Outcome-oriented Fitness Measurement of Personal Learning Environments

Felix Mödritscher

Institute for Information Systems and New Media,
Vienna University of Economics and Business
Augasse 2-6, 1090 Vienna, Austria
+43-1-31336-5277

felix.moedritscher@wu.ac.at

ABSTRACT

Personal learning environments (PLEs) comprise a new kind of learning technology which aims at putting learners into centre stage, i.e. by empowering them to design and use environments for their learning needs and purposes. While a lot of research and development is going on in realizing and providing technical PLE solutions, less effort is spent in examining the 'fitness' of PLEs. By fitness we refer to the property of a PLE that it is successfully used to achieve a goal. In this paper we attempt to formalize the PLE fitness by focusing on one specific aspect, namely on outcomes of PLE-based activities. For this purpose, we analyze a certain kind of PLE outcomes, i.e. publications, by measuring their impact and use real-world data harvested in the Web to propose a mathematical fitness model. Furthermore, we address factors characterizing the fitness of a publication as well as preliminaries of our approach. The paper concludes with pointing out related findings from other fields and possible future work on outcome-oriented PLE fitness measurement.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: Probability and Statistics: *Distribution functions, Time series analysis*, H.2.8 [**Information Systems**]: Database Applications: *scientific databases*, G.1.2 [**Mathematics of Computing**]: Approximation: *Nonlinear approximation*.

General Terms

Algorithms, Measurement, Experimentation.

Keywords

Personal Learning Environments, Scientific Publications, Citation History Analysis, Fitness Function, Gamma Distribution.

1. INTRODUCTION

According to Henri et al. [1], personal learning environments (PLEs) refer to "*a set of learning tools, services, and artifacts gathered from various contexts to be used by the learners*". Furthermore Van Harmelen [2] states that PLEs aim at empowering learners to design (ICT-based) environments for their activities so that they can connect to learner networks in order to collaborate on shared outcomes and acquire necessary (professional and rich professional) competences. In the last years a lot of work has been investigated in the development and application of new, PLE-related technologies (like apps, widgets or gadgets) and their underlying infrastructures (widget containers, personalized websites, mobile phones etc).

Considering the spreading of these technologies in society and the raising profits of leading companies in this sector (e.g. Apple or Google), they are highly successful. However less attention is paid to their usage as personal learning environments and their (positive and negative!) effects on lifelong learning. In order to formalize and examine the evolvability of PLEs, we build upon the notion of fitness, a concept given by evolutionary theory. By comparing the development, spreading, and utilization of PLEs – the technical infrastructures as well as their entities, e.g. tools and their features – to genetic evolution [3], a learning environment can be understood as a socio-technical system (*organism*) with its functionalities (*traits*). According to our initial definition, a PLE is a set of tools, services, artifacts, and peer actors, thus the fitness of a PLE refers to specific situations in which it is used and consequently to defined purposes (*fit-for-purpose*) as well as to the scope of a community and a context (*local fitness*).

Over time, PLEs can evolve, for instance specialize, according to situations in which certain features are used more frequently and others are ignored or even removed – as learners also demand new features, developers are part of this evolutionary process and implement them so that a PLE solution is being used in the future. Such processes bear a resemblance to the concept of natural selection [4]. In the context of this paper, fitness refers to a property describing PLE functionalities. Fitter PLE features (*genes*) become more common, i.e. a certain form of a feature (*allele; DNA sequence*) is used more frequently, spreads faster, or can even substitute other forms of the same functionality.

We explain these definitions through two examples for the evolution of software artifacts in praxis. A first example comprises a new way of providing recommendations. In the last few years many web applications have included recommendations which appear on typing in a term into the search field. Restricting these recommendations to the user's context (e.g. Facebook.com) or auto-completing the query on the basis of terms given by many other users (e.g. Google.com) seem to be two manifestations of this feature which will become more important in the future. So, the generic function “recommendations” has been specialized over time. In a second example a new researcher enters a scientific community on statistical mathematics. In this group of researchers a specific tool, namely the R software, is favored for teaching and research activities. Thus the new member is facing a tool with a high fitness factor within the community and can either work with this tool or try to establish some other software in this community, consequently opposing the R framework.

Overall, the idea of our approach is to consider PLEs as the outcomes of (collaborative, ICT-based) learning – which is also stated e.g. by Wild et al. [5] – and to formalize and examine their evolution over several generations. Unfortunately this would require detailed data about PLE-based activities over a long period of time – which is not easy to get and which we do not have. Therefore we propose to focus on certain aspects of PLE activities, namely on PLE outcomes in the form of scientific papers. We use the information on publications to model and analyze their fitness with respect to their scientific impact.

The rest of the paper is structured as follows. The next section elaborates our approach towards outcome-oriented fitness measurement as well as preliminaries and related work. Then, section 3 describes the stepwise development of a fitness function for PLE outcomes and examines different characteristics of this model. Section 4 summarizes findings as well as similarities to other fields, and discusses the approach towards its relevance for the PLE fitness, before an outlook on future work is given.

2. CONCEPTUAL APPROACH, PRELIMINARIES, AND RELATED WORK

As mentioned before, we consider scientific papers as typical PLE outcomes and use bibliographic data to examine and formalize their fitness. In a first step we have to clarify how publications and PLEs are related. In former research we have elaborated the notion and the most important concepts of PLE-based learning ecologies [6]. Figure 1 shows what PLE-based collaboration looks like. Learners are involved into different activities in which they try to achieve personal and group goals (e.g. publishing a paper to a journal). They use various tools to collaborate on shared artifacts. In the context of this paper, publications can be seen as typical outcomes of such activities, as they are created by one or more scientists using different tools – and even single-authored papers normally involve other actors in the background.

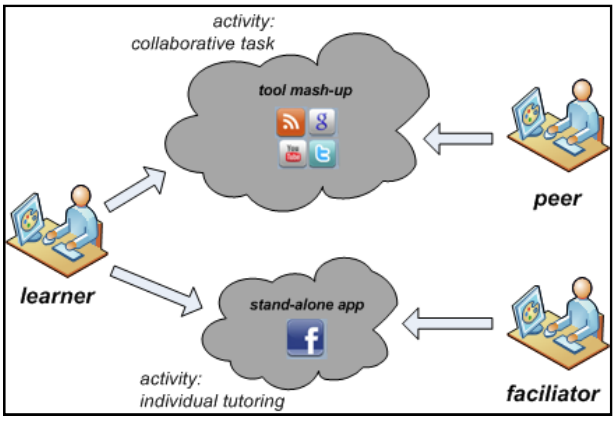

Figure 1. Example scenario for PLE-based collaboration.

On a theoretical level and putting the learner (actor) central stage, Klamma and Petrushyna [7] propose a model of learning ecologies which is based on the Actor-Network Theory (ANT) and describes five important entities of a PLE:

- **Processes:** *Activities* carried out for educational reasons, at workplace, or due to personal goals (e.g. a job task in a business process, attending a course for further education, or a spare time activity requiring the acquisition of new competences)
- **Media:** *Collection of learning resources* required for or created in these activities (e.g. the Wikipedia platform, learning objects repository, or simply the Internet)
- **Artifacts:** *Documents* and other (digital or real-world) artifacts collaboratively created and accessed by learners (e.g. Wiki articles or a joint paper)
- **Agents:** *Actors*, no matter if humans or software (e.g. peer learners or functionality provided by software)
- **Communities:** *People sharing the same environment*, e.g. in terms of having common interests, working on the same artifacts, being connected to the same actors (e.g. a group of learners trying to achieve a course goal or a special interest group for a specific topic)

In the scope of this paper, the PLE related to a publication can be described as follows. A scientific publication is an outcome of a PLE-based activity which involves several human agents in different roles (main author, co-authors, organizer/editor, reviewers, etc.) and using different tools (MS Word, email, conference/journal submission system, etc.). The whole publication process consists of various different activities, e.g. research, writing, and submission activities. Normally, a paper also addresses one or a few scientific communities which can be determined by the targeted journal or conference.

Realistically the PLE of a publication cannot be fully reconstructed any more, as the tools used and the interaction sequences were not tracked sufficiently. Thus, we examine the fitness (success) of papers towards their impact in scientific communities by analyzing the number of citations of different kind of publications over time. The analysis of citations and the citation history of papers is a well-explored field (cf. [8]). Furthermore shortcomings of citation analysis, like biased citing, secondary sources, variations in citation rates with disciplines or nationalities, and many more, are elaborated extensively [8, 9]. Yet, we consider these problems of citation analysis (similarly to the learning environment itself) as part of the outcome of PLE-based activities, being worth an in-depth analysis.

With respect to existing citation indices like CiteseerX (http://citeseerx.ist.psu.edu/), the ISI Web of Knowledge (http://www.isiwebofknowledge.com/), or the ACM Digital Library (http://portal.acm.org/), new tools such as Google Scholar (see http://scholar.google.com/) or community approaches like Mendeley (see http://www.mendeley.com/) provide new opportunities for citation analysis on the basis of large and topical data-sets (cf. upcoming section and [10]).

In the following we describe the development of an approach for formalizing the fitness (citation success) of papers and discuss characteristics of this fitness model.

3. MEASURING AND FORMALIZING THE FITNESS OF SCIENTIFIC PAPERS

First of all, we had to decide on the data source for the bibliographic data required for our approach. After inspecting possible platforms (CiteseerX, ISI Web of Knowledge, ACM Digital Library, Google Scholar, and Mendeley) we conducted a small evaluation study. Therefore, we selected four prominent (i.e. highly cited) publications for this brief evaluation, a well-known book on data mining and papers on booming topics in the Web (Semantic Web and the PageRank algorithm).

Table 1. Comparison of different citation indices (CiteseerX [CX], ISI Web of Knowledge [WoK], ACM Digital Library [ACM], Google Scholar [GS], and Mendeley [M]) on the basis of four highly cited papers and retrieved on February 8, 2011 (*) no. citations given by Scholar vs. sum of yearly citations, +) no. readers)

Publication on:	CX	WoK	ACM	GS*)	M+)
Data mining	n.a.	n.a.	n.a.	10700 /6035	61
Semantic Web	n.a.	1159	n.a.	10709 /8312	323
PageRank (1)	1301	n.a.	n.a.	3670/ 2949	44
PageRank (2)	2140	n.a.	1534	7245/ 5917	573

In Table 1 the comparison of different citation indices is shown. Overall, this statistic confirms the impressions of our inspection. For instance, the data quality of CiteseerX seems to be very poor, as it has no or faulty data on two of our selected publications. On the other hand, the ISI Web of Knowledge and the ACM Digital Library provides bibliographic data on a good quality level but the coverage seems to be poor. Mendeley is not a real citation index, as it rather contains usage data (no. readers) than citations. Yet, this data is interesting and valuable for our evaluation. In sum, we decided to use Google Scholar which contains significantly more and topical data-sets. Moreover, the quality of this data is on a reasonable level, which is also backed up by other evaluation studies, e.g. one on citation mining [11].

With respect to [12], citing a research paper follows the Poisson process, a stochastic process in which citations occur continuously and independently of each other. More precisely, the citation curve of a publication can be formalized by the convolution of two Poisson distributions, one describing the initial phase of a paper's uptake and another one representing its continuous aging process. As a simplification and to combine the two citation curves into one model, we propose to use the Gamma distribution to formalize the fitness of a paper according to its citations. The probability density function of a Gamma distribution is defined as follows [13]:

$$f(x;k,\theta) = \frac{x^{k-1}\,e^{-\frac{x}{\theta}}}{\theta^k\,(k-1)!} \; for\; k,\theta > 0 \; and \; x \geq 0$$

Different to former research which is based upon the Avramescu function [12] – a specialization of the Erlang distribution which itself is a special kind of Gamma distribution –, we use the Gamma distribution for formalizing the fitness of a paper, as it allows approximating the citation curve according to two parameters, the shape (k) and the scale (θ). Given the number of citations per year retrieved from Google Scholar, we use the citation history of prominent papers to develop a method for estimating these two parameters.

Figure 2 displays the citation curves of the four papers analyzed in Table 1. All of these publications are well cited and have sufficient data starting in the years 1998, 2001, and 2006. The book on data mining (green curve) is problematic, as it is the second edition and thus the citation history seems to be biased. However, the other three papers deal with important innovations in the field of computer science and are considered to be appropriate for developing a method for measuring the fitness of PLE outcomes.

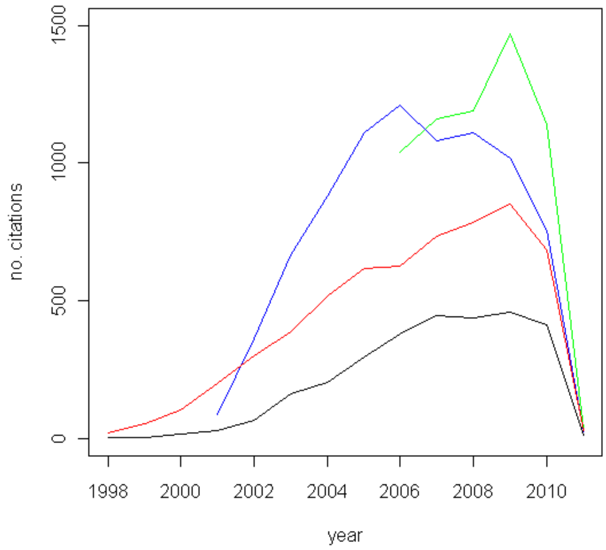

Figure 2. Citation curves of the four publications mentioned in Table 1 (data-sets taken from Google Scholar on February 8, 2011; green curve: data mining book, blue curve: Semantic Web paper, red and black curve: two papers on PageRank).

For developing our method to approximate the citation history according to a Gamma distribution, we used the second paper on PageRank (S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine”, 1998) because sufficient data is provided over a long period of time (see red curve in Figure 2). Basically, our fitness measurement method consists of three steps to approximate a given citation history: (1) determination of the mode, i.e. the value that occurs most frequently in the data-set; (2) parameter estimation of the shape and the scale with respect to minimizing the error rate of the given sample according to the probability density function (pdf) of the Gamma distribution; (3) visualization and evaluation of the approximated fitness curve.

The first step, the identification of the mode, is the one which is the trickiest and highly restricts our approach but it is also necessary. As we have only data-sets of the first years after publications appear, we decided to select the mode manually due to two facts. On the one hand, distribution fitting algorithms are based on the preliminary that the values are distributed over time – which is not the case for our data. Existing software, like the open source framework for statistical computing and graphics (R Project, see http://cran.r-project.org/), provide packages for estimating the parameters of Gamma distributions (cf. [14]), but they do not lead to useful results for our data. On the other hand, we have to assume that the mode is already included within the data-set available, which is also a necessary condition for our approximation method.

However, having the mode of the distribution gives us the possibility to estimate the two parameters (shape k, scale θ) on the basis on the following mathematical relationship (setting first derivation of pdf to 0):

$$x_{mode} = (k-1)\,\theta \;\rightarrow\; \theta = \frac{x_{mode}}{k-1}$$

In a second step, we used (n-2) values of our citation history for estimating the two parameters so that the error rate is minimal. It is recommended to not use the citation data of the last two years (here 2010 and 2011) because of publication and indexing delays, thus the number of citations is incomplete. Given the mode, we have a written a R function which numerically calculates the best values for k and θ by means of minimizing the error rate of the first m values of the citation history (with m being number of values to the mode) according to the following equation:

$$mean\ error(k,\theta) = \frac{1}{m}\sum_{i=1}^{m}\left| x_i \frac{f_{k,\theta}(m)}{x_m} - f_{k,\theta}(i)\right| \rightarrow min.$$

After calculating the parameters (e.g. k = 5.042 and θ = 2.968827 for the selected PageRank paper), the third step comprises evaluation (the relative error for these parameters is 7.85%) and a visualization of the approximated curve. Figure 3 shows the number of citations gathered from Google Scholar and the approximation according to the Gamma distribution.

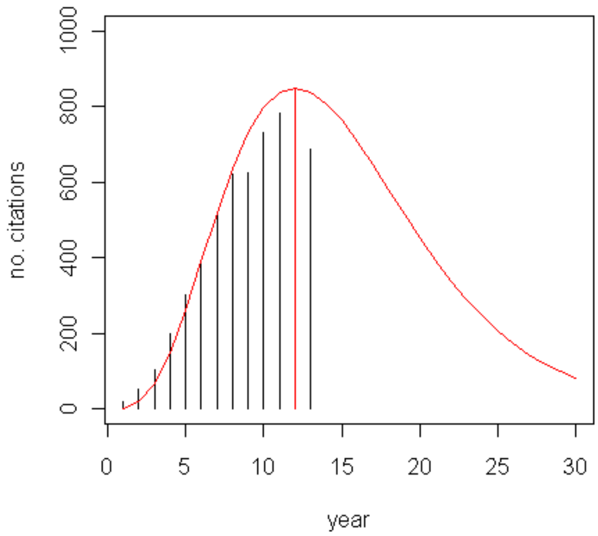

Figure 3. Gamma approximation for PageRank (2) paper from Figure 1 (x is the time axis starting with 1 as the publication year; red curve describes the Gamma pdf approximated according to the citation history).

In principle, we now can formalize the fitness of a PLE outcome by two numbers, the shape and the scale of the Gamma pdf. If based on sufficient data, this distribution of a publication’s citation history seems to be reasonable, as it starts to have impact after being published, reaches a peak some years in the future and then decreases again. The last phase can be argued by effects like more successful follow-up publications or aging of published knowledge. Overall, this fitness measurement enables comparing the success (impact) of publications to each other.

In the next step we analyzed the fitness of different publications: (a) the most frequently cited papers, i.e.

fundamental literature of a selected scientific community, (b) a successful follow-up paper by a lead researcher, (c, d) average (less successful) papers of the same author (single-authored and co-authored papers), and (e) the mostly cited paper of other researchers in a selected field. We used the bibliographic data of the adaptive hypermedia (AH) community, as this discipline is very young and most of the key publications are captured by the index of Google Scholar.

Table 2. Comparison of selected papers according to our fitness estimation method (data retrieved from Google Scholar on February 23, 2011)

Publication:	k	θ	norm. factor	rel. error
1. Brusilovsky, "Methods and techniques of adaptive hypermedia", 1996 [**1373 citations, 16 values**]	3.105	4.751	2336.03	12.54
2. Brusilovsky, "Adaptive hypermedia", 2001 [**1274 citations, 11 values**]	2.993	3.010	2043.25	9.68
3. Brusilovsky et al., "From adaptive hypermedia to the adaptive web", CACM, 2001 [**303 citations, 11 values**]	3.347	2.983	486.46	15.82
4. De Bra, Brusilovsky, "Adaptive hypermedia: from systems to framework", 1999 [**159 citations, 13 values**]	3.724	2.937	275.57	15.60
5. Brusilovsky, "Adaptive educational systems on the world-wide-web", 1998 [**174 citations, 14 values**]	6.648	1.062	141.29	24.92
6. De Bra et al., "AHAM: a Dexter-based reference model for adaptive hypermedia", 1999 [**326 citations, 13 values**]	3.372	2.530	394.38	22.73

Table 2 gives an overview of the comparison of papers being relevant for the assumptions (a-e). A first observation deals with the relative error of the approximation. Obviously the error decreases if more values per year are given. Particularly the last two publications are approximated moderately, as the relative error is above 20%. Yet, the approximation according to Gamma distribution works well, as also shown by the papers' fitness functions in Figure 4. As mentioned before, it is important to not consider the two latest years of the citation history retrieved due to publication and indexing delays. These values (2010, 2011) are also not visualized in the figure.

A second interesting observation concerns the shape parameter (k). A lower shape factor is an indicator for a fitter paper, i.e. a publication cited more often in a shorter period of time and reaching the citation peak earlier. Comparing the first two papers, both were published by the same author and on the same topic. Yet, the second one is cited nearly as much as the first one although being published 5 years later. Most probably, the second paper will outpace the first one in the next years, which can be concluded from the fitness functions shown in Figure 4. As we assume the fitness of a publication to be dependent on the community, we restrict the comparison of Gamma parameters to this scientific field. Thus, the shape calculated for the PageRank paper (Web researcher) cannot be set in direct relation with the shape factors of the AH papers.

Next to the speed of a paper's uptake, success can be also determined by the number of citations in general. Here, both scaling factors, the Gamma parameter θ (second column of Table 2) as well as the factor to normalize the citation history to the pdf of the Gamma distribution (third column), allow inferences on the quantity of citations. The first two papers are cited significantly more often than the papers 3 and 6 which in turn are more successful than the publications 4 and 5. However, both scaling factors dependent on the shape k that is why the fitness function of the first paper has a higher scale and a higher normalization factor but a lower peak.

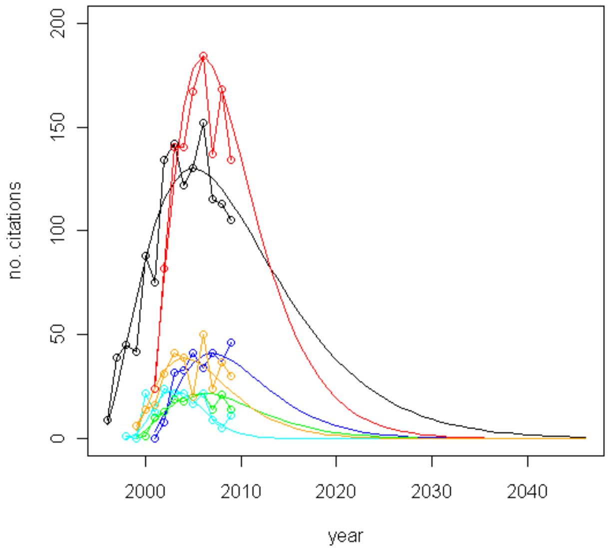

Figure 4. Fitness functions and citation histories (from the publication years to 2009) of the papers depicted in Table 2 (colors: 1. black, 2. red, 3. blue, 4. green, 5. cyan, 6. orange).

Overall, we have tackled a set of very diverse publications for which the fitness functions are visualized in Figure 4. The first two papers (scenario (a); black and red curve) are the most frequently cited papers of one of the lead researchers of the AH community. These two curves evidence that two very successful papers behave different in being cited within a community, i.e. that one publication can be fitter than another one and that preferential attachment [15] – a favored paradigm for emergent, networked structures – is not always valid.

The fitness of the third paper, a successful follow-up paper of the AH lead researcher (scenario c), is similar to the mostly cited paper of another (well-known) researcher in this scientific field (scenario e). The less successful papers (scenario d) are problematic as the approximation of the fitness curve does not work that good (high relative error). Most obviously, they are characterized by a shape which is growing slower. Particularly paper 5 has a shape of over 6, meaning that the data could be faulty or that the uptake of this work was that slow.

Addressing further issues that might have an influence on our fitness estimation method, [8, 9] give a comprehensive overview on problematic issues of citation analysis. Due to a lack of space and time, we have not addressed the phenomena of self-citations which we assume to be necessary to successfully 'initialize' the fitness of a paper. Concerning such influential factors, we refer to future work which could aim at differentiating between self-citations and citations by other researchers and examining the different fitness functions.

Finally it has to be outlined that our fitness estimation method also includes a model for predicting the future citation frequency. Given the data of the papers we have examined, this prediction worked fine for those citation histories going beyond the citation peak. On the other hand, this prediction is also based on the assumption that in the future no unforeseeable event concerning a publication (e.g. a rediscovery after a couple of decades) occurs. Here, our approach is restricted to the condition that the citation peak is given and that it is a global maximum.

4. CONCLUSIONS, RELATIONS TO OTHER FIELDS, AND FUTURE WORK

In this paper we have examined a very particular aspect of personal learning environments, namely publications as outcomes of distributed, collaborative, and technology-based activities. Precisely we have proposed a method for formalizing the fitness of such scientific content artifacts, i.e. the success in being taken up, on the basis of usage data (the number of citations) retrieved by a large and up-to-date citation index. Although being restricted by some hard conditions (sufficient data available; citation peak given and global maximum; dependency on a scientific community), the fitness measurement method seems to be valid and reasonable due to the following reasons.

On the one hand, approximation works fine for well-cited papers, as shown in the last section. On the other hand, citing scientific publications is a natural process for which the waiting times between Poisson distributed events are relevant [16], which can be characterized by a Gamma distribution. Similar processes can be observed in other areas, like weather forecast (estimating the likelihood of monthly rainfalls for draught monitoring [17]), insurance businesses (effect of risk factors, like rainfalls, on insurance claims [18]), medical treatment (time to treatment response in arthritis patients [19]), or modeling the distribution of fitness effects in evolutionary biology in general [20, 21, 22].

Although the connection between scientific publications and the PLEs leading to such artifacts is very vague, we think that the fitness model proposed in this paper is generally relevant for PLE-based activities, as other aspects of personal learning processes (e.g. tool usage or communication behavior) might underlie a similar lifecycle and a curve following a Gamma distribution. In particular the results of our research are relevant for those activities which aim at creating artifacts that should be extensively used by others. By applying our approximation method it is possible to compare the success of papers with each other and to predict their future performance. However, we see the work tackled in this paper as a first step only. Based on the fitness estimation method developed, next steps could address the fitness curves of publications according to different scientific communities (local fitness assumption), to the social networks of paper authors (co-author assumption), to self-citations (initialization assumption), to the novelty and quality of publications (fit-for-purpose assumption), or to other characteristics of such PLE outcomes.

Furthermore, future work could comprise a closer examination of the PLEs which led to high impact papers, i.e. by interviewing the authors of such publications. Additionally it would be valuable to develop a tool for (semi-)automatically calculating the fitness curve of user-selected papers. From the evaluation perspective it is necessary to examine papers of different scientific fields – if sufficient data is available – and to use data from other systems, i.e. real usage data on publications as captured e.g. by Mendeley (cf. author readership analysis available at http://readermeter.org).

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 231396 (ROLE project).

6. REFERENCES

[1] Henri, F., Charlier, B., and Limpens, F. 2008. Understanding PLE as an Essential Component of the Learning Process. In *Proc. of ED-Media* (Vienna, Austria, Jun 30-Jul 4, 2008). AACE, Chesapeake, VA, 3766-3770.

[2] Van Harmelen, M. 2008. Design trajectories: Four experiments in PLE implementation. *Interactive Learning Environments* 16, 1, 35-46.

[3] Futuyma, D.J. 2005. *Evolution*. Sinauer Associates, Sunderland, MA.

[4] Darwin, C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. William Cloves and Sons, London.

[5] Wild, F., Mödritscher, F., and Sigurdarson, S.E. 2008. Designing for Change: Mash-Up Personal Learning Environments. *eLearning Papers*, 2008(9).

[6] Mödritscher, F., and Petrushyna, Z. 2009. Model and Methodology for PLE-Based Collaboration in Learning Ecologies. *Deliverable D7.1/ID7.2*, ROLE consortium.

[7] Klamma, R., and Petrushyna, Z. 2008. The Troll Under the Bridge: Data Management for Huge Web Science Mediabases. In *Proc. of the 38.*

Jahrestagung der Gesellschaft für Informatik e.V. (GI), die INFORMATIK 2008 (München, Germany, Sept 8-13, 2008), Köllen Druck+Verlag GmbH, Bonn, 923-928.

[8] Smith, L.C. 1981. Citation analysis. *Library Trends*, 30, 83-106.

[9] MacRoberts, M.H., and MacRoberts, B.R. 1996. Problems of citation analysis. *Scientometrics*, 36, 435-444.

[10] Harzing, A.-W.K., and Van der Wal, R. 2008. Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics* 8, 1, 62-71.

[11] Afzal, M.T., Maurer, H., Balke, W., and Kulathuramaiyer, N. 2010. Rule based Autonomous Citation Mining with TIERL. *Journal of Digital Information Management (JDIM)* 8, 3, 196-204.

[12] Egghe, L., and Rousseau, R. 2000. The influence of publication delays on the observed aging distribution of scientific literature. *Journal of the American Society for Information Science*, 51, 2, 158-165.

[13] Choi, S.C., and Wette, R. 1969. Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. *Technometrics*, 11, 4, 683-690.

[14] Ricci, V. 2005. *Fitting distributions with R*. Technical report, retrieved from http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf (2011-02-25).

[15] Barabási, A.-L., and Albert, R. 1999. Emergence of Scaling in Random Networks. *Science*, 286, 5439, 509-512.

[16] Weisstein, E.W. 2005. Gamma Distribution. *MathWorld*, Wolfram Research, retrieved from http://mathworld.wolfram.com/GammaDistribution.html (2011-02-25).

[17] Husak, G.J., Michaelsen, J., and Funk, C. 2007. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *International Journal of Climatology*, 27, 7, 935-944.

[18] Yuen, K.C., Guo, J., and Wu, X. 2002. On a correlated aggregate claims model with poisson and erlang risk processes. *Insurance: Mathematics and Economics*, 31, 2, 205-214.

[19] Abrahamyan, L., Beyene, J., Feng, J.Y., Chon, Y., Willan, A.R., Schmeling, H., Horneff, G., Keystone, E.C., and Feldman, B.M. 2010. Response times follow lognormal or gamma distribution in arthritis patients. *Journal of Clinical Epidemiology*, 63, 12, 1363-1369.

[20] Nielsen, R., and Yang, Z. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Molecular Biology and Evolution*, 20, 8, 1231-1239.

[21] Loewe, L, and Charlesworth, B. 2006. Inferring the distribution of mutational effects on fitness in Drosophila. *Biology Letters*, 2, 3, 426-430.

[22] Gu, X. 2007. Stabilizing selection of protein function and distribution of selection coefficient among sites. *Genetica*, 130, 1, 93-97.