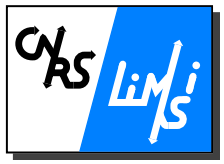
COMMUNITY-DRIVEN HIERARCHICAL FUSION OF NUMEROUS CLASSIFIERS

APPLICATION TO VIDEO SEMANTIC INDEXING

Hervé BREDIN



perso.limsi.fr/bredin
bredin@limsi.fr

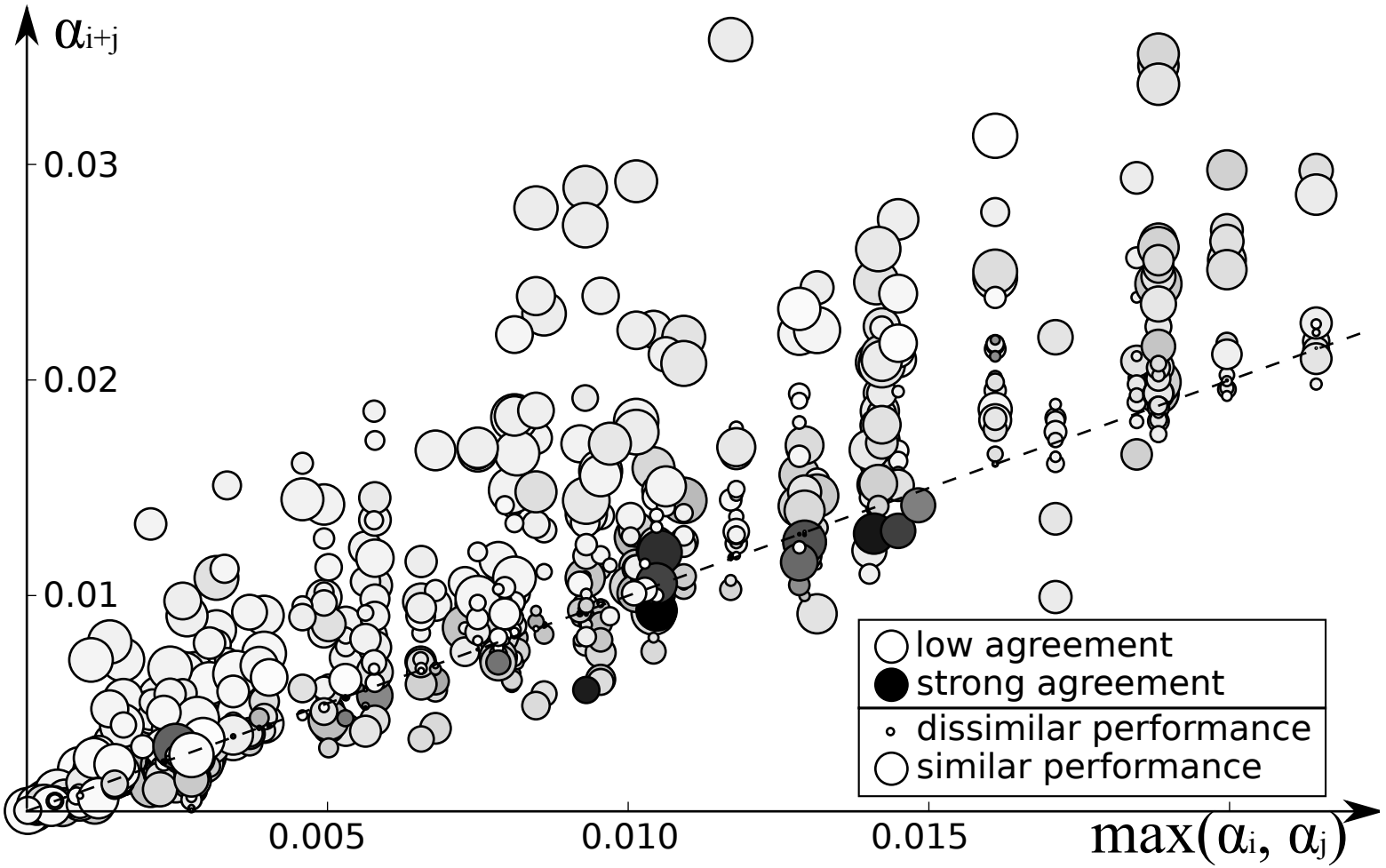
Late fusion (aka score fusion)

- N: number of shots in test set
- K: number of classifiers
  - x_{kn} : score returned by classifier k for shot n, the higher, the more likely to contain the concept.
  - α_k : performance obtained by classifier k, (inferred) average precision
- Performance-driven weighted score fusion

$$f(\mathbf{x}_n) = \sum_k \alpha_k \cdot x_{kn}$$

$$f(\mathbf{x}) = \alpha_i \cdot \mathbf{x}_i + \alpha_j \cdot \mathbf{x}_j$$

« Does combining two classifiers always yield better results than the two of them taken separately? »



Agreement between classifiers

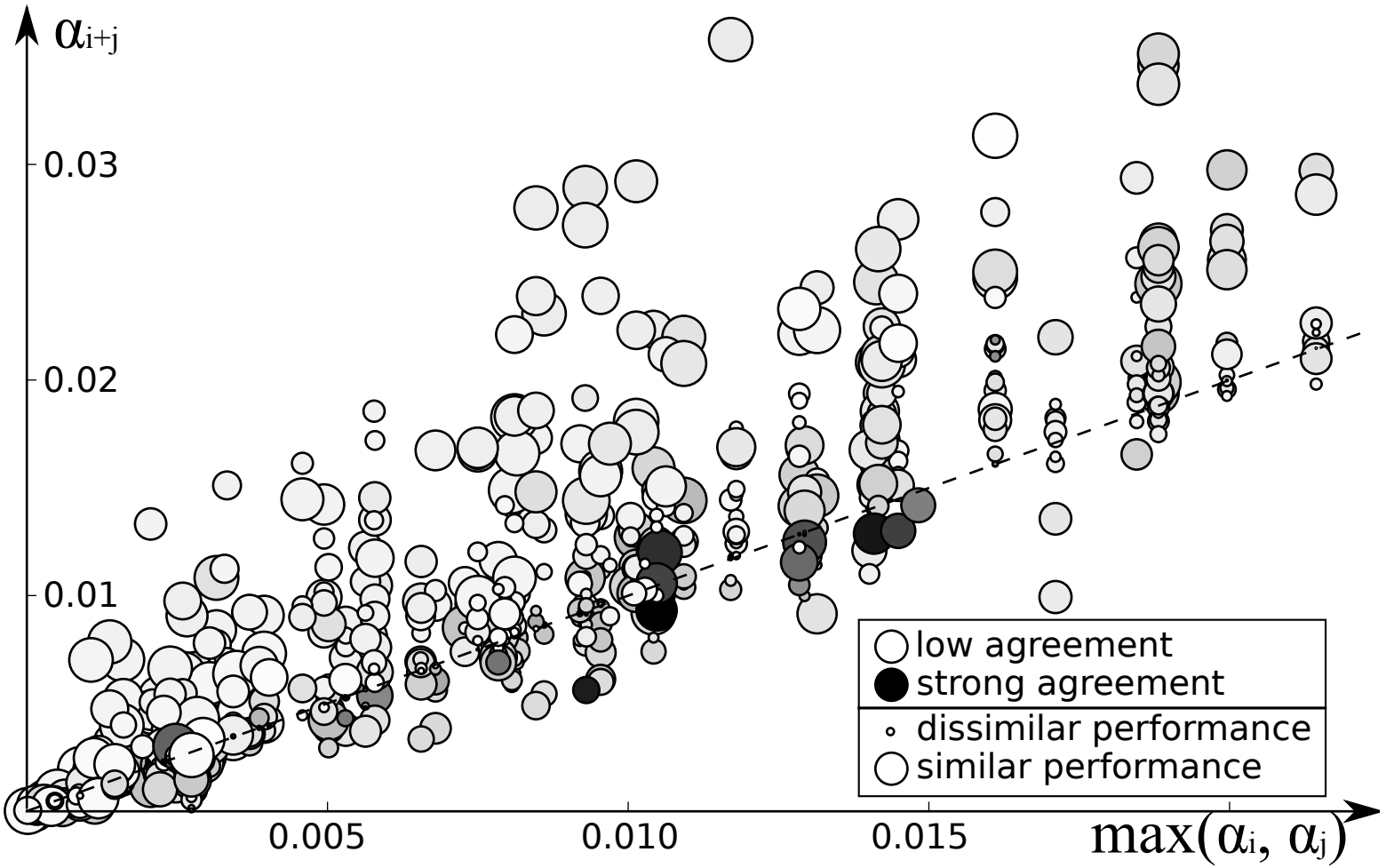
- Spearman rank correlation coefficient

$$\rho_{ij} = \frac{\sum_{n=1}^{n=\mathrm{N}} \left(r_{in} - \overline{r_i}\right)\left(r_{jn} - \overline{r_j}\right)}{\sqrt{\sum_{n=1}^{n=\mathrm{N}} \left(r_{in} - \overline{r_i}\right)^2 \sum_{n=1}^{n=\mathrm{N}} \left(r_{jn} - \overline{r_j}\right)^2}}$$

- $\rho_{ij} = 0$: classifiers are independent
- $\rho_{ij} = 1$: rankings are identical

$$f(\mathbf{x}) = \alpha_i \cdot \mathbf{x}_i + \alpha_j \cdot \mathbf{x}_j$$

« Does combining two classifiers always yield better results than the two of them taken separately? »



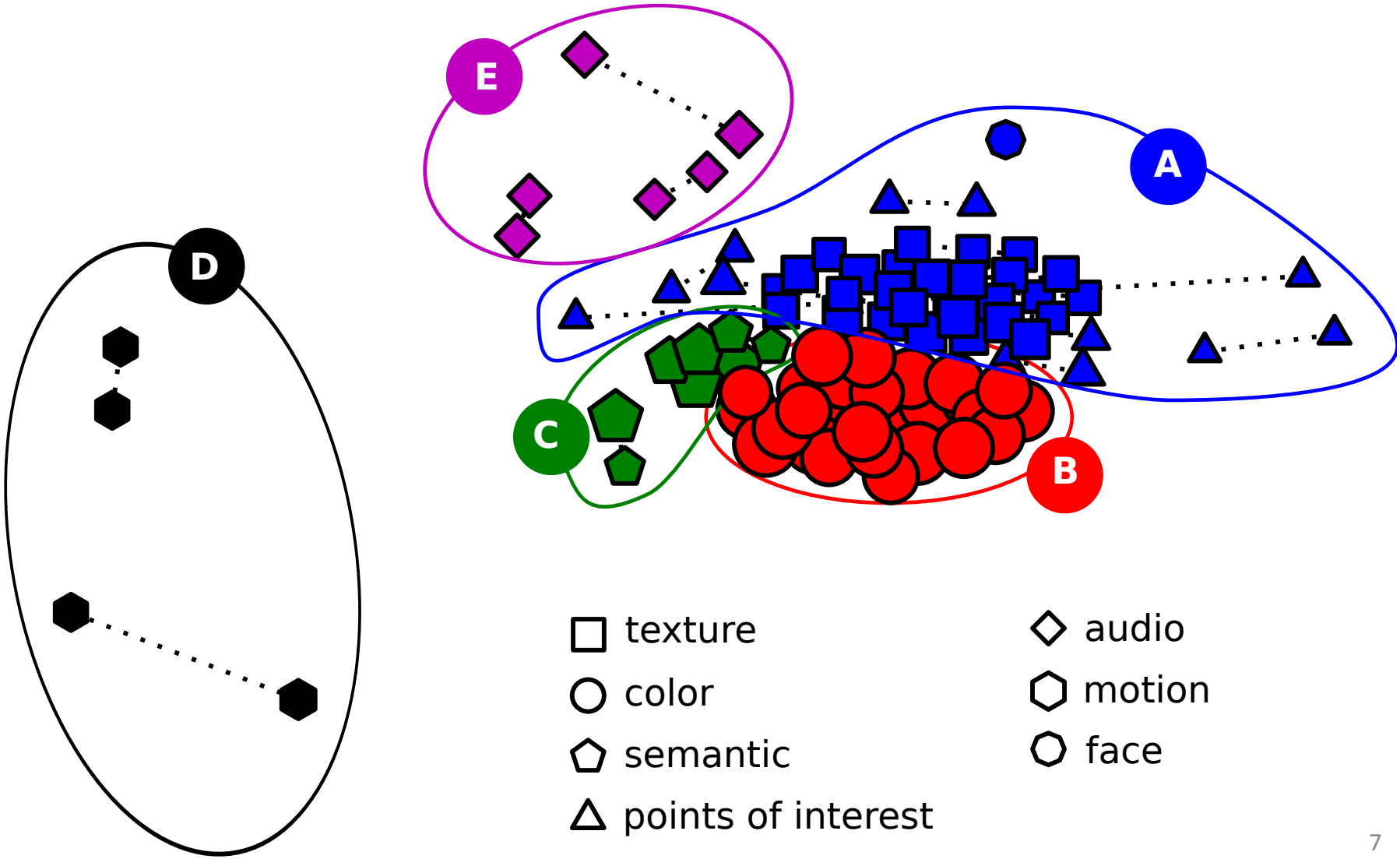
Communities of classifiers

- Graph of classifiers
  - One node per classifier
  - Complete undirected graph
  - Weights of edge (i, j)

$$A_{ij} = \max(0, \rho_{ij})$$

- Automatic community detection
  - Maximization of graph modularity
  - Louvain approach (Blondel et al., 2008)

Communities of classifiers



Community-driven hierarchical fusion

- Step 1 – community detection

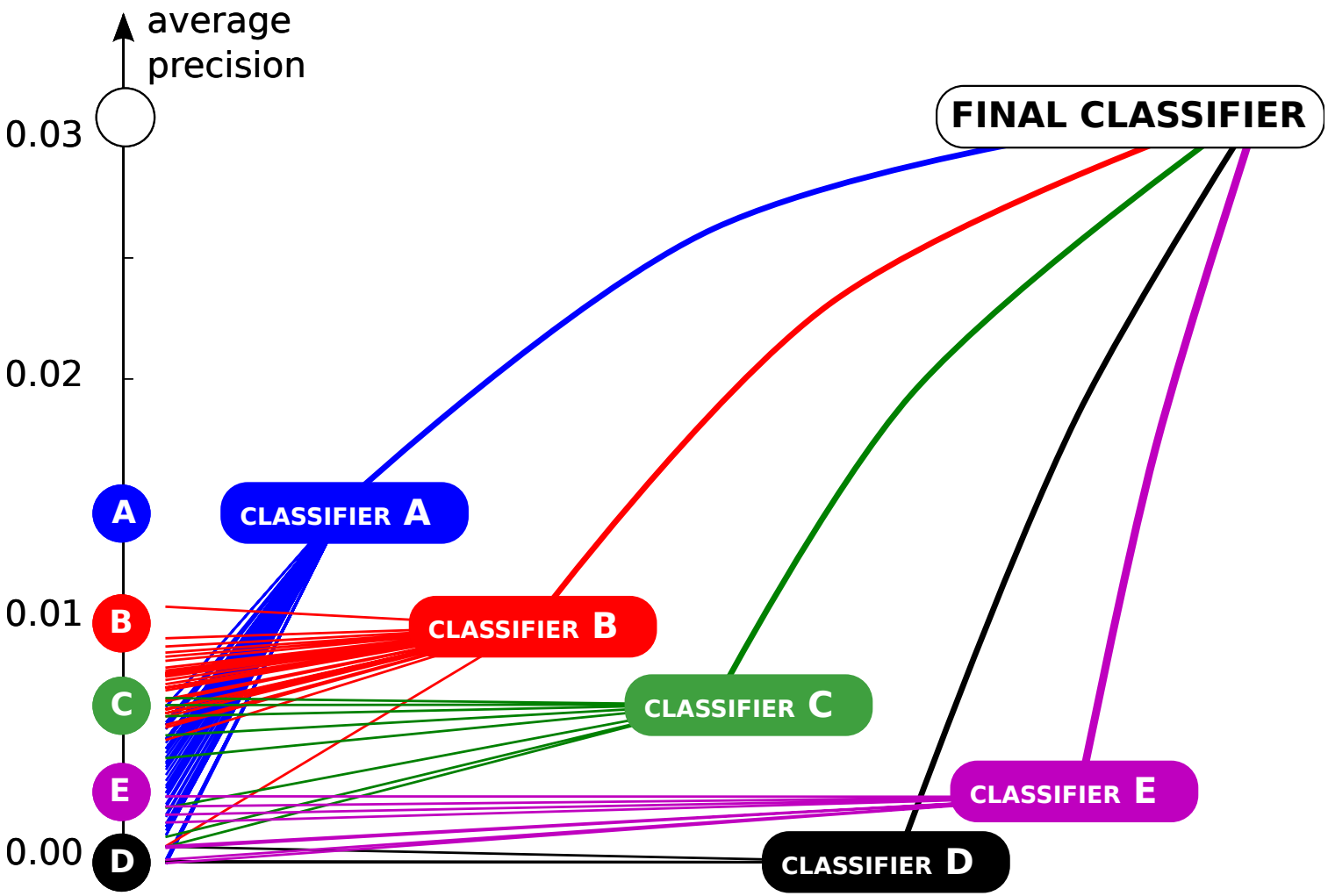
- Step 2 – intra-community fusion

$$\mathbf{x}_c = \sum_{k=1}^{k=\mathtt{K}} \delta_c(k)\,\widehat{\mathbf{x}_k}$$

- Step 3 – inter-community fusion

$$\mathbf{x} = \sum_{c=1}^{c=\mathtt{C}} \alpha_c \widehat{\mathbf{x}_c}$$

Community-driven hierarchical fusion

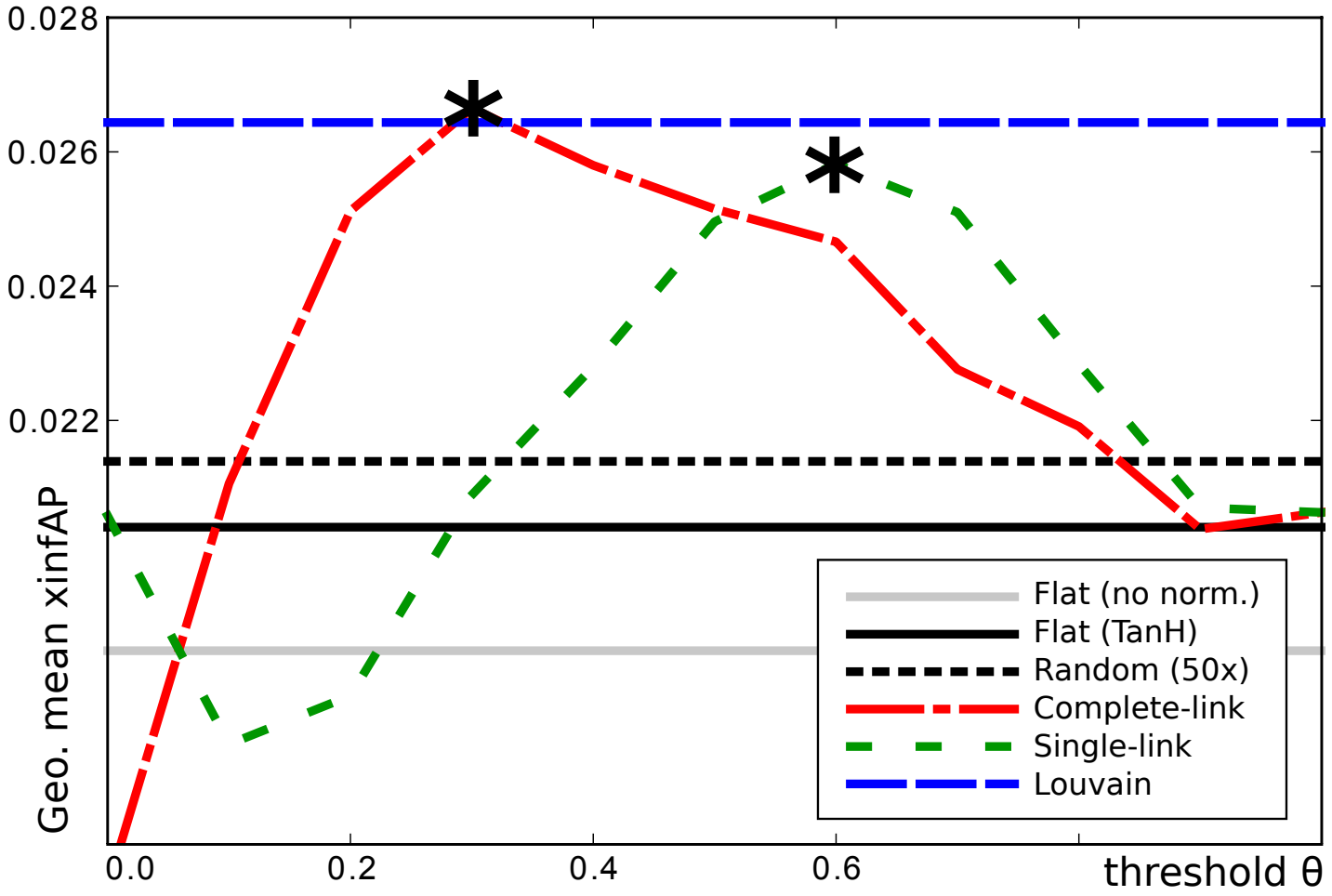


Experiments on TRECVid 2010

- Baseline – flat fusion ($f(\mathbf{x}_n) = \sum_k \alpha_k \cdot x_{kn}$)
- Contrastive hierarchical approaches
  - Random communities
  - Agglomerative clustering

Fusion	**Ari. mean xinfAP**	**Geo. mean xinfAP**
Flat (no norm.)	0.0595 (−3%)	0.0186 (−9%)
Flat (TanH)	**0.0614**	**0.0204**
Random (50×)	0.0618 (+1%)	0.0214 (+5%)
Complete-link*	0.0679 (+11%)	**0.0266** (+31%)
Single-link*	**0.0686** (+12%)	0.0258 (+27%)
Louvain	0.0634 (+3%)	**0.0264** (+30%)

Experiments on TRECVid 2010



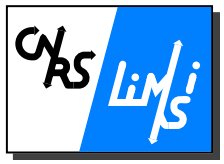
TRECVid 2011

- IRIM 4
  - Community-driven hierarchical fusion
  - MAP : **0.134**
- IRIM 1 (5th consortium/lab)
  - IRIM 4 + contextual reranking
  - MAP : **0.139**
- Quaero 1 : **0.153**
- Best MAP : **0.173** (Tokyotech/Canon)

COMMUNITY-DRIVEN HIERARCHICAL FUSION OF NUMEROUS CLASSIFIERS

APPLICATION TO VIDEO SEMANTIC INDEXING

Hervé BREDIN



perso.limsi.fr/bredin
bredin@limsi.fr

Modularity

$$Q = \frac{1}{\sum_{i,j} A_{ij}} \sum_{i,j} \left[A_{ij} - \frac{\sum_k A_{ik} \sum_k A_{kj}}{\sum_{i,j} A_{ij}} \right] \delta_{ij}$$

tanh score normalization

$$\widehat{x_{kn}} = \frac{1}{2}\left\{\tanh\left[0.01\left(\frac{x_{kn}-\mu_k}{\sigma_k}\right)\right]+1\right\}$$