AUTHOR QUERY FORM

	Journal: JOI	Please e-mail or fax your responses and any corrections to:
		E-mail: corrections.esch@elsevier.thomsondigital.com
ELSEVIER	Article Number: 184	Fax: +353 6170 9272

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: <u>click on the Q link to go</u> Please insert your reply or correction at the corresponding line in the proof
$\frac{Q1}{Q2}$	Please check affiliations. Please check the corresponding address, telephone/fax number of the corresponding author, and correct if necessary.

Thank you for your assistance.

G Model IOI 184 1–8

Journal of Informetrics xxx (2010) xxx-xxx



Contents lists available at ScienceDirect

Journal of Informetrics



journal homepage: www.elsevier.com/locate/joi

How and where the TeraGrid supercomputing infrastructure benefits science

Johan Bollen^{a,b,*}, Geoffrey Fox^{a,c}, Prashant Raj Singhal^a

Q1 ^a School of Informatics and Computing, Indiana University, United States

Center for Complex Networks and Systems Research, United States

^c Pervasive Technology Institute, United States 5

ARTICLE INFO

Article history: a

- 10 Received 15 June 2010
- Received in revised form 11 August 2010 11 Accepted 23 September 2010
- 12

ABSTRACT

We investigate how the benefits of the TeraGrid supercomputing infrastructure are distributed across the scientific community. Do mostly high-impact scientists benefit from the TeraGrid? Are some scientific domains more strongly represented than others in TeraGridsupported work? To answer these questions, we examine the relation between TeraGrid usage and scientific impact for a set of scientists whose projects relied to varying degrees on the TeraGrid infrastructure. For each scientist we measure TeraGrid usage expressed in terms of allocated Service Units (SU) vs. various indicators of their scientific impact such as the h-index, total citations, and citations per article. Our results show a significant correlation between scientific impact and TeraGrid usage. We furthermore examine the distribution of TeraGrid-related publications across various scientific journals. A superposition of these journals over an existing large-scale map of science shows how TeraGrid-supported work is mostly concentrated in Physics and Chemistry, with a lesser focus on biology.

© 2010 Elsevier Ltd. All rights reserved.

The TeraGrid integrates high-performance computers, data resources and tools, and high-end experimental facilities 13 around the country, including more than 2 petaflops (quadrillions of floating point operations) of computing capability 14 and more than 60 petabytes (quadrillions of bytes) of online and archival data storage with rapid access and retrieval over 15 high-performance networks. It is presently the world's largest, most comprehensive distributed cyberinfrastructure for open 16 scientific research.

The contributions of the TeraGrid to high-impact scientific work since its start are indisputable, but how are they dis-18 tributed? Here we investigate two basic questions with regards to TeraGrid usage patterns (Lawrence & Zimmerman, 2007; 19 Zimmerman & Finholt, 2006). First, do higher-impact scientists amongst TeraGrid users make more use of the TeraGrid 20 infrastructure, or a more tantalizing corollary; does TeraGrid use in that community lead to higher impact? Second, do all 21 scientific domains benefit equally from the TeraGrid's facilities or have some leveraged this infrastructure more efficiently 22 than others? 23

The TeraGrid accounting and allocation systems keep extensive records of the allocations and usage of its resources, along 24 25 with project codes and fields of science, and project-related publications for each Principal Investigator (PI). We collected the following data for 112 scientists that were allocated computing time on the TeraGrid in one quarterly meeting in 2009: 26

* Corresponding author at: Indiana University, School of Informatics and Computing, 919 E. 10th Street, Bloomington, Indiana, IN 47408, United States. Tel.: +1 812 856 1893; fax: +1 505 665 6452.

E-mail address: jbollen@indiana.edu (J. Bollen).

1751-1577/\$ - see front matter © 2010 Elsevier Ltd. All rights reserved. doi:10.1016/j.joi.2010.09.004

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx



Fig. 1. Cumulative distributions of SU allocations (left), Total Cites (middle) and h-index (right). Raw values and cumulative percentages for 25 highest ranking PIs are provided in Table 1.

Table 1

27

28

29

30

31

32

33

34

35

36

37

38

39

40

Raw values and cumulative percentages of SU allocated (SU) in millions (M), Total Cites (TC), and h-indices (hx) for 25 PIs sorted according to highest values for each indicator. The graphs in Fig. 1 shows cumulative distribution for all 112 PIs.

PI rank	SU(M)	$SU\geq\%$	TC	$TC \ge \%$	hx	$hx \ge \%$
1	36.00	11.48	22.53	12.55	68.00	4.09
2	34.97	22.63	8.02	17.02	45.00	6.80
3	25.00	30.60	7.08	20.97	43.00	9.39
4	20.02	36.98	6.87	24.80	42.00	11.92
5	15.00	41.76	6.80	28.59	38.00	14.21
6	12.00	45.59	6.27	32.08	37.00	16.44
7	8.03	48.15	6.24	35.56	36.00	18.60
8	8.00	50.70	6.11	38.96	36.00	20.77
9	7.50	53.09	5.34	41.94	36.00	22.94
10	6.50	55.16	4.84	44.63	34.00	24.98
11	5.33	56.86	4.65	47.22	33.00	26.97
12	5.00	58.45	4.64	49.81	33.00	28.96
13	4.20	59.79	4.07	52.08	30.00	30.76
14	4.01	61.07	3.93	54.27	29.00	32.51
15	4.00	62.35	3.83	56.40	28.00	34.20
16	4.00	63.62	3.79	58.52	26.00	35.76
17	3.95	64.88	3.77	60.62	26.00	37.33
18	3.81	66.10	3.27	62.44	26.00	38.89
19	3.65	67.26	3.00	64.10	25.00	40.40
20	3.55	68.39	2.85	65.69	24.00	41.84
21	3.11	69.38	2.82	67.26	23.00	43.23
22	3.10	70.37	2.51	68.66	23.00	44.61
23	3.08	71.35	2.46	70.03	22.00	45.94
24	3.05	72.33	2.28	71.30	21.00	47.20
25	3.00	73.28	2.23	72.54	21.00	48.46

1. The Service Units (SUs) that were allocated to the PI, defined as the sum of the CPU core-hours allocated across various TeraGrid resources.

2. A variety of indicators of scientific impact derived from the Publish or Perish tool¹ that collects citation statistics from Google Scholar (Harzing & van der Wal, 2008) to calculate among others the PI's total accumulated citations, citations per article, the PI's h-index (Hirsch, 2005), g-index (Egghe, 2006), and several others.

Care was taken to disambiguate author names to avoid duplicating citation counts.

A multiple regression analysis over all PI's (N = 112) indicates that Total Papers published (p < 0.001), h-index (p = 0.010), Total Cites (p = 0.006), E-index (p = 0.008), and Hirsch's M-index (p = 0.017) are statistically significant predictors of SU-allocations.² Since Total Cites and the h-index are presently some of the best characterized indicators of scientific impact we compare these to a PI's SU-allocation in the TeraGrid.

First we examine the cumulative distributions of SUs allocated, Total Cites and h-indices as shown in Fig. 1. In each individual graph in Fig. 1 we sort the PIs from left-to-right along the *x*-axis, from highest to lowest individual values. The *y*-axis value for a given PI is the sum of the values of all higher or equally ranked PIs (thus including the PI herself) normalized to a percentage of the total over all PIs. Raw values and percentages for the 25 highest ranking PIs are listed in Table 1.

Please cite this article in press as: Bollen, J., et al. How and where the TeraGrid supercomputing infrastructure benefits science. *Journal of Informetrics* (2010), doi:10.1016/j.joi.2010.09.004

2

¹ http://www.harzing.com/pop.htm.

² Multiple regression analysis: $R^2 = 0.369$, *F*-statistic = 3.75 on 15 and 96 DF, *p*-value = 3.521E - 05.

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx



Fig. 2. PI's SU-allocation vs. h-index (left), Total Cites (middle), and Project Total Cites (right).

Table 2

41 42

43

44

45

46

47

48

10

50

51

53

54

55

Summary of fit achieved by regression analyses of various combinations of Total Cites, h-index and Total Project Cites vs. SU allocations.

Model	R^2	F-statistic	<i>p</i> -value
SU allocations ~ h-index:	0.188	25.41	1.833E-06
SU allocations ~ Total Cites	0.218	30.57	2.192E-07
SU allocations ~ Project Cites	0.451	19.7	0.0001719
SU allocations \sim h-index, Total Cites	0.222	15.56	1.139E-06



Fig. 3. Distribution of a preliminary "Return on Investment" metric: ratio of Total Cites per SU Allocated for all PI.

We find strongly skewed distributions for all three indicators, but mostly so for cumulative SU allocations. The first 7 highest TeraGrid users combined (out of 112) receive 50% of total SU allocations, and the first 26 highest users combined are allocated 75% of total SU allocations. A similarly skewed distribution can be found for the cumulative distribution of Total Cites. The publications of the first 12 TeraGrid PIs receive 50% of all citations, and the first 27 receive 75% of all citations. The cumulative h-index distribution, produced by simply adding h-indices, follows a less skewed distribution, but here too we can see that that the 26 highest-ranked PIs represent 50% of accumulative h-index values.³

From these cumulative distributions we conclude that a small minority of TeraGrid PIs receives the majority of SU allocations but also generates the majority of scientific impact as indicated by Total Cites and h-index.⁴

Second we examine the relations between H-index, Total Cites and SU Allocation distribution. We find statistically significant Spearman rank-order correlations (denoted ρ) between SU-allocation vs. the h-index (ρ = 0.337, p < 0.001, N = 112), and SU-allocation vs. Total Cites (ρ = 0.357, p < 0.001, N = 112). A stronger correlation is found when we compare the total citations to articles published by all participants of a given TeraGrid project vs. its SU allocations (ρ = 0.626, p < 0.001, N = 27). These correlations are moderate, but highly statistically significant for the sample size.

The scatterplots in Fig. 2 summarize these comparisons.⁵ A linear regression line (blue) was added to visually highlight the correlation pattern, but not to suggest that it is in fact linear. Increased SU-allocations do seem to correspond to increased

³ H-indices are arguably not cumulative, but we want to show that the highest ranked PIs represent a disproportional amount of total scientific impact as measured by their h-indices.

⁴ We remind the reader that throughout this paper SU allocations refer to the one quarter's data we analyze here.

⁵ Our purpose is to look at statistical impact and not to evaluate individual projects. Therefore we have anonymized the data and these charts for analysis.

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx

h-indices and Total Citations of the Pls, indicating a positive relation between scientific impact and SU-allocations. This relation seems to be most reliable at the project level (Fig. 2-right).

A few things have to be noted about the observed correlations.

First, correlation is not proof of causation. This analysis does not tell us whether scientists within the TeraGrid community achieve higher impact through TeraGrid use or whether high impact scientists tend to make more use of TeraGrid facilities. A careful longitudinal analysis of the impact trajectories of particular scientists and their teams over time is required to make that determination. However it is not unreasonable to assume some degree of bidirectional interaction between TeraGrid use and scientific impact. SU allocation decisions may be shaped by perceptions of the PIs scientific reputation; in fact the h-indices are based on citation data that was recorded well before SU allocation decisions were made in 2009. Conversely, TeraGrid use may over time increase a PI's scientific impact by establishing a foundation for high-impact research.

Second, the linear regression has a poor fit in spite of the statistically significant correlations between Total Cites, hindices and Project Total Cites vs. SU allocations. This is also the case for a multiple linear regression analysis of Total Cites and h-index vs. SU allocations ($R^2 = 0.222$). Table 2 lists the results of these regression analyses which indicate a considerable amount of scatter, i.e. a significant number of TeraGrid users and project receive low SU allocations yet produce high impact science.

The mentioned scatter effect is shown in Fig. 3 where for each PI we calculate the ratio *T* of Total Cites (*TC*) per SU allocated (*SU*), i.e. $T = \log (TC/SU)$, and plotted *T* against SU allocated (log (*SU*)). The distribution in Fig. 3 (left) shows that



Fig. 4. Frequency distribution of journals that TeraGrid users most frequently publish in.

Please cite this article in press as: Bollen, J., et al. How and where the TeraGrid supercomputing infrastructure benefits science. *Journal of Informetrics* (2010), doi:10.1016/j.joi.2010.09.004

4

58

G Model JOI 184 1–8

74

75

76

77

78

70

80

81

82

83

84

ARTICLE IN PRESS

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx

T values increase as SU allocations increase (ρ = 0.356). Furthermore, we find several orders of magnitude in the variation among PIs in terms of their *T* values at different levels of SU allocation (R^2 = 0.124). This is further confirmed by the histogram in Fig. 3 (right) of the marginal distribution of *T* values which spans several orders of magnitude. Given the large variation of *T* values at different SU allocations, these results do not support a strategy of prioritizing SU allocations in favor of high-impact scientists.

To investigate whether TeraGrid supports scientific research across various domains, we analyze the frequency distribution of the journals that TeraGrid users publish in. The resulting distribution is shown in the log-normal plot of Fig. 4; it is highly skewed towards astrophysics, physics and biochemistry and seems to group journals in two classes that are separated by nearly an order of magnitude difference in publication numbers: one that includes the Astrophysics Journal, APS Meeting Abstracts and Physical Review D, and a distant second that includes the Journal of Computational Physics, Biochemistry and the Astrophysical Journal Supplements.

A similar focus on a limited set of scientific domain emerges when we superimpose the 150 journals that TeraGrid users 85 most frequently published in over the course of their careers on the MESUR (Bollen, Van de Sompel, & Rodriguez, 2008) 86 Map of Science Bollen et al. (2009). This map was derived earlier from large-scale scientific journal usage data collected 87 from some of the world's most significant publishers, aggregators and academic institutions Butler (2009). The resulting 88 map of TeraGrid domains is shown in Fig. 5. Two journals in the map are connected if they are frequently co-retrieved in 89 users' clickstreams. Journals are color-coded according to the JCR and Dewey Decimal subject classification of the journal 90 in question. The large diamonds indicate journals that are part of the TeraGrid journal set, the smaller circles correspond to 91 any other journal in the map. 92



Fig. 5. The journals that TeraGrid users most frequently published in superimposed on the MESUR clickstream map of science (TeraGrid journals are indicated by cliamonds, all other journals by circles).

G Model JOI 184 1–8

ARTICLE IN PRESS

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx

The map confirms a strong focus among TeraGrid users on Physics and Chemistry with Biology, Engineering and Geo/Astro-Physics situated in the margins. Although the social sciences and humanities are well represented in the MESUR map, they are absent from the set of TeraGrid journals. This is surprising since the social sciences seem to have recently experienced a surge of activity in applications of computational science to models of large-scale socio-cultural phenomena. This phenomenon is not manifested in TeraGrid usage, not even as a secondary relation to the primary clusters of interest in this map. In addition, we find no evidence of any significant level of activity in medicine, cognitive science and sociology. From our results we can draw the following conclusions:

- 1. TeraGrid usage is indeed significantly correlated with the scientific impact of its users, but the causal direction of this relation remains unclear.
- 2. Use of the TeraGrid is disproportionally oriented towards traditional scientific domains; it has not yet reached the full
 range of scientific domains that may benefit from large-scale super-computing infrastructure.

Analyses such as these could greatly benefit from the TeraGrid (and similar facilities) gathering and presenting user and 104 publication data in a systematic and automated fashion. The present study is based on data that corresponds to only one 105 quarterly allocation of SUs and can thus not resolve longitudinal effects such as the potential cause and effect between 106 scientific productivity and TeraGrid allocation size. The availability of more detailed, longitudinal data could resolve this 107 issue, and provide the basis for an expanded analysis that examines in addition the correlations between use of other modes 108 of computing (clouds, clusters) and scientific productivity across various scientific domains. Of great interest would be the 100 development of "Return of Investment metrics" similar to our T value (ratio of Total PI Citations and SU allocated) that could 110 provide indications of where investments in supercomputing infrastructure could best be directed to maximize scientific 111 productivity and impact. 112

113 Appendix A.

114

See Tables A.1–A.4.

Table A.1

Raw values and cumulative percentages of SU allocated (SU) in millions (M), Total Cites (TC), and h-indices (hx) for 112 PIs sorted according to highest values for each indicator. The graphs in Fig. 1 show cumulative distribution for all 112 PIs.

PI rank	SU(M)	$SU \ge \%$	ТС	$TC \ge \%$	hx	$hx \ge \%$
1	36.00	11.48	22.53	12.55	68.00	4.09
2	34.97	22.63	8.02	17.02	45.00	6.80
3	25.00	30.60	7.08	20.97	43.00	9.39
4	20.02	36.98	6.87	24.80	42.00	11.92
5	15.00	41.76	6.80	28.59	38.00	14.21
6	12.00	45.59	6.27	32.08	37.00	16.44
7	8.03	48.15	6.24	35.56	36.00	18.60
8	8.00	50.70	6.11	38.96	36.00	20.77
9	7.50	53.09	5.34	41.94	36.00	22.94
10	6.50	55.16	4.84	44.63	34.00	24.98
11	5.33	56.86	4.65	47.22	33.00	26.97
12	5.00	58.45	4.64	49.81	33.00	28.96
13	4.20	59.79	4.07	52.08	30.00	30.76
14	4.01	61.07	3.93	54.27	29.00	32.51
15	4.00	62.35	3.83	56.40	28.00	34.20
16	4.00	63.62	3.79	58.52	26.00	35.76
17	3.95	64.88	3.77	60.62	26.00	37.33
18	3.81	66.10	3.27	62.44	26.00	38.89
19	3.65	67.26	3.00	64.10	25.00	40.40
20	3.55	68.39	2.85	65.69	24.00	41.84
21	3.11	69.38	2.82	67.26	23.00	43.23
22	3.10	70.37	2.51	68.66	23.00	44.61
23	3.08	71.35	2.46	70.03	22.00	45.94
24	3.05	72.33	2.28	71.30	21.00	47.20
25	3.00	73.28	2.23	72.54	21.00	48.46
26	3.00	74.24	2.17	73.75	21.00	49.73
27	3.00	75.20	2.01	74.87	20.00	50.93
28	2.40	75.96	1.81	75.88	20.00	52.14
29	2.29	76.69	1.78	76.87	20.00	53.34
30	2.20	77.39	1.76	77.86	20.00	54.55

6

93

94

95

96

97

98

99

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx

Table A.2

Raw values and cumulative percentages of SU allocated (SU) in millions (M), Total Cites (TC), and h-indices (hx) for 112 PIs sorted according to highest values for each indicator. The graphs in Fig. 1 show cumulative distribution for all 112 PIs.

PI rank	SU(M)	$SU \ge \%$	TC	$TC \ge \%$	hx	$hx \ge \%$
31	2.18	78.09	1.70	78.81	20.00	55.75
32	2.00	78.72	1.69	79.75	19.00	56.89
33	2.00	79.36	1.62	80.65	19.00	58.04
34	2.00	80.00	1.49	81.48	19.00	59.18
35	2.00	80.64	1.36	82.24	19.00	60.33
36	2.00	81.27	1.33	82.98	18.00	61.41
37	2.00	81.91	1.15	83.62	18.00	62.49
38	1.91	82.52	1.15	84.26	18.00	63.58
39	1.90	83.13	1.15	84.90	18.00	64.66
40	1.80	83.70	1.15	85.54	18.00	65.74
41	1.75	84.26	1.15	86.18	17.00	66.77
42	1.66	84.79	1.14	86.82	17.00	67.79
43	1.60	85.30	1.13	87.45	16.00	68.75
44	1.60	85.81	1.05	88.03	15.00	69.66
45	1.50	86.29	1.04	88.61	15.00	70.56
46	1.50	86.77	0.90	89.11	15.00	71.46
47	1.50	87.24	0.89	89.61	15.00	72.37
48	1.50	87.72	0.85	90.08	15.00	73.27
49	1.50	88.20	0.85	90.55	14.00	74.11
50	1.30	88.62	0.84	91.02	14.00	74.95
51	1.20	89.00	0.84	91.49	13.00	75.74
52	1.20	89.38	0.74	91.90	13.00	76.52
53	1.20	89.76	0.68	92.28	13.00	77.30
54	1.20	90.15	0.68	92.66	12.00	78.03
55	1.14	90.51	0.65	93.02	12.00	78.75
56	1.08	90.85	0.63	93.37	12.00	79.47
57	1.04	91.19	0.60	93.71	12.00	80.19
58	1.01	91.51	0.59	94.04	12.00	80.92
59	1.00	91.83	0.58	94.36	12.00	81.64
60	1.00	92.15	0.56	94.67	12.00	82.36

Table A.3

Raw values and cumulative percentages of SU allocated (SU) in millions (M), Total Cites (TC), and h-indices (hx) for 112 PIs sorted according to highest values for each indicator. The graphs in Fig. 1 show cumulative distribution for all 112 PIs.

PI rank	SU(M)	$SU \ge \%$	TC	$TC \ge \%$	hx	$hx\geq\%$
61	1.00	92.47	0.55	94.98	11.00	83.02
62	1.00	92.79	0.49	95.25	11.00	83.68
63	1.00	93.10	0.45	95.50	11.00	84.35
64	0.93	93.40	0.42	95.74	10.00	84.95
65	0.93	93.69	0.39	95.95	10.00	85.55
66	0.82	93.96	0.39	96.17	10.00	86.15
67	0.80	94.21	0.37	96.38	10.00	86.75
68	0.80	94.47	0.35	96.57	10.00	87.36
69	0.75	94.71	0.34	96.76	10.00	87.96
70	0.73	94.94	0.34	96.95	9.00	88.50
71	0.70	95.16	0.34	97.14	9.00	89.04
72	0.70	95.38	0.34	97.33	9.00	89.58
73	0.70	95.61	0.33	97.51	8.00	90.07
74	0.69	95.83	0.32	97.69	8.00	90.55
75	0.62	96.03	0.31	97.86	8.00	91.03
76	0.60	96.22	0.31	98.03	8.00	91.51
77	0.60	96.41	0.29	98.20	7.00	91.93
78	0.54	96.58	0.28	98.35	7.00	92.35
79	0.50	96.74	0.26	98.50	7.00	92.78
80	0.50	96.90	0.23	98.62	7.00	93.20
81	0.50	97.06	0.21	98.74	6.00	93.56
82	0.50	97.22	0.21	98.86	6.00	93.92
83	0.50	97.38	0.20	98.97	6.00	94.28
84	0.50	97.54	0.18	99.07	6.00	94.64
85	0.45	97.68	0.18	99.17	6.00	95.00
86	0.44	97.82	0.17	99.27	6.00	95.36
87	0.44	97.96	0.16	99.36	6.00	95.73
88	0.43	98.10	0.15	99.44	6.00	96.09
89	0.43	98.23	0.13	99.51	6.00	96.45
90	0.40	98.36	0.13	99.58	6.00	96.81

J. Bollen et al. / Journal of Informetrics xxx (2010) xxx-xxx

Table A.4

Raw values and cumulative percentages of SU allocated (SU) in millions (M), Total Cites (TC), and h-indices (hx) for 112 PIs sorted according to highest values for each indicator. The graphs in Fig. 1 show cumulative distribution for all 112 PIs.

PI rank	SU(M)	$SU \ge \%$	TC	$TC \ge \%$	hx	$hx\geq\%$
91	0.40	98.49	0.12	99.65	5.00	97.11
92	0.40	98.62	0.11	99.71	5.00	97.41
93	0.40	98.75	0.09	99.76	4.00	97.65
94	0.32	98.85	0.07	99.80	4.00	97.89
95	0.30	98.94	0.05	99.83	4.00	98.13
96	0.30	99.04	0.05	99.86	4.00	98.37
97	0.30	99.13	0.05	99.88	3.00	98.56
98	0.29	99.23	0.04	99.90	3.00	98.74
99	0.28	99.32	0.03	99.92	3.00	98.92
100	0.27	99.40	0.02	99.94	3.00	99.10
101	0.26	99.49	0.02	99.95	3.00	99.28
102	0.26	99.57	0.02	99.96	2.00	99.40
103	0.23	99.64	0.02	99.97	2.00	99.52
104	0.21	99.71	0.02	99.98	2.00	99.64
105	0.20	99.77	0.01	99.99	2.00	99.76
106	0.18	99.83	0.01	99.99	1.00	99.82
107	0.15	99.88	0.01	100.00	1.00	99.88
108	0.10	99.91	0.00	100.00	1.00	99.94
109	0.10	99.94	0.00	100.00	1.00	100.00
110	0.10	99.97	0.00	100.00	0.00	100.00
111	0.05	99.99	0.00	100.00	0.00	100.00
112	0.04	100.00	0.00	100.00	0.00	100.00

References 115

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M. A., et al. (2009). Clickstream data yields high-resolution maps of science. 116

PLoS ONE, 4 (3), e4803. 117

Bollen, J., Van de Sompel, H., & Rodriguez, M. A. (2008, June). Towards usage-based impact metrics: First results from the MESUR project. In *Joint Conference* on *Digital Libraries (JCDL2006)* Pittsburgh, PA, (pp. 231–240). 118 119

120 Butler, D. (2009, March). Web usage data outline map of knowledge. Nature News

Agghe, L. (2006). Theory and practice of the g-index. Scientometrics, 69(1), 131–152. 121

Harzing, A. W. K., & van der Wal, R. (2008, June). Google scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, 8, 61–73. Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572. 122 123 Lawrence, K. A., & Zimmerman, A. (2007). Teragrid planning process report: June 2007 workshop for science gateways. Technical report: School of Information, University of Michigan http://deephine.lib.umich.edu/biteteese/2007 workshop for science gateways. Technical report: School of Information, 124

Zimmerman, A., & Finholt, T. A. (2006). Peragrid user work-shop – final report. Technical report, School of Information, http://www.crew.umich.edu/ 126 127 research/teragrid_user_workshop.pdf.

University of Michigan, http://deepblue.lib.umich.edu/bitstream/2027.42/61843/1/TeraGrid_ScienceGateways_Workshop_Report.pdf. 125