Proceedings of Machine Learning Research 222, 2023 ACML 2023

Supplementary Material for SANGEA: Scalable and Attributed Network Generation

Valentin Lemaire[1] VALENTIN.LEMAIRE@EURANOVA.EU
Youssef Achenchabe[2] YOUSSEF.ACHENCHABE@EURANOVA.EU
Lucas Ody[1] LUCAS.ODY@EURANOVA.EU
Houssem Eddine Souid[3] HOUSSEM.SOUID@EURANOVA.EU
Gianmarco Aversano[1] GIANMARCO.AVERSANO@EURANOVA.EU
Nicolas Posocco[1] NICOLAS.POSOCCO@EURANOVA.EU
Sabri Skhiri[1] SABRI.SKHIRI@EURANOVA.EU
[1] *Euranova, Rue Emile Francqui 4, 1435 Mont-Saint-Guibert, Belgium.*
[2] *Euranova, 146 Rue Paradis, 13006 Marseille, France.*
[3] *Euranova, 6 Les Berges du Lac 3, Tunis 2015, Tunisia.*

Editors: Berrin Yanıkoğlu and Wray Buntine

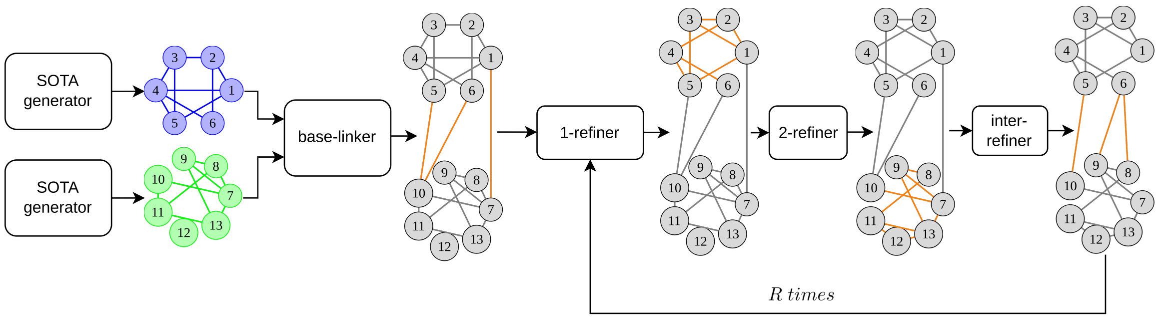

Figure 1: Full generation procedure of the SANGEA method

Parallelization of the Training Phases The training process is composed of independent steps that can be carried out in parallel: community partitioning is a prerequisite of all other phases and must be run as the first phase; phases 2 to 4 are independent and can be run in parallel: The base linker only uses information from the original graph and its partition to train the link predictor, and the same holds for the base refiner. Only phase 5 depends on the result of phase 4. Because in phase 5 the base refiner is copied to perform further training, phase 4 must be completed before phase 5 can start. However, we are fine-tuning $C+1$ models in phase 5. Each of those is independent of the others so they can all be trained in parallel.

1. Process time

1.1. SANGEA

Table 1 reports results on the execution time per epoch of the different parts of our *SANGEA* algorithm on four datasets of an increasing number of nodes and edges.

	Process time at training [s]				Process time at generation [s]	
	Init	Base linker	Base refiner	k-refiner	Base linking	Refinement step
Cora	3.25	1.12	1.92	0.81	2.14	0.74
IMDB	14.52	2.51	1.73	1.43	44.86	2.91
Amazon	134.13	12.06	37.97	3.59	28.48	6.02
Flickr	1286.15	19.24	219.87	5.46	954.35	224.75

Table 1: *SANGEA*'s execution time on increasing graph size. Training times reported in *sec/epoch* except for Init time which happens only once.

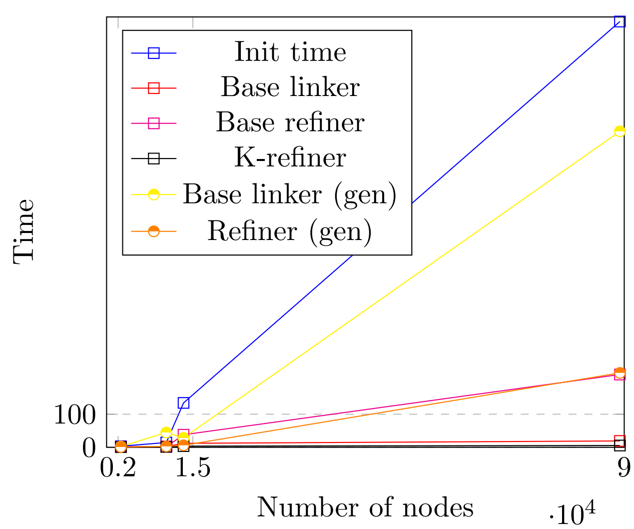

Figure 2: SANGEA's execution times per number of nodes

1.2. State-of-the-art approaches

	Process time [s]		
	Init time	Time per epoch	Generation time
NVDiff	2.52	0.60	2.56
GDSS	0.01	1.73	62.59
GVAEmm	2.70	0.47	0.03
GraphGen	5.85	0.92	2.39
BiGG	1.97	0.61	0.54

Table 2: Execution time on a subgraph from Cora (175 nodes 281 edges)

	Process time [s]		
	Init time	Time per epoch	Generation time
NVDiff	OOM	OOM	OOM
GDSS	OOM	OOM	OOM
GVAEmm	OOM	OOM	OOM
GraphGen	10186	2.45	49.22
BiGG	2.23	1.09	91.19

Table 3: Execution time of state-of-the-art approaches on Cora

	Process time [s]		
	Init time	Time per epoch	Generation time
NVDiff	OOM	OOM	OOM
GDSS	OOM	OOM	OOM
GVAEmm	OOM	OOM	OOM
GraphGen	OOM	OOM	OOM
BiGG	OOM	OOM	OOM

Table 4: Execution time on Amazon Computers

	Process time [s]		
	Init time	Time per epoch	Generation time
NVDiff	OOM	OOM	OOM
GDSS	OOM	OOM	OOM
GVAEmm	OOM	OOM	OOM
GraphGen	32538.84	5.81	161.17
BiGG	2.06	1.75	173.58

Table 5: Execution time on IMDB

	Process time [s]		
	Init time	Time per epoch	Generation time
NVDiff	OOM	OOM	OOM
GDSS	OOM	OOM	OOM
GVAEmm	OOM	OOM	OOM
GraphGen	Timeout	Timeout	Timeout
BiGG	OOM	OOM	OOM

Table 6: Execution time on Flickr

2. Generation quality

	Competitors			Ours			
	GraphGen	BiGG	Other SOTAs	SANGEA (GDSS)	SANGEA (GVAEmm)	SANGEA (GraphGen)	SANGEA (NVDiff)
MMD (gcn tr.)	-	-		**0**	-	-	**0**
MMD (gcn unt.)	-	-		**0**	-	-	0.15
WS spectral	**1.45e-4**	26e-4		40e-4	36.2e-4	3.28e-4	13.57e-4
WS deg. hist.	**3.52e-4**	28.3e-4		126e-4	11.4e-4	5.97e-4	18.08e-4
WS deg. centr.	3.35e-6	20.4e-6	OOM	**1.33e-6**	133e-6	9.21e-6	132.19e-6
WS clos. cent.	**4.56e-6**	141e-6		25.6e-6	95.7e-6	29.38e-6	178.55e-6
WS eigenv. cent.	**21.03e-6**	137e-6		230e-6	70e-6	166.38e-6	211.62e-6
WS clust. coeff.	**0.95e-4**	3.64e-4		17.2e-4	5.2e-4	1.82e-4	2.91e-4
AUROC (LP)	0.74	0.77		**0.90**	0.74	0.73	**0.90**

Table 7: Structural and attribute similarity results on CiteSeer

	Competitors			Ours			
	GraphGen	BiGG	Other SOTAs	SANGEA (GDSS)	SANGEA (GVAEmm)	SANGEA (GraphGen)	SANGEA (NVDiff)
MMD (gcn tr.)	-	-		8.97e-1	-	-	**0.96e-1**
MMD (gcn unt.)	-	-		9.3e-1	-	-	**4.9e-1**
WS spectral	**1.61e-4**	18.5e-4		46e-4	41.5e-4	4.07e-4	5.22e-4
WS deg. hist.	**2.06e-4**	14.3e-4		75e-4	10.4e-4	5.65e-4	14.6e-4
WS deg. centr.	**2.92e-6**	5.36e-6	OOM	35e-6	31.7e-6	13.33e-6	66.64e-6
WS clos. cent.	**4.74e-6**	48e-6		21e-6	47.5e-6	28.13e-6	91.02e-6
WS eigenv. cent.	**3.1e-5**	22.7e-5		18.3e-5	17e-5	24.97e-5	38.77e-5
WS clust. coeff.	**0.76e-4**	9.25e-4		14.1e-4	8.1e-4	2.4e-4	4.64e-4
AUROC (LP)	0.78	0.79		**0.93**	0.78	0.76	0.74

Table 8: Structural and attribute similarity results on Cora. SANGEA(GDSS) stands for *SANGEA* using GDSS as a community generator.

	Sa(GDSS)		Sa(NVDiff)	
	w.o. ref.	*w. ref.*	*w.o. ref.*	*w. ref.*
MMD (gcn tr.)	23.4e-1	**8.97e-1**	13.7e-2	**9.6e-2**
MMD (gcn untr.)	21.7e-1	**9.27e-1**	17.9e-1	**4.9e-1**
WS spectral	**46.1e-4**	**46.1e-4**	11.e-4	**5.22e-4**
WS deg. hist.	7.64e-3	**7.48e-3**	17.6e-3	**1.46e-3**
WS deg. cent.	5.49e-5	**3.50e-5**	9.25e-5	**6.66e-5**
WS clos. cent.	121e-6	**21e-6**	10.6e-5	**9.10e-5**
WS eigenv. cent.	15.5e-4	**1.83e-4**	13.1e-4	**3.88e-4**
WS clust. coeff.	1.67e-3	**1.41e-3**	35.8e-4	**4.64e-4**
AUROC (LP)	0.54	**0.93**	0.71	**0.74**

Table 9: Ablation study of the refinement process for the Cora dataset

	Sa(GDSS)		Sa(NVDiff)	
	w.o. ref.	*w. ref.*	*w.o. ref.*	*w. ref.*
MMD (gcn tr.)	2.72e-3	**0.00**	0.08	**0.00**
MMD (gcn untr.)	8.82e-4	**0.00**	0.37	**0.15**
WS spectral	**2.20e-3**	4.00e-3	4.87e-3	**1.36e-3**
WS deg. hist.	3.09e-2	**1.26e-2**	11.5e-3	**1.81e-3**
WS deg. cent.	192e-6	**1.33e-6**	1.34e-4	**1.32e-4**
WS clos. cent.	8.83e-5	**2.56e-5**	**1.00e-5**	17.9e-5
WS eigenv. cent.	9.30e-4	**2.30e-4**	**1.40e-4**	2.12e-4
WS clust. coeff.	3.29e-3	**1.72e-3**	12.4e-4	**2.91e-4**
AUROC (LP)	0.63	**0.90**	0.84	**0.90**

Table 10: Ablation study of the refinement process for the CiteSeer dataset

	Sa(GDSS)		Sa(NVDiff)	
	w.o. ref.	*w. ref.*	*w.o. ref.*	*w. ref.*
MMD (gcn tr.)	0.438	**0.305**	0.489	**0.379**
MMD (gcn untr.)	0.0631	**0.00210**	0.251	**0.0713**
WS spectral	43.2e-3	**5.66e-3**	9.78e-3	**3.50e-3**
WS deg. hist.	41.4e-4	**8.58e-4**	3.00e-3	**1.82e-3**
WS deg. cent.	**1.05e-5**	**1.05e-5**	5.40e-5	**5.40e-5**
WS clos. cent.	26.2e-6	**2.84e-6**	6.17e-5	**1.00e-5**
WS eigenv. cent.	31.2e-5	**4.12e-5**	4.24e-5	**3.59e-5**
WS clust. coeff.	45.6e-5	**1.51e-5**	17.3e-5	**2.07e-5**
AUROC (LP)	0.71	**0.76**	0.73	**0.74**

Table 11: Ablation study of the refinement process for the IMDB dataset

	Sa(GDSS)		Sa(NVDiff)	
	w.o. ref.	*w. ref.*	*w.o. ref.*	*w. ref.*
MMD (gcn tr.)	0.592	**0.419**	0.760	**0.375**
MMD (gcn untr.)	0.123	**0.0426**	0.195	**0.0329**
WS spectral	0.00155763	**1.68e-3**	1.75e-3	**1.74e-3**
WS deg. hist.	12.9e-5	**8.56e-5**	1.09e-4	**1.10e-4**
WS deg. cent.	**2.04e-6**	**2.04e-6**	10.2e-6	**9.88e-6**
WS clos. cent.	4.73e-6	**2.76e-6**	1.33e-5	**1.28e-5**
WS eigenv. cent.	**1.70e-5**	1.78e-5	3.73e-5	**3.33e-5**
WS clust. coeff.	8.14e-5	**4.24e-5**	5.97e-5	**5.78e-5**
AUROC (LP)	0.758	**0.814**	0.787	**0.833**

Table 12: Ablation study of the refinement process for the Amazon dataset

Hidden dimension of GNN	16	64	128	256
Layers in GNN	2	3	4	5
Embedding size of GNN	16	32	64	128
GNN aggregation type	sum	mean		
Hidden dimension of MLP	8	16	64	128
Layers in mlp	2	3		

Table 13: Parameters and model size ranges; those ranges formed the parameter space searched during hyper-parameter optimisation.