Appendix:

We have conducted a series of experiments to find the similarity and difference across several state-of-the-art graph processing frameworks/benchmark suites [12, 2, 3, 10, 9], in order to offer insights for architects working on hardware design for graph analytics applications, and to provide suggestion on benchmark selection. To present a clearer flow in the paper: **Demystifying Graph Analytics Frameworks and Benchmarks**, we choose not to include all the diagrams and tables, but provide the supplemental data in this appendix.
There are thirteen graph kernels from four frameworks/benchmark suites (in Table 1) and five real-world graphs (in Talbe 2) involved in our experiments.

Table 1: Graph kernels in the graph frameworks/benchmark suites

Graph Frameworks	Graph Kernels (Algorithm)
GraphMat [12]	**BFS** (Map to Sparse Matrix Vector Multiplication), **SSSP** (Map to Sparse Matrix Vector Multiplication), **PR** (Map to Sparse Matrix Vector Multiplication), **TC** (Map to Sparse Matrix Vector Multiplication), and 6 more
GAP [2, 3]	**BFS** (direction-optimizing [1]), **SSSP** (Δ-stepping [8]), **PR** (classic approach), **TC** (terminate intersection + descent degree), Connected Component and Betweenness Centrality
GraphBIG [10]	**BFS** (Top-down), **SSSP** (Dijkstra's Algorithm [4]), **PR** (Classic approach), **TC** (Schank's Algorithm [11]), and 25 more
Graph500 [9]	**BFS** (direction-optimizing [1]), and single-thread SSSP

Table 2: Graph Datasets for Shared Memory Graph System

dataset	#nodes	#edges	diameter	degree	type
Flickr(FLI) [7]	0.83M	9.84M	18	11.87	Directed
LiveJournal(LJ) [6]	4.85M	68.99M	16	14.23	Directed
Orkut(OK) [6]	3.07M	117.19M	9	76.28	Undirected
roadCA(CA) [5]	1.97M	2.76M	849	2.82	Undirected
roadTX(TX) [5]	1.38M	1.92M	1054	2.79	Undirected

To guide readers through this appendix and help link to the paper, we summaries the contents as follows:

- Table 3 to Table 8 shows the absolute values of six metrics (i.e., *Dara Movement Per Edge*, *Computation Per Edge*, *Energy Consumption Per Edge*, *IPC*, *L1D MPKI*, *L2 MPKI*). The correlation analysis, Principle Component Analysis (PCA), and other statistic analyses are all based on these data.
- Figure 1 presents 20 Kiviat diagrams for 20 test tasks (running one of four applications on one of five graphs is one task), and each Kiviat diagram summarizes 8 metrics for 3 or 4 implementations from different frameworks/benchmark suites. Figure 1t and 1r are included in the paper Section IV-B to illustrate the hardware metrics pitfalls.
- Figure 2 displays 20 scalability charts for 20 test tasks. Figure 2i and 2n are selected in paper Section IV-C to elaborate the findings regarding scalability.
- Figure 3 is the dendrogram for all the graph workloads studied. Dendrogram serves as a complement to the PCA scatter plot shown in paper Section IV-D, because it is not only showing what workloads are more closer to others, but also let users form a subset of workloads with specific budget.

Table 3: Data Movement Per Edge of Graph Applications on Different Graphs.

	BFS				**SSSP**			**PageRank**			**Triangle Counting**		
Graphs	GraphMat	Graph500	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG
FLI	139.15	18.18	2.73	21.57	319.75	147.32	542.19	235.32	135.96	615.17	2,688.39	195.37	1,775.98
LJ	110.98	16.65	2.41	20.15	383.62	55.97	471.44	293.21	114.54	293.63	735.40	41.92	312.39
OK	61.80	13.55	0.73	51.15	243.32	27.83	364.92	156.93	52.58	333.92	1,476.12	161.67	1,640.08
CA	4,567.69	147.60	24.83	142.78	8,409.26	40,675.98	828.99	521.50	524.09	791.52	2,418.86	8.24	68.18
TX	6,181.48	199.32	29.90	228.13	10,574.72	59,618.98	858.64	551.17	537.54	792.22	2,861.46	8.06	86.00

Table 4: IPC of Graph Applications on Different Graphs.

	BFS				**SSSP**			**PageRank**			**Triangle Counting**		
Graphs	GraphMat	Graph500	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG
FLI	0.44	0.58	0.31	0.04	0.86	0.02	0.08	0.68	0.36	0.05	0.96	0.99	0.67
LJ	0.57	0.54	0.25	0.04	1.26	0.03	0.09	1.24	0.35	0.07	0.73	0.82	0.30
OK	0.6	0.71	0.2	0.08	1.54	0.03	0.09	1.31	0.43	0.05	1.08	0.62	0.75
CA	1.38	0.33	0.05	0.04	0.88	0.03	0.07	1.10	0.23	0.07	0.39	0.40	0.02
TX	1.3	0.27	0.05	0.05	0.96	0.03	0.07	0.89	0.28	0.07	0.37	0.38	0.03

Table 5: L1D MPKI of Graph Applications on Different Graphs.

	BFS				**SSSP**			**PageRank**			**Triangle Counting**		
Graphs	GraphMat	Graph500	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG
FLI	10.87	33.83	15.19	116.23	21.46	19.55	147.87	18.99	104.32	91.08	3.37	5.58	10.43
LJ	12.77	29.91	35.46	132.45	22.52	24.49	130.04	23.22	94.80	97.43	8.15	8.69	16.20
OK	13.89	34.95	30.14	69.61	19.06	27.39	146.32	18.79	141.46	126.12	0.13	9.34	13.27
CA	18.44	55.63	51.56	53.76	24.22	15.16	122.03	17.82	31.09	59.13	5.14	18.69	57.69
TX	18.65	79.22	47.56	36.34	24.84	19.27	126.74	16.66	32.59	62.51	4.54	24.08	45.20

Table 6: L2 MPKI of Graph Applications on Different Graphs.

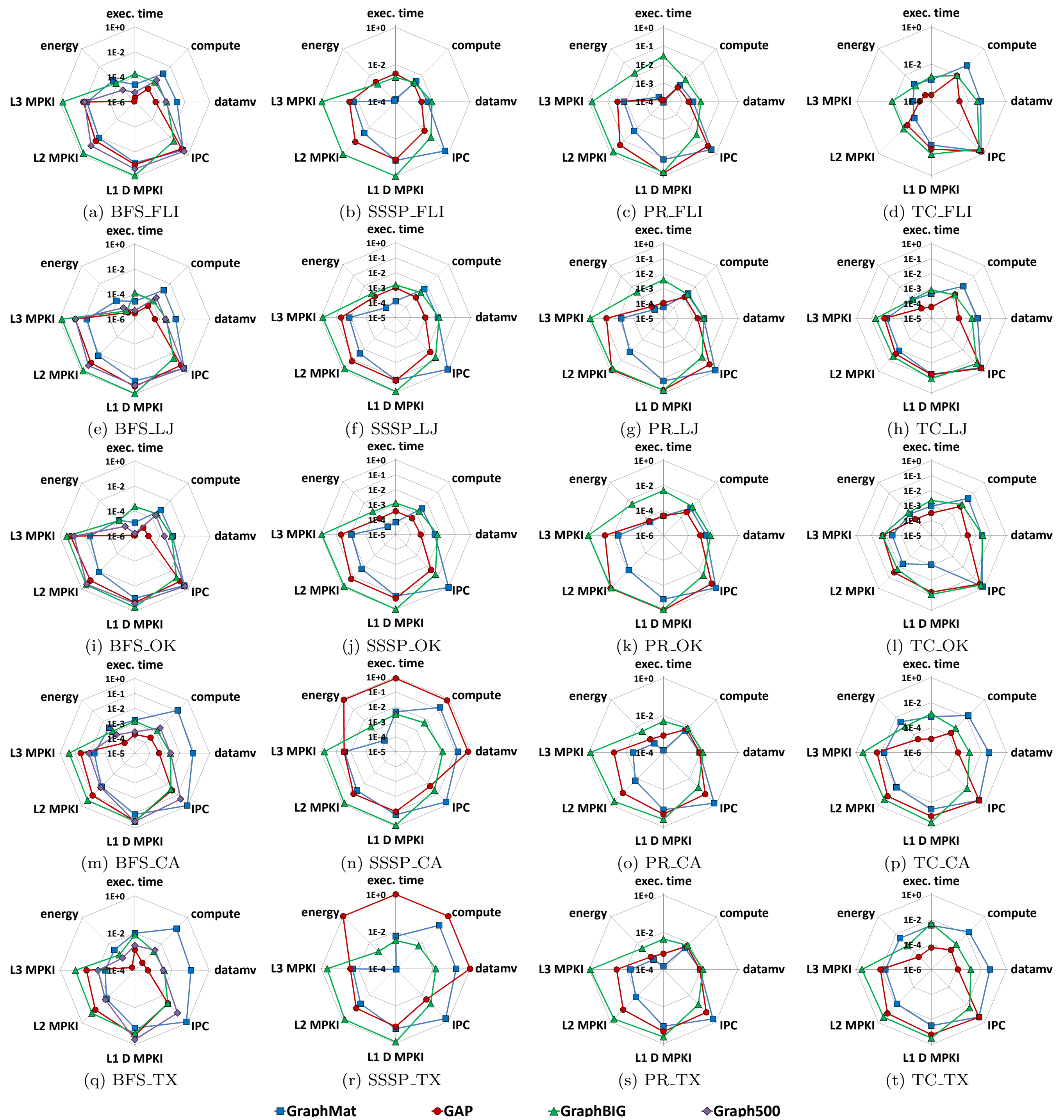
	BFS				**SSSP**			**PageRank**			**Triangle Counting**		
Graphs	GraphMat	Graph500	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG
FLI	0.84	6.49	1.91	47.08	1.76	8.42	72.79	1.25	14.58	47.86	0.14	0.48	0.92
LJ	0.98	12.67	6.43	50.25	1.80	9.95	49.49	1.09	56.15	45.27	0.87	1.72	3.01
OK	0.82	20.24	7.64	23.16	1.24	11.62	55.41	0.61	69.83	61.39	0.36	2.34	1.16
CA	1.07	1.24	7.26	21.38	3.42	7.10	53.86	1.01	8.94	39.32	0.63	6.36	14.95
TX	1.05	1.25	7.30	13.44	3.38	7.35	54.53	0.89	8.36	42.24	0.56	6.65	18.07

Table 7: Computation Per Edge of Graph Applications on Different Graphs.

	BFS				**SSSP**			**PageRank**			**Triangle Counting**		
Graphs	GraphMat	Graph500	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG
FLI	222.54	40.03	4.40	25.55	512.40	331.22	391.41	249.40	170.64	658.27	7,404.83	1,217.74	1,274.86
LJ	260.12	37.56	4.26	16.81	696.50	110.92	323.73	305.13	140.60	254.30	1,486.62	249.40	232.60
OK	120.43	31.80	1.26	45.69	415.63	49.28	230.36	160.71	57.95	259.47	4,212.12	755.10	1,126.68
CA	15768.00	346.97	42.00	184.40	21,163.55	98,523.89	740.63	580.80	809.61	967.69	2,134.86	23.83	78.66
TX	20701.32	475.85	51.80	388.38	27,873.43	138,704.65	771.28	606.82	830.77	952.81	2,516.21	23.20	94.42

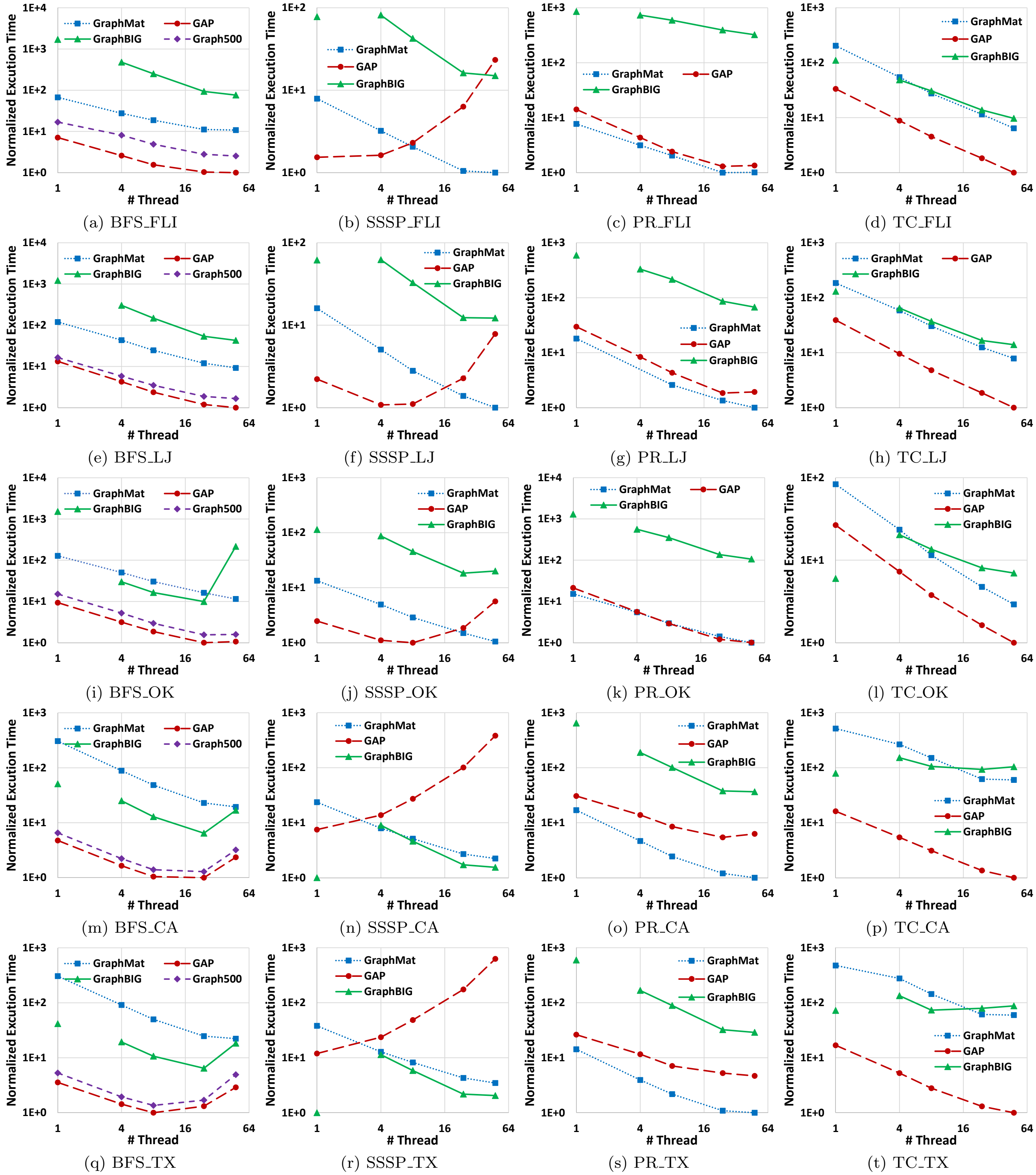
Table 8: Energy Consumption Per Edge of Graph Applications on Different Graphs (in Joules).

	BFS				**SSSP**			**PageRank**			**Triangle Counting**		
Graphs	GraphMat	Graph500	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG	GraphMat	GAP	GraphBIG
FLI	3.58E-06	2.96E-07	1.52E-08	1.74E-06	1.60E-06	4.18E-05	2.97E-05	2.96E-06	1.92E-06	2.01E-04	2.64E-05	3.80E-06	1.98E-05
LJ	1.52E-06	2.54E-07	7.52E-08	1.11E-07	1.11E-06	1.28E-05	2.47E-05	8.96E-07	1.72E-06	4.12E-05	7.44E-06	1.14E-06	8.09E-06
OK	8.04E-07	1.62E-07	1.64E-08	7.58E-07	7.43E-07	4.38E-06	1.93E-05	4.24E-07	5.41E-07	4.56E-05	1.37E-05	4.44E-06	1.89E-05
CA	3.44E-05	7.26E-06	1.25E-06	1.59E-05	1.53E-06	1.07E-02	2.92E-05	6.43E-06	1.32E-05	5.36E-05	3.86E-05	4.17E-07	1.15E-05
TX	4.73E-05	1.13E-05	2.15E-06	1.96E-05	1.18E-06	1.31E-02	2.68E-05	7.46E-06	1.21E-05	5.37E-05	4.54E-05	3.50E-07	6.36E-06



(a) BFS_FLI (b) SSSP_FLI (c) PR_FLI (d) TC_FLI
(e) BFS_LJ (f) SSSP_LJ (g) PR_LJ (h) TC_LJ
(i) BFS_OK (j) SSSP_OK (k) PR_OK (l) TC_OK
(m) BFS_CA (n) SSSP_CA (o) PR_CA (p) TC_CA
(q) BFS_TX (r) SSSP_TX (s) PR_TX (t) TC_TX

Figure 1: Kiviat diagrams of graph applications with different graphs. Data are collected on 48-core setup. BFS is the only kernel that has OpenMP implementation in Graph500.



(a) BFS_FLI (b) SSSP_FLI (c) PR_FLI (d) TC_FLI
(e) BFS_LJ (f) SSSP_LJ (g) PR_LJ (h) TC_LJ
(i) BFS_OK (j) SSSP_OK (k) PR_OK (l) TC_OK
(m) BFS_CA (n) SSSP_CA (o) PR_CA (p) TC_CA
(q) BFS_TX (r) SSSP_TX (s) PR_TX (t) TC_TX

Figure 2: Performance comparison and scalability of graph applications with different graphs. BFS is the only kernel that has OpenMP implementation in Graph500. Normalized with respect to lowest execution time in each chart.

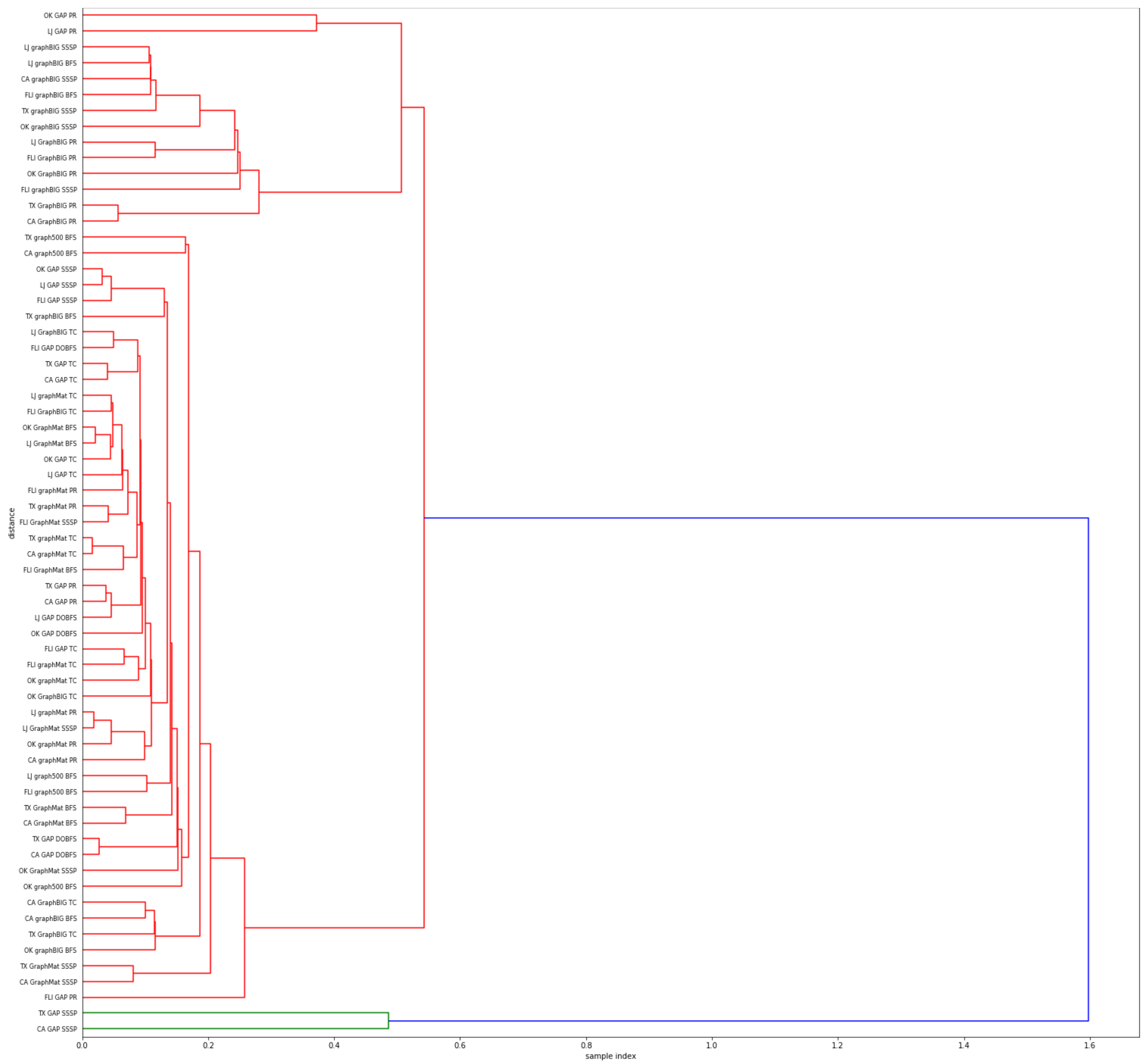


Figure 3: Hierarchical clustering of graph analytics workloads presented as a dendrogram, which is useful to form a subset with certain number of workloads.

References

[1] S. Beamer, K. Asanović, and D. Patterson. Direction-optimizing breadth-first search. *Scientific Programming*, 21(3-4):137–148, 2013.

[2] S. Beamer, K. Asanović, and D. Patterson. The gap benchmark suite. *arXiv preprint arXiv:1508.03619*, 2015.

[3] S. Beamer, K. Asanovic, and D. Patterson. Locality exists in graph processing: Workload characterization on an ivy bridge server. In *2015 IEEE International Symposium on Workload Characterization*, pages 56–65, Oct 2015.

[4] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numer. Math.*, 1(1):269–271, Dec. 1959.

[5] J. Leskovec and A. Krevl. {SNAP Datasets}:{Stanford} large network dataset collection. 2015.

[6] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.

[7] J. McAuley and J. Leskovec. Image labeling on a network: using social-network metadata for image classification. In *European conference on computer vision*, pages 828–841. Springer, 2012.

[8] U. Meyer and P. Sanders. Delta-stepping: A parallel single source shortest path algorithm. In *Proceedings of the 6th Annual European Symposium on Algorithms*, ESA '98, pages 393–404, London, UK, UK, 1998. Springer-Verlag.

[9] R. C. Murphy, K. B. Wheeler, B. W. Barrett, and J. A. Ang. Introducing the graph 500. *Cray User's Group (CUG)*, 19:45–74, 2010.

[10] L. Nai, Y. Xia, I. G. Tanase, H. Kim, and C.-Y. Lin. Graphbig: understanding graph computing in the context of industrial solutions. In *High Performance Computing, Networking, Storage and Analysis, 2015 SC-International Conference for*, pages 1–12. IEEE, 2015.

[11] T. Schank and D. Wagner. Finding, counting and listing all triangles in large graphs, an experimental study. In *Proceedings of the 4th International Conference on Experimental and Efficient Algorithms*, WEA'05, pages 606–609, Berlin, Heidelberg, 2005. Springer-Verlag.

[12] N. Sundaram, N. Satish, M. M. A. Patwary, S. R. Dulloor, M. J. Anderson, S. G. Vadlamudi, D. Das, and P. Dubey. Graphmat: High performance graph analytics made productive. *Proceedings of the VLDB Endowment*, 8(11):1214–1225, 2015.