

Your Name:

Your Andrew ID:

Homework 1

Collaboration and Originality

1. Did you receive help of any kind from anyone in obtaining your data for this assignment (Yes or No)?
It is not necessary to describe discussions with the instructor or TA.

If you answered Yes, provide the name(s) of anyone who provided help, and describe the type of help that you received.

2. Did you give help of any kind to anyone in obtaining their data for this assignment (Yes or No)?

If you answered Yes, provide the name(s) of anyone that you helped, and describe the type of help that you provided.

3. Are you the author of every word of your report (Yes or No)?

If you answered No:

- a. identify the text that you did not write,
- b. explain where it came from, and
- c. explain why you used it.

Your Name:

Your Andrew ID:

1 Experiment #1: Baseline 1

Entities (E_1, E_2)	Frequency			PMI	Phi-Square
	E_1	E_2	E_1, E_2		
Jamie Callan, Grace Hui Yang					
K Callan, Bruce Croft					
Jamie Callan, Dean Cain					
Teri Hatcher, Dean Cain					
Cylvia Hayes, Margi Hoffmann					

Sort all tables in descending order based on PMI

Experiment 1 investigates whether PMI and/or Phi-Square identify meaningful relationships among pairs of entities. Discuss and provide an analysis of your experimental results. You may wish to consider some of the issues listed below (i.e., these are example issues, not required issues).

- Do PMI and Phi-Square identify meaningful relationships among pairs of entities? What is the nature of the (real world) relationships that you observed?
- Do the metrics behave as you expected? If the metrics make mistakes, what might have caused the behavior that you observed?
- Do the metrics tend to agree or disagree? When they disagree, do you understand why?
- Do you consider one metric more effective in general, or more effective for high- or low-frequency entities?

Feel free to discuss other issues of your own choosing. The goal is for you to show that you have thought about and tried to analyze the experimental results that you obtained.

You should spend a few minutes determining whether there is actually a relationship between the two entities and what it might be (e.g., by doing a web search and examining a few documents). High PMI or phi-square doesn't necessarily mean that the pair of entities is related in the way that you expect.

2 Experiment #2: Baseline 2

Frequency of "mobile advertising" (T):

Entity E_2	Frequency		PMI	Phi-Square
	E	T, E		
Google				
Facebook				
Madvertise Media				
Sizmek				
Aditic				

Experiment 2 investigates whether PMI and/or Phi-Square identify meaningful relationships between topics and entities. Discuss and provide an analysis of your experimental results. You may wish to consider some of the issues listed below (i.e., these are example issues, not required issues).

- Do PMI and Phi-Square identify meaningful relationships among topics and entities? What is the nature of the (real world) relationships that you observed?
- Do the metrics behave as you expected? If the metrics make mistakes, what might have caused the behavior that you observed?
- Do the metrics tend to agree or disagree? When they disagree, do you understand why?
- Do you consider one metric more effective in general, or more effective for high- or low-frequency entities?
- Did the metrics behave the same way in Experiments 1 or 2, or are there differences? If there are differences, what might have caused them?

Feel free to discuss other issues of your own choosing. The goal is for you to show that you have thought about and tried to analyze the experimental results that you obtained.

3 Experiment #3: Assessing Relationships Between People

Entity 1

Entity name (E_1):

Frequency (E_1):

Why you picked this entity:

Sort this table in descending order of PMI					
	Entity E_2	Frequency		PMI	Phi-Square
		E_2	E_1, E_2		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Entity 2

Entity name (E_1):

Frequency (E_1):

Why you picked this entity:

Sort this table in descending order of PMI					
	Entity E ₂	Frequency		PMI	Phi-Square
		E ₂	E ₁ , E ₂		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Entity 3

Entity name (E₁):

Frequency (E₁):

Why you picked this entity:

Sort this table in descending order of PMI					
	Entity E ₂	Frequency		PMI	Phi-Square
		E ₂	E ₁ , E ₂		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Analysis

Experiment 3 investigates whether PMI and/or Phi-Square identify meaningful relationships between people. Note that Experiment 3 is similar to Experiment 1, except that in this experiment *you* chose the people, and you investigated a larger set of related people. Discuss and provide an analysis of your experimental results. You may wish to consider some of the issues listed below (i.e., these are example issues, not required issues).

- Discuss whether the behavior across Entity₁, Entity₂, and Entity₃ was similar or different. If there are interesting differences, what might have caused them?

- If the experiment exposed flaws in your choice of seed entities, how should you have chosen the entities now that you know more?
- Discuss how the results of Experiment 1 compare to the results of Experiment 3. If there are differences, describe them and discuss what might have caused them. If there are no differences, does this result provide information about the stability or consistency of the metric(s)?

Feel free to discuss other issues of your own choosing. The goal is for you to show that you have thought about and tried to analyze the experimental results that you obtained.

4 Experiment #4: Assessing Relationships Between Topics and Companies

Topic 1

Topic (T):

Frequency (T):

Why you picked this topic:

Sort this table in descending order of PMI					
	Entity E	Frequency		PMI	Phi-Square
		E	T, E		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Topic 2

Topic (T):

Frequency (T):

Why you picked this topic:

Sort this table in descending order of PMI					
	Entity E	Frequency		PMI	Phi-Square
		E	T, E		
1					
2					
3					

4					
5					
6					
7					
8					
9					
10					

Topic 3

Topic (T):

Frequency (T):

Why you picked this topic:

Sort this table in descending order of PMI					
	Entity E	Frequency		PMI	Phi-Square
		E	T, E		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Analysis

Experiment 4 investigates whether PMI and/or Phi-Square identify meaningful relationships between companies. Note that Experiment 4 is similar to Experiment 2, except that in this experiment *you* chose the people, and you investigated a larger set of related people. Discuss and provide an analysis of your experimental results. You may wish to consider some of the issues listed below (i.e., these are example issues, not required issues).

- Discuss whether the behavior across Entity₁, Entity₂, and Entity₃ was similar or different. If there are interesting differences, what might have caused them?
- If the experiment exposed flaws in your choice of seed entities, how should you have chosen the entities now that you know more?
- Discuss how the results of Experiment 2 compare to the results of Experiment 4. If there are differences, describe them and discuss what might have caused them. If there are no differences, does this result provide information about the stability or consistency of the metric(s)?

Feel free to discuss other issues of your own choosing. The goal is for you to show that you have thought about and tried to analyze the experimental results that you obtained.

5 Experiment 5: An Experiment of Your Own Design

Describe your experiment. For example, discuss the purpose of the experiment, the choices that you made, why you made those choices, and what you expected the results to be.

Describe and analyze your results. Discuss and provide an analysis of your experimental results. This is your experiment, so discuss the issues that make sense for your experiment. You may wish to consider some of the issues listed below (i.e., these are example issues, not required issues).

- Did your original hypothesis or expectations turn out to be correct?
- Did you learn anything about the subject of your experiment that you did not know before?
- Did you learn anything about how to use PMI and Phi-Square for co-occurrence analysis that you did not know before?

We are hoping to see a little creativity and originality in this section.

6 Effectiveness of the Different Types of Queries

Experiments 1-5 use queries to retrieve statistics for people, companies, topics, and co-occurrences. Discuss the effectiveness of the different types of queries in your experiments. You may wish to consider some of the issues listed below (i.e., these are example issues, not required issues).

- Discuss any trends that you observe about difference in the behavior of topics, people, or companies.
- Are there differences in the stability of the different types of queries? Do any of the differences seem significant?
- Did the different types of queries work as you expected, or were there surprises? If some didn't work as you expected, what might be the causes?
- Are there more effective methods of acquiring the statistics needed by PMI and Phi-Square? If so, what are they, and why do you consider them more effective?

Feel free to discuss other issues of your own choosing. The goal is for you to show that you have thought about the experimental results that you obtained.

There is some overlap between this section and observations that you may make in Experiments 1-5. However, in this section you have an opportunity comment on trends that span different experiments, which you can't do as easily in the earlier sections.