# Modern Optimization Methods for Big Data Problems
# ASSIGNMENT 1: Randomized Gossip Algorithms

**DEADLINE: Sunday, February 5, 2017 @ 22:00pm**

January 22, 2017

*Instructions:* Submit your solution electronically via Learn. Your submission should be a single Julia notebook, i.e., a single .ipynb file. Use a style similar to how the labs are written. Include explanatory text where needed. This is important and will be marked as well. Use any Julia kernel that works for you to produce the file. Start the notebook with a markdown cell containing the text

```
## Assignment 1 -- NAME SURNAME -- UUN
```

Clearly mark the start of each of the 8 problems by inserting a markdown cell with the text

```
## Problem X
```

where $X = 1, 2, \ldots, 8$.

This assignment is worth one third out of 50% towards your final mark in the course. You may discuss the problems with others if you wish, but you may not copy any code nor solution. Only verbal discussion is allowed.

1. [3 marks] Let $G = (V, E)$ be an undirected graph with nodes $V$ and edges $E$. The *adjacency matrix* of $G$ is the $n \times n$ symmetric matrix $B$ defined by

$$B_{ij} = \begin{cases} 1 & \text{if } \{i,j\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

   By $G(n, p) = (V, E)$ we denote the *Erdös-Rényi graph* with $n$ vertices ($V = \{1, 2, \ldots, n\}$) and a random set of edges $E$ formed via the following mechanism: for each pair of distinct vertices $\{i, j\} \in V \times V$, the probability that $e = \{i, j\} \in E$ is equal to $p$, independently from other edges. Clearly, as $p$ increases, the graph is more likely to contain more edges. If $p = 1$, then tall pairs of vertices are joined by an edge (i.e., the graph is *complete*).

   Write function `ER(n,p)`, with inputs $n$ and $p$, whose output is the adjacency matrix of $G(n, p)$.

2. [3 marks] Write function `Union(B1,B2)`, whose inputs $B1, B2$ are the adjacency matrices of some graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$, respectively, where $V = \{1, 2, \ldots, n\}$, and whose output is the adjacency matrix of the graph $G = (V, E_1 \cup E_2)$.

3. [3 marks] Write function `Cycle(n)`, whose output is the adjacency matrix of the cycle graph on $n$ vertices, typically denoted in the literature as $C_n = (V, E)$. This is the graph with $V = \{1, 2, \ldots, n\}$ and $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \ldots, \{n - 1, n\}, \{n, 1\}\}$.

4. [3 marks] Write function `AfromB(B)`, whose input is an adjacency matrix of a graph $G = (V, E)$ with $n$ vertices, and whose output is matrix $A \in \mathbf{R}^{|E| \times n}$, with entries in the set $\{0, 1, -1\}$, with rows corresponding to the edges of $G$. In particular, row $e = \{i, j\} \in E$ of the linear system $Ax = 0$, i.e., the equation $A_{e:}x = 0$, should represent the constraint $x_i = x_j$. Thus, the system $Ax = 0$ simply says that $x_i = x_j$ whenever nodes $i$ and $j$ are joined by an edge in $G$.

Recall that when the randomized Kaczmarz method is applied to the system $Ax = 0$, where $A$ is associated with a connected graph $G$ as explained above, and the initial iterate $x^0 = c$ is used, then $x_i^t \to \frac{1}{n}\sum_{i=1}^n c_i$ for all $i = 1, 2, \dots, n$ (in a probabilistic sense). That is, the values at all nodes converge to the average of the initial values. We have seen in the class that the RK method can then be interpreted as a standard *randomized gossip* algorithm: in each iteration we pick a random edge $i, j \in E$ and set $x_i^{t+1} \leftarrow \frac{x_i^t + x_j^t}{2}$, $x_j^{t+1} \leftarrow \frac{x_i^t + x_j^t}{2}$ and $x_l^{t+1} \leftarrow x_l^t$ for all other $l$.

5. [3 marks] Write function `rate(A)` whose input is matrix $A \in \mathbf{R}^{m \times n}$ and whose output is the scalar

$$\rho = 1 - \frac{\lambda_{\min}^+(A^\top A)}{\|A\|_F^2},$$

where $\|A\|_F$ is the Frobenius norm of $A$ and $\lambda_{\min}^+(C)$ denotes the smallest nonzero eigenvalue of matrix $C$. It turns out that $\rho$ is precisely the rate of the randomized Kaczmarz method applied to the system $Ax = b$ (indeed, the rate does not depend on $b$).

6. *[3 marks] The value $\rho$ represents the rate. However, the derived value $1/(1 - \rho)$ also has a very natural interpretation. Think about what it means and provide an explanation. You may give a theoretical explanation, or one you arrived at through numerical experimentation with the RK method.

7. [3 marks] Write a piece of code which lists the values $1/(1-\rho)$ for matrices $A$ formed as the output of the function `AfromB(B)` for input matrices $B$ which are the adjacency matrices of the graphs $G(n, p) \cup C_n$ for $n = 100$ and $p = 0.01, 0.02, \dots, 0.99, 1.00$. Do you see a pattern in the results? Comment on this.

8. *[4 marks] Let $\mathcal{M}$ be the set of matrices $A$ with $n$ columns (we do not restrict the number of rows) with the property that the solution space of the linear system $Ax = 0$ consists precisely of multiples of the vector $e = (1, 1, \dots, 1) \in \mathbf{R}^n$. There are many matrices with this property; one such matrix was described in Problem 4. As we have seen, application of the RK method to the system $Ax = 0$, started from $x^0 = c \in \mathbf{R}^n$, resulted in a randomized gossip algorithm for computing the average of the values $c_1, \dots, c_n$.

   The set $\mathcal{M}$ is important because for each $A \in \mathcal{M}$, the system $Ax = 0$ represents a way to encode the constraint $x_1 = x_2 = \cdots = x_n$. Hence, for each $A \in \mathcal{M}$, applying the RK method to the system $Ax = 0$, started from $x^0 = c$, leads to a different algorithm for computing the average of the private values $c_1, \dots, c_n$.

   (a) Fix $n = 100$ and $c = (1, 2, 3, \dots, 100) \in \mathbf{R}^{100}$. By numerical experimentation, suggest which matrices $A \in \mathcal{M}$ are best for computing the average of $c_i$ via RK in terms of the computational time needed by the RK method to output a solution of certain fixed accuracy (say $\|x - c\| \leq 10^{-2}$, where $\|s\| = \sqrt{\sum_i s_i^2}$ is the standard Euclidean norm). It is best to use a dedicated RK solver for these experiments. Comment on your findings.

   (b) Find $A \in \mathcal{M}$ such that the RK method, when applied to the system $Ax = 0$, started from $x^0 = c$, has a natural and interesting interpretation. Be creative. Explain why is the method you found interesting.