# Hetnet connectivity search provides rapid insights into how two biomedical entities are related

*This manuscript ([permalink](#)) was automatically generated from [greenelab/connectivity-search-manuscript@1638378](#) on November 13, 2019.*

## Authors

> **ⓘ Manuscript in preparation**
> The authorship information below is incomplete and preliminary. This notice will be updated once all contributors meeting <u>authorship criteria</u> have added themselves to `metadata.yaml` .

- **Daniel S. Himmelstein**
  ⓘ [0000-0002-3012-7446](#) · ◯ [dhimmel](#) · 𝕏 [dhimmel](#)
  Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America · Funded by GBMF4552

## Abstract

Hetnets, short for "heterogeneous networks", contain multiple node and relationship types and offer a way to encode biomedical knowledge. For example, Hetionet connects 11 types of nodes — including genes, diseases, drugs, pathways, and anatomical structures — with over 2 million edges of 24 types. Previously, we trained a classifier to repurpose drugs using features extracted from Hetionet. The model identified types of paths between a drug and disease that occurred more frequently between known treatments.

For many applications however, a training set of known relationships does not exist; Yet researchers would still like to know how two nodes are meaningfully connected. For example, users may want to know not only how metformin is related to breast cancer, but also how the GJA1 gene might be involved in insomnia. Therefore, we developed hetnet connectivity search to propose the most important paths between any two nodes.

The algorithm behind connectivity search identifies types of paths that occur more frequently than would be expected by chance (based on node degree alone). We implemented the method on Hetionet and provide an online interface at https://het.io/search. Several optimizations were required to precompute significant instances of node connectivity at scale. We provide an open source implementation of these methods in our new Python package named hetmatpy.

To validate the method, we show that it identifies much of the same evidence for specific instances of drug repurposing as the previous supervised approach, but without requiring a training set.

# Introduction

## Hetionet

## Rephetio

## Unsupervised connectivity search

# Results

## Hetmatpy Package

## DWPC null distribution

## Enriched metapaths

## Enriched paths

## Comparison to Rephetio

## Detecting Mechanisms of Action for Indications

Assess ability to predict paths in [https://github.com/SuLab/DrugMechDB](https://github.com/SuLab/DrugMechDB)

## Connectivity Search Webapp

## Use cases

# Discussion

## Methods

**Computing DWPCs with matrix multiplication**

**Permuted hetnets**

**Degree-grouping of node pairs**

**Gamma-hurdle distribution**

**Prioritizing metapaths for database storage**

**Rest API & backend**

**Webapp & Frontend**

**Realtime open science**

**Software & data availability**

# References