Evaluation in Computer Science.

J. Díaz (UPC)
Why to evaluate in Academia?

- Hiring
- Promotion
- Grant evaluation
Whom to evaluate in Academia?

- Institutions
- Research Groups
- Individual Scientists
- Individual Papers
- Journals
Whom to evaluate in Academia?

- Institutions
  - Research Groups.
  - Individual Scientists
    - Individual Papers
      - Journals
How to evaluate?

Research is too important to rely on subjective judgments.

Evaluation must be:

- Objective
- Simple
- Easy to implement and implement
- Minimal time consuming
How to evaluate?

Research is too important to rely on subjective judgments.

Evaluation must be:

- **Objective**
- Simple
- Easy to implement and implement
- Minimal time consuming
How to evaluate?

Used to be Peer Review, tends towards Bibliometrics as primary source.
How to evaluate?

Used to be Peer Review, tends towards Bibliometrics as primary source.

Is that the correct way?

Research is too important to measure its value using bibliometrics alone
Bibliometrics

Main and most used:

- **ISI-web-of-knowledge** (Thomson-Reuters) (subscription based)
  Evaluates journals, papers appeared in a DB of 9000 peer-reviewed journals, and researcher productivity for publications in journals, in all fields of science. It uses their own DB.

- **Scopus** (Elsevier) (subscription based)
  Indexes 16000 peer-reviewed journals in most fields of science.

- **Google Scholar** (free)
  Web search engine that retrieves information for all accessible documents in the web. It includes all documents available in the web. Prone to all kinds of errors: (use of Tech. Reports, problems with non-english names, ...)

Bibliometrics

- CiteSeer ($\text{CiteSeer}^X(\beta)$) (PennState) (free).
  It crawls open access DB, as DBLP. Evaluates and uses journal, conferences and books.
  Focuses on Computer and Information Science.
  Similar problems to Google Scholar.
Bibliometrics

- **CiteSeer (CiteSeerX(β)) (PennState) (free).**
  It crawls open access DB, as DBLP. Evaluates and uses journal, conferences and books.
  Focuses on Computer and Information Science.
  Similar problems to Google Scholar.

Other bibliometric search engines: Scirus, getCITED, etc...
Impact Factor (IF): (Garfield Factor)
The IF of J in 2008 = number of references during 2008 to papers published in J during the $t$ previous years, divided by the total number of papers published in J during the same period.
Impact Factor (IF): (Garfield Factor)
The IF of J in 2008 = number of references during 2008 to papers published in J during the $t$ previous years, divided by the total number of papers published in J during the same period.
In Garfield proposal $h = 2$, which is good for biological and medical journals, but it is equivocal for CS and Math.
Scopus: uses only journals and conferences after 1994.
ISI-web-of-knowledge: uses only journals and a 5 year period.
Recently, started to incorporate conferences.
Evaluation of Researcher

Based on counting number of citations of the researcher

The Hirsch number (H-index): A researcher has H-index=$h$ if $h$ of his papers are cited at least each $h$ times each.

Jorge Hirsh gives empirical evidence that all Nobel Prizes have large H-index.

Today’s trend is to use basically the $H$-index to evaluate CV at Universities and Research Institutions (tenures, promotions,...) and even to rate Universities.
Evaluation of Researcher

Computation of H-index:
- ISI-web-of-knowledge:
- Scopus
- Scholar Index (Google Scholar)
Evaluation of Researcher

Computation of H-index:
ISI-web-of-knowledge:
Scopus
Scholar Index (Google Scholar)
g-index, AW-index (age-weighted citation rate), etc..

Publish or Perish (uses Google Scholar as underling support)
A large body of literature has been developed that explains the dangers of using the described bibliometrics for evaluation of Journals, Researchers and Institutions.


The case of Computer Science

Computer science has as much to do with computers as astronomy has to do with telescopes.

*Edsger Dijkstra*

Computer Science is the only discipline that cannot be defined in a single sentence.

*Christos Papadimitriou*
The case of Computer Science: Peculiarities

The format and relevance of conferences in CS:
Big difference with prestigious conferences in related fields
Mathematics, Physics, etc.
The heterogeneity of the field: Informatics is science and engineering.
Could be analytical and experimental
Sometimes, it is difficult for people working in one area to understand work done in other areas of CS. For ex. people working in Algorithmics and in Algorithmic Engineering.
This has created small clusters of work with different methodologies and goals.
Conferences in CS

- **IJCAI** (Art. Intell.): 16%
- **ECAI** (Art. Intell.): 26%
- **ICSE** (Soft. Engineer.): 13%
- **ICDCS** (Distributed Comp.): 15%
- **OOPSLA** (Object Technology): 19%
- **POPL** (Prog. Lang.): 19%
- **AAMAS** (Agents): 22%
- **ICCAD** (Graphics): 26%
- **ICML** (Mach. Learning): 27%
- **ICRA** (Robotics): 43%
- **ICDM** (Data Mining): 20%
- **CVPR** (Vision- Recognition): 4%
- **INFOCOM** (Computer Communicat.): 17%
- **DAC** (Design autom.): 24%
- **IPDPS** (Parallel-Dist.): 26%
- **STOC** (Theory): 25%
- **SODA** (Theory): 29%
- **ICALP** (Theory): 33%
Conferences in CS

For CS, conferences should be fully introduced into the DB
Conferences in CS

For CS, conferences should be fully introduced into the DB. Scopus, Google already do, ISI-web-kn is starting to do it.

BUT
Conferences in CS

For CS, conferences should be fully introduced into the DB. Scopus, Google already do, ISI-web-kn is starting to do it. BUT

- Proliferation of conferences exchanging payment and acceptance (ex. WSEAS series)
- Even in the best conferences, plenty of wrong papers: Too many papers to handle by the same PC member in short time (use students as referees).
  For most papers, it is difficult to fit all details in 10 pages.
Example: A. Broder in a STOC-98 paper, used Markov chain Monte-Carlo methods to approximate hard counting problems.
Example: A. Broder in a STOC-98 paper, used Markov chain Monte-Carlo methods to approximate hard counting problems. M. Mihail in IPL-99, proved a key step in Broder’s paper was wrong.
Example: A. Broder in a STOC-98 paper, used Markov chain Monte-Carlo methods to approximate hard counting problems. M. Mihail in IPL-99, proved a key step in Broder’s paper was wrong. M. Jerrum and A. Sinclair in FOCS-99 fixed the convergence using a completely different proof.
Example: A. Broder in a STOC-98 paper, used Markov chain Monte-Carlo methods to approximate hard counting problems. M. Mihail in IPL-99, proved a key step in Broder’s paper was wrong. M. Jerrum and A. Sinclair in FOCS-99 fixed the convergence using a completely different proof.

*I consider that the paper of A. Broder, even wrong, is one of the most important contributions to TCS in the last 10 years*
Example: A. Broder in a STOC-98 paper, used Markov chain Monte-Carlo methods to approximate hard counting problems. M. Mihail in IPL-99, proved a key step in Broder’s paper was wrong. M. Jerrum and A. Sinclair in FOCS-99 fixed the convergence using a completely different proof.

I consider that the paper of A. Broder, even wrong, is one of the most important contributions to TCS in the last 10 years. Papers in good conferences must be considered mainly by their novel ideas. Most of the papers make it into long journal versions.
The IF in "Theory" CS Journals

<table>
<thead>
<tr>
<th>Mark</th>
<th>Rank</th>
<th>Abbreviated Journal Title (linked to journal information)</th>
<th>ISSN</th>
<th>Total Cites</th>
<th>Impact Factor</th>
<th>Immediacy Index</th>
<th>Articles</th>
<th>Cited Half-life</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>ACM COMPUT SURV</td>
<td>0360-0300</td>
<td>2169</td>
<td>5.250</td>
<td>1.083</td>
<td>12</td>
<td>8.2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>J ACM</td>
<td>0004-5411</td>
<td>4894</td>
<td>3.136</td>
<td>0.312</td>
<td>32</td>
<td>&gt;10.0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>IEEE T NEURAL NETW</td>
<td>1045-9227</td>
<td>7019</td>
<td>2.769</td>
<td>0.189</td>
<td>164</td>
<td>8.1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>HUM-COMPUT INTERACT</td>
<td>0737-0024</td>
<td>960</td>
<td>2.476</td>
<td>1.455</td>
<td>11</td>
<td>7.2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>IEEE T EVOLUT COMPUT</td>
<td>1089-778X</td>
<td>2203</td>
<td>2.426</td>
<td>0.130</td>
<td>46</td>
<td>5.7</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>REAL-TIME IMAGING</td>
<td>1077-2014</td>
<td>284</td>
<td>2.270</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>QUANTUM INF COMPUT</td>
<td>1533-7146</td>
<td>709</td>
<td>1.988</td>
<td>1.653</td>
<td>49</td>
<td>3.8</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>ACM T COMPUT SYST</td>
<td>0734-2071</td>
<td>779</td>
<td>1.917</td>
<td>0.091</td>
<td>11</td>
<td>&gt;10.0</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>FOUND COMPUT MATH</td>
<td>1615-3375</td>
<td>163</td>
<td>1.833</td>
<td>0.059</td>
<td>17</td>
<td>4.5</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>IEEE ACM T NETWORK</td>
<td>1063-6692</td>
<td>3689</td>
<td>1.831</td>
<td>0.202</td>
<td>124</td>
<td>7.5</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>J CRYPTOL</td>
<td>0933-2790</td>
<td>848</td>
<td>1.824</td>
<td>0.353</td>
<td>17</td>
<td>9.0</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>ARTIF LIFE</td>
<td>1064-5462</td>
<td>784</td>
<td>1.644</td>
<td>0.409</td>
<td>22</td>
<td>7.6</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>COMMUN ACM</td>
<td>0001-0782</td>
<td>8969</td>
<td>1.593</td>
<td>0.343</td>
<td>140</td>
<td>&gt;10.0</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>EVOL COMPUT</td>
<td>1063-6560</td>
<td>1327</td>
<td>1.575</td>
<td>0.000</td>
<td>19</td>
<td>8.5</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>CONNECT SCI</td>
<td>0954-0091</td>
<td>316</td>
<td>1.500</td>
<td>0.222</td>
<td>18</td>
<td>7.1</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>DISTRIB COMPUT</td>
<td>0178-2770</td>
<td>484</td>
<td>1.458</td>
<td>0.242</td>
<td>33</td>
<td>8.6</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>FUZZY SET SYST</td>
<td>0165-0114</td>
<td>7401</td>
<td>1.373</td>
<td>0.237</td>
<td>194</td>
<td>9.9</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>INT J HIGH PERFOM C</td>
<td>1094-3420</td>
<td>374</td>
<td>1.319</td>
<td>0.152</td>
<td>33</td>
<td>5.0</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>SIAM J COMPUT</td>
<td>0097-5397</td>
<td>4081</td>
<td>1.310</td>
<td>0.255</td>
<td>55</td>
<td>&gt;10.0</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>IBM SYST</td>
<td>0018-8670</td>
<td>1007</td>
<td>1.214</td>
<td>1.020</td>
<td>49</td>
<td>6.7</td>
</tr>
</tbody>
</table>
The IF of 3 main Theory CS Journals

Data from SCOPUS on 3 Theory Journals: JACM, SIAMJC, TCS
The IF of 3 main Theory CS Journals

Data from SCOPUS on 3 Theory Journals: JACM, SIAMJC, TCS
The IF of 3 main Theory CS Journals

Data from SCOPUS on 3 Theory Journals: JACM, SIAMJC, TCS

Number Papers Published
# The IF of CS Journals

## JOURNAL RANKINGS IN COMPUTER SCIENCE

<table>
<thead>
<tr>
<th>View</th>
<th>Journal</th>
<th>Papers</th>
<th>Citations</th>
<th>Citations Per Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BIOINFORMATICS</td>
<td>4,473</td>
<td>95,512</td>
<td>21.35</td>
</tr>
<tr>
<td>2</td>
<td>LECT NOTE COMPUT SCI</td>
<td>87,567</td>
<td>84,434</td>
<td>0.96</td>
</tr>
<tr>
<td>3</td>
<td>IEEE TRANS INFORM THEORY</td>
<td>3,352</td>
<td>33,517</td>
<td>10.00</td>
</tr>
<tr>
<td>4</td>
<td>COMPUT METHOD APPL MECH ENG</td>
<td>3,189</td>
<td>23,828</td>
<td>7.47</td>
</tr>
<tr>
<td>5</td>
<td>IEEE J SEL AREA COMMUN</td>
<td>1,886</td>
<td>22,257</td>
<td>11.80</td>
</tr>
<tr>
<td>6</td>
<td>IEEE TRANS COMMUN</td>
<td>2,718</td>
<td>22,007</td>
<td>8.10</td>
</tr>
<tr>
<td>7</td>
<td>IEEE TRANS ANTENNAS PROPAGAT</td>
<td>3,799</td>
<td>19,234</td>
<td>5.06</td>
</tr>
<tr>
<td>8</td>
<td>BMC BIOINFORMATICS</td>
<td>2,039</td>
<td>13,659</td>
<td>6.70</td>
</tr>
<tr>
<td>9</td>
<td>COMPUT CHEM ENG</td>
<td>2,265</td>
<td>12,825</td>
<td>5.66</td>
</tr>
<tr>
<td>10</td>
<td>THEOR COMPUT SCI</td>
<td>3,990</td>
<td>12,186</td>
<td>3.05</td>
</tr>
<tr>
<td>11</td>
<td>IEEE COMMUN MAG</td>
<td>1,774</td>
<td>11,599</td>
<td>6.54</td>
</tr>
<tr>
<td>12</td>
<td>SIAM J SCI COMPUT</td>
<td>1,290</td>
<td>10,822</td>
<td>8.39</td>
</tr>
<tr>
<td>13</td>
<td>IEEE ACM TRANS NETWORKING</td>
<td>914</td>
<td>10,692</td>
<td>11.70</td>
</tr>
<tr>
<td>14</td>
<td>COMMUN ACM</td>
<td>1,747</td>
<td>10,387</td>
<td>5.95</td>
</tr>
<tr>
<td>15</td>
<td>COMPUT STRUCT</td>
<td>2,334</td>
<td>9,855</td>
<td>4.22</td>
</tr>
<tr>
<td>16</td>
<td>COMPUT MATH APPL</td>
<td>3,094</td>
<td>8,161</td>
<td>2.64</td>
</tr>
<tr>
<td>17</td>
<td>IEEE COMMUN LETT</td>
<td>2,167</td>
<td>7,556</td>
<td>3.49</td>
</tr>
<tr>
<td>18</td>
<td>ACM TRANS GRAPHIC</td>
<td>808</td>
<td>7,219</td>
<td>8.93</td>
</tr>
<tr>
<td>19</td>
<td>INFORM SCIENCES</td>
<td>1,877</td>
<td>6,138</td>
<td>3.27</td>
</tr>
<tr>
<td>20</td>
<td>COMPUT AID DES</td>
<td>979</td>
<td>6,023</td>
<td>6.15</td>
</tr>
</tbody>
</table>
Evaluation of the individual research in CS

The list of the 20 highest H-number for CS scholars (Jens Palsberg):

(89) Hector Garcia-Molina (Stanford), (87) Jeffrey D. Ullman (Stanford), (82) Robert Tarjan (Princeton), (80) Deborah Estrin (UCLA), (79) C. Papadimitriou (Berkeley), (77) Don Towsley (U Mass, Amherst), (73) Ian Foster (Argonne N. L.), (71) Scott Shenker (Berkeley), (70) David Culler (Berkeley), (70) David Haussler (Santa Cruz) (68) Takeo Kanade (CMU), (61) Mario Gerla (UCLA), (61) Nick Jennings (U Southampton), (58) Luca Cardelli (Microsoft), (58) Anil K. Jain (Michigan S U), (57) Martin Abadi ( Microsoft), (57) Sushil Jajodia (George Mason U), (57) Leslie Lamport (Microsoft), (57) Demetri Terzopoulos (UCLA), (56) Randy H. Katz (Berkeley),
Turing awards

2007 Clarke-Emerson-Sifakis  1993 Hartmanis-Stearns  1979 K. Iverson
2004 Cerf-Kahn  1990 F.Corbato  1976 Rabin-Scott
1995 M. Blum  1981 E.F.Codd  1967 M.V.Wilkes
### Scientist Rankings in Computer Science

**Display items with at least:** 0 Citation(s)

**Sorted by:** Citations

<table>
<thead>
<tr>
<th>View</th>
<th>Scientist</th>
<th>Papers</th>
<th>Citations</th>
<th>Citations Per Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kumar, S</td>
<td>91</td>
<td>7,759</td>
<td>85.26</td>
</tr>
<tr>
<td>2</td>
<td>Nel, M</td>
<td>4</td>
<td>7,501</td>
<td>1,875.25</td>
</tr>
<tr>
<td>3</td>
<td>Tamura, K</td>
<td>8</td>
<td>7,477</td>
<td>934.62</td>
</tr>
<tr>
<td>4</td>
<td>Posada, D</td>
<td>6</td>
<td>7,014</td>
<td>1,169.00</td>
</tr>
<tr>
<td>5</td>
<td>Crandall, K</td>
<td>2</td>
<td>6,621</td>
<td>3,310.50</td>
</tr>
<tr>
<td>6</td>
<td>Huelserbeck, JP</td>
<td>3</td>
<td>6,395</td>
<td>2,131.67</td>
</tr>
<tr>
<td>7</td>
<td>Ronquist, F</td>
<td>3</td>
<td>6,395</td>
<td>2,131.67</td>
</tr>
<tr>
<td>8</td>
<td>Tarokh, V</td>
<td>33</td>
<td>3,280</td>
<td>99.39</td>
</tr>
<tr>
<td>9</td>
<td>Jakobsen, IB</td>
<td>2</td>
<td>3,238</td>
<td>1,619.00</td>
</tr>
<tr>
<td>10</td>
<td>Calderbank, AR</td>
<td>27</td>
<td>3,022</td>
<td>111.93</td>
</tr>
<tr>
<td>11</td>
<td>Rozas, I</td>
<td>4</td>
<td>2,286</td>
<td>571.50</td>
</tr>
<tr>
<td>12</td>
<td>Rozas, R</td>
<td>2</td>
<td>2,275</td>
<td>1,137.50</td>
</tr>
<tr>
<td>13</td>
<td>Tse, DNC</td>
<td>34</td>
<td>2,271</td>
<td>66.79</td>
</tr>
<tr>
<td>14</td>
<td>Seshadri, N</td>
<td>11</td>
<td>1,784</td>
<td>162.18</td>
</tr>
<tr>
<td>15</td>
<td>Jafarkhani, H</td>
<td>20</td>
<td>1,697</td>
<td>84.85</td>
</tr>
<tr>
<td>16</td>
<td>Hochwald, BM</td>
<td>16</td>
<td>1,548</td>
<td>96.75</td>
</tr>
<tr>
<td>17</td>
<td>Daly, M</td>
<td>3</td>
<td>1,518</td>
<td>506.00</td>
</tr>
<tr>
<td>18</td>
<td>Barrett, JC</td>
<td>2</td>
<td>1,486</td>
<td>743.00</td>
</tr>
<tr>
<td>19</td>
<td>Fry, B</td>
<td>2</td>
<td>1,478</td>
<td>739.00</td>
</tr>
<tr>
<td>20</td>
<td>Maller, I</td>
<td>1</td>
<td>1,476</td>
<td>1,476.00</td>
</tr>
</tbody>
</table>

Copyright © 2008 The Thomson Corporation
Person $X$ H-index:

*ISI-web-of-knowledge*: 48  
*Scopus*: 5  
*Scholar*: 42

\[ X = \]
Person X H-index:

ISI-web-of-knowledge: 48
Scopus: 5
Scholar: 42
On Journals and Conferences

- The **IF** is useful as a first approximation.
- DB should be pruned of self-references.
- Editorial Boards should be professionalized
  Ed. Boards members should receive a fee for the job
  Referees should receive an honorarium for a good job
  Strict honorability rules should be implemented for behavior in
  Editorial Boards and Program Committees.
- There should be an IF for good CS conferences
On Journals and Conferences

- The IF is useful as a first approximation.
- DB should be pruned of self-references.
- Editorial Boards should be professionalized
  Ed. Boards members should receive a fee for the job
  Referees should receive an honorarium for a good job
  Strict honorability rules should be implemented for behavior in
  Editorial Boards and Program Committees.
- There should be an IF for good CS conferences
- *The expertise of the people working in a specific field, is the
  most important measure of evaluation for a given Journal and
  Conference* (assuming the honesty of the researcher!)
Personal Conclusions 2: 
On Research Evaluation

- Bibliometrics as the $H$-number are good as a lower bound on the capacity of researchers, but it is pretty much helpless evaluating the research abilities above a certain threshold.
- Parameters as the $H$-number could be improved by returning to computing it in an ad-hoc manner (difficult).
- In Bibliometrics should incorporate a "negative" score for facts as excessive self-referencing and plagiarism (two of the modern plagues affecting CS research). This is best done by careful interpretation of the existing DATA.
- In CS recognition of seminal ideas could take a long time.
- The fashion for publishing as fast as possible and as much as possible should be reversed in CS. In that sense, the evaluation must focus in: novel ideas, steps forward, novel techniques, etc.
A "wrong paper" with interesting ideas that stimulate plenty of researchers is more important for the progress of CS than a paper who nobody cares even to validate its results.

The design of good useful software should be incorporated into the evaluation of researchers in CS.
Personal Conclusions 3:

On the Researchers

- The fashion of publishing as fast as possible and as much as possible should be reversed in CS. In the long run, a few good ideas will be better for your reputation than 100 irrelevant publications.

- As John McCarthy says: *Excessive self reference is often as dangerous as smoking.*

- Plagiarism and self-plagiarism is the most dangerous game for the self reputation.

- Concurrent multitasking of Editorial Boards and Program Committees is dangerous for the fairness of the process.
Thank you.