Semantic graphs and associative memories

Andrés Pomi[1,2,*] and Eduardo Mizraji[1]
[1]*Sección Biofísica, Facultad de Ciencias, Universidad de la República, Iguá 4225, Montevideo 11400, Uruguay*
[2]*Depto. de Biofísica, Facultad de Medicina, Universidad de la República, Gral. Flores 2125, Montevideo, Uruguay*
(Received 1 June 2004; published 23 December 2004)

Graphs have been increasingly utilized in the characterization of complex networks from diverse origins, including different kinds of semantic networks. Human memories are associative and are known to support complex semantic nets; these nets are represented by graphs. However, it is not known how the brain can sustain these semantic graphs. The vision of cognitive brain activities, shown by modern functional imaging techniques, assigns renewed value to classical distributed associative memory models. Here we show that these neural network models, also known as correlation matrix memories, naturally support a graph representation of the stored semantic structure. We demonstrate that the adjacency matrix of this graph of associations is just the memory coded with the standard basis of the concept vector space, and that the spectrum of the graph is a code invariant of the memory. As long as the assumptions of the model remain valid this result provides a practical method to predict and modify the evolution of the cognitive dynamics. Also, it could provide us with a way to comprehend how individual brains that map the external reality, almost surely with different particular vector representations, are nevertheless able to communicate and share a common knowledge of the world. We finish presenting adaptive association graphs, an extension of the model that makes use of the tensor product, which provides a solution to the known problem of branching in semantic nets.

DOI: 10.1103/PhysRevE.70.066136 PACS number(s): 89.75.Hc, 87.23.Ge

Graph-theoretic analysis has been increasingly used in the study of large and complex networks [1], as they emerge from social, technological, and biological domains. Once a graph structure has been defined, various measures can be applied to investigate, for example, average distance among nodes, degree distribution of nodes, and clustering. These techniques are general and widely applied in physics [2]. Recently, growing attention has been paid to diverse complex semantic networks, such as the Thesaurus, WordNet, and free-associations databases [3]. Certainly, the description of the structure of mental associations stored in human memories as concept or semantic graphs has a long tradition [4]. However, a theoretic frame that could explain how neural systems are able to hold graph structures as conceptual networks is lacking. The comprehension of the relations between neurobiology and cognition is the main goal of neural network theories. But the use of graphs in neural network theory has been classically confined to the study of connectivity among neurons or groups of neurons [5]. Here, we want to show that a particular kind of neural network model, distributed associative memories (also known as correlation matrix models), have a structural bond with the adjacency matrix of the graph of associations that could be relevant. In fact, the connection that we describe here allows a technical approximation to an ancient epistemological problem: How can individual brains, which map the external reality with different particular vector representations, communicate and share a common knowledge of the world? Moreover, we show that the mapping between matrix models of distributed memory and associative graphs is extensible to context-dependent associative memories that make use of tensor products to allow different associations to the same cue concept, solving by this way the problem of branching in semantic nets.

I. DISTRIBUTED ASSOCIATIVE MEMORY MODELS

In the field of cognitive neuroscience, the techniques of functional brain imaging [6] have contributed with important advances to the efforts to map cognitive processes on the human brain. The data of neuroimaging have shown that every cognitive activity is supported by patterns of activity of extended and distributed groups of neurons. This finding, in addition to the knowledge that the memory traces are stored superimposed and that memories are tolerant to diffuse damage [7], naturally point to the kind of models of cognitive activities named “distributed associative memories.” These models fit perfectly well with this vision emerging from the current functional neuroimaging.

The models of distributed associative memories, also known as “correlation matrix memories” [8], account in a natural way for some of the distinctive features of human memory (for a systematic approach to this theory we refer the reader to the classical book of Kohonen [9]; a recent physicist’s view on these models can be seen in [10]).

In distributed associative memory models, the units carrying cognitive meaning are large patterns of neural activities, represented by vectors. The memory traces are stored, distributed and superimposed. The basic operation of these models is the association, that is, the capacity of neuronal groups to respond to a certain configuration of neural activity with an associated pattern of neuronal activation. We might then have k such associations summarized by the mapping $\{\mathbf{f}_1 \rightarrow \mathbf{g}_1, \mathbf{f}_2 \rightarrow \mathbf{g}_2, \ldots, \mathbf{f}_k \rightarrow \mathbf{g}_k\}$, where $\mathbf{f}$ and $\mathbf{g}$ are real column vectors representing the patterns of neuronal activities acting

*Corresponding author. Email address: pomi@fcien.edu.uy

as stimulus and associated responses, respectively. In these models the patterns are distinguished by their angles (two colinear vectors represent the same pattern). So, if a set of stimulus vectors $\{\mathbf{f}_i\}$ are to be mutually distinguishable they must be linearly independent.

Including the set of k different, linearly independent, stimulus vectors $\{\mathbf{f}_i\}$ in the k columns of the matrix $\mathbf{F} =[\mathbf{f}_1,\mathbf{f}_2,\dots,\mathbf{f}_k]$, and the corresponding k associated response vectors in the columns of the matrix $\mathbf{G}=[\mathbf{g}_1,\mathbf{g}_2,\dots,\mathbf{g}_k]$, then, an associative memory able to store k arbitrary associations without distortion is given by the general expression [9]

$$\mathbf{M} = \mathbf{G}(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T, \tag{1}$$

where $(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T$ is the Moore-Penrose pseudoinverse, usually written as $\mathbf{F}^+$. Without loss of generality we can so far assume that the mapped input signals $\mathbf{f}_1\dots\mathbf{f}_k$ are orthonormal to one another [11]. In this case the expression of the matrix memory reduces to the simple form

$$\mathbf{M} = \mathbf{G}\mathbf{F}^T, \tag{2}$$

or, equivalently, $\Sigma\mathbf{g}_i\mathbf{f}_i^T$. Such memory, associative and content-addressable, stores the associations distributed and superimposed in the coefficients of the matrix $\mathbf{M}$. These coefficients can be related, in the first instance, to synaptic conductance modifications of the neurons represented by the rows of the matrix.

II. MEMORY MODEL FOR THE ASSOCIATIONS BETWEEN CONCEPTS

Now consider the utilization of an associative memory to model the store of associations between concepts from the real world, such as those represented in semantic nets as free associations databases [3]. The nervous system stores the concepts in a distributed form, as a certain pattern of neuronal activation [12].

We represent h distinct concepts by vectors chosen within an arbitrary orthonormal basis $\{\mathbf{w}\}$ that spans R^n, $n\geqslant h$. The dimension n corresponds to that of the biological domain and takes into account all the possible activations for any given concept. Our memory stores the associations elicited by each concept acting as a stimulus and giving as a result another concept. So, both stimulus vectors and associated responses—generically, $\{\mathbf{f}_i\}$ and $\{\mathbf{g}_i\}$—map concepts, and they belong to the same n-dimensional vector space.

A memory $\mathbf{M}_w$ that associates to each cue concept $\mathbf{w}_i$ another arbitrary concept $\mathbf{w}_j$, is written as a sum of h outerproducts $\mathbf{w}_i'\mathbf{w}_i^T$

$$\mathbf{M}_w = \sum_{i=1}^{h} \mathbf{w}_i'\mathbf{w}_i^T, \tag{3}$$

where $\mathbf{w}_i'=\mathbf{w}_j$ for $1\leqslant j\leqslant h$, or in the compact form $\mathbf{M}_w =\mathbf{G}_w\mathbf{F}_w^T$, where $\mathbf{F}_w$ is a rectangular matrix $(n\times h)$ and its columns are the h mutually orthogonal concepts ($\mathbf{F}_w =[\mathbf{w}_1,\mathbf{w}_2,\dots,\mathbf{w}_h]$). The set of h associated concepts corresponding to each of the stimulus concepts are chosen within the same set $\{\mathbf{w}\}$ and are packed in the matrix $\mathbf{G}_w(n\times h)$. Notice that $\mathbf{G}_w$ admits repetition within its columns inasmuch a same concept can be retrieved as a result of different stimulus concepts.

III. GRAPH OF ASSOCIATIONS

Graphs $\Gamma(V,E)$ are mathematical objects defined by an ordered pair of sets (V,E). The set V is the set of *vertices* and E, a subset of the set of unordered pairs of V, is the set of *edges* [13]. If the pairs of vertices are ordered, the edges of the graph are named arcs, represented by arrows, and the graph is named as a directed graph or a digraph. Every graph can be univocally determined by its adjacency matrix $\mathbf{A}(\Gamma) =[\mathbf{a}_{i,j}]$. The adjacency matrix of a graph is a square matrix (the same dimension of the vertex space V), whose entries take values 1 or 0, depending on the existence of an arc linking the pair of vertices ij.

Given a finite number of concepts, an associative memory is characterized by its capacity of responding to each concept with the evocation of another concept. An association is, therefore, an ordered relation between pairs of concepts. Then, the set of associations configures a digraph $\Gamma(V,E)$ in a vertex space V of possible concepts, being their arcs E the set of associations between concepts instructed in the memory. We name this digraph as "graph of associations."

In this graph of associations only one arrow outgoes from each vertex, so each vertex of the digraph has *outdegree* 1, which is the definition of a *functional digraph* [14]. As the vertices are in correspondence with concepts represented by real vectors of length l, such that any two vectors are orthogonal, the vertex space V is an orthonormal basis of a vector space of dimension equal to the number of vertices $|V|$. Let us call this a *complete* orthonormal representation [15] of the graph of associations.

Hence, distributed memories admit two complementary representations: (a) its matrix resulting from the sum of outerproducts of Eq. (3), strongly dependent on the particular code of concepts, and (b) its graph of associations. Figure 1 shows the distributed memory, the corresponding graph of associations and the adjacency matrix for an example of associations between only four different concepts.

IV. ADJACENCY MATRIX AS A MEMORY

Now we return to the matrix memory. $\mathbf{M}_w(n\times n)$ is dimensional redundant inasmuch as the different concepts were codified with n-dimensional vectors imposed by the biological information processing. Thus, for the analysis of the memory we only need to consider the abstract concept vector space. In this case it is always possible to do an orthogonal change of basis by mapping the biological n-dimensional representation into an abstract representation in the h-dimensional space of concepts.

We transform the vectors of the basis $\{\mathbf{w}\}$ to the standard basis of the concept space constituted by the column unit vectors of the h-dimensional identity matrix: $\mathbf{e}_1 =(1,0,\dots,0)^T$, $\mathbf{e}_2=(0,1,\dots,0)^T,\dots,$ $\mathbf{e}_h=(0,0,\dots,1)^T$. Rewriting the matrix memory in the standard basis gives the following sum of outerproducts:

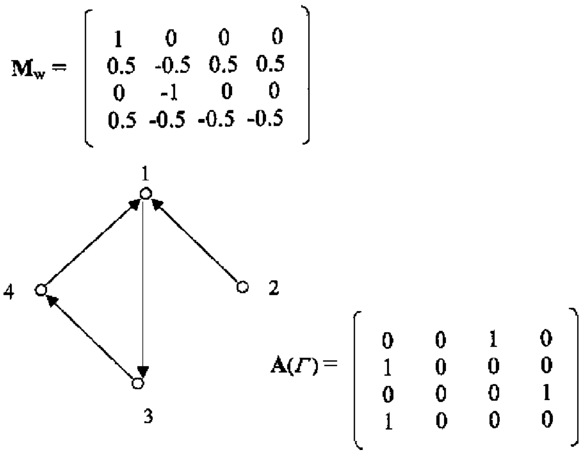

FIG. 1. A memory storing the associations $\{1\to 3, 2\to 1, 3\to 4, 4\to 1\}$ between four concepts coded by vectors of an orthonormal basis $\{\mathbf{w}\}$. The matrix memory $\mathbf{M}_{\mathrm{w}}=\mathbf{G}\mathbf{F}^{\mathrm{T}}=[\mathbf{w}_3\mathbf{w}_1\mathbf{w}_4\mathbf{w}_1]\times[\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3\mathbf{w}_4]^{\mathrm{T}}$ is generated by the sum of outerproducts $\mathbf{M}_{\mathrm{w}}=\mathbf{w}_3\mathbf{w}_1{}^{\mathrm{T}}+\mathbf{w}_1\mathbf{w}_2{}^{\mathrm{T}}+\mathbf{w}_4\mathbf{w}_3{}^{\mathrm{T}}+\mathbf{w}_1\mathbf{w}_4{}^{\mathrm{T}}$. Representing $\mathbf{w}_1\ldots\mathbf{w}_4$ by the corresponding columns of a normalized 4×4 Hadamard matrix, the matrix memory looks as in the figure. This memory can also be represented by its graph of associations. The four concepts are represented by the set of vertices $\mathrm{V}=\{\nu_1,\nu_2,\nu_3,\nu_4\}$ and the associations by the arcs $\mathrm{E}=\{\nu_1\nu_3,\nu_2\nu_1,\nu_3\nu_4,\nu_4\nu_1\}$. $\mathbf{A}(\Gamma)$ is the adjacency matrix of this graph.

$$\mathbf{M}_e=\sum_{i=1}^{h}\mathbf{e}_i'\mathbf{e}_i^T, \qquad (4)$$

where $\mathbf{e}_i'=\mathbf{e}_j$ for $1\leq j\leq h$. For the particular four-concept example of Fig. 1, we have

$$\mathbf{M}_e=\mathbf{e}_3\mathbf{e}_1^T+\mathbf{e}_1\mathbf{e}_2^T+\mathbf{e}_4\mathbf{e}_3^T+\mathbf{e}_1\mathbf{e}_4^T, \qquad (5)$$

which is another vector codification of the same set of associations represented by the graph of the figure.

Finally, observe that each term of the sum, i.e., each association, is a unit square matrix with 1 in the entry ji and 0 in the rest of the elements. Hence, each term of the sum provides an element of the transpose of the adjacency matrix of the graph of association

$$\mathbf{M}_e=\mathbf{A}(\Gamma)^T. \qquad (6)$$

This means that the adjacency matrix of the graph of associations is just the memory coded in the standard basis $\{\mathbf{e}\}$ of the h-dimensional space of concepts. Note that the transpose is the result of adopting in these models the usual convention of coding the neural activities with column vectors. Coding the concepts with row vectors instead of columns, the adjacency matrix of the graph of associations directly gives the very memory coded in the standard basis.

Remark that: (i) if a complete orthonormal representation of a graph is adopted, the adjacency matrix of the graph can be viewed as a sum of outerproducts of the unit vectors labeling the vertices $\mathbf{A}(\Gamma)=\Sigma_{i=1}^{h}\mathbf{e}_i\mathbf{e}_i'^T$, with $\mathbf{e}_i'=\mathbf{e}_j$, $1\leq j\leq h$; and (ii) if the graph is a functional digraph, then $\mathbf{A}(\Gamma)$ can be viewed as a matrix memory.

V. SPECTRUM OF THE MEMORY

When characterizing the dynamic properties of a neural system with recursion, the final behavior will depend on the spectrum of the matrix memory. Therefore, its determination becomes a desired goal in a real biological associative memory. This goal is, nevertheless, hard to attain, because we do not know the details of the neural codes. However, as long as the distributed associative memory model remains valid, the representation of a neural memory by the graph with the structure of stored associations provides with a simple method to obtain that spectrum.

Let us now consider the relation between the spectrum of the n-dimensional matrix model of the biological memory $\mathbf{M}_w$ and the spectrum of the h-dimensional memory of the abstract concept space $\mathbf{M}_e$. Since $\mathbf{M}_e=\mathbf{A}(\Gamma)^T$ and it is well known that a matrix and its transpose have the same eigenvalues [16], what we wish to relate is the spectrum of the matrix model of the biological associative memory with that of the adjacency matrix of the graph of associations. To resolve the point, observe that the change of basis from the neural n-dimensional concept vectors $\{\mathbf{w}\}$ to the abstract h-dimensional standard basis $\{\mathbf{e}\}$ can be viewed as occurring in two steps: $\{\mathbf{w}\}_n\to\{\mathbf{e}\}_n\to\{\mathbf{e}\}_h$.

This diagram represents these two steps: an orthogonal change of basis from $\{\mathbf{w}\}$ to $\{\mathbf{e}\}$, both n-dimensional, and a dimensional reduction from the biological to the conceptual space. In the first step we have $\mathbf{M}_w$ and a memory coded in the basis $\{\mathbf{e}\}$, but now also with n-dimensional unit vectors $\mathbf{M}_{e(n)}$. So, since $\mathbf{M}_w=\Sigma_{i=1}^{h}\mathbf{w}_i'\mathbf{w}_i^T$ and $\mathbf{w}_i=\mathbf{F}_w\mathbf{e}_i$, (where $\mathbf{F}_w$, the square matrix containing the n column vectors of the $\{\mathbf{w}\}$ orthogonal basis, is here acting as the matrix of change of basis), then the memory can be written as

$$\mathbf{M}_w=\sum_{i=1}^{h}\mathbf{F}_w\mathbf{e}_i'(\mathbf{F}_w\mathbf{e}_i)^T=\sum_{i=1}^{h}\mathbf{F}_w\mathbf{e}_i'\mathbf{e}_i^T\mathbf{F}_w^T=\mathbf{F}_w\mathbf{M}_{e(n)}\mathbf{F}_w^T. \qquad (7)$$

Since $\mathbf{F}_w$ is orthogonal, matrices $\mathbf{M}_w$ and $\mathbf{M}_{e(n)}$ are similar and share the spectrum of eigenvalues. Therefore,

$$\{\lambda(\mathbf{M}_w)\}=\{\lambda[\mathbf{M}_{e(n)}]\}. \qquad (8)$$

In the second step we will evaluate the relation between the spectrum of the matrix memories $\mathbf{M}_{e(n)}$ and $\mathbf{M}_{e(h)}$. First remember that $\mathbf{M}_{e(h)}=\mathbf{A}(\Gamma)^T$ and also that the adjacency matrix of a graph and its transpose have the same eigenvalues. Therefore, $\{\lambda[\mathbf{M}_{e(h)}]\}=\{\lambda[\mathbf{A}(\Gamma)]\}$. Adopting the complete orthonormal representation of the graph, both $\mathbf{A}(\Gamma)^T$ and $\mathbf{M}_{e(n)}$ are orthonormal labelings of the vertices of the same graph the only difference being the dimension of the unit vectors of the standard basis employed.

Labeling the h vertices with h columns of an n-dimensional identity matrix generates $(n\text{-}h)$ unconnected vertices in the graph, that are in correspondence to $(n\text{-}h)$ zero rows and columns in the adjacency matrix. Choosing the first h columns of this identity matrix as labels, the adjacency matrix of the graph [here equals $\mathbf{M}_{e(n)}$] attains the form

$$\mathbf{M}_{e(n)} = \begin{bmatrix} \mathbf{A}(\Gamma) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{9}$$

where the zeroes represent null matrices. It is evident that $\mathbf{M}_{e(n)}$ will have $(n\text{-}h)$ additional zero eigenvalues. Any other labeling of the vertices (obtained by choosing h different n-dimensional unit vectors of the standard basis) results in the permutation of rows and columns of the matrix $\mathbf{M}_{e(n)}$, an operation that preserves the spectrum and gives rise to isomorphic graphs [17].

Consequently, the eigenvalues of any given matrix memory will be the same as those of the adjacency matrix of its graph of associations plus $(n\text{-}h)$ zero eigenvalues.

VI. ADAPTIVE ASSOCIATIONS AND THEIR GRAPHS

In this section we present adaptive associative graphs as an extension of the model that we have already seen. These graphs correspond to memories that can associate different responses to the same given concept, depending on the neural context. In doing so, these graphs and their coresponding memories provide a solution to the known problem of *branching* in semantic nets; this is the practical problem of what to do if there is more than one link leaving a node.

The directed graphs that arise from experiments of free associations in humans [3] can present more than one arrow emerging from each concept. These associative digraphs with outgoing degrees larger than one are population graphs, resulting from the study of a sample of people. Each person responds to the cue concepts with a single association, so the graph of explored associations of an individual is a functional digraph similar to those we have just seen emerging from the classical associative memory models [18].

However, it is well known that a same human mind can retrieve diverse associations, depending on the context accompanying each cue. A human memory in a real nervous system can probably be regarded as a superposition of these kinds of memories and their associative graphs. Context-dependent associative models were proposed in 1989 to permit adaptive associations [19], providing a solution to the problem of branching. In these models making use of the tensor product, a particular associative memory can be extracted from a state of superposition in the same neural substrate, by the action of the context [20]. In these memories, a context is a vector $\mathbf{p}$ representing another pattern of neural activity that modulates a vector stimulus.

The semantic structure of this kind of memory can be represented by its corresponding graph of adaptive associations. This graph is a superposition of the previously described associative graphs. The associations corresponding to a particular context are distinguished in this superposed graph by different edge coloring (here we use Greek letters α and β for the contextual colorings). Let us suppose that the associations instructed in the example of Fig. 1 correspond to a certain context α of neural activity represented by a column vector $\mathbf{p}_\alpha$. Now consider another neuronal context β, represented by the vector $\mathbf{p}_\beta$, orthogonal to $\mathbf{p}_\alpha$. The same set of concept cues $\mathbf{F}=[\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3\mathbf{w}_4]$, but now in the presence of context $\mathbf{p}_\beta$, elicit the associated concepts $\mathbf{G}_\beta=[\mathbf{w}_2\mathbf{w}_4\mathbf{w}_1\mathbf{w}_3]$

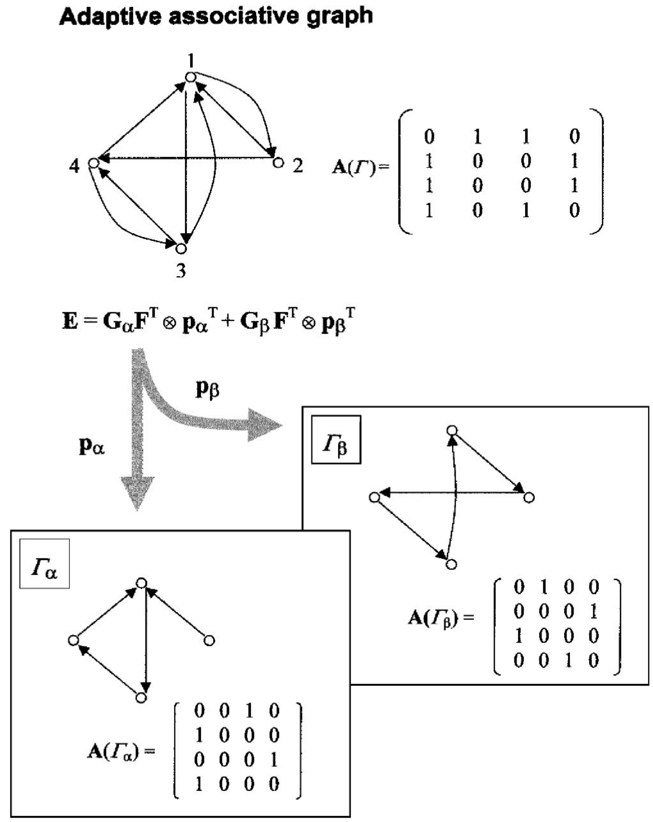

FIG. 2. Human memories can retrieve different associations depending on the context. The semantic structure of such a memory can be represented by a corresponding graph of adaptive associations. At the upper part, the figure shows a multigraph that is a superposition (the graph union) of two functional digraphs, the subgraphs Γ_α and Γ_β, and its adjacency matrix. The Greek letters α and β represent two different edge colorings, corresponding to associations in the presence of different contexts (see text). The context-dependent memory $\mathbf{E}$ stores both sets of associations. $\mathbf{F}$, $\mathbf{G}_\alpha$, and $\mathbf{G}_\beta$ are defined in the text. A given context vector $\mathbf{p}_\alpha$ or $\mathbf{p}_\beta$ alternatively extracts the memories $\mathbf{M}_\alpha=\mathbf{G}_\alpha\mathbf{F}^{\mathrm{T}}$, or $\mathbf{M}_\beta=\mathbf{G}_\beta\mathbf{F}^{\mathrm{T}}$. Their corresponding associative graphs Γ_α and Γ_β are shown with their adjacency matrices.

instead of the previous $\mathbf{G}_\alpha=[\mathbf{w}_3\mathbf{w}_1\mathbf{w}_4\mathbf{w}_1]$. To each one of these sets of associations corresponds an associative graph with one color Γ_α and Γ_β. The adaptive associative graph Γ is their superposition, and it is formally defined as their union $\Gamma=\Gamma_\alpha\cup\Gamma_\beta$. A memory instructed as

$$\mathbf{E} = \mathbf{G}_\alpha\mathbf{F}^T \otimes \mathbf{p}_\alpha{}^T + \mathbf{G}_\beta\mathbf{F}^T \otimes \mathbf{p}_\beta{}^T, \tag{10}$$

where $\otimes$ is the Kronecker product (also named direct or tensor product), stores both sets of associations. A given context $\mathbf{p}_\alpha$ or $\mathbf{p}_\beta$ alternatively extracts the memories $\mathbf{M}_\alpha$ or $\mathbf{M}_\beta$

$$\mathbf{E}(\mathbf{I} \otimes \mathbf{p}_\alpha) = \mathbf{G}_\alpha\mathbf{F}^T = \mathbf{M}_\alpha \tag{11}$$

where, as we have seen, $\mathbf{M}_\alpha$ is in correspondence with $\mathbf{A}(\Gamma_\alpha)$. Hence, looking at the graphs, the operation performed by the context is to dissect the subgraph with its same coloring.

Figure 2 shows the adaptive associative graph described in the previous paragraph and its adjacency matrix. This graph is decomposable in the two associative subgraphs Γ_α or Γ_β, depending on which context, $\mathbf{p}_\alpha$ or $\mathbf{p}_\beta$, is presented to the context-dependent associative memory $\mathbf{E}$.

VII. FINAL REMARKS

We have shown here that distributed associative memory models, also known as correlation matrix memories, admit a natural representation as graphs of association. We defined this associative graph and showed that the adjacency matrix of this graph is just the memory coded in the standard basis of the abstract conceptual space. Therefore, if the neural memory organization can be at some scale represented with this kind of model, our results fulfill the gap between two levels of networks: the neuronal wiring and the semantic nets. In addition, we want to point out that this result of semantic nets emerging from matrix associative memories unifies two representations considered until now as alternatives [21]: one of graphs of knowledge and another of concepts as multidimensional vector spaces.

The second finding we want to note is that we have obtained a code invariant. This result may be important, not only because it provides us with a possible neural support for semantic graphs, but also because it could enlighten the very hard question of the "shared world phenomenon." In what sense do we inhabit a common world? How can perceptual experiences be shared by different individuals provided that, almost surely, the distributed codes employed by two different brains would be different in a detailed level? The code vectors effectively used in the neural representation of a same portion of the world by individuals sharing the same maternal language are the result of modifications of synaptic weights that have an anecdotal character. Our result implies that, in associative memory models, both the graph of associations and its spectrum remain code invariant. This code invariance is not only referred to the election of the basis and its dimension, but also to the election of the labeling of each concept within the vectors of a given basis. We believe that this finding of code invariance of the graph and the spectral characterization of associative memories can provide with a pathway for the comprehension of the phenomenon of shared worlds.

Finally, this code invariance suggests the following. A neural memory is a dynamical system whose final state will depend on the spectrum of the matrix memory. To know the eigenvalues would usually imply the precise knowledge of the neural population coding for each concept, the neuronal connectivity, etc. All of these are extremely difficult tasks and practically unattainable at the neurobiological level. Nevertheless, the code invariance of the representation of the memory by its associative graph and its spectrum provides us with a quick method to predict the dynamics of the neural system, i.e., the sequence of associations, without the necessity of knowing the detailed biological implementation [22]. Actually, the knowledge of the relationships of the information stored in the memory enables us to straightforwardly write the adjacency matrix of the associative graph, find its spectrum, and predict the dynamics of the neural system. Hence, in the case that distributed associative memories remain acceptable models of human memories, an exciting (and also disturbing) possibility of a cognitive engineering emerges. With this cognitive engineering it would suffice to explore *in vivo* the semantic structure of a brain in order to, adding a vertex and some adjacencies or deleting certain key edges, alter the dynamics of the cognition in a desired way.

ACKNOWLEDGMENTS

We would like to thank our colleagues Julio A. Hernández and Luis Acerenza for their comments.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998); A.-L. Barabási and R. Albert, Science **286**, 509 (1999); S. H. Strogatz, Nature (London) **410**, 268 (2001).

[2] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001); R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[3] A. E. Motter, A. P. S. de Moura, Y.-C. Lai, and P. Dasgupta, Phys. Rev. E **65**, 065102(R) (2002); M. Sigman and G. A. Cecchi, Proc. Natl. Acad. Sci. U.S.A. **99**, 1742 (2002); M. Steyvers and J. B. Tenenbaum, Cogn. Sci. (to be published).

[4] For a comprehensive introduction to semantic networks and their history see M. Spitzer, *The Mind within the Net: Models of Learning, Thinking, and Acting* (MIT, Cambridge, 1999), Chap. 10.

[5] See for example F. Crick and C. Koch, Nature (London) **391**, 245 (1998); B. Jouve, P. Rosenstiehl, and M. Imbert, Cereb. Cortex **8**, 28 (1998); O. Sporns, G. Tononi, and G. M. Edelman, *ibid.* **10**, 127 (2000).

[6] M. E. Raichle, Proc. Natl. Acad. Sci. U.S.A. **95**, 765 (1998); Cold Spring Harb Symp. Quant Biol. **61**, 9 (1996); A. Martin, C. L. Wiggs, L. G. Ungeleider, and J. V. Haxby, Nature (London) **379**, 649 (1996).

[7] P. Fries, G. Fernández, and O. Jensen, Trends Neurosci. **26**, 123 (2003); K. L. Hoffman and B. L. McNaughton, Science **297**, 2070 (2002).

[8] These memory models were developed since 1972 principally by J. A. Anderson, Math. Biosci. **14**, 197 (1972); T. Kohonen, IEEE Trans. Comput. **C-21**, 353 (1972); and L. N. Cooper, in *Proceedings of the Nobel Symposium on Collective Properties of Physical Systems*, edited by B. and S. Lundquist (Academic, New York, 1973).

[9] T. Kohonen, *Associative Memory: A System Theoretical Approach* (Springer, New York, 1977).

[10] L. N. Cooper, Int. J. Mod. Phys. A **15**, 4069 (2000).

[11] Arguments concerning the orthogonality of neural vectors can be found in E. T. Rolls and A. Treves, *Neural Networks and Brain Function* (Oxford University Press, New York, 1998); T. Kohonen, *Self-Organizing Maps* (Springer-Verlag, New York, 1997), p. 239.

[12] Although the nervous system is an open system receiving a

continuous flow of information from the external world, as a first approximation one can assume that during the time we are exploring the memory of concepts, there is no incorporation of any new concept. This is a reasonable assumption for an adult.

[13] B. Bollobás, *Graph Theory: An Introductory Course* (Springer-Verlag, New York, 1979).

[14] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1972).

[15] M. Aigner and G. M. Ziegler, in *Proofs from the Book* (Springer-Verlag, Berlin, 2001), attribute to László Lovász the idea of using an *orthonormal representation of a graph* to compute the Shannon capacity of the 5-cycle graph C_5 in L. Lovász, IEEE Trans. Inf. Theory **25**, 1 (1979). It is a system of n unit vectors $\{\nu_i\}$ in a Euclidean space such that if i and j are nonadjacent vertices, then ν_i and ν_j are orthogonal.

[16] S. Barnett, *Matrices: Methods and Applications* (Oxford University Press, New York, 1992).

[17] D. M. Cvetkovic, M. Doob, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1979).

[18] The adjacency matrix of the population graph can be regarded as a Markovian transition matrix if the sum of coefficients of each row is forced to equal one. Such a matrix gives, for a given concept, a probability of different associations as the relative frequency of presentation of each association in the explored sample.

[19] E. Mizraji, Bull. Math. Biol. **51**, 195 (1989); E. Mizraji, A. Pomi, and F. Alvarez, BioSystems **32**, 145 (1994).

[20] A. Pomi-Brea and E. Mizraji, BioSystems **50**, 173 (1999).

[21] T. L. Griffiths and M. Steyvers, in *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (George Mason University, Fairfax, VA, 2002). On-line version: http://www-psych.stanford.edu/~gruffydd/papers/semrep.pdf

[22] A preliminar version of this result was communicated in A. Pomi and E. Mizraji, XIV International Biophysics Congress-IUPAB, Buenos Aires, 2002 (unpublished).