SCHELLING SIMULATOR

BLUE VS WHITE COLLAR BREAKDOWNS

BACKGROUND

White-collar workers are commonly known as “suit-and-tie workers”. Typically, white-collar workers work in service industries and often avoid manual labor. The blue-collar stereotype refers to any worker who engages in hard physical labor such as construction, maintenance, or mining.

We have decided to compare workers in New York City, NY to workers in Pittsburgh, PA. Pittsburgh was recently listed as a “blue-collar hotspot”, while New York City is widely known to have a large population of white-collar workers.

Through experimentation using the Schelling model, we will vary grid size, population density, population split representing groups of workers (white vs. blue-collar), and the number of “like” neighbors for workers to be satisfied. Pittsburgh has a 31.5% blue-collar population, while New York City has an 8.8% blue-collar population. We hypothesize that people in New York City with white-collar jobs are also more likely to want their friends to have white-collar jobs compared to those in Pittsburgh.

EXPERIMENT GOAL

The goal of this experiment is to observe how a change in grid size, population density, number of neighbors in a similar career path needed to be satisfied, and population split (reflecting, approximately, the ratio of blue-collar workers to white-collar workers) affects the average homophily ratio across trials.

HYPOTHESIS

With two classes, New York City, NY (larger grid size, denser population, higher number of neighbors in the same career path to be satisfied, 91.2/8.8 population split) will have a higher average homophily ratio than the city of Pittsburgh, PA (smaller grid size, less dense population, lower number of neighbors in the same career path to be satisfied, 68.5/31.5 population split).

METHODS

After continually adjusting the parameters such that the 100 trials can be successfully performed, we came up with the following parameters:

New York City:
Grid Size: 300 x 300
Population Size: 300 x 300 x .5
Number of neighbors in same labor division = 3
Percent Blue Collar Workers: 8.8%
Percent White Collar Workers: 91.2%

Pittsburgh:
Grid Size: 100 x 100
Population Size: 100 x 100 x .35
Number of neighbors in same labor division = 2
Percent Blue Collar Workers: 31.5%
Percent White Collar Workers: 68.5%

In reality, the respective populations of each city are much more dense than what was modeled in the Schelling simulator. For the purpose of this project, we have adjusted the population densities accordingly in order to achieve 100 trials.

In addition, we adapted the populate() method in SchellingSimulator.java to adjust for different population breakdowns between white and blue collar workers. To do this, we assigned each random grid cell to groups 1 and two 2 based on the percentage of a city's population split, thus ensuring that the grid would be populated according to each city's respective breakdown of white and blue collar workers.

Note: We used a seed of 100 each so that our results were consistent each time the simulator was run.

CONCLUSION

Our experiment resulted in an average homophily ratio of ~0.67 for New York City and ~0.44 for Pittsburgh, *validating our hypothesis*.

NEW YORK CITY

```
public SchellingVisualizer() {
    // create a new simulator for New York City
    int populationNYC = 300;
    simulator = new SchellingSimulator( seed: 100,  numClasses: 2,
            (int) (populationNYC * populationNYC * .5),  numNeighbors: 3, populationNYC, populationNYC);
```

```
/**
 * Adds the specified population to the grid according to blue vs white collar population percentages
 */
private void populateBlueWhiteCollar() {
    if (population > width * height)
        throw new RuntimeException("Only " + width * height + " cells exist!");

    for (int i = 0; i < population; i++) {
        int xChoice;
        int yChoice;
        do {
            xChoice = randGen.nextInt(width);
            yChoice = randGen.nextInt(height);

        } while (grid[xChoice][yChoice] != 0);

        double randNum = (double)randGen.nextInt( bound: 10);

        double percentWhiteCollar = 9.12;
        double percentBlueCollar = 10 - percentWhiteCollar;

        if (randNum > percentBlueCollar) {
            grid[xChoice][yChoice] = 1;
        }
        else {
            grid[xChoice][yChoice] = 2;
        }
    }
}
```

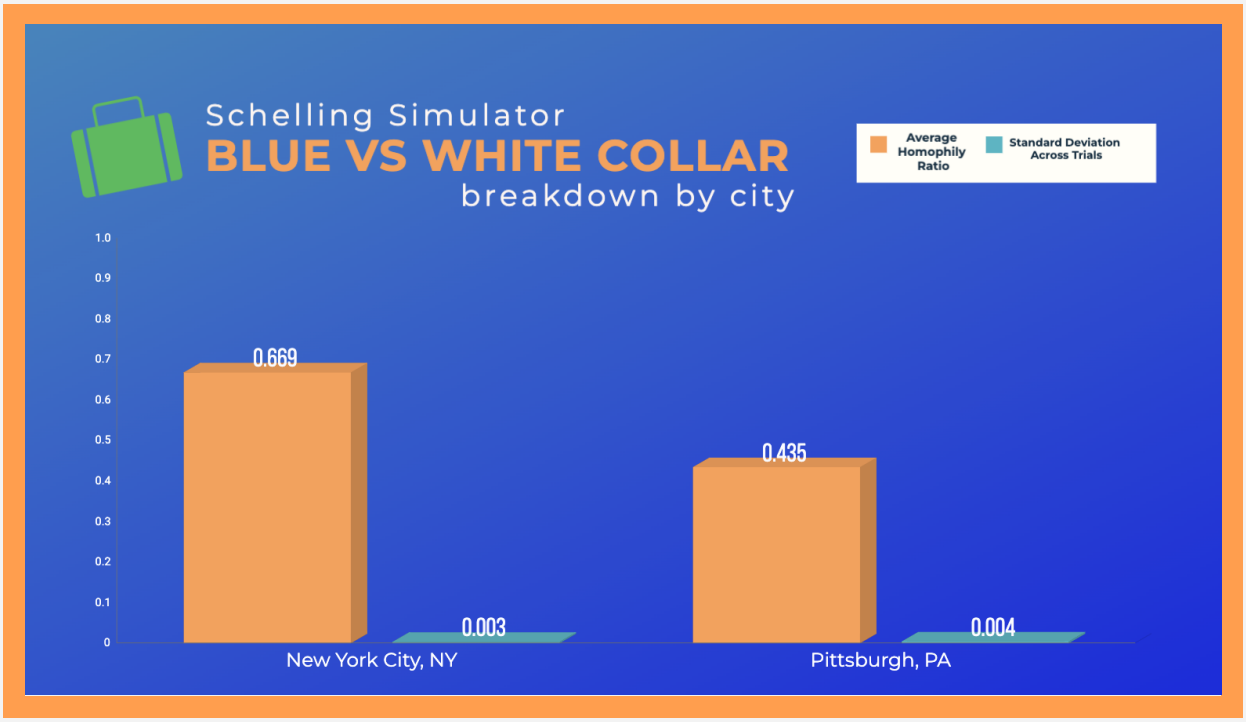
```
Average homophily ratio across trials: 0.6690625691103638
Standard deviation across trials: 0.0033108478079689804
```

PITTSBURGH

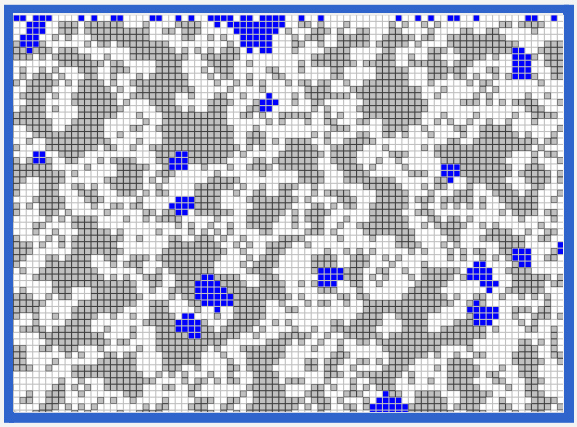
```
public SchellingVisualizer() {
    // create a new simulator for Pittsburgh
    int populationPitts = 100;
    simulator = new SchellingSimulator( seed: 100,  numClasses: 2,
        (int) (populationPitts * populationPitts * .35),  numNeighbors: 2, populationPitts, populationPitts);
```

```
/**
 * Adds the specified population to the grid according to blue vs white collar population percentages
 */
private void populateBlueWhiteCollar() {
    if (population > width * height)
        throw new RuntimeException("Only " + width * height + " cells exist!");

    for (int i = 0; i < population; i++) {
        int xChoice;
        int yChoice;
        do {
            xChoice = randGen.nextInt(width);
            yChoice = randGen.nextInt(height);

        } while (grid[xChoice][yChoice] != 0);

        double randNum = (double)randGen.nextInt( bound: 10);

        double percentWhiteCollar = 6.85;
        double percentBlueCollar = 10 - percentWhiteCollar;

        if (randNum > percentBlueCollar) {
            grid[xChoice][yChoice] = 1;
        }
        else {
            grid[xChoice][yChoice] = 2;
        }
    }
}
```

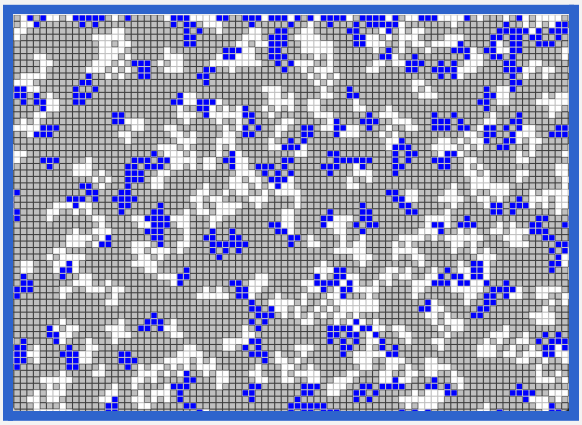
```
Average homophily ratio across trials: 0.43538682479645985
Standard deviation across trials: 0.004315993810076534
```



NEW YORK CITY, NY



PITTSBURGH, PA



Interestingly, in New York City, we found that the average homophily ratio for the blue collar population was slightly higher (~0.698) than for the white collar population (~0.63). Drawing from the real world, we hypothesized that this could be due to two blue collar employees being more similar on average than two white collar employees.

Take, for example, two blue collar professions. A mechanic and an electrician would likely be more similar than a physician and an accountant. We wanted to further investigate this difference in homophily ratios and see if this was true among other pairs of blue vs white collar professions. Thus we decided to perform a second experiment below using the vector space model.

DATA SOURCES

://www.bizjournals.com/pittsburgh/news/2021/05/11/pittsburgh-no1-for-blue-collar-wages.html
https://bluecollarjobs.us/2017/04/10/highest-to-lowest-share-of-blue-collar-jobs-by-state/

VECTOR SPACE MODEL

COSINE SIMILARITY

EXPERIMENT GOALS

The goal of this experiment is to compare blue versus white-collar job descriptions from Indeed.com.

HYPOTHESIS

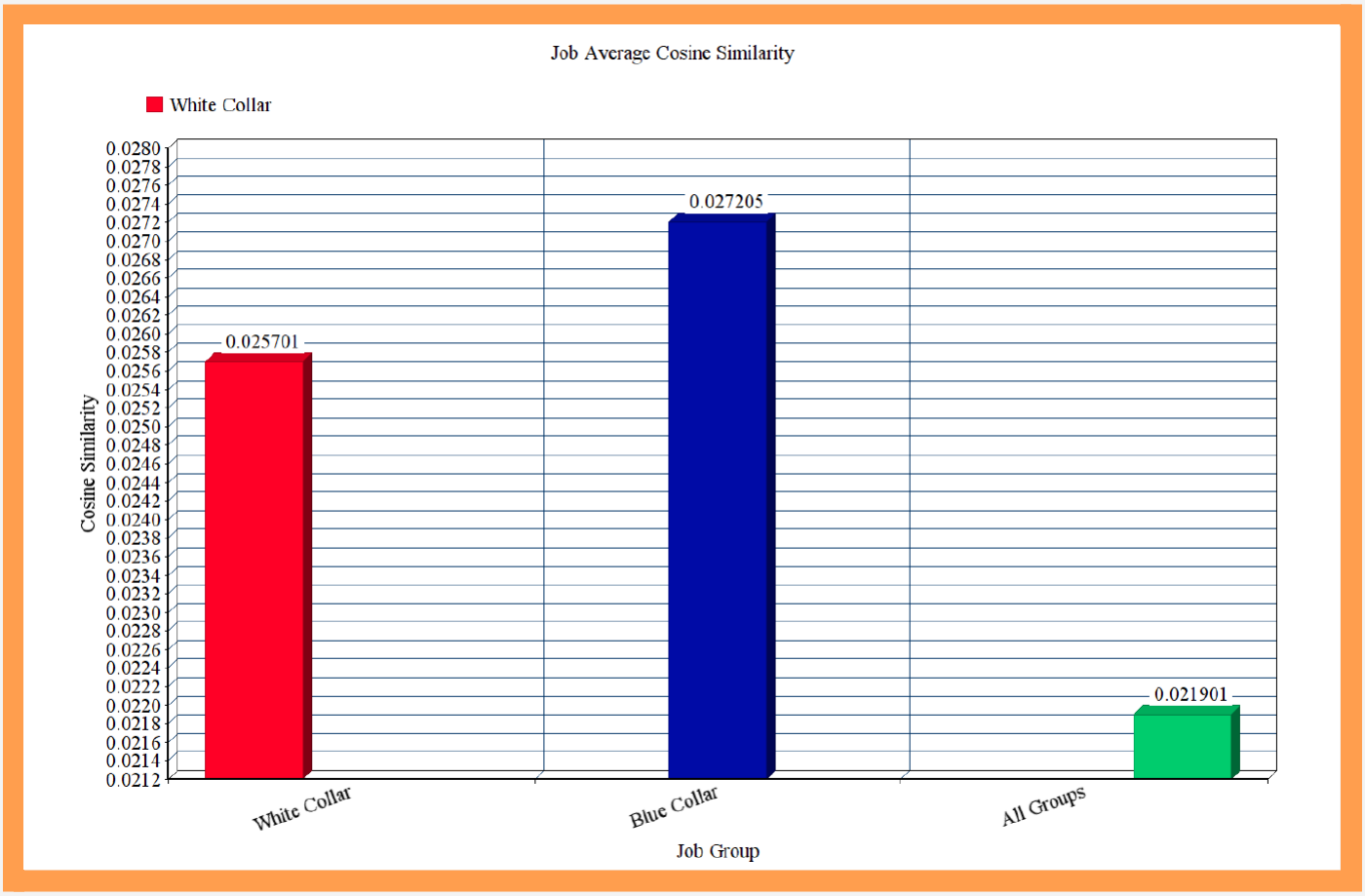
On average, Indeed job descriptions of two blue collar jobs will have a higher cosine similarity score than Indeed job descriptions of two white collar jobs. In other words, two blue collar job descriptions are more similar to each other than two white collar job descriptions on average.

Example: The job description for a mechanic and the job description for an electrician are more similar than the job description for a physician and the job description for an accountant.

CONCLUSION

Our hypothesis was proven because we found the average cosine similarity of blue collar jobs to be ~0.02720 whereas the average cosine similarity of white collar jobs was ~0.02570, *validating our hypothesis.* For comparison, the average cosine similarity of *all* job descriptions was found to be ~0.02190.

GRAPHS



CODE

```
//WHITE COLLAR
Document w1 = new Document( filename: "white_collar_1.txt");
Document w2 = new Document( filename: "white_collar_2.txt");
Document w3 = new Document( filename: "white_collar_3.txt");
Document w4 = new Document( filename: "white_collar_4.txt");
Document w5 = new Document( filename: "white_collar_5.txt");
Document w6 = new Document( filename: "white_collar_6.txt");
Document w7 = new Document( filename: "white_collar_7.txt");
Document w8 = new Document( filename: "white_collar_8.txt");

//BLUE COLLAR
Document b1 = new Document( filename: "blue_collar_1.txt");
Document b2 = new Document( filename: "blue_collar_2.txt");
Document b3 = new Document( filename: "blue_collar_3.txt");
Document b4 = new Document( filename: "blue_collar_4.txt");
Document b5 = new Document( filename: "blue_collar_5.txt");
Document b6 = new Document( filename: "blue_collar_6.txt");
Document b7 = new Document( filename: "blue_collar_7.txt");
Document b8 = new Document( filename: "blue_collar_8.txt");
```

```
ArrayList<Document> documents = new ArrayList<>();
//Adding white collar
documents.add(w1);
documents.add(w2);
documents.add(w3);
documents.add(w4);
documents.add(w5);
documents.add(w6);
documents.add(w7);
documents.add(w8);
//Adding blue collar
documents.add(b1);
documents.add(b2);
documents.add(b3);
documents.add(b4);
documents.add(b5);
documents.add(b6);
documents.add(b7);
documents.add(b8);

Corpus corpus = new Corpus(documents);
VectorSpaceModel vectorSpace = new VectorSpaceModel(corpus);
```

```
double avgWhiteCollarSimilarity = 0.0;
int count1 = 0;
double avgBlueCollarSimilarity = 0.0;
int count2 = 0;
double avgDifferenceSimilarity = 0.0;
int count3 = 0;
```

```
//Comparing all white collar descriptions with each other
for (int i = 0; i < 7; i++) {
    for (int j = i + 1; j < 7; j++) {
        Document doc1 = documents.get(i);
        Document doc2 = documents.get(j);
        System.out.println("\nComparing " + doc1 + " and " + doc2);
        double cosineSim = vectorSpace.cosineSimilarity(doc1, doc2);
        System.out.println(cosineSim);
        avgWhiteCollarSimilarity += cosineSim;
        count1++;
    }
}
```

```
//Comparing all blue collar descriptions with each other
for (int i = 7; i < documents.size(); i++) {
    for (int j = i + 1; j < documents.size(); j++) {
        Document doc1 = documents.get(i);
        Document doc2 = documents.get(j);
        System.out.println("\nComparing " + doc1 + " and " + doc2);
        double cosineSim = vectorSpace.cosineSimilarity(doc1, doc2);
        System.out.println(cosineSim);
        avgBlueCollarSimilarity += cosineSim;
        count2++;
    }
}
```

```
//Comparing all white collar descriptions and blue collar descriptions
for (int i = 0; i < 7; i++) {
    for (int j = 7; j < documents.size(); j++) {
        Document doc1 = documents.get(i);
        Document doc2 = documents.get(j);
        System.out.println("\nComparing " + doc1 + " and " + doc2);
        double cosineSim = vectorSpace.cosineSimilarity(doc1, doc2);
        System.out.println(cosineSim);
        avgDifferenceSimilarity += cosineSim;
        count3++;
    }
}
```

```
avgWhiteCollarSimilarity = avgWhiteCollarSimilarity/count1;
avgBlueCollarSimilarity = avgBlueCollarSimilarity/count2;
avgDifferenceSimilarity = avgDifferenceSimilarity/count3;

System.out.println("~~~~~~~~~~~~~~~~~~~~~~~~~~~~");
System.out.println("Average White Collar Similarity: "+avgWhiteCollarSimilarity);
System.out.println("Average Blue Collar Similarity: "+avgBlueCollarSimilarity);
System.out.println("Average Similarity between the groups: "+avgDifferenceSimilarity);
```

DATA SOURCES

Data is a set of .txt files with job descriptions from Indeed.com for various blue and white collar jobs (8 blue and 8 white)
https://www.indeed.com/career-advice/finding-a-job/difference-between-blue-and-white-collar-jobs