# VISUAL ANALYSIS & MINING OF SOCIAL MEDIA – PART 2

Dave King

HICSS – 48

2015

# Agenda

- Part 1
  - Definitions
  - Example 1: Quantification of Self
- Part 2
  - Visual Framework(s)
  - Example 2: Text analysis of Rap lyrics
  - Example 3: Social Network Analysis

# VISUALIZATION FRAMEWORK

# Types of Data Visualization

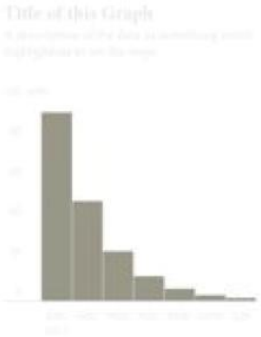
## *Statistical Graphics*

- The view from (Stephen) Few
  - Frequency distribution (e.g. histogram or bar chart)
  - Nominal comparison (e.g. bar chart)
  - Ranking (e.g. bar chart)
  - Part-to-whole (e.g. pie chart or bar chart)
  - Deviation (e.g. bar chart)
  - Time-series (e.g. line chart)
  - Correlation (e.g. scatter plot)
  - Geographic or geospatial (e.g. cartogram)

# Components of a data visualization

## Working parts

Several pieces work together to make a graph. Sometimes these are explicitly shown in the visualization and other times they form a visual in the background. They all depend on the data.

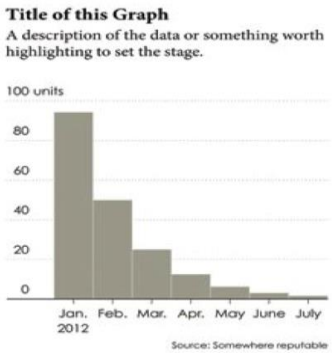


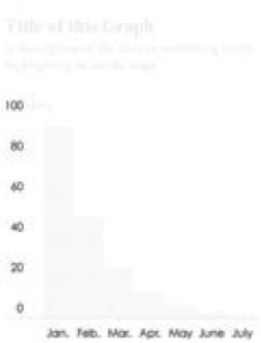
## Visual Cues

Visualization involves encoding data with shapes, colors, and sizes. Which cues you choose depends on your data and your goals.

## Coordinate System

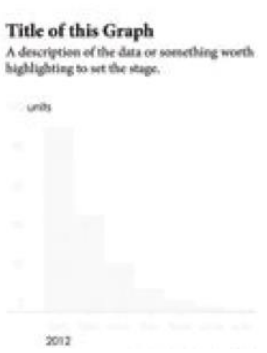
You map data differently with a scatterplot than you do with a pie chart. It's x- and y-coordinates in one and angles with the other; it's cartesian versus polar.



## Scale

Increments that make sense can increase readability, as well as shift focus.




## Context

If your audience is unfamiliar with the data, it's your job to clarify what values represent and explain how people should read your visualization.

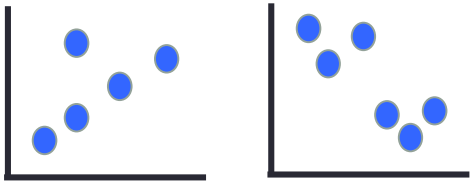
*Each visualization, regardless of where it is on the spectrum, is built on data and these five components.*

# Visual cues

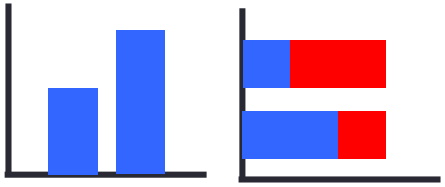
More Accurate

Less Accurate

Position

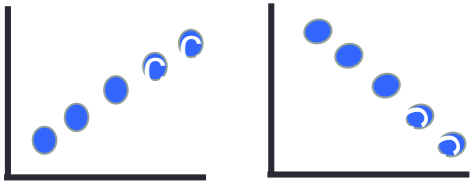


Length



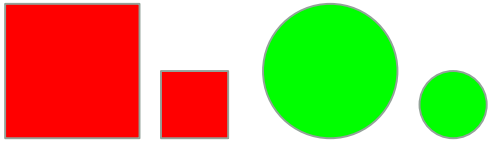
Angle

Direction

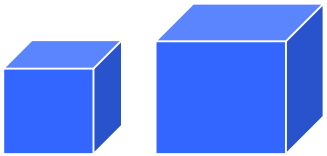


Shapes

Area



Volume



Color (Saturation)

Color (Hue)

*Cleveland & McGill*

# How many __’s are there?

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 9 | 2 | 5 | 8 | 5 | 0 | 4 | 1 | 5 | 9 | 0 | 1 | 0 | 0 |
| 2 | 2 | 3 | 6 | 1 | 0 | 6 | 0 | 6 | 7 | 0 | 7 | 6 | 9 | 5 |
| 9 | 4 | 8 | 4 | 6 | 1 | 4 | 2 | 9 | 2 | 8 | 6 | 4 | 6 | 8 |
| 0 | 9 | 1 | 3 | 1 | 1 | 7 | 6 | 7 | 5 | 2 | 5 | 9 | 4 | 1 |

**6’s**

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 9 | 2 | 5 | 8 | 5 | 0 | 4 | 1 | 5 | 9 | 0 | 1 | 0 | 0 |
| 2 | 2 | 3 | **6** | 1 | 0 | **6** | 0 | **6** | 7 | 0 | 7 | **6** | 9 | 5 |
| 9 | 4 | 8 | 4 | **6** | 1 | 4 | 2 | 9 | 2 | 8 | **6** | 4 | **6** | 8 |
| 0 | 9 | 1 | 3 | 1 | 1 | 7 | **6** | 7 | 5 | 2 | 5 | 9 | 4 | 1 |

**2’s**

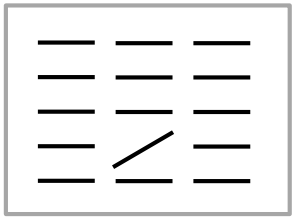
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 9 | 2 | 5 | 8 | 5 | 0 | 4 | 1 | 5 | 9 | 0 | 1 | 0 | 0 |
| 2 | 2 | 3 | 6 | 1 | 0 | 6 | 0 | 6 | 7 | 0 | 7 | 6 | 9 | 5 |
| 9 | 4 | 8 | 4 | 6 | 1 | 4 | 2 | 9 | 2 | 8 | 6 | 4 | 6 | 8 |
| 0 | 9 | 1 | 3 | 1 | 1 | 7 | 6 | 7 | 5 | 2 | 5 | 9 | 4 | 1 |

**1’s**

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 9 | 2 | 5 | 8 | 5 | 0 | 4 | 1 | 5 | 9 | 0 | 1 | 0 | 0 |
| 2 | 2 | 3 | 6 | 1 | 0 | 6 | 0 | 6 | 7 | 0 | 7 | 6 | 9 | 5 |
| 9 | 4 | 8 | 4 | 6 | 1 | 4 | 2 | 9 | 2 | 8 | 6 | 4 | 6 | 8 |
| 0 | 9 | 1 | 3 | 1 | 1 | 7 | 6 | 7 | 5 | 2 | 5 | 9 | 4 | 1 |

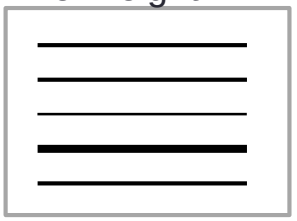
# Detection and Recognition

Line Orientation

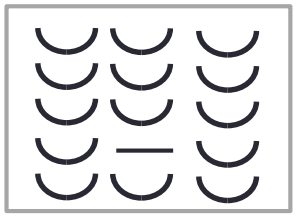


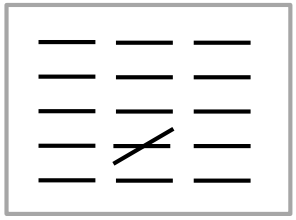
Line Length

Line Weight

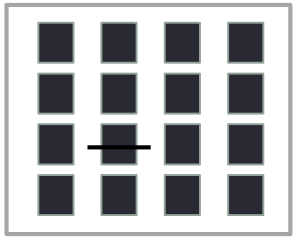


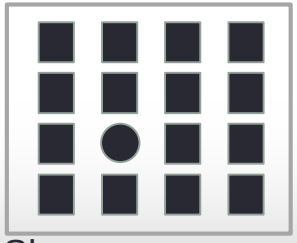
Curvature



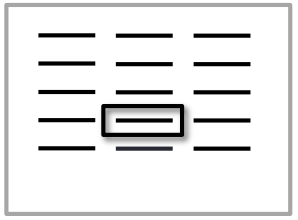


Added Markers

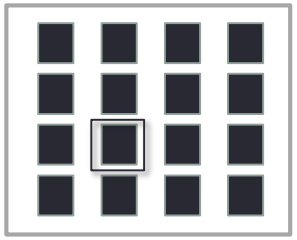




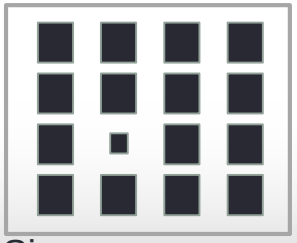
Shape

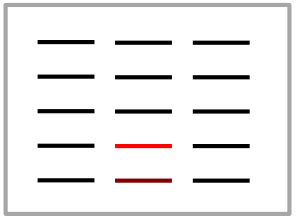


Enclosure

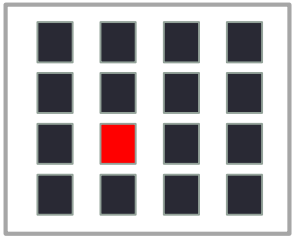


Size

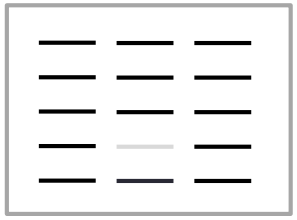




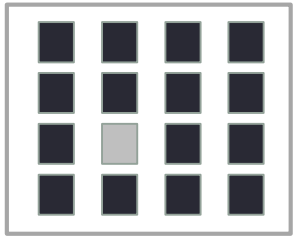
Color/Hue



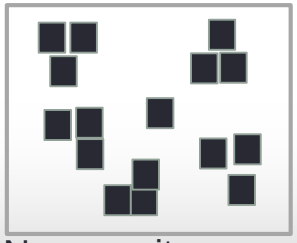
Sharpness



Intensity/Value



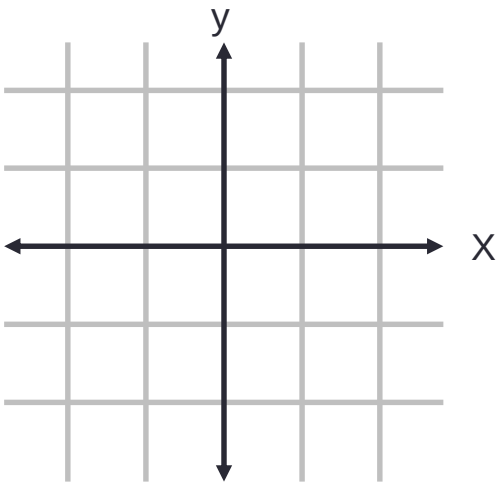
Numerosity



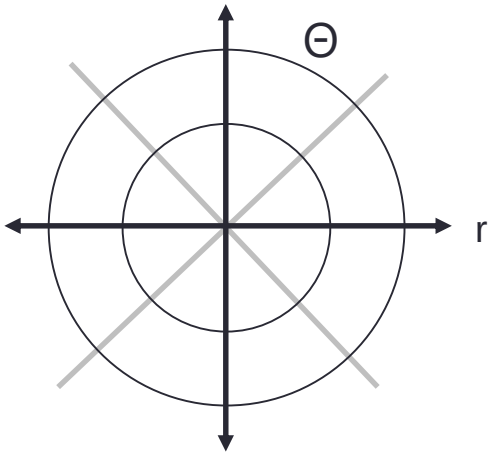
# Coordinate Systems

*Most common*

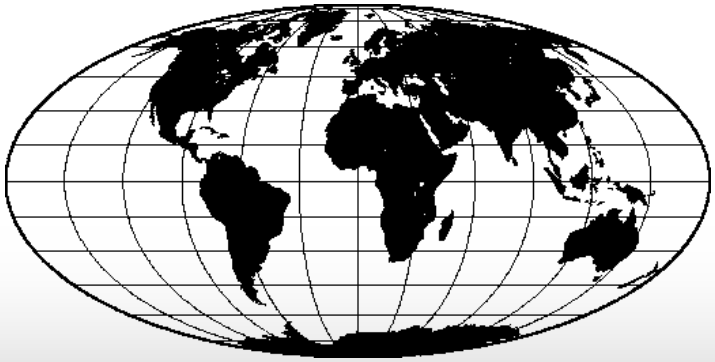
Cartesian




Polar



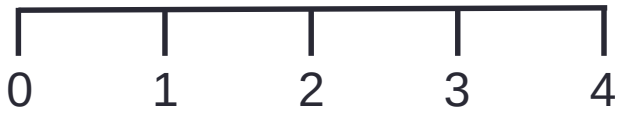

Geographic Projection



# Scale

*Most common*

Linear (even space)




Logarithmic (% change)


Categorical (discrete)

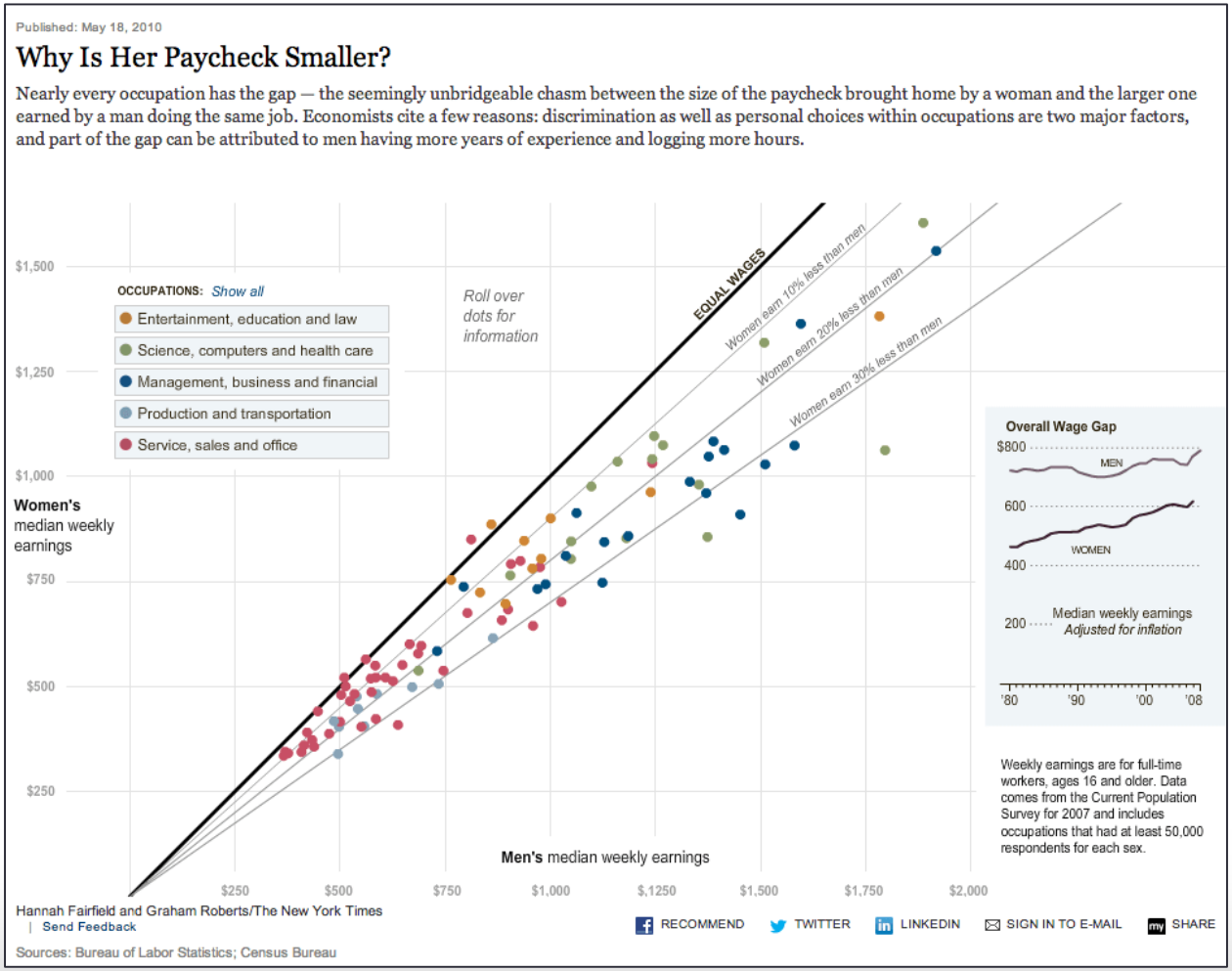

Ordinal (ranked categories)


Percent (of whole)


Time (hours, days, mnths, years …)

# The role of context

*1 word is worth …*



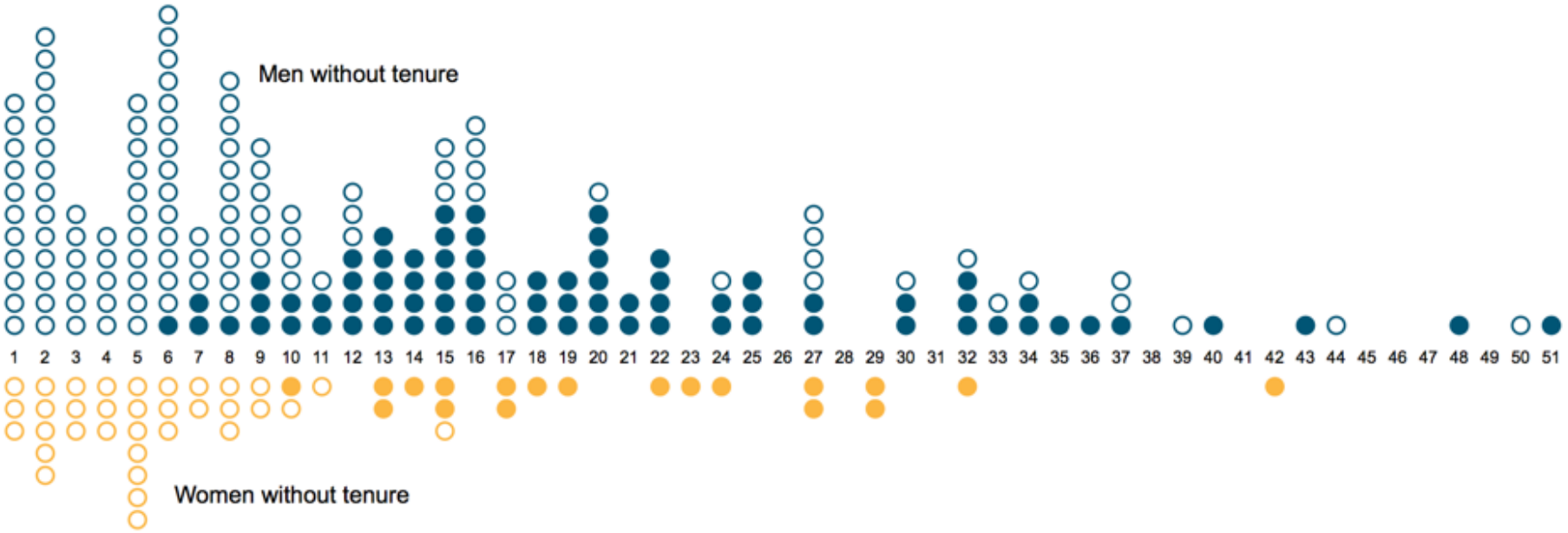
# Context

*1 word is worth …*

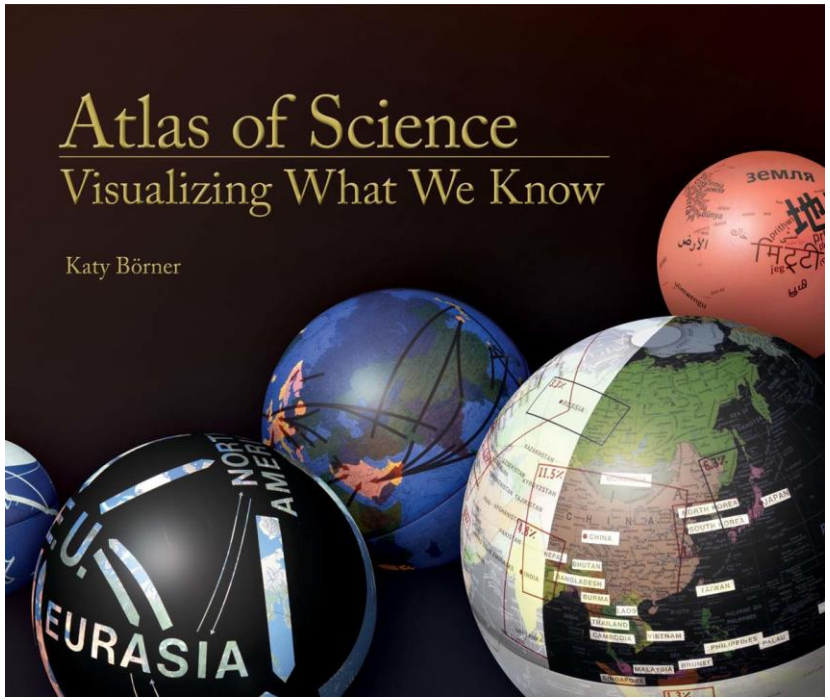
## The Tenure Pipeline at Harvard Business School

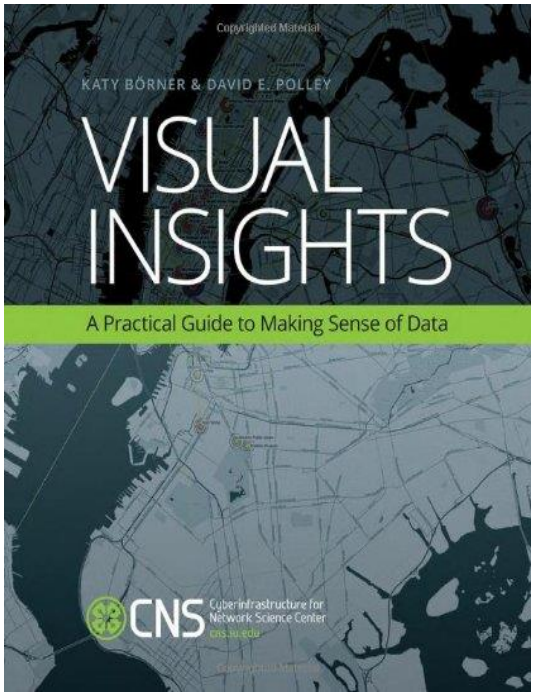
1 2 3 4 **5** NEXT >

But **the pipeline for women is small**. Fewer than a third of untenured faculty members are women, making it unlikely that the gender imbalance in tenured faculty will shift in the near future.



# Visualization Framework & Workflow

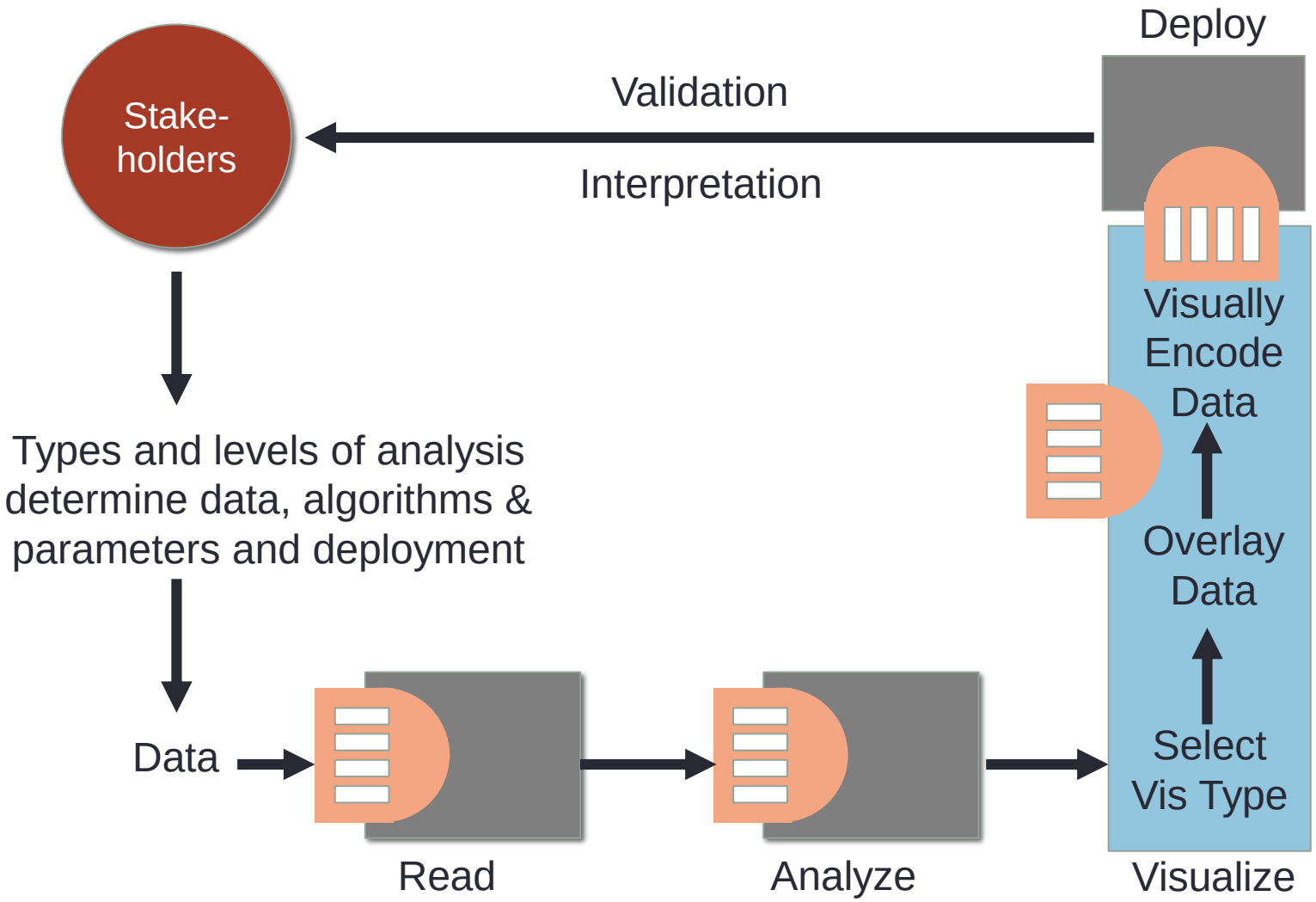



cns.iu.edu/home.html

ella.slis.indiana.edu/~katy/

# Visualization Framework & Workflow

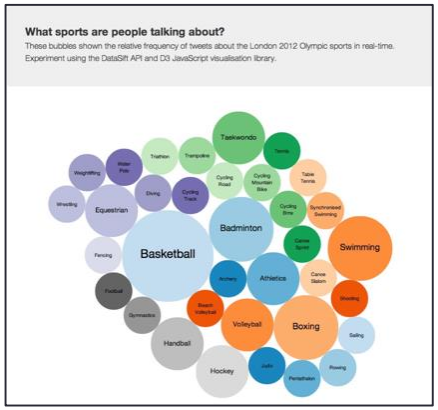


# Analysis: Type vs. Level

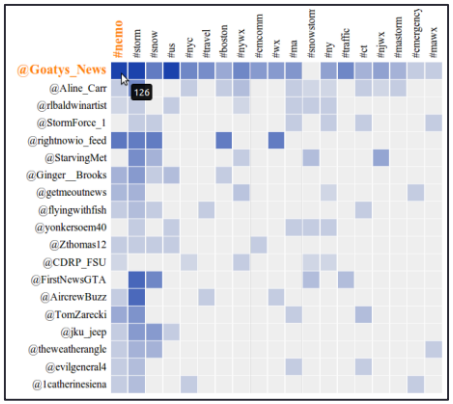
| Types of Analysis | Levels of Analysis | | |
|---|---|---|---|
| | Micro/Individual 1-100 records | Meso/Local 101-10,000 records | Macro/Global 10,000+ records |
| Statistical Analysis/Profiling | | | |
| Temporal Analysis (When) | | | |
| Geospatial Analysis (Where) | | | |
| Topical Analysis (What) | | | |
| Network Analysis (with Whom) | | | |

# Visualization Types

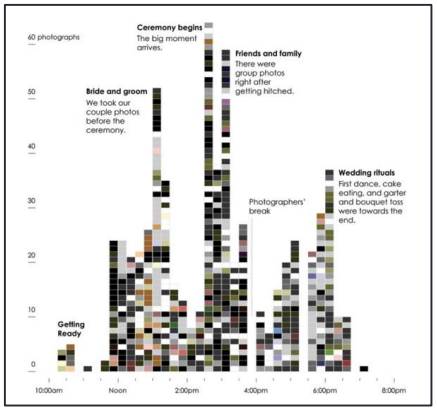
## Charts



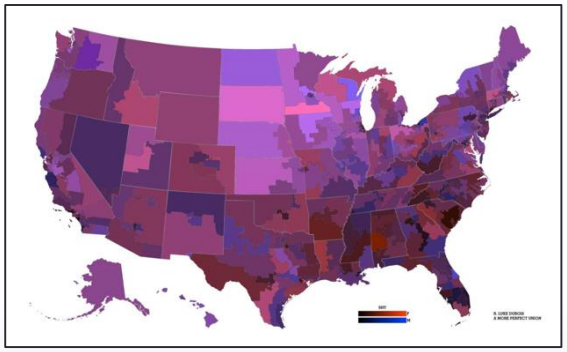

## Tables



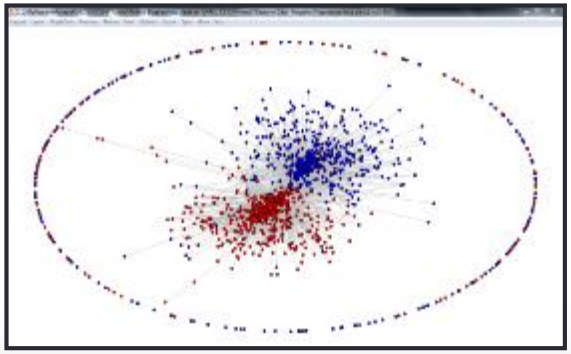

## Graphs




## Geospatial Maps



## Network Graphs

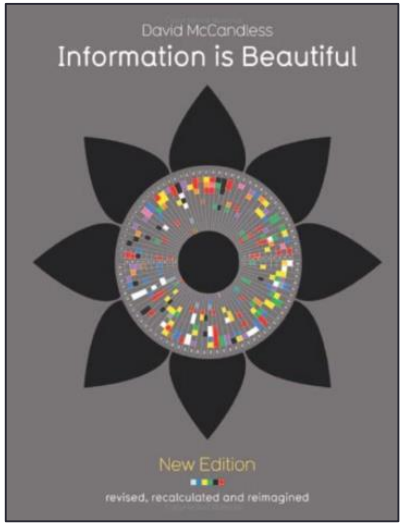


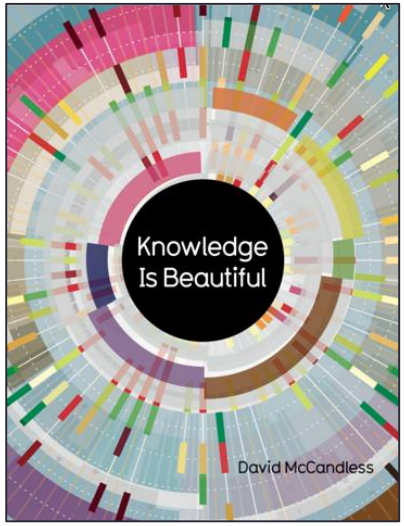
# EXAMPLE: RAP LYRICS

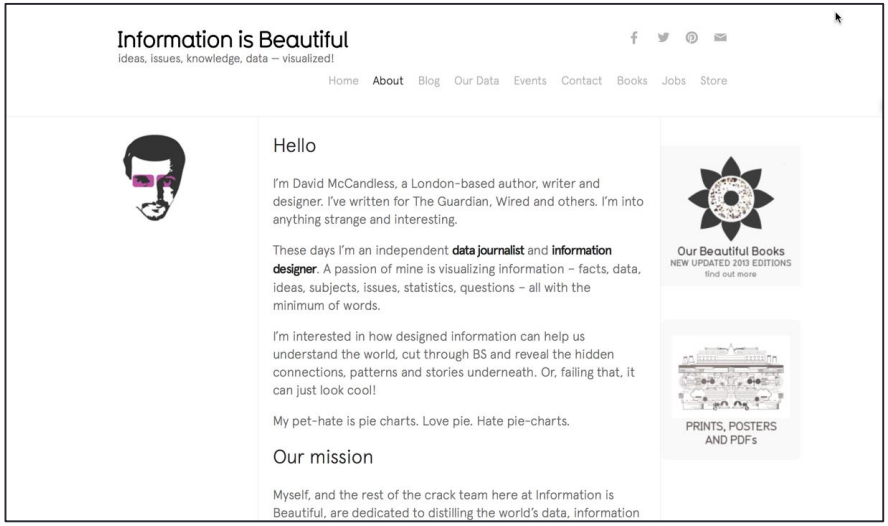
# Example: Textual Analysis of Rap Lyrics

# Example

*Analyzing the lyrics of Rap/Hip Hop songs*



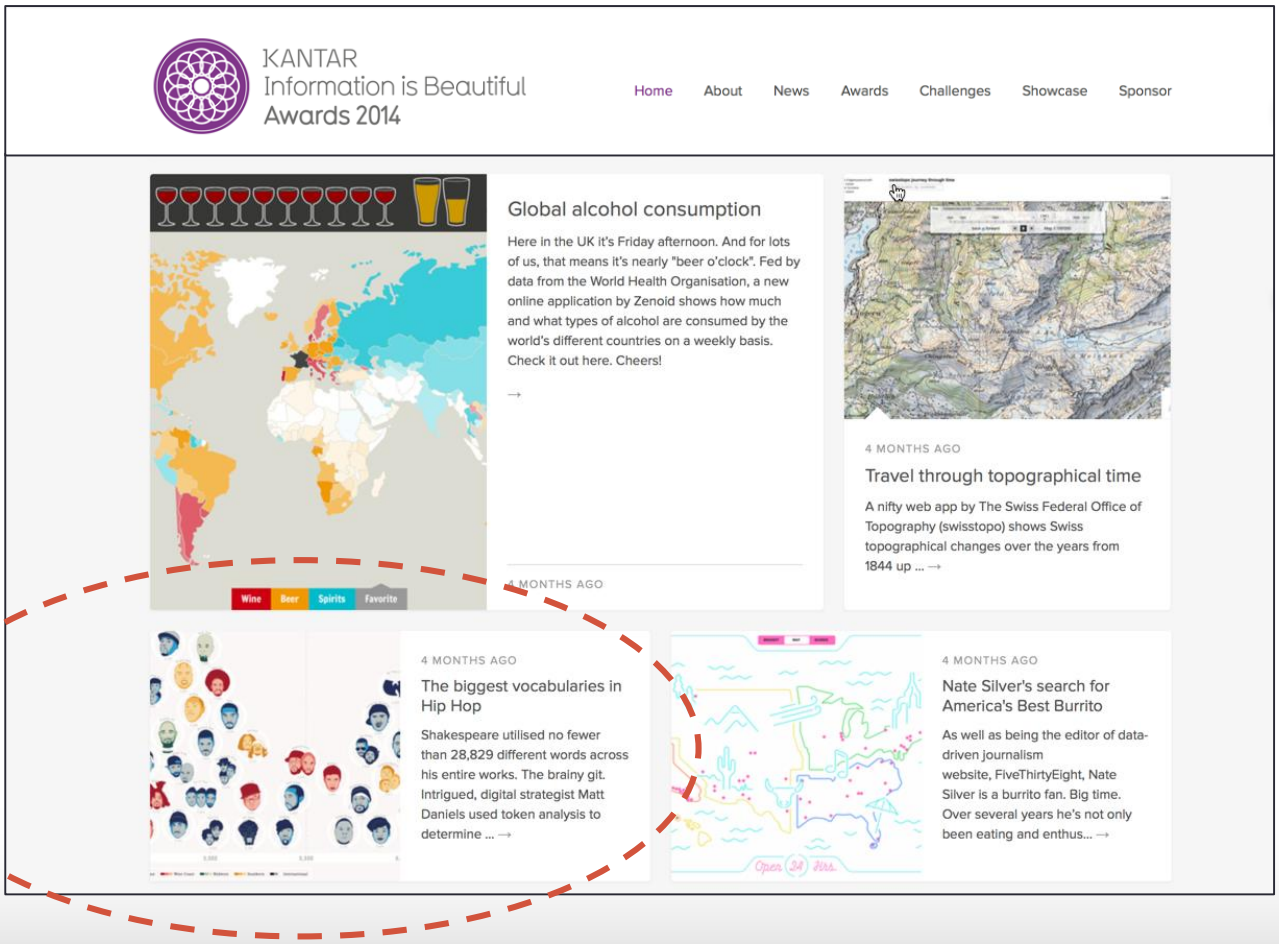




David McCandless – Informationisbeautiful.net

# Example

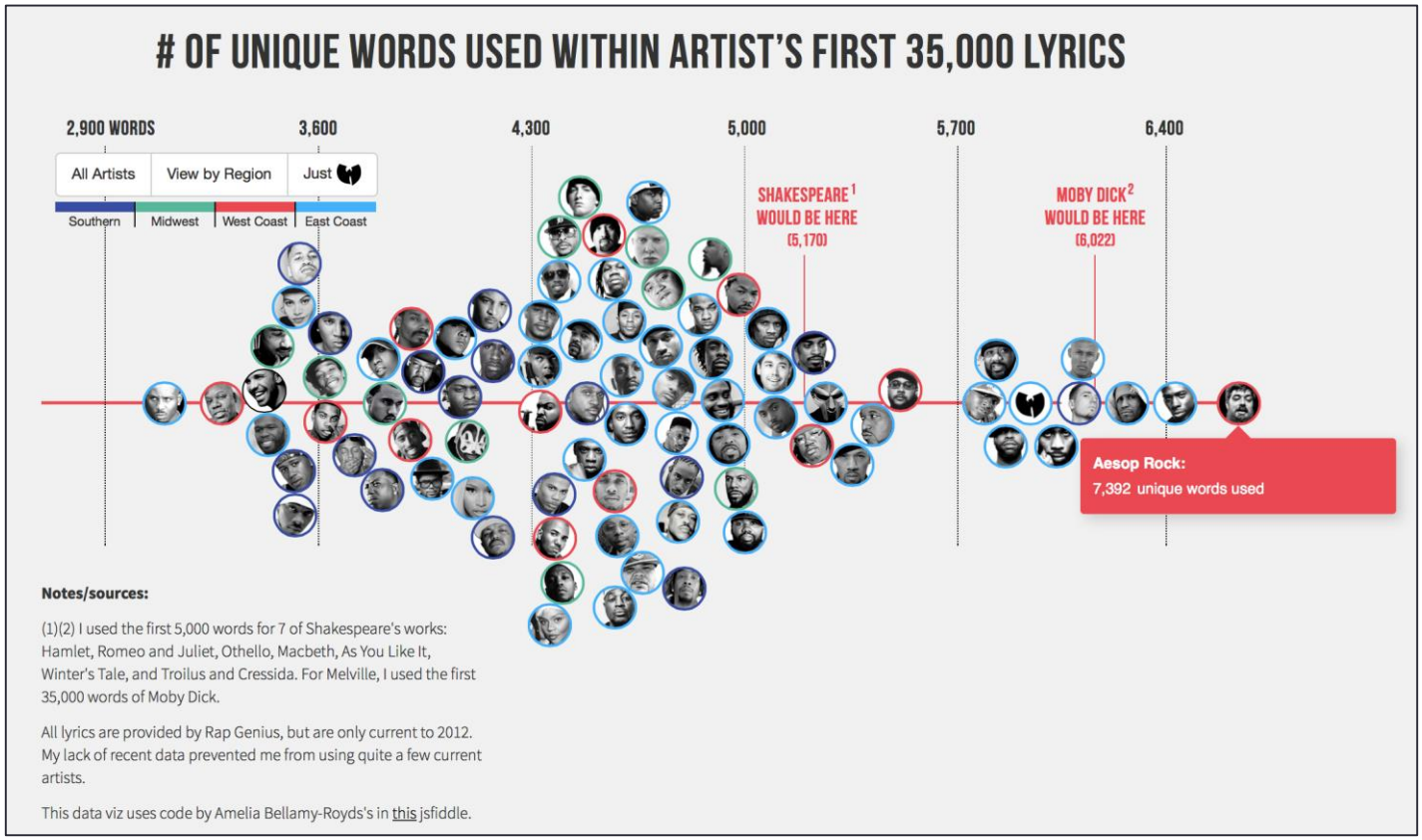
*Analyzing the lyrics of Rap/Hip Hop songs*




Informationisbeautifulawards.net

# Example: Lyrics of Rap Songs

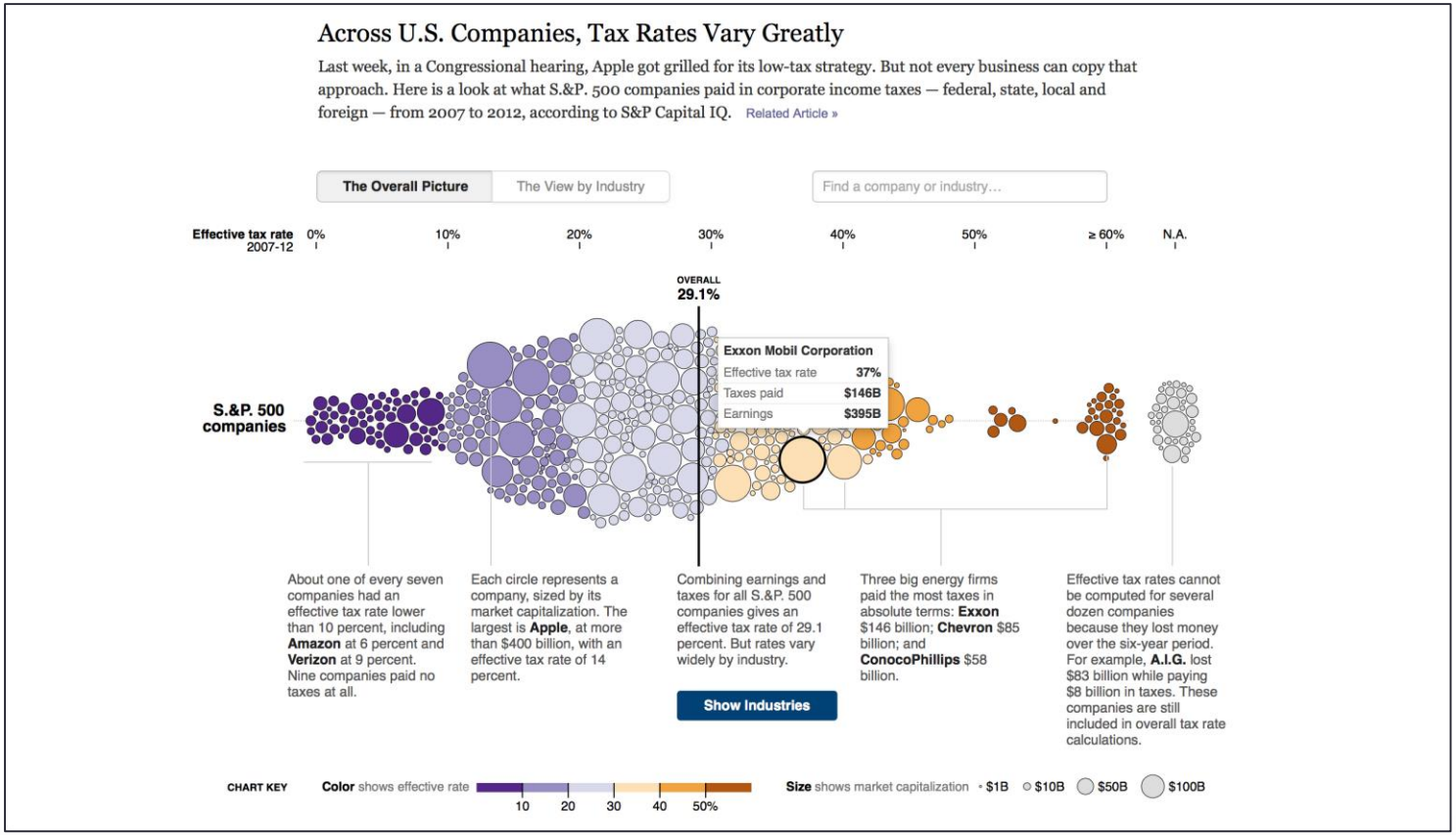
## *Biggest Vocabularies and Unique Words*




Research by: Matt Daniels, mdaniels.com
Visualization by: Amelia Bellamy-Royds, fiddle.jshell.net/6cW9u/8/

# Example: Lyrics of Rap Songs

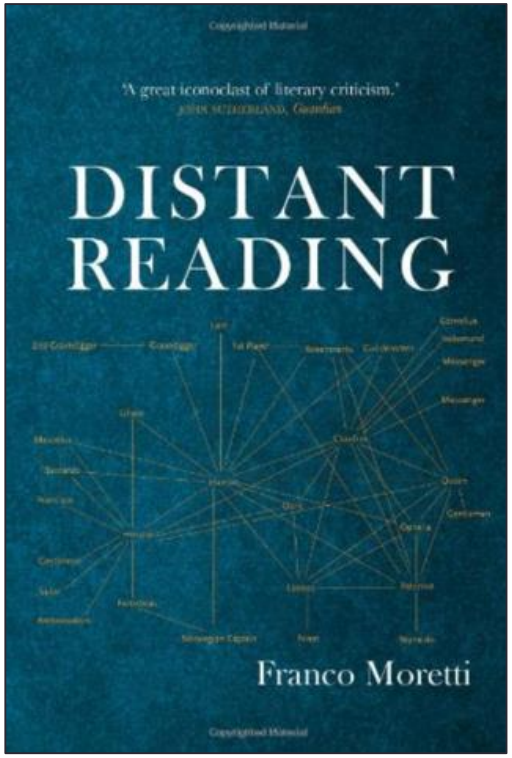
*Visualization inspired by…*

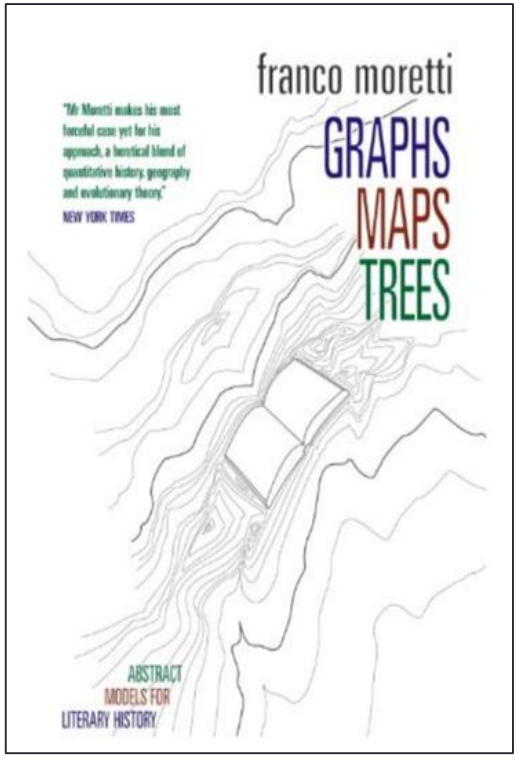



Inspired by: nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html?_r=2&

# Example: Lyrics of Rap Songs

*In the same vein as …*

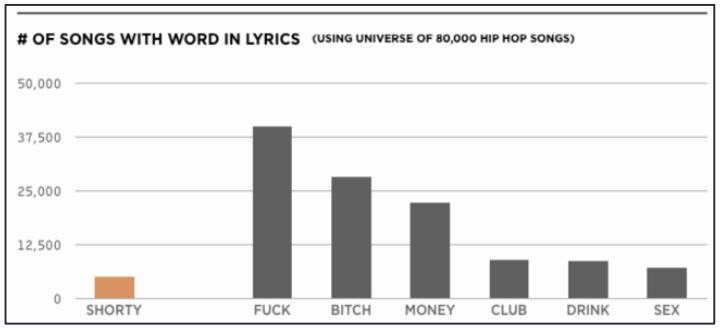


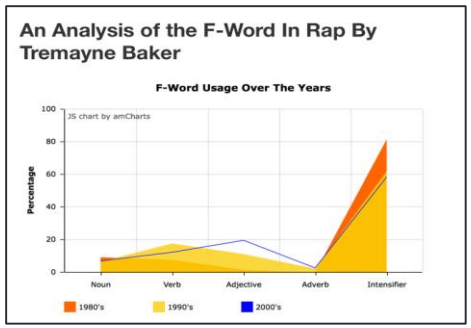




Literature scholars should stop reading books and start counting, graphing, and mapping them instead.

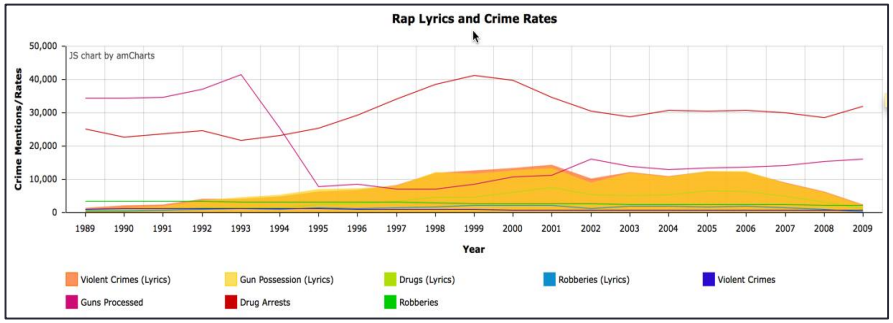
# Example: Lyrics of Rap Songs

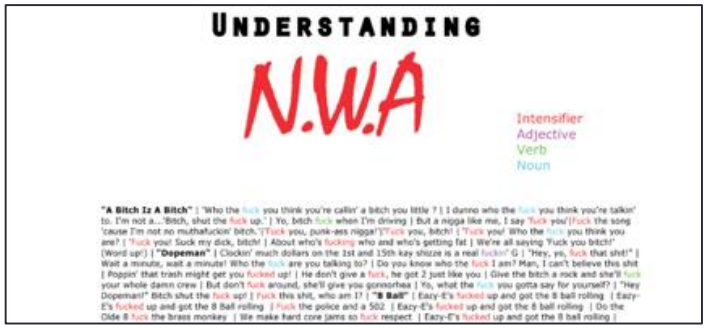
*Types of analyses – word or topic frequency*

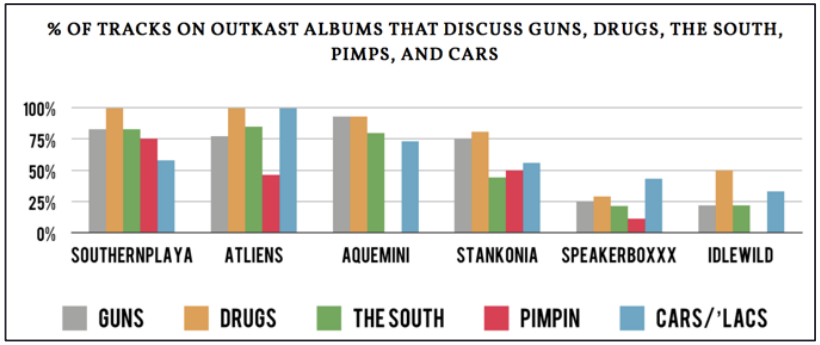


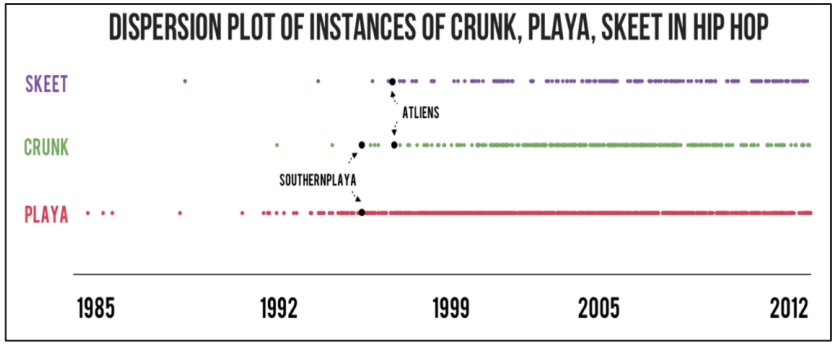







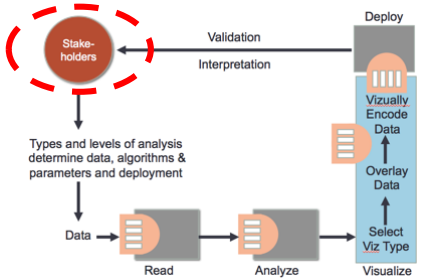



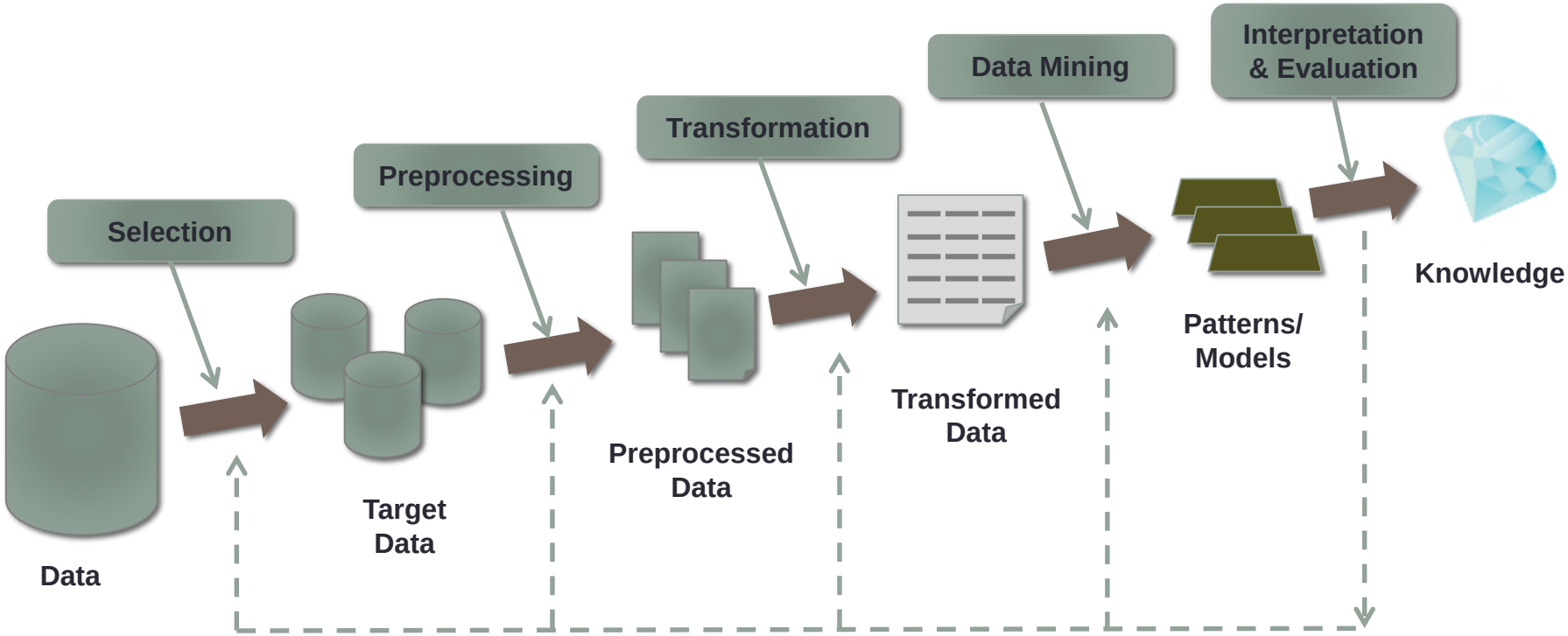
# Example: Lyrics of Rap Songs

*Textual analysis – the interest*



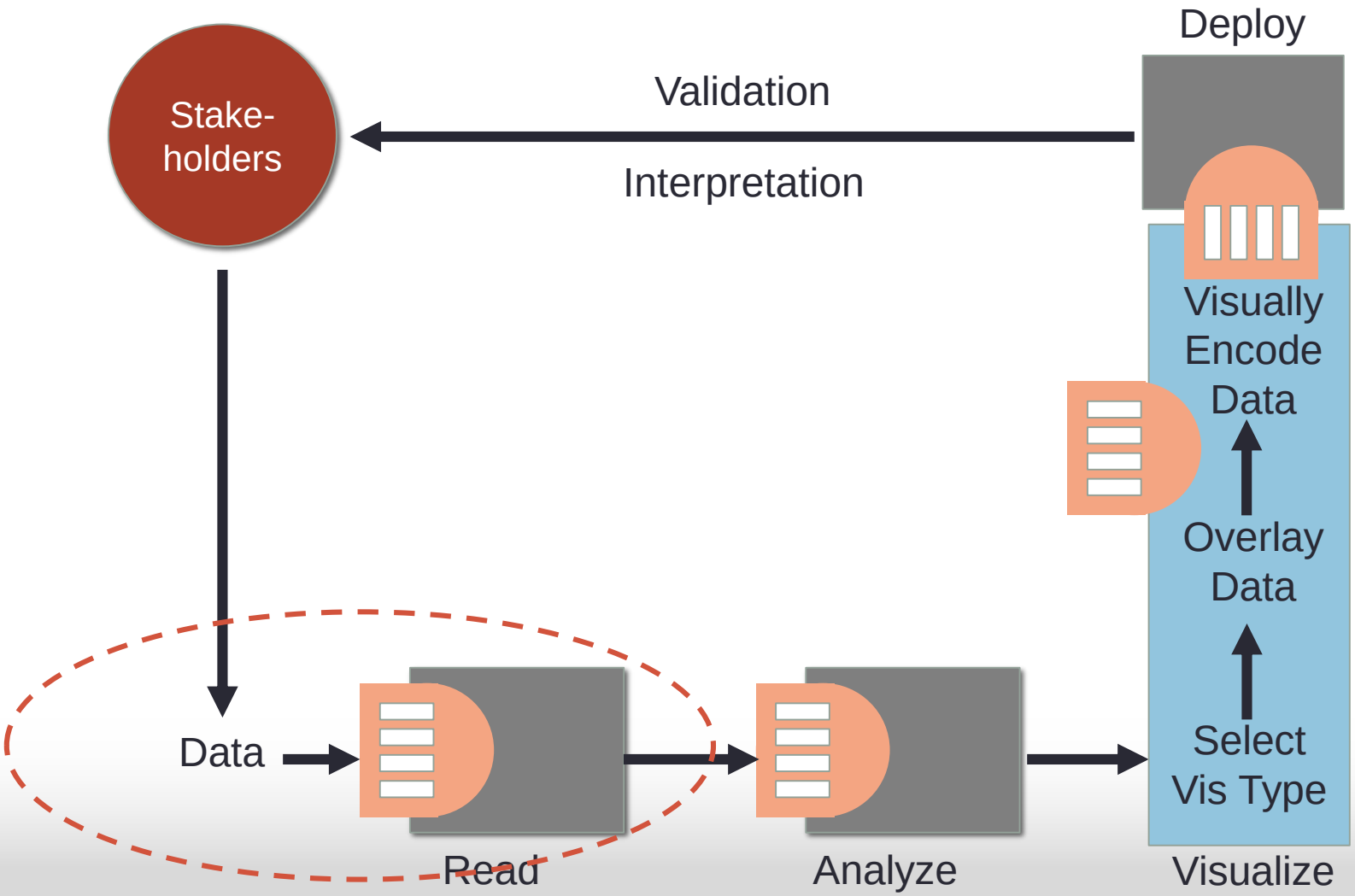

kickstarter.com/projects/1801076626/the-hip-hop-word-count-a-searchable-rap-almanac

# Example: Lyrics of Rap Songs
## *CRISP Data Mining Process*



# Example: Lyrics of Rap Songs

*Difficult part is turning text into analyzable data*



# Example: Lyrics of Rap Songs

## *Moving from lyrics to…*

Everybody just rock it, don't stop it
You gotta rock it, don't stop
Keep tickin' and tockin', work it all around the clock
Everybody keep rockin' and clockin' and shockin' and rockin', go house

Da leaders, lookin straight charming in our Giorgio Armani's
You wanna harm me and Nas you gots ta conquer through a whole army
The cee-lo rollers, money folders, sippin' Bolla, holdin mad payola
Slangin that Coke without the Cola
Me and black don't fake jacks but we might sling one
It ain't no shame in our game we do our thing son

*[Chorus – Eminem (repeat 2x):]*
'Cause I'm Slim Shady, yes I'm the real Shady
All you other Slim Shadys are just imitating
So won't the real Slim Shady please stand up,
Please stand up, please stand up?

They were jammin off a record that said it best:
"Now what you hear is not a test!"

Aowowowowowowowowowowowo!
A-hunga-hunga-hunga-hunga
Aowowowowowowowowowowowo!
Aowowowowowowowowowowowo!

Form(at) amenable to mathematical/statistical analysis and visual display

# Example: Lyrics of Rap Songs

## *Issues with rap/hip hop lyrics*


- No specified format
- Variable length
- Variable spelling
- Punctuation and non-alphanumeric characters
- No predefined content or predefined set of values
- Slang, made up and misspelled words.
- Repetitive content(e.g. choruses)
- Some sources embed annotation (who is singing, how often to repeat a phrase, …)

# Example: Lyrics of Rap Songs

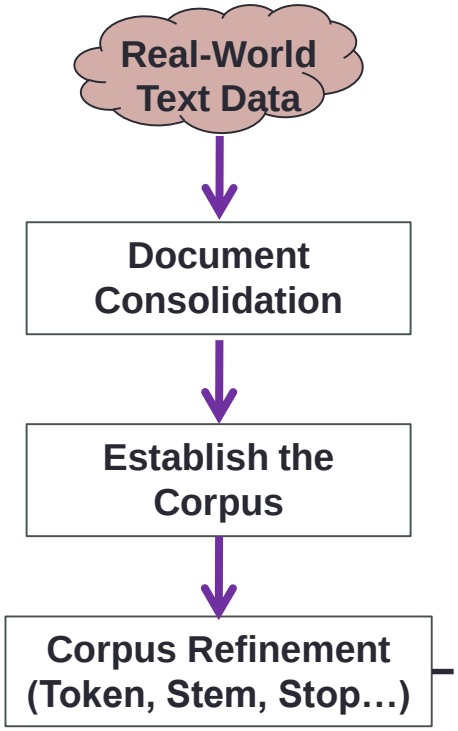
## *General approach*

The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods.

It is often assumed that documents are a *bag of words*, where order does not inform our analyses… If this assumption is unpalatable, we can retain some word order by including bigrams (word pairs) or trigrams.

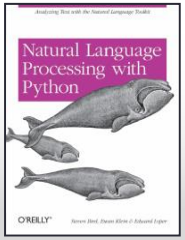
# Example: Lyrics of Rap Songs

## *Text conversion process*

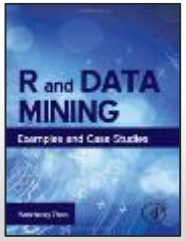


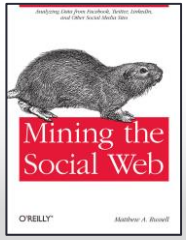


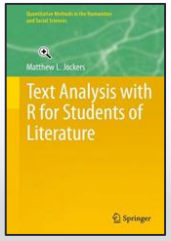
- **Tokenization** —Parse the text to generate terms. Sophisticated analyzers can also extract phrases from the text.
- **Normalize** — Convert them to lowercase.
- **Eliminate stop words** — Eliminate terms that appear very often (e.g. the, and, …).
- **Stemming** — Convert the terms into their stemmed form— remove plurals and different word forms (e.g. achieve, achieves, achieved – achiev) [note: word about synonyms – WordNet Synset]





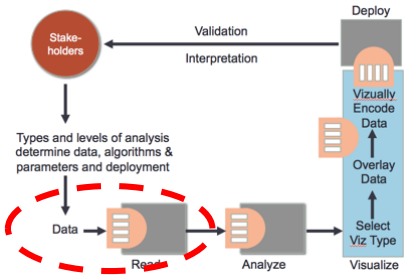


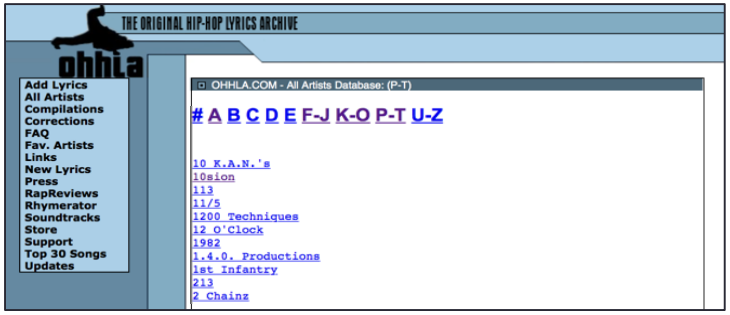



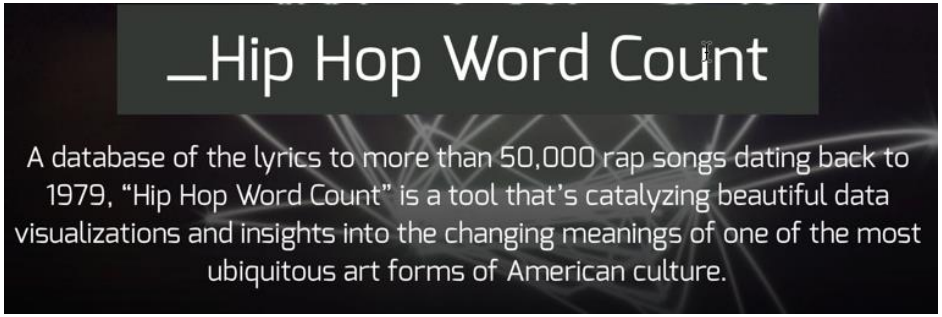




# Example: Lyrics of Rap Songs

## *Text conversion process - Corpus*

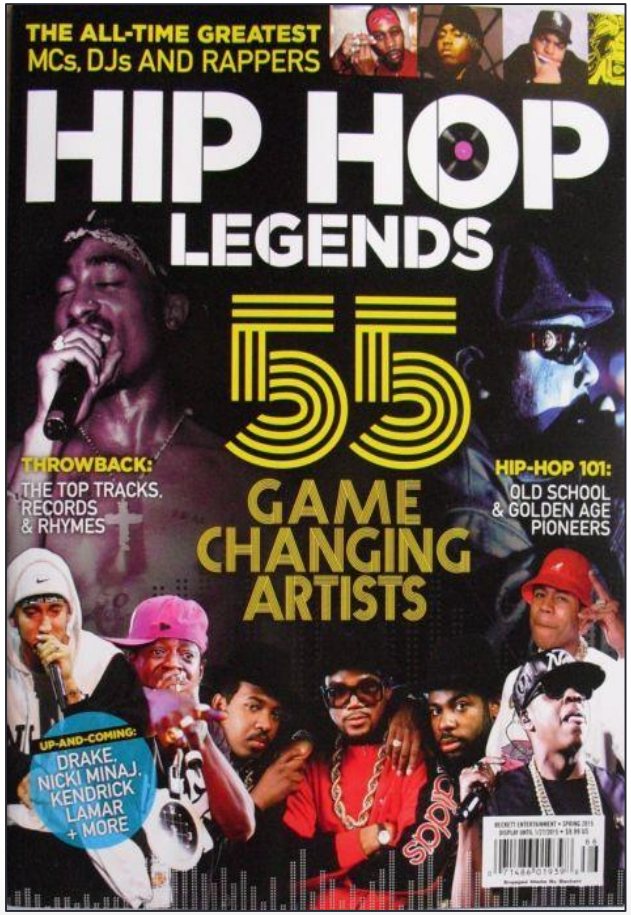


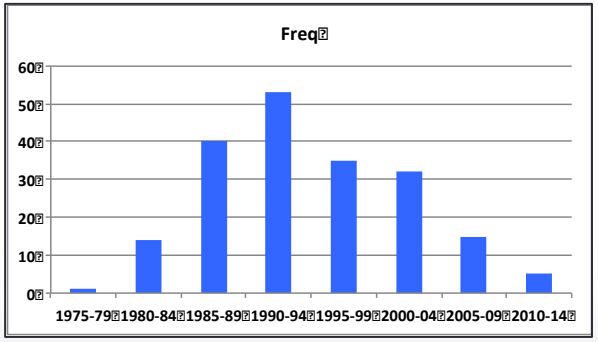

# Example: Lyrics of Rap Songs

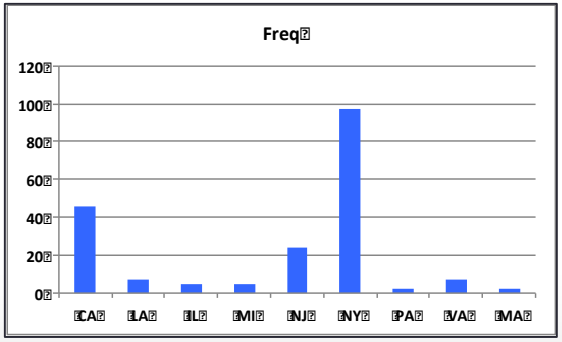
## *Text conversion process – Hip Hop corpus*




| 195 Hip Hot Artists (Individuals & Groups) | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Old School** | **15** | **Golden Age** | **86** | **Gangsta Rap** | **62** | **Millennium** | **32** |
| Afrika Bambaataa | 3 | 2 Live Crew | 6 | 2PAC | 5 | Eminem | 5 |
| Grandmaster Flash & The Furious Five | 1 | ATCQ | 4 | Dr. DRE | 6 | Jay-Z | 8 |
| Kook G Rap | 2 | Beastie Boys | 6 | Ice Cube | 8 | Kanye West | 5 |
| Kool Moe Dee & Treacherous Three | 2 | Big Daddy Kane | 2 | Ice-T | 6 | Lil Wayne | 7 |
| Kurtis Blow | 4 | Biz Markie | 2 | Lil' Kim | 5 | Missy Elliott | 7 |
| Sugar Hill Gang | 3 | Busta Rhymes | 7 | N.W.A | 6 | | |
| | | De La Soul | 4 | NAS | 5 | | |
| | | DJ Jazzy & The Fresh Prince | 2 | Puff Daddy, P. Diddy | 6 | | |
| | | Eric B. & Rakim | 7 | Snoop Dogg | 5 | | |
| | | Gang Starr | 2 | The Notorious Big | 5 | | |
| | | KRS-One | 4 | Wu-Tang Clan | 5 | | |
| | | LL Cool J | 9 | | | | |
| | | MC Hammer | 4 | | | | |
| | | Public Enemy | 7 | | | | |
| | | Queen Latifah | 6 | | | | |
| | | Run-D.M.C. | 7 | | | | |
| | | Salt-N-Pepa | 5 | | | | |
| | | Slick Rick | 2 | | | | |

**195 Songs By 40 Artists**



# Example: Lyrics of Rap Songs

## *Text conversion process – Hip Hop corpus*

```
Something's Got to Give
wish for peace between the races
Someday we shall all be one
Why fight yourself?
This one's called 'Rectify'
There's something coming to the su
There's fire all around but this i
I've seen better days than this on
I've seen better nights than this
Tension is rebuilding
Something's got to give
Something's got to give
Someday we shall all be one
Jesus Christ, we're nice
```

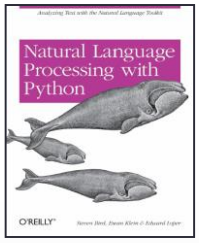
**Words - 69/45**
['wish', 'for', 'peace', 'between', 'the', 'races', 'Someday', 'we', 'shall', 'all', 'be', 'one', 'Why', 'fight', 'yourself', 'This', 'ones', 'called', 'Rectify', 'Theres', 'something', 'coming', 'to', 'the', 'surface', 'Theres', 'fire', 'all', 'around', 'but', 'this', 'is', 'all', 'illusion', 'Ive', 'seen', 'better', 'days', 'than', 'this', 'one', 'Ive', 'seen', 'better', 'nights', 'than', 'this', 'one', 'Tension', 'is', 'rebuilding', 'Somethings', 'got', 'to', 'give', 'Somethings', 'got', 'to', 'give', 'Someday', 'we', 'shall', 'all', 'be', 'one', 'Jesus', 'Christ', 'were', 'nice']

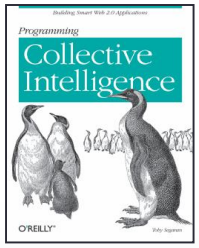
**Lowers - 69/44**
['wish', 'for', 'peace', 'between', 'the', 'races', 'someday', 'we', 'shall', 'all', 'be', 'one', 'why', 'fight', 'yourself', 'this', 'ones', 'called', 'rectify', 'theres', 'something', 'coming', 'to', 'the', 'surface', 'theres', 'fire', 'all', 'around', 'but', 'this', 'is', 'all', 'illusion', 'ive', 'seen', 'better', 'days', 'than', 'this', 'one', 'ive', 'seen', 'better', 'nights', 'than', 'this', 'one', 'tension', 'is', 'rebuilding', 'somethings', 'got', 'to', 'give', 'somethings', 'got', 'to', 'give', 'someday', 'we', 'shall', 'all', 'be', 'one', 'jesus', 'christ', 'were', 'nice']

**Alphas - 69/44**
['wish', 'for', 'peace', 'between', 'the', 'races', 'someday', 'we', 'shall', 'all', 'be', 'one', 'why', 'fight', 'yourself', 'this', 'ones', 'called', 'rectify', 'theres', 'something', 'coming', 'to', 'the', 'surface', 'theres', 'fire', 'all', 'around', 'but', 'this', 'is', 'all', 'illusion', 'ive', 'seen', 'better', 'days', 'than', 'this', 'one', 'ive', 'seen', 'better', 'nights', 'than', 'this', 'one', 'tension', 'is', 'rebuilding', 'somethings', 'got', 'to', 'give', 'somethings', 'got', 'to', 'give', 'someday', 'we', 'shall', 'all', 'be', 'one', 'jesus', 'christ', 'were', 'nice']

**NonStops - 42/30**
['wish', 'peace', 'races', 'someday', 'shall', 'one', 'fight', 'ones', 'called', 'rectify', 'theres', 'something', 'coming', 'surface', 'theres', 'fire', 'around', 'illusion', 'ive', 'seen', 'better', 'days', 'one', 'ive', 'seen', 'better', 'nights', 'one', 'tension', 'rebuilding', 'somethings', 'got', 'give', 'somethings', 'got', 'give', 'someday', 'shall', 'one', 'jesus', 'christ', 'nice']

**Stems - 42/28**
['wish', 'peac', 'race', 'someday', 'shall', 'one', 'fight', 'one', 'call', 'rectifi', 'there', 'someth', 'come', 'surfac', 'there', 'fire', 'around', 'illus', 'ive', 'seen', 'better', 'day', 'one', 'ive', 'seen', 'better', 'night', 'one', 'tension', 'rebuild', 'someth', 'got', 'give', 'someth', 'got', 'give', 'someday', 'shall', 'one', 'jesu', 'christ', 'nice']



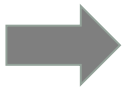

Analysis from Python NLTK Program

# Example: Lyrics of Rap Songs

## *Text conversion process – Hip Hop corpus*

```
Something's Got to Give
wish for peace between the races
Someday we shall all be one
Why fight yourself?
This one's called 'Rectify'
There's something coming to the
There's fire all around but this
I've seen better days than this
I've seen better nights than thi
Tension is rebuilding
Something's got to give
Something's got to give
Someday we shall all be one
Jesus Christ, we're nice
```

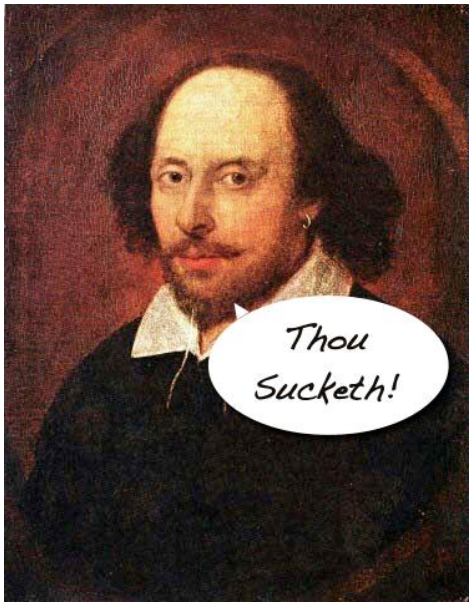
**"Bag of Words, Terms, Tokens, Stems, …"**



wordsCnts
{'all': 4, 'Christ': 1, 'rebuilding': 1, 'is': 2, 'Somethings': 2, 'surface': 1, 'yourself': 1, 'something': 1, 'seen': 2, 'Jesus': 1, 'fire': 1, 'Tension': 1, 'nights': 1, 'for': 1, 'peace': 1, 'fight': 1, 'better': 2, 'to': 3, 'Ive': 2, 'between': 1, 'got': 2, 'Why': 1, 'nice': 1, 'be': 2, 'we': 2, 'around': 1, 'shall': 2, 'This': 1, 'but': 1, 'ones': 1, 'coming': 1, 'one': 4, 'than': 2, 'Theres': 2, 'Someday': 2, 'give': 2, 'Rectify': 1, 'this': 3, 'wish': 1, 'days': 1, 'illusion': 1, 'races': 1, 'were': 1, 'the': 2, 'called': 1}
lowerCnts
{'someday': 2, 'all': 4, 'rectify': 1, 'is': 2, 'surface': 1, 'yourself': 1, 'ive': 2, 'something': 1, 'seen': 2, 'jesus': 1, 'nights': 1, 'christ': 1, 'for': 1, 'peace': 1, 'somethings': 2, 'fight': 1, 'better': 2, 'to': 3, 'between': 1, 'got': 2, 'illusion': 1, 'nice': 1, 'be': 2, 'we': 2, 'around': 1, 'than': 2, 'shall': 2, 'fire': 1, 'rebuilding': 1, 'but': 1, 'ones': 1, 'coming': 1, 'one': 4, 'why': 1, 'give': 2, 'this': 4, 'wish': 1, 'days': 1, 'tension': 1, 'races': 1, 'were': 1, 'the': 2, 'called': 1, 'theres': 2}
alphasCnts
{'someday': 2, 'all': 4, 'rectify': 1, 'is': 2, 'surface': 1, 'yourself': 1, 'ive': 2, 'something': 1, 'seen': 2, 'jesus': 1, 'nights': 1, 'christ': 1, 'for': 1, 'peace': 1, 'somethings': 2, 'fight': 1, 'better': 2, 'to': 3, 'between': 1, 'got': 2, 'illusion': 1, 'nice': 1, 'be': 2, 'we': 2, 'around': 1, 'than': 2, 'shall': 2, 'fire': 1, 'rebuilding': 1, 'but': 1, 'ones': 1, 'coming': 1, 'one': 4, 'why': 1, 'give': 2, 'this': 4, 'wish': 1, 'days': 1, 'tension': 1, 'races': 1, 'were': 1, 'the': 2, 'called': 1, 'theres': 2}
nonstopsCnts
{'someday': 2, 'rectify': 1, 'surface': 1, 'one': 4, 'ive': 2, 'something': 1, 'seen': 2, 'nights': 1, 'christ': 1, 'peace': 1, 'somethings': 2, 'fight': 1, 'better': 2, 'got': 2, 'illusion': 1, 'nice': 1, 'tension': 1, 'around': 1, 'shall': 2, 'fire': 1, 'ones': 1, 'coming': 1, 'jesus': 1, 'give': 2, 'wish': 1, 'days': 1, 'races': 1, 'rebuilding': 1, 'called': 1, 'theres': 2}
stemsCnts
{'someday': 2, 'ive': 2, 'jesu': 1, 'one': 5, 'rectifi': 1, 'seen': 2, 'surfac': 1, 'christ': 1, 'come': 1, 'someth': 3, 'there': 2, 'fight': 1, 'better': 2, 'call': 1, 'got': 2, 'nice': 1, 'tension': 1, 'around': 1, 'shall': 2, 'fire': 1, 'illus': 1, 'day': 1, 'give': 2, 'wish': 1, 'peac': 1, 'race': 1, 'night': 1, 'rebuild': 1}

# Example: Lyrics of Rap Songs

## *Text conversion process – structural features*




| | Words | Alphas | Stems |
|---|---|---|---|
| Total | 127643 | 120263 | 52094 |
| TotUniq | 12521 | 10182 | 7675 |
| NumLines | 16412 | 16412 | 16412 |
| Chars | 495063 | 487265 | 247340 |
| AvgWrd/Ln | 7.8 | 7.3 | 3.2 |
| AvgCh/Word | 3.9 | 4.1 | 4.7 |
| LexDiv | 10.2 | 11.8 | 6.8 |

| Author | WordLen | AvgWords/Sent | Words/Voca |
|---|---|---|---|
| austen-emma.txt | 4 | 21 | 26 |
| austen-persuasion.txt | 4 | 23 | 16 |
| austen-sense.txt | 4 | 24 | 22 |
| bible-kjv.txt | 4 | 33 | 79 |
| blake-poems.txt | 4 | 18 | 5 |
| bryant-stories.txt | 4 | 17 | 14 |
| burgess-busterbrown.txt | 4 | 14 | 12 |
| carroll-alice.txt | 4 | 16 | 12 |
| chesterton-ball.txt | 4 | 17 | 11 |
| chesterton-brown.txt | 4 | 19 | 11 |
| chesterton-thursday.txt | 4 | 16 | 10 |
| edgeworth-parents.txt | 4 | 18 | 24 |
| melville-moby_dick.txt | 4 | 24 | 15 |
| milton-paradise.txt | 4 | 52 | 10 |
| shakespeare-caesar.txt | 4 | 12 | 8 |
| shakespeare-hamlet.txt | 4 | 13 | 7 |
| shakespeare-macbeth.txt | 4 | 13 | 6 |
| whitman-leaves.txt | 4 | 35 | 12 |

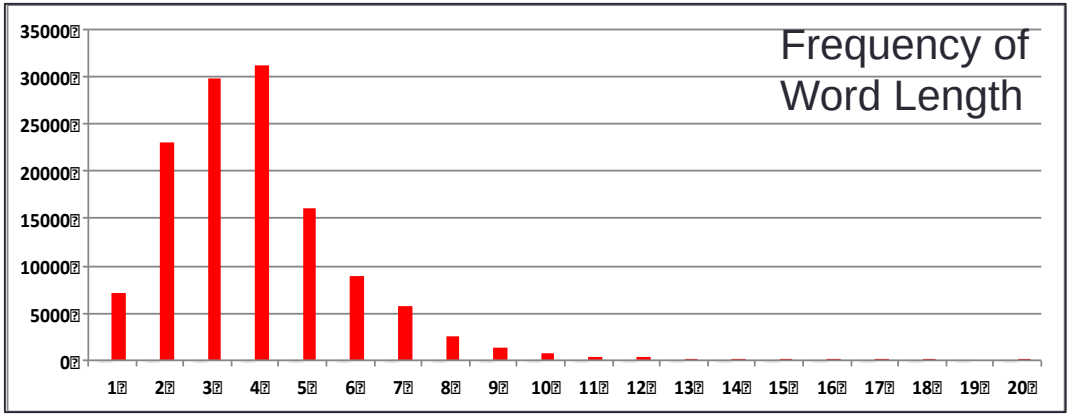
# Example: Lyrics of Rap Songs

## *Text conversion process – structural features*


…
Final answer "not to be", "not to be" is right!
Next question - to build winged shoes or autophagy
Silk screen band tees, take apart a vcr, ringer off, canned peas
Cabin fever mi amor
Patiently adhering to the chandelier ta key-in-door
To usher in the understated anarchy of leisureforce
Led a purple tongue and ratty caballeros
Up over the black rainbow into the house of mirrors
To become a thousand zeroes
Echoing a twisted alchemy, freak flags, fluttering to circadian free jazz
Sleep apnea scratching "bring that beat back"
I doze off, clothes on, noise in the feedbag
Shhh.. om nom nom, blinds drawn
Compost thrown to the spine pile, bygones, mangy
Intimately spaced pylons on a plot of inhospitable terrain
Hi mom!
…
(from Leisureforce)

# Example: Lyrics of Rap Songs

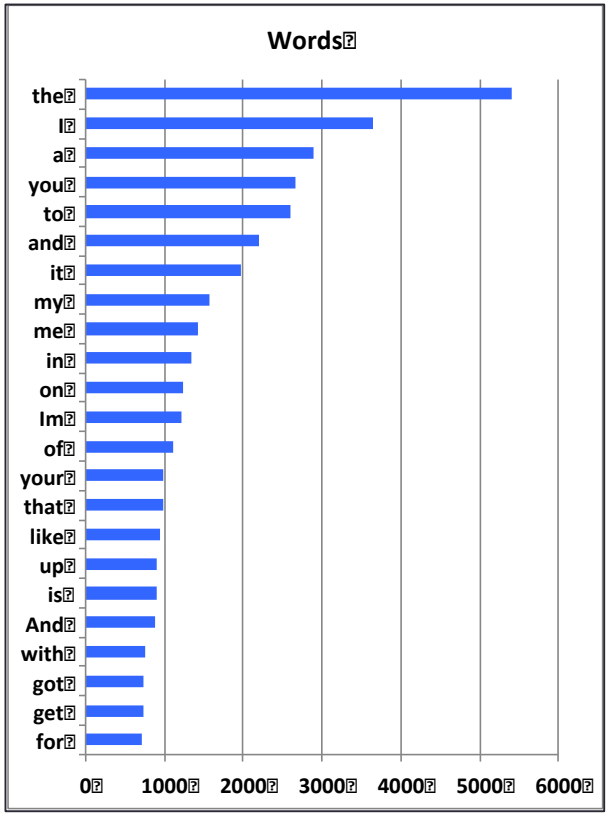
## *Text conversion process – word length*

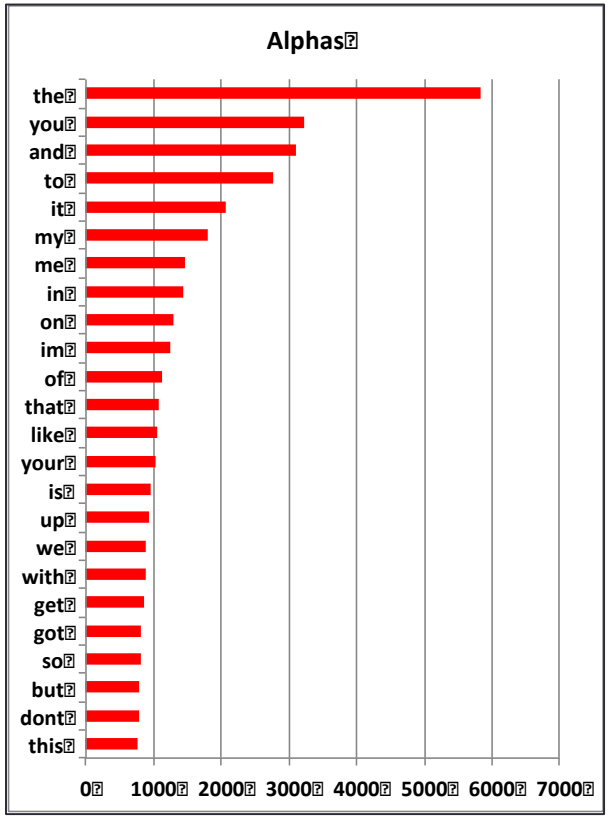


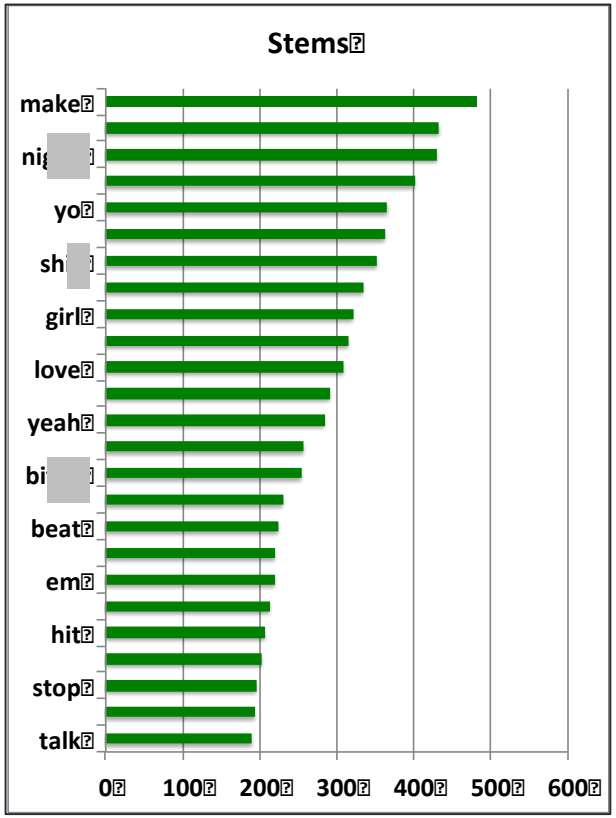

| Word | Freq | Length | Word | Freq | Length |
|---|---|---|---|---|---|
| EeeyiyiyiyiyeahtheDoggPoundsinthehouowwowse | 1 | 43 | Ratatatatatatatatata | 1 | 20 |
| EeeyiyiyiyiyahtheDoggPoundsinthehouowwse | 1 | 40 | Whatwhawhatwhatwhats | 1 | 20 |
| Wooooooooooooooooooooooeeeeeeeeeeeee | 1 | 39 | rrrrrrrrrrrrrrrah | 1 | 18 |
| Wowowowowowowowowowowowowowowow | 1 | 34 | EazymotherfuckingE | 1 | 18 |
| MisterImTooGoodToCallOrWriteMyFans | 1 | 34 | wowowowowowowowowo | 1 | 18 |
| supercalifragilisticexpialidocious | 1 | 34 | CharlieBakerDenver | 1 | 18 |
| Checkchcheckcheckcheckchcheck | 1 | 29 | Heyheyheyheyheyhey | 1 | 18 |
| checkchcheckcheckcheckchcheck | 1 | 29 | moneymakinmachines | 1 | 18 |
| yaauhhhaaaauhhhhahhhauhhhhll | 1 | 28 | YouknowhatImsayin | 1 | 17 |
| bowwowwowyippyyoyoyoyoyo | 1 | 26 | Wowowowowowowowow | 1 | 17 |
| Rockthaithaiotoyothaithai | 1 | 25 | youknowhatImsayin | 1 | 17 |
| Bowwowwowyippyyoyippyyay | 1 | 24 | BedfordStuyvesant | 1 | 17 |
| Workwaworkworkworkwawork | 1 | 24 | whoahhhhhhhhhhhh | 1 | 17 |
| Whythaithaiotoyothaithai | 1 | 24 | Checkacheckacheck | 1 | 17 |
| Aowowowowowowowowowowo | 1 | 24 | yaknowhatImsayin | 2 | 16 |
| ooooooooooooooooooooo | 1 | 23 | bowwowwowyippyyo | 1 | 16 |
| HeyyyeyyyeEYEYyyyEYYYY | 1 | 22 | crossedSyndicate | 1 | 16 |
| Ahungahungahungahunga | 1 | 21 | Incomprehensible | 1 | 16 |
| FUNNNNNNNNNNNNNNNNNN | 1 | 21 | doodoodoodadahhh | 1 | 16 |
| hippahoppadippaboppa | 1 | 20 | Micmicmicrophone | 1 | 16 |
| Awowowowowowowowowow | 1 | 20 | Extraterrestrial | 1 | 16 |
| GreGreGreGreGreatest | 1 | 20 | Aftermathematics | 1 | 16 |
| WHODIWHOOOOOOOOOOOOO | 1 | 20 | multimillionaire | 1 | 16 |

# Example: Lyrics of Rap Songs

## *Text conversion process – frequencies*

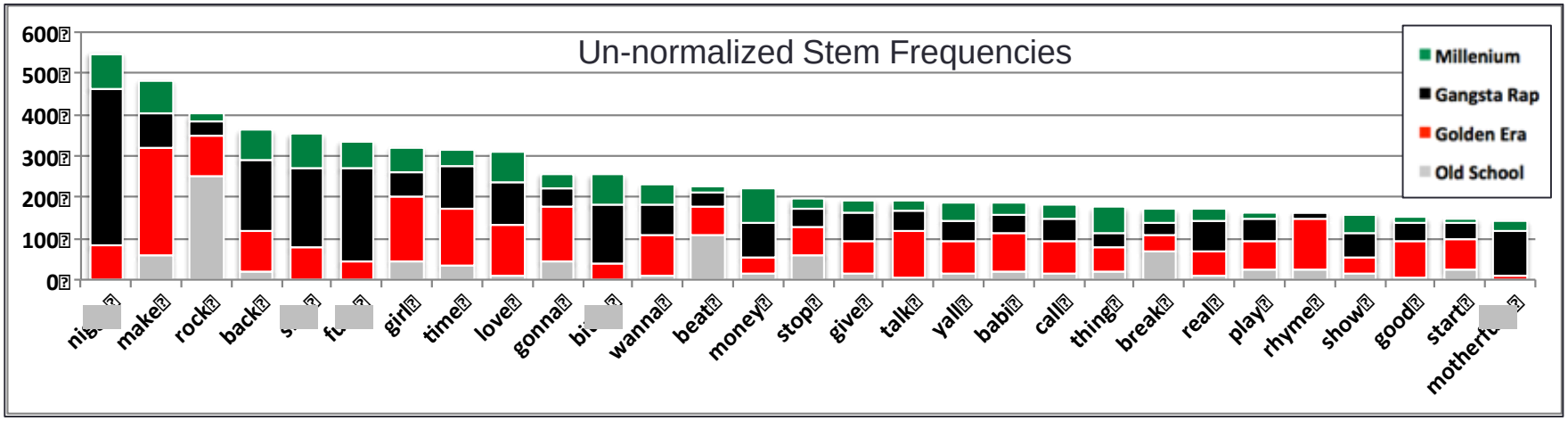


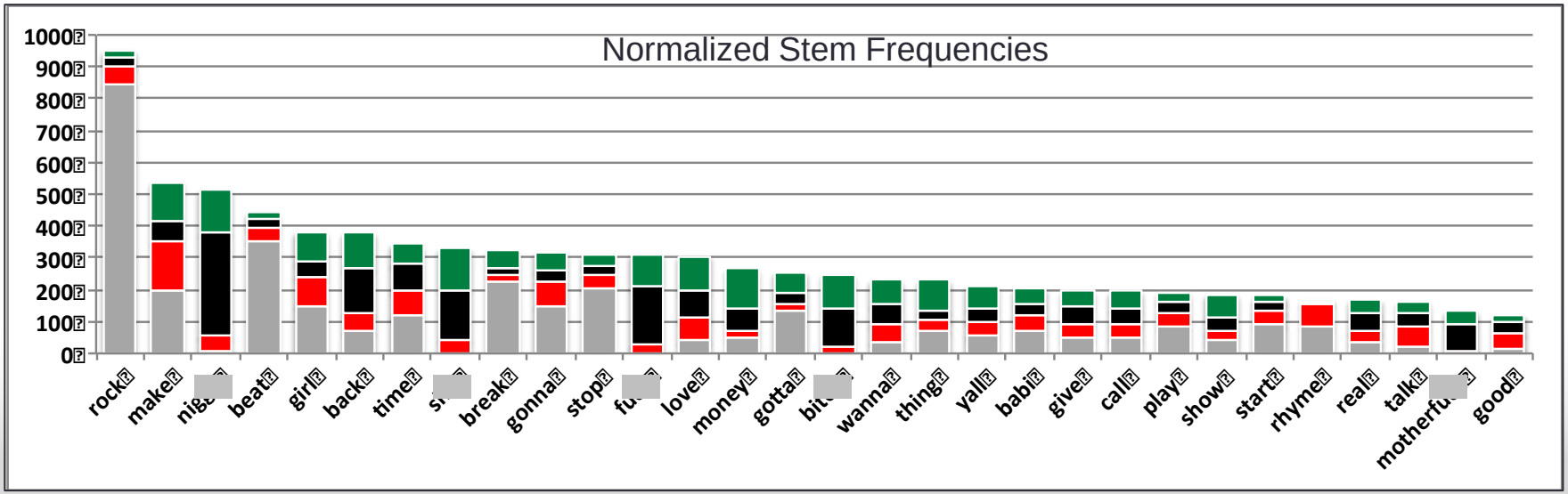



# Example: Lyrics of Rap Songs

## *Text analysis process – stem frequencies by era*

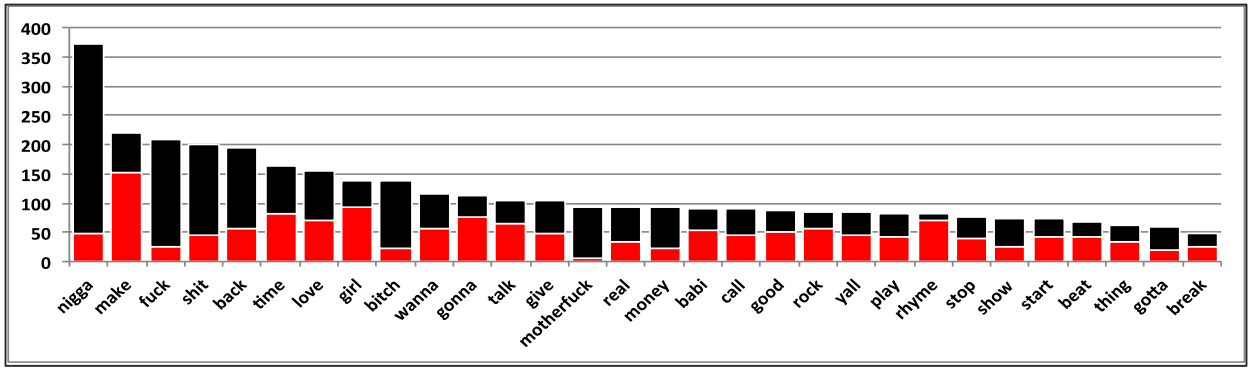


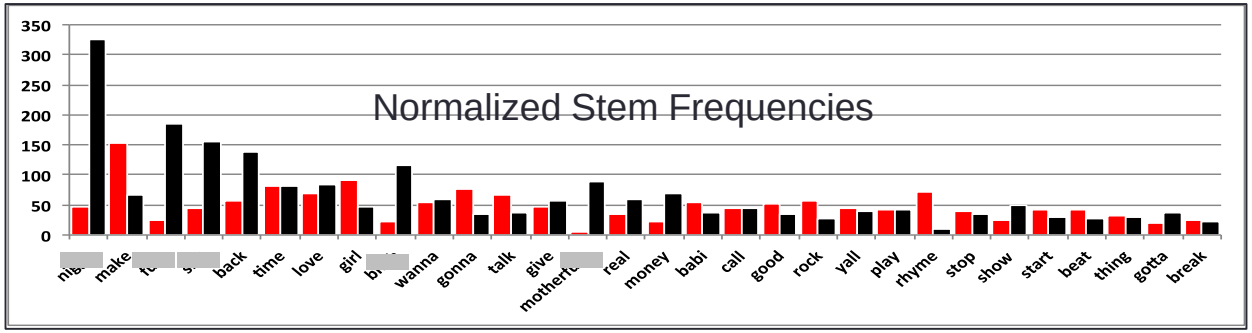



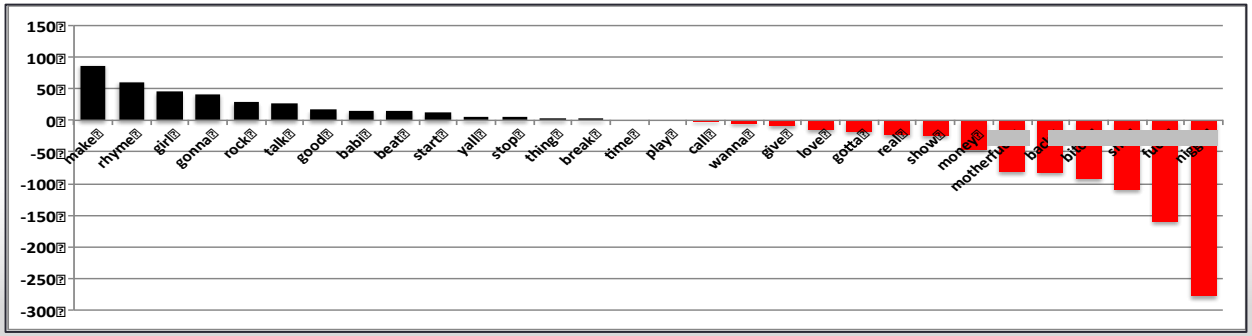
# Example: Lyrics of Rap Songs

## *Text analysis process – stem frequencies by era*

Normalized Stem Frequencies

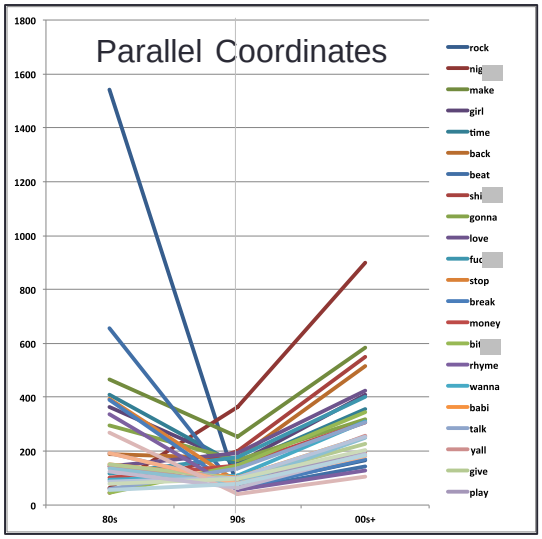


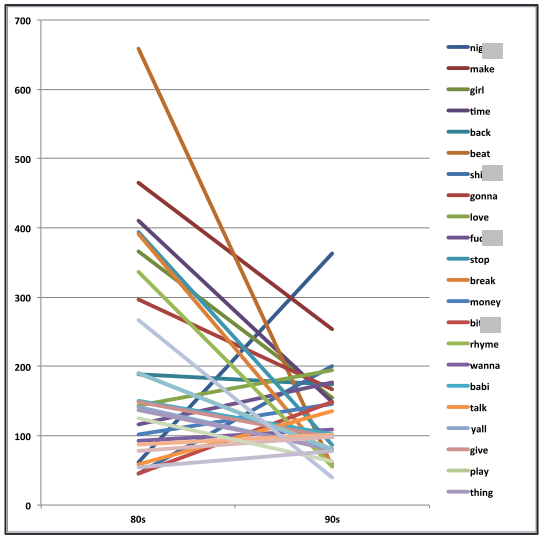



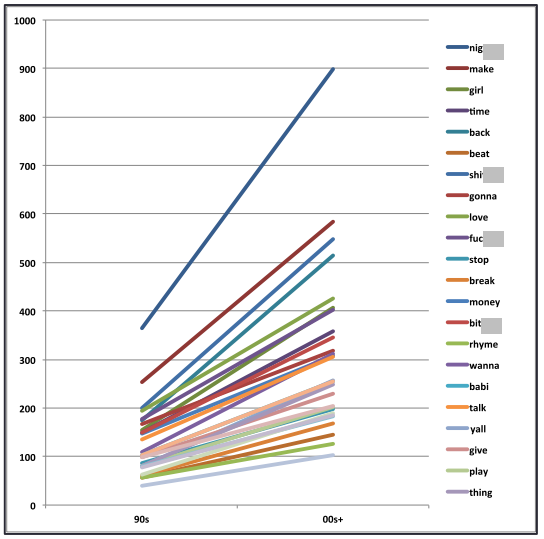



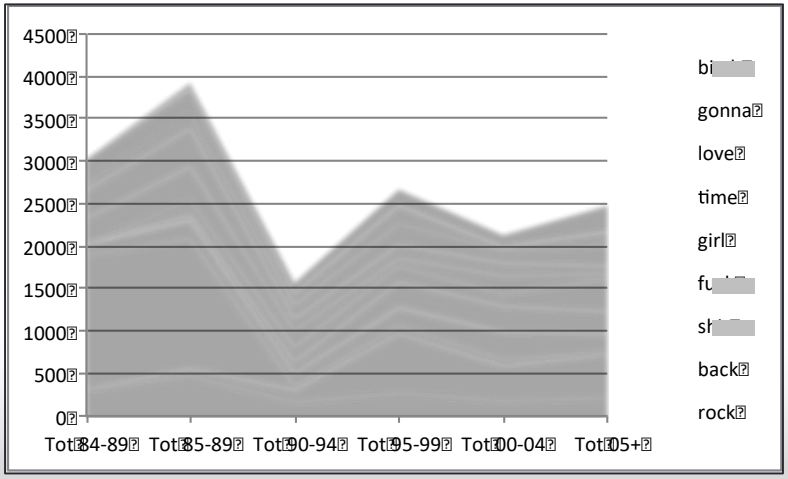

# Example: Lyrics of Rap Songs

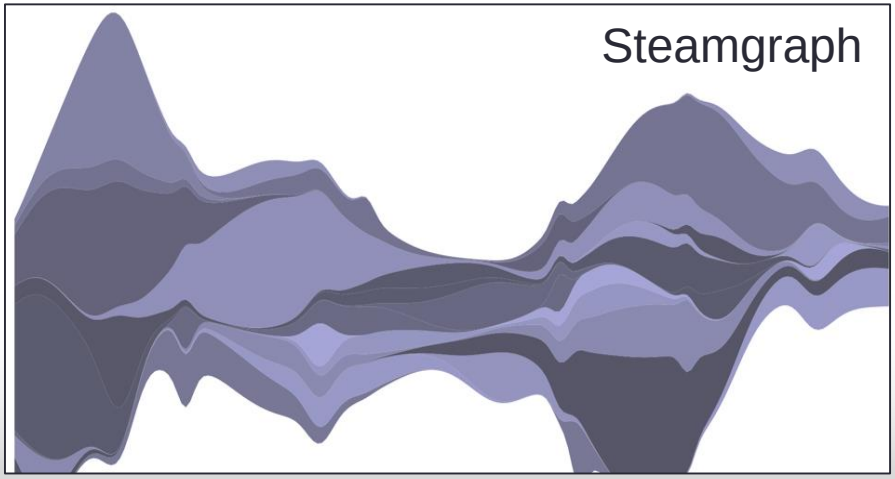
*Text analysis process – stem frequencies by time*



# Example: Lyrics of Rap Songs

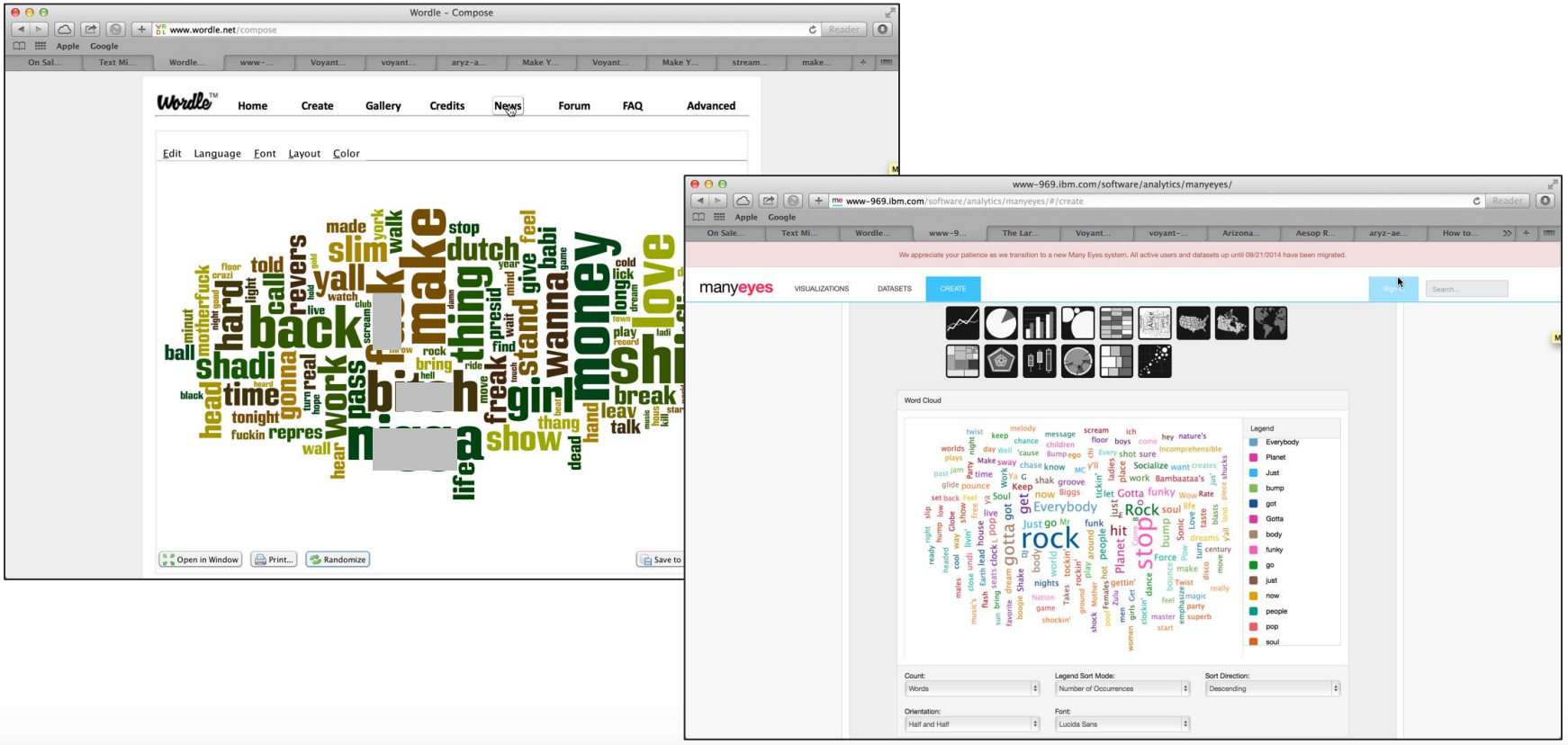
*Text analysis process – frequencies with less precision*




Word Clouds – They're everywhere along with subway maps, periodic charts…

# Example: Lyrics of Rap Songs

*Text analysis process – frequencies with less precision*




Word Clouds – They're everywhere along with subway maps, periodic charts…

# Example: Lyrics of Rap Songs

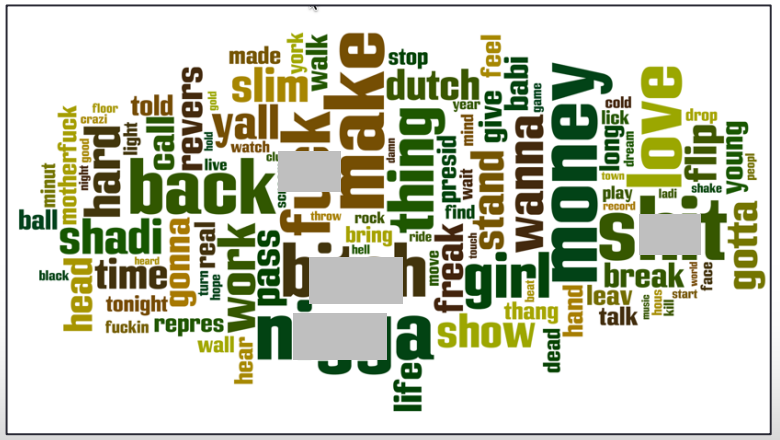
*Text analysis process – frequencies with less precision*

Old School


Golden Era

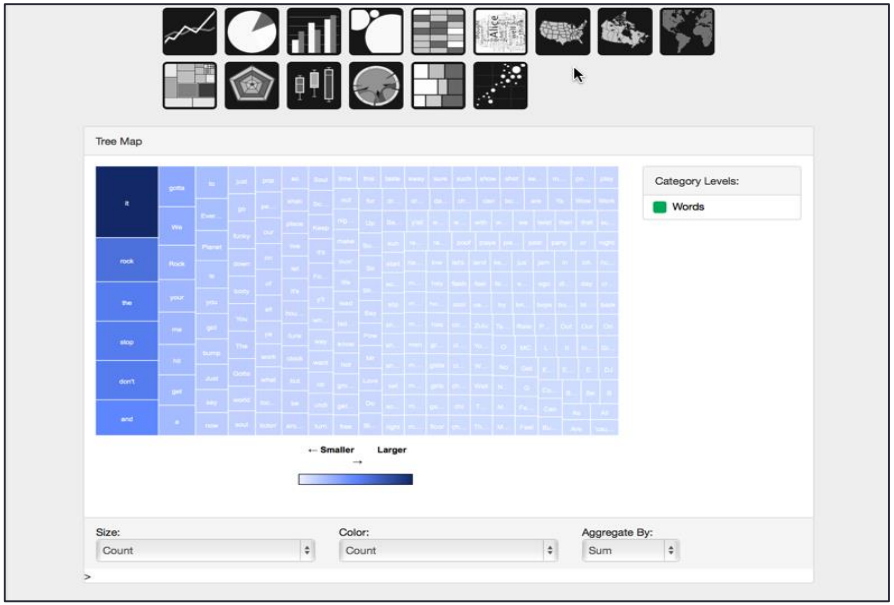

Gangsta


Millenium



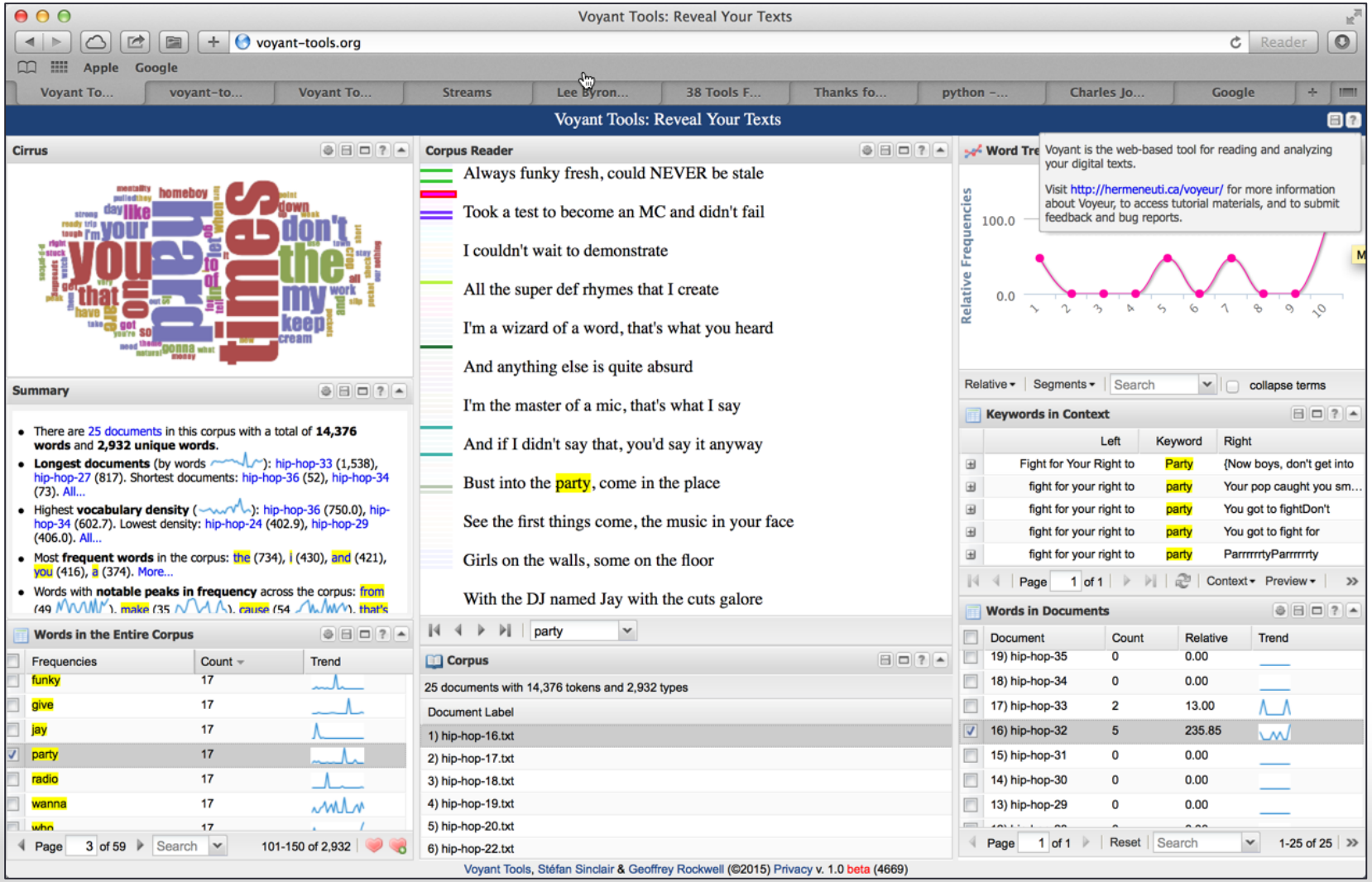
# Example: Lyrics of Rap Songs

*Text analysis process – frequencies with less precision*

# Example: Lyrics of Rap Songs

*Text analysis process – general interactive platform*

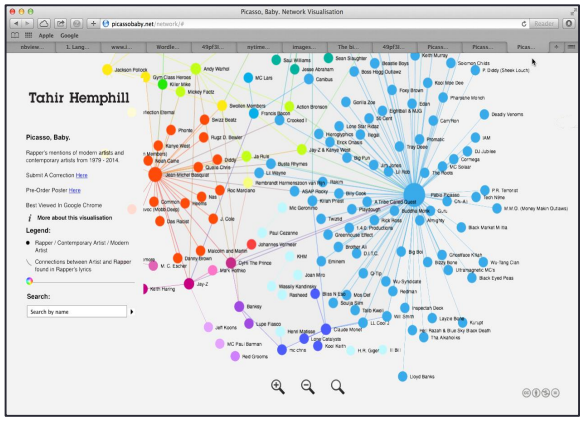


# Example: Lyrics of Rap Songs

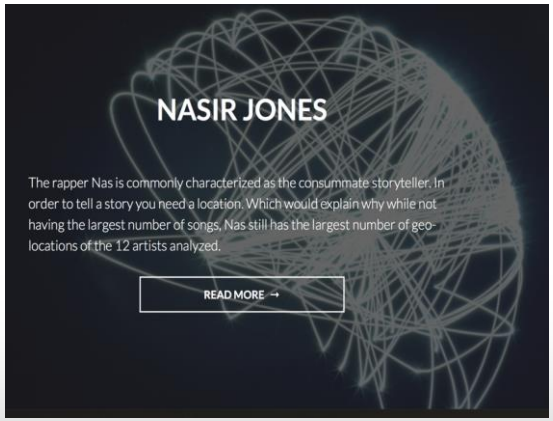
*Types of (visual) analysis*

Cluster/Classification

Network

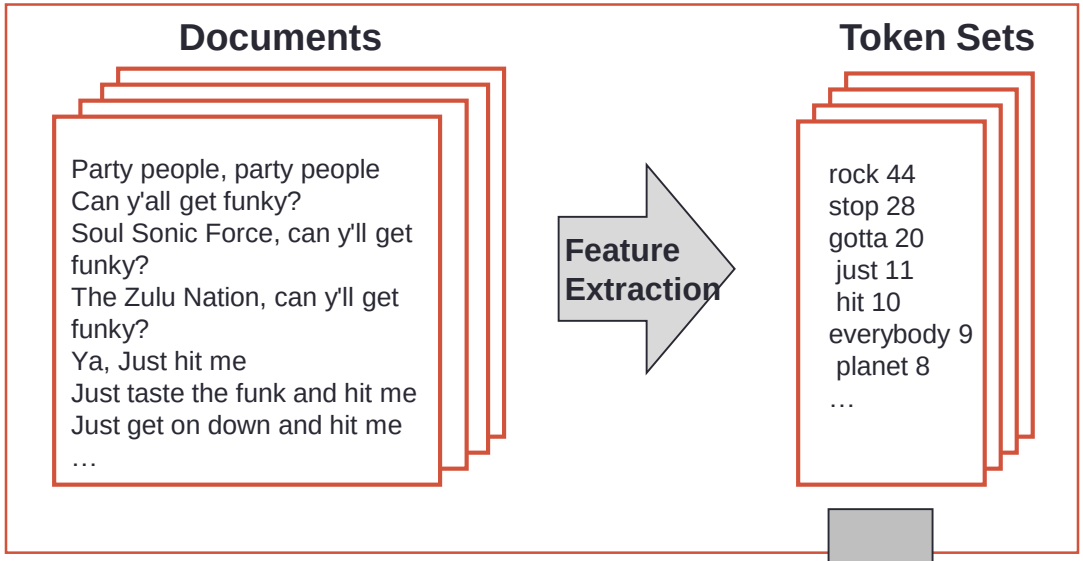



Geospatial



# Example: Lyrics of Rap Songs

## *Text analysis process – Role of Document-Term Matrix*




*Feature Extraction & Weighting*

**"Bag of Words, Terms or Tokens"**

**Doc/Token Matrix:**
**Vectors of Words, Terms or Tokens by Doc**

| | Token1 | Token2 | Token3 | Token4 | … |
|---|---|---|---|---|---|
| Doc1 | 1 | 2 | 2 | 4 | |
| Doc2 | 4 | 2 | 3 | 0 | |
| Doc3 | 1 | 1 | 1 | 0 | |
| Doc4 | 1 | 1 | 1 | 2 | |
| … | | | | | |

**"Bag of Words" (BOW) or Vector Space Model (VSM): Words or Tokens are attributes and documents are examples**

# Example: Lyrics of Rap Songs

## *Text analysis process – Document-Term Matrix*

| SongCode | make | nigga | rock | back | shit | fuck | … | girl | sell | suck | sweat | thug | uhuh | week | worri | TotalStems |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Song1-OS-NY-80s | 3 | 0 | 43 | 1 | 0 | 0 | … | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 193 |
| Song2-OS-NY-80s | 1 | 0 | 0 | 0 | 0 | 0 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 116 |
| Song3-OS-NY-80s | 3 | 0 | 8 | 0 | 0 | 0 | … | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 174 |
| Song4-OS-NY-80s | 3 | 0 | 1 | 0 | 0 | 0 | … | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 82 |
| Song5-OS-NY-80s | 0 | 0 | 1 | 1 | 0 | 0 | … | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 166 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| Song16-GA-NY-80s | 0 | 0 | 0 | 0 | 0 | 0 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 55 |
| Song17-GA-NY-80s | 2 | 0 | 2 | 3 | 0 | 0 | … | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 93 |
| Song18-GA-NY-80s | 2 | 0 | 6 | 1 | 0 | 0 | … | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 108 |
| Song19-GA-NY-80s | 1 | 0 | 10 | 1 | 0 | 0 | … | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 98 |
| Song20-GA-NY-80s | 0 | 0 | 6 | 1 | 0 | 0 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 103 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| Song102-GR-NJ-90s | 0 | 0 | 0 | 0 | 1 | 32 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 106 |
| Song103-GR-NJ-90s | 0 | 3 | 0 | 2 | 1 | 1 | … | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 98 |
| Song104-GR-NJ-80s | 1 | 0 | 3 | 5 | 5 | 0 | … | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 190 |
| Song105-GR-NJ-80s | 1 | 0 | 1 | 0 | 0 | 0 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 56 |
| Song106-GR-NJ-90s | 1 | 0 | 0 | 1 | 0 | 0 | … | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 101 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| Song164-ML-LA-00s | 9 | 4 | 0 | 6 | 1 | 0 | … | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 123 |
| Song165-ML-LA-00s | 7 | 0 | 0 | 1 | 3 | 2 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 84 |
| Song166-ML-LA-00s | 0 | 0 | 0 | 1 | 0 | 0 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 49 |
| Song167-ML-LA-00s | 0 | 1 | 1 | 0 | 4 | 0 | … | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 66 |
| Song168-ML-LA-00s | 0 | 6 | 0 | 3 | 2 | 1 | … | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 77 |
| TotalStems | 482 | 430 | 401 | 362 | 352 | 334 | | 321 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | |

| | |
|---|---|
| 0 | 0 |
| 1 | 0 |
| 2 | 1 |
| 3 | 1 |
| 1 | 0 |
| 1 | 2 |
| 0 | 0 |
| 0 | 0 |
| 13 | 17 |
| 8 | 3 |
| 1 | 0 |
| 0 | 11 |
| 12 | 10 |

| | |
|---|---|
| 3 | 2 |
| 1 | 6 |
| 3 | 1 |
| 3 | 0 |
| 0 | 1 |
| 0 | 0 |
| 3 | 4 |
| 2 | 0 |
| 2 | 1 |
| 13 | 11 |
| 12 | 1 |
| 0 | 0 |
| 5 | 3 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Song1-OS-N' | 3 | 0 | 43 | 1 | 0 | 0 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 28 | 0 | 0 | 1 |
| Song2-OS-N' | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 1 | 0 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Song13-OS- | 5 | 0 | 112 | 2 | 0 | 0 | 0 | 3 | 0 | 11 | 0 | 0 | 7 | 1 | 1 | 0 | 0 | 0 |
| Song14-OS- | 8 | 0 | 39 | 2 | 0 | 0 | 12 | 5 | 1 | 10 | 0 | 1 | 4 | 1 | 2 | 5 | 2 | 2 |

Which pair of song vectors is more correlated? (.05 vs .49)

Which pair of word vectors is more correlated? (.7 vs. .6)

# Example: Lyrics of Rap Songs

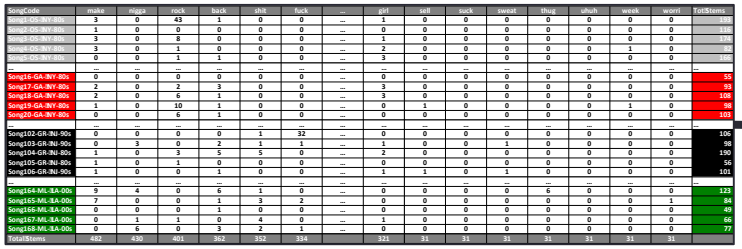
## *Text analysis process – Document-Term Matrix*

**Normalizing the data**

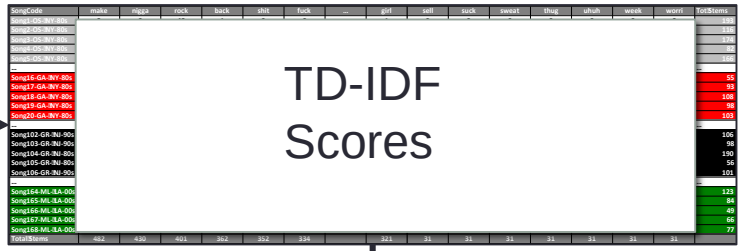
- Binary Frequencies: tf =1 for tf>0; otherwise 0
- Term Frequencies: tf(i,j)/Sum of tf(i,j) in Doc K
- Log Frequencies: 1 + log(tf) for tf>0; otherwise 0
- Normalized Frequencies: Divide each frequency by SQRT of Sum of Squares of the frequencies within the vector (column)
- Term Frequency–Inverse Document Frequency
  - TF: Freq of term for given doc/ sum of words for given doc
  - Inverse Document Frequency: log(N/(1+D)) where N is total number of docs and D is number with term
  - TF * IDF

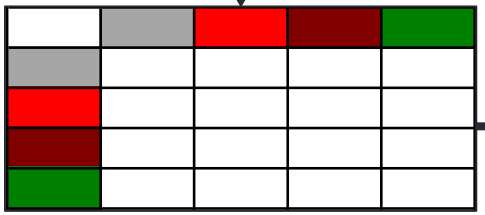
# Example: Lyrics of Rap Songs

## *Text analysis process – Clustering*



Doc-Term Matrix



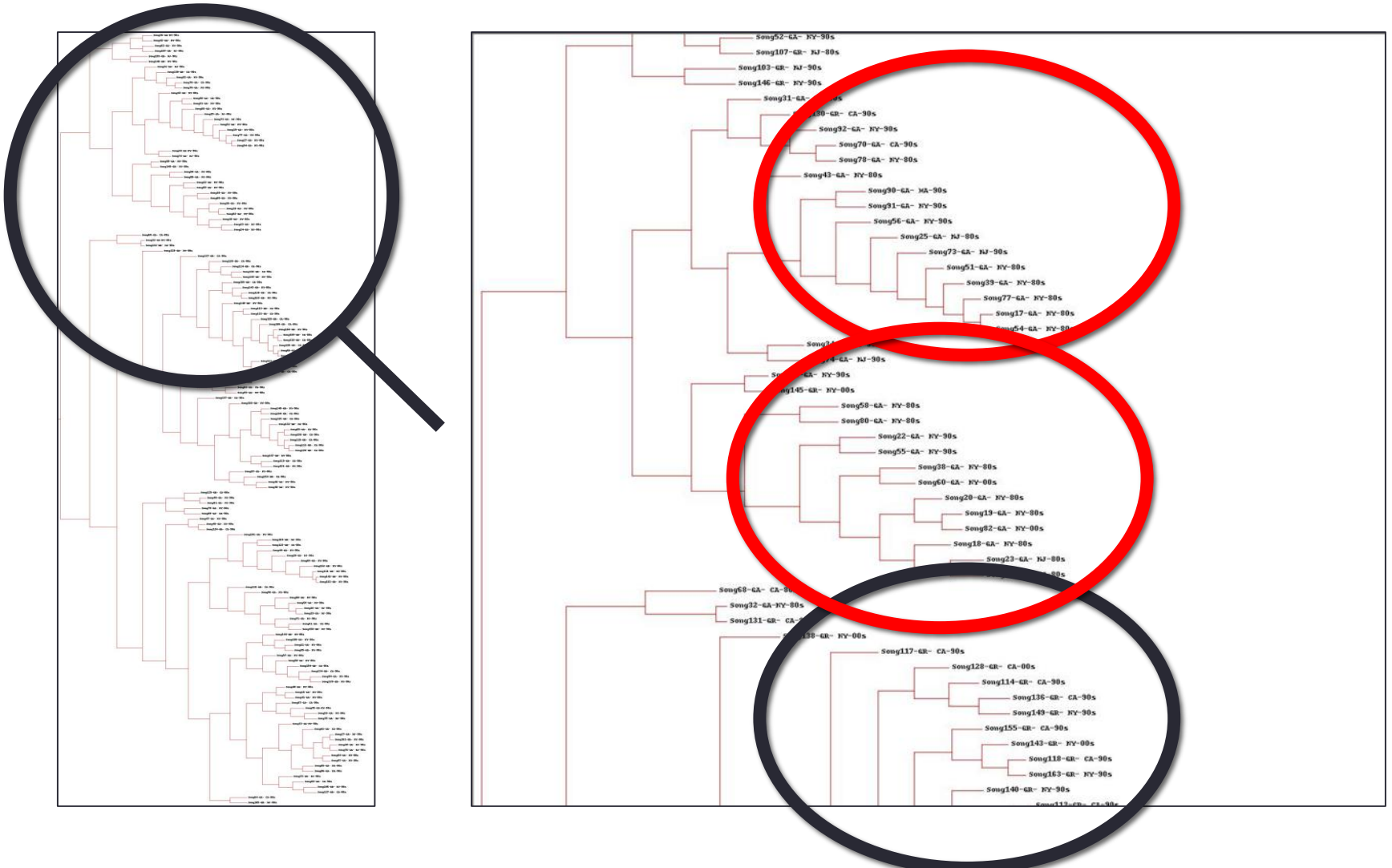



Correlation Matrix
(t x t or d x d)

Dendogram

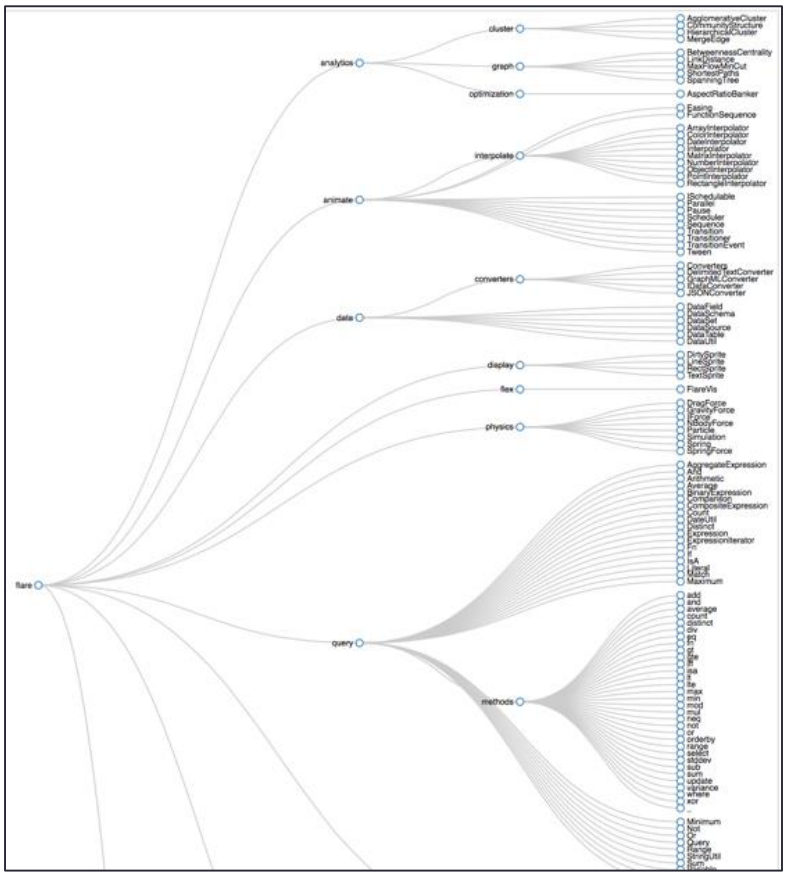
# Example: Lyrics of Rap Songs

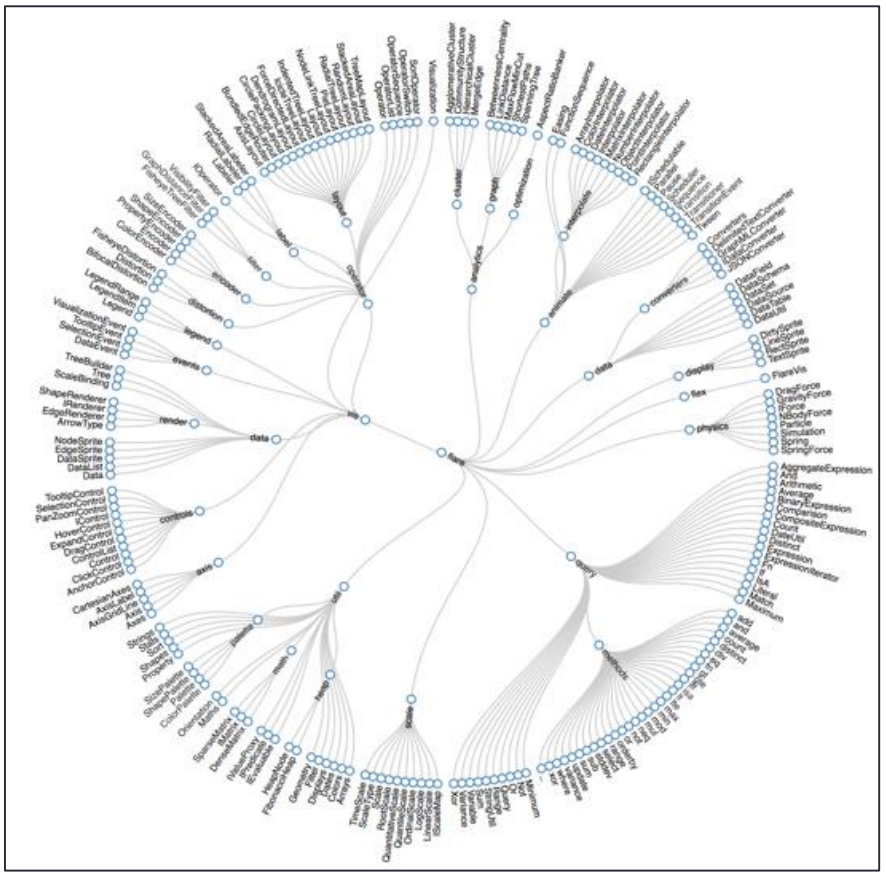
## *Text analysis process – Clustering*



# Example: Lyrics of Rap Songs

## *Text analysis process – Clustering*



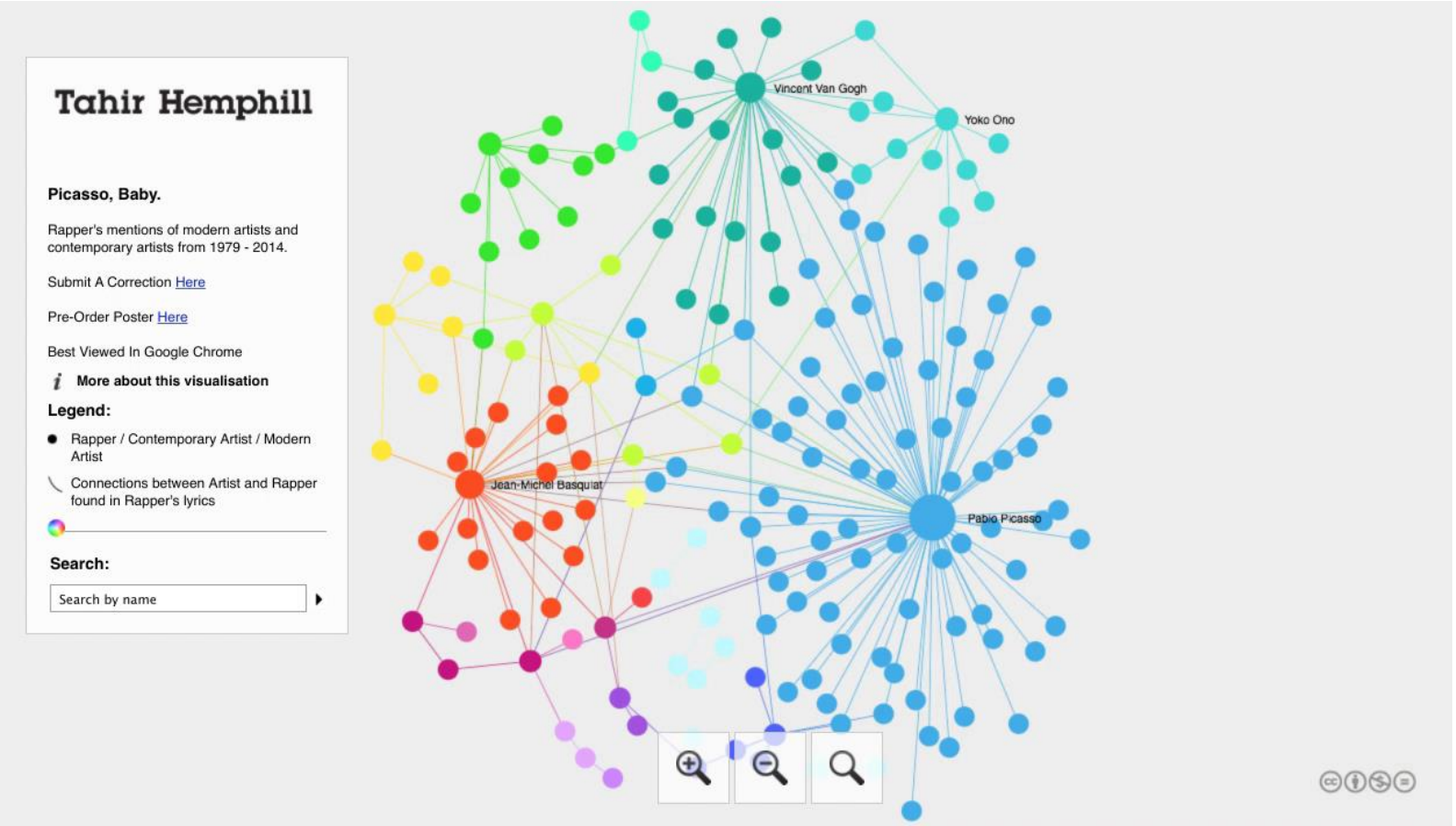
New Versions

# Example: Social Network Analysis

# EXAMPLE: SOCIAL NETWORK ANALYSIS

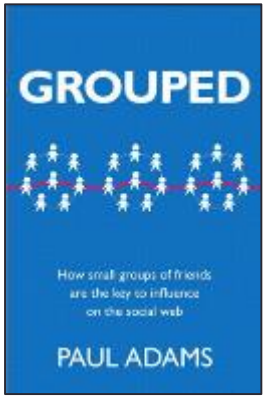
# Example: Lyrics of Rap Songs

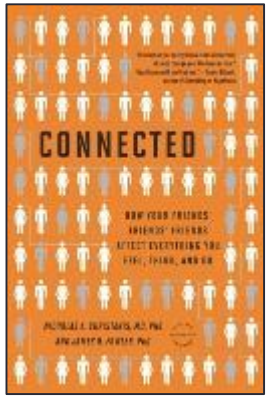
*Network analysis*

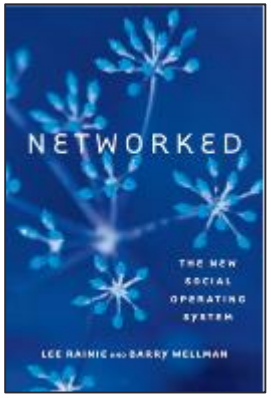


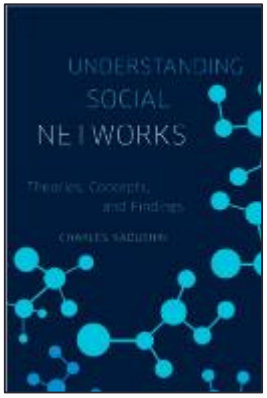
# Social Network Analysis

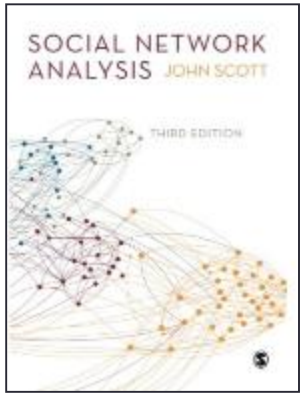
## *General Resources*

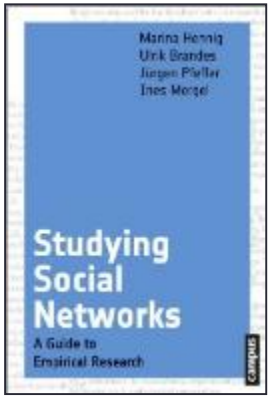


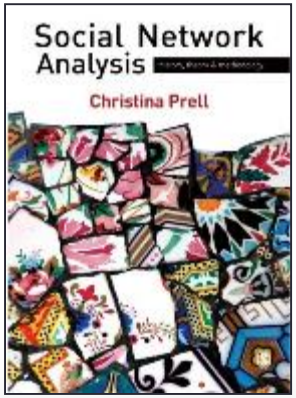



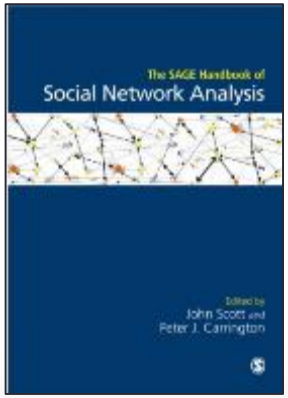



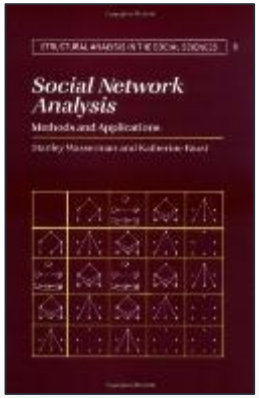



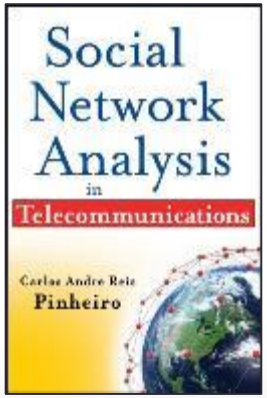



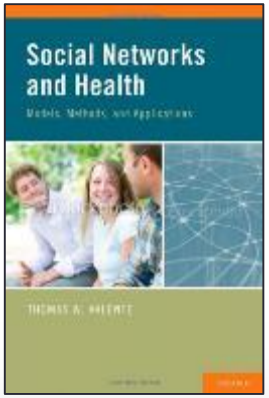



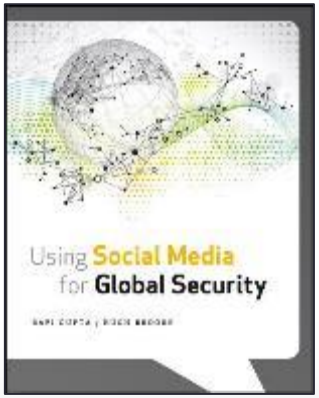



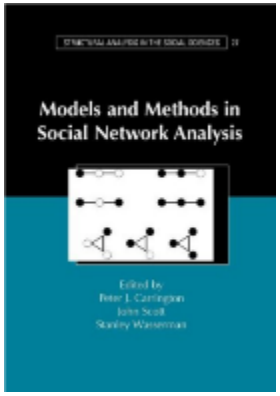



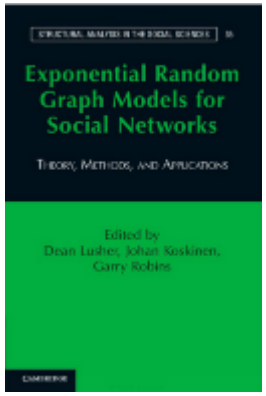



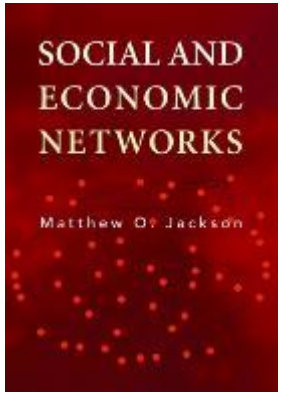



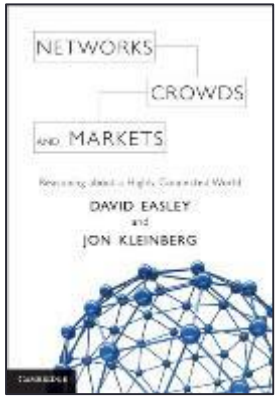
# Social Network Analysis

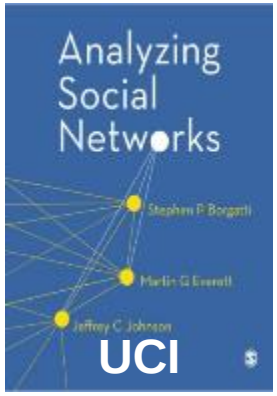
## *General Resources with Packages*

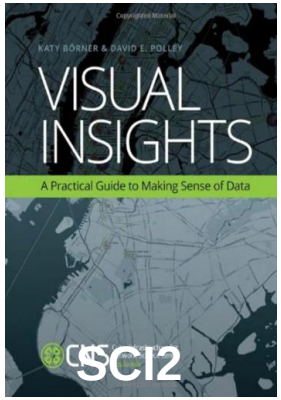


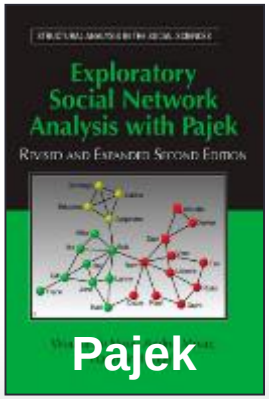



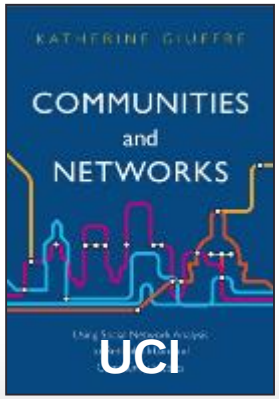



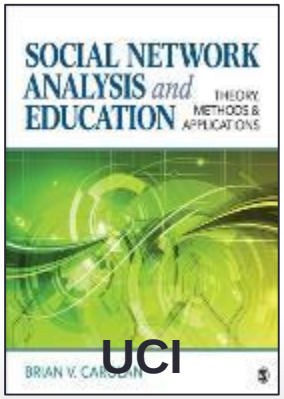



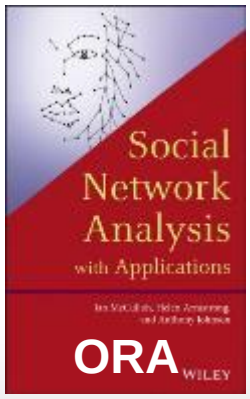



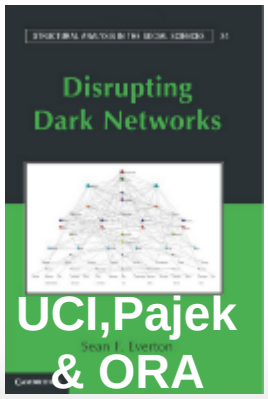



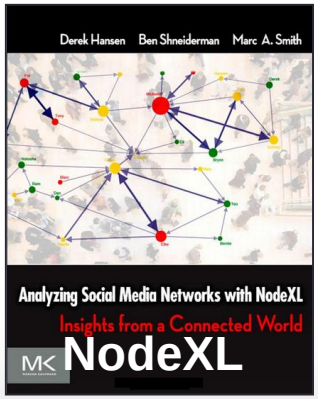



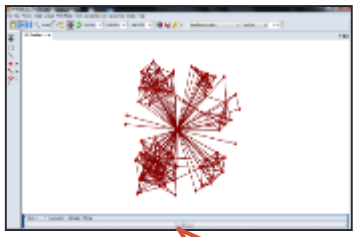
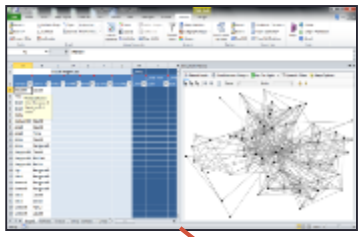
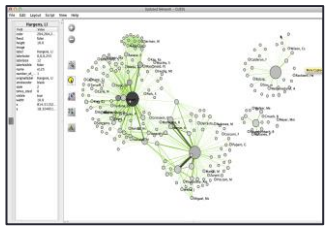



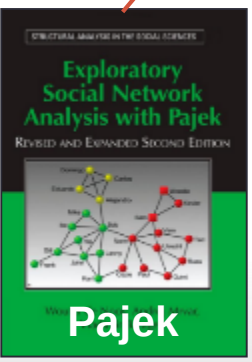



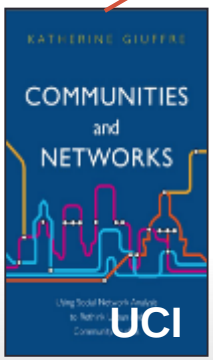
# Social Network Analysis

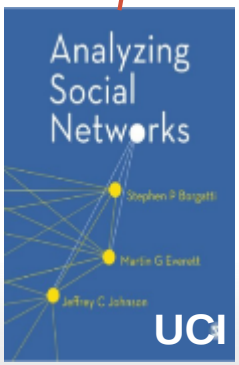
## *Social Network Packages*

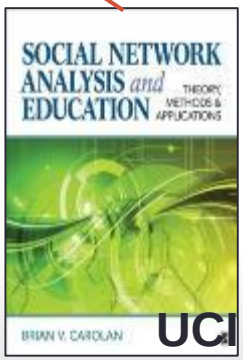


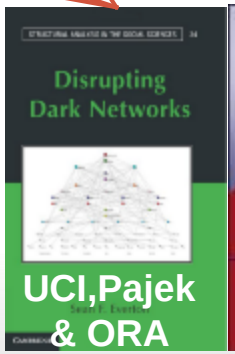




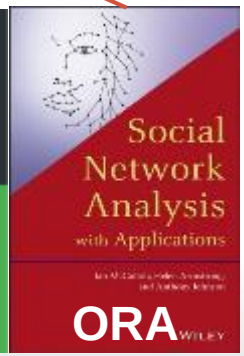


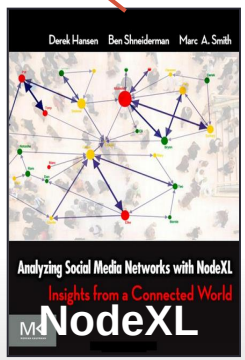



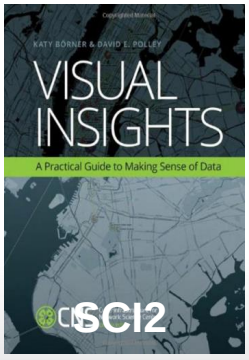



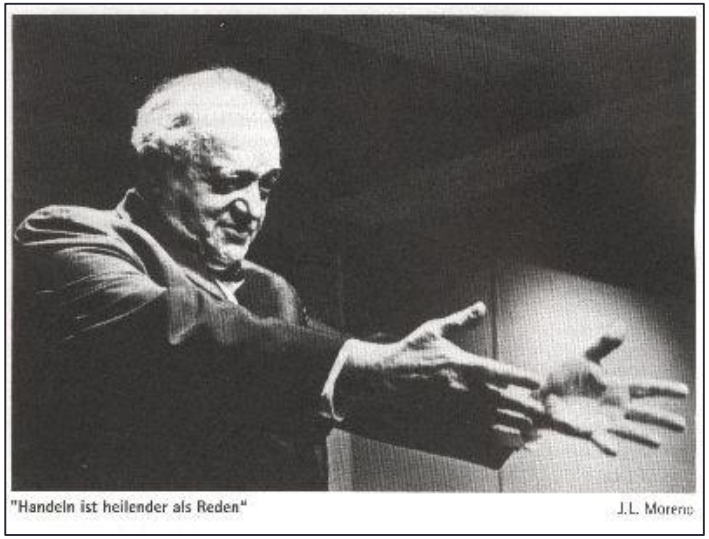



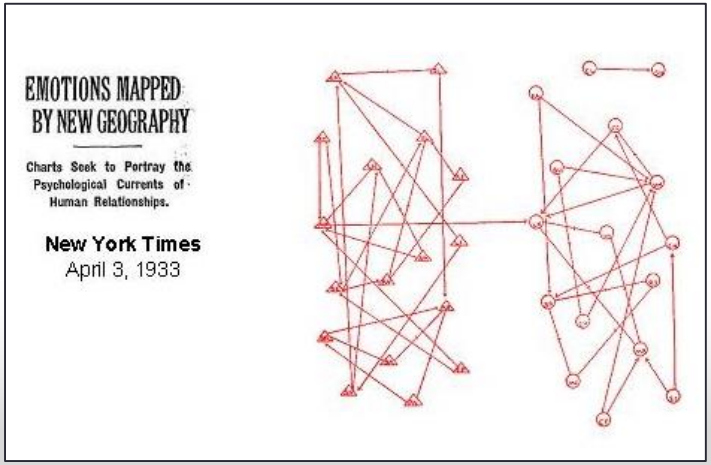



# Social Network Analysis

*Long tie to visualization*



"Handeln ist heilender als Reden" J.L. Moreno



EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the Psychological Currents of Human Relationships.

New York Times
April 3, 1933

| Decade | Scholar(s) | Innovations |
|---|---|---|
| 1900 | Simmel | Dyads and Triads |
| 1930 | Jacob Moreno | Sociometry, Sociograms |
| 1930 | Mayo & Warner | Hawthorne Study |
| 1940 | Forsyth & Katz | (Adjacency) Matrix |
| 1940 | Luce & Festinger | Matrix Algebra, n-cliques |
| 1940 | Bavelas | Centrality, Centralization |
| 1950 | Radcliff-Browne | Social Structure as a Network of Social Relations |
| 1950 | Harary & Norman | Graph Theory, Structural Balance |
| 1950 | Manchester School | Ego Networks |
| 1950 | Bott | Connectedness, Density |
| 1950 | Barnes | Social Network' |
| 1950 | Homans | Social Exchange |
| 1960 | James Davis | Clustering, Transitivity |
| 1960 | Coleman | Diffusion in Social Networks |
| 1960 | Milgram | Small world |
| 1970 | Blau | Homophily |
| 1970 | White | Block models, Vacancy Chains |
| 1970 | Granovetter | Weak ties |
| 1980 | Holland & Leinhardt | Exponential Random Graph Models |
| 1980 | Frank & Strauss | Markov dependency grfaphs |
| 1990 | Friedkin | Social Influence Network Theory |
| 1990 | Bonacich | Eigenvector centrality, Power centrality |
| 1990 | Putnam | Social capital |
| 1990 | Watts & Strogatz | Small world simulation |
| 2000 | Snijders & Huisman | Longitudinal network data |

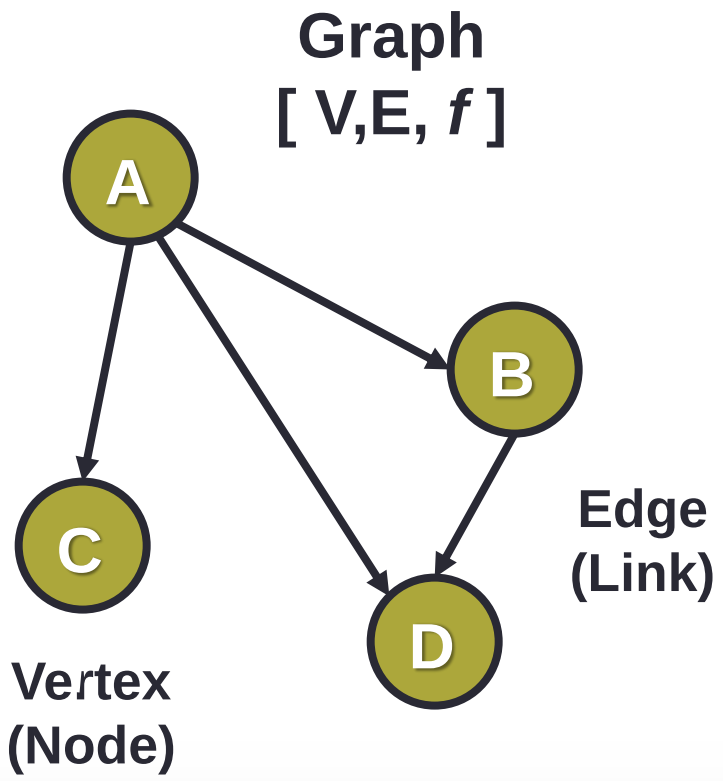
# Social Network Analysis

## *Key Elements*

**Vertices or Nodes**
*The "things"*

**Edges or Links**
*The "relationships"*

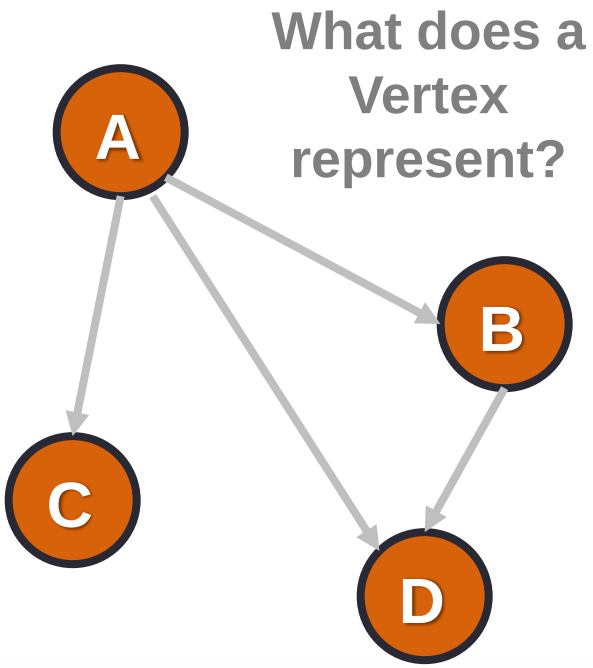
**Graph or Network**
*The set of vertices/nodes, edges/links and the relationship/function connecting them.*



# Social Network Analysis

## *Types of Nodes or Vertices*

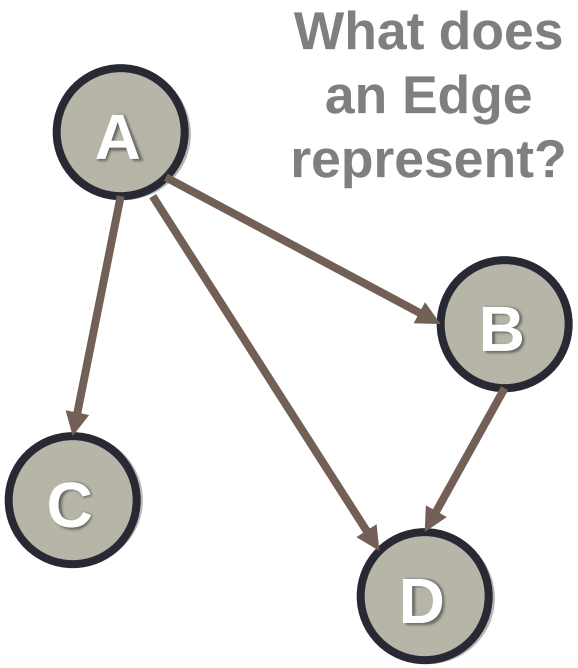
- Social Entities
  - People or social structures such as workgroups, teams, organizations, institutions, states, or even countries.
- Content
  - Web pages, keyword tags, or videos.
- Locations
  - Physical or virtual locations or events.
- Primary building blocks of social media
  - Friends in social networking sites, posts or authors in blogs, or pages in wikis.



# Social Network Analysis

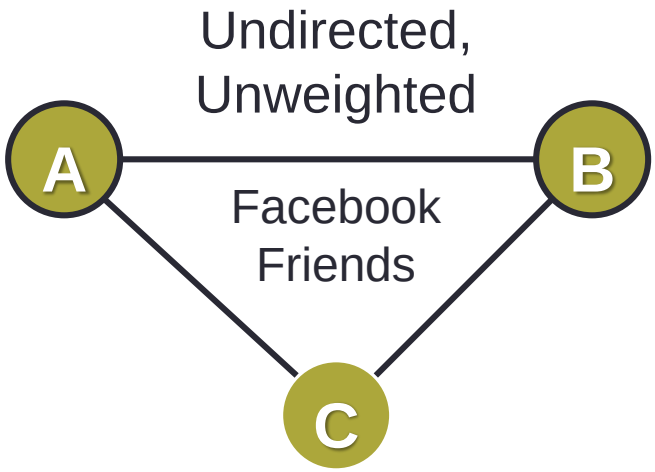
## *Types of Relations or Links*

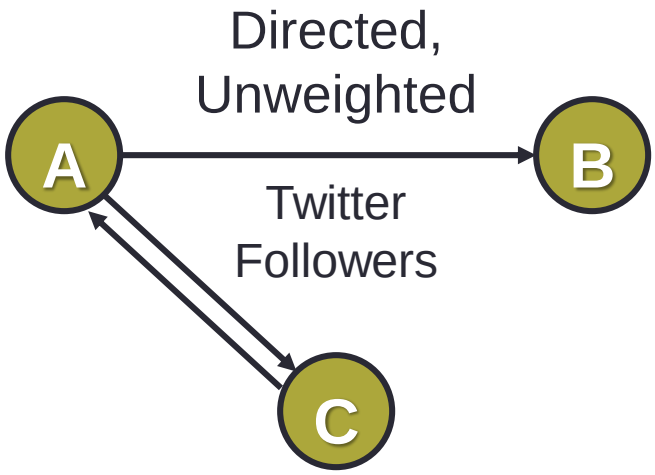
- Similarities
  - Location (same spatial and temporal space)
  - Participation (same club, same event, …)
  - Attributes (Age, gender, same attitudes, …)
- Relational Roles
  - Kinship (mother of, sibling of, …)
  - Other Roles (friend of, boss of, …)
- Relational Cognition
  - Affective (Likes, Hates…)
  - Perceptual (Knows, Knows of, …)
- Relational Events
  - Interactions (Sold to, talked to, helped, …)
  - Flows (Information, beliefs, money, …)

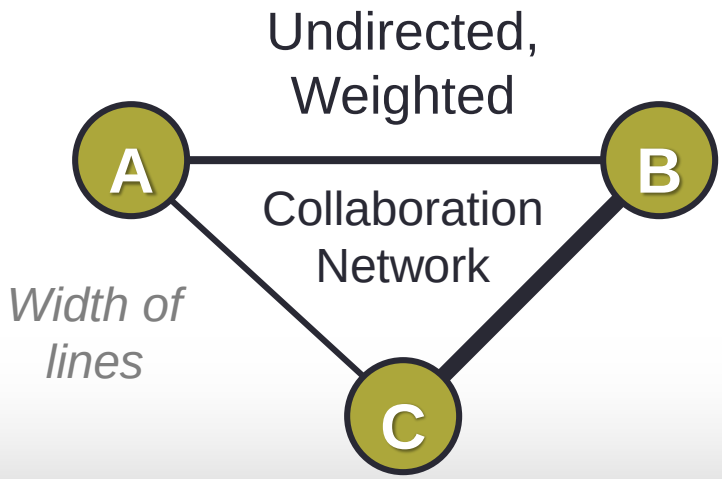


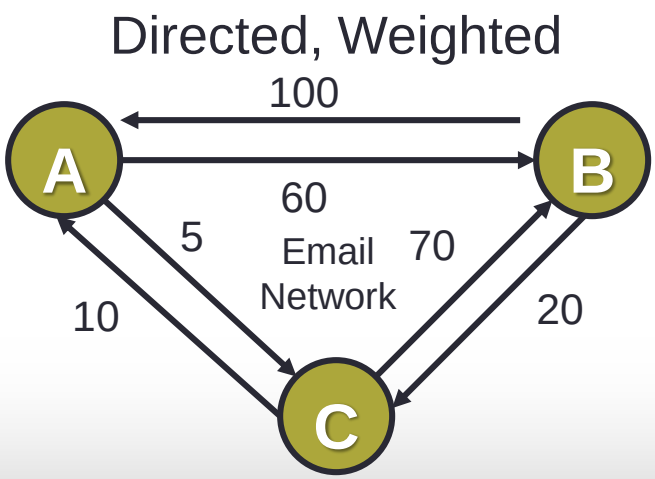
# Social Network Analysis

## *Types of Edges or Links*



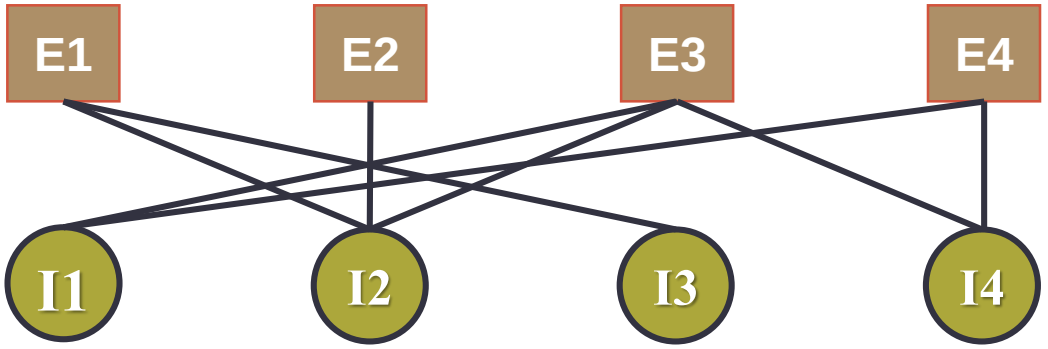



# Social Network Analysis

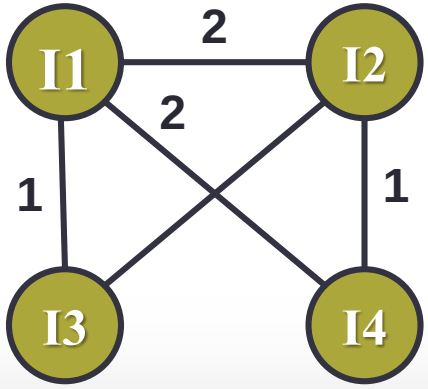
## *Bipartite or Bimodal Networks*

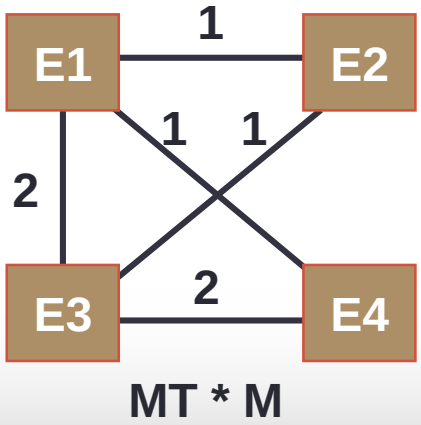
Bipartite or Bimodal Network
Linking individuals to events (participation, membership, topic, tag, post, …)

Examples:
Authors-to-papers (they authored)
Actors-to-Movies (they appeared in)
Users-to-Movies (they rated)
"Folded" networks:
Author collaboration networks
Movie co-rating networks




| | E1 | E2 | E3 | E4 |
|---|---|---|---|---|
| I1 | 1 | 0 | 1 | 1 |
| I2 | 1 | 1 | 1 | 0 |
| I3 | 1 | 0 | 0 | 0 |
| I4 | 0 | 0 | 1 | 1 |

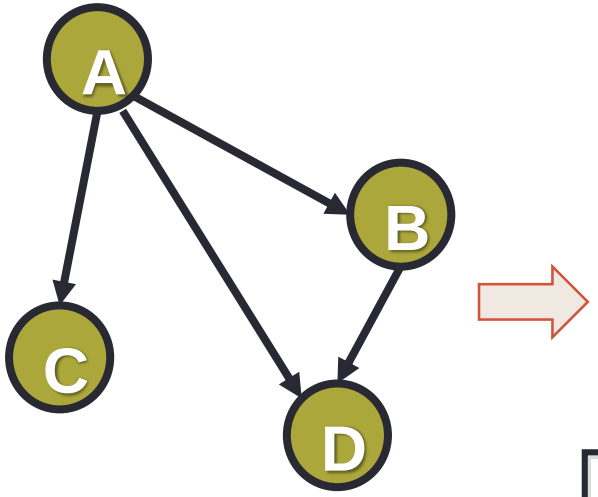


# Social Network Analysis

## *Types of Alternative Representations*

### Graph




### Edge List

| | |
|---|---|
| A | B |
| A | C |
| A | D |
| B | D |

### Adjacency Matrix

| | A | B | C | D |
|---|---|---|---|---|
| A | - | 1 | 1 | 1 |
| B | 0 | - | 0 | 1 |
| C | 0 | 0 | - | 0 |
| D | 0 | 0 | 0 | - |

### Adjacency List

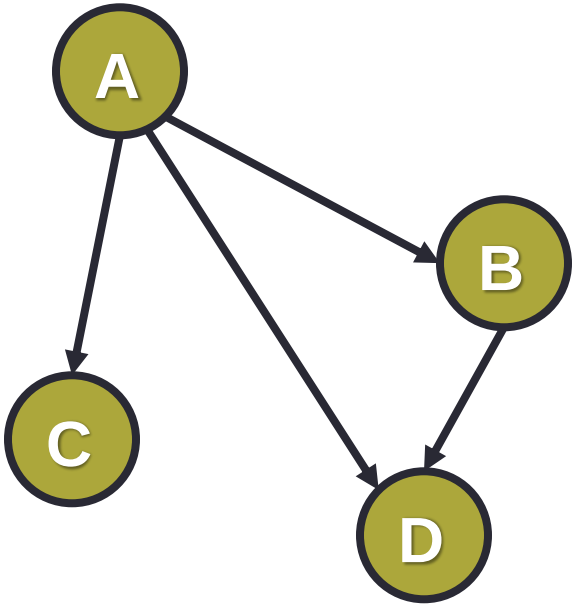
| | |
|---|---|
| A | B, C, D |
| C | D |

### XML

```
<Node>
    <Label> A </Label>
    <Connection> B </Connection>
    <Connection> C </Connection>
    <Connection> D </Connection>
</Node>
<Node>
    <Label> C </Label>
    <Connection> D </Connection>
</Node>
```

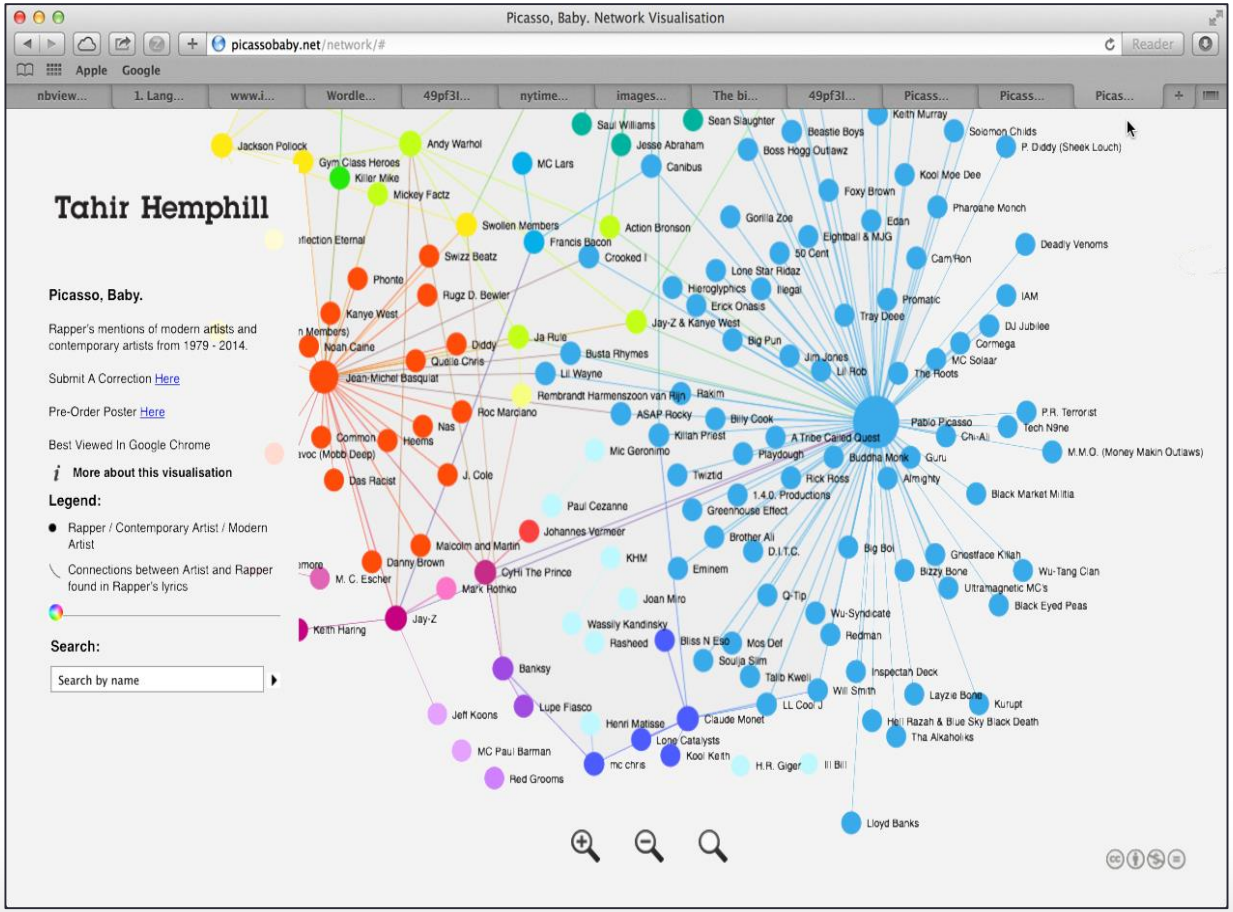
# Social Network Analysis

*How many variables can we represent?*



# Social Network Analysis

*How many variables in this network?*



# Social Network Analysis

*How many variables in this network?*

**Reducing the number of edges via pathfinder network scaling.**

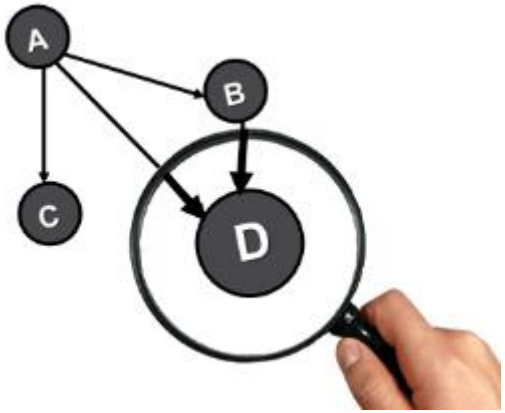

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

(*Mane & Börner, 2004*)

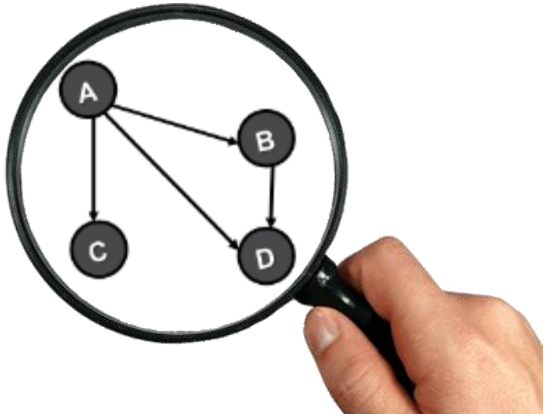
# Social Network Analysis

*Bifurcated measures*

## Local Measures




## Global Measures



# Social Network Analysis

## *Wide Variety of Metrics Available*

Measures Manager

Select Measures | Set Measure Inputs

Contains

| | Measure Title | Network Level | Node Level | Computation... | U |
|---|---|---|---|---|---|
| ☐ | Redundancy, Access | true | false | normal | fa |
| ☐ | Actual Workload | false | true | normal | fa |
| ☐ | Socio Economic Power, Agent | false | true | normal | fa |
| ☐ | Redundancy, Assignment | true | false | normal | fa |
| ☐ | Centrality, Authority | false | true | normal | tr |
| ☐ | Characteristic Path Length | true | false | normal | tr |
| ☐ | Speed, Average | true | false | normal | tr |
| ☑ | Centrality, Betweenness | false | true | normal | tr |
| ☑ | Network Centralization, Betweenness | true | false | normal | tr |
| ☐ | Centrality, Bonacich Power | false | true | normal | tr |
| ☐ | Capability | false | true | normal | tr |
| ☐ | Clique Count | false | true | normal | fa |
| ☑ | Centrality, Closeness | false | true | normal | tr |
| ☑ | Network Centralization, Closeness | true | false | normal | tr |
| ☑ | Density, Clustering Coefficient | true | true | normal | fa |
| ☐ | Cognitive Demand | false | true | normal | tr |
| ☐ | Cognitive Distinctiveness | false | true | normal | fa |
| ☐ | Cognitive Expertise | false | true | normal | fa |
| ☐ | Cognitive Resemblance | false | true | normal | fa |
| ☐ | Cognitive Similarity | false | true | normal | fa |
| ☐ | Breadth, Column | true | false | normal | fa |
| ☐ | Count, Column | true | false | normal | fa |
| ☐ | Centrality, Column Degree | false | true | normal | tr |

Select All | Select Visible

13 / 160 Selected, 160 / 160 Visible

OK | Close

# Social Network Analysis

## *Who is most influential*

| Measure | Definition | Interpretation | Reasoning |
|---|---|---|---|
| Degree | Number of edges or links. In degree- links in, Out-degree - links out | How connected is a node? How many people can this person reach directly? | Higher probability of receiving and transmitting information flows in the network. Nodes considered to have influence over larger number of nodes and or are capable of communicating quickly with the nodes in their neighborhood. |
| Betweenness | Number of times node or vertex lies on shortest path between 2 nodes divided by number of all the shortest paths | How important is a node in terms of connecting other nodes? How likely is this person to be the most direct route between two people in the network? | Degree to which node controls flow of information in the network. Those with high betweenness function as brokers. Useful where a network is vulnerable. |
| Closeness | 1 over the average distance between a node and every other node in the network | How easily can a node reach other nodes? How fast can this person reach everyone in the network? | Measure of reach. Importance based on how close a node is located with respect to every other node in the network. Nodes able to reach most or be reached by most all other nodes in the network through geodesic paths. |
| Eigenvector | Proporational to the sum of the eigenvector centralities of all the nodes directly connected to it. | How important, central, or influential are a node's neighbors? How well is this person connected to other well-connected people? | Evaluates a player's popularity. Identifies centers of large cliques. Node with more connections to higher scoring nodes is more important. |

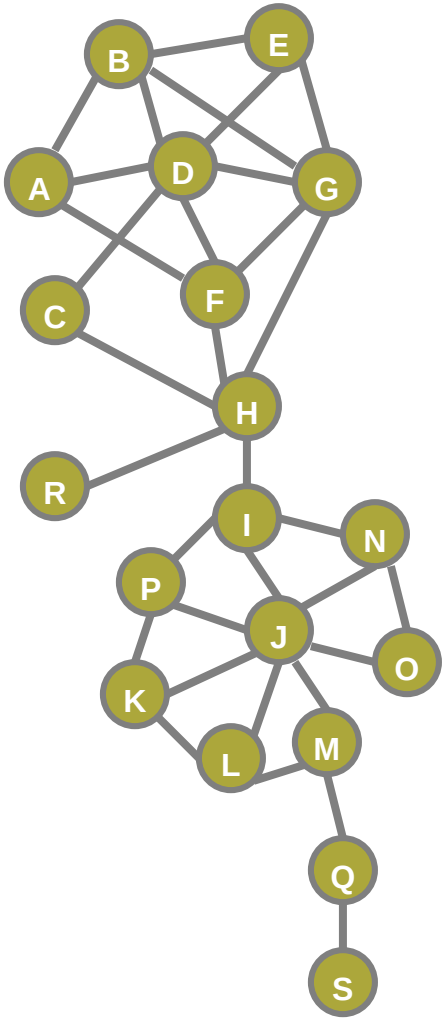
# Social Network Analysis

## *How well connected is the network?*

| Cohesion | Definition | Interpretation | Reasoning |
|---|---|---|---|
| Density | Ratio of the number of edges in the network over the total number of possible edges between all pairs of nodes | How well connected is the overall network? | Perfectly connected network is called a "clique" and has a density of 1. |
| Clustering | A node's *clustering coefficient* is the density of it's 1.5 degree egocentric network (ratio of connecting among ego's alters). For entire network it is the average of all the coefficients for the individual nodes. | What proportion of ego's alters are connected? More technically, how many nodes form triangular subgraphs with their adjacent nodes? | Measures certain aspects of "cliquishness." Proportion of you friends that are also friends with each other. Another way to measure is to determine (in a undirected) graph the ratio of the number of times that two links eminating from the same node are also linked. |
| Average Path Length (Distance) | Average number of edges or links between any two nodes (along the shortest path) | On average, how far apart are any two nodes? | This is synonymous with the "degrees of separation" in a network. |
| Diameter | Longest (shortest path) between any two nodes | At most, how long will it take to reach any node in the network? Sparse networks usually have greater diameters. | Measure of the reach of the network |
| Centralization | Normalize ratio of the sum of the variances of the centrality of each node from the most central node to the maximum sum possible | Indicates how unequal the distribution of centrality is in a network. | Measures how much variance there is in the distribution of centrality in a network. The measure applies to all forms of centrality. |

# Social Network Analysis

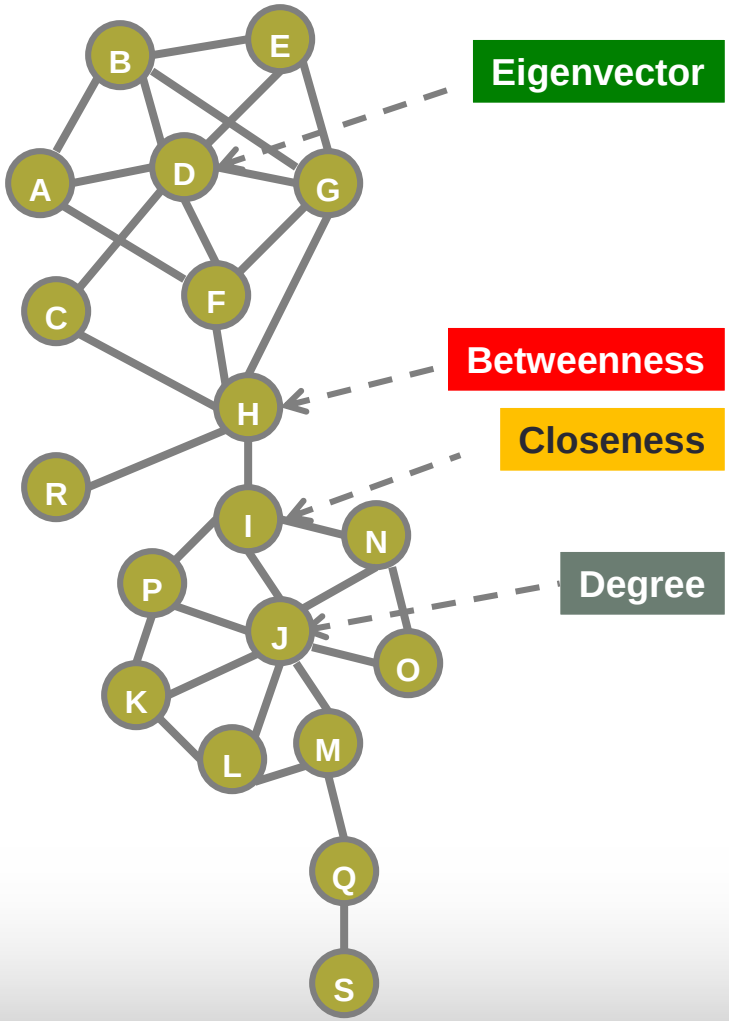
## *Generic Example*




- For each of the nodes what is it's
  - Degree Centrality
  - Betweenness Centrality
  - Closeness Centrality
  - Eigenvector Centrality
- For the entire network what is it's
  - Degree Centralization
  - Betweenness Centralization
  - Closeness Centralization
  - Eigenvector Centralization

# Social Network Analysis

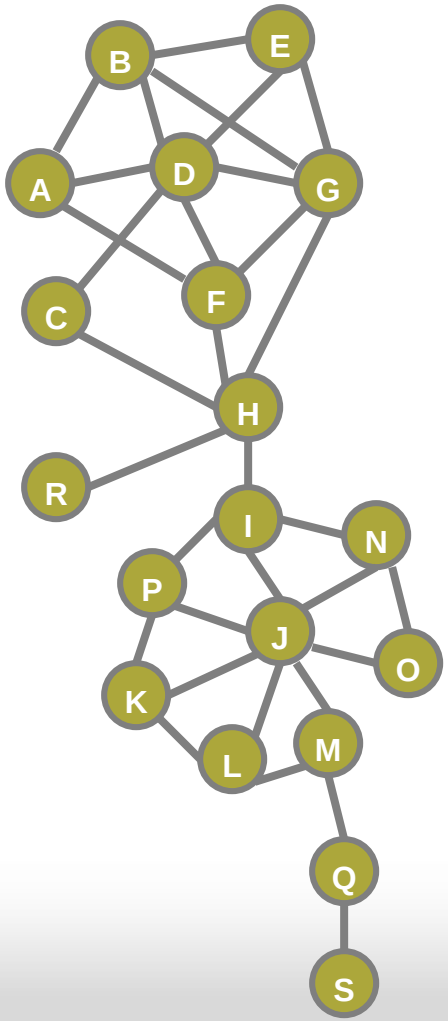
*Who's most influential?*




| Node | Degree | Normed Degree | Betweenness | Closeness | Eigen Vector |
|---|---|---|---|---|---|
| A | 3 | 0.17 | 0.00 | 0.29 | 0.29 |
| B | 4 | 0.22 | 0.01 | 0.30 | 0.36 |
| C | 2 | 0.11 | 0.03 | 0.35 | 0.18 |
| D | 6 | 0.33 | 0.04 | 0.31 | 0.46 |
| E | 3 | 0.17 | 0.00 | 0.29 | 0.30 |
| F | 4 | 0.22 | 0.11 | 0.36 | 0.35 |
| G | 5 | 0.28 | 0.19 | 0.37 | 0.43 |
| H | 5 | 0.28 | 0.58 | 0.45 | 0.28 |
| I | 4 | 0.22 | 0.53 | 0.46 | 0.13 |
| J | 7 | 0.39 | 0.43 | 0.43 | 0.12 |
| K | 3 | 0.17 | 0.00 | 0.32 | 0.06 |
| L | 3 | 0.17 | 0.01 | 0.33 | 0.05 |
| M | 3 | 0.17 | 0.21 | 0.33 | 0.04 |
| N | 3 | 0.17 | 0.03 | 0.38 | 0.07 |
| O | 2 | 0.11 | 0.00 | 0.31 | 0.05 |
| P | 3 | 0.17 | 0.03 | 0.38 | 0.08 |
| Q | 2 | 0.11 | 0.11 | 0.26 | 0.01 |
| R | 1 | 0.06 | 0.00 | 0.32 | 0.07 |
| S | 1 | 0.06 | 0.00 | 0.21 | 0.00 |

| Correlations | Degree | Betweenness | Closeness | Eigenvector |
|---|---|---|---|---|
| Degree | - | 0.57 | 0.59 | 0.59 |
| Betweenness | | - | 0.79 | 0.07 |
| Closeness | | | - | 0.13 |
| Eigenvector | | | | - |

# Social Network Analysis
# How cohesive is the network?




| Measure | Value |
|---|---|
| Network Size | 19 |
| Average Degree | 3.37 |
| Degree Centralization | 0.22 |
| Betweenness Centralization | 0.48 |
| Closeness Centralization | 0.27 |
| Eigenvector Centralization | 0.56 |
| Clustering Coefficient | 0.43 |
| Density | 0.19 |
| Average Distance | 3.06 |
| Diameter | 8 |
| Number of Unreachable Nodes | 0 |

| Node | Clustering |
|---|---|
| A | 0.67 |
| B | 0.67 |
| C | 0.00 |
| D | 0.40 |
| E | 1.00 |
| F | 0.50 |
| G | 0.50 |
| H | 0.10 |
| I | 0.33 |
| J | 0.29 |
| K | 0.67 |
| L | 0.67 |
| M | 0.33 |
| N | 0.67 |
| O | 1.00 |
| P | 0.67 |
| Q | 0.00 |
| R | NA |
| S | NA |

6

Frigyes Karinthy
1929

John Guare
1990

Stanley Milgram
1967

Duncan Watts
1998

**Six Degrees of Kevin Bacon (1994)**

# Social Network Analysis

*Who is the center of Hollywood?*

## The Center of the Hollywood Universe

Click on a name to see that person's table.

1. Harvey Keitel (2.848635)
2. Dennis Hopper (2.849329)
3. Robert De Niro (2.855810)
4. David Carradine (2.857729)
5. Martin Sheen (2.858291)
6. Udo Kier (2.859489)
7. Michael Madsen (I) (2.860010)
8. Donald Sutherland (I) (2.860447)
9. Michael Caine (I) (2.862189)
10. Eric Roberts (I) (2.867675)
11. Seymour Cassel (2.869415)
12. Malcolm McDowell (2.870208)
13. Max von Sydow (I) (2.872338)
14. Willem Dafoe (2.873805)
15. Samuel L. Jackson (2.873819)
16. Danny Trejo (2.876002)
17. John Hurt (2.878378)
18. Christopher Lee (I) (2.879217)
19. Harry Dean Stanton (2.880725)
20. Bruce Willis (2.886364)
21. Christopher Plummer (I) (2.886928)
22. John Malkovich (2.888575)
23. Morgan Freeman (I) (2.891003)
24. Christopher Walken (2.894212)
25. John Savage (I) (2.894873)

| Kevin Bacon Number | # of People |
|---|---|
| 0 | 1 |
| 1 | 2799 |
| 2 | 313045 |
| 3 | 1078865 |
| 4 | 276680 |
| 5 | 22296 |
| 6 | 2361 |
| 7 | 251 |
| 8 | 24 |

Total number of linkable actors: 1696322
Weighted total of linkable actors: 5099799
Average Kevin Bacon number: 3.006

| Kyra Sedgwick Number | # of People |
|---|---|
| 0 | 1 |
| 1 | 1353 |
| 2 | 229226 |
| 3 | 1083255 |
| 4 | 350117 |
| 5 | 29167 |
| 6 | 2845 |
| 7 | 331 |
| 8 | 27 |

Total number of linkable actors: 1696322
Weighted total of linkable actors: 5275476
Average Kyra Sedgwick number: 3.110

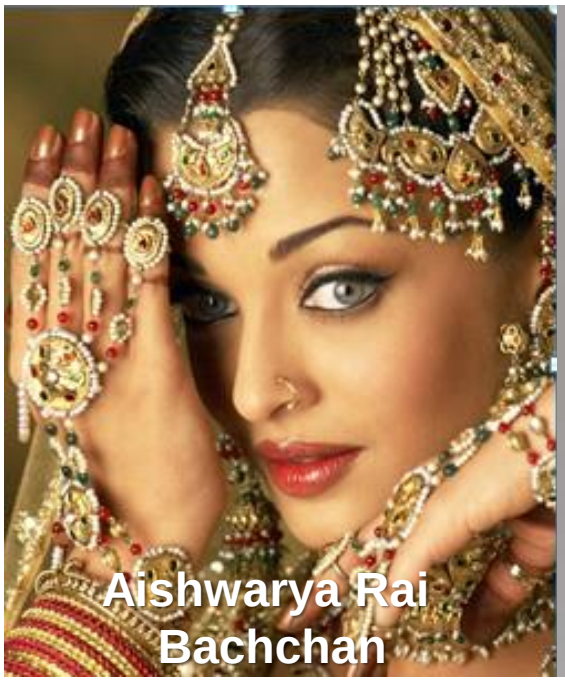
| Harvey Keitel (I) Number | # of People |
|---|---|
| 0 | 1 |
| 1 | 4128 |
| 2 | 454260 |
| 3 | 1051685 |
| 4 | 169704 |
| 5 | 14709 |
| 6 | 1679 |
| 7 | 141 |
| 8 | 15 |

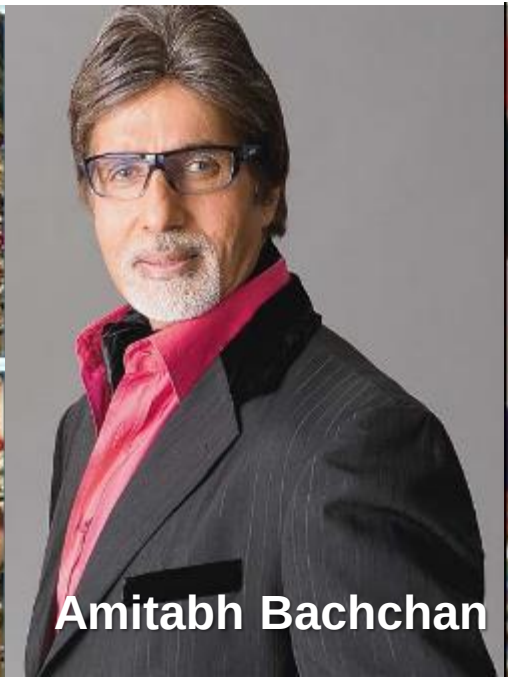
Total number of linkable actors: 1696322
Weighted total of linkable actors: 4831245
Average Harvey Keitel (I) number: 2.848

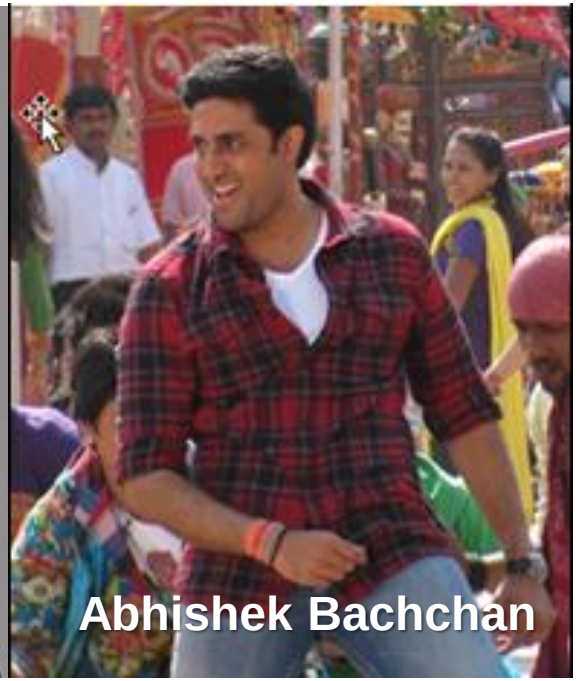
| John Savage (I) Number | # of People |
|---|---|
| 0 | 1 |
| 1 | 3766 |
| 2 | 408573 |
| 3 | 1073294 |
| 4 | 192718 |
| 5 | 16000 |
| 6 | 1764 |
| 7 | 195 |
| 8 | 11 |

Total number of linkable actors: 1696322
Weighted total of linkable actors: 4903703
Average John Savage (I) number: 2.891

# Social Network Analysis Bollywood?




जुदाई की छह डिग्री

judā'ī kī chaha ḍigrī

# Social Network Analysis

## *Source of Bollywood Data*

Amitabh Bachchan - Goo | List of Bollywood films - V

en.wikipedia.org/wiki/List_of_Bollywood_films#2010s

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

Print/export

Languages
हिन्दी
Português
తెలుగు

Edit links

# List of Bollywood films

From Wikipedia, the free encyclopedia

This article **does not cite any references or sources**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(September 2012)*

This is a list of films produced by the Bollywood film industry of Mumbai ordered by year and decade of release and also contains the top ten or forty superhit films of respective years as the case may be. Although "Bollywood" films are generally listed under the Hindi language, most are in mixed Hindi, Urdu and Punjabi and occasionally other languages. There is a range of mixtures from mostly Urdu to mostly Hindi to mostly Punjabi. Speakers of Hindi, Urdu, and Punjabi understand the mixed language usage of Bollywood thus extending the viewership to people all over the Indian subcontinent (throughout India and its neighboring countries). Here are some examples - Partly Hindi: *Om Shanti Om*, *Dhoom 2*, *No Entry* and *Kabhi Alvida Naa Kehna*, Partly Urdu: *Jodhaa Akbar*, *Fanaa*, *Saawariya* and *Kurbaan*, Partly Punjabi: *Singh Is Kinng*, *Jab We Met*, *Patiala House* and *Rab Ne Bana Di Jodi*. The film *Veer Zaara* is an equal mix of Hindi, Punjabi and Urdu.

**Contents** [hide]



*Alam Ara* (1931), the first Indian sound film

## 2010s [edit]

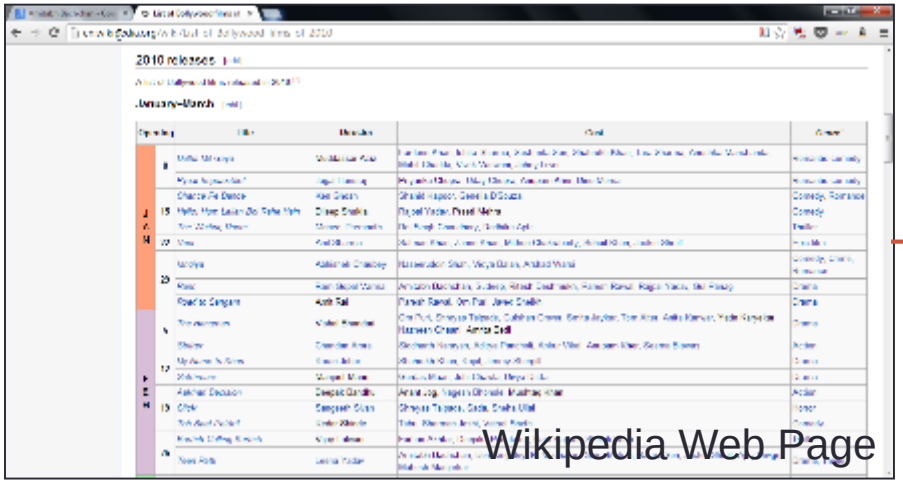
- List of Bollywood films of 2010
- List of Bollywood films of 2011
- List of Bollywood films of 2012
- List of Bollywood films of 2013
- List of Bollywood films of 2014

## 2000s [edit]

- List of Bollywood films of 2000

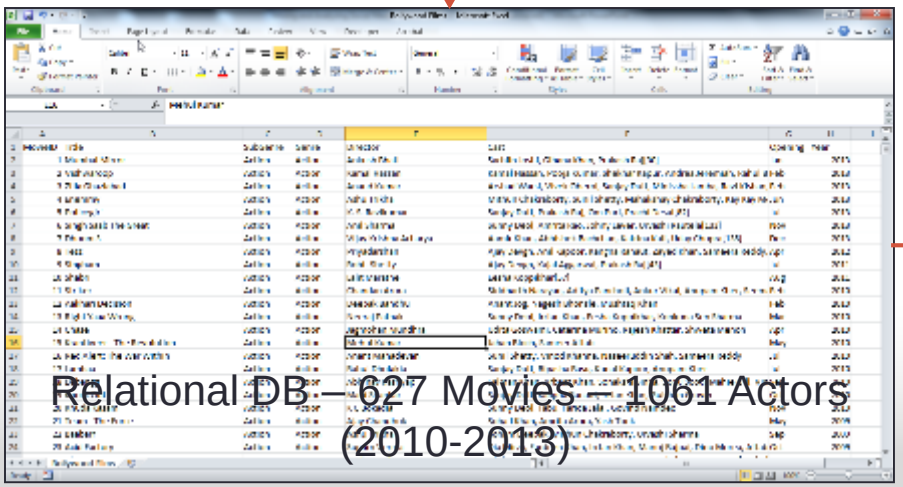
# Social Network Analysis

## *Source of Bollywood Data*



Wikipedia Web Page

HTML- Page Source



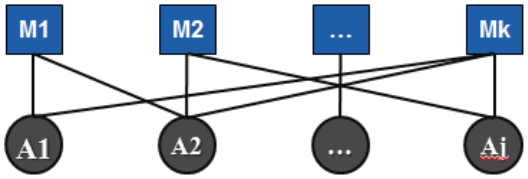
Relational DB – 627 Movies – 1061 Actors (2010-2013)

```
*Network bollywood.net [2-Mode]
*Vertices 1643 627
1 "Mumbai Mirror" 0.0000 0.0000 0.5000
2 "Vishwaroop" 0.0000 0.0000 0.5000
...
628 "A. K. Hangal" 0.0000 0.0000 0.5000
629 "Aamir Ali" 0.0000 0.0000 0.5000
...
*Arcs
*Edges
1 896 1
1 1220 1
...
627 856 1
627 1053 1
```

Pajek .NET File

# Social Network Data

## *Bipartite Network – Movies and Actors*



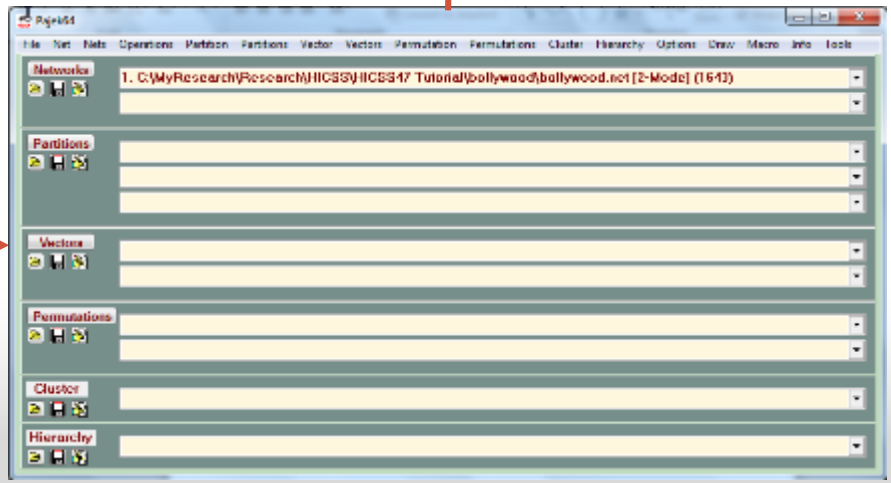

| | M1 | M2 | ... | Mk |
|---|---|---|---|---|
| A1 | 1 | 0 | | 1 |
| A2 | 1 | 1 | ... | 1 |
| ... | ... | ... | ... | ... |
| Aj | 0 | 1 | | 1 |

| | A1 | A2 | ... | Aj |
|---|---|---|---|---|
| A1 | - | 2 | ... | 1 |
| A2 | 2 | - | ... | 1 |
| ... | ... | ... | ... | ... |
| Aj | 1 | 1 | ... | - |

| | M1 | M2 | ... | Mk |
|---|---|---|---|---|
| M1 | - | 1 | ... | 2 |
| M2 | 1 | - | ... | 2 |
| ... | ... | ... | ... | ... |
| Mk | 2 | 2 | ... | - |

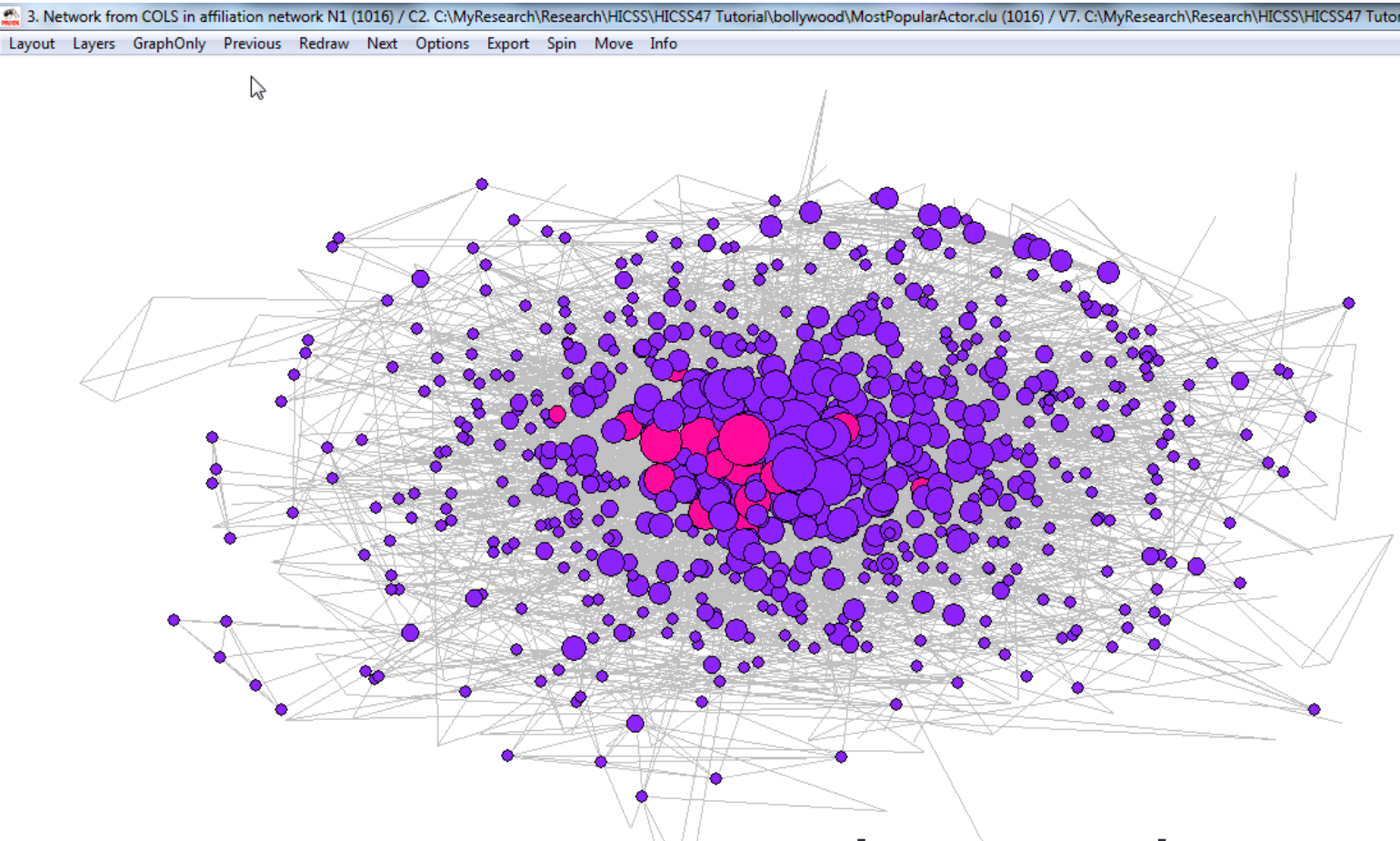
```
*Network bollywood.net [2-Mode]
*Vertices 1643 627
1 "Mumbai Mirror" 0.0000 0.0000 0.5000
2 "Vishwaroop" 0.0000 0.0000 0.5000
...
628 "A. K. Hangal" 0.0000 0.0000 0.5000
629 "Aamir Ali" 0.0000 0.0000 0.5000
...
*Arcs
*Edges
1 896 1
1 1220 1
...
627 856 1
627 1053 1
```

Pajek .NET File



# Social Network Analysis

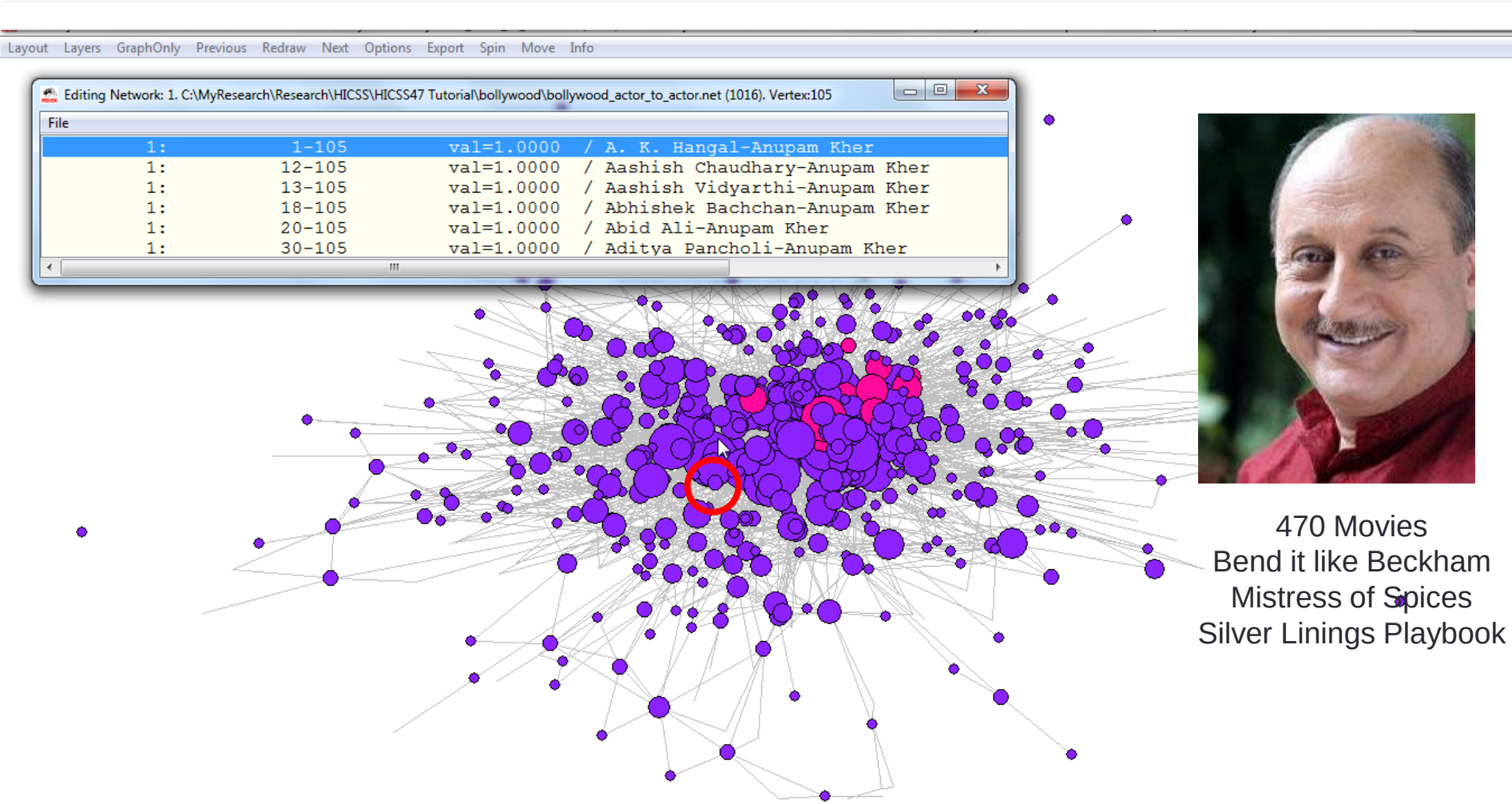
*And the answer is…*




Average Distance is 3.44

# Social Network Analysis

*and the center of Bollywood is*



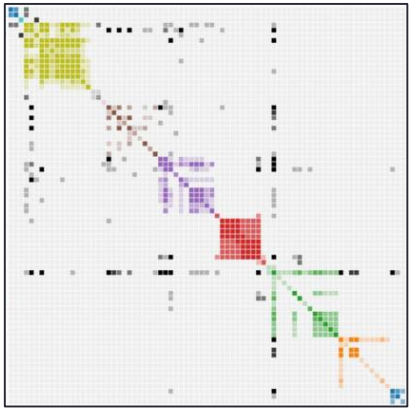



470 Movies
Bend it like Beckham
Mistress of Spices
Silver Linings Playbook

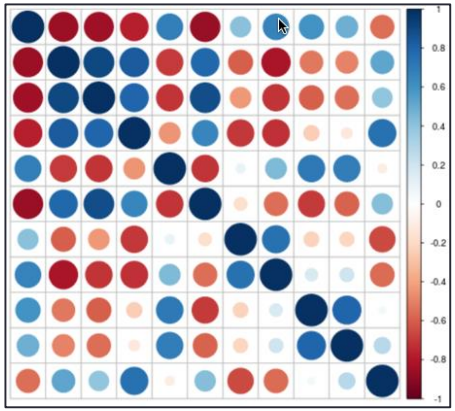
# Alternative Visualizations

*Once in matrix form …*

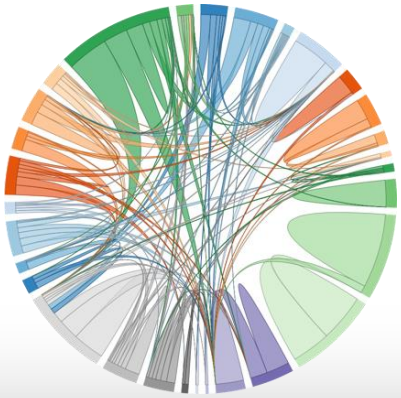
Co-occurrence



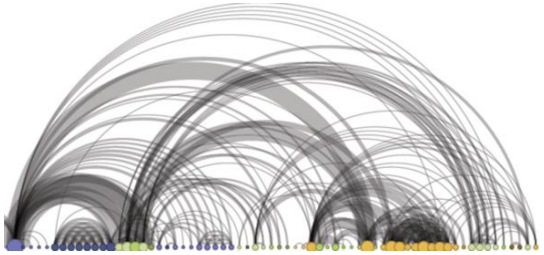
Corrplot




Chord



Arc



# Alternative Visualizations

## *Chord Diagram*