journals in languages other than English, or with specific foci. The scientific quality of journals is always evolving; a change in the reviewing process, for example, can lead to an improvement of a journal. Thus any grading should retain the possibility of updating and evolution; the grading proposed here is presented as our best attempt at assessing the current situation.

References

Arnold, D. N., & Fowler, K.K. (2011). Nefarious numbers. *Notices of the AMS* 58 (3), 434–437.

Towns, M.H., & Kraft, A. (2011). The 2010 Rankings of Chemical Education and Science Education Journals by Faculty Engaged in Chemical Education Research. *Journal of Chemical Education* 2012, 89 (1), 16–20.

Time Lag in Mathematical References

Thierry Bouche (Université de Grenoble, Saint-Martin d'Hères, France), Olaf Teschke (FIZ Karlsruhe, Berlin, Germany) and Krzyś Wojciechowski (University of Warsaw, Poland)

Results in mathematics do not just hold forever – we are also conscious of them for a long time. The Pythagorean Theorem still taught in schools may be an extreme example but also, in our specific research fields, specialists are usually aware of long-lasting conjectures which have influenced centuries of research or seemingly dead areas revived when looked upon by a new generation from a different angle.

Unfortunately, this very fascinating attribute of mathematical research has turned out to be a handicap in the scientometric age. When measures like impact factors came into use in the second half of the 20th century, they were initially limited to very recent data for very practical reasons: critical masses of references were generally not yet available or manageable for longer periods of time. As an effect, the computation of the usual impact factor is restricted to two years (some extensions go to the limit of at most five years). But what is lost? This leads to the natural question of the time lag for mathematics references, i.e. the average difference of the publication year of the citing and the cited article.

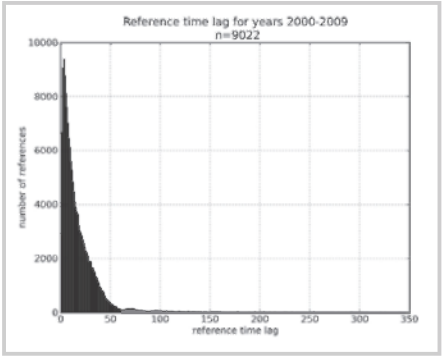
Surprisingly, this is not a question that is easy to answer. In an earlier issue, Rui Loja Fernandes[1] gave a good illustration in the example of the evaluation of the IST: Table 5 on p. 16 shows that both the aggregate cited and citing half-life in mathematics is longer than ten years on the sample of mathematics reflected in the ISI data. The recent progress of digital libraries over the last two decades contributes hitherto unavailable data to the pool and allows at least some glimpses into how mathematical information is alive through the decades.

Here, we start from two data sets which became only recently available: references extracted during the development of the European Digital Mathematics Library (EuDML)[2,3] and the set of references currently stored and identified in the zbMATH database.

Both sets are not ad hoc comparable: EuDML represents, with about 200,000 entries, a fraction of the literature now available in the public domain. For 44,817 articles (greater than 20%), references are now extracted from the digital full text but the data comes with the natural inaccuracies of automatic processing. Naturally, most of the cited articles are not contained in EuDML itself. On the other hand, the (about) 166,305 articles with references in zbMATH are a considerably smaller fraction of the total, greater than 3.2 million, items (just greater than 5%) but come along with the advantage of being mostly matched against the database, hence providing more accurate information.

However, there are also striking similarities: for both data sets, there are virtually no citing articles before 1890 and only very scarce cited articles before 1850. There is no surprise here. By now, references must obey at least rudimentary patterns (author, title, source, publication year) to be detected. This only came along with the appearance of scientific journals in larger numbers. (Note that the volumes of Euclid's *Elements* are rarely cited with their precise publication years).

Hence, when ignoring sparse data, we have a 120×140 years citation matrix, with zeroes below the obvious diagonal. (Actually, this is not so obvious in the case of zbMATH data: contrary to EuDML, there are several cases of negative time lag, which comes from the fact that "submitted/to appear"-articles in the references were identified in their final version in the zbMATH database, with a possibly delayed publication year. But this pertains to less than 0.1% of the references.) For easier representation, these data were grouped by decade of the citing articles; the typical picture can be seen below.



Publication years of references of EuDML articles published 2000–2009; typically, the numbers can be approximated by a power law distribution.

1 Evaluation of Faculty at IST – a Case Study. EMS Newsletter 84, 13–17.

2 http://eudml.org.

3 EMS Newsletter 76, June 2010, 11–16; ibid. 85, 57–58.

The following table shows both the average time lag and the median for EuDML and zbMATH references for the decades since 1890.

	EuDML avg	zBMATH avg	EuDML median	zBMATH median
1890–1899	*	9.88	*	8
1900–1909	*	9.07	*	7
1910–1919	*	10.07	*	7
1920–1929	*	13.07	*	10
1930–1939	*	11.13	*	8
1940–1949	16.39	15.74	12	12
1950–1959	11.46	12.66	8	8
1960–1969	10.39	11.14	7	8
1970–1979	10.10	10.80	7	8
1980–1989	12.40	12.18	9	10
1990–1999	14.99	13.28	11	10
2000–2009	17.41	14.64	12	11

Average and median time lag in mathematical reference per decade. *No figures derived from EuDML because only scarce data available.

One observation is a larger difference between EuDML and zbMATH data before 1930 and after 2000. The effect in the early decades is simply due to the fact that EuDML has very few articles with references in this period. Furthermore, the section of EuDML articles after 2000 is quite different from the earlier corpus because there are fewer sources which provide articles that recent.

When we exclude this, maybe the most striking observation is that both data sets show very similar patterns, with the main influences seeming not to be related to developments in mathematics but just to the two World Wars! The two local maxima in the '20s and '40s in the more comprehensive zbMATH data seem to be directly linked to the fact that there are many references bridging the war gaps. Consequently, the time lag is reduced in the following decades, reaching a minimum in the '70s. Interestingly enough, the time lag starts growing again after then, with no indication that the faster availability of information or the acceleration of academic activities in the last decades had any chance in stopping this process (on the contrary, the enhanced accessibility of older articles through digital libraries may actually have the effect of an increase of the citations with larger time lag). The most likely explanation for the minimum in the '70s is just that there were simply "not enough older papers there" to be cited due to wartime – an effect that is now slowly dissolving. It is still unclear when the growth of average time lag may come to an end but if we accept that the World War II gap came into full effect in time lag only three decades later, it may well be expected that the average time lag stabilises beyond 20 years.

Another question pertains to the small but visible differences in time lag for EuDML and zbMATH data for recent decades, where both services have a comparable magnitude of data and one would expect convergence. The higher time lag (by about 1.5 years over the last two decades) in EuDML seems to be influenced by two effects: firstly, zbMATH identifies more recent references which appear initially without a publication year and,

Average reference time lag per year in zbMATH data.

secondly, the zbMATH scope includes relatively more articles from areas like mathematical physics (MSC 70-86) or mathematics related to computer science (MSC 68), where the time lag in citations is significantly smaller than in the areas predominant in EuDML articles. On the other hand, the open access nature of EuDML articles seems to have no influence on time lag right now, which is not too surprising if we take into account that the average citation still goes back to the times when digital libraries had just begun.

Without doubt, these are just first estimates – there will be more data available soon and many questions (e.g. normalisations with respect to publication growth or citing behaviour, differences according to MSC areas, sources, etc.) are just at the beginning. Just one thing seems to be certain – things just start to be interesting when going more than 10 years back – a period of time which is usually omitted.

Thierry Bouche [thierry.bouche@ujf-grenoble.fr] is maître de conférences at Institut Fourier (Université Joseph-Fourier, Grenoble, France) and director of Cellule Mathdoc (a joint service unit of CNRS and UJF). He is scientific coordinator of the EuDML project, and member of the following committees: Electronic Publication Committee (EMS), Committee on Electronic Information and Communication (IMU), Conference on Intelligent Computer Mathematics steering committee.

Olaf Teschke [teschke@zblmath.fiz-karlsruhe.de] is member of the Editorial Board of the EMS Newsletter, responsible for the Zentralblatt Column.

Krzys Wojciechowski [k.wojciechowski@icm.edu.pl] is a software architect in the Interdisciplinary Centre for Mathematical and Computational Modelling of the University of Warsaw. His professional interests concentrate mostly on data mining, machine learning and big data processing. He currently holds the position of technical coordinator in the EuDML project [http://eudml.org].