CONCUR THROUGH TIME

Luca Aceto
ICE-TCS, Department of Computer Science,
Reykjavik University
Gran Sasso Science Institute, L'Aquila
luca@ru.is, luca.aceto@gssi.it

Pierluigi Crescenzi
Gran Sasso Science Institute, L'Aquila
pierluigi.crescenzi@gssi.it

1 Introduction

The 33rd edition of the International Conference on Concurrency Theory (CONCUR) will be held in Warsaw, Poland, in the period 16–19 September 2022. The first CONCUR conference dates back to 1990 and was one of the conferences organized as part of the two-year ESPRIT Basic Research Action 3006 with the same name. The CONCUR community has run the conference ever since and established the IFIP WG 1.8 "Concurrency Theory" in 2005 under Technical Committee TC1 Foundations of Computer Science of IFIP[1].

In light of the well-established nature of the CONCUR conference, and spurred by a data- and graph-mining comparative analysis carried out by the second author to celebrate the 50th anniversary of ICALP[2], we undertook a similar study for the CONCUR conference using some, by now classic, tools from network science. Our goal was to try and understand the evolution of the CONCUR conference throughout its history, the ebb and flow in the popularity of some research areas in concurrency theory, and the centrality of CONCUR authors, as measured by several metrics from network science, amongst other topics.

This article reports on our findings. We hope that members of the CONCUR community will enjoy reading it and playing with the web-based resources that accompany this piece. It goes without saying that the data analysis we present has to be taken with a huge pinch of salt and is only meant to provide an overview of the evolution of CONCUR and to be food for thought for the concurrency theory community.

The paper is organized as follows. Section 2 describes the data collection and mining software used for the analysis presented in our study. Section 3 details the evolution of the number of CONCUR papers and authors per year, and Section 4 reports on our findings related to the representation of female authors at the conference. We present data on the evolution of popular research topics in papers presented at CONCUR in Section 5 by analyzing the words

[1] See https://pure.tue.nl/ws/portalfiles/portal/4345371/589768.pdf and https://concurrency-theory.org/organizations/ifip for information on the ESPRIT project and the IFIP "Concurrency Theory" working group, respectively.

[2] See the presentation available at https://slides.com/piluc/icalp-50?token=fl3BBJ8j.

appearing in the paper titles. Section 6 is devoted to a study of the CONCUR collaboration graph. We conclude the article by applying several centrality measures from network science to identify the "most central figures" in the CONCUR community (Section 7).

2 Data collection and mining software

The data collection software has been developed in Java, mostly because this allowed us to take advantage of the Java library available on the DBLP web site[3]. (All the generated graphs are based on the DBLP XML file dated March 1, 2022, and up to the 2021 edition of CONCUR[4].) Note that, even if the first CONCUR took place in August 1990, the collected data include also the papers published in three events devoted to concurrency that took place in July 1984, October 1988, and September 1989, respectively[5]. The basic data mining software has been developed in Julia. Both the Java code and the Julia code are publicly available at the following GitHub repository: `https://github.com/piluc/ConferenceMining`.

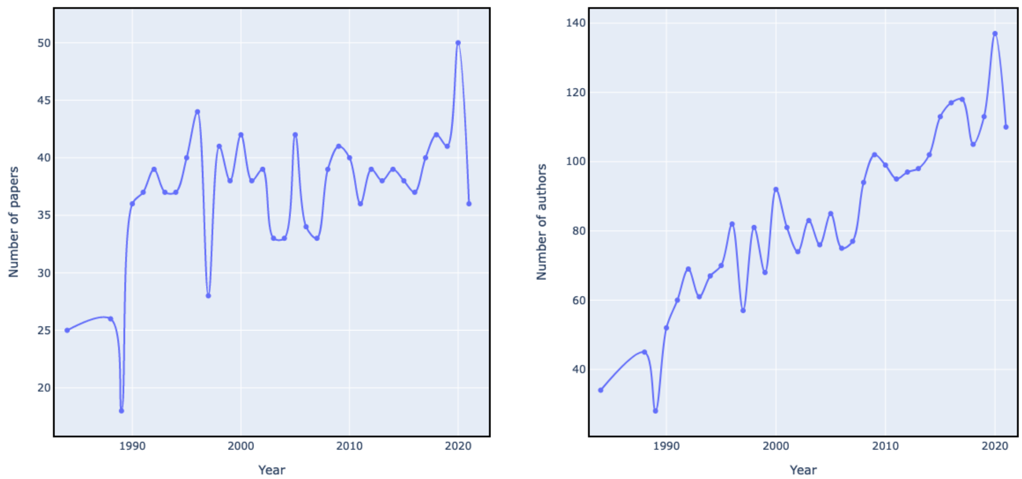

Figure 1: The evolution of the number of CONCUR papers (left) and the number of authors per year (right).

3 Evolution of paper and author numbers

The evolution of the number of CONCUR papers per year is shown in the left part of Figure 1, while the evolution of the number of authors per year is depicted in the right part of that figure. We observe that, while the number of papers per year has been rather stable (approximately

[3]See `https://dblp.org/faq/1474681.html`.

[4]The tables of contents of all the editions of CONCUR are available on the DBLP web site, starting from `https://dblp.org/db/conf/concur/index.html`. The structure of the DBLP XML file, instead, is described in M. Ley, "DBLP – Some Lessons Learned", *Proc. VLDB Endow.*, 2(2): 1493-1500 (2009).

[5]These three events, which predate the first CONCUR conference, are called *Concurrency: Theory, Language, And Architecture* (1989: Oxford, UK), *Concurrency* (1988: Hamburg, Germany), and *Seminar on Concurrency* (1984: Pittsburgh, PA, USA), respectively.

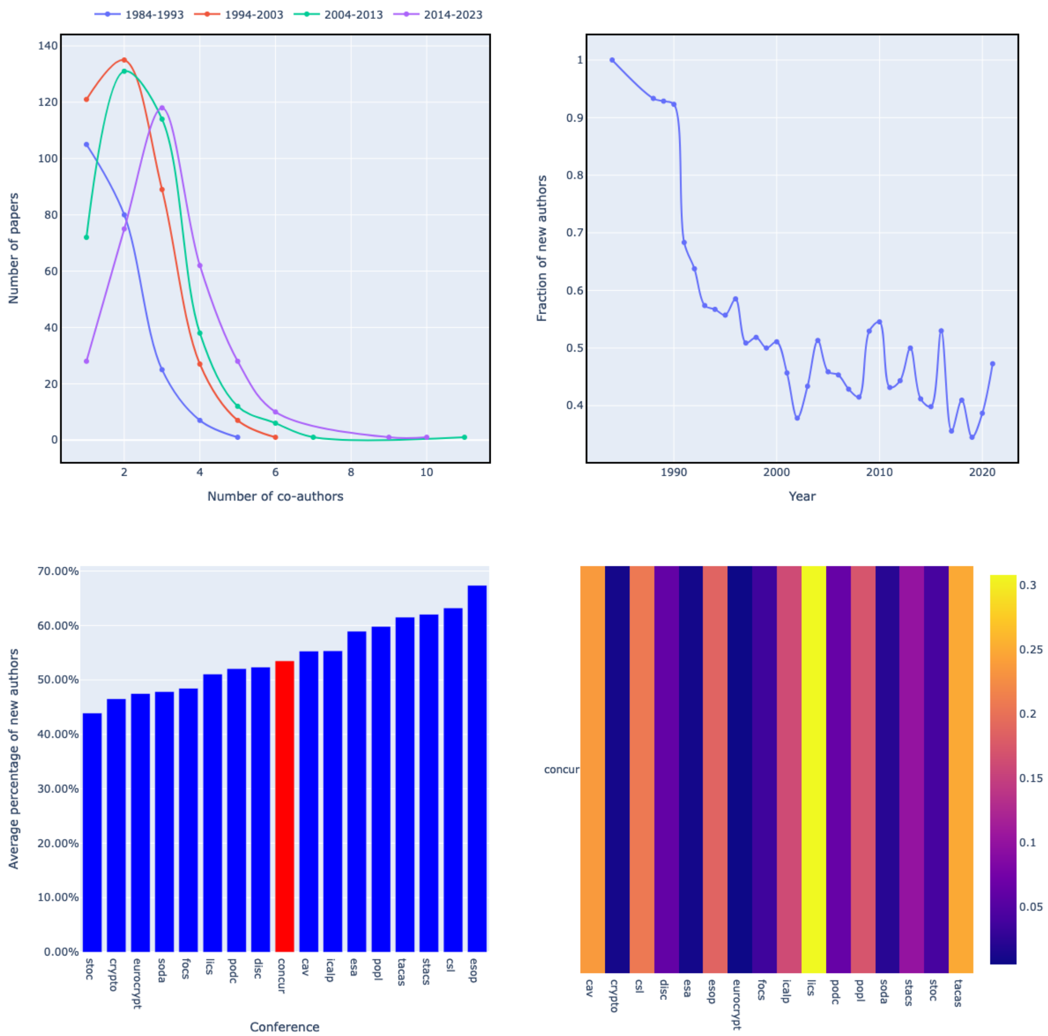

Figure 2: The evolution of the number of co-authors per decade (top left), the percentage of new authors per year (top right), the average number of new authors compared with 16 theoretical computer science conferences (bottom left), and the values of the Sørensen-Dice similarity index with respect to the same 16 conferences (bottom right).

38), the number of authors more than doubled (from 52 to 110). This is probably justified by the fact that the number of co-authors per paper has increased significantly over the years, as it is shown in the top left part of Figure 2. Indeed, while in the first decade the number of papers with a single author was the majority and the maximum number of co-authors was five, in the last decade the papers with two, three, and even four authors have become more popular than single-author papers. At the same time, the maximum number of co-authors has increased to ten. As indicated by a similar data- and graph-mining analysis for ICALP and other major conferences in theoretical computer science reported at `https://slides.com/piluc/icalp-50?token=fl3BBJ8j#/2/5`, papers authored by two to four researchers are now more frequent than singly-authored ones in all fields of the theory of computing.

The top right part of Figure 2 shows the evolution of the percentage of *new* distinct authors of the published papers per year. This percentage decreased and stabilized between 40% and

Figure 3: The evolution of the percentage of male and female authors per year (the two percentages are computed with respect to the number of authors for which the sex has been assigned). The percentage of authors with no sex assigned is also shown (with respect to the total number of authors).

50%. In other words, every year approximately half of the authors of the CONCUR conference are new authors. (Note that, in this analysis, we are not considering the co-authorship between authors, that is, we are not verifying whether the new authors have been "introduced" by an author who already published in the conference.) The percentage of new authors for several conferences in theoretical computer science is available at `https://slides.com/piluc/icalp-50?token=fl3BBJ8j#/2/3`. We find it noteworthy that the percentage of new authors for 11 of the conferences considered in that plot is above 50% (see also the bottom left part of Figure 2, where the bar corresponding to CONCUR is shown in red).

Finally, the bottom right part of the figure shows the values of the Sørensen-Dice index of similarity computed by comparing the set of CONCUR authors with the sets of authors for sixteen theoretical computer science conferences[6]. As it can be seen, the conference that is most similar to CONCUR is LICS (with Sørensen-Dice index approximately equal to 0.3), followed by TACAS (approximately 0.25), CAV (approximately 0.24), and CSL (approximately 0.21). The least similar conferences to CONCUR are, instead, EUROCRYPT, ESA, and CRYPTO (all below 0.01).

4 Sex analysis

The sex of CONCUR authors has been determined mostly by querying the web service available at `genderize.io` (which is based on first names only), and partly by manually searching

[6]Given two sets A and B, the Jaccard index $J(A,B)$ is equal to $\frac{|A\cap B|}{|A\cup B|}$, and the Sørensen-Dice index is equal to $\frac{2J(A,B)}{1+J(A,B)}$ (see T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons", *Kongelige Danske Videnskabernes Selskab.*, 5 (4): 1–34 (1948), and L.R. Dice, "Measures of the Amount of Ecologic Association Between Species", *Ecology*, 26 (3): 297–302 (1945)).

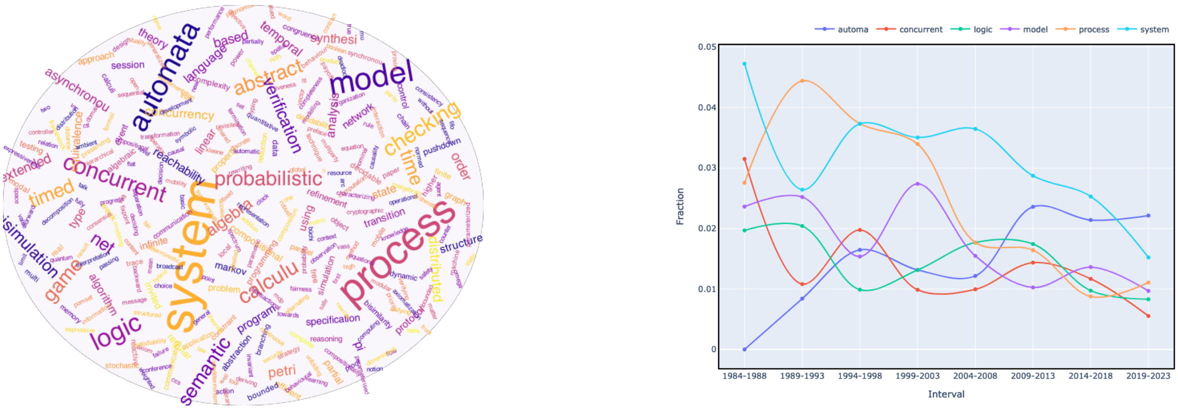

Figure 4: The word cloud corresponding to the words contained in the titles of CONCUR papers (left) and the evolution of fractions of occurrences per five-year interval of the six words globally most frequent (right).

the authors on the web. At the end of this phase, almost all authors have been assigned a sex (which should not be confused with their gender—see, for instance, the interview with Judith Butler at `https://www.youtube.com/watch?v=Bo7o2LYATDc`). Figure 3 shows the evolution of the percentages of male and female authors per year (the percentage of authors with no sex assigned is also shown). The percentage of female authors increased from approximately 6% to approximately 21% over the years. However, the number of women is still approximately only one fifth of the total number of authors, which maybe indicates that some reflections have to be done on this subject[7]. Note, however, that, as indicated by the data displayed at `https://slides.com/piluc/icalp-50?token=fl3BBJ8j#/3/1`, these numbers are consistent with the ones of many other theoretical computer science conferences, where the percentage of female authors was below 20% in 2021.

5 Topic analysis

The word cloud corresponding to the words contained in the titles of CONCUR papers is shown in the left part of Figure 4. As it can be seen, the words `automa`, `concurrent`, `logic`, `model`, `process`, and `system` are those that appear more frequently in the title of a CONCUR paper.

Of all the words contained in the titles of CONCUR papers in a certain time interval, the plot on the right part of Figure 4 shows what fraction of them are one of the above most frequent six words. It can be seen that `system` is almost always the most frequent one, while the other five words alternate and three of them have been the most frequent one in at least one time

[7]As mentioned in the recently published opinion article available at `https://www.scientificamerican.com/article/there-are-too-few-women-in-computer-science-and-engineering/`, which summarizes the main findings in the paper Allison Master, Andrew N. Meltzoff, and Sapna Cheryan, "Gender stereotypes about interests start early and cause gender disparities in computer science and engineering", *Proceedings of the National Academy of Sciences* 118 (48) e2100030118 (2021), `https://www.pnas.org/content/118/48/e2100030118`, sex-based stereotypes related to computer science and engineering seem to become entrenched early in life. Indeed, as reported in those studies, children and adolescents in the U.S. already believe that girls are less interested than boys in computer science and engineering. Experiments reported in the above-mentioned PNAS paper indicate that the culture in computer science and engineering contributes to excluding girls and women.

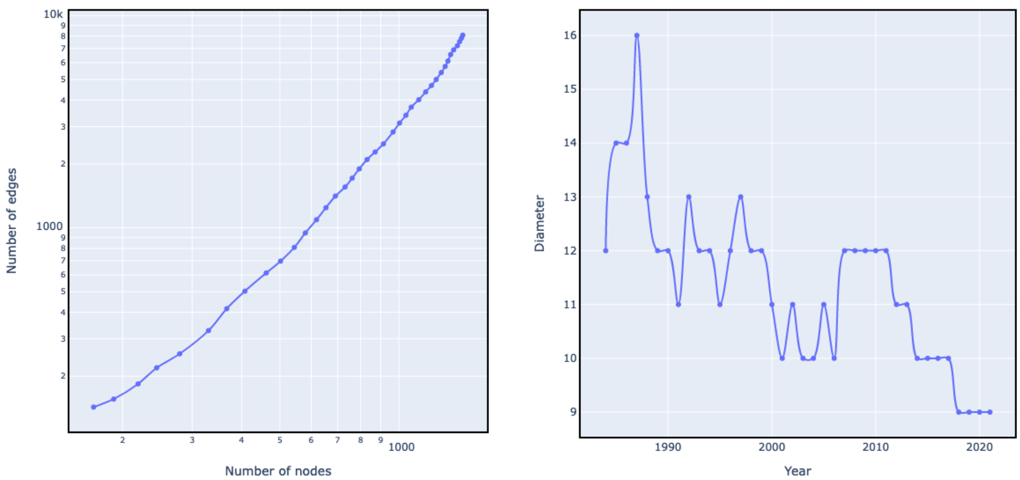

Figure 5: The densification (left) and the diameter shrinking (right) of the collaboration graph of CONCUR authors.

interval. The interested reader can see the evolution of all the words appearing in the title of some CONCUR paper at the web page `http://www.pilucrescenzi.it/concur/word_frequencies_5.html`, where it is also possible to compare the evolution of two different words.

6 Basic graph mining

The *static graph* (or collaboration graph) of CONCUR is an undirected graph whose nodes are the authors who presented at least one paper at CONCUR, and whose edges (a_1, a_2) correspond to two authors a_1 and a_2 who co-authored at least one paper (not necessarily presented at CONCUR). In other words, this graph is the subgraph of the DBLP graph induced by the set of CONCUR authors.

The static graph has 1451 nodes and 8086 edges. It is a sparse graph, since its density[8] is approximately equal to 0.008. It contains a giant connected component, which includes approximately 98% of all nodes.

Two phenomena that have been pointed out in the literature are the *densification* of a social network and the *shrinking* of its diameter[9]. In Figure 5, these two phenomena are represented in the left and the right part of the figure, respectively. Indeed, it can be seen how the number of edges increases more than linearly with respect to the number of nodes, and that the diameter decreases from 12 to 9 (even if the number of nodes increases).

We also compute the evolution of the degrees of separation, that is, the average distance

[8]The density of an undirected graph with n nodes and m edges is $\frac{2m}{n(n-1)}$, that is, the ratio of its number of edges with respect to the maximum number of possible edges. For a definition of most of the notions used in this section and in the next one and for a description of the used algorithms, we refer the interested reader to the lecture notes available at `https://github.com/piluc/GraphMining`.

[9]See J. Leskovec, J.M. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters", *ACM Trans. Knowl. Discov. Data*, 1:1, 2 (2007).

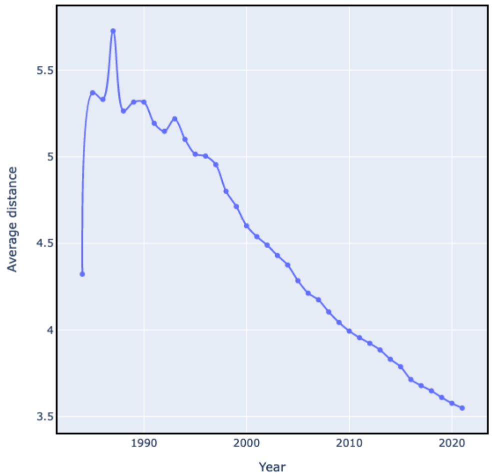

Figure 6: The evolution of the degrees of separation of the collaboration graph of CONCUR authors.

between any two authors in the largest connected component[10]. This evolution (which is similar to the evolution of the diameter) is shown in Figure 6. As it can be seen, the CONCUR community is quite a small world, in which the average distance is currently approximately 3.5.

7 Centrality measures

Centrality measures are a key tool for understanding social networks and are used to assess the "importance" of a given node[11]. In order to quantify the role played by CONCUR authors, we compute the following three different centrality measures on the largest connected component of the static graph.

Degree This is the number of neighbors (that is the number of coauthors).

Closeness This is the average distance from one author to all other authors of its connected component.

Betweenness This is the fraction of shortest paths, passing through one author, between any pair of other authors in its connected component.

In Table 1, we show the top ten CONCUR authors with respect to the above-mentioned three centrality measures in decreasing order. As expected, several authors appear in multiple lists: this is due to the well-known phenomenon of correlation between the centrality measures. It is also interesting to observe that the two female scientists included in the lists, namely Marta Z. Kwiatkowska and Catuscia Palamidessi, appear in the closeness and the betweenness lists. This indicates that they maybe do have fewer coauthors than other "central colleagues", but

[10]The study of the degrees of separation and of the so-called *small-world phenomenon* started with the experiment described in S. Milgram, "The Small World Problem", *Psychology Today*, 1:1, 61–67 (1967).

[11]See L.C. Freeman, "Centrality in social networks conceptual clarification", *Social Networks*, 1, 215—239 (1978).

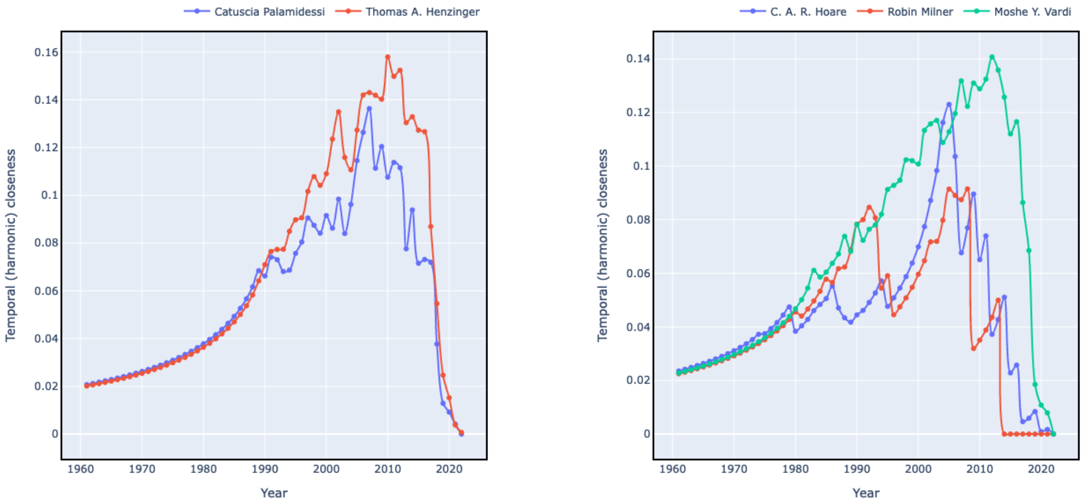

Figure 7: The evolution of the temporal harmonic closeness of Thomas A. Henzinger and Catuscia Palamidessi (left) and of Tony Hoare, Robin Milner, and Moshe Y. Vardi (right).

that their collaborations make them either quite close to the rest of the community or a sort of "bridge". Finally, it might be interesting to determine the centrality of an author by analysing the citation network. However, this network cannot be easily and precisely derived by using only the DBLP data, and other data repositories should be used (such as, for instance, the OpenAlex service available at `https://openalex.org/`).

Degree	**Closeness**	**Betweenness**
Thomas A. Henzinger	Kim G. Larsen	Kim G. Larsen
Kim G. Larsen	Moshe Y. Vardi	Thomas A. Henzinger
Moshe Y. Vardi	Thomas A. Henzinger	Moshe Y. Vardi
Axel Legay	Axel Legay	Javier Esparza
James Worrell	Joost-Pieter Katoen	Catuscia Palamidessi
Krishnendu Chatterjee	Luca Aceto	Axel Legay
Joost-Pieter Katoen	Javier Esparza	Joost-Pieter Katoen
Rupak Majumdar	Marta Z. Kwiatkowska	Luca Aceto
Jean-François Raskin	Catuscia Palamidessi	Rupak Majumdar
Javier Esparza	Rupak Majumdar	Scott A. Smolka

Table 1: The top-10 CONCUR authors with respect to three centrality measures

7.1 Temporal closeness

The *temporal graph* has the same set of nodes of the static graph, but the edges (a_1, a_2, y) correspond to two authors a_1 and a_2 who co-authored in year y at least one paper (not necessarily presented at CONCUR). In the case of this graph, we compute the *temporal closeness*, which is intuitively the area covered by the plot of the temporal harmonic closeness of an author[12].

[12] The *temporal harmonic closeness* of a node u at time t is defined as $\frac{1}{n-1}\sum_{v\neq u}\frac{1}{d_t(u,v)}$, where $d_t(u,v)$ is the time duration of the earliest arrival path starting no earlier than t (see P. Crescenzi, C. Magnien, and A. Marino,

For example, in the left part of Figure 7, the plot of the temporal harmonic closeness of Thomas A. Henzinger and of Catuscia Palamidessi are shown, while the right part depicts the temporal harmonic closeness of Tony Hoare, Robin Milner, and Moshe Y. Vardi. By computing the area covered by these two plots, we may conclude that the temporal closeness of Henzinger is higher than Palamidessi's one. The top ten CONCUR authors with respect to this centrality measure are Moshe Y. Vardi, Kim G. Larsen, Thomas A. Henzinger, Joost-Pieter Katoen, Javier Esparza, Orna Kupferman, Edmund M. Clarke, Ugo Montanari, Rocco De Nicola, and Marta Z. Kwiatkowska.

Several other notions of temporal centrality have been introduced in the literature in the last few years. For instance, the temporal analogue of the betweenness centrality has been deeply analyzed and, since such a measure cannot be efficiently computed even in the case of medium-sized graphs, approximation algorithms based on sampling techniques have been proposed[13]. We believe that it would be interesting to apply these algorithms to the temporal graph of the CONCUR collaborations.

"Finding Top-*k* Nodes for Temporal Closeness in Large Temporal Graphs", *Algorithms*, 13:9, 211, (2020)). Note that in a temporal graph a path is a sequence of edges such that each edge appears later than the edges preceding it.

[13]See S. Buß, H. Molter, R. Niedermeier, and M. Rymar, "Algorithmic Aspects of Temporal Betweenness", *KDD*, 2084–2092 (2020), and D. Santoro and I. Sarpe, "ONBRA: Rigorous Estimation of the Temporal Betweenness Centrality in Temporal Networks", *WWW*, 1579–1588, (2022).