



Classification Society

— Formerly the Classification Society of North America (CSNA) —

Fionn Murtagh, President
William D. Shannon, President-Elect
Hans-Friedrich Köhn, Newsletter Editor

Classification Society Newsletter (80), October 2008

Contents:

- (1) “The Future of Communicating Research” by Fionn Murtagh
- (2) Classification Society Annual Meeting 2009
- (3) Chikio Hayashi Awards Program
- (4) The CS Distinguished Dissertation Award 2009
- (5) Call for Papers: *Journal of Library Metadata*
- (6) Important Conference Dates

Editorial

We all would like to express our great appreciation to Mike Larsen for his many years of excellent service as the Classification Society Newsletter editor. The October 2008 edition features a new editor, Hans-Friedrich Köhn. He is Assistant Professor of Quantitative Psychology in the Department of Psychological Sciences at the University of Missouri-Columbia. After receiving his Diplom (= Masters) in Psychology from the University of Hamburg, Germany, he spent ten years in industry as a marketing researcher, then went back to grad school. He earned a Masters Degree in Statistics and a Doctoral degree in Quantitative Psychology from the University of Illinois, Champaign, Illinois (2007, advisor: Larry Hubert). His research concerns applications of combinatorial optimization methods to scaling/unfolding, clustering/tree-fitting, and order-constrained matrix decomposition problems, with special focus on the analysis of individual differences based on sets of multiple proximity matrices, as might be collected from different data sources in the context of cross-sectional or longitudinal studies. He also has worked on algorithms for the p -median clustering for partitioning large data sets and the clique partitioning problem.

Thanks again, Mike! Welcome aboard, Frieder!

Stan Sclove

Contact Information:

Hans-Friedrich Köhn
Department of Psychological Sciences
University of Missouri-Columbia
19 McAlester Hall
Columbia, MO 65211-2500
Work Phone: 573-884-7866
koehnh at missouri dot edu



1. “The Future of Communicating Research” by Fionn Murtagh

What does the future hold for research dissemination, new fora for scientific interaction, and publishing? What are the implications for the Classification Society? This is a general discussion piece, with an invitation to respond. More so, there are various proposals here that could well result in action. At the Classification Society Board meeting on June 8, 2008 it was agreed that we need to improve the information on clustering available on Wikipedia. Google “cluster analysis” and the top two returns are to Wikipedia, with information which could be enhanced and improved.

While the issue of getting new content onto Wikipedia was a starting point, I think it would be good to look at the broad context, including the following themes:

1. Preprint servers
2. Blogs
3. Citation ratings and implications for how we publish
4. Virtual communities
5. Wikipedia

(1) *Preprint Servers*

Preprints help to have one’s work noticed and cited. I am not aware of any issue on the part of publishers with preprints. The situation is different with regard to postprints, following publication. When initially on faculty way back in the early 1980s, I remember the importance of checking out the latest preprints from other institutions that were available in printed format, always with an institutional cover page, and with institute-based identifier information.

Should the Classification Society have a strong recommendation to avail of preprint servers, for example, [arXiv \(arXiv.org\)](http://arXiv.org)? Categories and hence sections in [arXiv](http://arXiv.org) for our sort of all-embracing work are tricky. However categories are not hugely important when free text search is available.

Current [arXiv](http://arXiv.org) sections are:

- (1) Physics (includes: Data Analysis, Statistics and Probability)
- (2) Mathematics
- (3) Nonlinear Sciences
- (4) Computer Science
- (5) Quantitative Biology
- (6) Statistics

The statistics archive (with sections Applications, Computation, Machine Learning, Methodology, Theory/Mathematical Statistics) is overseen by an advisory committee consisting of the following members: Larry Wasserman (coordinator), David Banks, Chad Schafer, Dave Higdon, Rob Strawderman, Susan Holmes and Jim Pitman; section Machine Learning features Classification, Graphical Models, High Dimensional Inference, and is moderated by [David Banks](http://arXiv.org).

I am not entirely at ease with these categories covering all of our work, even if we restrict attention to theory as opposed to applications. Nor am I entirely at ease with Machine Learning (i.e., supervised) covering unsupervised data analysis work.

(2) Blogs

Would it be of interest to propose a collective blog? Would it work? Consider for example,

<http://dialinf.wordpress.com>,

relating to: “A Dialogue on Infinity between a mathematician and a philosopher. In our dialogue, the mathematician is Alexandre Borovik and the philosopher is David Corfield. We expect a number of guests contribute to our blog; ...”. Included on this site is an extensive bibliography. Note that our Classification Society Service, distributed on CD, each year contains the list of “profile items” which constitute a very good introductory bibliography. Furthermore the scanned books on the CD constitute a valuable resource.

(3) Citation Ratings and Implications For How We Publish

The h-index (Hirsch index) of a researcher’s abilities as measured by citing of that person’s output has become pretty much all-pervasive. The trend towards citation-based assessment of scholarly work is led (in my opinion) by the life sciences. After all, NIH has led the charge in regard to open access through PubMed Central depositing requirements and one reason for this is the greatly evolved and structured nature of publishing in the life sciences. Recently (as I recall, in May 2008), a life sciences colleague asked me if I was also waiting with bated breath for the latest ISI Journal Citation rankings data. In his case, a movement up or down for the journal he is associated with would be hugely important. He described it as like a student waiting for an exam result. I found this an unusual view to hold.

Books are more cited than other works but a book citation rating, unlike journal citation ratings, is not something that is typically available. Furthermore the Open Access movement has a focus on journal publication and nothing else. Unlike the case of journal publishing, book postprint depositing has not been raised, nor has depositing of other research outputs — data, software, even conference proceedings. Or conference programs, or video or audio recordings of talks. Or the Figure (or plot), if one takes that as the paradigmatic output of science. (The Table is somewhat accounted for, especially in astronomy, through sharing of catalogs, i.e., tables of data. In bioinformatics there is also major interest in federated data stores.)

Could we steal a march on others by adopting a particular interest in research monograph publishing, given the headlong rush towards quantitative “metrics” of research that are, by and large, journal oriented only?

(4) Virtual Communities

Are all scholarly societies having a declining membership curve, and is the action to be found elsewhere, such as in society networking communities? A colleague (in astronomy, as it happens) was recently trying to persuade me (very cogently) to get involved in Second Life. But is Second Life even somewhat passé by now? What about LinkedIn and its groups? How about having short video snippets of our annual conference on YouTube?

(5) *Wikipedia*

At long last, I am coming to Wikipedia. What about collectively adding content to Wikipedia? For example, background and discussion of clustering algorithms, multivariate data analysis theory and practice, perhaps personalities, in particular early leaders in the field? Would some of our younger members take the lead on this? Individual members can contribute on their own, but do we additionally need encouragement to get a collective (federated, let's say) effort underway?

2. Classification Society Annual Meeting, 2009

The Classification Society Annual Meeting will be held at St. Louis, MO, June 11–13, 2009 (to be confirmed); organization: Bill Shannon, Washington University School of Medicine, St. Louis, MO (Tel. 314-454-8356; WSHANNON at dom dot wustl dot edu). Please submit proposals for invited sessions and contributions to Bill!

3. Chikio Hayashi Awards Program

The Chikio Hayashi Awards Program supports promising researchers who will present a paper on classification, data analysis and related areas at the IFCS 2009 conference, and who are in the early stages of their professional careers. Deadline: November 30, 2008. Information: <http://www.classification-society.org/cha>.

4. Classification Society Distinguished Dissertation Award 2009

The Classification Society Distinguished Dissertation Award is supported by Chapman and Hall/CRC. The Award will be for the best PhD (or approximately equivalent doctoral) dissertation nominated by an annual deadline. The theme is clustering, classification, related areas of data analysis, encompassing both associated theory and/or applications. The Award is administered by the Classification Society. An evaluation committee is set up by the Classification Society to evaluate the nominated theses. This committee should have between 3 and 5 members. The chair of the committee is appointed by the President of the Classification Society. The committee is selected by the committee chair, in conjunction with the President of the Classification Society, and is approved by the Classification Society Board.

In 2008-2009, the evaluation committee is: Samantha Prins (chair, Department of Mathematics & Statistics, James Madison University), Fionn Murtagh (Science Foundation Ireland; and Department of Computer Science, Royal Holloway, University of London), Douglas Steinley (Department of Psychological Sciences, University of Missouri-Columbia) and

Rebecca Nugent (Department of Statistics, Carnegie Mellon University). Members of the evaluation committee must declare any potential conflict of interest. A member of the evaluation committee cannot play any role in regard to evaluating his/her own PhDs, nor also in regard to nominations where there has been a direct collaborative link (for example, through joint publications). Nominations will be through the evaluation committee chair. It may be useful to have printed copies sent to committee members, but in any case web-accessible versions have to be made available. The evaluation panel may seek one or two external referee reports for each nominated dissertation. Typical criteria for nomination for the Award include: (1) the most innovative or impressive work in theory/methodology, or the most innovative or well developed application(s); and (2) the literature review has to be thorough.

Nominations are to be received by January 1 each year, from the author (PhD), their advisor/supervisor, or other related person. Nomination includes the name and contact points, an online copy of the dissertation, and a short description of why the dissertation merits the award (see above). A decision is made by the evaluation committee by April 1. The evaluation committee can recommend one award; an award and a runner-up; two joint winners; or other combinations if they think it appropriate. For the next submission deadline, January 1 2009, nominated PhDs have to have been successfully completed in the 2008 calendar year. Awards are ratified by the Classification Society Board. Publicity is made in the Classification Society and other newsletters, and on relevant websites. The short-list of nominated dissertations will also be openly published. For non-English language theses, it is requested to provide an extended abstract, and (where relevant) pointers to papers associated with the thesis that have been published. In 2009 the Award totals US\$500 in book vouchers from Chapman and Hall/CRC. The winner will be invited to make a presentation in a special plenary session at the Classification Society Annual Meeting. Up to \$500 of the Award winner's travel expenses to the Annual Meeting will be covered by the Classification Society.

Contact for information and nominations: prinssc at jmu dot edu.

5. Call for Papers: *Journal of Library Metadata*

The charter issue of the *Journal of Library Metadata* is now in circulation (quarterly; previously published as the *Journal of Internet Cataloging*, through *Vol 7, No. 4*). Under the editorship of Jeffrey Beall, Metadata Librarian and Assistant Professor, Auraria Library, University of Colorado (Denver), the new peer review journal from Haworth Press (now part of Taylor & Francis) covers research, theory and practice related to metadata in libraries of all types. The charter issue includes articles on "Dublin Core Metadata Harvested Through OAI-PMH", "Cataloging Images in Millennium: A Central Repository", "You Need My Metadata: Demonstrating the Value of Library Cataloging", "From Hanging Files to Digital Collection: Growing a Controlled Vocabulary for Added Functionality in the Online World". Editor Beall welcomes contact from potential authors, Jeffrey.Beall at cudenvr dot edu; free samples are available from marisa.starr at taylorandfrancis dot com.

6. Important Conference Dates

(1) IASC 2008 — Joint Meeting of the 4th World Conference of the IASC and the 6th Conference of the Asian Regional Section of the IASC on Computational Statistics & Data Analysis, December 5–8, 2008, Yokohama, Japan;
<http://jasp.ism.ac.jp/~iasc2008/>

(2) ICPR 2008 — 19th International Conference on Pattern Recognition, December 8–11, 2008, Tampa, Florida;
<http://www.icpr2008.org>

(3) NIPS 2008 — Twenty-Second Annual Conference on Neural Information Processing Systems, December 8–11, 2008, Vancouver, B.C., Canada;
<http://nips.cc/Conferences/2008/Program/>

(4) IICAI 2009 — Fourth Indian International Conference on Artificial Intelligence, December 16–18, 2009, Tumkur, India;
www.iiconference.org

(5) IFCS 2009 — International Federation of Classification Societies 2009 Conference, March 13–18, 2009, Dresden University of Technology, Dresden, Germany;
<http://www.ifcs2009.de>

(6) Symposium on Learning and Data Science, April 1–3, 2009, University of Paris-Dauphine, Paris, France;
<http://ceremade.communication-pro.fr/>

(7) IMPS 2009 — International Meeting of the Psychometric Society, July 20–24, 2009, St. Johns College, Cambridge, UK;
<http://www.psychometrika.org/meeting/2009/leaflet.pdf>

(8) COMPSTAT 2010 — 19th International Conference on Computational Statistics, August 22–27, 2010, Paris, France;
<http://www.compstat2010.fr/>