Collapsed Variational Bayesian Inference of the Author-Topic Model: Application to Large-Scale Coordinate-Based Meta-Analysis

Gia H. Ngo[1], Simon B. Eickhoff[2], Peter T. Fox[3], and B.T. Thomas Yeo[1]
[1]Department of Electrical and Computer Engineering, Clinical Imaging Research Centre
Singapore Institute for Neurotechnology & Memory Networks Programme, National University of Singapore, Singapore
Email: ngohgia@u.nus.edu, thomas.yeo@nus.edu.sg
[2]Institute for Clinical Neuroscience and Medical Psychology, Heinrich-Heine University Dusseldorf, Dusseldorf, Germany
Institute for Neuroscience and Medicine (INM-1), Research Center Julich, Julich, Germany
[3]Research Imaging Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
South Texas Veterans Health Care System, San Antonio, TX, USA

***Abstract*—The author-topic (AT) model has been recently used to discover the relationships between brain regions, cognitive components and behavioral tasks in the human brain. In this work, we propose a novel Collapsed Variational Bayesian (CVB) inference algorithm for the AT model. The proposed algorithm is compared with the Expectation-Maximization (EM) algorithm on the large-scale BrainMap database of brain activation coordinates and behavioral tasks. Experiments suggest that the proposed CVB algorithm produces parameter estimates with better generalization power than the EM algorithm.**

I. Introduction

Neuroimaging studies often report activation coordinates, which can then be used in meta-analyses to gain insights into mind-brain relationships [1, 2, 3]. Most coordinate-based meta-analyses require *expert grouping of task contrasts* to isolate specific cognitive processes of interests, such as internal mentation [4], working memory [5] or emotion [6]. However, expert grouping of task contrasts is non-trivial and introduces some degree of subjectivity. For example, while it is probably reasonable to group "Anti-Saccade" and "Stroop" tasks under a "response conflict" construct [7], differences between the two tasks (e.g., response modalities, response potency, etc) might not be eliminated by standard contrasts.

An alternative approach is to analyze large collections of brain activation coordinates without dividing task contrasts into psychological constructs *a priori* [8, 9, 10, 11]. A recent paper [12] utilized the author-topic (AT) model to encode the notion that a behavioral task engages multiple cognitive components, which are in turn supported by multiple overlapping brain regions. Each cognitive component was found to be supported by a network of strongly connected brain regions, suggesting that the brain consists of autonomous modules each executing a discrete cognitive function. Moreover, connector hubs appeared to enable modules to interact while remaining mostly autonomous [13].

The AT model [14] belongs to the class of models originally proposed to discover latent topics from documents [15]. In the context of neuroimaging, topic modeling has been used to identify latent structure of mental functions and their mapping to the human brain [16], as well as overlapping brain networks from resting-state fMRI [17]. Previous works have proposed algorithms to estimate the parameters of the AT model using collapsed Gibbs sampling [14] or Expectation Maximization (EM) [12]. In this work, we propose a novel Collapsed Variational Bayesian (CVB) inference algorithm to learn the parameters of the AT model.

The CVB algorithm was first proposed for the Latent Dirichlet Allocation (LDA) model [18]. Here we extend CVB to the AT model. By exploiting the Dirichlet-multinomial compound distribution, the CVB algorithm utilizes a less restrictive variational distribution than standard VB (SVB) inference, resulting in a tighter lower bound of the marginal data likelihood. Furthermore, the CVB algorithm marginalizes (collapses) the model parameters, and thus avoids using point estimates of the model parameters to estimate the posterior distribution of the latent variables like in EM. Therefore CVB might result in better parameter estimates than EM.

We applied the AT model to the BrainMap database [19] using EM and CVB. Experiments suggest that the proposed CVB algorithm improves upon the EM algorithm in terms of generalization power of the learned parameters.

II. Background

A. BrainMap database

The BrainMap database [19, 20] contained 2194 published studies at the time of download. Each study comprises one or more "experiments" (i.e., imaging contrasts) comparing two or more imaging conditions. Each experiment is associated with a set of reported activation coordinates in MNI152 space. Overall, our analysis utilized 83178 activation foci from 10499 experiments, categorized into 83 BrainMap-defined task categories.

B. Author-topic Model

The author-topic (AT) was first proposed to discover topics from a collection of documents written by a collection of authors [14]. The AT model was applied to the BrainMap database [12] by encoding the notion that behavioral tasks (analogous to authors) recruit multiple cognitive components (analogous to topics) associated with the activation of various brain locations (analogous to words).

Formally, the variables of the AT model are defined as follows. The dataset consists of $E = 10449$ experiments. The

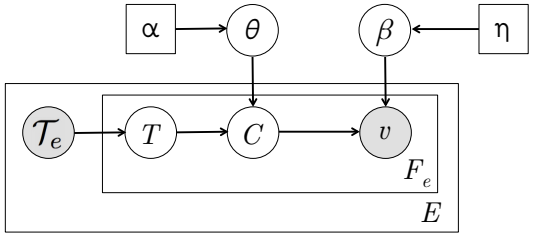

Fig. 1: Graphical model representation of the AT model [12]. Circles represent random variables. Squares represent non-random parameters. Edges represent statistical dependencies. Rectangles (plates) indicate variables within rectangles are replicated. For example, "E" in the outer plate indicates that there are E experiments. See text for detailed explanations.

e-th experiment reports a set of F_e activation foci, with the f-th activation focus located at voxel v_{ef} in MNI152 space. There are $V = 284100$ possible activation locations (MNI voxels). Each experiment is thus associated with an unordered set of F_e activation foci denoted as $\mathbf{v}_e$. The collection of all activation foci across all E experiments is denoted as $\mathbf{v} = \{\mathbf{v}_e\}$. Each experiment utilizes a set of tasks $\mathcal{T}_e$ consisting of one or more of the 83 BrainMap task categories. The collection of all tasks across all E experiments is denoted as $\mathcal{T} = \{\mathcal{T}_e\}$. Therefore the input data consists of $\{\mathbf{v}, \mathcal{T}\}$.

Assuming there are K cognitive components, the probability that a task recruits a component Pr(Component|Task) is denoted by a $|\mathcal{T}| \times K$ matrix θ. The probability that a component activates a voxel Pr(Voxel|Component) is denoted by a $K \times V$ matrix β. Each row of θ and β corresponds to the parameters of a categorical distribution, so each row sums to one. The model assumes symmetric Dirichlet priors with hyperparameters α and η on θ and β respectively.

Fig. 1 illustrates the statistical dependencies of the AT model. The activation foci $\mathbf{v}_e$ of the e-th experiment associated with the set of tasks $\mathcal{T}_e$ are independently and identically distributed. To generate the f-th activation focus v_{ef}, a task T_{ef} is sampled uniformly from $\mathcal{T}_e$. Conditioned on task T_{ef}, a component C_{ef} is sampled from the categorical distribution specified by the t-th row of θ. Conditioned on component $C_{ef} = c$, the activation location v_{ef} is sampled from the categorical distribution specified by the c-th row of β. We note that T_{ef} and C_{ef} are latent variables that are not directly observed. We denote $\mathbf{T} = \{T_{ef}\}$ and $\mathbf{C} = \{C_{ef}\}$ as the collection of all latent task and component variables across all experiments and foci.

C. Expectation-Maximization (EM) of AT model parameters

For fixed hyperparameters α and η, and given the activation foci $\mathbf{v}$ and tasks $\mathcal{T}$ of all experiments, the parameters Pr(Voxel|Component) β, and Pr(Component|Task) θ can be estimated using the EM algorithm, which iterates between the E-step and M-step until convergence [12].

In the $(i+1)$-th E-step, we compute the posterior probability γ_{eftc} that activation focus v_{ef} is generated by task $T_{ef} = t$ and component $C_{ef} = c$ as follows:

$$\gamma^{i+1}_{eftc} \propto \frac{\mathbb{I}(t \in \mathcal{T}_e)}{|\mathcal{T}_e|} \theta^i_{tc} \beta^i_{cv_{ef}} \tag{1}$$

where $\mathbb{I}(t \in \mathcal{T}_e)$ is equal to one if task $t \in \mathcal{T}_e$ and zero otherwise. θ^i and β^i are the estimates of θ and β from the i-th M-step. θ_{tc} is the probability that task t recruits component c and $\beta_{cv_{ef}}$ is the probability that component c activates location v_{ef}.

In the $(i+1)$-th M-step, the model parameters are updated using the posterior distribution estimates:

$$\begin{aligned} \theta_{tc} &\propto (\alpha - 1) + \sum_{e,f} \gamma^{i+1}_{efct} \\ \beta_{cv} &\propto (\eta - 1) + \sum_{e,f,t} \gamma^{i+1}_{efct} \mathbb{I}\left(v_{ef} = v\right) \end{aligned} \tag{2}$$

where $\mathbb{I}\left(v_{ef} = v\right)$ is one if the activated focus v_{ef} corresponds to location v in MNI152 space, and zero otherwise.

Observe that in the $(i+1)$-th E-step, the point estimates θ^i and β^i are utilized to compute the posterior distribution γ^{i+1}. This is avoided in CVB (Section III), thus potentially improving the quality of the estimates.

III. Collapsed Variational Bayesian inference

The CVB algorithm "mirrors" the EM algorithm in that we estimate the posterior distribution of the latent variables $(\mathbf{C}, \mathbf{T})$ (Sec III-A), which is then used to estimate the model parameters θ and β (Sec III-B).

A. Estimating posterior distribution of latent variables

To estimate the posterior distribution of the latent variables $(\mathbf{C}, \mathbf{T})$, we start by lower bounding the log data likelihood using Jensen inequality:

$$\log p(\mathbf{v} \mid \alpha, \eta, \mathcal{T}) = \log \sum_{\mathbf{C},\mathbf{T}} p\left(\mathbf{v}, \mathbf{C}, \mathbf{T} \mid \alpha, \eta, \mathcal{T}\right) \tag{3}$$

$$= \log \sum_{\mathbf{C},\mathbf{T}} q\left(\mathbf{C}, \mathbf{T}\right) \frac{p\left(\mathbf{v}, \mathbf{C}, \mathbf{T} \mid \alpha, \eta, \mathcal{T}\right)}{q\left(\mathbf{C}, \mathbf{T}\right)} = \log E_q \left(\frac{p\left(\mathbf{v}, \mathbf{C}, \mathbf{T} \mid \alpha, \eta, \mathcal{T}\right)}{q\left(\mathbf{C}, \mathbf{T}\right)} \right)$$

$$\geq E_{q(\mathbf{C},\mathbf{T})}\left[\log p\left(\mathbf{v}, \mathbf{C}, \mathbf{T} \mid \alpha, \eta, \mathcal{T}\right)\right] - E_{q(\mathbf{C},\mathbf{T})}\left[\log q\left(\mathbf{C}, \mathbf{T}\right)\right], \tag{4}$$

where $q\left(\mathbf{C}, \mathbf{T}\right)$ is any probability distribution. By finding the variational distribution $q(\mathbf{C}, \mathbf{T})$ that maximizes the lower bound (Eq. (4)), we indirectly maximize the data likelihood. The equality in the inequality Eq. (4) is achieved when $q(\mathbf{C}, \mathbf{T}) = p(\mathbf{C}, \mathbf{T}|\mathbf{v}, \alpha, \eta, \mathcal{T})$, but this is computationally intractable because of dependencies among the variables constituting $\mathbf{C}$ and $\mathbf{T}$. Using standard mean-field approximation [18], the posterior is approximated to be factorizable:

$$q(\mathbf{C}, \mathbf{T}) = \prod_{e=1}^{E} \prod_{f=1}^{F_e} q(C_{ef}, T_{ef}), \tag{5}$$

where $q(C_{ef}, T_{ef})$ is a categorical distribution:

$$q(C_{ef} = c, T_{ef} = t) = \phi_{efct}. \tag{6}$$

Maximizing the lower bound (Eq. (4)) with respect to ϕ and using the constraint $\sum_{c,t} \phi_{efct} = 1$, we get:

$$\phi_{efct} \propto \exp\left(E_{q(\mathbf{C}_{-ef},\mathbf{T}_{-ef})} \log p(\mathbf{v}, \mathbf{C}_{-ef}, \mathbf{T}_{-ef}, C_{ef} = c, T_{ef} = t \mid \alpha, \eta, \mathcal{T})\right), \tag{7}$$

where the subscript "$-ef$" means the corresponding variables with C_{ef} and T_{ef} are excluded.

By exploiting properties of Dirichlet-multinomial compound distribution [18], we can simplify Eq. (7) (details not shown) to become

$$\phi_{efct} \propto \exp[E_{q(\mathbf{C}_{-ef},\mathbf{T}_{-ef})}(\log(\eta + N^{-ef}_{\cdot\cdot cv_{ef}}) - \log(V\eta + N^{-ef}_{\cdot\cdot c\cdot}) - \log(K\alpha + N^{-ef}_{\cdot t\cdot\cdot}) + \sum_a \log(\alpha + N^{-ef}_{\cdot tc\cdot}))], \quad (8)$$

where N_{etcv} is the number of activation foci in experiment e generated by task t, cognitive component c, and located at brain location v. Dot indicates the corresponding indices are summed out. For example, $N^{-ef}_{\cdot\cdot c\cdot}$ is the number of foci that were generated from component c, excluding the f-th focus of the e-th experiment.

Following [18], we perform a second-order Taylor expansion of the $\log(.)$ functions about the means of the counts in Eq. (8), so the final update equation of ϕ is given by Eq. (9) at the top of the next page (details not shown).

Eq. (9) is iterated until convergence. The resulting ϕ is analogous to γ in the EM algorithm (Eq. (1)). Unlike the E-step (Eq. (1)), CVB estimates the posterior distribution ϕ without relying on point estimates of θ and β.

B. AT model parameters estimation

Given the estimate of posterior distribution ϕ (Eq. (9)), the parameters θ and β are estimated by the posterior mean [18]:

$$\theta_{tc} \propto \alpha + \sum_{e,f} \phi_{efct}$$
$$\beta_{cv} \propto \eta + \sum_{e,f,t} \phi_{efct} \mathbb{I}\left(v_{ef} = v\right). \quad (10)$$

Eq. (2) and Eq. (10) are not exactly the same due to differences between MAP (Eq. (2)) and Bayesian (Eq. (10)) estimation.

C. Comparison between standard VB (SVB) and CVB

We now discuss why CVB is theoretically superior to standard VB (SVB), such as [15]). We start with Eq. (4) and apply Jensen's inequality again:

$$\log p(\mathbf{v} \mid \alpha, \eta, \mathcal{T})$$
$$\geq E_{q(\mathbf{C},\mathbf{T})}\left[\log p\left(\mathbf{v}, \mathbf{C}, \mathbf{T} \mid \alpha, \eta, \mathcal{T}\right)\right] - E_{q(\mathbf{C},\mathbf{T})}\left[\log q\left(\mathbf{C}, \mathbf{T}\right)\right], \quad (11)$$
$$\geq E_{q(\mathbf{C},\mathbf{T})q(\theta,\beta)}\left[\log p\left(\mathbf{v}, \mathbf{C}, \mathbf{T}, \theta, \beta \mid \alpha, \eta, \mathcal{T}\right)\right] - E_{q(\mathbf{C},\mathbf{T})q(\theta,\beta)}\left[q(\theta,\beta)\right] - E_{q(\mathbf{C},\mathbf{T})}\left[\log q\left(\mathbf{C}, \mathbf{T}\right)\right], \quad (12)$$

where the first inequality (Eq. (11)) is the same as Eq. (4), and the second inequality (Eq. (12)) corresponds to SVB.

As discussed in Sec III-A, CVB proceeds by maximizing Eq. (11) (or Eq. (4)) with respect to $q\left(\mathbf{C}, \mathbf{T}\right)$, so CVB is equivalent to marginalizing (collapsing) θ and β before approximating the posterior of $\mathbf{C}$ and $\mathbf{T}$. In doing so, CVB avoids having to use the point estimates of θ and β to estimate the posterior of $\mathbf{C}$ and $\mathbf{T}$ like in EM (Eq. (1)).

By contrast, SVB maximizes Eq. (12) with respect to $q\left(\mathbf{C}, \mathbf{T}\right)$ and $q\left(\theta, \beta\right)$. The equality in Eq. (12) is achieved when $q\left(\theta, \beta\right) = p(\theta, \beta|\mathbf{v}, \mathbf{C}, \mathbf{T}, \alpha, \eta, \mathcal{T})$. In general $q\left(\theta, \beta\right) \neq p(\theta, \beta|\mathbf{v}, \mathbf{C}, \mathbf{T}, \alpha, \eta, \mathcal{T})$ for SVB and hence CVB provides a tighter lower bound of the data log likelihood $\log p(\mathbf{v}|\alpha, \eta, \mathcal{T})$. Therefore CVB is better than SVB in theory.

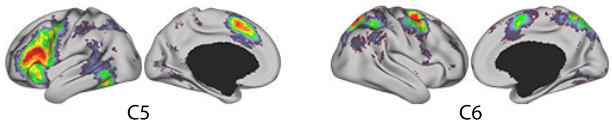

Fig. 2: Surface projection of β (Pr(Voxel|Component)) for two of twelve cognitive components estimated by CVB. Hot colors indicate voxels are likely to be activated by the corresponding components. The two cognitive components are very similar to components C5 and C6 in [12] and are labeled as such.

IV. Experiments

A. Cognitive components estimated by the CVB algorithm

We apply the CVB algorithm to the BrainMap database (Sec II-A). To be consistent with [12], the same parameters were used: $K = 12$ components, $\alpha = 50/K$ and $\eta = 0.01$. The posterior distribution ϕ is initialized randomly and updated using Eq. (9) until convergence. The parameters θ and β are computed using Eq. (10). The procedure is repeated with 100 random initialization of ϕ to produce 100 estimates of θ and β. Following [12], a final estimate was obtained by selecting the solution that was closest to all the other estimates.

The cognitive components estimated by CVB were similar to those estimated by EM [12]. Fig. 2 illustrates Pr(Voxel|Component) for two of the twelve components. The similarity in the estimates of CVB and EM algorithms suggests that both approaches are reasonable. The next section shows that the CVB estimates enjoy better generalization power.

B. Cross-validation of CVB and EM

The BrainMap database was randomly divided into $90\%-10\%$ non-overlapping training-test sets. We apply CVB and EM to estimate θ and β with the 90% training subset and compute the generalization power on the 10% test subset. The procedure is repeated 100 times with a different $90\%-10\%$ split of the database. Generalization power is measured using "perplexity", which measures how "unlikely" test data are according to the trained model [15]. Therefore higher perplexity corresponds to less likelihood of observing the test data and hence worse generalization power. The perplexity of a test set of M experiments is given by:

$$Perplexity(\mathcal{E}_{test}) = \exp\left(-\frac{1}{\sum_{e=1}^{M} F_e} \sum_{e=1}^{M} \log p(\mathbf{v}_e \mid \theta, \beta, \mathcal{T}_e)\right)$$
$$= \exp\left(-\frac{1}{\sum_e F_e} \sum_{e,f} \log \frac{1}{|\mathcal{T}_e|} \sum_{c,t\in\mathcal{T}_e} \beta_{cv_{ef}} \theta_{tc}\right) \quad (13)$$

where β and θ are the parameters estimates given by Eq. (2) for the EM algorithm and Eq. (10) for the CVB algorithm.

Like the previous section, the following parameters were used: $K = 12$ components, $\alpha = 50/K$ and $\eta = 0.01$. For each cross-validation run, the posterior distributions γ (for EM) and ϕ (for CVB) were randomly initialized such that $\phi = \gamma$. This ensures that any difference between the algorithms were not due to different initializations.

Figure 3 shows the histograms of perplexity values from the 100 cross-validation runs. The mean perplexities of CVB

$$\phi_{efct} \propto \left(\eta + E_q\left(N_{..cv_{ef}}^{-ef}\right)\right)\left(V\eta + E_q(N_{..c.}^{-ef})\right)^{-1}\left(K\alpha + E_q\left(N_{.t..}^{-ef}\right)\right)^{-1}\left(\alpha + E_q\left(N_{.tc.}^{-ef}\right)\right)$$
$$\times \exp\left[-\frac{\mathrm{Var}_q\left(N_{..cv_{ef}}^{-ef}\right)}{2\left(\eta + E_q\left(N_{..cv_{ef}}^{-ef}\right)\right)^2} + \frac{\mathrm{Var}_q\left(N_{..c.}^{-ef}\right)}{2(V\eta + E_q(N_{..c.}^{-ef}))^2} + \frac{\mathrm{Var}_q\left(N_{.t..}^{-ef}\right)}{2\left(K\alpha + E_q\left(N_{.t..}^{-ef}\right)\right)^2} - \frac{\mathrm{Var}_q\left(N_{.tc.}^{-ef}\right)}{2\left(\alpha + E_q\left(N_{.tc.}^{-ef}\right)\right)^2}\right] \quad (9)$$

Update equations for the posterior distribution of the latent variables under CVB. Similar to [18], the expectation and variance are computed by Gaussian approximation of the word count distributions. For example, $E_q(N_{..c.}^{-ef}) \approx \sum_{(e',f')\neq(e,f)} \sum_{t\in\mathcal{T}_{e'}} \phi_{e'f'ct}$ and $\mathrm{Var}_q(N_{..c.}^{-ef}) \approx \sum_{(e',f')\neq(e,f)}(\sum_{t\in\mathcal{T}_{e'}} \phi_{e'f'ct})(1 - \sum_{t\in\mathcal{T}_{e'}}\phi_{e'f'ct})$. Note that $\phi_{efct} = 0$ for $t \notin \mathcal{T}_e$.

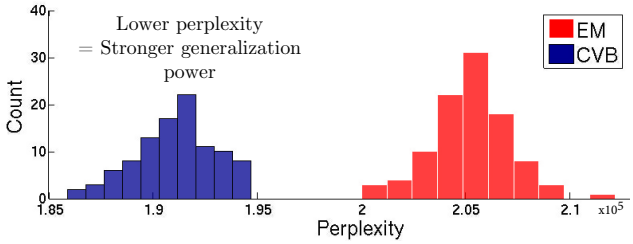

Fig. 3: Histogram of perplexity across 100 test sets. CVB has lower perplexity ($p \approx 0$) or better generalization power than EM.

and EM are 19.1×10^5 and 20.5×10^5 respectively. The improvement of CVB over EM is modest (8%), but highly significant ($p \approx 0$).

V. Conclusions

We proposed a CVB learning algorithm for the AT model. Experiments on a large-scale coordinate-based database show that the CVB algorithm provides parameter estimates with better generalization power than those of the EM algorithm. Therefore, the CVB algorithm can be helpful for quantifying relationships in any neuroimaging data that exhibit the hierarchical structure of the AT model. In addition, the proposed CVB algorithm can be utilized in other applications (e.g., natural language processing) using the AT model.

Acknowledgement

This work was supported by NUS Tier 1, Singapore MOE Tier 2 (MOE2014-T2-2-016), NUS Strategic Research (DPRT/944/09/14), NUS SOM Aspiration Fund (R185000271720), Singapore NMRC (CBRG14nov007, NMRC/CG/013/2013) and NUS YIA. Comprehensive access to the BrainMap database was authorized by a collaborative-use license agreement (http://www.brainmap.org/collaborations.html). BrainMap database development is supported by NIH/NIMH R01 MH074457.

References

[1] T. Yarkoni, R.A. Poldrack, T.E. Nichols, D.C. Van Essen, and T.D. Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nat. Methods*, vol. 8, pp. 665–670, 2011.

[2] P.T. Fox, J.L. Lancaster, A.R. Laird, and S.B. Eickhoff, "Meta-analysis in human neuroimaging: computational modeling of large-scale databases," *Annual review of neuroscience*, vol. 37, pp. 409–434, 2014.

[3] R.A. Poldrack and T. Yarkoni, "From brain maps to cognitive ontologies: informatics and the search for mental structure," *Annual review of psychology*, vol. 67, pp. 587–612, 2016.

[4] R.N. Spreng, R.A. Mar, and A.S. Kim, "The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis," *Journal of cognitive neuroscience*, vol. 21, pp. 489–510, 2009.

[5] C. Rottschy, R. Langner, I. Dogan, K. Reetz, A.R. Laird, J.B. Schulz, P.T. Fox, and S.B. Eickhoff, "Modelling neural correlates of working memory: a coordinate-based meta-analysis," *Neuroimage*, vol. 60, pp. 830–846, 2012.

[6] K.A. Lindquist, T.D. Wager, H. Kober, E. Bliss-Moreau, and L.F. Barrett, "The brain basis of emotion: a meta-analytic review," *Behavioral and Brain Sciences*, vol. 35, pp. 121–143, 2012.

[7] J. Duncan and A.M. Owen, "Common regions of the human frontal lobe recruited by diverse cognitive demands," *Trends in neurosciences*, vol. 23, pp. 475–483, 2000.

[8] S.M. Smith, P.T. Fox, K.L. Miller, D.C. Glahn, P.M. Fox, C.E. Mackay, N. Filippini, K.E. Watkins, R. Toro, A.R. Laird *et al.*, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 13 040–13 045, 2009.

[9] A.R. Laird, P.M. Fox, S.B. Eickhoff, J.A. Turner, K.L. Ray, D.R. McKay, D.C. Glahn, C.F. Beckmann, S.M. Smith, and P.T. Fox, "Behavioral interpretations of intrinsic connectivity networks," *Journal of cognitive neuroscience*, vol. 23, pp. 4022–4037, 2011.

[10] S.B. Eickhoff, D. Bzdok, A.R. Laird, C. Roski, S. Caspers, K. Zilles, and P.T. Fox, "Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation," *Neuroimage*, vol. 57, pp. 938–949, 2011.

[11] N.A. Crossley, A. Mechelli, P.E. Vértes, T.T. Winton-Brown, A.X. Patel, C.E. Ginestet, P. McGuire, and E.T. Bullmore, "Cognitive relevance of the community structure of the human brain functional coactivation network," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 11 583–11 588, 2013.

[12] B.T. Yeo, F.M. Krienen, S.B. Eickhoff, S.N. Yaakub, P.T. Fox, R.L. Buckner, C.L. Asplund, and M.W. Chee, "Functional specialization and flexibility in human association cortex," *Cerebral Cortex*, vol. 25, pp. 3654–3672, 2015.

[13] M.A. Bertolero, B.T. Yeo, and M. DEsposito, "The modular and integrative functional architecture of the human brain," *Proceedings of the National Academy of Sciences*, vol. 112, pp. E6798–E6807, 2015.

[14] M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers, "Learning Author-topic Models from Text Corpora," *ACM Trans. Inf. Syst.*, vol. 28, pp. 4:1–4:38, 2010.

[15] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[16] R.A. Poldrack, J.A. Mumford, T. Schonberg, D. Kalar, B. Barman, and T. Yarkoni, "Discovering relations between mind, brain, and mental disorders using topic mapping," *PLoS Comput Biol*, vol. 8, p. e1002707, 2012.

[17] B.T. Yeo, F.M. Krienen, M.W. Chee, and R.L. Buckner, "Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex," *Neuroimage*, vol. 88, pp. 212–227, 2014.

[18] Y.W. Teh, D. Newman, and M. Welling, "A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation," in *Advances in neural information processing systems*, 2006, pp. 1353–1360.

[19] P.T. Fox and J.L. Lancaster, "Mapping context and content: the BrainMap model," *Nat. Rev. Neurosci.*, vol. 3, pp. 319–321, Apr 2002.

[20] P.T. Fox, A.R. Laird, S.P. Fox, P.M. Fox, A.M. Uecker, M. Crank, S.F. Koenig, and J.L. Lancaster, "BrainMap taxonomy of experimental design: description and evaluation," *Hum Brain Mapp*, vol. 25, pp. 185–198, 2005.