Can Statistical Language Models be used to improve Spectrum Based Fault Localization Rankings?
(Full Spectra – Additional Material)

Anonymous Author(s)

SBFL ranking metric	$\overline{\mathcal{R}_\lambda(\Omega)}$				max improv.	$\overline{\mathcal{R}^*_\lambda(\Omega)}$				max improv.
	$\lambda = 1$	$\lambda = 0.98$	$\lambda = 0.9$	$\lambda = 0.5$		$\lambda = 1$	$\lambda = 0.98$	$\lambda = 0.9$	$\lambda = 0.5$	
AMPLE	7797.3	**7766.3**	7900.8	8603.6	0.4%	1094.8	824.0	**770.1**	822.5	29.7%
ANDERBERG	10077.4	**9632.7**	9653.0	9811.7	4.4%	597.0	**340.7**	345.1	413.4	42.9%
ARITHMETIC MEAN	14367.6	13597.6	13057.6	**11980.3**	16.6%	1072.9	811.7	756.5	**688.2**	35.9%
COHEN	14236.9	13460.9	12944.1	**11926.6**	16.2%	986.3	719.5	669.4	**649.7**	34.1%
DICE	10077.1	**9632.5**	9652.6	9811.3	4.4%	597.1	**340.8**	345.2	413.4	42.9%
EUCLID	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
FLEISS	20114.6	18442.8	17101.5	**14964.4**	25.6%	4762.1	3326.9	2601.0	**1842.7**	61.3%
GEOMETRIC MEAN	14736.6	13930.0	13328.6	**12107.7**	17.8%	1233.8	986.4	901.2	**759.7**	38.4%
GOODMAN	10096.3	**9644.8**	9659.7	9812.6	4.5%	702.0	428.9	**354.6**	414.3	49.5%
GP13	10204.9	**9761.4**	9791.6	10011.9	4.3%	889.2	582.9	**561.6**	619.9	36.8%
HAMANN	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
HAMMING ETC.	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
HARMONIC MEAN	14892.1	14069.1	13463.0	**12177.9**	18.2%	1264.1	1012.9	919.2	**765.7**	39.4%
JACCARD	10077.2	**9632.6**	9652.8	9811.4	4.4%	597.0	**340.8**	345.0	413.4	42.9%
KULCZYNSKI1	10077.3	**9632.6**	9652.9	9811.5	4.4%	597.0	**340.8**	345.2	413.4	42.9%
KULCZYNSKI2	10181.3	**9737.0**	9764.8	9971.1	4.4%	746.9	461.2	**446.1**	513.2	40.3%
M1	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
M2	10184.2	**9740.4**	9768.9	9979.5	4.4%	835.1	534.1	**510.0**	568.5	38.9%
OCHIAI	10108.4	**9663.4**	9685.6	9857.1	4.4%	610.2	353.2	**349.8**	419.6	42.7%
OCHIAI2	10200.3	**9751.7**	9770.8	9918.5	4.4%	789.0	**447.7**	449.4	521.3	43.3%
NAISH2 (OP2)	12902.7	12349.8	11976.7	**11155.1**	13.5%	942.5	709.4	**602.4**	624.3	36.1%
OVERLAP	10241.8	**9775.4**	9828.0	10203.4	4.6%	1067.7	**467.7**	502.4	821.6	56.2%
ROGERS & TANIMOTO	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
ROGOT1	20004.4	18364.9	17058.7	**14937.8**	25.3%	4719.3	3283.7	2561.3	**1805.7**	61.7%
ROGOT2	14881.4	14022.5	13386.9	**12096.1**	18.7%	1384.1	1074.8	862.6	**698.1**	49.6%
RUSSELL & RAO	10284.8	**9824.7**	9875.8	10230.1	4.5%	1198.0	**627.0**	646.0	897.1	47.7%
SCOTT	20004.4	18364.9	17058.7	**14937.8**	25.3%	4719.3	3283.7	2561.3	**1805.7**	61.7%
SIMPLE MATCHING	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
SOKAL	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
SØRENSEN-DICE	10077.3	**9632.6**	9652.8	9811.5	4.4%	597.0	**340.8**	345.1	413.4	42.9%
TARANTULA	10064.6	**9620.1**	9639.1	9796.2	4.4%	585.7	**335.4**	338.8	405.9	42.7%
WONG1	10284.8	**9824.7**	9875.8	10230.1	4.5%	1198.0	**627.0**	646.0	897.1	47.7%
WONG3	20455.7	18165.1	17076.5	**15799.9**	22.8%	10350.5	5056.8	3954.1	**3435.9**	66.8%
WONG2	23987.3	21911.0	20115.0	**17503.3**	27.0%	13655.8	8588.5	6102.9	**4253.7**	68.9%
ZOLTAR	11980.9	**11213.8**	11222.9	11292.0	6.4%	1575.2	**606.0**	608.9	655.8	61.5%

TABLE 1: OVERVIEW OF *ALL* EXAMINED SBFL METRICS WITH $\overline{\mathcal{R}_\lambda(\Omega)}$ AND $\overline{\mathcal{R}^*_\lambda(\Omega)}$ FOR $\lambda \in \{1.0, 0.98, 0.9, 0.5\}$ AND THE MAXIMUM IMPROVEMENTS FOR THE HIGHEST VALUES WITH REGARD TO $\lambda = 1$. HIGHEST RANKINGS ARE PRINTED WITH A BOLD FONT FOR EACH SET OF VALUES.

SBFL ranking metric	$\widetilde{\mathcal{R}}_\lambda(\Omega)$				max improv.	$\widetilde{\mathcal{R}}^*_\lambda(\Omega)$				max improv.
	$\lambda = 1$	$\lambda = 0.98$	$\lambda = 0.9$	$\lambda = 0.5$		$\lambda = 1$	$\lambda = 0.98$	$\lambda = 0.9$	$\lambda = 0.5$	
Ample	2843.0	**2796.0**	2978.0	3874.0	1.7%	**42.0**	43.0	44.0	75.0	0.0%
Anderberg	5321.0	3428.0	**3360.0**	4382.0	36.9%	27.0	**23.0**	26.0	40.0	14.8%
Arithmetic Mean	11064.0	8591.0	6714.0	**5572.0**	49.6%	27.0	**26.0**	28.0	41.0	3.7%
Cohen	11064.0	8412.0	6702.0	**5546.0**	49.9%	27.0	**26.0**	28.0	41.0	3.7%
Dice	5324.0	3428.0	**3360.0**	4382.0	36.9%	27.0	**23.0**	26.0	40.0	14.8%
Euclid	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Fleiss	21118.0	18014.0	15229.0	**11293.0**	46.5%	40.0	**39.0**	42.0	68.0	2.5%
Geometric Mean	11081.0	8653.0	6846.0	**5590.0**	49.6%	29.0	**27.0**	30.0	46.0	6.9%
Goodman	5324.0	3428.0	**3360.0**	4382.0	36.9%	27.0	**23.0**	26.0	40.0	14.8%
GP13	5286.0	**3863.0**	4063.0	5167.0	26.9%	33.0	**31.0**	34.0	55.0	6.1%
Hamann	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Hamming etc.	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Harmonic Mean	11426.0	8818.0	7057.0	**5590.0**	51.1%	27.0	**26.0**	27.0	45.0	3.7%
Jaccard	5320.0	3428.0	**3360.0**	4382.0	36.8%	27.0	**23.0**	26.0	40.0	14.8%
Kulczynski1	5325.0	3428.0	**3360.0**	4382.0	36.9%	27.0	**23.0**	26.0	40.0	14.8%
Kulczynski2	5274.0	**3900.0**	4079.0	4973.0	26.1%	26.0	**22.0**	23.0	40.0	15.4%
M1	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
M2	5286.0	**3865.0**	4062.0	5070.0	26.9%	34.0	**32.0**	34.0	58.0	5.9%
Ochiai	5453.0	3718.0	**3690.0**	4575.0	32.3%	28.0	**26.0**	27.0	43.0	7.1%
Ochiai2	5922.0	**3593.0**	3690.0	4482.0	39.3%	29.0	**27.0**	30.0	45.0	6.9%
Naish2 (Op2)	5630.0	4484.0	**4321.0**	5079.0	23.3%	33.0	**31.0**	34.0	55.0	6.1%
Overlap	5224.0	**3697.0**	3906.0	5285.0	29.2%	168.0	**49.0**	79.0	161.0	70.8%
Rogers & Tanimoto	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Rogot1	21118.0	17871.0	15195.0	**11296.0**	46.5%	33.0	**31.0**	33.0	55.0	6.1%
Rogot2	11686.0	8628.0	6707.0	**5365.0**	54.1%	27.0	**25.0**	**25.0**	44.0	7.4%
Russell & Rao	5307.0	**3969.0**	4203.0	5434.0	25.2%	203.0	**63.0**	88.0	189.0	69.0%
Scott	21118.0	17871.0	15195.0	**11296.0**	46.5%	33.0	**31.0**	33.0	55.0	6.1%
Simple Matching	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Sokal	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Sørensen-Dice	5324.0	3428.0	**3360.0**	4382.0	36.9%	27.0	**23.0**	26.0	40.0	14.8%
Tarantula	5321.0	3423.0	**3352.0**	4198.0	37.0%	29.0	**25.0**	26.0	45.0	13.8%
Wong1	5307.0	**3969.0**	4203.0	5434.0	25.2%	203.0	**63.0**	88.0	189.0	69.0%
Wong3	20454.0	16794.0	15619.0	**13717.0**	32.9%	6199.0	549.0	**443.0**	**443.0**	92.9%
Wong2	24776.0	22947.0	19705.0	**15753.0**	36.4%	9683.0	1162.0	1049.0	**856.0**	91.2%
Zoltar	18042.0	**7397.0**	**7397.0**	7403.0	59.0%	27.0	**23.0**	24.0	41.0	14.8%

Table 2: Overview of *all* examined SBFL metrics with $\widetilde{\mathcal{R}}_\lambda(\Omega)$ and $\widetilde{\mathcal{R}}^*_\lambda(\Omega)$ for $\lambda \in \{1.0, 0.98, 0.9, 0.5\}$ and the maximum improvements for the highest values with regard to $\lambda = 1$. Highest rankings are printed with a bold font for each set of values.

SBFL ranking metric	$\widetilde{\lambda_p}$, $(\overline{\lambda_p})$, [min, max]	$\widetilde{RI_{\mathcal{R}}}^{SBFL}$, $(\overline{RI_{\mathcal{R}}}^{SBFL})$, [min, max]	$\widetilde{RI_{\mathcal{R}}}^{LM}$, $(\overline{RI_{\mathcal{R}}}^{LM})$, [min, max]
AMPLE	0.98, (0.98), [0.98,0.98]	0.3%, (-0.4%), [-8.8%,4.1%]	58.3%, (56.8%), [34.7%,78.2%]
ANDERBERG	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.1%]	50.2%, (48.7%), [16.4%,70.4%]
ARITHMETIC MEAN	0.17, (0.18), [0.16,0.22]	18.6%, (18.6%), [3.8%,32.7%]	40.6%, (36.2%), [5.1%,51.5%]
COHEN	0.16, (0.17), [0.16,0.22]	18.3%, (16.5%), [-0.8%,32.7%]	40.4%, (36.3%), [5.1%,51.7%]
DICE	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.1%]	50.2%, (48.7%), [16.4%,70.4%]
EUCLID	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
FLEISS	0.06, (0.07), [0.06,0.1]	34.5%, (32.3%), [11.4%,43.1%]	23.6%, (21.0%), [-1.3%,41.1%]
GEOMETRIC MEAN	0.14, (0.16), [0.14,0.2]	20.7%, (20.6%), [7.2%,32.2%]	40.2%, (35.5%), [5.2%,51.1%]
GOODMAN	0.98, (0.98), [0.98,0.98]	4.3%, (3.1%), [-10.1%,17.1%]	50.2%, (48.6%), [16.4%,70.4%]
GP13	0.98, (0.98), [0.98,0.98]	4.1%, (2.8%), [-10.0%,16.3%]	48.7%, (47.5%), [16.5%,70.1%]
HAMANN	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
HAMMING etc.	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
HARMONIC MEAN	0.14, (0.15), [0.14,0.2]	21.0%, (21.1%), [6.8%,32.1%]	40.3%, (35.3%), [4.8%,51.1%]
JACCARD	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.1%]	50.2%, (48.7%), [16.4%,70.4%]
KULCZYNSKI1	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.1%]	50.2%, (48.7%), [16.4%,70.4%]
KULCZYNSKI2	0.98, (0.98), [0.98,0.98]	4.2%, (2.8%), [-10.0%,16.4%]	49.0%, (47.8%), [16.5%,70.4%]
M1	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
M2	0.98, (0.98), [0.98,0.98]	4.2%, (2.8%), [-10.0%,16.4%]	48.8%, (47.7%), [16.5%,70.1%]
OCHIAI	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.0%]	49.8%, (48.5%), [16.4%,70.4%]
OCHIAI2	0.98, (0.98), [0.98,0.98]	4.0%, (2.9%), [-10.1%,17.1%]	50.2%, (47.6%), [16.0%,68.8%]
NAISH2 (OP2)	0.28, (0.3), [0.22,0.44]	8.7%, (9.0%), [-7.0%,21.2%]	41.2%, (39.1%), [4.2%,59.2%]
OVERLAP	0.98, (0.98), [0.98,0.98]	4.3%, (3.2%), [-9.4%,17.1%]	48.8%, (47.3%), [16.7%,69.8%]
ROGERS & TANIMOTO	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
ROGOT1	0.06, (0.06), [0.06,0.08]	34.1%, (32.4%), [12.7%,43.5%]	24.3%, (21.6%), [-0.4%,41.8%]
ROGOT2	0.14, (0.15), [0.14,0.2]	23.3%, (22.1%), [7.6%,30.1%]	41.9%, (35.8%), [4.8%,51.1%]
RUSSELL & RAO	0.98, (0.98), [0.98,0.98]	4.3%, (3.0%), [-9.4%,16.7%]	48.3%, (46.9%), [16.6%,69.3%]
SCOTT	0.06, (0.06), [0.06,0.08]	34.1%, (32.4%), [12.7%,43.5%]	24.3%, (21.6%), [-0.4%,41.8%]
SIMPLE MATCHING	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
SOKAL	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
SØRENSEN-DICE	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.1%]	50.2%, (48.7%), [16.4%,70.4%]
TARANTULA	0.98, (0.98), [0.98,0.98]	4.3%, (2.9%), [-10.1%,17.1%]	50.3%, (48.9%), [16.4%,70.5%]
WONG1	0.98, (0.98), [0.98,0.98]	4.3%, (3.0%), [-9.4%,16.7%]	48.3%, (46.9%), [16.6%,69.3%]
WONG3	0.06, (0.06), [0.04,0.08]	28.0%, (25.6%), [11.2%,34.9%]	8.4%, (9.4%), [-1.0%,26.2%]
WONG2	0.02, (0.02), [0.02,0.02]	37.7%, (36.1%), [17.9%,44.5%]	4.2%, (4.4%), [-0.3%,9.7%]
ZOLTAR	0.98, (0.98), [0.98,0.98]	6.3%, (6.2%), [-8.7%,20.4%]	43.7%, (38.8%), [4.8%,61.8%]

TABLE 3: RESULTS OF THE 10-FOLD CROSS VALIDATION OF $\overline{\mathcal{R}}_{\lambda_p}(\Omega)$.

SBFL ranking metric	$\widetilde{\lambda_p}$, $(\overline{\lambda_p})$, [min, max]	$\widetilde{RI}_{\mathcal{R}*}^{SBFL}$, $(\overline{RI}_{\mathcal{R}*}^{SBFL})$, [min, max]	$\widetilde{RI}_{\mathcal{R}*}^{LM}$, $(\overline{RI}_{\mathcal{R}*}^{LM})$, [min, max]
AMPLE	0.76, (0.8), [0.76,0.9]	24.2%, (25.9%), [0.1%,50.5%]	85.2%, (83.8%), [63.4%,95.7%]
ANDERBERG	0.98, (0.97), [0.88,0.98]	31.6%, (31.0%), [1.8%,57.4%]	93.2%, (92.8%), [88.0%,97.2%]
ARITHMETIC MEAN	0.58, (0.58), [0.52,0.62]	36.9%, (29.7%), [-51.5%,65.5%]	88.2%, (85.6%), [64.4%,94.1%]
COHEN	0.58, (0.62), [0.58,0.76]	33.0%, (23.6%), [-51.9%,65.4%]	87.8%, (86.2%), [67.4%,94.1%]
DICE	0.98, (0.97), [0.88,0.98]	31.6%, (31.0%), [1.8%,57.4%]	93.2%, (92.8%), [88.0%,97.2%]
EUCLID	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
FLEISS	0.24, (0.24), [0.22,0.26]	68.2%, (63.0%), [20.2%,82.9%]	68.1%, (65.4%), [27.8%,82.7%]
GEOMETRIC MEAN	0.42, (0.43), [0.42,0.48]	39.9%, (31.3%), [-71.8%,64.0%]	87.2%, (84.1%), [53.3%,93.3%]
GOODMAN	0.88, (0.88), [0.86,0.92]	42.3%, (32.3%), [-9.6%,69.5%]	93.2%, (92.4%), [86.0%,96.9%]
GP13	0.78, (0.79), [0.78,0.86]	41.0%, (32.9%), [5.1%,51.7%]	88.5%, (88.2%), [81.4%,93.7%]
HAMANN	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
HAMMING ETC.	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
HARMONIC MEAN	0.44, (0.44), [0.44,0.48]	41.1%, (34.6%), [-51.0%,66.2%]	87.7%, (84.2%), [51.5%,94.0%]
JACCARD	0.98, (0.97), [0.88,0.98]	31.6%, (31.0%), [1.8%,57.4%]	93.2%, (92.8%), [88.0%,97.2%]
KULCZYNSKI1	0.98, (0.97), [0.88,0.98]	31.6%, (31.0%), [1.8%,57.4%]	93.2%, (92.8%), [88.0%,97.2%]
KULCZYNSKI2	0.88, (0.87), [0.8,0.9]	44.7%, (34.1%), [-16.6%,56.6%]	90.5%, (90.6%), [83.4%,96.7%]
M1	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
M2	0.81, (0.82), [0.78,0.88]	42.2%, (34.3%), [2.6%,53.5%]	89.6%, (89.3%), [81.2%,94.6%]
OCHIAI	0.88, (0.9), [0.88,0.98]	30.3%, (28.3%), [-7.8%,56.7%]	92.7%, (92.4%), [87.2%,97.5%]
OCHIAI2	0.98, (0.97), [0.88,0.98]	39.1%, (33.8%), [0.8%,59.9%]	91.8%, (90.7%), [81.9%,96.3%]
NAISH2 (OP2)	0.78, (0.78), [0.76,0.8]	32.0%, (33.6%), [4.2%,66.7%]	88.8%, (88.1%), [81.1%,93.6%]
OVERLAP	0.98, (0.98), [0.98,0.98]	56.8%, (56.7%), [45.5%,69.2%]	91.1%, (90.3%), [85.0%,94.3%]
ROGERS & TANIMOTO	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
ROGOT1	0.18, (0.19), [0.18,0.24]	67.8%, (63.2%), [21.3%,83.0%]	68.5%, (66.0%), [32.0%,83.2%]
ROGOT2	0.44, (0.44), [0.44,0.48]	49.2%, (43.0%), [-51.0%,75.3%]	88.3%, (85.4%), [51.4%,93.9%]
RUSSELL & RAO	0.98, (0.98), [0.98,0.98]	45.2%, (46.6%), [36.7%,55.8%]	86.4%, (86.8%), [84.0%,90.6%]
SCOTT	0.18, (0.19), [0.18,0.24]	67.8%, (63.2%), [21.3%,83.0%]	68.5%, (66.0%), [32.0%,83.2%]
SIMPLE MATCHING	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
SOKAL	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
SØRENSEN-DICE	0.98, (0.97), [0.88,0.98]	31.6%, (31.0%), [1.8%,57.4%]	93.2%, (92.8%), [88.0%,97.2%]
TARANTULA	0.98, (0.97), [0.88,0.98]	31.6%, (31.0%), [3.5%,59.0%]	93.3%, (92.9%), [87.8%,97.2%]
WONG1	0.98, (0.98), [0.98,0.98]	45.2%, (46.6%), [36.7%,55.8%]	86.4%, (86.8%), [84.0%,90.6%]
WONG3	0.4, (0.38), [0.26,0.4]	64.5%, (66.7%), [51.2%,80.2%]	27.7%, (27.9%), [-1.4%,50.4%]
WONG2	0.06, (0.07), [0.06,0.08]	70.4%, (71.4%), [58.0%,81.7%]	21.1%, (19.7%), [1.4%,34.4%]
ZOLTAR	0.94, (0.95), [0.94,0.98]	60.7%, (57.3%), [23.7%,84.1%]	89.0%, (87.7%), [66.9%,97.2%]

TABLE 4: RESULTS OF THE 10-FOLD CROSS VALIDATION OF $\overline{\mathcal{R}}_{\lambda_p}^*$ (Ω).

SBFL ranking metric	$\widetilde{\lambda_p}$, $(\overline{\lambda_p})$, [min, max]	$\widetilde{RI}_{\widetilde{\mathcal{R}}}^{SBFL}$, $(\overline{RI}_{\widetilde{\mathcal{R}}}^{SBFL})$, [min, max]	$\widetilde{RI}_{\widetilde{\mathcal{R}}}^{LM}$, $(\overline{RI}_{\widetilde{\mathcal{R}}}^{LM})$, [min, max]
AMPLE	0.97, (0.97), [0.94,1.0]	-0.1%, (0.9%), [-11.3%,23.2%]	85.7%, (78.9%), [47.4%,97.6%]
ANDERBERG	0.98, (0.97), [0.94,0.98]	3.7%, (10.5%), [-4.6%,35.1%]	92.1%, (77.9%), [29.3%,98.0%]
ARITHMETIC MEAN	0.43, (0.54), [0.42,0.92]	-33.9%, (-36.3%), [-139.4%,59.8%]	83.9%, (66.3%), [-5.5%,96.1%]
COHEN	0.46, (0.56), [0.42,0.92]	-37.7%, (-34.6%), [-139.6%,60.9%]	83.6%, (66.1%), [-5.5%,96.1%]
DICE	0.98, (0.97), [0.94,0.98]	3.7%, (10.5%), [-4.6%,35.1%]	92.1%, (77.9%), [29.3%,98.0%]
EUCLID	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
FLEISS	0.2, (0.21), [0.1,0.3]	56.3%, (54.7%), [29.4%,75.6%]	39.5%, (40.8%), [-4.6%,83.2%]
GEOMETRIC MEAN	0.43, (0.52), [0.4,0.9]	-28.2%, (-33.2%), [-134.6%,60.9%]	84.3%, (66.1%), [-5.6%,95.6%]
GOODMAN	0.98, (0.97), [0.94,0.98]	3.7%, (10.5%), [-4.6%,35.1%]	92.1%, (77.9%), [29.3%,98.0%]
GP13	0.98, (0.97), [0.94,0.98]	0.0%, (8.4%), [-8.1%,35.1%]	81.3%, (72.3%), [29.3%,97.7%]
HAMANN	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
HAMMING ETC.	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
HARMONIC MEAN	0.42, (0.52), [0.38,0.9]	-37.0%, (-30.6%), [-138.8%,60.3%]	83.7%, (64.7%), [-9.4%,95.9%]
JACCARD	0.98, (0.97), [0.94,0.98]	3.7%, (10.5%), [-4.6%,35.1%]	92.1%, (77.9%), [29.3%,98.0%]
KULCZYNSKI1	0.98, (0.97), [0.94,0.98]	3.7%, (10.5%), [-4.6%,35.1%]	92.1%, (77.9%), [29.3%,98.0%]
KULCZYNSKI2	0.98, (0.97), [0.94,0.98]	1.3%, (8.6%), [-6.3%,35.1%]	81.7%, (73.4%), [29.3%,97.8%]
M1	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
M2	0.98, (0.98), [0.98,0.98]	2.2%, (9.4%), [-5.4%,35.1%]	81.3%, (73.0%), [29.3%,97.8%]
OCHIAI	0.98, (0.97), [0.94,0.98]	1.7%, (9.6%), [-5.0%,35.1%]	91.7%, (76.7%), [29.3%,97.9%]
OCHIAI2	0.98, (0.97), [0.86,0.98]	4.1%, (10.1%), [0.7%,34.9%]	92.0%, (77.1%), [28.2%,97.9%]
NAISH2 (OP2)	0.8, (0.82), [0.66,0.94]	-6.4%, (-12.9%), [-54.5%,14.6%]	80.2%, (63.3%), [-5.7%,97.2%]
OVERLAP	0.98, (0.98), [0.94,0.98]	9.9%, (15.2%), [-0.4%,35.1%]	84.4%, (72.6%), [29.3%,92.7%]
ROGERS & TANIMOTO	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
ROGOT1	0.2, (0.21), [0.1,0.3]	60.5%, (56.8%), [29.6%,75.3%]	42.3%, (42.4%), [-4.5%,87.6%]
ROGOT2	0.42, (0.52), [0.4,0.94]	-57.0%, (-48.6%), [-138.8%,62.0%]	84.4%, (64.3%), [-12.4%,94.9%]
RUSSELL & RAO	0.98, (0.98), [0.94,0.98]	12.3%, (11.5%), [-24.0%,35.1%]	79.8%, (70.7%), [29.3%,92.2%]
SCOTT	0.2, (0.21), [0.1,0.3]	60.5%, (56.8%), [29.6%,75.3%]	42.3%, (42.4%), [-4.5%,87.6%]
SIMPLE MATCHING	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
SOKAL	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
SØRENSEN-DICE	0.98, (0.97), [0.94,0.98]	3.7%, (10.5%), [-4.6%,35.1%]	92.1%, (77.9%), [29.3%,98.0%]
TARANTULA	0.97, (0.97), [0.94,0.98]	1.9%, (11.7%), [-2.7%,35.1%]	91.5%, (78.5%), [29.3%,98.3%]
WONG1	0.98, (0.98), [0.94,0.98]	12.3%, (11.5%), [-24.0%,35.1%]	79.8%, (70.7%), [29.3%,92.2%]
WONG3	0.2, (0.2), [0.16,0.22]	40.1%, (42.4%), [22.3%,65.0%]	15.6%, (17.4%), [-0.2%,53.1%]
WONG2	0.04, (0.07), [0.02,0.18]	52.4%, (48.0%), [30.2%,62.4%]	8.1%, (6.6%), [-1.6%,15.4%]
ZOLTAR	0.48, (0.58), [0.32,0.98]	3.2%, (-22.7%), [-203.3%,55.5%]	79.4%, (67.6%), [10.0%,96.3%]

TABLE 5: RESULTS OF THE 10-FOLD CROSS VALIDATION OF $\widetilde{\mathcal{R}}_{\lambda_p}(\Omega)$.

SBFL ranking metric	$\widetilde{\lambda_p}$, $(\overline{\lambda_p})$, [min, max]	$\widetilde{RI}_{\widetilde{\mathcal{R}}^*}^{SBFL}$, $(\overline{RI}_{\widetilde{\mathcal{R}}^*}^{SBFL})$, [min, max]	$\widetilde{RI}_{\widetilde{\mathcal{R}}^*}^{LM}$, $(\overline{RI}_{\widetilde{\mathcal{R}}^*}^{LM})$, [min, max]
Ample	0.99, (0.96), [0.9,1.0]	0.0%, (-2.8%), [-21.9%,18.5%]	97.9%, (97.1%), [94.1%,98.7%]
Anderberg	0.98, (0.97), [0.94,0.98]	10.2%, (10.5%), [-42.5%,41.7%]	98.6%, (98.3%), [95.3%,99.6%]
Arithmetic Mean	0.98, (0.98), [0.98,0.98]	10.2%, (15.8%), [-6.5%,50.0%]	98.7%, (98.2%), [94.3%,99.6%]
Cohen	0.98, (0.97), [0.94,0.98]	10.2%, (15.7%), [-6.5%,50.0%]	98.7%, (98.2%), [94.3%,99.6%]
Dice	0.98, (0.97), [0.94,0.98]	10.2%, (10.5%), [-42.5%,41.7%]	98.6%, (98.3%), [95.3%,99.6%]
Euclid	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Fleiss	0.98, (0.99), [0.98,1.0]	1.7%, (3.5%), [-12.2%,23.3%]	97.5%, (97.5%), [94.7%,99.5%]
Geometric Mean	0.98, (0.98), [0.98,0.98]	10.5%, (16.9%), [-2.9%,56.3%]	98.7%, (98.1%), [94.3%,99.6%]
Goodman	0.98, (0.97), [0.94,0.98]	10.2%, (10.5%), [-42.5%,41.7%]	98.6%, (98.3%), [95.3%,99.6%]
GP13	0.98, (0.98), [0.94,0.98]	13.2%, (12.2%), [-30.5%,60.6%]	98.7%, (98.3%), [96.7%,99.5%]
Hamann	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Hamming etc.	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Harmonic Mean	0.98, (0.97), [0.9,1.0]	0.0%, (4.2%), [-32.5%,35.7%]	98.7%, (98.3%), [95.7%,99.6%]
Jaccard	0.98, (0.97), [0.94,0.98]	10.2%, (10.5%), [-42.5%,41.7%]	98.6%, (98.3%), [95.3%,99.6%]
Kulczynski1	0.98, (0.97), [0.94,0.98]	10.2%, (10.5%), [-42.5%,41.7%]	98.6%, (98.3%), [95.3%,99.6%]
Kulczynski2	0.98, (0.97), [0.9,0.98]	12.9%, (14.7%), [-7.4%,50.0%]	98.6%, (98.6%), [97.3%,99.7%]
M1	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
M2	0.98, (0.98), [0.98,0.98]	9.4%, (11.5%), [-11.4%,53.6%]	98.6%, (98.4%), [97.2%,99.5%]
Ochiai	0.98, (0.96), [0.92,0.98]	4.9%, (10.8%), [-42.5%,56.3%]	98.4%, (98.2%), [95.3%,99.7%]
Ochiai2	0.98, (0.97), [0.92,0.98]	9.5%, (15.8%), [-3.5%,56.3%]	98.7%, (98.1%), [93.7%,99.6%]
Naish2 (Op2)	0.98, (0.98), [0.94,0.98]	13.2%, (12.2%), [-30.5%,60.6%]	98.7%, (98.3%), [96.7%,99.5%]
Overlap	0.98, (0.98), [0.98,0.98]	67.1%, (66.7%), [43.2%,84.1%]	96.9%, (96.8%), [94.2%,99.4%]
Rogers & Tanimoto	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Rogot1	0.98, (0.97), [0.94,0.98]	8.3%, (8.8%), [-19.4%,28.6%]	97.9%, (97.8%), [94.2%,99.6%]
Rogot2	0.98, (0.97), [0.9,1.0]	0.0%, (8.0%), [-5.4%,35.7%]	98.7%, (98.4%), [96.8%,99.6%]
Russell & Rao	0.98, (0.98), [0.98,0.98]	75.9%, (73.8%), [48.6%,84.1%]	96.1%, (96.1%), [93.0%,99.2%]
Scott	0.98, (0.97), [0.94,0.98]	8.3%, (8.8%), [-19.4%,28.6%]	97.9%, (97.8%), [94.2%,99.6%]
Simple Matching	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Sokal	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Sørensen-Dice	0.98, (0.97), [0.94,0.98]	10.2%, (10.5%), [-42.5%,41.7%]	98.6%, (98.3%), [95.3%,99.6%]
Tarantula	0.98, (0.96), [0.9,0.98]	15.3%, (17.8%), [-6.2%,50.0%]	98.6%, (98.4%), [96.6%,99.7%]
Wong1	0.98, (0.98), [0.98,0.98]	75.9%, (73.8%), [48.6%,84.1%]	96.1%, (96.1%), [93.0%,99.2%]
Wong3	0.69, (0.69), [0.58,0.78]	90.4%, (88.5%), [71.1%,96.6%]	66.8%, (64.8%), [31.3%,92.0%]
Wong2	0.38, (0.43), [0.34,0.9]	91.1%, (90.6%), [81.1%,97.2%]	51.7%, (49.9%), [-4.9%,92.9%]
Zoltar	0.98, (0.96), [0.9,0.98]	8.5%, (12.7%), [-15.9%,50.0%]	98.6%, (98.6%), [97.3%,99.7%]

Table 6: Results of the 10-fold cross validation of $\widetilde{\mathcal{R}}_{\lambda_p}^*$ (Ω).

FIGURE 1: PLOTS OF $\overline{\mathcal{R}}_\lambda(\Omega)$ (SOLID) AND $\widetilde{\mathcal{R}}_\lambda(\Omega)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

FIGURE 2: PLOTS OF $\overline{\mathcal{R}}_\lambda(\Omega)$ (SOLID) AND $\widetilde{\mathcal{R}}_\lambda(\Omega)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.
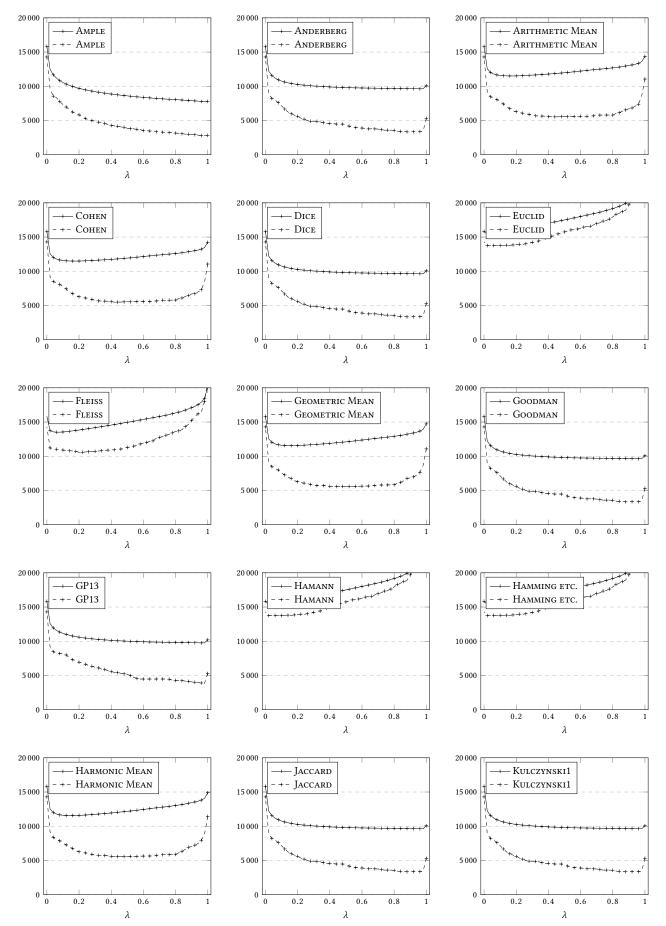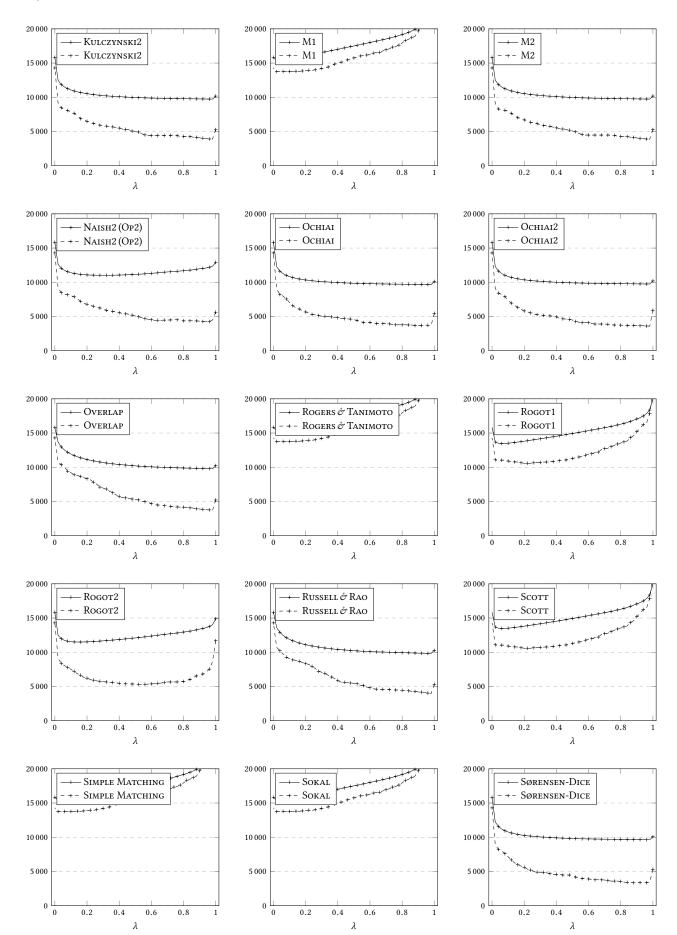
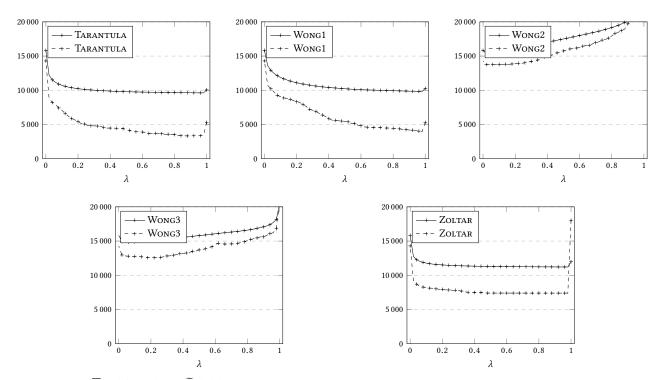**Figure 3: Plots of $\overline{\mathcal{R}}_\lambda(\Omega)$ (solid) and $\widetilde{\mathcal{R}}_\lambda(\Omega)$ (dashed) for different SBFL ranking metrics.**

Can Statistical Language Models be used to improve Spectrum Based Fault Localization Rankings?
(Full Spectra – Additional Material)

ISSTA, 2017, Santa Barbara

FIGURE 4: PLOTS OF $\overline{\mathcal{R}}^*_\lambda(\Omega)$ (SOLID) $\widetilde{\mathcal{R}}^*_\lambda(\Omega)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.
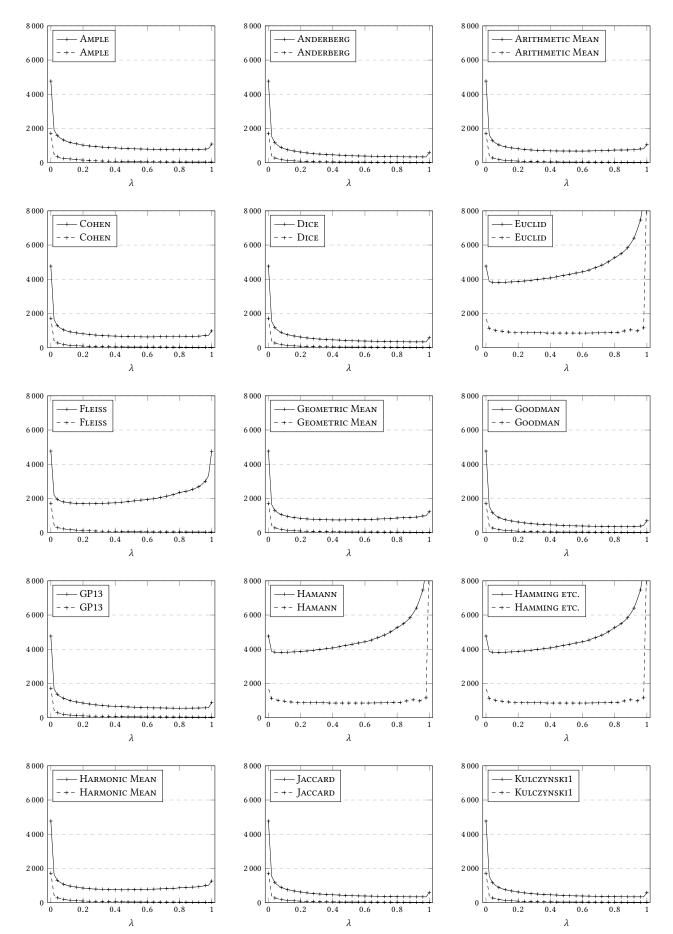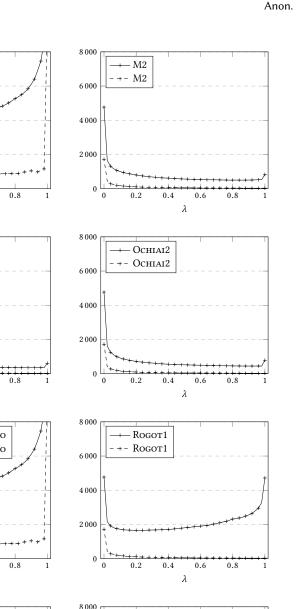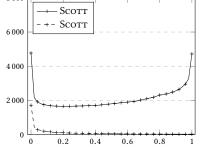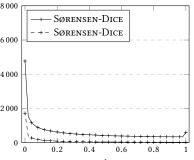
FIGURE 5: PLOTS OF $\overline{\mathcal{R}}^*_\lambda(\Omega)$ (SOLID) $\widetilde{\mathcal{R}}^*_\lambda(\Omega)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

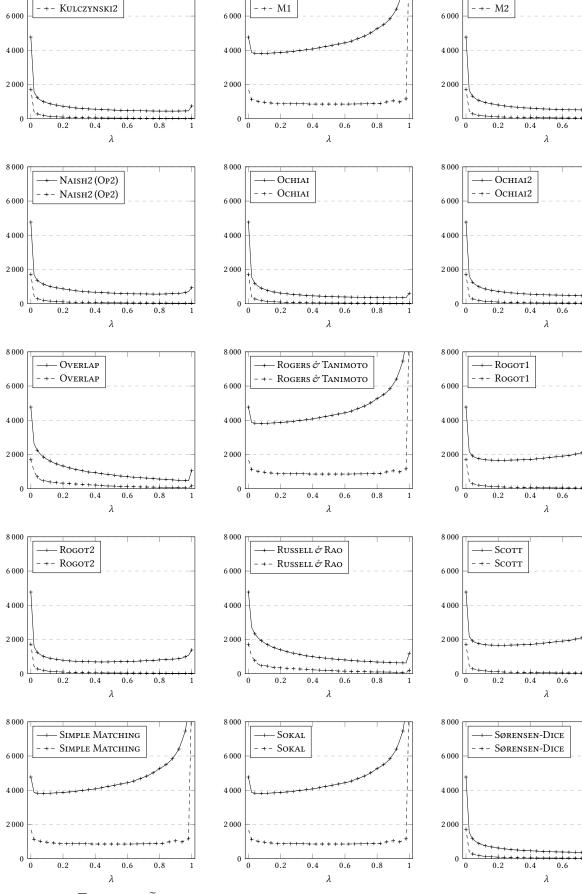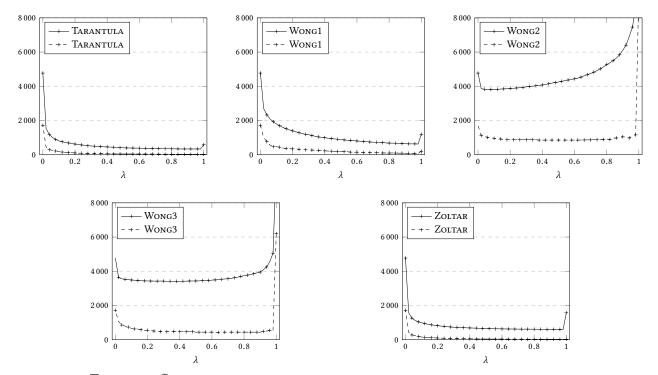**Figure 6: Plots of** $\overline{\mathcal{R}}^*_\lambda(\Omega)$ **(solid)** $\widetilde{\mathcal{R}}^*_\lambda(\Omega)$ **(dashed) for different SBFL ranking metrics.**