Generating new drug repurposing hypotheses using disease-specific hypergraphs

Ayush Jain,[1,2,†] Marie-Laure Charpignon,[1] Irene Y. Chen,[3,4] Anthony Philippakis[1], Ahmed Alaa [3,4]
[1] *Broad Institute of MIT and Harvard, Cambridge, MA, USA*
[2] *Duke University, Durham, NC, USA*
[3] *UC Berkeley, Berkeley, CA, USA*
[4] *UC San Francisco, San Francisco, CA, USA*
[†] *Corresponding e-mail: a.jain@duke.edu*

The drug development pipeline for a new compound can last 10-20 years and cost over $10 billion. Drug repurposing offers a more time- and cost-effective alternative. Computational approaches based on network graph representations, comprising a mixture of disease nodes and their interactions, have recently yielded new drug repurposing hypotheses, including suitable candidates for COVID-19. However, these interactomes remain aggregate by design and often lack disease specificity. This dilution of information may affect the relevance of drug node embeddings to a particular disease, the resulting drug-disease and drug-drug similarity scores, and therefore our ability to identify new targets or drug synergies. To address this problem, we propose constructing and learning disease-specific hypergraphs in which hyperedges encode biological pathways of various lengths. We use a modified node2vec algorithm to generate pathway embeddings. We evaluate our hypergraph's ability to find repurposing targets for an incurable but prevalent disease, Alzheimer's disease (AD), and compare our ranked-ordered recommendations to those derived from a state-of-the-art knowledge graph, the multiscale interactome. Using our method, we successfully identified 7 promising repurposing candidates for AD that were ranked as unlikely repurposing targets by the multiscale interactome but for which the existing literature provides supporting evidence. Additionally, our drug repositioning suggestions are accompanied by explanations, eliciting plausible biological pathways. In the future, we plan on scaling our proposed method to 800+ diseases, combining single-disease hypergraphs into multi-disease hypergraphs to account for subpopulations with risk factors or encode a given patient's comorbidities to formulate personalized repurposing recommendations.

Supplementary materials and code: `https://github.com/ayujain04/psb_supplement`

Keywords: Hypergraphs, Precision Medicine, Drug Repurposing, Disease Specificity

1. Introduction

The development of new drugs can take more than 15 years, from the discovery and pre-clinical phase to review by regulatory agencies.[1] Hence, repurposing drugs previously approved by the Food and Drug Administration or European Medicines Agency serves as a convenient alternative since they are already known to be safe in human populations. From a research

and development perspective, drug repurposing is a less risky enterprise. Indeed, following compound identification, repositioned drugs would generally hit the market in less than 10 years. Beyond time savings, this strategy brings significant cost savings, potentially reducing the average pharmaceutical pipeline's budget by over $5 billion compared to traditional drug development. To date, drug repurposing encompasses three main approaches: computational biomedicine,[2] biological experimentation, and their combination, e.g., through systems pharmacology.[3]

Computational approaches are both more time-effective and cost-effective than *in vitro* or *in vivo* biological experiments, which involve high-throughput screening or phenotypic screening based on animal and human models, respectively. Examples of available strategies include signature matching, genome-wide association studies, and the retrospective analysis of real-world clinical information.[4] Their use has been unlocked by the concurrent emergence of technical advances such as biological microarrays and the increase in data accessibility, as illustrated by the rapid growth of electronic health records and biobanks.[5]

Simultaneously, massive genomic databases and cell lines have yielded 20+ high-quality biological and biomedical knowledge graphs (KG) such as SPOKE[6] and PrimeKG[7] and aggregating platforms such as the KG-Hub[8] to ensure that the former can be shared and made interoperable for downstream graph machine learning tasks. Network-based methods for drug repurposing rely on the encoding of interactions between entities (i.e., drugs, diseases, proteins, biological functions) that can be heterogeneous (i.e., inhibition, binding). These representations can help address both predictive (e.g., polypharmacy side effects) and inferential (e.g., reasoning over causal pathways) questions. Prior graph representations such as the multi-scale interactome (MSI)[9] have proved useful in identifying agents that were previously repurposed and in formulating new potential drug repurposing candidates.

However, drug repurposing hypotheses output by algorithms or deep learning models deployed on KGs may appear as "black boxes." Yet structural and/or functional explanations are often desirable and necessary to understand the possible mechanisms of action underlying a predicted relationship between an existing drug and a disease – be it beneficial or detrimental. Further, KGs integrating various data sources are rarely disease-specific. Thus, they may result in overall drug similarities that do not hold for the pathology of interest or in spurious correlations. This concern is especially relevant for neurodegenerative diseases such as AD, given the presence of the blood-brain barrier[10] and differential gene expression levels and patterns in the brain – relative to other tissues – and across brain regions themselves.

Contributions. Hypergraphs have seen success in uncovering relationships in areas like marketing,[11] finance,[12] and computer vision.[13] Building upon this precedent in other disciplines, we propose disease-specific hypergraphs as the basis for data-driven drug repurposing. Importantly, hypergraphs allow encoding relationships among groups of nodes (i.e., hyperedges) rather than pairwise relationships (i.e., edges) only. In our study, hyperedges capture known biological pathways. First, we show that the properties of hypergraphs reflect relative disease complexity. Second, we transform disease-specific hypergraphs into weighted graphs where nodes encode biological pathways and weighted edges relate to the number of entities that

they have in common (e.g., the number of shared genes or proteins). With the intent of encompassing disease specificity, we focus on pathways that start with a drug entity and end with the disease entity of interest, irrespective of their length. Using a modified node2vec[14] algorithm, we learn the disease-specific embeddings of each hyperedge. In particular, we use these low-dimensional representations to find original biological pathways that are highly similar to those whose starting entity is a drug currently prescribed to treat the disease or mitigate its progression. Then, we pool the top k or k% candidate biological pathways for various values of k and analyze the distributions of starting drug entities and middle gene entities. Such prevalences help gain a mechanistic understanding of promising drug classes and targets for repurposing. We illustrate our proposed method in the context of Alzheimer's disease (AD), a multi-factorial disease of aging that still has no cure despite recent progress.[15] We demonstrate that our proposed method outputs candidate biological pathways that are topologically non-obvious, i.e., they do not have any entities in common with the reference pathways involving currently prescribed drugs, besides the end disease entity. To assess the utility and complementary of learning disease-specific pathway embeddings, we contrast these non-obvious suggestions with those of the MSI, compare the corresponding rank orderings, and validate our findings by mining the biomedical literature to find supporting evidence. Our comparative analysis and publication search reveals that certain candidates that were highly ranked (i.e., in the top 10%) by our hypergraph-based learning approach for AD drug repurposing and had supporting evidence in the literature were missed by the MSI (i.e., in the bottom 33% across all drugs). Going forward, our proposed framework can be scaled to derive novel drug repurposing hypotheses for each of the 800+ major diseases currently registered on the KG-Hub[8] (i.e., excluding rare and orphan diseases).

2. Methods

In this section, we describe our proposed approach, which encompasses three main parts: hypergraph construction, pathway/hyperedge representation learning, comparative analysis with the MSI,[9] and mining of the biomedical literature to find supporting evidence.

2.1. *Hypergraph Construction*

We built disease-specific hypergraphs by querying the Hetionet[16] knowledge graph, which comprises 1,522 drugs, 5,734 side effects, and 137 diseases, to extract significant[a] biological pathways connecting each drug present in the KG to the disease of interest. Hetionet is an existing state-of-the-art knowledge graph that incorporates 11 node types (e.g., gene, symptom), allowing for vast heterogeneity in the node composition of "metapaths" going from a compound to a disease, which we sample from to create disease-specific hypergraphs (Figure 1a). We query the Hetionet to retrieve all paths starting at one of the 1,522 drugs and ending at a disease of choice. These paths are further grouped into metapath categories based on the type and order of nodes present within the path. For example, a metapath category could be "drug-gene1-gene2-targetDisease" or "drug-gene2-similarDisease-targetDisease." We

[a]A more detailed definition of significance follows.

use the direct weighted path count and adjusted p-value defined by the Hetionet to quantify the significance of a path, relative to others within its metapath category. We include the top 10% most significant paths within each category to create our induced disease-specific subgraph. We reasoned that selecting only the most significant pathways would help mitigate the resulting number of false positives among drug repurposing candidates. The largest connected component is treated as the subgraph of interest (Figure 1b); other components, generally much smaller in size, are ignored. All existing biological pathways in the resulting subgraph are explicitly unified as hyperedges, creating a disease-specific hypergraph (Figure 1c). Lastly, we transformed our disease-specific hypergraph into a disease-specific graph where the nodes now correspond to the biological pathway hyperedges that originally constituted the hypergraph. Two biological pathway nodes are connected if they share another element in their path besides the start entity (drug) and the end entity (disease). The edge weight is defined by the number of shared elements $\mathbf{w}$, normalized between 0 and 1 using min-max scaling (Figure 1d) to enable comparisons of graph structures across diseases.

2.2. *Biological Pathway Hyperedge Representation Learning*

Given a specific disease of interest, our study aimed to identify biological pathways analogous – in a learned distinct dimensional subspace – to those associated with drugs currently used to treat it. In particular, we conducted a case study on Alzheimer's disease and considered medications prescribed to alleviate the associated symptoms and behavioral complications. We focused primarily on three compounds: donepezil, galantamine, and memantine,[17–19] approved by the FDA in 1996, 2001, and 2003, respectively. Our approach is disease-agnostic and can be readily extended to other diseases than AD, upon the supply of a list of compounds currently used in clinical practice or previously suggested as repurposing candidates and provided access to adequate computing resources.

Our methodology involved initiating a random walk of fixed length L on the transformed, weighted graph G_w delineated in Figure 1 (d), commencing from any of the biological pathway nodes whose first path element was one of the drugs currently prescribed against the disease of interest. We accounted for the presence of weighted edges by sampling neighboring nodes proportionally to the strength of the connection. The random walker began at a selected node, then proceeded iteratively to an adjacent node chosen uniformly at random among possibly duplicated neighbors, and repeated this process for a predetermined number of steps. For each eligible starting node in our weighted graph G_w, a random walk was initiated, with a fixed length set at L=80. Each start node-specific random walk was replicated R=10 times, in light of the vast heterogeneity of node types and the resulting variation in feasible trajectories.

We denote by v_i the position of the random walker at iteration $i = 1$. At each iteration i of the random walk, the probability of transitioning from biological pathway x to biological pathway y is expressed as:

$$P(v_i = y | v_{i-1} = x) = \frac{\text{weight of edge between } x \text{ and } y}{\text{sum of weights of all edges leaving } x} \tag{1}$$

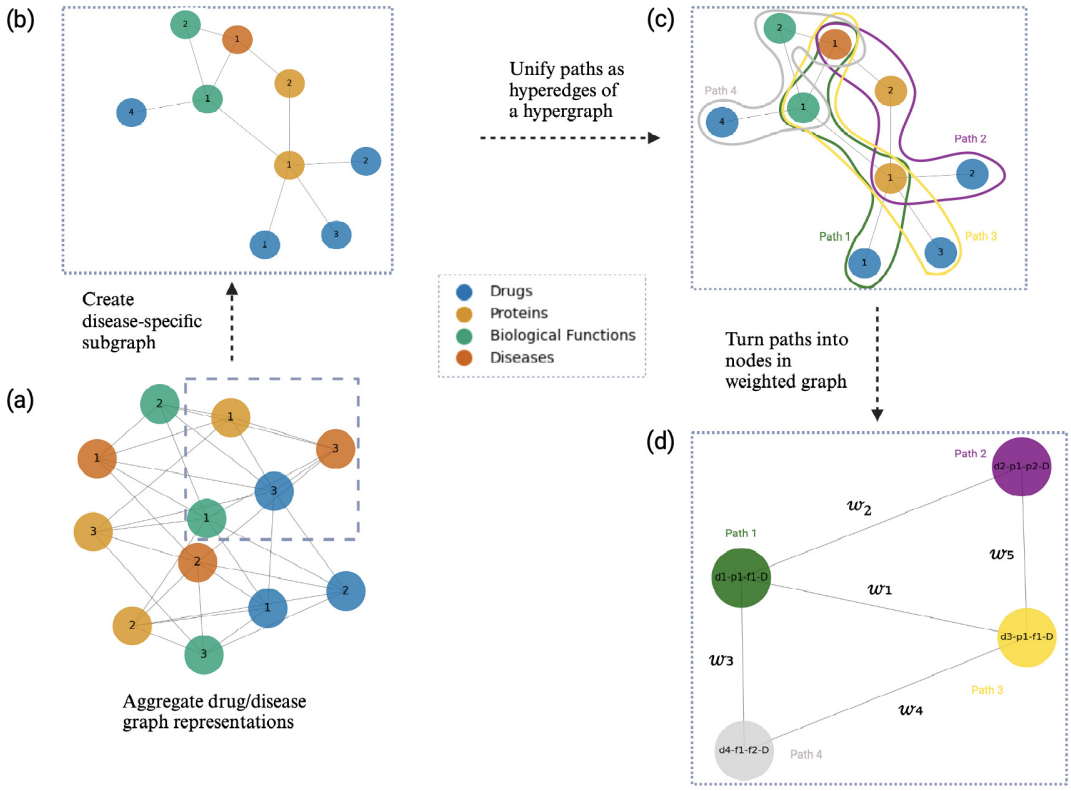


Fig. 1: Pipeline to derive disease-specific hypergraphs from existing KGs and learn contextual embeddings of biological pathways. (a) Full Hetionet graph with nodes of 11 types, including 1,522 drugs, 5,734 side effects, and 137 diseases. (b) Disease-specific subgraph, selecting only the biological pathways whose end node is the disease of interest. Of note, only the top 10% most significant pathways within each metapath category (as defined by their length and structure, e.g., drug-gene1-gene2-disease) are retained, based on a path importance score assigned by Hetionet.[16] (c) Disease-specific hypergraph unifying significant paths or hyperedges into a single structure. (d) Disease-specific weighted graph resulting from the transformation of the hypergraph described in (c). Each hyperedge in (c) becomes a node in (d) and nodes in (d) are connected if their corresponding biological pathways in (c) have at least another element in common, beyond the disease node. Each edge is assigned a weight $\mathbf{w}$, corresponding to the number of elements common to the two biological pathways. Note that the weight does not include the disease node, which all pathways present in a given disease-specific hypergraph intersect at, by design; similarly, the compound node at the start of each reference biological pathway does not contribute to the weight either, continuing our focus on learning biological similarities between the drugs).

2.2.1. *Skip-Gram Model*

We interpreted the resulting random walks as sentences, utilizing the Word2Vec Skip-Gram model provided by gensim to develop node embeddings for each biological pathway.[14] This model predicts context words (nodes within the same walk) given a target word (a node).

Applied to the context of our disease-specific weighted graph, the embeddings of biological pathways learned through this process encapsulate the local neighborhood structure of the nodes and are subsequently used for our pathway similarity search.

The Skip-Gram model's objective is to devise word representations that effectively predict surrounding words in a sentence or document.[20] Formally stated, given a sequence of training words $w_1, w_2, ..., w_T$, the model aims to maximize the average log probability obtained via the chain rule:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-k\leq j\leq k, j\neq 0}\log P(w_{t+j}|w_t) \tag{2}$$

where k denotes the size of the training context and T denotes the total number of training words. In linguistics, k often represents the typical length of a sentence; by analogy, in biology, it could encode the number of reactions occurring in cascade. Similarly, in linguistics, T often represents the size of the vocabulary, which can be language-specific; in biology, the total number of biological pathways involved is disease-specific. To alleviate the fact that pathway length can greatly vary, we guided the model to learn embeddings of fixed dimension $p = 64$ dimensions. We subsequently used cosine similarity as the metric to quantify similarity between any two biological pathways.

Our decision to utilize the Skip-Gram algorithm for learning embeddings was driven by our intent to infer semantic contextual relationships among biological pathways, given a specific disease. To learn the embeddings, we chose the random walk and skip-gram based approach to derive a first proof of concept of using a hypergraph structure and explicitly restricting it to a given disease.

2.3. *Methods for Evaluation*

Our approach to proposing repurposing hypotheses for a given disease of interest relies on identifying the top 10% of biological pathway embeddings having the highest cosine similarity with pathways initiating from any of the drugs known to mitigate or prevent disease progression and ending at the disease node. While pathways can include a variety of intermediary nodes (e.g., symptom, anatomical object, etc.), we selected exclusively those pathways with one or more gene intermediary nodes linking one of the 1,522 drug candidates to the considered disease. We reasoned that this feature would help focus our hypotheses on biologically plausible pathways and thus facilitate the interpretation of drug repurposing candidate rankings.

The data and methods of the multiscale interactome (MSI) were used as a baseline for comparative analysis. The MSI consists in a large biological KG with 1,566 drugs and 841 diseases and leverages a random walk approach to formulate repurposing hypotheses.

From our weighted graphs, we quantified the rank of each biological pathway in terms of its cosine similarity to a selected relevant pathway. We considered relevant pathways to be those whose starting drug entity was a drug currently indicated against the disease of interest. To obtain a single metric per drug, we either aggregated the cosine similarity scores of the pathways in which it was involved into an median value or used the pathway with

the maximum similarity. Then, we used this summary metric to rank all considered drugs and contrast our own repurposing suggestions with those of the MSI. We used these metrics and the relative rankings of psychoanaleptics in the MSI vs. our hypergraph to compare their AUC.

From the MSI, we derived rankings of the most similar drug pairs based on the rankings of the drugs most similar to disease-specific drugs, based on the cosine similarity of their 64-dimensional embeddings. We also established a rank-ordered list of drugs most similar to the disease of interest (e.g., AD), given the cosine similarity between drug and disease embeddings. Notably, while the rank of a drug as derived from the MSI is the output of either a single similarity score (with the disease's embedding) or a couple of scores only (with the embeddings of known drugs against this disease), its rank as derived from our hypergraph-based approach is the output of a much larger double-averaging operation, across all pathways starting at the drug of interest and those starting at a drug already known to target this disease.

We aimed at uncovering any potential blind spots in the MSI that our methodology might successfully uncover. To this end, for drugs that our approach ranked among the top 10%, we retrieved the corresponding MSI-derived rankings for comparison. For each pair, we computed the absolute difference in rank. In addition, we computed an aggregate similarity metric between the MSI and our approach, defined as the size of the overlap between the sets of drugs appearing in the top 10% under each.

To further validate our methodology and the relevance of the resulting pathway embeddings, we undertook a deeper analysis of the drug repurposing suggestions that most differed between the MSI and our proposed method, based on the difference in ranks. In particular, we searched the biomedical literature for biological and/or clinical evidence about drug repurposing suggestions that fell within the bottom third of the MSI's rank-ordered list (i.e., compounds ranked 1,032 to 1,522) while being in the top 10% of ours.

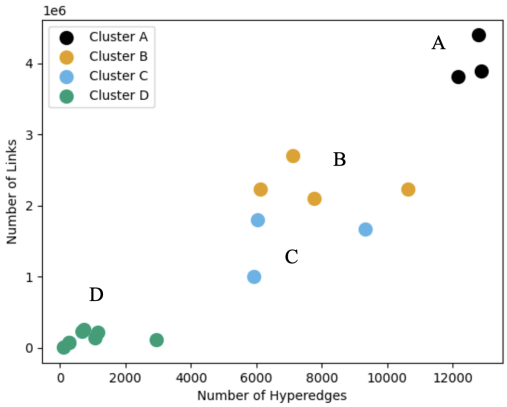
3. Results

We conducted several experiments on our disease-specific hypergraph, using a sample of 18 prevalent and/or incurable diseases. First, we computed summary statistics about these 18 disease hypergraphs and formed clusters that reflect known disease complexity (Section 3.1) Second, we learned disease-specific embeddings of biological pathways on these hypergraphs to identify potential drug targets. To interpret our findings, we explored the distribution of genes involved in the resulting pathways (Section 3.2 and Figure 3). Our intent was to confirm some of the repurposing hypotheses that emerged from the MSI and to formulate new ones. Third, we reviewed the literature to gather information about targets overlooked by the MSI but documented in prior studies; we summarize our findings (Section 3.3). We also evaluated the AUC differences between the MSI and our Alzheimer's disease specific hypergraph (Appendix B in our supplement)

3.1. *Hypergraph Construction Underlines Known Disease Complexity*

Utilizing our method, we constructed 18 disease-specific hypergraphs (Figure 2). For each, we computed the number of hyperedges (biological pathways) and the number of weighted

links among them; we mapped diseases on a scatterplot along these two dimensions. Using the k-means clustering algorithm (k=4), we grouped the 18 hypergraphs into four clusters. In addition to this visual representation, we quantified the number of protein nodes that each disease connects with in the MSI – a proxy to characterize disease complexity. Generally, more complex hypergraphs, with a larger number of hyperedges and links, were those of diseases involving a larger number of proteins. This suggests that the network properties of disease-specific hypergraphs could be leveraged to summarize their complexity and identify diseases that may be proximal based on their higher-order structure.The richness of the embeddings that we seek to learn will depend on the size of the underlying hypergraph; in particular, smaller hypergraphs may yield sparser embeddings. For instance, Figure 2 highlights a clear separation between chronic diseases such as Chronic Kidney Disease (CKD) and Coronary Heart Disease (CHD) and more complex diseases such as Rheumatoid Arthritis (RA) and Amyotrophic Lateral Sclerosis (ALS) involving auto-immune processes. Among the 18 hypergraphs, those of diseases in Cluster A boast more information to learn from and potentially uncover peripheral biological pathways of importance; this configuration prompted us to select one of the diseases in cluster A for our case study. Alzheimer's Disease (AD) was chosen due to its large and growing prevalence, as about 6.2 million Americans aged 65 and older are currently affected – a number which could rise to 13.8 million by 2060.[21]



Cluster ID	Diseases Present	Proteins connected in the MSI
A	Alzheimer's Disease (1), Rheumatoid Arthritis (2), ALS (3)	(1) 73 (2) 162 (3) 49
B	Lupus (4), Psoriasis (5), Asthma (6), Hepatitis B (7)	(4) 60 (5) 52 (6) 92 (7) 6
C	Hypertension (8), Type 2 Diabetes (9), Type 1 Diabetes (10),	(8) 151 (9) 78 (10) 9
D	CHD (11), CKD (12), Diabetic Retinopathy (13), Renal Failure (14), Parkinson's Disease (15), AIDS (16), Atrial Fibrillation (17), Vascular Dementia (18)	(11) 15 (12) 39 (13) 13 (14) 13 (15) 39 (16) 1 (17) 34 (18) N/A

Fig. 2: The scatter plot represents four disease clusters, based on two structural attributes of their respective disease-specific hypergraphs, constructed as outlined in sections 1(b-c). The location of a given cluster indicates the complexity of the higher-order hypergraph structures and often reflects disease complexity. Diseases known to be highly complex (e.g., AD) are positioned in the top right corner; conversely, diseases deemed to be of lower complexity (e.g., renal failure) are situated in the bottom left corner. While disease complexity is primarily defined by the number of currently known biological pathways involved, we also provide the number of proteins to which each disease is directly connected (i.e., one-hop neighbors) in the MSI.[9]

3.2. *Hypergraph Representation Learning Identifies Repurposing Targets in Accordance with the MSI*

Both the MSI and our AD specific hypergraph shared a 50% overlap in drug categories for their repurposing suggestions: psycholeptics, psychoanaleptics, and drugs used in diabetes management, all of which are supported to have repurposing targets to AD in the literature.[22,23] The MSI also brought attention to drugs acting on the renin-angiotensin system, sex hormones, and other nervous system drugs, thereby grouping together common co-morbidity targets for AD treatment.[24–27] In contrast, our AD-specific hypergraph focused more on antineoplastic agents, cardiac therapy, and ophthalmologicals, each of which has been associated with AD in the literature, as well.[28–30]

Figure 3 shows how our hypergraphs can be used to add a layer of explainability to existing knowledge graphs. The figure also compares the number of gene targets that each of the suggestions from both our method and the MSI contained. After finding the top 10% of drugs most similar to AD in the MSI, we found all gene intermediary nodes in the paths starting at these drugs and ending at AD in our hypergraph. We then compared the makeup of these pathways to the makeup of the top 10% most similar pathways to those of donepezil, memantine, or galantamine (ranked by highest cosine similarity to any of the three drugs).

It is important to note the discrepancy in the number of paths considered when generating Figure 3(a)-(g) and Figure 3(h)-(n). The former, involving 926 paths, includes all paths in the AD hypergraph that start with a drug in the top 10% cosine similarity to AD in the MSI. Conversely, the latter, with 574 paths, only encompasses the top 10% of pathways that exhibit the highest cosine similarity to paths initiating from donepezil, memantine, or galantamine.

These findings underscore the efficacy of our disease-specific hypergraph approach in targeting drugs with pathways highly similar to those of known pertinent drugs when identifying potential candidates for repurposing. Moreover, these outcomes provide initial validation to our hypergraph representation learning method, which will be further discussed in the following subsection.

3.3. *Hypergraph Representation Learning Identifies Drug Repurposing Targets Discounted from the MSI but Present in Literature*

Hypergraph representation learning suggested 7 drug repurposing targets out of its top 30 (23%) that the MSI discounted (rank of ≥ 1032 in either column (2) or (3) of Table 1 in the supplement). The 7 drugs were eplerenone (diruetic), fosphenytoin (cardiac therapy), exemestane (endocrine therapy), eperisone (muscle relaxants), protriptyline (psychoanaleptics), ethotoin (antiepileptics), and pentamidine (antiprotozoals). 4 out of 7 (eplernone, pentamidine, exemestane, and protriptyline) of these have literature supporting their potential efficacy against AD. For the remaining 3 out of 7 (eperisone, ethotoin, and fosphenytoin), we explored tangential literature to evaluate the suggestion and/or looked upon the path that this druge headed in hopes of understanding why the prediction was made. Refer to Table 1 in the supplement for the exhaustive list of the top 30 repurposing suggestions based on pathway similarity to donepezil, memantine, or galantamine in the AD hypergraph.

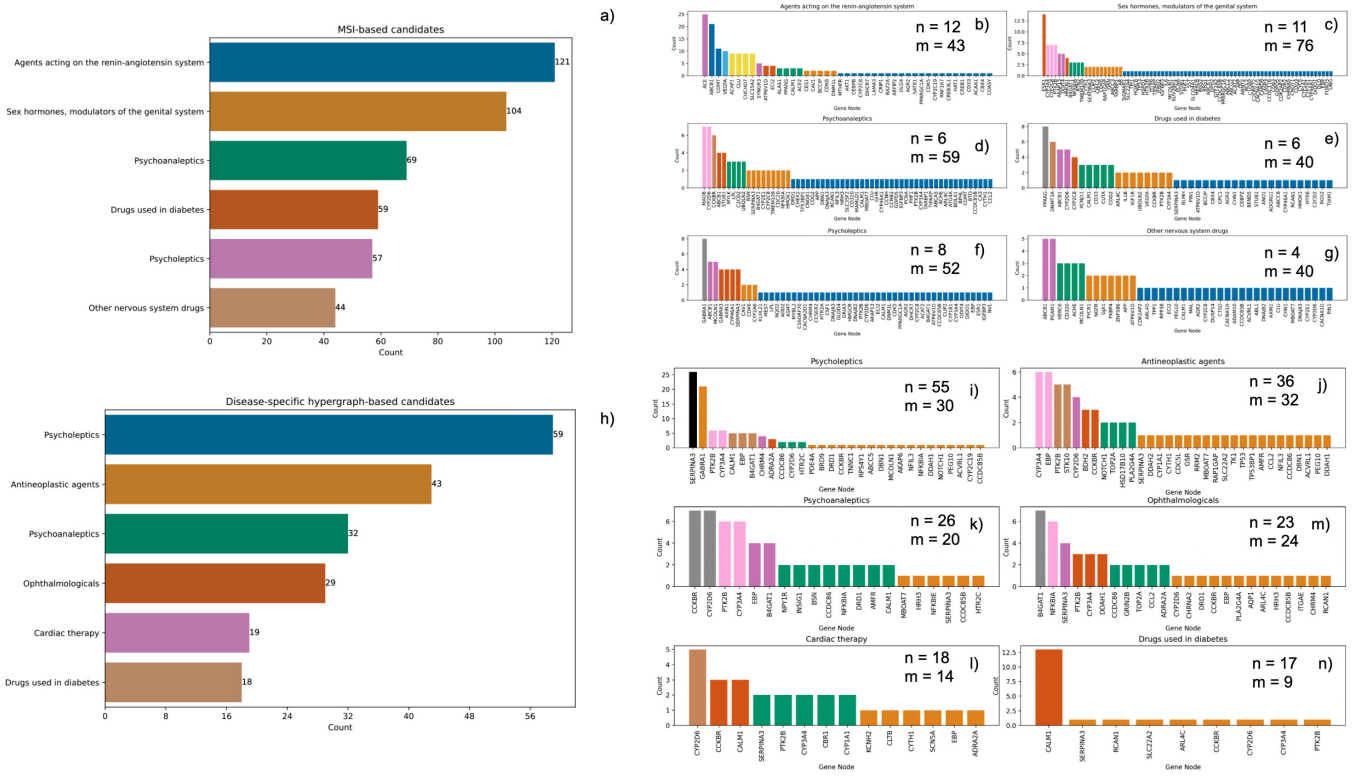


Fig. 3: (a)-(g) illustrate the number of gene targets within the paths of our AD hypergraph, originating from a drug node that ranks within the top 10% in terms of highest cosine similarity to the AD node in the MSI. (h)-(n) depict the number of gene targets in paths within our AD hypergraph that are within the top 10% in similarity to the pathways of donepezil, memantine, and galantamine, only considering paths with gene intermediary nodes.(a) presents the six categories with the most gene targets among the MSI's top 10% suggestions. Conversely, (h) displays the six categories with the most gene targets in the AD hypergraph's top 10% predictions based on similarity to donepezil, memantine, and galantamine. (b)-(g) further break down these top six categories from (a), demonstrating the count of each gene in the top 10% of predicted paths. Similarly, (i)-(n) break down the top six categories from (h), showing the count of each gene within the top 10% of paths similar to those of donepezil, memantine, and galantamine, as determined by cosine similarity. In these graphics, 'n' represents the number of unique drugs within this category, while 'm' signifies the number of unique gene targets in (b)-(g) and (i)-(n).

3.3.1. *Literature Review on 7 Targets Found by Hypegraph Representation Learning but Missed by MSI*

In this section, we delve into the literature that supports our hypotheses for drug repurposing. These drugs were identified as potential repurposing candidates, yet were overlooked by the MSI.

Eplernone has been observed to decrease brain damage, defined by cell death and cortical thinning, in a rat model.[31] Additionally, it is documented that eplernone enhances cognitive

function in a mouse model of AD.[32] Another study reinforces these findings, illustrating that eplernone can mitigate cognitive deficits in the hippocampus of spontaneously hypertensive rats.[33] These outcomes coincide with the established correlation between hypertension and dementia/AD.[34] Lastly, an *in silico* pharmacological assessment of eplernone proposed that the drug holds potential in treating AD.[35]

Exemestane exhibits efficacy when AD patients are concurrently dealing with cancer, suggesting that women diagnosed with breast cancer who underwent treatment with tamoxifen or exemestane exhibited fewer instances of AD.[36] A subsequent study characterizes the relationship between AD and cancer, demonstrating that exemestane is proficient at managing cancer when it co-occurs with AD.[37] Additional research has suggested exemestane as a potential therapeutic for Parkinson's Disease (PD),[38] a neurodegenerative disorder associated with AD.[39]

Protriptyline was found to have the highest inhibitory activity among 140 FDA approved nervous system drugs against the three primary AD targets: AChE, BACE-1, and Aβ aggregation.[40,41] A study using an AD rat model concluded that protriptyline reduces oxidative damage and improves spatial memory in AD mice.[42]

Pentamidine, in a mouse model of AD, was found to inhibit Aβ-induced gliosis and neuroinflammation in AD mice.[43] However, to our knowledge, this is the only publication endorsing its use in alleviating AD, likely because pentamidine is unable to cross the blood-brain barrier. However, recent developments in nose-to-brain methods could surmount this obstacle.[10]

Ethotoin (antiepileptic), to our knowledge, doesn't have explicit literature connecting it to AD. There is, however, a study that warns that antiepileptics could escalate stroke risk in AD patients.[44] This elucidates a current limitation of our approach: we currently do not differentiate between positive and negative drug pathways to a disease.

Fosphenytoin's affect on AD is not specifically discussed in the literature. However, it is a prodrug of phenytoin,[45] which inhibits hippocampal tissue degradation and consequently the progression of AD.[46]

Eperisone lacks direct literature linking it to AD, to the best of our knowledge. When examining the pathway, starting at eperisone and ending at AD that was suggested to have a high similarity to galantamine (see Table 1. in the supplement) in our hypergraph, we find: Eperisone-Triporlidine-CYP2D6-AD. This pathway shows that eperisone was connected to AD by way of similarity to triporlidine. Triporlidine has been observed to enhance NREM sleep in AD patients.[47]

4. Conclusion and Future Directions

Our disease-specific hypergraphs have proven useful for clustering diseases based on their known complexity, identifying potential drug repurposing targets alongside existing methods, and discovering promising repurposing targets overlooked by state-of-the-art methods. We found that the disease hypergraphs formed four clear groups when comparing the number of hyperedges to the number of links between these hyperedges (see Figure 2). Additionally, in Figure 2, we see more complex hypergraphs correlated with more known disease complexity, which we assessed by counting the protein-disease connections in the MSI.

We also demonstrated the value of this method in generating drug repurposing suggestions for Alzheimer's disease (AD). We saw a significant overlap with the suggestions from the MSI when looking at the top 10% of suggestions from both methods, especially among drug categories with the highest number of gene targets in the pathways.

Among our top 30 repurposing suggestions for AD, ranked by pathway cosine similarity, we focused on pathways with one or more protein/gene intermediary nodes, hoping to keep our results grounded in biological relevance. Each suggestion comes with the drug pathway that supports its potential use in treating AD (see Table 1 in the supplement for the full list).

Additionally, our method also identified promising repurposing pathways for AD that the MSI overlooked. In fact, 7 out of our top 30 suggestions ranked in the lower third of the MSI's suggestions. We found supporting evidence for these suggestions in the scientific literature, both from studies that directly tested the drugs and from related research.

Looking ahead, we plan to enhance this method in several ways. We aim to refine our hypergraph construction by merging disease hypergraphs of co-occurring diseases such as AD, Type 2 Diabetes, and Hypertension. We will explore ways to improve our pathway embeddings and, crucially, we will look beyond the literature review to other forms of validation, including evidence from electronic health records and experimental studies.

Future research will explore the use of power iteration, page rank, and page rank with teleportation for learning additional sets of embeddings for each biological pathway-disease pair. We aim to compare the resulting outputs to our current pathway embeddings pairwise and to assess the sensitivity of the downstream similarity scores. To derive more robust drug repurposing candidates for a specific disease, several embeddings of biological pathways could be combined to minimize dependence on a particular algorithm or parameter set and instead maximize confidence, across representation learning approaches. Further, we can experiment with more specific designs of a true positive, perhaps using literature for or against a drug in the context of a disease of interest.

Additionally, we plan on doing more sensitivity analyses upon the comparison metric and learning method, editing parameters like the dimensions of the embedding vector (p), distance of each random walk (L), and amount of random walks taken per each node (R). We also plan on comparing our results to the hypergraphs encompassing more paths. Now that we have outlined a proof-of-concept for this design on the top 10% of paths ending at a disease of interest, we can explore how the suggestions compare to hypergraphs built with the top 20%, 25%, etc.

References

1. H. Xue, J. Li, H. Xie and Y. Wang, Review of Drug Repositioning Approaches and Resources, *International Journal of Biological Sciences* **14**, 1232 (July 2018).
2. T. N. Jarada, J. G. Rokne and R. Alhajj, A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions, *Journal of Cheminformatics* **12**, p. 46 (July 2020).
3. S. Zhao and R. Iyengar, Systems pharmacology: network analysis to identify multiscale mechanisms of drug action, *Annual Review of Pharmacology and Toxicology* **52**, 505 (2012).
4. P. Wu, Q. Feng, V. E. Kerchberger, S. D. Nelson, Q. Chen, B. Li, T. L. Edwards, N. J. Cox, E. J.

Phillips, C. M. Stein, D. M. Roden, J. C. Denny and W.-Q. Wei, Integrating gene expression and clinical data to identify drug repurposing candidates for hyperlipidemia and hypertension, *Nature Communications* **13**, p. 46 (January 2022), Number: 1 Publisher: Nature Publishing Group.
5. G. S. Q. Tan, E. K. Sloan, P. Lambert, C. M. J. Kirkpatrick and J. Ilomäki, Drug repurposing using real-world data, *Drug Discovery Today* **28**, p. 103422 (January 2023).
6. J. H. Morris, K. Soman, R. E. Akbas, X. Zhou, B. Smith, E. C. Meng, C. C. Huang, G. Cerono, G. Schenk, A. Rizk-Jackson, A. Harroud, L. Sanders, S. V. Costes, K. Bharat, A. Chakraborty, A. R. Pico, T. Mardirossian, M. Keiser, A. Tang, J. Hardi, Y. Shi, M. Musen, S. Israni, S. Huang, P. W. Rose, C. A. Nelson and S. E. Baranzini, The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information, *Bioinformatics* **39**, p. btad080 (02 2023).
7. P. Chandak, K. Huang and M. Zitnik, Building a knowledge graph to enable precision medicine, *Scientific Data* **10**, p. 67 (February 2023), Number: 1 Publisher: Nature Publishing Group.
8. KG-Hub—building and exchanging biological knowledge graphs | Bioinformatics | Oxford Academic.
9. C. Ruiz, M. Zitnik and J. Leskovec, Identification of disease treatment mechanisms through the multiscale interactome, *Nature Communications* **12**, p. 1796 (March 2021), Number: 1 Publisher: Nature Publishing Group.
10. F. Rinaldi, L. Seguella, S. Gigli, P. N. Hanieh, E. Del Favero, L. Cantù, M. Pesce, G. Sarnelli, C. Marianecci, G. Esposito and M. Carafa, inPentasomes: An innovative nose-to-brain pentamidine delivery blunts MPTP parkinsonism in mice, *Journal of Controlled Release* **294**, 17 (January 2019).
11. A. M A and A. Rajkumar, Hyper-IMRANK: Ranking-based Influence Maximization for Hypergraphs, in *5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)*, CODS-COMAD 2022 (Association for Computing Machinery, New York, NY, USA, January 2022).
12. X. Ma, T. Zhao, Q. Guo, X. Li and C. Zhang, Fuzzy hypergraph network for recommending top-K profitable stocks, *Information Sciences* **613**, 239 (October 2022).
13. S. Bai, F. Zhang and P. H. S. Torr, Hypergraph convolution and hypergraph attention, *Pattern Recognition* **110**, p. 107637 (February 2021).
14. A. Grover and J. Leskovec, node2vec: Scalable Feature Learning for Networks (July 2016), arXiv:1607.00653 [cs, stat].
15. P. Scheltens, B. D. Strooper, M. Kivipelto, H. Holstege, G. Chételat, C. E. Teunissen, J. Cummings and W. M. v. d. Flier, Alzheimer's disease, *The Lancet* **397**, 1577 (April 2021), Publisher: Elsevier.
16. D. S. Himmelstein, M. Zietz, V. Rubinetti, K. Kloster, B. J. Heil, F. Alquaddoomi, D. Hu, D. N. Nicholson, Y. Hao, B. D. Sullivan, M. W. Nagle and C. S. Greene, Hetnet connectivity search provides rapid insights into how two biomedical entities are related (January 2023), Pages: 2023.01.05.522941 Section: New Results.
17. C.-C. Tan, J.-T. Yu, H.-F. Wang, M.-S. Tan, X.-F. Meng, C. Wang, T. Jiang, X.-C. Zhu and L. Tan, Efficacy and Safety of Donepezil, Galantamine, Rivastigmine, and Memantine for the Treatment of Alzheimer's Disease: A Systematic Review and Meta-Analysis, *Journal of Alzheimer's Disease* **41**, 615 (January 2014), Publisher: IOS Press.
18. M. Bond, G. Rogers, J. Peters, R. Anderson, M. Hoyle, A. Miners, T. Moxham, S. Davis, P. Thokala, A. Wailoo, M. Jeffreys and C. Hyde, The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (review of Technology Appraisal No. 111): a systematic review and economic model., *Health technology assessment (Winchester, England)* **16**, 1 (2012), Number: 21 Publisher: NIHR Journals Library.

19. A. Burns, M. Rossor, J. Hecker, S. Gauthier, H. Petit, H.-J. Möller, S. Rogers and L. Friedhoff, The Effects of Donepezil in Alzheimer's Disease – Results from a Multinational Trial1, *Dementia and Geriatric Cognitive Disorders* **10**, 237 (May 1999).
20. T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, Distributed Representations of Words and Phrases and their Compositionality (October 2013), arXiv:1310.4546 [cs, stat].
21. 2021 Alzheimer's disease facts and figures, *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **17**, 327 (March 2021).
22. M. Citron, Alzheimer's disease: treatments in discovery and development, *Nature Neuroscience* **5**, 1055 (November 2002), Number: 11 Publisher: Nature Publishing Group.
23. M. M. Atef, N. M. El-Sayed, A. A. M. Ahmed and Y. M. Mostafa, Donepezil improves neuropathy through activation of AMPK signalling pathway in streptozotocin-induced diabetic mice, *Biochemical Pharmacology* **159**, 1 (January 2019).
24. R. Loera-Valencia, F. Eroli, S. Garcia-Ptacek and S. Maioli, Brain Renin–Angiotensin System as Novel and Potential Therapeutic Target for Alzheimer's Disease, *International Journal of Molecular Sciences* **22**, p. 10139 (January 2021), Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.
25. IJMS | Free Full-Text | Brain Renin–Angiotensin System as Novel and Potential Therapeutic Target for Alzheimer's Disease.
26. L. Ghiadoni, Management of high blood pressure in type 2 diabetes: perindopril/indapamide fixed-dose combination and the ADVANCE trial [corrected], *Expert Opinion on Pharmacotherapy* **11**, 1647 (July 2010).
27. R. Li, J. Cui and Y. Shen, Brain sex matters: Estrogen in cognition and Alzheimer's disease, *Molecular and Cellular Endocrinology* **389**, 13 (May 2014).
28. M. Dahiya, A. Kumar, M. Yadav, P. Dhakla and S. Tushir, Therapeutic Targeting of Antineoplastic Drugs in Alzheimer's Disease: Discovered in Repurposed Agents, in *Drug Repurposing for Emerging Infectious Diseases and Cancer*, eds. R. C. Sobti, S. K. Lal and R. K. Goyal (Springer Nature, Singapore, 2023) pp. 329–345.
29. L. G. Howes, Cardiovascular Effects of Drugs Used to Treat Alzheimer's Disease, *Drug Safety* **37**, 391 (June 2014).
30. W. A. Fletcher, Ophthalmological aspects of Alzheimer's disease, *Current Opinion in Ophthalmology* **5**, p. 43 (December 1994).
31. X. Wang, Y. Zhu, S. Wang, Z. Wang, H. Sun, Y. He and W. Yao, Effects of eplerenone on cerebral aldosterone levels and brain lesions in spontaneously hypertensive rats, *Clinical and Experimental Hypertension* **42**, 531 (August 2020), Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10641963.2020.1723615.
32. L. Chen, R. Shi, X. She, C. Gu, L. Chong, L. Zhang and R. Li, Mineralocorticoid receptor antagonist-mediated cognitive improvement in a mouse model of Alzheimer's type: possible involvement of BDNF-H2S-Nrf2 signaling, *Fundamental & Clinical Pharmacology* **34**, 697 (2020), _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/fcp.12576.
33. Z. Lin, Y. Lu, S. Li, Y. Li, H. Li, L. Li and L. Wang, Effect of eplerenone on cognitive impairment in spontaneously hypertensive rats, *American Journal of Translational Research* **14**, 3864 (June 2022).
34. S. K. Raina, V. Chander, S. Raina, D. Kumar, A. Grover and A. Bhardwaj, Hypertension and diabetes as risk factors for dementia: A secondary post-hoc analysis from north-west India, *Annals of Indian Academy of Neurology* **18**, 63 (2015).
35. S. Hira, U. Saleem, F. Anwar, Z. Raza, A. U. Rehman and B. Ahmad, In Silico Study and Pharmacological Evaluation of Eplerinone as an Anti-Alzheimer's Drug in STZ-Induced Alzheimer's Disease Model, *ACS Omega* **5**, 13973 (June 2020), Publisher: American Chemical Society.
36. V. Das and M. Hajdúch, Randomizing for Alzheimer's disease drug trials should consider the

cancer history of participants, *Brain* , p. awad177 (May 2023).
37. J. Forés-Martos, C. Boullosa, D. Rodrigo-Domínguez, J. Sánchez-Valle, B. Suay-García, J. Climent, A. Falcó, A. Valencia, J. A. Puig-Butillé, S. Puig and R. Tabarés-Seisdedos, Transcriptomic and Genetic Associations between Alzheimer's Disease, Parkinson's Disease, and Cancer, *Cancers* **13**, p. 2990 (January 2021), Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.
38. H. J. Son, S. H. Han, J. A. Lee, E. J. Shin and O. Hwang, Potential repositioning of exemestane as a neuroprotective agent for Parkinson's disease, *Free Radical Research* **51**, 633 (June 2017), Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10715762.2017.1353688.
39. N. R. Jabir, C. K. Firoz, S. S. Baeesa, G. M. Ashraf, S. Akhtar, W. Kamal, M. A. Kamal and S. Tabrez, Synopsis on the Linkage of Alzheimer's and Parkinson's Disease with Chronic Diseases, *CNS Neuroscience & Therapeutics* **21**, 1 (2015), _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cns.12344.
40. S. B. Bansode, A. K. Jana, K. B. Batkulwar, S. D. Warkad, R. S. Joshi, N. Sengupta and M. J. Kulkarni, Molecular Investigations of Protriptyline as a Multi-Target Directed Ligand in Alzheimer's Disease, *PLOS ONE* **9**, p. e105196 (August 2014), Publisher: Public Library of Science.
41. Y. Wang, H. Wang and H.-z. Chen, AChE Inhibition-based Multi-target-directed Ligands, a Novel Pharmacological Approach for the Symptomatic and Disease-modifying Therapy of Alzheimer's Disease, *Current Neuropharmacology* **14**, 364 (May 2016).
42. V. Tiwari, A. Mishra, S. Singh, S. K. Mishra, K. K. Sahu, Parul, M. J. Kulkarni, R. Shukla and S. Shukla, Protriptyline improves spatial memory and reduces oxidative damage by regulating NFB-BDNF/CREB signaling axis in streptozotocin-induced rat model of Alzheimer's disease, *Brain Research* **1754**, p. 147261 (March 2021).
43. C. Cirillo, E. Capoccia, T. Iuvone, R. Cuomo, G. Sarnelli, L. Steardo and G. Esposito, S100B Inhibitor Pentamidine Attenuates Reactive Gliosis and Reduces Neuronal Loss in a Mouse Model of Alzheimer's Disease, *BioMed Research International* **2015**, p. e508342 (July 2015), Publisher: Hindawi.
44. T. Sarycheva, P. Lavikainen, H. Taipale, J. Tiihonen, A. Tanskanen, S. Hartikainen and A. Tolppanen, Antiepileptic Drug Use and the Risk of Stroke Among Community-Dwelling People With Alzheimer Disease: A Matched Cohort Study, *Journal of the American Heart Association* **7**, p. e009742 (September 2018), Publisher: American Heart Association.
45. S. M. Holliday, P. Benfield and G. L. Plosker, Fosphenytoin. Pharmacoeconomic implications of therapy, *PharmacoEconomics* **14**, 685 (December 1998).
46. V. Dhikav, Can phenytoin prevent Alzheimer's disease?, *Medical Hypotheses* **67**, 725 (2006).
47. A. Satpati, T. Neylan and L. T. Grinberg, Histaminergic neurotransmission in aging and Alzheimer's disease: A review of therapeutic opportunities and gaps, *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **9**, p. e12379 (2023), _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/trc2.12379.