

An introduction to the theory of citing by M. V. Simkin and V. P. Roychowdhury

Г. Мороз

18 июля 2020, ReproducibiliTea, Москва

Read before you cite

$$R = \frac{D}{T} \times \frac{N - T}{N - D}, \text{ где}$$

R – доля людей, прочитавших статью из процитировавших работу; N – общее количество цитирований работы T – количество работ, процитировавших статью с опечаткой D – количество работ, процитировавших статью с уникальной опечаткой

При анализе статьи [Kosterlitz and Thouless \(1973\)](#) $R = 0.22$.

Не то чтобы ребята первые придумали

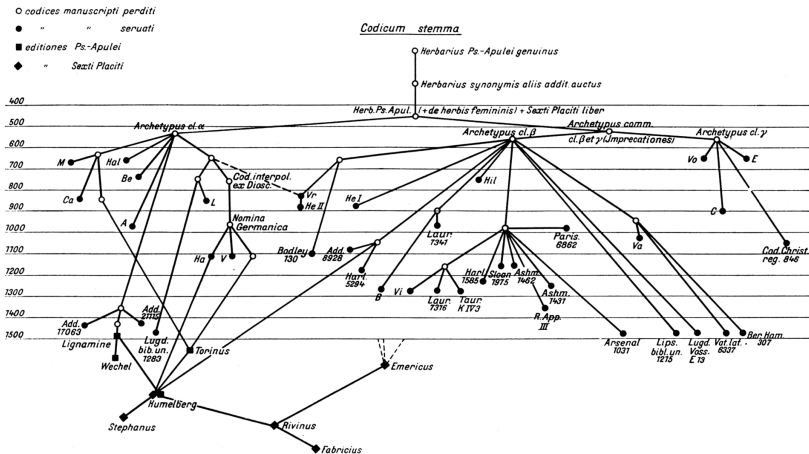


Схема распространения манускриптов Псевдо-Апулея *Herbarius*
 Musa et al. (1927)

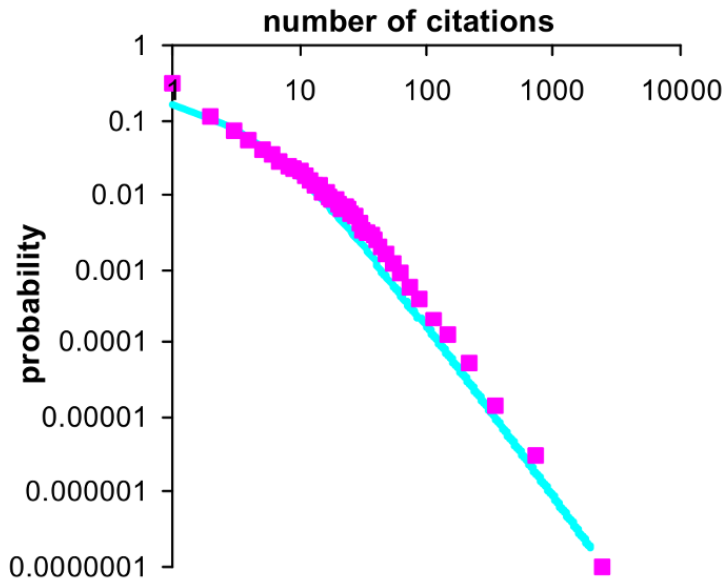
Copied citations create renowned papers

- Кажется, что чем больше цитирований – тем более выдающаяся работа
- Однако, если большинство цитирований скопировано из чужих списков литературы, значит, если работу уже цитируют, то ее с высокой вероятностью процитируют еще.
- Далее авторы обсуждают модель случайноцитирующего ученого (random-citing scientists, [Simkin and Roychowdhury \(2003\)](#)), который цитирует m случайных статей и копирует часть ссылок из этих статей, каждая с вероятностью p . Модель с параметрами $m = 3$ и $p = 1/4$, согласно работе, хорошо описывает данные:

$$N_K \sim \frac{1}{K^{1+(m \times p)^{-1}}} = \frac{1}{K^{2.(3)}}, \text{ где}$$

N_K – количество работ, которые процитировали K раз

Copied citations create renowned papers



References

- Kosterlitz, J. M. and Thouless, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6(7):1181–1203.
- Musa, A., Howald, E., and Sigerist, H. E. (1927). *Antonii Musae De herba vettonica liber: Pseudoapulei Herbarius: Anonymi De taxone liber: Sexti Placiti Liber medicinae ex animalibus etc.* Teubner.
- Simkin, M. V. and Roychowdhury, V. P. (2003). Copied citations create renowned papers? *arXiv preprint cond-mat/0305150*.