CONSMAG: Knowledge Representation Based Upon Consensus for Multiple Agents

Donald W. Dearholt, Senior Member, IEEE, Department of Computer Science, New Mexico State University, Las Cruces, NM 88003, 505-646-6228, dearholt@cs.nmsu.edu

Abstract *A multi-agent system designed to observe and remember patterns of co-occurrences in an open, dynamic system is described. The agents are organized according to function, with the low-level agents storing observed co-occurrences of key concepts or events in memory until being recycled for further service, thus simulating fixed short-term memory. The mid-level agents retain a short history (the length of this history is determined by the number of low-level agents available) of co-occurrence phenomena. The high-level executive agent computes three different, related associative networks as a graphical representation of the co-occurrences recorded by means of cumulative consensus of the mid-level agents and responds to queries. Thus the CONSMAG system effects a transformation from episodic memory to a class of associative networks representative of semantic memory. This learning paradigm supports cost estimates for learning any given associative network, including modification of one network into another, to simulate a novice becoming an expert.*

Keywords multi-agent systems, dynamic systems, adaptive systems, co-occurrence, clustering, discrete models, Pathfinder networks, consensus, data fusion, learning, costs of learning

1. Introduction and Overview of the Pathfinder Paradigm

Modeling dynamic co-occurrence phenomena by means of networks and clusters is important for several reasons. There are situations in which humans learn about the classification of entities based upon clustering or co-occurrence; and in machine learning, there are circumstances in which it is desirable to modify a system to account for a new set of events. The data for this sort of learning can be either ordered sets or unordered sets of clusters, and the resulting graphs are either directed or undirected, respectively. The initial challenge is to derive a meaningful dissimilarity matrix, so that matters of consensus and edge membership can be properly treated sequentially, as the sets of co-occurring entities appear. A second challenge is to suggest a set of agents capable of supporting the necessary computations, and establishing an organization for those agents which facilitates the needed communications, computations, and expected queries.

An important aspect of the model proposed has to do with the tasks and limitations of the agents. It will be assumed that there are three types of agents, with constraints on the numbers and capabilities of each; these constraints effectively provide a constraint on the total memory available for the system, thus limiting the bandwidth available for monitoring the dynamic system. Together, these agents provide the computational power required to count and remember co-occurrences, to generate three types of networks (unconstrained, Pathfinder, and single-link hierarchical networks) iteratively, and to respond to queries.

The Pathfinder paradigm was created to model aspects of human semantic memory [7]. Among the properties of Pathfinder networks are the clustering of similar concepts and the preservation of minimum-cost paths [6, 7]. The original Pathfinder networks are also always connected. The properties of the Pathfinder paradigm are extended to support the generation of a co-occurrence Pathfinder network (CPFN) which is not necessarily connected; such networks are constructed sequentially as an associative model of a dynamic phenomenon at successive time steps. This is an extension of the previous result that Pathfinder networks provide all information needed to construct a minimum method (single-link) hierarchical clustering [7, 8]. Because Pathfinder networks contain more information than the single-link hierarchical clustering scheme [7], and are not necessarily hierarchical, they can model less constrained phenomena. In the Pathfinder paradigm, edges violating a generalized triangle inequality are pruned, assuring that remaining edges are portions of shortest paths for the metric used.

The objective of this paper is to present a new way of viewing some dynamic phenomena, utilizing, in part, an extension of Pathfinder networks. This paradigm supports modeling open systems (in which the set of entities is not normally static), and provides a way of expressing and representing information on clustering of the entities of interest. Clustering techniques are usually applied to static phenomena, but learning paradigms require that constraint to be relaxed, so that dynamic phenomena can be modeled. The Council of Agents suggested and described herein can be viewed either as a computer-based system of multiple agents, or as a candidate model for human memory capable of facilitating the transition from episodic memory to semantic (associational) memory, in the spirit of Minsky's *Society of Mind* [10]. An earlier version of this paradigm for modeling co-occurrence phenomena, adapted for an evolutionary approach, was given in [5].

Throughout this paper, it will be assumed that the co-occurrences of unordered entities are counted, leading to undirected networks. The sequential application of co-occurrence data to a graph-generation procedure results in the modification of an existing network as a new set of co-occurrence terms is observed and taken into account. The resulting framework supports viewing the transformation from a sequential flow of concepts to an associative representation for memory in the form of one of three types of associative networks by means of multiple agents, and the resulting networks can be queried. Furthermore, the paradigm can also be viewed as a means of accomplishing data fusion from the various agents.

2. Properties of Co-occurrence and its Representation

It will be assumed that in the given domain, entities co-occur within an appropriate unit of time or space, or other unit of interest, such as some syntactic unit (e.g., a sentence or clause). Basic definitions follow.

Definition 1: The *entities* are the events or objects used to characterize the dynamic system, and they are denoted A, B, C, … For example, the entities can be key concepts, or events of interest.

Definition 2: A *co-occurrence unit* (c-unit) is a unit of space, time, or syntax (such as a phrase, clause, sentence, or abstract) within which entities are sampled and are considered to co-occur.

Units of time or syntax occur in a sequential order, and in each c-unit, a particular set of entities will be found to co-occur. To model such a phenomenon, it is necessary to be able to represent the collection of co-occurring entities over the set of c-units defining the object of interest (such as an abstract or other text). Formal languages may be viewed as describing a very structured form of co-occurrence of entities, in which order and sometimes the context of entities is important. This paper addresses those problems for which co-occurrence represents a less constrained phenomena, so that the order of the co-occurring entities within a c-unit is not important.

It is assumed that the sampling of the dynamic phenomenon occurs in discrete steps, which could be either of time or of syntactic units, and these steps will be denoted by t_i. Figure 1 illustrates the concepts outlined in this section, and attention will be directed primarily to the functions of the agents: the collection of co-occurrence data, the clustering procedure, the generation and representation of the clusters by means of any of three associative network representations, and responses to likely queries. Some additional definitions and notation follow.

Definition 3: A *co-occurrence term* (c-term) is a set of entities which co-occur within a c-unit, and thus warrants identification at a higher level of abstraction. A c-term is denoted by c_i. There may be more than one c-term for a c-unit; for example, suppose that the domain is sentences of text, in which each sentence corresponds to a c-unit. If clauses correspond to c-terms, then there can be multiple c-terms for each c-unit.

For some phenomena, it is possible to have multiple occurrences of an entity within a c-term. There are two choices in modeling co-occurrence in this situation: (1) keep track of the number of times each entity occurs within each c-term, or (2) consider that the appearance of an entity within a c-term is all that is important, so that multiple occurrences are considered to be equivalent to a single occurrence. In this paper, it will be assumed that it is adequate to consider that entities either co-occur or do not co-occur within a given c-term, so that repetitions within a c-term won't be counted.

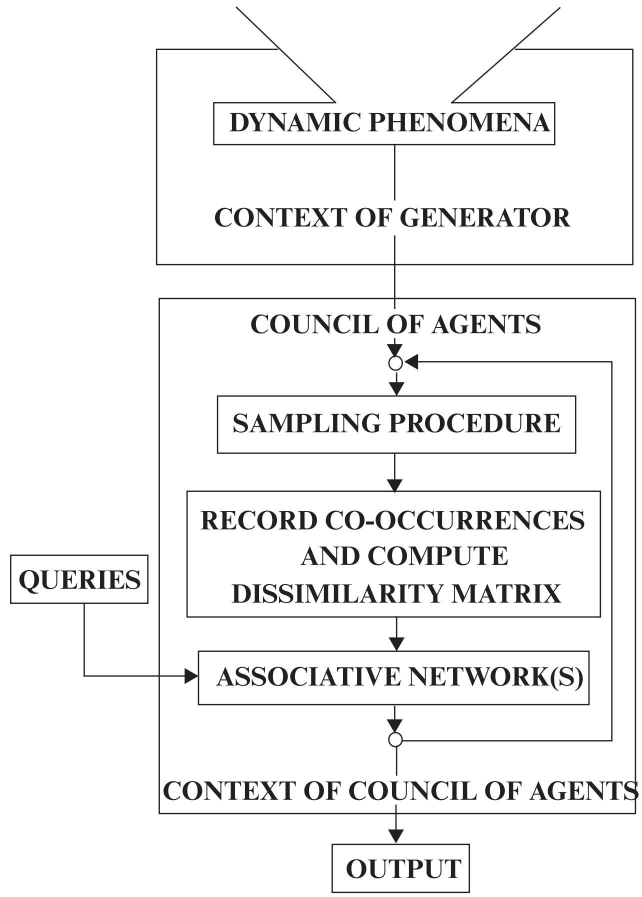


Figure 1 - The Dynamic System and Functions of Agents

3. The Council of Agents and Representation of Co-occurrences

The Council of Agents proposed is a heterogeneous set of agents, organized as a hierarchy having three levels, with the agents on each level having differing roles.

Definition 4: A *proxy agent* is able to observe co-occurrences of entities within one c-unit, and provides storage for a k-tuple $(t_i, c_p, c_q, \ldots)$, which contains a time stamp t_i and at least one c-term c_p representing co-occurrences within the given c-unit. We will assume that there are n proxy agents in the Council of Agents, and when all proxy agents have been utilized in remembering the observed co-occurrences, the one with the oldest memories (earliest time stamp) is recycled for reuse.

Definition 5: A *state agent* is a set of the stored observations of the set of proxy agents at a particular time, in which the most recent c-unit and corresponding c-term(s), and the next most recent, and the next most recent, until the memory width is exhausted (up to n), are considered a vector of adjacent observations of co-occurrences within the memory of the proxy agents.

Definition 6: *Mimer* is the high-level executive agent responsible for collecting and summing all c-terms from all state agents, for computing whatever networks are needed, and responding to queries, on behalf of the Council of Agents; he also performs interface functions. Mimer was a Norse god who was said to guard the beautiful fountain of knowledge; anyone who could get past him to drink from the fountain was said to have all knowledge available. It is likely that the functions performed by Mimer could be distributed among another set of agents, but it's not clear if that's advantageous or not.

Associative networks can be generated for each state of memory (state agent), and for Mimer, to provide different perspectives on the status of the model at any time. Indeed, Mimer can be viewed as computing the consensus of the state agents. The perspective offered by this paradigm is to view the memory of a proxy agent as being capable of recognizing and remembering entities co-occurring at a given time, and of a state agent being able to recall recent co-occurrences. The definitions which follow are intended to formalize these notions.

Definition 7: The memory of the Council of Agents and its subagents (the proxies, states, and Mimer) can be represented by a *scenario* at a particular time. A scenario is (initially, upon beginning anew in some domain) an n ¥ n triangular matrix of c-units, in which the contents of each cell of the matrix correspond to the c-term(s) of the corresponding c-unit. Each row of a scenario represents the c-units from the present toward the past, up to the maximum width n for the maximum number of proxy agents available. That is, the bottom row in the scenario is the most recent state of memory, and the top row is the oldest available state of memory. The leftmost column is the set of c-units observed at each time step, and time increases in descending the column. As soon as each proxy agent has been utilized, then the recycling of the proxy with the earliest memory makes it available for observing and recording the next c-term(s), and the scenario becomes rectangular from then onward.

An example will illustrate a scenario and ways of viewing the proxy and state agents. This example assumes that the sampling of the dynamic phenomena begins at time t_1, and continues thereafter; the snapshot shown is taken immediately after t_7. Please note that the Council of Agents does not have all the information shown available to it, since we will assume that there are only five proxy agents (and therefore five state agents), so that any five adjacent rows, representing five state agents, can be polled at the most recent time by Mimer, if it is desirable to do so.

Time	Memory				
t_1	(A, B, C)				
t_2	(D, E)	(A, B, C)			
t_3	(D, E, F)	(D, E)	(A, B, C)		
t_4	(B, D)	(D, E, F)	(D, E)	(A, B, C)	
t_5	(C, D)	(B, D)	(D, E, F)	(D, E)	(A, B, C)
t_6	(A, B, E)	(C, D)	(B, D)	(D, E, F)	(D, E)
t_7	(A, C, D)	(A, B, E)	(C, D)	(B, D)	(D, E, F)

Figure 2 - The Scenario After Seven c-units

The seven states of memory shown in the scenario in Figure 2 appear as the rows of a mostly triangular matrix. Assuming that the memory limitations are five proxy agents, then, if Mimer polls the state agents at time t_5, the state agents represented by rows labeled t_1 through t_5 are polled. The co-occurrence expression showing the numbers of times key concepts co-occur is:

$$C_5 = \{(A, B, C)5, (D, E)4, (D, E, F)3, (B, D)2, (C, D)\}$$

This co-occurrence expression is unconstrained by either hierarchical or generalized triangle inequality constraints. Counting co-occurrences in C_5, the co-occurrence matrix S_5 is obtained; the dissimilarity matrix D_5 is obtained by subtracting each entry in S_5 from 8 (the largest number of co-occurrences plus one):

$$S_5: \begin{bmatrix} - & 5 & 5 & 0 & 0 & 0 \\ 5 & - & 5 & 2 & 0 & 0 \\ 5 & 5 & - & 1 & 0 & 0 \\ 0 & 2 & 1 & - & 7 & 3 \\ 0 & 0 & 0 & 7 & - & 3 \\ 0 & 0 & 0 & 3 & 3 & - \end{bmatrix} \qquad D_5: \begin{bmatrix} - & 3 & 3 & 8 & 8 & 8 \\ 3 & - & 3 & 6 & 8 & 8 \\ 3 & 3 & - & 7 & 8 & 8 \\ 8 & 6 & 7 & - & 1 & 5 \\ 8 & 8 & 8 & 1 & - & 5 \\ 8 & 8 & 8 & 5 & 5 & - \end{bmatrix}$$

If Mimer polls state agents (Figure 2) at time t_7, then the state agents represented by rows t_3 through t_7 are polled. This requires the state agents to have additional memory beyond that currently resident in the proxy agents, since the co-occurrences of (D, E) and (A, B, C) are no longer in the memory of any proxy agent. Thus the memory storage requirement for the Council of Agents, in terms of c-units of storage, is $n + n - 1$, in which each of n proxy agents require one c-unit of storage. In addition, state agents must be able to provide $n - 1$ c-units of storage (when the scenario becomes rectangular), so that the c-units in the scenario which are not in storage of proxy agents at the time of sampling will be included. The number of key concepts supported in each c-term, and the number of c-terms supported in each c-unit, must be decided before actual storage allocation can be known.

4. The Morphology of Discrete, Dynamic Co-occurrences

Three levels of representation for co-occurrences will be described:

Definition 8: An *Unconstrained Co-occurrence Network* (UCN) is the most general representation, in which a snapshot of the system at a given time can be used to generate a co-occurrence matrix S. A network having edges whenever two entities co-occur, with the weight of each edge determined by the number of times the pair of entities defining the edge have co-occurred, provides an associative, graphical representation. The entities become the nodes.

Definition 9: A *Co-occurrence Pathfinder Network*, denoted as CPFN(r, q), is a network in which each co-occurring pair of entities is linked with an edge unless there is some shorter path between those entities. This requires the computation of dissimilarities using the r-metric and the q-parameter [6, 7]. Since edges don't connect entities which have never co-occurred, a CPFN(r, q) is not necessarily connected.

The CPFN can be computed by inverting the S matrix to a dissimilarity matrix D, and then removing all edges which violate a generalized triangle inequality for the r-metric selected [7]. Advantages of the CPFN as an intermediate level of representation are (1) consistency with geometric reality, in eliminating violations of the generalized triangle inequality, (2) sparser representation, typically having fewer distinct values of co-occurrence or edge weights than the UCN, (3) the capability of representing hierarchical or nonhierarchical clusterings, (4) the ability to withstand some noise (it is the entities having the fewest numbers of co-occurrences which are lost in the transition from the UCN to the CPFN), and (5) the preservation of clusters and minimum-cost paths between entities.

Definition 10: A *Hierarchical Co-occurrence Network* (HCN) provides a representation of the single-linkage clustering of the objects/entities of the scenario, using co-occurrence counts as the similarity metric (to be inverted before clustering [7]). The HCN is easily computed from either the UCN or the CPFN, and is usually shown as a dendrogram (tree).

This third level is obtained by further constraining the mathematical or graphical representation by applying hierarchical constraints (see Chapter 1 of [7]). The information in either the UCN or the CPFN is sufficient to generate the single-linkage hierarchical clustering, and this can be represented by a dendrogram or a tree network. These three levels of representation are illustrated in Figure 3.

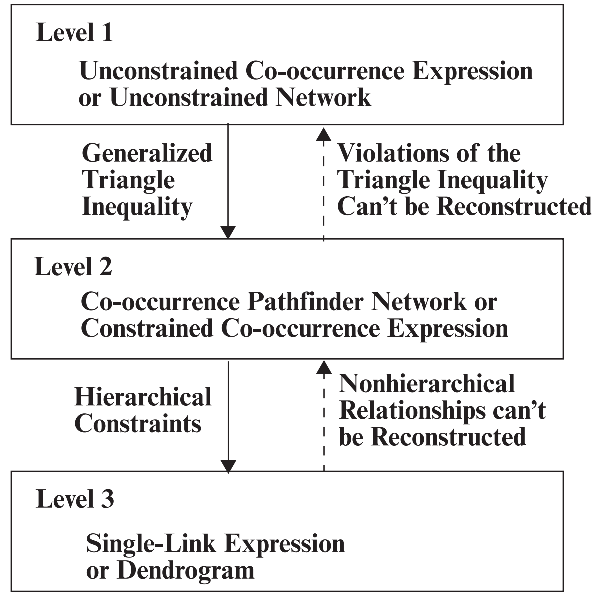
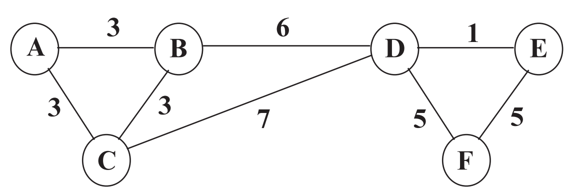


Figure 3 - Three Levels of Representation for Dynamic Phenomena

An example will illustrate a scenario and the graphical representations at three levels. For simplicity, it will be assumed that an adaptive agent begins monitoring the dynamic system at t_1, and sampling occurs at time steps t_i. The scenario after seven time steps (c-units) is shown in Figure 2, assuming there are five proxy agents. This example will be used to illustrate the three types of associative graphs that can be generated by Mimer, all based upon the dissimilarity matrix D_5, as shown above.

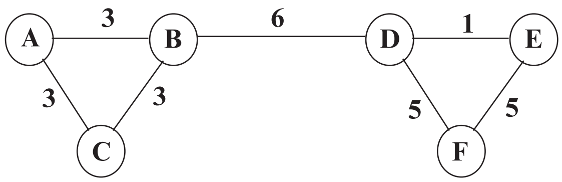
The network (unconstrained) corresponding to D_5 is shown in Figure 4.



CU = {(A, B, C)5, (D, E)4, (D, E, F)3, (B, D)2, (C, D)}

Figure 4 - The UCN and Expression Corresponding to D_5

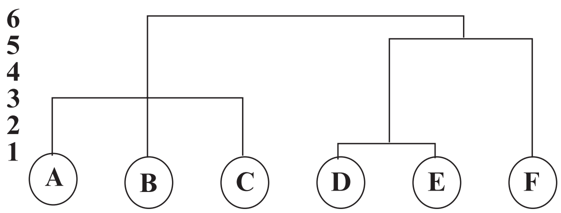
The $CPFN(r = \infty, q = 5)$ is obtained from the dissimilarity matrix D_5 by the usual generation procedure for a Pathfinder network, but requiring that entities co-occur before they can be directly linked. The CPFN is shown in Figure 5. Note that the edge between nodes C and D is missing, as it violates a (generalized) triangle inequality.



CCPFN = {(A, B, C)4, (D, E)4, (D, E, F)2, (B, D)}

Figure 5 - The CPFN and Expression for the Dissimilarity Matrix D_5

The hierarchical representation most easily available is the single-linkage representation, denoted by HCN. This is obtained by applying the single-linkage constraints to either the UCN or the CPFN [7,8]. The HCN dendrogram for this example is shown in Figure 6.



CSLHCS = {((A, B, C)3, ((D, E)4, F))}

Figure 6 - The HCN Dendrogram and Expression Corresponding to D_5

This example (Figures 4, 5, and 6) illustrates the three different levels of representation, corresponding to different assumptions and constraints. There are advantages and disadvantages in the representations and modeling power available at each level.

The associative data structures that are derived from the co-occurrence counts of the dynamic phenomenon provide the foundations for a variety of queries. While the scope of queries is a large one, it is sufficient for the purposes of this paper to note a few types of queries. First, queries could be related to the entities themselves--whether or not a particular entity is in one of the graphs, whether or not a list of entities co-occur, the list of entities in a particular cluster, or the dissimilarity of a pair of entities. Second, queries could be related to clusters--whether or not two clusters are directly linked, or the dissimilarity between two clusters. Third, queries could relate to aspects of the progression of the associative graphs as the dynamic process unfolds, related to the formation of clusters, to consensus, or other high-level aspects of the representation of the system.

Definition 11: A *canonical scenario* is a scenario in which

1. The i_{th} state of memory M_i contains i c-units, $1 \le i \le n$, where n is the number of proxy agents, and hence is the limit of memory of the scenario.

2. Each c-unit contains one or more distinct c-terms; that is, a given c-term can occur at most once within any c-unit. Furthermore, Λ, the null value, is an acceptable c-term.

3. All the proxies of the scenario are consistent; that is, the c-terms in each c-unit along any diagonal are identical.

A canonical scenario is triangular, with the number of rows and columns equal to n, the number of proxy agents. As time progresses, the limits of memory of n rows and n columns are reached. The canonical scenario is not extended beyond the triangular representation, however, as that is quite sufficient for a powerful theorem. Monitoring general co-occurrence phenomena, however, does normally result in the scenarios becoming rectangular as the proxy agents are recycled within a domain.

Theorem 1: For any network N having finite, integer-valued weights, there is a canonical scenario which will generate that network.

Proof: The proof is by construction. Let the weight values on the edges of the network N have the values $\{i_1, i_2, \ldots, i_m\}$, in increasing order, where

$$1 \le i_1 = weight_{min} \le i_2 \le \ldots \le i_m = weight_{max} = Max$$

Let the size of the canonical scenario be i_m memory states M_i, to represent sampling at times $t_1, t_2, \ldots, t_{i_m}$, in which each M_{i_j} has i_j c-units. This triangular scenario can be filled uniquely using the following procedure:

1. In the first (and only) c-unit of M_1, place the c-terms resulting from the co-occurrences having minimum weight value i_1. Thus, in the entire scenario, these co-occurrences appear in the longest diagonal, having the maximum number of c-units, and resulting in the minimum weight value in the network (after inverting the co-occurrence count to obtain the dissimilarity coefficients).

2. Proceeding in the order of increasing weight values from the network, the c-terms resulting from the edges having weights i_j are placed in the first c-unit of the memory state having cardinality $|i_j|$. Thus the total count of the requisite co-occurrences for these entries in the scenario is exactly $Max + 1 - i_j$, where Max is the value of the maximum weight. If there is some integer weight value i_p for which there are no edges labeled in the network, then enter Λ (the null value) into the first c-unit of M_{i_p} (the state of memory having cardinality $|i_p|$), and fill in the diagonals so they are consistent.

Summing the co-occurrences for each proxy (down the diagonals), the c-terms are found to co-occur the number of times required to generate the corresponding weights in the network. Furthermore, any pair of entities can co-occur at most once in each c-unit in a canonical scenario constructed in this fashion; this is consistent with the assumptions made on counting co-occurrences. *QED*

Corollary to Theorem 1: For any UCN, CPFN, or HCN having finite, integer-valued weights, there is a canonical scenario which will generate that network, since Theorem 1 holds for any network.

5. The Cumulative Consensus of Agents

Consensus is the formation of a structure to represent a set of similar structures. Mathematically, it is a way of computing a prototype for a set of objects [1]. Consensus has been applied to trees [2], sequences [4], and lattices [9]. The perspective in this paper provides a means of viewing the three associative graphical representations of the dynamic phenomena, constructed from a scenario, as a consensus of state agents in the scenario. Furthermore, this perspective invites extension to the more general case by providing a consensus operation to construct, for example, a $CPFN_{cons}$ which represents a set of CPFNs. Similar consensus operations can be used to construct UCNs and HCNs.

Definition 12: The *cumulative consensus* of a set of memory states is computed by summing the co-occurrences in n adjacent state agents, forming the resulting similarity matrix, inverting to form the dissimilarity matrix, and then computing the desired network. Such a consensus may be either *weighted* or *unweighted*; if weighted, then some c-units may be assigned a larger significance than others, and weighted accordingly. If the consensus is unweighted, then all c-units are weighted equally, and the cumulative co-occurrence counts are simply summed. In the example given earlier, the consensus was unweighted.

The UCN_{cons} is simply a graphical representation of the edges in the dissimilarity matrix, in which the weights in the matrix are also the weights assigned to the edges. No edges are pruned in this representation, so that every pair of co-occurring nodes/entities has a corresponding edge. This representation can be viewed as the consensus of the n adjacent states (rows) of the scenario which are used by Mimer.

The CPFN computed from the entire scenario can be viewed as the $CPFN_{cons}$ of all the $CPFN_i$ from the states (rows) of the scenario. That is, the state agents can be used to construct individual CPFNs, and the CPFN constructed by Mimer can be viewed as the cumulative consensus CPFN of the n adjacent state agent CPFNs.

The HCN_{cons} is the single-linkage hierarchical clustering representation for the scenario, and is useful for those cases where hierarchical information is desired. It is easy to compute from either the UCN_{cons} or the $CPFN_{cons}$ by examining the lowest-valued weights connecting subgraphs of either of the former [7].

Consensus can be viewed as a formal means of communication among agents, and this confers some interesting additional properties upon the results of the consensus.

Definition 13: The consensus of a set of objects is said to be *pareto* (or *paretian*) if and only if every object in the set having a certain property implies that the consensus object of the set also has that same property.

Theorem 2: Each unweighted, associative network resulting from a canonical scenario having the properties of a scenario designed to generate a specific UCN, CPFN, or HCN, as described in Theorem 1, is pareto in the presence of edges.

Proof: Such a scenario is triangular, and any given co-occurrence of entities is observed and recorded only once, by one proxy agent, during the training period. The proof will be in three parts. First, the UCN will be considered. By definition, the unconstrained network will contain all edges defined by co-occurring entities; thus any edge appearing in any state agent UCN will also be in the consensus UCN constructed by Mimer, so the theorem holds. Second, consider the CPFN. If each state agent (row of the scenario) has a given edge, then the co-occurrence count for that edge is not exceeded by any other edge in the cumulative count of co-occurrences. This is because the first proxy agent, recording co-occurrences in the first c-unit, is the only proxy agent providing input to the first state agent; it is these co-occurrences which must provide the common edge(s) of each state agent throughout the entire scenario. When inverted to form the dissimilarity matrix, the weight for that edge will be as small as any in the entire consensus CPFN, so that edge can not violate any generalized triangle inequality; it will not be pruned from the consensus CPFN. Last, the edges in the HCN, determined by single-linkage constraints, are composed of the edges having smallest weights recursively connecting clusters of increasing size until the entire set of nodes is included. The smallest weights in either the consensus UCN or the consensus CPFN will always be included in the consensus HCN, since the smallest weights correspond to the largest count of co-occurrences. *QED*

It is known that, without additional constraints, the CPFN and HCN from more general scenarios, computed as the cumulative consensus of state agents, are not necessarily pareto in edges. A more thorough discussion of consensus and the pareto property can be found in [1].

6. Integration of Interrelated Concepts Utilizing CONSMAG

The experiments of Bransford and Franks [3] showed that humans are capable of integrating the sequential experiences recorded in *episodic memory* to a memory having associational aspects, often referred to as *semantic memory*. Their primary experiment showed that people could take a string of sentences related by the concepts contained therein, and put them together without explicit thought or effort, into a story containing the concepts and actions depicted in the individual sentences. For example, subjects were given a set of sentences as follows:

> The ants were in the kitchen. The jelly was on the table. The ants ate the sweet jelly. The ants ate the sweet jelly which was on the table. The ants in the kitchen ate the jelly which was on the table.

In testing later, on similar and consistent sentences, the subjects were not able to remember well exactly which sentences they had heard in training, but had the highest confidence in having heard the sentence containing all four of the concepts--The ants in the kitchen ate the sweet jelly which was on the table. Other subsentences were regarded as less likely to have been in the training set in inverse proportion to the number of concepts contained therein. The subjects had integrated the information in the five sentences used for training into one consistent story, and thereafter, were not capable of distinguishing between those sentences used to form the representation in semantic memory and other sentences consistent with the story.

If we imagine the paragraph about ants being scanned, looking for the co-occurrences of key words, we can construct a scenario from the paragraph. Let A be the abbreviation for "ants", J for "jelly", K for "kitchen", and T for "table". Also suppose there are five proxy agents (and therefore five state agents), one for each sentence. Then the scenario is as follows:

(A, K)				
(J, T)	(A, K)			
(A, J)	(J, T)	(A, K)		
(A, J, T)	(A, J)	(J, T)	(A, K)	
(A, K, J, T)	(A, J, T)	(A, J)	(J, T)	(A, K)

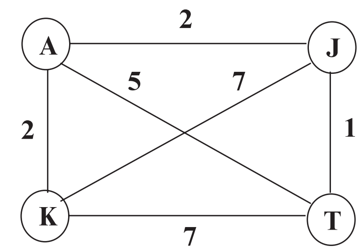
Taking the abbreviations in alphabetical order, the similarity (co-occurrence) and dissimilarity matrices are:

$$S = \begin{bmatrix} - & 6 & 6 & 3 \\ 6 & - & 1 & 7 \\ 6 & 1 & - & 1 \\ 3 & 7 & 1 & - \end{bmatrix} \qquad D = \begin{bmatrix} - & 2 & 2 & 5 \\ 2 & - & 7 & 1 \\ 2 & 7 & - & 7 \\ 5 & 1 & 7 & - \end{bmatrix}$$

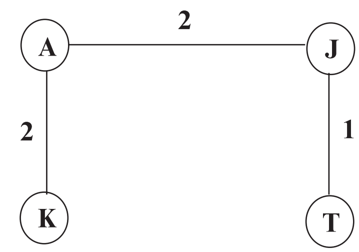
The UCN, $CPFN(r = \infty, q = 3)$, and the HCN (shown as a dendrogram) computed from the D matrix above results in the networks shown in Figure 7.

The UCN is simply all edges (the complete graph) over the set of vertices A, J, K, and T, using the corresponding values in the D matrix for weights. Note that the links connecting A and T, J and K, and K and T were pruned by the Pathfinder paradigm because they violate a (generalized) triangle inequality. The dendrogram for the HCN is most easily constructed from the CPFN, and has two levels of clustering--J and T at level 1, and the other nodes at level 2.

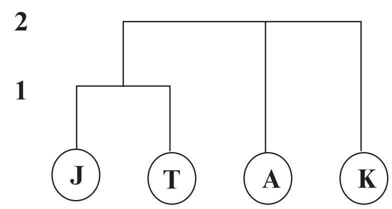
It is clear that CONSMAG provides a means of integrating a collection of co-occurrences over a set of concepts into a set of associational data structures which provides clusterings, minimum-cost paths, connectivities, and hierarchical information, and which can be queried. This can also be viewed as a means of data fusion which utilizes consensus. Consensus can be viewed as a formal means of communication between agents having computable results.



The UCN



The $CPFN(r = \infty, q = 3)$



The HCN Dendrogram

Figure 7 - The Associative Graphs for the Paragraph About the Ants

7. Resource and Cost Issues

The multi-agent paradigm described supports measures of cost for a given learning task in co-occurrence phenomena. The size of the Council of Agents (determined by the number of proxy agents) required to represent the learning task by means of Theorem 1 is one measure, for example. However, if more than one c-term is required in some of the c-units, then this places additional burdens upon the memory required. In addition, the number of key concepts or events in the scenario also affects the ultimate memory requirements. This corresponds to the number of nodes in any of the associative graphs computed. Finally, the number of pairs of concepts co-occurring in the scenario is a factor, since it is these pairs which correspond to edges in the UCN. Thus, in the spirit of the cost vector for Boolean networks (in which the numbers of NAND and NOR gates are often given), a cost vector for a learning scenario is represented by the following factors:

Cost = {| proxy agents|, Ave| c-terms/c-unit|, | key concepts|, | co-occurring pairs|}

It should be noted that the cost vector can be used for a difference graph, the latter denoting the graph which must be learned by a novice to become an expert, as well as for learning any given scenario from the beginning. For the former, ordinal transformations of weights are used to avoid negative weights. Costs usually form a partially ordered set, so direct cost comparisons may not always be possible.

8. Summary and Conclusions

It has been shown that a multi-agent paradigm for modeling co-occurrence phenomena is capable of transforming sequential co-occurrence phenomena into any of three associative graph representations. These data structures can be queried for a variety of relationships among the key concepts. This is accomplished by means of limited and predictable memory requirements, so that memory can be scaled to the problem at hand. Furthermore, the paradigm utilizes cumulative consensus among the state agents, which can be viewed as a formal means of communication, with computable results having known properties. The paradigm also supports cost estimates for any given co-occurrence learning task, by means of a cost vector which can be used to compare the (partially-ordered) costs of various learning tasks.

9. References

[1] J. P. Barthelemy and M. F. Janowitz, "A formal theory of consensus," *SIAM J. of Discrete Mathematics*, Vol. 4, pp. 305-322, 1991.

[2] J. P. Barthelemy and F. R. McMorris, "The median procedure for n-trees,"*J. Classification*, Vol. 3, pp. 329-334. 1986.

[3] J. D. Bransford and J. J. Franks, "The Abstraction of Linguistic Ideas," *Cognitive Psychology*, 2, 331-350, 1971.

[4] William H. E. Day and F. R. McMorris, "Critical comparison of consensus methods for molecular sequences," *Nucleic Acids Research*, Vol. 20, No. 5, 1992.

[5] D. W. Dearholt, "The Cumulative Consensus of Cognitive Agents in Scenarios: A Framework for Evolutionary Processes in Semantic Memory," *Proceedings of the Seventh International Conference of Evolutionary Programming (EP98)*, Invited Plenary Paper and Presentation, San Diego, CA, March; Springer, pp. 3 - 12, 1998.

[6] D. W. Dearholt, "Modeling Dynamic Systems Using Co-occurrence and Associative Networks," *Proceedings of the Tenth International Conference on Mathematical and Computer Modelling and Scientific Computing*, Boston, Massachusetts, July 5, 1995.

[7] D. W. Dearholt and R. Schvaneveldt, "Properties of Pathfinder Networks," In *Pathfinder Associative Networks: Studies in Knowledge Organization*, Ed. R. Schvaneveldt, Ablex, 1990.

[8] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, Vol. 32, No. 3, September 1967.

[9] B. Leclerc, "Medians and majorities in semimodular lattices," *SIAM J. Discrete Mathematics*, Vol. 3, pp. 266-276, 1990.

[10] M. Minsky, *The Society of mind*, Simon and Schuster, 1986.