barcelona|gse
graduate school of economics

Master Thesis

Topic:

Unsupervised learning in decision making

Author:	Domagoj Fizulic
	Felix Gutmann
Student number:	125604
	125584
Program:	M.S. Data Science
E-Mail:	domagoj.fizulic@barcelonagse.eu
	felix.gutmann@barcelonagse.eu

I Table of Contents

II List of Figures

III List of Tables

IV List of Listings

V List of mathematical symbols

Symbol	Meaning
a_t	Action at time t
$Q(a)_t$	Value function at time t
ϵ	Probability of exploration in epsilon greedy
α	Learning rate
τ	Softmax parameter
X	Random variable
$H(X)$	Entropy of a discrete random variable X
$d(\cdot, \cdot)$	Distance Function
$S(\cdot, \cdot)$	Similarity Function
$\mathbb{R}_0^+$	Positive real numbers including zero
$\mathcal{X}$	Data set
$\mathbf{W}$	Weighted adjacency matrix
d_i	Degree of node i
$\mathbf{D}$	Diagonal matrix of degrees
$\mathbf{L}$	Graph laplacian

VI List of abbreviations

Abbreviations	Description
IGT	Iowa gambling Task
RL	Reinforcement learning
CH	Choices
BBC	Blockwise bad choices
ENT	Cumulative entropy
BENT	Blockwise entropy
CC	Concatenated
NMI	Normalized mutual information score
ARI	Adjusted rand index
VM	V-Measure score

1 Introduction and conceptual approach

Decision-making is a cognitive process of selecting an option from a set of possible alternatives based on certain criteria.[Wang and Ruhe, 2007]

When analysing decision-making as a continuous process of interaction with the environment, learning becomes an important aspect.

Learning is a complex procedure. The learning procedure can affected due to social conditions... Due to their mental ability people should show different learning behaviour. The decision making process is studied by monitoring peoples sequential choices in a controlled experiment environment.
In *supervised learning* data are predicted by training a classifier based on features and corresponding categories. In comparison to that in *unsupervised learning* we don't know the ground truth. The objective is to discover natural clustering behaviour in the data. [Murphy, 2012, page 9 et. seqq.] and group objects into subsets, such that objects in those subsets are more closely related to each other [Hastie et al., 2001, page 501]. A vast class of clustering algorithms based on different approaches are proposed in the literature (e.g. hierarchical, and optimization based clustering). A detailed overview of the algorithms and corresponding distance and similarity concepts is provided in section 3.3.
This paper investigates, whether such unsupervised learning techniques can be used in the context of human decision making process to identify latent grouping. This paper is operates in the intersection of machine learning and cognitive science. To our knowledge this particular setting has not been studied before.
To approach this issue, we first set up a reinforcement learning based simulation framework to study theoretical boundaries of several clustering techniques and when they are applicable. Subsequently, we test our chosen methods on several real world data sets. Corresponding to our simulation framework we first apply clustering algorithms on a real n-armed bandit experiment.[1] A widely used approach to monitor human decision making process is the *Iowa gambling task*, where participants try maximize rewards by choosing from different deck with different reward structures.[2] Within that framework we analyse data from people with different criminal profiles. Furthermore, we use data from cocaine abusers.

[1] A detailed introduction is given in section 3.2
[2] There are existing several slightly different versions of test. [Steingroever et al., 2015] provides a data collection for from several sources giving a broad overview of different variations of the test.

The report has the following structure. In section two we provide a short overview of related literature in the field.
The following section three is dedicated to the theoretical foundation and the simulation. Therefore, we first provide knowledge of reinforcement and line out our experiment design in more detail. This section also includes an overview of applied algorithms, similarity and distance concepts. Finally, we study their simulation performance and try to identify situations where they are applicable.
The rest of this paper is dedicated to the data sets. Our data sets are studied before from a cognitive science perspective. We introduce relevant concepts such as expectancy valence models and compare the findings of the authors to the results we obtain from the clustering algorithms.
We try to keep the scope of this paper tight and there are some open questions left. Thus we dedicated another section to discussing some possible extensions. We close this paper with a final summary of our results. A mathematical formulation of the applied algorithms, applied algorithms, similarity measures and cluster evaluation techniques.[3]

2 Relevant Literature

There exists a rich literature in cognitive science on identifying different behavioural groups. As mentioned in the introduction a commonly applied tool is the Iowa Gambling Task experiment. It has been shown that individuals with pre-frontal brain damage and decision-making defects continue to choose disadvantageously even after they learned the optimal strategy [Bechara et al., 1997].
A broad overview on various results in the field can be found in [Steingroever et al., 2013]. Several studies identify especially specific drug-user groups, e.g. cocaine addicts [Stout et al., 2004], chronic cannabis users [Fridberg et al., 2010], heavy alcohol users(heavy drinkers) [Gullo and Stieger, 2011]. Furthermore, extensive set of research is focused around particular mental disabilities, e.g. Asperger's disorder [Johnson et al., 2006], psychopathic tendencies [Blair et al., 2001], bipolar disorder [Brambilla et al., 2012], schizophrenia [Martino et al., 2007] pathological gambling disorder [Cavedini et al., 2002], attention-deficit-hyperactivity disorder [Nirit Agaya, Eldad Yechiama, Ziv Carmelb, 2010]. Most popular reinforcement learning models for identifying behavioural differences between different disorders are Expectancy Valence model [Busemeyer and Stout, 2002] and Prospect Valence Learning model [Ahn et al., 2008].

[3]Since our data are fairly small we will not discuss complexity of the algorithms.

3 Theoretical Background and simulation experiments

This section is dedicated to a more detailed outline of our analysis approach and the results of our simulation experiments. The data we are analysing are gathered by observing peoples decisions over time. Hence, our data set are in the form $N \times M$ data set, where a row N is the number of individuals and M is the number of trials in the experiments. The data for each individual can be seen as a categorical time series. In terms of modeling there are two challenges. Some algorithms relying on euclidean distance while others operating on similarity concepts. On one hand we introduce how repress those data and introduce related distance and similarity concepts for the algorithms. The section has a rather qualitative character. The appendix provides in more detail a mathematical background to applied algorithms, distance concepts and related clustering evaluation techniques.

3.1 Experiment design and problem formulation

Figure 1 depicts our simulation design.[4] The objective is to obtain a set of sequential choices for a given parameter setting of the artificial agent. We first generate a set of rewards by sampling from $n-$vectors from a normal distribution ("*multi arm bandit*"). The agent processes those rewards by sequentially choosing from those $n-$vectors. We repeat this procedure for several parameter settings for the agents and keep track of those choices which will define our data set.

3.2 Reinforcement Learning background and multi arm bandits

In the following we provide some basic background of *Reinforcement Learning* (RL). The following definitions coming from [Sutton and Barto, 2012, chapter 1 and 2]. RL is a branch of *machine learning* try model how an artificial agents interact with its environment and learns from the process over time.

In our particular setting the agent is confronted with the task of choosing sequentially from a set of possible choices. The doesn't have any knowledge about the system a priori. Therefore, it has to learn the nature of the system by sequentially interacting with its environment and keeping tack of the obtained information. Because of the lack of examples it has to explore different possible actions to identify the best action.

[4] We implemented related coding for this project mainly in python. The code can be found on our github repository.

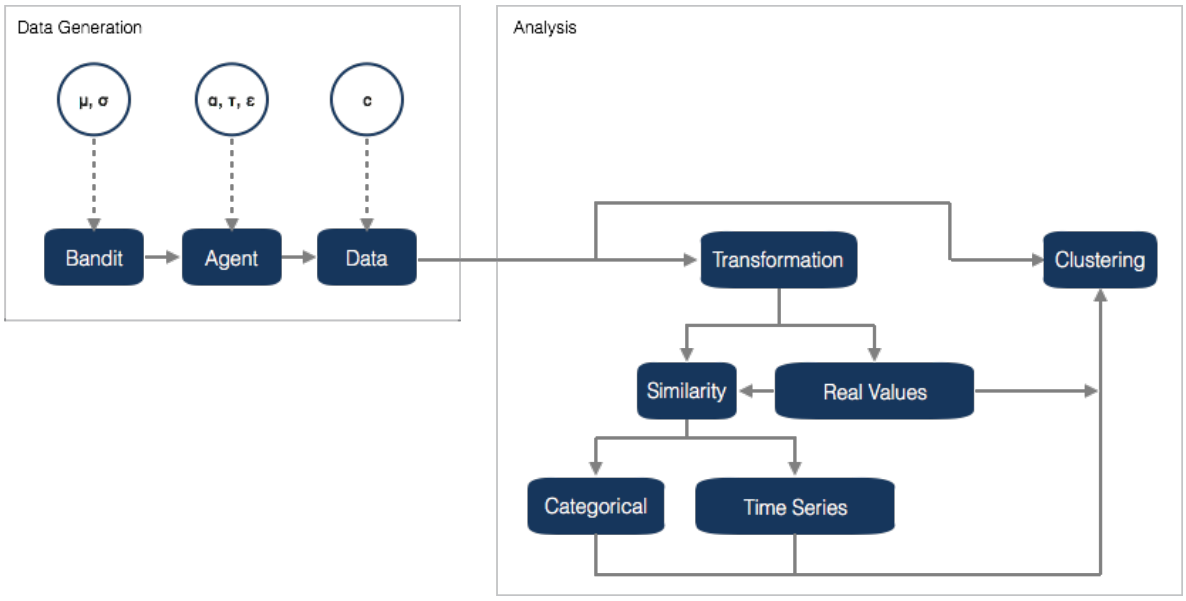


Figure 1: Flowchart experiment design

Hence, it is useful to deviate from time to time from the current optimal strategy. It is necessary for the agent to keep track and updates the value of a each state it discovered so far. This is done by defining a value function for each state or action a. Denote the value function of an action a as $Q_t(a)$ and define it as:

$$Q_t(a) := \frac{R_1 + R_2 + \cdots + R_{K_\alpha}}{K_\alpha} \tag{1}$$

where R denotes rewards. Hence the value function is the average over the rewards for a given state. The main task of the agent is to balance *exploration* and *exploitation* of the environment. There are two basic approaches to model this trade-off; An *"Epsilon-Greedy"* action selection method and *"Softmax"* selection method.
Considering epsilon greedy action selection method: The rule is to select the next action with the highest value function. However, to model exploration we introduce a random element to deviate from that greedy strategy with a certain probability denote by ϵ. In general we can define the next action selected by the epsilon greedy strategy as:

$$a_{t+1} = \begin{cases} \text{random action} & \text{, with probability } \epsilon \\ \arg\max_i Q_t(i) & \text{, with probability } 1-\epsilon \end{cases}$$

In the softmax action selection method each next action is sampled from with a certain probability coming *Boltzmann Distribution).* The probability for action a is computed

by:

$$P(a_t|X) = \frac{e^{\frac{Q_t(a)}{\tau}}}{\sum_i^K e^{\frac{Q_t(i)}{\tau}}}$$

In each iteration the next action of the agent is drawn with probability p_a:

$$a_{t+1} \sim p_a$$

After selecting an action the agent is updating its believe of the chosen action. One can show that equation (1) can be expressed in the following way:

$$Q(a)_{k+1} = Q(a)_k + \alpha \left[R(a)_k - Q(a)_k\right],$$

where α is a is the non negative *learning rate* defining how much the current action is affecting the believes.

3.3 Unsupervised learning methods, data handling and similarity

As mentioned we find two main challenges. First, our data have a categorical nature. Furthermore, the learning process also imposes a time series dependence on the data. A well studied approach clustering such data are *hidden markov models*. For example a decent research on that can be found in [Pamminger, 2007], [Pamminger and Fruhwirth-Schnatter, 2009] and [Pamminger and Fruehwirth-Schnattery, 2010]. However, we focus our attention only on partition based clustering algorithms.
Some of applied algorithms operate on different distance or similarity concepts. Given our data we have to think carefully think about distance and similarity measures to respect the nature of our data and algorithms.[5] In extension to only considering raw choices we might try to re-express our data to discriminate them according to different purposes. The first approach is to map the series of choices to a real valued series. One way is to consider *Shannon's Entropy* introduced by [Shannon, 1948] based on the empirical probability of the choices. For a discrete random variable X with probability p the entropy is defined as [MacKay, 2005, page 32]:

$$H(X) :== -\sum_{i=1}^{N} p_i \log_2 p_i \tag{2}$$

[5] A formal definition of the distance and similarity can be found in the appendix

Entropy gives measure on how random a random variable behaves.[6] Within the entropy framework we consider to types of entropies. For each time step we compute the entropy using choices done so far. We call this *cumulative entropy*. Furthermore, we might want to observe more clearly how individuals adapt their behaviour over time. Considering an experiment with 100 trials we then compute the entropy for set of ten choices. We call this *blockwise entropy*. Mapping choices to and entropy based data set is therefore aims to discriminate individuals by their level randomness in their behaviour.
A second approach is based on the experimental setting. We know within the framework there are a set of choices which are disadvantageous for the participant. Following e.g. [Yechiam et al., 2008] or [Ahn et al., 2008], we then compute block wise the ratio of disadvantageous choices. We call this *disadvantageous choice*.

Within our analysis we consider a broad selection of several clustering techniques and similarity concepts. Figure 2 shows an overview of the algorithms and their corresponding distance/similarity requirements. A technical description for all of them is provided in the appendix. According to our different data types and for our different algorithms

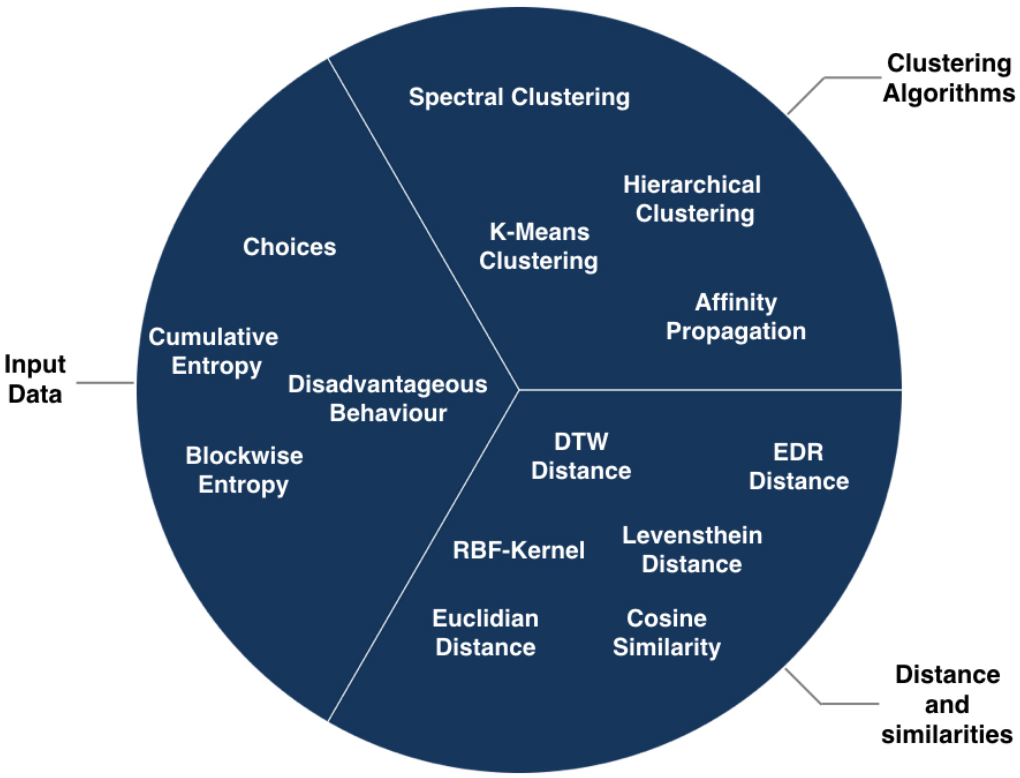


Figure 2: Input data, algorithms and proximity measures

[6] In case we get a value for X such that $p_i = 0$, we set $H(X) = 0$ [Bishop, 2006, page 49]

we apply the following distance and similarity concepts.[7]

3.4 Simulation results

The following table shows a snippet of our simulation results. We have a broad range o similarity measures. Some of them are task specific meaning being designed for either categorical or time series data or both.[8]

We also considered some specilised articles. [Garcia-Magarinos and Vilar, 2014] provide an algorithm to cluster categorical time series.

[7]Note that distance and similarity can be converted according to equation (3) in the appendix

[8]For example Levenstein distance (or edit distance) is a basic way to measure similarity between categorical sequences [**?**, page 1]Richter or [Gabadinho et al., 2009, page 2], however some authors stating that it might perform poorly in their task [Ren et al., 2011, page 3] or a poor measure at all [Morzy et al., , page 5]

Tau	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1, 0.3}	Spectral	DTW				x		0.136 (0.134)	0.094 (0.146)	0.135 (0.134)
	Spectral	EDR			x			0.134 (0.135)	0.097 (0.130)	0.134 (0.135)
	Spectral	Levenshtein	x					0.093 (0.134)	0.050 (0.128)	0.092 (0.134)
	Spectral	Cosine		x				0.113 (0.134)	0.082 (0.150)	0.113 (0.134)
	K-Means	Euclidian					x	0.093 (0.135)	0.050 (0.130)	0.092 (0.135)
{0.1, 0.5}	Spectral	RBF				x		0.331 (0.185)	0.328 (0.208)	0.331 (0.186)
	Spectral	RBF			x			0.322 (0.197)	0.301 (0.221)	0.321 (0.197)
	Spectral	EDR			x			0.315 (0.189)	0.328 (0.206)	0.315 (0.189)
	K-Means	Euclidian			x			0.313 (0.187)	0.261 (0.213)	0.312 (0.187)
	Spectral	EDR		x				0.226 (0.127)	0.131 (0.113)	0.223 (0.127)
	Spectral	Levenshtein	x					0.219 (0.144)	0.126 (0.121)	0.217 (0.144)
	K-Means	Euclidian					x	0.208 (0.143)	0.124 (0.122)	0.207 (0.143)
{0.1, 0.7}	Spectral	RBF				x		0.642 (0.200)	0.653 (0.206)	0.641 (0.200)
	Spectral	DTW			x			0.580 (0.209)	0.611 (0.205)	0.580 (0.201)
	K-Means	Euclidian			x			0.533 (0.248)	0.539 (0.275)	0.533 (0.249)
	K-Means	Euclidian					x	0.382 (0.181)	0.309 (0.237)	0.381 (0.180)
	Spectral	Levenshtein	x					0.350 (0.114)	0.241 (0.163)	0.347 (0.116)
	Spectral	Euclidian		x				0.306 (0.155)	0.198 (0.179)	0.303 (0.156)
{0.1, 1}	Spectral	DTW				x		0.938 (0.131)	0.942 (0.126)	0.938 (0.131)
	Spectral	Euclidian				x		0.811 (0.136)	0.833 (0.129)	0.811 (0.136)
	Ward	Euclidian				x		0.797 (0.205)	0.796 (0.218)	0.797 (0.205)
	Spectral	DTW			x			0.765 (0.197)	0.791 (0.184)	0.765 (0.197)
	K-Means	Euclidian			x			0.757 (0.180)	0.774 (0.185)	0.757 (0.180)
	K-Means	Euclidian					x	0.796 (0.218)	0.687 (0.265)	0.696 (0.219)
	Spectral	Overlap	x					0.600 (0.299)	0.571 (0.353)	0.598 (0.302)
	Spectral	Euclidian		x				0.335 (0.091)	0.220 (0.138)	0.332 (0.093)

Table 1: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 2, 4\}$, $\sigma = \{1, 1, 1\}$, $\alpha = 0.1$)

4 Data Analysis

Our simulation result suggested that we can cluster the decision behaviour under some give constraints. We apply our approach to different real data sets. In total we are considering three different data sets. As mentioned in the introduction the Iowa gambling task is popular way to monitor decision and learning process of individuals. s In the simulation setting we initially define participants with a certain set of parameters. There exist techniques to estimate those parameters from actual data.

4.1 Recover cognitive parameters

4.2 Prison data and approach

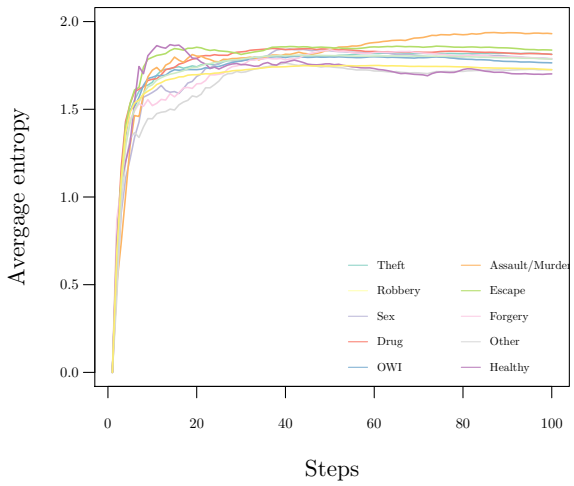
With acknowledgment to [HERE WE ADD THE GUY] we were provided with experimental data from [Yechiam et al., 2008]. We try to apply our methods to cluster different groups in the data. The participants had to perform the Iowa gambling task (detailed description see later on). [Steingroever et al., 2015] assembled a collection of data of healthy individuals performing the IGT test from various sources. We extended the data set we got with a set of healthy people, which we hope will behave significantly different than the individuals with criminal profile.

4.2.1 Data summary and experimental design

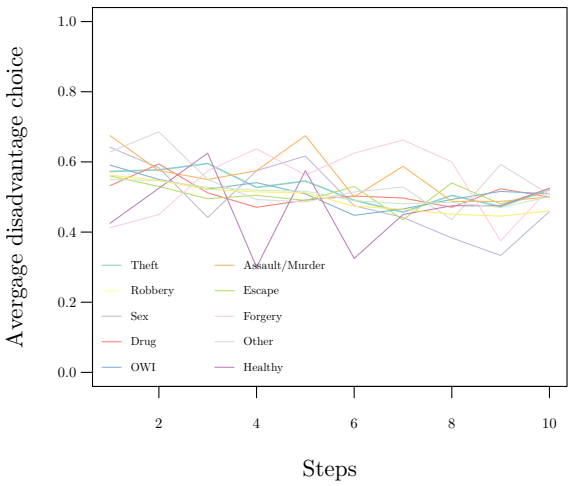
In particular we have data of 96 individuals with different criminal profile. Table 2 gives a broad summary of some demographics.

Propositi had to perform the *Iowa gambling task* (IGT), where they have to pick sequentially a card from four different decks. Two decks have distributions with negative expectations while two have positive. However, each deck has its own variance and a set of profits.
Figure 3 (a) shows the average cumulative entropy averaged across groups. We observe a random behaviour independent from group affiliation. However, normal people show the least random behaviour. Furthermore, Figure 3 (b) shows the average people of negative choices people chose over steps of ten periods averaged across groups.

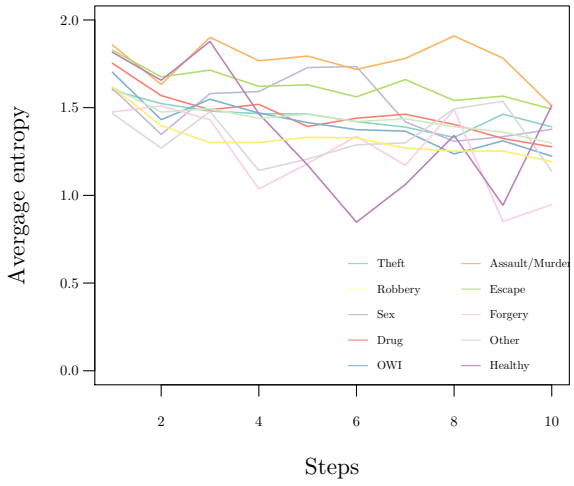
Yechiam et al. found three clusters by using the attention to recent outcomes(ARO) and attention to gains(AG) parameters from the Expectancy Valance model. The most distinct group were the Robbery convicts with the only negative attention to gains mean at -0.36 and the highest attention to recent outcomes mean 0.57. The second



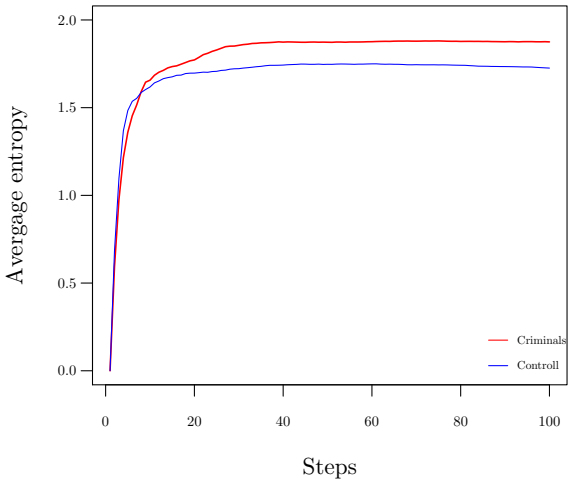
(a) Average Entropy across groups



(b) Blockwise picks from disadvantageous deck



(c) Average Entropy across groups



(d) Blockwise picks from disadvantageous deck

Figure 3: Average py and disadvantageous behaviour

Crime	**Count**	**Age**	**TABE Score**	**Education**	**Beta IQ**
Theft/Burglary	22	25.36 (7.03)	11.09 (1.29)	7.38 (3.34)	92.91 (14.37)
Robbery	6	24.17 (9.83)	11.00 (0.63)	9.22 (3.30)	96.50 (7.58)
Sex	17	33.41 (13.59)	10.97 (1.47)	9.15 (2.98)	99.65 (11.74)
Drug	22	30.91 (10.11)	11.64 (1.85)	9.06 (2.70)	100.36 (12.92)
OWI	4	38.75 (7.27)	10.88 (1.93)	7.12 (1.17)	94.25 (10.40)
Assault	10	27.20 (8.77)	12.30 (2.41)	7.62 (2.28)	94.50 (11.29)
Escape/ Failure To Appear	4	2.008 (5.60)	11.00 (1.35)	7.78 (3.21)	96.50 (14.18)
Vandalism	1	18.00 (NA)	11.00 (NA)	9.40 (NA)	90.00 (NA)
Forgery	7	34.57 (13.14)	10.93 (5.15)	9.83 (3.82)	100.71 (11.01)
Probabiton	1	38.00 (NA)	12.00 (NA)	6.30 (NA)	92.00 (NA)
Other	2	35.00 (9.90)	11.50 (0.00)	9.20 (4.67)	95.00 (5.66)

Table 2: Summary prison data (means with standard deviation in parenthesis)

cluster is made of assault and murder convicts with ARO of 0.26 and AG of 0.1. The third cluster is formed of all the remaining prisoner groups with ARO means between 0 and -0.1 and AG between 0.1 and 0.2.

4.2.2 Prison data results

Following the last section we provide a summary of our results.

Following the clusters given by the EV model, we tried pairwise clustering between robbers and other prisoner groups and assault offenders and other prisoner groups.

4.3 Multi-arm Bandit Data

The second data set we are considering are related to [Stojic et al., 2015]. In this setting we consider a 20-arm bandits situation. The data were gathered online were users were compensated with small amount of money. Four different distributional settings were given to different people. In total the data sets consists of 429 participants divided in 199 female and 229 male participants.[9] The average age 33.04 with standard deviation 11.75. Furthermore, the participants overall have a stronger higher education background where 261 participants have college degree and and 39 with graduate degree and Phd. 127 have a high school degree and 2 declined to answer.

[9]One did not wish to answer

We tried to identify different clusterings according to demographics within those four sub experiments. Our results show not a significantly different behaviour.

4.4 Huntington Data

[Stout et al., 2001] conducted research on behaviour of patients with Huntington disease (HD) and Parkinson disease (PD).[10] In the study there were considered 14 participants with HD and 22 participants with PD. They participants are required to don't have ongoing drug problems (including alcohol), free of other major diseases (physical or psychological nature). A group of 33 people serves as a control group. This group can further distinguished in younger (YHC) and older control participants (OHC). Table 3 gives a basic summary of the demographic of the participants.

Demographic	Hunting Disease	Parkinson Disease	YHC	OHF
Age	44.60 (11.70)	66.00 (8.30)	45.30 (10.60)	65.50 (10.70)
Education (years)	15.30 (2.30)	14.20 (2.90)	14.30 (2.10)	14.70 (2.40)
Sex	130.60 (10.10)	131.70 (7.60)	139.50 (2.20)	138.00 (4.60)
Years since diagnosis	4.10 (2.80)	7.70 (5.50)	NA (NA)	NA (NA)
Estimated age at diagnosis	40.50 (10.80)	58.30 (7.60)	NA (NA)	NA (NA)

Table 3: Shortened summary of HD, PD and control people (means with standard deviations in parenthesis)
Source: Summary of [Stout et al., 2001, page 3]

4.5 Cocaine Abusers data

We have data from several individuals of cocaine abusers. There are 12 individuals performing the IGT. The control group consist out of 14 participants. Candidates among the drug abusers were selected as active users with additional drug abusing past and without any known additional mental illness [**?**]. Table **??** gives a summary of demographic profile.

[10] The Huntington Disease is an illness of the central nervous system. The are are a wide range of associated symptoms including physical symptoms like uncontrollable muscular movements and clumsiness and psychological symptoms like minored concentration, short term memory lapses etc . Moreover a related possible symptom is Parkinson disease [Huntington's-Disease-Association, 2016]. Parkinson is chronic and progressive movement disorder. Also Parkinson goes along with physical, such as tremors and rigidity and psychological symptoms, such as memory problems, minored speed of thinking fear and anxiety (among others) [Parkinson's-Disease-Foundation, 2016]

demographic	Drug abusers	Control Group
Share of men	79%	100%
Age	36.90 (10.30)	30.00 (6.10)
Estimated IQ	105.00 (7.62)	93.70 (10.30)

Table 4: Demographic summary of cocaine abusers (means with standard deviations in parenthesis)

4.6 Cocaine Abusers data results

5 Discussion of results and possible extensions

Our analysis so far showed that people are not separate themselves. In general we assume that healthy participants and those with assumed decision making deficits show significantly different behaviour. However, we observe that their behaviour seem to be quite similar given our data and applied mappings. Furthermore , in most of the applied unsupervised techniques we as analysts have to to set the number of clusters we assume to be in the data (so in our case two for control group and patients with habits). We apply another algorithm called affinity propagation, which identifies the number of clusters itself (algorithm formulation see appendix). In general we find that the algorithm is assign , which suggest that there more natural clusters in the data than the one we assume due to their status labeled as healthy and ill.

6 Conclusion

7 List of Literature

[Ahn et al., 2008] Ahn, W.-K., Busemeyer, J. R., Wagenmakers, E.-J., and Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science*, 32(8):1376–1402.

[Bechara et al., 1997] Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1997). Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*, 275:1293–1295.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[Blair et al., 2001] Blair, R. J., Colledge, E., and Mitchell, D. G. V. (2001). Somatic markers and response reversal: is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Child Psychology*, 29(6):499–511.

[Boriah et al., 2008] Boriah, S., Chandola, V., and Kumar, V. (2008). Similarity measures for categorical data: A comparative evaluation. *SIAM*.

[Brambilla et al., 2012] Brambilla, P., Perlini, C., Bellani, M., Tomelleri, L., Ferro, a., Cerruti, S., Marinelli, V., Rambaldelli, G., Christodoulou, T., Jogia, J., Dima, D., Tansella, M., Balestrieri, M., and Frangou, S. (2012). Increased salience of gains versus decreased associative learning differentiate bipolar disorder from schizophrenia during incentive decision making. *Psychological Medicine*, pages 1–10.

[Brusco and Köhn, 2008] Brusco, M. J. and Köhn, H.-F. (2008). Clustering by passing messages between data points. *Science (New York, N.Y.)*, 319(5864):726; author reply 726.

[Busemeyer and Stout, 2002] Busemeyer, J. R. and Stout, J. C. (2002). L8 A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological assessment*, 14(3):253–262.

[Cavedini et al., 2002] Cavedini, P., Riboldi, G., Keller, R., D'Annucci, A., and Bellodi, L. (2002). Frontal lobe dysfunction in pathological gambling patients. *Biological Psychiatry*, 51(4):334–341.

[Everitt et al., 2009] Everitt, B. S., Landau, S., and Leese, M. (2009). *Cluster Analysis*. Wiley Publishing, 4th edition.

[Fratev et al., 1979] Fratev, F., Polansky, O. E., Mehlhorn, A., and Monev, V. (1979).

Application of distance and similarity measures. The comparison of molecular electronic structures in arbitrary electronic states. *Journal of Molecular Structure*, 56(C):245–253.

[Fridberg et al., 2010] Fridberg, D. J., Queller, S., Ahn, W. Y., Kim, W., Bishara, A. J., Busemeyer, J. R., Porrino, L., and Stout, J. C. (2010). Cognitive mechanisms underlying risky decision-making in chronic cannabis users. *Journal of Mathematical Psychology*, 54(1):28–38.

[Gabadinho et al., 2009] Gabadinho, A., Ritschard, G., Studer, M., and Nicolas, S. M. (2009). SUMMARIZING SETS OF CATEGORICAL SEQUENCES Selecting and visualizing representative sequences. *Methods*, pages 6–8.

[Garcia-Magarinos and Vilar, 2014] Garcia-Magarinos, M. and Vilar, J. A. (2014). *A framework for dissimilarity-based partitioning clustering of categorical time series*, volume 29.

[Gullo and Stieger, 2011] Gullo, M. J. and Stieger, A. A. (2011). Anticipatory stress restores decision-making deficits in heavy drinkers by increasing sensitivity to losses. *Drug and Alcohol Dependence*, 117(2-3):204–210.

[Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

[Hubert and Arabie, 1985] Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.

[Huntington's-Disease-Association, 2016] Huntington's-Disease-Association (2016). About huntington's disease.

[Johnson et al., 2006] Johnson, S. a., Yechiam, E., Murphy, R. R., Queller, S., and Stout, J. C. (2006). Motivational processes and autonomic responsivity in Asperger's disorder: evidence from the Iowa Gambling Task. *Journal of the International Neuropsychological Society*, 12:668–676.

[Luxburg, 2007] Luxburg, U. V. (2007). A Tutorial on Spectral Clustering. (March):1–32.

[MacKay, 2005] MacKay, D. J. C. (2005). *Information Theory, Inference, and Learning Algorithms David J.C. MacKay*, volume 100.

[Martino et al., 2007] Martino, D. J., Bucay, D., Butman, J. T., and Allegri,

R. F. (2007). Neuropsychological frontal impairments and negative symptoms in schizophrenia. *Psychiatry Research*, 152(2-3):121–128.

[Morzy et al.,] Morzy, T., Wojciechowski, M., and Zakrzewicz, M. Clustering sequences of categorical values.

[Murphy, 2012] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

[Nirit Agaya, Eldad Yechiama, Ziv Carmelb, 2010] Nirit Agaya, Eldad Yechiama, Ziv Carmelb, Y. L. (2010). Non-specific effects of Methylphenidate (Ritalin) on cognitive ability and decision- making of ADHD and healthy adults. (972).

[Pamminger, 2007] Pamminger, C. (2007). Bayesian Clustering of Categorical Time Series Using Finite Mixtures of Markov Chain Models Bayesian Clustering of Categorical Time Series Using Finite Mixtures of Markov Chain Models.

[Pamminger and Fruehwirth-Schnattery, 2010] Pamminger, C. and Fruehwirth-Schnattery, S. (2010). Model-based clustering of categorical time series. *Bayesian Analysis*, 5(2):345–368.

[Pamminger and Fruhwirth-Schnatter, 2009] Pamminger, C. and Fruhwirth-Schnatter, S. (2009). Bayesian Clustering of Categorical Time Series Using Finite Mixtures of Markov Chain Models. 43(0907).

[Parkinson's-Disease-Foundation, 2016] Parkinson's-Disease-Foundation (2016). About parkinson's disease.

[Ren et al., 2011] Ren, J., Cao, S., and Hu, C. (2011). A Hierarchical Clustering Algorithm Based on Dynamic Programming for Categorical Sequences. 5:1575–1581.

[Rosenberg and Hirschberg, 2007] Rosenberg, A. and Hirschberg, J. (2007). V-measure: A conditional entropy-based external cluster evaluation measure. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1(June):410–420.

[Serra and Arcos, 2014] Serra, J. and Arcos, J. L. (2014). An empirical evaluation of similarity measures for time series classification. *Knowledge-Based Systems*, 67:305–314.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(July 1928):379–423.

[Shirali and Vasudeva, 2005] Shirali, S. and Vasudeva, H. L. (2005). *Metric Spaces*. Springer.

[Shlens, 2014] Shlens, J. (2014). A Tutorial on Principal Component Analysis. *ArXiv*, pages 1–13.

[Steingroever et al., 2015] Steingroever, H., Fridberg, D. J., Horstmann, A., Kjome, K. L., Kumari, V., Lane, S. D., Maia, T. V., Mcclelland, J. L., Pachur, T., Premkumar, P., Stout, J. C., and Wetzels, R. (2015). Data from 617 Healthy Participants Performing the Iowa Gambling Task : A "Many Labs" Collaboration. *Journal of Open Psychology Data*.

[Steingroever et al., 2013] Steingroever, H., Wetzels, R., and Wagenmakers, E.-J. (2013). Absolute Performance of Reinforcement-Learning Models for the Iowa Gambling Task. *Decision*.

[Stojic et al., 2015] Stojic, H., Analytis, P. P., and Speekenbrink, M. (2015). Human behavior in contextual multi-armed bandit problems. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 1:1–6.

[Stout et al., 2004] Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., and Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review*, 11(4):742–747.

[Stout et al., 2001] Stout, J. C., Rodawalt, W. C., and Siemers, E. R. (2001). Risky decision making in Huntington ' s disease. pages 92–101.

[Sutton and Barto, 2012] Sutton, R. and Barto, A. (2012). Reinforcement Learning : An Introduction.

[Vinh et al., 2010] Vinh, N. X., Epps, J., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research*, 11:2837–2854.

[Wang et al., 2013] Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., and Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309.

[Wang and Ruhe, 2007] Wang, Y. and Ruhe, G. (2007). The cognitive process of decision making. *International Journal of Cognitive Informatics and Natural Intelligence*, 1(2):73–85.

[Yechiam et al., 2008] Yechiam, E., Kanz, J. E., Bechara, A., Stout, J. C., Busemeyer,

J. R., Altmeier, E. M., and Paulsen, J. S. (2008). Neurocognitive deficits related to poor decision making in people behind bars. *Psychonomic Bulletin & Review*, 15(1):44–51.

Appendix

Metrics and Similarities

This part of the appendix formally defines metrics and similarities and dissimilarities (proximity) used in this paper. We first define some basic general concepts followed by a description of the applied distance and similarity concepts.

Distances vs. Similarities

Let $\mathcal{X}$ be a dataset and let $\boldsymbol{x_i}, \boldsymbol{x_j}$ be two datapoints, such that $\boldsymbol{x_i}, \boldsymbol{x_j} \in \mathcal{X}$.

A distance function assign for pairs a points a non negative real number as distance. $d : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}_0^+$. Formally if the following properties are additionally staisfied the distance is also metric [Shirali and Vasudeva, 2005, page 28].

1. $d(\boldsymbol{x_i}, \boldsymbol{x_j}) \geq 0$
2. $d(\boldsymbol{x_i}, \boldsymbol{x_i}) \geq 0$
3. $d(\boldsymbol{x_i}, \boldsymbol{x_j}) = d(\boldsymbol{x_j}, \boldsymbol{x_i})$
4. $d(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq d(\boldsymbol{x_i}, \boldsymbol{x_j}) + d(\boldsymbol{x_j}, \boldsymbol{x_i})$

A distance can be seen as a measure for dissimilarity of two points [Everitt et al., 2009, page 35]. Besides distance some algorithms operate on a *similarity* matrix. Formally a similarity is a function $S : \mathcal{X} \times \mathcal{X} \mapsto [0, 1]$. Also for similarity we can define the following properties [Fratev et al., 1979, page 3]:

1. $0 \leq S(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq 1, \text{for } i \neq j$
2. $S(\boldsymbol{x_i}, \boldsymbol{x_i}) = 1$
3. $S(\boldsymbol{x_i}, \boldsymbol{x_j}) = S(\boldsymbol{x_j}, \boldsymbol{x_i})$

Once we have computed distance or a similarity we can compute for two data points we can use this information to transform it to a similarity or the distance vice versa [Boriah et al., 2008, page 4]:

$$S(\boldsymbol{x_i}, \boldsymbol{x_j}) = \frac{1}{1 + d(\boldsymbol{x_i}, \boldsymbol{x_j})} \quad \Leftrightarrow \quad d(\boldsymbol{x_i}, \boldsymbol{x_j}) = \frac{1}{S(\boldsymbol{x_i}, \boldsymbol{x_j})} - 1 \tag{3}$$

.0.1 Similarity measures for time series data

We use three different similarity measures for time series. E.g. [Wang et al., 2013] or [Serra and Arcos, 2014] provide an overview and empirical evaluation on common similarity measures for time series. The following definitions are taken from the latter. Empirical research suggest that simple euclidean distance for time series performs quite well and is hard to beat. Hence, the first distance measure for time series is simply euclidean distance between two time series. They might be converted to similarity based on equation 3. Let $\boldsymbol{x}, \boldsymbol{y}$ be two time series over N-periods. Then the L_2 distance between two time series is define as:

$$d_E(\boldsymbol{x}, \boldsymbol{y}) := \sqrt{\left(\sum_{i=1}^{N} (x_i - y_i)^2 \right)} \tag{4}$$

.0.2 Similarity measures for categorical data

Reffering to equation 4 we define a simple measure for the categorical aspect of the data. Note that [SOURCE] defines the simplest measure for categorical data. However, since the probability that two people behave exactly the same in our context is arguably zero we modify this concept slightly. So we relax that and define the overlap similarity just as the count of overlapping instances. This serves as a benchmark similarity for categorical data.

$$d_O(\boldsymbol{x}, \boldsymbol{y}) := \sum_{i=1}^{N} \mathbb{1}_{(x_i=y_i)} \tag{5}$$

In [Boriah 2008] we find a rich class of further categorical measures. To keep it concise we considered two of them. The first one is *Eskin* similarity measures.

$$d_O(\boldsymbol{x}, \boldsymbol{y}) := \begin{cases} 1 & \text{,if } \boldsymbol{x} = \boldsymbol{y} \\ \frac{n_k^2}{n_k^2+2} & \text{otherwise} \end{cases} \tag{6}$$

Furthermore, we consider lins similarity

$$d_{lin}(\boldsymbol{x}, \boldsymbol{y}) := \begin{cases} 2\log \hat{p}_k(X_k) & \text{,if } \boldsymbol{x} = \boldsymbol{y} \\ 2\log\left(\hat{p}_k(x_k) + \hat{p}_k(y_k)\right) & \text{otherwise} \end{cases} \tag{7}$$

Clustering Evaluation

In this section we formally derive and explain the applied clustering metrics. Evaluating clustering performance has some issues. The algorithm is producing labels. Nevertheless we generated the "true" labels the might not be comparable. A simple example we might consider the following situation. Let y denote the labels of data the data and y' the corresponding prediction such that $y, y' \in \{0, 1\}$. In a small example let our data points be like $y = (1, 1, 0, 0)$ and the corresponding prediction $y' = (0, 0, 1, 1)$. Obviously the clustering worked perfectly, however comparing "labels" would produce an accuracy of zero.

There a several clustering metrics, which respect such a situation. We consider a bunch of information based metrics. Most of the measures use some sort of entropy. The following concepts can be found in [Rosenberg and Hirschberg, 2007] and [Vinh et al., 2010] and as additional reading [Hubert and Arabie, 1985].

First we might introduce the contingency table.

	V_1	V_2	$\dots$	V_c	Σ
U_1	$n_{1,1}$	$n_{1,2}$	$\dots$		a_1
U_2	$n_{2,1}$	$\ddots$			a_1
$\vdots$	$n_{1,1}$	$\dots$			a_1
U_R	$n_{1,1}$	$\dots$			a_1
	b_1	b_2	$\dots$	b_c	N

Table 5: Contigency Table

N_{11}**:** Number of pairs in the same cluster

N_{00}**:** Number of pairs that are in different clusters in both v and u

N_{01}**:** Number of pairs that are in the same cluster in both u but different in v

N_{10}**:** Number of pairs that are in the same cluster in both v but different in u

$$RI(u,v) = \frac{N_{00} + N_{11}}{\binom{N}{2}} \tag{8}$$

$$ARI(u,v) = \frac{2\left(N_{00}N_{11} - N_{01}N_{10}\right)}{\left(N_{00} + N_{01}\right)\left(N_{01} + N_{11}\right) + \left(N_{00} + N_{10}\right)\left(N_{10} + N_{11}\right)} \tag{9}$$

$$H(u) = -\sum_{i=1}^{R} \frac{a_i}{N} \log \frac{a_i}{N} \tag{10}$$

$$H(v) = -\sum_{i=1}^{C} \frac{b_i}{N} \log \frac{a_i}{N} \tag{11}$$

$$H(u,v) = -\sum_{i=1}^{R}\sum_{j=1}^{C} \frac{n_{i,j}}{N} \log \frac{n_{i,j}}{N} \tag{12}$$

$$H(u|v) = -\sum_{i=1}^{R}\sum_{j=1}^{C} \frac{n_{i,j}}{N} \log \frac{n_{i,j}/N}{b_j/N} \tag{13}$$

$$H(v|u) = -\sum_{i=1}^{C}\sum_{j=1}^{R} \frac{n_{i,j}}{N} \log \frac{n_{i,j}/N}{b_j/N} \tag{14}$$

$$I(u,v) = \sum_{i=1}^{R}\sum_{j=1}^{C} \frac{n_{i,j}}{N} \log \frac{n_{i,j}/N}{a_i b_j/N} \tag{15}$$

$$\tag{16}$$

Normalized Info Score:

This is one example of a normalized version

$$NMI_{max}(u,v) = \frac{I(u,v)}{\max\left(H(u), H(v)\right)} \tag{17}$$

$$\begin{aligned} AMI_{max}(u,v) &= \frac{NMI_{max}(u,v) - \mathbb{E}\left[NMI_{max}(u,v)\right]}{1 - \mathbb{E}\left[NMI_{max}(u,v)\right]} \\ &= \frac{I(u,v) - \mathbb{E}\left[I(u,v)\right]}{\max\left(H(u), H(v)\right) - \mathbb{E}\left[I(u,v)\right]} \end{aligned} \tag{18}$$

$$\mathbb{E}\left[I(u,v)\right] = \sum_{i=1}^{R}\sum_{j=1}^{C} \sum_{n_{i,j}=\max(a_i+b_j-N,0)}^{\min(a,b)} \frac{n_{ij}}{N} \log\left(\frac{N n_{ij}}{a_i b_j}\right) \frac{a_i! b_j! (N-a_i)! (N-b_j)!}{N! n_{ij}! (a_i - n_{ij})! (b_j - n_{ij})! (N - a_i - b_j + n_{ij})!} \tag{19}$$

Homogeneity:

$$h = \begin{cases} 1 & \text{,if } H(u,v) = 0 \\ 1 - \frac{H(u|v)}{H(u)} & \text{otherwise} \end{cases} \tag{20}$$

Completeness:

$$c = \begin{cases} 1 & \text{,if } H(v,u) = 0 \\ 1 - \frac{H(v|u)}{H(v)} & \text{otherwise} \end{cases} \tag{21}$$

V Measure Score:

$$V_\beta = \frac{(1+\beta)hc}{\beta h + c} \tag{22}$$

.1 Algorithms

Principal Components and Multidimensional Scaling

Coming to principal component analysis. [Shlens, 2014] provides an excellent contribution considering both practical applications and theoretical background. The following derivations are from this source.

Spectral Clustering

For the spectral clustering algorithm we formally introduce some graph notation. If not stated otherwise the following derivation follows [Luxburg, 2007]. In the following we consider a weighted and simple undirected graph.

$$G = \{V, E\} \tag{23}$$

$$V = \{v_1, \ldots, v_n\} \tag{24}$$

$$E = \{e_1, \ldots, e_n\} \tag{25}$$

Furthermore let the graph has a weighted and symetric ($|V| \times |V|$) adjacency matrix, such that:

$$\boldsymbol{W} = \begin{cases} w_{i,j} & \text{,if } v_i v_j \in E \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

The *degree* of a node is defined as the sum of edge weights of connected nodes. Formally we denote the degree of node i as:

$$d_i := \sum_{j=1}^{n} w_{ij} = \sum_{i=1}^{n} w_{ij} \tag{27}$$

Using the last expression we define matrix $\boldsymbol{D}$ as the diagonal matrix of the degress

$$\boldsymbol{D} := diag(\boldsymbol{d}) \tag{28}$$

The algorithm works on the *Laplacian* matrix defined by:

$$\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{W} \tag{29}$$

Former versions of the algorithm are applied on the graph laplcian. However, there were proposed newer versions using the so called *normalized laplacian*. Since also the python version is using this package we will focus on this version of the algorithm. Following that the normalized graph laplacian is defined as:

$$\boldsymbol{L}_{norm} := \boldsymbol{D}^{1/2}\boldsymbol{L}\boldsymbol{D}^{1/2} = \boldsymbol{I} - \boldsymbol{D}^{1/2}\boldsymbol{W}\boldsymbol{D}^{1/2} \tag{30}$$

A Affinity Propagation

[Brusco and Köhn, 2008]

B Simulation Data

Tau	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1, 0.3}	Spectral	DTW				x		0.136 (0.134)	0.094 (0.146)	0.135 (0.134)
	Spectral	EDR			x			0.134 (0.135)	0.097 (0.130)	0.134 (0.135)
	Spectral	Levensthein	x					0.093 (0.134)	0.050 (0.128)	0.092 (0.134)
	Spectral	Cosine		x				0.113 (0.134)	0.082 (0.150)	0.113 (0.134)
	K-Means	Euclidian					x	0.093 (0.135)	0.050 (0.130)	0.092 (0.135)
{0.1, 0.5}	Spectral	RBF				x		0.331 (0.185)	0.328 (0.208)	0.331 (0.186)
	Spectral	RBF			x			0.322 (0.197)	0.301 (0.221)	0.321 (0.197)
	Spectral	EDR			x			0.315 (0.189)	0.328 (0.206)	0.315 (0.189)
	K-Means	Euclidian			x			0.313 (0.187)	0.261 (0.213)	0.312 (0.187)
	Spectral	EDR		x				0.226 (0.127)	0.131 (0.113)	0.223 (0.127)
	Spectral	Levensthein	x					0.219 (0.144)	0.126 (0.121)	0.217 (0.144)
	K-Means	Euclidian					x	0.208 (0.143)	0.124 (0.122)	0.207 (0.143)
{0.1, 0.7}	Spectral	RBF				x		0.642 (0.200)	0.653 (0.206)	0.641 (0.200)
	Spectral	DTW			x			0.580 (0.209)	0.611 (0.205)	0.580 (0.201)
	K-Means	Euclidian			x			0.533 (0.248)	0.539 (0.275)	0.533 (0.249)
	K-Means	Euclidian					x	0.382 (0.181)	0.309 (0.237)	0.381 (0.180)
	Spectral	Levensthein	x					0.350 (0.114)	0.241 (0.163)	0.347 (0.116)
	Spectral	Euclidian		x				0.306 (0.155)	0.198 (0.179)	0.303 (0.156)
{0.1, 1}	Spectral	DTW				x		0.938 (0.131)	0.942 (0.126)	0.938 (0.131)
	Spectral	Euclidian				x		0.811 (0.136)	0.833 (0.129)	0.811 (0.136)
	Ward	Euclidian				x		0.797 (0.205)	0.796 (0.218)	0.797 (0.205)
	Spectral	DTW			x			0.765 (0.197)	0.791 (0.184)	0.765 (0.197)
	K-Means	Euclidian			x			0.757 (0.180)	0.774 (0.185)	0.757 (0.180)
	K-Means	Euclidian					x	0.796 (0.218)	0.687 (0.265)	0.696 (0.219)
	Spectral	Overlap	x					0.600 (0.299)	0.571 (0.353)	0.598 (0.302)
	Spectral	Euclidian		x				0.335 (0.091)	0.220 (0.138)	0.332 (0.093)

Table 6: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 2, 4\}$, $\sigma = \{1, 1, 1\}$, $\alpha = 0.5$)

Tau	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1, 0.3}	Spectral	Overlap	x					0.103 (0.019)	0.002 (0.007)	0.087 (0.020)
	Spectral	EDR		x				0.096 (0.061)	0.016 (0.049)	0.088 (0.059)
	Spectral	EDR				x		0.068 (0.080)	0.008 (0.085)	0.063 (0.078)
	Spectral	Cosine			x			0.043 (0.067)	-	0.043 (0.067)
	K-Means	Euclidian					x	0.036 (0.041)	-	0.036 (0.041)
{0.1, 0.5}	Average	Euclidian		x				0.133 (0.115)	0.040 (0.077)	0.127 (0.115)
	K-Means	Euclidian	x					0.112 (0.124)	0.035 (0.085)	0.110 (0.122)
	Spectral	Euclidian				x		0.103 (0.096)	0.060 (0.092)	0.103 (0.095)
	Spectral	Cosine			x			0.080 (0.088)	0.038 (0.092)	0.080 (0.088)
	K-Means	Euclidian					x	0.079 (0.097)	0.034 (0.088)	0.079 (0.096)
{0.1, 0.7}	Spectral	DTW				x		0.361 (0.200)	0.352 (0.211)	0.361 (0.200)
	Spectral	EDR				x		0.223 (0.083)	0.087 (0.080)	0.217 (0.086)
	Ward	Euclidian				x		0.195 (0.143)	0.103 (0.135)	0.192 (0.143)
	K-Means	Euclidian		x				0.167 (0.123)	0.064 (0.114)	0.161 (0.123)
	K-Means	Euclidian	x					0.157 (0.127)	0.060 (0.116)	0.151 (0.127)
	Spectral	DTW			x			0.105 (0.140)	0.087 (0.179)	0.105 (0.140)
	K-Means	Euclidian					x	0.102 (0.183)	0.063 (0.201)	0.101 (0.183)
{0.1, 1.0}	Spectral	DTW				x		0.792 (0.174)	0.808 (0.170)	0.792 (0.174)
	Spectral	RBF				x		0.692 (0.164)	0.692 (0.184)	0.692 (0.164)
	K-Means	Euclidian				x		0.606 (0.188)	0.578 (0.233)	0.606 (0.189)
	Spectral	EDR		x				0.280 (0.198)	0.201 (0.215)	0.278 (0.198)
	Spectral	Levensthein	x					0.276 (0.101)	0.154 (0.113)	0.272 (0.111)
	Spectral	EDR			x			0.236 (0.196)	0.171 (0.184)	0.235 (0.195)

Table 7: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 2, 4\}$, $\sigma = \{1, 1, 1\}$, $\alpha = 1$)

Tau	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1, 0.3}	Spectral	RBF				x		0.290 (0.173)	0.279 (0.205)	0.279 (0.173)
	Spectral	EDR			x			0.284 (0.253)	0.279 (0.274)	0.284 (0.253)
	Spectral	Levensthein	x					0.275 (0.147)	0.195 (0.156)	0.272 (0.148)
	K-Means	Euclidean					x	0.260 (0.126)	0.200 (0.150)	0.258 (0.127)
	Spectral	Euclidean		x				0.215 (0.144)	0.126 (0.159)	0.211 (0.144)
{0.1, 0.5}	Spectral	DTW				x		0.824 (0.166)	0.837 (0.170)	0.824 (0.166)
	Spectral	DTW			x			0.636 (0.213)	0.661 (0.205)	0.636 (0.213)
	K-Means	Euclidean					x	0.532 (0.182)	0.543 (0.206)	0.531 (0.182)
	Spectral	Levensthein	x					0.411 (0.125)	0.333 (0.160)	0.410 (0.127)
	Spectral	Euclidean		x				0.242 (0.163)	0.132 (0.187)	0.238 (0.165)
{0.1, 0.7}	Spectral	DTW				x		1.00 (0.000)	1.00 (0.000)	1.00 (0.000)
	Spectral	DTW			x			0.726 (0.228)	0.750 (0.214)	0.726 (0.228)
	K-Means	Euclidean					x	0.721 (0.211)	0.741 (0.207)	0.721 (0.211)
	Spectral	Levensthein	x					0.333 (0.147)	0.222 (0.195)	0.329 (0.150)

Table 8: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 1, 2\}$, $\sigma = \{1, 1, 1\}$, $\alpha = 0.1$)

Tau	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1, 0.3}	Average	Euclidian		x				0.101 (0.051)	0.005 (0.023)	0.088 (0.050)
	Average	Euclidian				x		0.093 (0.060)	0.006 (0.024)	0.084 (0.057)
	Spectral	Overlap	x					0.091 (0.073)	0.007 (0.029)	0.084 (0.070)
	Spectral	EDR			x			0.071 (0.077)	0.025 (0.072)	0.070 (0.076)
	K-Means	Euclidian					x	0.057 (0.064)	0.014 (0.069)	0.057 (0.064)
{0.1, 0.5}	Spectral	RBF				x		0.256 (0.190)	0.221 (0.195)	0.255 (0.190)
	Spectral	Overlap	x					0.092 (0.079)	0.008 (0.041)	0.087 (0.075)
	Ward	Euclidian		x				0.092 (0.085)	0.007 (0.040)	0.088 (0.082)
	K-Means	Euclidian			x			0.080 (0.100)	0.043 (0.122)	0.079 (0.100)
	K-Means	Euclidian					x	0.062 (0.093)	0.026 (0.116)	0.062 (0.093)
{0.1, 0.7}	Spectral	DTW				x		0.618 (0.192)	0.627 (0.201)	0.617 (0.192)
	Spectral	Levensthein	x					0.252 (0.137)	0.163 (0.147)	0.249 (0.137)
	Spectral	RBF			x			0.215 (0.158)	0.207 (0.188)	0.215 (0.157)
	Spectral	DTW		x				0.205 (0.218)	0.192 (0.247)	0.205 (0.218)
	K-Means	Euclidian					x	0.185 (0.205)	0.165 (0.225)	0.185 (0.205)
{0.1, 1.0}	Spectral	RBF				x		0.952 (0.098)	0.960 (0.082)	0.952 (0.098)
	Spectral	DTW		x				0.489 (0.269)	0.491 (0.290)	0.488 (0.269)
	K-Means	Euclidian					x	0.453 (0.214)	0.427 (0.252)	0.452 (0.215)
	Spectral	DTW			x			0.406 (0.203)	0.403 (0.226)	0.406 (0.203)

Table 9: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 1, 2\}$, $\sigma = \{1, 1, 1\}$, $\alpha = 1$)

Tau	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1,0.3}	Spectral	Cosine			x			0.088 (0.116)	0.038 (0.105)	0.088 (0.115)
	Spectral	Cosine	x					0.087 (0.099)	0.019 (0.075)	0.084 (0.097)
	Average	Euclidian				x		0.083 (0.107)	0.026 (0.095)	0.080 (0.106)
	Spectral	Cosine		x				0.039 (0.055)	0.000 (0.068)	0.039 (0.055)
	K-Means	Euclidian					x	0.037 (0.057)	-0	0.037 (0.056)
{0.1,0.5}	Spectral	EDR				x		0.088 (0.087)	0.027 (0.078)	0.083 (0.086)
	Spectral	Overlap	x					0.066 (0.051)	0.026 (0.062)	0.065 (0.050)
	Spectral	Cosine			x			0.047 (0.064)	0.002 (0.067)	0.047 (0.063)
	Spectral	DTW		x				0.030 (0.045)	-0	0.030 (0.045)
	K-Means	Euclidian					x	0.029 (0.045)	-0	0.029 (0.045)
{0.1,0.7}	Average	Euclidian				x		0.087 (0.038)	0.007 (0.039)	0.076 (0.034)
	Spectral	Cosine	x					0.081 (0.076)	0.013 (0.044)	0.078 (0.072)
	K-Means	Euclidian			x			0.070 (0.074)	0.014 (0.061)	0.068 (0.072)
	Spectral	EDR		x				0.048 (0.054)	0.003 (0.063)	0.047(0.053)
	K-Means	Euclidian					x	0.041 (0.071)	-0	0.041 (0.071)
{0.1,0.9}	Spectral	Cosine	x					0.180 (0.110)	0.069 (0.081)	0.175 (0.109)
	Spectral	EDR		x				0.161 (0.125)	0.097 (0.129)	0.159 (0.124)
	K-Means	Euclidian			x			0.142 (0.127)	0.049 (0.097)	0.139 (0.126)
	Average	Euclidian				x		0.115 (0.034)	0.008 (0.016)	0.100 (0.037)
	K-Means	Euclidian					x	0.058 (0.108)	0.016 (0.115)	0.058 (0.108)
{0.5, 0.9}	Spectral	Cosine	x					0.169 (0.125)	0.075 (0.097)	0.164 (0.126)
	Spectral	EDR			x			0.158 (0.184)	0.111 (0.188)	0.157 (0.184)
	Spectral	EDR		x				0.150 (0.138)	0.103 (0.149)	0.148 (0.137)
	Average	Euclidian				x		0.104 (0.040)	0.003 (0.014)	0.091 (0.038)

Table 10: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 2, 4\}$, $\sigma = \{1, 1, 1\}$, $\tau = 0.1$)

Alpha	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1,0.3}	Spectral	Cosine				x		0.131 (0.114)	0.100 (0.133)	0.131 (0.113)
	K-Means	Euclidian	x					0.130 (0.199)	0.075 (0.212)	0.127 (0.199)
	Spectral	Cosine		x				0.119 (0.104)	0.068 (0.094)	0.118 (0.103)
	K-Means	Euclidian					x	0.113 (0.156)	0.089 (0.181)	0.113 (0.156)
	Spectral	RBF			x			0.101 (0.140)	0.073 (0.162)	0.101 (0.140)
{0.1,0.5}	Spectral	Cosine			x			0.199 (0.131)	0.141 (0.129)	0.198 (0.131)
	Spectral	Cosine		x				0.191 (0.157)	0.128 (0.148)	0.190 (0.157)
	Spectral	Euclidian				x		0.168 (0.183)	0.134 (0.186)	0.168 (0.182)
	K-Means	Euclidian					x	0.122 (0.121)	0.085 (0.122)	0.121 (0.121)
	Spectral	Cosine	x					0.063 (0.081)	0.033 (0.109)	0.063 (0.081)
{0.1,0.7}	Spectral	Cosine		x				0.277 (0.229)	0.222 (0.240)	0.276 (0.229)
	Spectral	Cosine				x		0.278 (0.272)	0.183 (0.254)	0.271 (0.183)
	K-Means	Euclidian			x			0.227 (0.178)	0.186 (0.176)	0.227 (0.177)
	Spectral	Cosine	x					0.190 (0.114)	0.152 (0.116)	0.189 (0.114)
	K-Means	Euclidian					x	0.178 (0.150)	0.147 (0.147)	0.177 (0.149)
{0.1,0.9}	Spectral	DTW				x		0.300 (0.232)	0.273 (0.254)	0.299 (0.232)
	Spectral	Cosine		x				0.211 (0.158)	0.159 (0.160)	0.210 (0.158)
	Spectral	Cosine			x			0.191 (0.117)	0.137 (0.127)	0.189 (0.118)
	Spectral	Cosine	x					0.164 (0.110)	0.115 (0.109)	0.162 (0.109)
	K-Means	Euclidian					x	0.138 (0.153	0.091 (0.176)	0.137 (0.153)

Table 11: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 2, 4\}$, $\sigma = \{1, 1, 1\}$, $\tau = 0.5$)

Alpha	Method	Similarity	C	BBC	E	EB	CC	NMI	ARI	VM
{0.1, 0.3}	Spectral	Overlap	x					0.103 (0.019)	0.002 (0.007)	0.087 (0.020)
	Spectral	EDR		x				0.096 (0.061)	0.016 (0.049)	0.088 (0.059)
	Spectral	EDR				x		0.068 (0.080)	0.008 (0.085)	0.063 (0.078)
	Spectral	Cosine			x			0.043 (0.067)	-0	0.043 (0.067)
	K-Means	Euclidian					x	0.036 (0.041)	-	0.036 (0.041)
{0.1, 0.5}	Average	Euclidian		x				0.133 (0.115)	0.040 (0.077)	0.127 (0.115)
	K-Means	Euclidian	x					0.112 (0.124)	0.035 (0.085)	0.110 (0.122)
	Spectral	Euclidian				x		0.103 (0.096)	0.060 (0.092)	0.103 (0.095)
	Spectral	Cosine			x			0.080 (0.088)	0.038 (0.092)	0.080 (0.088)
	K-Means	Euclidian					x	0.079 (0.097)	0.034 (0.088)	0.079 (0.096)
{0.1, 0.7}	Spectral	DTW				x		0.361 (0.200)	0.352 (0.211)	0.361 (0.200)
	Spectral	EDR				x		0.223 (0.083)	0.087 (0.080)	0.217 (0.086)
	Ward	Euclidian				x		0.195 (0.143)	0.103 (0.135)	0.192 (0.143)
	K-Means	Euclidian		x				0.167 (0.123)	0.064 (0.114)	0.161 (0.123)
	K-Means	Euclidian	x					0.157 (0.127)	0.060 (0.116)	0.151 (0.127)
	Spectral	DTW			x			0.105 (0.140)	0.087 (0.179)	0.105 (0.140)
	K-Means	Euclidian					x	0.102 (0.183)	0.063 (0.201)	0.101 (0.183)
{0.1, 1.0}	Spectral	DTW				x		0.792 (0.174)	0.808 (0.170)	0.792 (0.174)
	Spectral	RBF				x		0.692 (0.164)	0.692 (0.184)	0.692 (0.164)
	K-Means	Euclidian				x		0.606 (0.188)	0.578 (0.233)	0.606 (0.189)
	Spectral	EDR		x				0.280 (0.198)	0.201 (0.215)	0.278 (0.198)
	Spectral	Levensthein	x					0.276 (0.101)	0.154 (0.113)	0.272 (0.111)
	Spectral	EDR			x			0.236 (0.196)	0.171 (0.184)	0.235 (0.195)
	K-Means	Euclidian					x	0.099 (0.105)	0.052 (0.097)	0.098 (0.104)

Table 12: Simulation results (Setting: Rounds = 20, Size = 20, $\mu = \{0, 2, 4\}$, $\sigma = \{1, 1, 1\}$, $\tau = 1.0$)