Université catholique de Louvain

LOUVAIN RESEARCH INSTITUTE IN MANAGEMENT AND ORGANIZATIONS (LOURIM)

WORKING PAPER

2017/21

Modularity-driven kernel k-means for community detection

Felix Sommer,
François Fouss,
Louvain Research Institute in Management and Organizations

Marco Saerens,
ICTEAM

Louvain Research Institute in Management and Organizations
Working Paper Series
Editor: Prof. Frank Janssen
(president-lourim@uclouvain.be)

Modularity-driven kernel k-means for community detection

Felix Sommer, Louvain Research Institute in Management and Organizations
François Fouss, Louvain Research Institute in Management and Organizations
Marco Saerens, ICTEAM

Summary

Nowadays, k-means is probably the most well-known and most popular clustering method in existence. This work evaluates if a new, autonomous, kernel k-means approach for graph node clustering coupled with the modularity criterion can rival the well-established Lou- vain method. We test the algorithm on social network datasets of various sizes and types. The new method estimates the optimal kernel or distance parameters as well as the natural number of clusters in the dataset. Results indicate that this new black-box algorithm manages to perform on par with the Louvain method given the same input.

Keywords: clustering, graph theory, modularity, kernel k-means, community detection

JEL Classification: C38

The support of this work by the FNRS and FSR (UCL) is gratefully acknowledged.

Corresponding author:
Felix Sommer
Louvain Research Institute in Management and Organizations (LouRIM)
Université catholique de Louvain
Chaussée de Binche 151
B-7000 Mons, BELGIUM
Email : felix.sommer@uclouvain.be

Modularity-driven kernel *k*-means for community detection

Felix Sommer, François Fouss, Marco Saerens

LSM – LouRIM & ICTEAM
Université catholique de Louvain
Chaussée de Binche 151, 7000 Mons, Belgium
felix.sommer@uclouvain.be francois.fouss@uclouvain.be
marco.saerens@uclouvain.be

Abstract. Nowadays, *k*-means is probably the most well-known and most popular clustering method in existence. This work evaluates if a new, autonomous, kernel *k*-means approach for graph node clustering coupled with the modularity criterion can rival the well-established Louvain method. We test the algorithm on social network datasets of various sizes and types. The new method estimates the optimal kernel or distance parameters as well as the natural number of clusters in the dataset. Results indicate that this new black-box algorithm manages to perform on par with the Louvain method given the same input.

Keywords: clustering, graph theory, modularity, kernel *k*-means, community detection

1 INTRODUCTION

Identifying user- or item-clusters has been of interest in many fields for various reasons [46]. Specifically, in data mining and data analysis applications, grouping similar items or users together can lead to interesting and useful insights (see, e.g., [1]). Moreover, in social network analysis, researchers more and more use segmentation to analyze groups, direct contacts, indirect relationships, etc. Social network analysis shows connections between nodes with ties, edges, or links. The nodes can be individual users or items linked in the network; the ties or links are relationships and interactions [23].

A very popular clustering method is the *k*-means algorithm, as it is easy to understand, implement, and able to run with reasonably limited computational resources. To address the shortcomings of using a *k*-means on a graph structure, approaches such as kernel *k*-means were developed [47].

Kernel *k*-means can use various kernels derived from the distance-based or proximity-based data, and is particularly useful in community detection, that is, finding similar nodes in a given graph and grouping them together [47].

Today, the *k*-means algorithm is probably the most popular clustering algorithm in existence. In this work we evaluate if a simple kernel *k*-means coupled with modularity criterion—introduced by Newman and Girvan [38, 37]—can choose parameters and the natural number of clusters on its own, such that it manages to rival the Louvain method.

We show that with a properly chosen distance used for the clustering, *k*means indeed does just as well (but not better) than the Louvain method, when compared using statistical measures and rank-comparison in community detection.

Brief related work and contributions. Varying the graph node distance between a graph's nodes can have a substantial influence on clustering effectiveness. As shown in prior work [26, 44], modern graph distances can improve clustering solutions. However, clustering using a kernel *k*-means approach still suffers a major drawback, in that the *k*-means algorithm always requires that the number of clusters is given.

The modularity criterion is a standard measure for community detection in various fields, including social network analysis [16]. While modularity has its limitations (see, e.g., [19]), it is generally accepted that by maximizing modularity a reasonable good community structure can often be obtained [32]. Furthermore, there are attempts to alleviate its limitations [39, 32]. As such, modularity should be of interest for clustering tasks and indeed, there have been many proposals to use the modularity criterion in clustering (see, e.g., [5, 3]). However, combining the modularity criterion with recent developments in graph distances and applying this combination in order to empirically analyze the effects this could have on clustering has—to the best of our knowledge—not yet been done.

Following up on our prior work of empirically testing new distances [44], we address the discussed shortcomings of using kernel *k*-means clustering typically encountered. In this work we therefore:

- propose automatic kernel parameter tuning, in order to improve clustering quality, without requiring user interaction, using the modularity criterion,

- identify the natural number of clusters present in a dataset, using also the modularity criterion, without knowledge of the (correct) number of clusters beforehand,
- finally, compare this improved kernel *k*-means approach, using the kernels presented in our previous work [44], as well as additional kernels [26], to the Louvain method [2].

2 THEORETIC BACKGROUND

Let $G = (V,E)$ be a graph consisting of a set of n vertices V and an edge set E. A cost matrix $\mathbf{C}$ consists of non-negative scalars $c_{ij} \geq 0$, which represent the cost of following the edge linking nodes i and j. If not defined in our problem, we can compute a cost matrix from an adjacency matrix $\mathbf{A}$, using the relation $c_{ij} = 1/a_{ij}$, where a_{ij} are the elements in $\mathbf{A}$ indicating the affinity between nodes i and j. Also, we can derive G's Laplacian matrix as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where the diagonal matrix $\mathbf{D} = \mathbf{Diag}(\mathbf{A}^{\mathsf{T}}\mathbf{e})$ contains the column sums of $\mathbf{A}$ with a column vector full of ones, $\mathbf{e}$, and transpose $\mathbf{A}^{\mathsf{T}}$.

In the following we briefly present the kernels and distances used in our experiments. We transform distances and dissimilarities into a kernel $\mathbf{K}$ using the relationship $\mathbf{K} = -\frac{1}{2}\mathbf{H}\boldsymbol{\Delta}^{(2)}\mathbf{H}$, with the centering matrix $\mathbf{H} = \mathbf{I} - \frac{\mathbf{E}}{n}$ (see [4]); $\mathbf{E}$ is an $n \times n$ matrix full of ones and n is the number of nodes; $\boldsymbol{\Delta}^{(2)}$ is a square distance or dissimilarity matrix.

Shortest-path (SP) distance. This distance is one of the most popular distances available. It is defined as the distance of the shortest-path between two nodes. For details on this distance, please refer to the literature (e.g., [13]).

Randomized shortest-path (RSP) dissimilarity. This dissimilarity interpolates between the shortest-path distance (see, e.g., [29, 40, 49]) and half the commute-time distance based on a pure random walk. The random walker adopts a *randomized* strategy biased towards the paths with lowest cost ([29, 40], see also [20]). The randomized shortest-path cost corresponds to the *expected cost* over all paths connecting these two nodes. With the cost matrix $\mathbf{C}$, the transition probability matrix of the natural random walk $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$, and the inverse temperature $\beta = 1/T$ parameter we can define the matrix $\mathbf{W} = \exp(-\beta\mathbf{C}) \circ \mathbf{P}$, where $\circ$ marks element-wise multiplication. Using this matrix we define the fundamental matrix of the killed, but non-absorbing Markov Chain $\mathbf{Z} = (\mathbf{I}-\mathbf{W})^{-1}$, and furthermore the matrix $\mathbf{S} = (\mathbf{Z}(\mathbf{C} \circ \mathbf{W})) \div \mathbf{Z}$, where $\div$ marks

element-wise division. This step enables us to compute the expected cost of hitting walks as $\mathbf{C}^{-} = \mathbf{S} - \mathbf{e}\mathbf{d}_S^T$. Recall, that **e** is a column vector of ones and here, $\mathbf{d}_S$ is a vector of the diagonal elements of **S**. We can then symmetrize to obtain the dissimilarity matrix:

$$\mathbf{\Delta}^{RSP} = (\mathbf{C}^{-} + \mathbf{C}^{-T})/2. \quad (1)$$

Free energy (FE) distance. This distance is based on the minimum Helmholtz free energy or potential distance. This minimization of free energy can be considered as a regularized version of the Randomized shortest-path (RSP) dissimilarity. [29] Here, the random walker finds a compromise between the expected cost and predictability (entropy). We define the matrices **C**, **W**, **Z**, and parameter β as we did for the randomized shortest-path above and compute the free energy between all pairs of nodes $\mathbf{\Phi} = -\log(\mathbf{Z}\ \mathbf{Diag}(\mathbf{Z})^{-1})/\beta$. Symmetrization then yields the free energy distance matrix [22, 29]:

$$\mathbf{\Delta}^{FE} = (\mathbf{\Phi} + \mathbf{\Phi}^T)/2. \quad (2)$$

Sigmoid commute-time (SCT) similarity. This similarity or kernel has already proven useful in clustering applications [47, 48]. It is obtained by applying a sigmoid transformation [41] on the commute-time kernel. The kernel is derived from the average commute time a random walker takes, when starting in one node *i*, going to another node *j*, and then back to the starting node *i*, hence commuting. Using the Moore-Penrose pseudo-inverse $\mathbf{L}^+$ of the Laplacian matrix **L**, with a normalizing factor σ set to the standard deviation of the elements of $\mathbf{L}^+$, and α as a parameter (see also [47, 20]), we define it using the sigmoid function $s(x) = 1/(1 + \exp(-\alpha x/\sigma))$ as:

$$\mathbf{K}_{SCT} = s(\mathbf{L}^+) \quad (3)$$

Corrected commute-time (CCT) distance. As the commute-time similarity tends have certain drawbacks on large graphs, [45] proposed a correction term leading to this distance. Using a matrix similar to the modularity matrix, we define $\mathbf{M} = \mathbf{D}^{-\frac{1}{2}\left(\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^T}{\mathrm{vol}(G)}\right)}\mathbf{D}^{-\frac{1}{2}}$, where **d** is a vector of the diagonal elements of

D, $\mathrm{vol}(G)$ is the volume of the graph, i.e., the sum of all elements of the adjacency matrix **A**, and **I** is the identity matrix. As for the SCT kernel, α is a parameter and σ is a normalizing factor. We then derive the corrected commute-time kernel with sigmoid transformation as follows ([45], see also [20]):

$$\mathbf{K}_{CCT} = s(\mathbf{H}\mathbf{D}^{-\frac{1}{2}}\mathbf{M}(\mathbf{I}-\mathbf{M})^{-1}\mathbf{M}\mathbf{D}^{-\frac{1}{2}}\mathbf{H}) \quad (4)$$

Communicability (Com) kernel. This kernel is also known as the exponential diffusion kernel and is based on the communicability distance [30, 21, 18]. The communicability between a graph's two nodes *i* and *j* is the weighted sum of all walks starting at node *i* and ending in node *j*. More importance is attributed to shorter walks than to longer ones. The kernel is defined by the following equation:

$$\mathbf{K}_{Com} = \text{expm}(t\mathbf{A}), \qquad t > 0 \quad (5)$$

Matrix-Forest (For) kernel. This kernel is also known as the Regularized Laplacian kernel and is based on the regularization of the graph Laplacian [9, 10, 43]. It is defined as follows:

$$\mathbf{K}_{For} = (\mathbf{I} + t\mathbf{L})^{-1}, \qquad t > 0 \quad (6)$$

Heat kernel (Heat). This kernel is also known as the Laplacian exponential diffusion kernel [11]. The idea of the heat kernel stems from classical physics where the diffusion of, e.g., heat is described by the heat equation, from which this kernel can be derived:

$$\mathbf{K}_{Heat} = \text{expm}(-t\mathbf{L}), \qquad t > 0 \quad (7)$$

Plain walk kernel (Walk). This kernel is also known as the Neumann kernel and has originally been proposed to compute document similarity [27], but has equally been applied in the context of link analysis in social network analysis [25]. It is also closely related to the Katz similarity [28] and the walk distance [8]. Note that ρ is the spectral radius of the adjacency matrix **A**, i.e., the absolute value of **A**'s largest eigenvalue $\rho(\mathbf{A}) = \max_i(|\lambda_i|)$ for **A**'s eigenvalues $\lambda_1,\ldots,\lambda_n$:

$$\mathbf{K}_{Walk} = (\mathbf{I} - t\mathbf{A})^{-1}, \qquad 0 < t < \rho^{-1} \quad (8)$$

Logarithmic kernels. The four kernels presented just beforehand equally exist in logarithmic form. We refer to [6, 7, 8] as well as the indicated references for details. Note that ln indicates application of element-wise natural logarithm.

Log. Communicability [26]: $\mathbf{K}_{lCom} = \ln(\text{expm}(t\mathbf{A})), \qquad t > 0$ (9)

Log. Heat [26]: $\mathbf{K}_{lHeat} = \ln(\text{expm}(-t\mathbf{L})), \qquad t > 0$ (10)

Log. Forest [7]: $\mathbf{K}_{lFor} = \ln(\mathbf{I} + t\mathbf{L})^{-1}, \qquad t > 0$ (11)

$$\text{Log. Walk [8]:} \quad \mathbf{K}_{\text{lWalk}} = \ln(\mathbf{I} - t\mathbf{A})_{-1}, \qquad 0 < t < \rho_{-1} \qquad (12)$$

3 METHODOLOGY

In this section we present the clustering methods we investigate as well as their runtime parameters employed in the experiments. We test the following techniques:

- an enhanced kernel *k*-means approach, that employs the modularity criterion to estimate the optimal kernel or distance parameters as well as the natural number of clusters,
- the Louvain method [2], which also estimates the natural number of clusters on its own,

on 15 graph datasets, the smallest of which (Zachary's Karate club, [50]) contains 34 nodes. The largest graph (a Newsgroup graph, [34, 47] with five classes) contains 999 nodes. We analyze a total of nine Newsgroup datasets (400 to 999 nodes). The remaining are the classical Football [23] and political blogs [31] datasets, and finally, three artificial Lancichinetti-Fortunato-Radicchi (LFR) graphs [33].

3.1 Kernel *k*-means coupled with modularity criterion

This method is an extension of the kernel *k*-means algorithm. This extended method optimizes kernel parameters and automatically estimates the natural number of clusters present in the dataset. The kernel *k*-means algorithm itself is the same as in our previous work, see [44], and corresponds to a two-step iterative algorithm based on a distance—or dissimilarity—matrix instead of features. Given a meaningful, symmetric distance matrix **Δ**, containing the distances Δ_{ij}, the goal is to partition the nodes by minimizing the total *within-cluster sum of distances*. For details, see [47, 20].

The change between the standard kernel *k*-means and the kernel *k*-means coupled with modularity criterion is found in the two step approach of the latter, which is more easily explained using the flowchart in Figure 1. As shown in this flowchart, prior to the *actual clustering* step, the *optimize parameters* step is executed. For the present case we (very verbosely) test 77 kernel parameters and 2 to 18 clusters per kernel and dataset. We then identify the optimum by choosing the kernel parameter and number of clusters combination which shows the highest modularity. The chosen kernel

parameter and the identified number of clusters are passed to the kernel *k*-means algorithm, which is run in the *actual*

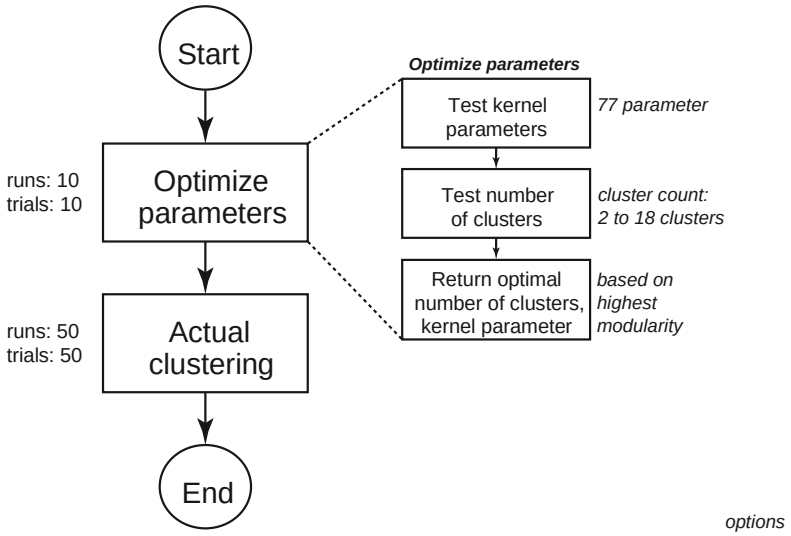

Fig.1. Kernel *k*-means coupled with modularity criterion flowchart.

clustering step. The results obtained from the *actual clustering* step are analyzed and discussed in the following Section 4.

Let us briefly elaborate on the details of the kernel *k*-means approach. For both the *optimize parameters* and the *actual clustering* steps it is important to keep in mind that as the kernel *k*-means approach depends on a random initialization of prototypes, it is necessary to test multiple trials with different initializations. Among these trials we keep the partition showing the lowest *within-cluster sum-of-distances*. It is furthermore necessary to average the results to account for variance in the results and we thus repeat the trials procedure multiple times. As shown in Figure 1 for the *optimize parameters* step, we execute 10 runs, consisting of 10 trials each, while for the *actual clustering* step, which generates the final results, we conservatively choose to take the average of 50 runs, each of which consists of 50 trials.

The kernel *k*-means coupled with modularity criterion approach is tested using the 13 different kernels, described in Section 2. The aforementioned 77 kernel parameters are tested for each of the 13 kernels. Note that all the compared methods find a natural number of clusters.

3.2 Louvain method

The idea behind the Louvain method [2] is to first perform an *iterative local optimization* (see, e.g., [17]) for seeking local minimum of a specific criterion (step 1) – in our case, the modularity criterion [38, 37, 36, 35]. Then, the second phase, called the *nodes aggregation* or *coarsening*, step whose purpose is to build a new agglomerated graph (step 2), is performed. These two steps are repeated until no further improvement of the employed criterion can be achieved. The details are described in the original work [2].

The Louvain method [2] was used as a baseline comparison method and it equally determines the natural number of clusters for each dataset by itself. We denote this method with the label LVN.

3.3 Evaluation methods

To compare the algorithms' performance, we compute three metrics for all datasets and methods: the adjusted rand index (ARI) [24], the normalized mutual information criterion (NMI) [12, 15] as well as the classification rate (CR). Furthermore, we run a Friedman-Nemenyi test-based [14, 15] multiple comparison test [15], which allows for the aggregation of results over multiple datasets, therefore comparing the differences for all algorithms simultaneously. Similarly, we perform two-by-two Wilcoxon signed-rank tests (see, e.g., [42]) to ascertain individual algorithms' performances in a one-on-one rank comparison to identify in a meaningful manner if one algorithm statistically significantly outperforms another. We compare the kernel *k*-means coupled with modularity criterion with its different kernels to the Louvain method.

4 RESULTS & DISCUSSION

Both clustering methods—kernel *k*-means coupled with modularity criterion and the Louvain method—were run on the graph-based community datasets and ARI, NMI, and CR metrics were computed for all of them. We give the raw numbers for the NMI in Table 1, the results for the ARI and CR show a very similar pattern and we thus omit those here. The results of the Friedman-Nemenyi multiple comparison test are shown in Figures 2 and 3, and discussed below, as are the results of the two-by-two Wilcoxon signed-rank tests.

Table 1. Normalized Mutual Information (NMI)

Dataset	SCT	CCT	FE	RSP	SP	Com	For	Heat	Walk	lCom	lFor	lHeat	lWalk	LVN
football	**0.908**	0.889	0.884	0.884	0.812	0.545	0.293	0.492	0.749	0.598	0.424	0.583	0.249	0.698
lfr1	0.976	**0.990**	0.981	0.983	0.890	0.052	0.027	0.051	0.003	0.475	0.500	0.511	0.485	0.962
lfr2	**1.000**	**1.000**	**1.000**	**1.000**	0.986	0.147	0.164	0.266	0.694	0.481	0.478	0.449	0.682	0.823
lfr3	**1.000**	0.997	**1.000**	**1.000**	0.989	0.087	0.130	0.159	0.502	0.520	0.379	0.376	0.729	0.827
news2cl1	0.608	**0.638**	0.584	0.581	0.434	0.018	0.032	0.032	0.249	0.291	0.243	0.253	0.462	0.573
news2cl2	0.356	0.396	0.359	0.346	0.296	0.008	0.027	0.100	0.394	**0.584**	0.123	0.207	0.233	0.432
news2cl3	0.579	0.584	0.590	0.598	0.616	0.333	0.053	0.117	0.357	**0.682**	0.553	0.611	0.496	0.586
news3cl1	0.697	0.706	0.702	0.696	0.660	0.096	0.032	0.125	0.245	0.489	0.471	0.478	**0.748**	0.699
news3cl2	0.656	0.689	**0.711**	0.706	0.572	0.068	0.035	0.040	0.294	0.345	0.343	0.281	0.572	0.661
news3cl3	0.642	0.605	0.681	0.678	**0.720**	0.067	0.031	0.033	0.380	0.295	0.255	0.248	0.423	0.673
news5cl1	0.641	0.643	0.648	0.651	0.614	0.087	0.019	0.032	0.284	0.259	0.273	0.291	0.504	**0.684**
news5cl2	0.616	0.630	**0.643**	0.633	0.596	0.185	0.026	0.024	0.291	0.345	0.214	0.234	0.374	0.634
news5cl3	0.573	**0.624**	0.615	0.571	0.480	0.034	0.021	0.021	0.369	0.300	0.337	0.281	0.383	0.572
polblogs	0.556	0.556	0.567	0.568	0.549	0.423	0.322	**0.641**	0.298	0.511	0.550	0.527	0.577	0.597
zachary	0.832	0.724	0.838	0.832	**1.000**	0.262	0.395	**1.000**	0.875	**1.000**	0.866	**1.000**	0.697	0.982

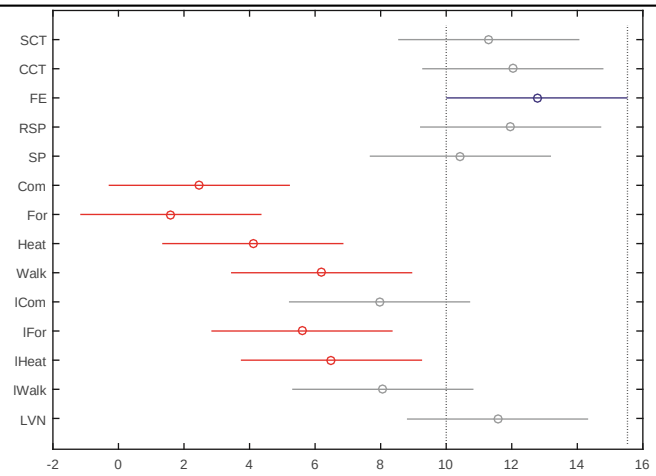

Fig.2. Friedman-Nemenyi multiple comparison test aggregating results of all datasets for the individual algorithms based on NMI. A higher rank, i.e., a bar shifted to the right, indicates better results; the length of the bars shows the significance interval. If these bars do not overlap the rank is considered significantly different.

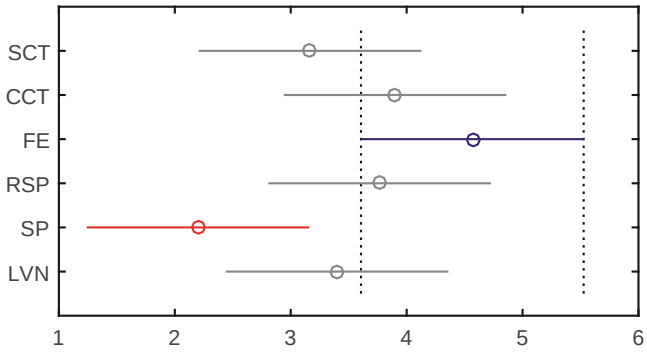

Fig.3. Friedman-Nemenyi multiple comparison test aggregating the results of all datasets for the six best algorithms. A higher rank, i.e., a bar shifted to the right, indicates better results; the length of the bars shows the significance interval. If these bars do not overlap the rank is considered significantly different.

In the following we present and discuss the results based on an analysis of the NMI. As mentioned before, the results from the NMI, ARI, as well as CR metrics show very similar patterns, and we thus restrict the analysis to NMIbased results.

It remains difficult to identify one "best" clustering distance. Not surprisingly, the results in Table 1 show that the datasets have a significant impact on clustering quality. Overall, it is however safe to say, that based on the multiple comparison using all methods—with the exception of the logarithmic communicability and logarithmic walk kernels—the modern FE, CCT, RSP, SCT kernels outperform the competition to a statistically significant degree. Comparing only these best kernels amongst each other (cf. Figure 3) confirms, that the most modern kernels (FE, CCT, and RSP) have an edge on the competition. A two by-two Wilcoxon signed-rank test, comparing the NMI for the FE-distance to the five best competitors (that is, SCT, CCT, RSP, SP and LVN) shows, that the FE-distance based kernel manages to outperform the SCT- and SP-distance to a statistically significant degree (both $p < 5 \times 10^{-2}$), but not the others (CCT, RSP and LVN).

Comparing the results in Table 1 to those in [44] shows that the lack of the true number of clusters has a detrimental effect on clustering quality, which also cannot be compensated by optimizing kernel parameters per dataset. This behavior should however be expected, as thus far no method has been found to deliver perfect clustering results, especially with limited to no knowledge on the data.

However, when comparing to the Louvain method, given the same information, i.e., the lack of the true number of clusters, the kernel *k*-means with modularity criterion approach manages to perform on par, or even outperform the Louvain method, albeit not to statistically significant degree. When comparing only the six best algorithms in a multiple comparison as shown in Figure 3, the simple FE kernel based *k*-means obtains results equivalent to the Louvain method.

5 CONCLUSION AND FUTURE WORK

This work presents a novel method for executing clustering in a black-box fashion, while making use of recent advances in graph distances. We compare 13 different graph kernels and the Louvain method. The *k*-means clustering performance depends on the dataset, but obtains results equivalent to, and thus rivals, the Louvain method almost all of the time, without needing additional input parameters. The method instead optimizes each kernel's parameters and predicts the number of clusters in the dataset based on modularity. The results are strongly influenced by the choice of kernel, with the most modern FE, CCT, and RSP kernels showing the best results.

Future work will focus on (1) refining kernel parameters' search space in order to increase estimation speeds, and (2) further tune the algorithm to better handle unknown datasets.

ACKNOWLEDGEMENTS

This work was supported in part by the FNRS and UCL (FSR) through a PhD scholarship. This work was also partially supported by the Immediate and the Brufence projects funded by Innoviris (Brussels Region). We thank these institutions for giving us the opportunity to conduct both fundamental and applied research.

Bibliography

[1] Berkhin, P.: A Survey of Clustering Data Mining Techniques. In: Grouping Multidimensional Data, pp. 25–71. Springer (2006)
[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment p. P10008 (2008)
[3] Bolla, M.: Penalized versions of the Newman-Girvan modularity and their relation to normalized cuts and k-means clustering. Physical Review E 84(1), 016108 (2011)
[4] Borg, I., Groenen, P.: Modern multidimensional scaling. Springer (1997)
[5] Brandes, U., Delling, D., Gaertler, M., G¨orke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: On modularity clustering. IEEE Transactions on Knowledge and Data Engineering 20, 172–188 (2008)
[6] Chebotarev, P.: A class of graph-geodetic distances generalizing the shortest-path and the resistance distances. Discrete Applied Mathematics 159(5), 295–302 (2011)
[7] Chebotarev, P.: The Graph Bottleneck Identity. Advances in Applied Mathematics 47(3), 403– 413 (2011)
[8] Chebotarev, P.: The Walk Distances in Graphs. Discrete Applied Mathematics 160(10–11), 1484–1500 (2012)
[9] Chebotarev, P., Shamis, E.: The matrix-forest theorem and measuring relations in small social groups. Automation and Remote Control 58(9), 1505–1514 (1997)
[10] Chebotarev, P., Shamis, E.: The forest metric for graph vertices. Electronic Notes in Discrete Mathematics 11, 98–107 (2002)
[11] Chung, F., Yau, S.T.: Coverings, heat kernels and spanning trees. Journal of Combinatorics 6, 163–184 (1998)
[12] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multi-modality image registration based on information theory. In: Information Processing in Medical Imaging. vol. 3, pp. 263–274 (1995)
[13] Cormen, T., Leiserson, C., Rivest, R., Stein, C.: Introduction to Algorithms. The MIT Press, 3rd edition edn. (2009)
[14] Daniel, W.: Applied Non-parametric Statistics. The Duxbury Advanced Series in Statistics and Decision Sciences, PWS-Kent Publishing Company (1990)
[15] Dem˘sar, J.: Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7, 1–30 (2006)
[16] Devooght, R., Mantrach, A., Kivimaki, I., Bersini, H., Jaimes, A., Saerens, M.: Random walks based modularity: application to semi-supervised learning. Proceedings of the 23rd International World Wide Web Conference (WWW 2014) pp. 213–224 (2014)
[17] Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. John Wiley & Sons (1973)
[18] Estrada, E.: The communicability distance in graphs. Linear Algebra and its Applications 436(11), 4317 – 4328 (2012)
[19] Fortunato, S., Barthelemy, M.: Resolution limit in community detection. Proceedings of the National Academy of Sciences of the United States of America 104(1), 36–41 (2007)
[20] Fouss, F., Saerens, M., Shimbo, M.: Algorithms and Models for Network Data and Link Analysis. Cambridge University Press (2016)
[21] Fouss, F., Yen, L., Pirotte, A., Saerens, M.: An experimental investigation of graph kernels on a collaborative recommendation task. Proceedings of the 6th International Conference on Data Mining (ICDM 2006) pp. 863–868 (2006)
[22] Franc¸oisse, K., Kivima¨ki, I., Mantrach, A., Rossi, F., Saerens, M.: A bag-of-paths framework for network data analysis. To appear in Neural Networks (2017)
[23] Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. In: Proceedings of the National Academy of Sciences. vol. 99, pp. 7821–7826 (2002)
[24] Hubert, L., Arabie, P.: Comparing partitions. Journal of Classification 2(1), 193–218 (1985) [25] Ito, T., Shimbo, M., Kudo, T., Matsumoto, Y.: Application of kernels to link analysis. In: Proceedings of the eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 586–592 (2005)
[26] Ivashkin, V., Chebotarev, P.: Do logarithmic proximity measures outperform plain ones in graph clustering? In: Proceedings of 6th International Conference on Network Analysis (2016)

[27] Kandola, J., Cristianini, N., Shawe-Taylor, J.: Learning semantic similarity. Advances in Neural Information Processing Systems (NIPS 2002) 15 pp. 657–664 (2002)

[28] Katz, L.: A new status index derived from sociometric analysis. Psychmetrika 18(1), 39–43 (1953)
[29] Kivim¨aki, I., Lebichot, B., Saerens, M.: Developments in the theory of randomized shortest paths with a comparison of graph node distances. Physica A: Statistical Mechanics and its Applications 393, 600–616 (2014)
[30] Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete structures. In: Proceedings of the 19th International Conference on Machine Learning (ICML 2002). pp. 315–322 (2002)
[31] Krebs, V.: New political patterns (2008), http://www.orgnet.com/divided.html
[32] Lancichinetti, A., Fortunato, S.: Limits of modularity maximization in community detection. Physical review E 84(6), 066122 (2011)
[33] Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community detection algorithms. Physical review E 78(4), 46–110 (2008) [34] Lang, K.: 20 newsgroups dataset, http://bit.ly/lang-newsgroups
[35] Newman, M.E.J.: Networks: An introduction. Oxford University Press (2010)
[36] Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Physical Review E 74(3), 036104 (2006)
[37] Newman, M.E.J.: Modularity and community structure in networks. In: Proceedings of the National Academy of Sciences. vol. 103, pp. 8577–8582 (2006)
[38] Newman, M.E.J., Girvan, M.: Finding and Evaluating Community Structure in Networks. Physical Review E 69, 026113 (2004)
[39] Reichardt, J., Bornholdt, S.: Detecting fuzzy community structures in complex networks with a Potts model. Physical Review Letters 93(21), 218701 (2004)
[40] Saerens, M., Achbany, Y., Fouss, F., Yen, L.: Randomized shortest-path problems: Two related models. Neural Computation 21(8), 2363–2404 (2009)
[41] Sch¨olkopf, B., Smola, A.J.: Learning with Kernels. The MIT Press (2002)
[42] Siegel, S.: Non-parametric Statistics for the Behavioral Sciences. McGraw-Hill (1956)
[43] Smola, A.J., Kondor, R.: Kernels and regularization on graphs. In: Warmuth, M., Scho¨lkopf, B. (eds.) Proceedings of the Conference on Learning Theory. pp. 144–158 (2003)
[44] Sommer, F., Fouss, F., Saerens, M.: Comparison of graph node distances on clustering tasks. In: International Conference on Artificial Neural Networks. pp. 192–201. Springer (2016)
[45] von Luxburg, U., Radl, A., Hein, M.: Getting lost in space: large sample analysis of the commute distance. In: Proceedings of the 23th Neural Information Processing Systems conference (NIPS 2010). pp. 2622–2630 (2010)
[46] Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Transactions on neural networks 16(3), 645–678 (2005)
[47] Yen, L., Fouss, F., Decaestecker, C., Francq, P., Saerens, M.: Graph nodes clustering based on the commute-time kernel. In: Proceedings of the 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2007). pp. 1037–1045 (2007)
[48] Yen, L., Fouss, F., Decaestecker, C., Francq, P., Saerens, M.: Graph nodes clustering with the sigmoid commute-time kernel: A comparative study. Data & Knowledge Engineering 68(3), 338–361 (2009)
[49] Yen, L., Mantrach, A., Shimbo, M., Saerens, M.: A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. In: Proceedings of the 14th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008). pp. 785–793 (2008)
[50] Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of Anthropological Research (33), 452–473 (1977)