Tracking the increasing complexity of Deep Learning Research

- Aniket Gupta, Liam Pavlovic, Vinit Nagap

Goal

- Track and visualize the trends in Deep Learning research across time and conferences.
- Visualize the interdependencies of sub-fields.
- Interactive visualizations to understand the evolution of sub-fields over time.
- Understand the trend in what sub-fields have been the most popular over time.

Background

Deep Learning Research

Exponential growth in research and application.

- Over 3,500 published in ICLR 2023
- Over 19,000 submissions in ICLR 2023
- Over 17,000 submissions in NeurIPS 2023

Challenge

Staying updated amidst increasing complexity.

- Research is coming out at an unprecedented rate
- Tracking developments in different sub-fields is virtually impossible

Solution

Dynamic visualizations and analysis tools for informed decision-making

- By visualizing the trends in growth of different sub-fields, we can methodically track research.
- This also provides an idea of how many people are actively working in a particular research area

Data

Source

The OpenReview Repository for the International Conference on Learning Representations scraped with OpenReview Scraper

Fields

Authors
List of author names

Title
Title of paper

Keywords
List of author-chosen keywords

Venue
Conference paper was accepted to

Link
Link to full text

Lowercase Applied

Domain & Corresponding Data Questions

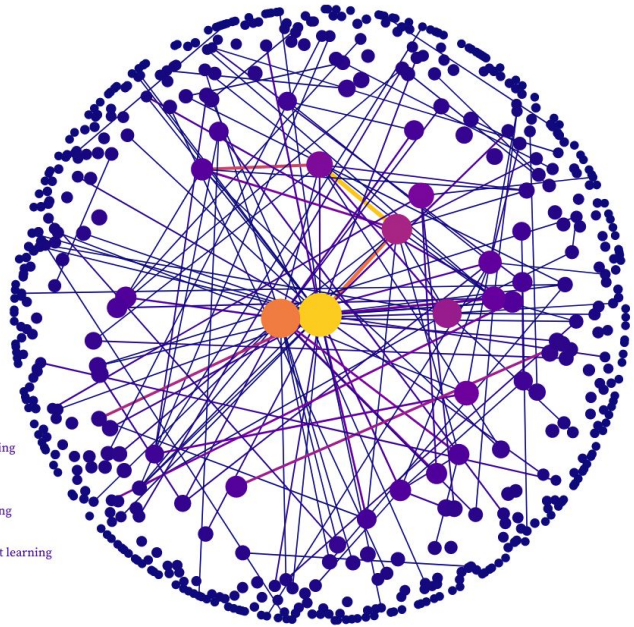
1. How has the overall volume of papers changed over the years?
    a. How does the total number of accepted papers at a specific ML conference vary by conference year?
2. How do different sub-fields intersect? What intersections have the most interest? How does interest in an intersection affect interest in the sub-field as a whole?
    a. What is the co-occurence rate between different pairs of keywords in a given year?
    b. Do keywords with a high level of co-occurrence have a similar number of associated publications?
    c. How do the co-occurrence relationships between keywords vary by conference year?
3. What are the most popular sub-fields in a given year? Do the most popular sub-fields change dramatically between consecutive years?
    a. For a given year, what are the top 10 most frequently occurring keywords?
    b. How do the top keywords vary between subsequent conference years?
4. Are there specific years where certain sub-fields experienced rapid growth/rapid decline?
    a. How does the total number of accepted papers containing a specific keyword vary by conference year?

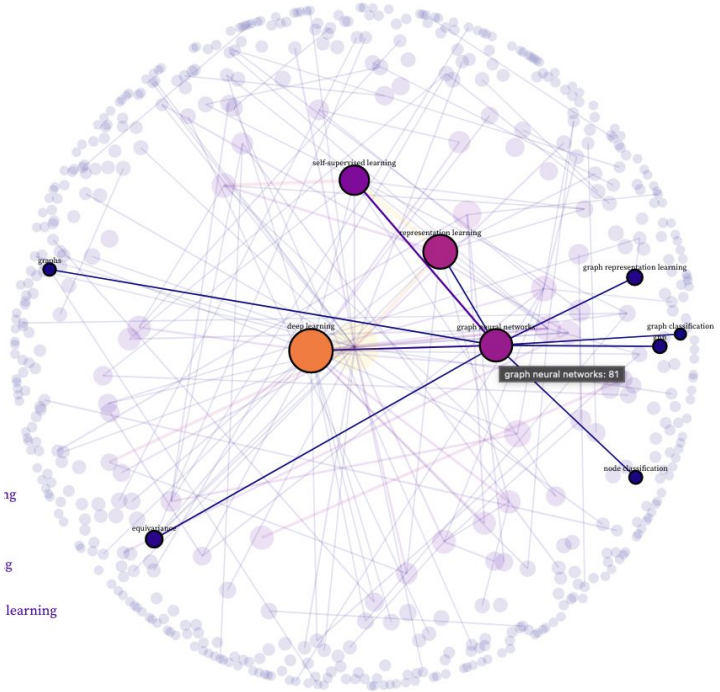
Graph View

Year 2022

Number of Shared Publications for Edge 3

1. reinforcement learning
2. deep learning
3. representation learning
4. graph neural networks
5. self-supervised learning
6. federated learning
7. robustness
8. generalization
9. transformer
10. contrastive learning
11. interpretability
12. transfer learning
13. continual learning
14. generative models
15. computer vision
16. optimization
17. neural networks
18. meta-learning
19. machine learning
20. deep reinforcement learning
21. adversarial robustness
22. few-shot learning
23. natural language processing
24. adversarial training
25. multi-agent reinforcement learning



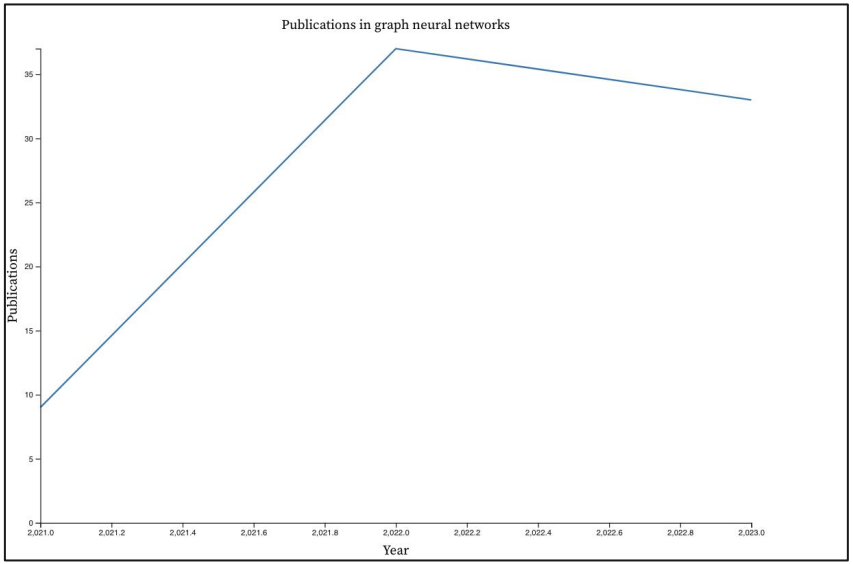


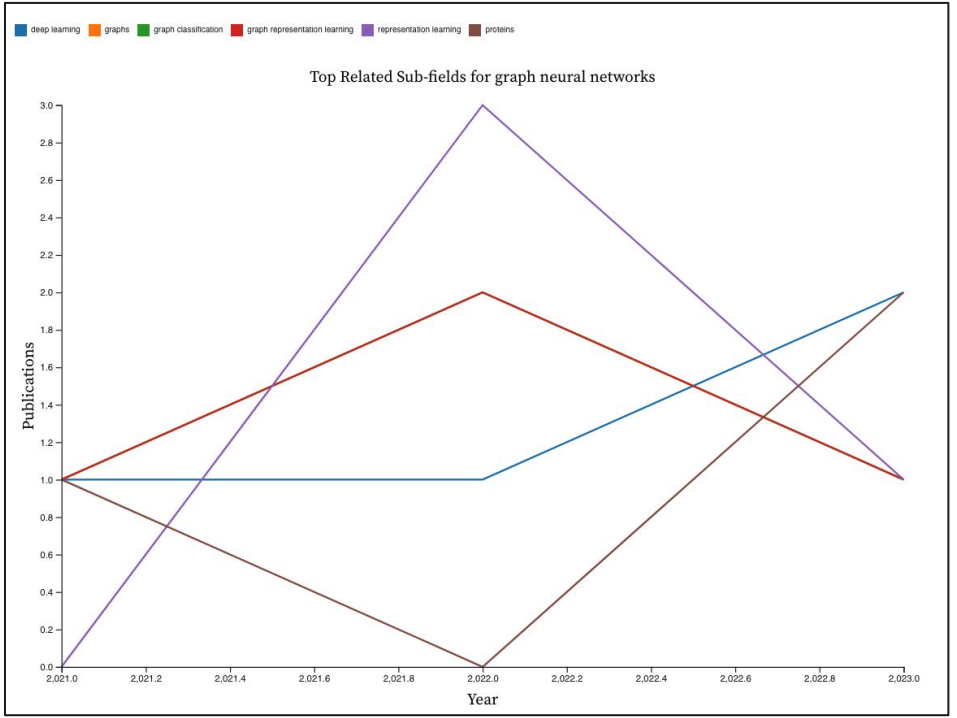
ıg

g

learning

Secondary Line Views

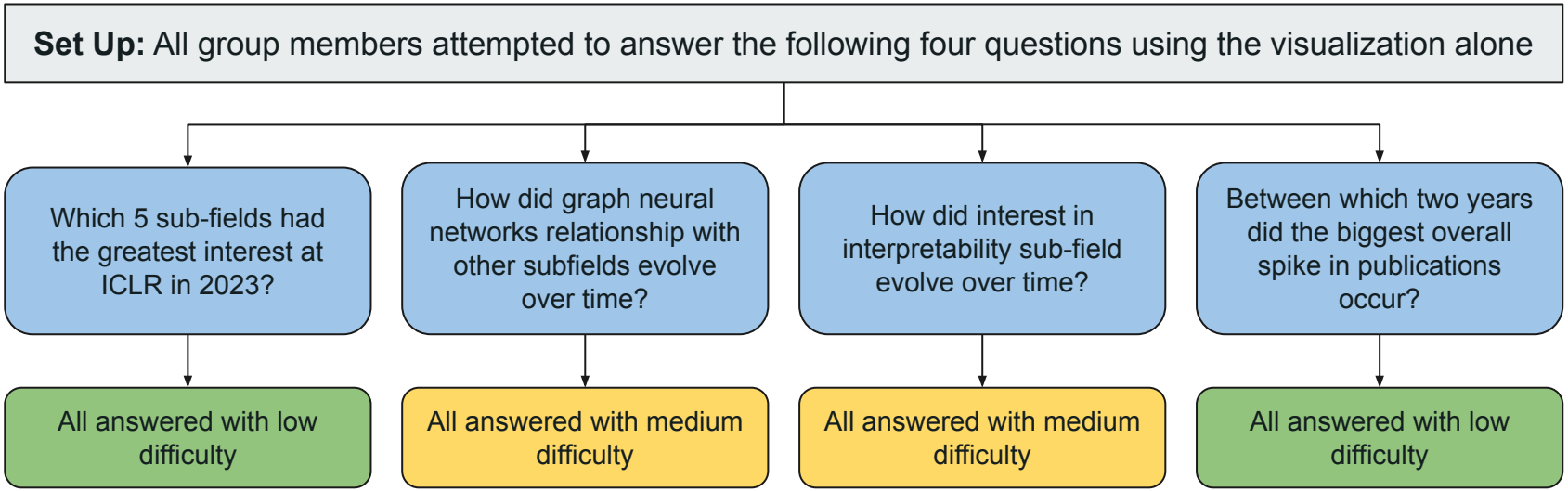




Findings

- Top 3 most popular fields are constant over time
  - Reinforcement Learning, Deep Learning and Representation learning are the top fields
  - Self-supervised learning also has significant interest.
- Overall volume of publications increased sharply from 2017
- Federated learning and GNNs experience sharp spike in interest recently, now top 4/6
- Very few intersections have strong (>2 publications) interest
  - The volume of single-publication intersections is very high
- Strong relationships between very popular and relatively smaller fields implies hierarchy

Evaluation


Set Up: All group members attempted to answer the following four questions using the visualization alone
Which 5 sub-fields had the greatest interest at ICLR in 2023?
How did graph neural networks relationship with other subfields evolve over time?
How did interest in interpretability sub-field evolve over time?
Between which two years did the biggest overall spike in publications occur?
All answered with low difficulty
All answered with medium difficulty
All answered with medium difficulty
All answered with low difficulty

Reflections & Lessons Learned

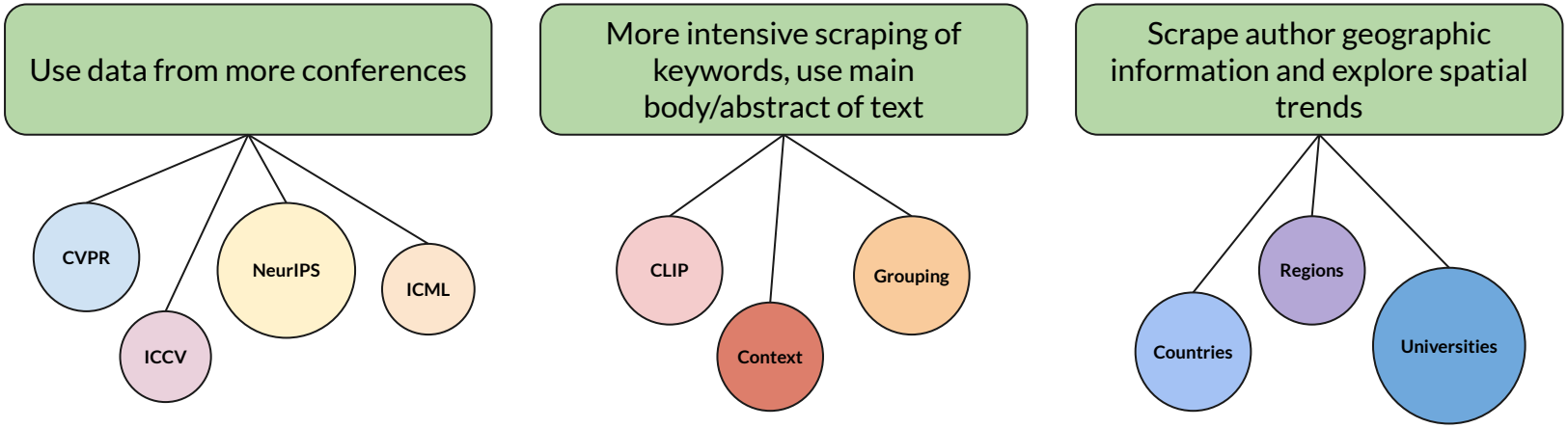
Control Interactivity and Number of Views

Redundantly encode information in multiple channels when possible

Data accessibility may be lower than anticipated

Be careful relying on attributes with subjective/varied interpretations (i.e. keywords)

Future Work


Use data from more conferences
CVPR
ICCV
NeurIPS
ICML
More intensive scraping of keywords, use main body/abstract of text
CLIP
Context
Grouping
Scrape author geographic information and explore spatial trends
Countries
Regions
Universities

Thank you!