

How Do Hyperedges Overlap in Real-World Hypergraphs? - Patterns, Measures, and Generators (ONLINE APPENDIX)

Anonymous Author(s)

ABSTRACT

This document provides supplementary information to the main paper "How Do Hyperedges Overlap in Real-World Hypergraphs? - Patterns, Measures, and Generators." The full results of the observations and experiments are provided.

ACM Reference Format:

Anonymous Author(s). 2021. How Do Hyperedges Overlap in Real-World Hypergraphs? - Patterns, Measures, and Generators (ONLINE APPENDIX). In . ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/1122445.1122456>

1 SUPPLEMENTARY EXPERIMENTS

We provide full results of the experiments presented in the main paper.

Degree distributions: Figure 1 shows the degree distributions in the real-world hypergraphs and the corresponding synthetic hypergraphs generated by different generative models. HYPERCL, HYPERLAP, and HYPERLAP⁺ accurately preserve the degree distributions of the most of the considered real-world hypergraphs, while HYPERPA and HYPERFF fail in many datasets.

Sampling collisions: Figure 2 shows how the average number of sampling trials for completing a hyperedge changes depending on the sizes of hyperedges in HYPERCL. Note that the number of sampling trials can be larger than the hyperedge size due to collisions. That is, the same node can be sampled multiple times. The difference between the average number of sampling trials and the hyperedge size is small in most cases, while the difference increases as the hyperedge size increases.

Full results of observations: In Table 1, we provide the full results regarding the following patterns:

- **Observation 1:** Egonets in real-world hypergraphs tend to have higher density than those in randomized hypergraphs.
- **Observation 2:** Egonets in real-world hypergraphs tend to have higher overlapness than those in randomized hypergraphs.
- **Observation 3:** The number of hyperedges overlapping at each pair of nodes tends to be more skewed in real-world hypergraphs than in randomized hypergraphs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

- **Observation 4:** The number of hyperedges overlapping at each triple of nodes tends to be more skewed in real-world hypergraphs than in randomized hypergraphs.
- **Observation 5:** Hyperedges in real-world hypergraphs tend to contain structurally more similar nodes, compared to those in randomized hyperedges.

Here, we use hypergraphs generated by HYPERCL as randomized hypergraphs.

Full results of comparisons: We provide full results of the comparison of hypergraph generators with respect to the reproduction of the five empirical patterns above.

- **Comparison 1:** The results regarding Observation 1 (i.e., the density of egonet) are provided in Table 3.
- **Comparison 2:** The results regarding Observation 2 (i.e., the overlapness of egonet) are provided in Table 4.
- **Comparison 3:** The results regarding Observation 3 (i.e., the number of hyperedges overlapping at each pair of nodes) are provided in Table 5. We show that their distributions are heavy-tailed in Table 8.
- **Comparison 4:** The results regarding Observation 4 (i.e., the number of hyperedges overlapping at each triple of nodes) are provided in Table 6. We show that their distributions are heavy-tailed in Table 8.
- **Comparison 5:** The results regarding Observation 5 (i.e., the homogeneity of hyperedges) are provided in Table 7. We show that their distributions are heavy-tailed in Table 8.

We also compare HYPERPA and HYPERCL with respect to all these patterns visually in Table 2.

Counterexamples regarding the Axioms 1, 2, and 3: In the main paper, we suggest the three axioms below, which any reasonable metrics of the degree of hyperedge overlaps should satisfy. We provide counter examples showing that, except for our proposed Overlapness, **none** of the baseline metrics (i.e., Intersection, Union Inverse, Jaccard Index, Overlap Coefficient, and Density) satisfy **all** the three axioms.

- **Axiom 1: number of hyperedges** The counter examples showing that some baseline metrics do not satisfy Axiom 1 are given in Table 9.
- **Axiom 2: number of distinct nodes** The counter examples showing that some baseline metrics do not satisfy Axiom 2 in are given in Table 10.
- **Axiom 3: sizes of hyperedges** showing that some baseline metrics do not satisfy Axiom 3 in are given in Table 11.

Note that, except for our proposed Overlapness, all baseline metrics do not satisfy at least one axiom.

117 **Results of examining macroscopic structural properties:** We
 118 demonstrate that HYPERLAP⁺ reproduces the following four macro-
 119 scopic structural patterns of real-world hypergraphs, which are
 120 suggested in [3].

- 121 • **Heavy-tailed degree distribution:** The degree distributions
 122 of real hypergraphs generally follow heavy-tailed distributions.
 123 The degree of a node v is defined as the number of hyperedges
 124 containing v .
- 125 • **Heavy-tailed hyperedge size distribution:** The distributions
 126 of hyperedge sizes are also heavy-tailed.
- 127 • **Heavy-tailed intersection size distribution:** The number of
 128 nodes that two hyperedges commonly have also fall under the
 129 class of heavy-tailed distributions.
- 130 • **Skewed singular values:** The singular values of the incidence
 131 matrices generally follow heavy-tailed distributions.

132 In Table 12, we provide the distributions regarding these four struc-
 133 tural properties in hypergraphs generated by different models and
 134 show that all the distributions from HYPERLAP⁺ are very similar to
 135 that of real-world hypergraphs. We also check in Table 13 that all
 136

137 the distributions from HYPERLAP⁺ are best fitted to heavy-tailed
 138 distributions rather than exponential distributions.

139 **Comparison of potential null models:** We compare HYPERCL,
 140 which we use as a null model, with two potential alternatives: the
 141 random bipartite graph generative model [1] and the hypergraph
 142 configuration model [2]. Note that HYPERCL preserves hyperedge
 143 sizes exactly and node degrees in expectation, while the configura-
 144 tion model preserves both exactly, and the bipartite graph gener-
 145 ative model preserves both in expectation. As shown in Table 14,
 146 empirically, these three models generate similarly realistic hyper-
 147 graphs.

REFERENCES

- [1] Sinan G Aksoy, Tamara G Kolda, and Ali Pinar. 2017. Measuring and modeling bipartite graphs with community structure. *Journal of Complex Networks* 5, 4 (2017), 581–603.
- [2] Philip S Chodrow. 2020. Configuration models of random hypergraphs. *Journal of Complex Networks* 8, 3 (2020), cnaa018.
- [3] Yunbum Kook, Jihoon Ko, and Kijung Shin. 2020. Evolution of Real-world Hypergraphs: Patterns and Models without Oracles. *arXiv preprint arXiv:2008.12729* (2020).

175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232

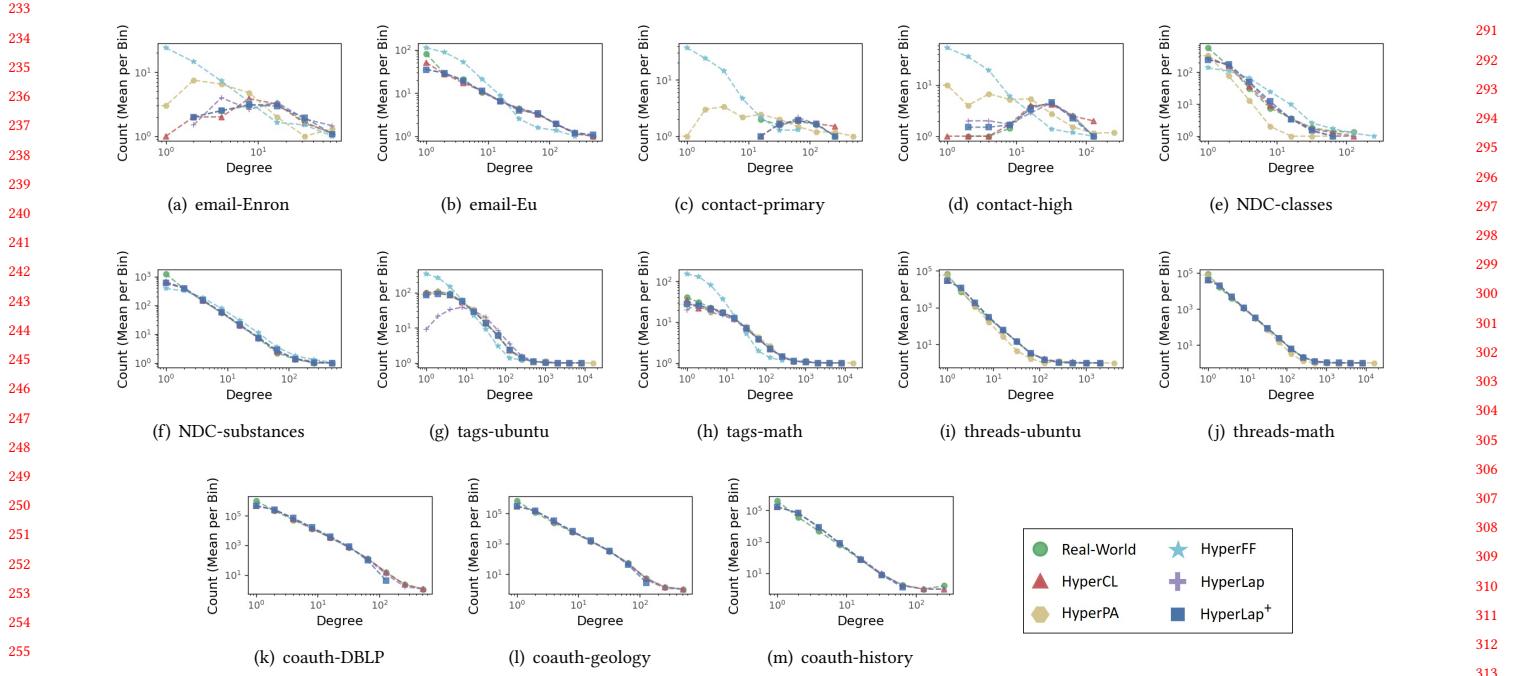


Figure 1: Degree distributions. The degree distributions in the real-world hypergraphs and the corresponding synthetic hypergraphs generated by different generative models. HYPERCL, HYPERLAP, and HYPERLAP⁺ accurately preserve the degree distributions of the most of the considered real-world hypergraphs, while HYPERPA and HYPERFF fail in many datasets.

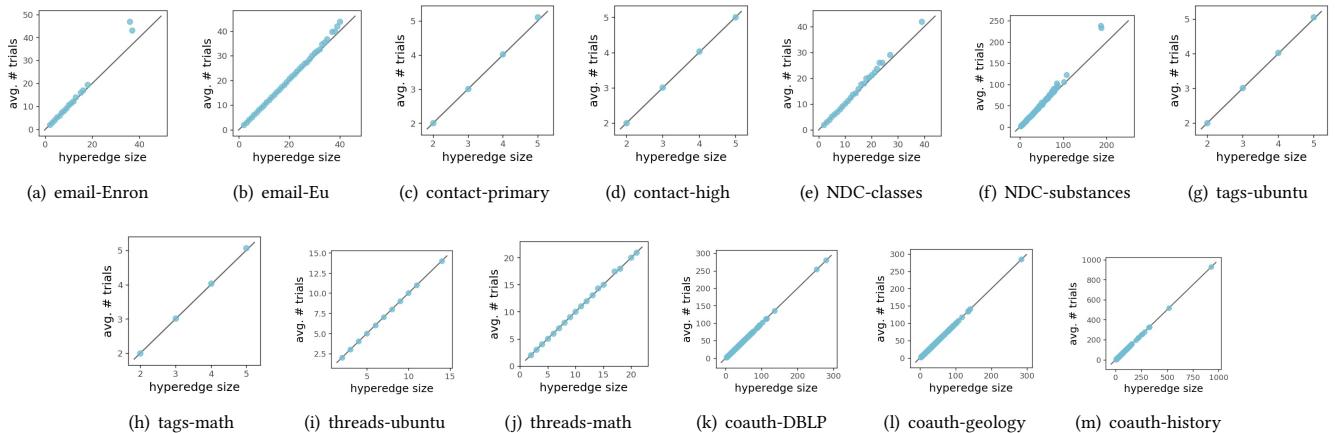
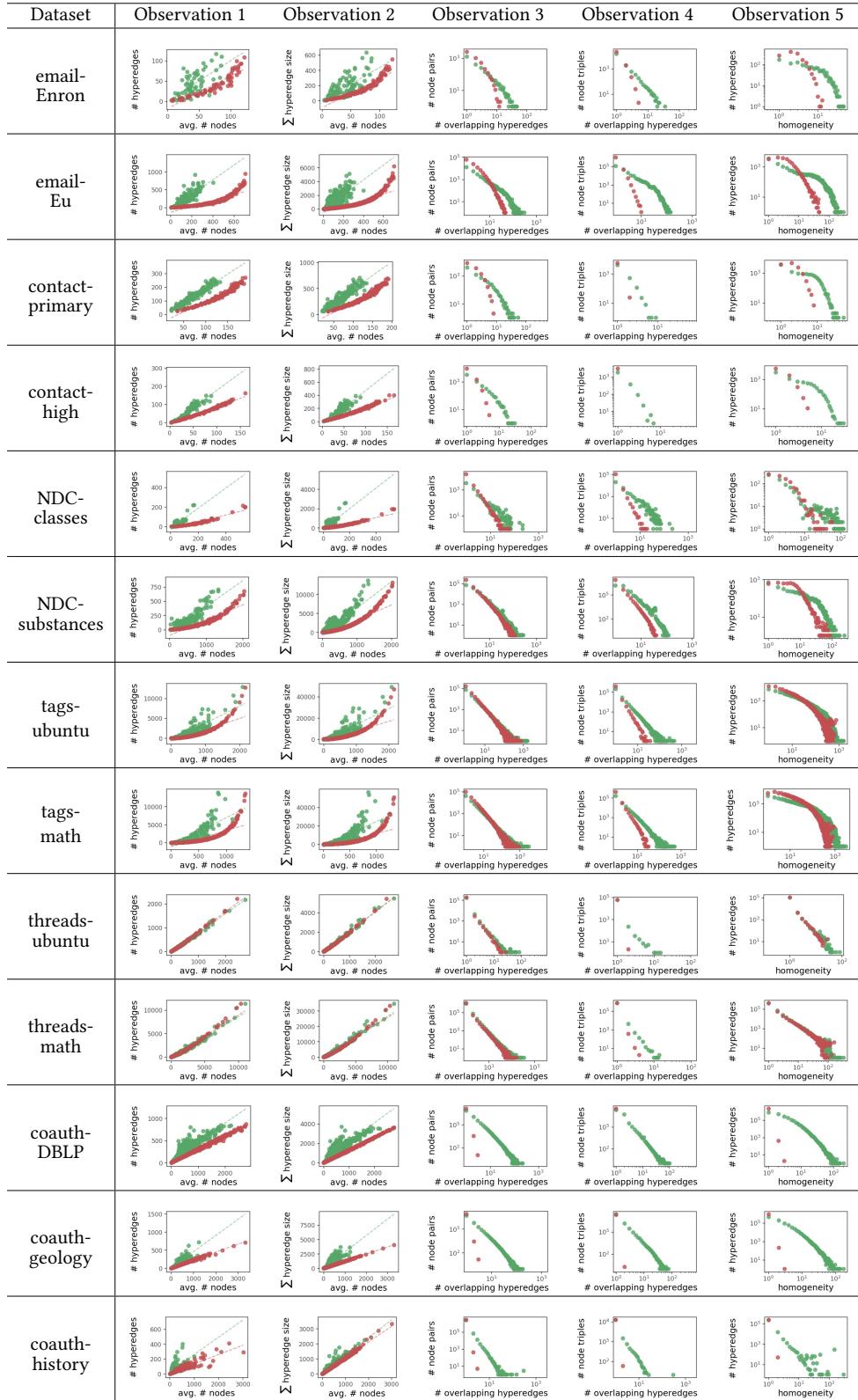


Figure 2: The average number of sampling trials for completing hyperedges with different sizes. The difference between the average number of sampling trials and the hyperedge size is small in most cases, while the difference increases as the hyperedge size increases.

Table 1: Full results of observations. The overlapping patterns of hyperedges in real-world hypergraphs (green) are distinguished from those in randomized hypergraphs (red) in thirteen datasets.



465
 466
 467

Table 2: The five empirical patterns in real-world hypergraphs (green) are reproduced accurately by HYPERLAP⁺ (blue), while HYPERCL (red) fails in many cases.

523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580

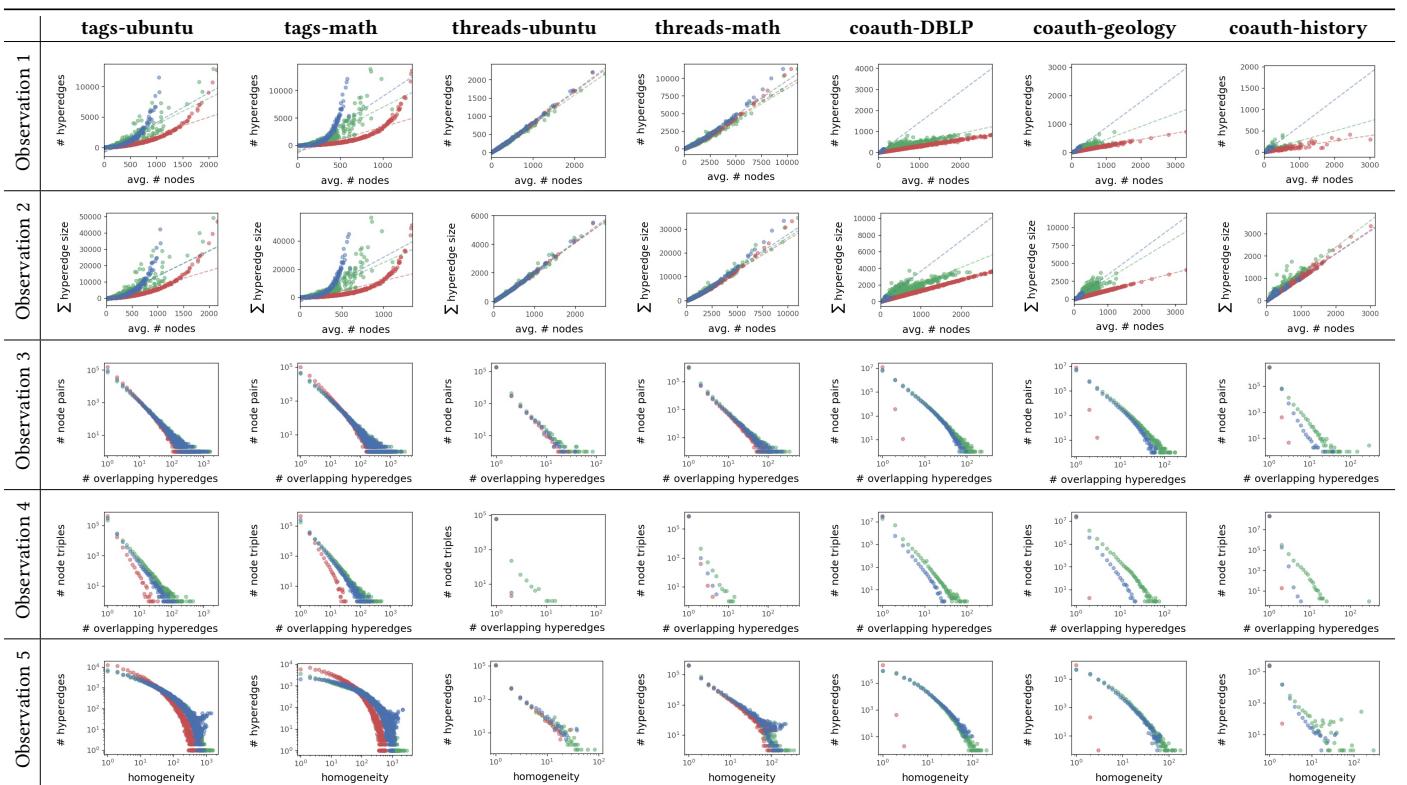
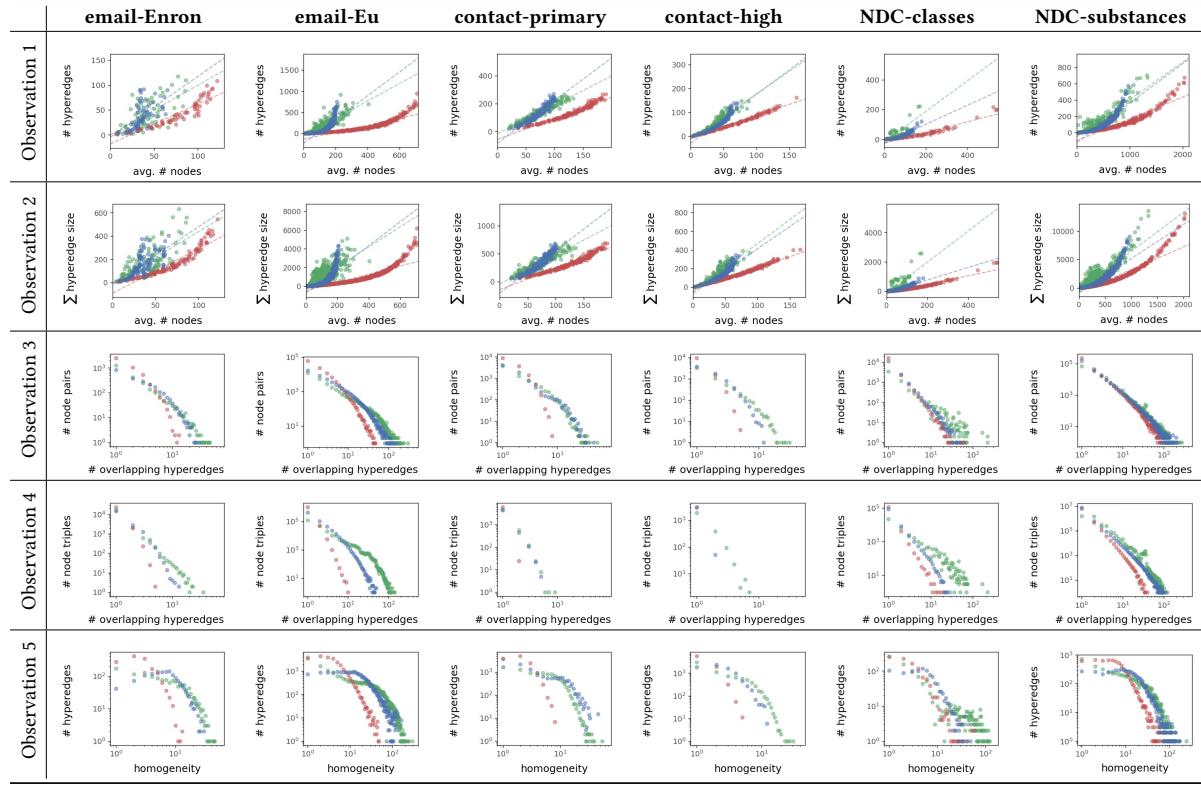
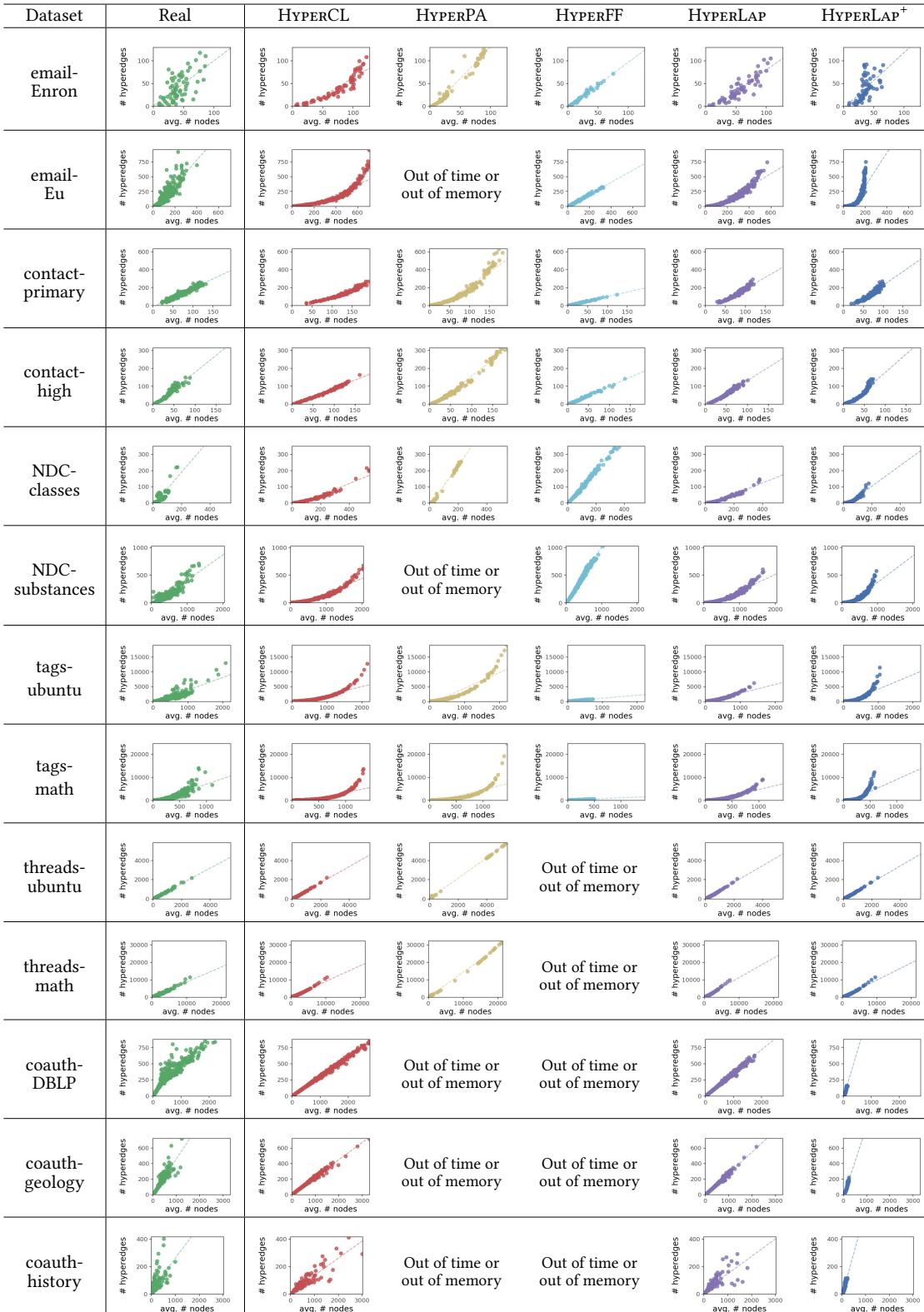


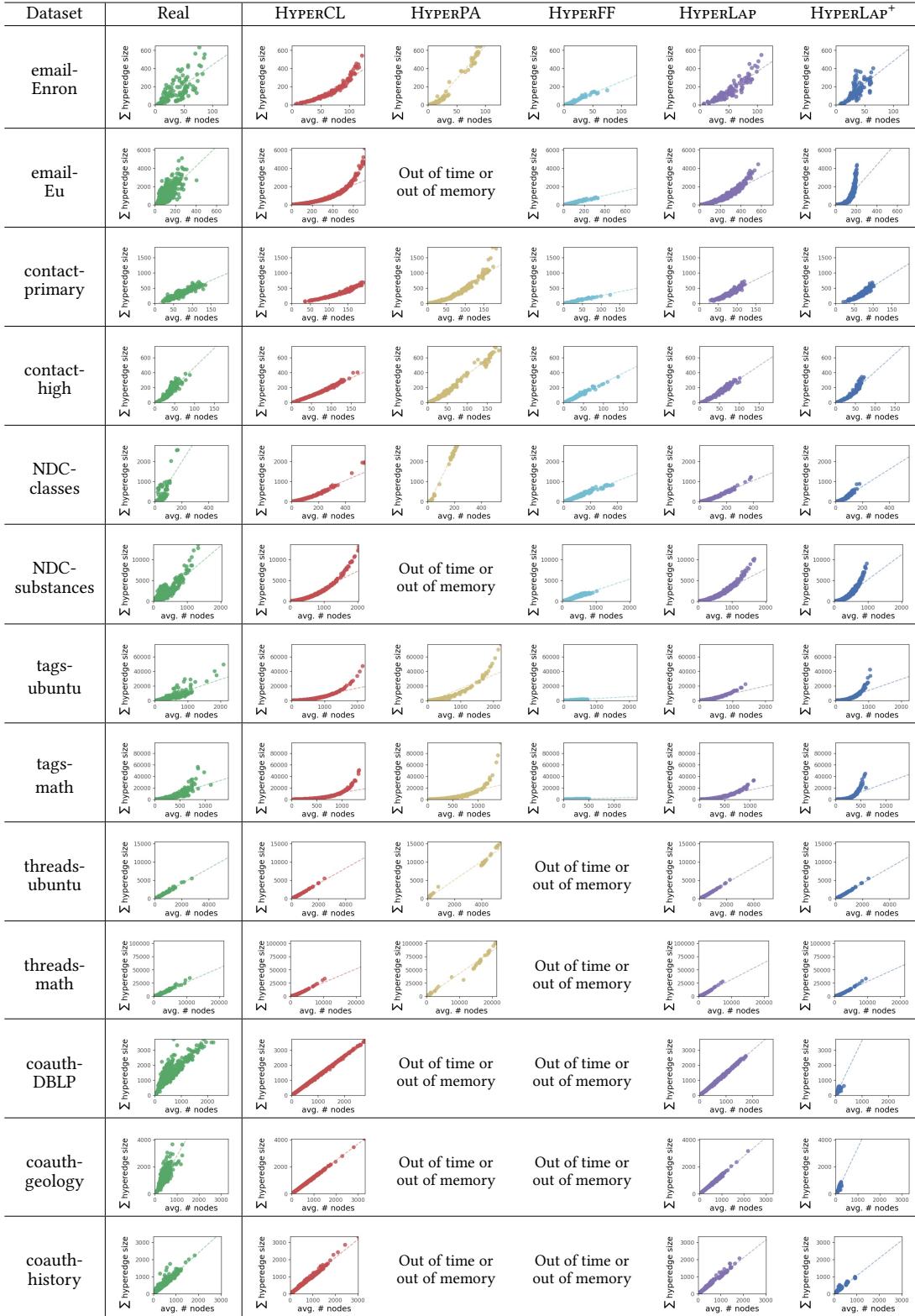
Table 3: Comparison 1. Comparison of hyperedge generators with respect to egonet density distributions. The slopes of the regression lines imply the average egonet density. The distribution generated by HYPERLAP and HYPERLAP⁺ look similar to the distributions generated by real-world hypergraphs in thirteen datasets.



697

Table 4: Observation 2. Comparison of hyperedge generators with respect to egonet overlapness distributions. The slopes of the regression lines imply the average egonet overlapness. The distributions generated by HYPERLAP and HYPERLAP⁺ look similar to the distribution generated by real-world hypergraphs in thirteen datasets.

701



755
 756
 757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

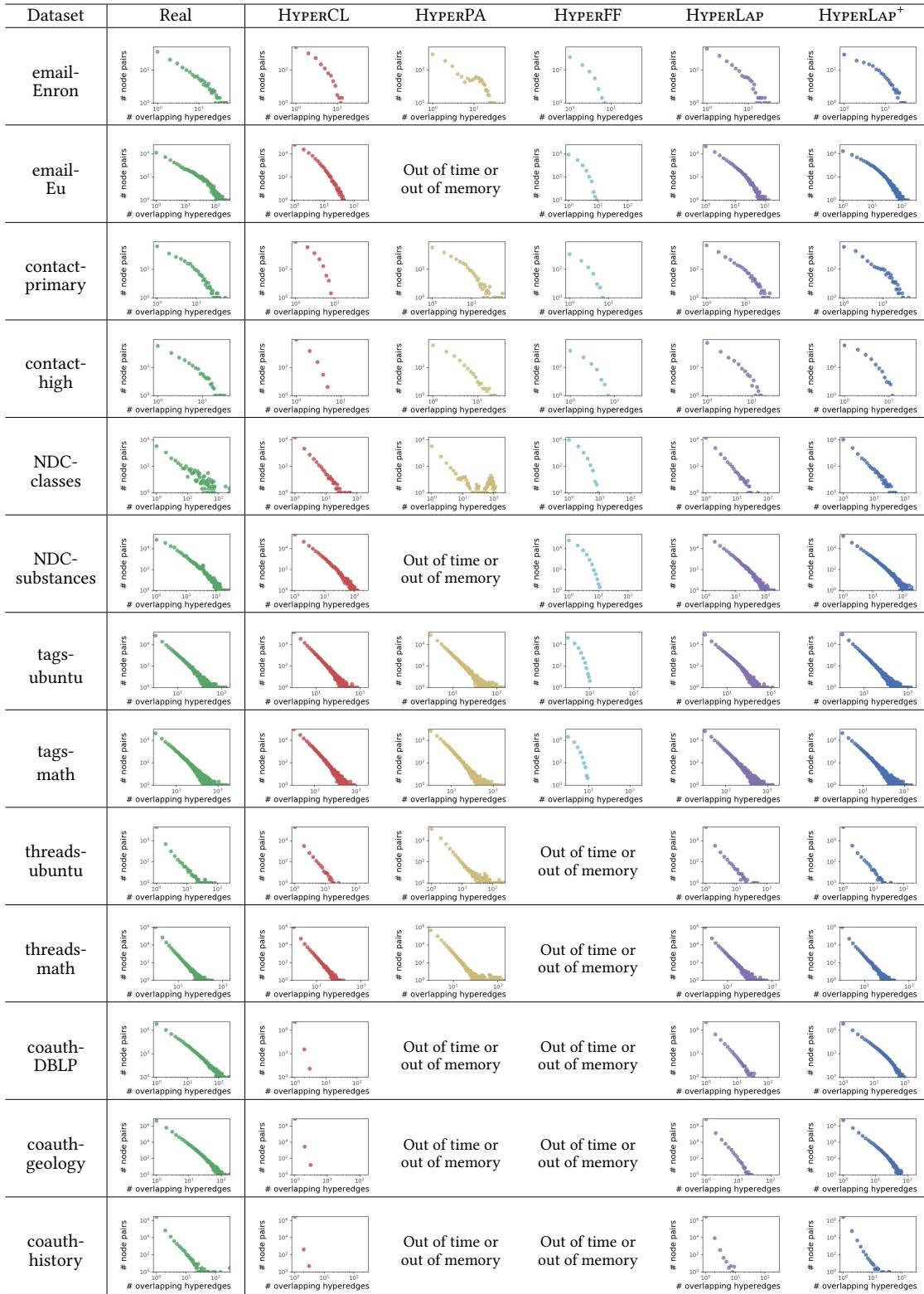
808

809

810

811

Table 5: Comparison 3. Comparison of hypergraph generators with respect to node-pair degree distributions, i.e., the number of hyperedges overlapping at each pair of nodes. The distribution is skewed with a heavy tail in real-world hypergraphs. This pattern is preserved well by HYPERLAP and HYPERLAP⁺.



929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000
 1001
 1002
 1003
 1004
 1005
 1006
 1007
 1008
 1009
 1010
 1011
 1012
 1013
 1014
 1015
 1016
 1017
 1018
 1019
 1020
 1021
 1022
 1023
 1024
 1025
 1026
 1027
 1028
 1029
 1030
 1031
 1032
 1033
 1034
 1035
 1036
 1037
 1038
 1039
 1040
 1041
 1042
 1043

Table 6: Comparison 4. Comparison of hypergraph generators with respect to node-triple degree distributions, i.e., the number of hyperedges overlapping at each triple of nodes. The distribution is skewed with a heavy tail in real-world hypergraphs. This pattern is preserved well by HYPERLAP and HYPERLAP⁺.

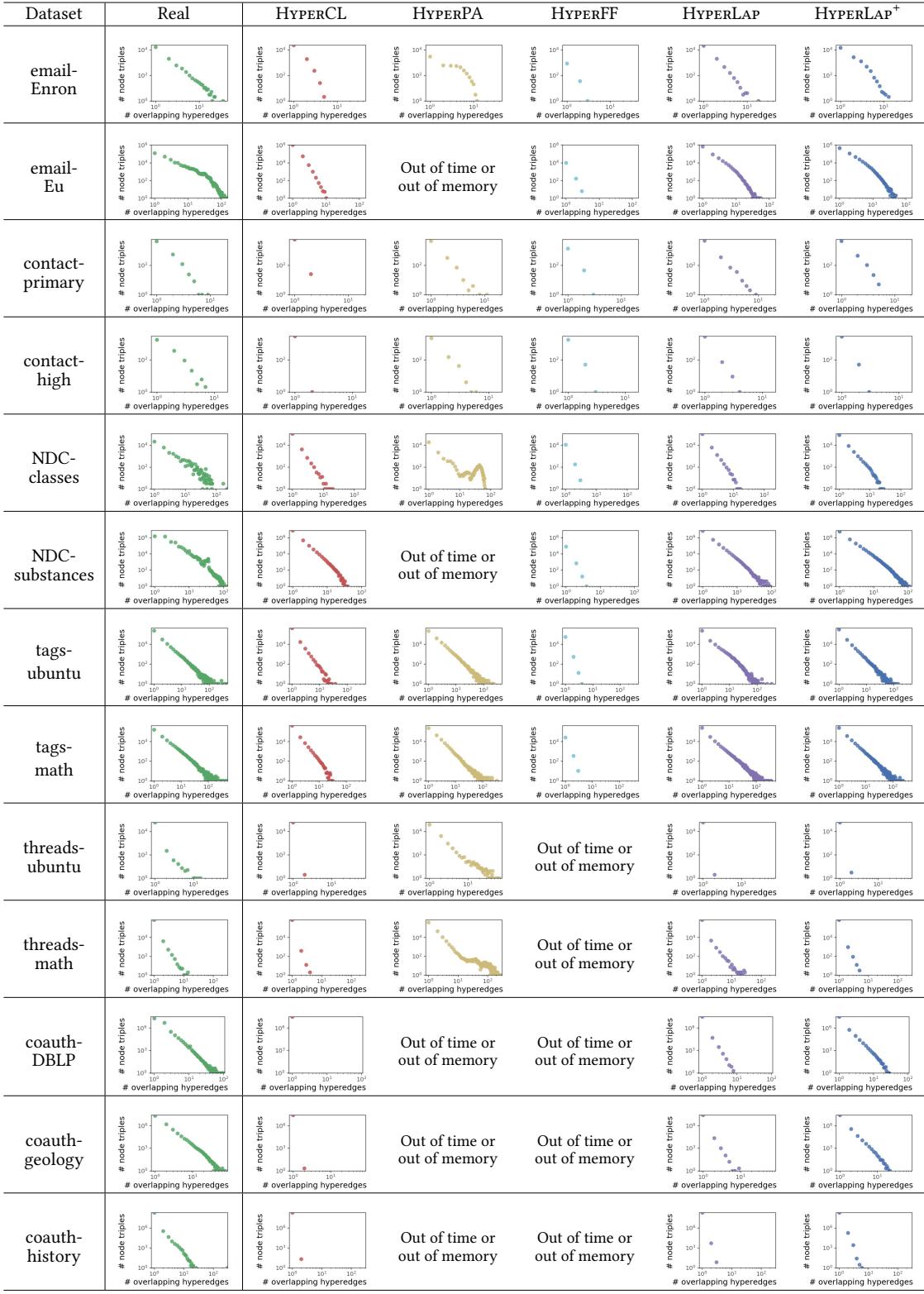


Table 7: Comparison 5. Comparison of hyperedge generators with respect to hyperedge homogeneity distributions. The distribution of real-world hypergraphs shows that hyperedges in them tend to contain structurally more similar nodes, which leads to be skewed with a heavy tail. This pattern also can be shown in the distributions generated by HYPERLP and HYPERLP⁺.



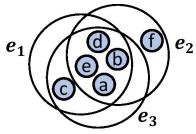
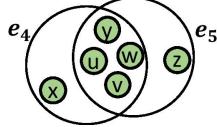
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159

1161
 1162 **Table 8: The heavy-tailed distributions.** The distributions of the number of hyperedges overlapping at each pair or triple of
 1163 nodes and homogeneity of hyperedges are all heavy-tailed. The loglikelihood ratios when fitting the distributions to each of
 1164 three heavy-tailed distributions (i.e., the power-law distribution, the truncated power-law distribution, and the log normal
 1165 distribution) against the exponential distribution support this claim. We report the maximum log-likelihood ratio, and a
 1166 positive value means that the distribution is closer to a heavy-tailed distribution than to the exponential distribution.
 1167

Dataset	Pair of Nodes					Triple of Nodes					Homogeneity							
	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺
email-Enron	4.22	0.11	0.13	0.23	1.03	-0.29	3.88	0.37	1.69	0.50	0.74	0.38	-0.26	0.53	1.14	2.18	1.34	-0.43
email-Eu	1.48	1.95	-	0.41	1.71	2.75	0.77	-0.18	-	1.24	1.25	2.11	0.91	2.63	-	4.21	3.93	3.13
contact-primary	1.40	-0.49	2.54	0.54	0.40	15.74	0.48	2.60	0.44	0.50	0.00	1.23	2.30	17.67	1.16	4.25	1.63	2.26
contact-high	0.81	1.06	1.34	0.15	0.38	-0.06	0.80	0.10	0.10	0.50	0.52	0.50	1.95	11.08	2.09	3.23	2.95	4.72
NDC-classes	15.74	4.09	3.42	0.55	6.47	14.42	31.53	1.69	5.03	1.24	1.76	7.92	0.39	3.10	1.26	4.94	2.78	0.87
NDC-substances	43.87	34.99	-	0.54	34.78	40.13	116.45	12.55	-	0.51	19.29	55.38	1.22	3.46	-	9.22	2.78	2.90
tags-ubuntu	41.55	38.19	33.01	1.66	49.54	43.70	17.84	4.85	22.82	0.51	3.12	15.57	2.26	5.77	-1.04	8.37	11.53	7.00
tags-math	4.49	54.94	12.59	0.27	49.14	45.60	29.26	8.59	13.98	1.60	29.34	23.12	2.62	1.92	3.70	5.40	13.06	6.56
threads-ubuntu	3.97	2.33	7.08	-	2.68	1.75	0.80	-1.41	9.10	-	-1.41	-1.72	7.70	2.14	46.51	-	6.56	4.25
threads-math	14.78	18.44	27.31	-	22.46	16.66	-0.09	0.76	0.18	-	3.96	0.96	9.00	9.66	-0.50	-	27.35	12.10
coauth-DBLP	22.47	1.57	-	-	3.32	74.95	5.84	0.00	-	-	0.79	6.73	4.31	156.83	-	-	4.53	25.23
coauth-geology	53.39	1.93	-	-	1.52	45.08	13.73	-1.41	-	-	1.53	1.10	5.52	101.59	-	-	41.95	8.06
coauth-history	3.91	1.12	-	-	0.89	1.77	1.42	-4.46	-	-	1.32	0.57	1.73	4.82	-	-	9.91	4.31

-: out of time (taking more than 10 hours) or out of memory

1193 **Table 9: An example showing that Intersection, Union Inverse, Jaccard Index, and Overlap Coefficient do not satisfy Axiom 1.**

Metric	\mathcal{E}'	\mathcal{E}
		
Intersection	$ e_1 \cap e_2 \cap e_3 = \{a, b, d, e\} = 4$	= $ e_4 \cap e_5 = \{u, v, w, y\} = 4$
Union Inverse	$\frac{1}{ e_1 \cup e_2 \cup e_3 } = \frac{1}{ \{a,b,c,d,e,f\} } = \frac{1}{6}$	= $\frac{1}{ e_4 \cup e_5 } = \frac{1}{ \{u,v,w,x,y,z\} } = \frac{1}{6}$
Jaccard Index	$\frac{ e_1 \cap e_2 \cap e_3 }{ e_1 \cup e_2 \cup e_3 } = \frac{ \{a,b,d,e\} }{ \{a,b,c,d,e,f\} } = \frac{4}{6}$	= $\frac{ e_4 \cap e_5 }{ e_4 \cup e_5 } = \frac{ \{u,v,w,y\} }{ \{u,v,w,x,y,z\} } = \frac{4}{6}$
Overlap Coefficient	$\frac{ e_1 \cap e_2 \cap e_3 }{\min(e_1 , e_2 , e_3)} = \frac{ \{a,b,d,e\} }{\min(5,5,5)} = \frac{4}{5}$	= $\frac{ e_4 \cap e_5 }{\min(e_4 , e_5)} = \frac{ \{u,v,w,y\} }{\min(5,5)} = \frac{4}{5}$
Density	$\frac{ \mathcal{E}' }{ \bigcup_{e \in \mathcal{E}'} e } = \frac{ \{e_1, e_2, e_3\} }{ \{a,b,c,d,e,f\} } = \frac{3}{6}$	> $\frac{ \mathcal{E} }{ \bigcup_{e \in \mathcal{E}} e } = \frac{ \{e_4, e_5\} }{ \{u,v,w,x,y,z\} } = \frac{2}{6}$
Overlapness	$\frac{\sum_{e \in \mathcal{E}'} e }{ \bigcup_{e \in \mathcal{E}'} e } = \frac{ e_1 + e_2 + e_3 }{ \{a,b,c,d,e,f\} } = \frac{15}{6}$	> $\frac{\sum_{e \in \mathcal{E}} e }{ \bigcup_{e \in \mathcal{E}} e } = \frac{ e_4 + e_5 }{ \{u,v,w,x,y,z\} } = \frac{10}{6}$

1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244
 1245
 1246
 1247
 1248
 1249
 1250
 1251
 1252
 1253
 1254
 1255
 1256
 1257
 1258
 1259
 1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276

Table 10: An example showing that Intersection, Jaccard Index, and Overlap Coefficient do not satisfy Axiom 2.

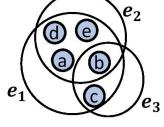
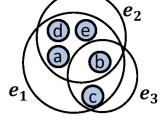
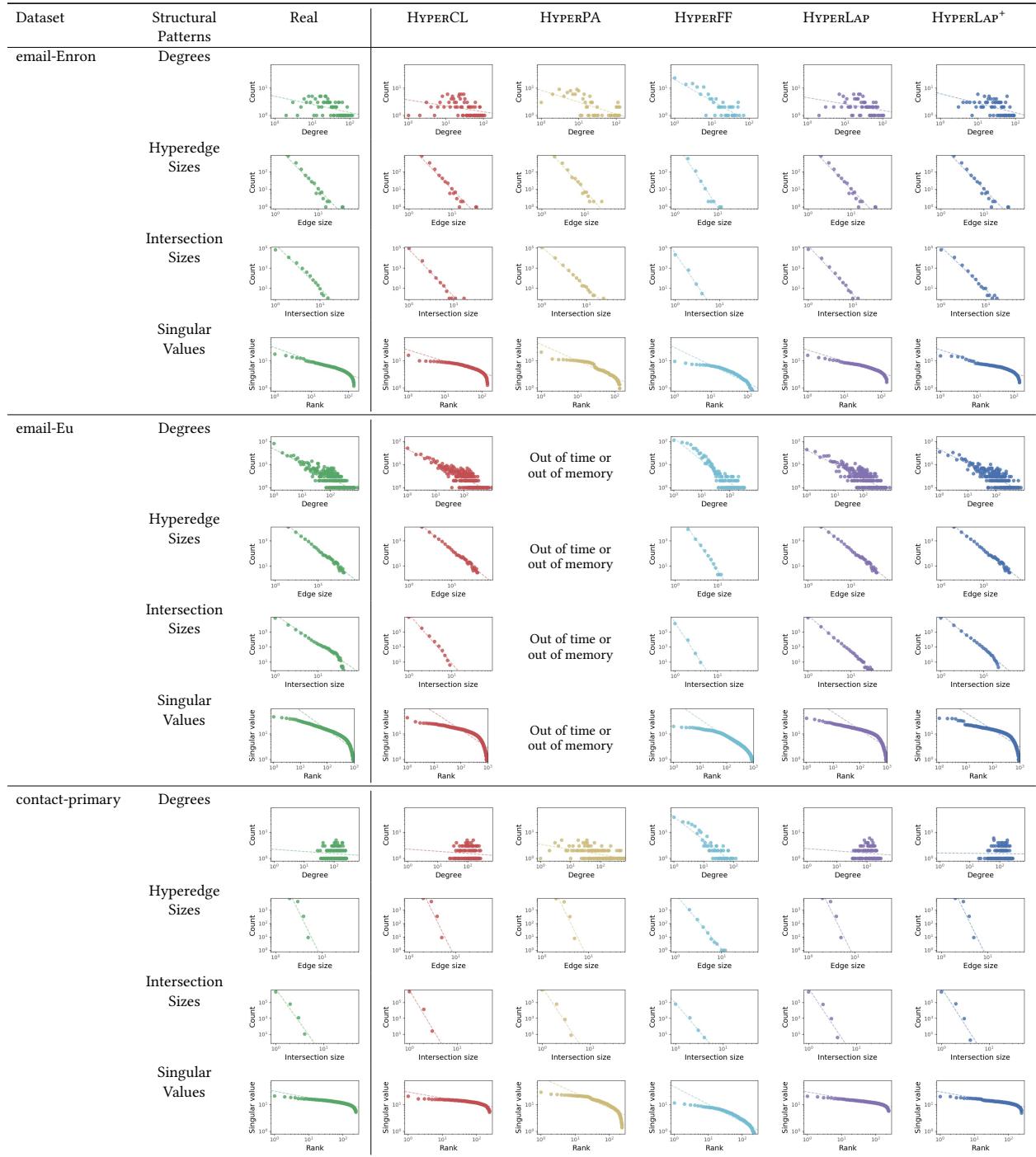
Metric	\mathcal{E}'	\mathcal{E}
	$ e_1 \cap e_2 \cap e_3 = \{b\} = 1$	$ e_4 \cap e_5 \cap e_6 = \{t, u\} = 2$
Intersection	$<$	
Union Inverse	$\frac{1}{ e_1 \cup e_2 \cup e_3 } = \frac{1}{ \{a,b,c,d,e\} } = \frac{1}{5}$	$>$
Jaccard Index	$\frac{ e_1 \cap e_2 \cap e_3 }{ e_1 \cup e_2 \cup e_3 } = \frac{ \{b\} }{ \{a,b,c,d,e\} } = \frac{1}{5}$	$<$
Overlap Coefficient	$\frac{ e_1 \cap e_2 \cap e_3 }{\min(e_1 , e_2 , e_3)} = \frac{ \{b\} }{\min(5,4,2)} = \frac{1}{2}$	$<$
Density	$\frac{ \mathcal{E}' }{ \bigcup_{e \in \mathcal{E}'} e } = \frac{ \{e_1, e_2, e_3\} }{ \{a, b, c, d, e\} } = \frac{3}{5}$	$>$
Overlapness	$\frac{\sum_{e \in \mathcal{E}'} e }{ \bigcup_{e \in \mathcal{E}'} e } = \frac{ e_1 + e_2 + e_3 }{ \{a, b, c, d, e\} } = \frac{11}{5}$	$>$

Table 11: An example showing that all the considered measures except for Overlapness do not satisfy Axiom 3.

Metric	\mathcal{E}'	\mathcal{E}
	$ e_1 \cap e_2 \cap e_3 = \{b\} = 1$	$ e_4 \cap e_5 \cap e_6 = \{w\} = 1$
Intersection	$=$	
Union Inverse	$\frac{1}{ e_1 \cup e_2 \cup e_3 } = \frac{1}{ \{a,b,c,d,e\} } = \frac{1}{5}$	$=$
Jaccard Index	$\frac{ e_1 \cap e_2 \cap e_3 }{ e_1 \cup e_2 \cup e_3 } = \frac{ \{b\} }{ \{a,b,c,d,e\} } = \frac{1}{5}$	$=$
Overlap Coefficient	$\frac{ e_1 \cap e_2 \cap e_3 }{\min(e_1 , e_2 , e_3)} = \frac{ \{b\} }{\min(5,4,2)} = \frac{1}{2}$	$=$
Density	$\frac{ \mathcal{E}' }{ \bigcup_{e \in \mathcal{E}'} e } = \frac{ \{e_1, e_2, e_3\} }{ \{a, b, c, d, e\} } = \frac{3}{5}$	$=$
Overlapness	$\frac{\sum_{e \in \mathcal{E}'} e }{ \bigcup_{e \in \mathcal{E}'} e } = \frac{ e_1 + e_2 + e_3 }{ \{a, b, c, d, e\} } = \frac{11}{5}$	$>$

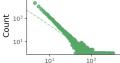
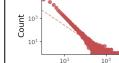
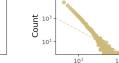
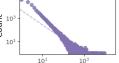
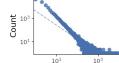
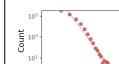
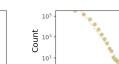
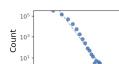
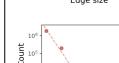
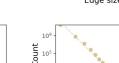
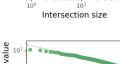
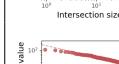
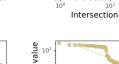
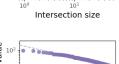
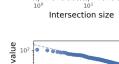
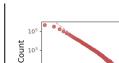
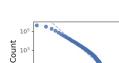
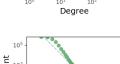
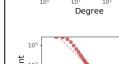
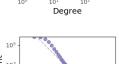
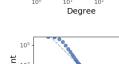
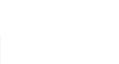
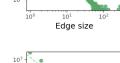
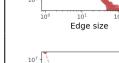
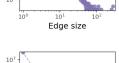
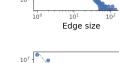
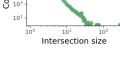
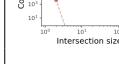
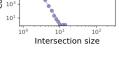
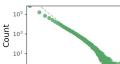
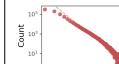
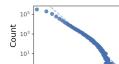
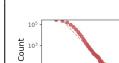
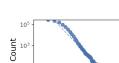
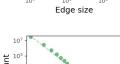
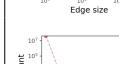
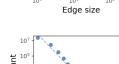
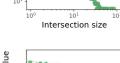
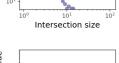
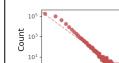
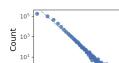
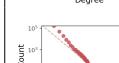
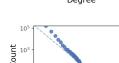
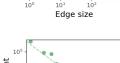
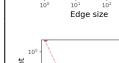
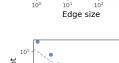
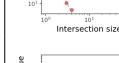
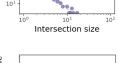
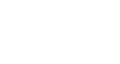
1393
 1394
 1395

Table 12: Macroscopic Structural Patterns. All four macroscopic structural properties of real-world hypergraphs [3] are reproduced accurately by HYPERLAP⁺ in all the datasets.





1625	Dataset	Structural Patterns	Real	HYPERCCL	HYPERPAA	HYPERRFF	HYPERLAP	HYPERLAP ⁺	1683
1626	tags-ubuntu	Degrees							1684
1627		Hyperedge Sizes							1685
1628		Intersection Sizes							1686
1629		Singular Values							1687
1630	tags-math	Degrees							1688
1631		Hyperedge Sizes							1689
1632		Intersection Sizes							1690
1633		Singular Values							1691
1634	threads-ubuntu	Degrees							1692
1635		Hyperedge Sizes							1693
1636		Intersection Sizes							1694
1637		Singular Values							1695
1638		Degrees							1696
1639		Hyperedge Sizes							1697
1640		Intersection Sizes							1698
1641		Singular Values							1699
1642		Degrees							1700
1643		Hyperedge Sizes							1701
1644		Intersection Sizes							1702
1645		Singular Values							1703
1646		Degrees							1704
1647		Hyperedge Sizes							1705
1648		Intersection Sizes							1706
1649		Singular Values							1707
1650		Degrees							1708
1651		Hyperedge Sizes							1709
1652		Intersection Sizes							1710
1653		Singular Values							1711
1654		Degrees							1712
1655		Hyperedge Sizes							1713
1656		Intersection Sizes							1714
1657		Singular Values							1715
1658		Degrees							1716
1659		Hyperedge Sizes							1717
1660		Intersection Sizes							1718
1661		Singular Values							1719
1662		Degrees							1720
1663		Hyperedge Sizes							1721
1664		Intersection Sizes							1722
1665		Singular Values							1723
1666		Degrees		<img					

1741	Dataset	Structural Patterns	Real	HYPERCL	HYPERPA	HYPERRFF	HYPERLAP	HYPERLAP ⁺	1799
1742	threads-math	Degrees				Out of time or out of memory			
1743		Hyperedge Sizes				Out of time or out of memory			
1744		Intersection Sizes				Out of time or out of memory			
1745		Singular Values				Out of time or out of memory			
1746	coauth-DBLP	Degrees			Out of time or out of memory	Out of time or out of memory			
1747		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1748		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1749		Singular Values			Out of time or out of memory	Out of time or out of memory			
1750	coauth-geology	Degrees			Out of time or out of memory	Out of time or out of memory			
1751		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1752		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1753		Singular Values			Out of time or out of memory	Out of time or out of memory			
1754	coauth-history	Degrees			Out of time or out of memory	Out of time or out of memory			
1755		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1756		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1757		Singular Values			Out of time or out of memory	Out of time or out of memory			
1758		Degrees			Out of time or out of memory	Out of time or out of memory			
1759		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1760		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1761		Singular Values			Out of time or out of memory	Out of time or out of memory			
1762		Degrees			Out of time or out of memory	Out of time or out of memory			
1763		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1764		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1765		Singular Values			Out of time or out of memory	Out of time or out of memory			
1766		Degrees			Out of time or out of memory	Out of time or out of memory			
1767		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1768		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1769		Singular Values			Out of time or out of memory	Out of time or out of memory			
1770	coauth-geology	Degrees			Out of time or out of memory	Out of time or out of memory			
1771		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1772		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1773		Singular Values			Out of time or out of memory	Out of time or out of memory			
1774	coauth-history	Degrees			Out of time or out of memory	Out of time or out of memory			
1775		Hyperedge Sizes			Out of time or out of memory	Out of time or out of memory			
1776		Intersection Sizes			Out of time or out of memory	Out of time or out of memory			
1777		Singular Values			Out of time or out of memory	Out of time or out of memory			
1778		Degrees			Out of time or out of memory	Out of time or out of memory	<		

1857
1858
1859
1860
1861
1862**Table 13: The heavy-tailed distributions regarding the four macroscopic structural properties [3]. The distributions of degrees, hyperedge sizes, intersection sizes, and singular values are all heavy-tailed in hypergraphs generated by HYPERLAP⁺. As in Table 8, we report the maximum log-likelihood ratios of heavy-tailed distributions against exponential distributions. We boldface positive values, which indicate that the corresponding distribution is closer to a heavy-tailed distribution than to the exponential distribution.**1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972

Dataset	Degrees					Hyperedge Sizes						
	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺
email-Enron	-2.97	-2.21	3.77	1.28	-3.19	0.51	2.46	2.46	2.55	1.45	2.46	2.46
email-Eu	0.98	1.21	-	5.47	0.40	-1.86	19.54	19.54	-	-0.19	19.54	19.54
contact-primary	1.89	1.23	-2.22	2.79	1.32	0.84	2.35	2.35	2.49	0.62	2.35	2.35
contact-high	1.14	0.13	4.31	3.16	1.18	0.61	12.77	12.77	12.60	0.31	12.77	12.77
NDC-classes	7.48	7.22	16.17	5.17	7.04	7.14	0.85	0.85	0.59	0.064	0.85	0.85
NDC-substances	1.95	1.87	-	8.88	2.85	1.26	4.80	4.80	-	0.24	4.80	4.80
tags-ubuntu	8.83	8.86	9.34	7.97	6.63	9.15	99.66	99.66	103.5	0.198	99.66	99.66
tags-math	5.32	5.57	3.94	6.74	5.00	5.50	104.7	104.7	102.2	0.94	104.7	104.7
threads-ubuntu	12.40	12.49	10.87	-	12.96	13.61	0.77	0.77	1.12	-	0.77	0.77
threads-math	36.91	31.33	11.02	-	32.92	31.20	1.02	1.02	2.07	-	1.02	1.02
coauth-DBLP	8.85	8.73	-	-	6.62	193.7	4.19	4.19	-	-	4.19	4.19
coauth-geology	144.5	124.9	-	-	151.1	143.17	5.77	5.77	-	-	5.77	5.77
coauth-history	8.13	9.47	-	-	10.80	13.98	2.38	2.38	-	-	2.38	2.38

:- out of time (taking more than 10 hours) or out of memory. Also, the range of hyperedge sizes is small in some datasets, which make some ratio not available.

Dataset	Intersection Sizes					Singular Values						
	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺
email-Enron	0.68	0.75	2.54	0.91	0.94	2.44	30.58	32.12	25.84	17.51	31.55	32.83
email-Eu	33.45	1.10	-	1.54	6.45	6.12	65.75	70.35	-	58.63	72.45	72.75
contact-primary	1.69	2.51	1.42	1.05	1.01	0.77	74.84	73.73	47.64	24.25	74.98	76.11
contact-high	2.20	0.28	1.33	1.04	0.50	3.67	70.23	68.21	48.60	28.60	68.82	67.63
NDC-classes	13.76	1.32	7.59	1.53	0.20	0.65	44.35	33.93	43.97	62.20	33.17	34.35
NDC-substances	36.31	7.89	-	0.51	13.64	8.24	80.90	80.75	-	121.3	80.85	80.77
tags-ubuntu	127.5	6.16	42.08	0.51	14.24	1.48	139.3	137.6	141.9	105.3	183.7	140.2
tags-math	174.5	10.04	45.86	0.51	31.95	3.10	119.4	121.8	127.9	80.45	126.6	122.2
threads-ubuntu	0.89	-1.41	1.54	-	-1.41	0.01	115.7	116.0	91.06	-	115.9	115.9
threads-math	3.92	8.86	-	-	0.49	0.86	189.6	184.7	-	-	188.8	188.7
coauth-DBLP	2.58	-60.3	-	-	1.05	2.35	178.4	173.5	-	-	164.8	151.4
coauth-geology	5.13	0.41	-	-	0.64	2.12	159.5	151.2	-	-	148.3	150.0
coauth-history	3.25	0.76	-	-	0.61	3.11	112.5	112.1	-	-	112.1	110.4

:- out of time (taking more than 10 hours) or out of memory. The range of intersection sizes is small in some datasets, which makes some ratios not available.

Dataset	Intersection Sizes					Singular Values						
	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺	Real	H-CL	H-PA	H-FF	H-LAP	H-LAP ⁺
email-Enron	0.68	0.75	2.54	0.91	0.94	2.44	30.58	32.12	25.84	17.51	31.55	32.83
email-Eu	33.45	1.10	-	1.54	6.45	6.12	65.75	70.35	-	58.63	72.45	72.75
contact-primary	1.69	2.51	1.42	1.05	1.01	0.77	74.84	73.73	47.64	24.25	74.98	76.11
contact-high	2.20	0.28	1.33	1.04	0.50	3.67	70.23	68.21	48.60	28.60	68.82	67.63
NDC-classes	13.76	1.32	7.59	1.53	0.20	0.65	44.35	33.93	43.97	62.20	33.17	34.35
NDC-substances	36.31	7.89	-	0.51	13.64	8.24	80.90	80.75	-	121.3	80.85	80.77
tags-ubuntu	127.5	6.16	42.08	0.51	14.24	1.48	139.3	137.6	141.9	105.3	183.7	140.2
tags-math	174.5	10.04	45.86	0.51	31.95	3.10	119.4	121.8	127.9	80.45	126.6	122.2
threads-ubuntu	0.89	-1.41	1.54	-	-1.41	0.01	115.7	116.0	91.06	-	115.9	115.9
threads-math	3.92	8.86	-	-	0.49	0.86	189.6	184.7	-	-	188.8	188.7
coauth-DBLP	2.58	-60.3	-	-	1.05	2.35	178.4	173.5	-	-	164.8	151.4
coauth-geology	5.13	0.41	-	-	0.64	2.12	159.5	151.2	-	-	148.3	150.0
coauth-history	3.25	0.76	-	-	0.61	3.11	112.5	112.1	-	-	112.1	110.4

:- out of time (taking more than 10 hours) or out of memory. The range of intersection sizes is small in some datasets, which makes some ratios not available.

Table 14: D-statistics between the distributions regarding Observations 1-5 in real-world hypergraphs and those in hypergraphs generated by three potential null models: HYPERCL, the bipartite graph generative model (Bipartite) [1], and the hypergraph configuration model (Configuration) [2]. Overall, all three models generate similarly realistic hypergraphs.

		Observation 1	Observation 2	Observation 3	Observation 4	Observation 5		
1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	email-Enron	HyperCL	0.545	0.517	0.143	0.089	0.498	2031
		Bipartite	0.580	0.487	0.122	0.095	0.518	2032
		Configuration	0.538	0.517	0.148	0.125	0.514	2033
1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	email-Eu	HyperCL	0.724	0.543	0.225	0.480	0.505	2034
		Bipartite	0.755	0.526	0.210	0.477	0.522	2035
		Configuration	0.732	0.545	0.225	0.482	0.506	2036
1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	contact-primary	HyperCL	0.896	0.867	0.196	0.137	0.430	2037
		Bipartite	0.966	0.760	0.129	0.130	0.388	2038
		Configuration	0.904	0.859	0.197	0.139	0.426	2039
1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	contact-high	HyperCL	0.948	0.874	0.277	0.210	0.423	2040
		Bipartite	0.993	0.877	0.195	0.207	0.411	2041
		Configuration	0.951	0.871	0.275	0.210	0.423	2042
1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	NDC-classes	HyperCL	0.694	0.302	0.273	0.376	0.274	2043
		Bipartite	0.714	0.305	0.274	0.375	0.304	2044
		Configuration	0.688	0.342	0.292	0.385	0.268	2045
1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	NDC-substances	HyperCL	0.451	0.321	0.272	0.521	0.377	2046
		Bipartite	0.464	0.320	0.273	0.525	0.408	2047
		Configuration	0.449	0.348	0.283	0.539	0.349	2048
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	tags-ubuntu	HyperCL	0.522	0.432	0.091	0.148	0.245	2049
		Bipartite	0.666	0.432	0.073	0.120	0.219	2050
		Configuration	0.524	0.439	0.094	0.147	0.238	2051
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	tags-math	HyperCL	0.496	0.460	0.095	0.209	0.337	2052
		Bipartite	0.578	0.439	0.074	0.170	0.302	2053
		Configuration	0.503	0.468	0.097	0.209	0.327	2054
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	threads-ubuntu	HyperCL	0.159	0.299	0.011	0.004	0.020	2055
		Bipartite	0.488	0.299	0.006	0.004	0.267	2056
		Configuration	0.076	0.013	0.012	0.004	0.022	2057
2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	threads-math	HyperCL	0.137	0.232	0.041	0.006	0.060	2058
		Bipartite	0.472	0.232	0.031	0.005	0.231	2059
		Configuration	0.153	0.078	0.042	0.006	0.062	2060
2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	coauth-DBLP	HyperCL	0.228	0.302	0.224	0.215	0.715	2061
		Bipartite	0.379	0.322	0.224	0.215	0.713	2062
		Configuration	0.215	0.307	0.224	0.215	0.715	2063
2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	coauth-geology	HyperCL	0.200	0.248	0.178	0.086	0.624	2064
		Bipartite	0.333	0.247	0.178	0.086	0.623	2065
		Configuration	0.171	0.227	0.178	0.086	0.624	2066
2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	coauth-history	HyperCL	0.087	0.316	0.033	0.001	0.154	2067
		Bipartite	0.281	0.318	0.033	0.001	0.229	2068
		Configuration	0.045	0.065	0.033	0.001	0.155	2069
2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030	Average	HyperCL	0.468	0.439	0.158	0.191	0.359	2070
		Bipartite	0.590	0.428	0.140	0.185	0.395	2071
		Configuration	0.458	0.391	0.162	0.196	0.356	2072