SUPPLEMENTARY MATERIAL

1 Time Proportion of Each Part of Algorithm

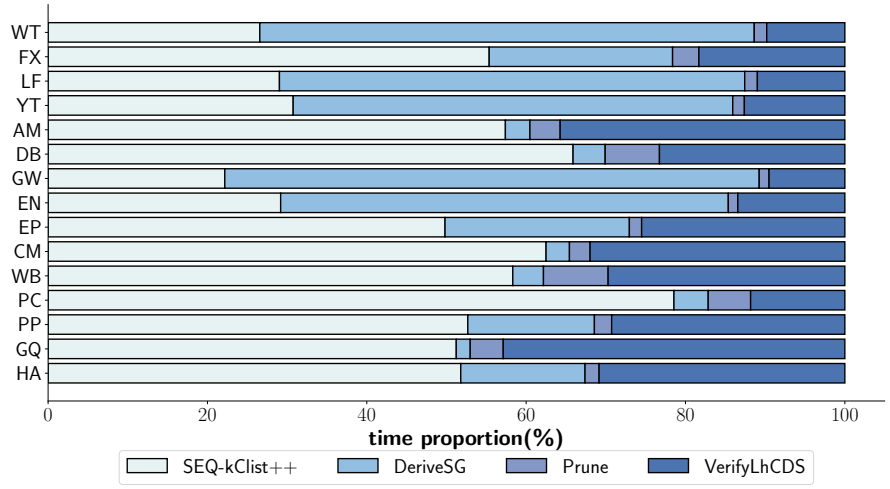


Figure 1: Time Proportion of Each Part of IPVAC

2 Basic Idea of Enumeration of Some Patterns

2.1 Star

The star with h vertices is called $(h-1)$-star. For each vertex v in a graph G, we enumerate all its neighbors, and then any instance containing v itself and $(h-1)$ neighbors of v is the $(h-1)$-star pattern.

For each vertex v in a graph, there are two different types of $(h-1)$-star that containing v. In the first type, v is the star-vertex, and in the second type, v is a tail-vertex. The sets of the two types are disjoint. Suppose v has y neighbors, and u is the i-th neighbor of v $(1 \leq i \leq y)$. The degree of u is z_i. We thus have $deg_G(v, \psi_{hx}) = \binom{y}{h-1} + \sum_{1 \leq i \leq y \wedge (h-1) \leq z_i} \binom{z_i - 1}{h-2}$.

2.2 Loop

The loop with h vertices is called h-loop. If h is even, the loop is an even loop; and if h is odd, the loop is an odd.

Even Loop. For each vertex v in a graph G, we first enumerate all the paths from v to its $\frac{h}{2}$-hop neighbors, and then organize these paths into r groups, each of which share the same $\frac{h}{2}$-hop neighbor. Then, for each group, any pair of paths can form an instance of the even loop pattern with h vertices.

Let the size of the i-th group be $y_i (1 \leq i \leq r)$. We thus have $deg_G(v, \psi_{hx}) = \sum_{1 \leq i \leq r \wedge y_i \geq 2} \binom{y_i}{2}$.

Odd Loop. For each vertex v in a graph G, we first enumerate all the paths from v to its $\frac{h-1}{2}$-hop neighbors and $\frac{h+1}{2}$-hop neighbors, and then organize these paths into r_a groups and r_b groups, each of which share the same $\frac{h-1}{2}$-hop neighbor and $\frac{h+1}{2}$-hop neighbor. Then, for each pair of groups, any pair of paths from group a and group b can form an instance of the odd loop pattern with h vertices.

Let the size of the i-th a group be $y_i (1 \leq i \leq r_a)$ and the i-th b group be $z_i (1 \leq i \leq r_b)$. We thus have $deg_G(v, \psi_{hx}) = \sum_{1 \leq i \leq r_a \wedge i \leq r_b \wedge y_i \geq 1 \wedge z_i \geq 1} \binom{y_i}{1}\binom{z_i}{1}$.

2.3 n-Triangle

The pattern with n triangle, $(2 + n)$ vertices and $(1 + 2n)$ edges is called n-triangle. For each edge e in a graph G, we enumerate all triangles containing e. Then, any n triangles of the calculated triangles that containing e can form an instance of the n-triangle pattern.

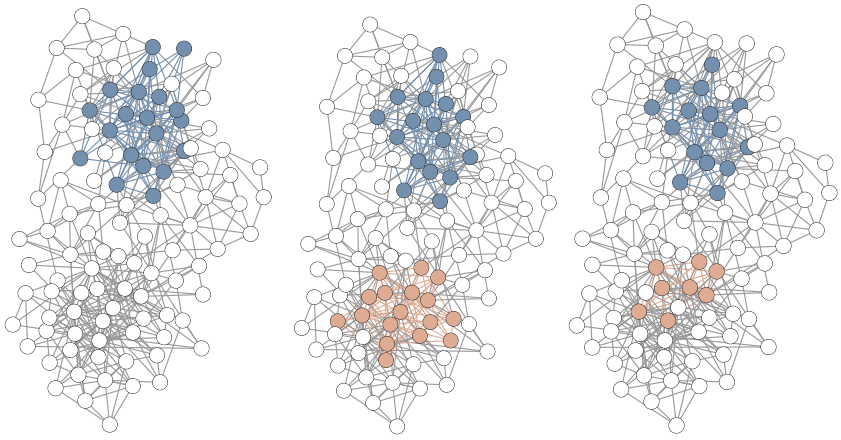
For each vertex v in a graph, there are two different types of n-triangle that containing v. In the first type, v is the vertex with $(n + 1)$ neighbors, and in the second type, v is the vertex with two neighbors. The sets of the two types are disjoint.

2.4 c3-Star

The pattern with a triangle, four vertices and four edges is called c3-star. For each vertex v in a graph G, we enumerate all triangles containing v. for each triangle that containing v, we enumerate all neighbors of v that is not contained in the triangle. Then, any instance with the triangle that containing v and the calculated neighbor of v is the c3-star pattern.

For each vertex v in a graph, there are three different types of c3-star that containing v. In the first type, v is the vertex with three neighbors; in the second type, v is the vertex with two neighbors; and in the third type, v is the vertex with only one neighbor. The sets of the three types are disjoint.

3 Visualization results of LhCDS with different h



(a) 3-clique (b) 4-clique (c) 5-clique

Figure 2: LhCDS case study on real network (the top-1 LhCDS: steelblue; the top-2 LhCDS: orange vertices)

4 All the Data in Our Paper

Table 1: Subgraph statistics of h-density and size on CA-CondMat ($h = 3$)

	Greedy		Top-k LhCDS	
top-k	size	h-clique density	size	h-clique density
1	30	109.63	30	109.63
2	69	77.81	19	51.00
3	23	53.17	17	40.00
4	19	51.00	16	35.00
5	24	41.37	13	22.00
6	124	41.37	13	22.00
7	17	40.00	12	18.33
8	18	37.22	18	15.83
9	23	35.17	9	9.33
10	32	35.00	9	9.33

Table 2: Subgraph statistics of h-density and size on CA-CondMat ($h = 5$)

	Greedy		Top-k LhCDS	
top-k	size	h-clique density	size	h-clique density
1	28	2736.07	28	2736.07
2	23	1463.00	23	1463.00
3	19	612.00	19	612.00
4	22	562.00	17	364.00
5	55	453.00	18	338.00
6	34	364.00	16	273.00
7	18	338.00	13	99.00
8	18	284.00	13	99.00
9	32	273.00	13	99.00
10	33	246.82	12	66.00

Table 3: Subgraph statistics of h-density and size on fb-pages-company ($h = 3$)

	Greedy		Top-k LhCDS	
top-k	size	h-clique density	size	h-clique density
1	32	100.25	32	100.25
2	43	89.70	43	89.70
3	24	70.67	24	70.67
4	32	53.16	21	49.14
5	21	49.14	20	47.80
6	20	47.80	19	44.63
7	19	44.63	40	42.08
8	40	42.08	18	40.44
9	18	40.44	16	29.31
10	25	37.72	16	28.00

Table 4: Subgraph statistics of h-density and size on fb-pages-company ($h = 5$)

	Greedy		Top-k LhCDS	
top-k	size	h-clique density	size	h-clique density
1	30	1918.97	30	1918.97
2	24	1046.58	24	1046.58
3	37	820.59	37	820.60
4	20	488.75	20	488.75
5	20	471.90	20	471.90
6	19	411.21	19	411.21
7	18	333.67	18	333.67
8	30	308.33	16	152.81
9	24	174.46	16	135.00
10	16	152.81	34	124.21

Table 5: Memory usage of algorithms (kb) with $k = 5$, $h = 3$

Dataset	**IPVAC**	**LTDS**
soc-hamsterster	54,268	53,108
CA-GrQc	31,544	52,044
fb-pages-politician	153,004	160,220
fb-pages-company	27,332	64,560
web-webbase-2001	18,128	31,536
CA-CondMat	125,544	170,696
soc-epinions	132,048	161,964
Email-Enron	583,648	688,596
loc-gowalla	1,261,932	2,143,860
DBLP	1,253,444	2,149,380
Amazon	1,023,348	798,236
soc-youtube	1,490,284	2,477,512
soc-lastfm	1,921,144	4,228,120
soc-flixster	3,246,136	8,443,976
soc-wiki-talk	7,311,900	9,277,812

Table 6: Running time of algorithms with different h (= 3,4,5) and k.

Dataset	h	Algorithm	k			
			5	10	15	20
fb-pages-company	3	basic	1.65	4.60	9.83	13.73
		fast	0.52	0.85	1.65	2.98
	4	basic	3.50	7.44	11.08	17.12
		fast	1.26	1.49	1.76	2.38
	5	basic	7.47	19.29	31.12	51.18
		fast	2.56	3.56	3.47	4.13
soc-hamsterster	3	basic	8.67	12.25	12.63	14.46
		fast	7.50	8.99	8.98	8.96
	4	basic	34.74	37.13	43.22	48.77
		fast	27.56	26.86	26.89	27.02
	5	basic	20.49	31.73	38.48	48.29
		fast	10.13	11.17	11.72	11.73
soc-epinions	3	basic	81.82	125.23	108.61	145.22
		fast	74.27	73.62	73.09	74.63
	4	basic	643.26	662.74	673.51	686.52
		fast	620.26	628.43	614.40	614.48
	5	basic	216.48	243.30	243.49	277.06
		fast	194.03	191.00	193.59	193.38
Email-Enron	3	basic	1445.47	1525.21	1548.91	1592.20
		fast	1369.84	1438.16	1440.30	1459.11
	4	basic	10040.42	10147.54	10289.65	10440.24
		fast	9926.20	9923.58	9910.32	9868.51
	5	basic	10354.91	10716.25	10786.24	10974.85
		fast	10184.32	9980.45	9759.69	10027.16
loc-gowalla	3	basic	5310.25	5819.97	5909.72	6225.08
		fast	5095.63	5188.29	5186.31	5224.33
	4	basic	26695.73	27388.21	27704.84	27902.67
		fast	26615.71	26338.37	26066.67	26231.87
	5	basic	23874.49	26000.95	26476.11	27361.03
		fast	22497.99	22054.19	22499.65	22070.97
CA-CondMat	3	basic	30.41	53.31	76.28	80.96
		fast	22.34	26.84	29.25	28.55
	4	basic	32.25	63.64	110.91	164.01
		fast	10.00	15.73	19.77	21.15
	5	basic	32.25	63.64	110.91	164.01
		fast	10.00	15.73	19.77	21.15
CA-GrQc	3	basic	1.08	2.31	6.07	8.69
		fast	0.35	0.47	0.63	0.54
	4	basic	18.25	36.74	45.37	66.73
		fast	8.71	9.02	9.66	7.81
	5	basic	104.16	191.15	299.52	372.63
		fast	37.29	33.90	35.72	36.79
Amazon	3	basic	1154.16	1169.51	1182.00	1218.40
		fast	937.88	952.93	946.26	946.54
	4	basic	245.87	258.92	263.26	278.65
		fast	239.62	238.13	241.10	234.59
	5	basic	45.88	54.49	57.87	63.37
		fast	24.22	24.38	24.59	23.88