Problem 7: Friends and Smokers

## Friends and Smokers (Undirected Graphical Model)

**Given:**

An undirected graph in which the nodes represent people and an edge connects two people if they are friends (assumed to be symmetric). We are told that all else being equal, friends are 3 times more likely to have the same smoking habit (both be smokers or both be non-smokers) as to have different smoking habits). The prior probability of being a smoker is 0.2.

The observed smoking habit status of a subset of the nodes in the graph

**Find:**

Query 1: For each of the unobserved nodes, the marginal posterior probability that that person is a smoker.

**Metrics:**

Metric 1: The sum, over all of the queried nodes, of the absolute difference between the computed and the true posterior probability.

Consider the following graph (available as problem-7-friends.csv):



Define a Markov Random Field as follows:

People have a prior probability of smoking of 0.2

All else being equal, it is 3 times as likely that two friends have the same smoking habit than different.

Query: Compute the marginal probability of smoking for each of the unobserved nodes given the six observed nodes:

|  |  |
| --- | --- |
| Node | Observation of Smokes |
| 1 | 1 |
| 2 |  |
| 3 |  |
| 4 | 0 |
| 5 |  |
| 6 | 0 |
| 7 |  |
| 8 | 1 |
| 9 | 1 |
| 10 |  |
| 11 |  |
| 12 |  |
| 13 |  |
| 14 |  |
| 15 | 1 |
| 16 |  |
| 17 |  |
| 18 |  |
| 19 |  |
| 20 |  |

Additional graphs can be generated using the code in problem-7-generator.R. This file also contains a very slow implementation of Gibbs sampling as a baseline method.

Query 1: For each of the unobserved nodes, compute the posterior marginal probability that that person is a smoker.

Metric 1: The sum, over all of the queried nodes, of the absolute difference between the computed and the true posterior probability.