

## SVD Visualization

### Overview

The goal of this visualization is help the user achieve a greater understanding of Singular Value Decomposition by viewing an application of the process and allowing the user to manipulate the view in order to see the data each each step of the calculation. In this visualization, Singular Value Decomposition is used for Latent Semantic Indexing to visualize the popular topics on your Facebook page.

### How it Works

The program contains a parser to construct a 2-Dimensional matrix from your Facebook page. It creates a row for each word and a column for each person. In the end, a matrix A is created with I words and j columns. A single cell  $A(i, j)$  represents the number of times word i was used by person j. Singular Value Decomposition (SVD) allows us to recreate the matrix A with fewer calculations. In this visualization, I use Cleve Moler's "Jama" Matrix Package (<http://math.nist.gov/javanumerics/jama/>) to run SVD over the matrix A. The process returns three matrices, U, S and  $V^T$ . SVD creates these values in such a way that the summation of their product (will call this matrix M) allows us to recreate the data with a relatively small number of calculations. The minimum number of summations required such that M contains an accurate representation of A is known as the "effective rank." If the dataset is large enough, the effective rank will be smaller than the j, the number of people. The particular usage of SVD in this application is similar to a process called Latent Semantic Indexing (LSI). For more information on LSI, visit: [http://en.wikipedia.org/wiki/Latent\\_semantic\\_indexing](http://en.wikipedia.org/wiki/Latent_semantic_indexing)

### How to Use It:

#### 1. Download the Application

Download the Zip file at: [www.artfordorks.com/SVDApp/Download.zip](http://www.artfordorks.com/SVDApp/Download.zip)

This file contains both the app and the native files it needs in order to run.

Be sure to keep all of the files in the same directory

#### 2. Get a Facebook Access URL

The access token will allow the application to parse your Facebook page. It expires after about an hour so you may need to get a new one every hour or so. To get an access token, do the following

- Get the access URL to your Facebook wall by visiting (and logging in if needed to):
- Visit: <http://developers.facebook.com/docs/reference/api/> (this will require that you login to your facebook account)
- Click the link that says "Profile feed (Wall)" and you will see a text version of your Facebook page.
- Copy the entire URL from that page and paste it somewhere where you can reference it again, it will look something like:  
*[https://graph.facebook.com/me/feed?access\\_token=2227470867|2.AQct-LoiaMCD0Xv7.3600.1309107600.0-3612050|rPQravZu-bvzRcZDKIAjVQR\\_YbQ](https://graph.facebook.com/me/feed?access_token=2227470867|2.AQct-LoiaMCD0Xv7.3600.1309107600.0-3612050|rPQravZu-bvzRcZDKIAjVQR_YbQ)*

#### 3. Run the Application

Enter the following command in the terminal, replace the bracketed information with your information.

```
java -Djava.library.path=[full path to lib/natives/] -jar [path to SVDApp.jar] "[url]" "[number of posts – something greater than 200 usually to be effective]"
```

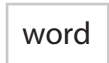

example entry:

```
java -Djava.library.path=/Users/lauradevendorf/Download/lib/natives/ -jar ./SVDApp.jar "https://graph.facebook.com/me/feed?access_token=2227470867|2.AQct-LoiaMCD0Xv7.3600.1309107600.0-3612050|rPQravZu-bvzRcZDKIAjVQR_YbQ" "200"
```

**Troubleshooting:**

1. Visualization runs out of memory: If this occurs, just reduce the number you enter in the command line.
2. Expired access token: repeat step 2 to obtain a new token and enter it into your command line.

## Symbol Mappings:

-  each word represents a word used on your wall
-  each circle represents a person who posted on your wall

word  
word  
word

the size of the font represents the frequency of this word in the original matrix, before SVD was applied. Small words are less frequent than larger words

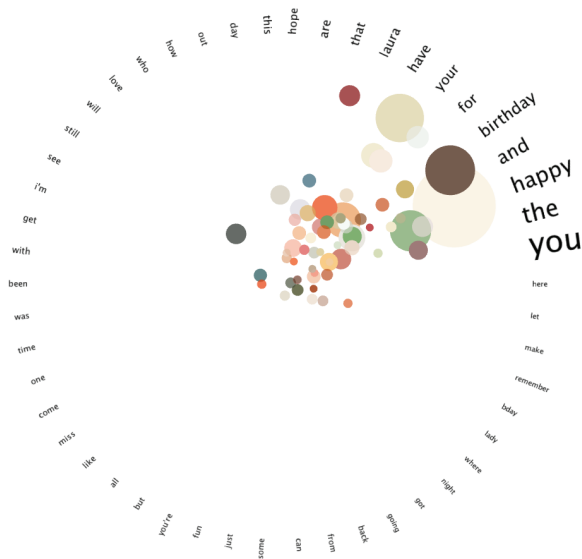


The size of the circle measures the activity of this person. This measure is calculated by total number of words used at the current rank

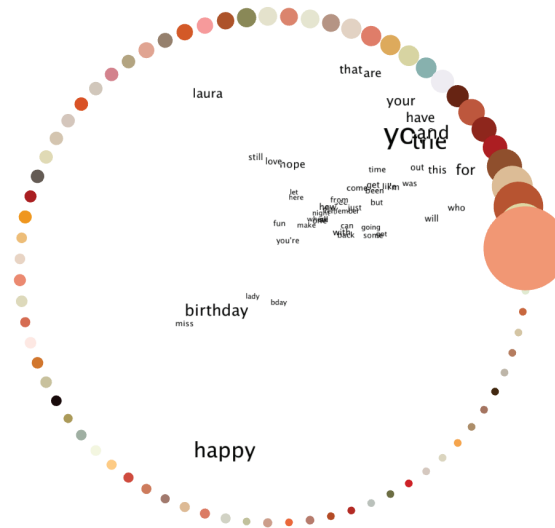
Each circle is arbitrarily assigned a color, one is assigned color per person

## Layout:

*view mode : words on ring, people in center*



*view mode : people on ring, words in center*



*An example on how to read the data:* In the chart on the left, you can see the the majority of people are attracted to the most popular words, naturally since there are the words they are using the most. In the data on the right, you can see that the word "happy" is far closer to the smaller "person" nodes. This tells us that people who aren't posting much are commonly using the word "happy" when they do. This is most likely because most people posted to me on my birthday and they're not people I interact with much at other times.

In both views, the position of the items in the center is determined by the item's attraction to other items in the ring. The strength of the attraction is determined by the frequency between that person and the word used, which is the value of  $M_{ij}$  at the given rank

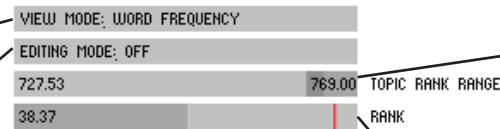
## Controls:

**View Mode Toggle:**  
Clicking this will toggle between view modes

**Editing Mode Toggle:**  
Allows to you toggle between editing and non-editing modes. In editing mode, clicking objects on the ring will remove them from the dataset. Use this to remove outliers from the data

**Sigma Plot:**  
This shows you the current rank's value of sigma as well as a history of sigma values by rank. The red line denotes the effective rank, where  $\sigma = 0$

**Current /Effective Rank Display:**  
This shows you the rank of information in the visualization as well as the effective rank.

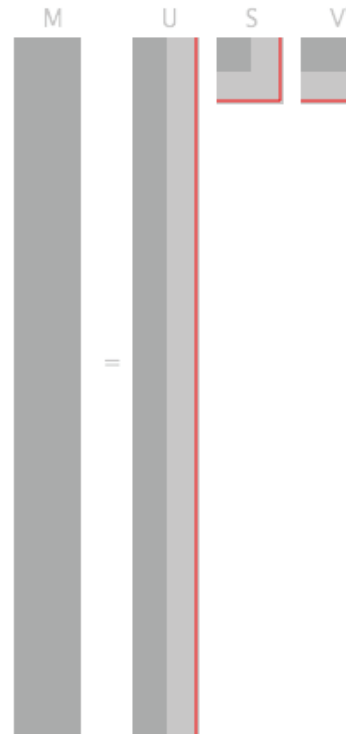


**Range Selection:**  
This control allows you to specify the range of words on the visualization. Words are sorted by popularity at the current rank and by default, the top 50 words are shown.



**Rank Slider:**  
This allows you to view the information at any rank. The red line denotes the effective rank. Changing the rank update the information on the screen.

Current Rank / Effective Rank  
**38 / 70**



**SVD Matrix Display:**  
This shows the rows and columns taken into account at each rank. The red line shows the effective rank. The Matrix  $M = U, S$  and  $V$  transpose. The lighter grey shows the entire matrix size and the darker grey shows the areas of the matrix that are taken into account for the current rank.

## Camera Controls:

**Zoom In / Out**  
Scroll up / down

**Reset Camera**  
Double Click

**Rotation**  
Click and drag in the direction you would like to rotate