1 Preliminaries

In this section, we introduce basic definitions and preliminaries models and theorem to fully understand our heterogeneous EHR graph model.

1.1 Definition

Definition 1 (Heterogeneous Network). *A heterogeneous network is defined as a graph $\boldsymbol{G} = (V, E, T)$, where each node $\boldsymbol{v}$ and each link $\boldsymbol{e}$ are represented by their mapping functions to a specific node and relation type $\phi(v) : V \to T_V$ and $\phi(e) : E \to T_E$. Where T_V and T_E denote the sets of node and relation types, and $|T_V| + |T_E| > 2$.*

Definition 2 (Heterogeneous Graph Learning). *Given a heterogeneous network $\boldsymbol{G}$, the task of heterogeneous graph learning is to learn a function mapping $f : V \to R^d$, that connects disparate type of nodes into a $d-dimensional$ uniform latent representation $X \in R^{|V| \times d}$, and $d \ll |V|$, that are able to capture the structural and semantic relations between them.*

Definition 3 (One-hop Connectivity). *One-hop connectivity in a heterogeneous network is the local pairwise connection between two consecutive vertices, which directly linked by an edge belongs to a relational type.*

1.2 Skip-gram Model

The skip-gram model [**?**] seeks to maximize the probability of observing the context neighborhood nodes given the center node:

$$\max_f \sum_{u \in V} logPr(N_c(u)|f(u)) \tag{1}$$

Where $N_c(u)$ is the neighborhood context nodes of the center node u, and $f(u)$ is the latent representation of u. It is crucial to note that skip-gram model tries to classify closely connected vertices as similar group, while maximizing the dissimilarity between vertices that are far away. This procedure is the essence for our heterogeneous graph embedding model on EHR data, since closely related vertices tends to have similar embedding representations, patients connected by common diagnoses are classified as similar group, so that their information could be shared and encoded in both their embedding representations.

1.3 Heterogeneous Skip-gram Model

EHR data is heterogeneous, including varies type of vertices, such as lab tests, diagnoses, prescriptions, and patient demographics. Each of these vertices encodes different information. Heterogeneous Skip-gram model [**?**] learns the latent expression of these different type of nodes by maximizing the probability of observing heterogeneous neighborhood given a center node:

$$\max \sum_{u \in V} \sum_{t \in T_V} logPr(N_t(u)|f(u)) \tag{2}$$

Where $N_t(u)$ is the heterogeneous neighborhood vertices of center node u, and $t \in T_V$ is the node type. The difference between heterogeneous skip-gram model and skip-gram model is by adding the node type in the summation term of the objective equation 1, so that the logic of the skip-gram model could still apply.

1.4 TransE

Different heterogeneous node types may have different representation dimensions. For example, the total number of diagnosis ICD codes is over 20,000. Lab test codes may contain hundreds to thousands of different items. Therefore, the dimensions of multi-hot initial representation for these diverse type of medical concepts are different. The traditional way for integrating these features is by concatenating them to form a final vector representation [**?**]. In heterogeneous graph learning, the goal is to classify the embedding representation of closely related medical concept together, so the latent representation of these different type of vertices must be the same, and projected into the same latent space. We apply TransE model [**?**] for doing this latent space projection as shown in Figure ... .

Here, we aim to relate different type of nodes by their relationship type. Specifically, two different types of nodes are connected by a relation type would be represented as a triple (head, relation, tail), denoted as (h, l, t). For example, one triple from EHR data could be $(patient, diagnosed, ICD)$, where *patient* is the head node, ICD is the specific diagnosis code attributed to the patient, and the relation between these two vertices is *diagnosed*.

This TransE model leverages the procedure by first projecting different type of node with different initial representation dimension into a same latent dimension space (where the dimension of this latent space can be customized), and these two different type projected nodes are linked by a relation type which is represented as a translation vector in that latent space. Both the projection matrix and the relational translation vector are learnable parameters in the deep learning system.