

# How Does Regression Test Prioritization Perform in Real-World Software Evolution?

Yafeng Lu<sup>†</sup>, Yiling Lou<sup>\*</sup>, Shiyang Cheng<sup>†</sup>, Lingming Zhang<sup>†</sup>, Dan Hao<sup>\*</sup>, Yangfan Zhou<sup>‡</sup>, Lu Zhang<sup>\*</sup>

<sup>†</sup>Department of Computer Science, The University of Texas at Dallas, USA

{yx131230,sxc145630,lingming.zhang}@utdallas.edu

<sup>\*</sup>Key Laboratory of High Confidence Software Technologies (Peking University), MoE, China

Institute of Software, School of EECS, Peking University, China

{louyiling,haodan,zhanglucs}@pku.edu.cn

<sup>‡</sup>School of Computer Science, Fudan University, China

zyf@fudan.edu.cn

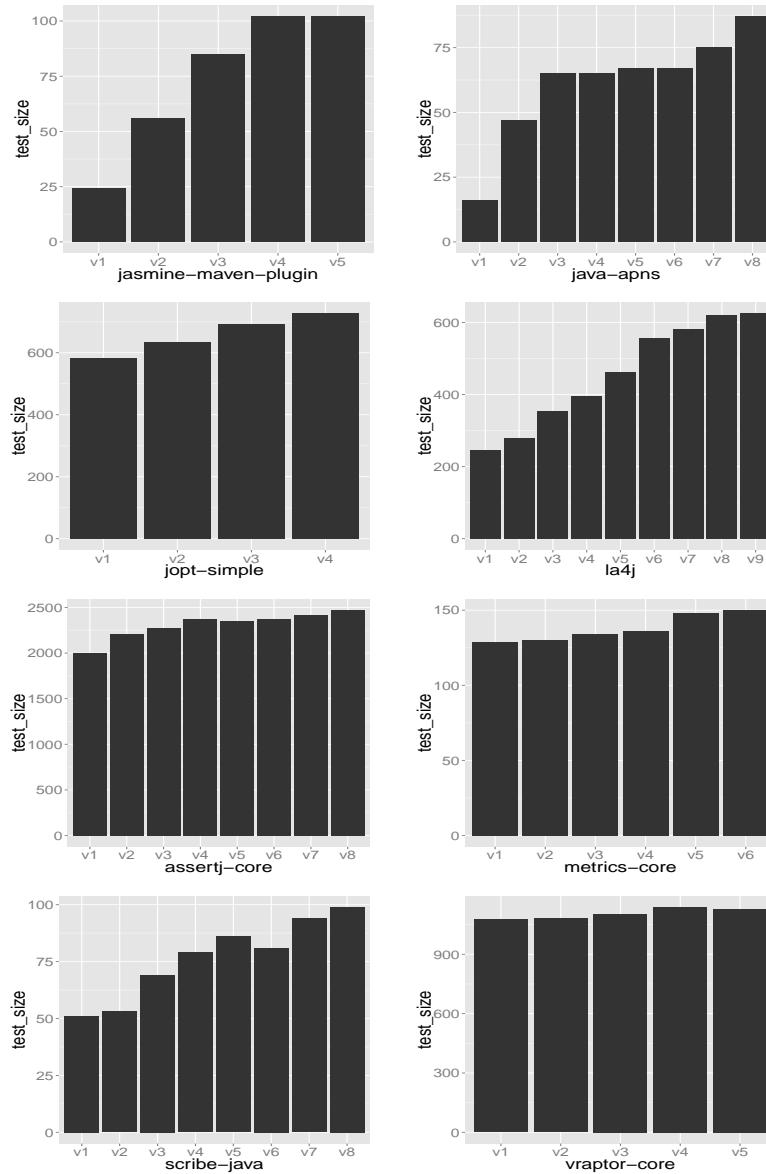


Fig. 1: Test size trend of each subject

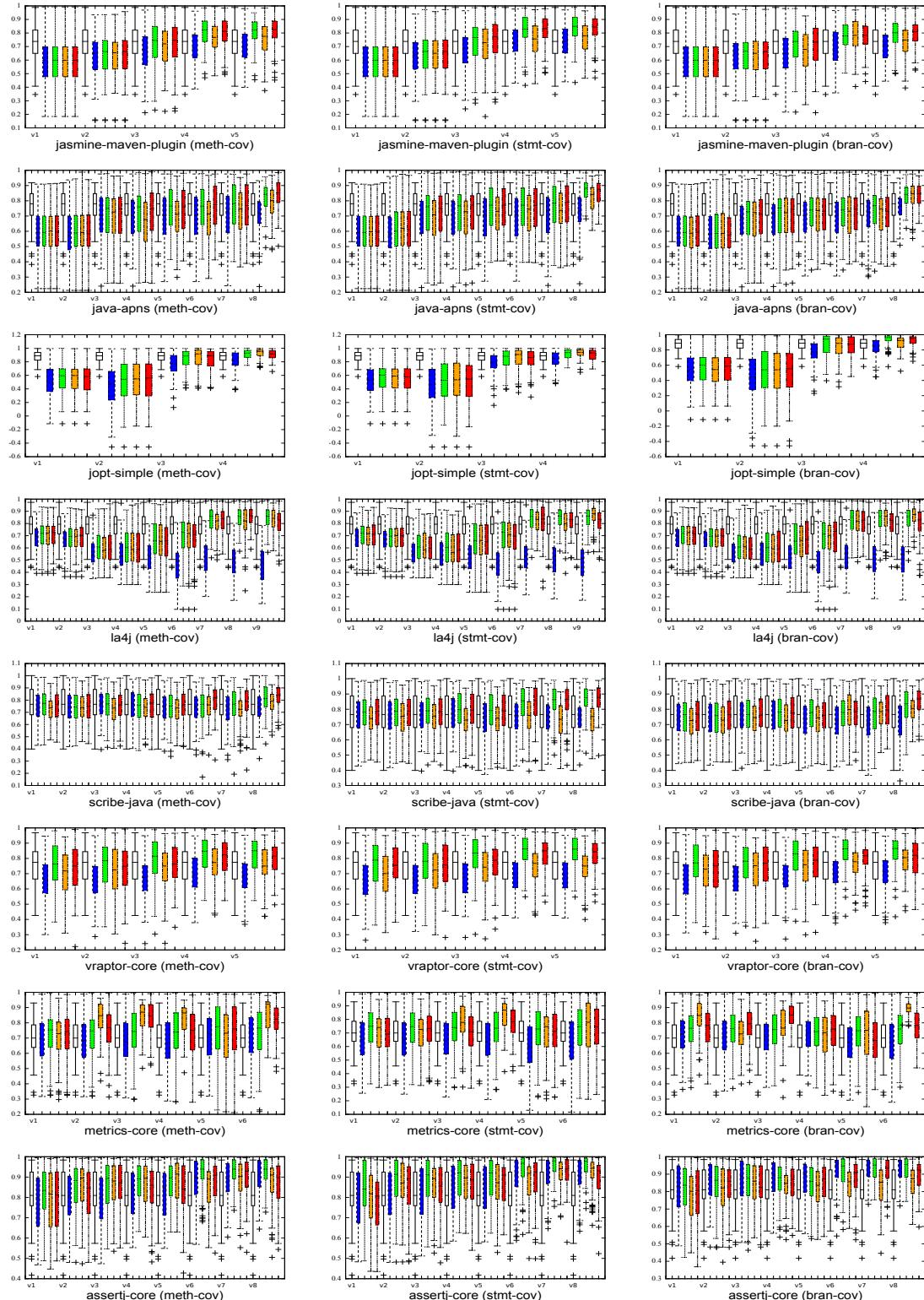


Fig. 2: Results for traditional test prioritization techniques based on Method, Statement and Branch coverage (with new tests, random ○, total ●, additional ●, ART ●, search based technique ●)

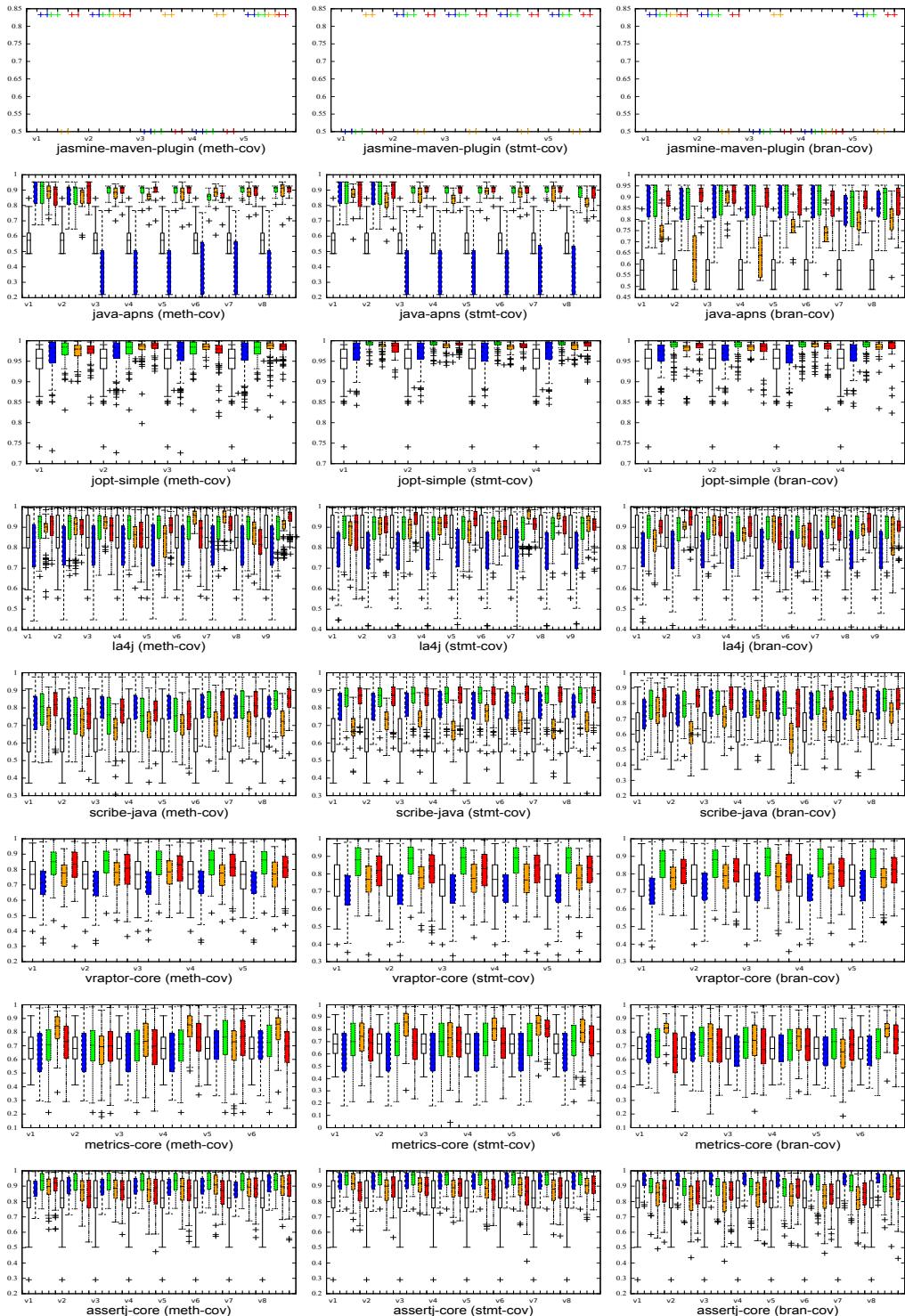


Fig. 3: Results for traditional test prioritization techniques based on Method, Statement and Branch coverage (excluding new tests, random  $\circlearrowleft$ , total  $\bullet$ , additional  $\bullet$ , ART  $\bullet$ , search based technique  $\bullet$ )

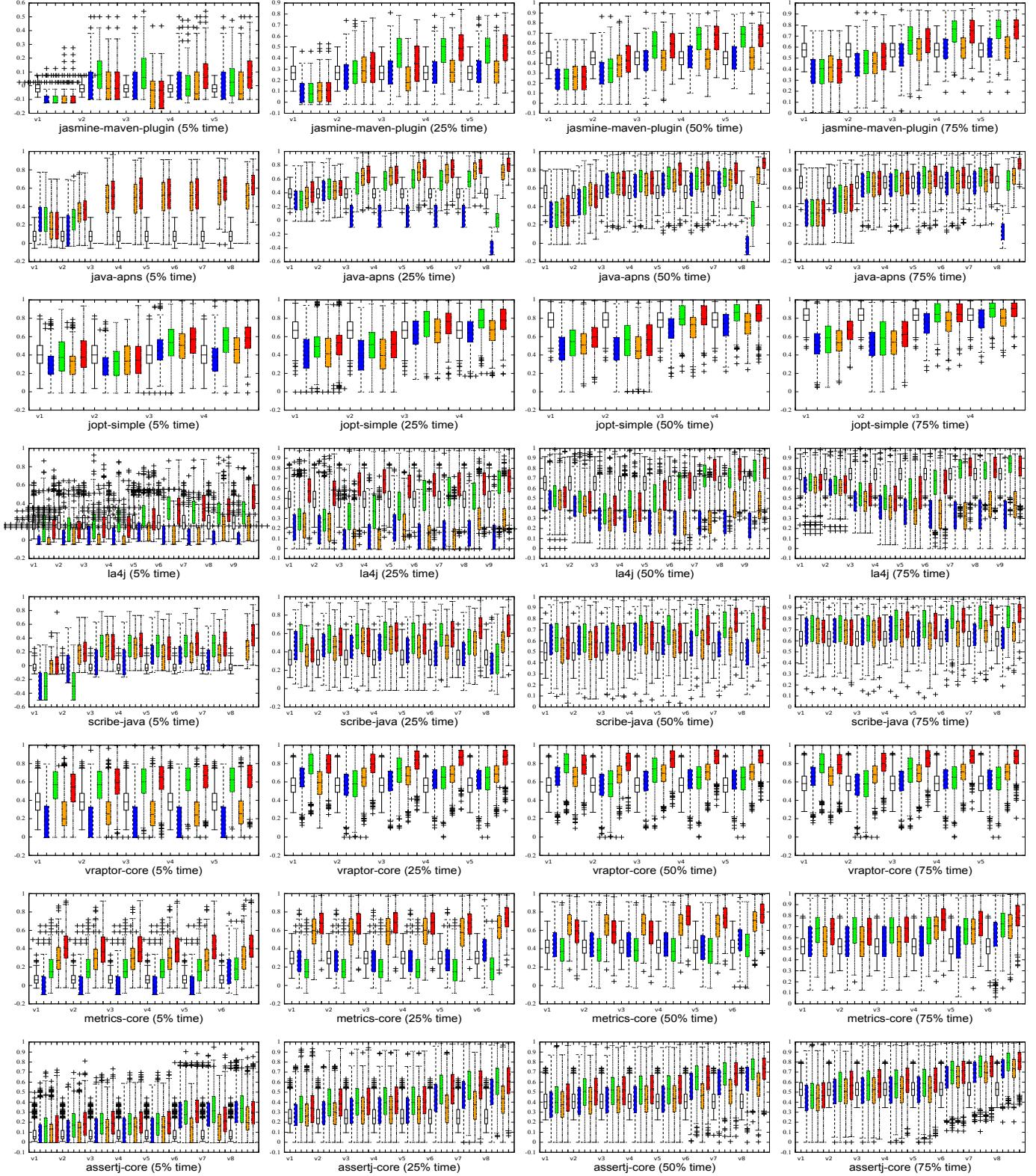


Fig. 4: Results for time-aware test prioritization techniques based on Statement coverage (with new tests, random ○, total ●, additional ●, ILP-total ●, ILP-additional ●)

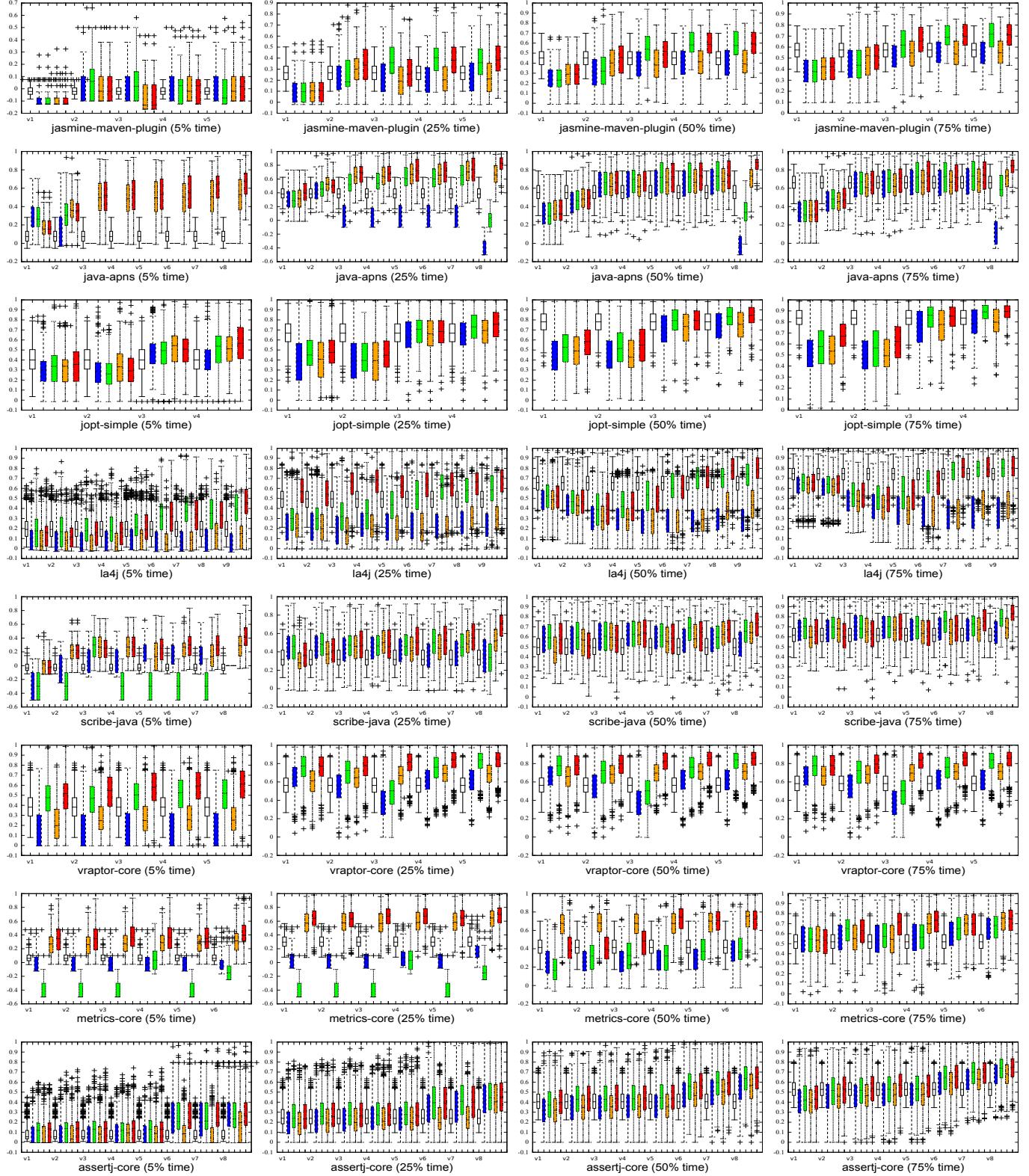


Fig. 5: Results for time-aware test prioritization techniques based on Method coverage (with new tests, random  $\circ$ , total  $\bullet$ , additional  $\bullet$ , ILP-total  $\bullet$ , ILP-additional  $\bullet$ )

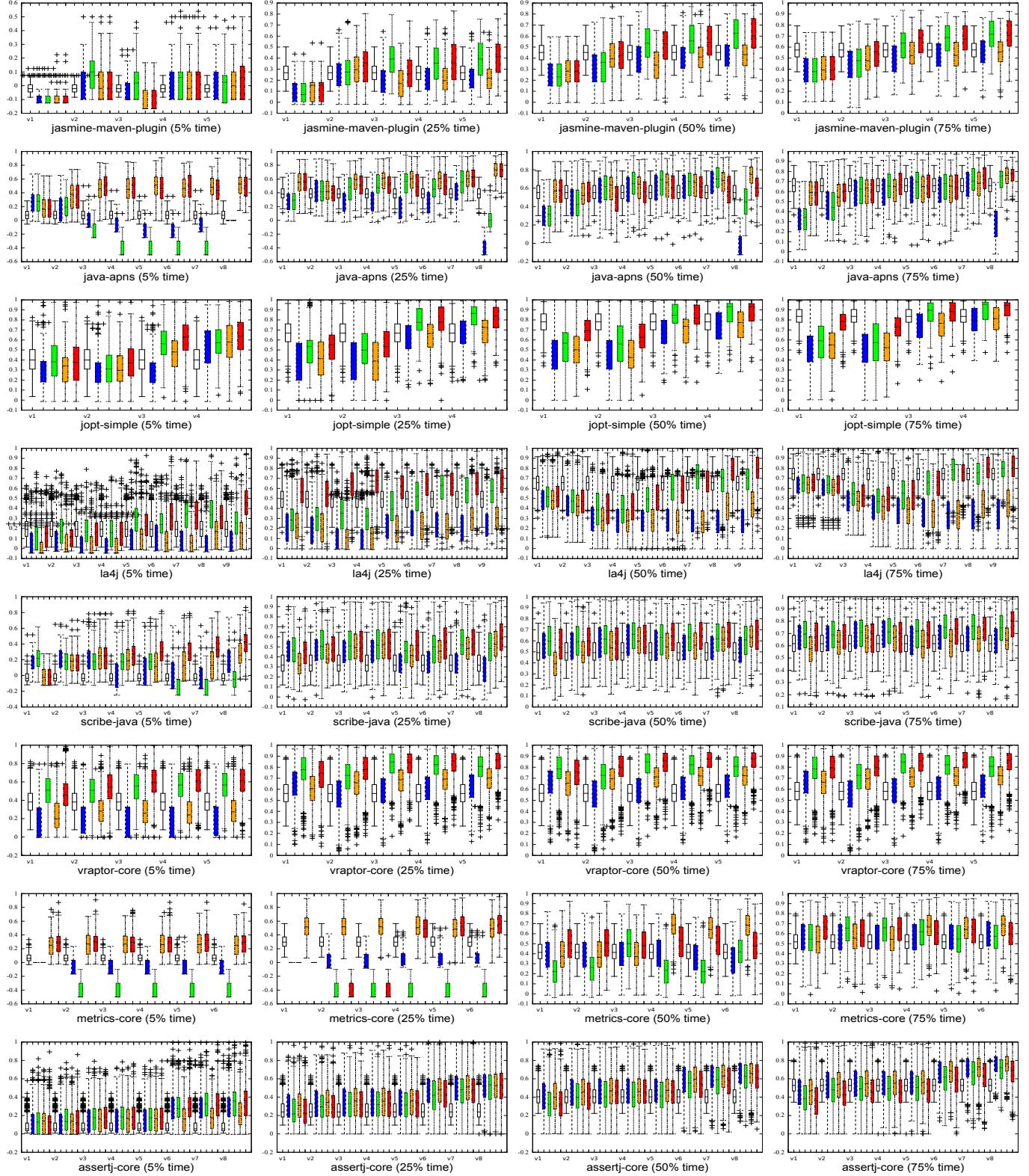


Fig. 6: Results for time-aware test prioritization techniques based on Branch coverage (with new tests, random  $\circ$ , total  $\bullet$ , additional  $\bullet$ , ILP-total  $\bullet$ , ILP-additional  $\bullet$ )

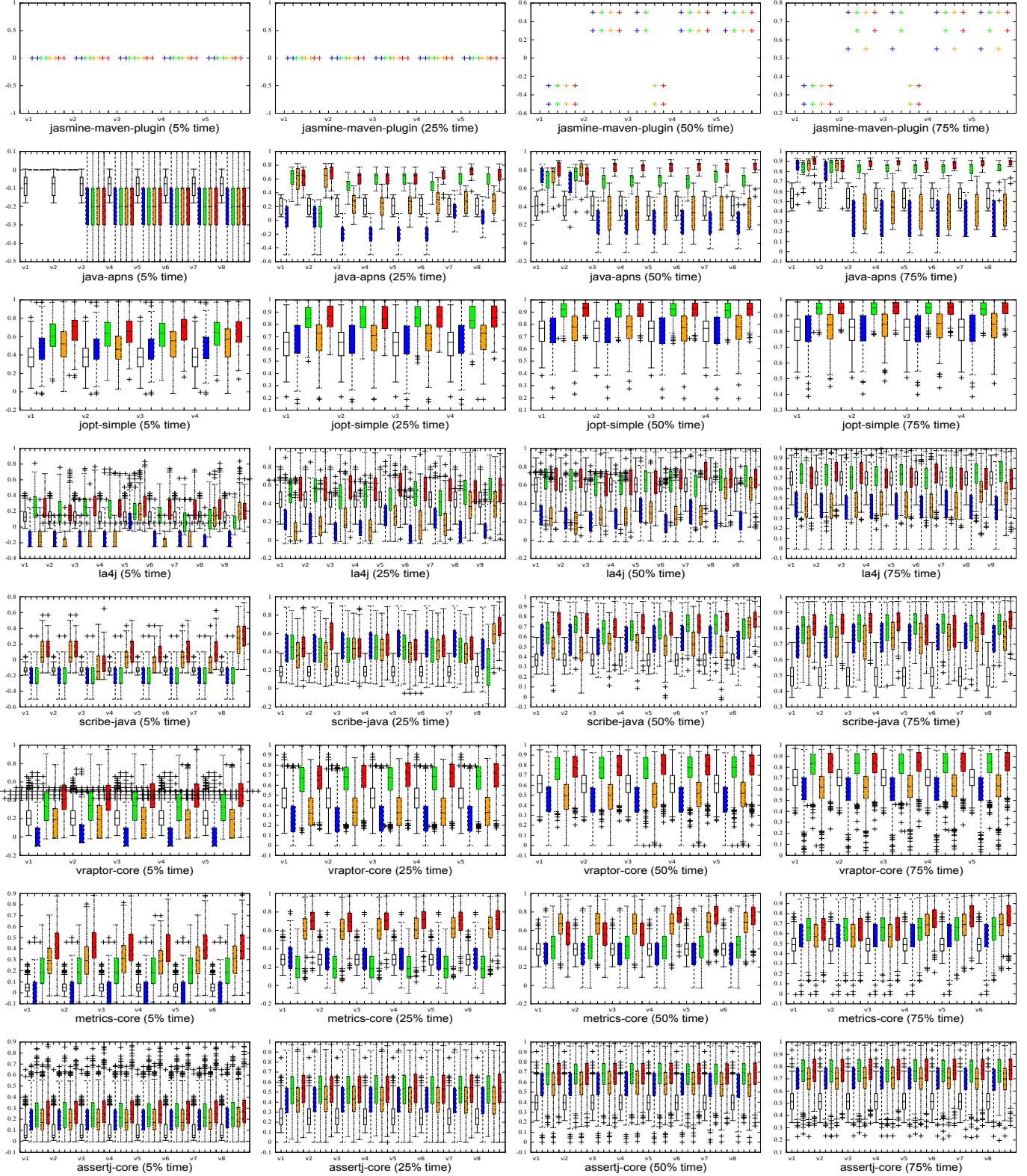


Fig. 7: Results for time-aware test prioritization techniques based on Statement coverage (excluding new tests, random ○, total ●, additional ○, ILP-total ○, ILP-additional ●)

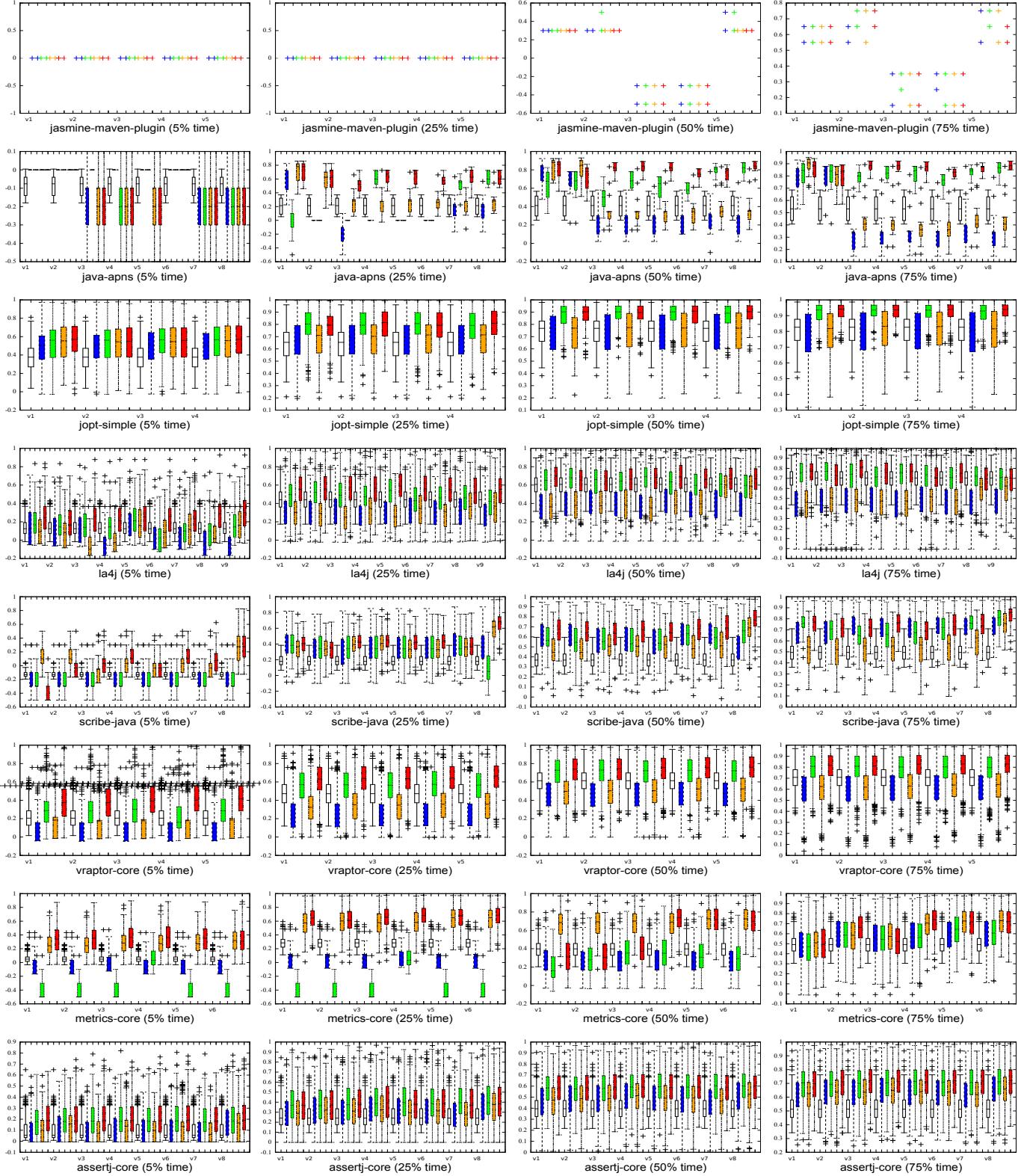


Fig. 8: Results for time-aware test prioritization techniques based on Method coverage (excluding new tests, random ○, total ●, additional ●, ILP-total ●, ILP-additional ●)

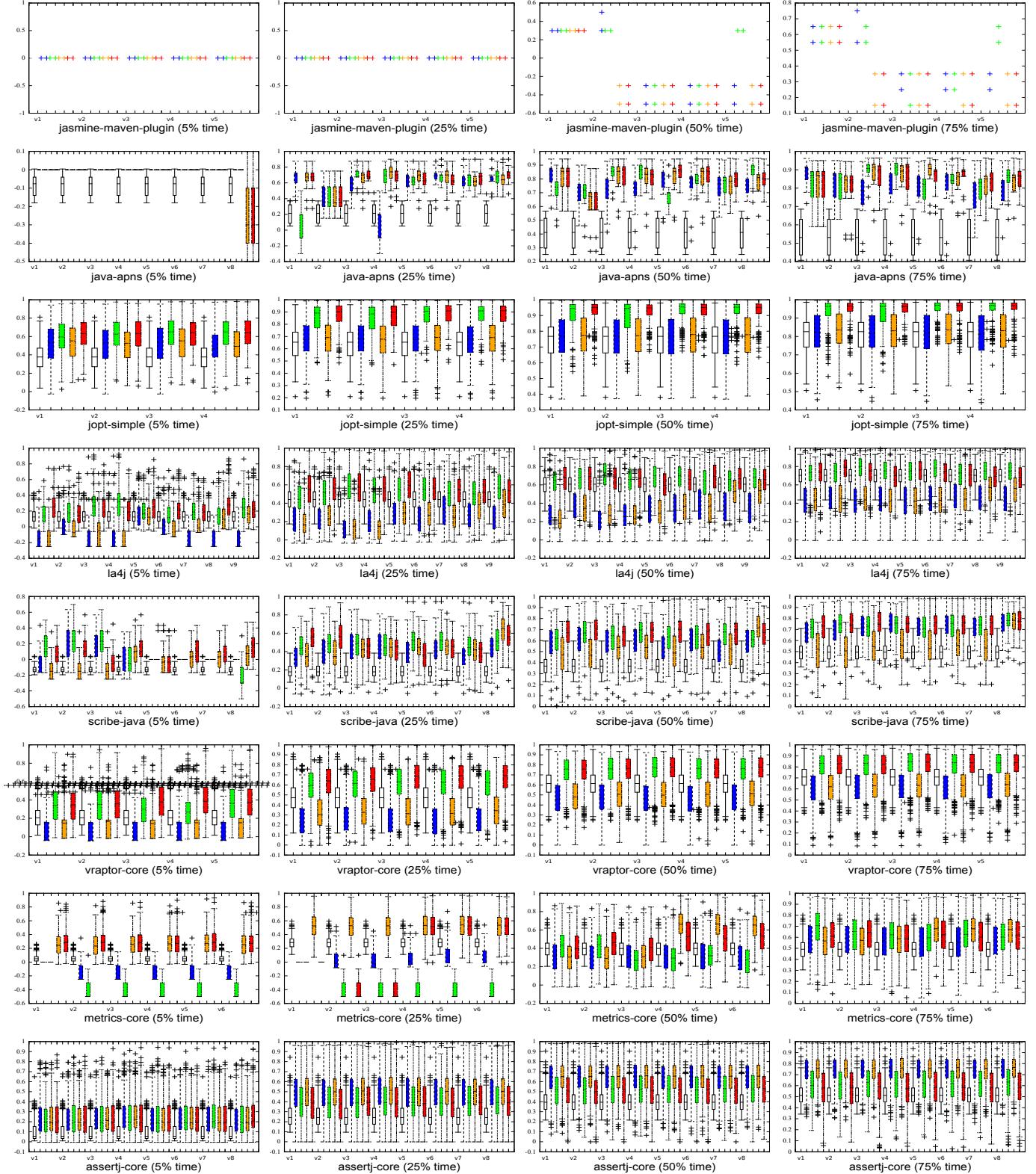


Fig. 9: Results for time-aware test prioritization techniques based on Branch coverage (excluding new tests, random  $\circ$ , total  $\bullet$ , additional  $\bullet$ , ILP-total  $\bullet$ , ILP-additional  $\bullet$ )



Tech	Sub	With new tests									Without new tests										
		p-value	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	v <sub>4</sub>	v <sub>5</sub>	v <sub>6</sub>	v <sub>7</sub>	v <sub>8</sub>	p-value	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	v <sub>4</sub>	v <sub>5</sub>	v <sub>6</sub>	v <sub>7</sub>	v <sub>8</sub>	v <sub>9</sub>	
Total	scribe-java	p<0.05	a	a	a	a	a	a	a	a	p<0.05	b	ab	a	a	a	a	a	a	-	
	java-apns	p<0.05	c	c	b	b	b	b	ab	a	0.988	a	a	a	a	a	a	a	a	-	
	jopt-simple	p<0.05	b	c	a	a	-	-	-	-	0.995	a	a	a	a	-	-	-	-	-	
	vraptor-core	p<0.05	b	ab	ab	a	-	-	-	-	0.682	a	a	a	a	a	-	-	-	-	
	assertj-core	p<0.05	d	cd	cd	bc	bc	ab	a	a	0.954	a	a	a	a	a	a	a	a	-	
	jasmine-maven	p<0.05	d	cd	bc	ab	a	-	-	-	p<0.05	a	a	b	b	a	-	-	-	-	
	la4j	p<0.05	a	a	b	b	bc	c	bc	bc	0.938	a	a	a	a	a	a	a	a	a	
	metrics-core	p<0.05	a	a	a	a	a	a	-	-	0.833	a	a	a	a	a	a	-	-	-	
Addit.	scribe-java	p<0.05	b	b	b	b	ab	ab	a	-	0.537	a	a	a	a	a	a	a	a	-	
	java-apns	p<0.05	c	c	b	b	b	b	a	-	0.999	a	a	a	a	a	a	a	a	-	
	jopt-simple	p<0.05	b	b	a	a	-	-	-	-	0.993	a	a	a	a	-	-	-	-	-	
	vraptor-core	p<0.05	b	b	ab	a	-	-	-	-	0.801	a	a	a	a	-	-	-	-	-	
	assertj-core	p<0.05	c	bc	bc	b	bc	a	a	a	0.562	a	a	a	a	a	a	a	a	-	
	jasmine-maven	p<0.05	c	c	b	a	a	-	-	-	p<0.05	a	a	b	b	a	-	-	-	-	
	la4j	p<0.05	b	b	c	c	bc	b	a	a	0.983	a	a	a	a	a	a	a	a	a	
	metrics-core	0.919	a	a	a	a	a	a	-	-	0.982	a	a	a	a	a	a	-	-	-	
ART	scribe-java	p<0.05	b	b	b	b	a	ab	a	-	p<0.05	a	c	ab	a	c	b	b	ab	-	
	java-apns	p<0.05	c	c	b	b	b	b	a	-	p<0.05	b	d	a	cd	b	bc	ab	ab	-	
	jopt-simple	p<0.05	b	b	a	a	-	-	-	-	p<0.05	b	ab	a	ab	-	-	-	-	-	
	vraptor-core	p<0.05	c	bc	abc	ab	a	-	-	-	0.918	a	a	a	a	a	-	-	-	-	
	assertj-core	p<0.05	b	ab	a	a	a	a	a	-	p<0.05	ab	b	b	ab	b	b	a	-	-	
	jasmine-maven	p<0.05	c	bc	b	a	a	-	-	-	p<0.05	a	b	a	b	b	-	-	-	-	
	la4j	p<0.05	c	c	d	d	cd	c	b	ab	a	p<0.05	d	a	a	bc	ab	cd	a	ab	bc
	metrics-core	p<0.05	b	cd	c	d	cd	a	-	-	p<0.05	a	b	b	b	c	a	-	-	-	
Gene.	scribe-java	p<0.05	b	b	b	b	b	b	a	-	p<0.05	ab	a	a	ab	b	ab	ab	ab	-	
	java-apns	p<0.05	c	c	b	b	b	b	b	a	0.778	a	a	a	a	a	a	a	a	-	
	jopt-simple	p<0.05	b	b	a	a	-	-	-	-	0.071	a	a	a	a	-	-	-	-	-	
	vraptor-core	p<0.05	b	ab	ab	a	a	-	-	-	0.760	a	a	a	a	a	-	-	-	-	
	assertj-core	p<0.05	d	cd	cd	bc	bc	ab	a	a	0.115	a	a	a	a	a	a	a	a	-	
	jasmine-maven	p<0.05	c	c	b	a	a	-	-	-	p<0.05	a	a	b	b	a	-	-	-	-	
	la4j	p<0.05	b	b	d	cd	bc	b	a	a	p<0.05	abc	a	bc	abc	cd	d	c	ab	bc	
	metrics-core	p<0.05	b	b	a	b	c	b	-	-	p<0.05	b	ab	ab	ab	ab	a	-	-	-	

TABLE III: ANOVA analysis and Tukey's HSD test among using Branch coverage of different versions for traditional prioritization















Sub	Tech	With new tests									Without new tests								
		$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$
jasmine-maven	Total	X(0.00)	X(0.00)	X(0.00)	O(0.94)	O(0.92)	-	-	-	-	O(0.35)	O(0.35)	O(0.35)	O(0.35)	O(0.35)	-	-	-	-
	Addit.	X(0.00)	X(0.00)	O(0.60)	✓(0.00)	✓(0.00)	-	-	-	-	O(0.35)	O(0.35)	O(0.35)	O(0.35)	O(0.35)	-	-	-	-
	ART	X(0.00)	X(0.00)	O(0.31)	✓(0.03)	✓(0.04)	-	-	-	-	O(0.35)	O(0.35)	O(0.35)	O(0.35)	O(0.35)	-	-	-	-
	Genetic	X(0.00)	X(0.00)	O(0.42)	✓(0.00)	✓(0.00)	-	-	-	-	O(0.35)	O(0.35)	O(0.35)	O(0.35)	O(0.35)	-	-	-	-
java-apns	Total	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.01)	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	Addit.	X(0.00)	X(0.00)	O(0.08)	O(0.13)	O(0.84)	O(0.82)	O(0.38)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	ART	X(0.00)	X(0.00)	X(0.00)	X(0.03)	X(0.02)	O(0.08)	O(0.06)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	Genetic	X(0.00)	X(0.00)	X(0.04)	O(0.15)	O(0.82)	O(0.75)	O(0.23)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
jopt-simple	Total	X(0.00)	X(0.00)	X(0.00)	X(0.00)	-	-	-	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-	-
	Addit.	X(0.00)	X(0.00)	O(0.14)	✓(0.00)	-	-	-	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-	-
	ART	X(0.00)	X(0.00)	O(0.38)	✓(0.00)	-	-	-	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-	-
	Genetic	X(0.00)	X(0.00)	X(0.03)	✓(0.00)	-	-	-	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-	-
la4j	Total	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	
	Addit.	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	X(0.00)	✓(0.03)	O(0.07)	O(0.07)	O(0.07)	✓(0.02)	O(0.07)	✓(0.03)	O(0.07)	O(0.07)
	ART	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.15)	✓(0.00)	O(0.63)	✓(0.00)	✓(0.00)	✓(0.00)	O(0.08)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.00)
	Genetic	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	X(0.00)	O(0.47)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.00)
scribe-java	Total	O(0.73)	O(0.53)	O(0.69)	O(0.88)	O(0.13)	O(0.09)	O(0.18)	X(0.03)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	Addit.	O(0.53)	O(0.60)	O(0.30)	O(0.10)	O(0.19)	✓(0.00)	✓(0.00)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	ART	O(0.06)	O(0.38)	O(0.23)	X(0.01)	X(0.03)	O(0.49)	✓(0.02)	✓(0.02)	O(0.02)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.01)	X(0.00)	✓(0.00)	✓(0.02)	✓(0.00)
	Genetic	O(0.44)	O(0.53)	O(0.31)	O(0.10)	O(0.24)	✓(0.00)	✓(0.00)	✓(0.00)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
vraptor-core	Total	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	-	-	-	-	X(0.00)	X(0.00)	X(0.00)	X(0.00)	X(0.00)	-	-	-	-
	Addit.	O(0.10)	✓(0.02)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-
	ART	X(0.02)	O(0.09)	O(0.52)	O(0.51)	O(0.99)	-	-	-	-	O(0.88)	O(0.19)	O(0.81)	O(0.56)	O(0.39)	-	-	-	-
	Genetic	O(0.43)	O(0.29)	O(0.08)	✓(0.00)	✓(0.00)	-	-	-	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	-	-	-
metrics-core	Total	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	X(0.00)	X(0.00)	-	-	-	✓(0.00)	✓(0.00)	X(0.00)	✓(0.00)	X(0.00)	-	-	-	-
	Addit.	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.05)	✓(0.00)	-	-	-	✓(0.03)	✓(0.03)	✓(0.03)	✓(0.03)	✓(0.02)	✓(0.03)	-	-	-
	ART	O(0.43)	O(0.91)	✓(0.00)	✓(0.00)	O(0.12)	✓(0.00)	-	-	-	✓(0.00)	✓(0.00)	✓(0.02)	✓(0.00)	X(0.00)	✓(0.00)	-	-	-
	Genetic	O(0.70)	✓(0.01)	O(0.54)	✓(0.00)	O(0.88)	✓(0.00)	-	-	-	O(0.92)	O(0.81)	O(0.20)	O(0.33)	✓(0.00)	O(0.16)	-	-	-
assertj-core	Total	O(0.26)	O(0.89)	O(0.64)	O(0.09)	O(0.10)	✓(0.00)	✓(0.00)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	Addit.	O(0.10)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-
	ART	O(0.58)	X(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	✓(0.00)	X(0.00)	X(0.00)	✓(0.00)	✓(0.00)	X(0.00)	✓(0.00)	✓(0.00)	-
	Genetic	X(0.01)	✓(0.00)	✓(0.04)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	✓(0.00)	-

TABLE XVIII: Wilcoxon tests between each studied technique and random technique based on Branch coverage for traditional prioritization(note: the values in parentheses are p values)











