exemplar

Data Fusion &
The Synthetic Meaning Project

A computational approach to detecting suspicious communications in very large input streams

August 2003

PROPRIETARY

exemplar

Variety of Data Types

- Sensor Data
  - *Numeric Measurements*
  - *Visual (Video)*
  - *GIS*
- Non-sensor Data
  - *Communications*
    - Cell Phone
    - E-Mail
  - *Contextual Data*
    - Who, what, where, when, how
    - Affiliations
    - Associates

Fusion

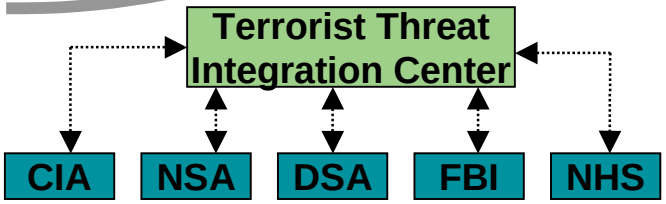
- Not all types of data can be “fused” creating a “multiverse” of systems
- The best machine for multi-viewpoint reasoning is the human brain
- Result:
  - *Fuse what you can – similar data types*
  - *Present the various “universes” as multiple viewpoints of the same phenomenon*
  - *Flag items based on interesting characteristics*
  - *Trigger other events (e.g. pull related data for the analyst to synthesize and assess)*
  - *Assign weights and measures to the various inputs as “levels of confidence”*

The Problem

ALERT!

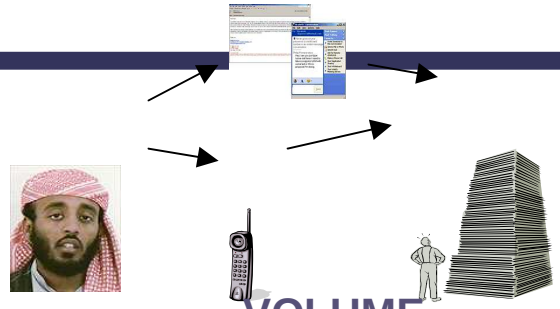
THE ANALYST'S PROBLEM

Timely,
- Discovery and interpretation of suspicious communications
- Threat identification
- Threat notification



INFORMATION SILOS

- Inability to efficiently and securely access, correlate, share and commonly assess information among agencies



VOLUME

- Digital communications are increasing exponentially
- *"The enemy will have to monitor all communications"*

…Ramzi bin al Shibh, 2002

LEGAL

- Privacy policies restrict access to useful information

"Buy Russian Apples"
…Jenny

"I have the apples"

LINGUISTIC AMBIGUITY

- Keyword avoidance, and frequent content changes mask threats

CURRENT NLP APPROACHES

- Reactive versus proactive
- Not adaptable to changing conditions until after the fact

A Current Approach To Threat Assessment

FIRST

- Identify Suspicious Communications
  - *Discriminating suspicious communications from a large corpus of textual inputs containing a prevalence of linguistic ambiguity requires:*
    - Matching lexical and grammatical information with *a priori* knowledge from one or more classes of information:
      - Social – Recognition that the participant(s) are themselves suspicious
      - Semantic - Suspicious keywords or concepts are detected
      - Temporal – Timing, flurry or lack of communications
      - Contextual – The circumstances or situation associated with the text
      - Spatial – Identifying that a communication's origination/destination is suspicious

THEN

- Analyze and Interpret the Communication's Content
  - *Human interpretation produces the best quality and consistency*
  - *Apply a variety of tools and operations specific to immediate needs of the analyst to help determine the meaning; analyst is integration point*
  - *Assess the probability, scale, and urgency of the threat*

The process is *reactive*, requires *a priori lexical and grammatical knowledge,* and, therefore, subject to a high rate of failure when confronted with frequent content changes and linguistic ambiguity

Synthetic Meaning Approach to Threat Assessment

FIRST

- Identify Suspicious Communications
  - New § *Compute the lexical, social, temporal, spatial and entity/relation information, as well as the semiotic attributes of a communication*
  - New § *Computationally create the corresponding semantic and semiotic network knowledge representation*
  - New § *Computationally* 'fingerprint' *the resultant semantic network pattern and classify it through best fit analysis with network categories*
  - New § *Create prioritized lists of probable suspicious communications*

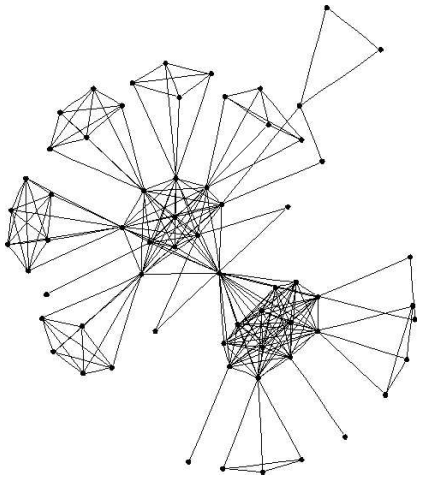
THEN

- Analyze and Interpret the Communication's Content
  - New § *Human interpretation produces the best quality and consistency*
  - § *The feature rich Analyst Workspace is the integration point and provides a full set of image processing tools for performing semantic, social, temporal, spatial and semiotic analysis, visualization, and interpretation*
  - § *Assess the probability, scale, and urgency of the threat*

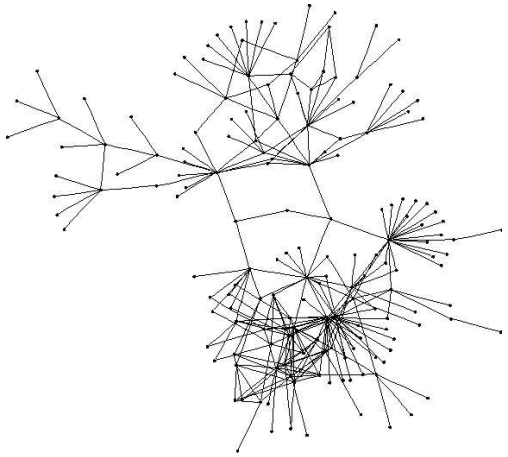
The process is *proactive*, dependent on the pattern that emerges in the form of a semiotically tagged semantic network knowledge representation of a communication; creation of a pattern of communication is *not* adversely effected by content changes and linguistic ambiguity

Prelude

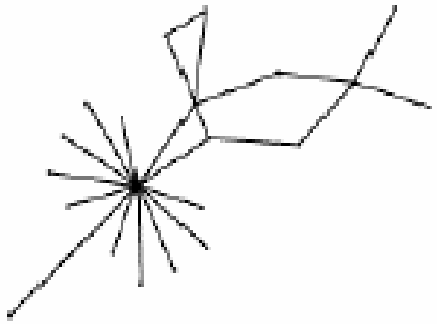
- A semantic network's ***Wireframe*** representation is a basic "***fingerprint***" of that semantic network's relational and conceptual properties



Family Tree



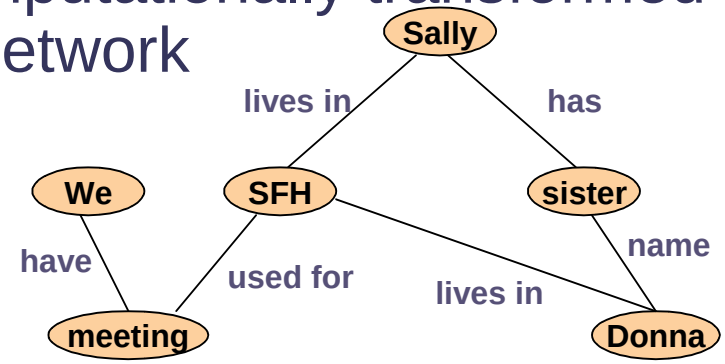
Protein



Conversation

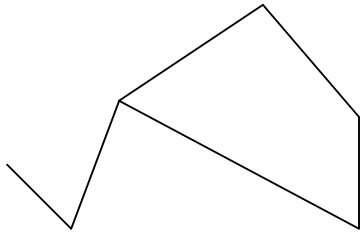
Prelude

- The sentence "*Sally lives in a single family home that we can use for meetings. Her sister, Donna, lives there also.*" can be computationally transformed into its representative semantic network



Graphic Frame View

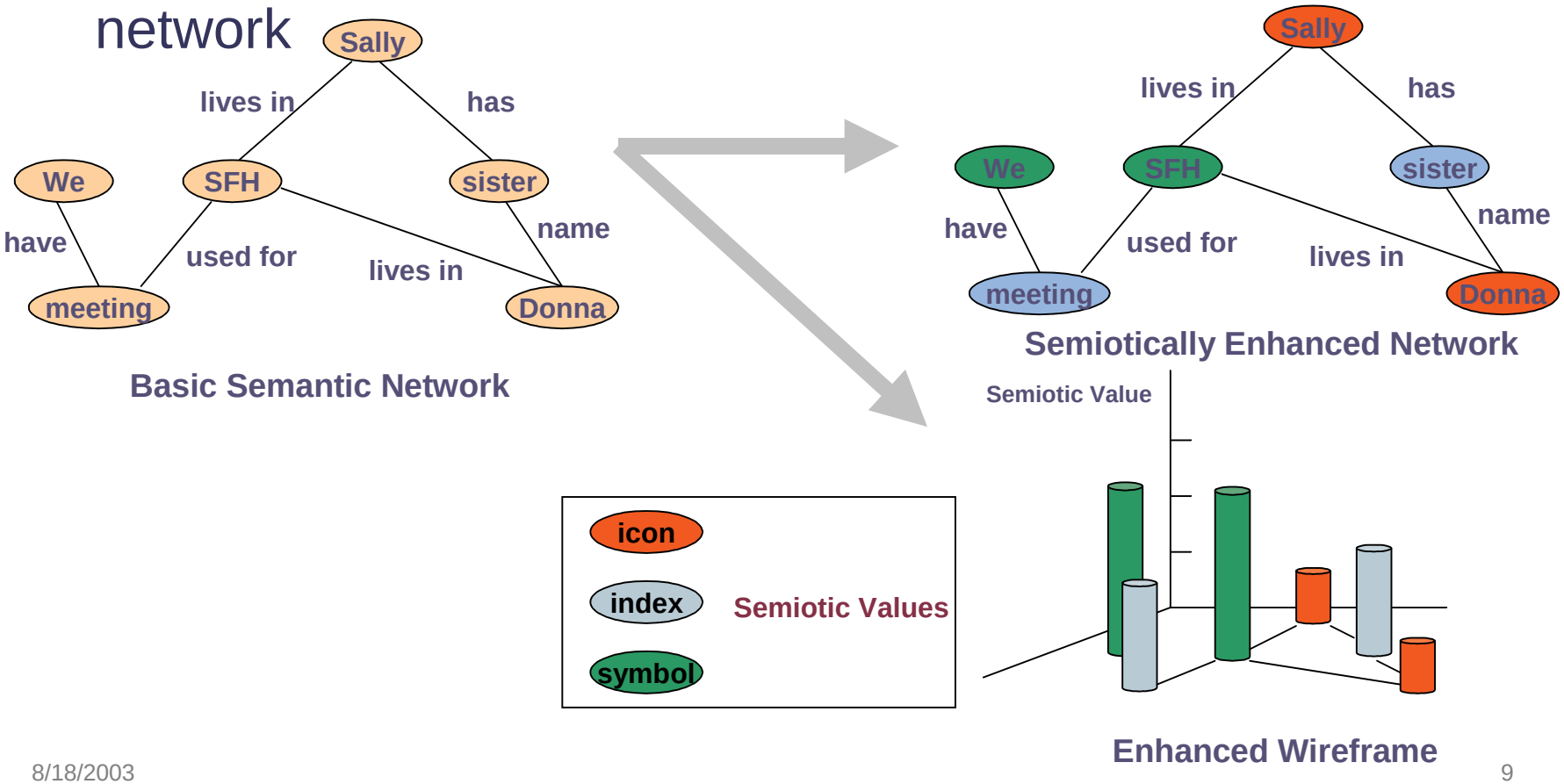
- The *Graphic Frame* view can then be transformed into its corresponding *Wireframe* visual equivalent



Wireframe View

Prelude

- *Semiotically* enhanced semantic networks add dimensionality and another relational identity to further aid in discriminating the unique "*fingerprint*" of a semantic network



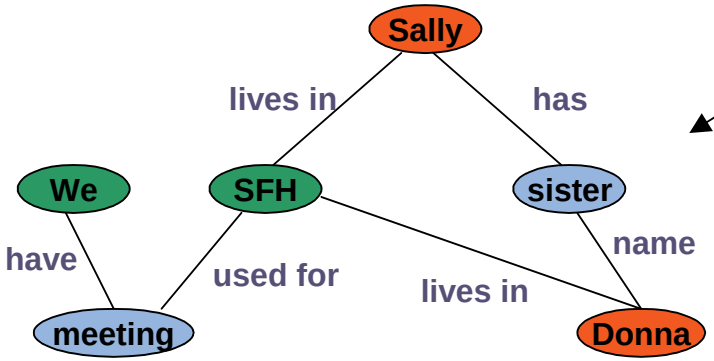
Prelude

- Computing the ***Semiotically Enhanced Network***

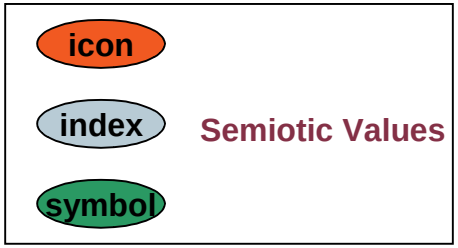
Sally lives in a Single Family Home that we can use for meetings. Her sister, Donna, lives there also.

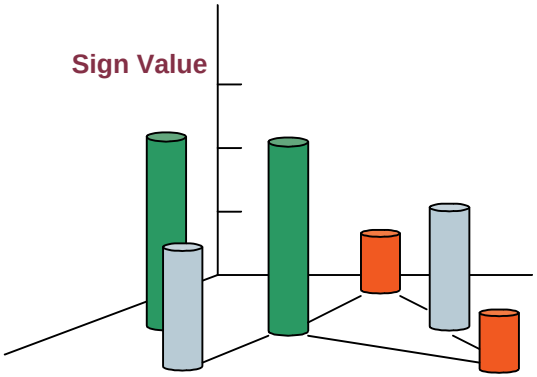
→ **Semiotic Interpretant Algorithm e.g. Sarbo / Farkas** →

SIGN CLASS	CONCEPT 1	RELATION	CONCEPT 2	SIGN CLASS
icon	Sally	lives in	SFH	symbol
symbol	SFH	used for	meeting	index
symbol	We	have	meeting	index
icon	Sally	has	Sister	index
index	Sister	name	Donna	icon
icon	Donna	lives in	SFH	symbol



Graphic Frame View





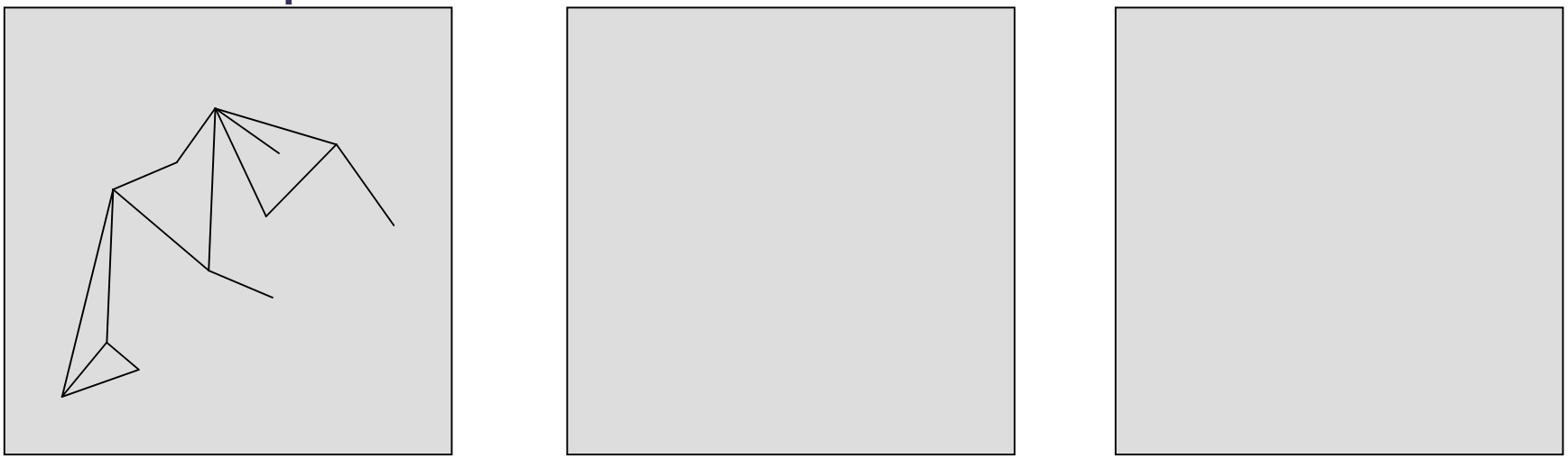
3D Sign Value View

Prelude

- Together, the ***Basic Semantic Network***, and the ***Enhanced Semantic Network*** give a dimensionality and visual structure to communications that
  - § *provide additional discriminatory features far exceeding those of current systems*
  - § *provide graphical analytic features at both the semantic and semiotic layers*
  - § *enables analysis at multiple levels of abstraction*
  - § *facilitates combining subnet structures into larger, more meaningful constellations of networks*
  - § *is topologically different than a lexical or grammatical construct*
  - § *identifies and categorizes* "units of meaning" *of the communication*

A Preliminary Example

- “Begin with the end in mind” – Stephen Covey
- Which semantic network pattern demonstrates the least amount of connectivity among its concepts?

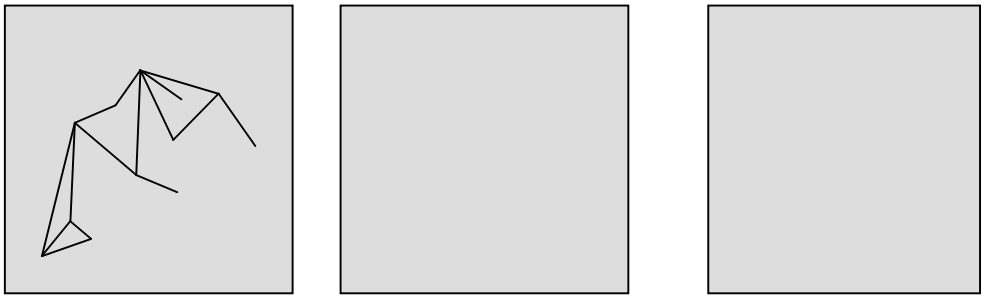


A **B** **C**

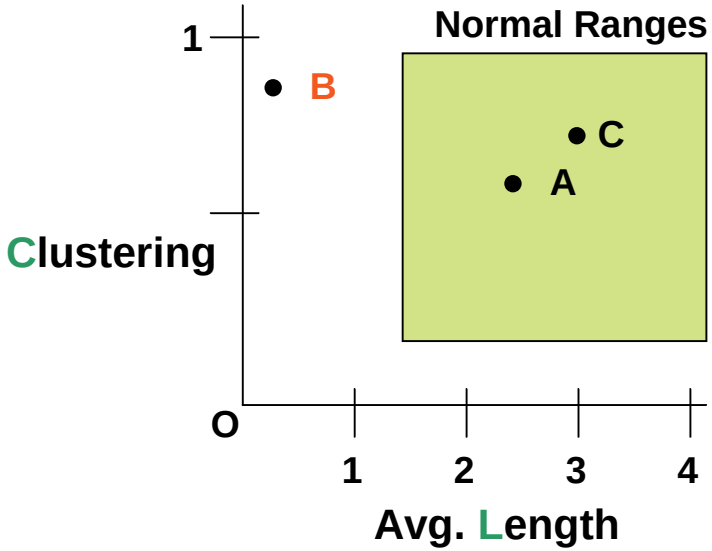
Semantic Network Representations of Three Textual Communications

A Preliminary Example

- Computing a simple structural fingerprint from the semantic networks



A	B	C
L = 2.31 C = 0.64	L = 0.25 C = 0.89	L = 2.91 C = 0.77



ATTRIBUTE DEFINITIONS

L – Average node-to-node distance in a network

C – Clustering coefficient; the tendency of nodes within a network to form highly connected clusters

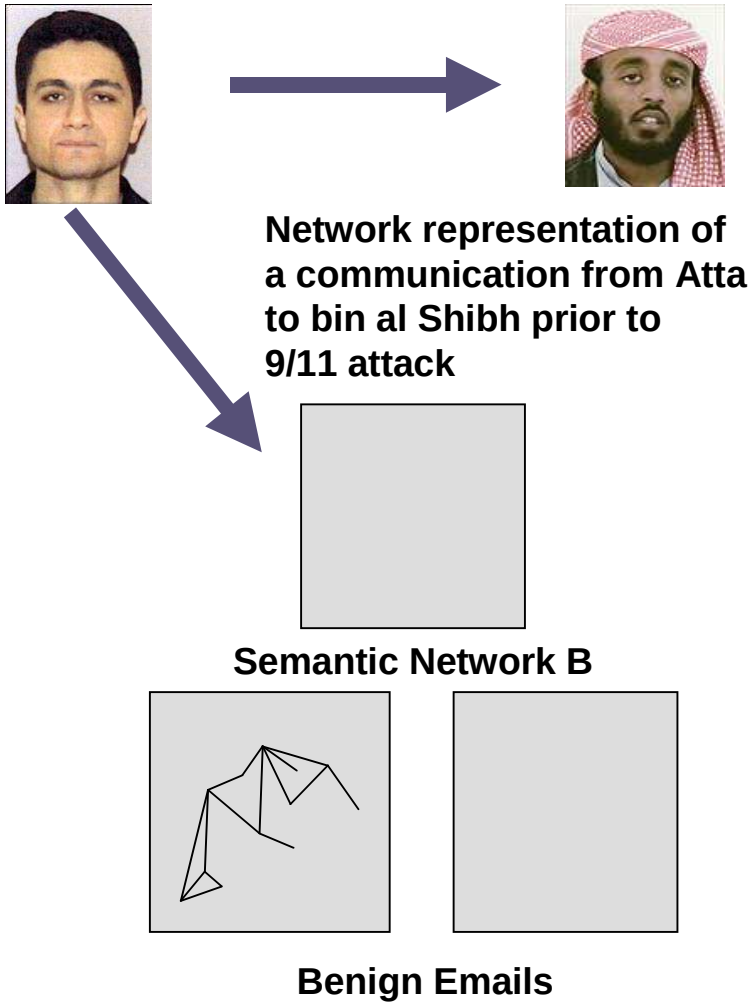
INTERPRETATION

'B" is composed of a fragmented set of highly connected clusters; i.e., very fragmented communication. It falls outside of 'normal' ranges and is therefore suspicious.

A Preliminary Example

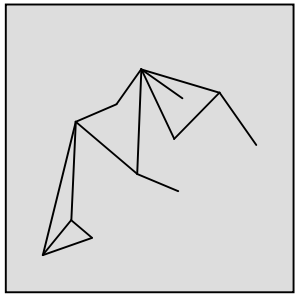
- By picking Semantic Network B
  - §*An abnormal, or potentially suspicious network was identified as requiring further analysis*
  - §*The analyst workload was reduced and focused*

This represents the essence of the proposed Synthetic Meaning Project

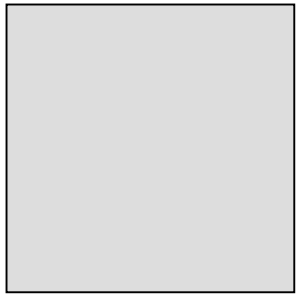


A Preliminary Example

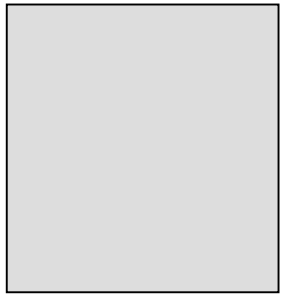
- **The communications and their basic semantic networks**
- **Note - *there are not any suspicious key words***
- ***This project does not assume that it would always be this easy***



Larry,
Thanks for all your help. I've heard from Bruce and emailed him my resume which he's going to pass along to Infovision. Thanks for the great reference. If you could mail it to me that would be wonderful. I will talk to you later. Tell Dave not to work you too hard!
Pamela



Jenny,
The first semester starts in three weeks. Nothing has changed. Everything is fine. There are good signs and encouraging ideas. Two high schools and two universities. Everything is going according to plan. This summer will surely be hot. I would like to talk to you about a few details. Nineteen certificates for private study and four exams.
Regards to the professor. Goodbye.



Larry,
Mom is feeling better. We went to lunch at Pala. They have a great buffet. Of course we ate too much. And, we gambled too much. We lost 25 dollars but had a good time. I'll have mom call you tonight.
Dad

Synthetic Meaning Project

- Primary Project Assertions
  - § *Meaningful, distinct patterns can be detected in communications*
  - § *These patterns are revealed through combinations of semantic, social, temporal, lexical, contextual spatial, and semiotic information*
  - § *Patterns exist even in the face of linguistic ambiguity*
  - § *Patterns can be grouped into classes or categories*
  - § *Patterns can be used as the basis for effective classification of communications and efficient analysis*

SensorNet Applicability

- Example
  - *Major event (Super Bowl, Indy 500, etc.)*
  - *Recent communications have been identified as suspicious*
  - *Data related to communications' participants is flagged for priority comparison/correlation with incoming live data (e.g. surveillance photos at stadium)*
- Example
  - *Suspicious communications are traced to a geographic location*
  - *Sensors in the area have their polling increased*
  - *Number of mobile sensors are increased*
  - *Density of a certain type of sensor is increased*
- Use Semantic Networks and/or other tools (associative memory engine) to show relationships between disparate data types and concepts, helping an analyst with "multiverse" analysis
- May feed into intelligent agents/grid to initiate collection of data from additional systems
- May change storage/bandwidth concepts due to analyst need to create, manipulate and save several instances of the data and their related models

Conclusion

Thank You

Exemplar Solutions, LLC

P.O. Box 270897

San Diego, CA 92198

philip.porreca@exemplarllc.com

(858) 312-1126