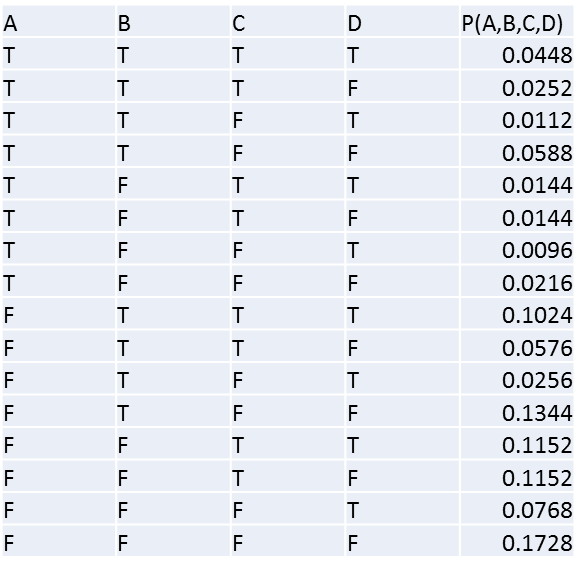
**NAME:**

**COLLABORATOR(S):**

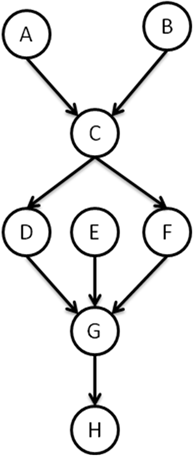
**CS 583 – Assignment 1**

1. We are given the following joint distribution over the random variables A, B, C, and D. Please answer the following questions. Feel free to use the attached excel file.



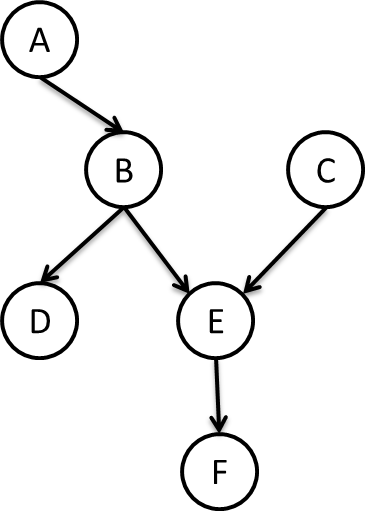
* 1. Are A and B independent?
  2. Are A and C independent?
  3. Are A and C independent given B?
  4. Are A and D independent?
  5. Are A and D independent given C?

1. We are given the random variables *X*2, *X*3, …, *Xn*, and *Y*2, *Y*3, …, *Y*m. Answer the following questions.
   1. Assuming every variable is binary, how many independent parameters are needed to represent P(*X*2, *X*3, …, *Xn*, *Y*2, *Y*3, …, *Ym*)?
   2. Assuming every variable has three possible values, how many independent parameters are needed to represent P(*X*2, *X*3, …, *Xn*, *Y*2, *Y*3, …, *Ym*)?
   3. Assuming each *Xi* has *i* possible values and similarly every *Yi* has *i* possible values, how many independent parameters are needed to represent P(*X*2, *X*3, …, *Xn*, *Y*2, *Y*3, …, *Ym*)?
   4. Assuming every variable is binary, how many independent parameters are needed to represent P(*Y*2, *Y*3, …, *Ym* | *X*2, *X*3, …, *Xn*)?
   5. Assuming every variable has three possible values, how many independent parameters are needed to represent P(*Y*2, *Y*3, …, *Ym* | *X*2, *X*3, …, *Xn*)?
   6. Assuming each *Xi* has *i* possible values and similarly every *Yi* has *i* possible values, how many independent parameters are needed to represent P(*Y*2, *Y*3, …, *Ym* | *X*2, *X*3, …, *Xn*)?
2. We have a domain with *n* variables, each of which are binary. One of them is the class variable and the rest *n*-1 are feature variables. How many independent parameters are need for a naïve Bayes representation?
3. We are given the following the Bayesian network. Please answer the following questions.



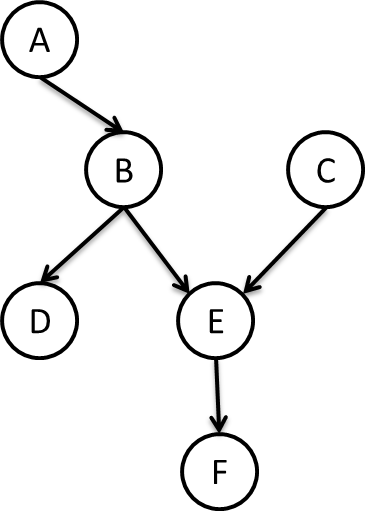
* 1. Write down the joint distribution as a factorization over this Bayesian network.
  2. Assuming each variable is discrete and can take *n* possible values, how many independent parameters are needed for this Bayesian network?
  3. Are the following independence statements true or false?
     1. A ⊥ B
     2. A ⊥ B | C
     3. A ⊥ B | F
     4. A ⊥ B | G
     5. A ⊥ B | E
     6. A ⊥ B | H
     7. A ⊥ H
     8. A ⊥ H | F
     9. A ⊥ H | D, F
     10. D ⊥ F
     11. B ⊥ E
     12. B ⊥ E | F
     13. B ⊥ E | F, H

1. We have a P over the variables A, B, C, D, E, and F. We would like to build a Bayesian network that is a minimal I-Map for P. In reality, you have access to P, which you can query for independencies, but for the purposes of this problem, we will assume P has the following structure. Create a minimal I-Map for P, using the following variable orders.



* 1. C, A, B, E, D, F
  2. D, B, A, E, C, F
  3. F, E, D, C, B, A
  4. F, A, C, E, D, B

1. How many I-equivalent structures are there to the following structure? Draw them.



1. Download Hugin Lite from <https://www.hugin.com/index.php/hugin-lite/> and the student network from GitHub <https://github.com/CS583pgm/S2020/blob/master/other/student.net>. Answer the following queries:
   1. P(SAT = low) = ?
   2. P(SAT = low | Difficulty = difficult) = ?
   3. P(SAT = low | Grade = a) = ?
   4. P(SAT = low | Grade = a, Difficulty = difficult) = ?
   5. P(SAT = low | Intelligence = low) = ?
   6. P(SAT = low | Grade = a, Difficulty = difficult, Intelligence = low) = ?
2. Find a dataset that is interesting to you at <https://www.openml.org/>. Learn the structure and parameters of a Bayesian network using Hugin. Discretize the features as needed during the process. Note that Hugin Lite is limited to 50 variables and 500 cases, which means you may need to subsample the data before you feed it to Hugin. Run a few probabilistic queries and report them. In your brief report, describe the dataset and the domain, and include a screenshot of the structure. Save your Hugin file as a .net file and attach it.