Workflows for Visualization

An example with bigger dataset

Here the visualization workflow from Dalziel et al. 2013. Human mobility patterns predict divergent epidemic dynamics among cities. PRSB 280.

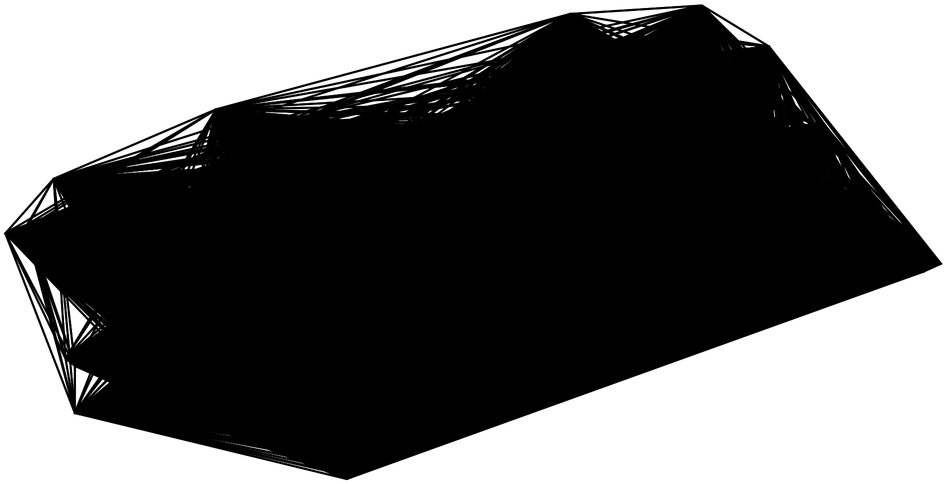
```
rm(list = ls())
load("../data/TorontoCommuting.Rdata")
ls()
```

```
## [1] "A" "d" "N" "x" "y"
```

These data consist of commuting flows between 743 census tracts within the city of Toronto, connecting the home and work locations of 3.8 million workers. These are organized in the matrix A, such that element in the ith row and jth column corresponding to the number of people who live in census tract i and work in census tract j. The latitude and longitude of the census tracts are stored in the vectors x and y. The matrix d supplies the pairwise distance between census tracts. N is the population size of each census tract.

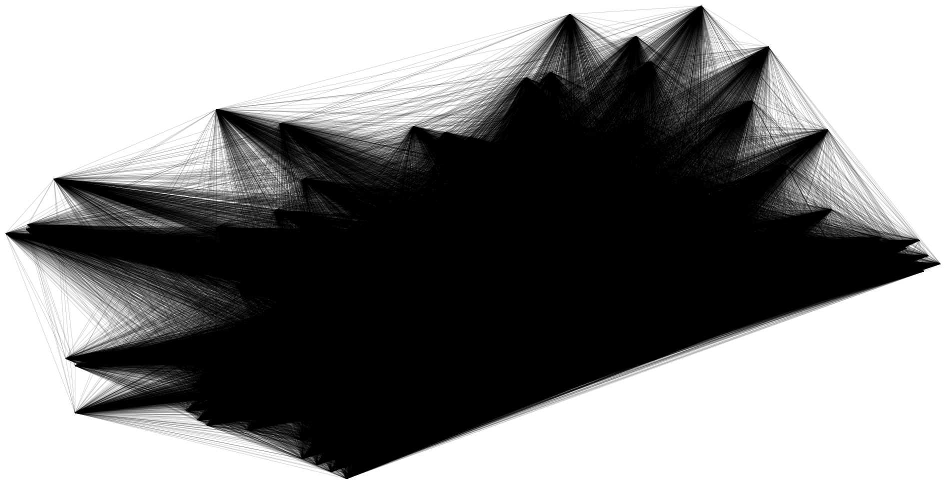
Here is a barebones network plot where each edge represents a commuting flow between a pair of locations

```
# Assign home and work x and y locations to each element of the commmuting matrix
n <- nrow(A)
hx <- matrix(x,n,n)
hy <- matrix(y,n,n)
wx <- t(hx)
wy <- t(hy)


# Empty plot
plot(range(x), range(y),
     type="n",
     xlab="",
     ylab="",
     bty="n",
     xaxt="n",
     yaxt="n")

# Draw line segments for each commuting flow
segments(hx, hy, wx, wy)
```



What a mess. The problem is that there are too many lines and they are too thick to see what is going on. A quick cleanup step is to make each of the lines very thin, and to make them almost transparent. We control line color using the `rgb()` function, which specifies colors using the red, green, blue, alpha system, where alpha is transparency. Let's also plot the most important (ie biggest) flows last, so they don't get obscured by all the tiny ones.

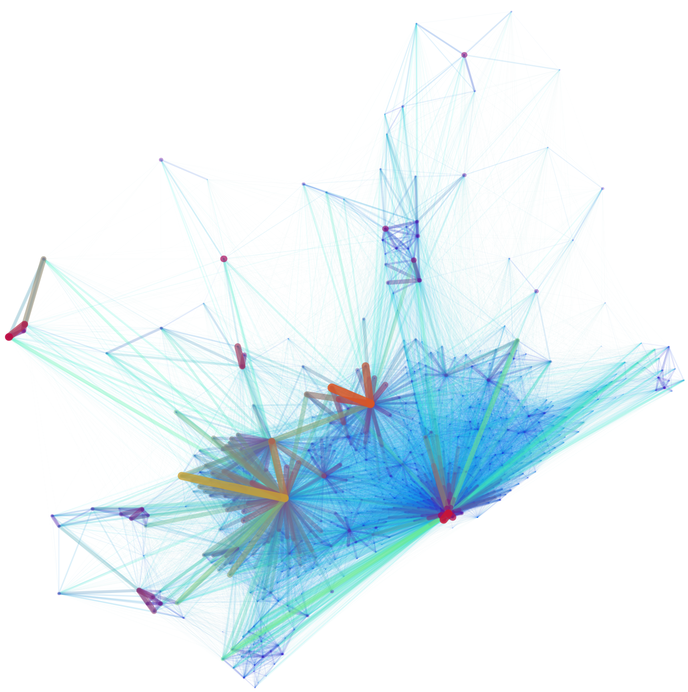
```
# Assign home and work x and y locations to each element of the commmuting matrix
n <- nrow(A)
hx <- matrix(x,n,n)
hy <- matrix(y,n,n)
wx <- t(hx)
wy <- t(hy)


# Specify edge widths and colors
edge_width <- 0.1
edge_color <- rgb(red = 0, green = 0, blue = 0, alpha = 0.1)     # black, nearly transparent


# Sort by flow so thickest segments are plotted last
idx <- order(A)
idx <- idx[idx>0]


# Empty plot
plot(range(x), range(y),
     type="n",
     xlab="",
     ylab="",
     bty="n",
     xaxt="n",
     yaxt="n")

# Draw line segments for each commuting flow
segments(hx[idx], hy[idx], wx[idx], wy[idx],
         lwd = edge_width,
         col = edge_color)
```



That's better, but the internal stucture is still obscured by all the "ink" that's being drawn, even after we made it mostly transparent. There are so many lines on top of each other, that tranparency alone isn't enough to let them be distinguished. The next step is to use variable widths and colors to distinguish edges. We are interested in the pattern of the biggest flows, so lets map the flow volume and distance to colors.

```
# Plotting parameters
max_edge_width <-  5
min_edge_width <- 0.1
max_alpha <- 1
min_alpha <- 0


# Assign home and work x and y locations to each element of the commmuting matrix
n <- nrow(A)
hx <- matrix(x,n,n)
hy <- matrix(y,n,n)
wx <- t(hx)
wy <- t(hy)


# Standardize commuting flows and distance in A in preparation for mapping to colors
standardized_flow <- 1-exp(-0.0015*A)
standardized_distance <- 1-exp(-0.07*d)
standardized_distance[is.na(standardized_distance)] <- 0


# Specify red, green, blue and transparency channels for the edge colors
# as functions of flow volume and distance
red <- standardized_flow
green <- standardized_distance
blue <- 1 - standardized_flow
alpha <- (max_alpha - min_alpha) * standardized_flow
alpha <- alpha - min(alpha) + min_alpha


# Set the edge colors and widths
edge_color <- rgb(red, green, blue, alpha)
edge_width <- standardized_flow*(max_edge_width-min_edge_width)
edge_width <- edge_width + min_edge_width
```

```
# Sort by flow so thickest segments are plotted last
idx <- order(A)
idx <- idx[idx>0]


# Plot
par(fin = c(4,4))
par(mai = c(0.1,0.1,0.1,0.1))
plot(range(x), range(y),
     type="n",
     xlab="",
     ylab="",
     bty="n",
     xaxt="n",
     yaxt="n")

segments(hx[idx], hy[idx], wx[idx], wy[idx],
         lwd = edge_width[idx],
         col = edge_color[idx])
```



Look what was hiding in there! As a last step we will make a legend, by applying the same color mapping rules we used on the plot to a grid of flow volumes.

```
legendMinX <- 40
legendMaxY <- 70
legendWidth <- 12
legendHeight <- 40
legendXResolution <- 8
legendYResolution <- 10
maxEdgeWidth <-  18
minEdgeWidth <- 0.1
```

```
legendXSequence <- seq(legendMinX,legendMinX+legendWidth,length.out=legendXResolution)
legendYSequence <- seq(legendMaxY-legendHeight,legendMaxY,length.out=legendYResolution)

legendStandardizedDistanceSequence <- legendXSequence - min(legendXSequence)
legendStandardizedDistanceSequence <- legendStandardizedDistanceSequence/max(legendStandardizedDistanceS
legendStandardizedFlowSequence <- legendYSequence - min(legendYSequence)
legendStandardizedFlowSequence <- legendStandardizedFlowSequence/max(legendStandardizedFlowSequence)

out <- expand.grid(legendXSequence,legendYSequence)
legendX <- out[,1]
legendY <- out[,2]

out <- expand.grid(legendStandardizedDistanceSequence,legendStandardizedFlowSequence)
legendStandardizedDistance <- out[,1]
legendStandardizedFlow <- out[,2]

legendRed <- legendStandardizedFlow
legendGreen <- 1-legendStandardizedDistance
legendBlue <- 1 - legendStandardizedFlow
#legendAlpha <- (maxAlpha-minAlpha)*legendStandardizedFlow
#legendAlpha <- legendAlpha - min(legendAlpha) + minAlpha
legendAlpha <- 1

legendCol = rgb(legendRed,legendGreen,legendBlue,legendAlpha)
legendEdgeWidth <- legendStandardizedFlow*maxEdgeWidth + minEdgeWidth

#segments(legendX,legendY,legendX,legendY,col=legendCol,lwd=legendEdgeWidth)

plot(range(legendX),range(legendY),type="n",axes=FALSE,xlab="",ylab="")
segments(legendX,legendY,legendX,legendY,col=legendCol,lwd=legendEdgeWidth)
```

Thicker lines with hotter colors correspond to more flow volume. The long distance flows go from green to yellow as volume increases, while the short distance flows go from blue to red.