Supplementary Material for
On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study
by G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent and M. E. Houle
Data Mining and Knowledge Discovery 30(4): 891-927, 2016, DOI: 10.1007/s10618-015-0444-8

Shuttle (version#01)

This dataset has been preprocessed in different variants in the literature. We follow the procedure of Zhang et al. [1], using classes 1, 3, 4, 5, 6 and 7 as inliers and class 2 as outlier, selecting 1000 inliers vs. 13 outliers (class 2). The selection of instances is based on the test set. The processed dataset consists of 1013 instances represented in 9 attributes, with 13 outliers (1.28%) and 1000 inliers (98.72%).

References:

[1] K. Zhang, M. Hutter, and H. Jin. A new local distance-based outlier detection approach for scattered real-world data. In Proc. PAKDD, pages 813-822, 2009.

Download all data set variants used (328.2 kB). You can also access the original data. (shuttle.tst, [1] only uses test set)

- Normalized, without duplicates
- Not normalized, without duplicates

Normalized, without duplicates

This version contains 9 attributes, 1013 objects, 13 outliers (1.28%)

Download raw algorithm results (8.4 MB) Download raw algorithm evaluation table (46.3 kB)

Best Parameters

The following table contains the best (overall and per-method) results for each method and evaluation measure (when the same score was achieved twice, only the smallest k is given).
The Maximum F1-Measure is complimentary in addition to the measures in the original publication.

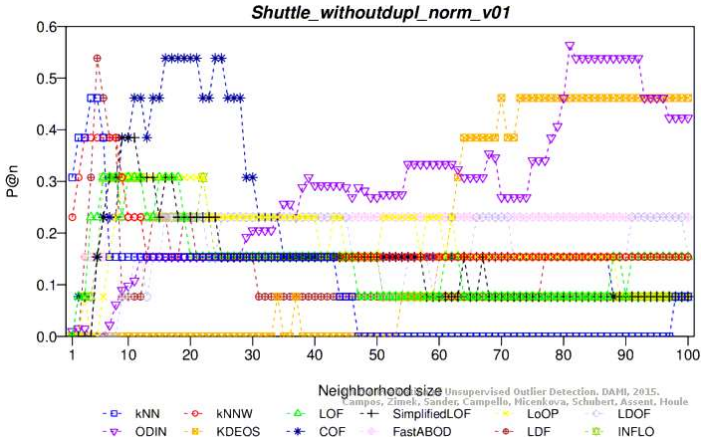
Algorithm	k	P@n	Adj. P@n	AP	Adj. AP	Max-F1	Adj. MF1	ROC AUC
KNN	**3**	0.38462	0.37662	0.38600	0.37802	0.58537	0.57998	0.98908
KNN	**4**	0.46154	0.45454	0.38494	0.37695	0.55319	0.54738	0.98746
KNNW	**3**	0.38462	0.37662	0.30430	0.29526	0.51429	0.50797	0.91754

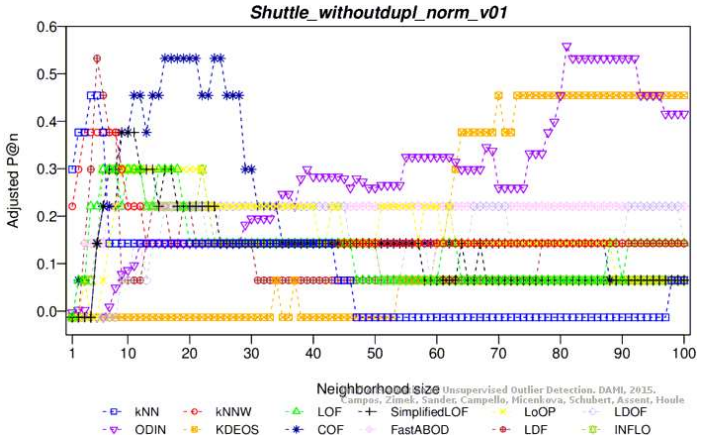
KNNW	**6**	0.38462	0.37662	0.37464	0.36651	0.52941	0.52329	0.98592
KNNW	7	0.38462	0.37662	0.37166	0.36349	0.48649	0.47981	0.98615
LOF	**6**	0.30769	0.29869	0.31305	0.30412	0.56410	0.55844	0.92869
LOF	**9**	0.23077	0.22077	0.37864	0.37056	0.66667	0.66233	0.98962
SimplifiedLOF	**9**	0.38462	0.37662	0.29411	0.28494	0.46154	0.45454	0.98100
SimplifiedLOF	**11**	0.38462	0.37662	0.33752	0.32891	0.53333	0.52727	0.98692
SimplifiedLOF	**15**	0.23077	0.22077	0.31031	0.30135	0.55000	0.54415	0.98546
LoOP	**9**	0.30769	0.29869	0.20080	0.19041	0.35897	0.35064	0.95631
LoOP	**20**	0.30769	0.29869	0.34661	0.33812	0.57895	0.57347	0.98692
LDOF	**16**	0.23077	0.22077	0.12299	0.11159	0.23077	0.22077	0.92469
LDOF	**57**	0.15385	0.14285	0.25882	0.24919	0.47059	0.46371	0.97623
LDOF	**62**	0.23077	0.22077	0.26086	0.25125	0.42424	0.41676	0.97662
LDOF	**77**	0.15385	0.14285	0.25075	0.24101	0.39286	0.38496	0.97754
ODIN	**46**	0.26923	0.25973	0.38724	0.37927	0.55814	0.55240	0.98885
ODIN	**56**	0.33333	0.32467	0.38988	0.38195	0.55556	0.54978	0.98835
ODIN	**81**	0.56410	0.55844	0.36564	0.35739	0.59259	0.58730	0.94096
FastABOD	**7**	0.23077	0.22077	0.19030	0.17978	0.26471	0.25515	0.82708
FastABOD	**82**	0.23077	0.22077	0.23937	0.22948	0.35714	0.34879	0.83646
FastABOD	**96**	0.23077	0.22077	0.24103	0.23117	0.35714	0.34879	0.83754
FastABOD	**100**	0.23077	0.22077	0.24086	0.23099	0.35714	0.34879	0.83815
KDEOS	**67**	0.38462	0.37662	0.29315	0.28396	0.45714	0.45009	0.98108
KDEOS	**70**	0.46154	0.45454	0.31823	0.30937	0.50000	0.49350	0.98062
KDEOS	**98**	0.46154	0.45454	0.42922	0.42180	0.60000	0.59480	0.96577
KDEOS	**100**	0.46154	0.45454	0.43398	0.42662	0.60000	0.59480	0.96600
LDF	**5**	0.53846	0.53246	0.43072	0.42332	0.61538	0.61038	0.97462
LDF	**8**	0.38462	0.37662	0.43793	0.43062	0.70270	0.69884	0.99223
INFLO	**6**	0.30769	0.29869	0.20006	0.18967	0.36667	0.35843	0.88362
INFLO	**13**	0.23077	0.22077	0.31486	0.30596	0.59459	0.58932	0.94100
INFLO	**15**	0.23077	0.22077	0.34284	0.33430	0.59459	0.58932	0.98623
INFLO	**19**	0.23077	0.22077	0.33360	0.32494	0.53659	0.53056	0.98631
COF	**16**	0.53846	0.53246	0.40798	0.40028	0.59459	0.58932	0.98969
COF	**19**	0.53846	0.53246	0.45899	0.45196	0.64865	0.64408	0.99177
COF	**20**	0.53846	0.53246	0.46190	0.45491	0.64865	0.64408	0.99200

Plots

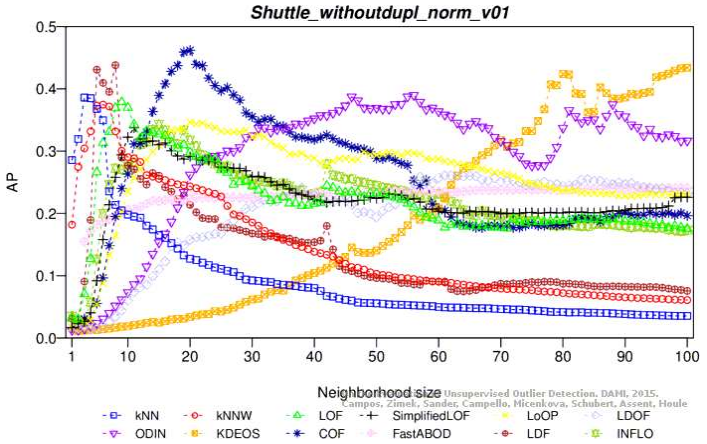
Precision at *n*

Adjusted precision at *n*



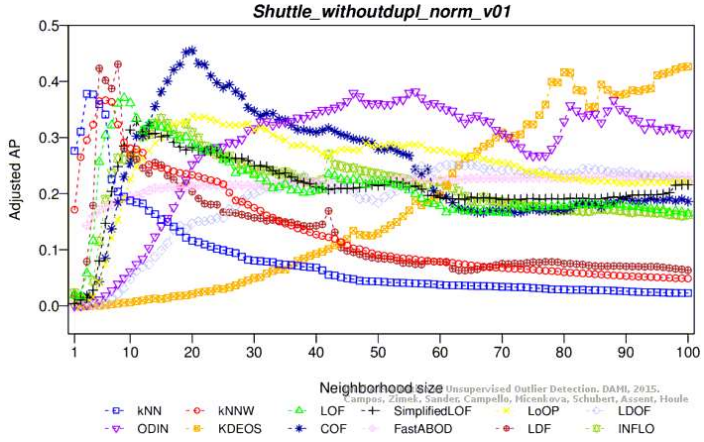


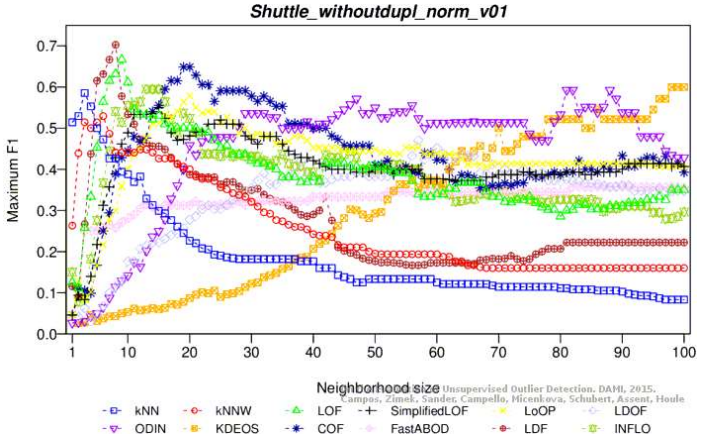
Average precision



Adjusted average precision

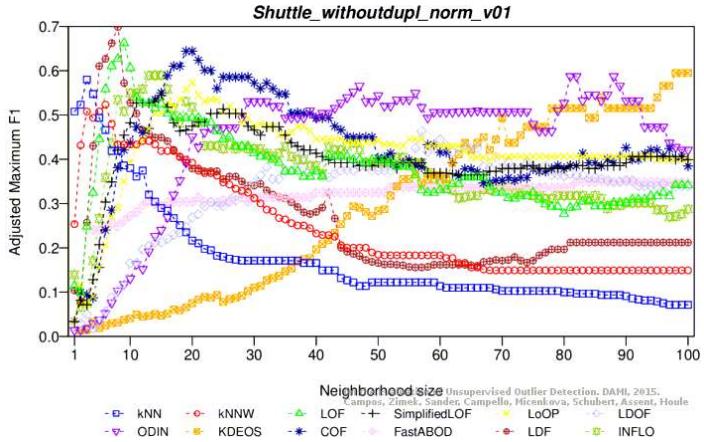
Maximum F1 score

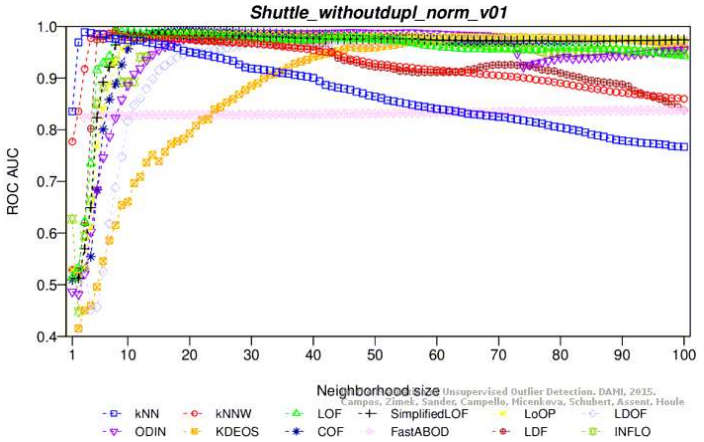




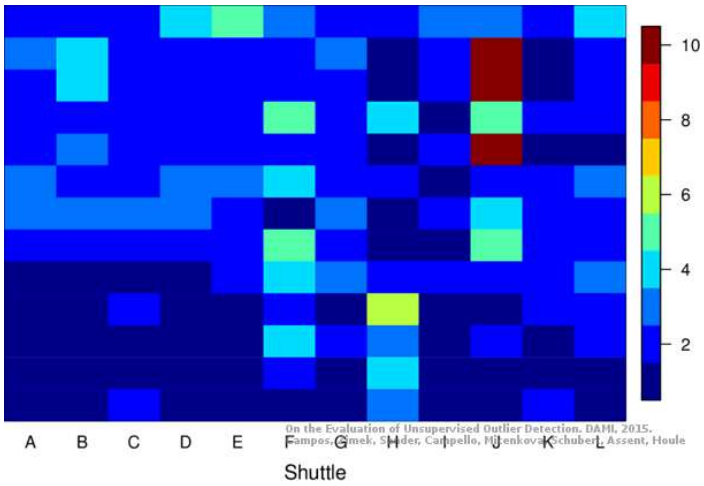
Adjusted maximum F1 score

ROC AUC





Diversity



A: KNN, B: KNNW, C: LOF, D: SimplifiedLOF, E: LoOP, F: LDOF
G: ODIN, H: KDEOS, I: COF, J: FastABOD, K: LDF, L: INFLO

Not normalized, without duplicates

This version contains 9 attributes, 1013 objects, 13 outliers (1.28%)

Download raw algorithm results (8.3 MB) Download raw algorithm evaluation table (44.5 kB)

Best Parameters

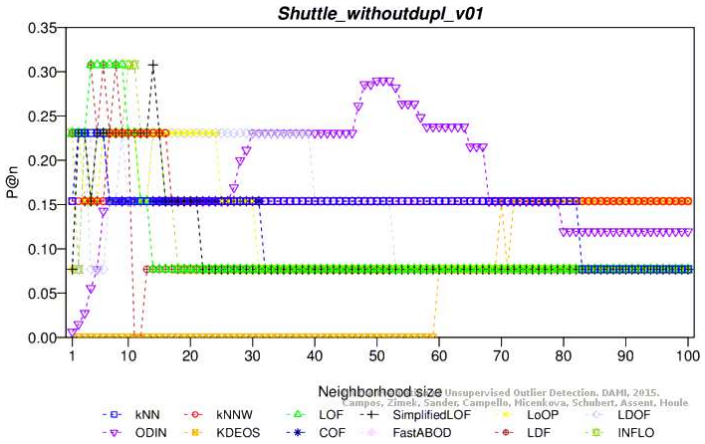
The following table contains the best (overall and per-method) results for each method and evaluation measure (when the same score was achieved twice, only the smallest k is given).
The Maximum F1-Measure is complimentary in addition to the measures in the original publication.

Algorithm	k	P@n	Adj. P@n	AP	Adj. AP	Max-F1	Adj. MF1	ROC AUC
KNN	**2**	0.23077	0.22077	0.14835	0.13728	0.25926	0.24963	0.80992
KNN	**5**	0.23077	0.22077	0.18879	0.17824	0.30380	0.29475	0.96496
KNNW	**7**	0.23077	0.22077	0.15511	0.14413	0.25806	0.24842	0.93515
KNNW	**9**	0.23077	0.22077	0.15912	0.14819	0.25000	0.24025	0.94362
KNNW	**13**	0.23077	0.22077	0.15514	0.14416	0.24000	0.23012	0.94585
LOF	**4**	0.30769	0.29869	0.11507	0.10356	0.30769	0.29869	0.62969
LOF	**21**	0.07692	0.06492	0.18285	0.17223	0.39130	0.38339	0.95715
LOF	**79**	0.07692	0.06492	0.20549	0.19516	0.45455	0.44745	0.93685
LOF	**98**	0.07692	0.06492	0.20225	0.19188	0.47619	0.46938	0.92308
SimplifiedLOF	**14**	0.30769	0.29869	0.17533	0.16461	0.35714	0.34879	0.93385
SimplifiedLOF	**26**	0.07692	0.06492	0.18432	0.17372	0.31579	0.30689	0.96754
SimplifiedLOF	**93**	0.07692	0.06492	0.22147	0.21135	0.47619	0.46938	0.95954
SimplifiedLOF	**95**	0.07692	0.06492	0.22065	0.21052	0.48780	0.48115	0.95746
LoOP	**2**	0.23077	0.22077	0.09070	0.07888	0.28571	0.27643	0.69569
LoOP	**26**	0.15385	0.14285	0.17358	0.16284	0.31579	0.30689	0.95992
LoOP	**97**	0.07692	0.06492	0.20922	0.19894	0.44444	0.43722	0.95369
LoOP	**98**	0.07692	0.06492	0.20959	0.19932	0.44444	0.43722	0.95385

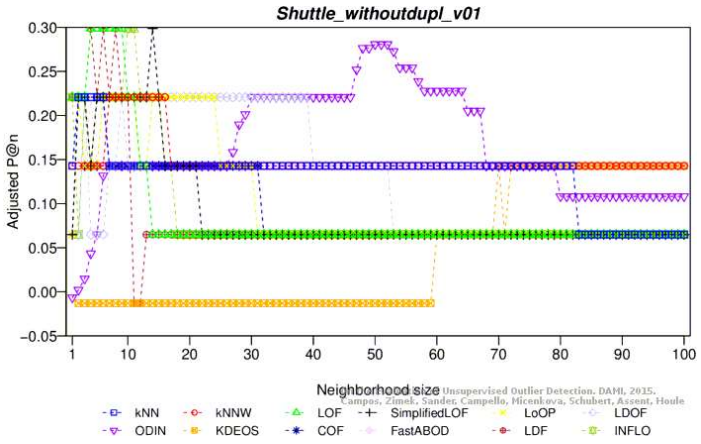
LDOF	**3**	0.23077	0.22077	0.09675	0.08501	0.30000	0.29090	0.51015
LDOF	**47**	0.15385	0.14285	0.18127	0.17063	0.33333	0.32467	0.96377
LDOF	**91**	0.07692	0.06492	0.18936	0.17882	0.40909	0.40141	0.95969
LDOF	**95**	0.07692	0.06492	0.19040	0.17988	0.40000	0.39220	0.95938
ODIN	**47**	0.26154	0.25194	0.25108	0.24135	0.45000	0.44285	0.93788
ODIN	**50**	0.28994	0.28071	0.25128	0.24154	0.43902	0.43173	0.93050
FastABOD	**3**	0.15385	0.14285	0.06976	0.05767	0.18750	0.17694	0.63615
FastABOD	**4**	0.23077	0.22077	0.08641	0.07454	0.24000	0.23012	0.62692
KDEOS	**52**	0.00000	-0.01300	0.07039	0.05830	0.21212	0.20188	0.89346
KDEOS	**70**	0.15385	0.14285	0.07662	0.06462	0.15385	0.14285	0.86031
KDEOS	**84**	0.15385	0.14285	0.11446	0.10295	0.23529	0.22535	0.85262
KDEOS	**89**	0.15385	0.14285	0.15401	0.14301	0.23529	0.22535	0.85615
LDF	**4**	0.30769	0.29869	0.13649	0.12526	0.33333	0.32467	0.63923
LDF	**8**	0.30769	0.29869	0.25542	0.24574	0.45000	0.44285	0.96023
LDF	**10**	0.23077	0.22077	0.21041	0.20014	0.40816	0.40047	0.96662
LDF	**48**	0.07692	0.06492	0.19742	0.18699	0.46154	0.45454	0.92331
INFLO	**10**	0.30769	0.29869	0.12542	0.11405	0.30769	0.29869	0.83577
INFLO	**50**	0.07692	0.06492	0.19598	0.18553	0.36667	0.35843	0.96585
INFLO	**68**	0.07692	0.06492	0.20272	0.19236	0.39024	0.38232	0.96054
INFLO	**97**	0.07692	0.06492	0.19693	0.18649	0.46512	0.45816	0.93685
COF	**2**	0.23077	0.22077	0.07538	0.06336	0.26087	0.25126	0.61927
COF	**5**	0.15385	0.14285	0.09903	0.08732	0.27586	0.26645	0.66462
COF	**26**	0.15385	0.14285	0.15235	0.14133	0.25000	0.24025	0.95038

Plots

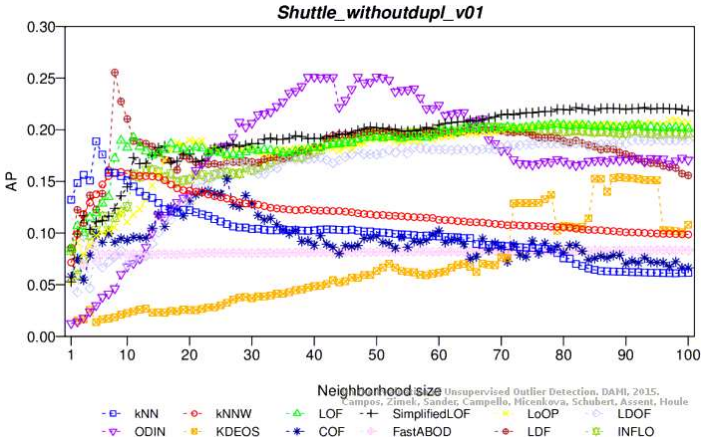
Precision at *n*



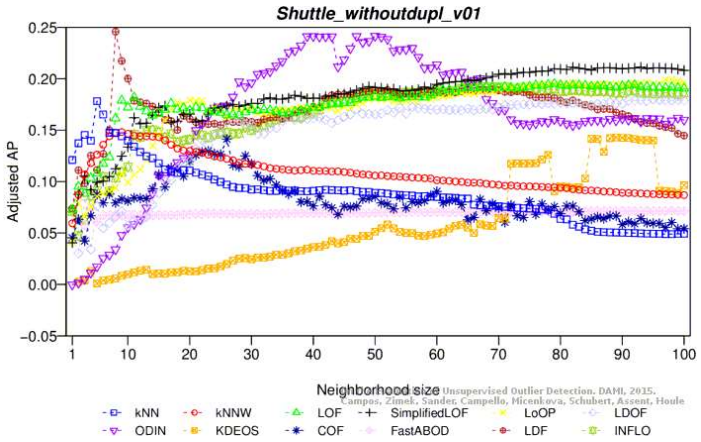
Adjusted precision at *n*



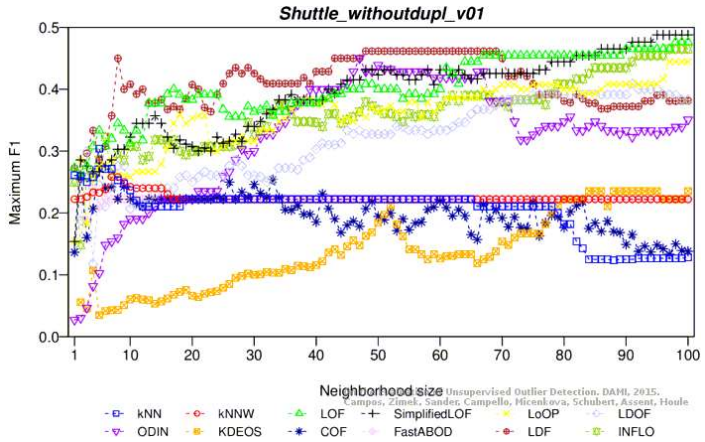
Average precision



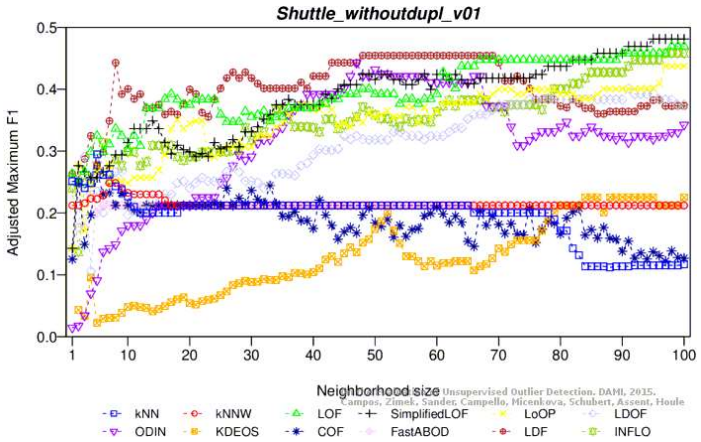
Adjusted average precision



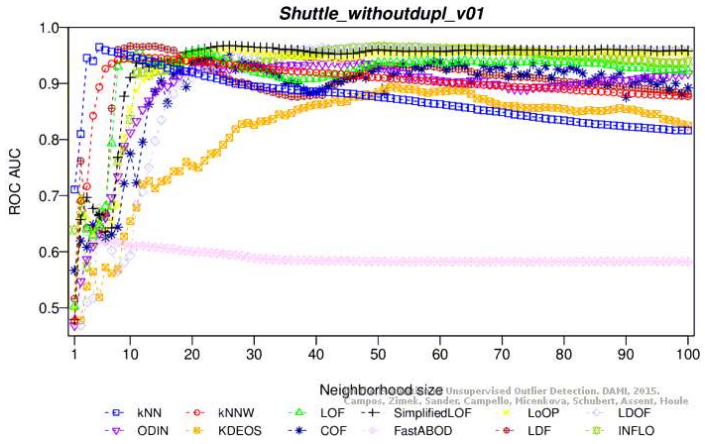
Maximum F1 score



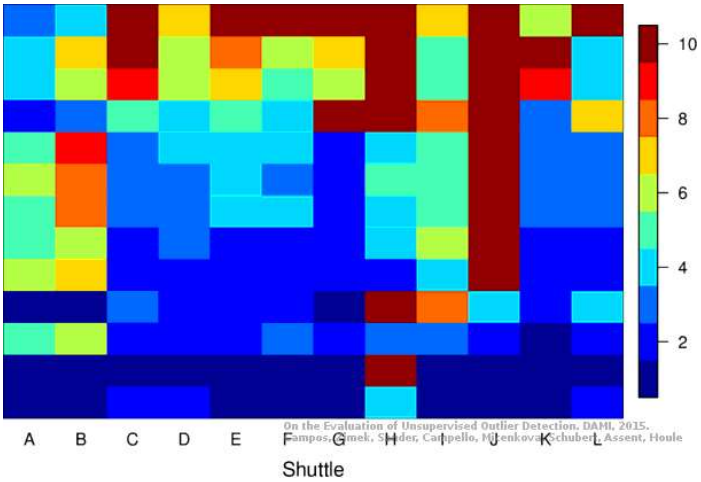
Adjusted maximum F1 score



ROC AUC



Diversity



A: KNN, B: KNNW, C: LOF, D: SimplifiedLOF, E: LoOP, F: LDOF
G: ODIN, H: KDEOS, I: COF, J: FastABOD, K: LDF, L: INFLO

File generated: 2016-07-05T21:48:12