What Do Computing Interns Discuss Online? An Empirical Analysis of Reddit Posts

Saheed Popoola, Ashwitha Vollem, Isaac Kofi Nti

University of Cincinnati

next lives here

University of CINCINNATI

Background

- Internships are key to gain real-world experience and meet expectations of workforce
- Internship experience varies
- Existing research often use surveys or questionnaires
  - Limit scale/ diversity
- Social media platforms provide wider reach

next lives here

University of CINCINNATI

2024

550 MILLION posts

2.72 B comments

91M Reddit daily active users

2M subreddits

UNDERSTAND INTERN DISCUSSIONS TO INFORM EDUCATION

Contributions

- Use case on the use of Reddit for educational purposes.
- Publicly available dataset on computing internships
  - 143,912 posts (June 2009 to December 2023)
  - Python scripts for data collection and analysis
- Themes in internship discussion

next lives here

University of CINCINNATI

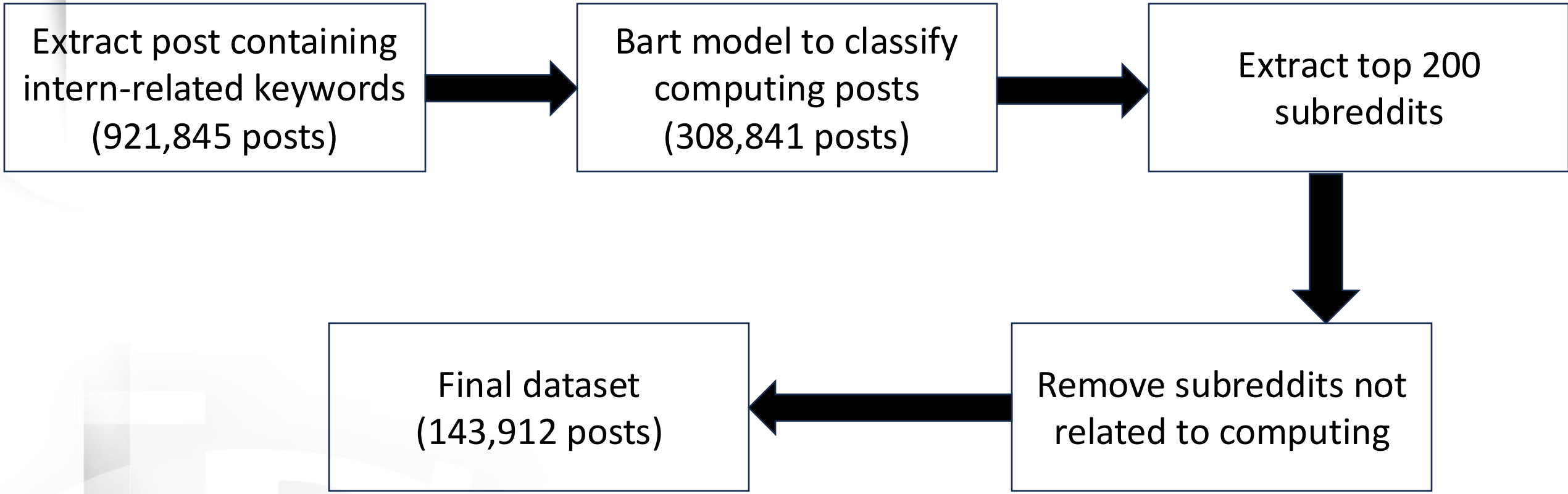
Research Questions

- RQ1: What themes appear in computing intern discussions?
- RQ2: What sentiments do interns express?
- RQ3: What domains do posts fall into?

next lives here

University of CINCINNATI

Data Collection



next lives here

University of CINCINNATI

Methodology

- Topic Modeling
  - Latent Dirichlet Allocation (LDA) - Gensim
  - Manual coding
- Sentiment Analysis.
  - VADER + TextBlob
  - Classification: positive / negative / neutral

next lives here

University of CINCINNATI

Results

- RQ1: What themes appear in computing intern discussions?
- RQ2: What sentiments do interns express?
- RQ3: What domains do posts fall into?

next lives here

University of CINCINNATI

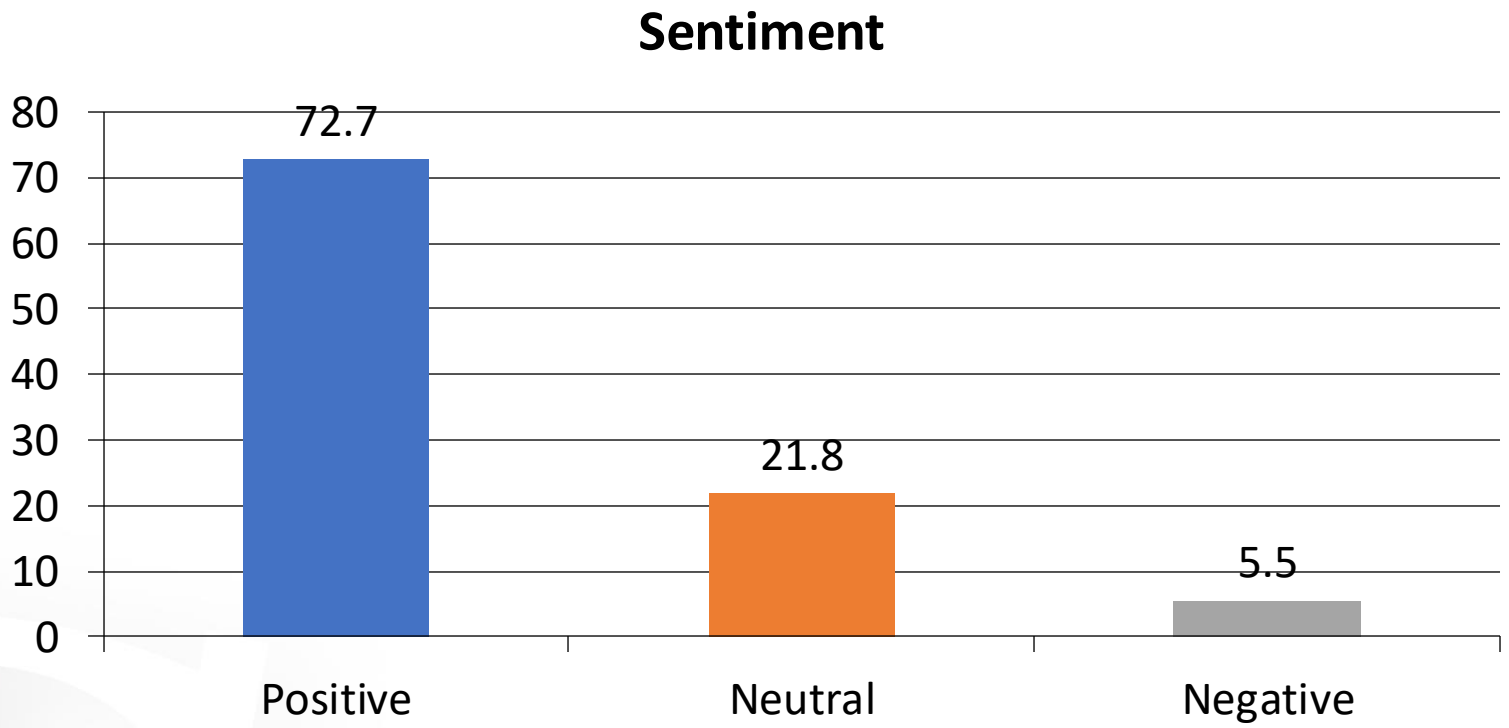
Themes (RQ1)

- Academics & admissions
- Professional career
- Core internship tasks
- Social interactions
- Entertainment

next lives here

University of CINCINNATI

Sentiment Distribution



next lives here

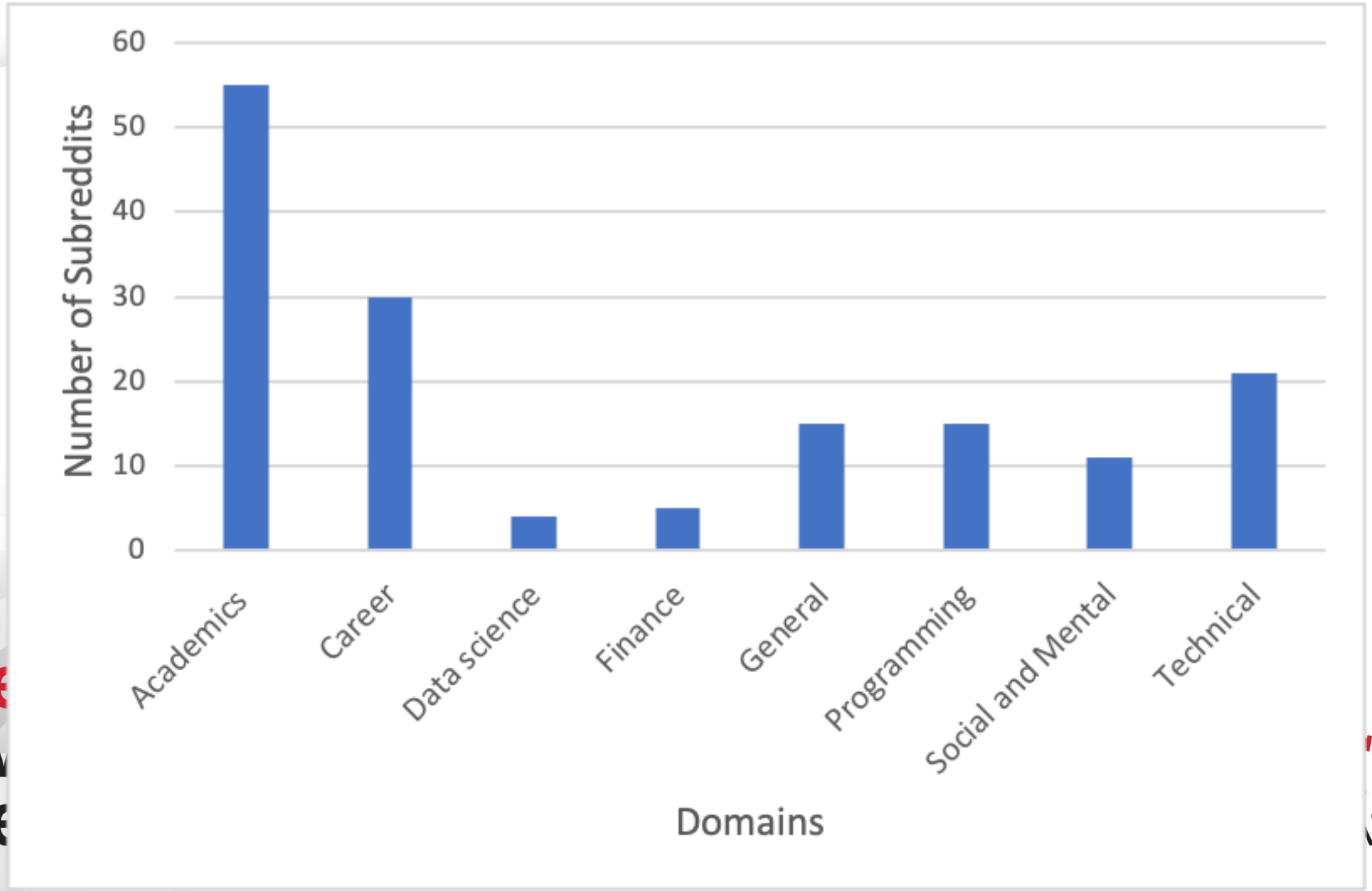
University of CINCINNATI

Domains(RQ3)

- Reddit is composed of subreddits
- Each subreddit targets a subject area
- Every post is associated with exactly one subreddit
- Subreddits analysis can provide insight into the subject area of intern’s discussion

next
lives
here

University of
CINCINNATI


60
50
40
30
20
10
0
Number of Subreddits
Academics
Career
Data science
Finance
General
Programming
Social and Mental
Technical
Domains

Implications

next
lives
here

University of
CINCINNATI

Strengthen
career
preparation

SOFT
SKILLS

INDUSTRY
PARTNERSHIPS

?

Threats to Validity

- Reddit only
- Possible misclassification
- No time-series separation
- Anonymity limits demographics

next lives here

University of CINCINNATI

Future Work

- Other forums e.g., StackOverflow
- Time-series/ demographics analysis
- Analyze replies to posts
- Compare computing vs non-computing interns
- Adopt transformer models (e.g., BertTopic) for the analysis

next lives here

University of CINCINNATI

Conclusion

- Reddit reveals major intern themes
- Mostly positive sentiment
- Broad life and career focus
- Peer mentorship and support
- Enhanced industry collaboration
- Can inform curriculum and support systems

next lives here

University of CINCINNATI