

---

# **pinnacle**

***Release 0.1***

**Marcelo Lares**

**Oct 30, 2020**



**CONTENTS:**

<b>1</b>	<b>User’s guide</b>	<b>3</b>
1.1	Installation . . . . .	3
1.2	API usage . . . . .	3
<b>2</b>	<b>Configuration</b>	<b>5</b>
<b>3</b>	<b>Investigadores del IATE</b>	<b>7</b>
<b>4</b>	<b>Informe del estado del IATE</b>	<b>9</b>
<b>5</b>	<b>Metricas de producción científica</b>	<b>15</b>
<b>6</b>	<b>Comparación con otros Institutos</b>	<b>17</b>
<b>7</b>	<b>PINNACLE module</b>	<b>19</b>
<b>8</b>	<b>pub_dataviz module</b>	<b>21</b>
<b>9</b>	<b>Indices and tables</b>	<b>23</b>
	<b>Python Module Index</b>	<b>25</b>
	<b>Index</b>	<b>27</b>





This is the documentation for PINNACLE (Publications from an Institute: Numbers, Networks, Authors and Citations Looking for Excellence.)

It aims at exploring the publication metrics, relations and networks for researchers at an institute, in order to design strategies to increase impact and productivity.



## USER'S GUIDE

The code is based on the `ads` package, that allows to perform queries through a python code. Objects in `pinnacle` include the `Article` class from `ads`.

We aim at exploring the publication metrics, relations and networks for researchers at an institute, in order to design strategies to increase impact and productivity.

The process basically consists on creating four datasets:

- `pub_auth_all`
- `pub_auth_top`
- `pub_inst_all`
- `pub_inst_top`

along with entries with data for the institute and the authors:

- `history`
- `staff`

### 1.1 Installation

To install the [latest stable](#) version from the Python Package Index:

```
pip install pinnacle_pub
```

To install the development version, [download the last version](#) of the code from the GitHub page, and run:

```
pip install .
```

### 1.2 API usage

The analysis of a set of authors can be made easily with the following steps.

First, load the package:

```
from pinnacle import pinnacle
```

For the configuration, edit a configuration file following the template in the `set` directory, and load it with the `Parser`.

```
ini = 'my_institute.ini'
config = Parser(ini)
df = pinnacle.inst_adsentries(config)
```

The set of the names of the researchers can be loaded with different tools, from Excell spreadsheets, CSV files, or entered manually as lists.

```
DF = staff.download_inst(staff_staff)
staff.reduce_article_list(DF)
staff.eliminate_repeated('bibcode')
staff.journal_quality()
staff.load_history(nstaff, int(columns[1]))
staff.save_inst()
```

It not necessary to make this every time, since pickle files are saved. After the first run, the dataset can be loaded simply using:

```
staff.load_inst()
```

The number of papers per author per year can be obtained in a dataframe or an XLSX file with:

```
staff.save_table()
```

Finally, plotting tools are available with the `pub_dataviz` class, that inherits the data from a `inst_adsentries` class:

```
viz = pub_dataviz.pub_dataviz(df)
```

There are several plots that can be obtained. The full set is produced with:

```
viz.plot_all()
```

Or, alternatiely, individual plots using the function in the class.



## CONFIGURATION

All the parameters for the analysis of a set of authors (e.g., an Institute) can be maintained in one place with a configuration file.

These files are loaded and used as follows:

```
ini = 'configuration_file.ini'
config = Parser(ini)
df = pinnacle.inst_adsentries(config)
```

The content of a configuration file has several sections (the name of these sections should not be changed).

The first section contains the name for a given experiment. This is used for the creation of data files and plots filenames:

```
# _____
[experiment] # EXPERIMENT ID

# Experiment ID. Useful to compare and save experiments.
# A directory will be created with this name under [out]dir_output
experiment_ID = IATE_ALL
```

The directories and filenames for data sources and outputs are set in the files section:

```
# _____
[files] # DIRECTORIES AND FILE NAMES

# directory for data
dir_data = ../dat

# directory for plots
dir_plot = ../plt/iate/

# reload ADS query?
qreload = False

# overwrite pickle files?
clobber = False

# maximum number of ADS entries
nrowsmax = 200

# filename for staff
fname_staff = staff_iate_iate.csv
```

(continues on next page)

(continued from previous page)

```
# load from file ("file") or download from ADS ("ads")
data_source = file

# roots for data filenames
fname_pub_auth_all = pub_auth_all
fname_pub_auth_top = pub_auth_top
fname_pub_inst_all = pub_inst_all
fname_pub_inst_top = pub_inst_top
fname_pub = pub
```

The selection of authors by affiliation keywords or keyphrases is made with a number of strings in the `keys` section:

```
# _____
[keys] # KEY STRINGS FOR AFFILIATION FILTERING

# add as many strins as needed, but use a different index (aff1, aff2, ...)
aff1 = IATE
aff2 = Córdoba
aff3 = Laprida 854
aff4 = X5000BGR
aff5 = Universidad Nacional de Córdoba
aff6 = Instituto de Astronomía Teórica y Experimental
```

Finally, some flags for the behavior of the standard output:

```
# _____
[UX] # USER EXPERIENCE

# Show progress bars
# options: Y/N
show_progress = y

# Show messages for partial computations
# options: Y/N
verbose = y

# Return objects (N: only write to files)
# options: Y/N
interactive = n
```

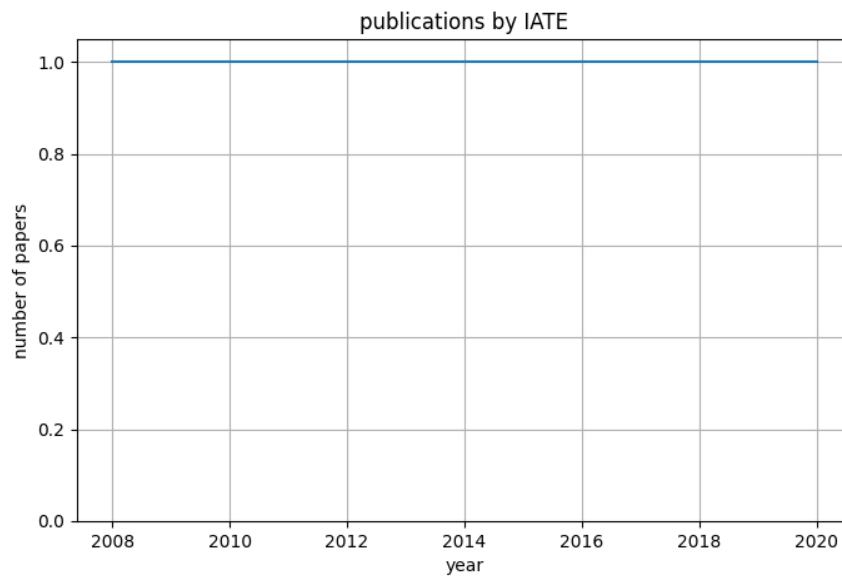
**INVESTIGADORES DEL IATE**

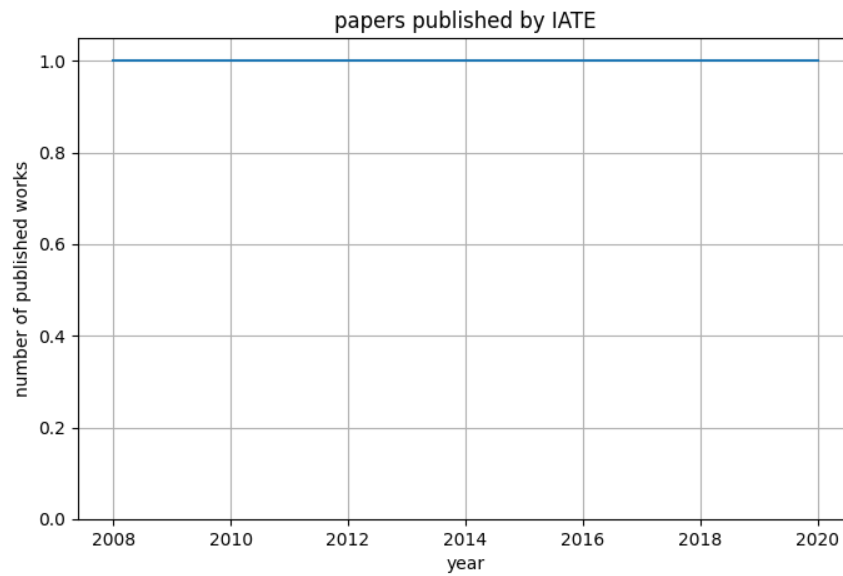
Abadi, Mario Alonso, María Victoria Beaugé, Cristian Bornancini, Carlos Ceccarelli, María Laura Cécere, Mariana Coenda, Valeria Díaz, María Eugenia Dominguez, Mariano Donzelli, Carlos Ferreiro, Diego Lambas, Diego Garcia Giuppone, Cristian Gonzalez, Elizabeth Gurovich, Sebastián Lares, Marcelo Luparello, Heliana Martínez, Héctor Julián Merchán, Manuel Muriel, Hernán O'Mill, Ana Laura Parisi, María Celeste Paz, Dante Racker, Juan Diego Figueroa, Cinthia Ragone Ruiz, Andrés Schmidt, Eduardo Schneider, Matías Sgró, Mario Agustín Stasyszyn, Federico Valotto, Carlos Neme, Luis Vega Villarreal D'Angelo, Carolina Susana Yaryura, Yamila Zandivarez, Ariel



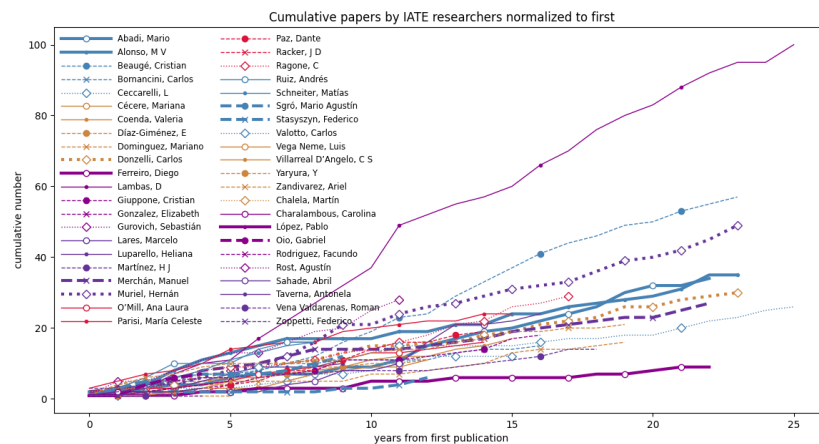
## INFORME DEL ESTADO DEL IATE

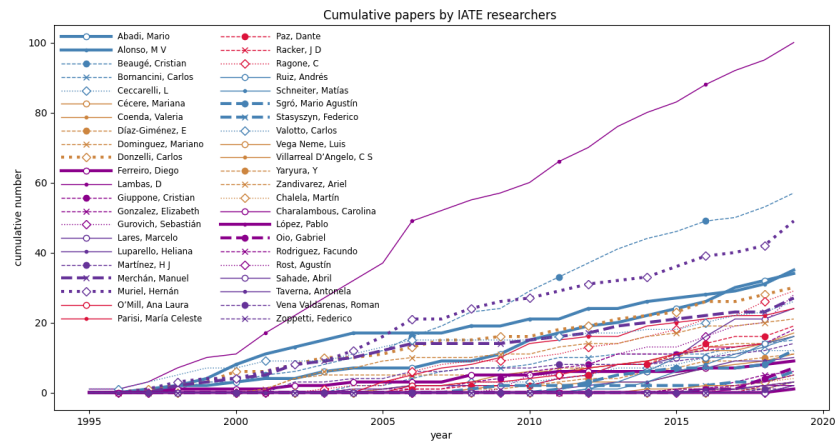
Numero de papers con autores investigadores, para todas las publicaciones y para las revistas más importantes (i.e., excluyendo proceedings y BAAA):



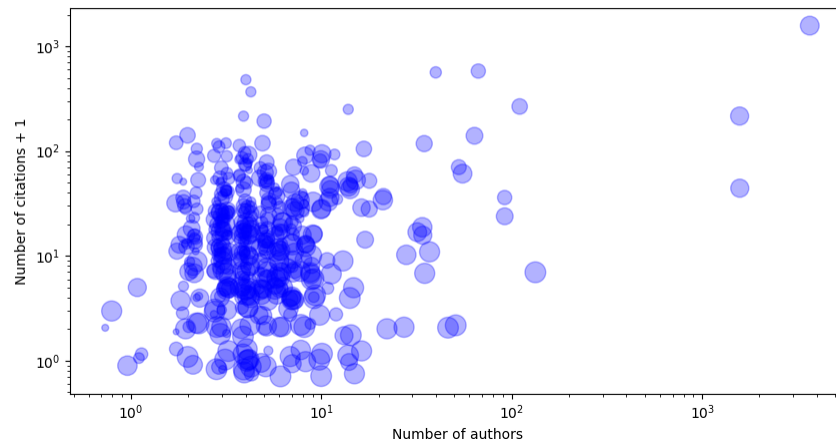


Evolución de las publicaciones por autor (hay papers que aparecen para varios autores), contando el tiempo en años o contando desde el año de la primera publicación.



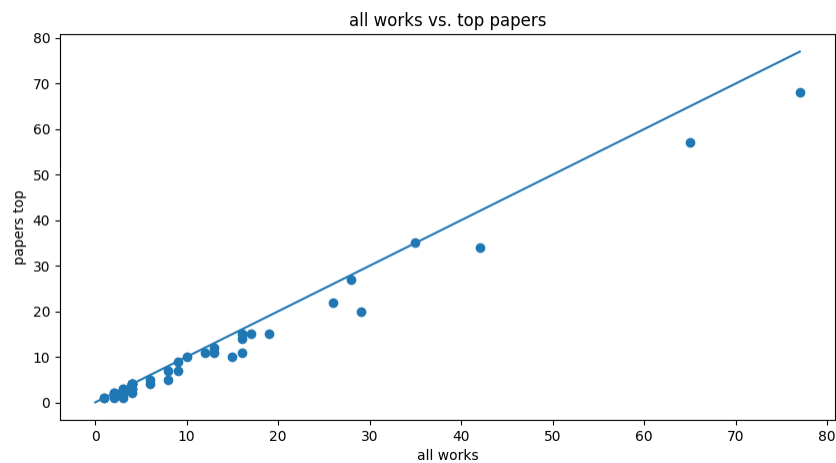


Relación entre la cantidad de citas y el número de autores:

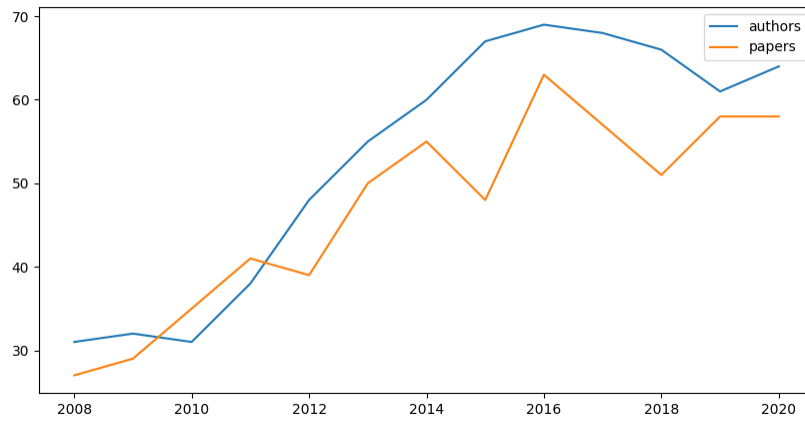


en general, los papers con más autores tienen más citas.

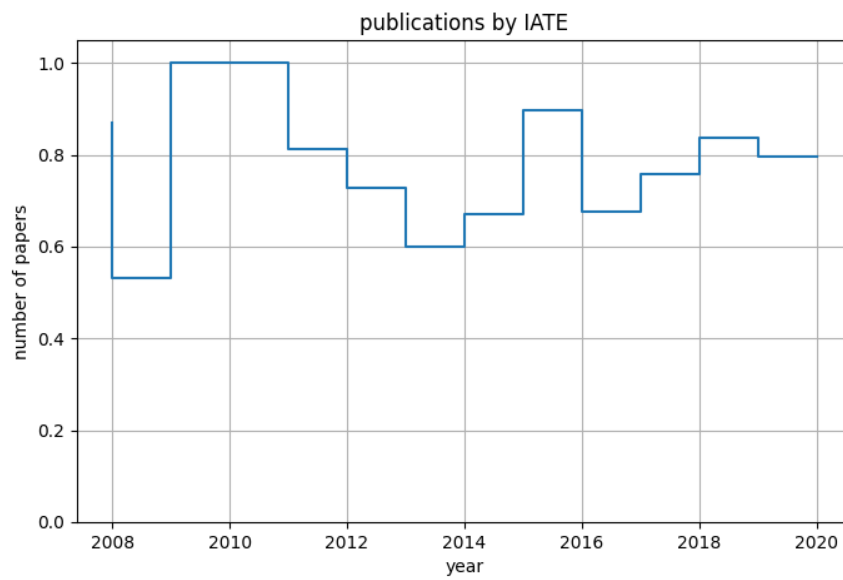
Relación entre la cantidad de publicaciones y el número de papers en revistas top.



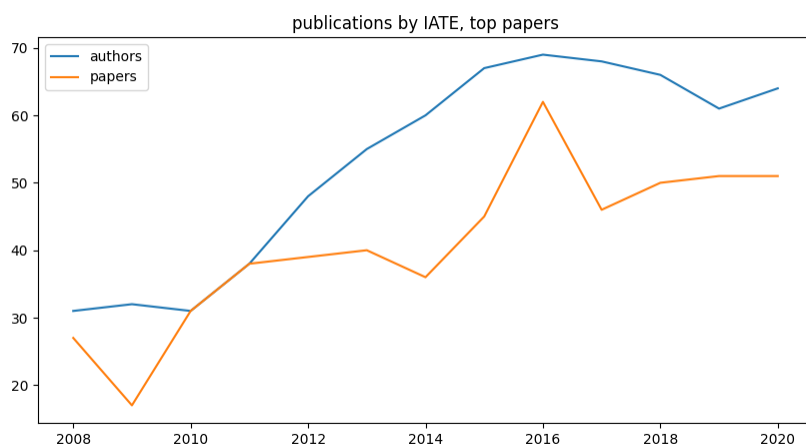
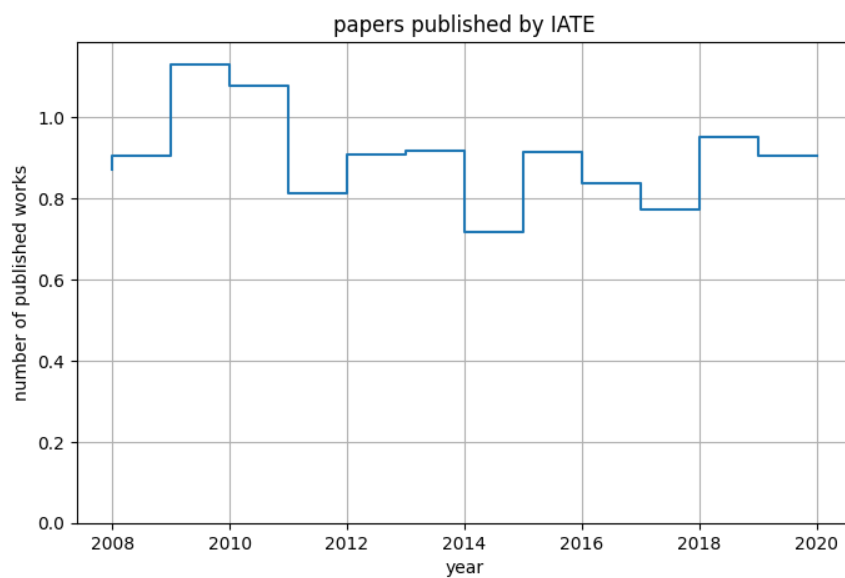
Evolución de la cantidad de autores en función del tiempo:



se puede ver que hay un leve incremento en la cantidad de autores de los papers.









## **METRICAS DE PRODUCCIÓN CIENTÍFICA**



## COMPARACIÓN CON OTROS INSTITUTOS



## PINNACLE MODULE

**PINNACLE: (Publications from an Institute: Numbers, Networks, Authors and Citations Looking for Excellence)**

Software to analyze publication metrics on author sets.





## PUB\_DATAVIZ MODULE

```
class pub_dataviz.pub_dataviz (inst)
    Bases: object

    authors_citations_years (top=True)
        Plot a scatter of number of authors and number of citations

        Parameters top (bool) – Use all works or papers from selected journals

    cumulative_per_author (top=False, normalize_first=False)
        Parameters
            • top (bool) – Use all works or papers from selected journals
            • normalize_first (bool) – Normalize to the year of the first publication

    nauth_npprs (top=True)

    number_authors (top=True)
        Plot a scatter for the number of authors as a function of time

        Parameters top (bool) – Use all works or papers from selected journals

    papers_histogram (top=False, per_auth=False, quality=5)
        Papers_histogram: histogram of the years of publications

        Parameters top (bool) – If True, paper in selected journals are used, otherwise, all papers.

    papers_histogram2 (top=False, per_auth=False)
        Papers_histogram: histogram of the years of publications

        Parameters top (bool) – If True, paper in selected journals are used, otherwise, all papers.

    plot_all ()
        Make all the plots.

    top_proceedings ()
        Plot a scatter of number of publications vs number of papers
```



## INDICES AND TABLES

- `genindex`
- `modindex`
- `search`



## PYTHON MODULE INDEX

### p

pinnacle, [19](#)  
pub\_dataviz, [21](#)



## INDEX

### A

`authors_citations_years()`  
(*pub\_dataviz.pub\_dataviz method*), 21

### C

`cumulative_per_author()`  
(*pub\_dataviz.pub\_dataviz method*), 21

### M

`module`  
    *pinnacle*, 19  
    *pub\_dataviz*, 21

### N

`nauth_npprs()` (*pub\_dataviz.pub\_dataviz method*),  
21  
`number_authors()` (*pub\_dataviz.pub\_dataviz*  
*method*), 21

### P

`papers_histogram()` (*pub\_dataviz.pub\_dataviz*  
*method*), 21  
`papers_histogram2()` (*pub\_dataviz.pub\_dataviz*  
*method*), 21  
`pinnacle`  
    *module*, 19  
`plot_all()` (*pub\_dataviz.pub\_dataviz method*), 21  
`pub_dataviz`  
    *module*, 21  
`pub_dataviz` (*class in pub\_dataviz*), 21

### T

`top_proceedings()` (*pub\_dataviz.pub\_dataviz*  
*method*), 21