Ecole Polytechnique Federale de Lausanne

A Network tour of Data Science
Project Proposal
A.Weber, L.Gauchoux, L.Loiselle

1 Introduction

Satellites are omnipresent in space and are used for a vast array of applications. We sometime hear about big satellite projects, such as the ESA Iridium or the International Space Station, but the scale and the scope of the current satellite infrastructure is not well known.

In this project, we will attempt to "democratize" satellites by analysing unclassified satellite information provided by space-track.org. On this website a wealth of information can be found: satellites currently in orbit, historical satellite launches and decay, satellite positions updated every 30 days and space debrits.

This dataset gives us broad possibilities on how we can create networks and analyse them to extract meaningfull informations.

2 Data Acquisition and Exploration

The dataset comes from the website "space-track.org". To extract the data, we needed to create an account, but this is free and easy to do.

From the website, we extracted datasets that we found meaningfull for the task at hand. Those datasets are as follow:

- **Satellite current TLE:** The Two Line Element (TLE) format describes the main satellites information. It provides current orbital status and position. It also gives the satellite designator that can be used to extract more informations about the satellite.
- **Satellite in orbit info:** This file is given in a csv format. It gives all the satellites currently in orbits. It gives the NORAD_CAT_ID, so it can easily be linked to the TLE. Also, it gives information on the satelitte name, the launch date and the country of launch which will be valuable in our analysis.
- **Decayed satellites info:** This file is given in a csv format. It gives the historical decay of satelitte and their launcher since the SPUTNIK 1 launch. It gives orbital information, satellite size, name, NORAD_CAT_ID, launch date, decay date and country of origin.
- **Satellite statistics by country:** This file is given in a csv format. It gives information per country on various metrics: the number of satellites in orbit, the number of debris, the number of rocket bodies, the same information for the decayed objects and totals per country. This is interesting to give a per country view of the space programs without having to scrape the other (big) datasets.

The csv dataset extraction is very simple since we can import it directly into a panda dataframe.

The TLE dataset extraction will be more complicated, but because it is given in a standardized format, simple regex parsing will allow us to extract the data without too much trouble.

The complete dataset size is currently 7.9 MB, which we consider to be a size easily manageable on our personal computers, which is a big advantage. This contains all of the datasets described in the list above.

3 Data Exploitation

We are no astrophysicists, but we can use Python and a quick search on TLE and orbital analysis provided us with a library that we can use to get meaningfull informations from the provided TLE: PyEphem. The library allows direct analysis of TLE data (without even parsing it! We might not need regex after all!).

Now, we can use the tools viewed in the NTDS course to use the data, group it into network and analyse it.

We mentionned in the introduction that we wanted to "democratize satellites". How will we achieve this goal?

3.1 Network analysis

We will use networkx to create networks with the gathered information.

- **Network of satellites:** The first network we will create is a combined network of past and present satellites. Each node in this network will be a satellite, identified by their Norad ID. An additional "type" attribute will be added to identify those nodes as "satellites", because we will add other types of nodes afterwards.
- **Historical satellite overview:** For each year since the start of the space age, we will create a node. This type of node will have "year" as their type identifier. We will link every satellites launched during that year to this node with an edge named "launch year". We will do the same if the satellite decayed during that year, but with an edge named "decay year". We will add a node "Future" to which we will link the undecayed satellites. With this network, we will be able to show how the satellite programs evolved with time by showing the graph.
- **Orbital categorization of satellites:** We will also create a network of satellites by type of orbit. This can be easily done by using the satellite orbit information and following standard rules to determine in which category it lies (Geosynchronous (GEO), Medium Earth Orbit (MEO), Low Earth Orbit (LEO), Highly Elliptical Orbit (HEO). The filtering methodology can be found on the space-track.org website. For each of those orbits, we will create a node with the orbit abbreviation as name and the propriety "orbit" as type identifier.
- **Country of origin:** We will create a node for each country in the country dataset, the type identifier will be the country itself.
- **Constellations:** This network will create nodes identifying the main satellite constellations (Iridium, Orbcomm, Globalstar, GPS, etc.). It will use "constellation" as a type identifier. This type of network is interesting to show the scale of the modern satellite networks, especially if it is combined with a projection on a map.

The various tools learned in the NTDS class could be used on those networks to provide information. Tables that resume those findings would be a good tool to explain the relations in the graph created.

3.2 Satellite mapping

By combining the informations from the networks and the TLE, we will be able to project satellites on a world map. For now, we will project the current satellites on the world map, which correspond to the most recent TLE that we acquired.

One thing that would be interesting is to provide one such projection per decade since the start of the space age. This would clearly show the evolution of the satellites through time. Also, those satellites could be classified by type of orbit to show the developments in each of these domains.

The Folium python library will be used to do those mappings.

4 Data Analysis

Other than the network creation and the mapping that was described above, we could use the tools seen in class to cluster the different satellites.

Particularly, from their orbital information, it would be nice to see if we can use the spectral graph theory to cluster the satellites into their respective orbit, whithout using the orbital description.

Another analysis that would be interesting is using the epidemic model of networks to see if it can represent the development of satellite technology through time per country.

During the project, we might also find other meaningfull ways to analyse the data.