

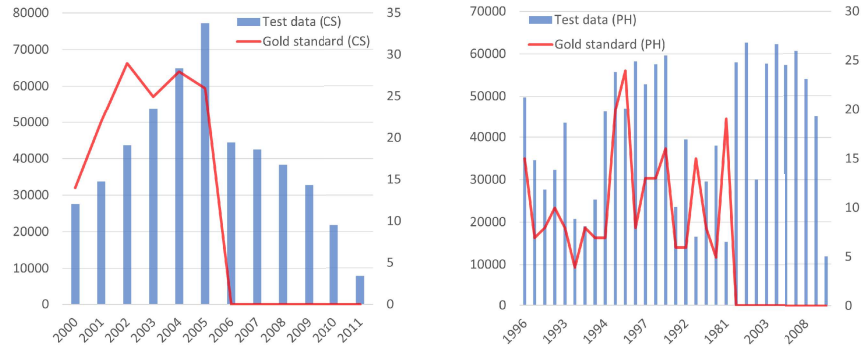
Appendix A Supplementary materials

Table A1: Selected computer science conferences and the number of papers awarded for long lasting impact in the test data (CS) set.

Conference	#P	Conference	#P	Conference	#P	Conference	#P	Conference	#P
SIGIR	13	SIGMETRICS	6	SIGMOD	4	ISCA	3	KDD	1
SIGCOMM	13	ICML	5	CVPR	3	PLDI	3	NIPS	1
ICCV	10	ICFP	5	EMNLP	3	INFOCOM	2	NSDI	1
NAACL	10	POPL	5	OSDI	3	UIST	2	SOSP	1
ICSE	10	AAAI	4	SIGPLAN	3	ISSTA	2	SIGSOFT	1
ASE	7	GECCO	4	STOC	3	ACL	1		
FAST	6	OOPSLA	4	VLDB	3	FOCS	1	<i>In total</i>	143

Table A2: High impact papers recognised in the test data (PH) set.

Recognition	#P
Nobel Prize	43
PRL Milestone	12
Nobel Reference	191
<i>In total</i>	246



(a) Test data (CS), historical time 2005 (b) Test data (PH), historical time 2001

Fig. A1: Distribution of the gold standard papers and overall papers by year in test data sets.

Table A3: Spearman's rank correlation between citation count-based benchmarks.*Results on the arXiv dataset*

	HCC	FCC	wFCC	HCC_t	FCC_t	wFCC_t
HCC	1	0.46(0.44,0.47)	0.42(0.41,0.44)	0.93(0.92,0.93)	0.24(0.22,0.26)	0.23(0.21,0.24)
FCC		1	0.97(0.97,0.97)	0.55(0.53,0.56)	0.91(0.91,0.92)	0.89(0.89,0.89)
wFCC			1	0.52(0.51,0.54)	0.9(0.9,0.9)	0.92(0.92,0.93)
HCC_t				1	0.46(0.44,0.47)	0.44(0.42,0.45)
FCC_t					1	0.98(0.98,0.98)
wFCC_t						1

Results on the MAG (ID) dataset

	HCC	FCC	wFCC	HCC_t	FCC_t	wFCC_t
HCC	1	0.53(0.49,0.57)	0.47(0.42,0.51)	0.95(0.95,0.96)	0.1(0.05,0.15)	0.06(0.0,0.11)
FCC		1	0.94(0.93,0.94)	0.58(0.54,0.62)	0.81(0.79,0.83)	0.74(0.71,0.76)
wFCC			1	0.53(0.48,0.56)	0.79(0.77,0.81)	0.83(0.81,0.85)
HCC_t				1	0.26(0.21,0.31)	0.22(0.16,0.27)
FCC_t					1	0.94(0.94,0.95)
wFCC_t						1

Results on test data (CS)

	HCC	HCC_t	FCC	FCC_t
HCC	1	0.97(0.97,0.97)	0.6(0.6,0.6)	0.39(0.39,0.4)
HCC_t		1	0.62(0.62,0.62)	0.48(0.48,0.49)
FCC			1	0.92(0.92,0.92)
FCC_t				1

Results on test data (PH)

	HCC	HCC_t	FCC	FCC_t
HCC	1	0.99(0.99,0.99)	0.34(0.34,0.34)	0.11(0.11,0.12)
HCC_t		1	0.33(0.33,0.34)	0.15(0.15,0.16)
FCC			1	0.67(0.67,0.68)
FCC_t				1

Table A4: Spearman's rank correlation between citation network-based benchmarks.*Results on the arXiv dataset*

	FPR	wFPR	FPR+A	FPR+V	FPR+T
FPR	1	0.99 (0.98,0.99)	0.93 (0.92,0.93)	0.71 (0.7,0.72)	0.89 (0.88,0.89)
wFPR		1	0.91 (0.91,0.92)	0.7 (0.69,0.71)	0.88 (0.88,0.89)
FPR+A			1	0.68 (0.67,0.69)	0.83 (0.82,0.83)
FPR+V				1	0.63 (0.62,0.65)
FPR+T					1

Results on the MAG (ID) dataset

	FPR	wFPR	FPR+A	FPR+V	FPR+T
FPR	1	0.98 (0.98,0.98)	0.93 (0.92,0.94)	0.89 (0.88,0.9)	0.93 (0.92,0.94)
wFPR		1	0.91 (0.9,0.92)	0.88 (0.86,0.89)	0.92 (0.92,0.93)
FPR+A			1	0.83 (0.81,0.85)	0.87 (0.85,0.88)
FPR+V				1	0.84 (0.82,0.85)
FPR+T					1

Results on test data (CS)

	FPR	FPR+A	FPR+V	FPR+T
FPR	1	0.94 (0.94,0.94)	0.88 (0.88,0.89)	0.79 (0.79,0.79)
FPR+A		1	0.83 (0.83,0.83)	0.78 (0.78,0.78)
FPR+V			1	0.7 (0.7,0.7)
FPR+T				1

Results on test data (PH)

	FPR	FPR+A	FPR+V	FPR+T
FPR	1	0.92 (0.91,0.92)	0.98 (0.98,0.98)	0.89 (0.89,0.89)
FPR+A		1	0.91 (0.91,0.91)	0.83 (0.83,0.83)
FPR+V			1	0.89 (0.89,0.89)
FPR+T				1

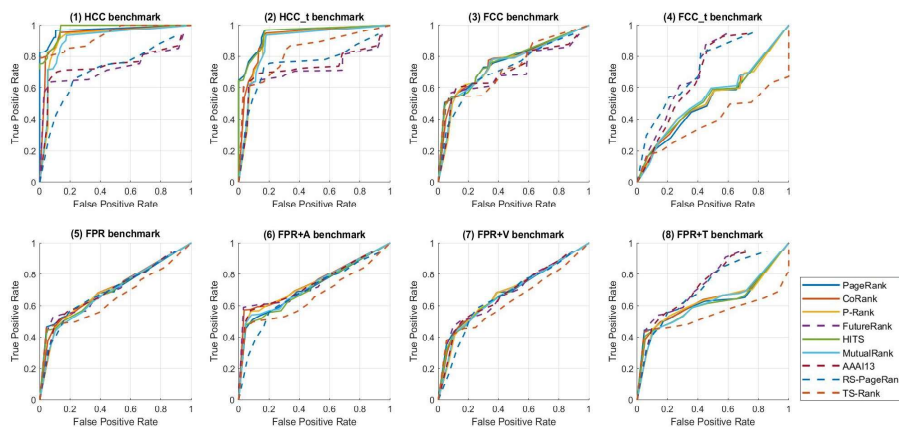


Fig. A2: ROC of ranking algorithms against different benchmarks. Results on MAG (ID) dataset.

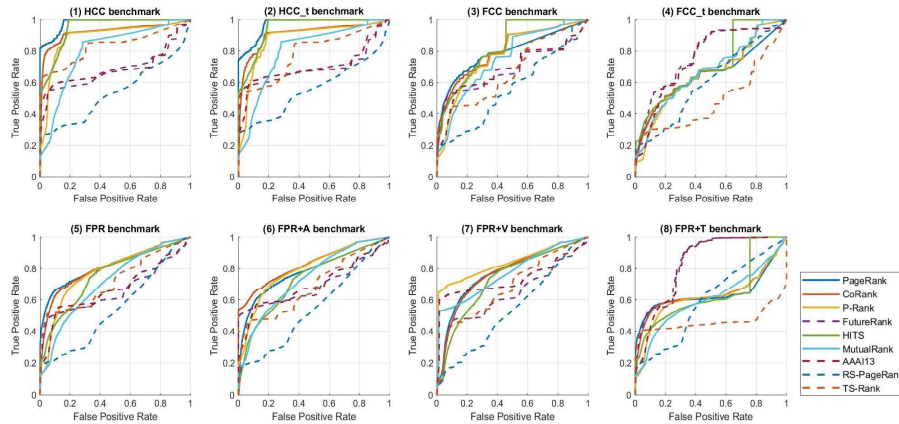


Fig. A3: ROC of ranking algorithms against different benchmarks. Results on test data (CS).

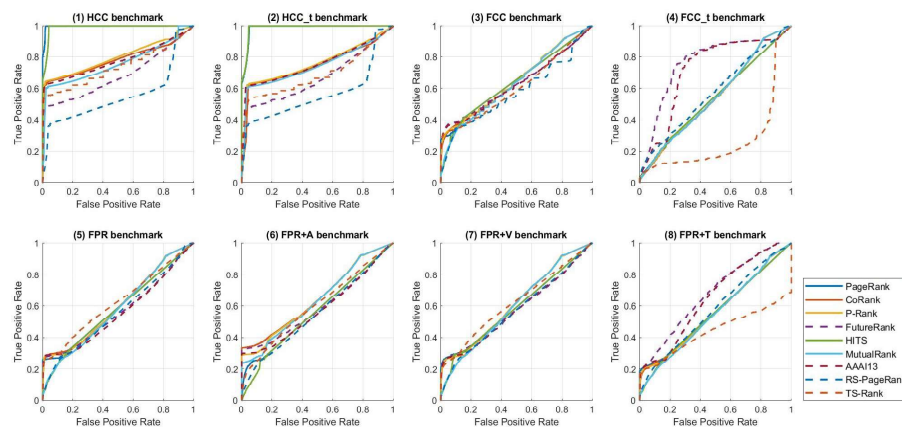


Fig. A4: ROC of ranking algorithms against different benchmarks. Results on test data (PH).