

GRAPH DIFFUSION DISTANCE : A DIFFERENCE MEASURE FOR WEIGHTED GRAPHS BASED ON THE GRAPH LAPLACIAN EXPONENTIAL KERNEL

David K. Hammond¹, Yaniv Gur² and Chris R. Johnson²

¹ NeuroInformatics Center, University of Oregon

² SCI Institute, University of Utah

ABSTRACT

We propose a novel difference metric, called the graph diffusion distance (GDD), for quantifying the difference between two weighted graphs with the same number of vertices. Our approach is based on measuring the average similarity of heat diffusion on each graph. We compute the graph Laplacian exponential kernel matrices, corresponding to repeatedly solving the heat diffusion problem with initial conditions localized to single vertices. The GDD is then given by the Frobenius norm of the difference of the kernels, at the diffusion time yielding the maximum difference. We study properties of the proposed distance on both synthetic examples, and on real-data graphs representing human anatomical brain connectivity.

Keywords : Graph Laplacian, Graph diffusion, connectome matrix

1. INTRODUCTION

Many interesting modern scientific problems involve the study of connectivity networks, where the relevant data to be analysed consist of the connection strengths between objects of interest. In many cases, these connection strengths represent the propensity for some quantity (e.g. neural activation level, information, opinions) to be transmitted from one graph vertex to another. In some cases, varying experimental conditions and/or methodologies of graph construction may produce multiple sets of estimated or observed connection strengths for a fixed set of vertices. One example of this is the study of estimating brain anatomical connectivity, where one may vary the subject and/or algorithm parameters for the brain graph construction. In such a setting, one may view each such network produced as a single observation, and it is natural to seek quantitative measures of how similar two networks are. Such a measure may also be used for “dissimilarity embedding”, a technique for extracting features relevant for machine learning by measuring distances to a set of prototype exemplar graphs [1].

In this paper, we study the problem of how to quantify the differences between two sets of connectivity strengths for a set of network vertices. We assume that the connectivity relationship is symmetric, so the networks may be represented as symmetric weighted graphs. We also assume that the two graphs vertices are already in correspondence, and thus in this work do not address the graph-matching problem [2]. For example, graphs with vertices already in correspondence are produced when varying algorithm parameters for estimating brain connection graphs, when the underlying anatomical regions defining vertices are held constant.

This work was funded in part by the NIH/NCRR Center for Integrative Biomedical Computing, P41RR12553

We propose a novel technique for comparing weighted graphs that is related to the diffusion maps framework [3]. Our approach is motivated by the desire to measure how changes in the graph structure affect transmission (of activation level, information, etc.) throughout the graph. We model such transmission in a very general way by heat diffusion, in particular we consider the diffusion pattern graph-signals generated by solving the heat equation with initial conditions isolated to single vertices. The graph diffusion distance is given by averaging the norms of the difference of these diffusion patterns for the two graphs. In addition to providing a difference metric between graphs, the graph diffusion distance can be used to analyse the relative importance of edges upon the overall diffusion structure of a single graph, by examining perturbations due to edge deletion. We explore several properties of both the graph diffusion distance and this edge deletion perturbation, on both synthetic examples and actual anatomical brain connectivity graphs.

Other authors have explored related approaches for computing distances between graphs, notably [4] who construct distances by examining random walks at multiple timescales. Conceptually, our approach is similar to [4], however our approach is simpler, and avoids using an explicit hierarchical graph partition employed to compress the random-walk operators.

2. GRAPH DIFFUSION DISTANCE

The motivating principle behind our approach is the idea that two weighted graphs are similar if they enable similar patterns of information transmission. There are many systems in various scientific applications that may be viewed as a network supporting some transmission process, the details of which may depend heavily on the specific application. As an example, for anatomical brain connectivity graphs, one can view neural excitation as a quantity that can be transmitted. Exactly modeling the transmission of information across the brain in a biophysically plausible manner would require a detailed model of coupled populations of neural networks.

In this work we seek to employ a simple, universal model for network transmission that roughly approximates a wider variety of application specific networks. We will model generic transmission in graphs by heat diffusion. For each vertex i in the graph, given a set of edge weights, we generate a diffusion pattern centered around i by initializing with a localized delta impulse at that vertex, and allowing diffusion to proceed for some time t . Different adjacency matrices will generate different diffusion patterns; the graph diffusion distance is based on measuring the average norm of the differences between such patterns for any two adjacency matrices.

We fix our notation as follows. Let A_1 and A_2 be weighted adjacency matrices for N vertices, so both A_1 and A_2 are symmetric, non-negative, $N \times N$ real matrices with zeros along the principle

diagonal. Note that we are not restricting ourselves to unweighted (binary) graphs. We will make frequent use of the (unnormalized) graph Laplacian operator [5], defined by $L_n = D_n - A_n$ (for $n = 1, 2$), where D_n is a diagonal degree matrix for the adjacency A_n , i.e. $(D_n)_{i,i} = \sum_{j=1}^N (A_n)_{i,j}$.

To describe the diffusion process on a graph with adjacency A (for convenience, we suppress the subscript), we let $v(t) \in \mathbb{R}^N$ be a time-varying vector representing the value of the quantity that is undergoing diffusion at each vertex. The edge weights $a_{i,j}$ describe the conductivity between vertices, so that for two vertices i and j , the quantity $a_{i,j}(v_i(t) - v_j(t))$ represents the flux from vertex i to vertex j across the edge connecting them. Summing over these fluxes for each vertex yields $v'_j(t) = \sum_i a_{i,j}(v_i(t) - v_j(t))$. It is straightforward to verify that this may be written as

$$v'(t) = -Lv(t) \quad (1)$$

where L is the graph Laplacian corresponding to A . With initial conditions $v^{(0)}$ at time $t = 0$, equation 1 has the analytic solution $v(t) = \exp(-tL)v^{(0)}$. Here $\exp(-tL)$ is an $N \times N$ matrix-valued function of t , known as the Laplacian exponential diffusion kernel [6]. We now consider letting $v^{(0)} = e_j$, where $e_j \in \mathbb{R}^N$ is the unit vector with all zeros except in the j^{th} component. Running the diffusion up to time t gives the diffusion pattern $\exp(-tL)e_j$, which is precisely the j^{th} column of $\exp(-tL)$.

We are now ready to define the graph diffusion distance. The columns of the Laplacian exponential kernels, $\exp(-tL_1)$ and $\exp(-tL_2)$, describe the different diffusion patterns centered at each vertex generated by diffusion up to time t under the two different sets of weighted edges. Computing the sum of squared differences between these patterns, summed over all the vertices, yields

$$\begin{aligned} \xi(A_1, A_2; t) &= \sum_{i,j} ((\exp(-tL_1))_{i,j} - (\exp(-tL_2))_{i,j})^2 \\ &= \|\exp(-tL_1) - \exp(-tL_2)\|_F^2 \end{aligned} \quad (2)$$

where $\|\cdot\|_F$ is the matrix Frobenius norm. This defines a family of distance measures depending on the diffusion time t . The graph diffusion distance is given by $\sqrt{\xi}$ at the time of maximal difference, i.e. $d_{gdd}(A_1, A_2) = \max_t \sqrt{\xi(A_1, A_2; t)}$.

Given the spectral decomposition $L = V\Lambda V'$, the Laplacian exponential may be computed by

$$\exp(-tL) = V \exp(-t\Lambda) V', \quad (3)$$

where for Λ , $\exp(-t\Lambda)$ is diagonal with i^{th} entry given by $e^{-t\Lambda_{i,i}}$. We compute $d_{gdd}(A_1, A_2)$ by first diagonalizing L_1 and L_2 , then, a straightforward application of (3) and (2) allows computation of $\xi(A_1, A_2; t)$ for any fixed t . Finally, we optimize over t by a line search to give $d_{gdd}(A_1, A_2)$.

For completeness, we mention here that later we will be comparing the GDD to the simpler edge difference distance, d_{edd} , defined for two adjacency matrices by

$$d_{edd}(A_1, A_2) = |A_1 - A_2|_F. \quad (4)$$

2.1. Properties of GDD

The GDD is a metric, in the strict mathematical sense, i.e.

Proposition 2.1 For any $N \times N$ adjacency matrices A, B, C

- i) $d_{gdd}(A, B) \geq 0$, and $d_{gdd}(A, B) = 0$ iff $A = B$
- ii) $d_{gdd}(A, B) = d_{gdd}(B, A)$
- iii) $d_{gdd}(A, C) \leq d_{gdd}(A, B) + d_{gdd}(B, C)$

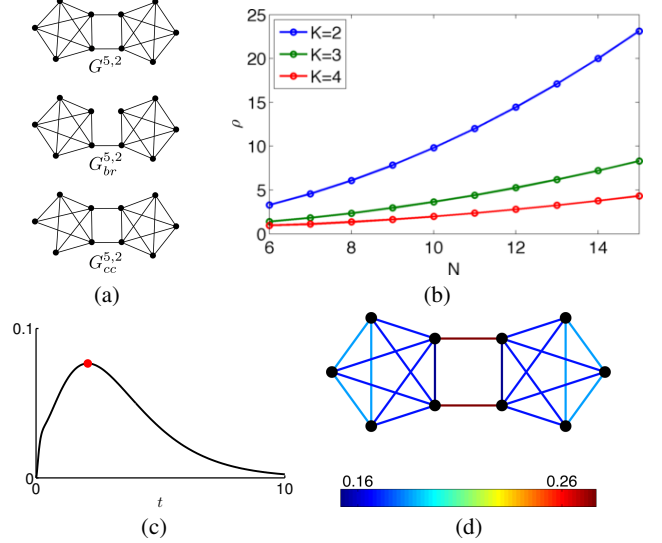


Fig. 1. (a) Barbell graph, and single-edge perturbations, for $N = 5$, $K = 2$. (b) Plot of ratio $d_{gdd}(G^{N,2}, G_{br}^{N,2})/d_{gdd}(G^{N,2}, G_{cc}^{N,2})$ vs N . (c) Plot of $\xi(t)$ for $A_1 = G^{5,2}$, $A_2 = G_{cc}^{5,2}$, red dot indicates maximum, corresponding to $d_{gdd}(A_1, A_2)^2$. (d) Values of normalized edge deletion perturbation, on edges of $G^{5,2}$.

Proof Consider the mapping $\phi : A \rightarrow e^{-tA}$ taking A into $C([0, \infty), \mathbb{R}^{N \times N})$, the space of continuous functions from non-negative real numbers to $N \times N$ matrices. First note that ϕ is one-to-one, as follows: $\phi(A) = \phi(B)$ implies $e^{-tA} = e^{-tB}$ for all $t \geq 0$, then differentiating gives $-Ae^{-tA} = -Be^{-tB}$, and letting $t \rightarrow 0$ shows $A = B$.

Next note that $d_{gdd}(A, B) = \sup_{t \geq 0} \|\phi(A)(t) - \phi(B)(t)\|_F$, so the GDD can be written in terms of the supremum norm, using the fact the $\|\cdot\|_F$ is a proper norm. That d_{gdd} is a metric follows from the properties of the supremum norm.

We note some simple properties of ξ . First, at $t = 0$, the diffusion patterns are still equal to their initial conditions for both A_1 and A_2 , and are thus all equal, which implies $\xi(A_1, A_2; 0) = 0$. Secondly, for any connected graph, i.e. a graph where any two vertices can be connected by some path with nonzero edge weights, as $t \rightarrow \infty$ each diffusion pattern will converge to the constant vector $(1/N, 1/N, \dots, 1/N)^T$. This implies that if A_1 and A_2 are both connected, then $\lim_{t \rightarrow \infty} \xi(A_1, A_2; t) = 0$ (see Figure 1(c)).

Finally, we note an interesting connection between the GDD and $|L_1 - L_2|_F$, the Frobenius norm of the difference of the graph Laplacians. This quantity is closely related to the edge difference distance, specifically $|L_1 - L_2|_F^2 = d_{edd}(A_1, A_2)^2 + \sum_i ((d_1)_i - (d_2)_i)^2$, where $(d_n)_i = (D_n)_{i,i}$ is the weighted degree of vertex i for graph n (for $n = 1, 2$). We have seen that $\xi(t)$ grows from zero at the origin before decaying, and that the GDD is determined by its maximum value. Interestingly, $|L_1 - L_2|_F$ is related to the growth of ξ at the origin, in particular

Proposition 2.2 $\xi(t)$ satisfies $\xi(0) = 0$, $\xi'(0) = 0$, and $\xi''(0) = 2\|L_1 - L_2\|_F^2$, where the derivatives are understood as the right-hand limits $\lim_{t \rightarrow 0^+} \xi'(t)$, $\lim_{t \rightarrow 0^+} \xi''(t)$.

Proof $\xi(0) = 0$ was shown previously. Using the matrix relation $\|X\|_F^2 = \text{tr}(X^T X)$, and that e^{-tL} is symmetric for symmetric L ,

we express $\xi(t) = \text{tr}((e^{-tL_1} - e^{-tL_2})^2)$. Differentiating w.r.t t and simplifying yields

$$\xi'(t) = \text{tr}[-2L_1e^{-2L_1t} - 2L_2e^{-2L_2t} + e^{-L_2t}L_1e^{-L_1t} + L_2e^{-L_2t}e^{-L_1t} + L_1e^{-L_1t}e^{-L_2t} + e^{-L_1t}L_2e^{-L_2t}]$$

$$\xi''(t) = \text{tr}[4L_1^2e^{-2L_1t} + 4L_2^2e^{-2L_2t} - 2L_2e^{-L_2t}e^{-L_1t}L_1 - 2L_1e^{-L_1t}e^{-L_2t} - e^{-L_2t}e^{-L_1t}L_1^2 - L_2^2e^{-L_2t}e^{-L_1t} - L_1^2e^{-L_1t}e^{-L_2t} - e^{-L_1t}e^{-L_2t}L_2^2]$$

Evaluating at $t = 0$ shows $\xi'(0) = 0$, and $\xi''(0) = 2\text{tr}[(L_1 - L_2)^2] = 2\|L_1 - L_2\|_F^2$

2.2. Edge Deletion Perturbation

In addition to comparing two graphs, the GDD may be used to study the relative importance of edges within a single graph by examining the effects of deleting edges. Given an adjacency matrix A , we set $\bar{A}^{(m,n)}$ to be the adjacency matrix with the edge from vertex m to n deleted, i.e. $\bar{A}_{i,j}^{(m,n)} = A_{i,j}$ unless $(i,j) = (m,n)$ or $(i,j) = (n,m)$, in which case the entry is zero. We then define the normalized edge deletion perturbation (EDP) to be $\chi(m,n) = d_{gdd}(A, \bar{A}^{(m,n)})/A_{m,n}$ if (m,n) is a nonzero edge in the original graph, otherwise setting $\chi(m,n) = 0$ when $A_{m,n} = 0$.

We emphasize that the fact that the edge deletion perturbation shows non-trivial behavior is intimately connected with the property that the GDD is constructed from diffusion patterns probing the global structure of the graph. In contrast, consider the edge difference distance d_{edd} defined in equation (4). It is easy to see that deleting a single edge with weight w results in $d_{edd}(A, \bar{A}^{(m,n)}) = \sqrt{2}w$, independent of any other properties of the graph. Therefore, the equivalent normalized EDP would be constant for every edge, and would not reveal any information about structure within the graph.

3. EXAMPLES

As a demonstration of both the GDD and the edge deletion perturbation, we explore their behavior on both a class of simple example graphs, and on real-data derived weighted graphs that represent human brain anatomical connectivity.

3.1. Synthetic graphs

We first construct a simple class of synthetic unweighted graphs we call ‘‘barbell graphs’’ where the behavior of heat diffusion is intuitively easy to understand. We form the unweighted $N - K$ barbell graph $G^{N,K}$ (for $k < N$) by taking the union of two completely connected graphs on N vertices, and then joining K of the vertices in one such component to K in the other component (see Fig. 1(a)). We perturb $G^{N,K}$ by deleting either one of the K ‘‘bridge’’ edges (yielding $G_{br}^{N,K}$), or one of the edges from one of the original completely connected N subgraphs (yielding $G_{cc}^{N,K}$)².

Diffusion on these barbell graphs is straightforward to understand qualitatively. Heat will diffuse quickly within each of the two completely connected N -subgraphs, and only more slowly between these two N -subgraphs. Intuitively, we expect that if K is small relative to N , deleting one of the bridge edges should have a larger

²Specifically, to form $G_{cc}^{N,K}$ we remove an edge connecting vertices that are not incident to any of the bridge edges, which is possible if $K < N - 1$

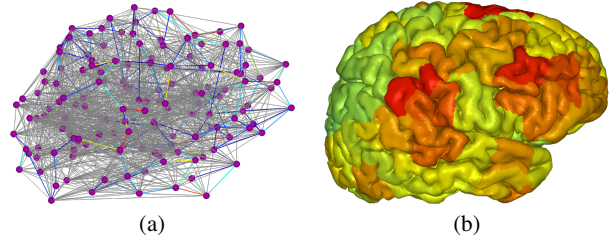


Fig. 2. (a) Normalized edge-deletion perturbation, for brain connectivity graph (thresholded to show only top 10%). (b) Normalized EDP averaged over all edges incident to each vertex, rendered on cortical surface.

effect on the overall diffusion than deleting an edge within each N -subgraph, as such a deletion will have a large effect on the coupling between these two N -subgraphs. Examining the ratio $\rho(N, K) = \frac{d_{gdd}(G^{N,K}, G_{br}^{N,K})}{d_{gdd}(G^{N,K}, G_{cc}^{N,K})}$ allows quantitative study of this effect. These results are shown in Figure 1 (b), where the ratio ρ for fixed K increases with increasing N , and in particular is greater than 1. In contrast, the edge difference distance is completely blind to the difference between these two types of edges, as $d_{edd}(G^{N,K}, G_{br}^{N,K}) = d_{edd}(G^{N,K}, G_{cc}^{N,K}) = \sqrt{2}$ independent of N, K .

The above ratios are formed from single values of the edge deletion perturbation (for these binary weighted graphs, normalizing the EDP by edge weight has no effect). We show an image of the EDP rendered on the edges of $G^{5,2}$ in Fig. 1(d). As can be seen, the bridge edges, as well as edges within the connected components but incident to vertices incident to the bridge edges, have elevated values for the EDP indicating their greater importance to the overall diffusion structure of the graph.

3.2. Brain Connection Graphs

Brain tissue is organized into regions of grey matter, containing neuron cell bodies, that are interconnected by white matter fibers containing axonal projections. Using diffusion-weighted MRI imaging, it is possible to estimate the directionality of these white matter fibers, and reconstruct white matter tract streamlines which estimate their spatial extent. By applying this process (tractography), and counting the number of estimated fibers that connect different segmented grey matter regions, one can estimate a weighted graph which represents anatomical brain connectivity. The development of tractography methods, as well as analysis of anatomical connectome matrices, are very active current research areas [7].

We explore both the edge-deletion perturbation for a single brain connectome, as well as use the GDD to study the degradation of the connectome as the number of underlying diffusion weighted measurements is reduced. Briefly, the tractography used in this work relies on estimating a fiber orientation distribution (FOD) and extracting the dominant diffusion directions (fiber orientations) at each voxel. We combine two approaches that complement each other and have been proven to effectively achieve these tasks [8, 9]. To reconstruct the FODs we use the CT-ODF formulation proposed in [8], which enables the estimation of FODs as positive-definite higher-order tensors by solving a non-negative least squares (NNLS) problem. Second, since this estimation technique outputs the tensor unique coefficients, the tensor decomposition approach proposed in [9] can directly compute the fiber orientations using the tensor coefficients. Fiber streamlines are then computed by integrating

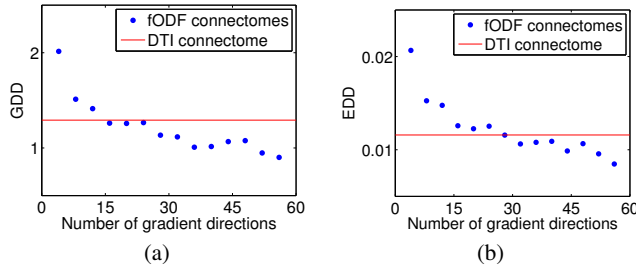


Fig. 3. (a) GDD between A_{60} and A_n , where n is the number of diffusion weighted gradient directions used to compute the FODs. (b) Similar, but using EDD.

those directions using the deterministic streamlining algorithm implemented in the Camino software package [10].

The connectome matrix vertices correspond to patches of the cortical surface. Briefly, we compute a mesh representation of the outer cortical surface from a T1-weighted MRI image, partition it into 150 roughly equal-area patches (75 per hemisphere), align the tracts with the computed cortical surface, and then estimate the connectome edge weights by counting the number of tract streamlines connecting each pair of cortical patches. Further details are described in [11].

We show the normalized edge-deletion perturbation based on the GDD for the connectome for a single subject in Figure 2 (a). Here, for visualization purposes, we show colormap values only for edges whose EDP values exceeded the 90th percentile, other edges are shown in grey. To assist interpretation, for each vertex we compute the average EDP over all edges incident to this vertex, producing a graph signal which can be rendered on the cortical surface as shown in Figure 2 (b). The connectome EDP shows interesting structure, with elevated values near the temporal-parietal junction and in the frontal lobe. Human cortex is broadly divided into dorsal and ventral areas based on cytoarchitectonics (the dorsal division with a greater concentration of pyramidal cells, the ventral division with a greater concentration of layer IV granular cells reflecting the thalamic sensory input). Intriguingly, the areas showing elevated EDP in Figure 2 (b) correspond to two known regions where these dorsal and ventral cytoarchitectonics are mixed [12].

As a final use of the graph diffusion distance, we explore the degradation in the quality of the brain connectome graphs computed using the FOD approach, as the number of diffusion weighted measurements are reduced. As the FODs are computed from a 4th order tensor model that is estimated from the diffusion measurements at each voxel, it is possible to estimate them from a reduced number of diffusion measurements, but at the price of reduced accuracy. We explore the effects of this tradeoff on the connectome matrices in Fig. 3 by showing plots of both $d_{gdd}(A_n, A_{60})$ and $d_{edd}(A_n, A_{60})$, where A_n is the connectome matrix computed based on using only n of the original 60 diffusion weighted images for the FOD reconstruction. Additionally, we measure the distances between A_{60} and the connectome A_{DTI} computed with standard diffusion tensor based tractography [10].

While the overall behavior of both the GDD and EDD is similar, showing larger values as n decreases, there are differences between the two figures. In particular, the GDD shows the error between A_{60} and A_n to be less than that between A_{60} and A_{DTI} for $n \geq 16$, while for the EDD this is obtained only for $n \geq 28$. This demonstrates that GDD and EDD show different sensitivities to the pertur-

bations induced in the connectome matrix due to downsampling the available number of diffusion weighted measurements.

4. DISCUSSIONS / CONCLUSIONS

In this paper, we proposed a novel metric to measure distance between two brain weighted graphs based on the Laplacian exponential diffusion kernel. This graph diffusion distance is computed by searching for a diffusion time t that maximizes the value of the Frobenius norm of the difference between the diffusion kernels. We have described several mathematical properties of the novel metric, and explored its behavior on both simple synthetic graphs and on weighted graphs representing anatomical brain. Future work will include exploring the use of the graph diffusion distance for machine learning applications in anatomical connectivity analysis, as well as the use of the edge-deletion perturbation as a method for feature extraction.

5. REFERENCES

- [1] K Riesen and H Bunke, *Graph Classification and Clustering Based on Vector Space Embedding*, World Scientific, 2010.
- [2] D Conte, P Foggia, C Sansone, and M Vento, “Thirty years of graph matching in pattern recognition,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 03, pp. 265–298, 2004.
- [3] B Nadler, S Lafon, R Coifman, and I Kevrekidis, “Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators,” in *NIPS*, 2005, pp. 955–962.
- [4] Jason D Lee and Mauro Maggioni, “Multiscale analysis of time-series of graphs,” *International Conference on Sampling Theory and Applications (SampTA)*, 2011.
- [5] F K Chung, *Spectral Graph Theory*, vol. 92 of *CBMS Regional Conference Series in Mathematics*, AMS Bookstore, 1997.
- [6] F Fouss, K Francoise, L Yen, A Pirotte, and M Saerens, “An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification,” *Neural Networks*, vol. 31, no. 0, pp. 53 – 72, 2012.
- [7] M Kaiser, “A tutorial in connectome analysis: topological and spatial features of brain networks,” *Neuroimage*, vol. 57, no. 3, pp. 892–907, 2011.
- [8] Y Weldeslassie, A Barmpoutis, and M. S. Atkins, “Symmetric positive-definite cartesian tensor orientation distribution functions (CT-ODF),” in *MICCAI’10*, 2010, pp. 582–589.
- [9] F Jiao, Y Gur, C R Johnson, and S Joshi, “Detection of crossing white matter fibers with high-order tensors and rank-k decompositions,” in *IPMI’11*, 2011, pp. 538–549.
- [10] P. A. Cook, Y. Bai, S. Nedjati-Gilani, K. K. Seunarine, M. G. Hall, G. J. Parker, and D. C. Alexander, “Camino: Open-source diffusion-MRI reconstruction and processing,” in *14th Scientific Meeting of the ISMRM*, 2006.
- [11] D Hammond, B Scherrer, and S Warfield, “Cortical graph smoothing : a novel method for exploiting DWI-derived anatomical brain connectivity to improve EEG source estimation,” *IEEE Transactions on Medical Imaging*, In press.
- [12] D Eidelberg and AM Galaburda, “Inferior parietal lobe: Divergent architectonic asymmetries in the human brain,” *Archives of Neurology*, vol. 41, no. 8, pp. 843–852, 1984.