Matthew Baker

Zhi Ying Chen

Bryan Persaud

<https://www.aminer.org/citation>

In order to combine text processing and graph networks techniques we will explore a citation dataset. The data set contains text abstract, authors, year, venue, and title for each paper in the dataset.

The text processing will be performed on the abstract text (and maybe title) to extract some sort of information like topic identification. We can tag topics and possibly cluster them (k-means).

Bipartite projection can be used to cluster topics or show topic connections or other info extracted from the text processing. The text processing can be added to the graph to weight edges (e.g. how many times a topic is shared on a single paper, etc).