# Summary and Problem Definition

The intent of this research is to perform analyses on bibliographic metadata, including measures of centrality, community detection, and referential-relations. The topical data was gathered by searching the Scopus resource for the following keywords: "social network, surveillance, disease control." The search returned 928 articles between the years 1983-2020.

Discussion of the Research Design, Measurement and Network Methods Employed

The collection of 928 articles included various metadata including references cited by each of those articles. These references were used to build edges (connections) from articles returned from the search to their references, initially as directed then decomposed for two further reviews. The "co-citation" graph searches for pairs or groups of publications which are both referenced by the same publication (indegrees). The "coupling" graph identifies groupings that make reference to the same source (outdegrees).
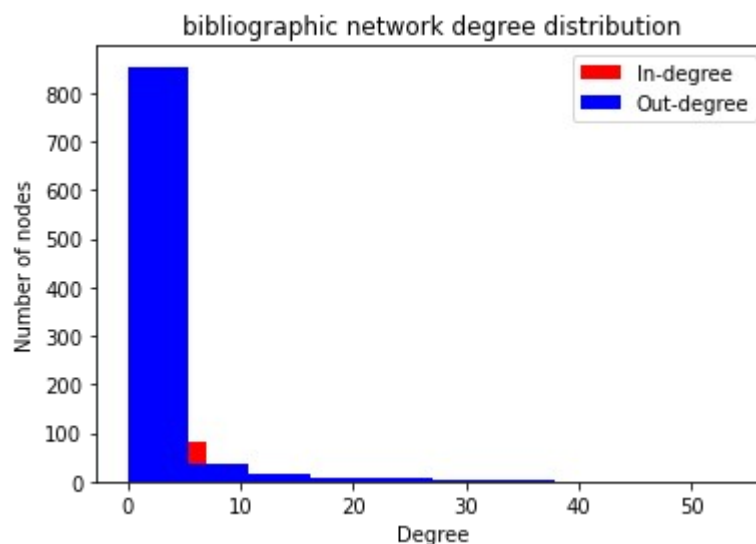
Barbasi & Albert (1999) recognized that "new vertices attach preferentially to sites that are already well connected", which is readily apparent in the subject network.

For each of the graphs mentioned, the following centrality measures were calculated: closeness, betweenness, eigenvector, Katz, PageRank, load, and current flow. The following community algorithms were utlized for detection: Girvan-Newman, Asynchronous Label Propagation, Louvain, Label Propagation, and Fluid.

Overview of Programming Work

All of these analyses were performed using python, with graph algorithms provided by the NetworkX. The dataset was manually downloaded from Elsevier's Scopus database (Scopus, n.d.). This provided a comma delimited set of publications with metadata. The metadata was then loaded into a dataframe for inspection & determination of necessary parsing. Functions were developed for generating graph visualizations using holoviews and NetworkX. "NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks." (Hagberg et al., 2008) Further functions were written using the various methods available within NetworkX to support community detection, centrality measures on nodes, and other network-wide statistics such as density, transitivity, and clustering.

Review of Results

Each of the three methodologies revealed one major cluster and two smaller but still significant clusters of referential relationship. This pattern is likely to continue and potentially increase in the future as more widely recognized publication increase their "virality", similarly to the compounding results of pagerank. The communities detected across the three networks were fairly consistent between three (Girvan-Newman) and thirty-one (Asynchronous Label Propagation) communities.

References

Barabasi, A.-L. & Albert, R. (1999) Science 286, 509–512.

Scopus. (n.d.). Retrieved March 18, 2021, from https://www.scopus.com/

Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008