Comput Math Organ Theory (2012) 18:328–339
DOI 10.1007/s10588-012-9126-x

SI: DATA TO MODEL

Extracting socio-cultural networks of the Sudan from open-source, large-scale text data

Jana Diesner · Kathleen M. Carley · Laurent Tambayong

Published online: 13 July 2012

Keywords Social networks · Automated text analysis · Sudan · Dynamic network analysis

1 Introduction and background

This paper was originally presented at the annual conference of the Sudan Studies Association (SSA) at Purdue University in 2010, and has been adapted for this publication. At the SSA meeting, we had the opportunity to present our work on a computer-supported methodology for extracting and analyzing over-time network data based on open source text data about the Sudan, and to receive feedback from subject matter experts on the Sudan. Our presentation was part of a panel on "Data analyses of Sudan's reality". In this panel, the collaborators from a multi-university research initiative that focuses on rapid ethnographic retrieval reported on their work.

The field of Network Analysis provides theories and methods for representing and examining the structure, functioning and dynamics of interactions between social agents, infrastructures and information (Wasserman and Faust 1994;

J. Diesner
UIUC, Urbana Champaign, IL, USA
e-mail: jdiesner@illinois.edu

K.M. Carley (✉)
ISR, SCS, Carnegie Mellon University, Wean 5130, Pittsburgh, PA 15213, USA
e-mail: kathleen.carley@cs.cmu.edu

L. Tambayong
California State University at Fullerton, Fullerton, CA, USA
e-mail: ltambayong@gmail.com

National Research Council 2005). To give an example, in a *social network*, social agents are represented as nodes and the connections between them as edges. The edges can denote various types of relationships, such as kinship, cooperation and antagonism. The types of entity classes considered for a network can be extended to other classes that denote not only the who (social agents, i.e. individuals and groups), but also the what (tasks, event), where (location), why (opinions, emotions, beliefs), how (resources, information) and when (time) of actual and fictional events. Based on the concept of *socio-technical systems* (Emery and Trist 1960), data representing connections between social agents and infrastructures are also called *socio-technical networks*. Such multi-mode networks are also referred to as meta-networks (Carley 2002). Real-world networks can come at any scale and are typically complex, i.e. featuring a plethora of interactions between nodes, and dynamic, i.e. constantly changing their shape and behavior.

Network analysis is employed for a number of purposes including:

– Visualizing networks and their dynamics, which can for example facilitate group discussions about complex societal issues (Freeman 2000; Hämmerli et al. 2006).
– Formally describing the structure and functioning of networks, e.g. in terms of network metrics and key entities (Wasserman and Faust 1994).
– Testing and developing hypotheses, theories and models (Carrington et al. 2005).
– Providing input to further computations, e.g. simulations of social processes such as the reaction of a network to interventions (Carley 1991; Enders and Su 2007).

Besides these common uses, network analysis has also been adopted for studying covert networks of subs-state and non-state actors from various geopolitical backgrounds (Erickson 1981; Humphreys 2005; Carley et al. 2007).

Network data can be collected through a variety of methods, such as experiments (Milgram 1967), surveys (Ryan and Gross 1943; Krackhardt 1987) and (participatory) observations (Mitchell 1969). Modern information and communication technologies, such as the internet, mobile phones and social networking platforms, have facilitated and expedited the generation, collection and storage of network data (Eagle and Pentland 2006). In spite of these advances, direct or remote access to real-world networks is still hard to impossible for covert and past networks, such as illicit organizations and bankrupt companies (Malm et al. 2008; Klerks 2001). In these situations, one type of data that has been used for constructing network data are unstructured, natural language text data. Examples for text data sources include transcripts of court hearings, communication data such as emails, reports from subject matter experts, and material from online sources (Diesner and Carley 2005; Schrodt et al. 2001; Baker and Faulkner 1993; Popp and Poindexter 2006; Howlett 1980).

2 Data

The data set for this project is a corpus of about 32,000 text documents published by the Sudan Tribune.[1] This corpus was collected from publically available online

[1] http://www.sudantribune.com/.

sources by the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. All articles are in English. Each file has a time stamp, which represents the date on which the article was published. One limitation with these time stamps is that the event which an article reports on might have happened prior to the publication date.

3 Method

The methodology for extracting network data from text data herein presented involves the steps outlined below. This procedure is also referred to as the data to model process (D2M) (Carley et al. 2011b). The network data extraction, management, and analysis procedures were conducted in AutoMap (Carley et al. 2011a), a natural language processing toolkit, and ORA (Carley et al. 2011c), a software for network analysis.

First, we downloaded text documents from the Sudan Tribune and collected additional information on Sudanese key actors from ethnographers. Second, we cleaned and deduplicated the retrieved text data. Steps one and two are described in more detail for the general situation of collecting and preparing text data from public online sources in Diesner (2012).

Third, we identified and classified nodes from the text data (for details on steps 3 and 4 see also Diesner and Carley 2005). We considered the following node classes: people, organizations, locations, resources, and conflicts. Carrying out this step in AutoMap requires a thesaurus, which maps text terms to instances of nodes. In order to identify these text terms and assigning them to the appropriate node class, we used a combination of natural language processing (NLP) techniques, thesauri and bi-gram lists that we had previously generated in AutoMap and manually vetted, location information from gazetteers, and predefined list of node class instances that subject matter experts had provided. The employed NLP techniques included the automated detection and classification of nodes (Diesner and Carley 2008) and the identification of terms with a high cumulative or weighted frequency.[2] Once the potential set of nodes was extracted, we further revised this set in order to map the various spellings and mentions of one and the same node to unique and unambiguous node names. For example, the terms “Omar Hassan Ahmad al-Bashir”, “Omar al-Bashir”, and Omar al-Beshir” were mapped to “Omar al-Bashir”. This process is also referred to as entity resolution or co-reference resolution.

Forth, we linked the nodes into edges by using a proximity-based approach which basically connects any pair of nodes that co-occurs within a user-specified window into an edge (Danowski 1993). In AutoMap, the parameters available for specifying this window include the text unit, such as sentence or paragraph, and the number of words. Based on qualitative pre-tests, we chose any two adjacent sentences as the text unit, and a window size of seven. The analysis results are impacted by these coding choices (Diesner 2012). Based on our review of the resulting network data, we

[2]For weighed term frequency we used the tf-idf metric.

performed additional iterations of cleaning and reference resolution on the thesauri and network data. In other words, we iteratively looped back to step 3.

Once the coding material and procedure had been developed, revised and finalized, we ran the D2M process by using a script that executes AutoMap and ORA. This script is part of the original D2M technology reported on in this issue. The work done for this project was used for assessing and improving the D2M process and technology.

As for output from the D2M process, we generated one network data file per calendar year for the years of 2003 to 2008.[3] Each of these files contains all of the nodes and edges found in any of the articles published in a given year. The edge weights in these data represent the number of times that a link had been observed.

4 Results

The results we present provide a high level depiction of activities in the Sudan. These findings should be treated as indicative, not definitive, as the results need to be further validated, e.g. by subject matter experts. Thus, these outcomes should be interpreted as showing more what can be learned by using the D2M methodology than as providing a detailed understanding of the Sudan.

Network data can be analyzed in various ways, e.g. by assessing their visual representations (Freeman 2000) or graph level and node level metrics (for details on metrics mentioned herein see Wasserman and Faust 1994; Carley et al. 2011c). We focus our discussion on the latter. Figures 1, 2, 3 and 4 show the size of the agent and organization networks as well as a selected set of metrics over time; indicating that connectivity between social agents increases and decreases much more strongly than the number of network participants. This observation is manifested in the density metric, which is calculated as the number of actual connections over possible connections, and is typically fairly low for social networks. Our networks also exhibit this sparseness. Components are strongly connected portions of graphs and typically represent tightly interwoven sub-groups. In our results, the number of components seems to be strongly correlated with the number of nodes. Fragmentation describes the proportion of isolates, i.e. nodes that are not connected to any other node. More than half of the social agents in both networks are isolates; indicating that the majority of actors are not co-mentioned with nodes of the same type within in the same text window in the underlying data.

Centrality metrics express different notions of prominence and power: Graph-level degree centralization is based on the degree centrality of each node, which is measured as the number of direct links per node. The high degree centralization values in the Sudan networks suggest that on average, many individuals and organizations act as active power hubs. One explanation for this observation is that those social agents who are not isolates co-occur with a variety of other social agents, i.e. they appear in different social contexts.

[3]Broken down by years, we used 2003 (2932), 2004 (6943), 2005 (3828), 2006 (3828), 2007 (5815), 2008 (9266) text files.

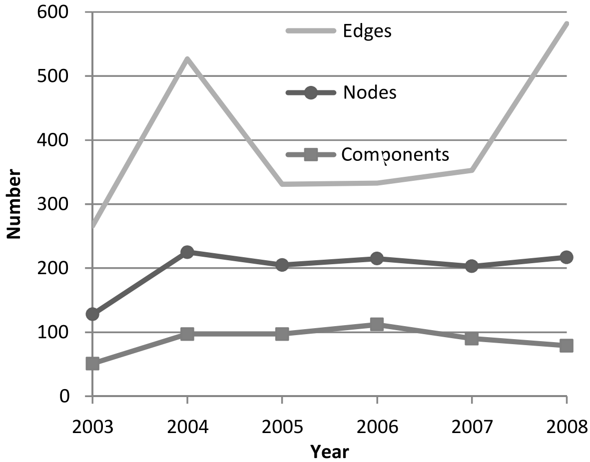


Fig. 1 Network size over time, agent network

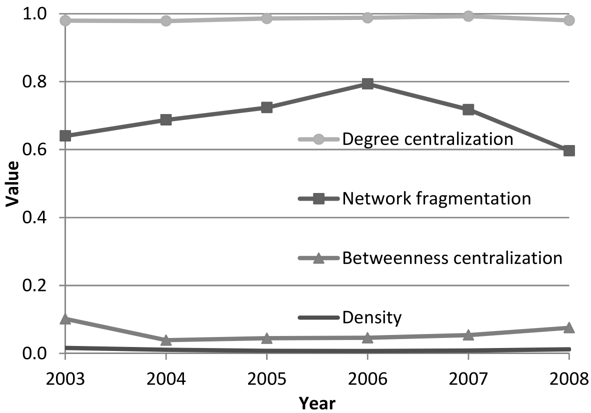


Fig. 2 Network metrics over time, agent network

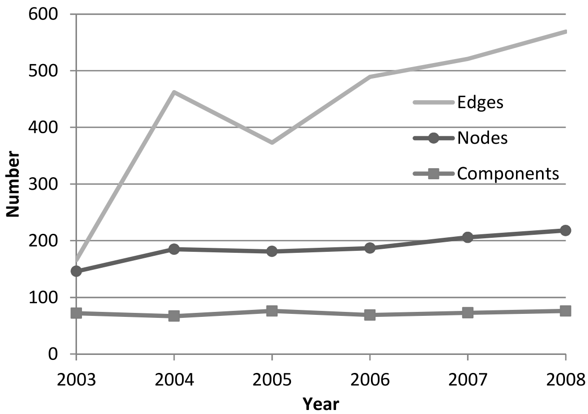
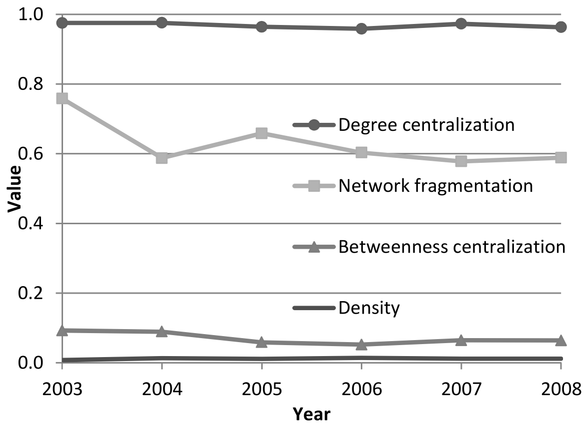


Fig. 3 Network size over time, organization network

Fig. 4 Network metrics over time, organization network



The graph-level metric of betweenness centralization is also based on the betweenness centrality scores per node. These scores describe how often a node is positioned on the shortest path between all other pairs of nodes in a network. Scoring high on this metric means that a node has the power to act as a broker or gatekeeper of contacts, social capital or information. The low betweenness centralization values in the Sudan networks suggest that most social agents are not located on efficient paths between other network participants, and therefore are not in the position to act as boundary spanners or brokers between different sets of actors.

Figures 5 and 6 show the key players according to a selected set of network metrics in ranked order over time. These results were generated by using the following procedure:

1. Identify relevant metrics given the research context. We selected the following ones:
   - Degree centrality (prestige, action) and betweenness centrality (control, brokerage) as introduced above.
   - Eigenvector centrality: a node scores high on this metric if it is connected to many other highly connected nodes. An example would be a lobbyist or a diplomat.
   - Clique Count: the number of distinct cliques to which a node belongs. A clique is a maximally complete subgraph, and implies strong cohesion among its members.
   - Simmelian ties: computed as the normalized number of nodes to which a node has a Simmelian tie. Simmelian ties are directed links in closed triads, which are considered as stable social structures (Krackhardt 1998).
2. For each node (social agent) in each network, compute the value per metric, sort nodes by their rank, and output the N highest ranking nodes per metric. We set N to 10.
3. For each of the top ten nodes per metric, count how often they occur across the five considered metrics. For the figures shown below, we decided to consider only

Degree Centrality	03	04	05	06	07	08	Betweenness Centrality	03	04	05	06	07	08	Eigenvector Centrality	03	04	05	06	07	08
omar hassan ahmad al bashir	3	2	1	1	1	1	omar hassan ahmad al bashir	1	1	1	2	2	1	omar hassan ahmad al bashir	4	3	1	1	1	1
ali osman mohammed taha	1	1	4	2	3	2	zulkepli bin marzuki	2	10	11	10	1	2	ali osman mohammed taha	1	1	3	2	4	4
khalil jarraya	4	4	2	4	4	5	john garang	3	6	6	12	3	3	khalil jarraya	5	4	2	3	3	2
john garang	2	3	3	6	6	10	khalil jarraya	4	3	3	1	8	11	john garang	2	2	4	5	6	6
salva kiir mayardit	17	15	5	3	2	3	ahmad zerfaoui	12	7	4	6	4	10	salva kiir mayardit	16	15	5	4	2	3
samira shahbandar	10	5	13	5	12	4	mustafa mohamed fadhil	14	11	2	4	6	8	redendo cain dellosa	3	5	10	13	14	13
zulkepli bin marzuki	9	6	8	8	10	6	george bush	6	13	7	8	5	6	zaini zakaria	13	9	7	6	5	11
redendo cain dellosa	5	8	11	15	14	13	djamel moustfa	16	2	12	13	12	4	djamel moustfa	7	7	13	11	7	9
djamel moustfa	6	10	6	12	8	9	ali osman mohammed taha	9	9	5	5	7	13	colin powel	9	6	6	15	15	14
zaini zakaria	13	9	12	7	5	12	salva kiir mayardit	10	15	9	3	9	5	zulkepli bin marzuki	12	8	8	7	13	12
george bush	12	12	9	9	7	11	samira shahbandar	8	5	13	9	11	9	george bush	10	12	15	8	8	10
colin powel	8	7	7	14	15	15	zaini zakaria	11	8	8	14	13	12	hassan al turabi	6	13	14	12	11	5
hassan al turabi	7	13	14	13	11	7	kofi annan	15	4	10	7	14	14	samira shahbandar	11	10	11	9	12	7
ahmad zerfaoui	15	14	15	11	9	8	al sadeq al mahdi	13	14	15	11	10	7	al sadeq al mahdi	8	11	12	14	10	8
kofi annan	16	11	9	10	13	14	redendo cain dellosa	5	12	14	15	15	15	kofi annan	15	14	9	10	9	15
al sadeq al mahdi	11	13	15	15	11	12	colin powel	7	16	16	14	15	17	mustafa mohamed fadhil	17	15	12	13	13	14
mustafa mohamed fadhil	14	14	13	10	14	13	hassan al turabi	17	15	13	15	16	14	ahmad zerfaoui	14	17	16	12	15	13

Clique Count	03	04	05	06	07	08	Simmelian Ties	03	04	05	06	07	08
omar hassan ahmad al bashir	1	1	1	1	1	1	omar hassan ahmad al bashir	1	1	1	1	1	1
zulkepli bin marzuki	3	3	12	3	2	2	khalil jarraya	3	2	3	2	6	4
khalil jarraya	2	4	5	2	6	8	john garang	2	4	2	6	2	7
ahmad zerfaoui	7	10	7	4	4	3	george bush	3	5	3	3	2	9
djamel moustfa	9	2	6	9	9	4	ali osman mohammed taha	3	5	6	3	2	8
ali osman mohammed taha	5	5	2	6	4	10	salva kiir mayardit	11	15	5	3	2	2
mustafa mohamed fadhil	12	7	4	5	7	6	hassan al turabi	7	12	8	10	7	9
john garang	3	9	3	13	8	11	zulkepli bin marzuki	11	3	13	6	11	2
salva kiir mayardit	16	15	11	9	3	5	al sadeq al mahdi	7	8	12	13	7	9
samira shahbandar	12	8	14	8	12	7	kofi annan	10	8	6	8	9	14
zaini zakaria	11	6	9	11	12	14	colin powel	6	8	8	13	11	14
george bush	10	13	7	7	10	12	ahmad zerfaoui	11	14	8	10	9	4
al sadeq al mahdi	12	12	14	11	11	9	samira shahbandar	11	11	13	8	11	4
redendo cain dellosa	7	10	9	15	15	15	mustafa mohamed fadhil	11	13	8	10	11	12
hassan al turabi	12	13	13	14	14	12	djamel moustfa	11	7	13	13	11	12
colin powel	6	13	14	16	15	17	redendo cain dellosa	9	12	13	13	11	15
kofi annan	16	14	13	14	15	15	zaini zakaria	11	11	13	13	11	14

Fig. 5 Key individuals ranked (decreasing) by node-level SNA measures

Degree Centrality	03	04	05	06	07	08	Betweenness Centrality	03	04	05	06	07	08	Eigenvector Centrality	03	04	05	06	07	08
sudan gov	1	1	2	3	1	2	sudan gov	1	1	2	4	2	3	sudan gov	1	2	3	2	2	3
african union	8	4	4	1	2	1	women	3	4	3	2	4	5	women	3	3	1	1	1	6
janjaweed	6	2	3	2	5	4	action churches together	2	2	1	6	5	7	janjaweed	4	1	2	3	3	4
un security council	9	3	1	4	3	3	regime	11	5	4	3	1	1	african union	8	5	6	4	4	1
women	4	6	5	5	6	8	dinka	4	8	11	5	6	13	un security council	7	4	4	5	8	2
regime	5	9	7	9	8	7	coalition prov. authority	15	15	13	1	3	2	human rights watch	5	6	5	9	5	8
students	2	12	8	7	11	6	janjaweed	12	6	5	9	8	10	amnesty international	2	8	11	6	6	10
human rights watch	12	5	6	6	9	10	african union	8	7	6	12	9	12	regime	9	9	7	10	10	7
amnesty international	3	7	13	8	7	13	un security council	14	10	7	7	13	4	action churches together	6	10	10	12	9	11
khartoum government	7	8	10	15	12	11	khartoum government	13	9	8	14	7	8	students	11	12	9	7	11	12
coalition prov. authority	15	15	14	13	4	5	fur	9	3	14	15	14	6	fur	12	11	8	8	12	13
dinka	13	13	12	11	10	9	nuer	6	12	9	10	11	14	coalition prov. authority	15	13	14	13	7	5
action churches together	10	10	9	12	14	14	students	7	11	15	8	12	11	khartoum government	13	7	13	14	13	9
nuer	11	14	15	10	13	11	amnesty international	10	14	12	13	10	9	nuer	10	15	15	11	15	14
fur	14	11	11	14	15	15	human rights watch	5	13	10	11	15	15	dinka	14	14	12	15	14	15

Clique Count	03	04	05	06	07	08	Simmelian Ties	03	04	05	06	07	08
sudan gov	1	1	1	6	1	1	african union	2	3	2	1	5	4
women	2	4	2	1	2	4	janjaweed	2	1	2	1	2	10
janjaweed	3	2	3	5	6	6	sudan gov	1	2	5	8	1	3
african union	7	3	4	4	4	5	women	2	3	1	5	5	4
regime	11	6	6	2	5	2	regime	8	6	6	3	2	1
un security council	3	8	5	3	7	9	un security council	8	5	4	3	7	2
action churches together	7	6	7	8	8	8	action churches together	2	8	8	6	8	8
dinka	3	12	11	10	10	7	human rights watch	2	6	6	9	10	11
human rights watch	10	4	8	9	9	14	coalition prov. authority	8	15	15	6	2	6
coalition prov. authority	15	15	15	7	3	3	amnesty international	2	10	10	11	13	11
students	11	11	10	11	11	10	dinka	8	10	13	12	8	8
nuer	3	14	13	12	14	11	nuer	8	13	13	12	11	6
amnesty international	7	13	13	13	12	11	students	8	14	10	9	11	11
khartoum government	11	10	9	15	13	11	fur	8	10	9	15	14	14
fur	11	9	11	14	15	15	khartoum government	8	9	12	14	14	14

Fig. 6 Key organizations ranked (decreasing) by node-level SNA measures

those nodes for which these sums were higher than four for people and three for organizations.[4] This procedure resulted in a set of 17 agents and 15 organizations.
4. Output nodes by decreasing rank along with their rank per year (Figs. 5 and 6). In these figures, the lighter the background of a cell, the higher the rank.

Our results indicate that even though the graph-level network metrics change considerably from year to year, the set of key players (agents and organizations) and their rankings are fairly robust, even across different node-level network metrics. Looking at specific agents, we see that "Omar Hassan Al-Bashir", president of the Sudan during all years considered, has maintained his leading role in the agent network. In contrast to that, the prominence of vice president "Ali Osman Taha" has slightly decreased. At the same time, political leaders from South Sudan, especially "Salva Kiir", who replaced "John Garang" as the president of South Sudan in 2005, were able to increase his importance and establish their power.

The set of key agents also entails several people who are associated with various military separatist or terrorist groups. For example, "Redondo Cain Dellosa", who had been detained for his involvement in a ferry bombing outside of Manila in 2004, is associated with the Abu Sayyaf group. "Ahmad Zerfaoui" has been identified by a Security Council Committee from the United Nations as the leader of a subgroup which operates in the Sahel-Sahara region and is a member of the Al-Qaeda network (Committee).

Among the key agents are also religious leaders: "Hassan al Turabi" was instrumental in manifesting the sharia law in Northern Sudan, and has been imprisoned several times since 2004. The prominence of "Sadiq al Mahdi", who is a descendant of the Mahdi who successfully led the Mahdist War against the British colonizers from 1881 to 1885, but whose current activities might be minimal, shows the strong legacy of religious leaders—at least in news coverage.

The networks of organizations are dominated by Sudanese governmental groups. Other types of prominent organizations include military groups and armed forces such as the "Janjaweed", international and external organizations such as the "African Union" and the "United Nations Security Council", and NGOs concerned with justice and human rights, such as "Amnesty International" and "Human Rights Watch". There are also three native tribes among the key groups, namely the "Dinka", "Nuer" and "Fur". In the Sudan, tribes have a strong impact on many dimensions of society, such as culture and language, economy, politics and defense. In fact, many of the political leaders of Sudan originate from the named tribes.

5 Limitations and expansion of work

The validation of network data and analysis results for networks that are hard to access is challenging. Typically, validation is performed by comparing the data and results against previously created ground truth data, or consulting with subject matter

[4] For organizations, we disregarded "islamist" as it seemed too broad and "university" as is correlated with "students".

experts (SMEs). The support from an ONR grant allowed us to collaborate with Dr. Richard Lobban, who is a professor of anthropology and African studies at Rhode Island College and a renowned SME on Africa. Dr. Lobban and his team examined the tribal affiliation networks that we had extracted from the text data and marked the changes that they considered necessary in order to correct for false negatives and positives in the data. The analyses presented herein consider these changes.

Several of the research tasks presented as future work in the original version of this paper have been realized since then: Building upon the ground truth data verified by the SMEs and the D2M technology presented herein, we have later compared the tribal affiliation networks from the SMEs (ground truth data) against network data extracted from the content of news articles and from index terms for these articles (Diesner 2012). We have also extended the D2M technology and methodology to extract and consider a larger set of node classes, including locations, resources, knowledge, events, tasks, time and attributes, and to assign a specificity value[5] and additional sub-categories to nodes (Diesner 2012). Examples for sub-categories of nodes of type organization are "corporate", "political" and "religious".

Finally, it is crucial to note that some (key) social agents might have been erroneously identified due to their similarity in spelling to nodes of the same or other types. In order to disambiguate these nodes, reference resolution techniques beyond the manual vetting of thesauri need to be employed.

Acknowledgements This research was supported in part by ARI W91WAW07C0063, GMU MURI FA9550-05-1-0388, and ONR N00014-08-1-1186. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Institute, the Air Force Office of Sponsored Research, the Office of Naval Research, or the United States Government. We also thank our research partners Jeffrey Johnson and Tracy van Holt, both from East Carolina University, and Richard Lobban from Rhode Island College for their continuous help on this project. Special thanks also to Frank Kunkel, CMU, for his help with this project. We are furthermore grateful to the feedback from the attendees of our presentation at the 29th Annual Conference of the Sudan Studies Organization (SSA).

References

Baker WE, Faulkner RR (1993) The social organization of conspiracy: illegal networks in the heavy electrical equipment industry. Am Sociol Rev 58(6):837–860

Carley KM (1991) Designing organizational structures to cope with communication breakdowns: a simulation model. Organization & Environment 5(1):19–57

Carley KM (2002) Smart agents and organizations of the future. In: Lievrouw L, Livingstone S (eds) The handbook of new media: social shaping and consequences of ICTs. Sage, Thousand Oaks, pp 206–220

Carley KM, Columbus D, Bigrigg M, Kunkel F (2011a) AutoMap User's guide 2011. Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-11-108

Carley KM, Diesner J, Reminga J, Tsvetovat M (2007) Toward an interoperable dynamic network analysis toolkit. Decis Support Syst 43(4):1324–1347. Special Issue Cyberinfrastructure for Homeland Security

Carley KM, Lanham M, Martin M, Morgon G, Schmerl B, van Holt T, Bigrigg MW, Garlan D, Johnson J, Kunkel F (2011b) Rapid ethnographic assessment: data to model. Paper presented at the HSCB Focus 2011: Integrating Social Science Theory and Analytic Methods for Operational Use

[5] The specificity value can be "specific", e.g. for the concept "Khartoum" or "generic", e.g. for the concept "city".

Carley KM, Reminga J, Storrick J, Columbus D (2011c) ORA User's guide 2011. Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-11-107
Carrington PJ, Scott J, Wasserman S (2005) Models and methods in social network analysis. Cambridge University Press, Cambridge
Committee SC QI.Z.168.04. Ahmad Zerfaoui. Security Council Committee pursuant to resolutions 1267 (1999) and 1989 (2011) concerning Al-Qaida and associated individuals and entities. http://www.un.org/sc/committees/1267/NSQI16804E.shtml
Danowski JA (1993) Network analysis of message content. Prog Commun Sci 12:198–221
Diesner J (2012) Uncovering and managing the impact of methodological choices for the computational construction of socio-technical networks from texts. CMU-ISR-12-101 (forthcoming)
Diesner J, Carley KM (2005) Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis. In: Narayanan VK, Armstrong DJ (eds) Causal mapping for information systems and technology research: approaches, advances, and illustrations. Idea Group Publishing, Harrisburg, pp 81–108
Diesner J, Carley KM (2008) Conditional random fields for entity extraction and ontological text coding. J Comput Math Organ Theory 14:248–262
Eagle N, Pentland A (2006) Reality mining: sensing complex social systems. Pers Ubiquitous Comput 10(4):255–268
Emery F, Trist E (1960) Socio-technical systems. Manag Sci, Models Tech 2:83–97
Enders W, Su X (2007) Rational terrorists and optimal network structure. J Confl Resolut 51(1):33–57
Erickson B (1981) Secret societies and social structure. Soc Forces 60:188–210
Freeman L (2000) Visualizing social networks. J Soc Struct 1(1). Available: http://www.cmu.edu/joss/content/articles/volume1/Freeman.html
Hämmerli A, Gattiker R, Weyermann R (2006) Conflict and cooperation in an actors' network of Chechnya based on event data. J Confl Resolut 50(2):159–175
Howlett J (1980) Analytical investigative techniques: tools for complex criminal investigations. Police Chief 47(12):42–45
Humphreys M (2005) Natural resources, conflict, and conflict resolution. J Confl Resolut 49(4):508–537
Klerks P (2001) The network paradigm applied to criminal organizations: theoretical nitpicking or a relevant doctrine for investigators. Connections 24(3):53–65
Krackhardt D (1987) Cognitive social structures. Soc Netw 9:109–134
Krackhardt D (1998) Simmelian ties: super strong and sticky. In: Power and influence in organizations, pp 21–38
Malm A, Kinney J, Pollard N (2008) Social network and distance correlates of criminal associates involved in illicit drug production. Security J 21(1–2):77–94
Milgram S (1967) The small world problem. Psychol Today 2(1):60–67
Mitchell J (1969) The concept and use of social networks. In: Mitchell J (ed) Social networks in urban situations: analyses of personal relationships in central African towns. Manchester University Press, Manchester, pp 1–50
National Research Council CoNSfFAA (2005) Network science. National Academies Press, Washington
Popp R, Poindexter J (2006) Countering terrorism through information and privacy protection technologies. IEEE Secur Priv 4(6):18–27
Ryan B, Gross N (1943) The diffusion of hybrid seed corn in two Iowa communities. Rural Soc 8(1):15–24
Schrodt P, Simpson E, Gerner D (2001) Monitoring conflict using automated coding of newswire reports: a comparison of five geographical regions. Paper presented at the PRIO/Uppsala University/DECRG high-level scientific conference on identifying wars: systematic conflict research and its utility in conflict resolution and prevention, Uppsala, Sweden, 8–9 June
Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

Jana Diesner is an Assistant Professor at University of Illinois Urbana Champaign, Graduate School of Library and Information Science. She will receive her Ph.D. in Computation, Organizations and Society from the Institute for Software Research Department in the School of Computer Science at Carnegie Mellon University. Her work is at the nexus of network analysis, natural language processing and machine learning. She advances computational methods and technologies that help people to understand the interplay and co-evolution of information and the structure, functioning and dynamics of socio-technical networks. She is particularly interested in covert information and covert networks.

Kathleen M. Carley is a Professor of Computation, Organizations and Society in the Institute for Software Research Department in the School of Computer Science at Carnegie Mellon University and the director of the center for Computational Analysis of Social and Organizational Systems. She received her Ph.D. in Sociology from Harvard and her undergraduate degrees from the Massachusetts Institute of Technology. Her research interests are in dynamic network analysis, agent-based modeling, social networks, and information diffusion and belief dispersion. In addresses her interests she takes an interdisciplinary approach drawing on advances in machine learning, organization science, social-psychology and cognitive science. She and her team developed ORA, AutoMap and Construct.

Laurent Tambayong is currently a faculty member at Department of Information Systems and Decision Science at Mihaylo College of Business Economic at California State University at Fullerton. He received his Ph.D. in Mathematical Behavioral Science from the University of California, Irvine. His research interests are in network and complex systems, with specific applications for social, organizational, decision, and information sciences. These include game theory, network economics, network structure and formation analysis, as well as strategic behavior and stochastic modeling.