Online Appendix: Further Results

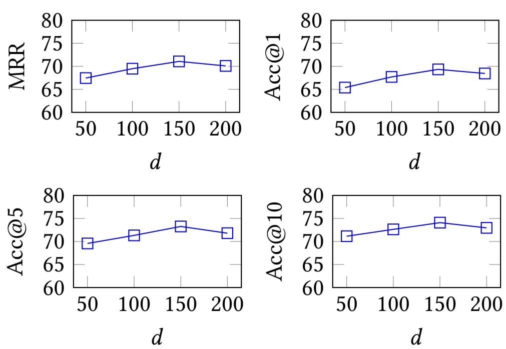
A Ablation Study

Table II shows the results of CAPER when the value of M is 2, 3, 4, and 5, respectively. The results consistently demonstrate that not only each design choice in CAPER is effective, but also CAPER integrating all of them is the most effective.

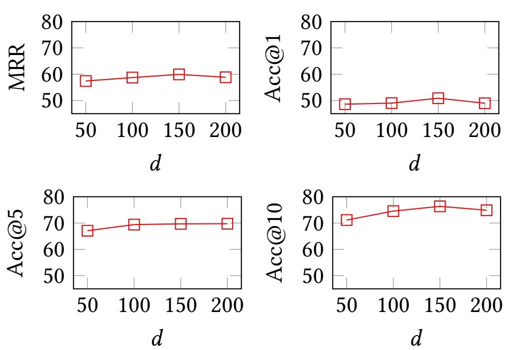
B Granularity Analysis

We conducted an additional experiment where we updated the model and predicted the next company and position every six months instead of every year. In this scenario, the average number of careers per user is 1.52. For comparison, we selected TACTP, the best-performing method among the competitors. As shown in Table I, CAPER significantly and consistently outperforms TACTP in all metrics, which coincides with the trend at the 1-year time granularity. These findings indicate that CAPER is more suitable for adjusting to rapid changes in career movement.

C Parameter Analysis



(a) Company Prediction



(b) Position Prediction

Figure I: The effects of d on the accuracies of CAPER when the value of M is 1. The accuracies of CAPER do not fluctuate a lot as d varies, but usually perform the best with d being around 150.

To understand the learning stability of CAPER, we carefully analyze the change of accuracy according to different values for the embedding dimensionality (*i.e.*, d). Figures I-(a) and -(b) show the accuracy changes of CAPER in our dataset with varying d for the company and position prediction, respectively. From Figure I, we see that the accuracies of CAPER do not fluctuate a lot as d varies, but usually perform the best with d being around 150.

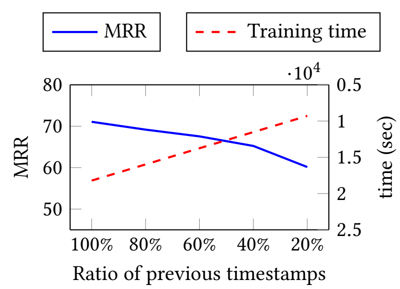
D Efficiency Analysis

Table III: Training and inference times of CAPER and TACTP

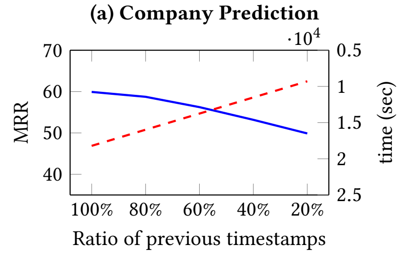
	Training and Inference Time (sec)	Avg. Inference Time per User (msec)
CAPER	18,213.48	0.89
TACTP	49,931.93	1.74

We conducted experiments that measure the training and inference times to show the efficiency of CAPER. For comparison, we selected TACTP, the best-performing method among the competitors. All the experiments were conducted in Ubuntu 20.04 LTS running on Nvidia v100×1ea and 7 vCPUs with 70GB of RAM. Table III shows that CAPER surpasses TACTP in terms of time efficiency during both the training and inference phases. Considering the substantial improvement of CAPER in accuracy over TACTP (refer to Table 3), we can confidently assert that CAPER holds advantages in both effectiveness and efficiency over TACTP.

E Computational Overhead Analysis



(a) Company Prediction



(b) Position Prediction

Figure II: The effects of adjusting the ratio of previous timestamps (*i.e.*, l).

We conducted experiments to demonstrate how reducing the value of l used to train each temporal embedding can mitigate the computational overhead of CAPER without compromising accuracy. All experiments were conducted on Ubuntu 20.04 LTS, running on

Table I: Accuracies of CAPER and TACTP when the time granularity is six months

(a) Company Prediction

Methods	MRR					Acc@1					Acc@5					Acc@10				
	M=1	M=2	M=3	M=4	M=5	M=1	M=2	M=3	M=4	M=5	M=1	M=2	M=3	M=4	M=5	M=1	M=2	M=3	M=4	M=5
TACTP	58.42	53.84	49.35	48.49	47.94	57.43	52.11	48.46	46.92	46.31	59.76	54.49	50.79	49.59	48.31	60.74	56.06	51.73	**51.41**	48.50
CAPER	**70.88**	**57.52**	**51.01**	**49.28**	**49.30**	**68.63**	**56.13**	**49.25**	**48.03**	**48.56**	**72.90**	**58.54**	**51.33**	**50.42**	**49.23**	**73.55**	**59.53**	**52.62**	51.15	**50.42**

(b) Position Prediction

Methods	MRR					Acc@1					Acc@5					Acc@10				
	M=1	M=2	M=3	M=4	M=5	M=1	M=2	M=3	M=4	M=5	M=1	M=2	M=3	M=4	M=5	M=1	M=2	M=3	M=4	M=5
TACTP	36.49	35.25	34.78	33.76	32.53	27.23	27.10	26.28	25.13	24.05	43.13	42.16	40.52	39.49	39.21	52.44	51.18	50.04	50.02	48.06
CAPER	**59.51**	**48.10**	**42.84**	**42.75**	**43.89**	**50.25**	**40.05**	**35.10**	**35.06**	**35.13**	**69.11**	**56.14**	**49.81**	**50.03**	**50.34**	**76.01**	**62.12**	**56.26**	**56.27**	**56.04**

an Nvidia v100 GPU and 7 vCPUs with 70GB of RAM. Figure II shows the trade-off between the accuracy and training time of CAPER by changing the ratio of previous timestamps used. The x-axis indicates the ratio of previous timestamps actually used for training, and the y-axis does the training time (right) and accuracy (left). The results show that reducing the ratio of timestamps to 80% leads to a relatively small decrease in MRR, while the training time is reduced by approximately 12%. This indicates that it is possible to improve efficiency without a substantial loss in model accuracy.

F Scalability Analysis

To test the scalability of CAPER, we measured the execution time of CAPER by increasing the sampling ratio, *i.e.*, 20%, 40%, $\cdots$, 100%, of training users. All the experiments were conducted in Ubuntu 20.04 LTS running on Nvidia v100×1ea and 7 vCPUs with 70GB RAM. The result shows that the execution time was around 9, 17, 26, 35, and 43 minutes when we have about 20% (66K), 40% (133K), 60% (199K), 80% (266K), and 100% (333K) users. This result indicates that CAPER has linear scalability as the number of users increases.

Table II: Ablation studies for CAPER when the values of M are 2, 3, 4, and 5

(a) Company Prediction (M=2)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	53.01	51.12	5.94	57.83	48.03	57.36	58.14	50.35	53.78	57.74	53.64	58.14
Acc@1	49.82	44.12	2.97	56.82	40.36	56.49	57.09	48.64	52.59	56.68	50.71	57.09
Acc@5	56.45	56.07	7.13	58.70	55.43	58.09	59.05	52.11	54.97	58.11	56.67	59.05
Acc@10	58.68	56.77	10.09	59.54	51.09	58.84	59.93	53.28	55.77	59.50	58.86	59.93

(b) Position Prediction (M=2)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	37.67	42.03	2.17	12.38	34.35	12.73	48.66	29.29	41.26	46.89	42.29	48.66
Acc@1	22.72	27.17	0.05	6.16	22.70	7.16	40.62	15.92	37.56	39.11	32.03	40.62
Acc@5	56.58	54.97	2.17	16.77	52.11	16.1	56.56	26.06	44.28	50.01	52.97	56.56
Acc@10	63.01	58.12	7.26	20.94	53.28	20.12	62.47	36.64	47.70	60.92	60.29	62.47

(c) Company Prediction (M=3)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	45.40	42.71	5.39	50.10	42.78	49.62	51.35	43.68	46.59	50.02	45.09	51.35
Acc@1	42.44	39.45	2.60	49.25	35.50	48.88	49.44	42.26	45.61	49.12	42.14	49.44
Acc@5	48.46	48.44	6.41	50.72	49.39	50.18	51.75	45.09	47.48	50.22	48.04	51.75
Acc@10	50.66	49.02	9.19	51.52	44.92	50.86	52.87	46.14	48.21	51.47	50.29	52.87

(d) Position Prediction (M=3)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	33.23	37.70	1.97	10.51	30.00	10.33	43.28	25.64	36.57	42.17	37.25	43.28
Acc@1	19.60	23.58	0.05	5.99	20.28	6.43	35.74	14.15	33.19	34.73	27.75	35.74
Acc@5	49.80	47.48	1.84	13.93	45.09	13.89	50.27	23.54	39.14	44.53	46.58	50.27
Acc@10	56.58	50.10	6.76	19.36	46.14	17.76	56.61	31.94	42.45	55.76	54.39	56.61

(e) Company Prediction (M=4)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	43.95	42.36	5.44	48.85	43.19	48.41	49.70	42.63	45.43	48.77	41.50	49.70
Acc@1	41.00	39.92	2.70	48.04	36.40	47.70	48.24	41.24	44.48	47.91	38.07	48.24
Acc@5	46.97	47.24	6.43	49.45	49.51	48.91	50.75	43.98	46.29	48.91	44.95	50.75
Acc@10	49.19	47.83	9.16	50.22	44.59	49.58	51.57	44.98	47.01	50.18	47.60	51.57

(f) Position Prediction (M=4)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	33.31	38.19	1.92	10.12	29.25	9.69	43.15	25.02	36.28	42.68	36.31	43.15
Acc@1	19.67	22.99	0.05	5.86	20.12	5.86	35.62	13.93	32.94	35.41	26.74	35.62
Acc@5	49.85	46.29	1.71	13.22	43.98	13.47	50.42	23.19	38.82	44.87	46.58	50.42
Acc@10	56.67	48.87	6.75	19.14	44.98	17.27	56.53	31.14	42.10	56.03	53.91	56.53

(g) Company Prediction (M=5)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	44.10	42.65	5.46	49.18	43.56	48.71	49.43	42.84	45.65	49.09	34.96	49.43
Acc@1	41.13	40.25	2.66	48.38	36.50	48.01	48.58	41.45	44.70	48.23	30.47	48.58
Acc@5	47.14	47.47	6.50	49.79	49.86	49.22	50.09	44.21	46.52	49.23	39.66	50.09
Acc@10	49.36	48.09	9.30	50.52	44.90	49.85	50.88	45.22	47.25	50.52	43.15	50.88

(h) Position Prediction (M=5)

Metrics	(i) EQ2		(ii) EQ3	(iii) EQ4-1				(iv) EQ4-2				CAPER
	KG	HetG	w/o GCN	All	w/o $\tilde{u}_x$	w/o $\tilde{c}_x$	w/o $\tilde{p}_x$	w/o GRNN	RNN	GRU	Transformer	
MRR	33.53	38.54	1.89	10.08	29.36	9.67	44.03	25.25	36.53	43.08	33.19	44.03
Acc@1	19.83	23.15	0.05	5.83	20.26	5.79	36.63	14.03	33.20	35.85	23.15	36.63
Acc@5	50.20	46.52	1.64	13.25	44.21	13.47	50.77	23.31	39.07	45.28	43.07	50.77
Acc@10	56.99	49.12	6.87	19.08	45.22	17.26	56.58	31.40	42.34	56.35	52.22	56.58