

## **Functional**

See Phase 1 API design and meeting recording with Jonathan.

## **Technical**

Currently there are two modes of plagiarism checking in SSID depending on whether or not useFingerprints option is enabled. When the useFingerprints option is enabled, the submissions are passed through the winnowing fingerprinting algorithm to eliminate unlikely matches to reduce the needed number of pairwise comparisons before the submissions are put into the original engine.

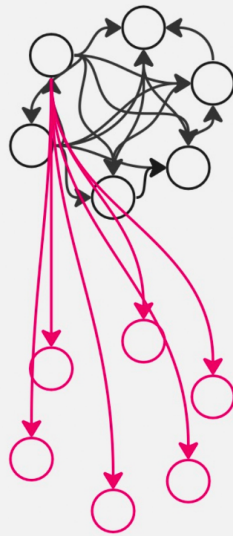
As such, in terms of changes needed for the engine for the reference-checking function, suppose we have  $N$  submissions of current semester and  $M$  submissions of the past semesters

- § If useFingerprints is disabled: We can compare  $N$  submissions with  $M$  submissions but don't compare within  $M$  submissions.

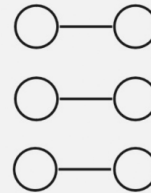
- § If useFingerprints is enabled (winnowing): Since it is a black box, we can probably only put all  $N + M$  submissions into the engine

## Without useFingerprints

Current Semester



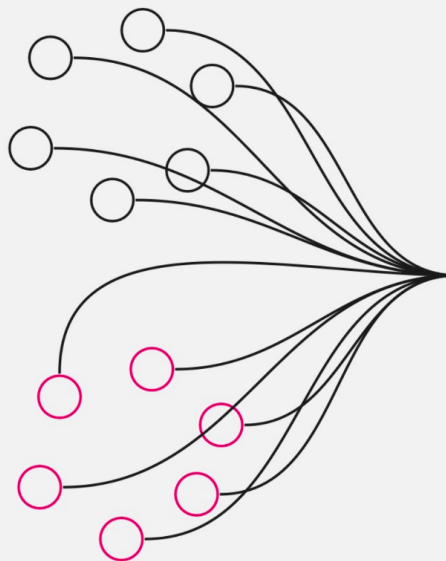
Suspicious pairs



Past Semester (as reference)

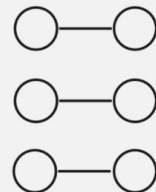
## With useFingerprints

Current Semester



Black box  
 $O(N^2)$  with smaller  
constant factor,  
where  $N$  is number  
of submissions

Suspicious pairs



Past Semester (as reference)