

1 Logarithmic vs. plain measures

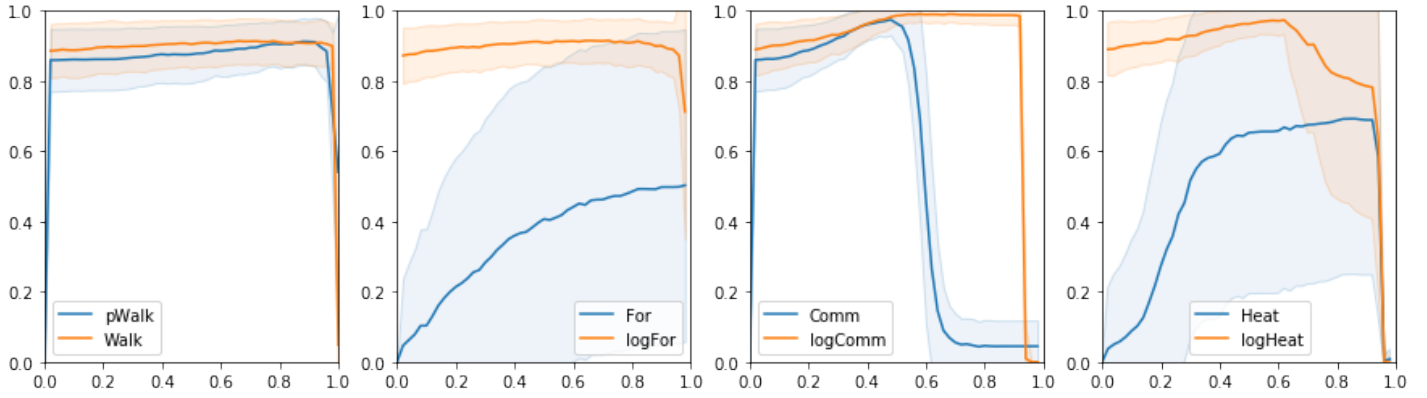


Рис. 1: Logarithmic vs. plain measures for $G(100, (2)0.2, 0.05)$

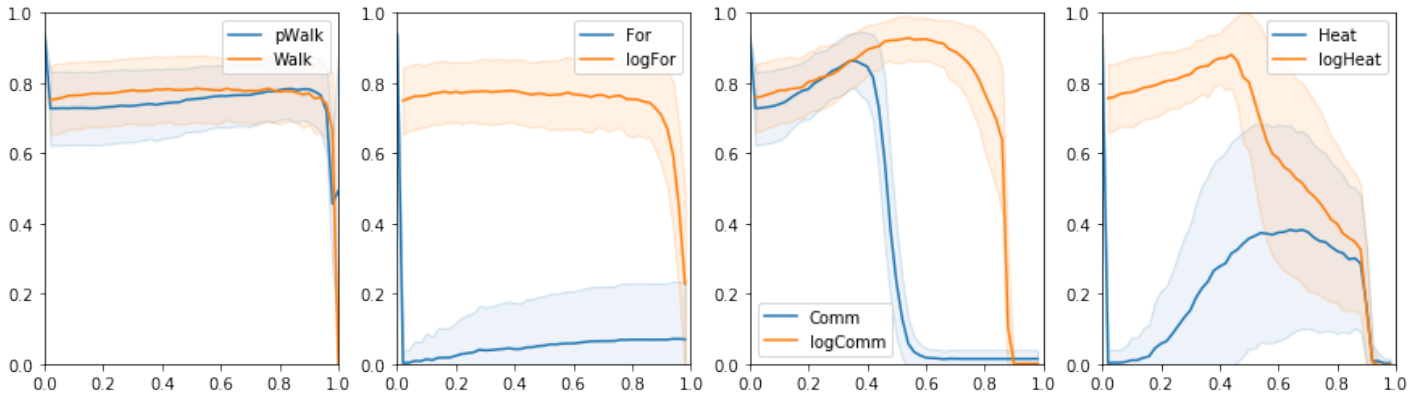


Рис. 2: Logarithmic vs. plain measures for $G(100, (3)0.3, 0.1)$

2

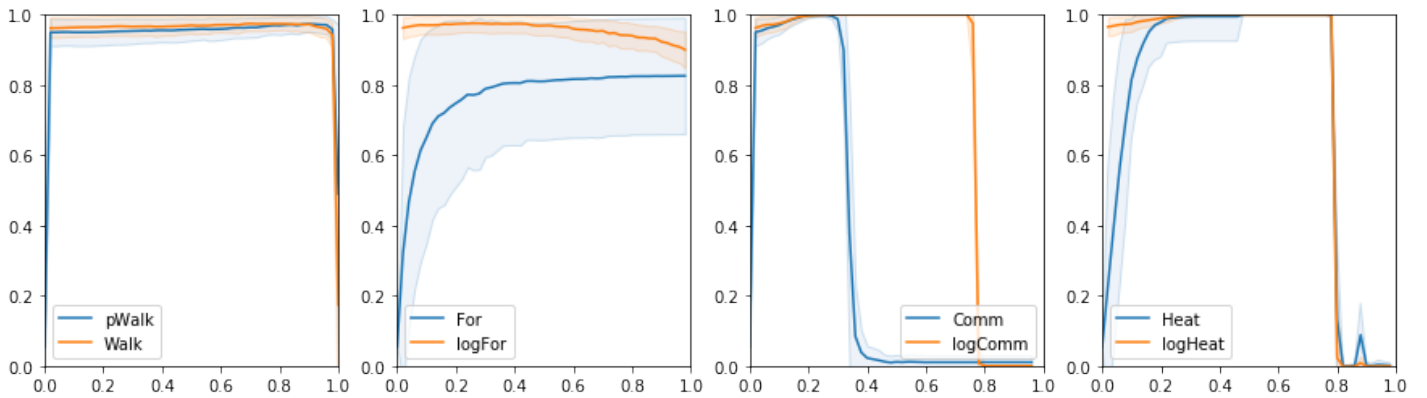


Рис. 3: Logarithmic vs. plain measures for $G(200, (2)0.3, 0.1)$

2 Competition by Copeland's score

Nodes	100	100	100	100	200	200	200	200	Sum
Classes	2	2	4	4	2	2	4	4	of
p_{out}	0.1	0.15	0.1	0.15	0.1	0.15	0.1	0.15	scores
logComm	383	547	476	-66	301	565	592		2798
Comm	249	150	308	418	291	212	325		1953
SCCT	316	299	166	44	293	392	412		1922
logHeat	308	314	180	-264	301	321	343		1503
pWalk	-81	26	56	418	-105	-155	6		165
SCT	-74	36	78	44	47	-24	44		151
Heat	221	-342	-456	418	295	205	-478		-137
RSP	-96	4	62	-272	-32	-85	-30		-449
Walk	-90	-26	78	-222	-149	-125	-26		-560
logFor	-92	-44	-24	-264	-63	-92	-32		-611
FE	-202	-64	-44	-224	-135	-169	-134		-972
For	-387	-566	-456	418	-525	-574	-478		-2568
SP-CT	-455	-334	-424	-448	-519	-471	-544		-3195

(a) optimal parameters

Nodes	100	100	100	100	200	200	200	200	Sum
Classes	2	2	4	4	2	2	4	4	of
p_{out}	0.1	0.15	0.1	0.15	0.1	0.15	0.1	0.15	scores
logComm	413	568	448	356	332	568	598	598	3881
SCCT	269	274	136	78	340	391	423	360	2271
logHeat	318	183	290	142	340	273	202	98	1846
Comm	168	151	222	172	286	258	333	178	1768
SCT	58	92	46	90	26	45	38	104	499
logFor	-114	60	56	110	-55	-115	4	88	34
Walk	-84	-10	132	86	-140	-85	30	66	-5
pWalk	-125	-40	54	74	-163	-79	-2	-14	-295
FE	-198	-27	-27	120	-120	-186	-66	32	-472
RSP	-151	-1	-8	78	-138	-179	-106	-16	-521
Heat	299	-341	-502	-490	340	154	-417	-515	-1472
SP-CT	-463	-345	-320	-228	-558	-462	-446	-396	-3218
For	-390	-564	-588	-588	-490	-583	-591	-583	-4377

(b) 90th percentiles

Таблица 1: Copeland's scores of the measure families on random graphs

3 Reject curves

Measure (kernel)	$G(100, (2)0.3, 0.05)$ Opt. parameter, ARI	$G(100, (2)0.3, 0.1)$ Opt. parameter, ARI	$G(100, (2)0.3, 0.15)$ Opt. parameter, ARI
pWalk	0.86, 0.9653	0.90, 0.8308	0.66, 0.5298
Walk	0.86, 0.9664	0.74, 0.8442	0.64, 0.5357
For	1.00, 0.5816	0.98, 0.3671	0.00, 0.2007
logFor	0.62, 0.9704	0.56, 0.8542	0.52, 0.5541
Comm	0.38, 0.9761	0.32, 0.8708	0.26, 0.5661
logComm	0.68, 0.9838	0.54, 0.9466	0.62, 0.7488
Heat	0.86, 0.6128	0.86, 0.5646	0.78, 0.2879
logHeat	0.52, 0.9827	0.40, 0.8911	0.28, 0.5561
SCT	0.74, 0.9651	0.62, 0.8550	0.64, 0.5531
SCCT	0.36, 0.9834	0.26, 0.9130	0.22, 0.6626
RSP	0.99, 0.9712	0.98, 0.8444	0.98, 0.5430
FE	0.94, 0.9697	0.94, 0.8482	0.86, 0.5460
SP-CT	0.28, 0.9172	0.34, 0.6782	0.42, 0.4103

Таблица 2: Optimal family parameters and the corresponding ARI's

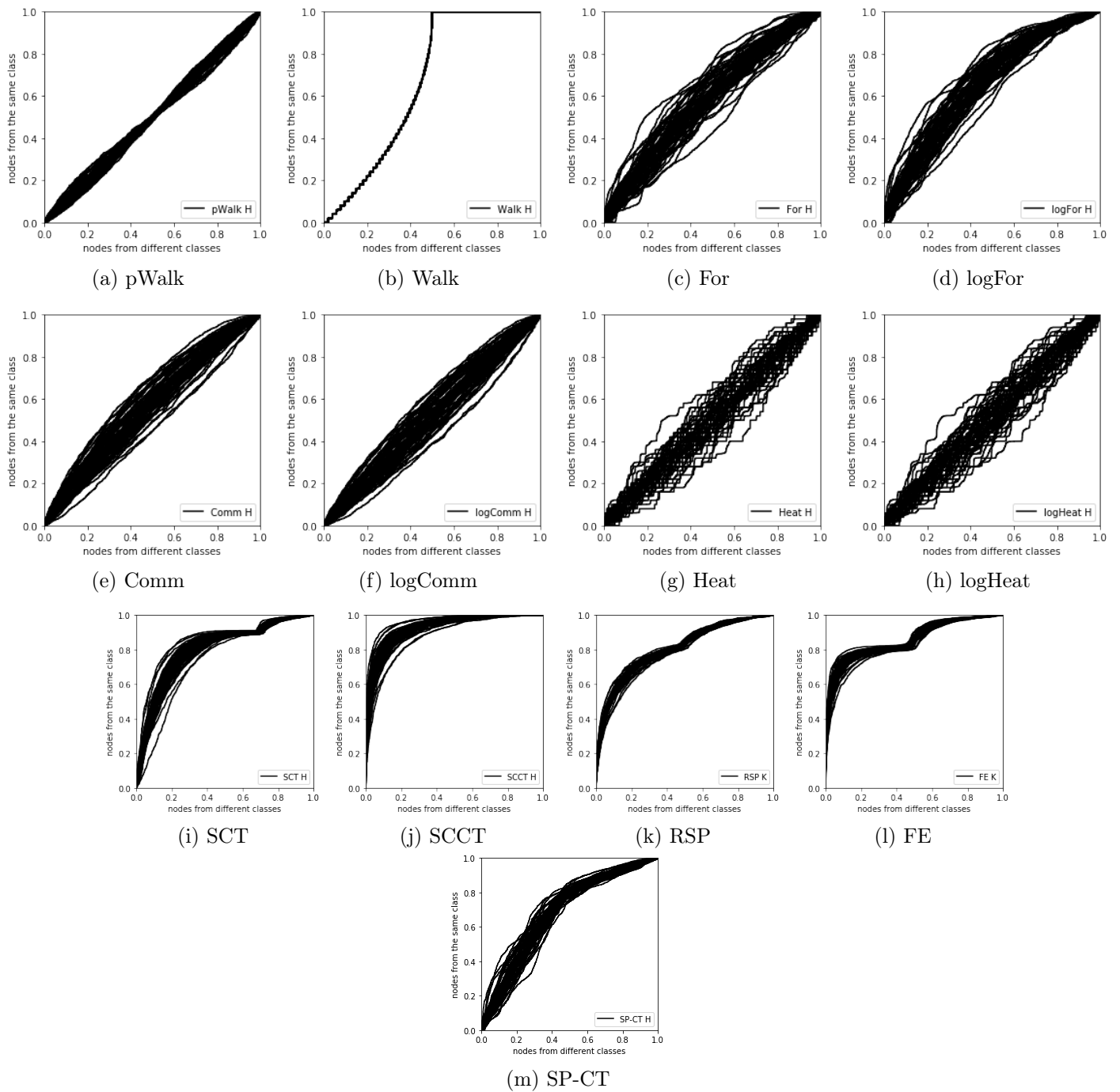
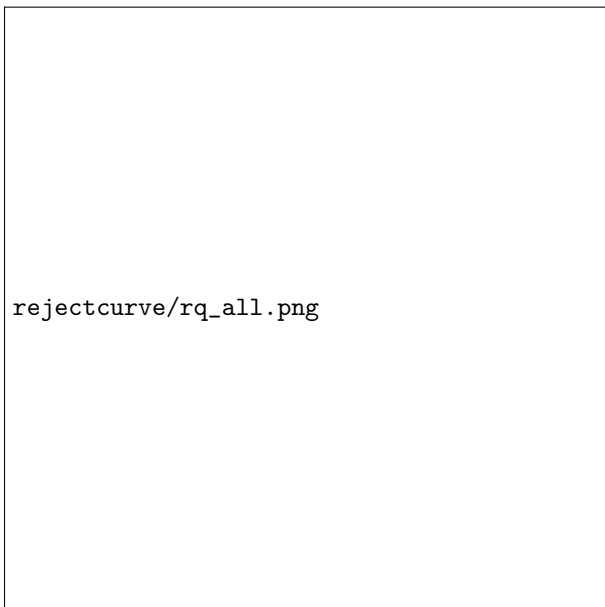


Рис. 4: Reject curves for the graph measures under study



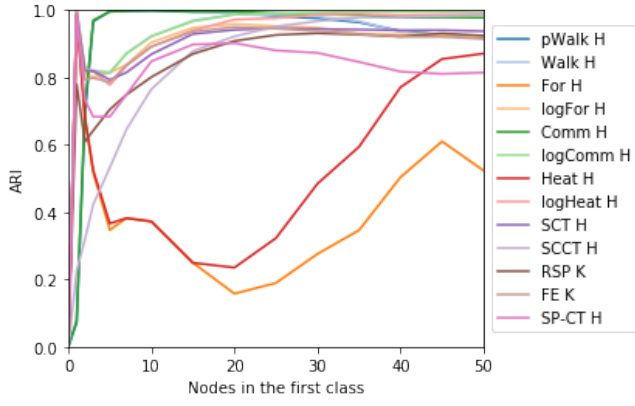
(a) All families



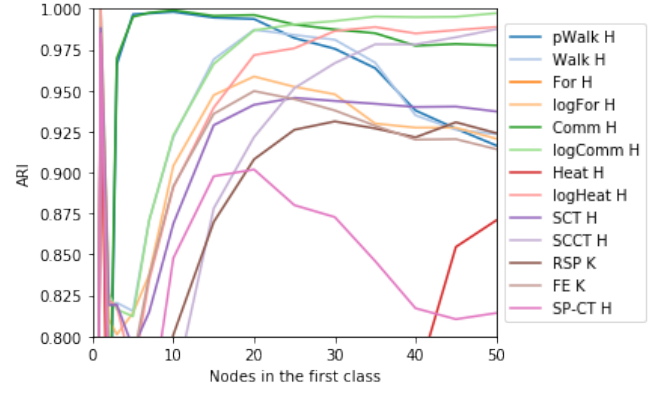
(b) The four families leading in Table ??(a)

Рис. 5: Average reject curves

4 Graphs with classes of different sizes



(a) All families



(b) Leading families

Рис. 6: Graphs with two classes of different sizes: clustering with optimal parameter values

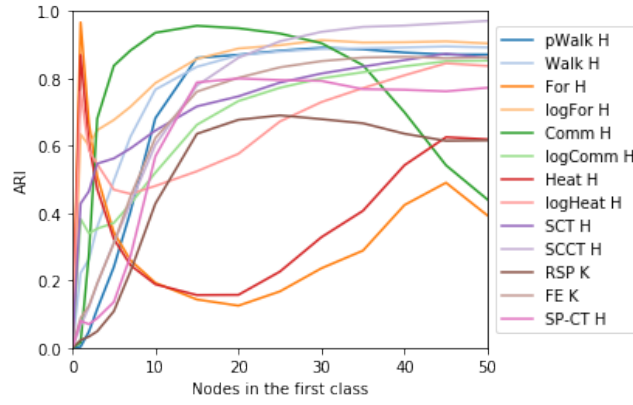


Рис. 7: Graphs with two classes of different sizes: random parameter values

$$P = \begin{pmatrix} 0.30 & 0.20 & 0.10 & 0.15 & 0.07 & 0.25 \\ 0.20 & 0.24 & 0.08 & 0.13 & 0.05 & 0.17 \\ 0.10 & 0.08 & 0.16 & 0.09 & 0.04 & 0.12 \\ 0.15 & 0.13 & 0.09 & 0.20 & 0.02 & 0.14 \\ 0.07 & 0.05 & 0.04 & 0.02 & 0.12 & 0.04 \\ 0.25 & 0.17 & 0.12 & 0.14 & 0.04 & 0.40 \end{pmatrix}.$$

AddFigures/sixClusters.png

.45.45ARI

of various measure families on a structure with 6 classes

5 Cluster analysis on several classical datasets

Dataset family	Dataset name	Number of nodes	Number of classes
Football	football	115	12
Political books	polbooks	105	3
Zachary	Zachary	34	2
Newsgroup	news_2cl_1	400	2
	news_2cl_2	400	2
	news_2cl_3	400	2
	news_3cl_1	600	3
	news_3cl_2	600	3
	news_3cl_3	600	3

Таблица 3: Overview of the datasets in the experiments



(a) football



(b) polbooks



(c) Zachary



(d) news_2cl_1



(e) news_2cl_2



(f) news_2cl_3



(g) news_2cl_4

	football	polbooks	Zachary	news_2cl_1	news_2cl_2	news_2cl_3	news_3cl_1	news_3cl_2	news_3cl_3
SCCT	-12	12	1	7	10	10	12	12	12
logComm	-1	5	1	12	12	12	10	10	10
logHeat	-1	1	1	7	3	8	2	8	8
FE	-1	-2	1	2	-1	6	8	0	0
RSP	-1	10	1	0	6	1	4	-2	-2
Walk	-1	5	1	4	8	-4	0	4	4
logFor	-1	-6	1	10	3	1	-4	6	6
SP-CT	-1	8	1	-3	-1	4	6	-4	-4
SCT	-1	-10	1	-3	-4	-2	-2	2	2
Comm	12	-6	1	-6	-6	-6	-6	-8	-8
pWalk	10	-6	1	-8	-8	-8	-8	-6	-6
Heat	-1	1	1	-10	-10	-11	-11	-10	-10
For	-1	-12	-12	-12	-12	-11	-11	-12	-12

Таблица 4: Copeland's scores of the measure families for the best parameter values

	football	polbooks	Zachary	news_2cl_1	news_2cl_2	news_2cl_3	news_3cl_1	news_3cl_2	news_3cl_3
logComm	0	10	3	10	12	8	4	10	10
SCCT	-10	8	3	12	8	12	6	12	12
FE	0	3	3	2	4	8	12	2	2
Walk	0	12	3	4	10	-4	0	4	4
logFor	0	3	3	8	4	-2	-2	6	6
SP-CT	0	3	3	0	-1	8	8	-2	-2
logHeat	0	-7	3	6	-1	2	2	8	8
RSP	-12	3	-8	-2	4	4	10	0	0
SCT	0	-7	3	-8	-4	0	-4	-4	-4
pWalk	11	-4	3	-6	-8	-8	-6	-7	-7
Comm	11	-12	3	-4	-6	-6	-8	-7	-7
Heat	0	-2	-11	-11	-11	-11	-11	-11	-11
For	0	-10	-11	-11	-11	-11	-11	-11	-11

Таблица 5: Copeland's scores of the measure families for 80th percentiles