



TIMERS: Error-Bounded SVD Restart on Dynamic Networks

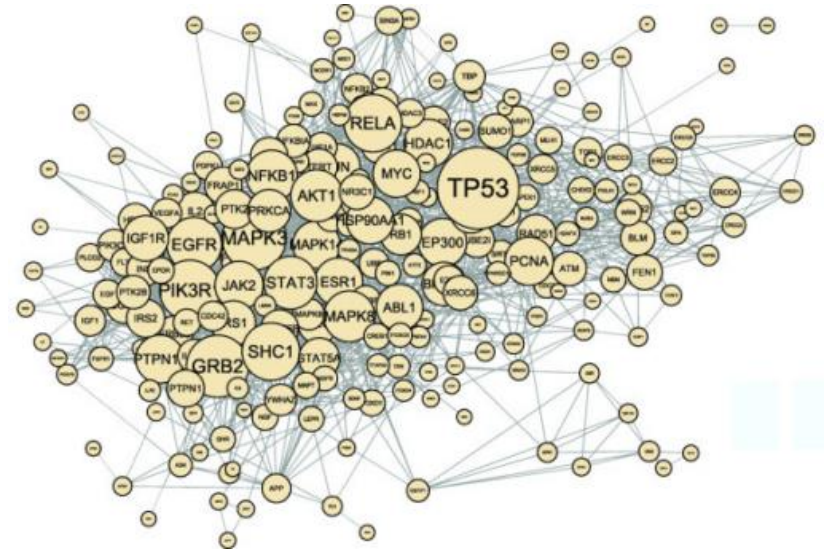
Ziwei Zhang Peng Cui Jian Pei Xiao Wang Wenwu Zhu
Tsinghua U Tsinghua U Simon Fraser U Tsinghua U Tsinghua U



Networks



Social Network



Biology Network



Traffic Network

Matrix Representation

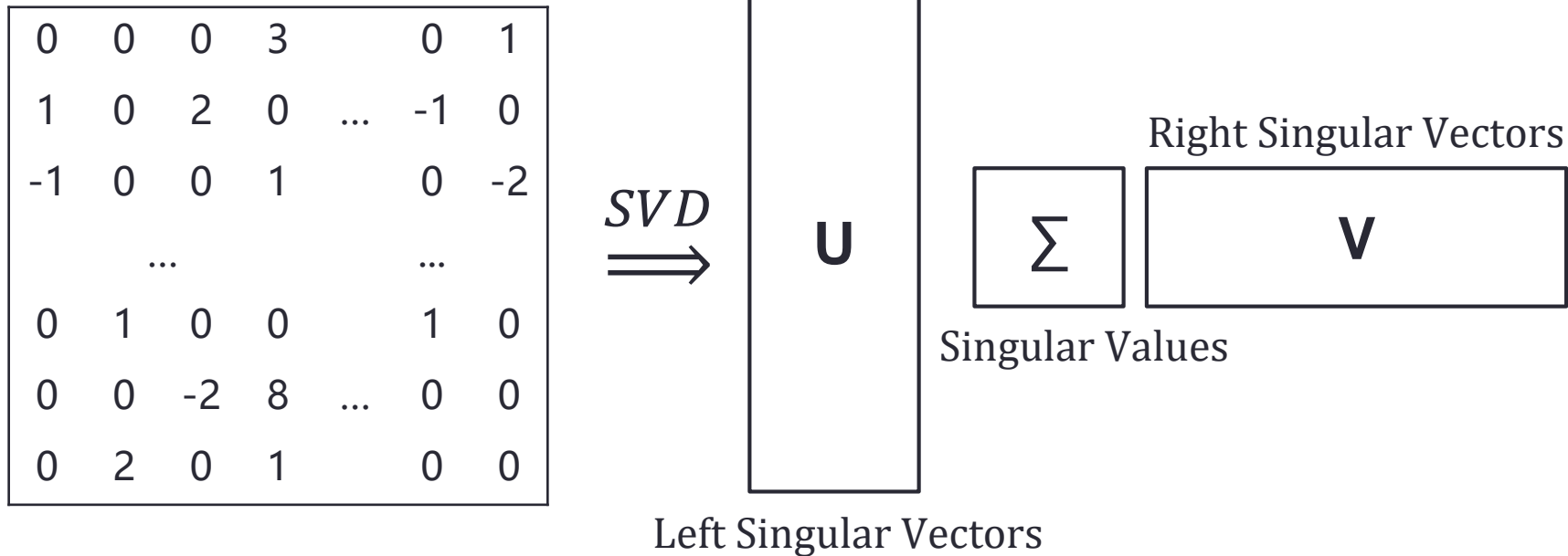


0	0	0	3		0	1
1	0	2	0	...	-1	0
-1	0	0	1		0	-2
				...		
0	1	0	0		1	0
0	0	-2	8	...	0	0
0	2	0	1		0	0

Adjacency Matrix

- ❑ Other variants: high-order similarity matrix, transition matrix, Laplacian matrix, etc.

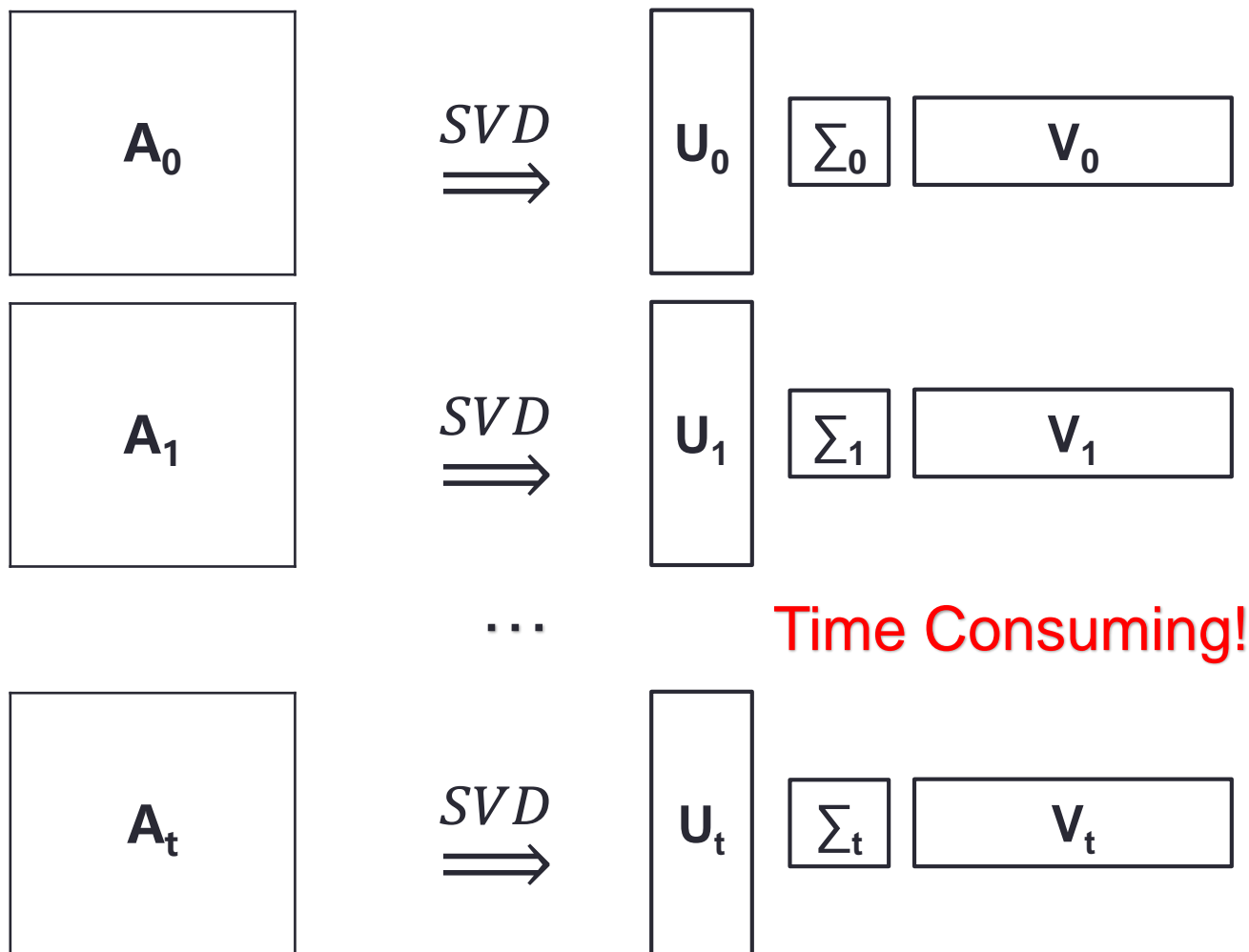
Network Analysis: SVD



- ❑ Applications: network embedding, link prediction, characterizing network parameters, etc.
- ❑ Computationally expensive procedure

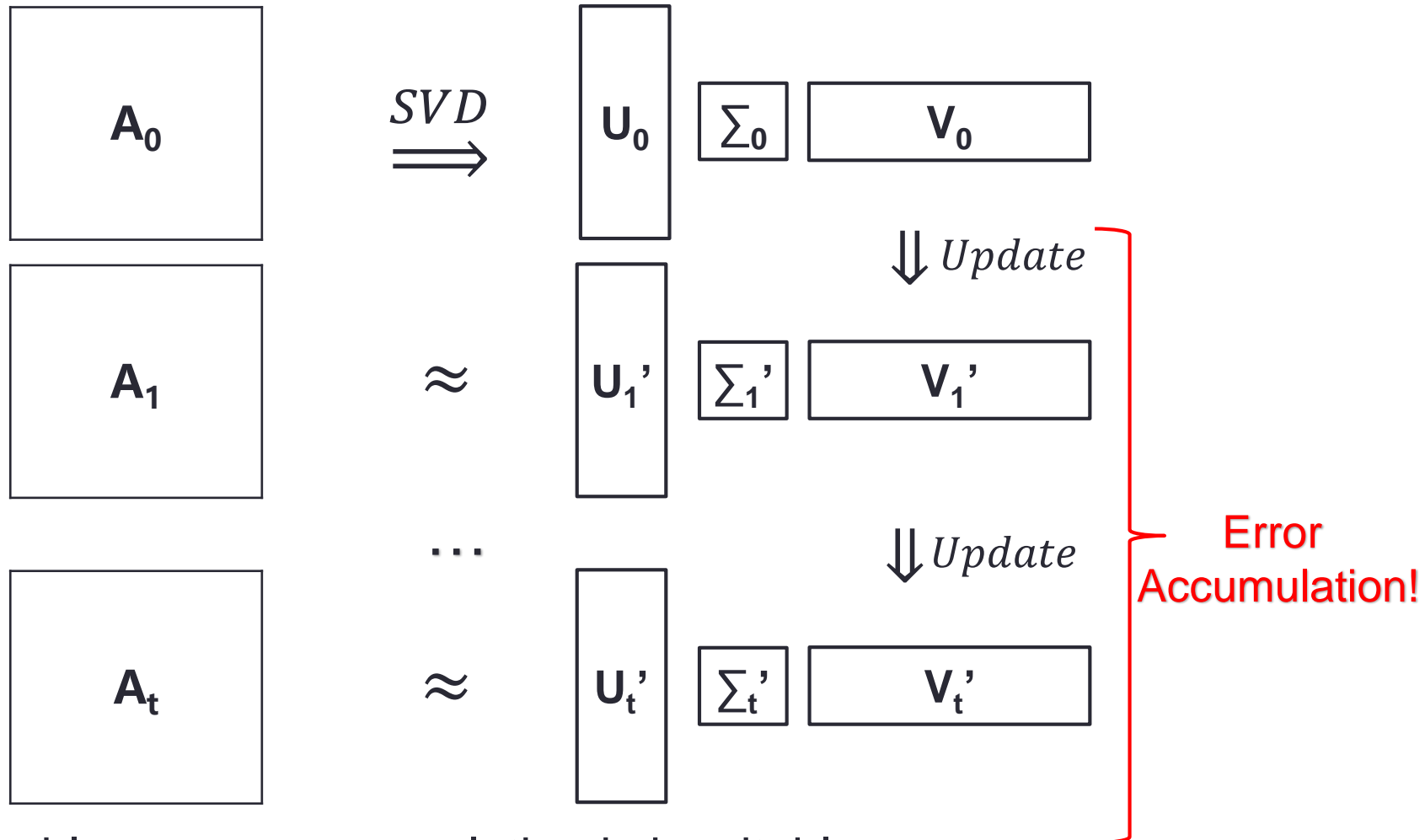
Dynamic Networks

- Networks are dynamic in nature
 - SVD results need to be updated accordingly



Incremental SVD

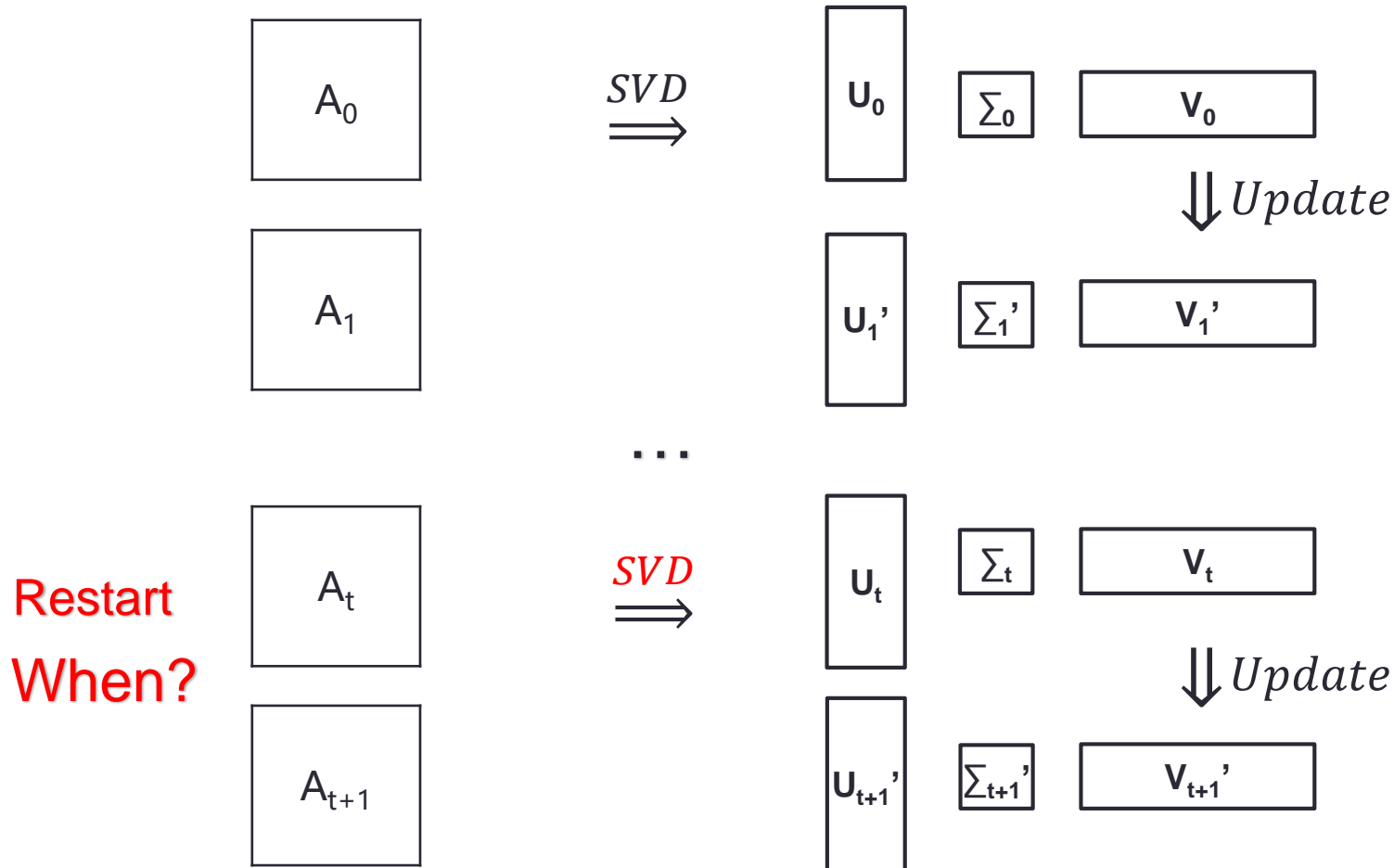
- Incremental SVD is proposed by inducing approximation



- Problem: error accumulation is inevitable

SVD Restarts

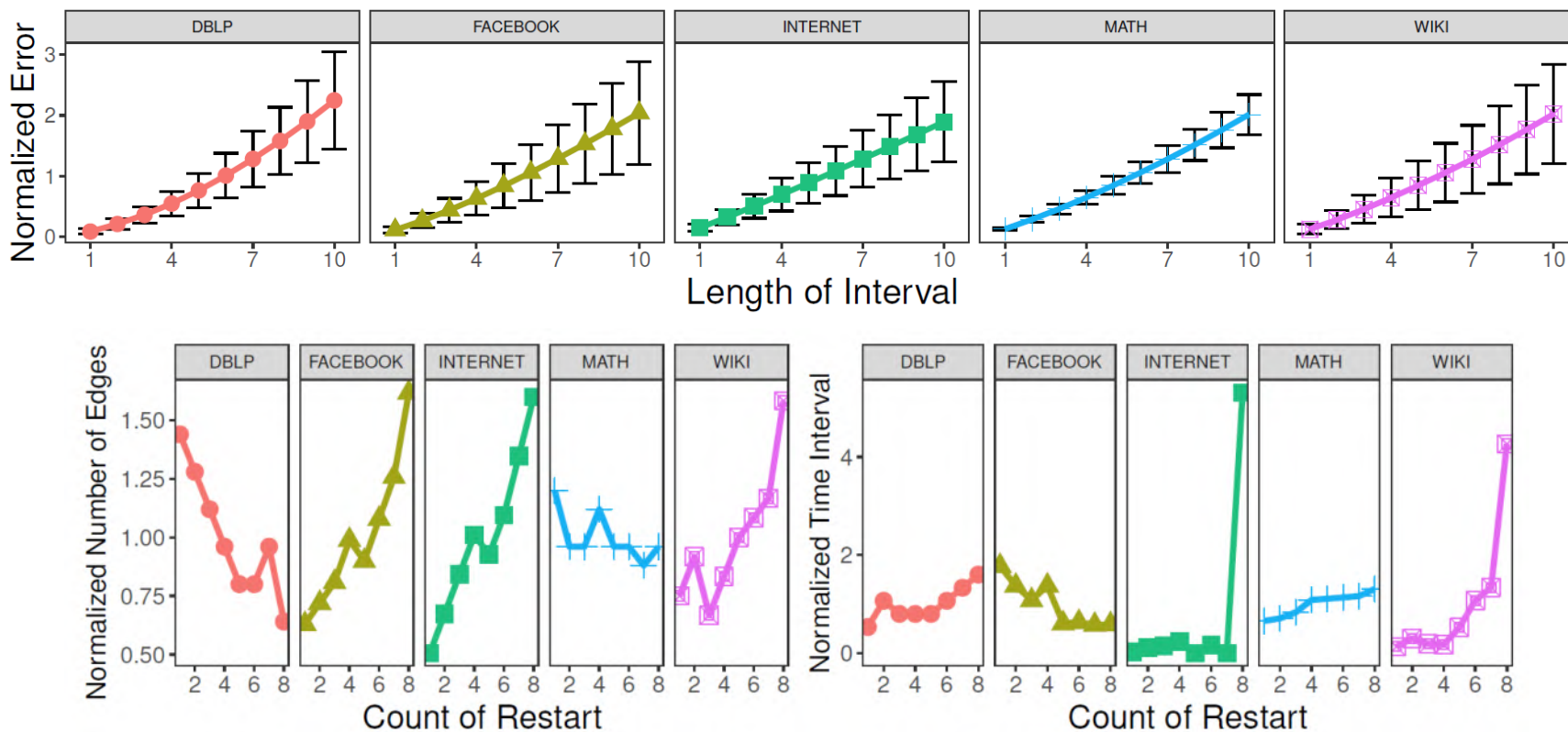
- Solution: restart SVD occasionally



- What are the appropriate time points?

Naïve Solution

- Naïve solution: fixed time interval or fixed number of changes
- Difficulty: error accumulation is not uniform
 - Validated by preliminary studies



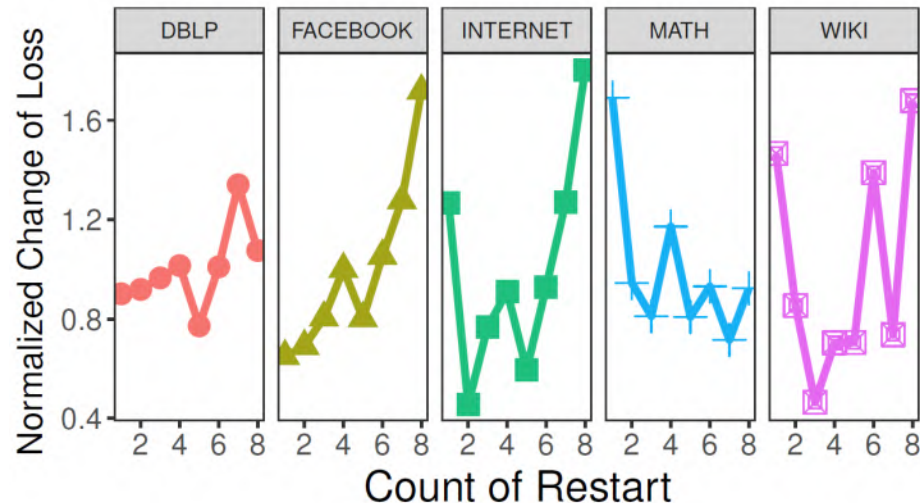
Existing Method

- Existing method: monitor loss (Chen and Candan, KDD 2014)
- Loss in SVD:

$$\mathcal{J} = \|S - U\Sigma V^T\|_F^2$$

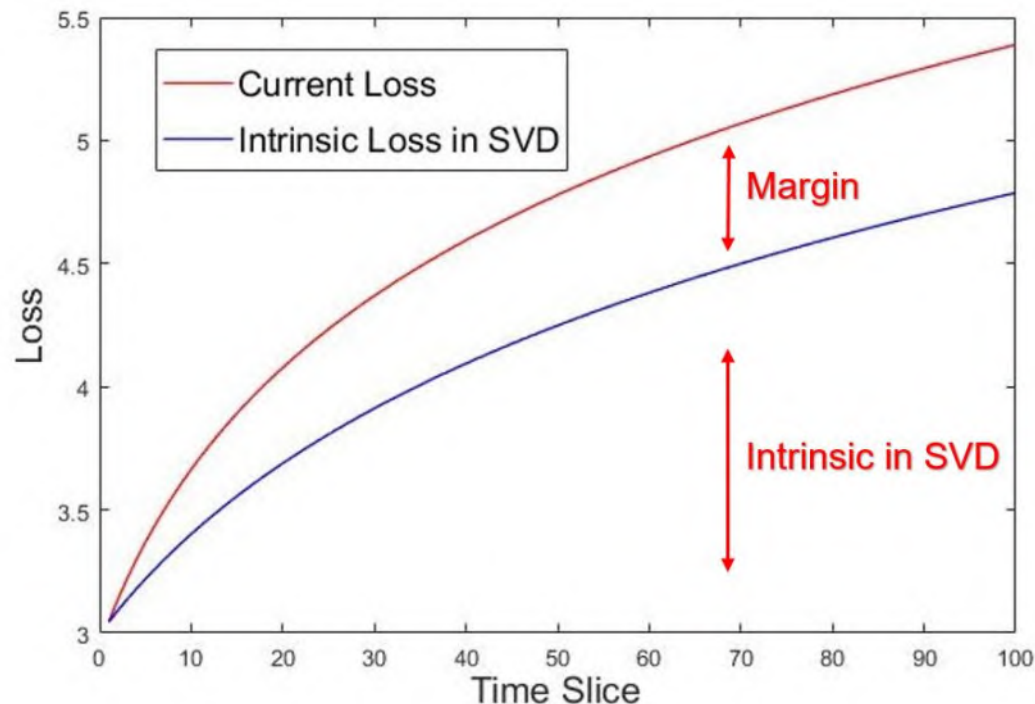
S : target matrix, $[U, \Sigma, V]$: results of SVD

- Problem: loss includes approximation error and intrinsic loss in SVD
- Preliminary study results:



Framework: Monitor Margin

- Observation: the **margin** between the current loss and intrinsic loss in SVD is the actual accumulated error
- Current loss: $\mathcal{J} = \|S - U\Sigma V^T\|_F^2$
- Intrinsic loss: $\mathcal{L}(S, k) = \min_{U^*, \Sigma^*, V^*} \|S - U^* \Sigma^* V^{*T}\|_F^2, k: \text{dimensionality}$



Framework: Monitor Margin

- Framework: monitor the maximum margin
- Formulation: constrained optimization

$$\min_{c_1, \dots, c_T} \sum_{t=1}^T c_t$$

$$s.t. \quad \mathcal{G}(\mathbf{S}_0 \dots \mathbf{S}_T, [\mathbf{U}_t, \mathbf{\Sigma}_t, \mathbf{V}_t], 1 \leq t \leq T) \leq \Theta$$

- c_t : whether to restart; \mathcal{G} : evaluating the margin; Θ : threshold

$$\mathcal{G} = \max_{1 \leq t \leq T} \frac{\mathcal{J}(t) - \mathcal{L}(\mathbf{S}_t, k)}{\mathcal{L}(\mathbf{S}_t, k)}$$

- Intuition: keep the maximum margin within a threshold while reducing the number of restarts

Solution: Lazy Restarts

- Lazy restarts: restart only when the margin exceeds the threshold
- Problem: intrinsic loss is hard to compute
 - Direct calculation has the same time complexity as SVD
- Relaxation: an upper bound on margin
 - A lower bound on intrinsic loss $\mathcal{L}(S, k)$

$$\mathcal{L}(\mathbf{S}_t, k) \geq B(t) \Rightarrow \frac{\mathcal{J}(t) - \mathcal{L}(\mathbf{S}_t, k)}{\mathcal{L}(\mathbf{S}_t, k)} \leq \frac{\mathcal{J}(t) - B(t)}{B(t)}.$$

- $\mathcal{J}(t)$: current loss; $\mathcal{L}(S_t, k)$: intrinsic loss; $B(t)$: bound of intrinsic loss

A Lower Bound of SVD Intrinsic Loss

- Idea: use matrix perturbation

Theorem 1 (A Lower Bound of SVD Intrinsic Loss). *If \mathbf{S} and $\Delta\mathbf{S}$ are symmetric matrices, then:*

$$\mathcal{L}(\mathbf{S} + \Delta\mathbf{S}, k) \geq \mathcal{L}(\mathbf{S}, k) + \Delta tr^2(\mathbf{S} + \Delta\mathbf{S}, \mathbf{S}) - \sum_{l=1}^k \lambda_l, \quad (9)$$

where $\lambda_1 \geq \lambda_2 \dots \geq \lambda_k$ are the top- k eigenvalues of $\nabla_{S^2} = \mathbf{S} \cdot \Delta\mathbf{S} + \Delta\mathbf{S} \cdot \mathbf{S} + \Delta\mathbf{S} \cdot \Delta\mathbf{S}$, and

$$\Delta tr^2(\mathbf{S} + \Delta\mathbf{S}, \mathbf{S}) = tr((\mathbf{S} + \Delta\mathbf{S}) \cdot (\mathbf{S} + \Delta\mathbf{S})) - tr(\mathbf{S} \cdot \mathbf{S}).$$

- Intuition: treat changes as a perturbation to the original network
- Results: need to calculate top- k eigenvalues of $\mathbf{S} \cdot \Delta\mathbf{S} + \Delta\mathbf{S} \cdot \mathbf{S} + \Delta\mathbf{S} \cdot \Delta\mathbf{S}$

Time Complexity Analysis

Theorem 2. *The time complexity of calculating $B(t)$ in Eqn (13) is $O(M_S + M_L k + N_L k^2)$, where M_S is the number of the non-zero elements in $\Delta \mathbf{S}$, and N_L, M_L are the number of the non-zero rows and elements in ∇_{S^2} respectively.*

- If every node has a equal probability of adding new edges, we have: $M_L \approx 2d_{avg}M_S$, where d_{avg} is the average degree of the network .
- For Barabasi Albert model (Barabási and Albert 1999), a typical example of preferential attachment networks, we have: $M_L \approx \frac{12}{\pi^2} [\log(d_{max}) + \gamma] M_S$, where d_{max} is the maximum degree of the network and $\gamma \approx 0.58$ is a constant.

□ Conclusion: the complexity is only **linear** to the **local dynamic changes**

TIMERS: Theoretically Instructed Maximum-Error-bounded Restart of SVD

Algorithm 1 TIMERS: Theoretically Instructed Maximum-Error-bounded Restart of SVD

Require: Static Adjacency Matrix \mathbf{A}_0 , Dynamic Changes $\Delta\mathbf{A}_1 \dots \Delta\mathbf{A}_T$, Similarity Function $\mathcal{S}(\cdot)$, Dimensionality k , Error Threshold Θ , Incremental SVD method $\mathcal{F}(\cdot)$

Ensure: SVD results in each time slice $[\mathbf{U}_t, \mathbf{\Sigma}_t, \mathbf{V}_t]$

```

1: Calculate the initial similarity  $\mathbf{S}_0 = \mathcal{S}(\mathbf{A}_0)$ 
2: Calculate SVD on  $\mathbf{S}_0$  to obtain  $[\mathbf{U}_0, \mathbf{\Sigma}_0, \mathbf{V}_0]$ 
3: for  $t$  in  $1:T$  do
4:   Calculate the similarity change in that time slice  $\Delta\mathbf{S}_t$ 
5:   Use  $\mathcal{F}(\cdot)$  to get updated SVD results  $[\mathbf{U}_t, \mathbf{\Sigma}_t, \mathbf{V}_t]$ 
6:   Compute loss  $\mathcal{J}(t)$  and bound  $B(t)$  using (13)
7:   if  $\frac{\mathcal{J}(t) - B(t)}{B(t)} > \Theta$  then
8:     Calculate SVD on  $\mathbf{S}_t$  to obtain new  $[\mathbf{U}_t, \mathbf{\Sigma}_t, \mathbf{V}_t]$ 
9:   end if
10:  Return  $[\mathbf{U}_t, \mathbf{\Sigma}_t, \mathbf{V}_t]$ 
11: end for

```

- ❑ **General** across different types of networks and dynamic scenarios
 - ❑ Weighted/unweighted, signed/unsigned networks
 - ❑ Add/delete nodes/edges, adjust edge weights
- ❑ **Flexible** to cooperate with any incremental SVD method

Experimental Setting

- ❑ Five real dynamic networks:
 - ❑ FACEBOOK, MATH, WIKI: social networks
 - ❑ DBLP: citation network
 - ❑ INTERNET: autonomous system connections
- ❑ Types: weighted/unweighted, signed/unsigned

Table 1: The Statistics of Datasets

Dataset	Nodes	Static E	Evolving E	Type
FACEBOOK	52804	962654	632188	SN
MATH	13586	600000	138408	W
WIKI	28223	1000000	375270	W,S
DBLP	28331	200000	79436	W
INTERNET	32077	111644	196270	W,SN

W = Weighted, S = Signed, SN = Static Network marked

Experimental Setting

□ Baselines:

- Heu-FL: restart SVD after a fixed number of edges changed
- Heu-FT: restart SVD after a fixed amount of time passed
- LWI2 (Chen and Candan, KDD 2014): restart SVD whenever the reconstruction loss exceeds a preset threshold

□ Tasks:

- Approximation error
 - Fixed number of restarts
 - Fixed maximum error
- Applications
 - Link prediction
 - Eigenvalue tracking
- Analysis

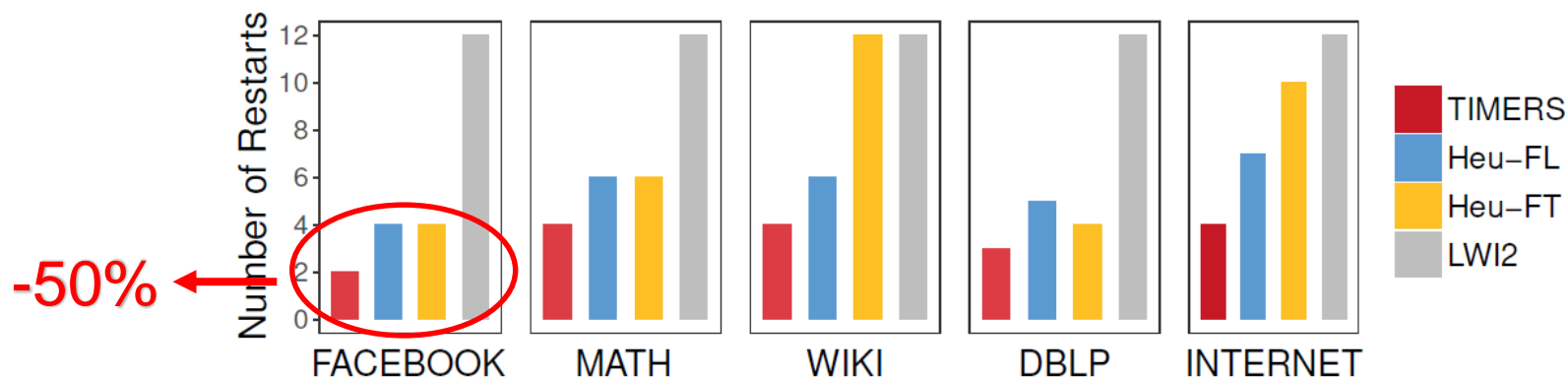
Experimental Results: Approximation Error

□ Fixing number of restarts

Dataset	$avg(r)$				$max(r)$			
	TIMERS	LWI2	Heu-FL	Heu-FT	TIMERS	LWI2	Heu-FL	Heu-FT
FACEBOOK	0.005	0.020	0.009	0.011	0.014	0.038	0.025	0.023
MATH	0.037	0.057	0.044	0.051	0.085	0.226	0.117	0.179
WIKI	0.053	0.086	0.071	0.281	0.139	0.332	0.240	0.825
DBLP	0.042	0.110	0.053	0.064	0.121	0.386	0.198	0.238
INTERNET	0.152	0.218	0.196	0.961	0.385	0.806	0.647	1.897


□ Fixing maximum error

27%~42% Improvement



Experimental Results: Applications


□ Link prediction:



Dataset	TIMERS	LWI2	Heu-FL	Heu-FT
FACEBOOK	1.54*	4.27	2.21	2.81
MATH	1.14*	2.68	1.31	1.29
WIKI	1.06*	3.44	1.63	4.13
DBLP	0.18*	0.32	0.22	0.27
INTERNET	11.27*	23.36	16.98	34.30

*: outperform other methods at 0.005 level paired t-test in 10 runs.

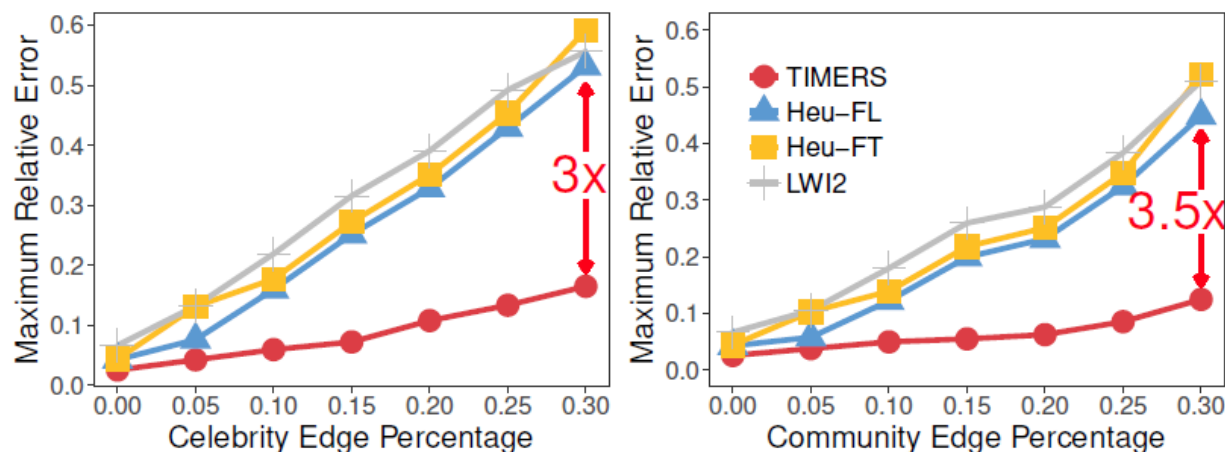
□ First eigenvalue tracking:



Dataset	TIMERS	LWI2	Heu-FL	Heu-FT
FACEBOOK	0.66	0.94	1.11	1.53
MATH	2.27	5.03	4.92	4.80
WIKI	15.45	18.42	17.73	97.08
DBLP	8.31	11.42	13.92	24.66
INTERNET	4.18	12.56	7.95	57.38

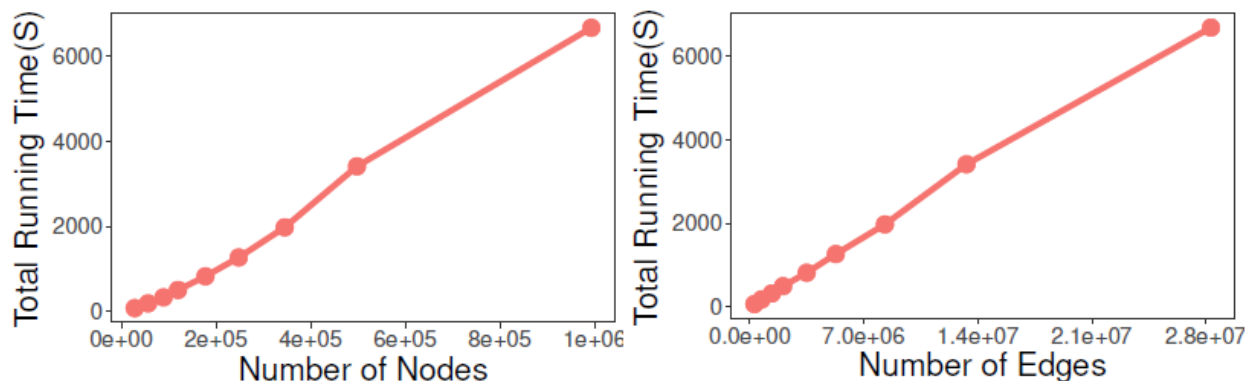
Experimental Results: Analysis

- Syntactic networks: simulate drastic changes in the network structure



- Robust to sudden changes

- Linear scalability



Conclusion

- ❑ When to **restart SVD** for **dynamic networks**
 - ❑ Key to prevent error accumulation for incremental SVD
- ❑ Framework: **monitor margin**
 - ❑ Constrained optimization and lazy restarts
 - ❑ **A lower bound of SVD Intrinsic Loss** to reduce time complexity
 - ❑ Linear time complexity to the local dynamic changes
- ❑ Extensive experiments on several dynamic networks
 - ❑ **50% improvement** in approximation error
 - ❑ **Robust** to sudden changes in the network structure
- ❑ Future direction: generalize to directed networks and non-square matrices (e.g. user-product)

Thanks!

Ziwei Zhang, Tsinghua University

zw-zhang16@mails.tsinghua.edu.cn

<https://cn.linkedin.com/in/zhangziwei>

Reference:

- [1] Zhang, Z; Cui, P; Pei, J; Wang, X; Zhu, W. TIMERS: Error-Bounded SVD Restart on Dynamic Networks. AAAI, 2018.