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 ond how we can create networks and analyse them to extract meaningfull informations.

This document provides information on how we will manage this data, from the acquisition, exploration and exploitation point of view.

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. The extracted file gives the current TLE for 16731 satellites or debrits in orbit. The file is given in a text format, but it is standardized, so it will be easily scraped with regexes. A more extensive TLE dataset is provided on the website, we hitted rate limiting when downloading it, but theorically, it countains the TLE of all satellites since the beginning of the space age, with various degrees of precisions. For now we will not extract the complete historical TLE dataset, but it is an option if we want to do a more complete/complex analysis.

- **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.

Also, because we have access to every satellite names, we could use the wikipedia API to extract more information on each satellite if needed. Though, for now the dataset seems complete enough for our needs.

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. It should be enough to provide us with a vast range of potential analysis.

If we wanted to expand those dataset, we can always access the complete historical TLE dataset. The complete historical TLE dataset is said to contain 97 million entries, with 69 bytes per row on 2 rows, so approximatly 13 GB. To keep the dataset manageable, we will not use the complete historical TLE dataset. Though, we could provide historical tracking of satellites over a restricted range of time or a restricted quantity of satellites if we find it usefull during our project.

Additional data management could be done to separate payload, launchers and debris from the TLE. Such categorization would be usefull for the exploitation of the data. We could use the website api: http://heavens-above.com/SatInfo.aspx?satid="OUR_SAT_ID" to extract those informations for a particular item in the dataset.

3 Data Exploitation

Now that we know the dataset we will use, how can we extract meaningfull information from it?

After all, we are not astrophysicists, so how can we use the orbital informations provided?

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!). A good tutorial is provided here to get ground latitude and longitude from the TLE: goo.gl/AtpcpJ. We can also have the evelation of the satellite from the TLE.

Lets resume the information we have: name, country, launch site, date of launch, date of decay and norad id of the satellites from the decayed and in orbit satellite datasets. In addition, we have information by country of their satellites, launchers and debris, in orbit and decayed. Finally, from the TLE, we have orbital information for specific dates, which allows us to get the satellite latitude, longitude and height at a given point in time.

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. This network will be unconnected in its first stage. Our different analysis will use this base network and add edges to it to create a more complex graph. We will refer this unconnected network as the Network of Satellite Nodes (NSN).

- **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. Having labelled edges will allow us to control the drawing with more precision. We could also group the decayed satellites by years of operation. This could show how the satellite industry has evolved to create long lasting satellites.

- **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 those 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. This network will allow us to show the distribution of satellites depending on their orbit. Also, when we do satellite projection later on, we will be able to use it to select only a category of orbits to project on the map.

- **Country of origin:** We will create a node for each country in the country dataset, the type identifier will be country. Each satellite will have an edge linking it to its respective country. This will allow us to show the country distribution of the satellites.

- **Type of satellite:** This is optional, but we could also classify the objects by their type (debris, launcher or satellite). This would require web crawling as we have said above.

- **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.

3.3 Various 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.

4 Conclusion

In this project proposal, we have described a methodology to tell the story of the space age by using readily available datasets and libraries, combined with the tools learned in the NTDS course.

This analysis will provide information on satellite development through time, country, type and constellations. It will also show the distribution of those satellites by projecting them on the world map. Finally, we will use network analysis tool to caracterize the network at hand and to attempt to find a way to simulate its progression through time.

In this proposal, we exposed methods that could be used. The extent of the methods used in the final project could vary depending on our results. Especially, we might find better ways of expressing the information than in the described methodology.