

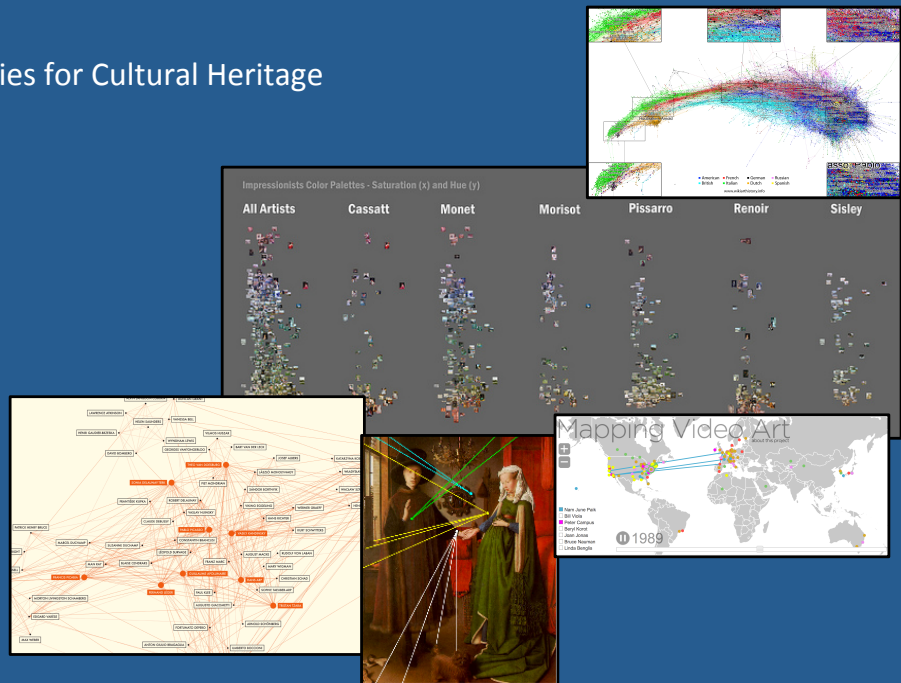
The “Art” of Digital Art History

Diane M. Zorich

Consultant – Digital Strategies for Cultural Heritage

dzorich@mindspring.com

@dzorich



The Digital World Of Art History

Princeton University

June 26, 2013

Let me begin by thanking Colum Hourihane of the Index of Christian Art (<http://ica.princeton.edu/>) and Trudy Jacoby of the Visual Resources Collection in the Department of Art and Archaeology (<http://www.princeton.edu/visualresources/>) at Princeton for organizing this conference on the *Digital World of Art History* and for inviting me to take part.

Slide image credits (clockwise from top right):

Art History on Wikipedia, a Macroscopic Observation. *Wiki Art History*. May 19, 2012.
<http://vsem.ec.tuwien.ac.at/wikiarthistory/index.php?entry=entry120519-174955>

Manovich, Lev. Visualizations of Impressionist artists - color palette comparisons (part 1). *Software Studies*. April 14, 2012
<http://lab.softwarestudies.com/2012/04/visualizations-of-impressionist-artists.html>

Mapping Video Art. Visualization by Shilpan Bhagat and Jasone Carone. January 13, 2013.
<http://www.visualizing.org/visualizations/mapping-video-art>.

Analysing the consistency of vanishing points in a painting: The Arnolfini Portrait (by Jan van Eyck (circa 1395-1441). In “Portrait Bringing Pictorial Space to Life: Computer Techniques for the Analysis of Paintings.” 2002. Antonio Criminisi, Martin Kemp and Andrew Zisserman. <http://research.microsoft.com/apps/pubs/default.aspx?id=67260> Figure 3, p. 6.

Network diagram of artists relationships from “Inventing Abstraction: 1910-1925.”
<http://www.moma.org/interactives/exhibitions/2012/inventingabstraction/?page=connections>.



I am not an art historian, so I am guessing the reason I was invited to speak here today is because of a report I wrote for the Kress Foundation on community perceptions of digital scholarship and digital art history. I've spoken and written about this report in many other venues so I won't go into its findings here (the report can be downloaded from Kress Foundation website at <http://www.kressfoundation.org/research/Default.aspx?id=35379>)

What I want to do today is to use this report as a springboard for a larger issue: refining the notion or the concept of "digital art history."

When I was conducting the interviews that form the basis of the Kress report, I used the phrase "digital art history" to represent art historical research, scholarship and/or teaching using new media technologies. I intentionally used such a broad definition for two reasons:

1. In 2011, when I began carrying out the research for this report, digital art history was really just gaining awareness as concept across the community, and a firm and more concise definition hadn't yet gelled.
2. I was interviewing professionals from all areas of the art history community – scholars, librarians, conservators, research center administrators, funders, information technology professionals, etc. Each professional had his/her unique slant or view on digital art history. I wanted to be inclusive in my interviews, and a broad definition allowed me to do this.

But in the year that the report has been out, I have gotten some interesting pushback on the findings of the report that makes me think the time has come to refine our concept of digital art history.

This pushback takes two forms:



On the surface, I agree with both these comments. It is true that art historians are using digital technologies in their research. And most art history research centers have digitized portions of their resources and make these resources freely available to scholars. But I disagree with the assumption underlying these comments – i.e., that digital art history is “art history in a digital environment”. I think there is more to it than that. I think it is time to reconsider what we mean by digital art history and move toward a more sophisticated understanding of it. Only then can we have more rigorous discussions about it.

Johanna Drucker, an artist and Professor of Information Studies at the University of California, Los Angeles, gets to the heart of this in her recent article in the journal *Visual Resources* (Volume 29, Numbers 1–2, March–June 2013, pp. 5-13) entitled, “*Is There a Digital Art History?*” She calls current efforts in art history “digitized art history”, not digital art history (Italics mine). The former represents the extensive digital access and delivery of images of works of art, but it is not digital art history because these changes have “not had a ripple effect on the intellectual foundations of art history.”

Several people I interviewed for the Kress report agree with Drucker’s assessment. They acknowledge the value of technology in identifying and delivering resources, but felt these capabilities simply addressed the mechanics of research while doing nothing to transform the nature of it.

I think the reason for these sentiments is that art history has been slow at adopting the **computational methodologies and analytical techniques** that are enabled by new technologies. And until it does so, art historians will never really be practicing digital art history in the more meaningful sense that Drucker implies. They will only be moving their current practices to a digital platform, not using the methodologies unique to this platform to expand art history in a transformational way.

Today I want to talk about some of these new computational methodologies to show how they can lead to new forms of exploration, analysis and scholarship that would truly be transformative for the discipline. In the interest of time I am only going to talk about three of these methodologies – visualization, network analysis, and topic modeling – but I chose these three because I strongly suspect they have some potential for the discipline.

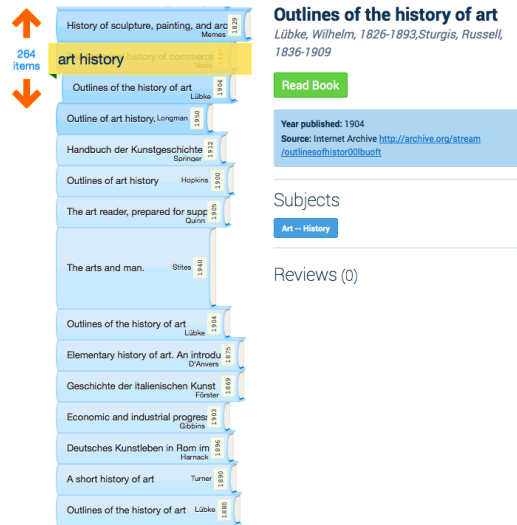
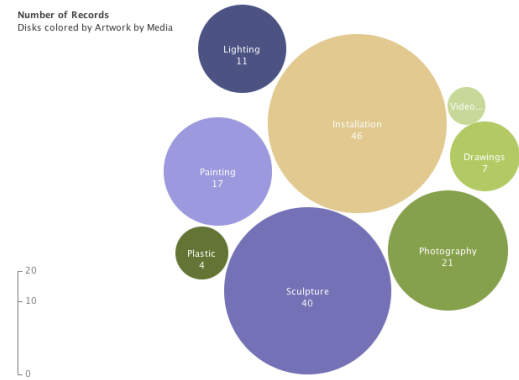
Before I go further, I want to make it clear at the outset that I don't mean to imply in any way that these methodologies should replace current art history methodologies: they are additions or extensions to current methodologies in the discipline. Art historians and other scholars who have used these new computational methodologies emphasize how they often they use them *with* traditional methods in a complementary fashion.

Visualization

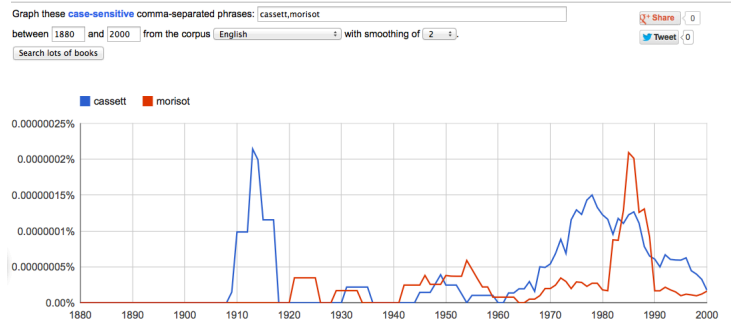
Visualization is a technique that uses computational methods to organize, display and interpret large amounts of information visually in order to reveal patterns that cannot be deduced from a close reading.



Number of Records
Disks colored by Artwork by Media

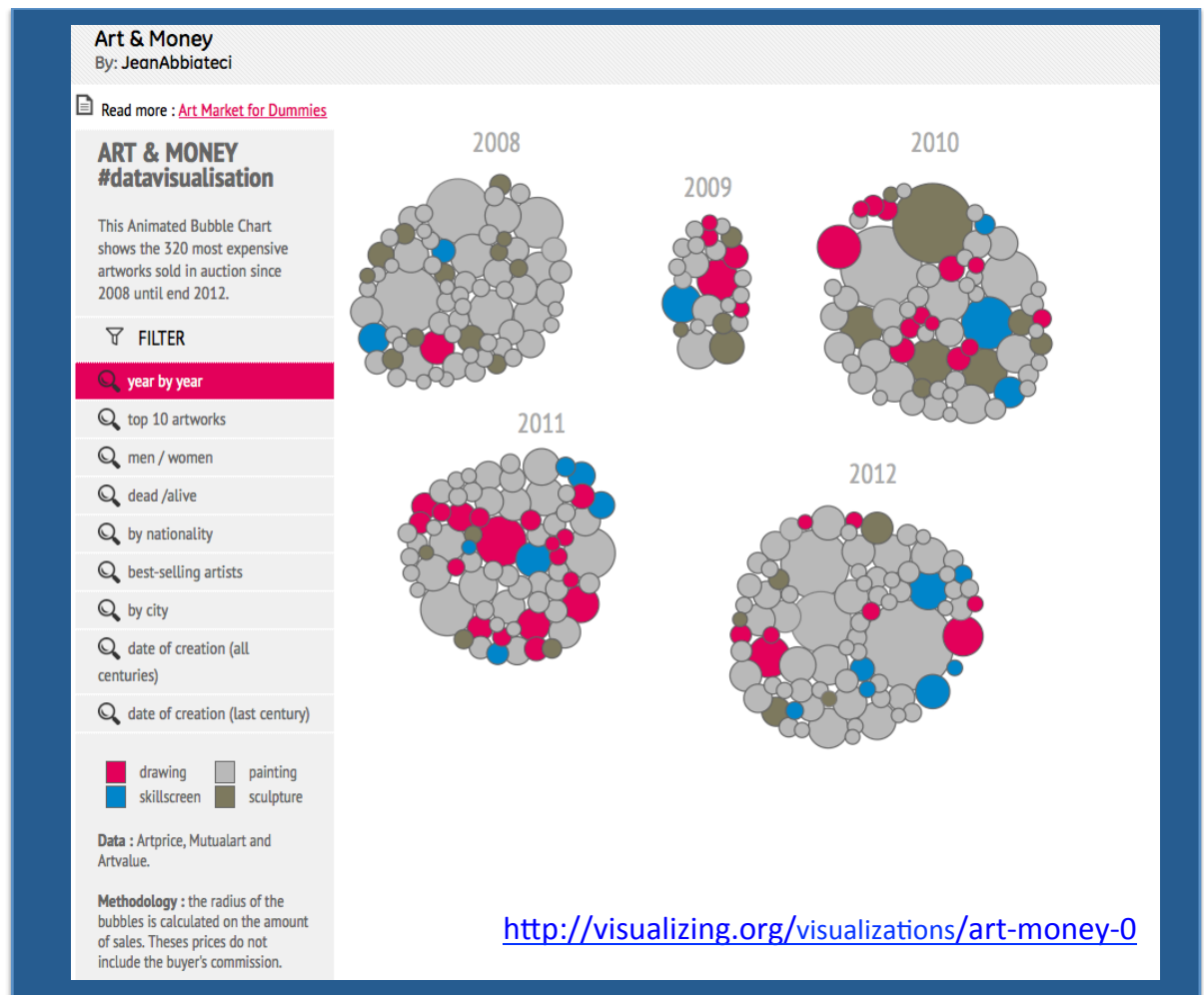


Google books Ngram Viewer



Visualizations take many forms and can use any type of graphic element to convey information. They can be simple line graphs such as the one that Google's N-Gram Viewer (seen on your lower right) generates, or they can build on metaphors drawn from our physical world, such as the one on your lower left ("Stacklife" - <http://stacklife.harvard.edu/> - created by the Library Innovation Lab at Harvard University for use in visualizing the books available through the Digital Public Library of America.)

The visualizations on this slide are really examples of what I would call low-hanging fruit: three of them I created on the fly, using freely available software that required absolutely no programming on my part. All I did was feed in some information.



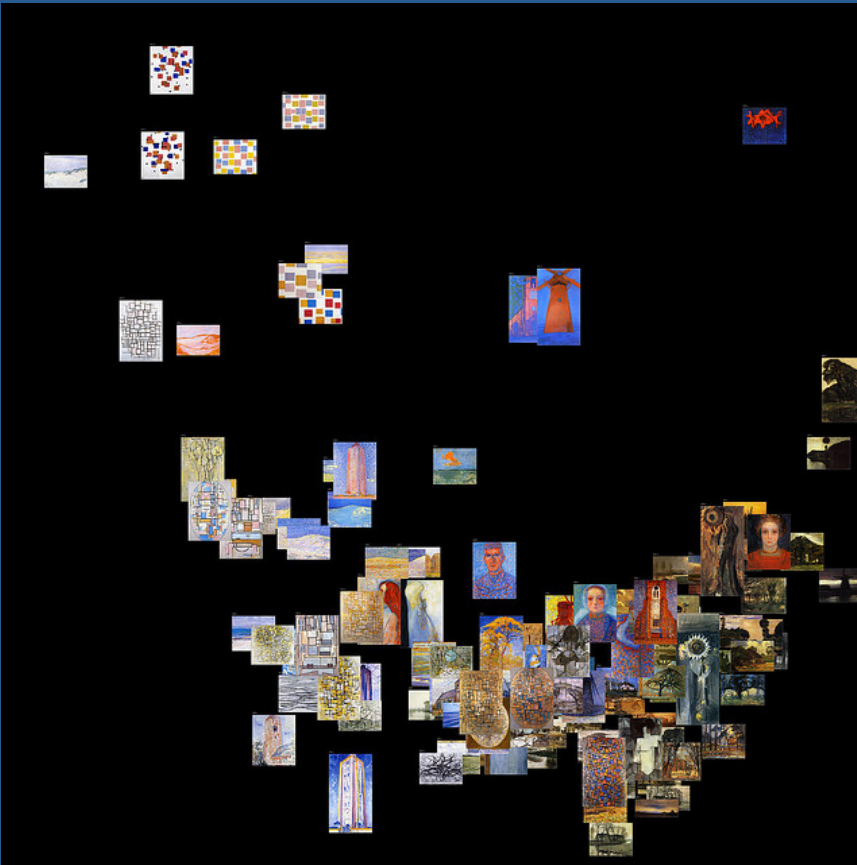
But data visualization is more than just a tool that gives us a nice way to look at something. It allows us to visually comprehend information in ways that facilitate interpretation and prompt new lines of inquiry.

This particular visualization entitled “Art and Money” (<http://visualizing.org/visualizations/art-money-0>) by Jean Abbateci gives you a better sense of this. This visualization shows 320 of the most expensive works of art sold at auction in the years 2008 – 2012. By interacting with this visualization (which is animated and filters the information in various ways – click on the link to interact with the visualization), you can reorganize the information in different ways to reveal patterns that can lead to further inquiry.

For example, if you click on the filter “date of creation, by all centuries”, you see that 20th century works by far form the bulk of art sales in this sample. Why is this? Are 20th century works seen as a better investment by those who buy high-end art? Perhaps the buyers’ market favors 20th century art for aesthetic reasons? Or perhaps there just weren’t that many 14th-19th century art works available for sale during the 2008-2012 period that this visualization covers?

By exploring the information in this visualization, you may see patterns such as this that prompt you to ask new questions or look deeper to look for underlying reasons behind the patterns. You “see” aspect of the data in ways that wouldn’t be obvious from a “line-by-line” analysis.

Cultural Analytics



Mondrian Paintings
N=128
(1905-1917)

Manovich, Lev. Mondrian vs Rothko: footprints and evolution in style space. Software Studies, June 29, 2011 <http://lab.softwarestudies.com/2011/06/mondrian-vs-rothko-footprints-and.html>

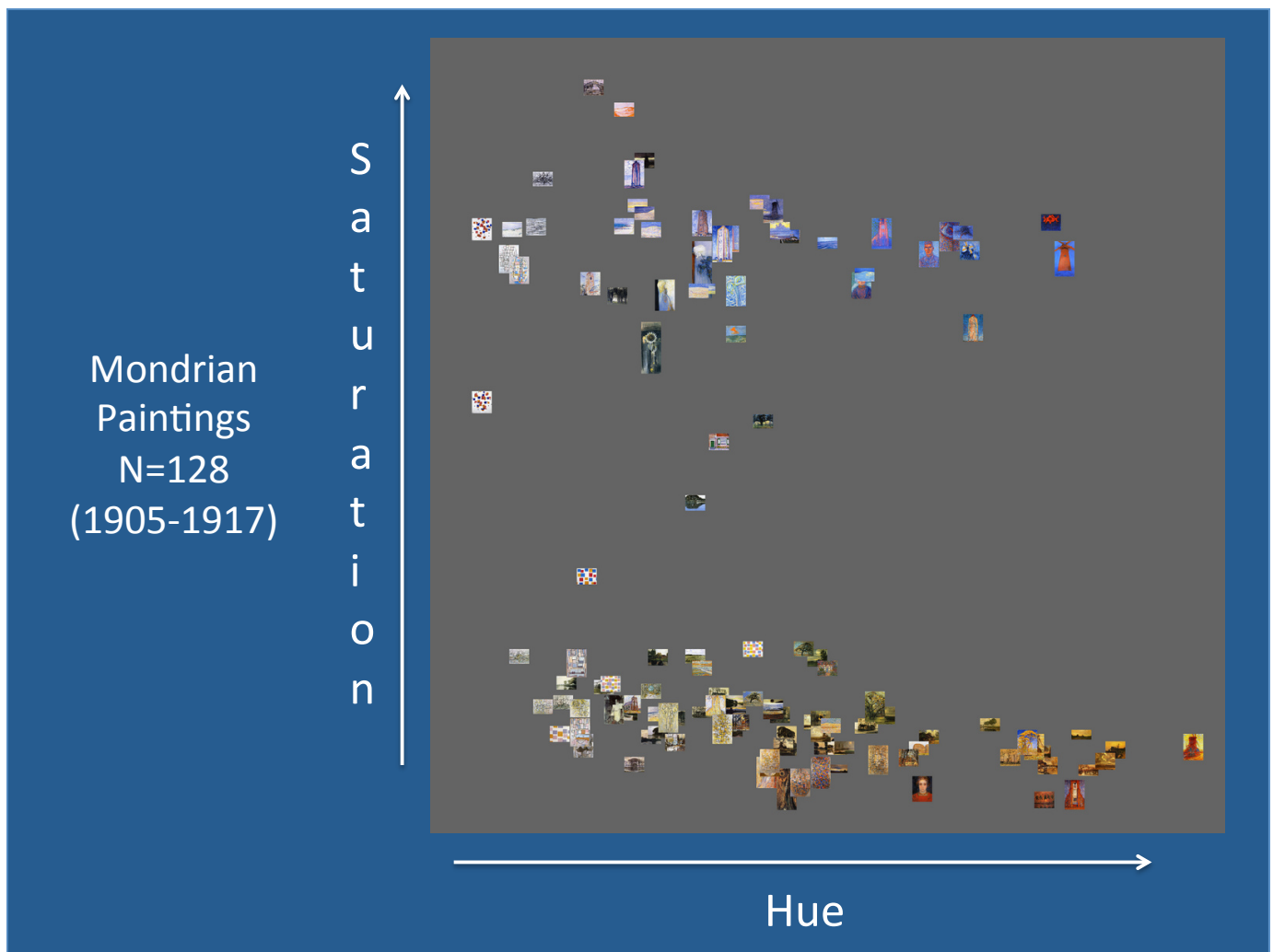
Another visualization technique getting a lot of attention is one put forth by Lev Manovich, a computer scientist at the City University of New York (CUNY). Manovich calls his work “*cultural analytics*”, (see <http://lab.softwarestudies.com/2008/09/cultural-analytics.html>) and it is receiving attention because he is applying it to image sets – particularly images of works of art.

Manovich and his students use a standard statistical technique called Principal Component Analysis (PCA) to analyze 60 different visual features of a work (called “image features”, comprised of things such as color, texture, lines, shapes, etc.). In the example illustrated here, Manovich shows the results of this analysis when applied to 128 Mondrian paintings created between 1905-1917.

The result is displayed as a scatter plot, but instead of the dots you see in typical scatter plots the points are the actual paintings. The paintings are organized or plotted, if you will, by visual similarity, based on 60 different visual features.

Manovich calls visualizations such as this one “footprints” of the artist’s work. He talks about how this approach lets you see in a new way – “the parts of the space of visual possibilities (that the artist) explored,” the relative distributions of their works – the denser areas, the sparser areas, the presence or absence of clusters, etc. You can see here how Mondrian’s more geometric works cluster together,

how his more representational works cluster together, where the outliers are, etc. (See <http://lab.softwarestudies.com/2011/06/mondrian-vs-rothko-footprints-and.html>)



Another interesting aspect of this type of analysis is that you can also map images of the works of art by any combination or subset of the 60 image features. In this example, Manovich and his team mapped the same 128 Mondrian works by saturation (i.e., the intensity of the color) and hue (the colors as defined on the color spectrum).

Although choosing these parameters may seem odd, and I am not sure Manovich had any particular reason for doing so beyond experimentation, the result is interesting: almost all of these 128 Mondrian works fall into two groups: those dominated by yellow and orange (bottom) and those dominated by blue and violet (top). Manovich undertook a similar analysis with Rothko's works (which I have not shown here but you can see at <http://lab.softwarestudies.com/2011/06/mondrian-vs-rothko-footprints-and.html>) and you see an entirely different patterning.

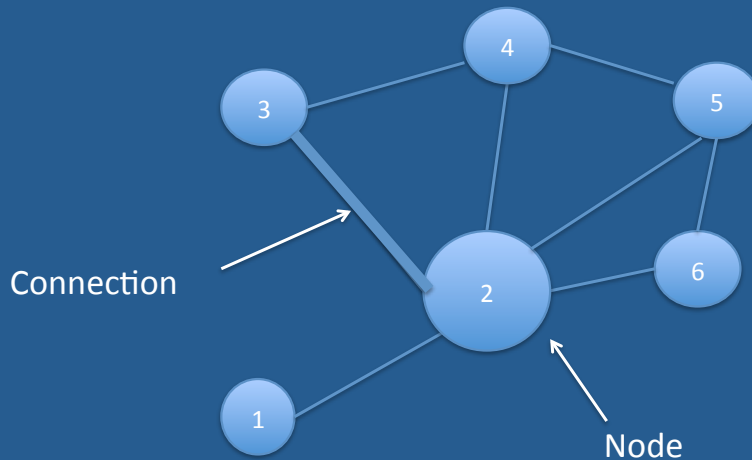
What does this all mean? Manovich offers his thoughts, which have been received by art historians with various levels of enthusiasm. But for me, the bottom line is that unless you have an amazing visual memory, it is hard to see patterns of this nature across the oeuvre of an artist.

I think Manovich's explorations of visual patterning in works of art do offer possibilities for useful art historical insights, but this type of analysis needs more direct input from art historians. They need to bring their research questions, the rigor of their intellectual and analytical frameworks, and their knowledge, to see what can be learned from this type of analysis.

Network Analysis

The second computational method I want to talk about is *Network Analysis*. This is a methodology for analyzing relationships between entities – i.e., objects, people, and places – almost anything you are interested in. The underlying assumption of this method is that there are relationships between entities, and these relationships matter: they give meaning, context, and greater understanding to a data set than could be otherwise be seen from a close reading.

Network Analysis Model



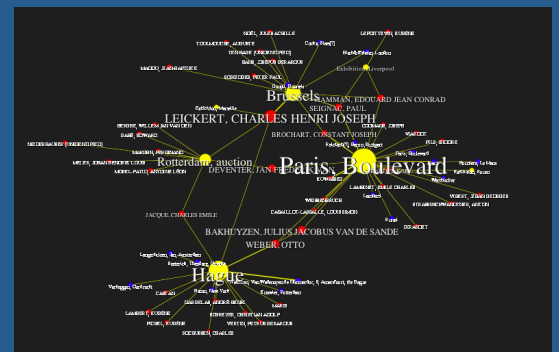
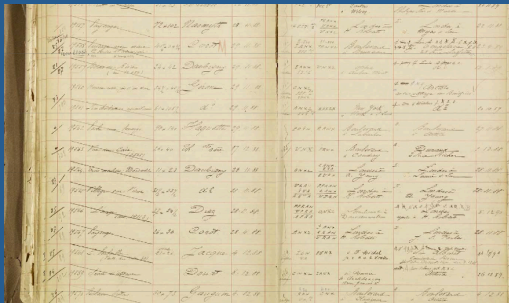
In network analysis, relationships are usually displayed in a model such as this – a circle and stem diagram. The circles are called “nodes”, and these represent the entities you are looking at – the people, places, things, etc. The lines between the nodes are called “connections” and they elucidate the relationships between entities.

The more connections a node has, the larger it is portrayed in the diagram. So in the diagram shown here, Node #2 is larger than the other nodes because it has connections with five other nodes, while Node #1 is smaller than all the others because it has only one relationship or connection.

In addition, the strength of a relationship (as determined by things such as frequency of contacts, or transactions between entities) is indicated by the thickness of the lines. So, for example, in life we have good friends and we have acquaintances. Our connection with our good friends would be indicated by a thick line (such as shown here between Nodes #2 and #3), while our connection with an acquaintance would be indicated with a thinner line.

Local/Global: Mapping Nineteenth-Century London's Art Market

Pamela Fletcher and Anne Helmreich
Nineteenth Century Art Worldwide



Art historians Pamela Fletcher and Anne Helmreich used network analysis methodology in their study of art markets of the mid- to late 19th century. They took the sales transaction information (data) from the stock books of the Goupil publishing and art firm - shown on your upper left- and analyzed it using network analysis software tools. The results are shown in the illustration on the lower right in a network analysis model, i.e., the circle and stem diagrams I talked about just a moment ago.

Helmreich and Fletcher's analysis is very rich and extensive, and I can't begin to do it justice here. But briefly, they interpret the art market as a series of relationships between artist, buyer, dealer and location – these are the entities that form the nodes (circles). The connections or relationships in this market – i.e., the lines between the nodes – are transactions conducted between these entities, i.e., the act of consigning, exchanging, selling, or buying a work of art.

Fletcher and Helmreich results are published online (see <http://www.19thc-artworldwide.org/index.php/autumn12/fletcher-helmreich-mapping-the-london-art-market>) - in fact, they can only be viewed online because the network analysis diagrams are interactive: you can zoom in or drill down through the network of relationships to look at things in more detail and study or investigate individual pathways of the relationships. You can do a close reading, or a distant reading.

Why did Helmreich (who did the network analysis portion of this research) explore network analysis and decide to apply it in this instance? She explains that she needed a way to see the whole “distant

view”, to “shift from the individual, telling transaction” and analyze the behavior and history of what was clearly a complex market.

The results revealed many interesting and surprising changes in the growth of the Goupil firm and the market, and influences in both that she could not have inferred from reviewing the stock books alone. It also put forth new questions that she hopes to continue to investigate.

Topic Modeling

The third methodology I want to talk about is *Topic Modeling*. This is a text-mining technique that uses statistical methods to look at words in huge text corpora - corpora that are too extensive for any human being to read through - to identify and uncover themes or categories (called “topics”) in the corpora.

Historians have applied topic modeling to things such as historical newspapers, which are rich but underutilized resources (their volume and coverage make it impossible for any individual to comb through them and accurately gauge the content.) Topic modeling also been used in certain disciplines to explore the historiography of these disciplines.

Important side note: One of the key figures in topic modeling is here in the Princeton University Computer Science department - Professor David Blei (see <http://www.cs.princeton.edu/~blei/>). Those of you interested in this methodology should seek him out. He has written and spoken a lot about this method to many different audiences.

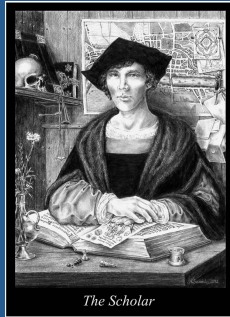
3 Components



Evan Bench, Flickr, CC-BY

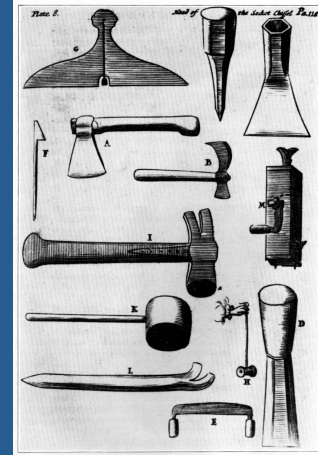
Corpus

Someone familiar
with the Corpus



The Scholar
Katherine Bramley, CC BY-
NC-ND 3.0 US

Tool
(Topic Modeling Software)



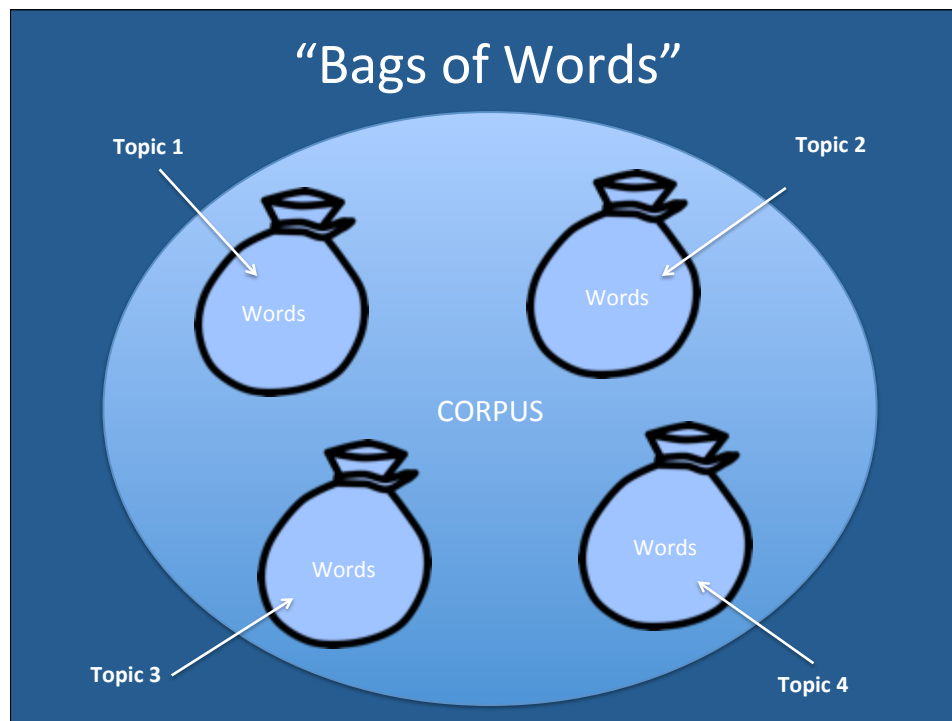
Topic modeling requires three components:

- A large corpus (e.g., a print run of a newspaper; all the articles ever published in a journal, or in all the journals of a particular discipline, etc.)
- A topic modeling tool that identifies the topics and give you a way to visualize your results (e.g. MALLET)
- Someone who is familiar with the corpus. For example, if you are using topic modeling on journals in a particular discipline, you probably want to bring on board a scholar in that discipline who is familiar with these journal and their general content.

Slide Image credits:

(Pile of Books) Evan Bench - Flickr <http://www.flickr.com/photos/austinevan/1225274637/> cc-BY

(Scholar) Katherine Bramley- <http://crimsongriffin28.deviantart.com/art/The-Scholar-327721392> CC BY-NC-ND 3.0 US



Topic modeling is often referred to as a “bags of words” approach to text mining. This slide shows a really basic schematic diagram of its key components.

The topic modeling software extracts from the corpus various “bags of words”. But these words are not just a random assortment; rather they are words that are likely to appear together with some frequency and within some proximity to one another. These groupings of words are called “topics”.

Now the software **doesn’t** identify the topic that the words represent. This is where the scholar comes in: you need someone who has some familiarity with the corpus to infer the topic that the words represent.

Example “Mining The Dispatch”

WORDS

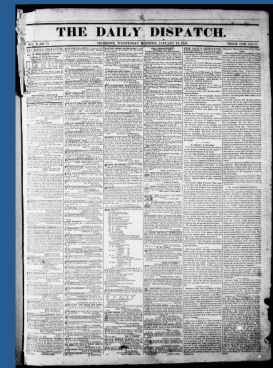
Negro, years, reward,
boy, man, named, jail,
delivery, give, left,
black, paid, pay, ran,
color, Richmond,
subscriber, high,
apprehension, age,
ranaway, free, feet,
delivered

TOPIC

Fugitive slave
advertisements

CORPUS

Richmond Daily
Dispatch



Although I would have liked to show you an example of topic modeling applied to art historical materials, I have not been able to identify anyone in the discipline who has used this approach. So my example comes from the discipline of history.

Dr. Robert K. Nelson, Director of the Digital Scholarship Lab at the University of Richmond, in a project entitled, “*Mining the Dispatch*,” (<http://dsl.richmond.edu/dispatch/>) examines the print run of a newspaper – the *Richmond Daily Dispatch* – which ran six days a week from 1850-1903 under various names. Nelson did topic modeling on only a subsection of this print run: a five year period starting from the eve of Lincoln’s election in 1860 through December 1865. This subset alone encompassed more than 112,000 articles consisting of approximately 24 million words!

Using a popular topic modeling software program (MALLET), a number of topics were generated from this corpus. On this slide, in the circle on your left, is an example of one of those topics. Consider this one of the “bags of words” words identified as a “topic” by the program.

Now here is where the researcher’s familiarity with the corpus comes into play. When I look at these words, I am not sure what topic is being reflected beyond some inference to slavery. But Nelson looked at these words and immediately inferred that these were words commonly used in fugitive slave advertisements.

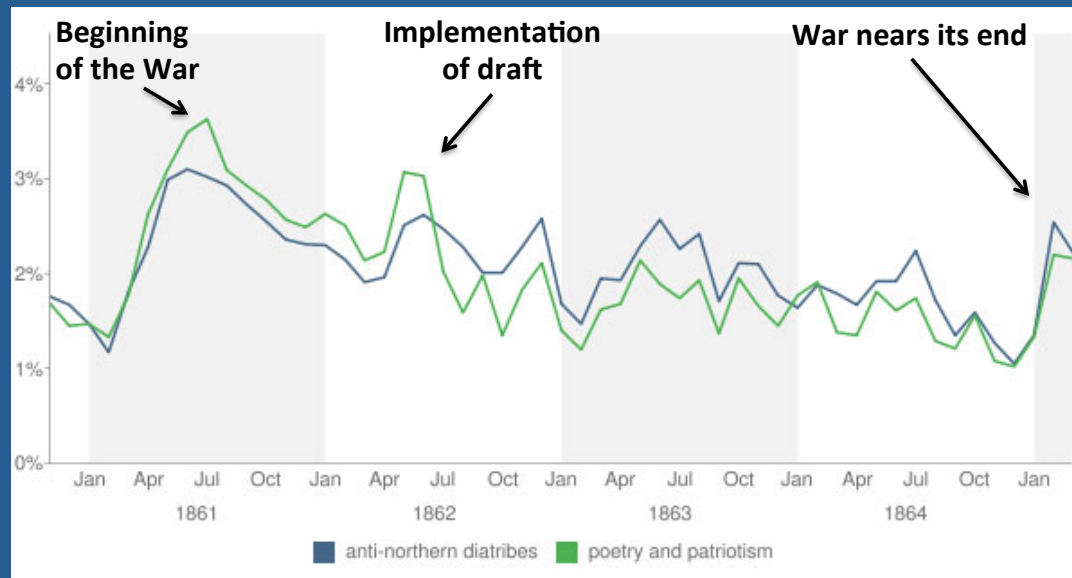
Now when you think about it, I suppose you could guess that a southern newspaper, publishing during the time of the Civil War, would have fugitive slave advertisements in it. This would seem to be one of the more obvious topics. But if you were trying to guess all the types of topics that such a newspaper might cover, you couldn't possibly guess them all. You would certainly miss the more subtle topics.

Moving along with this analysis, when you get your topics, what's next? Once again, the scholar brings in his/her special knowledge and skill set to do a closer reading or analysis.

Let me give you another example from Nelson's work.

Two Other Topics

“Anti-Northern Diatribes”
“Poetry and Patriotism”



Nelson, Richard K. “Of Monsters and Men – And Topic Modeling.” *The New York Times*, Opinionator, May 29, 2011
<http://opinionator.blogs.nytimes.com/2011/05/29/of-monsters-men-and-topic-modeling/>

Two particular topics generated by the topic modeling software that elicited Nelson’s further interest were “Anti-Northern Diatribes” and “Poetry and Patriotism.”

Nelson wanted to see when these topics appeared over the course the newspaper’s Civil War print run so he mapped their occurrence and found something really interesting – essentially, these two topics often appear together in parallel, i.e., if one topic appeared in the paper at any given moment, the other was likely to appear too. Nelson notes that association of these two topics would be difficult if not impossible to see without topic modeling because the corpus is just too large for an individual to “see” this.

In interpreting this overlap, Nelson felt that there was a logic to the correlation between these two topics or themes. Briefly, he explains it as a phenomenon you often see in military conflict. To convince people to fight in a war, groups often demonize the other side (i.e., the “anti-Northern diatribes”). But they also must extol the importance of their own cause, and glorify the valor of their own men who may die at the hands of the enemy (i.e., “the poetry and patriotism”).

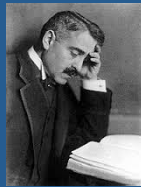
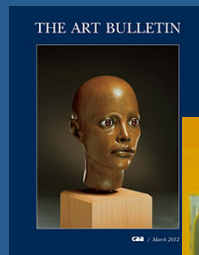
Nelson wondered whether the highpoints seen on the chart mapped to any significant events taking place at the time. And sure enough, when you do this mapping you see a rise in the prevalence of these two topics that corresponds with key moments in the war when you might expect these two themes to become prominent.

I have given you a quick and rather facile summary of what is a very detailed analysis, but I do think Nelson's use of topic modeling is fascinating and offers an intriguing look into the types of insights that can be gained. He uses this methodology to identify themes, and then uses his historian's framework to interpret those themes.

Nelson, Richard K. Of Monsters and men – And Topic Modeling-The New York Times, Opinionator, May 29, 2011
<http://opinionator.blogs.nytimes.com/2011/05/29/of-monsters-men-and-topic-modeling/of-monsters-men-and-topic-modeling>.

Could Topic Modeling Be Used in Art History?

Possible Corpora



So, to turn this back to art history...

Could topic modeling be useful in art history? The discipline certainly has its share of large corpora...

Examples:

The Getty Portal (<http://portal.getty.edu/portal/landing>)

A relatively new project, the Portal is collecting digitized versions of published art historical texts (book and articles). When it collects a critical mass of materials, I think it would be a good candidate for topic modeling projects.

Journals in the discipline

CAA's journals, *The Art Bulletin* and the *Art Journal*, have been in existence for over seventy years and thus offer a long publishing horizon with many articles that might be usefully examined via topic modeling.

The oeuvre of icons in the field

I am guessing that the extensive writings of some of the key icons in the field – e.g., perhaps Warburg, Panofsky, Gombrich, etc. – might be usefully explored via topic modeling.

Oral histories

Of the many interesting things I learned in conducting the Kress study, one of the most surprising was the number of oral histories that have been collected by art history research repositories. There are literally *thousands* of oral histories of key figures in the field (artists, art historians, art dealers and others important to the discipline) distributed around these repositories that I think would be ripe for topic modeling. (As an aside, if newspapers are the underutilized resource of the historian, I think oral histories may be the underutilized resource of the art historian....)

Images?

At a digital art history symposium hosted by the Institute of Fine Arts in December of 2012 (see <http://www.nyu.edu/gsas/dept/fineart/research/mellon/mellon-digital.htm>), Professor Daniel Rockmore (Professor of Mathematics at Dartmouth) wondered aloud whether it might be possible to conduct topic modeling on images. I am not sure how this would work since topic modeling is a text mining technique, but I am guessing that one possible way to pursue this is by tagging a critical mass of images of works of art in a consistent manner and then applying topic modeling to these tags. This is a total assumption on my part, but I think it is an area worth exploring with some topic modeling experts.

The “Art” of Digital Art History



Banksy (2008)

I am going to stop here, and close with some cautions, as epitomized here in a more visual way by Banksy.

I am proposing that the phrase *digital art history* be used more rigorously to refer to the research and scholarship that results from methodologies that are associated with computational technologies. But just because you *can* use a tool doesn't mean you *should*.

Also, identifying which of these computational methodologies offer the greatest potential for your research – which will help you “see” an art historical problem in a new light, which will bring forth new questions to pursue – is more of an art than a science. As in traditional art historical practice, you have to choose the right methodology for your research needs.

But, I do believe that it is the use of these computational methodologies for art historical scholarship that is really what *digital art history* is all about. Simply using online resources or communication channels does little to **further** art history's intellectual frameworks, while the research and scholarship generated from computational methodologies *does* have the potential to do this - to expand art historical practice in a transformational way.

Thank you.