

This analysis is proposed to continue the focus on NetworkX application to parse through the academic citational dataset that was extracted from Scopus Document Search website and take a look at citations on the selected topic from the network analysis standpoint.

“Deep Learning” is the key word used for document search from last 30 days because it is a prevalent and quite active academic field that can assure dataset extracted to satisfy our requirement in terms of size and recency. 2,000 documents out of 8,518 returned search results were extracted initially and fed into the process and returned more than 2,500 nodes and 3,800 edges. To reduce computation complexity, the dataset was trimmed down to 1,001.

Five attributes including Author, Article Title, Year, Source Title, and Include Reference were extracted. Title and References are used to identify the nodes with directional edges defined by citational relationship. Three types of relationships have been explored with graph analysis based on NetworkX.

Citation Network Overview. There are ultimately 1,432 nodes and 2,021 edges identified and generated in the network of citation which is generated based on 1,001 academic documents

Network Graph of Citation on Deep Learning

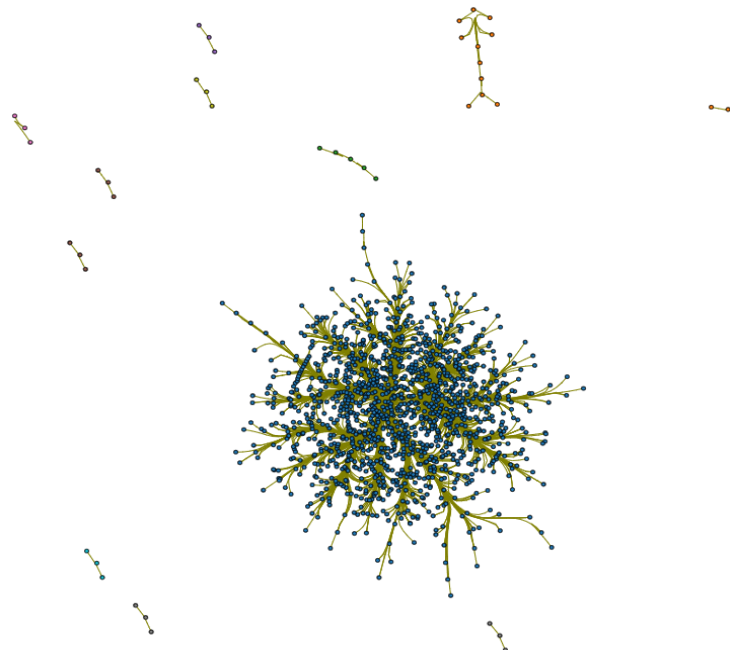


Figure 1. Citational Network Graph - Deep Learning

25 Giran-Newman communities have been detected with one large (community 0) community involving the majority of documents while the rest of 24 concentrating on niche study topics as shown in Figure 1. Also, the distribution of citation numbers was investigated. The in-degree values of nodes have been sorted and plotted into a histogram graph as shown as Figure 2, which clearly demonstrates skewed distribution to the right with a long tail, aligning with Price's theory articulated in Newman's book. The paper that's cited most has been cited in total 39 times.

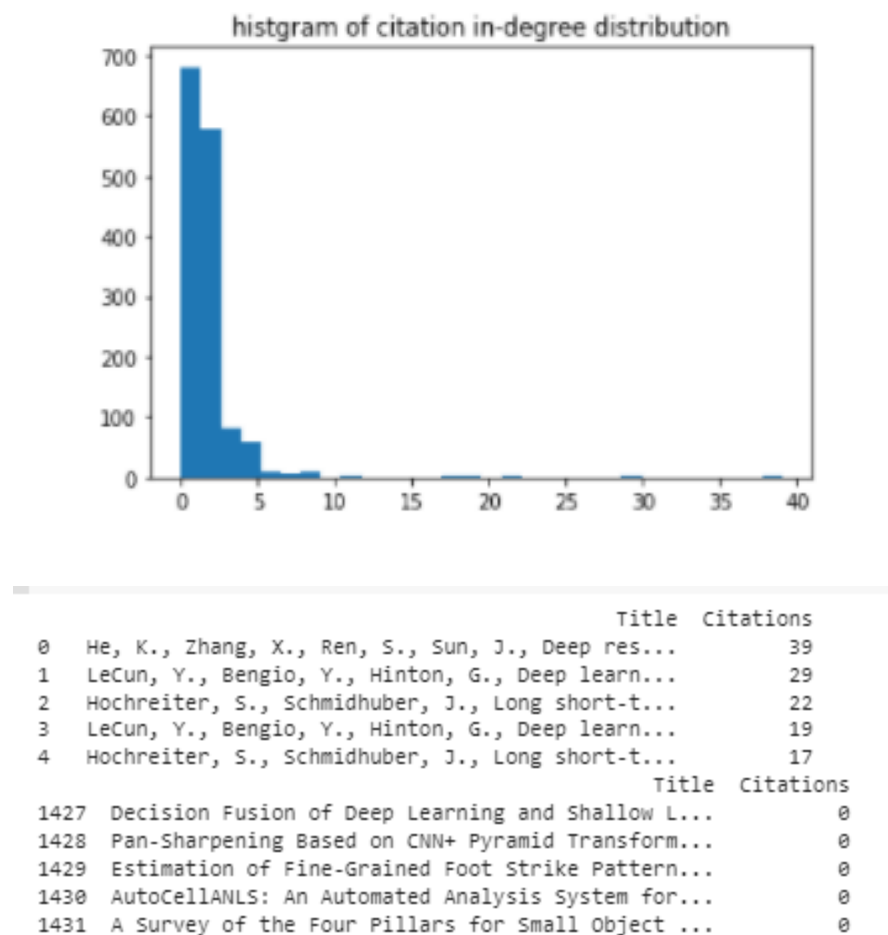


Figure 2. Distribution of In-Degrees of Nodes

Co-Citation Network. This phase focuses the lens on documents that are being cited by multiple other documents jointly and the relationship in-between. There are 738 nodes and 3,487 edges detected, which is plotted as shown in Figure 3. 8 Girvan-Newman communities are identified. We can tell the majority of them are fully-connected, meaning cross-referenced in the largest community with several sparse small ones. This

probably partially indicates that the papers that have been cited more are closely interconnected as they cross-referenced to each other.



Figure 3. Co-Citation Network Graph

Bibliographic Coupling Network. Another view being explored is the bibliographic coupling which focuses on documents that cite other documents and their interconnections. There are 679 nodes and 3,661 edges detected, which is plotted as shown in Figure 4. 24 Girvan-Newman communities are identified.

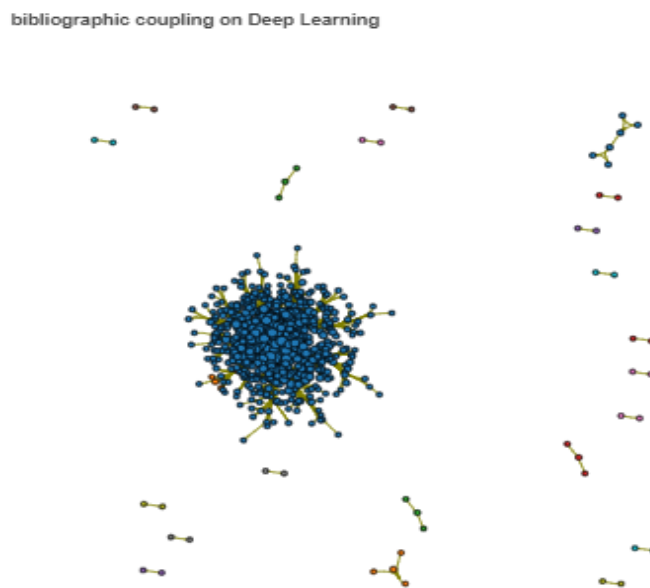


Figure 4. Bibliographic Coupling Network