
DiffuPy Documentation

Release 0.0.4-dev

Josep Marín-Llaó, Sergi Picart Armada, Daniel Domingo-Fernández

Jul 05, 2021

CONTENTS:

- 1 Installation 3**
 - 1.1 Requirements 3
- 2 Command Line Interface 5**
- 3 Constants 7**
- 4 Databases 9**
 - 4.1 Network Dumps 9
 - 4.2 Custom-network formats 10
 - 4.3 Custom-network example 10
- 5 Visualization 13**
 - 5.1 Input mapping 13
 - 5.2 Validations 14
- 6 PathMe Harmonization 17**
- 7 Disclaimer 19**
- 8 References 21**
- Python Module Index 23**
- Index 25**

DiffuPath is an analytic tool for biological networks that connects the generic label propagation algorithms from [DiffuPy](#) to biological networks encoded in several formats such as Simple Interaction Format (SIF) or Biological Expression Language (BEL). For example, in the application scenario presented in the paper, we use three pathway databases (i.e., KEGG, Reactome and WikiPathways) and their integrated network retrieved from [PathMe](#)¹ to analyze three multi-omics datasets. However, other biological networks can be imported from the Bio2BEL ecosystem².

Installation is as easy as getting the code from [PyPI](#) with `python3 -m pip install diffupath`. See the [installation](#) documentation.

See also:

- Documented on [Read the Docs](#)
- Versioned on [GitHub](#)
- Tested on [Travis CI](#)
- Distributed by [PyPI](#)

¹ Domingo-Fernandez, D., Mubeen, S., Marin-Llao, J., Hoyt, C., *et al.* Hofmann-Apitius, M. (2019). PathMe: Merging and exploring mechanistic pathway knowledge.. *BMC Bioinformatics*, 20:243.

² Hoyt, C. T., *et al.* (2019). Integration of Structured Biological Data Sources using Biological Expression Language. *bioRxiv*, 631812.

INSTALLATION

The latest stable code can be installed from [PyPI](#) with:

```
$ python3 -m pip install diffupath
```

The most recent code can be installed from the source on [GitHub](#) with:

```
$ python3 -m pip install git+https://github.com/multipaths/diffupath.git
```

Required to install the latest PathMe version directly from GitHub:

```
$ python3 -m pip install git+https://github.com/PathwayMerger/PathMe.git
```

For developers, the repository can be cloned from [GitHub](#) and installed in editable mode with:

```
$ git clone https://github.com/multipaths/diffupath.git
$ cd diffupath
$ python3 -m pip install -e .
```

1.1 Requirements

diffupath requires the following libraries:

```
networkx (>=2.1)
pybel (0.13.2)
biokeen (0.0.14)
click (7.0)
tqdm (4.31.1)
numpy (1.16.3)
scipy (1.2.1)
scikit-learn (0.21.3)
pandas (0.24.2)
openpyxl (3.0.2)
plotly (4.5.3)
matplotlib (3.1.2)
matplotlib_venn (0.11.5)
bio2bel (0.2.1)
pathme
diffupy
```


COMMAND LINE INTERFACE

The following commands can be used directly use from your terminal:

1. Download a database for network analysis.

The following command generates a BEL file representing the network of the given database.

```
$ python3 -m diffupath database network --database=<database-name>
```

To check the available databases, run the following command:

```
$ python3 -m diffupath database ls
```

2. Run a diffusion analysis

The following command will run a diffusion method on a given network with the given data

```
$ python3 -m diffupath diffusion run --network=<path-to-network-file> --input=<path-to-  
↪data-file> --method=<method>
```


CONSTANTS

Constants of DiffuPath.

```
diffupath.constants.DEFAULT_DIFFUPATH_DIR = '/home/docs/.diffupath'
    Default DiffuPath directory

diffupath.constants.OUTPUT_DIFFUPATH_DIR = '/home/docs/.diffupath/output'
    Default DiffuPath output directory

diffupath.constants.ensure_output_dirs()
    Ensure that the output directories exists.

diffupath.constants.BY_METHOD = 'method'
    raw

diffupath.constants.KEGG_NAME = 'kegg'
    KEGG

diffupath.constants.REACTOME_NAME = 'reactome'
    Reactome

diffupath.constants.WIKIPATHWAYS_NAME = 'wikipathways'
    WikiPathways

diffupath.constants.MIRTARBASE_NAME = 'mirtarbase'
    MirTarBase

diffupath.constants.SIDER_NAME = 'sider'
    SIDER

diffupath.constants.PHEWAS_NAME = 'pewascatalog'
    PhewasCatalog

diffupath.constants.HSDN_NAME = 'hsdn'
    HSDN

diffupath.constants.DDR_NAME = 'ddr'
    DDR

diffupath.constants.DRUGBANK_NAME = 'drugbank'
    DrugBank

diffupath.constants.GENE_ONTOLOGY_NAME = 'go'
    Gene Ontology

diffupath.constants.DATABASES = ['kegg', 'reactome', 'wikipathways', 'mirtarbase',
'sider', 'pewascatalog', 'hsdn', 'ddr', 'drugbank', 'go']
    Databases available for download in DiffuPath
```


DATABASES

In this section, we describe the types of networks (databases) you can select to run diffusion methods over. These include the following and are described in detail in this section⁰:

- Select a network representing an individual biological database
- Select multiple databases to generate a harmonized network
- Select from one of four predefined collections of biological databases representing a harmonized network
- Submit your own network^{†0} from one of the accepted formats

4.1 Network Dumps

Because of the high computational cost of generating the kernel, we provide links to pre-calculated kernels for a set of networks representing biological databases.

Database	Description	Reference	Download
DDR	Disease-disease associations	¹	ddr.json
DrugBank	Drug and drug target interactions	²	drugbank.json
Gene Ontology	Hierarchy of tens of thousands of biological processes	³	go.json
HSDN	Associations between diseases and symptoms	⁴	hsdn.json
KEGG	Multi-omics interactions in biological pathways	⁵	kegg.json
miRTarBase	Interactions between miRNA and their targets	⁶	mirtarbase.json
Reactome	Multi-omics interactions in biological pathways	⁷	reactome.json
SIDER	Associations between drugs and side effects	⁸	sider.json
WikiPathways	Multi-omics interactions in biological pathways	⁹	wikipathways.json

⁰ Please note that all networks available through DiffuPath have been generated using PyBEL v.0.13.2.

^{†0} If there are duplicated nodes in your network, please take a look at this [Jupyter Notebook](#) to address the issue.

¹ Menche, J., et al. (2015). Disease networks. *Uncovering disease-disease relationships through the incomplete interactome*. Science, 347(6224), 1257601.

² Wishart, D. S., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082.

³ Ashburner, M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1), 25–9.

⁴ Zhou, X., Menche, J., Barabási, A. L., & Sharma, A. (2014). Human symptoms–disease network. *Nature communications*, 5(1), 1–10.

⁵ Kanehisa, et al. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs.. *Nucleic Acids Res.* 45,D353–D361.

⁶ Huang, H. Y., et al. (2020). miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic acids research*, 48(D1), D148–D154.

⁷ Fabregat, A et al. (2016). The Reactome Pathway Knowledgebase, *Nucleic Acids Research* 44. Database issue: D481–D487.

⁸ Kuhn, M., et al. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Research*, 44(D1), D1075–D1079.

⁹ Slenter, D.N., et al. (2017). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research*, 46(D1):D661–D667.

If you would like to use one of our predefined collections, you can similarly download pre-calculated kernels for sets of networks representing integrated biological databases.

Collection	Database	Description	Download
#1	KEGG, Reactome and WikiPathways	-omics and biological processes/pathways	pathme.json
#2	KEGG, Reactome, WikiPathways and DrugBank	-omics and biological processes/pathways with a strong focus on drug/chemical interactions	pathme_drugbank.json
#3	KEGG, Reactome, WikiPathways and MirTarBase	-omics and biological processes/ pathways enriched with miRNAs	pathme_mirtarbase.json

4.2 Custom-network formats

You can also submit your own networks in any of the following formats:

- [BEL](#) (.bel)
- [CSV](#) (.csv)
- [Edge list](#) (.lst)
- [GML](#) (.gml or .xml)
- [GraphML](#) (.graphml or .xml)
- [Pickle](#) (.pickle)
- [TSV](#) (.tsv)

Minimally, please ensure each of the following columns are included in the network file you submit:

- Source
- Target

Optionally, you can choose to add a third column, “Relation” in your network (as in the example below). If the relation between the **Source** and **Target** nodes is omitted, and/or if the directionality is ambiguous, either node can be assigned as the **Source** or **Target**.

4.3 Custom-network example

Source	Target	Relation
A	B	Increase
B	C	Association
A	D	Association

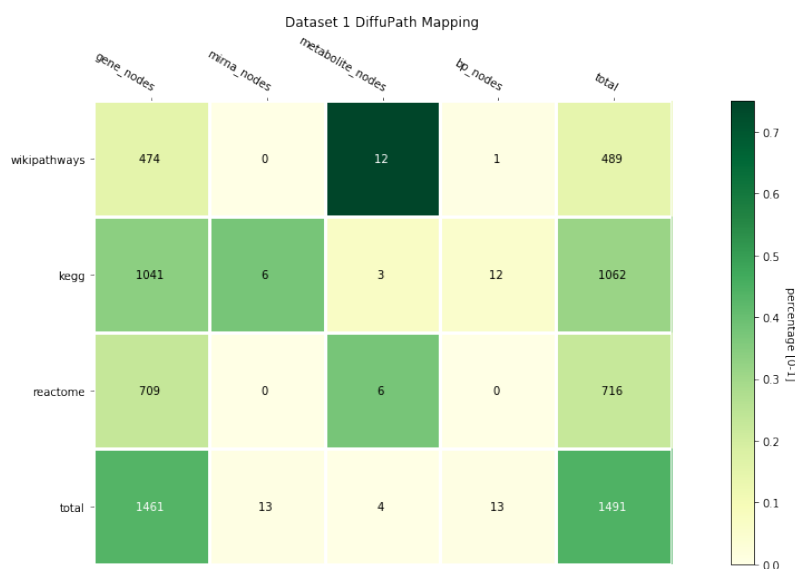
You can also take a look at our [sample networks](#) folder for some examples networks.

4.3.1 References

VISUALIZATION

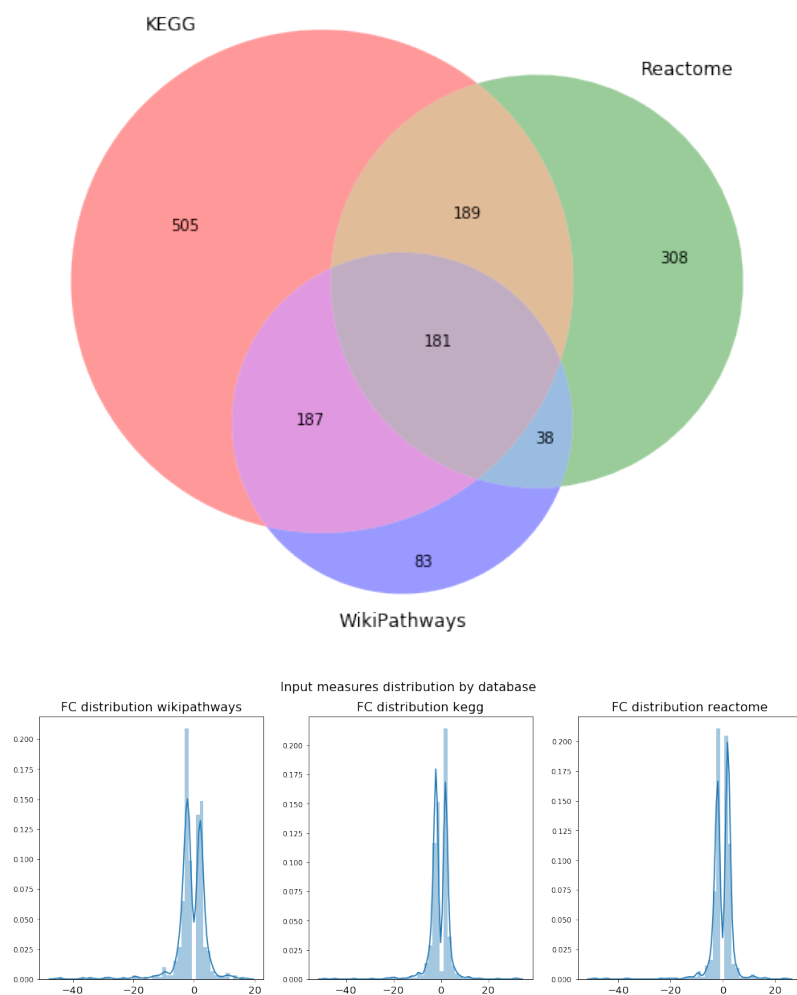
5.1 Input mapping

Even though it is not relevant for the input user usage, taking into account the input mapped entities over the background network is relevant for the diffusion process assessment, since the coverage of the input implies the actual entities-scores that are being diffused. In other words, only the entities whose labels match an entity in the network will be further processed for diffusion.



To visualize the mapping statistics heatmap, use the following function:

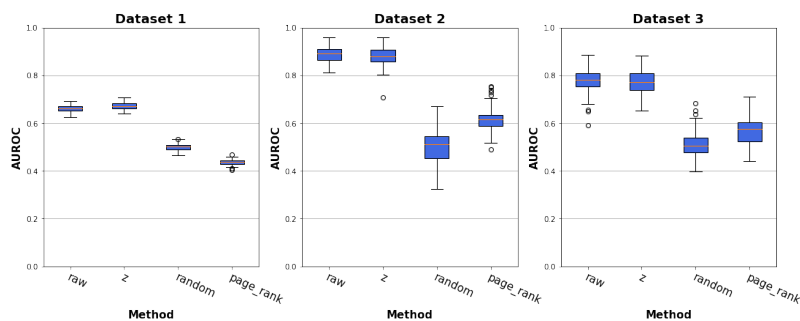
Further data views can be rendered for the input data mapping, such as VennDiagram to explore the overlap or distribution bloxplot:



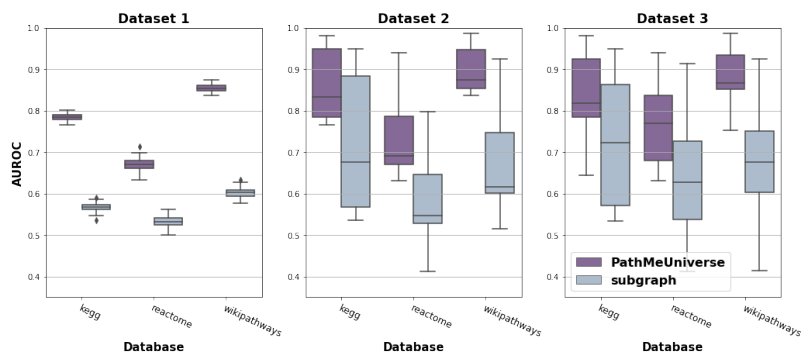
5.2 Validations

To visualize the metrics derived from validation experiments, you can plot metric Boxplots for repeated holdouts or iterated cross validation and its statistical tests and Barcharts with its threshold line:

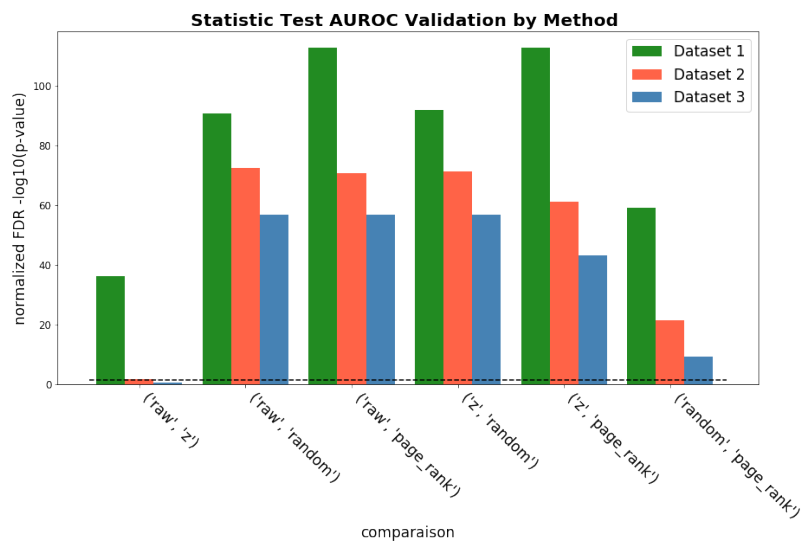
Two dimensional BLOXPLOT:



Three dimensional BLOXPLOT:



Statistical test BARCHART:



PATHME HARMONIZATION

DISCLAIMER

DiffuPath is a scientific software that has been developed in an academic capacity, and thus comes with no warranty or guarantee of maintenance, support, or back-up of data.

REFERENCES

PYTHON MODULE INDEX

d

`diffupath.constants`, [7](#)

INDEX

B

BY_METHOD (in module *diffupath.constants*), 7

D

DATABASES (in module *diffupath.constants*), 7

DDR_NAME (in module *diffupath.constants*), 7

DEFAULT_DIFFUPATH_DIR (in module *diffupath.constants*), 7

diffupath.constants
module, 7

DRUGBANK_NAME (in module *diffupath.constants*), 7

E

ensure_output_dirs() (in module *diffupath.constants*), 7

G

GENE_ONTOLOGY_NAME (in module *diffupath.constants*), 7

H

HSDN_NAME (in module *diffupath.constants*), 7

K

KEGG_NAME (in module *diffupath.constants*), 7

M

MIRTarBASE_NAME (in module *diffupath.constants*), 7

module
diffupath.constants, 7

O

OUTPUT_DIFFUPATH_DIR (in module *diffupath.constants*), 7

P

PHEWAS_NAME (in module *diffupath.constants*), 7

R

REACTOME_NAME (in module *diffupath.constants*), 7

S

SIDER_NAME (in module *diffupath.constants*), 7

W

WIKIPATHWAYS_NAME (in module *diffupath.constants*), 7