Searching for “Spitzer” In the Archives: Drawing New Conclusions from the “Troopergate” Scandal Using Archival Records and Social Network Analysis

CATHERINE STOLLAR PETERS
Department of Informatics, University at Albany

Abstract: In the Spring of 2009, I analyzed a corpus of documents related to the 2007 “Troopergate” scandal in New York State. The documents were gathered by the Albany County District Attorney’s Office during a the investigation of Darren Dopp, Governor Eliot L. Spitzer’s Director of Communications, to determine if he had perjured himself or committed other crimes during a 2007 inquiry into the “Troopergate” case. The documents are publicly available, not under copyright and contain no legally sensitive material. Using text indexing software (the Lemur Toolkit) and social network analysis techniques (and UCInet software), I analyzed 8,562 archival documents in the series.

My research fulfills three goals. First, I demonstrate improved access methods to digital archival documents through sophisticated indexing as compared to the simple posting lists created by native PDF searching. Second, I use archival documents to create visualizations of relationships over time between important actors in the “Troopergate” scandal including Eliot Spitzer, Joseph Bruno, Darren Dopp and other Executive Chamber staffers. Third, this project employs methods of document analysis that can be applied to digital records for use in archival appraisal, description, arrangement, and use. Although my research was conducted using scanned images of documents, all of the techniques I used can be applied to born-digital documents in future investigations of description and use of born-digital archival materials.

Introduction

During his short fourteen-month tenure (January 2007 – March 2008) as Governor of New York, Eliot Spitzer and his administration suffered through two major scandals. The most memorable scandal, involving Spitzer’s extra-martial tryst with a call girl, caused a political uproar eventually leading to his resignation. Prior to the call girl scandal, Spitzer and officials from his administration were also involved in another scandal called “Troopergate” due to the involvement of the New York State Police. Six separate state and local governmental entities investigated Spitzer and officials in his administration as a result of the “Troopergate” scandal. This paper discusses further research conducted using archival electronic records gathered during investigations into the “Troopergate” scandal to reveal potential relationships previously undisclosed and to demonstrate novel approaches to research using archives and research within the archival discipline.

Purpose

This paper describes research conducted using information retrieval and social network analysis techniques on two bodies of archival electronic records. Archival records in electronic form provide researchers with innovative new methods of establishing relationships between individuals. An extension of the same methods of information retrieval and social network analysis as presented in this paper could be applied to archival electronic records to demonstrate relationships between records (like actors in a network) potentially useful for electronic records appraisal, arrangement and description. Although this research demonstrates a case study of innovative use of electronic records from a political scandal, application of the concepts of the research could be broadly applied to digital archivy.
Scope

At the heart of each investigation into alleged wrongdoings of the Spitzer administration during the “Troopergate” scandal are thousands of records detailing conversations, meetings, requests, financial transfers, directives and other interactions by the governor and his staff. Each investigation gathered similar record types: emails, interview transcripts, articles, sworn statements, handwritten notes and state policies. The Albany County District Attorney’s records gathered during Investigation “D” were accessioned by the New York State Archives, assigned the series and accretion number B1989-08 and are available for research. The Commission on Public Integrity records are available with the Notice of Reasonable Cause at http://www.nyintegrity.org/enforcement/norc.html. There are overlaps of records between and within the documents gathered by both investigations. These records are the raw materials used in this paper to create a social network of actors in the scandal. The records demonstrate proven interactions between individuals and, when analyzed using social network analysis techniques, reveal new relationships within the network.

Methods

Two main methods were employed during this research: information retrieval and social network analysis. Information retrieval is a field of study interested in searching for documents, searching for information and finding metadata about documents. In this project, two pieces of software were used to search records from the “Troopergate” scandal including Adobe Reader and the Lemur toolkit (http://www.lemurproject.org/). The default software most researchers will use to sort through the records is Adobe Reader because of its ubiquity, familiarity by users and simple graphical user interface. Compared to the open source indexing and retrieval software the Lemur Toolkit, Adobe Reader produced slow search results (over 10 minutes for a search on “Spitzer” compared to under 0.8 seconds for the Lemur Toolkit), was less efficient and had limited advanced searching potential. Through information retrieval and the use of itemized lists of correspondence created for the records from the Commission on Public Integrity, I created a matrix documenting email and memo exchange between individuals within the network. The matrix was used to create a social network that could be analyzed.

Social Network Analysis focuses on “relationships among social entities and, on the patterns and implications of these relationships. (Wasserman and Faust, 1999, p. 4.) It is a research methodology with roots in sociology, statistics and data analysis that focuses on interdependent, not independent, actors (or nodes) within a network, the ties between the actors that allow for flow of resources in the network and models of the network that focus on individual constraints or opportunities and general patterns of social interaction. (Wasserman and Faust, 1999, p. 5.)

This paper describes research on electronic archival records as a case study for using records to document interactions between individuals and the context of creation of records. Archivists can also use information retrieval and social network analysis in the archival environment to demonstrate relationships between records for appraisal, arrangement and description.

Case Study Findings and Results

Background

A background understanding of the “Troopergate” scandal will be helpful to understand the results of this research. The next few paragraphs provide the context in which records gathered during investigations were created.
The “Troopergate” scandal originally began with an exposé of then-Senate Majority Leader Joseph Bruno’s use of state aircraft for non-state related travel and ended with evidence of politically motivated surveillance of Senator Bruno using State Police by some Spitzer administration officials. The scandal broke with a July 1, 2007 Times Union article by James Odato (Odato, 2007) reporting Senator Bruno’s misuse of State Aircraft and State Police chauffeuring vehicles for transportation to political fundraisers. On July 5, 2007 journalist Fred Dicker published an article in the New York Post (Dicker, 2007) presenting new allegations of wrongdoings by Spitzer administration officials and potentially Spitzer himself. In the article, Senator Bruno claimed that State Police were documenting and scrutinizing his travel beyond legal requirements on the instruction of Spitzer administration officials. New State Police record-keeping policies implemented in April 2007 coincided with increasingly difficult relations between the Governor and Senator and produced a number of records relating to Senator Bruno’s travel. Darren Dopp, the Governor’s Communications Director, shared the newly “required” records with both Odato and Dicker (as described in a later email from Glen Miner to Dicker on July 11, 2007.) It is important to note that the Public Information Officer for the State Police Glen Miner refused to share records with Dicker\(^1\), who requested the records orally, and instead, Miner followed previously established FOIL procedures requiring request for records in writing. Dopp’s eagerness to share records about Bruno’s travel with reporters was at the heart of the “Troopergate” scandal.

After the Times Union and New York Post articles appeared in early July 2007, a wave of investigations into the actions of Bruno, Spitzer administration officials and the State Police began. Six governmental entities launched multiple inquiries into the scandal including the State Attorney General, the Albany County District Attorney, the Commission on Public Integrity (formerly the Ethics Commission), the Senate Committee on Investigations and Government Operations, the State Inspector General and the State Commission of Investigation. A timeline from July 2007 to August 2009 detailing multiple investigations, often by the same entity, is included in Figure 1.

**Figure 1. Timeline of “Troopergate” Investigations**

- AG= Attorney General of New York State
- DA= District Attorney of Albany County
- CPI=Commission on Public Integrity
- SCI&GO= Senate Committee on Investigations and Government Operations
- IG= Inspector General of New York State
- CI= State Commission of Investigation

The final outcomes of each investigation varied. Although the Attorney General and the Inspector General agreed on the findings in their investigations, the District Attorney published an initial report

---

\(^1\) Glen Miner describes his response to Fred Dicker’s oral request for records in an email to Governor’s Counsel Paul Larrabee on March 14, 2007 and in an email directly to Dicker on July 11, 2007.
concluding no real acts of misconduct by the Spitzer administration. The July 23, 2007 final report of the Attorney General detailed the following key findings.

- The Governor’s office intended to gather information on Senator Bruno’s travel for purposes of disseminating a negative perception of Bruno to the press and the general public. Under direction of the Governor’s Liaison to the State Police (Howard), the Superintendent of the State Police (Felton) deviated from standard practices and required the State Police to gather and create records of Senator Bruno’s requests for transportation, his schedule and changes to his schedule.
- The Acting Superintendent of the State Police (Felton), the Governor’s liaison (Howard) and the Governor’s Communications Director (Dopp) should be evaluated for wrongdoings and reprimanded. (Dopp was forced to resign after the publication of the Attorney General’s report.)
- Senator Bruno’s travel during the days in question was in accordance to State regulations, but the regulations are overly permissive. (Office of Attorney General, 2007)

The Inspector General conducted an investigation contemporaneous to the Attorney General’s investigation and concurred with the final report (Office of the Inspector General, 2007). Differing from the Attorney General’s and the Inspector General’s opinions, on September 21, 2007, the Albany County District Attorney published a final report from its first investigation into the “Troopergate” scandal declaring “no unauthorized exercise of official function” was committed by any member of the Governor’s office or the State Police” (Office of Albany County District Attorney, 2007).

Another governmental entity, the Commission on Public Integrity (known initially as the Ethics Commission) launched a separate investigation in August of 2007 then stopped for a short term while the Albany County District Attorney, on request of the Commission, investigated whether or not Communications Director (Dopp) issued a false statement to the Office of Attorney General on July 22, 2007. The Albany County District Attorney’s second investigation (“Investigation D”) was the source for one of the bodies of emails used in this research. Investigation D concluded with a report declaring Dopp did not commit perjury but may have committed two misdemeanors: Offering a False Instrument for Filing in the Second Degree (penal Law Section 175.30) and Making an Apparent Sworn False Statement in the Second Degree (Penal Law Section 210.35). (Albany County District Attorney, 2008) The report stressed the need for honest testimony from Dopp to settle the “Troopergate” investigations and granted immunity to Dopp in exchange for his testimony. Investigation “D” also inquired into potential crimes committed by Executive Chamber lawyer David Nocenti under Executive Law Section 135-a relating to notary publics who practice fraud while performing duties. The District Attorney found no “viable prosecution of Nocenti under Executive Law Section 135-a.”

The Commission on Public Integrity finally issued a Notice of Reasonable Cause report on July 24, 2008, a year after the initial scandal and after the resignation of Spitzer in March 2008. (Commission on Public Integrity, 2008.) The Commission found the Governor’s Liaison (Howard), Acting State Police Superintendent (Felton), Secretary to the Governor (Baum) and Communications Director (Dopp) guilty of violating Public Officers Law Section 74(3)(h). Baum and Howard admitted to violating section 74. In addition to violating Public Officers Law Section 74(3)(h), Felton and Dopp, who contested their charges, were also found by the Commission to have knowingly and intentionally violated Section 73(3)(d) and, in Felton’s case, Section 73(3)(c). Dopp was found to have “initiated and directed a course of conduct that (i) caused the improper creation of documents by the State Police, that were made to appear as if they were official documents, for the purpose of providing those documents to the Times Union; and (ii) caused the improper collection of otherwise confidential information from the State Police, sometimes on a real time basis, documenting the times and locations of Senator Bruno’s activities in New York City during May and June 2007” (Commission on Public Integrity, 2008). The records used in the Commission on Public Integrity’s investigation are the second body of documents used to research the social network of the “Troopergate” scandal in this paper.
Records

The two bodies of records that were used in this research were from the Albany County District
Attorney’s (DA) office and the Commission on Public Integrity. Both sets of records were PDFs of
printed email correspondence, typed interviews and depositions and other printed exhibits acquired during
investigations into the scandal. Although the records differed slightly between the archival collection
from the Commission on Public Integrity and the DA’s office, most correspondence records, except for a
few emails, were duplicated in both investigations.

The records differed in organization and content. Digitized documents accessioned from the Albany
County DA’s Investigation “D” inquiry were scanned one printed page per PDF file (a 30 page interview
transcript would be scanned as 30 files), were named using numbers according to the box where each
physical page was stored and contained duplicates. Search was the only method of access to the records.
The Commission on Public Integrity records contained multi-page PDFs for unique legal exhibits and
interviews, did not contain duplicates, were arranged by type of records and included item-level listings
for records including senders and receivers for email or named interviewees for legal interviews. Item
level lists containing metadata for individual records served as a browse-friendly list for accessing files.
These lists, which required additional arrangement and description work by clerks at the Commission on
Public Integrity, were crucial to using the records because of the limitations of full-text search on similar
records used by the DA’s office.

Limitations of the scans of the DA’s records caused issues with searching, the only method of access to
the record. The records gathered for the DA’s investigation contained some records that were scanned
askew which contributed to inaccurate character recognition during OCR and lead to poor retrieval
results. Records also contained typos due to officials sending quick emails or using their blackberries
carelessly that contributed to inaccurate OCRing as well. Ideally, retention of original electronic records
instead of scans of printouts would reduce OCR errors since the text is already electronic. Spelling errors,
however, would need to be addressed in smart indexing tools that could compensate for common typing
mistakes.

Poor retrieval results were also due to the main software used for document retrieval: Adobe Reader. Poor
access to the records because of search limitations forced a heavier reliance on the item level lists created
by the Commission on Public Integrity. Reliance on the list implies that archival researchers may benefit
greatly from archival description, for some highly desired electronic records, especially when full-text
search is limited due to scanning difficulties of printed documents.

This paper will demonstrate characteristics of the “Troopergate” scandal social network through the use
of social network analysis methods. I limited data used to prepare the matrix to the 303 unique emails and
memoranda sent between March 2007 and August 2007 between actors in the network gathered during
investigations into the scandal. I limited the data because I wanted to analyze documented relationships
between actors that could not be disputed. Emails and memos are well suited for social network research;
they accurately document sender, receiver and date of communication. The corpus of communication
used for my research is not the complete body of communication for each actor, rather select emails and
memoranda gathered because of their importance to investigating the scandal will demonstrate the
complete network of the “Troopergate” scandal.
Creating the Social Network

After gathering both sets of records, I segregated the emails and memos from the rest of the records to create a sociomatrix of the 47 individuals involved in the “Troopergate” scandal. A sociomatrix is essentially a spreadsheet with all the actors in the network as columns and row headings (senders are the columns and receivers are the rows.) I recorded existence of a relationship documented in a memo or an email between two actors in the cell intersecting the sender column and receiver row. For each email, I counted a symmetric relation between sender and receiver. If Darren Dopp sent an email to Eliot Spitzer, who used his personal email address and the pseudonym Laurence, both Dopp and Spitzer would demonstrate the same relationship with each other. The original matrix was valued, meaning I counted the number of emails between actors and added the total to their respective cells in the matrix. I also allowed diagonal values in the sociomatrix, because Dopp’s emails to himself were part of the body of records. Eventually I dichotomized the matrix, turning each numerical value into a binary value of the relationship exists or it does not exist, in order to perform a number of analytic functions.

Analyzing the Results

I created the sociogram of the network using the sociomatrix and the graphics program NetDraw (Figure 2). This graph demonstrates a number of interesting characteristics of the network. First, the most remarkable tie between Executive Chamber staff and the State Police occurs between the Governor’s State Police Liaison Howard and Acting Superintendent of Police Felton. Although it makes sense that the liaison between the Police and Spitzer administration would be one of the few connections between the two, the relationship between the two actors supports the Commission on Public Integrity’s findings that the two worked together to share information about Bruno’s travel arrangements with other Spitzer administration officials. Howard’s and Felton’s roles as gatekeepers for information between Spitzer officials and the State Police are demonstrated by their connection in Figure 2 and by an analytical values taken from UCINET that measure “gatekeeper attributes” (measured as flow betweeness in social network analysis.) When reviewing the numerical quantification of centrality using UCINET social analysis software, only Anderson (with a value of 593.5), the Spitzer Press official who communicated with reporters for the Administration has similar importance as a gatekeeper in the network as compared to Felton (550.6) and Howard (714.6).

Other interesting relationships revealed through the social network show potentially close ties between Howard and Spitzer. Howard is part of a triad (a group of three actors connected with each other to make a triangle shape in the sociogram) with Baum and Dopp. Both Baum and Dopp have strong relations to Spitzer, presenting a noticeable absence of ties between Howard and Spitzer. The absence of a tie between Howard and Spitzer despite the strong relationships between Spitzer and Baum, Spitzer and Dopp, Howard and Baum and Howard and Dopp indicates two potential forbidden triads, where two actors closely relate to a third but do not relate to each other, as described by Granovetter. (1973) Dopp makes conflicting statements about Spitzer’s awareness of the records gathered about Bruno’s travel and subsequent revelation of those records by Dopp. In his testimony to the Commission on Public Integrity on October 11, 2007, while Spitzer was still in office, Dopp claimed that Spitzer was not aware of the exact records collected about Bruno that were shared with the press and that they were shared under the belief that the records were public according to FOIL rules. Later, Dopp describes Spitzer’s lack of concern for FOIL procedures and his political motivation for sharing records about Senator Bruno during Dopp’s testimony to the Office of the District Attorney on February 5, 2008. The sociogram would indicate the close relationship between Spitzer, Baum, Dopp and Howard indicating Spitzer’s awareness of surveillance of Bruno by the State Police. These relationships are further explored in the SoNIA created time-sequence representation of the network, which I will discuss further in this paper.
Figure 2: Sociogram (Graph) of the “Troopergate” Network

An intriguing aspect of the visualization of the “Troopergate” social network is the centrality of Dicker, a reporter for the New York Post, between Dopp and Miner. Miner is the Public Information Officer for the New York State Police and was the first person Dicker contacted to request records from the State Police concerning Bruno’s travel. According to emails in early July 2007, Dicker requested records from Miner who would not release the records because Dicker failed to follow established FOIL procedures. When Miner would not offer records without proper FOIL request, Dicker emailed Dopp and Dopp circumvented FOIL protocol and supplied the requested records to Dicker. Circumventing FOIL was one of the charges against Dopp in the Commission of Public Integrity’s findings.

A fourth revelation from my analysis came from the structural equivalence measure, a measure to determine if any actor in the network has the same connections as other actors thereby demonstrating similar positions in the network, to determine if any actors filled the same or similar roles as Dopp and Felton, since they faced the most serious charges from multiple investigations. I used the Profile algorithm to find similarity measures between the actors in the network using UCINET. Figure 3 shows the diagram as produced by UCINET and NetDraw. From the structural equivalence measures of the network, no other actor is structurally equivalent to Dopp (position 7 in the diagram) nor is another position in the model as centrally connected as Dopp. Nocenti (in position 14) is closely connected to a number of other positions, but not as centrally connected as Dopp. This measure follows closely with the results of a number of investigations placing most of the responsibility of the scandal with Dopp. Missing

---

2 The colors represent employee affiliation. The red squares represent actors who work for the Executive Chamber. The gray squares are reporters. A yellow node indicates Office of Attorney General Staff. State Police are denoted by blue squares. The District Attorney and his staff are shown in cyan. Each node is labeled by the actor’s last name.
from this analysis, however, are the implied relationships discovered in the forbidden triad analysis between Spitzer, Dopp and Baum and Howard because implied relationships cannot be measured using UCINET.

**Figure 3: Structural Equivalence Measure Diagram**

A fifth finding from the network demonstrates the centrality of Dopp, Nocenti and Anderson. Similar to results of the blockmodel and measures of structural equivalence, centrality measures of the network show Dopp as the most central node with the highest measures of degree and closeness centrality. According to UCINET analysis Dopp, Nocenti and Anderson are the most central actors in the “Troopergate” matrix. One cause for these centrality similarities might be that the matrix includes all of the interactions between actors. Actors involved in final stages of the scandal, after it had already been publicized, could hold the same centrality values as actors involved in the scandal while the scandal was taking place. Further investigation of the matrix using the dimension of time is warranted.

**Figure 4. Centrality Measures for Most Central Actors in the “Troopergate” Matrix**

<table>
<thead>
<tr>
<th>Actor</th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
<th>Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson</td>
<td>39.130</td>
<td>46.939</td>
<td>31.160</td>
<td>55.878</td>
</tr>
<tr>
<td>Baum</td>
<td>15.217</td>
<td>46.000</td>
<td>11.829</td>
<td>36.789</td>
</tr>
<tr>
<td><strong>Dopp</strong></td>
<td><strong>36.957</strong></td>
<td><strong>51.685</strong></td>
<td><strong>36.254</strong></td>
<td><strong>57.902</strong></td>
</tr>
<tr>
<td>Howard</td>
<td>13.043</td>
<td>42.593</td>
<td>34.571</td>
<td>15.759</td>
</tr>
<tr>
<td>Larrabee</td>
<td>10.870</td>
<td>41.818</td>
<td>5.402</td>
<td>30.205</td>
</tr>
<tr>
<td>Maloney</td>
<td>23.913</td>
<td>43.810</td>
<td>8.720</td>
<td>46.059</td>
</tr>
<tr>
<td>Miner</td>
<td>10.870</td>
<td>36.220</td>
<td>9.597</td>
<td>8.670</td>
</tr>
<tr>
<td>Nocenti</td>
<td>36.957</td>
<td>46.465</td>
<td>22.936</td>
<td>56.887</td>
</tr>
<tr>
<td>Pope</td>
<td>10.870</td>
<td>40.351</td>
<td>4.348</td>
<td>29.574</td>
</tr>
</tbody>
</table>
Interested in investigating the matrix using temporal dimensions, I used SoNIA (Social Network Image Animator) a social network animation tool created by Dan McFarland and Skye Bender-deMoll at Stanford University to create an animated visual of the “Troopergate” network. Since the emails were created between March 2007 and August 2007, I created 12 time slices, two for each half-month. Slice 1 included March 1 to March 15, slice 2 included March 16 to 31, and so forth. Since SoNIA requires directed graphs, because lines have to be drawn from one node to another, I rewrote the matrix to include the first email from one actor to another in each time slice. So if Dopp emailed Spitzer on March 1, the line would originate with Dopp and would go to Spitzer in the first time slice. If Spitzer emailed back in the same timeframe, I would not draw another arc since the relationship in that timeframe had already been established. The colors in the movie represent employment affiliation. State Police employees are blue, Executive Chamber staff members are red, Office of Attorney General staff are yellow/orange, reporters are gray and District Attorney staff members are cyan nodes. The animation is included in this paper as TroopergateFinal.mov.

The SoNIA animation begins with the nodes in the network arranged in a random pattern. The nodes begin to pull together as connections to other nodes emerge over time. The final time slice of the movie reveals a social network diagram similar to the one created by NetDraw, but animation developing the final network reveals previously absent characteristics of the network because it emphasizes the time at which relationships are developed.

The time based nature of the network reveals when actors first made connections with one another revealing more about the sequential events of the “Troopergate.” One of the first arcs in the network is between Howard and Felton resulting from emails concerning Senator Bruno’s travels in mid-March 2007. At the same time slice, an arc appears between Miner, Information Officer for the State Police, and Executive Chamber lawyer Larrabee. Miner informs Larrabee in March that reporter Fred Dicker is requesting information outside of normal FOIL procedures from Miner’s staff member Swoboda. In the third and fourth time slices, State Police staff members send emails discussing Bruno’s travel in April and some connections between Executive Chamber staff members appear. In the fifth time slice representing April 16 to April 30, two notable ties are established. Baum, the Governor’s Secretary, connects with Howard and Spitzer. This makes a strong argument for the forbidden triad that could likely exist between Spitzer and Howard, or at least an awareness of Baum and Howard’s and Dopp and Howard’s relationships and awareness of surveillance Bruno’s activities, since both strong relationships start at the same time in this network. In early May (time slice 6) a notable group between Dopp, Baum, Anderson, Spitzer and Howard forms. Dopp also becomes more central from this time slice forward. In the eighth time slice starting on June 16 and ending June 30, just before the first article appears in the Times Union, the author of the article Odato joins actors in the Executive Chamber. Dopp’s centrality is established by the ninth time slice representing early July (during which the scandal broke) and continues in the sociogram despite his forced resignation in time slice 10. None of the other actors who are as central to the network as Dopp is as demonstrated by the UCINET values of centrality have established themselves as central figures in the network. Their centrality is established after July 1-July 15, when the “Troopergate” scandal had already started and records had already been improperly gathered and shared.

Future Research

Through the use of archival electronic records, information retrieval and social network analysis, this paper demonstrates innovative uses of electronic records, methods to show interactions between individuals and methods to describe the context of record creation. Analysis of the records in the case study show how Spitzer, Baum, Dopp, Howard and Felton interacted with each during the “Troopergate” Scandal and how new techniques reveals potential new relationships between actors in the network that
were not revealed through manual analysis of the correspondence alone. The case study uses electronic records to demonstrate relationships that facilitate improper sharing of documents and information for political gain.

The research conducted using records obtained during investigations into the “Troopergate” scandal as described in this report shed new light on the social interactions of actors in the scandal and application of social network analysis using archival records. This research also demonstrates ways in which archivists and archival researchers can incorporate methods from other disciplines into archival research. Future use of information retrieval and social network analysis methods in archival research could allow archivists to appraise, arrange and describe archival records. Through information retrieval techniques, electronic records could be indexed by author beyond use of file metadata alone. One indexing technique might take documents written by a known author and compare them to documents of unknown authorship to find algorithmic or lexical similarities potentially automating document appraisal. Additionally, cluster algorithms in both information retrieval and social network analysis might be used to arrange documents by author preventing the need for archivists to manually “arrange” electronic records. Emphasizing interdisciplinary research could bring a range of new ideas and methods to archival work. Application of new methods of research would help archivists envision potential uses of electronic records by scholars and innovative options for archival practice in the future.

References


