

Big Data HW #5

Will Clark & Matthew DeLio
41201-01

May 13, 2015

1 Plotting the Actors' Network

In this section we plot the actors' network as a whole (see Figure 1 on the following page). For this plot, the node-size remained constant with the vertex color varying based on an actor's degree of connectedness (more blue for less connected and more red for well connected). The data-set has 7015 actors in it. The degree of connectedness in the set of actors is shown in Figure 2 on page 3. This histogram shows that most actors have worked with between 40 and 60 other actors, although the heavy tail in this distribution indicates that many have worked with far more. For instance "Dobtcheff, Vernon" has worked with 378 others.

2 Kevin Bacon

Kevin Bacon's network can be visualized in Figure 3 on page 3. In each of these plots, the node-size of each actor is sized according to their degree within the overall actors' network. That is, larger degree actors have bigger nodes. The coloring scheme used was to label 3rd-degree connections blue, 2nd-degree green, 1st-degree red, with Kevin Bacon's node colored black (see R-Snippet 1 on page 4 for code used to generate these plots).

In Figure 3a we see his 1st-degree connections and note that while Kevin Bacon is relatively well connected with 96 connections, 22.68% of his connections have a higher degree than he does. In Figure 3b we visualize 2nd-degree connections and note that of the 2128 connections, 27.99% have a higher degree. Also because of the way R rendered the visualization, we cannot really see any 1st degree conenctions. In Figure 3c we visualize 3rd-degree connections and see a much better visual dispersion of the three types of connections. Of the 5980 connections, 31.23% have more connections than Kevin Bacon.

3 Most Common and Connected Actors

We can determine the most common actors by summing up the columns of the actors' matrix and sorting the result. The actor appearing in the most films is **Zivojinovic, Velimir 'Bata'**, who

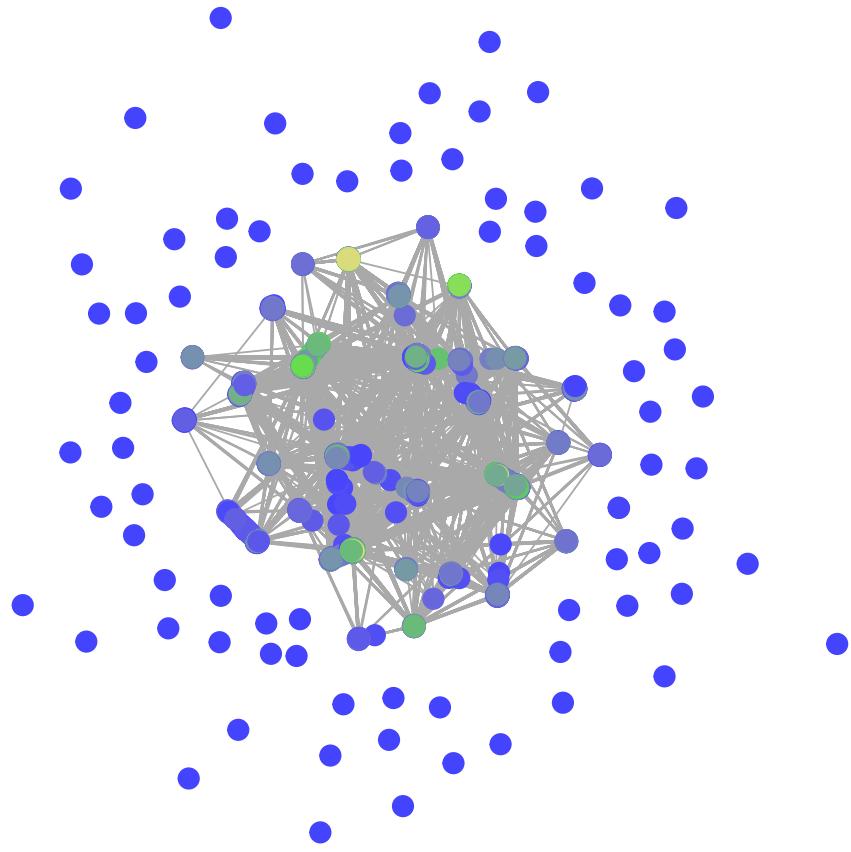


Figure 1: Visualization of Actors' Network

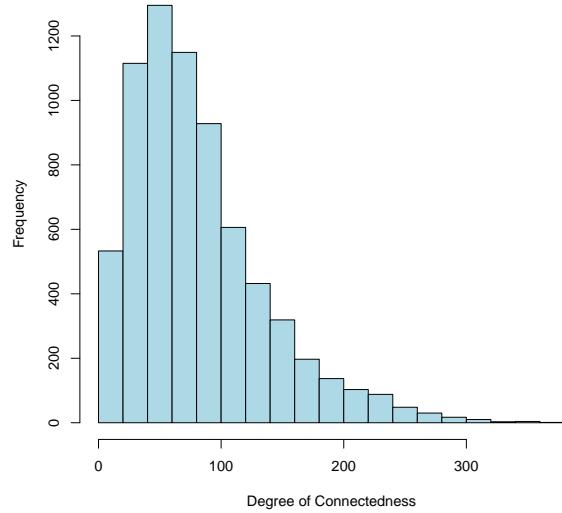


Figure 2: Histogram of Actors' Connectedness

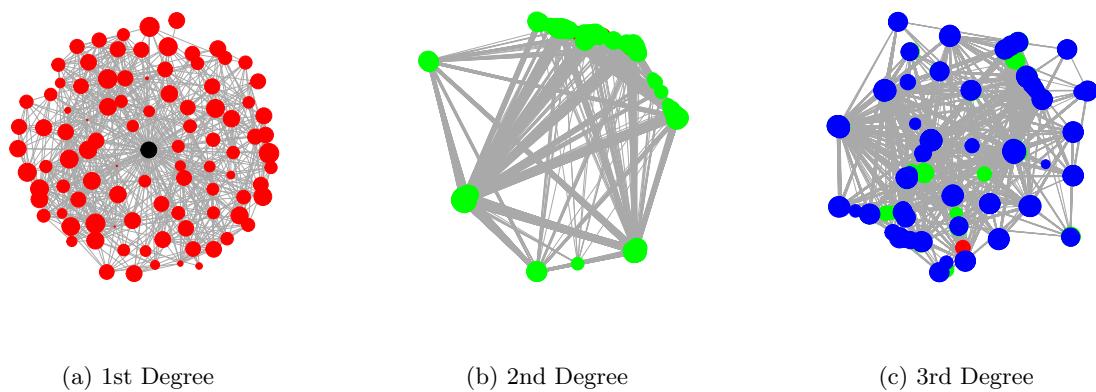


Figure 3: Kevin Bacon's Network

appeared in 57 films. The most connected actor is **Dobtcheff, Vernon**, who has been in at least one film with 378 other actors.

Here, we randomly select two actors **Auger, Claudine** and **Dray, Albert**. The shortest path between them contains 2 vertices: Auger, Claudine was in Secrets de la princesse de Cadignan, Les with Arditi, Pierre ;, Arditi, Pierre was in Adieu blaireau with Dray, Albert ;

4 Appendix

R-Snippet 1.

```
# Generate actor vertex-sizes based on the degree of connectedness on the
# overall actor network. These constants were found by manually tuning the
# size histogram.
actorSizes <- 3*log(0.5*degree(actnet)+1)+1

# Color the nodes for the 1st-degree plot
V(actor1st)$color <- "red"
V(actor1st)[actor]$color <- "black"
V(actor1st)$frame.color <- V(actor1st)$color
V(actor1st)$size <- actorSizes[V(actor1st)]
plot(actor1st, vertex.label=NA, edge.curved=F, margin=0)

# Color the nodes for the 2nd-degree plot
V(actor2nd)$color <- "green"
V(actor2nd)[V(actor1st)]$color <- "red"
V(actor2nd)[actor]$color <- "black"
V(actor2nd)$frame.color <- V(actor2nd)$color
V(actor2nd)$size <- actorSizes[V(actor2nd)]
plot(actor2nd, vertex.label=NA, edge.curved=F, margin=0)

# Color the nodes for the 3rd-degree plot
V(actor3rd)$color <- "blue"
V(actor3rd)[V(actor2nd)]$color <- "green"
V(actor3rd)[V(actor1st)]$color <- "red"
V(actor3rd)[actor]$color <- "black"
V(actor3rd)$frame.color <- V(actor3rd)$color
V(actor3rd)$size <- actorSizes[V(actor3rd)]
plot(actor3rd, vertex.label=NA, edge.curved=F, margin=0)
```