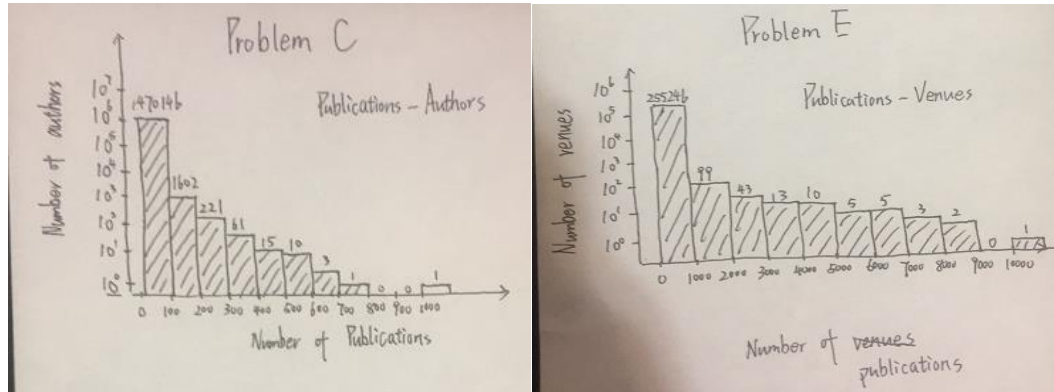


## Problem 1:

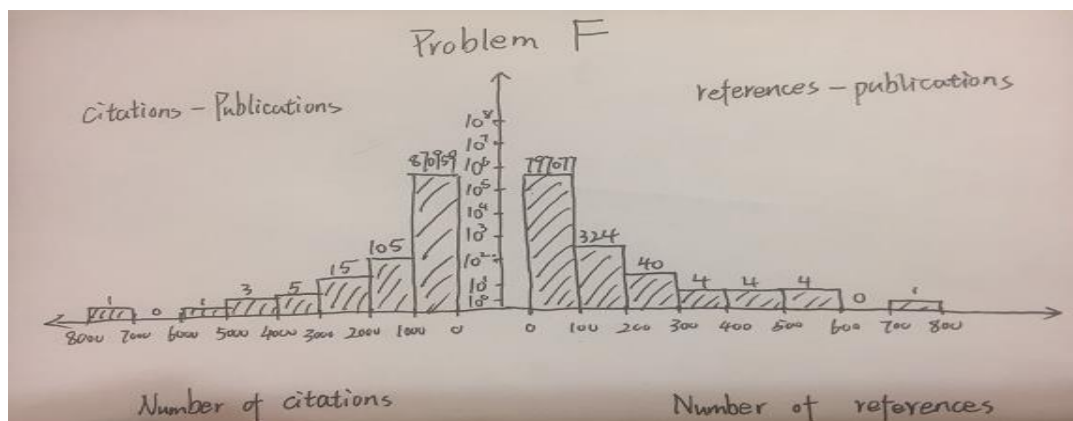
- A: Number of publications: 1976815      Number of authors: 1472722  
 Number of venues: 255681      Number of citations: 7250326
- B: Number of venues' names associated "Principles and Practice of Knowledge Discovery in Databases": 212
- C:



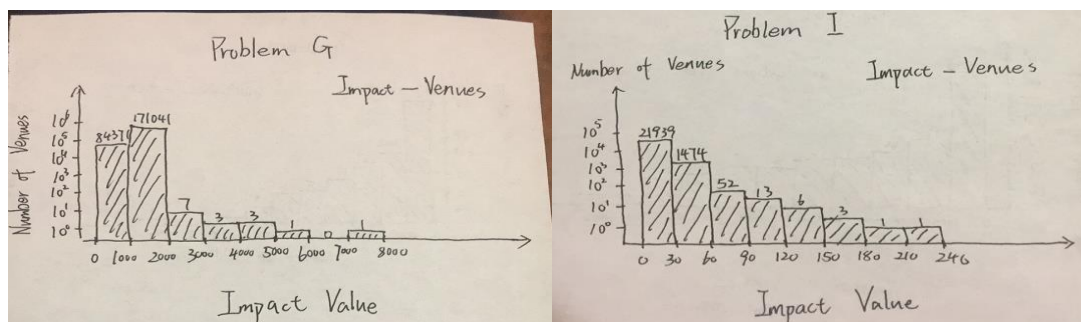
- D: min = 1.0      max = 1054.0      mean = 3.292066220126897      Q1 = 1.0  
 median = 1.0      Q3 = 2.0      stddev = 8.81585364516579
- E: min = 1.0      max = 11227.0      mean = 7.746405039404605      Q1 = 1.0  
 median = 1.0      Q3 = 1.0      stddev = 83.42253083953304

Most Publications venue: IEEE Transactions on Information Theory

- F: The publication with the largest number of citations:  
 Computers and Intractability: A Guide to the Theory of NP-Completeness: 7753
- The publication with the largest number of references:  
 Cited References: 772



G:



H: The venue with the highest apparent impact factor:

Computers and Intractability: A Guide to the Theory of NP-Completeness: 7753.0

I: The venue with the highest apparent impact factor:

Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications 214.82608695652175

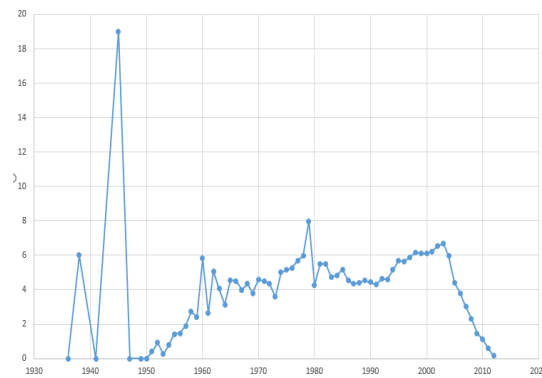
The Publications List of Max Impact Venue:

Index: 340170 340171 340167 340180 340183 340182 340178 340174 340181 340184 340185 340187 340166  
340169 340179 340177 340173 340172 340165 340168 340164 340176 340175

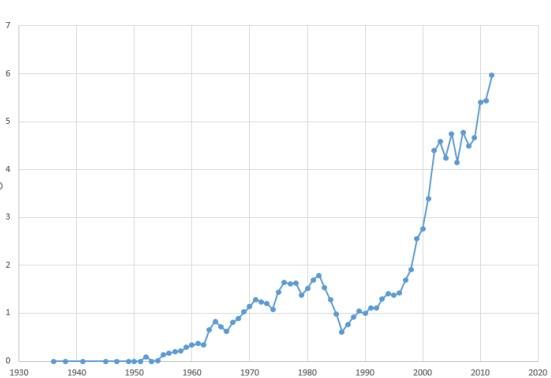
mean: 214.82608695652175 median: 46

J:

Citations:



References



## Problem 2:

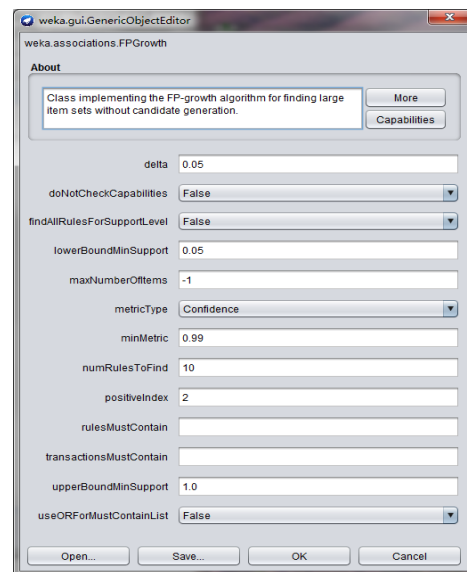
B: About 12 sec

C: About 11 sec Instances: 990002 Attributes: 41270

D: weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.99 -D 0.05 -U 1.0 -M 0.05

1. [i11=1, i218=1, i148=1]: 50098 ==> [i6=1]: 49866 <conf:(1)> lift:(1.64) lev:(0.02) conv:(84.4)

2. [i11=1, i148=1]: 55759 ==> [i6=1]: 55230 <conf:(0.99)> lift:(1.63) lev:(0.02) conv:(41.3)



E: Running1 6 sec Running2 5 sec Running3 5 sec Running4 5 sec Running5 5 sec mean: 5.2 sec

## Problem 4:

5-NN Performance: MINST accuracy: 96.93% 20NG accuracy: 73.83%