

Diversified Stress Testing of RDF Data Management Systems

Güneş Aluç

Olaf Hartig

M. Tamer Özsu

Khuzaima Daudjee



UNIVERSITY OF
WATERLOO

This presentation is sponsored in part by the *Semantic Web Science Association (SWSA)*.

Questions

- Are popular SPARQL benchmarks really suitable for **stress-testing** RDF data management systems?
- Are popular RDF data management systems really as **flawless** as they appear on existing benchmarks?

Contributions

Waterloo SPARQL Diversity Test Suite
(WatDiv)



<http://db.uwaterloo.ca/watdiv/>

Contributions

Waterloo SPARQL Diversity Test Suite
(WatDiv)



<http://db.uwaterloo.ca/watdiv/>

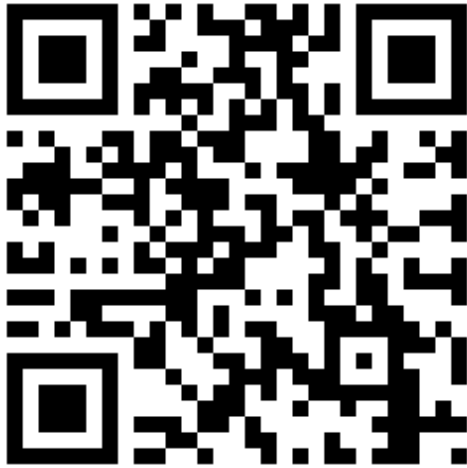
Measures to Evaluate Diversity
in SPARQL Workloads

Structural

Data-driven

Contributions

Waterloo SPARQL Diversity Test Suite
(WatDiv)



<http://db.uwaterloo.ca/watdiv/>

Measures to Evaluate Diversity
in SPARQL Workloads

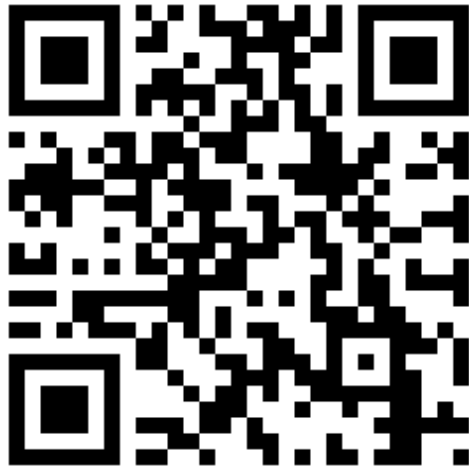
Structural

Data-driven

Analysis of WatDiv and Popular
SPARQL Benchmarks

Contributions

Waterloo SPARQL Diversity Test Suite
(WatDiv)



<http://db.uwaterloo.ca/watdiv/>

Measures to Evaluate Diversity
in SPARQL Workloads

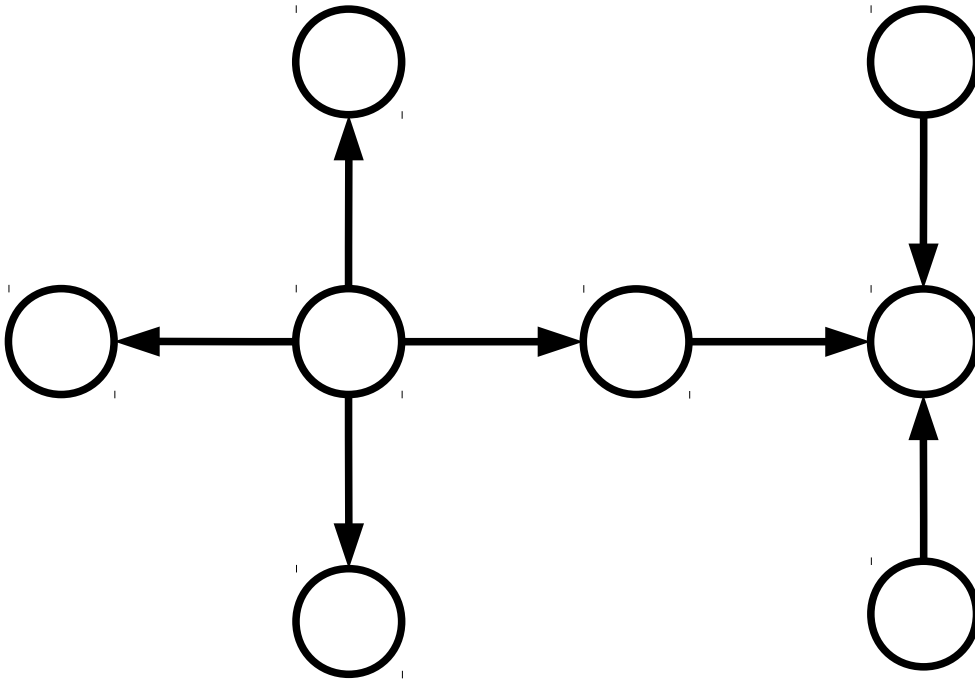
Structural

Data-driven

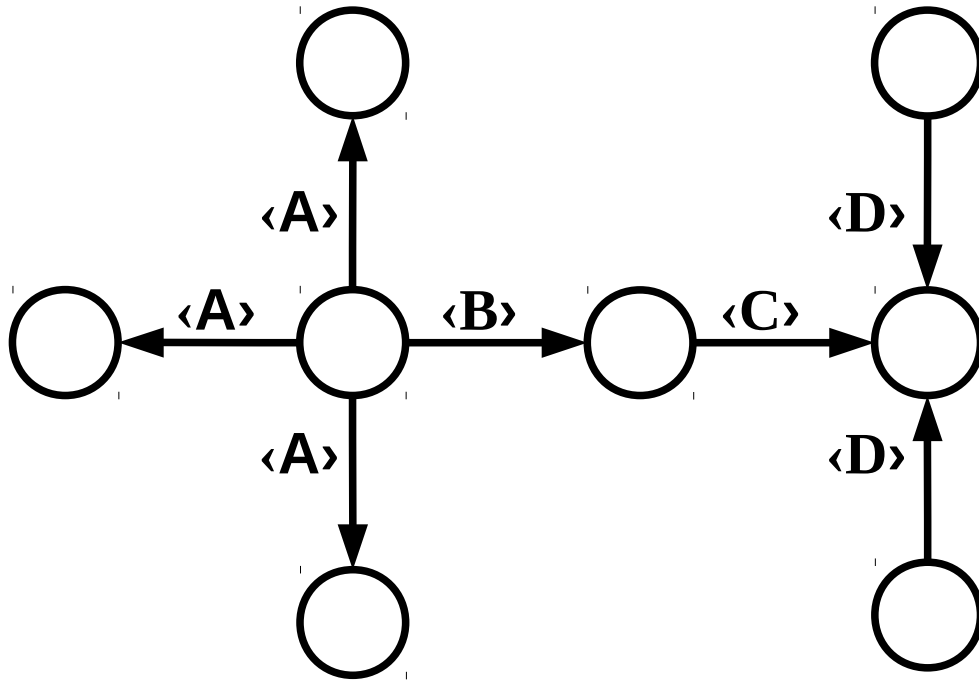
Analysis of WatDiv and Popular
SPARQL Benchmarks

Evaluation of Popular RDF
Data Management Systems
using WatDiv

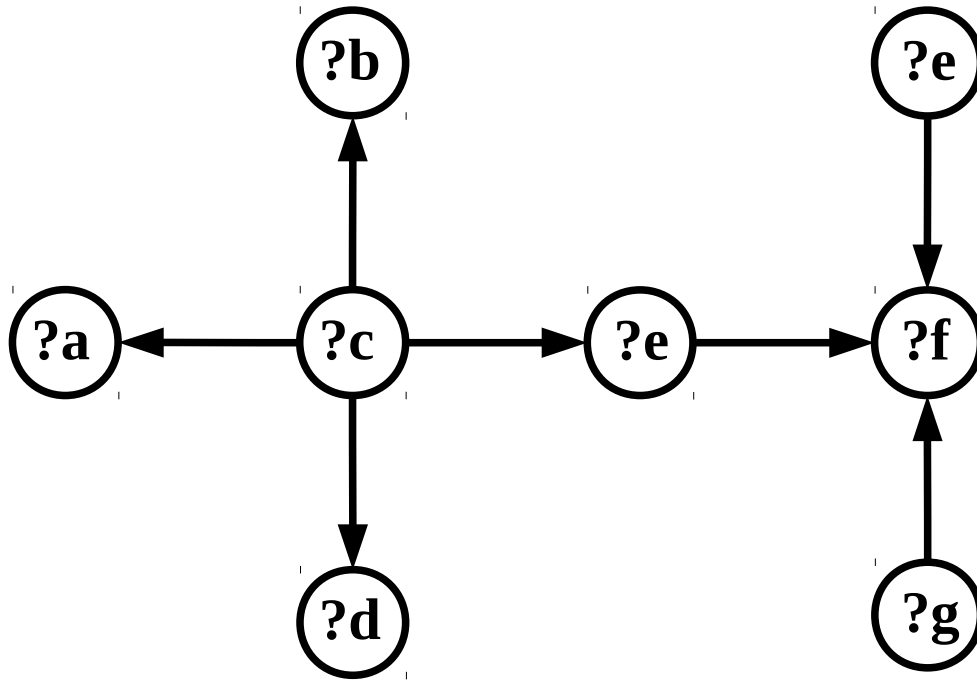
Structural Features



Structural Features

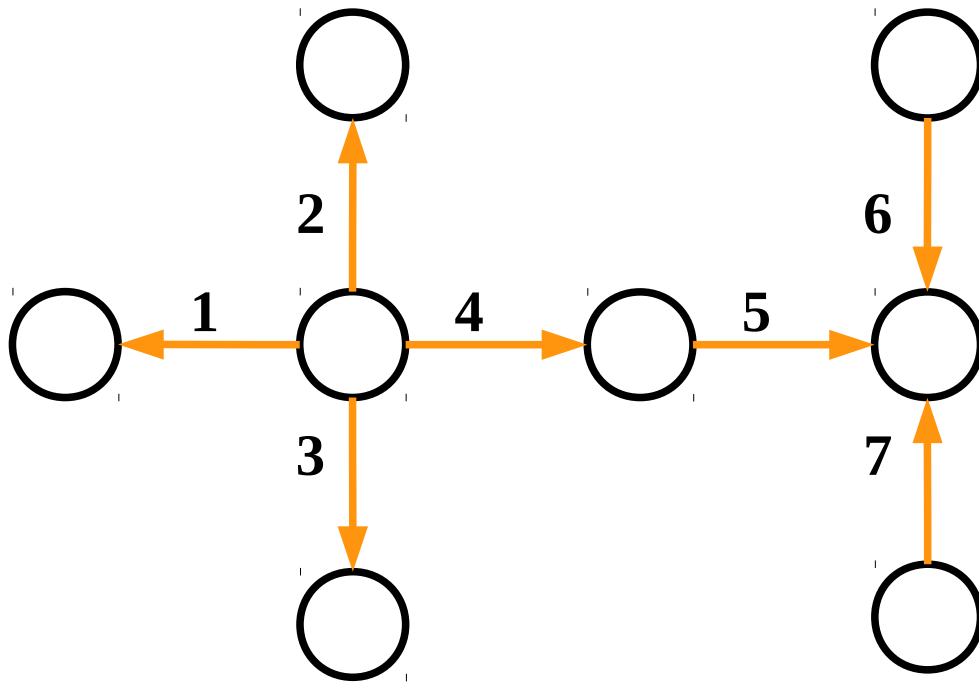


Structural Features



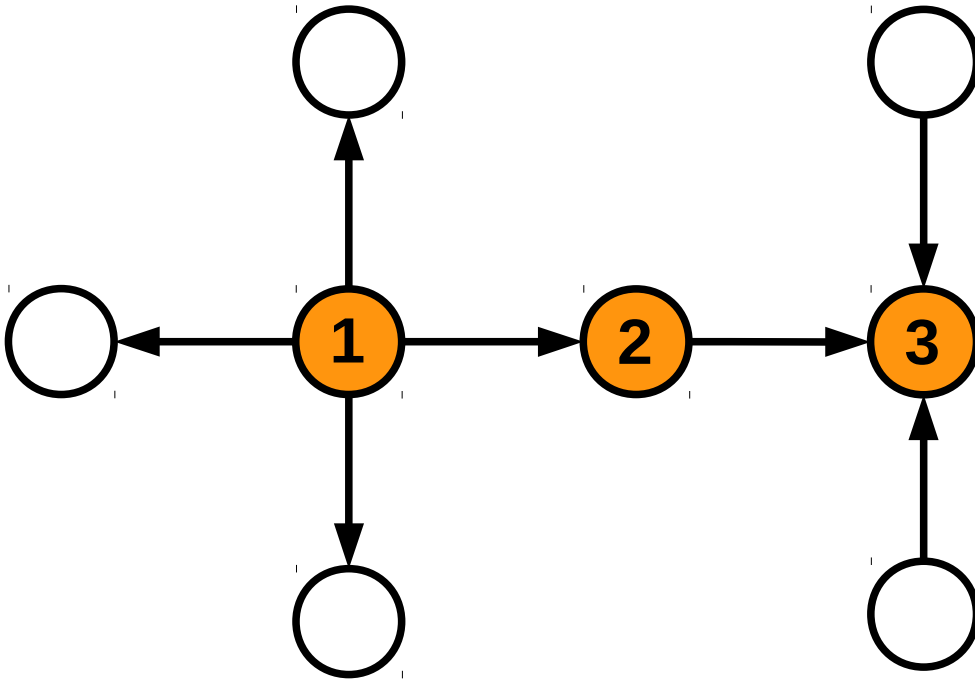
Structural Features

[Triple Pattern Count]



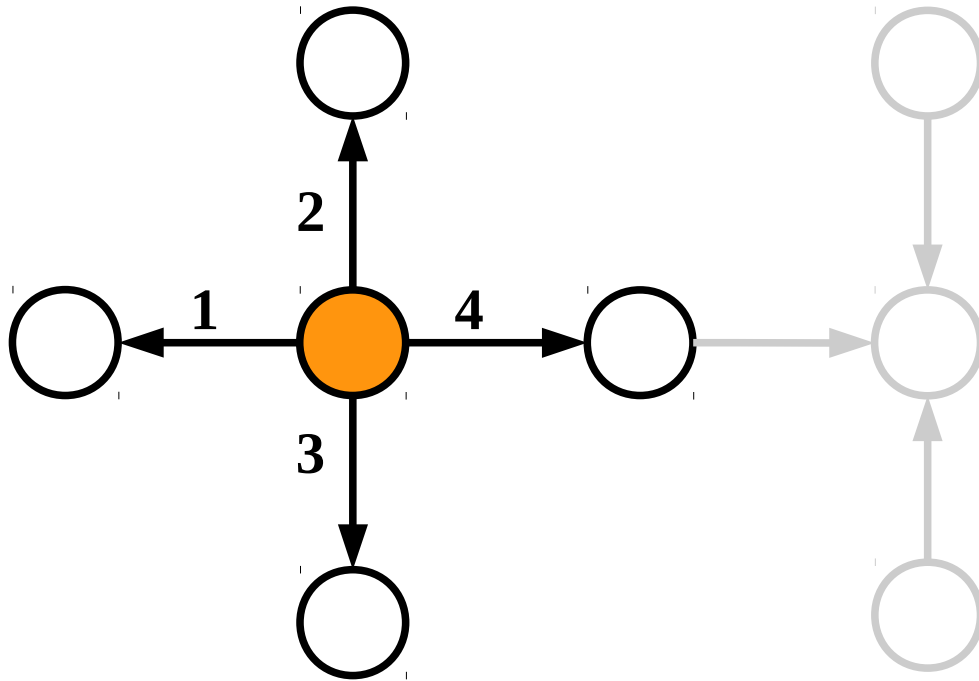
Structural Features

[Join Vertex Count]



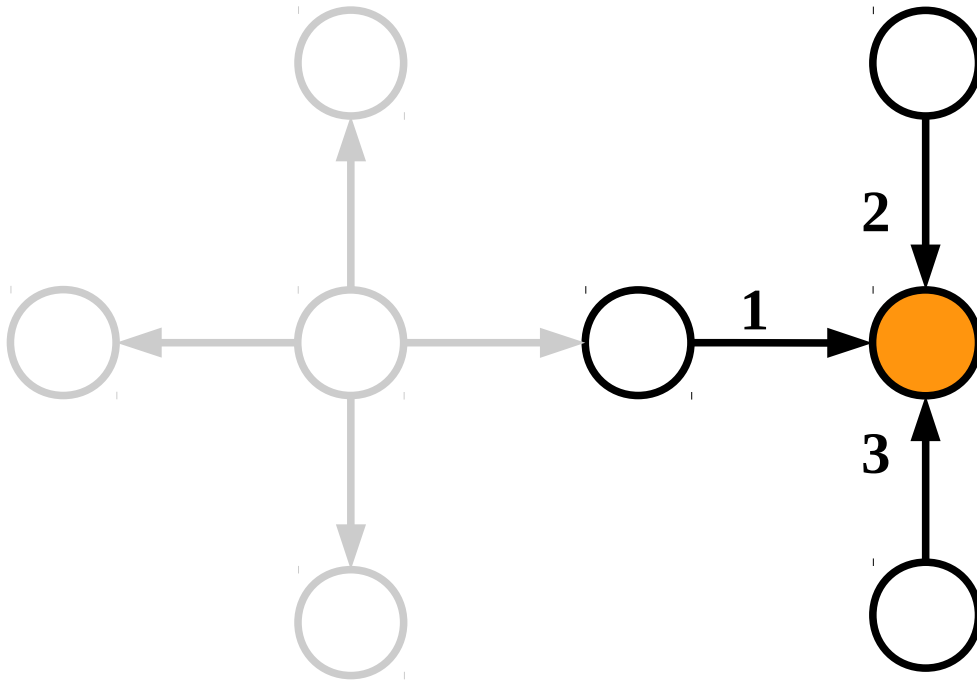
Structural Features

[Join Vertex Degree]



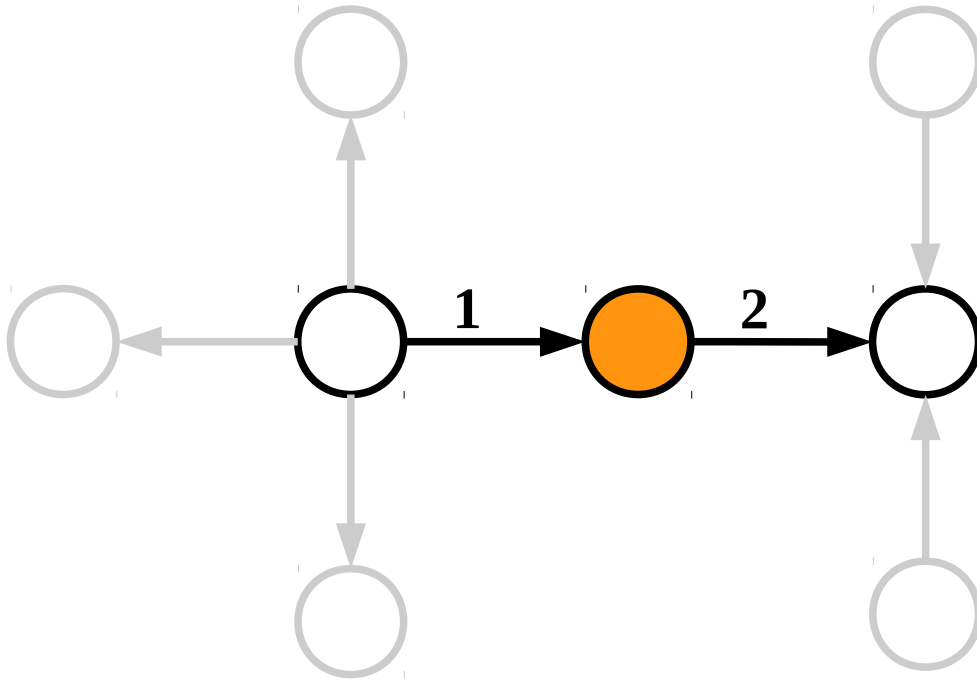
Structural Features

[Join Vertex Degree]

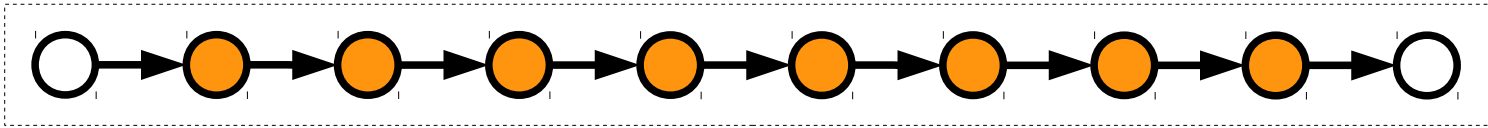


Structural Features

[Join Vertex Degree]

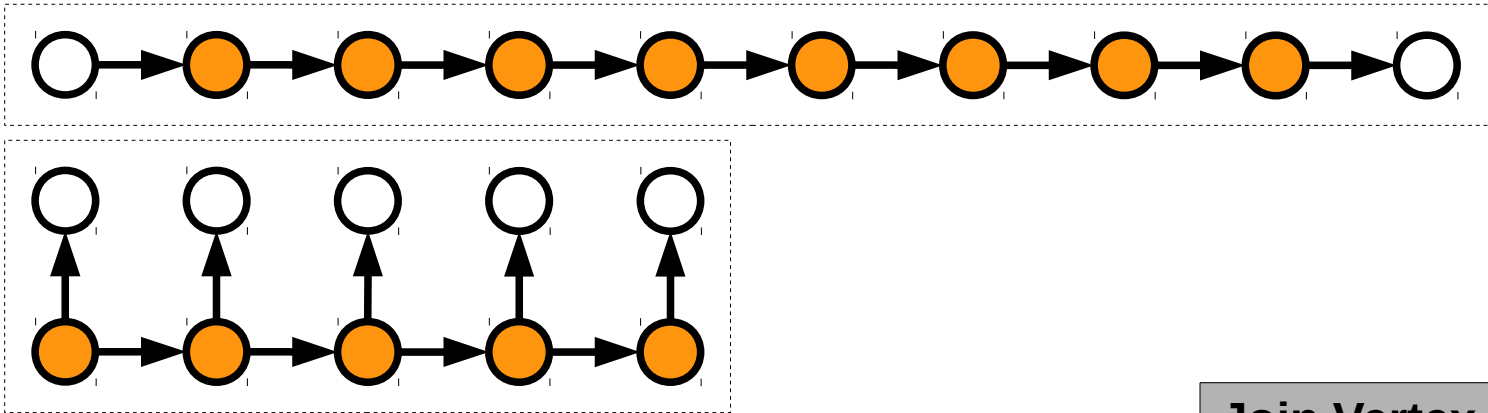


Structural Features



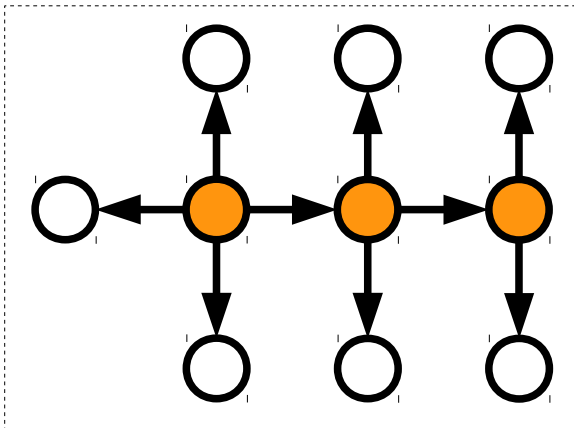
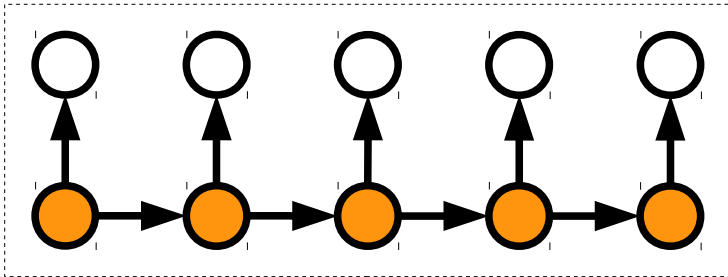
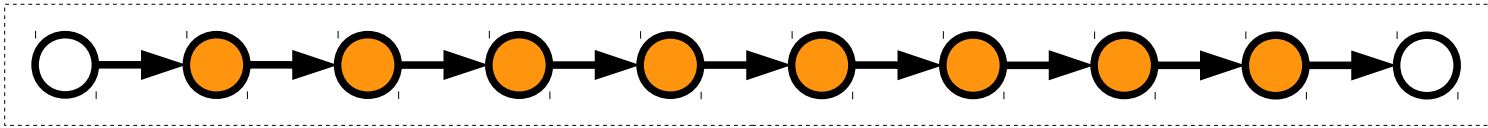
Join Vertex Count	Mean Join Vertex Degree
8	2.0

Structural Features



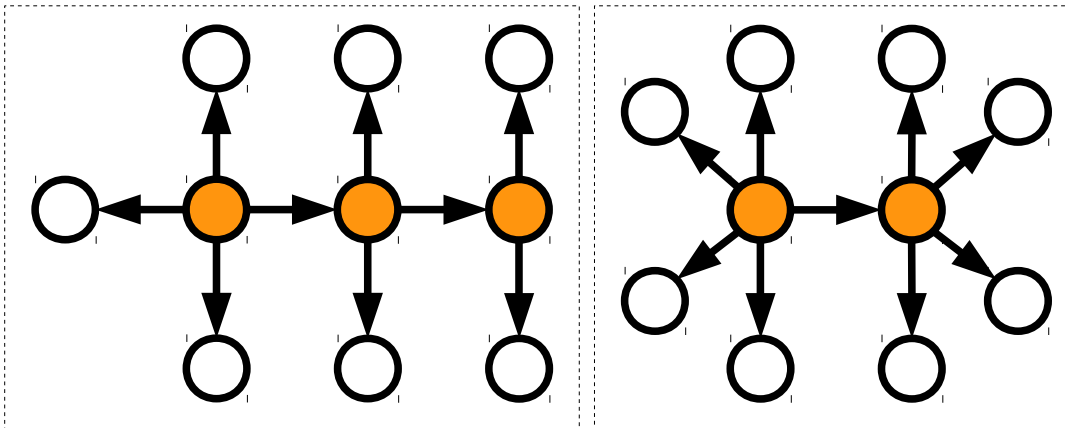
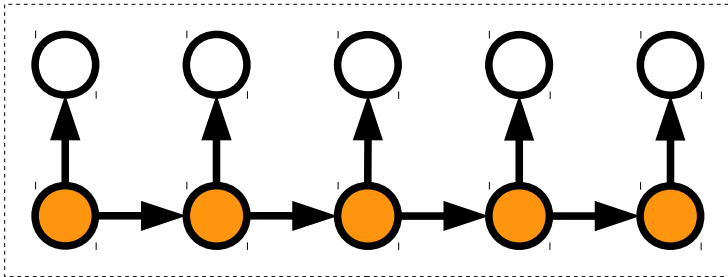
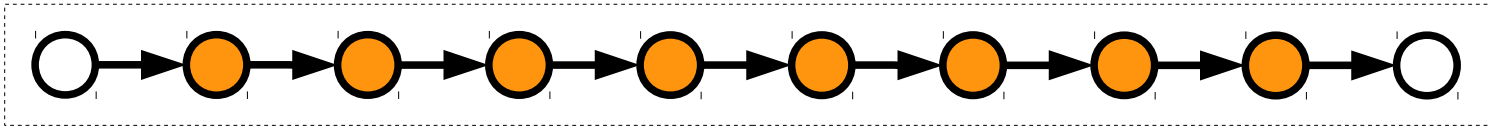
Join Vertex Count	Mean Join Vertex Degree
8	2.0
5	2.6

Structural Features



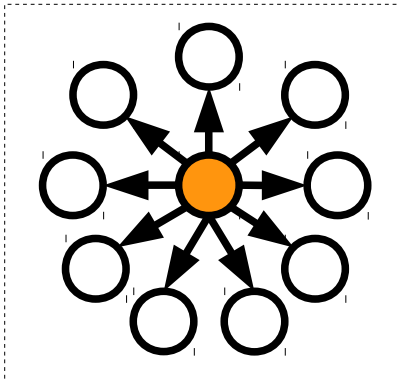
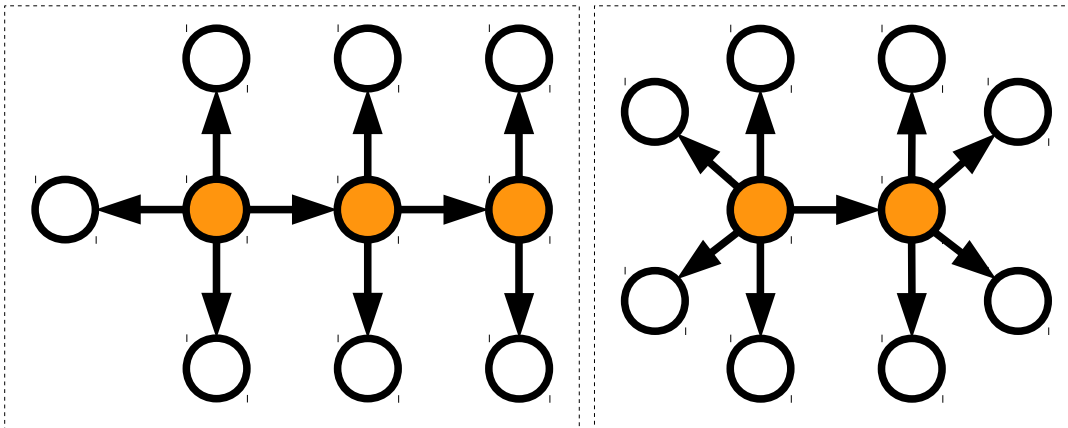
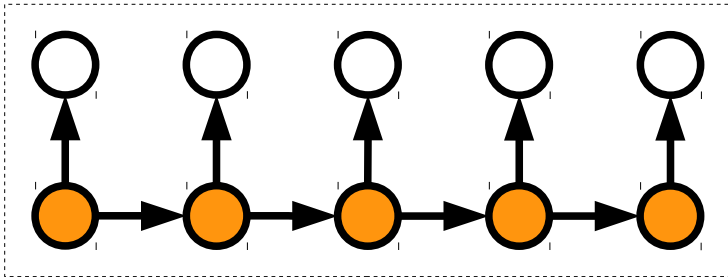
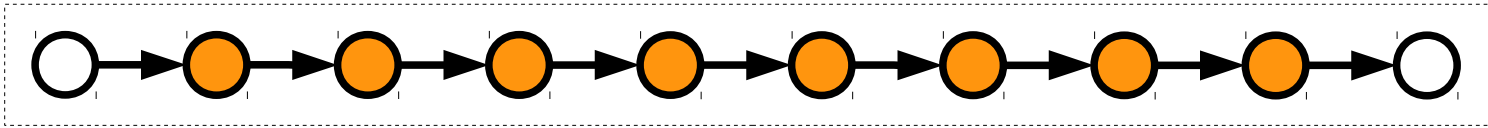
Join Vertex Count	Mean Join Vertex Degree
8	2.0
5	2.6
3	~3.7

Structural Features



Join Vertex Count	Mean Join Vertex Degree
8	2.0
5	2.6
3	~3.7
2	5.0

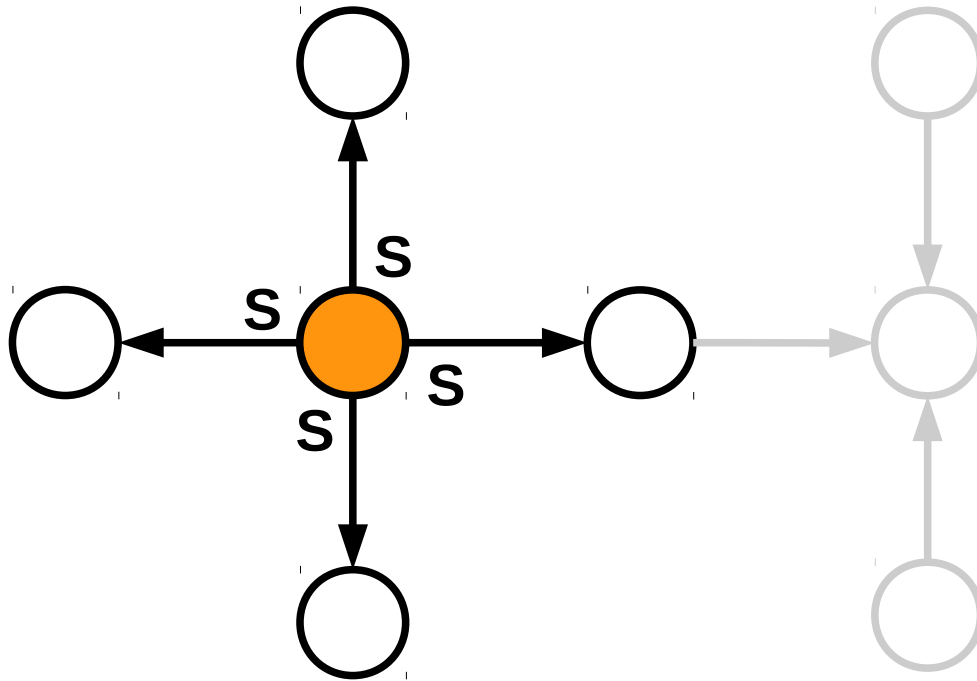
Structural Features



Join Vertex Count	Mean Join Vertex Degree
8	2.0
5	2.6
3	~3.7
2	5.0
1	9.0

Structural Features

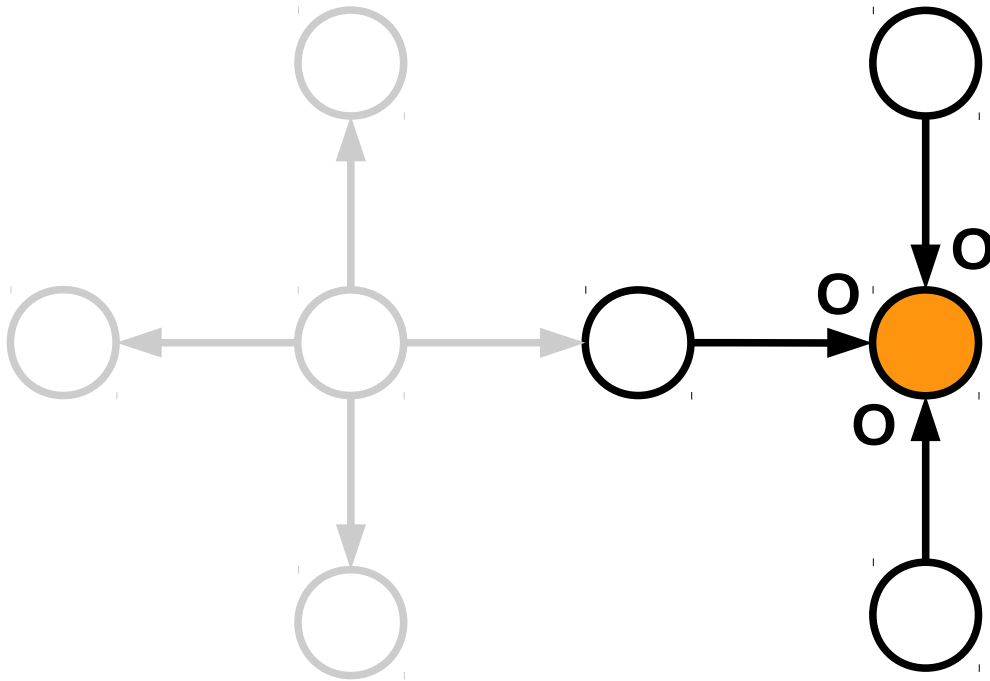
[Join Vertex Type]



SS⁺ Type

Structural Features

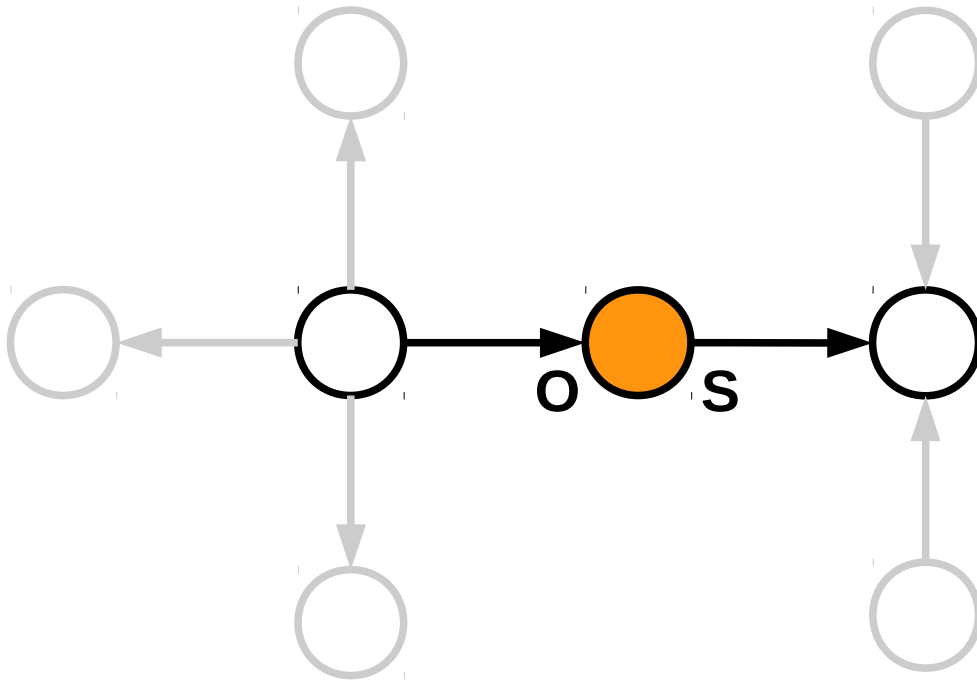
[Join Vertex Type]



OO⁺ Type

Structural Features

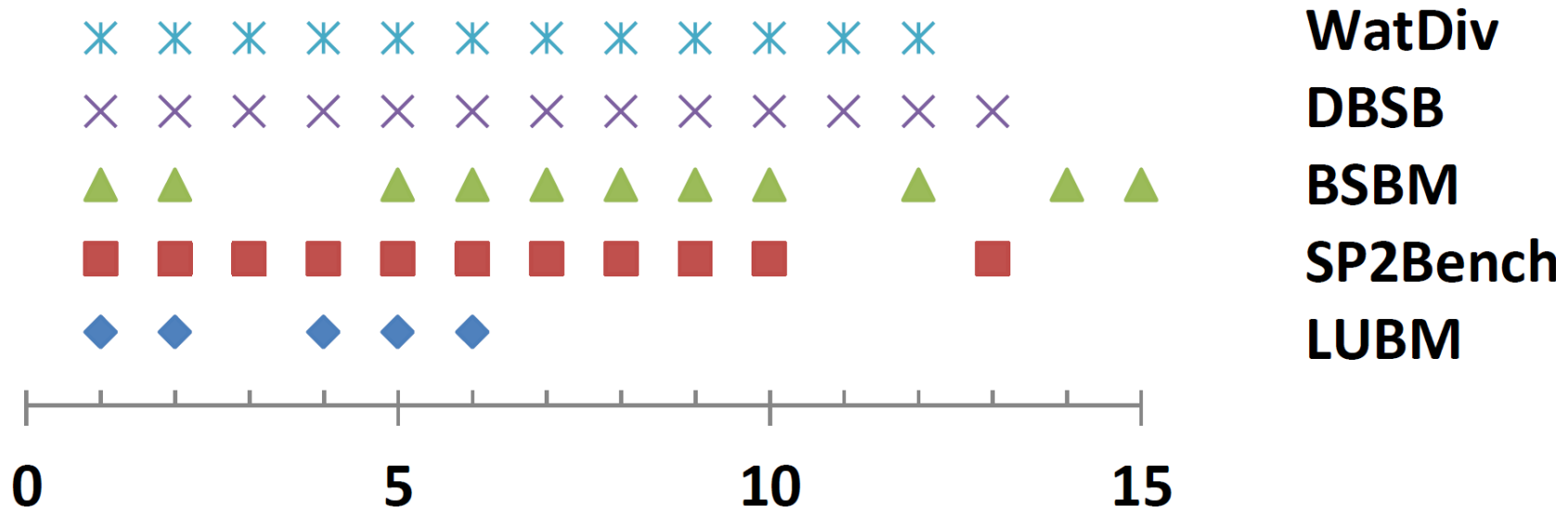
[Join Vertex Type]



SO⁺ Type

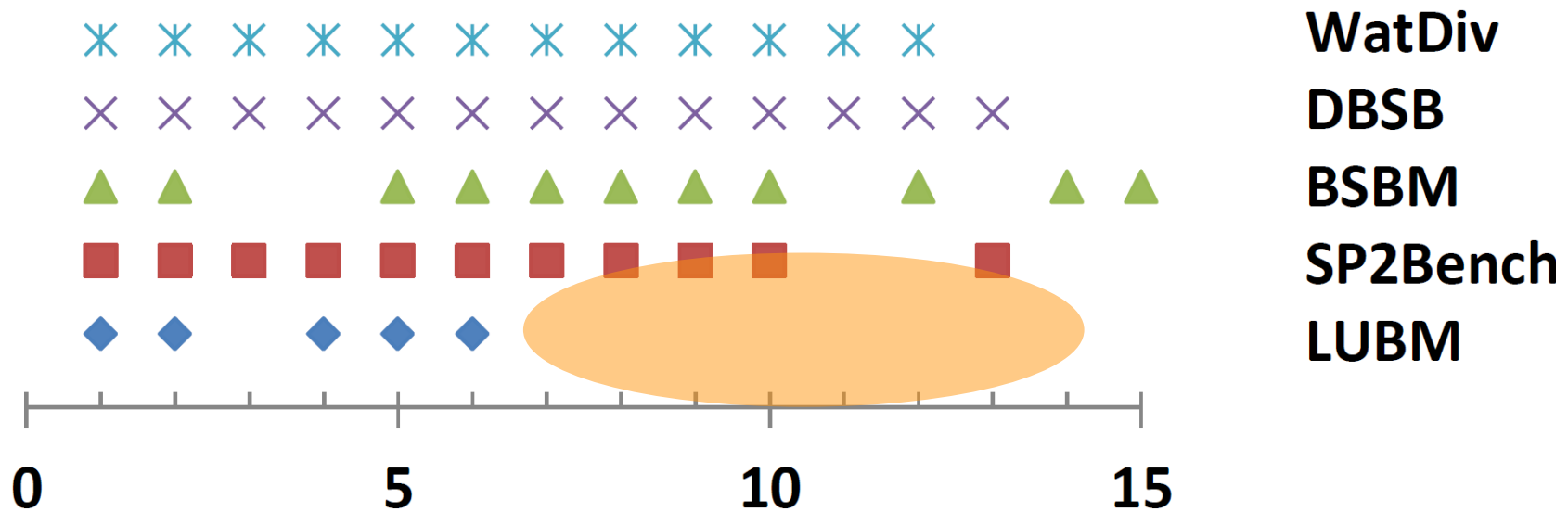
How Diverse are SPARQL Benchmarks?

[Triple Pattern Count]



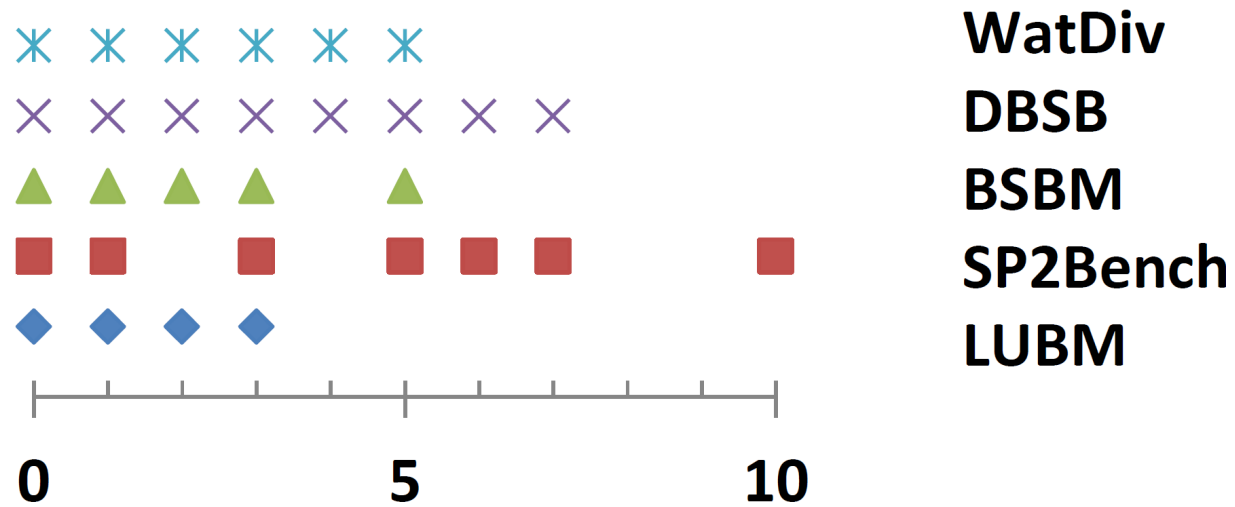
How Diverse are SPARQL Benchmarks?

[Triple Pattern Count]



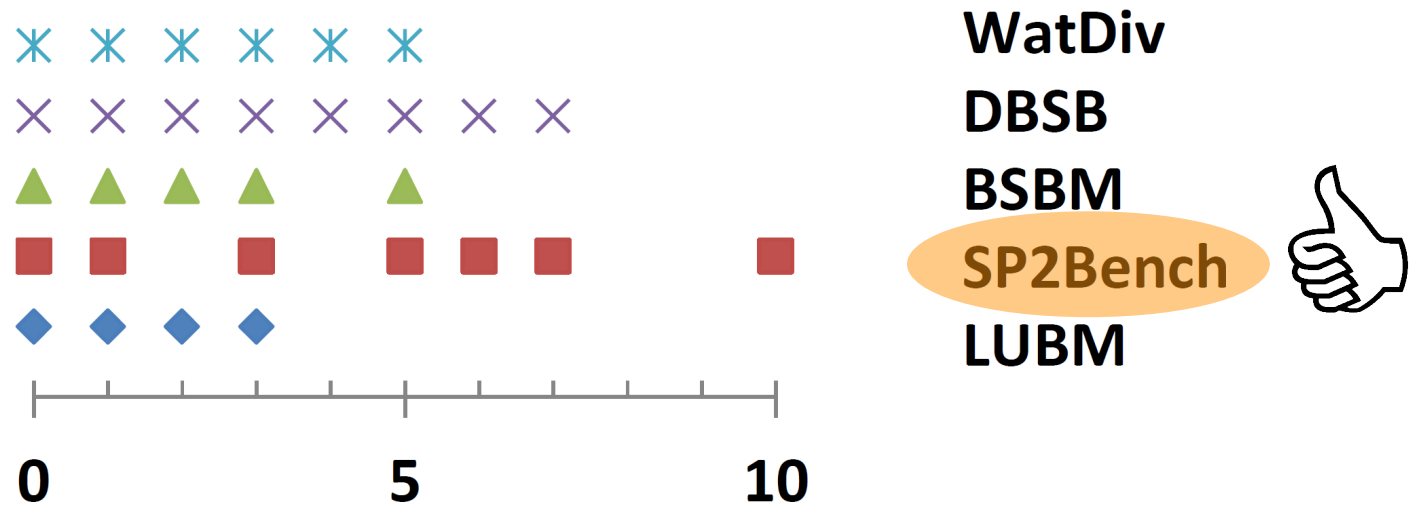
How Diverse are SPARQL Benchmarks?

[Join Vertex Count]



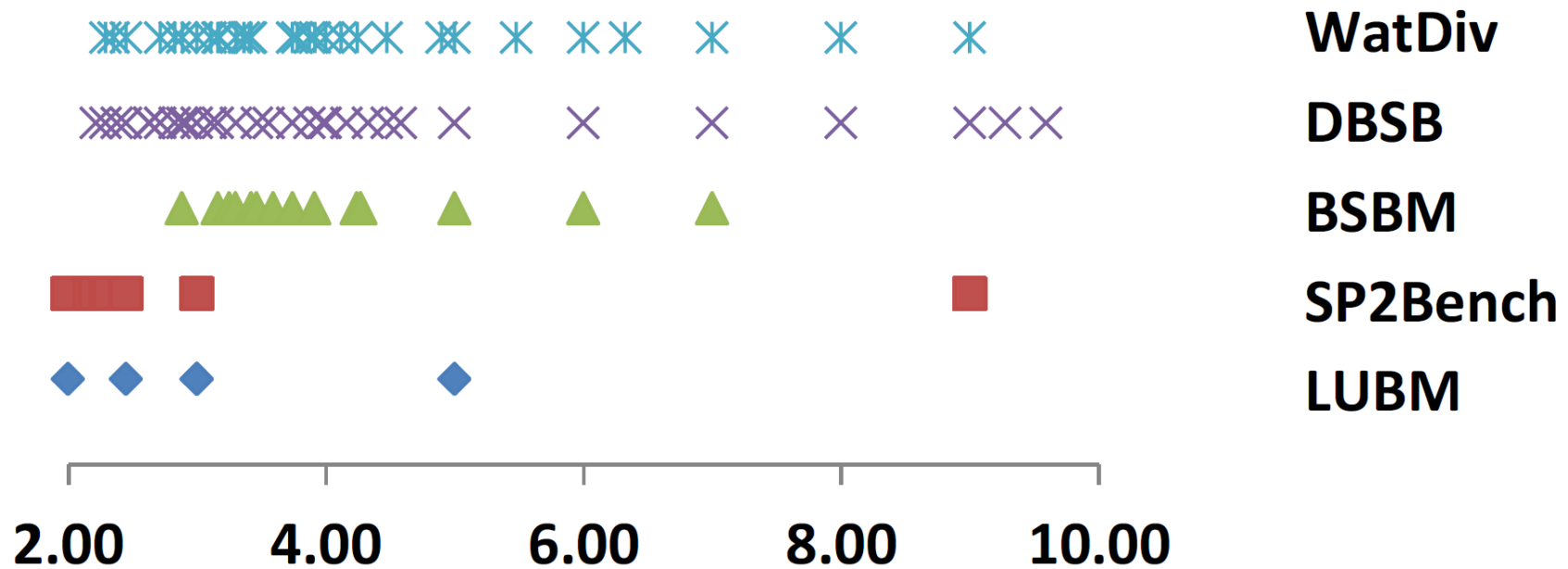
How Diverse are SPARQL Benchmarks?

[Join Vertex Count]



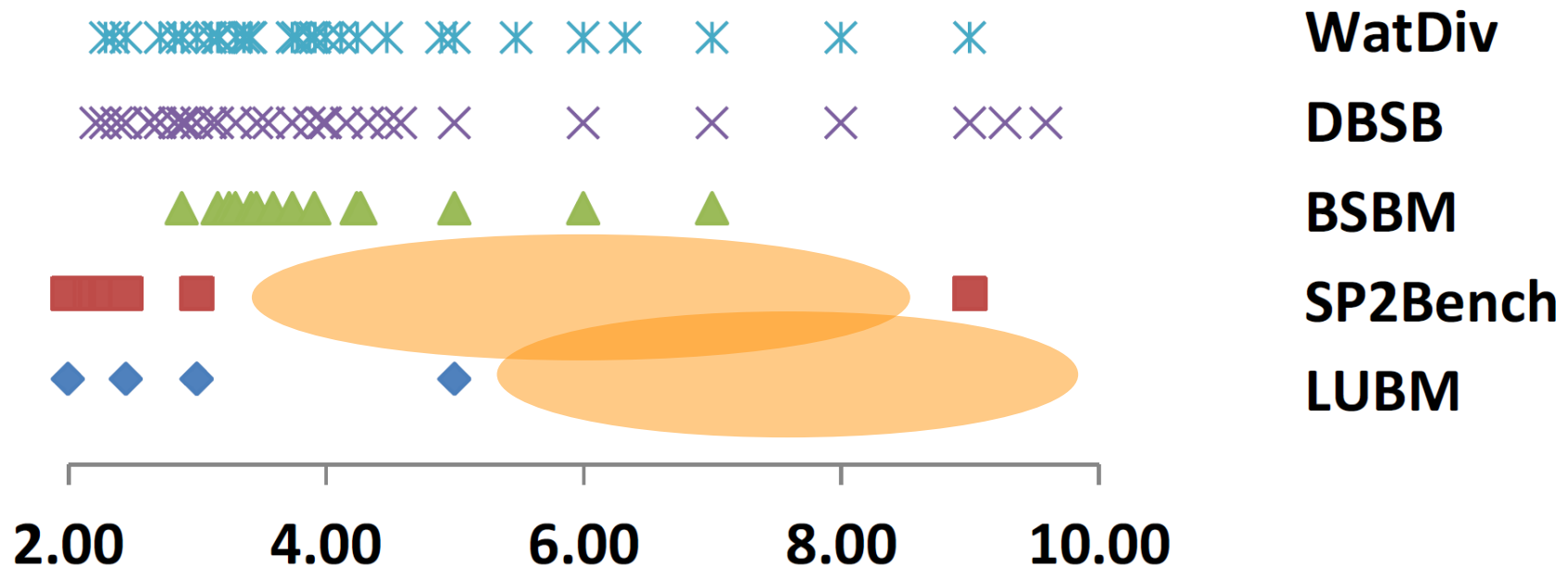
How Diverse are SPARQL Benchmarks?

[Join Vertex Degree – mean]



How Diverse are SPARQL Benchmarks?

[Join Vertex Degree – mean]



How Diverse are SPARQL Benchmarks?

[Join Vertex Type – % Queries w/in Workload]

	SS ⁺	OO ⁺	SO ⁺
LUBM	78.6 %	0.0 %	42.9 %
SP ² Bench	81.0 %	33.3 %	57.1 %
BSBM	84.8 %	5.6 %	52.8 %
DBSB	41.1 %	4.4 %	5.4 %
WatDiv	61.3 %	18.0 %	61.3 %

How Diverse are SPARQL Benchmarks?

[Join Vertex Type – % Queries w/in Workload]

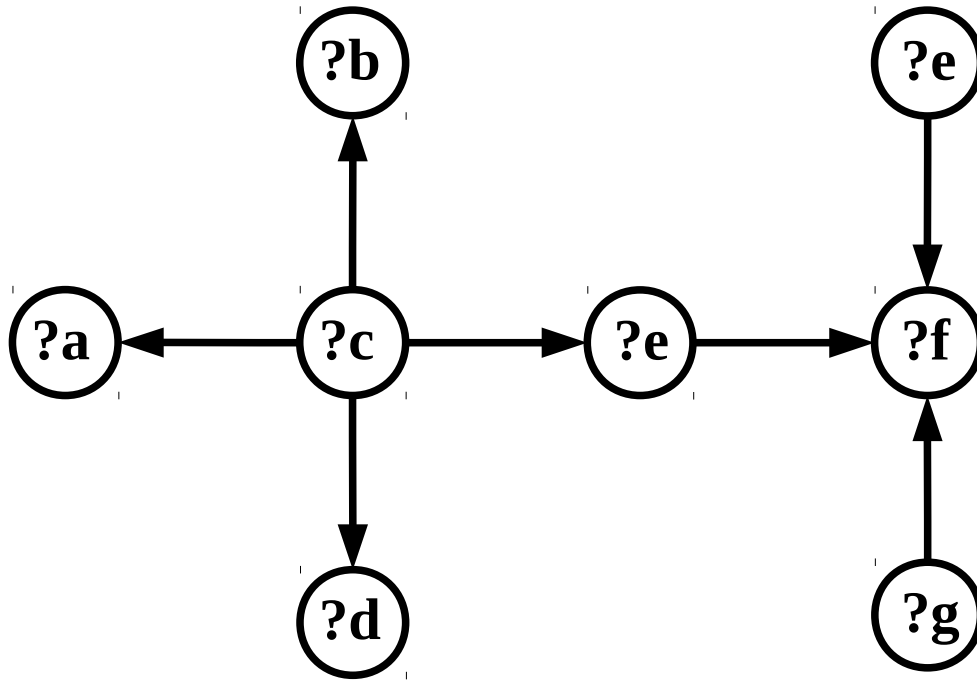
	SS ⁺	OO ⁺	SO ⁺
LUBM	78.6 %	0.0 %	42.9 %
SP ² Bench	81.0 %	33.3 %	57.1 %
BSBM	84.8 %	5.6 %	52.8 %
DBSB	41.1 %	4.4 %	5.4 %
WatDiv	61.3 %	18.0 %	61.3 %

Data-Driven Features

- Why are data-driven query features important?
 - *Why are structural features not sufficient?*
 - *Why is analysis based purely on the data not sufficient?*

Data-Driven Features

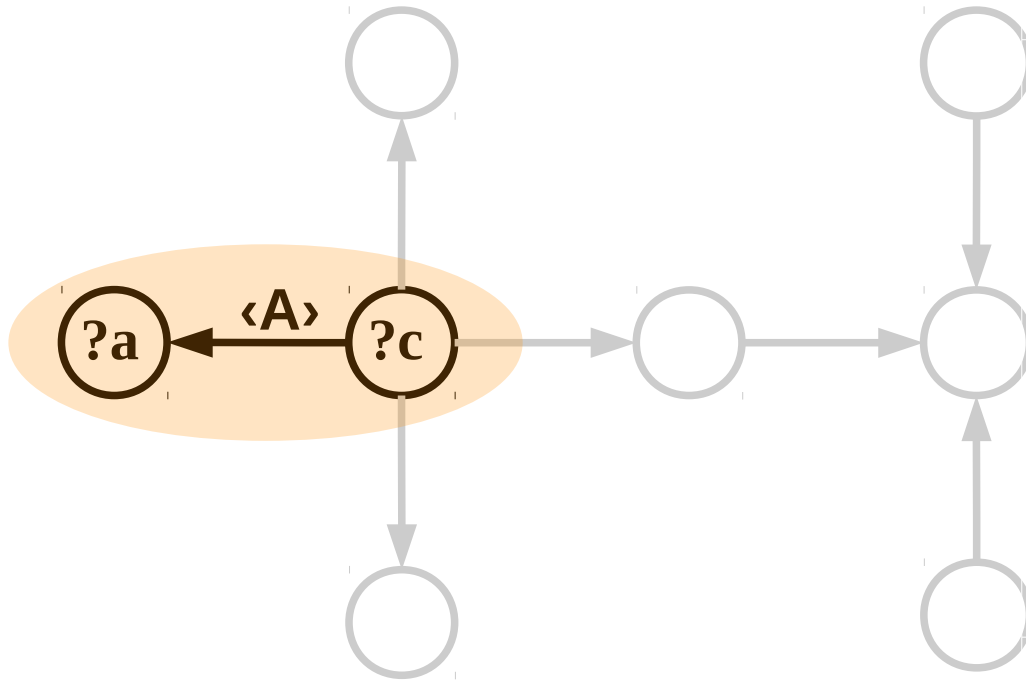
[Result Cardinality]



	?a	...	?g
1			
2			
.			
.			
.			
k			

Data-Driven Features

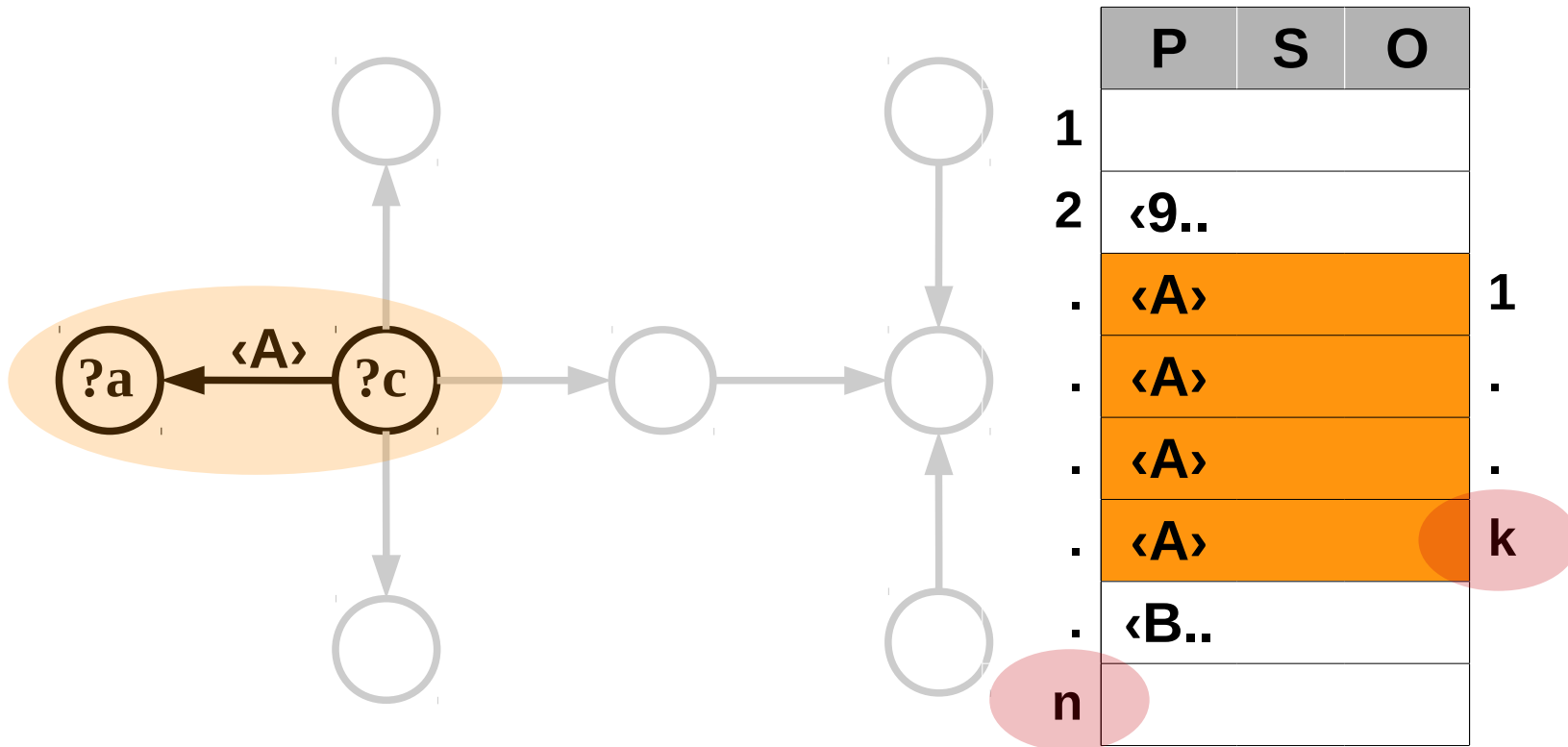
[Filtered Triple Pattern (f-TP) Selectivity]



P	S	O
<9..		
<A>		
<A>		
<A>		
<A>		
<B..		

Data-Driven Features

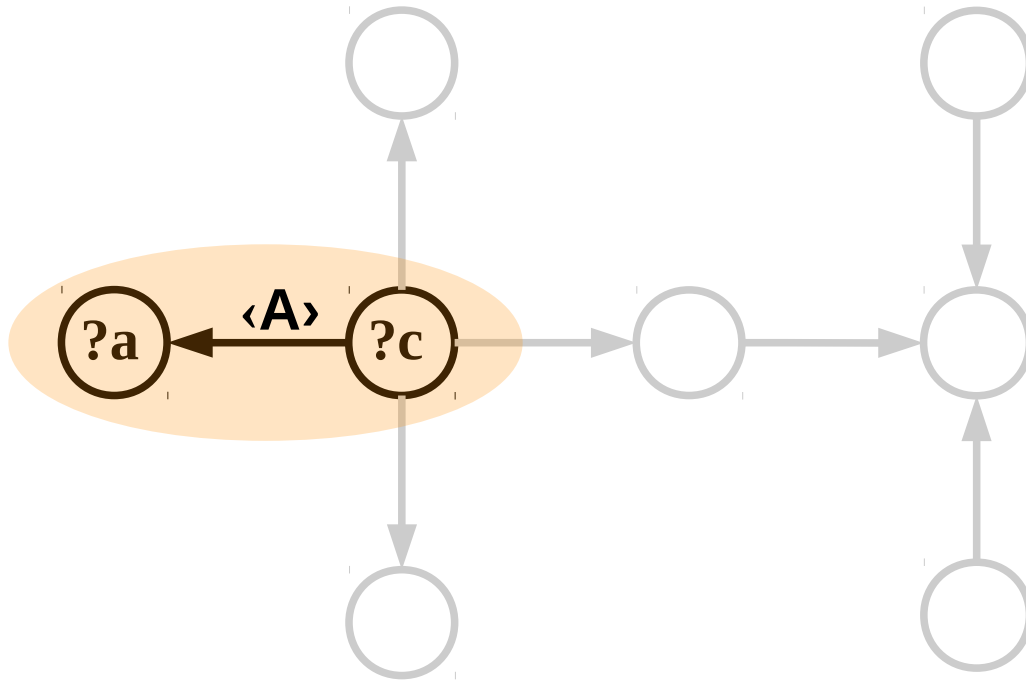
[Filtered Triple Pattern (f-TP) Selectivity]



$$\text{f-TP Selectivity} = k / n$$

Data-Driven Features

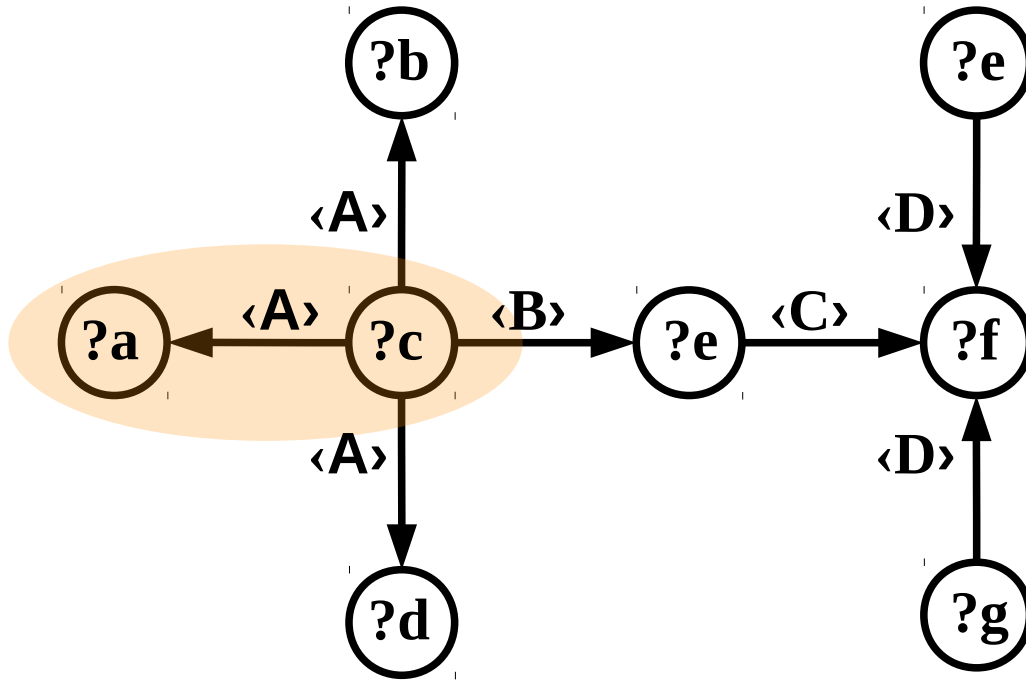
[f-TP Selectivity, **BGP-Restricted**]



P	S	O
		<9..
		<A>
		<A>
		<A>
		<A>
		<B..

Data-Driven Features

[f-TP Selectivity, **BGP-Restricted**]



P	S	O
	<9..	
	<A>	
	<A>	
	<A>	
	<A>	
	<B..	

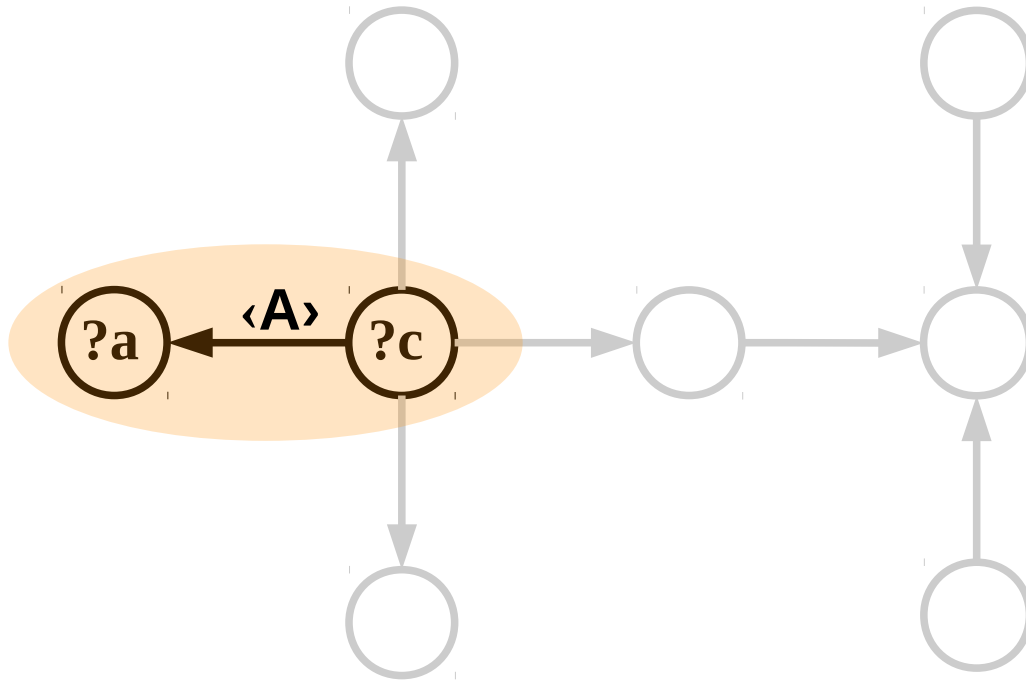
?a	...	?g

BGP-Restricted

f-TP Selectivity = |blue| / |orange|

Data-Driven Features

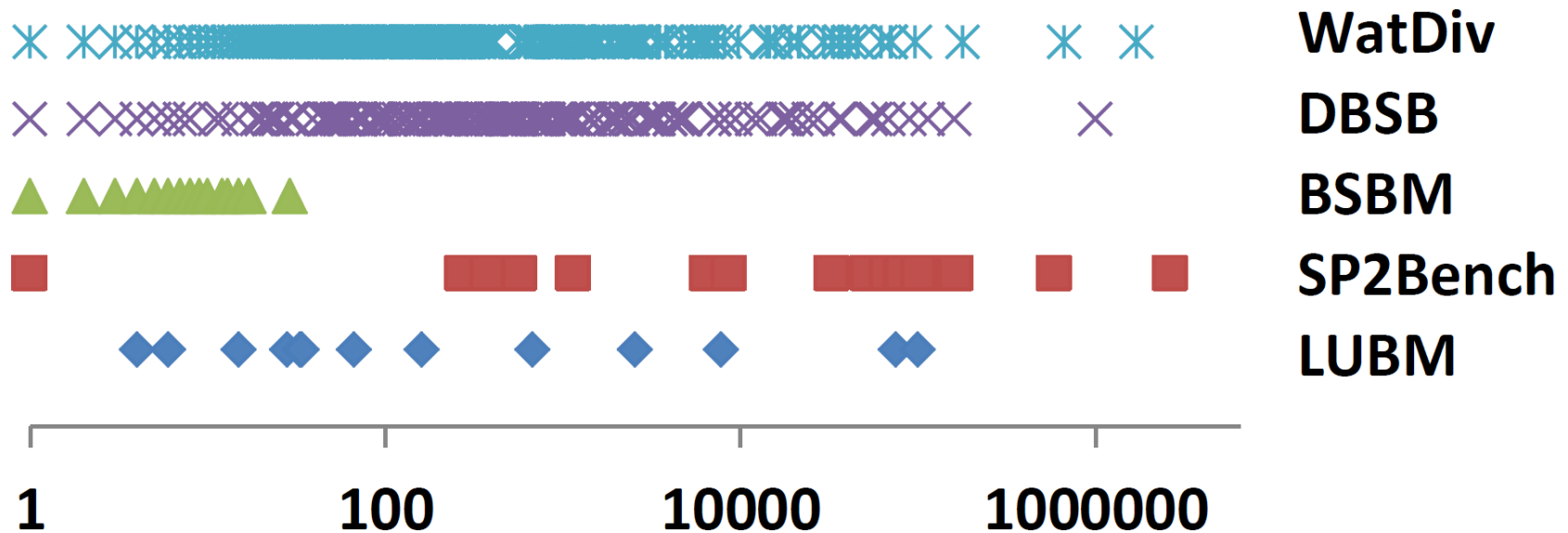
[f-TP Selectivity, **Join-Restricted**]



P	S	O
		<9..
		<A>
		<A>
		<A>
		<A>
		<B..

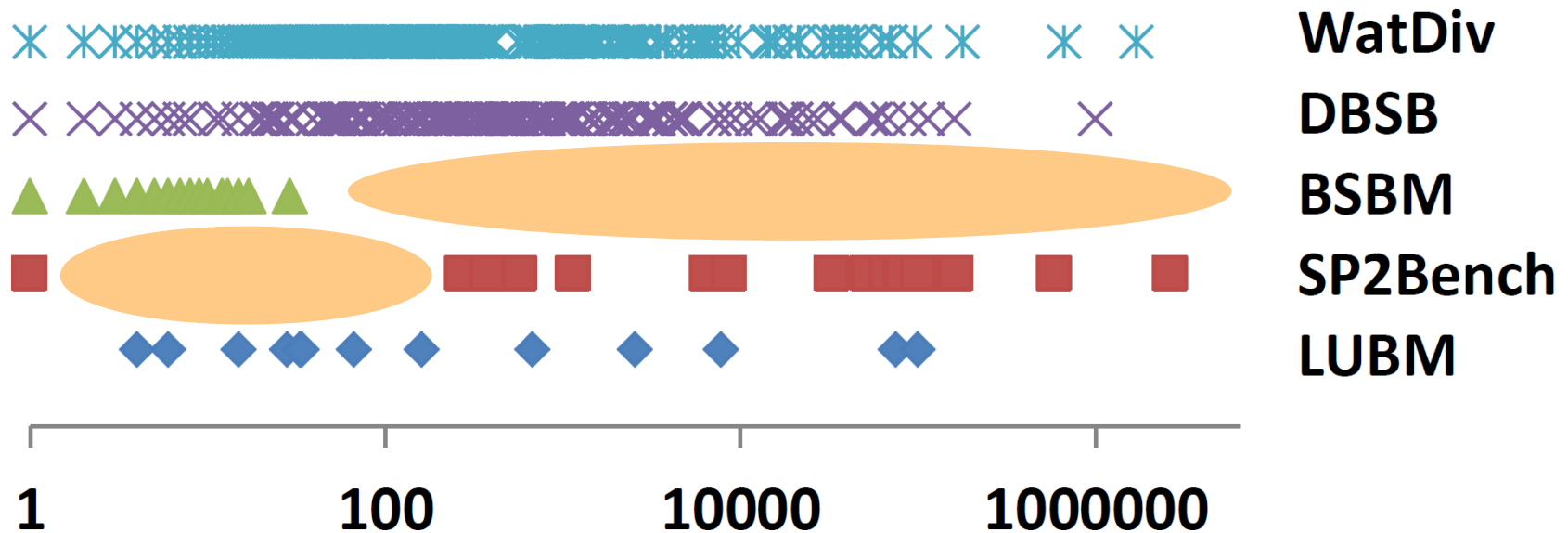
How Diverse are SPARQL Benchmarks?

[Result Cardinality]



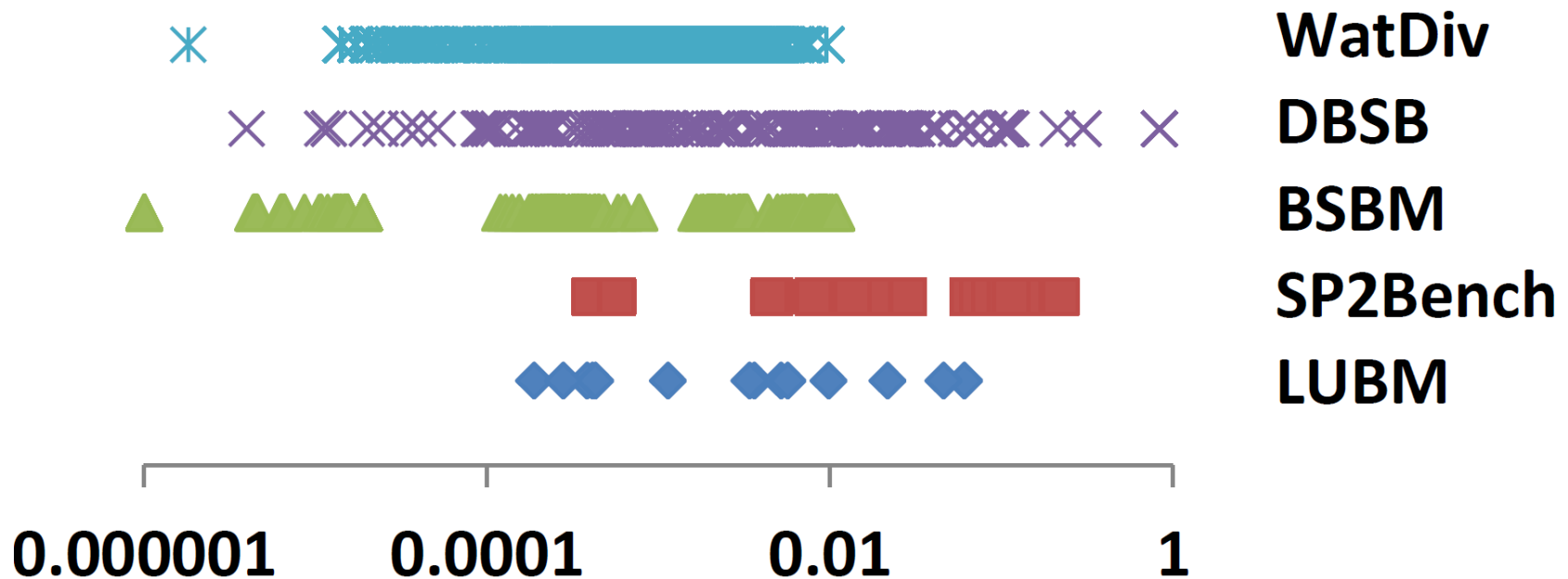
How Diverse are SPARQL Benchmarks?

[Result Cardinality]



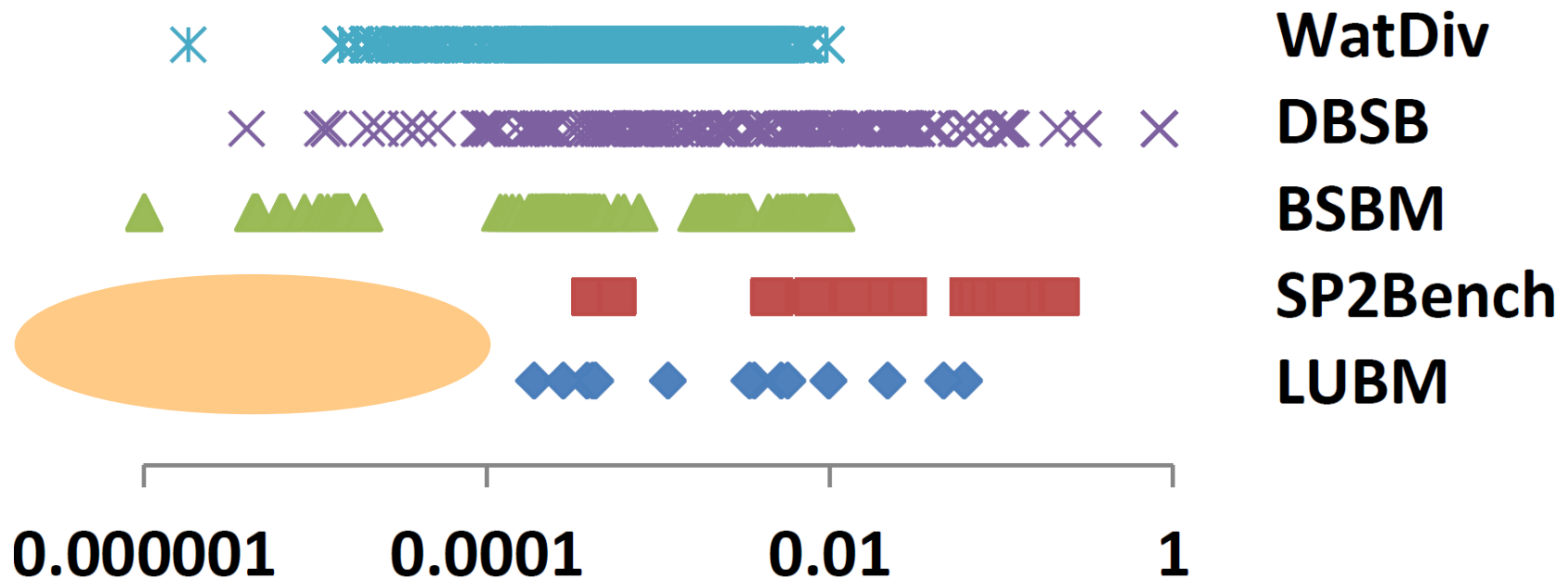
How Diverse are SPARQL Benchmarks?

[f-TP Selectivity – mean]



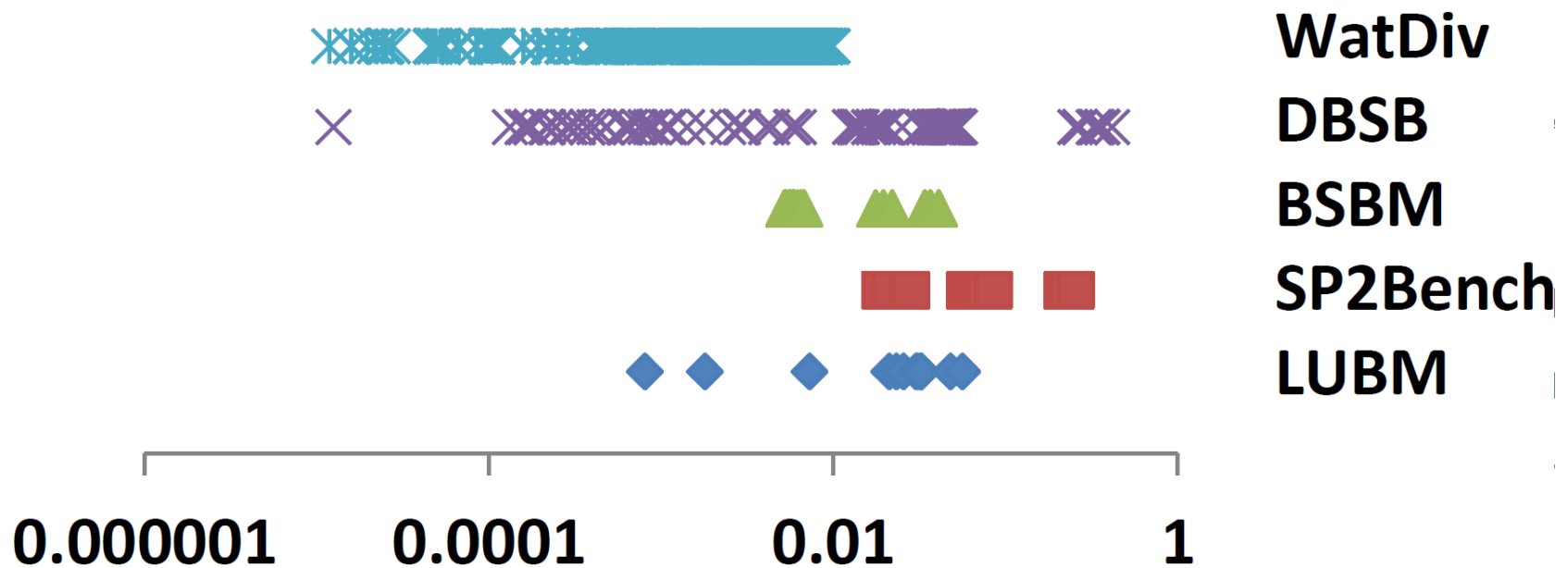
How Diverse are SPARQL Benchmarks?

[f-TP Selectivity – mean]



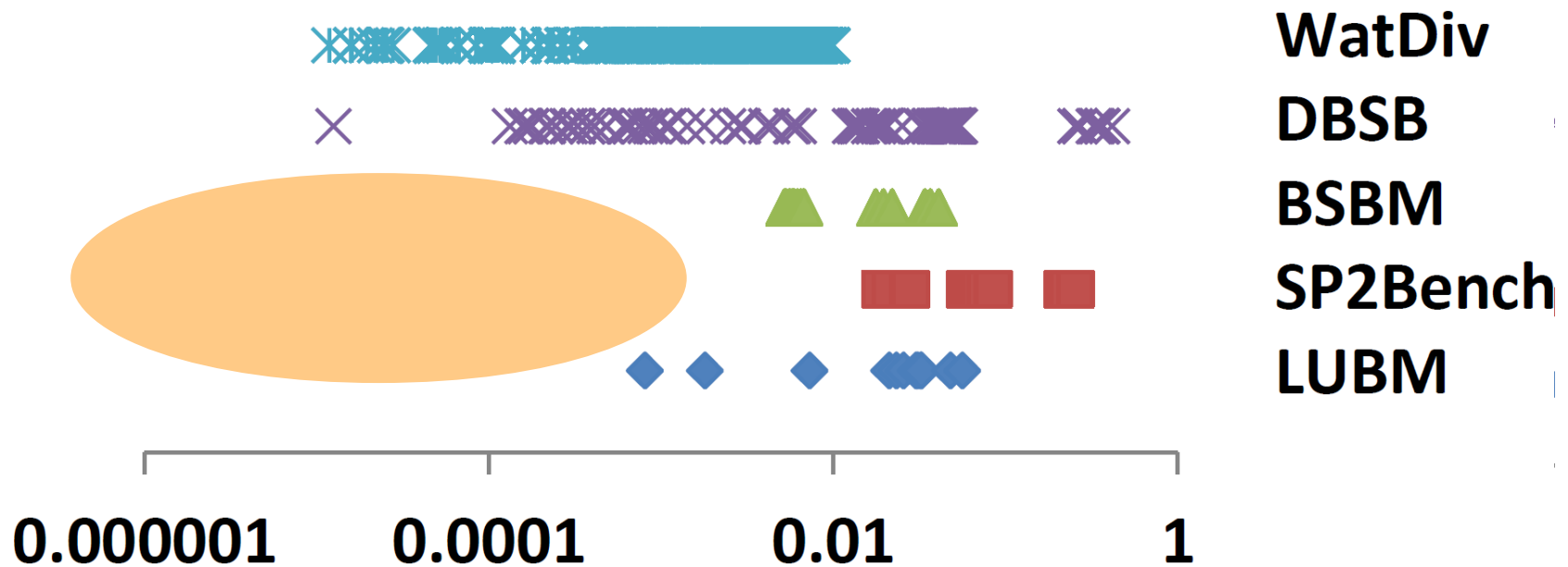
How Diverse are SPARQL Benchmarks?

[f-TP Selectivity – *stdev*]



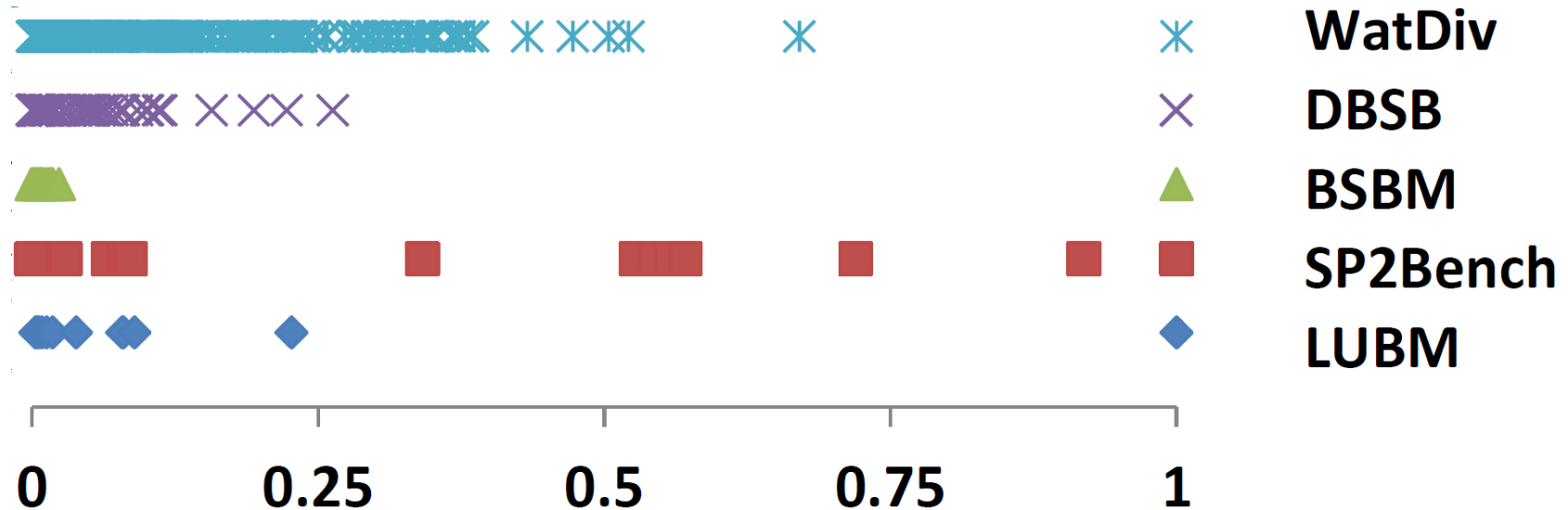
How Diverse are SPARQL Benchmarks?

[f-TP Selectivity – **stdev**]



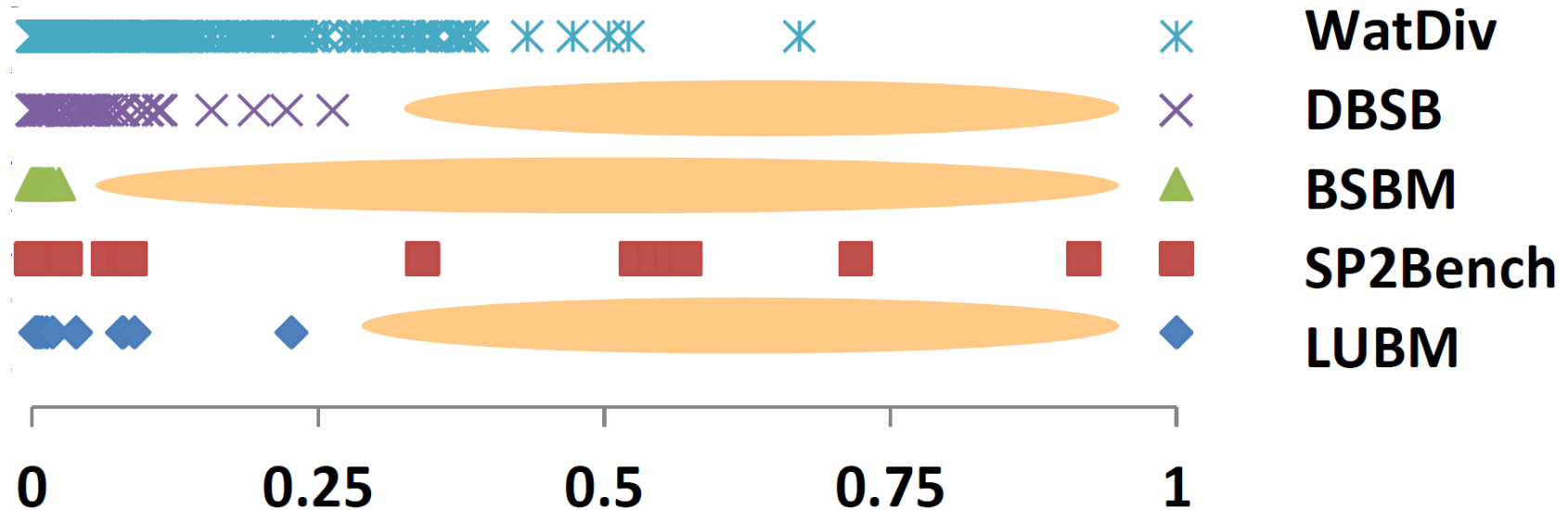
How Diverse are SPARQL Benchmarks?

[BGP-Restricted f-TP Selectivity – mean]



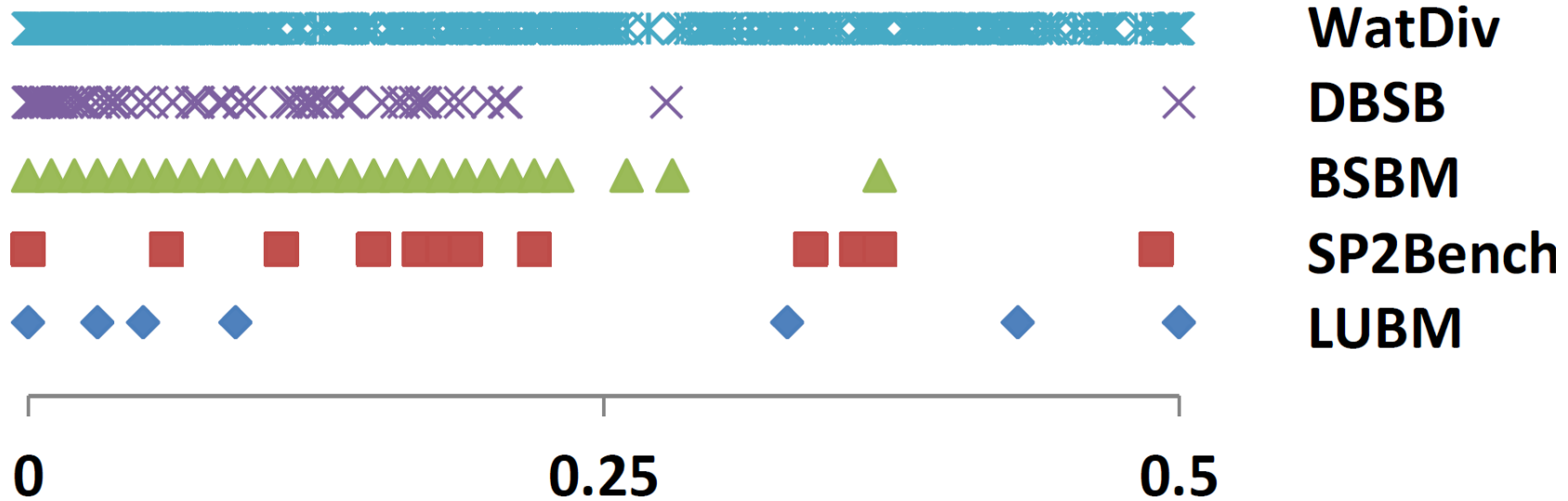
How Diverse are SPARQL Benchmarks?

[BGP-Restricted f-TP Selectivity – mean]



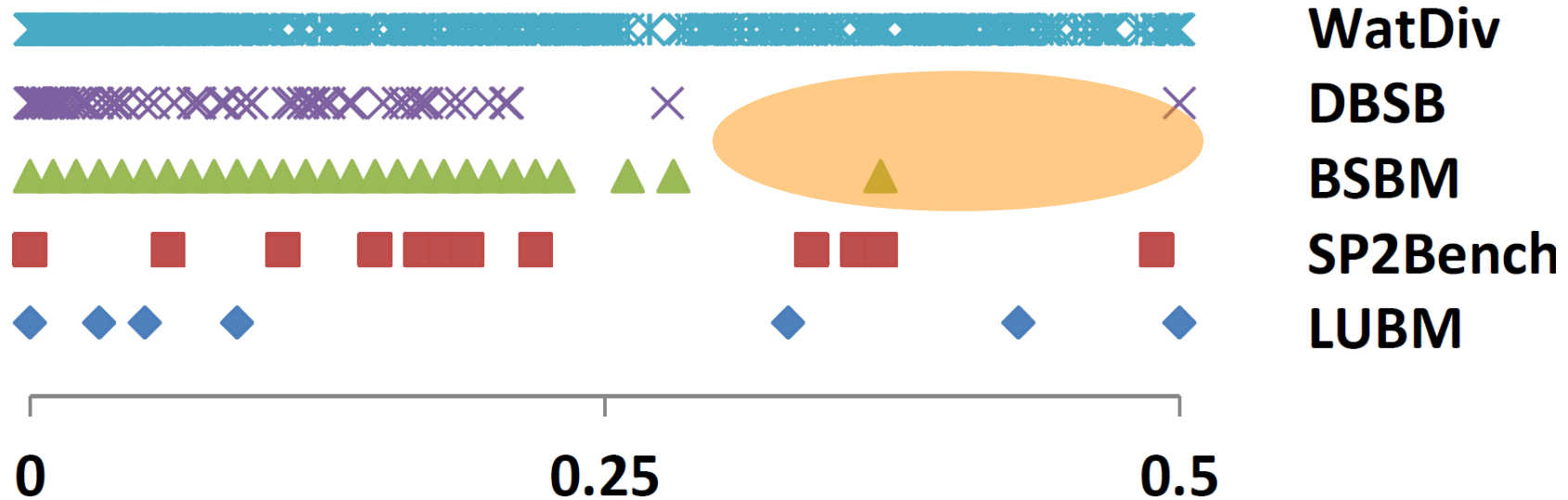
How Diverse are SPARQL Benchmarks?

[BGP-Restricted f-TP Selectivity – stdev]



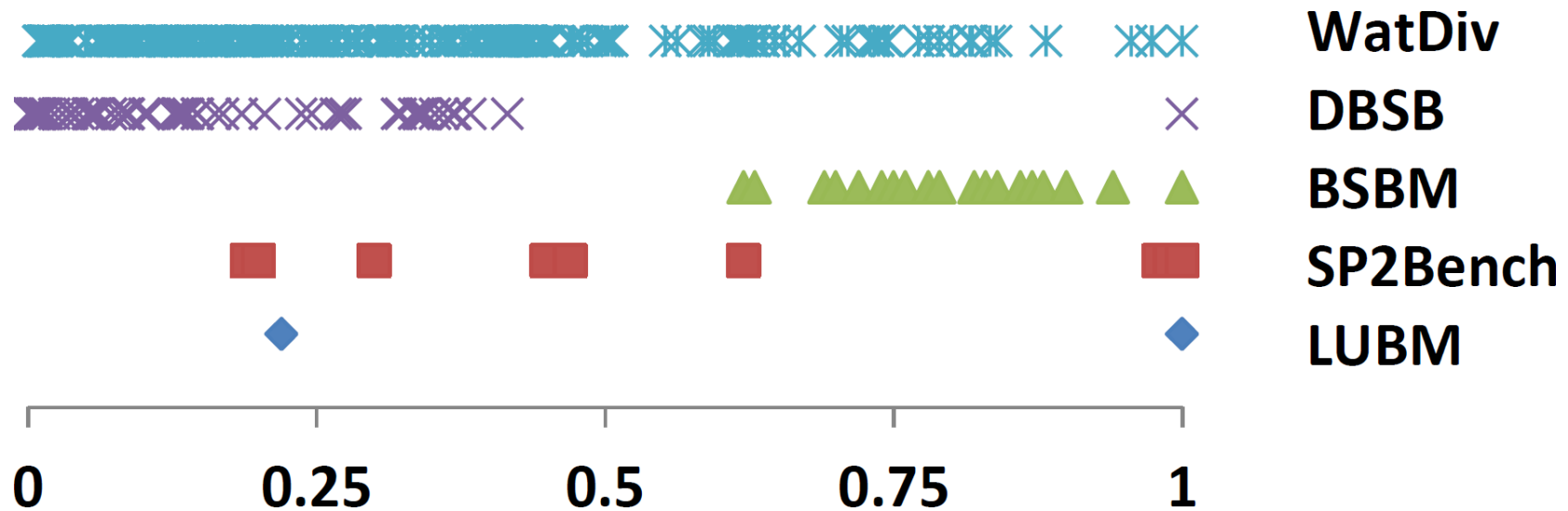
How Diverse are SPARQL Benchmarks?

[BGP-Restricted f-TP Selectivity – stdev]



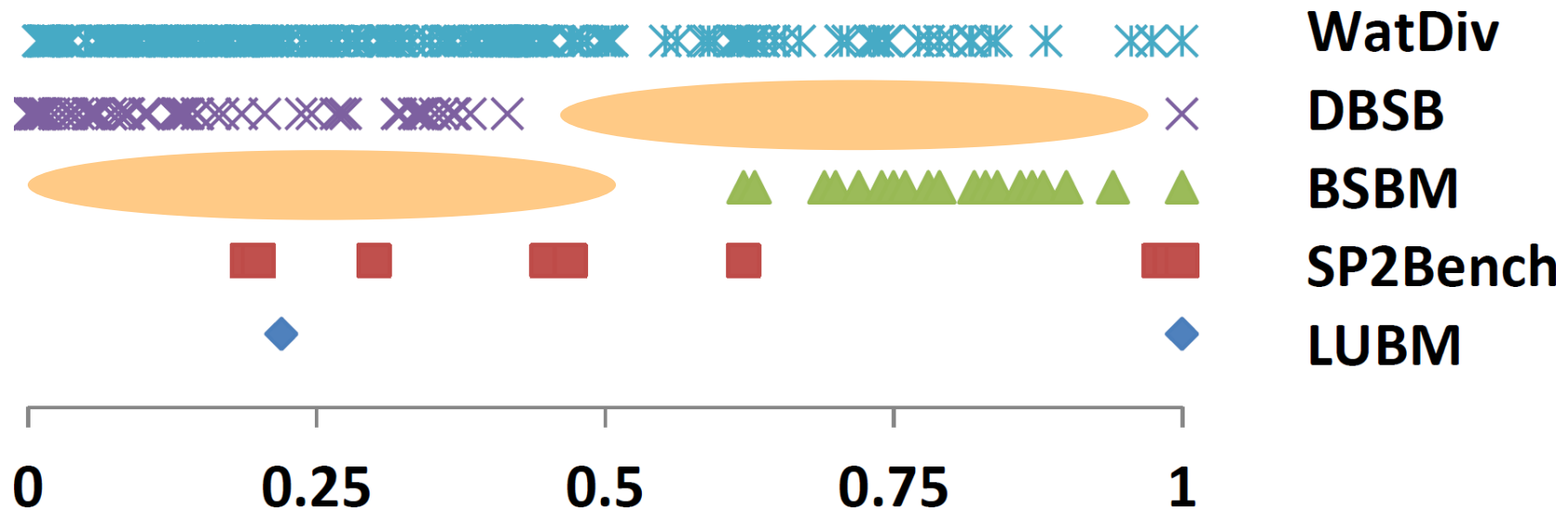
How Diverse are SPARQL Benchmarks?

[Join-Restricted f-TP Selectivity – mean]



How Diverse are SPARQL Benchmarks?

[Join-Restricted f-TP Selectivity – mean]



How Robust are Systems across WatDiv Workloads?

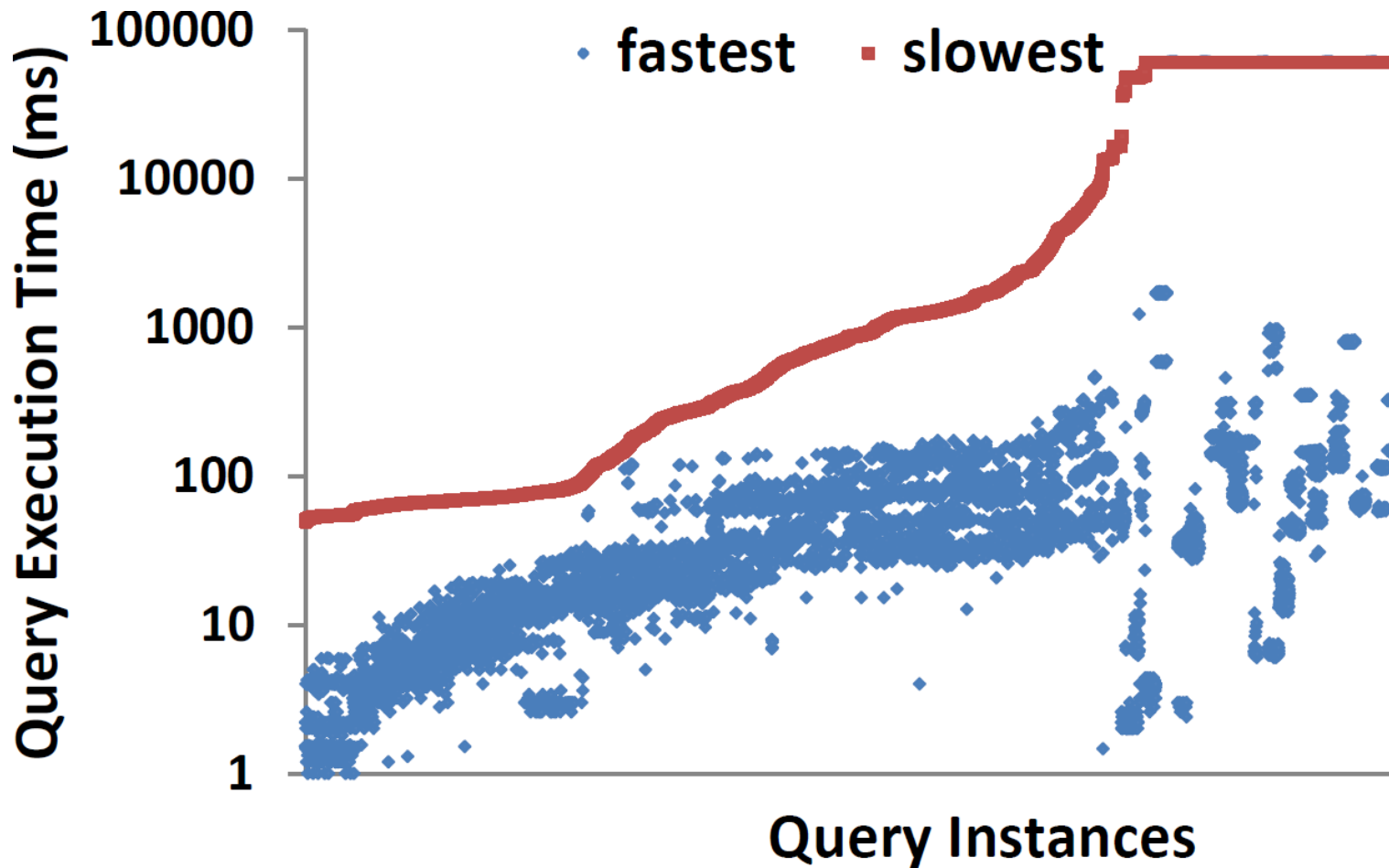
BAD NEWS

	RDF-3x	VOS [6.1]	VOS [7.1]	MonetDB	4Store
<i>fastest</i>	20.9 %		22.6 %	56.5 %	
<i>10¹x slower</i>	60.9 %	59.1 %	54.8 %	31.3 %	53.0 %
<i>10²x slower</i>	13.9 %	40.0 %	20.0 %	2.6 %	21.7 %
<i>10³x slower</i>	3.5 %	0.9 %	1.7 %	6.1 %	15.7 %
<i>10⁴x slower</i>			0.9 %	3.5 %	7.0 %
<i>10⁵x slower</i>			0.9 %	3.5 %	7.0 %

WatDiv 100M triples, stress testing workload (12500 queries)

How Robust are Systems across WatDiv Workloads?

BAD NEWS

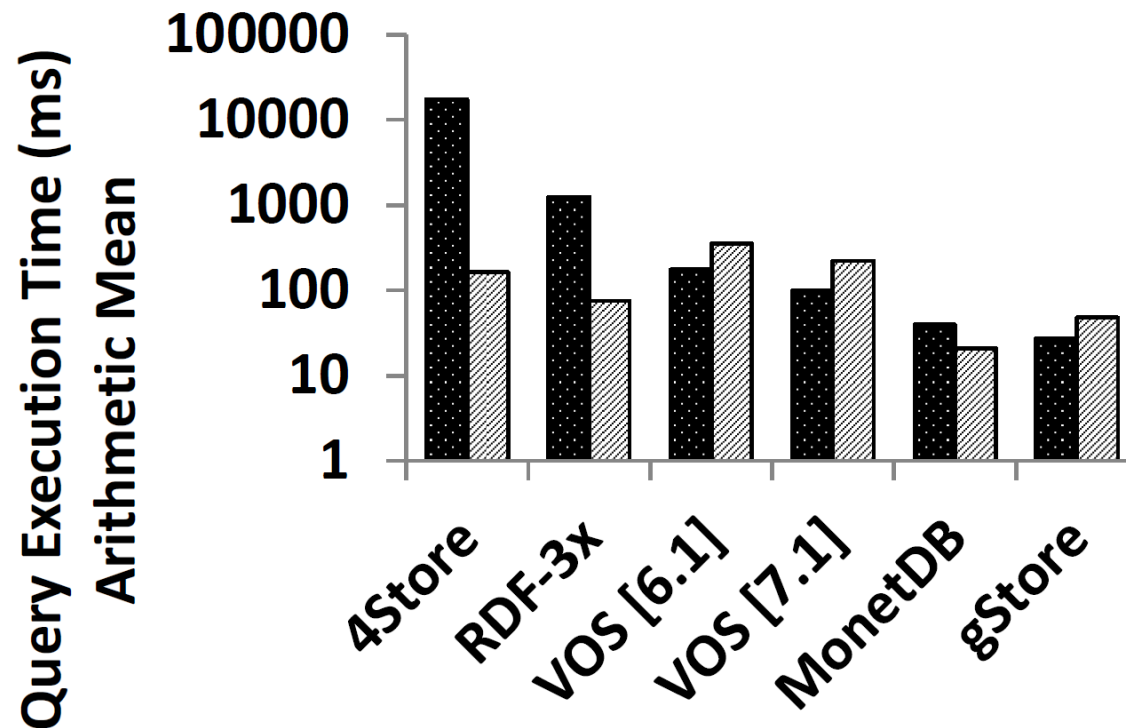


WatDiv 100M triples, stress testing workload (12500 queries)

How Robust are Systems across WatDiv Workloads?

[BGP-restricted f-TP selectivity – mean]

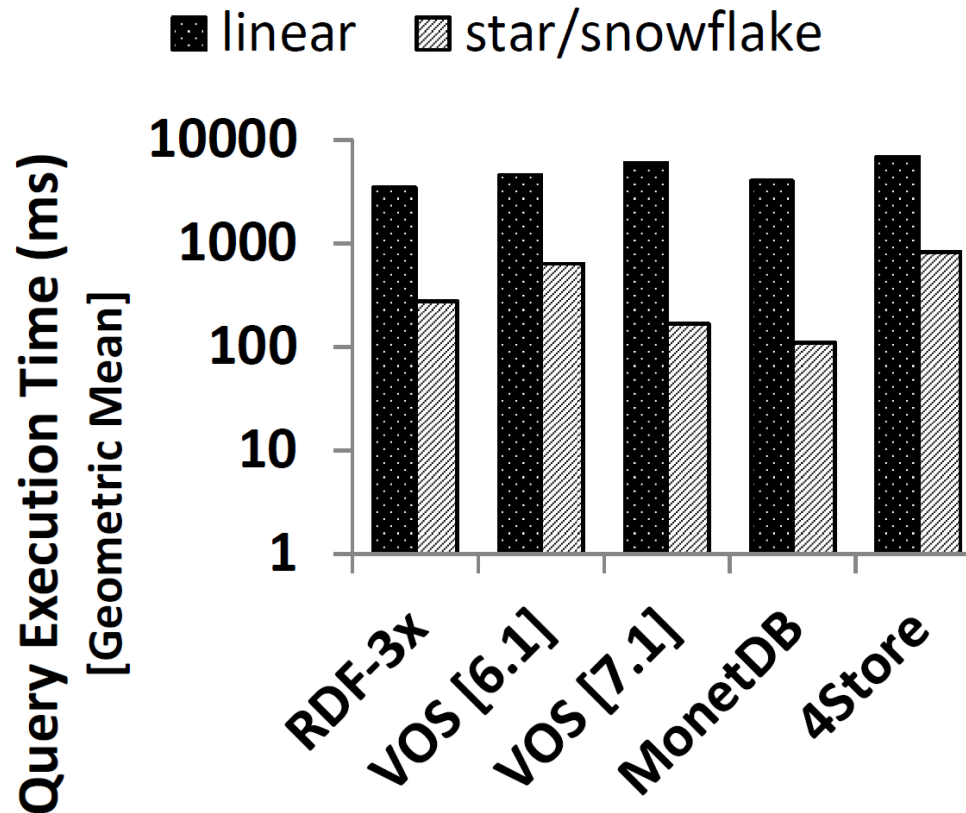
■ [0.000, 0.003) ▨ [0.030, 1.000)



WatDiv 100M triples, queries w/ single join vertex, result cardinality ≤ 2000

How Robust are Systems across WatDiv Workloads?

GOOD NEWS



WatDiv 10M triples

linear = { mean join vertex degree ≤ 3.0 , join vertex count ≥ 3 }

star/snowflake = { mean join vertex degree ≥ 5.0 , join vertex count ≤ 2 }

Conclusions

- Are popular SPARQL benchmarks really suitable for **stress-testing** RDF data management systems?
 - **No**, but they are useful for FILTER/OPT/UNION
- Are popular RDF data management systems really as **flawless** as they appear on existing benchmarks?
 - **No**

Questions

Waterloo SPARQL Diversity Test Suite
(WatDiv)



<http://db.uwaterloo.ca/watdiv/>