

connectedcomponents

January 31, 2022

The Graph Structure of Public Software Development

Antoine Pietri(1), Guillaume Rousseau(2) and Stefano Zacchiroli(3)

1. Inria, Paris, France. antoine.pietri@inria.fr
2. Université de Paris, Paris, France. guillaume.rousseau@u-paris.fr
3. LTCI, Tlcom Paris, Institut Polytechnique de Paris, Paris, France. stefano.zacchiroli@telecom-paris.fr

TODO: Add citation string

1 Replication Package : Connected Component Distributions

```
[19]: from pathlib import Path
import matplotlib.pyplot as plt
import numpy as np
import common

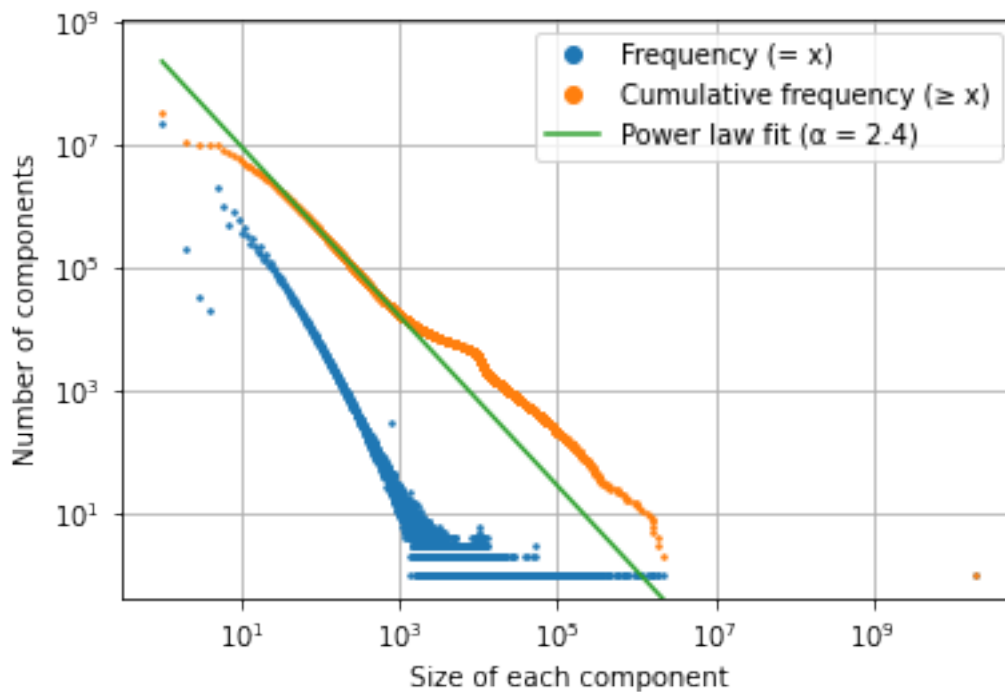
DATASET = Path('../experiments/connectedcomponents')

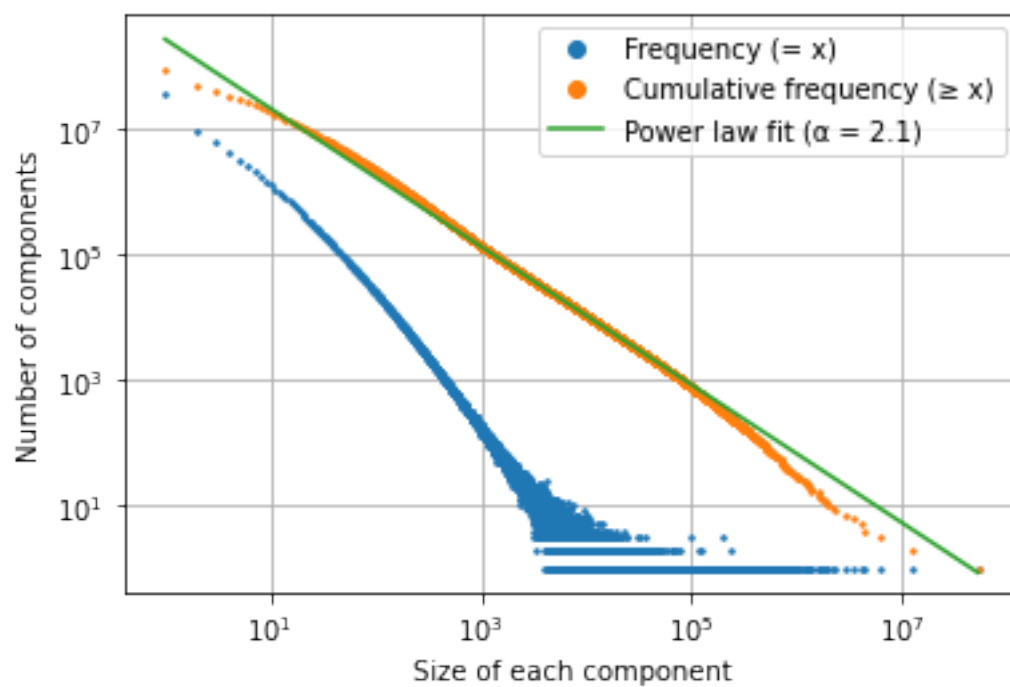
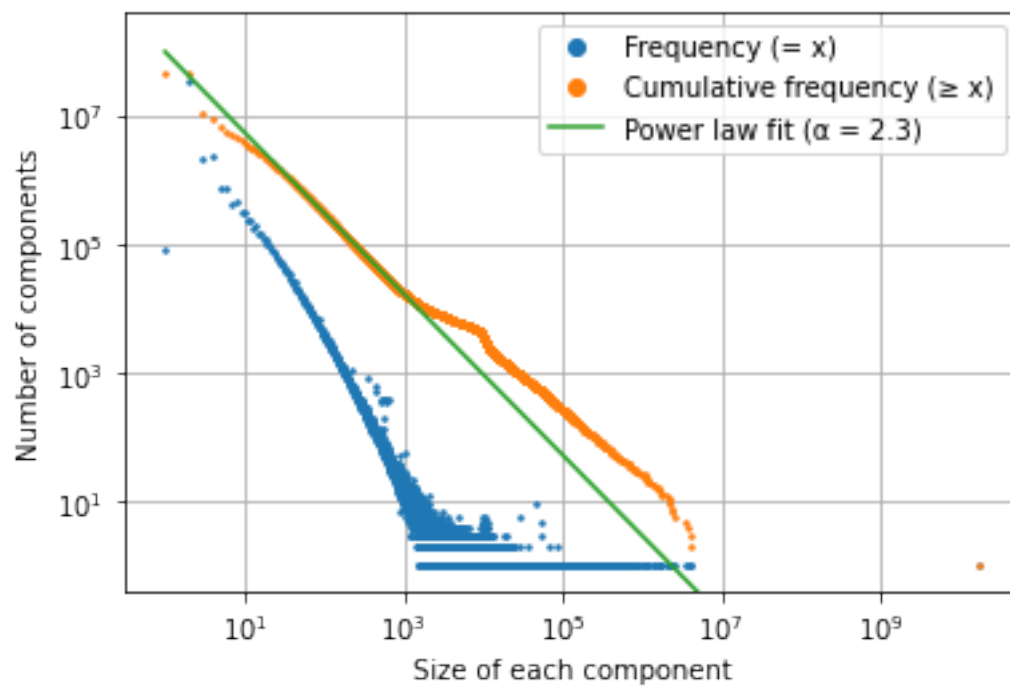
def show_distribution(name):
    title = "Connected components ({}).format(common.LAYERS[name])

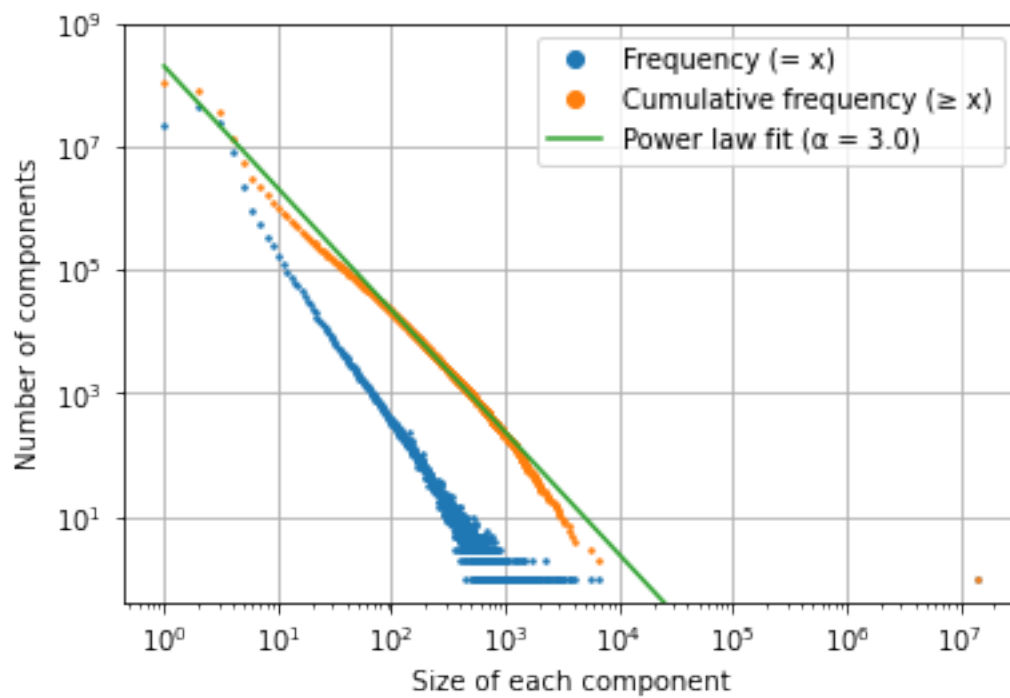
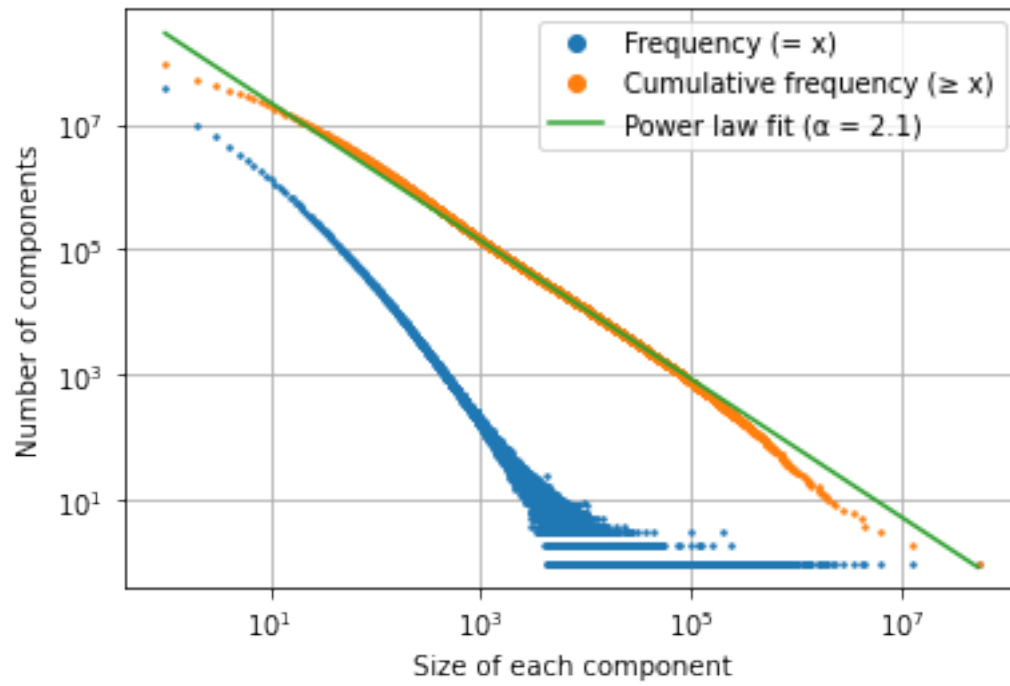
    x, y = common.load_text_distribution(DATASET / name / 'distribution.txt')
    dist = common.Distribution(x, y, "Size of each component", "Number of_
    components", title)
    print("{}: number of CC = {}, size largest = {}, percent of nodes in_
    largest = {}, percent <100 node = {}".format(
        name,
        sum(y),
        x[-1],
        x[-1] / sum(degree * count for degree, count in zip(x, y)) * 100,
        sum(y[0:100]) / sum(degree * count for degree, count in zip(x, y)) * 100,
    ))
    common.plot_all(dist, 'figures/connectedcomponents', name)
```

```
[20]: show_distribution('full')
      show_distribution('dir+cnt')
      show_distribution('rev')
      show_distribution('rel+rev')
      show_distribution('ori+snp')
```

full: number of CC = 33104255.0, size largest = 18902683142.0, percent of nodes in largest = 97.78561816828444, percent <100 node = 0.16930583000190183
 dir+cnt: number of CC = 46286502.0, size largest = 16565521611.0, percent of nodes in largest = 97.15599193231185, percent <100 node = 0.2696940075401386
 rev: number of CC = 88031649.0, size largest = 51543944.0, percent of nodes in largest = 2.6078706710155517, percent <100 node = 4.341575100539041
 rel+rev: number of CC = 88040059.0, size largest = 52176239.0, percent of nodes in largest = 2.617954146945862, percent <100 node = 4.30500753137543
 ori+snp: number of CC = 108342722.0, size largest = 13841855.0, percent of nodes in largest = 4.81813911862127, percent <100 node = 37.70532116061812







[]:

[]:

[]: