Contrasting Publications in Design and Scientific Research

Gabriel Wurzer[1], Wolfgang E. Lorenz[2], Tomo Cerovsek[3], Bob Martens[4]
[1,2,4] TU Wien [3] University of Ljubljana
[1,2,4] {gabriel.wurzer|wolfgang.lorenz|bob.martens}@tuwien.ac.at [3] tomo.cerovsek@gmail.com

This paper explores the differences between `design' and `science' papers published at eCAADe conferences through use of automatic classification. The latter is conducted using a set of differentiating criteria (e.g. number of figures determines a paper to be either `design' or `science') which are calibrated with the help of a manual selection of papers from eCAADe 2015 as ground truth. Results show that we predict 83% of the papers correctly; experiments using data from eCAADe 2014 until eCAADe 2016 furthermore show the stability of our results. However, we are not so much after the development this automatic classification but rather want to characterize the two research cultures of design and science. This is achieved by taking a close look at the differentiating criteria, which can inform tools such as ProceeDings over possible future directions and adaptation needs.

Keywords: *Differentiation, Design, Science, ProceeDings, CumInCAD*

INTRODUCTION

Architectural conferences are inevitably caught between two very different research cultures: They must bring both "science" and "design" together, a process which is not seldomly done by having two proceedings volumes (one for the former, one for the latter). However, both ways differ with regards to the content (e.g. more or less use of figures or references). Another important difference is the way of referencing and giving credit (think: impact). To carefully characterize and contrast these two different cultures, with the expressed intent of being able to further develop and adapt the content guidelines of conferences in our field, we have conducted a study in which content content analysis was applied to papers:

- We exported all papers from ProceeDings (Wurzer, Martens and Grasl 2014), the web-based paper editing system eCAADe uses since 2014, and built up a publication graph (see section 'Base Data'). Since ProceeDings stores each paper in a structured manner (e.g. figure, paragraph, enumeration, bullet point, reference to an article, etc.), there was a treasure-trove of semantic input which could be used to build up criteria that distinguish science from design. Additional metadata was retrieved from CuminCAD (Cumulative Index on CAD; Cerovsek and Martens 2016) so as to provide not only data about each individual paper but also about papers being referenced.

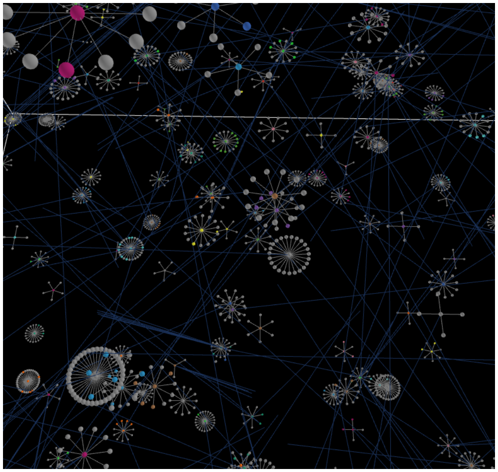

Figure 1
Graph representation of papers (NetLogo 3D showing papers and references as nodes, citations as edges).

- In order to automatically tag each paper as 'science' or 'design', all papers of eCAADe 2015 were first classified manually (see section 'Calibration', subsection Ground Truth). Most of these papers do not fit directly into these categories but are mixtures (n=148; science=24; design=22; other=102; see Figure 2). However, since the manual selection was carried out in a double-blind process, it is clear that only the most characteristic and undisputed examples were selected as ground truth.

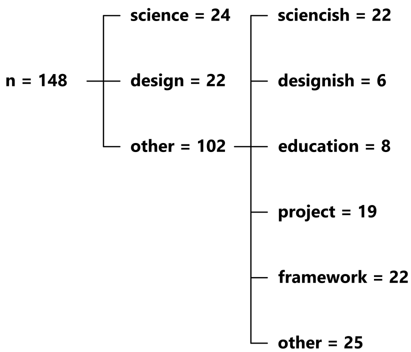

Figure 2
Decomposition of all eCAADe 2015 papers into science, design and other.

- Everything that is measurable in a paper and the graph it is embedded into can potentially serve as basis for a classification. However, there were far too many of these variables (in exact terms: 110) to be able to analyze each of these individually, even though much work was done in a statistical pre-study. Instead, we used a brute-force parameter sweep to find all variables which can serve as differentiating criteria (see section 'Calibration', subsection 'Fitting'). Only criteria which have predictive certainty - i.e. at least 65% of all papers can be explained using that measure, are selected as basis for classification.
- Applying all selected criteria yields a list of tuples of the form *(classification, certainty)*. Through application of majority voting, we find the most probable classification for each paper (see section 'Automatic Classification'). Furthermore, an overall measure for the quality of this classification is given by the mean of all certainties of the winning class. Application of this method on a publication graph built up from all papers of eCAADe 2014 and 2015 shows that our method can classify 83% of all papers within the ground truth correctly (see section 'Results'). The rest of the papers are classified as 57% science, 43% design. Applying our measure further, we find a fairly constant rate of 55% science vs. 45% design to be present in eCAADe even if we additionally take all papers of eCAADe 2016 into our publication graph.

However, the reason behind this paper lies not so much in the development of an automatic classification for science or design but rather in the question of *what motivations we can infer from the classifying variables*. Ample space is devoted to the discussion of these factors, since they can serve as basis for future development for tools such as ProceeDings (see section 'Discussion', subsections 'What scientists want' and 'What designers want').

BASE DATA

This paper is based on structured data originating both from ProceeDings as well as CumInCAD:

- ProceeDings stores each paper as xml file. The different kinds of paragraphs (paragraph, bullet point, enumeration, figure, table, etc.) are technically represented as tags. These may be highly structured - as in the case of references or they might contain only text (see example further down in the text). It is the job of the analysis program to transform each paper into a graph representation (see Figure 1): For each paper, a node is allocated. The node is attributed with all metadata, such as: authors, the sequence of paragraph types used (e.g. title subtitle abstract keywords ...) and lengths of these paragraphs, in characters. Furthermore, all references of the paper are also allocated as nodes. For references that also exist in CumInCAD, there is additional metadata which can be attributed (id, authors, year, title, source, keywords, series etc.). However, the usual case is that we know only what was entered in the references, i.e. authors, title, source and year.

```
block with nested structure:
<inproceedings>
  <authors>
    <author>
      <initials>JRR</initials>
      <surname>Tolkien</surname>
    </author>
  </authors>
  ...
</inproceedings>

block containing only text:
<title>This is a test</title>
```

- Papers are linked to their references via edges. A reference is said to be *shared* if more than one paper references it. Since we deal with multiple conference year (eCAADe 2014-2016), it can also be that a paper references a previously-published paper. In that case, we check whether there is a common author in both paper, so as to account for self-citations (which can also be thought of as a continuation of work).
- In a further analysis step, our approach looks at each paragraph of a paper. It performs a content analysis, which considers each paragraph's length in proportion to the total paper body (or, in case of the references, the reference section). Furthermore, it counts the amount of standard headings, i.e. "introduction", "related work", "previous work", "background", "contribution", "idea", "conclusion(s)", "discussion", "motivation", "research question" and "methodology" - which are typical elements of every scientific paper and can thus distinguish a design paper from a scientific one to some extent.
- Manual classification into design or science is another piece of metadata that can be attributed to paper nodes. An in-depth description of the classification's logic is given in the next section (see 'Calibration', subsection 'Ground Truth').

CALIBRATION

Ground Truth

To classify papers into 'science' or 'design' is a hard task. We performed a double-blind selection process on all papers of eCAADe 2015 and only took papers into either category if both reviewers agreed. The logic behind the selection was as follows:

- Scientific papers must present an idea or hypothesis which is proven in the paper's main body. The presentation of the proof must be reproducible, i.e. there must be enough information regarding the process and input data (if this in the scope of a paper) such that it is clear how the authors came to a certain conclusion. Supporting material - such as algorithms, tables and illustrative figures, is indicative for such a kind of work. Furthermore, obtained results must be discussed and set into a perspective (think: relevance, quality of results and so on). Failing to produce one of these points makes a paper "sciencish" but not scientific (see Figure 2).

- Papers that are purely "design" bring forward inspirational work - typically in the form of built structures which are presented as such. They are not concerned with formal proof but present original work following an artistic logic. Most often, these papers can be spotted already by regarding the images - which will typically be photos of pavillons, installations, exhibitions and so forth. Furthermore, papers of this category often cite related artists in their figure captions (e.g. "work by ....", "photo by ....") but not in the references. Any attempt to put such a paper on a more formal basis - such as description of technology used, structural analysis of the end product(s) or so forth, makes a paper "designish" but not "design" in our classification (see Figure 2).

Further categories included project descriptions ("we did this, afterwards we did that"), education ("the students produced....; the workshop was about ....") and frameworks ("using our program, it is now possible to ...."). Even then, we were unable to capture all the knowledge that is present in the eCAADe conference, and have thus included an "other/other" category (see Figure 2).

Our classification stated that 24 papers were clearly "scientific" and 22 papers clearly "design" in the above sense. All further categorization was conducted for the sake of completeness, but did not enter our analysis. Papers other than design or science were marked as unclassified, so as to exclude them from the ground truth.

Fitting

How does one select measures that have a high predictive quality with regards to whether a paper is "science" or "design"? The naive approach would be to come up with some intuitive rules, such as: Papers in design have a high number of graphics while scientific papers are generally text-based. However, such an approach turned out to be not adequate for classifying all papers, and further analysis showed that the variables we have are largely insignificant if we talk about statistics: 46 papers (ground truth) are not enough to determine a classification based e.g. on median, lower quartile and upper quartile. The second difficulty lay in the mere number of variables we can measure (110), which proved to be difficult to handle manually. Thus, we used an automatic variable selection and fitting process which gave us good results despite our limited set of ground truth samples:

- We assume that there is a differentiating threshold between science and design for every variable. This threshold is two-fold (refer to Figure 3): It is either so that all scientific papers have less than a certain value and all design papers have more or are equal, or vice versa.

Figure 3
Finding a threshold such that the amount of matched papers is maximal.

- To find the best threshold, we have to search through all possible values a variable might take. If its domain is [0, 1], our algorithm starts at 0 and ends at 1, incrementing by 0.01 (1 percentent) in every step . In all other cases, we have to search in between *(min - 1)* and *(max + 1)* since the variable is in the domain [0 ... N]. The best threshold is the one that can categorize the highest amount of papers found in the ground truth correctly. A formal description of this algorithm is given in due course as pseudocode.

```
======================================
PSEUDOCODE
======================================
fit all variables() {
  for each variable {
    if domain is [0..1] {
      fit(variable,0,1,0.01)
    } else { // domain is [0..N]
      fit(variable,
          min value of variable - 1,
          max value of variable + 1,
          1)
    }
  }
}

fit(variable, start, end, step) {
  n = number of papers classified '
   ↪ design' or 'science'
  best threshold := start
  best correctness := 0
  best estimate := science left,
   ↪ design right
  best certainty := 0
  for threshold := start to end
   ↪ incremented by step {

    // TRY SCIENCE LEFT, DESIGN RIGHT:
    correct = 0
    for each paper {
      seems to be science := variable
       ↪ < threshold
      seems to be design  := variable
       ↪ >= threshold
      if seems_to_be_science is not
       ↪ seems_to_be_design {
        // (CANNOT BE BOTH SCIENCE AND
         ↪  DESIGN)
        if (seems_to_be_science and
         ↪ ground truth says science)
         ↪  or
           (seems_to_be_design and
            ↪ ground truth says
            ↪ design) {
           // (PREDICTION AGREES WITH
            ↪ GROUND TRUTH)
           correct := correct + 1
        }
      }
    }
    if currect > best-correctness {
      best threshold := threshold
      best correctness := correct
      best estimate := science left,
       ↪ design right
      best certainty := best
       ↪ correctness / n
    }

    // TRY DESIGN LEFT, SCIENCE RIGHT:
    correct = 0
    for each paper {
      seems to be design := variable <
       ↪  threshold
      seems to be science  := variable
       ↪  >= threshold
      if seems to be science is not
       ↪ seems to be design {
        // (CANNOT BE BOTH SCIENCE AND
         ↪  DESIGN)
        if (seems to be science and
         ↪ ground truth says science)
         ↪  or
           (seems to be design and
            ↪ ground truth says
            ↪ design) {
          // (PREDICTION AGREES WITH
           ↪ GROUND TRUTH)
          correct := correct + 1
        }
      }
    }
    if currect > best-correctness {
      best threshold := threshold
      best correctness := correct
      best estimate := design left,
       ↪ science right
      best certainty := best
       ↪ correctness / n
    }
  }

  // (result is in best threshold,
  // best estimate, best certainty)
}
```

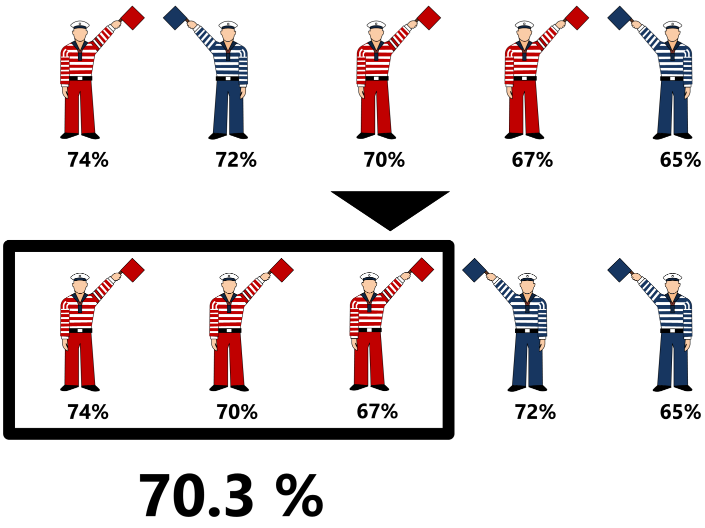

Figure 4
Papers are evaluated by a set of criteria, each yielding a category ('design' or 'science') and a certainty in percent. The overall category is the one with most votes, the overall certainty is given by the mean of certainties in that category.

- Depending on the amount of papers where prediction and ground truth match, we get a "certainty". Technically, this is the number of matching papers divided by all papers (see previous algorithm). By ranking these certainties as in Table 1, the final set of measures is determined. A measure that is 50% certain would make no sense, since it would equally predict and mis-predict. Thus, we have taken only measures above 65% certainty into account.

AUTOMATIC CLASSIFICATION

Evaluating all criteria given in Table 1 yields tuples of the form *(classification, certainty)* for every one of them (see Figure 4 for a visualization). Through majority voting, we can then obtain the overall classification. Furthermore, all measures of that class have a certainty, the mean of which forms the overall certainty for the classification.

This certainty is less informative than one would assume: It is typically the mean of the certainties of all criteria - which in our calibration amounts to 68%. Deviations from this mean indicate a better or lesser degree of prediction, which could be used to further indicate the quality of the classification e.g. to end-users.

RESULTS

Table 2 gives an overview of the experiments conducted with our data. In the calibration scenario, 46 papers from eCAADe 2015 were classified with 83% accuracy. The experiment was repeated for all other papers of eCAADe 2015. Since this dataset does not contain any papers which are purely 'design' or 'science', one would assume that the classification gives neither of both. However, this is not the case since the classification *will* categorize into 'design' or 'science' regardless of whether that makes sense, which is certainly one of the shortcomings of that approach. In that case, 57% of the papers were classificed as 'science' and the rest as design. This is a typical rate, which we also saw when repeating the experiment with data from eCAADe 2016 (see again Table 2).

Table 1
Measures ranked by certainty.

Measure	Criterion	Rule	Domain	Certainty
1	Share of figures vs. other blocks in paper body	if value < 0.17 then science else design	[0..1]	74%
2	Number of references	if value < 16 then design else science	[0..N]	72%
3	Characters in paper body divided by number of figures plus 1	if value < 2500 then design else science	[0..N]	72%
4	Measure 3 divided by number of characters in body	if value < 0.109 then design else science	[0..1]	70%
5	Number of figures in paper	if value < 9 then science else design	[0..N]	70%
6	Share of heading 1 vs. other blocks in paper body	if value < 0.119 then science else design	[0..1]	70%
7	Number of article references	if value < 5 then design else science	[0..N]	70%
8	Share of heading 1 containing standard titles* vs. total number of heading 1	if value < 0.34 then design else science	[0..1]	70%
9	Number of characters in the whole paper	if value < 22500 then design else science	[0..N]	67%
10	Number of characters in the references section	if value < 2000 then design else science	[0..N]	67%
11	Number of tables in paper body	if value = 0 then design else science	[0..N]	67%
12	Share of characters in abstract vs. characters in the paper's body	if value < 0.06 then science else design	[0..1]	67%
13	Number of inproceedings references	if value < 2 then design else design	[0..N]	67%
14	Number of blocks in paper	if value < 65 then design else science	[0..N]	65%
15	Number of characters in paper body	if value < 18500 then design else science	[0..N]	65%
16	Characters in paper body divided by number of figures and tables plus 1	if value < 2500 then design else science	[0..N]	65%
17	Measure 16 divided by number of characters in body	if value < 0.109 then design else science	[0..1]	65%
18	Number of figures plus number of tables	if value < 9 then science else design	[0..1]	65%
19	Number of formulas	if value = 0 then design else science	[0..N]	65%
20	Share of the number of characters contained in paragraphs vs. total characters in paper body	if value < 0.67 then science else design	[0..1]	65%
21	Share of characters in figure captions vs. chars in paper body	if value < 0.09 then design else science	[0..1]	65%
22	Number of thesis references	if value = 0 then design else science	[0..N]	65%
23	Share of article references vs. number of references	if value < 0.27 then design else science	[0..1]	65%
24	Share of chars contained in article references vs. chars of reference section	if value < 0.28 then design else science	[0..1]	65%
25	Share of references available in CuminCAD vs. number of references	if value < 0.23 then science else design	[0..1]	65%

Scenario	Base Data	Investigated Papers	Science	Design	Remarks
1	graph consisting of eCAADe 2014 and 2015 papers together with all referenced papers.	all papers which were manually attributed 'design' or 'science' (ground truth, n=46)	22 (48%)	24 (52%)	comparison of manual classification with automated classification: 38/46 papers categorized correctly (83% accuracy). mean certainty 68%.
2	same as Scenario 1.	all papers excluding the ones of scenario 1 (n=231)	131(57%)	100 (43%)	mean certainty: 68%
3	same as Scenario 1.	all papers (n=277)	153 (55%)	124 (45%)	mean certainty 68%.
4	graph of eCAADe 2015 and 2016 papers together with their references	all papers (n=289)	160 (55%)	129 (45%)	mean certainty 67%
5	eCAADe 2014, 2015 and 2016.	all papers (n=418)	231 (55%)	187 (45%)	mean certainty 67%

Table 2
Conducted experiments (combination between loaded base data and nodes selected for analysis)

DISCUSSION

Even though we can predict 83% of all papers in eCAADe 2015 correctly, it is clear that our calibration data is not statistically significant and we were not able to use that as general classification. However, in the scope of eCAADe 2015, we can nevertheless interpret the measures that helped us match all papers as intent of the authors, and infer some general remarks for the whole conference. If and to what extent these statements may also be applicable to other "CAAD" conferences remains to be studied in the future.

What designers want

A picture gallery. The unusually high share of figures in the paper body (>= 0.17) tells us that there is a need for many figures. ProceeDings is currently very limited with regards to its abilities in layout, since figures may only be placed in a column or at the top/bottom of the page, spanning two columns. A further option which we also provide, the positioning of figures on an own page, might be a good match for this urge to put many figures into a paper: Combining many pictures into a single figure by use of an image editing software [or perhaps a "picture gallery" block which is to be implemented in the future] could let authors use even more photographic material without having to care about placement.

Credits in addition to references. To say that authors of design papers do not reference (number of references < 16) would blatantly ignore the fact that referencing works differently in those cases. Looking through these papers, one can see that most figure captions contain a part mentioning some source (typically the designer or architect who has created the structure or artwork being displayed). Thus, it might be beneficial to include an additional citation type [perhaps named *credit*], which can be counted in the same way as references. Another observation is that designers use more references from CumInCAD, so it might be beneficial to include an "insert reference from CumInCAD" functionality into ProceeDings in the future.

Less text in the paper body, more in the abstract. It is misleading to think that design papers have less text. The measure *(characters in a paper divided through the number of images plus 1)* only states that ratio between text and figures is somewhat smaller that in those papers, which is clear when one thinks about the purpose of such papers: Scientists use figures merely as illustration, while designers put figures in because they are the actual content and source of inspiration for others. However, the typical guidelines of conferences such as eCAADe state that figures should be used only in the former sense. It might be possible to devise another type of publication [maybe called a *graphical entry*] which capitalizes on the visual expression sought by designers. Other conferences furthermore have short papers accompanying posters, which could turn out as possible route in that sense. A further measure supporting this observation is the length of the abstract in comparison with the rest of the paper: Designers seem to like longer abstracts stating their motivation, while scientists try to use that only as a summary and expand the whole paper in the introduction.

Less structure, more freedom. There is a lack of standard headings when it comes to design papers. From looking through the papers, it is evident that this is a result of not presenting research in the classical sense - from 'introduction' to 'conclusions' - but rather describing the development of a final product and its reception e.g. in exhibitions. As eCAADe is also about education, it might be reasonable to adapt the content guidelines so as to accommodate reports of this nature. This might in turn also benefit the papers which we categorized as "project descriptions" or "education", which have a similar way of presenting their content.

What scientists want

Text, text, text. Scientific papers have a high share of text in comparison with other material in the paper body (< 0.17) . The text import wizard of ProceeDings is clearly geared towards this way of writing, as is LaTeX: A low amount of text makes it difficult for to position figures, which is causing major complications during layouting.

Tables and formulas. Formulas are indicative for research papers, as are tables for giving supporting evidence . Both paragraph types are currently quite limited, since ProceeDings currently converts formulas from ASCIIMath [1] into LaTeX for compilation and allows only images as tables, which prohibits floating across multiple pages ["longtable" in LaTeX terms]. The first shortcoming could be alleviated by using the graphics generated by the ASCIIMath preview directly, the second by using a genuine table paragraph type. This type already exists in ProceeDings but was considered a bit too clumsy (it comes with a table editor that is fairly intuitive, however, rows and columns have to be added consecutively and it takes quite a time to paste text into these). One could offer both possibilities an have the authors choose either of them, and perhaps also implement an import from spreadsheets.

A lot of references. Scientific papers excel at their use of references: In almost every category (notably except URLs), scientists use more and higher-quality references (articles and proceedings rather than books). Currently, the import of references into ProceeDings is limited to BibTeX, which may be a limiting factor. However, one could invest some effort into text recognition, since it can be expected that authors would want to copy-and-paste citations from the web into our system. Another option would be to directly connect to citation indices such as CumInCAD, Google Scholar or CiteSeer, in order to get citation metadata quicker and in a more accurate fashion.

RELATED WORK

Instead of presenting the related work in between the introduction and the idea, we have opted to put it here since the reader now knows about the extent of our work and can thus relate to this section more easily.

Autodesk Research has produced a similar research effort in their paper "Citeology - Visualizing Paper Genealogy" (Matejka, Grossman and Fitzmaurice 2012). The authors establish a visualization in which a selected paper can be traced back to the work it cites and the work these predecessors cite and also forth. In contrast to that effort, we have concentrated on content analysis and inference of some remarks concerning our conference, without the intended goal to propose a novel computational approach.

Cerovsek and Martens (2004) have used CumInCAD to build up a (chronological) citation graph and to infer statements on scientific information exchange, for example that the CAAD field uses "many more references" (p. 15) than other fields (taken from Open Citation Project 2002), and that most scientific ideas evolve over a period of 2 years (Figures 7 and 8 on p. 11).

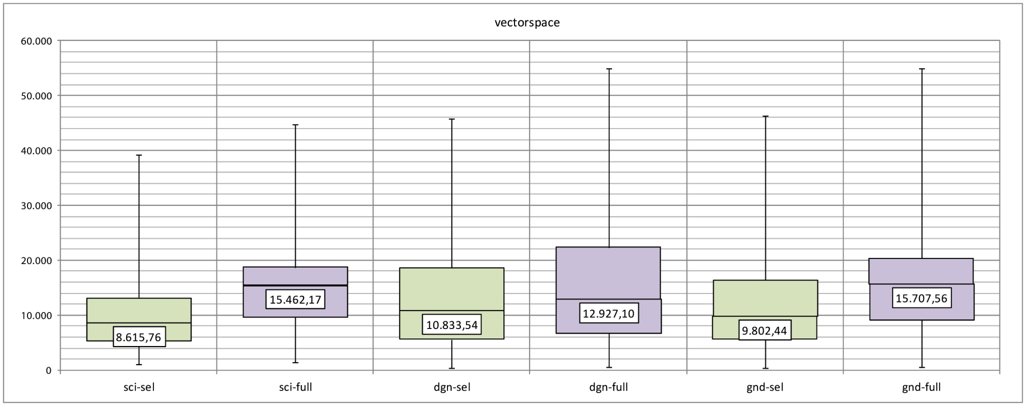

Figure 5
Relative euclidean distance between ground truth papers (within sci=science, dgn=design, or gnd=all) using our measures to build up a vector space (sel=top 25 selected measures as in Table 1, full=all 110 measures).

FUTURE DIRECTIONS

One may obtain a vector for each paper by applying either all measures or just the ones selected for classification as given in Table 1. The resulting vector space may be used to calculate the euclidean distance between papers of the ground truth (see Figure 5). Results show that distances between science papers are less spread than design ones, and change significantly depending on whether we use all measures or just the ones selected for classification (see median of sci-sel vs. sci-full in Figure 5). As outlook, one could furthermore perform clustering, identify representative parsimonious "aggregated paper features" or conduct shopping basket analysis (people who read paper x also read papers ...), which would certainly yield interesting results.

CONCLUSIONS

We know from practice that the research cultures of design and science need different modes of reporting about their work. To put this statement on a firmer basis, we have been conducting a study on all papers of eCAADe 2014-2016, using ProceeDings and CumInCAD as our data basis. Through use of automatic categorization into science and design, we were able to extract metainformation for both types. The results have been discussed and interpreted as possible future extension points for the ProceeDings system. As side-effect, we are now also able to categorize papers into 'design' and 'science' volumes, which is typical at least for the eCAADe conference.

ACKNOWLEDGEMENTS

The team would like to acknowledge the contribution of Claudia Lorenz who invested valuable time in numerous discussions regarding statistical relevance.

REFERENCES

Cerovsek, T and Martens, B 2016 'CumInCAD 2.0: A Redesigned Scalable Cloud Deployment', *Proceedings of eCAADe 2016 (Volume 1)*, Oulu, pp. 485-492

Cerovsek, T and Martens, B 2004, 'On the Extended Use of Citations in CAAD', in Van Leeuwen, JP and Timmermans, HJP (eds) 2004, *Developments in Design and Decision Support Systems in Architecture and Urban Planning*, Eindhoven University of Technology, ISBN 90-6814-155-4, Eindhoven, pp. 1-17

Matejka, J, Grossman, T and Fitzmaurice, C 2012 'Citeology - Visualizing Paper Genealogy', *Proceedings of CHI 2012*

Citations Project, Open 2004, *Reference Linking and Citation Analysis for Open Archives*, http://opcit.eprints.org/

Wurzer, G, Martens, B and Grasl, T 2014 'ProceeDings - A web-based word processor automating the production of conference proceedings', *Proceedings of eCAADe 2014 (Volume 1)*, Newcastle, pp. 11-20

[1] http://asciimath.org