

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

Anonymous Author(s)

SBFL ranking metric	$\overline{\mathcal{R}}_\lambda(\Omega_f)$					max improv.	$\overline{\mathcal{R}}_\lambda^*(\Omega_f)$					max improv.
	$\lambda = 1$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 0.3$			$\lambda = 1$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 0.3$		
AMPLE	900.6	819.5	807.8	804.9	10.6%	504.3	252.0	216.4	200.4	60.3%		
ANDERBERG	750.9	689.5	694.3	714.1	8.2%	238.9	180.3	171.5	170.0	28.9%		
ARITHMETIC MEAN	703.6	666.2	678.8	704.4	5.3%	238.7	178.7	169.6	168.5	29.4%		
COHEN	746.1	688.4	693.7	713.9	7.7%	239.0	180.1	171.2	169.9	28.9%		
DICE	750.5	689.2	694.2	714.1	8.2%	239.0	180.3	171.5	170.0	28.9%		
EUCLID	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
FLEISS	698.6	667.9	686.5	718.5	4.4%	273.8	189.8	177.9	174.1	36.4%		
GEOMETRIC MEAN	701.1	665.7	679.6	705.7	5.1%	236.9	180.4	172.4	170.8	27.9%		
GOODMAN	750.4	689.3	694.3	714.1	8.1%	238.8	180.1	171.3	169.9	28.8%		
GP13	1019.1	902.1	878.6	860.5	15.6%	538.0	260.6	224.3	201.2	62.6%		
HAMANN	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
HAMMING ETC.	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
HARMONIC MEAN	699.4	665.1	679.2	705.3	4.9%	243.7	186.5	179.6	174.7	28.3%		
JACCARD	750.7	689.4	694.3	714.1	8.2%	239.0	180.3	171.5	170.0	28.9%		
KULCZYNSKI1	750.8	689.3	694.3	714.1	8.2%	239.0	180.3	171.5	170.0	28.9%		
KULCZYNSKI2	969.5	858.5	837.9	823.9	15.0%	392.4	218.4	194.9	181.4	53.8%		
M1	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
M2	975.6	861.6	839.2	825.9	15.3%	482.6	241.9	212.1	194.9	59.6%		
OCHIAI	816.2	736.6	731.0	739.5	10.4%	252.5	183.6	174.9	172.5	31.7%		
OCHIAI2	704.2	667.1	680.5	706.4	5.3%	236.1	180.1	172.4	170.7	27.7%		
NAISH2(OP2)	1018.9	902.6	879.3	861.4	15.5%	537.5	261.5	226.3	202.3	62.4%		
OVERLAP	1096.8	972.1	961.0	951.3	13.3%	720.8	255.6	242.6	236.1	67.2%		
ROGERS & TANIMOTO	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
ROGOT1	690.1	657.1	673.9	703.9	4.8%	249.6	178.6	169.6	168.4	32.5%		
ROGOT2	699.2	665.0	679.2	705.3	4.9%	243.7	186.5	179.6	174.7	28.3%		
RUSSELL & RAO	1187.1	1022.5	1002.0	982.2	17.3%	854.2	293.2	264.8	248.3	70.9%		
SCOTT	690.1	657.1	673.9	703.9	4.8%	249.6	178.6	169.6	168.4	32.5%		
SIMPLE MATCHING	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
SOKAL	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
SØRENSEN-DICE	750.8	689.4	694.3	714.1	8.2%	239.0	180.3	171.5	170.0	28.9%		
TARANTULA	724.1	676.5	685.0	708.8	6.6%	227.4	177.6	168.7	168.9	25.8%		
WONG1	1187.1	1022.5	1002.0	982.2	17.3%	854.2	293.2	264.8	248.3	70.9%		
WONG3	800.6	752.9	758.1	769.4	6.0%	307.4	212.7	197.2	187.5	39.0%		
WONG2	699.6	670.2	687.9	718.0	4.2%	246.3	179.4	170.9	171.8	30.6%		
ZOLTAR	877.4	783.6	773.3	774.2	11.9%	330.1	197.5	183.8	173.2	47.5%		

TABLE 1: OVERVIEW OF ALL EXAMINED SBFL METRICS WITH $\overline{\mathcal{R}}_\lambda(\Omega_f)$ AND $\overline{\mathcal{R}}_\lambda^*(\Omega_f)$ FOR $\lambda \in \{1.0, 0.7, 0.5, 0.3\}$ 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	$\tilde{\mathcal{R}}_{\lambda}(\Omega_f)$					max improv.	$\tilde{\mathcal{R}}_{\lambda}^*(\Omega_f)$				
	$\lambda = 1$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 0.3$			$\lambda = 1$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 0.3$	
AMPLE	218.0	230.0	224.0	232.0	0.0%	33.0	15.0	16.0	19.0	54.5%	
ANDERBERG	200.0	215.0	213.0	222.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
ARITHMETIC MEAN	218.0	228.0	222.0	226.5	0.0%	25.0	13.0	14.0	17.0	48.0%	
COHEN	200.5	222.0	219.0	224.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
DICE	200.0	215.0	213.0	222.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
EUCLID	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
FLEISS	201.0	222.0	222.0	235.0	0.0%	32.0	13.0	14.0	17.0	59.4%	
GEOMETRIC MEAN	220.0	228.0	223.5	229.0	0.0%	26.0	12.0	13.0	17.0	53.8%	
GOODMAN	200.0	215.0	213.0	222.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
GP13	259.0	234.0	229.0	232.5	11.6%	31.0	14.0	16.0	19.0	54.8%	
HAMANN	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
HAMMING ETC.	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
HARMONIC MEAN	215.0	219.0	215.0	220.0	0.0%	26.0	12.0	14.0	17.0	53.8%	
JACCARD	200.0	215.0	213.0	222.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
KULCZYNSKI1	200.0	215.0	213.0	222.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
KULCZYNSKI2	239.0	232.0	226.0	230.5	5.4%	24.0	12.0	13.0	17.0	50.0%	
M1	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
M2	235.0	227.0	220.5	226.0	6.2%	32.0	14.0	17.0	19.0	56.3%	
OCHIAI	200.0	222.5	221.0	229.5	0.0%	26.0	12.0	14.0	17.0	53.8%	
OCHIAI2	210.0	220.5	219.0	225.0	0.0%	26.0	12.0	13.0	17.0	53.8%	
NAISH2 (Op2)	259.0	234.0	229.0	233.0	11.6%	31.0	14.0	16.0	19.0	54.8%	
OVERLAP	584.0	302.0	283.0	279.0	52.2%	142.0	24.0	26.0	27.0	83.1%	
ROGERS & TANIMOTO	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
ROGOT1	200.0	217.0	216.0	222.0	0.0%	26.0	13.0	14.0	17.0	50.0%	
ROGOT2	215.0	219.0	215.0	220.0	0.0%	26.0	12.0	14.0	17.0	53.8%	
RUSSELL & RAO	465.0	311.0	289.5	283.0	39.1%	190.0	29.0	28.0	28.0	85.3%	
SCOTT	200.0	217.0	216.0	222.0	0.0%	26.0	13.0	14.0	17.0	50.0%	
SIMPLE MATCHING	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
SOKAL	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
SØRENSEN-DICE	200.0	215.0	213.0	222.0	0.0%	25.0	13.0	14.0	17.0	48.0%	
TARANTULA	200.0	216.5	215.5	223.0	0.0%	27.0	13.0	14.0	19.0	51.9%	
WONG1	465.0	311.0	289.5	283.0	39.1%	190.0	29.0	28.0	28.0	85.3%	
WONG3	200.0	224.0	214.0	224.0	0.0%	33.0	15.0	14.0	17.0	57.6%	
WONG2	200.0	213.5	215.5	218.0	0.0%	27.0	14.0	15.0	17.0	48.1%	
ZOLTAR	215.0	231.5	228.0	230.0	0.0%	25.0	14.0	14.0	17.0	44.0%	

TABLE 2: OVERVIEW OF ALL EXAMINED SBFL METRICS WITH $\tilde{\mathcal{R}}_{\lambda}(\Omega_f)$ AND $\tilde{\mathcal{R}}_{\lambda}^*(\Omega_f)$ FOR $\lambda \in \{1.0, 0.7, 0.5, 0.3\}$ 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	$\widetilde{\lambda}_p, (\lambda_p), [\min, \max]$	$\widetilde{RI}_{\overline{R}}^{SBFL}, (RI_{\overline{R}}^{SBFL}), [\min, \max]$	$\widetilde{RI}_{\overline{R}}^{LM}, (RI_{\overline{R}}^{LM}), [\min, \max]$
AMPLE	0.34, (0.33), [0.2,0.44]	7.0%, (-2.1%), [-62.1%,33.8%]	13.4%, (15.0%), [-16.0%,57.7%]
ANDERBERG	0.67, (0.67), [0.6,0.74]	7.6%, (4.0%), [-23.9%,31.8%]	23.8%, (27.8%), [-16.9%,73.8%]
ARITHMETIC MEAN	0.82, (0.81), [0.7,0.86]	2.1%, (5.4%), [-8.9%,27.6%]	25.7%, (28.8%), [-21.2%,77.0%]
COHEN	0.66, (0.67), [0.62,0.76]	6.4%, (4.4%), [-22.8%,31.0%]	24.7%, (28.1%), [-17.5%,74.1%]
DICE	0.67, (0.67), [0.6,0.74]	7.6%, (4.1%), [-22.1%,31.8%]	23.8%, (27.9%), [-16.9%,74.2%]
EUCLID	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
FLEISS	0.85, (0.84), [0.76,0.88]	3.0%, (5.3%), [-8.0%,25.3%]	28.2%, (28.1%), [-24.2%,74.6%]
GEOMETRIC MEAN	0.84, (0.83), [0.7,0.88]	2.8%, (5.6%), [-5.2%,26.1%]	27.8%, (29.1%), [-16.8%,77.1%]
GOODMAN	0.67, (0.67), [0.6,0.74]	7.6%, (3.9%), [-23.9%,31.8%]	23.8%, (27.9%), [-16.9%,73.8%]
GP13	0.14, (0.15), [0.08,0.26]	17.2%, (4.3%), [-84.2%,41.0%]	9.2%, (9.1%), [-17.4%,39.8%]
HAMANN	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
HAMMING ETC.	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
HARMONIC MEAN	0.84, (0.85), [0.8,0.92]	2.3%, (5.1%), [-4.9%,23.1%]	32.8%, (29.7%), [-3.6%,76.6%]
JACCARD	0.67, (0.67), [0.6,0.74]	7.6%, (3.9%), [-23.9%,31.8%]	23.8%, (28.8%), [-16.9%,73.8%]
KULCZYNSKI1	0.67, (0.67), [0.6,0.74]	7.7%, (4.1%), [-22.1%,31.8%]	23.8%, (27.9%), [-16.9%,74.2%]
KULCZYNSKI2	0.22, (0.22), [0.1,0.4]	12.7%, (2.3%), [-90.7%,38.2%]	9.9%, (11.6%), [-19.0%,42.5%]
M1	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
M2	0.24, (0.23), [0.14,0.4]	14.9%, (4.5%), [-72.1%,39.6%]	10.5%, (12.1%), [-19.9%,48.4%]
OCHIAI	0.5, (0.52), [0.42,0.64]	5.8%, (1.5%), [-35.6%,35.6%]	19.1%, (23.7%), [-15.1%,69.1%]
OCHIAI2	0.84, (0.83), [0.7,0.86]	3.1%, (5.4%), [-6.6%,26.5%]	29.8%, (29.3%), [-18.1%,76.8%]
NAISH2 (Op2)	0.14, (0.15), [0.08,0.26]	17.0%, (4.1%), [-84.4%,40.9%]	9.2%, (9.1%), [-17.6%,39.7%]
OVERLAP	0.0, (0.03), [0.0,0.24]	16.0%, (12.7%), [-18.1%,31.9%]	0.0%, (-1.1%), [-9.6%,0.0%]
ROGERS & TANIMOTO	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
ROGOT1	0.84, (0.83), [0.72,0.88]	3.3%, (5.3%), [-9.5%,26.5%]	27.6%, (29.1%), [-23.5%,73.8%]
ROGOT2	0.84, (0.85), [0.8,0.92]	2.3%, (5.1%), [-5.0%,23.1%]	32.8%, (29.7%), [-3.6%,76.6%]
RUSSELL & RAO	0.0, (0.0), [0.0,0.0]	18.5%, (17.9%), [-4.7%,45.9%]	0.0%, (0.0%), [0.0%,0.0%]
SCOTT	0.84, (0.83), [0.72,0.88]	3.3%, (5.3%), [-9.5%,26.5%]	27.6%, (29.1%), [-23.5%,73.8%]
SIMPLE MATCHING	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
SOKAL	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
SØRENSEN-DICE	0.68, (0.67), [0.6,0.74]	7.6%, (3.9%), [-23.9%,31.8%]	23.8%, (27.8%), [-16.9%,73.8%]
TARANTULA	0.76, (0.74), [0.66,0.8]	6.4%, (4.1%), [-20.9%,26.3%]	36.8%, (30.1%), [-20.4%,73.9%]
WONG1	0.0, (0.0), [0.0,0.0]	18.5%, (17.9%), [-4.7%,45.9%]	0.0%, (0.0%), [0.0%,0.0%]
WONG3	0.73, (0.73), [0.6,0.84]	3.2%, (3.1%), [-14.9%,26.6%]	14.6%, (19.1%), [-12.4%,73.8%]
WONG2	0.85, (0.84), [0.76,0.9]	2.8%, (5.3%), [-8.8%,26.7%]	27.4%, (27.7%), [-24.5%,66.0%]
ZOLTAR	0.42, (0.4), [0.26,0.46]	10.2%, (6.8%), [-44.6%,35.1%]	15.2%, (17.6%), [-10.3%,61.4%]

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

SBFL ranking metric	$\widetilde{\lambda}_p, (\lambda_p), [\min, \max]$	$\widetilde{RI}_{\overline{R}^*}^{SBFL}, (\overline{RI}_{\overline{R}^*}^{SBFL}), [\min, \max]$	$\widetilde{RI}_{\overline{R}^*}^{LM}, (\overline{RI}_{\overline{R}^*}^{LM}), [\min, \max]$
AMPLE	0.14, (0.15), [0.12,0.2]	61.9%, (60.1%), [25.8%,80.4%]	17.6%, (16.5%), [-20.9%,43.8%]
ANDERBERG	0.32, (0.37), [0.26,0.52]	23.0%, (22.3%), [-21.4%,65.7%]	27.9%, (20.8%), [-30.0%,58.4%]
ARITHMETIC MEAN	0.32, (0.37), [0.28,0.54]	29.0%, (24.3%), [-37.8%,66.6%]	27.4%, (22.0%), [-27.0%,58.3%]
COHEN	0.31, (0.37), [0.28,0.52]	25.4%, (22.7%), [-22.3%,66.7%]	27.9%, (21.4%), [-29.3%,58.5%]
DICE	0.32, (0.37), [0.26,0.52]	23.0%, (22.3%), [-21.4%,65.9%]	27.8%, (20.8%), [-30.0%,58.4%]
EUCLID	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
FLEISS	0.32, (0.33), [0.24,0.4]	39.2%, (33.5%), [-44.2%,70.6%]	22.8%, (22.3%), [-24.2%,62.3%]
GEOMETRIC MEAN	0.36, (0.35), [0.2,0.4]	30.4%, (26.6%), [-44.6%,59.7%]	26.2%, (24.2%), [-24.6%,60.6%]
GOODMAN	0.32, (0.37), [0.26,0.52]	23.0%, (22.3%), [-21.4%,65.7%]	28.1%, (20.9%), [-29.2%,58.4%]
GP13	0.14, (0.13), [0.08,0.14]	68.7%, (65.4%), [34.1%,79.1%]	17.9%, (17.8%), [-11.4%,53.3%]
HAMANN	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
HAMMING ETC.	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
HARMONIC MEAN	0.3, (0.3), [0.28,0.36]	29.0%, (28.8%), [-15.2%,52.6%]	26.9%, (23.8%), [-26.0%,58.9%]
JACCARD	0.32, (0.37), [0.26,0.52]	23.0%, (22.3%), [-21.4%,65.9%]	27.9%, (20.9%), [-30.0%,58.4%]
KULCZYNSKI1	0.32, (0.37), [0.26,0.52]	23.0%, (22.3%), [-21.4%,66.0%]	27.8%, (20.8%), [-30.0%,58.4%]
KULCZYNSKI2	0.14, (0.15), [0.14,0.16]	57.9%, (54.6%), [26.9%,76.5%]	21.0%, (23.1%), [-14.0%,52.7%]
M1	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
M2	0.16, (0.15), [0.12,0.16]	65.4%, (61.8%), [34.6%,76.1%]	19.4%, (19.0%), [-14.8%,53.3%]
OCHIAI	0.34, (0.34), [0.2,0.4]	29.5%, (26.6%), [-28.8%,64.1%]	29.0%, (22.5%), [-29.5%,60.3%]
OCHIAI2	0.36, (0.35), [0.2,0.4]	27.1%, (25.3%), [-53.7%,61.5%]	26.5%, (23.6%), [-23.7%,60.6%]
NAISH2 (Op2)	0.14, (0.13), [0.08,0.14]	68.5%, (65.3%), [33.8%,79.0%]	17.7%, (17.7%), [-11.6%,52.9%]
OVERLAP	0.08, (0.08), [0.0,0.16]	73.6%, (67.3%), [34.4%,81.3%]	-0.8%, (-1.1%), [-5.7%,3.6%]
ROGERS & TANIMOTO	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
ROGOT1	0.34, (0.38), [0.3,0.52]	33.9%, (27.8%), [-44.2%,69.7%]	27.3%, (23.4%), [-23.0%,61.7%]
ROGOT2	0.3, (0.3), [0.28,0.36]	29.0%, (28.8%), [-15.2%,52.6%]	26.9%, (23.8%), [-26.0%,58.9%]
RUSSELL & RAO	0.0, (0.01), [0.0,0.14]	78.0%, (72.7%), [55.7%,84.8%]	0.0%, (-0.9%), [-9.1%,0.0%]
SCOTT	0.34, (0.38), [0.3,0.52]	33.9%, (27.8%), [-44.2%,69.7%]	27.3%, (23.4%), [-23.0%,61.7%]
SIMPLE MATCHING	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
SOKAL	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
SØRENSEN-DICE	0.32, (0.36), [0.26,0.52]	23.5%, (22.5%), [-21.4%,65.9%]	27.8%, (21.1%), [-29.3%,58.4%]
TARANTULA	0.34, (0.38), [0.26,0.5]	24.3%, (18.2%), [-23.1%,65.4%]	28.9%, (21.0%), [-31.8%,60.7%]
WONG1	0.0, (0.01), [0.0,0.14]	78.0%, (72.7%), [55.7%,84.8%]	0.0%, (-0.9%), [-9.1%,0.0%]
WONG3	0.18, (0.19), [0.16,0.26]	42.0%, (40.1%), [12.2%,63.6%]	25.8%, (18.8%), [-18.9%,43.9%]
WONG2	0.4, (0.39), [0.26,0.4]	34.4%, (23.8%), [-51.6%,68.9%]	29.7%, (23.0%), [-21.2%,65.9%]
ZOLTAR	0.15, (0.18), [0.14,0.3]	47.0%, (43.3%), [-9.1%,80.1%]	22.7%, (23.0%), [-21.9%,53.2%]

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

SBFL ranking metric	$\widehat{\lambda}_p, (\overline{\lambda}_p), [\min, \max]$	$\widehat{RI}_R^{SBFL}, (\overline{RI}_R^{SBFL}), [\min, \max]$	$\widehat{RI}_R^{LM}, (\overline{RI}_R^{LM}), [\min, \max]$
AMPLE	0.84, (0.76), [0.42,1.0]	0.3%, (7.2%), [-19.6%,31.7%]	17.0%, (14.1%), [-35.0%,59.8%]
ANDERBERG	1.0, (0.95), [0.5,1.0]	0.0%, (0.6%), [-8.6%,12.4%]	11.9%, (7.5%), [-168.7%,77.0%]
ARITHMETIC MEAN	0.68, (0.71), [0.4,1.0]	0.0%, (-6.1%), [-60.8%,9.5%]	15.1%, (-3.4%), [-206.3%,63.4%]
COHEN	1.0, (0.83), [0.46,1.0]	0.0%, (-1.5%), [-8.8%,2.5%]	14.6%, (-1.4%), [-207.2%,77.3%]
DICE	1.0, (0.95), [0.5,1.0]	0.0%, (0.6%), [-8.6%,12.4%]	11.9%, (7.5%), [-168.7%,77.0%]
EUCLID	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
FLEISS	1.0, (0.95), [0.46,1.0]	0.0%, (2.8%), [0.0%,27.5%]	9.5%, (-9.9%), [-247.0%,76.7%]
GEOMETRIC MEAN	0.66, (0.71), [0.42,1.0]	0.0%, (-5.6%), [-64.2%,11.7%]	14.1%, (-1.2%), [-161.5%,62.7%]
GOODMAN	1.0, (0.95), [0.5,1.0]	0.0%, (0.6%), [-8.6%,12.4%]	11.9%, (7.5%), [-168.7%,77.0%]
GP13	0.44, (0.44), [0.4,0.48]	15.4%, (10.7%), [-39.2%,50.7%]	16.5%, (17.4%), [-28.4%,59.9%]
HAMANN	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
HAMMING ETC.	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
HARMONIC MEAN	0.43, (0.56), [0.34,1.0]	0.0%, (-2.8%), [-75.0%,22.5%]	24.0%, (17.5%), [-38.8%,58.4%]
JACCARD	1.0, (0.95), [0.5,1.0]	0.0%, (0.6%), [-8.6%,12.4%]	11.9%, (7.5%), [-168.7%,77.0%]
KULCZYNSKI1	1.0, (0.95), [0.5,1.0]	0.0%, (0.6%), [-8.6%,12.4%]	11.9%, (7.5%), [-168.7%,77.0%]
KULCZYNSKI2	0.46, (0.55), [0.22,1.0]	0.0%, (0.9%), [-49.5%,38.7%]	13.5%, (12.0%), [-37.0%,60.2%]
M1	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
M2	0.44, (0.44), [0.4,0.46]	14.4%, (11.1%), [-37.1%,50.8%]	24.0%, (21.1%), [-23.7%,60.1%]
OCHIAI	1.0, (0.92), [0.48,1.0]	0.0%, (2.3%), [-5.3%,15.5%]	11.5%, (2.6%), [-151.2%,77.0%]
OCHIAI2	0.93, (0.77), [0.42,1.0]	0.0%, (-4.9%), [-60.2%,12.2%]	14.3%, (3.2%), [-155.1%,63.6%]
NAISH2 (Op2)	0.44, (0.44), [0.4,0.48]	15.4%, (10.7%), [-41.4%,50.7%]	16.5%, (17.5%), [-28.4%,59.9%]
OVERLAP	0.06, (0.08), [0.0,0.34]	36.7%, (34.3%), [-16.5%,76.4%]	0.0%, (-1.9%), [-25.3%,7.5%]
ROGERS & TANIMOTO	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
ROGOT1	1.0, (0.95), [0.48,1.0]	0.0%, (-0.6%), [-6.0%,0.0%]	13.8%, (-4.6%), [-248.9%,77.3%]
ROGOT2	0.44, (0.56), [0.34,1.0]	0.0%, (-2.3%), [-75.0%,22.5%]	23.5%, (18.0%), [-38.8%,58.4%]
RUSSELL & RAO	0.05, (0.09), [0.02,0.3]	39.5%, (32.5%), [-30.5%,75.1%]	-0.1%, (-2.3%), [-34.4%,11.0%]
SCOTT	1.0, (0.95), [0.48,1.0]	0.0%, (-0.6%), [-6.0%,0.0%]	13.8%, (-4.6%), [-248.9%,77.3%]
SIMPLE MATCHING	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
SOKAL	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
SØRENSEN-DICE	1.0, (0.95), [0.5,1.0]	0.0%, (0.6%), [-8.6%,12.4%]	11.9%, (7.5%), [-168.7%,77.0%]
TARANTULA	1.0, (0.95), [0.5,1.0]	0.0%, (1.3%), [0.0%,12.5%]	13.4%, (6.4%), [-169.4%,80.7%]
WONG1	0.05, (0.09), [0.02,0.3]	39.5%, (32.5%), [-30.5%,75.1%]	-0.1%, (-2.3%), [-34.4%,11.0%]
WONG3	1.0, (0.84), [0.44,1.0]	0.0%, (-0.5%), [-29.3%,20.0%]	13.5%, (10.9%), [-47.9%,73.1%]
WONG2	1.0, (0.95), [0.52,1.0]	0.0%, (4.5%), [0.0%,30.9%]	21.0%, (1.0%), [-249.6%,78.0%]
ZOLTAR	1.0, (0.83), [0.4,1.0]	0.0%, (-2.9%), [-31.6%,9.5%]	16.2%, (-4.3%), [-216.0%,73.4%]

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

SBFL ranking metric	$\widetilde{\lambda}_p, (\lambda_p), [\min, \max]$	$\widetilde{RI}_{\widetilde{R}^*}^{SBFL}, (RI_{\widetilde{R}^*}^{SBFL}), [\min, \max]$	$\widetilde{RI}_{\widetilde{R}^*}^{LM}, (RI_{\widetilde{R}^*}^{LM}), [\min, \max]$
AMPLE	0.62, (0.61), [0.4,0.74]	34.0%, (31.5%), [-25.0%,76.7%]	51.3%, (40.4%), [-29.5%,61.5%]
ANDERBERG	0.66, (0.66), [0.62,0.76]	40.0%, (38.7%), [13.2%,76.5%]	59.5%, (56.0%), [38.5%,70.8%]
ARITHMETIC MEAN	0.66, (0.66), [0.62,0.74]	38.4%, (37.8%), [9.5%,76.2%]	58.3%, (55.4%), [38.5%,70.8%]
COHEN	0.66, (0.66), [0.62,0.74]	39.1%, (38.2%), [9.5%,76.7%]	58.3%, (55.8%), [38.5%,70.8%]
DICE	0.66, (0.66), [0.62,0.76]	40.0%, (38.7%), [13.2%,76.5%]	59.5%, (56.0%), [38.5%,70.8%]
EUCLID	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
FLEISS	0.74, (0.73), [0.66,0.78]	30.9%, (32.3%), [-1.4%,71.3%]	46.2%, (42.1%), [-11.4%,63.6%]
GEOMETRIC MEAN	0.68, (0.68), [0.6,0.76]	33.6%, (32.4%), [-5.9%,71.2%]	51.5%, (49.6%), [9.1%,70.8%]
GOODMAN	0.66, (0.66), [0.62,0.76]	40.0%, (38.7%), [13.2%,76.5%]	59.5%, (56.0%), [38.5%,70.8%]
GP13	0.66, (0.65), [0.6,0.7]	36.9%, (36.4%), [0.0%,76.3%]	52.3%, (46.6%), [20.7%,61.7%]
HAMANN	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
HAMMING ETC.	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
HARMONIC MEAN	0.62, (0.64), [0.58,0.78]	24.4%, (31.2%), [3.1%,78.9%]	58.1%, (51.1%), [6.8%,63.6%]
JACCARD	0.66, (0.66), [0.62,0.76]	40.0%, (38.7%), [13.2%,76.5%]	59.5%, (56.0%), [38.5%,70.8%]
KULCZYNSKI1	0.66, (0.66), [0.62,0.76]	40.0%, (38.7%), [13.2%,76.5%]	59.5%, (56.0%), [38.5%,70.8%]
KULCZYNSKI2	0.62, (0.62), [0.58,0.66]	25.0%, (28.8%), [-11.8%,79.1%]	56.4%, (52.8%), [27.6%,69.2%]
M1	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
M2	0.66, (0.67), [0.62,0.76]	33.3%, (27.4%), [-21.9%,74.2%]	39.2%, (40.5%), [0.0%,61.5%]
OCHIAI	0.66, (0.68), [0.62,0.74]	37.8%, (37.1%), [-7.4%,72.6%]	51.1%, (52.1%), [9.1%,73.8%]
OCHIAI2	0.74, (0.7), [0.6,0.76]	31.7%, (32.9%), [-5.9%,71.2%]	50.0%, (50.3%), [13.6%,70.8%]
NAISH2 (Op2)	0.66, (0.65), [0.6,0.7]	36.9%, (36.4%), [0.0%,76.3%]	52.3%, (46.6%), [20.7%,61.7%]
OVERLAP	0.65, (0.66), [0.52,0.82]	76.4%, (77.7%), [58.5%,93.7%]	7.9%, (9.3%), [-1.9%,30.8%]
ROGERS & TANIMOTO	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
ROGOT1	0.66, (0.66), [0.62,0.74]	39.1%, (36.7%), [4.8%,76.7%]	57.2%, (54.8%), [38.5%,70.8%]
ROGOT2	0.62, (0.64), [0.58,0.78]	24.4%, (31.2%), [3.1%,78.9%]	58.1%, (51.1%), [6.8%,63.6%]
RUSSELL & RAO	0.16, (0.38), [0.02,0.82]	86.4%, (84.5%), [69.8%,91.1%]	0.0%, (-5.4%), [-54.5%,4.6%]
SCOTT	0.66, (0.66), [0.62,0.74]	39.1%, (36.7%), [4.8%,76.7%]	57.2%, (54.8%), [38.5%,70.8%]
SIMPLE MATCHING	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
SOKAL	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
SØRENSEN-DICE	0.66, (0.66), [0.62,0.76]	40.0%, (38.7%), [13.2%,76.5%]	59.5%, (56.0%), [38.5%,70.8%]
TARANTULA	0.78, (0.78), [0.78,0.8]	41.4%, (36.3%), [4.3%,69.0%]	52.6%, (49.2%), [0.0%,71.8%]
WONG1	0.16, (0.38), [0.02,0.82]	86.4%, (84.5%), [69.8%,91.1%]	0.0%, (-5.4%), [-54.5%,4.6%]
WONG3	0.52, (0.53), [0.4,0.7]	30.5%, (30.7%), [-13.5%,76.0%]	49.5%, (42.7%), [-34.1%,63.6%]
WONG2	0.61, (0.62), [0.58,0.7]	32.8%, (29.4%), [-21.9%,73.1%]	51.8%, (48.8%), [27.3%,63.1%]
ZOLTAR	0.66, (0.63), [0.5,0.66]	29.2%, (33.4%), [-13.2%,77.4%]	58.4%, (53.4%), [20.7%,68.2%]

TABLE 6: RESULTS OF THE 10-FOLD CROSS VALIDATION OF $\widetilde{R}_{\lambda_p}^*$ (Ω_f).

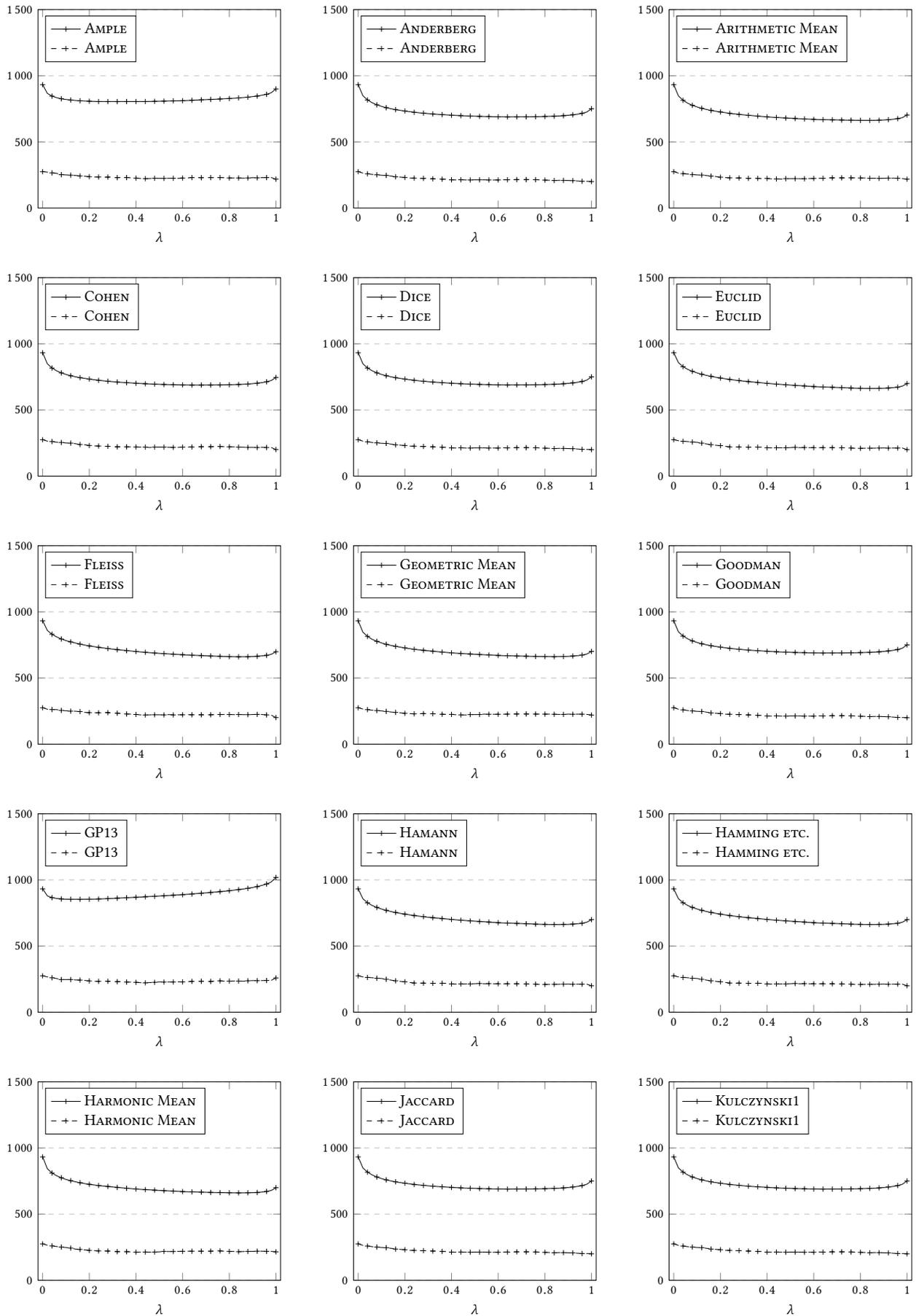


FIGURE 1: PLOTS OF $\bar{\mathcal{R}}_\lambda(\Omega_f)$ (SOLID) AND $\tilde{\mathcal{R}}_\lambda(\Omega_f)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

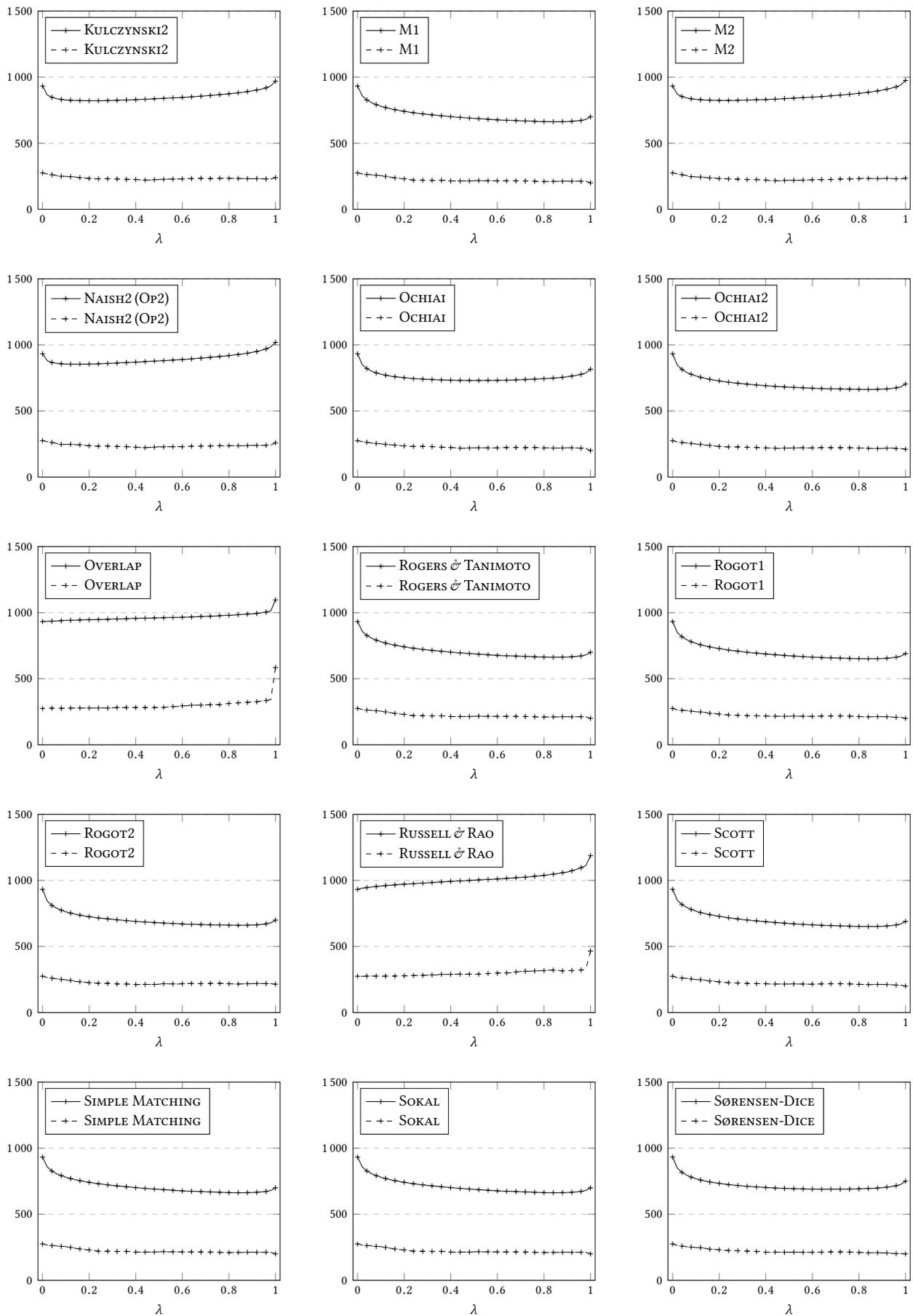


FIGURE 2: PLOTS OF $\bar{\mathcal{R}}_{\lambda}(\Omega_f)$ (SOLID) AND $\bar{\mathcal{R}}_{\lambda}(\Omega_f)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

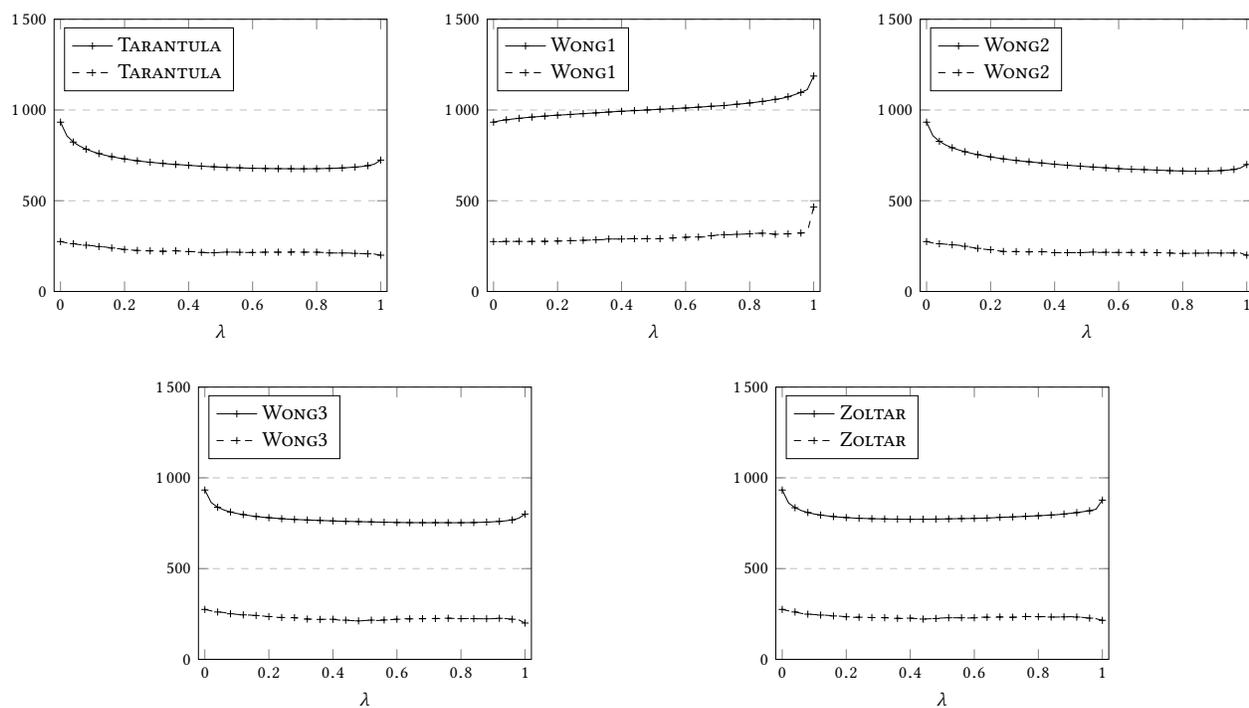


FIGURE 3: PLOTS OF $\bar{\mathcal{R}}_\lambda(\Omega_f)$ (SOLID) AND $\tilde{\mathcal{R}}_\lambda(\Omega_f)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

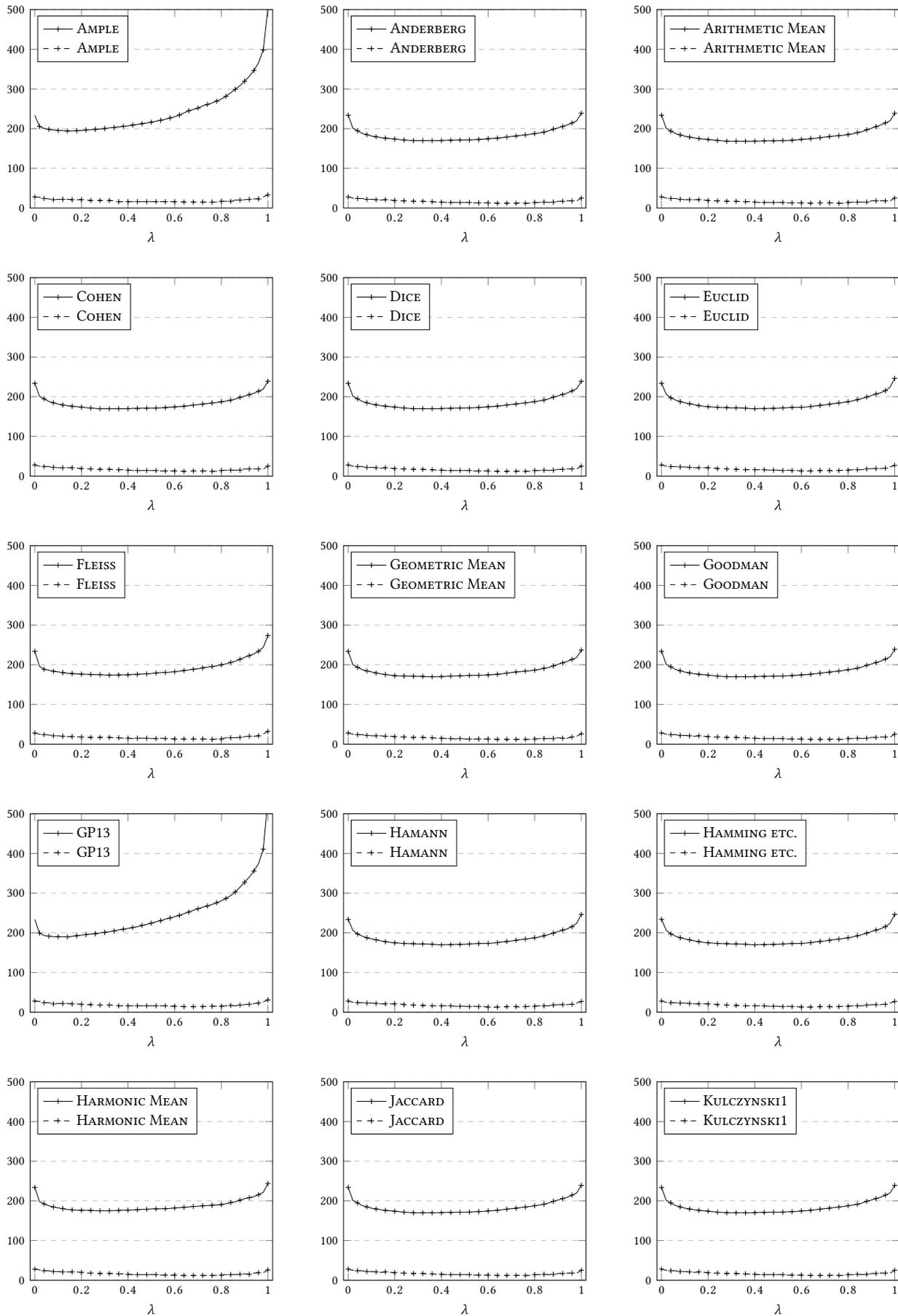


FIGURE 4: PLOTS OF $\bar{\mathcal{R}}_{\lambda}^*(\Omega_f)$ (SOLID) $\bar{\mathcal{R}}_{\lambda}(\Omega_f)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

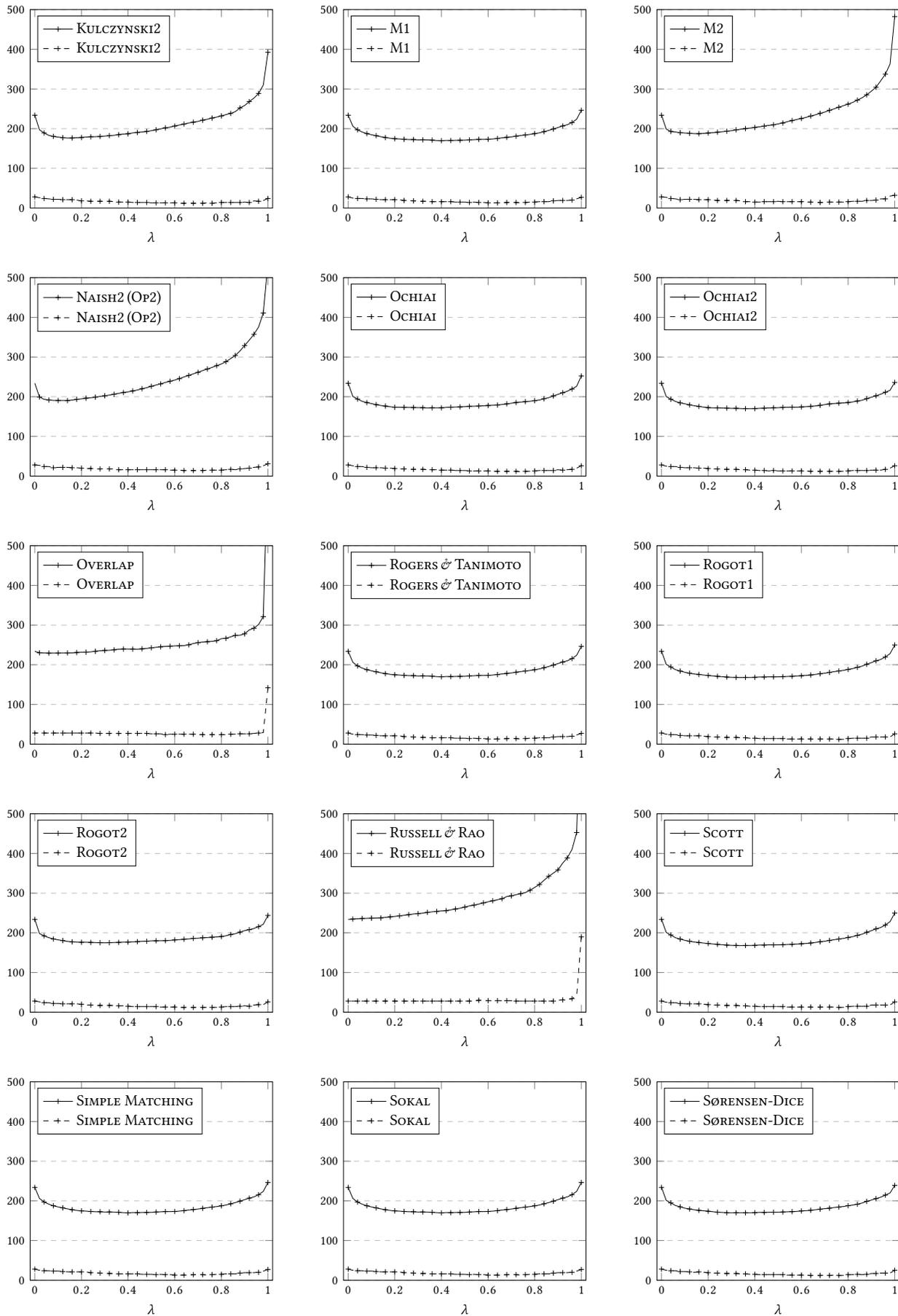


FIGURE 5: PLOTS OF $\bar{\mathcal{R}}_{\lambda}^*(\Omega_f)$ (SOLID) $\bar{\mathcal{R}}_{\lambda}(\Omega_f)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.

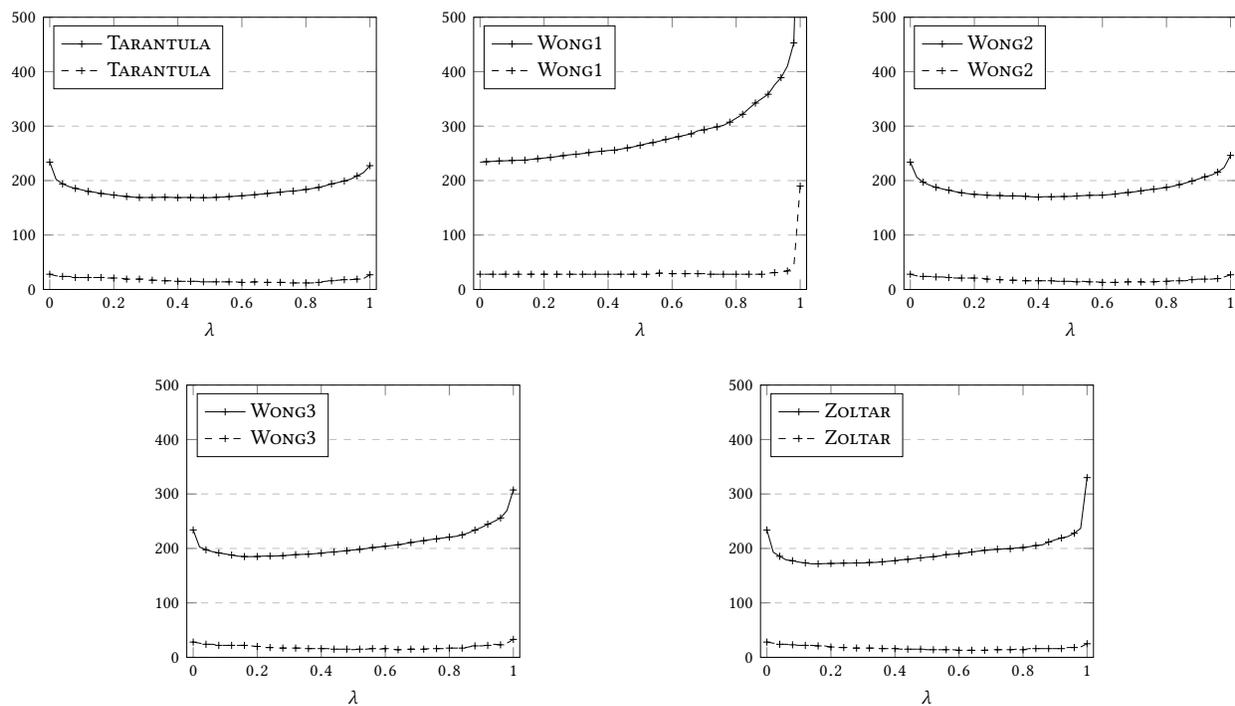


FIGURE 6: PLOTS OF $\bar{\mathcal{R}}_{\lambda}^*(\Omega_f)$ (SOLID) $\bar{\mathcal{R}}_{\lambda}^{\dagger}(\Omega_f)$ (DASHED) FOR DIFFERENT SBFL RANKING METRICS.