The Modular Structure of an Ontology: an Empirical Study

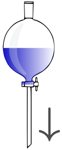
Chiara Del Vescovo Bijan Parsia Uli Sattler *Thomas Schneider*

School of Computer Science, University of Manchester, UK

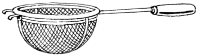
DL, 7 May 2010 WoMo, 11 May 2010

Three questions

What is my ontology about?



How many modules does my ontology have?



↓

How do we identify relevant modules?

And now . . .

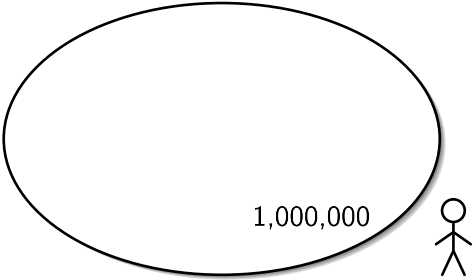
What is my ontology about?

How many modules does my ontology have?

How do we identify relevant modules?

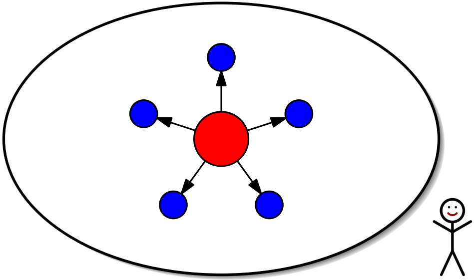
What is my ontology about?

We can't inspect all its axioms.



What is my ontology about?

We can inspect its modular structure, obtained a posteriori.



We bet Robert Stevens

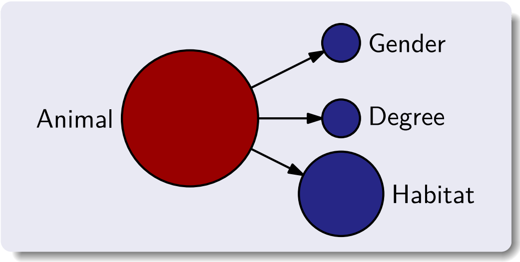
- Ontology about periodic table of the chemical elements
- Logical structure $\approx$ intended modelling?
  - What is its modular structure?
  - What are its main parts?

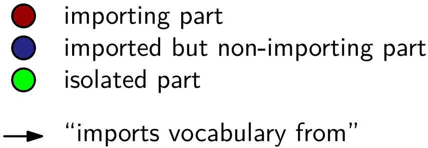
We bet Robert Stevens

- Ontology about periodic table of the chemical elements
- Logical structure ≈ intended modelling?
  - What is its modular structure?
  - What are its main parts?
- Challenge: *automatic* partition into meaningful modules

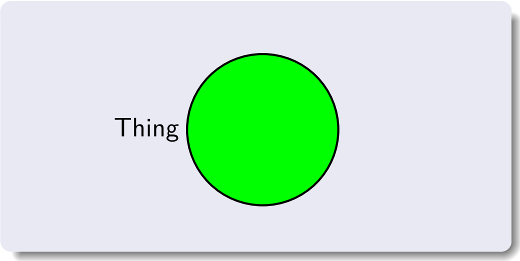
Modular structure with existing tools

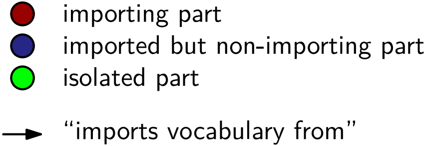
Partition of **Koala** via E-connections in Swoop





Partition for ontology **Periodic**





Locality-based modules (LBMs)

Module extraction service

Input: ontology $\mathcal{O}$; set Σ of terms from $\mathcal{O}$

Output: subset $\mathcal{M} = \text{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

Guarantee: for all axioms α with $\text{sig}(\alpha) \subseteq \Sigma$:

$$\mathcal{O} \models \alpha \quad \text{iff} \quad \mathcal{M} \models \alpha \qquad \textbf{(Coverage)}$$

Locality-based modules (LBMs)

Module extraction service

Input: ontology $\mathcal{O}$; set Σ of terms from $\mathcal{O}$

Output: subset $\mathcal{M} = \mathrm{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

Guarantee: for all axioms α with $\mathrm{sig}(\alpha) \subseteq \Sigma$:

$$\mathcal{O} \models \alpha \text{ iff } \mathcal{M} \models \alpha \qquad \textbf{(Coverage)}$$

Modules providing coverage

- encapsulate knowledge about the *topic* Σ
- are important for modular import/reuse:
  "Give me all that $\mathcal{O}$ knows about the topic Σ"
- are hard to extract if minimality is required

Locality-based modules (LBMs)

Module extraction service

Input: ontology $\mathcal{O}$; set Σ of terms from $\mathcal{O}$

Output: subset $\mathcal{M} = \mathsf{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

Guarantee: for all axioms α with $\mathsf{sig}(\alpha) \subseteq \Sigma$:

$$\mathcal{O} \models \alpha \quad \text{iff} \quad \mathcal{M} \models \alpha \qquad \textbf{(Coverage)}$$

LBMs

- provide coverage and therefore encapsulation
- are not always minimal, but often of reasonable size
- can be efficiently computed
- have important robustness properties

Locality-based modules (LBMs)

Module extraction service

Input: ontology $\mathcal{O}$; set Σ of terms from $\mathcal{O}$

Output: subset $\mathcal{M} = \text{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

Guarantee: for all axioms α with $\text{sig}(\alpha) \subseteq \Sigma$:

$$\mathcal{O} \models \alpha \quad \text{iff} \quad \mathcal{M} \models \alpha \qquad \textbf{(Coverage)}$$

General remarks:

- Often $\text{sig}(\mathcal{M}) \neq \Sigma$
- Different seed signatures can lead to the same module

$$\begin{array}{ccc} & \text{sig}(\mathcal{M}) & \\ / & & \backslash \\ \Sigma_1 & \cdots & \Sigma_n \end{array}$$

Modular structure via LBMs

Want to extract *all (relevant)* LBMs in order to:

- obtain a finer-grained analysis
- guide users in choosing the right topic(s)
- draw conclusions on characteristics of an ontology:

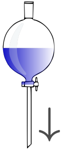
Modular structure via LBMs

Want to extract *all (relevant)* LBMs in order to:

- obtain a finer-grained analysis
- guide users in choosing the right topic(s)
- draw conclusions on characteristics of an ontology:
  - To which extent does $\mathcal{O}$ cover its topics?
  - How strongly are certain terms connected in $\mathcal{O}$?
  - What is the axiomatic richness of $\mathcal{O}$?
  - Does $\mathcal{O}$ have superfluous parts?
  - Agreement between logical and intended intuitive modelling?

And now . . .

What is my ontology about?



How many modules does my ontology have?

How do we identify relevant modules?

Obvious lower and upper bounds

- Ontologies of size n can have between 1 and 2^n modules.
- (We're working on tighter bounds.)

Obvious lower and upper bounds

- Ontologies of size n can have between 1 and 2^n modules.
- (We're working on tighter bounds.)
- Do real-life ontologies fall into the worst case?

An algorithm that extracts all modules

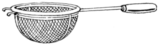
Results when applied to two small ontologies:

Ontology	#Ax	#Terms	**#mods**	Theor. Max.	Time
Koala	42	25	**3660**	33 554 432	9s
Mereology	44	25	**1952**	33 554 432	3min

An algorithm that extracts all modules

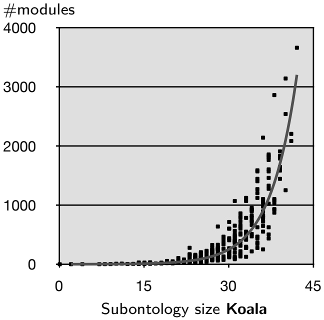
Results when applied to two small ontologies:

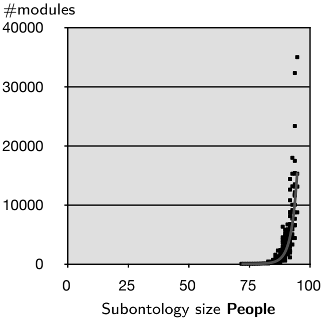
Ontology	#Ax	#Terms	**#mods**	Theor. Max.	Time
Koala	42	25	**3660**	33 554 432	9s
Mereology	44	25	**1952**	33 554 432	3min

1. Are 3660 and 1952 “exponential” numbers? 
2. How to filter for interesting modules? 

Modularisation of subontologies

- Modularised randomly generated parts of 8 ontologies
- Example growth of module numbers:



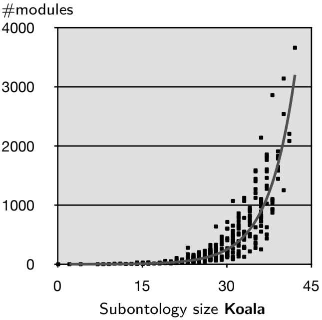


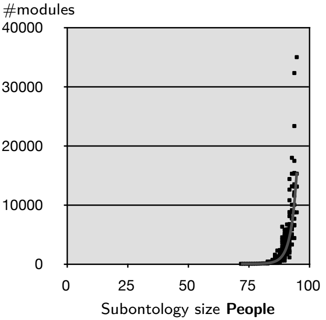
Trendline equation: $y = O(1.2^x)$ $y = O(1.5^x)$

Confidence: 0.90 0.96

Modularisation of subontologies

- Modularised randomly generated parts of 8 ontologies
- Example growth of module numbers:





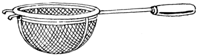
Trendline equation: $y = O(1.2^x)$ $y = O(1.5^x)$ **Exponential!**

Confidence: 0.90 0.96

And now . . .

What is my ontology about?

How many modules does my ontology have?



↓

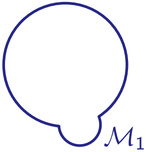
How do we identify relevant modules?

Unification of similar modules

- Identify sets $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathfrak{M}$ with $\bigcup \mathfrak{M}$ and $\bigcap \mathfrak{M}$

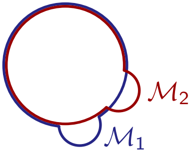
Unification of similar modules

- Identify sets $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathfrak{M}$ with $\bigcup \mathfrak{M}$ and $\bigcap \mathfrak{M}$



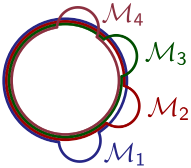
Unification of similar modules

- Identify sets $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathfrak{M}$ with $\bigcup \mathfrak{M}$ and $\bigcap \mathfrak{M}$



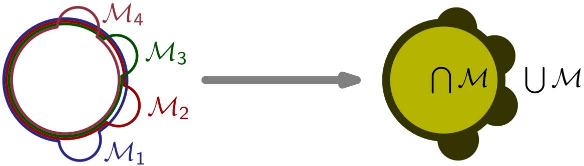
Unification of similar modules

- Identify sets $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathfrak{M}$ with $\bigcup \mathfrak{M}$ and $\bigcap \mathfrak{M}$



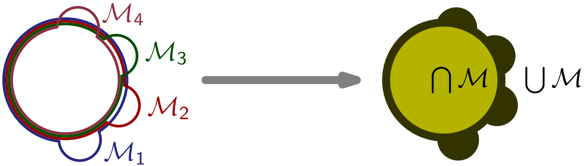
Unification of similar modules

- Identify sets $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathfrak{M}$ with $\bigcup \mathfrak{M}$ and $\bigcap \mathfrak{M}$



Unification of similar modules

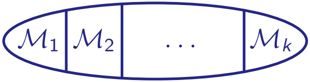
- Identify sets $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathfrak{M}$ with $\bigcup \mathfrak{M}$ and $\bigcap \mathfrak{M}$



- Outcome for **Koala**: no significant reduction in module numbers

Genuine and fake modules

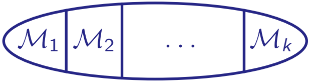
- Identify modules that are an agglomeration of other modules and contain no new information
- $\mathcal{M}$ is *fake* if there is partition $\mathcal{M} = \mathcal{M}_1 \uplus \cdots \uplus \mathcal{M}_k$ with pairwise disjoint $\mathrm{sig}(\mathcal{M}_i)$.



- All other modules are *genuine*.

Genuine and fake modules

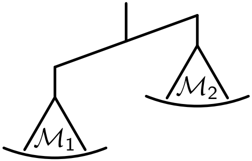
- Identify modules that are an agglomeration of other modules and contain no new information
- $\mathcal{M}$ is *fake* if there is partition $\mathcal{M} = \mathcal{M}_1 \uplus \cdots \uplus \mathcal{M}_k$ with pairwise disjoint $\mathrm{sig}(\mathcal{M}_i)$.



- All other modules are *genuine*.
- Outcome for **Koala**:
  66% of the 3660 modules are genuine.

Weight analysis

. . . by scalesperson Chiara ;-)



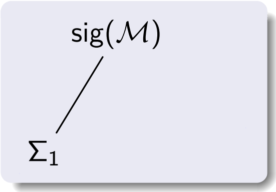
Weight analysis

- Number of terms in the module m

$\mathrm{sig}(\mathcal{M})$

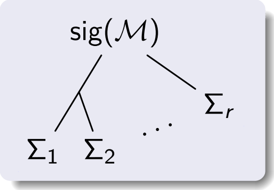
Weight analysis

- Number of terms in the module m
- Minimal size of seed signatures s



Weight analysis

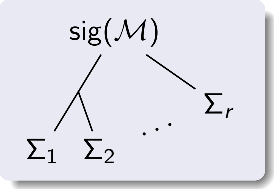
- Number of terms in the module m
- Minimal size of seed signatures s
- Number of different minimal seed signatures r



Weight analysis

- Number of terms in the module m
- Minimal size of seed signatures s
- Number of different minimal seed signatures r

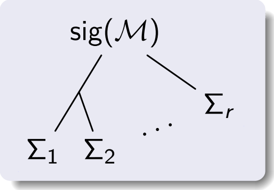
$\rightsquigarrow$ PullingPower($\mathcal{M}$)



Weight analysis

- Number of terms in the module m
- Minimal size of seed signatures s
- Number of different minimal seed signatures r

$\rightsquigarrow$ PullingPower($\mathcal{M}$) $\frac{m}{s}$

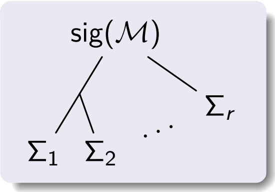


Weight analysis

- Number of terms in the module m
- Minimal size of seed signatures s
- Number of different minimal seed signatures r

$\leadsto$ PullingPower($\mathcal{M}$) $\frac{m}{s}$

$\leadsto$ Cohesion($\mathcal{M}$)

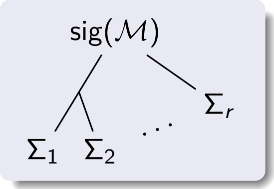


Weight analysis

- Number of terms in the module m
- Minimal size of seed signatures s
- Number of different minimal seed signatures r

$\rightsquigarrow$ PullingPower($\mathcal{M}$) $\frac{m}{s}$

$\rightsquigarrow$ Cohesion($\mathcal{M}$) $\frac{r}{s}$



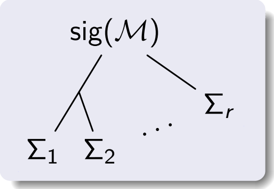
Weight analysis

- Number of terms in the module m
- Minimal size of seed signatures s
- Number of different minimal seed signatures r

$\rightsquigarrow$ PullingPower($\mathcal{M}$) $\frac{m}{s}$

$\rightsquigarrow$ Cohesion($\mathcal{M}$) $\frac{r}{s}$

Weight($\mathcal{M}$) $w = \frac{r \cdot m}{s^2}$



Example: **Koala**

$\text{Animal} \sqsubseteq\ \geqslant 1\ \text{hasHabitat}$
$\text{Animal} \sqsubseteq\ =1\ \text{hasGender}$
$\text{DryEucalyptForest} \sqsubseteq \text{Forest}$
$\text{Female} \equiv \exists\ \text{hasGender}.\{\text{female}\}$
$\text{Forest} \sqsubseteq \text{Habitat}$
$\text{GraduateStudent} \sqsubseteq \text{Student}$
$\text{GraduateStudent} \sqsubseteq \exists\ \text{hasDegree}.(\{\text{BA}\} \sqcup \{\text{BS}\})$
$\text{Koala} \sqsubseteq \exists\ \text{hasHabitat}.\text{DryEucalyptForest}$
$\text{Koala} \sqsubseteq \text{Marsupials}$
$\text{Koala} \sqsubseteq \exists\ \text{isHardworking}.\{\text{false}\}$
$\text{KoalaWithPhD} \equiv \text{Koala} \sqcap \exists\ \text{hasDegree}.\{\text{PhD}\}$
$\text{Male} \equiv \exists\ \text{hasGender}.\{\text{male}\}$
$\text{Marsupials} \sqsubseteq \text{Animal}$
$\text{Marsupials} \sqsubseteq \neg\ \text{Person}$
$\text{Parent} \equiv \text{Animal} \sqcap \geqslant 1\ \text{hasChildren}$
$\text{Parent} \sqsubseteq \text{Animal}$
$\text{Person} \sqsubseteq \text{Animal}$
$\text{Person} \sqsubseteq \neg\ \text{Marsupials}$
$\text{Quokka} \sqsubseteq \exists\ \text{isHardworking}.\{\text{true}\}$
$\text{Quokka} \sqsubseteq \text{Marsupials}$
$\text{Rainforest} \sqsubseteq \text{Forest}$
$\text{TasmanianDevil} \sqsubseteq \text{Marsupials}$
$\text{University} \sqsubseteq \text{Habitat}$
$\exists\ \text{hasChildren} \sqsubseteq \text{Animal}$
$\top \sqsubseteq \forall\ \text{hasChildren}.\text{Animal}$
$\exists\ \text{hasDegree} \sqsubseteq \text{Person}$

$\top \sqsubseteq \forall\ \text{hasDegree}.\text{Degree}$
$\top \sqsubseteq\ \leqslant 1\ \text{hasGender}$
$\exists\ \text{hasGender} \sqsubseteq \text{Animal}$
$\top \sqsubseteq \forall\ \text{hasGender}.\text{Gender}$
$\exists\ \text{hasHabitat} \sqsubseteq \text{Animal}$
$\top \sqsubseteq \forall\ \text{hasHabitat}.\text{Habitat}$
$\top \sqsubseteq\ \leqslant 1.\text{isHardWorking}$
$\text{BA} : \text{Degree}$
$\text{BS} : \text{Degree}$
$\text{MA} : \text{Degree}$
$\text{PhD} : \text{Degree}$
$\text{female} : \text{Gender}$
$\text{male} : \text{Gender}$

$\text{MaleStudentWith3Daughters} \equiv \text{Student} \sqcap \forall\ \text{hasChildren}.\text{Female} \sqcap \exists\ \text{hasGender}.\{\text{male}\} \sqcap =1\ \text{hasChildren}$
$\text{Student} \equiv \text{Person} \sqcap \exists\ \text{hasHabitat}.\text{University} \sqcap \exists\ \text{isHardworking}.\{\text{true}\}$

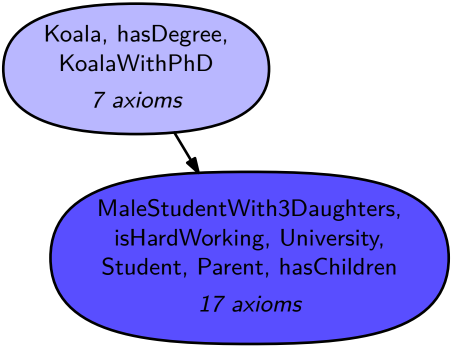
Example: **Koala**

MaleStudentWith3Daughters,
isHardWorking, University,
Student, Parent, hasChildren

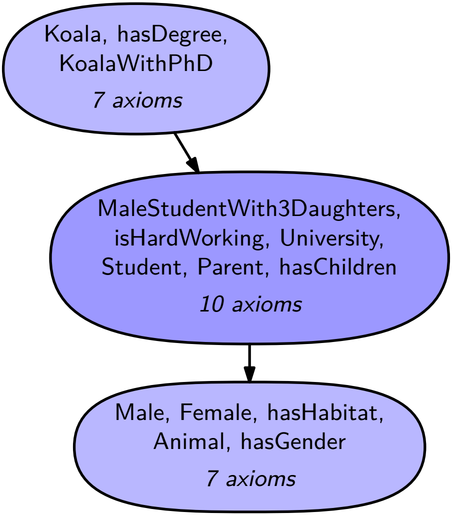
17 axioms

{Student, Parent}
{Student, hasChildren}
{MaleStudentWith3Daughters}
{hasChildren, University, isHardWorking}
{Parent, University, isHardWorking}

Example: **Koala**



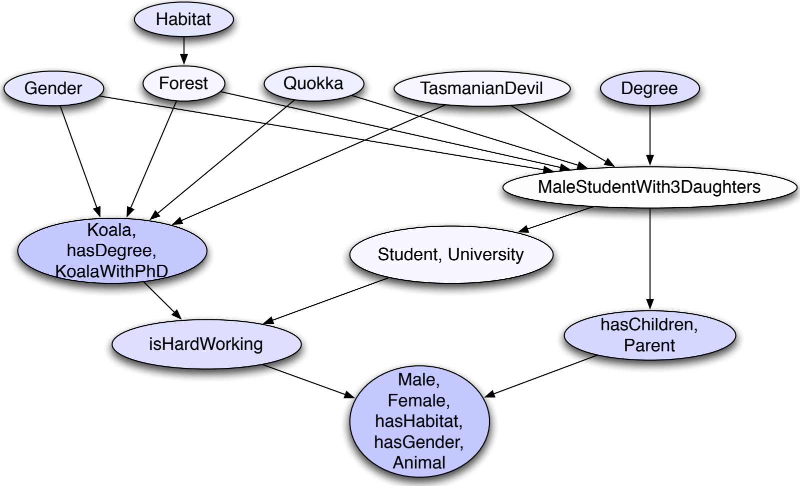
Example: **Koala**



Example: **Koala**

After the first 12 heaviest modules . . .

Example: **Koala**



Work in progress

Is it necessary for the weight analysis to compute all modules?

Work in progress

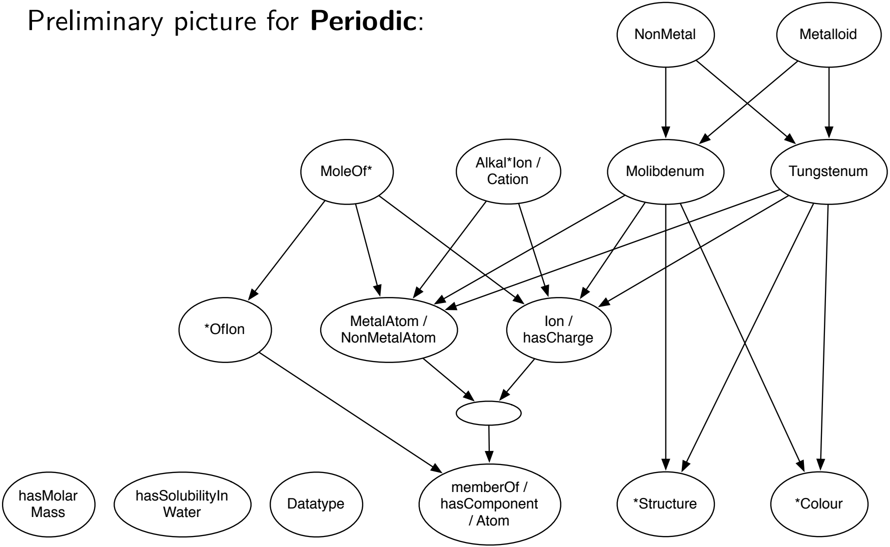
Is it necessary for the weight analysis to compute all modules?

Remember: Weight($\mathcal{M}$) $w = \frac{r \cdot m}{s^2}$

$\rightsquigarrow$ search within modules with very small seed signatures

Work in progress

Preliminary picture for **Periodic**:



Outlook

- Find heaviest modules without computing all modules
- How many modules can ontologies have?
- Relation module number $\leftrightarrow$ justificatory structure

Outlook

- Find heaviest modules without computing all modules
- How many modules can ontologies have?
- Relation module number $\leftrightarrow$ justificatory structure

Thank you.