

Applying modality and equivalence concepts to pattern finding in social process-produced data

Robert A. Hanneman · Christian R. Shelton

Received: 24 June 2010 / Accepted: 2 August 2010 / Published online: 5 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Large amounts of detailed transactional information are generated by ongoing social processes. For managing and mining such data, we treat them as “objects” and “relations”. These ideas strongly parallel the way that social network analysts conceive of social structure. Modality (roughly, distinguishing multiple classes of social actors or nodes in networks) and equivalence classes (roughly, distinguishing general patterns in the ways that objects in classes are related to one another or to objects in other classes) have proven to be very useful in helping social network analysts to think about complex relational structures among social objects. Dimensional and generalized “block models” of multi-modal social networks provide tools for designing searches to identify patterns. The ideas are illustrated by descriptions of how a number of social process-produced data might be approached, including bibliographic databases, communications logs, virtual communities, and economic transactions.

1 Introduction

Every minute of every day, huge amounts of data generated by ongoing social interactions are deposited in digital databases. These records are remarkable collections of “trace evidence” (Webb et al. 1966) produced by social

processes for their own purposes. While social scientists have always “mined” archives of records (e.g., manuscript censuses, newspapers, roll calls of votes, mortality registers) as “non-reactive” ways of understanding patterns of social structure, the current era is unique in the amount of all social transactions that are documented, the accuracy of these records, and the sheer volume of data. Not surprisingly, the “mining” of digital archives and transaction logs is a very rapidly growing enterprise within and without the social sciences (e.g., new journals such as *Social Network Analysis and Mining* 2010).

There are remarkable similarities between the ways that some social scientists think about the information in such digital archives as “pictures” of social structure, and the languages and logics of the computer scientists, database engineers, and others who have designed and built them. To date, however, communication between these two groups has been fairly limited. Most social scientists speak the languages of information sciences badly, if at all. The arcane languages and conceptual schema of the social sciences may be both unfamiliar, and seemingly irrelevant, to engineers. Some of the goals of database engineers (search, optimizing processes, assessing reliability) are quite different from those of social scientists. Other goals are quite similar (e.g., finding regular patterns and abstracting generalizations), though social scientists and engineers often do not understand what is a meaningful pattern in the same way.

At one point, the gulf between the two cultures is not so wide. On the computer science/engineering side, “social computing” seeks to build architectures to support social transactions: and (usually implicitly) use theories of social structures. From the social science side, the field of social network analysis has extensive experience in formally modeling and analyzing the kinds of data that are

R. A. Hanneman (✉)
Department of Sociology, University of California,
Riverside, USA
e-mail: robert.hanneman@ucr.edu

C. R. Shelton
Department of Computer Science and Engineering,
University of California, Riverside, USA
e-mail: cshelton@cs.ucr.edu

being produced by social computing, but little experience in exploiting the flood of data that has become available.

We are going to look at one small part of how social sciences (particularly, social network analysis) and social computing might inform one another. First, we will look at a very concrete example from the two perspectives. Next, we discuss some strong parallels between critical concepts of data structures and social networker's conceptual schemes. The ways that social network analysts look at social computing data and what they want to know from it are, in some instances, quite similar to some of the goals of data mining. Two particular ideas from network analysis are then explored: modality (roughly, thinking about heterogeneous classes of social objects and their relations) and equivalence (roughly, what we mean when we say that two objects are similar to one another in terms of their relational patterns). Following this, we explore some examples of how the concepts might be (or, in a few cases, have been) applied to mining social process-produced data.

2 Bibliographic data mining and the evolution of scientific communities

Relational databases of periodical literature are now a critical part of the infrastructure of doing research work. For sociologists, bases such as *Sociological Abstracts* (ProQuest 2010) and *Web of Science* (ISI 2010) are everyday tools of the trade. To the information scientist, the key issues are entry, storage, search, and reporting architectures and algorithms. To the social scientist, the database is an archive of trace evidence deposited by social actors in the process of producing "knowledge".

As a data object, a periodical literature database could be organized as a single table with a row for each new article that appears. Each row might contain a number of fields (e.g., first author, second author, journal–volume number–pages, date of publication, keywords, abstract, text body, and references). One could mine the database by specifying unions and intersections of sets of values on multiple attributes of the records to produce lists.

This description would make most database designers wince: it is an inefficient architecture that would make it difficult and slow to extract useful information. However, the "traces" left by many very important social processes are recorded in essentially this form of cumulating lists of transactions as they occur. E-mail logs, lists of searches conducted by visitors to Amazon, contributions to blogs or to virtual communities (multi-user games, open-source programming communities), sales records, and stock trades are some examples. These "data structures" are very much like the marriage registers, birth and death records, crime reports, voting roll calls, and other documentary archives

that have been mined by social scientists. Many other very important data collections about social processes are simply aggregated transactions, for example annual tables of trade flows of commodities among nations.

To make mining more efficient, databases of periodical literature actually use object-oriented and relational concepts for their organization. Rather than a single table of transactions, each with many attributes, the data are organized in relational databases composed of multiple tables, connected by indexing attributes. One might have a table of authors, one of journal titles, one of articles (which might contain the abstract, body, and references), and one of keywords. Individual authors might be linked to other authors (co-authorship), to one or more articles (authorship), which appeared in one or more journals at a particular time, with various combinations of keywords. For most bibliographic databases, articles are also indexed to other articles by way of cited author or cited article relations. This is the familiar relational database containing multiple indexed tables with a variety of one-to-many, many-to-one, and many-to-many relations among objects in the various tables.

Bibliographic data miners exploit the relationality of the objects in the data tables in a number of ways. A few examples suffice: the extent to which the articles published in one journal (over some period of time) cite articles published in other journals form a network of directed journal-to-journal citation ties. Eigenvector centrality of journals in this network is called the "impact factor" of a journal and is critical to its desirability, its value as social capital in the career attainment of scientists, its advertising rates, etc. We may trace co-authorship patterns (author–author networks that count the number of times authors have written together), co-citation (the number of times that one author cites another in their articles), the prominence of particular authors, which articles cite which others (to find critical paths and key contributions in the development of discourse), and so on. One particularly clever application of this type is a recent work by Chen (2006) that identifies "research fronts" based on bursts (and other factors) in two-mode article/key-term networks.

Now let us take the rather different perspective of a social scientist studying the development of science, who is seeking to exploit these data. To the social scientist (e.g., Collins 1998), the information in the database are "trace evidence" of an ongoing social process that produced the data. As a "thick description" or narrative analysis, the analyst sees a complex process of co-evolution involving heterogeneous social agents interacting to "construct" social reality (very sorry if that phrase causes immediate headaches).

Roughly, the process looks something like this. Individual scientists (each of whom has a history), become

interested in topics, interact with other scientists (at the same workplace, in professional associations), are influenced by the published work of others, and create a new article. They may form direct ties of working together to produce one or more articles, or work together indirectly (by citing one another). The new items produced cite previous articles, and so on. At the same time, journal editors shape the process by seeking high-impact contributions; themes and research problems evolve through combination and division. In short, it is a complex process of co-evolution in which scientists, specific articles, research problem areas, venues of publication, and institutions where work occurs all shape the connections of the “web of science” as it changes over time.

Traditional “history of science” treats the process as an unfolding narrative of individuals, events, places, and texts co-determining and influencing one another. Social science approaches to the same types of data attempt to find patterns and commonalities in repeated similar causal chains by identifying types of individuals, events, places, and texts that frequently co-evolve in similar ways.

The perspective of the information scientist and the social scientist in looking at the same bibliographic data would seem to be very different. But there are some fundamental ideas in common.

3 Shared concepts

Many sociologists (particularly, social network analysts) might describe their perspectives as “object” or “agent” oriented, and focused on “relational structures”. Perhaps somewhat surprisingly, they appreciate these terms in the same general way as computer scientists, though both groups have elaborations of the basic ideas that go in somewhat different directions.

For the sociologist, the “particles” that make up the relational structures they study can easily be seen as objects in very much the same sense intended by object-oriented programming:

“Object-oriented programming (OOP) is a programming paradigm that uses “objects” – data structures consisting of data fields and methods together with their interactions – to design applications and computer programs. Programming techniques may include features such as data abstraction, encapsulation, modularity, polymorphism, and inheritance.” (Wikipedia 2010)

The most obvious kind of a “social object,” of course, is an individual human being. Persons have social identities described by attributes (e.g., race, sex). Persons also have what social scientists are wont to call “agency”, which is

strongly analogous to the OOP notion of “methods.” That is, persons have capacities to initiate behavior, and particularly behavior that creates, modifies, or deletes relations to other objects in the object class of persons, and to objects in other classes.

When thinking systematically about social structure as composed of objects and relations, sociologists usually recognize some classes of “social” objects that are not people. Rather uncontroversial are the notions that “events” and “organizations” are social things with attributes and agency. “Events” are interactions that have their own emergent attributes and are recognized by the actors (named, having shared meanings); for example, a research article might be thought of as an “event”. The article has attributes (length, topic, co-authors, citations, etc.), a name in itself, and a “social life” of its own that is not reducible to the attributes of the agent(s) that produced it. “Organizations” (couples, families, small informal groups, large formal organizations, whole nations, etc.) are also recognized as socially meaningful and have attributes and methods that are unique to their class.

More controversial, but regarded by many sociologists as very useful, is the idea of treating cultural objects as social objects. Identities, categories, and symbols (e.g., “engineer” or “American flag”) are shared meanings that have attributes. Sometime cultural objects may also be thought of as having “methods” in that they have logical and/or scripted relations to other cultural objects (“black” evokes the superclass “minority”).

Sociologists often name what they study as “social structure” or “patterns of social relations.” Again, there is a strong analogy between the social science use of “relations” and the sense of the term when it is used to describe databases as structures of objects connected by indexing attributes or methods. Social objects (i.e., people, events, organizations, cultural objects) are classes, and the patterns of relations among elements of a class, or between elements of different classes, are “social structure”. The most explicit statement of this view of social structure is in social network analysis, where a social network is a set of social actors and relations connecting them.

The complexity of the social sciences lies primarily in the kinds of relations that are seen as connecting social actors. There is certainly no consensus within or between social sciences on classifying types of social relations. Social network analysis identifies two very abstract classes of relations: directed and “bonded”. Directed relations or ties between two social actors indicate the conserved flow of some quantity from one to the other. A husband may direct money to a wife (and/or vice versa). “Bonded” relations or ties between two actors indicate that both are equally embedded in an “emergent social fact”. A husband and a wife share the relation of “married”.

Object-oriented programming can trace its roots back to the language Simula, which was designed primarily to support discrete-event simulations. It is perhaps, therefore, not surprising that sociologists' particle view aligns well with programmers' object view. Objects can be viewed as a codification of procedural knowledge about a simulation. They define precisely what can be done (and how) to or with an entity. Much as we note that some sociologists treat more abstract concepts (like a profession) as an object with relationships, computer programmers also came to embrace object-oriented design beyond the straightforward simulation interpretation. Object-oriented programming dominates modern software design as a way to organize and simplify the design of algorithm and data structures even when the base "objects" are not as tangible as physical people, machines, or places.

The relationships that exist in object-oriented programming are more flexible and less structured than those of relational databases. In the former, relationships (methods) are arbitrary algorithms that can query, modify, or create one or more objects based on one or more other objects. These are used (in a simulation-type program) to enact the series of interactions between objects that constitute the simulation. In the latter (relational databases), relationships are links between the attributes of two or more tables. The relationships link disparate data about the same entity. In our context, object-oriented relationships model simulation narratives, while database relationships model social graphs. Put differently, the object view of the world describes how events transpire (how the social network is created). The database view of the world describes the evidence trail left by those events (how to describe the end social network result). In statistics, a similar dichotomy exists between generative and discriminative models for data.

There is a great deal that social scientists, and particularly social network analysts, could learn from serious conversations with information scientists about the nature of "objects" and "relational data structures". But the two fields do have a great deal in common at a very basic level. Both work with "structures" that are composed of "relations" (which have attributes) among "objects" (which have attributes).

The design and mining of relational data structures that are used to capture transactions of social processes are often approached by information scientists without thinking explicitly about the "social structures" that produce the data. Social scientists think quite a lot about the processes of social structures that produce "data", but often fail to think about social structure as a data structure. Information scientists are comfortable with data structures and algorithms to extract information from them. Beyond searches and lists, though, what might we want to know about process-produced social computing data?

4 A social network analysis approach to relational object data structures

The social networks perspective sees "social structure" as patterns of relations among social actors. These patterns are represented as graphs or directed graphs with nodes as social actors (who may have "color" spectra representing their attributes) and edges or arcs representing relations. Formal graphs have unambiguous translations into matrix representations. The "mining" or analysis of social network data consists of operations on these matrices to identify features of the graphs that are of theoretical interest, such as the "centrality" of nodes, graph "centralization", the distribution of shortest paths among objects, or partitions of nodes into classes based on similarities in their relational structures.

The notions of "modes" in social network analysis, and the kinds of relations they imply, are the basic conceptual tools that social network analysts use to think about how to organize complex relational data structures. There are many and varied tools for summarizing the patterns in the data (e.g., Wasserman and Faust 1994; Hanneman and Riddle 2005; Scott 1991). For current purposes, we are going to focus on the problem of identifying (or testing hypotheses about) partitions of the data based on relational equivalence of social actors.

4.1 Modality and kinds of relations

A large part of social network analysis focuses on the very simple data structure of a single relation connecting all elements of a class of social agents to other members of the same class. One can imagine a matrix of scientists by scientists, with elements containing the count of the number of articles on which they were co-authors. Structures that connect elements in a class to elements in the same class are labeled "one-mode" structures. In our example, scientists could be connected to scientists in multiple (multiplex) relations such as "friendship", "co-authorship", and "co-citers" "located at the same institution". Articles could be connected to articles in one or more single-mode relations (one article cites another, two articles share authors, two articles appear in the same journal, etc.) Similarly, other classes of social actors could be connected in single-mode relations (institutions to institutions, journals to journals, etc.).

Another data structure maps (one or more) relations between social agents of different types. The "two-mode" structure (e.g., scientists by articles, mapping who authored which) is rectangular. Two-mode data structures are also frequently called "co-occurrence", "actor-event", or "affiliation" matrices. For some examples: authors are located at particular institutions; articles appear in

particular journals; articles contain particular keywords. The principle can easily be extended to imagine “multi-mode” data structures that relate all modes simultaneously. These multi-mode structures consist of collections of rectangular matrices that may be processed simultaneously.

A third common type of data structure, an “attribute” matrix, maps variables or attributes to the social agents in a class, giving the nodes “color”. We might show the relation between scientists and the attributes of gender, ethnicity, number of prior publications, institution of employment, etc. In a multi-modal social network, there could be a separate attribute matrix for each mode (scientists have attributes, journals have attributes, institutions have attributes, articles have attributes, etc.).

It can be argued that “color” should always be represented as affiliation, rather than as an “attribute”. A person’s gender, for example, is really an “affiliation” of a person with a cultural category or symbolic object and not something that is unique and wholly nested within that individual. As a practical matter, it is often more insightful and useful to “color” nodes by attributes and use attributes as partitions. At a deeper level, though, many attributes of actors are actually better thought of as “affiliations” of actors with cultural symbols or meanings (rather than unobservable unique internal states). When the goal of analysis is to find equivalence classes, as discussed below, it is often better to treat “attributes” of nodes as “affiliations” between two modes.

Relations in a single-mode matrix may be symmetric (represented as a simple graph with edges) or asymmetric (represented as a directed graph with arc). For example, the count of co-authorships between pairs of scientists is necessarily symmetric; the citation of articles by articles is necessarily asymmetric (though there may be reciprocal co-citation). Social action, however, is initiated by an individual and directed toward another. Thinking about social process suggests that one-mode social relations are best seen as directed and asymmetric. Symmetric relations among the elements of a mode of social actors can almost always be seen as induced from an affiliation matrix. For example, co-authorship ties between scientists might be thought of as induced by affiliation of each scientist with the same object in another mode (the article class).

Relations between two modes are, by definition, asymmetric. Each row (e.g., scientist) is affiliated with one or more columns (the affiliation may be binary, multi-valued, ordinal, or a measure of “strength” of the affiliation). The data produced by social processes then can be represented as some number of rectangular arrays of directed relations between the elements of each mode, and between the elements of each pair of modes. The arrays are linked by the indexes of the elements of each mode. The resulting data structures can be thought of as N -dimensional matrices, or

hypergraphs representing relations among several modes simultaneously.

Having structured the information, what data do we want to extract from it?

4.2 Mining social process-produced data: equivalence

In querying a database, we are locating data objects that satisfy (or are similar to) as a set of criteria: “Show me all the books by Joseph Conrad, and are currently in print in paperback”. It is easy to see such a query as asking about the attributions of a single mode of objects (books, in this case).

If we think about databases as relational structures or networks, however, the query might be understood a bit differently: “show me all book objects that have the relation “written by” to objects in the class “authors” with the attribute “Joseph Conrad”, AND have the relation “true” to the object in the class “publication statuses” with the value “in print”. We might imagine a three-way data array of authors by books by publication statuses, and ask to see the index values of all columns in the “books” dimension for the “row” “Joseph Conrad” in the author dimension AND the row “in print” in the publication status slice (that is, a specific value in the mode author; a specific value in the mode publication status; and any non-zero value in the mode book).

Making sense of complex relational data left by social processes can be seen as finding objects that are similar to some prior hypothesis about relational equivalence (in a confirmatory analysis) or similar to one another (in an exploratory analysis). The book “Lord Jim” and the book “Nostromo” are “similar”, in relational terms, because they are elements of the mode “book” that have an “authored by” tie to the element “Joseph Conrad” in the mode “authors”.

Sometimes, “mining” databases is an exercise in finding a particular object. More commonly, the goal is to identify sets of objects that are similar. Sets of people who have similar patterns of relationships to other people, organizations, and cultural symbols are actual or potential “communities” or “market segments”. Identifying the demography of such communities and the relational characteristics that define them may be critical in reaching or influencing them. From a social scientific point of view, theory consists of understanding and explaining classes of social objects, and not individuals. The identification and delineation of classes of “similar” objects is critical to making or testing theory.

But, what do we mean by “similar?” Social network analysts have given a good deal of thought to what it means for two social actors to be “similar” or “equivalent” in relational terms (Everett 1994). Here, we will focus on the

two most widely used definitions of relational similarity: structural and regular equivalence.

Structural equivalence was first explicitly defined by Lorrain and White (1971) and is described in Batagelj et al. (2004) as: “Units are structurally equivalent if they are connected to the rest of the network in identical ways.” Put even more simply, two nodes are structurally equivalent if they have exactly the same pattern of ties to all other nodes. Structural equivalence is the strongest form of equivalence: exact equality in the pattern of relational ties. In practice, approximate structural equivalence is often used. There are numerous commonly used measures of approximate structural equivalence: correlation, Hamming distance, Euclidean distance, etc.

Almost all queries and methods of pattern finding (components analysis, cluster analysis, MDS, correspondence analysis) use some algorithm to locate dimensions, clusters, or classes of structurally equivalent nodes in graphs. In doing so, we are locating “substitutable” or “identical” nodes on the basis of their patterns of ties with other nodes. Almost all data mining, whether based on relational or attribute approaches, has used structural equivalence. Despite this, regular equivalence may be a more useful definition of relational similarity for studying social computing and data archives.

The first formal statement of relational regular equivalence is usually attributed to White and Reitz (1983). Regular equivalence, described in Batagelj et al. (2004) as “...two units are regularly equivalent if they are equally connected to equivalent others”. The core idea is also sometimes understood with regard to the mathematics of coloring graphs. In graph coloring (Chung 1997), two nodes in a graph are regularly equivalent (have the same color) if they have the same spectra (have at least one relation with an element of each the same set of other classes).

In social network theory, the idea of regular equivalence is tied to the notion of a social role. Consider a table that shows a list of adult women as rows, and minor children as columns. A cell contains a 1 if a particular child is the offspring of a particular parent, and zero otherwise. Using structural equivalence, no reduction of the rows is possible, as each mother has a unique set of specific children; reduction of the columns is possible, however, by grouping together the multiple children of a particular mother. Viewing the same data from the perspective of regular equivalence produces a different result. In this case, the adult women may be partitioned into two groups: those who have children, and those who do not. The minor children cannot be partitioned: each child has a relational tie to a member to the class of adult women who have children, and none has any tie to any of the adult women without children.

Regular equivalence is a “more relaxed” idea of similarity between nodes than is structural equivalence. In many cases, the goal of pattern finding and data mining is actually to find partitions that are regularly equivalent, not structurally equivalent. Regular equivalence is used to identify classes of actors who have similar “roles”. That is, they have similar patterns of ties to similar others. When we identify words or phrases as “equivalent” in the coding dictionary of content analysis, we are using regular equivalence; when we identify nations as “semi-peripheral” in the world system, we are using regular equivalence. Most social science theory is stated in terms of actors who are regularly equivalent (e.g., “elite”, “parent”). In mining the archives of social transactions or social media, our interest is often on finding groups of actors who are regularly equivalent: it may be more useful to identify all purchasers who bought any (or all) books by Joseph Conrad, rather than those who purchased a particular title.

Algorithms and methods for testing hypotheses or identifying regularly equivalent partitions in relational data are not as highly developed as those of structural equivalence. Probably, the most commonly used approach is “block modeling” (Breiger et al. 1975). In block modeling, the rows, columns, and slices of multi-modal graphs are permuted to locate blocks of cells that contain particular patterns of ties. One very useful example of the major types of blocks (or types of equivalence) is given by Doreian et al. (1994). See (Fig. 1).

The power of generalized block modeling in two modes can be illustrated rather simply. In the “core-periphery” view of economic relations in the world system, “core” nations export heavily to all other core nations. This would be a “complete” block of ties. “Peripheral” nations do not export to one another. This would be a “null” block of ties. Core nations each export to a sub-set of peripheral nations that fall within its sphere of influence, but not to all peripheral nations, generating a regular equivalence block. Peripheral nations export to some, but not all core nations, generating another regular equivalence block. We would seek the best-fitting partition of the asymmetric trade data that had a zero block and a complete block on one diagonal and regular equivalence blocks on the off-diagonal.

5 Social network analysis of multi-mode relational object data

The information produced by social processes can be structured into multi-mode relational data. In these data structures, the goals of mining, generally, are identifying sets of cases in each mode that are equivalent (in either the structural or regular sense) with respect to the cases in each other mode.

Until fairly recently, social network analysts usually worked with multi-mode data by analyzing it one mode at a time. There can be great power in this approach.

Suppose that we were “mining” a data set of e-mail messages, and examining only the two modes of “sender” and “receiver.” A rectangular array of senders and receivers is constructed (which would contain many, but not necessarily all of the same agents), and the presence/absence or number of messages in each dyad would be constructed. We could induce a matrix of the senders who were similar to other senders by counting the number (or volumes) of messages they sent to the same receivers. We could also induce a matrix of similarities among the senders by indexing the extent to which they received messages (or message volumes) from the same senders. Each of these “one-mode” square arrays could be thought of as a bonded (simple, undirected) graph. Conventional network techniques could be used to identify central actors and graph sub-structures (e.g., the “modular” community approach of Newman 2006). Senders or receivers could be classified into groups or clusters based on similarities in the specific others to whom they directed messages, or to which other “types” of senders (or receivers) they were tied to. That is, the senders can be classified into either positions (structurally equivalent nodes) or roles (regularly equivalent nodes).

A great deal of interesting and useful information can be extracted by transforming the relational data for all pairs of modes into single-mode similarities. We can find senders who are similar in terms of the receivers that they send to; we can find receivers who are similar in terms of who is sending them messages. In each of these analyses, though, we are implicitly treating one mode as “independent” and the other as “dependent.” The process we are describing, however, is co-evolutionary, with both sending and receiving being dependent. A two-mode analysis would be more appropriate.

To date, there are two main approaches to two (and multi) mode relational data. One approach is to apply the technique “correspondence analysis”, “singular value decomposition”, “multi-modal factoring” type (Faust 2005). These approaches partition the total pooled variance (e.g., variance across senders in their profile of receivers along with variance across receivers in their profiles of senders). The result is a dimensional decomposition of the variance that can be used to scale both modes simultaneously, and can be used to identify clusters of senders and receivers who are “close” to one another. These are extremely useful outcomes (some examples are given below). Unfortunately, only structural equivalences can be considered, at least in existing software.

The alternative approach is generalized block modeling (Doreian et al. 2004). Senders would be classified into partitions based on their profiles of ties to partitions of receivers, and vice versa. For example, we might identify a partition of message senders who directed communications

at all others (spammers), partitions that communicated only with members of their own group, a partition of receivers who did not send, and so on. We might have a prior hypothesis about the number of sending and receiving partitions and the kinds of equivalences that described their relations; or we might explore the data for best-fitting partitions and equivalences. The generalized block-modeling approach provides the greatest fidelity to modeling processes among heterogeneous modes of social actors. Unfortunately, existing software is very limited (two modes, small numbers of cases in each mode).

In the next sections, we will provide some examples and some speculations about ways in which casting problems as multi-modal relational networks has been and/or may be of use in understanding data produced by ongoing social processes.

6 Illustrations of modality and equivalence in social process-produced data

Any set of social processes that produce documentation (preferably time stamped!) in the form of transaction records could be treated as a relational data structure and analyzed using network analytic tools. A good deal of such work has been done, and we are not attempting a survey here. Because of both conceptual and software limitations, we have yet to take full advantage of the approach. A few illustrations will serve to highlight some of potentials and current limitations.

6.1 Bibliographic databases

In his survey article on scientific networks, White (2011) demonstrates that the multi-mode, co-evolutionary perspective is becoming the dominant approach in scientific studies of the structure and performance (e.g., quality, quantity, creativity, and breakthrough innovations) of knowledge communities.

Automated bibliographic databases, built from data scraped from the web, are also available. Citeseer and Google Scholar are two such examples. These represent an interesting bridge between the object (or particle or narrative) view of networks and the relational (or graph) view. These databases are built automatically using algorithms that reason about the possible objects to produce a relational database. As a more concrete example, the software wishes to link references in papers to the papers they reference. However, misspellings, differences in citation formats, and omitted information make such a matching ambiguous. To resolve the ambiguity, the algorithm must reason about different possible explanations (sequences of events describing how the data, the text of the papers

scanned, might have been generated) to find the one that is most probable (see Pasula et al. 2003, for example). The result is a hypothesized relational database connecting papers to their citations (a one-mode two-way matrix). In doing so, the algorithm has reasoned about “higher level” structures that might have led to the paper.

The volume of information that is available in digital form in bibliographic databases is quite stunning and growing very rapidly. One popular resource for literature in biomedicine, popularly known as “Medline” (National Institutes of Health 2010), currently contains about 19 million citations from a broad range of periodical literature in bio-medical fields. Each record contains authors, titles, abstracts, many full texts, keywords, venue of publication, date of publication, and other standard fields. A collaborator of the authors of this paper has developed software to mine records for additional data (such as the institutional affiliation of authors). Content analysis and perhaps even sentiment mining tools could be developed for characterizing texts.

A number of the fields in these data records are very reasonably conceptualized as modes of social actors. Authors and articles are obvious, but important: author–author ties by direct collaboration or citation are staples. When these affiliation networks are examined through time, the rise and fall of article impact, author status, critical paths, and community structure (e.g., how does the size of the giant component evolve?) can be described. Many such analyses exist, though they explore only very small parts of the available data and rely entirely on structural, rather than regular, equivalence notions.

Still to be explored are the effects of other active social agents. Journals and their editors play active roles in shaping the development of fields. Institutions (universities, laboratories, etc.) affect the likelihood of collaboration. Topics (keywords) are combined and re-combined to elaborate existing specialties and stake claims to new leading edges. Emerging empirical work is exploring some of these less traveled paths and is finding evidence of very complex co-evolutionary dynamics.

Structural equivalence analyses of such multi-modal data would yield particular combinations of authors/venues/keywords/articles that are at particular locations in graphs (high closeness centrality, high betweenness centrality). Regular equivalence analyses would seek to identify parallel and similar structures in, perhaps, varying scientific fields or historical contexts.

6.2 Text and narrative mining: integrating content analysis with network analysis

The method of content analysis is to create classes of objects (text strings) that have some form of relation with

other objects (text strings) and study the pattern for the resulting semantic network. The most obvious and oldest approach is to treat words as objects, and to count the number of times they appear within a defined distance from one another in a text as undirected tie strength: simple co-occurrence of words is using the notion of structural equivalence. Generally, however, content analysis seeks to create or identify regular equivalence classes. For example, a tie exists if any of the words in the set {pony, horse, pinto...} are within a given distance of any of the terms {ride, mount...}.

Commonly, equivalence is imposed by the analyst based on conceptual schema and deep knowledge of the problem. The validity of results, however, depends on the coders and consensus about the dictionary. And, until the dictionary is developed, content analysis of text is slow, somewhat unreliable, and expensive. Processing large volumes of text traffic in anything resembling real time remains a major challenge if such a schema is to be generated and applied by humans.

Mining large volumes of texts and multiple coding of the same text to create databases of equivalences is one approach. Google’s efforts in developing language translators by building equivalences from multiple translations of the same text and direct comparisons of web contents (e.g., the same content posted in a Web site in German and English) is one feasible approach based on structural equivalence. Alternatively, it might be possible to apply algorithms for identifying approximate regular equivalence classes. Regular equivalence reductions would not yield good textual translations; they would, however, be rather more useful for uncovering meanings and implications of text. Given the huge and rapidly growing volume of text content available in digital form, the development of content-analyzing engines is a major growth area (dictionaries, natural language recognition, neural networks, etc.).

Now, consider some complexities. Rather than a single text, suppose that we were working with multiple texts, or considering parts of a text produced by different actors or texts produced by different actors. Perhaps the texts are “directed”: for example, in a conversation, thread in a discussion board, or e-mail stream. Perhaps, and usually, the texts are temporally ordered.

Imagine if we could define a multi-modal data structure of class (words) by class (words) produced by class (actor) directed to class (actor), at class (time). We can now, potentially, partition the total joint variance, or propose and fit equivalence block models to the entire structure. Why would one? Word prevalence and word adjacency may well be contingent depending on the sending and receiving actors, and may vary systematically as the discourse develops. When texts are examined in an attempt to identify unknown authors or their attributes (the writer was

raised in the southern USA, for example), multi-modal mining occurs.

The same kinds of notions of treating parts of texts as objects and examining them relationally have been applied to whole narratives. Beginning, perhaps, with the work of Heise (1989) and Corsaro and Heise (1990), narratives are treated as series of “events” (each of which has affiliated sources, targets, and other attributes), which are ordered by the relations of logically necessary and sufficient conditions for the occurrence of other events. Mining the structure of narratives, identifying logical peculiarities, and comparing accounts of the same events by different actors in historical research have generated a (very limited number of) quite interesting results (Griffin 1993).

Formal analysis of narratives (and the related study of event sequences) has not been cast in network-relational terms. Heise’s “event” objects, however, can easily be seen as one mode in a relational structure with which authors and targets are affiliated. The structure of narratives as event sequences themselves can be cast as networks and mined for structural and regular equivalences, which would identify characteristic sequences that might vary by author or other affiliated traits.

6.3 Cognitive social structure

An early, but still very useful, application of multi-mode analysis is that of “cognitive social structure” (Krackhardt 1987). Data of this type consist of collecting information about the relational structure of a number of objects, as understood by a number of perceivers. For example, the patterns of which persons “liked” which other persons might be reported by each person in a group. Which products or ideas were seen as “similar” (and viewed with positive or negative sentiment) by consumer raters are logically parallel data. The data are three-mode: source of a “liking” relation; target of a “liking” relation; and the rater.

It is possible to examine which raters are similar to which others in terms of the similarities of the “maps” they draw of who likes whom. One could evaluate which actors were similar as sources of liking, based on the profiles of their targets, or (alternatively) based on the degree of similarity in the ratings of this by raters. Identifying how social actors view the connections among social objects and identifying types of persons who “construct” different mental maps is an important problem. From a practical point of view, identifying distinctive communities of people who share similar cognitive maps and understanding those maps can much more carefully target appeals and actions. The identification and understanding of the dynamics of group identity formation and patterns of similarity in “social construction” are core theoretical problems in the academic social sciences.

In this example, the sources and targets of liking are two modes of social actors. Even though the two modes contain the same elements, they are not the same mode, because the relation of “liking” is asymmetric. The third mode also has the same index of actors, but is “ratings”. We might treat the “rating” as an “event”: an emergent symbolic, cultural characterization, or perception of social structure. This generates a network structure in which k events (where k is the index of group members) each “affiliate” sources and targets of their liking. As a structural equivalence problem, we would like to know: which actors are perceived by raters as having similar targets of their liking? Which actors are perceived by raters as being similar in terms of which actors like them? And, which perceivers have similar maps of who likes whom? One might seek a further reduction of the modes into regular equivalence categories: are there “kinds” of sources of liking relations who have different spectra across “kinds” of targets of liking, as perceived by “kinds” of perceivers? “Individual differences scaling”, three-way clustering, and multiple correspondence analysis can be applied to data of this type, if we perceive the questions of interest to be similar to structural equivalence (e.g., Arabie et al. 1987).

“Cognitive social structure” types of data have been collected and used in focus group and survey/interview research in marketing for some time, though the analysis is rarely multi-modal. In the current period, there is the possibility of applying the same logic of analysis to “sentiment mining”. If attributes of the raters are known (as in on-line rating panels), the emergent regular or structural equivalence classes of raters and perceivers can be profiled to identify social types or market segments.

6.4 Virtual communities: e-mail, blogs, WWW, Social-Networking, Net-games, Open-source Development Communities

Social processes occurring in “new media” leave logs of transactions. Because transaction records are already in forms that are fairly easily machine processed, and because the volumes of data available are huge (exceeding by orders of magnitude the volumes of old-media documents), a great deal of effort is currently directed toward their analysis. The largest part of the effort to exploit data sources of this type so far has been by information scientists and researchers in complex network dynamics. A large part of this work has treated the data as networks, and has applied network analysis tools (often from engineering and physics, more than social sciences). Much of the effort has focused on problems of search, robustness, and other aspects of network topology. Some work has been done on more traditionally sociological topics such as identifying communities, core-periphery structures, central nodes, and

the like. For instance, Zheleva et al. (2009) describes a co-evolution model for explaining links and groups in data from Flickr. The two key concepts that we have explored at some length in this paper, modality and equivalence, have not yet been extensively applied in new media studies. Here are a few ideas of how they might be.

E-mail, phone, discussion board, texting, and blog transactions are routinely archived in digital form. These “traces” reveal the structure (and often the content, as well) of very large volumes of one-to-one and one-to-many communications among social actors. Typically, the records exist in transactional form with fields that record information which can be treated as multiple modes and attributes.

An e-mail object, for example, has a source, one or more destinations (of various types), a subject line, often some indication that it is part of a thread (RE, FWD), and a text (it may also contain embedded links). The time and location from which it was sent, the path it followed, and the content and attachments are often available. Senders and receivers are “affiliated” with messages (and form either regular or structural equivalence classes). The text mining of subject lines and/or message texts can produce regular or structural partitions that are symbolic/cultural contexts within which senders, receivers, and particular sets of messages are affiliated (again, either regularly, structurally, or both). The time stamping of traffic could, in principle, allow the characterization and analysis of the “shape” of the multi-modal space and the characterizing of the trajectories (direction and speed) of topics, sources, and senders.

Social networking sites, URL–URL linking in the WWW, logs of games, blogging, and open-source software development communities are some examples of virtual communities that are self-selecting affiliation structures, logically parallel to “voluntary associations” in traditional social science studies (Cress et al. 1997). Virtual communities are many-to-many structures that are created by affiliation and have a bi-partite network structure. They may also embed direct connections between individual agents and direct connections and or affiliations among event/symbolic/organizational social agents. Some work has been developed in this space, using network approaches. The notions of modes and equivalences may provide some interesting new directions. Two very brief speculations are as follows.

Open-source software development processes and communities have been studied (primarily by computer scientists), in part because of the large quantity of high quality data produced by the documentation of such collaborations (e.g., Sourceforge 2010). Participants affiliate with one or more projects, taking roles in creating, revising, and assembling components of software programs.

Within a project, actors affiliate with one or more components, which are themselves “affiliated” with other components (code segments depend on other code segments). The entire structures evolve over time, driven by the internal logic of the task, but also by the social logics of leadership, status seeking, cooperation, and altruism. Structural analysis enables us to understand particular projects; regular analysis could provide more general insight into the commonalities of successful and failed communities.

Multiple user interactive games such as Warcraft (2010) are hosted on servers that log all transactions. Some of the communities are huge (millions of participants), and the transaction logs are almost incomprehensibly large. These communities (like social networking sites) are of interest both in themselves as new social phenomena, and because they are naturally occurring experiments in network dynamics, exchange dynamics, and other structured interactions. Games are particularly intriguing because human participants may construct one (or more) identities, and create and affiliate/disaffiliate with both long-term (e.g., “kingdom”) and short-term (e.g., “quest”) symbols and organizations. The symbolic and organizational classes evolve by both affiliation and by selection dynamics within their own mode (two “quests” may join forces to fight a battle). The structural equivalences of these multi-mode structures may be important for the information they provide us about evolving network topologies. Regular equivalence structures might tell us something deeper about more general patterns and dynamics by which communities and more complex social structures are constructed and de-constructed.

6.5 Policy networks and politics

The relational network perspective has, particularly recently, been advancing rapidly in political science and political sociology. It is quite easy to see records of political acts (e.g., voting, making donations) from a relational network perspective. The votes of citizens for candidates and initiatives accumulate over election cycles; the votes of legislators on bills accumulate over a session; the votes of justices on courts accumulate over time. More generally, any set of recorded preferences by multiple actors toward multiple objects might be treated as an affiliation structure.

Traditional approaches to such archival data are to use attributes of actors (donors, voters, legislators, judges) to predict their orientation toward (or affiliation with) particular outcomes (candidates, bills, court cases). Increasingly, however, political analysts have become more sensitive to the non-independence of these events across actors and time. The relational perspective can provide

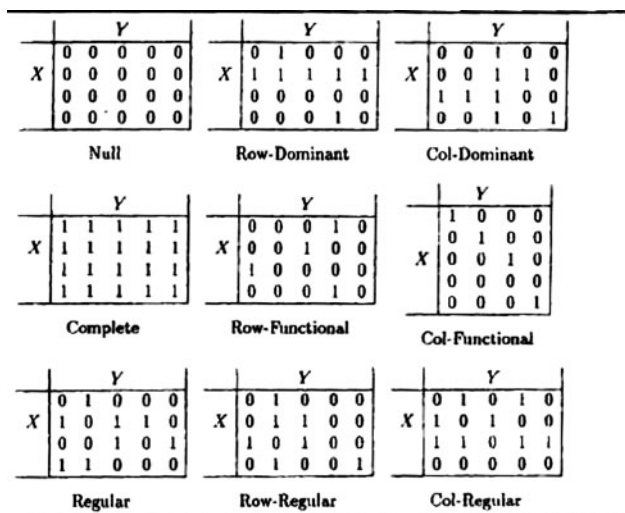


Fig. 1 Relational blocks in generalized block modeling. Source: Doreian et al. (1994), p. 6

some different insights to such complex processes than conventional statistical approaches.

Bowler and Hanneman (2006) examined the data archive collected by the Secretary of State of California on donor’s contributions for and against 59 ballot initiatives over the period 2000–2004. Donors and initiatives can be cast as two modes of social actors that are brought together into a co-evolving relationship by the act of donations. Donors who support the same sets of initiatives come to perceive themselves as a “community” or “social movement”. The initiatives that are supported by coherent sets of constituents are perceived to be part of larger policy

issues or ideologies. Past collaboration may breed future cooperation among donors; as multiple initiatives become seen as part of an agenda (e.g., California’s “Proposition 13” and the “tax revolt”), they may spawn new initiatives. Figure 2 shows a mapping of major donors (those who gave more than \$1 M US to more than one initiative), and a mapping of the initiatives in the joint “policy space”.

The analysis suggests both dimensionality (the authors interpret the dimensions as liberal/conservative and statist/anti-statist), and clustering (e.g., labor unions and Democratic political groups often co-donate). This analysis pools across several election cycles. Treating each election as an additional “mode”, a contextual event, might improve the analysis and would be able to show how the policy space and the donor space co-evolved. The analysis also relies on structural equivalence (donors are similar to the extent that they had the same profile of ties across 59 initiatives; the initiatives are similar to the degree that they were supported by exactly the same individual donors). Greater insight might be possible by seeking patterns of regular equivalence: “types” of donors based on similarities in their profiles of support for “types” of initiatives, and vice versa.

Many political (and other) data showing the orientations of actors toward ideas or objects are increasingly available in digital form, though the mining is often a challenge. Roll calls and court decisions are two such traditional data that could be approached with multi-mode equivalence analysis. Attitude surveys, blog content, web-page text, and other media could also be mined to develop co-evolutionary portraits of constituencies and meanings.

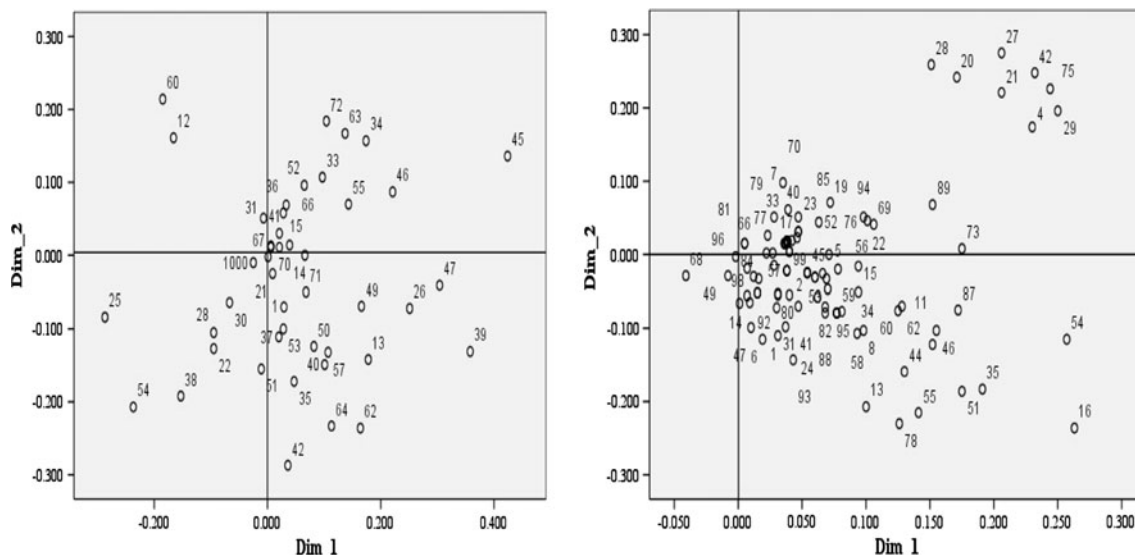


Fig. 2 California ballot initiatives (left) and major multi-campaign donors (right) in joint space

6.6 Business directories

Business directory databases and annotated mailing and e-mail lists are produced by social agents who map the economic and occupational social space. Organizational ecology attempts to apply principles and theories of human ecology (Hawley 1986) to populations of usually formal, but sometimes voluntary, organizations. Carroll and Hannan (2000) are perhaps the key figures with regard to formal organizational ecologies. Cress et al. (1997) are the leading figures with regard to organizational ecologies of voluntary associations.

The core idea here is that organizations that perform particular specialized functions locate in non-random ways in human settlements. One reason that they make non-random choices is the presence of other functions in particular places. Places also have independent attributes that make them differentially attractive for different organizational functions (e.g., they are located on a river).

One supposes that one may identify patterns of organizational density that define “types” of organizational communities. One may also identify “types” of places that select for varying mixes of organizational types. That is, populations of organizations and populations of settlements “co-evolve” by the processes of affiliation (birth, death, change in function, migration). The analogy to biological ecology is extremely strong, so the notions of modality and equivalence could easily be applied to biological and ecological co-evolutionary processes.

Business directories and listings are routinely produced as adjunct to processes of marketing (e.g., finding all the dentists in Omaha, if you have a new dental instrument to sell). Directories such as Reference USA (2008) list about 13 million enterprises and give a number of attributes to them (primary product, approximate sales volume, location, etc.). The data are stored in a relational form, with some classes or primary keys: location, primary product.

Individual establishments are affiliated with particular communities and also with the social construct of a “primary product” or “industry”. Individual establishments have attributes (size, ownership form, etc.) that may be used as partitions or colors. Communities, as well, have attributes (e.g., connection to logistics networks, population size, political centrality) that may partition them and shape affiliation processes. Industries have theoretically important attributes (e.g., capital cost barriers to entry, scope of market, location in commodity chains) that may color their dynamics.

One theory of organizational and community ecology hypothesizes a “central place” hierarchy, in which (possibly following a power-law distribution) both functions and places form nested hierarchies. Figure 3 shows a display of this joint hierarchy for the state of New Mexico in 2004.

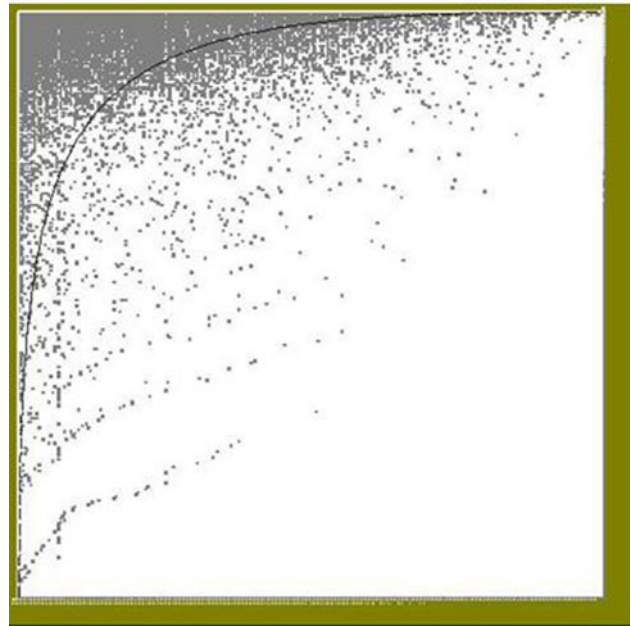


Fig. 3 Organizational/community nestedness in New Mexico, 2004. Source: data from Reference USA; analysis by the author using NestCalc (Atmar and Patterson 1993)

It is clear from the nestedness diagram that a simple scale-free network does not apply to the joint affiliation. A structural block model might do at least as well in fitting the data. More interesting, however, would be a regular reduction of the data: are their substitutable sets of organizations present in varying combinations across multiple, qualitatively different types of communities?

6.7 Trade dynamics in world systems

Patterns in volumes of trade in commodities among national economies are of interest for a number of national strategic, economic, and trade policy, and social science theoretical reasons.

The data are stored in a relational database compiled by the International Monetary Fund (IMF 2010), from national government’s reports and surveys. The data describe (aggregated by year) the volume of flow from each of a large number of nations, to each of a large number of nations, of each of a large number of commodities. The basic relational structure is a four-dimensional many-to-many relation: each sending nation may send volumes of many commodities to many receiving nations at many points in time. The modes here are exporters and importers, and they are connected by a crossed relation of commodities and time.

From a blocking or clustering perspective, we are interested in identifying (or modeling) sets of exporting nations that are regularly equivalent with regard to

importing nations and exporting nations that are regularly equivalent with regard to exporting nations. We also want to know what commodities are regularly equivalent. We might hypothesize that some nations at some points in time are producers and consumers of high-tech goods, for example, and we are interested in seeing whether nations change roles as importers and exporters of various types of commodities over time. The analysis then involves the trajectories of equivalence classes of importers, exporters, and commodities in time.

An exemplary analysis of trade tables very similar to this description was accomplished by Smith and White (1992). Proceeding from a world systems perspective, they sought to identify blocks of nations (at each of three points in time, separately), which were regularly equivalent in exporting 15 commodities (chosen as indicators of core commodities from a prior factor analysis of a large number of commodities that identified five dimensions of commodity flows). Smith and White proceeded by producing measures of regular equivalence for pairs of nations across the 15 commodity flow tables simultaneously, and then used block modeling to identify five blocks. The dynamic dimension was studied by looking at the mobility of individual nations from one regular block to another between time 1 and 2, and between time 2 and 3.

The approach of Smith and White identifies two modes (exporters and importers) as social actors, and sees them as having 15 relations at each of three points in time. We could just as easily treat this as 45 relations. A fully simultaneous blocking of the data would allow that blocks of exporters might have different members than blocks of importers, that the 15 commodities could be blocked into a smaller set of classes, and that these blockings might change over the three time periods.

The notion of treating exporters (or originators of transactions) and importers (receivers of transactions) as separate modes, allowing that the variance of each mode might contribute different amounts to the total variance, has recently been pursued by Boyd et al. (2010). These authors, rather than seeking regularly equivalent blocking (as did Smith and White), fit a core-periphery model (a blocking with high density of ties among members of the core, low tie density among members of the periphery, and agnostic about ties in the off-diagonal blocks), using structural equivalence.

Many economic transactions are now documented in digital form in real time (e.g., consumer purchases, stock market trades). The social scientist's approach of seeking patterns of co-evolution among buyers, sellers, and commodities by applying regular and structural equivalence analysis to multi-mode affiliations could, in principle, be applied to produce entirely new understandings of the dynamics of economic behavior.

7 Conclusion

We have argued that many social processes produce records of transactions, and this is particularly true of digitally mediated interactions. Conceptualizing the social processes that produce these records as co-evolutionary processes of relation making among and within heterogeneous agent classes (modes, in network analysis terms) may be helpful in structuring the flows of information into useful data. Analytic tools for working with relational data may be powerful approaches to mining such data structures. In particular, the strategy of identifying similarities in relational patterns within and between agent classes on the basis of structural and regular equivalence can produce useful insights into complex and voluminous data.

The relationship between the social process and the resulting transaction record mirrors that of the distinction between generative and discriminative models in statistics. In some cases, a block model directly reveals structure in the underlying social process. However, often the transaction record does not record all interactions between entities. In this case, the relationship between the observed record's structure and the generative process's structure is more complex. We believe that better understanding of the implications of a social process structure on the resulting transaction record (and the reverse) to be important to fully mining the growing wealth of social data.

Many "structural" social scientists have ways of thinking about social process-produced data that are highly compatible (at a broad level) with the ways that computer and other information scientists think. The notions of "objects" containing data structures and "methods" are highly compatible with the notions of "social actors" and "ties". Both structural social scientists and information scientists tend to view phenomena as complex, co-evolving, relational processes. There is a great deal in common to build on as we approach the study of the increasingly large volumes of process-produced digital data that document more and more of social life.

Two conceptual distinctions that are of great importance in social network analysis, the notions of modes and types of equivalences, may provide bridges between the skills and expertise of computer and information scientists with data structures, and the knowledge that social scientists have of the processes that produce the data. A number of examples have been briefly explored of data produced by social transactions and how they might be (or in some cases have been) approached using modes and equivalences. The study that has been done is primitive when compared to the potentials. There is a great deal more that could be done in the domains that we have briefly explored and in many others. To exploit this potential, increasingly close collaboration between social and information sciences will be necessary.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Arabie P, Carroll JD, DeSarbo WS (1987) Three-way scaling and clustering. Sage Publications, CA
- Atmar W, Patterson B (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96:373–382
- Batagelj V, Mrvar A, Ferlioj A, Doreian P (2004) Generalized blockmodeling with Pajek. *Metodološki Zvezki* 1(2):455–467
- Bowler S, Hanneman R (2006) Just how pluralist is direct democracy? The structure of interest group participation in ballot proposition elections. *Polit Res Q* 59(4):557–568
- Boyd JP, Fitzgerald WJ, Mahutga M, Smith DA (2010) Computing continuous core/periphery structures for social relations data with MINRES/SVD. *Soc Networks* 32:125–137
- Breiger RL, Boorman SA, Arabie P (1975) An algorithm for blocking relational data, with applications to social network analysis. *J Math Psychol* 12:328–383
- Carroll GR, Hannan MT (2000) The demography of corporations and industries. Princeton University Press, Princeton
- Chen C (2006) CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inform Sci Technol* 57(3):359–377
- Chung FRK (1997) Spectral graph theory. American Mathematical Society, Washington
- Collins R (1998) The sociology of philosophies: a global theory intellectual change. Harvard University Press, Cambridge, MA
- Corsaro W, Heise D (1990) Event structure models from ethnographic data. In: Clogg C (ed) *Sociological methodology*. Basil Blackwell, Cambridge MA, pp 1–57
- Cress DM, Miller McPherson J, Rotolo T (1997) Competition and commitment in voluntary membership: the paradox of persistence and participation. *Sociol Perspect* 40(1):61–79
- Doreian P, Batagelj V, Ferligoj A (1994) Partitioning networks based on generalized concepts of equivalence. *J Math Sociol* 19(1): 1–27
- Doreian P, Batagelj V, Ferligoj A (2004) Generalized blockmodeling of two-mode network data. *Soc Networks* 26:29–53
- Everett MG (1994) Regular equivalence: general theory. *J Math Sociol* 19(1):29–52
- Faust C (2005) Using correspondence analysis for joint displays of affiliation networks. In: Carrington PJ, Scott J, Wasserman S (eds) *Models and methods in social network analysis*. Cambridge University Press, Cambridge, pp 77–97
- Griffin LJ (1993) Narrative, event-structure analysis, and causal interpretations in historical sociology. *Am J Socio* 98:1094–1133
- Hanneman RA, Riddle M (2005) Introduction to social network methods. University of California, Riverside, Riverside, CA. <http://faculty.ucr.edu/~hanneman/>
- Hawley A (1986) Human ecology: a theoretical essay. University of Chicago Press, Chicago
- Heise D (1989) Modeling event structures. *J Math Sociol* 14:139–169
- Institute for Scientific Information (ISI) (2010) ISI Web of science. Thompson Reuters
- International Monetary Fund (2010) Direction of trade statistics. Publications Services, I.M.F., Washington, DC (annual)
- Krackhardt D (1987) Cognitive social structures. *Soc Networks* 9:104–134
- Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. *J Math Sociol* 1:49–80
- Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103(23):8577–8582
- Pasula H, Marth B, Milch B, Russell S, Shpitset I (2003) Identity uncertainty and citation matching. *Adv Neural Inform Process* 15:1401–1408
- ProQuest (2010) CSA sociological abstracts. Ann Arbor, MI
- Reference USA (2008) Reference USA. US Businesses Database. <http://www.referenceusa.com/>
- Scott J (1991) *Social network analysis: a handbook*. Sage, London
- Smith DA, White DR (1992) Structure and dynamics of the global economy: network analysis of international trade 1965–1980. *Soc Forces* 70(4):857–893
- Social Network Analysis and Mining (2010) *J Web*. <http://www.springer.com/springerwiennewyork/computer+science/journal/13278>
- Sourceforge (2010) Find and develop open-source software. <http://Sourceforge.net/>
- United States. National Institutes of Health and United States National Library of Medicine (2010) PubMed Home. <http://www.ncbi.nlm.nih.gov/pubmed>
- Warcraft (2010) World of Warcraft Community Site. <http://www.worldofwarcraft.com/index.xml>
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press, Cambridge
- Webb EJ, Campbell DT, Schwartz RD, Sechrest L (1966) Unobtrusive measures: nonreactive research in the social sciences. Rand McNally, Chicago
- White H (2011) Scientific and scholarly networks. In: Carrington P, Scott J (eds) *Handbook of social network analysis*. Sage, London (forthcoming)
- White DR, Reitz KP (1983) Graph and semigroup homomorphisms on networks of relations. *Soc Networks* 5:193–234
- Wikipedia (2010). http://en.wikipedia.org/wiki/Object_oriented_language. Accessed 3 May 2010
- Zheleva E, Sharara H, Getoor L (2009) Co-evolution of social and affiliation networks. In: 15th ACM SIGKDD conference on knowledge discovery and data mining