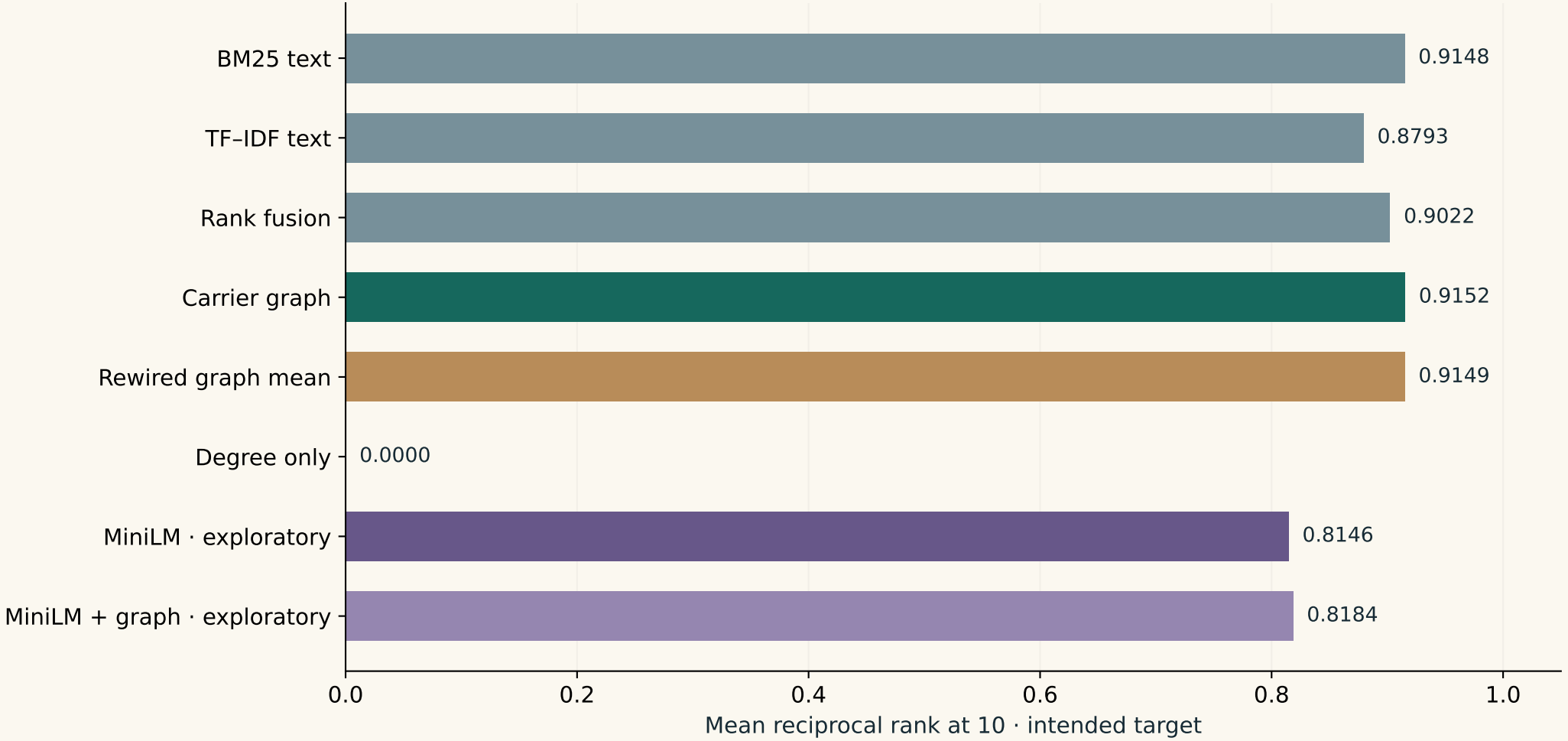
Does the right concept rise to the top?

240 held-out queries · 80 families · one intended concept per query



Carrier R01 · Frozen settings selected on 40 separate development families.
Partial relevance labels: other retrieved concepts are not automatically wrong. Bars have a common zero baseline.