

A Sampling-Based Tool for Scaling Graph Datasets

ICPE2020 11th ACM / SPEC

International Conference on Performance Engineering

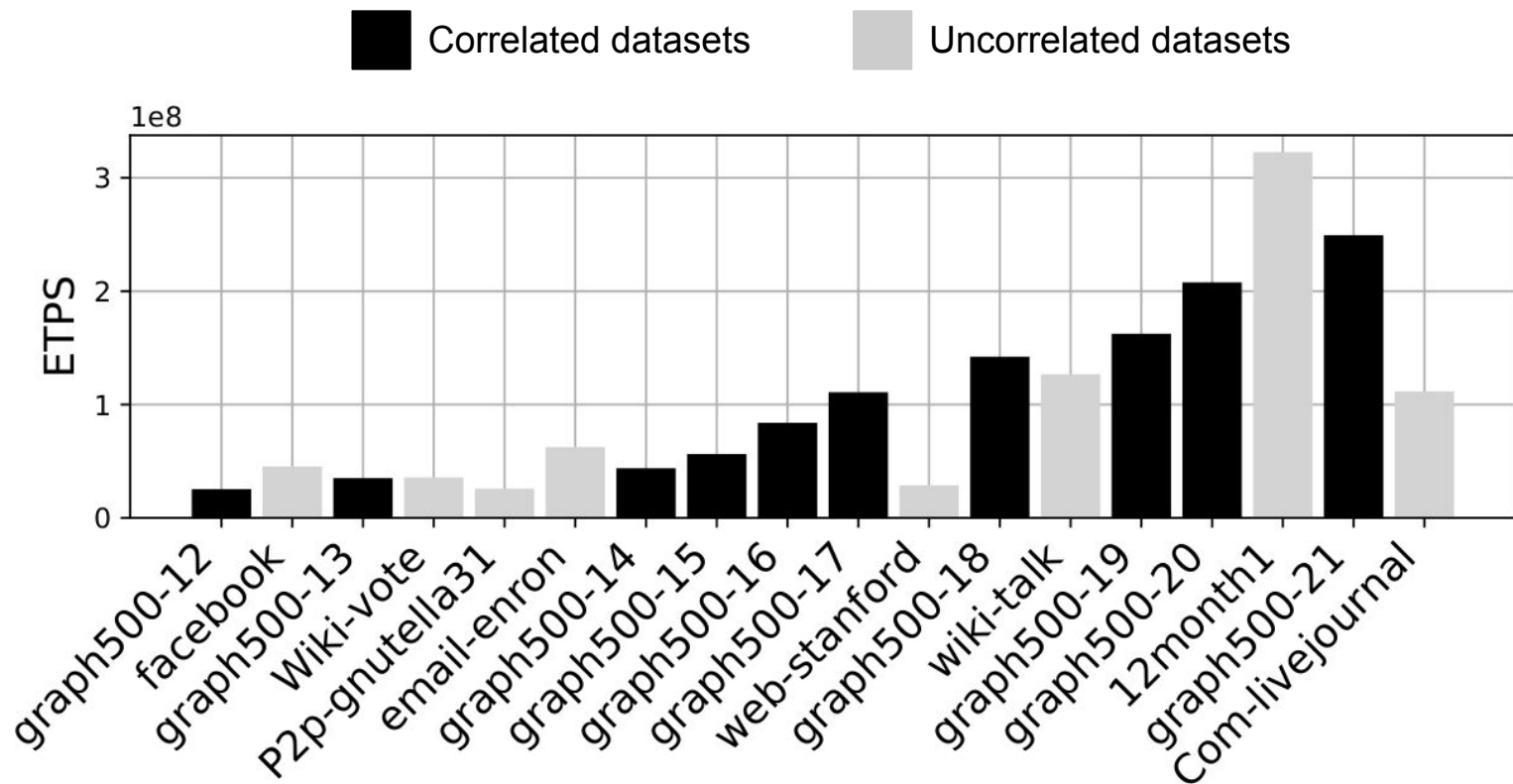
Ahmed Musaafir, Alexandru Uta, Henk Dreuning, Ana-Lucia Varbanescu

Vrije Universiteit Amsterdam & University of Amsterdam

Context

- Graph datasets
 - Used in different domains (e.g., logistics, biology, social networks, infrastructure networks)
- Graph processing
 - Different graph processing platforms: Giraph, GraphMat, Gunrock, etc.
- Graph analytics benchmarking
 - Platform, Algorithm, Dataset, Hardware
 - No in-depth evaluation or performance analysis
 - Which properties of the graph dataset affect performance?

Context



Problem

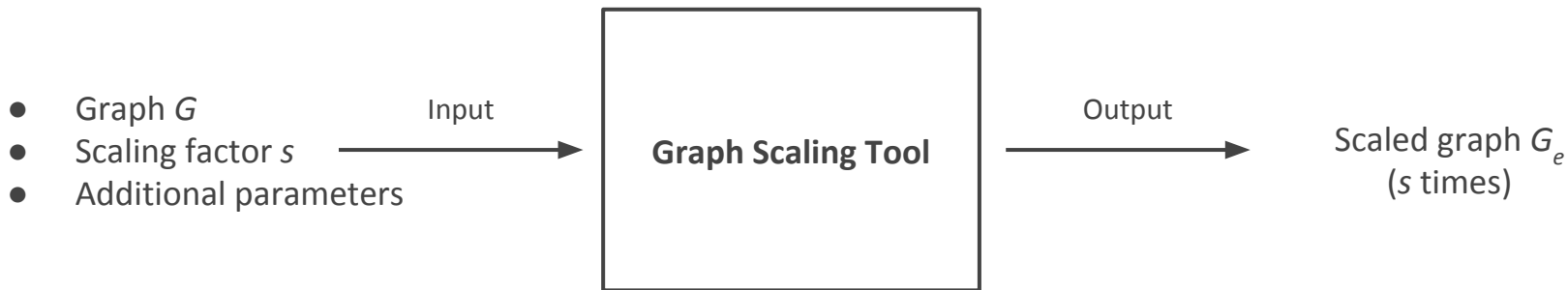
- Lack of representative graph datasets
- Synthetic graph generators
 - Generate a graph from scratch
 - Allow controlling specific graph properties only
- Graph archives
 - Few types of graphs
 - Small collection and size

Solution

- Graph scaling
- Control certain graph properties
 - Predict and tune the properties of scaled-up graphs based on models, guidelines
- Tool to generate diverse families of graphs fast

Solution

- Graph scaling
- Control certain graph properties
 - Predict and tune the properties of scaled-up graphs based on models, guidelines
- Tool to generate diverse families of graphs fast



Scaling Down

Scaling Down: Graph Sampling

- Node-based Sampling
 - Node Sampling
- Edge-based Sampling
 - Random Edge Sampling
 - Totally-Induced Edge Sampling (TIES)
- Traversal-based Sampling
 - Random Walk
 - Forest Fire

Scaling Down: Graph Sampling

- Node-based Sampling
 - Node Sampling
- Edge-based Sampling
 - Random Edge Sampling
 - Totally-Induced Edge Sampling (TIES)
- Traversal-based Sampling
 - Random Walk
 - Forest Fire

	Random Node	Random Edge	TIES	Random Walk	Forest Fire
<i>Scale (fraction based on)</i>	<i>V</i>	<i>E</i>	<i>V</i>	<i>V/E</i>	<i>V/E</i>
Degree distribution	+	+	++	+	+
Diameter	+	+	-	+	+
Connectivity	--	-	-	--	-
Avg. Clustering Coefficient (C.C.)	+	-	++	+	+
Avg. Shortest Path Length (S.P.L.)	+	-	++	+	++

Property preservation quality per sampling algorithm, represented as likelihood from low (--) to high (++)

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Down: Results

	Com-Orkut G (original)	Gs 0.8	Gs 0.5	Gs 0.3
#Nodes	3,072,441	2,457,952	1,536,220	921,733
#Edges	117,185,083	108,686,099	73,626,482	42,194,208
Avg. degree	76.28	88.44	95.85	91.55
Diameter	9	9	10	8
Density	2.48e-05	3.59e-05	6.24e-05	9.93e-05
Components	1	7	17	36
Avg. Clustering Coeff.	0.16	0.15	0.15	0.14
Avg. Shortest path	4.19	4.05	3.97	3.95

Scaling Up

Scaling Up: Existing work

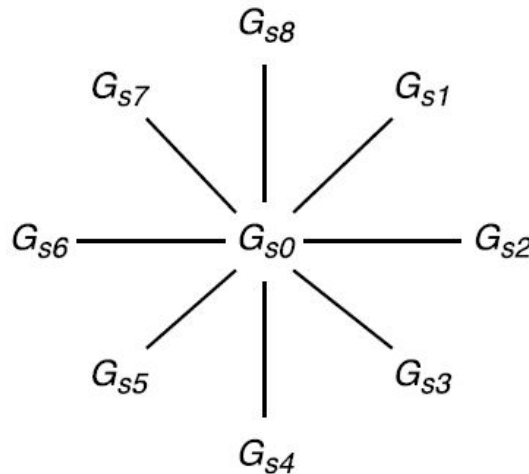
- Graph generators
 - Datagen, Graph500, R-MAT
- Graph evolution algorithms
 - Focus on evolving the graph
- Graph scalers
 - GScaler, ReCoN, Musketeer

Scaling Up: Method

- Obtain samples G_i of the original graph G
- Interconnect the different samples

Scaling Up: Method

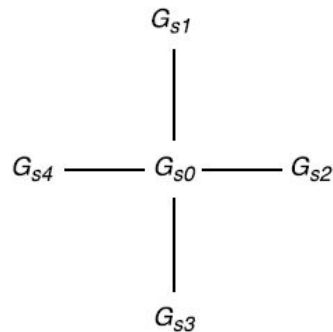
- Obtain samples G_i of the original graph G
- Interconnect the different samples
- Example: scale up a graph 4.5 times
 - Sample size: 0.5
 - Results in 9 different samples



Example of scaling up a graph
 $G_{s0...8}$ = Sampled versions of the graph

Scaling Up: Method

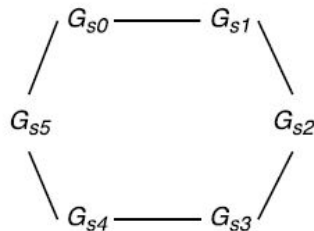
- Interconnection topologies
 - Star; Chain; Ring; Fully-connected
- Selecting bridge vertices
 - Random; High-degree
- Multi-edge interconnections
 - n number of interconnections
 - Directed; undirected



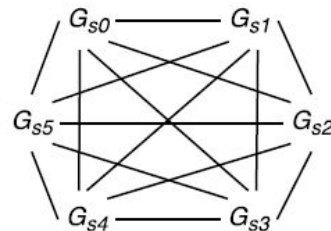
a. star



b. chain



c. ring



d. fully connected

Scaling Up: Impact on properties

- Different parameters
 - Interconnection topologies
 - Selecting bridge vertices
 - Multi-edge interconnections
- Sampling algorithm
 - Sample size
- Scaling factor
- Dataset

Scaling Up: Measuring the quality of graph output

- Given the same parameters, the properties of the expanded graph should be predictable.
- Models & guidelines
- "In case you want to have the scaled-up graph with a larger diameter, choose a chain topology with a single random bridge".

Scaling Up: Measuring the quality of graph output

- Given the same parameters, the properties of the expanded graph should be predictable.
- Models & guidelines
- "In case you want to have the scaled-up graph with a larger diameter, choose a chain topology with a single random bridge".

Maximum diameter:

$$\sum_1^n \max D(S_i) + (n-1)$$

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

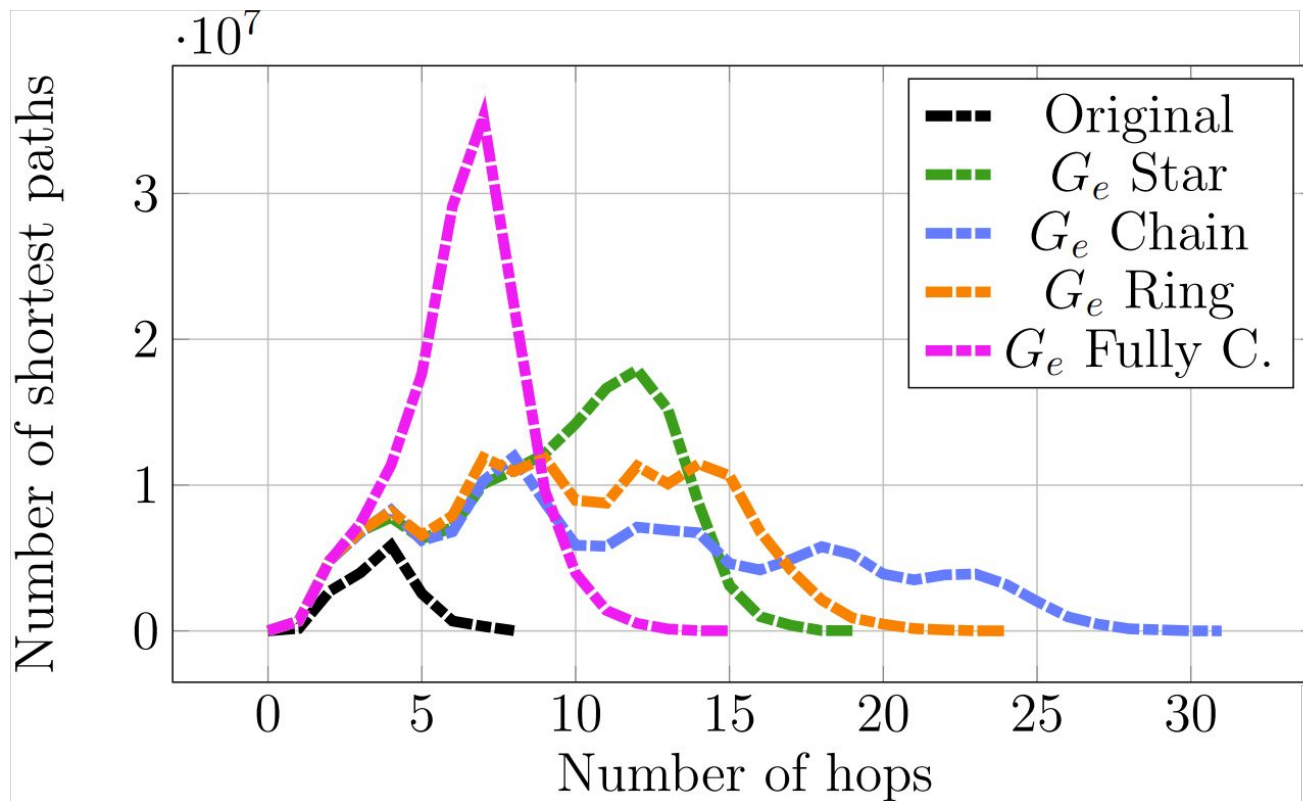
Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results

	FB G (original)	G x3	G x3	G x3	G x3	G x3
Sample size	-	0.5	0.5	0.5	0.5	0.5
Topology	-	Star	Chain	Fully Connected	Star	Star
Bridge	-	Random	Random	Random	Random	High-degree
#Interconnection	-	1	1	1	45,000	45,000
#Nodes	4,039	12,117	12,114	12,114	12,114	12,115
#Edges	88,234	339,497	340,091	339,777	559,798	560,168
Avg. degree	43.69	56.04	56.15	56.09	92.42	92.48
Diameter	8	19	31	15	6	6
Density	1.10e-2	4.62e-3	4.63e-3	4.63e-3	7.62e-3	7.63e-3
Components	1	7	9	7	2	10
Avg. Clustering Coeff.	0.62	0.63	0.63	0.63	0.31	0.46
Avg. Shortest path	3.69	9.26	11.79	6.35	2.65	2.92

Scaling Up: Results



Case studies

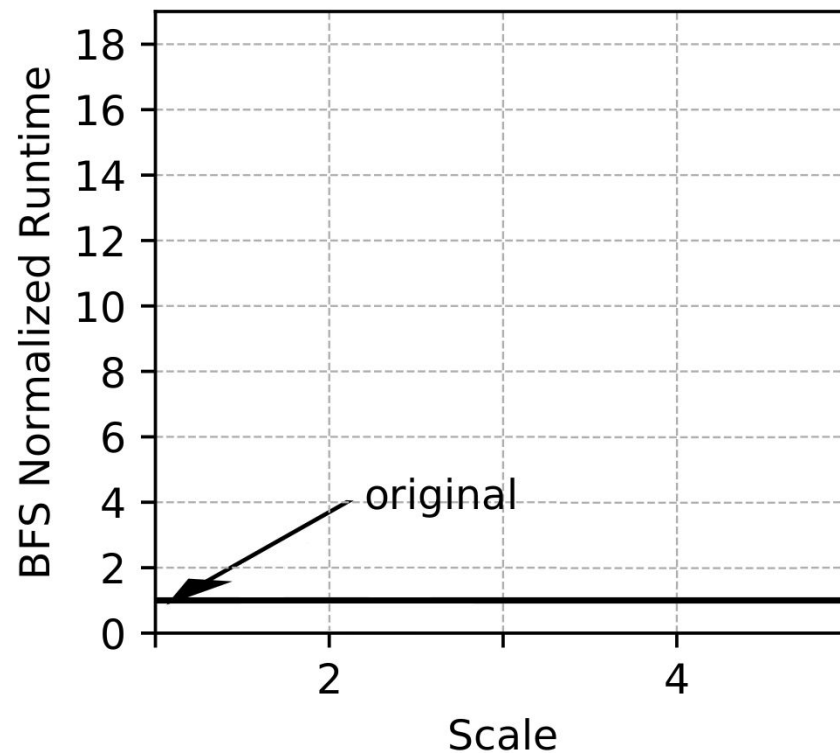
- Generate families of scaled-up graphs
 - Used for performance analysis
 - How do properties of the scaled-up graphs impact performance?
- Processing time of Breadth-First Search and PageRank
 - Using GraphMat and the Graphalytics benchmark suite
- Two datasets
 - Available in graph archives: "com-livejournal" and "12month1"

Case studies: Scale-up Configuration

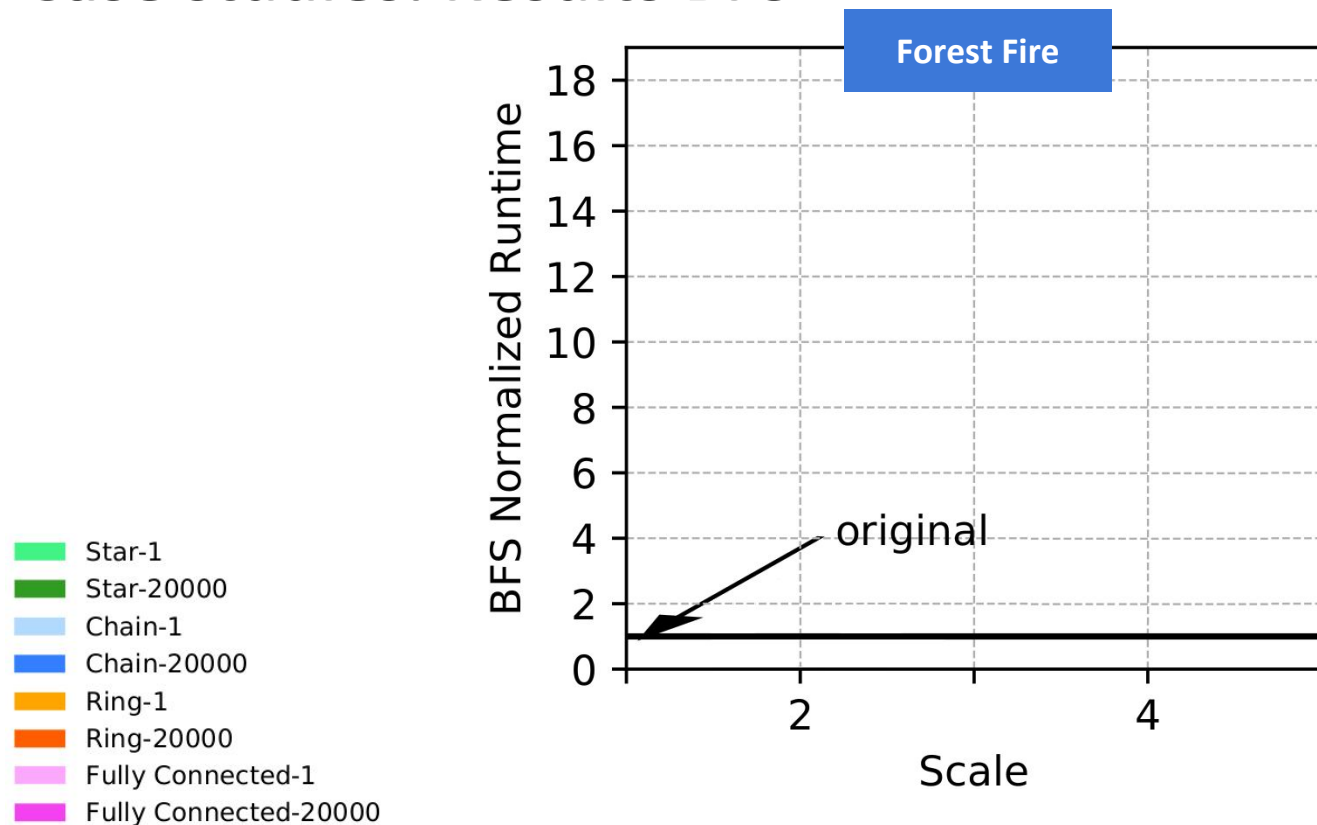
- Fixed parameters
 - 0.5 sample size
 - Random bridge interconnections
- Combining parameters
 - Scaling factor: 2, 4, 8*
 - Topologies: star, chain, ring, fully-connected
 - Number of interconnections: 1, 20,000
 - Different sampling algorithms and graph copies
- Sampling algorithms
 - Total-Induced Edge Sampling (TIES)
 - Forest-Fire

**only for graph copies*

Case studies: Results BFS

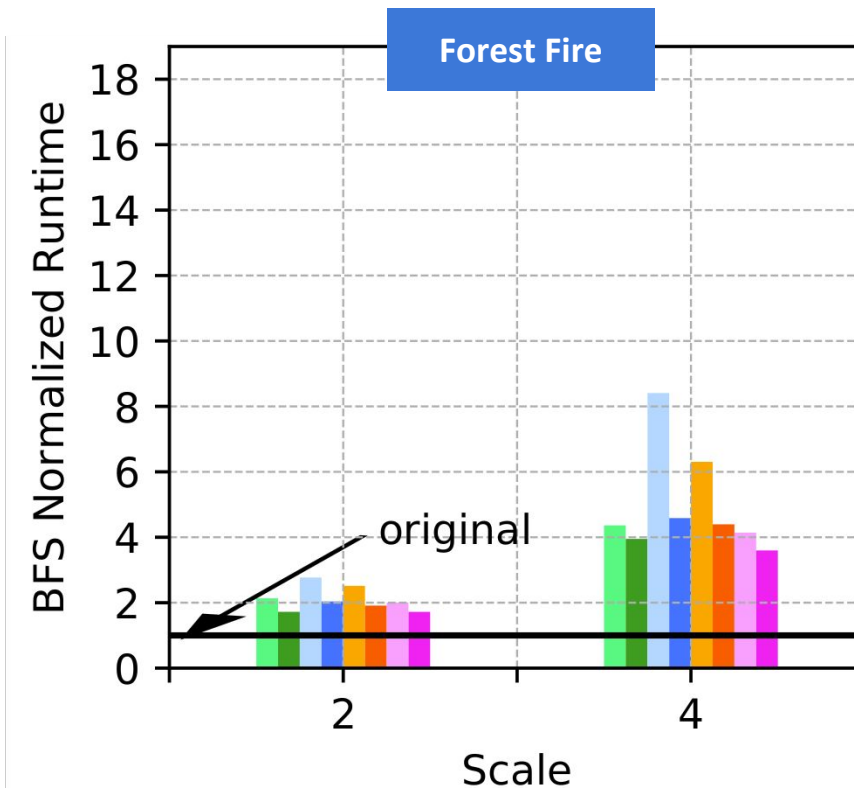


Case studies: Results BFS

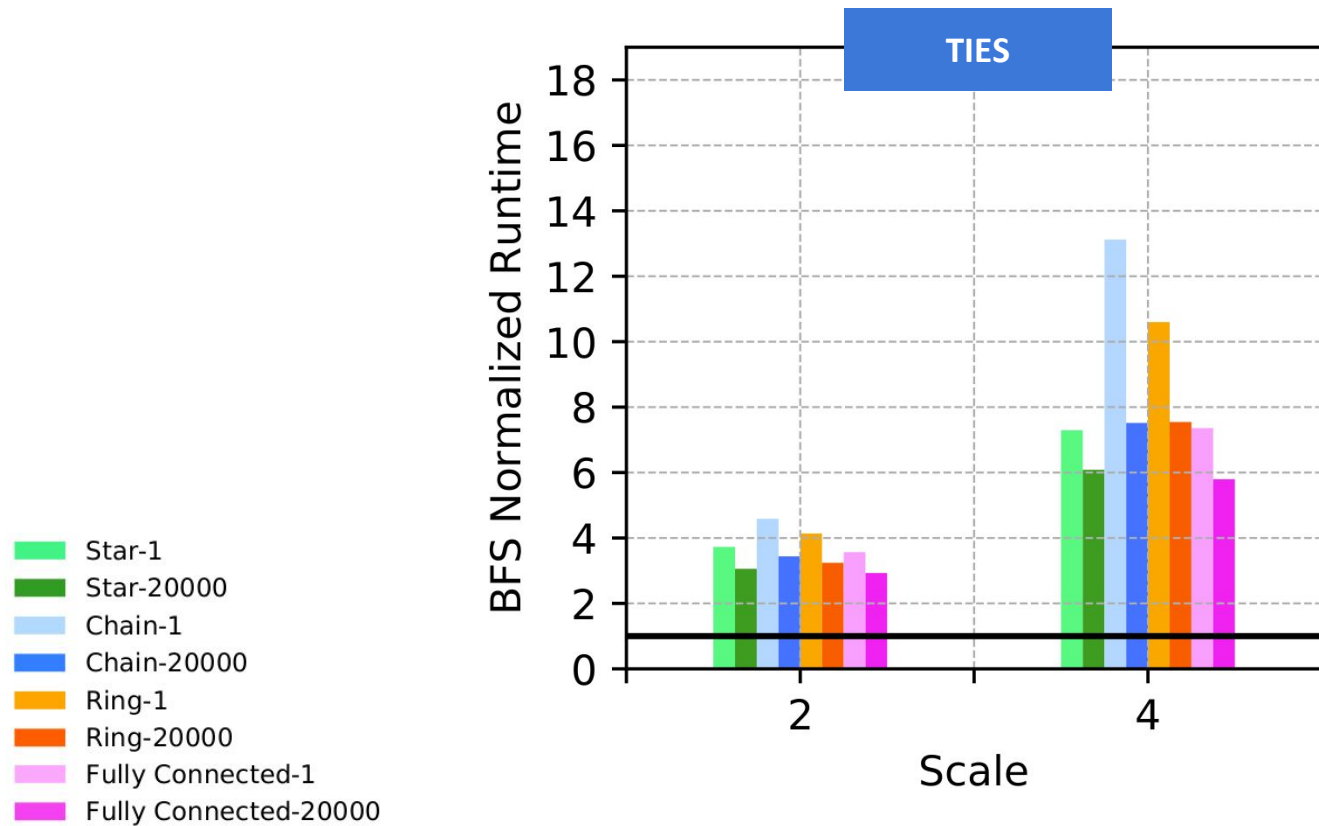


Case studies: Results BFS

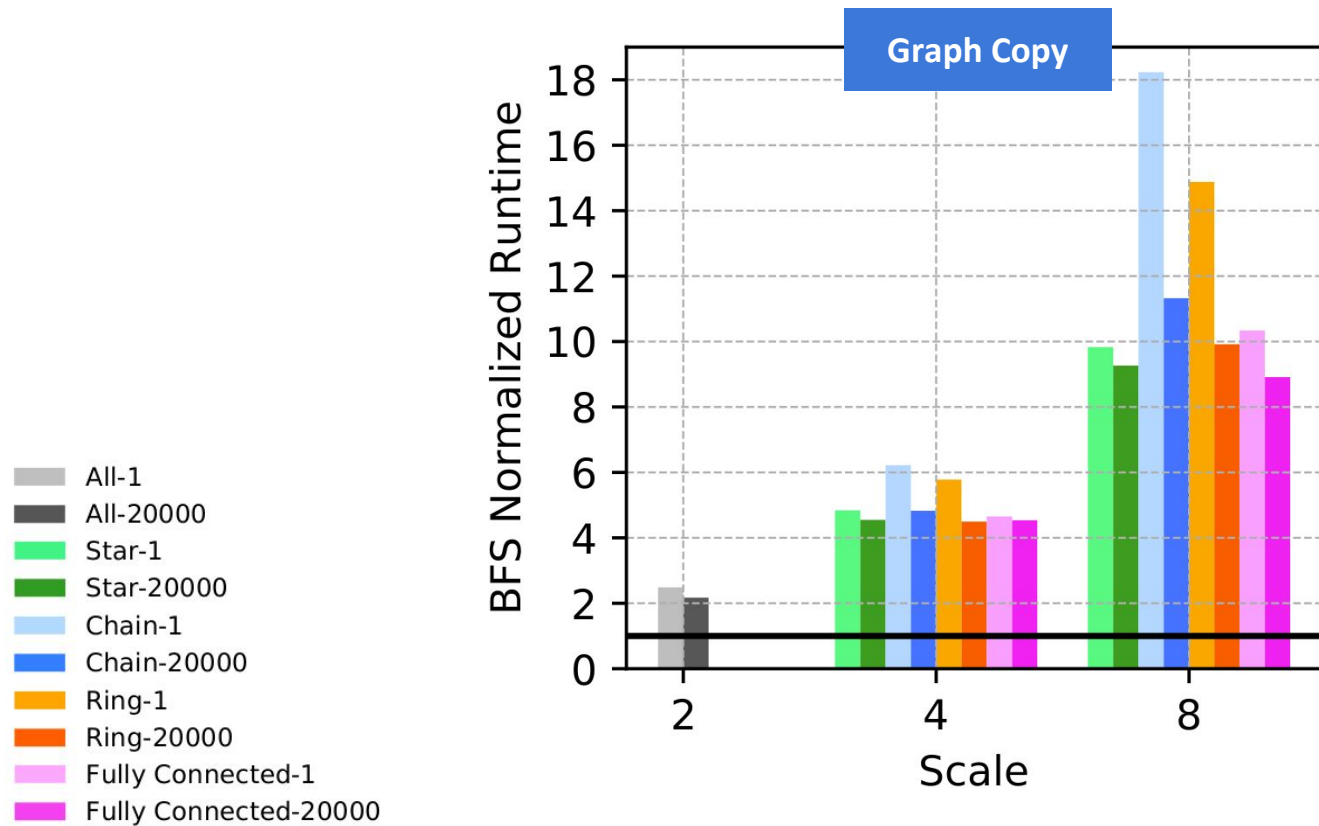
- Star-1
- Star-20000
- Chain-1
- Chain-20000
- Ring-1
- Ring-20000
- Fully Connected-1
- Fully Connected-20000



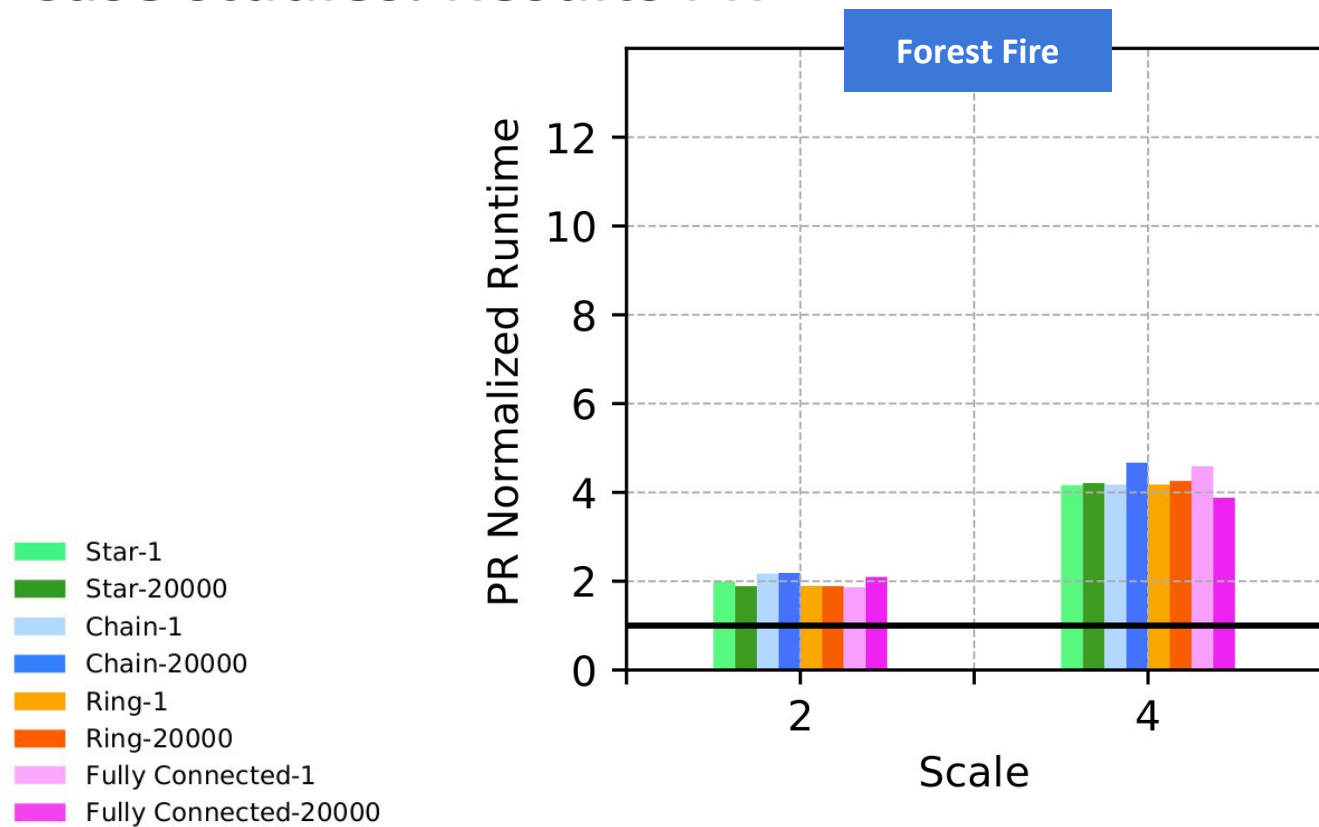
Case studies: Results BFS



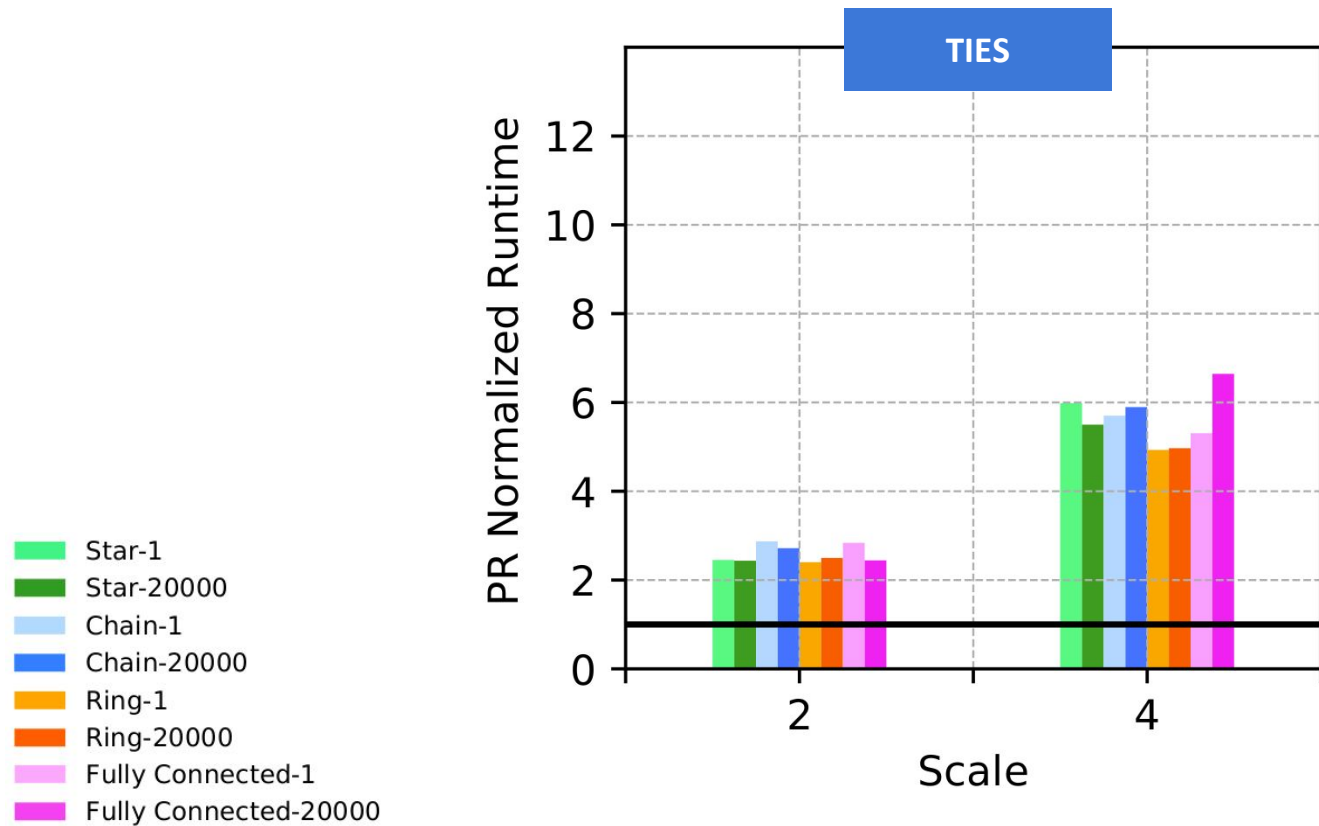
Case studies: Results BFS



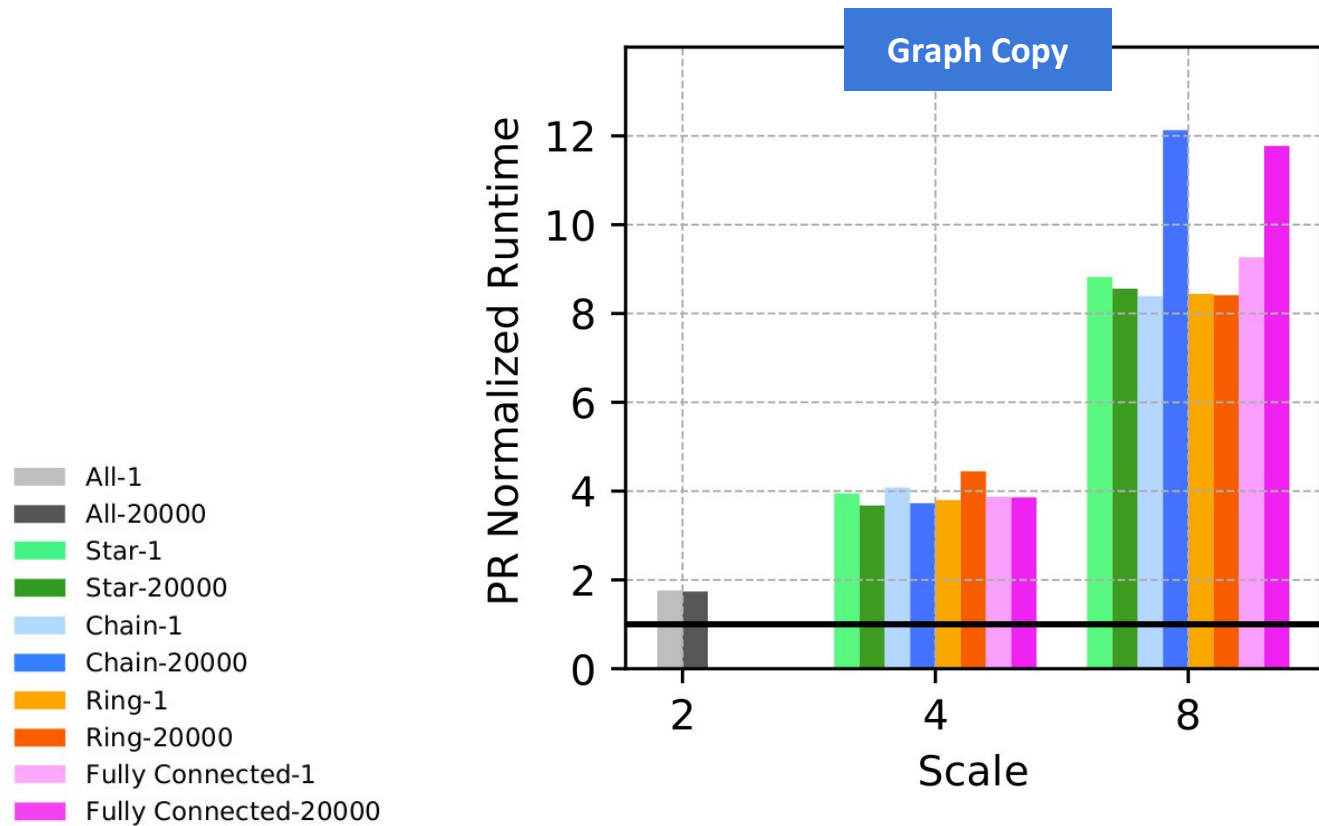
Case studies: Results PR



Case studies: Results PR



Case studies: Results PR



Tool Implementation

- Performance
 - Single node/thread
 - Parallel & Distributed Implementation
- Auto-tuner
- Tool publicly available

github.com/amusaafir/graph-scaling

Conclusion

- Sampling-based method for scaling graph datasets
 - Obtain families of similar graphs
 - Certain properties controlled by user requirements
- Validated our tool on a set of different graph datasets
- Diverse graph families used for understanding graph processing behavior