

AparoDB

We have generated a dataset containing graphs obtained according to stochastic (**Erdős-Rényi**) and scale-free models (**Barabási** and **Forest-Fire**), starting from small (100 vertices) and sparse graphs up to large (20,000 vertices) and dense graphs both with few or many labels. We used the python library **networkx** (for Erdős-Rényi graphs) and **igraph** (for Bababasi and Forest-Fire graphs). Altogether, our benchmark consists of 1,008 target graphs.

In the main directory the dataset is divided into two parts, containing equal graphs in the formats of **RI** (RI_data) and **VF** (VF3_data). For both folders there are three sub-folders, one for each **model**.

In this document you describe the structure only of the graphs in the RI format, for the VF format the structure is the same. At the moment the dataset contains **only direct graphs** (VF3 supports only direct graphs), but its structure is predisposed to a future addition of indirect graphs

The path of each graph (target or pattern) fully describes the features of the related graph. For each model, path and features changes.

We have generated **2 target for each possible graph** (considering all features). The file name of each target graph contains an *id* to distinguish it from the other generated graph with the same attributes and a *parameter* indicating the label range (in percentage with respect to the number of target vertices).

To each target is associated a directory with the same name (and the same path), containing all the **patterns** extracted from that graph. Each pattern is in a directory that indicates its number of vertices (variable according to the model and the size of the target) and its name indicates its density (0.1, 0.5 or ≈ 1).

Example:

targetID_labelRange.gfd → target1_10.gfd

RI_data/Erdos/directed/20000/0.1/target1_10.gfd → Erdos target graph with 20.000 vertex, p 0.1 and 10% label range.

RI_data/Erdos/directed/20000/0.1/target1_10/... → directory of pattern graphs (all related to this specific target)

RI_data/Erdos/directed/20000/0.1/target1_10/4/0.1_sub_0.gfd → pattern graphs whit 4 vertex and 0.1 density

Erdős-Rényi

The features of the target graphs based on Erdős-Rényi model are described in the following table; we have generated 464 targets in total.

<i>Number of vertices</i>	<i>Probability (p)</i>	<i>Label range</i>
100, 200, 500, 1.000,2.000, 5.000	0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4	5%, 10%, 20%, 30%
10.000	0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4	0.1%, 1%, 10%
20.000	0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3	0.1%, 1%, 10%

From these targets, up to 240 pattern graphs were extracted for each target with 4, 8, 16, 24, 32, 64, 128, 256 and 3 different pattern densities (10 patterns for each pattern density, 30 for each number of vertices). In total, 111.360 pattern graphs based on Erdős-Rényi model were generated.

Example:

RI_data/ErDOS/directed/NumberOfNode/p/targetID_LabelRange.gfd

RI_data/ErDOS/directed/20000/0.1/target1_10.gfd

Barabasi

This dataset has 384 target graphs having different number of vertices, different number of outgoing degree and different number of labels in relation to the number of vertices.

<i>Number of vertices</i>	<i>Degree (m)</i>	<i>Power</i>	<i>Label range</i>
200, 500	1, 2, 3, 6	0.5, 1, 1.5	1%, 10%
1.000, 5.000, 10.000, 20.000	1, 2, 3, 6	0.5, 1, 1.5	0.1%, 1%, 10%

28.800 patterns were extracted by varying the number of vertices (4, 8, 24, 32, 64) and density (5 patterns for each density, 15 for each number of vertices).

Example:

RI_data/Barabasi/directed/NumberOfNode/m/target-powerP-ID_LabelRange.gfd

RI_data/ErDOS/directed/20000/1/target-power1-1_10.gfd

ForestFire

For this model 160 target graphs were generated by varying the number of vertices, the forward burning probability and the label range.

<i>Number of vertices</i>	<i>Forward burning probability</i>	<i>Label range</i>
200, 500,	0.1, 0.3, 0.5, 0.7, 0.9	1%, 10%
1.000, 5.000, 10.000, 20.000	0.1, 0.3, 0.5, 0.7, 0.9	0.1%, 1%, 10%

12.000 patterns were extracted by varying the number of vertices (4, 8, 24, 32, 64) and density (5 patterns for each density, 15 for each number of vertices).

Example:

RI_data/ErDOS/directed/NumberOfNode/p/targetID_LabelRange.gfd

RI_data/ErDOS/directed/20000/0.1/target1_10.gfd