Chains of Fame from Catherine of Siena to Jan Karel Lenstra

Jan Korst Gijs Geleijnse Verus Pronk Emile Aarts

Philips Research Laboratories, HTC 34, 5656 AE Eindhoven, The Netherlands
email: {jan.korst,gijs.geleijnse,verus.pronk,emile.aarts}@philips.com

1 Introduction

For those of us who have had the opportunity to discuss with Jan Karel Lenstra matters extending beyond mathematics will have discovered that Jan Karel takes an interest in more subjects than mathematical models and algorithms only. Most profound is his interest and deep knowledge of classical music in all its dimensions. His collection of more than 2000 CDs is quite amazing. It covers the entire time span of classical music ranging from the early anonymous middle-aged flute pieces up to work by contemporary composers such as Louis Andriessen. Also his interest in books deserves mentioning. His extensive personal collection of books reveals his large interest in a broad range of subjects. In addition to their content he also collects books for their design. He has become truly fascinated by typography and typesetting and takes great interest in various forms of information representation and visualization. Over the years he has extended this passion for shape and form from a passive to an active level by applying his insights and view to his own publications.

In addition to all these fine qualifications, which show that Jan Karel is a truly well-educated person, there is another remarkable characteristic, which has to do with famous people. One could say that he has this "thing" with fame. Using his well-developed internal reference guide, Jan Karel is quite outspoken when it comes to assessing peoples' fame, leading to disapproval of those who in his opinion do not deserve fame as well as to full-hearted support of those who have no fame but deserve it. This issue about fame inspired us to write a short contribution for this book of friends with the subject "Chains of Fame".

2 Hunting for famous dead people on the web

One of the more playful activities we carried out over the last few years is collecting information on historical figures by web information extraction - the activity that we usually refer to ourselves by 'hunting for famous dead people on the web'.

We first collected a large set of historical figures that lived during the past millennium by simply querying Google with queries of the type "(y_1 - y_2)", with y_1 and y_2 integers satisfying $1000 \leq y_1 < y_2 \leq 2005$ and $15 \leq y_2 - y_1 \leq 114$. For each of these approximately 100,000 queries, the at most 1000 text snippets that Google returns are analyzed to identify person names that directly precede the query issued. In this way, query *"(1685 - 1750)"* results in text snippets including *'Johann Sebastian Bach (1685 - 1750)'*. Candidate person names are simply sequences of one to four capitalized words (allowing additional infixes as *'van der'*). These candidate person names are next analysed to filter out non-person names (see Table 1). Next we quantified the fame of each remaining person, which is simply defined by the estimated page count that Google returns on the query that includes the last name and the years of birth and death. For example, the fame of Johann Sebastian Bach is chosen to be the estimated page count on query *"Bach (1685 - 1750)"*, assuming that this expression uniquely identifies a historical figure. Note that 'last name' is the last word in the word sequence. The above strategy resulted in a large collection of historical figures ranked by fame. In addition, we extracted additional personal details as gender, nationality, and one or more professions for each of the the top 10,000 persons as well as a corresponding image. This allowed us to make ranked sublists of, for example, Dutch painters of the 17th century.

person name	*non-person name*
Art Blakey	Art Deco
Mae West	West Virginia
Amy Beach	Miami Beach
HP Lovecraft	HP Inkjet

Table 1: Person names and non-person names are sometimes hard to distinguish.

1. Blaise Pascal (1623 - 1662)	9. John von Neumann (1903 - 1957)
2. Carl Friedrich Gauss (1777 - 1855)	10. Henri Poincaré (1854 - 1912)
3. Leonard Euler (1707 - 1783)	11. Alfred North Whitehead (1861 - 1947)
4. John Maynard Keynes (1883 - 1946)	12. Georg Cantor (1845 - 1918)
5. Pierre-Simon Laplace (1749 - 1827)	13. Joseph Fourier (1768 - 1830)
6. Alan Turing (1912 - 1954)	14. Joseph Louis Lagrange (1736 - 1813)
7. Pierre de Fermat (1601 - 1665)	15. David Hilbert (1862 - 1943)
8. George Boole (1815 - 1864)	16. Bernhard Riemann (1826 - 1866)

Table 2: Top-16 of persons that have as best scoring profession 'mathematician'.

As an example we give in Table 2 the top-16 of persons that have as best scoring profession 'mathematician'. If we would give the top of persons that have as one of their professions 'mathematician', then Rene Descartes and Isaac Newton would have been number 1 and 2 and also Kepler, Kelvin, Huygens, Lichtenberg, Husserl and Caroll would have been in the top-20. For further technical details, the reader is referred to [1,2].

For this special occasion, we decided to focus on the year of Jan Karel's birth: 1947. We first have a look at famous people that died in 1947 and next look at chains of fame ending in 1947.

3 The year 1947

Table 3 gives the top-20 persons that died in 1947. It is a remarkable list, not only containing two famous mathematicians (Whitehead and Hardy), but also two nobel-prize winning physicists (Planck and Lenard), two female writers (Cather and Huch), two French painters (Bonnard and Marquet), two composers (Hahn and Casella), two psychologists (Lewin and Janet) and two persons of infamous reputation (Crowley and Capone).

1. Henry Ford (1863 - 1947)	11. Kurt Lewin (1890 - 1947)
2. Max Planck (1858 - 1947)	12. Willa Cather (1873 - 1947)
3. Alfred North Whitehead (1861 - 1947)	13. Ricarda Huch (1864 - 1947)
4. Aleister Crowley (1875 - 1947)	14. Arthur Machen (1863 - 1947)
5. Pierre Bonnard (1867 - 1947)	15. Al Capone (1899 - 1947)
6. Godfrey Harold Hardy (1877 - 1947)	16. Philipp Lenard (1862 - 1947)
7. Reynaldo Hahn (1875 - 1947)	17. Albert Marquet (1875 - 1947)
8. Nicholas Roerich (1874 - 1947)	18. Jozef Tiso (1887 - 1947)
9. Alfredo Casella (1883 - 1947)	19. Pierre Janet (1859 - 1947)
10. Ernst Lubitsch (1892 - 1947)	20. Wolfgang Borchert (1921 - 1947)

Table 3: The 20 persons that died in 1947 with the highest fame.

4 Chains of fame ending in 1947

Having a large collection of historical figures, it is tempting to look for peculiarities such as persons that have the same birth and death years, such as Christopher Wren (1632 - 1723) and Antony van Leeuwenhoek

(1632 - 1723). Other famous pairs are Claude Debussy / Gustav Klimt (1862 - 1918), James Joyce / Virginia Woolf (1882 - 1941), Pablo Picasso / Ludwig von Mises (1881 - 1973) and Winston Churchill / William Somerset Maugham (1874 - 1965). Relating to 1947, probably the most interesting pair is Irving Fischer / Stanley Baldwin (1867 - 1947). The latter was three times prime minister of the United Kingdom. The former was a well-known American economist of high reputation. However, a few days before the great stock market crash in 1929, he famously predicted "the stock prices have reached what looks like a permanently high plateau." This and other personal details in this paper are all taken from Wikipedia.

Another peculiarity is that of perfect chains, which are defined as follows. Let M denote the collection of persons, where for each person $i \in M$ his or her birth year $b(i)$, death year $d(i)$ and fame $f(i)$ are given.

Definition 1 [chains].
A sequence $(p_1, p_2, \ldots, p_n)$ of persons from M is called a chain if and only if

$$d(p_i) \leq b(p_{i+1}) \qquad i = 1, 2, \ldots, n-1.$$

A chain is called perfect if and only if $d(p_i) = b(p_{i+1})$ for $i = 1, 2, \ldots, n-1$. □

We are interested in finding chains that have a high overall fame. Since the strength of a chain is determined by its weakest link, we choose as primary objective to maximize the minimum fame, i.e., we look for chains that maximize $\min_{i=1,\ldots,n} f(p_i)$. As a secondary objective function, when multiple chains have the same maximum minimum fame, we want to select from these a chain that maximizes the sum of fame, i.e., that maximizes $\sum_{i=1,\ldots,n} f(p_i)$.

Problem 1 [chain problem].
Given is a collection M of historical persons, where for each person $i \in M$ birth year, death year, and fame are given by integers $b(i)$, $p(i)$, and $f(i)$, respectively. Also, given are start year s, end year e, and upperbound u. Now, the problem is to find a chain $(p_1, p_2, \ldots, p_n)$ satisfying

$$\begin{array}{rcl} b(p_1) & = & s \\ d(p_n) & = & e \\ d(p_i) & \leq & b(p_{i+1}) \leq d(p_i) + u, \quad i = 1, \ldots, n-1 \end{array}$$

that maximizes as primary objective $\min_{i=1,\ldots,n} f(p_i)$, and as secondary objective $\sum_{i=1,\ldots,n} f(p_i)$. □

Again, focusing on Jan Karel's birth year, we look for chains ending in 1947, i.e., the chains for which Jan Karel can be the next link. Table 4 gives an example of an optimal chain for $s = 1547$, $e = 1947$ and $u = 0$.

Miguel de Cervantes (1547 - 1616)
John Wallis (1616 - 1703)
Jonathan Edwards (1703 - 1758)
Horatio Nelson (1758 - 1805)
Hans Christian Andersen (1805 - 1875)
Aleister Crowley (1875 - 1947)

Table 4: An optimal perfect chain from 1547 to 1947.

Of these, John Wallis has the minimum fame. Wallis was a British mathematician who is given partial credit for the development of modern calculus. Amongst others, he is still known for the Wallis' product, given by

$$\prod_{n=1}^{\infty} \frac{(2n)(2n)}{(2n-1)(2n+1)} = \frac{\pi}{2}.$$

Figure 1 gives the results for $s = 1347$, $e = 1947$, and different values of u. We observe that larger values of u generally result in higher values of $\min_{i=1,\ldots,n} f(p_i)$. However, the maximum minimum fame cannot increase beyond $\min(\max_{i \in M, b(i)=1347} f(i), \max_{i \in M, d(i)=1947} f(i))$. When this value has been reached, we can still improve the secondary objective.

It is intriguing to see that most of these chains start with saint Catherine of Siena and end with Aleister Crowley, the British occultist, writer, philosopher, and mystic, who is infamously dubbed "The Wickedest Man In the World." Catherine of Siena is the patron saint of fire prevention and of Italy. Most of Catherine's

body is currently interred in Rome (Basilica Santa Maria sopra Minerva), except for her head and right thumb (Siena), and her left foot (Venice).

5 Computational complexity

The chain problem can be solved by using dynamic programming. We assume that the persons in M are given ordered by birth year.

Let us first consider the case $u = 0$, i.e., no gaps are allowed between successive persons. By using an array of d years, with $d = e - s$, we can simply determine for all years $y \in \{s, s+1, \ldots, e\}$ the score of the best chain (if one exists) from s to y in $O(m+d)$ time, where m denotes the number of persons in M. By storing for each year, a pointer to the last person in this best chain (to backtrack to this person's birth year), we can easily reconstruct an optimal chain by backtracking from e. Note that this gives a pseudo-polynomial running time, since d is not polynomial in the size of the problem input. An alternative approach using a balanced binary tree can be shown to run in $O(m \log m)$ time.

The case $u > 0$ is somewhat more complicated. Using again an array of d years, we now can determine for each year y the score of the best chain from s to y in $O(u \cdot m + d)$ time. Here, however, we have to store for each year y two pointers, relating to the best chain ending at y with a person and the best chain ending at y with a gap. Also, for this case, an alternative approach exists that runs in $O(m \log m)$ time.

6 Conclusion

Some of us have had the opportunity to work with Jan Karel for more than two decades in a variety of roles both professional and personal. In hindsight, they have been good years with many highlights. If one would have to mention one single aspect that has played a significant role in our personal interaction and which can be marked as truly outstanding, then this would undoubtedly be Jan Karel's well-developed sense for quality. With great perseverance he applied the most strong quality measures not only to himself but also to others with whom he was working, and it is our full conviction that his persistent quality control has led to many outstanding results. Hence, we would like to conclude by expressing the wish that Jan Karel should be widely recognized for his intrinsic high-quality standard, and we sincerely hope that this will further increase his fame.

References

1. Korst, J., G. Geleijnse, N. de Jong, and M. Verschoor, Ontology-based information extraction from the world wide web, Chapter 10, in: W. Verhaegh, E. Aarts, and J. Korst (Eds.), *Intelligent Algorithms in Ambient and Biomedical Computing*, Springer, 2006.
2. Geleijnse, G., and J. Korst, Creating a Dead Poets Society: Extracting a social network of historical figures from the web, in: K. Aberer et al. (Eds.), *Proceedings of the 6th International Semantic Web Conference*, LNCS 4825, pp. 156-168.

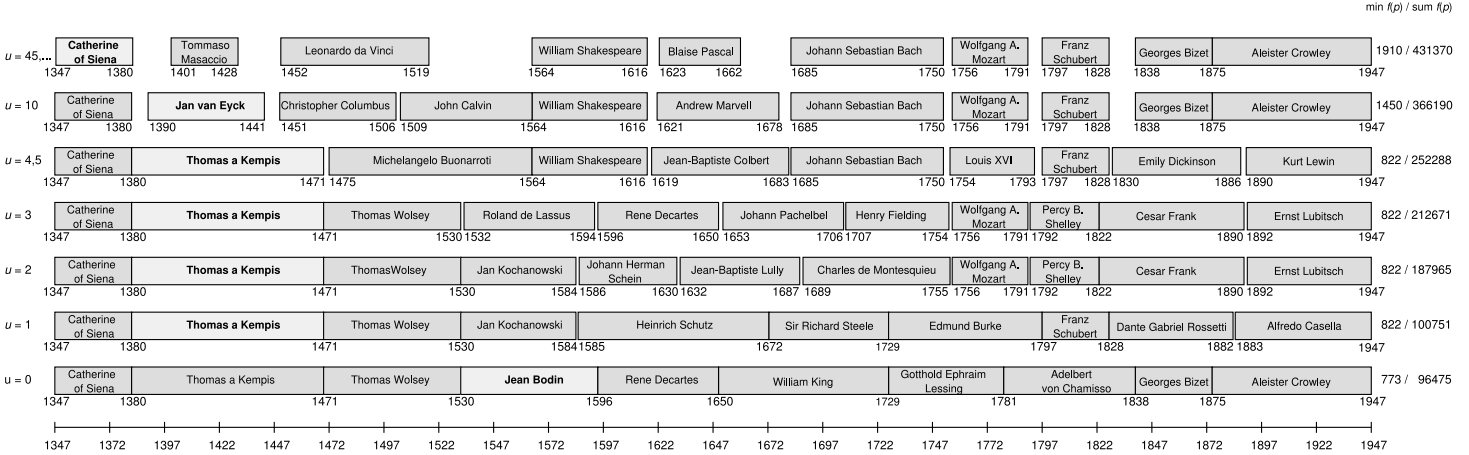

Figure 1: Optimal chains for different values of u. The weakest link is given in bold.