FAIR Data Infrastructure for computation

Mapping out the Space of Density Functionals

Nathan Daelman

José M. Pizarro, J.F. Rudzinski, L.M.Ghiringhelli, M. A. L. Marques, S. Botti
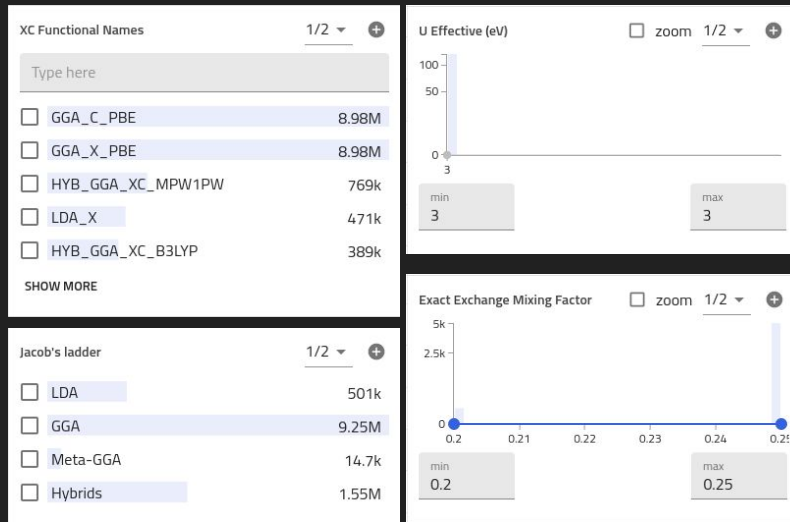
Motivation

Density Functional Theory in NOMAD

- over 12 M entries
- 88 subtypes registered

Heterogeneity: boom or bane?

- Analysis: insights in relevant physics
- Benchmarking

Interoperable BUT zoo of functionals

Need for a Knowledge Graph

LibXC

- basis for NOMAD XC Functional naming convention
- library for evaluating (semi-local) functionals
- supported by over 20 *ab inititio* packages
- v5.2.3 contains 671 entries
- **not a database!**

First Step: FAIRifying

- already **A**ccessible
- make information **F**indable
- *graph-based* paradigm: functional descriptors and *connections*
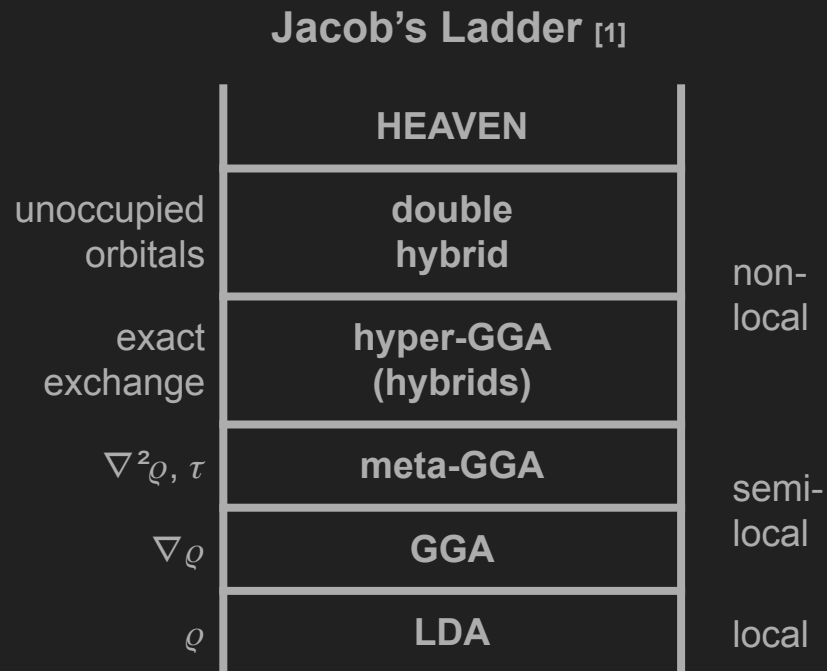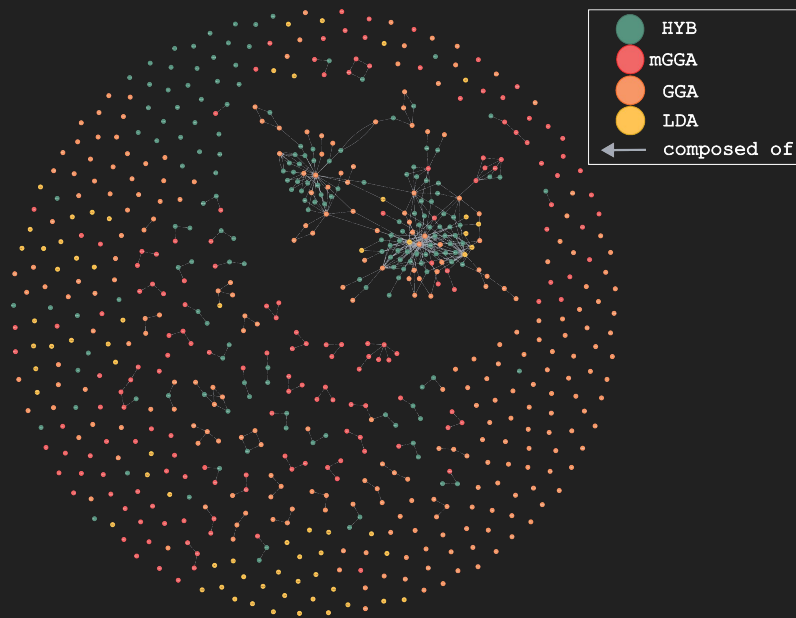
XC Functional Names	1/2 ▾ ⊕
Type here	
☐ GGA_C_PBE	8.98M
☐ GGA_X_PBE	8.98M
☐ HYB_GGA_XC_MPW1PW	769k
☐ LDA_X	471k
☐ HYB_GGA_XC_B3LYP	389k
SHOW MORE	

Building a Knowledge Graph

Jacob's Ladder [1]

	HEAVEN	
unoccupied orbitals	**double hybrid**	non-local
exact exchange	**hyper-GGA (hybrids)**	
$\nabla^2\varrho, \tau$	**meta-GGA**	semi-local
$\nabla\varrho$	**GGA**	
ϱ	**LDA**	local

Legend:
- HYB
- mGGA
- GGA
- LDA
- ← composed of

[1] J. P. Perdew, AIP Conference Proceedings 577, 1-20 (2001), doi: 10.1063/1.1390175.

Adding References

Legend:
- HYB
- mGGA
- GGA
- LDA
- Ref
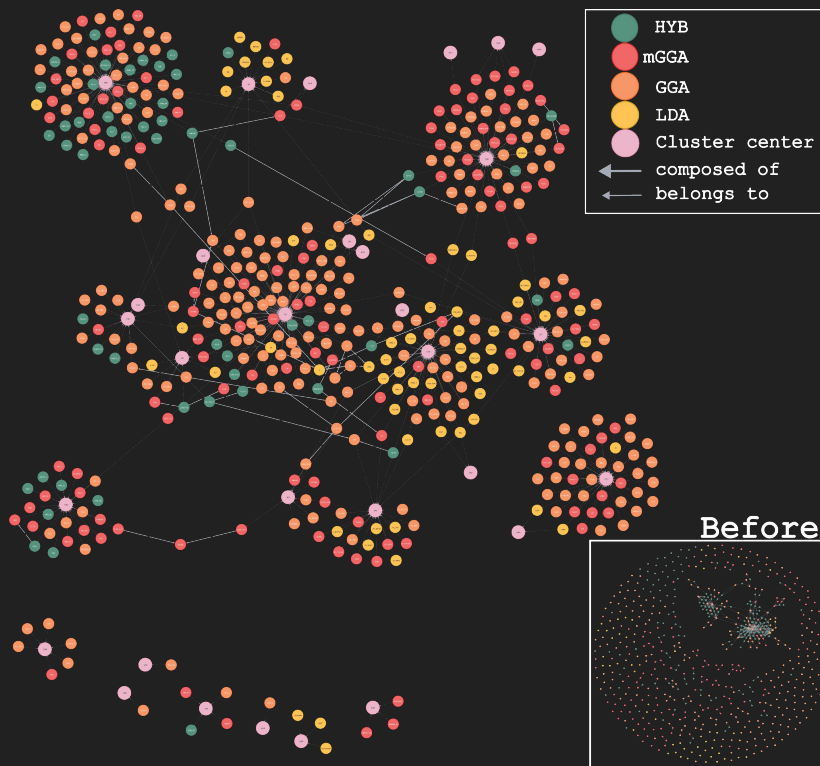- ← composed of
- ← references

Add structure via references

- 1st order: 701 (see figure)
- 2nd order: 5927

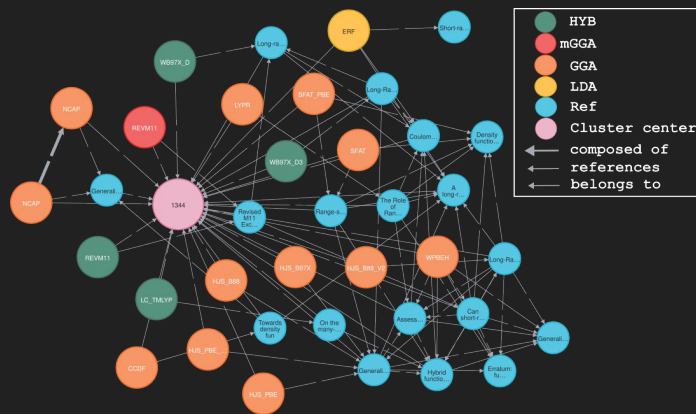**Collective consciousness** of the community: link what is proper

Clustering

Legend:
- HYB
- mGGA
- GGA
- LDA
- Cluster center
- ← composed of
- ← belongs to

Before:

Approach
- create references subgraph
- apply Louvain clustering
- connect cluster centres back to functionals:

🔴 → (🔵) → 🟡🟠🔴🟢

Semantics: do these clusters have meaning?

Adding Semantics: PageRank

Interpreting a cluster's theme
- Articles in this case: 19
- Articles in the whole set overall: 5937
- Filter via the PageRank algorithm

[1]

Ranked reading list:
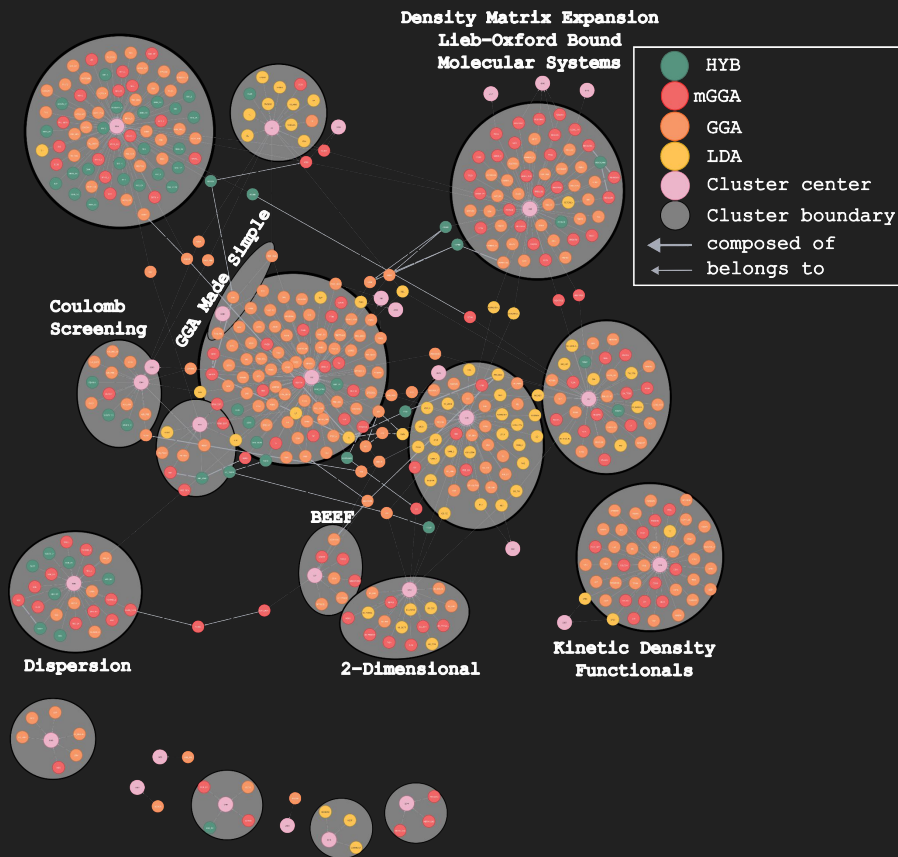- most talked articles
- most representative for a theme

[1] http://infolab.stanford.edu/~backrub/google.html

Adding Semantics: PageRank

PageRank	Publication	Authors
1.02	*Coulomb*-attenuated exchange energy density functionals	PMW Gill, RD Adamson, JA Pople
0.63	Hybrid functionals based on a *screened Coulomb* potential	J Heyd, GE Scuseria M Ernzerhof
0.60	A long-range-corrected time-dependent density functional theory	Y Tawada, T Tsuneda, S Yanagisawa
0.35	Assessment and validation of a *screened Coulomb* hybrid density functional	J Heyd, GE Scuseria

reduce to **max. 4 articles / cluster**
so approx. **80 in total**

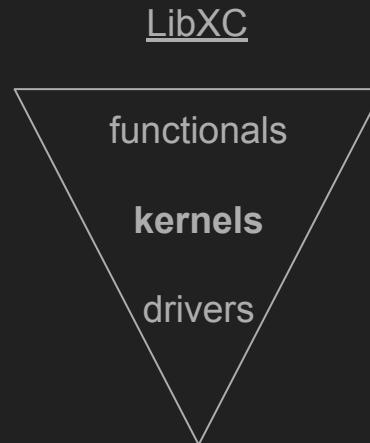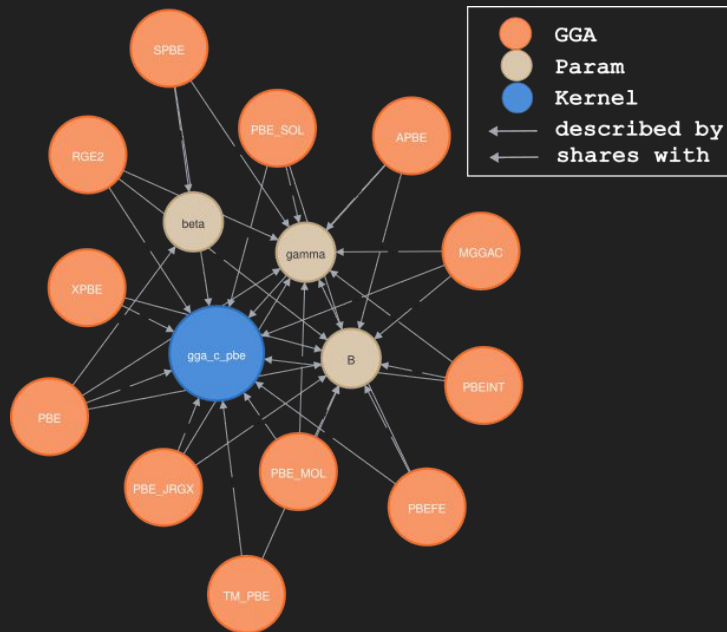# Adding Semantics

Categories
- Coulomb Screening
- 2-Dimensional
- Kinetic Density Functionals
- Dispersion
- BEEF / VdW

Thematic clusters + connections
- Density Matrix Expansion
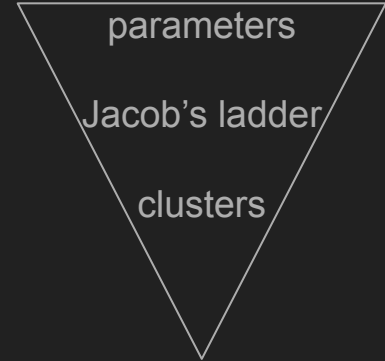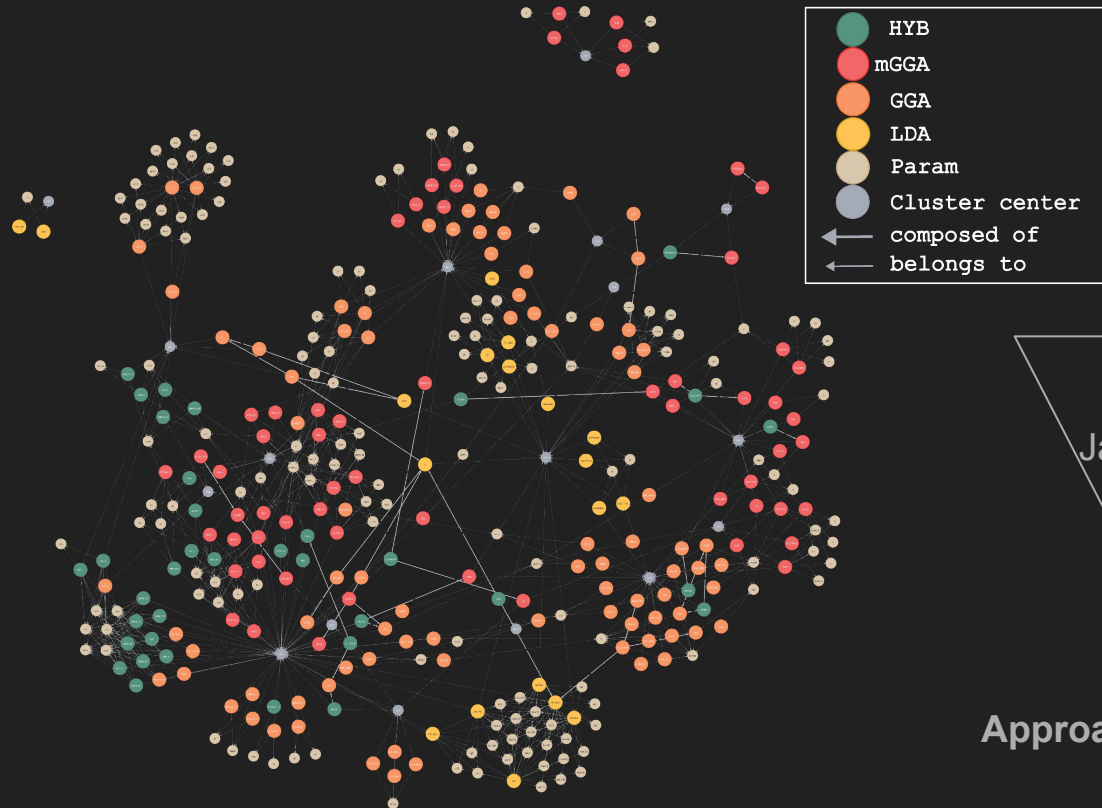- PBE and co.

Adding Semantics: Common Parameters

Fig.: PBE correlation kernel neighbourhood

functionals

kernels

drivers

Semantics
- described by: re-parametrization
- shares with: more common variables

Each **Param** has a story behind it

10

Adding Semantics: Common Parameters

Conclusions

Utility of **FAIR** knowledge bases / graphs

Future work

- make knowledge accessible to NOMAD users
  - composition
  - re-parametrization
  - clusters (categories, families)
  - extensions
- extend knowledge base
  - more semantic *references*: extract the *why*?
  - include other repositories? e.g. libvdwxc, Quantum ESPRESSO