

# Explain — LinkExplorer 0.1 documentation

## Contents

<b>1</b>	<b>Load</b>	<b>2</b>
1.1	Select dataset . . . . .	2
1.2	Select explanation file . . . . .	2
1.3	Wait for the data to load . . . . .	2
<b>2</b>	<b>Explore</b>	<b>3</b>
2.1	View entities of dataset . . . . .	3
2.2	Filter entites based on type . . . . .	4
2.3	Search based on entity label . . . . .	4
2.4	Get additional information on entities . . . . .	5
<b>3</b>	<b>Explain</b>	<b>6</b>
3.1	Select prediction task . . . . .	6
3.2	View ranked predictions . . . . .	6
3.3	Explore rules that generated a prediction . . . . .	7
3.4	Explore instantiatons of rules with variables . . . . .	9



# 1 Load

## 1.1 Select dataset

The screenshot shows the 'Select a dataset' interface in LinkExplorer. At the top, there is a navigation bar with 'LinkExplorer', 'Overview', 'Entities', 'Feedback', and a 'Load' button. The main content area is titled 'Select a dataset' and includes a 'Load custom endpoint' button. A list of datasets is shown on the left, with 'OpenBioLink 2021' selected. To the right, a description of the OpenBioLink2020 Dataset is provided. At the bottom, there is a search bar and a 'Next' button.

Dataset Name	Version
WN18RR	
YAGO3-10	
<b>OpenBioLink</b>	<b>2021</b>
Hetionet	1.0
PheKnowLator	v2.0.0
None	

The OpenBioLink2020 Dataset is a highly challenging biomedical benchmark dataset containing over 5 million positive and negative edges. The test set does not contain trivially predictable, inverse edges from the training set and does contain all different edge types, to provide a more realistic edge prediction scenario.

## 1.2 Select explanation file

The screenshot shows the 'Select an explanation file' interface in LinkExplorer. At the top, there is a navigation bar with 'LinkExplorer', 'Overview', 'Entities', 'Feedback', and a 'Load' button. The main content area is titled 'Select an explanation file' and includes a 'Load local explanation' button. A list of explanation files is shown on the left, with 'NRNO' selected. To the right, the configuration for the selected file is displayed, including 'Rule learning configuration', 'Clustering configuration', and a 'Comment'. At the bottom, there is a search bar and 'Previous' and 'Done' buttons.

File Name	Method
MAX	Non-redundant Noisy-OR
<b>NRNO</b>	<b>Non-redundant Noisy-OR</b>

**Rule learning configuration**

- SNAPSHOTS\_AT = 1000
- WORKER\_THREADS = 22
- POLICY = 2
- REWARD = 5
- EPSILON = 0.1
- THRESHOLD\_CONFIDENCE = 0.0001
- MAX\_LENGTH\_CYCLIC = 3
- MAX\_LENGTH\_ACYCLIC = 1

**Clustering configuration**

- STRATEGY = random
- ITERATIONS = 10000
- RESOLUTION = 10
- SEED = 0

**Comment**

These results were retrieved by applying the Non-redundant Noisy-Or aggregation method with a random seed to the OpenBioLink dataset based on edge-based

## 1.3 Wait for the data to load

This might take a few seconds depending on the size of the dataset. You will be redirected automatically.

Select an explanation file Load local explanation

MAX	Non-redundant Noisy-OR	<b>Rule learning configuration</b> SNAPSHOTS_AT = 1000 WORKER_THREADS = 22 POLICY = 2 REWARD = 5 EPSILON = 0.1 THRESHOLD_CONFIDENCE = 0.0001 MAX_LENGTH_CYCLIC = 3 MAX_LENGTH_ACYCLIC = 1
NRNO	Non-redundant Noisy-OR	
		<b>Clustering configuration</b> STRATEGY = random ITERATIONS = 10000 RESOLUTION = 10 SEED = 0
		<b>Comment</b> These results were retrieved by applying the Non-redundant Noisy-Or aggregation method with a random

Search

Previous Done

## 2 Explore

### 2.1 View entities of dataset

LinkExplorer Overview Entities Feedback Dataset: obl Explanation: max Load other

All types

< 1 2 3 4 5 ... > >>

PUBCHEM.COMPOUND:53325460	((1S,3R)-3-((3S,4S)-3-Methoxy-tetrahydro-2H-pyran-4-ylamino)-1-(2,2,2-trifluoroethyl)cyclopentyl)(4-(4-(trifluoromethyl)pyridin-2-yl)piperazin-1-yl)methanone
PUBCHEM.COMPOUND:46887993	((2R,4S)-4-(3-Chlorophenoxy)pyrrolidin-2-yl)(4-cyclobutyl-1,4-diazepan-1-yl)methanone
PUBCHEM.COMPOUND:44600841	((2R,4S)-4-(4-Chloro-3-methylphenoxy)pyrrolidin-2-yl)(4-cyclobutyl-1,4-diazepan-1-yl)methanone
PUBCHEM.COMPOUND:44439344	((2S,4S)-4-(4-(4-Nitrophenyl)-1,4-diazepan-1-yl)pyrrolidin-2-yl)(thiazolidin-3-yl)methanone
PUBCHEM.COMPOUND:44439342	((2S,4S)-4-(4-Phenylpiperidin-1-yl)pyrrolidin-2-yl)(thiazolidin-3-yl)methanone
PUBCHEM.COMPOUND:46885523	((3R,4S)-4-(2,4-Difluorophenyl)-1-(pyridazin-3-yl)pyrrolidin-3-yl)((3S,4S,5R)-4-hydroxy-3,5-dimethyl-4-(pyridin-2-yl)piperidin-1-yl)methanone
PUBCHEM.COMPOUND:44442978	((3R,4S)-4-(4-Chlorophenyl)-1-isopropylpyrrolidin-3-yl)(4-(1-(pyridin-3-ylmethylamino)methyl)cyclohexyl)piperazin-1-yl)methanone

## 2.2 Filter entites based on type

LinkExplorer Overview Entities Feedback Dataset: obl Explanation: max Load other

All types Quicksearch

Type to filter...

All types Anatomy Disease Drug GO Gene Pathway Phenotype

OMPOUND:53325460	((1S,3R)-3-((3S,4S)-3-Methoxy-tetrahydro-2H-pyran-4-ylamino)-1-(2,2,2-trifluoroethyl)cyclopentyl)(4-(4-(trifluoromethyl)pyridin-2-yl)piperazin-1-yl)methanone
OMPOUND:46887993	((2R,4S)-4-(3-Chlorophenoxy)pyrrolidin-2-yl)(4-cyclobutyl-1,4-diazepan-1-yl)methanone
OMPOUND:44600841	((2R,4S)-4-(4-Chloro-3-methylphenoxy)pyrrolidin-2-yl)(4-cyclobutyl-1,4-diazepan-1-yl)methanone
OMPOUND:44439344	((2S,4S)-4-(4-(4-Nitrophenyl)-1,4-diazepan-1-yl)pyrrolidin-2-yl)(thiazolidin-3-yl)methanone
PUBCHEM.COMPOUND:44439342	((2S,4S)-4-(4-Phenylpiperidin-1-yl)pyrrolidin-2-yl)(thiazolidin-3-yl)methanone
PUBCHEM.COMPOUND:46885523	((3R,4S)-4-(2,4-Difluorophenyl)-1-(pyridazin-3-yl)pyrrolidin-3-yl)((3S,4S,5R)-4-hydroxy-3,5-dimethyl-4-(pyridin-2-yl)piperidin-1-yl)methanone
PUBCHEM.COMPOUND:44442978	((3R,4S)-4-(4-Chlorophenyl)-1-isopropylpyrrolidin-3-yl)(4-(1-((pyridin-3-yl)methylamino)methyl)cyclohexyl)piperazin-1-yl)methanone

## 2.3 Search based on entity label

LinkExplorer Overview Entities Feedback Dataset: obl Explanation: max Load other

Anatomy liver

UBERON:0013138	coronary ligament of liver
UBERON:0013139	ligament of liver
UBERON:0015796	liver blood vessel
UBERON:0006729	liver perisinusoidal space
UBERON:0016478	liver stroma
UBERON:0002107	liver
UBERON:0004647	liver lobule
UBERON:0001280	liver parenchyma
UBERON:0001116	quadrate lobe of liver

## 2.4 Get additional information on entities

LinkExplorer Overview Entities Feedback
Dataset: obi Explanation: max Load other

Anatomy  
**UBERON:0002107 : liver**

**Description**

An exocrine gland which secretes bile and functions in metabolism of protein and carbohydrate and fat, synthesizes substances involved in the clotting of the blood, synthesizes vitamin A, detoxifies poisonous substances, stores glycogen, and breaks down worn-out erythrocytes[GO].

**Online Resource**

<https://identifiers.org/UBERON:0002107>

**Predictions**

Heads	Tails
? can be underexpressed in liver	liver is a ?
? can be overexpressed in liver	
? is part of liver	
? is expressed in liver	

**Known links**

and explore connections to other entities

Known links	
▼ Outgoing edges	
is a	<a href="#">exocrine gland</a>
is a	<a href="#">abdomen element</a>
is part of	<a href="#">hepatobiliary system</a>
▼ Incoming edges	
<a href="#">ARPIN-AP352 readthrough</a>	is expressed in
<a href="#">NT5C1B-RDH14 readthrough</a>	is expressed in
<a href="#">PHOSPHO2-KLHL23 readthrough</a>	is expressed in
<a href="#">FPGT-TNNI3K readthrough</a>	is expressed in
<a href="#">RPL17-C18orf32 readthrough</a>	is expressed in
<a href="#">TGIF2-RAB5IF readthrough</a>	is expressed in
<a href="#">GIMAP1-GIMAP5 readthrough</a>	is expressed in
<a href="#">PMF1-BGLAP readthrough</a>	is expressed in
<a href="#">TLCD4-RWDD3 readthrough</a>	is expressed in

### 3 Explain

#### 3.1 Select prediction task

The screenshot shows the LinkExplorer interface for the 'Anatomy' prediction task. The main title is 'UBERON:0002107 : liver'. Below this, there is a 'Description' section with text: 'An exocrine gland which secretes bile and functions in metabolism of protein and carbohydrate and fat, synthesizes substances involved in the clotting of the blood, synthesizes vitamin A, detoxifies poisonous substances, stores glycogen, and breaks down worn-out erythrocytes[GO]'. There is an 'Online Resource' section with the URL 'https://identifiers.org/UBERON:0002107'. The 'Predictions' section is divided into 'Heads' and 'Tails'. The 'Heads' section lists four predictions: '? can be underexpressed in liver', '? can be overexpressed in liver', '? is part of liver', and '? is expressed in liver'. The 'Tails' section lists one prediction: 'liver is a ?'. There is also a 'Known links' section at the bottom.

#### 3.2 View ranked predictions

... and their confidences. A checkmark means that this entity is in the test set (Note: this does not mean entities without a checkmark are false)

The screenshot shows the ranked predictions for the relationship 'can be overexpressed in' between '?' and 'liver'. The interface displays a table of results with columns for the entity name, a checkmark, and the confidence score.

Entity	Checkmark	Confidence
C-X3-C motif chemokine receptor 1		0.59375
cytochrome P450 family 2 subfamily C member 18	●	0.58576
cytochrome P450 family 2 subfamily S member 1		0.58576
alcohol dehydrogenase 1A (class I), alpha polypeptide	●	0.58576
alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide		0.58576
cytochrome P450		0.58576

### 3.3 Explore rules that generated a prediction

LinkExplorer Overview Entities Feedback Dataset: obi Explanation: max Load other

cytochrome P450 family 2 subfamily C member 18 can be overexpressed  
in liver  
Confidence: 0.58576  
because

Rule	Confidence	Correctly predicted	Predicted	
aldehyde oxidase 1 is in an interaction with cytochrome P450 family 2 subfamily C member 18	0.58571	41	65	
NADPH is catalyzed by cytochrome P450 family 2 subfamily C member 18	0.44218	65	142	
cytochrome P450 family 2 subfamily C member 18 associated with Triphosphopyridine nucleotide	0.42056	90	209	
cytochrome P450 family 2 subfamily C member 18 associated with 4-(Methylnitrosamino)-1-(β-pyridyl)-1-butanone	0.37838	14	32	
cytochrome P450 family 2 subfamily C member 18 can be underexpressed in frontal cortex	0.29116	214	730	
Hydrogen is catalyzed by cytochrome P450 family 2 subfamily C member 18	0.26767	231	858	
A is catalyzed by cytochrome P450 family 2 subfamily C member 18	0.24513	302	1227	

... and retrieve details about rules

back
Details x

---

### Rule

can be overexpressed in(X,liver) <= is in an interaction with(aldehyde oxidase 1,X)

```

graph TD
    AO1(aldehyde oxidase 1) -- "is in an interaction with" --> X((X))
    X -.-> |"can be overexpressed in"| liver((liver))
            
```

(Interactive graph)

-----> Predicted

Confidence	Correctly predicted	Predicted
0.58571	41	65

---

### Instantiations

cytochrome P450 family 2  
subfamily C member 18

can be overexpressed in

→

is in an interaction with

liver

cytochrome P450 family 2  
subfamily C member 18

Close



### 3.4 Explore instantiations of rules with variables

The screenshot shows the LinkExplorer interface. At the top, it says "liver is a endoderm-derived structure" with a confidence of 0.12918. Below this, it says "because" and lists four rules with their respective confidence, correctly predicted, and predicted counts. A modal window titled "Instantiations" is open, showing the rule "liver is a endoderm-derived structure" and its instantiations for variable A, such as "A = lobe of liver", "A = biliary ductule", etc.

Rule	Confidence	Correctly predicted	Predicted
liver shares tail/s of (is part of) with B B is a endoderm-derived structure	0.12918	259	2000
A is part of liver A is a endoderm-derived structure	0.06783	136	2000
liver shares tail/s of (is a) with B B is a endoderm-derived structure	0.04090	82	2000
A is part of liver	0.01845	37	2000

  

Instantiation
A = lobe of liver
A = biliary ductule
A = liver papillary process
A = ligament of liver
A = liver reticuloendothelial system
A = liver stroma
A = capsule of liver
A = triangular ligament of liver
A = ventral liver lobe

© Copyright 2021, Simon Ott.

Built with Sphinx using a theme provided by Read the Docs.