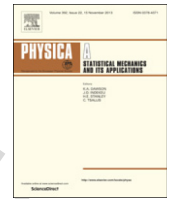




Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Q1 The double power law in human collaboration behavior: The case of Wikipedia

Q2 Okyu Kwon^a, Woo-Sik Son^a, Woo-Sung Jung^{b,*}

^a Division of Mathematical Models, National Institute for Mathematical Sciences, Daejeon 305-390, South Korea

^b Department of Industrial and Management Engineering and Department of Physics, Pohang University of Science and Technology, Pohang 790-784, South Korea

HIGHLIGHTS

- We study inter-event time distribution of editing behavior on Wikipedia.
- We observe a double power law distribution for the inter-editing behavior at the population level.
- We observe a single power law distribution at the individual level.
- We reveal that the synchronized editing behavior among users plays a key role.

ARTICLE INFO

Article history:

Received 30 December 2014
Received in revised form 13 November 2015
Available online xxxx

Keywords:

Wikipedia
Bursty behavior
Double power law

ABSTRACT

We study human behavior in terms of the inter-event time distribution of revision behavior on Wikipedia, an online collaborative encyclopedia. We observe a double power law distribution for the inter-editing behavior at the population level and a single power law distribution at the individual level. Although interactions between users are indirect or moderate on Wikipedia, we determine that the synchronized editing behavior among users plays a key role in determining the slope of the tail of the double power law distribution.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The current availability of enormous amounts of digital information on human activities facilitates the analysis of temporal activity patterns of various human dynamics. Bursty behavior, which leads to a heavy-tailed inter-event or response time distribution, has been observed in a wide range of human dynamics; these include communication patterns in mail [1–6], short message correspondence [7], mobile phone records [8–10], web surfing [5,11–13], library loans [5], printing requests [14], networked games [15], online chatting [16], and file downloads [17].

To the best of our knowledge, there are three approaches to studying bursty behavior. The first approach investigates empirical patterns in terms of inter-event (or response) time distributions [1,3–5,7]. Although there has been a longstanding controversy regarding the power law behavior of the inter-event time distribution, various universality classes in the power law distribution have been reported from various temporal human behaviors [5]. Moreover, empirical evidence has revealed that the bimodal distribution is neither completely Poisson nor exhibits power law behavior [7]. The second approach determines the origins of bursty behavior. A queueing model for a to-do list with a highest-priority-first protocol [3,5] is the

* Corresponding author.

E-mail address: wsjung@postech.ac.kr (W.-S. Jung).

Table 1

The most active and least active articles of the 418 featured articles.

Article	# of edits	# of users
Michael Jackson	27,188	6437
Jesus	26,920	6621
The Beatles	21,718	6631
Aelle of sussex	402	182
Japanese aircraft carrier Soryu	341	156
Japanese aircraft carrier Hiryu	313	169

best-known model that explains heavy tails in inter-event time distributions. Periodic cycles of human activity also provide a good alternative explanation to the heavy tails [6,18]. The last approach suggests a new definition or rescaling method for the time interval in terms of hidden scaling patterns and universality [19–21]. Underlying similarities between surface letters and e-mail are successfully revealed by rescaling time intervals [19]. Furthermore, a new timing method to eliminate activity heterogeneity in temporal human behavior was recently introduced [20,21].

Wikipedia is an online collaborative encyclopedia that has grown exponentially in terms of its number of users and information since 2001 [22]. It is considered to be a success story of low-cost collaborative knowledge systems. Anyone with access to this online encyclopedia can edit almost all of its articles. Wikipedia also provides a revision history for its articles that includes when an article was revised and who revised it. Therefore, the revision history is one of the best tools for determining temporal activity patterns of collaborative human behavior on the Internet.

In this work, we investigate the distribution of inter-editing time intervals between two successive edits of Wikipedia articles at both the individual and population levels. Interestingly, at the individual level, a single power law distribution with an exponent of $\alpha \approx 1$ is observed. However, at the population level, a double power law distribution is observed. The latter distribution exhibits two exponents, $\alpha \approx 1$ for shorter time intervals and $\alpha \approx 2$ for longer time intervals. Previous studies on the distribution of inter-event times between two consecutive events for e-mail, mobile phones, and short message correspondence [6–8] provide empirical results for the temporal activity pattern of a single person when there is explicit correspondence with other individuals. In contrast to direct communication, editors of Wikipedia articles interact indirectly with other individuals; this interaction is implicitly mediated by the articles that are continuously revised by Wikipedia users. Our study reveals new evidence and insight on the temporal patterns of collective human behavior when there are indirect or moderate interactions between users.

2. Methods

Wikipedia's editors determine the featured articles that are considered the best articles on Wikipedia; as of March 26, 2014, there were 4200 featured articles out of the total 4,479,603 articles on the English Wikipedia web page. In this study, we only considered featured articles. Each article displays heterogeneous statistical properties in terms of the number of users and editing frequency according to its popularity [23]; however, the time of the first editing event differs greatly from article to article; thus, we only focused on the 418 featured articles that had their first editing event in 2001 to obtain some homogeneity. Although these articles have existed for approximately 13 years, their degree of activity differs in terms of the number of users involved in their development.

To clearly compare patterns in the probability distribution of inter-editing time intervals according to their degree of editing activity, we selected the three most active articles and the three least active articles, in terms of number of edits, from the 418 featured articles. The topics of the most active articles were “Michael Jackson”, “Jesus”, and “The Beatles”; the topics of the three least active articles were “Aelle of Sussex”, “the Japanese aircraft carrier, Soryu”, and “the Japanese aircraft carrier, Hiryu”. Table 1 shows the total number of edits and the number of unique users for each article.

To study the effects of collective editing behavior, we measured the probability distribution of inter-editing time intervals at both the individual and population levels of same article. For example, the temporal edit pattern of “the Japanese aircraft carrier, Hiryu” is shown in Fig. 1. The edit history of each individual user was separated from the entire edit history for a single article. At the population level, we measured inter-editing time intervals using the complete editing history of all users. At the individual level, we measured inter-editing time intervals using each individual user's edit history.

3. Results and discussion

Patterns of inter-editing time intervals at the individual level. To observe patterns in the inter-editing time interval, τ , at the individual user level, we extracted a single user's editing history for a particular article and measured the inter-editing time interval of that user. We collected the inter-editing time intervals of all single users who participated in collective editing behavior for the same article. Using these time intervals, we obtained a distribution of inter-editing time intervals at the individual level for each single article. As shown in Fig. 2, the probability distributions follow a power law distribution with exponent $\alpha \approx 1$. Notice that they are identical for all six articles, even though their number of edits and users differ.

Patterns of inter-editing time intervals at the population level. There are various perspectives of understanding collective behavior between users in editing Wikipedia articles, especially editorial wars and disputes are well known and studied

Download English Version:

<https://daneshyari.com/en/article/7377188>

Download Persian Version:

<https://daneshyari.com/article/7377188>

[Daneshyari.com](https://daneshyari.com)