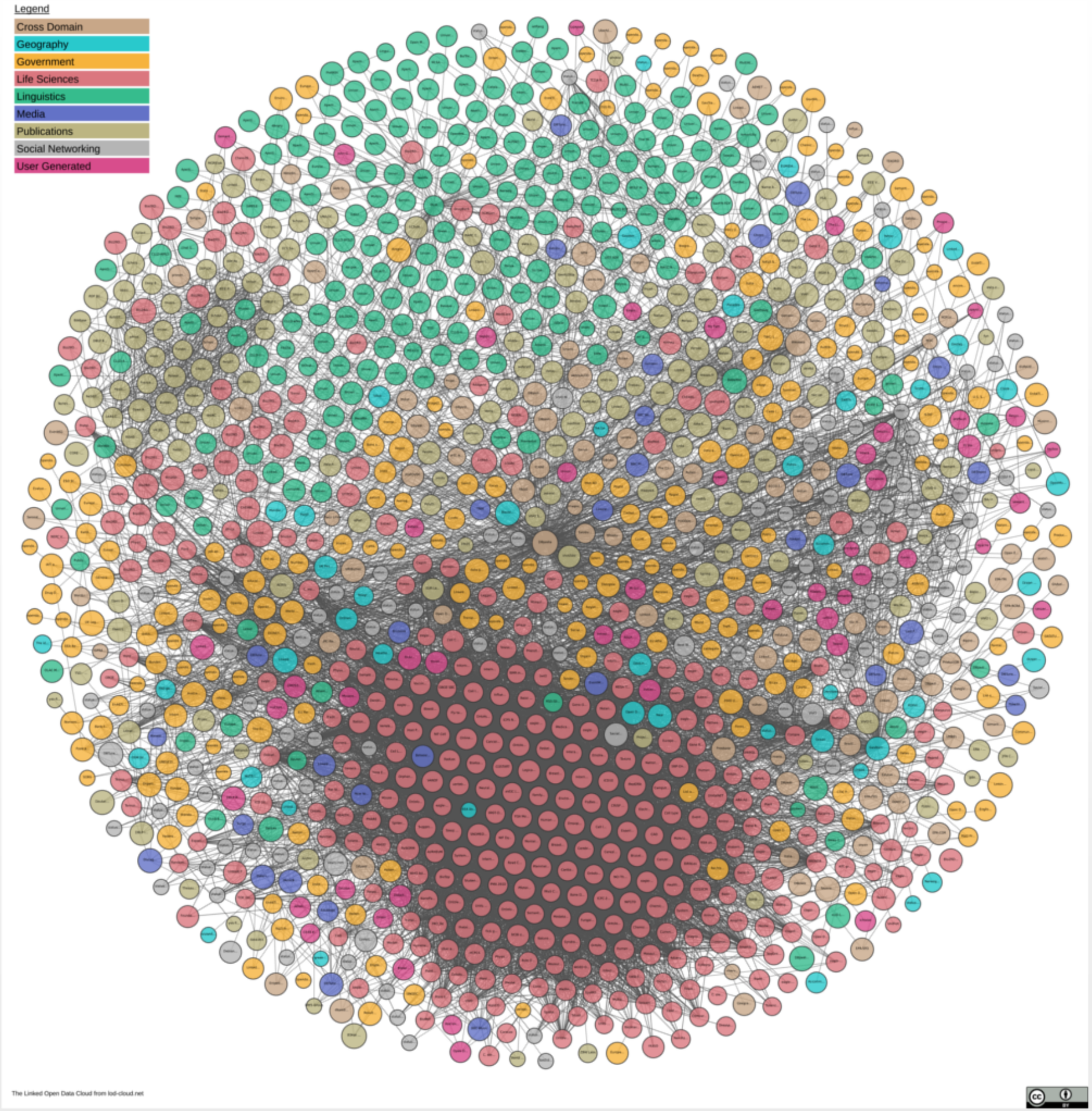
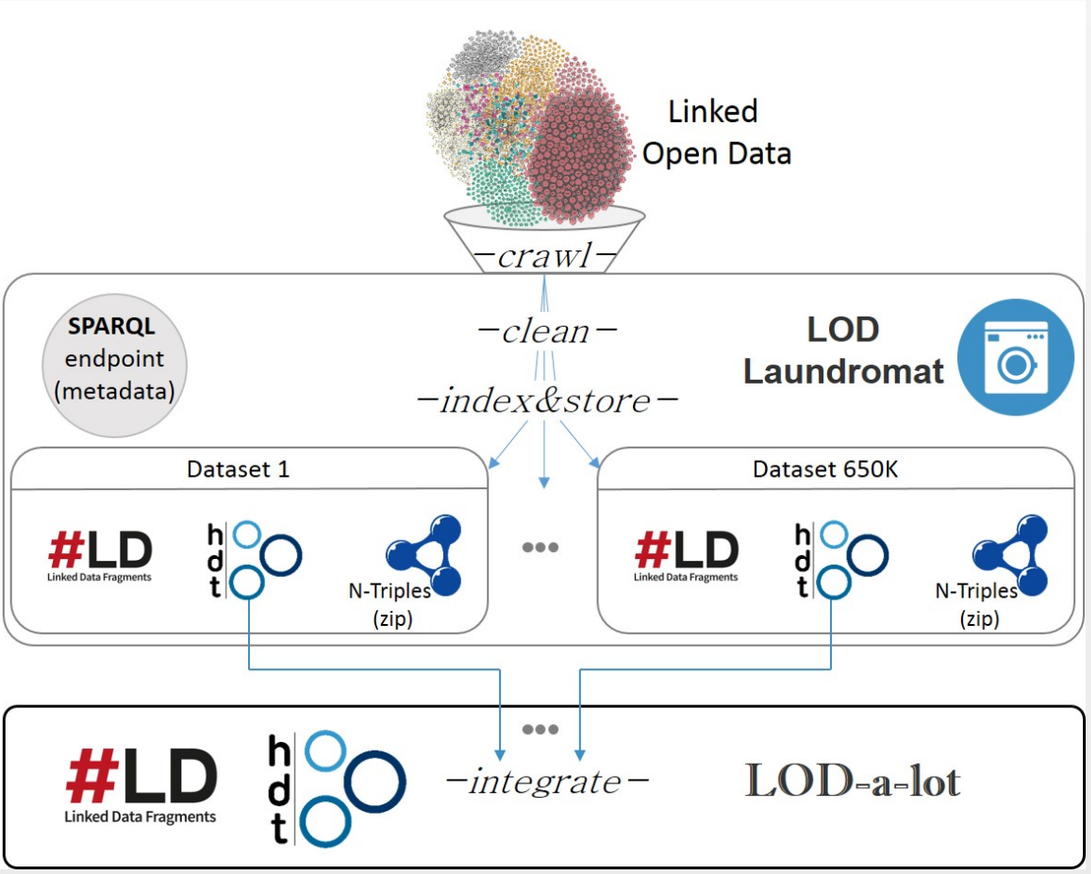
Observing LOD using Equivalent Set Graphs: it is mostly flat and sparsely linked

Luigi Asprino[1,2], Wouter Beek[3], Paolo Ciancarini[2], Frank van Harmelen[3] and Valentina Presutti[1,2]

1. STLab, ISTC-CNR, Rome, Italy
2. University of Bologna, Bologna, Italy
3. Dept. of Computer Science, VU University Amsterdam, NL

29/10/2019 - ISWC


Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
The Linked Open Data Cloud from lod-cloud.net



~ 28B unique triples

crawled from 650k datasets

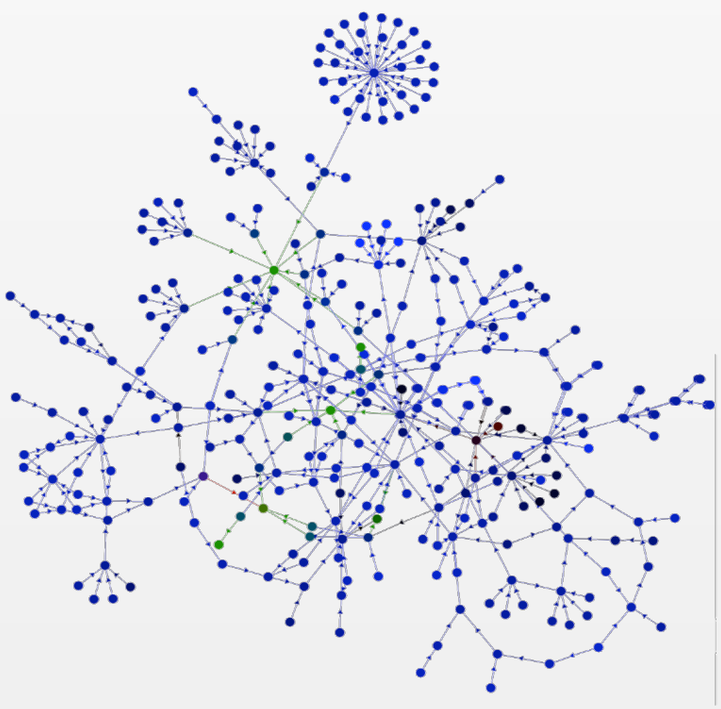
(Fernández et al. 2017)

http://lod-a-lot.lod.labs.vu.nl/

Research Questions

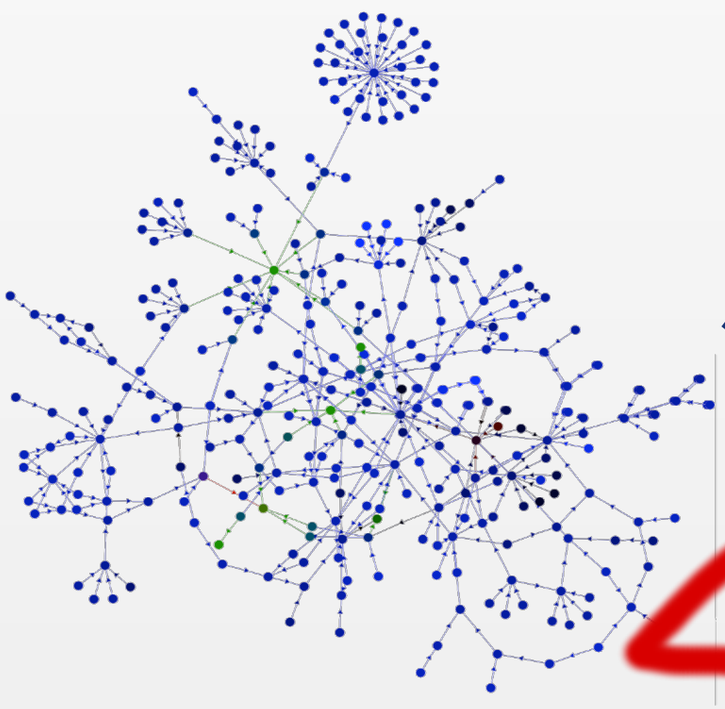
Research Questions

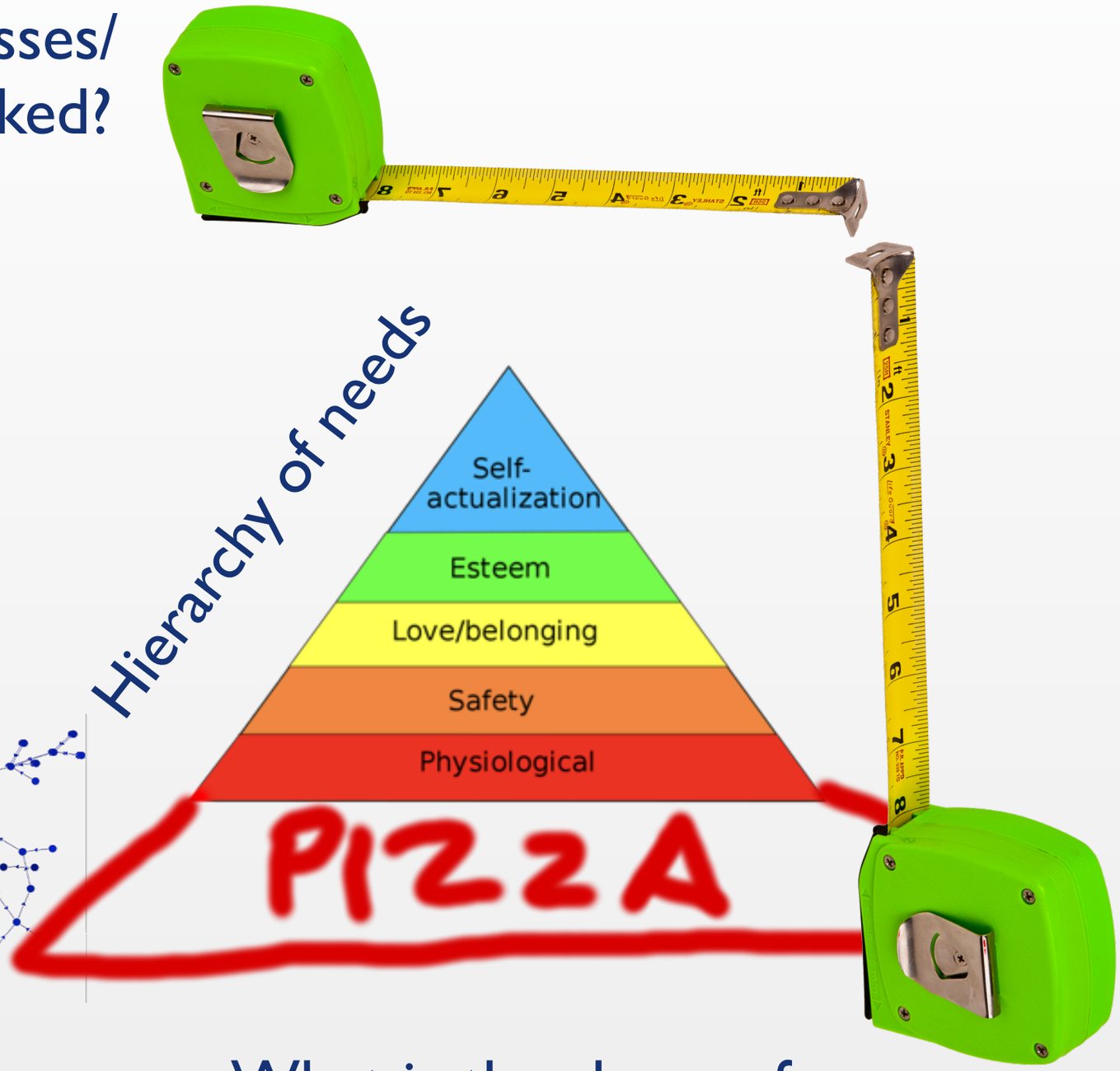
How much are classes/
properties interlinked?



Research Questions

How much are classes/ properties interlinked?



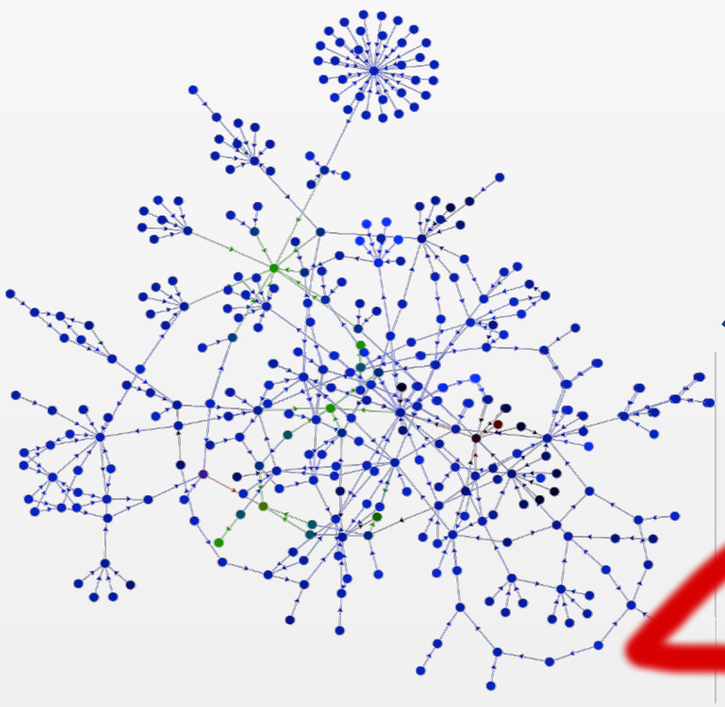


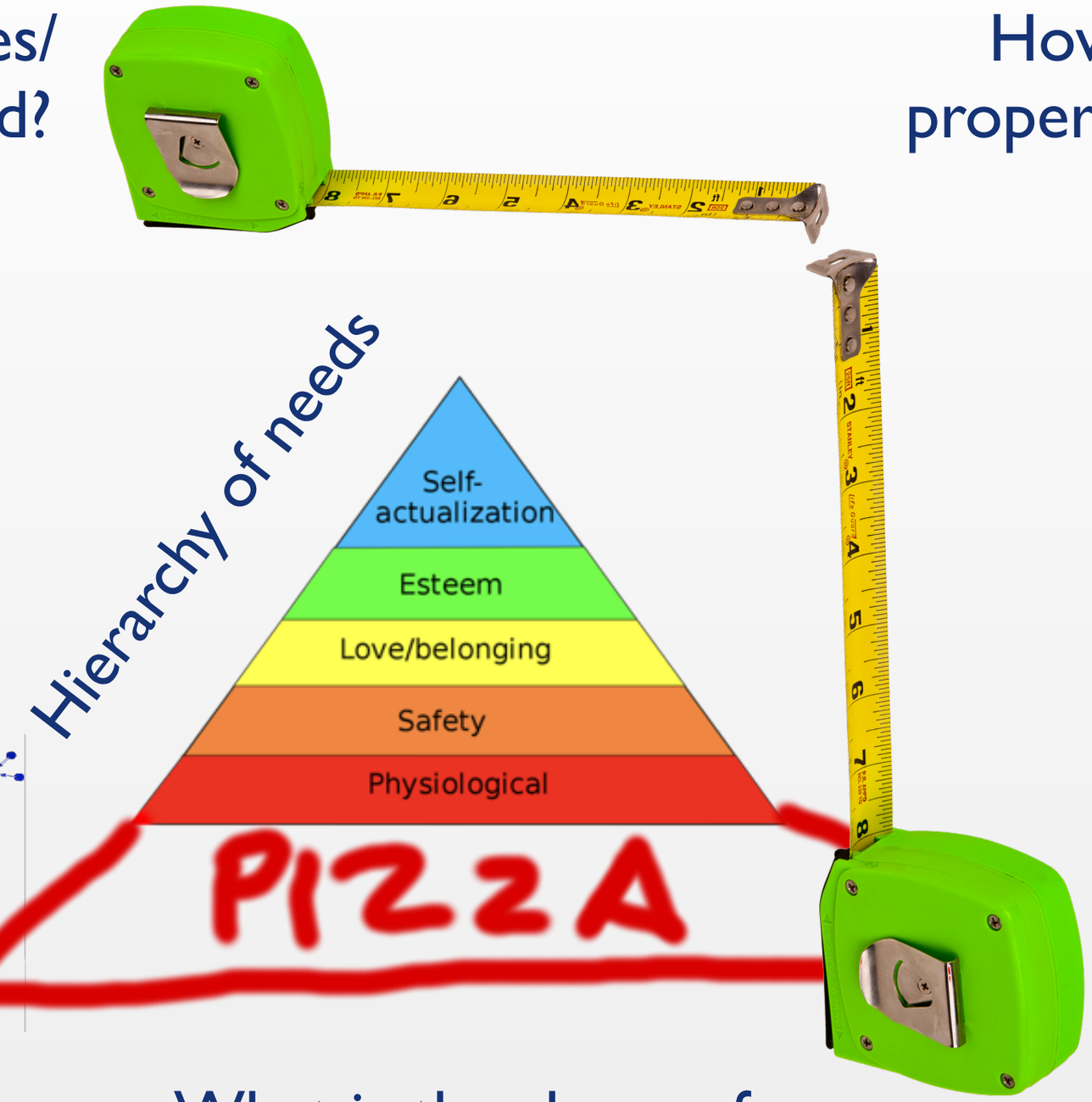
What is the shape of the ontological hierarchies?

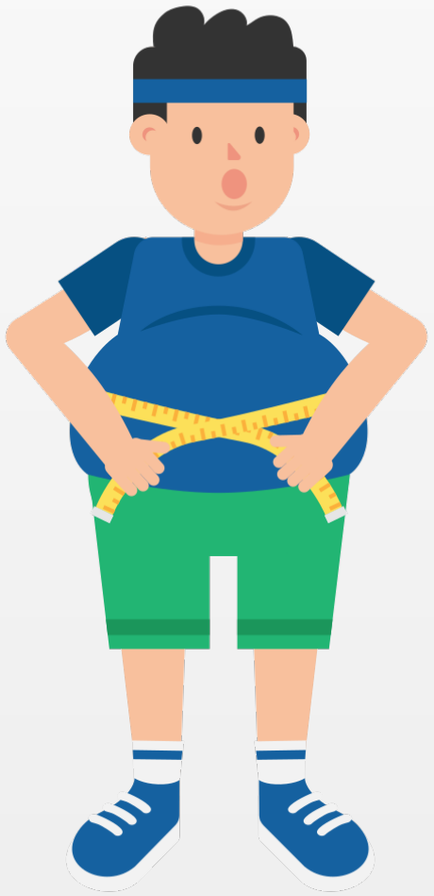
Research Questions

How much are classes/ properties interlinked?

How are classes / properties instantiated?







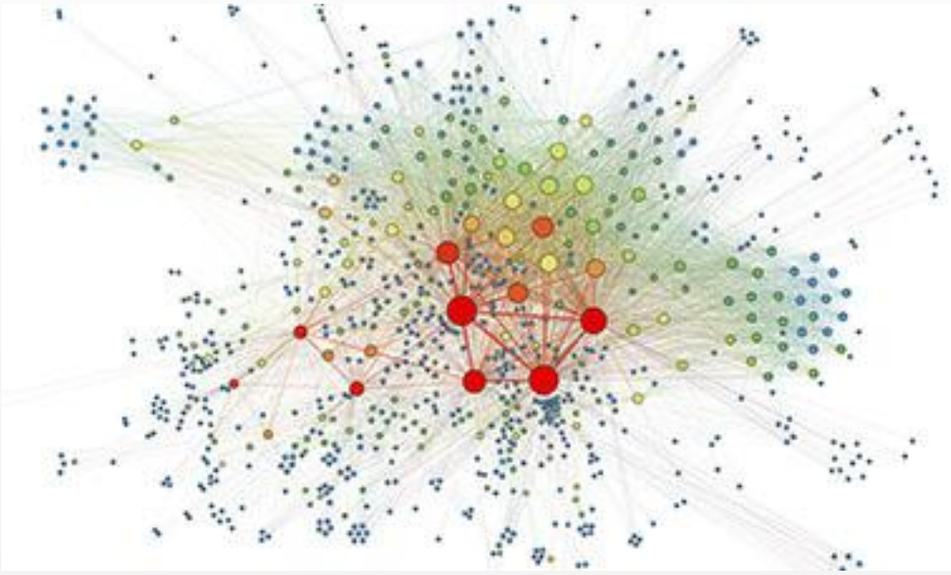
What is the shape of the ontological hierarchies?

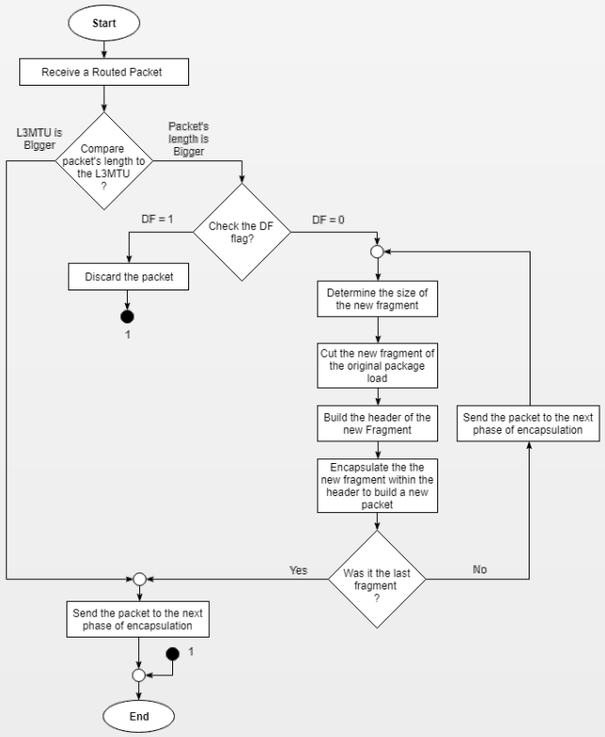
Related Work

- Early stage of SW: (Gil 2004), (Ding 2006), (Wang 2006), (Theoharis 2008), (Ding 2010)
- sameAs: (Ding 2010), (Halpin 2010), (Beek 2018)
- Blank Nodes: (Mallea 2011)
- Coherence of large datasets (Paulheim 2015)
- LOD Quality and adoption of best practices (Schmachtenberg 2014), (DeBattista 2018)
- Empirical Analysis of Knowledge patterns (Nuzzolese 2011)
- Foundational distinctions (Asprino 2018)

Contributions

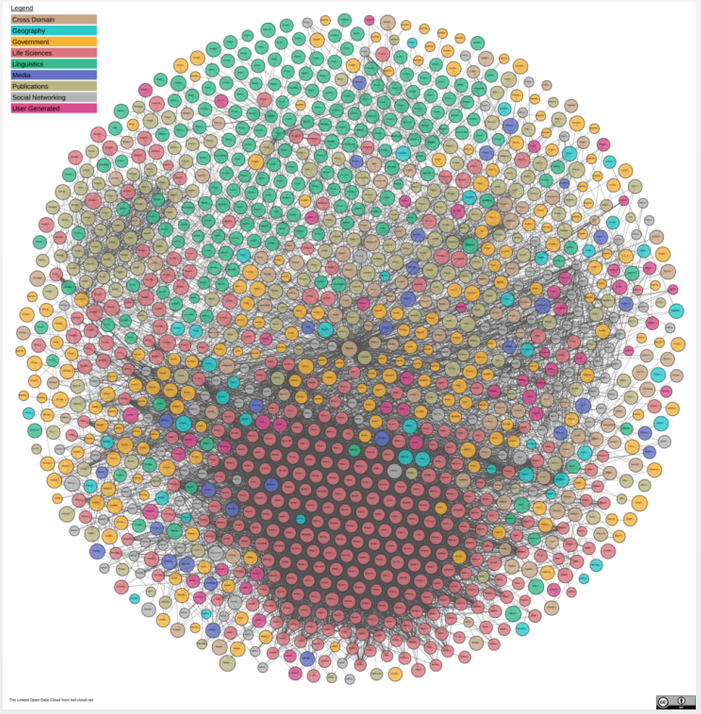
Equivalence Set Graph





Algorithm

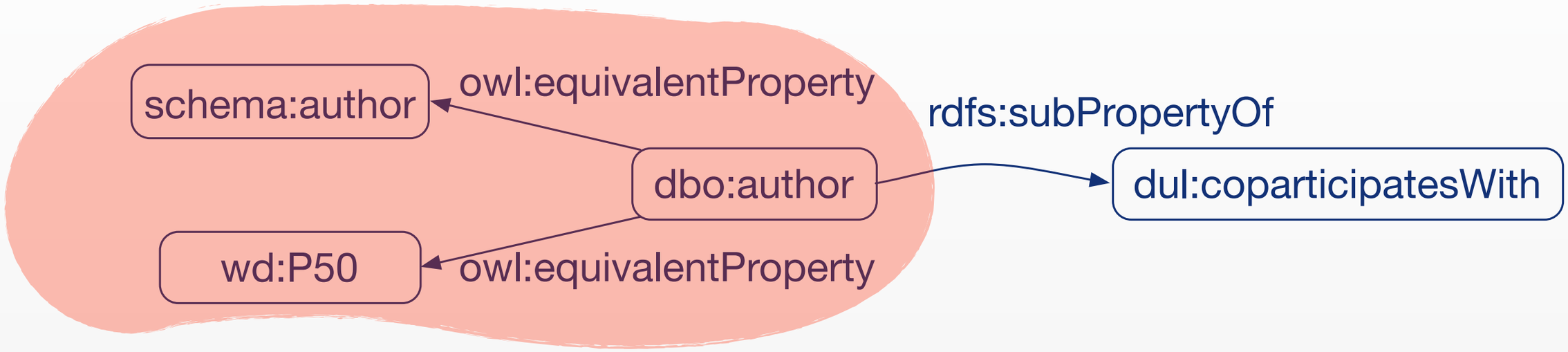
Metrics



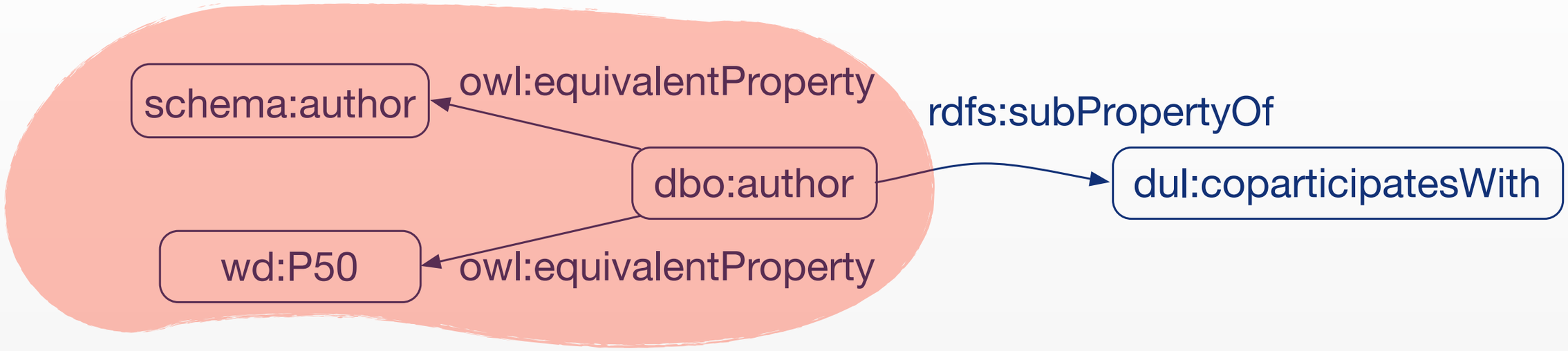
LOD Analysis

Equivalence Set Graph

Equivalence Set Graph

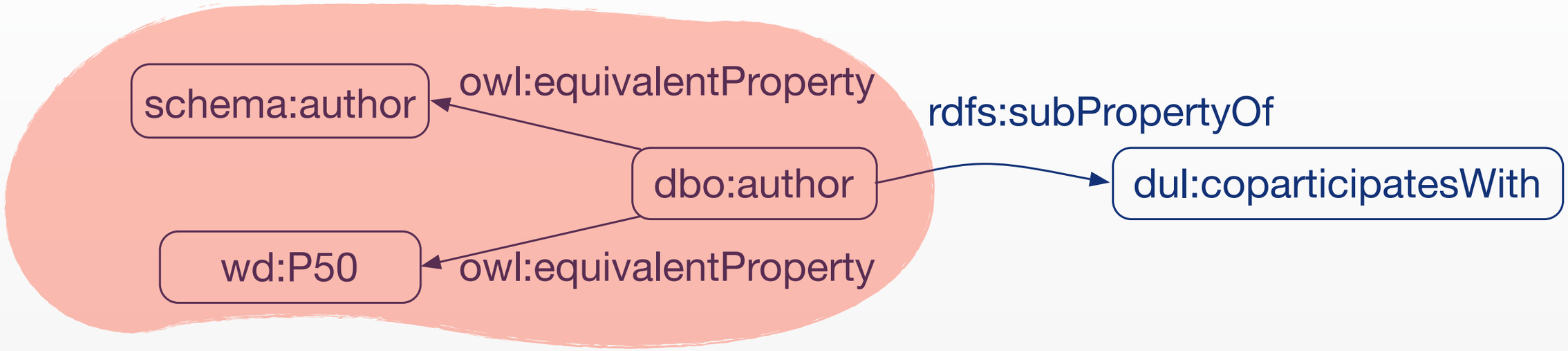


Equivalence Set Graph



Equivalence	Specialization
Reflexive	Reflexive
Symmetric	Anti-Symmetric
Transitive	Transitive

Equivalence Set Graph

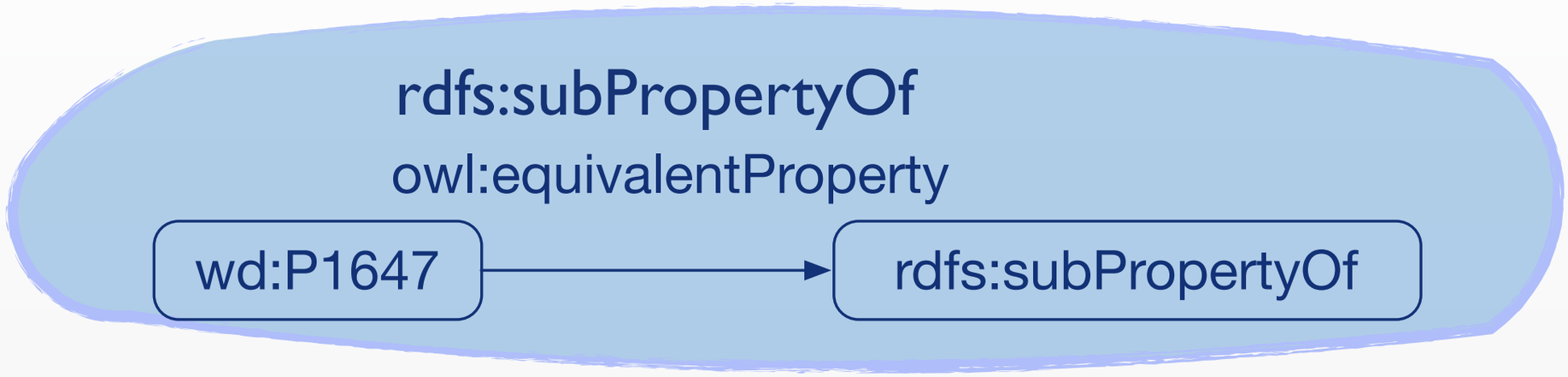


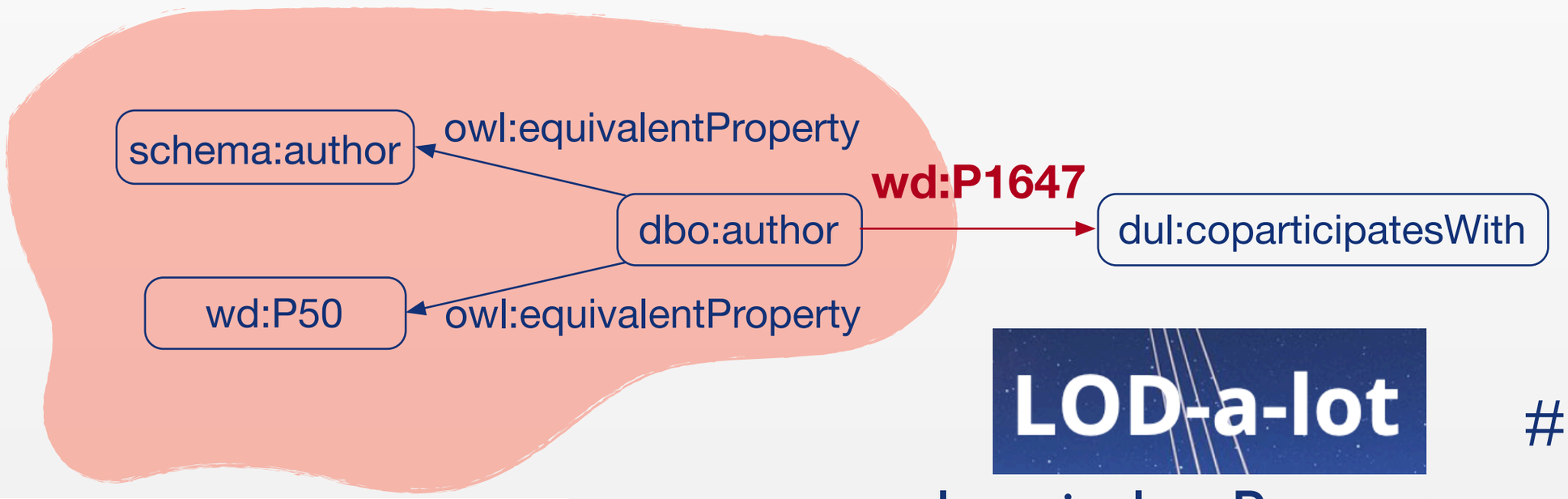
Equivalence	Specialization
Reflexive	Reflexive
Symmetric	Anti-Symmetric
Transitive	Transitive

LOD-a-lot

# Properties	~ 1.3 M
# Equivalent Sets	~ 1.3 M
Ratio	.997
Edges	~ 147 K

Equivalence/Specialization of standard properties





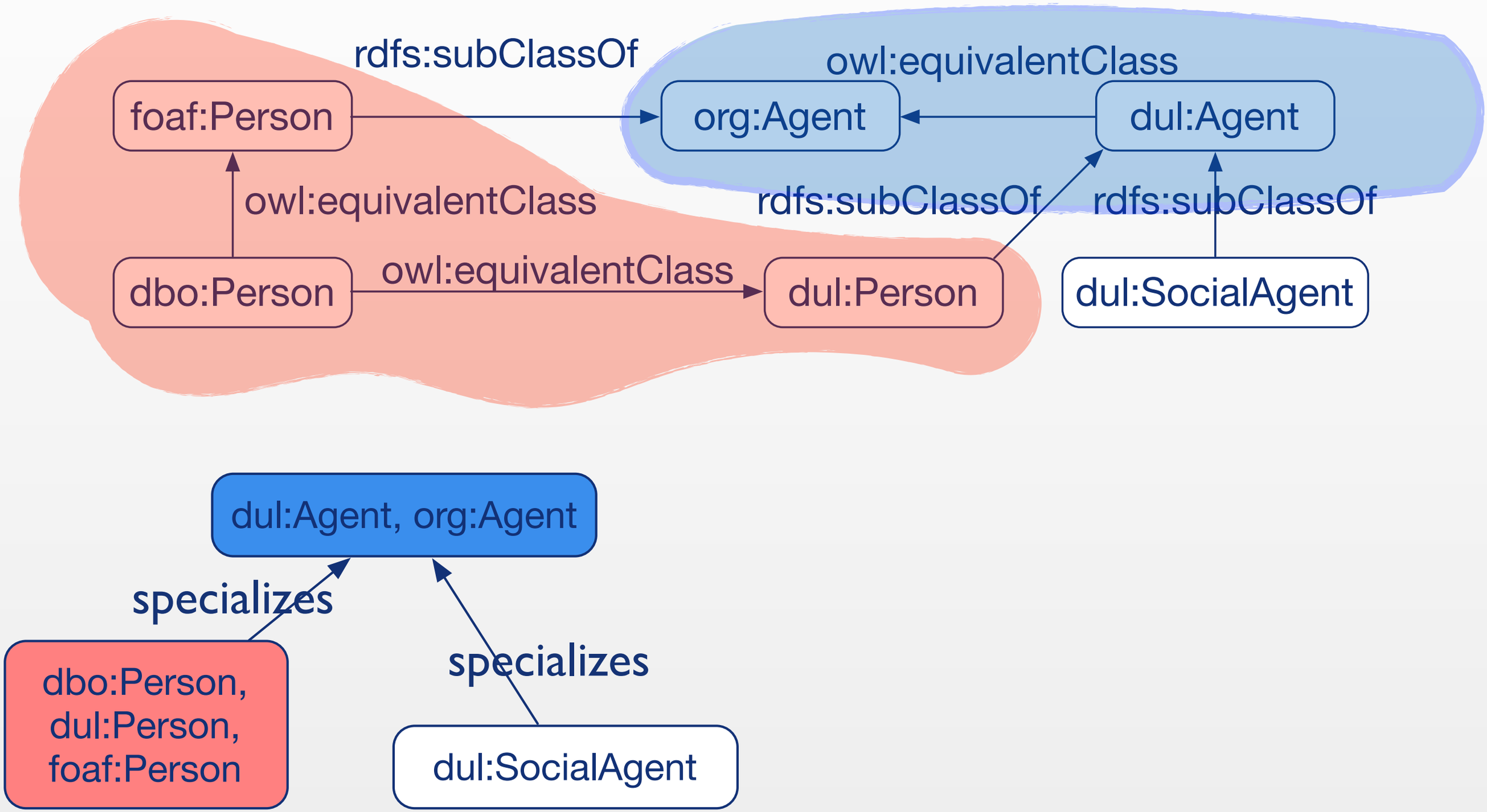
wd:P1647,
owl:equivalentProperty

schema:author,
wd:P50,
dbo:author
specializes
dul:copartipatesWith

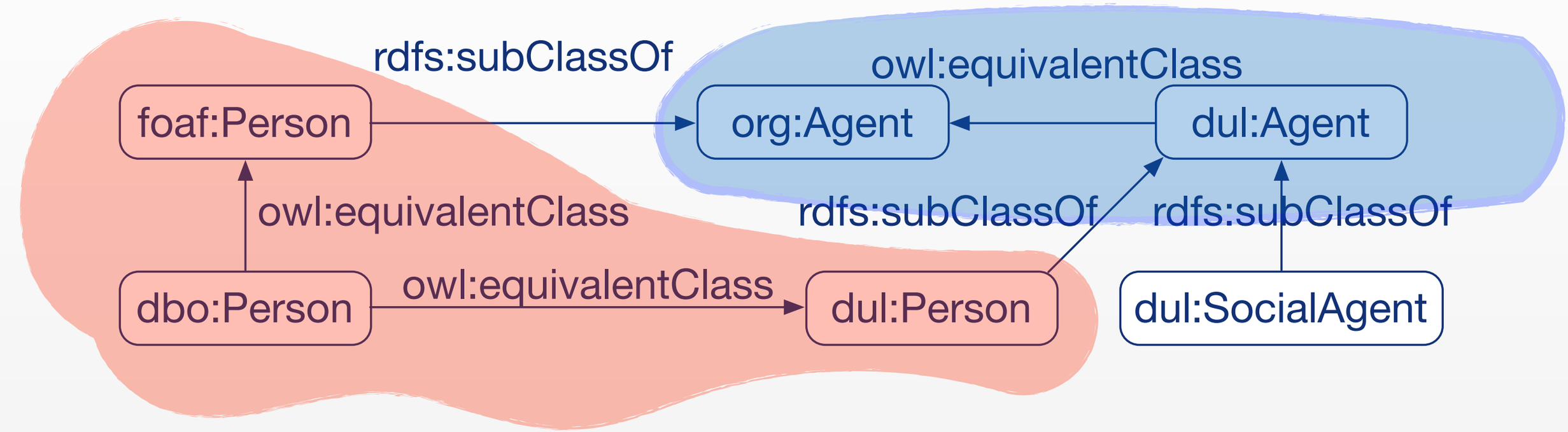
LOD-a-lot

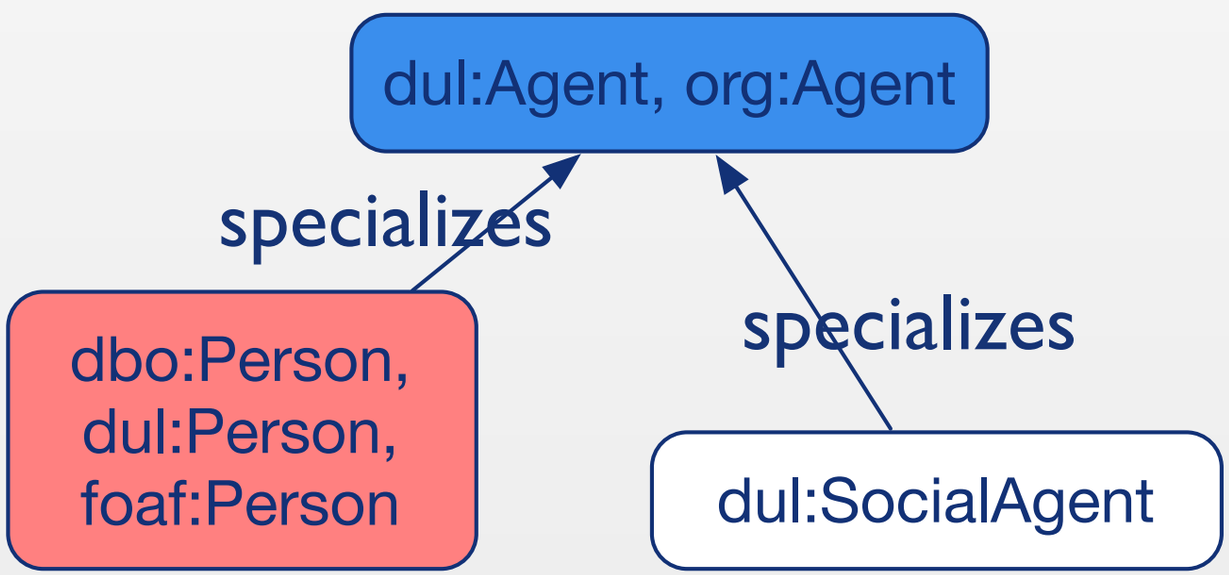
	# of ⊑
owl:equivalentProperty	0
rdfs:subPropertyOf	447
owl:equivalentClass	1
rdfs:subClassOf	379

Equivalence Set Graph



Equivalence Set Graph





LOD-a-lot

# Classes	~ 4.8 M
# Equivalent Sets	~ 4 M
Ratio	.831
Edges	~ 5 M

Computing ESG

Computing ESG

dbo:author owl:equivalentProperty schema:author .

Computing ESG

dbo:author owl:equivalentProperty schema:author .

dbo:author,
schema:author

Computing ESG

dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author,
schema:author

wd:P150
gnd:author

Computing ESG

dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author owl:equivalentProperty bibtex:hasAuthor .

dbo:author,
schema:author

wd:P150
gnd:author

Computing ESG

dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author owl:equivalentProperty bibtex:hasAuthor .

dbo:author,
schema:author,
bibtex:hasAuthor

wd:P150
gnd:author

Computing ESG

dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author owl:equivalentProperty bibtex:hasAuthor .

dbo:author owl:equivalentProperty gnd:author .

dbo:author,
schema:author,
bibtex:hasAuthor

wd:P150
gnd:author

Computing ESG

dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author owl:equivalentProperty bibtex:hasAuthor .

dbo:author owl:equivalentProperty gnd:author .

dbo:author,
schema:author,
bibtex:hasAuthor,
wd:P150
gnd:author

Computing ESG

dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author owl:equivalentProperty bibtex:hasAuthor .

dbo:author owl:equivalentProperty gnd:author .

dbo:author rdfs:subPropertyOf dul:coparticipatesWith .

dbo:author,
schema:author,
bibtex:hasAuthor,
wd:P150
gnd:author

Computing ESG

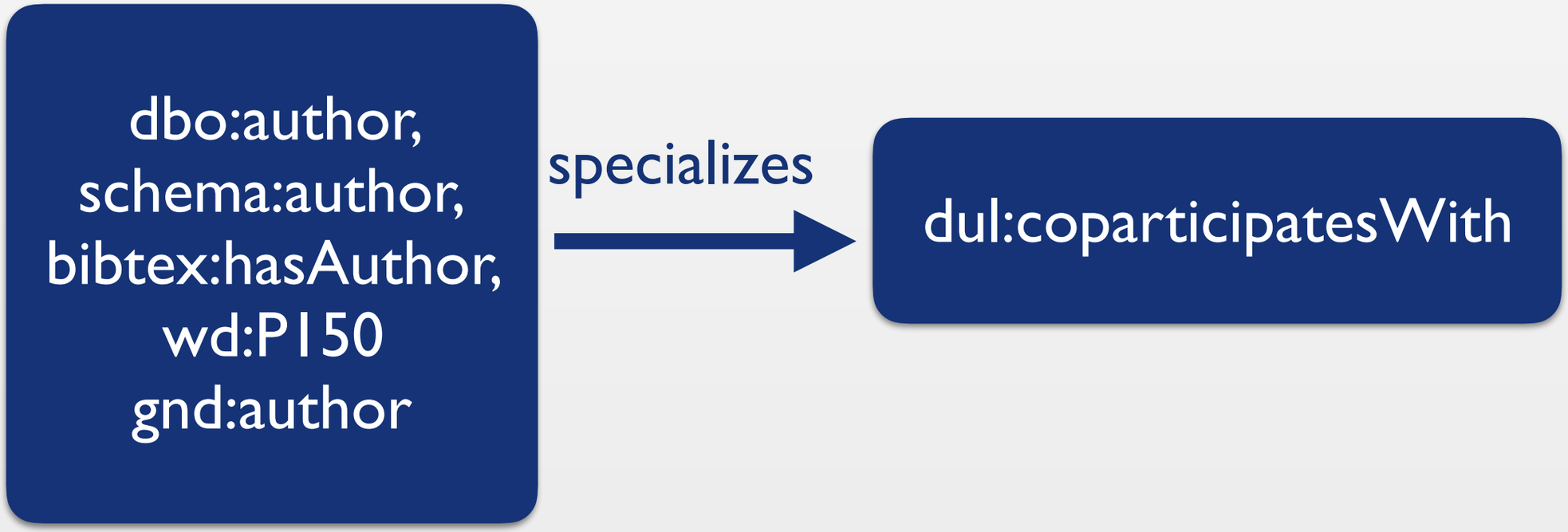
dbo:author owl:equivalentProperty schema:author .

wd:P150 owl:equivalentProperty gnd:author .

dbo:author owl:equivalentProperty bibtex:hasAuthor .

dbo:author owl:equivalentProperty gnd:author .

dbo:author rdfs:subPropertyOf dul:coparticipatesWith .



Time Complexity: $O(n_e^2+n_s)$

Space Complexity: $O(n^2)$

	ESG Computing time (on this laptop)
Properties	~ 4 hrs
Classes	~ 14 hrs

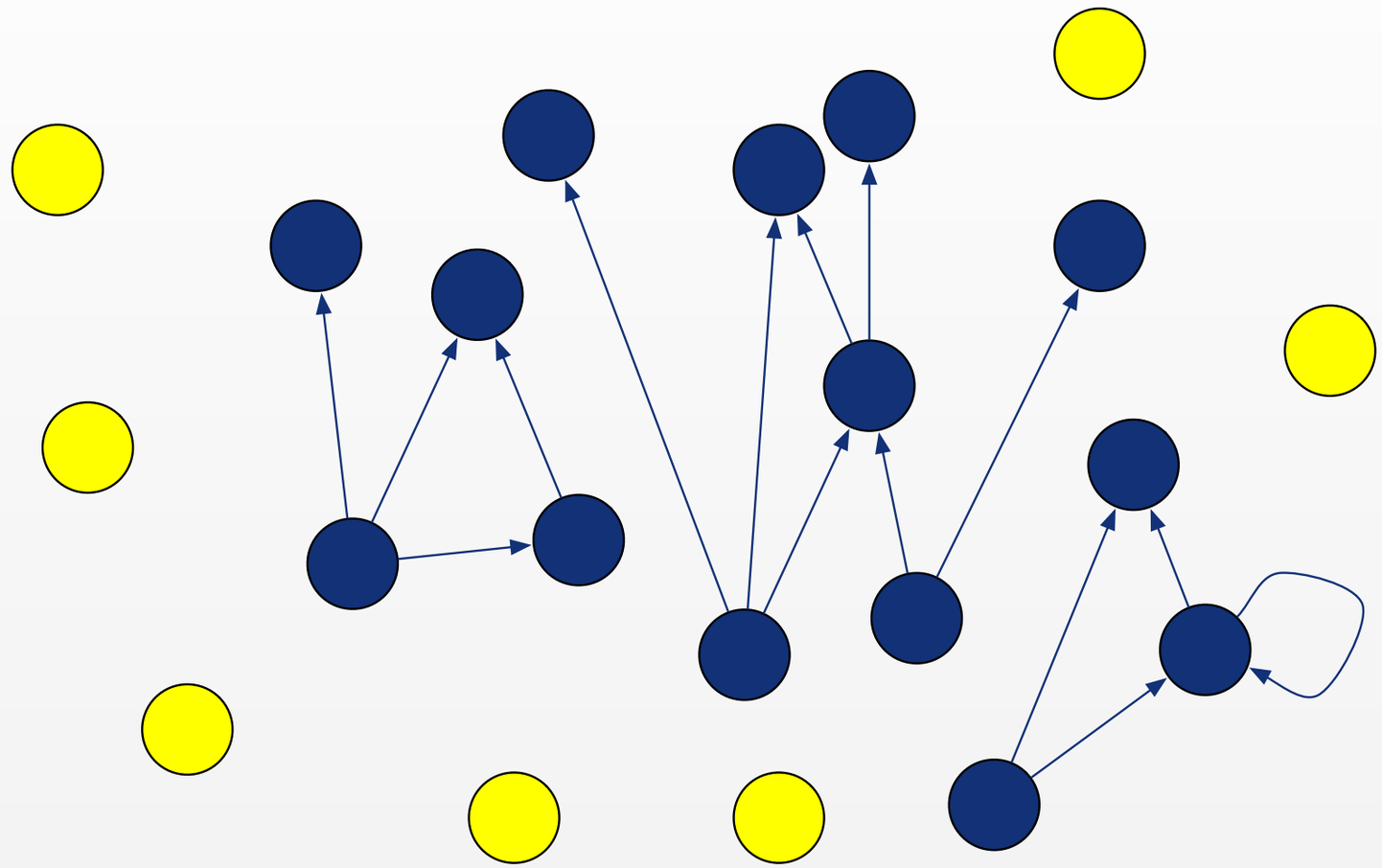
SCAN ME

RocksDB

SNAP

https://github.com/luigi-asprino/edwin

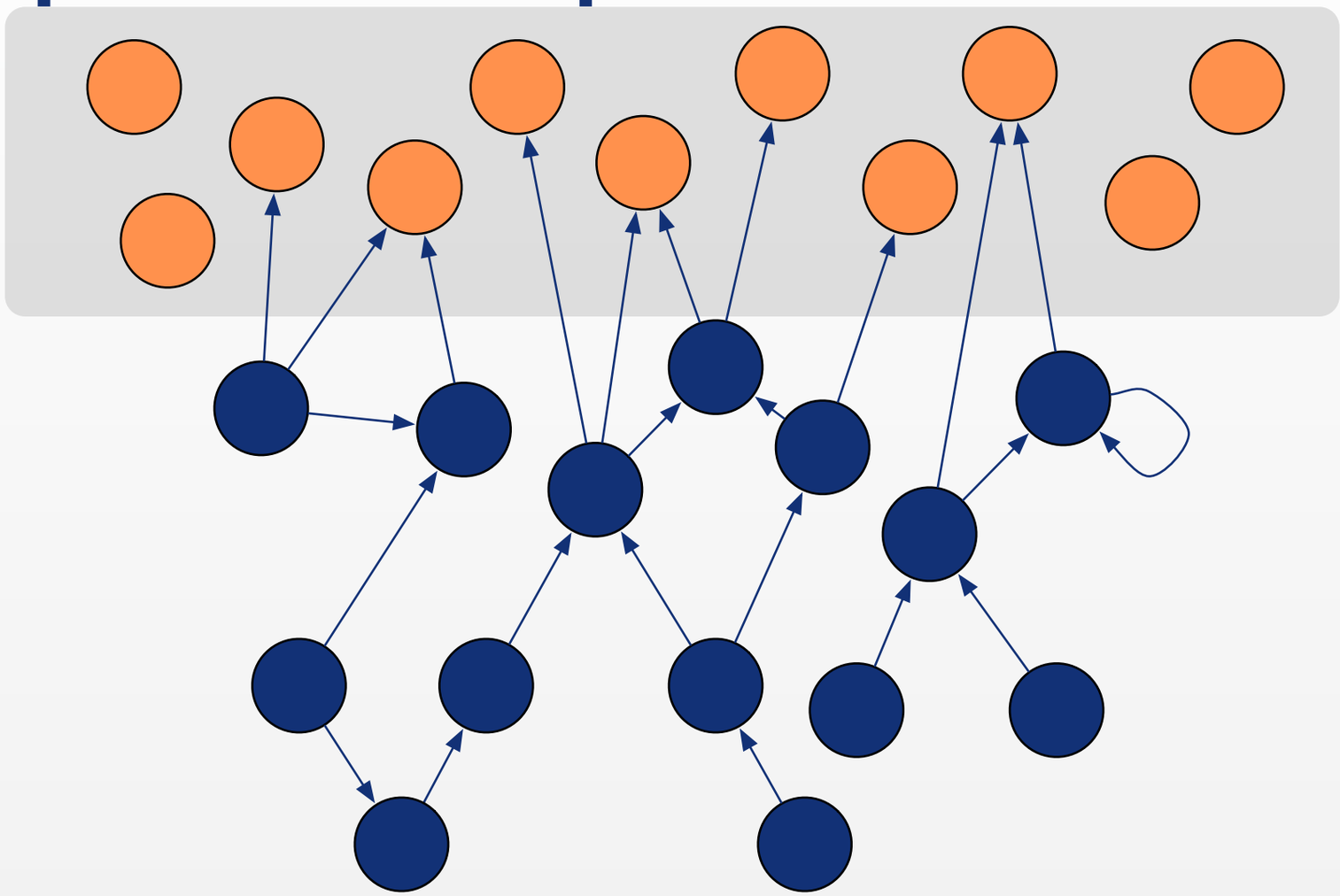
Isolated Equivalence Sets



LOD-a-lot

	# Isolated Equivalence Sets	% of ESs
Properties	~ 1.15 M	~ 88%
Classes	~ 288 K	~ 7 %

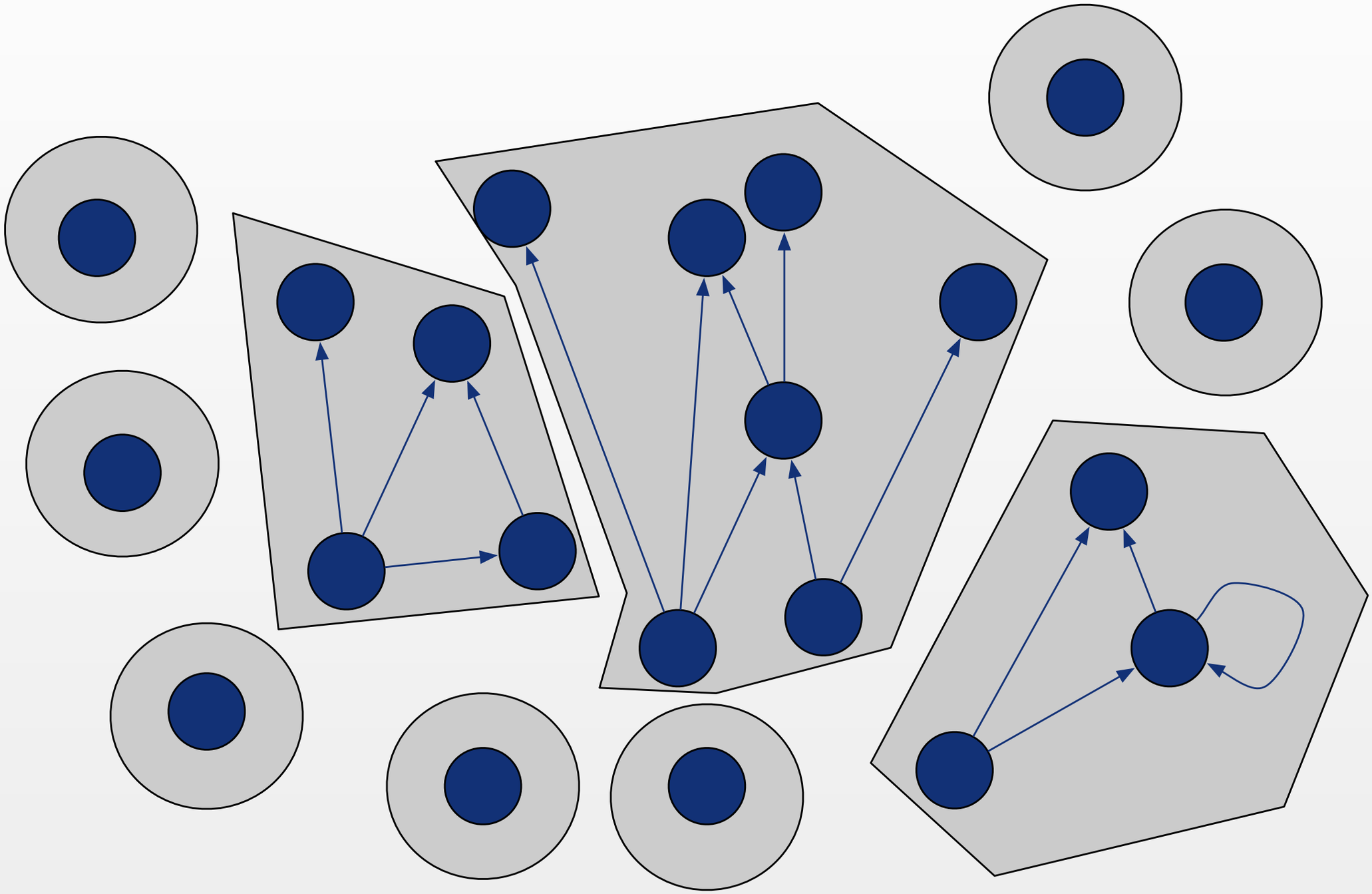
Top Level Equivalence Sets



LOD-a-lot

	# Top Level Equivalence Sets	% of ESs
Properties	~ 1.18 M	~ 90%
Classes	~ 1.28 M	~ 32 %

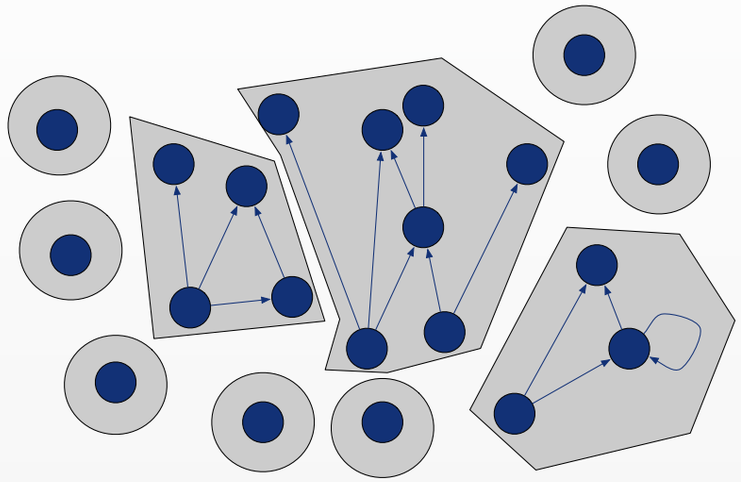
Weakly Connected Components

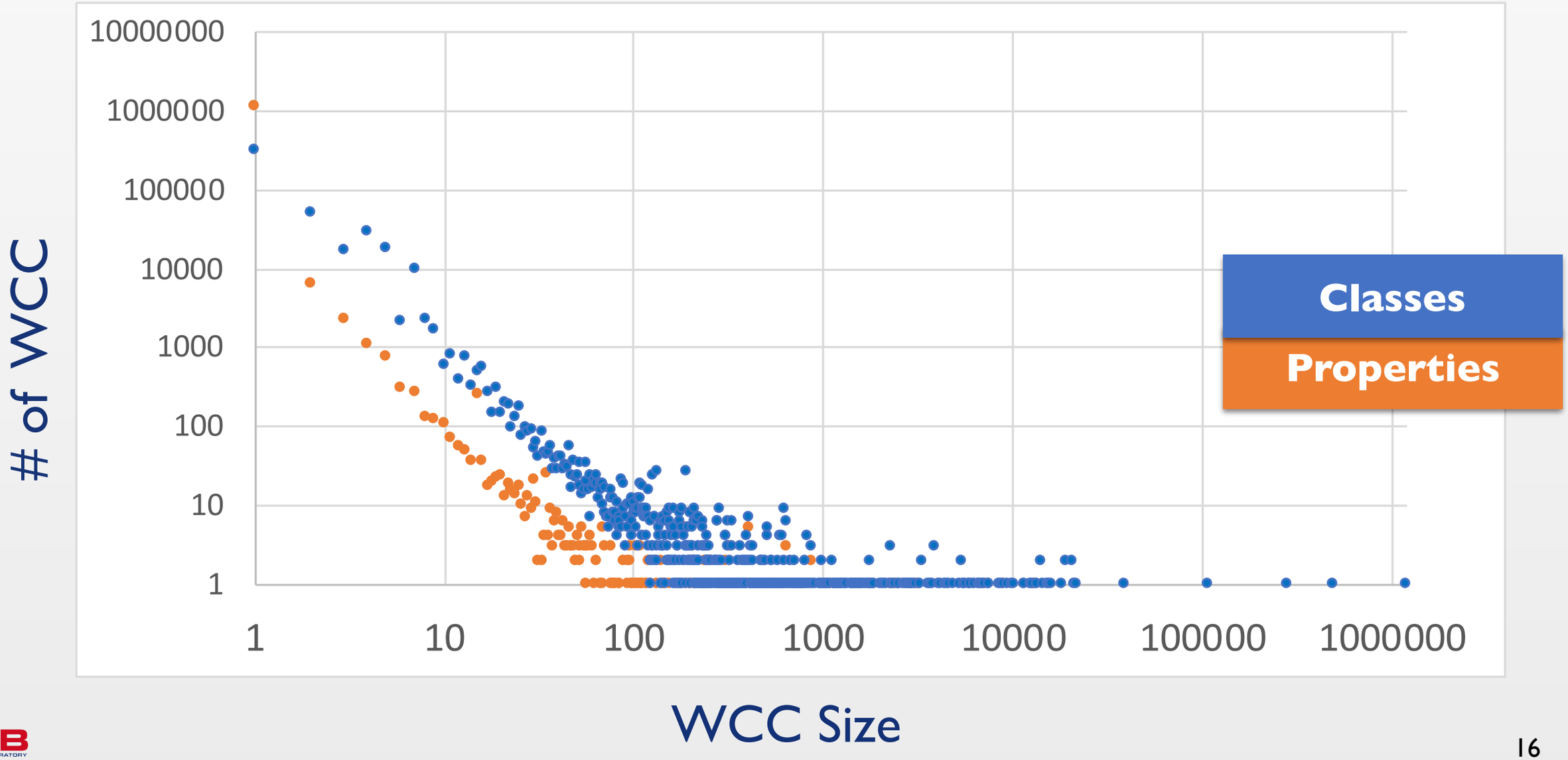


Weakly Connected Components

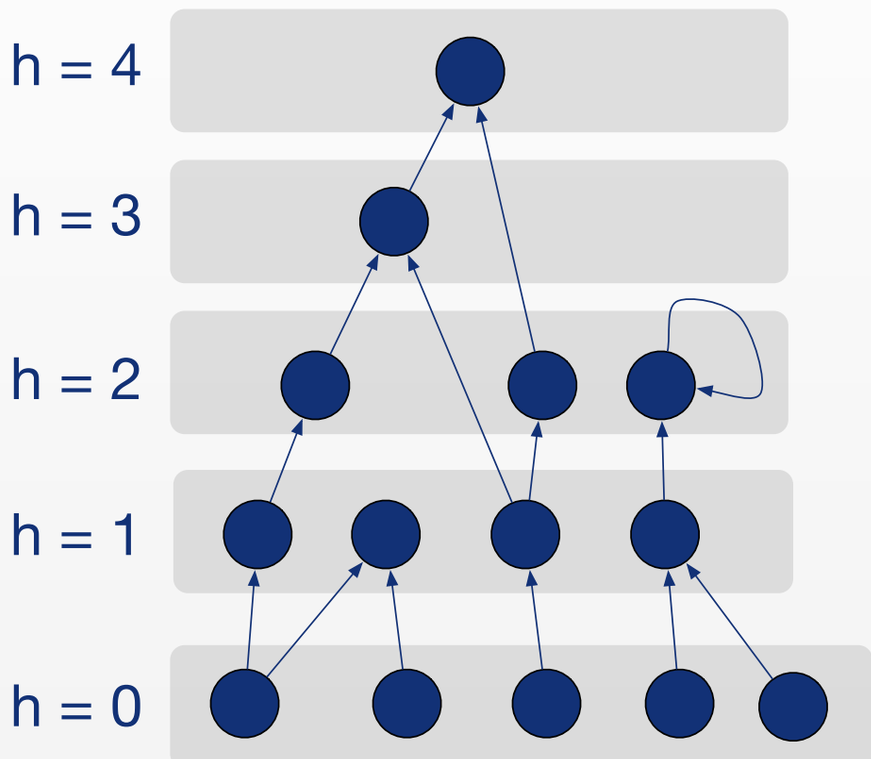
LOD-a-lot

	#WCC
Properties	~ 1.17 M
Classes	~ 449 K



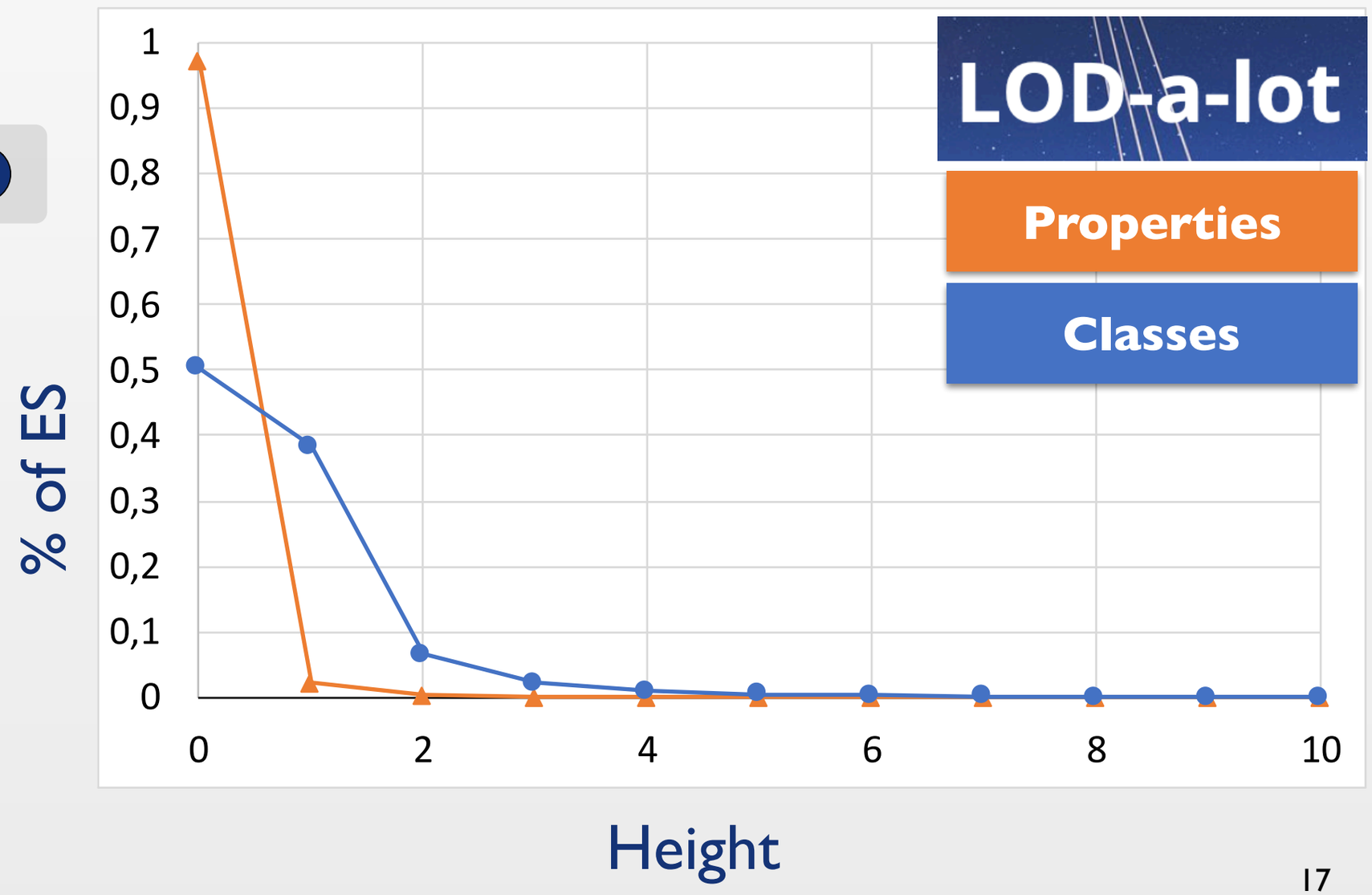


Height of Equivalence Sets

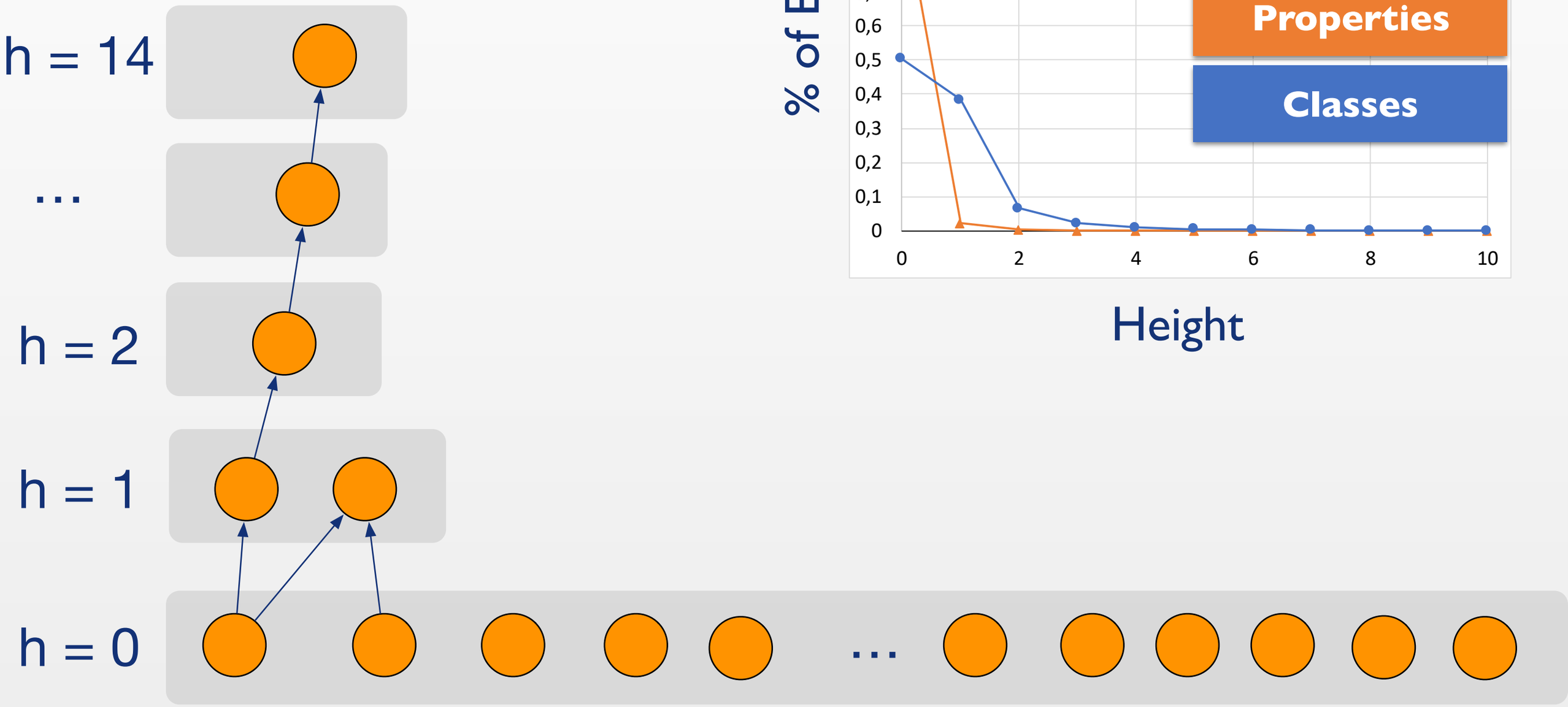


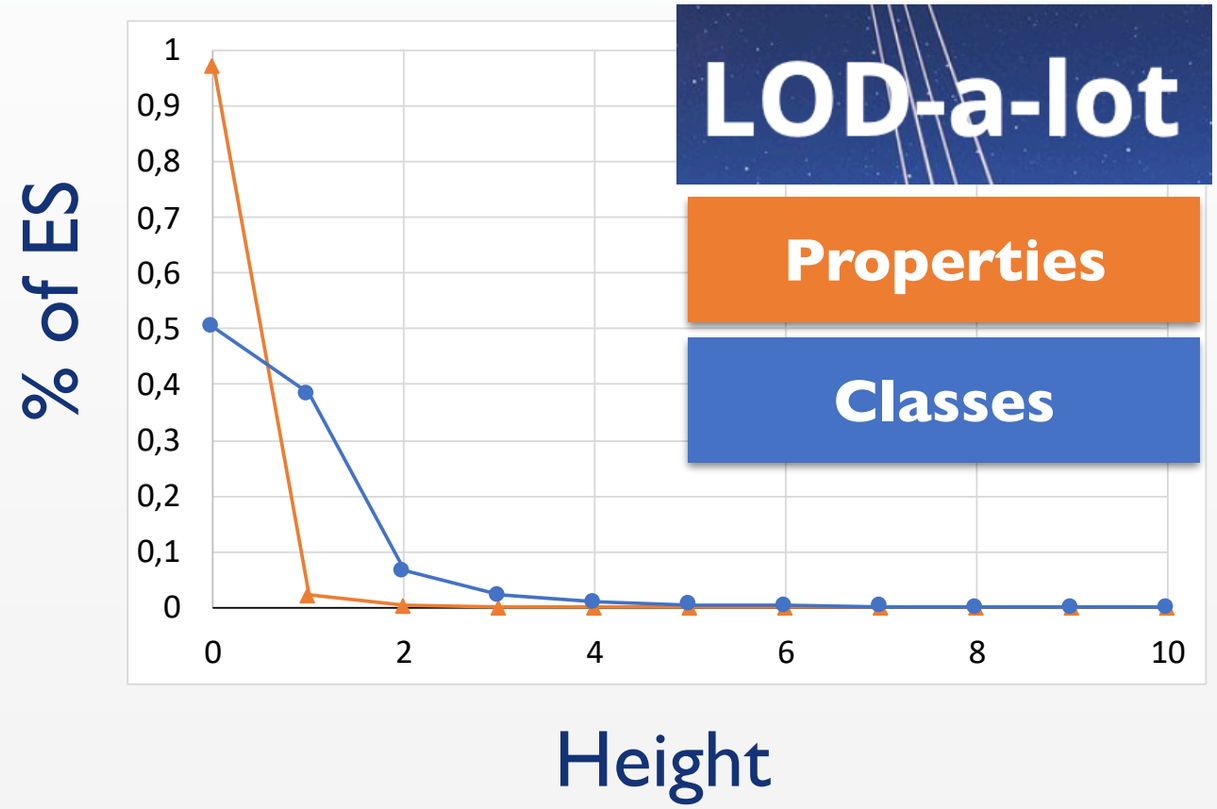
H_{max} =14

H_{max} = 77

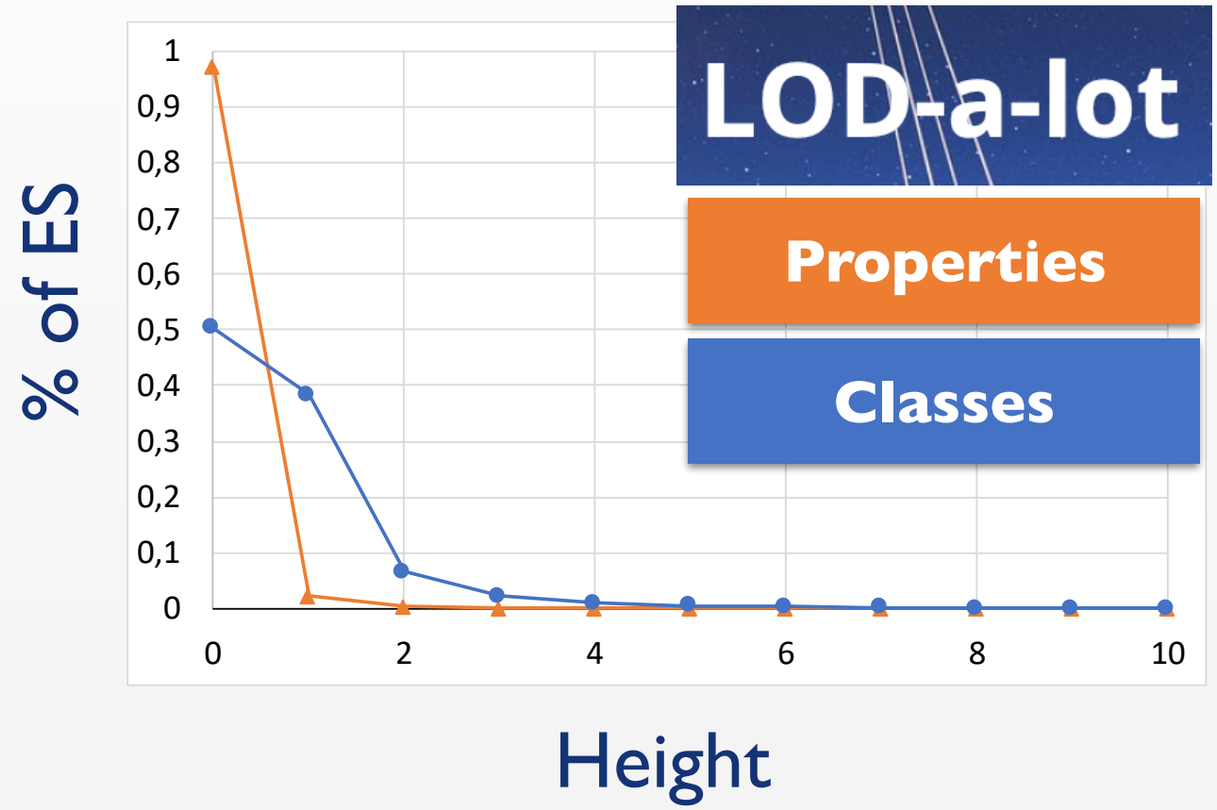


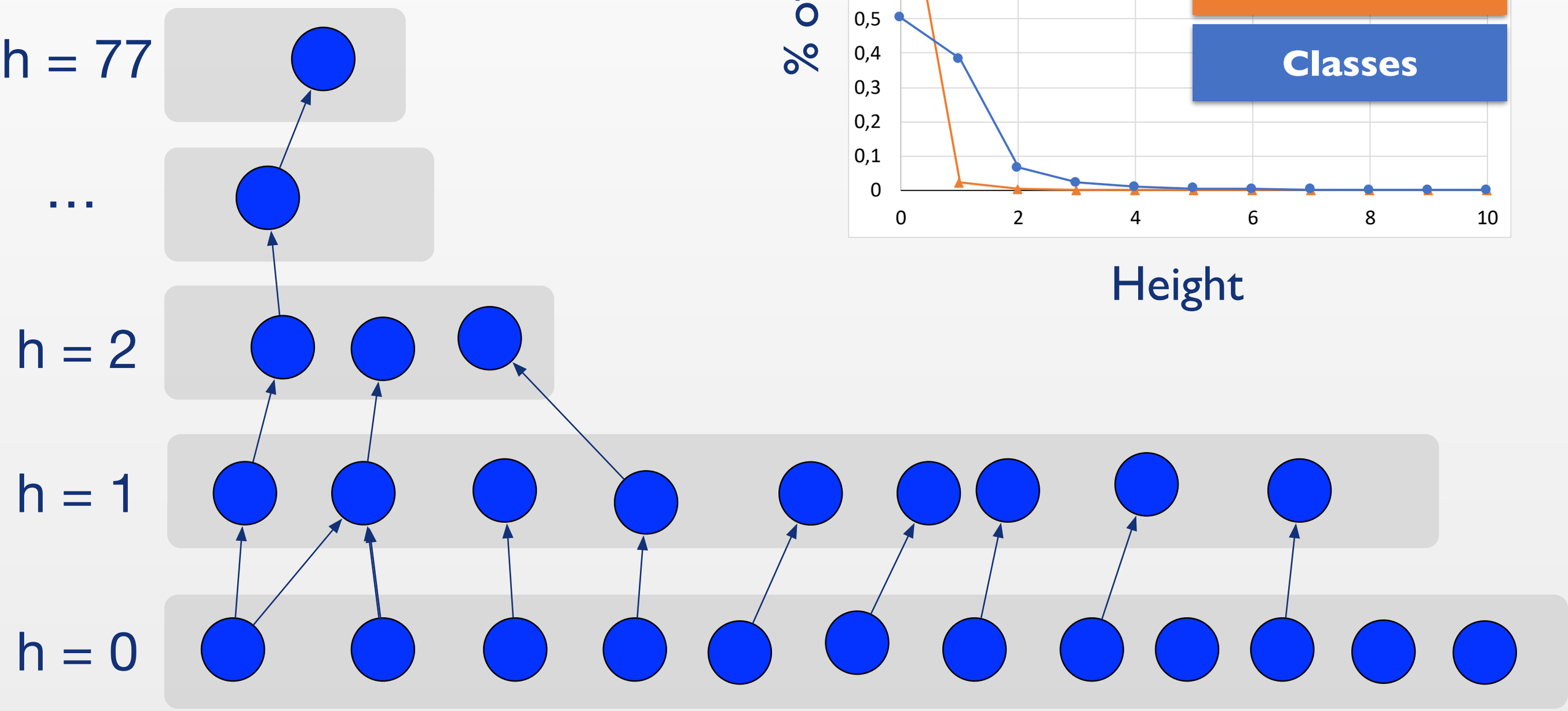
Shape of the property hierarchy



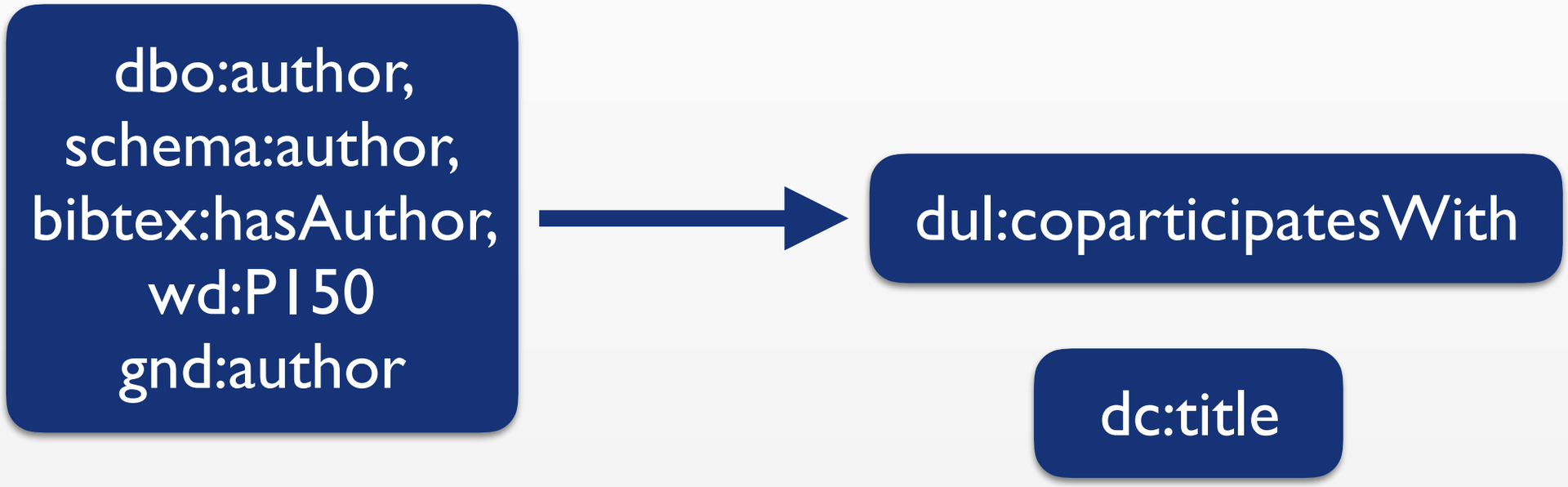


Shape of the class hierarchy

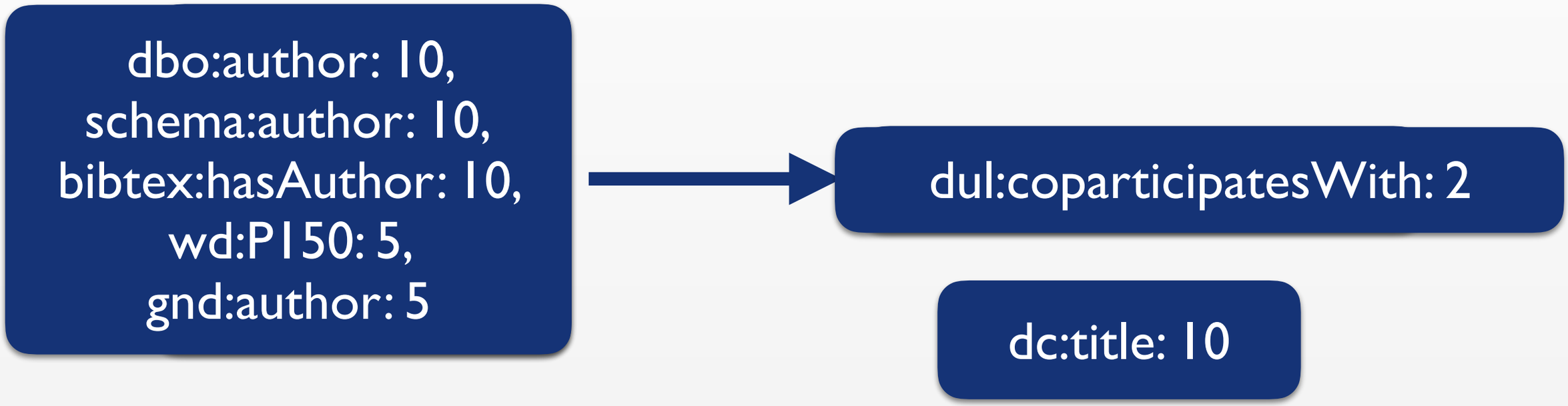




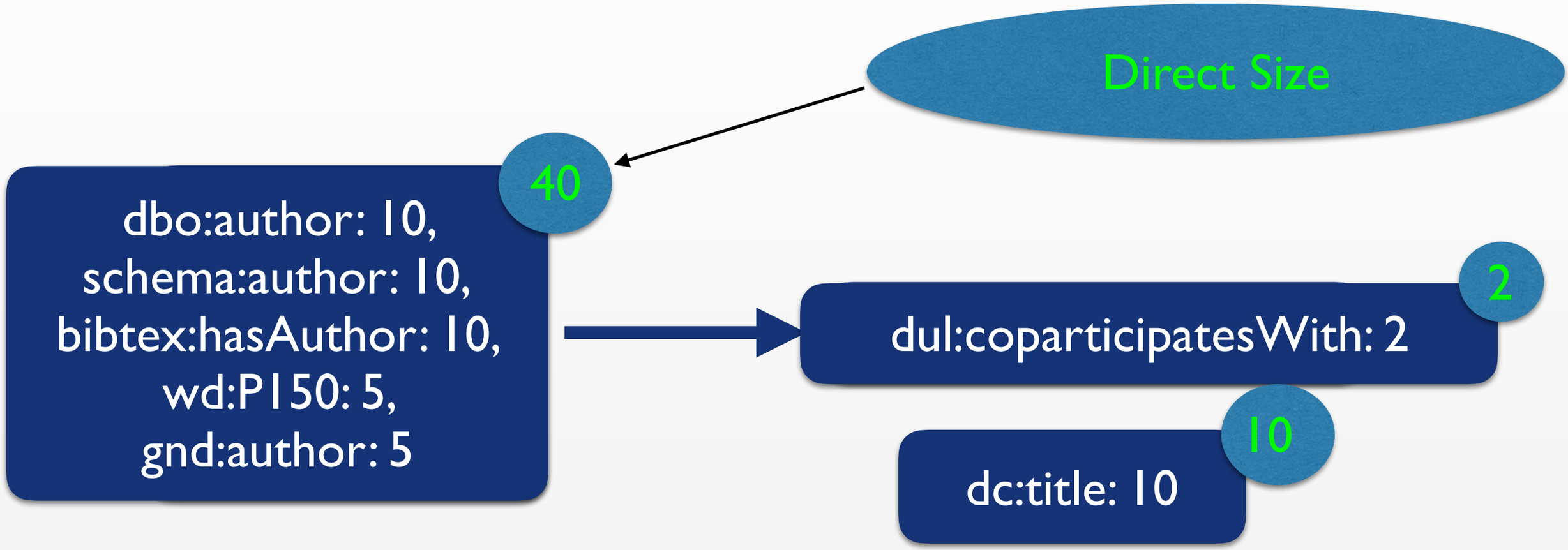
Extensional Size



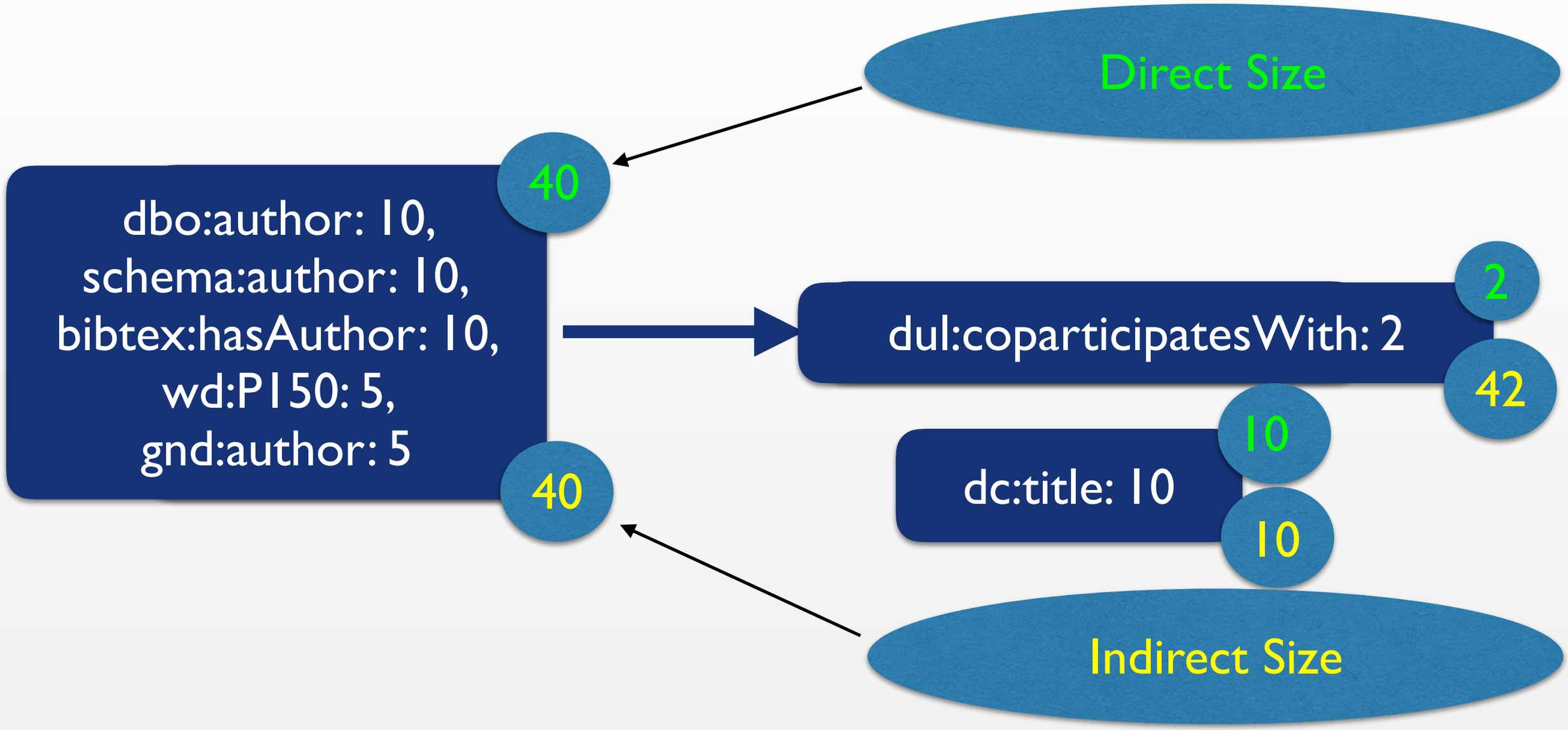
Extensional Size



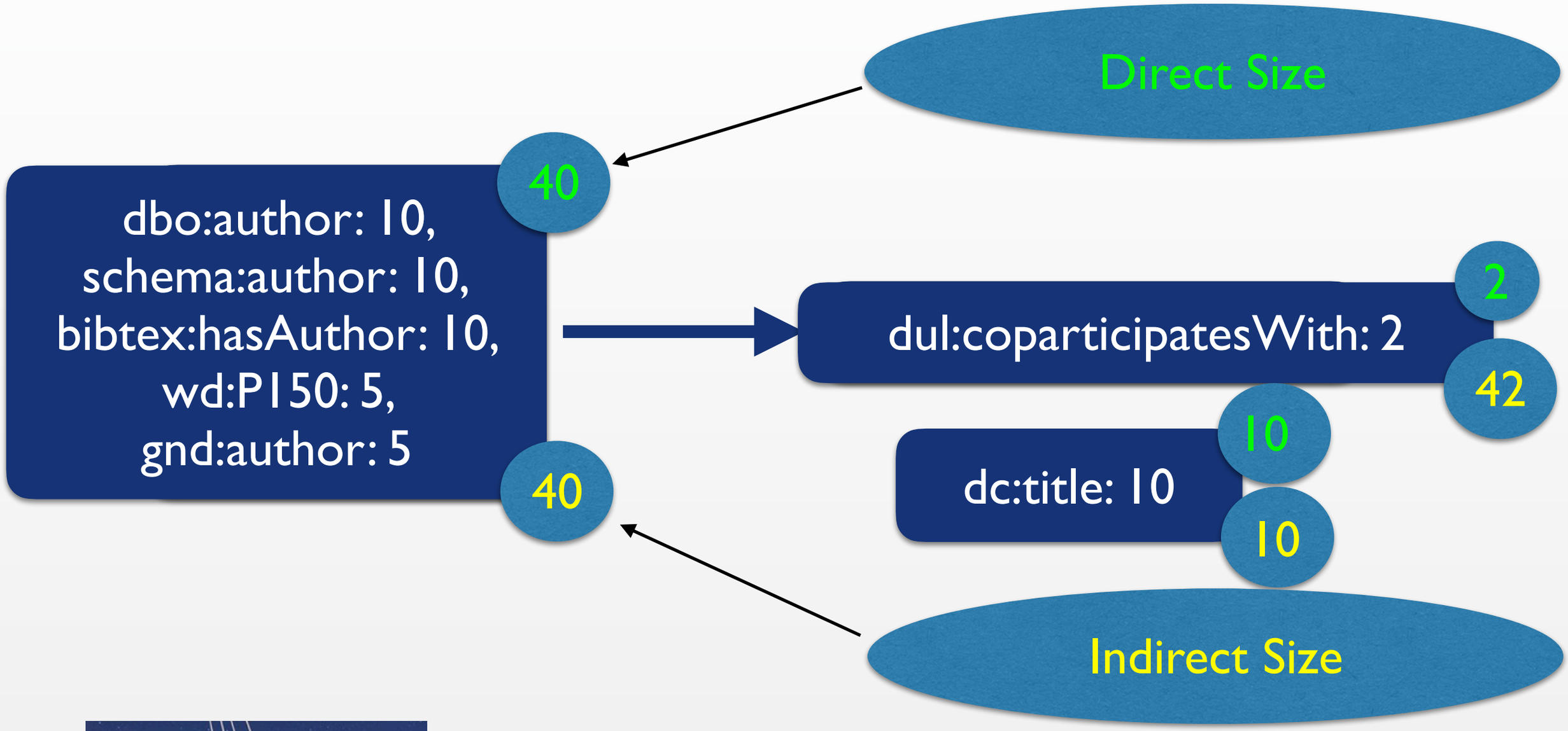
Extensional Size



Extensional Size



Extensional Size



LOD-a-lot	# Empty ESs (Indirect Size)	% of ESs
Properties	~ 132 K	~ 10 %
Classes	~ 3 M	~ 75 %

CONCLUSIONS

A.

B.

C.

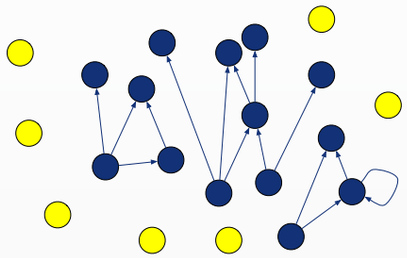
Interlinking of Properties and Classes

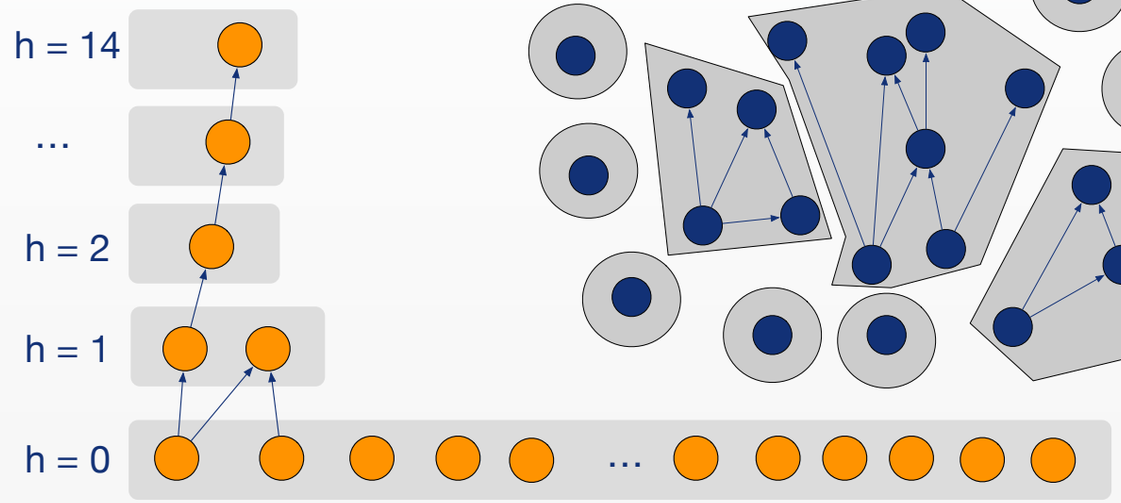
# Properties	~ 1.3 M
# Equivalent Sets	~ 1.3 M
Ratio	.997
Edges	~ 147 M
Isolated ESs	~ 1.15 M
#WCC	~ 1.17 M

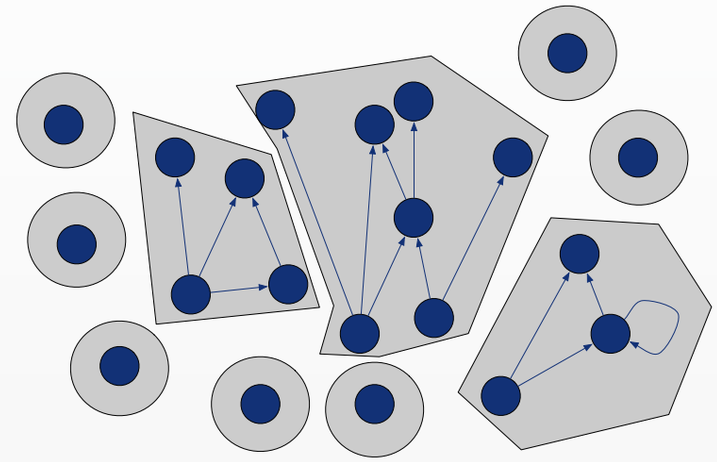
# Classes	~ 4.8 M
# Equivalent Sets	~ 4 M
Ratio	.831
Edges	~ 5 M
Isolated ESs	~ 288K
#WCC	~ 449 K

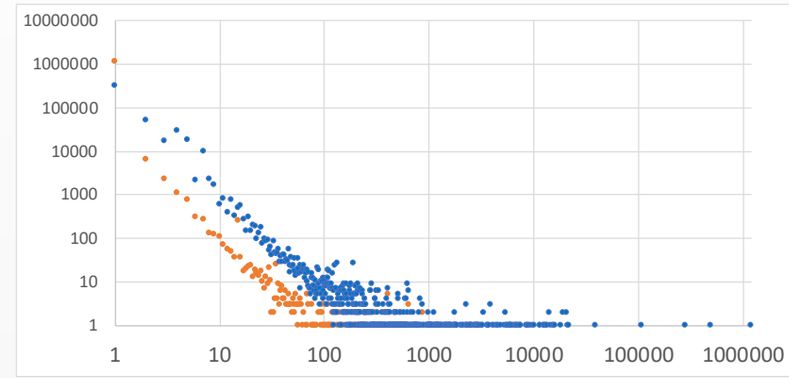
Equivalence and Specialization are sparsely used

Shape of the LOD hierarchies

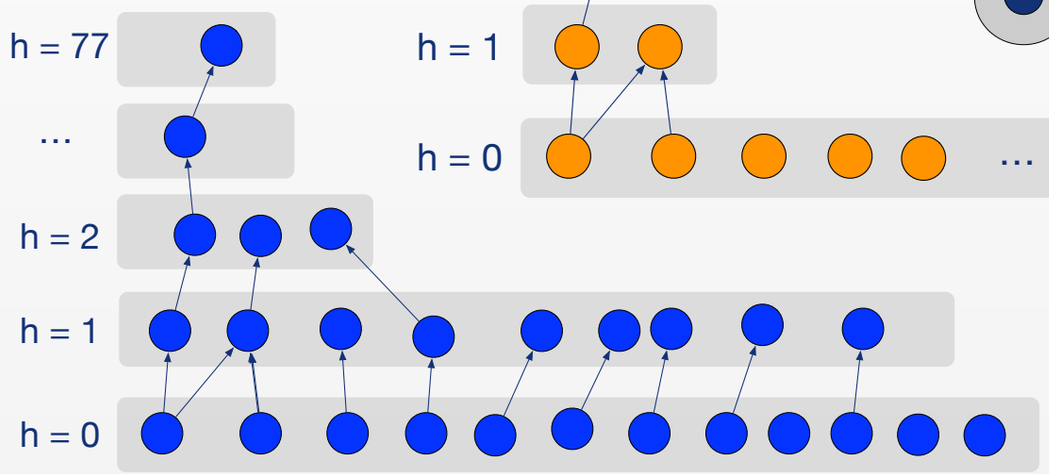


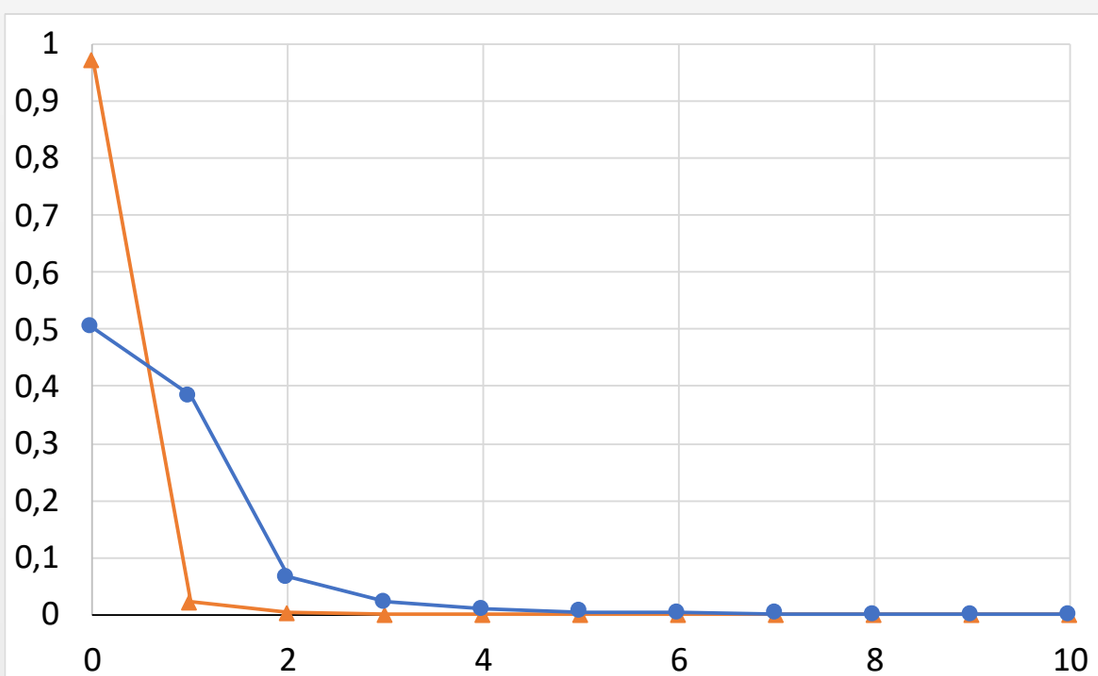






WCC





Height

Hierarchies are mostly flat

Instantiation of classes and properties

	# Empty ESs (Indirect Size)	% of ESs
Properties	~ 132 K	~ 10 %
Classes	~ 3 M	~ 75 %

Most of properties are instantiated

Many classes are unistantiated

Future Work

thank you
danke
謝謝
ngiyabonga
شكراً جزيلاً
teşekkür ederim
gracias
tapadh leat
спасибо
Баярлалаа
рахмат
faafetai lava
kiitos
dankie
dhanyavad
nandri
nanni
bedankt
hvala
mauruuru
köszönöm
gràcie
bayarlalaa
enkosi
dziękuję
obrigado
sobodi
dēkuji
mèsi
didi madloba
sagolun
kam sah hamnida
তোমাকে ধন্যবাদ
rahmat
najis tuke
terima kasih
감사합니다
xièxie
ありがとう
sukriya
chnorakaloutioun
gratias ago
gràcies
sulpáy
kop khun krap
taiku
tanemirt
rahmet
grazie
ευχαριστώ
diolch
merci
go raibh maith agat
arigatô
takk
dakujem
trugarez
dhanyavadagalu
shukriya
merçé
мерси
mercè
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chokrane
murakoze
tenki
obrigada
asante
manana
хвала
mahalo
paldies
grazzi
matondo
misaotra
dank je
akun
dankon
aciŭ
spas
welalin
tack
barka
kia ora
mersi
vinaka
спасибі
blagodaram

References

(Asprino 2018) L. Asprino, V. Basile, P. Ciancarini, and V. Presutti. "Empirical Analysis of Foundational Distinctions in Linked Open Data" @ IJCAI 2018

(Beek 2018) W. Beek, J. Raad, J. Wielemaker, and F. van Harmelen. "sameAs.cc: The Closure of 500M owl: sameAs Statements" @ ESWC 2018

(DeBattista 2018) Debattista, J., Lange, C., Auer, S., & Cortis, D. (2018). Evaluating the quality of the LOD cloud: An empirical investigation. Semantic Web, (Preprint), 1-43.

(Ding 2006) L. Ding and T. Finin. "Characterizing the Semantic Web on the Web" @ ISWC 2006

(Ding 2010) L. Ding, J. Shinavier, Z. Shangguan, and D. McGuinness. "SameAs Networks and Beyond: Analyzing Deployment Status and Implications of owl: sameAs in Linked Data" @ ISWC 2010

(Fernández 2017) J. Fernandez, W. Beek, M. Martinez-Prieto, and M. Arias. "LOD-a-lot - A Queryable Dump of the LOD Cloud" @ ISWC 2017

(Gil 2004) R. Gil, R. Garc´ıa, and J. Delgado. "Measuring the semantic web" @AIS SIGSEMIS

(Halpin 2010) H. Halpin, P. Hayes, J. McCusker, D. McGuinness, and H. Thompson. "When owl:sameAs Isn't the Same: An Analysis of Identity in Linked Data" @ ISWC 2010

(Mallea 2011) A. Mallea, M. Arenas, A. Hogan, and A. Polleres. "On Blank Nodes" @ ISWC 2011

(Nuzzolese 2011) AG Nuzzolese, A Gangemi, V Presutti, P Ciancarini. Encyclopedic knowledge patterns from wikipedia links @ ISWC 2011

(Paulheim 2015) H. Paulheim and A. Gangemi. "Serving DBpedia with DOLCE". @ ISWC 2015

(Schmachtenberg 2014) Schmachtenberg, M., Bizer, C., & Paulheim, H. (2014, October). Adoption of the linked data best practices in different topical domains. @ ISWC 2014

(Theoharis 2008) Y. Theoharis, Y. Tzitzikas, D. Kotzinos, and V. Christophides. "On Graph Features of Semantic Web Schemas" In TKDE

(Wang 2006) T. Wang, B. Parsia, and J. Hendler. "A Survey of the Web Ontology Landscape" @ISWC 2006