Soft clustering of age-related cortical features

ANGELA LOMBARDI, PH.D. INFN BARI

SABINA TANGARO, PH.D. INFN BARI

Summary

- Motivations
- Framework
- Future developments

Motivations

- ❑Properties of MRI scanners such as field strength, manufacturer, gradient nonlinearity, subject positioning, and longitudinal drift have been long understood to increase bias and variance in the measurement of brain volume changes, regional cortical thickness, voxel-based morphometry, and structural, functional, and diffusion images in general.

- ❑**Non-biological confounders** typically have a priori **unpredictable effects**, thus compromising consistency and reproducibility of the downstream analyses across studies.

- ❑Harmonization techniques are required to remove scanner and site effects, while **preserving the variability associated with biology**.

Motivations

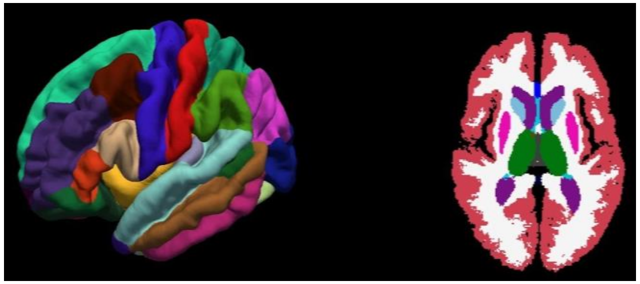
- In the **supervised classification of human connectome** data, subjects are usually grouped based on high-level clinical categories (e.g., patients and controls, age matched groups, etc.), and typical approaches aim at deducing a decision function from the labeled training data.
- **Unsupervised analysis is blind to any phenotypic factors** and aims at finding subgroups of subjects/features with similar characteristics.
- In general the population of healthy subjects is usually **highly heterogeneous so stratification** of groups may be a useful preprocessing stage, in order to exploit the knowledge of the structure of data.

Framework

STATISTICAL HARMONIZATION	COMPLEX NETWORK MODELING	SOFT CLUSTERING
Removes unwanted variation associated with site and preserves biological associations in the data	Create associations between each couple of subjects of the dataset (multivariate approach).	Stratification of groups with homogeneous features.

Dataset

383 MALE CONTROL subjects from 17 sites (ABIDE DATASET)
Age range 7-43; mean=17; std=7;



DESIKAN ATLAS
31 ROIs for hemisphere

372 CORTICAL FEATURES:

- **SurfArea (mm^2)**
- **GrayVol (mm^3)**
- **ThickAvg (mm)**
- **ThickStd (mm)**
- **MeanCurv (mm^{-1})**
- **CurvInd**

39 SUBCORTICAL FEATURES
9 GLOBAL FEATURES

Di Martino, Adriana, et al. "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism." *Molecular psychiatry* 19.6 (2014): 659.

Statistical Harmonization

y_{ijv} n x 1 vector of FS features for imaging site i, for participant j and feature v, for a total of (k + 1) sites, n participants and V features.

X p × n matrix of biological covariates of interests

Z k × n matrix of site indicators

$$y_{ijv} = \alpha_v + X_{ij}^T \beta_v + Z_{ij}^T \theta_v + \delta_{iv} \varepsilon_{ijv}$$

α_v is the average feature for the reference site for feature v;
θ_v is the k × 1 vector of the coefficients associated with the site indicators Z for feature v
β_v is the p × 1 vector of coefficients associated with X for feature v
δ_{iv} descrive the multiplicative site effect of the j-th site on the feature v.

Fortin, Jean-Philippe, et al. "Harmonization of multi-site diffusion tensor imaging data." *Neuroimage* 161 (2017): 149-170.

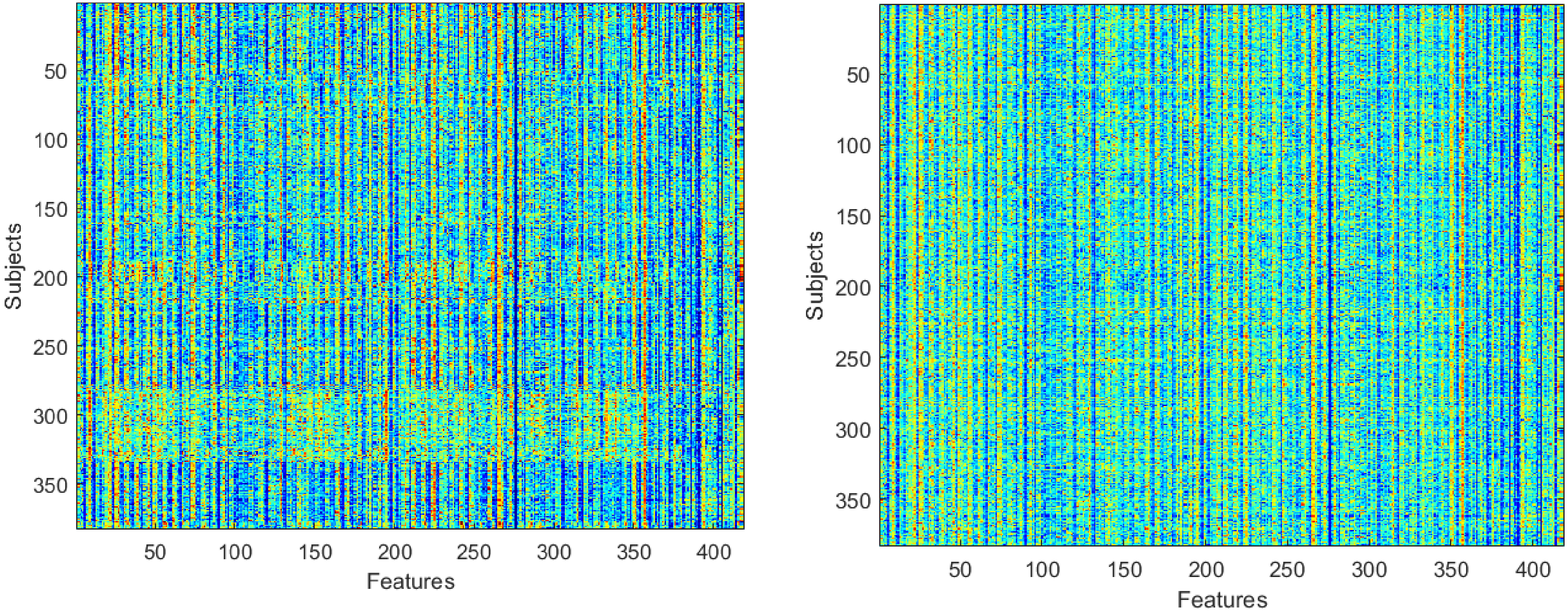
Statistical Harmonization

(1) it models **site-specific scaling factors**
(2) it uses **empirical Bayes** to improve the estimation of the site parameters.

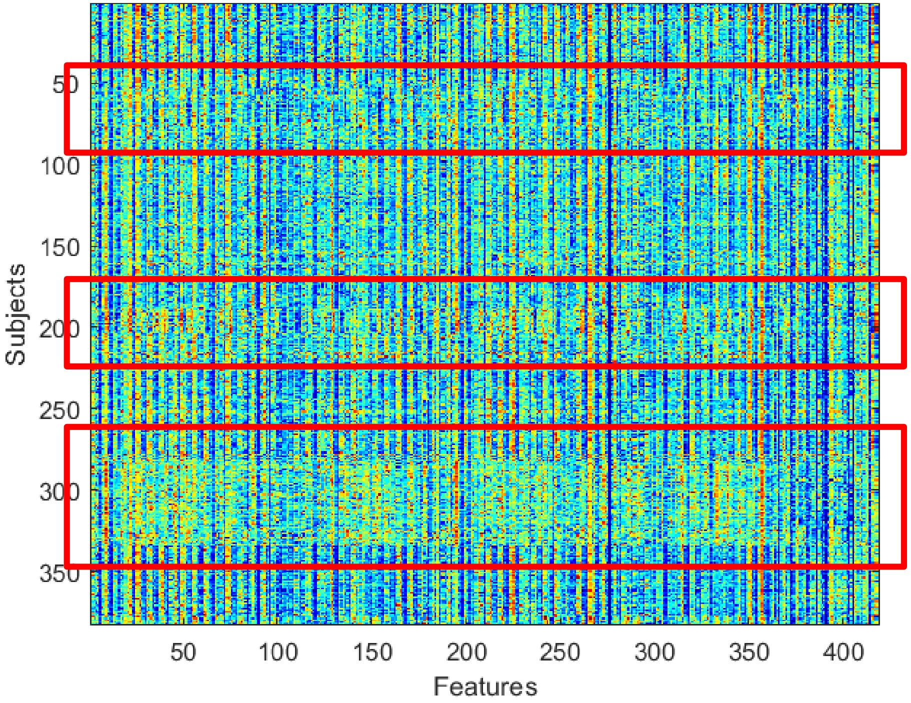
It posits a unique linear model of location and scale at each feature, making the assumption that scanners (or sites) have both an additive and multiplicative effects on the data.

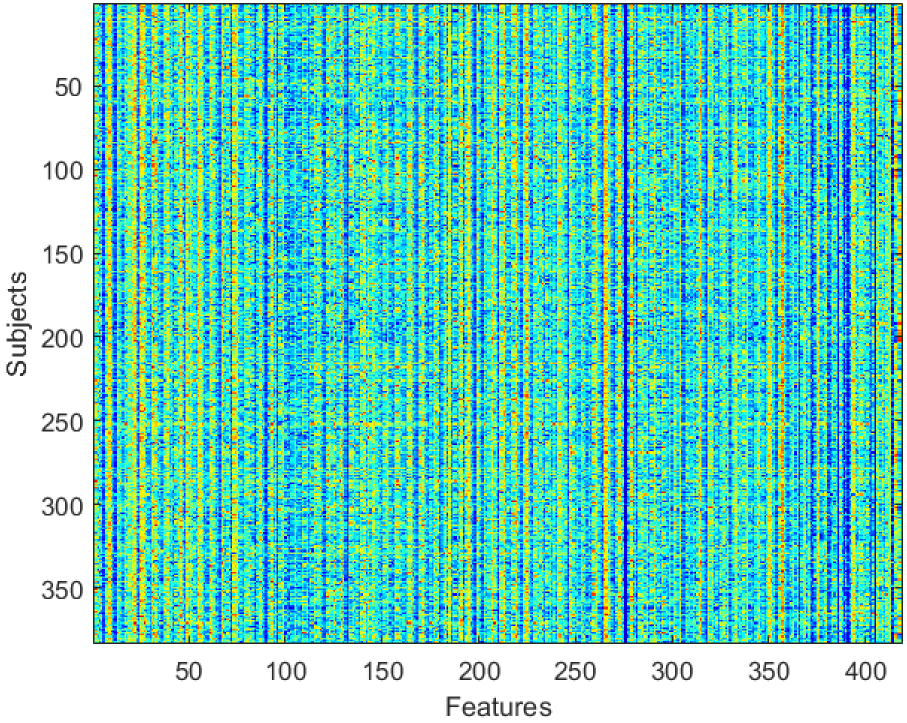
- The model assumes that the expected values of the imaging feature measurements can be modeled as **a linear combination of the biological variables and the site effects**, whose error term is modulated by additional site-specific scaling factors.

Statistical Harmonization

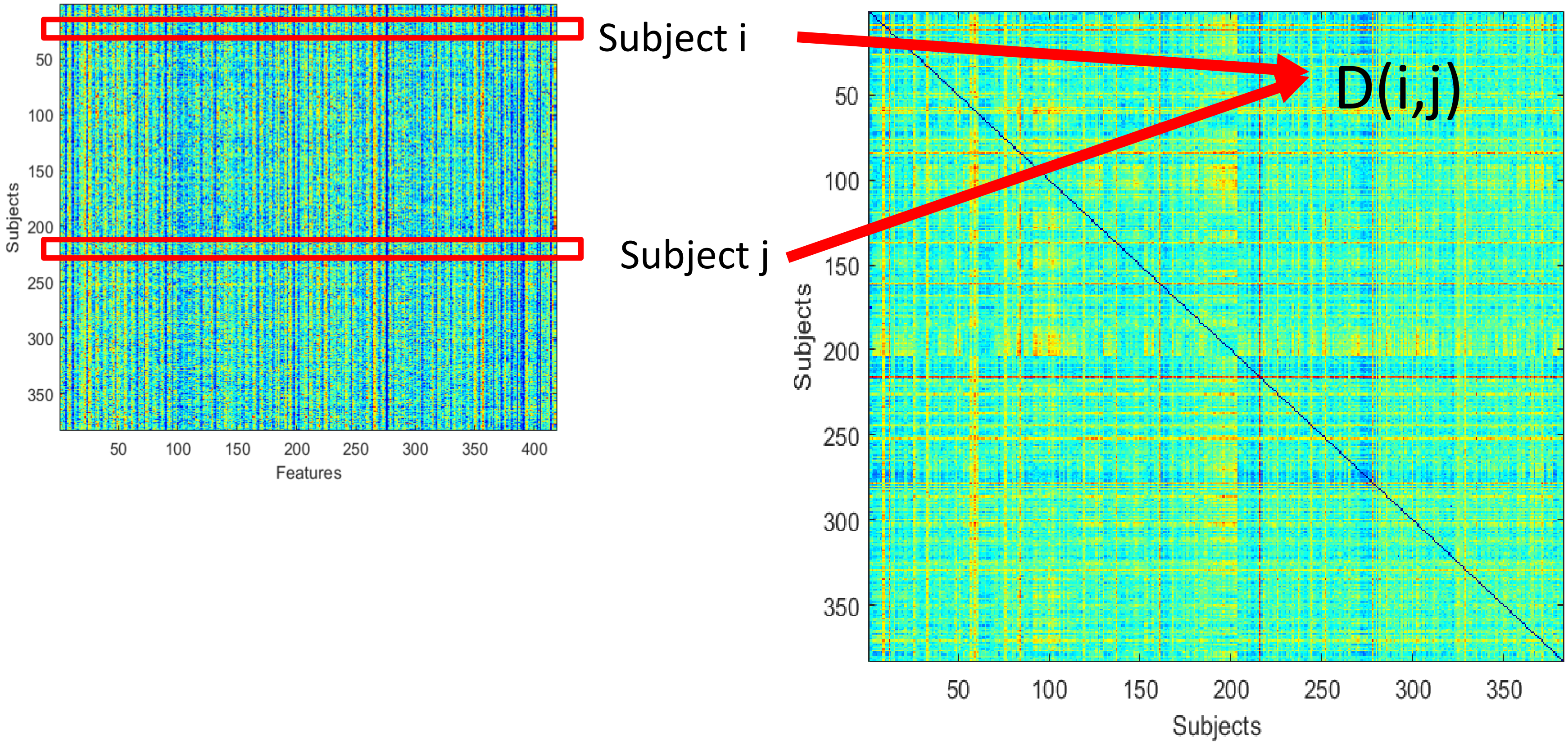


Statistical Harmonization





Complex Network modeling



Subject i

Subject j

D(i,j)

Community Detection

- Communities in networks often overlap such that nodes simultaneously belong to several groups. Meanwhile, many networks are known to possess hierarchical organization, where communities are recursively grouped into a hierarchical structure. However, the fact that many real networks have **communities with pervasive overlap, where each and every node belongs to more than one group**, has the consequence that a global hierarchy of nodes cannot capture the relationships between overlapping groups.

- Link community detection reinvent communities as groups of links rather than nodes. In contrast to the existing literature, which has entirely focused on grouping nodes, link communities naturally incorporate overlap while revealing hierarchical organization.

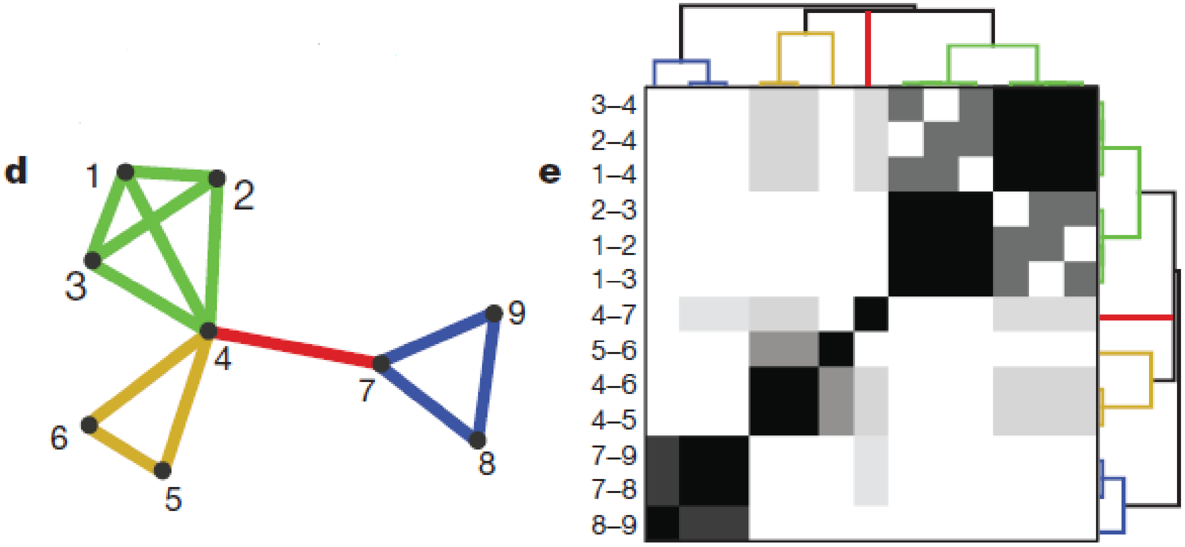
Community Detection

1.Hierarchical clustering with a similarity between links is used to build a dendrogram where each leaf is a link from the original network and branches represent link communities.

2.In this dendrogram, links occupy unique positions whereas nodes naturally occupy multiple positions, owing to their links. We extract link communities at multiple levels by cutting this dendrogram at various thresholds.

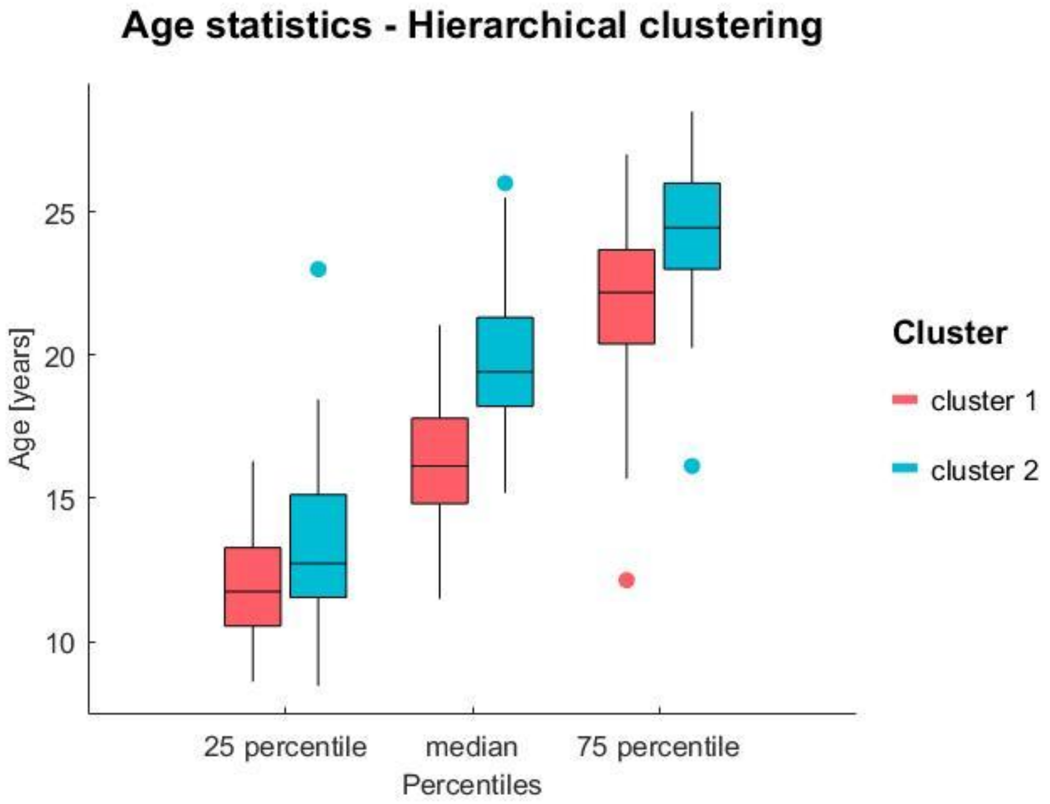
3.In order to obtain the most relevant communities it is necessary to determine the best level at which to cut the tree. For this purpose, a natural objective function, the partition density, D, based on link density inside communities is used; unlike modularity, D does not suffer from a resolution limit.

Community Detection



Ahn, Yong-Yeol, James P. Bagrow, and Sune Lehmann. "Link communities reveal multiscale complexity in networks." *nature*466.7307 (2010): 761.

Age-related soft clusters



Future developments

1. Features' optimization
2. Analysis of age-related trajectories of selected features
3. Application to neurodevelopmental disorders

Questions?

angela.lombardi@ba.infn.it

sonia.tangaro@ba.infn.it