

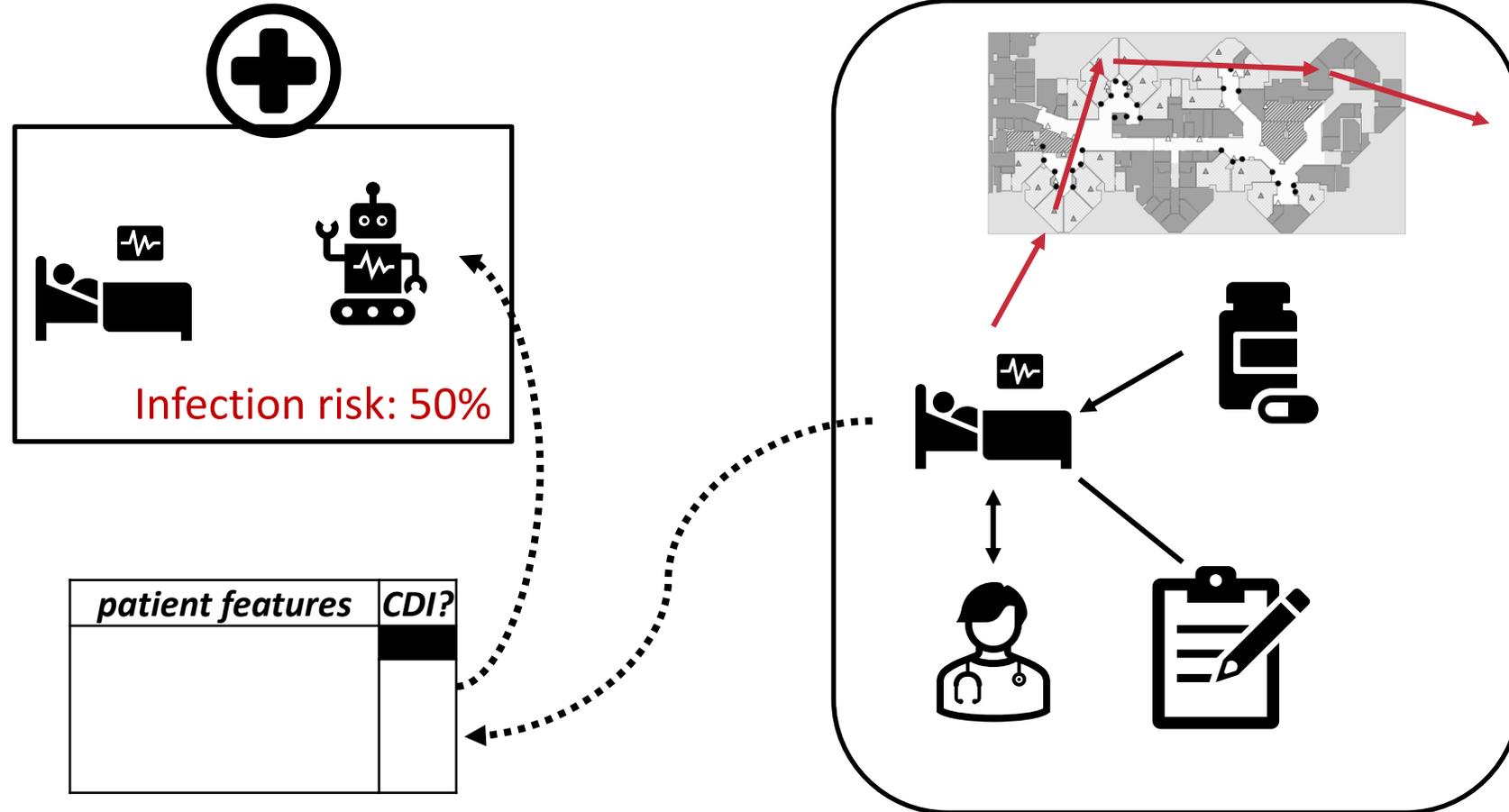
# Dynamic Healthcare Embeddings for Improving Patient Care

Presenter: Hankyu Jang

Co-authors: Sulyun Lee, D.M.Hasibul Hasan, Philip M. Polgreen, Sriram V. Pemmaraju, Bijaya, Adhikari

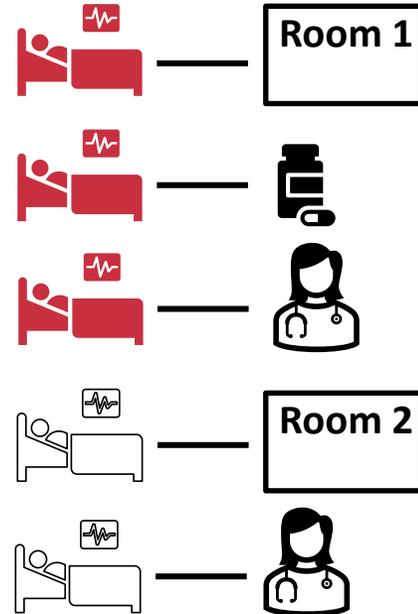
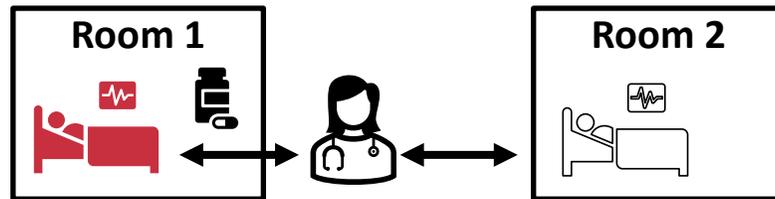
\*Funded by CDC Mind Healthcare Network grant and NSF grant

# Motivation for learning representations of patients



# What we want to capture in the embeddings

Elevated risk of a patient to additional infection



Domain knowledge

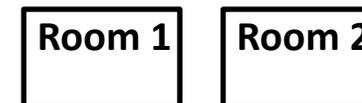
Same specialty



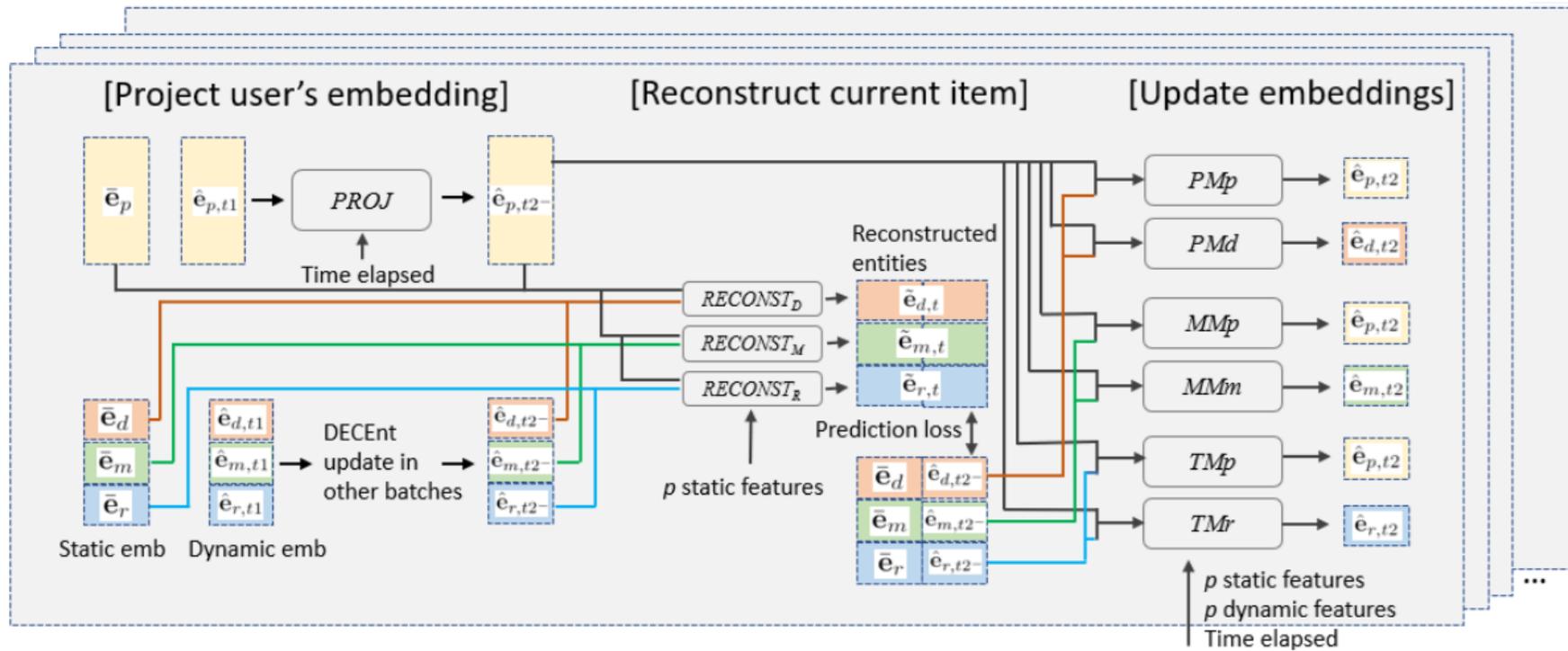
Similar medications



Nearby rooms



# DECEnt: Dynamic Embedding of healthCare Entities



DECEnt is a general-purpose, unsupervised embedding method for dynamic and heterogeneous interactions

- DECEnt preserves information on the interaction via interaction type specific autoencoder
- DECEnt guides the embedding to capture domain knowledge

# Problem formulation

## Given

- A set  $S$  of time-stamped interactions among healthcare entities
- Static networks  $G_{room}$ ,  $G_{med}$ , and  $G_{doc}$
- Dynamic and static attributes of patients  $\hat{p}$  and  $p$

## Learn dynamic embeddings for each time $t$

- $\hat{e}_{u,t}$  for each entity  $u \in \mathcal{P}$
- $\hat{e}_{v,t}$  for each entity  $v \in \mathcal{D} \cup \mathcal{M} \cup \mathcal{R}$

## Such that

- $f(\hat{e}_{u,t}, \hat{e}_{v,t})$  encodes information to be predictive of  $v$
- $\text{dist}(\hat{e}_{v,t}, \hat{e}_{v',t})$  of the same type  $v$  and  $v'$  reflects the distance between the two in  $G_{entitytype}$

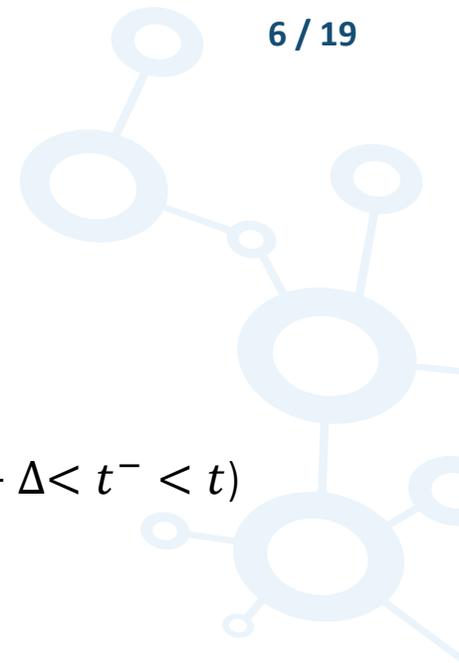
## Notations

- $\mathcal{P}$  : a set of patients
- $\mathcal{D}$  : a set of doctors
- $\mathcal{M}$  : a set of medications
- $\mathcal{R}$  : a set of rooms
- $G_{doc}$ : a graph of doctors
- $G_{med}$ : a graph of medications
- $G_{room}$ : a graph of hospital rooms
- $\hat{e}_{u,t}$ : dynamic embedding of  $u$  at time  $t$



# DECEnt – high level overview

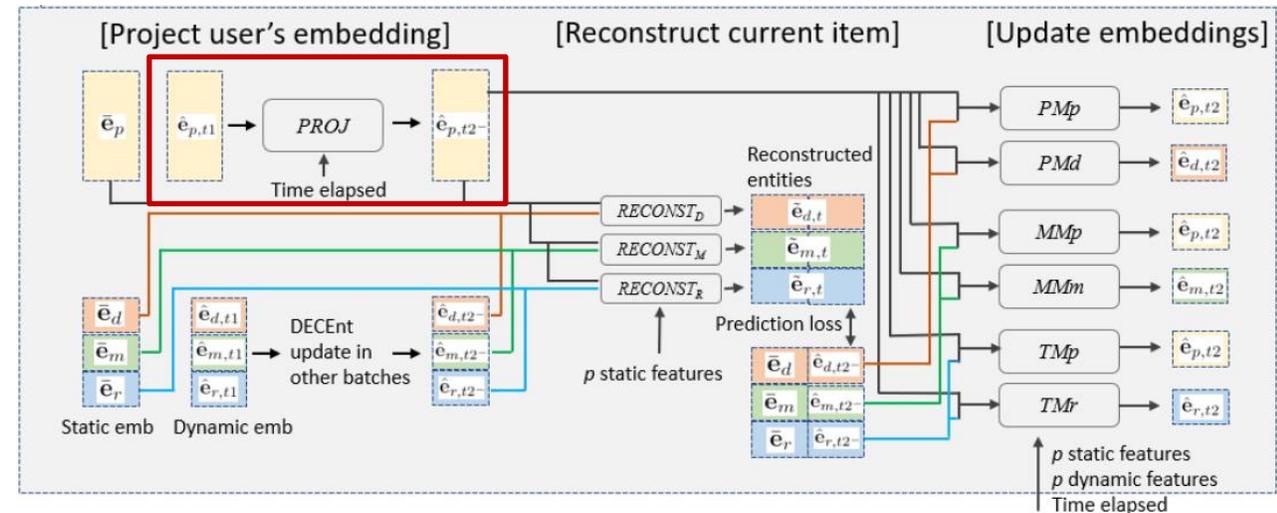
- The learning process is *patient centric*
- Whenever an interaction (e.g., patient  $p$  – doctor  $d$ ) occur at time  $t$ , we
  - project patient  $p$ 's embedding at the time of  $t - \Delta$  ( $p$ 's previous interaction time) to  $t^-$  for  $(t - \Delta < t^- < t)$
  - update  $p$ 's embedding at time  $t^-$  to time  $t$  via the module  $PM_p$
  - update  $d$ 's embedding at time  $t^-$  to time  $t$  in the module  $PM_d$
- While ensuring
  - Temporal consistency: dynamic embeddings do not change drastically
  - Reconstruction: we preserve each interaction information by reconstructing the embedding of the entity
  - Domain knowledge: we guide that similar nodes (in terms of domain knowledge) to have similar embeddings



# Project patient's embedding

- For an interaction made at time  $t + \Delta$ , we need patient's dynamic embedding at time  $t + \Delta$
- However, the patient's embedding would have changed from the patient's previous interaction at time  $t$
- To model this, we project the patient embedding at time  $t$  to  $t + \Delta$  by a projection module [+]

$$\hat{\mathbf{e}}_{p,t+\Delta} = (1 + \mathbf{W} \times \Delta) * \hat{\mathbf{e}}_{p,t}$$

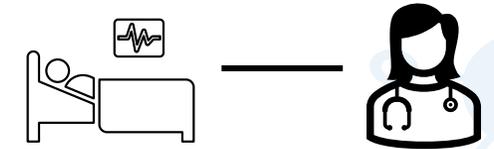


# Update dynamic embeddings of patient and entity

- For a physician interaction  $(p, d, t) \in \mathcal{PR}$ , we update dynamic embeddings of patient and physician
- We model this via a co-evolutionary neural network (*physician module*)

$$\hat{\mathbf{e}}_{p,t} = \sigma \left[ \mathbf{W}_p^{PM} [\hat{\mathbf{e}}_{p,t-} \mid \hat{\mathbf{e}}_{d,t-} \mid \Delta_{p,t} \mid \mathbf{p}_p \mid \hat{\mathbf{p}}_{p,t}] + \mathbf{B}_p^{PM} \right]$$

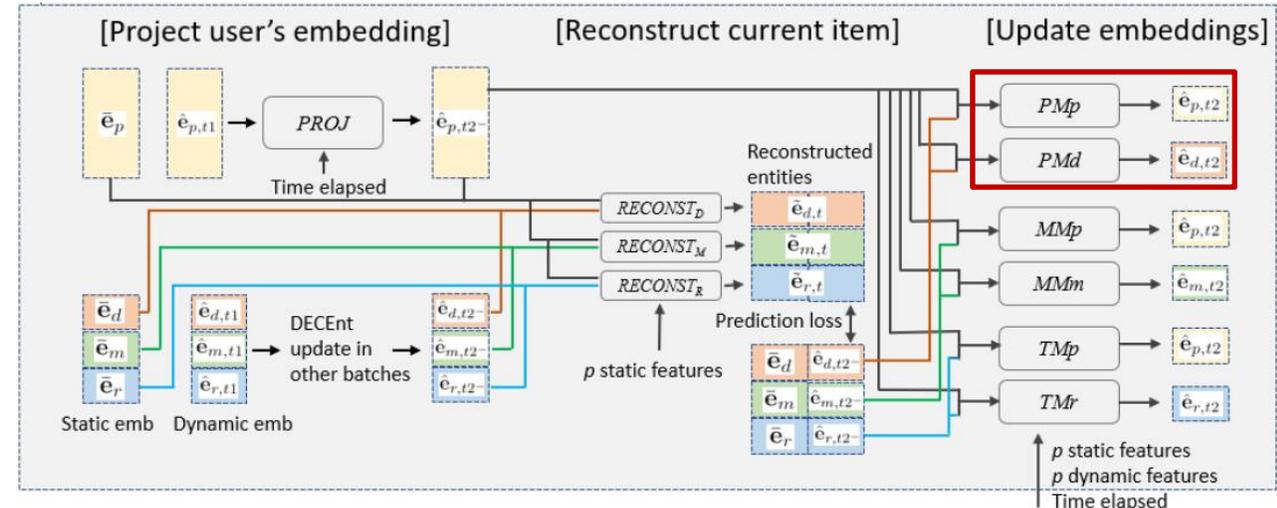
$$\hat{\mathbf{e}}_{d,t} = \sigma \left[ \mathbf{W}_d^{PM} [\hat{\mathbf{e}}_{d,t-} \mid \hat{\mathbf{e}}_{p,t-} \mid \Delta_{d,t} \mid \mathbf{p}_p \mid \hat{\mathbf{p}}_{p,t}] + \mathbf{B}_d^{PM} \right]$$



- The *medication module* and the *transfer module* are defined accordingly
- We don't want the dynamic embeddings to change drastically, so we minimize the *temporal consistency loss*, defined as the following:

$$L_{temp} = \sum_{(p,e,t) \in \mathcal{S}} \|\hat{\mathbf{e}}_{p,t} - \hat{\mathbf{e}}_{p,t-}\|_2 + \|\hat{\mathbf{e}}_{e,t} - \hat{\mathbf{e}}_{e,t-}\|_2$$

where  $\mathcal{S}$  denotes the set of interaction tuples



# Reconstruct static and dynamic embedding of entity

- For a physician interaction  $(p, d, t) \in \mathcal{PR}$ , we preserve the information by reconstructing the dynamic and static embedding of the doctor
- We model this via an autoencoder (**reconstruction module**)
  - Our autoencoder is different from vanilla auto encoder that it takes patient information as input as well

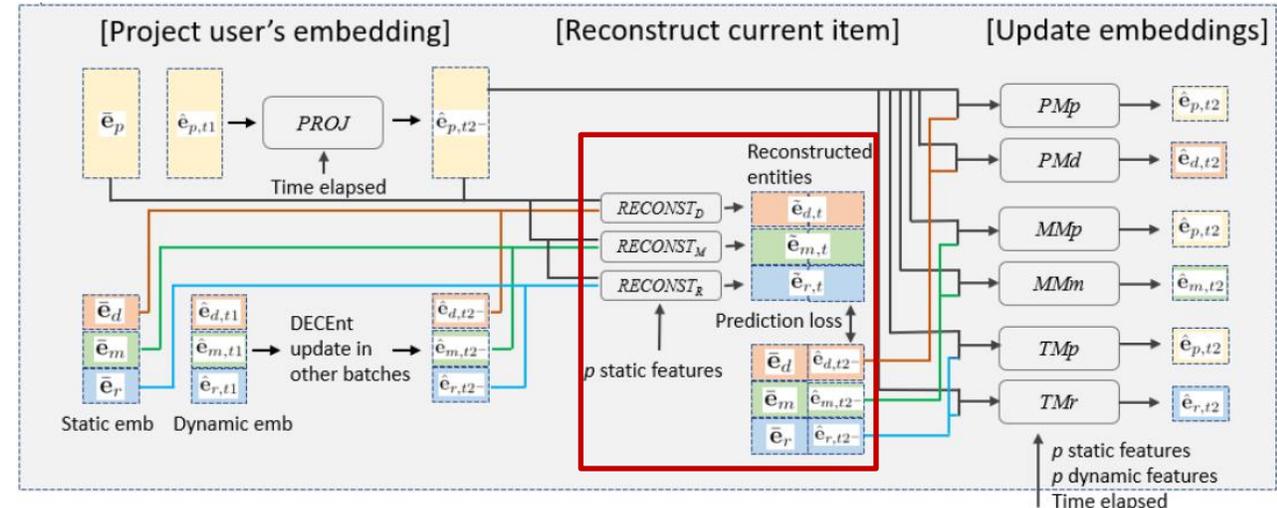
$$\tilde{\mathbf{e}}_{d,t_2^-} = \mathbf{W}_d \left[ \hat{\mathbf{e}}_{p,t_2^-} \mid \bar{\mathbf{e}}_p \mid \mathbf{p}_p \mid \hat{\mathbf{e}}_{d,t_2^-} \mid \bar{\mathbf{e}}_d \right] + \mathbf{B}_d$$

- The **reconstruction module** for medication and room interaction are modeled accordingly
- Reconstruction loss

$$L_{reconst} = \sum_{(p,d,t) \in \mathcal{PR}} \|\tilde{\mathbf{e}}_{d,t^-} - [\hat{\mathbf{e}}_{d,t^-} \mid \bar{\mathbf{e}}_d]\|_2$$

$$+ \sum_{(p,m,t) \in \mathcal{MD}} \|\tilde{\mathbf{e}}_{m,t^-} - [\hat{\mathbf{e}}_{m,t^-} \mid \bar{\mathbf{e}}_m]\|_2$$

$$+ \sum_{(p,r,t) \in \mathcal{TR}} \|\tilde{\mathbf{e}}_{r,t^-} - [\hat{\mathbf{e}}_{r,t^-} \mid \bar{\mathbf{e}}_r]\|_2$$



# Capture domain knowledge

- We ensure that the entities known to be similar as per domain knowledge to have similar embeddings
- We compute Laplacian matrix for each entity graph (edges denote similarity)
- Then, compute the graph Laplacian based *domain specific loss*

$$\begin{aligned}
 L_{dom} = & \lambda_{dom}^D \sum_{t \in [0, T], d \in \mathcal{D}} \hat{\mathbf{e}}_{d,t}^T \mathbf{L}_{doc} \hat{\mathbf{e}}_{d,t} \\
 & + \lambda_{dom}^M \sum_{t \in [0, T], m \in \mathcal{M}} \hat{\mathbf{e}}_{m,t}^T \mathbf{L}_{med} \hat{\mathbf{e}}_{m,t} \\
 & + \lambda_{dom}^R \sum_{t \in [0, T], r \in \mathcal{R}} \hat{\mathbf{e}}_{r,t}^T \mathbf{L}_{room} \hat{\mathbf{e}}_{r,t}
 \end{aligned}$$

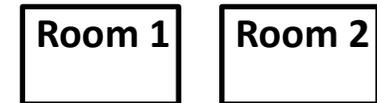
Same specialty



Similar medications



Nearby rooms



# Overall loss and training

- We jointly train all the modules, and the overall loss is the weighted aggregation of
  - $L_{temp}$ : Temporal consistency loss
  - $L_{reconst}$ : Reconstruction loss
  - $L_{dom}$ : Domain specific loss
- Training details
  - Adam optimizer. learning rate (1e-3) decay (1e-5)
  - Dynamic embedding size: 128
  - Epochs: 1000 (early stopping. patience 10 epochs)

# Data

- Three months of timestamped patient-entity interaction data at hospital, captured from
  - Electronic health records
  - Admission-discharge-transfer logs

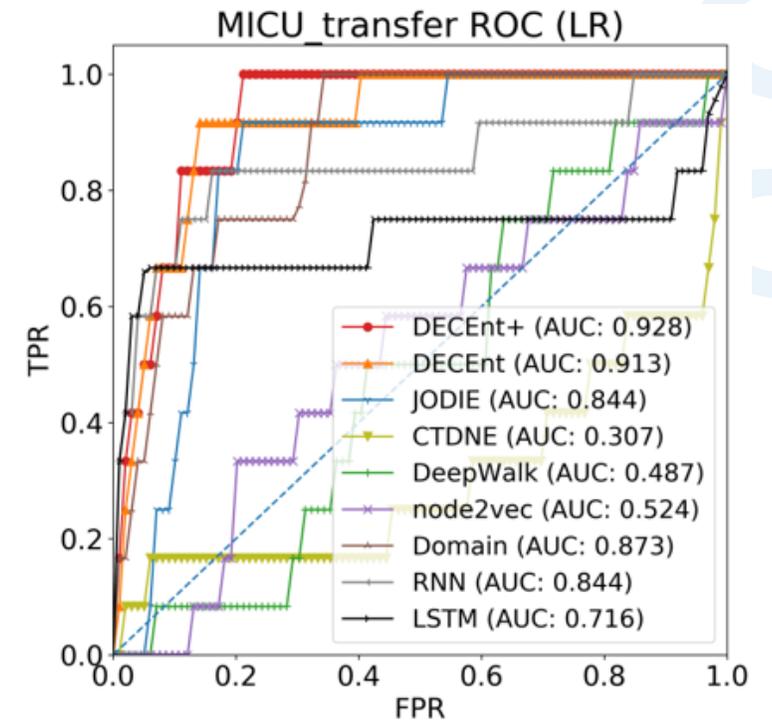
	Unique counts (in K)	Interaction with patients (in K)
<b>Patient</b>	6.5	-
<b>Doctor</b>	0.6	23.1
<b>Medication</b>	0.7	349.3
<b>Room</b>	0.6	16.7

## Entity Graphs

- $G_{doc}(\mathcal{D}, E_{doc})$  : Graph of doctors, where  $(d_1, d_2) \in E_{doc}$  are based on the proximity of specialty of doctors  $d_1$  and  $d_2$
- $G_{med}(\mathcal{M} \cup M, E_{med})$  : Hierarchy (tree) of medications  $\mathcal{M}$  and their subtypes ( $M$ ). Each leaf  $m \in \mathcal{M}$  in  $G_{med}$  denote medication
- $G_{room}(\mathcal{R}, E_{room})$  : Graph representation of hospital rooms.  $(r_1, r_2) \in E_{room}$  denote the physical proximity of rooms  $r_1$  and  $r_2$

# Application1: MICU transfer prediction

- Forecast whether a patient is at risk of transfer to a Medical Intensive Care Unit (MICU)
- MICU provides care for patients at a critical stage
  - an early indication of potential MICU transfer helps hospital officials allocate resources better
- Binary classification problem
  - Instance: Patient at time  $t$
  - Feature: Patient's embedding
  - Label: Indication of patient transfer into MICU the next day



# Application 2: CDI Prediction

- Early detection of an healthcare associated infection (HAI) is critical to prevent infection spread
- Clostridioides difficile infection (CDI) is one of a common HAI, increases mortality risk of patients with weakened immune system
- Binary classification problem
  - Instance: Patient at time  $t$
  - Feature: Patient's embedding
  - Label: indication of patient getting CDI within the next 3 days [+]

Method	AUC		
RNN	0.56 (0.119)		
LSTM	0.585 (0.103)		
-	LR	RF	MLP
DOMAIN	0.655 (0.123)	0.709 (0.104)	0.582 (0.137)
DEEPWALK	0.494 (0.087)	0.487 (0.093)	0.492 (0.103)
NODE2VEC	0.453 (0.098)	0.43 (0.106)	0.478 (0.1)
CTDNE	0.463 (0.101)	0.528 (0.079)	0.483 (0.116)
JODIE	0.552 (0.192)	0.377 (0.177)	0.469 (0.176)
DECENT	0.732 (0.069)	0.711 (0.08)	0.668 (0.082)
DECENT +	<b>0.736 (0.064)</b>	0.717 (0.078)	0.664 (0.091)

<sup>a</sup>The value in bold denotes best performance

# Application 3: Mortality and Case Severity Risk Prediction

- Predicting case severity and mortality while the patient is in the hospital has many applications
  - Personalized patient care
  - Resource allocation
- The Agency for Health Research and Quality (AHRQ) performs mortality and severity analysis of inpatient visits across hospitals in US
- Multi-class classification problem
  - Instance: Patient on the last day of the visit
  - Feature: Patient's embedding
  - Label: one of 'minor', 'moderate', 'major', 'extreme'

Method	Mortality	Severity
RNN	0.276 (0.039)	0.31 (0.032)
LSTM	0.289 (0.033)	0.308 (0.026)
DOMAIN	0.22 (0.017)	0.258 (0.007)
DEEPWALK	0.172 (0.034)	0.192 (0.019)
NODE2VEC	0.172 (0.02)	0.196 (0.009)
CTDNE	0.184 (0.019)	0.199 (0.007)
JODIE	0.143 (0.039)	0.193 (0.014)
DECENT	0.421 (0.027)	0.34 (0.014)
DECENT+	<b>0.428 (0.022)</b>	<b>0.349 (0.015)</b>

<sup>a</sup>The value in bold denotes best performance

F1 Macro

# Case study on doctor embeddings: dispersion score

- We compute *dispersions* for subsets of healthcare entities
- We partition doctors  $\mathcal{D}$  into disjoint sets. E.g.,
  - $\mathcal{D}_1$ : set of general internal medicine physicians
  - $\mathcal{D}_2$ : set of Otolaryngologists (ears, nose, throat)
  - ...
- Then, we compute pairwise doctor dispersion between  $\mathcal{D}_i$  and  $\mathcal{D}_j$

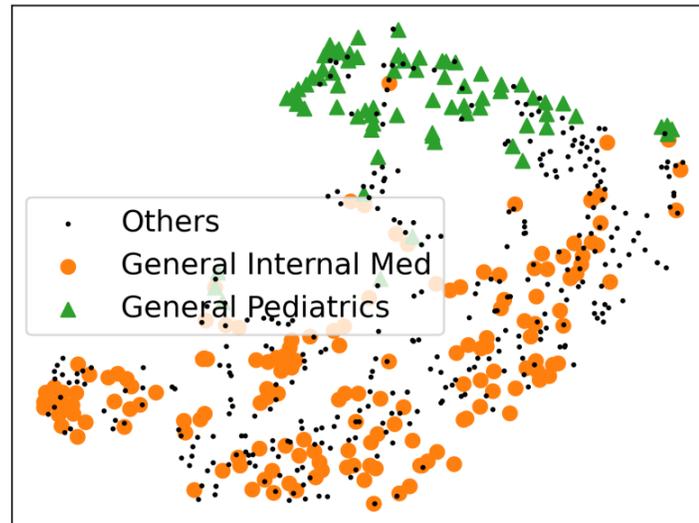
$$disp_{\mathcal{D},t}(i,j) = \frac{\sum_{d \in \mathcal{D}_i, d' \in \mathcal{D}_j} \|\hat{\mathbf{e}}_{d,t} - \hat{\mathbf{e}}_{d',t}\|_2}{|\mathcal{D}_i| \cdot |\mathcal{D}_j|}$$

- The dispersion score provides a measure of avg. distance between doctors from different specialties in the computed embeddings

- Doctors that interacts with wide variety of patients have *small* dispersion score with others. E.g.,
  - General internal medicine
  - Anesthesia
- Doctors that (i) tend not to require referral to other specialties or (ii) their unit is far away from other units have *large* dispersion score with others. E.g.,
  - Family Practice
  - Otolaryngology (ears, nose, throat)
  - Pediatrician

# Case study on doctor embeddings: 2-d projection

- 2-d projection [+] of doctor embeddings of
  - General internal medicine (**small** dispersion score with others)
  - Pediatrics (**large** dispersion score with others)



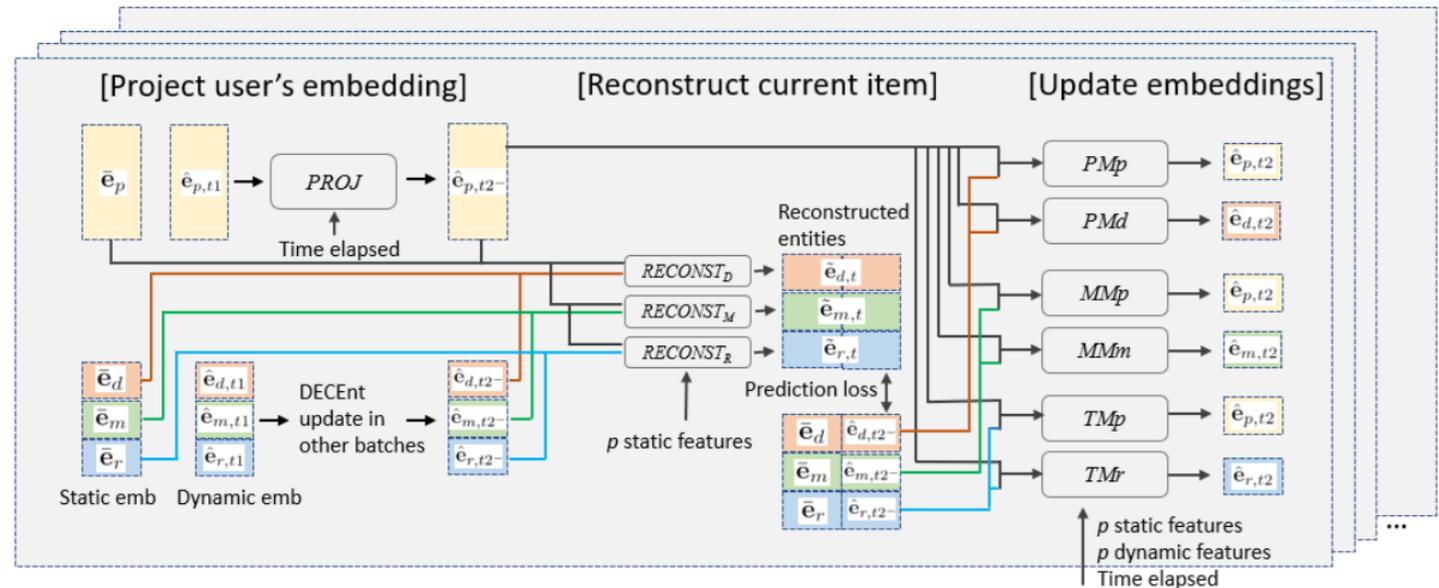
# Conclusion

- We propose DECEnt, a method for learning embeddings of nodes in dynamic, heterogeneous interactions
- DECEnt enables several healthcare predictive modeling applications
  - Adverse event
  - Early detection of healthcare associated infection
  - Case severity and mortality prediction
- DECEnt outperforms state-of-the-art baselines in all the healthcare predictive modeling tasks we consider
- Our embeddings are interpretable and meaningful



# Special thanks to

- Sulyun Lee
- D.M.Hasibul Hasan
- Prof. Philip M. Polgreen
- Prof. Sriram V. Pemmaraju
- \*Prof. Bijaya Adhikari



Back up slides



# Problem formulation

## Given

- A set  $S$  of time-stamped interactions among healthcare entities
- Static networks  $G_{room}$ ,  $G_{med}$ , and  $G_{doc}$
- Dynamic and static attributes of patients  $\hat{p}$  and  $p$

Learn dynamic embeddings for each time  $t$

- $\hat{e}_{u,t}$  for each entity  $u \in \mathcal{P}$
- $\hat{e}_{v,t}$  for each entity  $v \in \mathcal{D} \cup \mathcal{M} \cup \mathcal{R}$

## Such that

- $f(\hat{e}_{u,t}, \hat{e}_{v,t})$  encodes information to be predictive of  $v$
- $\text{dist}(\hat{e}_{v,t}, \hat{e}_{v',t})$  of the same type  $v$  and  $v'$  reflects the distance between the two in  $G_{entitytype}$

## Notations

- $\mathcal{P}$  : a set of patients
- $\mathcal{D}$  : a set of doctors
- $\mathcal{M}$  : a set of medications
- $\mathcal{R}$  : a set of rooms
- $G_{doc}(\mathcal{D}, E_{doc})$  : Graph of doctors, where  $(d_1, d_2) \in E_{doc}$  are based on the proximity of specialty of doctors  $d_1$  and  $d_2$
- $G_{med}(\mathcal{M} \cup M, E_{med})$  : Hierarchy (tree) of medications  $\mathcal{M}$  and their subtypes ( $M$ ). Each leaf  $m \in \mathcal{M}$  in  $G_{med}$  denote medication
- $G_{room}(\mathcal{R}, E_{room})$  : Graph representation of hospital rooms.  $(r_1, r_2) \in E_{room}$  denote the physical proximity of rooms  $r_1$  and  $r_2$



# Dynamic embedding methods and limitation

- The core of the learning methods for dynamic embedding is temporal link prediction [+]
  - Where given an interaction of a user-item (e.g., user purchasing an item)
  - Predict the next item that the use is likely to purchase (predicting user's decision)
- This temporal link prediction doesn't seem to be that applicable in learning patient embedding
  - E.g., a patient's next interaction entity is not based on the patient's decision
  - Rather, a better way would be to preserve the current information
  - We embed the information using autoencoder
- We have domain knowledge to consider for patient embeddings
  - We need some flexibility (e.g., a regularizer) in the dynamic embedding
  - We model this by graph Laplacian regularization on the graphs of healthcare entities

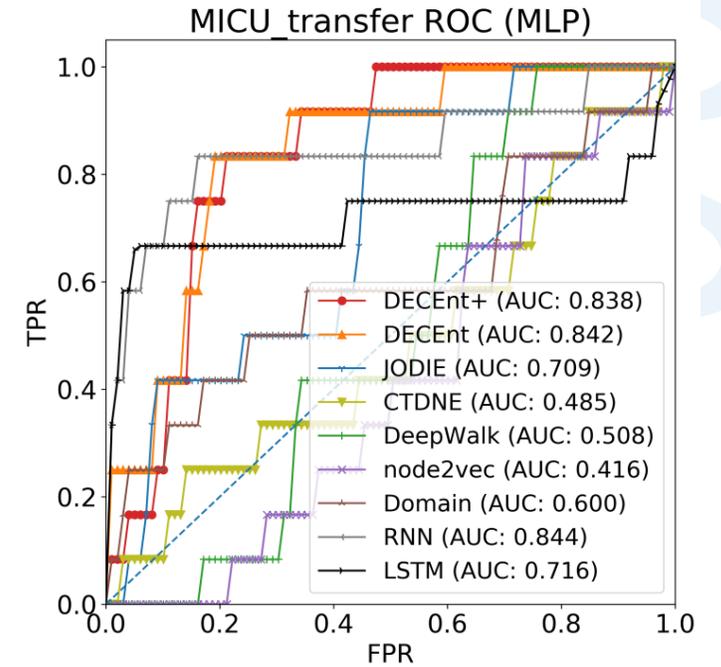
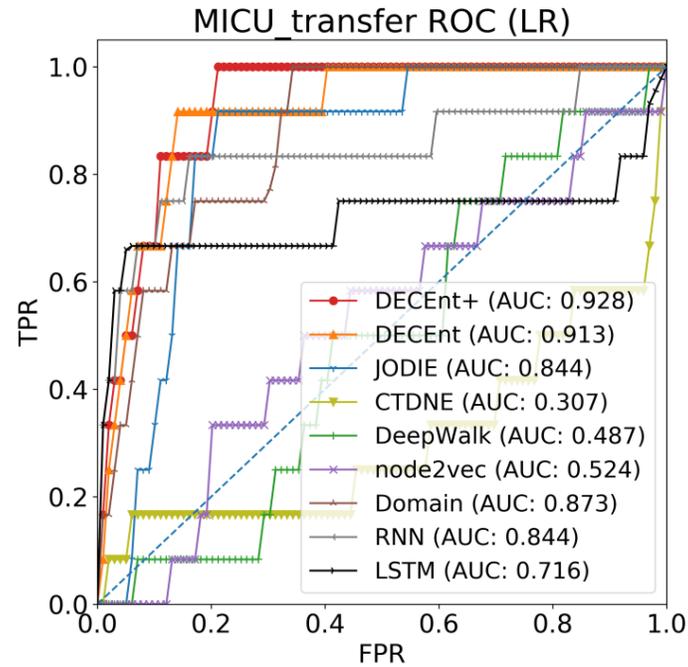
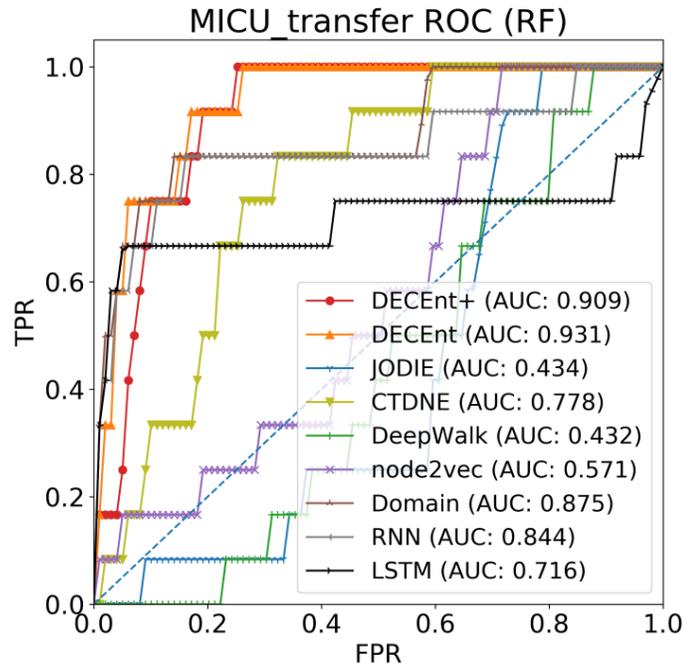


# Discussion & limitation

- Limitation in the patient–doctor interaction data
  - Currently, the data is processed from procedure data
  - We can extract richer set of data using clinical notes
- Incorporate clinical notes data in the learning procedure



# Application1: MICU transfer prediction



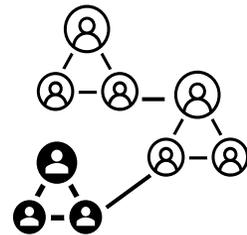
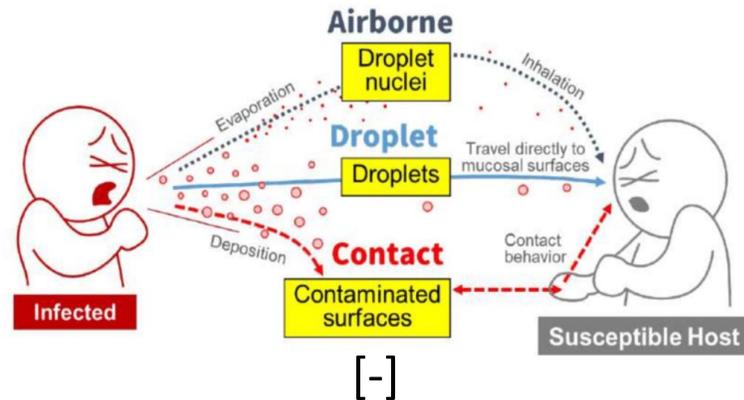
# HAIs are threat to patients

- Healthcare-associated infections (HAIs): infections that occur during care
- Each year, ~ 4% of patients in the US are diagnosed with infection during their care in the hospital [\*]
- Patients in hospitals are typically susceptible to HAIs
  - Healthcare facilities are interested in preventing HAIs
- However, there are **challenges** in designing effective interventions

## Realistic disease modeling and simulation due to

Complex nature of disease

Heterogeneity in contacts



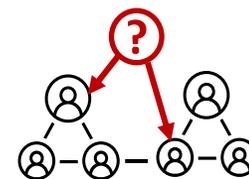
## Missing infections

Symptomatic  
(not recorded)



Infection Source

Asymptomatic  
(Latent spreaders)



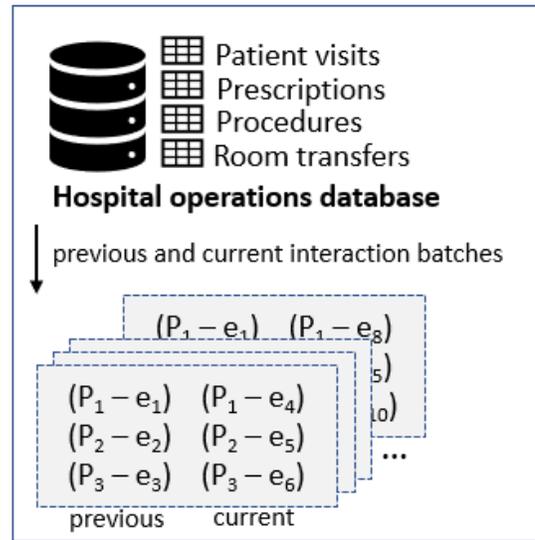
## Other challenges

- Disease specific feature engineering
- Complexity of simulations
- Environment-mediated infectious diseases
- Non-tree like cascade

[\*] Centers for Disease Control and Prevention (CDC), "Healthcare-associated infections (hais)," <https://www.cdc.gov/winnablebattles/report/HAIs.html>.

[-] Gameiro da Silva, M. An analysis of the transmission modes of COVID-19 in light of the concepts of Indoor Air Quality. Doi: 10.13140. 2020

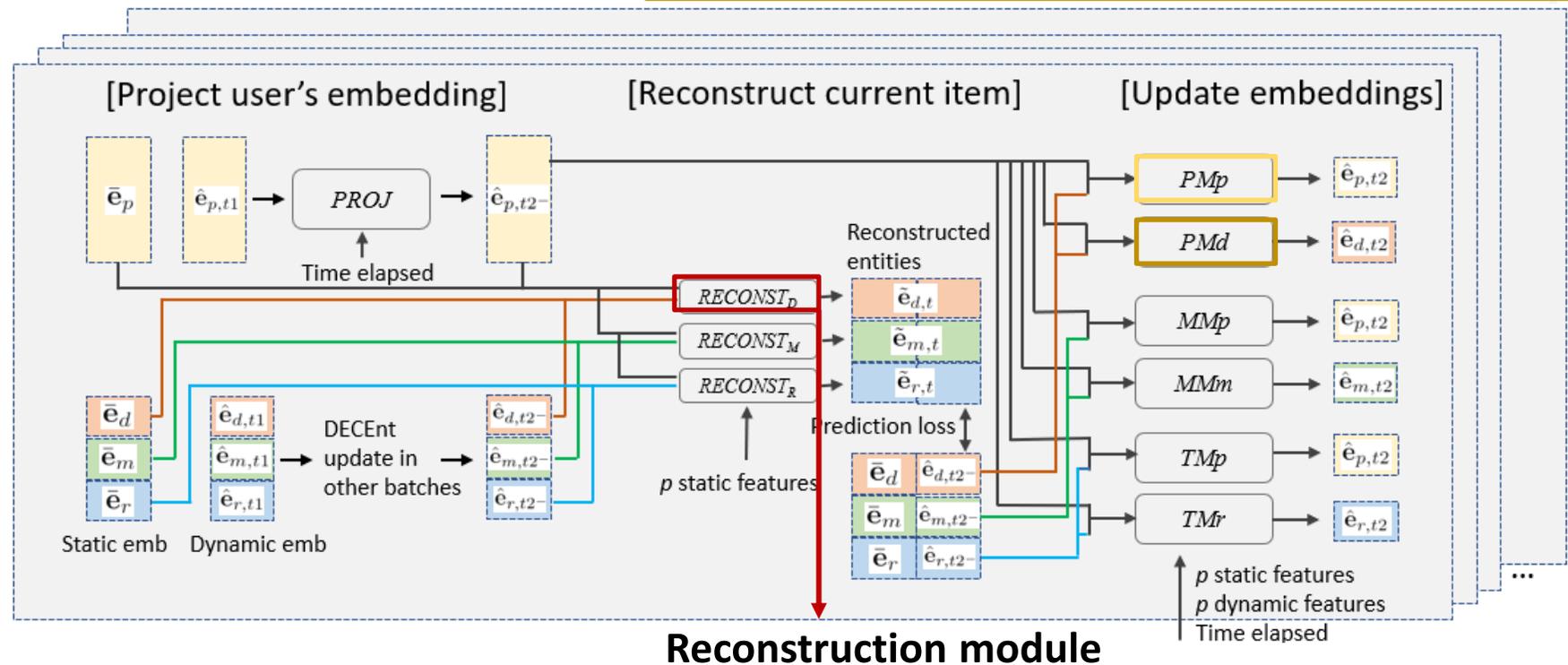
# Paper: Dynamic healthcare embeddings for improving patient care



## Physician module

$$\hat{e}_{p,t} = \sigma \left[ \mathbf{W}_p^{PM} [\hat{e}_{p,t-} \mid \hat{e}_{d,t-} \mid \Delta_{p,t} \mid \mathbf{p}_p \mid \hat{\mathbf{p}}_{p,t}] + \mathbf{B}_p^{PM} \right]$$

$$\hat{e}_{d,t} = \sigma \left[ \mathbf{W}_d^{PM} [\hat{e}_{d,t-} \mid \hat{e}_{p,t-} \mid \Delta_{d,t} \mid \mathbf{p}_p \mid \hat{\mathbf{p}}_{p,t}] + \mathbf{B}_d^{PM} \right]$$



## Reconstruction module

$$\tilde{e}_{d,t_2} = \mathbf{W}_d \left[ \hat{e}_{p,t_2-} \mid \bar{e}_p \mid \mathbf{p}_p \mid \hat{e}_{d,t_2-} \mid \bar{e}_d \right] + \mathbf{B}_d$$

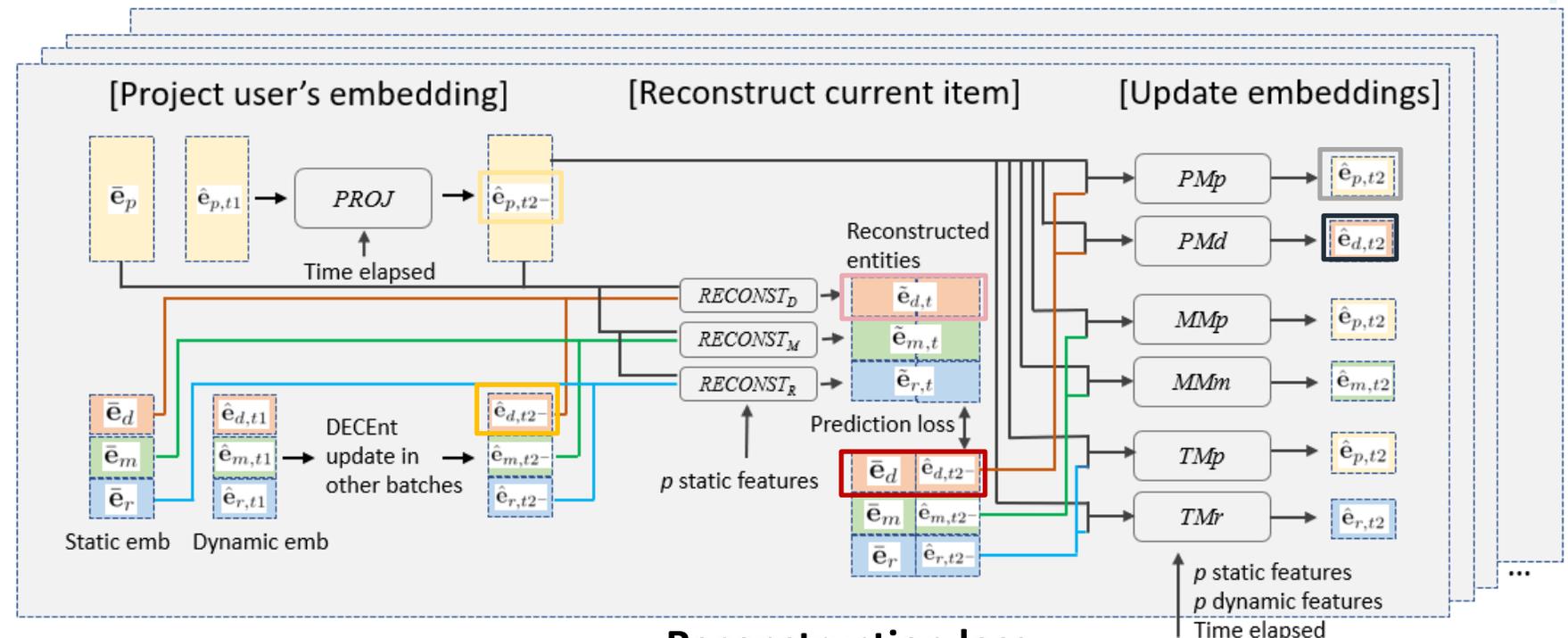
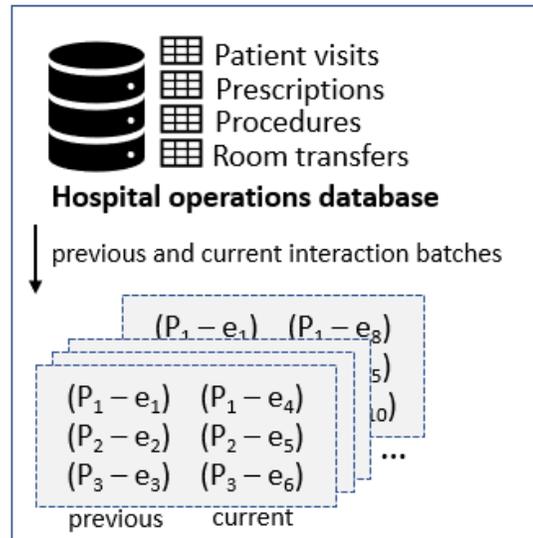
# Paper: Dynamic healthcare embeddings for improving patient care

## Domain specific loss

$$\lambda_{dom}^D \sum_{t \in [0, T], d \in \mathcal{D}} \hat{e}_{d,t}^T \mathbf{L}_{doc} \hat{e}_{d,t}$$

## Temporal consistency loss

$$L_{temp} = \sum_{(p,e,t) \in \mathcal{S}} \|\hat{e}_{p,t} - \hat{e}_{p,t-}\|_2 + \|\hat{e}_{e,t} - \hat{e}_{e,t-}\|_2$$

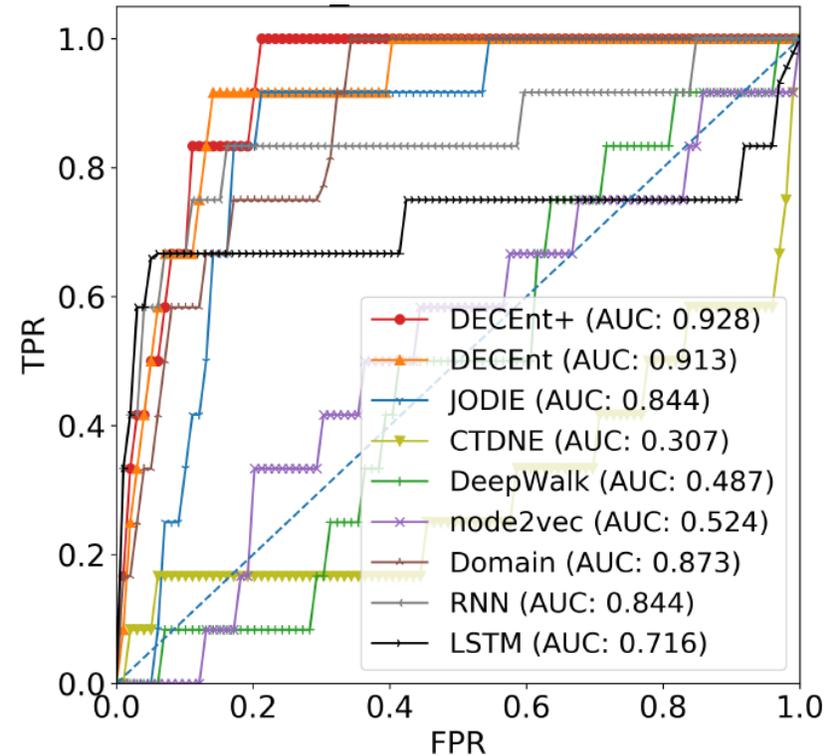


## Reconstruction loss

$$\sum_{(p,d,t) \in \mathcal{PR}} \|\tilde{e}_{d,t-} - [\hat{e}_{d,t-} | \bar{e}_d]\|_2$$

# Paper: Dynamic healthcare embeddings for improving patient care

## MICU transfer prediction



## CDI prediction

Method	CDI
RNN	0.56 (0.119)
LSTM	0.585 (0.103)
DOMAIN	0.655 (0.123)
DEEPWALK	0.494 (0.087)
NODE2VEC	0.453 (0.098)
CTDNE	0.463 (0.101)
JODIE	0.552 (0.192)
DECENT	0.732 (0.069)
DECENT +	<b>0.736 (0.064)</b>

<sup>a</sup>The value in bold denotes best performance

## Mortality and severity prediction

Method	Mortality	Severity
RNN	0.276 (0.039)	0.31 (0.032)
LSTM	0.289 (0.033)	0.308 (0.026)
DOMAIN	0.22 (0.017)	0.258 (0.007)
DEEPWALK	0.172 (0.034)	0.192 (0.019)
NODE2VEC	0.172 (0.02)	0.196 (0.009)
CTDNE	0.184 (0.019)	0.199 (0.007)
JODIE	0.143 (0.039)	0.193 (0.014)
DECENT	0.421 (0.027)	0.34 (0.014)
DECENT+	<b>0.428 (0.022)</b>	<b>0.349 (0.015)</b>

<sup>a</sup>The value in bold denotes best performance