**High-energy physics theory citation network**

**Dataset information**

Arxiv HEP-TH (high energy physics theory) citation graph is from the e-print [arXiv](http://arxiv.org/) and covers all the citations within a dataset of 27,770 papers with 352,807 edges. If a paper *i* cites paper *j*, the graph contains a directed edge from *i* to *j*. If a paper cites, or is cited by, a paper outside the dataset, the graph does not contain any information about this.

The data covers papers in the period from January 1993 to April 2003 (124 months). It begins within a few months of the inception of the arXiv, and thus represents essentially the complete history of its HEP-TH section.

The data was originally released as a part of [2003 KDD Cup](http://www.cs.cornell.edu/projects/kddcup/).

|  |  |
| --- | --- |
| Dataset statistics | |
| Nodes | 27770 |
| Edges | 352807 |
| Nodes in largest WCC | 27400 (0.987) |
| Edges in largest WCC | 352542 (0.999) |
| Nodes in largest SCC | 7464 (0.269) |
| Edges in largest SCC | 116268 (0.330) |
| Average clustering coefficient | 0.3120 |
| Number of triangles | 1478735 |
| Fraction of closed triangles | 0.04331 |
| Diameter (longest shortest path) | 13 |
| 90-percentile effective diameter | 5.3 |

**Source (citation)**

* J. Leskovec, J. Kleinberg and C. Faloutsos. [Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations](http://www.cs.cmu.edu/~jure/pubs/powergrowth-kdd05.pdf). ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2005.
* J. Gehrke, P. Ginsparg, J. M. Kleinberg. [Overview of the 2003 KDD Cup](http://www.cs.cornell.edu/home/kleinber/kddcup2003.pdf). SIGKDD Explorations 5(2): 149-151, 2003.

**Files**

|  |  |
| --- | --- |
| File | Description |
| [cit-HepTh.txt.gz](https://snap.stanford.edu/data/cit-HepTh.txt.gz) | Paper citation network of Arxiv High Energy Physics Theory category |
| [cit-HepTh-dates.txt.gz](https://snap.stanford.edu/data/cit-HepTh-dates.txt.gz) | Time of nodes (paper submission time to Arxiv) |
| [cit-HepTh-abstracts.tar.gz](https://snap.stanford.edu/data/cit-HepTh-abstracts.tar.gz) | Paper meta information (see below) |

**Example of paper meta information**

Paper: hep-th/0002031

From: Maulik K. Parikh

Date: Fri, 4 Feb 2000 17:04:51 GMT (10kb)

Title: Confinement and the AdS/CFT Correspondence

Authors: D. S. Berman and Maulik K. Parikh

Comments: 12 pages, 1 figure, RevTeX

Report-no: SPIN-1999/25, UG-1999/42

Journal-ref: Phys.Lett. B483 (2000) 271-276

\\

We study the thermodynamics of the confined and unconfined phases of

superconformal Yang-Mills in finite volume and at large N using the AdS/CFT

correspondence. We discuss the necessary conditions for a smooth phase

crossover and obtain an N-dependent curve for the phase boundary.

\\