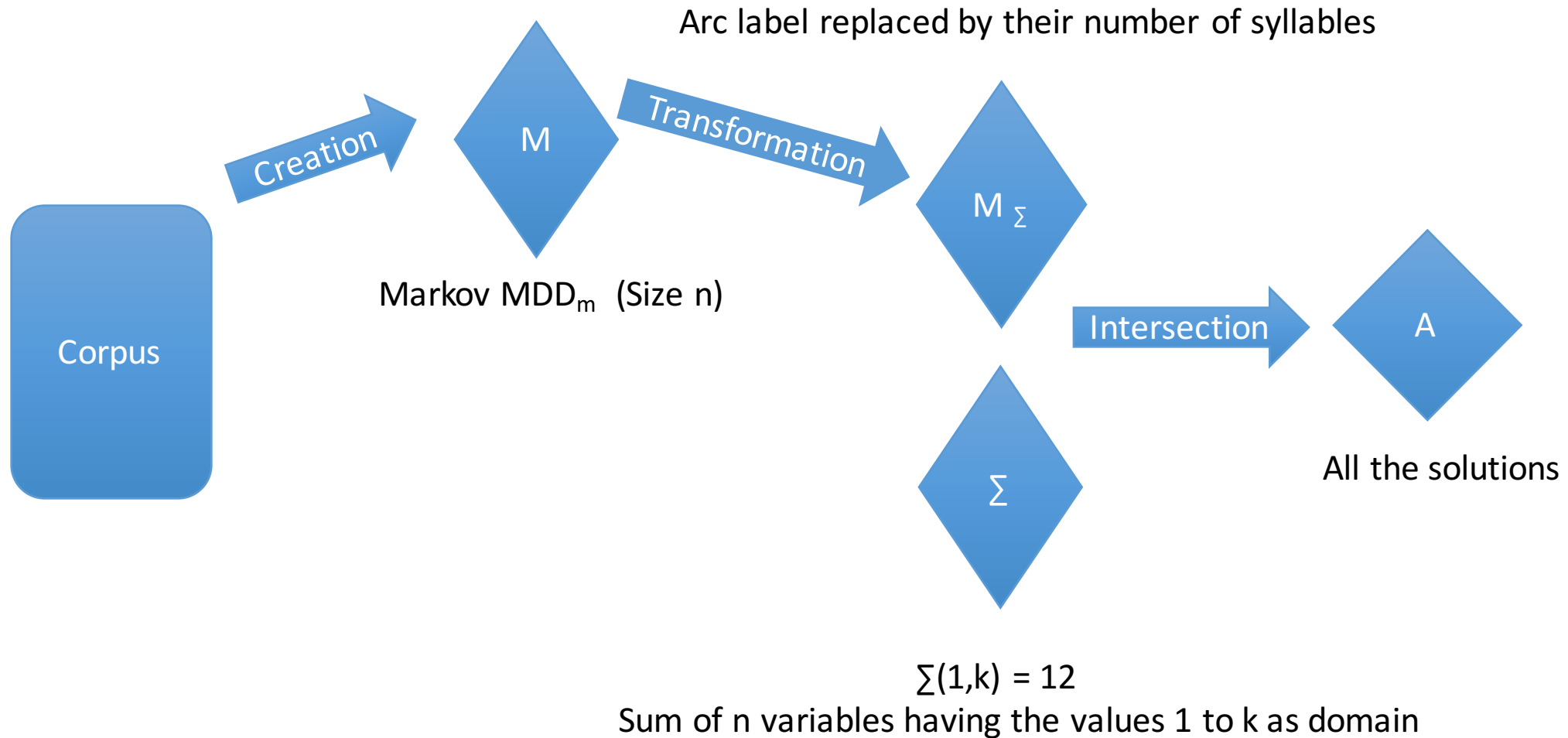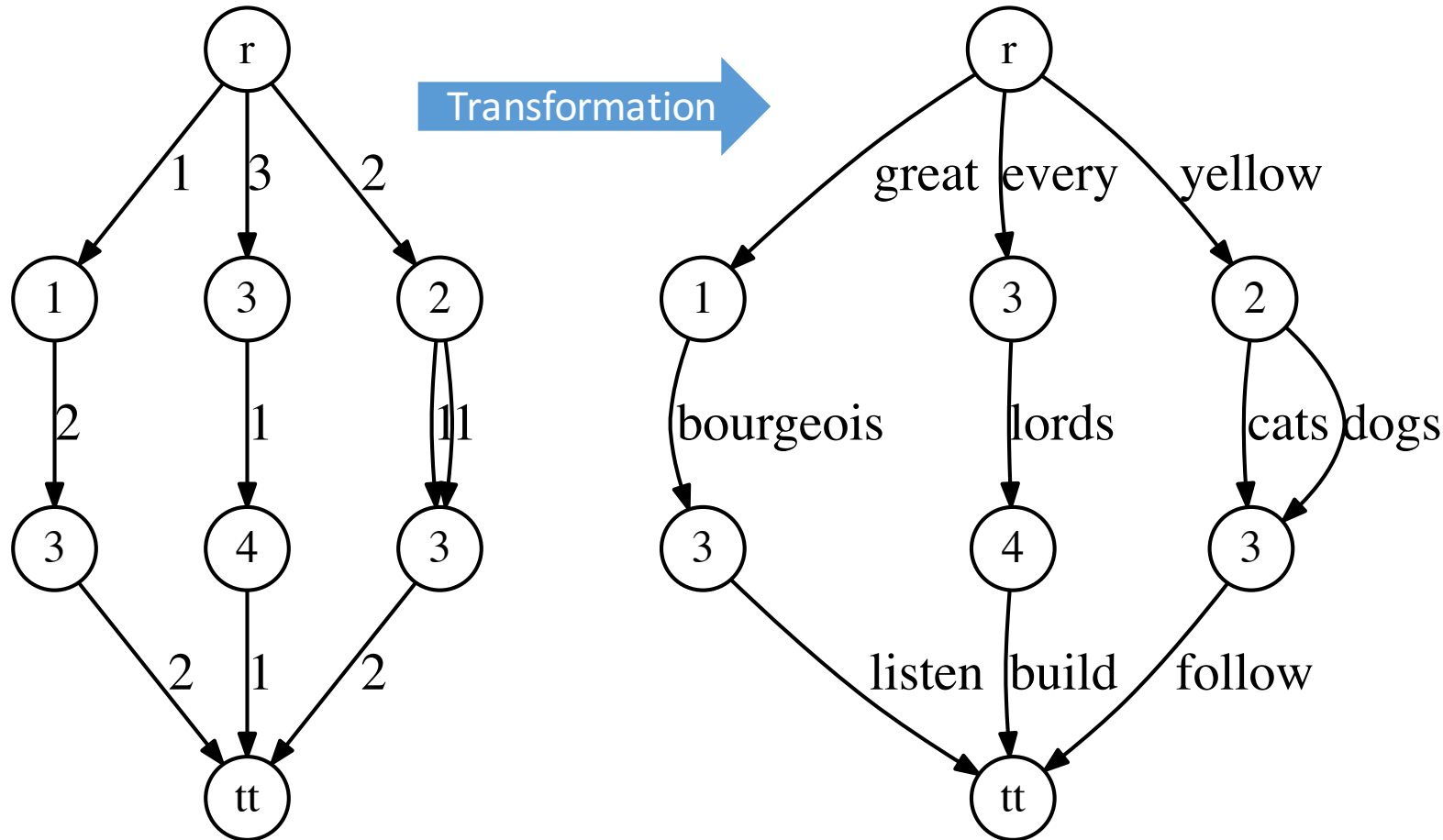
Intersection

Conversion

Transformation

- Solutions:
  - Sequences of size 2 from the corpus
  - Sum of syllables equal 5
    - (Usually 12)

- Now Sample !
  - Big MDDs
    - Millions of nodes
  - Huge amount of solution
    - 10^{90} in our problems

Conclusion

- Learn to **model and solve** some problems

- Using **standalone** methods

- Using basics of **modeling** and solvers

- Using some advance **modeling tools**

- Provided good model for some **hard problems**

Thank You

- Questions?
  - For real ☺

Combinatorial Generation Problems

Guillaume Perez