Potential Human Rights Uses of Network Analysis and Mapping

A report to the Science and Human Rights Program of the American Association for the Advancement of Science

Skye Bender-deMoll

April 28, 2008

Abstract

This report investigates potential new tools and existing applications of network analysis and network mapping to assist or facilitate human rights work. It provides a very brief overview of some network concepts, quick introductions to a number of relevant fields of research, and some specific examples of how people are currently using network tools for academic and applied work. The examples serve as an overview and entry point to the research areas. As this is a developing and fragmented field, classification is difficult. However, some common points exist and a few conclusions are presented. Some of the risks and challenges of network research are discussed, along with criteria for evaluating potential future projects. Finally, several possible projects are proposed.

Contents

1 Background.

The AAAS Science and Human Rights Program is investigating the application of network analysis and network mapping to human rights problems because these emerging technologies appear to offer new approaches for analysis, communication, human rights, and strategic planning. If these tools prove useful, it is important that their potential benefits are available not only to industry and national interests, but also to groups engaged in the global struggle to fully realize the fundamental human rights outlined in the UN declarations. The goals of the report are to explore the basic concepts and practices involved in network analysis and mapping, and to determine how the relevant techniques can best be disseminated to the human rights community and utilized at all levels, from grassroots organizations in the field, to policy research and implementation.

2 A quick introduction to networks.

How is network data different from other kinds of data?

The key feature that distinguishes *network* approaches from other kinds of data collection is their focus on relationships. Standard *tabular* or *attribute* data usually list columns of items or observations along with their properties or values. A tabular example would be a list of names showing age, gender, employer, income, etc. (see Table 1).

Table 1: Example of tabular data

Name	Age	Income
Jane	31	$50,000
Maria	25	$32,000
Chu	40	$46,000

By contrast, network data describe *the ties between entities*, such as people, concepts, or organizations. These relationships are usually expressed in pairs: Jane and Maria are linked by their friendship, Maria is Chu's employer, etc. (See Table 2).

Networks allow representation of one-to-many and many-to-many relationships between items. This makes networks a useful way to store information about complex issues such as human rights, where orderly linear relationships seldom occur and it is important to track how many things are interconnected.

Table 2: Example of relational data

Name	Relationship	Name
Jane	*friend of*	Maria
Maria	*boss of*	Chu
Chu	*sent email to*	Jane
Jane	*sent email to*	Maria

Figure 1 is a nice example: it shows a network of alliance relations between a number of different groups. This kind of relational data, and far more complicated examples, can now be studied systematically using new approaches and tools for analysis. Interestingly, there appears to be some common properties and structures that show up in many different kinds of networks. In some cases these structures can provide insight about the underlying processes that produce them.

What exactly does *network* mean here?

Network is a term that has both technical and common-use meanings in multiple fields. The meanings are closely related, but the connotations can be different enough to cause confusion. In a human rights context, *network* usually refers to groups of individuals or organizations having some form of common interest and relationships to each other such as "the network of women's-rights NGOs." In this report, *network* will usually be used in a more specific sense, referring to a collection of data that describe relationships (such as linkages, associations, or contacts) between entities (such as organizations, people, or documents). This use of network has the sense of not only naming a set of entities with something in common, but of specifically enumerating the relations between them so that patterns and structures can be revealed or utilized. This sense is also distinct from *networking*, meaning "schmoozing" or strategically socializing (although these kinds of social relationships can be described using networks). In fact, all kinds of relationships can be considered networks, some more usefully than others. For example, if we are considering a group of women's-rights NGOs, we could speak about the following networks:

- *funding network* – the linkages of grants and contracts between organizations.
- *communications network* – the pattern of email and phone calls between NGOs.

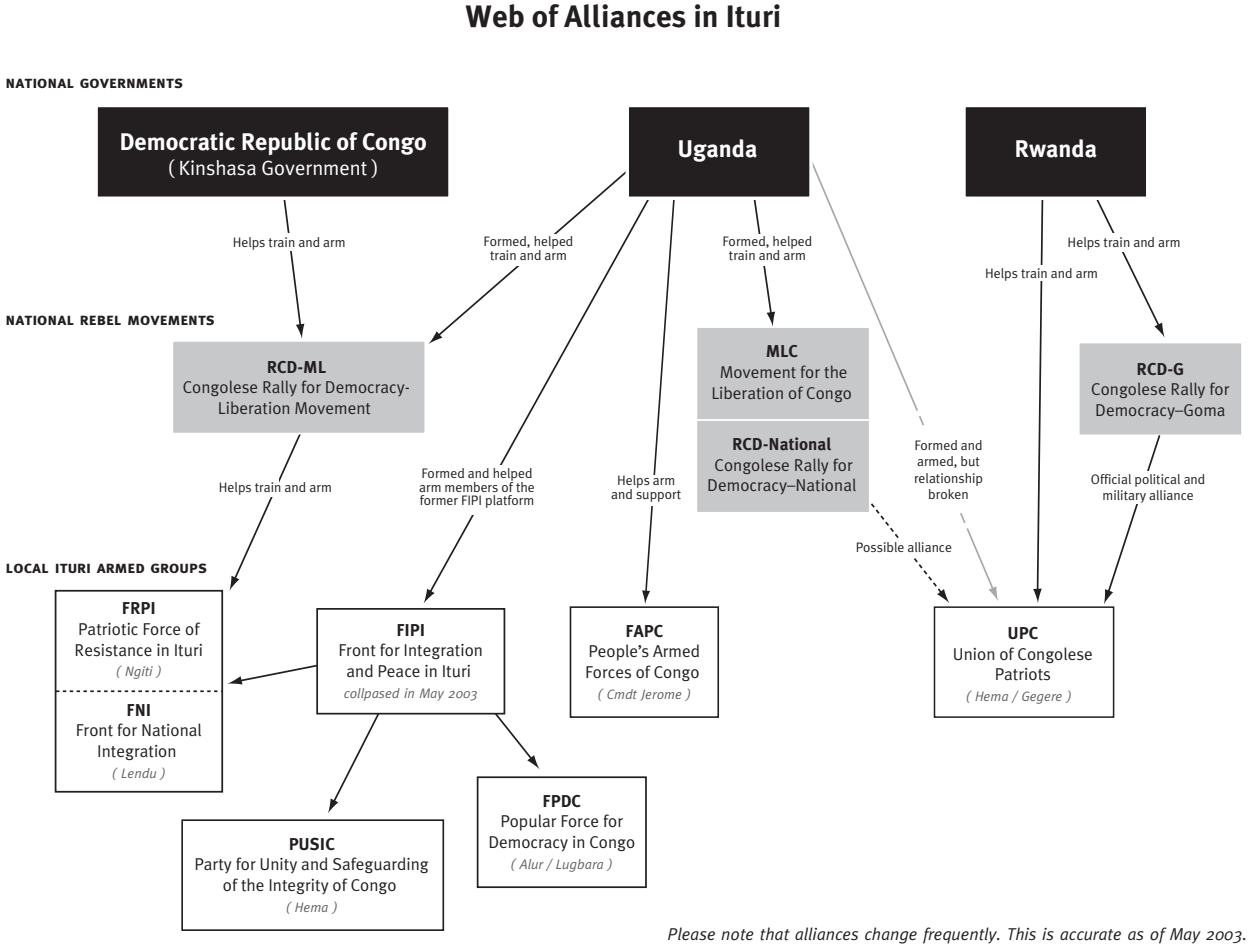


Fig. 1: An example of a complicated set of relationships displayed simultaneously. This diagram, from a Human Rights Watch report on violence in the Congo [1] shows countries, organized movements, and other armed groups with the major alliance relationships between them indicated by labeled arrows.

- *collaboration network* – which groups are working together on joint projects .
- *support network* – the allies that each group can call on for aid.

If all organizations were identical, there would be little use in describing the relations between them—it would be much more useful to lump them all into the same category. In real life, however, organizations do vary not only in size and budget, but in how well connected they are, whom they work with, and how closely integrated they are with the groups they are aiding. Sometimes organizations have huge budgets but spend them inefficiently because they are not getting good information. There may also be groups with few resources that are crucial to the network because they are sources everyone goes to for information. Network analysis provides some tools for looking at some of these subtleties: not only what groups "have," but where they are "positioned."

Representing relationships.

In order to talk about positioning and relationships, network researchers use some specialized terminology. This can be confusing at first, but at the same time it can serve as a useful shorthand for describing patterns of linkages. The entities (people, organizations, doc-

uments, concepts) that are linked are usually referred to as *nodes* and the relationships between the nodes are called *ties*, *links*, or *edges*. Sometimes it is useful to think of links as having *directionality*. For example, if Node A gives money to Node B, the "gives money to" relation would usually be considered a *directional tie* as it describes some sort of transaction or one-way flow. In pictures of networks, directionality is usually shown with an arrow. An example of an *undirected tie* would be the relationship "collaborates with." If Node A and Node B are engaged in some kind of mutual exchange, it doesn't make sense to describe this relationship as directional.

In some situations, it is useful to describe relationships as having different strengths: some people give more money, some friendships are closer than others. When the links of a network have weights assigned to them, the network is called a *weighted* or *valued network*. In some cases, it is useful to assign names or tags (labels or categories) to the nodes and links of a network. These are usually referred to as attributes.

Networks can be used to describe many kinds of relationships between various levels of actors and entities, both abstract and concrete. There are a few broad categories (with exceptions):

Transmission networks. Something actually "flows" along the network. It can be material (water, electricity) or intangible (money, news) but it is generally measureable. In a transmission network there is usually some kind of non-abstract physical linkage, or at least an exchange between the nodes—something which can be "broken."

Interaction networks. Patterns of contact, or perhaps transmission (email, disease). The link is usually an "event" (something nameable with a specific time or range of time associated with it) or a description of a pattern of events. In interaction networks there is often something (such as information, material, or resources) that can be passed between nodes during contact.

Attributional networks. An expression or statement of a relationship. Ties are an acknowledgment of social connection, a named reference, recognition of influence. Trust and power within organizations, friendship between people, or the citations authors include at the end of a document can be thought of as attributional links. Although these relations are "real" they are often very difficult to measure as they are usually part of an individual's knowledge. For this reason they are often called *cognitive networks*. The links in these networks can be quite loose about time, the nodes that are linked don't have to be "alive" at the same times, or even exist at the same level of abstraction.

Affiliation networks. "Belonging to" relationships. Nodes are linked if they share a group, category, or some specific property of interest. Affiliation networks can be entirely abstract: the ties are really more like shared categories, such as co-occurence in a document. These include networks in which nodes are linked by correlation or other measurements of similarity.

These categorizations can be elusive and imprecise; most examples have some elements of each. But is still important to think carefully about what category is being used when doing analysis because the precise interpretations of network measurements will vary depending on the type of network.

3 What can you do with networks? How do you analyze them?

The examples so far have all been of trivially small networks. When a collection of relational information is large enough to be useful it usually becomes complicated and bulky to work with. Although there are definitely examples of groups usefully employing networks without using formal analysis, network data is usually manipulated on computers using specialized software programs.[1] Some familiarity with the subject is generally required to make good assessments about networks because accurate conclusions can be counter-intuitive; many standard statistical techniques cannot be appropriately applied. But whatever the methods, the basic tasks and goals are often quite similar.

Identify communities.

A very common analysis task on networks is to discover the "natural groups" or *communities* based on connections between the nodes in the network. There are various technical definitions of what constitutes a community on a network, but the basic idea is to find *clusters* of nodes that have more connections to one another than they do to outsiders. Some definitions of communities allow individuals to be members

[1] The International Network for Social Network Analysis (INSNA) maintains a useful web page listing of network analysis software at http://www.insna.org/INSNA/soft_inf.html

of multiple groups at the same time. An example of a network clustering task is finding *communities of practice*: groups of individuals who work together or in the same field and may share similar knowledge and perspectives.

Locate important nodes or ties.

Depending on the type of network, some nodes may have relatively more important (or "powerful") positions. The definition of importance is quite context dependent; in some cases it may mean having many ties, but it is usually not so simplistic. In some situations important nodes may be defined as being very well-integrated or *central* to the network, or as having a *bridging* or *gatekeeper* role between two communities. These ties can be important because their removal may break the network into parts. There are software procedures for calculating various measures of importance for nodes and links in a network, but the decision about which one to use depends on what the researcher wants to learn about the network. An example would be locating opinion leaders by using networks of document citations — in this case the important nodes would likely be those with high *centrality*.

Discover roles and positions.

Sometimes what is interesting about a network is not which node is most important, but rather which nodes have the most similar patterns of connections. It is possible to run software procedures to locate nodes or groups of nodes that have *structurally equivalent* positions—meaning that they have similar links to nodes that are themselves similar.

Uncover hidden connections.

When a lot of information about separate individual relationships is assembled into a shared structure it may be possible to find *indirect paths*: a series of connections between nodes that are not themselves directly linked. It may also be possible to discover alternate routes that were not previously used and would not otherwise be apparent.

Make tangible representations for communication.

If a network is not too dense or too large, it can be mapped out into a visual form using a number of computer programs. Nodes are often labeled so that viewers can orient themselves and bring their knowledge to bear when interpreting the network map. Colors and shapes are often used to indicate attributes of the nodes and edges. Visualizing relationships in an iconic form gives people something concrete to refer to in discussions and a means of exchanging knowledge about something very complicated.

Facilitate discussion.

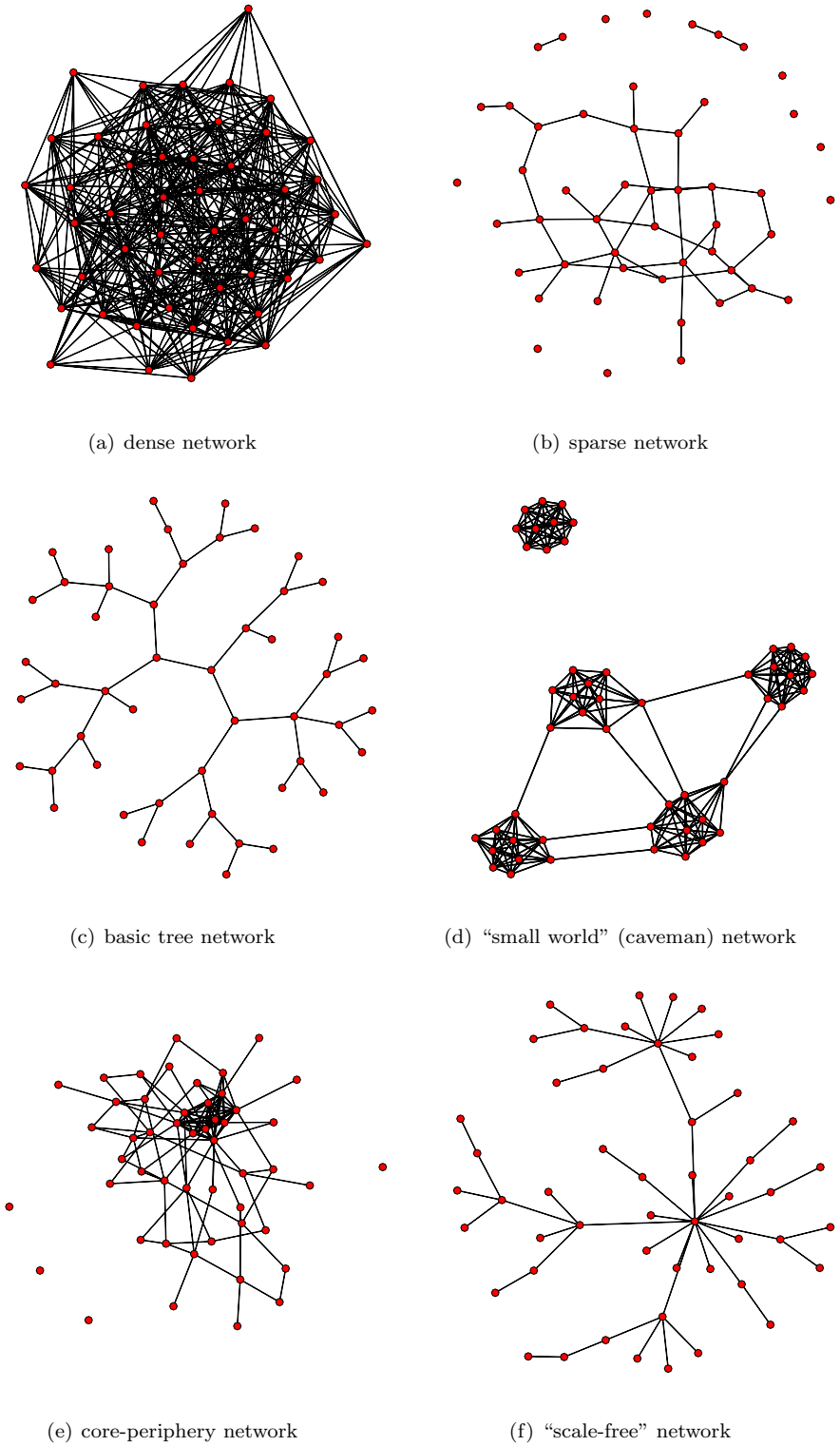
In some cases, the actual data that is collected is less important that the process of acquiring the data. This can be true when the collection process takes place in a group discussion that puts relationships into a shared form so that participants can discuss and argue about them.

A few more key concepts for talking about networks.

Density In *dense* or *highly-connected* networks each node has a very large number of connections, and tends to be linked to most of the other nodes in the network (Fig. 3a). In low-density or *sparse* networks it is still possible that some nodes have many connections, but overall most of the nodes are not directly tied to one another (Fig. 3b). Most of the networks that are interesting and useful to work with are sparse networks. Networks that are very sparse may have *isolates* (nodes with no ties) and *disconnected components*—groups of nodes that are tied to each other but do not have links to other parts of the network.

Distance In networks, this is usually thought of in terms of the number of steps it takes to "travel" along the ties of the network from one node to another, with each tie counting as one step. The *neighborhood* of a node is the group of other nodes that are *local*; they can be reached by searching a very small distance along the network, and are perhaps even directly linked to the source node. Networks that have dense local neighborhoods are described as having a high degree of *clustering*.

Centrality There are many different centrality measures that make important distinctions about what is measured, but the basic idea is to express how much a node is "in the middle" of the network: how well connected and integrated it is. Again, the appropriate centrality measure depends on the researcher's interests and the category of network.



(a) dense network

(b) sparse network

(c) basic tree network

(d) "small world" (caveman) network

(e) core-periphery network

(f) "scale-free" network

Fig. 2: Generated examples of various types of networks, each with 50 nodes.

In addition to the above properties that describe links between groups of nodes and ties, there are also some descriptive classifications of larger relationship structures. The following concepts provide more refined definitions about the shapes of the overall pattern of ties in a network.

Trees are hierarchies in which each node may have multiple *children* below them but only one *parent* above them (as in a traditional org chart). Figure 3c shows a tree drawn without the hierarchical positioning of nodes on the page. *Cross-linking* ties (links between branches) will not appear in a tree network. Pure trees are not found very often in naturally-occurring networks, but they are frequently used in classification systems or any situation where a strict hierarchy is imposed. It is more common to find tree-like networks in which a few cross-linking ties between nodes on different branches do appear.

Small-world is a term used to describe networks that have many small, tightly-knit groups—each of which has a few ties linking it to other groups (Fig. 3d). In other words, small-world networks are locally dense but have sparse *short-cut* ties to link groups that would otherwise be remote or isolated. This leads to some surprising properties, and may be responsible for the often quoted "six degrees of separation" between any two people.[2] Many naturally occurring networks have small-world properties.

Core-periphery describes networks that are dense in the middle and sparse on the edges (Fig. 3e). This means a *core* group of a few densely interconnected nodes, around which are many more nodes that have some links to the core but few links to each other. These in turn are surrounded by a much larger fringe of *periphery* nodes that may have only one or two ties, very few of which are directly linked to the core.

Scale-free and **Power-law** are terms used to describe networks in which there are very few nodes having huge numbers of connections, some nodes with moderate connectivity, and many nodes having very few ties (Fig. 3f). Many natural networks have this feature of rare "popular" nodes and frequent "ordinary" nodes, and in fact often have similar proportions of nodes in each of the categories. This relationship—that the number of nodes of each kind found in a network goes down as the number of ties for each kind goes up—can be described by a type of mathematical formula called a "power law", hence the confusing name.

[2]Several famous studies have shown that even people who have never met are often connected in a social network by a chain of less than six acquaintances.

4 Summary of relevant research areas.

There are a number of overlapping disciplines and areas of practice that employ networks. In some fields, networks are an integral part of the theoretical framework, while in others they are simply a useful data structure or set of software procedures. Most research programs draw techniques from several of these areas. Because the research on networks is taking place in fields ranging from physics to philosophy, there can be different terms that refer to essentially the same concepts. It can be quite challenging to make sense of the soup of three letter acronyms for each subfield. See Table 3 for an overview; a more detailed discussion of areas most relevant for network mapping follows.

Graph Theory.

A very formal and mathematical field of networks, *graph theory* deals with networks as *graphs* or sets of points (more formally called *vertices* or *nodes*) linked by *edges* or *arcs*. Graph theory is somewhat like the "geometry" of networks, in that it uses various axioms and derived theorems about the properties of graphs to help with the investigation of networks. Although much of the work in the field is fundamental to social network analysis, it is probably not immediately relevant here. One of the key tools used by researchers in the past was the representation of relational data in a matrix form, which allowed techniques from linear algebra to be employed for analysis (see Table 4).

Social Network Analysis (SNA).

The phrase "social network analysis" refers to the application of network methodologies to human's social relations and groups. SNA developed out of sociological and anthropological traditions of representing social structure. There is a great deal of overlap and exchange of methods with formal graph theory, network theory, anthropology, organizational studies, and mathematical sociology. SNA practitioners attempt to make quantitative investigations of behavior patterns, focusing on the role of group structure and relational aspects of society, with less attention on individuals' attributes. While SNA is sometimes criticized as being more of a set of concepts or a toolbox

Table 3: Alphabet soup: network subfields

Graph Theory	The formal geometry of networks
SNA	Social Network Analysis
Social Networking	Websites for managing friends and associates
ONA	Organizational Network Analysis
DNA	Dynamic Network Analysis (for networks that change over time)
DNV	Dynamic Network Visualization (animations of networks that change over time)
Network Viz	Network mapping and visualization
Link Analysis	Military term for networks
Data Mining	Surfing the database for surprising cases
KM	Knowledge Management (networks of who knows what)
ABM	Agent-Based Modeling (technique for simulation)
ANT	Actor-Network Theory (philosophical extensions of networks)
Social Capital	Networks as resources
Power Mapping	Mapping power relationships (also Power Analysis)
Issue Networks	Networks for policy analysis (also Policy Networks)

	Jane	Maria	Chu	Eduardo
Jane	0	1	0	1
Maria	1	0	1	1
Chu	0	1	0	1
Eduardo	0	1	0	0

Table 4: Example of a *socio-matrix* used in graph theory to store a network. The "1"s in each row indicate the presence of a tie to the corresponding column, so Jane has links to Maria and Eduardo. Not all links have to be symmetric, in this example Eduardo does not link back to Jane.

of techniques rather than an integrative social theory, the field brings fifty years of insight into the application of network concepts to real problems. A few of the interesting key threads include:

Strength of weak ties is a deservedly-famous observation by Mark Granovetter that individuals tend to have strong ties to similar people who may have similar resources and information [2]. It is the rarer and "weaker" ties (casual acquaintances or contacts) that often connect one to more diverse groups of people, thereby providing access to less common resources and points of view. Thus, although your close friend may work harder to help you get a new job, it is likely to be an acquaintance that actually gets you a useful lead.

Diffusion of innovations is a subfield that focuses on how ideas and influence move through groups. A key observation is that when new ideas or products appear, most people don't learn about them or decide to use them based on widely broadcast messages. Individuals are instead more likely to learn new things from a few of their friends who are "early adopters." Thus, "late adopters" are those who only start a new practice once there is a clear majority of people already engaged. Everett Rogers [3] proposes that this tends to produce a regular pattern in the adoption rates of a new technology or practice in a population. More recent insights have been developed from the study of simulated epidemics (of either diseases or ideas) spreading through networks.

Triadic closure is based on the observation by sociologist George Simmel that the sets of relations among any three people (a triad) tend to change in predictable ways. In short, my friend's friend is likely to become my friend, my friend's enemy may become my enemy, and if my two of friends are enemies, soon only one will be my friend. Although this process may only apply to certain kinds of attributional networks, the basic idea is that a low-level process between pairs and small groups can often create recognizable signature patterns in the overall network structure.

Although SNA encompasses some powerful techniques, there are a number of serious limitations. (Table 5 gives a quick summary.) Usually the kinds of networks organizations want to investigate are quite difficult to actually measure. For example, it can be hard enough to come up with a clear definition of "support" or "power" between organizations. But even if a

good definition can be found, the concept may not be directly measurable. It may be necessary to find measurable relationships that can be used as a proxy to infer the relations of interest. To put it more concretely, suppose an organization wanted to analyze its internal friendship network. There is no way for the organization to look inside it's members' heads to find out how they feel about each other (fortunately!) The organization could ask each of its members a set of questions about each other, but there are many reasons why the answers might not yield the necessary information. It is quite possible that each person might have a different definition of friendship. Another option would be for the organization to frame the definition of friendship as behaviors that can be observed objectively: How often do members visit? Do pairs of members exchange gifts, offer support, and participate in the same events? These kinds of measurement issues are by no means unique to networks; they come up in most situations when attempts are made to measure things in the real world.

Table 5: Some challenges for using SNA

Measurement	Distinguishing between the relations we are interested in and what can actually be measured.
Multiplicity	Choosing among the multiple kinds of relationships to measure.
Boundaries	Where does the network end? Who should be included?
Missing data	Getting enough responses to make the network sufficiently accurate.

Part of the problem is that it is not really correct to speak of "The Network" linking a group of nodes—there are always many different kinds of networks that can be constructed. This is because between almost any pair of entities there are a multiplicity of relationship types and levels of analysis where relationships can occur. For example, two organizations may be related because:

- one funds the other,
- they collaborate on a project,
- people in the organizations send email to each other,
- their websites cross link,
- they perform similar kinds of work,
- they attend the same conferences,
- they are mentioned in the same news article,
- employees move between them,

or any of an innumerable number of ways—some trivial, some significant. Each of these relationship types could be used to construct a network. While this provides a great deal of flexibility, it can be challenging to determine what network or bundle of networks is appropriate to answer the questions at hand. Other methodological and ethical difficulties are discussed below in the section on issues and risks. More information about SNA can be found in an easy-to-read overview by John Scott [4] and in a very useful online textbook by Hanneman and Riddle [5].

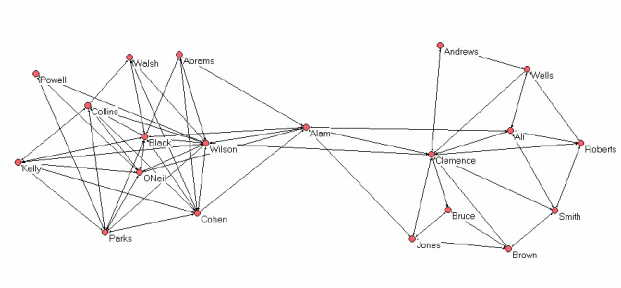
Organizational Network Analysis (ONA).

According to many of its practitioners, "organizational network analysis" is essentially the same field as SNA, merely repackaged with terminology that is more acceptable to business clients. ONA focuses strongly on relationships between individuals either inside of or between companies and is frequently used as a management tool [6]. There are a number of commercial consultancies offering services in this domain, several of which make rather extreme claims about its utility, although a few are well-regarded. Cross and Borgatti [6] give a good overview of their experience working with a number of organizations facing integration and collaboration challenges. Figure 3 shows an example of the kind of organizational integration problem ONA can assist with. Organizational network consultant Valdis Krebs has good reviews and case studies in general business language on his website.[3]

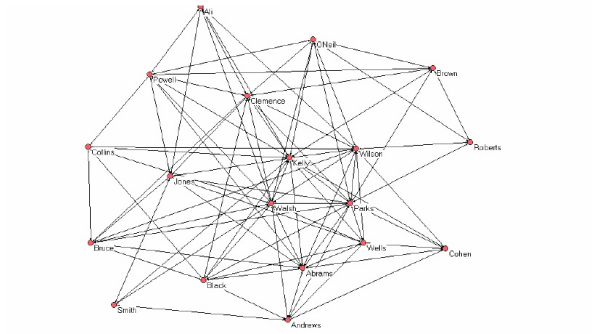
Network Mapping / Visualization.

Network mapping is a means of visually presenting relational information that is too complicated to show with simple tables and cannot be sensibly displayed as a geographic map. Network mapping is a specific tool used as a component of multiple fields, as well as something of a micro-discipline in itself. The goal is to create a "map" that makes the relationships between various entities (such as people or organizations) both visible and informative. The challenge is that, unlike

[3]Website with well-written ONA case studies: http://www.orgnet.com/cases.html



(a) Pre-intervection



(b) Post-intervention (nine months later)

Fig. 3: Network of information-sharing ties within an expert consulting group showing an initial division at time (a) alleviated through creation of a number of new initiatives, eventually resulting in a more cohesive network at time (b). From Cross and Borgatti [6]

maps based on geographic positioning, in most situations we need some kind of system of rules to decide where to position the elements.

The *sociogram*, or diagrammatic map using points and lines to display social relationships (employed in many figures in this report), is generally credited to Jacob Moreno's work from the early 1930s. Recent increases in computing power have made it possible to use computer programs to automatically produce network layouts. A number of procedures have been developed and applied to the problem, some from the field of *graph drawing*. (The term *graph* here is the technical term referring to a set of points and lines.) Other techniques like *Multi-Dimensional Scaling* have been lifted from wider statistics fields, as well as several purpose-built layout procedures. The goal of most layout techniques is to find a way to position the nodes on the page so that the viewer can read it like a geographic map, meaning that distances between nodes on the page correspond to distances measured across the network.

However, just as some distortion must be introduced when squishing the three dimensional globe of the earth onto a flat map, most network maps must distort some of the relationships they present to the viewer in a network map. Several studies have shown that viewers' interpretations of network maps—such as who is most important or best connected—are strongly impacted by the relative positioning of the nodes on the page [7]. For most networks, especially dense networks with lots of ties, it is never possible to create a perfectly accurate two-dimensional map. Thus it is important to decide which information should be omitted or compressed, and to make the decision in a rigorous way.

The key distinction between a network map and a flow chart is that the relative positions of elements in a network map are assumed to be meaningfully based on the relationship structure, rather than placed at the discretion of the designer to convey other types of information. John Emerson has written an elegant review of some of the issues in creating information graphics for advocacy [8], although most of the network examples in the booklet are of the flowchart type. Many examples of network visualizations are included in later sections of this report, and there are also several online collections that demonstrate diverse approaches.[4]

[4] http://www.visualcomplexity.com is a popular gallery of networks and network-related graphics.

Dynamic Network Analysis.

Dynamic or *longitudinal networks* are networks in which the set of relationships or the membership (which nodes are in the network) changes over time. They are used to describe processes in which the structures of relationships do not remain constant over the time period under analysis. Thus, dynamic network analysis often deals with multiple "waves" of data, each describing the network at a subsequent time step. Visualization of dynamic networks is frequently done using animated movies[5] to show the structure shifting over time [9].

Link Analysis.

Link Analysis seems to be a preferred term for network analysis when used in the U.S. intelligence community. Confusingly, it is also used as a term to describe the study of internet hyperlinks between web pages. The intelligence sense of the term focuses on connections between individuals, membership or communication with subversive groups, or participation in events. While there is little public information about this kind of work, numerous anecdotal reports suggest that this might be a fairly crucial element of various proposed or operational government data mining programs. If relational information from many sources (such as intelligence reports, travel records, possibly phone records, internet use, and bank transactions as well) can be collected in centralized databases, it may be possible to "connect the dots" to locate criminal suspects. Additionally, if the databases include substantial proportions of the population, it could conceivably be possible to search the collection for activity patterns that might suggest new suspects.

Some security experts are highly critical of these approaches, suggesting that network approaches are most effective when applied to specific suspects using existing legal frameworks, rather than investigating an entire population at once [10]. Otherwise, too many resources will be wasted chasing false positives instead of investigating legitimate threats. More importantly, any potential security gains may not offset the serious privacy and civil liberties issues raised.

Knowledge Management (KM).

The main focus of Knowledge Management is mapping out and cataloging information in an organization: who knows what, who seeks advice from whom, and where information resides. KM also includes work on developing systems for cataloging and linking the documents used inside organizations. These categorization systems are often called *ontologies* and take the form of a hierarchical tree, or a specific kind of tree-like network of categories and sub-categories. KM frequently uses network analysis of communication patterns to track how various branches of an organization coordinate with each other. Sometimes *semantic networks* are considered an element of KM. Semantic networks (also known as *knowledge graphs* [11]) are collections of relationships between terms (or phrases) that occur in discussions, literature in a specific research area or a collection of business documents. The associations of terms and phrases appearing in a semantic network can serve as a rough summary of the key elements in a text.

Agent-Based Modeling (ABM).

Agent-Based Modeling describes a class of techniques for producing models or simulations to aid in understanding various processes or even occasionally to make predictions. Agent-based computer simulations are characterized by large numbers of objects or *agents* that follow relatively simple rules for interacting with one another and some representation of an environment. Even though the agents are simple, sometimes very complex behaviors emerge at the population level, often in unexpected ways. A classic example of this is "ant foraging." Ants are very good at finding food. Once they find it, many ants will work together to bring it home, apparently in a highly organized and coordinated way. Using software tools, researchers have constructed models wherein simulated "ants" follow basic rules:

1. If you are home, go out and wander randomly.
2. If you find food, bring it home, and leave a "scent trail."
3. If you find a scent trail, follow it until you find food or are at home.

Surprisingly, just these few rules are enough for fictional ants to show foraging patterns that appear very similar to real ants.

ABM can be contrasted with another common modeling technique, using *differential equations* to describe how various aggregate properties of the model interrelate and change over time. For example, a model of a predator-prey relationship in an ABM framework might include many "fox" and "rabbit" agents that "eat" and "reproduce" in a virtual arena

[5] Several examples of dynamic network animations are at http://www.stanford.edu/group/sonia/examples/index.html

where their population can be monitored. A differential equation model would describe how the fox population (an aggregate measure) is dependent on the rabbit population (as an aggregate measure) and vice versa. ABM models tend to be more complicated and more difficult to analyze, but often exhibit complicated interdependencies and surprising behaviors suggestive of those in the real world. These features are often overlooked when a system is modeled solely using aggregate quantities like population.

One way of categorizing agent-based models is to divide them up into two classes: those that are intended to be predictive and those that are more metaphoric. Metaphoric models are based on simple sets of assumptions about a problem; the fewer "bells and whistles," the better. The models can answer questions are along the lines of "could interactions following these simple rules produce patterns of behavior that have some of the same qualities as we observe in real data?" It is generally a mistake to do any modeling of real processes or to make major generalizations from these kinds of models. Just because a feature appears in both the model and real life doesn't mean that both have occurred through similar processes. It only means that processes as simple as those in the model would be sufficient to create the real-world properties. An advantage of these kinds of models is that they usually don't require inputs of real data to be a useful aid to understanding.

Predictive agent-based models, on the other hand, require extensive and detailed input data describing all of the properties in the model. The process to be modeled must be fairly well understood so that all the necessary rules and details can be included. Predictive models tend to be vastly more complex and difficult to build, but if they can be made sufficiently realisitic they can be used to test a number of "what if" scenarios.

Although there are many examples of ABM using networks, the modeling framework is not directly related to networks. It is often included in discussions of networks because they are both tools of Complexity Science, which focuses on problems that have surprising or *chaotic* (complicated, not easily described with summary statistics) behaviors. An important author is Robert Axelrod, who has done work with interesting models of conflict and cooperation [12].

Actor Network Theory (ANT).

Actor Network Theory uses network concepts to create a descriptive sociology of society and technology. Somewhat like the philosophical side of networks, ANT emerged from the sociology of science and technology studies and the work of Michel Callon and Bruno Latour. Practitioners of ANT describe the world in terms of heterogeneous networks: associations between the people, concepts, machines, and technologies necessary for a task, event, or innovation to take place. All of these elements are considered *actors* in the ANT perspective, so the approach is somewhat broader than just looking at relationships between human stakeholders. ANT also suggests that all actors are themselves made up of networks, and uses the term *translation* to describe the assembly and mobilization of elements into a network which is stable enough to be considered an actor in its own right. An example would be a project: at a certain point an assembly of people, tools, and resources coalesces into something identifiable: people begin to refer to the project as having "its" own relationships and goals.

One way ANT can be used is as a theoretical framework for analyzing the processes that lead to a particular point of view becoming the accepted norm, or specific project (such as a transit system) being enacted where others fail. ANT is sometimes criticized as being an incoherent framework due to a lack of consistency in how practitioners apply it. However, minus the philosophical perspective, ANT's basic methodologies do not seem very different from other kinds of network analysis. The basic techniques seem to involve: identifying key people and conducting interviews; analysis of texts produced by organizations; or diagraming out the various people, resources, values, and technologies necessary to achieve a specific task.

Social Capital.

Some researchers find it useful to think of the knowledge, relationships, access, privileges, and rights of individuals and organizations as a form of capital (somewhat analogous to financial capital). *Social capital* can be increased by social and political maneuvering and relationship building. After it is accrued, it can then be employed to facilitate collective action, or permit people who are better-positioned to move towards their goals more easily. A form of power analysis, social capital focuses on the kinds of power people have by virtue of their position in networks. From the social capital perspective, a key ingredient to an individual's power is the degree to which he or she fills *structural holes*—gaps in the relations among groups. These people can act as brokers because all communication between the groups must pass through them. In some contexts, social capital is used as shorthand to refer to how well-structured the networks are within a

community. Groups that can act effectively and adapt quickly to new challenges often have strong networks of trust and good communication patterns, and these groups can be said to have high social capital. Although social capital can be a useful concept, it is also something of a catch-all term, with various authors disagreeing about its precise definition and how literally to interpret the analogy to financial capital.

Power Mapping / Power Analysis.

Most generally, *power mapping* is a technique for creating shared representations of the relative power relations between entities. It is a way of collecting and representing participants' views of power structures, collapsing the complexity into simpler diagrammatic forms and organizing information to focus on the most important relationships. Diagrams can be constructed in various ways, but a popular method involves asking participants to position tokens corresponding to various parties in some sort of space defined so that the location on the page is an indicator of their power and/or position on issues. Power mapping can be used to structure a participatory discussion with issue stakeholders. It is also as means to share knowledge and help a group quickly identify key allies and opponents. One advantage of power mapping is that it requires little technology. A whiteboard or single sheet of paper can suffice. It focuses on quickly collecting qualitative and relative information about parties, and therefore does not require detailed numerical data to be effective. In some cases the approach has been extended to include collection of relational information, making it a form of network mapping. Power mapping is used by many organizations as a community analysis tool (Fig. 4), and appears to have been independently discovered, or perhaps derived from social psychology subjects such as field theory. Several aid agencies are investigating possible uses of power analysis to predict and understand the impacts of interventions.

Issue / Policy Networks.

A policy network is one type of analytical framework for policy analysis. It assumes that groups of organizations and political actors who are interested in an issue, problem, or area of public policy may have relatively stable interdependencies—whether supportive alliances or between antagonists. If these relationships are represented as networks, it becomes possible to study cohesion and positioning of the various groups using network analytic techniques and visualization [13]. If the organizations have some goals in common, they may either function as a coordinated network of advocates, or be relatively independent. Sometimes *issue network* is used more specifically to refer to a relatively non-hierarchical means of negotiating policy by groups operating in loose coalition because they each do not have enough power to directly enact policy. Data for policy network analysis is often collected using interviews or extensive reviews of relevant documents such as news articles and lobbying records.

Data Mining and Text Mining.

A rough definition of data mining would be: gathering large amounts of information from diverse sources into a single well-structured database system to permit analysis and exploration using statistical techniques and computerized procedures. Data mining is distinguished from other modes of data-driven analysis in that the researchers usually do not have specific hypotheses they are trying to prove or disprove. Rather, they are searching for insight, trying to discover general patterns, or sifting for strange or anomalous cases that might merit further investigation.

Text mining is a subset of techniques used when dealing with unstructured text, meaning text that that has not already been carefully tagged or entered into appropriate fields in a database so that it can be easily and appropriately searched. This often requires very sophisticated programs such as textitsentence parsers to determine what the various parts of the text refer to. One of the most common problems in text mining is *entity resolution*: recognizing and matching partial or alternate spellings of names, and distinguishing among references to different people who have the same name. Text mining is also used to construct *semantic networks*: linkages of terms and concepts found within collections of documents that summarize some of the most important relationships.

Data mining in general need not involve any network concepts, but its application in recent years has often been to search for patterns of relationships. The term is often used to refer to programs initiated by the U.S. national security agencies to search vast quantities of consumer information and communications records for traces or patterns of terrorist activity (also Link Analysis). These programs have raised multiple civil rights and privacy concerns due to their comprehensiveness and lack of focus on particular identifiable suspects, resulting in the term gaining serious negative connotations. A limitation of data mining techniques is that they can only be applied to information in electronic form so applications to historical documents or

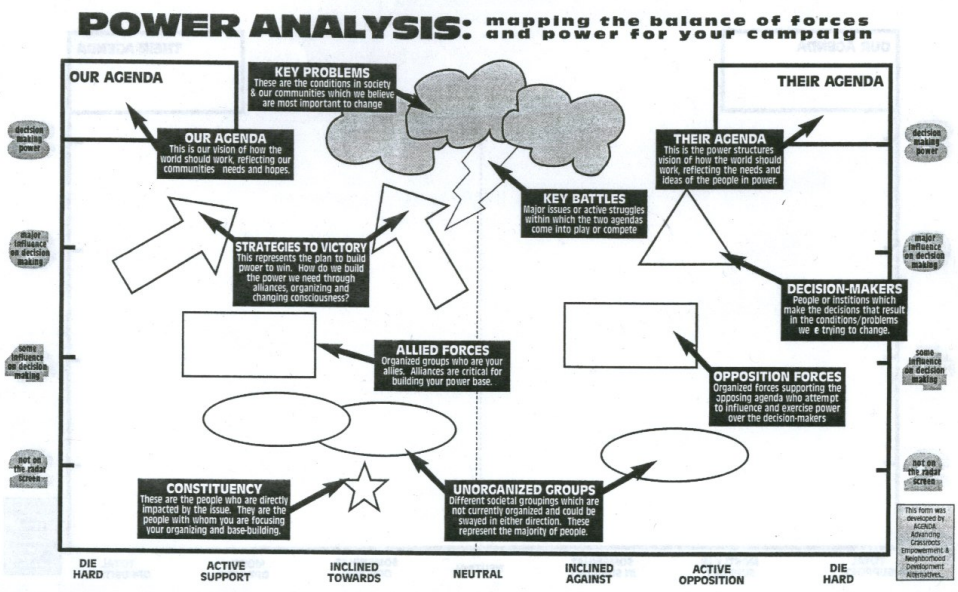


Fig. 4: Sample of a power analysis worksheet developed by AGENDA (Action for Grassroots Empowerment and Neighborhood Development Alternatives), a South Los Angles community organization. Icons for various groups can be arranged so that their horizontal position indicates support/opposition, and the vertical position shows the perceived power of the organization.

archives in developing countries generally require massive coding/data entry efforts.

Social Networking.

The recent increase in use of the phrase "social network" is due in a large part to the popularity of *social networking* websites such as Facebook, Bebo, MySpace and Friendster. Although these sites are loosely based around the concept of networks (and certainly use some network data structures internally), they tend to have very limited support for network analytics. Instead, they serve as tools to support socializing, allowing users to create profiles for themselves including various personal attributes (location, pictures, favorite music) and create mutually ratified friendship links with other profiles. Once linked, users normally gain access to their friends' profiles to read and leave comments, and in some cases receive updates when friends edit or post to their profile.

There are also similar sites (LinkedIn, Orkut, VisiblePath) that are designed for business use. These sites provide ways for individuals to network (in the verb sense) or to determine a chain of acquaintances who can provide an introduction to a needed contact. VisiblePath seems to provide integrated contact management and is one of the few such websites with SNA credibility. There are also sites that specialize in creating customized social networks (Ning) that can be set up for internal use. Several non-profit technology providers have begun integrating networking capabilities into their campaign management software.

5 Ways of collecting and storing network data.

Collecting relational data often requires more work than tabular or time-series data. Fortunately, there are many places where suitable data has already been collected but is not recognized or has not yet been structured to make the network aspects obvious.

Re-purposing existing data.

Many kinds of data are already relational and only require formatting before they can be used. Emails for example, include both the sender and recipient address and can be considered instances of linking events in communications networks. In situations where an organization's entire email collection has been archived, it is possible to construct very detailed pictures of communication patterns (see Fig. 5). Grants and funding information present in an organization's re-

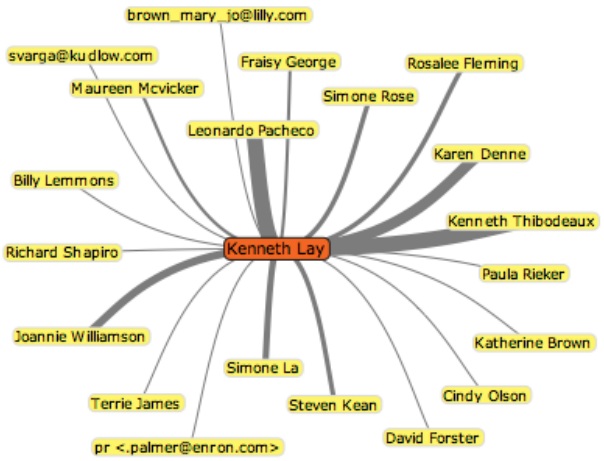


Fig. 5: Kenneth Lay's email communication ego network extracted from the corpus of emails subpoenaed from Enron. Image from the Enron Email Explorer [14]

ports are another form of relational data, but to construct a network it is usually necessary to collect reports from many organizations. In some cases, the data collected by investigative human rights organizations may contain a wealth of relational information if the original data can be located and in a relatively unaggregated form (i.e. the data that was used for research, usually not what appears in reports).

Surveys, interviews and online questionnaires.

When network data are to be collected about a community of people, it is often possible to design a survey or interview process to ask each individual to provide information about their own links. The collection of individual *ego networks* can then be assembled into a single collective description of the community network. However, as with most survey research, it is rarely possible to get 100% of individuals to respond, which can cause serious validity problems in the resulting networks. Survey data collection is quite time-consuming for both subjects and administrators, so it is crucial to be very clear about exactly what kinds of relationships are going to be examined. Cross and Borgatti give a concise table with examples of questions to ask to collect specific types of relationships [6].

In situations where it is difficult to gain access to an entire community, or where the members of a community are not known before beginning a study, it is possible to get some idea of the network structure using *snowball sampling*. In snowball sampling (as in the closely related *link tracing*) a few individuals are chosen as "seed" nodes and asked about their ties to other nodes, each of whom are then asked to name more nodes. This process usually provides a sample that "snowballs" larger and larger until researchers either run out of resources or the limits of the community are located. When this technique is applied to the network of hyperlinks on the internet, it is usually called *link crawling*.

Participant observation and expert opinion.

Much of the initial research on networks was done as part of enthographic studies. The networks produced were often a single researcher's impression of the relevant relationships. Although it is usually a good idea to have data that are more objective than a single individual's point of view, there are also situations where getting experts' descriptions of the relationships can be effective—especially if the experts are themselves members of the network. In situations where there are moderately-sized collections of documents, experts or trained reviewers can often extract some amount of relational information by reading and coding the texts for certain associations. For example, reading complicated texts such as reports or archival documents and extracting ties such as "Politician A *met with* Politician B."

Storing network data.

No matter what method is used for collection, all data should be stored in a standardized way. Network data do not need exotic forms of storage; usually a basic text file, spreadsheet, or simple database will be sufficient. However, some care should be put into setting up the data. Usually it is helpful to keep the information about nodes separate from information about ties. It is crucial to use a consistent identification code (*id*) or standardized names for nodes to make sure the data can be linked up correctly. For example, if the id for a node is "Public, John Q." it must always appear exactly the same way. "Public, J.Q." or "John Quincy Public" will not match correctly. This is a challenge, because most real data is full of misspellings, missing initials, and other complications (the "entity resolution problem" mentioned earlier).

Unfortunately, there is not yet a consensus on computer data formats for networks.[6] Therefore it is usually best to collect the data in whatever form is most convenient, and then translate it for a specific computer program.

[6] A wiki listing various network-related data formats is located at http://netwiki.amath.unc.edu/DataFormats/

Difficult or unusable data.

There are, of course, vast amounts of data that are not useful for creating networks, even if they are very appropriate for other forms of analysis. Many techniques used to process and summarize tabular data (reporting averages, sector totals, etc.) remove the necessary relational details. Basically, to use a piece of data relationally, one needs to be able to answer "who does it belong to?" or "whom did it come from?" Being able to link the information is more important than knowing "how much" or "how big." For example, a spreadsheet that gives totals for how much money each organization spent is interesting, but not helpful for constructing a network. To make a network, it is necessary to know exactly whom they paid money *to*, and in enough detail to determine if two organizations are linked—not just how much they spent in the same sector. Non-relational attribute data are still incredibly helpful for other kinds of statistical analysis—just not as useful for building networks.

Relational information that is not standardized is very hard to work with; the links need to be indicated by a well-defined set of terms or categories. For this reason, text in narrative formats—one of the most understandable ways of explaining basic relationships—is mostly unusable for constructing networks until it has be processed and coded. Data that are not in electronic form are also inaccessible for numerical analysis until they have been processed and interpreted by real people, although they may still be helpful for qualitative analysis.

6 Existing practice: current human rights-related use of network techniques.

Although network concepts are not at all alien to humanitarian work, their application in human rights work is somewhat rarer. Many organizations do explicitly or implicitly use networks, and there are many existing projects and studies that have applied networks and network mapping to problems closely linked with human rights. This section outlines some current case studies, highlights how these employ the concepts previously discussed, and why they are relevant. Some of the groupings are topical and others methodological.

Tracking flows of people and goods.

Networks can be used to represent the patterns of flows between sets of entities. They are, therefore, a useful analysis structure for looking at data about of the movement of people and goods between countries. Some nice analyses have been done using economic trade data [15] between countries to help understand what predicts the flows of global resources. Data on immigration and refugee diasporas are available from several organizations[7] and have been used to construct maps that show both geography and network relations.[8] Figure 6 shows a striking presentation of the spread of displaced survivors from New Orleans after Hurricane Katrina using data extracted from the internet. In general, the geographic flows of people often mirror the ties of migrants' social networks, as displaced people relocate to cities where friends and family members already live. An Asia Foundation / USAID report on human trafficking from Cambodia [16] makes recommendations for studying social networks of victims, employers, and recruiters to help understand possibilities for intervening in involuntary migration flows.

Key tools used: Network modeling, text mining, SNA, network mapping.

Why relevant: The movement of resources among organizations focusing in a region or an issue can be very difficult to understand. Network maps can provide a means of succinctly communicating the overall patterns to a viewer. Network analysis can often locate crucial weakness or control points for the flows.

Law and legal networks.

Network perspectives are being employed as an aid to track the diffusion and acceptance of precedent in the court systems. The basic technique relies on using some of the major commercial law search and citation databases (Lexis-Nexis, Westlaw, etc.) to locate groups of cases and rulings related to a given topic area; the precedent citations are treated as the links in the network. Most of these applications deal with U.S. state and federal courts. In some projects, the network aspect is really only a sampling strategy to aid in collecting relevant documents. Other projects explicitly analyze communities and make predictions based on attributes.

These methodologies seem fairly well developed and effective. James Fowler's work on citation networks

[7] A database of migration and refugee displacement is available from http://www.migrationdrc.org/research/typesofmigration/global_migrant_origin_database.html

[8] For an interactive example, see http://www.niceone.org

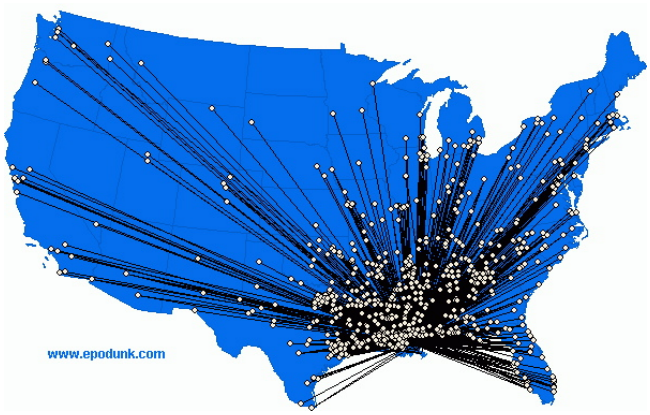


Fig. 6: Combined geographic and link map of the Katrina diaspora with data extracted from blog posting and internet "safe lists." Image from the organization EPodunk [17].

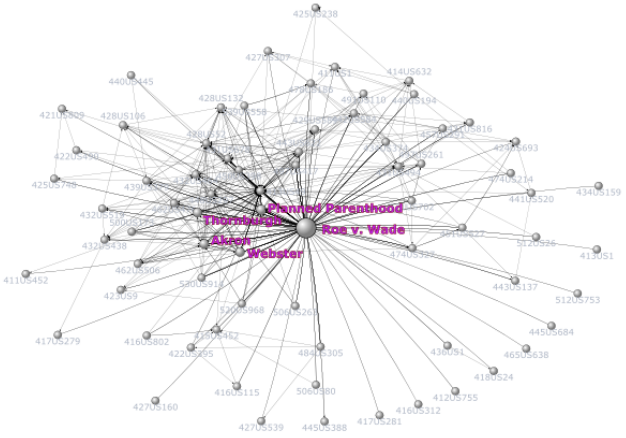


Fig. 7: Map of the network of legal citation references between U.S. Supreme Court abortion-rights cases. Image from [18].

and U.S. Supreme Court precedent shows that the citation network scores are able to reproduce some of the rankings given by legal experts. The scores "...corroborate qualitative assessments of which issues and cases the Court prioritizes and how these change over time" [18] (see Fig. 7). A group at the Department of Government and Politics at the University of Maryland has been working for the past several years on an NSF grant to develop methodology and visualization software for this kind of analysis [19]. Some of their work deals with civil rights, contrasting the behavior of courts considering affirmative action and regulatory takings cases [20].

Other types of legal documents can be examined in similar ways, such as analyzing the links between countries that are co-signatories to treaties and international agreements. Figure 8 shows an example of the network of countries and major human rights treaties growing over time.

One advantage of using legal documents is that a great deal of the data-coding legwork has already been done by the commercial information providers (some of the relevant features are built in to their search systems). Overall, the U.S. legal system has very well-defined systems for recording the process of decision-making and the interests of external organizations (via briefs and filings) as well as referencing. Presumably this holds true in jurisdictions in most developed countries.

Key tools used: Text mining, locating important nodes with network analysis, diffusion of innovations, visualization.

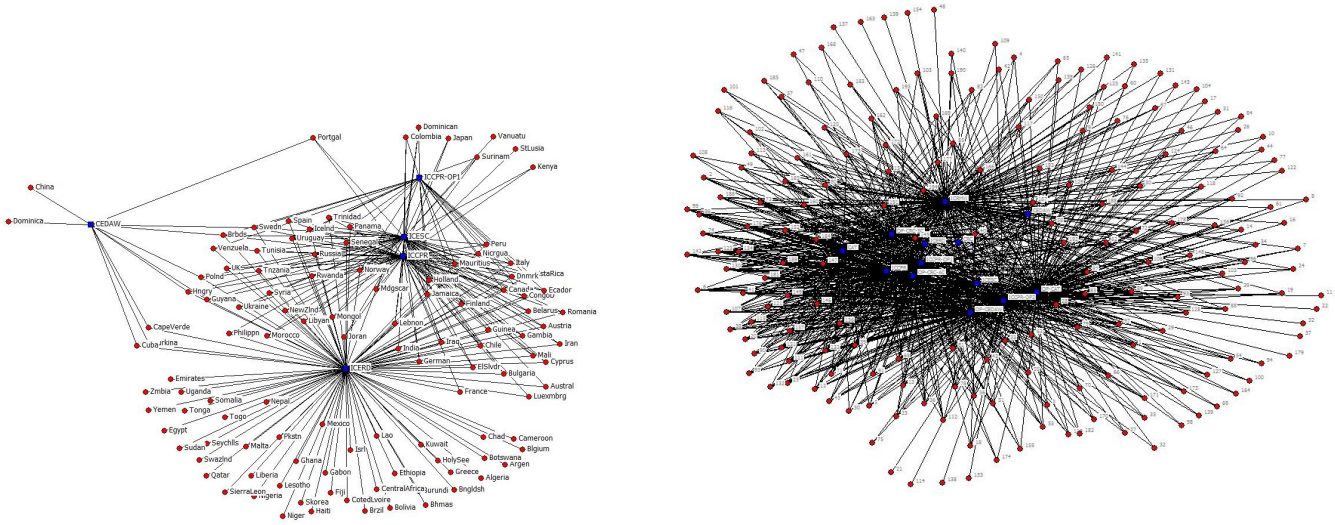
Why relevant: Network mapping may make some of the complexities of the legal system more transparent to the public. It might also be interesting to assess (and communicate diagrammatically to an audience without a law background) the impact of specific landmark human-rights rulings in various jurisdictions and legal traditions. For example, do citations of human-rights treaties appear in federal courts?

Modeling social structure and change.

Networks are often employed as a conceptual framework for studying social structures, social movements, and other complicated cultural processes. Unfortunately, a review of the literature from these fields is not within the scope of this report. However, it is worth mentioning one concrete human rights use of social structure models: validating reports of abuses. The process is quite complicated, but in situations where researchers are able to obtain good estimates of social network properties (such as average family size and numbers of personal acquaintances) it is possible to use statistical techniques to cross-check the numbers of cases reported by multiple sources. In some situations, it is even possible to give estimates of how many cases have not been reported [22]. These methods have been used to corroborate the numbers of people with AIDS/HIV+ in the US [23].

Key tools used: Statistical models of social structure, research and data from Social Psychology, Sociology, Political Science and Anthropology.

Why relevant: Familiarity with common structures of power and social change processes can lead to more accurate assessments of situations and the design of interventions that are more likely to produce the desired positive effects.



(a) by 1980

(b) by 2007

Fig. 8: Signatory countries (red nodes) to major human rights treaties (blue nodes) at two points in time. Figures from Kwon and Barnett [21].

Political contribution networks.

Some jurisdictions in the U.S. have requirements that candidates who are running for office must disclose information about the individuals and organizations that contribute to their campaign. Although this data has traditionally been summarized in a tabular manner, there is a great deal of relational information that can be extracted as well. In the last few years there has been a surge of academic research using network approaches along with campaign finance data to examine political contributions and influence. In races for federal offices, candidates must file reports with the Federal Elections Commission that detail the job title, employer, and home address of each of their contributors. As most of the major contributions come from individuals with important job positions (president, CEO, or VP) this affiliation information is often used alongside filings from corporations' Political Action Committees to measure corporate support for individual politicians. The contributions made by politicians to each other's campaigns are also recorded in this data, making it possible to examine support networks within Congress as well [24].

There are a number of websites in the U.S. that provide access to these data.[9] There are also fledgeling initiatives to create the necessary data structures

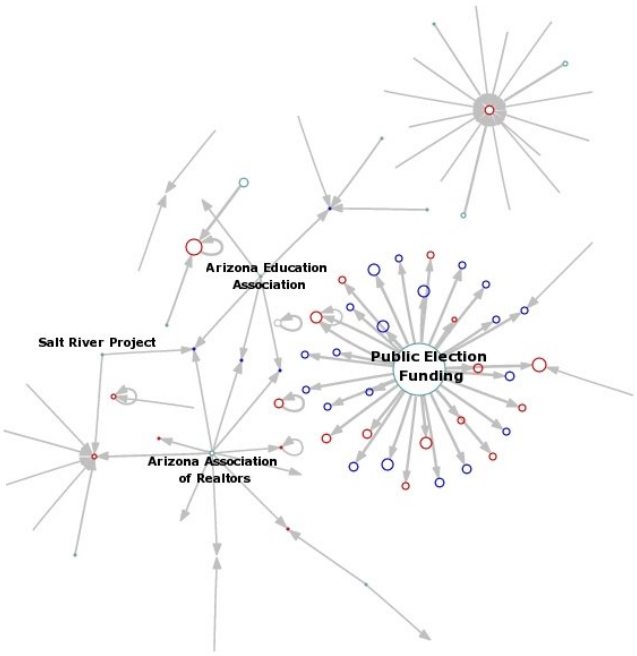


Fig. 9: Network map of contributions greater than $500 to 2006 Republican (red circles) and Democratic (blue circles) candidates for Arizona State Senate. The cluster of candidates on the right have agreed to receive public campaign funding. From Unfluence.net [25].

[9]Some examples are the non-profits The National Institute on Money in State Politics, the Center for Responsive Politics and Maplight.org, winner of the 2007 NetSquared competition for best non-profit tech project.

along with increasing pressure on government offices to release data in more accessible forms. Some U.S. data about budget "earmarks" are available, and information on government contracts will start being released in 2008. Internationally, it appears that at least some information is available for Australia, Brazil, Canada, Chile, Mexico, Ireland, New Zealand, and the UK, as well as others. Several recent websites (see Fig. 9) have attempted to put these data into a more generally accessible format using network mapping. Similar approaches can be applied to other funding domains where information on financial resources flows are available.

Key tools used: Text mining, relational databases, network analysis, visualization.

Why relevant: Tools which increase government transparency can aid the layperson in understanding the financial workings of their political systems. They could help encourage fair elections and more transparent political processes. When contribution information is cross-referenced with politicians' voting behavior on human rights-related issues, it becomes possible to make some judgments about where individual companies or industry sectors stand on human rights.

Voting similarity and structure.

Roll-call vote analysis is another technique used to make politicians' behavior transparent (one with a fairly long tradition). Patterns of voting can be discovered by collecting the series of voting decisions (yea, nay, abstain/absent) for each member of a voting body into the rows of a table. The similarity of each person's voting profile can be assessed by comparing the rows using a distance algorithm to count how often the votes match up. This produces a matrix showing how similar each candidate is to every other candidate. While not directly interpretable, this matrix can be fed into network analysis techniques to discover groups of legislators who tend to vote as a block. The matrix can also be drawn as a network such that the voting blocks will appear as tight clusters. There are other related statistical techniques that can be used to produce related kinds of "spatial" analysis and voting polarization measures. Some of the most influential work in this area has been done by Kenneth Poole, who maintains a website on roll-call analysis that includes interesting animations of U.S. Congress over time.[10]

Not surprisingly, what usually appears across multiple methods of analyzing the U.S. Congress is the division between Republican and Democratic Party members. But intriguingly, Aleks Jakulin's analyses[11] show finer-grained structures as well: blocks of voters within the party structures. Similar tools have been applied to analyze trends in voting by UN member states [27].

Votes are not the only tangible record of the political process. If a legislative body records the names of members who sponsor or endorse bills, it is possible to construct networks in which legislators are linked by the number of bills they sponsor together. These *co-sponsorship* networks can also be analyzed for coalition structures. *Co-voting* networks can also be constructed in which individuals are linked with a tie weighted by the number of times they voted together. This is a subtle distinction from overall voting similarity. Since most legislators vote on most bills, these networks tend to be very dense. However, the networks can be *thresholded* (weak ties with weights below a "threshold" value are removed from the network, leaving a backbone of strong ties) to make more subtle structures visible. See Figure 10 for a an illustrative software example. Another mode of analysis is to look at the overlap in committee membership, creating a network in which the strengths of the links are proportional to the number of legislators on both committees (Fig. 11).

Key tools used: Visualization, cluster analysis.

Why relevant: This kind of analysis could potentially be used by groups to understand where the split points are in political coalitions, where groups within governments stand with respect to various policy decisions. A transparent political system helps with voting and the democratic process by allowing citizens to make informed decisions. Related techniques can give clear comparisons of how politicians actually voted on issues —focusing attention on actions instead of rhetoric.

Corporate interlocks and ownership analysis.

There have been many efforts to develop relatively objective tools to measure and understand the structures of control and ownership in the commercial world. These endeavors often have difficulty obtaining good data, partially because such information can be quite sensitive. One approach is to look at how different firms are related by investigating who sits on their respective boards. The names of board members of

[10] VoteView voting analysis site: http://voteview.com/

[11] Aleks Jakulin's website has good visual examples of roll call vote structure analysis: http://kt.ijs.si/aleks/Politics/index.htm

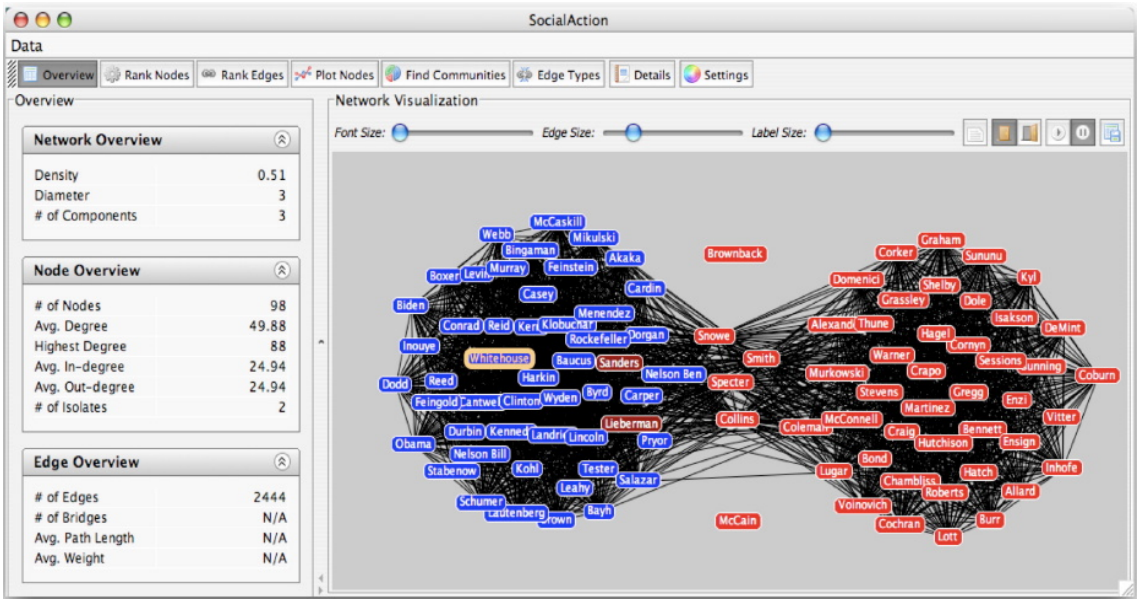


Fig. 10: Screenshot from SocialAction software [26] showing the co-voting network of U.S. Senators in 2007. Republicans are colored red, Democrats blue, and Independents maroon (180 shared-vote threshold).

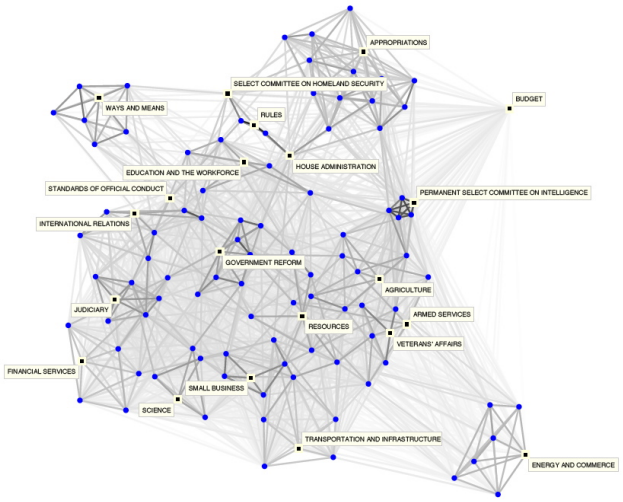


Fig. 11: A co-membership network of committees (squares) and subcommittees (circles) of the 107th U.S. House of Representatives. Ties between committees are weighted by the number of Representatives who serve on both committees. Standing and select committees are labeled. Subcommittees tend to be closely tied to their main committee and are therefore left unlabeled. Image from [28].

publicly traded corporations are usually public information. Two firms are said to have an *interlock* when one person is a member of both boards. The set of interlocks can be represented as a network. When this information is collected for a large number of companies, it frequently becomes apparent that a relatively small pool of individuals control most of the largest firms. An early and quite influential website, They Rule,[12] presented interactive board interlock networks for publicly traded companies extracted from Securities and Exchange Commission (SEC) data. However, whether or not these interlocks actually imply strong relationships (such as coordinated strategies) between companies is a matter of debate.

Interlock analyses such as these can be used in other situations to identify individuals who are holding positions in multiple organizations that may lead to a conflict of interest. Network consultant Valdis Krebs relates an interesting account of how a network analysis approach to ownership relations was used to uncover a conspiracy of deceptive real-estate transactions by a group of slumlords [29]. This example demonstrates how an economic justice organization was able to successively unravel an extended family network that linked the holders of the buying and selling companies, even including the lending organization providing the financing. The activist group discovered

[12] Corporate interlock website by Josh On: http://www.theyrule.net.

the scheme by using the snowball technique with public record data they were able to obtain. While in some ways this is not very different than ordinary sleuthing, the group apparently used visualization techniques to help them organize the search and presumably (although Krebs does not say this) to help explain their case to a jury, resulting in a conviction.

Academic research on the networks and hierarchies of ownership relations has been conducted in several countries. When focused on specific industries and repeated for multiple points in time, this can illuminate shifts in industry structures and trends in diversification or consolidation. An example of this is a project at UC Berkeley that uses semi-automated parsing of ownership filings with the SEC, a technique they applied to analyze consolidation in the communications industry [30]. A very intelligible visualization of a similar data set shows the various media industry mergers and spin-offs as a branching timeline (Fig. 12). In this case, the timeline data is still relational, but because the number of merger relations are very low, it can be displayed in a simple tree form that reads very differently from other network diagrams.

Key tools used: Relational databases, text mining, dynamic networks, visualization.

Why relevant: Large corporations and industry coalitions tend to have a great deal of influence on national policies. If interlocks indicate information-sharing and potential for coordinated action, analyzing networks in a specific industry exhibiting problematic behavior (labor / human rights abuses, poor environmental practices) may reveal opportunities for strategic action to correct the problem. In some situations, understanding ownership patterns can expose networks of influence and help fight corruption. Consolidation is an area of concern in many industries, but particularly in the mass media because a lack of diversity in sources can impact public access to information and opinion. Network graphics may be a useful way to raise awareness around these issues.

Network mapping for project diagnostics and strategic planning.

Several large aid agencies have been investigating power analysis and networks to help them understand the details of local power and politics that often lead to unintended results in aid distribution and interventions. At the same time, researchers have been field testing various network-based approaches for analysis and strategy that can be utilized by the communities faced with various issues. In these examples, network research mostly serves as a participatory tool, a means of community self-assesment. The techniques are often attempted in partnership with NGOs (such as The Institute for Food Policy Research) that are interested in the methodology as an evaluation tool.

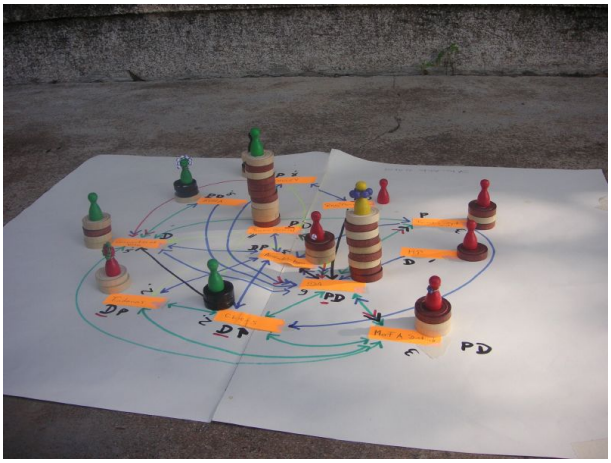


Fig. 13: A Net-Map produced by a single individual during an interview for the White Volta Basin Board evaluation in Ghana. Tokens represent individuals and groups, height of towers shows relative power, and colors of arrows represent types of relationships. Image from [33].

The *Power Mapping Toolkit* developed by Eva Schiffer [32] makes use of tokens which can be used to represent various actors, their relative power, positioning, and modes of influence. The toolkit has recently expanded to become the *Net-Map Toolbox* [33] in which the tokens are arranged on paper and participants are led through an interview process to draw the network of relevant links between the actors (see Fig. 13). It has mostly been used for individual interviews or small group discussions around resource management and local political stratagey. Schiffer explains that the processes is quite time intensive, requiring 1-2 hours per interviewee, but frequently produces valuable detailed discussion among participants [34].

A similar approach, the *tactical map* has been developed by the Center for Victims of Torture and is being tested by a number of organizations worldwide [35]. Tactical maps appear to function as flow-charts of problem antecedents and impacts (Fig. 14), with less standardization of tie types than other kinds of network diagrams. The nodes are also quite heterogeneous, with many different kinds of entities appearing on the same diagram. The nodes tend to be archetypical, representing classes of people and social structures as opposed to identifiable individuals and groups. The maps serve to focus group discussion on specific re-

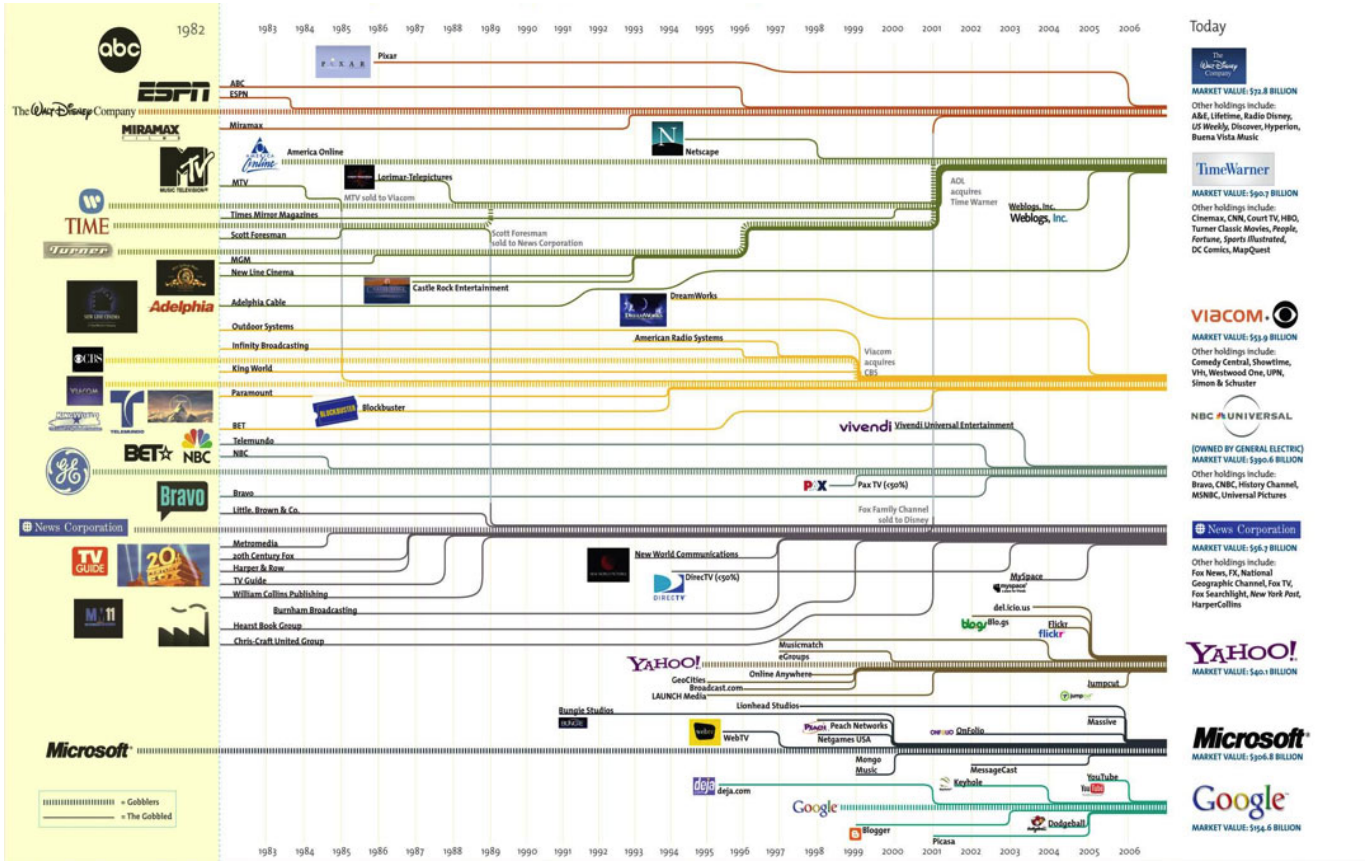


Fig. 12: *And Then There Were Eight: 25 years of media mergers, from GE-NBC to Google-YouTube*. Sometimes relational information can be elegantly displayed without drawing a network. A good example is this hybrid timeline-network map of mergers by Dmitry Krasny [31]. Chronology is indicated horizontally, from 1982 on the left to 2006 on the right. The point where a company's line joins the main group line indicates the date of the merger.

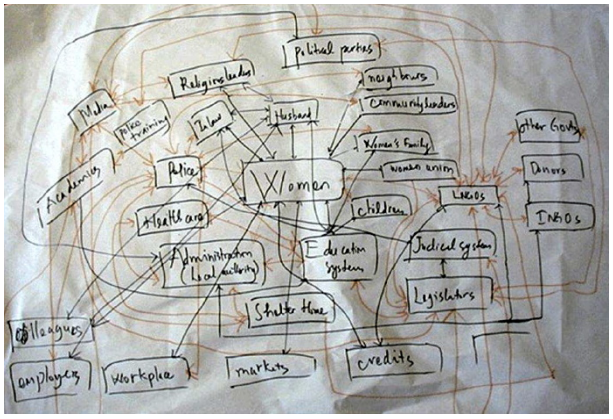


Fig. 14: Hand-drawn *tactical map* of relationships around the issue of violence against Vietnamese women. From a CVT training workshop in Thailand [35].

Fig. 16: A warm-up exercise for participants in a network mapping workshop for a Colombian farmer collective. Colored wool was used to represent the various communication paths by which participants received invitations to the workshop. [38].

lationships in order to determine potential points of intervention in the network that could positively alter outcomes.

The unusual aspect of the net-map and tactical mapping techniques is that the relational data is actually collected through a visualization process—the reverse of most of the other approaches. Darren Noy's network based survey of the homeless policy field [36] relied on extensive roster-driven interviews with representatives of each stakeholder organization. (Key informants were first interviewed to compile lists of organizations to include on the rosters.) The resulting network maps were shared with the groups only when the study was complete, and the the groups' names were disguised in the final versions. This was important because most organizations in the study are engaged in long-term power struggles with each other to shape public policy. Anonymizing in this way served to protect some of the strategic information in the networks, helped to encourage broader participation in the study, and to focus readers' attention on the overall patterns and structures revealed, rather than on the positioning of individual organizations.

Following a mode closer to participant-observation, Biggs and Matsaert's actor-oriented evaluations in Namibia [37] used only the researchers' description of the networks in which they were embedded. Their reports detail their perspective on patterns of communication about farming practices among groups they were working with. They also included recommendations about relationships that should be developed—cells that should be "filled in" on the sociomatrix. (See Figure 15.)

A pair of rural studies in Colombia [38] (Fig. 16) and Bolivia [39] used a group discussion process to design a network questionnaire which was then administered. The resulting networks were used in additional discussions to plot strategy. This approach is being further developed under the heading of *Participatory Impact Pathways Analysis* which includes network mapping as one component.[13] Louise Clark, a key author of the Bolivian study, provides an example of the utility of this approach for giving an overview of the influences at play in the participants' local economy. Her description of the network of information links in the coffee supply chain is shown in Figure 17. Local communities' reaction to the studies ranged from pessimistic (a distracting waste of time and resources), to incredibly enthusiastic. Clark claims:

> The maps created by this tool have great potential to provide immediate information about what is happening on the ground. The sociograms or visual graphics are able to capture the attention and imagination of rural actors whose literacy levels are sometimes very low, which may act as a disincentive to participate in research projects or read written reports. The maps can help outsiders to identify key actors in the local network, which can be a good guide for deciding who to involve in a project or at least who to consult during the initial stages in the hope

[13] More information about the PIPA approach is available at: http://boru.pbwiki.com/Draw+network+maps

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		KFSRE team members	Farmers with cattle	Farmers without cattle	Commercial Farmers	San Bushmen	Producer Org members	Foreign NGO staff	Extension staff	NDC staff	Mashare ADI staff	Traders	NOLIDEP staff	Forestry staff	Researchers (plant)	Researchers (livestock)	Fisheries	University staff	Southern Commercial Farmers	Project staff from other regions	Foreign Res. Org. members	Farmers outside region	DFID Plant Science Researchers
A	KFSRE team members		*	*			*	*	*		*		*	*	*	*				*	*		*
B	Farmers with cattle	*	**	**	*	*	**	**	**	*		*	**		*	*						*	*
C	Farmers without cattle	*	*	**		*		*	*													*	*
D	Commercial Farmers		*	*	**	*	**			**		*?	*?									*	
E	San Bushmen	*		*		**		*														*	
F	Producer Org members	*	**	*	*		*	**	*			?			*								
G	Foreign NGO staff	*	**	**		*	**	*	*		**			*	*					*		*	
H	Extension staff	*	*	*			*	*		*	*	*	*	*	*	*							
I	NDC staff		*		**							*?											
J	Mashare ADI staff	*						*	**				*	*	**	**				*			*
K	Traders											*											
L	NOLIDEP staff	*	**						**		*					**							
M	Forestry staff						*																
N	Researchers (plant)	*							*		*				*	*			**		**		*
O	Researchers (livestock)	*							*		*		**		*	**			**		*		
P	Fisheries																						
Q	University staff																						
R	Southern Commercial Farmers		*	*	*	*							*		*	*							
S	Project staff from other regions	*						*															
T	Foreign Research Org members	*													*	*							**
U	Farmers outside region		*	*	*	*																	
V	DFID Plant Science Researchers	*	?	?	?										*						**		

Information flow indicated by * (more stars indicate higher level of flow).
Areas to strengthen research capability are shaded. Question marks indicate areas of uncertainty.

(a) linkage matrix

(b) network diagram

Fig. 15: (a) Biggs and Matsaert's "linkage matrix" (sociomatrix) assessment of research linkages among farming research groups in Namibia [37]. (b) A network diagram representation of the same relationship data.

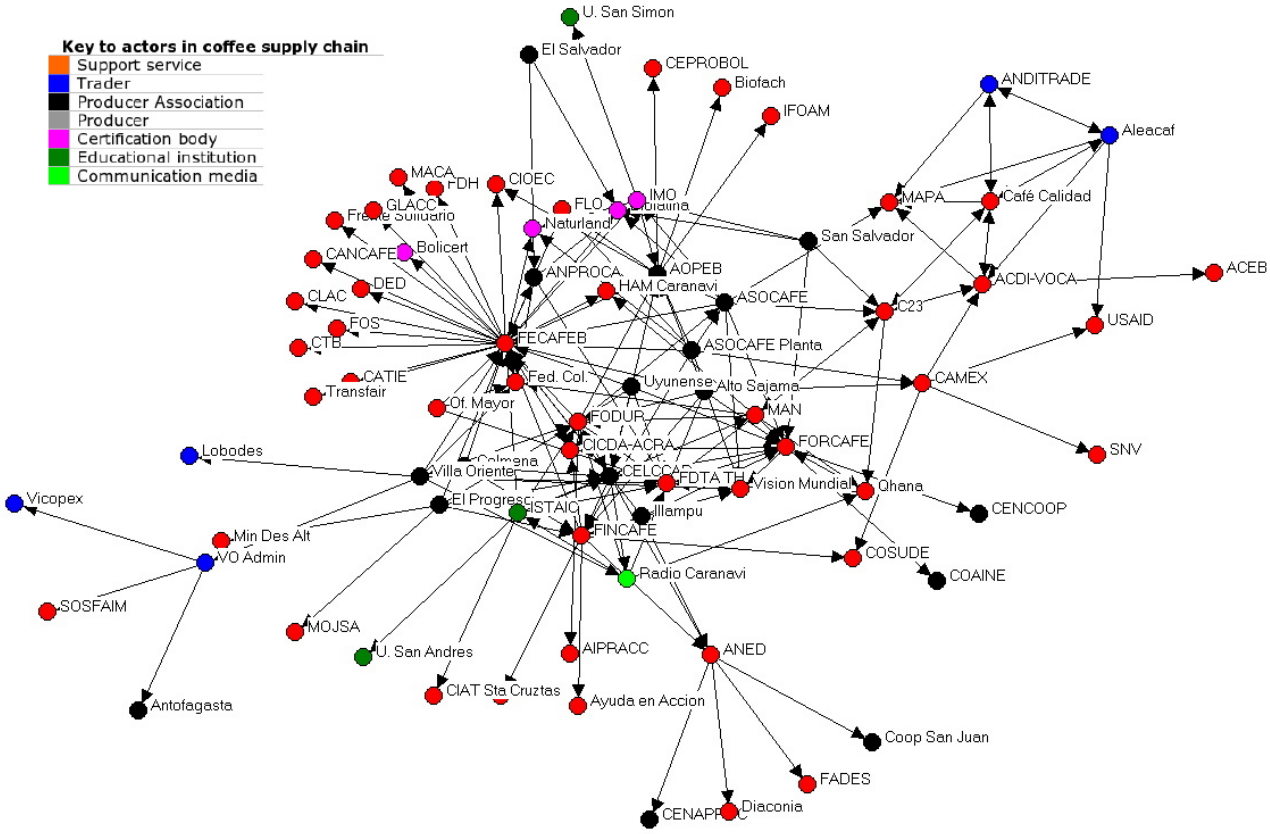


Fig. 17: A sociogram of the network of relationships in Carnavi Bolivia [39] generated from responses to the question: "Who provides you with information related to the coffee supply chain?" The "star" of coffee supply chain information collection and distribution is FECAFEB (Federation of Bolivian Coffee Exporters) in the upper left. The cluster of actors on the upper right are various projects funded by MAPA (Market Assistance for Poverty Alleviation), USAID's Alternative Development program—which supports coffee production as part of their coca eradication strategy. From the sociogram, it is clear that these projects have not integrated into the wider coffee network, and are not trusted by the central cluster of actors. (Clark, personal communication.)

> that involving them in planning and implementation will lead to better participation and more rapid appropriation of innovations. This visual material is also a good starting point for encouraging key actors to discuss their relationships and the benefits of working together to construct a network based on common needs. [40]

Each of these authors seemed optimistic about the potential uses of the technique for helping the NGOs and the community groups. However, all of the trials involved relatively small groups so it is not clear how well this approach would scale for larger projects. Several authors remark that in order to be effective, the network maps must not be too dense, requiring that the relationship questions be limited in number and closely targeted to yield sparse networks. (See [6] for good examples of targeted questions.)

It is also interesting that in some situations the idea of anonymization was not well received:

> On the African village level, people have no concept of anonymizing something and when I told a traditional chief that I would tell no one that it was him who said this and that, he was seriously insulted, because he wanted to be recognized for what he told me. (Eva Schiffer, personal communication.)

Key tools used: Network visualization, data collection, extensive interviewing, policy analysis.

Why relevant: Participatory network mapping could be an effective tool for human rights organizations to use in strategic analysis, particularly for helping grassroots groups to identify and articulate power structures and obstacles.

Evaluation of connectivity within and between organizations.

Network mapping and analysis has been used for several years as a tool for understanding communication patterns within and between organizations. These types of studies mostly take the form of one organization (or researcher) attempting to gain a rigorous overview of the relationships between actors in some domain in which the researcher may not be a direct participant. There are active fields of research which apply network analysis to evaluate relationships between NGOs in the realm of policy networks by examining lobbying, coordinated strategies, or coalition formation. Such analyses can also examine coordination among NGOs involved in international aid intervention projects.[14] Similar techniques are used to examine the relations within groups that help or hinder organizational effectiveness. To give a few examples:

- Researcher Jessica Crowe uses interviews with carefully selected members of small rural towns in Washington state to construct leadership interlock networks of community organizations. These networks were used as measures of the town's cohesiveness in order to examine the relationship between network structure and the amount of community economic self-development [42].

- A small scale but fairly sophisticated network-based study of cooperation relationships among eighteen "civil society" NGOs in postwar Croatia in 2000 and 2002 found a general decline of collaboration between organizations, and from this draws some specific lessons for international aid funders [43].

- A study of the relationship networks among aid agencies responding to floods in Mozambique in 2000 [44] attempts to determine whether organizations' "potential for communicative activity" was reflected in increased aid beneficiary numbers.

- On the inter-organizational level, it appears that network mapping techniques may be useful even within organizations that maintain strict and well-defined hierarchies, such as the U.S. Air Force. Colonel Elisabeth Strines argues that some degree of "horizontal" communication is very important for efficiency, and that mapping out the unofficial cross-unit ties as networks can be helpful in identifying soldiers with leadership qualities [45].

There are definitely some potential pitfalls to using SNA for these types of evaluations. To use the Mozambique floods paper as an example: The authors were interested in coordination between aid groups, so for their analysis they constructed *affiliation networks* in which organizations were linked if they worked on projects in the same district. This is a creative approach because information on project locations was readily available and actual communication events among groups would have been very difficult to measure. A concern with procedure is that although the

[14] Evaluation consultant Rick Davies is a strong proponent of the use of network analysis to evaluate aid interventions [41], and is also a co-developer of the Impact Pathways network approach mentioned above.

district affiliation can indeed be represented as a network, it may not be closely related to the communication network at all. Although it may be likely that groups working in the same district would need to communicate, the assumption that district affiliation can be used as a proxy for a communication network was a very big leap to make, and seriously weakened the paper's SNA-based claims. For affiliation networks to be useful they need to be closely related to the substantive relationships researchers are making hypotheses about.

An example of a typical use of affiliation networks is an exploratory study of relationships between actors involved in communications and information policy [46]. The study asked participants (identified from organizations' websites and lists of conference attendees) to name the events (conferences) they had attended as individuals and the organizations they worked with. This information was used to construct various types of affiliation networks: people linked by communication, mutual attendance at events, etc. To an outsider unfamiliar with the field, the network image that seemed most informative was a map of individuals linked by the number of organizations they named as important to their advocacy work. If we assume that people doing similar work rely on similar organizations, one way to think of this network is that it gives a picture of how similar the work of each individual is. Indeed, the authors were able to identify dense clusters corresponding to people working on human rights campaigns and internet development advocacy. No detailed conclusions were drawn from the analysis, since by the author's own account the work is fairly preliminary.

Another creative use of affiliation networks as a means of examining organizational relations appears in the research of Michael Heaney and Fabio Rojas on co-mobilization by U.S. anti-war groups [47]. Their study surveyed carefully selected samples of participants in large-scale public demonstrations during 2004-2005, asking them (among other things) which organizations had contacted them to participate in an event. After aggregating the data across five protests in several cities, they were able to construct a network showing which organizations are linked via protesters who are on multiple email lists. (Fig. 6). Not surprisingly, the major sponsors of the protests appear very centrally in the network. Individuals involved in the demonstrations agreed that many of the other apparent clusters made sense as well. Because the authors also asked the participants about their other activities, they were able to draw some conclusions about the tactics employed by the different segments of the anti-war movement with regards to participation in partisan support, lobbying, etc.

Key tools used: Surveys, affiliation networks, communication networks, SNA, policy networks.

Why relevant: Gaining an overview of the contacts and alliances between organizations could help groups navigate politics more effectively. It also could help outsiders understand the jurisdictions and similarities among organizations.

Social networking for visibility and fundraising.

Social networking sites are wildly popular—they now make up a large portion of total internet traffic. Some demographic groups[15] seem to prefer social networking sites to email or instant messaging for communicating with friends. Use of these sites is not limited to individuals: many organizations now maintain profiles as well, making it possible for supporters to demonstrate affiliation by linking to them.

Although the networking sites are corporately owned (serving as quasi-public space such as a shopping mall) and generally subject to strict terms of use, a number of the sites have made key aspects of their technology public so that software developers can write small programs to be displayed on users' pages. Several non-profits, advocacy groups, and political campaigns are using social networking sites to gain exposure and attract new members. For some users, a link to an organization's profile serves as the online equivalent of a bumper-sticker or lapel-pin awareness campaign. More advanced networking tools have been used to raise funds or mobilize turnout for campaigns. An example is a small widget to be displayed on a person's profile page (see Fig. 19) in order to show how much that individual contributed and also how much they have indirectly contributed by convincing their friends to give—a form of campaign finance "bundling" for the masses. Amnesty International has over 366,000 members on Facebook,[16] and maintains multiple "Facebook Causes": specific campaigns that members of the service can join and display. The users of social networking sites tend to value irreverent and ironic content, often making it tricky to pitch serious issues. For example, the membership in the "Foundation for the protection of Swedish underwear models" cause is still considerably larger than that of any human rights group. But there does ap-

[15] Various social groups have different preferred social networking portals depending on age, nationality, and other factors [48].

[16] Amnesty International's Facebook organizational profile: http://apps.facebook.com/causes/beneficiaries/161.

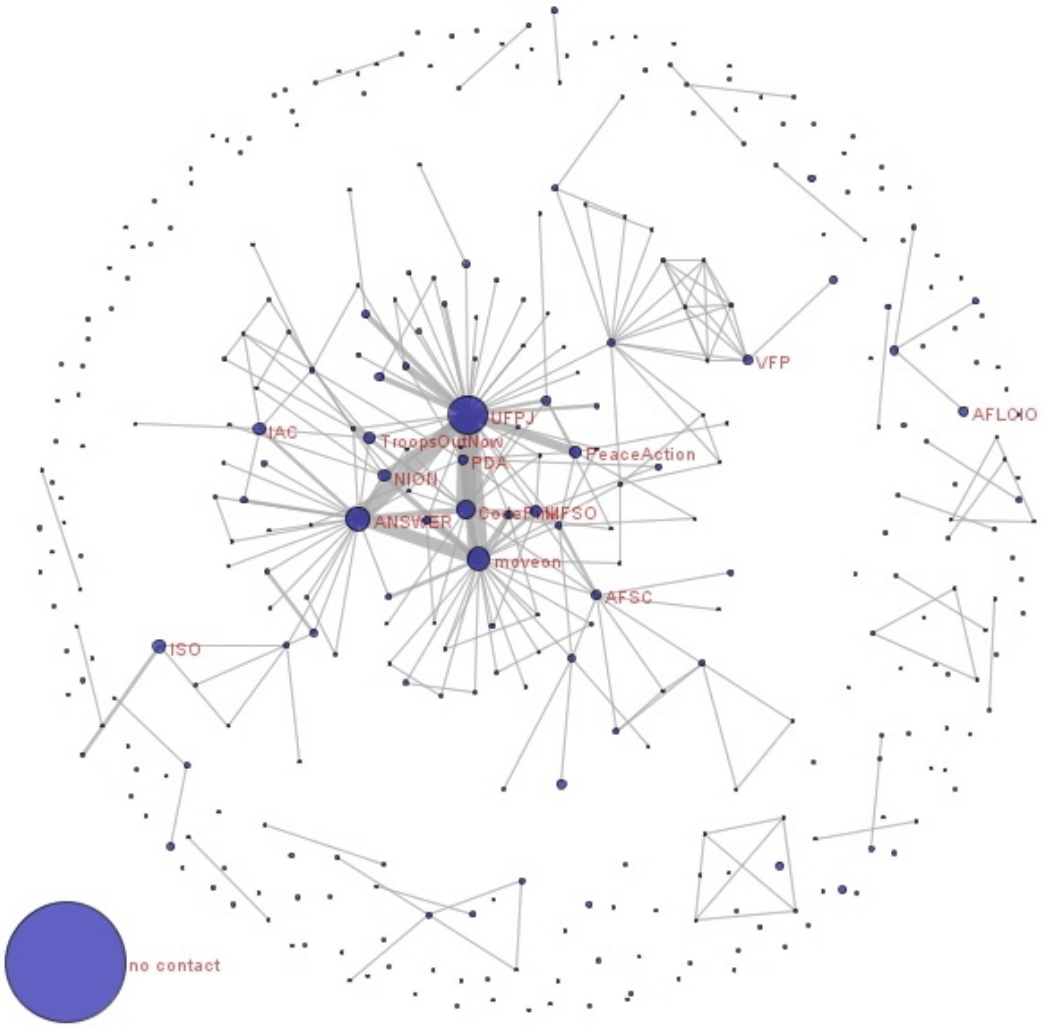


Fig. 18: Co-contact network of groups that encouraged people to attend large-scale anti-war demonstrations in 2004-2005. The size of each node indicates the number of respondents at the demonstrations who said they were contacted by that group. Thickness of connecting lines indicates the number of respondents contacted by both groups. The "no contact" node corresponds to individuals who did not list any organization, possibly receiving information from other sources such as media or friends. Data from [47].

Fig. 19: An example of a user's "wall of fame" from the Facebook Causes application [49]. The images is displayed on a user's main profile page and automatically updated when they or their friends make contributions.

pear to be a difference in the level of financial commitment: Amnesty's Facebook members contributed considerably more, over $14,000 by April 2007, and several other human rights campaigns seem quite financially successful as well.

Many organizations evidently believe that there is large outreach potential due to the sheer numbers of users involved in networking sites and are willing to investigate new techniques. A few (such as the Buma Campaign) have tried online organizing of real-world demonstrations, with mixed results and fairly low turnout. As the ability to post and share multimedia content via profile pages is now widespread, there is also some potential for use as a vehicle for distributing first-hand information about human rights events as they occur. An example of a rich-content website with a social networking component as well as a strong human rights focus is Witness's The Hub,[17] which hosts user posted human rights videos with comments and view tracking.

Another potential use that has received some minor attention is that human-rights workers or political dissidents who are potential targets of violence or disappearance can use networking sites that permit posting messages via SMS (cell phone text messages) to inform others of their continued well-being and local activity. Hopefully the person's activity would be not clandestine, as the information would be available to opponents as well as supporters.

Key tools used: Social networking, information about friendship networks.

Why relevant: Many groups are already using social networking as a tool for awareness, outreach and fundraising. Yet there may also be potential benefits in creating specialized groups or services for use as communications tools within the community of human rights groups and NGOs.

Public health and disease spread.

Network models are being used to understand how the structure of human contact networks aid or discourage the spread of disease. There are multiple studies of sexual contact and intravenous drug-sharing networks that have yielded some intriguing results. Although detailed personal relationship data is difficult and expensive to collect, it can be combined with sophisticated statistical modeling of networks to explore the population-wide consequences of individual behaviors. One finding is that the threshold for epidemics (which are formed when the networks of

[17] Human rights video sharing and networking website: http://hub.witness.org

contacts are dense enough for the disease to spread quickly) is highly dependent on subtle variations in network properties [50] (Fig. 20). If the network structure can be altered to reduce the number and size of the connected components, the disease is unable to jump quickly between clusters and the epidemic is slowed or halted. This is encouraging, because it suggests some ways that people can modify their behavior such that relatively minor changes can yield huge gains [51]. Uganda has pioneered this approach, implementing an education program called "zero grazing" that encourages people to avoid having multiple simultaneous partners (or to use protection outside their primary relationships) as an effective alternative to abstinence. The program has been partially credited with Uganda's success in reducing the extent of the HIV epidemic.

Key tools used: Network statistics, modeling and simulation, diffusion, ego network sampling.

Why relevant: Network models may suggest more effective ways to apply limited resources to achieve effective health care policy. Some models of information and innovation diffusion are very similar to disease models, suggesting some possibilities for communication contact structures that might encourage the spread of desirable practices.

Innovation spread and adoption.

Studies of innovation diffusion and adoption generally focus on patterns of information spread by communication contacts among people, or between media outlets and people. The mathematics of innovation diffusion are closely related to those of disease spread, except that the "epidemics" are of ideas rather than viruses. There are many studies that use social network analysis to help track the diffusion and adoption of practices such as health procedures, HIV prevention, contraception, and farming innovations.

An excellent example is a paper evaluating the effectiveness of family planning programs in a small village in Madagascar [52]. The programs train local community members to teach about and distribute contraceptive products. Normally, similar programs are evaluated with aggregate measures of contraceptive use and awareness, which make it hard to determine why some programs do well and others do not. The Madagascar study employed a network-based evaluation, interviewing most of the community members about whom they asked for advice and information, in addition to determining how well-informed they were. Analysis of the family planning advice network showed that both of the members who had been trained were centrally located, indicating that many people went to them for advice. Also, the closer other villagers were to them on the network, the more they knew about family planning. Surprisingly, it was the people who had primary advice ties to contacts outside the village that were most likely to actually *use* family planning.

Ideas, opinions, and news items are also types of innovations, and in some cases they can be tracked as they appear in mass media and spread through blogs. It is interesting to consider the ways that the structure of social and communications networks impacts the spread of opinion and influence. Krebs has published a review of recent work on political conversations and voting decisions [53], concluding that telling others how you intend to vote is quite likely more important than actually voting yourself.

Key tools used: Interviews, network analysis, text mining, network modeling.

Why relevant: Understanding the spread of information and social practices can also aid in the design of positive or negative interventions.

Literature overview and large corpus document searching.

Network visualization is often proposed as a solution for creating visual summaries of quantities of text or collections of documents too large for a single person to review. Ideally the network representation can serve as a map to help users explore an unfamiliar field and quickly identify the most important sources to read, determine who are the most influential opinion leaders, or gain a quick overview of the main concepts and relationships.

There are a variety of techniques used to generate network representations of text. As mentioned in the section on legal networks, there are multiple research groups with tools under development that use network representations of citation patterns to analyze large numbers of documents and determine which are most influential. In some cases the network structure produced by the citation network can be aggregated to show the broader patterns of communities and disciplines (see Fig. 21). A citation-based approach can work well for documents such as academic papers which have fairly standardized structures and include detailed bibliographies, but is much less effective for other kinds of text.

In some situations, the text is semi-structured. Many types of documents include abstracts, lists of keywords, category codes, indexes, or other kinds of markup that provide some help in summarizing their contents. These categorizations and keywords can of-

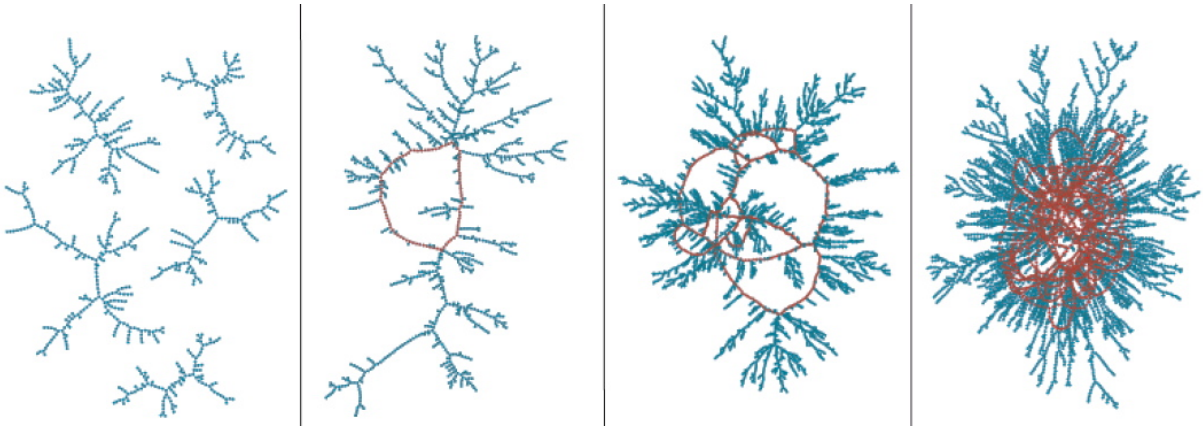


Fig. 20: A series of simulated networks showing the size of possible epidemics as the number of people with three partners is increased (from left to right). Ties represent possible paths of infection. The red ties indicate "robustly connected" regions where multiple paths connect people—greatly increasing the likelihood of an epidemic. The average number of partners per person increases from 1.68 on the left to 1.87 on the right. This relatively small change causes a big change in the size of the largest component (largest number of people who could be reached by an epidemic), increasing it from 215 to 6433 people [50].

ten be used to construct affiliation networks in which links between the keywords or categories are established when they appear in documents together. This approach can give a good picture of how the categories are related, but only at a fairly coarse level, since it is usually not possible to distinguish between the types of relationships described in the text that created the co-occurences of the keywords. In some cases, when the terms of interest are fairly unique (or have a limited number of alternate spellings), it is possible to use a small number of search queries across a set of documents to find when terms appear together, or to locate passages for a human reviewer to code. A combination of the above approaches have been successfully used to extract ownership relationships between firms in the communications industry [30].

As mentioned above, finding useful relationships in unstructured text is difficult. One method is to employ techniques from computerized *natural language processing* (NLP) to automatically "read" text and tag certain kinds of relationships according to sentence structure. Although effectiveness has improved dramatically in the last several years, computer programs are generally not yet able to pick up on some of the important subtleties that human readers intuitively understand. However, there are certainly cases where tools that have been carefully refined for a specialized domain are able to perform well enough to be useful. For example, although "understanding" text is very hard for a computer, it is sometimes feasible for well-designed software to correctly locate and code specific instances of narrowly defined relationships, even making the crucial distinctions between examples such as "A tortured B" and "A was tortured by B."

There are also a number of ways to construct networks based on similarities between documents. The core idea is to count the frequencies with which various terms occur within a set of documents, or within blocks of text. By comparing the number of terms that match in every pair of documents (and correcting for how frequently the terms occur in general), it is possible to measure how related each of the documents are to one another. This set of relationships can be treated as a network. Usually, to draw this kind of network it is necessary to simplify it by removing some of the less important ties. This can be done either by choosing a cutoff point and discarding the weak relations (thresholding), or with more sophisticated techniques such as calculating *pathfinder networks* [55] that preserve only the more direct similarity ties between documents. This general approach of calculating relationships (and occasionally deriving networks) between documents, terms, and concepts (defined as related groups of terms) is usually called *latent semantic analysis*.

Another possibility is to distribute the search and classification task across a large number of well-trained people. One of the best examples of human rights text mining (and of data mining in general) is a program headed by a former SHRP employee, Patrick Ball. Several projects, most recently with the Hu-

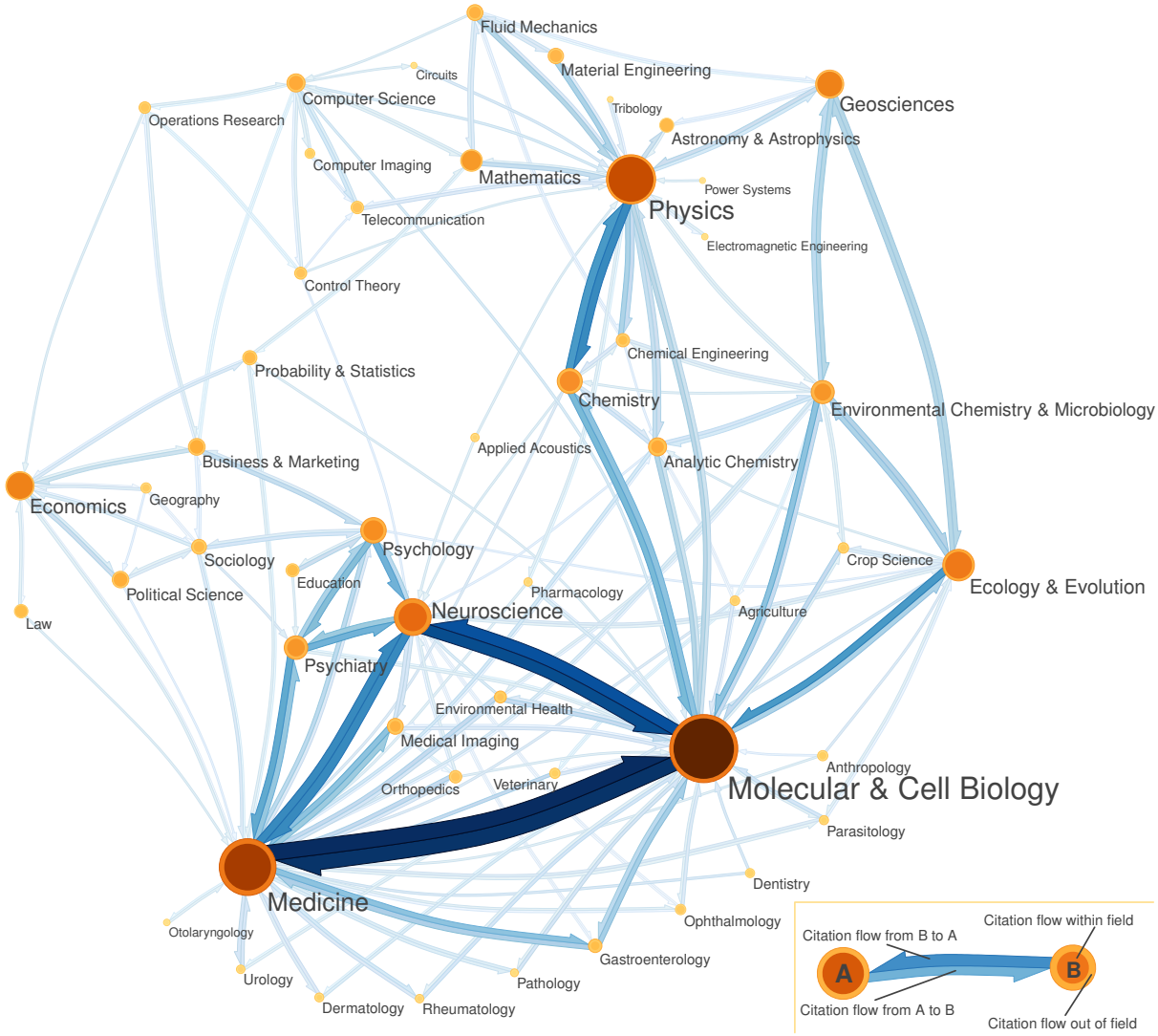


Fig. 21: A network map of the citation relationships between academic fields created by Carl Bergstorm and Martin Rosvall [54]. Orange circles represent fields, with larger, darker circles indicating larger field size. The size and darkness of arrows indicates the citation volume between the fields. The map was created using 6,434,916 links among 6,128 journals using data from Thomson Scientific's 2004 Journal Citation Reports. The journals were assigned to fields using an information flow methodology. For visual simplicity, only the most important links are shown. The map is online at http://www.eigenfactor.org/map/maps.htm.

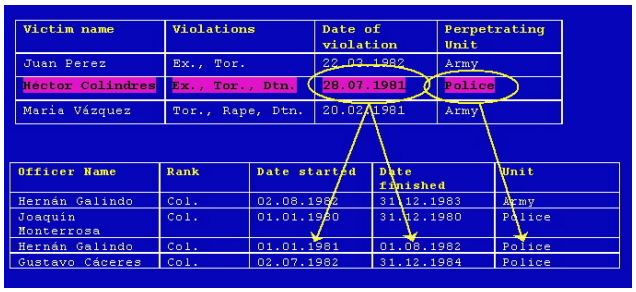


Fig. 22: Diagram of using a relational database to link event data from multiple sources. The date and other identifying information in victim testimonies can be automatically matched with career records to determine which individuals may have been responsible for abuses involving a specific place and time period. Illustration from Patrick Ball [56].

man Rights Data Analysis Group[18] (HRDAG), have used relational databases to compile large amounts of testimony and other kinds of human-rights documentation in order to make solid statistical arguments about human rights abuses. These projects required large data-entry efforts to get the documents appropriately coded in an electronic form. For many analyses, HRDAG employs a carefully designed vocabulary of event definitions which human rights workers use to transcribe the abuses they find while reading through each document. In some situations, it is then possible to cross-link the dates and locations of incident reports with military records in order to discover the names of individuals responsible for the rights violations (see Fig. 22). A single person would never be able to read enough of the material to make these connections, but the combination of rigorous information management and good database design makes the analyses possible. Interestingly, although many HRDAG projects appear to contain a wealth of relational information, it does not appear that any network analyses have been done with them.

Key tools used: Text mining, affiliation networks, attributional networks.

Why relevant: Could be useful for organizations that need to search, extract relationships, or locate influential people in large collections of primary source material (victim histories, bureaucratic files). Can also be used to locate relevant influential research literature.

[18] Benetech HRDAG http://www.hrdag.org/.

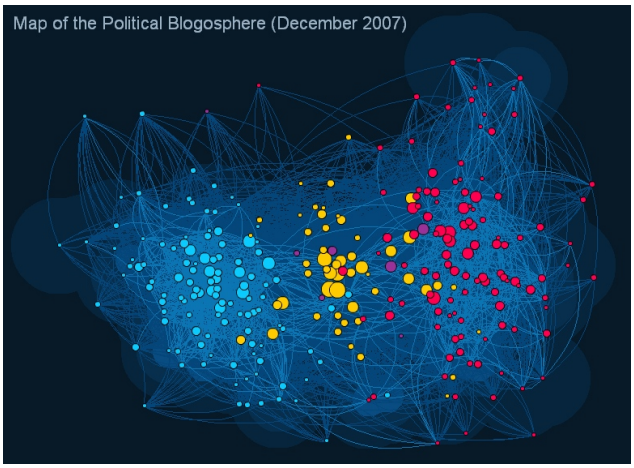


Fig. 23: Screenshot of an interactive network map of U.S. political blog connections with only important links shown. Progressive sites have been categorized as blue, Independent as purple, Conservatives are red, and general mass media outlets are shown in yellow. From Presidential Watch 08 [58].

Website link analysis.

Almost as soon as the web became widely accepted, a research practice of studying the structure of hyperlinks between websites emerged. The assumption is that organizations use links not only to guide viewers to related information, but also to position themselves with respect to other sites and groups. If this is true, then the pattern of links between organizations' web pages, sometimes combined with limited content analysis (text mining) of the pages themselves, can be used as a proxy for the network of affiliations between the organizations. This information is much cheaper and easier to collect on a large scale than interviews with organizations. Links can be mapped out automatically, either by starting from a few "seed" sites central to an issue or community of interest and using software to read, record, and 'crawl' the pages to discover new sites, or by using a roster of organizations and examining the links between them [57].

Web-link analysis has been used to study social movements. Garrido and Halavais map out the links between websites related to the Zapitista movement and calculate various SNA measures [59]. On the other end of the political spectrum Luca Tateo [60] maps out the relationships between Italian extreme-right organizations. Govcom.org has built an interesting online public tool called *Issue Crawler*[19] that can

[19] The Issue Crawler tool can be found at http://govcom.org.

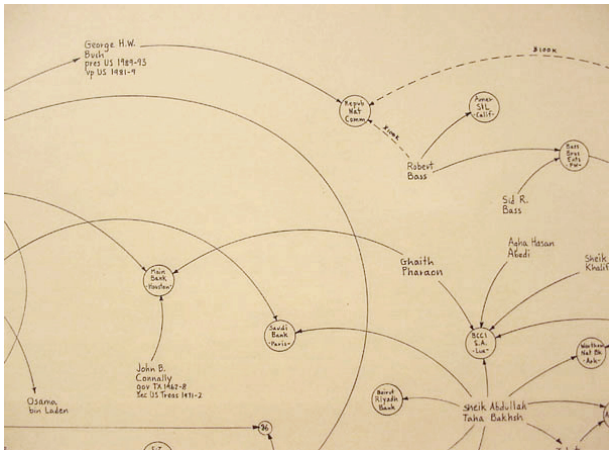


Fig. 24: A fragment of a drawing by Mark Lombardi entitled *George W. Bush, Harken Energy, and Jackson Stephens, ca. 1979-90 (5th Version)* depicting a set of related individuals, companies, and events extracted from the news.

automate the process of collecting web link relationships for policy analysis, including some rudimentary visualization. The group has produced several studies based on the Issue Crawler, such as one that combines keyword and web-link analysis in an effort to understand the marginalized position of Israeli peace and justice groups who are campaigning against Israel's construction of a massive wall in Palestine [61]. When these types of tools are used to look at the link structures in larger political domains, a common finding is a very strong segregation between sites with left and right political orientations, with small numbers of cross links and a few intermediate sites (see Fig. 23).

News-mining and event data.

When breaking news is published about current events, it is usually fragmented and incomplete—almost by definition. Frequently, only after the fact is it possible to gather the various conflicting accounts, determine the sequence of events, make an attempt to sift out the pattern of relationships that lead to a conflict, reconcile stories, and "read between the lines" of each side's self-flattering narration. Networks can be used as a framework to store and organize some of the complicated relational information needed to represent complicated events with many actors. In some cases, a network map can be used to communicate the kinds of summary information that are difficult to convey with narrative alone. An often-cited example of the application of network diagrams to reveal complicated facts is the collection of hand-drawn news-event maps by artist Mark Lombardi [62]. Lombardi carefully researched news stories on topics ranging from the Chicago Mob to the Iran-Contra scandal, collecting detailed notes about key individuals, events, and relationships, eventually rendering them as elegant poster-sized node-link diagrams in pen and ink (see Fig. 24). Several media outlets have also generated maps showing the links between entities appearing in current events, attempting to extract meaningful relationships from automated coding or human-tagged news feeds and blog contents and presenting them in network form. A reasonably good example, is Libero's news map (Fig. 25) which displays a network of entities, mostly people and place names, associated with a given search term over a specified time period.[20] Google News takes an alternate approach: grouping news stories by their similarity to form hierarchical clusters of topics. Newsmap[21] also provides a visualization of the news clusters, displaying current topics as sets of nested rectangles corresponding to each story. Confusingly, this style of visualization is called a *treemap*.

These kinds of data are available for political science as well. Political scientist Gary King has released an extensive database of worldwide news events [63], each event is summarized in the data as "Actor A does something to Actor B"—the actors are the major political entities in 450 countries—and "does something to" is coded using an ontology (well-defined vocabulary) of 200 types of actions. Although the data are automatically coded from the "lead" sentences of Reuters news reports by VRA,[22] King's analysis seems to show that the computer procedure performs very well when the classifications are tested against those of human experts. Other researchers have used the data for projects such as examining networks of NGOs and countries that are linked by events [64]. All authors seem aware that the events found in news sources will be subject to certain biases and will not provide a completely accurate picture of what is actually going on.

Not surprisingly, these approaches have been used to search for patterns of human-rights related events. Rather than using an automated tagging process, the FAST[23] project of the Swiss Peace Foundation em-

[20] "Human rights" search example on Libero's news browser: http://arianna.libero.it/graph/search/abin/graphnews?query=HUMAN+RIGHTS (results will vary).

[21] Newsmap Google News browser: http://www.marumushi.com/apps/newsmap/.

[22] Virtual Research Associates sells analysis and tools for news wire data: http://vranet.com/.

[23] FAST—Frühanalyse von Spannungen und Tatsachenermittlung. A German acronym for Early Analysis of Tensions and Fact-finding.

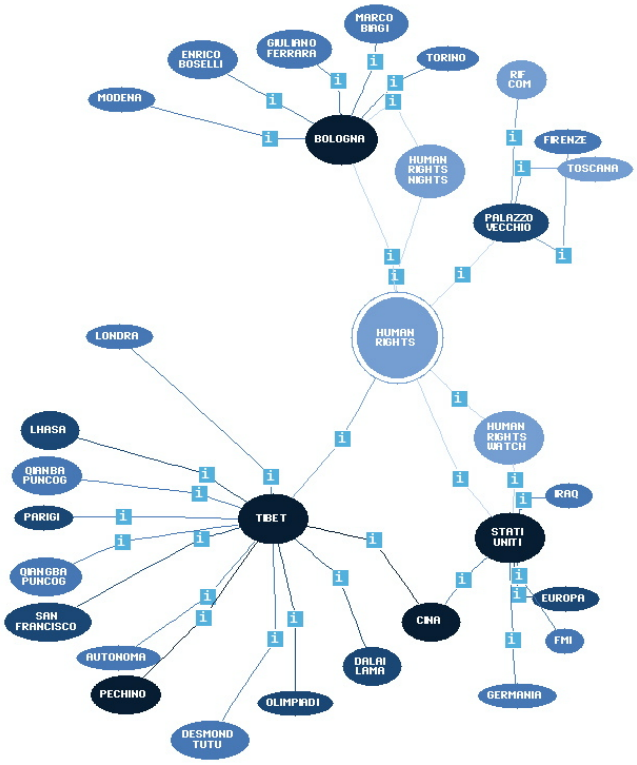


Fig. 25: An example of a network of relationships extracted by the Italian-language news service Libero. The map shows entities found in news stories from April 9, 2008 related to the term "Human Rights". Clicking on the ''i" boxes on each line loads a list of news stories establishing the link between the nodes. The cluster at the top corresponds to an ongoing "Human Rights Nights" festival in Bologna. The protests in multiple cities about the Olympic games and Chinese repression in Tibet show up in the groups of nodes on the bottom left.

ploys a large a team of local professionals in twenty countries to monitor, code, and report on local news stories and events. Hämmerli and others [65] constructed networks using this event and actor data from Chechnya between 2002 and 2005. They used various SNA measures on the networks to determine the relative importance of various groups in the conflicts. The coding of the data made it possible to examine conflictive and cooperative interactions independently, but the authors did not employ any dynamic network analysis.

Key tools used: Text mining, dynamic networks, actor networks, policy networks.

Why relevant: Extracting networks using specific human rights events may make it possible to communicate quick snapshots of key relationships within a region or issue, and perhaps track how the structures change over time.

7 Conclusions from case studies.

In the last several years there has been an explosion of the use of network methodologies and general awareness of network concepts in multiple research areas. It is difficult to draw any comprehensive conclusions from work drawn from so many methodological backgrounds. Even an effective categorization remains elusive, as individual case studies in similar areas often use very different techniques, and the same techniques are applied in areas that are only superficially related to each other. The following are a few key points related to network mapping gleaned from this review:

- Although many people are advocating that network techniques will help a great deal with evaluation tasks, there have not been any large scale systematic studies comparing the various pilot projects. Most projects have been fairly small, both in sizes of networks and numbers of participants.

- The majority of projects appear to be doing more diagnosis than assessment. This may be partially because in most cases there is no "known standard" for comparing assessments. Also, researchers tend to be cautious because the data collection is not rigorous or comprehensive enough to be safely used for evaluation at this time.

- Although a few of the academic studies show high methodological maturity, much of the work still seems quite exploratory. In several papers the

insight appears to come more from the in-depth interviews and data collection process, with the network analysis component serving as a parallel approach.

- A major impediment to use and acceptance of the techniques is the amount of time researchers and participants must invest to get good quality network data. The cost of software tools are a smaller barrier, but a significant one in some countries.
- In many cases the data collection process can be integrated with discussion and planning in such a way that the meetings themselves have direct value for the participants. This makes the addition of a network component much less burdensome—especially if the results are immediately available to stimulate discussion.
- There seems to be a split between those who employ network maps, and those who do more formal types of network analysis. Although a few of the concepts underlying network measures are employed frequently, the measures themselves seem to be mostly used by trained specialists and academics—a pattern similar to the use of high-level statistical concepts.
- Reactions to the network maps range from unrealistically enthusiastic to dismissive. There is a lack of clear guidelines on how to create and read network maps (reflecting a lack of consensus in the visualization research community).
- In most cases the visual quality of the maps produced is actually quite low. This is partially due to the lack of clarity in the output of the widely available software tools. It might be possible to improve the utility of network mapping considerably by providing people with more legible tools.
- Techniques (such as Net-Map) that collect relationship data through a visualization process seem quite effective. However, because dense networks are difficult to visualize, the results visual data collection may be somewhat distorted towards simple networks. A rigorous comparison with other collection techniques would be helpful to explores their relative biases.
- Tightly focused relationship questions, and/or clearly defined tie types seemed to improve the utility of the network maps. Including many different kinds of ties in the same diagram tends to reduce legibility.
- Although the inclusion of names and other identifiers on network maps can provide serious risks to the named entities, the visualizations that appear to be most effective are those that include enough contextual information for viewers to construct explanations for the structure they see, as well as locate "known" positions or "landmarks" in the maps.
- It appears that there are several existing collections of human rights-related data containing relational information that could be used to test the utility of visualization techniques.

8 Some issues and risks: what network analysis can not and should not do.

The use of network analysis has raised a number of justifiable ethical concerns [66]. Although it is a powerful and promising methodology, it has also received a lot of undeserved hype. Network analysis will not magically locate people who are responsible for human rights abuses. It is only useful as a technique when high quality data about relationships are available. This data may be very difficult or impossible to collect even when the analysis involves subjects who are willing and able to participate in good faith. Network data is challenging to analyze. Thinking relationally can be difficult, and the appropriate conclusions are frequently counterintuitive—this can be especially true for people who are highly skilled at working with data in tabular form. Many standard statistical techniques can not be appropriately applied to network data because the relations in networks violate some of the assumptions of "event independence" that underly standard statistical tests. Although there are "network-safe" variants of some statistics techniques [67], many of these methods are still under development and are difficult to understand and use. Table 6 gives a quick summary of some common issues.

Obstacles for human rights use of network mapping and analysis.

The information needed to do a network analysis effectively is usually fairly sensitive, requiring individuals to divulge personal details and opinions. Organizations are often asked to disclose their allies, enemies, and funders. This information can be difficult to collect even in situations where everyone involved is already convinced of the benefit of creating a shared

Table 6: Common errors by users of network mapping and risks for participants

Oversimplification and misreading	Viewers are not used to thinking critically about network images. Like any statistical graphics, they can be easily manipulated to convey a viewpoint that would not hold up well to rigorous analysis.
Making evaluations from incomplete data	Network data is highly sensitive to missing cases, changing a few ties can dramatically alter a map. Viewers tend to make evaluations (Who has the most links? Who is the biggest?) rather than searching for patterns.
Lack of anonymity	Network data is usually not aggregated so identities are not disguised by being lumped into an average. Even when names are not included, they can often be deduced by those who know enough about the topic of the network.
Using wrong measures for network type	Measures must be matched appropriately with network types and research questions—betweenness scores for an affiliation network should not be interpreted the same was as those for a communication network.

representation. In many imaginable human rights scenarios, the individuals or organizations in the network may be the subjects of an investigation or involved in an ongoing conflict. They may be completely unwilling to divulge information that could threaten their objectives or security, and may work to prevent data from being collected by refusing to participate, giving misleading information, destroying documents or intimidating researchers.

Ethical questions.

The collection of relational data for use in network mapping or analysis raises a number of complicated ethical questions. The positional information contained in an analysis can be of strategic importance, and in some instances the names of the individuals or organizations involved may be sensitive information. Kadushin [68] gives an example of waiting to begin a study on U.S. intellectual elites because of concerns that it might be used to politically target individuals suspected of being against the Vietnam war.

Although some of these concerns can be dealt with by using the informed-consent and sensitive data collection designs employed by other modes of survey research, there are some additional challenges for network data. For example, although the names and identifying information can be removed before making results public, they must be retained within a research project to permit the appropriate linking of data. Even when anonymization is employed, it is often possible to "reverse-engineer" the network—rediscovering the names by combining outside sources with the attribute and linkage information. Also, survey respondents are usually asked to provide identifying information about the individuals they have links to. These people will be included in the study (and perhaps put at risk) without having consented to become participants.

Even information collected with positive intentions can be misused at a later point in time. Network techniques can be used to target activists and human-rights workers in the same way that they can be applied to bring abusers to justice using data uncovered in police-state archives. There are a number of technical safeguards which can be employed to mitigate the risks of archiving sensitive data—such as one-way encryption of names and using numerically coded coupons for recruitment in snowball sampling—but the problems need to be solved politically as well as technically.

In some cases, data are gathered from publicly available sources. Although such procedures may be very defensible legally, it is still likely that the individuals or organizations appearing in the data did not imagine that their information would be used in such a

fashion. The combination of data from multiple public sources can also reveal unexpected features that put individuals at risk. In most cases the risks may be fairly diffuse or merely embarrassing, but if the generated metrics and maps are going to be used to evaluate individuals and organizations (perhaps for firing, promotion, or grant making), the potential impacts should be discussed with the participants before starting the study. In situations where network techniques are being used to initiate criminal proceedings or make an individual a military target, the researcher is placed in a difficult ethical position. The U.S. military, for example, is reportedly deploying anthropologists to perform network mapping as part of its operations in Afghanistan. The American Anthropological Association has expressed various ethical concerns about the program and members' involvement [69]. A very pragmatic concern is that as information about the utility of network mapping and analysis becomes widespread—or if there is a perception that the data being collected will be used in an evaluative fashion—respondents will have incentives to misrepresent their relationships. For example, study participants may report on who they "ought" to talk to instead of their actual communications partners.

Data sensitivity.

Another crucial element to remember in network mapping is that it is often highly sensitive to missing or intentionally obscured data—far more so than traditional aggregate measures such as averages. This means that conclusions about who the most central nodes are, or the identification of the most important connecting ties may change dramatically with the inclusion or elimination of other data elsewhere in the network. In many cases, a single link change can reconfigure a network. It is important to carefully consider the sampling strategies and quality of the data that are used to construct a network. For example, if a single actor provides most of the relationship information, it would not be surprising to discover that they appear very centrally in the resulting network.

Determining the appropriate boundaries of a network under investigation can be a dismayingly arbitrary task. Geographic boarders are usually not particularly relevant for networks, and social groups tend to be hard to delineate (without using a network). For example, to produce a communication network based on email connections within a company, the employee roster might be a natural list of names to with with to start. But what about the employees in marketing who send lots of emails to prospective clients? Should the clients be included? Or what about spouses who might unintentionally communicate important information? It is important to remember that any network being studied is usually a fragment of a much larger network. There are usually not enough resources to track down all the possible relationships, so there will always be some nodes with ties extending "outside" of the network.

With any census procedure for an unknown population, it is difficult to know if everyone who meets a certain criteria has responded. Even when it is feasible to know ahead of time who should be included in a network, it is usually not possible to get everyone to reply. In most surveys involving humans, an 80% response rate is considered fortunate, but even this much missing data may cause validity problems for a network. One of the issues with networks is that missing data can skew data in very counterintuitive ways [70].

9 Evaluation criteria for projects.

The overall purpose of this report is to outline some potential research directions for human rights applications of networks and network mapping. The following sections present some suggestions about ways to classify and consider the various advantages and goals of possible projects.

Research and development required: deployability.

Some proposals could be implemented with known technology and existing tools, requiring mostly user training to become useful. On the other end of the spectrum, there are promising directions that would require significant basic research, or at least adaptation and development of techniques from other fields. Generally, development implies adding an element of risk, as a solution may turn out to be more difficult than expected or even not possible. When contrasted with mature fields such as cartography, there are lots of unresolved technical issues for network analysis and visualization. When geographic maps became computerized through the use of tools such as GIS (Geographic Information Systems) the problems mostly required technical engineering solutions—there was already hundreds of years of map-making experience to draw on. For networks, the techniques and technology are being invented simultaneously and the field is full of surprises.

Who can directly use the project: immediacy vs. complexity.

Projects designed to identify or mitigate human rights issues can be enacted at various levels of abstraction away from the effects of the problems. The most concrete and immediate projects would produce tools that could be used by people who are being directly impacted to improve their immediate situation. The next level might be projects that can help grassroots associations of impacted people—groups that are working to support one another in the affected areas. This would include education programs and other tools that assist with articulation and communication around issues. A bit more abstract would be solutions assisting direct aid groups and support agencies. At a further level of remove would be facilitating "big picture" policy analysis by human rights organizations and government entities. Perhaps the farthest level of all would be tools for the academic consultants and subcontractors who perform analyses and highly specialized research tasks.

Tools that are going to be of immediate use to large numbers of people need to be very robust and easy to understand, whereas those that are intended for use by specialists can require more training and skilled interpretation. These are often difficult trade-offs: training users is expensive, but creating good documentation and a tool that is stable and polished enough to be self explanatory can also take a tremendous amount of time.

Types of use: action – analysis – communication.

In addition to who uses a tool, it is helpful to consider how it will be used to ensure that it is designed appropriately. Although the following categories may not be mutually exclusive, they outline a continuum of types of intended uses. On one end are tools created to help people take action. These would be tools to support organizing work, perhaps serving a communication function among activists in the field or between organizations working on an issue. In the middle would be projects that facilitate documentation and collection of information—such as cataloging and linking abuse accounts. These are projects that help with fact-finding and evidence gathering. On the other end are analytical tools, projects that facilitate scenario development, long-term strategy, and planning. There is also a need for tools to help communicate with the general public about the discoveries made as a result of these analyses. Good tools are needed at all of these levels, but they may not be the same tools.

Resources required.

The resources question has several parts. It is important to consider the funding and staffing required to:

- design and develop,
- distribute and use (costs for end users),
- maintain and update

any of the proposed projects. For projects involving data, it is important to remember that some kinds of data may go out of date or quickly become inaccurate. These projects should be designed with this limitation in mind, or should budget for ongoing maintenance to keep the information updated and useful. Software tools tend to be expensive to develop or customize and cheap to distribute. But it is important to remember that software is rarely "stand-alone"—free software may still require expensive computer equipment to operate. Again, for many complex technical projects, the quality of the documentation is almost as important as the tool itself. For data-driven projects, the quality and completeness of the input data will greatly impact the length and duration of other project phases. As it is often difficult to appropriately assess the quality of data before a project begins, it is important to plan for re-evaluation when problems are encountered.

Feasibility.

Some projects may have no technical barriers, but would be impossible to complete because they would be perceived as politically sensitive, or because they require unrealistic expectations of cooperation among organizations that are competing for funding and political attention. Or projects might simply not be appropriate for SHRP to spearhead and would be better pursued by another organization. This kind of assessment is not within the scope of this report.

Scaleability.

It often the case with data-driven projects that an approach that works quite well on a small scale does not generalize well when it is expanded to larger communities. It is always important to consider how a project might fare as the numbers of users increases. This is especially true for network related projects,

partially due to some of the computational requirements necessary for working with networks. Many of the visualization and analysis techniques become dramatically more difficult as the size of the networks increases. This means that some things which are possible to calculate in a few seconds or minutes for hundreds or thousands of nodes may take months or simply be impossible to calculate for networks with tens of thousands of nodes. This is often a problem with software proposals: a demo which appears elegant and attractive can become completely unusable when faced with the complexities and scale of real data. When working with large networks, it is important to limit the choices of metrics carefully to those that can be feasibly calculated.

Conversely, there are many people-based projects that will only work well when larger numbers of people are actively participating. Extra planning may be needed on how to make a project useful in the interim period before it reaches the critical number of engaged users to make it sustainable.

Risks to participants.

It is important to consider how a project may create risks for participants—whether voluntarily or involuntarily included. The section "Some issues and risks" above discusses some of the issues in more detail.

10 Project recommendations.

The following examples are suggestions of projects that appear to be feasible applications of current network-related technology to human rights domains. A summary of key elements of each proposal is presented in Table 7. Of course, the actual relevance of each of these proposals for the work of human rights organizations would need to be evaluated by practitioners in the field. This list clearly does not exhaust the project possibilities. One way to solicit new proposals and approaches would be to curate and distribute human rights data to academics and publicize human rights-related problems as fields for research and technology development.

Project A: Develop an network-based assessment and strategy curriculum.

Network visualization appears to be an effective way of eliciting the tacit knowledge of a group and putting it in a tangible form for analysis and strategic action. The main goal of this project would be to collect, evaluate, and redistribute best practices from humanitarian organizations that already gather and map relational information. The organizations that are developing relevant approaches seem to be working independently with little opportunity to share experience and insight. In order to move the methodology forward, it would be useful to compile the experience that exists, identify the gaps, and help groups coordinate their research on developing techniques to fill them. The development process should involve cycles of testing with real applications and re-evaluation until solid procedures have been documented and can be shared with interested human rights groups.

The first stages of the work should be done in partnership with an aid agency or human rights organization. The existing practitioners and/or academic or commercial SNA consultants could be hired to work with the partner organization, with the understanding that the consultant's role is to document what they do. The project team would contact researchers and practitioners in order to integrate their experiences with the information from the consultants. (This might include arranging meetings of existing practitioners.) The goal of the team would be to produce several non-technical manuals, including templates for surveys, with information on how to acquire the necessary tools. These might take the form of "cookbooks" for relevant basic methods. Hopefully the resulting analysis methods would rely mostly on mapping instead of advanced statistics that could only be employed by experts. A good example to start with would be Clark's network mapping manual [40]. The project could have a continuing role, serving as a resource and coordination point for organizations who are using the tools.

The project would require at least 2-4 years to complete because there should be multiple design and testing iterations. The output would be mostly documentation. This means it would be expensive to develop, but cheap to distribute. Learning to use the techniques would require investments of time, but little or no capital outlay by the organizations that employ it. Ideally, the manuals would require little specialized knowledge and training—beyond the ability to use a computer. The risks are mostly in determining if the methods already used for small groups can be scaled up and made rigorous, while at the same time robust enough to be employed by non-experts. Staffing the project would likely require some expensive consultants in the early phases, a dedicated lead program person with writing and research skills, and probably an additional assistant. It would be difficult to determine what the costs of trials would be, hopefully the

Table 7: Overview of project recommendations, comparing evaluation criteria

	Project A	Project B	Project C	Project D	Project E	Project F
Duration	2-4 years	2 years, ongoing	1-2 years, ongoing	2-4 years	2-3 years	6 mo. - 1 year
Type of use	Planning, analysis	Communication, education	Communication, education	Organizational diagnoses, planning	Advocacy, analysis	Communication, analysis
R & D needed	Medium	Low	Low	High	Very high	Very low
R & D risk	Low risk	Low risk	Medium risk	Medium risk	High risk	Low risk
Who served	Grassroots orgs	Public, NGOs, policy groups	Public, law orgs	Grassroots, NGO, policy groups	NGOs	Public, policy groups
Development cost	High	Medium	Medium	Very High	Very High	Low
Use cost	Low	Very low	Very low	Medium	Unknown	Very low
Maintenance cost	Very low, mostly coordination	Medium, if updating	Medium, if updating	High	None	None
Risks to participants	Low	Moderate	Very low	Unknown	Moderate	Low

partner organization would be contributing resources as well.

The end results of the project would probably be utilized by grassroots organizations and NGOs planning medium- to long-term-human rights projects requiring problem analysis and strategy. The data collection phase will probably always be too time-expensive to be useful for those who are immediately affected by issues or organizations needing to make rapid decisions. A concern about this proposal is that the organizations already working to develop tools in this area—The Institute for Food Policy Research (Net-Map), The Center for Victims of Torture (tactical mapping), Centro Internacional de Agricultura Tropical (visualization manual), and the PIPA project—may see a new project as a competitor. This might be alleviated if the organizations and academics can see direct benefits from participating with each other in a larger project. Fortunately, those practitioners that are aware of each other have spoken positively of each other's work.

Project B: Create an interactive network atlas of human rights organizations displaying key projects and their relationships.

An atlas can be an excellent tool to give an overview—the big picture of who is doing what where. Because many organizations performing human rights work are working on local projects that also require international attention, geography may not be the best basis for such map. It should be feasible to instead create a who-works-with-whom map of the world of human rights NGOs and projects. Ideally, this map could be created as a web application built from a database of organization descriptions, contact info, and collaboration relationships, making the atlas into an interactive directory. Although a traditional directory[24] goes a long way towards providing this information, it doesn't show the reader the context—who the important players are and which organizations are doing similar kinds of of work. Furthermore, if the database was updated frequently, it would eventually accumulate a picture of how the relationships change over time in response to world events.

This project would have two main components: a) collecting and curating data, b) visualization and interaction design. The challenge for the data portion will be to assemble, clean, classify and standardize the relevant information. It may be difficult to determine which organizations should be included, and collecting the relationship data would require participating organizations to see the value in it despite the politics of turf issues and concerns about what it might reveal. It would be important to allocate some funds to maintain and keep the information up to date. Since the number of entities and relations will most likely be greater than what can be effectively put on the screen at the same time, the primary obstacle for the visualization side will be to find sensible ways to filter and aggregate the data.

The organizational communication necessary for the data collection could be done by a project manager, someone detail-focused having enough familiarity with the domain to do the classification. A network visualization specialist and/or graphic designer would be needed. In the web-based format, this project would also require a good web application developer. As the initial data collection might be quite time consuming, the project might take as long as two years to complete.

Project C: Global map of human rights law and legal precedent.

A birds-eye perspective on human rights law would help convey the history and relevance of human rights perspectives in global legal systems. This is a close relative of project B so it could also be done as an interactive website and or a series of graphic posters. The goal would be to combine both time and relational elements to show the spread of legal influence. Hopefully this would reveal the ripple effects and impacts of specific legal decisions as they are absorbed into practice in each country and court system. This would be an educational resource, and potentially a legal resource and strategy planning tool as well.

Carrying out the project would require compiling a multi-national database of legal arguments that cite primary human rights documents. Some data could probably be purchased from commercial providers, but will require a lot of normalization and matching. The project would need a very good graphics person or user interface designer, a specialist to manage data cleaning and research, and an international legal expert with expertise or connections to get information about the bodies of law in multiple nations. The bulk of the work would be in the research phase. A finished version would probably take 1-2 years, with some additional work each year to update the project.

[24] An example is a UN NGO directory: http://www.unog.ch/80256EE60057E07D/(httpPages)/3101491B86487F6D80256EFC0061DFD9?OpenDocument

Project D: Relationship collection and visualization software.

Creating a network map on a computer usually requires many complex operations using different programs to collect, format, and display data. Constructing and distributing a software tool that is designed from the outset to facilitate interview-based data collection—with integrated network maps—would greatly streamline the process. Technically, such a program may not be that difficult to create, as it is likely that it could be built from existing software components and libraries. The hard part would be to do a really good job designing the software so that it helps guide users to collect and display the right types of information. Projects with similar goals in other domains have had limited success, partially because it is easy to get bogged down in adding too many of the wrong "bells and whistles" and lose track of what the users need. In order to succeed in producing a genuinely useful tool, this project would need to aim for a modest set of specific features designed to support those tasks directly relevant to human rights data visualization work.

Building such a tool would require an experienced software project manager and two good programmers with good user interface design skills and familiarity with networks and relevant software libraries. The project would also need to have access to someone with expertise in survey design and administration, and would be best executed in partnership with an NGO or agency that would be a likely user. It would probably take 2-4 years to release a well-designed stable tool with useful documentation.

Project E: Exploratory work with unstructured data from an archive of human rights material.

The area of application that might have the largest potential impact would be combining text mining and network techniques in order to extract and present new facts or evidence from existing data that are not normally treated in a relational way. Several human rights organizations have been involved in archiving large document caches from repressive regimes, or have collected quantities of testimony from victims. It may be possible to quickly and accurately extract overviews of relationships that are buried in the data and present them in an understandable way. Or to use relationships mentioned in testimony to aid in matching up various cases.

The first stage would be partnering an experienced rights organization that has primary source data with a specialist researcher to determine what questions the organizations has that could be addressed, and what features are available in the data. The next step would be adapting and applying the extraction and coding tools to the specific problems. The final stage would be to evaluate the analysis results to determine if new insight has actually been gained. This would also be the riskiest proposal, because it would rely on a number of unproven technologies and the form of the outcome is much less certain. It would probably take 2-3 years for this project to produce useful results. However, it is likely that researchers will be able to determine if a specific partnership or data set is likely to be productive in the first few months of work.

Project F: Create a "fingerprint" map of interconnected rights abuses for a specific country or issue.

It would be interesting to determine if there are human rights violations that tend to occur together in characteristic clusters. This could be helpful to see how the various established rights are interlinked, and to determine if specific issues or events in a certain region display "signature" patterns of violations. There are many existing archives released by human rights investigations or hearings that contain collections of established cases, each one tagged with the list of human rights that were violated. From this data it should be possible to construct affiliation networks in which each of the nodes corresponds to a human right, and the strength of the links between pairs of nodes are established by the number of cases in which those rights were both violated. Another possibility would be to generate the networks using subsets of the relationships that have been collected in the news event archives about a region or conflict. The networks would of course be subject to the same uncertainties, missing information and biases as the archives they are based on, but the resulting maps might provide useful summaries for education and communication.

A good team for this kind of work would be a database and information specialist, a network analyst, and a graphic designer. Depending on how well structured the data sources are, it might be possible to produce examples of such maps in six months to year.

References

[1] Human Rights Watch Africa Division. Ituri: "Covered in Blood" ethnically targeted violence

in northeastern DR Congo. *Human Rights Watch Report*, 15(11 (A)), July 2003. Online at http://www.hrw.org/reports/2003/ituri0703/.

[2] M.S. Granovetter. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[3] E.M. Rogers. *Diffusion of Innovations*. Simon and Schuster, 2003.

[4] J. Scott. *Social Network Analysis: A Handbook*. Sage Publications, 2000.

[5] Robert A. Hanneman and Mark Riddle. *Introduction to Social Network Methods*. University of California, Riverside, Riverside, CA, 2005. published in electrontic form at http://www.faculty.ucr.edu/~hanneman/nettext/.

[6] R. Cross, S.P. Borgatti, and A. Parker. Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, 44(2):25–46, 2002. Online at https://webapp.comm.virginia.edu/SnaPortal/portals%5C0%5Cmaking_invisible_work_visible.pdf.

[7] J. Blythe, C. McGrath, and D. Krackhardt. The effect of graph layout on inference from social network data. *Graph Drawing: Symposium on Graph Drawing, Gd'95, Passau, Germany, September 20-22, 1995: Proceedings*, 1996.

[8] John Emerson. *Visualizing Information for Advocacy: An Introduction to Information Design*. Tactical Technology Colletive, January 2008. online at http://www.tacticaltech.org/infodesign.

[9] S. Bender-deMoll and D.A. McFarland. The art and science of dynamic network visualization. *Journal of Social Structure*, 7(2), 2006.

[10] Jeff Jonas and Jim Harper. *Effective Counterterrorism and the Limited Role of Predictive Data Mining*. CATO Institute, Washington, D.C., December 2006. Online at http://www.cato.org/pub_display.php?pub_id=6784.

[11] R. Popping. Knowledge graphs and network text analysis. *Social Science Information*, 42(1):91, 2003. avalible online at http://ics.uda.ub.rug.nl/FILES/root/Articles/2003/P.

[12] Robert Axelrod. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, 1997.

[13] U. Brandes, P. Kenis, J. Raab, V. Schneider, and D. Wagner. Explorations into the visualization of policy networks. *Journal of Theoretical Politics*, 11(1):75, 1999.

[14] Trampoline Systems. Trampoline Enron Explorer, 2006. http://enron.trampolinesystems.com/.

[15] L. Krempel and T. Plümper. Exploring the dynamics of international trade by combining the comparative advantages of multivariate statistics and network visualizations. *Journal of Social Structure*, 4(1), 2003. Online at http://www.cmu.edu/joss/content/articles/volume4/KrempelPlumper.html.

[16] Annuska Derks, Roger Henke, and Ly Vanna. *Review of a Decade of Research On Trafficking in Persons, Cambodia*. Center for Advanced Study / The Asia Foundation, May 2006. Online at http://www.asiafoundation.org/pdf/CB_TIPreview.pdf.

[17] Brad Edmondson, Laurie Bennett, and Daniel Shorter. American diaspora. Article at Epodunk.com http://www.epodunk.com/top10/diaspora/, 2007.

[18] J.H. Fowler and S. Jeon. The authority of supreme court precedent. *Social Networks*, 30:16–30, 2008.

[19] Wayne McIntosh, Ken Cousins, and Steve Simon. cite-it: mapping precedent across the American court system, 2005. online at http://www.bsos.umd.edu/gvpt/CITE-IT/.

[20] Rebecca Thorpe and Michael Evans. Beyond the double standard: Comparing US supreme court decision making in property and civil rights cases. *Paper presented at the Annual Meeting of the Midwest Political Science Association*, April 2006. http://www.bsos.umd.edu/GVPT/CITE-IT/Documents/Thorpe%20etal%202006%20Beyond%20t%20double%20standard.pdf.

[21] K. Kwon and G.A. Barnett. Mapping international agreement on human rights treaties, 1980-2006: An application of network analysis. *Sunbelt Social Networks Conference*, January 2008.

[22] Daniel Guzmán, Tamy Guberek, Amelia Hoover, and Patrick Ball. *Missing People in Casanare*. Human Rights Data Analysis Group, Palo Alto, California, November 2007. Online at http:

//www.hrdag.org/resources/publications/casanare-missing-report.pdf.

[23] E.C. Johnsen, H.R. Bernard, P.D. Killworth, G.A. Shelley, and C. McCarty. A social network approach to corroborating the number of AIDS/HIV+ victims in the US. *Social Networks*, 17(3-4):167–187, 1995.

[24] Brendan Nyhan and Michael Tofias. Passing the bucks: The Congressional contribution network 1985-2006. Online at http://www.duke.edu/~bjn3/nyhan-tofias-mpsa08.pdf, April 2008.

[25] Skye Bender-deMoll and Greg Michalec. Unfluence: Mapping political contribution networks. http://unfluence.net.

[26] Adam Perer and Ben Shneiderman. Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008. http://hcil.cs.umd.edu/trs/2008-03/2008-03.pdf.

[27] Lloyd Paulette. Mapping the world order: A reassessment of huntington's clash of civilizations (unpublished). Dept. Sociology, Indiana University, 2008.

[28] M.A. Porter, P.J. Mucha, M.E.J. Newman, and A.J. Friend. Community structure in the United States House of Representatives. *Physica A: Statistical Mechanics and its Applications*, 386(1):414–438, 2007. Online at http://www.maths.ox.ac.uk/~porterm/papers/conglong.pdf.

[29] Valdis Krebs. Uncloaking a slumlord conspiracy, 2008. http://www.orgnet.com/slumlords.html.

[30] Kim Norelen, Gabriel Lucas, Mike Gebbie, and John Chuang. EVA: Extration, visualization and analysis of the telecommunications and media ownership network. Research report, University of California, Berkeley online at http://denali.berkeley.edu/eva/its-paper.pdf, 2001.

[31] Eric Klinenberg. Breaking the news. *Mother Jones*, March/April 2007. (See "And Then There Were Eigh" subsection) Online at http://www.motherjones.com/news/feature/2007/03/breaking_the_news.html.

[32] E. Schiffer. The Power Mapping Tool: A method for the empirical research of power relations. IFPRI Discussion paper 00703 http://www.ifpri.cgiar.org/pubs/dp/ifpridp00703.pdf, May 2007.

[33] Eva Schiffer. Net-Map Toolbox, influence mapping of social networks, 2008. http://netmap.ifpriblog.org/.

[34] E. Schiffer, J. Hauck, and M. Abukari. *Influence Network Mapping: Mapping linkages of Water Users' Associations in IFAD-supported LACOSREP in Northern Ghana*. International Food Policty Research Institute, 2007. Online at http://netmap.files.wordpress.com/2008/01/abukari_hauck_schiffer_07_netmap_wuas_ghana.pdf.

[35] The Center for Victims of Torture. *New Tactics in Human Rights Project*. The Center for Victims of Torture, 2007. http://www.newtactics.org/sites/newtactics.org/files/resources/Using_the_tactical_map_WDA.pdf.

[36] Darren Noy. Understanding the San Francisco homeless policy field. Paper for Society for the Study of Social Problems 2006 Annual Meeting, Section 45: Housing Policies and Social Justice in Communities, 2006.

[37] S. Biggs and H. Matsaert. An actor-oriented approach for strengthening research and development capabilities in natural resource systems. *Public Administration and Development*, 19(3):231–262, 1999.

[38] B. Douthwaite, A. Carvajal, S. Alvarez, E. Claros, and L.A. Hernández. Building farmers' capacities for networking (Part I): Strengthening rural groups in Colombia through network analysis. *Knowledge Management for Development Journal*, 2(2), 2006.

[39] L. Clark. Building farmers' capacities for networking (Part II): Strengthening agricultural supply chains in Bolivia using network analysis. *Knowledge Management for Development Journal*, 2(2), 2006.

[40] Louise Clark. *Network Mapping as a Diagnostic Tool (English translation)*. Centro Internacional de Agricultura Tropical, La Paz, Bolivia, 2006.

[41] Rick Davies. Network perspectives in the evaluation of deveopment interventions: More than a

metaphor. Online at http://www.mande.co.uk/docs/nape.pdf, November 2003.

[42] J.A. Crowe. In search of a happy medium: How the structure of interorganizational networks influence community economic development strategies. *Social Networks*, 29(4):469–488, 2007.

[43] M.L. Doerfel and M. Taylor. Network dynamics of interorganizational cooperation: the Croatian civil society movement. *Communication Monographs*, 71(4):373–394, 2004.

[44] S. Moore, E. Eng, and M. Daniel. International NGOs and the role of network centrality in humanitarian aid operations: A case study of coordination during the 2000 Mozambique floods. *Disasters*, 27(4):305–318, 2003.

[45] Elisabeth J. Strines. *Social Network Mapping: A New Tool for the Leadership Toolbox*. PhD thesis, Air University Maxwell Air Force Base, 8037 Washington Road Alexandria, VA 22308, April 2002. Online at http://handle.dtic.mil/100.2/ADA420742.

[46] Convergence Center. *Movement in the Making? Social Network Analysis of the WSIS Civil Society*, 2005. Online at http://dcc.syr.edu/ford/mim/SNA-WSIS.pdf.

[47] Michael T. Heaney and Fabio Rojas. Partisans, nonpartisans, and the antiwar movement in the United States. *American Politics Research*, 35(5), September 2007. Online at http://plaza.ufl.edu/mtheaney/AntiwarMovement.pdf.

[48] d. m. boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 2007. http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html.

[49] Facebook. Facebook Causes. http://www.facebook.com/apps/application.php?id=2318966938.

[50] Martina Morris, Steven Goodreau, and James Moody. Sexual networks, concurrency, and STD/HIV. In King K. Holmes, P. Frederick Sparling, and Walter E. Stamm, editors, *Sexually Transmitted Diseases, 4th Edition*. McGraw-Hill, New York, 2006.

[51] M. Morris and M. Kretzschmar. A microsimulation study of the effect of concurrent partnerships on HIV spread in Uganda. *Math Pop Studies*, 8(2):109–133, 2000.

[52] K. Stoebenau and T.W. Valente. Using network analysis to understand community-based programs: A case study from highland Madagascar. *International Family Planning Perspectives*, 29(4):167–174, 2003.

[53] Valdis Krebs. It's the conversations, stupid! the link between social interaction and political choice, 2004. http://www.extremedemocracy.com/chapters/Chapter%20Nine-Krebs.pdf.

[54] M. Rosvall and C.T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118, 2008. online at http://octavia.zoology.washington.edu/publications/RosvallAndBergstrom08.pdf.

[55] R.W. Schvaneveldt. *Pathfinder associative networks: studies in knowledge organization*. Ablex Publishing Corp. Norwood, NJ, USA, 1990.

[56] Patrick Ball, Greg Walton, Drunken Master, and Reid Fleming. *Hacktivism and Human Rights: Using Technology to Raise the Bar*, July 2001. Presentation and panel discussion at DEF CON 9 conference. http://www.cultdeadcow.com/panel2001/hacktivism_panel.php.

[57] H.W. Park and M. Thelwall. Hyperlink analyses of the world wide web: A review. *Journal of Computer-Mediated Communication*, 8(4), 2003. Online at http://jcmc.indiana.edu/vol8/issue4/park.html.

[58] Linkfluence. Presidential watch 08 interactive website. http://presidentialwatch08.com/index.php/map/.

[59] M. Garrido and A. Halavais. Mapping networks of support for the Zapatista movement. *Cyberactivism: Online Activism in Theory and Practice*, 2003.

[60] L. Tateo. The Italian extreme right on-line network. an exploratory study using an integrated social network analysis and content analysis approach. *Journal of Computer-Mediated Communication*, 10(2), 2005. Online at http://jcmc.indiana.edu/vol10/issue2/tateo.html.

[61] Richard Rogers and Anat Ben-David. *The Palestinian-Israeli peace process and transnational issue networks: The complicated place of the Israeli NGO*. Govcom.org, 2005. Online at http://www.govcom.org/publications/full_list/rogers_ben-david_1.pdf.

[62] R.C. Hobbs. *Mark Lombardi: Global Networks*. Independent Curators International, 2003.

[63] G. King and W. Lowe. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57(03):617–642, 2003. Online at polmeth.wustl.edu/retrieve.php?id=105.

[64] Seung-Joon Jun. The international structure of non-government-organization activity. Presentation at the INSNA Sunbelt XXVIII conference, St. Pete Beach, FL, January 2008.

[65] A. Hammerli, R. Gattiker, and R. Weyermann. Conflict and cooperation in an actors' network of Chechnya based on event data. *Journal of Conflict Resolution*, 50(2):159, 2006.

[66] S.P. Borgatti and J.L. Molina. Toward ethical guidelines for network research in organizations. *Social Networks*, 27(2):107–117, 2005.

[67] Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1), 2008.

[68] C. Kadushin. Who benefits from network analysis: ethics of social network research. *Social Networks*, 27(2):139–153, 2005.

[69] The American Anthropological Association. Board statement on human terrain systems. http://aaanewsinfo.blogspot.com/2007/11/aaa-board-statement-on-hts.html.

[70] P.V. Marsden. Network data and measurement. *Annual Review of Sociology*, 16(1):435–463, 1990.

Contact information:

Skye Bender-deMoll (author of this report) is an independent researcher and consultant on network visualization and dynamic networks.
Contact: mailto:skyebend@skyeome.net
Website: http://skyeome.net/wordpress/

AAAS Science and Human Rights Program (SHRP) works with scientists to "advance science and serve society" through human rights. The Program carries out its mission by engaging individual scientists and scientific associations in human rights efforts; applying scientific tools and technologies to enhance human rights work; bringing human rights standards to the conduct of science; and promoting the human right to enjoy the benefits of scientific progress.
Contact: mailto:shrp@aaas.org
Website: http://shr.aaas.org/