# Les Miserables

*2018/2/5*

## Stage 1

This dataset I am using, collected by D.E. Knuth, contatins a network of characters in Victor Hugo's novel, Les Miserables. The original measurements are charaters' names and connections among the characters. Connections with less important characters might be the sources of noise in this case. The data is fully enumerated. Below is a undecorated plot of the data.

```
head(graph1)
```

```
## 6 x 77 sparse Matrix of class "dgCMatrix"
##    [[ suppressing 77 column names 'Napoleon', 'MlleBaptistine', 'MmeMagloire' ... ]]
##
## Napoleon        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## MlleBaptistine  . . 2 . . . . . . 2 . . . . . . . . . . . . . . . . . . . . .
## MmeMagloire     . 2 . . . . . . . 2 . . . . . . . . . . . . . . . . . . . . .
## CountessDeLo    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## Geborand        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## Champtercier    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
##
## Napoleon        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## MlleBaptistine  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## MmeMagloire     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## CountessDeLo    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## Geborand        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## Champtercier    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
##
## Napoleon        . . . . . . . . . . . . . . 2 . . .
## MlleBaptistine  . . . . . . . . . . . . . . 2 . . .
## MmeMagloire     . . . . . . . . . . . . . . 2 . . .
## CountessDeLo    . . . . . . . . . . . . . . 2 . . .
## Geborand        . . . . . . . . . . . . . . 2 . . .
## Champtercier    . . . . . . . . . . . . . . 2 . . .
```

```
plot(graph1)
```
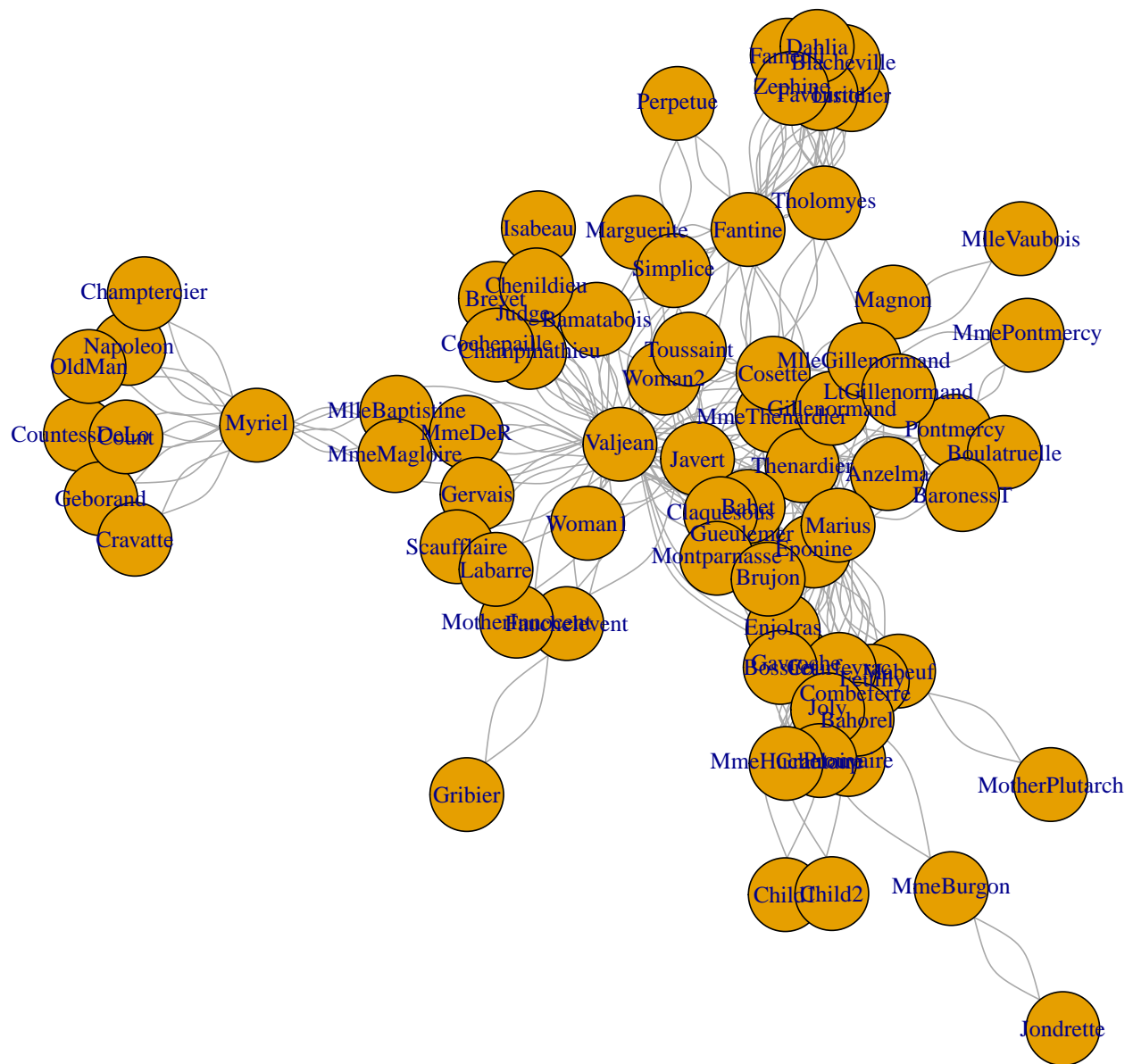
As one can easily observed from above, the plot is not very useful in explaining the relationships between charaters.

# Stage 2

As shown in below, a vertex is represented by each character and an edge attributeis a connection between one charater and another(demonstrated below). Attributes are the edge weights. Here I used visulizations to validate the network graph.

```
#vertex
V(graph1)
```

```
## + 77/77 vertices, named, from 9db2d8e:
## [1] Napoleon        MlleBaptistine   MmeMagloire       CountessDeLo
## [5] Geborand        Champtercier     Cravatte          Count
```

```
##  [9] OldMan           Valjean          Marguerite       MmeDeR
## [13] Isabeau          Gervais          Listolier        Fameuil
## [17] Blacheville      Favourite        Dahlia           Zephine
## [21] Fantine          MmeThenardier    Thenardier       Cosette
## [25] Javert           Fauchelevent     Bamatabois       Perpetue
## [29] Simplice         Scaufflaire      Woman1           Judge
## [33] Champmathieu     Brevet           Chenildieu       Cochepaille
## [37] Pontmercy        Boulatruelle     Eponine          Anzelma
## + ... omitted several vertices
```

*#edge*
**E**(graph1)

```
## + 508/508 edges from 9db2d8e (vertex names):
##  [1] Napoleon       --Myriel      MlleBaptistine--Myriel
##  [3] MmeMagloire    --Myriel      MlleBaptistine--MmeMagloire
##  [5] CountessDeLo   --Myriel      Geborand      --Myriel
##  [7] Champtercier   --Myriel      Cravatte      --Myriel
##  [9] Count          --Myriel      OldMan        --Myriel
## [11] Valjean        --Labarre     MmeMagloire   --Valjean
## [13] MlleBaptistine--Valjean      Valjean       --Myriel
## [15] Valjean        --Marguerite  Valjean       --MmeDeR
## [17] Valjean        --Isabeau     Valjean       --Gervais
## [19] Listolier      --Tholomyes   Fameuil       --Tholomyes
## + ... omitted several edges
```

# Stage 3

### 1st visualization

For this plot, I calculated the edge betweenness for each notion first to see which relationships are the most important. The goal is to show important relationships on the plot. I indicated the betweenness by the size of the vertex, a larger vertex size indicates a character with strong betweenness. Also, a thicker edge line indicates an important relationship.

According the plot, it is obvious that Valijean is the major character and the most important character in the novel. Valijean's relationship with Bishop Myriel is indeed one of the most important relationships in the story.The advantage of this plot is that main relationships and important characters are obviously seen.

```
plot(graph2,
    vertex.label = gsub(" ", "\n", V(graph2)$name),
    vertex.frame.color = "gray",
    vertex.label.color = "black",
    vertex.label.cex = got_between * 0.001,
    vertex.frame.color = "gray",
    vertex.size = got_between * 0.01 + 1,
    edge.width = betweenness * 0.01,
    edge.arrow.size = 0.5
)
```
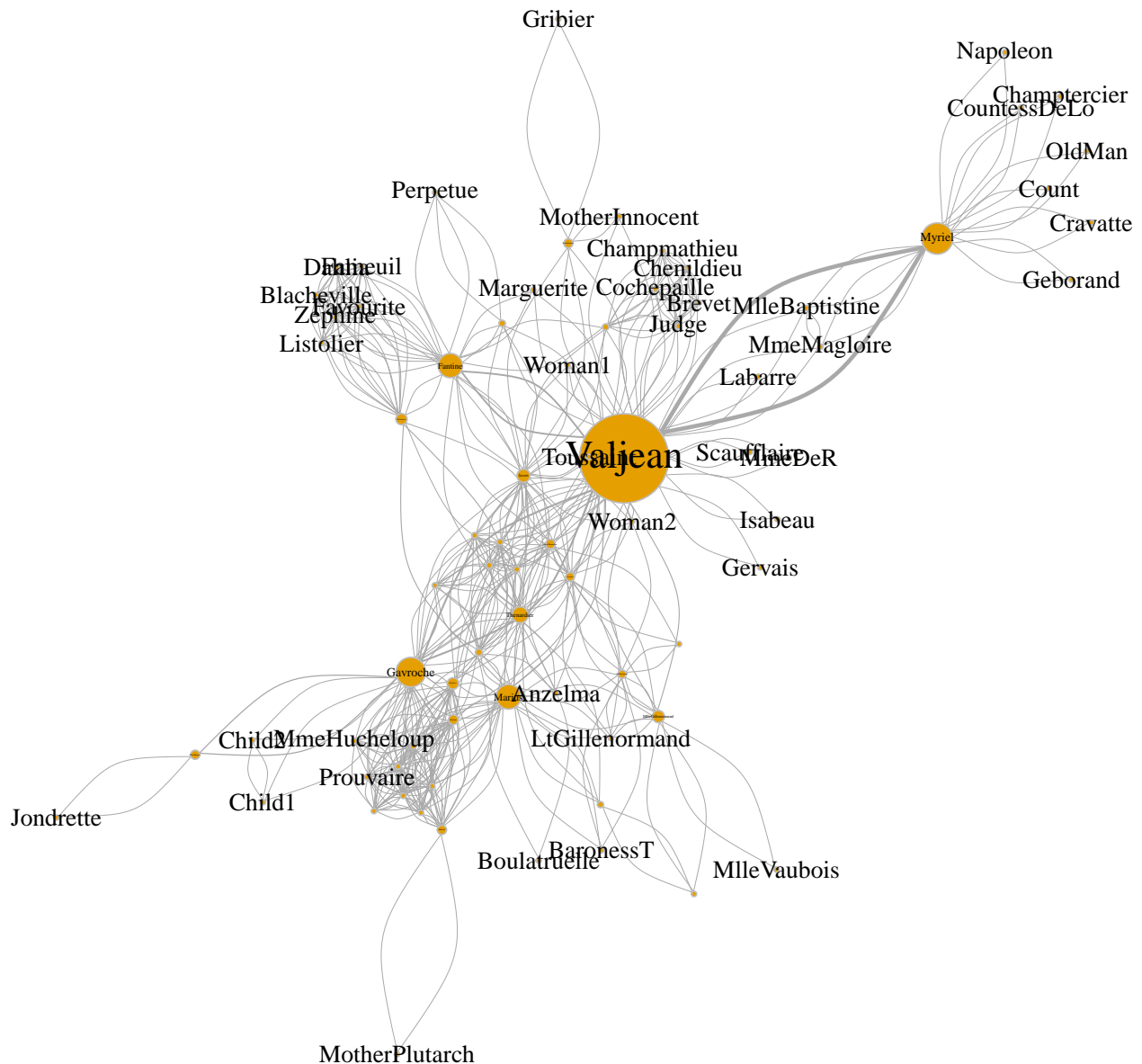
## 2nd visualization

For the second visualization, I looked into betweenness of vertex in different clusters. Clusters are shown by different colors in the plot. This plot can easily tell which characters co-appeared in the novel. Also, it shows clusters on a fixed scale together with charaters' names, which is very easy to interpret. This plot is very useful especially when examining details.

```
igraph.options(vertex.size = 3,label.size=3, edge.arrow.size = 0.5)
ceb <- cluster_edge_betweenness(graph2)
dendPlot(ceb, mode="hclust")
```