A Phenomenon of Optimal Scale of Science Cooperation

Yue Chen[1], Xianwen Wang[2], Deming Lin[3], Zeyuan Liu[4]

[1] *chenyuedlut@163.com*
Dalian University of Technology, Webometrics-Informetrics-Scientometrics-Econometrics Lab (WISE Lab), 116085 Dalian (China)

[2] *xianwenwang@dlut.edu.cn*
Dalian University of Technology, Webometrics-Informetrics-Scientometrics-Econometrics Lab (WISE Lab), 116085 Dalian (China)

[3] *vpopeye@163.com*
Dalian University of Technology, Webometrics-Informetrics-Scientometrics-Econometrics Lab (WISE Lab), 116085 Dalian (China)

[4] *liuzy@vip.163.com*
Dalian University of Technology, Webometrics-Informetrics-Scientometrics-Econometrics Lab (WISE Lab), 116085 Dalian (China)

Introduction

Scientific collaboration would be expected to increase both the quality and the quantity of the scientific output(Wallmark et al., 1973), While some empirical studies are not consistent with it like citations is usually viewed as the indicator to decide the quality of scientific output.

Based on the previous work, we proposed the following scientific questions:

- Is the Weibull distribution suitable to measure the scale of science cooperation?
- Can Kernel Density Estimation, a comparatively convenient function, be used to explain the relationship of the scientific productivity and scale of science cooperation? If yes, what are the optimal scale and reasonable scale?
-Is the optimal scale of science collaboration positively correlated to discipline biology connotation?

Data and methods

The database product SCI from the Institute for Scientific Information (ISI, Philadelphia, USA) was used as the data source. We chose the subcategory of *Nanoscience & Nanotechnology* as the sample dataset to answer the first two questions mentioned above. We use the data from five subcategories of *Physics/ Nuclear*, *Physics/Applied*, *Chemistry/Applied*, *Microbiology*, *Nanoscience & Nanotechnology* to answer the third question. All the data were retrieved and collected from the top 20 journals according to the 5-year Impact Factor in those five subcategories respectively.

This paper involved two density distribution functions, which are Weibull Probability Density Function and Kernel Density Estimation.

Results

We take the dataset of *Nanoscience & Nanotechnology* as the sample, and use Matlab to generate the following Figures (Figure 3). Figure (a) shows the original relationship curve between the number of co-authors and papers. Figure (b), (c) and (d) show the fitting of the density distribution of sample data with the Weibull Probability Density Distribution in 2000, 2005 and 2010 respectively.

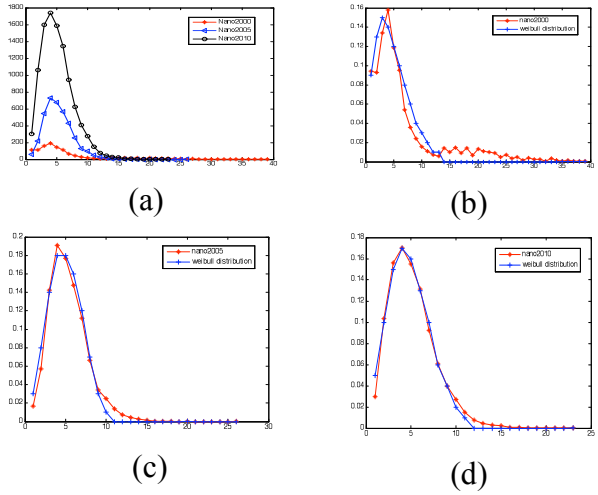

Figure 1. Weibull Probability Density distribution vs Sample data from 2000, 2005, and 2010

Because Kernel Density Estimation can relax assumptions on the underlying distribution and can model any distribution to higher levels of accuracy as we mentioned above, we applied this method in modeling the same dataset NANOSCIENCE & NANOTECHNOLOGY). The fitting result showed that it is reasonable to use KDE to analyse the scientific growth and collaboration scale.

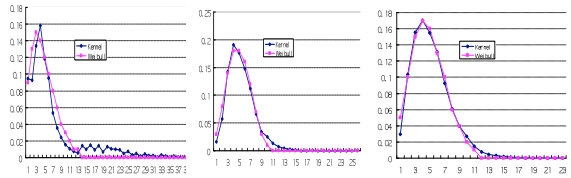

Figure 2. Kernel Density Estimation vs Weibull Probability Density distribution in 2000, 2005, and 2010

	Scale/Density (a)	Scale/Density (b)	Scale/Density (c)
K	3.8795/0.1423	3.8861/0.2376	4.0542/0.1695
W	3.1661/0.1454	4.7393/0.186	4.2354/0.1683

Scale of scientific cooperation, involves optimal scale and reasonable scale(Fig.2).We assume the optimal scale refers to the number of co-authors opposite to the peak, and the reasonable scale is decided by $P(\alpha \le x \le \beta) = \int_{\alpha}^{\beta} f(x)dx$(x means the cooperation group size), if $P(\alpha \le x \le \beta) > 80\%$, then $[\alpha, \beta]$ refers reasonable size(Fig.5)(briefly in Tab.2).

Table 2. optimal scale and reasonable scale in 5 subject categories

	optimal scale *(Density)*	*Reasonable scale*	*Range*	*Proportion*
Physics, Nuclear	3.5886 (0.1413)	3.5886-7.9373	4.3487	82.9%
Physics, Applied	4.0342 (0.1546)	0.7167-7.3518	6.6351	80.5%
Nanoscience & Nanotechnology	4.0542 (0.1695)	1.5317-7.5857	6.054	81.3%
Chemistry, Applied	3.393 (0.1141)	2.3255-7.6628	5.3373	84.9%
Microbiology	3.8307 (0.1197)	1.2736-9.5842	8.3106	82.5%

Conclusion

- The Weibull distribution is suitable to cooperation group size distribution.
- Kernel Density Estimation curve fits with the Weibull Density Distribution curve perfectly(R2>0.9) . As a comparatively convenient function, Kernel Density Estimation can be used as a model to study the relationship between scientific productivity and group size.
- The optimal scale and reasonable scale are calculated, the result didn't show correlation between collaboration group size and discipline biology connotation obviously.

Knowledgments

We appreciate Dr. Liang Liming, Dr.Zhang Dongling, Dr. Xie Caixia, this paper is inspired by their previous work mostly. Thanks for Dr.H.Kretschmer, she send her relevant research results to us for reference timely. Thanks for Dr. Guo Hanning, she made contribution to paper's English translation.
The research is sponsored by National Natural Science Foundation of China under Grant 70773015 and fundamental Research Funds for the Central Universities of China under Grant 851014.

References

Crane, D. (1969). Social structure in a group of scientists: A test of the" invisible college"

hypothesis. *American Sociological Review* 34, 335-352.
Goffman, W. & Warren, K. (1980). Scientific information systems and the principle of selectivity: Praeger New York.
Kretschmer, H. (1985). Cooperation structure, group size and productivity in research groups. *Scientometrics* 7, 39-53.
Kretschmer, H. & Kretschmer, T. (2007). Lotka's distribution and distribution of co-author pairs' frequencies. *Journal of Informetrics* 1, 308-337.
Lawani, S. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics* 9, 13-25.
Liang, L., Kretschmer, H., Guo, Y. & deB. Beaver, D. (2001). Age structures of scientific collaboration in Chinese computer science. *Scientometrics* 52, 471-486.
Lotka, A. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16, 317-323.
Needham, J. & Gwei-Djen, L. (1986). *Science and civilisation in China: Biology and biological technology. Botany*. Cambridge: Cambridge University Press.
Scott, D. (1992). Multivariate density estimation: theory, practice, and visualization. New York: Wiley-Interscience.
Seglen, P. & Aksnes, D. (2000). Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research. *Scientometrics* 49, 125-143.
Stankiewicz, R. (1979). The size and age of Swedish academic research groups and their scientific performance. In: Andrews, F. M. (ed.) *Scientific Productivity: The Effectiveness of Research Groups in Six Countries*. Cambridge: Cambridge University Press, 191.
Wallmark, J., Eckerstein, S., Langered, B. & Holmqvist, H. (1973). The increase in efficiency with size of research teams. *IEEE Transactions on Engineering Management* 20, 80-86.
Xie, C. (2006). Scientometric Study on Scientific Collaboration Pattern and Its Funcation. Dalian: Dalian University of Technology.
Zhang, D. (2009). Bibliometric Analysis of Output and Collabration of China's Scientific Literature. Dalian: Dalian University of Technology.