iFUNDit: Visual Profiling of Fund Investment Styles

by

Bon Kyung KU

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Master of Philosophy
in Computer Science and Engineering

August 2020, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

BonKyung

Bon Kyung KU

27 August 2020

iFUNDit: Visual Profiling of Fund Investment Styles

by

Bon Kyung KU

This is to certify that I have examined the above MPhil thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.

Prof. Huamin QU, Thesis Supervisor

Prof. Dit-Yan YEUNG, Department Head

Department of Computer Science and Engineering, HKUST

27 August 2020

Acknowledgments

Foremost, I would like to express my deepest appreciation to my supervisor **Prof. Huamin QU** for his support and guidance throughout my two-years research journey. My official connection to Prof. QU started as I joined his project on Pulse of HKUST for my final year project. Under his supervision, I cultivated my interest in data visualization which attributed to my decision in pursuing further studies. I sincerely appreciate his earnest yet encouraging advice which always enchanted me with courage and motivation. As an admirer of his generous and open-minded characteristics, being under his mentorship was a delightful experience and I am much obliged for his kindness.

I would like to show my respect to **Prof. Xiaojun MA** and **Prof. Dimitris PAPADOPOULOS** for serving on my defense committee.

I wish to share my gratitude to my lab mates who have been friendly and courteous throughout. I owe special thanks to **Aowu YU**, who has been a good friend and a charismatic leader for Pulse of HKUST. His humble and eager-to-learn attitude helped me to reflect on myself. I also want to show my gratitude to my dearest and wonderful members of Pulse of HKUST, **Wai TONG** and **Leni YANG**, with whom I shared joys and hardships, and memorable moments.

I extend my appreciation to **Rong ZHANG**, **Yong WANG** and **Abishek PURI** with whom I spent nights together working on research papers. Furthermore, I would like to thank my Korean lab mates **Cheol Su JEON** and **Noah KIM** for their accompany, and my roommate **Se Hwan KIM** for his endless support. I also would like to express my gratitude to **Justin KWOK** who offered tremendous support towards Pulse of HKUST.

I feel sincerely indebted for my family who always has been supportive of my decisions and for putting exceeding efforts to provide perfect environment for me. Being born into such a harmonious family is a blessing for me.

Lastly, I would like to thank **Eun CHOI** for her unshakable support and boundless love that made me strong throughout the tough journey.

Contents

Title Page	i
Authorization Page	ii
Signature Page	iii
Acknowledgments	iv
Table of Contents	v
List of Figures	vii
List of Tables	ix
Abstract	x
1 Introduction	1
2 Literature Review	4
2.1 Fund Investment Style	4
2.2 Financial Data Visualization	5
3 Design Requirements	8
3.1 Background	8
3.2 Task Analysis	8
4 System Overview	11
5 Data Models	13
5.1 Terminology	13
5.2 Investment Style Factors	13

5.3 Financial Indicators 16
5.4 Unsupervised Clustering Algorithm 16

6 Visual Design 18
6.1 Manager View and Fund View 18
6.2 Cluster View 21
6.3 List View 22
6.4 Detail View 22
6.5 Temporal View 25

7 Evaluation 27
7.1 Fund-level investigation 27
7.2 Manager-level investigation 30
7.3 Domain Expert Interview 33

8 Discussion 35

9 Conclusion 37

Reference 38

Publication 44

List of Figures

4.1 *iFUNDit* is a web-based application that is comprised of three modules: database module, data-processing module and visualization module. Raw data from RQData API is preprocessed into fund- and manager-level data in the database module. Data-processing module handles on-demand data manipulation needed for visualizations. Views organized into two levels, Multiple-Fund level and Single-Fund level, allows top-down investigation: from exploration of distribution of multiple funds to in-depth analysis of single fund. 12

6.1 *Components of iFUNDit.* Manager View (B) shows the distribution of fund managers' performance attributes, such as return and risk. Fund View (C) displays the distribution of funds' performance attributes. Cluster View (D) projects the crowdedness of fund investment styles. List View (E) lays out the selected funds and managers. Detail View (F) visualizes the evolution of funds' investment styles. Temporal View (G) displays the return of funds and benchmarks. Control Panel (A) supports interactive exploration with different parameters. 19

6.2 Proposed design (top) and alternative design (bottom) for links connecting three views. The projection of points onto the y-axis using straight line is straightforward, whereas the curved lines might mislead the projected values. Comparing (a) and (c), we can observe that the curved lines intercept y-axis at a position which are deviated from the actually position of projections. This might introduce perception error and lead to misinterpretation of values. Similar observations can be found in (b) and (d). 20

6.3 *Explanation of our glyph design.* Our glyph comprises of two main parts. Radar chart (top) encodes 10 style-factor values of a fund. Circular sectors (bottom) encodes information about 11 sector-factors. Two components encoding two different aspects of investment style are combined to form our glyph design that represent the overall investment style of a fund. Grey text labels describe the visual encoding. 23

6.4 *Alternative design for glyph.* Two donut charts with different sector holding ratio have different colors but has similar shapes (A). Frequent change of positions of each donut makes it hard to observe sector-wise changes. In addition, donut chart do not encode zero values which makes it difficult to spot empty sectors. Fix-oriented circular sectors (B) avoids these issues by providing a clear visual distinction of sector changes and empty sectors. 25

7.1 Identifying funds that are similar to a given investment style. The user brushed the given investment style in the Cluster View (A), to show all funds with similar styles in the Fund View. The details of these funds were examined in the Detailed View (B). The Temporal View shows the performance of the funds, which assists users to evaluate the investment styles not only in different time periods, but also between different funds at the same time period. 28

7.2 A fund manager was on the manager lists of two funds with drastically different investment styles. The fund manager was a big ticket who has a high return and manage high assets. It was suspected that the manager was not the actual manager of both funds. 31

7.3 Detailed analysis revealed that relations between Manager A and Fund E/F. Manager A was the sole manager of Fund E. Therefore the investment style of Fund E represented that of Manager A. On the other hand, Manager A was one of the five managers of Fund B. The findings implied that Manager A was not the actual manager of Fund F. 32

List of Tables

5.1 Explanation of financial investment terminology 13

iFUNDit: Visual Profiling of Fund Investment Style

by Bon Kyung KU
Department of Computer Science and Engineering, HKUST
The Hong Kong University of Science and Technology

Abstract

The analysis of fund investment style is crucial for both fund managers and investors. It reveals the underlying investment strategy of a fund, which determines its performance. A clear profiling of a fund's investment strategy provides other fund managers with invaluable insights to enhance their investment strategies, and help investors assess the suitability of the fund investments regarding style preference and risk management. However, analyzing a fund's investment styles is challenging, as it requires a comprehensive analysis of high dimensional temporal data that are often difficult to explore, even for experienced fund managers.

To address this issue, we propose *iFUNDit*, an interactive visual analytic system for fund investment style analysis. The system decomposes funds' various attributes into two categories, namely, the performance attributes and the investment style factors; and visualize them in a set of coupled visualizations: a distribution view to display performance attributes of funds and managers, a cluster view to show the grouping of investment styles in the market, and a detail view to delineate the temporal evolution of investment style. The system provides a holistic overview of fund data and facilitates a streamlined analysis of investment styles at the fund and manager level. We demonstrate the effectiveness of the system through interviews with domain experts, and case studies by using the China mutual fund data set.

Chapter 1

Introduction

A mutual fund is an investment tool that allows many investors to pool money together to purchase securities. Compared to direct investment in individual securities, a mutual fund has outstanding advantages in terms of professional investment management and risk diversification, making it a popular investment choice. The mutual fund industry reached over $51.4 trillion in total assets in 2019 [1].

Investment style is the method used by a fund manager in selecting investments for a portfolio. Investment style is based on several factors and typically tends to be based on parameters such as risk preference, growth vs. value orientation, and/or market capitalization (cap). The investment style of a mutual fund is crucial for fund managers and investors [2] as it reveals the underlying investment strategy of a fund, which determines its performance. It should be noted that style investing is also an investment approach that rotates among different styles for successful investing. As opposed to investing in individual securities, style investors can decide to make portfolio allocation decisions by placing money in broad categories of assets, such as “large-cap”, “growth”, “international”, or “emerging markets”. Thus a clear profiling of the fund investment provides invaluable insights for managers to enhance strategies, and for investors to assess the investment better.

The analysis of fund investment style is challenging. To capture the underlying investment strategy, analysts evaluate a fund from a variety of aspects, such as the types of stocks, the economic sectors, and the trading frequency. The process generates a large amount of high dimensional temporal data, which is often difficult to navigate even for experienced fund managers. According to the feedback of the domain experts who collaborated with us in this study, in practice, fund managers often rely on dedicated departments in their institutions to investigate and summarize the strategies of competitor funds for them.

The evaluation of fund investment styles is even more inaccessible for investors due to the complexity of the analysis; therefore, they have to rely on the name or the prospectus of a fund to identify its investment style. Unfortunately, neither can reflect the actual investment style of a fund, because the strategy of a fund is determined by its fund managers who may adjust their strategies from time to time. In addition, it is not uncommon for a fund to change its managers and, as a result of this, its investment strategy changes completely.

Visualization has been employed to improve the analysis process of fund management. Currently, financial institutes still rely on basic visual representations, such as line graphs or other basic charts to perform an analysis. Yet, they are inefficient for handling the high dimensional temporal data that is required for investment-style analysis.

To improve the analysis of fund investment styles, three major challenges need to be resolved. First, the visual presentation of fund investment style need to be defined. Second, the comparison between different funds from the perspectives of both the fund performance and style should be enabled. Third, an interactive exploratory visualization system is required to enable a customized exploration subject to different criteria.

To address the issues above, we propose *iFUNDit* an interactive visual analytics system for fund investment style analysis. The system aims to assist fund managers and investors in investigating fund investment styles intuitively and efficiently. In particular, we propose a set of coupled visualizations: a distribution view to display the attributes of funds and managers, a cluster view to present the investment-style crowdedness in the market, and a detailed view to visualize the evolution of a fund's investment style. We combined the multi-factor Barra model [3] and the GICS economic sector categorization [4, 5], together with performance attributes, to characterize fund investment styles. The system provides a holistic overview of fund data, and facilitates the analysis on two levels: the Multiple-Fund level, which enables the distribution analysis of funds using different attributes, and the Single-Fund level, which supports a comparative analysis of investment styles and performance. We evaluated the effectiveness of *iFUNDit* through case studies with domain experts from financial institutes by using the China mutual fund dataset.

The major contributions of our study are as follows:

- A visual presentation of fund investment style, that displays the evolution of a style, and supports detailed comparisons between different styles.

- An interactive visual analytics system, that streamlines the analysis of fund investment style. It supports analysis on the bi-partite relations of funds and managers, from the perspectives of performance and investment style. The system emphasizes showing references and benchmarks, so that the analysis can be conducted in a comparative context which is critical for financial data analysis.
- A set of comprehensive case studies with domain experts from different financial institutes, along with the use of the China mutual fund data set, to demonstrate the effectiveness of the system.

Chapter 2

Literature Review

2.1 Fund Investment Style

Mutual fund investment styles are important signals for investors as it directly attributes to funds' performances. Due its significance, monumental efforts were put into inventing different methods of investment style analysis. Depending on the nature of investment style descriptions that result from the analysis, the analysis is be either qualitative or quantitative.

In a qualitative analysis, investigators analyze fund managers through their reports, speeches, or by conducting interviews with them in order to infer their investment styles. In the industry, people have adopted quasi-quantitative evaluation methods to categorize funds' investment style. One widely accepted method has been developed by Morningstar, Inc. [6]. It constructs a set of criteria for the analysis of an investment style. Based on the criteria, its analysts evaluate a fund and give a certain score, which eventually categorizes the fund into a three-by-three Style-Box matrix with a qualitative description. Compared to quantitative factor models, this method is more accessible to the public. It is convenient to qualitatively label a fund with a preset investment style, but this falls short concerning a detailed analysis of funds' investment style.

Quantitative research interprets styles through various factor models. Sharp, et al. introduced the classic CAPM one-factor model [7, 8] to evaluate the return of stocks, based on which a return-based style analysis RBSA model was proposed [9]. Fama and Fench et al. developed another classic three-factor Fama-Fench model to explain stock return [2, 10]. The model was then utilized to infer funds' investment style [11]. Carhart et al. added a momentum factor to the Fama-Fench model; thus, proposing a four-factor model [12]. Recently, Bar Rosenberg developed a multi-factor model, referred to as the Barra Risk Factor Analysis model, to interpret stock return from more dimensions [3, 13]. Depending on specific markets, the Barra Risk model constructs a comprehensive set of factors to evaluate stock returns, and can also be used to analyze the investment style of a stock portfolio. The Barra Risk model has been widely used for this, and it has been updated for different markets [4]. For example, in the China stock market, the CNE-5 model which contains 10 style factors was proposed by MSCI in 2012 [14]. Compared with other factor models, the Barra model interprets an investment style from more dimensions, which offers more explanatory options for fund managers and investors to portray an investment style. In this study, we adopted the CNE-5 model for investment style analysis.

Based on these models, people adopted return-based and holding-based analyses for funds' investment style [15, 16]. The former approach built models based on the funds' performance, such as return, volatility, etc. The latter focused on the funds' stock holdings. Researchers also proposed various measures to evaluate the managerial skills and investment styles of fund managers, such as Reliance on Public Information [15], Active Share [17], time-varying skills [18], etc.

The above-mentioned methods are designed for the analysis of a single fund. In order to compare between different funds or evaluate a fund in a different time period, visualization techniques are indispensable.

2.2 Financial Data Visualization

Fund data used for this study is multivariate time-series data, which is composed of performance data, holdings data and sector data over several quarters. General multivariate time-series techniques have been studied in the past decades for various applications [19–24].

For applications in the finance context, some study focused on improving classic visual forms such as scatter plot, line charts, and matrix in order to contain more information. StockViz [25] utilized scatter plots in spiral arrangement to visualized the historical stock prices of individual companies. Matthias et al. [26] used line chart with segmented background and color encoding to display stock returns. Ziegler et al. [27] proposed pixel-based performance matrix to visualize volatility and return of funds in long-term investment. Xuanwu et al. visualized factor data in quantitative investment portfolio with radar charts and line charts [28].

In order to increase the information contained in a two dimensional space, visualization in the form of maps have been developed. A number of work were based on heat map techniques [29–31]. Alsakran et al. proposed density-based distribution map and tile-based parallel coordinate system [29] to visualize multivariate financial data [32]. Csallner et al. and Jungmeister et al. [33] adopted tree map graph to visualize the stock holding diversity of mutual funds [34]. Xiong et al. showed the performance of funds in geographic maps [35]. Lei et al. analyzed the volatility of stock market in a ring-shape visualization design [36].

Researchers also explored techniques to visualize fund data in three dimensional space. Dwyer et al. [37] visualized the relation graph of fund manager portfolio data in three dimensional space to show the changing investment behavior of fund managers. The authors also used multidimensional scaling techniques to map high dimensional data of fund holdings, to two dimensional spaces, and then mapped time to the third dimension, to show the movement of fund manager's stock portfolios [38, 39]. Comprehensive surveys about visualization techniques were conducted. Aigner et al. summarized time-series related visualization for general applicaiton [40]. Ko et al conducted a survey on visual analysis approach for financial data [41]. FinanceVis.net [42] summarized finance-related visualizations before 2014 in an interactive system.

In this study, in addition to the investment style of funds, the bi-partite relationships between funds and managers are also visualized. Techniques regarding bi-partite visualization were reported in various applications [43–45]. However, these techniques are not suitable to handle time-series data and cannot be simply applied in our study.

In summary, for fund related applications, most of the previous visualization work focused on either individual stocks, or on an aggregated fund level. The data was either about the performance, or stock holdings. However, in order to analyze the investment style, not only the performance metrics, but also the stock holdings as well as style factors from both the perspectives of funds and fund managers are also important. Yue et al proposed a system [28] to visualize the style factors in portfolio management. However, it uses only the return as the performance measure of portfolios, and does not support an analysis from the perspective of portfolio managers. In order to conduct a comprehensive analysis of funds' investment styles, we proposed to visualize two separated but related data sets simultaneously. One is the investment style factors, the other is the performance metrics. These two data sets need be visualized in a coupled visual form in order to give an informative holistic view of funds' investment style. To the best of our knowledge, there is no specific design that achieves the requirements in fund investment scenario.

Chapter 3

Design Requirements

3.1 Background

A consensus about investment is that there is no "best" investment choice. High return is often associated with high risk. The balance between them is made based on personal judgement and preferences. In addition, to evaluate an investment style, many features are involved. According to our survey and the feedback from domain experts, there is no standard model to characterize and evaluate an investment style.

As a result of the these inherent characteristics, different people evaluate funds/managers from different perspectives, using different metrics. Professional institutions use various in-house systems to evaluate funds/managers. For individual investors, the evaluation is even more challenging. They have to rely on information scattered on the Internet, often without understanding all the features about a fund/manager, to make their judgements, which can often be misled by information such as advertisement and promotions.

Therefore, it would be beneficial to build an evaluation system for individual investors as well as professional institutions, to assist them in understanding the investment styles of funds/managers, and improving the efficiency of their evaluation process.

3.2 Task Analysis

We collaborated with six domain experts in this study. E_1 is a senior fund manager with more than ten years of experience in a top fund institution in Mainland China. He manages funds with over \$10 billion CNY asset. E_2 is a finance researcher who studies funds. E_3 (a co-author of this paper) and E_4 are financial product managers. E_2, E_3, and E_4 are from RiceQuant, a financial data and service provider whose clients cover a variety of mutual funds, private funds, securities institutes and banks in Mainland China. These experts have extensive experience with financial institutes and investors, and have a clear understanding of industry requirements. We also consulted two stock traders, E_5 and E_6, in order to have a better understanding of the stock market.

To ensure that the system fulfills domain users' requirements, we conducted a series of structured interviews with the experts to identify their primary concerns. After several iterations of discussions and system developments, we have decomposed the requirements to a list of tasks from three levels: single-fund level, multiple-fund level and system level.

Single-fund level:

- **T1** *How to characterize a fund investment style and visualize it?* It involves many features to profile an investment style, such as the types of stocks, the sector of economics, the trading frequency. It is crucial to characterize and further visualize the investment styles in an intuitive way, such that investors can easily understand them.
- **T2** *What is the temporal evolution of the investment style of a fund?* The investment style of a fund is intrinsically dynamic. A fund manager may adjust the investment style periodically. The visualization should capture the dynamic changes of the investment style in different time periods.
- **T3** *How to evaluate the correlation between the fund investment style and its performance?* It is crucial to evaluate a fund's performance, such as return and risk. These performance metrics are the results of an investment style, and often have a great influence on investors' final investment choice. The visualization should provide a clear mapping between an investment style and its performance metrics, so that investors and fund managers can easily explore the correlation between them.

Multiple-fund level:

- **T4** *How to find similar/different funds effectively, in terms of investment styles or performance metrics?* It is critical for investors to be able to select funds with different investment styles to diversify the investment risks. It is also beneficial for fund managers to be able to group funds with similar investment styles together, so that they can study the investment styles systematically. The system should be able to cluster funds according to their investment styles and performance metrics.
- **T5** *How to compare different investment styles effectively?* Conventionally, it is difficult

to compare two investment styles in detail, because it involves many features to characterize an investment style. The visualization should not only visualize an investment style in great detail, but also enable an effective comparison between different investment styles.

T6 *How to compare both the performance metrics and investment style of different funds, in the context of reference?* Benchmark is critical for evaluating financial products. It gives investors a sense of how good or bad a fund is. For example, comparing to just saying "A fund has a return of 30%.", it is more informative to say "A fund has a return that exceeds 75% of all funds in the market." for decision-making.

T7 *What is the relation between funds and managers?* There is a bi-partite relation between fund managers and funds. A manager can run many funds. Do these funds have a similar investment style? On the other hand, a fund can also have many managers. Do these managers have a consistent style? Who determines the final investment style of the fund? Showing the relation between the fund managers and funds helps investors and analysts to identify which manager dominates the investment style of a fund, and therefore allow them to evaluate the fund more comprehensively.

System level:

T8 *How can the users explore different attributes of a fund according to users' needs?* Different users put emphasis on different aspects when they evaluate a fund. An investor may value more on the return of a fund, while a fund manager may pay closer attention to a certain factor in an investment style, such as the capitalization size of the stocks, in order to analyze the style from a certain perspective. The system should enable users to explore different criteria to evaluate a fund through intuitive interactions.

Chapter 4

System Overview

iFUNDit is a web-based application that is comprised of three modules: database module, data-processing module and visualization module. Figure 4.1 shows the overview of the system pipeline. The database module employs MongoDB to store both the raw data from RQData API and the processed data. Raw data refers to unprocessed data collected from RQData API that consists of full records of Chinese mutual funds on asset allocation, asset values, holdings, financial indicators and manager records and other related information such as daily stock prices and daily stock factor exposures.

According to the regulations, a stock-oriented fund should hold at least 60 percent of stock in its asset allocation. Therefore, funds that holds stock positions less than 60 percent have been filtered out in this study. As a result, the dataset consists of 2398 stock-oriented funds.

Processed data consists of temporal data at fund- and manager-level which is calculated from the raw data. It includes fund-level information on quarterly factor exposure, quarterly sector positions and daily financial indicators and daily asset values. From the raw data which is only provided at fund-level, the manager-level information on quarterly factor exposure, quarterly sector positions and daily financial indicators and daily asset values were calculated with using weighted average of associated funds. In addition to fund- and manager-level information, processed data also consists of mapping information of fund to manager and vice versa.

The data-processing module, which is implemented using Python Pandas and scikit-learn, handles data manipulation involved in the system such as aggregation of temporal data, filtering and clustering algorithm discussed in Section 5.4.

The visualization module is used for communication of data to the users in various intuitive format. The system consists of six interactive views that function in harmony to

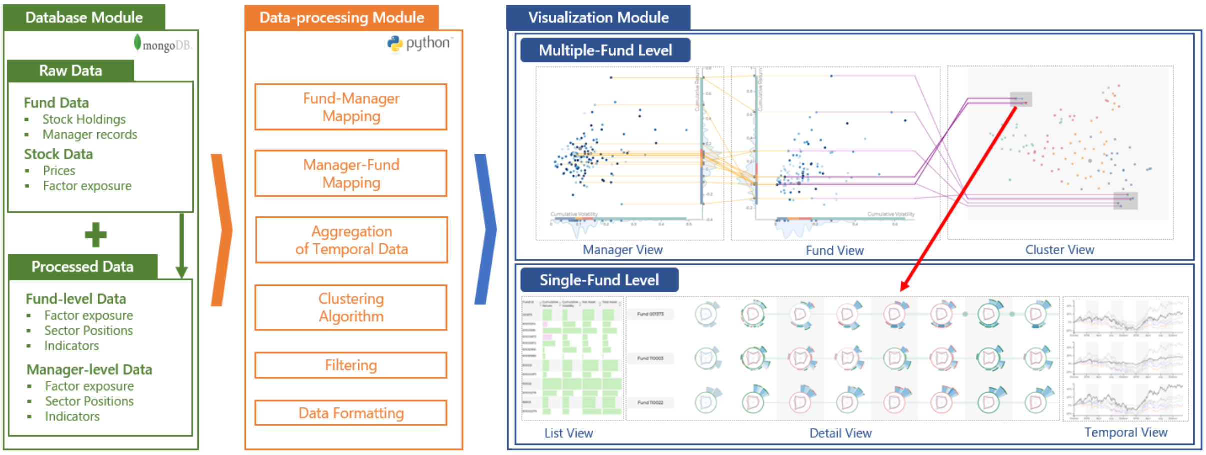

Figure 4.1: *iFUNDit* is a web-based application that is comprised of three modules: database module, data-processing module and visualization module. Raw data from RQ-Data API is preprocessed into fund- and manager-level data in the database module. Data-processing module handles on-demand data manipulation needed for visualizations. Views organized into two levels, Multiple-Fund level and Single-Fund level, allows top-down investigation: from exploration of distribution of multiple funds to in-depth analysis of single fund.

provide visual-assisted investigation on funds investment styles and performances. The Manager View (Fig. 6.1B) and the Fund View (Fig. 6.1C) summarizes the distribution of managers and funds in two-dimensional space. The Cluster View (Fig. 6.1D) displays the results of tSNE clustering algorithm on funds using different style features discussed in Section 5.2. The Lists View (Fig. 6.1E) displays comparison and sorting of selected funds based on aggregated values of performance metrics. The Detail View (Fig. 6.1F) allows in-depth inspection of investment style changes of funds. The Temporal View (Fig. 6.1G) is used with the Detail view to find style-performance correlations.

Chapter 5

Data Models

5.1 Terminology

Terminology used in financial investment are reasonably straightforward. However, for clarification, Table 5.1 shows the list of terminology used in the papers and its contextual meaning.

Term	**Explanation**
Risk Factor	A set of generalizable market factors that impact return.
Factor Exposure	A measure of how much a stock or a portfolio is exposed to a certain risk factor.
Equity Fund	A portfolio is a grouping of financial assets such as stocks.
Cumulative Return	Return in percentage gain or loss, with respect to initial value.
Maximum Drawdown	The maximum drop of the asset value in a given period. It is used to measure the risk.
Turnover rate	Turnover rate of a fund. It measures the difference of a fund's stock holdings between two quarters.
Sector	A large segment of the economy. Not to be confused with the sector of a circle. For sector of a circle, we will explicitly name it as a *circular sector*.
CSI300 Index [46]	A capitalization-weighted stock market index that follows the top 300 stocks in Shanghai and Shenzhen stock exchange.

Table 5.1: Explanation of financial investment terminology

5.2 Investment Style Factors

In this study, in order to characterize funds' investment style, we proposed to combine the 10 style-factors from the Barra China Equity Model (CNE5) [14], and 11 sector-factors according to the Global Industry Classification Standard (GICS) [4, 5, 47], to construct a 21-factor metric to infer funds' investment style.

The CNE5 model uses state-of-the-art risk assessment methodology to captures the short-term and long-term dynamics of the China market. It help investors align the risk model with their investment processes. The factor model [48] uses an assumption that there exists a set of K common factors that drive stocks return. Then the equation for the return of stock could be written as follows:

$$r_i^t = \sum_{k=1}^{K} X_{ik} f_k^t + \varepsilon_i^t, \qquad i = 1, 2, \ldots, M;\ t = 1, 2, \ldots, T \tag{5.1}$$

where,

r_i^t = Return of stock i at time t

f_k^t = Return of the factor k at time t

X_{ik} = Exposure of stock i on factor k, for the time period $t = 1, 2, ..T$.

ε_i^t = Stock's specific return, which cannot be explained by the factors at time t

Based on Equation 5.1, multivariate linear regression [49] is used to estimate the factor exposure $\{X_{ik}\}_{i=1,2,...,M;\, k=1,2,...,K}$.

For a single fund F consisting of N stocks, the return of fund F at time t is the weighted average of individual stock returns:

$$R_F^t = \sum_{i=1}^{N} w_i r_i^t, \tag{5.2}$$

where,

w_i = Weight of stock i

The fund's exposure to factor k is given by the weighted average of the stock exposure:

$$X_k^F = \sum_{i=1}^{N} w_i X_{ik}, \tag{5.3}$$

Because different factors has different scales, when we use the factor exposure in practice, we need to standardize the raw exposure to have mean = 0 and standard deviation

= 1. Moreover, to make sure the market index has 0 exposure to every factors, we need to deduct the capitalization weighted average of the raw exposure from the raw exposure before standardization. Thus the factor exposure of the CNE5 is calculated as follows:

$$X_{ik} = \frac{X_{ik}^{Raw} - \mu_k}{\sigma_k} \tag{5.4}$$

where,

$$\mu_k = \sum_{i=1}^{N} w_i^{Cap} X_{ik}^{Raw} w_i^{Cap} = \frac{Market\ capitalization\ of\ stock\ i}{\sum_{j=1}^{N} Market\ capitalization\ of\ stock\ j} \tag{5.5}$$

Factor exposure of the 10 style factors is used to measure how much a fund is exposed to a certain style factor, which infers the investment style of the fund. Positive factor exposure means the fund exposes to a certain style factor more than the market index, while negative factor exposure means the fund exposes to a certain style factor less than the market index. Zero factor exposure means the fund has the same exposure as the market index to a certain style factor. For example, a positive *Size* factor exposure means the fund allocates more asset in stocks with large market capitalization, or "large-cap" in short, while a negative *Size* one means the fund prefers small-cap stocks. A high *Book to Price* factor exposure indicates that the fund implements a "value" investment strategy, which means it buys stocks whose market price is "underrated", while low *Book to Price* shows that the fund uses a "growth" strategy and invests in stocks that are "overpriced" but have potential to grow even bigger.

There are in total 10 style factors to evaluate the investment style from different perspectives. The 10 factors are: *Beta*, *Book to Price*, *Earning Yield*, *Growth*, *Leverage*, *Liquidity*, *Momentum*, *Non-linear Size*, *Residual Volatility*, and *Size*.

In addition to the 10 style factors, we constructed 11 sector factors according to the Global Industry Classification Standard (GICS) [4, 5, 47], in order to characterize a fund investment style in more detail. The sector factors indicate to which economic sector a stock belongs. We used one-hot encoding to indicate whether a stock belongs to a specific sector or not.

In summary, the 10 style factors and 11 sector factors together reflect what types of stocks the fund invest in. The 21 factors together characterize the investment style of a fund.

5.3 Financial Indicators

Maximum Drawdown refers to the the maximum downwards drop of asset value of a fund between a given time period. Since normal volatility measure that is commonly used to take into account of both upwards and downwards volatility, maximum drawdown is an indicator that better represents the downward volatility [50].

$$Maximum\ Drawdown = \max_{\tau \in (0,T)} \left\{ \max_{t \in (0,\tau)} P(t) - P(\tau) \right\} \tag{5.6}$$

Turnover Rate is a measurement that reflects how frequently the holdings are bought and sold by a fund between a given time period [51].

$$Turnover\ Rate = \frac{min(Buy, Sell)}{(Initial\ Asset\ Value + Final\ Asset\ Value)/2} \tag{5.7}$$

5.4 Unsupervised Clustering Algorithm

In order to project high dimensional style features of fund into two-dimensional space, we utilized two unsupervised dimensionality reduction and clustering algorithms, t-SNE [52] and MDS [53]. Considering that that our emphasis is on the clustering based on local similarities of funds, t-SNE is an effective method to produce visually distinguishable cluster structure while preserving local similarities and achieve stable convergence to global optimal values. Due to stochastic nature of t-SNE, the orientation of clustering results might appear different, however, it still appears to produce consistent community structure. MDS was chosen as another clustering option since MDS would preserve the proximity or distance between clusters unlike t-SNE where proximity between clusters does not possess any significant value. While t-SNE could yield cluster structures that visually distinguishable, proximity-preserving MDS could be helpful for situations that require careful examination of distance of clusters.

Parameter settings of the clustering algorithms could pose significant impact on the results. However, it is expected that the users of our system are from non-technical background. Therefore, we have carried out various experiments on our dataset and tuned the parameters following the advice of a well-established guideline [54].

In terms of usability of clustering algorithms in the system, both algorithms run at a considerable speed. In general cases involving less than 500 data points, both algorithms runs in less than three seconds in a laptop with four CPUs and one built-in GPU. While running on full 2398 data points, t-SNE runs in ten seconds but MDS runs in four minutes.

$$P = Cluster(V) \tag{5.8}$$

A fund with N style features could be represented as a vector v_i with size N. Given a set of fund vectors V, clustering algorithm calculates P, where V is a collection of v_i and P is a collection of p_i which is the relative position of each v_i in two-dimensional space.

Chapter 6

Visual Design

iFUNDit consists of six coordinated views: the Manager View (Fig. 6.1B), the Fund View (Fig. 6.1C), the Cluster View (Fig. 6.1D), the List View (Fig. 6.1E), the Detail View (Fig. 6.1F) and the Temporal View (Fig. 6.1G). We have carefully designed the system framework by following Shneiderman's mantra [55]: *"Overview first, zoom and filter, then details on demand."*

The system provides an overview of the performances and styles of funds through the Manager View, the Fund View and Cluster View (Figs. 6.1B,C,D). Upon user interaction, system allows filtering of funds according to different criteria and zoom-in to check details that are displayed in the List View, the Detail View and the Temporal View (Figs. 6.1E,F,G). The linked analysis across different views allow users to carry out top-down investigation, hence fulfilling the requirements discussed in Section 3.2.

For the visual designs of *iFUNDit*, we have adopted design principles that make use of references or benchmark, and relative scales such as percentile. This is to ensure that our system is consistent with the nature of the domain context where there is no ground truth for classifying a fund as *good or bad*. In the context of financial investment, evaluation techniques often involve a comparison to some generalizable benchmark fund selection methods or ranking techniques which can help determine whether a fund is *better or worse*. Therefore, in our system, the performances and investment styles of funds are presented in a comparative context, making it more convenient for fund investment decision-makings.

6.1 Manager View and Fund View

The Manager View (Fig. 6.1B) and the Fund View (Fig. 6.1C) provide overviews of distribution of all funds in two-dimensional space and facilitate convenient exploration

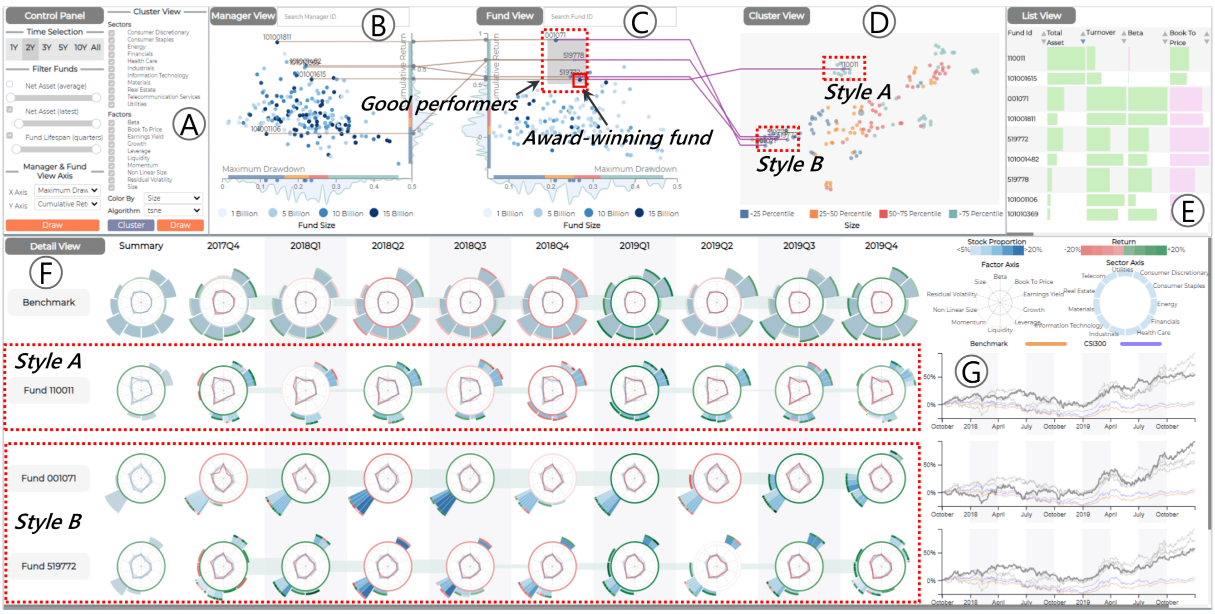

Figure 6.1: *Components of iFUNDit.* Manager View (B) shows the distribution of fund managers' performance attributes, such as return and risk. Fund View (C) displays the distribution of funds' performance attributes. Cluster View (D) projects the crowdedness of fund investment styles. List View (E) lays out the selected funds and managers. Detail View (F) visualizes the evolution of funds' investment styles. Temporal View (G) displays the return of funds and benchmarks. Control Panel (A) supports interactive exploration with different parameters.

of the relative positions of funds and managers in entire mutual fund market (**T4**). The Manager View and the Fund View are used in combination with the Cluster View to visualize the mapping between the performance and the investment styles (**T4**). The Manager View, the Fund View and the Cluster View, which are the Multiple-Fund level visualizations, act as the entry point to the system.

Using an augment scatterplot with selectable axes, the Manager View and the Fund View allow intuitive exploration of the data distribution from various perspectives. Each axis of the scatterplot is equipped with a colored percentile ribbon, a density plot and projections of coordinate values of selected funds which displays the positions of selected funds in the distribution. This helps users to understand the rankings of funds in the distribution (**T6**). Direct interpretation of absolute numerical values often bring ambiguity, especially in the context of financial metrics. Thus, the Manager View and the Fund View assist users in gaining a clear evaluation of a fund's performance with regards to its competitors. The color of each circle encodes the *latest net asset size* managed by

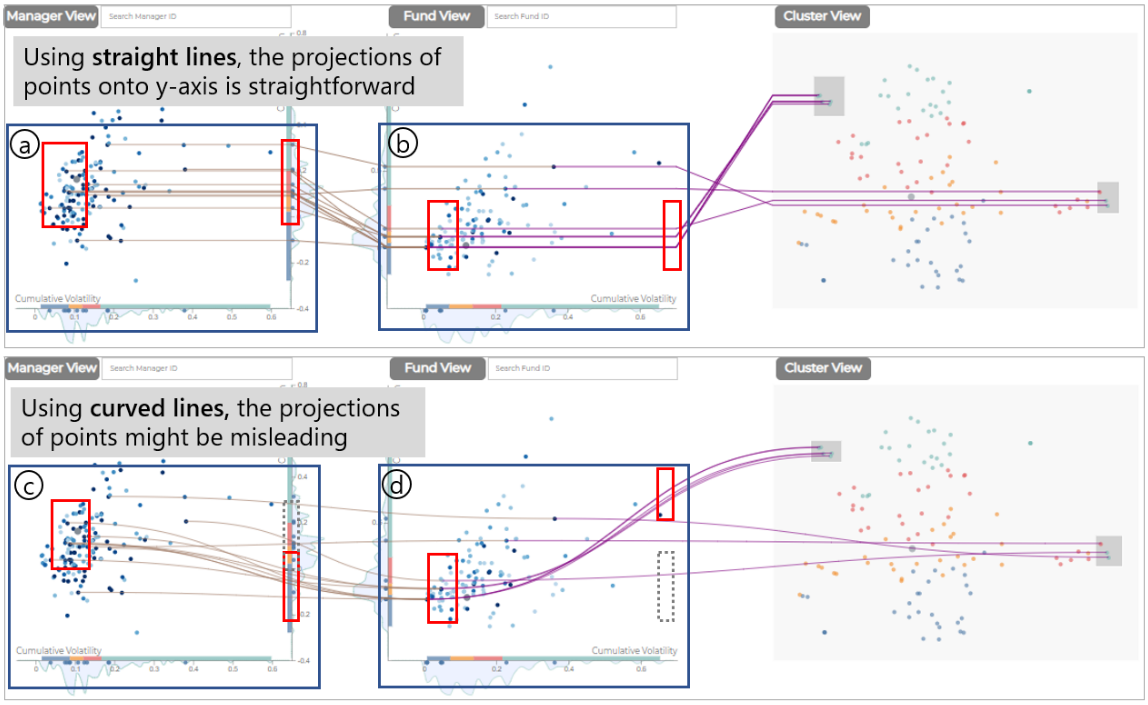

Figure 6.2: Proposed design (top) and alternative design (bottom) for links connecting three views. The projection of points onto the y-axis using straight line is straightforward, whereas the curved lines might mislead the projected values. Comparing (a) and (c), we can observe that the curved lines intercept y-axis at a position which are deviated from the actually position of projections. This might introduce perception error and lead to misinterpretation of values. Similar observations can be found in (b) and (d).

funds and managers as the asset size is an important characteristic that describes a fund. Y-axis of Manager View and Fund View are aligned side-by-side which allows projection of y-coordinate values between the two views, in form of parallel coordinates.

The links between the Manager View, the Fund View and the Cluster View enable an easy reference to the associated managers and funds across multiple views, and provide visual cues for mapping between performances of managers and funds, and between fund performances and investment styles (**T4, T7**). The selection of a manager or fund in one of the three views will highlight the links connecting the associated funds and managers across the three views. The selection of manager and fund could be done by brushing on any of the three views or by direct queries using search boxes embedded on top of both view. Along with the selection, the corresponding manager(s) and fund(s) will also be added to the List View, the Detail View and the Temporal View, which are Single-Fund level visualizations.

Justification: We carefully designed the Manager View and Fund View and some other alternative designs are also considered. Here we further clarify our design choices.

1) *Use of colored percentile ribbons.* An alternative method for visualizing percentiles is using a boxplot. However, design of boxplot causes visual clutter when it is placed on the axes because the lines and boxes of the boxplots obstruct ticks of the axes. Hence, we use color channel to encode percentile ranges using the percentile ribbons which is more distinguishable. The color scheme is consistent with the percentile color scheme used in the Cluster View, minimizing the color diversity.

2) *Use of color to encode asset size of funds and managers.* In Manager View and Fund View, we used the color of a circle to encode the asset size of a fund instead of the area of a circle. This contradicts with the common visualization practice of using magnitudinal channel to encode quantitative values. This decision was made due to the nature of the fund market distribution, where the majority of the funds dwell near the "average", forming a distribution with high kurtosis. Using circles of varying sizes in these dense regions can result in severe visual clutters and make each circle indistinguishable. Therefore, we use color to encode the asset sizes.

3) *Use of parallel-coordinate style links to show the mapping between adjacent views* The current design for links indicates the mapping of values between adjacent views. Alternative design which uses horizontal tangent curved links can effective soothe the overlapping of links, making it easier to track individual lines. However, curved lines might intercept the axis in positions that is deviated from the actual positions of the projection of points onto the axis (Fig. 6.2). This could introduce confusion in interpreting the coordinate mapping between views, since their coordinate values are not fully preserved. Using straight lines that project onto the axes directly can provide more accurate mapping of coordinates values. Main objective of the links between the views is to provide a visual cue, in terms of clustering of links, that maps values between the views. Therefore, we put priority on producing clustering of links that incorporate accurate coordinate values rather than the visibility to track individual links.

6.2 Cluster View

The Cluster View (Fig. 6.1D) displays the clustering structure of funds with respect to their investment styles. Multidimensional style features of funds are projected onto a 2D plane by using unsupervised clustering algorithm aforementioned in Section 5.4. In the context of investment styles, there are no predefined classifications and it brings difficulties in identifying investment style clusters (**T4, T5**). To address this issue, the Cluster View supports customizable features and color labels, which maximizes the efficiency in carrying out data-driven cluster analysis.

In the Control Panel (Fig. 6.1A), the users can choose either t-SNE or MDS as the desired clustering algorithm. The user can customize the combinations of features used for clustering algorithm and select the feature for color labels to be used for the cluster. For color labels, we transformed the continuous numerical values to four percentile categories: [0, 25), [25, 50), [50, 75), [75,100). The reason for the categorization was because the use of continuous labels might pose ambiguity when identifying clusters, and we also considered the financial context where relative values are more important than the absolute values.

6.3 List View

The List View (Fig. 6.1E) displays the details of other performance metrics of funds and managers, which are not covered in the Manager View and the Fund View. The List View allows a comparison and sorting of the selected funds from various perspectives (**T6, T8**). The performance metrics of managers associated with the funds are also shown, enabling a convenient comparison of managers within a fund or across different funds (**T7**).

In the List View, each row represents a fund with aggregated values of different metrics. The magnitudes of values are encoded using horizontal bars, which allows an intuitive comparison across rows and columns. The color of bars encodes the sign of the values (i.e. positive or negative). The rows can be sorted according to desired columns. Rows are expandable and collapsible on click to display the associated managers of funds. Row arrangement of the List view is synced with the order of funds in the Detail View and this allows different arrangements of funds for style comparison.

6.4 Detail View

The Detail View (Fig. 6.1F) displays the details of funds in terms of investment style, performance and holdings (**T1**). The Detail View depicts the temporal evolution of a fund (**T2**), for example, how the different aspects of a fund change over time? Through an observation of the style and performance changes of a fund, the users can estimate

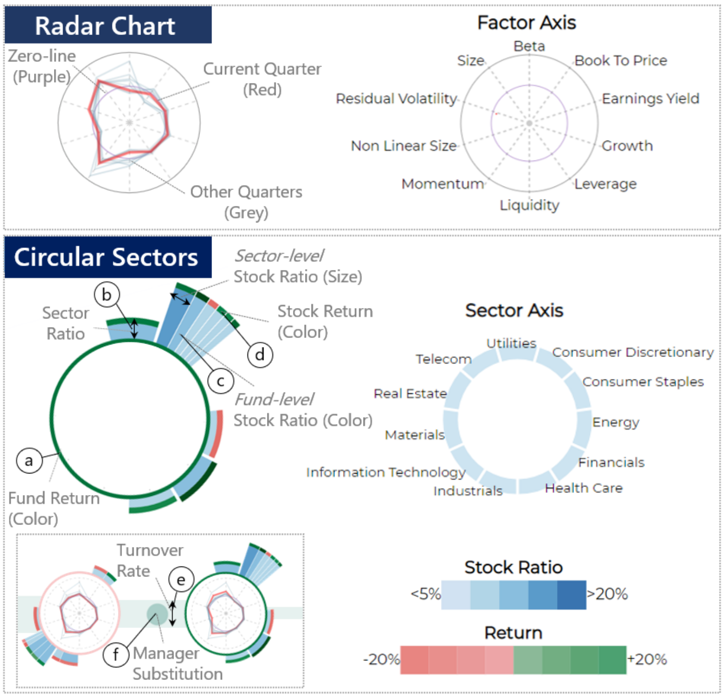

Figure 6.3: *Explanation of our glyph design.* Our glyph comprises of two main parts. Radar chart (top) encodes 10 style-factor values of a fund. Circular sectors (bottom) encodes information about 11 sector-factors. Two components encoding two different aspects of investment style are combined to form our glyph design that represent the overall investment style of a fund. Grey text labels describe the visual encoding.

how the style changes have influenced the fund performance (**T3**). The Detail View also provides insights into the rationale behind style change of a fund. Style change of a fund could be due to the adaptation of existing manager to market situations or substitution of the manager. In addition to single-fund analysis, the Detail View allows in-depth comparisons of investment style between funds.

In the Detail View, each row displays a fund's investment style and performance along time. Within each row, each glyph represents the style and performance of each quarter. As mentioned in Section 3, the investment style of a fund is represented by 10 style-factors and 11 sector-factors. The performance is represented using quarter return in two granularity: fund-level and individual stock-level.

Glyph consists of two main components: *center radar chart* and *circular sectors*. *Center radar chart* (Fig. 6.3) encodes 10 style-factors of a fund. The red highlighted radar represents the current quarter and the grey lines represent other quarters. Since magnifying the differences from radar charts that are arranged side-to-side could be difficult, we deliberately added radar charts of other quarters which could act as direct references for comparison. *Colored ring* (Fig. 6.3a) between the radar chart and the circular sectors encodes the fund's overall quarter return. *Circular sectors* (Fig. 6.3) encode various information about 11 sectors. In order to avoid confusion and to emphasise that two parts of the glyph have distinct encoding, we added a gap to detach the two parts and used non-aligned axes. Each circular sector (Fig. 6.3b) is oriented in a designated directions, each representing a sector, and its height encodes the sector ratio. Each circular sector is partitioned, with each partition representing an individual stock holding. Size of each partition encodes sector-level stock holding ratio. Each partition is divided into its inner (Fig. 6.3c) and outer part (Fig. 6.3d). The color of the inner part of the partition encodes the fund-level stock holding ratio. The color of the outer part of the partition encodes the quarter return of the stocks. Hovering over the circular sectors displays the sector-specific stocks details such as stock ID, quarter return and holding ratio. The glyphs are connected by turnover bridges (Fig. 6.3e) and the thickness of the bridges encodes turnover rates. Any substitution of manager is marked by a circle (Fig. 6.3f) on the turnover bridges. Hovering over the circle displays the details of the respective manager.

The uppermost row in the Detail View displays the benchmark which is the average values of all stock-based funds operated in the time period. Benchmark allows users to observe the overall trend of the market and make comparison to the market average (**T6**).

Justification: We carefully designed the Detail View and our detailed considerations for the key design choices are as follows:

1) *Fix-oriented circular sectors to encode sectors.* An alternative design to visualize the eleven sectors is by using a donut chart. In a donut chart, different sectors are encoded using categorical colors. The sector holding ratios are encoded as the size of the donut. Given that the sector holding ratios always add up to 100 percent, donut chart is an effective visualization design for representing such proportional values. However, the design requirement for the glyph needs it to effectively visualize the investment style over time, which means that the two glyphs with different sectors must be differentiable. It is possible that two donut charts with totally different sector ratios to have indistinctive shape except colors. The frequent change of positions of each sector donut makes it hard to observe sector-wise changes. Additionally, the donut chart does not encode zero values, which makes it difficult to spot empty sectors. Fix-oriented circular sectors avoids these issues by providing a clear visual distinction of sector changes and empty sectors. Also,

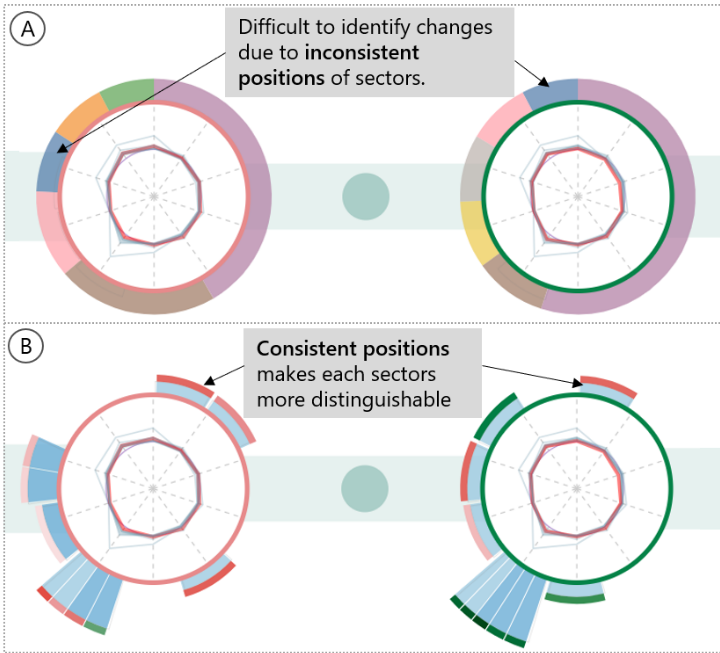

Figure 6.4: *Alternative design for glyph.* Two donut charts with different sector holding ratio have different colors but has similar shapes (A). Frequent change of positions of each donut makes it hard to observe sector-wise changes. In addition, donut chart do not encode zero values which makes it difficult to spot empty sectors. Fix-oriented circular sectors (B) avoids these issues by providing a clear visual distinction of sector changes and empty sectors.

it provides additional insights into the contributions of each stock to the fund's overall return, which are encoded by the outer part of the circular sector.

2) *Turnover bridges.* Although turnover rate is calculated every quarter, turnover bridges are placed between two adjacent quarters, as the word *"turnover"* itself indicates how the fund holdings change *between* two adjacent quarters.

6.5 Temporal View

The Temporal View (Fig. 6.1G) shows the performance of the fund during the selected time period. In the Detail View, the performance of a fund, in terms of quarterly return, is visualized to support a direct comparison of performance and style. Although it can act as direct reference for quarterly performance and style, it is not sufficient for identifying temporal patterns and fund-wise comparison. Hence, the Temporal View complements the Detail View by providing direct visual cues on the performance comparison between funds and against the benchmarks (**T3, T6**). In the Temporal View, the selected fund is highlighted with dark grey, and other funds are in light grey. The benchmark and CSI300

Index are in orange and purple respectively.

Justification: Multiple line charts with more than 5 lines can bring severe visual clutters, due to the dense crossing lines and color diversity. It has a bad impact on effectively exploring the temporal evolution of funds and conducting fund comparisons. To alleviate this issue, we adopted minimal color usage, using only four colors. Comparison between a fund and the benchmarks could be done within a single window, comparing the dark grey with orange and purple line. Comparison between multiple funds can be done using multiple windows, comparing the dark grey lines.

Chapter 7

Evaluation

In this section, we present three case studies to illustrate the practical usage of the system. The main objective of *iFUNDit* is to assist user in profiling different investment styles of funds, and in performing comparative analysis between different funds. To evaluate the system, we conducted case studies with a senior fund manager (E_1), a financial researcher E_2, and two product managers from RiceQuant (E_3, E_4). E3 is a co-author.

7.1 Fund-level investigation

What are the investment styles of good funds? The fund manager, E_1, would like to analyze the investment styles of top performing funds on the market to gain some insight (**T1, T3**). This was a routine he performed regularly at work. Normally, the analysis was performed by a dedicated department in the fund institution he served in. The team analyzed the good performing funds and competitor funds, and summarize reports to E_1.

E_1 specified his investigation scope to funds with asset size over $3 Billion and with at least two-year history, and decided to study their two-year performances (**T8**). He selected the two axes for the Manager View and the Fund View as *cumulative return* and *maximum drawdown*, which are the two key performance attributes he prioritized (**T8**). E_1 then selected all 21 styles features to cluster funds, and used *Size* factor as the color label in the Cluster View (**T8**). He chose *Size* factor as he would like to focus on the capital sizes of the stocks that the funds invest in. He chose the t-SNE algorithm because he was interested in the local clustering structure. E_1 set these parameters in the Control Panel (Fig.6.1A), and clicked "Draw" buttons to plot the distributions of funds that satisfied his requirements in the Fund View, Manager View, and Cluster View.

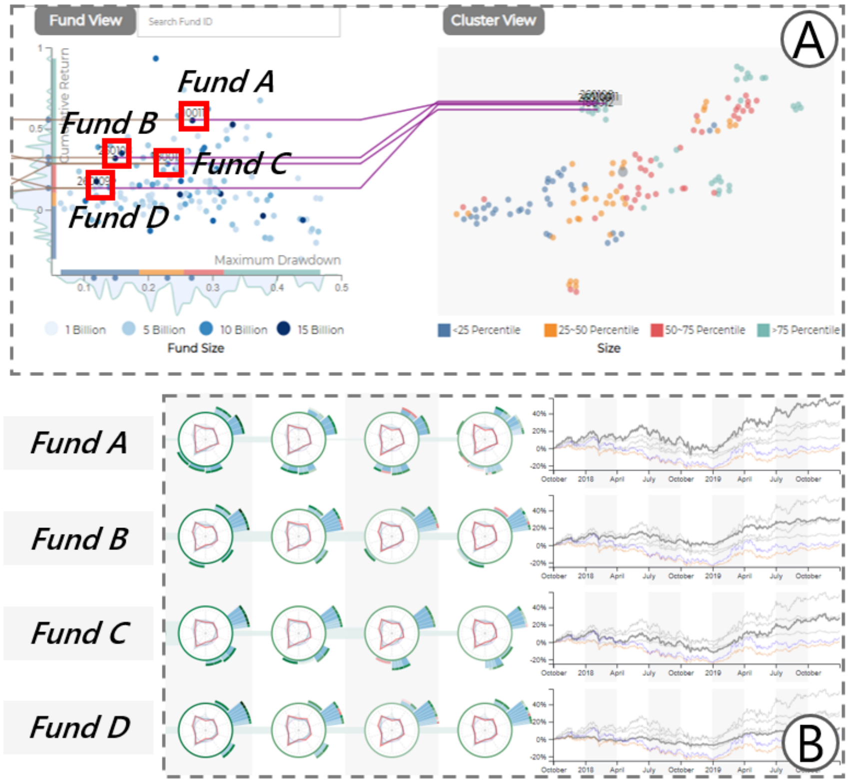

Figure 7.1: Identifying funds that are similar to a given investment style. The user brushed the given investment style in the Cluster View (A), to show all funds with similar styles in the Fund View. The details of these funds were examined in the Detailed View (B). The Temporal View shows the performance of the funds, which assists users to evaluate the investment styles not only in different time periods, but also between different funds at the same time period.

E_1 observed several funds with relatively high return and medium risk in the distribution (**T6**), hinted by the quantile ribbons. He brushed the region in the Fund View (as shown in Fig.6.1B) to select the funds for detailed investigation (**T2**). Upon brushing, names of selected funds and managers were displayed and the links connect these funds in the Fund View and the Cluster View (**T3**). From the name of the fund, E_1 recognized one fund, Fund A, which received the *Fund-of-the-Year-2020 Award* from Morningstar[6] in March 2020.

E_1 then examined the Cluster View and it revealed that the investment style of these funds fell in two distinct clusters, as shown in Fig.6.1D. From the relative positions and color labels of the Cluster View, E_1 recognized that the two clusters have contrasting investment behavior in terms of *Size* factor (**T1, T5, T7**). The top cluster, colored in Cyan, belonged to the upper 75 percentile group of the *Size* factor, which inferred that the funds mainly invested in large-cap stocks (Style A). On the other hand, the bottom cluster, colored in Blue, belonged to lower 25 percentile group which indicated that it invested in small-cap stock (Style B). Surprisingly, Fund A solely fell in the cluster of Style A, while the other funds fell into Style B.

After observing how Fund A, an awarding-winning fund, separated from the others, E_1 decided to investigate what made Fund A special. So, E_1 proceeded to the Detail View for more comprehensive analysis of investment styles (**T2**). E_1 observed that the glyphs of Fund A showed clear difference from other funds in general (**T1, T5**). In particular, E_1 identified that Fund A focused on the *Consumer Staple* and *Consumer Discretionary* sector while other funds focused in the *Information Technology* and *Material* sector. E_1 mentioned that it was interesting to find that Fund A did not invest in the most popular Information Technology sector in the market, as shown by the Benchmark glyphs, the first row in the Detail View (**T6**). Instead, it pursued its own investment strategy. E_1 then read the turnover bridges and noticed that Fund A consistently had low turnover rates, which were indicated by the thin bridges, showing that Fund A pursued low-frequency trading and other funds had relatively high turnover rates (**T1, T5**).

E_1 then headed to the Temporal View, where he discovers that Fund A had always outperformed the CSI300 Index (purple) and others selected funds also outperformed CSI300 Index most of the time. E_1 also found that while the market average (orange) always underperformed CSI300 in terms of cumulative return in the last two years (**T2**).

In order to review other aspects of the funds, E_1 used the List View to sort the funds using various features. After inspecting at different angles, E_1 found that Fund A had significantly high asset under management and high historical return.

By summarizing the findings from the system, E_1 was able to implicate that Fund A pursued its own distinguishing investment strategy. It employed a low-frequency trading style and favored large-cap stocks in its unique sector selections. It was interesting to reveal that the Fund A's actual investment style was contradictory to what its name suggested, a "medium-to-small-cap" fund. E_1 was impressed that *iFUNDit* could help him to shortlist an awarding-winning fund from thousands of funds and also assisted him in profiling its investment style, with intuitive and effective interactions.

Looking for similar funds. E_1 was then curious about whether there were funds similar to the investment style of Fund A (**T4**), and how they performed. He brushed the neighboring nodes of Fund A in the Cluster View. Three other funds were highlighted in the Fund View, as shown in Fig. 5. E_1 evaluated their investment styles in the Detail View. From the glyphs, E_1 discovered that these funds had similar investment styles. The orientation of the circular sectors showed that all of them invested heavily in the *Consumer Staples* sector. The shape of the center radar charts revealed that these funds mainly invested in large-cap stocks which is indicated by the sharp bulge in the direction of the *Size* factor axis.

Interestingly, E_1 said that he was well aware of the actual investment style of Fund B, because the fund manager of Fund B was once his colleague. E_1 confirmed that the actual investment style of the Fund B was indeed similar to that of Fund A. E_1 was impressed that *iFUNDit* could discover Fund B given Fund A. He mentioned that it normally requires a lot of domain knowledge and comprehensive investigation in order to identify similar funds with given criteria. He did not expect to be able to accomplish the task with simple interactions in *iFUNDit*.

E_1 noticed that all the funds similar to Fund A had good performances. Their cumulative return were all above 50% percentile of the market, and the downward risk (measured by maximum drawdown) were lower than 50% percentile of the market (**T4, T6**). The Temporal View showed that all funds consistently outperformed the CSI300 and the market average in the past two years. This finding gave E_1 a solid reference that could help him to adjust his own investment style. From the perspective of investors, the finding could give the investors a reference of what kind of funds to invest in.

7.2 Manager-level investigation

Fund manager is one the most critical elements that determine the investment style and performance of a fund. In many circumstances, investing in a fund is essentially betting on its fund managers. However, identifying the actual manager of a fund can be trickier than it sounds. It is not uncommon for a fund to have multiple managers, in which case the actual managers are difficult to identify. This is because fund institutes sometimes put their famous fund managers to the manager list of many funds, especially newly launched funds, in order to promote those funds and attract investors. However, these famous managers may not actually manage the fund and hence not affect the investment style. In practice, in order to identify whether a manager actually manages a fund, analysts

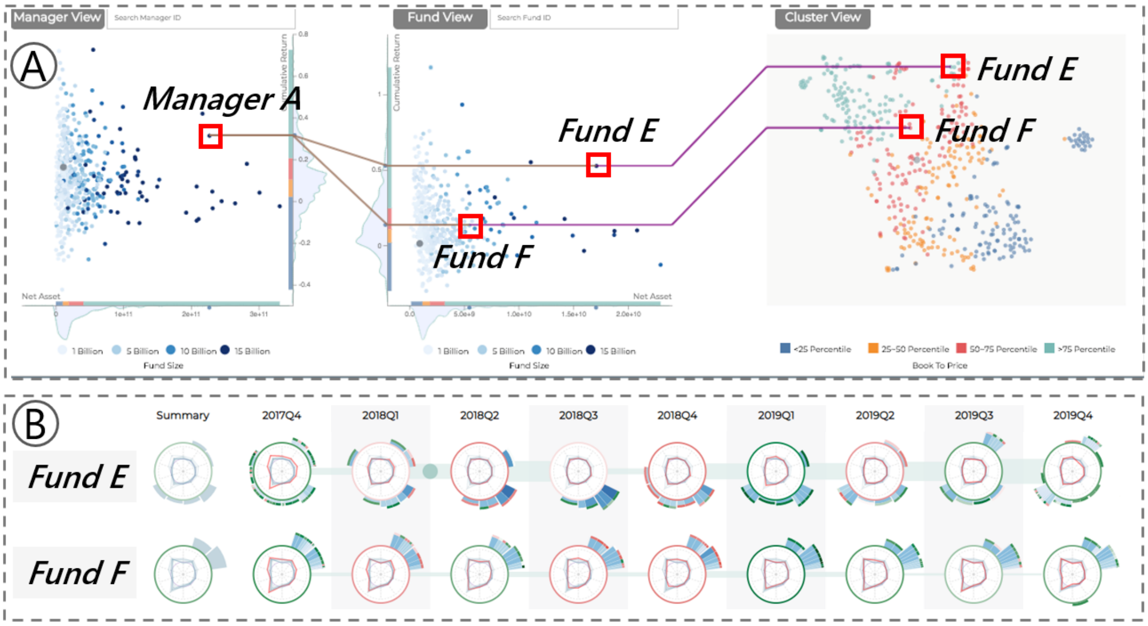

Figure 7.2: A fund manager was on the manager lists of two funds with drastically different investment styles. The fund manager was a big ticket who has a high return and manage high assets. It was suspected that the manager was not the actual manager of both funds.

gather information from various resources such as the manager's talks, news, and conduct interviews with the manager. The product managers E_2, E_3, and E_4, would like to use *iFUNDit* to look for clues about the actual manager of a fund (T6), in order to provide guidance for analysts' investigation.

E_2, E_3 and E_4 investigated on fund managers who manage funds with high net asset (**T7**). These managers are the famous ones who are assigned to the manager lists of many funds in order to attract investors. Therefore, E_4 set the "net asset" for the X-axis in the Manager View and Fund View, and "cumulative return" for the Y-axis (**T8**). By this setting, the managers with large net assets were plotted towards the right side of the Manager View.

E_4 brushed a manager on the right side of Manger View, whose cumulative return ranked above 75 percentile indicated by the top percentile ribbon on the Y-axis, to observe the funds under his names (**T7**) . As shown in the Fig.7.2, the Manager A had two funds under his name. However, the two funds had very different investment styles, indicated by their distinct positions in the Cluster View (**T4, T5**). This was confirmed in the Detail View. Fund E diversified its stocks in many sectors, which was indicated by the many non-empty circular sectors in the glyphs. In the contrary, Fund F almost invested solely in two sectors, which were not the focusing sectors of Fund E. In addition, Fund E had high

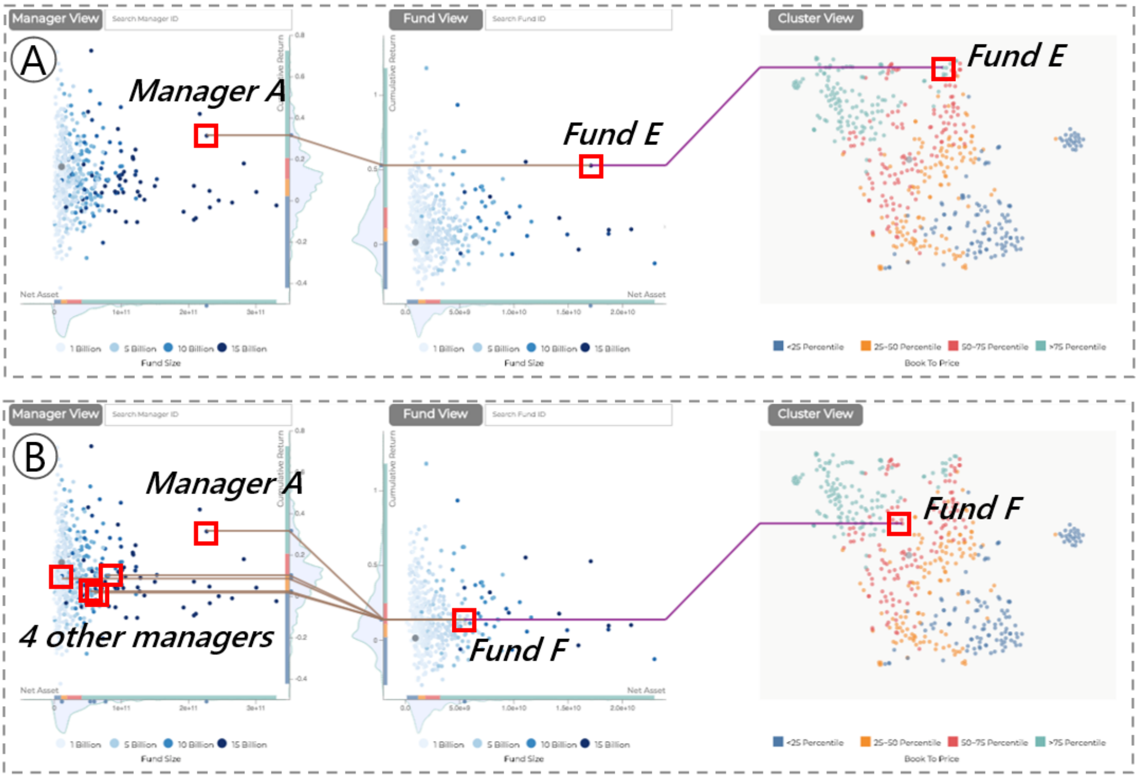

Figure 7.3: Detailed analysis revealed that relations between Manager A and Fund E/F. Manager A was the sole manager of Fund E. Therefore the investment style of Fund E represented that of Manager A. On the other hand, Manager A was one of the five managers of Fund B. The findings implied that Manager A was not the actual manager of Fund F.

turnover rate (indicated by the thick bridge between glyphs) and the investment style changed from time to time (indicated by the varying glyph shapes) (**T1, T2**); while Fund F had low turnover rate and maintained a consistent investment style during the period. These observations showed that Manager A was managing two funds with drastically different investment styles.

According to the senior fund manager E_1, it is unlikely that a fund manager adopts completely different investment styles at the same time. Therefore, E_4 suspected that Manager A was not the actual manager of both funds. To verify his hypothesis, he brushed the two funds to check their managers. The results are shown in Fig.7.3. It showed that Fund E had only Manager A as its sole manager, which infers that its investment style was the actual investment style of Manager A. On contrary, Fund F had 5 managers. Considering the investment style of Fund F which is drastically different from Fund E, it was very likely that Manager A was not the actual manager of the Fund F. It could be also observed from the Manager View and the Fund View that, the Manager A and his fund had greater return than the other managers and Fund F (**T6**). Therefore, E_2, E_3 and E_4 highly suspected that the fund institution assigned Manager A to Fund F in order to promote the fund. E_2, E_3 and E_4 claimed that these findings could give solid guidance for the analysts in further investigation on the manager and the fund.

7.3 Domain Expert Interview

We conducted interviews with experts E_1-E_4, collected the their feedback and summarized it as follows.

Usability All experts appreciated the capability of *iFUNDit*. E_1 and E_2 suggested that the system would greatly boost their working efficiency. Currently, an investment style is characterized by many charts and tables, as well as qualitative descriptions, which are hard to summarize or compare. With our system, E_1 and E_2 stated that they could investigate fund investment styles more efficiently. E_1 commented that, the system could help him to quickly identify any changes in the investment styles. E_3 and E_4 stated that our system could help him to show investors the differences between various funds and explain the investment styles more clearly.

Visual Design Experts agreed that the visual design shows the details of an investment style clearly. They could inspect an investment style from different perspectives conveniently, and compare different styles efficiently with the help of the glyph design. However, they also mentioned the glyph design took them the most time to understand in the first place. They appreciated that our system combines an innovative glyph design and traditional visualization techniques together, and found a balance in between. They praised the idea of separating critical performance attributes from investment style factors, and visualizing these attributes with familiar visualization techniques such as scatter plot, line charts and tables, which were intuitive and effective. They appreciated how the system emphasizes reference and benchmark in all views. They particularly like the percentile ribbon and density plot in the Manager View and Fund View. E_1 mentioned that, "*In industry, fund managers' KPI were measured according to their rankings in the market. The system efficiently shows the relative position of a fund or a manager in the market, which is crucial for our evaluation.*"

Interaction Experts commented that the system was intuitive to use. E_1 and E_4

commented that, "*The system is intuitive to use. This is critical when you are going to commercialize the system.*" E_1 and E_2 especially like the selectable axes in the Manager View and Fund View, and selectable color encoding in the Cluster View, which enables them to explore the distribution of factors in the market.

Improvement Experts mentioned that the glyph design took them the most time to learn. The glyph could be complicated for investors who do not have domain knowledge. They stated that it was difficult to memorize the labels of all axes. Therefore, it would be more convenient if the system could enlarge a glyph when users have mouse hovered on it. They also would like to see the system integrates textual information such as news and fund reports, in order to evaluate funds more conveniently without referring to external systems.

Chapter 8

Discussion

The case studies and domain expert interviews demonstrate the effectiveness of the system. Domain experts confirm that the system is a successful attempt to profile fund investment style in an effective way, and has a good potential to create impact in the industry. Although we used China mutual fund data for the demonstration, the system and the proposed analytics pipeline are applicable to mutual funds in the global market, as well as to private funds if the data is accessible. In addition to Barra models, the system can also be applied to other evaluation models with similar number of factors. The system's effectiveness and generalizability have been recognized by domain experts. We have been working with RiceQuant to deploy our system on their production platform, which serves overall 300 mutual fund and private fund institutions, as well as over 120,000 individual users. We are going to collect more feedback from users and improve our system accordingly. Some of the possible improvements are discussed below.

One major issue of the system is the scalability of the Fund/Manager View. Due to the nature of the fund market, the performance attributes of many funds/managers dwell near the mean of the overall distribution. This causes visual clutter in the two views, which can make it difficult to brush a certain fund/manager near the center of the distribution. In order to alleviate this issue, we use color to encode the circle size in the two views, and implement a filter and direct query function to help users to select funds/managers more easily. Nevertheless, it would be even more convenient for users to brush more precisely by allowing them to zoom in the Fund/Manager View.

The other limitation is the that the system does not explicitly label a fund with a certain investment style. This is because the investment style of a fund can be measured from different perspectives and interpreted differently. For example, a fund can be labeled as "large-cap" from the perspective of stock caps, or labeled as "value" from the perspective of the of stock value orientation. The objective of the study is to profile an investment style in detail, instead of labeling it qualitatively. Nevertheless, it would be more convenient to have a certain labels for investors to understand.

Chapter 9

Conclusion

In this work, we propose a visual analytics system, *iFUNDit*, to address the two major challenges of fund investment style analysis: an intuitive presentation of fund investment style and an effective approach to evaluate it. We categorize the funds' attributes into two groups, namely, the performance attributes, and the investment style factors; and present them cohesively to achieve a streamlined analysis of investment style at both fund and manager level. The system emphasizes on showing references and benchmarks, so that the analysis can be conducted in a comparative context, which is critical for finance data analysis. By using the China mutual fund data set, we conducted a set of comprehensive case studies with domain experts to validate the usefulness and effectiveness of the system.

In the future, in addition to addressing the aforementioned issues, we plan to conduct research on characterizing the personalities of investors and fund managers to better guide fund investment and management in behavioral finance study [56, 57]. Specifically, we can use the investment style details that are obtained in the system, to profile the personalities of fund managers and label them accordingly. We would also like to integrate a Know-Your-Customer process in which the investment preference of an investor is profiled. With the profiling of investors, funds and managers, it is feasible to develop a recommendation system to match investors with funds according to their personality traits.

References

[1] “The International Investment Fund Association,” https://www.iifa.ca/.

[2] E. F. Fama and K. R. French, “Luck versus skill in the cross-section of mutual fund returns,” *The journal of finance*, vol. 65, no. 5, pp. 1915–1947, 2010.

[3] R. C. Grinold and R. N. Kahn, “Active portfolio management,” pp. 73–185, 2000.

[4] “MSCI,” https://www.msci.com.

[5] “Global industry classification standard,” https://www.spglobal.com/marketintelligence/en/documents/112727-gics-mapbook_2018_v3_letter_digitalspreads.pdf.

[6] “Morningstar StyleBox,” https://www.morningstar.com/InvGlossary/morningstar_style_box.aspx.

[7] W. F. Sharpe, “Capital asset prices: A theory of market equilibrium under conditions of risk,” *The journal of finance*, vol. 19, no. 3, pp. 425–442, 1964.

[8] ——, “The capital asset pricing model: a “multi-beta” interpretation,” in *Financial Dec Making Under Uncertainty*. Elsevier, 1977, pp. 127–135.

[9] ——, “Determining a fund’s effective asset mix, investment management review,” *December*, vol. 59, p. 69, 1988.

[10] E. F. Fama and K. R. French, “The cross-section of expected stock returns,” *the Journal of Finance*, vol. 47, no. 2, pp. 427–465, 1992.

[11] J. L. Davis, “Mutual fund performance and manager style,” *Financial Analysts Journal*, vol. 57, no. 1, pp. 19–27, 2001.

[12] M. M. Carhart, “On persistence in mutual fund performance,” *The Journal of finance*, vol. 52, no. 1, pp. 57–82, 1997.

[13] J. Bender, R. Briand, D. Melas, and R. Subramanian, “Foundations of factor investing,” 2013.

[14] “The barra china equity model (cne5),” https://www.msci.com/www/research-paper/the-barra-china-equity-model/014459336.

[15] M. Kacperczyk and A. Seru, “Fund manager use of public information: New evidence on managerial skills,” *The Journal of Finance*, vol. 62, no. 2, pp. 485–528, 2007.

[16] B. Ding and R. Wermers, “Mutual fund performance and governance structure: The role of portfolio managers and boards of directors,” *Available at SSRN 2207229*, 2012.

[17] M. Kacperczyk, S. V. Nieuwerburgh, and L. Veldkamp, “Time-varying fund manager skill,” *The Journal of Finance*, vol. 69, no. 4, pp. 1455–1484, 2014.

[18] K. M. Cremers and A. Petajisto, “How active is your fund manager? a new measure that predicts performance,” *The review of financial studies*, vol. 22, no. 9, pp. 3329–3365, 2009.

[19] R. D. Peng, “A method for visualizing multivariate time series data,” 2008.

[20] J. Wang, L. Gou, H.-W. Shen, and H. Yang, “Dqnviz: A visual analytics approach to understand deep q-networks,” *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 288–298, 2018.

[21] L. Stopar, P. Skraba, M. Grobelnik, and D. Mladenic, “Streamstory: exploring multivariate time series on multiple scales,” *IEEE transactions on visualization and computer graphics*, vol. 25, no. 4, pp. 1788–1802, 2018.

[22] H. S. Garcia Caballero, M. A. Westenberg, B. Gebre, and J. J. van Wijk, “V-awake: A visual analytics approach for correcting sleep predictions from deep learning models,” in *Computer Graphics Forum*, vol. 38, no. 3. Wiley Online Library, 2019, pp. 1–12.

[23] J. Sansen, G. Richer, T. Jourde, F. Lalanne, D. Auber, and R. Bourqui, “Visual exploration of large multidimensional data using parallel coordinates on big data

infrastructure," in *Informatics*, vol. 4, no. 3. Multidisciplinary Digital Publishing Institute, 2017, p. 21.

[24] J. Wang, X. Liu, H.-W. Shen, and G. Lin, "Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 81–90, 2016.

[25] A. Sawant, "Stockviz: Analyzing the trend of stocks in major auto, oil, consumer, and technology companies." 01 2009, pp. 278–284.

[26] M. Schaefer, F. Wanner, R. Kahl, L. Zhang, T. Schreck, and D. Keim, "A novel explorative visualization tool for financial time series data analysis," in *VAW2 011: The Third International UKVAC Workshop on Visual Analytics*, 2011.

[27] H. Ziegler, M. Jenny, T. Gruse, and D. A. Keim, "Visual market sector analysis for financial time series data," in *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2010, pp. 83–90.

[28] X. Yue, J. Bai, Q. Liu, Y. Tang, A. Puri, K. Li, and H. Qu, "sportfolio: Stratified visual analysis of stock portfolios," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 601–610, 2019.

[29] J. Alsakran, Y. Zhao, and X. Zhao, "Visual analysis of mutual fund performance," in *2009 13th International Conference Information Visualisation*. IEEE, 2009, pp. 252–259.

[30] H. Ziegler, T. Nietzschmann, and D. A. Keim, "Visual analytics on the financial market: Pixel-based analysis and comparison of long-term investments," in *2008 12th International Conference Information Visualisation*. IEEE, 2008, pp. 287–295.

[31] ——, "Relevance driven visualization of financial performance measures," in *EuroVis2007*, 2007, pp. 19–26.

[32] J. Alsakran, Y. Zhao, and X. Zhao, "Tile-based parallel coordinates and its application in financial visualization," in *Visualization and Data Analysis 2010*, vol. 7530. International Society for Optics and Photonics, 2010, p. 753003.

[33] W.-A. Jungmeister and D. Turo, "Adapting treemaps to stock portfolio visualization," Tech. Rep., 1992.

[34] C. Csallner, M. Handte, O. Lehmann, and J. Stasko, "Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps," in *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)*. IEEE, 2003, pp. 203–208.

[35] F. Xiong, E. Prakash, and K. Ho, "Er modeling and visualization of large mutual fund data," *Journal of Visualization*, vol. 5, no. 2, pp. 197–204, 2002.

[36] S. T. Lei and K. Zhang, "A visual analytics system for financial time-series data," in *Proceedings of the 3rd International Symposium on Visual Information Communication*. ACM, 2010, p. 20.

[37] T. Dwyer and P. Eades, "Visualising a fund manager flow graph with columns and worms," in *Proceedings Sixth International Conference on Information Visualisation*. IEEE, 2002, pp. 147–152.

[38] T. Dwyer, "A scalable method for visualising changes in portfolio data," in *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24*. Australian Computer Society, Inc., 2003, pp. 17–25.

[39] T. Dwyer and D. R. Gallagher, "Visualising changes in fund manager holdings in two and a half-dimensions," *Information Visualization*, vol. 3, no. 4, pp. 227–244, 2004.

[40] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visualizing time-oriented data—a systematic view," *Computers & Graphics*, vol. 31, no. 3, pp. 401–409, 2007.

[41] S. Ko, I. Cho, S. Afzal, C. Yau, J. Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and D. S. Ebert, "A survey on visual analysis approaches for financial data," in *Computer Graphics Forum*, vol. 35, no. 3, 2016, pp. 599–617.

[42] M. Dumas, M. J. McGuffin, and V. L. Lemieux, "Financevis.net-a visual survey of financial data visualizations," in *Poster Abstracts of IEEE Conference on Visualization*, vol. 2, 2014.

[43] H. Zhou, P. Xu, and H. Qu, "Visualization of bipartite relations between graphs and sets," *Journal of Visualization*, vol. 18, no. 2, pp. 159–172, 2015.

[44] P. Xu, N. Cao, H. Qu, and J. Stasko, "Interactive visual co-cluster analysis of bipartite graphs," in *2016 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 2016, pp. 32–39.

[45] G. Y.-Y. Chan, P. Xu, Z. Dai, and L. Ren, "V i b r: Visualizing bipartite relations at scale with the minimum description length principle," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 321–330, 2018.

[46] "Csi 300 index," http://www.csindex.com.cn/uploads/indices/detail/files/en/145_000300_Fact_Sheet_en.pdf.

[47] "Ricequant," https://www.ricequant.com.

[48] M. Glantz and R. L. Kissell, *Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era*. Academic Press, 2013.

[49] A. C. Rencher and W. F. Christensen, "Chapter 10, multivariate regression–section 10.1, introduction," *Methods of multivariate analysis, Wiley Series in Probability and Statistics*, vol. 709, p. 19, 2012.

[50] "Definition of maximum drawdown," https://www.investopedia.com/terms/m/maximum-drawdown-mdd.asp.

[51] "US Securities and Exchange Commission," https://www.sec.gov/.

[52] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[53] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[54] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-sne effectively," *Distill*, 2016. [Online]. Available: http://distill.pub/2016/misread-tsne

[55] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *The Craft of Information Visualization*. Elsevier, 2003, pp. 364–371.

[56] D. De Bortoli, N. da Costa Jr, M. Goulart, and J. Campara, "Personality traits and investor profile analysis: A behavioral finance study," *PloS one*, vol. 14, no. 3, 2019.

[57] L. J. Belcher, "Prior perceptions, personality characteristics and portfolio preferences among fund managers: an experimental analysis," *Journal of Behavioral Finance*, vol. 11, no. 4, pp. 239–248, 2010.

List of Publications

1. Aoyu Wu, **Bon Kyung Ku**, Furui Cheng, Xinhuan Shu, Abishek Puri, Yifang Wang and Huamin Qu, “Pulse: Toward a Smart Campus by Communicating Real-time Wi-Fi Access Data”, Workshop on Visualization for Communication, the IEEE Visualization Conference, Berlin, Germany, 2018.
2. Abishek Puri, **Bon Kyung Ku**, Yong Wang and Huamin Qu, “RankBooster: Visual Analysis of Ranking Predictions”, EuroVis2020 - Short Papers, Norrkoping, Sweden, 2020.