52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Transient Response Functions for Graph Structure Addressable Memory

Weibo Gong
Complex Systems Modeling and Control Laboratory,
Department of Electrical and Computer Engineering,
University of Massachusetts, Amherst, MA 01002, U.S.A.

Abstract— **While neurons in the brain are individually quite slow, collectively they can recognize concrete objects as well as abstract concepts very quickly. Motivated by this puzzling fact we propose biologically plausible algorithms that are capable of quickly recognizing similar graph structures. Graphs are combinatorial constructions and pose serious challenges to similarity testing. In this paper we use the transient behavior of random walk over graphs to compare their spectral resolution. We collect data from intrinsically parallel random walks to form a graph response function as an effective measure of graph similarity. Our algorithm could be a solution to the long standing mystery of content addressability in the brain.**

I. Motivations

This era faces the challenge of understanding human intelligence and while its full understanding may take a bit more time, there is growing evidence that it is now feasible to develop brain-like concept-abstraction algorithms using current knowledge in neuroscience, mathematics, computer science and, in particular, in spatial-time signal conversion.

Human intelligence is believed to be based on neuronal network representations of objects and concepts. In [13] it was suggested that such networks can be abstracted as weighted graphs with random walks modeling the effect of neuronal spiking. In doing so we hope to model the fundamental mechanisms of concept abstraction, a cornerstone of high intelligence.

Concept abstraction is the process of searching for common features from a collection of instances. If graphs are the mathematical representations of instances, it follows that efficient graph similarity testing algorithms are at the core of abstraction. The random walk algorithms discussed in this paper generate a single time function to summarize the features of a graph and then use this function to access similar graphs over a fabric of random connections.

While significant research has been devoted to graph similarity finding algorithms, none has seized on the fact that humans and many other animals recognize visual objects quickly [28], [13]. Such immediate recognition is a result of the tremendous pressures that survival of the fittest has imposed over millions of years of evolution. For example, the fish brain is capable of quickly recognizing the identity of other fish, either as predator or prey, despite the variability in visual orientation. Humans recognize familiar faces immediately, with very little effort. However, we have difficulty in reconstructing faces from memory in the absence of the visual stimulus. If facial features can be encoded onto a graph (via the graph connectivity structure and edge weights), it follows that our brain can then represent a highly-abstracted facsimile of the face as a graph. This model can provide important guidance in developing general concept abstraction algorithms since Nature repeats successful mechanisms.

The above discussion suggests the following constraints for plausible graph structure access algorithms:

- Quick recognition of objects despite slow neuron responses;
- Most neurons have multiple inputs (dendrite branches) with only one output (axon);
- Slight rotation, stretching, or siding of an image does not alter recognition;
- The brain consumes very little energy performing tasks comparing to the current computer algorithms;
- The spatial/temporal signal conversion mechanisms should be plausible for emerging from randomness.

We believe quickness is the key to meeting all the above. In fact it is a hunch that the brain ability of quick recognition and sudden illumination [9] may have a deep connection to the small-time asymptotic results in the theory of diffusion on manifolds that led to the thoughts in the current paper [12], [13]. Consequently our algorithms differ from the other existing algorithms in several important aspects:

- We view the random walk over the representation graph as a dynamic system and we use the initial response to distinguish graphs;
- In our algorithms the earlier the random walk data come in the more important they are, enabling an early decision that achieves the quickness of recognition;
- Our algorithms are intrinsically rotational invariant and robust against other small variations;
- Our algorithms are trivially parallel in that they execute the same simple operations on every edge of the graph and the results collected from all the edges are simply added together to give the final output.

In short, our algorithms use the transient response of a graph to the initial distribution of random walkers to characterize the graph structure. This is analogous to the transient response description of linear dynamic systems. For large and complex graphs the intrinsic parallelism as well as the transient behavior comparison enable fast similarity testing. Since the motion modes of the random walk and their

[0]This work is supported in part by the United States National Science Foundation Grant EFRI-0735974, CNS-1065133 and CNS-1239102, and Army Research Office Contract W911NF-08-1-0233 and W911NF-12-1-03.

strength are governed by the spectral resolution (the eigenvalues and the eigenvectors) of the graphs [26], it is meaningful to use a similarity criterion based on the spectral resolution. A simple example is the similarity testing of graphs that represent independently generated images of white noise. These random graphs have similar spectra resolutions, and are visually similar despite their independence.

Random walk over complex graphs is analogous to diffusions on manifolds in some aspects [22], [1]. The latter has a rich history. Although subtleties always exist, many results in continuous random walks have similar counterparts in general graphs. Some continuous analogies of random walk on a complex graph are useful for us to gain insights into the fundamental nature of the problem. For example one could approximate the behavior of a two dimensional planar graph by a real positive function $f(x, y)$ in a bounded domain in $\mathbb{R}^2$ and use the function to approximate the denseness of the connectivity in a neighborhood. This allows us to use effective similarity testing methods for functions $f(x, y)$ and $g(x, y)$ in the context of graph structure comparison. A well-known method to carry out spatial function similarity testing is the Radon transform widely used in computer tomography(CT). In CT one collects the integrals of the functions along straight lines and applies the Radon transform to reconstruct the original function. If $f(x, y)$ and $g(x, y)$ differ, the line integral data would tell the differences. Radon transform does not reduce the data amount in the case of exact reconstruction. However for practical purposes a small portion of the line integral data would suffice to recognize the significance of the difference. We now try to develop a similar algorithm for complex graphs.

Integrating $f(x, y)$ along a straight line does not have a direct analogue for a graph. However we can think that the normalized $f(x, y)$ values represent an annihilation field of random walk particles. One can run random walk paths through the domain and collect the transient behavior of the random walks, hoping such transient behavior data would contain enough information about the field $f(x, y)$. In the case of graphs such random walks are discrete walks over the graphs and the transient behavior of the random walks could reveal enough graph structure information for similarity testing purposes. In the continuous random walk scenario, the Feynman-Kac formula for the diffusion type partial differential equations tells us that random walk paths specified by certain stochastic differential equations can be used to find the solution of the corresponding partial differential equations (PDEs). The eigen structure information of the PDE operators on the one hand governs the motion modes and their strength of the random walks and on the other hand is closely related to the geometric information of the domain and the field. In the next section we briefly review an intuitive picture of the Feynman-Kac formula to help us thinking of the analogue of the Radon transform.

It should be noted here that even though the analogue between graphs and continuous domains is very useful, there is a critical advantage of using graphs to represent concrete images as well as abstract concepts. The advantage lies in the fact that even just a few links between subgraphs allow random walk to travel from one to another. This provides the capability of automatically testing the similarity for subgraphs lying in different places (for example different regions of the brain). In fact, our graph similarity testing algorithms provide a plausible mechanism for a graph to find a similar graph as a counterpart or as an extension, in the spirit of content addressable memory. It has long been postulated that human memory is largely content addressable [3], [23]. However the concrete mechanism for content addressability in the brain remains a mystery. Assuming concepts in the brain are represented by the connectivity structure of weighted graphs, we propose an algorithmic approach for content addressable memory in the brain. The cognitive scientist, author of the book "How the mind works" [24] Steven Pinker referred to our proposal as a "graph structure addressable" scheme in a discussion, which we use.

II. Analogies in continuous random walk and system transient analysis

A major inspiration of our development is the connection between the quickness of recognition in the brain and the small-time asymptotics results of diffusion on manifolds [12], [13]. Consider a closed domain in a 2D Euclidean space denoted as D with its smooth boundary ∂D. Suppose we have an annihilation field $V(x)$ on this domain and we would like to run a diffusion like random walk over the domain and collect the dynamic movement statistics in order to figure out the geometric features of $V(x)$. This task is not much different from the computer tomography scenario mentioned before, except that instead of using the straight lines like in the Radon transform case we try to use the zigzag random walk paths. Feynman's path integral approach provides a hint. Suppose that the random walks started from $x \in D$ at time t. The quantity of the random walker (or the diffusive stuff) at (x, t) is denoted as $u(x, t)$. We want to know if the dynamics of $u(x, t)$ would tell us about the shape of the field $V(x)$. More specifically, suppose that $D \subset \mathbb{R}^d$ is a smoothly bounded domain and that $u(x, t)$ solves the initial-boundary value problem

$$u_t = \frac{1}{2}\Delta u - V(x)u, \quad x \in D, \quad t > 0 \tag{1}$$

$$u(x, 0) = f(x), \quad x \in D \tag{2}$$

$$u(x, t) = 0, \quad x \in \partial D, \quad t > 0. \tag{3}$$

The Feynman-Kac representation of the solution is

$$u(x, t) = E[f(X_t^x)\mathbb{I}_{\gamma^x > t} \exp(-\int_0^t V(X_s^x)ds)] \tag{4}$$

where $X_t^x = x + W_t$ and γ^x is the first time X_t^x hits the boundary.

Intuitively the formula can be understood as follows. First assume that $V(x) = 0$. Although $f(x)$ is the initial "stuff" to be diffused along the paths generated by $X_t^x = x + W_t$, it can also be viewed as the terminal stuff (at time t) to be diffused backwards in time. In other words the paths generated by the Wiener process W_t establish a path fabric in the domain

using continuous zigzag paths. The random walking stuff can be seen as moving on this path fabric either forward in time or backward in time. The weights of the paths remain the same regardless of whether the random walk is forward or backward in time. Thus when the diffusive stuff started at time t reaches position x at time 0, the expected arrival stuff $E[f(X_t^x)\mathbb{I}_{\gamma^x>t}]$ gives the solution $u(x,t)$. The partial differential equation above can be subjected to the spectral decomposition that leads to the method of obtaining the spectral resolution information of the domain, as mentioned before. In fact people have studied the heat content function, defined as $g(t) = \int_D u(x,t)dx$, for obtaining the relevant information about the domain in, e.g., [27], [22], [26]. When $V(x) \neq 0$ one uses Duhamel's principle for the superposition over time and the above thought carries over.

The above formula suggests several possibilities for graph similarity testing algorithms. We can either let $V(x) = 0$ and use the formula $u(x,t) = E[f(x + W_t)\mathbb{I}_{\gamma^x>t}]$ to suggest a pure diffusion over graph with the initial data reflecting the graph connectivity, or to use $V(x)$ to model the connectivity of the graph, or to combine both the initial condition and the potential function to model the graph connectivity. We will discuss some concrete algorithms in later sections.

Complex graphs are computationally demanding for brute force type of comparisons but practical graph similarity testing algorithm needs to be very fast. Can quickness be achieved via zigzag random walk paths? The "small-time asymptotics" results in [27], [22], [26] suggest that this is possible. Historically the small-time asymtpotics about diffusion was considered by the famous mathematician Mark Kac. He gave a lecture in 1965 and subsequently a paper in 1966 with the same colorful title "Can one hear the shape of a drum?" [18]. The key observation in Mark Kac's paper is that for smoothly bounded domains in 2 dimensional Euclidean space, heat diffusion from Dirac delta sources could be used to evaluate important geometric parameters of the domain "immediately". More precisely Kac considered the following diffusion equation on a closed domain $D \subset E^2$ with a smooth boundary ∂D:

$$\frac{\partial \rho(y,x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho(y,x,t)}{\partial y^2}. \tag{5}$$

The initial condition is $\rho(y,x,0) = \delta(y-x)$ and the boundary condition is $\rho(y,t) = 0$ for $y \in \partial D$.

His small-time asymptotic expression ($\sim$) for the heat kernel is

$$\sum_{k=1}^{\infty} e^{-\lambda_k t} \sim \frac{\|D\|}{2\pi t} - \frac{L}{4}\frac{1}{\sqrt{2\pi t}} + \frac{1}{6}(1-r) \quad as \quad t \to 0$$

where $\lambda_k, k = 1, 2, \cdots$ the eigenvalues of the diffusion operator, $\|D\|$ the area of the domain, L the length of the boundary ∂D and r the number of smooth holes inside the domain. If we know any finite segment, in particular a small early segment of the time function $\sum_{k=1}^{\infty} e^{-\lambda_k t}$ then theoretically we can determine $\|D\|, L, r$ and possibly other geometric parameters of the domain. We emphasize that this expression says that not only we can "hear" some important shape parameters of the drumhead, we can hear them "immediately". The practical implications of this immediacy have not been investigated much.

The fact that the system structural information can be obtained "immediately" via a scalar function is actually a familiar one in control and systems theory. Consider a linear time invariant system

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} \tag{6}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector and $\mathbf{u} \in \mathbb{R}^m$ is the control input vector. For the initial condition response ($\mathbf{u}(t) = 0$) we have (assuming that the eigenvalues of A are all distinct)

$$\mathbf{x}(t) = \mathbf{v}_1 e^{\lambda_1 t} + \mathbf{v}_2 e^{\lambda_2 t} + \cdots + \mathbf{v}_n e^{\lambda_n t} \tag{7}$$

where λ_i is the i-th eigenvalue of the system matrix A and $\mathbf{v}_i$ is the corresponding eigenvector. The initial condition response for

$$\mathbf{x}(0) = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \cdots + \alpha_n \mathbf{v}_n \tag{8}$$

is

$$\mathbf{x}(t) = \alpha_1 e^{\lambda_1 t}\mathbf{v}_1 + \alpha_2 e^{\lambda_2 t}\mathbf{v}_2 + \cdots + \alpha_n e^{\lambda_n t}\mathbf{v}_n. \tag{9}$$

If we let the system output be a scalar $y(t) = \mathbf{c}\mathbf{x}(t)$ with $\mathbf{c} = [1, 1, \cdots, 1]$ then we have

$$y(t) = \alpha_1 e^{\lambda_1 t}\mathbf{c}\mathbf{v}_1 + \alpha_2 e^{\lambda_2 t}\mathbf{c}\mathbf{v}_2 + \cdots + \alpha_n e^{\lambda_n t}\mathbf{c}\mathbf{v}_n. \tag{10}$$

In principle one can recover all the eigenvalues and the coefficients $\alpha_i \mathbf{c}\mathbf{v}_i$ from a small segment of the $y(t)$. In particular one can use an early segment for this purpose, leading to a quick identification of the critical system parameters including the motion modes and their strengthes. This is analogous to the above diffusion scenario. In both cases the relations between the dual descriptions in frequency domain and time domain of a dynamic system plays the critical role. In the diffusion case the spatial structure of the potential function or the shape of a domain can be represented in the Fourier decomposition. Such frequency domain description of the "system" is converted to a time function description for the purpose of real-time similarity testing. More generally we consider an m-dimensional Riemannian manifold (M, g) and the associated Dirichlet Laplace-Beltrami operator $-\Delta_M$ acting in $L^2(M, dx)$, where dx is the volume measure on M induced by the metric g. Let $u : M \times [0, \infty) \to \mathbb{R}$ be the unique solution of

$$\frac{\partial u}{\partial t} = \Delta_M u, \quad t > 0, \tag{11}$$

with the initial condition

$$u(x, 0) = 1 \tag{12}$$

and the boundary condition

$$u(x,t) = 0, \quad x \in \partial M, t > 0. \tag{13}$$

Using the spectral resolution $\{\lambda_k, \phi_k\}$ of Δ_M, where ϕ_k denotes the eigenfunction corresponding to the eigenvalue

λ_k with $\lambda_1 < \lambda_2 < \cdots$, the Dirichlet heat kernel for M can be written as

$$p(x,y,t) = \sum_{k=1}^{\infty} e^{-\lambda_k t}\phi_k(x)\phi_k(y) \tag{14}$$

and the heat content in the domain is

$$\begin{aligned} h(t) &= \int_M u(x,t)dx = \int_M \int_M p(x,y,t)dydx \\ &= \sum_{k=1}^{\infty} e^{-t\lambda_k}\left(\int_M \phi_k(x)dx\right)^2. \end{aligned} \tag{15}$$

It is worth noting that $\int_M \phi_k(x)dx = \int_M \phi_k(x)u(x,0)dx$ is the Fourier coefficient of the initial condition $u(x,0) = 1$ in the coordinate system $\{\phi_1, \phi_2, \cdots\}$. In other words the domain's Fourier coefficients are carried by the exponential motion modes in the heat content to represent the shape information. Due to the unitary property of the Fourier transform the heat content coefficients are robust against small variations in the initial distribution. The graph analogy of this is important for our algorithms to be plausible.

In general if the operator involved in the diffusion type of equations has distinct real eigenvalues, the time function $h(t) = \int_D u(x,t)dx)$ is an exponential sum of the form $\sum_{i=1}^{n} \alpha_i e^{-\lambda_i t}$. Thus, in principle, one can extract the information about the parameters $\alpha_i, \lambda_i, i = 1, \cdots, n$ from a small initial segment of $h(t)$. However it is worth noting that [10], [26] give examples where "isoheat" (namely $h(t)$ are the same) does not imply isospetral (the eigenvalues are the same). These examples involve unusual constructions and should not concern us for our purposes.

The reason we say "in principle" or "theoretically" is that the noise in the $h(t)$ estimates would seriously affect the numerical comparisons. However when applied to random walk on complex graphs one can execute many independent walks simultaneously and the law of large numbers will help to battle the noise. More importantly, similarity testing is often asking only for a best match, and not the exact values of the time function. The large deviation principle ensures that the "order ranking" converges exponentially fast compared to the inverse square root convergence of the function values asserted by the central limit theorem [15], [11].

We now turn to the description of a graph similarity testing algorithm that resembles the continuous situation where an uniform initial condition and a potential function or a metric are used.

III. A GRAPH SIMILARITY TESTING ALGORITHM

Our graph is denoted as $G(V,E,W)$ where V is the set of the graph vertex with $|V| = n = N^2$, E is the set of the edges and W is the set of the weights on the edges. In general the weights may not be symmetrical. We now consider the graph $G = (V,E,W)$ with a nonempty set of boundary vertexes $V_B \subset V$. The adjacency matrix of the graph G is denoted as $A = [w_{v,u}]$ with w_{vu} being the weight of the edge from vertexes u to v. The degree matrix is $D = diag\,[d_u]$ with $d_u = \sum_v w_{vu}$. If we let a discrete time random walker to start from a vertex u and move to vertex v with probability w_{vu}/d_u then the movements can be described by a Markov chain

$$P_{k+1} = MP_k \tag{16}$$

with $M = AD^{-1}$. The boundary vertexes are corresponding to the absorbing states of the Markov chain.

Since our graph G is supposed to reflect realistic neuronal networks (at a high level) rather than a pure mathematical construction, we assume that the eigenvalues of the matrix M are distinct and write

$$M = \Phi\Lambda\Phi^{-1} \tag{17}$$

where Λ is a diagonal matrix with the eigenvalues of M as the diagonal entries. In fact we can write the above equation in the following "spectral decomposition":

$$M = \sum_{k=1}^{n} \lambda_k \phi_k \psi_k^T. \tag{18}$$

Here λ_k is the k-th smallest eigenvalue of M, ϕ_k and ψ_k^T are the corresponding column and row eigenvector, respectively. Note that M may not be symmetric and the eigens may have complex conjugate pairs.

Suppose our random walk starts with a uniform initial condition, namely

$$p_0 = [1/n, 1/n, ..., 1/n]^T, \tag{19}$$

the probability vector for the Markov chain evolves as

$$p_{i+1} = Mp_i = \cdots = M^{i+1}p_0 \tag{20}$$

where the first element of p_{i+1} is the sum of the first row of M^{i+1}/n.

We call the function

$$g_{i+1} = [r_1, \cdots, r_n]p_{i+1} \tag{21}$$

the *graph response (function)* (grf) with

$$r_u = \begin{cases} 1 & \text{if } u \in V/V_B \\ 0 & \text{if } u \in V_B \end{cases}. \tag{22}$$

In other words $[r_1, \cdots, r_n]$ is a vector with its components indicating whether the corresponding graph vertex belongs to the boundary vertex set V_B. The importance of the graph response is due to

$$\begin{aligned} g_{i+1} &= \textstyle\sum_{u\in V, v\in V} \sum_{k=1}^{n} \lambda_k^{i+1}\phi_k(u)\psi_k^T(v) \\ &= \textstyle\sum_{k=1}^{n} \alpha_k \lambda_k^{i+1} \end{aligned} \tag{23}$$

where $\alpha_k = \sum_{u\in V, v\in V} \phi_k(u)\psi_k^T(v)$ with $\phi_k(u)$ and $\psi_k^T(v)$ denote the corresponding eigenvector component. As can be seen if one has enough number of the function g_i values then in principle one can recover the spectral parameters $\alpha_k, \lambda_k, k = 1, \cdots, n$ which describe the dynamics of the random walk and reflect the spectral structure of the graph. We emphasize that the values of the function g_i can be estimated from a simulation of the Markov chain with $M = AD^{-1}$. However such simulation is not efficient numerically. Firstly, at each time step of the simulation some vertexes may

emit a large numbers of walkers. Secondly the function g_i might drop very quickly in i before we harvest enough g_i values with sufficient accuracy. The well-known lazy walk scheme is suitable here for dealing with both issues.

The random walk that governed by the following transition matrix is called a lazy walk on graph:

$$M_L = (1-\delta)I + \delta M = (1-\delta)I + \delta AD^{-1}. \quad (24)$$

If we execute the lazy walk N steps in $[0,t)$ with the step size $\delta = t/N$ then we have

$$M_L^N = [I - \frac{t}{N}(D-A)D^{-1}]^N. \quad (25)$$

One can see that M_L^N would converge to a meaningful limit. The spectrum for the Markov chains M and M_L determines each other for a given step size δ. Since AD^{-1} is diagnolizable, so is $I - \delta LD^{-1}$ and there exists an nonsingular matrix Φ_L such that

$$M_L^N = \Phi_L \Lambda_L^N \Phi_L^{-1}. \quad (26)$$

In fact $\Phi_L = \Phi$ and Λ and Λ_L are uniquely related:

$$\Lambda_L = [(1-\delta)I + \delta\Lambda] = [I - \delta(I-\Lambda)]. \quad (27)$$

Furthermore (26) can be written as

$$M_L^N = \sum_{i=0}^{N} \lambda_{L,k}^i \phi_k \psi_k^T \quad (28)$$

and we can base our similarity testing on the graph response for the lazy walk

$$g_{L,i+1} = \sum_{k=1}^{n} \alpha_k \lambda_{L,k}^{i+1}. \quad (29)$$

The algorithm is to collect all the transitions of the random walkers from one vertex to another for all the vertex pairs in a Markov chain simulation. In other words our algorithm provides a Monte Carlo estimate of the graph response. Since each random walker walks independently, we are benefitting from the law of large numbers in terms of the estimates variance when the number of the graph vertexes is large. We note that our algorithm is trivially parallel. We also note that due to STDP the networks in the brain are mostly directional. However they can be modeled by asymmetric weighted graphs with very small weights on some edges. The eigens are likely to be in complex conjugate pairs, resulting in decayed oscillatory grfs. In fact the decayed oscillatory grfs would have more discriminative power than the purely decayed ones. A rotationally asymmetric connection scheme would be needed to take advantage of this.

The above basic algorithm can have the following generalization that might be of interest in neuroscience and psychology on issues involving attentions. Suppose we would like to emphasize more the traffic from a heavily connected vertex to a less connected vertex. Here the connectivity of a vertex v is described by the total edge weight dv. It is reasonable to insert a multiplicative factor $(dv/du)^\eta, \eta > 0$ for the traffic from v to u. We call this factor the "ridge factor" since its purpose is to emphasize the connections from heavily connected regions to the less connected ones. In the matrix notations this insertion is amount to change the i-th step graph response expression from $g_i = rM^i p_0$ to $g_i = rD^{-\eta}M^i D^\eta p_0$. The choice of η provides a tuning mechanism for classification since in some cases the traffic from the highly connected vertex to the lighter ones is indicative of the graph feature, such as in the case of identifying a sketch of a person with a photo.

IV. The impact of the boundary setting

For a Markov chain M with $m_{ij} = \text{Prob}(j \to i)$ with some absorbing states we can renumber the states such that the transient states come first. Then the transition matrix can be written as the following block matrix:

$$M = \begin{bmatrix} A & 0 \\ B & I \end{bmatrix}.$$

The graph vertexes corresponding to the columns of I are absorbing nodes. In other words the random walkers arrive at one of these nodes will not move any more. On the other hand a random walker started from nodes corresponding to the columns of A, for example j, could move to either a non-absorbing node i by the transition probability a_{ij} or an absorbing node k by the transition probability b_{kj}. However once the random walker arrives at an absorbing node it goes to itself with probability one.

The selection of the boundary nodes can be seen as the selection of the columns with self loops for the corresponding states. It can be seen that for a very large matrix such choices would not affect the spectral decomposition much. Specifically the eigenvalue interlacing theorems [14] promises that a few link changes would not affect the spectra of complex graphs much.

Absorbing nodes or boundary nodes are important for the graph response functions to pick up the transients that reflect the graph structure. One should select such nodes away from the center of the activities. For example if there are several densely connected clusters in the graph then the absorbing nodes should not be too close to them since the traveling paths of the random walkers should cover these clusters as much as possible. The choice of the absorbing nodes may affect the eigenvectors and thus the α_ks. A good practical choice in image processing tasks would be to use the frames of the images for absorbing nodes when the image is mapped to a graph. On the other hand the neuronal network formed in the brain for objects or concepts might have natural boundaries when components form neuronal cliques.

V. Similarity testing using derivative of the graph response function

While it is easy to normalize the graph response functions for different graphs for value comparisons on computers, it is quite implausible in biological systems such as the brain to have quantities that are exactly the same. The derivative

of the graph response function can be used instead. Since

$$g(t) = \sum_{k=1}^{N} \alpha_k e^{-\lambda_k t} \tag{30}$$

we have

$$g'(t) = \sum_{k=1}^{N} -\lambda_k \alpha_k e^{-\lambda_k t}. \tag{31}$$

In the derivative expressions we see that the impact of the fast modes are emphasized more. These fast modes represent the most important initial flows. For example in image processing they represent the flow on the edge of contrast where the graph vertexes have heavier weights. In fact we postulate that these fast modes also play more important roles in comparing abstract concepts represented by neuronal networks. We also note that derivative comparison allows graphs with different sizes to be tested as similar. This is important for a smaller graph to test similar with a much larger graph in the context of graph structure addressable memory. The smaller graph could serve as the "address" and the large ones as the "content". The content graph could have much more connections to other graphs than the address graph. However we note comparing derivatives demands more accuracy. The arguments at the end of section II is important in this context. Moreover we also note that errors in memory access do happen, especially in the brain.

VI. Random network as similarity testing fabric

In order to test for the similarity, the graph response functions or derivatives need to be synchronized in time. This is not an issue for sequential comparison on computers. However, in the brain, such synchronization poses a challenge for neuronal circuits.

A plausible mechanism is to carry out such a comparison over a large, randomly connected network via spike-time dependent plasticity (STDP) [21], [2], [8]. STDP adjusts the strength of connections between neurons in the brain. With STDP, repeated presynaptic spike arrival a few milliseconds before postsynaptic action potentials leads in many synapse types to long-term potentiation (LTP) of the synapses, whereas repeated spike arrival after postsynaptic spikes leads to long-term depression (LTD) of the same synapse. We refer to these two parts as the casual and anti-causal parts, respectively. We suggest that the graph similarity testing occurs in a path fabric made of a huge number of random connections. Here the path fabric refers to the paths between nodes in the random network. The times it takes a spike to travel down any such path are randomly distributed and cover many different time durations. Graphs to be tested for similarity send spikes with variable bursts, modulated by their respective grfs or derivatives, along many of the paths in the network. Two or more similar graphs would be able to meet at a place where the synchronization occurs. Once synchronized, similar grfs or derivatives would strengthen the connection because the spikes from one graph would arrive just ahead of the ones from the other graph most of the time, and the STDP causal part will be in action. Different grfs or derivatives would have spikes from one graph ahead of the other some of the time and behind at other times, resulting in a weak or no connection.

This mechanism assumes only the random path fabric and STDP. It provides possible explanations for some important phenomena in the brain at an algorithmic level. The random fabric for similarity testing allows the brain to check the similarity of the graph representing the incoming signal with thousands of stored graphs (presumably in the subconscious) in parallel. The strongest similarity found would connect the incoming graph with the stored counterparts and generate stronger spikes. These spikes would pass the threshold of the subconscious and break into the conscious level. The possibility of comparing many pairs of graphs in parallel demands a quick similarity testing algorithm, such as the ones outlined in this paper. Memory system with such parallel similarity finding can be referred to as graph structure addressable memory. We want to note that our proposed mechanism allows the process of similarity finding and consolidation to be constantly active in subconscious while consuming little energy, thanks to the fact that only the small initial segments of the grf or derivative need to be compared. On the other hand, the similarity consolidation should diminish when there is enough dissimilarity between subgraphs. We hypothesize that such subgraphs form the basic units of concepts that one may refer to as "conceplets". Similarity consolidation in subconscious could also happen for highly abstract concepts with the same mechanism, resulting into neuron cliques such as the ones discussed in [19].

To elaborate on using our algorithm for memory searching, we outline a plausible framework of the memory process in the brain. When the brain consolidates a piece of memory, henceforth referred to as a "record" (which is loosely similar to the concept used in [23]), the brain constructs graphs for the components of the record. These component graphs are connected due to the simultaneity of the components' occurrences. Random walks of neuron spikes generate a grf (or derivative, henceforth omitted) for each component graphs and the summed grf elicits random graphs that generate similar grf. The latter is retained as the "head" (again loosely similar to the concept in [23]) of the record. During a recall demanded by either external or inside stimulations the head graph would generate its grf and search for similar graphs. The synchronization of the grfs in such a search is provided by the random networks in the brain. We emphasize that such searching is parallel in the sense that one head graph grf could be compared with thousands of record graphs simultaneously over the brain random network. When the matched record graph is found, the head and the record graphs generate spike streams strong enough to break into the conscious level of the brain. Sometimes some record graph components might not be reached during the initial search, a plausible explanation for a scenario such as when one finds a memory on the "tip of the tongue," but cannot fully recall it. In the context of our grf matching scheme, during this "tip of the tongue" scenario, the brain searches for a component graph in certain areas and tries to join it

into generating the summed grf of the waiting parts of the record. Again, such trial and error processes carried out in parallel for many possible component graphs in the area. The specification of the searched area is usually very clearly demanded by the conscious thought. Finally we would note that in this framework the memory hierarchy is formed by the connection strength.

VII. Connections to related topics

A. Diffusion on manifolds and graphs

The line of works with heat diffusion over manifolds and graphs [18], [27], [5], [7], [22], [6], [10] revealed the beautiful connections between the random walk dynamics and the geometry of its environment. In particular the small-time series expansion coefficients of the heat content have been shown to contain more geometric information about the domain than the eigenvalues alone in many cases. These studies use diffusion or random walk as a mathematical bridge between the theories of differential operators and geometry. In contrast we take the usage of random walk more seriously and literally, and propose a framework in which the random walk of neuronal spikes IS the mechanism for graph structure addressable memory in the brain. The algorithms can also be implemented on computers to build brain-like concept abstraction capabilities.

B. Scale space theory in image processing

When using random walk to gather information of the graph structure it is also important to know which kinds of random walks are suitable in the sense that the pattern of the random walk should not introduce new structural features into the signal. Previous research on scale space theory has clarified this issue. In the area of image processing it has been noted by Linderberg and others [20] that real world objects are composed of different structures at different spacial scales. As such it is important to generate representations at different spacial scales from the signals that originated from real world objects. Under a set of assumptions called the scale-space axioms that include linearity, shift invariance, semi-group structure, non-enhancement of local extreme, scale invariance and rotational invariance, it can be shown that the Gaussian kernel is the unique smoothing kernel that fits the bill. Equivalently the scale-space elements can be defined as the diffusion equation

$$\frac{\partial}{\partial t}L(x,y;t) = \frac{1}{2}\Delta L(x,y;t) \tag{32}$$

with the initial condition $L(x,y;0) = f(x,y)$ the given image. This is consistent with the random walk scheme in our proposed graph similarity testing algorithm. The adjustment of the ridge factor is related to a position dependent diffusion coefficients $D(x)$, whose impact to the scale space theory is under investigation.

C. Fokker-Planck operator

There are also interesting connections to the Fokker-Planck equation which could also serve as a continuous analogue for random walk over graphs. Fokker-Planck equation describes the "heat density" $\rho(t,x)$ diffusion under a drifting potential $U(x), x \in \Omega$ with Ω a smoothly bounded domain. The equation is

$$\frac{\partial \rho(t,x)}{\partial t} = \langle \nabla, \rho(t,x)\nabla U(x) + \nabla \rho(t,x)\rangle \tag{33}$$

with a Dirichlet condition for a smooth boundary $\partial\Omega$. The corresponding sample path equation,$dx = -\nabla U(x) + \sqrt{2}dw$, describes a continuous random walk in the domain. Under suitable conditions the classical gauge transformation enables us to write out the eigendecompostion for the solution of (33) as [25]

$$\rho(t,x) = \sum_{i=0}^{\infty} a_i e^{-\lambda_i t}\phi_i(x) \tag{34}$$

where λ_i, ϕ_i are the i-th eigen pairs for the Fokker-Planck operator in (33). An analogous "graph response (function)" $g(t)$ would be

$$g(t) = \int_\Omega \rho(x)dx = \sum_{i=0}^{\infty} e^{-\lambda_i t}\int_\Omega a_i\phi_i(x)dx = \sum_{i=0}^{\infty}\gamma_i e^{-\lambda_i t} \tag{35}$$

which can be used to describe the eigen structural information of the domain.

We note that sometimes one can relate the drifting force and the diffusion term coefficients so that the Fokker-Planck operator becomes a pure diffusion operator in a Riemannian manifold with a metric calculated from the above terms [4]. This suggests that the graph connectivity analogue in continuous random walks can either be the potential function or the Riemannian metric with possibly different insights.

D. Dynamic system identification

The topic of recovering the shape of a domain from the Laplacian eigenvalues belongs to the spectral inversion problem and has a long history. In fact Kac [18] briefly reviewed this history. It is interesting for us in systems and control to note that originally Hendrik Lorentz formulated the problem as a wave problem in 1910. The issue concerned the formation of standing electromagnetic waves in an enclosure with a perfectly reflecting surface. The conjecture was that for a 2D domain (a membrane) the number of Laplacian eigenvalues less than λ would approach $\|D\|\lambda/2\pi$ when $\lambda \to \infty$ which was proved by Herman Weyl. In the 1960s Mark Kac treated the problem using diffusion theory, making use of the fact that both waves and diffusions are the acting of the same Laplacian operator. While in the wave approach the steady state behavior was the focus of analysis, in the diffusion approach of Kac the subject was the transient behavior of the diffused "stuff". In control engineering there are two main approaches to system identification for a linear time invariant system, namely either to use the sinusoidal input to obtain steady state frequency domain data, or to use a step input or the initial condition for transient time domain data. The latter is less practical in engineering practices due to the noise. However as we discussed earlier it turns out that the time domain transient method becomes quite useful in the identification of a shape or a graph. The drawback of

the transient behavior method can be effectively overcome with the parallel execution of the random walk algorithm on many paths of the domain or the graph.

It seems that much can be learned from the connections between the graph similarity testing algorithms and dynamic system identification. In particular, one notices that the heat diffusion equation on a graph is a linear system, where the graph Laplacian plays the role of the system dynamics matrix A in $\dot{x} = Ax + Bu$. Indeed we used the heat diffusion $\partial h_t/\partial t = -L_n h_t$ to collect the initial condition response. It also helps us to think about how an external input signal could affect activity in the brain. We emphasize that this discussion is based on a linear system model. In fact we would like to suggest that, at the level of algorithmic memory addressing, a linear dynamic system might be easier to code in DNA for robust reproduction than a nonlinear dynamic system. That said, nonlinear dynamics might dominate in other levels.

VIII. Conclusions

The line of mathematical works on the relation between diffusion and geometry pioneered by Mark Kac [18] enlightened our search for possible mechanisms of fast recognition with slow neurons. The line of neuroscience works headed by V.S. Ramachandran on cross wiring for different types of stimuli motivated the idea of using weighted graph as universal representations in the brain [17]. Combining these two thoughts we propose an algorithmic mechanism for (similar) content addressable memory. The main ingredients include (1) Quickness is the key to such a memory mechanism; (2) Random walk dynamics summarize graph structure into a time function; (3) Random connection fabric enables time signal synchronization; (4) Similarity order ranking converges very fast with noisy estimates; (5) Similarity testing for many graphs in parallel is trivially feasible.

The topic of testing the similarity of heavily asymmetric weighted graphs that represent the causality among two or more "static" concept graphs is not discussed in this paper. However the principle mechanism summarized above should be applicable to provide a structure access scheme for such graphs as well. The ability to find similarity in abstract relations could help explain the intriguing "analogical reminding" phenomenon and other related brain behaviors that are trademarks of high intelligence [16]. More importantly, it may help empower computers with such capabilities.

IX. Acknowledgments

The author has benefited from discussions on the topic over the years with the following people (in alphabetical order): Neal Anderson, John Baillieul, Andrew Barto, Roger Brockett, Christos Cassandras, Leon Chua, John Doyle, Richard Ellis, Zheng Fang, Jen Gong, Xiaohong Guan, Dev Gupta, Yu-Chi Ho, Christopher Hollot, Robert Jackson, Qingshan Jia, Rama Janaswamy, Bo Jiang, Jieqi Kang, Patrick Kelly, James Lee, Erik Learned-Miller, Steve Papa, Steven Pinker, Shan Lu, Donald Towsley, Marinos Vouvakis, Jie Wang, Sheng Xiao, Dan Yang, Ping Yuan, Hongkun Zhang, Qianchuan Zhao, Quan Zheng.

References

[1] M. Belkin and P. Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. In *18th Annual Conference on Learning Theory, COLT2005*, pages 486–500. Springer, 2005.

[2] G. Bi and M. Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–10472, 1998.

[3] D.G. Bobrow and D.A. Norman. Some principles of memory schemata. In *D. G. Bobrow and A. M. Collins(eds) Representation and Understanding: Studies in Cognitive Science*, 1975.

[4] R.W. Brockett. Notes on stochastic processes on manifolds. In *Systems and Control in the Twenty-First Century*, pages 75–101. Birkhauser, Boston, 1997.

[5] F.R.K. Chung. *Spectral Graph Theory*. CBMS Lecture Notes. AMS Publication, 1997.

[6] F.R.K. Chung. The heat kernel as the pagerank of a graph. *Proceedings of the National Academy of Sciences*, 2007.

[7] F.R.K. Chung and S.-T. Yau. A combinatorial trace formula. *Tsing Hua Lectures on Geometry and Analysis, International Press*, pages 107–116, 1997.

[8] Y. Dan and M. Poo. Spike timing-dependent plasticity of neural circuits. *Neuron*, 44:23–30, 2004.

[9] M. Gell-Mann. *The Quark and the Jaguar: Adventures in the Simple and the Complex*. Henry Holt and Company, LLC, 1994.

[10] P. Gilkey. Heat content, heat trace, and isospectrality. *Contemporary Math*, 491:115–124, 2009.

[11] P. Glynn and S. Juneja. Ordinal optimization: A nonparametric framework. *Proceedings of the 2011 Winter Simulation Conference*, 2011.

[12] W. Gong. Are stochastic dynamics the foundation of intelligence? In *http://www.ieeecss-oll.org/lecture/are-stochastic-dynamics-foundation-intelligence*. 50th IEEE Conference on Decision and Control, 2011.

[13] W. Gong. Can one hear the shape of a concept? In *Proceedings of 31st Chinese Control Conference*, pages 22–26, 2012.

[14] W.H. Haemers. Interlacing eigenvalues and graphs. *Linear Algebra and its Applications*, 226-228:593–616, 1995.

[15] Y.-C. Ho, Q. Zhao, and Q. Jia. *Ordinal Optimization: Soft Optimization for Hard Problems*. Springer, 2007.

[16] D. Hofstadter and E. Sander. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books, 2013.

[17] E.M. Hubbard and V.S. Ramachandran. Neurocognitive mechanisms review of synesthesia. *Neuron*, 48:509520, 2005.

[18] M. Kac. Can one hear the shape of a drum? *American Mathematical Monthly*, 73(4, part2):1–23, 1966.

[19] L. Lin, R. Osan, and J.Z. Tsien. Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes. *Trends in Neurosciences*, 29(1):48–57, 2006.

[20] T. Linderberg. On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance. *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, 1997.

[21] H. Markram, J. Lbke, M. Frotscher M., and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.

[22] P. McDonald and R. Meyers. Isospectral polygons, planar graphs and heat content. *Proceedings of the American Mathematical Society*, 131(11):3589–3599, 2003.

[23] J. Morton, R. H. Hammersley, and D.A. Bekerian. Headed records: A model for memory and its failures. *Cognition*, 20(1):1 – 23, 1985.

[24] S. Pinker. *How the Mind Works*. Norton, 1997.

[25] H. Risken. *The Fokker-Planck Equation*. Springer, 1989.

[26] M. van den Berg, E.B. Dryden, and T. Kappeler. Isospectrality and heat content. *arXiv: 1304.403v1 [math.SP] 15 April 2013*, 2013.

[27] M. van den Berg and P. Gilkey. Heat content asymptotics of a riemannian manifold with boundary. *J. Funct. Anal.*, 120:48–71, 1994.

[28] L.A. Zager and G.C. Verghese. Graph similarity scoring and matching. *Applied Mathematics Letters*, 21(1), 2008.