



SHANGHAI JIAO TONG UNIVERSITY, CHINA

Project Introduction

Development of authors

**Albert Wu
F1403030
5140219226**

OUTLINE

- **Motivation**
- **Implementation**
- **Future Work**

OUTLINE

- **Motivation**
- Implementation
- Future Work

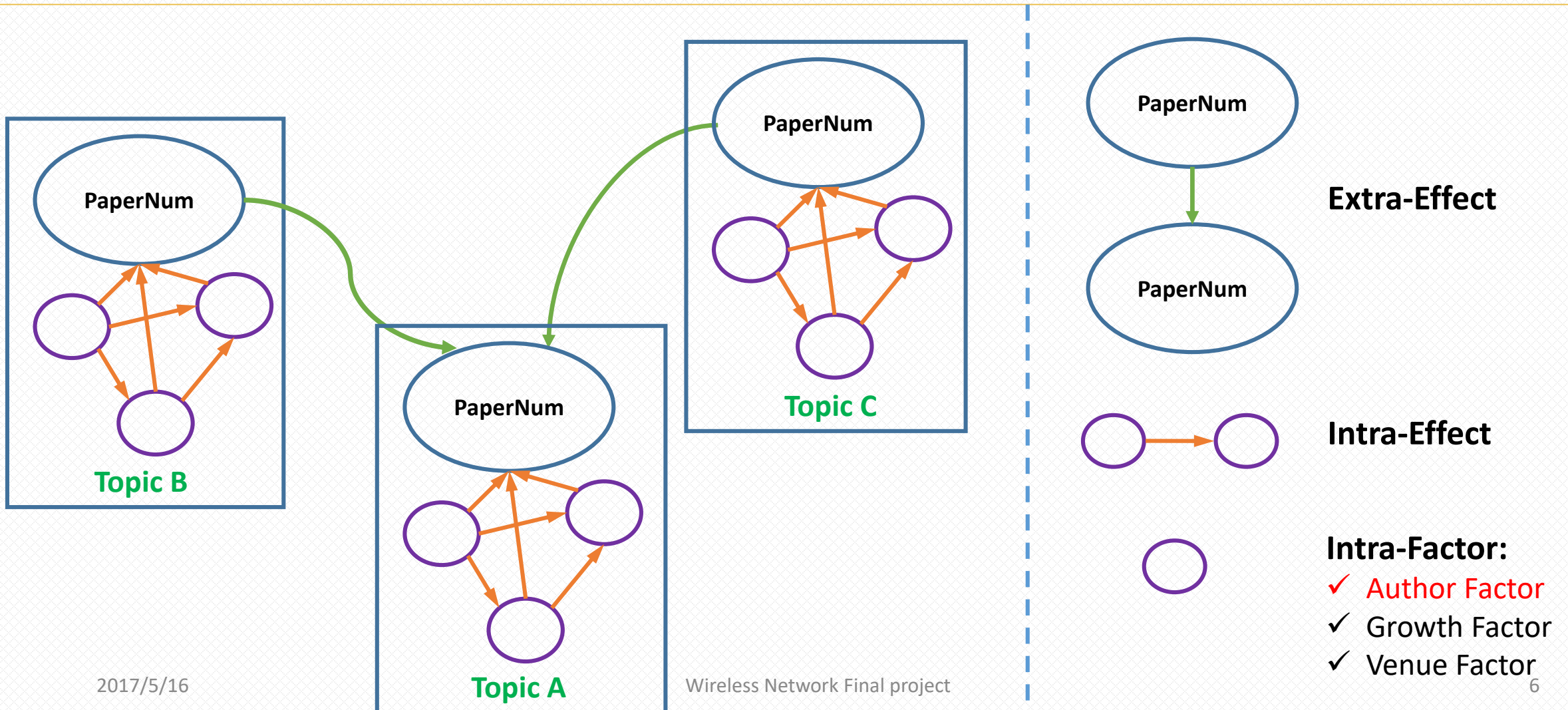
Motivation

- Current work mainly focus on the **total number** of authors and number of papers for hot topic prediction, ignoring the **difference** between individuals and **mutual influence** between topics.

AuthorDiv

- Mainly focus on how to divide authors into four categories based on the real distribution.

Prediction Model



OUTLINE

- Motivation
- **Implementation**
- Future Work

Dataset

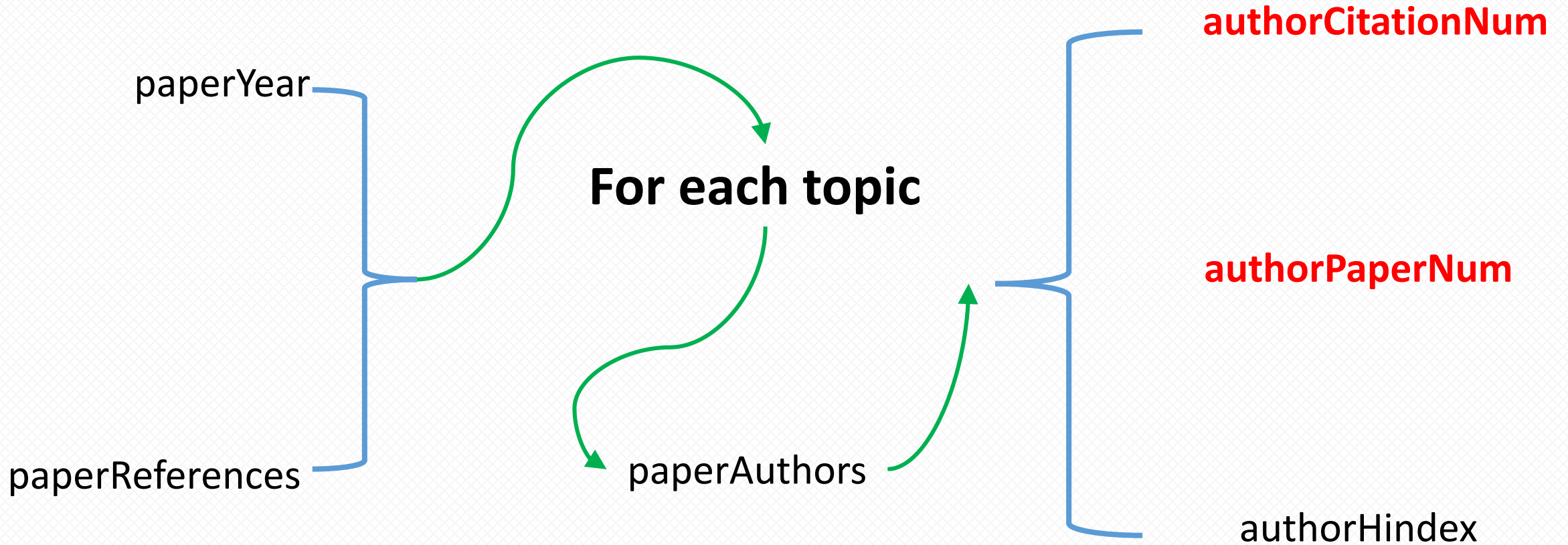
Provided by Microsoft Academic Graph

- ✓ 12,644 topics about CS
- ✓ 14.4 million authors
- ✓ 30 million papers
- ✓ 52.3 million citations

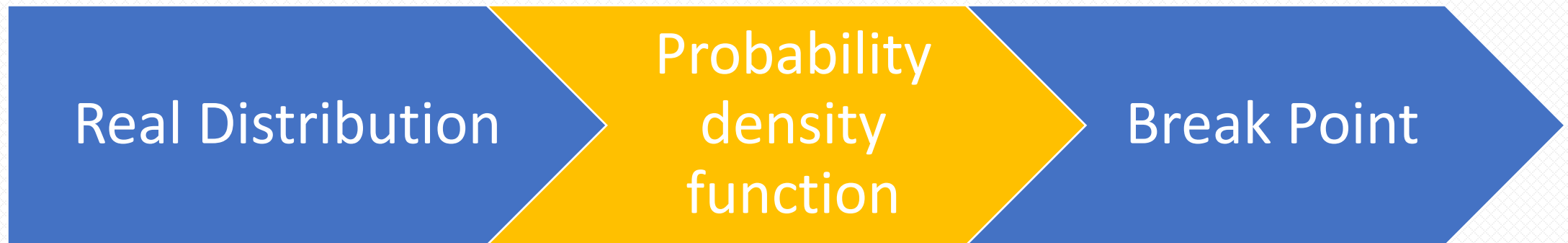
```
1 00000060$$$ 5E3BDA99
2 000002FE$$$ 781F8D75
3 00000698$$$ 7DB78D20
4 0000082C$$$ 7FF508C2
5 00000939$$$ 7D75C2AB
6 00000EBA$$$ 7A34C7AF
7 00000EBA$$$ 7B74BB2A
8 000011C0$$$ 80CF80C1
9 00001561$$$ 77B3BE94
10 00001561$$$ 7906CD45
11 0000197A$$$ 77AEC636
12 000019D7$$$ 77E0F5EE
13 00001C38$$$ 75C0DA77
14 00001EB2$$$ 7B4DE6AB
15 0000228E$$$ 7CAAC425
16 000022C1$$$ 5B82EDAD
17 00002350$$$ 7BF94F5F
18 000024C7$$$ 7D593412
```

```
1 18A42870 76A01B95 77EE0AAC 7B5525E6 7DF8018A
2 776541B8 7F99B9C0 7DDEBAEB 85012304 7FBF477E
3 7D84920E 7D8CADD A 861DE2D2 7FED5A1B 73733E08
4 7DAD9F37 7ED2E4A6 34C1B39E
5 7DC0DBAE 7DAE1EA7 7E5ADCBB 848B2BEE 7E4AC57B
6 7FB37798 80664857 77B5D4B8 7CC4E273 7E42C80A
7 7FEA53CE 7D8D61FC 7E5513A3 7E86066D 7D04E92A
8 801AFCD1 4B4FFF45 78CE6452 80DA0F8C 722A525E
9 81016058 7E957EFA 765A8458 80F610EA
10 59C6DD99 80D03E5E 0B6C614B 855BC0A7
11 701D6DBE 7E7A6BEC 78BC4A6A 7EBFC84B 8492E93D
12 787FC19A 685FF104 7643B67A 75D12EFF 7CDB7E84
13 7E24CA66 7ED2F440 00D8A9C0 60E13B28 8175709C
14 7F2053F6 7F0CAD9A 7E8E3A75 7DA2AF00 84816D12
15 7F4F2BC9 81626FD9 751466B9 80A445EC
16 803B40FE 8156DCC0 7AE9CFD4 8017E739 7A4783E5
17 8088387E 843075FB 7EBEFC6B 7B8F28ED 7A2409D9
18 816B88DI 7FE4954F
```

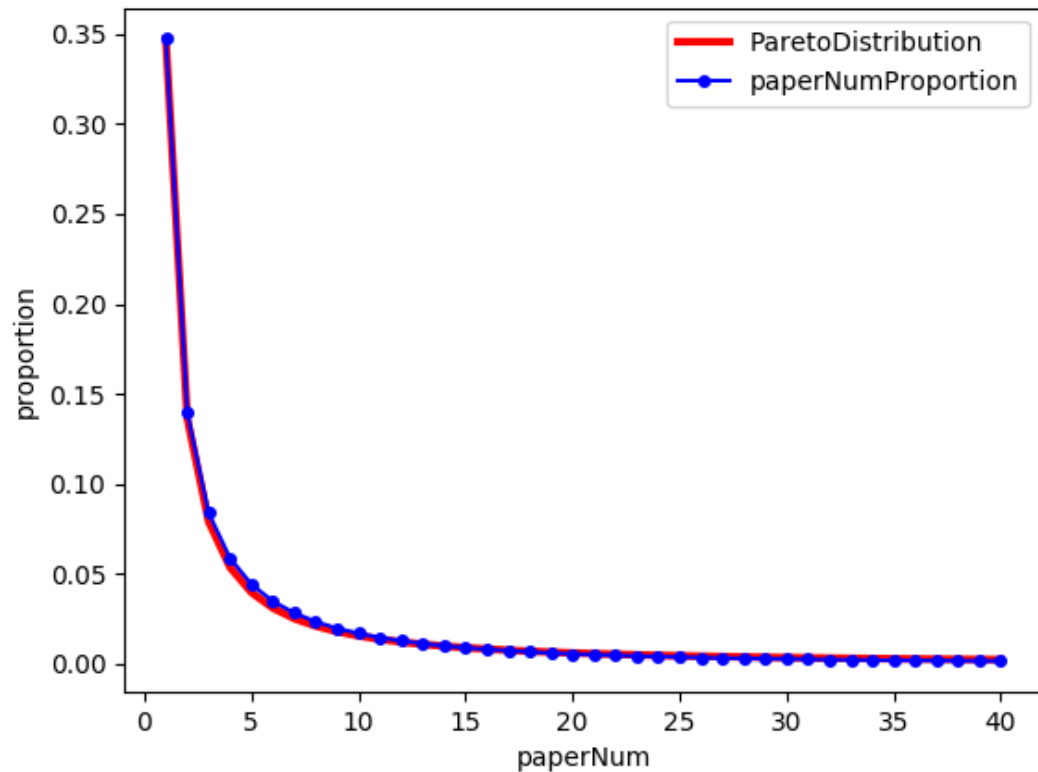

AuthorInfo in each Topic



Structure



Real Distribution & Probability density function



The standard pareto distribution is to describe the distribution of a random variable, the probability that X is greater than some number x is given by

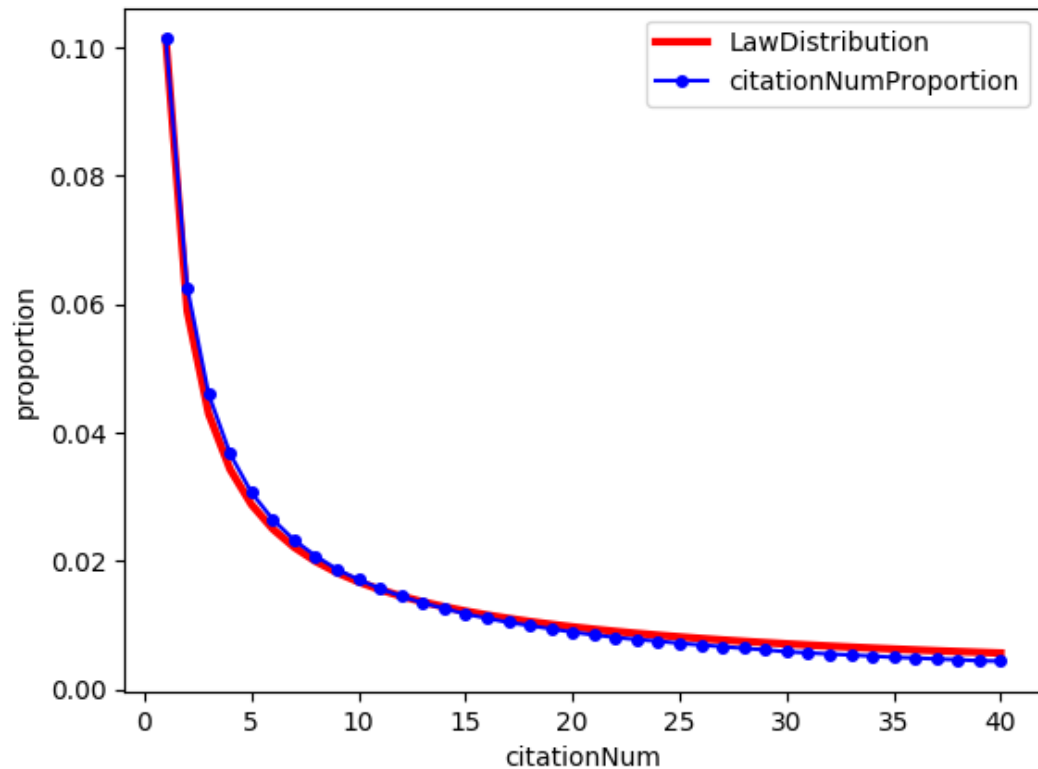
$$\bar{F}(x) = Pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m \\ 1 & x < x_m \end{cases}$$

where x_m is the minimum possible value of X , and α is a positive parameter called pareto index. So the probability density function of X followed is

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases}$$

As Fig. 2 presents, we get the conclusion that probability density function of paper-num obeys the pareto distribution with pareto index $\alpha_{\text{paper}} = 0.347$.

Real Distribution & Probability density function



And for the probability density function of citation-num, it obeys the law distribution with attenuation coefficient $\beta_{\text{citation}} = 0.782$

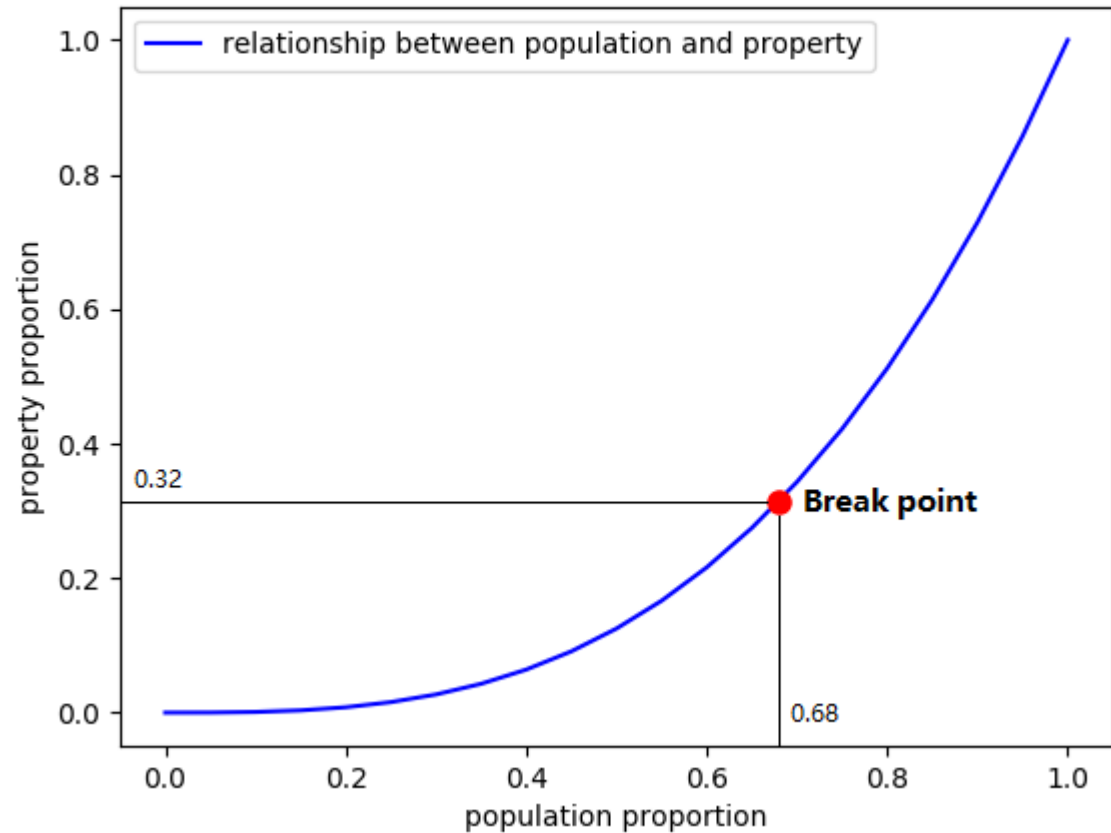
$$f_X(x) = \begin{cases} \left(\frac{x_m}{x}\right)^\beta & x \geq x_m \\ 0 & x < x_m \end{cases}$$

Break Point (pareto principle)

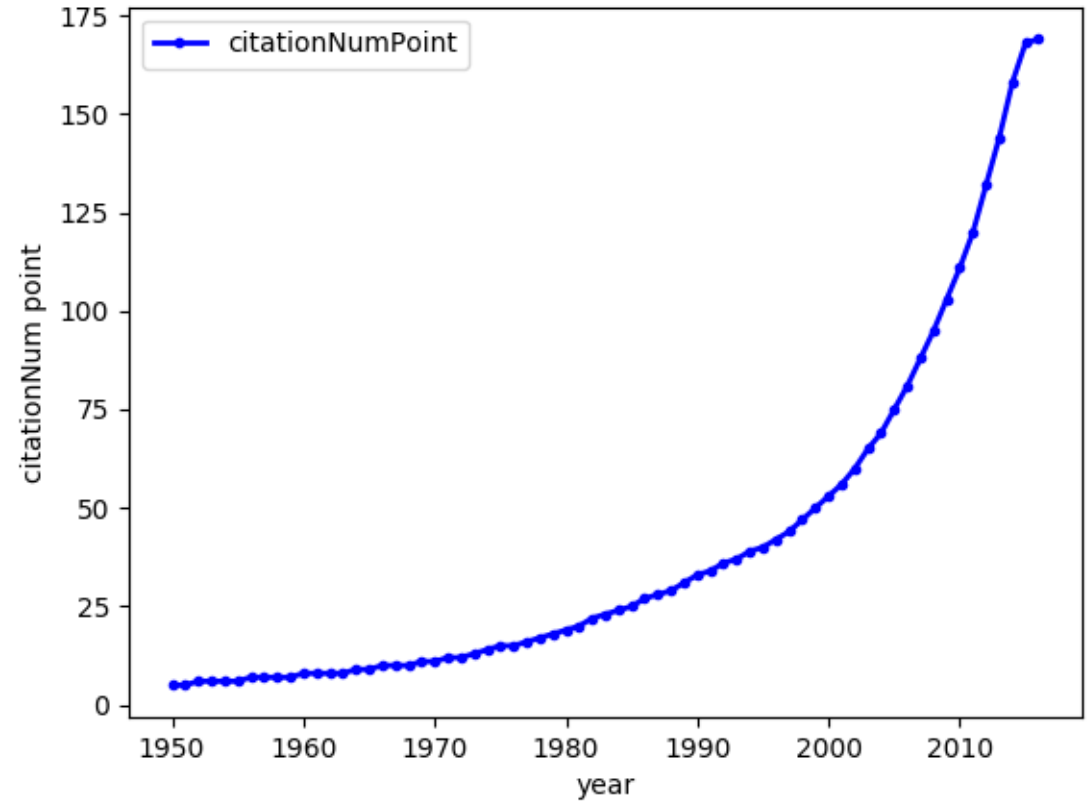
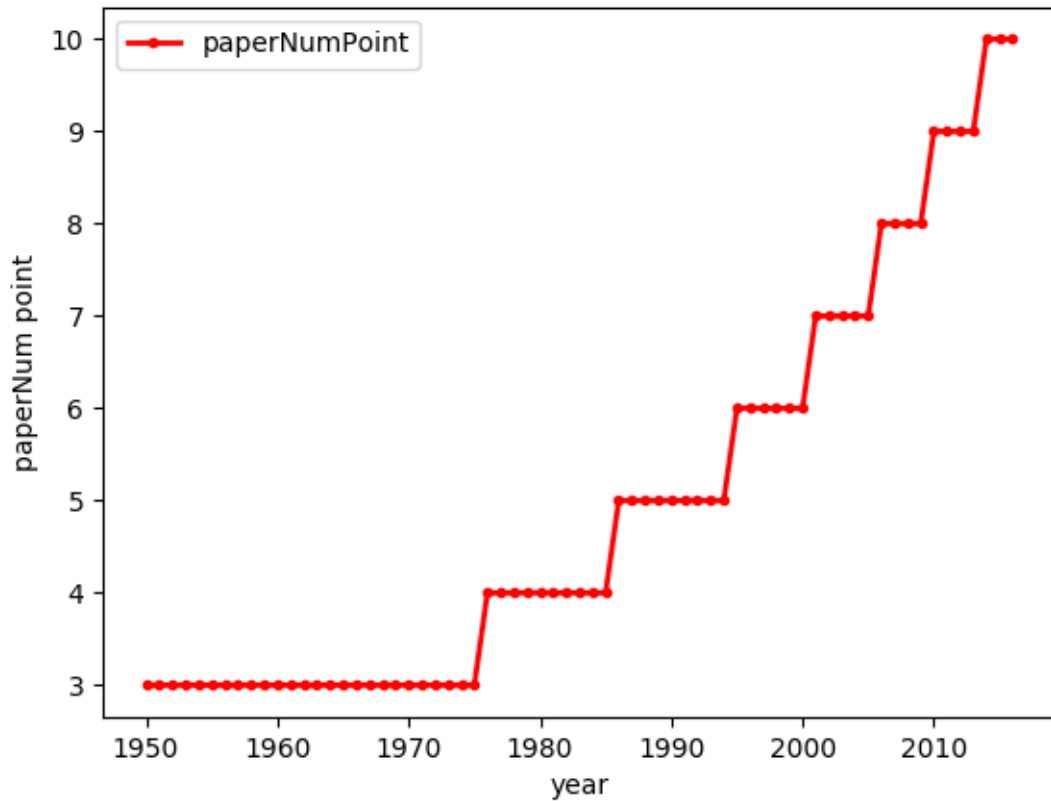
Based on pareto distribution and power distribution, Pareto principle states that for many events, roughly 80% of the efforts come from 20% of the cases.

For example, as Fig. 3 presents, let's define function $f(x)$ as the proportion of total property and x as the proportion of total population from the poor to the rich, and the break point x_{point} satisfies that

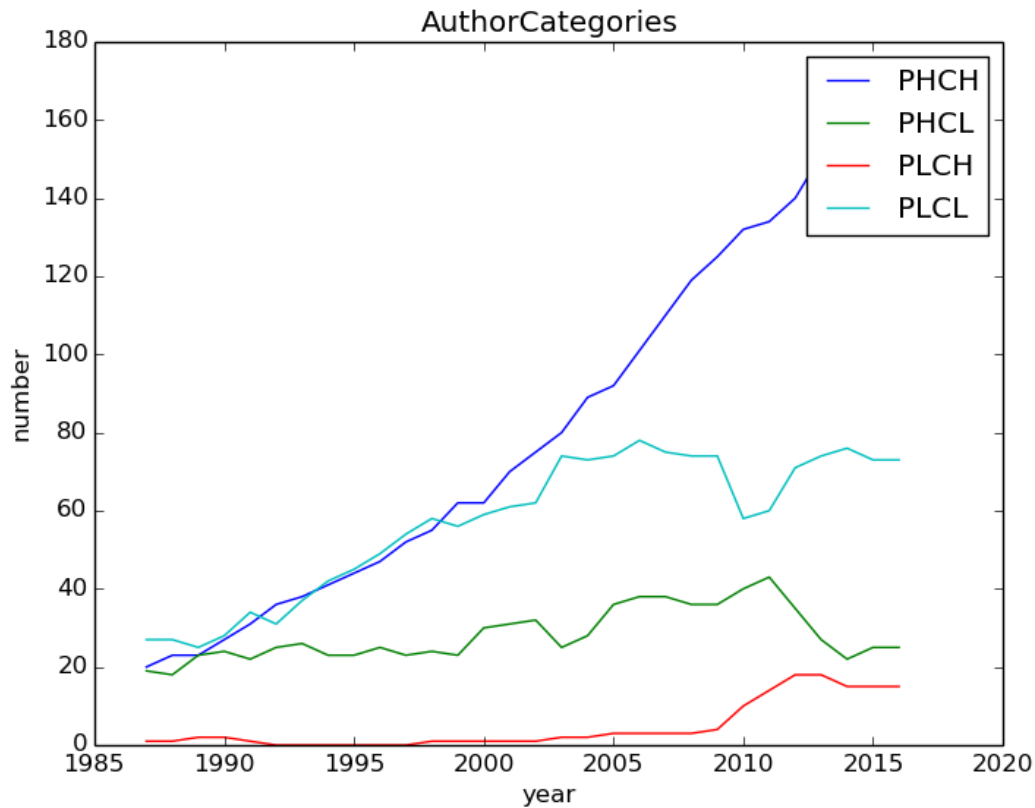
$$x_{point} + f(x_{point}) = 1$$



Break Point (pareto principle)



Development of authors of four types



Based on this result, we classify the authors into four types, PHCH, PHCL, PLCH and PLCL, and count the number for each type in each year, which takes us to the conclusion that the number of PHCH is increasing steadily, the number of PHCL and PLCH is increasing with complementary trend, which means when one is increasing, the other is declining relatively.

OUTLINE

- Motivation
- Implementation
- **Future Work**

Future Work

- Relationship between the number of author and the population of topic
- Define more characteristics of author
- Taking other methods for finding break point

Q & A



THANKS

