

Identificación de los ficheros de las muestras:

Estos ficheros, junto a otros, se generan con la aplicación Gly a partir del fichero descargado de Hyperleda. <http://leda.univ-lyon1.fr/fullsql.html> (output as CSV separator other ;)

HyperLEDA is a database and a collection of tools to study the physics of galaxies and cosmology. The project, started in 1983, is currently maintained by a collaboration between [Observatoire de Lyon \(France\)](#) and [the Special Astrophysical Observatory \(Russia\)](#).

Hay tres formatos:

1. G_1_10_444_8_16_15_70_600000_1000000_muestra_tot_meA.csv
2. G_1_10_444_8_16_15_70_600000_1000000_muestra_70_tot_meA.csv
3. G_1_10_444_8_16_15_70_600000_1000000_galaxias_intA.csv

Formato 1. La muestra de entrenamiento contiene el 60% de las observaciones y la de prueba el 40% restante. Contienen coordenadas (x, y, z) de celdas donde la unidad es el ancho de celda, además, se indica si la celda está ocupada (columna E) y si es así por cuantas galaxias (columna n).

G_1_10_444_8_16_15_70_600000_1000000_muestra_tot_meA.csv

G_1

G_0 distancias de luminosidad

G_1 distancias comoviles

_10

Diámetro del universo en Gal

444

Número de celdas por eje: X, Y, Z

Delimitación de la corona esférica:

8_16_

Entre 6 y 14 horas de ascensión recta

15_70

Entre 15 y 70 grados de declinación

600000_1000000

Entre 600 y 1000 millones de años luz respecto a la tierra

muestra_tot_meA

muestra_tot_meA = muestra de entrenamiento hyperleda
muestra_tot_mpA = muestra de prueba hyperleda
muestra_tot_mzA = agregado de las dos anteriores
muestra_tot_az_meA = muestra de entrenamiento generada al azar
muestra_tot_az_mpA = muestra de prueba generada al azar
muestra_tot_az_mzA = agregado de las dos anteriores
muestra_tot_azaj_meA = muestra de entrenamiento generada al azar ajustada por la distancia
muestra_tot_azaj_mpA = muestra de prueba generada al azar ajustada por la distancia
muestra_tot_azaj_mzA = agregado de las dos anteriores

Formato 2. Es igual que el anterior salvo que añade el porcentaje de observaciones en la muestra de entrenamiento:

G_1_10_444_8_16_15_70_600000_1000000_muestra_70_tot_meA.csv

70

La muestra de entrenamiento contiene el 70% de las observaciones.

Formato 3. Similar a los anteriores.

G_1_10_444_8_16_15_70_600000_1000000_galaxias_intA.csv

Contiene tantas observaciones como galaxias hay en el sector. Para cada observación (galaxia) registra las coordenadas de la celda en la que está incluida, todas las observaciones contienen un 1 en la columna de ocupación (columna E) y un 0 o un 1 el que corresponde al número de galaxias de los formatos 1 y 2 (columna n), el 1 significa que la galaxia pertenece a la muestra de entrenamiento y el 1 a la de prueba. También puede contener otros datos (objname;pgc;al2000;de2000;r)