北京工業大學
BEIJINGUNIVERSITYOFTECHNOLOGY

Relationship between Team Diversity and Innovation Performance in Interdisciplinary Research Teams within the Field of Artificial Intelligence: Decision Tree Analysis

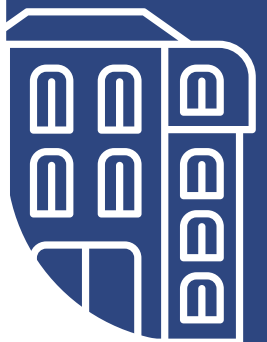
Presentation: Chenchen Huang

Authors: Junwan Liu, Chenchen Huang, Shuo Xu

Beijing University of Technology, Beijing, China

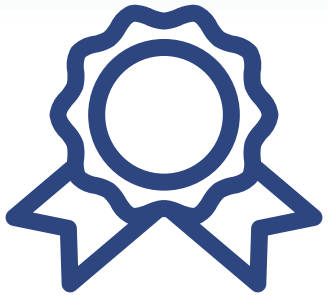
24 April 2024

Content



01

Background



Background

Transformation of Scientific Team Models Driven by Globalization and Market Competition

- Impact of **globalization** and **market competition** on the scientific research field.
- Trend towards the transition **from single-discipline** teams to **diversified** teams.

Interdisciplinary Research as an Inevitable Choice

- Rise of interdisciplinary research.
- Advantages of interdisciplinary teams in addressing **complex scientific questions**.

Importance of Effective Interdisciplinary Collaboration

- **Crucial role** of interdisciplinary collaboration in driving scientific and technological progress.
- Contributions of diversity in disciplines and backgrounds towards **knowledge and innovative outcomes**.

Background

Current Research on Diversity

- Demographic diversity has no significant relationship with team performance.
- Positive relationship between task-related diversity and team performance.

Classification of Diversity

- Demographic Diversity
- Task-Related Diversity

Crucial Factor

- Diversity is an essential impact factor affecting the success of interdisciplinary research teams.

Limitations of Existing Research

- The **nonlinear** relationship between demographic diversity and task-related diversity in interdisciplinary research teams' performance is unclear.

Challenges and Issues with Diversity in Interdisciplinary Research Teams

- Potential problems arising from diversity: **gender differences**, **team conflicts**, **collaboration difficulties**.

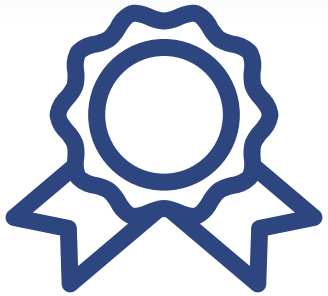
Background

» Research Questions

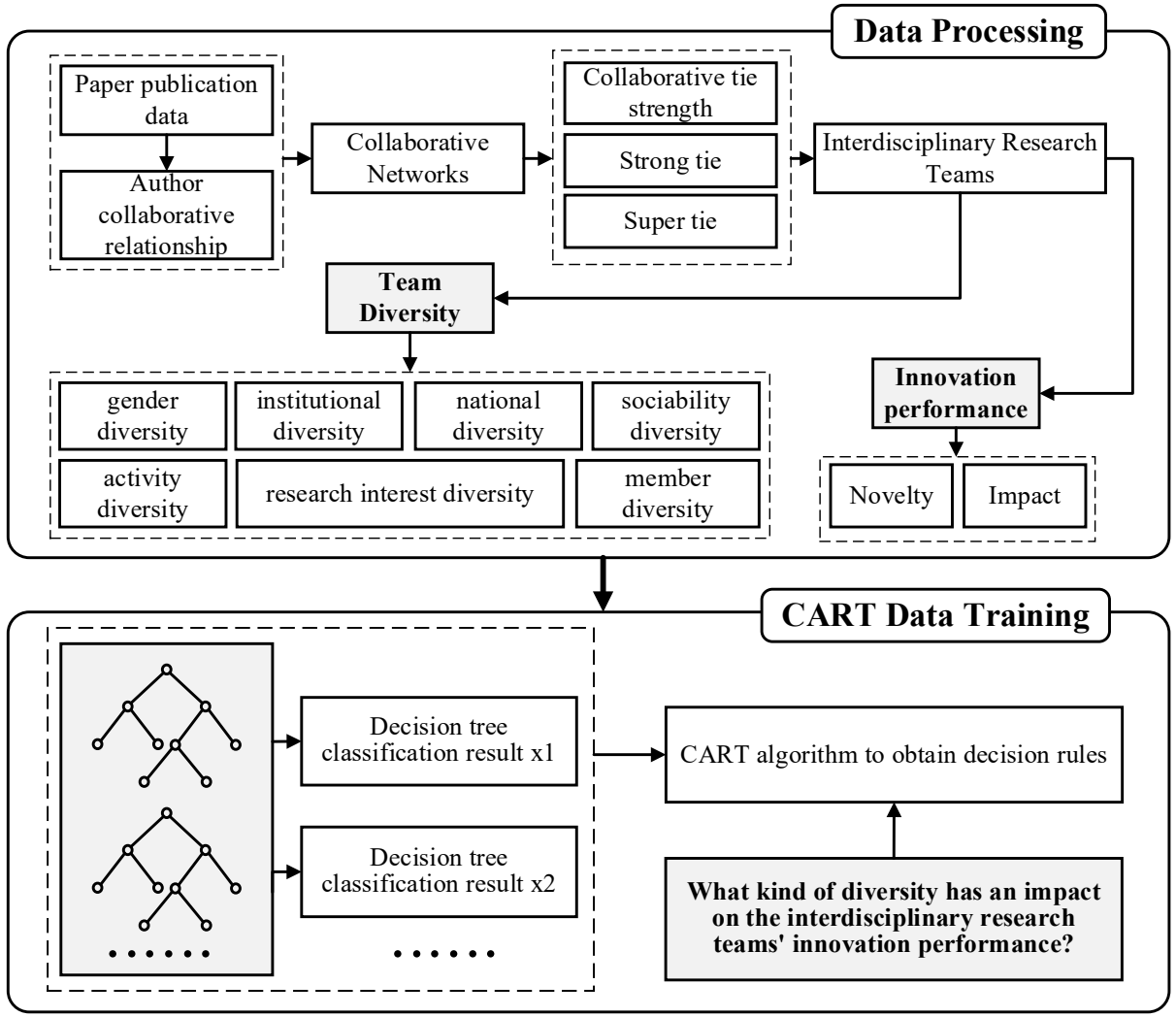
- **RQ1:** What is the **complex linear relationship** between demographic diversity, task-related diversity, and the innovation performance of interdisciplinary research teams?
- **RQ2:** What **characteristic combinations** promote high levels of team innovation performance?
- **RQ3:** Which diversity should researchers **prioritize to enhance the innovation performance** of interdisciplinary research teams?

02

Methodology



Methodology——Framework



Research Framework

◆ Data Processing

1. Identification of Interdisciplinary Research Teams.
2. Measurement of Diversity Index.
3. Measurement of Team Innovation Performance.

◆ CART Data Training

1. Exploration of the Impact of Team Diversity on Innovation Performance Using the CART Model.

Methodology—Data Processing

Data Processing

Data collection

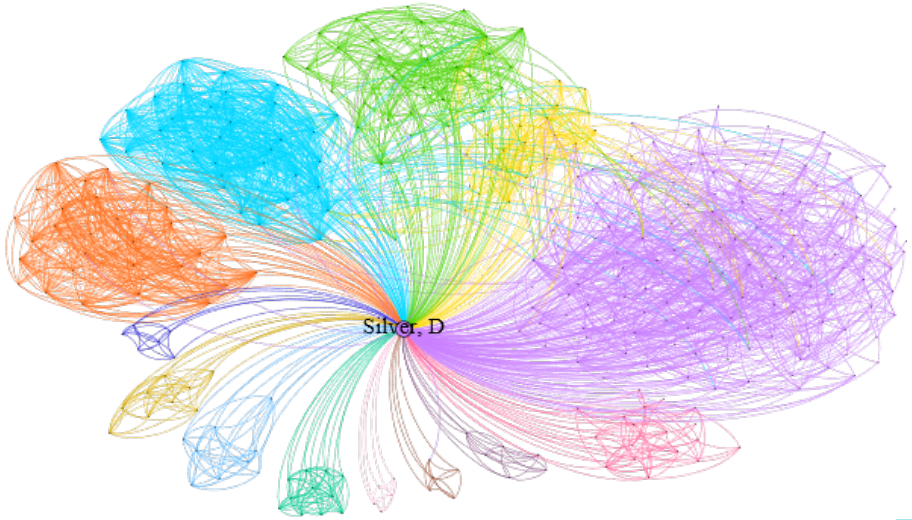
» **Data Source:** The AMiner website's 2023 AI 2000 annual list of the most influential scholars.

» **Data Content:** Information on 195 scholars and their core collaborators.

Community Detection

» **Methodology:** Community detection in the scholar co-occurrence network using the Louvain algorithm.

- Using Silver, D, and other selected scholars as **examples** to demonstrate the co-occurrence network, with different colors representing various communities and **purple** highlighting the community affiliated with the target scholars.



Collaboration Network of Silver, D scholar

Methodology—Data Processing

Core Collaborators Identification

» Calculation of Collaborative Tie Strength

- Calculating the collaborative tie strength of nodes within each community.

» Screening of Core Collaborators

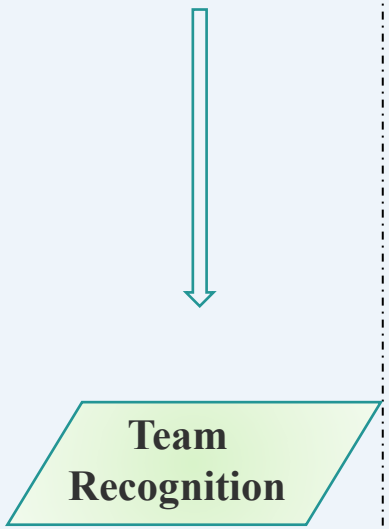
- Identification of core collaborators among the 195 selected scholars.
- **Threshold Determination:** Calculation of the super tie for each community.
- **Characteristics of Super Tie:** Long-term, stable, high-intensity, close ties, and enduring duration.
- **Definition of Core Collaborator:** A member whose collaborative tie strength exceeds the threshold.

» Formula for Calculating Super Tie

- $\langle K_i \rangle = S_i^{-1} \sum_{j=1}^{S_i} K_{ij}$
- $K_i^c = (\langle K_i \rangle - 1) \ln S_i$

Where the collaborative tie strength K_{ij} is defined as the cumulative number of papers co-authored by the selected scholar in community i and scholar j over the time between their first and last paper. S_i represents the number of different co-authors of selected scholars in the community i. $\langle K_i \rangle$ represents the average collaborative tie strength K_{ij}. Each scholar j with $K_{ij} > K_i^c$ is labelled as a super tie collaborator of community i.

Methodology—Data Processing



» Identification of Research Teams and Core Members

- **Identification Results:** A total of 195 research teams and their 1,217 core members have been determined.

» Interdisciplinary Verification

- **Verification Methodology:** Utilizing a mapping approach that aligns member affiliations with disciplinary classifications.
- **Detailed Steps:**
  - Extraction of secondary institutions from each member's address.
  - Retention of disciplinary terms within the institution names.
  - Matching of these terms to the disciplinary fields in the OECD classification scheme.
- **Verification Outcomes:**
  - 165 teams have members from two different disciplinary backgrounds.
  - 27 teams have members from three different disciplinary backgrounds.
  - 3 teams have members from four different disciplinary backgrounds.

Methodology—Data Processing

Team Recognition

» Disciplinary Distribution of Team Members

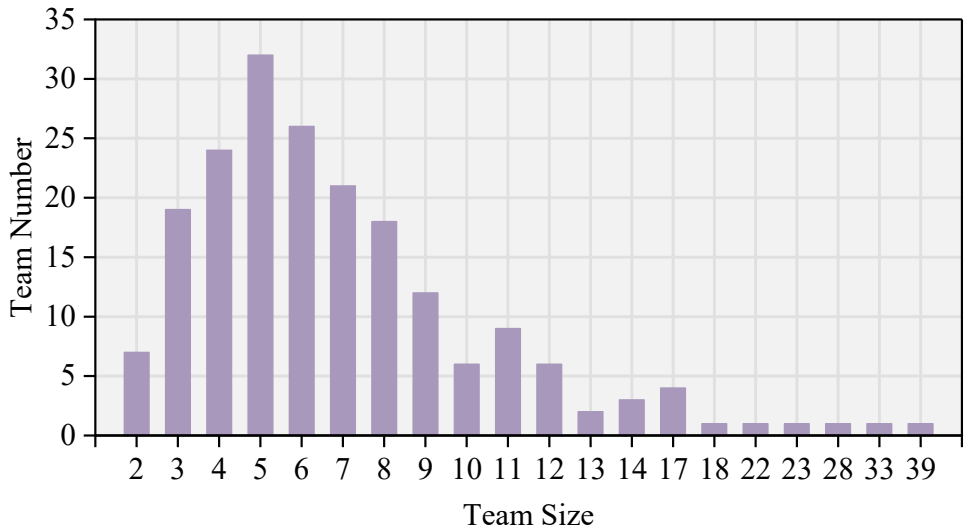
- **Overview of Disciplinary Distribution:** Presentation of the percentage breakdown of members across various disciplines.
- **Dominant Disciplines:** Computer and Information Science (71.18%), Electrical, Electronic, and Information Engineering (21.80%).

» Team Size Distribution

- Maximum team size is 57, 85% of teams are less than 10 people.

Distribution of Members from Different Disciplinary Fields

Discipline	Percentage of members
Computer and information science	71.18%
Electrical engineering, electronic engineering, information engineering	21.80%
Environmental engineering	2.03%
Nano-technology	1.52%
Physical science	1.09%
Health science	0.94%
Clinical medicine	0.51%
Mathematics	0.36%
Materials engineering	0.22%
Medical engineering	0.22%
Basic medicine	0.14%



Size Distribution of Teams

Methodology—Data Processing

Variables

» Innovation Performance

- **Novelty:** $\text{Commonness}_{ijt} = \frac{observed\ number\ of\ pairs_{ijt}}{expected\ number\ of\ pairs_{ijt}} = \frac{N_{ijt}}{\frac{N_{it}}{N_t}\times\frac{N_{jt}}{N_t}\times N_t} = \frac{N_{ijt}\times N_t}{N_{it}\times N_{jt}}$
- **Impact:** Based on forward citations.

» Demographic Diversity

- **Gender diversity**
- **National diversity**
- **Institutional diversity**

$$H = 1 - \sum_{i=1}^{n} P_i^2$$

- **Task-related Diversity**
- **Sociability diversity** $sociability(A) = 1 + \sum_{each\ coauthor(c)} ln(\#copaper_c)$
- **Activity diversity** $\text{activity(A)} = \sum_{each\ year(n) in\ recent\ N\ years} \text{IS}(G_n) \times \text{weight(n)}$
- **Research interest diversity** $P_A(\text{t}) = \frac{\#papers\ of\ A\ belong\ to\ topic\ t}{\#all\ paper\ of\ A}$

$$\text{research interest diversity(A)} = -\sum_{t\in all\ topic\ of\ A} P_A(t) log P_A(t)$$

- **Member diversity** $\text{member diversity} = \frac{relation_i}{coauthor_i(non-duplicate)}$

Methodology—CART Data Training

CART Data Training

Idea of a CART Decision Tree

» Limitations of Traditional Regression Models

- **Challenges:** Difficulties in handling multivariate and nonlinear relationships.
- **Methodological Constraints:** Reliance on the least squares method, requiring rigorous hypothesis testing and variable control.
- **Potential Issues:** Inaccuracies in regression analysis due to improper selection or omission of control variables.

• Advantages of the CART Model

- **Characteristics:** A classification and regression technique generated through the regression of fork decision properties.
- **Flexibility:** Builds decision trees through recursive binary splitting, offering adaptability.
- **Robustness:** Automatically selects partitioning rules based on the actual data distribution and characteristics, eliminating the need for strict hypothesis testing.
- **Capturing Complex Relationships:** Able to capture non-linear relationships and higher-order interactions.

Methodology—CART Data Training

Design of the CART Decision Tree

» Model Construction

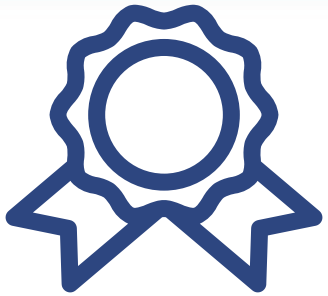
- **Dataset partitioning:** Splitting the dataset into training and testing samples in an 8:2 ratio.
- **Model selection:** Employing the CART model alongside benchmark models such as C4.5, Random Forest, and Gradient Boosting Tree.

» Model Performance Optimization and Evaluation

- **Grid search approach:** Systematically exploring various hyperparameter combinations to determine optimal configurations.
- **Performance metrics:** Achieving model accuracies exceeding 0.6, with the CART model demonstrating superior performance.
- **CART model accuracy:** Demonstrating 0.73 and 0.68 accuracy in assessing the novelty and impact of the research teams.

03

Results



Model Result Analysis

- Interdisciplinary research teams demonstrate a **higher proportion** of high innovation performance ratings in terms of **impact**.
- Activity diversity and member diversity emerge as **key factors** influencing both novelty and impact.
- The confidence coefficients for most decision rules **exceed** 60%.

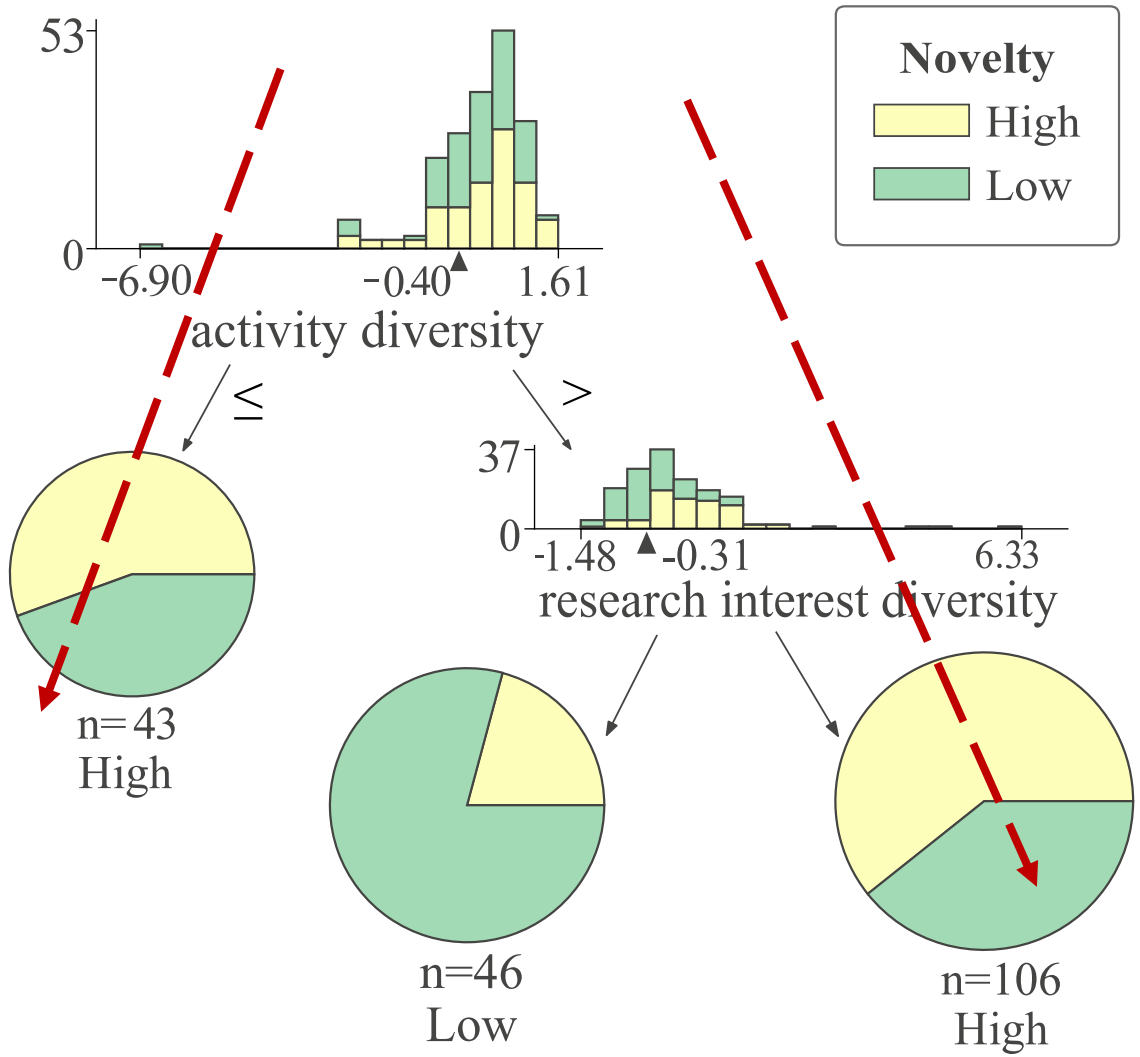
Innovation Performance Decision Rules

	Demographic diversity				Task-related diversity			Decision results	Support	Confidence
	GD	ID	ND	SD	AD	RID	MD			
Novelty	-	-	-	-	> -0.40	> -0.31	-	Low	29.00%	61.00%
	-	-	-	-	> -0.40	<= -0.31	-	High	12.00%	79.00%
	-	-	-	-	<= -0.40	-	-	High	14.00%	66.00%
Impact	-	<= -0.72	-	-	-	<= 0.47	<= -0.08	Low	9.00%	78.00%
	-	> -0.72	-	-	-	<= 0.47	<= -0.08	High	31.00%	60.00%
	-	-	-	-	-	> 0.47	<= -0.08	Low	7.00%	86.00%
	-	-	-	-	-	<= -0.29	> -0.08	High	8.00%	93.00%
	-	-	-	-	-	(-0.29,0.08]	> -0.08	Low	5.00%	70.00%
	-	-	-	-	-	(0.08,0.47]	> -0.08	High	19.00%	74.00%

Results

Model Result Analysis

- Data partitioning based on the **core feature** of "activity diversity".
- Improving team innovation performance when activity diversity is low.
- When activity diversity is high, increased research interest diversity helps teams achieve high innovation performance.
- Interdisciplinary research teams exhibiting high activity diversity are significantly affected by research interest diversity.

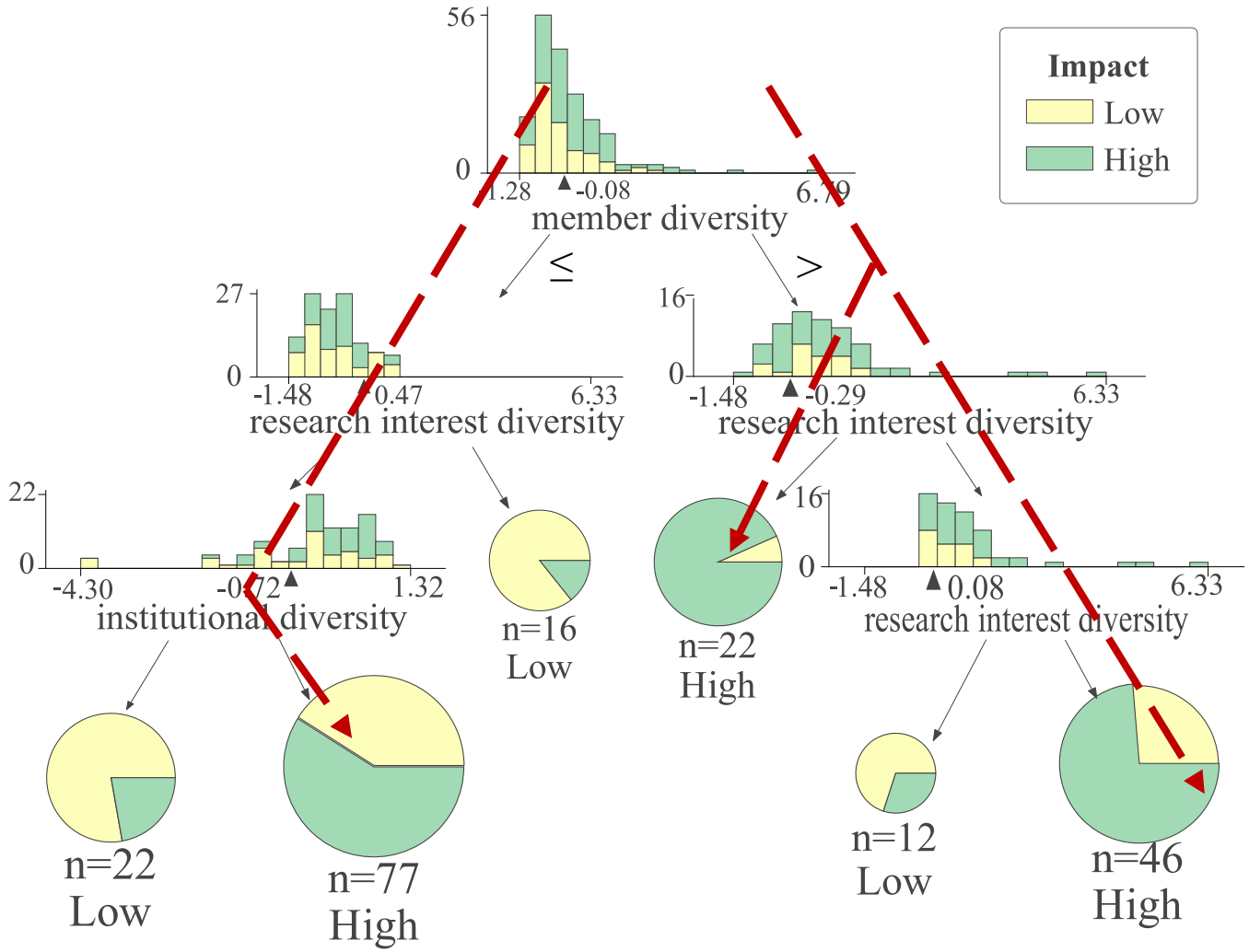


Decision Tree for Team's Novelty

Results

Model Result Analysis

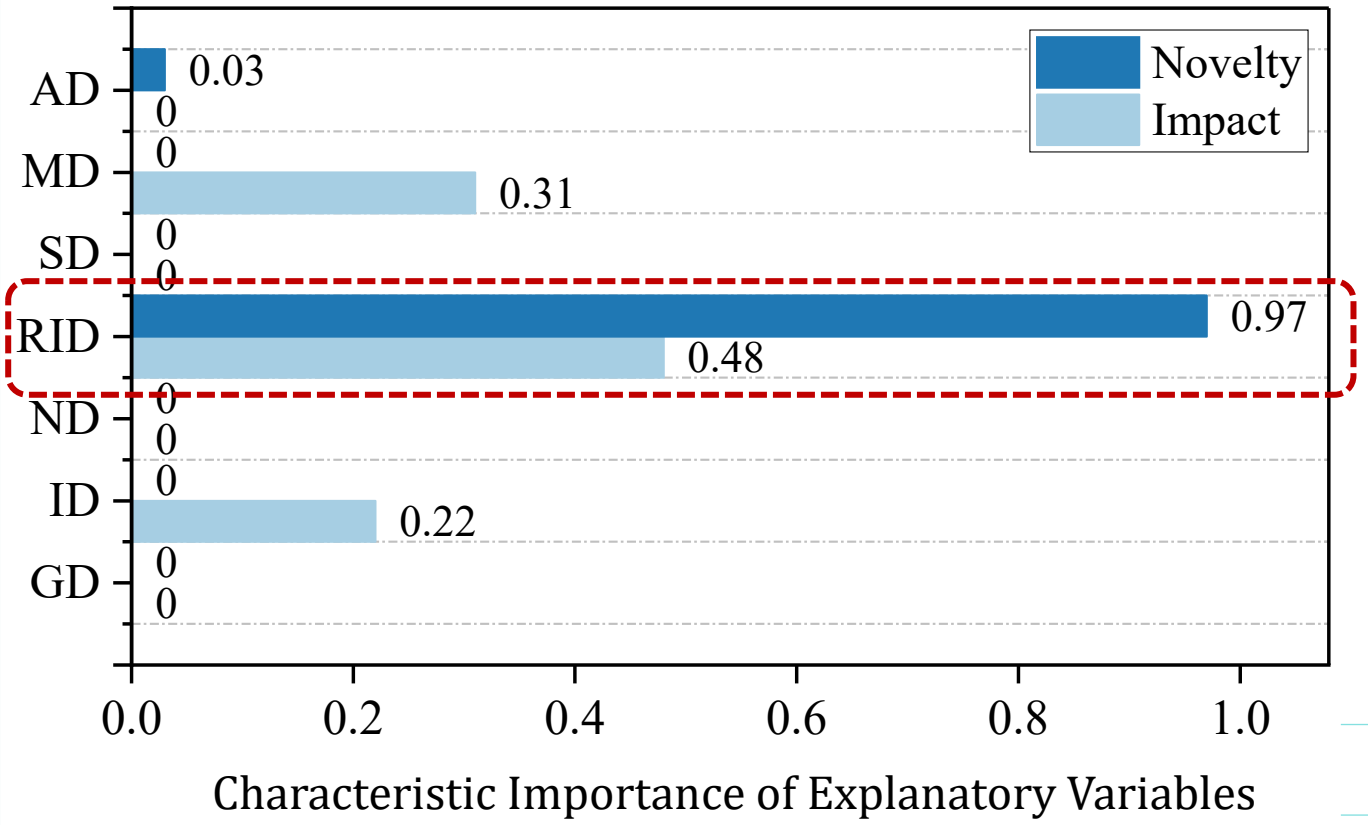
- Data partitioned into two primary branches based on the **core feature** of "member diversity".
- When the member diversity is higher than -0.08, interdisciplinary research teams need to control the research interest diversity within a **suitable range** to achieve high innovation performance.
- When member diversity is below -0.08, institutional diversity becomes a key factor influencing high innovation performance.



Decision Tree for Team's Impact

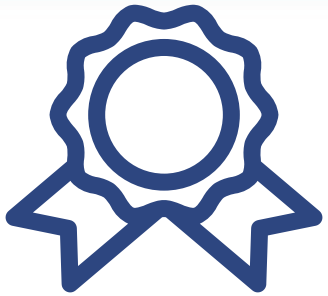
Feature Importance Analysis

- Factors Influencing Team Novelty:
  - Research interest diversity has the highest characteristic importance, at 0.97.
  - Activity diversity has a characteristic importance of 0.03.
- Factors Influencing Team Impact:
  - Research interest diversity has the highest characteristic importance, at 0.48.
  - Member diversity and institutional diversity follow closely, at 0.30488 and 0.22 respectively.
- Research interest diversity is **most strongly associated** with the innovation performance of interdisciplinary research teams.



04

Conclusion



Conclusion

1. The study reveals a **U-shaped relationship** between activity diversity and the team's innovation performance in terms of "novelty". This relationship is significantly influenced by the research interest diversity.

2. In terms of "impact" innovation performance, increasing member diversity and **managing the range of** research interest diversity can be beneficial.

3. Research interest diversity emerges as the **most significant determinant** of innovation performance in interdisciplinary research teams.

Contributions

1 We proposed a **framework** for studying the impact of team diversity on the innovation performance of interdisciplinary research teams.

2 We applied CART model to the study of team innovation performance, providing **multiple pathways** through which team diversity affects the innovation performance of interdisciplinary research teams.

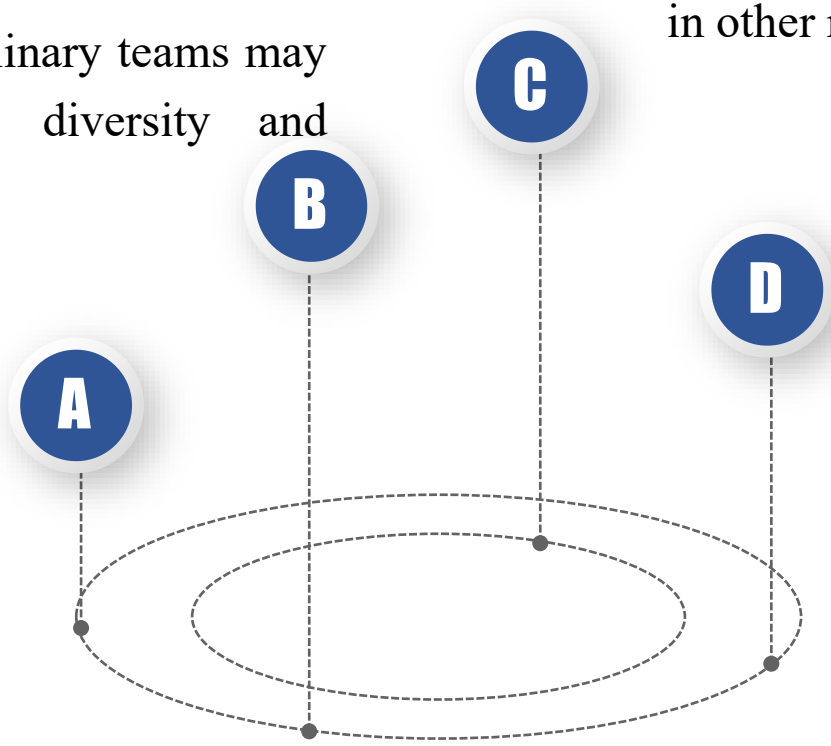
Limitations

Insufficient Sample Representativeness

- The sample of 195 interdisciplinary teams may not adequately reflect the diversity and complexity of AI teams.

Scope of Research

- The study is solely focused on the AMiner platform in AI, limiting its scope and generalizability.



Limited Scope of Team Diversity Research

- This study focuses on diversity within interdisciplinary teams and ignores diversity factors in other research team activities.

Overlooking Dynamics of Member Turnover

- Our current approach to identifying team members fails to capture the dynamism of member turnover.

北京工業大學
BEIJINGUNIVERSITYOFTECHNOLOGY

THANKS

Presentation: Chenchen Huang

Authors: Junwan Liu, Chenchen Huang, Shuo Xu

Beijing University of Technology, Beijing, China

24 April 2024