The Modular Structure of an Ontology: Atomic Decomposition

Chiara Del Vescovo[1] Bijan Parsia[1]
Uli Sattler[1] *Thomas Schneider*[2]

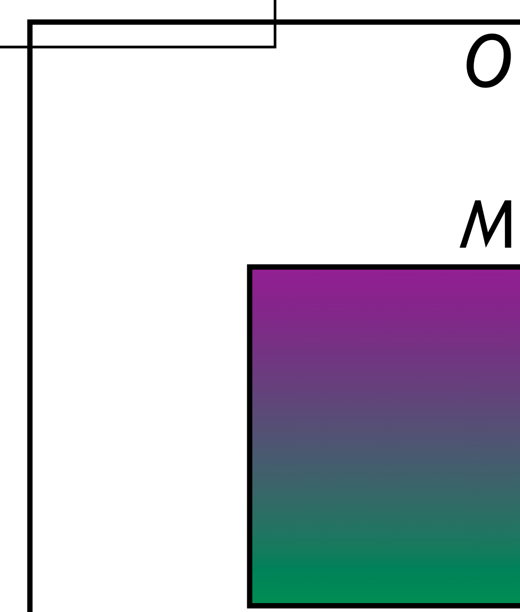
[1]The University of Manchester, UK
[2]Universität Bremen, Germany

Ontologies & Modules

- An **ontology** is a finite set of axioms in a (description) logic

- A **module** $M(\Sigma,O) \subseteq O$ encapsulates knowledge w.r.t. a signature Σ: $M \equiv^{c}_{\Sigma} O$

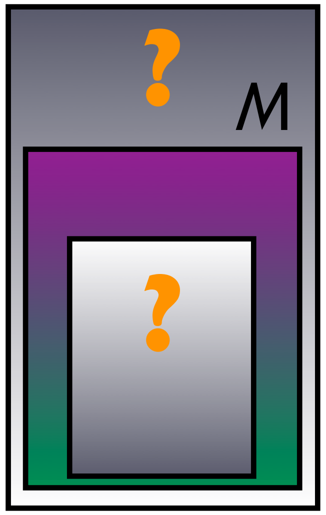
> i.e., for all $C \sqsubseteq D$ with $\text{sig}(C \sqsubseteq D) \subseteq \Sigma$:
> $O \models C \sqsubseteq D$ iff $M(\Sigma,O) \models C \sqsubseteq D$



M({part}, Mereology.owl) = {Trans: part,
part InverseOf: PartOf,
Trans: partOf}

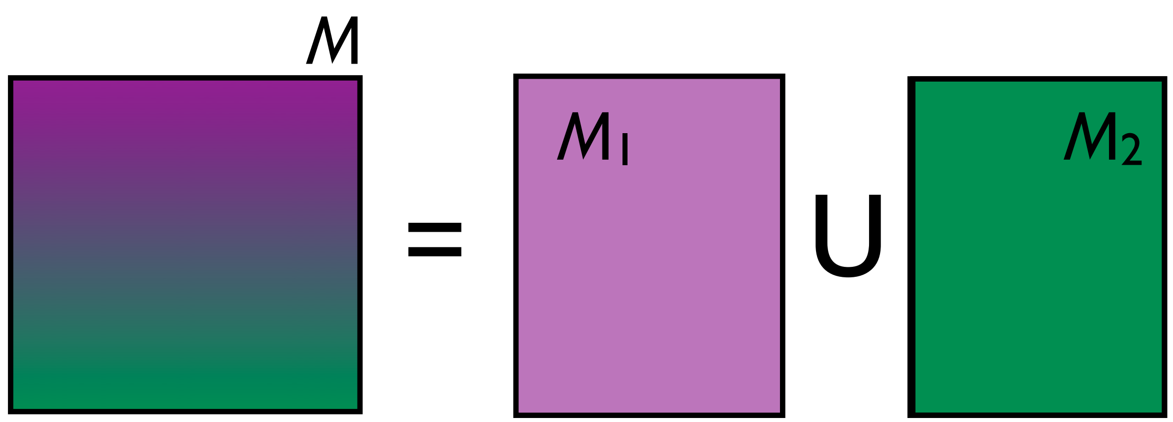
Modular *Structure*

- Modules are great...if you know your (seed) signature...
  - and for "module local" tasks such as reuse
- *Single* module extraction does *not* help if you
  - do *not* know the *right* seed signature
  - want to understand *other* modules
  - want to understand *axiom dependency structure*



- To analyse the *modular structure* of the ontology:
  - **significant** modules
  - **significant** relations between modules
  - ...which reveals logical dependence between axioms

Are all modules significant?

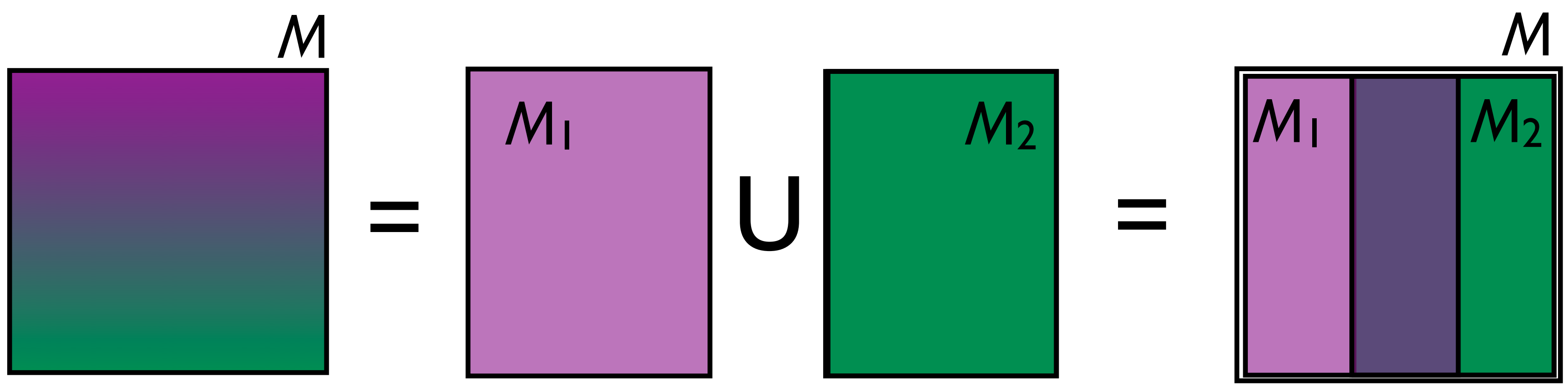


- To understand M, one must
  - understand the dependency structure of M_1
  - understand the dependency structure of M_2
  - **nothing** else: M_1 and M_2 have no further dependencies

➡ M is **not** significant: it is a ***fake*** module

  - Thus, M_1 and M_2 may be "significant"
  - knowing that M is "only" a union is important

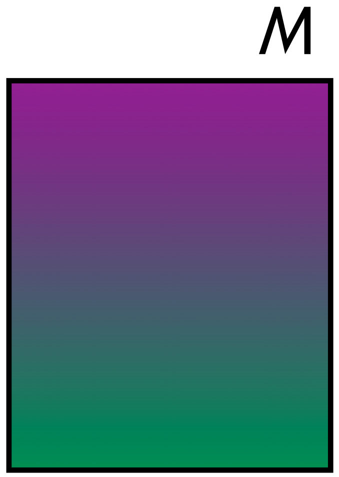
Are all modules significant?



- To understand M, one must
  - understand the dependency structure of M_1
  - understand the dependency structure of M_2
  - **nothing** else: M_1 and M_2 have no further dependencies
- ➡ M is **not** significant: it is a ***fake*** module
  - Thus, M_1 and M_2 may be "significant"
  - knowing that M is "only" a union is important

Are all modules significant?

- Consider a module *M* that is ***not fake***
- To understand *M*, one has to understand *M* **as a whole**
  - all axioms in *M* logically interact
  - in different ways – but interact
- Not fake implies significant: ***genuine***



Ratio of Fake to Genuine

- Given a set of genuine modules
  - unions lead to fake modules,
  - ➡ the space of fake modules is large (exponential)
  - but not *every* union of genuine modules is a module
- The cardinality of the set of all modules can and does grow exponentially in the size of *O*
  - See D., P., S., S., KR 2010 & WoMO 2010
- *Is module growth primarily due to trivial combinations?*
  - are most modules **fake**?

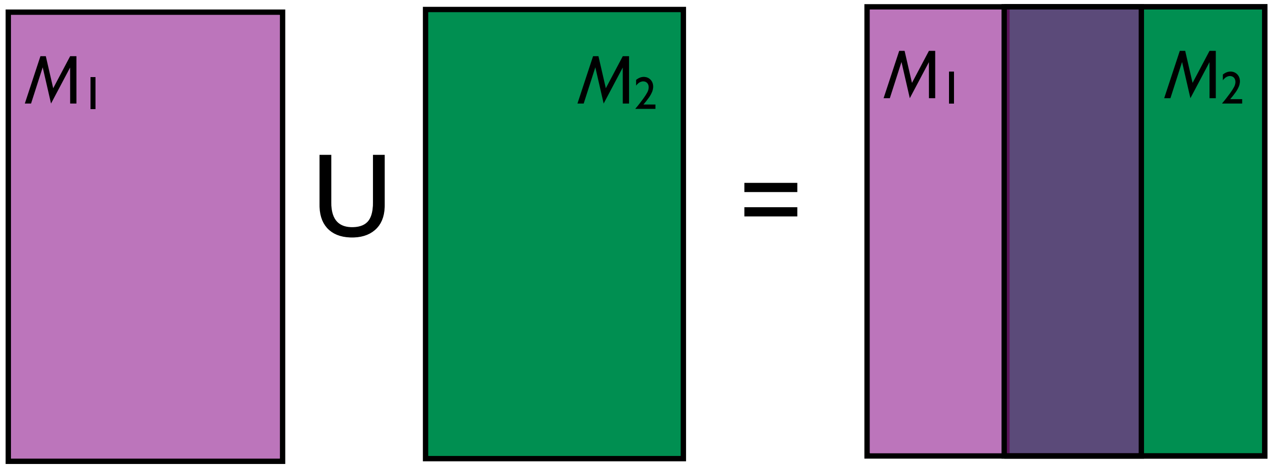
Yes!

Theorem 1: Each genuine module is the smallest module for some axiom $\alpha \in O$.

- The family of genuine modules is linear in $|O|$
  - ★ Most modules are fake!
- Proof exploits properties of modules
  - uniqueness, monotonicity, self-containedness, …
  - which are satisfied by all locality-based modules

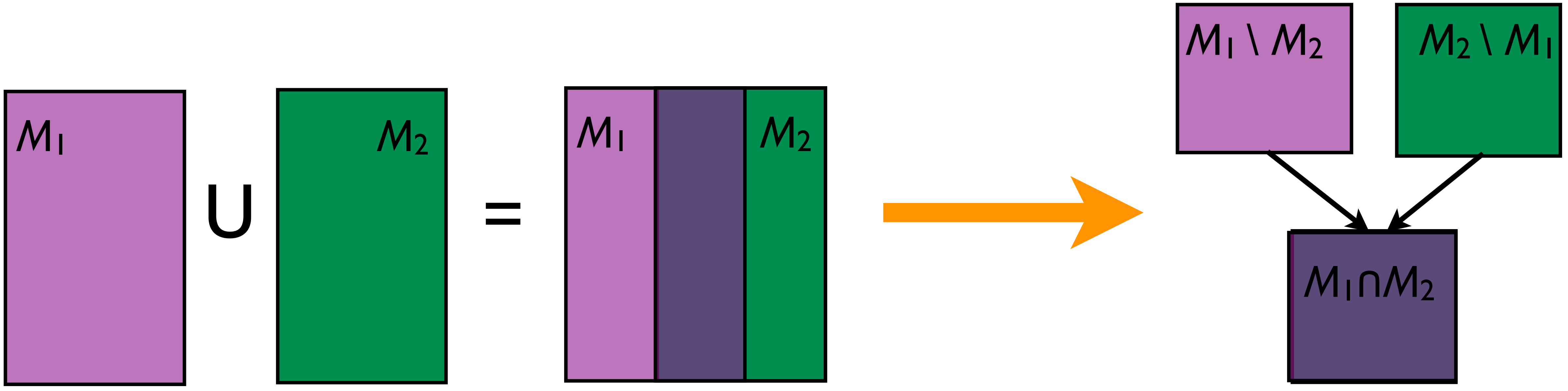
Relations between Modules

- Genuine modules may overlap
- This exposes significant logical dependence between axioms:
  - axioms in $M_1 \setminus M_2$ depend on axioms in $M_1 \cap M_2$

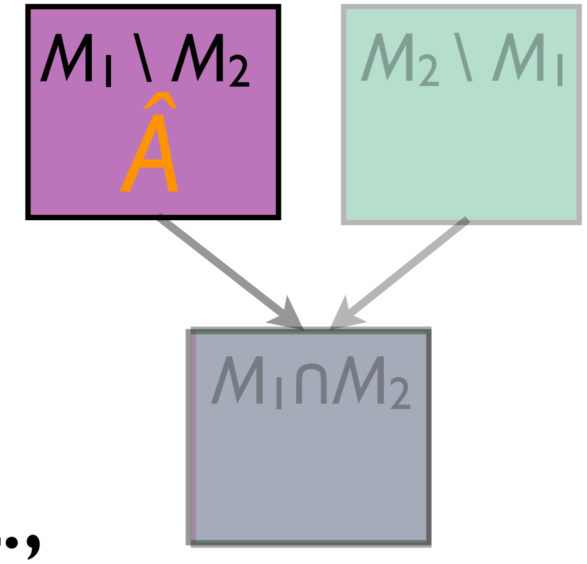


Relations between Modules

- Genuine modules may overlap
- This exposes significant logical dependence between axioms:
  - axioms in $M_1 \setminus M_2$ depend on axioms in $M_1 \cap M_2$



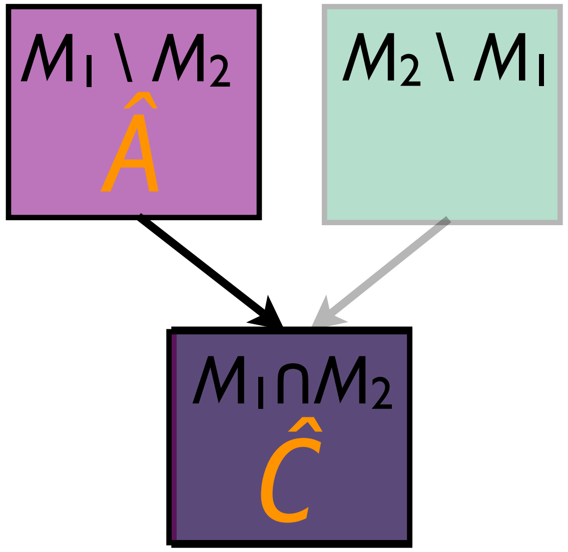
Atoms



- $\hat{A} \subseteq O$ is an **atom** if it is a maximal set s.t., for each module M, either $\hat{A} \subseteq M$ or $\hat{A} \cap M = \varnothing$.
  - ➡ The smallest module for an axiom α contains the whole atom to which α belongs!
  - ➡ Axioms in an atom are logically interdependent
  - ➡ Any two atoms are disjoint
  - ➡ The family of atoms is a partition of the ontology
    - Only linearly many atoms
  - ➡ Each GM is a disjoint union of atoms

Proposition: There is a 1-1 correspondence between genuine modules and atoms.

Atomic Decomposition



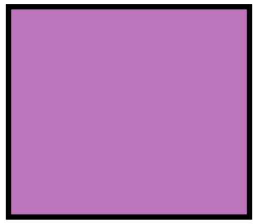
- Dependence between atoms:
  - $\hat{A} \succcurlyeq \hat{C}$ if, for each M: $\hat{A} \subseteq M$ implies $\hat{C} \subseteq M$
  - Axioms in $\hat{A}$ logically depend on axioms in $\hat{C}$

> Theorem 2: The relation $\succcurlyeq$ is reflexive, antisymmetric, and transitive.

- a Hasse diagram exposes 2 logical dependencies amongst axioms in atoms & between atoms

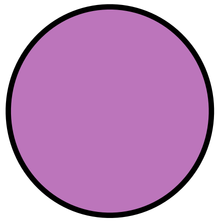
Mereology Ontology

42 axioms
1952 modules



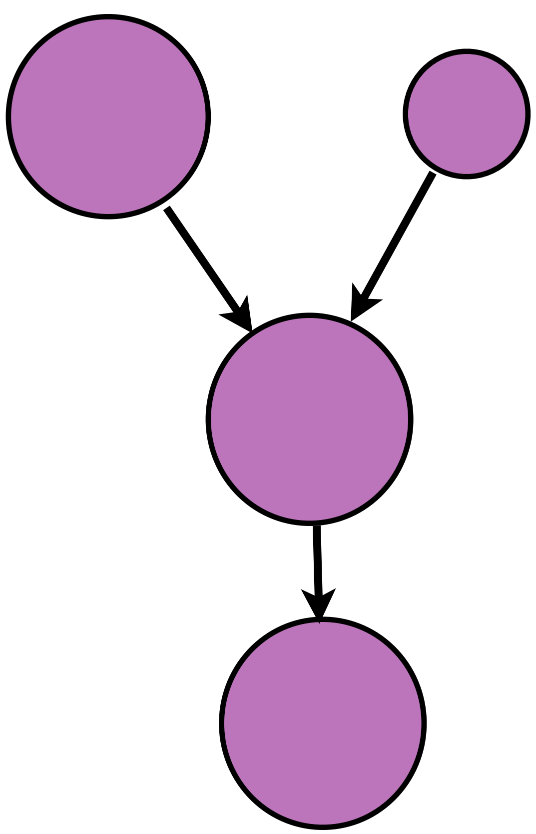
Mereology Ontology

42 axioms
1952 modules

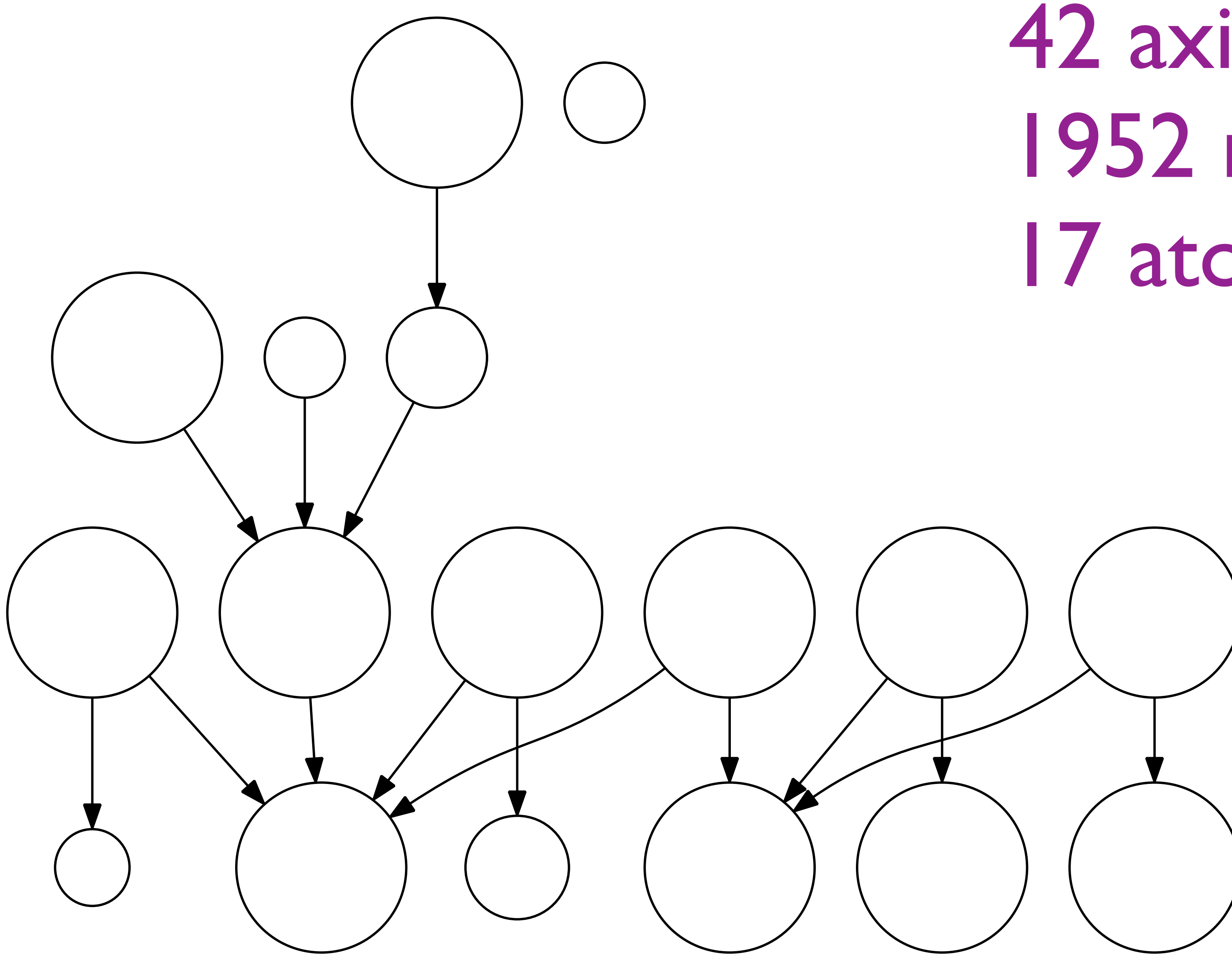


Mereology Ontology

42 axioms
1952 modules



Mereology Ontology



42 axioms
1952 modules
17 atoms/GMs

- Can we compute all genuine modules?
  - and all atoms
  - with their dependencies?
  - ...without computing all modules?!

Yes!

- Remember:

> Theorem 1: Each genuine module is the smallest module for some axiom $\alpha \in O$.

- extract $M(\text{sig}(\alpha), O)$
  - $\leq$ linearly many module extractions
- AD induced by the comparison of GMs
  - quadratic procedure

In Reality?

- We have decomposed 181 OWL ontologies from NCBO BioPortal
- Decomposability: average
  - nr. axioms/atom: 1.73
  - max nr. axioms/atom: 86
  - nr. axioms/GM: 66
  - max nr. axioms/GM: 143

Future Work

- More on dependency of axioms
  - between atoms and **sets** of atoms
- Labels for atoms
  - different labels for different tasks
- Applications
  - All Module Count
  - Fast Module Extraction
  - Topicality for Ontology Comprehension: see ICCS 2011
  - . . .

Thank you! – Questions?

Decomposability Issues

Ontology $\mathcal{O}$ (ID in BioPortal)	$\#\mathcal{O}$	#max Atom	#Eq. axs.	#Disj. axs.
Nanoparticle Ontology (1083)	16,267	6,425	42	6,106
Breast Tissue Cell Lines Ontology (1438)	2,734	2,201	0	7
IMGT Ontology (1491)	1,112	729	38	594
SNP Ontology (1058)	3,481	598	30	210
Amino Acid Ontology (1054)	477	445	8	190
Comparative Data Analysis (1128)	804	434	8	190
Family Health History (1126)	1,091	378	0	1
Neural Electromagnetic Ontologies (1321)	2,286	259	21	0
Computer-based Patient Record Ontology (1059)	1,454	238	18	20
Basic Formal Ontology (1332)	95	89	13	41
Ontology of Medically-related Social Entities (1565)	138	100	17	41
Ontology for General Medical Science (1414)	194	102	17	41
Cancer Research and Mgmt Acgt Master (1130)	5,435	3,796	16	42