UNITED STATES
PATENT AND TRADEMARK OFFICE
uspto

McGill | DESAUTELS

Optimizing Patent Examination @ the United States Patent and Trademark Office

An Organizational Networks Analysis

ORGB 672 | Organizational Networks
Final Presentation | Group Project

Group 3

Jennifer Liu
Yu Lu
Meriem Mehri
Moiz Shaikh
Ammad Sohail

DATE: APRIL 23, 2024
PREPARED FOR PROF. ROMAN GALPERIN
WINTER 2024 (2)

McGill | DESAUTELS

UNITED STATES PATENT AND TRADEMARK OFFICE uspto

Problem Statement

The USPTO faces challenges in managing patent examination durations and ensuring fairness. Our project employs advanced network and demographic analyses to address these issues.

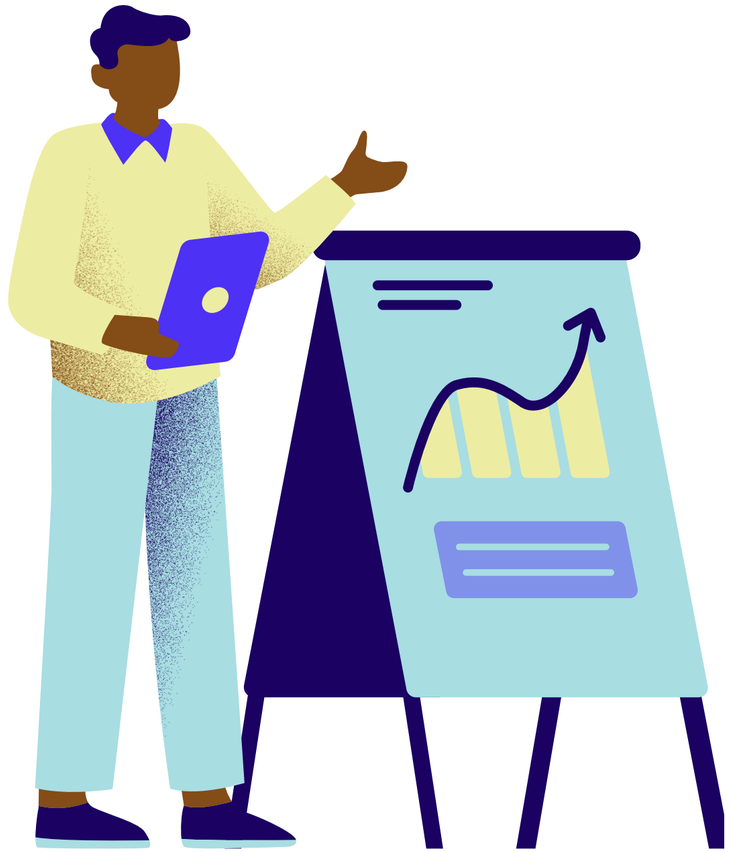
Project Objectives

- **Assess the impact of network dynamics and examiner collaboration on patent processing times.**
- **Examine the role of race, gender, and ethnicity in examination procedures for identifying biases.**
- **Formulate targeted strategies to enhance process efficiency and uphold equitable practices.**

Using data on over 10,000 patent examiners, we apply statistical and machine learning techniques to explore how social structures and demographic factors influence examination outcomes. Our goal is to enhance USPTO's operational efficiency and fairness, ultimately supporting innovation and economic growth.

Agenda

1. In-Depth Overview of the Analytical Approach
2. Influence of Social & Organizational Dynamics on Processing Times
3. Network Structure's Effect on Patent Processing Efficiency
4. Examining the Impact of Race & Ethnicity on Patent Review
5. Strategic Insights & Business Recommendations for the USPTO



McGill | DESAUTELS

UNITED STATES PATENT AND TRADEMARK OFFICE uspto

Data Sources Overview

Primary Data Source: U.S. Patent and Trademark Office

Analysis Period: Patent Applications from the Year 2000 Onwards*

**filling dates*

Data Structure

The data includes detailed records from the USPTO, primarily focused on examiners and patent applications. Sources are diverse, encompassing various public datasets, with modifications for manageability. The data focuses on specific TCs (1600, 1700, 2100, 2400) and includes 15 key variables relevant to our analysis.

- **Composition: Each row represents a single (individual) patent application.**
- **Details Included:**
  - *Application Number, Application Outcomes, Filing Date*
  - *Examiner's Details (inferred demographic attributes: gender; race)*
  - *Classification Details (USPC class and subclass), Patent Number, Patent Issue Date*
  - *Abandon Date, Disposal Type, Application Status Code & Date, Technology Center (tc)*

Figure 1. Data Dictionary: Detailed Variable Descriptions for USPTO Patent Dataset

Table A-1: List of variables included in application_data

Variable Name	Description	Type	Formatting
application_number	Application Number	str14	%14s
filing_date	Filing or 371 (c) Date	float	%td
invention_subject_matter	Invention Subject Matter	str3	%-3s
application_type	Application Type	str7	%-7s
examiner_name_last	Examiner's Family Name	str17	%-20s
examiner_name_first	Examiner's Given Name	str12	%-20s
examiner_name_middle	Examiner's Middle Name	str12	%-20s
examiner_id	Unique Examiner Identifier	str5	%9s
examiner_art_unit	Group Art Unit	str6	%-6s
uspc_class	Invention U.S. Classification	str3	%-3s
uspc_subclass	Invention U.S. Subclassification	str6	%-6s
confirm_number	Confirmation Number	int	%12.0f
customer_number	Customer number	str6	%-6s
atty_docket_number	Attorney Docket Number	str25	%-20s
appl_status_code	Application Status Code	int	%8.0f
appl_status_date	Status Date	float	%td
file_location	Location (where the file currently is)	str5	%-5s
file_location_date	Location Date	int	%td
earliest_pgpub_number	Earliest Publication No.	str15	%-15s
earliest_pgpub_date	Earliest Publication Date	int	%td
wipo_pub_number	WIPO Publication Number	long	%12.0f
patent_number	Patent Number	str7	%-10s
patent_issue_date	Issue Date of Patent	float	%td
abandon_date	Date of Abandonment	float	%td
disposal_type	Disposal Type	str4	%9s
invention_title	Title of Invention	str600	%-20s
small_entity_indicator	Entity Status	byte	%8.0f
aia_first_to_file	AIA (First Inventor to File)	byte	%8.0f

Purpose

The primary purpose of the data is to enable a detailed analysis of the **dynamics involved in patent processing** and to assess how **organizational** and **social factors** impact examination outcomes. Our analysis goals are centered around enhancing the understanding of **disparities in these outcomes**, with a focus on improving both the efficiency and fairness of operations at the USPTO. Our approach aims to ensure **a more equitable system** that also operates **more effectively**, aligning with broader organizational objectives.

For efficient data management in our project, we used several **R packages** each tailored to specific needs: **lubridate** for handling dates effectively, **stringr** for operations on text data, and **skimr** for quick data overviews. Demographic attributes can be estimated using the **gender** package to infer gender from names and the **wru** package for predicting race or ethnicity.

McGill | DESAUTELS

UNITED STATES PATENT AND TRADEMARK OFFICE uspto

Analytical Approach

Our analytical approach evaluates patent examiner data to dissect the impact of social connections and demographic factors on patent processing times. The process includes data curation, demographic analysis, network graph construction, and advanced regression analysis. We aim to uncover insights into the efficiency and equity of the USPTO's operations, thus refining the patent examination system.

1

Data Loading & Preparation
Import patent application data from Parquet and CSV files; Load R libraries: **arrow**, **dplyr**, **ggplot2**, etc.

2

3

Data Review and Preprocessing
Use head() to preview data structure. Estimate gender and race using gender and wru packages.

Data Integration & Cleaning
Merge gender and race data into main dataset.
Clean dataset by removing unnecessary data.

4

Statistical Analysis and Visualization
Calculate tenure using lubridate.
Analyze demographic distribution and create visualizations (bar, density plots).

5

Network Analysis
Visualize examiner interaction networks with **tidygraph** and **ggraph**; Calculate centrality measures (degree, betweenness).

6

Advanced Modeling
Apply linear regression to model relationships; Evaluate model effectiveness using statistical measures.

7

Actionable Recommendations
Use findings to enhance USPTO workflow and decision-making.

Part 1: Influence of Social & Organizational Factors on Patent Processing Times

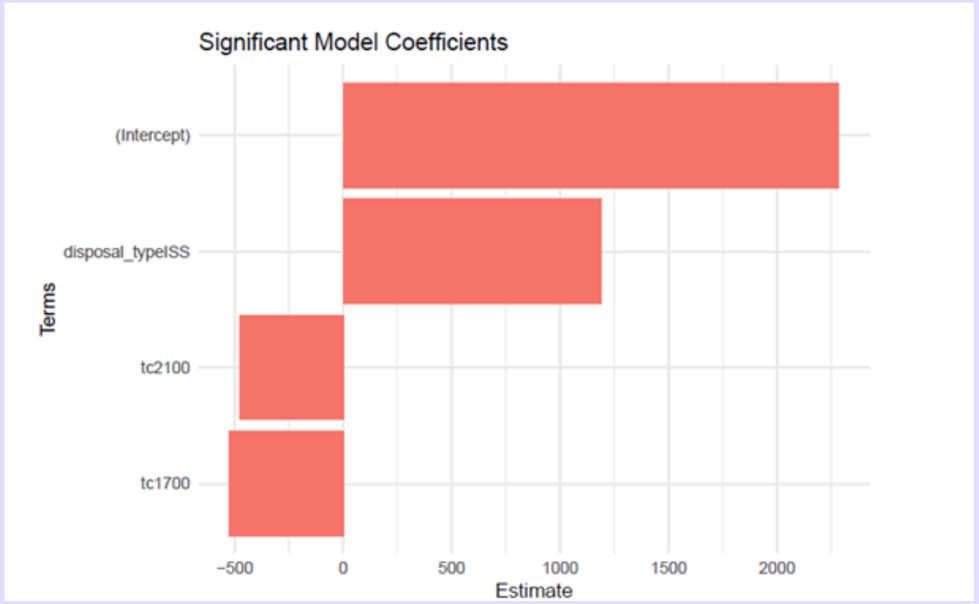
Figure 2. Gender Distribution across Workgroups

- **Workgroup 216** demonstrates a **gender imbalance** with more male than female examiners, though less pronounced than in **workgroup 247**.
- **Workgroup 247** shows a stark gender disparity, with a significantly higher count of male examiners.
- These gender distribution patterns suggest underlying organizational dynamics that **may influence patent processing times.**
- The data prompts further investigation into the USPTO's hiring practices, workgroup assignments, and overall culture towards gender diversity.

*Our findings underscore the intricate relationship between **social structures within the USPTO** & **patent processing efficiency**. The distinct **gender distributions across workgroups** raise critical questions about the impact of organizational culture on operational outcomes. Addressing these disparities is not just a matter of fairness but also **operational effectiveness**. By fostering a **balanced** and **inclusive** workplace, the USPTO can leverage diverse perspectives, enhance decision-making, and potentially streamline patent examination processes.*

Part 1: Influence of Social & Organizational Factors on Patent Processing Times

Figure 2. Advanced Regression Analysis Outcomes



- Interaction between **degree centrality** and **examiner demographics** has minimal effect on patent processing times, with an **Adjusted R-squared of only 0.00054**.
- **Disposal type** and **technology center** classifications significantly improve the model's explanatory power when considering demographic interactions.
- **Closeness centrality** emerges as a vital factor, with the model displaying the highest **Adjusted R-squared value at 0.163,** indicating its significance in processing time

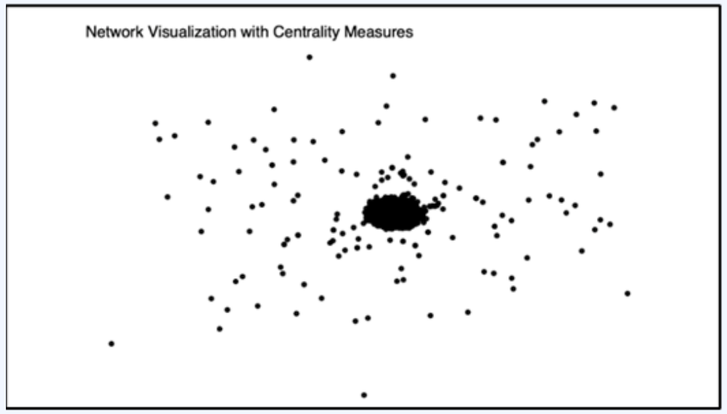
The advanced regression analysis reveals that demographics and network centrality have limited influence on USPTO patent processing times. Efficiency is more strongly dictated by case-specific procedures rather than by the social or organizational network. Thus, improving workflow and examiner network access is key to a more effective and fair patent examination process. **This finding sets the stage for Part 2, where we'll delve deeper into the structural aspects of the USPTO's network to further explore their influence on patent examination efficiency.**

McGill | DESAUTELS

UNITED STATES PATENT AND TRADEMARK OFFICE uspto

Part 2: Analysis of the Role of Network Structure in Patent Processing Efficiency

Figure 3. Centrality in Examiner Networks: Visualizing Influence and Connectivity



Network structure and centrality analysis reveal the impact on USPTO patent processing efficiency, highlighting the potential benefits of tenure-based training and mentorship. Integrating these findings with case-specific information deepens our understanding of what drives efficiency.

- Depicts **USPTO examiner network structure**, highlighting key individuals and the distribution of connectivity.
- Degree centrality is suggested by the density of connections around certain nodes, indicating those examiners with active roles in knowledge sharing and potentially in mentoring.
- The visualization implies a network with central individuals who may facilitate the transfer of information, a characteristic often associated with betweenness centrality, although this specific metric is not directly depicted in the figure.

- The plot compares tenure length distribution between **workgroups 216** and **247**.
- There is a peak in tenure density for **workgroup 216** around the 2,000-day mark.
- Both workgroups exhibit significant peaks at around 6,000 days, suggesting the prevalence of long-tenured examiners with extensive experience within these groups.
- The noticeable peak at 6,000 days for both workgroups signifies that these groups are likely composed of individuals who may serve as knowledge hubs and are integral to the organizational memory of the USPTO.

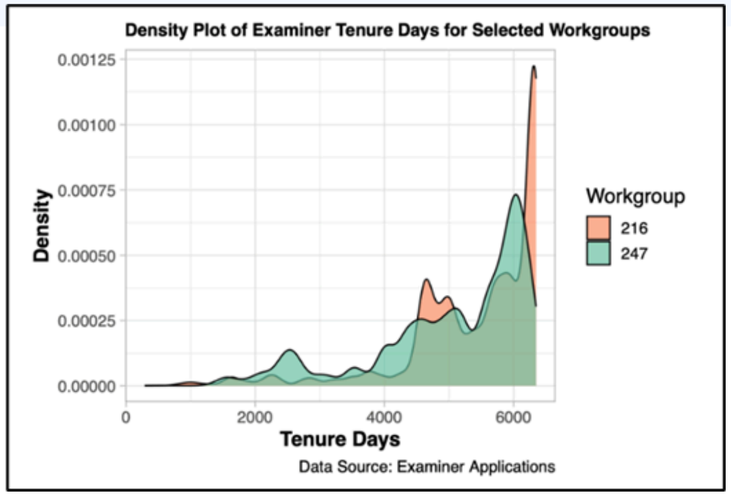
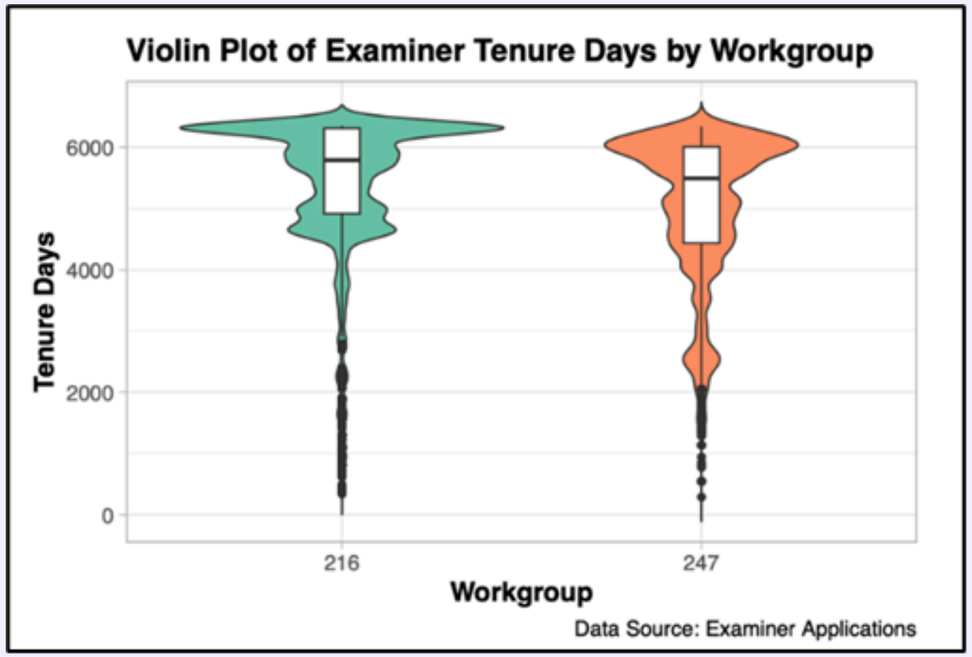


Figure 4. Examiner Tenure Days Density Plot by Workgroup

Part 2: Analysis of the Role of Network Structure in Patent Processing Efficiency

Figure 5. Violin Plot of Examiner Tenure Days by Workgroup



- This plot displays the distribution of examiner tenure days within two key USPTO workgroups, **216** and **247**.
- **Workgroup 216 and 247:** Both display a prominent peak in tenure density at around 6,000 days.
- **Implication:** This denotes a significant contingent of long-tenured examiners within the USPTO, suggesting a well-established expertise within these groups.

Note: *The shape of the plot for each workgroup reflects the density and spread of tenure days, encapsulating the diverse experience within the agency.*

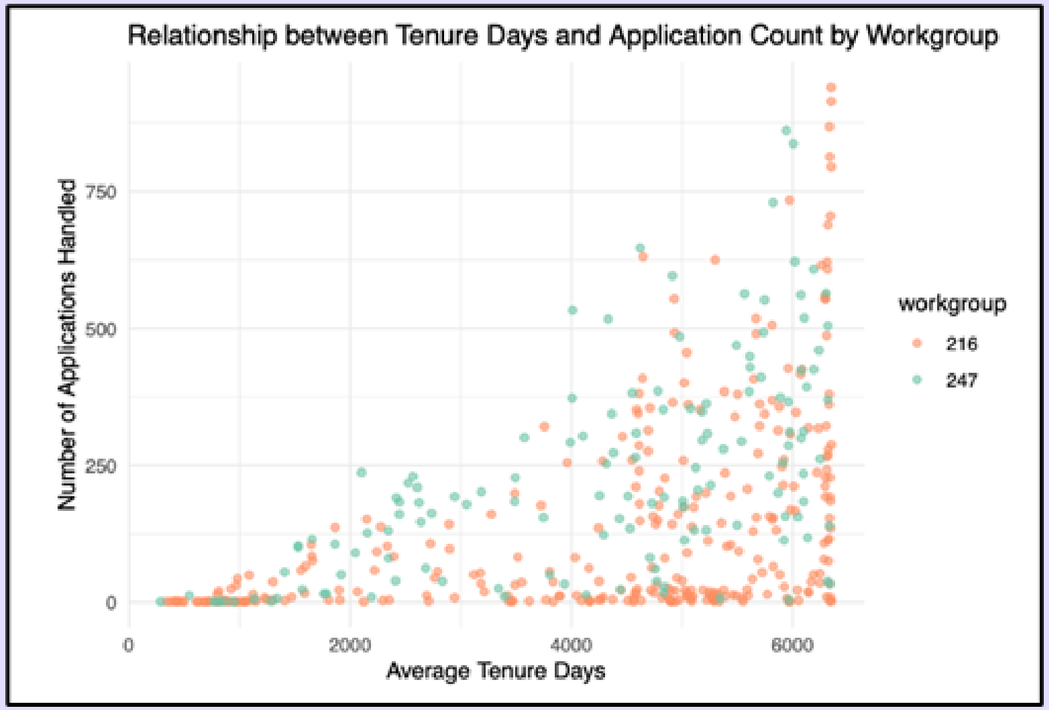
- The violin plot's distinct peaks reveal a notable concentration of long-tenured examiners in both Workgroups 216 and 247, indicating a substantial accumulation of expertise within the USPTO. This suggests a strong foundation of seasoned professionals capable of influencing the organizational network and patent examination process.
- For Workgroup 216, the extensive spread across the tenure spectrum reveals a diverse collective of examiners, ranging from the newly initiated to the deeply experienced. This variety underpins a network that is both rich in emerging insights and grounded in established knowledge. Meanwhile, the pronounced peak for Workgroup 247 underscores a wealth of long-standing expertise, which could represent a deep-seated proficiency within this group.

These tenure patterns are reflective of the network's structure, where longevity may enhance an examiner's centrality and inform their ability to navigate the network effectively, potentially boosting their contribution to the patent examination process.

These insights suggest that mentorship and training initiatives could leverage the diverse tenure landscape to optimize patent processing efficiency, particularly by empowering mid-tenured examiners in workgroup 216 and utilizing the vast experience present in workgroup 247.

Part 2: Analysis of the Role of Network Structure in Patent Processing Efficiency

Figure 6. Scatter Plot of Application Count vs. Examiner Tenure Days by Workgroup

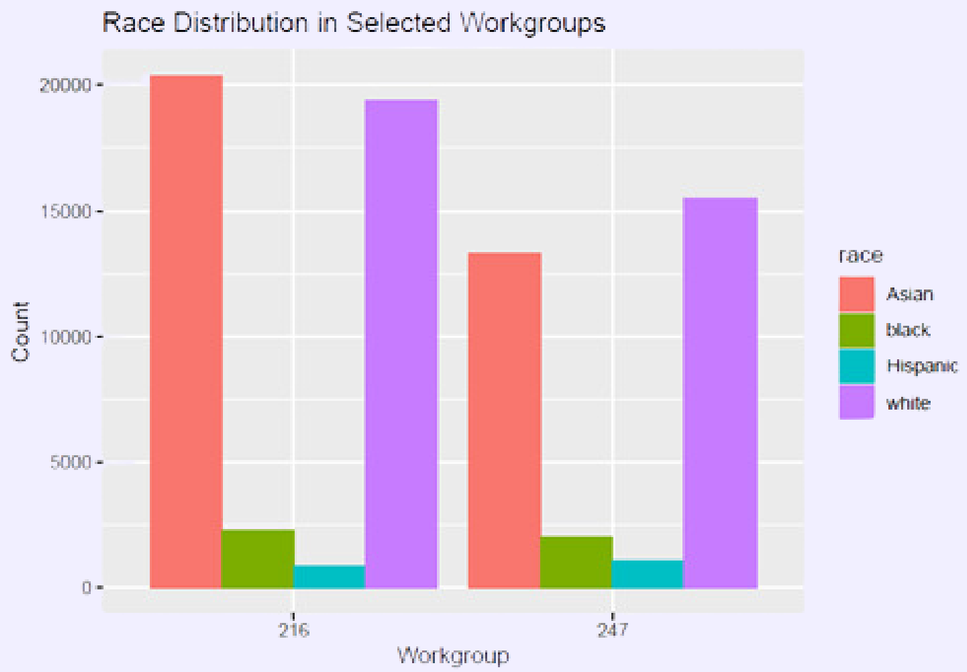


- **Variability in Examiner Experience:** The scatter plot for workgroup 216 shows a diverse range of tenure days, suggesting varied examiner experience levels, while **workgroup 247** indicates a concentration of longer-tenured examiners.
- **Examiner Workload Distribution: Workgroup 247** appears to have examiners with greater experience handling more applications, in contrast to the broader workload spread among workgroup 216's examiners.
- **Complex Dynamics of Tenure and Workloa**d: **A non-linear pattern emerges in the data**, highlighting a complex interplay between an examiner's tenure and the number of applications they handle, transcending simple linear relationships.
- **Network Centrality and Influence:** A modest **positive correlation (0.137)** between tenure and degree centrality indicates that, although tenure is associated with network centrality, both seasoned and new examiners can be influential within the USPTO's network.

The scatter plot, in conjunction with the network analysis, emphasizes the importance of leveraging examiner networks and addressing structural factors to optimize patent processing efficiency across the USPTO.

Part 3: Influence of Race & Ethnicity on Patent Examination Processes

Figure 7. Evaluating Racial Diversity across Workgroups



Both workgroups are predominantly white, with Asians as the second most common race. There is an underrepresentation of black and Hispanic individuals in both groups, which may point towards a lack of racial diversity.

- The bar chart compares racial diversity within workgroups 216 and 247, highlighting a significant underrepresentation of Black and Hispanic individuals in both.
- Workgroup 216 demonstrates a more equitable racial distribution, suggesting a wider range of perspectives which may contribute to a more comprehensive examination process.
- The stark contrast in racial diversity in workgroup 247 could imply a narrowed perspective in the patent examination process, potentially impacting decision-making and innovation.

The findings from our analysis underscore the intricate interplay between structural dynamics, procedural nuances, and examiner demographics in shaping patent processing efficiency at the USPTO.

- Structural elements & procedural practices outweigh examiner demographics in determining patent processing efficiency at the USPTO.
- Our regression analysis underscores the significance of enhancing closeness centrality and considering case specifics for process improvement.
- USPTO emphasizes internal communication enhancements and technology center-specific strategies to foster a fair and inclusive work environment, prioritizing diversity for innovation sustainability.

Business Implications & Strategic Recommendations

Network Efficiency

Strengthen Roles: Boost centrality in the examiner network.
Promote Mentorship: Accelerate knowledge transfer & innovation dissemination.

Collaboration & Integration

Enhance decision-making and processing times through collective cross-cluster expertise.
Establish roles to connect disparate groups for a cohesive integrated decision & action-making process.

Strategic Impact

USPTO Operations: Optimized interactions and resource use for faster, consistent outcomes.
Economic & Social: Speedier patent processing boosts economic growth and ensures equitable innovation access.

Tailored Approaches

Focus on technology centers with specialized facilitation for specific efficiency gains.
Diverse Expertise: Utilize examiner diversity to enrich the process across different cases.

Final Summary & Key Takeaways

Multifaceted Strategic Framework

Our framework is designed to acknowledge and structural and social fabric of the USPTO.

By closely aligning with the organization's diverse demographic profile and the specific requirements of various technological sectors, we aim to create a harmonious, efficient environment that fosters innovation and supports the diverse needs of its stakeholders. This tailored approach ensures that operational intricacies are finely tuned to enhance the overall patent examination process, promoting a culture of inclusivity and precision.

Leveraging Examiner Tenure

Use visual analysis of examiner tenure to bolster examination processes and ensure a steady innovation stream.

Knowledge Transfer Initiatives

Promote effective skill and knowledge dissemination across the organization to reinforce innovation and operational excellence.

Mentorship & Resource Adjustment

Implement mentorship programs that connect experienced examiners with newcomers and adjust resources based on tenure for targeted support.

Regular Network Analyses

Conduct analyses to pinpoint gaps and dependencies, maintaining the agency’s resilience and connectivity.

Final Reflections: *This project offered a comprehensive insight into USPTO's workings, emphasizing the value of examiner experience, knowledge sharing, resource allocation, and network understanding in driving efficiency and fairness. It not only guided us to innovative approaches but also deepened our appreciation for the crucial roles of mentorship and collaboration. Our analytical skills have grown alongside our strategic outlook, highlighting the critical elements that nurture innovation within the organization."*

Questions?

Thank you!

Appendix

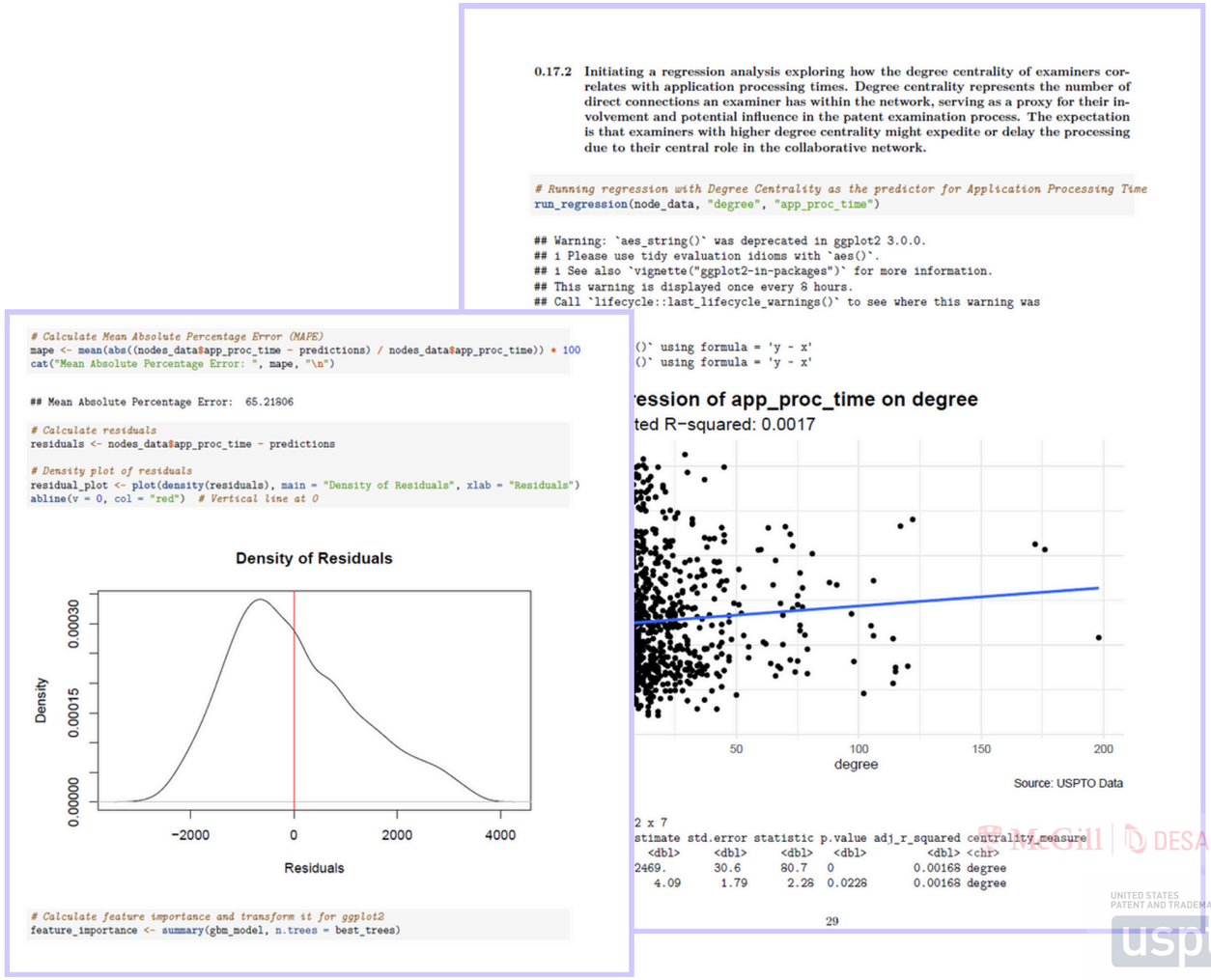
Appendix

Appendix

Appendix 1. Code File: An Overview

Our code file serves as the backbone of our analysis, structured to navigate through various aspects of the USPTO's patent examination process. Organized into distinct sections mirroring the project's key objectives, the code file delves into empirical analyses, advanced regression techniques, and network structure explorations. While not all figures are directly referenced, the code file encompasses a comprehensive exploration of the data, incorporating diverse methodologies to uncover insights.

Beyond the figures presented, the code file undertakes additional explorations, including data cleaning, transformation, and visualization techniques to ensure a thorough examination of the dataset. Through a systematic approach, the code file facilitates a deeper understanding of the USPTO's operational dynamics and underscores the multifaceted nature of patent processing efficiency.



0.17.2 Initiating a regression analysis exploring how the degree centrality of examiners correlates with application processing times. Degree centrality represents the number of direct connections an examiner has within the network, serving as a proxy for their involvement and potential influence in the patent examination process. The expectation is that examiners with higher degree centrality might expedite or delay the processing due to their central role in the collaborative network.

```
# Running regression with Degree Centrality as the predictor for Application Processing Time
run_regression(node_data, "degree", "app_proc_time")
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

```
()` using formula = 'y ~ x'
()` using formula = 'y ~ x'
```

ression of app_proc_time on degree
ted R-squared: 0.0017


```
# Calculate Mean Absolute Percentage Error (MAPE)
mape <- mean(abs((nodes_data$app_proc_time - predictions) / nodes_data$app_proc_time)) * 100
cat("Mean Absolute Percentage Error: ", mape, "\n")
```

```
## Mean Absolute Percentage Error:  65.21806
```

```
# Calculate residuals
residuals <- nodes_data$app_proc_time - predictions

# Density plot of residuals
residual_plot <- plot(density(residuals), main = "Density of Residuals", xlab = "Residuals")
abline(v = 0, col = "red")  # Vertical line at 0
```



```
# Calculate feature importance and transform it for ggplot2
feature_importance <- summary(gbm_model, n.trees = best_trees)
```

```
2 x 7
stimate std.error statistic p.value adj_r_squared centrality_measure
  <dbl>     <dbl>     <dbl>   <dbl>         <dbl> <chr>
2469.       30.6      80.7   0             0.00168 degree
   4.09      1.79      2.28  0.0228        0.00168 degree
```

29