BACHELOR'S PROJECT

Comparing subjective psychedelic experiences across substances using the BERTopic model

Aleksander Moeslund Wael[a]

Rebekah Brita Baglini[b]

[a] Corresponding author.
Student no. 202005192
Email: 202005192@post.au.dk
Cognitive Science BA
School of Communication and Culture
Faculty of Arts, Aarhus University

[b] Project supervisor.
Email: rbkh@cc.au.dk
School of Communication and Culture
Faculty of Arts, Aarhus University

Summary (abstract): The recent surge in research on the effects and treatment possibilities of psychedelic substances has helped to establish a foundation of psychedelic neuroscience, psychology and psycho-pharmacology. With LSD and psilocybin being the main substances used in psychedelic therapy research, the field still lacks a clear understanding of how psychedelic substances differ in terms of acute and subjective effects. A way of understanding the effects of psychedelic substances is to analyze unstructured experience reports, which are available at online forums such as Erowid. The current study follows the preliminary success of others in applying NLP methods to unstructured experience reports, with the goal of gauging the differences in substance effects. Topic modelling is performed on a corpus of 4219 experience reports spanning six psychedelic substances (LSD, psilocybin, mescaline, Ayahuasca, DMT and 2C-B) using the BERTopic algorithm. The topics were interpretable and revealed differences between substances in regards to perceptual effects and cognitive effects. Lack of controlled variables, importance of context (*set* and *setting)* and limitations of methods do not allow for strong conclusions of causality, but call for more controlled studies to be done on the acute effects of psychedelic substances.

Keywords: psychedelics, NLP, topic modelling, BERTopic

AARHUS UNIVERSITY

Table of Contents

1 Introduction

1.1 Psychedelic substances

The substances studied in this paper have seen various classifications throughout the history of research and recreational use. Today, the substances are most commonly referred to as “psychedelics”, meaning "mind manifesting" (Nichols, 2004) or "mind revealing" (Carhart-Harris & Goodwin, 2017). This term is often used interchangeably with hallucinogenics (referring to the albeit rare occurence of hallucinations), psychotomimetics ("psychosis mimicking") and entheogenics ("god within") drugs (Nichols, 2004). For simplicity, the term "psychedelic" is used in this paper, as seems to be the standard for current scientific research.

1.2 History of psychedelics

The natural occurrence and effects of psychedelic substances have likely played a key role in development of various religions and philosophic thought throughout history (Nichols, 2004). There is anthropological and ethnopharmacological evidence of psychedelic substance use in ancient India (a substance called *soma*), ancient Greece, Aztec and Maya Cultures (e.g. *teonanacatl* and *ayahuasca*), Australia, Tanzania, and by Native North Americans (*Peyote* cactus) to name a few (Nichols, 2016). Considering the described effects of psychedelic substances, it is not hard to imagine ancient peoples gazing at a bonfire, sunset or starry night sky after ingestion, discussing with awe and wonder (and, perhaps, existential dread) what divine powers lie behind the observable world.

More recently in the West, lysergic acid diethylamide (LSD) became almost synonymous with the '60s and '70s hippie culture, with musicians and the festival scene embracing the substance due to it heightening feelings of connectedness, love, freedom and creativity (Nichols, 2016). Psychedelic culture has inspired new genres of psychedelic music (DeRogatis, 2003), art (Grunenberg & Harris, 2005), literature (Dickins, 2012) and film (Gallagher, 2004). It was also the anti-war, anti-government, anti-convention aspects of this movement that partly led to the passage of the U.S Controlled Substances Act of 1970, under which several known psychedelics (e.g. LSD, psilocybin and mescaline) became Schedule 1 controlled substances, immediately halting all research on these substances (Nichols, 2016). Following a two-decade hiatus, psychedelic research was once again active during the '90s in Germany, Switzerland and U.S.A.

Combining the period before 1970 and the period from the '90s until today, a foundational body of research on psychedelics now exists in the fields of neuroscience, psychology and psycho-pharmacology (Carhart-Harris & Goodwin, 2017).

1.3 The effects of psychedelics

What characteristics does a psychedelic experience contain to be deemed mind manifesting or mind revealing? A psychedelic experience can well be categorized as a pharmacologically induced altered state of consciousness (ASC). Charles Tart defines these states as "alternate patterns or configurations of experience, which differ qualitatively from a typical baseline state." (Garcia-Romeu & Tart, 2013). According to Jaffe (1990), what distinguishes psychedelics from other drugs is *"their capacity reliably to induce states of altered perception, thought, and feeling that are not experienced otherwise except in dreams or at times of religious exaltation."* (Nichols, 2016).

Generally, the effects of psychedelics can be categorised accordingly: somatic, perceptual, psychic and cognitive (Nichols, 2004; Preller & Vollenweider, 2016). Although effects vary from user to user, common reported somatic effects include dizziness, tremors, increased heartrate, nausea (sometimes resulting in vomiting) and blurred vision, particularly when the first effects of the substance set in, known as the "come-up" phase.

The perceptual, psychic and cognitive effects are really at the core of what makes the experience "psychedelic". Simply put, these effects are alterations in visual, auditory and haptic perception, alterations of cognitive faculties such as memory, attention, learning, creativity and language production, all leading to potential experiences of depersonalization, altered thought processes and dreamlike feelings (Nichols, 2004; Preller & Vollenweider, 2016).

Effects of visual perception include perceiving different or more intense colors, alterations in shapes, perceived movement and/or animacy of objects and difficulty in gauging distances. Auditory perception is often altered in intensity; some sounds are perceived louder or sharper, and misperceptions of sound can also occur. Haptic perception is sometimes altered in the somatosensory sense; the texture, shape and weight of objects are altered, and sensations of warmth and cold appear at hands and feet; and sometimes in the sense of proprioception, where balance and sense of body size (such as arm length) is altered (Preller & Vollenweider, 2016).

The cognitive effects of psychedelics are mainly observed in working memory, attention, learning, creativity and language production. Users find it harder to connect thoughts after ingesting LSD (Preller & Vollenweider, 2016), perhaps a result of impairment in working memory and attention, which is also found in studies with psilocybin (Carter et al., 2005). Furthermore, performance in tests such as Raven's Progressive Matrices and the Stroop test is impaired, reaction time is consistently slowed, and attentional tracking is impaired as well. It is found that response inhibition is decreased whilst under the effects of psilocybin (Gouzoulis-Mayfrank et al., 2002). The perceptual and cognitive changes are, perhaps, what leads to a reported increase in creativity, although there are methodological limitations in studying this claim (Preller & Vollenweider, 2016).

Assessing the subjective effects of psychedelics has mostly been done through introspective questionnaires, starting with the Hallucinogen Rating Scale (HRS) (Strassman et al., 1994), followed by Dittrich's Abnormal Mental States (APZ) questionnaire (Dittrich, 1994), which was further edited to the OAV scales (Bodmer et al., 1994) and the five-dimensional altered states ofconsciousness (5D-ASC) questionnaire (Dittrich et al., 2006). The OAV scales were revised by Studerus et al. (2010), increasing dimensionality to 11 lower-order scales: experience of unity, spiritual experience, blissful state, insightfulness, disembodiment, impaired control and cognition, anxiety, complex imagery, elementary imagery, audio-visual synesthesia, and changed meaning of percepts (Nichols, 2016).

1.4 Treatment with psychedelics

Both before the Controlled Substances Act of 1970 and since the revival of psychedelic research, a primary motivation of the field has been the potential therapeutic and medicinal benefits of treatment with psychedelic substances.

Some success has been achieved with psychedelic-assisted treatment for anxiety and depression. Grob et al. (2011) found a reduction in anxiety (STAI trait anxiety subscale scores) at 1 and 3 months post treatment with psilocybin, and an improvement of mood at 6 months. Gasser et al. (2014) found positive outcomes in the long-term when applying LSD-assisted psychotherapy in patients with anxiety related to a life-threatening disease or diagnosis. Sanches et al. (2016) found an improvement in depressive symptoms lasting up to 21 days following a single dose of DMT (in the form of Ayahuasca) (Nichols et al., 2017).

Alongside a plethora of anecdotal reports and single-subject studies (Nichols, 2016), psilocybin has been administered in patients with obsessive-compulsive disorder (OCD), showing improvements within 24 hours post treatment (Moreno et al., 2006), though there are methodological concerns with these findings (Nichols et al., 2017).

Perhaps the most succesful application of psychedelics is in treating various forms of addiction. A sum of these have focused on alcoholism, e.g. treating with LSD (Chwelos et al., 1959) and psilocybin (Bogenschutz et al., 2015). In a meta-analysis of six studies using LSD to treat alcoholism (Krebs & Johansen, 2012), a significant decrease in alcohol misuse was found in patients who were administered varying doses of LSD compared to control groups. Another study focused on opioid addiction and found less usage of heroin in the LSD-receiving group compared to the control group post treatment, significant at all examined time windows up to a year post treatment (Savage & McCabe, 1973). Johnson et al. (2014) conducted a pilot study on treating tobacco addiction with cognitive behavioural therapy (CBT) alongside administering psilocybin. They confirmed abstinence in 80% of participants (n = 15) at 6 months post quitting smoking, and a follow-up study (Johnson et al., 2017) confirmed abstinence in 67% of participants at 12 months post quitting smoking. This is to be compared with the current most effective medication for smoking cessation Veranicline, which has an approx. 35% succes rate at 6 months post quit-date (Nichols et al., 2017).

Importantly, a note should be made on the importance of context when administering psychedelic substances. As such, many of the promising results come from studies that provide extensive psychological support for the participants before, during and after partaking in the experiment. This includes several preparation meetings with professionals, creating an optimal therapeutic environment with lighting, music and aesthetics, and always having someone to compassionately support the participant before, during and after the study. This is important to ensure a correct *"set"* and *"setting"*; set is the mindset (e.g. expectations and assumptions) of the user, and setting is the physical environment surrounding the user (Carhart-Harris et al., 2018). Administering psychedelics without this psychological support may limit the effects of the treatment or potentially worsen the condition of a patient. The potential of psychedelic treatment lies in leveraging the psychedelic state by combining it with methods such as cognitive therapy and attentional-bias training (Carhart-Harris & Goodwin, 2017).

1.5 Differences between substances

Psychedelics are commonly associated with three types of chemical structures; tryptamines (e.g. psilocybin, N,N-dimethyltryptamine (DMT) and 5-MeO-DMT), phenethylamines (e.g. mescaline and 2C-B) and lysergamides (e.g. LSD) (Nichols, 2016).

It is widely accepted that psychedelic effects are mainly caused by agonist or partial agonist activity at the serotonin 5-HT2A receptor, which has given rise to the name "serotonergic psychedelics" (Nichols, 2016). Some define psychedelic substances as "compounds with appreciable serotonin 2A receptor agonist properties that can alter consciousness in a marked and novel way" (Carhart-Harris & Goodwin, 2017). It is important to note that all psychedelic effects are not necessarily explained by this activity; tryptamines such as 5-MeO-DMT, LSD and psilocybin have also been found to act as 5-HT1A agonists (Coyle et al., 2012), LSD has a strong interaction with the five dopamine receptors (D_1, D_2, D_3, D_4 and D_5) , mescaline hits an adrenergic receptor (α_{2C}) (Ray, 2010) and DMT binds to σ-1 receptors (Fontanilla et al., 2009) and inhibits the serotonin transport (SERT) (Cozzi et al., 2009). Although the evidence is inconclusive, psychedelic effects are still thought to emerge as a result of direct receptor actions (Nichols, 2016).

Whether differences in receptor binding profiles result in different subjective effects has scarcely been studied. Holze et al. (2022) compared the acute effects of LSD and psilocybin in a double-blind, placebo-controlled study and found no significant difference of the substances at high doses, rated on the 5D-ASC scale. There was a significant difference between low dosage psilocybin and low dosage LSD, although this could be due to faulty dosage calculation rather than actual differences in drug effects, which also poses a challenge for psychedelic drug comparison (although research has been done on dosing, see e.g. Griffiths et al. (2011) & Hasler et al. (2004)). To this author's knowledge, the difference in acute effects of psychedelic substances such as mescaline, psilocybin, LSD, DMT, and Ayahuasca have not been studied in a framework such as the one used by Holze et al. (2022).

Griffiths et al. (2019) studied perceived God encounters across different psychedelics (LSD, psilocybin, Ayahuasca and DMT) using a survey-based approach and the Mystical Experience Questionnaire (MEQ30).

They found no difference between psilocybin and LSD, substantial difference between Ayahuasca and psilocybin/LSD, little difference between Ayahuasca and DMT and substantial difference between DMT and psilocybin/LSD. The results are likely influenced by differences in demographics between groups and the surrounding culture for substances like Ayahuasca, which is often used in a social and spiritual/religious context. Though the importance of context is paramount, the authors suggest that *N,N*-dimethyltryptamine (DMT, also the psychedelic compound in Ayahuasca) *"produces a unique profile of effects that is phenomenologically distinct from two widely used classic psychedelics (psilocybin and LSD)"* (Griffiths et al., 2019).

Furthermore, anecdotal evidence supporting differences in subjective effects of various psychedelic substances has been extensively reported (Zamberlan et al., 2018). Perhaps the largest and most renowned works on the effects of psychedelic drugs, in which the authors report on the subjective effects of over 200 psychoactive substances, are *PiHKAL* (A. Shulgin & Shulgin, 1992) and *TiHKAL* (Shulgin & Shulgin, 1997).

1.6 Using unstructured reports to study subjective effects of psychedelics

While a firm understanding of the difference in acute effects between psychedelic substances has yet to be established, it would be reasonable to explore various methods to further our knowledge of the matter. An increasing number of papers have been published which include an analysis of already available, retrospectively written experience reports from the Erowid Experience Vaults[1] (Coyle et al., 2012; Hase et al., 2022; Tagliazucchi, 2022; Zamberlan et al., 2018).

A recent review by Tagliazucchi (2022) focused on the effects of psychedelic drugs on speech organization and the semantic content of experience reports from Erowid. A point is made in the paper that, despite difficulties in extracting meaningful and quantitative data from unstructured texts, the information embedded in written reports outweigh that of questionnaires, and that *"natural language reports obtained during the acute effects of psychedelics open a new dimension of analysis beyond the possibilities of psychometric questionnaires"*. Hase et al. (2022) compared linguistic profiles of experience reports from Erowid between six psychedelic drug categories using latent semantic analysis (LSA) and various rating scales from the LIWC2015 library. They found language profiles of drug categories to differ in, among others, emotional intensity and analytical thinking.

[1] https://www.erowid.org/experiences/, 22nd of December, 2022.

Coyle et al. (2012) used a random-forest classifier on a subset of the experience reports from Erowid, where they achieved a 51.1% estimated accuracy on a 10-class classification task, classifying drug class based on report. Zamberlan et al. (2018) used a dataset of binding affinity profiles for various psychedelics combined with experience reports from Erowid, and found a correlation between similarity of binding affinity profiles of drugs and the semantic similarity of written reports associated with the drugs.

These examples show that many have turned to the field of natural language processing to analyze unstructured written reports. Though the mentioned papers have achieved some success, the use of natural language processing as a tool to study subjective experiences between psychedelic substances remains to be demonstrated (Tagliazucchi, 2022). Natural language processing (NLP) has already seen many applications in adjacent fields such as psychiatry, where it has been used to detect linguistic differences in schizophrenia spectrum disorders (Tang et al., 2021), predict suicide risk/ideation (Cook et al., 2016) and discover symptoms that are sometimes missed by professionals (Rezaii et al., 2022).

1.7 The current study

To further the research on the subjective effects of psychedelic substances, the approach of this paper is to use novel NLP methods on retrospective written reports of psychedelic experiences acquired from the Erowid Experience Vaults. Building on the success of Coyle et al. (2012) in predicting drug class, this framework utilizes topic modeling to interpret the difference between classes. The methods of analysis include calculating TF-IDF (term frequency-inverse document frequency) to assess word importance across six psychedelic substance categories: LSD, psilocybin (mushrooms), mescaline, Ayahuasca, DMT and 2C-B. Furthermore, per substance topic modeling is performed using the cutting-edge BERTopic topic model (Grootendorst, 2022), giving insight into the content of subjective reports across psychedelic substances. The work presented here contributes to the overall understanding of the subjective effects of psychedelic substances and the efficacy of applying NLP methods in this field.

2 Methods

2.1 Procedure

2.1.1 Dataset

The dataset of written reports was acquired from the Erowid Experience Vaults[2]. The vaults are an *"attempt to collect, catalog, and publish the wide variety of experiences people have with psychoactive plants and chemicals"*[3]. As of today[4], the vaults contain approx. 39.000 experience reports, spanning across experiences with ~950 different substances. Each entry consist of a title, author title, the substance type and a report on the writer's experience whilst influenced by the substance. Some reports also include information on dose, body weight, route of administration, year of experience, gender and age at the time of experience. To be viewable in the vaults, reports must be read and approved by at least two members of the Erowid Reviewing Crew. Reports are rejected if they are believed to be falsified, impossible to read or do not address the effects of the substance (Erowid Reviewing Crew, 2002).

The data was collected by scraping each report site using the *rvest* package (Wickham, 2022) for R. Only title, substance and report were scraped, as these were the only parameters available for all entries, which means that parameters like dose and route of administration are ignored in the analysis. Many entries were combination substances, i.e. experiences with taking multiple substances at the same time. These were removed to assure only single-substance entries remained. Authors report under pseudonymous handles which were not collected. As reports could countain personal or identifying information, thorough measures where taken to fulfill GDPR regulations. Analysis was performed on a local server and data will be deleted on the 1st of March, 2023. The final dataset contains 4219 reports, describing experiences from six substance groups: LSD, psilocybin (mushrooms), mescaline, Ayahuasca, DMT and 2C-B.

LSD (1130 reports)

Lysergic acid diethylamide (LSD) is a synthesized lysergamide. It is taken orally as a blotter under the tongue.

[2] https://www.erowid.org/experiences/, 22nd of December, 2022.
[3] https://www.erowid.org/experiences/exp_about.cgi, 22nd of December, 2022.
[4] Visited on 22nd of December, 2022

Psilocybin (mushrooms) (1783 reports)

Magic mushrooms contain the psychedelic compound psilocybin. Mushroom type was not accounted for, and approx. 45 different types of fungi are included in this category.

Mescaline (355 reports)

3,4,5-trimethoxyphenethylamine (mescaline), either in the form of mescaline-containing cacti (such as *San Pedro* or *Peyote*) or synthesized mescaline.

Ayahuasca (120 reports)

Ayahuasca is a brew or tea which contains N,N-dimethyltryptamine (DMT) and a monoamine oxidase inhibitor (MAOI) (Nichols, 2016).

DMT (652 reports)

N,N-dimethyltryptamine (DMT) is usually smoked.

2C-B (179 reports)

4-Bromo-2,5-dimethoxyphenethylamine (2C-B) is a psychedelic phenethylamine derivative belonging to the 2C-family, which effects have not yet been fully studied in humans, but is structurally close to mescaline (Papaseit et al., 2018).

Whilst inspecting the reports, some where found to be written in Italian. The *cld3* package (Ooms, 2022) was used to detect all non-english reports which were subsequently excluded.

2.2 Analysis

2.2.1 TF-IDF weighted word clouds

As a first measure of analyzing the experiences of different psychedelic substances, the term frequency-inverse document frequency (TF-IDF) was calculated for all six substance groups in R using the *tidytext* package (Silge & Robinson, 2016). This statistic measures the importance of each word in a document (experience report), importance being high when the word occurs often in the document but rarely in other documents. In this case, words with a high score are more defining for the reported experience (Karabiber, 2021). The term frequency per word is calculated as thus:

$$TF = \frac{no.of\ times\ the\ term\ appears\ in\ the\ document}{no.of\ terms\ in\ the\ document}$$

The inverse document frequency is calculated as thus:

$$IDF = log\left(\frac{no.of\ documents\ in\ the\ corpus}{no.of\ documents\ in\ the\ corpus\ containing\ the\ term}\right)$$

TF-IDF is calculated by multiplying the term frequency with the inverse document frequency:

$$TF - IDF = TF * IDF$$

A TF-IDF score is calculated for each word in each document. The per-substance-group TF-IDF scores were visualized in six word clouds (Figure 1) created in R with the *ggwordcloud* package (Le Pennec & Slowikowski, 2019).

2.2.2 BERTopic topic modelling

BERTopic is an algorithm which uses transformer-based embeddings, clustering techniques and a variation of TF-IDF to create interpretable topics. It leverages the more context-based word and sentence representations of BERT language model(s), which allow for semantically similar texts to be close in vector space, opposed to the conventional bag-of-words topic models such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003; Grootendorst, 2022). Topic modeling with BERTopic was done in Python using the *bertopic* package (Grootendorst, 2022).

There are four steps in generating topics with BERTopic: 1) Converting documents to embedding representations, 2) reducing dimensionality of embeddings, 3) clustering the embeddings and 4) extracting topic representations (Grootendorst, 2022). Topics were generated for each of the six substance categories, resulting in six different topic models. The document embeddings were extracted using all-mpnet-base-v2[5], a pre-trained all-purpose Sentence-Transformers model trained on a dataset of over 1 billion training pairs (Reimers & Gurevych, 2019). The dimensionality of embeddings was reduced using the UMAP algorithm (McInnes et al., 2018). Documents were then clustered using the HDBSCAN algorithm (McInnes et al., 2017). To derive topics from clusters, important words were extracted using class-based TF-IDF (c-TF-IDF).

[5] https://huggingface.co/sentence-transformers/all-mpnet-base-v2, 22nd of December, 2022

By combining the documents pertaining to each cluster into one, the clusters are considered single documents. Then TF-IDF is calculated, resulting in multiple important words per cluster/topic. This is effectively the topic representation in the BERTopic algorithm (Grootendorst, 2020).

Preprocessing

Since BERTopic uses document embeddings, data should not be preprocessed by means of tokenization, lemmatization or removal of punctuation and digits, which is of paramount importance in typical topic modeling frameworks. It is also advised to keep stop words, because these provide important contextual information which affect the embeddings (Grootendorst, 2020). Stop words can be removed after the model is fit, so as to not show up in the topic representations. This was done for all models, using a combined stop word list consisting of common english stop words, contractions and psychedelic-specific stop words.

For each substance-category, the topic representations were visualized in three different ways: 1) A bar chart which displays top words per topic, 2) a plot of the topics and document embeddings reduced to 2D-space and 3) a topic similarity matrix. All plots were created with the *bertopic* package (Grootendorst, 2022).

Hyperparameter tuning

The goal of topic modeling is to understand what topics are present in the documents. Thus, tuning the topic model is done to further this understanding by maximizing the interpretability of topic representations. In the BERTopic model, various parameters can be adjusted to achieve this, including cluster size, sample size and n-gram range. For all topic models, bigrams were used as these maximized interpretability. Across all substance groups, it was found that 4-5 clusters (topics) yielded best interpretability.

2.3 Results

2.3.1 TF-IDF weighted word clouds

The six word clouds can be seen in Figure 1.

LSD

The LSD group contains important words which can be sub-grouped into psychic/mystical effects (e.g. *epiphanies, thesis, dreams, extraordinary, flashbacks, philosophers, theory*), somatic effects (e.g. *pain, headaches, migraine, itching*) and narrative experience/activity words (e.g. *tunes, temple, chair, yoga, people*).

Psilocybin (mushrooms)

The psilocybin group is mostly dominated by words related to preparation and route of administration (e.g. *swallow, eating, truffles, seeds, ounce, milk, cakes*). Some words also relate to effects of the substance, both somatic (e.g. *migraines, pain, cluster, sore, submit, veils*) and cognitive/psychic (e.g. *OCD, neuron*).

Mescaline

The majority of the most important words for the mescaline group relate to the preparation and route/method of administration (e.g. *swallow, powdered, ingestion, intestinal, inflection, chew, cuttings, buttons, cut, trays, powder, drinks, extraction, boiled, sift*). This is perhaps because the preparation process for cacti is explained in reports.

Ayahuasca

The Ayahuasca group is dominated by words pertaining to the ritual or traditions around the brew (e.g. *cult, shaman, ceremony, sessions, ritual, leaders, retreats*). Locations are also present such as *Peru*, which carries cultural weight. Some words also suggest uncomfortable experiences (e.g. *harassed, assault, forcibly*).

DMT

The most important words for experiences in the DMT group are related to altered perception and psychic effects (e.g. *forget, concious, dream, hyperspace, simulation, merging, pattern*). There are also words related to route/method of administration (e.g. *product, insufflation*) and substance type (e.g. *tryptamine, benzos, oil*). A few specific words give insight into the narrative experiences, such as *bongo, snakes, deer,* and *jester*. The word *pineal* is also present, perhaps due to the false notion that enough DMT is naturally produced in the pineal gland to cause psychedelic effects (Nichols, 2018).

2C-B

The most important words in the 2C-B group are largely related to route/method of administration (e.g. *nasal, foil, pill, capsules, mucus, dosage, snorting*) and somatic effects of the substance (e.g. *cough, tremors, seizure, shivering*). Other substances (e.g. *molly* and *nexus*) are also important in some documents, perhaps due to comparison between substances.

Figure 1. Six word clouds based on the TF-IDF weightings across substance groups.

2.3.2 BERTopic topic modeling

The following paragraphs are summations of the top term bar charts for each substance category. All corresponding 2D-document-topic plots and similarity matrices for the substance groups can be found in appendix A through F.

LSD

The topic model for the LSD-group yielded five topics. Topic 0 is interpreted as the *time* topic, including bigrams such as *started feel, time trip, time time, 20 minutes, hour half, hour trop* and more. This suggests that a focus of LSD-reports is the time or chronology of the experience, perhaps commenting on different phases of the experience. Topic 1 is interpreted as an *activity* topic, is it seems to include activities which were performed during the experience such as *decided walk, listening music, remember thing, looked mirror, started walking, left house* and others.

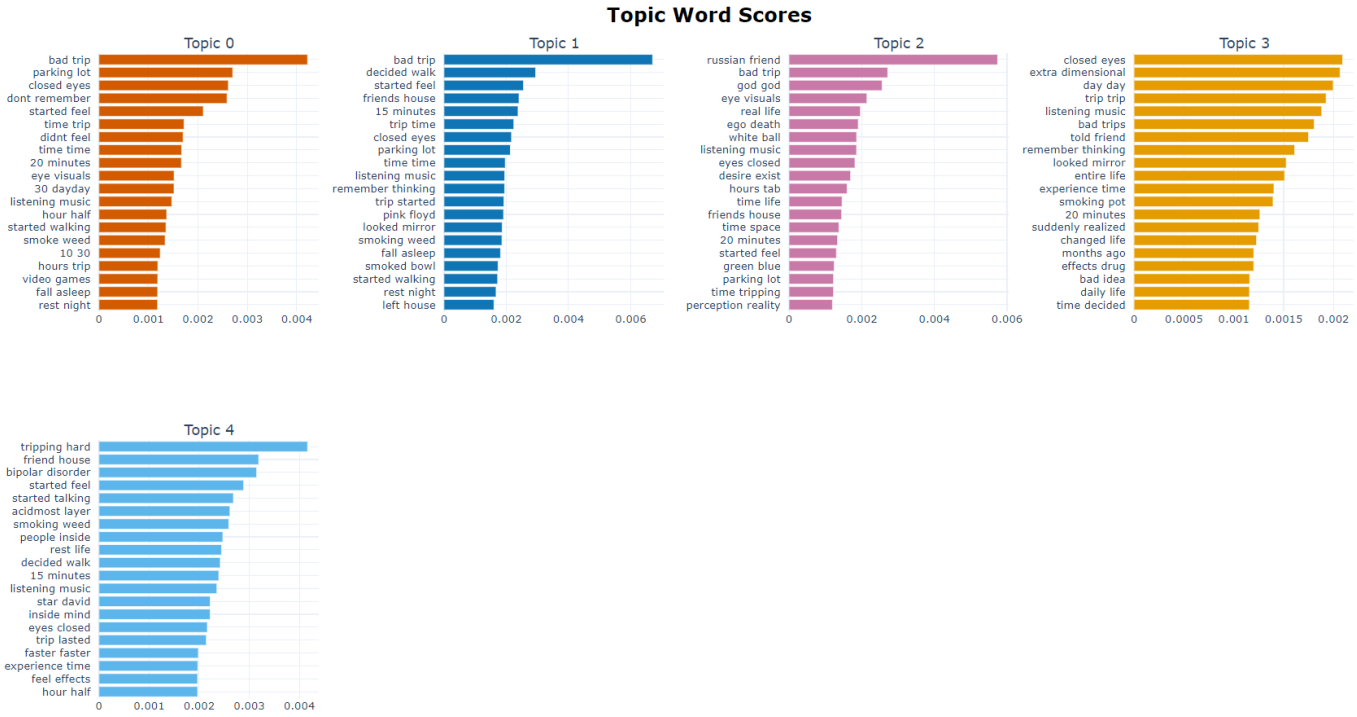


Figure 2. Top terms for topics from the LSD topic model determined by c-TF-IDF.

The activities are mostly everyday activities, perhaps suggesting that the setting of LSD-experiences is often a relaxed, everyday location. Topic 2 and topic 3 could both be interpreted as *existence* topics.

They both contain several bigrams which describe abstract, existential or mystical thoughts, e.g. *god god, real life, ego death, desire exist, time life, perception reality* in topic 2 and *extra dimensional, entire life, experience time, changed life, daily life* in topic 3. These topics inform of the mind altering or psychedelic effects of LSD. Topic 4 is less interpretable, being a mix of the aforementioned topics. There is high similarity between all five topics, as can be seen in the similarity matrix in appendix A.

Psilocybin (mushrooms)

The topic model for the psilocybin (mushrooms) dataset produced five topics after tuning. Similarity between topics from this model is very high, and there are only few words in each topic which do not overlap with other topics. Topic 0 is a *time* topic. It includes words pertaining to the progress or chronology of the experience, e.g. *20 minutes, time time, trip started, started thinking, trip time, time trip* etc. Topic 1 is concerned with *existence*, including bigrams such as *psychedelic experience, time space, entire life, meaning life, time life* and *life changing*. Topics 2 are a mix of *existence* and *perceptual effects*, including the bigrams *eye visuals, ego death, tripping hard* in topic 2 and *ego death, love love, field vision* in topic 3. Topic 4 is concerned more with people, with bigrams such as *people people, sober people, friend house* and *looked friend*, suggesting a social element is present for experiences with mushrooms.

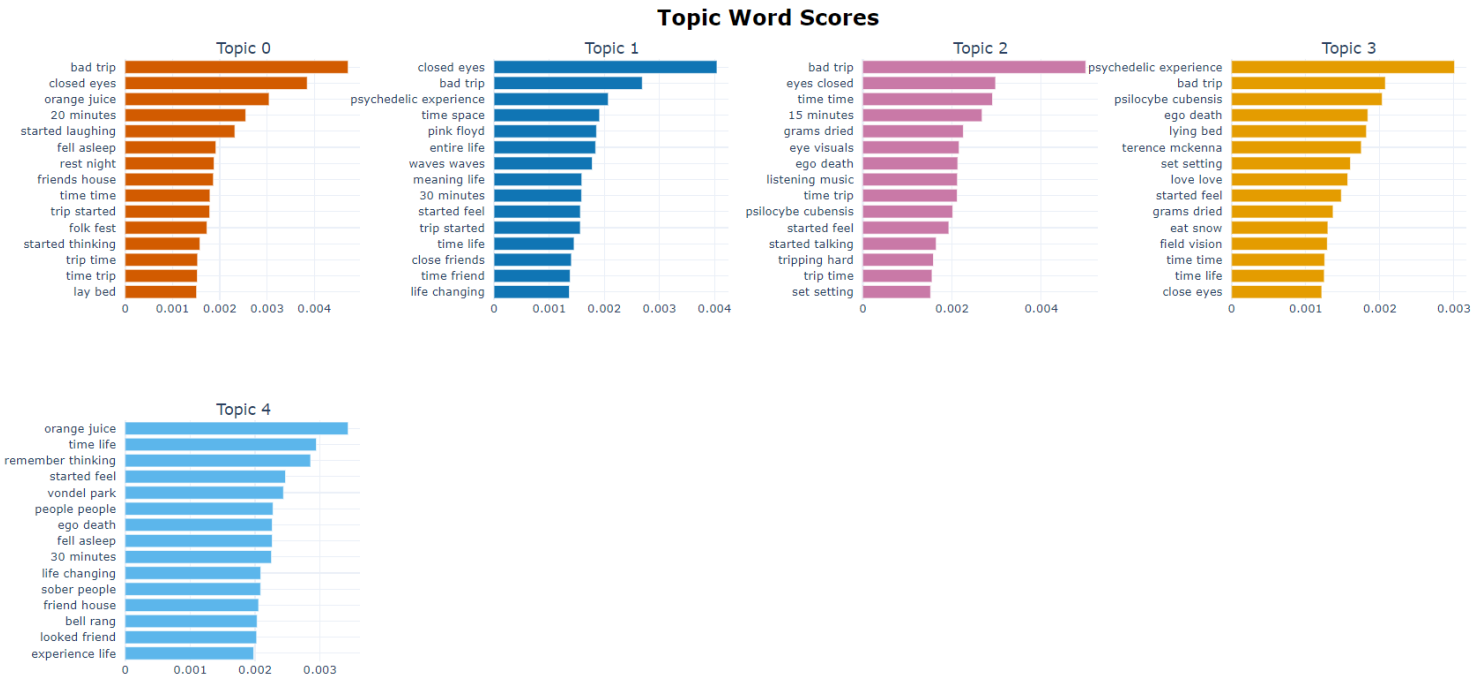


Figure 3. Top terms for topics from the psilocybin topic model determined by c-TF-IDF.

Mescaline

The topic model for the mescaline dataset produced four topics. Topic 0 is a *time* topic, as was also present in the LSD and psilocybin groups. This topic includes bigrams such as *00 hrs, started feel, experience time, 45 minutes* and *half hour*. Topic 1 can be interpreted as a *preparation* topic. All top bigrams in this topic are associated with preparation of mescaline, most of them associated with the preparation of cacti as can be seen it the bigrams *dark green, green flesh, plant material, stock pot, green liquid, low heat* and *green layer*. The other bigrams (*lemon juice, distilled water* and *grapefruit juice*) are perhaps informative of how the substance was administered. Topic 2 and 3 are *activity* topics, including *closed eyes, spirit walk, water drum, set setting* in topic 2 and *eyes closed, lying bed, fell asleep*. Topic 3 also includes some effects or *existence* bigrams such as *visual effects, beauty life* and *body energy*.

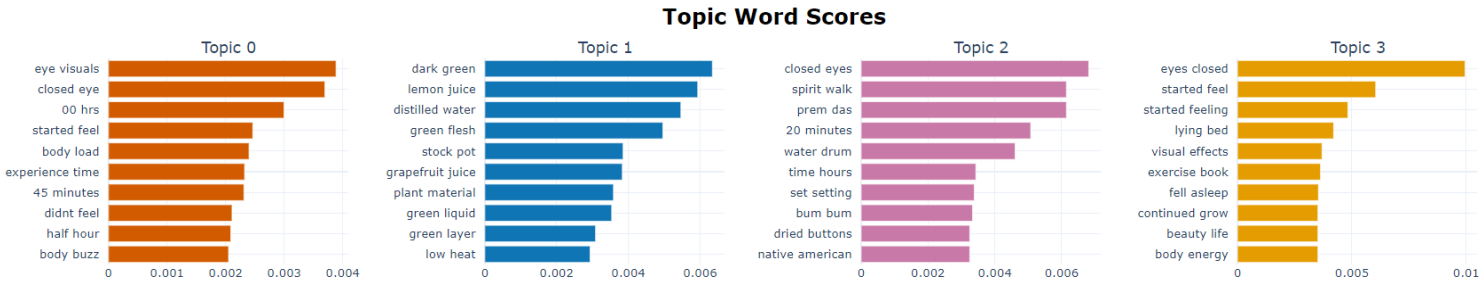


Figure 4. Top terms for topics from the mescaline topic model determined by c-TF-IDF.

Ayahuasca

The topic model for the Ayahuaca-group yielded four topics. The similarity score for topics from this model are relatively low compared to the other substance-groups. Topic 0 is mostly related to *ceremony*, including bigrams such as *agua florida* (a fragrance used by shamans in ceremonial settings), *shaman apprentice, fellow tripper, effects medicine, time drank, blowing tobacco* and *hours ceremony*. This suggests that many of the experiences have taken place in native ceremonial settings. Topic 1 is related to *feeling*, as the word *feel* appears several times in this topic, e.g. *remember feeling, feel sick, feel feel, day feeling* and *started feel*. Perhaps this suggests that the immediate somatic effects of the Ayahuasca brew (i.e. nausea) are often focused on during experiences. Topic 2 is related to *existence and religion*, including bigrams such as *santo daime* (Brazilian religion which uses Ayahuasca in ceremonies), *sacred geometry, divine mind, blessings blessings* and *spirit earth*, which inform of the psychedelic effects of the substance.

Topic 3 is less interpretable, but is related to the substance itself (e.g. *aya brew, psychotria viridis*) and the set/setting (e.g. *experience started, setting mood, experience familiar*).

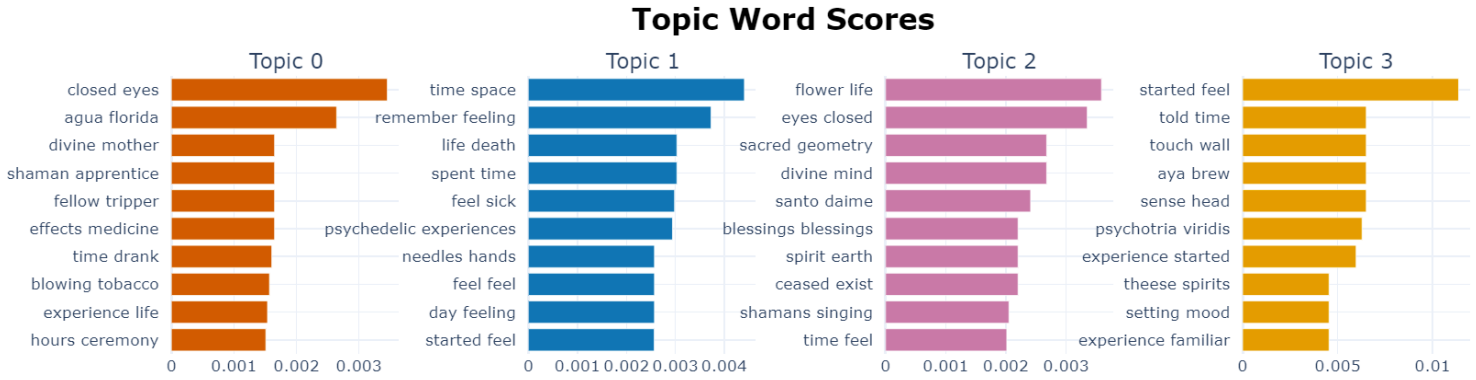


Figure 5. Top terms for topics from the Ayahuasca topic model determined by c-TF-IDF.

DMT

The topic model for the DMT dataset yields four topics. Topic 0 is related to the experience of time, with bigrams such as *entire life, lasted minutes* and *trip time*. There are also words related to body feeling, such as *eyes closed, deep breaths* and *entire body*. Topic 1 is mostly related to perceptual effects, including *eyes closed, patterned light, visual field, feel body* and *geometrical shapes*. Topics 2 and 3 are *existence* topics, relating to mystical or existential words such as *fourth dimension, individuated consciousness, realms existence, experience life* and *totality existence* in topic 2, and *experience life, psychedelic experience* and *experience time* in topic 3, which informs of the psychedelic effects of DMT.

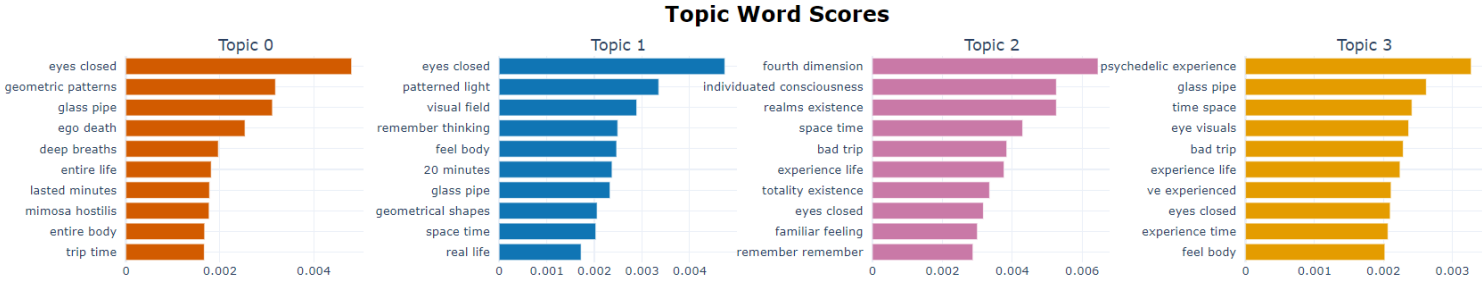


Figure 6. Top terms for topics from the DMT topic model determined by c-TF-IDF.

2C-B

The topic model for the 2C-B dataset produced four topics. Topic 0 is related to *trip* or *tripping*, which is concerned with the intensity or onset of effects. This can be seen in bigrams such as *bad trip, started feel, trip report, trip started* and *tripping hard*. Topic 1 is concerned with perceptual effects, such as *body load, depth perception, visual effects, eye visuals* and *music sounds*. This suggests that perceptual effects, such as visual and auditory alterations, are, perhaps, more present for this substance. Topic 2 also includes some perceptual effect bigrams, but is mostly related to time (e.g. *decided time, half hour, minutes hour, 15 minutes* and *40 minutes*). Topic 3 relates to effects as well, but with more focus on *experience*, as this word is frequent in the top terms (e.g. *psychedelic experience, experience experience, effects note* and *experiences substance*).

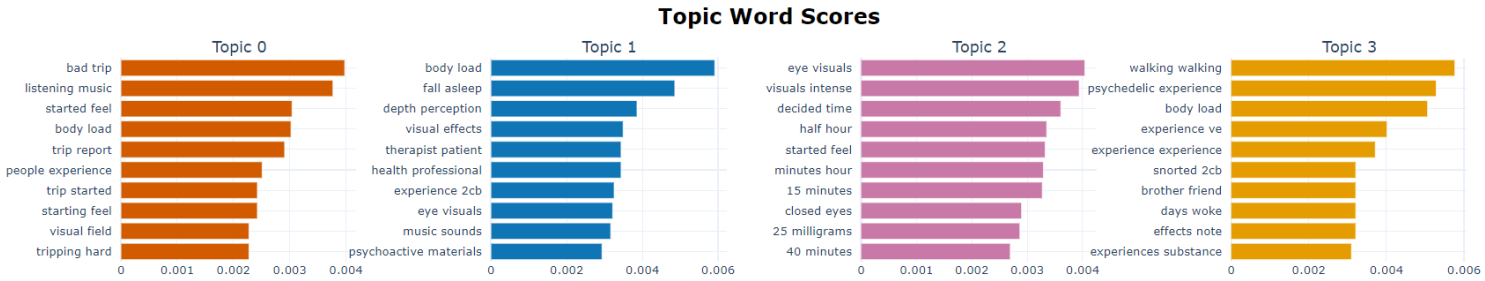


Figure 7. Top terms for topics from the 2C-B topic model determined by c-TF-IDF.

3 Discussion

3.1 Similarities between topics and groups

The TF-IDF-weighted word clouds provide a first glance at the contents of reports across substance-groups. Many of the words and themes of the word clouds can be gauged in the topic models, but the points in which they differ is of interest. Most notably, for the psilocybin group, the TF-IDF weightings revealed more importance for words pertaining to route of administration than was observed in the topic representations. Interestingly, the highest TF-IDF weightings for the 2C-B group were solely route of administration words, whereas these were scarcely present in the topic representations. In comparing the topic model outputs, it becomes apparent that reports generally include similar topics. All substance groups except Ayahuasca (and to a lesser degree DMT and 2C-B) include a *time* topic, which focuses on the chronology of the experience, which likely includes time of administration, onset of effects and perhaps when effects wear off.

Although it does not inform directly of the effects of substance, including time in experience reports could be valuable for gauging effect duration between substances, which could be the aim of further research. Perhaps the absence of time bigrams in the Ayahuasca group is due to the setting in which users administer the substance, as is suggested by the *ceremony* topic, though it is notable that *time* bigrams are also scarce (but not absent) in the DMT group, as DMT is the psychedelic compound in Ayahuasca.

Furthermore, most models produced an *existence* topic, which informs of the cognitive effects of psychedelics. The bigrams in these topics suggest altered though processes and novel (or altered) perspectives on beliefs, existence and/or daily life. They vary across substances; LSD and psilocybin share bigrams relating to life and *ego death*, Ayahuasca has a more spiritual/religious focus as seen by *divine mind, spirit earth* and *sacred geometry*, and DMT, although similar, seems more related to consciousness and dimensions. This echoes some of the findings of Griffiths et al. (2019), who found no difference between psilocybin and LSD and little difference between Ayahuasca and DMT when studying God encounter experiences. The 2C-B topics are notable in this regard for not having a single bigram relating to the cognitive effects of the substance. Perhaps this substance does not invoke the same degree of thought alteration, but it stands out in terms of *perceptual effects* bigrams, as is explained in the following paragraph.

Psilocybin, DMT and 2C-B have a *perceptual effects* topic. These topics include bigrams that inform in some way of the perceptual effects present during the experience. The perceptual effects topics for psilocybin and DMT mostly include visual alterations, whereas the 2C-B model produced two *perceptual effects* topics which include both visual, haptic and auditory effect bigrams.

LSD and mescaline share an *activity* topic, which is named so as the bigrams describe the activities presumably performed during the experience. These probably play a role in the narrative structure of the experience reports, supposing that it is structured chronologically. These are less informative of the effects of the substances, but perhaps they give insight into the actions frequently performed by users, i.e. walking, lying in bed and listening to music.

The apparent spiritual character of all Ayahuasca topics would seem to make this substance-group quite unique, as it is the substance with most ritual and practice attached.

Perhaps the same can be said for mescaline, as this is also used in religious ceremonies, but it is not as clear from the topic representations. The topics for mescaline do not contribute much to understanding the cognitive or perceptual effects, perhaps because a substantial part of the experience reports are dedicated to preparation of the substance, as seen by the *preparation* topic.

3.2 Limitations of data

The use of unstructured self-reported texts poses some problems when applying NLP methods. Although the reports in the Erowid Experience Vaults are moderated, reports suffer from bad grammar and misspellings, which changes how the data is processed by the NLP models. Furthermore, some reports were observed which only contained recipes for preparing substances, without any detail as to the experience with the substance. This introduces unwanted noise in the data, as the reports are assumed to provide insight into the experience of taking the substance.

Furthermore, the main limitation of using unstructured self-reported texts is the lack of controlled variables. Since the reports are only moderated based on the content of the reports, it is impossible to know whether the writer actually took the substance, if the dose was correct, what the quality of the substance was, etc. Since the reports should be written retrospectively, the content also relies on the memory of the writer, and there are no rules for reporting experiences which happened many years back. The free-form format also relies on the writer's ability to express themselves verbally, which can be difficult as experiences are complex.

Another limitation to the data is an observed pattern in the experience reports; reports are often written chronologically and include things that happen leading up to the experience, i.e. how much the writer had slept, what they had eaten, what their mood was, whether they had tried the substance before etc. These passages do not inform of the effects of substances when performing topic modelling, so they are viewed as noise in the data.

3.3 Limitations of BERTopic modeling

Although the topic representations from the six BERTopic models provide insight into the experiences of the substances, the lack of controlled variables does not allow for strong conclusions on causal relationships.

For instance, whether the spiritual character of Ayahuasca topics is due to properties inherent in the brew or due to the shamanistic set and setting is yet to be understood. It is also based on the viewers capability of interpreting the topic representations, which is an inherently subjective affair. Although BERT embeddings leverage the context of a sentence, the topic representations are still prone to misinterpretation.

4 Conclusion

The differences in acute effects between psychedelic substances with different binding affinity profiles is yet to be fully understood. The vast datacorpus of unstructured psychedelic experience reports have revealed differences across substances when applying psychometric evaluations (Griffiths et al., 2019), NLP methods such as LSA and ML-classifcication (Hase et al., 2022; Coyle et al., 2012) and in comparing reports with binding affinity profiles for substances (Zamberlan et al., 2018). The aim of this project was to further explore the efficacy of NLP methods, more specifically topic modelling, as a means of understanding the subjective experience of six psychedelic substances. The TF-IDF word clouds and topic representations produced by the six BERTopic models were interpretable and differ across substance groups, highlighting the differences in which users report their experience based on the administered substance. Based on the topics, reports might focus more on perceptual effects, cognitive effects, activities or preparation across substance groups. This may suggest differences in the subjective effects of psychedelic substances, but the lack of controlled variables in unstructured reports, importance of set and setting and limitations of topic modelling call for more research to be done in the field. What topic models do reveal, though, are observable differences in which users report the experiences of psychedelic substances.

5 References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bodmer, I., Dittrich, A., & Lamparter, D. (1994). Altered states of consciousness–their common structure and assessment. *Welten Des Bewusstseins Bd. 3, Experimentelle Psychologie, Neurobiologie Und Chemie*, 45–58.

Bogenschutz, M. P., Forcehimes, A. A., Pommy, J. A., Wilcox, C. E., Barbosa, P. C., & Strassman, R. J. (2015). Psilocybin-assisted treatment for alcohol dependence: A proof-of-concept study. *Journal of Psychopharmacology*, *29*(3), 289–299.

Carhart-Harris, R. L., & Goodwin, G. M. (2017). The Therapeutic Potential of Psychedelic Drugs: Past, Present, and Future. *Neuropsychopharmacology*, *42*(11), Article 11. https://doi.org/10.1038/npp.2017.84

Carhart-Harris, R. L., Roseman, L., Haijen, E., Erritzoe, D., Watts, R., Branchi, I., & Kaelen, M. (2018). Psychedelics and the essential importance of context. *Journal of Psychopharmacology*, *32*(7), 725–731. https://doi.org/10.1177/0269881118754710

Carter, O. L., Burr, D. C., Pettigrew, J. D., Wallis, G. M., Hasler, F., & Vollenweider, F. X. (2005). Using psilocybin to investigate the relationship between attention, working memory, and the serotonin 1A and 2A receptors. *Journal of Cognitive Neuroscience*, *17*(10), 1497–1508.

Chwelos, N., Blewett, D. B., Smith, C. M., & Hoffer, A. (1959). Use of d-lysergic acid diethylamide in the treatment of alcoholism. *Quarterly Journal of Studies on Alcohol*, *20*(3), 577–590.

Cook, B. L., Progovac, A. M., Chen, P., Mullin, B., Hou, S., & Baca-Garcia, E. (2016). Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Computational and Mathematical Methods in Medicine*, *2016*.

Coyle, J. R., Presti, D. E., & Baggott, M. J. (2012). *Quantitative Analysis of Narrative Reports of Psychedelic Drugs* (arXiv:1206.0312). arXiv. https://doi.org/10.48550/arXiv.1206.0312

Cozzi, N. V., Gopalakrishnan, A., Anderson, L. L., Feih, J. T., Shulgin, A. T., Daley, P. F., & Ruoho, A. E. (2009). Dimethyltryptamine and other hallucinogenic tryptamines exhibit substrate behavior at the serotonin uptake transporter and the vesicle monoamine transporter. *Journal of Neural Transmission*, *116*(12), 1591–1599.

DeRogatis, J. (2003). *Turn on your mind: Four decades of great psychedelic rock*. Hal Leonard Corporation.

Dickins, R. J. (2012). *The Birth of Psychedelic Literature: Drug Writing and the rise of LSD Therapy 1954–1964*. https://ore.exeter.ac.uk/repository/handle/10036/4207

Dittrich, A. (1994). Psychological aspects of altered states of consciousness of the LSD type: Measurements of their basic dimensions and prediction of individual differences. *Pletscher A, Ladewig D (Eds)*, *50*, 101–118.

Dittrich, A., Lamparter, D., & Maurer, M. (2006). 5D-ASC: Questionnaire for the assessment of altered states of consciousness. A short introduction. *Psychologisches Institut Fur Beratung Und Forschung, Zurich, Switzerland*.

Erowid Reviewing Crew. (2002, October). *Experience Report Reviewing: The Good, the Bad and the Ugly*. https://www.erowid.org/experiences/exp_info1.shtml

Fontanilla, D., Johannessen, M., Hajipour, A. R., Cozzi, N. V., Jackson, M. B., & Ruoho, A. E. (2009). The hallucinogen N, N-dimethyltryptamine (DMT) is an endogenous sigma-1 receptor regulator. *Science*, *323*(5916), 934–937.

Gallagher, M. (2004). Tripped out: The psychedelic film and masculinity. *Quarterly Review of Film and Video*, *21*(3), 161–171.

Garcia-Romeu, A. P., & Tart, C. T. (2013). Altered states of consciousness and transpersonal psychology. *The Wiley-Blackwell Handbook of Transpersonal Psychology*, 121–140.

Gasser, P., Holstein, D., Michel, Y., Doblin, R., Yazar-Klosinski, B., Passie, T., & Brenneisen, R. (2014). Safety and efficacy of lysergic acid diethylamide-assisted psychotherapy for anxiety associated with life-threatening diseases. *The Journal of Nervous and Mental Disease*, *202*(7), 513.

Gouzoulis-Mayfrank, E., Thelen, B., Maier, S., Heekeren, K., Kovar, K.-A., Sass, H., & Spitzer, M. (2002). Effects of the hallucinogen psilocybin on covert orienting of visual attention in humans. *Neuropsychobiology*, *45*(4), 205–212.

Griffiths, R. R., Hurwitz, E. S., Davis, A. K., Johnson, M. W., & Jesse, R. (2019). Survey of subjective "God encounter experiences": Comparisons among naturally occurring experiences and those occasioned by the classic psychedelics psilocybin, LSD, ayahuasca, or DMT. *PLOS ONE*, *14*(4), e0214377. https://doi.org/10.1371/journal.pone.0214377

Griffiths, R. R., Johnson, M. W., Richards, W. A., Richards, B. D., McCann, U., & Jesse, R. (2011). Psilocybin occasioned mystical-type experiences: Immediate and persisting dose-related effects. *Psychopharmacology*, *218*(4), 649–665.

Grob, C. S., Danforth, A. L., Chopra, G. S., Hagerty, M., McKay, C. R., Halberstadt, A. L., & Greer, G. R. (2011). Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Archives of General Psychiatry*, *68*(1), 71–78.

Grootendorst, M. (2020, October 6). *Topic Modeling with BERT*. Medium. https://towardsdatascience.com/topic-modeling-with-bert-779f7db187e6

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Grunenberg, C., & Harris, J. (2005). *Summer of love: Psychedelic art, social crisis and counterculture in the 1960s* (Vol. 8). Liverpool University Press.

Hase, A., Erdmann, M., Limbach, V., & Hasler, G. (2022). Analysis of recreational psychedelic substance use experiences classified by substance. *Psychopharmacology*, *239*(2), 643–659. https://doi.org/10.1007/s00213-022-06062-3

Hasler, F., Grimberg, U., Benz, M. A., Huber, T., & Vollenweider, F. X. (2004). Acute psychological and physiological effects of psilocybin in healthy humans: A double-blind, placebo-controlled dose–effect study. *Psychopharmacology*, *172*(2), 145–156.

Holze, F., Ley, L., Müller, F., Becker, A. M., Straumann, I., Vizeli, P., Kuehne, S. S., Roder, M. A., Duthaler, U., Kolaczynska, K. E., Varghese, N., Eckert, A., & Liechti, M. E. (2022). Direct comparison of the acute effects of lysergic acid diethylamide and psilocybin in a double-blind placebo-controlled study in healthy subjects. *Neuropsychopharmacology*, *47*(6), Article 6. https://doi.org/10.1038/s41386-022-01297-2

Jaffe, J. H. (1990). Drug addiction and drug abuse. *Goodman and Gilman's The Pharmacological Basis of Therapeutics (Goodman AG, Rall TW, Nies AS, and Taylor P Eds)*, *8th ed*, 522–573.

Johnson, M. W., Garcia-Romeu, A., Cosimano, M. P., & Griffiths, R. R. (2014). Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. *Journal of Psychopharmacology*, *28*(11), 983–992.

Johnson, M. W., Garcia-Romeu, A., & Griffiths, R. R. (2017). Long-term follow-up of psilocybin-facilitated smoking cessation. *The American Journal of Drug and Alcohol Abuse*, *43*(1), 55–60.

Karabiber, F. (2021). *TF-IDF — Term Frequency-Inverse Document Frequency – LearnDataSci*. https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/

Krebs, T. S., & Johansen, P.-Ø. (2012). Lysergic acid diethylamide (LSD) for alcoholism: Meta-analysis of randomized controlled trials. *Journal of Psychopharmacology*, *26*(7), 994–1002.

Le Pennec, E., & Slowikowski, K. (2019). *ggwordcloud: A Word Cloud Geom for "ggplot2"* (R package version 0.5.0). https://CRAN.R-project.org/package=ggwordcloud

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, *2*(11), 205.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *ArXiv Preprint ArXiv:1802.03426*.

Moreno, F. A., Wiegand, C. B., Taitano, E. K., & Delgado, P. L. (2006). Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. *Journal of Clinical Psychiatry*, *67*(11), 1735–1740.

Nichols, D. E. (2004). Hallucinogens. *Pharmacology & Therapeutics*, *101*(2), 131–181. https://doi.org/10.1016/j.pharmthera.2003.11.002

Nichols, D. E. (2016). Psychedelics. *Pharmacological Reviews*, *68*(2), 264–355. https://doi.org/10.1124/pr.115.011478

Nichols, D. E. (2018). N, N-dimethyltryptamine and the pineal gland: Separating fact from myth. *Journal of Psychopharmacology*, *32*(1), 30–36.

Nichols, D. E., Johnson, M., & Nichols, C. (2017). Psychedelics as Medicines: An Emerging New Paradigm. *Clinical Pharmacology & Therapeutics*, *101*(2), 209–219. https://doi.org/10.1002/cpt.557

Ooms, J. (2022). *Google's Compact Language Detector 3* (R package version 1.4.4). https://CRAN.R-project.org/package=cld3

Papaseit, E., Farré, M., Pérez-Mañá, C., Torrens, M., Ventura, M., Pujadas, M., De la Torre, R., & González, D. (2018). Acute pharmacological effects of 2C-B in humans: An observational study. *Frontiers in Pharmacology*, *9*, 206.

Preller, K. H., & Vollenweider, F. X. (2016). Phenomenology, structure, and dynamic of psychedelic states. *Behavioral Neurobiology of Psychedelic Drugs*, 221–256.

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (arXiv:1908.10084). arXiv. https://doi.org/10.48550/arXiv.1908.10084

Rezaii, N., Wolff, P., & Price, B. H. (2022). Natural language processing in psychiatry: The promises and perils of a transformative approach. *The British Journal of Psychiatry*, *220*(5), 251–253.

Sanches, R. F., de Lima Osório, F., Dos Santos, R. G., Macedo, L. R., Maia-de-Oliveira, J. P., Wichert-Ana, L., de Araujo, D. B., Riba, J., Crippa, J. A. S., & Hallak, J. E. (2016). Antidepressant effects of a single dose of ayahuasca in patients with recurrent depression: A SPECT study. *Journal of Clinical Psychopharmacology*, *36*(1), 77–81.

Savage, C., & McCabe, O. L. (1973). Residential psychedelic (LSD) therapy for the narcotic addict: A controlled study. *Archives of General Psychiatry*, *28*(6), 808–814.

Shulgin, A., & Shulgin, A. (1992). *PiHKAL: A chemical love story*. Transform Press Berkeley, CA, USA.

Shulgin, A. T., & Shulgin, A. (1997). *TIHKAL: The continuation* (Vol. 546). Transform press Berkeley.

Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *Journal of Open Source Software*, *1*(3), 37. https://doi.org/10.21105/joss.00037

Strassman, R. J., Qualls, C. R., Uhlenhuth, E. H., & Kellner, R. (1994). Dose-response study of N, N-dimethyltryptamine in humans: II. Subjective effects and preliminary results of a new rating scale. *Archives of General Psychiatry*, *51*(2), 98–108.

Studerus, E., Gamma, A., & Vollenweider, F. X. (2010). Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PloS One*, *5*(8), e12412.

Tagliazucchi, E. (2022). Language as a window into the altered state of consciousness elicited by psychedelic drugs. *Frontiers in Pharmacology*, 900.

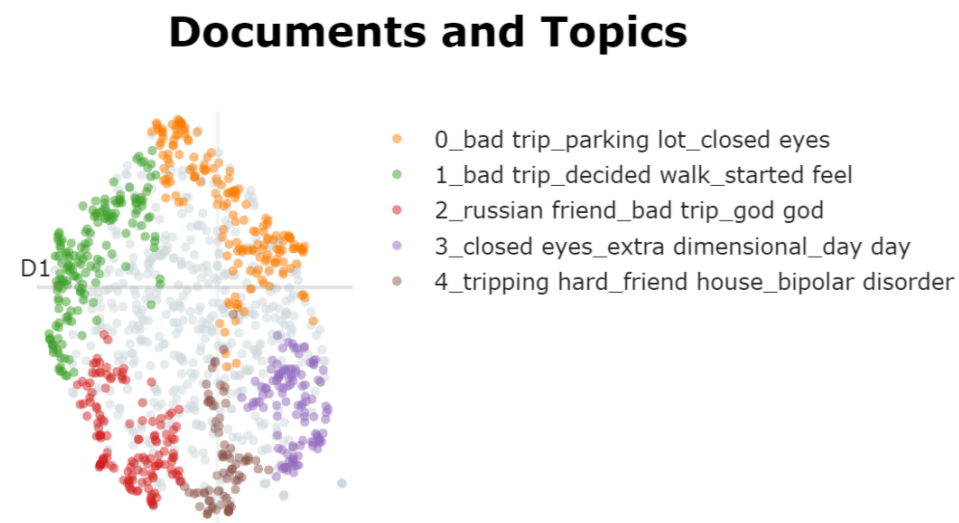
Tang, S. X., Kriz, R., Cho, S., Park, S. J., Harowitz, J., Gur, R. E., Bhati, M. T., Wolf, D. H., Sedoc, J., & Liberman, M. Y. (2021). Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *Npj Schizophrenia*, *7*(1), 1–8.

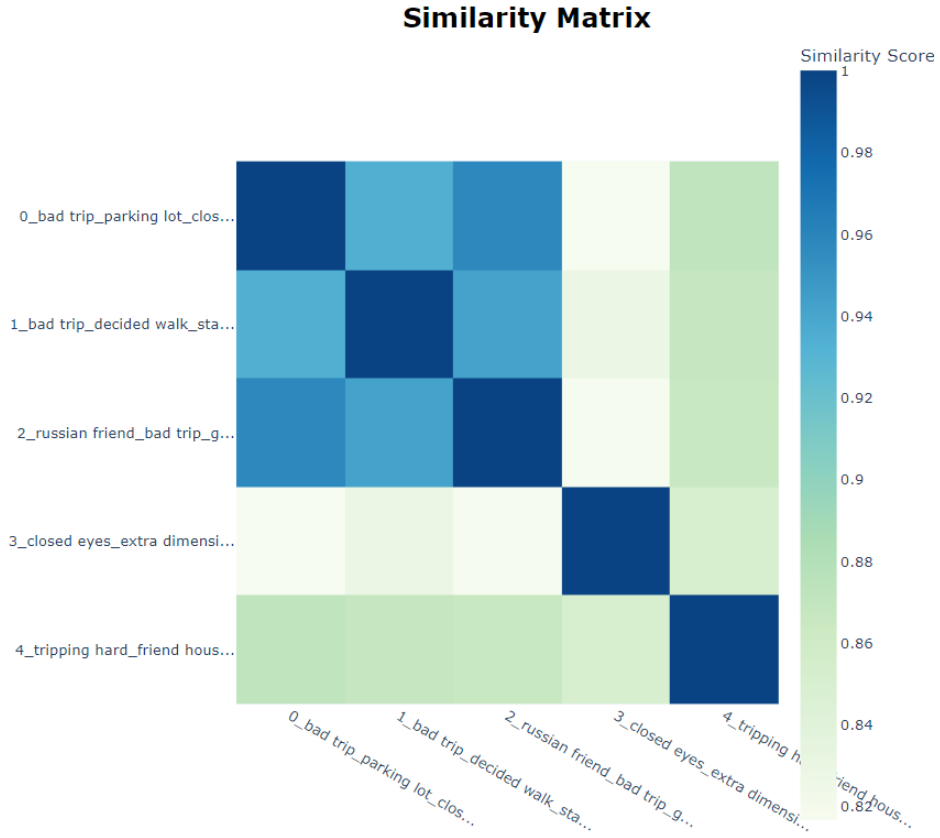
Wickham, H. (2022). *rvest: Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org/

Zamberlan, F., Sanz, C., Martínez Vivot, R., Pallavicini, C., Erowid, F., Erowid, E., & Tagliazucchi, E. (2018). The varieties of the psychedelic experience: A preliminary study of the association between the reported subjective effects and the binding affinity profiles of substituted phenethylamines and tryptamines. *Frontiers in Integrative Neuroscience*, *12*, 54.

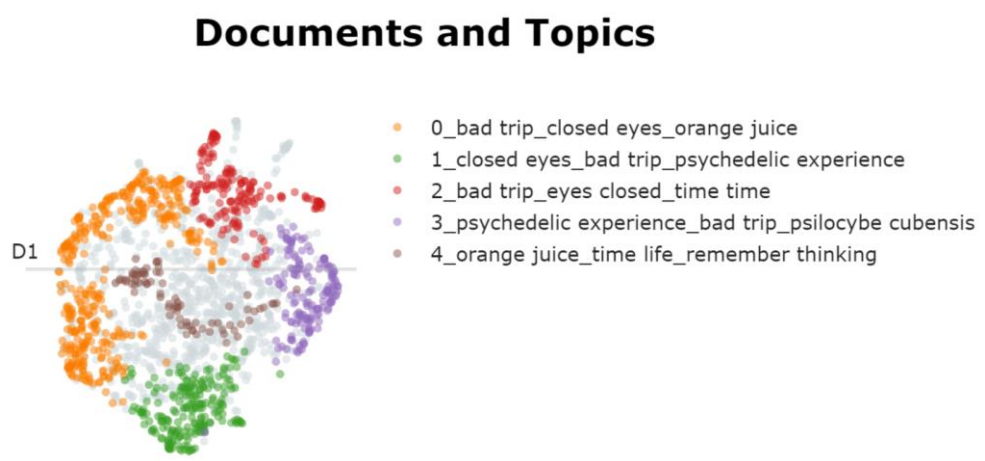
6 Appendices

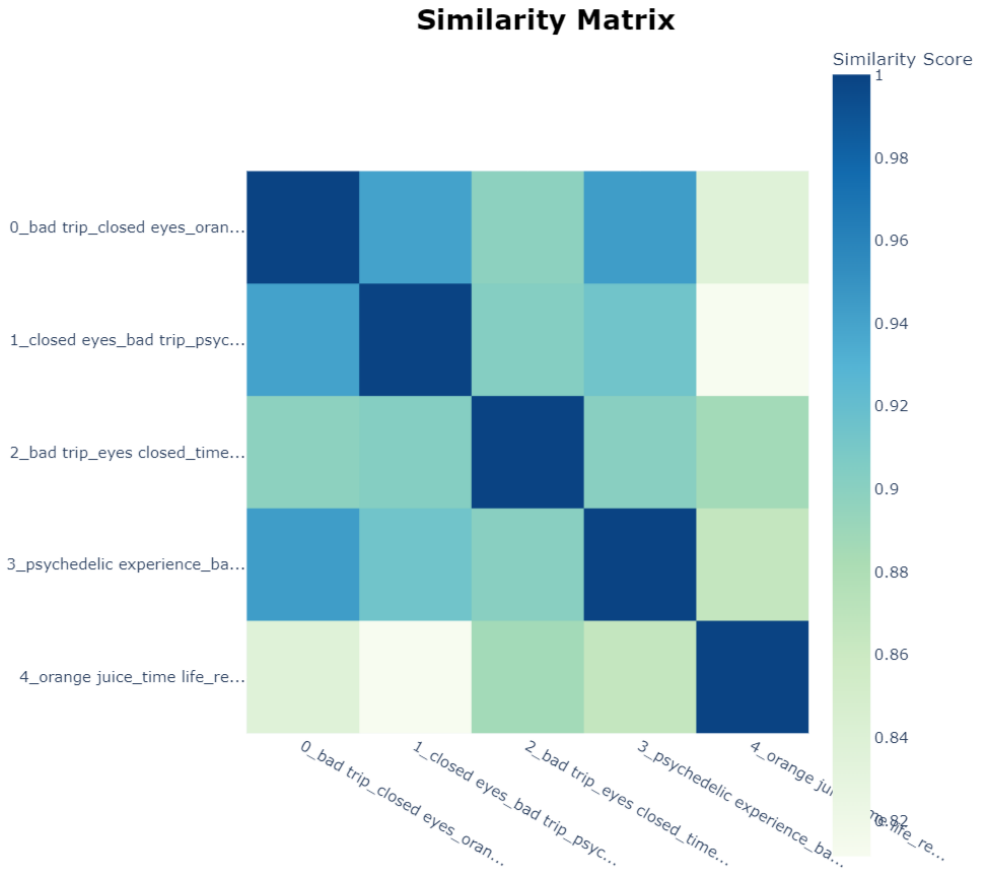
6.1 **Appendix A.** 2D document-topics plot & topic similarity matrix for LSD-group



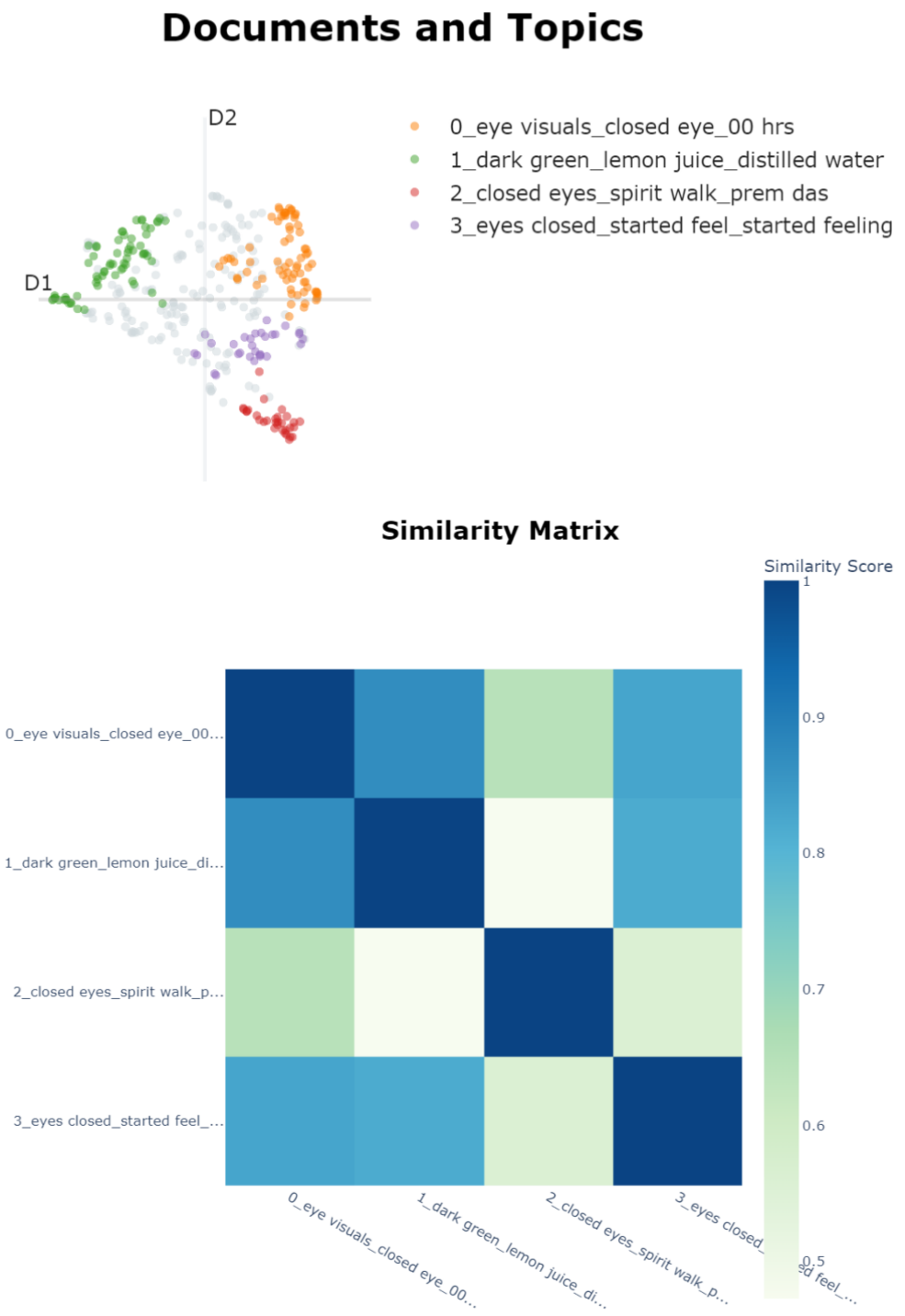


6.2 **Appendix B.** 2D document-topics plot & topic similarity matrix for psilocybin-group

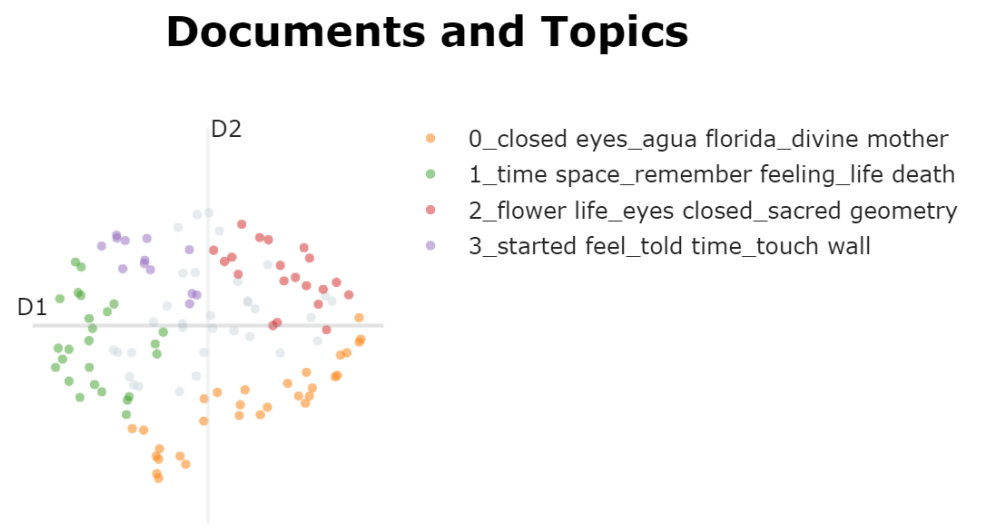


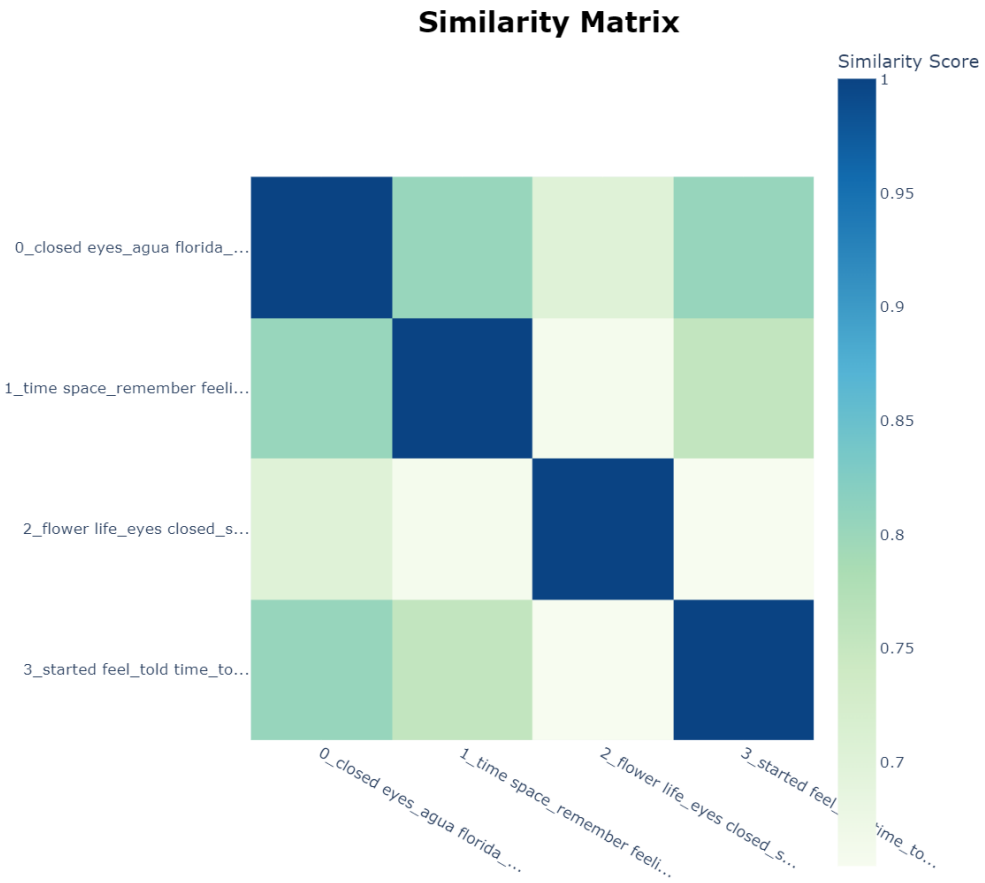


6.3 **Appendix C.** 2D document-topics plot & topic similarity matrix for mescaline-group

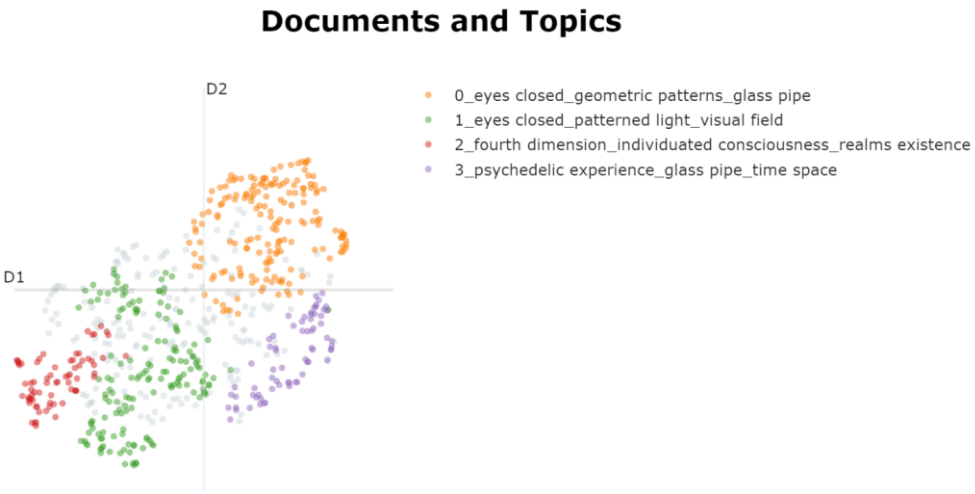


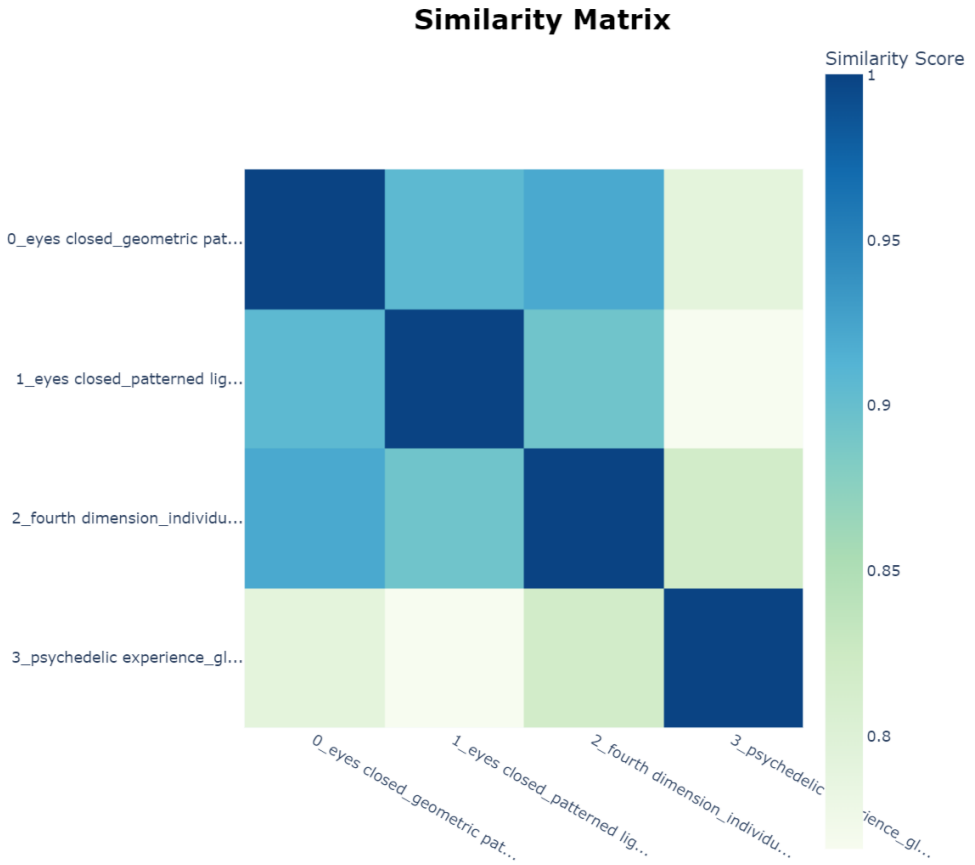
6.4 **Appendix D.** 2D document-topics plot & topic similarity matrix for Ayahuasca-group





6.5 **Appendix E.** 2D document-topics plot & topic similarity matrix for DMT-group





6.6 **Appendix F.** 2D document-topics plot & topic similarity matrix for 2C-B-group