

How Much and When Do We Need Higher-order Information in Hypergraphs? A Case Study on Hyperedge Prediction (Supplementary)

SE-EUN YOON, HYUNGSEOK SONG, KIJUNG SHIN, and YUNG YI, Korea Advanced Institute of Science and Technology, South Korea

1 INTRODUCTION

This is the supplementary document for the paper published in TheWebConf 2020: *How Much and When Do We Need Higher-order Information in Hypergraphs? A Case Study on Hyperedge Prediction*. Here we provide: full statistics of datasets (Section 2), full experiment results in all test settings (Section 3), and additional experiments (Section 4).

2 DATASET STATISTICS

The hypergraph datasets are originally timestamped. However, we counted the number of unique hyperedges and weighted them by the number of occurrences. We considered hyperedges with number of nodes up to 10. For hyperedge prediction, our goal was to predict size 4, 5, and 10 hyperedges. In Table 1 we provide the number of all unique hyperedges and target size hyperedges.

We defined the n -projected graph (n -pg), which reflects the n -way relationships of the original hypergraph. Since we considered hyperedges of up to 10 nodes, we used projections of up to $n = 9$. We provide the statistics of n -pgs of each dataset, where the number of nodes are in Table 2 and edges in Table 3.

3 EXPERIMENT RESULTS

We considered 24 different types of candidate sets: 8 ways of constructing the negative set (2 types of negative hyperedges and 4 types of class imbalance); 2 different target hyperedge sizes (4 and 5); for size 4 hyperedge prediction, 2 to 3-order gain; for size 5 hyperedge prediction, 2 to 3-order gain and 3 to 4-order gain. We provide the results in all 24 settings.

3.1 Performance gains

Our performance metric was the AUC-PR, and we computed its percentage gain from order n to $n + 1$. We examined the prediction performance of 6 different features, which are geometric mean (GM), harmonic mean (HM), arithmetic mean (AM), common neighbors (CN), Jaccard’s coefficient (JC), and Adamic-Adar index (AA). We show in Tables 4 to 11 the results for each of the 8 combinations of negative hyperedge types.

3.2 Performance comparisons

We provide plots that compare the absolute performance between order n and $n + 1$. That is, for size 4 prediction, we compared order 2 and 3; for size 5 prediction, we compared order 2 and 3, order 3 and 4. We averaged the performance across features. Figure 1 shows the results for all possible combinations. We can see that there is increase of performance, and this increase is less for 3 and 4 than for 2 and 3. We also note that the absolute performance is lower for cliques and larger class imbalances. Gain is larger for smaller absolute performance.

3.3 Interpretations

We investigated the correlation of 2 to 3 performance gains with edge density in 3-pg, conditional entropy, and mutual information. We display the results for all cases in Figure 2. We can see its negative correlation with conditional entropy, and positive correlation with the other two measures.

4 ADDITIONAL EXPERIMENTS

We perform additional experiments on: running time for generating n -pgs, (Figure 12), number of n -way interactions (Table 13) and their correlation with diminishing returns (Figure 3).

Table 1. # Unique hyperedges in hypergraphs

Dataset	All	Size 4	Size 5	Size 10
email-Enron	1491	138	63	11
contact-primary-school	12704	347	9	0
contact-high-school	7818	222	7	0
NDC-classes	901	125	94	25
NDC-substances	8167	535	500	305
email-Eu	24223	2294	1359	188
threads-ask-ubuntu	166995	4560	1117	1
congress-bills	57887	7764	5780	2837
DAWN	137417	29829	15690	1158
tags-ask-ubuntu	147222	59158	25475	0
tags-math-sx	170476	50892	29244	0
threads-math-sx	595648	49198	16402	83
coauth-MAG-History	891296	19120	8775	1025
coauth-MAG-Geology	1189770	159509	99500	7480
coauth-DBLP	2454734	419434	205970	6032

Table 2. # Nodes in n -projected graphs

Dataset	2-pg	3-pg	4-pg	5-pg	6-pg	7-pg	8-pg	9-pg
email-Enron	143	1210	2871	4261	4369	3150	1594	537
contact-primary-school	242	4548	1243	45	0	0	0	0
contact-high-school	327	2785	751	35	0	0	0	0
NDC-classes	1140	3556	7227	10110	10352	7618	3908	1315
NDC-substances	2647	26367	77607	121113	126192	92521	47094	15779
email-Eu	967	16552	45841	72605	76556	56342	28799	9697
threads-ask-ubuntu	90045	101113	39055	14782	5368	1750	456	81
congress-bills	1715	176665	808669	1347305	1373563	971212	478088	155394
DAWN	2262	88643	468811	801356	735793	471601	216952	67096
tags-ask-ubuntu	3021	126518	248596	112478	0	0	0	0
tags-math-sx	1627	88367	222853	118792	0	0	0	0
threads-math-sx	153657	801423	586961	313315	154359	67389	23664	5967
coauth-MAG-History	443845	569179	654471	685680	583921	378280	177840	56149
coauth-MAG-Geology	1053582	4048699	6100315	6534570	5331402	3276336	1460431	438942
coauth-DBLP	1796914	6716702	8288725	7315793	5248130	3001570	1287801	378122

Table 3. # Edges in n -projected graphs

Dataset	2-pg	3-pg	4-pg	5-pg	6-pg	7-pg	8-pg	9-pg
email-Enron	1442	8916	25938	43960	47640	33894	15484	4140
contact-primary-school	8317	15417	2286	90	0	0	0	0
contact-high-school	5818	7110	1428	70	0	0	0	0
NDC-classes	3727	21885	61176	104150	115020	82761	37548	9792
NDC-substances	26973	234240	729012	1266080	1392090	995358	451108	118296
email-Eu	21465	143238	440916	771340	851595	610995	277928	73044
threads-ask-ubuntu	186955	181881	116046	64840	30435	11172	2940	504
congress-bills	178647	2439960	8117514	13784890	14634345	10120068	4448080	1132776
DAWN	97046	1456683	4917996	7484560	7188420	4637850	1942360	476856
tags-ask-ubuntu	132703	838107	874056	254750	0	0	0	0
tags-math-sx	91685	748644	936774	292440	0	0	0	0
threads-math-sx	1083531	2184567	2174994	1707610	1089660	534177	183792	38808
coauth-MAG-History	723382	2101608	4226058	5924860	5748105	3789534	1624140	409140
coauth-MAG-Geology	4241817	18870564	40067280	54239790	49990305	31354260	12842900	3104676
coauth-DBLP	7123888	26398201	46071251	54373952	46348672	27871398	11142960	2649988

Table 4. Clique 1:1

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	2.08	0.59	-0.92	2.23	-1.42	2.69	5.99	14.5	6.66	17.88	1.38	15.18	1.39	-0.89	-0.58	4.03	1.17	6.4
email-Eu	0.17	-0.01	1.51	21.99	3.72	20.92	0.97	-0.02	1.86	21.71	4.0	21.58	0.1	-0.22	1.27	0.06	4.53	0.27
contact-primary-school	0.32	0.59	3.98	62.56	52.55	61.79	0.0	0.0	0.0	68.34	139.3	40.02	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	3.43	-0.08	15.98	56.37	56.62	58.67	0.0	0.0	0.0	153.07	26.7	92.7	0.0	0.0	0.0	0.0	78.34	15.06
NDC-classes	0.1	0.98	-25.22	0.56	3.82	0.34	-30.93	8.48	-11.8	0.41	4.98	0.83	-0.88	0.96	-21.07	-1.73	2.55	-1.59
NDC-substances	3.08	-0.02	-2.89	8.52	8.97	7.58	22.26	0.62	1.37	12.43	18.76	12.23	3.22	-0.01	4.47	0.06	1.09	-0.41
DAWN	0.01	0.01	4.27	21.93	8.82	21.01	0.01	-0.0	3.65	28.15	12.14	26.9	0.13	0.0	-1.37	0.21	1.78	0.48
congress-bills	1.29	-0.45	1.99	47.26	8.62	42.66	2.41	0.0	8.01	45.98	29.02	43.94	-0.01	0.43	-0.02	0.08	-0.32	0.03
tags-ask-ubuntu	0.05	-0.01	5.66	24.75	8.75	23.36	0.01	0.01	7.01	28.87	15.25	27.42	0.0	-0.05	-0.23	0.05	0.47	0.08
tags-math-sx	0.08	0.11	5.51	18.02	11.32	16.98	0.01	0.0	5.91	17.51	10.49	16.49	0.0	0.14	0.46	0.08	0.61	0.08
threads-ask-ubuntu	0.01	14.26	2.22	1.94	1.7	1.41	-0.0	0.86	3.59	2.37	2.44	1.86	-0.0	11.36	-0.06	-0.02	-0.05	-0.02
threads-math-sx	0.06	1.86	7.02	6.75	6.16	5.92	0.01	-0.32	10.27	7.81	7.35	6.86	0.0	1.14	-0.0	-0.0	-0.21	-0.0
coauth-MAG-History	-44.48	-5.72	0.29	0.75	0.79	0.72	-41.76	0.02	1.97	1.08	1.2	0.9	2.95	47.78	0.7	-0.17	0.11	-0.05
coauth-MAG-Geology	-26.63	0.08	13.74	3.44	6.37	3.34	-29.12	0.63	76.28	4.63	6.25	4.09	24.48	0.03	3.37	0.08	0.57	0.2
coauth-DBLP	-16.31	0.21	22.68	2.62	4.78	2.55	-21.36	0.55	26.29	4.14	4.91	3.52	16.83	0.13	2.97	0.09	0.43	0.17

Table 5. Clique 1:2

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	7.23	-0.03	-0.0	6.58	-7.68	7.72	5.41	3.44	5.29	4.13	0.41	6.9	-1.59	0.24	-3.68	-0.98	0.32	-1.41
email-Eu	0.0	0.0	1.41	40.41	4.84	38.55	0.83	0.06	2.45	42.35	4.55	41.8	0.39	-0.36	1.25	-0.03	9.5	0.18
contact-primary-school	1.69	-0.53	14.91	125.29	103.65	123.39	0.0	0.0	5.35	153.12	162.02	101.52	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	3.35	-0.02	15.91	97.63	127.81	95.3	0.0	-5.63	14.83	349.76	215.28	349.76	0.0	5.96	0.0	0.0	21.76	0.0
NDC-classes	-0.13	1.53	-8.79	0.86	8.16	0.69	-39.6	10.88	-14.23	0.75	5.77	1.71	-0.64	8.22	1.17	-3.58	1.57	-3.42
NDC-substances	0.71	-0.01	-0.2	7.02	1.13	4.29	23.78	0.93	2.07	19.94	27.9	18.04	66.05	-0.05	7.73	-2.74	-1.24	-1.99
DAWN	0.02	0.01	7.7	42.98	14.24	41.07	0.02	0.0	5.22	52.9	19.96	50.69	0.19	-0.0	-0.46	0.19	3.58	0.71
congress-bills	2.77	-0.57	4.79	90.87	14.64	82.78	4.27	-0.01	13.57	81.35	43.59	77.6	0.0	1.83	0.02	0.12	-0.34	0.02
tags-ask-ubuntu	0.08	-0.1	10.1	49.92	14.86	46.64	0.02	0.02	12.21	59.31	30.82	56.22	0.0	-0.4	-0.29	0.11	0.94	0.17
tags-math-sx	0.14	0.18	10.27	34.16	19.81	32.13	0.04	0.01	9.41	33.73	19.79	31.67	0.0	0.06	1.15	0.12	1.21	0.14
threads-ask-ubuntu	0.01	15.96	2.8	2.65	2.38	1.82	-0.0	-0.61	5.3	3.09	3.24	2.39	-0.0	15.26	-0.05	0.0	-0.06	-0.0
threads-math-sx	0.06	1.68	9.08	8.07	7.28	7.01	0.01	-0.45	15.3	9.37	8.83	8.18	0.0	1.17	0.0	-0.0	-0.25	-0.0
coauth-MAG-History	-48.97	-7.45	-3.05	1.08	1.32	1.05	-40.99	0.06	2.08	1.54	1.99	1.3	0.23	75.06	1.53	-0.22	0.19	-0.07
coauth-MAG-Geology	-38.32	0.22	18.74	4.1	10.22	4.41	-38.8	1.17	142.51	6.46	10.07	6.13	35.27	0.14	4.39	-0.05	0.88	0.55
coauth-DBLP	-26.2	0.39	41.42	4.13	8.43	4.14	-32.83	1.03	47.84	6.17	8.63	5.58	34.32	0.24	4.56	0.12	0.71	0.42

Table 6. Clique 1:5

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	2.22	0.3	2.98	23.15	-3.46	27.77	7.79	6.36	4.75	54.18	0.89	65.79	-1.66	-0.93	-3.5	1.67	-1.02	-1.92
email-Eu	-0.04	-0.0	2.44	90.53	7.5	86.73	-1.49	-0.09	2.99	91.25	8.03	90.08	1.38	-1.33	0.82	1.63	16.68	2.17
contact-primary-school	2.78	0.96	26.69	277.38	239.91	270.25	0.0	1.69	6.12	472.16	222.85	439.44	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	7.01	-1.24	40.67	266.17	259.16	261.82	0.0	0.0	4.31	801.92	174.66	796.04	0.0	5.96	0.0	0.0	43.66	0.0
NDC-classes	0.19	2.2	-21.35	1.33	11.52	2.25	-62.3	18.65	-32.79	0.77	8.03	2.83	7.73	15.72	-1.74	-3.42	2.38	-1.72
NDC-substances	-3.47	0.02	0.65	5.64	-21.36	2.03	15.39	1.93	0.82	13.8	9.88	12.28	117.92	-0.04	8.99	0.73	1.25	0.26
DAWN	0.07	0.03	15.87	102.26	23.07	97.34	0.08	0.0	6.08	121.78	33.69	116.86	0.26	0.0	4.64	-0.39	8.29	0.18
congress-bills	5.36	-0.89	9.35	192.57	16.2	173.57	8.12	-0.04	22.78	165.38	49.78	158.34	-0.13	3.32	-0.08	4.48	-1.06	0.12
tags-ask-ubuntu	0.2	-0.3	18.18	117.08	16.39	106.94	0.02	0.03	17.79	135.85	58.14	127.15	0.0	-1.14	-0.1	0.36	2.55	0.56
tags-math-sx	0.3	0.22	20.94	77.07	34.97	71.99	0.06	0.01	16.35	78.71	40.27	73.46	0.0	0.5	3.96	0.34	3.0	0.37
threads-ask-ubuntu	0.01	12.91	2.75	2.73	2.58	1.77	-0.0	-1.91	6.49	3.08	3.66	2.28	-0.0	13.16	-0.06	0.01	-0.07	-0.0
threads-math-sx	0.04	1.01	9.79	7.61	7.02	6.48	0.01	-0.16	21.43	8.55	8.27	7.33	0.0	0.86	0.0	0.0	-0.21	-0.01
coauth-MAG-History	0.01	-9.29	-4.99	1.48	2.26	1.55	0.07	0.16	2.3	2.08	3.34	1.89	0.32	110.03	2.62	-0.26	0.36	-0.1
coauth-MAG-Geology	1.0	0.41	22.64	14.16	16.05	13.58	0.86	2.41	323.09	7.05	16.17	12.63	-23.28	0.28	7.11	0.25	1.41	0.47
coauth-DBLP	-42.92	0.71	85.96	4.51	15.59	13.81	1.31	2.21	101.23	7.94	16.0	12.68	-17.03	0.45	8.42	0.16	1.29	0.36

Table 7. Clique 1:10

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	12.67	0.49	-0.15	33.89	-1.57	35.78	14.37	22.95	5.95	10.73	-2.98	17.01	-1.53	-0.59	-0.26	5.04	1.59	-2.05
email-Eu	0.69	-0.13	1.67	153.81	9.16	148.09	-2.44	-0.78	4.04	158.79	13.93	157.73	0.64	-0.32	1.23	2.01	18.86	2.31
contact-primary-school	6.42	1.21	49.2	495.94	413.35	484.73	0.0	0.0	6.63	708.56	361.79	627.66	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	15.16	-0.87	78.62	515.17	455.54	507.13	0.0	0.0	14.14	1623.33	221.51	1617.75	0.0	5.96	3.9	0.0	104.37	0.0
NDC-classes	0.18	4.23	-44.7	1.55	10.91	2.25	44.28	16.62	-37.82	0.28	10.07	5.6	6.77	16.95	-2.31	-4.14	2.0	-1.5
NDC-substances	-4.95	-0.02	0.47	0.57	-40.98	-3.54	10.73	2.46	0.16	0.61	-17.02	14.18	158.1	0.02	7.07	-1.81	-0.47	-2.99
DAWN	0.15	0.04	21.34	197.97	30.48	187.62	0.23	0.0	3.48	220.79	42.8	212.33	0.49	0.0	14.04	-0.85	17.31	-0.54
congress-bills	7.92	-0.99	14.53	328.76	16.49	294.16	11.84	-0.03	30.86	271.64	48.55	259.22	-0.07	4.98	0.26	0.93	0.57	0.16
tags-ask-ubuntu	0.24	-0.51	23.09	216.47	14.07	192.03	0.07	0.02	20.84	244.72	80.37	225.89	0.0	-1.13	2.96	0.85	5.5	1.35
tags-math-sx	0.46	0.18	32.38	137.4	46.53	127.25	0.13	0.01	21.35	146.02	60.64	135.54	0.0	0.63	9.73	0.67	5.86	0.74
threads-ask-ubuntu	0.0	10.05	2.47	2.34	2.34	1.48	-0.0	-1.76	6.51	2.56	3.1	1.76	-0.0	9.62	-0.07	0.01	-0.05	0.0
threads-math-sx	0.03	0.44	8.52	6.01	5.61	5.1	0.0	0.42	23.48	6.63	6.56	5.65	0.0	0.61	0.01	0.0	-0.15	0.0
coauth-MAG-History	0.0	-8.48	-1.81	1.69	3.0	1.94	0.08	0.13	3.0	2.32	4.43	2.37	0.16	137.94	2.36	-0.23	0.53	-0.1
coauth-MAG-Geology	0.93	0.36	22.93	15.39	19.76	15.34	0.79	3.78	603.02	14.56	20.52	14.65	-30.08	0.17	8.14	-0.1	1.68	0.3
coauth-DBLP	-53.43	0.9	145.31	16.89	21.99	16.73	1.32	3.52	175.24	16.17	22.82	15.44	-24.05	0.58	11.06	-0.08	1.69	0.3

Table 8. Star 1:1

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	-1.11	-0.11	2.77	21.99	14.78	20.62	0.0	5.42	2.13	48.12	29.7	48.66	0.0	-7.96	0.16	3.04	0.86	1.98
email-Eu	0.07	0.05	1.41	26.43	16.69	24.54	0.06	-0.05	1.78	27.2	13.35	24.92	-0.0	1.4	0.86	-0.43	1.08	-0.34
contact-primary-school	0.1	2.56	4.64	41.12	20.12	39.4	0.0	0.0	0.0	0.0	6.78	0.0	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	0.03	-5.6	10.77	30.39	33.36	30.26	0.0	0.0	0.0	181.41	170.75	178.61	0.0	0.0	0.0	0.0	0.0	0.0
NDC-classes	-28.51	0.05	-0.07	2.76	-0.07	6.86	-4.99	-0.38	23.49	13.48	-28.82	6.95	-0.93	-0.3	-4.28	-1.28	-2.19	0.29
NDC-substances	-0.4	-1.04	-1.35	6.85	-3.97	4.69	0.03	-3.02	-3.9	0.92	1.32	-0.5	3.05	1.34	-0.21	2.73	0.01	1.28
DAWN	0.0	0.03	2.16	13.37	10.02	12.86	0.0	0.0	2.25	14.48	10.79	13.89	0.02	0.01	-1.24	0.03	0.06	0.1
congress-bills	0.9	0.59	2.26	27.78	15.8	25.03	0.6	-0.0	3.25	25.22	13.48	22.83	0.01	3.2	0.08	0.13	1.11	0.07
tags-ask-ubuntu	0.01	0.91	2.24	14.71	9.93	13.99	0.0	0.35	2.07	16.68	10.0	15.85	-0.0	0.6	0.0	0.0	0.21	0.0
tags-math-sx	0.03	0.71	2.12	12.55	12.62	11.72	0.0	0.03	1.79	13.04	11.64	12.08	0.0	0.7	0.02	0.01	0.39	0.01
threads-ask-ubuntu	-38.51	1.44	-42.45	1.51	0.13	1.6	-66.36	0.13	17.55	0.07	-0.02	0.04	49.07	1.14	0.01	0.0	0.0	0.0
threads-math-sx	0.02	-0.39	5.09	0.35	1.02	0.48	0.0	-0.4	7.35	0.12	0.76	0.27	0.0	0.04	0.02	0.05	-0.01	0.0
coauth-MAG-History	-29.95	-0.16	-40.62	-0.12	0.02	-0.07	-23.98	-2.02	-34.62	-0.07	-0.02	-0.05	5.07	1.53	8.27	0.01	0.01	0.04
coauth-MAG-Geology	41.97	0.51	2.42	0.37	-0.02	0.1	45.62	-0.29	1.89	0.27	-0.11	-0.0	2.63	4.48	0.81	-25.05	0.02	-19.39
coauth-DBLP	21.5	3.2	0.46	-0.08	-0.18	-0.32	26.94	0.12	0.03	-6.95	-14.34	-0.37	1.07	7.4	1.19	6.7	-0.53	-11.0

Table 9. Star 1:2

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	0.88	-1.77	-1.62	50.74	32.62	46.28	-0.95	5.23	-1.72	84.94	70.75	83.36	-0.07	0.49	-0.52	1.79	7.59	1.67
email-Eu	-0.01	0.08	2.07	47.27	26.37	43.69	0.28	-0.06	4.86	48.01	25.54	43.68	-0.0	2.23	1.66	-0.57	2.31	-0.47
contact-primary-school	0.39	1.66	8.2	71.03	40.4	69.59	0.0	0.0	0.0	71.01	19.06	0.0	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	0.59	4.09	17.42	69.68	63.34	68.45	0.0	0.0	0.0	2.56	308.26	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NDC-classes	-39.75	-4.61	0.55	1.58	-0.42	6.97	-5.98	-1.42	34.63	11.31	-34.1	14.93	2.61	-0.08	-7.95	4.4	4.24	3.48
NDC-substances	-0.45	-0.51	-2.66	14.26	-0.74	10.43	-0.25	-0.48	-5.37	7.74	0.4	5.68	2.71	2.28	3.28	4.7	-1.44	2.91
DAWN	0.02	0.08	4.27	26.16	19.08	25.19	0.01	0.01	3.97	28.12	20.22	27.01	0.03	0.02	-1.92	0.02	0.26	0.14
congress-bills	1.51	0.78	4.03	54.84	31.03	49.67	1.12	0.11	6.49	52.51	26.47	47.91	0.01	4.9	-0.01	0.27	1.15	0.12
tags-ask-ubuntu	0.02	1.98	4.01	28.06	18.59	26.63	0.0	0.4	3.52	32.66	21.33	31.14	0.0	1.09	-0.0	0.0	0.36	0.01
tags-math-sx	0.04	0.92	4.28	21.63	21.25	20.19	0.0	0.06	3.52	23.81	21.77	22.09	0.0	1.19	0.09	0.02	0.68	0.02
threads-ask-ubuntu	-54.9	0.48	-56.79	1.52	0.18	-61.1	-74.76	-0.67	21.13	-1.32	0.02	0.04	0.0	0.84	0.02	0.0	0.01	0.0
threads-math-sx	0.02	-0.68	7.41	0.5	1.35	0.68	0.0	-0.33	9.18	0.12	0.83	0.3	0.0	0.04	0.04	0.06	-0.0	0.01
coauth-MAG-History	-46.72	-1.78	-54.46	-0.19	0.02	-0.17	-30.35	-4.84	-0.15	-0.19	-0.02	-0.12	0.28	20.68	-1.43	0.04	0.02	0.06
coauth-MAG-Geology	57.83	1.09	1.94	0.3	-0.03	0.02	67.24	-0.3	1.87	0.22	-0.14	-0.04	3.15	6.72	1.15	-36.11	0.02	-30.3
coauth-DBLP	28.44	4.62	0.51	-0.12	-0.18	-0.35	38.98	0.36	-0.0	-9.83	-23.36	-0.38	3.71	10.92	1.86	7.41	-1.08	-20.89

Table 10. Star 1:5

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	-5.07	0.14	-2.25	92.31	36.06	85.01	0.07	13.41	8.08	195.35	84.82	185.09	0.0	-0.85	-3.37	-0.43	18.24	0.0
email-Eu	-0.2	-0.12	3.3	97.01	48.09	88.51	-0.6	0.59	6.13	117.55	55.64	108.53	0.27	1.93	1.42	0.54	5.72	0.59
contact-primary-school	0.37	2.95	13.31	176.81	95.16	170.27	0.0	-19.85	0.0	197.38	11.72	60.04	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	2.22	3.23	45.49	153.38	142.08	152.8	0.0	0.0	30.43	2.56	133.22	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NDC-classes	-64.65	2.4	1.72	1.66	-0.35	4.89	16.0	-4.23	37.32	12.12	-45.64	10.91	-2.56	3.81	-2.76	6.39	-14.93	0.08
NDC-substances	-0.27	-0.61	-4.05	24.95	-2.81	22.68	-0.14	-0.52	-6.51	10.56	-0.84	8.9	5.17	2.54	5.51	5.71	-0.73	4.39
DAWN	0.05	0.07	9.25	57.9	37.11	55.49	0.02	0.01	8.36	63.99	41.56	61.6	0.03	0.06	-2.1	-0.07	0.65	0.09
congress-bills	3.19	1.21	7.9	121.5	64.54	109.71	2.61	0.14	13.62	117.18	59.28	107.3	0.04	8.93	0.16	0.99	2.71	0.64
tags-ask-ubuntu	0.07	2.87	9.62	65.11	36.44	61.43	0.01	0.52	7.41	76.02	44.94	71.99	-0.0	2.25	-0.05	0.03	0.97	0.0
tags-math-sx	0.08	1.36	9.55	46.53	37.87	43.34	-0.01	0.08	7.86	50.43	42.0	46.61	-0.0	2.62	0.3	0.04	1.56	0.04
threads-ask-ubuntu	-72.76	-0.77	-72.91	1.57	0.18	1.34	-85.56	-3.18	16.01	0.04	-0.01	0.02	98.66	0.28	0.05	0.0	0.0	0.0
threads-math-sx	0.01	-1.11	9.32	0.49	1.49	0.72	0.0	-0.25	13.16	0.19	0.98	0.41	0.0	0.01	0.03	0.07	-0.01	0.01
coauth-MAG-History	-69.3	-8.0	0.27	-0.21	0.01	-0.2	-43.35	-8.45	-0.82	-0.22	-0.02	-0.14	6.28	25.65	-0.33	0.01	0.02	0.06
coauth-MAG-Geology	70.12	2.35	-0.08	0.16	-0.03	-0.02	91.79	0.31	1.16	0.13	-0.12	-0.04	6.5	9.31	3.18	-50.6	0.02	-42.75
coauth-DBLP	30.22	6.33	0.17	-0.13	-0.14	-0.32	45.91	0.77	0.04	-13.19	-40.17	-0.31	11.37	15.67	3.07	4.94	-0.93	-39.9

Table 11. Star 1:10

Dataset	Size 4 prediction						Size 5 prediction											
	2 to 3 gain (%)						2 to 3 gain (%)						3 to 4 gain (%)					
	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA	GM	HM	AM	CN	JC	AA
email-Enron	-0.29	0.11	0.03	146.62	61.25	136.85	0.94	20.44	-0.95	290.71	114.69	284.58	0.15	-12.38	-0.65	-1.24	40.45	1.24
email-Eu	-0.23	-0.36	4.84	179.15	78.08	165.82	-0.66	1.16	9.65	197.94	80.12	182.04	0.67	2.62	3.72	1.5	11.06	1.64
contact-primary-school	0.83	2.31	26.38	352.97	191.32	341.63	0.0	0.0	0.0	363.57	15.84	231.52	0.0	0.0	0.0	0.0	0.0	0.0
contact-high-school	4.1	0.73	62.55	276.95	268.14	270.9	0.0	0.0	44.58	2.56	146.33	0.0	0.0	5.96	0.0	0.0	0.0	0.0
NDC-classes	-16.41	2.37	6.82	1.58	0.08	4.64	25.08	-1.22	29.13	11.93	-56.56	13.5	4.6	4.26	-0.09	7.19	-21.12	-1.25
NDC-substances	-0.47	-0.42	-1.9	45.49	-2.65	42.64	-0.0	-2.58	-2.34	15.7	-0.78	14.22	-0.03	2.01	-1.82	5.72	-0.22	5.4
DAWN	0.1	0.06	15.34	109.31	53.4	104.23	0.05	0.01	12.52	117.02	65.69	112.87	0.06	0.06	-0.75	-0.22	1.5	-0.09
congress-bills	5.4	0.96	13.31	220.02	94.64	199.34	4.62	0.03	22.55	211.05	93.92	194.34	0.06	12.26	0.23	1.59	3.5	1.06
tags-ask-ubuntu	0.13	3.64	15.56	121.91	54.16	113.52	0.01	0.45	10.38	142.59	73.4	133.63	0.0	3.74	0.03	0.0	2.15	0.21
tags-math-sx	0.15	1.43	16.18	79.17	52.63	73.45	-0.01	0.09	11.77	87.29	61.81	80.33	-0.0	4.1	0.92	0.11	3.08	0.13
threads-ask-ubuntu	-83.81	-1.74	-82.3	1.55	0.13	0.06	-91.55	-3.4	15.73	0.03	0.01	0.02	77.37	0.2	0.21	0.0	0.0	0.0
threads-math-sx	0.01	-1.41	10.38	0.53	1.3	0.77	0.0	-0.17	14.0	0.03	0.84	0.32	0.0	-0.0	0.1	0.11	-0.0	0.02
coauth-MAG-History	-82.43	-8.23	-0.13	-0.21	0.01	-0.17	-62.3	-10.2	-0.28	-0.2	-0.02	-0.11	3.74	23.56	0.76	0.04	0.02	0.05
coauth-MAG-Geology	65.97	3.41	-2.43	0.08	-0.03	-0.02	96.11	0.58	0.17	0.07	-0.09	10.87	10.55	5.35	-60.83	0.02	-55.89	
coauth-DBLP	22.64	7.75	-0.11	-0.09	-0.11	-0.25	37.39	1.04	0.27	-16.5	-55.24	-0.22	22.48	18.06	4.59	1.84	-0.99	-56.23

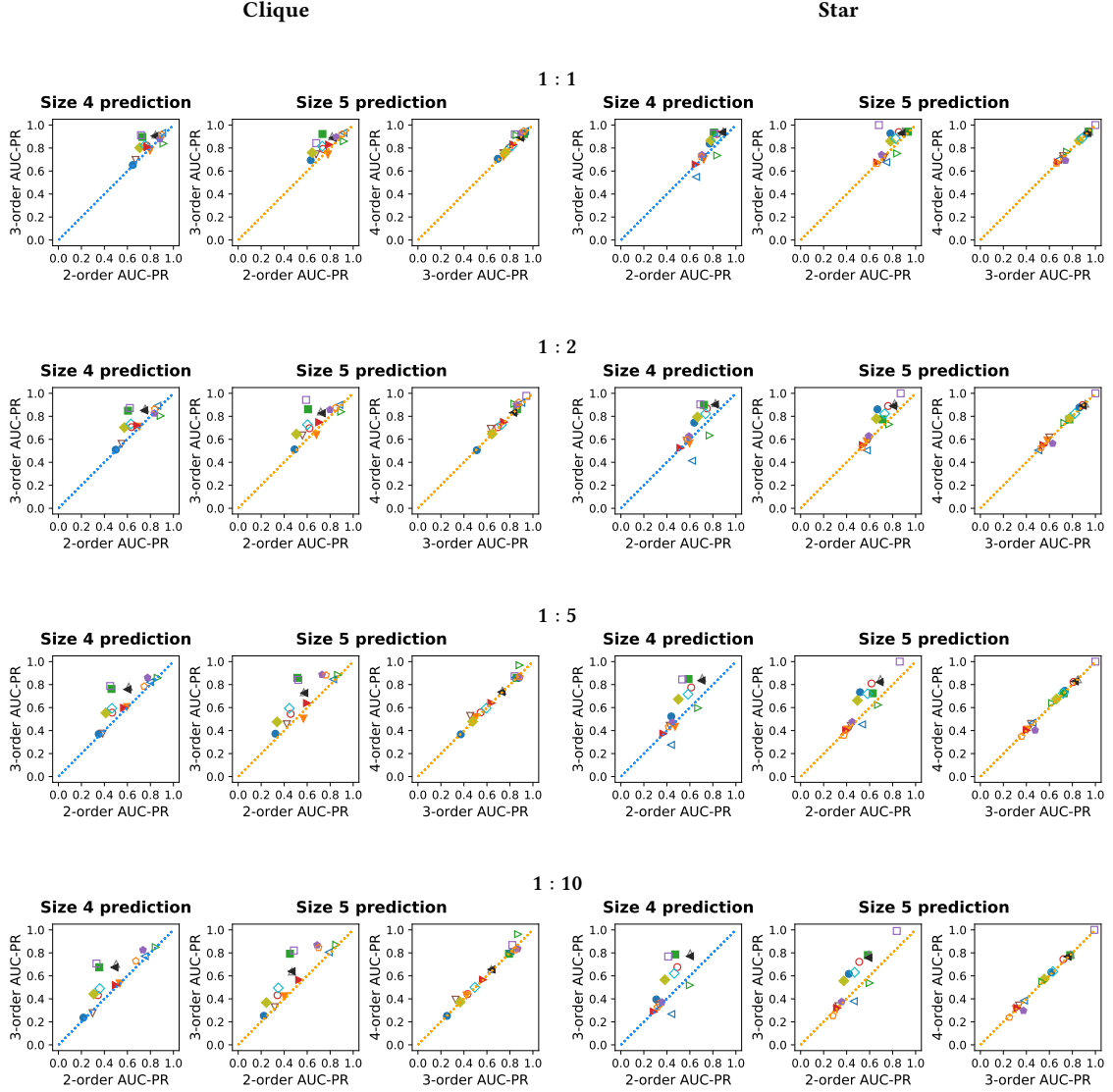


Fig. 1. Performance comparisons between order n and $n+1$. We list the results of size 4 and 5 hyperedge prediction, for every possible combination of negative hyperedges (cliques and stars) and class imbalances (1:1, 1:2, 1:5, 1:10). Performance is higher for order $n+1$, but the gain is smaller for 3 to 4 than 2 to 3.

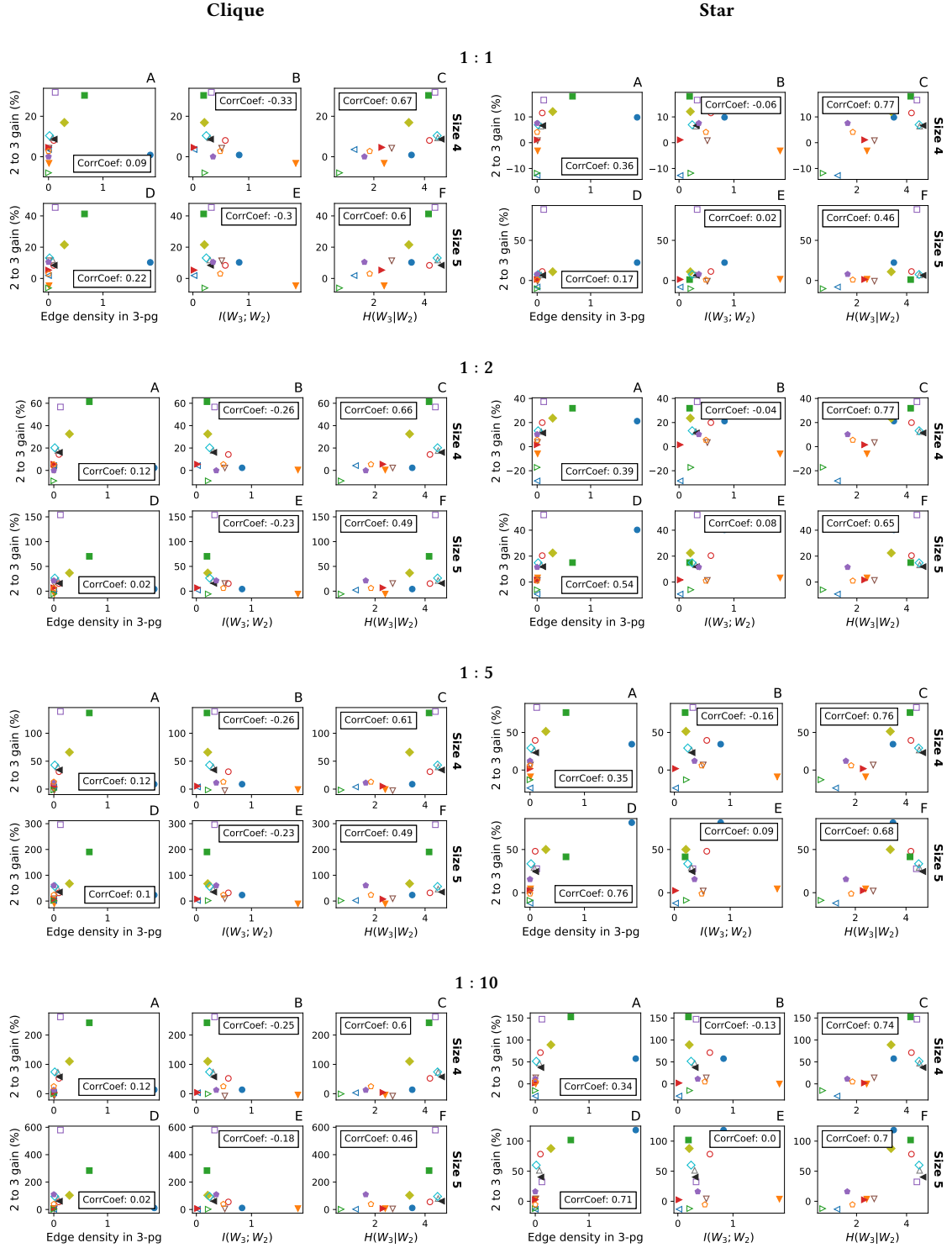


Fig. 2. Correlation between 2 to 3 gain and various measures. The measures are edge density in 3-pg (A, D), mutual information (B, E), and conditional entropy (C, F). Results are shown for every possible combination of negative hyperedges (cliques and stars) and class imbalances (1:1, 1:2, 1:5, 1:10). We observe negative correlation in (B, E) and positive correlations in (A, D) and (C, F).

Table 12. Running time (in seconds) of generating n -projected graphs on Intel Xeon CPU E5-2630 v4

Dataset	2-pg	3-pg	4-pg	5-pg	6-pg	7-pg	8-pg	9-pg
email-Enron	0.01	0.06	0.17	0.46	0.46	0.32	0.1	0.02
contact-primary-school	0.06	0.06	0.01	0.0	0.0	0.0	0.0	0.0
contact-high-school	0.03	0.03	0.01	0.0	0.0	0.0	0.0	0.0
NDC-classes	0.02	0.12	0.5	0.9	1.14	0.75	0.25	0.05
NDC-substances	0.12	1.16	4.96	11.82	15.53	10.47	3.03	0.62
email-Eu	0.22	1.28	3.67	7.82	9.05	5.79	2.16	0.39
threads-ask-ubuntu	0.8	1.11	0.49	0.32	0.19	0.09	0.04	0.02
congress-bills	1.35	12.96	50.08	116.0	152.25	101.61	42.97	7.2
tags-ask-ubuntu	1.47	4.6	5.75	1.25	0.13	0.03	0.01	0.01
tags-math-sx	1.84	5.47	6.06	1.89	0.16	0.03	0.02	0.02
threads-math-sx	3.87	13.74	12.83	10.74	7.33	4.22	1.23	0.31
coauth-MAG-History	3.96	9.69	24.27	43.63	52.62	33.79	12.82	2.42
coauth-MAG-Geology	21.9	104.07	274.7	642.02	715.52	449.94	169.86	29.91
coauth-DBLP	39.94	158.4	384.34	618.01	665.97	387.19	152.9	24.98

Table 13. # n -way relationships

Dataset	2-way	3-way	4-way	5-way	6-way	7-way	8-way	9-way
email-Enron	1442	2972	4323	4396	3176	1614	553	115
contact-primary-school	8317	5139	381	9	0	0	0	0
contact-high-school	5818	2370	238	7	0	0	0	0
NDC-classes	3727	7295	10196	10415	7668	3941	1341	272
NDC-substances	26973	78080	121502	126608	92806	47398	16111	3286
email-Eu	21465	47746	73486	77134	56773	29095	9926	2029
threads-ask-ubuntu	186955	60627	19341	6484	2029	532	105	14
congress-bills	178647	813320	1352919	1378489	975623	481908	158860	31466
DAWN	97046	485561	819666	748456	479228	220850	69370	13246
tags-ask-ubuntu	132703	279369	145676	25475	0	0	0	0
tags-math-sx	91685	249548	156129	29244	0	0	0	0
threads-math-sx	1083531	728189	362499	170761	72644	25437	6564	1078
coauth-MAG-History	723382	700536	704343	592486	383207	180454	58005	11365
coauth-MAG-Geology	4241817	6290188	6677880	5423979	3332687	1493060	458675	86241
coauth-DBLP	7123888	8799400	7678541	5437395	3089911	1327209	397962	73610

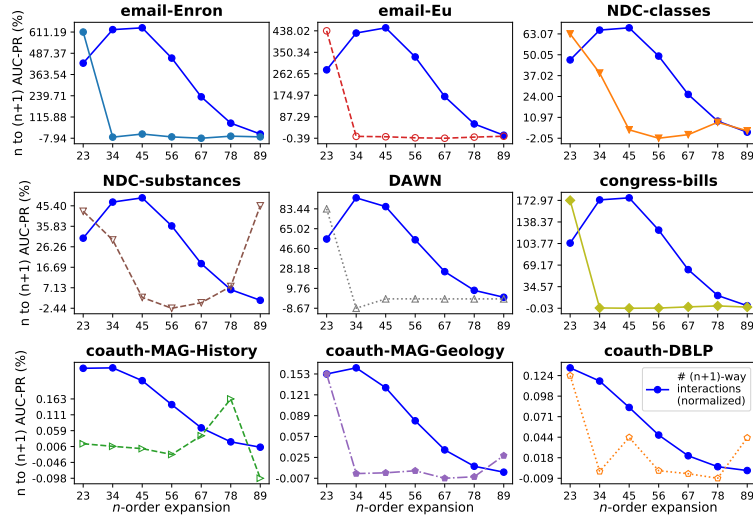


Fig. 3. Can diminishing returns be explained by the number of interactions? To see this, we plot the percentage performance gains of n to $(n+1)$ -way interactions and the number of $(n+1)$ -way interactions in the same axes. For visual convenience, numbers of interactions are normalized (see Table 13 for exact values). We can see that interactions reduce when $n > 4$, which seems to explain the diminishing returns. However, in small n , interactions actually increase while gains significantly decrease. Thus, number of interactions do not strongly explain diminishing returns.