

Enhanced pathogen detection through metagenome assembly reference optimization.

Kianna Hales
Sponsor: Microsoft Research | Project Premonition | Jonathan Carlson
Advisor: Jevin West

Project Premonition aims to detect pathogens before they cause outbreaks.



Collect mosquitos with robots



Sequence DNA in the sample



Detect pathogens with metagenomics

Key Words

Metagenome	a genome sample containing the genomes of multiple organisms
Assembly	identifying the individual genomes represented in a metagenome
Reference	known genomes used for assembly & stored in a digital library

Data

1.4x10¹² base pairs (bp)
600,000 genomes

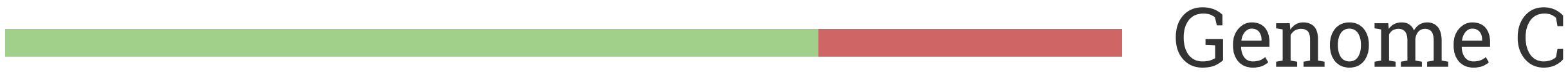
Human genome size (bp)

All reference genomes size (bp)

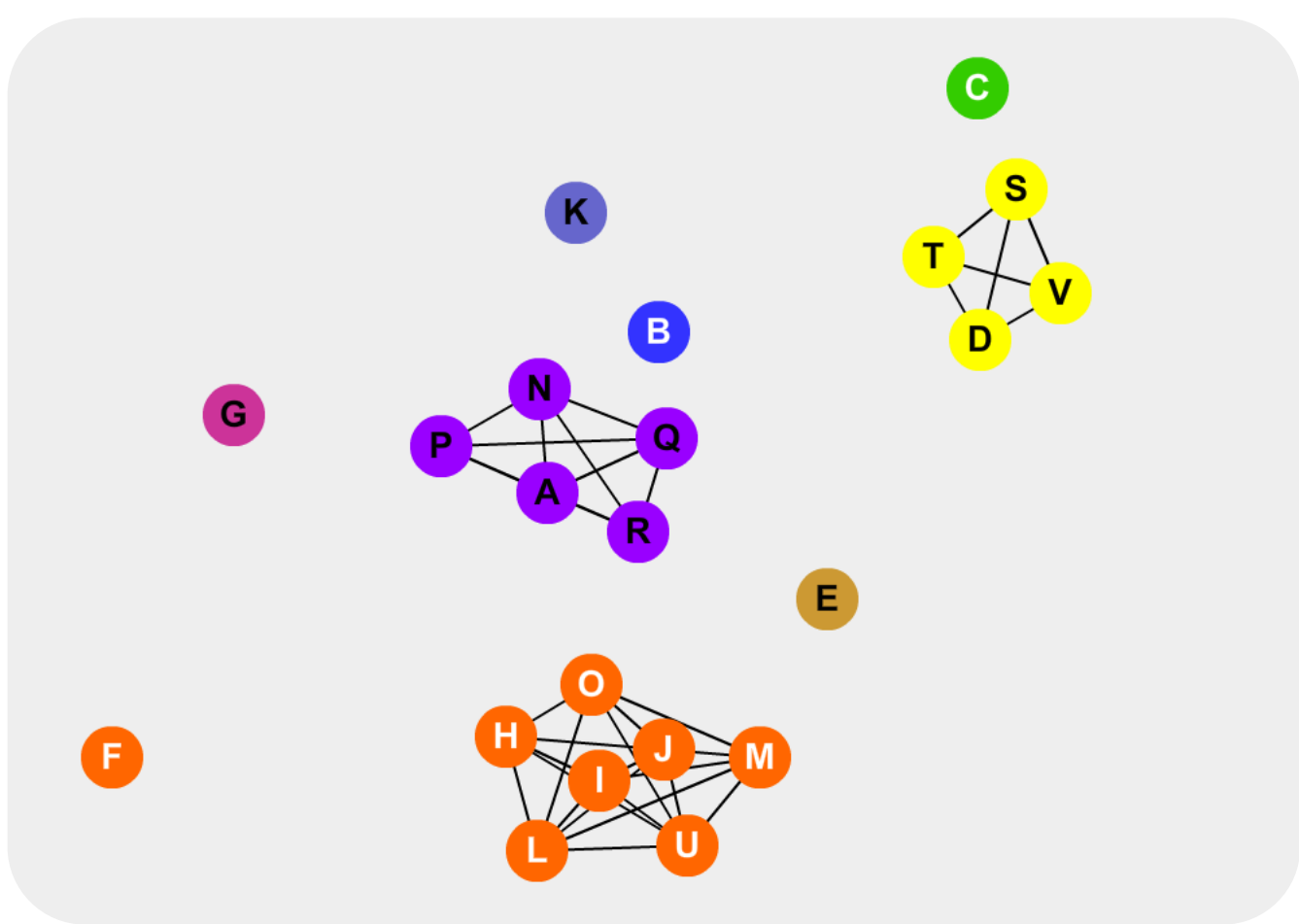
Redundancy



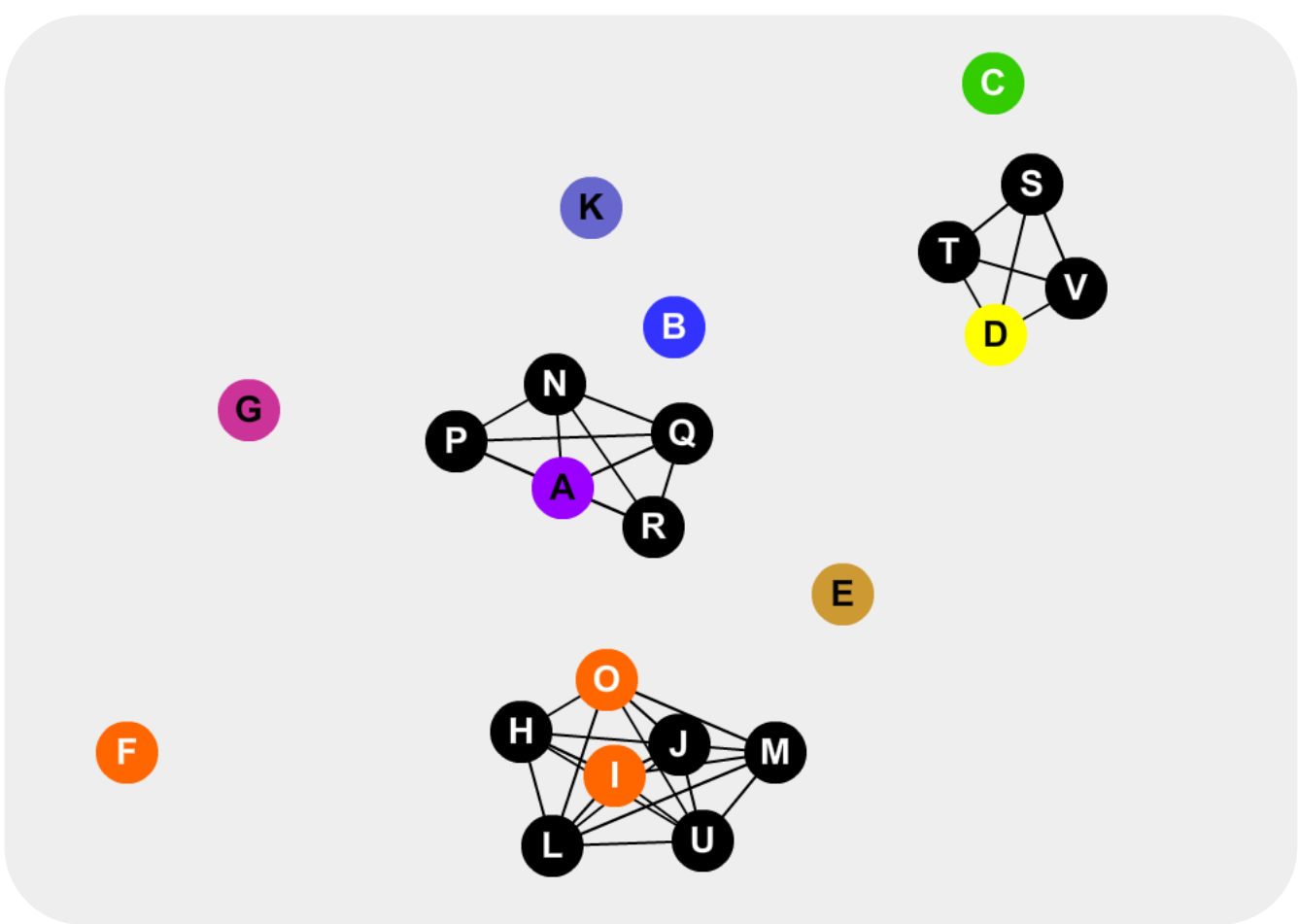
Contamination



Optimization



Remove
Redundancy



Remove
Contamination

