ORGN –672
Organization Network:
Analyzing the Impact of Organizational, Social, and Network Factors on Patent Prosecution Length at the USPTO

Group 4: Mahrukh Shamas, Aasna Shah, Kazuya Hayashi, Tony Xu, Yichen Yu

Agenda

Transform key findings
to actionable results

01
Data Transformation

02
EDA Insights

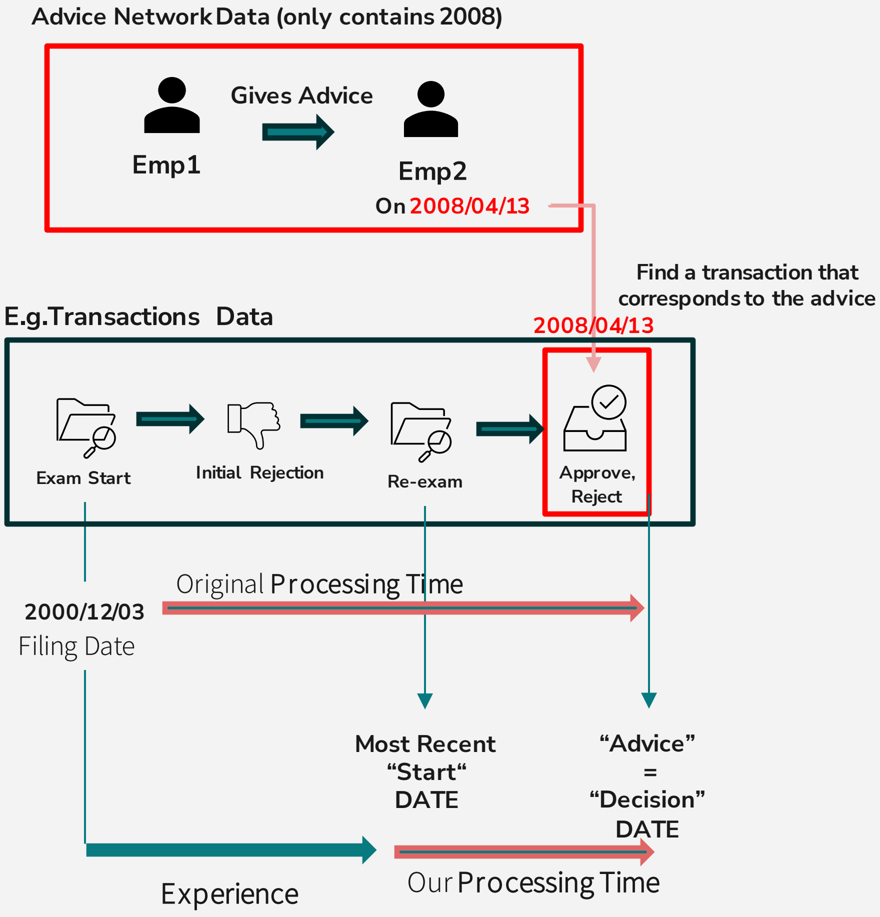
03
Model Development

04
Key Findings

05
Recommendations

- Feature selection
- Model Evaluation

DATA TRANSFORMATION



Steps

- Only restricted to worker presented in the advice_network (edge_sample.csv) data
- Only restricted to applications ENDED in the advice date recorded in the advice_network data
- Filter the data to event code = "'MN/=', 'MCTNF', 'MCTFR" (Only approved and rejected applications are kept). Marked this date as application end date.
- For each application, find the 'DOCK' event and record its date. Marked it as the application start date.

Reasoning

- Each application has dozens of transactions. **Some applications go though many more steps** and time than others.

Additional Steps

- Calculate duration between the examiner's start date and the application record date (most recent "DOCK" date) to create the '**experience**' feature.
- Replace the tenure feature since it stays constant for an examiner, despite the record date of the application.

EDA – Part I

Figure 1: Network **Betweenness** by Gender
- Grey dots: unidentified gender
- Larger Node size = **high betweenness** (>75 pct)

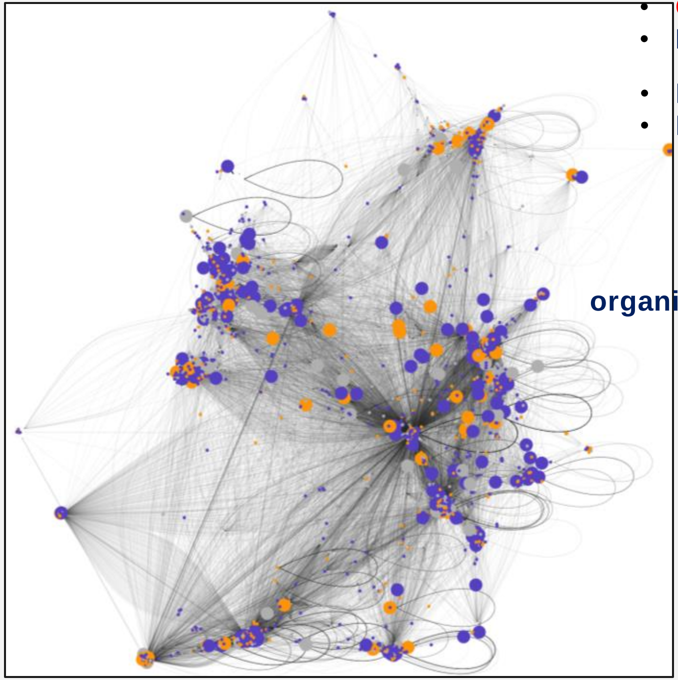
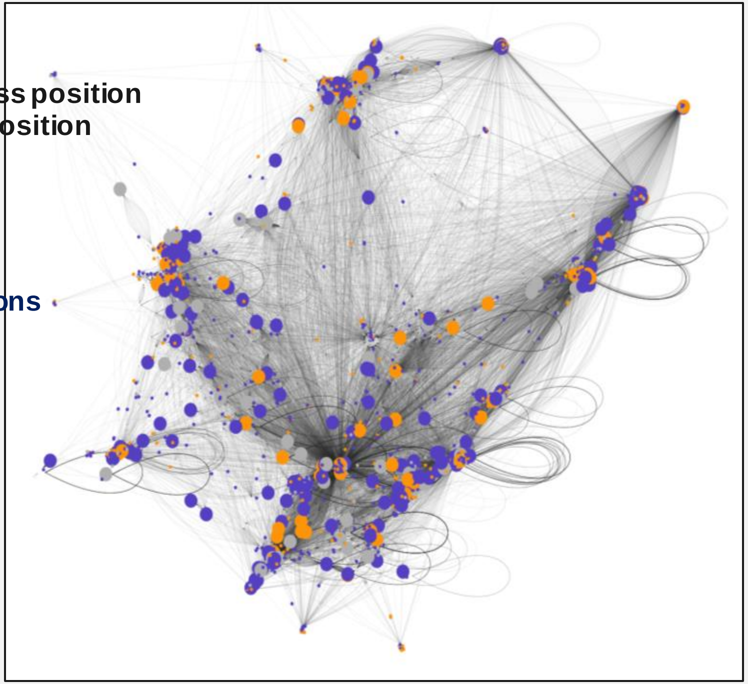


Figure 2: Network **In-Degree** Centrality by Gender
- Grey dots: unidentified gender
- Larger Node size = high in-degree centrality (>75pct)



- **Orange dots**: women
- **Purple dots**: men

- **More Men in high betweenness position**
- **More Men in high in-degree position**

More Men are in organizationally "important" positions

EDA – Part II

Figure 3: Network Betweenness by Race

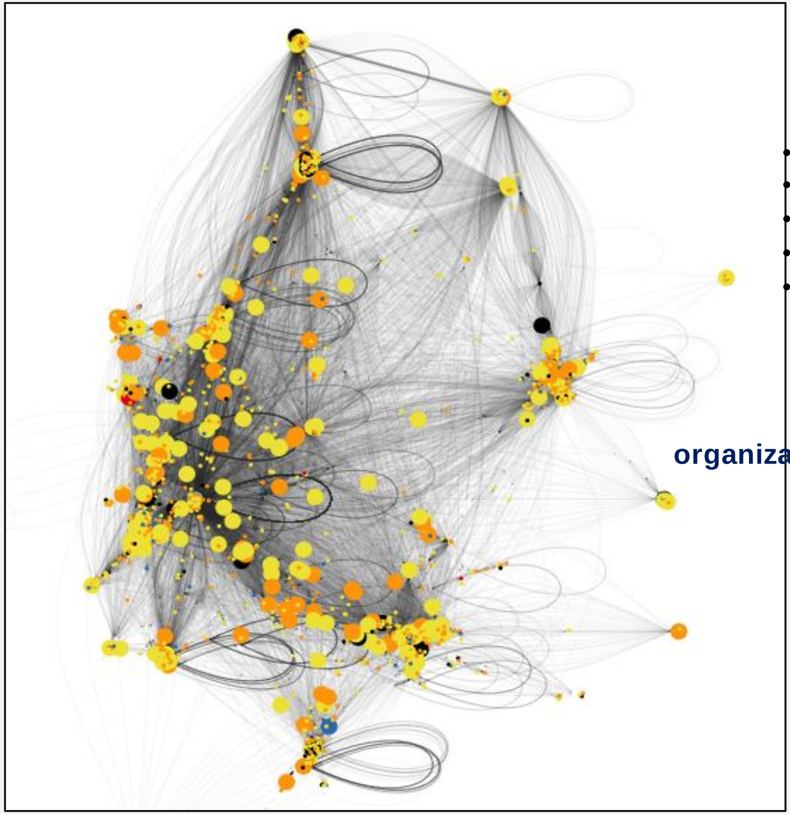
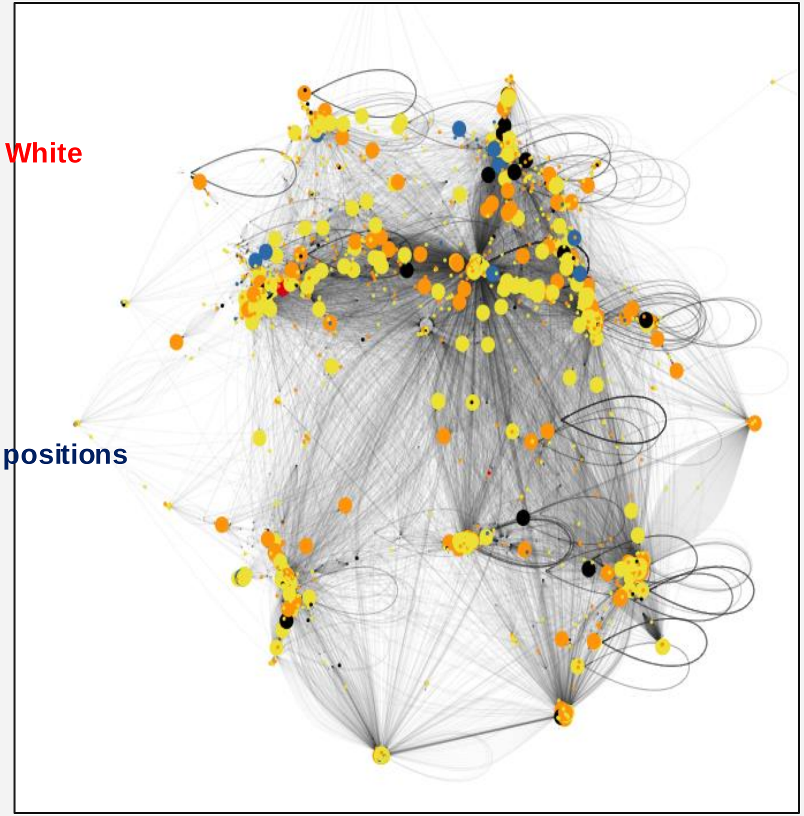


Figure 4: Network In-Degree Centrality by Race



- Light yellow dots: White
- Black dots: Black
- Orange dots: Asian
- Blue dots: Hispanic
- Red dots: Other

⬇

Race = White are in organizationally "important" positions

⬇

Men & White

EDA – Part III

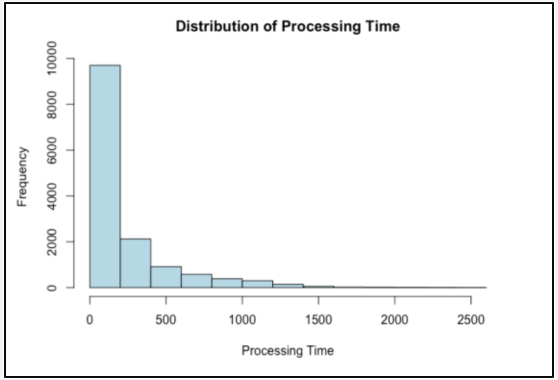


Figure 6: Distribution of Processing Time

Proc Time by Gender & Race

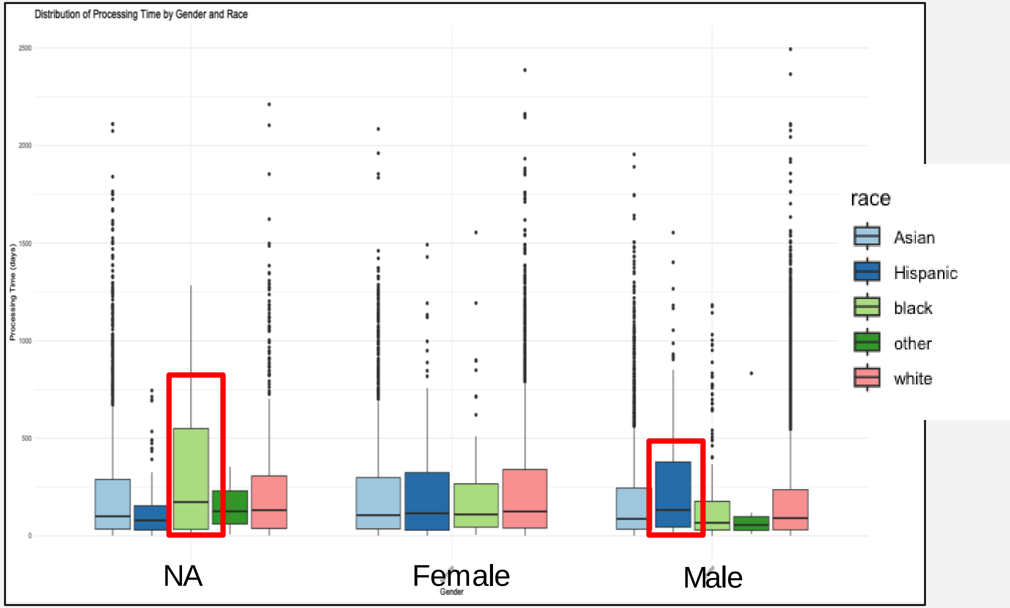
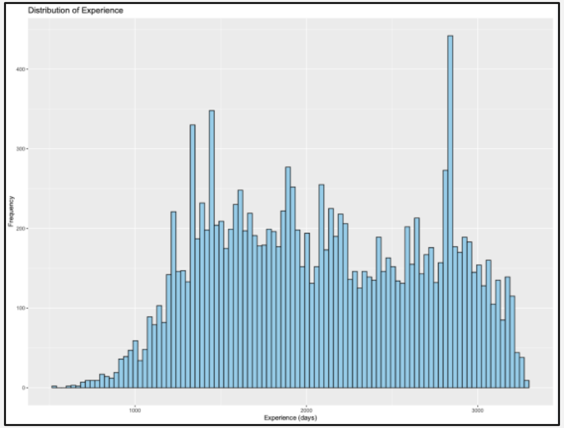


Figure 7: Distribution of Processing Time by Gender and Race

EDA – Part IV

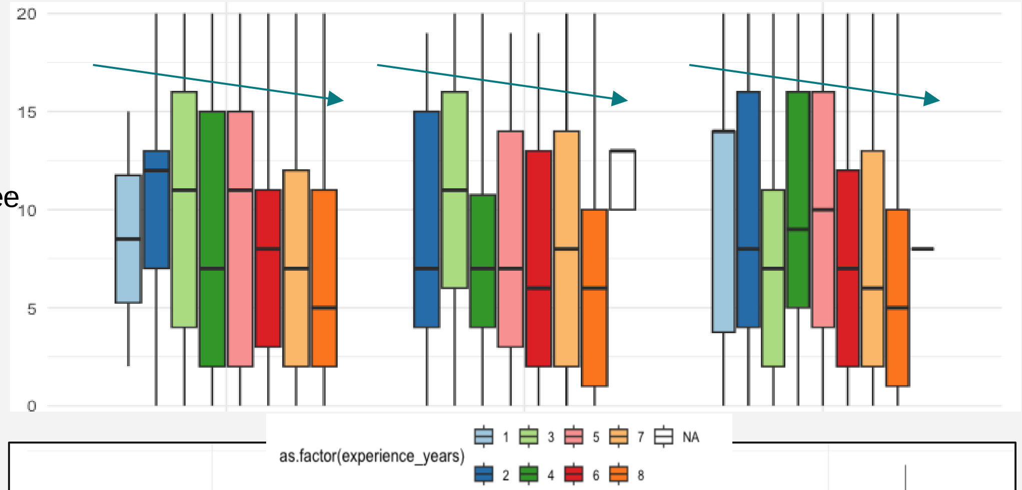
Figure 8: Distribution of Experience in (days)



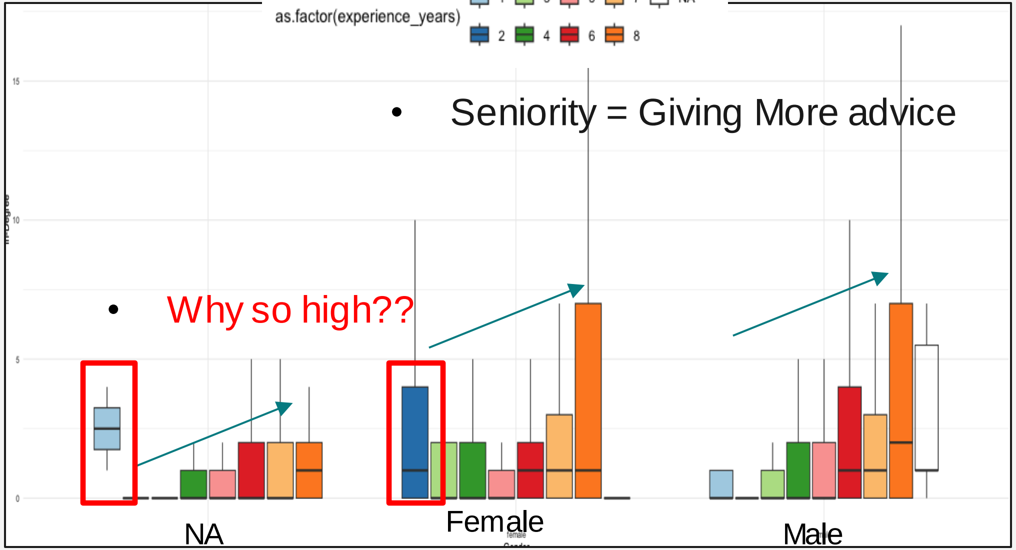
Experience by Gender & Race

Similar Distribution for all gender & race

Out-Degree



In-Degree



- Seniority = Giving More advice
- Why so high??

EDA Insights – Summary

01

Network Central Roles: Men and individuals identified as white or Asian predominantly occupy central positions within the network, indicating they are key in the distribution of advice.

02

Advice Dissemination: Men & White and Asian groups are particularly active in giving advice, which aligns with their larger node sizes indicating greater in-degree centrality.

03

Processing Time and Experience: Processing times and experience levels show notable variability across gender and race, with some demographic groups experiencing faster processing times and others displaying a broad range of experience within the organization.

04

Gender and Experience Dynamics: Young females and NA exhibit unusually high in-degree centrality, indicating they might be substantial recipients of potentially from peers of similar demographic profiles.

05

Seniority and Network Engagement: While in-degree centrality tends to rise with seniority, suggesting more experienced individuals are often advice recipients, out-degree centrality surprisingly declines, demonstrating a reduced tendency for these experienced individuals to offer advice.

Linear Regression V1

Use the features identified (appendix) to conduct the linear regression.

Linear Regression V2

Standardize the continuous variables to see practical significance (effect size)

```
lm(formula = processing_time ~ gender * in_degree + gender *
    out_degree + gender * betweenness + race * in_degree + race *
    out_degree + race * betweenness + experience + disposal_type +
    advice_count_final + start_year + workgroup, data = train_data)
```

- Variables of Interest = gender, network features, advice_count
- Control Vars = application start_year, workgroup, disposal_type

Base Effects				
Term	Estimate	Std. Error	t value	Pr(>|t|)
Experience	-0.8	0.005	-177	***
Advice_Count	-1.83	0.602	-3.01	**
male	-20.1	4.4	-4.6	***
Hispanic	74.1	14.8	5	***
Black	-15	12.5	-1.2	
Other	100.6	46.8	2.2	*
White	-12.5	4.2	-3	**
in_degree	-0.48	0.21	-2.23	*
out_degree	-0.03	0.08	-0.37	
between	0.0016	0.0004	4.18	**
Gender Effects and Interaction Terms				
male:out_degree	0.24	0.07	3.52	***
Race Effects and Interaction Terms				
in_degree:Black	-3.2	1	-3.2	**
in_degree:White	1.15	0.16	7.06	***
between: Hispanic	-0.007	0.0016	-4.52	***
between:Black	0.007	0.0028	2.81	**
between:White	-0.0024	0.0003	-7.73	***

Full summary in Appendix

Linear Regression V1

Most important influencer identified (Ref=Female, Asian)

- Experience (+1 day = -0.8days)
- Total Advice received for the application (+1 = -1.83 days)
- Gender (Men = -20days)
- Race (Hispanic = +74 days)
- In-Degree of examiner +1
  - -0.48 days for Asian, Hispanic, Other
  - -3.68 for Black
  - +0.67 for White
- Out-degree insignificant but heterogeneity between gender
- Betweenness of examiner +1 =
  - + 0.001 for Asian, Other
  - + 0.008 for Black
  - - 0.006 for Hispanic
  - -0.0008 for White

Linear Regression V2

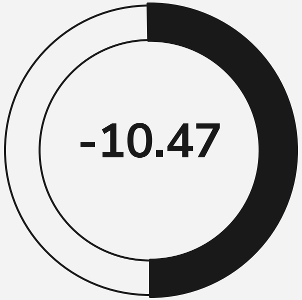
Main Difference compared to V1: Standarizing continuous variables to see how 1 standard deviation change in X affects the application time. **We Focus especially on the effects of network features.**

One std increase in those variables =

Previously 0.001 days !

Experience

Highly significant – Suggest more experience correlates with lower approval time.



In_Degree

Significant – Suggest more central individuals manage to process faster.

Betweeness

Significant – Suggests individuals who act as bridges may face processsing delay.

Research Questions Explored in Our Analysis

1. What are the organizational and social factors associated with variation in patent application processing time?

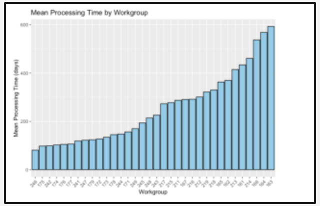
2. How does network structure (i.e., examiner collaboration, communication) influence patent examination outcomes?

3. Does examiner gender and/or race/ethnicity correlate with disparities in processing time, career trajectory, or attrition?

Question 1: Understanding Variations in Patent Application Processing Times

Key Concepts:

1 Organizational Impact



- Specific workgroup dynamics, such as leadership styles and policy enforcement, significantly dictate processing times.
- Visual data analysis from EDA shows a clear performance gradient among different workgroups, indicating operational influences.

2 Social and Individual Factors

- Examiner experience emerged as a key factor; more seasoned examiners process applications quicker.
- Gender and racial demographics are closely tied to processing times, with disparities suggesting underlying social dynamics.

3 Network Centrality

- A strong network within the USPTO correlates with shorter processing times, demonstrating the importance of internal communication.
- However, these benefits are not uniformly distributed; gender and race play a role in the extent of advantage gained from network centrality.

4 Interactions and Inequities

- Intersection of advice-giving network behavior with examiner demographics reveals efficiency variances, highlighting possible biases.
- White examiners and males show disproportionate benefits from network centrality, indicating potential systemic inequities.

5 Implications

- To enhance efficiency and promote fairness in processing times, it is essential to consider the interplay between organizational frameworks and the individual characteristics of examiners, including their roles in the network's communication structure.
- Implementing changes should aim at not only improving efficiency but also at ensuring equitable processes across all examiner groups.

Question 2: The Impact of Network Structure on Patent Processing

Key Concepts

- **Betweenness Centrality:** Examiners with high betweenness centrality often serve as vital links between unconnected parts of the network. However, this central position increases their workload, leading to longer processing times for their own patent applications.
- **In-degree Centrality:** The number of direct advice relationships an examiner has within the network correlates inversely with processing times. Those who are sought after for advice can leverage the collective knowledge and expertise of their network, allowing them to process applications more rapidly.
- **Advice Network Dynamics:** The structure and interactivity of the advice network have a profound impact on patent processing outcomes. High engagement within the network during key decision-making periods enables examiners to process applications with greater speed and accuracy.

Findings on Network Structure

1. **Central Positions Increase Workload:** Examiners with high betweenness centrality face extended processing times.
2. **Receiving Advice Speeds Up Processes:** A negative correlation between in-degree centrality and processing times is observed.
3. **Giving Advice Neutral in Effect:** Out-degree centrality does not significantly alter examiner processing speeds.

Demographic Dynamics

- **Gender:** The relationship between gender and network centrality indicates that social perceptions and roles within the network could affect processing times differently for male and female examiners.
- **Race:** Network benefits such as advice and support are not uniformly experienced across different racial groups, suggesting systemic disparities within the network's social dynamics.

Implications for Process Efficiency

- **Strengthening Network Ties:** Proposals for improving connectivity to expedite processing.
- **Fair Resource Distribution:** Ensuring equitable access to advice and support for all examiners.

Question 3: The Impact of Gender and Race/Ethnicity on Patent Examination Processes

Key Insights

- **Racial and Ethnic Influence on Processing Times:** Minority examiners experience longer processing times than white counterparts, indicating disparities in workload and access to resources.
- **Network Structure and Racial Integration:** Lower in-degree centrality for underrepresented racial groups implies less received advice, contributing to slower processing times.
- **Social Dynamics and Racial Perception:** Interaction of racial factors with network centrality shows uneven benefits across racial lines, pointing towards implicit biases in advice-sharing practices.

Implications for Policy and Practice

Enhancing Network Connectivity:

- Implementation of structured mentorship and peer-support programs specifically designed to strengthen professional networks for minority examiners.
- Development of initiatives that promote regular interactions and knowledge exchange sessions across diverse groups, aiming to integrate underrepresented examiners into the central fabric of the advice network.

Ensuring Equitable Workload Distribution:

- Monitoring and auditing of case assignments to identify and correct any patterns of bias, ensuring that all examiners receive a fair and balanced distribution of work, irrespective of racial or ethnic background.
- Adoption of transparent criteria for case allocation that accounts for diversity and expertise, promoting a balanced distribution of complex and straightforward cases across all examiners.

Gender Disparities

Importance of Addressing Disparities:

- Recognition that systemic racial and ethnic disparities in patent examination not only affect individual examiners but also the overall integrity and effectiveness of the USPTO's processes.
- Acknowledgment that improving fairness and efficiency in patent examination requires proactive measures to address these disparities.

Recommended Interventions:

- Establishment of comprehensive diversity and inclusion training programs aimed at raising awareness of unconscious biases and fostering a culture of inclusivity within the USPTO.
- Creation of formal channels and platforms that facilitate equal access to advice, resources, and organizational support for examiners from diverse racial and ethnic backgrounds.

Recommendations

Enhancing Examiner Training and Support:
Experiences correlates with shorter approval time →Programs focusing on transferring these learnings to new staff members.

Addressing Disparities in Processing Times:
Disparities in processing times based on gender and race → Programs focusing on promoting diversity, equity, and inclusion as well as knowledge sharing.

Leveraging Workgroup Dynamics:
Significant variation in processing times across different workgroups →Examiner rotation program and cross-department discussion seminars.

Fostering a Collaborative Culture:
Network effects influence the processing time a lot → Strengthen collaboration among examiners, promoting inclusive behaviors

Conclusion

Strategic Approach

- Holistic Methodology: Analysis demonstrates the synergy between human and organizational factors.
- Potential Benefits: Enhanced efficiency and reduced disparities, benefiting stakeholders and contributing to global economic growth and technological advancement.

Key Pillars for Implementation

- Leadership Commitment & Examiner Engagement: Essential for sustained implementation of recommendations to improve knowledge sharing and collaboration.
- Continuous Improvement: Viewing the recommendations as a dynamic journey, adapting to the evolving needs of inventors, businesses, and society.

Future Outlook

- Redefining Excellence: Leveraging data, investing in people, and fostering collaboration for setting new standards in intellectual property protection.

By embracing these recommendations and committing to a transformative journey, the USPTO can position itself as a global leader in patent examination, setting a new benchmark for efficiency, equity, and innovation.

Thank you!
Any Questions?

APPENDIX

More EDA Insights

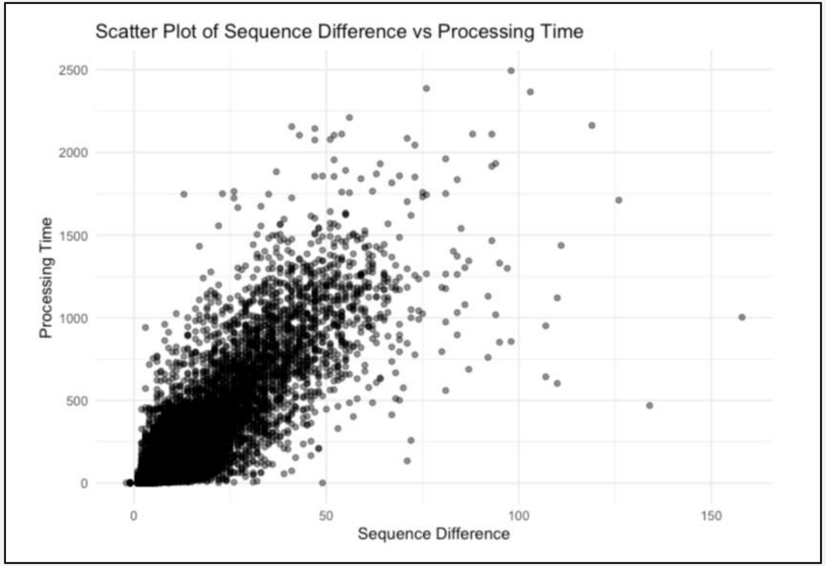


Figure 14: Mean Processing Time by Workshop

The Impact of Sequence Difference on Patent Processing Time

Graph Analysis:

- The scatter plot reveals a dispersed correlation between 'sequence_diff' and processing time.
- An increase in 'sequence_diff' appears to be associated with longer processing times, indicating that more events may lead to delays in the patent approval or rejection process.

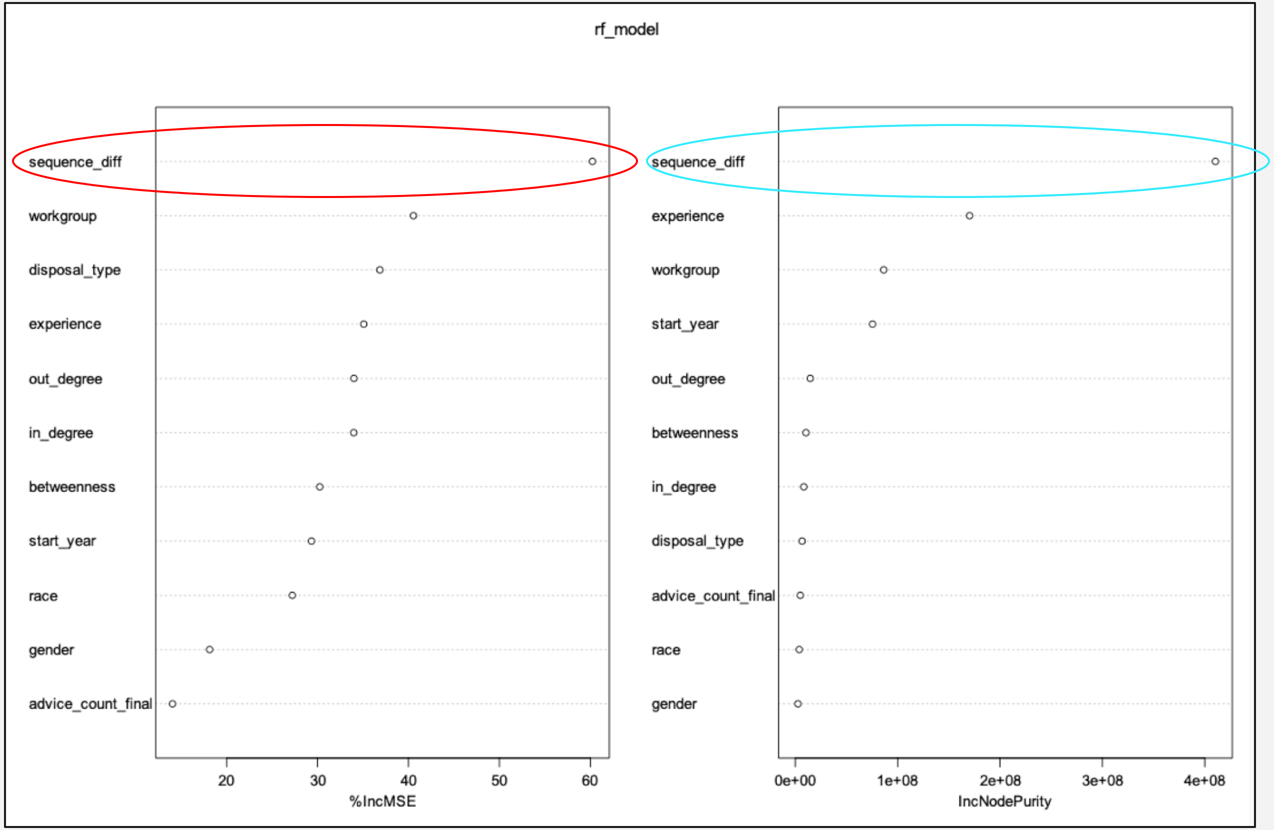
Key Observations:

- A significant concentration of data points towards the lower end of 'sequence_diff' suggests a common minimum number of processes.
- Outliers with high 'sequence_diff' values hint at complex cases that require an extended processing time.

Conclusions:

The 'sequence_diff' is a crucial variable for understanding the heterogeneity in patent processing times. By accounting for the number of events, we can better predict and potentially streamline the patent prosecution process.

Random Forest Feature Selection



Use Random Forest to visualize feature importance and select feature for Linear Regression

Most important features include:

- Sequence Difference.
- Examiner Experience
- Work group in.
- Year started
- Social Network Features

Linear Regression V1

```
Coefficients: (2 not defined because of singularities)
                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)               2.546e+03  1.444e+01  176.371  < 2e-16 ***
gendermale               -2.011e+01  4.375e+00   -4.596 4.37e-06 ***
in_degree                -4.788e-01  2.148e-01   -2.229 0.025861 *
out_degree               -2.769e-02  7.595e-02   -0.365 0.715464
betweenness               1.638e-03  3.913e-04    4.184 2.89e-05 ***
raceblack                -1.501e+01  1.249e+01   -1.201 0.229645
raceHispanic              7.405e+01  1.484e+01    4.991 6.12e-07 ***
raceother                 1.006e+02  4.678e+01    2.150 0.031620 *
racewhite                -1.249e+01  4.184e+00   -2.986 0.002833 **
experience               -7.988e-01  4.506e-03 -177.277  < 2e-16 ***
disposal_typerejected    -1.495e+01  3.645e+00   -4.101 4.16e-05 ***
advice_count_final       -1.885e+00  6.041e-01   -3.121 0.001809 **
start_year2001           -3.380e+02  4.479e+00  -75.452  < 2e-16 ***
start_year2002           -6.294e+02  5.487e+00 -114.708  < 2e-16 ***
start_year2003           -9.638e+02  6.521e+00 -147.792  < 2e-16 ***
start_year2004           -1.226e+03  7.596e+00 -161.426  < 2e-16 ***
start_year2005           -1.464e+03  1.011e+01 -144.744  < 2e-16 ***
start_year2006           -1.692e+03  1.675e+01 -101.045  < 2e-16 ***
gendermale:in_degree     -2.261e-01  2.187e-01   -1.034 0.301160
gendermale:out_degree     2.357e-01  6.707e-02    3.515 0.000442 ***
gendermale:betweenness    8.211e-04  4.250e-04    1.932 0.053365 .
in_degree:raceblack      -3.219e+00  9.950e-01   -3.235 0.001221 **
in_degree:raceHispanic    1.360e+00  8.437e-01    1.612 0.106899
in_degree:raceother      -3.875e-01  4.937e+01   -0.008 0.993738
in_degree:racewhite       1.152e+00  1.631e-01    7.063 1.76e-12 ***
out_degree:raceblack      1.741e+00  2.967e-01    5.869 4.55e-09 ***
out_degree:raceHispanic  -1.365e+00  2.958e-01   -4.615 3.98e-06 ***
out_degree:raceother             NA         NA       NA       NA
out_degree:racewhite      1.811e-02  7.014e-02    0.258 0.796313
betweenness:raceblack     7.790e-03  2.766e-03    2.816 0.004869 **
betweenness:raceHispanic -7.214e-03  1.596e-03   -4.521 6.24e-06 ***
betweenness:raceother            NA         NA       NA       NA
betweenness:racewhite    -2.366e-03  3.059e-04   -7.734 1.16e-14 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 122.3 on 8396 degrees of freedom
Multiple R-squared:  0.844, Adjusted R-squared:  0.8428
F-statistic: 720.8 on 63 and 8396 DF,  p-value: < 2.2e-16
```

Linear Regression V2

```
Coefficients: (2 not defined because of singularities)
                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)               2.546e+03  1.444e+01  176.371  < 2e-16 ***
gendermale               -2.011e+01  4.375e+00   -4.596 4.37e-06 ***
in_degree                -4.788e-01  2.148e-01   -2.229 0.025861 *
out_degree               -2.769e-02  7.595e-02   -0.365 0.715464
betweenness               1.638e-03  3.913e-04    4.184 2.89e-05 ***
raceHispanic              7.405e+01  1.484e+01    4.991 6.12e-07 ***
raceblack                -1.501e+01  1.249e+01   -1.201 0.229645
raceother                 1.006e+02  4.678e+01    2.150 0.031620 *
racewhite                -1.249e+01  4.184e+00   -2.986 0.002833 **
experience               -7.988e-01  4.506e-03 -177.277  < 2e-16 ***
disposal_typerejected    -1.495e+01  3.645e+00   -4.101 4.16e-05 ***
advice_count_final       -1.885e+00  6.041e-01   -3.121 0.001809 **
start_year2001           -3.380e+02  4.479e+00  -75.452  < 2e-16 ***
start_year2002           -6.294e+02  5.487e+00 -114.708  < 2e-16 ***
start_year2003           -9.638e+02  6.521e+00 -147.792  < 2e-16 ***
start_year2004           -1.226e+03  7.596e+00 -161.426  < 2e-16 ***
start_year2005           -1.464e+03  1.011e+01 -144.744  < 2e-16 ***
start_year2006           -1.692e+03  1.675e+01 -101.045  < 2e-16 ***
gendermale:in_degree     -2.261e-01  2.187e-01   -1.034 0.301160
gendermale:out_degree     2.357e-01  6.707e-02    3.515 0.000442 ***
gendermale:betweenness    8.211e-04  4.250e-04    1.932 0.053365 .
in_degree:raceHispanic    1.360e+00  8.437e-01    1.612 0.106899
in_degree:raceblack      -3.219e+00  9.950e-01   -3.235 0.001221 **
in_degree:raceother      -3.875e-01  4.937e+01   -0.008 0.993738
in_degree:racewhite       1.152e+00  1.631e-01    7.063 1.76e-12 ***
out_degree:raceHispanic  -1.365e+00  2.958e-01   -4.615 3.98e-06 ***
out_degree:raceblack      1.741e+00  2.967e-01    5.869 4.55e-09 ***
out_degree:raceother             NA         NA       NA       NA
out_degree:racewhite      1.811e-02  7.014e-02    0.258 0.796313
betweenness:raceHispanic -7.214e-03  1.596e-03   -4.521 6.24e-06 ***
betweenness:raceblack     7.790e-03  2.766e-03    2.816 0.004869 **
betweenness:raceother            NA         NA       NA       NA
betweenness:racewhite    -2.366e-03  3.059e-04   -7.734 1.16e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 122.3 on 8396 degrees of freedom
Multiple R-squared:  0.844,  Adjusted R-squared:  0.8428
F-statistic: 720.8 on 63 and 8396 DF,  p-value: < 2.2e-16
```

Regression Assumptions Check

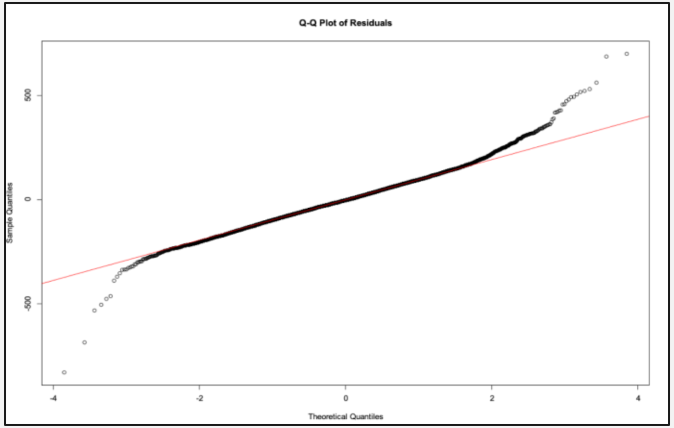


Figure 18: Q-Q Plot of Residuals

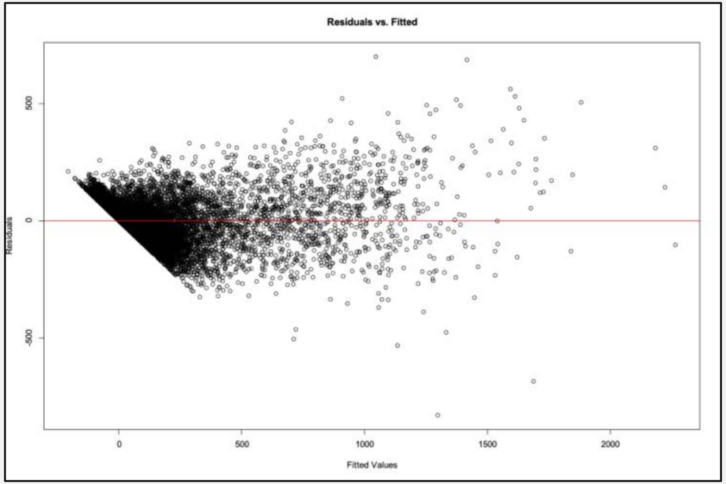


Figure 17: Scatter Plot of Residuals vs. Fitted of Model 2

	GVIF	Df	GVIF^(1/(2*Df))
gender	1.109651	1	1.053400
race	1.350837	4	1.038306
experience	8.992470	1	2.998745
in_degree	1.656975	1	1.287236
out_degree	1.256723	1	1.121037
betweenness	1.715267	1	1.309682
disposal_type	1.349295	1	1.161592
sequence_diff	2.628070	1	1.621132
advice_count_final	1.063972	1	1.031490
start_year	13.505028	6	1.242247
workgroup	3.626073	33	1.019709

Figure 16: Multicollinearity Table