Gleb Lanin

Influence and Citation: A Research Study of Publications in the HEP-TH Archive

In the world of scientific publications, the number of citations to an article is an important indicator of its importance and impact. However, beyond simply counting the number of citations, of interest is also the density of these citations and their relationship to temporal changes in the scientific community. In this article, we examine the dynamics of citations in theoretical physics (HEP-TH) by analysing publication and citation data from 1992 to 2002.

Number of Publications by Year

We started by analysing the number of publications by year to understand general trends in scientific activity. The graph showing the number of publications for each year shows that there is an upward trend, reaching a peak of 5,313 papers in 2001. However, there is a significant drop to 882 papers in 2002, which is due to incomplete data for 2002, as only the first two and a half months of data are available. So we can expect the trend to continue to increase.

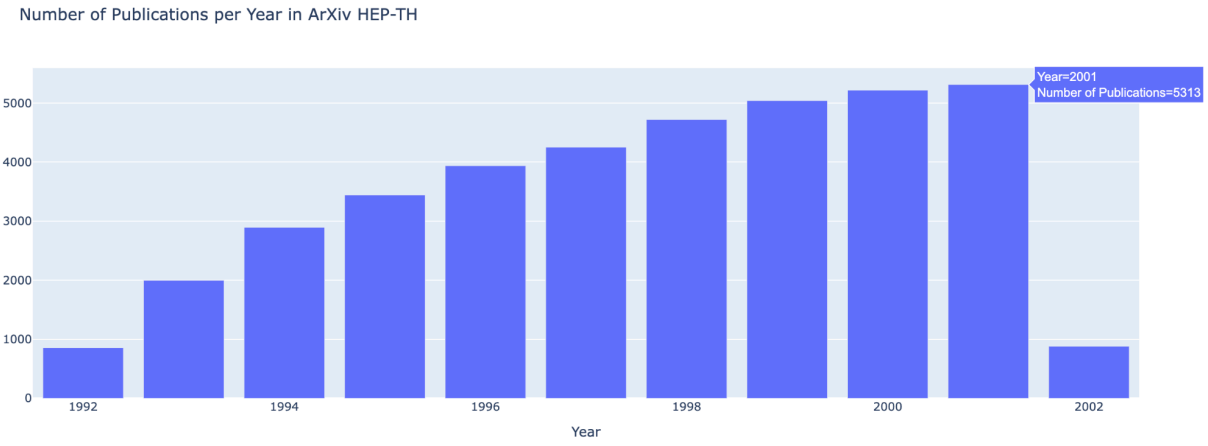

Figure 1: Number of publications by year

Citation density

Next, we investigated how the citation density per article changed by year. In 2002, despite a significant decrease in the number of publications, the citation density remained high at 5.03, slightly higher than the previous year's figure of 4.68. This trend shows that despite fewer new publications, older articles continued to have a strong impact, which may indicate that the scientific trends are stabilising. And also, it suggests that the number of papers per year does not affect the citation density as much.

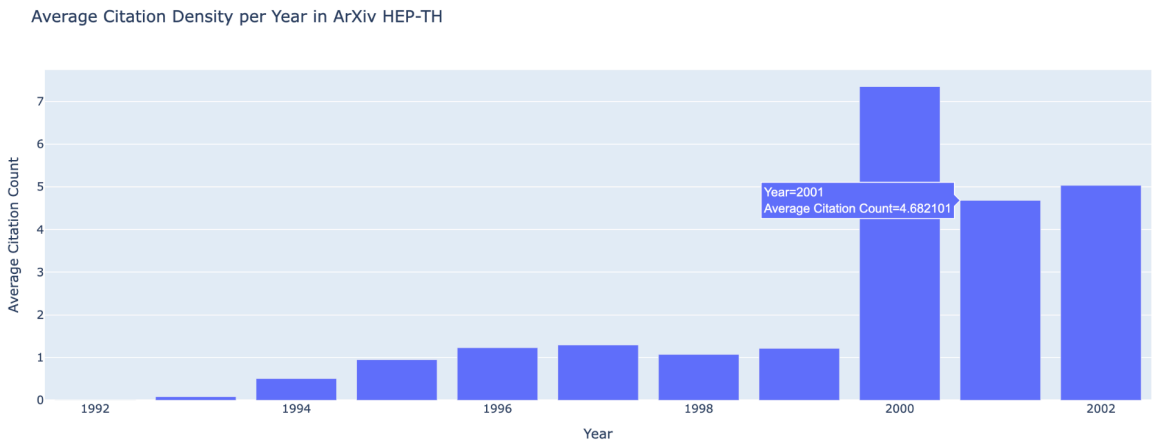

Figure 2: Average citation density by year

Also, between 1999 and 2000 there was a significant jump in citation density, which can be attributed to advances in technology and access to digital libraries. At this time, online archives such as arXiv were already in heavy use, enabling researchers to share knowledge and find key papers more quickly. Improved search capabilities and access to information likely accelerated the pace of integration of new ideas and facilitated the creation of more citable articles, increasing the impact of published work

Top 10 Most Cited Works

When we examined the top 10 most cited papers by year, we paid special attention to 1999 and 2000, when there was a sharp jump in citation density, which could be due to the emergence of key research in theoretical physics. Also, it is interesting to see the difference between the first and last places for each of the 1999 and 2000 top 10s. It can be seen that in 1999 the first places have more citations than in 2000, but at 8-10 places in 1999 the number of citations drops dramatically when 2000 still keeps the bar high.

Top 10 most cited papers in 1999:		Top 10 most cited papers in 2000:	
PaperId	CitationCount	PaperId	CitationCount
9904207	282	112044	242
9903205	251	3160	242
9910219	211	7191	197
9912249	201	9148	196
9905210	168	2091	186
9903225	130	2075	182
9903224	119	3136	179
9909205	111	5031	179
9907209	109	9103	174
9912287	99	5048	159

Table 1: Top 10 most cited papers by year

Centrality on the Web: Degree and Betweenness

To analyse the citation structure in more depth, we applied network analysis methods, calculating degree centrality and betweenness centrality for publications in 2002. Since this year has the least amount of data (which will allow the metrics to be calculated quickly) and a fairly high citation density (which suggests that its structure is the same as previous years' papers)

Degree centrality allows us to assess the importance of a paper based on its connections with other papers. In 2002, the top 10 papers for this metric show that papers linked to multiple other articles had the greatest impact. For example, article 202021 - “Strings in flat space and pp waves from $\mathcal{N} = 4$ Super Yang Mills” by David Berenstein, Juan Maldacena and Horatiu Nastase had the highest degree of centrality (0.0518), indicating its significant presence in the citation network.

Top 10 papers by Degree Centrality in 2002:		
PaperId	**DegreeCentrality**	**BetweennessCentrality**
202021	0.051825	0.000205
201253	0.037938	0.000273
202109	0.032785	0.00038
201081	0.031782	0.000035
203048	0.022334	0.000084
202111	0.020759	0.000078
203101	0.0199	0.000085
202179	0.01947	0.0001
202157	0.016034	0.000036
202186	0.015891	0.000107

Table 2: Top 10 articles by Degree centrality in 2002.

Betweenness centrality assesses how often an article acts as a ‘bridge’ between other articles in the citation network. Article 202127 - “De Sitter Holography and the Cosmic Microwave Background” by Finn Larsen, Jan Pieter van der Schaar, Robert G. Leigh, which ranks first in betweenness centrality (0.00093), can be considered an important bridge between different strands in a research network, providing a flow of information between multiple other papers. Article 202109 - “Exactly solvable model of superstring in Ramond-Ramond plane wave background” by R.R. Metsaev and A.A. Tseytlin also appeared in both top 10s, emphasising its importance for both direct citations and indirect influence in a research network.

Top 10 papers by Betweenness Centrality in 2002:		
PaperId	**DegreeCentrality**	**BetweennessCentrality**
202127	0.006872	0.00093
203018	0.011739	0.000826
202048	0.005297	0.000818
201004	0.011167	0.000763
201215	0.006013	0.000707
201174	0.006729	0.000659
203019	0.010594	0.000655
202124	0.009592	0.000545
202109	0.032785	0.00038
203082	0.013744	0.000341

Table 3: Top 10 articles on Betweenness Centrality in 2002.

Founding Articles: Top 10 Founders

The final chord of our research was the rubric that we decided to introduce in the style of Rolling Stone magazine. We could not help but pay attention to the founding articles, which played the role of progenitors in the scientific environment, despite their low number of primary citations. Article **110055** - “Strings, Branes and Extra Dimensions” by Stefan Forste released in 2001, although it had only 5 primary citations, had an indirect impact on 3080 papers. This is akin to top articles or albums on music charts that may not become bestsellers but will influence generations of researchers. Article **106178** - “Nonabelian D-branes and Noncommutative Geometry” by Robert C. Myers is also interesting: despite only 4 primary citations, it has influenced over 1600 other papers. These articles play the role of not outright bestsellers of scientific thought, but provide the foundation for the development of entire scientific fields.

PaperId	Year	Initial Citations	Indirect Citation
110055	2001	5	3080
7170	2000	78	2246
101126	2001	60	2034
203048	2002	14	2013
106048	2001	220	1858
201253	2002	39	1852
106178	2001	4	1667
8241	2000	39	1642
108172	2001	36	1586
7195	2000	10	1538

Table 4: Top 10 progenitors and their citations

Conclusion

In conclusion, our results show that the number of citations does not always directly correlate with the impact of an article. Some papers remain little known in the early years, but then become major ‘pillars’ of the scientific discipline, and through secondary citations their influence continues to grow. It is important to remember that scientific citation is not only about direct citations, but also about the deep connection between research, which manifests itself through the network of citations and the construction of scientific ideas.

All the code can be found here:
https://github.com/Booshlya/Study-of-Publications-in-the-HEP-TH-Archive