ORF 350: Assignment 6

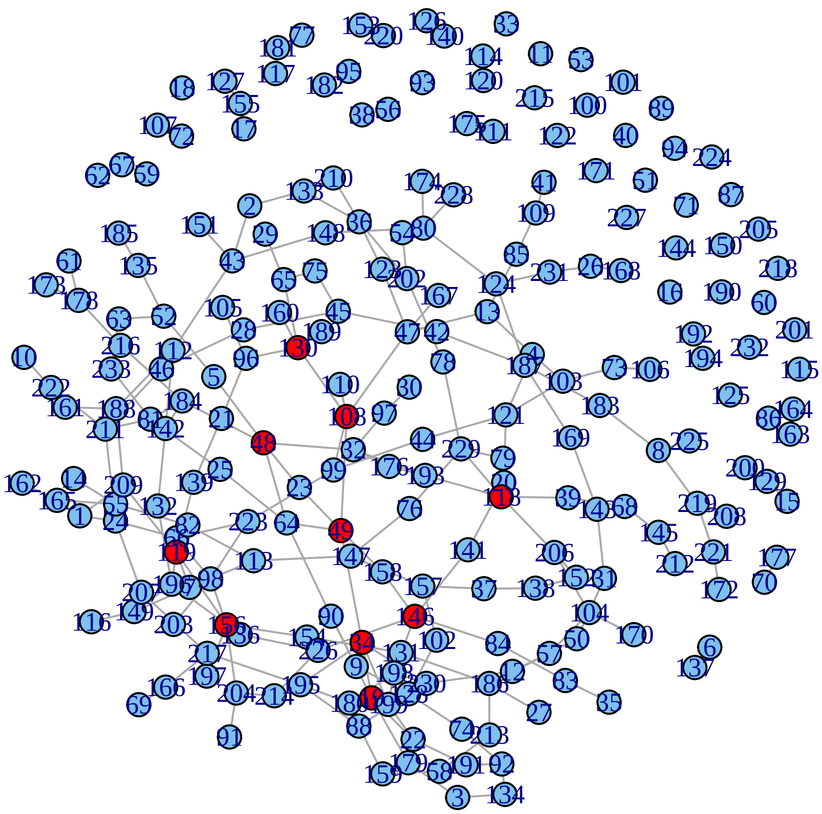
David Fan

5/11/2017

Question 1: Nationwide GDP Growth Correlation (30 points)

```
setwd("/Users/dfan/Dropbox/School/Sophomore Year/Spring 2017/ORF 350/Assignments/HW6")
load("gdp.Rdata")  # loads gdp object

### 1.1 remove countries with no GDP data across all 14 years
toRemove <- numeric(length = 0)
for (i in 1:nrow(gdp)) {
    if (length(which(is.na(gdp[i, ]))) == ncol(gdp)) {
        toRemove <- c(toRemove, i)
    }
}
gdp <- gdp[-toRemove, ]

# replace any NA's in the remaining countries with the mean
# of the row (not including NA values)
for (i in 1:nrow(gdp)) {
    toReplace <- which(is.na(gdp[i, ]))
    if (length(toReplace) > 0) {
        gdp[i, toReplace] <- mean(as.numeric(gdp[i, ]), na.rm = TRUE)
    }
}

### 1.2
set.seed(1)  # for reproducibility
M <- matrix(0, nrow = nrow(gdp), ncol = nrow(gdp))
for (i in 1:nrow(gdp)) {
    # lasso is alpha = 1. Use tuning parameter lambda = 1 in lieu
    # of cv
    model <- glmnet(t(gdp[-i, ]), t(gdp[i, ]), family = "gaussian",
        alpha = 1, lambda = 1)
    model_coef <- coef(model)[-1]  # don't want the intercept
    neighbors <- which(model_coef != 0)
    # adjust for countries shifted based on the current response
    # variable
    neighbors[which(neighbors >= i)] <- neighbors[which(neighbors >=
        i)] + 1
    M[i, neighbors] <- 1
}
# Apply the AND rule
for (i in 1:nrow(M)) {
    for (j in 1:ncol(M)) {
        if (M[i, j] != M[j, i]) {
            M[i, j] <- 0
            M[j, i] <- 0
        }
```

```r
    }
}

# visualize graph
graphplot = function(X) {
    ag = graph.adjacency(X, mode = "undirected")
    V(ag)$colors = ifelse(degree(ag) < 5, "SkyBlue2", "red")
    par(mai = c(0, 0, 0, 0))
    plot.igraph(ag, vertex.color = V(ag)$colors, vertex.size = 6,
        vertex.label.cex = 0.8, layout = layout_nicely(ag))
}

graphplot(M)
```



```r
paste("Red nodes:")
```

```
## [1] "Red nodes:"
```

```r
rownames(gdp)[which(rowSums(M) >= 5)]
```

```
##  [1] "Bulgaria"                        "Central Europe and the Baltics"
##  [3] "Cyprus"                          "Czech Republic"
##  [5] "Cambodia"                        "Libya"
##  [7] "St. Lucia"                       "Latvia"
##  [9] "Montenegro"                      "Nigeria"
```

Question 2: Preliminary Theories (25 points)

2.1

Global independence does not imply conditional independence: consider two independent coin flips $C_1, C_2 \in \{H, T\}$. Let X be the event that at least one of the coin flips is heads. $p(C_1 = H, C_2 = H|X) \neq p(C_1 = H|X) * p(c_2 = H|X)$. Note that $p(X) = \frac{3}{4}$ since the sample space is {HH, HT, TH, TT} and three of the possibilities have at least one heads. $p(C_1 = H|X) = \frac{p(X|C_1=H)P(C_1=H)}{P(X)} = \frac{1*1/2}{3/4} = \frac{2}{3}$ and $p(C_2 = H|X) = \frac{p(X|C_2=H)P(C_2=H)}{P(X)} = \frac{1*1/2}{3/4} = \frac{2}{3}$. Counting the sample space, $p(C_1 = H, C_2 = H|X) = \frac{1}{3} \neq \frac{2*2}{3*3} \neq \frac{4}{9}$.

Conditional independence does not imply global independence: Consider $y_1 = \beta x + \epsilon_1$ and $y_2 = \beta x + \epsilon_2$. Epsilons are independent Gaussians in the linear model, so for two different responses under the same model, ϵ_1 and ϵ_2 are independent conditioned on x. $y_1|x$ is just a constant + ϵ_1, and $y_2|x$ is just a constant + ϵ_2, so y_1 and y_2 are just independent Gaussians conditioned on x. y_1 and y_2 are obviously not globally independent since they depend on the same covariates.

2.2

1) False. Node set {3} does not separate node sets {5} and {7}. Removing {3} leaves {5} and {7} still connected.
2) True. Node set {8} separates node sets {7} and {9}. Removing {8} leaves {7} and {9} disconnected since the only edge from {9} is to {8}.
3) True. $X\backslash$\{1,8\} separates {1} and {8} since removing all other node sets leaves {1} and {8} disconnected. Thus, X_1 and X_8 are conditionally independent given $X\backslash$\{1,8\}.
4) True. X_4 separates {1} and {8} since removing node set {4} leaves {1} and {8} disconnected. Thus, X_1 and X_8 are conditionally independent given X_4.
5) False. Since {5} does not separate {1} and {3}, {1} and {3} are not conditionally independent given {5}.

Question 3: The Gaussian Graphical Model (25 points)

3.1

$\Theta\Sigma = I$, and this matrix multiplication leads to four equations. We only need two of them for this proof.

$$\begin{aligned} 1)\ \Theta_{AA}\Sigma_{AA} + \Theta_{AA^c}\Sigma_{A^cA} &= I \\ 2)\ \Theta_{AA}\Sigma_{AA^c} + \Theta_{AA^c}\Sigma_{A^cA^c} &= 0 \end{aligned}$$

Multiplying both sides of 1) by Θ_{AA}^{-1}, we get

$$\Theta_{AA}^{-1} = \Sigma_{AA} + \Theta_{AA}^{-1}\Theta_{AA^c}\Sigma_{A^cA}$$

Using equation 2:

$$\begin{aligned} \Theta_{AA^c}\Sigma_{A^cA^c} &= -\Theta_{AA}\Sigma_{AA^c} \\ \Theta_{AA^c} &= -\Theta_{AA}\Sigma_{AA^c}\Sigma_{A^cA^c}^{-1} \end{aligned}$$

Substituting in:

$$\begin{aligned} \Theta_{AA}^{-1} &= \Sigma_{AA} - \Theta_{AA}^{-1}\Theta_{AA}\Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}\Sigma_{A^cA} \\ \Theta_{AA}^{-1} &= \Sigma_{AA} - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}\Sigma_{A^cA} \end{aligned}$$

3.2

We want to prove that $X_j \perp X_K | X \setminus \{j,k\} \leftrightarrow \Theta_{jk} = 0, \forall j \neq k$. $A = \{1\}$ and $A^c = \setminus\{1\}$.

From right to left: Let $A = \{j,k\}$. Since $\Theta_{jk} = 0\ \forall\ j \neq k$, Θ_{AA} is diagonal and thus Θ_{AA}^{-1} is diagonal. Then, $X_A|X_{A^c} \sim N(..., \Theta_{AA}^{-1})$. X_j and X_k are jointly Gaussian and two Gaussian variables are independent if and only if their covariance is 0, which is given by $\Theta_{jk} = 0\ \forall\ j \neq k$.

From left to right: Since X_j and X_k are conditionally independent given $\setminus\{j,k\}$, and are jointly Gaussian, their covariance matrix Θ_{AA}^{-1} must be diagonal. If Θ_{AA}^{-1} is diagonal, then so is Θ_{AA}. And this is true because $\Theta_{jk} = 0\ \forall\ j \neq k$.

3.3

$$\begin{aligned} X_A &= \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c} + \epsilon \\ \epsilon &= X_A - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c} \\ E(\epsilon) &= E(X_A) - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c} \\ &= \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c} - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c} \\ &= 0 \end{aligned}$$

So, $\epsilon \sim N(0, \Theta_{AA}^{-1})$ since $Var(\epsilon) = \Sigma_{AA} - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}\Sigma_{A^cA}$. $X_A \sim N(\Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c}, \Theta_{AA}^{-1})$.

Show $\epsilon \perp X \setminus 1$: Let $A = \{1\}$ and $A^c = \setminus\{1\}$.

$$
\begin{aligned}
Cov(\epsilon, X_{A^c}) &= Cov(X_A - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c}, X_{A^c}) \\
&= Cov(X_A, X_{A^c}) - Cov(\Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}X_{A^c}, X_{A^c}) \\
&= \Sigma_{AA^c} - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}Cov(X_{A^c}, X_{A^c}) \\
&= \Sigma_{AA^c} - \Sigma_{AA^c}\Sigma_{A^cA^c}^{-1}\Sigma_{A^cA^c} \\
&= 0
\end{aligned}
$$

Show $\beta = -\Theta_{11}^{-1}\Theta_{\setminus 1,1}$:

$$
\begin{aligned}
\Sigma\Theta &= I \\
\Sigma_{A^cA}\Theta_{AA} + \Sigma_{A^cA^c}\Theta_{A^cA} &= 0 \\
\Sigma_{A^cA}\Theta_{AA} &= -\Sigma_{A^cA^c}\Theta_{A^cA} \\
\Sigma_{A^cA^c}^{-1}\Sigma_{A^cA} &= -\Theta_{A^cA}\Theta_{AA}^{-1} \\
&= -\Theta_{AA}^{-1}\Theta_{A^cA} \\
&= \beta
\end{aligned}
$$

Note that Θ_{AA} and Θ_{AA}^{-1} are scalars.

Question 4: The Ising model (20 points)

4.1

Without loss of generality, we only consider $X_j = 1$ and $X_k = 1$. Since X_j and X_k are conditionally independent, $P(X_j = 1, X_k = 1|X \setminus \{j,k\}) = P(X_j = 1|X \setminus \{j,k\})P(X_k = 1|X \setminus \{j,k\})$.

$$
\begin{aligned}
P(X_j = 1, X_k = 1|X \setminus \{j,k\}) &= \frac{P(X_j = 1, X_k = 1, X \setminus \{j,k\})}{\sum\limits_{X_j, X_k \in \{\pm 1\}} P(X_j = 1, X_k = 1, X \setminus \{j,k\} = x \setminus \{j,k\})} \\
&Let\ A_1 = P(X_j = 1, X_k = 1, X \setminus \{j,k\}) \\
&Let\ A_2 = P(X_j = 1, X_k = -1, X \setminus \{j,k\}) \\
&Let\ A_3 = P(X_j = -1, X_k = 1, X \setminus \{j,k\}) \\
&Let\ A_4 = P(X_j = -1, X_k = -1, X \setminus \{j,k\}) \\
P(X_j = 1, X_k = 1|X \setminus \{j,k\}) &= \frac{A_1}{A_1 + A_2 + A_3 + A_4} \\
P(X_j = 1|X \setminus \{j,k\}) &= \frac{\sum\limits_{X_k \in \{\pm 1\}} P(X_j = 1, X_k = x_k, X \setminus \{j,k\}}{\sum\limits_{X_j, X_k \in \{\pm 1\}} P(X_j = x_j, X_k = x_k, X \setminus \{j,k\} = x \setminus \{j,k\}} \\
&= \frac{A_1 + A_2}{A_1 + A_2 + A_3 + A_4} \\
P(X_k = 1|X \setminus \{j,k\}) &= \frac{A_1 + A_3}{A_1 + A_2 + A_3 + A_4}
\end{aligned}
$$

Since $P(X_j = 1, X_k = 1|X \setminus \{j,k\}) = P(X_j = 1|X \setminus \{j,k\})P(X_k = 1|X \setminus \{j,k\})$:

$$
\begin{aligned}
\frac{A_1}{A_1 + A_2 + A_3 + A_4} &= \frac{(A_1 + A_2)(A_1 + A_3)}{(A_1 + A_2 + A_3 + A_4)^2} \\
A_1(A_1 + A_2 + A_3 + A_4) &= (A_1 + A_2)(A_1 + A_3) \\
A_1 A_4 &= A_2 A_3 \\
P(X_j = 1, X_k = 1, X \setminus \{j,k\})P(X_j = -1, X_k = -1, X \setminus \{j,k\}) &= \\
P(X_j = 1, X_k = -1, X \setminus \{j,k\})P(X_j = -1, X_k = 1, X \setminus \{j,k\}) \\
\frac{1}{Z} exp(\beta_j + \beta_k + \beta_{jk} + \sum_{i \neq j,k} \beta_i x_i + \sum_{j<k} \beta_{jk} x_j x_k) * \\
\frac{1}{Z} exp(-\beta_j - \beta_k + \beta_{jk} + \sum_{i \neq j,k} \beta_i x_i + \sum_{j<k} \beta_{jk} x_j x_k) &= \\
\frac{1}{Z} exp(\beta_j - \beta_k - \beta_{jk} + \sum_{i \neq j,k} \beta_i x_i + \sum_{j<k} \beta_{jk} x_j x_k) * \\
\frac{1}{Z} exp(-\beta_j + \beta_k - \beta_{jk} + \sum_{i \neq j,k} beta_i x_i + \sum_{j<k} \beta_{jk} x_j x_k) \\
2\beta_{jk} &= -2\beta_{jk}
\end{aligned}
$$

This can only be true if $\beta jk = 0$ for $j \neq k$.

Now we prove the opposite direction: $\beta_{jk} = 0 \to X_j \perp X_k | X \setminus \{j,k\}$. Without loss of generality, we let $j = 1$ and $k = 2$. Since X_j and X_k are conditionally independent given $X \setminus \{j,k\}$, $P(X_j = x_j, X_k = x_k|X \setminus \{j,k\}) = P(X_j = x_j|X \setminus \{j,k\})P(X_k = x_k|X \setminus \{j,k\})$.

Let $A = P(X_3 = x_3, X_4 = x_4...X_d = x_d)$

Let $K = P(X_1 = x_1, X_2 = x_2...X_d = x_d)$

Let $L = P(X_1 = x_1, X_3 = x_3...X_d = x_d)$

Let $M = P(X_2 = x_2, X_3 = x_3...X_d = x_d)$.

$$P(X_j = x_j, X_k = x_k | X \setminus \{j,k\}) = P(X_j = x_j | X \setminus \{j,k\}) P(X_k = x_k | X \setminus \{j,k\})$$
$$\frac{K}{A} = \frac{L}{A} * \frac{M}{A}$$
$$AK = LM$$

Let $A_1 = P(X_j = 1, X_k = 1, X \setminus \{j,k\})$

Let $A_2 = P(X_j = 1, X_k = -1, X \setminus \{j,k\})$

Let $A_3 = P(X_j = -1, X_k = 1, X \setminus \{j,k\})$

Let $A_4 = P(X_j = -1, X_k = -1, X \setminus \{j,k\})$

Let $L_1 = P(X_1 = x_1, X_2 = 1, X_3 = x_3...X_d = x_d)$

Let $L_2 = 1 - L1$

Let $M_1 = P(X_1 = 1, X_2 = x_2...X_d = x_d)$

Let $M_2 = 1 - M_1$.

So, $K(A_1 + A_2 + A_3 + A_4) = (L_1 + L_2)(M_1 + M_2)$. This must hold for all possible $x_1, x_2, ..., x_d$. We consider four cases:

Case 1) $x_1 = 1, x_2 = 1$
Then $L_1 = A_1$, $M_1 = A_1$, $L_2 = A_2$, $M_2 = A_3$ and $K = A_1$ for all $x_3...x_d$. So we have $A_1(A_1+A_2+A_3+A_4) = (A_1 + A_2)(A_1 + A_3)$ and thus $A_1A_4 = A_2A_3$.

Case 2) $x_1 = 1, x_2 = -1$
Then $L_1 = A_1$, $M_1 = A_2$, $L_2 = A_2$, $M_2 = A_4$ and $K = A_2$ for all $x_3...x_d$. So we have $A_2(A_1+A_2+A_3+A_4) = (A_1 + A_2)(A_1 + A_4)$ and thus $A_1A_4 = A_2A_3$.

Case 3) $x_1 = -1, x_2 = 1$
Then $L_1 = A_3$, $M_1 = A_1$, $L_2 = A_4$, $M_2 = A_3$ and $K = A_3$ for all $x_3...x_d$. So we have $A_3(A_1+A_2+A_3+A_4) = (A_3 + A_4)(A_1 + A_3)$ and thus $A_1A_4 = A_2A_3$.

Case 4) $x_1 = -1, x_2 = -1$
Then $L_1 = A_3$, $M_1 = A_2$, $L_2 = A_4$, $M_2 = A_4$ and $K = A_4$ for all $x_3...x_d$. So we have $A_4(A_1+A_2+A_3+A_4) = (A_3 + A_4)(A_2 + A_4)$ and thus $A_1A_4 = A_2A_3$.

In all four cases, it is sufficient to show $A_1A_4 = A_2A_3$ to show that $AK = LM$ for all $x_1...x_d$. Thus, if $A_1A_4 = A_2A_3$, then $X_j \perp X_k | X \setminus \{j,k\}$ since $\beta_{jk} = 0$ as we showed in the first part of 4.1.

4.2

Let $X \setminus \{j\} = X_{j^c}$.

$$
\begin{aligned}
P(X_j = 1 | X_{j^c} = x_{j^c}) &= \frac{P(X_j = 1, X_{j^c} = x_{j^c})}{P(X_j = 1, X_{j^c} = x_{j^c}) + P(X_j = -1, X_{j^c} = x_{j^c})} \\
&= \frac{1}{1 + \frac{P(X_j = -1, X_{j^c} = x_{j^c})}{P(X_j = 1, X_{j^c} = x_{j^c})}} \\
&= \frac{1}{1 + \frac{A}{B}} \\
A &= \frac{1}{Z} exp(\sum_{i \neq j} \beta_i x_i + \sum_{i<k, a \neq j} \beta_{ik} x_i x_k - \beta_j - \sum_j \beta_{jk} x_k) \\
B &= \frac{1}{Z} exp(\sum_{i \neq j} \beta_i x_i + \sum_{i<k, a \neq j} \beta_{ik} x_i x_k + \beta_j + \sum_j \beta_{jk} x_k)
\end{aligned}
$$

So $\frac{1}{1+\frac{A}{B}} = \frac{1}{1+exp(-2(\beta_j + \sum_{j \neq k} \beta_{jk} x_k))}$