Diversified Recommendation on Graphs: Pitfalls, Measures, and Algorithms

Onur Küçüktunç[1,2]
Erik Saule[1]
Kamer Kaya[1]
Ümit V. Çatalyürek[1,3]

[1]Dept. Biomedical Informatics
[2]Dept. of Computer Science and Engineering
[3]Dept. of Electrical and Computer Engineering
The Ohio State University

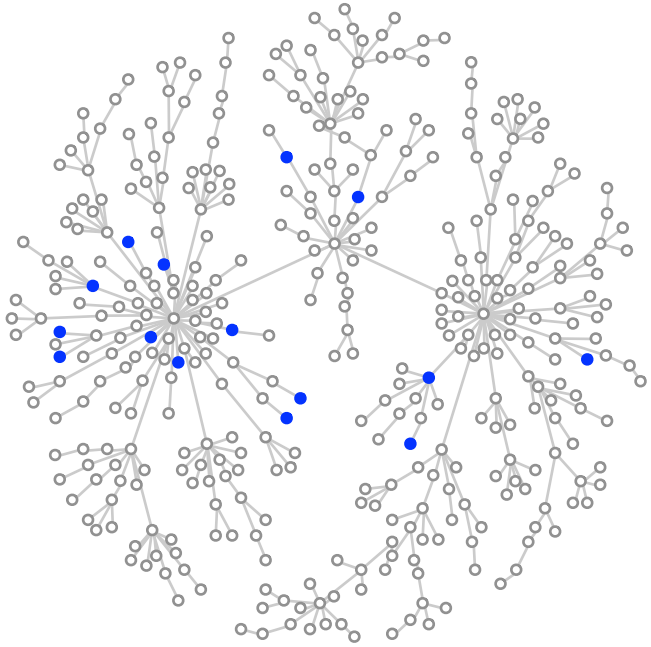
WWW 2013, May 13–17, 2013, Rio de Janeiro, Brazil.

Outline

- Problem definition
  - Motivation
  - Result diversification algorithms
- How to measure diversity
  - Classical relevance and diversity measures
  - Bicriteria optimization?!
  - Combined measures
- Best Coverage method
  - Complexity, submodularity
  - A greedy solution, relaxation
- Experiments

Problem definition

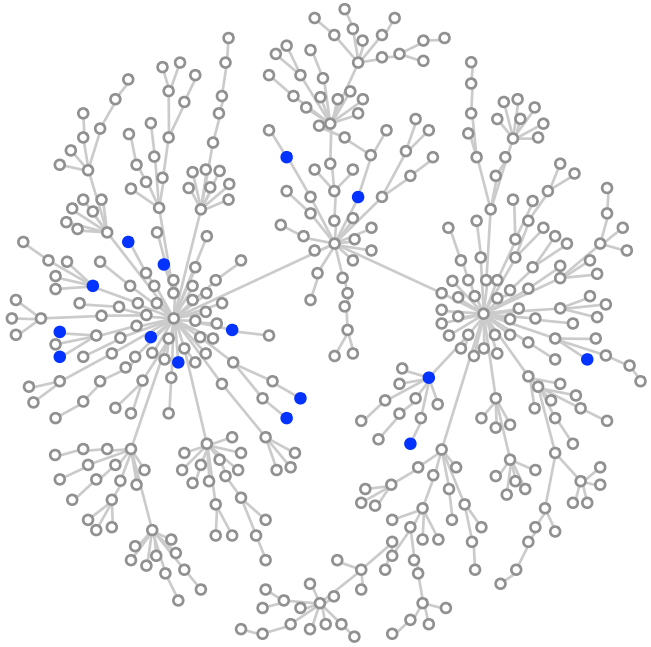
Let $G = (V, E)$ be an undirected graph.
Given a set of m seed nodes
$\mathcal{Q} = \{q_1, \ldots, q_m\}$ s.t. $\mathcal{Q} \subseteq V$,
and a parameter k, return top-k items
which are relevant to the ones in $\mathcal{Q}$,
but diverse among themselves, covering
different aspects of the query.



	Online shopping	**Academic**	**Social**
$G = (V, E)$	product co-purchasing	paper-to-paper citations; collaboration network	friendship network
$\mathcal{Q} \subseteq V$	• one product • previous purchases • page visit history	• paper/field of interest • set of references • researcher himself/herself	• user himself/herself • set of people
$\mathcal{R} \subset V$	product recommendations "you might also like…"	references for related work new collaborators	friend recommendations "you might also know…"

Problem definition

Let $G = (V, E)$ be an undirected graph.
Given a set of m seed nodes
$\mathcal{Q} = \{q_1, \ldots, q_m\}$ s.t. $\mathcal{Q} \subseteq V$,
and a parameter k, return top-k items
which are relevant to the ones in $\mathcal{Q}$,
but diverse among themselves, covering
different aspects of the query.



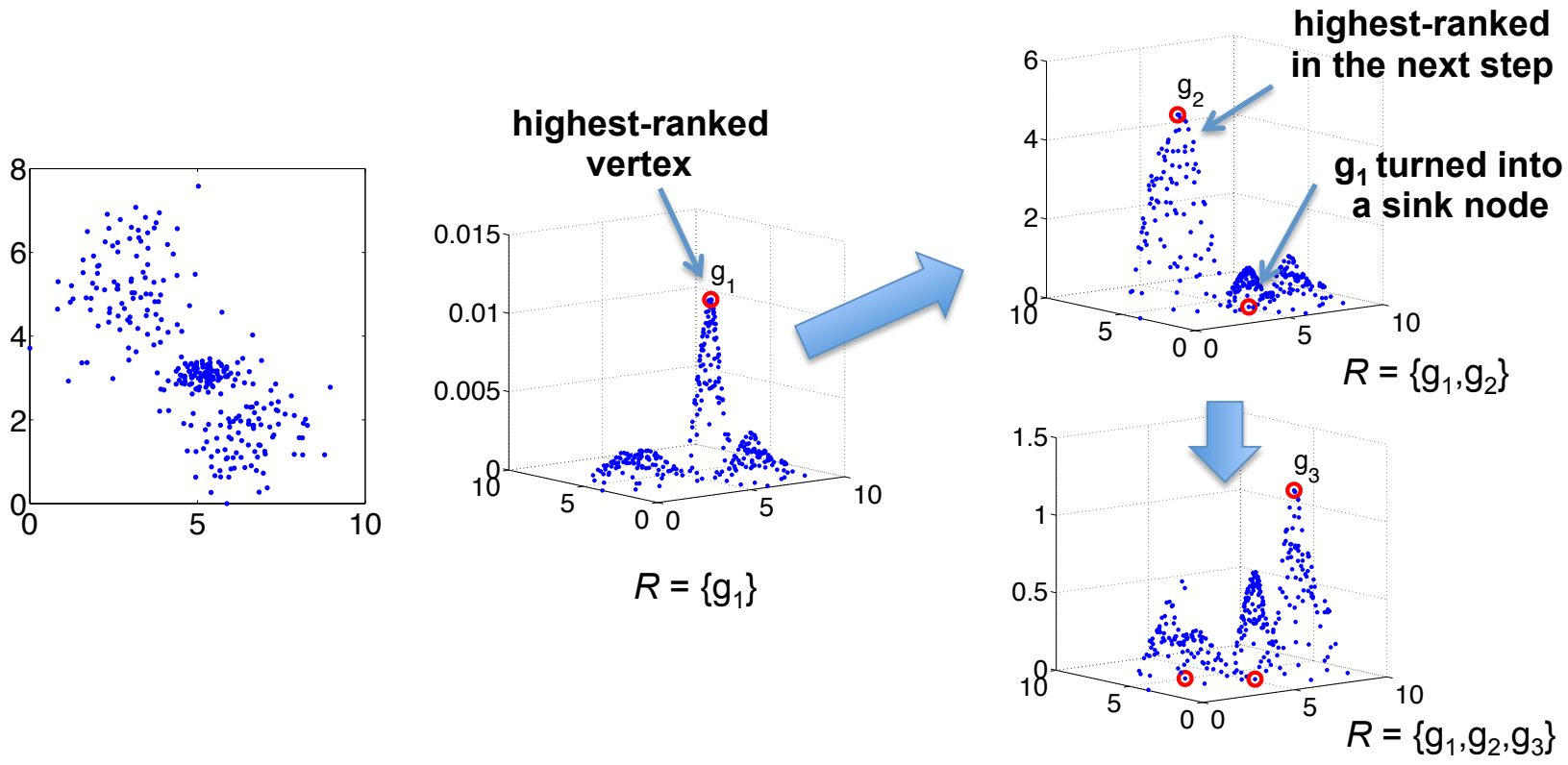
- We assume that the graph itself is the only information we have, and **no categories or intents are available**
  - *no comparisons to intent-aware algorithms [Agrawal09,Welch11,etc.]*
  - *but we will compare against intent-aware measures*

- Relevance scores are obtained with **Personalized PageRank (PPR)** [Haveliwala02]

$$p^*(v) = \begin{cases} 1/m, & \text{if } v \in \mathcal{Q} \\ 0, & \text{otherwise.} \end{cases}$$

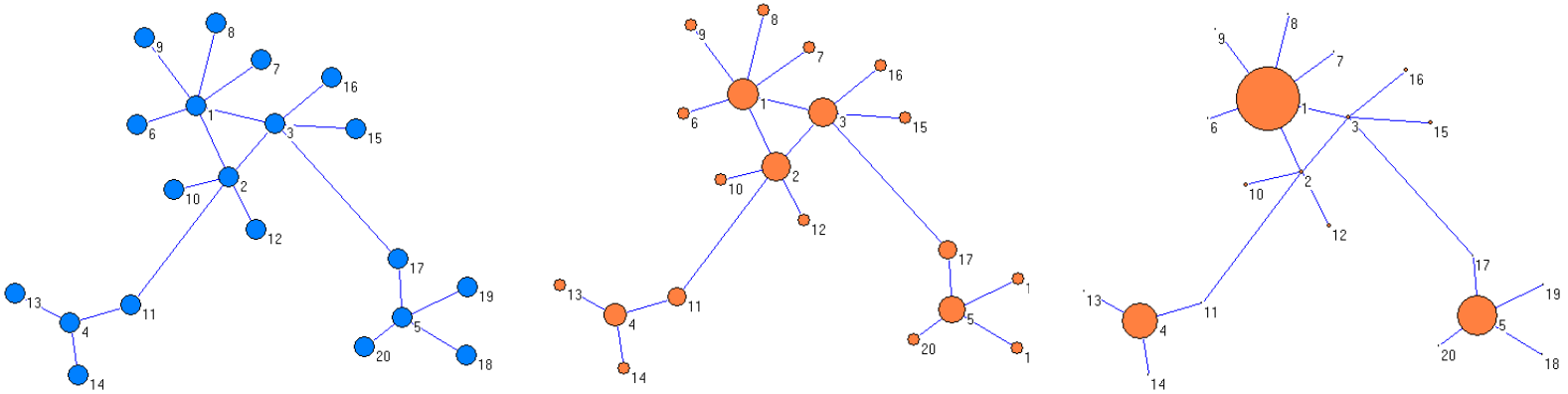
Result diversification algorithms

- GrassHopper [Zhu07]
  - ranks the graph *k* times
    - turns the highest-ranked vertex into a sink node at each iteration



Result diversification algorithms

- GrassHopper [Zhu07]
  - ranks the graph *k* times
    - turns the highest-ranked vertex into a sink node at each iteration
- DivRank [Mei10]
  - based on vertex-reinforced random walks (VRRW)
    - adjusts the transition matrix based on the number of visits to the vertices (*rich-gets-richer* mechanism)



sample graph **weighting with PPR** **diverse weighting**

Result diversification algorithms

- GrassHopper [Zhu07]
  - ranks the graph *k* times
    - turns the highest-ranked vertex into a sink node at each iteration
- DivRank [Mei10]
  - based on vertex-reinforced random walks (VRRW)
    - adjusts the transition matrix based on the number of visits to the vertices (*rich-gets-richer* mechanism)
- Dragon [Tong11]
  - based on optimizing the *goodness* measure
    - punishes the score when two neighbors are included in the results

Measuring diversity

Relevance measures

- Normalized relevance

$$rel(S) = \frac{\sum_{v \in S} \pi_v}{\sum_{i=1}^{k} \hat{\pi}_i}$$

- Difference ratio

$$diff(S, \hat{S}) = 1 - \frac{|S \cap \hat{S}|}{|S|}$$

- nDCG

$$\mathrm{nDCG}_k = \frac{\pi_{s_1} + \sum_{i=2}^{k} \frac{\pi_{s_i}}{\log_2 i}}{\hat{\pi}_1 + \sum_{i=2}^{k} \frac{\hat{\pi}_i}{\log_2 i}}$$

Diversity measures

- l-step graph density

$$dens_\ell(S) = \frac{\sum_{u,v \in S, u \neq v} d_\ell(u, v)}{|S| \times (|S| - 1)}$$

- l-expansion ratio

$$\sigma_\ell(S) = \frac{|N_\ell(S)|}{n}$$

where

$$N_\ell(S) = S \cup \{v \in (V - S) : \exists u \in S, d(u, v) \leq \ell\}$$

Bicriteria optimization measures

- aggregate a relevance and a diversity measure
- [Carbonell98]

$$f_{MMR}(S) = (1-\lambda)\sum_{v \in S} \pi_v - \lambda \sum_{u \in S} \max_{\substack{v \in S \\ u \neq v}} sim(u,v)$$

- [Li11]

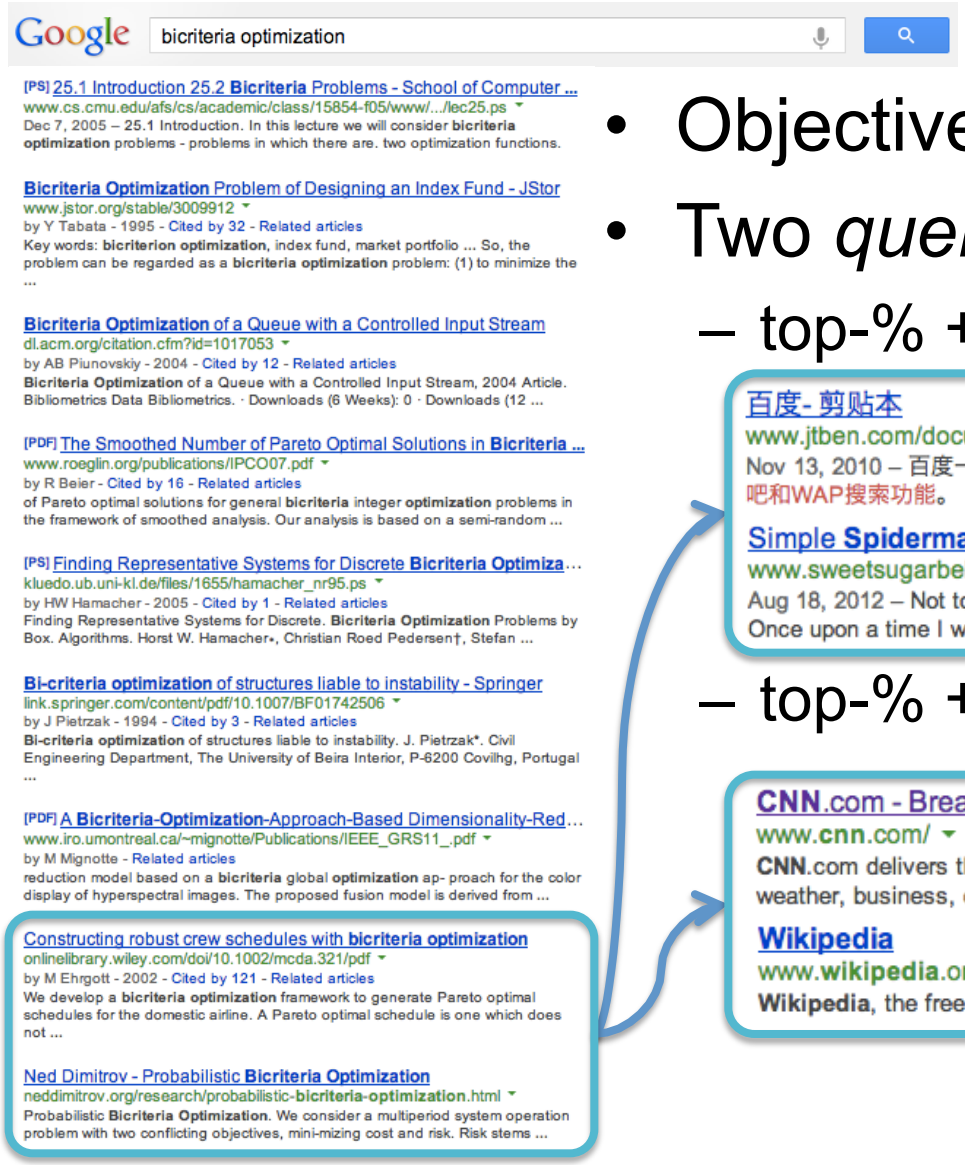
$$f_L(S) = \sum_{v \in S} \pi_v + \lambda \frac{|N(S)|}{n}$$

- [Vieira11]

$$f_{MSD}(S) = (k-1)(1-\lambda)\sum_{v \in S} \pi_v + 2\lambda \sum_{u \in S} \sum_{\substack{v \in S \\ u \neq v}} div(u,v)$$

- max-sum diversification, max-min diversification, k-similar diversification set, etc. [Gollapudi09]

Bicriteria optimization is not the answer

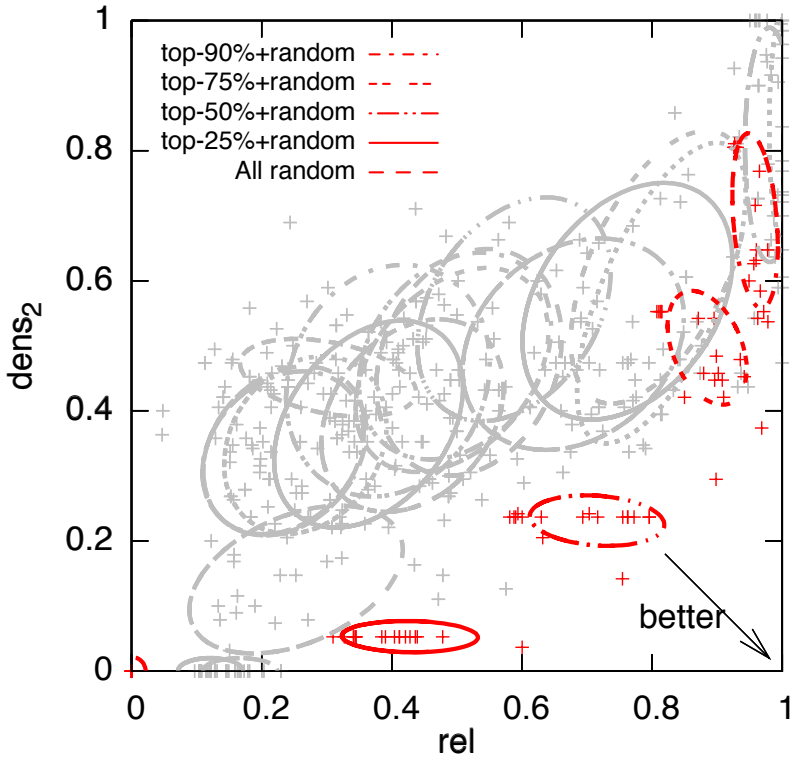


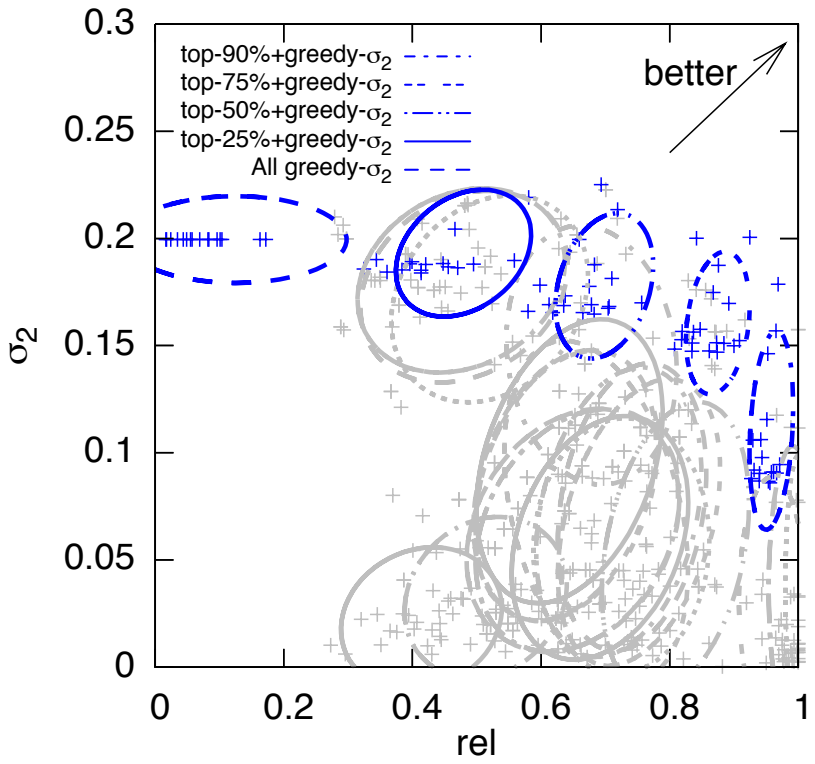
- Objective: diversify top-10 results
- Two *query-oblivious* algorithms:
  - top-% + random

  - top-% + greedy-σ_2

Bicriteria optimization is not the answer

- **normalized relevance** and **2-step graph density**





- evaluating result diversification as a bicriteria optimization problem with
  - a **relevance** measure that ignores **diversity**, and
  - a **diversity** measure that ignores **relevancy**.

A better measure? Combine both

- We need a combined measure that tightly integrates **both** *relevance* and *diversity* aspects of the result set
- *goodness* [Tong11]

penalize the score when two results share an edge

$$f_G(S) = \underbrace{2\sum_{i \in S} \pi_i}_{\text{max-sum relevance}} - \underbrace{d \sum_{i,j \in S} \mathbf{A}(j,i)\pi_j}_{\text{penalize the score when two results share an edge}} - (1-d)\sum_{j \in S} \pi_j \sum_{i \in S} p^*(i)$$

  - downside: highly dominated by relevance

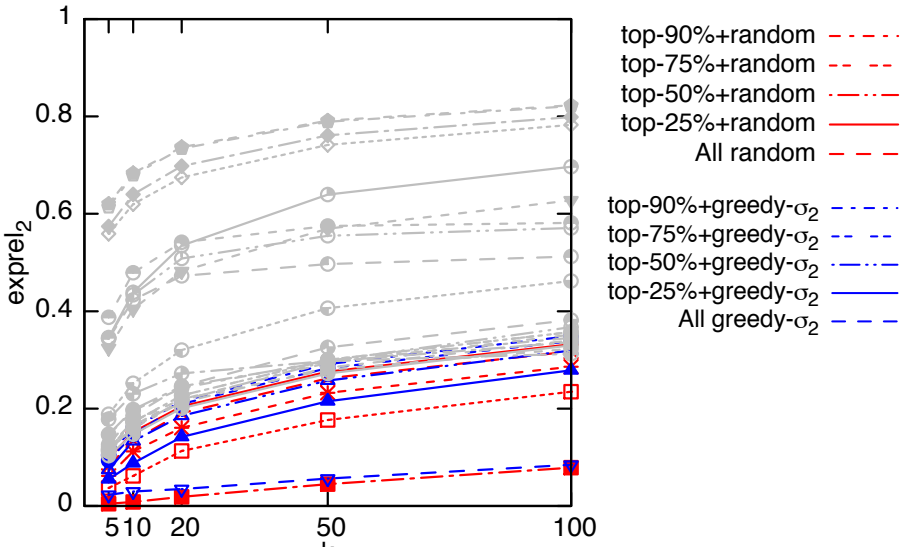
Proposed measure: *l*-step expanded relevance

- a combined measure of
  - *l*-step expansion ratio (σ_2)
  - relevance scores (π)
- quantifies: **relevance** of the *covered region* of the graph
- do some sanity check with this new measure

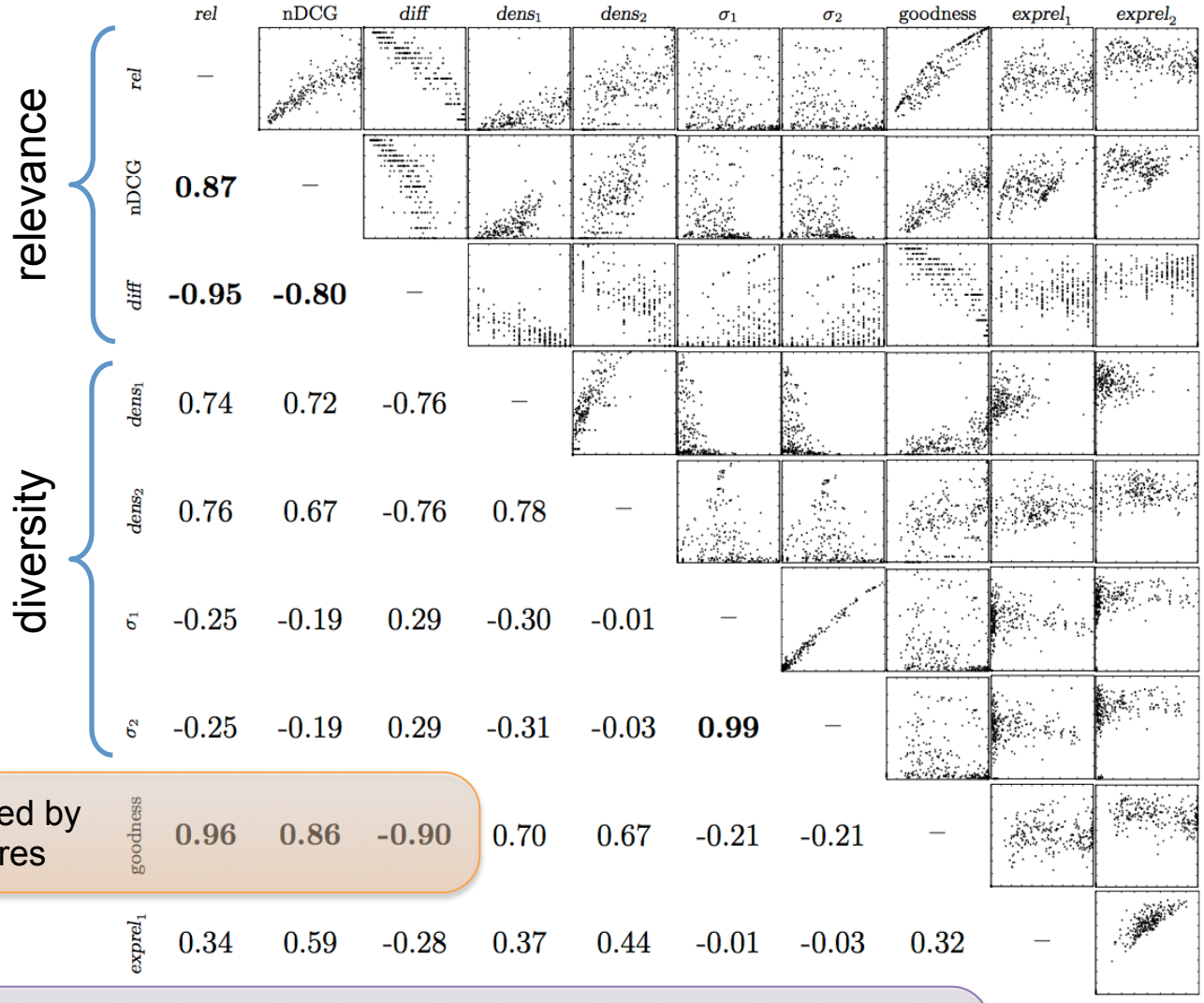
ℓ-step expanded relevance:

$$exprel_\ell(S) = \sum_{v \in N_\ell(S)} \pi_v$$

where $N_\ell(S)$ is the ℓ-step expansion set of the result set S, and π is the PPR scores of the items in the graph.



Correlations of the measures



	rel	nDCG	diff	$dens_1$	$dens_2$	σ_1	σ_2	goodness	$exprel_1$	$exprel_2$
relevance	rel	–								
	nDCG	**0.87**	–							
	diff	**-0.95**	**-0.80**	–						
diversity	$dens_1$	0.74	0.72	-0.76	–					
	$dens_2$	0.76	0.67	-0.76	0.78	–				
	σ_1	-0.25	-0.19	0.29	-0.30	-0.01	–			
	σ_2	-0.25	-0.19	0.29	-0.31	-0.03	**0.99**	–		
	goodness	**0.96**	**0.86**	**-0.90**	0.70	0.67	-0.21	-0.21	–	
	$exprel_1$	0.34	0.59	-0.28	0.37	0.44	-0.01	-0.03	0.32	–
	$exprel_2$	0.20	0.37	-0.13	0.17	0.33	0.26	0.23	0.21	**0.86**	–

goodness is dominated by the relevancy measures

exprel has no high correlations with other relevance or diversity measures

Proposed algorithm: **Best Coverage**

- Can we use ℓ**-step expanded relevance** as an objective function?
- Define: $exprel_\ell$-*diversified top-k ranking* $(\mathrm{DTR}\ell)$

$$S = \operatorname*{argmax}_{\substack{S' \subseteq V \\ |S'|=k}} exprel_\ell(S')$$

ALGORITHM 1: `BestCoverage`

Input: k, G, π, ℓ
Output: a list of recommendations S
$S = \emptyset$
while $|S| < k$ **do**
  $v^* \leftarrow \operatorname{argmax}_v g(v, S)$
  $S \leftarrow S \cup \{v^*\}$
return S

- **Complexity:** generalization of *weighted maximum coverage problem*
  - NP-hard!
  - but $exprel_l$ is a submodular function (Lemma 4.2)
  - a greedy solution (Algorithm 1) that selects the item with the *highest marginal utility*

    $g(v, S) = \sum_{v' \in N_\ell(\{v\}) - N_\ell(S)} \pi_{v'}$

    at each step is the best possible polynomial time approximation (proof based on [Nemhauser78])

- **Relaxation:** computes BestCoverage on highest ranked vertices to improve runtime

ALGORITHM 2: `BestCoverage (relaxed)`

Input: k, G, π, ℓ
Output: a list of recommendations S
$S = \emptyset$
$\textsc{Sort}(V)$ w.r.t π_i non-increasing
$S1 \leftarrow V[1..k']$, i.e., top-k' vertices where $k' = k\bar{\delta}^\ell$
$\forall v \in S1, g(v) \leftarrow g(v, \emptyset)$
$\forall v \in S1, c(v) \leftarrow \textsc{Uncovered}$
while $|S| < k$ **do**
  $v^* \leftarrow \operatorname{argmax}_{v \in S1} g(v)$
  $S \leftarrow S \cup \{v^*\}$
  $S2 \leftarrow N_\ell(\{v^*\})$
  for each $v' \in S2$ **do**
    if $c(v') = \textsc{Uncovered}$ **then**
      $S3 \leftarrow N_\ell(\{v'\})$
      $\forall u \in S3, g(u) \leftarrow g(u) - \pi_{v'}$
      $c(v') \leftarrow \textsc{Covered}$
return S

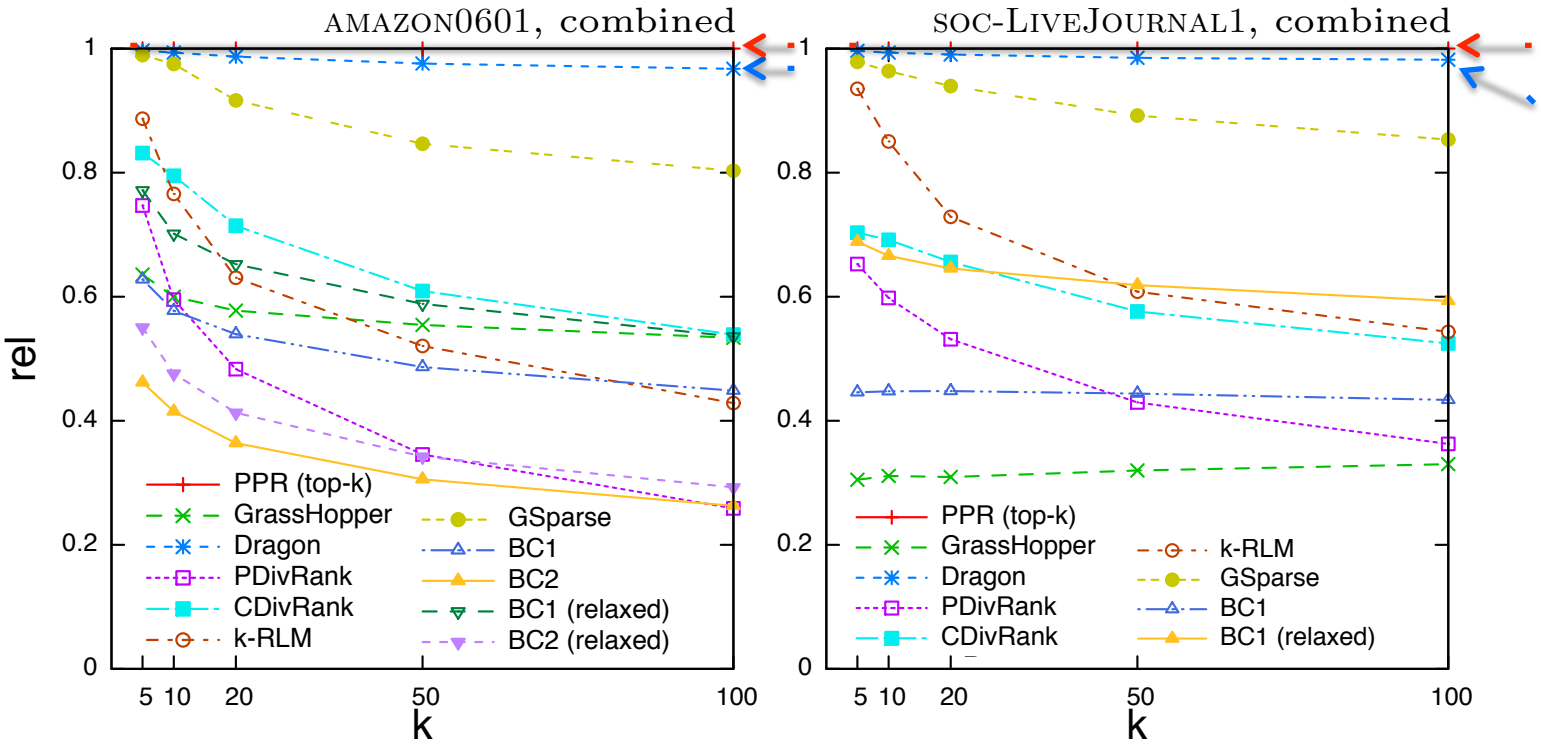
Experiments

- 5 target application areas, 5 graphs from SNAP

Dataset	$\|V\|$	$\|E\|$	$\bar{\delta}$	D	$D_{90\%}$	CC
AMAZON0601	403.3K	3.3M	16.8	21	7.6	0.42
CA-ASTROPH	18.7K	396.1K	42.2	14	5.1	0.63
CIT-PATENTS	3.7M	16.5M	8.7	22	9.4	0.09
SOC-LIVEJOURNAL1	4.8M	68.9M	28.4	18	6.5	0.31
WEB-GOOGLE	875.7K	5.1M	11.6	22	8.1	0.60

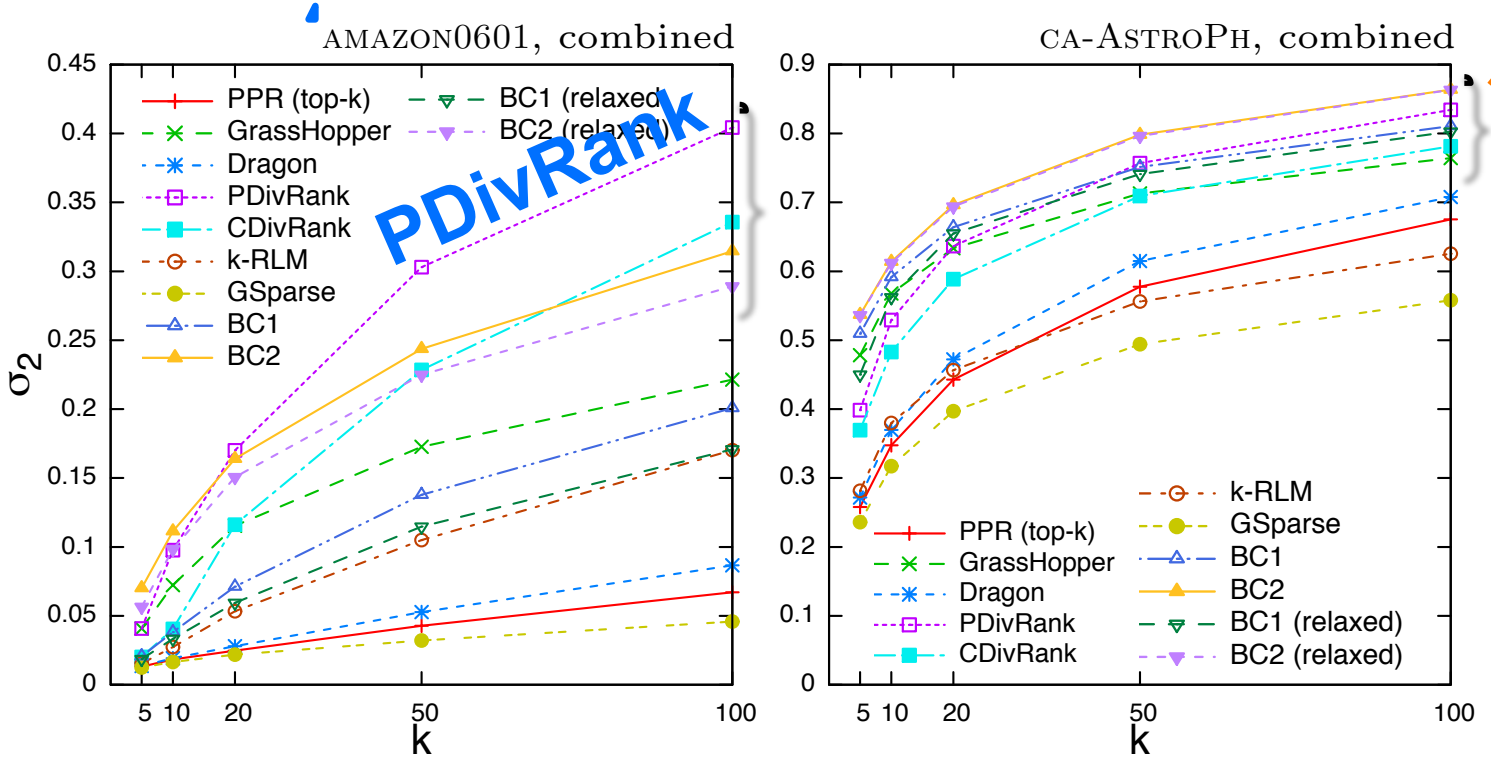
- Queries generated based on 3 scenario types
  - one random vertex
  - random vertices from one area of interest
  - multiple vertices from multiple areas of interest

Results – relevance



- Methods should trade-off relevance for better diversity
- Normalized relevance of top-*k* set is always 1
- DRAGON always return results having 70% similar items to top-*k*, with more than 80% *rel* score

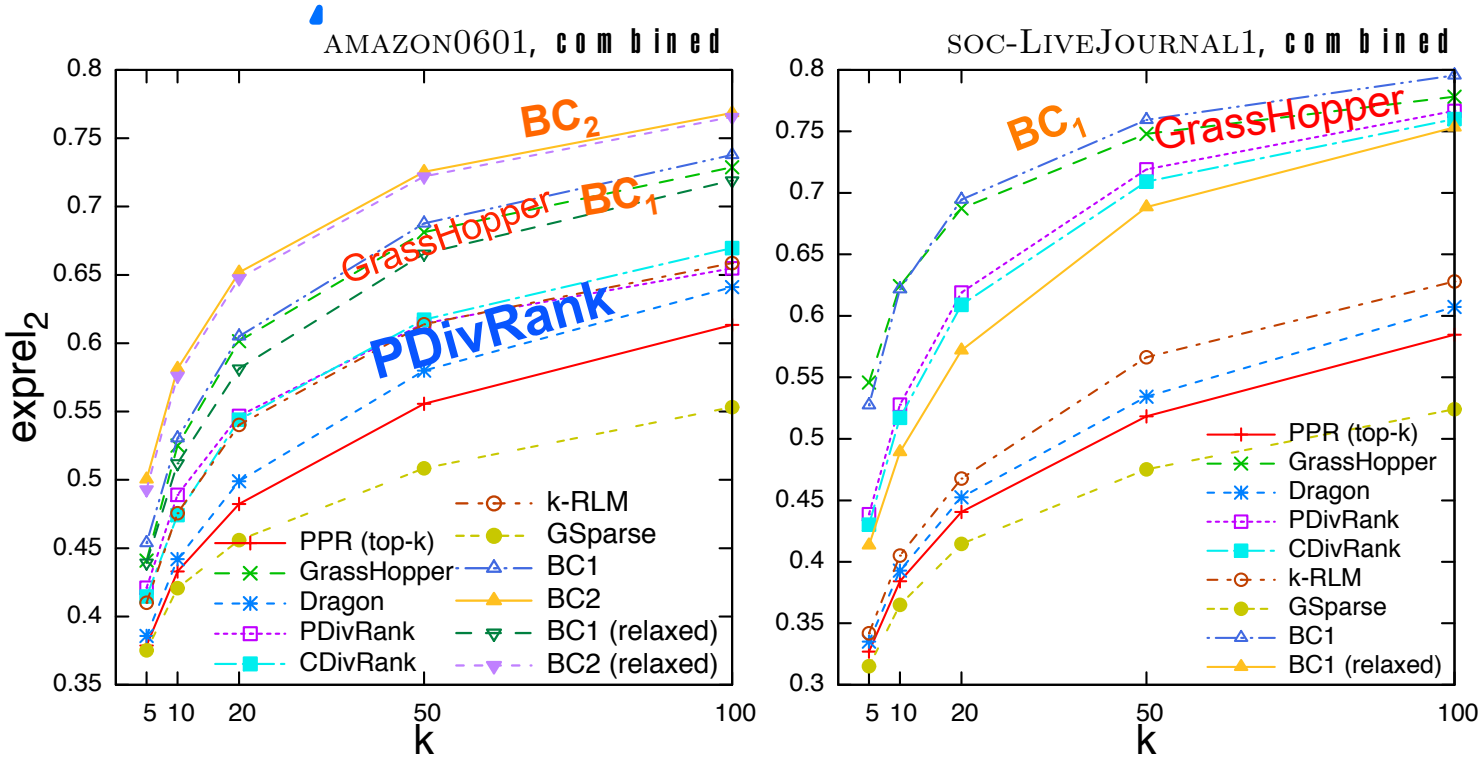
Results – coverage



BestCoverage and DivRank variants

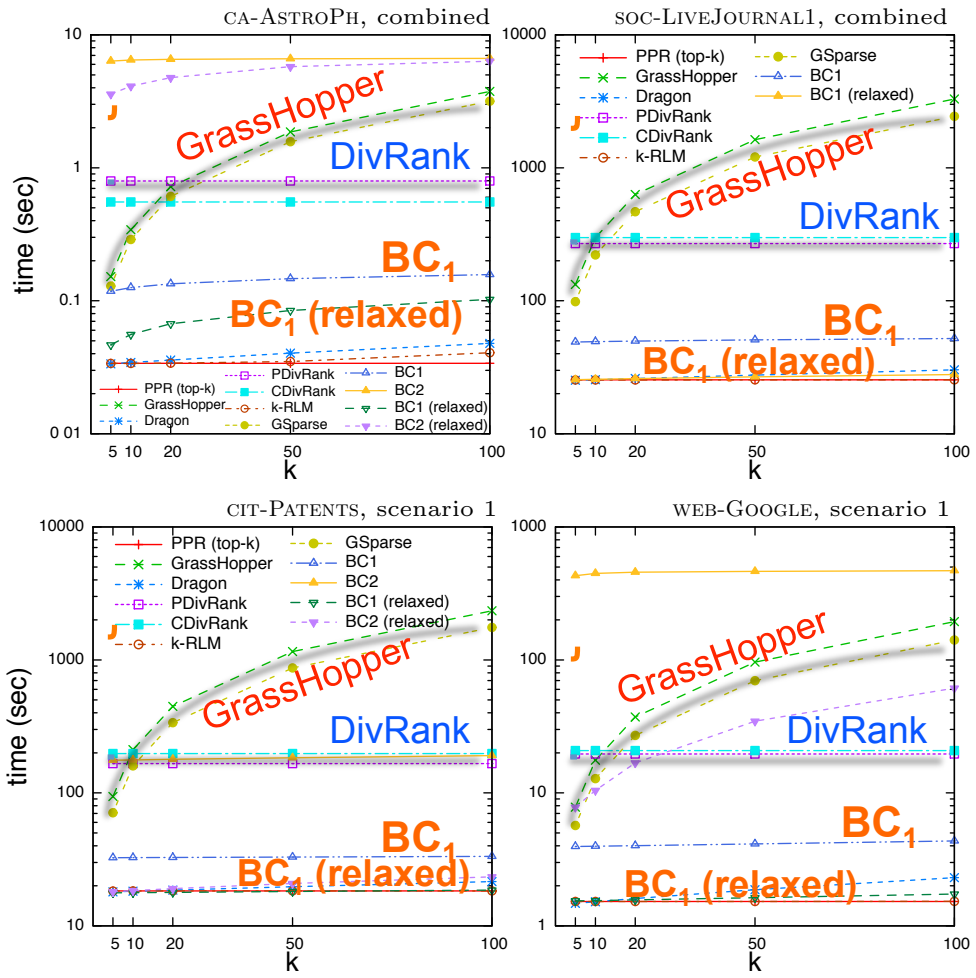
- *l*-step expansion ratio (σ_2) gives the *graph coverage* of the result set: *better coverage = better diversity*
- BestCoverage and DivRank variants, especially **BC_2** and **PDivRank**, have the highest coverage

Results – expanded relevance



- combined measure for relevance and diversity
- BestCoverage variants and GrassHopper perform better
- Although **PDivRank** gives the highest coverage on amazon graph, it fails to cover the relevant parts!

Results – efficiency



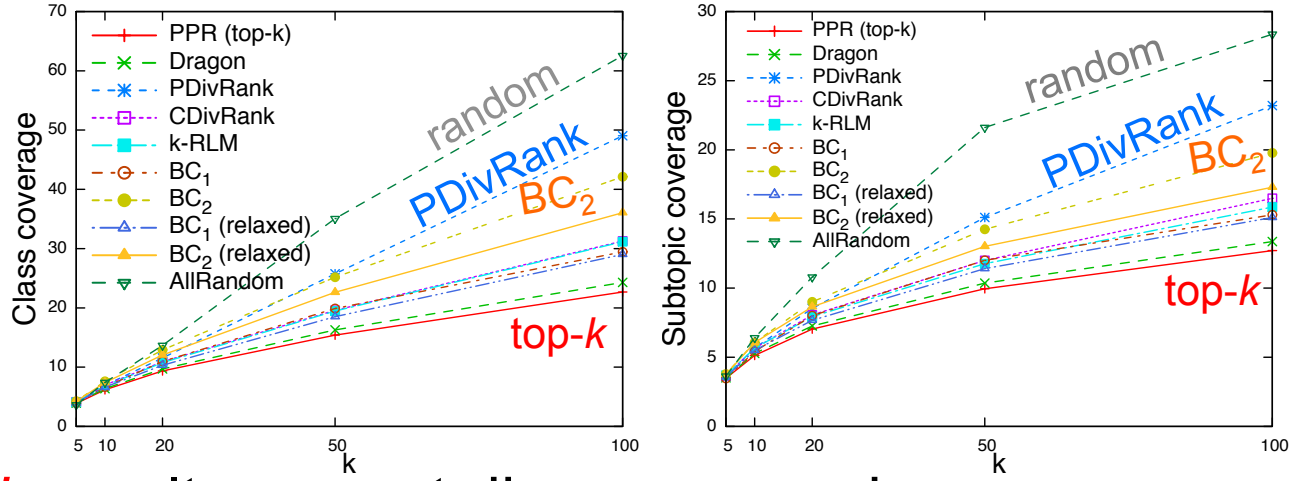
- BC_1 always performs better, with a running time less than, DivRank and GrassHopper
- BC_1 (relaxed) offers reasonable diversity, with a very little overhead on top of the PPR computation

Results – intent aware experiments

- evaluation of *intent-oblivious* algorithms against *intent-aware* measures
- two measures
  - group coverage [Li11]
  - *S-recall* [Zhai03]
- cit-Patent dataset has the categorical information
  - 426 class labels, belong to 36 subtopics

Results – intent aware experiments

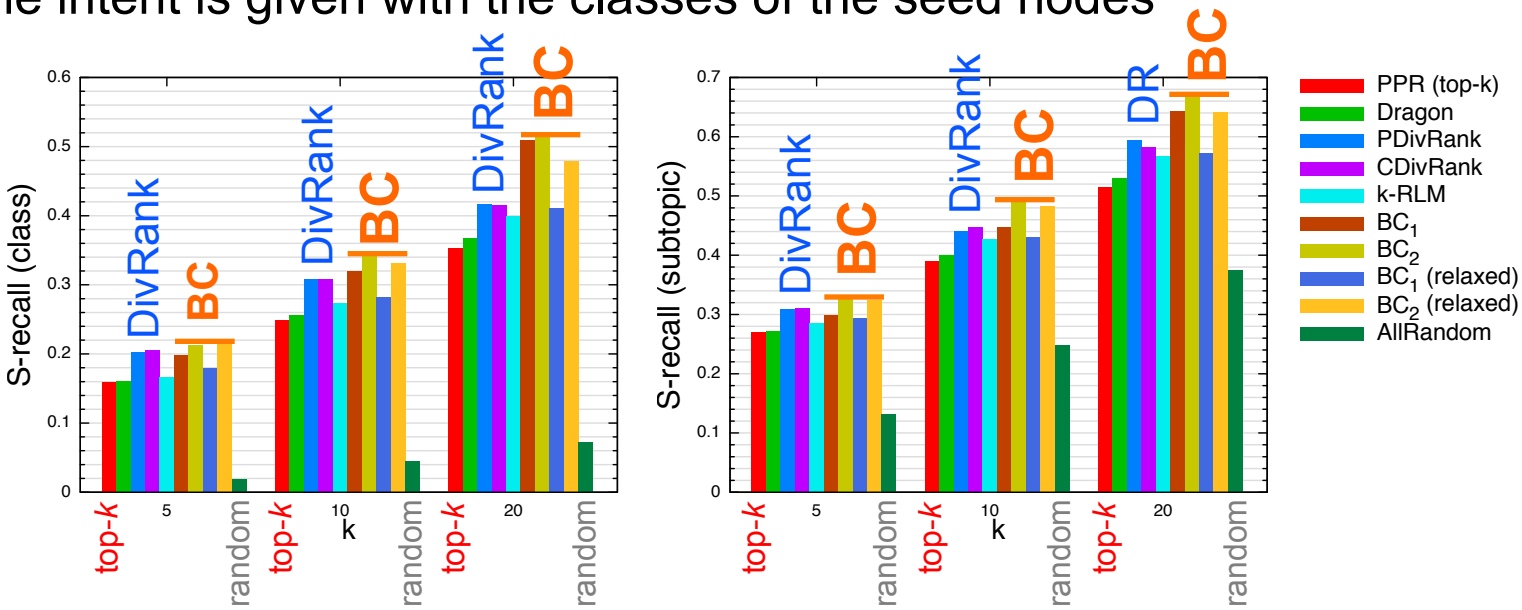
- group coverage [Li11]
  - How many different groups are covered by the results?
  - omits the actual intent of the query



- top-*k* results are not diverse enough
- AllRandom results cover the most number of groups
- PDivRank and BC_2 follows

Results – intent aware experiments

- *S-recall* [Zhai03], *Intent-coverage* [Zhu11]
  - percentage of relevant subtopics covered by the result set
  - the intent is given with the classes of the seed nodes



- AllRandom brings irrelevant items from the search space
- top-*k* results do not have the necessary diversity
- BC_2 variants and BC_1 perform better than DivRank
- BC_1 (relaxed) and DivRank scores similar, but BC_1r much faster

Conclusions

- Result diversification should not be evaluated as a bicriteria optimization problem with
  - a **relevance** measure that ignores **diversity**, and
  - a **diversity** measure that ignores **relevancy**
- ***l*-step expanded relevance** is a simple measure that combines both relevance and diversity
- **BestCoverage**, a greedy solution that maximizes $exprel_l$ is a (1-1/e)-approximation of the optimal solution
- BestCoverage variants perform better than others, its relaxation is extremely efficient
- *goodness* in DRAGON is dominated by relevancy
- DivRank variants implicitly optimize expansion ratio

Thank you

- For more information visit
  - http://bmi.osu.edu/hpc
- Research at the HPC Lab is funded by