



In the Digital World: Practice and Reflection of Chinese Young Auditors Driving Change with Technology

Applicant: Yunhui Liu

中华人民共和国
审计署

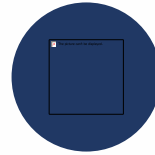
CONTENTS

- 1. Background**
- 2. Practical Cases**
- 3. Realistic Examination**
- 4. Pathways and Outlook**

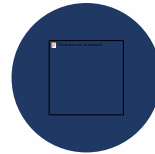
01

Background

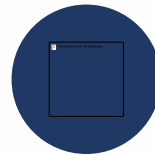
Background



Impacts and Requirements of the Digital World on Auditing



Advancement of the “Strengthening Auditing through Technology” Strategy



The Unique Role of Young Auditors

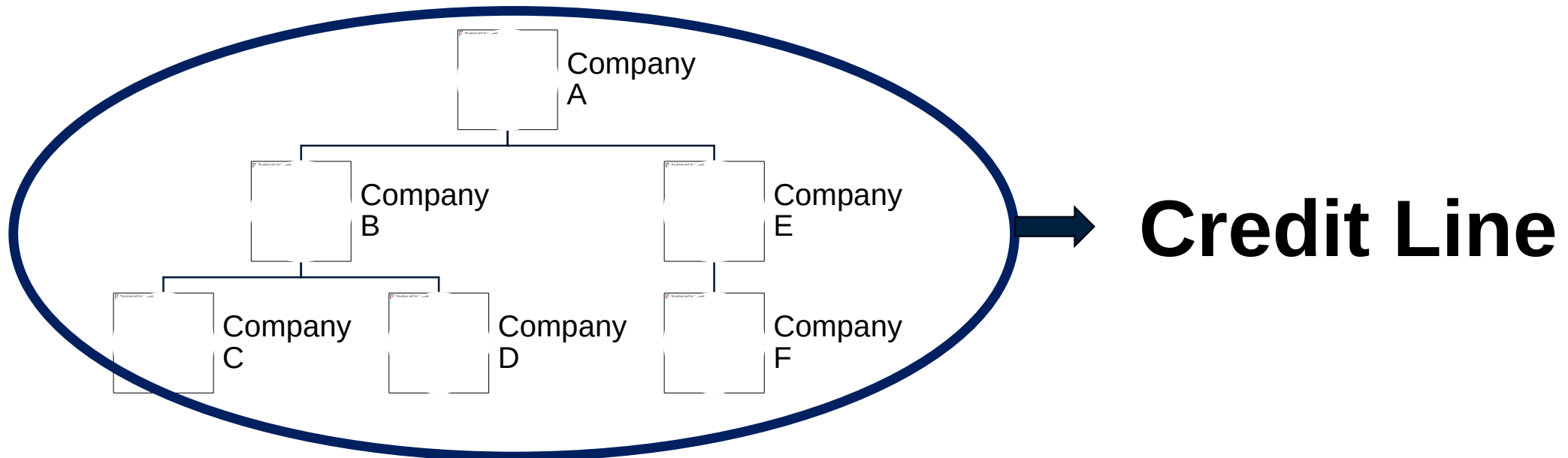
02

Practical Cases

Practical Cases

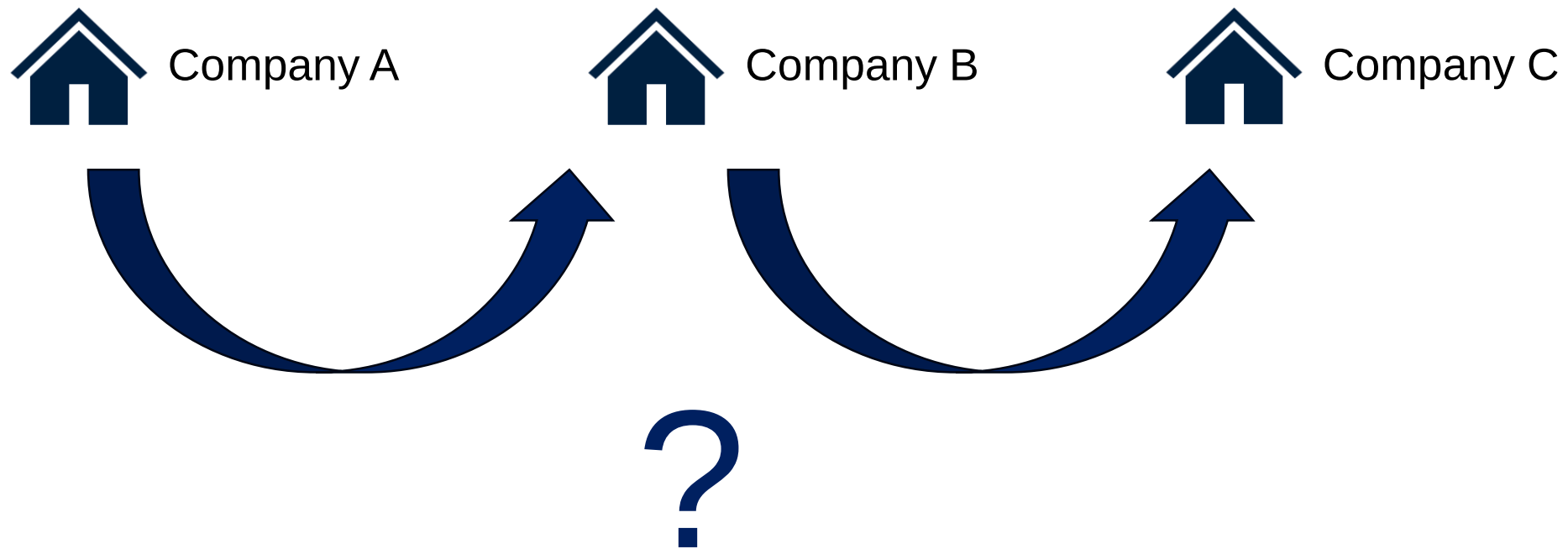
CASE 1: Using Python-based Community Detection Algorithms to Identify Hidden “Fraud Groups” in Lending.

Commercial banks usually adopt unified credit granting to determine the maximum comprehensive credit line for group customers to achieve centralized risk control. The so-called “group” customers often refer to a group of corporate legal entity clients with control relationships based on equity connections.



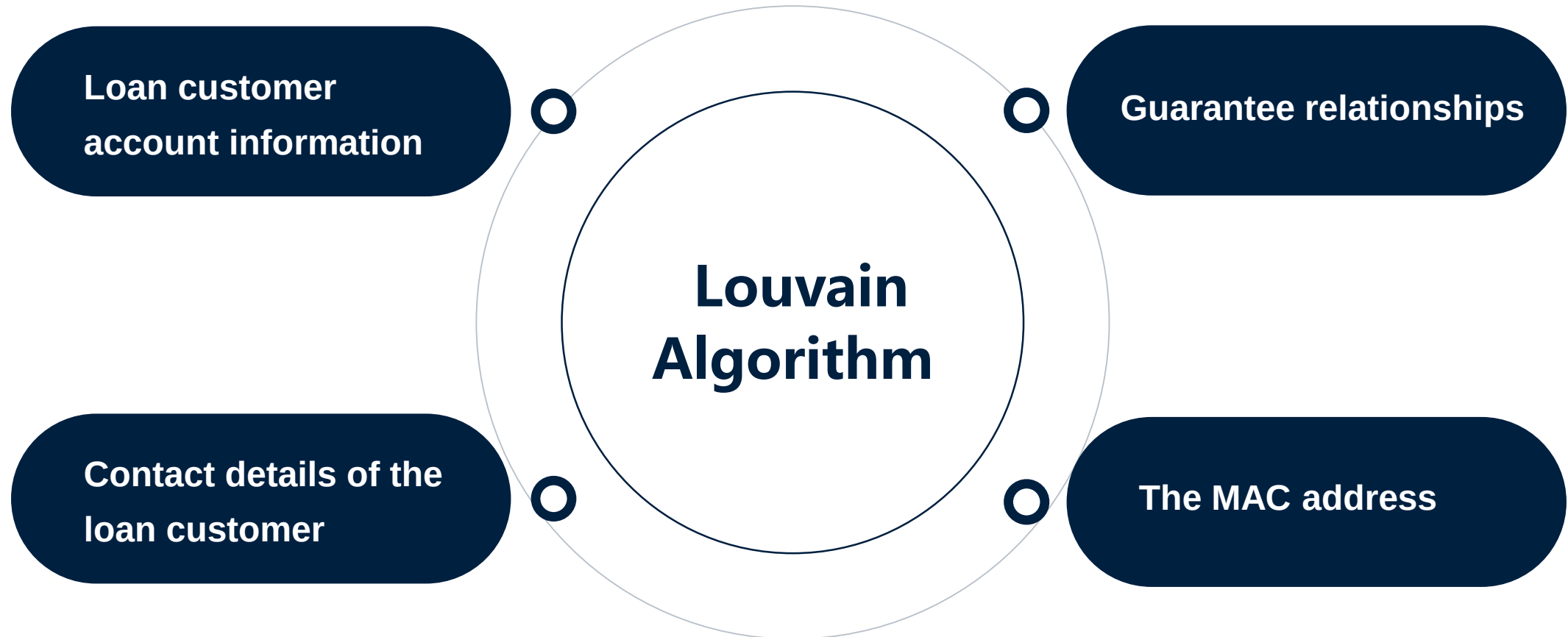
Practical Cases

CASE 1: Using Python-based Community Detection Algorithms to Identify Hidden “Fraud Groups” in Lending.



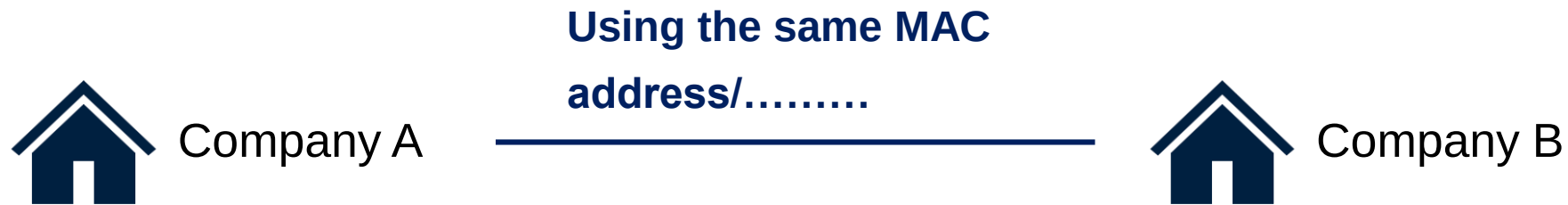
Practical Cases

CASE 1: Using Python-based Community Detection Algorithms to Identify Hidden “Fraud Groups” in Lending.



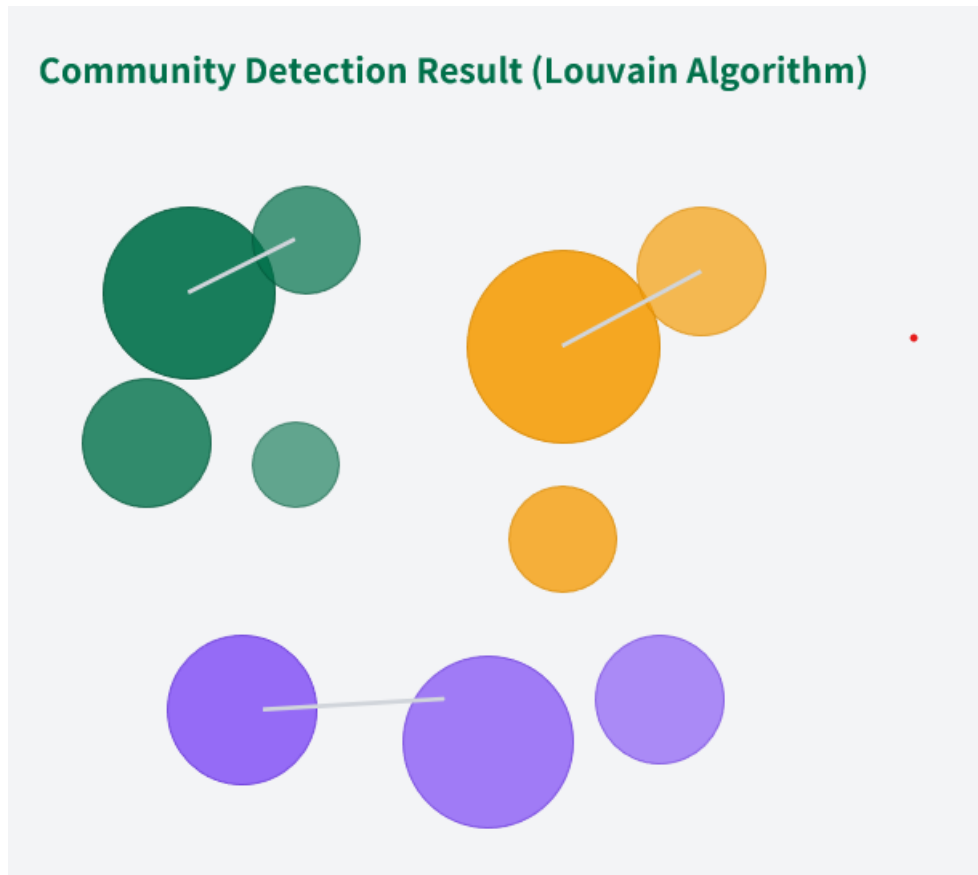
Practical Cases

CASE 1: Using Python-based Community Detection Algorithms to Identify Hidden “Fraud Groups” in Lending.



Practical Cases

CASE 1: Using Python-based Community Detection Algorithms to Identify Hidden “Fraud Groups” in Lending.



- **Nodes:** Represent individual loan applicants or businesses.
- **Edges:** Represent hidden connections, such as shared contact information, guarantor relationships, or common device fingerprints (e.g., MAC addresses).
- **Communities:** Nodes clustered in the same color form a community, indicating a high likelihood of belonging to the same fraud ring.

CASE 2: Leveraging Text Mining Techniques to Quickly Detect “Bid Rigging” Issues

Background

Practical Cases

Realistic
Examination

Pathways
& Outlook



Regulation on the Implementation of the Bidding Law of the People's Republic of China

If the tender documents of different bidders are abnormally consistent or their tender prices show regular differences, the bidders may be found to have colluded with each other in bidding.

Case Study 2: The Analytical Approach

01

Data Collection & Preprocessing

Data Source: Collected 5 years of tender docs for Company A.

Conversion: Transformed PDF/Word into structured, machine-readable text.

02

Text Similarity Calculation

Tools (Python): TF-IDF (importance) & Cosine Similarity (content match).

Process: Performed pairwise comparison of all docs to generate similarity scores.

03

Anomaly Detection & Deep Dive

Thresholding: Flagged docs with unusually high similarity scores for review.

Metadata Analysis: Examined creator info to find collusion evidence.

Case Study 2: Analysis of Tender Document Similarity

Tender Document A (Bidder 1)

Project Overview & Technical Scheme

12,000m² office building | RC frame structure with pile foundation. Construction period strictly controlled at **12 months**.

Phase-wise Implementation Plan

3-phase workflow (Foundation → Structure → Handover). Weekly progress reports submitted to supervision unit for full transparency.

Quality & Schedule Guarantee

Complies with national standards. **Rich experience in public projects** ensures on-time delivery with zero delays.

Tender Document B (Bidder 2)

Project Overview & Technical Scheme

12,000m² office building | RC frame structure with pile foundation. Construction period strictly controlled at **12 months**.

Phase-wise Implementation Plan

3-phase workflow (Foundation → Structure → Handover). Weekly progress reports submitted to supervision unit for full transparency.

Quality & Schedule Guarantee

Complies with national standards. **Rich experience in public projects** ensures on-time delivery with zero delays.



92%+ Text Similarity

Identical structure and phrasing across documents.



Suspected Bid-rigging

High repetition and consistent expressions indicate collusive behavior.

03

Practical Review

The Foundation of Insight: Overcoming Data Governance Challenges



Data Source Diversity & Inconsistency

Audit work involves data from a vast array of systems (finance, tax, industry, banking, social security, enterprise ERP, etc.).

- "YYYYMMDD" vs. "YYYY-MM-DD"
- "yuan" vs "ten thousand yuan"



Data Quality Issues

Some source data suffer from missing values, outliers, duplicate records, and other quality issues, further increasing the complexity of data governance.

- "XX Co., Ltd." vs "XX Limited Liability Company" vs "XX Company"

Bridging the Gap: Challenges in Comprehensive Data Analysis



Insufficient Integration of Business & Tech

Insufficient integration of business and technology and lack of experience among young auditors.

When results contradict business experience, inexperience can hinder the ability to interpret findings or identify root causes.



Insufficient Application of New Tech

insufficient application of cutting-edge technologies such as artificial intelligence and room for improvement in intelligence.

In practice, data analysis in current auditing still mainly relies on traditional methods such as descriptive statistics, rule matching, and outlier detection.

04

Pathways & Outlook

➤ Pathways & Outlook

Background

Practical Cases

Realistic
Examination

Pathways
& Outlook



● Focus on the Essence

In advancing technology applications, we must always remain clear that technology is a means and auditing is the goal.

● Embrace Innovation:

As Chinese President Xi Jinping has said, “Only innovators can advance and become strong.”



THANKS

Applicant: Yunhui Liu

中华人民共和国
审计署