

data_prep

November 5, 2022

1 Data Acquisition

The very first step of any computational experiment is always obtaining data. Most of the times this is the very tedious part and it pays off to spend a lot of time understanding, cleaning and refining your data set. For the sake of simplicity, we will take a few shortcuts in this tutorial, but we will try to hint at possibilities to improve the process as we go on.

1.1 Papyrus Data Set

The [Papyrus](#) is a curated data set that tries to make our lives easier by aggregating publically available data under one roof and already does a few things for us, such as standardizing the compounds, aggregating and transforming activity values and giving our data overall structure that is suitable for data analysis and machine learning. However, there are also downsides to using such data set. For example, the fact that this is a curated data set means that it needs to be updated each time new information becomes available so you might not always have the most recent data to work with. In addition, sometimes the form of standardization of the data set might not be suitable for your case so you still need to spend some time reading about how the data is curated and if valuable information is not lost. For example, Papyrus does not include stereochemistry by default so if you want that kind of data (i.e. for docking), you have to retrieve it separately.

In this tutorial, we will be using a tool (`papyrus-scaffold-visualizer`) that enables easy extraction of data from Papyrus, but also has some simple chemical space depiction features. First, we need UniProt accession keys of our protein of interest. For the purpose of this tutorial, it will be the [Adenosine A2A receptor](#):

```
[2]: # mount google drive
from google.colab import drive
drive.mount('/content/drive')

# define work directory to store data
DATA_ROOT = '/content/drive/MyDrive/DrugExDemo/' # or wherever you want the
↳ generated files to live on your GoogleDrive
import os
os.makedirs(DATA_ROOT, exist_ok=True)
os.chdir(DATA_ROOT)

# fetch pretrained model
os.makedirs("./data/drugex/models/pretrained/", exist_ok=True)
```

```

! wget -nc -P './data/drugex/models/pretrained/' 'https://zenodo.org/record/
↪7096859/files/DrugEx_v2_PT_Papyrus05.5.zip'
! unzip -n './data/drugex/models/pretrained/DrugEx_v2_PT_Papyrus05.5.zip' -d './
↪data/drugex/models/pretrained/DrugEx_v2_PT_Papyrus05.5'

# install dependencies
! git clone https://github.com/martin-sicho/drugex-demo
! pip install -r drugex-demo/requirements.txt

# verify where we are working
os.getcwd()

```

Drive already mounted at /content/drive; to attempt to forcibly remount, call `drive.mount("/content/drive", force_remount=True)`.
File './data/drugex/models/pretrained/DrugEx_v2_PT_Papyrus05.5.zip' already there; not retrieving.

Archive: ./data/drugex/models/pretrained/DrugEx_v2_PT_Papyrus05.5.zip
fatal: destination path 'drugex-demo' already exists and is not an empty directory.
Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/public/simple/>
Collecting drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master
Cloning https://github.com/martin-sicho/DrugEx-CDDG.git (to revision master)
to /tmp/pip-install-zbvevyps/drugex_a10de01965014311bce80b641e9bf436
Running command git clone -q https://github.com/martin-sicho/DrugEx-CDDG.git
/tmp/pip-install-zbvevyps/drugex_a10de01965014311bce80b641e9bf436
Collecting papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0
Cloning https://github.com/martin-sicho/papyrus-scaffold-visualizer.git (to revision v0.2.0) to /tmp/pip-install-zbvevyps/papyrus-scaffold-visualizer_04eb96643a704b0ca20bad06f1426025
Running command git clone -q https://github.com/martin-sicho/papyrus-scaffold-visualizer.git /tmp/pip-install-zbvevyps/papyrus-scaffold-visualizer_04eb96643a704b0ca20bad06f1426025
Running command git checkout -q ff4f2e885a3973f90a0d9864dfa00abed493f78d
Requirement already satisfied: molplotly==1.1.4 in
/usr/local/lib/python3.7/dist-packages (from -r drugex-demo/requirements.txt (line 2)) (1.1.4)
Requirement already satisfied: mols2grid==1.0.0 in
/usr/local/lib/python3.7/dist-packages (from -r drugex-demo/requirements.txt (line 3)) (1.0.0)
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.7/dist-packages (from drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.21.6)
Requirement already satisfied: scikit-learn>=1.0.2 in
/usr/local/lib/python3.7/dist-packages (from drugex@

```

git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-
demo/requirements.txt (line 1)) (1.0.2)
Requirement already satisfied: pandas>=1.2.2 in /usr/local/lib/python3.7/dist-
packages (from drugex@ git+https://github.com/martin-sicho/DrugEx-
CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.3.5)
Requirement already satisfied: torch>=1.7.0 in /usr/local/lib/python3.7/dist-
packages (from drugex@ git+https://github.com/martin-sicho/DrugEx-
CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.12.1+cu113)
Requirement already satisfied: matplotlib>=2.0 in /usr/local/lib/python3.7/dist-
packages (from drugex@ git+https://github.com/martin-sicho/DrugEx-
CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.2.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages
(from drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r
drugex-demo/requirements.txt (line 1)) (4.64.1)
Requirement already satisfied: rdkit-pypi in /usr/local/lib/python3.7/dist-
packages (from drugex@ git+https://github.com/martin-sicho/DrugEx-
CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (2022.9.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages
(from drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r
drugex-demo/requirements.txt (line 1)) (1.2.0)
Requirement already satisfied: optuna in /usr/local/lib/python3.7/dist-packages
(from drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r
drugex-demo/requirements.txt (line 1)) (3.0.3)
Requirement already satisfied: gitpython in /usr/local/lib/python3.7/dist-
packages (from drugex@ git+https://github.com/martin-sicho/DrugEx-
CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.1.29)
Requirement already satisfied: xgboost in /usr/local/lib/python3.7/dist-packages
(from drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r
drugex-demo/requirements.txt (line 1)) (0.90)
Collecting papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master
  Cloning https://github.com/OlivierBeq/Papyrus-scripts.git (to revision master)
  to /tmp/pip-install-zbvevyps/papyrus-scripts_49c51cbe007a461d93e61729128767b1
  Running command git clone -q https://github.com/OlivierBeq/Papyrus-scripts.git
  /tmp/pip-install-zbvevyps/papyrus-scripts_49c51cbe007a461d93e61729128767b1
Requirement already satisfied: sklearn in /usr/local/lib/python3.7/dist-packages
(from papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-
scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (0.0)
Collecting prodec@ https://github.com/OlivierBeq/ProDEC/tarball/master
  Using cached https://github.com/OlivierBeq/ProDEC/tarball/master
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing wheel metadata ... done
Collecting upsetplot@ https://github.com/OlivierBeq/UpSetPlot/tarball/master
  Using cached https://github.com/OlivierBeq/UpSetPlot/tarball/master
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-
packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-

```

sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt
(line 4)) (2.23.0)

Requirement already satisfied: natsort in /usr/local/lib/python3.7/dist-packages
(from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-
sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt
(line 4)) (5.5.0)

Requirement already satisfied: mordred in /usr/local/lib/python3.7/dist-packages
(from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-
sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt
(line 4)) (1.2.0)

Requirement already satisfied: swifter in /usr/local/lib/python3.7/dist-packages
(from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-
sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt
(line 4)) (1.3.4)

Requirement already satisfied: pystow in /usr/local/lib/python3.7/dist-packages
(from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-
sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt
(line 4)) (0.4.6)

Requirement already satisfied: tabulate in /usr/local/lib/python3.7/dist-
packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-
sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt
(line 4)) (0.8.10)

Requirement already satisfied: psutil in /usr/local/lib/python3.7/dist-packages
(from prodec@
https://github.com/OlivierBeq/ProDEC/tarball/master->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-
visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (5.9.3)

Requirement already satisfied: orjson in /usr/local/lib/python3.7/dist-packages
(from prodec@
https://github.com/OlivierBeq/ProDEC/tarball/master->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-
visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (3.8.1)

Requirement already satisfied: ipykernel in /usr/local/lib/python3.7/dist-
packages (from molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2))
(5.3.4)

Requirement already satisfied: dash>=2.0.0 in /usr/local/lib/python3.7/dist-
packages (from molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2))
(2.6.2)

Requirement already satisfied: plotly>=5.0.0 in /usr/local/lib/python3.7/dist-
packages (from molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2))
(5.5.0)

Requirement already satisfied: jupyter-dash>=0.4.2 in /usr/local/lib/python3.7/dist-packages (from molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (0.4.2)

Requirement already satisfied: werkzeug>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (2.2.2)

Requirement already satisfied: nbformat in /usr/local/lib/python3.7/dist-packages (from molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (5.7.0)

Requirement already satisfied: ipywidgets<8,>=7 in /usr/local/lib/python3.7/dist-packages (from mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (7.7.1)

Requirement already satisfied: jinja2>=2.11.0 in /usr/local/lib/python3.7/dist-packages (from mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (3.1.2)

Requirement already satisfied: Flask>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (2.2.2)

Requirement already satisfied: flask-compress in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (1.13)

Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (2.0.0)

Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (5.0.0)

Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (2.0.0)

Requirement already satisfied: itsdangerous>=2.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (2.1.2)

Requirement already satisfied: importlib-metadata>=3.6.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (4.13.0)

Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (8.1.3)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=3.6.0->Flask>=1.0.4->dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (3.9.0)

Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=3.6.0->Flask>=1.0.4->dash>=2.0.0->molplotly==1.1.4->-r drugex-

demo/requirements.txt (line 2)) (4.1.1)
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (7.9.0)
Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (5.1.1)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (3.0.3)
Requirement already satisfied: widgetsnbextension~=3.6.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (3.6.1)
Requirement already satisfied: ipython-genutils~=0.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.2.0)
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from ipykernel->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (6.1.12)
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.7/dist-packages (from ipykernel->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (5.1.1)
Requirement already satisfied: prompt-toolkit<2.1.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (2.0.10)
Requirement already satisfied: backcall in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.2.0)
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (4.4.2)
Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (4.8.0)
Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (57.4.0)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (2.6.1)
Requirement already satisfied: jedi>=0.10 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-

demo/requirements.txt (line 3)) (0.18.1)

Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.7.5)

Requirement already satisfied: parso<0.9.0,>=0.8.0 in /usr/local/lib/python3.7/dist-packages (from jedi>=0.10->ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.8.3)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.7/dist-packages (from jinja2>=2.11.0->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (2.1.1)

Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (1.5.6)

Requirement already satisfied: retrying in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (1.3.3)

Requirement already satisfied: ansi2html in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (1.8.0)

Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (2.8.2)

Requirement already satisfied: cycycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (0.11.0)

Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.4.4)

Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.0.9)

Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=1.2.2->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (2022.5)

Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from plotly>=5.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (1.15.0)

Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.7/dist-packages (from plotly>=5.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (8.1.0)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-toolkit<2.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.2.5)

Requirement already satisfied: scipy>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn>=1.0.2->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.7.3)

Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn>=1.0.2->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.1.0)

Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-packages (from widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (5.5.0)

Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (5.6.1)

Requirement already satisfied: pyzmq>=17 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (23.2.1)

Requirement already satisfied: jupyter-core>=4.4.0 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (4.11.2)

Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (1.8.0)

Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.13.3)

Requirement already satisfied: ptyprocess in /usr/local/lib/python3.7/dist-packages (from terminado>=0.8.1->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.7.0)

Requirement already satisfied: brotli in /usr/local/lib/python3.7/dist-packages (from flask-compress->dash>=2.0.0->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (1.0.9)

Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.7/dist-packages (from gitpython->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (4.0.9)

Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.7/dist-packages (from gitdb<5,>=4.0.1->gitpython->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (5.0.0)

Requirement already satisfied: networkx==2.* in /usr/local/lib/python3.7/dist-packages (from mordred->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (2.6.3)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgets

nbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.4)

Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (5.0.1)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (1.5.0)

Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.6.0)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.8.4)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.7.1)

Requirement already satisfied: fastjsonschema in /usr/local/lib/python3.7/dist-packages (from nbformat->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (2.16.2)

Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.7/dist-packages (from nbformat->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (4.3.3)

Requirement already satisfied: importlib-resources>=1.4.0 in /usr/local/lib/python3.7/dist-packages (from jsonschema>=2.6->nbformat->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (5.10.0)

Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/dist-packages (from jsonschema>=2.6->nbformat->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (22.1.0)

Requirement already satisfied: pyparsing!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /usr/local/lib/python3.7/dist-packages (from jsonschema>=2.6->nbformat->molplotly==1.1.4->-r drugex-demo/requirements.txt (line 2)) (0.18.1)

Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (from bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets<8,>=7->mols2grid==1.0.0->-r drugex-demo/requirements.txt (line 3)) (0.5.1)

Requirement already satisfied: cmaes>=0.8.2 in /usr/local/lib/python3.7/dist-packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (0.8.2)

Requirement already satisfied: sqlalchemy>=1.3.0 in /usr/local/lib/python3.7/dist-packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.4.42)

Requirement already satisfied: colorlog in /usr/local/lib/python3.7/dist-

packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (6.7.0)
 Requirement already satisfied: PyYAML in /usr/local/lib/python3.7/dist-packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (6.0)
 Requirement already satisfied: alembic>=1.5.0 in /usr/local/lib/python3.7/dist-packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.8.1)
 Requirement already satisfied: cliff in /usr/local/lib/python3.7/dist-packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.10.1)
 Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (21.3)
 Requirement already satisfied: Mako in /usr/local/lib/python3.7/dist-packages (from alembic>=1.5.0->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.2.3)
 Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.7/dist-packages (from sqlalchemy>=1.3.0->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.1.3.post0)
 Requirement already satisfied: PrettyTable>=0.7.2 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.4.1)
 Requirement already satisfied: stevedore>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (3.5.2)
 Requirement already satisfied: pbr!=2.1.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (5.11.0)
 Requirement already satisfied: autopage>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (0.5.1)
 Requirement already satisfied: cmd2>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (2.4.2)
 Requirement already satisfied: pyperclip>=1.6 in /usr/local/lib/python3.7/dist-packages (from cmd2>=1.0.0->cliff->optuna->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (1.8.2)
 Requirement already satisfied: pickle5 in /usr/local/lib/python3.7/dist-packages (from pystow->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (0.0.12)

Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from rdkit-pypi->drugex@ git+https://github.com/martin-sicho/DrugEx-CDDG.git@master->-r drugex-demo/requirements.txt (line 1)) (7.1.2)

Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (3.0.4)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (2.10)

Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (1.24.3)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (2022.9.24)

Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (1.5.0)

Requirement already satisfied: dask[dataframe]>=2.10.0 in /usr/local/lib/python3.7/dist-packages (from swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (2022.2.0)

Requirement already satisfied: fsspec>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (2022.10.0)

Requirement already satisfied: toolz>=0.8.2 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (0.12.0)

Requirement already satisfied: partd>=0.3.10 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (1.3.0)

```
Requirement already satisfied: locket in /usr/local/lib/python3.7/dist-packages
(from partd>=0.3.10->dask[dataframe]>=2.10.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer@ git+https://github.com/martin-sicho/papyrus-scaffold-
visualizer.git@v0.2.0->-r drugex-demo/requirements.txt (line 4)) (1.0.0)
```

[2]: `'/content/drive/MyDrive/DrugExDemo'`

```
[3]: !pip install git+https://github.com/martin-sicho/DrugEx-CDDG.git@master
!pip install git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.
↪git@v0.2.0
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting git+https://github.com/martin-sicho/DrugEx-CDDG.git@master
  Cloning https://github.com/martin-sicho/DrugEx-CDDG.git (to revision master)
  to /tmp/pip-req-build-10sapv1n
  Running command git clone -q https://github.com/martin-sicho/DrugEx-CDDG.git
  /tmp/pip-req-build-10sapv1n
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.7/dist-
packages (from drugex==3.2.0) (1.21.6)
Requirement already satisfied: scikit-learn>=1.0.2 in
/usr/local/lib/python3.7/dist-packages (from drugex==3.2.0) (1.0.2)
Requirement already satisfied: pandas>=1.2.2 in /usr/local/lib/python3.7/dist-
packages (from drugex==3.2.0) (1.3.5)
Requirement already satisfied: torch>=1.7.0 in /usr/local/lib/python3.7/dist-
packages (from drugex==3.2.0) (1.12.1+cu113)
Requirement already satisfied: matplotlib>=2.0 in /usr/local/lib/python3.7/dist-
packages (from drugex==3.2.0) (3.2.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages
(from drugex==3.2.0) (4.64.1)
Requirement already satisfied: rdkit-pypi in /usr/local/lib/python3.7/dist-
packages (from drugex==3.2.0) (2022.9.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages
(from drugex==3.2.0) (1.2.0)
Requirement already satisfied: optuna in /usr/local/lib/python3.7/dist-packages
(from drugex==3.2.0) (3.0.3)
Requirement already satisfied: gitpython in /usr/local/lib/python3.7/dist-
packages (from drugex==3.2.0) (3.1.29)
Requirement already satisfied: xgboost in /usr/local/lib/python3.7/dist-packages
(from drugex==3.2.0) (0.90)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-
packages (from matplotlib>=2.0->drugex==3.2.0) (0.11.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in
/usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex==3.2.0)
(3.0.9)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex==3.2.0)
```

(2.8.2)

Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->drugex==3.2.0)

(1.4.4)

Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from
kiwisolver>=1.0.1->matplotlib>=2.0->drugex==3.2.0) (4.1.1)

Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-
packages (from pandas>=1.2.2->drugex==3.2.0) (2022.5)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-
packages (from python-dateutil>=2.1->matplotlib>=2.0->drugex==3.2.0) (1.15.0)

Requirement already satisfied: scipy>=1.1.0 in /usr/local/lib/python3.7/dist-
packages (from scikit-learn>=1.0.2->drugex==3.2.0) (1.7.3)

Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.7/dist-packages (from scikit-learn>=1.0.2->drugex==3.2.0)
(3.1.0)

Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.7/dist-
packages (from gitpython->drugex==3.2.0) (4.0.9)

Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.7/dist-
packages (from gitdb<5,>=4.0.1->gitpython->drugex==3.2.0) (5.0.0)

Requirement already satisfied: PyYAML in /usr/local/lib/python3.7/dist-packages
(from optuna->drugex==3.2.0) (6.0)

Requirement already satisfied: cliff in /usr/local/lib/python3.7/dist-packages
(from optuna->drugex==3.2.0) (3.10.1)

Requirement already satisfied: colorlog in /usr/local/lib/python3.7/dist-
packages (from optuna->drugex==3.2.0) (6.7.0)

Requirement already satisfied: alembic>=1.5.0 in /usr/local/lib/python3.7/dist-
packages (from optuna->drugex==3.2.0) (1.8.1)

Requirement already satisfied: cmaes>=0.8.2 in /usr/local/lib/python3.7/dist-
packages (from optuna->drugex==3.2.0) (0.8.2)

Requirement already satisfied: importlib-metadata<5.0.0 in
/usr/local/lib/python3.7/dist-packages (from optuna->drugex==3.2.0) (4.13.0)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-
packages (from optuna->drugex==3.2.0) (21.3)

Requirement already satisfied: sqlalchemy>=1.3.0 in
/usr/local/lib/python3.7/dist-packages (from optuna->drugex==3.2.0) (1.4.42)

Requirement already satisfied: importlib-resources in
/usr/local/lib/python3.7/dist-packages (from
alembic>=1.5.0->optuna->drugex==3.2.0) (5.10.0)

Requirement already satisfied: Mako in /usr/local/lib/python3.7/dist-packages
(from alembic>=1.5.0->optuna->drugex==3.2.0) (1.2.3)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-metadata<5.0.0->optuna->drugex==3.2.0) (3.9.0)

Requirement already satisfied: greenlet!=0.4.17 in
/usr/local/lib/python3.7/dist-packages (from
sqlalchemy>=1.3.0->optuna->drugex==3.2.0) (1.1.3.post0)

Requirement already satisfied: stevedore>=2.0.1 in
/usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex==3.2.0)

(3.5.2)
Requirement already satisfied: cmd2>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex==3.2.0) (2.4.2)
Requirement already satisfied: autopage>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex==3.2.0) (0.5.1)
Requirement already satisfied: pbr!=2.1.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex==3.2.0) (5.11.0)
Requirement already satisfied: PrettyTable>=0.7.2 in /usr/local/lib/python3.7/dist-packages (from cliff->optuna->drugex==3.2.0) (3.4.1)
Requirement already satisfied: attrs>=16.3.0 in /usr/local/lib/python3.7/dist-packages (from cmd2>=1.0.0->cliff->optuna->drugex==3.2.0) (22.1.0)
Requirement already satisfied: pyperclip>=1.6 in /usr/local/lib/python3.7/dist-packages (from cmd2>=1.0.0->cliff->optuna->drugex==3.2.0) (1.8.2)
Requirement already satisfied: wcwidth>=0.1.7 in /usr/local/lib/python3.7/dist-packages (from cmd2>=1.0.0->cliff->optuna->drugex==3.2.0) (0.2.5)
Requirement already satisfied: MarkupSafe>=0.9.2 in /usr/local/lib/python3.7/dist-packages (from Mako->alembic>=1.5.0->optuna->drugex==3.2.0) (2.1.1)
Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from rdkit-pypi->drugex==3.2.0) (7.1.2)
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting git+https://github.com/martin-sicho/papyrus-scaffold-visualizer.git@v0.2.0
Cloning https://github.com/martin-sicho/papyrus-scaffold-visualizer.git (to revision v0.2.0) to /tmp/pip-req-build-1h0qkc34
Running command git clone -q https://github.com/martin-sicho/papyrus-scaffold-visualizer.git /tmp/pip-req-build-1h0qkc34
Running command git checkout -q ff4f2e885a3973f90a0d9864dfa00abed493f78d
Collecting papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master
Cloning https://github.com/OlivierBeq/Papyrus-scripts.git (to revision master) to /tmp/pip-install-7rj0bbwu/papyrus-scripts_80bb0def3e3e48f3a74c3169522551fc
Running command git clone -q https://github.com/OlivierBeq/Papyrus-scripts.git /tmp/pip-install-7rj0bbwu/papyrus-scripts_80bb0def3e3e48f3a74c3169522551fc
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from papyrus-scaffold-visualizer==0.2.0) (1.21.6)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from papyrus-scaffold-visualizer==0.2.0) (1.3.5)
Requirement already satisfied: sklearn in /usr/local/lib/python3.7/dist-packages (from papyrus-scaffold-visualizer==0.2.0) (0.0)
Requirement already satisfied: rdkit-pypi in /usr/local/lib/python3.7/dist-packages (from papyrus-scaffold-visualizer==0.2.0) (2022.9.1)
Requirement already satisfied: molplotly in /usr/local/lib/python3.7/dist-packages (from papyrus-scaffold-visualizer==0.2.0) (1.1.4)
Collecting prodec@ https://github.com/OlivierBeq/ProDEC/tarball/master

```

Using cached https://github.com/OlivierBeq/ProDEC/tarball/master
Installing build dependencies ... done
Getting requirements to build wheel ... done
Preparing wheel metadata ... done
Collecting upsetplot@ https://github.com/OlivierBeq/UpSetPlot/tarball/master
Using cached https://github.com/OlivierBeq/UpSetPlot/tarball/master
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (2.23.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.2.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (4.64.1)
Requirement already satisfied: natsort in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (5.5.0)
Requirement already satisfied: mordred in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.2.0)
Requirement already satisfied: swifter in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.3.4)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.0.2)
Requirement already satisfied: xgboost in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.90)
Requirement already satisfied: pystow in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.4.6)
Requirement already satisfied: tabulate in /usr/local/lib/python3.7/dist-packages (from papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.8.10)
Requirement already satisfied: psutil in /usr/local/lib/python3.7/dist-packages (from prodec@ https://github.com/OlivierBeq/ProDEC/tarball/master->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (5.9.3)
Requirement already satisfied: orjson in /usr/local/lib/python3.7/dist-packages (from prodec@ https://github.com/OlivierBeq/ProDEC/tarball/master->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (3.8.1)
Requirement already satisfied: matplotlib>=2.0 in /usr/local/lib/python3.7/dist-packages (from upsetplot@

```

<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (3.2.2)
Requirement already satisfied: cycycler>=0.10 in /usr/local/lib/python3.7/dist-
packages (from matplotlib>=2.0->upsetplot@
<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (0.11.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in
/usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->upsetplot@
<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->upsetplot@
<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (1.4.4)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.7/dist-packages (from matplotlib>=2.0->upsetplot@
<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (2.8.2)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from
kiwisolver>=1.0.1->matplotlib>=2.0->upsetplot@
<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (4.1.1)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-
packages (from pandas->papyrus-scaffold-visualizer==0.2.0) (2022.5)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-
packages (from python-dateutil>=2.1->matplotlib>=2.0->upsetplot@
<https://github.com/OlivierBeq/UpSetPlot/tarball/master>->papyrus_scripts@
git+<https://github.com/OlivierBeq/Papyrus-scripts.git>@master->papyrus-scaffold-
visualizer==0.2.0) (1.15.0)
Requirement already satisfied: dash>=2.0.0 in /usr/local/lib/python3.7/dist-
packages (from molplotly->papyrus-scaffold-visualizer==0.2.0) (2.6.2)
Requirement already satisfied: ipykernel in /usr/local/lib/python3.7/dist-
packages (from molplotly->papyrus-scaffold-visualizer==0.2.0) (5.3.4)
Requirement already satisfied: plotly>=5.0.0 in /usr/local/lib/python3.7/dist-
packages (from molplotly->papyrus-scaffold-visualizer==0.2.0) (5.5.0)
Requirement already satisfied: nbformat in /usr/local/lib/python3.7/dist-
packages (from molplotly->papyrus-scaffold-visualizer==0.2.0) (5.7.0)
Requirement already satisfied: werkzeug>=2.0.0 in /usr/local/lib/python3.7/dist-
packages (from molplotly->papyrus-scaffold-visualizer==0.2.0) (2.2.2)
Requirement already satisfied: jupyter-dash>=0.4.2 in
/usr/local/lib/python3.7/dist-packages (from molplotly->papyrus-scaffold-

visualizer==0.2.0) (0.4.2)

Requirement already satisfied: flask-compress in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (1.13)

Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (5.0.0)

Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.0.0)

Requirement already satisfied: Flask>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.2.2)

Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/lib/python3.7/dist-packages (from dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.0.0)

Requirement already satisfied: Jinja2>=3.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (3.1.2)

Requirement already satisfied: itsdangerous>=2.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.1.2)

Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (8.1.3)

Requirement already satisfied: importlib-metadata>=3.6.0 in /usr/local/lib/python3.7/dist-packages (from Flask>=1.0.4->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (4.13.0)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=3.6.0->Flask>=1.0.4->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (3.9.0)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.7/dist-packages (from Jinja2>=3.0->Flask>=1.0.4->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.1.1)

Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (1.5.6)

Requirement already satisfied: ansi2html in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (1.8.0)

Requirement already satisfied: retrying in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (1.3.3)

Requirement already satisfied: ipython in /usr/local/lib/python3.7/dist-packages (from jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0)

(7.9.0)
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.7/dist-packages (from plotly>=5.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (8.1.0)

(8.1.0)
Requirement already satisfied: brotli in /usr/local/lib/python3.7/dist-packages (from flask-compress->dash>=2.0.0->molplotly->papyrus-scaffold-visualizer==0.2.0) (1.0.9)

Requirement already satisfied: traitlets>=4.1.0 in /usr/local/lib/python3.7/dist-packages (from ipykernel->molplotly->papyrus-scaffold-visualizer==0.2.0) (5.1.1)

Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.7/dist-packages (from ipykernel->molplotly->papyrus-scaffold-visualizer==0.2.0) (5.1.1)

Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from ipykernel->molplotly->papyrus-scaffold-visualizer==0.2.0) (6.1.12)

Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.7.5)

Requirement already satisfied: backcall in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.2.0)

Requirement already satisfied: jedi>=0.10 in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.18.1)

Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (57.4.0)

Requirement already satisfied: prompt-toolkit<2.1.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.0.10)

Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (4.4.2)

Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (4.8.0)

Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.6.1)

Requirement already satisfied: parso<0.9.0,>=0.8.0 in /usr/local/lib/python3.7/dist-packages (from jedi>=0.10->ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.8.3)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-toolkit<2.1.0,>=2.0.0->ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.2.5)

Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.7/dist-packages (from jupyter-client->ipykernel->molplotly->papyrus-scaffold-visualizer==0.2.0) (23.2.1)

Requirement already satisfied: jupyter-core>=4.6.0 in /usr/local/lib/python3.7/dist-packages (from jupyter-client->ipykernel->molplotly->papyrus-scaffold-visualizer==0.2.0) (4.11.2)

Requirement already satisfied: networkx==2.* in /usr/local/lib/python3.7/dist-packages (from mordred->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (2.6.3)

Requirement already satisfied: fastjsonschema in /usr/local/lib/python3.7/dist-packages (from nbformat->molplotly->papyrus-scaffold-visualizer==0.2.0) (2.16.2)

Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.7/dist-packages (from nbformat->molplotly->papyrus-scaffold-visualizer==0.2.0) (4.3.3)

Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/dist-packages (from jsonschema>=2.6->nbformat->molplotly->papyrus-scaffold-visualizer==0.2.0) (22.1.0)

Requirement already satisfied: pyrsistent!=0.17.0,!0.17.1,!0.17.2,>=0.14.0 in /usr/local/lib/python3.7/dist-packages (from jsonschema>=2.6->nbformat->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.18.1)

Requirement already satisfied: importlib-resources>=1.4.0 in /usr/local/lib/python3.7/dist-packages (from jsonschema>=2.6->nbformat->molplotly->papyrus-scaffold-visualizer==0.2.0) (5.10.0)

Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.7/dist-packages (from pexpect->ipython->jupyter-dash>=0.4.2->molplotly->papyrus-scaffold-visualizer==0.2.0) (0.7.0)

Requirement already satisfied: pickle5 in /usr/local/lib/python3.7/dist-packages (from pystow->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.0.12)

Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from rdkit-pypi->papyrus-scaffold-visualizer==0.2.0) (7.1.2)

Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.24.3)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (2022.9.24)

Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (3.0.4)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (2.10)

Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn->papyrus_scripts@

git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (3.1.0)

Requirement already satisfied: scipy>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.7.3)

Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.5.0)

Requirement already satisfied: ipywidgets>=7.0.0 in /usr/local/lib/python3.7/dist-packages (from swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (7.7.1)

Requirement already satisfied: bleach>=3.1.1 in /usr/local/lib/python3.7/dist-packages (from swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (5.0.1)

Requirement already satisfied: dask[dataframe]>=2.10.0 in /usr/local/lib/python3.7/dist-packages (from swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (2022.2.0)

Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (from bleach>=3.1.1->swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.5.1)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (21.3)

Requirement already satisfied: toolz>=0.8.2 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.12.0)

Requirement already satisfied: fsspec>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (2022.10.0)

Requirement already satisfied: partd>=0.3.10 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.3.0)

Requirement already satisfied: pyyaml>=5.3.1 in /usr/local/lib/python3.7/dist-packages (from dask[dataframe]>=2.10.0->swifter->papyrus_scripts@git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (6.0)

Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from

```

ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (3.0.3)
Requirement already satisfied: ipython-genutils~=0.2.0 in
/usr/local/lib/python3.7/dist-packages (from
ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (0.2.0)
Requirement already satisfied: widgetsnbextension~=3.6.0 in
/usr/local/lib/python3.7/dist-packages (from
ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (3.6.1)
Requirement already satisfied: locket in /usr/local/lib/python3.7/dist-packages
(from partd>=0.3.10->dask[dataframe]>=2.10.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (1.0.0)
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-
packages (from
widgetsnbextension~=3.6.0->ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (5.5.0)
Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-
packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.0.0->sw
ifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (1.8.0)
Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-
packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.0.0->sw
ifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (5.6.1)
Requirement already satisfied: terminado>=0.8.1 in
/usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension
~=3.6.0->ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (0.13.3)
Requirement already satisfied: pandocfilters>=1.4.1 in
/usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgets
nbextension~=3.6.0->ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (1.5.0)
Requirement already satisfied: entrypoints>=0.2.2 in
/usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgets
nbextension~=3.6.0->ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (0.4)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-
packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets
>=7.0.0->swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-

```

```
scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.7.1)
Requirement already satisfied: mistune<2,>=0.8.1 in
/usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgets
nbextension~=3.6.0->ipywidgets>=7.0.0->swifter->papyrus_scripts@
git+https://github.com/OlivierBeq/Papyrus-scripts.git@master->papyrus-scaffold-
visualizer==0.2.0) (0.8.4)
Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-
packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets
>=7.0.0->swifter->papyrus_scripts@ git+https://github.com/OlivierBeq/Papyrus-
scripts.git@master->papyrus-scaffold-visualizer==0.2.0) (0.6.0)
```

```
[4]: acc_keys = ["P18031"] # Adenosine receptor PTP1B (https://www.uniprot.org/
    ↪uniprotkb/P29274/entry)
name = "PTP1B_LIGANDS" # name of the file to be generated
quality = "medium" # choose minimum quality from {"high", "medium", "low"}
```

These will be our basic settings for the data extraction from Papyrus. In short, we will get compound structures and their measured activities on the A2A receptor that should be of low quality or higher as specified by Papyrus' data curation workflow. For the purpose of this tutorial, we are interested in all available data (low and higher quality), but in some other applications you might want a more stringent filter. We suggest you read the documentation of Papyrus to learn how these quality classes are determined.

The Papyrus class exposes some basic configuration features with the help of the more powerful `papyrus_scripts` package. It will fetch the latest available version of Papyrus data automatically for us as specified. Here, we are asking only for the latest Papyrus data files without definitions of stereochemistry and all data will be placed in the `./data` directory:

```
[5]: from scaffviz.data.papyrus import Papyrus

# fetches latest version of Papyrus if not already available and filters out
    ↪A2A data for us
papyrus = Papyrus(
    data_dir="./data", # data storage directory
    stereo=False, # do not include stereochemistry
    version='latest'
)
```

No data was transferred yet, but once we attempt to filter for our data, the data should start downloading:

```
[6]: dataset = papyrus.getData(
    acc_keys,
    quality,
    name,
    use_existing=True # use existing data set if it was already compiled before
    ↪to speed things up
)
```

Latest version: 05.5

Number of files to be downloaded: 6

Total size: 118MB

Downloading version 05.5: 0% | 0.00/118M [00:00<?, ?B/s]

Using existing data from ./data/PTP1B_LIGANDS.tsv...

The Papyrus data will only be fetched once if the latest version is already available in the ./data folder. Plus, thanks to the `use_existing` parameter the created data set file will also be reused once extracted. You can find it here:

```
[7]: dataset.path
```

```
[7]: './data/PTP1B_LIGANDS.tsv'
```

It is no special file and you can either use `pandas` to load it manually into a `DataFrame` in your environment or get it straight from the instance returned by the `getData()` method:

```
[8]: dataset.asDataFrame().head()
```

```
[8]:
```

	Activity_ID	Quality	source	\
0	AADVGRQCQZZZNH_on_P18031_WT	High	ChEMBL30	
1	AAGZFGSRAZMRCS_on_P18031_WT	High	ChEMBL30	
2	AATACZKOPAGNPL_on_P18031_WT	High	ChEMBL30	
3	AAWBMDOJDQKGGQ_on_P18031_WT	High	ChEMBL30	
4	ABDQNRWCWGSGBQ_on_P18031_WT	High	ChEMBL30	

	CID	\
0	CHEMBL363338;44397840;CHEMBL363338;44397840;CH...	
1	CHEMBL511000;CHEMBL511000;44563281;CHEMBL51100...	
2		CHEMBL4160013
3		CHEMBL3759405
4		CHEMBL197929

	SMILES	connectivity	\
0	CC(C)CC(CC(=O)C(Cc1ccc(OC(F)(F)C(=O)O)cc1)NC(=...	AADVGRQCQZZZNH	
1	CC1(C)CCC2(C(=O)Nc3ccc(C(=O)O)cc3)CCC3(C)C(=CC...	AAGZFGSRAZMRCS	
2	CCOc1ccc(OC(=O)CSc2nnc(-c3cc(O)c(O)cc3)o2)cc1	AATACZKOPAGNPL	
3	CCC0c1ccc(-c2cccc(-c3noc(Cc4c[nH]c5cccc45)n3)...	AAWBMDOJDQKGGQ	
4	O=C(O)C0c1ccc(S(=O)(=O)N(Cc2ccc(C(F)(F)P(=O)(O...	ABDQNRWCWGSGBQ	

	InChIKey	\
0	AADVGRQCQZZZNH-UHFFFAOYSA-N	
1	AAGZFGSRAZMRCS-UHFFFAOYSA-N	
2	AATACZKOPAGNPL-UHFFFAOYSA-N	
3	AAWBMDOJDQKGGQ-UHFFFAOYSA-N	
4	ABDQNRWCWGSGBQ-UHFFFAOYSA-N	

	InChI	\
0	InChI=1S/C45H47F2N3O10/c1-27(2)22-30(41(48)53)...	
1	InChI=1S/C37H53NO4/c1-32(2)18-20-37(31(42)38-2...	
2	InChI=1S/C18H16N2O6S/c1-2-24-12-4-6-13(7-5-12)...	
3	InChI=1S/C25H22N4O2/c1-2-14-30-19-12-10-17(11-...	
4	InChI=1S/C21H20F2N08PS2/c22-21(23,33(27,28)29)...	

	InChI_AuxInfo	target_id	...	\
0	"AuxInfo=1/1/N:1,3,37,36,38,49,56,48,55,35,39,...	P18031_WT	...	
1	"AuxInfo=1/1/N:1,3,34,35,28,22,40,12,17,11,18,...	P18031_WT	...	
2	"AuxInfo=1/0/N:1,2,24,5,27,6,26,23,18,11,17,4,...	P18031_WT	...	
3	"AuxInfo=1/0/N:1,2,25,24,11,26,23,10,12,7,30,6...	P18031_WT	...	
4	"AuxInfo=1/1/N:31,30,16,27,17,26,7,35,8,34,32,...	P18031_WT	...	

	Descriptor_MorganFP_1016	Descriptor_MorganFP_1017	Descriptor_MorganFP_1018	\
0	0.0	0.0	1.0	
1	0.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	

	Descriptor_MorganFP_1019	Descriptor_MorganFP_1020	Descriptor_MorganFP_1021	\
0	1.0	0.0	0.0	
1	1.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	

	Descriptor_MorganFP_1022	Descriptor_MorganFP_1023	TSNE_1	TSNE_2
0	0.0	0.0	29.735216	-4.313994
1	0.0	0.0	-0.945723	55.179680
2	0.0	0.0	-0.630247	-24.217190
3	0.0	0.0	17.455896	-43.069510
4	0.0	0.0	37.058834	5.535724

[5 rows x 1056 columns]

This is a table you can already work with, but the `dataset` object is an instance of the `scaffviz.data.dataset.DataSetTSV` class that supports some more cheminformatics features that might be of interest:

```
[9]: type(dataset)
```

```
[9]: scaffviz.data.dataset.DataSetTSV
```

For example, we can easily attach predefined or our own molecular descriptors if we need to (however, note that `Papyrus` class itself also allows you to define descriptors to fetch with the data):


```
[10]: from scaffviz.clustering.descriptors import MorganFP

# add Morgan fingerprints to the data set -> will be used to calculate the
# t-SNE embedding for chemical space visualization
dataset.addDescriptors([MorganFP(radius=2, nBits=1024)], recalculate=False)
dataset.getDescriptors()
```

```
[10]:
```

	Descriptor_MorganFP_0	Descriptor_MorganFP_1	Descriptor_MorganFP_2	\
0	0.0	1.0	0.0	
1	0.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	
...	
1767	0.0	1.0	0.0	
1768	0.0	0.0	0.0	
1769	0.0	0.0	0.0	
1770	0.0	0.0	0.0	
1771	0.0	1.0	0.0	

	Descriptor_MorganFP_3	Descriptor_MorganFP_4	Descriptor_MorganFP_5	\
0	1.0	0.0	0.0	
1	0.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	
...	
1767	0.0	0.0	0.0	
1768	0.0	0.0	0.0	
1769	0.0	0.0	0.0	
1770	0.0	0.0	0.0	
1771	0.0	0.0	0.0	

	Descriptor_MorganFP_6	Descriptor_MorganFP_7	Descriptor_MorganFP_8	\
0	0.0	0.0	0.0	
1	0.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	
...	
1767	0.0	0.0	0.0	
1768	0.0	0.0	0.0	
1769	0.0	0.0	0.0	
1770	0.0	0.0	0.0	
1771	0.0	0.0	0.0	

Descriptor_MorganFP_9 ... Descriptor_MorganFP_1014 \

0	0.0	...	0.0
1	0.0	...	0.0
2	0.0	...	0.0
3	0.0	...	0.0
4	0.0	...	0.0
...
1767	0.0	...	0.0
1768	0.0	...	0.0
1769	0.0	...	0.0
1770	0.0	...	0.0
1771	0.0	...	0.0

	Descriptor_MorganFP_1015	Descriptor_MorganFP_1016	\
0	0.0	0.0	
1	0.0	0.0	
2	0.0	0.0	
3	0.0	0.0	
4	0.0	0.0	
...	
1767	0.0	0.0	
1768	0.0	0.0	
1769	0.0	0.0	
1770	0.0	0.0	
1771	0.0	0.0	

	Descriptor_MorganFP_1017	Descriptor_MorganFP_1018	\
0	0.0	1.0	
1	0.0	0.0	
2	0.0	0.0	
3	0.0	0.0	
4	0.0	0.0	
...	
1767	0.0	0.0	
1768	0.0	0.0	
1769	0.0	0.0	
1770	0.0	0.0	
1771	0.0	0.0	

	Descriptor_MorganFP_1019	Descriptor_MorganFP_1020	\
0	1.0	0.0	
1	1.0	0.0	
2	0.0	0.0	
3	0.0	0.0	
4	0.0	0.0	
...	
1767	1.0	0.0	
1768	1.0	0.0	

1769	0.0	0.0
1770	1.0	0.0
1771	0.0	0.0

	Descriptor_MorganFP_1021	Descriptor_MorganFP_1022 \
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0
...
1767	0.0	0.0
1768	0.0	0.0
1769	0.0	0.0
1770	0.0	0.0
1771	0.0	0.0

	Descriptor_MorganFP_1023
0	0.0
1	0.0
2	0.0
3	0.0
4	0.0
...	...
1767	0.0
1768	0.0
1769	0.0
1770	0.0
1771	0.0

[1772 rows x 1024 columns]

If we ever want to use this class again to modify our data, we can recreate it like so:

```
[11]: from scaffviz.data.dataset import DataSetTSV

DataSetTSV('data/PTP1B_LIGANDS.tsv').asDataFrame().head()
```

```
[11]: Activity_ID Quality source \
0 AADVGRQCQZZZNH_on_P18031_WT High ChEMBL30
1 AAGZFGSRAZMRCS_on_P18031_WT High ChEMBL30
2 AATACZKOPAGNPL_on_P18031_WT High ChEMBL30
3 AAWBMDOJDQKGGQ_on_P18031_WT High ChEMBL30
4 ABDQNRWCWGSGBQ_on_P18031_WT High ChEMBL30

CID \
0 CHEMBL363338;44397840;CHEMBL363338;44397840;CH...
```

1	CHEMBL511000;CHEMBL511000;44563281;CHEMBL51100...
2	CHEMBL4160013
3	CHEMBL3759405
4	CHEMBL197929

	SMILES	connectivity	\
0	CC(C)CC(CC(=O)C(Cc1ccc(OC(F)(F)C(=O)O)cc1)NC(=...	AADVGRQCQGZZZNH	
1	CC1(C)CCC2(C(=O)Nc3ccc(C(=O)O)cc3)CCC3(C)C(=CC...	AAGZFGSRAZMRCS	
2	CCOc1ccc(OC(=O)CSc2nnc(-c3cc(O)c(O)cc3)o2)cc1	AATACZKOPAGNPL	
3	CCC0c1ccc(-c2cccc(-c3noc(Cc4c[nH]c5cccc45)n3)...	AAWBMDOJDQKGGQ	
4	O=C(O)COc1ccc(S(=O)(=O)N(Cc2ccc(C(F)(F)P(=O)(O...	ABDQNRWGSGBQ	

	InChIKey	\
0	AADVGRQCQGZZZNH-UHFFFAOYSA-N	
1	AAGZFGSRAZMRCS-UHFFFAOYSA-N	
2	AATACZKOPAGNPL-UHFFFAOYSA-N	
3	AAWBMDOJDQKGGQ-UHFFFAOYSA-N	
4	ABDQNRWGSGBQ-UHFFFAOYSA-N	

	InChI	\
0	InChI=1S/C45H47F2N3O10/c1-27(2)22-30(41(48)53)...	
1	InChI=1S/C37H53N04/c1-32(2)18-20-37(31(42)38-2...	
2	InChI=1S/C18H16N2O6S/c1-2-24-12-4-6-13(7-5-12)...	
3	InChI=1S/C25H22N4O2/c1-2-14-30-19-12-10-17(11-...	
4	InChI=1S/C21H20F2N08PS2/c22-21(23,33(27,28)29)...	

	InChI_AuxInfo	target_id	...	\
0	"AuxInfo=1/1/N:1,3,37,36,38,49,56,48,55,35,39,...	P18031_WT	...	
1	"AuxInfo=1/1/N:1,3,34,35,28,22,40,12,17,11,18,...	P18031_WT	...	
2	"AuxInfo=1/0/N:1,2,24,5,27,6,26,23,18,11,17,4,...	P18031_WT	...	
3	"AuxInfo=1/0/N:1,2,25,24,11,26,23,10,12,7,30,6...	P18031_WT	...	
4	"AuxInfo=1/1/N:31,30,16,27,17,26,7,35,8,34,32,...	P18031_WT	...	

	Descriptor_MorganFP_1016	Descriptor_MorganFP_1017	Descriptor_MorganFP_1018	\
0	0.0	0.0	1.0	
1	0.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	

	Descriptor_MorganFP_1019	Descriptor_MorganFP_1020	Descriptor_MorganFP_1021	\
0	1.0	0.0	0.0	
1	1.0	0.0	0.0	
2	0.0	0.0	0.0	
3	0.0	0.0	0.0	
4	0.0	0.0	0.0	

	Descriptor_MorganFP_1022	Descriptor_MorganFP_1023	TSNE_1	TSNE_2
0	0.0	0.0	29.735216	-4.313994
1	0.0	0.0	-0.945723	55.179680
2	0.0	0.0	-0.630247	-24.217190
3	0.0	0.0	17.455896	-43.069510
4	0.0	0.0	37.058834	5.535724

[5 rows x 1056 columns]

1.2 Chemical Space Visualization

Now, we can do our first data analysis of this data set, which will be a visualization of the chemical space covered by the compounds tested against the A2A receptor in public literature. In order to do that, we will use another feature of **scaffviz**, which is interactive plotting with the **molplotly** package. We can configure the visualization in many different ways, but for now we just want to use the Morgan fingerprints we calculated to calculate a t-SNE embedding to place similar compounds in the same parts of the plot. We also color the points based on the pChEMBL median value extracted from the Papyrus data set. The higher this value, the more affinity/effect the compound should have on the A2A receptor. Generally, hits with pChEMBL of 6.5 and higher are considered as having significant biological effect:

```
[12]: from scaffviz.depicition.plot import Plot
      from scaffviz.clustering.manifold import TSNE

      # make an interactive plot that will use t-SNE to embed the data set in 2D (all
      # current descriptors in the data set will be used)
      plt = Plot(dataset, TSNE())

      # start the server, you can open the plot in the browser
      plt.plot(
          recalculate=False,
          color_by="pchembl_value_Median",
          color_style="continuous",
          color_continuous_scale="rdylgn",
          card_data=["pchembl_value_Median", "all_doc_ids", "source"],
          title_data="Activity_ID",
          viewport_height=800,
          port=8888
      )
```

<IPython.core.display.Javascript object>

This plot you are seeing is running as a separate web application embedded into this one. So if you are running this notebook on localhost:8888, you can navigate to localhost:9090 to see the full depiction on its own page. You should be able to see clusters of similar molecules and regions of chemical space with active/inactive molecules should be discernible. Therefore, such visualization could be useful to assess how many molecules of what activity we have in the data set and get an idea how similar they are to each other. It is also possible to identify potential activity cliffs

(regions where very similar molecules have vastly different activity).

In the [next part of the tutorial](#), we will show a simple workflow where we use this data to train a simple machine learning model that can learn how to predict whether a compound is likely to be active or inactive on A2A receptor.