Milestone 2

By Liam Pavlovic, Aniket Gupta, and Vinit Nagap

Goal Statement

Deep learning, as a research field, has seen significant growth, both in complexity and volume of publications, during the last decade. This rapid growth makes it difficult for researchers in the field to keep up with current developments, leading to feelings of inadequacy and burnout. We want to visualize how the field has developed over the last decade so researchers can better understand how the field grows and impacts the people working in it. We particularly want to capture how different sub-fields, and their relationships with other sub-fields, develop.

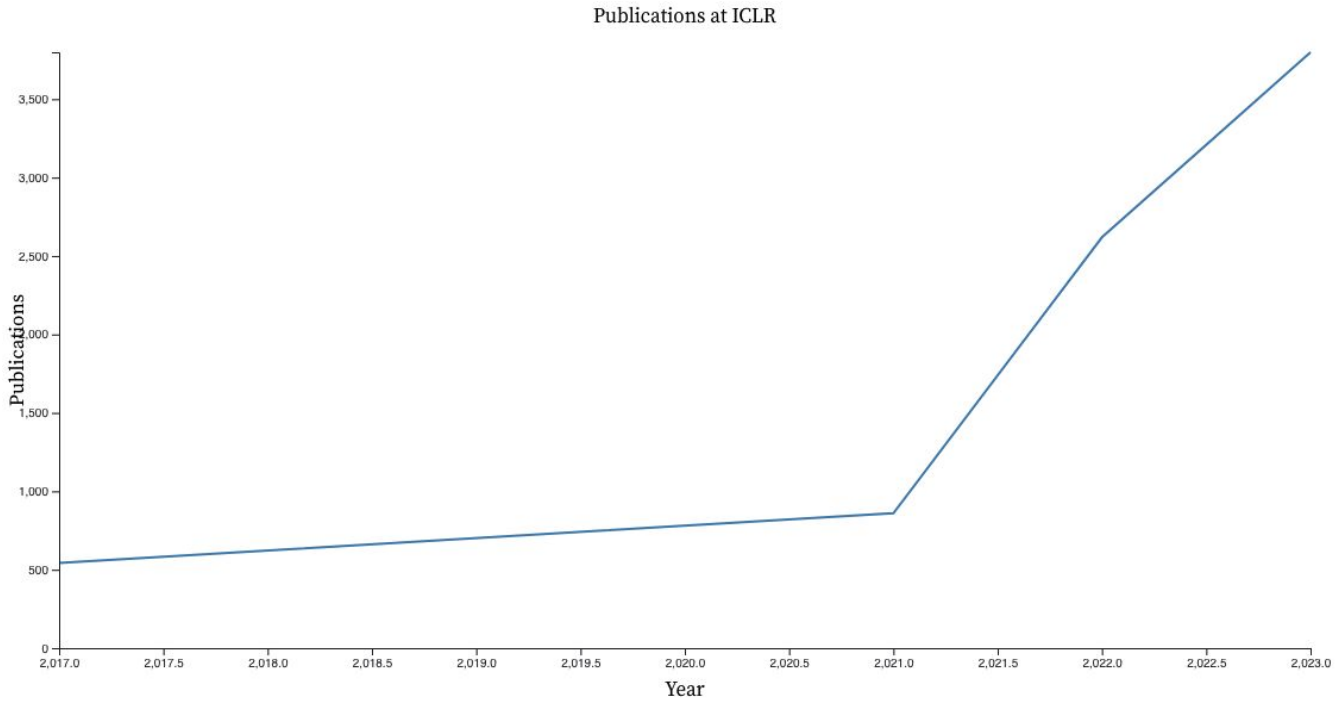
Questions

How does the total number of accepted papers at a certain ML conference vary by conference year?

What is the co-occurence rate between different pairs of keywords?Are there any groups of keywords that all have a high co-occurrence rate between each other? How do these co-occurrence relationships vary by conference year? Do keywords with a high level of co-occurrence have a similar number of associated publications?

For each year, what are the top 10 most frequently occurring keywords? How do these top keywords vary by conference year? How does the total number of accepted papers containing a specific keyword vary by conference year?

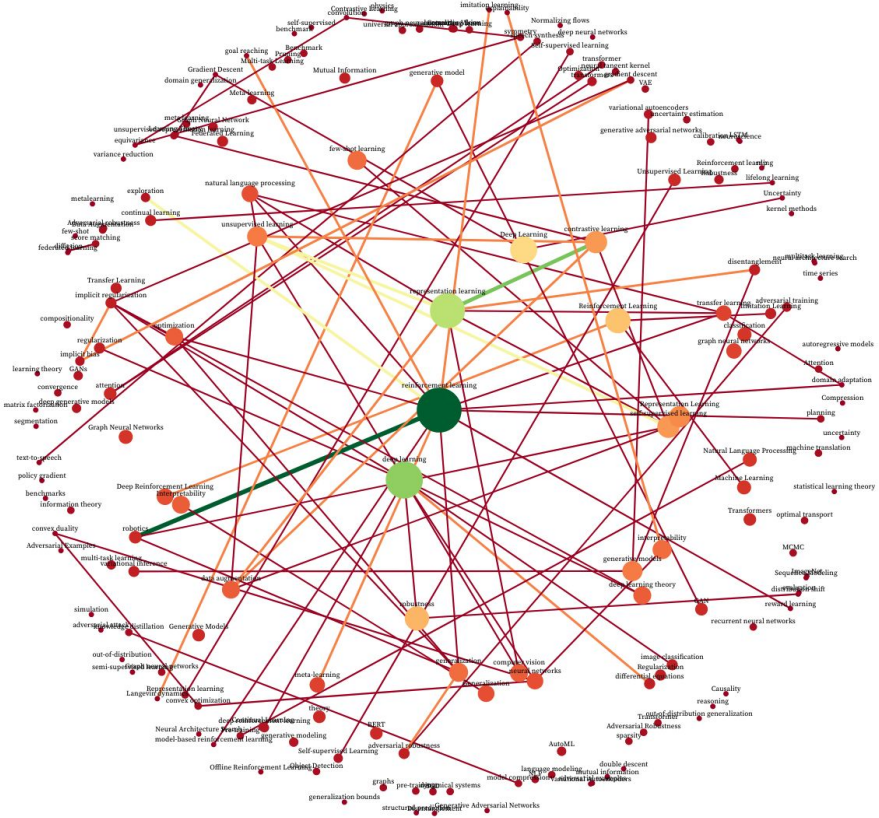
How does the total number of accepted papers at a certain ML conference vary by conference year?



This line graph exactly shows to relationship between number of publications and publication year for the conference ICLR

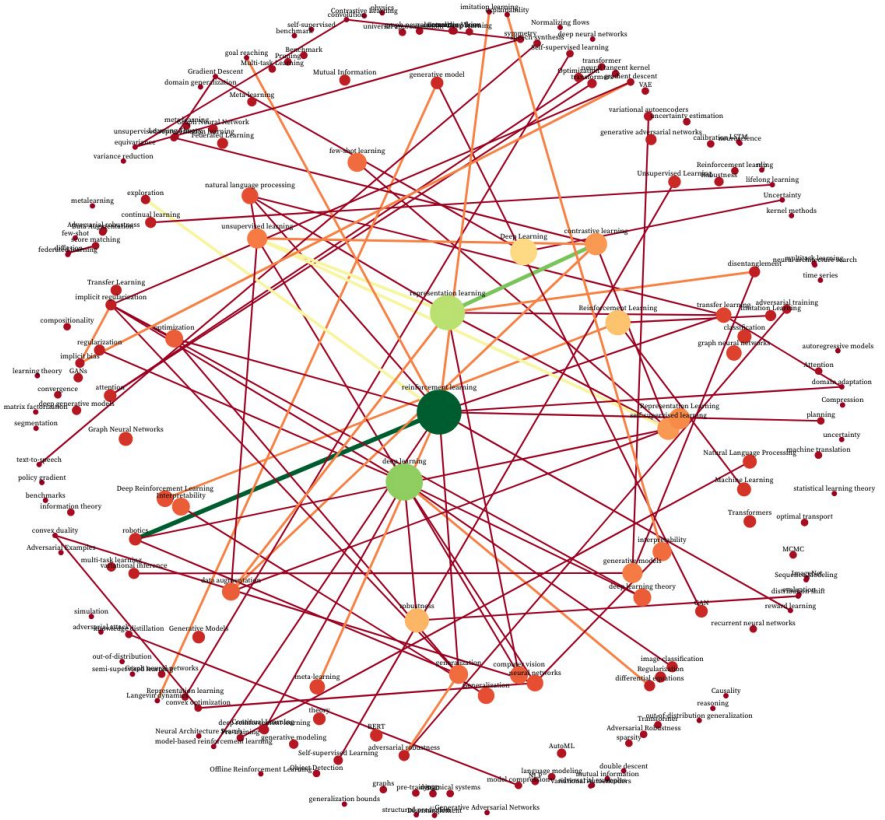
What is the co-occurence rate between different pairs of keywords in a given year?

By looking at the line width and color of the edges between the nodes, the viewers can determine the level of co-occurence between the two keywords at ICLR in 2021



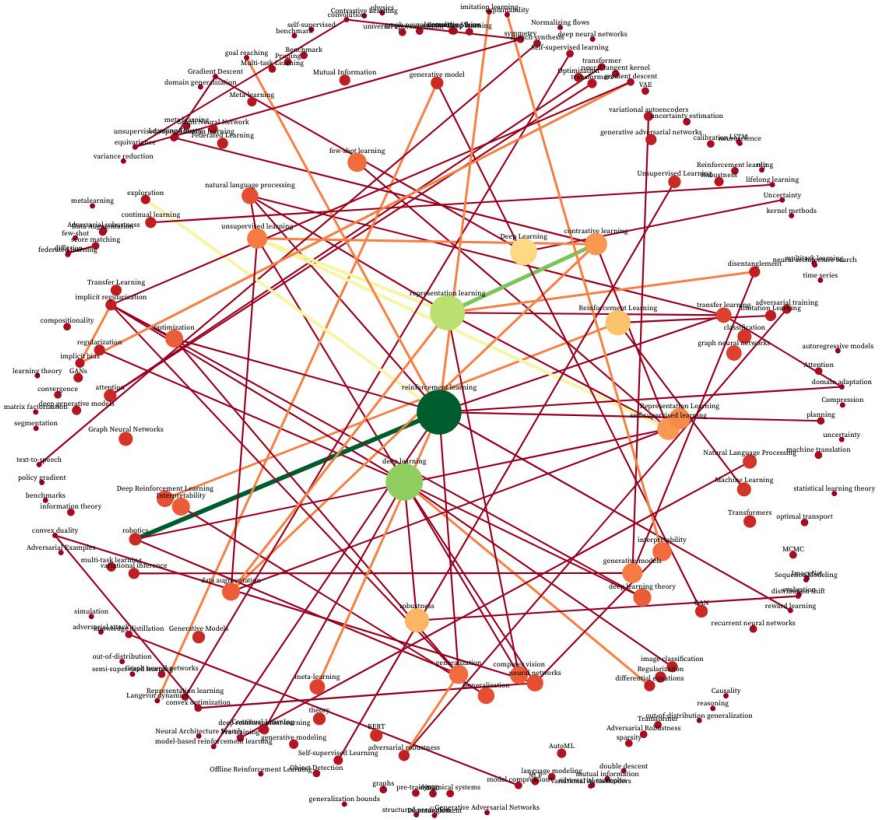
Do keywords with a high level of co-occurrence have a similar number of associated publications?

Many of the largest nodes in the graph have edges (even some bold edges) with much smaller nodes. Therefore the answer to this question is generally no for ICLR in 2021.



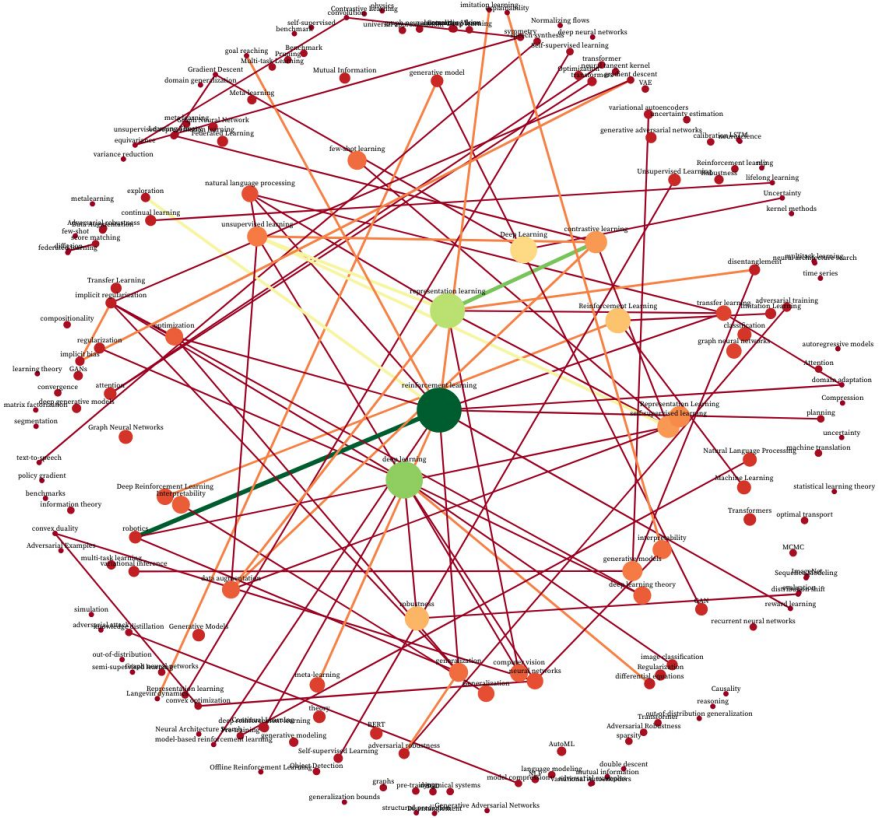
For a given year, what are the top 10 most frequently occurring keywords?

By examining this graph, the user can identify the most frequently occurring keywords by the nodes size, color, and proximity to the center of the graph. For ICLR in 2021, the top nodes are reinforcement learning, representation learning, deep learning, and robustness



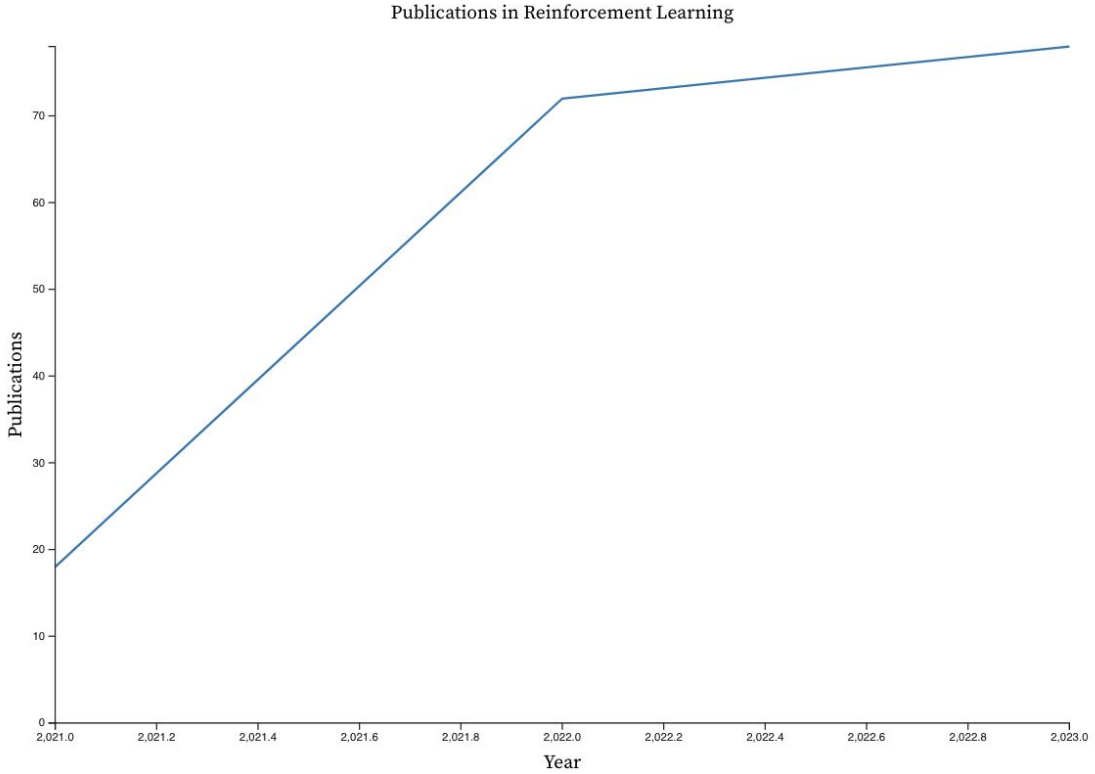
How do the top keywords vary between subsequent conference years?

The user can vary the year depicted in this graph via a slider and observe how the central nodes change. Major changes for a single subfield can be inspected more closely in the line graph view (depicted in next slide) by double clicking on the node.



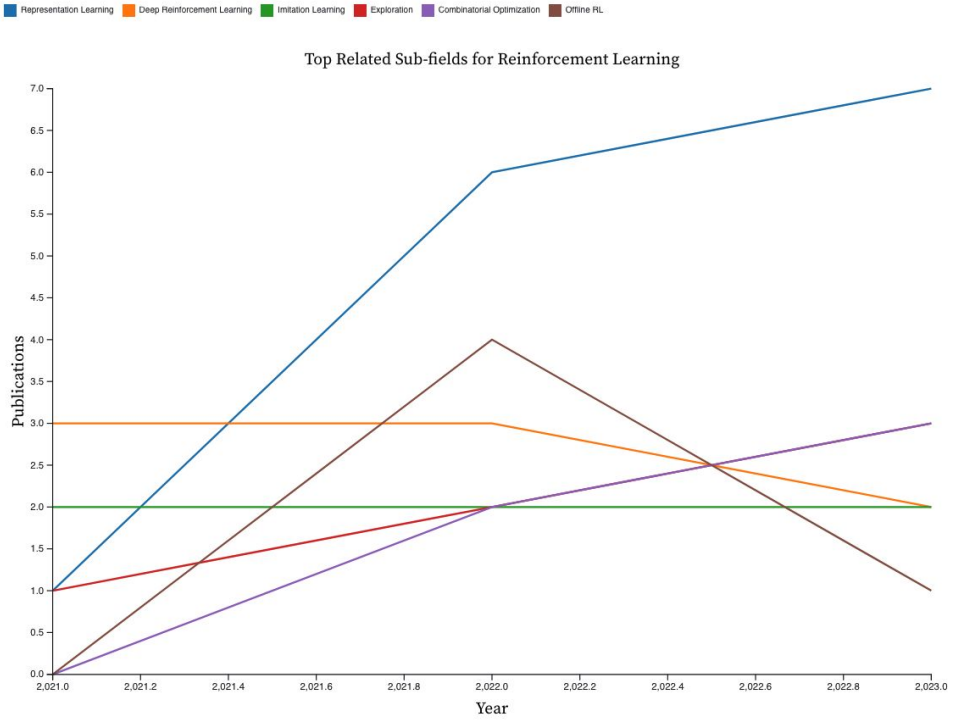
How does the total number of accepted papers containing a specific keyword vary by conference year?

From this plot, the user can see that reinforcement learning experienced rapid growth from 2021-2022 and tapered off after. The user can alter the keyword this plot focuses on via interaction.



How do the co-occurrence relationships between keywords vary by conference year?

This plot depicts how reinforcement learning's relationship with its most related sub-fields evolves over time. For example, we can see that its relationship with Offline RL peaked in 2022 and then fell off. The user can change the focus of this plot via interaction



Future Work

1. Visual refinements
2. Altering node placement method to prevent edge crossing and node overlap as much as possible
3. Adding double click and year slider interactivity

Evaluation Plan

Come up with more specific instantiations of our data questions (i.e. what subfields have the closest relationship with reinforcement learning and how do their relationship with it vary over time?) and attempt to answer them using only our visualization.

Video

https://drive.google.com/file/d/1zzkAiixJjRKPJ2trJRYkQIqRgY_Nfcx3/view?usp=sharing