


2005

## In Praise of Pattern

Stephen Ramsay

*University of Nebraska at Lincoln*, [sramsay.unl@gmail.com](mailto:sramsay.unl@gmail.com)

Follow this and additional works at: <http://digitalcommons.unl.edu/englishfacpubs>

 Part of the [English Language and Literature Commons](#), and the [Film and Media Studies Commons](#)

---

Ramsay, Stephen, "In Praise of Pattern" (2005). *Faculty Publications -- Department of English*. 57.  
<http://digitalcommons.unl.edu/englishfacpubs/57>

This Article is brought to you for free and open access by the English, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications -- Department of English by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

## In Praise of Pattern

Stephen Ramsay  
University of Nebraska  
sramsay@uga.edu

*All following figures for this text (which are referred to as “plates” by the author) are available online at:*

*<[http://texttechnology.humanities.mcmaster.ca/ramsay\\_figures/](http://texttechnology.humanities.mcmaster.ca/ramsay_figures/)>.*

### Abstract

*The exploration of pattern may be usefully regarded as the strongest point of intersection between the computational strictures of text analysis and the open-ended interpretive landscape of literary studies. Seeing computational analysis in literary studies as a quest for interpretations inspired by pattern can, moreover, lead to a change in the perception of text analysis among more mainstream literary critics by moving the hermeneutical justification of the activity away from the denotative realm of science and toward the more broadly rhetorical and exegetical practices of the humanities. This article presents the author’s creation of StageGraph -- a tool for the visualization of dramatic structure -- as a “narrative of process” and uses it to consider the implications of such a change against the broader backdrop of literary studies. By embracing a more humanistic vision of computational work, the author argues, text analysis can take its rightful place in the spectrum of critical tools.*

KEYWORDS: Text analysis, Graph theory, Literary criticism, Drama, Critical theory.

If computational text analysis is to move beyond its current status as a narrow specialization and join the broader discourse of the humanities, it must find ways to occupy a useful space in the landscape of interpretive inquiry. This will undoubtedly involve the development of tools that enhance the interpretive process beyond mere finding aids and pre-interpretive organizations (such as word frequency generators and concordances). But more importantly, it will mean active resistance against the perception that we are out to provide scientific solutions to interpretive problems. Such resistance need not imply the abandonment (or concealment) of the rigorously mathematical procedures that often underlie our

activities. The apparently objective realms of mathematics and computing can be made to fit with the insistently subjective processes of interpretation precisely because computational processes, when motivated by interpretive questions, are already fully aligned with the imperatives of humanistic inquiry.

A claim like this is difficult to justify by pointing to a tool and enumerating the results it produced, because doing so is likely to reinforce the impression that matters accomplished with a computer must perforce lay claim to scientific objectivity. Even when we acknowledge a tool's limitations—pointing, for example, to margins of error or flaws in the data—we still run the risk of portraying the activity as essentially a quest for factual information. To see how a tool might be otherwise, we have to look both to the purpose for which a tool was built and the context in which it was developed. To this end, I'd like to present my own work on the development of a text analysis tool, not as a presentation of results, but as a narrative of process. In doing so, I hope to record not just how an English professor approaches mathematics and computation, but to show how one cannot help but remain the former while doing the latter.

The tool in question is called StageGraph, and while we might say that its purpose is to facilitate the study of structural properties in the plays of Shakespeare, such a statement already misleads by substituting the result for the process. When I started building StageGraph, I wasn't seeking answers to questions about Shakespeare. I was instead investigating the mathematics that underly computation itself—in particular, that branch of discrete mathematics known as graph theory.

Graph theory has been described as “the geometry of position,” and its theorems have applications in areas ranging from the topology of computer networks to the structure of molecules. The field is said to have been inaugurated by the eighteenth-century Swiss amateur mathematician Leonhard Euler, who had originally set out to do nothing more than provide a clever solution to an interesting puzzle. The puzzle involved the Prussian city of Königsberg, which was divided into four regions (including one island) by the river Pregel. The four regions were connected to one another by seven bridges, and the townspeople, who were fond of taking walks about the city on Sunday afternoons, had wondered for a long time if it were possible to wander about the town crossing each bridge only once and end up back where you started (Rosen 475). Everyone who had tried (including Euler himself) had failed in the attempt.

Euler was able to prove that the proposed journey is impossible.

In fact, he was able to provide a formula that could determine whether such a journey was possible for any arbitrary configuration of bridges and landmasses. His proof came about from his having discerned some basic properties of such journeys (that, for example, one would need to have an even number of bridges to get on and off a given land mass without going over a bridge twice). The article in which the solution appears is widely considered a masterpiece of mathematical explanation, in part for its lucid distillation of the problem down to its essential features. To me, the most interesting part of the article is the pictures.

Figure 1—a fair copy of the original—is not so much a map as a schematic visualization of certain salient aspects of the data. It would be of little use to you if you were trying to make your way around Königsberg. He has removed the streets, the buildings, and the topography; replaced the bridges and rivers with cartoonish representations; and labeled the landmasses with letters. A modern graph theorist would draw an even sparer diagram, but one that is not far different in its essentials from the one Euler drew.

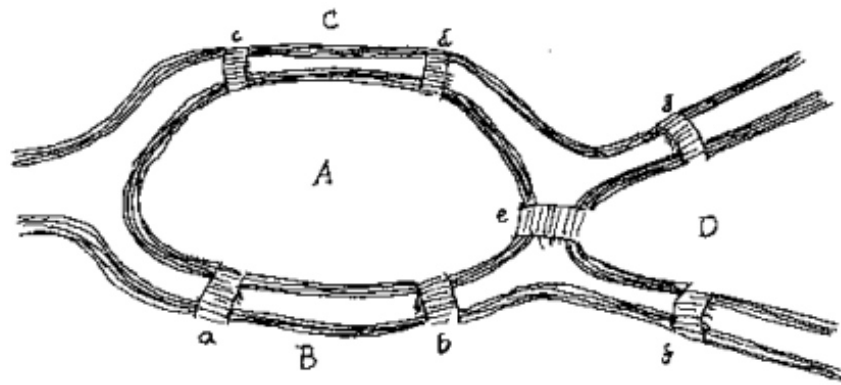


Figure 1: Schematic diagram of the bridges of Königsberg.

Euler uses his drawing to demonstrate the terms of his proof, but the diagram is not itself the proof. The diagram is a visualization tool intended to spur one to insight about what is essential to the problem. It appears as an aid to the reader, but it is clear that it must have been the principal aid used by the prover himself. One imagines Euler lingering over his drawing first without any idea of its true relation to the problem at hand—perhaps tracing his finger over the bridges in an attempt to take the journey with less exertion than one would need on foot. The proof itself isn't there, but the drawing provides Euler with the “noticings” he needs

in order to get to the proof.

One might argue that visualization always serves this purpose, but the rhetorical context in which a visualization occurs can easily obscure the heuristic nature of the evidence. Most of the visualizations one sees in text analysis are there to demonstrate the facts of the case—to prove to the reader that things cluster this way or that, that there are indeed more instances of this feature than of that feature. Relatively few of them are there to offer the reader the open possibilities of interpretive insight. And this is odd, when we consider that the kinds of texts that interest humanists are solidly of the latter variety—less concerned with proving a point, and far more concerned with allowing the reader the intellectual latitude to see something new.

I became interested in the idea of graphing structures and with the notion of having the computer generate visualizations that could help me to notice new things, and so I proceeded to do something that has been several times declared methodologically illegitimate by text analysis practitioners: I went on a fishing expedition.<sup>1</sup> I tried to think of things that naturally form themselves into graph structures. Many things do. Any set with repeating elements can be represented as a graph, since such structures can easily be construed as forming a network of relationships. It occurred to me that the scenes in a play can be viewed this way, and so I decided to write a program that could generate graphs of the scene changes in Shakespeare.

I need to make clear the utter absence of a research question or hypothesis at this point in my investigations. I had only a vague idea of what the graphs would look like, and no idea at all regarding their usefulness for literary study. I can also be justly accused of having first found a technology and then having gone in search of some way to apply it. Were I a scientist, all of this would have made for a dubious methodology indeed—the sort of history one would want to subject to significant revision on a grant application. But I think we need to ask whether it is likewise a dubious methodology in the context of the humanities.

Consider the following research methodology. I read a novel. I notice things about it that confuse or intrigue me. I remember similar things in other novels, and before long, I am actively seeking further instances. I begin to suspect that there might be something to my original impression, and I start to think of ways to make sense of it all. I discover people who disagree with my emerging sense of things, and I find myself engaging their thoughts and trying to square them with my own. Finally, I write a

polished critical article on D. H. Lawrence's high modernist literary critique of the Victorian marriage plot.

My colleagues in English studies might offer a number of censorious assessments of my conclusions, but surely they would have no problem at all with my having stumbled, somewhat fortuitously, upon a pattern in Lawrence that I hadn't been looking for in advance, drew conclusions about its relation to a set of Victorian novels that did not include all Victorian novels, and deliberately scoured through the critical literature looking for a fight. They would all recognize—somewhat sheepishly, perhaps—that this is how most humanistic journeys proceed. Some might argue that a good humanities research methodology should help to take the serendipity out of this process. But when it comes to exegetical work, serendipity often is the process.

I spent several months working on what came to be known as StageGraph. As a literary critic, I was sustained by the hope that I would see something new. As someone who has spent many years programming computers, I was sustained by the hope that the entire thing could be automated. The former hope came alive in large part because the latter hope failed utterly.

To begin with, I discovered that my ability to instruct the computer concerning the nature of a scene was significantly hampered by the fact that I myself wasn't sure what constituted a scene location. Things are easy, when, as in *Twelfth Night*, Shakespeare proceeds from the Duke's palace (Act I, Scene i) to the Seacoast (Act I, Scene ii). It's a lot more difficult to say what a scene is when Shakespeare says (as he does in *The Tempest*) that it takes place in "another part of island" or (as in *As You Like It*) "a room in the palace." Is it the same part of the island we were looking at the last time we were at the other part of the island, or a new part? Can we assume that the Duke is speaking to his Lords in the same place where Celia and Rosalind have lately had their tête-à-tête, or would it make more sense to have it occur in a different place? What, after all, is a "place" in a play? A nineteenth-century performance on the proscenium stage at Covent Garden might have let us know by a physical change of scenery (or perhaps not, in a production of *Love's Labour's Lost* where the entire play takes place in various parts of "The King of Navarre's Park"). What we know of Shakespeare's own stage would lead us to believe that such matters were left mainly to the audience, there being nothing but changes in character and costume to suggest a change of scene. And, of course, editors of Shakespeare's plays do not uniformly agree on the scene

divisions. *Antony and Cleopatra*, a play notable for its rapid scene changes and wide-ranging settings, had no typographical indications of scene divisions at all until 1709. And so, faced with this computationally intractable buffet of confusion, I did what any good humanist scholar would do: I guessed.

Or rather, I chose—armed with a new awareness of the hidden complexities in a matter so simple and informed by several years spent studying Shakespeare and the scholarship surrounding him. Had I not been so intent on getting StageGraph written, I might have written an article detailing the various complexities of the scene dynamics in Shakespeare's plays and its implications for interpreting Shakespeare. I would then have been allowed to let all of these complexities hang in mid air without privileging any particular decision. My point might then have been merely to detail the various consequences of each possibility—perhaps only to marvel at the diverse nature of the exegetical landscape. But I was trying to teach a computer to grapple with these matters, and the computer forces one to choose. It will not figure it out, and it will not accept half measures.

I built StageGraph as best as I could, which meant trying to be as transparent and as consistent as I could about the choices I had made. In so doing, I was merely following one of the research standards of our discipline. If you are going to make a claim based on a computational process, you must be very clear about the steps you took and the assumptions you made. If, however, the goal is not merely to allow claims but to facilitate them, we must be prepared to create tools that offer the user the ability to choose steps and assumptions. For StageGraph, this clearly implies that the final version should permit the user to define things at many different levels of granularity, under a variety of conditions, and with enough documentation to allow the user to grasp the assumptions. This does not imply that the software should be neutral, as many tools and web sites in digital humanities try to be. It cannot be neutral in this regard, since there is no level at which assumption disappears. It must, rather, assert its utter lack of neutrality with candor, so that the demonstrably non-neutral act of interpretation can occur.

I have spent many hours pouring over directed graphs of Shakespeare's plays, and I find the activity utterly absorbing. I am struck by the number plays that end with a scene that occurs nowhere else in the play. *The Comedy of Errors* (an early farce, [link 1](#)), *Richard II* (a history, [link 2](#)), and *Cymbeline* (a late romance, [link 4](#)) end this way, while so many other

plays end at some location that had occupied a position of centrality in the narrative. Battles sometimes appear (e.g. in *Coriolanus*, link 5) as strange limbs extending off the main trunk of the scene structure, but in others they appear more integrated (*Antony and Cleopatra*, link 6). The graphs do not depict the plays in terms of metric space, and yet one cannot help but be struck by the scenes that the graph layout algorithm chooses to centralize: The Garter Inn (*Merry Wives of Windsor*, link 7), Eastcheap (*Henry IV, Part 1*, link 8), and “A room in the prison” (*Measure for Measure*, link 9). Some are extremely linear (*Julius Caesar*, link 10), others start out linear and then divide into more expansive formations (*King Lear*, link 11). The first (link 12), second (link 13), and third (link 14) parts of the *Henry VI* cycle show this process happening over the course of, well, history.

StageGraph also allows one to display the subgraph formed by the movement of individual characters, so that one can, for example, track Antony’s path through the play (link 15) or Cleopatra’s (link 16). One can even do them both at the same time (link 17)—coloring one red and the other yellow, with shared scenes colored orange. At times, we notice features that would be extremely difficult to discern from linear reading. Graphing Antony and Cleopatra’s movement through the play, for example, reveals that they meet only once outside of Alexandria in a scene that is conventionally labeled either “A plain near Actium” or (following Plutarch) “Taenarum.” It is, in fact, the only scene in the entire play for which we cannot deduce a clear location from internal evidence.

What I do not see in these graphs, however, are clear instances of objective data that serve to adjudicate some humanistic problem. Some of these things have been noticed before by others (though I suspect some have not). We might want to say that the computational element—the totalizing algorithms upon which StageGraph relies—provides what Susan Hockey termed “concrete evidence to support or refute hypotheses or interpretations which have in the past been based on human reading and the somewhat serendipitous noting of interesting features” (Hockey 66), but it seems to me that I am using the computer precisely to do such reading and such noting. What, after all, is “concrete evidence” in interpretation? If it is the sort of thing that arrests discussion (by irrefutably proving that certain kinds of exegetical acts are either indisputable or impossible), it is also likely to be something too banal to merit attention. One does not, after all, “refute” a complex idea about a question of influence or of broad historical analogy; “challenge,” “modify,” “contend with” are the more operative terms.



A couple of years ago, I had the good fortune to be invited by the Mathematics Department at Middlebury College in Vermont to give a talk on graph theory and Shakespeare. You can well imagine my terror. It is one thing to talk about graph theory to English professors. It is another thing entirely to talk to graph theorists about graph theory (when you are an English professor). After all, these are the people who really are dealing with the calculus of fact and truth. Surely they would want to see proof and refutation, and I (who have been championing the legitimacy of a less scientific approach to digital humanities for several years) would be forced to defend my use of their techniques in the unruly pursuit of interpretive maybes.

There were several very striking aspects to that visit—not the least of which was their eagerness to communicate their deep love of my subject (by which they meant literary study, not digital humanities). Several of them obviously worried that they were too much the amateurs as literary critics to engage in a serious discussion of literature—an irony made all the more profound by my constant fear that I was going to forget the proper definition of graph planarity while speaking with them about math. I had hoped that even if the talk wasn't a great success, that perhaps they could help me think more clearly about how to visualize the graph structures.

When it came time for the talk, I stood up and made a number of obsequious gestures intended to communicate my deep humility toward the project of real mathematics—a pursuit with which, I tried to assure them, my own work should not be confused. I then proceeded to make a few very guarded observations about Euler's theorems, discuss the state of the art in scholarship on Shakespeare's dramatic structure, and showed them my graphs.

When it was over, there were a dozen hands in the air. "How wide is *Hamlet*?" "What is the chromatic number for *A Midsummer Night's Dream*?" "Have you run any shortest path algorithms on *Henry V*?" I had been showing them how I look at these graphs, but it was clear that they saw them in a very different light. To start with, they cared very little about the visual aspects. For them, the visualization was just a convenient way to present a set of mathematical statements and properties. They were far more interested in the properties themselves, and in finding out if there were any suggestive patterns to be gleaned from an analysis of the differences. More astonishing still, though, was the fact they seemed to feel that "suggestive pattern" (a vague term if ever there was one) was a perfectly

worthy goal for all of this. Some of them had devoted their adult lives to proving things about the properties of acyclic directed graphs, and yet they found the notion of proving things about Shakespeare to be slightly dispiriting. My standard arguments about the methodological differences between the sciences and the humanities met with a somewhat puzzled “How else would it be?”, and I was once again confirmed in my belief that the exaggerated epistemology that has come to be known as “scientism” is far more prevalent among humanists than scientists.<sup>2</sup>

I knew I could modify StageGraph so that it could generate graph properties, but a table full of such properties seemed far less likely to yield the sorts of epiphanies I was getting with the graphs. I also wasn't sure that the invariably low-level properties encapsulated in measurements like degree number and graph diameter were going to have any relation to the larger structures in which I was interested. In a sense, low-level graph properties seemed like a step backward. The graphs interested me in part because they seemed to provide holistic evidence of pattern. Knowing that, for example, *Antony and Cleopatra* had more scenes than *Twelfth Night* seemed to come back to the banality of the demonstrable. Would I not have objective evidence of a number of things about which no one particularly cared? Shortly afterward, I had the good fortune to present this problem to a colleague who is at once a mathematician, a computer scientist, and a digital humanist. He told me that I had a classic data mining problem.

I had been following research in data mining casually for some time. It appealed to me tremendously, because I saw in its quest for “suggestive pattern” a natural analogue to the sort of thing I was trying to do as a literary critic. One introduction to the subject states its purpose as follows:

Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. The idea is to build computer programs that sift through databases automatically, seeking regularities or patterns. Strong patterns, if found, will likely generalize to make accurate predictions on future data. Of course, there will be problems. Many patterns will be banal and uninteresting. Others will be spurious, contingent on accidental coincidences in the particular dataset used. And real data is imperfect: some parts are garbled, some missing.

Anything that is discovered will be inexact: there will be exceptions to every rule and cases not covered by any rule. Algorithms need to be robust enough to cope with imperfect data and to extract regularities that are inexact but useful. (Witten xvix)

And there you have it. Find interesting patterns and regularities in data that is generally held to be of the deepest significance, but which nonetheless contains the spurious, the contingent, the inexact, the imperfect, and the accidental in a state of almost guaranteed incompleteness. I have read few more accurate descriptions of the central task of literary criticism.

As it happened, I and several colleagues at other institutions were in the middle of applying for a large grant to fund the creation of web-based text analysis tools for existing text archives,<sup>3</sup> and the project group included a number of very skilled data miners from the National Center for Supercomputing Applications at the University of Illinois. We decided to conduct data mining experiments with StageGraph as part of the prototype.

I began by rewriting StageGraph so that it could generate five properties for every graph: the number of unique scene locations, the total number of scenes, the number of single-instance scenes, the number of loops (scene locations that appear consecutively), and the number of switches (consecutive scene locations with an intervening location). My isolation of these five properties wasn't arbitrary. These were the five simple things that I seemed to notice most often, even though I wasn't really able to correlate them into groups. I also wasn't at all sure how to gauge the relative importance of each feature (important in what regard?). I turned, therefore, to naive Bayesian analysis—a method of classification that tries to capture the changing probability that something is or is not a particular class of thing based on previous classifications. I decided to take the five properties I knew about and see if they could be used to classify Shakespeare's plays into the four traditional genre categories given to the plays: comedy, tragedy, history, and romance.

In this phase of the project, I was greatly assisted by Bei Yu—a graduate student at the University of Illinois who works on data mining problems at NCSA.<sup>4</sup> She ran several standard data mining algorithms on my data, including decision-tree generation, naive Bayes, and a technique she had worked on with two other researchers (Jun Wang and Les Gasser): shaded similarity matrices formed using concept tree clustering.

Her preliminary results were (she thought) somewhat discouraging. The comedies cluster together very well, as do the histories, but romance is very hard to distinguish from tragedy. That's a somewhat banal statement for a literary critic, since people have been saying precisely this of these plays for several hundred years. But I was struck by the fact that a fairly simple algorithm operating on extremely low-level properties managed to run up against the same conundrum. Besides this, the computer had taken one of the most abstruse and persistent questions in genre study (what characterizes the comedic genre) and reduced it to a simple test: if it has fewer than 9.5 scenes, it must be a comedy.

But there was nothing at all banal about the shaded similarity matrices. Shaded similarity matrices have been a popular tool in data visualization for some time. The basic idea is that you take a set of data measurements for a set of classes and use it to create another table that expresses degrees of proximity among those measurements (using something like Euclidean distance or product-moment correlation—similar to what many text analysis practitioners use for factor analysis). You then reorganize those values so that “more similar” values are adjacent to one another, assign each value a color, and arrange them in a grid. Wang, Yu, and Gasser modified the usual technique by replacing the standard correlation measurements with concept-based clustering (the concepts, in this case, being the genre categories).

Bei and I eventually arrived at the visualization displayed in link 18.

The plays are listed along the Y-axis and colored according to their various genres (red for comedy, green for history, purple for tragedy, and yellow for romance). We can imagine the X-axis as having the same list of plays going from left to right (the X-axis names are omitted in the diagram). Each square in the matrix shows how similar one play is to another by the brightness of the color. The diagonal shows consistently dark values, since each play is presumed to be perfectly similar to itself (*pace* McGann).

Yu's concept clustering algorithms had first found a broad split between plays with more than 12.5 single-incident scenes and plays with fewer than 12.5 such scenes. When we indicate that split on the matrix, and reorganize the plays according to how strongly they resemble one another in terms of that factor, we start to see different patterns developing (link 19). The comedies are clustering together strongly, but the differences between history and tragedy are becoming intermingled with one another

and re-assembled into substructures. Some of the histories are starting to seem more tragic (which makes sense for plays like *Richard III*), though some are likewise clustering toward the comedies (which makes sense, perhaps, for late plays like *Henry VIII*). The romances, too, are beginning to move as one would expect them to: the tragi-comic ones are moving toward tragedy, and the comedo-tragic ones are moving toward comedy.

Add another set of factors (number of scenes), and the alignments begin to get finer (link 20). The most comic of the comedies are clustering at one end, and the more tragic of the tragedies are clustering at the other.

Add all the factors, and things really begin to look sensible (link 21). The comedies are, for the most part, clustering together, and the tragedies are, for the most part, clustering together. There are several curious anomalies, but the clear cases seem to have been adjudicated properly. Moreover, the curious anomalies appear to conform to some of the more famous critical statements about the plays. For example, one very influential critic has argued that both *Othello* and *Romeo and Juliet* resemble comedy, without making any mention of “low-level” structural features (such as scene loops and switches).<sup>5</sup>

One is tempted—almost behooved—to make sense of the entire arrangement. Indeed, we might almost convince ourselves that such matters as “number of scenes” and “number of loops” prove something essential about Shakespeare’s genres. But this, of course, is nonsense. These methods and visualizations prove nothing at all. Indeed, to assert that these extremely low-level features are somehow constitutive of genre would be to perpetrate a ham-fisted abuse of statistics and a grotesque parody of scientific method simultaneously. But what, then, does all of this do?

It forces us to move our eyes over Shakespeare’s plays as Euler’s eye must once have moved over the bridges of Königsberg. I have never thought (at least in structural terms) of the ways in which *The Tempest* resembles comedy, the ways that *All’s Well that Ends Well* (one of the infamous “problem comedies”) resembles (of all things) history, or the ways in which *Henry V* resembles tragedy. The fact that I’m being led in such directions by low-level structural features—some of which barely register in one’s consciousness during the ordinary act of reading—raises an obvious question: How do the low-level matters of dramaturgy relate to the high-level matters of genre?

Now, if we say nothing more than that (for example) “comedies tend to be smaller and more compact than tragedies,”—a statement which, to our embarrassment, is not too far from the sort of thing we text analysts

are wont to advance (tentatively, pending further research) in our scholarly journals—we are both begging the question and avoiding it at the same time. We are begging the question, because we don't really know what "small and compact" means in the context of theatrical art; we are avoiding the question, because even if we are prepared to define smallness and compactness, we haven't said anything at all about *why* this might be so. These are interpretive questions, and in the end, they are the only kind of questions that really matter in the humanities. I would like to say that *The Tempest* resembles comedy because it overlays a more complex narrative upon the formal structure of the marriage plot—a world of tight social networks, resolved separations, and filial adjacencies far more conformable to a stage full of repeating locations and compact dynamics. I would like to say that *All's Well That Ends Well* resembles history because the problem element of this problem comedy entails separations that are never resolved, thus demanding far fewer repetitions and much greater scene diversity. I would like to say that *Henry V* resembles tragedy because in this case the "history" being told is the education of Prince Hal—an education which, in classic tragic form, involves sudden realizations that the small, closed world of experience is perhaps (as in *Lear*) not as it had first appeared. And why not? If the goal was to prove something, such comments are beyond the pale. But surely that wasn't (and isn't) the goal. The goal is to say something new, provocative, noteworthy, challenging, inspiring—to put these texts back into play as artifacts reconstituted by re-reading.

I began this article by pointing to an anxiety, and I have come around, perhaps, to the same anxiety in reverse. If we interpret our graphs and tables in this way, are we not introducing humanistic inquiry where it doesn't belong; inflicting an inappropriate humanism upon the cherished positivism of scientific inquiry? Well, yes. Text analysts use sophisticated software and complicated mathematical techniques to elucidate texts. We believe this is a fascinating and deeply engaging intellectual pursuit. Yet we manage to scare off (or appear irrelevant to) most of our colleagues in the more mainstream arenas of humanistic inquiry. Some might say that the solution is to alleviate our colleagues' anxiety by educating them on the ways of sophisticated software and complicated mathematics, but in the end, I wonder if it is not our own anxiety that lies at the root of the problem. We are so careful with our software and with our mathematics—so eager to stay within the tightly circumscribed bounds of what the data "allows"—that we are sometimes afraid (or we forget) that all of this

is meant to lead us to that area of inquiry where such caution and such tentativeness has no place. Not because interpretation must be careless and bold, but because it has to risk the perils of subjectivity in order to keep true to its own objectives—objectives which, in the context of literary study, seldom involve the amassing of verified facts.

## Notes

<sup>1</sup> See, for example, Rosanne Potter's comment that "The *sine qua non* for avoiding data inundation is a firm resolve to go to the data only when one is testing a clearly stated hypothesis" [333].

<sup>2</sup> Of course, one might persuasively argue that mathematics was originally a humanistic pursuit, and that it remains so despite the efforts of universities to place it within the organizational structure as a branch of the sciences.

<sup>3</sup> This project was eventually funded, and became The Nora Project (<http://www.noraproject.org>)

<sup>4</sup> National Center for Supercomputing Applications (<http://www.ncsa.uiuc.edu/>)

<sup>5</sup> See Snyder, *op cit*.

## Works Cited

- Gadamer, Hans-Georg. *Truth and Method*. Trans. Joel Weinsheimer and Donald G. Marshall. Continuum, New York, 1996.
- Hockey, Susan. *Electronic Texts in the Humanities*. Oxford UP, Oxford, 2000.
- Lubar, Steven. "'Do Not Fold, Spindle, or Mutilate:': A Cultural History of the Punch Card." *Journal of American Culture*, 15.4:43-55, 1992.
- Potter, Rosanne G. "From Literary Output to Literary Criticism: Discovering Shaw's Rhetoric." *Computers and the Humanities*, 23:333-40, 1989.
- Rosen, Kenneth H. *Discrete Mathematics and Its Applications*. McGraw Hill, Boston, 4th edition, 1999.
- Snyder, Susan. *The Comic Matrix of Shakespeare's Tragedies*. Princeton UP, Princeton, 1979.
- Wang, Jun, Bei Yu, and Les Gasser Wang. "Concept Tree Based Clustering Visualization with Shaded Similarity Matrices." *Proceedings of the 2002 IEEE International Conference on Data Mining*, 697-700, 2002.
- Witten, Ian H. and Eibe Frank. *Data Mining: Practical Machine Tools and Techniques with Java Implementations*. Kaufmann, San Francisco, 1999.