

# Link mining for contextual prioritization of tactical information

R. E. Marmelstein<sup>1</sup>, M. G. Devito<sup>1</sup>, C. Mungo<sup>2</sup>,  
G. Papanagopoulos<sup>2</sup>, K. Patel<sup>2</sup> & M. H. Linderman<sup>3</sup>

<sup>1</sup>*East Stroudsburg University, East Stroudsburg, Pennsylvania, USA*

<sup>2</sup>*Army Research, Development, and Engineering Command,  
Picatinny Arsenal, New Jersey, USA*

<sup>3</sup>*Air Force Research Laboratory – Information Directorate,  
Rome, New York, USA*

## Abstract

To address the problem of information overload in military Command and Control (C2), we have prototyped a Tactical Information Prioritization System (TIPS). TIPS performs link mining on relational, C2 databases in order to dynamically determine the relevance of a given report to a generic military unit of interest. Emerging C2 database schemas, such as C2 Integrated Exchange Data Model (C2IEDM), support this goal by providing multiple, overlapping pathways to relate disparate battlespace objects. By leveraging the C2IEDM schema, we have developed a set of queries to mine context-sensitive relationships for a given report/unit pairing. These findings are then interpreted by a Bayesian Network to prioritize a set of reports for transmission to the specified unit. Our testing shows that the TIPS recommendations approximate those of group of human analysts.

*Keywords: context awareness, command and control, link mining, Bayesian Networks, information prioritization, C2IEDM.*

## 1 Background

Communication is always a difficult task; this is especially true in combat. Warfighters must receive the information necessary for them to accomplish their mission in a timely manner. Conversely, an excess of information can degrade unit effectiveness and jeopardize mission success. Supplying the right



information is also important because of other constraints, such as short-duration communication “windows” or limited bandwidth. For these reasons, it is important that information be properly prioritized. Ideally, communications to units operating in the battlespace have the following characteristics:

- The information sent must be relevant to the unit’s context.
- The information must be sent in a timely manner.
- Both information overload & withhold must be avoided.
- Information must be sent according to its priority—subject to human factors and bandwidth constraints.

With the above in mind, it is highly desirable to have an automated capability to mine *contextual linkages* between a Reported Data Item (RDI) and some military Unit of Interest (UOI). Contextual linkages are operationally pertinent relationships between the two objects; the number and nature of those relationships determines the *strength* of the linkage. Finding numerous strong, contextual linkages between the objects makes it highly likely that the report is relevant to the unit. To accomplish this in a general sense, the context assessment algorithm must have no a priori knowledge or assumptions about either the RDI or the UOI. Mining contextual linkages to determine relevance distinguishes this approach from others merely search for semantic relationships based on a lexicon [4].

## 2 Approach

TIPS is an automated system to prioritize the transmission of operational messages to a given military unit [8,9]. The system is made up of three essential components: a robust database schema that allows for multiple linkages between the RDI and the UOI, a set of tests to detect the presence of these linkages for a given unit/message pair, and a Bayesian Network (BN) to compute a priority based on the number and type of discovered linkages. We now look at each of these components in turn.

### 2.1 C2IEDM

The C2IEDM is an emerging standard data model for military command and control developed the under auspices of the Multilateral Interoperability Programme (MIP). The mission of the MIP is to enable interoperability and advance digitization within NATO to support multinational, combined and joint operations [11]. The purpose of the C2IEDM is to model the information that commanders need to exchange for land-based combat operations.

The overall C2IEDM is composed of three data models: Conceptual, Logical, and Physical. The Conceptual Data Model supports general concepts such as actions, organizations, materiel, personnel, features, facilities, locations, etc. The Logical Data Model decomposes (via entity-relationship diagrams) the high level concepts into specific information that is regularly used at the staff level. The



Physical Data Model provides the specifications that define the corresponding database schema. In short, the C2IEDM schema can describe virtually any land battlefield entity, condition, task, or relationship.

## 2.2 Linkage tests

The purpose of the linkage tests is to gather evidence to determine if the RDI should be forwarded to the UOI. The tests themselves are a series of in-depth database queries, which are designed to do the following:

- Establish the degree of validity of the RDI. This is based on message characteristics such as source, accuracy, and freshness (time).
- Probe the existence and extent of the relationship between the RDI and the UOI.
- Determine the proximity of the UOI to locations or object referenced in the RDI

The current set of linkage tests is listed in Table 1. The C2IEDM provides a highly comprehensive and robust framework for specifying relationships between battlespace objects; as such, there are a large number of potential relationship pathways to test. These linkages are formed based on the content of the report. If the report refers to an action task, a capability, an object-item, and a command association, then linkages between the RDI and these objects must eventually form within the database to document these references. Once formed, their relationship (either direct or indirect) to the unit of interest can be established. A direct relationship exists if the RDI refers to some aspect of the UOI itself. An indirect relationship exists if the RDI refers to another object-item that is associated with the UOI.

For example, the RDI may refer to another unit which shares an action task with the UOI. In this case, message may still be of interest to the UOI, even though it is not directly about that unit. Naturally, the speed at which these linkages form within the database is an issue. The more linkages exist, the more accurate the determination of relevance for an RDI/UOI pair. Unfortunately, it may take many hours, if not longer, for the proper linkages to form. This reinforces the need for an overlapping, robust set of tests. In case some linkages will form quickly, while others will not. It also highlights the need for the linkage tests to be run multiple times after receipt, until the number of links reaches steady state or the message has been sent.

Ideally, messages are placed within a priority queue awaiting transmission. While it is in the queue, its priority can be re-evaluated, thus changing its position in the queue. New relationships may increase the priority, while the aging of the RDI might decrease its priority. The implementation of a queue is straightforward, but efficient mechanisms for triggering re-evaluations have not yet been investigated.



Table 1: Mapping of linkage tests to BN nodes.

#	UOI Linkage Test Description	Affected BN Nodes
1	What is the appraisal of the RDI?	<ul style="list-style-type: none"> <li>• Report Integrity</li> <li>• Source Reliability</li> <li>• Source Quality</li> <li>• Credibility</li> <li>• Freshness</li> <li>• Intel Quality</li> <li>• Report Appraisal</li> </ul>
2	Is the UOI already aware of the RDI?	<ul style="list-style-type: none"> <li>• Related Organization</li> <li>• Report Appraisal</li> </ul>
3	Was the RDI generated in response to the UOI's request?	<ul style="list-style-type: none"> <li>• Action Task – Primary</li> <li>• Action Task – Support</li> <li>• Data Criticality</li> </ul>
4	Is the RDI linked to the UOI's action task?	<ul style="list-style-type: none"> <li>• Action Task – Primary</li> <li>• Action Task – Support</li> </ul>
5	Is the RDI linked to the UOI's status?	<ul style="list-style-type: none"> <li>• Related Status</li> </ul>
6	Does the RDI refer to an association that impacts the UOI?	<ul style="list-style-type: none"> <li>• Related Organization</li> </ul>
7	Does the RDI refer to a change in capability that impacts the UOI?	<ul style="list-style-type: none"> <li>• Related Capability</li> </ul>
8	Does the RDI reference the UOI's call sign?	<ul style="list-style-type: none"> <li>• Related Organization</li> </ul>
9	Does the RDI refer to a location close to the UOI?	<ul style="list-style-type: none"> <li>• Related Location</li> </ul>
10	Does the RDI refer to a context that is related to the UOI?	<ul style="list-style-type: none"> <li>• Context – Direct</li> <li>• Context – Indirect</li> </ul>
11	Does the RDI refer to a change in holding that impacts the UOI?	<ul style="list-style-type: none"> <li>• Related Unit Holding</li> </ul>
12	Does the RDI refer to a weather event that impacts the UOI?	<ul style="list-style-type: none"> <li>• Weather Related</li> <li>• Related Location</li> </ul>
13	Does the RDI refer to an obstacle that may impede the UOI?	<ul style="list-style-type: none"> <li>• Obstacle Related</li> <li>• Related Location</li> </ul>
14	Does the RDI refer to a target that is related to the UOI's action task?	<ul style="list-style-type: none"> <li>• Action Task – Primary</li> <li>• Action Task – Support</li> <li>• Message Type</li> </ul>
15	Does the RDI indicate a change in status of a required capability?	<ul style="list-style-type: none"> <li>• Action Task – Primary</li> <li>• Related Capability</li> <li>• Message Type</li> </ul>
16	Does the RDI have content intended for units subordinate to the UOI?	<ul style="list-style-type: none"> <li>• Related Organization</li> <li>• Message Type</li> </ul>

Table 1: Continued.

#	UOI Linkage Test Description	Affected BN Nodes
17	Does the RDI contain information about a resource required for the UOI's mission?	<ul style="list-style-type: none"> <li>• Action Task – Primary</li> <li>• Related Resource</li> <li>• Related Organization</li> <li>• Message Type</li> </ul>
18	Does the RDI contain an assessment of a context related to the UOI?	<ul style="list-style-type: none"> <li>• Context Direct</li> <li>• Message Type</li> </ul>
19	Does the RDI deal with the status of a context related to the UOI?	<ul style="list-style-type: none"> <li>• Action Task – Primary</li> <li>• Context Direct</li> <li>• Related Organization</li> <li>• Message Type</li> </ul>
20	Is the RDI related to the organizational structure of the UOI?	<ul style="list-style-type: none"> <li>• Related Organization</li> <li>• Message Type</li> </ul>

### 2.3 TIPS Bayesian Network

BNs are directed acyclic graphs composed of nodes and arcs which model a generalized Probability Distribution Function (PDF) over some domain. The set of nodes within the BN are equivalent to random variables that define the domain. Each variable may be discrete (having a finite number of countable states) or continuous. The arrows represent dependencies between random variables and are sometimes interpreted as causal relationships. Within each node, a local PDF is defined; this encodes the conditional probabilities for that nodes based on the state of parent nodes. For each node, its PDF defines how its state is conditionally affected by the state of the parent nodes; the source of this

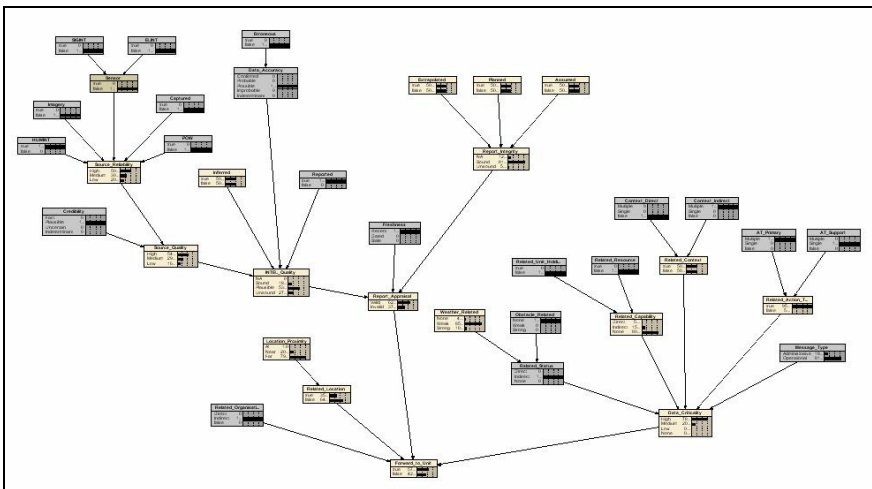


Figure 1: TIPS Bayesian Network.

belief can be either expert knowledge or probabilities learned from data. If a node has no incoming arcs, its PDF defines the apriori likelihood of each state. When we make observations about relationships gleaned from the C2 database, we insert these into the BN as evidence. The BN utilizes this evidence to compute conditional probabilities of the remaining nodes (for which no direct evidence exists). As such, the BN is an extremely useful tool for reaching a decision given what we expect and know to be true within some domain. A more detailed tutorial on BNs can be found in [3].

The findings generated by the linkage test provide the raw data for the BN to interpret. The results of each query are injected as evidence into the nodes of a specially designed BN shown in Figure 1. In addition to summarizing the linkage tests, Table 1 shows the mapping between each test and the BN nodes they affect. It should be noted that the BN structure was not learned from data. Rather, it reflects a domain expert's guidance for prioritizing the RDI given set of evidence.

### 3 Evaluation

The goal of the evaluation process was to compare the message transmission priorities generated by TIPS to those assigned by C2 analysts. To evaluate the utility of TIPS, we developed a scenario for a fictional military operation (code named Operation Fly Swatter). We describe this scenario using a sequence of thirteen (13) reports; these which detail the time-phased evolution of events in the scenario. A baseline C2IEDM database was created to reflect the order of battle at the start of the scenario; this instance primarily consisted of the participating units (both friendly and hostile), including their characteristics and associations. Based on the content of each report, we updated the database instance to reflect the impact on the tactical situation.

This process yielded thirteen database instances for evaluation—one for each report. The reports were evaluated by a set of ten (10) analysts, each with varying levels of experience. For each report, a transmission priority, from 0 (lowest) to 100 (highest) is assigned for each friendly unit in the scenario. The priorities are based on the perceived urgency that unit should receive a particular report. The analysts were instructed to assess the priority based on the scenario state (on message receipt) and the content of the message itself.

The differences in report priority ratings between TIPS and the human analysts are summarized in Table 2. The differences were the result of several different types of comparisons:

- Compared to the intended priorities of the scenario designer
- Compared to the findings of all ten (10) participating analysts
- Compared to the findings of the expert analysts (those with five years or more of military experience); there were four (4) in this category
- Compared to the consensus estimates of all analysts, as well, the expert analysts.



Table 2: Comparison of assigned message priorities (TIPS vs. human analysts).

Comparison Type	Statistic			
	Raw Assignments		Normalized Assignments	
	( $\mu$ )	( $\sigma$ )	( $\mu$ )	( $\sigma$ )
TIPS vs. Intended	14.03	9.41	4.64	6.44
TIPS vs. Analysts-All	11.81	8.05	2.91	4.75
TIPS vs. Analysts-Expert	19.97	15.30	9.50	13.54
TIPS vs. Consensus-All	11.06	6.70	1.98	3.38
TIPS vs. Consensus-Expert	14.90	9.86	5.16	7.09

The last comparison category was necessitated by the high standard deviation of analyst assigned priorities. For this experiment, those assignments that had a standard deviation of twenty (20) or less for a given response were treated as “consensus” answers. In the expert group, consensus existed on 52% of the assignments; overall, there was consensus on 18% of the assignments. For each type of comparison, we computed the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) statistics for using both raw and normalized priority differences. The raw differences were based on the actual priority percentages assigned. We also normalized to mitigate relatively small differences in message priorities. For the normalized differences, assigned priorities were considered the same if they fell into a certain range window (size =  $\pm 12.5$ ); differences greater than this had 12.5 subtracted from their absolute value.

These results show that TIPS did a credible job in assigning priorities to the various messages. While the raw comparison statistics may look high, it is important to note that there was a substantial amount of difference of opinion on priorities between the analysts; this is indicated by the relatively high standard deviation statistic. On those messages where consensus did exist, the differences with TIPS were substantially lower. When the results were normalized for smaller difference in priorities, the differences were under 5 in every category but expert. In short, on this relatively small experiment, the performance of TIPS is comparable to that of a human analyst. It is important to note that we do not advocate that TIPS replace humans—rather, we believe that it can be a decision aid for complex, rapidly changing situations.

## 4 Related work

The TIPS system encompasses two streams of active research: link mining and Bayesian classification. Link mining refers to data mining techniques that explicitly consider relationships between data objects when building predictive or descriptive models of the linked data [5]. Traditional data mining techniques, such as association rule induction or cluster analysis, attempt to mine data from a collection of independent instances of the single relation. In contrast, link mining techniques analyze the links between heterogeneous data sets (found in relational databases or on the World Wide Web) to draw inferences about the data. Perhaps one of the most widely known examples of link mining is the Page

Rank algorithm [1] to prioritize the applicability of web pages to a given search criteria.

There are several sub-areas of link mining research that have direct applicability to TIPS. These include the development of metrics to assess the similarity of two objects based on links to intermediate objects [6]. Another area is link prediction, which assesses the likelihood of a link existing between two entities, based on object attributes and other observed links [5]. One potential barrier to applying these techniques is when too few links are available to make a prediction. If the prior probability of a link is quite small, this causes a conservative bias in link prediction.

While domain independent linkage mining algorithms are inherently appealing, they rely almost entirely on the existing database structure. This becomes a problem when the database is too sparsely populated to make reliable statistical inferences. Other challenges to the problem of linkage prediction include the complexity of the database schema and the distance (from a schema perspective) between the database objects undergoing the similarity test. Thus as the schema complexity grows, an exponential increase in the number of training instances is required; this phenomenon is a variation of Bellman's "curse of dimensionality" [1]. In contrast, TIPS makes explicit use of domain knowledge to identify and weight the importance of intermediate linkages in determining the strength of the relationship between database objects.

Within TIPS, BNs are the central mechanism for characterizing the relevance of a message to a given unit. One application that is similar is the use of BNs to evaluate if a given email is spam or not. One such an application is the SpamBayes [10]. SpamBayes is conceptually similar to TIPS problems, in that some messages can be thought of as legitimate while others are not (spam). These applications utilize email attributes, including sender, subject, and keywords within the message as the basis for making the legit/spam determination. SpamBayes must be trained to be effective. Further, the training instances must be balanced for each in that for a given category (such as Viagra), there must be equal numbers of legitimate vs. spam emails for training purposes. To retain effectiveness, this balance must be maintained as the number of categories increases. Per the "curse of dimensionality", the BN will require exponentially more training data as the number of categories grows. Similar Bayesian approaches have been developed to detect spam in short text messages [7] and web pages [12].

In contrast to these approaches, TIPS employs a predefined BN which uses the presence or absence of linkages to the RDI as features. Further, TIPS BN is built with the aid of a domain expert, as opposed to learning it from the data. Instead of directly analyzing the content of the report, TIPS assumes that the appropriate linkages have already been created. While this is a reasonable assumption, it is also expected that not all linkages will exist when the RDI is inserted in the database. As noted earlier, the number of linkages will increase over time. Thus, it is unlikely that TIPS will immediately identify a given report as relevant to the UOI.



## 5 Summary

In this paper, we have introduced our work on the TIPS to extract context from C2 databases for the purpose of report prioritization. One of the most powerful advantages of the TIPS approach is the potential to keep military units properly informed with minimal human intervention. Consider the effort required for a human C2 analyst to quickly prioritize the importance of every report for each unit. The difficulty of this task is compounded by the dynamic nature of warfare—facts are continually being added or changed. This means that a report that seemed innocuous might in fact be important due to subtle (second or third order) relationships not readily apparent to an overworked analyst. Or that a less-than-fresh report might become more important to a unit over time, as additional facts become known. In short, situational awareness is critical—maintaining it is key to both defeating the adversary and avoiding friendly fire tragedies.

Overall, we were pleased that the priorities generated by TIPS approximated the priorities that human analysts assigned to each message. The demonstration in this regard highlights the potential of TIPS to streamline human-in-the-loop message processing and improve the timely flow of relevant information to the Warfighter. To date, however, the testing of TIPS has been on a synthetic database for the C2IEDM. This is primarily because the C2IEDM is an emerging standard and there are few examples of operationally compliant databases that use this mode. A key aspect of our planned (future) work with Air Force Research Laboratory on TIPS is to transition the technology from C2IEDM to a more mature database schema, such as the Air Operations Data Base (AODB). In addition, we will also evaluate the scalability of TIPS and develop approaches to minimize the number and frequency of RDI/UIO comparisons as the C2 database changes.

## References

- [1] Bellman, R. E. and Dreyfus, S.E. *“Applied Dynamic Programming”*. Princeton University Press, Princeton, NJ, 1962.
- [2] Brin, Sergey and Page, Lawrence. *“The anatomy of a large-scale hypertextual Web search engine”*. In Proceedings of the seventh international conference on World Wide Web 7: 107–117. (1998)
- [3] Charniak E. *“Bayesian Networks Without Tears”*, AI Magazine, pp. 50-63, 1991.
- [4] Culotta, A., McCallum, A., Betz, J. *“Integrating Probabilistic Extraction Models and Data Mining to Discover Relations and Patterns in Text”*. Proceedings of the Human Language Technology Conference of the NAACL (Jun 2006), pages 296–303. June 2006.
- [5] Getoor, Lisa. *“Link mining: a new data mining challenge”*. SIGKDD Explor. Newsletter. 5, 1, 84–89. (July 2003)
- [6] Jeh, Glen and Widom, Jennifer. *“SimRank: A measure of structural-context similarity”*. In Proceedings of the Eighth ACM SIGKDD International



- Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada. (July 2002)
- [7] Gómez Hidalgo, J. M., Bringas, G. C., Sández, E. P., and García, F. C. "*Content based SMS spam filtering*". In Proceedings of the 2006 ACM Symposium on Document Engineering (Amsterdam, The Netherlands, October 10 - 13, 2006). DocEng '06. ACM, New York, NY, 107-114. (October 2006)
  - [8] Marmelstein, Robert E. "*Extracting Context from the Command and Control Information Exchange Data Model*". Air Force Summer Faculty Fellowship Program (AFSFFP) – Final Report submitted to AFRL/RISB. (July 2006)
  - [9] Marmelstein, Robert E. "*Tactical Information Prioritization System*". Final Report submitted to The Research Foundation of State University of New York Institute of Technology (SUNY-IT), Utica, NY on behalf of AFRL/RISB. (Dec 2006)
  - [10] Meyer, T and Whateley, B. "*SpamBayes: Effective open-source, bayesian based, email classification system*". In Proceedings of the First Conference on Email and Anti-Spam (CEAS). (2004)
  - [11] Multilateral Interoperability Programme (MIP) website, [www.mip-site.org](http://www.mip-site.org).
  - [12] Ntoulas, A, Najork, M., Manasse, M., Fetterly, D. "*Detecting Spam Web Pages through Content Analysis*". In Proceedings of the World Wide Web (WWW) Conference, Edinburgh, Scotland. (2006)