

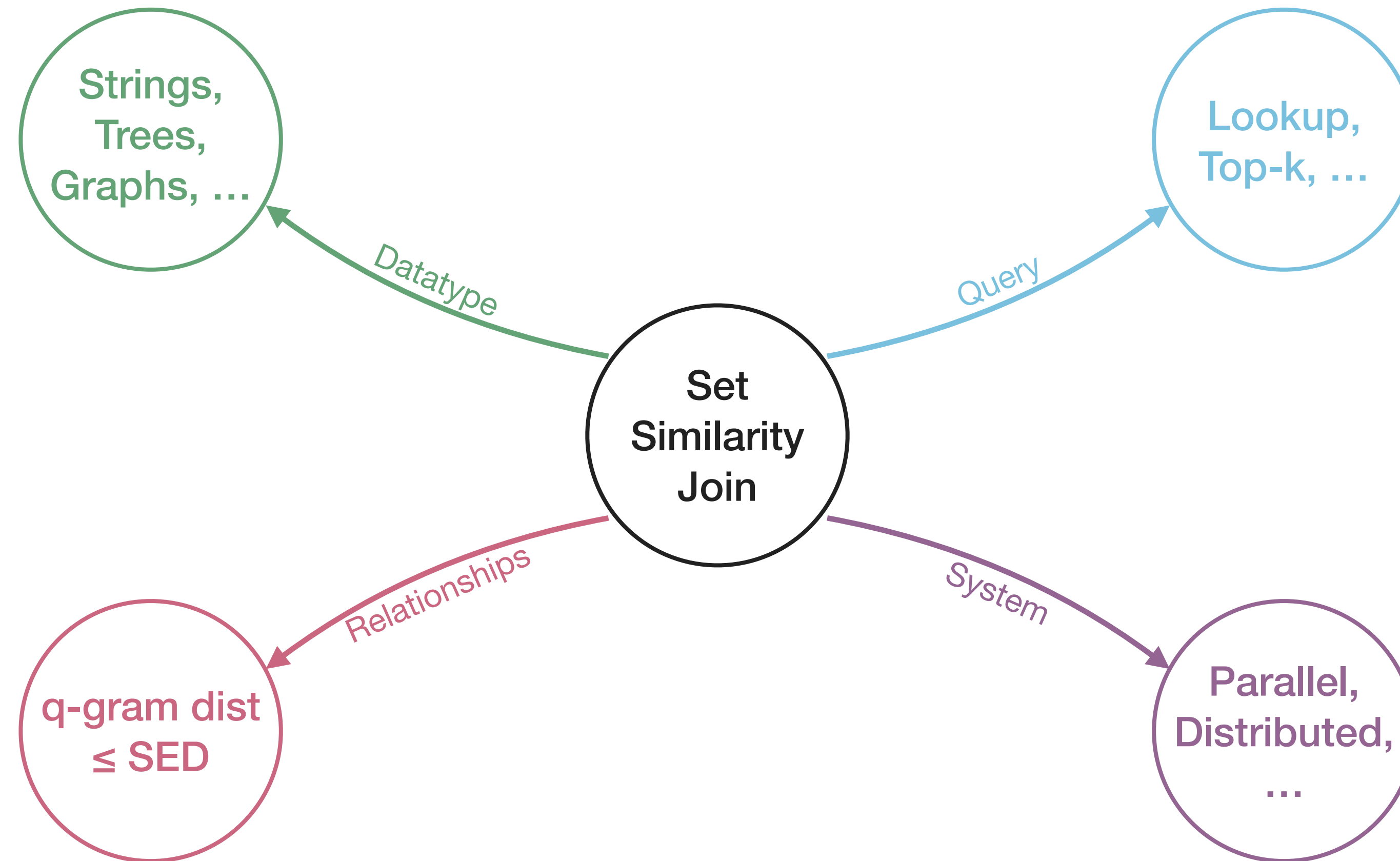
PS Ähnlichkeitssuche in großen Datenbanken

Final Session

Thomas Hütter, WS 2020

What's next ...

... in terms of similarity search?



Set Similarity Joins ...

... in a parallel system?

- Challenges:
 - Neither the sets nor their elements have an order.
 - Similar sets may be on different cluster nodes.
 - Sending data across the network is expensive.

Collection R of sets $r_i \in R$:

$|r_1| = \{a, b, c\}$

$|r_2| = \{a, c, d, e, s\}$

$|r_3| = \{b, d, e, s, t, x\}$

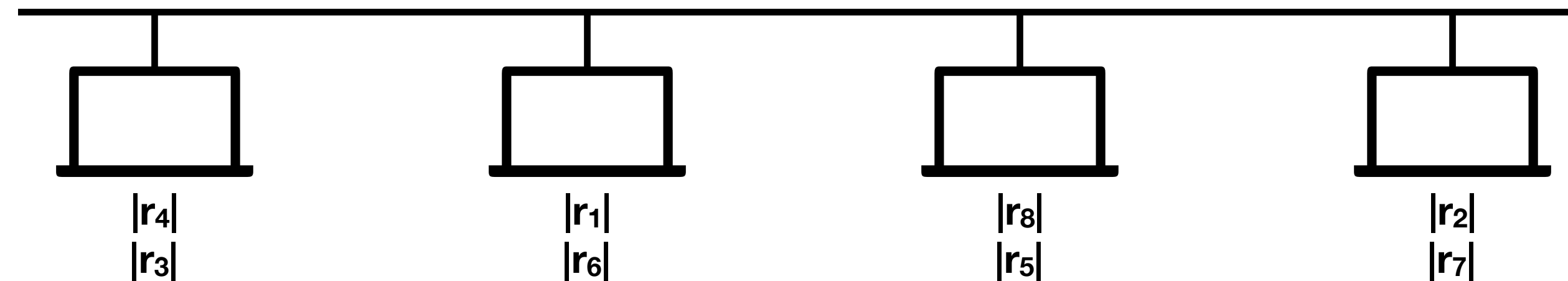
$|r_4| = \{a, b, m, n, s, u, v\}$

$|r_5| = \{d, h, i, m, n, t, x\}$

$|r_6| = \{a, e, g, k, t, u, v\}$

$|r_7| = \{b, c, e, i, n, s, t, v, w, x\}$

$|r_8| = \{a, c, d, k, m, t, u, v, w, x\}$



PS Ähnlichkeitssuche in großen Datenbanken

Evaluation

- PLUS Online (<https://online.uni-salzburg.at/>) → Ressourcen → Evaluierungen
- Direct feedback

Grading

Final

- Task 4 will be graded tomorrow.
- Comments until Tuesday, 26.01.2021, 16:55.
- The final grade will be in PLUS Online on Wednesday, 27.01.2021.