1 Materials

1.1 Gene Ontology (GO) and Gene Ontology Annotation (GOA)

The Gene Ontology Consortium [5] provides a structured and formal representation of biological knowledge, describing gene and gene product attributes across all eukaryotes, as described by GO and GOA. The consortium defines a dynamic, controlled vocabulary of all biological knowledge to explain gene and protein roles in all eukaryotes. GO terms are classified in three domains: biological process (BP), molecular function (MF), and cellular component (CC). Each GO term has four types of relations to others: is a, part of, has part, or regulates (including positively regulates and negatively regulates). Each GO domain can be represented in DAG format, where nodes represent GO terms and directed edges represent the relation between GO terms. To integrate three domains into one DAG for further steps, we made a fake root (GO:0000000) to connect to the root nodes of three domains. The GOA Database, which is also provided by the Gene Ontology Consortium, provides the knowledge of genes in specific species via GO terms. All annotations include an evidence code to support the relationship between GO term and gene, according to how annotation was gained, such as by an experiment, phylogenetic inference, computational analysis, author statement, electronic annotation, or curator statement.

1.2 Interaction Database

STRING [16] is a database of known and predicted protein-protein interactions. This database has collected interaction information from more than 2,000 species from five main sources: genomic context predictions, high-throughput lab experiments, co-expression, automated text mining, and previous knowledge in databases. In the database, interaction types are classified into 7 types: activation, binding, catalysis, expression, inhibition, post-translational modification, and reaction. The interaction also has a confidence score, which indicates the approximate probability that a predicted link exists (ranging from 0 to 1,000). We chose three species to evaluate the interaction predictors: human (H. sapiens), mouse (M. musculus) and yeast (S. cerevisiae).

HumanNet v2 [6] provides a human gene network for disease research, integrating diverse types of data information, such as protein-protein interaction, co-citation, co-essentiality, co-expression, pathway database, protein domain profile associations, gene neighborhood, phylogenetic profile association, and interlogs from other species, including 17,929 genes and 525,537 links. A link between genes has negative log-likelihood scores (LLSs) that measure the probability of interaction. They provide various versions of the gene-gene network. Among them, we chose to use HumanNet_XN (fully extended network) because it is recommended for studies requiring the most comprehensive networks.

2 Methods

2.1 GO version and processing

Gene Ontology can be downloaded from 'http://geneontology.org/docs/download-ontology/’. For our study, we used go.obo (version 1.2, released/2019-02-27). These files contain the core GO ontology not filtered relationship, including has_part. Therefore, the processed GO corpus contains a small number of cycles in the graph, which makes it impossible to obtain the GO level, so we removed one of part_of or has_part link randomly in these cases.
2.2 Semantic similarity measures

2.2.1 Resnik method [14]

The information content of a GO term is calculated by the negative log probability of its occurrence, which is defined as the frequency of the term in GO corpus. Resnik method is defined as:

\[ IC(t) = -\log(freq(t)) \]  

\[ \text{sim}_{\text{Resnik}}(t_1, t_2) = IC(MICA), \]

where the most informative common ancestor (MICA) of two GO terms denotes their closest common ancestor term.

2.2.2 Wang method [17]

Given a GO term \( A \), the DAG of \( A \) and its ancestors are defined as \( DAG_A = (A, T_A, E_A) \), where \( T_A \) denotes the set of GO terms in \( DAG_A \), and \( E_A \) denotes the set of edges in \( DAG_A \). The S-value of a GO term in \( DAG_A \), which infers the contribution of a GO term \( t \) to the semantics of GO term \( A \), is defined as:

\[ S_A(t) = \begin{cases} 1 & \text{if } t = A \\ \max \{w_e \cdot S(t') | t' \in \text{children}(t)\} & \text{if } t \neq A, \end{cases} \]

where \( w_e \) is the semantic contribution factor for edge \( e \in E_A \) (\( 0 < w_e < 1 \)). Given \( DAG_A \) and \( DAG_B \) for GO terms \( A \) and \( B \) respectively, the semantic similarity between two terms is defined as:

\[ \text{sim}_{\text{Wang}}(A, B) = \frac{\sum_{t \in T_A \cap T_B} (S_A(t) + S_B(t))}{\sum_{t \in T_A} S_A(t) + \sum_{t \in T_B} S_B(t)} \]

2.2.3 GOGO [19]

The similarity of GO terms in GOGO is calculated in the similar way as the Wang method. The additional operation in the methodology is to assign a semantic contribution factor \( w_e \). In GOGO, the semantic contribution factor is defined as:

\[ w_e = 1/(c \cdot nc(t)) + d, \]

where \( nc(t) \) is the total number of children for GO term \( t \), and \( c \) and \( d \) are constant parameters. In practice, \( c \) is assigned 0.67, and \( d \) is assigned as 0.4, 0.3, and 0.2 for \( is\ a, \ part\ of, \ and\ regulates \), respectively.

2.3 Transitive Closure

In graph theory, transitive closure of a graph is a graph which contains an edge \((\text{vertex}_i, \text{vertex}_j)\) whenever there is a directed path from \( \text{vertex}_i \) to \( \text{vertex}_j \) [15]. Therefore, the transitive closure of a directed graph indicates its reachability relation. We will use a concept of \( n\)-step reachable, which means that there is a directed path from \( \text{vertex}_i \) to \( \text{vertex}_j \) within \( n \) edges. In other words, 1-step reachable node means out-neighbors, while the whole step reachable node means the out-neighbors in transitive closure.

2.4 Similarity Measurement

- **Best-Match Average** The semantic similarities between genes can be computed from the idea of mixing the similarity of GO term pairs, such as Average (Avg) [8], Maximum (Max) [7], Average Best-Matches (ABM) [13, 10] and Best-match Average (BMA) [17, 7]. Among them, BMA was reported as the best approach [19]. The formulas of BMA and ABM were used interchangeably in previous studies [17, 13, 7, 3, 10, 11, 12, 4, 19]. We therefore followed the definition in GOSemSim [18]. Given two genes \( \text{gene}_1 \) and \( \text{gene}_2 \) annotated with \( \{GO_{11}, GO_{12}, ..., GO_{1m}\} \) and \( \{GO_{21}, GO_{22}, ..., GO_{2n}\} \) respectively, the BMA similarity defined as:

\[ \text{BMA}(\text{gene}_1, \text{gene}_2) = \frac{\sum_{i=1}^{m} \max_{1 \leq j \leq n} \text{sim}(GO_{1i}, GO_{2j}) + \sum_{j=1}^{n} \max_{1 \leq i \leq m} \text{sim}(GO_{1i}, GO_{2j})}{m + n} \]


• **Cosine similarity** In the embedding methods based on Word2Vec in the Euclidean space, cosine similarity are used for calculating similarity between vectors.

\[
\text{cosine\_similarity}(u, v) = \frac{u \cdot v}{\|u\| \|v\|}
\]  

(7)

• **Poincaré similarity** There is no concept of similarity in hyperbolic space [9]. Instead, we applied a monotonic decreasing function to poincaré distance to invert the rank order of distance, as it would play a role of similarity. The following function was used to our embeddings at the evaluation steps that required the similarity of vectors.

\[
\text{poincaré\_similarity}(u, v) = \frac{1}{1 + d(u, v)}
\]

(8)

where \(d(u, v)\) is a poincaré distance between \(u\) and \(v\).

### 2.5 Evaluation Metrics

• **AUC**, a Receiver operating characteristic (ROC) curve is widely used for evaluating prediction models. It plots True Positive Rate (TPR) against False Positive Rate (FPR).

\[
\text{TPR} = \frac{TP}{TP + FN}
\]  

(9)

\[
\text{FPR} = \frac{FP}{FP + TN}
\]  

(10)

where TP, FP, TN and FN are the number of true positives, false positives, true negatives, and false negatives respectively. The AUC of ROC stands for the area under the ROC curve. This evaluation metric allow us to compare the prediction models more formally and precisely.

• **mRank**, a mean rank (mRank) of DAG is defined as:

\[
N = \sum_i \text{count(nbr}(i)\text{))}
\]

(11)

\[
m\text{Rank} = \frac{1}{N} \sum_i \sum_{j \in \text{nbr}(i)} \left(\text{obs\_rank}_i(j) - \text{exp\_rank}_i(j)\right),
\]

(12)

where \(\text{nbr}(i)\) denotes the out-neighbors of vertex \(i\), and \(\text{obs\_rank}_i(j)\) and \(\text{exp\_rank}_i(j)\) denote an observed and expected distance-based rank of vertex \(j\) respectively from vertex \(i\).

• **mAP**, a mean Average Precision (mAP) of DAG is defined as:

\[
m\text{AP} = \frac{1}{n} \sum_i \text{AP\_score}(i)
\]

(13)

where \(\text{AP\_score}(i)\) denotes an average precision score of the rankings from vertex \(i\).

• **Hamming loss**, in the multilabel classification, hamming loss is defined as a fraction of labels that are incorrectly predicted.

\[
\frac{1}{NL} \sum_{j=1}^{N} \sum_{i=1}^{L} \text{xor}(y_{i,j}, z_{i,j})
\]

(14)

where \(N\) is the number of samples, \(L\) is the number of multiple labels, \(y_{i,j}\) is the target, \(z_{i,j}\) is the prediction, and \(\text{xor}()\) is the exclusive or operator.
2.6 Statistical test for model comparisons

- **Permutation test for comparing ROC curves.** [1],[2] proposed the statistical test to compare ROC curves. It is applied to testing for the ROC curves from the GO link reconstruction and the binary interaction prediction tasks, mainly between the classifier derived from HiG2Vec and the one classifier derived from OPA2Vec. We used one-tailed test, where alternative hypothesis is HiG2Vec’s AUC is higher than other’s, and selected 0.05 as a significant level.

- **Wilcoxon signed-rank test for comparing prediction errors.** non-parametric paired test is applied to testing for the prediction errors from the GO level reconstruction and the interaction score prediction tasks between the predictor derived from HiG2Vec and the one derived from OPA2Vec. We used one-tailed test, where alternative hypothesis is HiG2Vec’s error is lower than other’s, and selected 0.05 as a significant level.
3 Results

Table S1: The number of the matched GO terms with k nearest neighbors for the three GO terms at the top of the hierarchy.

|         | GO:0008150 | GO:0003674 | GO:0005575 | Sum          |
|---------|-------------|-------------|-------------|--------------|
| Human   |             |             |             |              |
| Count   |             |             |             |              |
| # of neighbors (k) | 32 | 19 | 21 | 72 |
| HiG2Vec | 6 (18.75%) | 2 (10.53%) | 4 (19.05%) | 12 (16.67%) |
| Matched GO terms |             |             |             |              |
| within k nearest neighbors |             |             |             |              |
| EL embeddings | 1 (3.13%) | 2 (10.53%) | 4 (19.05%) | 7 (9.72%)   |
| OPA2Vec  | 1 (3.13%) | 1 (5.26%) | 4 (19.05%) | 6 (8.33%)   |
|Onto2Vec | 3 (9.38%) | 0 (0%)     | 0 (0%)     | 3 (4.17%)   |
| Mouse   |             |             |             |              |
| Count   |             |             |             |              |
| # of neighbors (k) | 32 | 19 | 21 | 72 |
| HiG2Vec | 7 (21.88%) | 2 (10.53%) | 4 (19.05%) | 13 (18.06%) |
| Matched GO terms |             |             |             |              |
| within k nearest neighbors |             |             |             |              |
| OPA2Vec  | 1 (3.13%) | 2 (10.53%) | 4 (19.05%) | 7 (9.72%)   |
|Onto2Vec | 2 (6.25%) | 0 (0%)     | 0 (0%)     | 2 (2.78%)   |
| Yeast    |             |             |             |              |
| Count   |             |             |             |              |
| # of neighbors (k) | 32 | 19 | 21 | 72 |
| HiG2Vec | 8 (25.00%) | 2 (10.53%) | 4 (19.05%) | 14 (19.44%) |
| Matched GO terms |             |             |             |              |
| within k nearest neighbors |             |             |             |              |
| EL embeddings | 2 (6.25%) | 3 (15.79%) | 8 (38.10%) | 13 (18.06%) |
| OPA2Vec  | 0 (0%)     | 2 (10.53%) | 3 (14.29%) | 5 (6.94%)   |
|Onto2Vec | 0 (0%)     | 0 (0%)     | 0 (0%)     | 1 (1.39%)   |

A. GO embedding

![GO embedding](image)

B. Gene embedding

![Gene embedding](image)

Figure S1: HiG2Vec’s GO embedding and gene embedding on the 2-dimensional space. (A) shows how the location of GO terms are changed until converged in the Poincaré ball, where the fake root is blue and the three root GO terms (GO:0008150, GO:0003674, and GO:0005575) are red. (B) shows the final location of GO terms and genes in the space after gene embedding for each species.
Table S2: Results of GO link reconstruction for GO embeddings with dimensionality changes

| Dim | Cross-validation | Fully observed data |
|-----|------------------|---------------------|
|     | BP               | MF                  | CC      | All    | BP               | MF                  | CC      | All    |

**Human**

- **HiG2Vec**
  - **AUC**
    - 10: 0.8408, 0.7285, 0.7015, 0.8224
    - 20: 0.8399, 0.7316, 0.6964, 0.8219
    - 50: 0.8384, 0.7339, 0.6954, 0.8203
    - 100: 0.8382, 0.7332, 0.6934, 0.8200
    - 200: 0.8344, 0.7294, 0.6907, 0.8162
    - 1,000: 0.8247, 0.7192, 0.6849, 0.8067

- **OPA2Vec**
  - 200: 0.6393, 0.4687, 0.5378, 0.6226
  - 20: 0.8547, 0.7519, 0.6844, 0.8132
  - 1,000: 0.8271, 0.7101, 0.6815, 0.8076

- **Onto2Vec**
  - 200: 0.6410, 0.4758, 0.5359, 0.6268
  - 20: 0.8547, 0.7519, 0.6844, 0.8132
  - 1,000: 0.8271, 0.7101, 0.6815, 0.8076

**Mouse**

- **HiG2Vec**
  - **AUC**
    - 10: 0.8410, 0.7179, 0.6957, 0.8214
    - 20: 0.8402, 0.7222, 0.6910, 0.8210
    - 50: 0.8390, 0.7249, 0.6907, 0.8196
    - 100: 0.8400, 0.7229, 0.6898, 0.8204
    - 200: 0.8369, 0.7185, 0.6864, 0.8170
    - 500: 0.8332, 0.7135, 0.6844, 0.8132
    - 1,000: 0.8271, 0.7101, 0.6815, 0.8076

- **OPA2Vec**
  - 200: 0.6443, 0.4758, 0.5359, 0.6268
  - 20: 0.8547, 0.7519, 0.6844, 0.8132
  - 1,000: 0.8271, 0.7101, 0.6815, 0.8076

**Yeast**

- **HiG2Vec**
  - **AUC**
    - 10: 0.8545, 0.7574, 0.6667, 0.8364
    - 20: 0.8538, 0.7550, 0.6624, 0.8357
    - 50: 0.8547, 0.7620, 0.6629, 0.8366
    - 100: 0.8548, 0.7617, 0.6576, 0.8362
    - 200: 0.8549, 0.7604, 0.5988, 0.7641
    - 500: 0.8490, 0.7508, 0.6538, 0.8299
    - 1,000: 0.8429, 0.7441, 0.6474, 0.8240

- **OPA2Vec**
  - 200: 0.6256, 0.4050, 0.3930, 0.6073
  - 20: 0.8547, 0.7620, 0.6629, 0.8366
  - 1,000: 0.8429, 0.7441, 0.6474, 0.8240

*Bold denotes a significant model by permutation test comparing with OPA2Vec.*
| Dim | GOonly | Human | Mouse | Yeast |
|-----|--------|-------|-------|-------|
|     | mRank | mAP   | mRank | mAP   | mRank | mAP   | mRank | mAP   |
| 10  | 67.68  | 0.6143 | 2422.28 | 0.2868 | 2049.59 | 0.2763 | 453.60 | 0.4282 |
| 20  | 66.03  | 0.6219 | 2384.77 | 0.2966 | 2022.91 | 0.2876 | 446.25 | 0.4374 |
| 50  | 67.78  | 0.6093 | 2362.44 | 0.2895 | 1996.39 | 0.2816 | 441.82 | 0.4283 |
| 100 | 69.02  | 0.6040 | 2308.14 | 0.2832 | 1965.27 | 0.2759 | 430.35 | 0.4211 |
| 200 | 70.40  | 0.6020 | 2281.01 | 0.2804 | 1955.69 | 0.2743 | 425.00 | 0.4181 |
| 500 | 72.75  | 0.6017 | 2297.76 | 0.2811 | 1948.40 | 0.2766 | 434.46 | 0.4192 |
| 1,000 | 75.22 | 0.5992 | 2864.92 | 0.3201 | 2518.79 | 0.3081 | 580.94 | 0.4308 |
| OPA2Vec | 200 | - | - | 14312.74 | 0.0180 | 41417.21 | 0.0114 | 25940.85 | 0.0105 |
| Onto2Vec | 200 | - | - | 19541.39 | 0.0014 | 20989.66 | 0.0014 | 16658.71 | 0.0016 |
| EL embeddings | 50 | - | - | 33624.19 | 0.0007 | - | - | 26120.19 | 0.0009 |
Figure S2: GO link reconstruction for embeddings of mouse. (A) is a ROC curve using entire domains and each (B),(C) and (D) are ROC curves when using only biological process (BP), molecular function (MF) and cellular component (CC) domain respectively.

Figure S3: Hierarchy reconstruction for embeddings of mouse. (A) is log2 transformed mRank and (B) is mAP when reconstructing within n-step reachable nodes.
Figure S4: GO link reconstruction for embeddings of yeast. (A) is a ROC curve using entire domains and each (B),(C) and (D) are ROC curves when using only biological process (BP), molecular function (MF) and cellular component (CC) domain respectively.

Figure S5: Hierarchy reconstruction for embeddings of yeast. (A) is log2 transformed mRank and (B) is mAP when reconstructing within n-step reachable nodes.
Figure S6: GO level reconstruction for every GO embeddings. (A,B) GOonly(using HiG2Vec), (C-G) Human, (H-K) Mouse, (L-P) Yeast
### Table S4: Results of GO level reconstruction for GO embeddings with dimensionality changes

|         | GOonly Human |          |        |         |          |        |        |
|---------|--------------|----------|-------|---------|----------|-------|-------|
| Dim     | R-squared    | RMSE     | R-squared | RMSE    |          |       |       |
| 10      | 0.2014       | 2.5804   | 0.2161 | 2.5566 (p=1.000) |
| 20      | 0.3411       | 2.3440   | 0.3411 | 2.3439 (p=1.000) |
| 50      | 0.4443       | 2.1526   | 0.4301 | 2.1800 (p=3.650e-86) |
| HiG2Vec | 100          | 0.4700   | 2.1022 | 0.4484 | 2.1447 (p=8.558e-132) |
|         | 200          | 0.4713   | 2.0997 | 0.4468 | 2.1478 (p=4.830e-148) |
|         | 500          | 0.4898   | 2.0626 | 0.4677 | 2.1069 (p=1.557e-257) |
|         | 1000         | 0.4948   | 2.0525 | 0.4826 | 2.0771 (p=6.642e-265) |
| OPA2Vec | 200          | -        | -     | 0.3224 | 2.3865  |
| Onto2Vec| 200          | -        | -     | 0.1881 | 2.6029  |
| EL embeddings | 50 | -        | -     | 0.2069 | 2.5716  |

### Table S5: A detail of neural network architectures of interaction prediction.

| Layer type | Size of output | Remarks |
|------------|----------------|---------|
| Input      | 2 * d          | Concatenation of two embedding vectors |
| Layer 1    |                |         |
| Fully Connected | d   |         |
| BatchNorm  | d             |         |
| ReLU       | d             |         |
| Layer 2    |                |         |
| Fully Connected | GO level reconstruction: 1 | d → 1 |
|             | Binary prediction: 1 | d → 1 |
|             | Score prediction: 1  | d → 1 |
|             | Type prediction: count(types) | d → count(types) |

* Bold denotes a significant model by Wilcoxon signed-ranked test comparing with OPA2Vec

*d denotes a dimensionality of embeddings
Table S6: Results of binary interaction prediction using the semantic similarity measures with BMA approach

| Model | Domain | STRING_Human | STRING_Mouse | STRING_Yeast | HumanNet_XN |
|-------|--------|---------------|--------------|--------------|-------------|
| Resnik | BP | 0.8234 | **0.7329** | 0.8189 | 0.7799 |
| | MF | 0.6379 | 0.6636 | 0.6531 | 0.6840 |
| | CC | 0.7306 | 0.7114 | 0.8129 | 0.7194 |
| Wang | BP | 0.8244 | 0.7242 | 0.8151 | 0.7799 |
| | MF | 0.5871 | 0.6297 | 0.6714 | 0.6232 |
| | CC | 0.7295 | 0.6952 | **0.8491** | 0.7122 |
| GOGO | BP | **0.8483** | 0.7224 | 0.8252 | **0.7950** |
| | MF | 0.4967 | 0.6331 | 0.7036 | 0.4984 |
| | CC | 0.4979 | 0.7066 | 0.8439 | 0.4894 |

Table S7: STRING binary interaction prediction of the embedding methods

| Embedding | STRING_Human | HumanNet_XN |
|-----------|--------------|-------------|
| Dim | BMA | DIST | NN | BMA | DIST | NN |
| HiG2Vec | 10 | 0.8049 | 0.6923 | 0.8145 (p=1.000) | 0.7811 | 0.7025 | 0.7795 (p=1.000) |
| | 20 | 0.8047 | 0.6894 | 0.8638 (p=1.000) | 0.7808 | 0.6947 | 0.8179 (p=1.000) |
| | 50 | 0.8032 | 0.6936 | 0.9167 (p=1.000) | 0.7797 | 0.6959 | 0.8555 (p=1.000) |
| | 100 | 0.8035 | 0.6946 | 0.9492 (p=1.000) | 0.7802 | 0.6981 | 0.8761 (p=1.000) |
| | 200 | 0.8030 | 0.6930 | 0.9683 (p=1.000) | 0.7802 | 0.6976 | 0.8892 (p=1.000) |
| | 500 | 0.8033 | 0.6936 | **0.9806 (p=0.018)** | 0.7803 | 0.6966 | 0.8989 (p=1.000) |
| | 1,000 | 0.8114 | 0.6902 | **0.9837 (p=0.000)** | 0.7855 | 0.6938 | 0.9050 (p=1.000) |
| OPA2Vec | 200 | 0.8426 | 0.7550 | 0.9796 | 0.8219 | 0.7743 | 0.9086 |
| Onto2Vec | 200 | 0.6418 | 0.5218 | 0.7980 | 0.6502 | 0.5422 | 0.6482 |
| Gene2Vec | 200 | - | 0.6138 | 0.8999 | - | 0.6476 | 0.7837 |
| EL embeddings | 50 | 0.7860 | 0.9116 | 0.9713 | 0.7539 | 0.7912 | 0.8480 |

| Embedding | STRING_Mouse | STRING_Yeast |
|-----------|--------------|--------------|
| Dim | BMA | DIST | NN | BMA | DIST | NN |
| HiG2Vec | 10 | 0.7660 | 0.6541 | 0.8008 (p=1.000) | 0.8673 | 0.7443 | 0.8270 (p=1.000) |
| | 20 | 0.7664 | 0.6521 | 0.8506 (p=1.000) | 0.8663 | 0.7488 | 0.8908 (p=1.000) |
| | 50 | 0.7654 | 0.6555 | 0.9113 (p=1.000) | 0.8665 | 0.7503 | 0.9393 (p=1.000) |
| | 100 | 0.7664 | 0.6558 | 0.9449 (p=1.000) | 0.8666 | 0.7560 | 0.9645 (p=1.000) |
| | 200 | 0.7667 | 0.6550 | 0.9686 (p=1.000) | 0.8666 | 0.7564 | 0.9760 (p=1.000) |
| | 500 | 0.7667 | 0.6554 | **0.9827 (p=0.000)** | 0.8663 | 0.7554 | 0.9789 (p=1.000) |
| | 1,000 | 0.7711 | 0.6524 | **0.9846 (p=0.000)** | 0.8662 | 0.7531 | 0.9812 (p=1.000) |
| OPA2Vec | 200 | 0.7884 | 0.7860 | 0.9807 | 0.8732 | 0.7958 | 0.9871 |
| Onto2Vec | 200 | 0.6128 | 0.5159 | 0.7717 | 0.7760 | 0.5150 | 0.8546 |
| EL embeddings | 50 | - | - | - | 0.8351 | 0.9232 | 0.9686 |

* Bold denotes a significant model by the permutation test comparing with OPA2Vec
Table S8: STRING interaction score prediction of the embedding methods

| Dim  | HiG2Vec       | HumanNet_XN       | OPA2Vec       | Onto2Vec       | Gene2Vec       | EL embeddings |
|------|---------------|-------------------|---------------|----------------|----------------|---------------|
|      | R-squared     | RMSE  | R-squared     | RMSE  | R-squared     | RMSE  |
| 10   | 0.0987        | 304.26(p=1.000) | 0.0784        | 0.98(p=1.000) | 0.3616        | 0.82(p=1.033e-79) |
| 20   | 0.1607        | 293.61(p=1.000) | 0.1245        | 0.96(p=1.000) | 0.2376        | 0.90         |
| 50   | 0.2553        | 276.57(p=1.000) | 0.1953        | 0.92(p=1.000) | 0.2136        | 0.91         |
| 100  | 0.3626        | 255.86(p=1.000) | 0.2462        | 0.89(p=1.000) | 0.3084        | 0.86         |
| 200  | 0.4822        | 230.62(p=1.000) | 0.2981        | 0.86(p=1.000) | 0.3054        | 0.86         |
| 500  | 0.6137        | 199.19(p=0.000) | 0.3616        | 0.82(p=0.173) | 0.3812        | 0.83         |
| 1000 | 0.6555        | 185.36(p=0.000) | 0.3941        | 0.80(p=1.033e-79) | 0.3616 | 0.82 |
|      | OPA2Vec       | 200 | 0.5272        | 220.29        | 0.3672        | 0.82         |
|      | Onto2Vec      | 200 | 0.2977        | 268.57        | 0.1527        | 0.95         |
|      | Gene2Vec      | 200 | 0.4964        | 227.35        | 0.2376        | 0.90         |
|      | EL embeddings | 50  | 0.3881        | 249.19        | 0.2136        | 0.91         |

* Bold denotes a significant model by Wilcoxon signed-ranked test comparing with OPA2Vec

Table S9: STRING interaction type prediction of the embedding methods

| Dim  | HiG2Vec       | OPA2Vec       | Onto2Vec       | Gene2Vec       | EL embeddings |
|------|---------------|---------------|----------------|----------------|---------------|
|      | Loss Macro F1 Micro F1 | Loss Macro F1 Micro F1 | Loss Macro F1 Micro F1 | Loss Macro F1 Micro F1 |
| 10   | 0.1361 0.4739 0.7701 | 0.1181 0.8027 0.3545 | 0.1008 0.4371 0.7780 | 0.0362 0.5616 0.9250 |
| 20   | 0.1205 0.5251 0.7931 | 0.1033 0.4236 0.8271 | 0.0794 0.4862 0.8377 | 0.0362 0.5616 0.9250 |
| 50   | 0.0928 0.5611 0.8379 | 0.0771 0.4669 0.8685 | 0.0428 0.5303 0.9123 | 0.0362 0.5616 0.9250 |
| 100  | 0.0698 0.6371 0.8768 | 0.0544 0.5540 0.9008 | 0.0273 0.6050 0.9435 | 0.0362 0.5616 0.9250 |
| 200  | 0.0481 0.7120 0.9141 | 0.0363 0.6169 0.9370 | 0.0223 0.6299 0.9539 | 0.0362 0.5616 0.9250 |
| 500  | 0.0320 0.7632 0.9425 | 0.0235 0.6633 0.9588 | 0.0208 0.6656 0.9571 | 0.0362 0.5616 0.9250 |
| 1000 | 0.0254 0.7941 0.9543 | 0.0172 0.6882 0.9697 | 0.0118 0.8360 0.9756 | 0.0362 0.5616 0.9250 |
|      | OPA2Vec       | 200 | 0.0373 0.7401 0.9330 | 0.0268 0.6339 0.9532 | 0.0164 0.6780 0.9661 |
|      | Onto2Vec      | 200 | 0.0624 0.6819 0.8896 | 0.0522 0.5711 0.9102 | 0.0390 0.5895 0.9193 |
|      | Gene2Vec      | 200 | 0.0508 0.7011 0.9100 | - - - | - - - |
|      | EL embeddings | 50  | 0.0866 0.5803 0.8456 | - - - | 0.0362 0.5616 0.9250 |
References

[1] A. I. Bandos, H. E. Rockette, and D. Gur. A permutation test sensitive to differences in areas for comparing roc curves from a paired design. *Statistics in Medicine*, 24(18):2873–2893, 2005.

[2] T. M. Braun and T. A. Alonzo. A modified sign test for comparing paired ROC curves. *Biostatistics*, 9(2):364–372, 10 2007.

[3] F. M. Couto and M. J. Silva. Disjunctive shared information between ontology concepts : application to Gene Ontology. pages 1–16, 2011.

[4] R. Ehsani and F. Drabløs. TopoICSim : a new semantic similarity measure based on gene ontology. *BMC Bioinformatics*, pages 1–14, 2016.

[5] T. Gene and O. Consortium. Gene Ontology : tool for the. 25(may):25–29, 2000.

[6] S. Hwang, C. Y. Kim, S. Yang, E. Kim, T. Hart, E. M. Marcotte, and I. Lee. HumanNet v2: human gene networks for disease research. *Nucleic Acids Research*, 47(D1):D573–D580, 11 2018.

[7] S. Jain and G. D. Bader. An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology. 2010.

[8] P. W. Lord, R. D. Stevens, A. Brass, and C. A. Goble. Investigating semantic similarity measures across the gene ontology: The relationship between sequence and annotation. *Bioinformatics*, 19(10):1275–1283, 2003.

[9] M. Margenstern. *Cellular Automata in Hyperbolic Spaces: implementation and computations. Vol. 2*. Advances in unconventional computing and cellular automata series. Old City Publishing, 2008.

[10] G. K. Mazandu and N. J. Mulder. A Topology-Based Metric for Measuring Term Similarity in the Gene Ontology. *Advances in Bioinformatics*, 2012:1–17, 2012.

[11] G. K. Mazandu and N. J. Mulder. Information Content-Based Gene Ontology Semantic Similarity Approaches : Toward a Unified Framework Theory. 2013, 2013.

[12] A. Nagar and H. Al-Mubaid. A hybrid semantic similarity measure for gene ontology based on offspring and path length. *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2015*, (lc):1–7, 2015.

[13] C. Pesquita, D. Faria, H. Bastos, A. E. N. Ferreira, A. O. Falcão, and F. M. Couto. Metrics for GO based protein semantic similarity : a systematic evaluation. 16:1–16, 2008.

[14] P. Resnik. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research*, 11:95–130, 1999.

[15] S. Skiena. *Implementing discrete mathematics - combinatorics and graph theory with Mathematica*. Addison-Wesley, 1990.

[16] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. v. Mering. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1):D607–D613, 11 2018.

[17] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen. A new method to measure the semantic similarity of GO terms. *Bioinformatics*, 23(10):1274–1281, may 2007.

[18] G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang. GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 26(7):976–978, 2010.

[19] C. Zhao and Z. Wang. GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Scientific Reports*, 8(1):1–10, 2018.