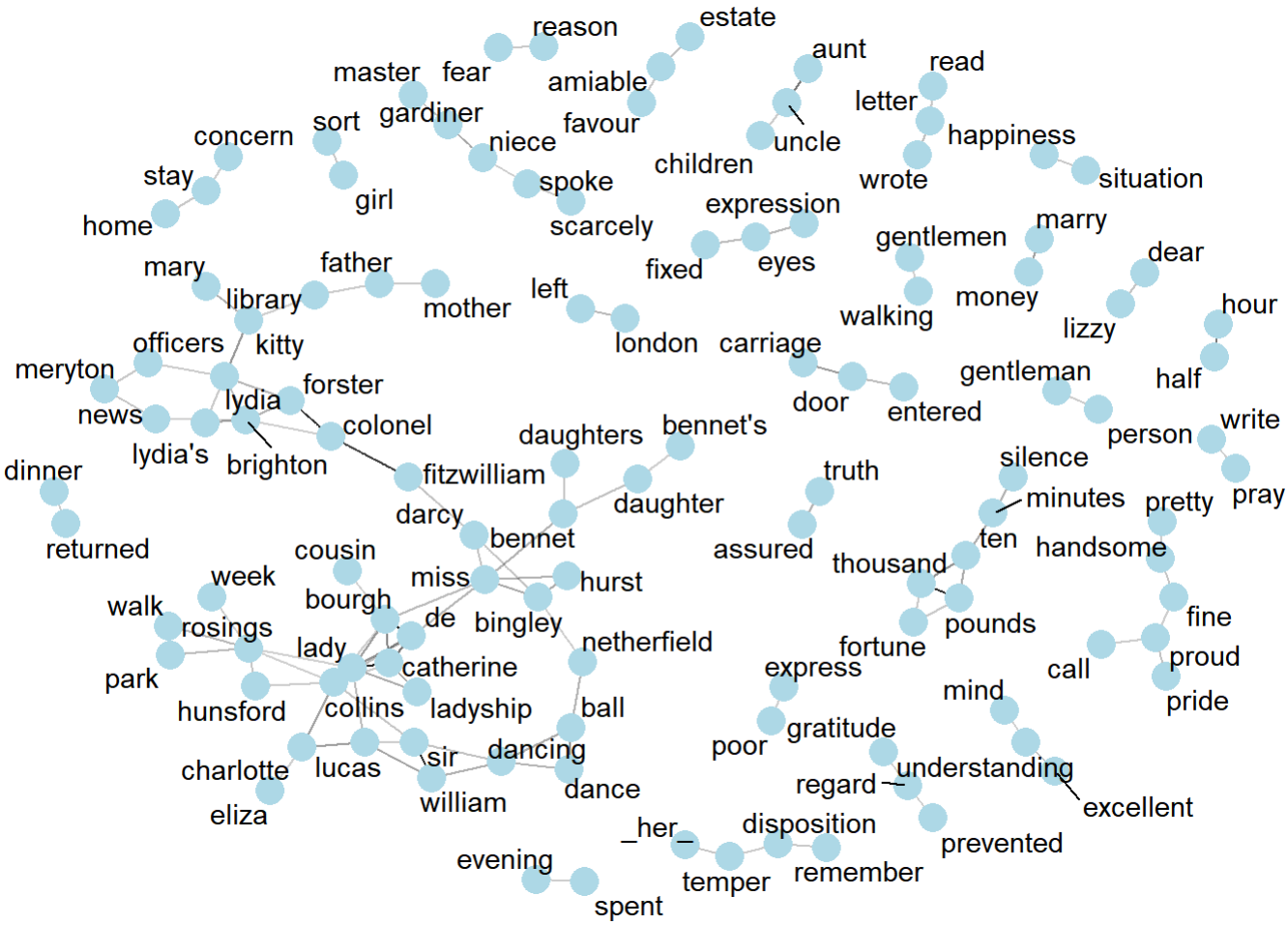


- 바이그램 분석과 달리 여기서의 관계는 방향성을 띄고있는 것이 아니라 상호관계적임을 주의해야한다.
- “colonel”이나 “fitzwilliam”과 같이 바이그램 짝 중에 거의 대부분을 차지하는 이름이나 제목의 짝이 흔하다는 점을 확인할 수 있고, 서로 인접해서 출현하는 단어의 조합도 확인할 수 있다.

Summary

-이번 장에서는 개별 단어를 분석하여 문장이나 책 한권에 적용하는 것이 아닌 단어간의 관계 및 연결을 탐색하는 방법들을 살펴보았다.

-관계를 보는 방법으로는 엔그램이 포함될 수 있으며, 이는 어떤 단어가 다른 단어 뒤에 나타나는 경향이 있는지와, 서로 인접한 곳에 출현하는 단어 간의 상관을 볼 수 있게 하는 방법임을 확인할 수 있었다.

-이러한 관계를 연결망으로 시각화하여 확인할 수 있었으며, 이러한 연결망 시각화는 관계에 대해 관심있게 볼때에 쓰기 적절한 도구임을 알 수 있었다.

igraph option

```
library(tidygraph)
```

```
## Warning: package 'tidygraph' was built under R version 4.0.4
```

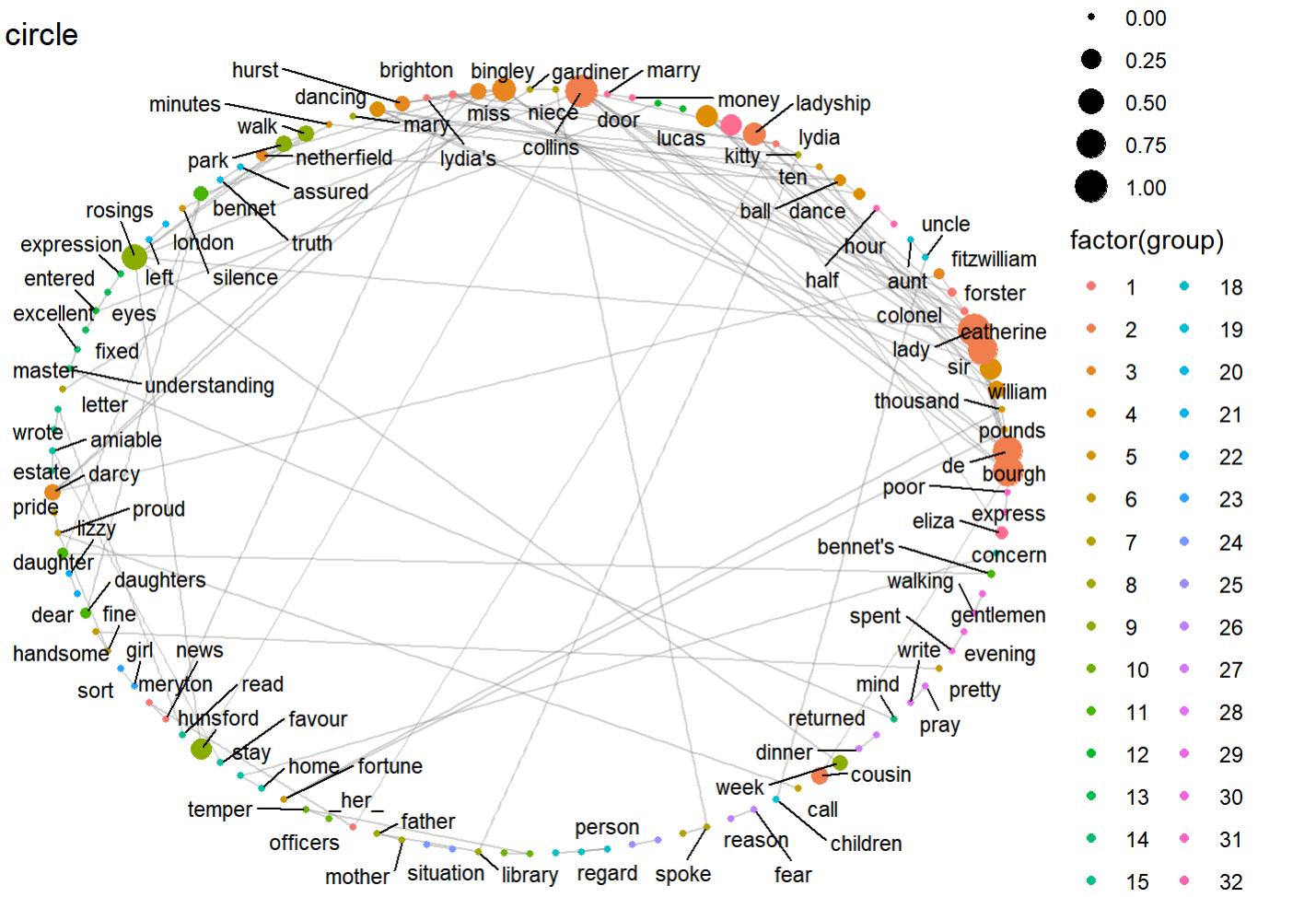
```
set.seed(2021)
igraph_layouts <- c('circle', 'graphopt', 'grid', 'mds', 'nicely', 'fr')

network.btw.data <- word_cors %>%
  filter(correlation > .15) %>%
  graph_from_data_frame() %>%
  as_tbl_graph(directed=F) %>%  # 방향성이 있으면 TRUE, 하지만 여기서는 방향성 없이 보고싶음.
  mutate(eigen = centrality_eigen(), # 중심성 계산.
         group = group_infomap())  # 그룹화 시키는 과정.
```

- `layout="circle"` : 원형의 network plot을 그림.
- `layout="graphopt"` : 반복수행으로 최적의 그림을 표현.
- `layout="grid"` : 직사각형의 network plot을 그림.
- `layout="mds"` : 거리행렬에 대해 Multidimensional scaling을 실시한 후, 그림으로 표현.
- `layout="nicely"` : simple algorithm에 대해 적절한 형태의 그림으로 표현.
- `layout="fr"` : Fruchterman and Reingold의 force-directed layout algorithm.

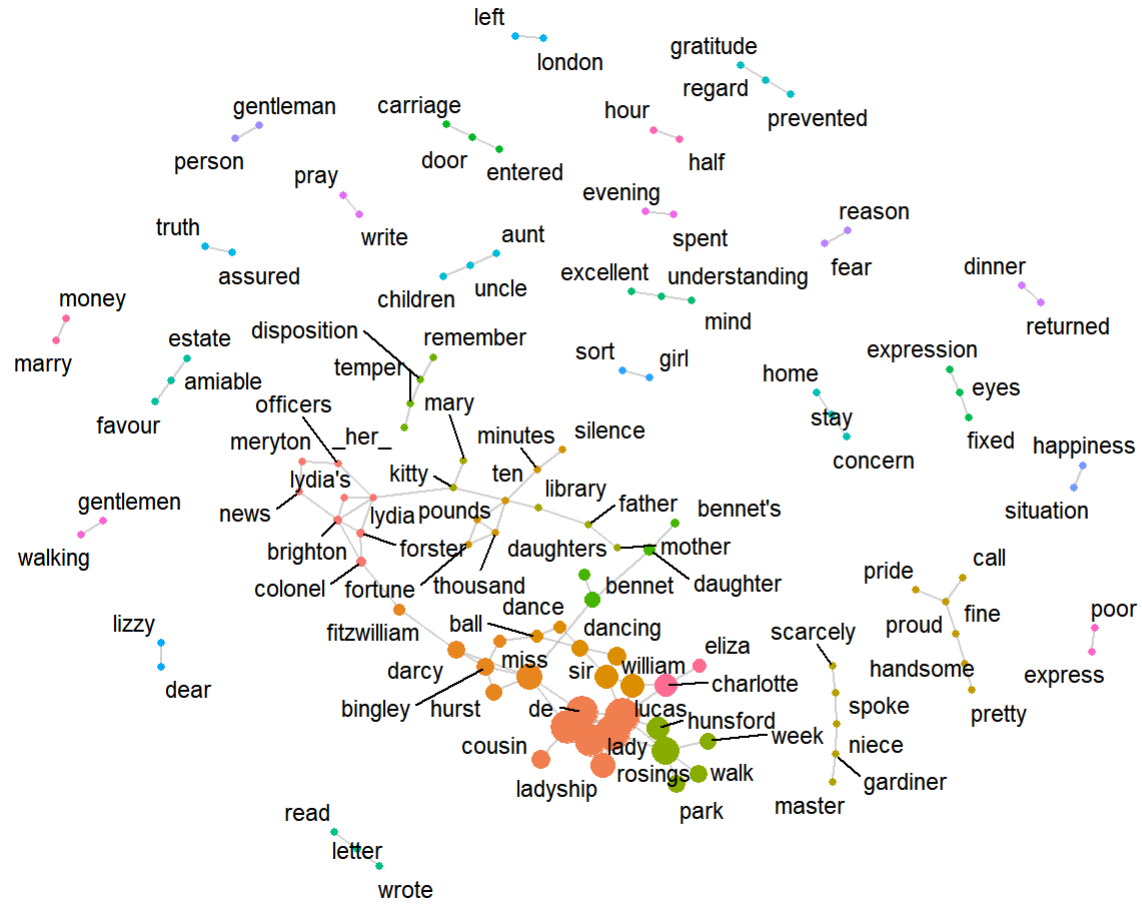
```
for (i in seq_len(length(igraph_layouts))) {
  p<-network.btw.data %>%
    ggraph(layout=igraph_layouts[i])+
    geom_edge_link(color='gray50', alpha=.2) +
    geom_node_point(aes(color=factor(group), size=eigen)) +
    geom_node_text(aes(label=name), size=3, repel=TRUE) +
    theme_graph() +
    theme(legend.position='none')+
    ggtitle(igraph_layouts[i])+
    theme_void()
  plot(p)
}
```

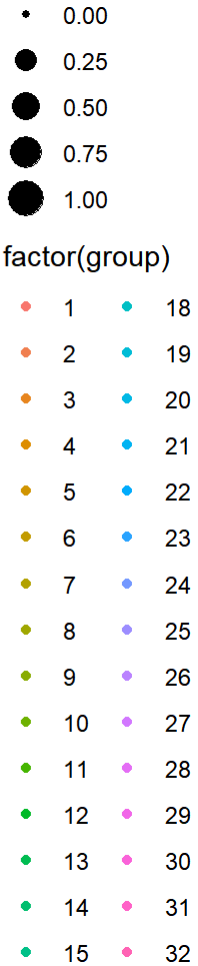
```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



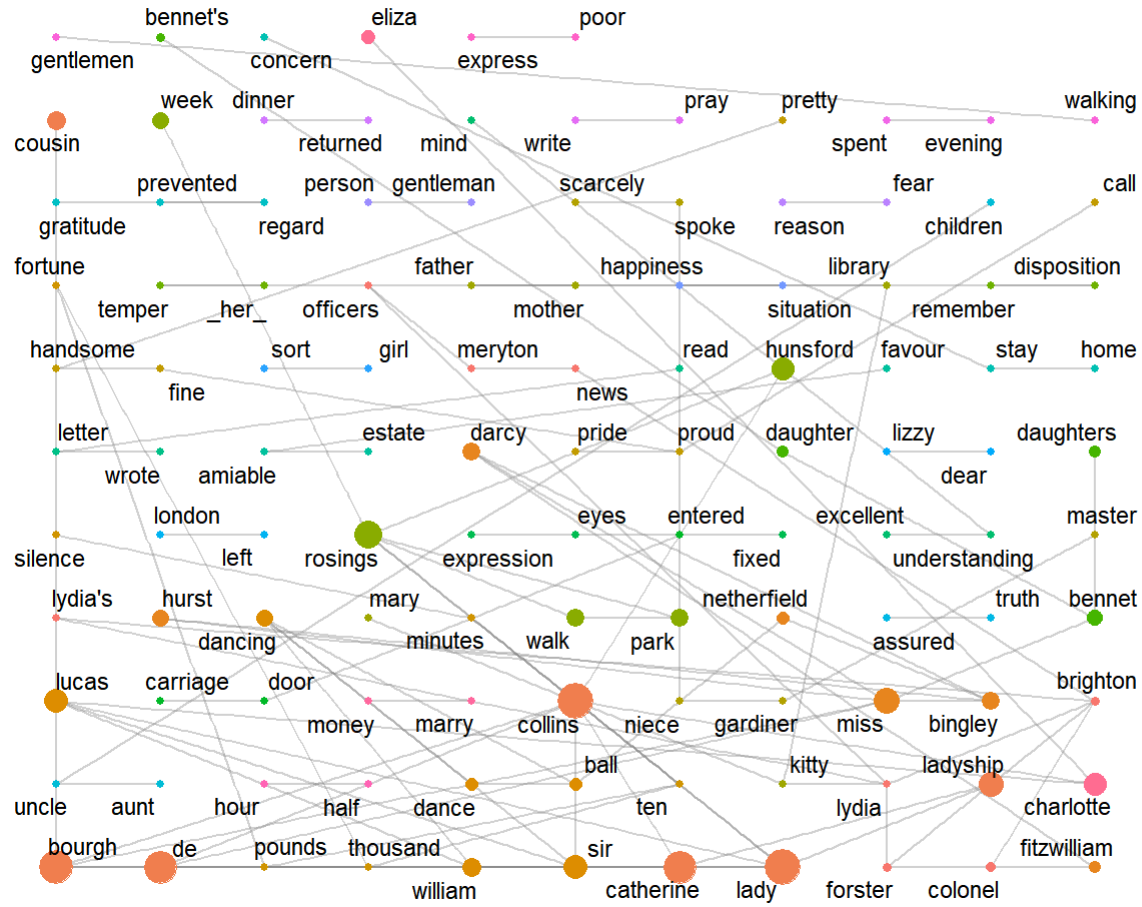

```
## Warning: ggrepel: 4 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

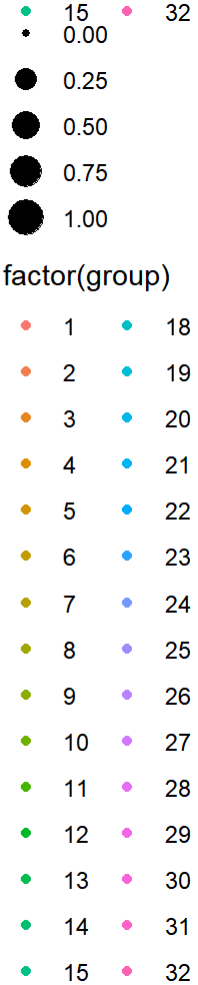
graphopt





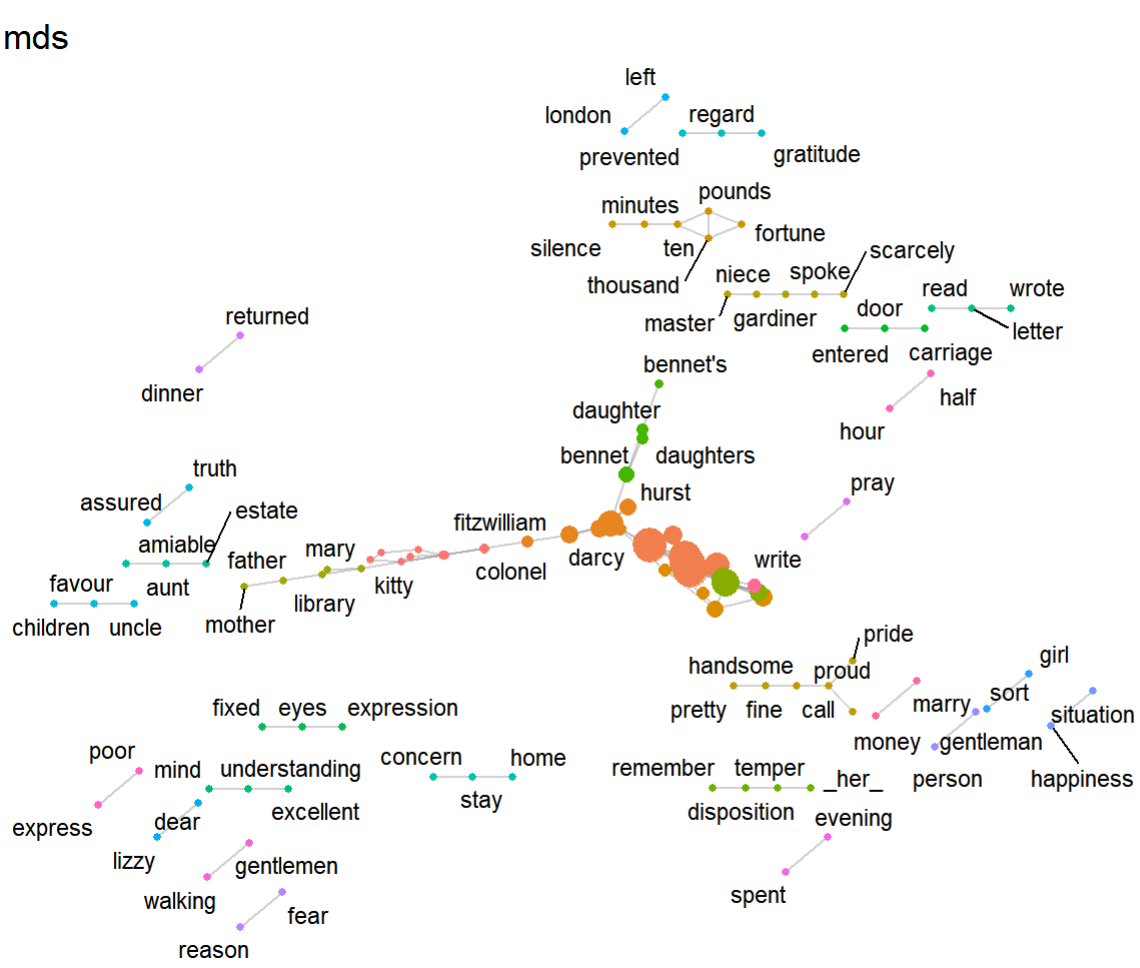
grid





```
## Warning: ggrepel: 30 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

mds

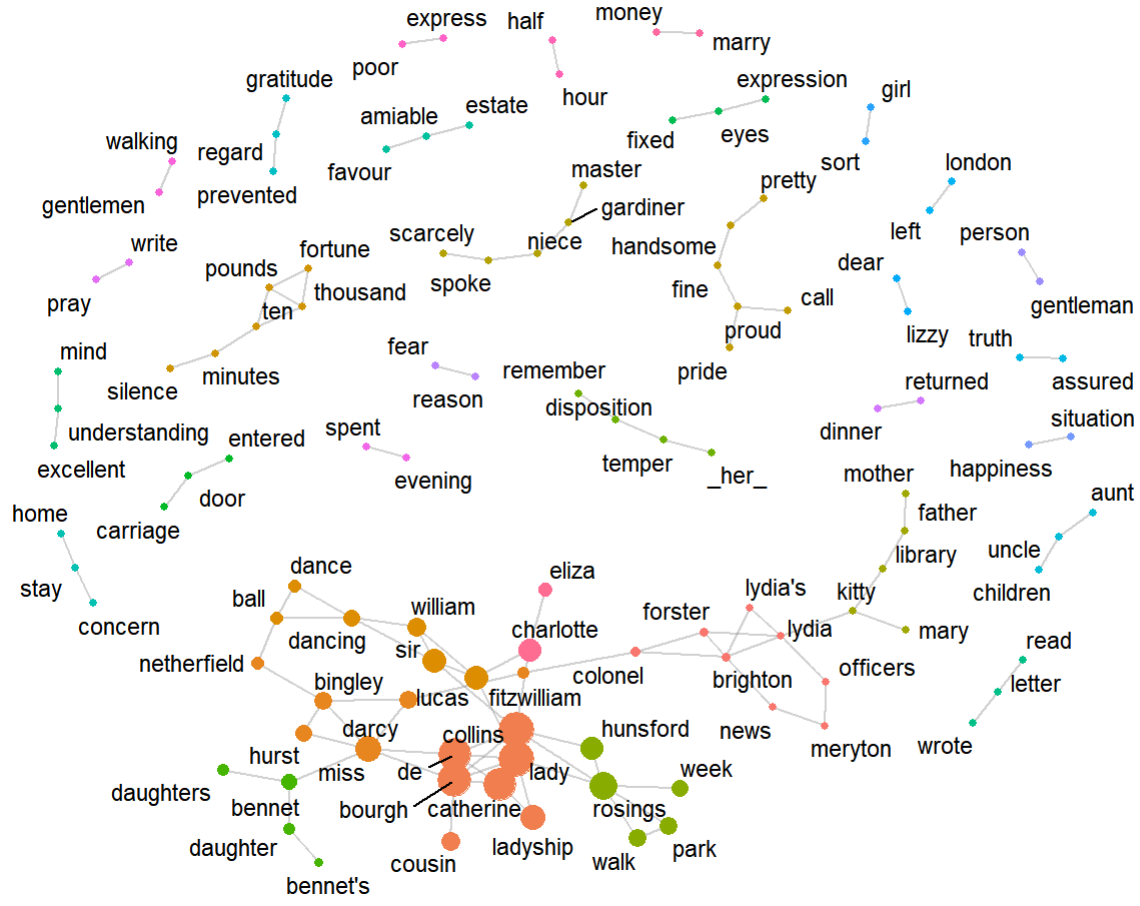


0.00
0.25
0.50
0.75
1.00

factor(group)

1 18
2 19
3 20
4 21
5 22
6 23
7 24
8 25
9 26
10 27
11 28
12 29
13 30
14 31
15 32

nicely

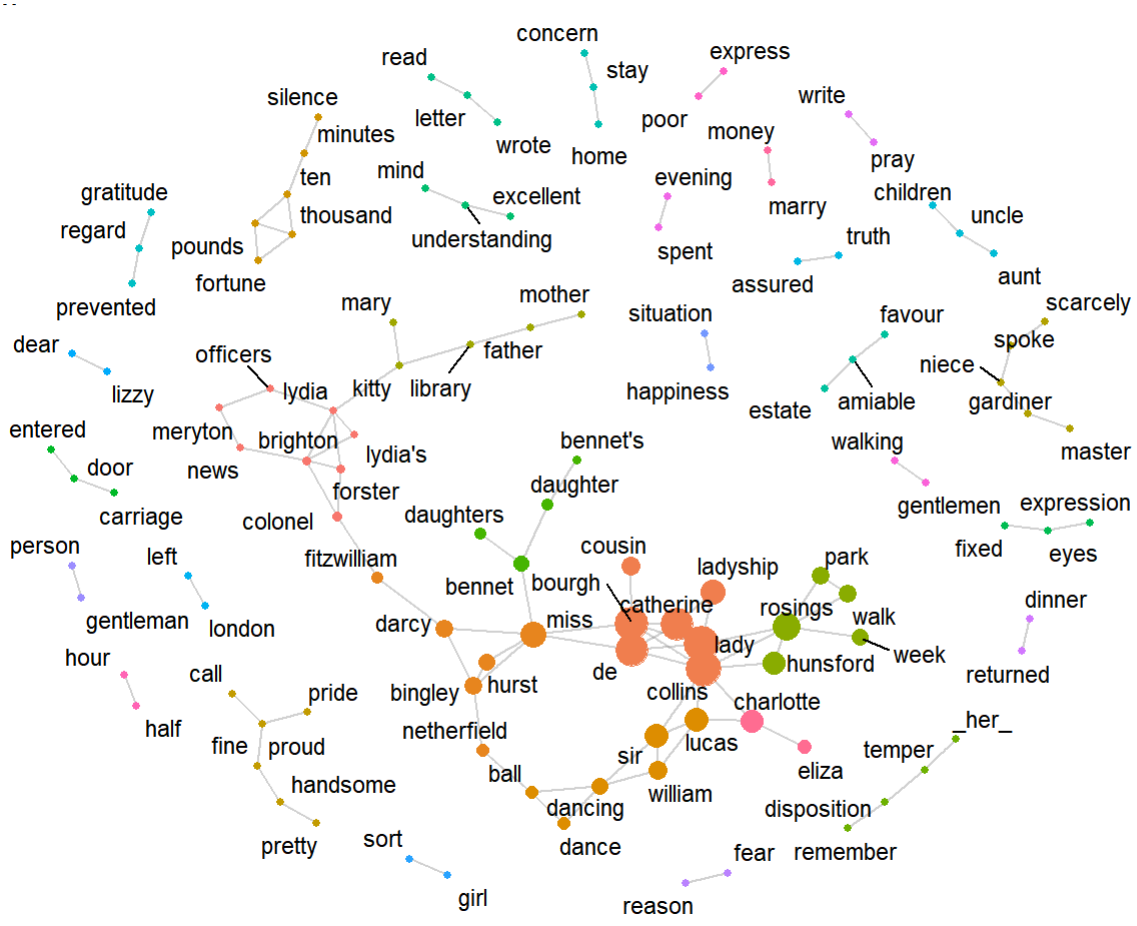


0.00
0.25
0.50
0.75
1.00

factor(group)

1 18
2 19
3 20
4 21
5 22
6 23
7 24
8 25
9 26
10 27
11 28
12 29
13 30
14 31
15 32

fr


concern
read
stay
express
silence
letter
poor
write
minutes
wrote
money
home
pray
ten
mind
evening
children
gratitude
excellent
marry
uncle
thousand
regard
understanding
truth
pounds
spent
aunt
fortune
assured
prevented
mary
mother
situation
favour
scarcely
dear
father
spoke
officers
niece
lizzy
lydia
kitty
library
happiness
estate
amiable
gardiner
entered
meryton
brighton
bennet's
walking
master
door
news
lydia's
daughter
forster
gentlemen
expression
carriage
colonel
daughters
fixed
eyes
person
cousin
ladyship
park
fitzwilliam
left
bennet
bourgh
dinner
catherine
rosings
walk
gentleman
london
darcy
miss
lady
week
hour
de
hunsford
returned
call
pride
bingley
hurst
collins
charlotte
half
netherfield
her
fine
proud
lucas
temper
sir
eliza
handsome
ball
william
dancing
disposition
pretty
sort
dance
fear
remember
girl
reason
0.25
0.50
0.75
1.00
factor(group)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32