Mengxi Yi
Klaus Nordhausen *Editors*

Robust and Multivariate Statistical Methods

Festschrift in Honor of David E. Tyler

Springer

Robust and Multivariate Statistical Methods

David E Tyler 1986 in Paris and 2015 in Myrtle Beach Published with the kind permission of

Mengxi Yi • Klaus Nordhausen
Editors

Robust and Multivariate Statistical Methods

Festschrift in Honor of David E. Tyler

Springer

Editors
Mengxi Yi
School of Statistics
Beijing Normal University
Beijing, China

Klaus Nordhausen
Department of Mathematics and Statistics
University of Jyväskylä
Jyväskylä, Finland

ISBN 978-3-031-22686-1 ISBN 978-3-031-22687-8 (eBook)
https://doi.org/10.1007/978-3-031-22687-8

Mathematics Subject Classification: 62Hxx, 62F35, 62E20, 62J07

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

It is an honor and a pleasure to contribute this biographical note to the Festschrift dedicated to David (Dave) Tyler on the occasion of his pending retirement from Rutgers University.

I met Dave in the late 1980s through our mutual friend and colleague Javier Cabrera. Our frequent conversations regarding statistics and robustness led to the idea that workshops in robust statistics would be beneficial to the rapidly changing field of statistics. As a result, our first conference on *Robustness and Data Analysis* was held at Princeton University in 1994 with an outstanding list of invited speakers that included John Tukey and Frank Hampel. The success of this first conference encouraged us to continue with international meetings over more than a decade, and Dave's contributions were vital to our creation of the International Conference on Robust Statistics (ICORS) workshops. Today, with the effort and guidance of a younger generation of statisticians, these ICORS workshops continue to thrive.

But Dave's involvement in these conferences went well beyond the usual organizational stage. His deep understanding of statistical issues and his conviction that robust statistics is not merely a subfield of statistics, but rather a school of thought motivated by the realities of data analysis, provided a clear and firm foundation for these meetings. In every conference, Dave interacted extensively with the participants by exchanging ideas and engaging in discussions for the future direction of robust statistics as well as by stressing the importance to further integrate the concept of robustness into the everyday practice of data analysis.

Dave's insight and depth of understanding regarding statistical issues is attested by his many high-quality research publications. From his first paper, "Asymptotic theory of eigenvectors," published in the *Annals of Statistics* in 1981, to his latest research work on robust covariance matrices, his broad knowledge and independent approach are reflected in the long list of research papers published in the most prestigious statistical journals. Taken together, these papers illustrate the wide range of Dave's interests and his continuing influence in statistics, particularly in the areas of multivariate statistics and robustness. His international academic reputation in these areas is evidenced by his countless invitations to conferences, seminars, and special courses at major academic institutions worldwide. His distinguished

academic credentials led to appointments of associate editor at some of the most prestigious statistical journals, including the *Annals of Statistics,* the *Journal of the Royal Statistical Society B,* and the *Journal of Multivariate Analysis.*

Dave's supportive and collaborative attitude toward his colleagues and his deep statistical knowledge earned him the respect and appreciation of major international researchers in statistics, as is documented by his joint publications with distinguished scholars around the world. In addition, his generosity with his ideas is reflected in the numerous doctoral students that he inspired and supervised with their PhD dissertations. The contributors of this Festschrift, many of whom had the privilege to have worked with Dave and to have benefitted from his knowledge as well as his company, enthusiastically offered their research papers for this volume, and this constitutes a testimony of their appreciation and admiration for his stature as a scholar and for him as a person.

Dave's childhood and adolescence merit comment. From the third oldest child of an impoverished family with ten children to a PhD in statistics from Princeton University to a Distinguished Professorship at Rutgers University, Dave's life and career have been quite unusual. He is the only member of his family to have achieved an advanced university degree.

Born and raised in Pittsburgh, Pennsylvania, Dave spent his early school years in Catholic schools, where he developed an early interest in mathematics. When he was 11 years old, due to family circumstances, he and his siblings were sent to a Catholic orphanage, and he remained there for two years. While poverty was a defining state, his mother provided a stabilizing influence in the family. After graduating from an urban public high school, Dave was admitted to Indiana University of Pennsylvania, where he earned a BA in mathematics in 1972 and where later, in 2015, he was awarded a Distinguished Alumni Award for his career accomplishments. In 1972, he continued with graduate studies at the University of Massachusetts, Amherst, and earned a Master's degree from the Mathematics Department in 1974. That same year, at age 24, he married Coleen McCullough, an aspiring young artist of similar religious and social background. Dave then pursued a doctoral program at Princeton University, where in 1979 he was awarded a PhD in statistics. After Princeton, he served as Assistant Professor at the University of Florida (1978,1979) and Old Dominion University (1979–1983). Finally, in 1983 he joined the Statistics Department of Rutgers University, where he was named Distinguished Professor of Statistics in 2004.

Among Dave's striking personal characteristics are modesty, humanity, and total honesty. This was evident not only during my work with him, but also in the social setting, where I met Coleen, an accomplished artist with whom I had a lasting friendship, and his son, Ed. The interaction between our families made me appreciate Dave's human dimension in addition to his outstanding scholarship. Moreover, firmly bound to his modest origins but dedicated to the field of statistics, Dave complemented his colorful personality with numerous interests and activities such as swimming, basketball, chess, hiking, biking, and boating, among others.

On the occasion of his pending retirement from Rutgers University, the institution where he spent most of his career, I wish Dave many more productive years and I look forward to enjoying the pleasure of his professional and personal company for many more years to come.

Philadelphia, PA, USA Luisa Fernholz
June 2022

Preface

We are honoured and delighted to edit this Festschrift dedicated to David (Dave) E. Tyler, Distinguished Professor of Statistics at Rutgers University. The idea for this Festschrift was born around the occasion of Dave's 70th birthday and coming retirement from Rutgers to celebrate his outstanding career with many significant contributions to the field of statistics, especially in the areas of multivariate and robust methods.

Dave has a remarkable research career, which he started in 1978, after obtaining his PhD from Princeton University, as an assistant professor at the University of Florida. Via the Old Dominion University, he came in 1983 to Rutgers University, where he currently is a distinguished professor in statistics. In 1994, Dave was elected as an IMS fellow for his distinctive contributions in statistics regarding his independent work on M-estimation of scatter. In particular, most of his work was supported by various grants from, e.g. National Science Foundation (NSF). Dave has a reliable intuition and ability to identify interesting and challenging research questions which are of general importance and relevance. Then he develops his ideas in an insightful as well as rigorous manner addressing all possible details. His attention to detail, while keeping an eye on the big picture and the relevant questions, has been passed on to early career stage researchers, with whom he collaborated and mentored. It is therefore also no surprise that all seven PhD students of Dave embarked on their careers in academia, most of whom are now associate and full professors at universities around the world.

Contributed by Dave's students, friends, coauthors and colleagues, this book includes 22 peer-reviewed papers. The topics of the contributions are mainly motivated by the research interests of Dave. Accordingly, the book consists of four parts. Part I begins with an analysis of Dave's publication and coauthor networks, followed by a review article on Dave's famous *Tyler's shape estimator*. Parts II and III, as the main body of this book, cover some recent advances in multivariate and robust methods. The final part, Part IV, includes some various other topics such as supervised learning and normal extremes.

Speaking of these cutting-edge articles, we would like to express our gratitude to the efforts and patience of all contributors in the publishing process, especially

because of the Covid-19 pandemic that disrupted most contributors' routine of work. Despite of those disruptions, upon joint work of authors and referees, we have reached a milestone with very interesting papers. We would like to thank therefore all contributors, who submitted their original and high-quality work to this Festschrift for Dave, and the referees, without whose generous help we would not have made it in time, given the tight schedule. We would like to thank also Veronika Rosteck and Daniel Ignatius from Springer who provided help and assistance whenever needed.

Finally, we want to salute Dave again for his intellectual contributions as well as his help as a mentor and as a friend. May Dave stay healthy and continue advancing the knowledge and boundaries of statistics!

Beijing, China — Mengxi Yi
Jyväskylä, Finland — Klaus Nordhausen
July 2022

Acknowledgements

The Editors would like to thank all the following referees for their excellent work:

Suchin Aeron	Andreas Alfons	Aurore Archimbaud
Andreas Artemiou	Marco Avella	James O. Berger
Germain van Bever	Ana Bianco	Eva Cantoni
Raymond Carroll	Raymond Chambers	Christophe Croux
Lutz Dümbgen	Kamila Facevicova	Hank Flury
Dominique Fourdeinier	Marc Hallin	Zifei Han
Jana Jureckova	Shogo Kato	John Kent
Arnab Kumar Laha	Gaorong Li	Chuan Liu
Markus Matilatinen	Peter McCullagh	Jaakko Nevalainen
Hannu Oja	Esa Ollila	Davy Paindaveine
Una Radojicic	Peter Rousseeuw	Gerta Rucker
Marcelo Ruiz	Anne Ruiz-Gazen	Richard L. Smith
Sara Taskinen	Beavers Traymon	Daniel Vogel
Changxi Wang	Christian Weiss	Ami Wiesel
Ines Wilms	Victor Yohai	Teng Zhang
Ting Zhang	Guang Zhu	

Contents

List of Contributors

Ana M. Bianco Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina

Gracila Boente Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina

Howard Bondell School of Mathematics and Statistics, University of Melbourne, Parkville, Australia

Javier Cabrera Department of Statistics, Rutgers University, Piscataway, NJ, USA

Gonzalo Chebi Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina

Xinyan Chen Department of Statistics, Rutgers University, Piscataway, NJ, USA

Kenneth Edward Cherasia Department of Statistics, Rutgers University, Piscataway, NJ, USA

Lutz Dümbgen Institute of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland

Luisa T. Fernholz Department of Statistics, Temple University, Philadelphia, PA, USA

Robert Fernholz Intech Corp., Princeton, NJ, USA

Daniel Fischer Applied Statistical Methods, Natural Resources Institute Finland (Luke), Jokioinen, Finland

Gabriel Frahm Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

Katrin Gysel SAKK Kompetenzzentrum, Bern, Switzerland

Marc Hallin ECARES and Département de Mathématique, Université libre de Bruxelles, Brussels, Belgium

John T. Kent Department of Statistics, University of Leeds, Leeds, UK

Hyon-Jung Kim Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland

John Kolassa Department of Statistics, Rutgers University, Piscataway, NJ, USA

Manuel Koller Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
Seminar für Statistik, ETH Zürich, Zürich, Switzerland

Ginette Lafit Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium

Thibault Laurent Toulouse School of Economics, French National Centre for Scientific Research, Toulouse, France

Peter Meer Department of Electrical Engineering, Rutgers University, Piscataway, NJ, USA

Camille Mondon Département de Mathématiques et Applications, Ecole Normale Supérieure, Paris, France

Gilles Mordant Institute for Mathematical Stochastics, Universität Göttingen, Göttingen, Germany

Javier Nogales Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain

Klaus Nordhausen Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

Hannu Oja Department of Mathematics and Statistics, University of Turku, Turku, Finland

Esa Ollila School of Electrical Engineering, Aalto University, Espoo, Finland

Davy Paindaveine ECARES and Mathematics Department, Université libre de Bruxelles, Brussels, Belgium

Daniel Pena Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain

Fabrice Perler Bundesamt für Gesundheit BAG, Leibefeld, Switzerland

Setareh Ranjbar Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland

Elvezio Ronchetti Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland

Marcelo Ruiz Department of Mathematics, Universidad Nacional de Río Cuarto, Río Cuarto, Argentina

Anne Ruiz-Gazen Toulouse School of Economics, University of Toulouse 1 Capitole, Toulouse, France

Yodit Seifu Bristol-Myers Squibb, Berkeley Heights, NJ, USA

Stefan Sperlich Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland

Werner A. Stahel Seminar für Statistik, ETH Zürich, Zürich, Switzerland

David S. Stoffer Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA

William E. Strawderman Department of Statistics, Rutgers University, Piscataway, NJ, USA

Sara Taskinen Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

Christine Thomas-Agnan Toulouse School of Economics, University of Toulouse 1 Capitole, Toulouse, France

Thomas Verdebout Mathematics Department, Université libre de Bruxelles, Brussels, Belgium

Daniel Vogel MEDICE Arzneimittel Pütter GmbH & Co. KG, Iserlohn, Germany
Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, UK

Stuart J. Watt Mirador Analytics, Melrose, UK

Anna Wiedemann Department of Psychiatry, University of Cambridge, Cambridge, UK

Han Xiao Department of Statistics, Rutgers University, Piscataway, NJ, USA

Mengxi Yi School of Statistics, Beijing Normal University, Beijing, China

Victor J. Yohai Department of Mathematics and Instituto de Calculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Weichang Yu School of Mathematics and Statistics, University of Melbourne, Parkville, Australia

Ruben Zamar Department of Statistics, University of British Columbia, Vancouver, BC, Canada

Dewei Zhong Anheuser-Busch Companies, New York, NY, USA

Yijun Zhu Department of Statistics, Rutgers University, Piscataway, NJ, USA

Stavros Zinonos Cardiovascular Institute of New Jersey, RWJMS, New Brunswick, NJ, USA

Part I
About David E. Tyler's Publications

An Analysis of David E. Tyler's Publication and Coauthor Network

Daniel Fischer, Klaus Nordhausen, and Mengxi Yi

Abstract David E. Tyler can look back on an impressive career with many significant contributions to robust and multivariate statistical methods. In this paper we attempt to quantify his scientific impact by having a closer look at his publications and by analyzing his coauthor network.

Keywords Bibliography · Community detection

1 Introduction

David (Dave) E. Tyler is a driving force in the development of robust and multivariate statistical methods with an impressive publication record. In this article, we will give a brief overview of Dave's publications and analyze his coauthorship network as well. As Dave is still active in research, we anticipate and expect to see still many more significant contributions from him. Here, however, we consider only his publications until May 2022. By then, Dave has published 82 scientific papers[1] as listed in Appendix A.1, which could be roughly classified into statistical theory, methodology, application, and comments such as reviews and discussions.

[1] For the purpose of this paper we ignored Dave's applied papers resulting from consulting but included also methodological papers which are so far only available on Arxiv.

D. Fischer
Natural Resources Institute Finland (Luke), Applied Statistical Methods, Jokioinen, Finland
e-mail: Daniel.Fischer@luke.fi

K. Nordhausen
Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland
e-mail: klaus.k.nordhausen@jyu.fi

M. Yi (✉)
School of Statistics, Beijing Normal University, Beijing, China
e-mail: mxyi@bnu.edu.cn

M. Yi, K. Nordhausen (eds.), *Robust and Multivariate Statistical Methods*,
https://doi.org/10.1007/978-3-031-22687-8_1

Below, we refer to these works using the numbers provided in the Appendix and all the citation data are based on the Web of Science on 5.5.2022 (URL: http://apps.webofknowledge.com). Citation details were downloaded from Semantic Scholar (URL: https://www.semanticscholar.org).

Dave is an expert in robust statistics, especially in M-estimation, having also significant influence in other areas such as signal processing. His most cited work [10] has been cited 380 times, which is rather high considering the general low frequency in citations in the area of statistics. In this paper, Dave proposes a new M-estimator of scatter which is nowadays quite popular and often referred to as Tyler's M-estimator or Tyler's shape matrix and it is still being actively studied in statistics and signal processing and, for example, reviewed in Taskinen et al. (2023).

Meanwhile, Dave has contributed to the field of multivariate analysis, directional data analysis, spectral analysis of time series, and functional data analysis. For instance, his most cited methodological work [58] (with 76 citations) suggested the invariant co-ordinate selection (ICS) procedure to better explore multivariate data. An R package developed accordingly introduces this method to a broader audience; see [54] and the [R1] in Appendix A.2, which lists the R packages where Dave is involved in. Methodologically applied in the area of psychometrics, computer vision, and signal processing, Dave's work has become more and more appreciated beyond the community of statistics. Dave's contribution to the academia also embodies in some review and discussion papers that gain significant attention from peer researchers, among which the most cited work [63] has been referred 261 times so far.

Dave obtained his PhD from Princeton University 1979 for his dissertation entitled "Redundancy Analysis and Associated Asymptotic Distribution Theory" supervised by Lawrence S. Mayer. This makes Dave an academic descendant of various famous statisticians who worked among others on multivariate methods and robust nonparametric methods, topics which Dave developed further in his career. A pruned version of Dave's academic genealogical tree is given in Fig. 1 which lists also the seven students who Dave supervised so far.

As Dave's academic life has been devoted to the development of statistical theory and methodology, most of his works are published in highly ranked statistical journals including *The Annals of Statistics, Biometrika, Journal of the Royal Statistical Society Series B,* and *Journal of Multivariate Analysis*. Based on the titles and abstracts of the publications considered here we provide in Fig. 2 the corresponding word cloud, see, e.g., Seifert et al. (2008), which shows the most frequent 100 words after removing the standard stop words of the English language as defined by the package `tm` and the words *abstract, keywords, can, also,* and *given*. As the font size and color reflects the frequency of a word, Fig. 2 shows clearly that Dave's research interest centers around multivariate data and scatter matrices and emphasizes the robust aspect, such as the breakdown point. Theoretically, Dave considers especially asymptotic properties and the distribution of estimators. Notice also that Dave's publication record shows surprisingly consistency on the studies of M-estimation of scatter matrices, where Dave has published numerous single authored papers and left a strong influence on statistics and signal processing till now and beyond.

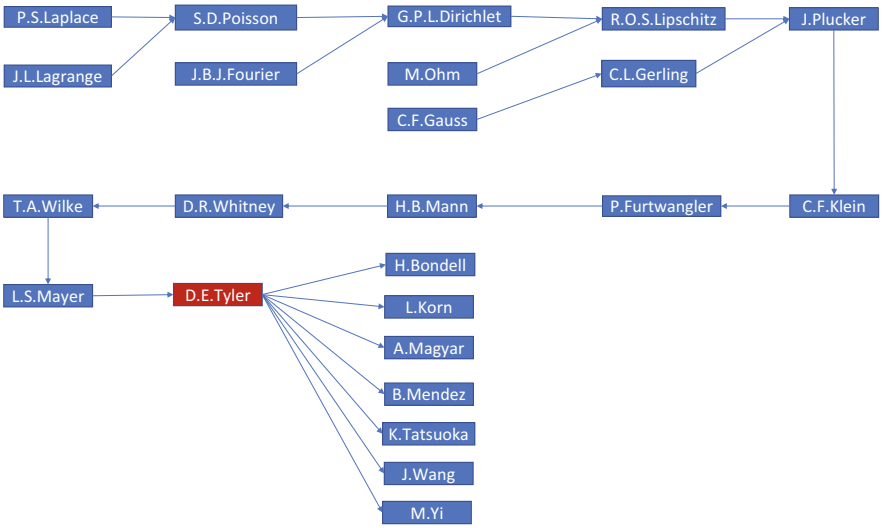

Fig. 1 Extract of Dave's genealogical tree including his seven academic children

Fig. 2 Word cloud based on the abstracts and titles of Dave's publications

In the following sections, we apply community detection methods to investigate Dave's coauthorship networks and the impact of Dave's work in the community.The analysis is done in R (R Core Team 2022), using the R packages `bibtex` (Francois 2014), `igraph` (Csardi & Nepusz 2006; Kolaczyk & Csardi 2014), `circlize` (Gu 2014), `wordcloud` (Fellows 2018), `tm` (Feinerer & Hornik 2020), and `rworldmap` (South 2011).

2 David Tyler's Coauthor Network

We first review some basic facts of network theory. A network graph $G = (V, E)$ consists of a set V of vertices or nodes and a set E of edges or links. The number of vertices $n = |V|$ is the *order* of the network G and the number of edges $m = |E|$ is its *size*. Two vertices are *neighbors* or *adjacent* if they are connected by an edge. Networks are *undirected* if there is no ordering in the vertices defining an edge and are *weighted* if a real number is associated with each of the edges. If vertices are not allowed to be connected to themselves, the graph G is called a *simple* graph. A network can be partitioned into several subgraphs, where $G_r = (V_r, E_r)$ is called a *subgraph* of $G = (V, E)$ if $V_r \subset V$ and $E_r \subset E$.

One of the most important subgraphs is the *egocentric* network. This is a network created by selecting an ego-node and all of its connections. First, we are interested to build Dave's direct-coauthor network $G = (V, E)$ based on his publications [1]–[82]. Thus, Dave is the ego vertex, his direct coauthors are the other vertices, and two distinct authors are connected by an edge if they have written one joint paper with Dave. Table 1 presents Dave's collaboration frequencies, which shows that Dave has one collaborator with whom he has written 8 papers. Dave has 21 single authored papers that do not contribute to the network. In the remaining 61 publications, Dave has 51 coauthors.

The network of Dave's direct coauthors is visualized in Fig. 3. Here, no information about how his coauthors work together with him was used, it rather visualizes Dave's blossoming levels of collaboration, as each joint publication between coauthors is visualized with an own edge, the closer therefore an coauthor is in Fig. 3 to Dave, the more often he/she collaborated with him. This indicates that his inner circle of coauthors with more than 5 joint papers consists of L. Dümbgen, P. Meer, K.Nordhausen, H. Oja, and E. Ollila; the first four coauthors have published 6 papers with Dave and the last author has published 8 papers.

Table 1 Number of times Dave collaborated with coauthors for his publications. The value zero corresponds to single author papers

Number of joint papers	0	1	2	3	5	6	8
Frequency	21	29	11	5	1	4	1

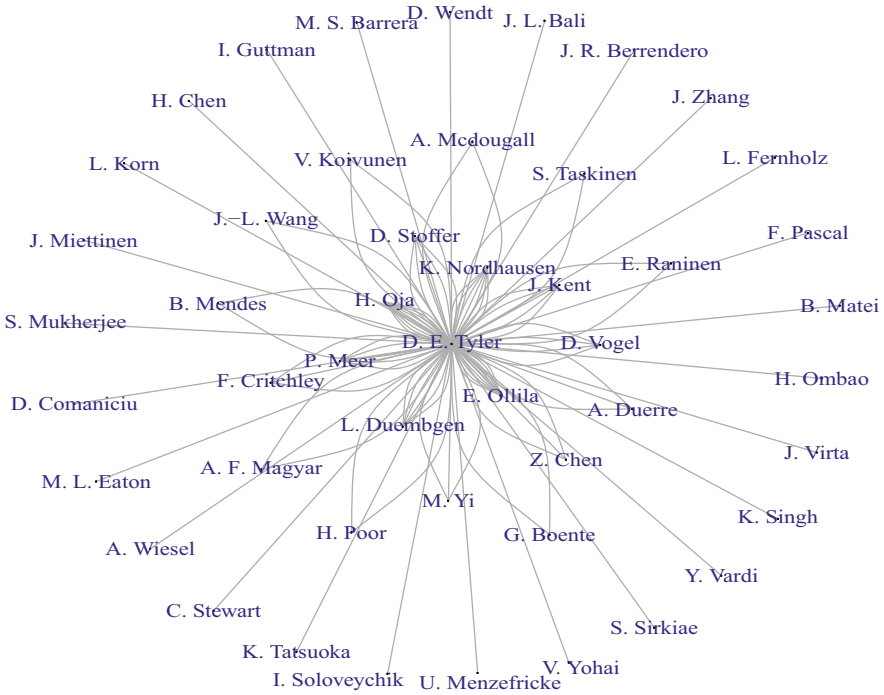

Fig. 3 Network of Dave's direct coauthors. Each connection corresponds to a collaboration for a joint paper

Next, we review some results on community detection methods. As the whole information about a network can be stored in matrix form, we could define the $n \times n$ *adjacency matrix* $\boldsymbol{A}$, for a network $G = (V, E)$, as follows

$$A_{ij} = \begin{cases} 1, & \text{if } \{i, j\} \in E, \\ 0, & \text{otherwise.} \end{cases}$$

For a simple network, the diagonal elements of the adjacency matrix are all zero. And the matrix $\boldsymbol{A}$ will be symmetric for undirected networks. If G is a weighted network, then A_{ij} represents the weight of the edge between i and j. Note also that the *degree* k_i of a vertex i, i.e., the number of its neighbors, can be given by $k_i = \sum_{j=1}^{n} A_{ij}$.

In order to detect significant community structure, or to identify good partitions of a network, it is useful to have a quality function to assess the goodness of a graph partition. In this way, the largest number given by the quality function means the partition is best. One of the most popular quality function is the *modularity* used in Newman (2006). It is based on the idea of finding divisions of the network in which the actual minus the expected number of edges over all pairs of vertices that belong

to the same cluster is highest. Let c_i be the cluster or community to which vertex i belongs, the modularity is defined as

$$Q = \frac{1}{2m} \sum_{i}^{n} \sum_{j}^{n} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

where m is the total number of edges of the network and $\delta(c_i, c_j) = 1$ if $c_i = c_j$ and 0 otherwise. The goal is then to decide the optimal number of partitions and to label the vertices by maximizing the modularity. Many methods have been suggested, in the literature, for this optimization problem; see, for example, in Fortunato (2010) for an overview. In the following we will describe the *multi-level modularity optimization algorithm* of Blondel et. al. (2008).

The algorithm consists of two phases. Consider initially to assign a different community to each node of the network. Then put each node to the community for which it gain maximum of the modularity. Repeat this process until no further improvement can be made. The second phase starts by regarding each community, found in the first phase, as a vertex and builds the network based on these nodes and links. The process stops until there are no more changes or a maximum of the modularity is attained.

By using the above-described community detection method, we build Dave's community graph, Fig. 4, based on Dave's direct coauthors. Here, in addition to the Dave's publication information, we also included relevant all coauthor publications that contribute an edge to the network to get the correct weights for the edges between the different coauthors. That means, the network visualizes the connections between the coauthors based on all joint papers and not only based on joint papers with Dave. In total we can identify eight different communities within this network; see Table 2. For instance, Community 7 corresponds to Dave's work on robust functional methods, Community 6 to Hannu Oja's group, Community 4 to computer vision groups, and Community 5 on his work in signal processing.

After considering the network of Dave's direct coauthors, we extend our search to include as well the coauthors from the coauthors. Also in this network, we take the direct connections between coauthors into consideration, so that we had to look at Dave's peers that are even three nodes away. For that extensive search Semantic Scholar granted us a API access and the search resulted in 495,864 peers with 253.5 million connections in total. For this network, we filter then to peers that are two levels away only; see Fig. 5. Here, we visualize $n = 2755$ nodes and $m = 364{,}469$ edges. And Table 3 lists the 5 most influential authors in each community. It becomes obvious that Dave is not part of a tiny community but has via his connections a huge reach into the scientific community. This is indicated that many of the communities detected are not directly linked. Crude interpretations for some of communities are possible. Community 3 could be dependent data like time series, Community 5 the British school of statistics and Community 15 multivariate nonparametric statistics while Community 12 could be summarized as signal processing.

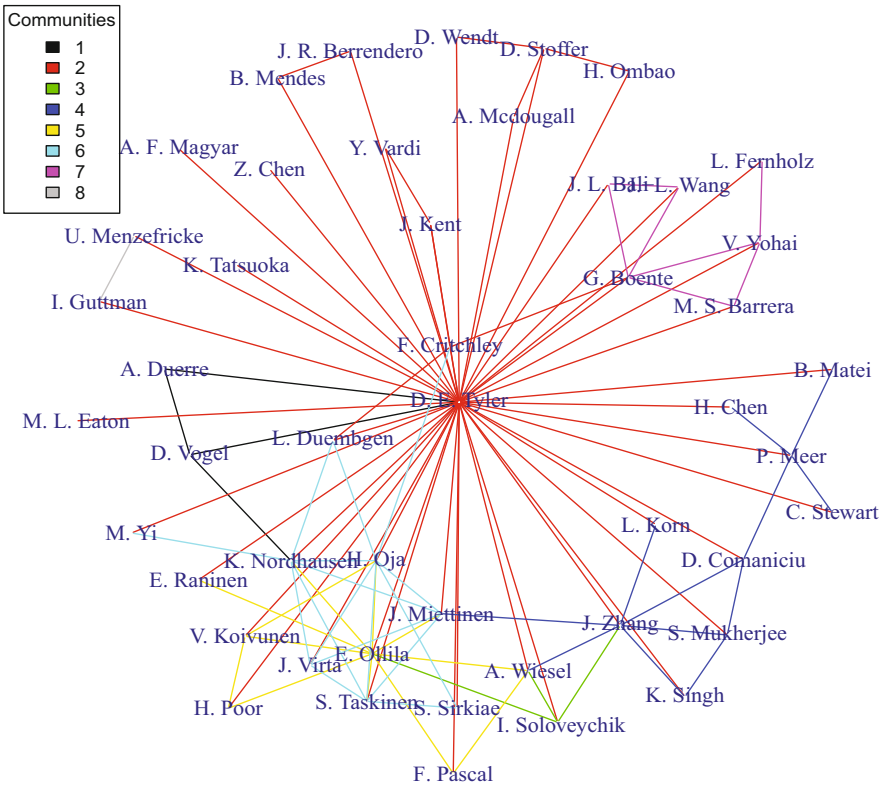

Fig. 4 Community Network of Dave's direct coauthors, where members belonging to the same community are connected by edges in the same color

Table 2 Communities detected by using Dave's direct coauthors. The table below lists the most prominent members of the communities

Community	1	2	3	4	5
1	D. Vogel	A. Duerre			
2	D. E. Tyler	D. Stoffer	J. Kent	F. Critchley	A. Mcdougall
3	I. Soloveychik	A. Wiesel			
4	J. Zhang	P. Meer	D. Comaniciu	S. Mukherjee	K. Singh
5	E. Ollila	V. Koivunen	H. Poor	F. Pascal	E. Raninen
6	H. Oja	S. Taskinen	K. Nordhausen	J. Miettinen	J. Virta
7	G. Boente	V. Yohai	M. S. Barrera	J. L. Bali	J.-L. Wang
8	I. Guttman	U. Menzefricke			

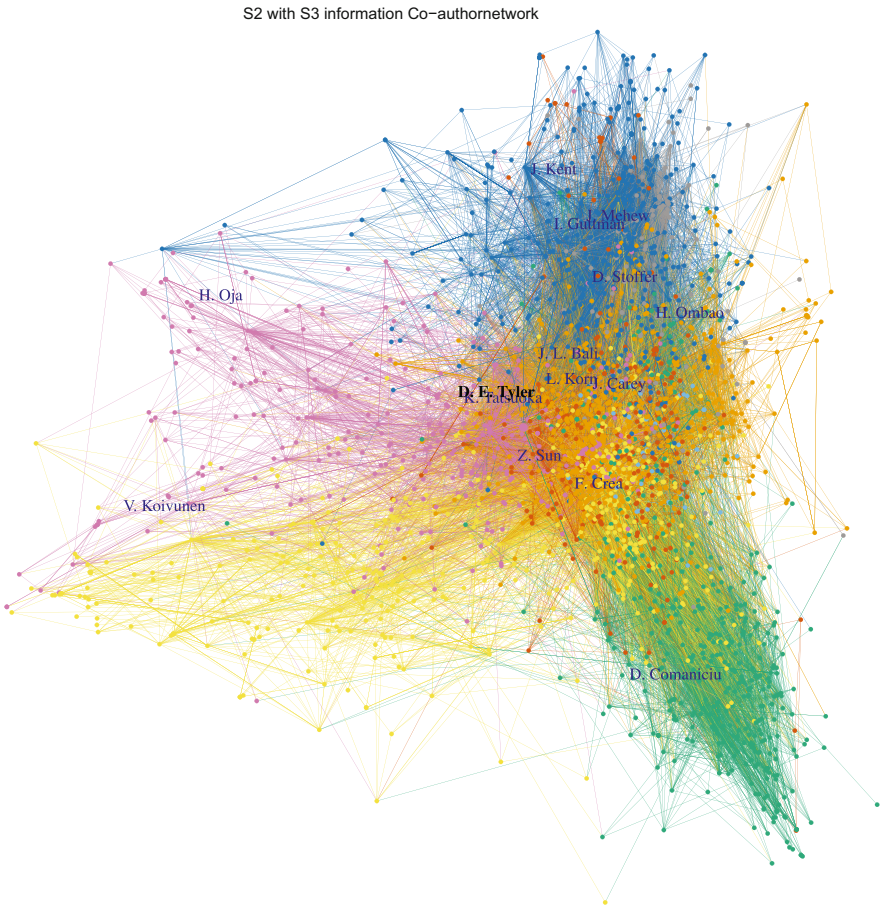

Fig. 5 Community Network of Dave's coauthors' coauthors. Top representatives from each community are highlighted with names

3 David Tyler's Influence

When considering the network spanned by Dave's coauthors' coauthors one can suspect that Dave's ideas have a wide reach. While it is quite a challenge to measure a researcher's impact we make an attempt by looking at the citations Dave's work got in Semantic Scholar. Based on this data, Dave's papers received in total 3739 citations from 2993 different authors. Hereby, his most citing peers are F. Pascal (184), K. Nordhausen (159), H. Oja (158), A. Breloy (99), D. Paindaveine (99), and E. Ollila (96). Further, his citations originate from 433 different journals, with *IEEE Transactions on Signal Processing* (179), *Journal of Multivariate Analysis* (146), *Signal Processing* (57), *Computational Statistics & Data Analysis* (55), and *IEEE Signal Processing Letters* (48) being the journals that contain the most citations

Table 3 Communities with more than 4 members detected by using Dave's coauthors–coauthors and giving their most prominent members

Community	Size	1	2	3	4	5
1	273	Z. Sun	Y. Chen	X. Luo	R. Sharma	Y. Pan
3	70	D. Stoffer	R. Fried	M. Thiel	R. Kliegl	M. Ding
4	223	F. Crea	K. Fox	H. Katus	S. Achenbach	S. Blankenberg
5	453	J. Kent	K. Mardia	F. Critchley	P. Diggle	C. Williams
6	86	L. Korn	S. Alimokhtari	S. Maberti	C. Weisel	A. Winer
7	70	J. L. Bali	L. Otero	E. Quel	P. Ristori	J. Salvador
8	47	I. Guttman	O. P. Aggarwal	D. Newman	H. Shapiro	M. Klamkin
9	393	H. Ombao	J. Williams	M. John	C. Brown	H. Zhao
10	35	J. Carey	P. Liedo	N. Papadopoulos	F. Molleman	B. Katsoyannos
11	343	D. Comaniciu	B. Georgescu	A. Kamen	P. Meer	J. Hornegger
12	241	V. Koivunen	F. Pascal	J. Ovarlez	A. Wiesel	H. Poor
14	163	K. Tatsuoka	C. Stewart	S. Adams	H. Martus	A. Olaharski
15	278	H. Oja	S. Taskinen	J. Miettinen	K. Nordhausen	L. Capranica
16	71	J. Mehew	S. Stehlíková	F. Gasperoni	F. J. G. Perez	V. Yohai

Table 4 Number of citations to Dave's work from different fields of sciences

Field	Citations	Field	Citation
Mathematics	2507	Psychology	9
Computer Science	1550	History	7
Medicine	145	Geology	6
Engineering	122	Art	4
Physics	53	Chemistry	3
Economics	45	Business	2
Biology	40	Political Science	2
Environmental Science	18	Sociology	2
Geography	16		
Materials Science	12		
Philosophy	12		

to Dave's work, which shows that his ideas are especially in signal processing of interest. Classifying these journals[2] into scientific fields, according to the system of Semantic Scholar, shows that these journals represent 19 main fields of study, ranging from Mathematics and Computer Science of Engineering, Biology and Physics towards Philosophy, History and Art; see Table 4 the frequencies. From Table 4, we find that Dave's most influential area is Mathematics, specifically in Statistics. Interesting to note is that his methods could also be applied in Art.

However, considering the huge amount of data that was collected to create these Figures, we relied heavily on an automatic data collection. Here it is possible that we missed for some authors a few publications, in case there is no consistent and traceable affiliation history available. Also, it might also happen that some wrong publications were assigned to authors based on a name mix-up.

A more detailed view of how Dave impacts the work of others could be revealed when we look at the keywords of the citing articles. For the most frequent ones ($\geq$20 occurrences), we create again another word cloud; see Fig. 6. Here, it is easily noticeable from the word cloud that statistical terms stand out.

While above we considered the total number of citations it is of course of interest to see how these papers distribute over Dave's 82 papers which are here under consideration. We visualize the corresponding information in a circos plot, a visualization type that is typically used in comparative genomics to show links between different chromosomes, see Krzywinski et al. (2009). In the circos plot given in Fig. 7 we therefore order the papers chronological and give a line from each paper to the year in which it was cited. As it is usual in mathematical sciences it usually takes some time before a paper gets cited. The blue lines in the figure correspond to citations of paper [10] where Dave introduced his famous shape matrix. The paper seems to have gotten increased interest starting from 2004 and interestingly then its popularity increased in three year cycles. Inquired which

[2] Note that one journal can be assigned by the system to several fields of science.

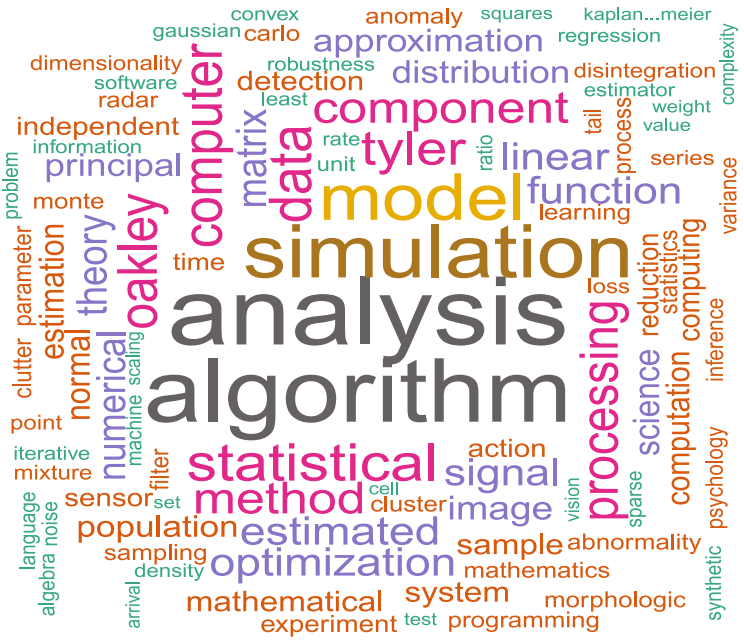

Fig. 6 Word cloud based on the keywords of articles citing Dave's papers

papers besides [10] have most important contributions, Dave listed his 10 papers [1, 5, 14, 18, 19, 24, 36, 58, 60, 63] which are marked in red in the figure. It is quite clear from the figure that all these papers have a steady impact over time but are less influential than the shape paper. Another feature of the figure is Dave's collaboration pattern, the comparison between collaborative papers and single authored ones. Interesting to see there is that while in the first 10 years of his career Dave mainly worked alone, after that almost all publications are made in collaborations. While this might be partly due to a shift in scientific conventions we believe it also reflects Dave's increased popularity and that Dave is an excellent collaborator willing to share with others his deep insight and many ideas as when looking of the list of collaborators one can see that many of them especially in the last 10 years are in much earlier career stages.

Also with the increase in collaborations Dave's output seems to have increased considerable since the mid-nineties. Notable in this context is also that when considering what we consider the current location of his coauthors that they are spread quite around the globe as visualized in Fig. 8.

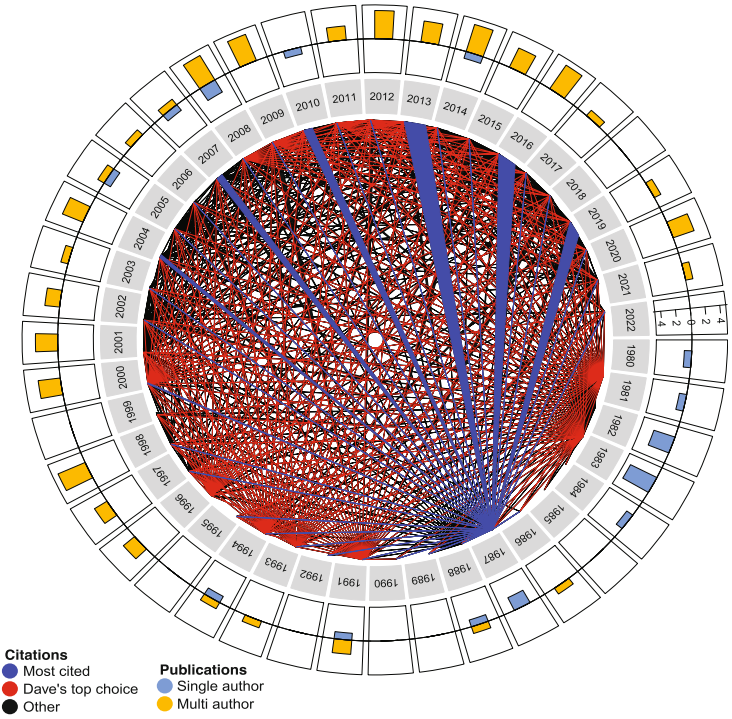

Fig. 7 Circos plot showing the influence of David's paper over time. The outer ring indicates the publications per year (blue are single author publications, yellow are multi-author papers). The red lines indicate citations of David's top 10 important papers and the blue lines are citations to Dave's shape matrix paper [10]

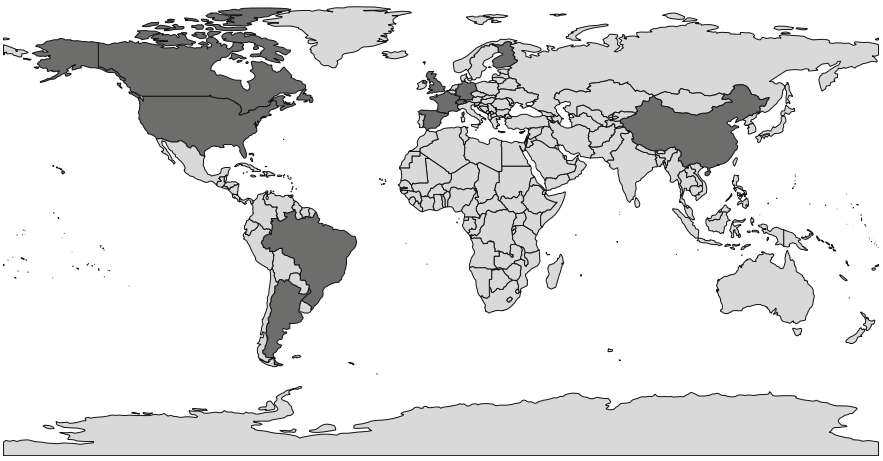

Fig. 8 Current location of Dave's coauthors, where the corresponding countries are in black