

# Molecular network and polypharmacy Dataset



**Brief description:** The use of multiple drugs, termed polypharmacy, is common to treat patients with complex diseases or co-existing medical conditions. However, a major consequence of polypharmacy is a much higher risk of adverse side effects for the patient. Polypharmacy side effects emerge because of drug-drug interactions, in which activity of one drug may change, favorably or unfavorably, if taken with another drug. The approach constructs a multimodal graph of protein-protein interactions, drug-protein target interactions, and the polypharmacy side effects, which are represented as drug-drug interactions, where each side effect is an edge of a different type.

**Dataset Type:** Nominal and Numerical data.

**Files Format:** csv files and tsv files.

**Number of Files:** 8 files

**Last Access:** 18/5/2023

## ❖ Dataset files

bio-decagon-combo.csv			
Details		Columns	Columns Description
Description	Polypharmacy side effects in the form of (drug A, side effect type, drug B) triples.	<u><b>STITCH 1</b></u>	ID of drug A in STITCH Database Platform.
Number of Rows	4649441 rows		
Number of Columns	4 columns	<u><b>STITCH 2</b></u>	ID of drug B in STITCH Database Platform.
Missing Values	No	<u><b>Polypharmacy Side Effect</b></u>	ID of side effect as identified in Drug Bank Database.
Type	csv	<u><b>Side Effect Name</b></u>	Name of Side Effect related to identified ID.
Source	SNAP Library (Stanford University)		

bio-decagon-effectcategories.csv			
Details		Columns	Columns Description
Description	Side effect categories	<u>Side Effect</u>	ID of polypharmacy side effect as identified in DrugBank database.
Number of Rows	561 rows	<u>Side Effect Name</u>	Name of side effect related to identified ID.
Number of Columns	3 columns		
Missing Values	No	<u>Disease Class</u>	Classification of disease associated with side effect infection.
Type	csv		
Source	SNAP Library (Stanford University)		

bio-decagon-mono.csv			
Details		Columns	Columns Description
Description	Side effects of individual drugs in the form of (drug A, side effect type) tuples.	<u>STITCH</u>	ID of drug in STITCH Database Platform
Number of Rows	174977 rows	<u>Individual Side Effect</u>	ID of polypharmacy side effect as identified in Drug Bank Database.
Number of Columns	3 columns	<u>Side Effect Name</u>	Name of side effect related to identified individual side effect ID.
Missing Values	No		
Type	csv		
Source	SNAP Library (Stanford University)		

bio-decagon-ppi.csv			
Details		Columns	Columns Description
Description	Protein-protein interaction network between gene A and gene B.	<u>Gene 1</u>	ID of gene A at HGNC (HUGO Gene Nomenclature).

Number of Rows	715612 rows	<b><u>Gene 2</u></b>	ID of gene B at HGNC (HUGO Gene Nomenclature).
Number of Columns	2 columns		
Missing Values	No		
Type	csv		
Source	SNAP Library (Stanford University)		

bio-decagon-targets.csv			
Details		Columns	Columns Description
Description	Drug-target protein associations.	<b><u>STITCH</u></b>	ID of drug in STITCH Database Platform.
Number of Rows	18690 rows		
Number of Columns	2 columns		
Missing Values	No	<b><u>Gene</u></b>	ID of gene A at HGNC (HUGO Gene Nomenclature).
Type	csv		
Source	SNAP Library (Stanford University)		

bio-decagon-targets-all.csv			
Details		Columns	Columns Description
Description	Drug-target protein associations culled from several curated databases.	<b><u>STITCH</u></b>	ID of drug in STITCH Database Platform.
Number of Rows	131034 rows		
Number of Columns	2 columns		
Missing Values	No	<b><u>Gene</u></b>	ID of gene A at HGNC (HUGO Gene Nomenclature).
Type	csv		
Source	SNAP Library (Stanford University)		

disease_associations.tsv			
Details		Columns	Columns Description
Description	Diseases associated to genes from DisGeNET	<u>diseaseId</u>	UMLS concept unique identifier
Number of Rows	30170 rows	<u>diseaseName</u>	Name of the disease
Number of Columns	7 columns	<u>diseaseType</u>	The DisGeNET disease type: disease, phenotype and group
Missing Values	Yes	<u>diseaseClass</u>	The MeSH (Medical Subject Headings) disease class(es)
Type	tsv	<u>diseaseSemanticType</u>	The UMLS(Unified Medical Language System) Semantic Type(s) of the disease
Source	DisGeNET	<u>NofGenes</u>	Number of genes associated to the disease
		<u>NofPmids</u>	Number of publications associated to the disease.

gene_associations.tsv			
Details		Columns	Columns Description
Description	Genes associated to Diseases from DisGeNET	<u>geneId</u>	NCBI Entrez Gene Identifier
Number of Rows	21671 rows	<u>geneSymbol</u>	Official Gene Symbol
Number of Columns	9 columns	<u>DSI</u>	The Disease Specificity Index for the gene
Missing Values	Yes	<u>DPI</u>	The Disease Pleiotropy Index for the gene
Type	tsv	<u>PLI</u>	The probability for the gene of being loss-of-function intolerant, provided by the GNOMAD consortium
Source	DisGeNET	<u>protein_class</u>	Protein Class identifier according to the Drug Target Ontology

		<b><u>protein_class_name</u></b>	Protein Class according to the Drug Target Ontology
		<b><u>NofDiseases</u></b>	Number of diseases associated to the gene
		<b><u>NofPmids</u></b>	Number of publications associated to the gene