Figures of Merit*

Martin Tompa
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, New York 10598

*A preliminary version of these results was presented at the 29^{th} Annual Symposium on Follies of Computer Science, White Plains, New York, October 24, 1988.

1 In the Spotlight

> *Watch out for the fellow who talks about putting things in order! Putting things in order always means getting other people under your control.*
>
> – *Denis Diderot, 1796*

In Theoretical Computer Science it is customary to alphabetize the names of coauthors on the title page of collaborative publications. For those of us whose name will never appear before the phrase *et al.*, this can be cause for concern.

Given the custom of alphabetizing, and under the assumption that surnames are distributed uniformly over the alphabet, one would expect to see such collaborations as

$$\boxed{\text{Beeri, Mendelzon, Sagiv, and Ullman [4]}}\ .$$

On the other hand, something seems amiss with the collaboration

$$\boxed{\text{Vishkin and Wigderson [36]}}\ .$$

After all, if Uzi Vishkin chose a coauthor at random, what is the probability that that coauthor's surname would occur later in the alphabet than Vishkin? Armed with this example, we make the following definition.

Definition: In a collaboration of $X_0 < X_1 < \cdots < X_k$, the *spotlight factor* of X_0 is

$$\star(X_0) = (1 - .X_0)^k \quad .^*$$

In words, the spotlight factor is the probability that k coauthors chosen independently at random will all have surnames later in the alphabet than X_0; the lower the spotlight factor, the more impressive the achievement of the first author.

---

*In this definition, the notation "$\mathbf{.}$Vishkin", for example, is of course the radix 27 fraction, where $a = 1$, $b = 2$, $\ldots$, $z = 26$, and blanks and punctuation represent 0.

$\star(X_0)$	X_0 *et al.*
0.1889	Ogden, Riddle, & Rounds [20]
0.1719	Vishkin & Wigderson [36]
0.1640	Paul, Seiferas, & Simon [24]
0.12685	Wong & Yao [37]
0.12680	Wood & Yap [38]
0.1214	Karmarkar, Karp, Lipton, Lovász, & Luby [14]
0.0919	Ruzzo, Simon, & Tompa [28]
0.0851	Schwartz, Sharir, & Siegel [30]
0.0664	Paul, Pippenger, Szemerédi, & Trotter [21]
0.0255	Santoro, Sidney J., Sidney S., & Urrutia [29]

Table 1: The Top Ten Spotlight Factors

Example 1 *In the aforementioned collaboration of Vishkin and Wigderson [36],*

$$\begin{aligned} \star(\textit{Vishkin}) &= 1 - .\textit{Vishkin} \\ &= 1 - \left(\frac{22}{27^1} + \frac{9}{27^2} + \frac{19}{27^3} + \cdots\right) \\ &\approx 0.1719 \end{aligned}$$

In contrast, the less exotic collaboration of Beeri, Mendelzon, Sagiv, and Ullman [4] achieves a more modest spotlight factor:

$$\begin{aligned} \star(\textit{Beeri}) &= (1 - .\textit{Beeri})^3 \\ &= \left(1 - \left(\frac{2}{27^1} + \frac{5}{27^2} + \frac{5}{27^3} + \cdots\right)\right)^3 \\ &\approx 0.7756\Box \end{aligned}$$

Table 1 lists the top ten spotlight factors in the Theoretical Computer Science community. A two-author collaboration that exceeds the record of 0.0255 would have to be at least Zippel and Zuck, which would achieve $\star$(Zippel) $\approx$ 0.0238. At the other extreme of the alphabet, Adleman would have to collaborate with 84 other coauthors in order to achieve a spotlight

Paul, Pippenger, Szemerédi, & Trotter [21]
Paul, Seiferas, & Simon [24]
Paul & Reischuk [23]
Paul, Prauß, & Reischuk [22]
Paul & Tarjan [25]
Paul, Tarjan, & Celoni [26]

Table 2: A Sampling of Wolfgang Paul's Collaborations

factor of 0.0246, and he would have to ensure that the list didn't include Aanderaa or Abelson.

Note that the only researcher who appears twice in the top ten of Table 1 is Wolfgang Paul. As corroboration of the statistical significance of the spotlight factor, a random sampling of Paul's publications is listed in Table 2.

This section closes with a particularly reprehensible form of the spotlight phenomenon, namely, those professors who willfully choose their graduate student advisees with surnames later in the alphabet than theirs, hoping to cash in at the time of future collaborations. Some examples are given in Table 3.

There is one known instance in which a resourceful Ph.D. student named Yehuda outspotlighted his advisor Shimon Even. When it came time to publish the results of their collaboration, Even announced his inevitable intention of being first author. Yehuda responded by legally changing his name to Bar-Yehuda [3].

2 Out of the Spotlight

One of the most pernicious effects of haste is obscurity.

– Samuel Johnson, 1752

After completing our careful study of the spotlight factor, we were settling back to rest on our laurels when the collaboration

Advisor	*Advisee*
Paul	Prauß
	Reisch
	Reischuk
	Rollig
	Schnitger
Shamir	Snir
	Upfal
Shiloach	Vishkin
Tompa	Venkateswaran
	Woll
Ullman	Yannakakis

Table 3: Reprehensible Form of Spotlight Factor

Brassard & Crépeau [8]

came to light. Something seemed biased about this collaboration, but in this case the fault cannot be Brassard's, who is likely to be first author in any collaboration. Closer inspection revealed Crépeau to be the culprit. After all, if Crépeau chose a coauthor at random, what is the probability that that coauthor's surname would upstage his by as little as Brassard's did? Armed with this example, we make the following definition.

Definition: In a collaboration of $X_0 < X_1 < \cdots < X_k$, the *coefficient of obliviousness* of X_i is

$$¿(X_i) = (.X_i - .X_0)^i,$$

for $1 \leq i \leq k$.

In words, the coefficient of obliviousness is the probability that i coauthors chosen independently at random will all have surnames that precede X_i as narrowly as does X_0; the lower the coefficient, the more oblivious X_i is to fame.

Example 2 *In the aforementioned collaboration of Brassard and Crépeau [8],*

$$¿(\textit{Crépeau}) \quad = \quad .\textit{Crépeau} - .\textit{Brassard}$$

$$
\begin{aligned}
&= \frac{1}{27^1} + \frac{0}{27^2} + \frac{4}{27^3} - \frac{3}{27^4} + \cdots \\
&\approx 0.0372
\end{aligned}
$$

In contrast, our standard example of Beeri, Mendelzon, Sagiv, and Ullman [4] achieves a more modest coefficient of obliviousness:

$$
\begin{aligned}
¿(\textit{Ullman}) &= (.\textit{Ullman} - .\textit{Beeri})^3 \\
&= \left(\frac{19}{27^1} + \frac{7}{27^2} + \frac{7}{27^3} - \frac{5}{27^4} - \cdots\right)^3 \\
&\approx 0.3635\Box
\end{aligned}
$$

Table 4 lists the top coefficients of obliviousness in the Theoretical Computer Science community.

3 The Fundamental Theorem

There's a sucker born every minute.

– *attributed to Phineas T. Barnum*

Comparing Tables 1 and 4, it is apparent that many of the collaborations that occur in the former also occur in the latter, and that in all cases the constants are much smaller in the latter. This leads us to the

Fundamental Theorem of Collaborative Sociology: If $X_0 < X_1 < \cdots < X_k$ agree to collaborate, X_k's obliviousness to fame exceeds X_0's appetite for fame.

Proof: $(.X_k - .X_0)^k < (1 - .X_0)^k$ $\Box$

4 Monotone Erdös Number

These studies raise the question of a natural variation of the well known "Erdös number". Define a directed graph $G = (V, E)$, where V is the set of all researchers, and $(u, v) \in E$ if and only if there is some publication in which u appears earlier in the list of coauthors than v.

¿(X_k)	*al. et* X_k
0.0372	Brassard & Crépeau [8]
0.0369	Borodin & Cook [6]
0.0364	Vishkin & Wigderson [36]
0.0334	Ladner & Lynch [16]
0.0278	Bhatt & Cai [5]
0.0185	Alon & Azar [2]
0.0169	Garey, Graham, & Johnson [10]
0.0151	Ogden, Riddle, & Rounds [20]
0.0149	Kung & Leiserson [15]
0.0148 *	Paul, Seiferas, & Simon [24]
0.0095	Aggarwal & Anderson [1]
0.0086	Shamir & Snir [31]
0.0071	Solovay & Strassen [32]
0.0062	Paul, Prauß, & Reischuk [22]
0.0050	Paul, Pippenger, Szemerédi, & Trotter [21]
0.0043	Ruzzo, Simon, & Tompa [28]
9.30×10^{-4}	Santoro, Sidney J., Sidney S., & Urrutia [29]
2.23×10^{-4}	Kahn, Klawe, & Kleitman [13]
2.00×10^{-4}	Lenstra A., Lenstra H., & Lovász [19]
6.57×10^{-5}	Schwartz, Sharir & Siegel [30]
2.24×10^{-5}	Leighton & Leiserson [18]
1.64×10^{-5}	Karmarkar, Karp, Lipton, Lovász, & Luby [14]
1.40×10^{-6}	Brassard & Bratley [7]
3.58×10^{-7}	Yao A. & Yao F. [39]
3.39×10^{-7} †	Goldreich, Goldwasser, & Micali [11]
8.56×10^{-13} ‡	Lenstra A., Lenstra H., & Lovász [19]
4.17×10^{-15} §	Vazirani U. & Vazirani V. [35]
1.53×10^{-15}	Plumstead B. & Plumstead J. [27]

*A popular conjecture at this time was that the limit of ¿(X) was $\left(\frac{5 \ln \pi}{27}\right)^e \approx$ 0.01475.

† ¿(Goldwasser)

‡ ¿(Lenstra H.)

§A long-standing conjecture held that this was the quantum of obliviousness, based on the remarkable coincidence that it is Planck's constant when measured in electron volts – seconds. This conjecture was finally disproved by Joan Boyar [personal communication].

Table 4: The Top Coefficients of Obliviousness

1.	Erdös, *Graham*, & Szemerédi [9]
2.	Graham, *Lawler*, Lenstra, & Rinnooy Kan [12]
3.	Lawler, *Tarjan*, & Valdez [17]
4.	Tarjan & *Vishkin* [33]
5.	Vishkin & *Wigderson* [36]

Table 5: Monotone Erdös Numbers

Definition: The *monotone Erdös number* of X is the length of a longest directed path in G between Paul Erdös and X.†

Table 5 illustrates that Wigderson's monotone Erdös number is 5. We know of no one in the Theoretical Computer Science community with a greater finite value.

5 Conclusions

> *Figures often beguile me, particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli would often apply with justice and force: "There are three kinds of lies: lies, damned lies, and statistics."*
>
> – *Mark Twain, 1924*

This paper introduced the following new statistics:

1. Spotlight factor
2. Coefficient of obliviousness
3. Monotone Erdös number
4. Johnson number

Definition: The *Johnson number* $\mathcal{J}(X)$ is the number of statistics that X has studied before David Johnson.

†The ordinary Erdös number is the length of a *shortest* path in the undirected version of G. In contrast, the challenge in this new definition is to find some X with as great a finite value as possible.

Theorem: $\mathcal{J}(\text{Tompa}) = 4$.

Acknowledgements

I thank the numerous contributors who inundated me with suggested additions after the Follies presentation. I can only wish that my serious research would stimulate half as much enthusiasm in the community.

References

[1] Alok Aggarwal and Richard J. Anderson. A Random *NC* algorithm for depth first search. *Combinatorica*, 8(1):1–12, 1988.

[2] Noga Alon and Y. Azar. Parallel comparison algorithms for approximation problems. In *29th Annual Symposium on Foundations of Computer Science*, pages 194–203, White Plains, NY, October 1988. IEEE.

[3] R. Bar-Yehuda and S. Even. A linear-time approximation algorithm for the weighted vertex cover problem. *Journal of Algorithms*, 2:198–203, 1981.

[4] C. Beeri, A. O. Mendelzon, Yehoshua Sagiv, and Jeffrey D. Ullman. Equivalence of relational database schemes. In *Conference Record of the Eleventh Annual ACM Symposium on Theory of Computing*, pages 319–329, Atlanta, GA, April-May 1979.

[5] Sandeep N. Bhatt and Jin-Yi Cai. Taking random walks to grow trees in hypercubes. *Journal of the ACM*, 40(3):741–764, July 1993.

[6] Allan Borodin and Stephen A. Cook. A time-space tradeoff for sorting on a general sequential model of computation. *SIAM Journal on Computing*, 11(2):287–297, May 1982.

[7] Gilles Brassard and P. Bratley. *Algorithmics: Theory and Practice.* Prentice-Hall, Inc., 1988.

[8] Gilles Brassard and Claude Crépeau. Non-transitive transfer of confidence: A perfect zero-knowledge interactive protocol for SAT and beyond. In *27th Annual Symposium on Foundations of Computer Science*, pages 188–195, Toronto, Ontario, October 1986. IEEE.

[9] P. Erdös, Ronald L. Graham, and Endre Szemerédi. On sparse graphs with dense long path. *Comp. and Maths. with Appls.*, 1:365–369, 1975.

[10] Michael R. Garey, R. L. Graham, and David S. Johnson. Some *NP*-complete geometric problems. In *Conference Record of the Eighth Annual ACM Symposium on Theory of Computing*, pages 10–22, Hershey, PA, May 1976.

[11] Oded Goldreich, Shaffi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, 1986.

[12] Ronald L. Graham, Eugene L. Lawler, Jan K. Lenstra, and A. H. G. Rinnooy Kan. Optimization and approximation in deterministic sequencing and scheduling: A survey. In P. L. Hammer, E. L. Johnson, and B. H. Korte, editors, *Discrete Optimization II, Annals of Discrete Mathematics*, 5, pages 287–326. North-Holland Publishing Company, 1979.

[13] Jeff D. Kahn, Maria M. Klawe, and D. Kleitman. Traditional galleries require fewer watchmen. *SIAM Journal on Algebraic and Discrete Methods*, 4:194–206, 1980.

[14] N. Karmarkar, Richard M. Karp, Richard J. Lipton, László Lovász, and Michael Luby. A Monte Carlo algorithm to approximate the permanent. Technical report, University of Toronto, 1988.

[15] H. T. Kung and Charles E. Leiserson. Algorithms for VLSI processor arrays. In *Symposium on Sparse Matrix Computations and Their Applications*, November 1978.

[16] Richard E. Ladner and Nancy A. Lynch. Relativization of questions about log space computability. *Mathematical Systems Theory*, 10(1):19–32, 1976.

[17] Eugene L. Lawler, Robert E. Tarjan, and J. Valdes. Analysis and isomorphism of series parallel digraphs. (Eventually appeared as [34]), 1982.

[18] F. Thomson Leighton and Charles E. Leiserson. Wafer-scale integration of systolic arrays. In *23rd Annual Symposium on Foundations of Computer Science*, pages 297–311, Chicago, IL, November 1982. IEEE.

[19] A. K. Lenstra, H. W. Lenstra, Jr., and László Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, December 1982.

[20] William F. Ogden, W. E. Riddle, and W. C. Rounds. Complexity of expressions allowing concurrency. In *Conference Record of the Fifth Annual ACM Symposium on Principles of Programming Languages*, pages 185–194, Tucson, AZ, January 1978.

[21] Wolfgang J. Paul, Nicholas J. Pippenger, Endre Szemerédi, and W. T. Trotter. On determinism versus non-determinism and related problems. In *24th Annual Symposium on Foundations of Computer Science*, pages 429–438, Tucson, AZ, November 1983. IEEE.

[22] Wolfgang J. Paul, E. J. Prauß, and Rüdiger Reischuk. On alternation. *Acta Informatica*, 14:243–255, 1980.

[23] Wolfgang J. Paul and Rüdiger Reischuk. On alternation II. *Acta Informatica*, 14:391–403, 1980.

[24] Wolfgang J. Paul, Joel I. Seiferas, and Janos Simon. An information theoretic approach to time bounds for on-line computation. *Journal of Computer and System Sciences*, 23:108–126, 1981.

[25] Wolfgang J. Paul and Robert E. Tarjan. Time-space trade-offs in a pebble game. *Acta Informatica*, 10:111–115, 1978.

[26] Wolfgang J. Paul, Robert E. Tarjan, and J. R. Celoni. Space bounds for a game on graphs. *Mathematical Systems Theory*, 10(3):239–251, 1977. Correction, *ibid.* 11(1):85, 1977.

[27] B. Plumstead and J. Plumstead. Bounds for cube coloring. *SIAM Journal on Algebraic and Discrete Methods*, 6(1), January 1985.

[28] Walter L. Ruzzo, Janos Simon, and Martin Tompa. Space-bounded hierarchies and probabilistic computations. *Journal of Computer and System Sciences*, 28(2):216–230, April 1984.

[29] N. Santoro, J. B. Sidney, S. J. Sidney, and J. Urrutia. Geometric containment and vector dominance. *Theoretical Computer Science*, 53:345–352, 1987.

[30] J. T. Schwartz, M. Sharir, and A. Siegel. An efficient algorithm for finding connected components in a binary image. Technical Report TR 154, New York University, 1985.

[31] E. Shamir and Marc Snir. On the depth complexity of formulas. *Mathematical Systems Theory*, 13:301–322, 1980.

[32] R. Solovay and V. Strassen. A fast Monte-Carlo test for primality. *SIAM Journal on Computing*, 6(1):84–85, March 1977.

[33] Robert E. Tarjan and Uzi Vishkin. An efficient parallel biconnectivity algorithm. *SIAM Journal on Computing*, 14(4):862–874, November 1985.

[34] J. Valdes, Robert E. Tarjan, and Eugene L. Lawler. The recognition of series parallel digraphs. *SIAM Journal on Computing*, 11:298–313, 1982.

[35] Umesh Vazirani and Vijay Vazirani. Random polynomial time is equal to slightly-random polynomial time. In *26th Annual Symposium on Foundations of Computer Science*, pages 417–428, Portland, OR, October 1985. IEEE.

[36] Uzi Vishkin and Avi Wigderson. Trade-offs between depth and width in parallel computation. *SIAM Journal on Computing*, 14(2):303–314, May 1985.

[37] C. K. Wong and A. C. Yao. A combinatorial optimization problem related to data set allocation. *Rev. Francaise Automat. Informat. Recherche Operationnelle Ser. Bleue*, 10.5 (suppl.):83–95, 1976.

[38] D. Wood and Chee-Keng Yap. Computing a convex skull of an orthogonal polygon. In *Proceedings of the First Annual ACM Symposium on Computational Geometry*, pages 311–315, 1985.

[39] A. C. Yao and F. F. Yao. A general approach to geometric queries. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, pages 163–168, Providence, RI, May 1985.