Naukkarinen, Ossi; Bragge, Johanna

Aesthetics in the age of digital humanities

Aesthetics in the age of digital humanities

Ossi Naukkarinen & Johanna Bragge

Aesthetics in the age of digital humanities

Ossi Naukkarinen[1] and Johanna Bragge[2]⋆
[1]Department of Art, Aalto University School of Arts, Design and Architecture, Helsinki, Finland; [2]Department of Information and Service Economy, Aalto University School of Business, Helsinki, Finland

Abstract

One of the most difficult but yet unavoidable tasks for every academic field is to define its own nature and demarcate its area. This article addresses the question of how current computational text-mining approaches can be used as tools for clarifying what aesthetics is when such approaches are combined with philosophical analyses of the field. We suggest that conjoining the two points of view leads to a fuller picture than excluding one or the other, and that such a picture is useful for the self-understanding of the discipline. Our analysis suggests that text-mining tools can find sources, relations, and trends in a new way, but it also reveals that the databases that such tools use are presently seriously limited. However, computational approaches that are still in their infancy in aesthetics will most likely gradually affect our understanding about the ontological status of the discipline and its instantiations.

Ossi Naukkarinen, PhD, is Head of Research and Vice Dean at the Aalto University School of Art, Design and Architecture, Finland. He has published books and articles on various themes in aesthetics, including environmental art, everyday aesthetics, and mobile aesthetics, in journals such as *Contemporary Aesthetics*, *Aisthesis*, and *Nordic Journal of Aesthetics*. His publications have also been translated into Spanish, Slovenian, Italian and Chinese.

Johanna Bragge holds a PhD in Management Science from the Helsinki School of Economics and works as Senior University Lecturer of Information Systems Science at Aalto University School of Business. Her research interests include research profiling with text-mining tools, e-collaboration, service co-creation, and crowdsourcing. Her research has been published, among others, in the *Journal of the AIS*, *IEEE Transactions on Professional Communication*, *Expert Systems with Applications*, *Futures*, *Group Decision and Negotiation*, and *Journal of Business Research*.

Traditionally, well-informed conceptions about the field of aesthetics have been formed by studying it for a long time and carefully—by reading and writing books and articles, teaching and following lectures, and taking part in academic discussions in conferences and learned societies. This is still quite a normal and reasonable approach, and knowledge attained through it cannot be achieved in any other way. The more one studies, the broader and more detailed a picture one has.

However, there is no universally accepted definition of aesthetics. We can probably agree that aesthetics has something to do with the arts, beauty, and other aesthetic values, as well as with art criticism in the broadest possible sense. As soon as one goes any further, philosophical ponderings and disagreements arise. What kinds of studies of the arts actually belong to the field of aesthetics, and what kinds are outside it? Where are the differences between art history and aesthetics?

⋆Correspondence to: Johanna Bragge, Department of Information and Service Economy, Aalto University School of Business, Runeberginkatu 22–24, FI-00100 Helsinki, Finland. Email: johanna.bragge@aalto.fi

Is aesthetics always a philosophical discipline, and what does that mean? Should we include non-academic publications such as memoirs or exhibition reviews in the field if they deal with the same themes as academic aesthetics papers? What are the latest trends and which themes are fading away? Such questions are acutely relevant when one designs, for example, an introductory book or course for undergraduate students.

Forming a comprehensive interpretation of any academic field is becoming more and more demanding all the time, because the number of publications, scholars, and institutions becomes higher each year. One simply cannot master all the different languages and traditions in which aesthetic issues are addressed, and a single scholar can never get a hold of everything published in his or her field. In fact, he or she cannot even access a tiny fraction of it, since in general the growth rate of traditional scientific publishing has been increasing for the last 50 years, and the number of publications using new channels such as open-access journals is growing rapidly.[1] The latest studies show that the growth of global scientific publication output has been exponential between 1980 and 2012.[2] The same trend can also be seen in the research of aesthetics, as presented in Figure 1. In the data set we analyzed from the Web of Science (WoS) for this article, the rate of growth has been steady and surging since the turn of the millennium.

How can we make sense of aesthetics in this situation? Well-informed understanding of one's field is still expected of professionals, after all; one is not supposed to focus on some narrow area only, without the ability to link one's specialty to a wider field. Like in many other contexts, that is the reason why it is reasonable to make use of the computa-
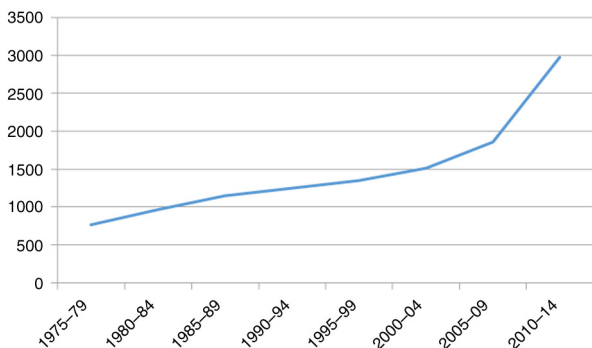

Figure 1. Increase in research articles in aesthetics in the Web of Science.

tional tools that we have nowadays. So far, aestheticians have not been very active in using these for clarifying the nature of their own field.[3] Our aim is to open up some possibilities and thus take aesthetics closer to the so-called Digital Humanities.[4] Furthermore, we want to show that using such tools is not as easy and straightforward as one could assume, but it requires specialized skills.

THE BIG PICTURE

Computational analyses always need data to be analyzed with the help of algorithms that define what the computational processes will do and present to us. For this essay, we have used the publication data provided by WoS.

Thomson Reuters' WoS database is the "gold standard" by which many governments in countries such as the USA, the UK, and Australia evaluate their national R&D performance.[5] It was also the first database that started indexing the cited references of publications, as early as the 1960s, thus allowing various (co-)citation analyses to be conducted, based on Eugene Garfield's original idea from 1955.[6] WoS is also used as standard by researchers for bibliometric studies, as the publications it indexes are stringently pre-inspected for quality, and the data it provides is consistently organized in the database. To summarize, as WoS is one of the best known, largest, and most influential academic databases, it is interesting to see first what kind of image it offers of aesthetics.

It is a known issue that arts and humanities (A&H) research is not as well covered in WoS as natural sciences—the indexing of A&H started much later, in 1975—although the situation has been improving lately.[7] We will return to some of the problems related to WoS and other similar databases, such as Elsevier's Scopus, soon. In any case, as WoS is arguably one of the most important academic databases, aestheticians cannot afford to ignore it. At least, we have to understand how it functions. If the picture seems to be distorted, we have to understand why, and try to find better tools and databases. In the present situation, where such databases have a huge impact on our academic communities, we cannot just omit them.

The data we consult does not tell us anything as such, and we cannot even start searching for relevant information without making active selections. When we created a picture of aesthetics

using WoS, we had to narrow down our approach, as will be explained soon. In addition, we chose three software tools to represent and organize the core results: VantagePoint, VOSviewer, and Leximancer.[8] VantagePoint is a professional text-mining tool for discovering and organizing information in search results from literature or patent databases. Besides advanced data cleaning functions, it makes it possible to quickly find answers to the questions of who, what, when, and where, helping the researcher to clarify relationships and find patterns. The second tool, VOSviewer, also analyzes bibliometric literature data, but its core focus is on visualizing the bibliometric networks, composed, for example, of journals, authors or key terms appearing in abstracts, based on co-citation, bibliometric coupling or co-occurrence relations. Leximancer is an automated content analysis tool that can be used to find prominent themes and concepts from any kind of textual data, whether bibliometric or otherwise. We used it to analyze the full texts of the *British Journal of Aesthetics* in 5-year periods.

The time span we analyzed was 1975–2014. The A&H citation index starts at 1975, and at the time of conducting the study, we were halfway through 2015. In addition, the span covers exactly 40 years, and thus allows long-term trend analyses to be conducted, for example, by 10-year periods.

If one simply types "aesthetics" in the basic search field of WoS, which searches for the term in titles, abstracts, and keywords, the search results (22,957 publications as of August 4, 2015) largely, at around 55%, feature publications other than A&H ones, such as life sciences and biomedicine from the other citation indexes. Figure 2 shows the division by scientific domain, as well as by more detailed research area in the A&H domain.[9]

This, in fact, is interesting as such: unlike we philosophers of aesthetics might believe, a large group of people addressing aesthetic issues seems to be operating outside our circles, even if our and their discourses seldom meet. If this is the case in academic contexts such as WoS and other similar databases, what is the situation outside academia? We will briefly return to this question at the end of this essay, but, all in all, the issue of how "our" and "their" aesthetics are related would actually deserve a study of its own.

This time, however, we wanted to keep the focus closer to what we think is the humanistic tradition of aesthetics. For this, we restricted the search to only the A&H citation index. Even that database initially seemed too large, as the same "aesthetics" search brought up many seemingly irrelevant research areas, such as radiology, nuclear medicine, and medical imaging. However, we looked into some of those and found that they can actually
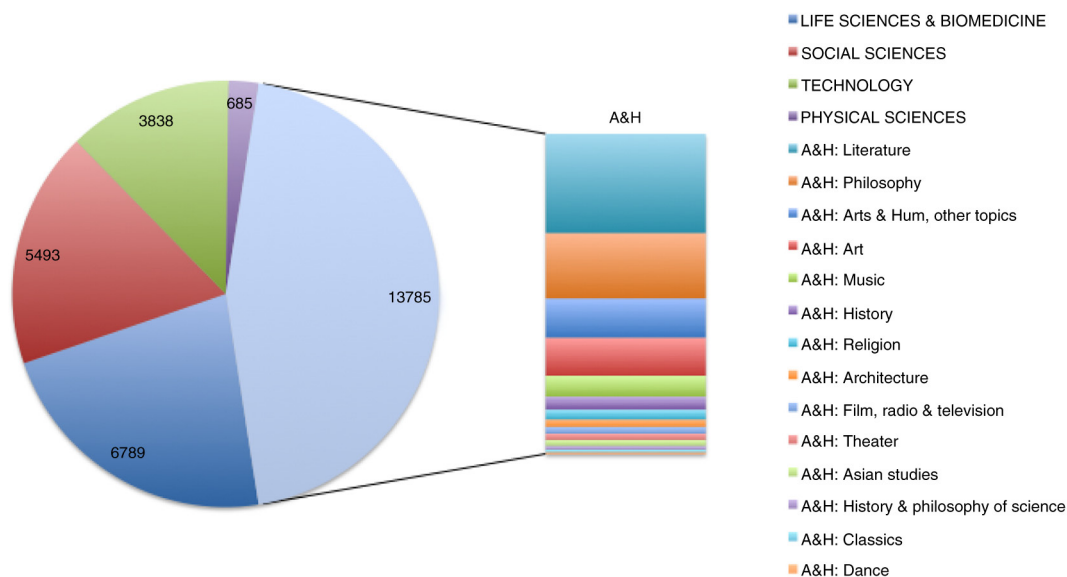

Figure 2. Aesthetics publications by scientific domain (in capital letters; A&H = Arts & Humanities) and research area in the Web of Science.

include interesting publications. For instance, they showed that the radiologic aesthetics of human body parts or organs have inspired many artists to create works of art, indicating that radiology is perhaps becoming a more common approach in the field of contemporary art. Thus, we decided to include all results from the A&H index that had "aesthetics" in the title, abstract or keywords. In addition, we included all publications from journals that are specific to aesthetics. The aesthetics journals that are indexed in A&H by WoS include the following: *British Journal of Aesthetics* (BJA), *Journal of Aesthetics and Art Criticism* (JAAC), *Journal of Aesthetic Education* (JAE), *International Review of the Aesthetics and Sociology of Music* (IRASM), *Revue d'esthétique*, *Psychology of Aesthetics Creativity and the Arts*, *Estetika – the Central European Journal of Aesthetics*, and *Zeitschrift für Ästhetik und allgemeine Kunstwissenschaft*.

Had we chosen to focus on more specifically defined research areas in A&H, it would have required more active definition, and there is no single solution for that. This means that computing is necessarily combined with a philosophical analysis of what aesthetics is. For example, it is quite reasonable to state that aesthetic issues are most probably dealt with in publications listed under research areas such as art, literature, and philosophy, because aesthetics is often related to the themes of art, criticism, and beauty and is emphatically philosophical in nature. On the other hand, if one chose some other set of fields, the search results would be somewhat different. If one assumes a more Baumgartian stance, understanding aesthetics as something close to "a science of sensitive knowing" (*scientia cognitionis sensitivae*), one would probably include more publications and fields closer to psychology; and emphasizing evolutionary, neuroscientific, or environmental branches of aesthetics could lead to including more fields of natural sciences. This means that one's pre-understanding necessarily guides what one finds from the data that is available. It is evident that there is no single, objective, and neutral way of selecting the relevant fields when doing a more focused analysis.

The aesthetics search in A&H index, including the eight domain journals mentioned above, resulted in 21,919 publications (as of June 18, 2015). As our purpose is to illustrate especially academic research in aesthetics, we refined the results to

include only full-length journal and conference articles, thus excluding, for example, book reviews, letters, and notes. This choice was guided by the category options WoS offers, and our final search result was 11,814 articles.

The results based on our selections show, first, that even if there are some self-evident forums of aesthetics, such as *BJA*, *JAAC*, and *JAE*, issues related to aesthetics are addressed in surprisingly many sources, some of which were previously unknown to us. In total, there were altogether 1,517 different journals or other sources listed as publishing aesthetics articles. This means that we might need to broaden our own understanding of the field, of its publication channels, and of who is actually working in it. Of course, this data analysis only suggests some possibilities and opens questions, and we have to study the phenomenon better by other means, including plenty of good old-fashioned reading. We have to find out whether the publications based on our search really are relevant to aesthetics, and whether the text-mining tools produce truthful results when making more detailed analyses. In any case, the point is that we would not have seen the new possibilities in the same way without the data analysis, and at least some of the new sources will probably turn out to be important.

On the other hand, it is striking that many journals that we think are relevant and interesting for the field are missing (not indexed) in WoS: *Journal of Aesthetics & Culture*, the Italian *Aisthesis*, the US-based *Contemporary Aesthetics*, *Journal of Aesthetics and Phenomenology*, and *The Nordic Journal of Aesthetics*, for example, not forgetting some of the perhaps lesser-known publications, such as *The Journal of Aesthetics and Protest*, *Aisthesis—International Journal of Art and Aesthetics in Management and Organisational Life*, and *Korean Journal of Aesthetics*. This is due to the very strict indexing principles of WoS. It is evident that one cannot blindly trust the computed results, but one needs to be aware of the database restrictions.

The data also shows that 93% of the articles are single-authored and reveals who are the most active and prominent scholars in the field. There are no big surprises. The top authors who have published most articles are all internationally familiar names. The top 10 are, respectively, Noël Carroll, Richard Shusterman, Peter Kivy, Robert Stecker, Stephen Davies, Jerrold Levinson,

Harold Osborne, Stanislav Tuksar, Malcolm Bud, and Joseph Margolis (all men!)—the only surprise perhaps being Stanislav Tuksar, the Croatian music scholar.

We had more or less assumed a list of this kind, but now we have evidence for our belief, and we can also see in more detail how much and where these scholars have actually published, and how many citations they have received for the articles (see Table 1). This, in turn, gives others a reference point: if someone wants to be active and visible in aesthetics, where and how often should one present one's ideas? In this data set, Carrol has 47 articles and Margolis 20, the other top authors something between this, and by far the most important publication forums are the *Journal of Aesthetic Education, Journal of Aesthetics and Art Criticism* and *British Journal of Aesthetics* – except for Tuksar, who has mostly published in the *International Review of the Aesthetics and Sociology of Music*, for which he is editor-in-chief. So, it might be a good idea to aim at these journals and publish at least some 20 articles, which is naturally not that easy.

The list of top *cited* authors, which is collected from the reference lists of our final sample of 11,814 articles, looks a little different, due to the fact that classics of philosophy, such as Immanuel Kant, are still commonly cited in the field.

However, all but one of the top-10 authors also appear among the top-60 cited authors. Table 2 presents the top 50 most cited authors, based on the number of publications in which they have been cited.[10] The table also divides the number of citing publications temporally into four decades. It is interesting to see that most of the top cited authors have an ascending trend in citations, but there are also some whose curve is descending. The top authors appearing in Table 1 have been shown in bold in Table 2 for easier detection; Robert Stecker and Harold Osborne are not shown as they are at places 57 and 60, respectively. In addition, Stanislav Tuksar's rank is 1558, with 15 sample publications in which he is cited.

Bibliometric studies typically analyze and visualize author networks via their co-authorship relations, revealing "scholarly communities." However, in the case of aesthetics and in the humanities in general, co-authorship analyses are not sensible, as our data shows that 93% of the articles are single-authored. To discover relations, one can instead conduct other types of network analyses, for example by cross-correlating authors with the help of commonly used title words or through the authors they refer to in their articles. Figures 3 and 4 illustrate two examples of such cross-correlation analyses. The most prolific authors are placed on the map based on the authors they cite

Table 1. Top-10 authors

Rank	Author	Number of articles	Percentage published in BJA, JAE or JAAC	Total cites for the articles	Avg. cites for the articles	Author's h-index for the articles[a]	Rank in top cited authors list
1	Carroll, Noël	47	85	368	7.83	12	13
2	Shusterman, Richard	39	79	185	4.74	7	27
3	Kivy, Peter	31	90	88	2.84	5	23
4	Stecker, Robert	31	97	145	4.68	7	57
5	Davies, Stephen	29	93	164	5.65	8	34
6	Levinson, Jerrold	28	93	249	8.89	8	11
7	Osborne, Harold	27	100	63	2.33	5	60
8	Tuksar, Stanislav	24	0	14	0.58	3	1558
9	Budd, Malcolm	20	90	105	5.25	6	45
10	Margolis, Joseph	20	95	79	3.95	6	41

[a]Hirsch's h-index: An author has index h, if h of his N_p papers have at least h citations each, and the other (N_p – h) papers have less than h citations each.

Table 2. Top 50 most cited authors in the 11,814 aesthetics articles, by decade

Rank	Number of publications in which author is cited		1975–1984	1985–1994	1995–2004	2005–2014
1	996	Kant Immanuel	118	182	283	413
2	617	Adorno Theodor W.	65	127	131	294
3	548	Benjamin Walter	32	92	132	292
4	547	Goodman Nelson	126	151	144	126
5	512	Danto Arthur	54	129	157	172
6	466	Hegel G. W. F.	82	115	109	160
7	448	Barthes Roland	45	90	113	200
8	448	Beardsley Monroe C.	129	128	97	94
9	448	Foucault Michel	23	70	115	240
10	433	Derrida Jacques	25	90	104	214
11	424	**Levinson Jerrold**	4	59	152	209
12	420	Wittgenstein Ludwig	75	98	124	123
13	405	**Carroll Noël**		38	140	227
14	401	Walton Kendall L.	30	66	115	190
15	378	Dewey John	48	74	92	164
16	377	Heidegger Martin	39	86	99	153
17	373	Nietzsche Friedrich	29	85	112	147
18	372	Gombrich Ernst	91	98	93	90
19	363	Wollheim Richard	63	93	106	101
20	357	Aristotle	65	70	85	137
21	347	Dickie George	73	95	94	85
22	341	Deleuze Guilles	5	29	96	211
23	329	**Kivy Peter**	17	72	110	130
24	328	Bourdieu Pierre	12	30	85	201
25	315	Scruton Roger	40	70	84	121
26	311	Freud Sigmund	29	63	75	144
27	282	Plato	48	58	74	102
28	269	Hume David	37	42	84	106
29	258	Langer Suzanne	87	66	45	60
30	257	Schiller Friedrich	41	49	67	100
31	249	Jameson Fredric	13	32	54	150
32	243	Lyotard Jean-François	3	53	71	116
33	242	Gadamer Hans-Georg	38	58	59	87
34	235	**Davies Stephen**	4	21	71	139
35	235	Eagleton Terry	10	40	62	123
36	226	Arnheim Rudolph	39	64	44	79
37	217	Eco Umberto	23	53	53	88
38	215	Merleau-Ponty Maurice	38	36	55	86
39	212	Marx Karl	57	39	39	77
40	207	Dahlhaus Carl	27	59	56	65
41	202	**Margolis Joseph**	51	66	44	41
42	200	Goethe Johann Wolfgang von	38	43	52	67
43	199	Collingwood Robin	52	54	52	41
44	198	Sartre Jean-Paul	43	40	45	70
45	196	**Budd Malcolm**	1	22	72	101
46	193	Habermas Jürgen	19	55	57	62
47	192	**Shusterman Richard**	5	29	55	103
48	188	Eliot Thomas S.	59	34	47	48
49	178	Currie Gregory		7	51	120
50	177	Baudelaire Charles	29	24	54	70

Top authors from Table 1 are indicated in bold font.
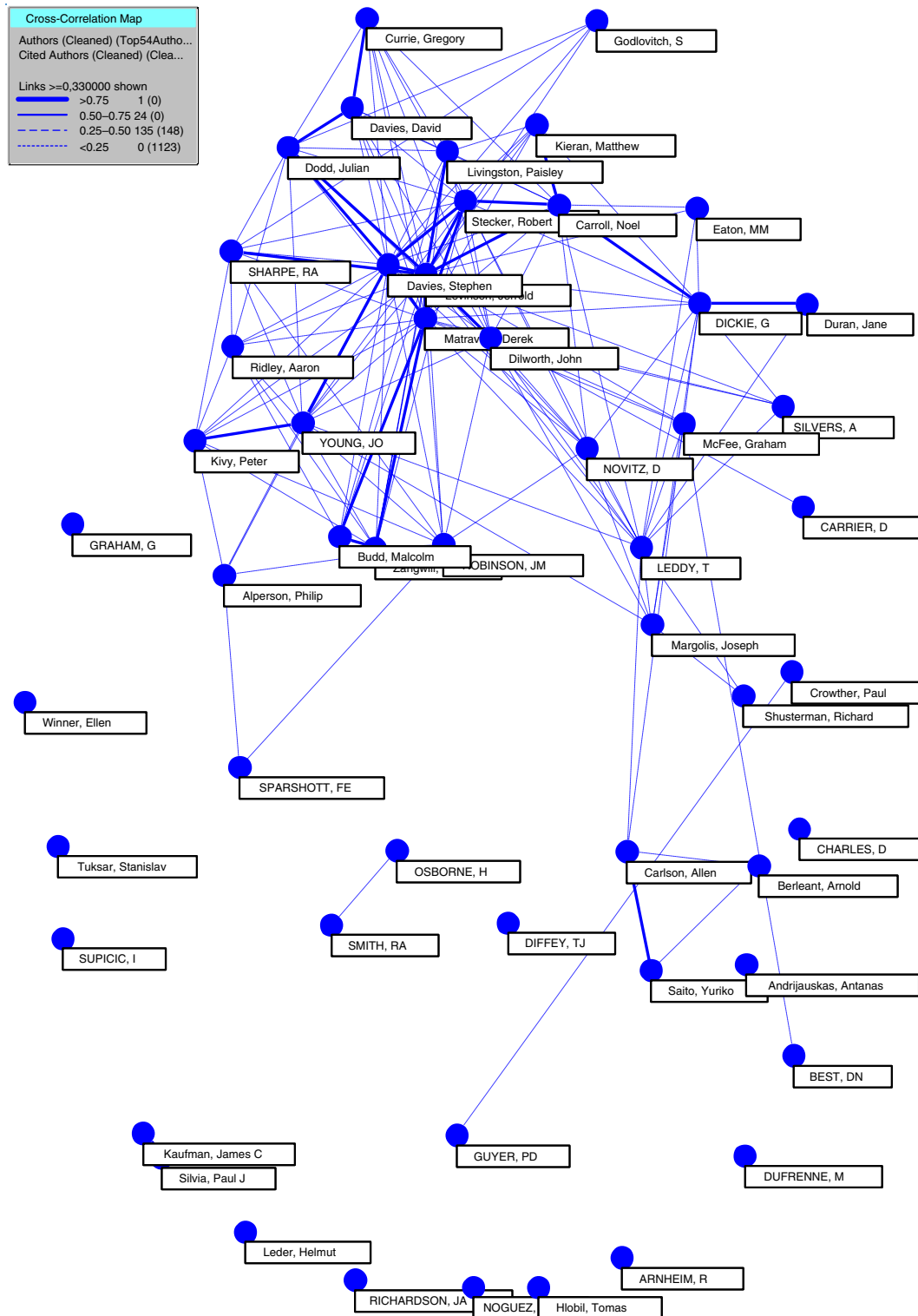
Figure 3. Cross-correlation map of top-54 authors vs. cited authors (top 1072). Jerrold Levinson's label appears underneath that of S. Davies, and Nick Zangwill's label underneath M. Budd.

(Figure 3) or on the title words they use (Figure 4). The correlations are shown as links between author nodes: the thicker the link lines, the greater the correlation between any two authors (see legends in the upper left hand corners).[11] One can also study the basis of the correlation using the tool online: when hovering the mouse above any author node, the tool will present information showing
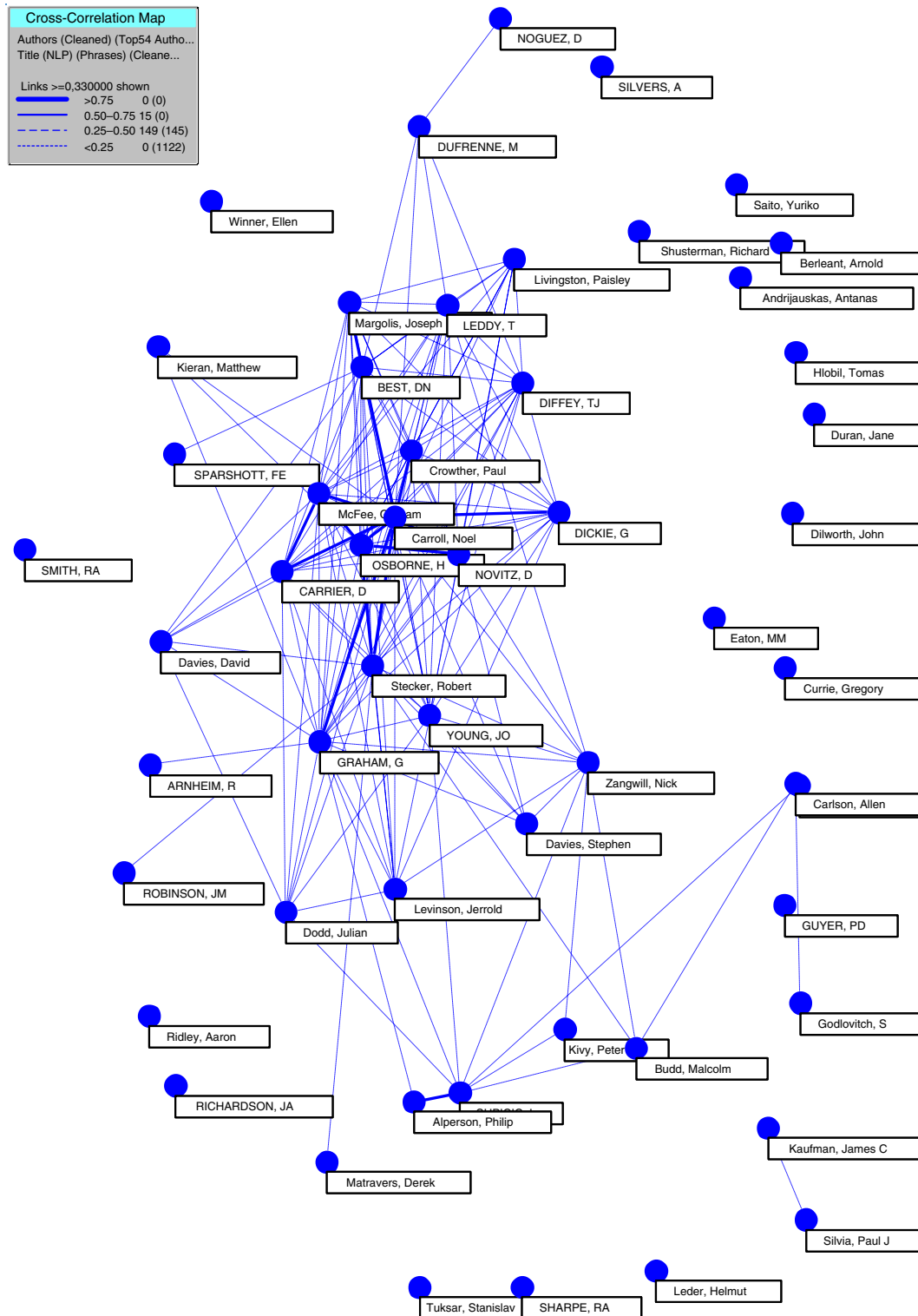
Figure 4. Cross-correlation map of top-54 authors' vs. title phrases (all; processed with Natural Language Processing NLP). Ivo Supicic's label appears underneath that of Philip Alperson.

the values for the cross-correlated field. To take rather an easy example, the tool shows that Joseph Margolis most often cites (besides his own works) Nelson Goodman, Arthur C. Danto, Jacques Derrida, Willard Van Orman Quine, and Donald Davidson; for a knowledgeable reader, this kind of information immediately says something about his approach.

If the data shows that two or more authors are closely related and we had not realized that before, now we have a reason to examine *how* they are related. This, again, requires consultation of the actual publications, but the text-mining tool has given us a reason to do that, as it gives an indication of the nature of the relationships. Without the tool, we would never have detected all such relations. Robert Stecker, for example, seems to be very well connected in many directions; how exactly and what this indicates is a matter for further analysis. On the other hand, it is interesting that the pictures do not show a stronger relation between authors such as Arnold Berleant and Yuriko Saito, even if we know from other sources that they have often addressed related topics and closely co-operated in other ways, for example, in the e-journal *Contemporary Aesthetics*; again, the results must be read critically.

One interesting result is the heat map (Figure 5) of the most common themes, as seen through the frequency and co-occurrence of the terms used in titles and abstracts (when stop words such as "and," "it," etc. are excluded).[12] The warmer the color, and the larger the font size, the more often the terms appear in the sample. For example, the term "politic" appears in the hot red area and in medium-large font, and the data behind it indicates that the term appears in the

title or abstract of 758 publications (counted only once if it appears in both). The proximity of terms indicates that they often appear in the same titles or abstracts. The map helps us to quickly see the most usual themes or issues addressed in aesthetics.

Again, the map requires interpretation and further study. As it shows that, for example, "politic" is a frequently used term in the field, this might mean that if one wants to be a credible aesthetician, one has to pay close attention to it (and its variations political, politics, etc.), even if one had not been very interested in it before. Without data analysis, one would not have as good an idea of *how* common it is, and one would not have an equally good reason to study what kinds of issues are addressed and who is active under its umbrella. Its 758 hits can be compared with the other large topics appearing on the map: music 1,579, philosophy 1,091, beauty 590, poetry 519, and performance 423 hits.

Furthermore, the map indicates how widespread interest is in the sub-fields in which I or someone else is specialized. This helps in relating sub-fields to each other, and provides one approach to the question of how to make sense of the relative weight of sub-fields within the whole field. It is interesting that some relatively new but possibly trending sub-fields, such as "everyday aesthetics," do not (yet)
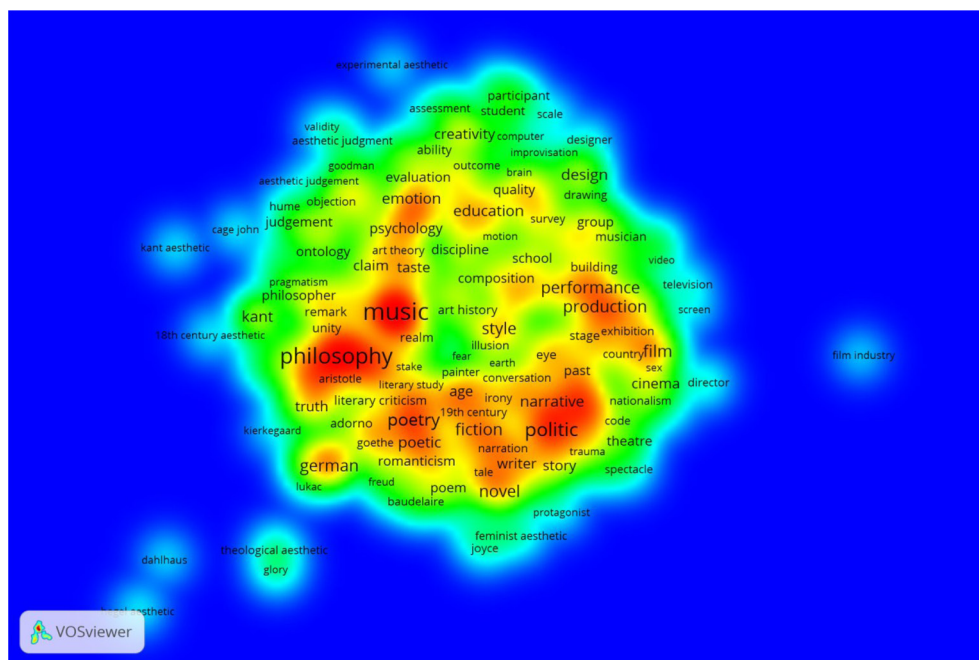

Figure 5. Co-occurrence map of terms in titles and abstracts (all publication types included).

manifest in the previous maps and analyses at all. This might be related to the fact that databases include plenty of old materials, and new themes are necessarily less visible in comparison. However, the tools enable searches for such topics of interest within various data fields, such as authors' key-words, title words/phrases, or abstract words/phrases. One can see that title phrases related to everyday aesthetics are currently occupying the following places in the ranked title word list cover-ing the whole time range of 1975–2014: everyday life (rank 260 with 16 publications), everyday aesthetics (rank 1173 with 4 publications), and everyday aesthetic experience (rank 9658 with 1 publication from 2014).

Figure 6 presents a co-citation analysis of jour-nals, as visualized using VOSviewer. Two journals are said to be co-cited if there is a third journal that cites both journals. The larger the number of journals by which two journals are co-cited, the stronger the co-citation relation between the two journals is.[13] For Figure 6, all journals with at least 30 citations (650) are included in the analysis, even if, due to reasons of clarity, only some of the journal titles are visible. One can see three "hot" areas on the heat map, illustrated by warmer red and yellow colors. The largest concentration is around the core of aesthetics, and this is featured by citations to JAAC, BJA, and JAE. The second center, on the left, is about communication research, and the third relates to publications on psychological issues.

It is also interesting to see how the field has changed over time. The bubble chart produced using VantagePoint (Figure 7) shows the temporal development of the top-15 words or phrases derived via statistical Natural Language Processing (NLP)[14] from the titles of the publications, pre-sented in alphabetical order. Note that the search word "aesthetics" was removed from the figure, as it appears in most titles. From the figure, we can immediately see that aesthetics articles are most often related to arts (in general) and then to music. Moreover, political and ethical topics have visibly increased their prominence during the last few years. With these types of figures, we can also easily detect when certain terms first emerged in the titles during the 40-year sample period, espe-cially regarding the less common terms and so-called emerging terms (not shown in the figure).

It is possible to analyze the temporal devel-opment in time sequences longer than a year, too. Table 3 presents the same top-15 title phrases in table/numerical format across four decades.

All such general results are worth paying atten-tion to when trying to figure out what aesthetics is and how it has changed. Of course, one must know the field rather well already in advance, because otherwise one cannot focus one's search and pay attention to relevant further questions, which are often more or less philosophical in nature. For example, if the analysis suggests a relation between two authors, it is by no means simple and straightforward to say what kind of
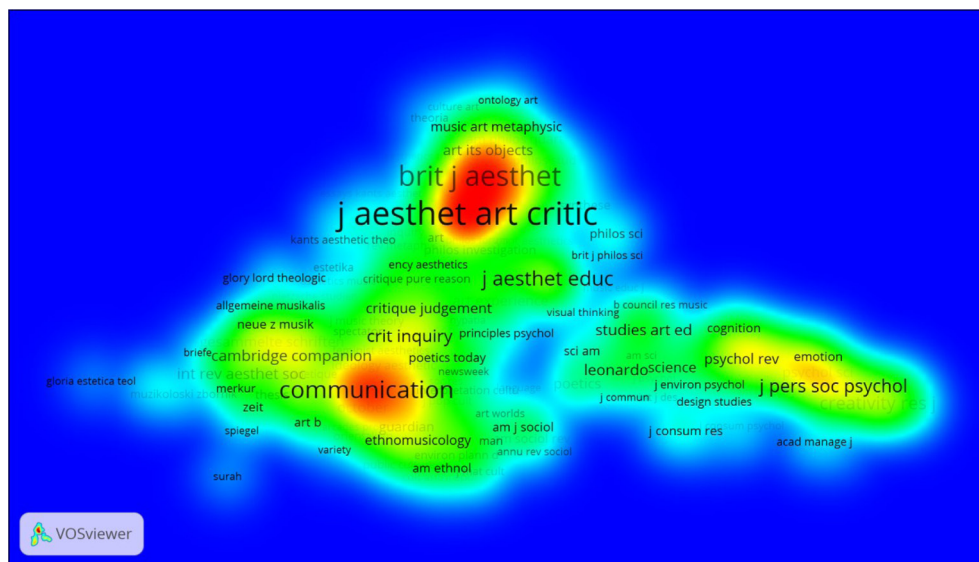

Figure 6. Co-citation analysis of journals (all publication types included).
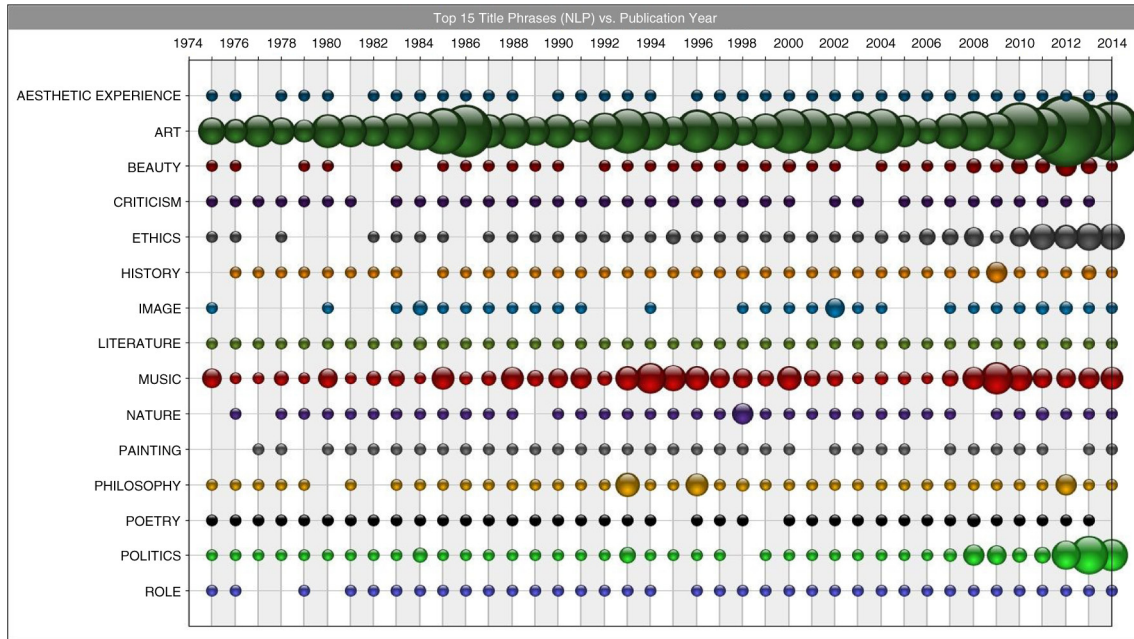
Figure 7. Bubble chart of top-15 title words or phrases.

relation that is. Only if one has enough understanding of the field, can one ponder different alternatives.

In addition, while such tools represent the results as frequency lists, figures, and temporal matrices, as soon as one learns to understand them, they are a very effective way of conveying information; one can see by a single glance much more than by reading a longish text. To our minds, information graphics in the form of science maps and research landscapes have been an under-used possibility in aesthetics. However, it is fairly easy to produce very informative images that could also be used in introductory books and other presentations.

Table 3. Temporal development of top-15 title words/phrases

		# Records in total for decade	1734	2393	2855	4832
Rank	# Records	Title word or phrase	1975–1984	1985–1994	1995–2004	2005–2014
1	896	Art	179	220	226	271
2	422	Music	76	117	110	119
3	220	Politics	23	39	40	118
4	197	Ethics	10	21	51	115
5	175	Philosophy	21	43	54	57
6	158	Beauty	11	24	41	82
7	136	History	14	31	33	58
8	135	Literature	32	31	26	46
9	129	Nature	21	25	46	37
10	116	Poetry	22	28	26	40
11	95	Image	13	19	25	38
12	92	Role	9	29	23	31
13	88	Aesthetic experience	17	13	26	32
14	87	Criticism	30	22	17	18
15	87	Painting	12	24	21	30

"Aesthetics" as the search word is removed from the table from the first row, as well as common research words such as "note," "reply," and "reflections."

A CASE STUDY

Inside the large amount of data, there are more straightforward cases, and it is wise to focus more closely on them. For example, we can tentatively assume that everything that has been published in the *British Journal of Aesthetics* are cases of aesthetics. We can accept that without trying to define what aesthetics is. Instead, we can simply see what there is and take that as one landscape of aesthetics. As we know that BJA is one of the main forums of discussion in the field (see previous section), the picture it offers is probably highly relevant more generally, too.

One could naturally try to read every volume of BJA published since 1960, but even if that might not be completely impossible, it would be an extremely time-consuming job. Furthermore, it is doubtful whether the reader could ever attain similar results to a computer, even if she read the material several times. Computers can do their tricks quickly, and as we have the titles, abstracts, keywords, and other bibliometric data ready at hand, the text-mining tools and algorithms can reveal patterns, trends, relationships, and emerging topics from the data. The advanced text-mining tools are, in practice, analogous to statistical software designed for numerical data.

Using VOSviewer, we can show that BJA looks like this. The map in Figure 8 is based on the title words from all BJA publications; words appearing at least five times are included; not all are visible. BJA has its own profile compared to the field at large.

In addition, as the data set for BJA is smaller, we can drill deeper and use automated content analysis tools such as Leximancer to detect major themes and concepts based on the full texts, not only on titles, abstracts, and key words. In Figure 9, we illustrate the full-text analyses for three separate time periods (1996–1999, 2005–2009, 2010–2014). We had access to PDF documents from 1996 onwards, but the maps do not cover the period of 2000–2004 and the January 2005 issue, as those PDFs were secured and not readable by the text-mining tool.

In the maps, each concept (grey node) is defined by a list of statistically weighted words from the full texts, the comparison of which enables the depiction of associations (closeness and links) between the concepts.[15] Node size indicates the frequency of a concept's appearance. To aid interpretation, the concepts cluster into higher-level themes (colored circles) when the map is generated, and the themes are automatically named according to the largest concept node they include. Colors are heat-mapped to indicate importance, with the most prominent cluster appearing in red, the next most prominent in brownish orange, and
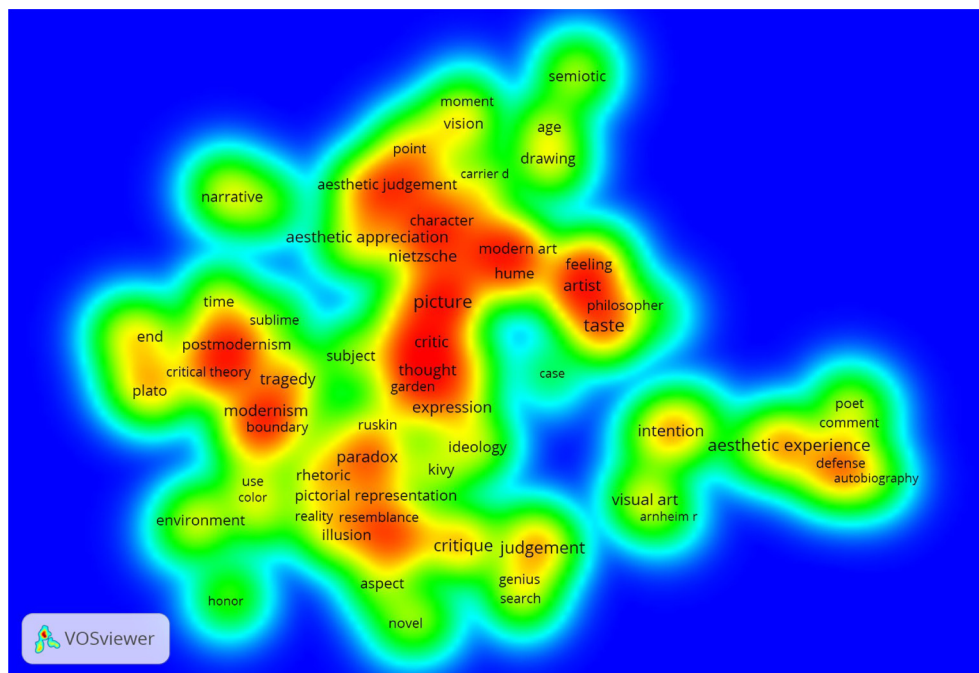

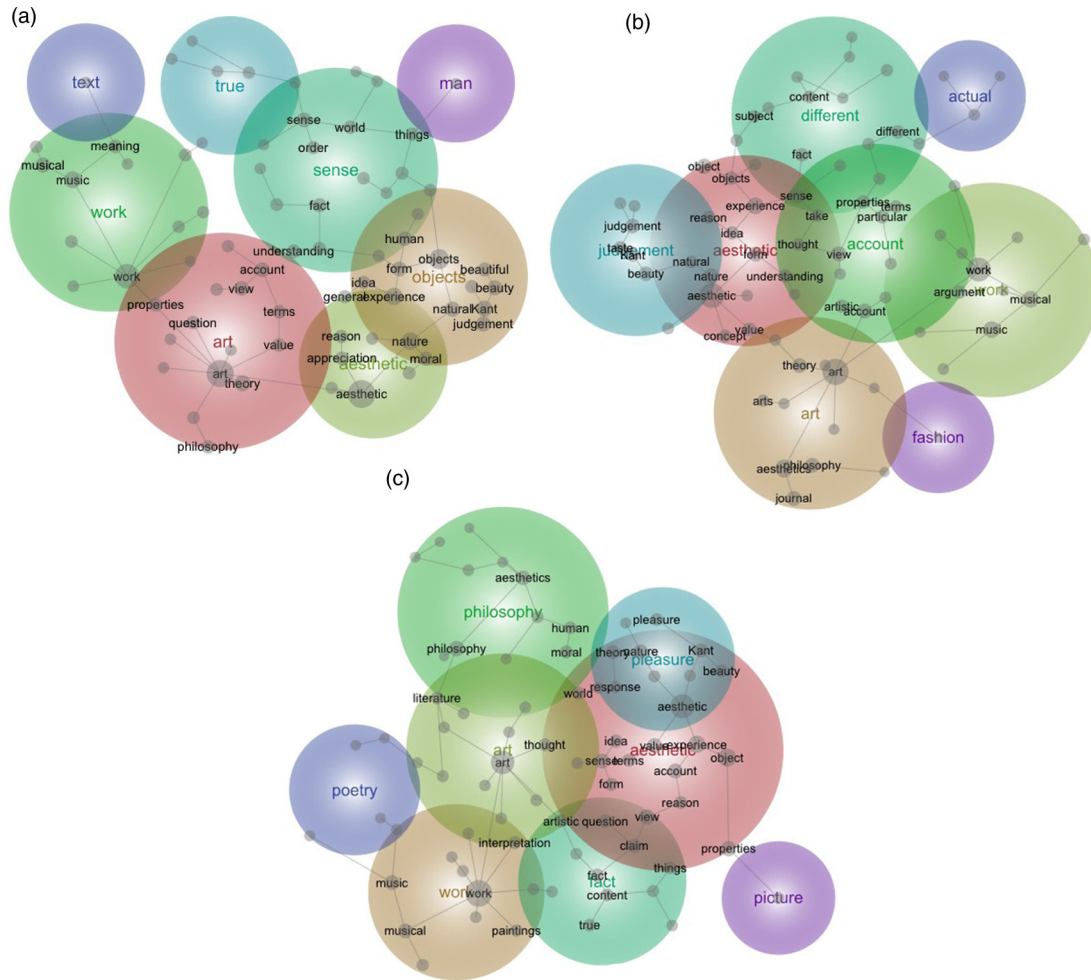Figure 8. Co-occurrence map of BJA title words.

Figure 9. (a) BJA full-text analysis from 1996 to 1999. The themes, in order of importance, are art, object, aesthetic, work, sense, trust, text, and man. (b) BJA full-text analysis from 2005 to 2009. The themes, in order of importance, are aesthetic, art, work, account, different, judgment, actual, and fashion. (c) BJA full-text analysis from 2010 to 2014. The themes, in order of importance, are aesthetic, art, philosophy, fact, pleasure, poetry, and picture.

so on, according to the color organization system that Leximancer deploys. Note that the figures portray only the most prominent node names for reasons of visibility.

It is easy to see that there are some themes, such as "art" and "work," that remain over decades, but others, such as "poetry," gain more interest in certain periods. On the other hand, themes such as "fashion" and "man" seem somewhat dubious and force one to dig deeper to see in what sense and way the concepts have been used. The tool enables the analyst to drill down to all text excerpts in which a certain word or word pair appears, to aid in the interpretation.

In principle, it would be fairly easy to make comparisons using Leximancer, or the other tools used here, between BJA and other journals, such as

the *Journal of Aesthetics* and *Art Criticism* (or any other digital data set). This would take some time, but the basic principles would not change.

PROBLEMS TO SOLVE

Analyzing aesthetics through WoS and BJA offers some useful insights, as we have seen, but there are limitations as well.

We already mentioned that many important sources are missing from WoS. Missing sources include journals, too, but the most evident lack is monographs, which are still very important in aesthetics, as well as in other fields of the humanities. This data does not tell us what the most referred books are, what themes those books address, and how they form groups. Most probably, such data sets will gradually be provided,

while more and more books are being digitized, but, for the time being, they are not common. Of course, even now, normal library databases have some information on books (titles, authors, publishers, short descriptions, key words), but that is far from a potential set of full-text databases offering cross-referential information. Thus, present-day possibilities offered by WoS and other similar databases for analyzing the field of aesthetics are seriously limited. For example, authors such as Arthur C. Danto and Yuriko Saito have important articles, but their books are probably at least as influential, which cannot be seen very easily through WoS. One possibility is to look for the cited reference information of the article data downloaded from WoS. However, that data is utterly messy, as the information is not uniformly entered into the database (meaning that the same book might have several instantiations with slightly differing indexing), and it is much more challenging and time-consuming to clean that data than the core bibliometric data of the main articles. Nevertheless, the analyst can gain preliminary insights even from the messy data, although reporting any strict statistics would be highly questionable.

WoS is also dominated by publications and authors writing in English. In the data set that we analyzed, more than 73% of the publications are written in English, 12% in French, and the other 15% in 24 other languages. BJA, naturally, is all in English. However, it is not reasonable to think that aesthetic issues would only be addressed in English, especially because many of them are highly dependent on culture and language. In the future, we need digital databases that better cover several languages. There are active communities of aesthetics using German, Polish, Slovenian, Finnish, Swedish, Japanese, Chinese, Spanish, Turkish, and several other languages. Finnish, for example, does not exist in the data set at all. How can we make different languages and cultures visible and comparable? At the moment, there are no good databases for that.

Moreover, some of the typical bibliometric analyses are clearly designed for the natural sciences, where many practices are somewhat different than in the humanities. For example, the tools offer co-authorship views, because it is typical in the sciences to publish in groups. In the humanities, in turn, it is still common to publish alone. As mentioned, in our data set, some 93% of the publications are single-authored, which is 10 times more than in many fields of the sciences and 2.5 times more than in the social sciences in general.[16] Thomson Reuters' ScienceWatch presents interesting field-specific statistics on single-authorship and how it has consistently decreased from 1981 to 2012: from 33% to 11%, considering all scientific articles indexed by WoS. The number of single-authored articles has, as such, remained rather stable, around 140,000 per year, during the 30 years, but the number of multi-authored papers has exploded at the same time, from 440,000 to 1.3 million.

For aesthetics, it might also be interesting to analyze pictures and sounds, but these text-mining tools cannot handle them; they are completely language-based. There are computational tools in domains other than text-mining that can be used to analyze pictures and sounds, but space does not allow us to present them here.[17]

Another issue related to the visual communication of text-mining results is that many tools that are available simply provide certain standard visualization options without too much explaining why they are of the kind that they are. Studies in information graphics, however, have again and again shown that there are no neutral ways of visualizing data and that different solutions in choosing colors, columns, links, lines, arrows and other visual means lead to completely different understandings of the questions addressed, and there are numerous alternatives that can be developed.[18] This is why visual options provided should be explicated in detail, which is not always the case. In the context of aesthetics, of course, also the aesthetic quality of visual presentations would be a theme worth explicating but in this article we simply wanted to give examples of the means available and not take a stand on their aesthetic worth. Figures 3 and 4, for example, would probably benefit from better graphic design, both aesthetically and otherwise. All in all, data visualization is a very potential option also for aesthetics but it must be developed much further from the level that has been exemplified in this article.

Yet another kind of problem is that WoS and other academic databases are not free, but only affiliated academic people have easy access to them. This is not an open and democratic situation. Moreover, there are license restrictions even for users with a user's license: systematic downloading

of bibliometric data or full texts is not allowed in large quantities. Some journals have also used a secure PDF format during some years, so that it is practically impossible to make full-text analyses of those materials. At best, one has to ask for special permission for that.

The most difficult nut to crack is to see what should be seen as data for making sense of aesthetics at large. Which sources should be included? WoS does not clearly cover everything, even if it is a very big data set, and neither do other databases. Moreover, even if aesthetic issues were dealt with in the sources analyzed, included in WoS or elsewhere, the word "aesthetics" is not always used. How can we find such cases, then? What are the best search terms and what do they actually bring up? The word "art," for example, can lead us to sociological and economic studies of the arts, as well as to essays on "art of war"—such sources potentially being irrelevant to aesthetics.

In addition, in the case of BJA, can we really trust that *everything* in it represents aesthetics? This, in particular, requires philosophical clarity: how do we interpret terms, concepts, and categories, as well as their limits, borders, and changes? It is far from self-evident which expressions can refer to the field of "aesthetic issues" and how—which should have been clear at least since Frank Sibley's classical analyses of aesthetic concepts. That is exactly why there is no automatic and simple way of using and analyzing databases, but search processes must be combinations of advanced computational methods and deep philosophical understanding of the field in question. Answers will eventually get better as we become more experienced.

In the end, we will end up discussing the ontology of aesthetics: how does it exist? As books and articles, for sure. But also in other ways? Does it have non-linguistic manifestations and how can we detect them? At least they do not exist in databases such as WoS, which leads us to say something about other possibilities related to computational approaches.

FURTHER POSSIBILITIES

Standard academic databases are limited in many ways, as we saw. Another option for making sense of academic aesthetics is to use online resources. Space does not allow us to explore this in more detail here, but the options available include Google Scholar and Google Books Ngram Viewer, as well as Wikipedia and its categories, which are gradually being formed by its users.[19]

In so-called altmetrics, all in all, the goal is to find alternative metrics for understanding academic activities.[20] Altmetrics is a subset of scientometrics, and it denotes "the study and use of scholarly impact measures based on activity in online tools and environments."[21] Although traditional scientometrics is heavily focused on citations for recording the impact of academic research, the outstanding rise of social media has exposed several new channels for tracking the impact.[22] Altmetrics is an interesting development currently taking its early steps, as it illuminates the impact of scholarly studies on the general public rather than just the academic community.[23] These metrics can be categorized in five general classes, listed in increasing order of importance: viewed, downloaded/saved, discussed, recommended, and cited.[24] Altmetrics utilizes, for example, microblogs, online reference managers such as Mendeley, blogs, social networking platforms, repositories like Github, domain-specific data from arXiv, access measures on publishers sites like PLoS, and user ratings on books, for example, from Goodreads.[25] Although alternative metrics currently present one of the most popular research topics in scientometrics,[26] it also has some problems, as listed by John Mingers and Loet Leydesdorff: "1) Altmetrics can be gamed by 'buying' likes or tweets; 2) there is little by way of theory about how and why altmetrics are generated (this is also true of traditional citations); 3) a high score may not mean that the paper is especially good, just on a controversial or fashionable topic; and 4) because social media is relatively new it will under-represent older papers."[27]

If we operate in altmetrics, we have to—again—ask which "hits" are actually cases of aesthetics, which are only somehow (loosely) related, and which are something else. We have to consider our search principles very carefully, when navigating the whole open internet. What kinds of terms will bring up relevant data? Are we looking for philosophical texts that are close to academic aesthetics but for some reasons excluded from the traditional academic publications, such as blog texts on Tom Leddy's *Aesthetics Today* (aesthetic-stoday.blogspot.com) and video presentations, or

perhaps tweets? Do we include pictures? Artists' activities? Or networks and groups of things rather than individual cases? Whatever we are looking for, we need to have suitable tools, and in the present situation the tools are more and more often computational. At the moment, there are no dominant, well-established tools in altmetrics, but a buzz of competing and developing ones. Still, aestheticians should follow what happens in that area. New tools appear all the time, and one can find several articles that review their features.[28]

It is possible that the computational digital world is changing our way of seeing what is a "work" or "piece" or "case" of aesthetics. We are not necessarily focusing on clear-cut cases, objects, events or authors, but relational networks or "clouds" of phenomena, even if this might not be so evident to us. The situation is probably more or less parallel with the one that David Joselit describes in his book *After Art*, which addresses the situation of contemporary visual arts and architecture. According to him, it is not that easy or even possible to see a clear difference between original artworks and all kinds of digital derivatives of and references to them; the internet and the dominant search procedures guide us to see the network they form together. In his words: "As I have argued, what now matters is not the production of new content but its *retrieval*, in intelligible patterns through acts of *reframing, capturing, reiterating*, and *documenting*. What counts, in other words, is how widely and easily images connect: not only to messages, but to other social currencies like capital, real estate, politics, and so on."[29]

Likewise in aesthetics, there might be more or less clear cases, related ones, derivatives, and so on; and what may count on many occasions is how they interact and form bigger, ever changing wholes. In such wholes, some nodes tend to attract more attention than others. However, on the internet, even such nodes are not single, clear-cut objects but relational networks within larger networks supporting them. It is not a single article or book that becomes visible alone, but everything that is attached to it in the digital network or cloud. By this logic, the article or author who attracts most connections (references) easily seems to be the most important. And in fact, often such articles and authors indeed are very important, because connections and relations are based on the fact that

readers or other users find them useful and want to tell others about them.[30]

One aspect of this situation is the importance of searchability, that is, how easily something can be searched and found in the digital net. Computational tools can only search and find objects and relations that are "visible" to them, which, again, is defined by the algorithms they are programmed to follow. Often, such tools do not find single, clear-cut cases, even if they were very interesting and important in some other ways. Very often, also, users of such tools do not really have to understand in detail how the tools function. We can use them without knowing *exactly* what they do and do not do. This, however, must make us extra careful when assessing what they actually find and show us, and why. Strong visibility in such searches does not necessarily mean that a scholar or a book is automatically better, more valuable or important than something that has a lower "searchability rate" (in this particular data set) and that has not yet been found. The value of scholars and publications is something we still have to evaluate by more complex, peer-review processes, too.

We could also leave the academic world behind and try to see what aesthetics is elsewhere. Then, we have an even more complex field to navigate. The simple test of googling "aesthetics" and comparing the image search with the text search shows that the former relates "aesthetics" to beautiful (white) women and body-builder men, the latter to philosophical definitions of the term "aesthetics," among literally millions of other things. How are these two interrelated? In any case, non-academic cases of aesthetics, on the internet and elsewhere, by far outnumber anything academic aestheticians can ever even imagine producing. Aesthetic values and issues are actively noticed and dealt with by various actors and in numerous ways, and academic, philosophical approaches are a tiny minority in the broad field. The top 10 actors in academic aesthetics found in WoS are unknown to the wider public. It is healthy to remember this. This theme, of course, would require a study of its own.

CONCLUSIONS

When one nowadays wishes to understand one's own discipline, aesthetics or otherwise, it is wise to make use of the latest computational tools

and combine them with the more traditional understanding of the field. There are several tools available, and the best, most comprehensive results will be achieved if one does not settle for one or two, but compares several points of view with each other.

All of these provide a slightly different picture of aesthetics. This, in itself, is an interesting result and worth presenting to students and readers of introductory books, for example. And it becomes even more interesting when one tries to argue which of them are more accurate, which less. Why am I for some of them? If the field is this big, why do I tend to focus on some of its parts?

The full picture can never be achieved, but making use of computational tools is one current route that we simply must follow,[31] even if there are many problems to solve. They will not substitute philosophical analyses, but will complement them and actually make them even more necessary. Computational approaches also force us to consider what is nowadays the ontological status of the field. Where and how does it exist? A short answer is, we think, that aesthetics is a social information network that is constantly growing and changing. What this means, in more detail, must be answered in another article.[32]

Notes

1. Peder Olesen Larsen and Markus von Ins, "The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index," *Scientometrics* 84 (2010): 573–603.

2. Lutz Bornmann and Rüdiger Mutz, "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References," *Journal of the Association for Information Science and Technology* 66 (2015): 2215–22.

3. However, aesthetics and computational approaches have been combined in other ways. For example, computational methods have been used for analyzing and even creating art works and other aesthetically interesting objects, and aesthetic features of such computational procedures have also been studied. In both areas, the MIT Press and the journal *Leonardo* have been active for a long time. See, for example, Paul A. Fishwick, ed., *Aesthetic Computing* (Cambridge, MA: The MIT Press, 2008). A recent, more specific example is an article on using computer vision to find beauty in low-attention photos stored on Flickr: Rossano Schifanella, Miriam Redi, and Luca Aiello, "An Image is Worth More than a Thousand Favorites: Surfacing the Hidden Beauty of Flickr Pictures" (Proceedings of the Ninth International AAAI Con-

4. ference on Web and Social Media ICWSM, Oxford, UK, May 26–29, 2015).

5. A many-sided description of the fast-growing field of the digital humanities is Debates in the Digital Humanities, both as a book edited by Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), a book series starting in 2016, and as an open-access online platform at http://dhdebates.gc.cuny.edu/about (accessed September 29, 2015). For a bibliometric review, see Loet Leydesdorff and Alkim Salah, "Maps on the Basis of the Arts & Humanities Citation Index: The Journals *Leonardo* and *Art Journal* versus 'digital humanities' as a topic," *Journal of the American Society for Information Science and Technology*, 61 (2010): 787–801.

5. Alan Porter and Scott Cunningham, *Tech Mining. Exploiting New Technologies for Competitive Advantage* (Hoboken, NJ: Wiley, 2005), 357.

6. Eugene Garfield, "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas," *Science* 122 (1955): 108–11.

7. Loet Leydesdorff, Björn Hammarfelt, and Alkim Salah, "The Structure of the Arts & Humanities Citation Index: A Mapping on the Basis of Aggregated Citations among 1,157 Journals," *Journal of the American Society for Information Science and Technology* 62 (2011): 2414–26; and Björn Hammarfelt, "Using Altmetrics for Assessing Research Impact in the Humanities," *Scientometrics* 101 (2014): 1419–30.

8. More information on the software is provided on their websites: VantagePoint http://www.thevantagepoint.com, VOSviewer http://www.vosviewer.com and Leximancer http://www.leximancer.com. See also Porter and Cunningham, *Tech Mining*; Nees Jan Van Eck and Ludo Waltman, "Visualizing Bibliometric Networks," in *Measuring Scholarly Impact: Methods and Practice*, eds. Ying Ding, Ronald Rousseau, and Dietmar Wolfram (Cham: Springer, 2014), 285–320; and David Thomas, "Searching for Significance in Unstructured Data: Text Mining with Leximancer," *European Educational Research Journal* 13 (2014): 235–56.

9. The categorization of research areas into scientific domains is done based on information provided by WoS at https://images.webofknowledge.com/WOKRS511B5/help/WOS/hp_research_areas_easca.html (accessed June 23, 2015).

10. The list includes only one female author, Suzanne Langer. This gender imbalance and its reasons would deserve a study of its own.

11. A Multi-Dimensional Scaling algorithm proprietary to VantagePoint determines the location of each author on the map. The x- and y-axes of the maps have no specific meaning. The algorithm simply tries to reduce an N-dimensional representation to two dimensions, seeking to maintain authors with a high degree of similarity (correlation) in close proximity to each other. Generally speaking, authors who are

close to each other are more similar than those that are farther apart. However, the presence or absence of a line (and the thickness of the line) between any two authors is a more appropriate measure of proximity, since it implies a relatively high correlation between them.

12. Note that this term map and subsequent VOSviewer maps are constructed with the larger search data set of 21,919 texts, including articles and all other publication types from WoS A&HCI. This is because the refining of publication types took place only in VantagePoint after the raw data was downloaded from WoS. With VOSviewer, it is not possible to refine the raw data like that. A threshold of 25 was used when constructing the map for Figure 5. This means that terms that appear in at least 25 titles or abstracts are included in the co-occurrence map, and only some of the approx. 950 qualified terms are visible, to avoid clutter.

13. Eck and Waltman, "Visualizing Bibliometric Networks."

14. On NLP, see Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, MA: The MIT Press, 1999).

15. Thomas, "Searching for Significance in Unstructured Data."

16. Christopher King, "Single-Author Papers: A Waning Share of Output, But Still Providing the Tools for Progress", ScienceWatch, Thomson Reuters, September 2013 http://sciencewatch.com/articles/single-author-papers-waning-share-output-still-providing-tools-progress (accessed August 17, 2015).

17. See, for example, Schifanella, Redi, and Aiello, "An Image is Worth More than a Thousand Favorites." See also the software Culture Cam at culturecam.eu.

18. For example, Katy Börner, *Atlas of Science—Visualizing What We Know* (Cambridge, MA: The MIT Press, 2010); Sandra Rendgen et al., *Information Graphics* (Cologne: Taschen, 2012); Edward R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. (Cheshire, CT: Graphics Press, 2001).

19. Anne-Will Harzing, "A Preliminary Test of Google Scholar as a Source for Citation Data: A Longitudinal Study of Nobel Prize Winners," *Scientometrics* 94 (2013): 1057–75; and Anne-Wil Harzing, *Publish or Perish, 2007*, software available at http://www.harzing.com/pop.htm (accessed September 29, 2015).

20. Lutz Bornmann, "Alternative Metrics in Scientometrics: A Meta-Analysis of Research into Three Altmetrics," *Scientometrics* 103 (2015): 1123–44.

21. Jason Priem, "Altmetrics," in *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. eds. Blaise Cronin and Cassidy R. Sugimoto (London: MIT Press, 2014), 263–88, specifically on 266.

22. John Mingers and Loet Leydesdorff, "A Review of Theory and Practice in Scientometrics," *European Journal of Operational Research* 246 (2015): 1–19.

23. Ibid.

24. Jennifer Lin and Martin Fenner, "Altmetrics in Evolution: Defining and Redefining the Ontology of Article-Level Metrics," *Information Standards Quarterly* 25 (2013): 21–26.

25. Bornmann, "Alternative Metrics in Scientometrics."

26. Ibid.

27. Mingers and Leydesdorff, "A Review of Theory and Practice in Scientometrics," 15.

28. Manuel Jesus Cobo, et al., "Science Mapping Software Tools: Review, Analysis and Cooperative Study among Tools," *Journal of the American Society for Information Science and Technology* 62 (2011): 1382–402; YunYun Yang, et al., "Text Mining and Visualization Tools–Impressions of Emerging Capabilities," *World Patent Information* 30 (2008): 280–93; Stefanie Haustein et al., "Coverage and Adoption of Altmetrics Sources in the Bibliometric Community", *Scientometrics* 101 (2014): 1145–63; and Katrin Weller, "Social Media and Altmetrics: An Overview of Current Alternative Approaches to Measuring Scholarly Impact", in *Incentives and Performance: Governance of Knowledge-Intensive Organizations*, eds. Isabell M. Welpe et al. (Cham: Springer, 2015), 261–76.

29. David Joselit, *After Art* (Princeton, NJ: Princeton University Press, 2013), 55–6.

30. More generally speaking, Joselit's analysis is related to various network points of view that have been developed in philosophy and social sciences over the last decades by theorists as various as Michel Foucault, Gilles Deleuze, Bruno Latour, Luc Boltanski, Manuel Castells, Duncan Watts, and Mark Granovetter, among many others.

31. Ivan Zupic and Tomaz Cater, "Bibliometric Methods in Management and Organization," *Organizational Research Methods* 18 (2015): 429–72.

32. Casey Haskins offers one interpretation of this network in his article "Aesthetics as an Intellectual Network," *The Journal of Aesthetics and Art Criticism* 69 (2011): 297–308.