

# A Capitalized Title: Something about a great Discovery

Liyang Wang<sup>1\*</sup>

**1 Department of Anthropology, University of Washington, Seattle, Washington, USA**

**\* E-mail:** liyang15@uw.edu

## 0.0.1 Introduction

My potential questions for this quantitative research project as following:

1. What is the trend of research of social inequality in archaeology in the past decades? Is it popular in any particular time period?
2. How often social inequality relates to culture contact? Especially for European contact. Also, does any area show preference of such research, such as North America, Southeast Asia, Oceania, or Africa.
3. For the method, what is the relationship between the term “inequality” and other archaeological evidence, including burials, ceramics, post holes, or glass beads.

In order to run the codes for my questions, I will use the following two libraries:

```
library(JSTORr)
# multiple_archives <- JSTOR_unpack_multiple_archives(mydir = "data/one_gram")
# save(multiple_archives, file= "data/one_gram/multiple_archives.RData")
load("data/one_gram/multiple_archives.RData")
# library(devtools)
# download and install the package (do this only once ever per computer)
# install_github("benmarwick/JSTORr")

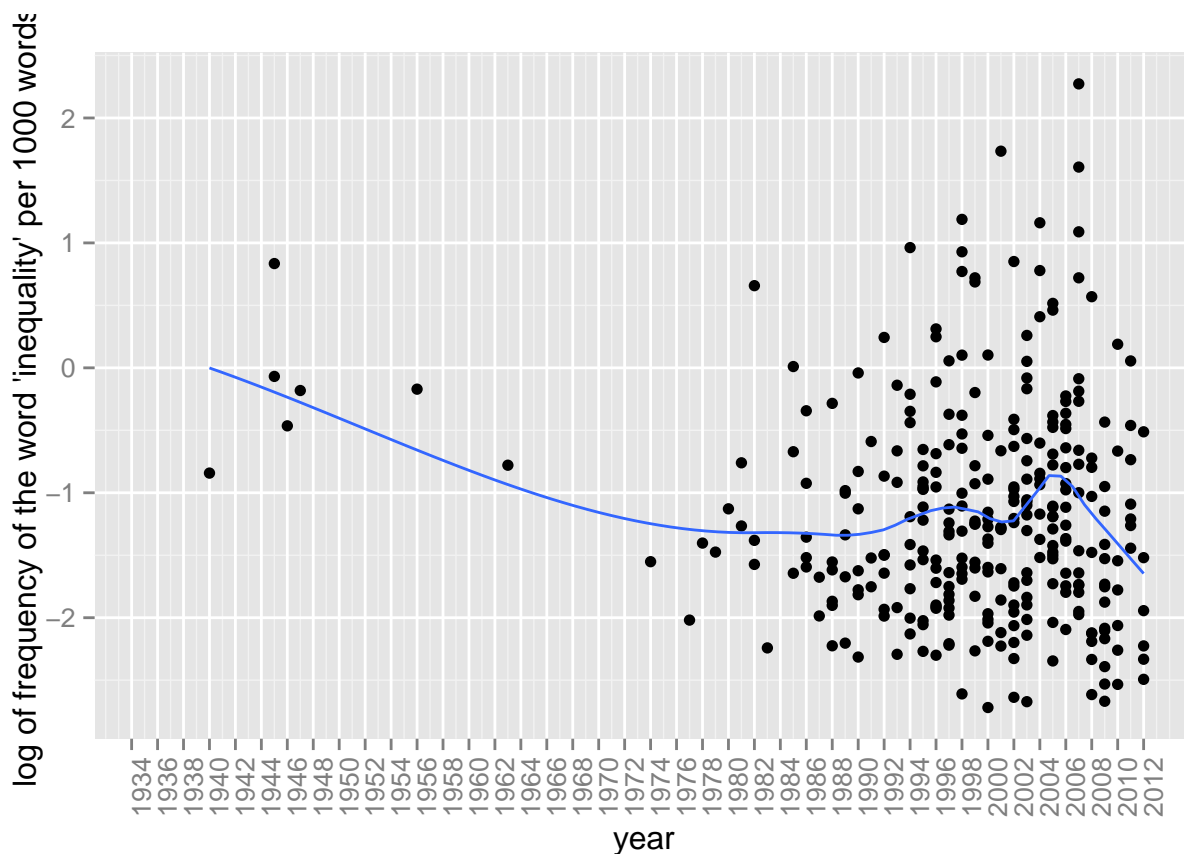
# setwd("data/two_gram")
# bigrams <- JSTOR_unpack2grams()
# save(bigrams, file = "bigrams.RData")
load("data/two_gram/bigrams.RData")

nouns_Asia <- JSTOR_dtmofnouns(multiple_archives, word = "asia", sparse =1, POSTag = TRUE)
nouns_Ame <- JSTOR_dtmofnouns(multiple_archives, word = "america", sparse =1, POSTag = TRUE)
nouns_Oce <- JSTOR_dtmofnouns(multiple_archives, word = "oceania", sparse =1, POSTag = TRUE)
nouns_Afr <- JSTOR_dtmofnouns(multiple_archives, word = "africa", sparse =1, POSTag = TRUE)
```

For the first question of the trend of social inequality, the plots show that the topic of inequality becomes common after maybe 1992, and then has become popular around 2006. The term “social” is commoner than the term “inequality”. I think the reason is that this term “social” is too general. The third graph shows that social and inequality were decussated together after 1978, but slightly goes down around 1986, and then popular again from 1994. Based on the results, I can further explore the exact articles with the word inequality.

The article with strongest correlation of the relationship of “social” and “inequality” as following: Sandra Montn Subas, 2007, Interpreting Archaeological Continuities: An Approach to Transversal Equality in the Argaric Bronze Age of South-East Iberia. World Archaeology, Vol. 39, No. 2, The Archaeology of Equality, pp. 246-262. Taylor & Francis, Ltd

```
\begin{CodeChunk} \begin{CodeInput} # one word over time inequality <- JSTOR_1word(multiple_archives,
“inequality”) inequality$plot \end{CodeInput}
```



```
inequality_tbl <- inequality$word_by_year
inequality_tbl[which.max(inequality_tbl$word_by_year), ]
```

Empty data.table (0 rows) of 3 cols: word, ratio, year, V2

```
# two words over time
two_soc_ine <- JSTOR_2words(multiple_archives, "social", "inequality")
two_soc_ine_tbl <- two_soc_ine$words_by_year
two_soc_ine_tbl[which.max(two_soc_ine_tbl$value), ]
```

year variable value V2 1: 1998 social 38.05955 10.2307125040

```
# correlation of words over time
cor_soc_ine <- JSTOR_2wordcor(multiple_archives, "social", "inequality")
soc_ine_paper <- two_soc_ine_tbl[(two_soc_ine_tbl$year == 2007), ]
soc_ine_paper[which.max(soc_ine_paper$value), ]
```

year variable value V2 1: 2007 social 29.81969 10.230740026656

\end{CodeChunk}

If I use “social inequality” as a term, the plots shows that such discussion only exists after 1978, becomes popular around 2000, and slightly goes down at 2002. In archaeology, social inequality is one of the indexes for social complexity. Therefore, I will also examine the discussion of social complexity in order to have a general understanding of such topic. The discussion of “social complexity” begins from 1947, becomes popular at around 1978, slightly goes down at 1990 and 2006, and then popular again at

2010. It seems like there is not big change in the discussion of these two terms. If we examine these two terms over time, the plot shows that they have different pattern of frequency. When social inequality was frequently discusses, the frequency of social complexity becomes lower. This might indicates the terms of social inequality and social complexity might have some pattern of relationship, which is to some extent mutually exclusive. Social inequality could be viewed as more specific approach to social complexity.

```
# one word over time
social_inequality <- JSTOR_1bigram(bigrams, "social inequality")
social_inequality_tbl <- social_inequality$data
social_inequality_tbl[which.max(social_inequality_tbl$bigram_ratio), ]

      bigram_ratio year
1:  0.001287416 2001

bi_socialcomplexity <- JSTOR_1bigram(bigrams, "social complexity")
# correlation of words over time
bi_ineq_comp <- JSTOR_2bigrams(bigrams, "social inequality", "social complexity")
```

For the second question about the correlation between social inequality and European contact, the plot of two terms shows that there is a similar trend before 1994, but after 1994 there is no clear pattern shows the relationship. When we further examine the correlation between these two terms, the plot indicates that there is only an article shows the relatively strong correlation at 1986. In addition to European contact, I also examine the correlation between social inequality and social complexity. The plot shows that they are strongly correlated at 1987, 1998, and 2007, which shows the periodic pattern of discussion. This plot also explains the pattern in the former plot of word frequency over time.

```
# two words over timw
bi_ineq_euro <- JSTOR_2bigrams(bigrams, "social inequality", "european contact")
# correlation of words over time
bicor_ineq_euro <- JSTOR_2bigramscor(bigrams, "social inequality", "european contact")
bicor_comp_ineq <- JSTOR_2bigramscor(bigrams, "social complexity", "social inequality")
```

I further examine the topic of inequality in terms of different areas, such as Asia, Oceania, America, and Africa. The results show that in Asia, most articles contain inequality exist after 1978, and the term reached the highest frequency at 2004, but declined later at 2009. For Oceania, the trend shows there are two periods with higher frequency, which are 1996 and 2004. For America, there are more articles refer to inequality, and this term was popular around 2005. The pattern of frequency in Africa looks similar to the pattern in America, and also shows the highest frequency at 2005. Although these four plots have different trends of the frequency of inequality, it seems this term was popular between 2004 and 2005 among these different areas.

```
# subset the data to different area
Asi_data <- JSTOR_subset1grams(multiple_archives, "asia")
Oce_data <- JSTOR_subset1grams(multiple_archives, "oceania")
Ame_data <- JSTOR_subset1grams(multiple_archives, "america")
Afr_data <- JSTOR_subset1grams(multiple_archives, "africa")
# one word, inequality, in different area
ineq_Asi <- JSTOR_1word(Asi_data, "inequality")
ineq_Oce <- JSTOR_1word(Oce_data, "inequality")
ineq_Ame <- JSTOR_1word(Ame_data, "inequality")
ineq_Afr <- JSTOR_1word(Afr_data, "inequality")
```

For the third questions, these plots show the relationship between the term ‘inequality’ and other archaeological evidence, including beads, burials, and pottery. For the beads, it indicates that a couple of articles have strong positive correlation around 1990. In addition, there is no much discussion until 1976. For the burials, it also shows that the discussion of burial and inequality become common after the publication of the article in 1976. Besides, it seems the articles could be divided into two groups according to the extent of correlation. For the pottery, there is no clear correlation until 2006. After 2006, some articles show the positive correlation.

```
# correlation of inequality and evidence over time
cor_ineq_buri <- JSTOR_2wordcor(multiple_archives, "burial", "inequality")
cor_ineq_pot <- JSTOR_2wordcor(multiple_archives, "pottery", "inequality")
cor_ineq_bead <- JSTOR_2wordcor(multiple_archives, "bead", "inequality")
```

TBA

```
# subset samples
nouns<- JSTOR_dtmofnouns(multiple_archives, word = NULL, sparse =1, POSTag = TRUE)
# Principal component analysis for articles containing inequality
PCA_inequality <- JSTOR_clusterbywords(nouns, "inequality", f = 0.005)
PCA_inequality <- JSTOR_clusterbywords(nouns_Asia, "inequality", f = 0.005)
PCA_inequality <- JSTOR_clusterbywords(nouns_Ame, "inequality", f = 0.005)
PCA_inequality <- JSTOR_clusterbywords(nouns_Oce, "inequality", f = 0.005)
# Two words
PCA_social_ineq <- JSTOR_clusterbywords(nouns, c("inequality", "social"), f = 0.005)
```

The frequency of words containing “inequality” shows that the focus changes with time. The focus of paper shift from population, organization, and mound to ritual, power and pottery, which indicates there is a trend of anthropological thinking from 2006. Besides, The term settlement could be observed in each period from 1996 until now, and it becomes more popular over time.

```
# Top words containing 'inequality' over time
# exclude some irrelevant words
custom_stopwords <- c('archaeology', 'university', 'research', 'evidence', 'journal', 'world', 'site')
ineq_nouns <- JSTOR_dtmofnouns(multiple_archives, word = 'inequality', sparse =1, POSTag = TRUE)
ineq_freqwords <- JSTOR_freqwords(multiple_archives, ineq_nouns, custom_stopwords, n = 5)
con_nouns <- JSTOR_dtmofnouns(multiple_archives, word = 'contact', sparse =1, POSTag = TRUE)
con_freqwords <- JSTOR_freqwords(multiple_archives, con_nouns, custom_stopwords, n = 10)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.