Large-Scale Community Detection on Speaker Content Graphs

Stephen H. Shum, William M. Campbell, Douglas A. Reynolds

MIT CSAIL, Cambridge, MA, USA
sshum@csail.mit.edu

MIT Lincoln Laboratory, Lexington, MA, USA
{wcampbell, dar}@ll.mit.edu

I. Motivation

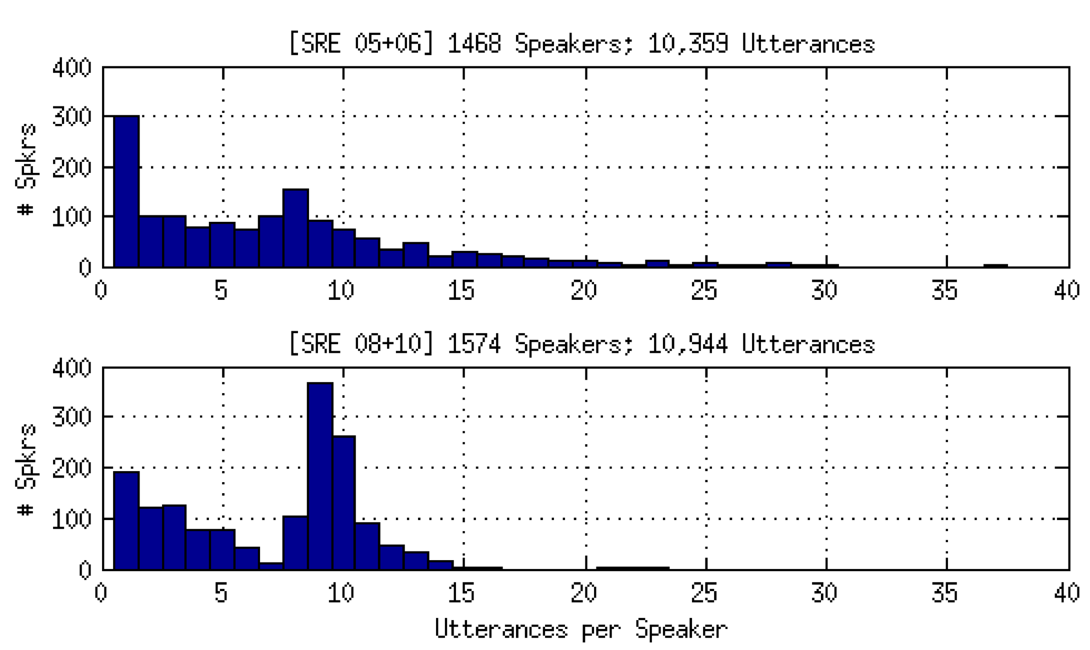
- Large collections of audio data
  - ✓ Storing and operating on O(N^2) inter-utterance relationships does not scale well as N → ∞.
  - ✓ Summarize data using a <u>sparse</u> graph, where only O(N) relationships are stored and operated upon.
  - ✓ What can we learn about the structure of the data?
- *K*-NN Speaker Content Graphs [1]
  - ✓ Nodes i, j and affinity matrix W, where

$$W_{i,j} = \begin{cases} e^{-d^2(\mathbf{m}_i,\mathbf{m}_j)/\sigma^2} & \text{if } \exists \text{ edge } (i,j) \\ 0 & \text{otherwise.} \end{cases}$$

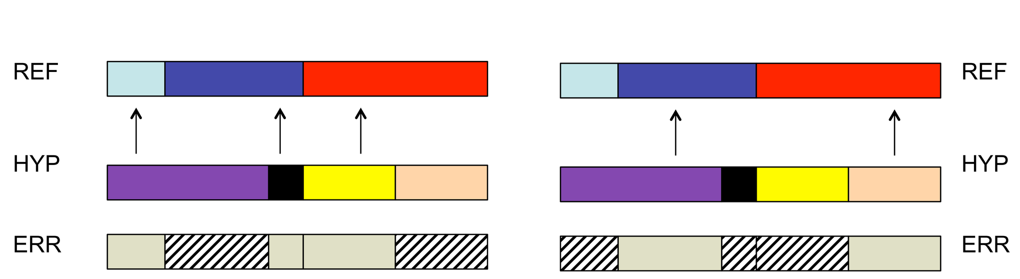
  and $d(\mathbf{m}_i, \mathbf{m}_j)$ corresponds to the Euclidean distance between two speaker GMM supervectors, $\mathbf{m}_i$ and $\mathbf{m}_j$.
  - ✓ Connect **m** with only its top-*K* nearest neighbors.

$$K \in \{2, 5, 10, 25, 50, 100\}$$

II. Experiment Data & Details



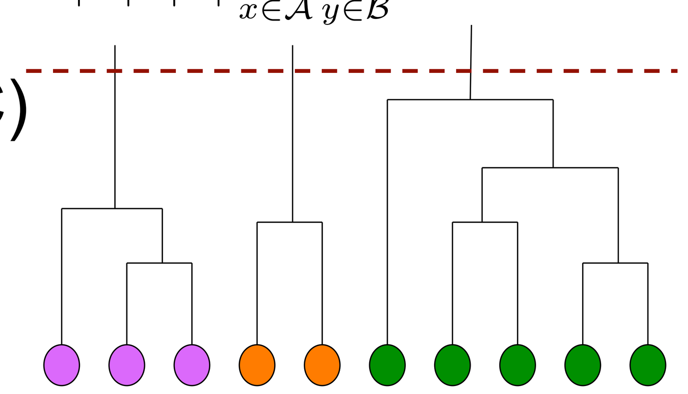
- Evaluation Protocol
  - ✓ Enforce a one-to-one mapping between reference and hypothesized clusters.
  - ✓ Greedily pick the mapping that results in smallest amount of clustering error.
  - ✓ Results in a single number that can be used to compare clustering performance across all algorithms and parameters.



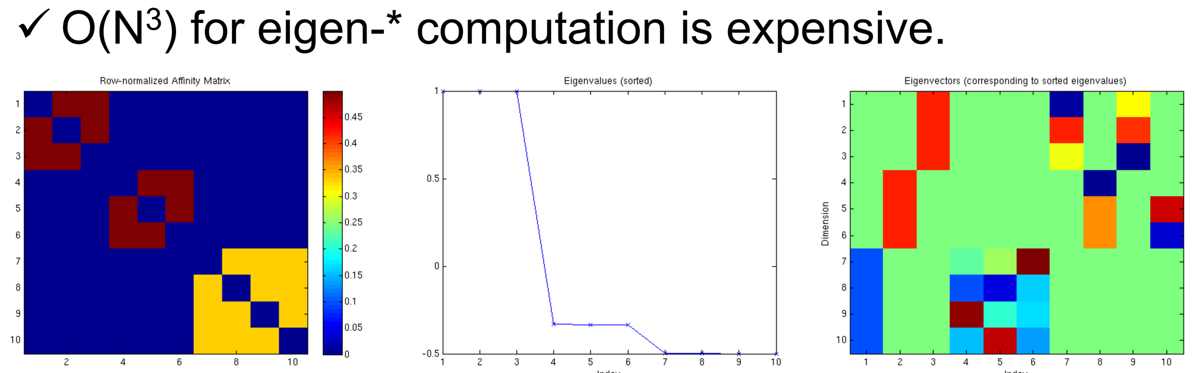
III. Graph Clustering Algorithms

- Agglomerative Hierarchical Clustering (AHC)

$$\frac{1}{|\mathcal{A}|\cdot|\mathcal{B}|}\sum_{x\in\mathcal{A}}\sum_{y\in\mathcal{B}} d(x,y) \le \theta$$

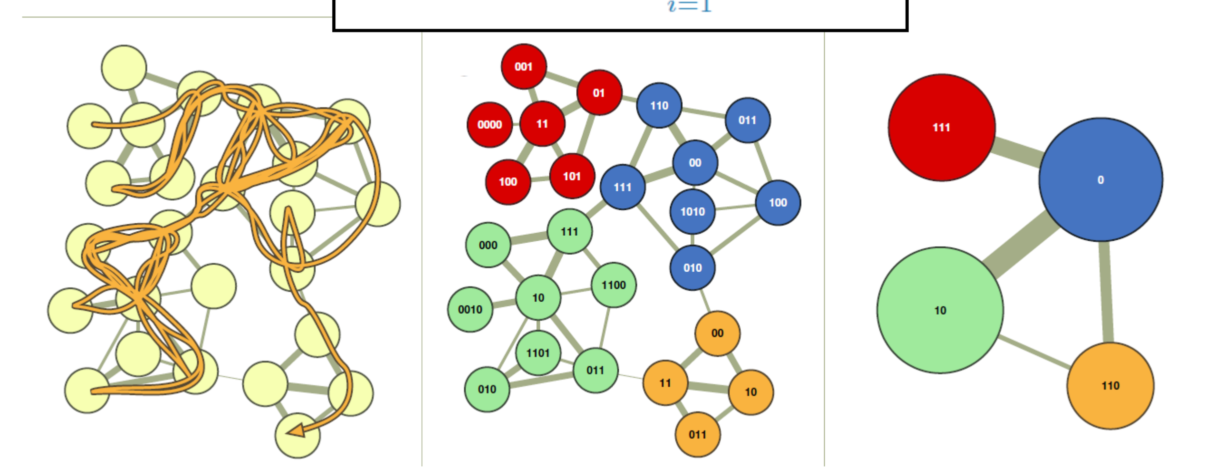


  - ✓ <u>Oracle benchmark</u>, #clusters used as stopping criterion.
- Modularity Optimization (CNM) [2]
  - ✓ Measure of the edge density within communities compared to the density between communities.
- Spectral Clustering (NJW) [3]
  - ✓ Use K largest eigenvectors of normalized affinity matrix (N x N) to project data onto lower dimensional space (i.e., M dimensions) before running K-means.
  - ✓ O(N^3) for eigen-* computation is expensive.

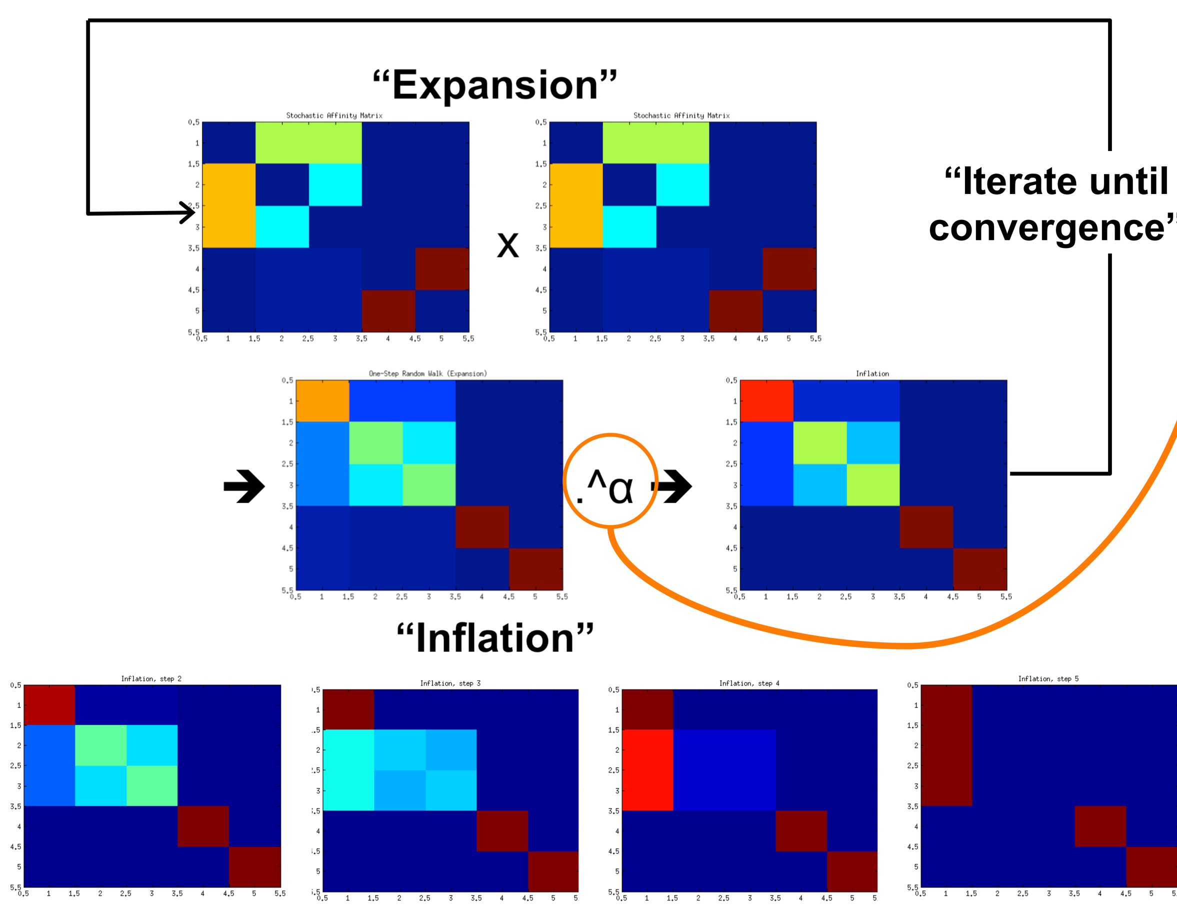


- Infomap [4]
  - ✓ Two-level compression of a random walk sequence, weighing between- and within-cluster entropies.

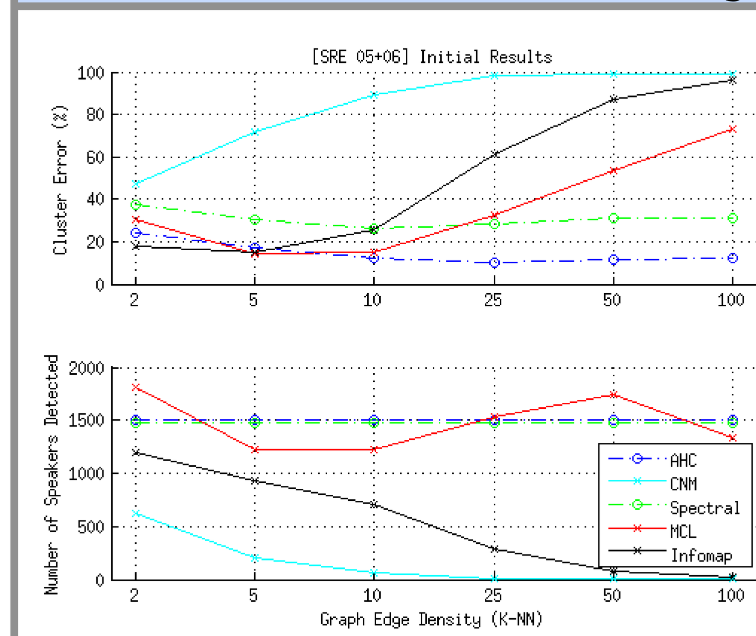
$$L(\mathsf{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p^i_{\circlearrowright} H(\mathcal{P}^i).$$

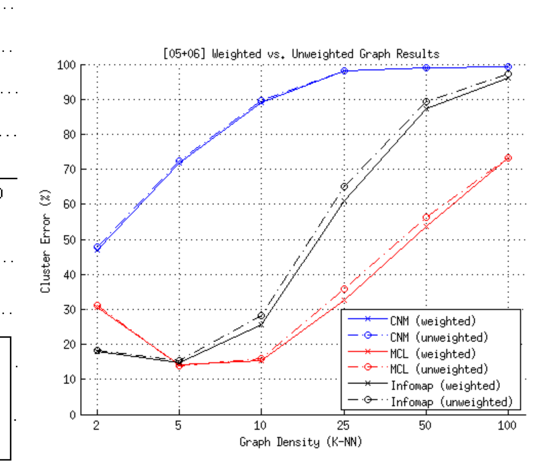


- Markov Clustering (MCL) [5]
  - ✓ "State" of clusters stored in normalized affinity matrix.



IV. Preliminary Results





- Unimportance of edge weight! (above right)
- CNM hindered by resolution limit
  - ✓ Evidenced by how few clusters are detected.
- Spectral-NJW uses #clusters as input
  - ✓ Still worse than AHC.
- MCL & Infomap
  - ✓ Do well at finding many small clusters.

V. System Refinements

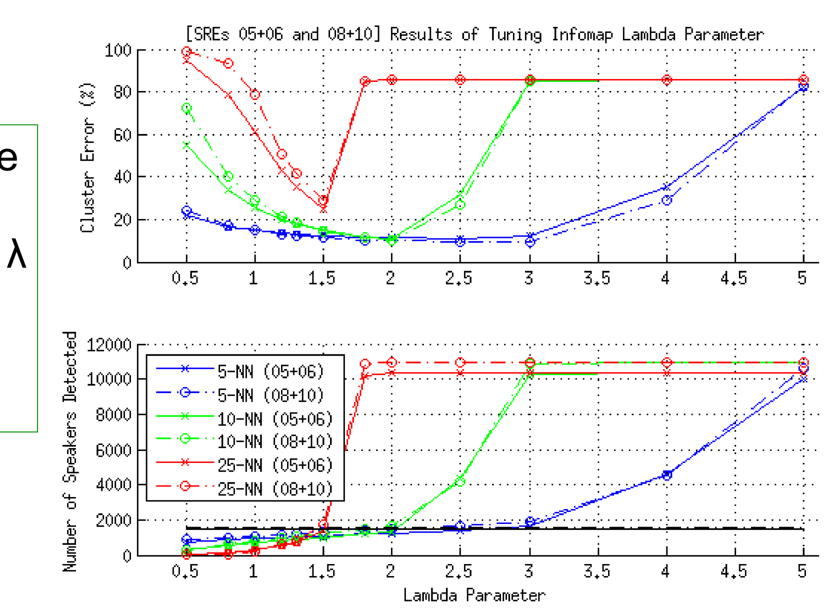
- Parameter Tuning
  - ✓ Infomap-λ

$$L(\mathsf{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \lambda \sum_{i=1}^{m} p^i_{\circlearrowright} H(\mathcal{P}^i).$$

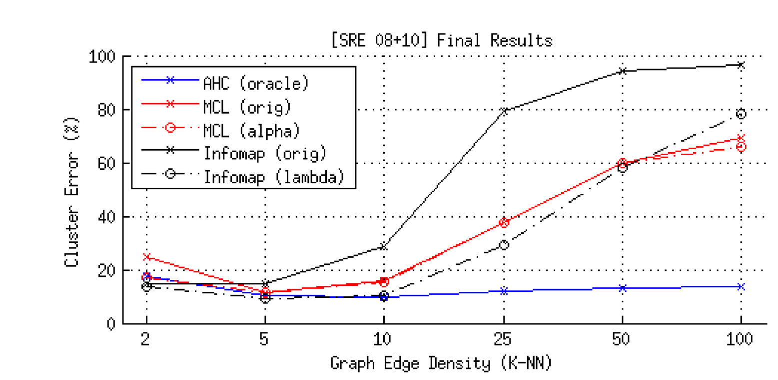
λ > 1 yields more, smaller clusters.

λ < 1 yields fewer, larger clusters.

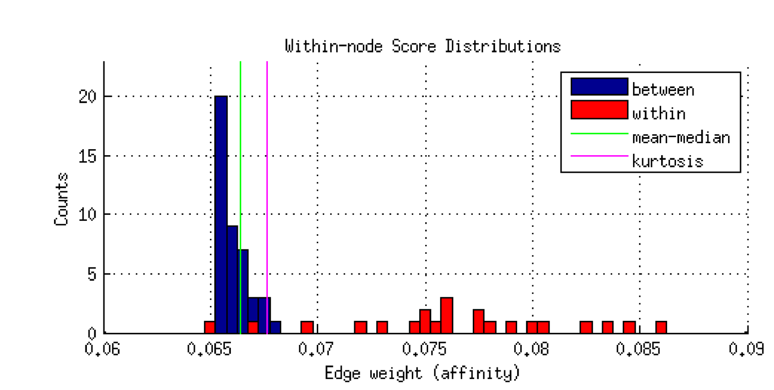
For different edge densities, there exists a value for λ that optimizes clustering performance.



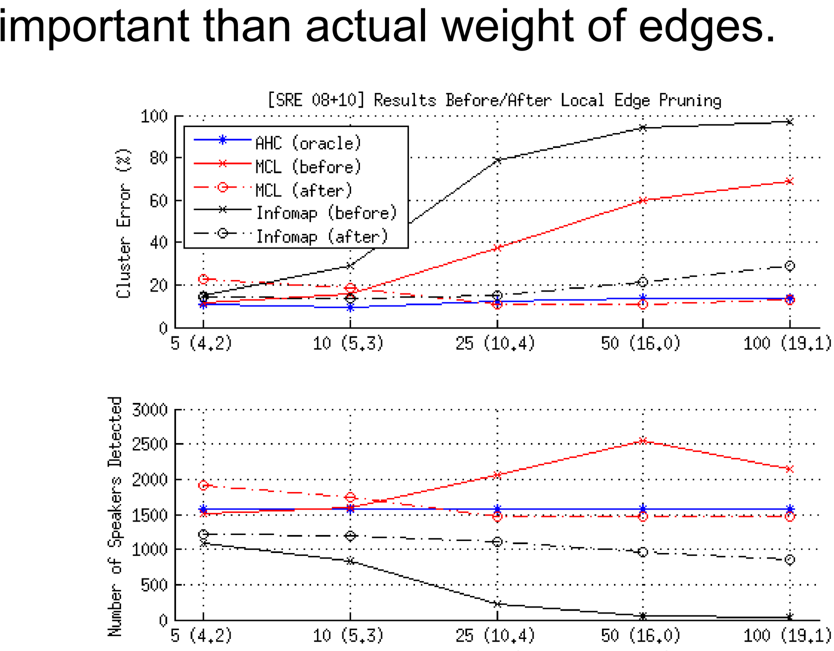
  - ✓ MCL-α
    - ➢ Faster convergence and more clusters estimated as α → ∞, and vice versa.
    - ➢ Resulting performance change is insignificant.



- Local Node Refinements



  - ✓ Knowing correct # neighbors seems to be more important than actual weight of edges.



Selected References

[1] Z. Karam and W. Campbell, "Graph Embedding for Speaker Recognition," in *Proceedings of Interspeech*, 2010, pp. 2742-2745.

[2] A. Clauset, M. Newman, and C. Moore, "Finding Community Structure in Very Large Networks," *Physical Review E*, 2004.

[3] A. Ng, M. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an Algorithm," in *Proceedings of NIPS*, 2001.

[4] M. Rosvall and C. Bergstrom, "Maps of Random Walks on Complex Networks Reveal Community Structure" in *Proceedings of NAS*, 2008.

[5] S. van Dongen, *Graph Clustering by Flow Simulation*, Ph.D. Thesis, University of Utrecht, 2000.

This work was sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.