Exercise 3 - Analysis of examiner demographics and advice networks at the USPTO

Sheida Majidi

2024-03-31

Introduction

In this analysis, I delve into the demographics of patent examiners and their advice networks within the United States Patent and Trademark Office (USPTO).

Data loading and preprocessing

First, I load the data and prepare it for the analysis:

```
applications = read_csv("/Users/sheidamajidi/Desktop/Winter2024/Winter2024-2/ORGB672/
Exercises/Exercise 3/app_data_starter.csv")

## Rows: 2018477 Columns: 21
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (11): application_number, examiner_name_last, examiner_name_first, exam...
## dbl   (5): examiner_id, examiner_art_unit, appl_status_code, tc, tenure_days
## date  (5): filing_date, patent_issue_date, abandon_date, earliest_date, late...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

edges = read_csv("/Users/sheidamajidi/Desktop/Winter2024/Winter2024-2/ORGB672/Exercis
es/Exercise 3/edges_sample.csv")

## Rows: 32906 Columns: 4
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (1): application_number
## dbl  (2): ego_examiner_id, alter_examiner_id
## date (1): advice_date
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Demographic analysis

Gender estimation

I'll estimate the gender of examiners based on their first names to explore gender diversity.

Race estimation

Next, I'll estimate the racial background of the examiners using their last names.

```
## Predicting race for 2020

## Proceeding with last name predictions...

## ℹ All local files already up-to-date!

## 701 (18.4%) individuals' last names were not matched.
```

Tenure calculation

Tenure at the organization could provide insights into experience levels and employee turnover.

```
## Warning: There were 2 warnings in `mutate()`.
## The first warning was:
## ℹ In argument: `appl_status_date = ymd_hms(appl_status_date)`.
## Caused by warning:
## ! All formats failed to parse. No formats found.
## ℹ Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

Workgroup analysis

I'll focus on two workgroups for a detailed comparison of their demographics.

Network analysis

Advice network creation

Now, let's examine the advice networks to understand the interaction patterns among examiners.

```
## Warning in graph_from_data_frame(d = edges_df, vertices = vertices_df, directed
## = TRUE): In `d' `NA' elements were replaced with string "NA"

## Warning in graph_from_data_frame(d = edges_df, vertices = vertices_df, directed
## = TRUE): In `vertices[,1]' `NA' elements were replaced with string "NA"

## [1] 4212
```

Centrality scores

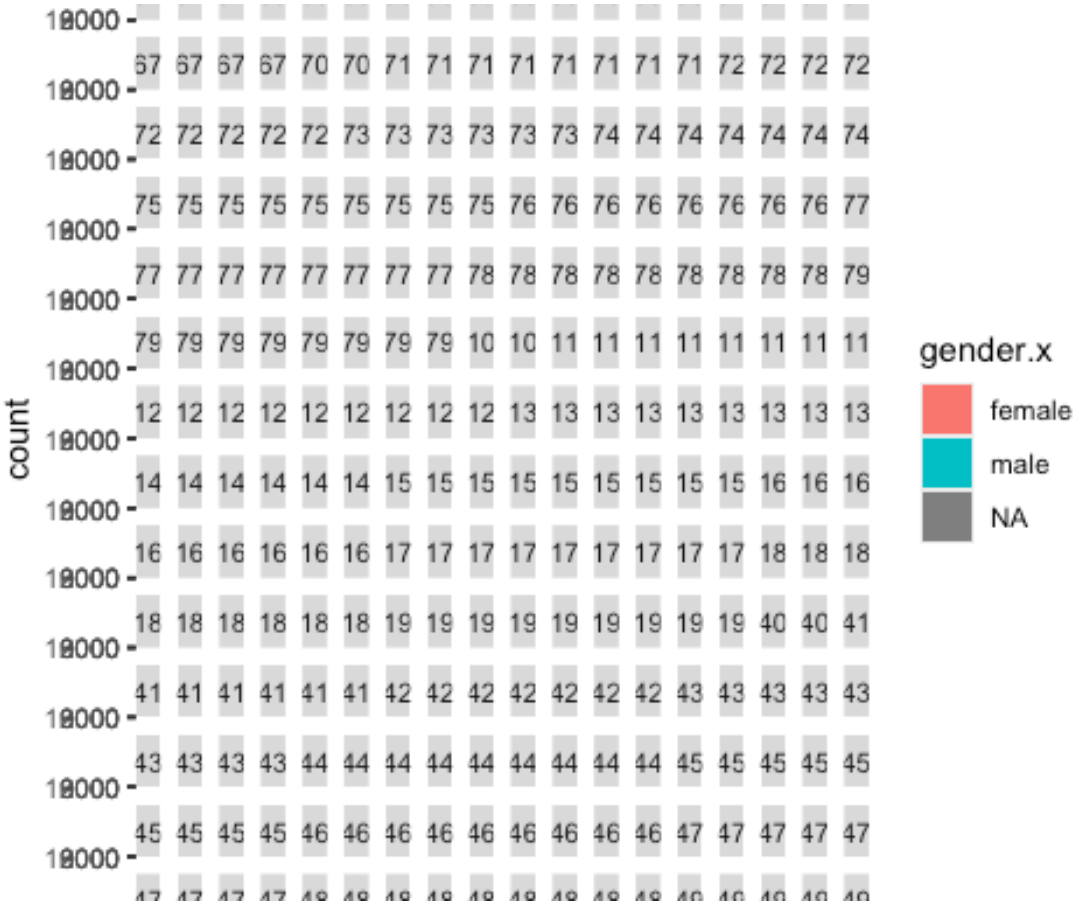
Centrality scores will help us identify key influencers and collaborators in the network.

```
## Warning in data.frame(examiner_id = as.numeric(V(advice_network)$name), : NAs
## introduced by coercion
```

Results and discussion

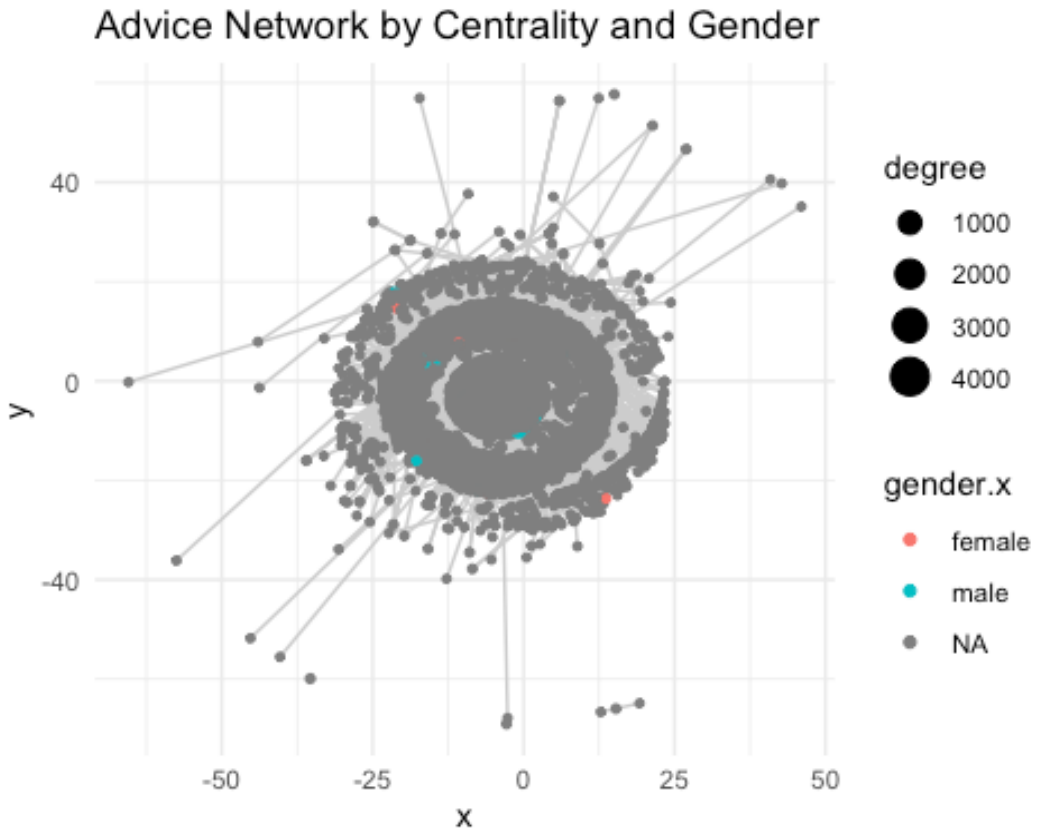
Let's delve into the results to uncover insights about the demographics and network dynamics of USPTO examiners.

Demographic distribution



Network visualization

Visualizing the advice network to see how examiners interact within and across workgroups.



Demographic statistics

```
## # A tibble: 3 × 2
##   gender.x   Count
##   <chr>      <int>
## 1 female    571227
## 2 male     1143391
## 3 <NA>      303859

## # A tibble: 5 × 2
##   race.x     Count
##   <chr>      <int>
## 1 Asian     591644
## 2 Hispanic   58856
## 3 black      89559
## 4 other       1891
## 5 white    1276527
```

gender.x <chr>	Count <int>
female	571227
male	1143391
NA	303859

race.x <chr>	Count <int>
Asian	591644
Hispanic	58856
black	89559
other	1891
white	1276527

5 rows

Network centrality measures

```
##   Average_Degree Median_Degree Average_Closeness Median_Closeness
## 1       15.62488             4         0.2310755     0.0002387205
##   Average_Betweenness Median_Betweenness
## 1            868.2317                  0
```

Conclusion

Taking a look at the USPTO's data, it's clear we've got a mix of examiners from various racial backgrounds, with white and Asian colleagues being particularly prevalent. There's room for growth in representation across all groups, especially among Hispanic and Black communities. The gender balance also leans more towards men, and there's a noticeable portion of the data where gender details are missing. It would be beneficial to delve into why that's the case and work towards a more balanced representation.

When it comes to the advice network within the organization, it's interesting to see how knowledge and support flow between colleagues. Many have a small, close-knit circle of contacts they interact with, but a few key individuals stand out as central nodes, forming the backbone of the network. These pivotal players are integral to the spread of information, although it's worth noting that the overall connectivity could be improved. The majority of examiners aren't regularly bridging communication between others, with only a select few serving as frequent connectors.

Overall, these insights highlight an opportunity to foster stronger connections across the board, encouraging a richer tapestry of interaction and collaboration. Enhancing the network and striving for greater diversity could lead to a more dynamic and inclusive environment, which I believe could greatly benefit our collective creativity and efficiency.