Neuroinformatics on Bornholm

Finn Årup Nielsen

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Lyngby, Denmark

Email: fn@imm.dtu.dk
WWW: http://www.imm.dtu.dk/~fn

ABSTRACT

1) Simple text analysis and modeling of spatial data can be combined: Data from the BrainMap database containing functional neuroimaging studies can be examined for outliers/novelty by simple kernel density estimation of the spatial data ("Talairach coordinates" of activation) conditioned on words and phrases from the labeling of the spatial data. The present application ranks the individual spatial data according to novelty and a manual investigation finds a number of interesting errors and non-conforming terminologies.

2) Term-based text analysis and link analysis can be combined: Terms and the scientific citations were extracted from the functional neuroimaging journal "NeuroImage" and simple link analysis on the citations by singular value decomposisiton identifies hubs and authorities and clusters of similar authors and their subfield. Singular value decomposition of the term-based data together with the citation data enables context sensitive search for authors: Finding the authority related to a term or a set of terms, e.g., from the title or the abstract of a scientific article. This gives a possibility for automatic reviewer identification.

OVERVIEW

- Spatial + text mining.
  - Modeling of data from the BrainMap database
  - Spotting of errors and interesting events in a database.
  - 3D visualization with the Brede matlab toolbox.
  - Technical report, article, abstract submitted to HBM2001
- Text + link analysis.
  - Modeling with data from a single functional neuroimaging journal "NeuroImage" from Academic Press.
  - Simple linear algebra: Author cocitation analysis, journal cocitation analysis, context sensitive authorities, ...
  - Visualization
  - Abstract submitted to HBM2001
- Other ideas

COMBINING SPATIAL AND TEXT MINING

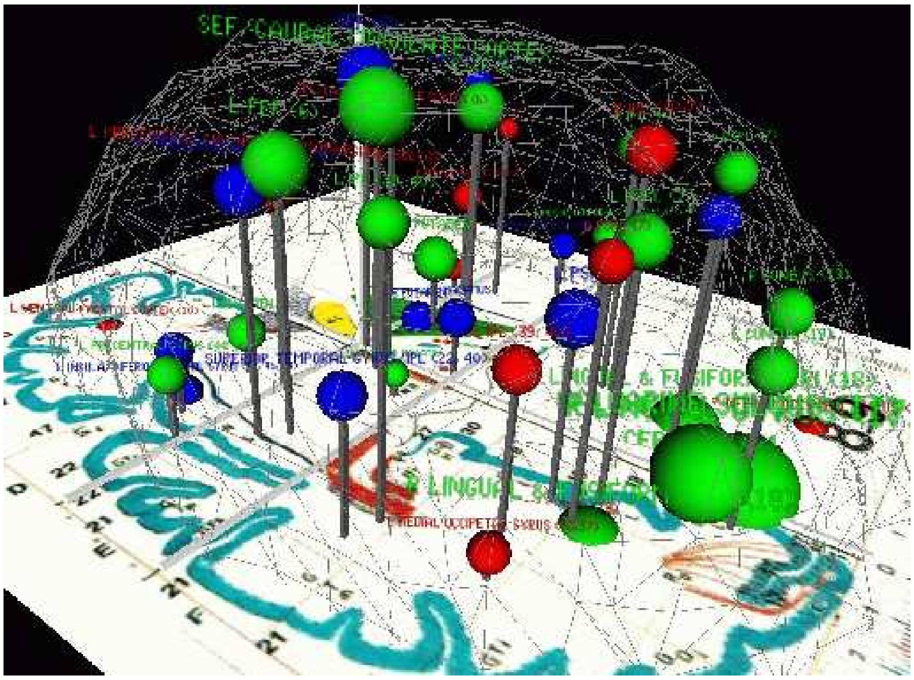

Figure 1: Locations from (Law, 1996).

- Data from BrainMap database with data from functional neuroimaging studies.
- 3 different data types + behavior
  - Paper (225): Bibliographic information
  - Experiment (771): Description of the paradigm
  - Location (7683): Talairach coordinate, anatomical and Brodmann labels (3935 location with this)
- Anatomical labels: “lateral superior parietal” $\rightarrow$ {“lateral”, “superior”, ”parietal”, “lateral superior”, “superior parietal”, “lateral superior parietal”}

MODELING THE BRAINMAP "LOCATIONS"

- Regard the "locations" as being generated from a distribution $p(\mathbf{x})$, where $\mathbf{x}$ lives in 3D Talairach space. Modeled with a mixture density:

$$\hat{p}(\mathbf{x}) = \sum_{k}^{K} w_k \tilde{p}_k(\mathbf{x} - \boldsymbol{\mu}_k) \tag{1}$$

- First try: Finite Gaussian mixture modeled with K determined by AIC and the covariance and center determined on different parts of the data set (Hansen et al., 2000).
- Second try: "kernel methods" (kernels centered on each object: $K = N$) with homogeneous Gaussian kernel (covariance $\sigma_k = \sigma_1 = \sigma_2 = \ldots = \sigma_K$) and homogeneous weights $w_k = w_1 = w_2 = \ldots = w_K = N^{-1}$:

$$\hat{p}(\mathbf{x}) = N^{-1} \sum_{n}^{N} \tilde{p}_n(\mathbf{x}) \tag{2}$$

$$\tilde{p}_n(\mathbf{x}) = (2\pi\sigma^2)^{-3/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu}_n)^2\right) \tag{3}$$

Optimize kernel width σ with leave-one-out cross-validation.

DENSITIES FOR “BEHAVIOR”

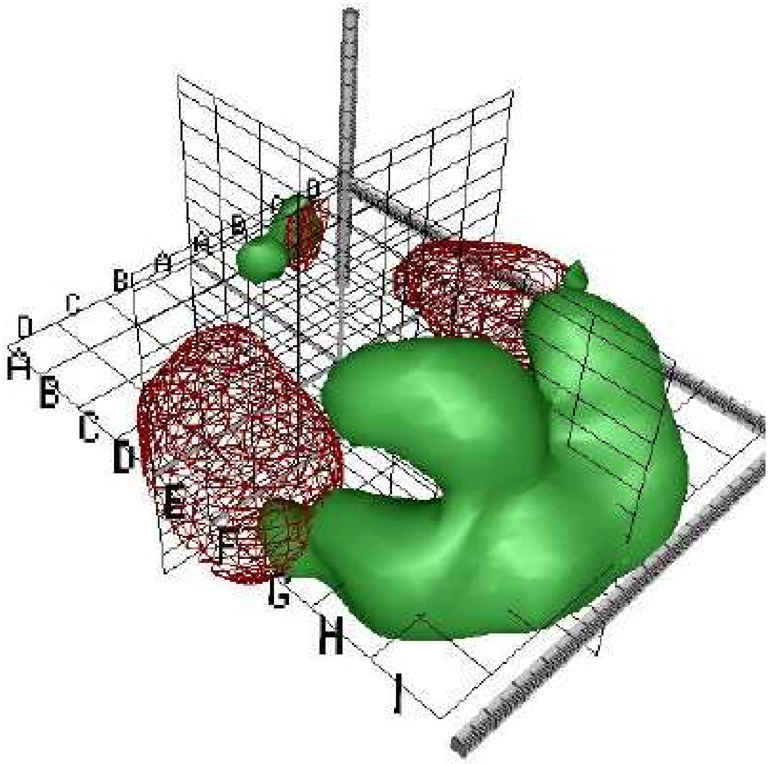

Figure 2: Densities from “Vision” and “Audition” behavioral paradigm.

- Condition on behavior: $p(\mathbf{x}|\text{behavior} = \text{"vision"})$.
- Application? Problem:
  - Behavior annotation is on the level of the experiment
  - Abstract other bibliographic information is on yet a higher level.

PDFS FOR LOCATION LABELS

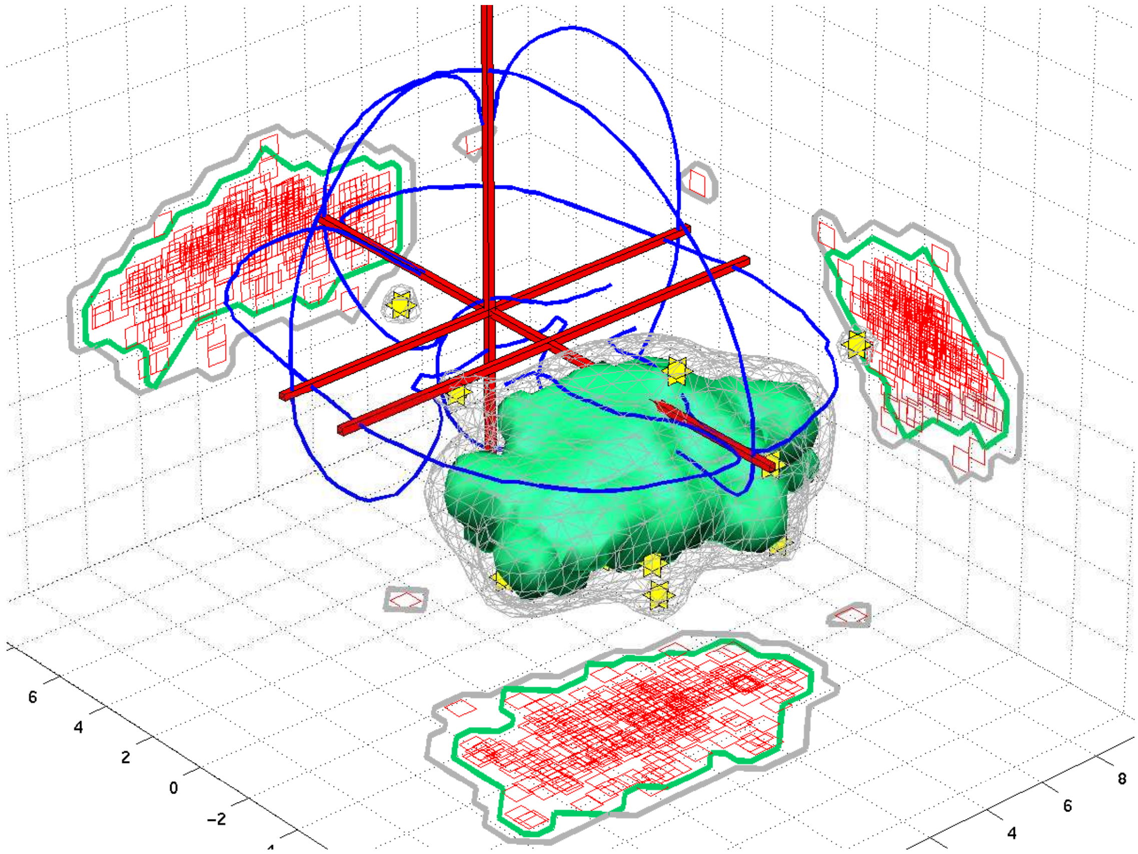

Figure 3: Densities from cerebellum locations.

- Condition on anatomical label: $p(\mathbf{x}|\text{structure} = \text{"cerebellum"})$.
- Individual labels for location.
- Novelty index from the density value.
- 2-stage heuristics: Excluding 5% most extreme

"MANUAL" INVESTIGATION OF OUTLIERS

No.	BrainMap	x	y	z	BrainMap label	Comment	Reference
1	267, 2, 1	−5	7	540	SMA	Millimeter and centimeter for z-coordinate confused during BrainMap entry	(Buckner et al., 1996, table 4, entry 1)
2	29, 10, 8	48	−23	−51	Lateral superior parietal	Resolved: Transcription mistake.	(Corbetta et al., 1993, table 5)
3	141, 1, 10	35	150	28	Dorsolateral prefrontal	Millimeter and centimeter for y-coordinate confused during BrainMap entry	(Kosslyn et al., 1994, tabel 2, entry 10)
4	249, 1, 59	−31.8	48.1	2.2	Subgyral frontal lobe	Correct	S. K. Brannan, 1997, Unpublished
5	280, 1, 9	24	−70	−24	Dorsal parietal cortex	Is labeled "Right cerebellum" in the article	(Schlösser et al., 1998, table 1, entry 9)
6	4, 2, 7	−6	42	−8	Cerebellum — superior anterior	Not possible to find the foci in the article.	(Petersen et al., 1988)
7	280, 1, 7	38	24	−8	Dorsolateral parietal	Is labeled "Right orbitofrontal cortex" in the article	(Schlösser et al., 1998, table 1, entry 7)
8	249,1,29	−2	26	16	Limbic Lobe	Correct	S. K. Brannan, 1997, Unpublished
9	277, 3, 3	−50	−42	−14	Inferior frontal gyrus, posterior	Is labeled "inferior temporal gyrus posterior (area 37)" in the article	(Owen et al., 1996, table 2, entry 3)
10	115, 2, 5	−38	54	0	Middle temporal gyrus	Not resolved.	(Shaywitz et al., 1995, page 155)
11	19,2,17	24	−47	38	Frontal	Not resolved	(Pardo et al., 1991, Table 1a, entry 17)
12	47,4,1	−36	32	28	Medial frontal lobe	Not investigated	(George et al., 1994)
13	65, 2, 23	57	26	45	Anterior cingulate	Millimeter and centimeter for x-coordinate confused during BrainMap entry.	(O'Sullivan et al., 1994, table 4, entry 10)
14	52, 1, 2	36	−46	36	Inferior frontal gyrus	Probably misunderstanding of the text during entry. The foci is around supramarginal gyrus and the denoted "BA40".	(Becker et al., 1994, page 287)
15	61, 1, 12	−24	42	4	Temporal/insular	Resolved: Transcription mistake.	(Tulving et al., 1994, table 1)

Table 1: BrainMap outliers. The entries are ordered according to novelty. The second column indicates the paper, experiment and location identifier of the BrainMap database. The third to fifth column are x, y and z with the "reported" coordinates from BrainMap (not the corrected "Talairach 1988" coordinates).

COMBINING TERM AND LINK ANALYSIS

- Working with data from “NeuroImage”: 325 web-pages which were abstracts with references (not full text: usually not allowed by publisher)
- Extracted with perl/matlab: Title (words in title), names of the authors cited, . . .
- Examples of preprocessing:
  - Names: Surname + one initial. 11 variations of “Richard S. J. Frackowiak”.
  - Journal titles: “3-stemming”, e.g., “Hum Bra Map”, except for certain prefixes “neuropsy-” 14 variation of “Human Brain Mapping”.
- Example on matrices:
  - $\mathbf{X}(N \times P)$: Document × author matrix: “Links”.
  - $\mathbf{Y}(N \times Q)$: Document × term matrix

AUTHOR COCITATION ANALYSIS

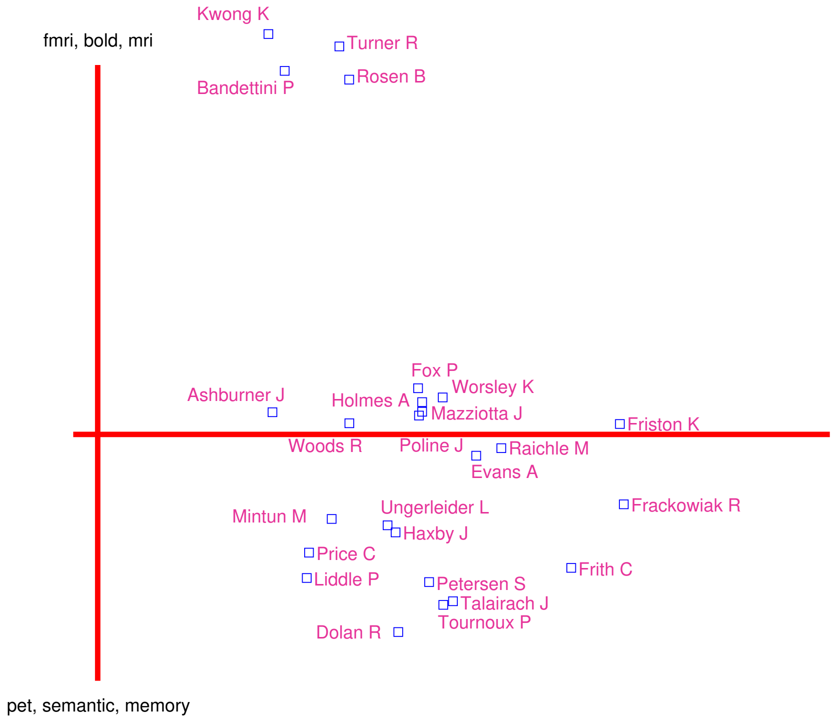

Figure 4: First and second principal component with most cited authors.

- $\mathbf{X}(N \times P)$: Document $\times$ cited author matrix, i.e., the extracted reference list for each article. $N = 325$ documents, $P = 11558$ cited authors.
- "Author cocitation analysis" (McCain, 1990; White and Griffith, 1981). Cluster analysis, multidimensional scaling, here SVD:
$$\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{V}^{\mathsf{T}}. \tag{4}$$
- First principal component (Kleinberg, 1997): $\mathbf{u}_1$: Hub documents (cites many authorities), $\mathbf{v}_1$: Authoritative author (is cited by many hubs)

AXIS LABELING IN ACA

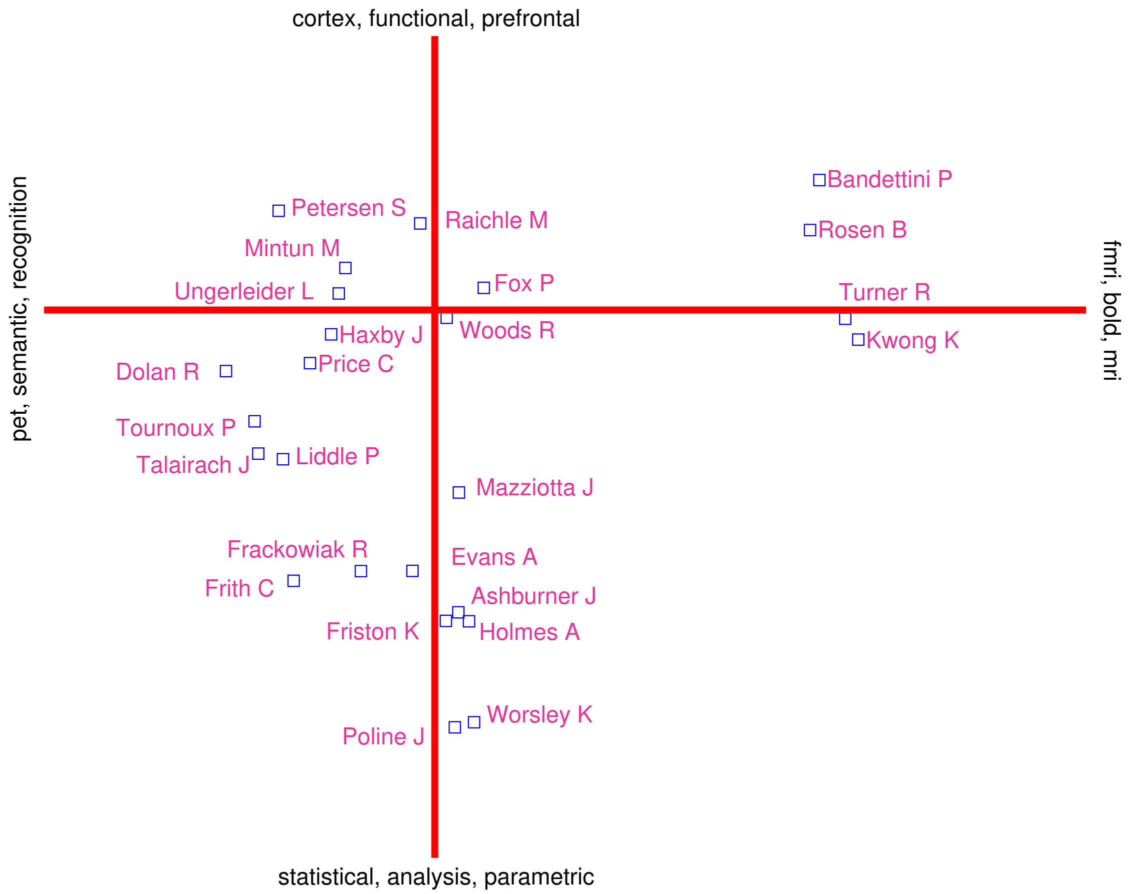

Figure 5: Second and fourth principal component with most cited authors.

- $\mathbf{Y}(N \times Q)$: Document $\times$ term matrix. Terms from the title of the article.
- Project the terms on the eigenauthor space:

$$\mathbf{Z} = \mathbf{Y}^{\mathsf{T}}\mathbf{U} \tag{5}$$

- Axis label for the ith eigenauthor is the (three) largest elements with the highest score in the ith column of $\mathbf{Z}$.

CLUSTERING OR SOMETHING ...

1	2	3	4	5	6
Raichle M	Rosen B	Friston K	Haxby J	Woods R	Roland P
Petersen S	Turner R	Frackowiak R	Ungerleider L	Passingham R	Cherry S
Buckner R	Bandettini P	Frith C	Mintun M	Rao S	Mesulam M
Grasby P	Kwong K	Evans A	Petrides M	Brooks D	Toga A
Grady C	Glover G	Talairach J	D'Esposito M	Mazoyer B	Zatorre R
Miezin F	Jezzard P	Worsley K	Gore J	Binder J	Zilles K
Fletcher P	Belliveau J	Tournoux P	Noll D	Fink G	Van Essen D
Horwitz B	Kim S	Mazziotta J	Cohen J	Grafton S	Weinberger D
Posner M	Hyde J	Poline J	Goldman-Rakic P	Deiber M	Arndt S
Maisog J	Ugurbil K	Holmes A	Meyer E	Hammeke T	Gazzaniga M
Wise R	Kennedy D	Fox P	McCarthy G	Frost J	Lancaster J
Houle S	Weisskoff R	Dolan R	Koeppe R	Hallett M	Ledberg A
Fiez J	Cohen M	Liddle P	Aguirre G	Weiller C	Schlaug G
Shallice T	Wong E	Ashburner J	Smith E	Kleinschmidt A	Steinmetz H
Alpert N	Josephs O	Price C	Minoshima S	Sadato N	Eriksson L
Herscovitch P	Menon R	Heather J	Corbetta M	Silbersweig D	Matelli M
Kapur S	Ogawa S	Neelin P	Jonides J	Kawashima R	Pandya D
Tulving E	Williams S	Watson J	Nobre A	Seitz R	Galaburda A
Paulesu E	Brady T	Marrett S	Shulman G	Tzourio N	Meyer J
Rapoport S	Dale A	Phelps M	Paus T	Allison T	Peters T
McIntosh A	Jesmanowicz A	Carson R	Zarahn E	Nixon P	Schleicher A

Figure 6: NMF of document-cited author matrix.

- SVD axes
- Non-negative matrix factorization (NMF) (Lee and Seung, 2001)
- Clusters in this data set
  - Clear fMRI cluster: Bruce Rosen, Robert Turner; "fMRI", "BOLD", "MRI".
  - Haxby, Glover: "Memory", "prefrontal", "working" opposing "motor", "human", "activation" with Mazziotta and Roland.
  - Methodological: Karl J. Friston, Keith J. Worsley; "Statistical", "Analysis"

JOURNAL COCITATION ANALYSIS

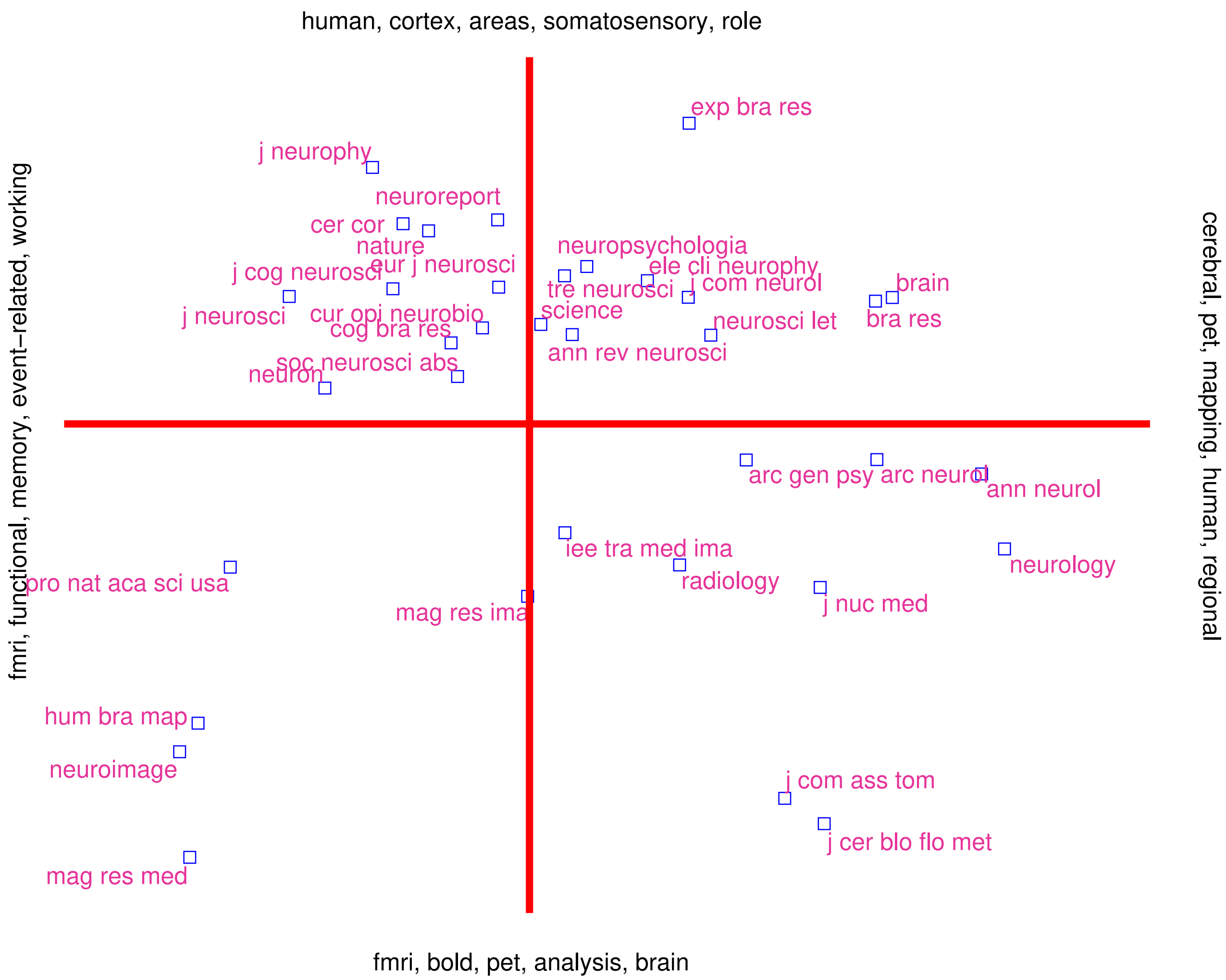

Figure 7: Second and third principal component with most cited journals from NeuroImage.

- $\mathbf{X}(N \times P)$: Document $\times$ cited journal matrix.
- Cocitation analysis (Carpenter and Narin, 1973)
- Clusters: functional neuroimaging, “doctors quadrant” (neurology, radiology, nuclear medicin, psychiatry?), psychology, “neuroscience”.

CONTEXT SENSITIVE AUTHORITIES

Non-context	Nielsen(fmri)	Hansen(fmri)
Frackowiak R	Friston K	Friston K
Friston K	Frackowiak R	Frackowiak R
Frith C	Turner R	Turner R
Raichle M	Rosen B	Rosen B
Evans A	Raichle M	Frith C
Talairach J	Frith C	Holmes A
Tournoux P	Holmes A	Raichle M
Worsley K	Bandettini P	Bandettini P
Petersen S	Kwong K	Worsley K
Mazziotta J	Worsley K	Poline J
Holmes A	Poline J	Evans A

- Singular value decomposition of the term matrix (latent sematic analysis) (Deerwester et al., 1990):

$$\mathbf{Y} = \mathbf{U}\mathbf{L}\mathbf{V}^{\mathsf{T}} \tag{6}$$

- Project a term vector $\mathbf{v}$ onto $\mathbf{y}$-space with a suitable generalizable subspace:

$$\mathbf{y} = \mathbf{U}f(\mathbf{L})\mathbf{v}^{\mathsf{T}} \tag{7}$$

- Aplication the document weighting on the authors:
  - Either with an extra SVD ("Nielsen"):

$$\mathbf{X}^* = \mathbf{y}\mathbf{1}^{\mathsf{T}} \odot \mathbf{X} \tag{8}$$

    ...and SVD on $\mathbf{X}^*$.
  - Or a direct projection ("Hansen"):

$$\mathbf{a} = \mathbf{X}^{\mathsf{T}}\mathbf{y} \tag{9}$$

VISUALIZATION OF LINKS

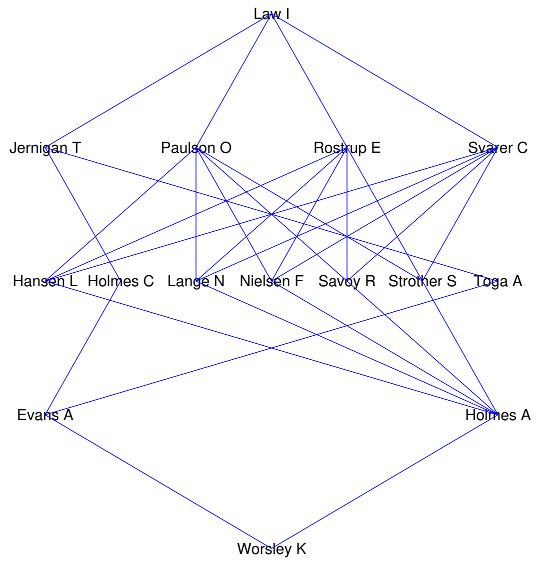

Figure 8: Co-author path between Keith J. Worsley and Ian Law.

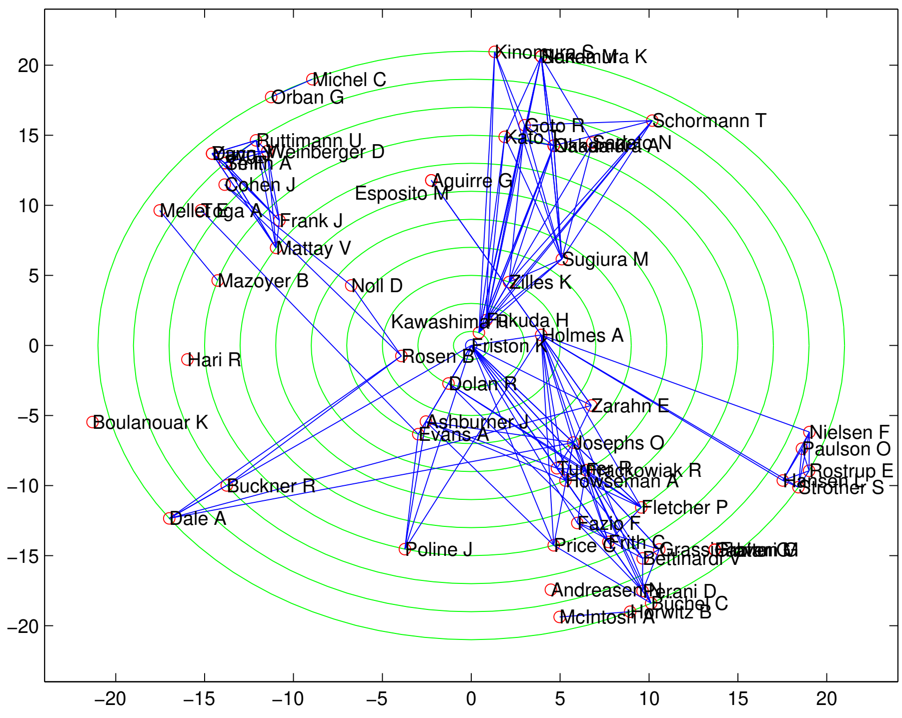

Figure 9: Co-authorship "targetgram" inspired from group psychology.

OTHER IDEAS

- "Find related volumes": Query with a list of locations or a voxel-volume, return similar BrainMap experiments.
- Perform canonical (correlation) analysis between $\mathbf{X}$ (BrainMap paper $\times$ word terms) matrix and $\mathbf{Y}$ (BrainMap paper $\times$ voxels = voxelization of locations) matrix.
- Asymmetries between the hemispheres.
- Canonical (correlation) analysis between term and link matrix.

References

Becker, J. T., Mintun, M. A., Diehl, D. J., Dobkin, J., Martidis, A., Madoff, D. C., and Dekosky, S. T. (1994). Functional neuroanatomy of verbal free recall: A replication study. *Human Brain Mapping*, 1(4):284–292.

Buckner, R. L., Raichle, M. E., Miezin, F. M., and Petersen, S. E. (1996). Functional anatomic studies of memory retrieval for auditory words and visual pictures. *The Journal of Neuroscience*, 16(19):6219–6235.

Carpenter, M. P. and Narin, F. (1973). Clustering of scientific journals. *Journal of the American Society for Information Science*, 24(6).

Corbetta, M., Miezin, F. M., Shulman, G. L., and Petersen, S. E. (1993). A PET study of visuospatial attention. *The Journal of Neuroscience*, 13(3):1202–1226.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

George, M. S., Ketter, T. A., Parekh, B. A., Gill, D. S., Huggins, T., Marangell, L., Pazaglia, P. J., and Post, R. (1994). Spatial ability in affective illness: differences in regional brain activation during a spatial matching task ($H_2^{15}O$ PET). *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 7:143–153.

Hansen, L. K., Sigurdsson, S., Kolenda, T., Nielsen, F. Å., Kjems, U., and Larsen, J. (2000). Modeling text with generalizable Gaussian mixtures. In *Proceedings of ICASSP'2000*. Turkey.

Kleinberg, J. M. (1997). Authoritative sources in a hyperlinked environment. Technical Report 10076(91892), IBM.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L., and Anderson, A. K. (1994). Identifying objects seen from different viewpoints, a PET investigation. *Brain*, 117:1055–1071.

Law, I. (1996). *Saccadic Eye Movements, A Functional Brain Mapping Approach*. PhD thesis, Department of Neurology, Rigshospitalet, The University of Copenhagen, Copenhagen, Denmark.

Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization.

McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, 41(6):433–443.

O'Sullivan, B. T., Roland, P. E., and Kawashima, R. (1994). A PET study of somatosensory discrimination in man. Microgeometry versus macrogeometry. *European Journal of Neuroscience*, 6(1):137–148.

Owen, A. M., Milner, B., Petrides, M., and Evans, A. C. (1996). Memory for object features versus memory for object location: A positron-emission tomography study of encoding and retrieval processes. *Proceedings of the National Academy of Sciences USA*, 93(17):9212–9217.

Pardo, J. V., Raichle, M. E., and Fox, P. T. (1991). Localization of a human system for sustained attention by positron emission tomography. *Nature*, 349(6304):61–63.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331:585–589.

Schlösser, R., Hutchinson, M., Joseffer, S., Rusinek, H., Saarimaki, A., Stevenson, J., Dewey, S. L., and Brodie, J. D. (1998). Functional magnetic resonance imaging of human brain activity in a verbal fluency task. *Journal of Neurology, Neurosurgery, and Psychiatry*, 64(4):492–495.

Shaywitz, B. A., Pugh, K. R., Constable, R. T., Shaywitz, S. E., Bronen, R. A., Fulgright, R. K., Shankweiler, D. P., Katz, L., Fletcher, J. M., Skudlarski, P., and Gore, J. C. (1995). Localization of semantic processing using functional magnetic resonance imaging. *Human Brain Mapping*, 2:149–158.

Talairach, J. and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publisher Inc, New York.

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I. M., Habib, R., and Houle, S. (1994). Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proceedings of the National Academy of Sciences of the USA*, 91:2012–2015.

White, H. D. and Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32:163–171.