RoundTripRank: Graph-based Proximity with Importance and Specificity

Yuan Fang [1,2] **Kevin C.-C. Chang** [1,2] **Hady W. Lauw** [3]

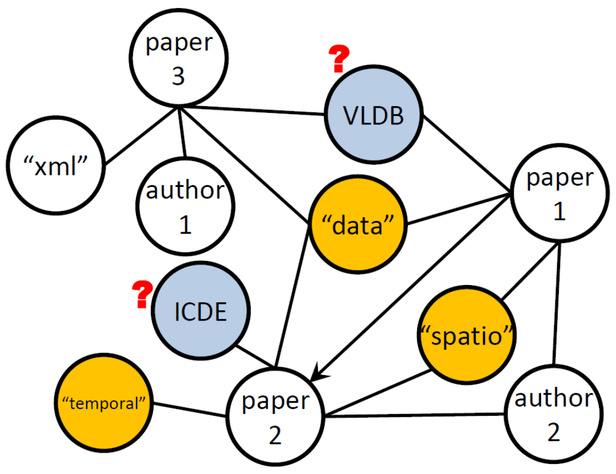
[1] University of Illinois at Urbana-Champaign, USA
[2] Advanced Digital Sciences Center, Singapore
[3] Singapore Management University, Singapore

arise.adsc.com.sg

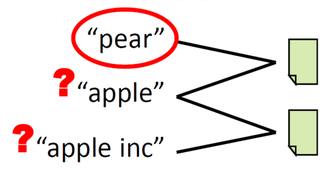
The Data and Information Systems Laboratories at The University of Illinois at Urbana-Champaign — Large Scale Information Management and Mining

Problem: *Proximity Ranking on the Graph*

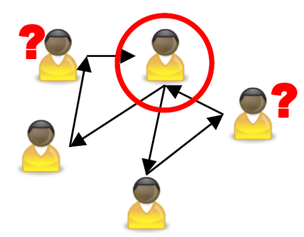
Citation network



Query log graph



Social network



Example Applications

- Find expert reviewers for a paper
- Find matching venues for some keywords
- Find relevant pages for a search phrase
- Find similar search phrases of a given phrase

Observations: *Importance & Specificity and Their Flexible Trade-offs*

Query

"spatio", "temporal", "data"

Matching venues by P-PageRank

SIGMOD	Intl Conference
VLDB	Intl Conference
ICDE	Intl Conference

Reasonable, but what's missing?

background study?
report prelim findings?

More *specific* venues?

Spatio-Temporal Databases	Springer Book
Spatio-Temporal Data Mining	Intl Workshop
Temporal Aspects in Information Systems	Working Conference

A *balanced* mixture of venues?

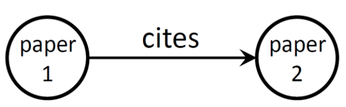
important	VLDB	Intl Conference
specific	Spatio-Temporal Databases	Springer Book
balanced	ACM SIGSPATIAL/GIS	Intl Conference

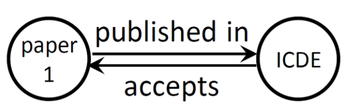
Observation 1:
Most tasks require both importance and specificity.

Observation 2:
The desirable trade-off varies from task to task.

Challenge 1: *How to unify importance & specificity?*

Citation analogy for importance/specificity

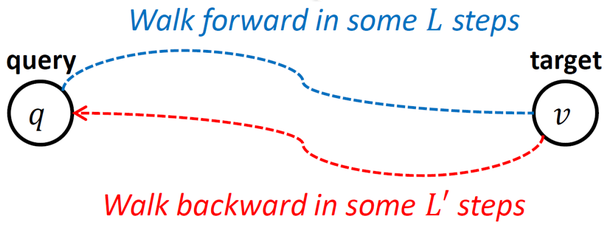




Importance of v w.r.t. q

Specificity of v w.r.t. q

Walk forward in some L steps



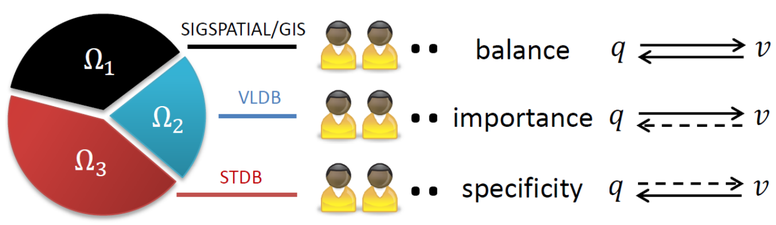
Walk backward in some L' steps

RoundTripRank

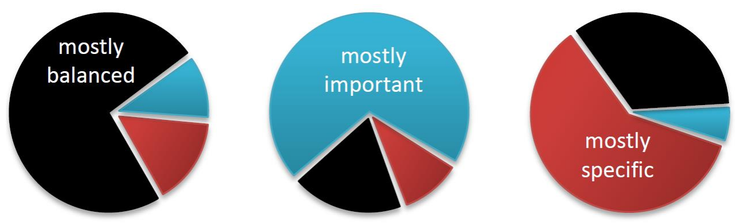
$$r(q,v) \triangleq p(W_L = v | W_0 = W_{L+L'}, W_0 = q)$$

Challenge 2: *How to customize their trade-offs?*

Hybrid surfers Ω



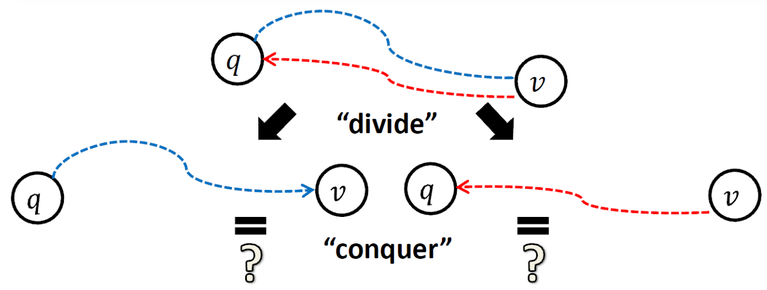
Customization



RoundTripRank+ $r_\Omega(q,v) \triangleq$

$$p(x = v | \forall \omega \in \Omega: W_0^\omega = W_{L+L'}^\omega = q, W_L^\omega = x)$$

Challenge 3: *How to compute the proximity efficiently?*



Divide and conquer

$$r(q,v) \propto p(W_L = v | W_0 = q) \times p(W_{L'} = q | W_0 = v)$$

F-Rank: $f(q,v)$ (reachability **f**rom q)

T-Rank: $t(q,v)$ (reachability **t**o q)

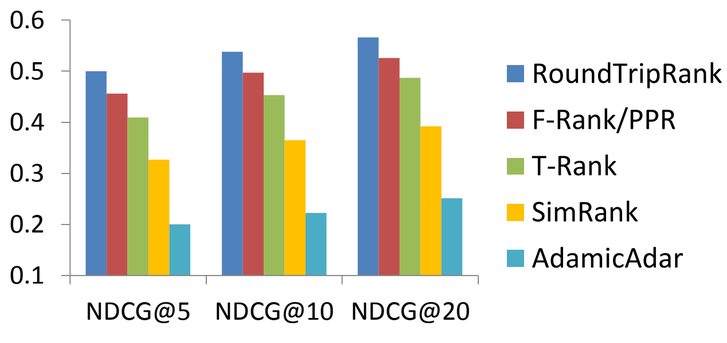
RoundTripRank: $r(q,v) \propto f(q,v)t(q,v)$

RoundTripRank+: $r_\Omega(q,v) \propto f(q,v)^{1-\beta} t(q,v)^\beta$

Specificity bias: $\beta = \frac{|\Omega_1|+|\Omega_3|}{2|\Omega_1|+|\Omega_2|+|\Omega_3|} \in [0,1]$

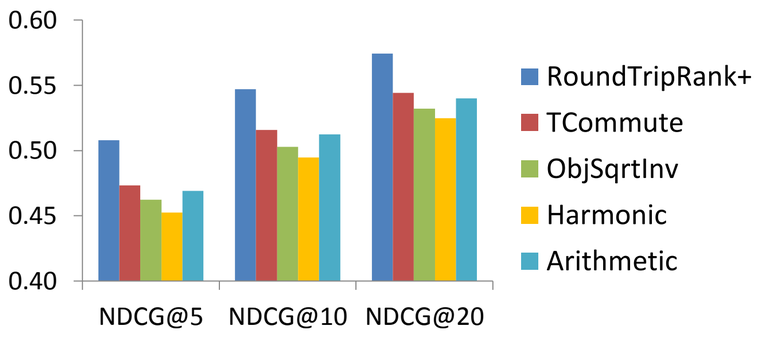
Finding 1: *Both importance & specificity are needed.*

RoundTripRank vs. mono-sensed proximity



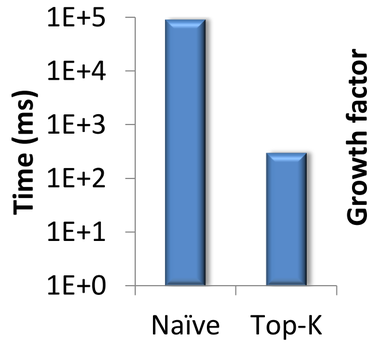
Finding 2: *Customizable trade-offs are needed.*

RoundTripRank+ vs. dual-sensed proximity



Finding 3: *Our computational model enables efficient ranking.*

Efficiency



Scalability