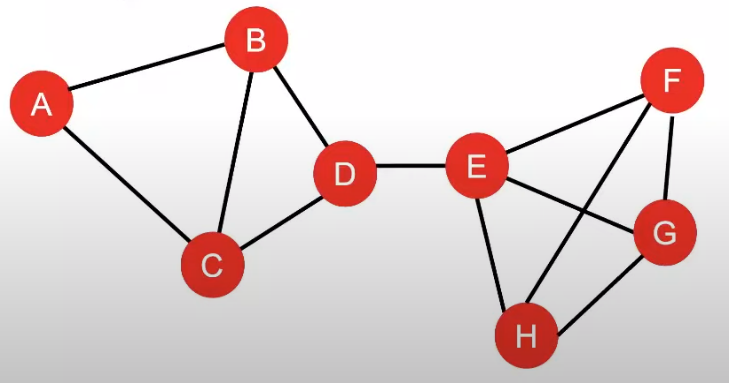
Graph Level Features and Graph Kernels

Graph-Level Features

* Goal: We want features that characterize the structure of an entire graph

For example: How do we describe the structure of the graph below:

Background: Kernel Methods:

* Kernel methods are widely used for traditional ML for graph-level prediction.
* Idea: Design kernels instead of feature vectors
* A quick introduction to Kernels:
* Kernel K(G, G’) measures similarity b/w data
* Kernel matrix K = (K(G, G’)) must always be positive semidefinite (i.e., has positive eigenvals)
* Once the kernel is defined, off-the-shelf ML model, such as kernel SVM, can be used to make predicitons

Graph Kernels: Measure similarity between two graphs:

1. Graphlet kernel
2. Weisfeiler-Lehman kernel
3. Other kernels are also proposed in the literature:

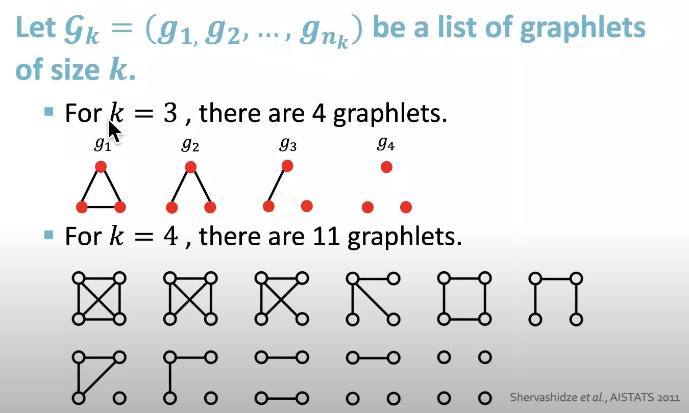
* Random-walk kernel
* Shortes-path graph kernel

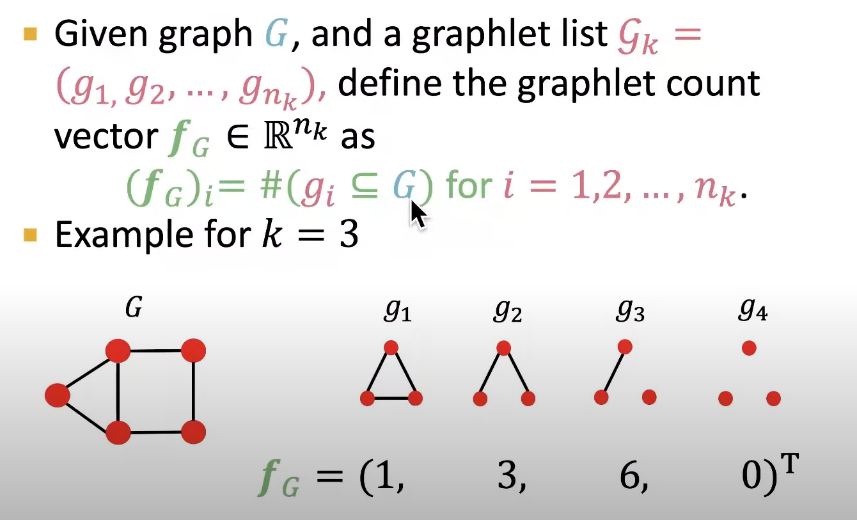
Graphlet Kernel:

Key idea --- Count the number of graphlets in a graph

Note: Definitions of graphlet here is slightly different from node-level features

The two differences are:

* Nodes in graphlets here do not need to be connected (allows for isolated notes)
* The graphlets here are not rooted

Graphlet Features:

A screenshot of a math problem

Description automatically generatedGraphlet Kernel:

Introduction to the graphlet kernel

The graphlet kernel is a method used in machine learning to compare different graphs. In order to compare 2 graphs, the graphlet kernel looks at a smaller structures within each graph, known as graphlet. These graphlets are effectively small graphs themselves, and can range from 2 node, one-edge pairs, to much larger structures. By comparing the similarity and frequency of these graphlets within the two graphs, the graphlet kernel can asses how similar the larger graphs are to each other.

Applications of the graph Kernel

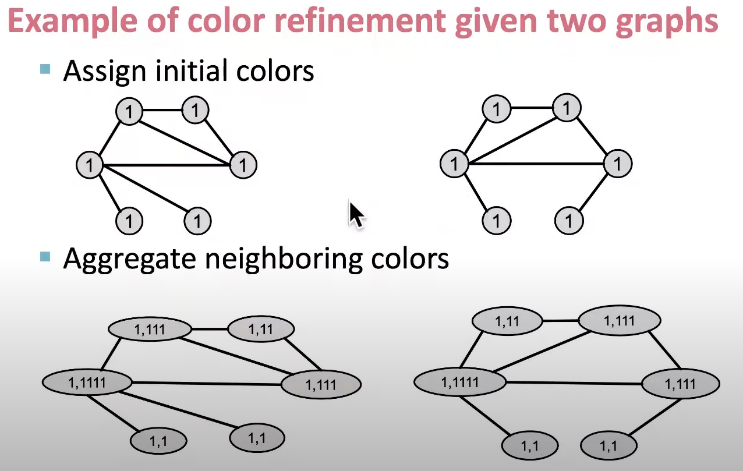
The graphlet kernel has a wide variety of applications, particularly in bioinformatics and social network analysis. In bioinformatics, it can be used to compare different molecular structures, helping to identify similar compounds or predict the properties of new ones. In social network analysis, the graphlet kernel can be used to identify similar patterns of relationship, which can be useful in understanding the structure an dynamics of social groups.

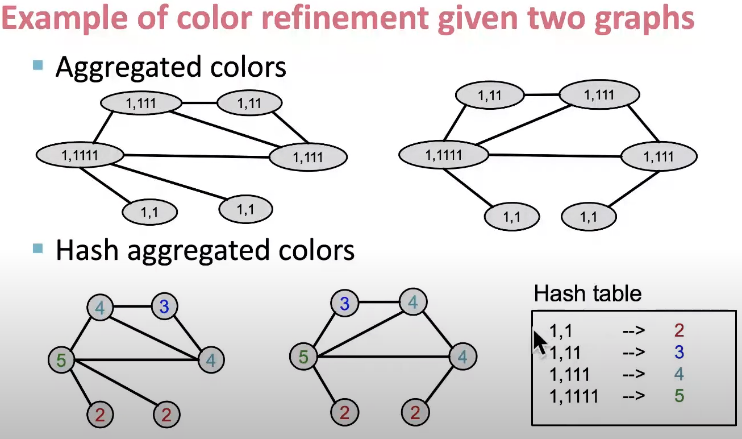
Benefits of Using Graphlet kernel

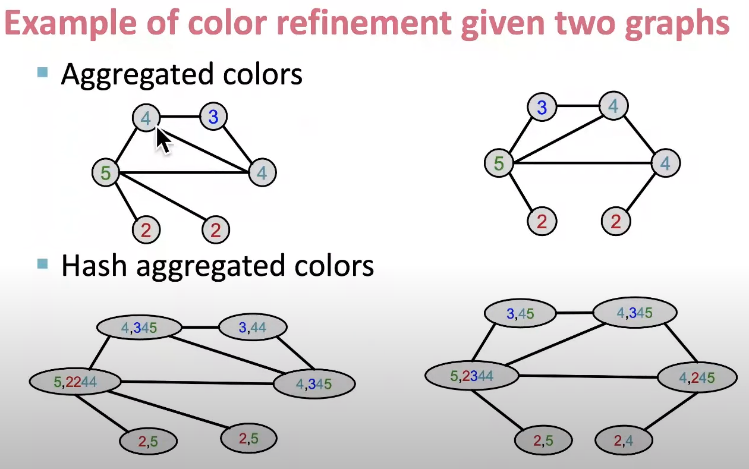
There are several benefits of using Graphlet Kernel:

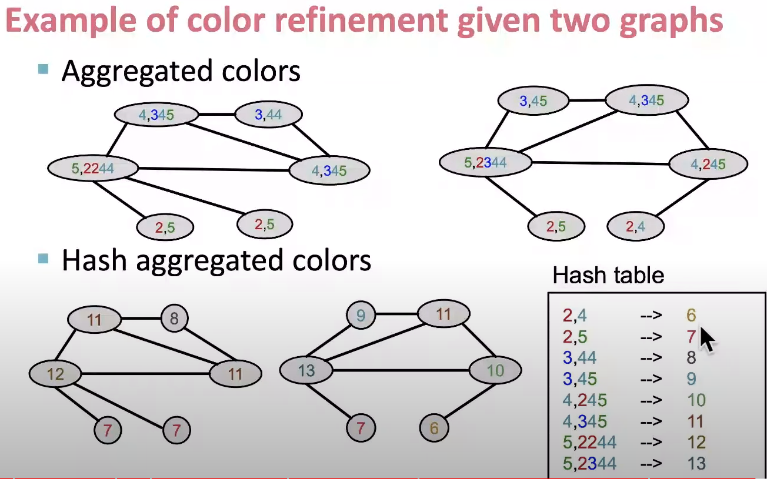
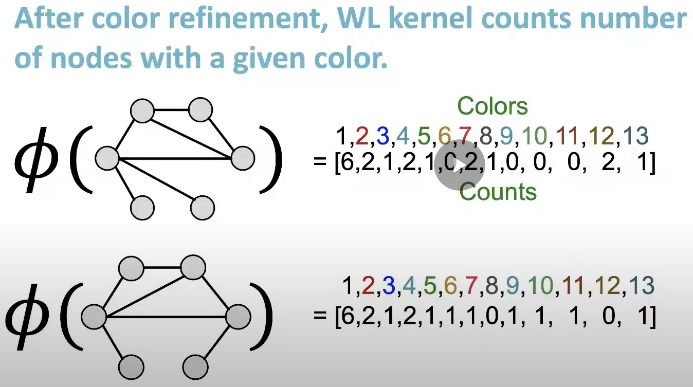
1. It provides a robust and flexible method for comparing graphs, which can be much more useful than simply comparing nodes and edges.
2. It can handle graphs of different sizes and structures, making it widely applicable. Finally, by focusing on smaller substructures, the graphlet kernel can identify similarities that might be missed by other method.

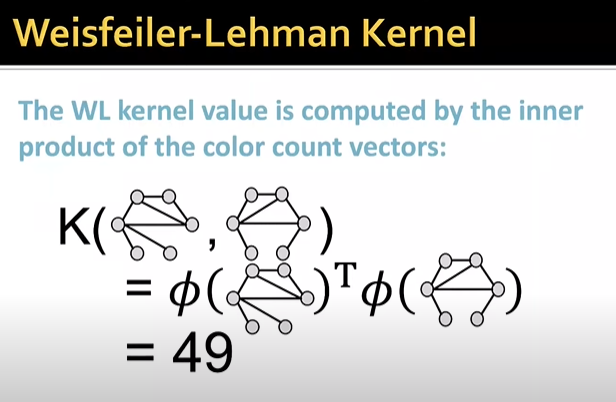
Weisfeiler-Lehman kernel

****Color Refinement

****

****

****

****

**Weisfeiler-Lehman Kernel**

WL kernel is computationally efficient

The time complexity of color refinement at each step is linear in (#edges), since it involves aggregation neighboring colors.

When computing a kernel value, any colors appeared in the two graphs need to be tracked.

Thus #(colors) is at most the total number of nodes.

Counting colors takes linear-time w.r.t #(nodes).

In total, time complexity is linear in #(edges)