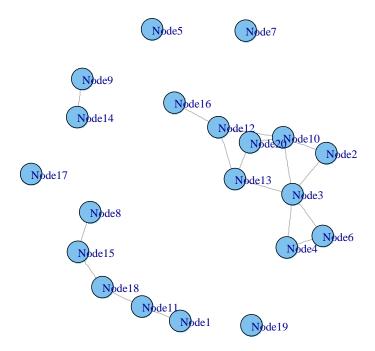# Problem 7: Friends and Smokers

Introduction

This problem involves 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.

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.

$$P(S) \propto \exp\left[\sum_i \phi(s_i) + \sum_{(i,j) \in Friends} \psi(s_i, s_j)\right]$$

$$\phi(s_i = 1) = \log 0.2 = -1.60943$$

$$\phi(s_i = 0) = \log 0.8 = -0.22314$$

$$\phi(s_i = 0, s_j = 0) = 0.549306$$

$$\phi(s_i = 1, s_j = 1) = 0.549306$$

$$\phi(s_i = 1, s_j = 0) = -0.549306$$

$$\phi(s_i = 0, s_j = 1) = -0.549306$$

Queries and Metrics

Query: Compute the marginal probability of smoking for each of the unobserved nodes given the six observed nodes (which represent the observed smoking habit status):

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.

Submission

The metric value should be computed for each elapsed time step (by calling the provided code or by implementing yourself). The metric value should be reported for several elapsed time steps. The number of elapsed time steps should be sufficient to establish an "informative profile".

For further details regarding submission of the metric and your code, please refer to the main CP4 problem description document, e.g. PPAML-Challenge-Problem-4.pdf.

Sample output for this problem has been provided in the "sampleoutput" folder:

```
problem-7-query-1-metric-1.csv
```

Notes:

Further details on this problem can be found in the provided sample solution, e.g.

```
ppaml-cp4/solutions/problem7
```