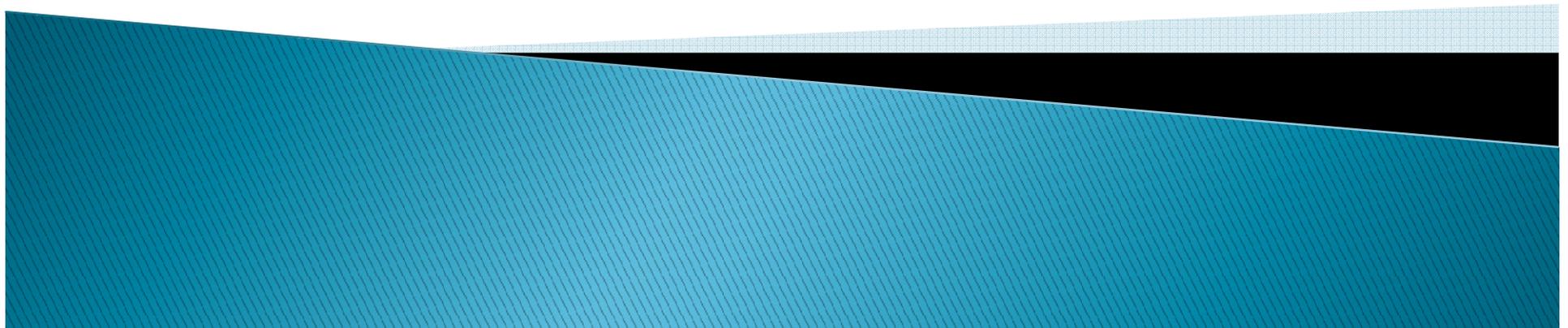


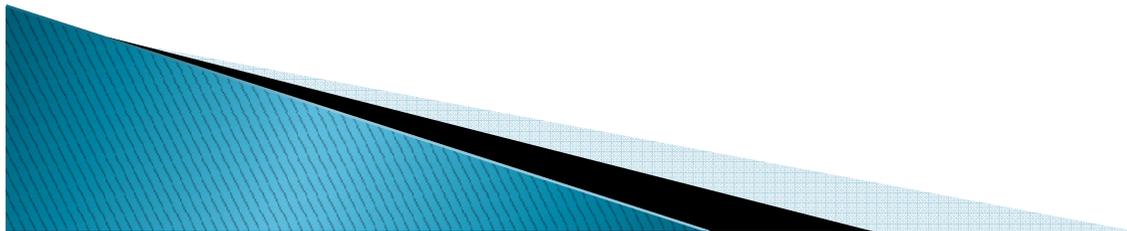
# **The Text Mining Analysis on Public Opinion Focus toward Macau's Gambling Industry**

Shianghau Wu  
Macau University of Science and Technology



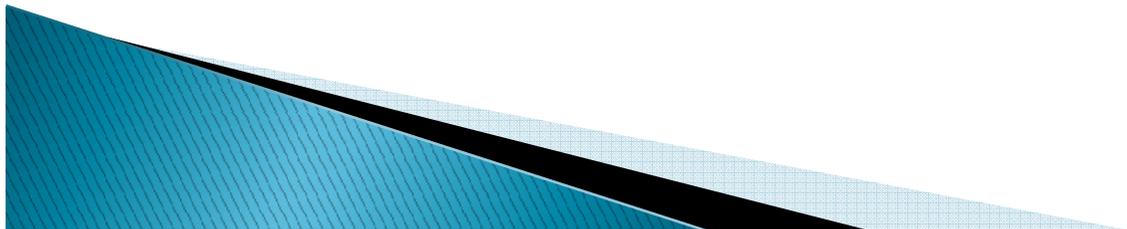
# Introduction

- ▶ The study applied the text mining analysis to extract the keywords in order to comprehend the public opinion focus on Macau's gambling industry.
- ▶ The study analyzed keywords of Chinese news reports in two time frames, September to December 2011 and June to September 2012 and made the comparison to explore the difference of the keywords in two different time frames.



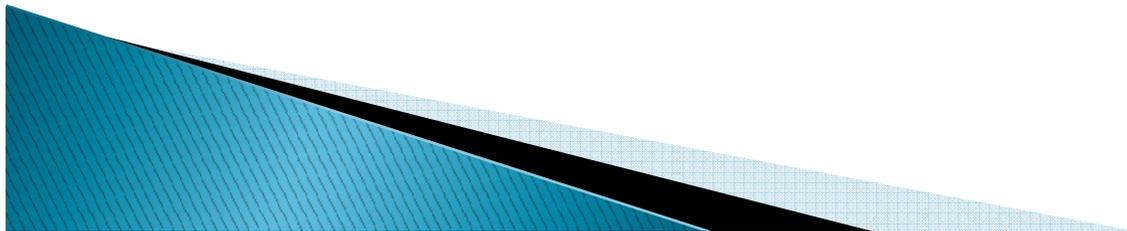
# Research Material

- ▶ All documents used in this study were accessed from the Wisers Information Portal(<http://wisearch.wisers.net>). In this study, we analyzed the Chinese news reports related to Macau's gambling industry ( 澳門博彩業 ).
- ▶ The study purposed to use the text mining method to analyze the keywords in order to get the public opinion focus in different time frames.



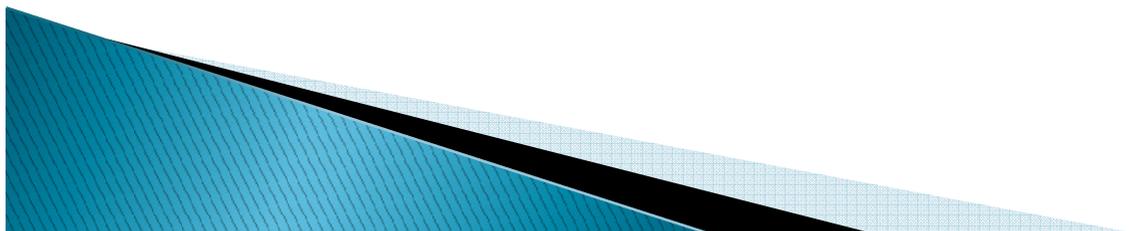
# Text Mining

- ▶ Text mining is one of the data mining methods, which learn from samples of past experience. In the text mining method, the text will be processed and transformed into a numerical representation.
- ▶ The text mining method is widely applied to information management on websites, biological data and customer relationship management.



# Research Design

- ▶ The text mining method of the study was firstly implemented by means of the text processing function (“tm” package) of the R language.

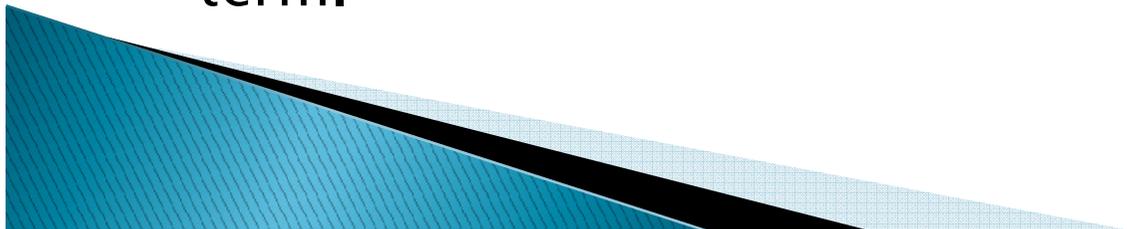


# Text Mining Process

- ▶ Parsing: The text and the structure were extracted and represented in a data structure.
  - ▶ Number Filter: The procedure filtered all terms consisting of numbers only.
  - ▶ The Punctuation Erasure: The procedure removed all punctuation marks.
  - ▶ Stop Words Filter: The procedure filtered all stop words.
  - ▶ Chinese Words Segments Analyzer: The study used the “mmseg4j” Chinese words segments analyzer (<http://code.google.com/p/mmseg4j/>) to extract the keywords and set the minimum characters as two in each keyword.
- 

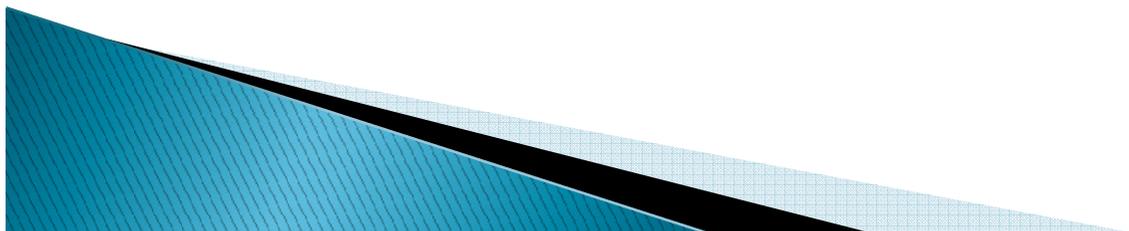
# Text Mining Process

- ▶ Frequencies and Keywords extraction: After the above-mentioned processes, all irrelevant terms can be filtered out. To compute these frequencies two nodes are available, one to compute the well known term frequency *tf*, relative or absolute, and one to compute the inverse document frequency *idf*. The step got the results of containing the computed frequencies.
- ▶ Visualization: The study applied the “wordcloud” package in R language. It provides a typical tagcloud with some additional options, such as different arrangements of terms, *i.e.* size sorted, alphabetic, inside out, adjustable minimal and maximal font size, or transparency of terms etc. The node requires an input data table consisting of a bag of word with a additional column containing a weight or score of each term.



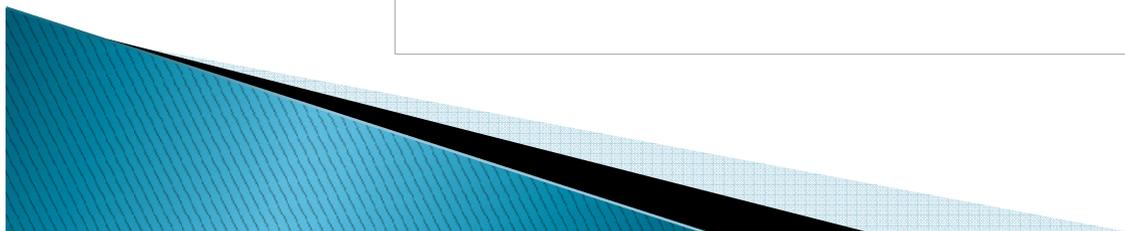
# Data Collection

- ▶ The study assembles the Chinese news reports in two time frames, 238 news reports from September to December 2011 and 260 news reports from June to September 2012 and made the comparison to explore the difference of the keywords in two different time frames.
- ▶ The goal of this study is to compare the keywords in two time frames in order to compare the focus change.



# Tag Clouds

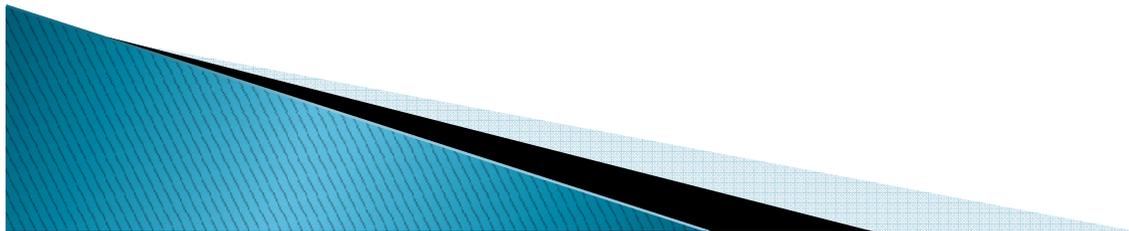
(September to December 2011)



# Tag Clouds

## (September to December 2011 )

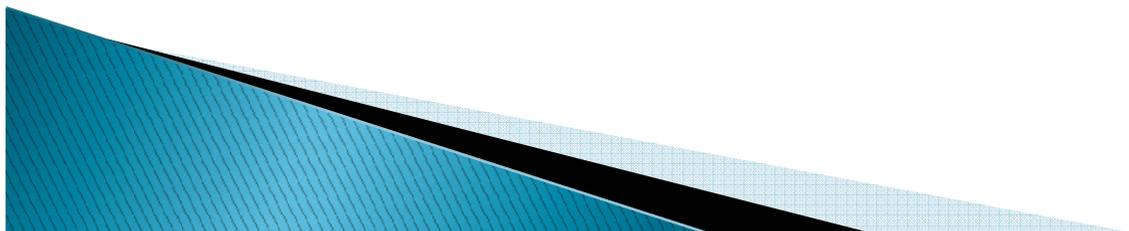
- ▶ We would know the major keywords from September to December 2011 news reports were special administrative region government (特区政府), golden holidays week (黄金周), industrial diversification (多元化), investors (投资者) and Singapore(新加坡).





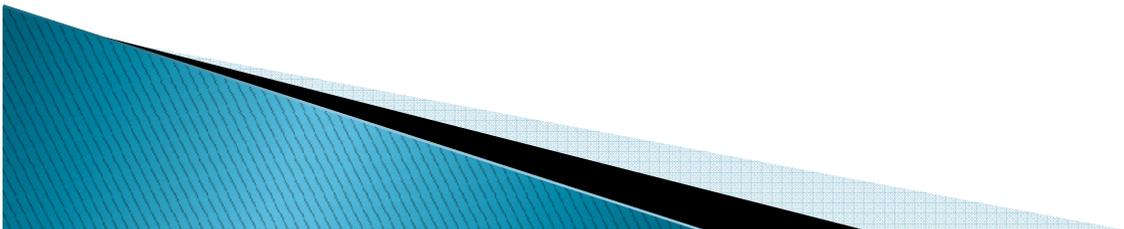
# Tag Cloud of Keywords from June to September 2012

- ▶ We would know the important keywords from Chinese news reports from June to September 2012 were ebitda (earnings before interest, taxes, depreciation, and amortization), special administrative region government (特区政府), investors (投资者), resorts (度假村) and junkets (中介人).

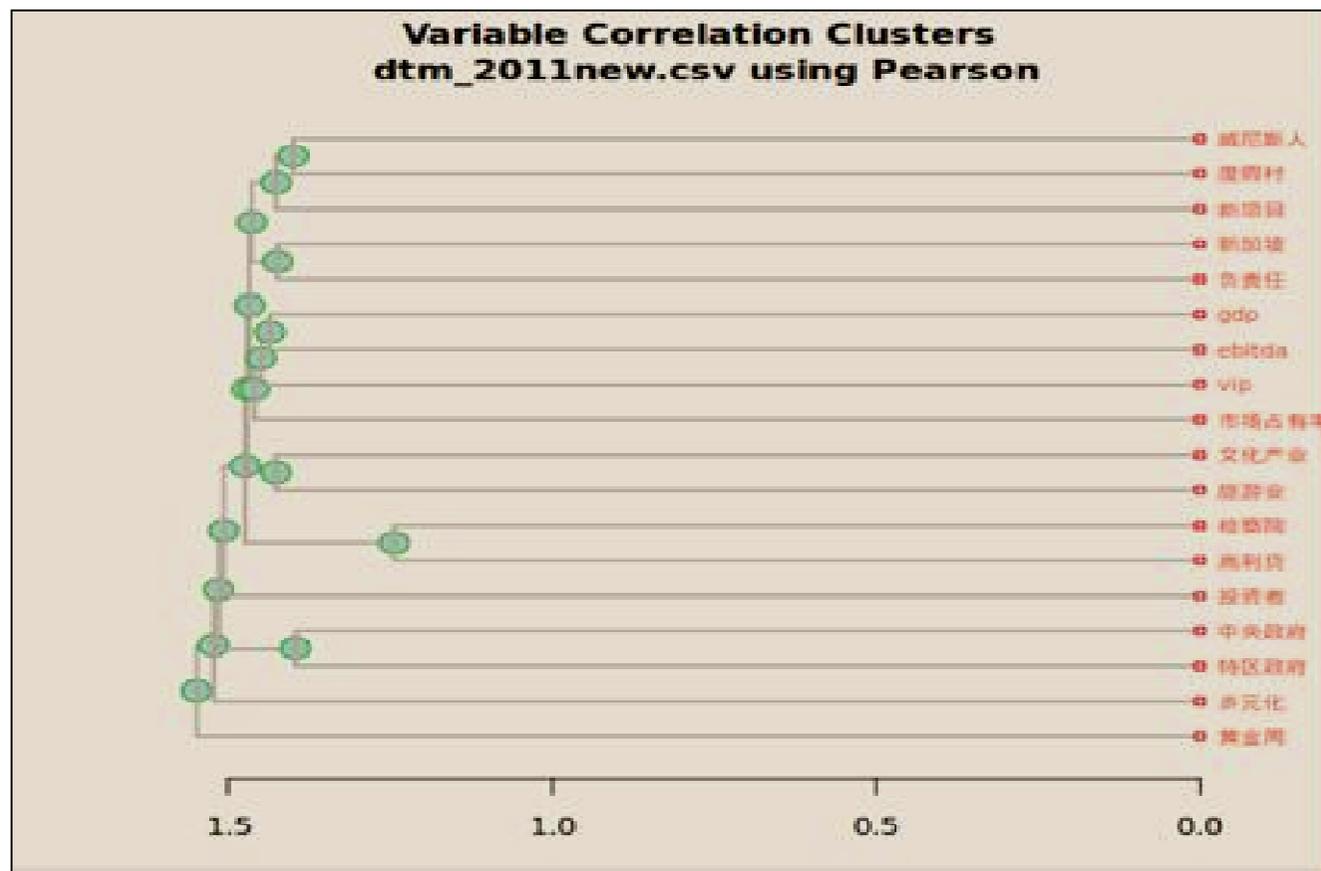


# Hierarchical Clustering

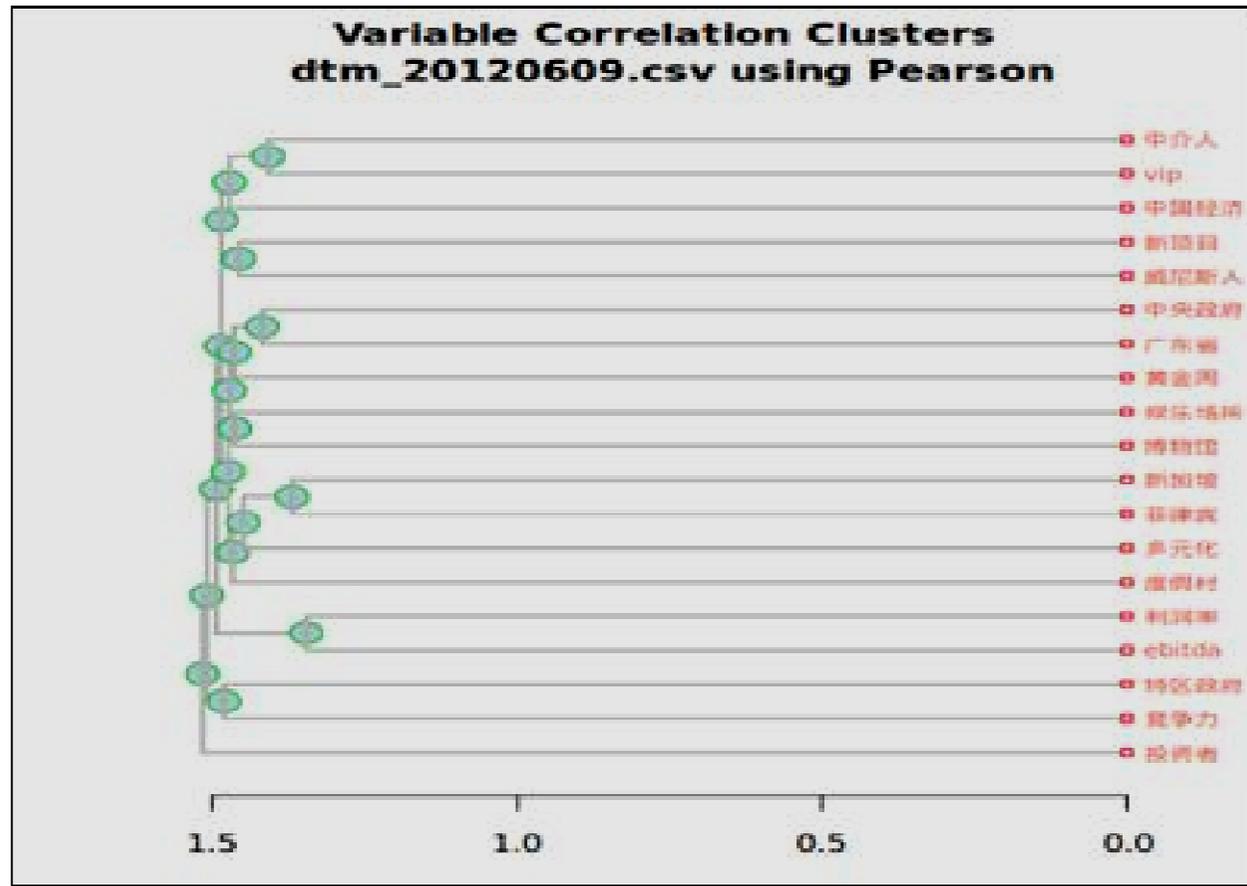
- ▶ The study calculated the term frequency  $tf$ , relative or absolute, and the inverse document frequency  $idf$  to choose the first twenty keywords and uses the “rattle” package of the R software to make the hierarchical clustering analysis according to Pearson correlation coefficient between each keyword or paragraph.
- ▶ The study used the Pearson correlation coefficient between each keyword to make the hierarchical clustering.



# Hierarchical Clustering Graphs of keywords from September to December 2011 News Reports

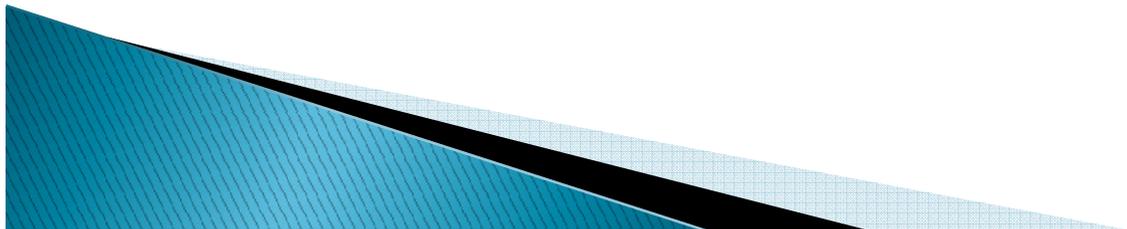


# Hierarchical Clustering Graphs of keywords from News Reports of June to September 2012



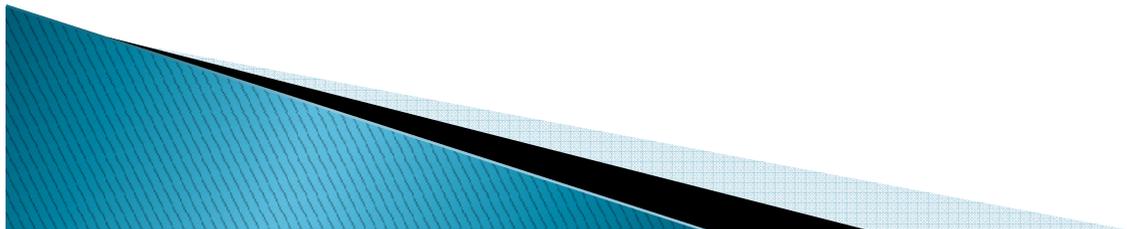
# Hierarchical Clustering Graphs

- ▶ It indicated that the keyword "Venetian" (威尼斯人) related to "resorts"(度假村) and "new projects" (新项目) in keywords of September to December 2011, and also related to "new projects" in keywords of June to September 2012.
- ▶ On the other hand, the keyword "vip" related to "ebitda", "GDP" and "market share" (市场占有率) in keywords of September to December 2011, but related to "junkets" (中介人) and "Chinese economy" (中国经济) in keywords of June to September 2012.
- ▶ The result showed the news reports related to Macau's gambling industry had the different emphasis in different time frames.

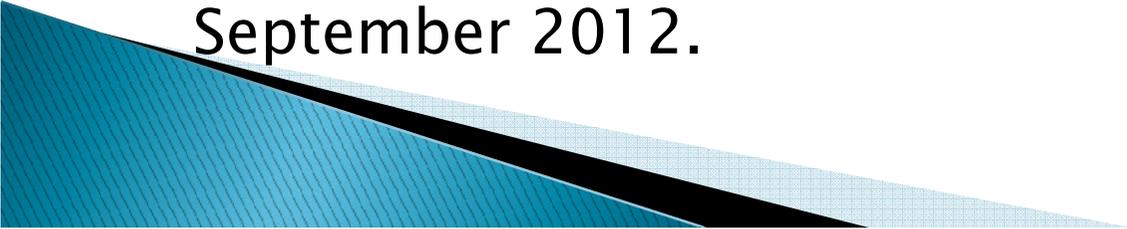


# Discussion

- ▶ The study found that major keywords the major keywords from September to December 2011 news reports were special administrative region government (特区政府), golden holidays week (黄金周), industrial diversification (多元化), investors (投资者) and Singapore(新加坡).
- ▶ The study also found that the important keywords from Chinese news reports from June to September 2012 were ebitda (earnings before interest, taxes, depreciation, and amortization), special administrative region government (特区政府), investors (投资者), resorts (度假村) and junkets ( 中介人 )

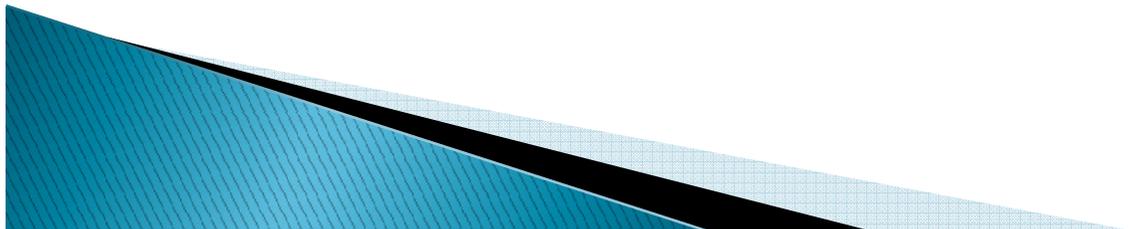


# Discussion

- ▶ The study also used the Pearson hierarchical clustering analysis and concluded that it indicated that the keyword "Venetian" (威尼斯人) related to "resorts"(度假村) and "new projects" (新项目) in keywords of September to December 2011, and also related to "new projects" in keywords of June to September 2012.
  - ▶ On the other hand, the keyword "vip" related to "ebitda", 'gdp" and "market share" (市场占有率) in keywords of September to December 2011, but related to "junkets" (中介人) and "Chinese economy" (中国经济) in keywords of June to September 2012.
- 

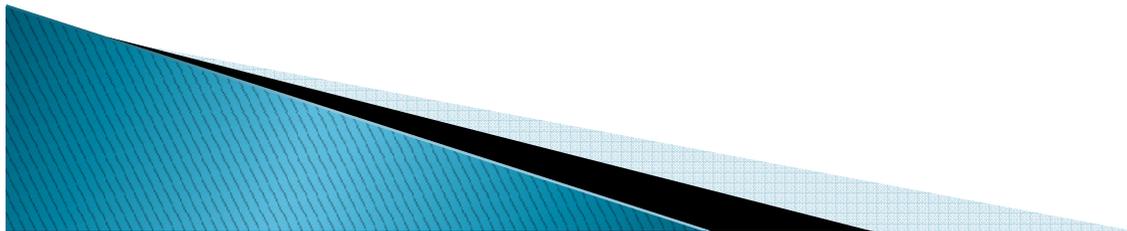
# Discussion

- ▶ The study also found the keywords of negative elements, such as “usury” (高利贷), was appeared in keywords of September to December 2011 but was not appeared in 2012.
- ▶ It can be induced that the public opinion focus changed from Macau's industrial diversification and negative impacts of gambling industry to economic issues; while the new projects of gaming resorts remained the focus of news reports in both of the time frames.



# Conclusion

- ▶ The contribution of the study was the development of new literature survey method in gambling studies to identify the keywords and the relationship among the keywords of Chinese news reports in two time frames, September to December 2011 and June to September 2012.
- ▶ The major conclusion implied that the public opinion focus changed from Macau's industrial diversification and negative impacts of gambling industry to economic issues; while the new projects of gaming resorts remained the focus of news reports in both of the time frames. Although the study showed the preliminary results of keywords analysis, it offered us more insights on the change of public opinion focus.



Thanks!

