TEMPORAL GRAPH PATTERN MINING

PRESENTED BY

EUGENE YANG EY120@GEORGETOWN.EDU

MOTIVATION

- Find the relationship between clinical activities
- Encode these information as feature

DEPENDENCY PATTERNS IN CLINICAL PATHWAYS

Lin, Fu-ren, et al. "Mining time dependency patterns in clinical pathways." International Journal of Medical Informatics 62.1 (2001): 11-25.

BASIC SETUP

- An activity record(process log data) of the process P_i can be written as a 4-tuple

$$\langle P_i, A_{ij}, T_{s,ij}, T_{e,ij} \rangle$$

- If activity A_{ij} starts right after A_{ik} without any other activities in between, we say $\boldsymbol{A_{ij}}$ is depends on $\boldsymbol{A_{ik}}$
- We can then construct a directed acyclic dependency graph

$$G_p = (V_p, E_p)$$

where V_p is the activities with pseudo start/end node and E_p is the Boolean dependency between every 2 activities.

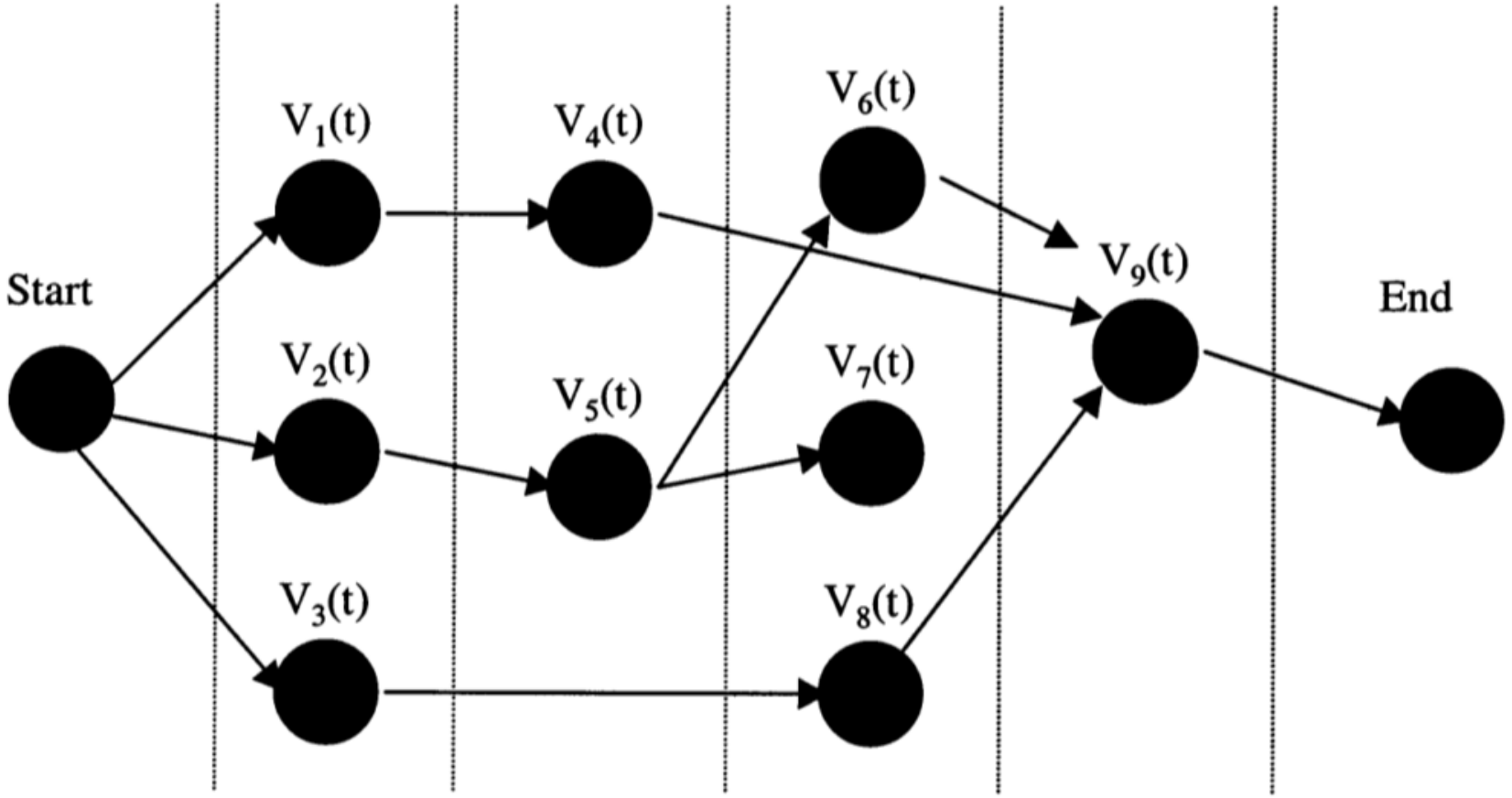
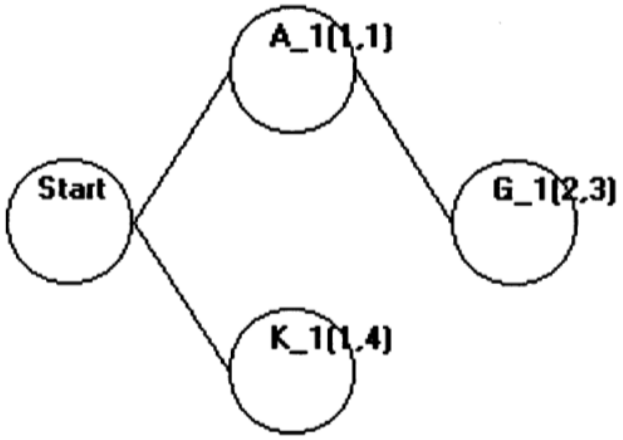


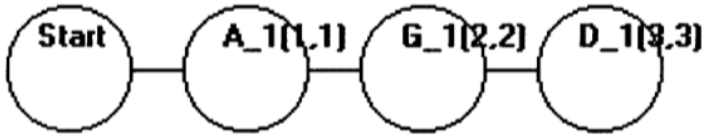
Fig. 1. Sample of dependency graph.

AGGREGATING THE INFORMATION

- A Large Graph LG_k is set of aggregated graphs with at most k nodes
- Algorithm
  1. Create activity nodes by counting the activity with support over certain threshold (i.e. creating LG_1).
  2. By looking at dependency between every 2 activities, create LG_2 by adding pairs with certain minimum support.
  3. Construct LG_k by adding edges that have not included in LG_{k-1} from LG_2



(a)



(b)

Fig. 2. An example of LG_4.

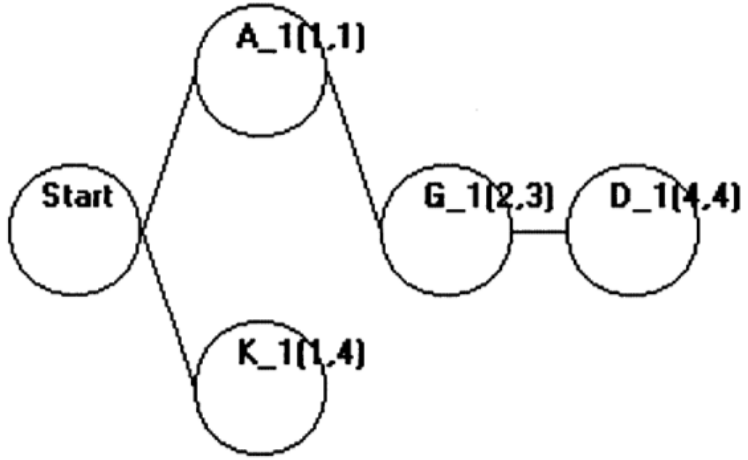


Fig. 3. An example of LG_5.

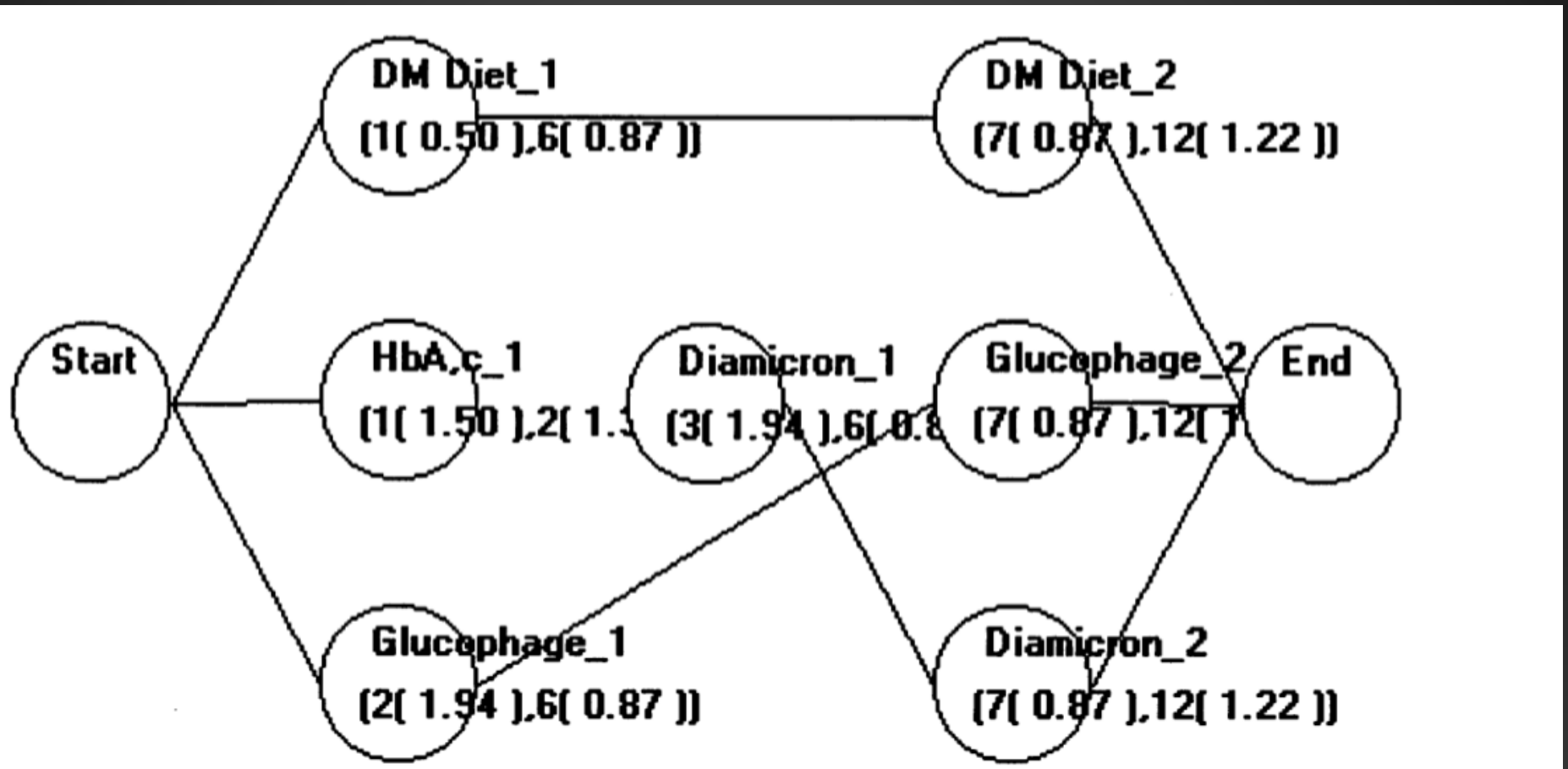


Fig. 6. An example of clinical pathways.

EXAMPLE ON REAL DATA

- Adding some runtime activities
  - E.g. blood pressure
- Using training/testing to test the validity and stability of this method

$$Sim(P_i, P_j) = \frac{2 \times |E_i \cup E_j|}{|E_i| + |E_j|}$$

$$Predicion\ Accuracy = \frac{1}{n} \sum_{i=1}^{n} \max_{P_j \in Training} Sim(P_i, P_j)$$

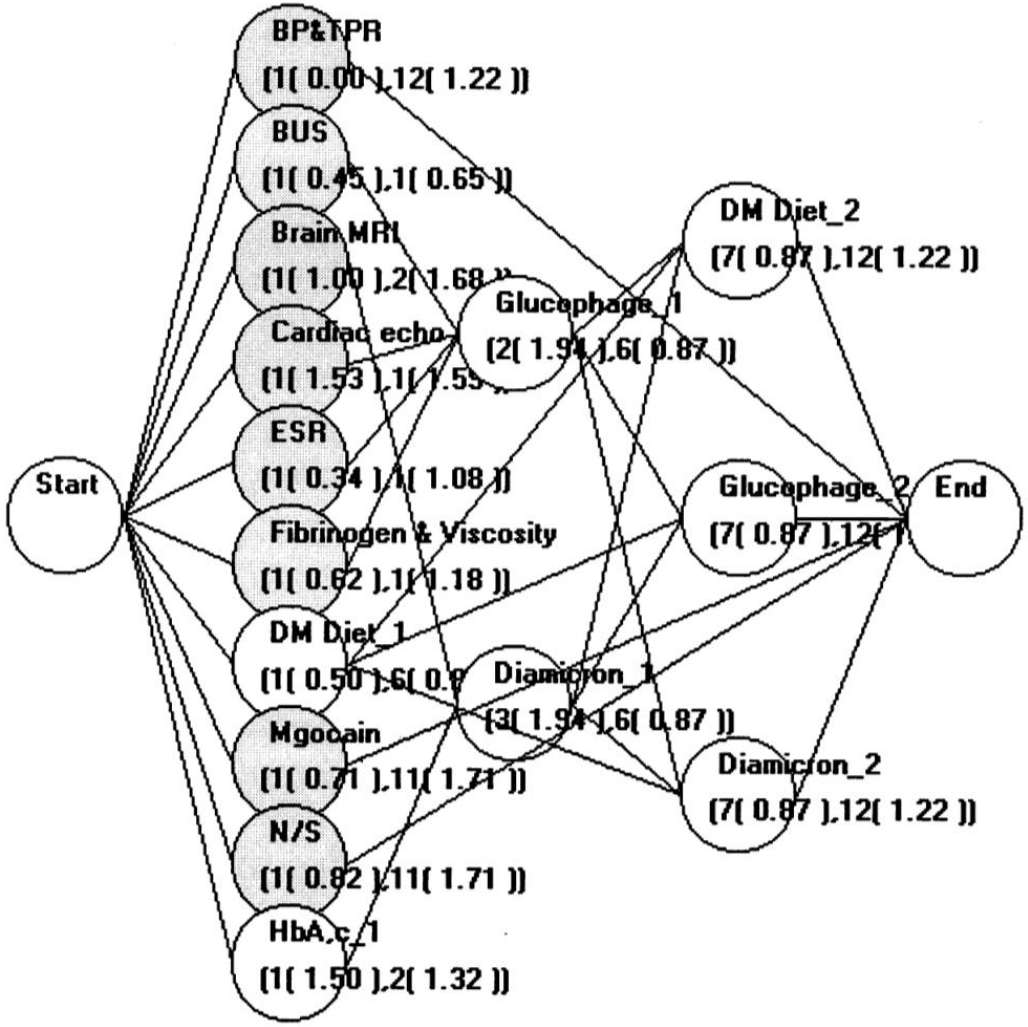


Fig. 7. The clinical pathways with routine activities.

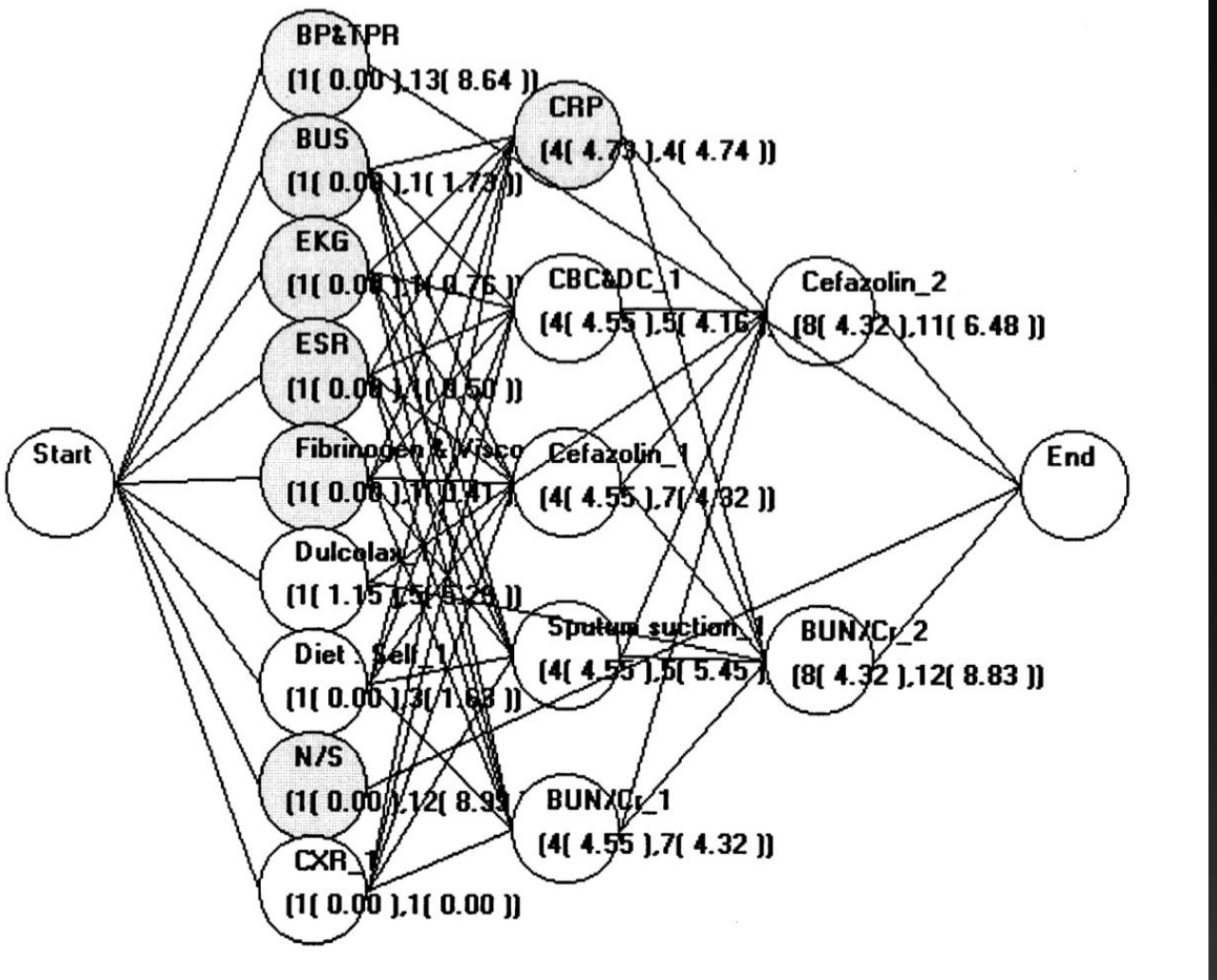


Fig. 8. A more complex case.

Table 1
Results of mining time dependency graphs

Number of vertex	Number of paths (before maximal)	Number of paths (after maximal)	Average support of large graph (%)
2	278	29	7
3	712	94	5
4	848	115	4
5	615	107	4
6	273	92	3
7	72	29	3
8	14	5	3
9	2	2	3

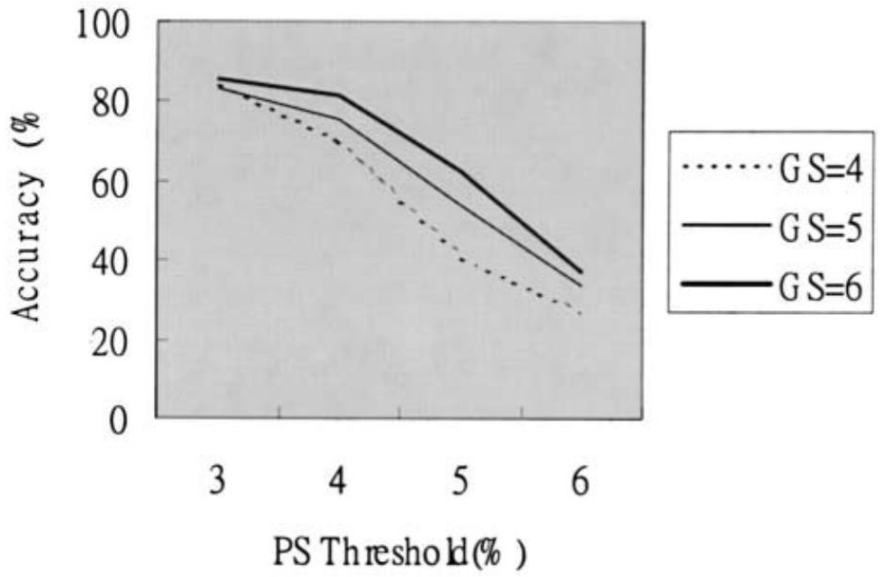


Fig. 9. The prediction accuracy under different *GS* and *PS*.

TEMPORAL PHENOTYPING WITH GRAPH BASED FRAMEWORK

Liu, Chuanren, et al. "Temporal phenotyping from longitudinal electronic health records: A graph based framework." *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015.

BASIC SETUP

- Motivation: try to embed relationship between events as features
- Construct a weighted directed temporal graph of event sequence
  - E.g. diagnosis, medication, lab test, etc.
- Weight Function

$$W_{ij}^{n} = \frac{1}{L_n} \sum_{1 \le p \le q \le L_n} \left[x_{np} = i \wedge x_{nq} = j \right] \kappa(t_{nq} - t_{np})$$

where $\kappa(\cdot)$ is a non-increasing function(here we choose as Exponential distribution function if input is greater than constant Δ)

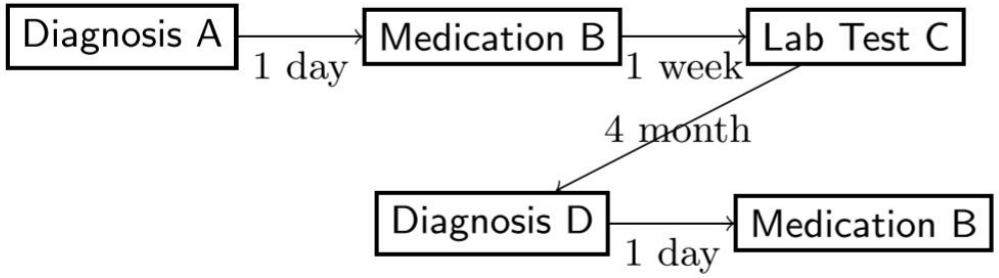


Figure 1: One example of medical event sequence of one subject (potential patient).

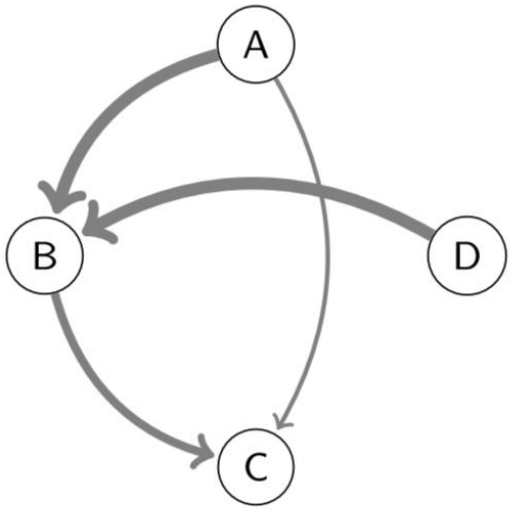


Figure 2: The temporal graph of event sequence in Figure 1.

BASIS DECOMPOSITION OF GRAPH

- Decompose graph weight matrix into aggregation of basis
- Use index of basis as feature

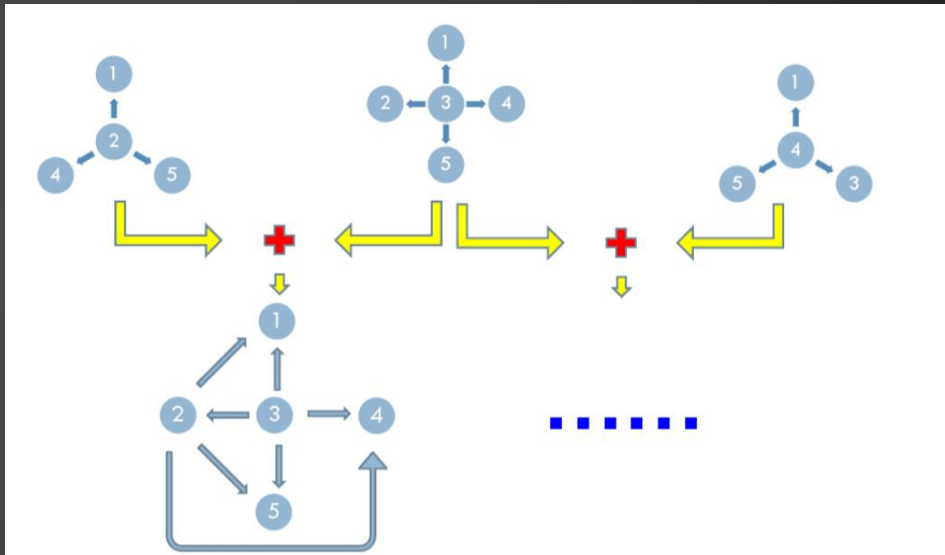


Figure 3: Example of temporal phenotyping.

BASIS DECOMPOSITION OF GRAPH

- For weight matrix W^n,

$$W^n = \sum_{k=1}^{K} A_{nk} B^k$$

where $A \in \mathbb{R}^{N \times K}$

- We minimize the estimation error

$$\mathcal{J}(A, B) = \frac{1}{2} \sum_{n=1}^{N} \left\| W^n - \sum_{k=1}^{K} A_{nk} B^k \right\|_F^2$$

where $\|\cdot\|_F$ is the matrix Frobenius norm.

REGULARIZATION

- Like most of the ML minimization problem, we can we can add regularization term to objective function

$$\mathcal{J}(A,B) = \frac{1}{2}\sum_{n=1}^{N}\left\|W^n - \sum_{k=1}^{K} A_{nk}B^k\right\|_F^2 + \lambda\Omega(A)$$

where regularization function $\Omega(A) \geq 0$ and some constant $\lambda \geq 0$

REGULARIZATION

- Similarity based regularization

$$\Omega(A) = \frac{1}{2}\sum_{n_1,n_2}\frac{1}{2}\left\|A_{n_1} - A_{n_2}\right\|^2 S_{n_1 n_2} = \frac{1}{2}tr(A'LA)$$

  where S is the similarity symmetric matrix and $L = D - S$

- Model based regularization

$$\Omega(A) = -\frac{1}{|\mathcal{L}|}\sum_{n\in\mathcal{L}}\log \Pr(A_n, Y_n|\mathcal{H})$$

  where $\mathcal{L}$ is the training set and Y_n is the correspondent label

REGULARIZATION

- We can easily embed a discriminative model under this setup, e.g. logistic regression

$$\Pr(A_n, Y_n|\mathcal{H}) = \frac{1}{1+\exp(-Y_n f(A_n))}$$
$$\mathcal{H}: A_n \longmapsto f(A_n) = \mathrm{A_n}\Theta + \theta$$

- Or hinge loss(SVM)

$$\Omega(A_n) = \frac{1}{|\mathcal{L}|}\sum_{n\in\mathcal{L}} \max(0, 1 - Y_n f(A_n))$$

OPTIMIZATION

- Similarity based regularization

$$A \leftarrow proj_{splx}\left(A - \alpha \frac{\partial \mathcal{J}}{\partial A}\right)$$

$$\min_{B_{ij}^*} \frac{1}{2} \sum_{n=1}^{N} \left(W_{ij}^n - \sum_{k=1}^{K} A_{nk} B_{ij}^k \right)^2$$

OPTIMIZATION

- Model based regularization
- The loss function space could be convex but not differentiable(e.g. SVM)
- Apply proximal gradient optimization

Algorithm 1 The algorithm updating A_n for $n \in \mathcal{L}$ with hinge loss regularization

Initialize A_n
repeat
  $A_n \leftarrow A_n - \alpha(A_n\langle B \otimes B\rangle - \langle W^n \otimes B\rangle)$
  $A_n \leftarrow A_n + Y_n \operatorname{proj}_{[0,\alpha\frac{\lambda}{|\mathcal{L}|}]}(\frac{1-Y_n f(A_n)}{\|\Theta\|^2})\Theta'$
  $A_n \leftarrow \operatorname{proj}_{\text{splx}}(A_n)$
until Convergence

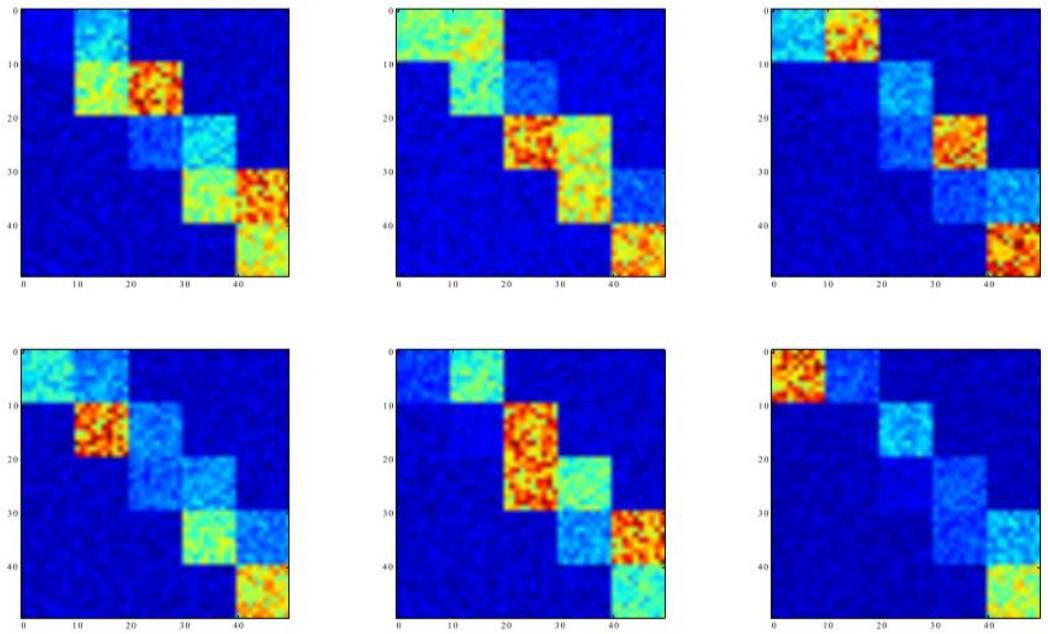


Figure 4: Examples of the synthetic data. Every sample is a 50x50 matrix generated by convex combination of the basis in Figure 5. The combination coefficients are first generated from uniform distribution within [0,1] and then normalized.

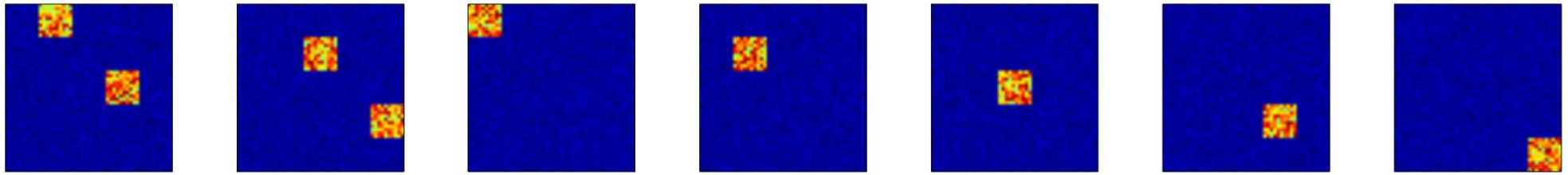


Figure 5: The basis for generating the synthetic data. Every base is a 50x50 matrix. The background noise are with values randomly generated from uniform distribution in [0,0.1]. The foreground blocks are 10x10 with values generated from uniform distribution in [0,1].

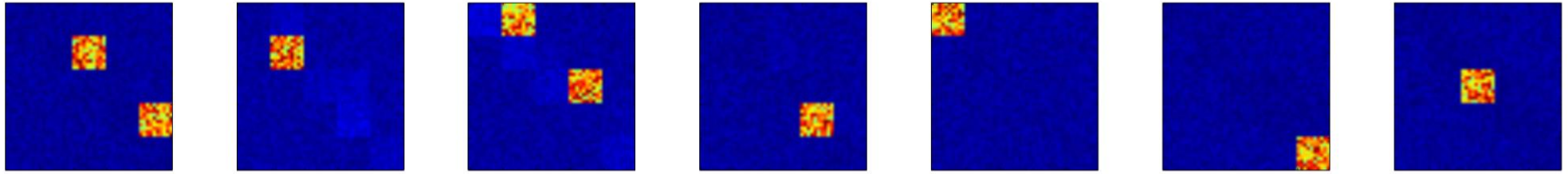


Figure 6: The basis learned by our algorithm without any regularizations. By comparing them with the basis in Figure 5 we can see they exactly match on the foreground patterns. There are only slight differences on the background noise.

Metric	Graph	Temporal Phenotyping			
		Un	Sim	Logit	Hinge
AUC	0.78±0.05	0.82±0.02	0.79±0.02	**0.84±0.01**	0.81±0.02
AUPR	0.64±0.04	0.71±0.02	0.72±0.05	**0.76±0.01**	0.74±0.03
ACC	0.87±0.08	0.87±0.01	0.89±0.04	0.89±0.02	**0.89±0.01**

Table 1: The classification performance over 10-fold cross validation on synthetic data.

REAL EXAMPLE

- Congestive Heart Failure(CHF)
- One-year hospitalization prediction after CHF confirmation
- Prediction of CHF

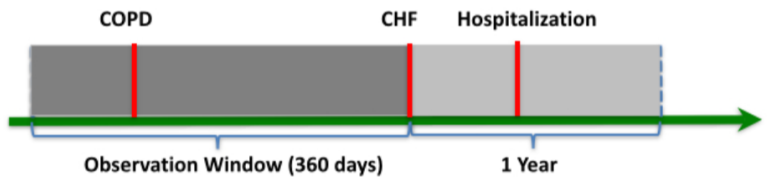


Figure 7: Experimental setting of hospitalization prediction.

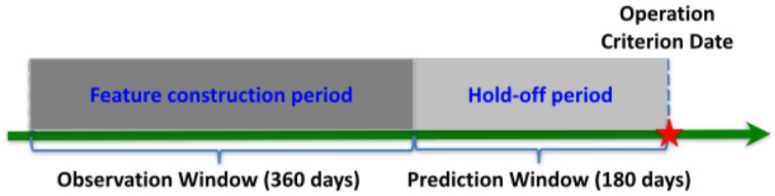
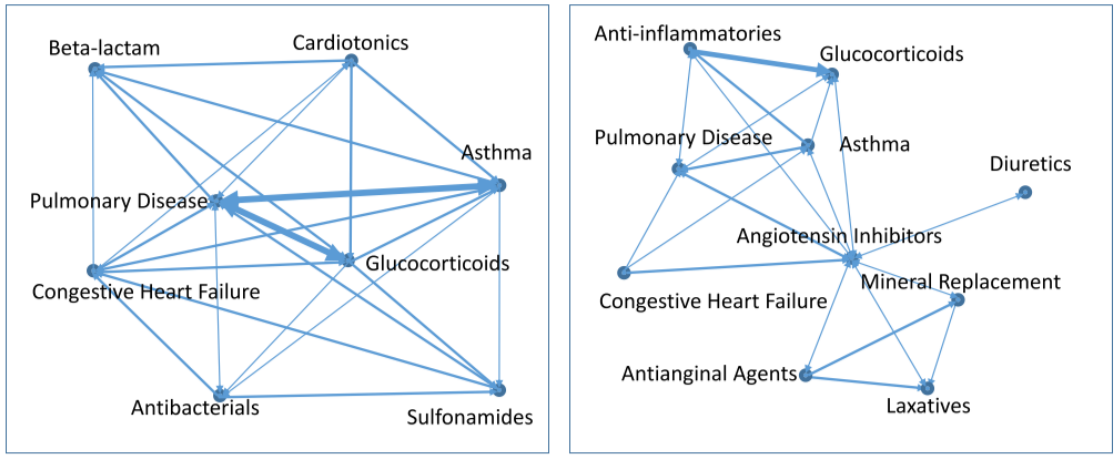
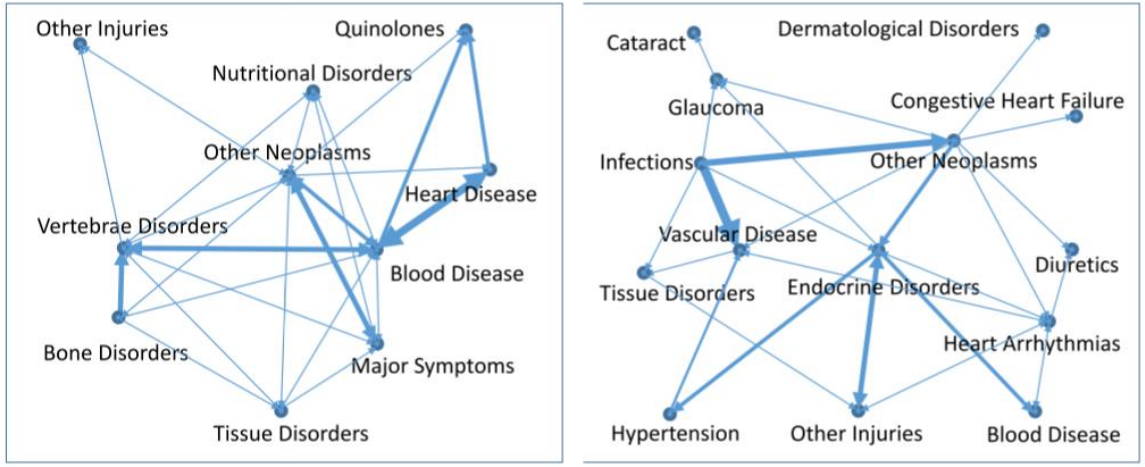


Figure 8: Experimental setting of CHF early prediction.



(a) Case instance. (b) Control instance.

Figure 9: Temporal graph examples of a case and control patient in hospitalization prediction data.

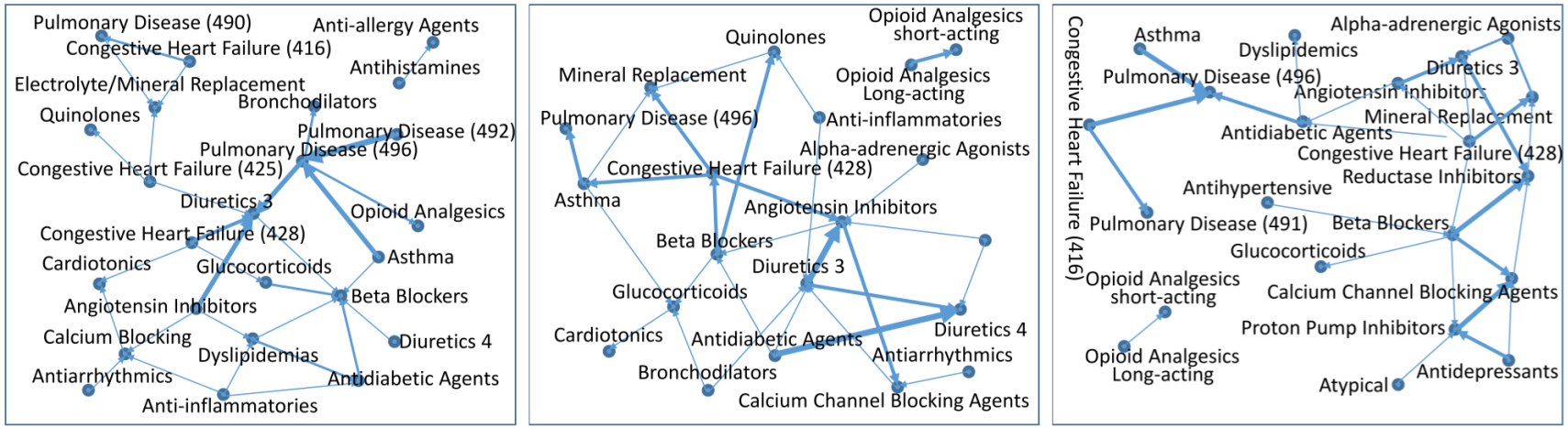


(a) Case instance. (b) Control instance.

Figure 10: Temporal graph examples of a case and control patient in CHF prediction data.

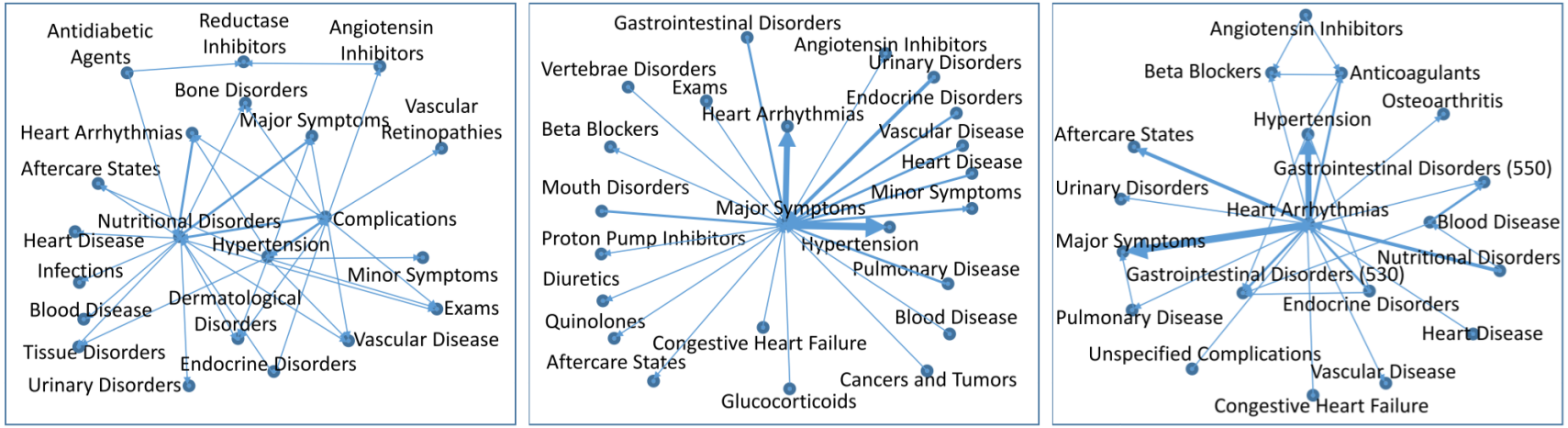
Data	Metric	AVR	BPS	TES	Temporal Phenotyping			
					Un	Sim	Logit	Hinge
CHF	AUC	0.70±0.03	0.69±0.04	0.67±0.02	0.71±0.02	0.69±0.03	**0.72±0.01**	0.72±0.04
	APR	0.41±0.05	0.52±0.06	0.37±0.04	0.62±0.01	0.60±0.04	**0.65±0.01**	0.62±0.03
	ACC	0.76±0.02	0.77±0.08	0.77±0.02	0.77±0.02	0.78±0.02	0.79±0.01	**0.80±0.04**
Hospitalization	AUC	0.56±0.11	0.67±0.05	0.65±0.06	0.73±0.08	0.71±0.10	**0.73±0.06**	0.69±0.10
	APR	0.32±0.09	0.58±0.13	0.38±0.07	0.64±0.04	0.65±0.15	**0.67±0.12**	0.64±0.16
	ACC	0.65±0.11	0.75±0.05	0.73±0.08	0.76±0.07	**0.80±0.07**	0.79±0.04	0.77±0.05

Table 2: The classification performance over 10-fold cross validation on two real-world data sets.



(a) Phenotype I (b) Phenotype II (c) Phenotype III

Figure 13: Example of temporal phenotypes of hospitalization prediction data. The number following a drug name indicates the strength of the drug, i.e., it is used to treat CHF of which stage. The three digits in the parentheses correspond to the first three digits of ICD-9.



(a) Phenotype I (b) Phenotype II (c) Phenotype III

Figure 14: Example of temporal graph basis of CHF prediction data. The three digits in the parentheses correspond to the first three digits of ICD-9.

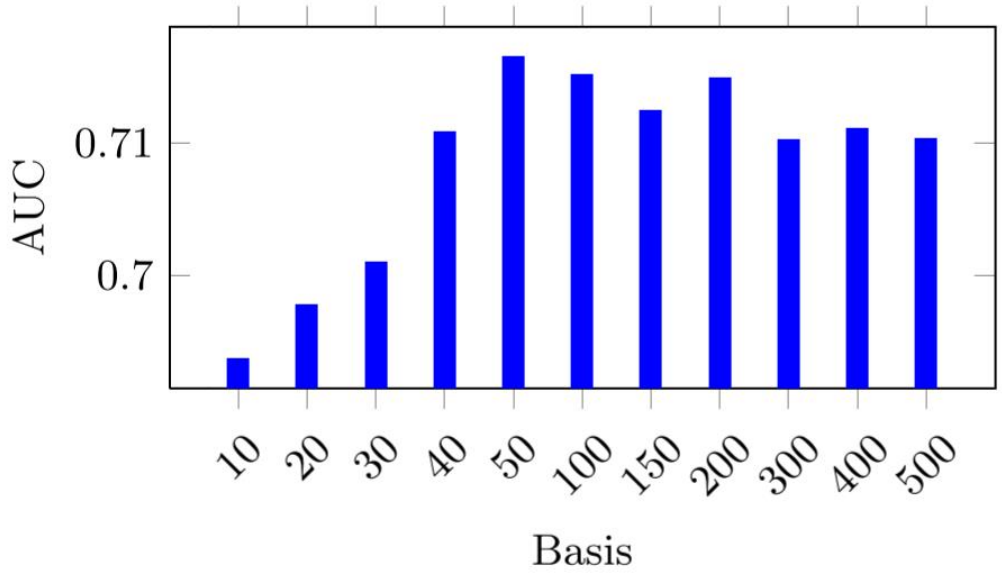


Figure 12: The AUC of phenotyping representations with different number of basis. Averaged with 10 runs of 10-fold cross validation.

QUESTIONS?