

Supervised Link Prediction



Used algorithms and tools

- Matplotlib — AUC plotting
- Scikit Learn — Logistic Regression, AUC score
- Shortest nodes distance — BFS
- DFS — WCC (number, largest, its relative size)
- Structure — adjacency list (using maps)
- Percentile — map (distance => number of times)
- Formulas from the task, etc.



Results (part 1, RA)

Nodes number	167	167
Edges number	82927	82,927
Unique edges number	3251	
Density	0.23454296226823462	
WCCs number	1	
LWCC size	167	167
LWCC relative size	1	
LWCC average clustering coefficient	0.5918632085486949	
Assortativity coefficient	-0.2939402578188977	-0.295 177
LWCC exact d / r / p	5 3 3	5, -, (2.246 15)
LWCC estimated (random) d / r / p	4 1 3, 4 0 3, 4 1 3, 4 1 3 4 0 3	
LWCC estimated (snowball) d / r / p	5 3 3, 5 3 3, 5 3 3, 5 3 3 5 3 3	



Results (part 1, UC)

Nodes number	1899	1,899
Edges number	59835	59,835
Unique edges number	13838	
Density	0.007678601848568738	
WCCs number	4	
LWCC size	1893	1,893
LWCC relative size	0.9968404423380727	
LWCC average clustering coefficient	0.10974567163130962	
Assortativity coefficient	-0.1877757871466803	-0.187 776
LWCC exact d / r / p	8 4 4	8, -, (3.657 93)
LWCC estimated (random) d / r / p	5 1 4, 5 1 4, 5 1 4, 6 1 4, 5 1 4	
LWCC estimated (snowball) d / r / p	6 4 3, 6 4 3, 7 4 3, 6 4 4, 6 4 4	



Results (part 1, DG)

Nodes number	30398	30,398
Edges number	87627	87,627
Unique edges number	86312	
Density	0.00018682089211248675	
WCCs number	373	
LWCC size	29652	29,652
LWCC relative size	0.9754589117705112	
LWCC average clustering coefficient	0.005419329606719914	
Assortativity coefficient	0.0064967849033495265	0.004 633 88
LWCC exact d / r / p		12, -, (5.401 91)
LWCC estimated (random) d / r / p	7 2 6, 8 2 6, 8 2 6, 8 2 6, 8 2 6	
LWCC estimated (snowball) d / r / p	9 7 5, 9 7 5, 9 7 5, 10 7 5, 10 7 5	



Results (part 1, MA)

Nodes number	24818	24,818
Edges number	506550	506,550
Unique edges number	199973	
Density	0.0006493596467263749	
WCCs number	104	
LWCC size	24668	24,668
LWCC relative size	0.9939559996776534	
LWCC average clustering coefficient	0.3139940182622459	
Assortativity coefficient	-0.18356859055404015	-0.215 181
LWCC exact d / r / p		9, -, (3.668 56)
LWCC estimated (random) d / r / p	6 0 4, 6 2 4, 5 1 4, 5 2 4, 5 2 4	
LWCC estimated (snowball) d / r / p	7 5 3, 7 5 3, 7 5 4, 6 5 3, 7 5 3	



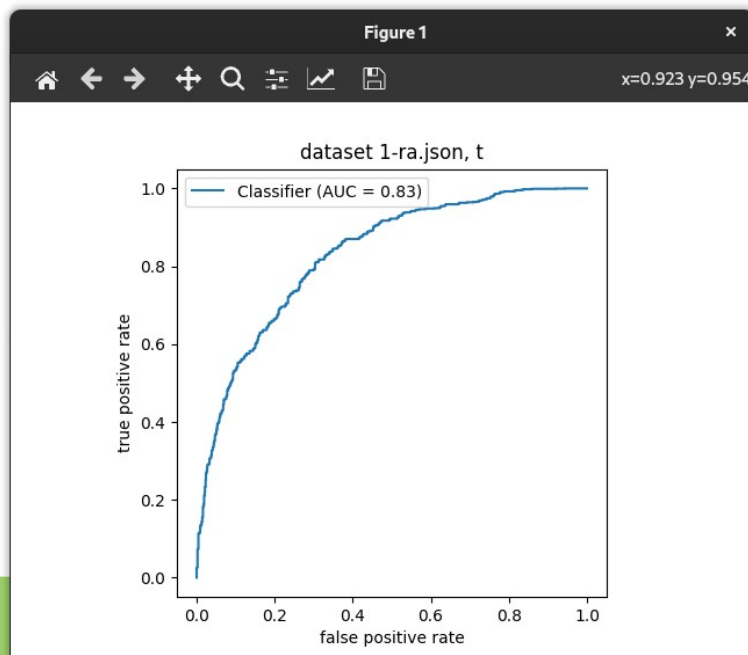
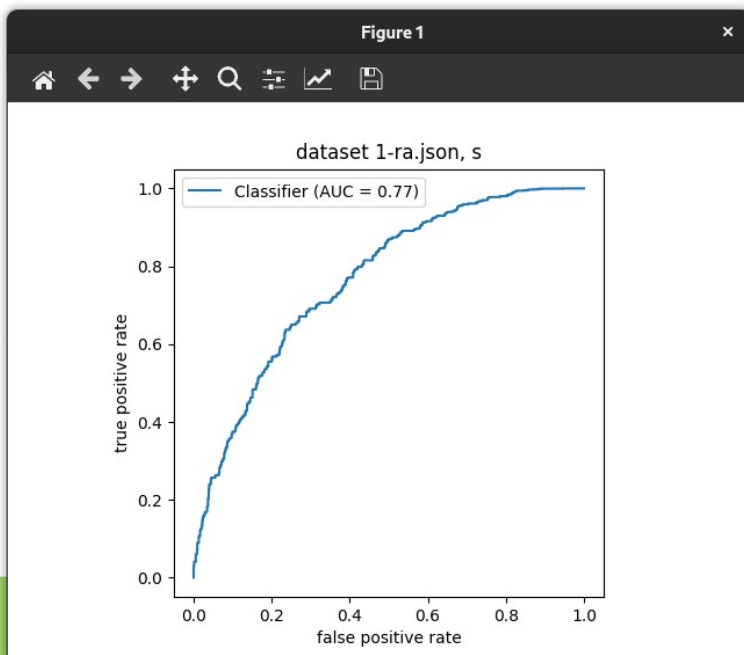
Results (part 1, AS)

Nodes number	159316	159,316
Edges number	964437	964,437
Unique edges number	508003	
Density	0.00004002950418644955	
WCCs number	4250	
LWCC size	152599	152,599
LWCC relative size	0.9578385096286626	
LWCC average clustering coefficient	0.11410799080556859	
Assortativity coefficient	-0.13234390518975114	-0.155 792
LWCC exact d / r / p		13, -, (4.482 80)
LWCC estimated (random) d / r / p	7 2 5, 7 2 5, 7 2 5, 8 2 5, 7 2 5	
LWCC estimated (snowball) d / r / p	9 7 4, 9 7 4, 10 7 4, 10 7 4, 9 7 4	



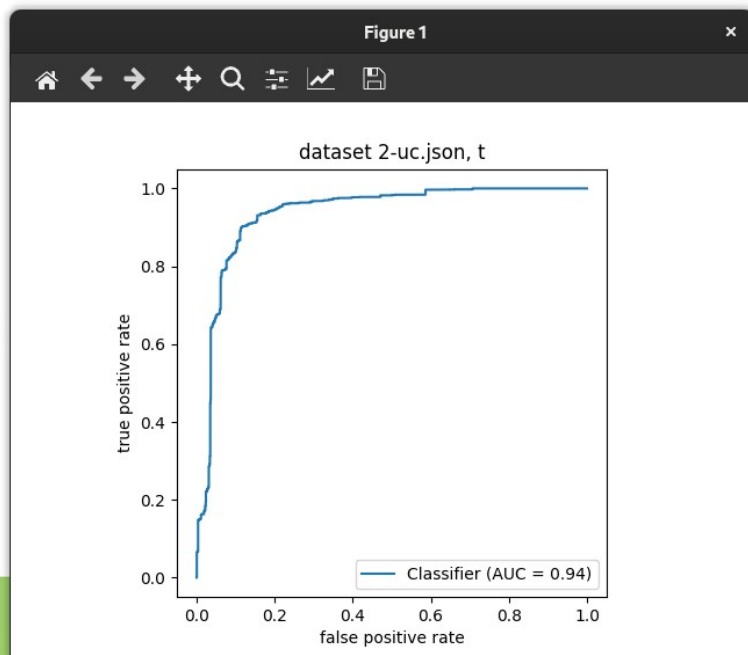
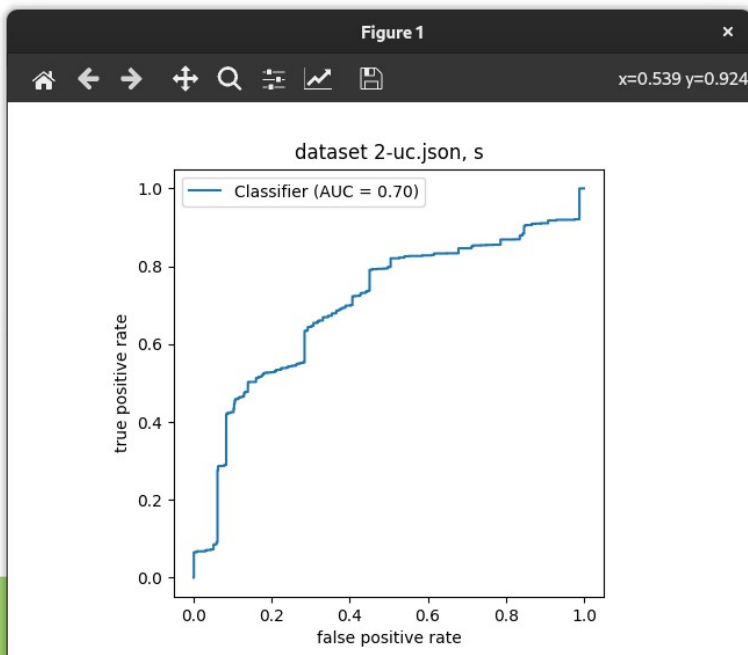
Results (part 2, RA)

	My results	Paper results
Static AUC	0.7660381700000001	0.864
Temporal AUC	0.8288871	0.852



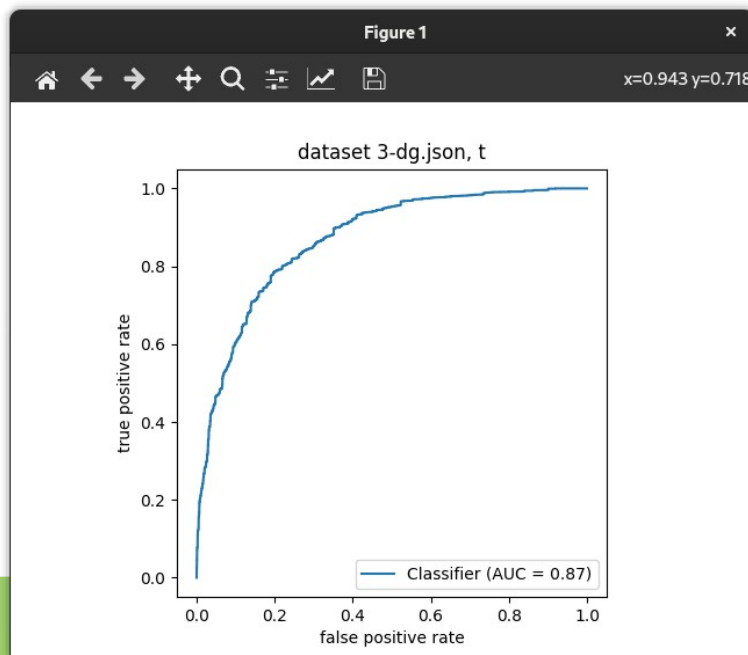
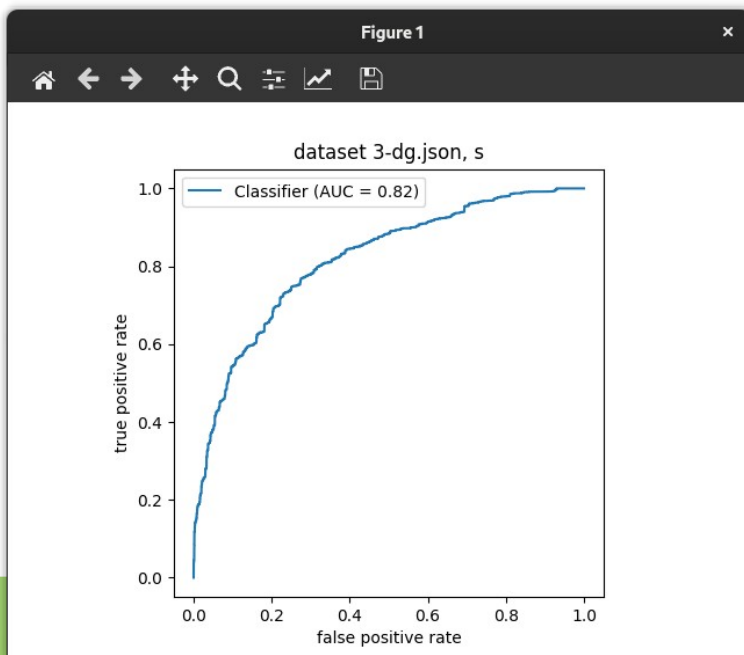
Results (part 2, UC)

	My results	Paper results
Static AUC	0.699394115	0.731
Temporal AUC	0.9357742	0.744



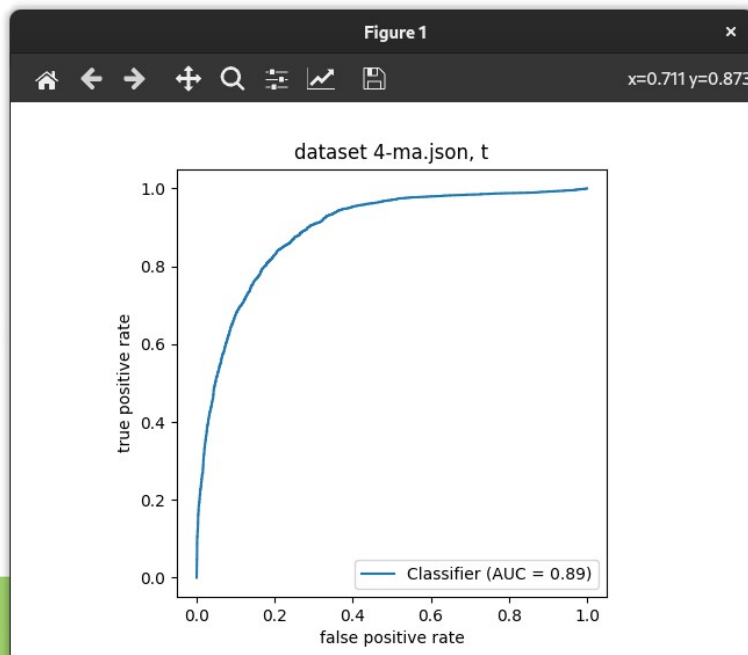
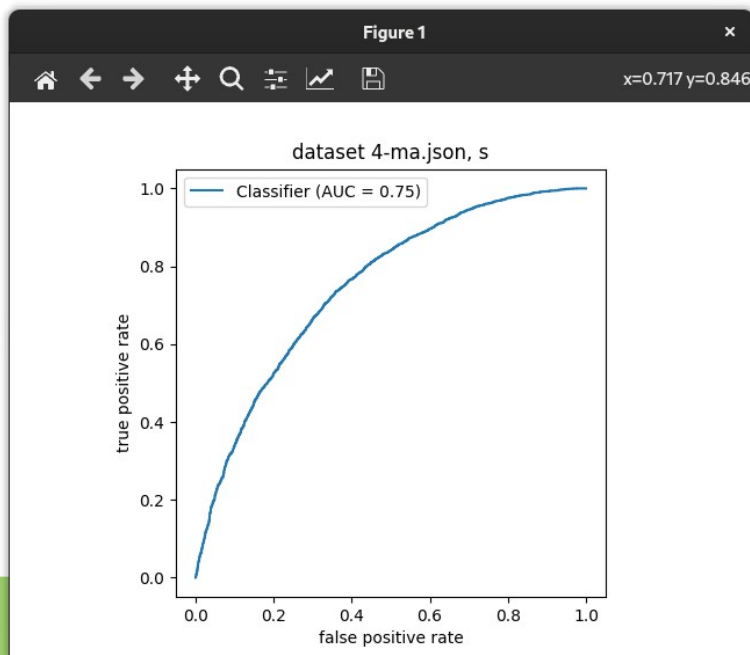
Results (part 2, DG)

	My results	Paper results
Static AUC	0.81619254	0.837
Temporal AUC	0.8695438799999999	0.866



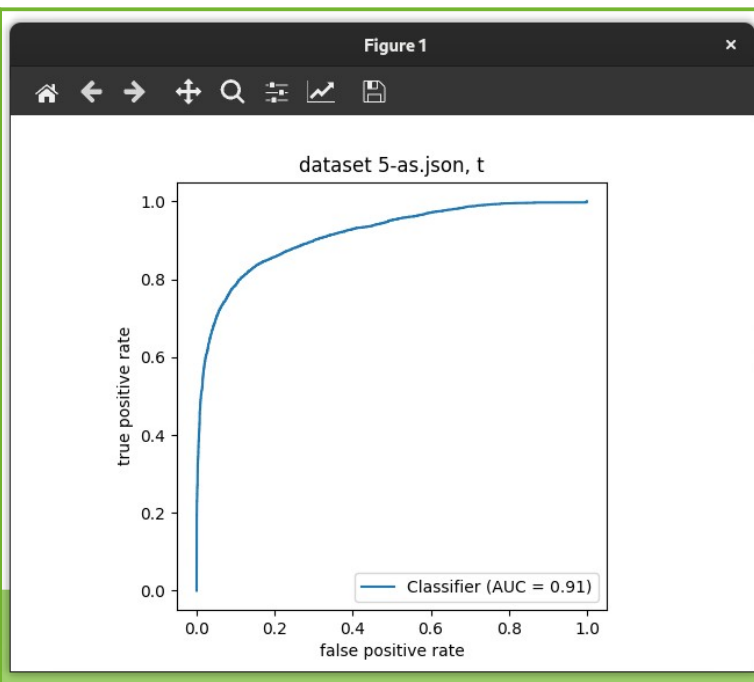
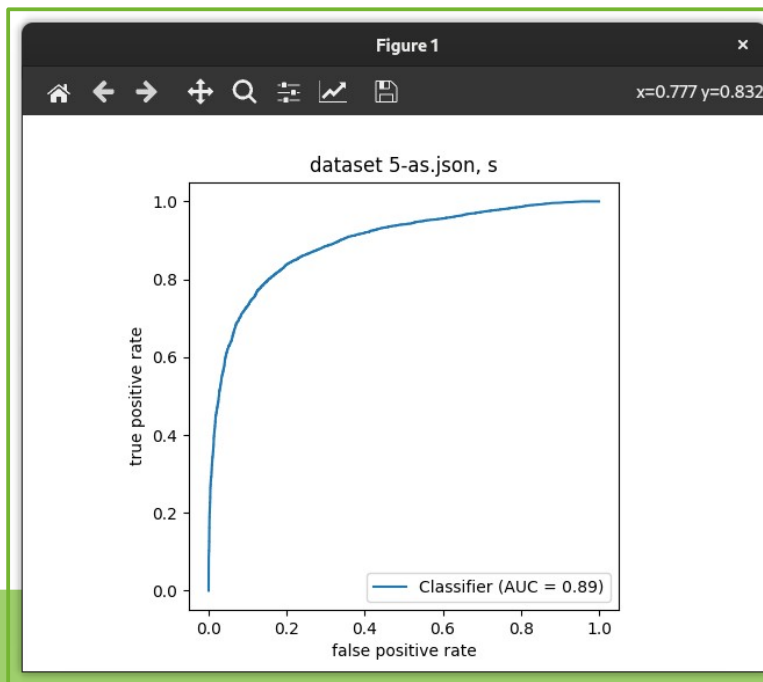
Results (part 2, MA)

	My results	Paper results
Static AUC	0.7494679399999999	0.859
Temporal AUC	0.8924620200000001	0.909



Results (part 2, AS)

	My results	Paper results
Static AUC	0.89480297	0.937
Temporal AUC	0.91464895	0.959



Conclusions

- For me, the temporal topological features worked better than the static ones
- The reason of the weird behavior of the UC dataset might be in its small number of nodes and small density
- The random nodes method estimates percentiles better, while radius and diameter are estimated better by the snowball method

