

Alain Lelu  
Martine Cadot  
alelu@orange.fr

Supplementary material to the winning contribution to Neutral Cluster Benchmarking Challenge of IFCS2019  
«Evaluation of text clustering methods and their dataspace embeddings: an exploration»  
[hal-02116493v3]

## Links to the original codes used in the study

### K-Means (KM):

**Litekmeans.m** by Deng Cai - ©1993-2002

<[https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/1c38ca31-840b-4491-ac07-bd3ef7b18ad3/ccec48ae-2dc7-451d-a705-3790374f6b5f/previews/FSLib\\_v6.2.1\\_2018/lib/litekmeans.m/index.html](https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/1c38ca31-840b-4491-ac07-bd3ef7b18ad3/ccec48ae-2dc7-451d-a705-3790374f6b5f/previews/FSLib_v6.2.1_2018/lib/litekmeans.m/index.html)>

See: Deng Cai, "Litekmeans: the fastest matlab implementation of kmeans," Available at:

< <http://www.zjucadcg.cn/dengcai/Data/Clustering.html>, 2011 >

### Agglomerative Hierarchical Clustering (HCa, HCw):

<<https://octave.sourceforge.io/statistics/function/linkage.html> >

See: Ward, J. H., Hierarchical Grouping to Optimize an Objective Function, *J. Am. Statist. Assoc.* 1963, 58, 236-244, <<http://iv.slis.indiana.edu/sw/data/ward.pdf>>

### NMF:

**nmf\_kim.m** by Jingu Kim

< <http://www.cc.gatech.edu/~jingu>>

See: H. Kim, H. Park, D.-Z. Du, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares", *Proceedings of the IASTED International Conference on Computational and Systems Biology (CASB2006)*, pp. 95-100, Nov. 2006.

### LDA:

**runFastLDA.m** by Hanhuai Shan (2012)

<[https://www-users.cs.umn.edu/~shan/mmnmb\\_code.html](https://www-users.cs.umn.edu/~shan/mmnmb_code.html)>

See: I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, M. Welling. "Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation." *KDD 2008*.

### Louvain:

**cluster\_jl.m** by Antoine Scherrer

We used Waltman-Eck modularity option: in modularity.m replace MOD=... by MOD=MOD+Ec/Et-gamma\*(Et/m);

<[https://github.com/epfl-ts2/gspbox/blob/master/3rdparty/Community\\_BGLL\\_Matlab/cluster\\_jl.m](https://github.com/epfl-ts2/gspbox/blob/master/3rdparty/Community_BGLL_Matlab/cluster_jl.m)>

See "Fast unfolding of community hierarchies in large networks", Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre <<http://arxiv.org/abs/0803.0476>>

### Infomap:

**mcsd\_afg.m** by Erwan Le Martelot 09/11/11

<[http://www.elemartelot.org/res/downloads/cd/matlab/mcsd\\_afg.m](http://www.elemartelot.org/res/downloads/cd/matlab/mcsd_afg.m)>

Fast multi-scale detection of communities using Arenas et al.'s parameter described in (Arenas, Fernandez, Gomez; Analysis of the structure of complex networks at different resolution levels, *New Journal of Physics*, 2008