Several datasets provide comprehensive information on scholarly publications, including mutual citations and associated metadata. Here are some notable examples:

1. \*\*DBLP Citation Network Dataset\*\*: This dataset encompasses 629,814 papers and 632,752 citation relationships. Each paper is detailed with attributes such as abstract, authors, publication year, venue, and title. It's particularly useful for clustering analyses incorporating network structures and supplementary information. https://paperswithcode.com/dataset/dblp?utm\_source=chatgpt.com

2. \*\*CiteSeer Citation Network\*\*: Extracted from the CiteSeer digital library, this dataset represents publications as nodes and citations as directed edges. It facilitates studies on citation dynamics and scholarly influence. https://networkrepository.com/citeseer.php?utm\_source=chatgpt.com

3. \*\*Cora Dataset\*\*: Comprising 2,708 scientific publications categorized into seven classes, the Cora dataset includes a citation network with 5,429 links. Each publication is represented by a binary word vector, indicating the presence or absence of specific words from a 1,433-word dictionary. This dataset is widely used for tasks like node classification and clustering. citeturn0search4

4. \*\*unarXive Dataset\*\*: Derived from arXiv's LaTeX sources spanning 1991 to July 2020, unarXive offers one million papers in plain text, 63 million citation contexts, and a citation network comprising 16 million connections. The dataset's high-quality extraction process ensures accurate full-text and citation data, making it valuable for natural language processing applications.

https://paperswithcode.com/dataset/unarxive?utm\_source=chatgpt.com

5. \*\*ACM Citation Network\*\*: This dataset includes metadata for over 2 million articles published by the Association for Computing Machinery (ACM), detailing reference lists for each article. It's instrumental for analyzing citation patterns within computer science research. citeturn0search9

6. https://paperswithcode.com/dataset/pubmed

These datasets are instrumental for research in bibliometrics, citation analysis, and the study of scholarly communication patterns.