RUTGERS

Rutgers Business School
Newark and New Brunswick

If You Can't Measure It, You Can't Manage It: Cybersecurity Risk Measurement Through Textual Imagification

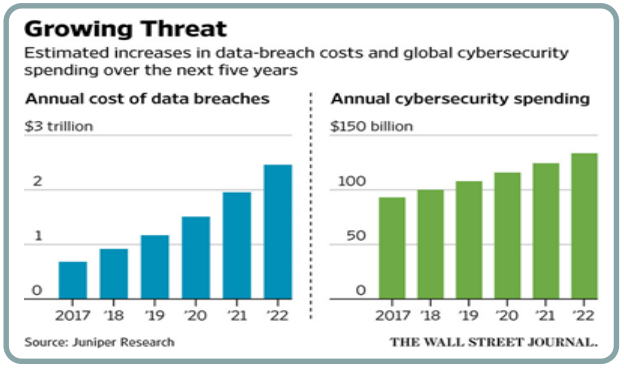
Arion Cheong

Dr. Soo Hyun Cho

Dr. Won Gyun No

Dr. Miklos Vasarhelyi

RESEARCH OBJECTIVES



- ✓ Identify types of cybersecurity risks.
- ✓ Compare a firm's cybersecurity risks with its peers.
- ✓ Disentangle firm-specific and industry general cybersecurity risks.

Overall measure

RESEARCH OBJECTIVES

- Need to find comparable industry peers.
- Need to classify types of cybersecurity risks.
- Need to calculate an overall cybersecurity risks.

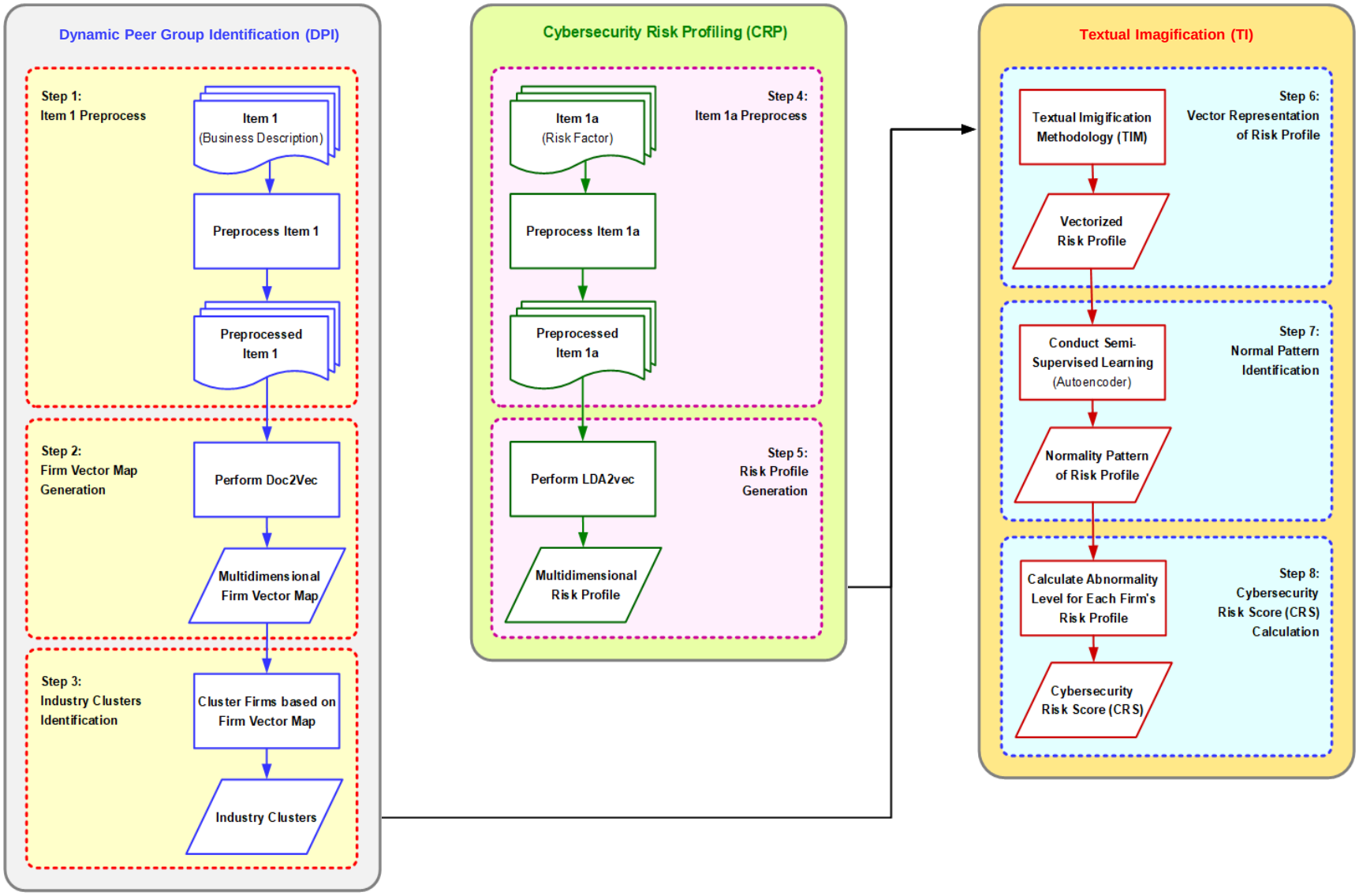
- Limitation of existing industry classification such as SIC
- Cybersecurity risks - Textual data
- Multicollinearity - Correlations among various types of cybersecurity risks

RESEARCH OBJECTIVES

- Limitation of existing industry classification such as SIC → Dynamic Peer Group Identification (DPI)
- Cybersecurity risks - Textual data → Cybersecurity Risk Profiling (CRP)
- Multicollinearity - Correlations among various types of cybersecurity risks → Textual Imigification(TI)

To introduce a new quantitative method for gauging cybersecurity risks in the hope of encouraging new streams of future research that might utilize our proposed measures.

GENERAL FRAMEWORK



RESEARCH OBJECTIVES

- Limitation of existing industry classification such as SIC → Dynamic Peer Group Identification (DPI)
- Cybersecurity risks - Textual data → Cybersecurity Risk Profiling (CRP)
- Multicollinearity - Correlations among various types of cybersecurity risks → Textual Imigification Methodology (TIM)

To introduce a new quantitative method for gauging cybersecurity risks in the hope of encouraging new streams of future research that might utilize our proposed measures.

DYNAMIC PEER GROUP IDENTIFICATION (DPI)

- **Limitation of Existing Industry Classification such as SIC, NAICS, and Fama-French.**
  - **May not properly describe firms' businesses: Created by statistics agencies for information aggregation.**
  - **Old: Established in 1930s.**
  - **Business operation changes during the business cycle.**
    ex) Apple – Telecom Equipment, Electronic Computers

Current Approach

Google

ITEM 1. BUSINESS

Overview

As our founders Larry and Sergey wrote in the original founders' letter, "Google is not a conventional company. We do not intend to become one." That unconventional spirit has been a driving force throughout our history -- inspiring us to do things like rethink the mobile device ecosystem with Android and map the world with Google Maps. As part of that, our founders also explained that you could expect us to make "smaller bets in areas that might seem very speculative or even strange when compared to our current businesses." From the start, the company has always strived to do more, and to do important and meaningful things with the resources we have.

Alphabet is a collection of businesses -- the largest of which is Google. It also includes businesses that are generally pretty far afield of our main internet products in areas such as self-driving cars, life sciences, internet access and TV services. We report all non-Google businesses collectively as Other Bets. Our Alphabet structure is about helping each of our businesses prosper through strong leaders and independence.

Text Mining using Specific Keywords

Toy Dog

=

Dog Toy

The methodology do not understand the sentence contextually (i.e., order of the words do not matter).

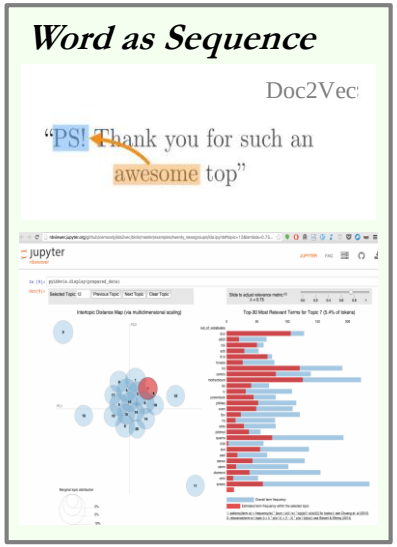
DYNAMIC PEER GROUP IDENTIFICATION (DPI)

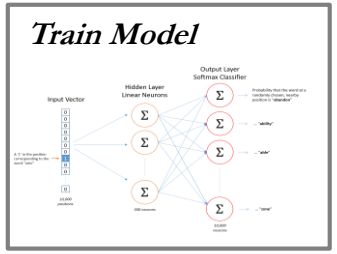
- **Limitation of Existing Industry Classification such as SIC, NAICS, and Fama-French.**
  - **May not properly describe firms' businesses: Created by statistics agencies for information aggregation.**
  - **Old: Established in 1930s.**
  - **Business operation changes during the business cycle.**
    ex) Apple – Telecom Equipment, Electronic Computers

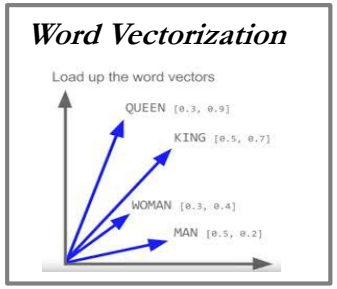
Current Approach

New Approach

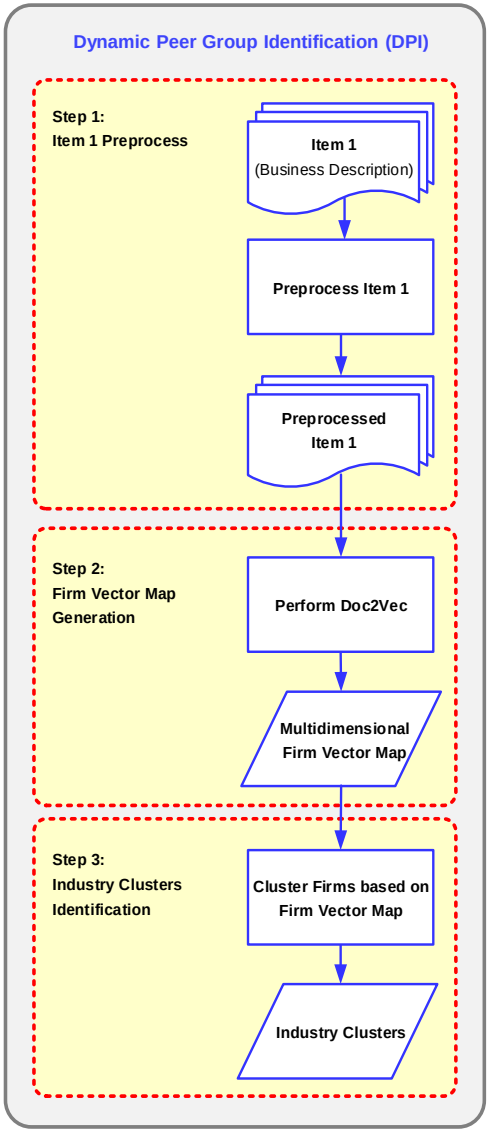
Distributed Representation Methods with Word Embedding

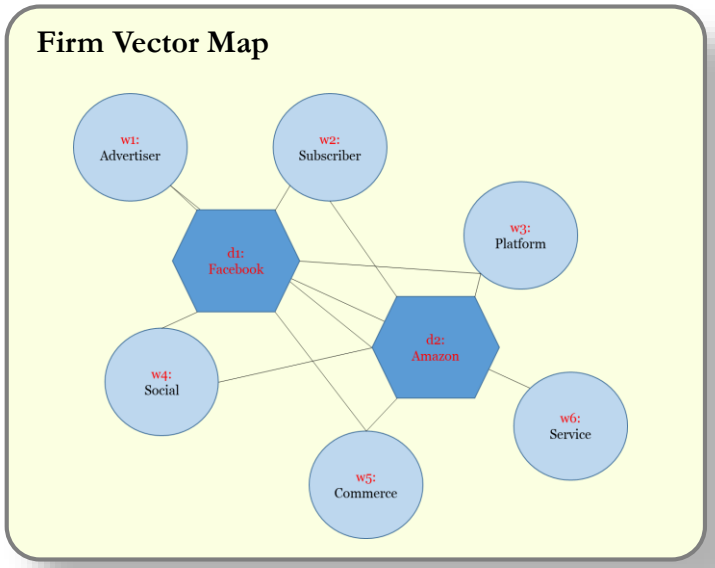


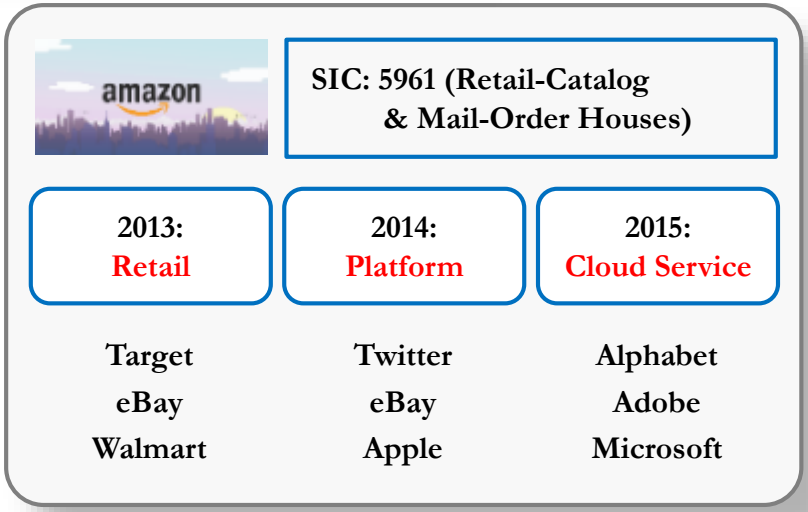




DYNAMIC PEER GROUP IDENTIFICATION (DPI)







RESEARCH OBJECTIVES

- Limitation of existing industry classification such as SIC → Dynamic Industry Classification (DIC)
- Cybersecurity risks - Textual data → Cybersecurity Risk Profiling (CRP)
- Multicollinearity - Correlations among various types of cybersecurity risks → Textual Imigification Methodology (TIM)

To introduce a new quantitative method for gauging cybersecurity risks in the hope of encouraging new streams of future research that might utilize our proposed measures.

CYBERSECURITY RISK PROFILING (CRP)

Firm Disclosures
(e.g., 10-K)

ITEM 1A. RISK FACTORS

Our operations and financial results are subject to various risks and uncertainties, including but not limited to those described below, which could adversely affect our business, financial condition, results of operations, cash flows, and the trading price of our common and capital stock.

Risks Related to Our Businesses and Industries

We face intense competition. If we do not continue to innovate and provide products and services that are useful to users, we may not remain competitive, and our revenues and operating results could be adversely affected.

Our businesses are rapidly evolving, intensely competitive, and subject to changing technologies, shifting user needs, and frequent introductions of new products and services. Competing successfully depends heavily on our ability to accurately anticipate technology developments and deliver innovative, relevant and useful products,

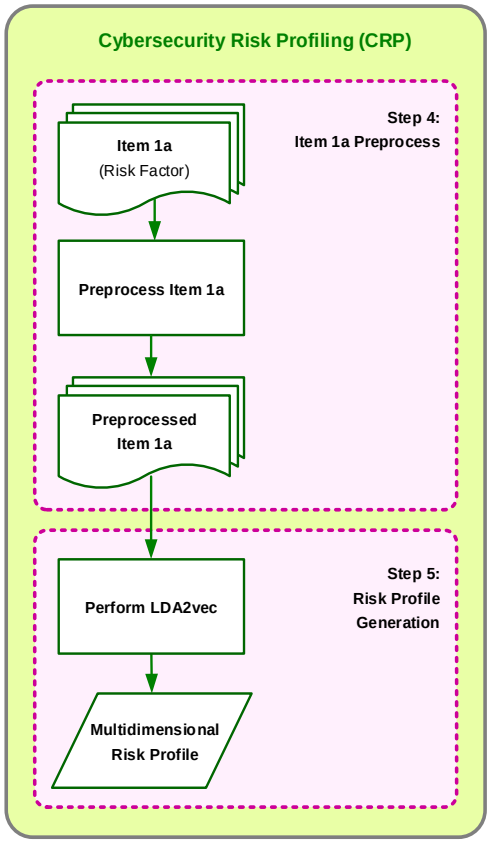
Data privacy and security concerns relating to our technology and our practices could damage our reputation and deter current and potential users or customers from using our products and services. Bugs or defects in our products and services have occurred and may occur in the future, or our security measures could be breached, resulting in the improper use and/or disclosure of user data, and our services and systems are subject to attacks that could degrade or deny the ability of users and customers to access, or rely on information received about, our products and services. As a consequence, our products and services may be perceived as being insecure, users and customers may curtail or stop using our products and services, and we may incur significant legal, reputational, and financial exposure.

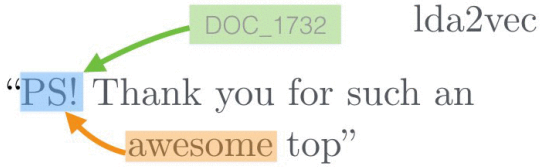
From time to time, concerns are expressed about whether our products, services, or processes compromise the privacy of users, customers, and others. Concerns about our practices with regard to the collection, use, disclosure, or security of personal information or other data privacy related matters, even if unfounded, could damage our reputation and adversely affect our operating results. Our policies and practices may change over time as users' and customers' expectations regarding privacy and their data changes.

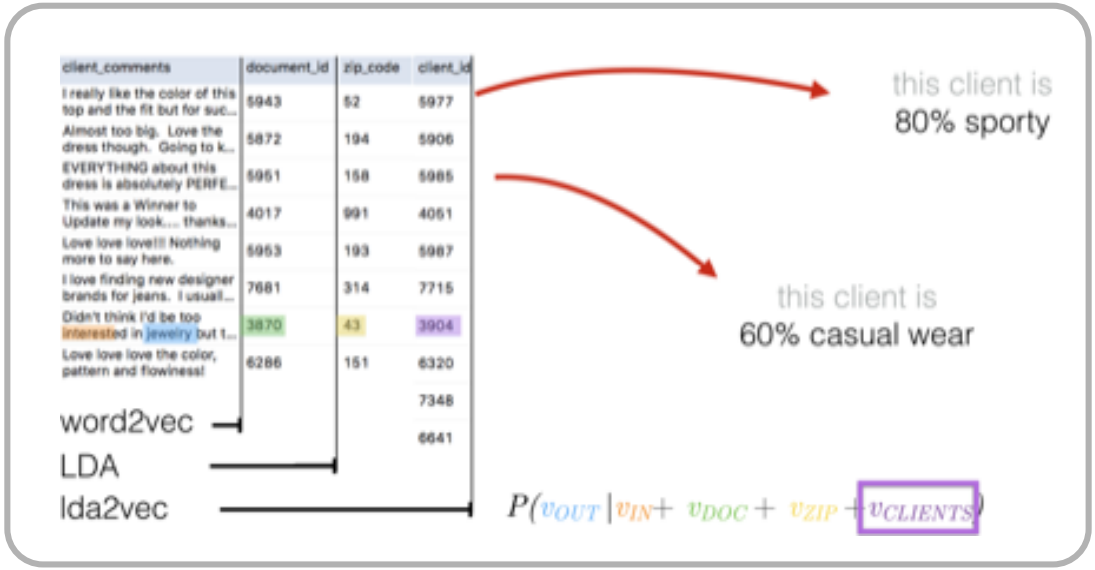
Our products and services involve the storage and transmission of users' and customers' proprietary information, and bugs, theft, misuse, defects, vulnerabilities in our products and services, and security breaches expose us to a risk of loss of this information, improper use and disclosure of such information, litigation, and other potential liability. Systems failures, compromises of our security, failure to abide by our privacy policies, inadvertent disclosure that results in the release of our users' data, or in our or our users' inability to access such data, could result in government investigations and other liability, legislation or regulation, seriously harm our reputation and brand and, therefore, our business, and impair our ability to attract and retain users. We expect to continue to expend significant resources to maintain state-of-the-art security protections that shield against bugs, theft, misuse or security vulnerabilities or breaches.

- Publicly available document from the qualified source
- Cover all U.S public firms.
- Updated annually allowing us to capture the changes

CYBERSECURITY RISK PROFILING (CRP)

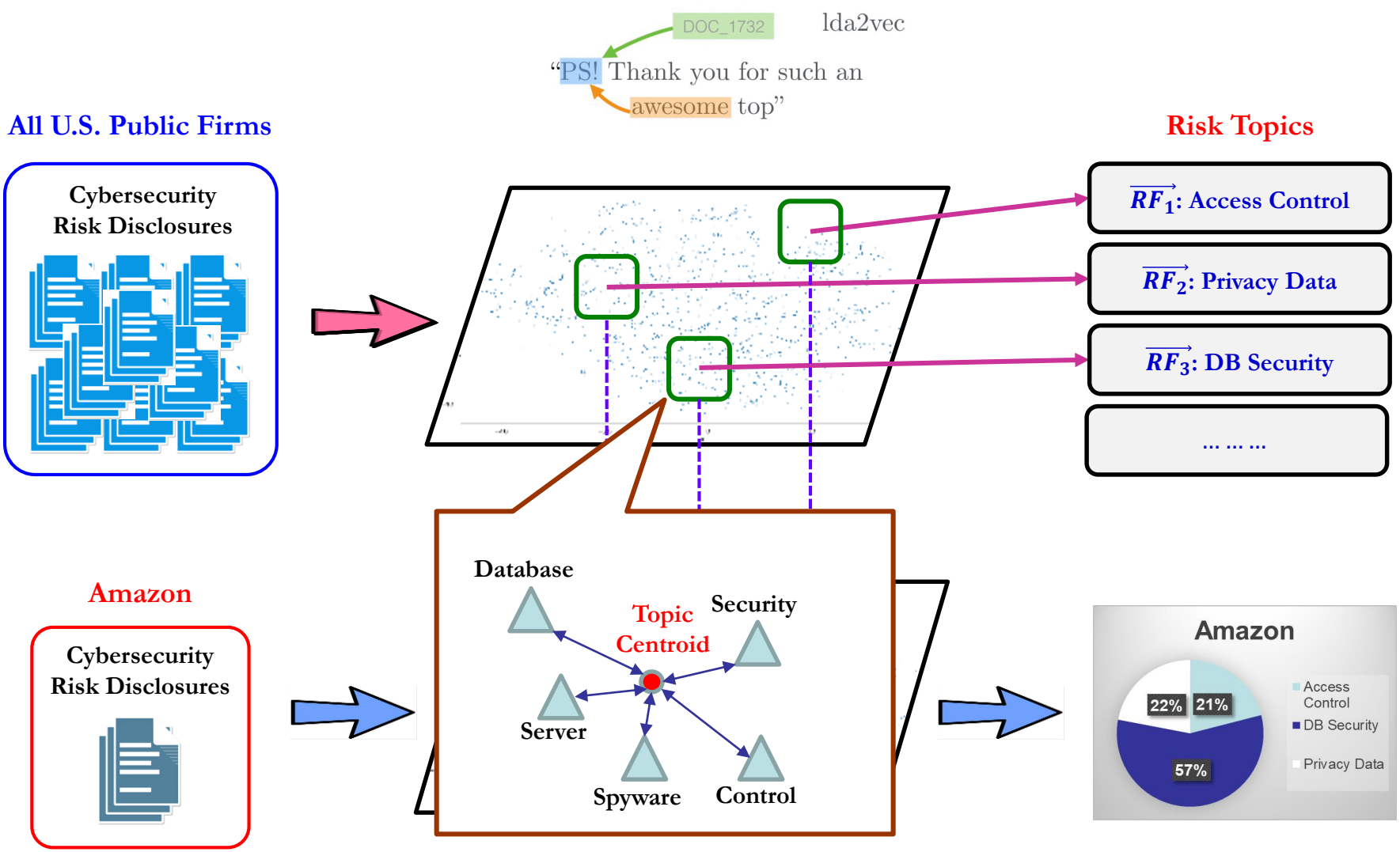






LDA2Vec globally finds topics for documents based on multiple attributes (e.g., stock price, location, size etc.).

CYBERSECURITY RISK PROFILING (CRP)



RESEARCH OBJECTIVES

- Limitation of existing industry classification such as SIC → **Dynamic Peer Group Identification (DPI)**
- Cybersecurity risks - Textual data → **Cybersecurity Risk Profiling (CRP)**
- Multicollinearity - Correlations among various types of cybersecurity risks → **Textual Imigification (TI)**

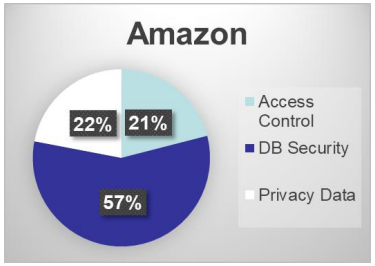
To introduce a new quantitative method for gauging cybersecurity risks in the hope of encouraging new streams of future research that might utilize our proposed measures.

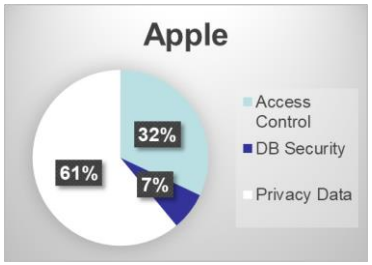
TEXTUAL IMIGIFICATION (TI)

- **Multicollinearity Issue**

Regression Analysis ?

Correlations among various cybersecurity risk topics





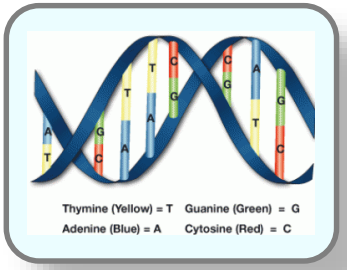
Access Control	$\overrightarrow{\lambda_1}$: 21%	<	$\overrightarrow{\lambda_1}$: 32%
Privacy Data	$\overrightarrow{\lambda_2}$: 22%	>	$\overrightarrow{\lambda_2}$: 7%
DB Security	$\overrightarrow{\lambda_3}$: 57%	<	$\overrightarrow{\lambda_3}$: 61%

Overall Cybersecurity Risk Score	??? out of 1000	?	??? out of 1000

TEXTUAL IMIGIFICATION (TI)

Entity (Human)

Subject of Analysis (DNA Sequences)



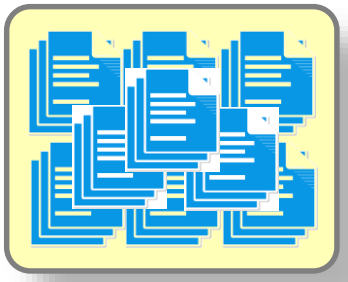
Imagify DNA Sequences

Probability of Cancer

FDA-Google DeepVariant

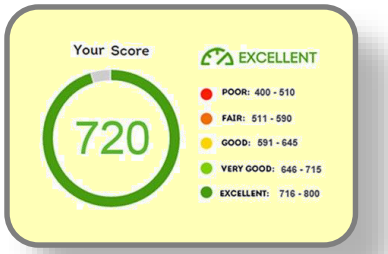
Entity (Company)

Subject of Analysis (Cybersecurity Risk Disclosures)

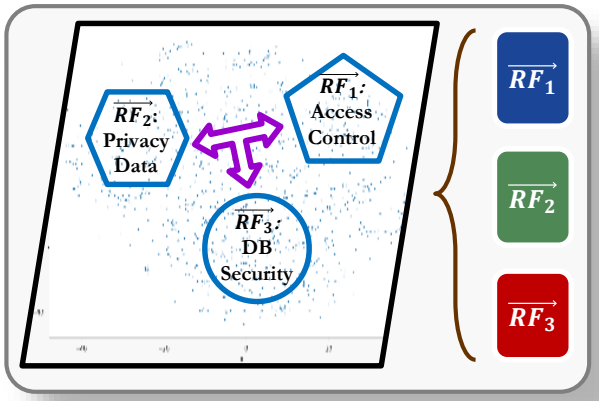


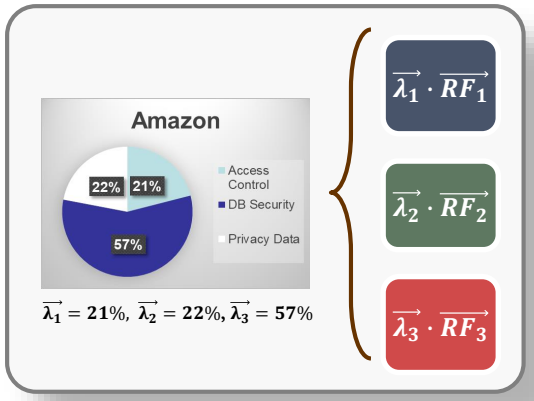
Imagify Cybersecurity Risk Topics

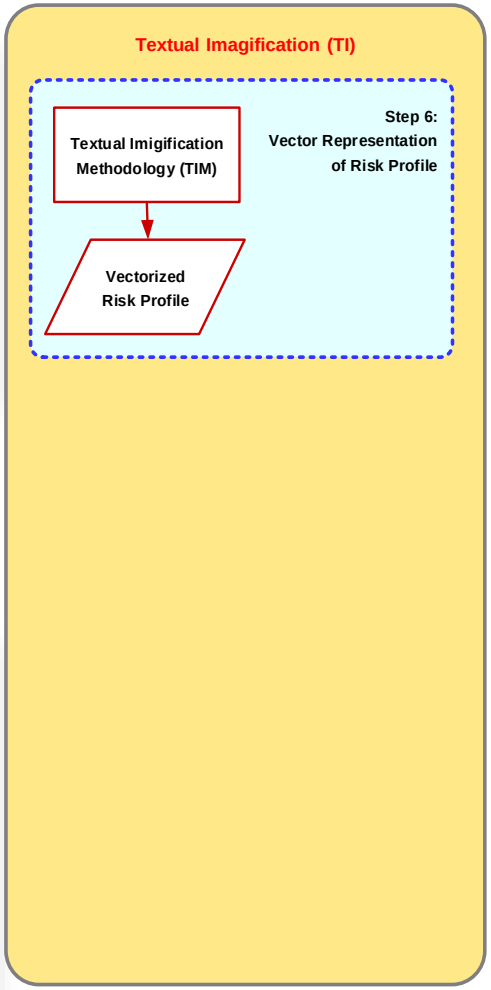
Cybersecurity Risk Metric



TEXTUAL IMIGIFICATION (TI)

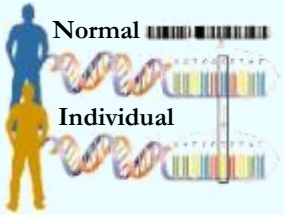






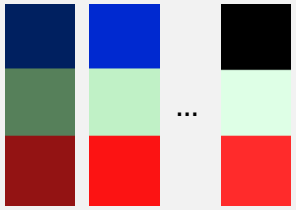
TEXTUAL IMIGIFICATION (TI)

Imagify DNA Sequences

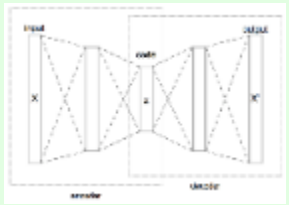


Probability of Cancer

Images of Non-breached Firms



Autoencoder (Semi-supervised)



Normal Pattern of Cybersecurity Risk Image

Cybersecurity Risk Metric

87.9%

amazon

Textual Imagification (TI)

Textual Imigification Methodology (TIM)

Step 6: Vector Representation of Risk Profile

Vectorized Risk Profile

Conduct Semi-Supervised Learning (Autoencoder)

Step 7: Normal Pattern Identification

Normality Pattern of Risk Profile

Calculate Abnormality Level for Each Firm's Risk Profile

Step 8: Cybersecurity Risk Score (CRS) Calculation

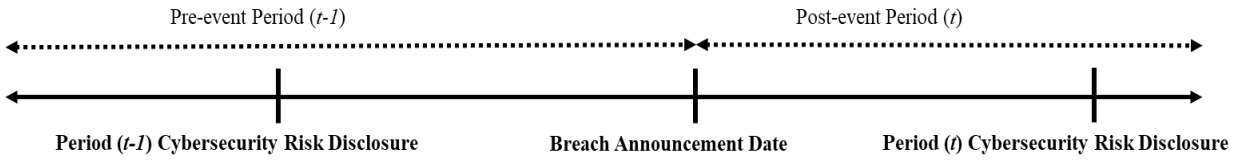
Cybersecurity Risk Score (CRS)

APPLICATION

❖ Market Reaction

$$CAR_{i,t} = \beta_0 + \beta_1 CRS_{i,t-1} + \beta_2 Breach_{i,t} + \beta_3 CRS_{i,t-1} \times Breach_{i,t}$$

$$+\beta_4 PastBreach_{i,t} + \beta_5 TA_{i,t} + \beta_6 Sales_{i,t} + \sum \beta_i Industry_{i,t} + \epsilon_{i,t}$$

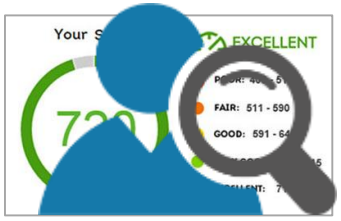


Variable	*Coefficients*	*t-statistics*	*p*
Intercept	0.0238	1.726	0.086
CRS	-0.0594***	-2.454	0.015
Breach	-0.0366***	-3.032	0.003
*CRS * Breach*	0.0934***	2.639	0.009
Industry Effects	Included		
Adjusted R^2	0.021		
No. of Observations	206		

Conclusion

Firm Disclosures
(e.g., 10-K)

Doc2Vec
LDA2Vec
Autoencoder



COMMENTS & SUGGESTIONS

Platform QR Code

BUSINESS.RUTGERS.EDU/CYBERSECURITYRISK