Generating new drug repurposing hypotheses using disease-specific hypergraphs (supplementary materials)

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.

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

1. Appendix

Appendix A. Table of top 10 closest paths to Donepezil, Memantine, or Galantamine

In Table **??** we show the top ten closest pathways to all pathways of donepezil, memantine, or galantamine, resulting in 30 potential repurposing suggestions for AD. We further evaluated

the drugs with the largest difference in ranks in Section 3.3 of manuscript through exhaustive literature review to papers linking the drug to AD.

Appendix B. ROC Curves: Hypergraph vs. MSI

To provide further comparison of how our method of formulating repurposing targets holds against MSI, we evaluated the AUROC of each method. The metric used was either median or max cosine similarity for the hypergraph or cosine similarity of each drug node and the Alzheimer's disease node in the MSI.

The median cosine similarity for each drug was calculated by taking the median of the cosine similarity values of all paths associated with that drug compared to all paths starting with galanatamine, donepezil, and memantine.

In equation form:

$$\text{MedianCosine}(D) = \text{median}\left(\cos(D, P_1), \cos(D, P_2), \ldots, \cos(D, P_n)\right)$$

Where:

- D represents a drug.
- $\cos(D, P_i)$ represents the cosine similarity between drug D and path P_i.
- n is the number of paths associated with drug D.

The max cosine similarity for each drug was calculated by taking the highest cosine similarity value among all paths associated with that drug compared to all paths starting with galanatamine, donepezil, and memantine. This gives us an indication of the most similar path for each drug, which might be important when considering potential drug candidates or understanding drug-path interactions.

In equation form:

$$\text{MaxCosine}(D) = \max\left(\cos(D, P_1), \cos(D, P_2), \ldots, \cos(D, P_n)\right)$$

Where each value is defined the same as the median cosine calculation.

A true positive was defined as the drug category being "psychoanaleptics." Most of the drugs currently prescribed for Alzheimer's disease fall in this category, including galantamine, donepezil, and memantine. The classification came from the Drugbank data provided by the MSI paper to ensure consistency of drug classification.[?] Drugs that were part of any other category were considered false positives.

In figure **??** (a), we see that when using the median cosine similarity, as defined above, for each drug our method performs on par with the MSI in giving higher preference in psychoanaleptics in the repurposing suggestions. Then, in figure **??** (b) we show how our method outperforms the MSI by having a 0.14 higher AUC (0.75 AUC in our method vs. 0.61 in the MSI).

Table A1: For each drug's biological pathway leading to AD, we calculated a cosine similarity with every other drugs' pathway in our AD hypergraph. The top 10 results for each drug are included below in the "Top 10 closest paths" column. The table is structured to show the top 10 results to all donepezil paths first, then galantamine, then memantine. Within each subsection of the table (donepezil, galantamine, or memantine), the rows are listed from highest overall cosine similarity to any donepezil's, memantine's, or galantamine's path (depending on the section) to lowest. (1) shows the rank (by cosine similarity) of the closest path to the individual pathway that was queried; e.g. a rank of 2 means that the closest path ranked 2nd most similar to its corresponding queried path (pathway-to-pathway similarity in our hypergraph). (2) provides the rank of the cosine similarity between donepezil's, galantamine's, or memantine's embedding in the MSI and the drug starting the respective closest path's embedding in the MSI (drug-to-drug similarity in the MSI). (3) provides the rank of the cosine similarity between the drug starting the respective closest path and the AD node embedding in the MSI (drug-to-targetDisease similarity in the MSI)

Queried Paths	Top 10 Closest Paths	Cosine Similarity	(1)	(2)	(3)
Donepezil–NFKBIA–B4GAT1	Rimexolone–NFKBIA–B4GAT1	0.998569	1	620	750
Donepezil–NFKBIA–B4GAT1	Testosterone Propionate–NFKBIA-B4GAT1	0.998527	2	620	830
Donepezil–ARL4C–RCAN1	Gliclazide–ARL4C–RCAN1	0.998504	1	123	945
Donepezil–NFKBIA–B4GAT1	Medroxyprogesterone Acetate–NFKBIA–B4GAT1	0.998501	3	93	343
Donepezil–ARL4C–RCAN1	Dexfenfluramine–ARL4C–RCAN1	0.998491	2	915	115
Donepezil–NFIL3–DBN1	Amcinonide–NFIL3–DBN1	0.998409	1	620	630
Donepezil–NFKBIA–B4GAT1	Methyltestosterone–NFKBIA–B4GAT1	0.998212	4	101	561
Donepezil–NFKBIA–B4GAT1	Pentamidine–B4GAT1	0.998146	5	1091	1410
Donepezil–NFKBIA–B4GAT1	Methylprednisolone–NFKBIA–B4GA	0.998073	6	544	630
Donepezil–NFKBIA–B4GAT1	Fluocinolone Acetonide–NFKBIA	0.997991	7	544	630
Galantamine–MBOAT7–BSN	Alfacalcidol–MBOAT7–BSN	0.997171	1	497	552
Galantamine–MBOAT7–BSN	Protriptyline–MBOAT7–BSN	0.996930	2	1035	510
Galantamine–MBOAT7–BSN	Ethotoin–MBOAT7–BSN	0.995548	3	1035	1041
Galantamine–CYP2D6–CCKBR	Eperisone–Triprolidine–CYP2D6	0.989089	1	1350	1354
Galantamine–CYP2D6–CCKBR	Enzalutamide–Bicalutamide–CYP2D6	0.987268	2	620	603
Galantamine–CYP2D6–CCKBR	Methylnaltrexone–CYP2D6–CCKBR	0.987103	3	759	203
Galantamine–CYP2D6–CCKBR	Ranitidine–CYP2D6–CCKBR	0.986098	4	21	20
Galantamine–CYP2D6–CCKBR	Acebutolol–CYP2D6–CCKBR	0.986035	5	846	972
Galantamine–CYP2D6–CCKBR	Chlordiazepoxide–CYP2D6–CCKBR	0.985522	6	620	93
Galantamine–CYP2D6–CCKBR	TetraBSNazine–CYP2D6–CCKBR	0.985514	7	544	131
Memantine–GRIN1–PTK2B	Desoximetasone–PTK2B	0.999134	1	372	291
Memantine–GRIN1–PTK2B	Telithromycin–CYP3A4–PTK2B	0.999018	2	123	877
Memantine–NFKBIA–B4GAT1	Colchicine–NFKBIA–B4GAT1	0.998943	2	544	561
Memantine–NFKBIA–B4GAT1	Exemestane–NFKBIA–B4GAT1	0.998885	3	1490	1540
Memantine–GRIN1–PTK2B	SIMEPREVIR–CYP3A4–PTK2B	0.998872	3	351	954
Memantine–GRIN1–PTK2B	Riluzole–PTK2B	0.998854	4	759	621
Memantine–GRIN1–PTK2B	Eplerenone–CYP3A4–AMFR	0.998814	5	1490	877
Memantine–GRIN1–PTK2B	Fosphenytoin–CYP3A4–PTK2B	0.998777	6	1431	1177
Memantine–NFKBIA–B4GAT1	Potassium Chloride–SLC12A2–B4GAT1	0.998780	4	846	495
Memantine–NFKBIA–B4GAT1	Medroxyprogesterone Acetate–NFKBIA–B4GAT1	0.998764	5	93	621

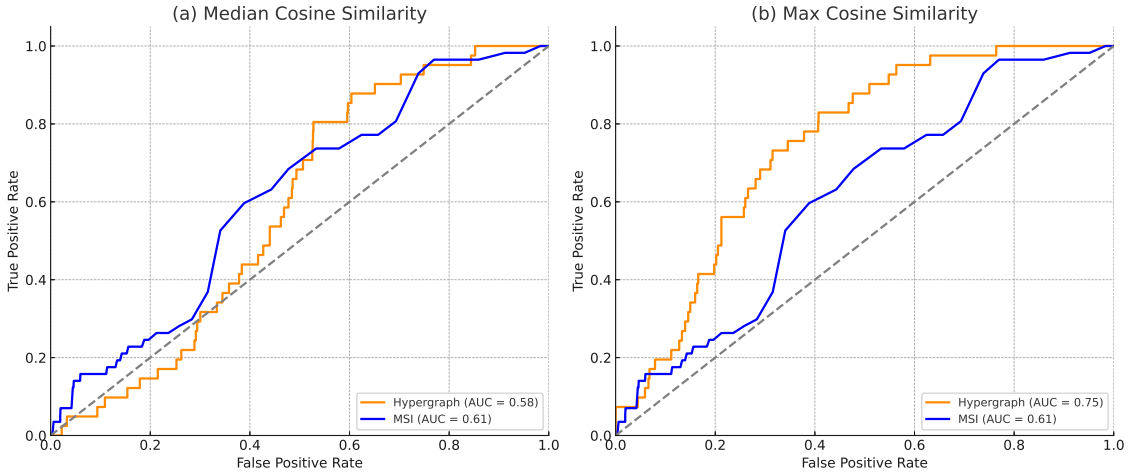

Fig. B1: (a) shows the ROC in orange of the median cosine similarity of a drug's path and galantamine's, donepezil's, or memantine's paths that end in Alzheimer's disease and each drugs cosine similarity to the MSI in blue. In (a) we see that the AUC of the median cosine similarity is roughly on par with that of the MSI. (b) displays the maximum cosine similarity of a drug's path and galantamine's, donepezil's, or memantine's paths that end in Alzheimer's disease in orange and each drugs cosine similarity to the MSI in blue. In (b) we see that when using the maximum cosine similarity, our method outperforms the MSI.