Some Intelligence Analysis Problems and Their Graph Formulations

Rich Colbaugh*† **Kristin Glass†** **Jim Gosler***

*Sandia National Laboratories
†New Mexico Institute of Mining and Technology

Abstract

This paper considers three important classes of intelligence analysis questions, shows how these questions can be naturally formulated as graph problems, and demonstrates that analysis based on this formulation yields insights and understanding which would be difficult to obtain using other methods. In order to make explicit the relevance and practical utility of graph methods for intelligence questions, we organize our discussion around specific real world problems of current interest in the counterproliferation and counterterrorism domains.

1. Introduction

The application of graph analysis methods to intelligence analysis problems is of significant current interest, with some advocates claiming these methods have the potential to revolutionize the way intelligence analysis is practiced [e.g., 1-5]. However, less has been said concerning specific, tangible benefits of graph-based analytics beyond generic assertions about scalability and efficiency when dealing with the vast data sets now ubiquitous in national security applications. While some analysts in specialized network-centric domains recognize the utility of graph methods in their work, little has been done to present to the broader community a cogent discussion of the value of these methods for mainstream, real world analysis questions.

In graph analysis, the system of interest is represented as a *graph*, which is a set of items with connections between them [e.g., 6]. The items are called *vertices* and are depicted as nodes which may be linked to each other with *edges*. For example, the graph on the left of Figure 1 is a social network, in this case reflecting collaboration, so that the vertices (squares) represent people and an edge (line) connecting two individuals indicates that they have collaborated on a project. In the graph on the right of Figure 1 the vertices are dolphins and edges connect pairs of dolphins which are observed to interact frequently; thus the graph is a model for the social network of this population of dolphins.

One objective of graph analysis is to deduce from vertex connectivity patterns something about the properties and behaviors of the underlying system. For instance, examination of the

social network on the left of Figure 1 reveals that some individuals collaborate with many others while most people collaborate with just a few others (some of the highly connected individuals are drawn as yellow vertices in the graph). One might suspect that highly connected individuals exert disproportionate influence in a social network, so that identifying them would provide insight into the network's behavior. It can also be seen that the graph is "clumpy": people form groups and tend to interact with members of their own group more frequently than they do with individuals from other groups. These groups of individuals for which intra-group edge densities are significantly higher than inter-group densities are referred to as *graph communities* and are ubiquitous in social networks [e.g., 6]. Again, it is plausible that this community structure might say something meaningful about the activities and characteristics of this population of people.

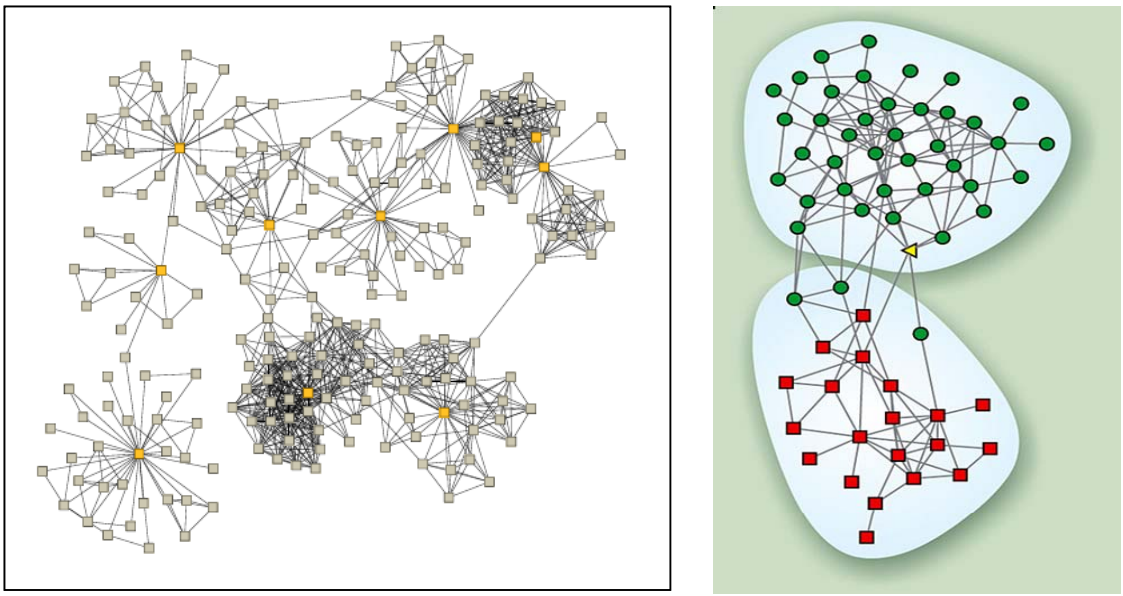


Figure 1: Social networks. The graph at left is a social network of collaborating people, in which vertices (squares) are individuals and edges (lines) connect pairs who have collaborated on a project. In the graph at right the vertices are dolphins and edges indicate social ties.

As a simple illustration of the way analysis of graph structure can indeed provide useful insights regarding a system of interest, consider the task of inferring the social group structure of the dolphin population mathematically, by computing a partition of the vertices of the dolphin graph into groups which have high intra-group edge density and low inter-group density (see, e.g., [6] for details regarding such partitioning algorithms). The results of this computation are shown as two encircled groups of vertices in Figure 1. Note that these groups are identified

using only the graph's link structure to estimate which vertices belong together. Interestingly, in the present case, "ground truth" is available regarding the dolphin groups, because this population split into two groups upon departure of the dolphin denoted by the yellow triangle; the groups resulting from this split are indicated by green and red vertex colors in Figure 1. It is seen that the groupings obtained via graph analysis are in good agreement with the actual social groupings of the dolphins,

Graph analysis, then, attempts to glean meaningful insights concerning interdependent systems from analysis of the vertex connectivity patterns of the system's graph representation. Because understanding connections between people, events, information, resources, and so on is so central to intelligence analysis, it is natural to suspect that graph-based methods may be useful in this field. We believe that graph analysis tools do indeed have the potential to substantially impact intelligence analysis. However, we also believe that in order for these tools to gain acceptance, the benefits which can be expected from their application must be made explicit. Thus a key step in the process of persuading analysts to add graph analysis tools to their "toolbox" is to illustrate the power of these tools with problems they care about.

Toward that end, this paper examines three broad and important classes of intelligence analysis questions – assessment, discovery, and prediction – and demonstrates that these questions can be naturally formulated as graph problems and that such a formulation enables development and implementation of powerful new analytic techniques. To keep the discussion as concrete as possible, the bulk of the paper consists of analyses of specific problems from these question classes; these case studies include an assessment of a potential bioweapons (BW) research facility, a discovery-based evaluation of the threat posed by "nontraditional" nuclear weapons (NW), and a predictive analysis of Islamic radicalization processes.

2. Overview of problem classes of interest

As indicated above, we consider intelligence analysis questions from three important and representative domains. *Assessment* questions of interest include: Are the objectives of a given facility what is claimed? What would be useful indicators that these objectives have changed? Who are the collaborators of a person of interest? What are the plans and intentions of this group of collaborators? The specific, real world problem of this type we consider in some depth involves assessing the activities and capabilities of a suspected "dual use" BW research facility. A key feature of this class of questions is the clandestine nature of the behavior of interest: if the target is indeed engaged in illicit activity then it is likely that attempts will be made to obscure this fact. Graph-based methods are well suited for dealing with adversary deception. The basic

idea is that, while it can be straightforward for an adversary to distort observables which are *directly-related* to an activity, it is much more difficult to maintain a consistent deception or distortion of all *indirectly-related* observables. Graph analysis, with its emphasis on understanding interdependencies, provides an excellent framework within which to identify and interpret subtle, indirect relationships between entities.

Our focus in the *discovery* domain is “unknown unknowns” sorts of questions. For example: Is there an A.Q. Khan of BW? Is there a plausible threat associated with “nontraditional” NW? Are there new technical and/or proliferation threats emerging in the WMD domain? What can be discovered about *our* secrets by an astute adversary? A key characteristic of such questions is their vague and general nature; often we are simply too ignorant about a situation to ask the kinds of well-posed questions which are amenable to standard analysis methods. Here again we concentrate on a specific problem – the threat posed by nontraditional NW ideas and substate actors – with the goal of providing a concrete illustration of the power of graph analysis for discovery-oriented questions.

A natural way to approach discovery questions is to conduct computational exploration of vast sets of possibly relevant data and identify “unusual” structure and behavior in these data. These potentially interesting features can be brought to the attention of analysts, who then carry out more focused examination, obtaining insights and providing guidance for the next round of computational exploration. This iterative, interactive analytic process enables rapid, informed refinement of broad discovery-oriented questions and generation of target-relevant knowledge and understanding. An important element of this process is the use of graph representations for the data and questions of interest, as this provides a framework of sufficient semantic richness and structure to enable deep, relevant discovery while maintaining computational tractability.

Prediction questions are ubiquitous in intelligence applications. Standard questions are of the form: How is a given threat situation likely to evolve? What would be the associated consequences? More operationally-oriented questions include: What are the likely effects of various intervention options? What additional data collection might be useful for predictive analysis? How can we move from a reactive to a proactive posture with a given adversary or threat? In the predictive analysis case study we investigate the possibility of anticipating Islamic radicalization-related processes, and show that a novel “early warning” capability for Muslim mobilization can be obtained through the application of graph analysis concepts and methods.

Predictive analysis for the complex systems of interest in intelligence analysis is, of course, very difficult. We propose that a crucial step in scientifically grounded predictive analysis is to explicitly consider *predictability* of the problem of interest before attempting to make predictions.

A key function of predictability assessment is the role it plays in identifying those observables which are most useful for prediction, and often this analysis yields surprising results. For example, we have shown that for many social processes the intrinsic features of the process, such as the “quality” of the various options in a social choice situation, are *not* useful observables; this is one reason standard prediction methods often perform poorly. Interestingly, the weak dependence of outcome on intrinsic characteristics often implies exploitable relationships between the ultimate outcome and other observables, and it is these exploitable relationships that can be used to formulate good predictions. We show below that formal, computationally tractable methods for assessing predictability and forming predictions can be developed by adopting a graph-based analytic framework for system modeling and analysis.

3. BW research assessment

Consider the task of assessing the activities, capabilities, and intentions of an organization suspected of being involved in BW research, development, and production [e.g., 7,8]. A crucial element of this assessment is identifying the research activities and trends of scientists who are associated with the organization and may be involved in a covert program. Perhaps surprisingly, the scientific papers written by the organization’s researchers can be a useful source of information for this analysis. At first glance it may seem that open publications would be of little value in assessing clandestine research – it is clearly unlikely, for instance, that researchers will publish a paper detailing their latest approach to weaponizing anthrax. However, we have shown that scientific publications invariably contain genuine, if distorted, information regarding the activities of authors [9].

The basic idea is straightforward: while it may be fairly easy to avoid publishing information which is directly informative regarding a covert program, it would be extremely difficult to suppress all indirectly-related information. For one thing, publishing a paper of sufficient significance to pass peer-review is a challenging undertaking. Thus, while it may be possible to produce a paper which contains important new scientific information and yet doesn’t disclose results specific to covert activities, it is much harder to distort the content of a sequence of papers in a manner consistent with the way scientific research evolves, or to maintain a set of co-authors whose work matches this deception. Of course, it can be quite challenging to detect these indirect signatures of illicit activity in massive publication data sets and to extract from them a reliable assessment of an organization’s true research activities.

Here we provide an indication of what can be deduced regarding research activity through graph analysis of indirect data by focusing on one high potential source of information: the

graph of interconnected authors, papers, and concepts in a scientific database (see Figure 2). As our interest in this case study is detecting BW research, we concentrate on the biological sciences literature and in particular the PubMed database; however, the proposed approach can be applied to any reasonably mature scientific discipline. Consider the situation in which an organization claims to be conducting research in some legitimate field in the biosciences, and there is suspicion that the organization's scientists may in fact be conducting BW research. The following computational graph analytic procedure provides an effective means to evaluate the relative merits of these competing hypotheses.

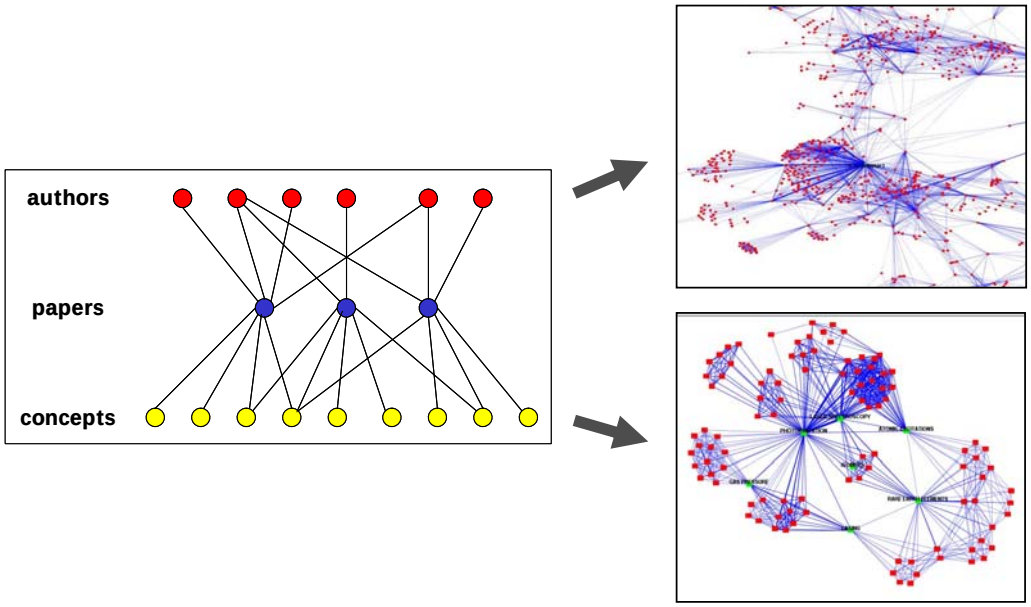


Figure 2: Network of authors, papers, and concepts in a scientific publication database. Left figure shows a cartoon of the graph, and right figures illustrate the way this graph can be projected to yield a co-author graph (top) or concept-concepts graph (bottom).

Input: name of organization of interest and identities of one or more key individuals associated with the organization.

Procedure:

1. Use the key individuals to "grow" in PubMed a co-author graph G_{CA}, in which vertices are authors and edges connect co-authors, and then apply to G_{CA} the graph analysis tools summarized below to identify collaborating groups of scientists.
2. Select a collaborating group of interest associated with the given organization, assemble the set of papers for the group, and extract from these papers the corresponding Medical Subject Headings (MeSH) terms (these are key words assigned to papers archived in the

PubMed database). Bin the MeSH terms by year and assign the labels $\{MeSH\}_{group}(t)$ to the resulting time series of term sets.

3. Build a set of MeSH terms corresponding to the claimed "good" activities of the organization, denoted $\{MeSH\}_G$, and sets of MeSH terms corresponding to the K suspected/hypothesized "bad" (BW) activities, labeled $\{MeSH\}_{B1}$, …, $\{MeSH\}_{BK}$.
4. For each year, compute the "distances" between the group's MeSH terms $\{MeSH\}_{group}(t)$ and the sets $\{MeSH\}_G$, $\{MeSH\}_{B1}$, …, $\{MeSH\}_{BK}$. Note that specifying a rational definition for distance between a pair of MeSH terms is easy because MeSH is a taxonomy of biological science concepts (e.g., one can simply count the number of hops required to get from one term to the other in the hierarchy). Moreover, this distance is meaningful because the MeSH taxonomy encodes considerable biological science information and the terms are assigned by independent experts.
5. Estimate the level of intensity with which the group under study is pursuing the "good" and "bad" activities, as a function of time, by using the time series of MeSH distances obtained in Step 4 together with a model for scientific production (e.g., a hidden Markov model). If the relative "bad to good" interest/activity level exceeds a given threshold then flag the group for further study.
6. Repeat Steps 2 – 5 for another organization group of interest.

We now provide additional details concerning the above procedure and its implementation. In Step 1, the co-author graph G_{CA} is "grown" automatically using an iterative procedure. The process is initialized by defining the key individuals to be vertices in G_{CA}. The process then retrieves from PubMed all of the papers written by these individuals, extracts the co-authors of these papers, and adds these co-authors to the set of vertices of G_{CA}; all co-author relationships reflected in these papers are encoded as edges in G_{CA}. The papers written by these newly identified authors are then retrieved from PubMed, *their* co-authors are extracted, and the process continues until the graph's growth reaches the desired scope.

Once G_{CA} is constructed, graph analysis methods can be applied to identify collaborating scientists. One useful approach to this problem is to first identify "communities" of co-authors, that is, groups of authors whose intra-group edge densities are significantly greater than their inter-group edge densities; these communities provide a first approximation to the collaborating groups (see Figure 1 and [6,10]). This approximation can then be refined in various ways, for instance by assigning a strength to each co-author relationship (e.g., based on the number of papers co-authored or number of co-authors on each paper) and then adding/deleting members from a group through considerations of collaboration intensity [9].

Step 2 involves extracting papers' MeSH terms and bining these by year. This process is straightforward because of the well-defined structure of the PubMed database, and yields the time series of sets of MeSH terms corresponding to the activity of the group of interest, denoted $\{MeSH\}_{group}(t)$. The construction, in Step 3, of the sets of MeSH terms $\{MeSH\}_G$, $\{MeSH\}_{B1}$, …, $\{MeSH\}_{BK}$ associated with the claimed (good) activity and the K BW hypotheses can be completed in collaboration with target and BW experts. Note that the process can be implemented in a semi-automated manner by leveraging the well-defined structure of MeSH.

In Step 4, distances are computed between the group's MeSH terms $\{MeSH\}_{group}(t)$ and the sets $\{MeSH\}_G$, $\{MeSH\}_{B1}$, …, $\{MeSH\}_{BK}$ for each year t. As indicated above, taxonomic distance (i.e., number of hops) can be used to define proximity for pairs of MeSH terms. The distance between a paper from $\{MeSH\}_{group}(t)$ and the MeSH term set for a given hypothesis, say $\{MeSH\}_{Bi}$, can then be obtained by simply averaging over the pair-wise distances between the paper's MeSH terms and the terms in $\{MeSH\}_{Bi}$. In order to obtain a measure of the distance between the two sets $\{MeSH\}_{group}(t)$ and $\{MeSH\}_{Bi}$, we average appropriately over the paper-set distances. For instance, meaningful and robust results can be obtained by identifying a pre-specified fraction of the papers in $\{MeSH\}_{group}(t)$ that are closest to the set $\{MeSH\}_{Bi}$ and then computing the average paper-set distance for this subset of papers.

Step 5 estimates the level of activity of the group in the "good" and "bad" domains, as a function of time, based on the time series of MeSH distances obtained in Step 4. We form this estimate by building a hidden Markov model (HMM) for scientific production and then using this model to map MeSH distance times series to research activity level. For instance, one useful implementation employs an HMM which has four states, corresponding to the research activity levels which would be expected of: 1.) no workers, 2.) an individual (e.g., postdoctoral) investigator, 3.) a team of researchers under the direction of a single Principal Investigator (PI), and 4.) a multi-PI program. Associated with each HMM state is a probability distribution characterizing the likelihood of observing the various MeSH distances when a group is in that activity state; these distributions can be obtained through analysis of PubMed data on biological sciences research. This HMM then permits a time series of MeSH distances for a particular hypothesis to be mapped to the most probable sequence of research activity levels for the group for that hypothesis (e.g., using the standard Viterbi algorithm [11]). Additional details regarding results obtained using this procedure may be found in [9].

As an illustration of the efficacy of this analytic approach, we briefly summarize a retrospective study of the Institute of Applied Microbiology in Obolensk, Russia. During the period of our study (1970 through the mid-1990s), this laboratory was ostensibly conducting

research on infectious diseases of importance to the USSR (e.g., TB). However, it is now known (e.g., [12]) that for a portion of this period the laboratory was involved in BW research, for example working to genetically engineer antibiotic-resistant strains of plague and tularemia and to develop bacteria-toxin combinations (e.g., myelin toxin and plague). Indeed, around 1980 an aggressive BW research program was initiated at the institute, and this program went on until about 1992, when Russian President Boris Yeltsin decreed that the BW work would stop. An application of the graph-based analysis method summarized above to the Obolensk laboratory easily identifies this BW research activity despite the aggressive "denial and deception" program implemented by the Soviets during most of this period. For instance, the plots of MeSH distances shown in Figure 3 reveal the dramatic increase in "bad" biology research relative to "good" biology research during the period 1980-1992. It is noted that the results of the HMM analysis described in Step 5 (not shown) are even more definitive.

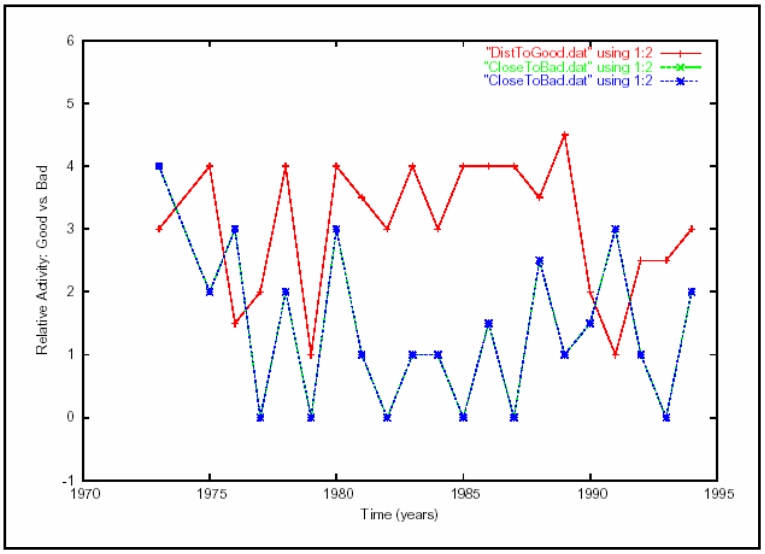


Figure 3: Sample result from Obolensk study. The plot shows MeSH distance v. time, with red depicting distance between institute terms and $\{MeSH\}_G$ and blue showing distance between institute terms and $\{MeSH\}_B$. Note that the change in research focus which occurred in the 1980s is clearly visible.

We note in closing that we have applied the proposed BW research assessment method to targets of current interest to the intelligence community. As in the Obolensk study summarized above, this application of graph analysis has produced results which would have been very difficult to obtain using standard intelligence analysis approaches [9].

4. NW threat discovery

There is concern within the national security community over the possibility that a substate group could acquire a NW [e.g., 13-16]. One particularly alarming scenario involves a group avoiding the significant challenges associated with obtaining weapons-grade Uranium or Plutonium and instead acquiring one of the "nontraditional" nuclear materials that can be used to construct a fission bomb [16]. For example, materials exist that represent attractive alternatives as NW materials, particularly for substate groups, as their weaponization-related properties are comparable to Uranium 235 and they may possess advantages from the perspective of material acquisition (e.g., fewer controls, easier separation). It is crucial to objectively assess the importance of this threat. Key questions include: Are substate groups interested in nontraditional approaches to NW (NT-NW)? Can technically sound NT-NW approaches be discovered by a non-expert adversary? Note that an affirmative answer to the latter question significantly increases the importance of the threat.

In what follows we demonstrate that progress can be made in addressing these questions by adopting a graph-based, discovery-oriented approach to analysis. It is natural to formulate the first question as a discovery problem: Can we discover *some* group which is interested in *any* NT-NW methodology? Interestingly, the second question becomes discovery-oriented if we place ourselves in the position of the adversary and ask: Can we discover a technically correct approach to developing a NT-NW without access to specialized NW expertise or classified information? Graph-based methods are particularly well-suited for such discovery problems. Most approaches to discovery involve searching for "unusual" features in problem-relevant data and then exploring possible explanations and implications of such features. Representing the discovery problem and associated data as graphs enables a semantically rich description and characterization of "unusual"; for instance, graphs can be used to model a heterogeneous collection of objects, their attributes, their relationships with each other, and the way the objects and their interdependencies change over time. As a consequence, even subtle anomalies in a system can be detected by analyzing the system's graph representations, leading to enhanced breadth and depth in the discovery process.

Consider first the problem of determining whether there is interest in NT-NW on the part of any substate group which poses a plausible threat. There is, of course, an enormous amount of information available in both classified and open sources on the activities and intentions of various groups. In fact, terrorist and extremist groups are increasingly using the www to recruit members, attract resources, disseminate ideologies, and even plan and coordinate operations, so that open, Internet-based sources can be surprisingly valuable. In view of this availability of

vast sets of *potentially* relevant data, a key challenge is to develop methods for sorting through these data sets and extracting information which is sufficiently relevant and reliable to be of *genuine* interest.

Consider the problem of exploring the www in search of reliable evidence of interest in NT-NW on the part of a substate group which poses a realistic threat. We adopt the following computational graph analytic approach to this problem:

Input: a collection of hyperlinked web objects (e.g., the www itself, or a subset of the www such as the Dark Web collection [17]) and a set of keywords associated with NT-NW (note that such keywords are readily obtainable via interaction with domain experts [16]).

Procedure:

1. Construct a graph G_{web} from the collection of hyperlinked web objects by defining two types of vertices – web pages and the concepts which appear in them – and two kinds of edges – hyperlinks between web pages and links connecting concepts to pages which contain them.
2. Explore the graph G_{web} using a "web crawler", that is, a program which browses a collection of hyperlinked documents in a systematic manner by "following" hyperlinks. Retain those web pages which are deemed potentially relevant based on a simple (multilingual) keyword query, denoting this set of documents D_{pr}.
3. Perform a simple translation of any non-English documents in the set D_{pr}.
4. Identify task-relevant documents in D_{pr} by:
    - finding concepts which are relevant to the discovery problem and are not already keywords (e.g., concepts that co-occur significantly with keywords in the documents in D_{pr});
    - retaining those documents which contain significant occurrences of keywords and these newly identified concepts; denote this set of documents by D_r.
5. Assess the "authoritativeness" of the web pages hosting the documents in D_r through analysis of the hyperlink structure in the neighborhood of each web page [e.g., 6,10], and collect the documents from authoritative web pages in a set D_{ra}. Note that this step is critical because many web pages claim affiliation with groups representing plausible threats, and it is necessary to distinguish those which actually represent the ideas and activities of such groups from those that do not.
6. Nominate the set of documents D_{ra} to intelligence analysts for further study.

We now provide additional details concerning the above procedure and its implementation. Step 1 typically requires little effort, as collections of hyperlinked web objects already have a

well-defined graph structure and only “concept vertices” and their associated links must be added. Nevertheless, this step is important because the resulting graph representation of the data is convenient for subsequent analysis and, moreover, the process of collecting the data in stable archives is essential given the volatility of the web presence of some of the groups of interest (e.g., terrorist and extremist groups); see [17] for a discussion of this latter point.

Steps 2 and 3 are by now standard operations in web mining applications, and many “off the shelf” tools exist which can perform these tasks. In this study we employ tools developed by the Artificial Intelligence Laboratory at the University of Arizona, which is also the group generating and managing the Dark Web archive [17]. In Step 4 we first form a document-concept bipartite graph and “project” this graph, based on concept co-occurrence, to obtain a concept-concept graph; see Figure 2 and [e.g., 6,9,10,18] for details. This concept-concept graph can then be analyzed to discover new problem-relevant concepts, for instance by identifying concepts which are closely related to the keywords in the graph. Given the set of relevant keywords and concepts, assessing the relevance of documents containing them is straightforward [e.g., 17].

There are a number of ways to assess the “authoritativeness” of a web page based on hyperlink topology and/or page content. In Step 5 we employ a variant of the popular HITS algorithm [e.g., 10]. The basic idea behind HITS is to simultaneously identify “hubs”, which are web pages that reliably hyperlink to pages which are authoritative sources of information in a given domain, and “authorities”, which are those pages consistently linked to by the hubs. Our version of HITS accounts for particular properties of the hyperlink structure found in web pages associated with extremist and radical groups [18]. Finally, Step 6 initiates an iterative interaction between intelligence analysts and the computational analysis performed in Steps 2-5.

As an illustration of the efficacy of this analytic process, we briefly summarize our search for evidence of interest in NT-NW by substate groups. The analysis uses as its data source the www itself, with “seed” web pages obtained from the Dark Web archive [17], and uncovers authoritative web pages expressing interest in NT-NW. For instance, the analysis identifies one Al Qaeda-related site that provides a wealth of useful, technically sound information on nuclear weapons theory and design. Additionally, this site encourages the search for nuclear bomb materials which are “… effective alternative[s] to Uranium and available on the market” (see Figure 4). Further details regarding this and other findings obtained during this investigation are given in [18].

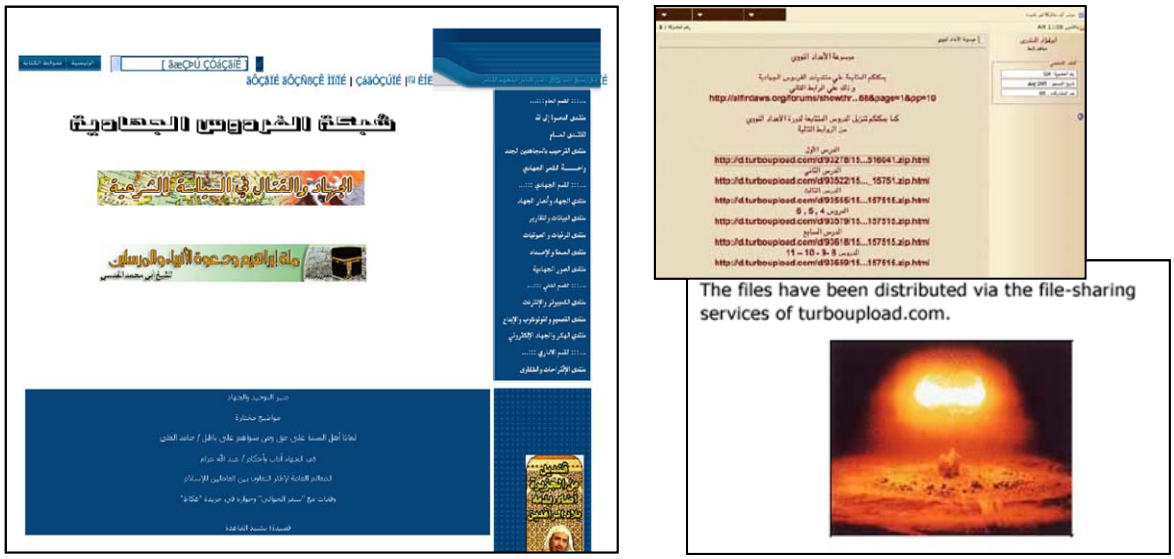


Figure 4: Screenshots from AQ-related website which provides detailed information on NW, in Arabic and English, including endorsement of the advantages of NT-NW methods.

Having discovered evidence that individuals associated with groups that pose a plausible threat are interested in NT-NW, we turn our attention to the question of *discoverability* of NT-NW information. Thus we place ourselves in the position of a technically sophisticated, but non-expert, substate group, and explore the extent to which detailed, technically correct information on NT-NW design and development can be found in open sources. We begin this exploration with a focus on the www, and then remark on the discoverability of such information in other open sources (e.g., scientific publications) at the end of the section.

Consider, then, the problem of finding detailed, reliable NT-NW information on the www. We adopt the following computational graph analytic approach:

Input: the www and a set of generic keywords associated with NT-NW (e.g., basic NW terms like ‘criticality’ and ‘fission’, together with a list of candidate NW materials, such as all radioactive elements).

Procedure:

1. Select a set of web pages which are “NW-related”, to serve as seeds for the computational exploration of the www.

2. Explore the www, starting from the seed web pages, using a web crawler. Retain those web pages which are deemed potentially relevant based on a simple (multilingual) keyword query, denoting this set of documents D_{pr}.
3. Perform a simple translation of any non-English documents in the set D_{pr}.
4. Assess the relevance of the documents in D_{pr} and the authoritativeness of the web pages hosting these documents using the approaches summarized earlier in this section; denote the resulting set of relevant, authoritative documents by D_{ra}.
5. Generate a concept/document bipartite graph G_{cd} from the documents in the set D_{ra}.
6. Perform a structured, graph-based anomaly detection study of G_{cd} and identify candidate "interesting science" concept groups.
7. Subject the set of interesting science concept groups to further (e.g., manual) analysis.

We now provide additional details regarding the above procedure and its implementation. Note first that in this study we are careful to employ only modest levels of *a priori* NW expertise, to ensure that we obtain a fair assessment of the discoverability of NT-NW information by non-experts. Thus, for instance, the keywords which serve as input to the procedure are assembled using less domain knowledge than is provided in the (extremist) web page depicted in Figure 4. Similarly, the NW-related seed web pages we use in Step 1 are fairly obvious choices: the home pages of several university physics and nuclear engineering departments and a few national laboratories [18].

Steps 2 – 5 are implemented exactly as described in the procedure presented earlier in this section. In Step 6, information encoded in the graph G_{cd} is used to identify groups of concepts, and associated web sites, which represent anomalous science activity. The motivation here is that anomalous research and development activity in the NW domain conducted by reputable organizations may provide useful insight into novel NW ideas, including NT-NW. For example, it is straightforward to define distance between a pair of concepts in G_{cd} using concept co-occurrence (because the extent to which two concepts appear together in documents provides a measure of their proximity) [9,10,18]. Once these concept-concept distances are computed, we search for pairs of concepts which are treated as closely related by one authoritative organization (i.e., exhibit significant co-occurrence in that organization's documents) and distant by other such organizations. This behavior is potentially interesting because it may indicate a technically sound relationship between concepts relevant to NW which is not generally acknowledged. Finally, in Step 7, the set of potentially interesting NW concepts identified in Step 6 is examined in greater detail.

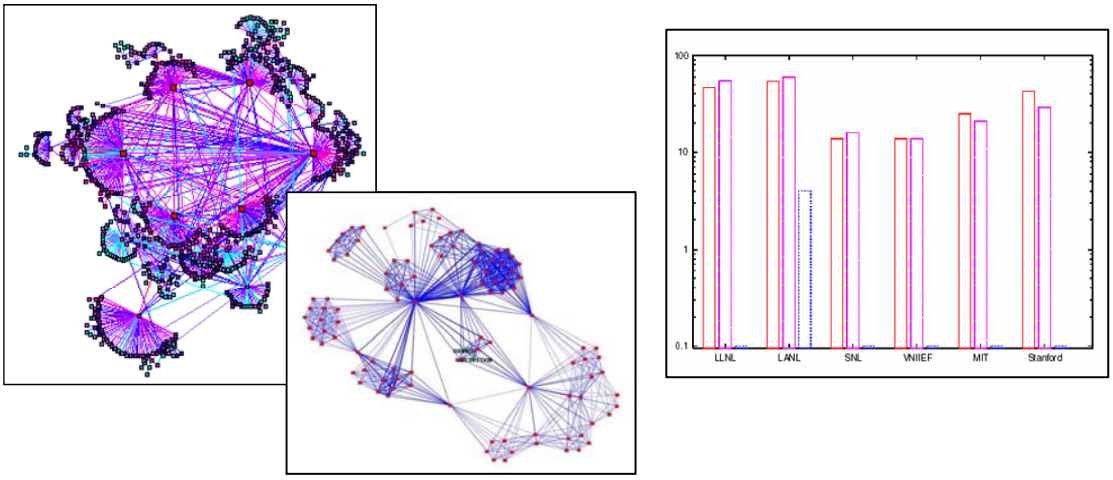


Figure 5: Graph-based discovery. Figures on left depict illustrative examples of a webpage graph and concept-concept graph. Bar chart on right is output of the "interesting science" algorithm; the anomaly associated with discussions related to a (particular) NT nuclear material and NW (blue bar) is clearly visible.

As an illustration of the effectiveness of the procedure outlined above, we briefly summarize our analysis of the discoverability of accurate NT-NW information on the www. Implementing the analysis procedure as described in Steps 1–7 demonstrates that detailed, technically sound NT-NW information is readily discoverable, even by non-experts. For instance we find that, as expected (and can be seen from Figure 5), the concepts Uranium (red bar) and Plutonium (pink bar) are treated as closely related to NW by all authoritative sources in our collection. However, only one source treats another NT nuclear material (blue bar) as being closely related to NW concepts (see Figure 5). This is potentially interesting, as it may indicate that this material would be a useful for NW but that this fact is considered sensitive and not widely acknowledged. Examining this possibility using simple www search queries quickly shows that this is, in fact, the case, and yields detailed information regarding design and construction of a NW based on this NT material [18]. We mention in closing that an analogous study using computational exploration of scientific publications rather than the www produces similar results.

5. Early warning for Muslim mobilization

There is considerable interest within the national security community to understand Islamic radicalization processes and to leverage this understanding to develop methods for anticipating radicalization [e.g., 19-22]. For instance, it would be very useful to identify reliable, practically measurable "early indicators" of impending radicalization, as this would enable more timely and

effective allocation of limited analytic, diplomatic, military, and economic resources. Predicting the evolution of complex social systems, such as radicalization processes, is quite difficult. In fact, recent experiments suggest that there are fundamental limits to the predictability of many social processes, at least if the predictions are to be formed using standard methods [23-27]]. This motivates the search for new, scientifically sound, practically implementable methods for early warning and prediction for social processes.

As indicated in Section 2, a crucial step in our approach to predictive analysis is to assess the predictability of the process of interest. Among other things, predictability assessment enables identification of those measurables which are most useful for prediction. We expect this function to be important here, as radicalization appears to emerge from a complex interplay between intrinsic features of the situation, such as regime oppressiveness or perceived legitimacy, and the collective dynamics which define the way individual beliefs and preferences become collective actions. Consequently there are myriad measurables associated with radicalization which *may* have predictive power, and identifying which (if any) actually do is both challenging and critical for successful prediction. We also expect that, given a context within which the potential for Islamic radicalization exists, social network dynamics play a critical role in determining the viability and ultimate extent of any emerging radicalization. Therefore it is natural to adopt a graph-based approach to modeling and analyzing radicalization phenomena. As shown below, this approach allows identification of a new and potentially valuable early indicator of radicalization-related Muslim mobilization.

In what follows, we provide a brief summary of our approach to assessing predictability and identifying measurables which are useful for prediction, introduce a graph-based framework within which to model social dynamics, and then apply these methodologies to the problem of early warning analysis for mobilization related to Islamic radicalization. Additional details concerning the proposed approach to predictive analysis and social dynamics modeling may be found in [25,26], while background on situational awareness and warning analysis for radicalization is given in [19-21,27].

We formulate prediction problems as questions about the expected dynamics of a network of interest. Given a social process, a set of candidate process measurables, and the behavior about which predictions are to be made, we represent the process as a stochastic dynamical system evolving on some state space, relate the observables to process characteristics (e.g., parameters) and/or states, and encode the behavior of interest in terms of the satisfaction of reachability conditions.

As an illustrative example, consider the problem of predicting ultimate market share in a cultural market (e.g., for music or films) in which “buzz” about products propagates through various social networks. If, in a market containing two products with indistinguishable “intrinsic appeal”, it is possible for one of the products to achieve a dominant market share, we might view the market to be unpredictable. Conversely, a predictable market would be one in which market shares of indistinguishable products evolve similarly and market shares of superior products are typically larger than those of inferior ones. Prediction, of course, then involves estimating the ultimate market share of a product of interest, perhaps based on measures of appeal. In our formulation, market share dominance by product A is associated with a region of market share state space, and the condition that A eventually achieves such dominance and simultaneously possesses an appeal that is indistinguishable from product B is expressed as the satisfaction of a state space reachability condition.

To formulate prediction questions in terms of reachability, the behavior about which predictions are to be made is used to define the system *state space subsets of interest* (SSI), while the particular set of candidate measurables under consideration allows identification of the *candidate starting sets* (CSS), that is, the set of states and system parameters which represent initializations that are equivalent under the assumed observational capability. Predictability assessment then involves determining which SSI can be reached from CSS. If the network’s reachability properties are incompatible with the prediction goals – if, for instance, “hit” and “flop” are both reachable from a single CSS – then the given prediction question should be refined in some way. Possible refinements include relaxing the level of detail to be predicted (by redefining the SSI) or using additional measurables to resolve the CSS. If and when a predictable situation is obtained, the problem of forming robust, useful predictions can be addressed. This problem is also naturally studied within the reachability framework, as it involves determining the most likely evolution of the network and quantifying the uncertainty associated with this estimate.

The preceding discussion motivates the need to develop a rigorous, tractable methodology for assessing reachability of complex network processes in the presence of uncertainty regarding the network. Particularly desirable are methods that can be implemented both in stochastic settings, where probabilistic characterizations of the uncertainty are available, and nonstochastic situations, for instance where bounds on the uncertainty are known.

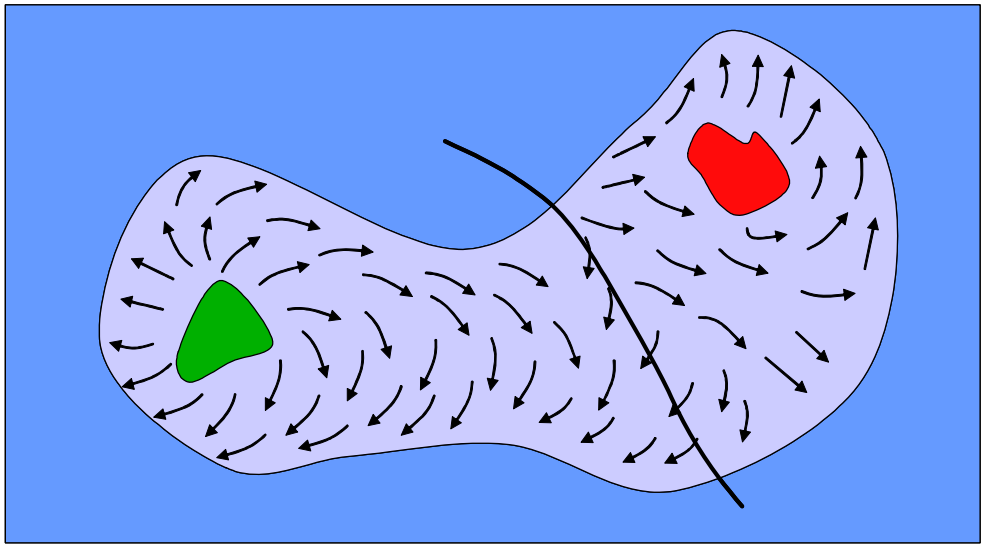


Figure 6: Cartoon illustrating reachability analysis using proposed approach.

We now briefly summarize an approach to reachability assessment which possesses these characteristics. Consider a network with state space X, and let X_0, X_u, and D denote the sets of initial states, "undesirable" states, and admissible parameter values, respectively. Suppose we wish to show that no system trajectory starting from X_0 can evolve to the set X_u for any admissible parameters. We adopt an analysis methodology which is analogous to the one underlying Lyapunov function-based stability analysis [e.g., 28]: we seek a scalar "altitude" function of the network state A(x) which permits reachability to be deduced *without computing system trajectories*. The basic idea is depicted in Figure 6. If we can find a function A(x) with a "level curve" (say, the set of states for which A(x) = 0) that separates the green and red sets of states and on which the system flow points from red to green, then we can conclude that the red set (X_u) is not reachable from the green set (X_0); moreover, we reach this conclusion without explicitly computing the set of states reachable from the green set.

This altitude function methodology can be extended for application to stochastic reachability problems, in which an estimate of the *probability* of reaching an SSI from a given initial set is desired. More specifically, suppose we can find a nonnegative function A(x) which is 1.) bounded from above by γ on the set of initial states X_0, 2.) bounded from below by 1 on the set of undesirable states X_u, and 3.) nonincreasing in expectation along all system trajectories (e.g., for all network parameters in D). In this case it can be proved that γ is an upper bound on the probability of reaching X_u from X_0 [25,26]; lower bounds on this reach probability can be computed analogously [25,26].

The trick, of course, is to find such a function A(x) or prove that no such function exists. Recent work in semidefinite programming (SDP) and semialgebraic geometry [29] provides a powerful framework within which to perform this analysis. Briefly, convex relaxation is used to transform the search for A(x) into a sums of squares (SOS) optimization problem, which can be solved using SDP. Software for solving SOS programs is available as a third-party Matlab toolbox [30], so that this analysis is convenient to implement. Moreover, the approach is tractable: the computational complexity of the associated SOS program grows polynomially in the dimension of the network's state space X and parameter space D.

We now turn to the problem of identifying useful, practically measurable early indicators of Islamic mobilization. It is widely recognized that collective dynamics and social network effects are central to mobilization phenomena (e.g., recruitment, attracting support, development and dissemination of ideas, expressive action) [2,5,19-22,31,32]. However, there has been little done to rigorously explore the implications of these dynamics for anticipating mobilization. Here we show how graph-based analysis, and in particular the stochastic reachability formulation of network predictability summarized above, can be leveraged to identify patterns in data which appear to be useful predictors of mobilization. A key step in this analysis process is the formulation of appropriate social dynamics models.

Recent work has clearly demonstrated the importance of capturing social network effects when modeling social processes [e.g., 10]. However, detailed information concerning the relevant social networks is often not available in national security applications. These facts have motivated our development of a class of models for social network processes which employs three modeling scales:

- a *micro-scale*, for modeling the behavior of individuals;
- a *meso-scale*, which represents the collective dynamics within social network communities via simple models for these interactions;
- a *macro-scale*, which characterizes the interaction between the social network communities.

Social network communities are localized social settings, determined by kinship, workplace environment, or physical neighborhood, in which interaction between individuals can be modeled as "fully mixed" – that is, all pairwise interactions between individuals within a social community are equally likely. A schematic of the basic framework is given in Figure 7. Note that this approach simultaneously alleviates the need for detailed social network data, because interactions within social communities are modeled as fully mixed, and captures the important social network structure by modeling the way communities are related. Moreover, the

characterization of intra-community and inter-community dynamics implicit in the proposed multi-scale model is well supported by social science research [e.g., 33].

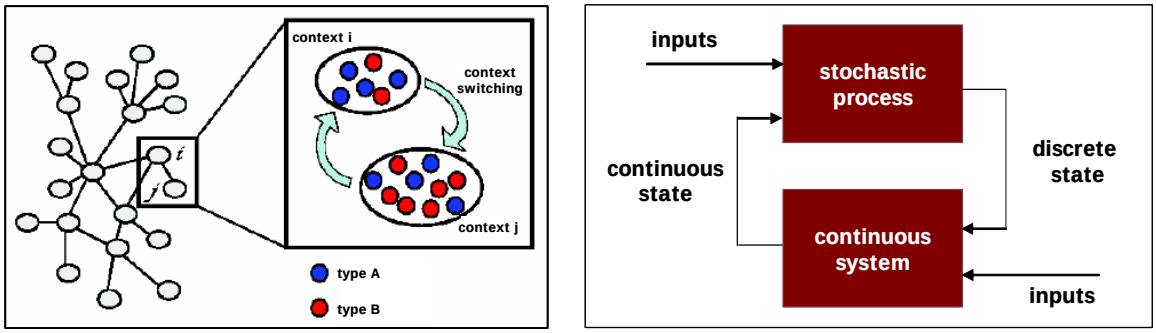


Figure 7: Multi-scale model for social network processes. Cartoon at left illustrates the basic model structure and block diagram at right depicts a hybrid dynamical system representation of the model.

We develop and analyze multi-scale social network dynamics models using the hybrid dynamical system formalism [34]. Briefly, hybrid systems are feedback interconnections of continuous dynamics, such as the dynamics of individuals exchanging ideas within a social community, and discrete dynamics, capturing for instance the switching behavior encountered when an individual from one community moves to another and introduces an idea which is novel in the latter community (see Figure 7). An advantage of representing multi-scale social dynamics using a hybrid system framework is that the resulting models are strongly structured, enabling rigorous and tractable analysis. For example, the altitude function approach to predictability analysis summarized above can be conveniently conducted using hybrid system models [25,26].

The social system modeling and predictive analysis methodologies summarized above are now applied to the problem of identifying early indicators of Islamic mobilization. We begin by briefly examining the early warning problem for *social movements*, a general class of social processes which includes radicalization as well as other types of social action. This broader setting is reasonably well-characterized both empirically and theoretically [27,31,32], and thus provides the opportunity to identify candidate early indicators in a principled way; these candidates indicators can then be tested for relevance to Islamic mobilization.

Reachability-based predictability assessment of several social movements (e.g., the emergence of the Swedish Social Democratic Party, the Leipzig “Monday demonstrations”) shows that a useful distinguisher between successful and unsuccessful movements is the

degree to which movement-related activity shows *early diffusion* across multiple social communities. In particular, this measurable has significantly more predictive power than *volume* of such activity and also more power than the various system intrinsics which are ordinarily considered (e.g., movement ideology, political environment) [27]. This result suggests social network dynamics are critical to social movement success. Moreover, the analysis shows that features of these dynamics which may be useful early indicators of movement success are practically measurable; for instance, diffusion across social communities may be inferred from analysis of online (i.e., www-based) data using only incomplete information regarding the relevant social networks. These results indicate the significant potential utility of the proposed network-oriented approach to predictive analysis for such processes.

We now investigate whether diffusion across social communities is a useful early indicator for Islamic mobilization. The study focuses on Muslim reaction to six recent incidents, each of which appeared at their outset to have the potential to trigger significant mobilization activities:

- publication of photographs and accounts of prisoner abuse at Abu Ghraib in Spring 2004;
- publication of cartoons depicting the prophet Mohammad in the Danish newspaper *Jyllands-Posten* in September 2005;
- distribution of the DVD “I was blind but now I can see” in Egypt in October 2005;
- the lecture given by Pope Benedict XVI in September 2006 in which he quoted controversial material concerning Islam;
- Salman Rushdie being knighted in June 2007;
- republication of the “Danish cartoons” in various newspapers in February 2008.

Recall that the first Danish cartoons event ultimately led to substantial Muslim mobilization, including massive protests and considerable violence, and that the Egypt DVD event also resulted in significant Muslim mobilization and violence. In contrast, Muslim outrage triggered by Abu Ghraib, the pope lecture, the Rushdie knighting, and the second Danish cartoons event all subsided quickly with essentially no violence. Therefore, taken together, these six events provide a useful setting for testing whether the extent of early diffusion across social communities can be used to distinguish nascent Islamic mobilization events which become large and self-sustaining (and potentially violent) from those that quickly dissipate.

A central element in the proposed approach to early warning analysis is the measurement, and appropriate processing, of social dynamics associated with the process of interest. In the present study we use online social activity as a proxy for “real world” diffusion of mobilization-relevant information. More specifically, we use blog discussions as our primary data set. The

“blogosphere” is modeled as a graph composed of two types of vertices, the blogs themselves and the concepts which appear in them. Two blogs are linked if a post in one hyperlinks to a post in the other, and a blog is linked to a concept if the blog contains (a significant occurrence of) that concept; see Figure 8 for a cartoon of the basic setup and a real world example of a blog hyperlink graph. Among other things, this blog model enables the identification of blog graph communities, that is, groups of blogs for which the intra-group edge densities are significantly higher than the inter-group edge densities [e.g., 10]. These blog communities serve as one of our proxies for social communities.

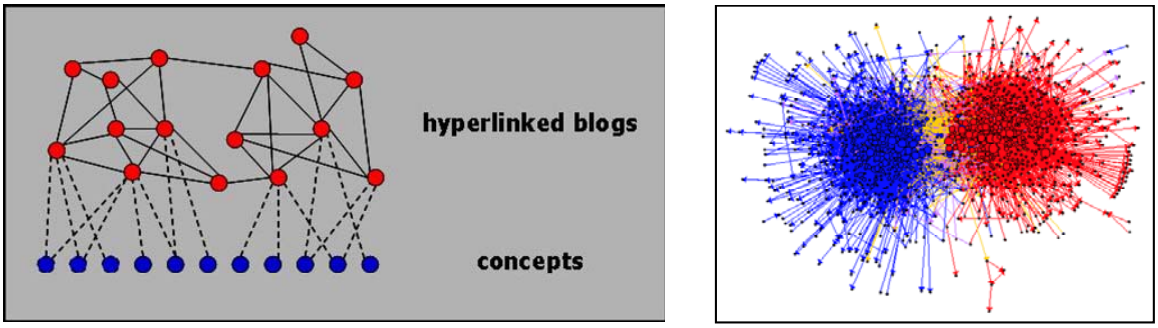


Figure 8: Blog graph model. Schematic on left depicts the basic graph structure, in which blogs (red vertices) can be connected to each other via hyperlinks (solid edges) and also connected to concepts (blue vertices) they contain. Blog graph on right corresponds to political blogs; note that in this graph liberal (blue) and conservative (red) blogs form two distinct graph communities.

We propose the following procedure for warning analysis.

Input: a potential “triggering” event of interest.

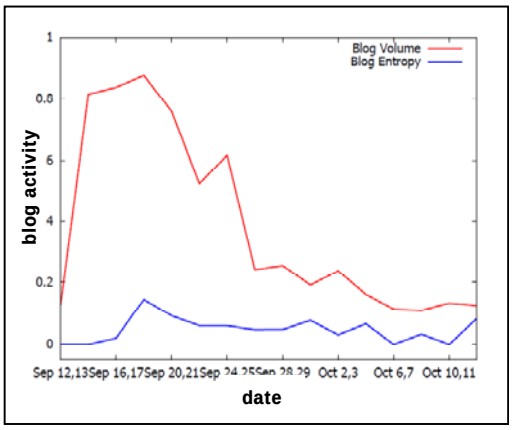
Procedure:

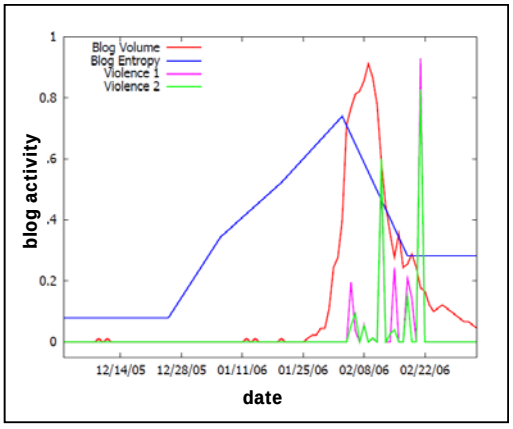
1. Use key words/concepts associated with the triggering event to collect relevant blog posts and build the associated blog graph.
2. Identify the blog communities (e.g., graph community-based, language-based).
3. Construct the post volume time series for each blog community. Compute post/community entropy (PCE) time series associated with the post volume time series. We define the PCE for a given sampling interval t as $PCE(t) = -\Sigma_i f_i(t) \log(f_i(t))$, where $f_i(t)$ is the fraction of total relevant posts during interval t which occur in blog community i.

4. Construct a *synthetic* ensemble of PCE time series from (actual) post volume dynamics using an empirically-grounded social diffusion model (see [25]).
5. Perform *motif detection*: compare the actual PCE time series to the synthetic ensemble series to determine if the early diffusion of activity across communities is “excessive”. Flag events with excessive early diffusion for further (e.g., manual) analysis.

We now provide a few additional details concerning this procedure. Step 1 is by now standard, and various off-the-shelf tools exist which can perform this task. In Step 2, graph-based blog communities are identified through community detection applied to the blog graph [6,10], and language communities are defined on the basis of post language. In Step 3, post volume for a given community i and sampling interval t is obtained by counting the number of relevant posts made in the blogs comprising community i during interval t. Given the post volume time series obtained in Step 3, Step 4 involves the construction of an ensemble of PCE time series that would be expected under “normal circumstances”, that is, if Muslim reaction to the triggering event diffused from a small “seed set” of initiators according to standard social diffusion models [25]. Finally, motif detection in Step 5 is carried out by searching for time periods, if any, during which the actual PCE time series exceeds the mean of the synthetic PCE ensemble by at least two standard deviations.

Sample results of applying the proposed early warning approach to Islamic mobilization events are shown in Figure 9. It can be seen that early diffusion of discussions across blog communities is, indeed, an indicator that the associated mobilization event will be large. Such diffusion is observed in the mobilization associated with the first Danish cartoons and Egypt DVD events and not with the other four events, and the observed early diffusion is excessive relative to the synthetic ensembles. For instance, in the case of the first Danish cartoons event the PCE of relevant discussions (blue curve) experiences a dramatic increase a few weeks before the corresponding increase in volume of blog discussions (red curve); this latter increase, in turn, takes place before any violence (see Figure 9). In contrast, in the case of the pope event, PCE of blog discussions is small relative to the cartoons event, and any increase in this measure lags discussion volume. Similar curves are obtained for the other four events. More importantly, the proposed motif detection process also yields the expected result: motifs are found only for the Danish cartoons and Egypt DVD events, and these motifs precede significant blog volume and real world violence. Note that qualitatively similar results are obtained for the graph-based and language-based definitions of social community. This case study suggests that early diffusion of mobilization-related activity (here blog discussions) across disparate social communities may be a useful early indicator of significant mobilization events.





Time series motif analysis

Event	Motif
Danish cartoons 1:	1/1—1/26/2006.
Egypt DVD release:	10/2–10/9/2005.
Abu Ghraib story:	none.
Pope lecture:	none.
Rushdie knighting:	none.
Danish cartoons 2:	none

Figure 9: Sample results for Islamic mobilization case study. The time series plots at the top correspond to the pope event (left) and first Danish cartoons event (right). In each plot, the red curve is blog volume and the blue curve is blog entropy; the Danish cartoon plot also shows two measures of violence (cyan and magenta curves). Note that while the data are scaled to allow multiple data sets to be graphed on each plot, the scale for entropy is consistent across plots to enable cross-event comparison. The table at the bottom summarizes the results of the motif analysis study.

6. Conclusions

This paper illustrates the power and relevance of graph analytics for intelligence analysis by considering three real world case studies drawn from the domains of counterproliferation and counterterrorism. More specifically, the paper examines three broad and important classes of intelligence analysis questions – assessment, discovery, and prediction – and demonstrates that these questions can be naturally formulated as graph problems and that such a formulation enables development and implementation of effective new analytic techniques.

We begin by discussing a classic assessment question: What are the true activities and capabilities of a suspected dual use BW research facility? We show that graph analysis, with its focus on capturing and understanding interdependencies, provides an excellent means of identifying and interpreting the *indirect* relations and effects which are so difficult for an

adversary to conceal when conducting a clandestine program. As a consequence, the application of graph analytic methods enables evidence regarding the actual activities of an organization or facility to be extracted even in the presence of aggressive denial and deception. The efficacy and practical utility of the proposed approach is illustrated through a retrospective study of the Institute of Applied Microbiology in Obolensk, Russia.

We next consider “discovery-oriented” analysis questions, that is, questions whose vague and ill-posed nature reflects our general ignorance concerning the domain of interest. The particular discovery question addressed in this paper focuses on the threat posed by NT-NW and substate groups. We find that graph-based methods are particularly well-suited for such problems. Representing discovery questions as graph problems enables a semantically rich description and characterization of the attributes and behaviors of the system of interest; consequently, “unusual” features of the system, even subtle ones, can be readily detected and interpreted. The utility of graph-based methods of discovery is illustrated through computational exploration of the www for evidence of interest in NT-NW approaches by substate groups and discoverability of technically sound NT-NW information by non-experts.

Finally we consider predictive analysis, and in particular the problem of developing an early warning capability for Islamic radicalization processes. A key component of predictive analysis is explicit consideration of the predictability of the problem of interest. We demonstrate that computational methods for assessing predictability and identifying measurables with predictive power can be developed by adopting a graph-based analytic framework. The power of the proposed approach to predictive analysis is illustrated through development of a novel, high potential algorithm for early warning analysis for Islamic mobilization.

Much work must be done if we are to realize the considerable potential of graph analysis approaches to intelligence analysis problems. For example, scientifically grounded predictive analysis for systems of interest to the intelligence community is still in its infancy, and we believe this domain represents a fertile topic of inquiry. Another key research area appears to be the use of high performance computing methods for the sorts of graph calculations which arise in intelligence problems. And, of course, it is essential to explore the implementation of graph analytics in other intelligence analysis domains. Among others, we believe that the counterintelligence area may be particularly well-suited for graph analysis. For instance, this paper has shown the utility of graph methods for addressing deception, discovery, and warning/prediction, and these are all central concerns in counterintelligence [35].

7. Acknowledgements

This work was supported by the U.S. Department of Defense (DoD), the U.S. Department of Homeland Security (DHS), and the Laboratory Directed Research and Development Program of Sandia National Laboratories (SNL). We gratefully acknowledge many helpful discussions on the topics of this paper with Chip Willard of DoD, Joe Kielman of DHS, and Rick Contreras of SNL. We thank the reviewers and editors of the Journal of Intelligence Community Research and Development for suggestions and feedback which improved the paper significantly.

8. References

1. Coffman, T., S. Greenblatt, and S. Marcus, "Graph-based technologies for intelligence analysis", *Communications of the ACM*, Vol. 47, pp. 45-47, 2004.
2. Sageman, M., *Understanding Terror Networks*, University of Pennsylvania Press, 2004.
3. Chen, H., C. Yang, M. Chau, and S. Li (Editors), *Intelligence and Security Informatics*, Lecture Notes in Computer Science, Springer, Berlin, 2009.
4. Skillicorn, D., *Knowledge Discovery for Counterterrorism and Law Enforcement*, CRC Press, Boca Raton, FL, 2009.
5. Memon, N., J. Farley, D. Hicks, and T. Rosenorn (Editors), *Mathematical Methods in Counterterrorism*, Springer, New York, 2009.
6. Newman, M., *Networks: An Introduction*, Oxford University Press, 2010.
7. Henderson, D., "The looming threat of bioterrorism", *Science,* Vol. 283, pp. 1279-82, 1999.
8. Block, S., "The growing threat of biological weapons", *American Scientist*, Vol. 89, pp. 2-11, 2001.
9. Colbaugh, R. and K. Glass, "Discovery and analysis in large, incomplete data: proliferation examples", Presentation (with Supplementary Materials) at 2005 ITIC Knowledge Discovery and Dissemination Workshop, Washington, DC, November 2005.
10. Newman, M., "The structure and function of complex networks", *SIAM Review*, Vol. 45, pp. 167-256, 2003.
11. Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, Vol. 77, pp. 257-286, 1989.
12. Alibek, K., *Biohazard*, Random House, NY, 1999.
13. Workshop on Role of Nuclear Physics Research Community in Combating Terrorism, U.S. Department of Energy, Washington, DC July 2002.
14. Allison, G., *Nuclear Terrorism*, Times Books, New York, 2004.
15. Bunn, M. and A. Wier, "Terrorist nuclear weapon construction: How Difficult?", *Annals of the AAPSS*, Vol. 607, pp. 133-149, 2006.
16. Personal communication with seniors at the National Security Agency, Spring 2005.
17. Chen, H., "Dark Web Terrorism Research", http://ai.bpa.arizona.edu/research/terror, 2007.

18. Colbaugh, R., K. Glass, and J. Gosler, “CBRN threat assessment via complex networks analysis”, Presentation to Deputy Director, National Security Agency, September 2005.

19. Post, J., K. Ruby, and E. Shaw, “The radical group in context: An integrated framework for the analysis of group risk for terrorism”, *Studies in Conflict and Terrorism*, Vol. 25, pp. 73-100, 2002.

20. Silber, M. and A. Bhatt, “Radicalization in the west: The homegrown threat”, New York City Police Department Report, 2007.

21. Radicalization: Forecasting and Warning, CENS-GFF Workshop Report, Singapore, February 2008.

22. Ackerman, G., et al., “Anticipating rare events: Can acts of terror, use of WMD, or other high profile acts be anticipated?”, White Paper, November 2008.

23. Salganik, M., P. Dodds, and D. Watts, “Experimental study of inequality and unpredictability in an artificial cultural market”, *Science*, Vol. 311, pp. 854-856, 2006.

24. Colbaugh, R. and K. Glass, “Predictability and prediction of social processes”, *Proc. 4th Lake Arrowhead Conference on Human Complex Systems*, Lake Arrowhead, CA, April 2007.

25. Colbaugh, R. and K. Glass, “Predictive analysis for social processes”, Sandia National Laboratories SAND Report 2009-0584, January 2009.

26. Colbaugh, R. and K. Glass, “Predictive analysis for social processes I: Multi-scale hybrid system modeling, and II: Predictability and warning analysis”, *Proceedings of the 2009 IEEE Multi-Conference on Systems and Control*, Saint Petersburg, Russia, July 2009.

27. Colbaugh, R. and K. Glass, “Early warning analysis for social diffusion events”, *Proceedings of the 2010 IEEE International Conference on Intelligence and Security Informatics*, Vancouver, Canada, May 2010.

28. Sontag, E., *Mathematical Control Theory*, Second Edition, Springer, NY, 1998.

29. Parrilo, P., *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD dissertation, California Institute of Technology, Pasadena, CA, 2000.

30. SOSTOOLS, http://www.cds.caltech.edu/sostools/, 2007.

31. Della Porta, D. and M. Diani, *Social Movement*, Second Ed., Blackwell, Oxford, UK, 2006.

32. Easley, D. and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

33. Watts, D., P. Dodds, and M. Newman, “Identity and search in social networks”, *Science*, Vol. 296, pp. 1302-1305, 2002.

34. Bemporad, A., A. Bicchi, and G. Buttazzo, eds., *Proc. 10th International Conference on Hybrid Systems: Computation and Control,* Pisa, Italy, April 2007.

35. Gosler, J.R., “Counterintelligence: Too narrowly practiced”, in *Vaults, Mirrors, and Masks: Rediscovering US Counterintelligence*, Georgetown University Press, 2008.