Social Data Science: Machine Learning & Econometrics

Exercise class 7

April 3, 2020

Todays quick warmup

Q: The number of primes less than n, $\pi(n)$ are well approximated by the logarithmic integral

$$\pi(n) \approx Li(n) = \int_2^n \frac{1}{\log t} dt \tag{1}$$

Write a function `Integrate(f,a,b)` that uses `scipy.integrate.quad` to integrate `f` from `a` to `b`. Use this function to plot $Li(n)$ along 100 values of $n \in [10, 10^{10}]$.

Pick a stylesheet for your matplotlib figure from here (or make your own):

https://github.com/topics/matplotlib-style-sheets

and properly format your figure with axis labels, a legend etc.

Todays quick warmup

This code should get you started.

```
from scipy import integrate
from numpy import log, linspace
import matplotlib.pyplot as plt

def Integrate(f, a, b):
    I, _ = integrate.quad(f, a, b)
    return I
```

Last lecture in a nutshell

- Networks are everywhere - also in social science!
- In mathematics networks are called *graphs*, they consist of *vertices* and *edges*, they are written $G = (V, E)$. In network science we usually use the words *network*, *nodes* and *links*.
- Some networks are *undirected*, others are *directed*. Some are *static*, others are *temporal*.
- We can encode a static network in a size $|V|^2$ matrix called the *adjacency matrix*. (how does the adjacency matrix of an undirected network look? What changes in a directed network?)
- There is a host of statistics to calculate in networks, e.g. average degree $\langle k \rangle$, path-length statistics, clustering statistics etc.