

- Numerical labels are hard to read for Figure 2 both main panel and inset. Rewrite them with LaTeX?
- Add table to define abbreviations such as LSD, DMT, ...
- Axes for Figure 5 should all range from -1 to 1 .
- Include scree plot and silhouette score plot for Figure 5
- Make Table of Bonferonni corrected p-values for significant correlations in Figure ??

Methods

Acquisition of data We used Scrappy [1] to scrape text from all fora in Lycaeum in December 2015. The analyses in this paper reflect the content of Lycaeum at that time.

Preprocessing of data

Analysis of data To quantify how similar the drugs were in two documents we used the Jaccard similarity (citation). The Jaccard similarity between two objects is the ratio of the components those objects share in common to the number of unique elements in either object. It is the ratio of intersection of the sets of components of two objects to the union. We chose Jaccard similarity over cosine distance because cosine distance considers two objects with binary features similar if they both lack the same feature. It did not seem reasonable to consider to posts related if they both failed to mention one of 802 drugs.

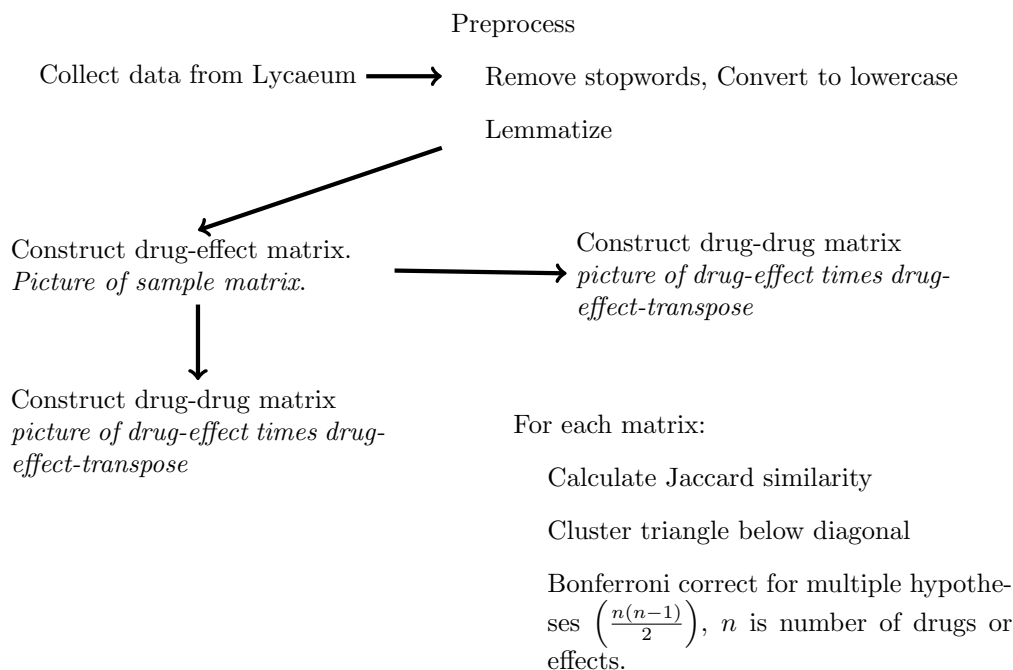


Figure 1: So much bob.

Results

Patterns of drug use Figure 2 shows the twenty most frequent substances mentioned in the Lycaeum corpus. The main panel of Figure 2 shows the top $20/161$ most mentioned substances across the Lycaeum corpus. We include “sound”, a common “coingestant” with LSD, when we write *substance*. It is not surprising that LSD is mentioned more often than caffeine or ethanol. Lycaeum aims to provide information about psychoactive substances.

Figure 3 shows the pattern of co-mentions of pairs of substances. In Figure 3 the color of the ij th square represents the Spearman correlation between number of documents mentioning drugs i and j in the corpus. To determine which correlations were significant we calculated the adjusted two-tailed p-value for each correlation coefficient for each combination of drugs. We adjusted each p-value by applying the Bonferroni correction, assuming that we were testing for 47 hypotheses (drug-drug pairs) simultaneously. Table X lists the combinations that occurred statistically significantly more frequently than would be expected by chance.

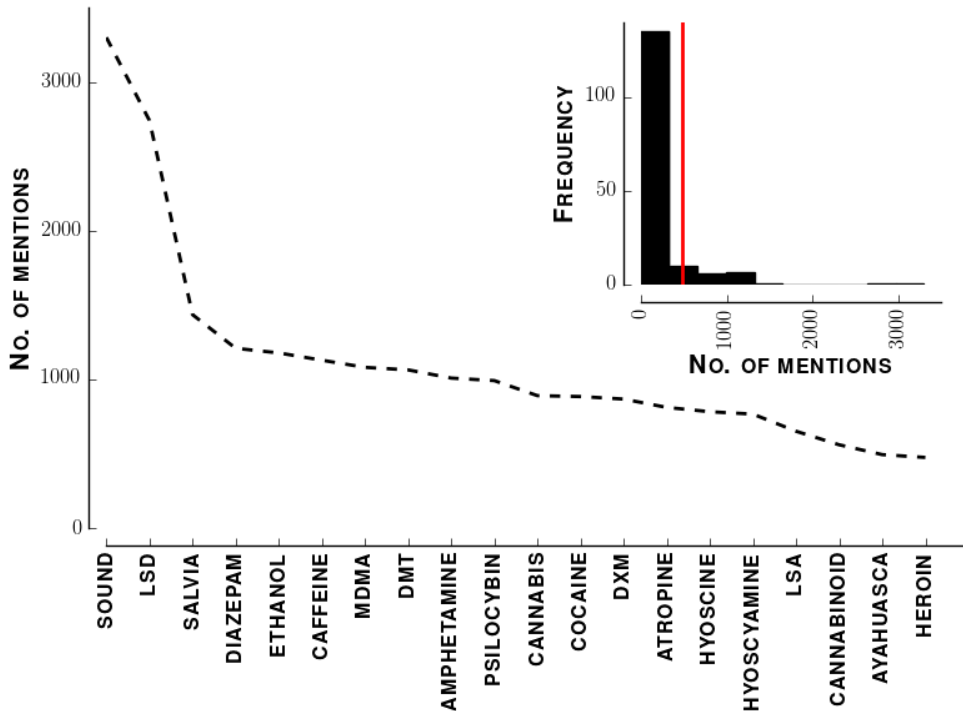


Figure 2: **Overall frequency of substance mentions.** 20 most prevalent substances are shown. Refer to Table X for abbreviations. **Inset:** Distribution of number of mentions of drugs across corpus. Main figure is derived from portion of histogram to the right of the vertical red line.

Patterns of effects Figure 4 shows the twenty most frequent effects mentioned in the Lycaeum corpus. Table 1 presents the definitions of some effects in Figure 4 that may be unfamiliar to the reader. Nearly all drugs mentioned in the Lycaeum corpus have multiple effects. For some, such as the opioids, anticholinergics, and stimulants, the dose determines which effect predominates. We did not attempt to infer dose from Lycaeum. There is no way to verify reported doses. One way to better understand dosage is to compare the doses mentioned across multiple websites similar to Lycaeum.

Term	Connotation
Entactogen	Feelings of belonging, familiarity, emotional openness; also called empathogen
Entheogen	Altered consciousness with religious, spiritual overtones
Hallucinogen	Perception of something without any sensory input
Psychedelic	Alters perception and cognition

Table 1: Terms used in x-axis of Figure 4

We hypothesized that users mix psychoactive substances to achieve distinct combinations of effects. We assumed that discussions on Lycaeum, at least in the aggregate, faithfully report that mixing. To test this hypothesis we investigated whether some pairs or groups of effects were mentioned together more frequently

than others.

To identify we calculated the frequency with which pairs of effects were mentioned

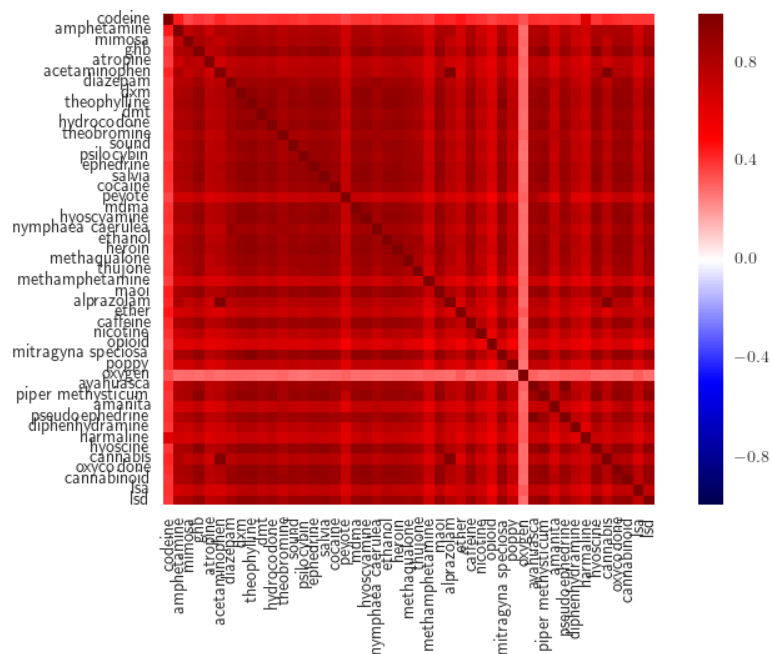


Figure 3: **Co-mentions of drugs.** Each square of the heat map shows the Spearman correlation between the substances indicates in the intersecting row and column. Warmer colors designate stronger positive correlation. Cooler, stronger negative correlation.

References

- [1] Daniel Myers and James W McGuffee. Choosing scrapy. *Journal of Computing Sciences in Colleges*, 31(1):83–89, 2015.

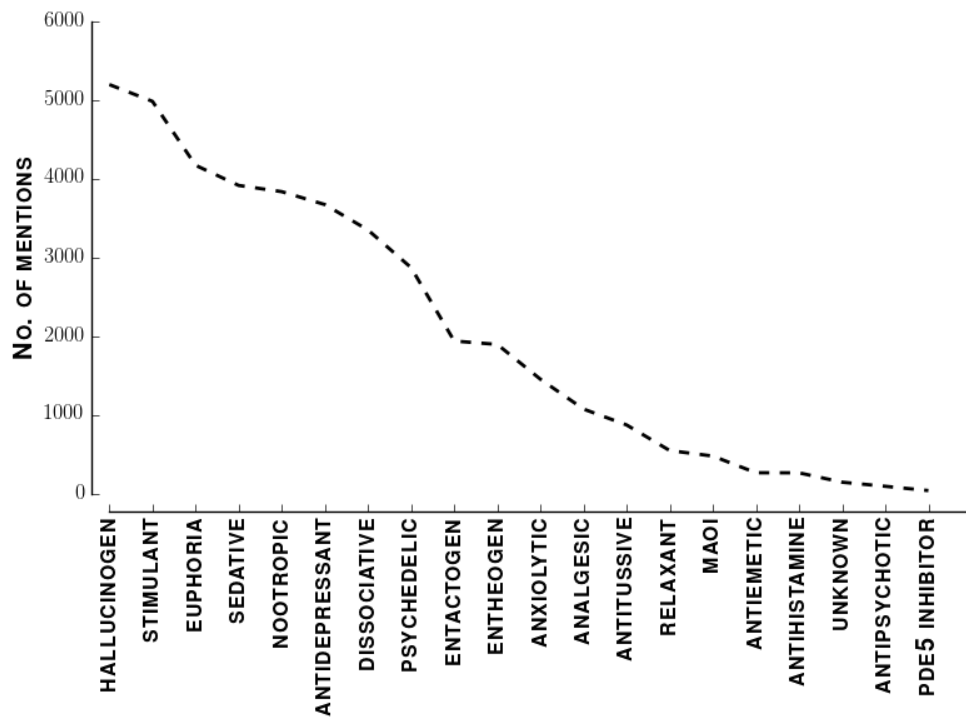


Figure 4: Overall frequency of substance effects.

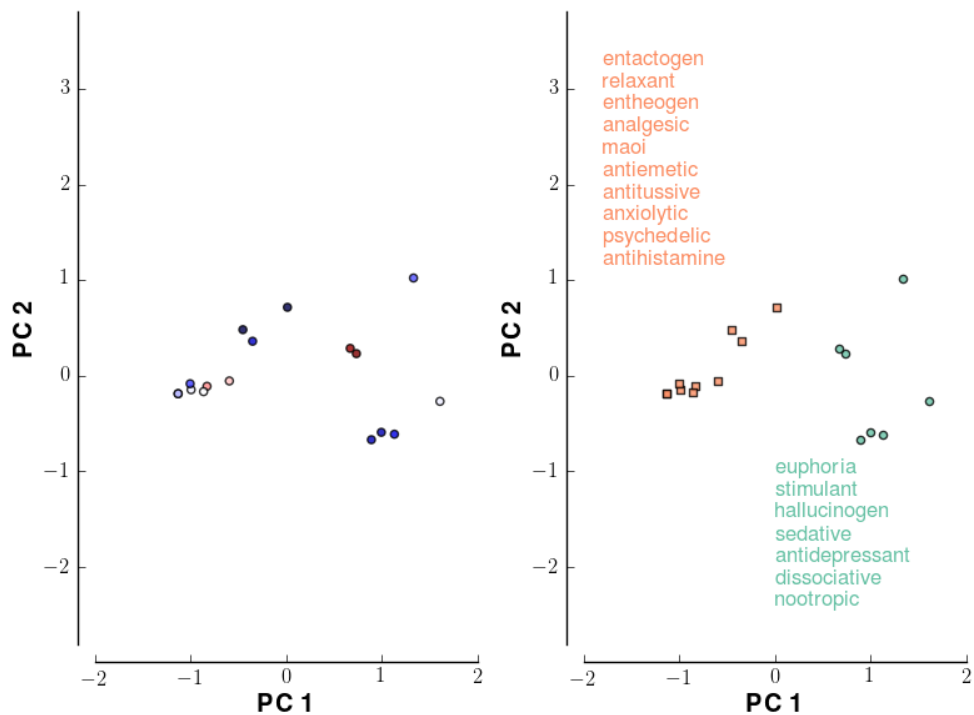


Figure 5: **Scatter plot of drug class combinations.** **Left:** Projection of drug effect combinations on first two principal components. **Right:**