









| 순위 | 곡 | | 아티스트 |
|--------|--|---|------|
| 1 - |  |  SCH Big Data Engineering | 서민지 |
| 2 - |  |  Exploratory Data Analysis | 정선경 |
| 3 - |  |  Feat(woo ji young) | 한예찬 |



1. 데이터 소개 및 수집 방법
2. 노래 재생 시간과 가사 길이의 상관 분석
3. 시대에 따른 단어의 변화
4. 동시 출현 단어 네트워크 분석
5. 요약 및 결론



장르 | 발라드/댄스/팝

| 종합 | 발라드 | 댄스 | 팝 | 포크/인디

최신 곡 | 최신 앨범 | 주요 아티스트 | 장르소개

☐ ☒ 듣기 | ☐ + 재생목록에 추가 | ☐ 내 앨범에 담기 | ☐ 다운로드 | ☐ 전체 듣기 (재생목록 추가)

| 곡 | 아티스트 |
|--|-------------------|
| <input type="checkbox"/> 아파도 사랑이라서 | 손정수 |
| <input type="checkbox"/> 초능력 나의 그대 | Kei(러블리즈) |
| <input type="checkbox"/> 문과감성 | 덕인 |
| <input type="checkbox"/> 영하 오도 | 코넷 |
| <input type="checkbox"/> 계단 | ME NONE |
| <input type="checkbox"/> 헤어지는 중 - 편지 [주영스트] | 셀리아 김 (Celia Kim) |
| <input type="checkbox"/> 해줄 수 없는 일 - 거미 [주영스트] | 셀리아 김 (Celia Kim) |



아티스트 손정수, 셀리아 김 (Celia Kim)

참여 정보 **보컬** 손정수, 셀리아 김 (Celia Kim) | **작곡** 필승불패, LACONIC

| **작사** 필승불패 전체 보기 ▾

앨범 아파도 사랑이라서

재생시간 03:54

고음질 FLAC 16bit

가사

아파도 아파도 난 사랑이라서
 죽어도 놓을 수가 없는 사랑아
 이대로 포기 못해 어떻게 널 보내니
 생각만 해도 난 아파
 처량하구나 우리 사랑은
 보잘 것이 없나보다
 치이고 채여 쓰러진대도
 너와 나는 변할게 없잖아
 아파도 아파도 난 사랑이라서
 죽어도 놓을 수가 없는 사랑아
 이대로 포기 못해 어떻게 널 보내니
 생각만 해도 난 아파
 거스를 수가 없나보구나



```
library(httr)
library(rvest)
library(stringr)

bal_url<-"https://music.bugs.co.kr/genre/kpop/ballad/ballad?tabtype=2" #발라드 url
hip_url<-"https://music.bugs.co.kr/genre/kpop/rnh/hnp?tabtype=2" #힙합 url

bal_ly_url<-GET(bal_url) %>% read_html() %>% html_nodes("a.trackInfo") %>% html_attr("href")
hip_ly_url<-GET(hip_url) %>% read_html() %>% html_nodes("a.trackInfo") %>% html_attr("href")
```

```
> head(bal_ly_url)
[1] "https://music.bugs.co.kr/track/31615201?wl_ref=list_tr_08"
[2] "https://music.bugs.co.kr/track/31615045?wl_ref=list_tr_08"
[3] "https://music.bugs.co.kr/track/31615034?wl_ref=list_tr_08"
[4] "https://music.bugs.co.kr/track/31613858?wl_ref=list_tr_08"
[5] "https://music.bugs.co.kr/track/31612748?wl_ref=list_tr_08"
[6] "https://music.bugs.co.kr/track/5661676?wl_ref=list_tr_08"
```

```
for(url in bal_ly_url){

  name<-GET(url) %>% read_html() %>% html_nodes("#container > header > div > h1") %>%
    html_text() %>% str_remove_all("\r\n\t") %>% str_remove_all("\t")

  song<-GET(url) %>% read_html() %>%html_nodes("table.info tbody tr time") %>% html_text()

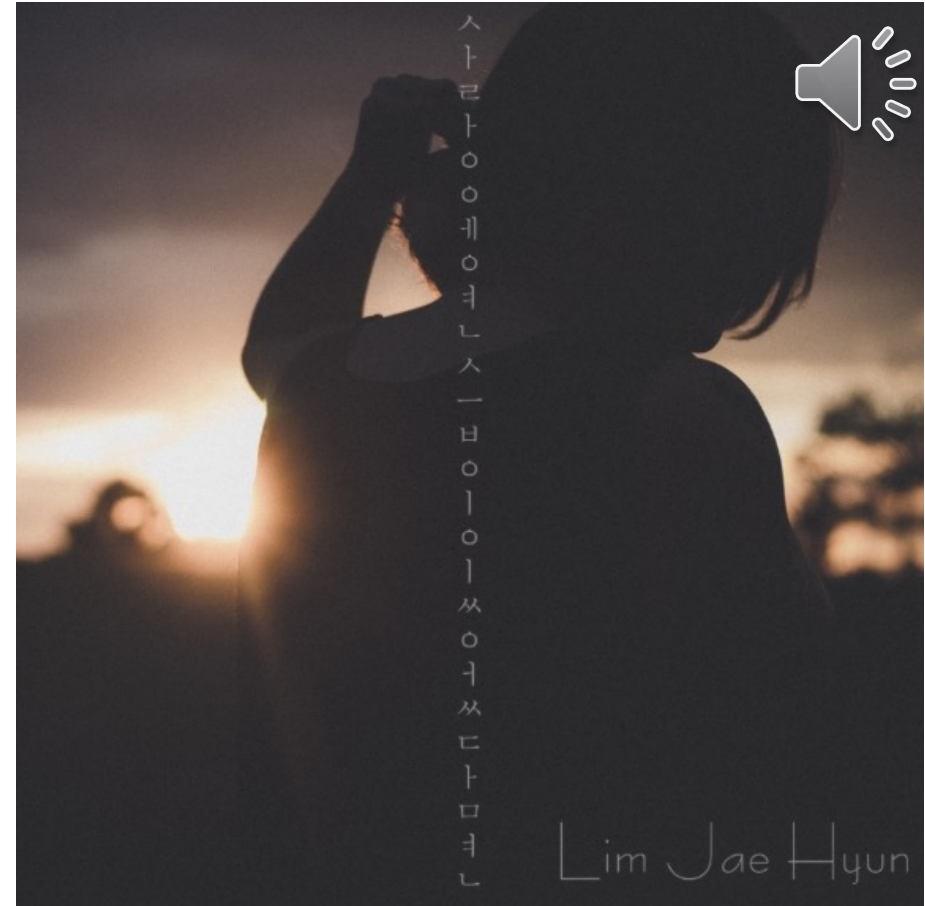
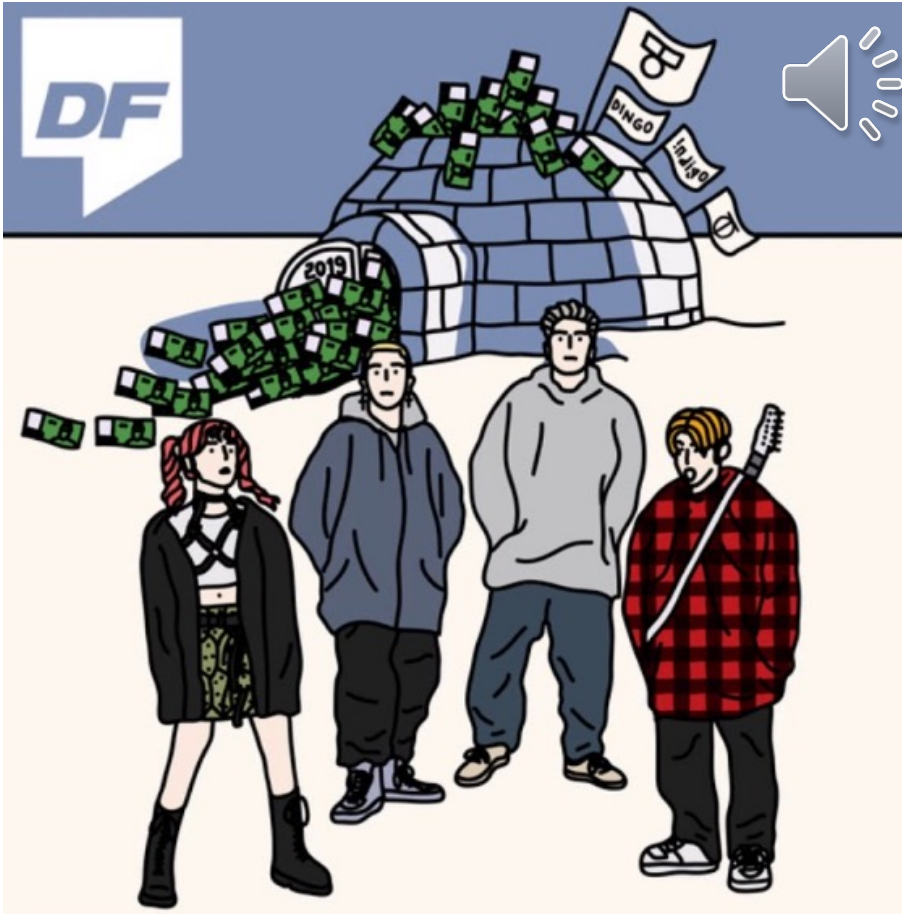
  ly<-GET(url) %>% read_html() %>%html_nodes("div.lyricsContainer xmp") %>% html_text() %>%
    str_replace_all("\r\n","") %>% str_length(.) %>% paste0("o")

  ballad<-rbind(ballad,as.data.frame(t(c(name,song,ly))))
}
```

URL 추출



노래가사,
노래 재생 시간,
가사 길이 등 추출





```
names(ballad)<-c("name","song_len","ly_len")
names(hiphop)<-c("name","song_len","ly_len")

ballad<-ballad %>% .[str_length($.ly_len)>1,]
hiphop<-hiphop %>% .[str_length($.ly_len)>1,]

ballad$ly_len<-str_remove(ballad$ly_len,"o") %>% as.numeric()
hiphop$ly_len<-str_remove(hiphop$ly_len,"o") %>% as.numeric()

bal_pos<-as.POSIXlt.factor(ballad$song_len,format="%M:%S")
ballad$song_len<-bal_pos$min*60+bal_pos$sec

hip_pos<-as.POSIXlt.factor(hiphop$song_len,format="%M:%S")
hiphop$song_len<-hip_pos$min*60+hip_pos$sec
```

데이터 전처리 및 형 변환

| | name | song_len | ly_len |
|-----|---|----------|--------|
| 1: | 아퍼 (Feat. Kid Milli, Lil tachi, 김승민, NO:EL, C JAMM) | 315 | 2165 |
| 2: | 위자드 | 241 | 1155 |
| 3: | How you feel (Feat. Karacin JR.) (Prod. AKIKAT) | 196 | 1339 |
| 4: | 페인트 (Feat. Rose de Penny) | 192 | 803 |
| 5: | 개새끼 (Feat. Sycho) | 201 | 1183 |
| 6: | 서울의 달 | 214 | 1148 |
| 7: | 취 | 193 | 1080 |
| 8: | 놀자 | 167 | 656 |
| 9: | These days | 250 | 1688 |
| 10: | 락앤롤 베이비 (Feat. G2, 창모) | 251 | 1683 |
| 11: | Enterprizin' | 240 | 756 |
| 12: | 다이어트 안하면 다이어 | 129 | 519 |
| 13: | color lens (feat. BLUEberii) | 203 | 1966 |
| 14: | 채워 | 160 | 699 |
| 15: | Fool (feat. Bradystreet) | 213 | 1444 |
| 16: | Diamonds Eternal (Feat. Jeremy Quest, Kinshaa wish) | 311 | 2032 |
| 17: | Atlantis | 204 | 1130 |
| 18: | zoom (dani REMIX) | 167 | 666 |
| 19: | BAND | 245 | 2572 |
| 20: | WMB (where's My Boy) | 250 | 1345 |

| | name | song_len | ly_len |
|-----|-------------------|----------|--------|
| 1: | 늦바람 | 279 | 498 |
| 2: | 이별지화 | 263 | 576 |
| 3: | starlight | 232 | 310 |
| 4: | 너는 내게 달 | 251 | 438 |
| 5: | 나 사랑이 잘 안돼서 | 233 | 443 |
| 6: | 좋겠다 | 152 | 269 |
| 7: | 이걸 사랑이라고 불러도 괜찮겠조 | 285 | 327 |
| 8: | 어쩌면 우리 | 215 | 364 |
| 9: | 너란 바다 | 205 | 620 |
| 10: | In Your Eyes | 172 | 632 |
| 11: | 추억이라 쓰고, 사랑이라 읽자 | 286 | 536 |
| 12: | 호수공원 | 176 | 390 |
| 13: | 너를 내일 만나면 | 254 | 319 |
| 14: | 벚꽃이 진다 해도 | 171 | 342 |
| 15: | 그대에게 가는 길 | 236 | 378 |
| 16: | 마음별 | 281 | 372 |
| 17: | 옛날동화 | 187 | 422 |
| 18: | 꿈 | 233 | 383 |
| 19: | 그대라서 | 217 | 387 |
| 20: | 솔직하게 말해서 나 | 200 | 420 |



```
> cor.test(ballad$song_len,ballad$ly_len)

Pearson's product-moment correlation

data: ballad$song_len and ballad$ly_len
t = 0.51237, df = 23, p-value = 0.6133
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3015567  0.4811659
sample estimates:
cor
0.1062322
```

```
> cor.test(hiphop$song_len,hiphop$ly_len)

Pearson's product-moment correlation

data: hiphop$song_len and hiphop$ly_len
t = 1.67, df = 44, p-value = 0.102
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.0496797  0.4990669
sample estimates:
cor
0.2441393
```

노래의 재생 시간과 가사 길이 간의 상관성은 발라드(0.1062), 힙합(0.2441)으로 두 장르가 약한 양의 상관 관계를 가진다고 생각 할 수 있습니다.

더 정확한 검증을 위해 cor.test함수를 사용하면 유의 수준 0.05에서 유의 확률 값이 발라드(0.6133), 힙합(0.102)로 두 장르 모두 1종 오류를 범하지 않아 귀무가설을 채택하게 됩니다.

따라서 노래의 재생 시간과 가사 길이 간의 상관성은 없다고 볼 수 있습니다.

귀무가설 : 상관관계가 없다
대립가설 : 상관관계가 있다



2000년대



1980년대



1990년대



1960년대



1970년대



1940년대



1950년대



2000년대



1940년대



```

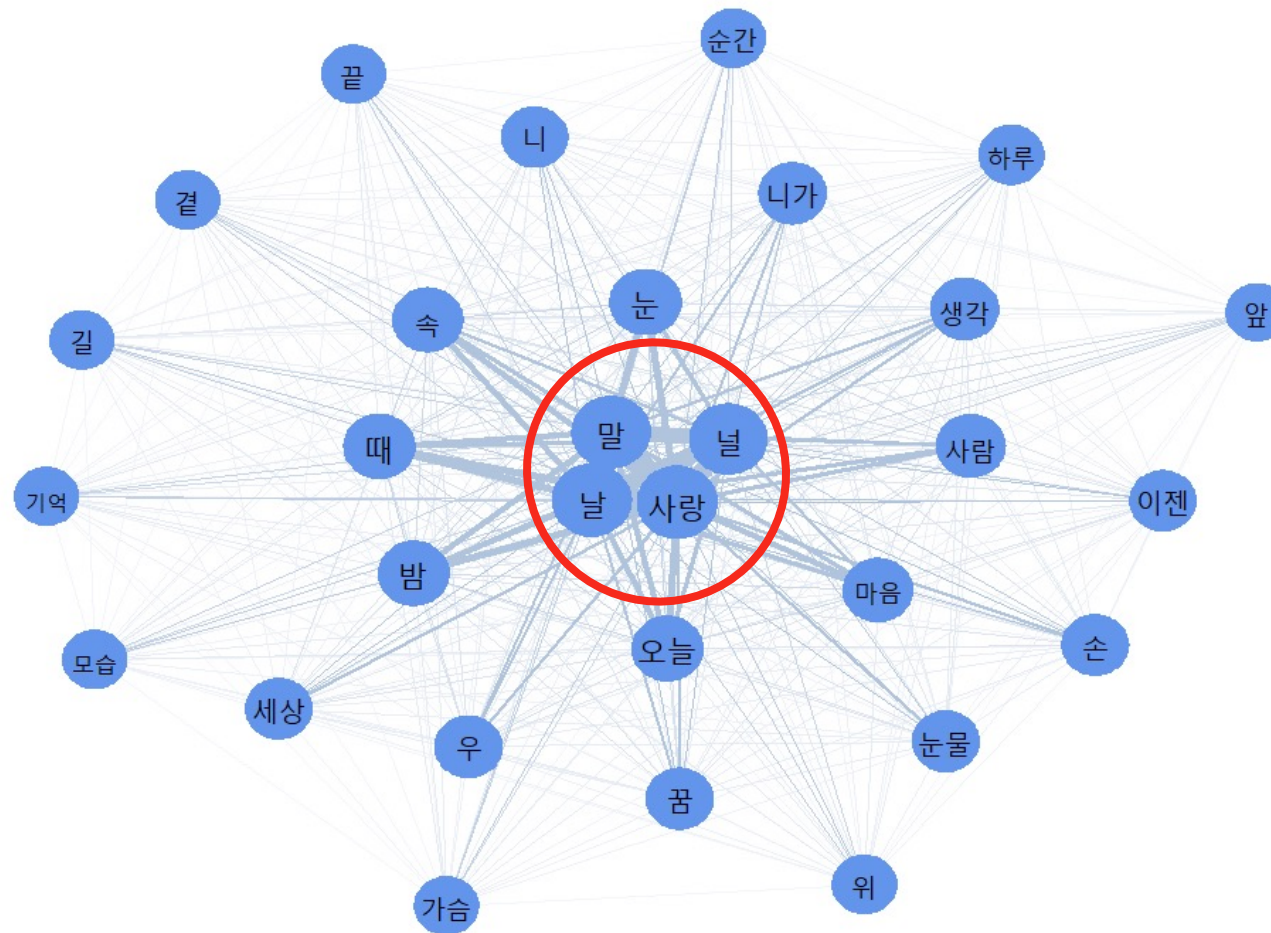
tt<-paste(unlist(SimplePos22(season12)))
alldta<-str_match_all(tt,"[가-힣]+/[N][C]|[가-힣]+/[N][Q]+")%>%unlist()
N<-str_replace_all(alldta,"/[N][C]", "")%>%
  str_replace_all("/[N][Q]", "")%>%unlist()
head(N)
alldta2<-str_match_all(tt,"[가-힣]+/[P][V]+|[가-힣]+/[P][X]+|[가-힣]+/[P][A]+|[가-힣]+/[M][A]+")%>%unlist()
PNM<-str_replace_all(alldta2,"/[P][V]", "")%>%
  str_replace_all("/[P][A]", "")%>%
  str_replace_all("/[M][A]", "")%>%
  str_replace_all("/[P][X]", "")%>%unlist()
DtaCorpusNC<-Corpus(VectorSource(N))
myTdmNC<-TermDocumentMatrix(DtaCorpusNC,control = list(wordLength=c(4,10),
                                                         removePunctuation=T,removeNumbers=T,weighting=weightBin))

findFreqTerms(myTdmNC,lowfreq = 10)
mtNC<-as.matrix(myTdmNC)
mtrowNC<-rowSums(mtNC) ##각 단어 별 합계를 구함
mtNC.order<-mtrowNC[order(mtrowNC, decreasing=T)] #
freq.wordsNC<-mtNC.order[61:90]
freq.wordsNC<-as.matrix(freq.wordsNC)
co.matrix <- freq.wordsNC %*% t(freq.wordsNC)
qgraph(co.matrix, labels=rownames(co.matrix),
       diag=FALSE, layout='spring', threshold=3,border.color="Cornflower Blue",color="Cornflower Blue",
       vsize=log(diag(co.matrix)) *0.4,edge.color="Light Steel Blue",label.cex=1.5)

```

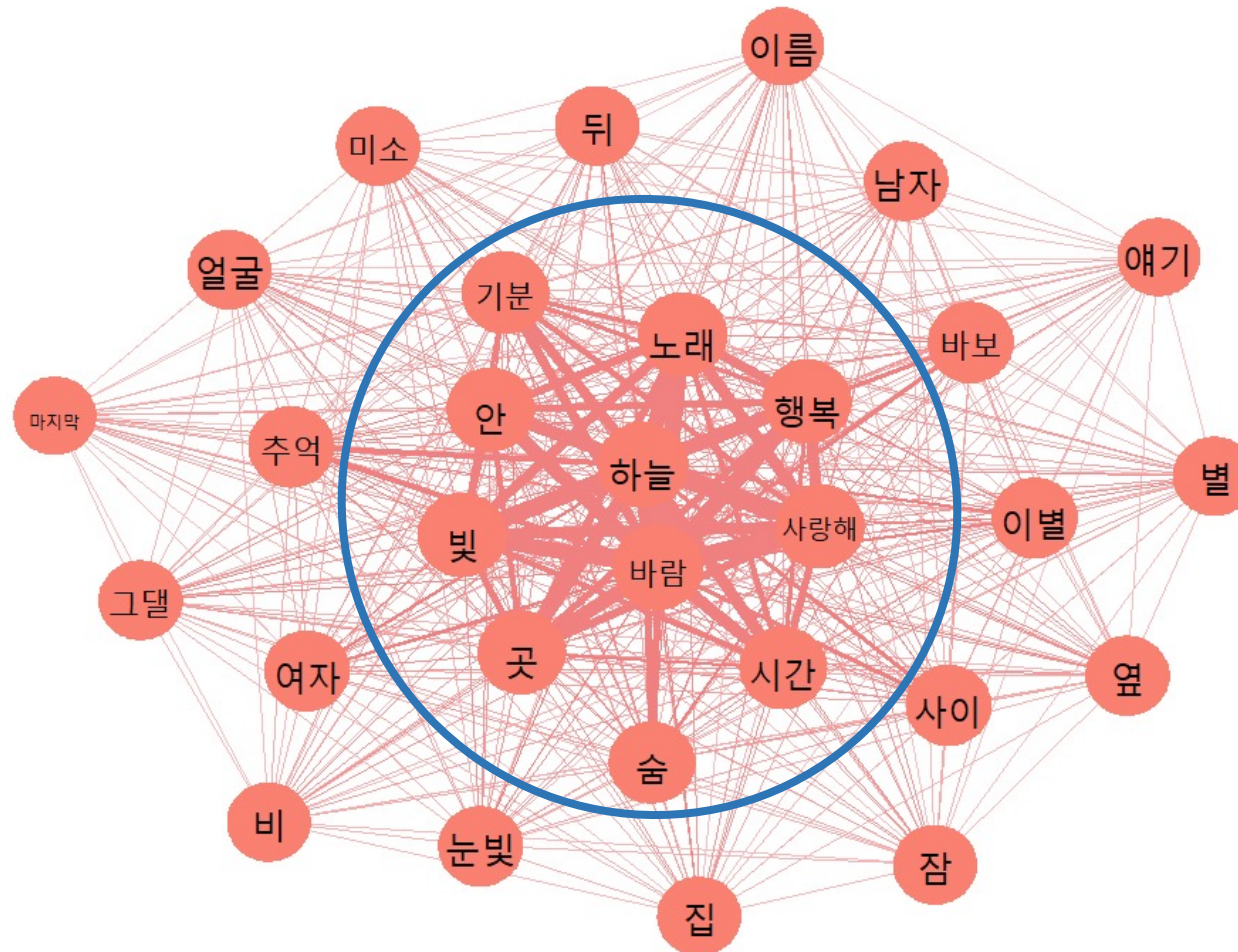


Top[1-30]



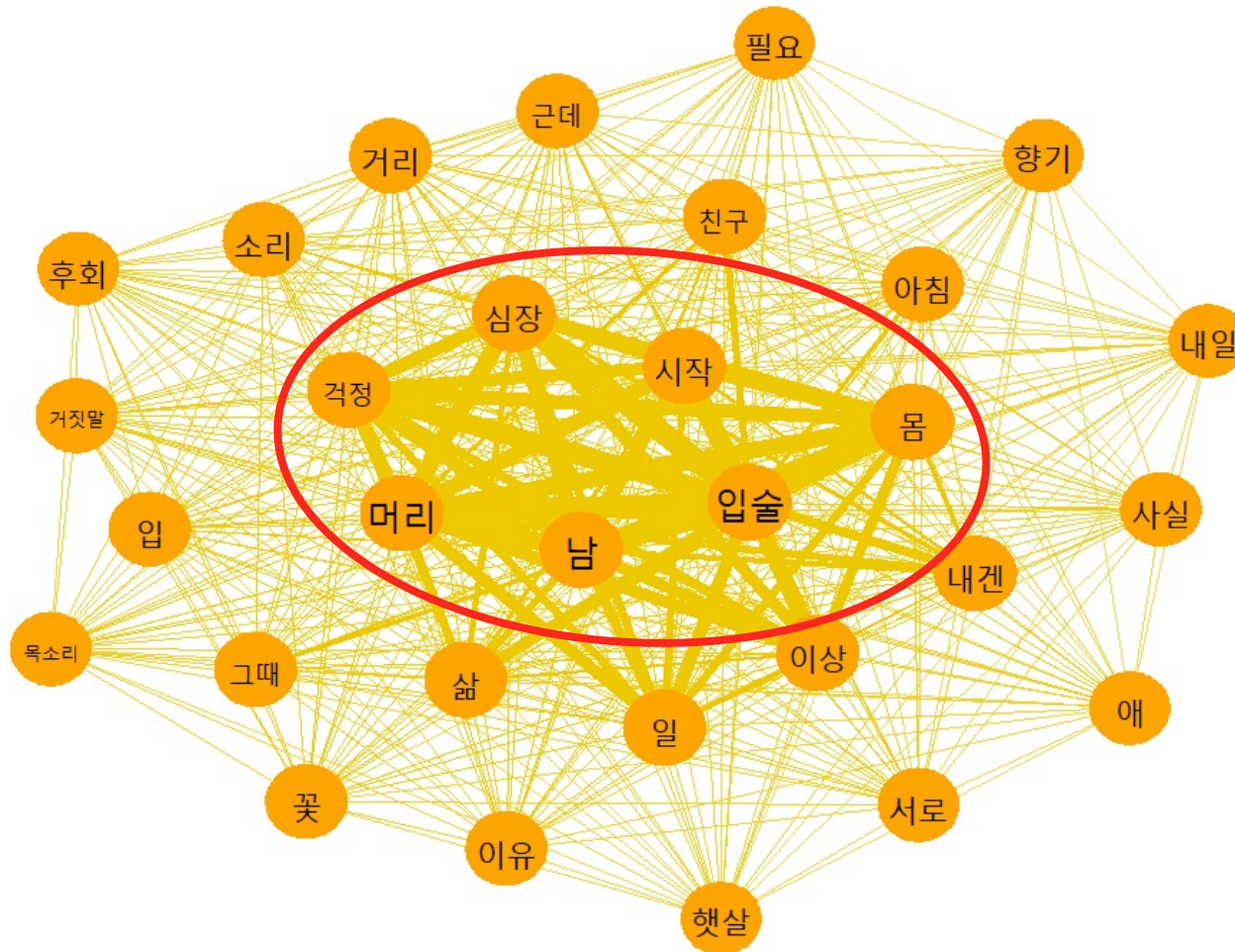


Top[31-60]





Top[61-90]





- 노래 재생 시간과 가사의 길이는 상관성이 없다.
(BUT, 장르별 차이는 존재한다)
- 4,50년대에는 고전 단어를 많이 사용했고,
90년대와 00년대에는 외국어와 외래어가 자주 사용되었다.
- 동시 출현 단어의 특정 조합을 발견할 수 있었다.