

1. Centrality Extension (Task 1)

Three centrality metrics were employed to evaluate the influence of directors within the network: degree, eigenvector, and betweenness centrality. Degree centrality identifies directors who serve on numerous boards and maintain extensive networks across various companies. Eigenvector centrality emphasizes directors who are not only connected but also linked to other prominent individuals—thereby embedding them within elite networks. Betweenness centrality was newly incorporated as it identifies directors who function as intermediaries between otherwise isolated groups. These individuals possess the structural authority to introduce companies to new networks and are ideally suited for facilitating high-level acquisition opportunities. This directly aligns with the venture capital fund's objective of having the company acquired by a well-connected public firm.

2. Code Repair (Task 2)

Fix – Boolean Mapping Clarification

The initial code that transformed software background flags from text to Boolean values was ambiguous and prone to errors. It has been replaced with a more explicit and comprehensible method that clearly associates 't' and 'f' with True and False. This eliminates confusion and simplifies the transformation of tty scores—potentially enhancing their value in facilitating a tech-driven acquisition. Considering the venture capital fund's aim of securing a sale to a large U.S.-based public companies, particularly in the technology sector, investigating this feature could assist in identifying directors who are both technically proficient and well-positioned to connect the company with appropriate buyers.

Fix – Subgraph Selection Clarity

The previous method for identifying the largest connected component of the graph was both fragile and lacked documentation. The new solution employs a more transparent and robust technique, clearly indicating that the largest component is being deliberately extracted. This clarity is crucial for ensuring that centrality analysis is performed on a properly connected graph.

Fix – Feature Scaling Naming

The designation of the variable used for scaled features was misleading, suggesting it encompassed only demographic data. It has been renamed and restructured to accurately represent that it includes both demographic and centrality features. This enhancement improves clarity and facilitates easier modification or reuse of the code in future analyses.

Fix – Graph Construction Documentation

The logic behind graph construction was not documented, which made it challenging to comprehend the network's structure. A comment has been included to clarify that the graph is bipartite, with nodes symbolizing companies and directors, and edges denoting board memberships. This documentation aids readers in understanding the foundational structure that underpins the network analysis.

Some other codes were removed such as hard lookups of individuals which served no purpose to the broader analysis.

3. Exploring an Unused Feature (Task 3)

The `software_background` column has been recognized as an intriguing and unused feature. It signifies whether a director possesses a background in software or technology. This variable could be utilized to investigate whether tech-savvy directors hold a more central position in the network or belong to specific clusters, which could be strategically advantageous for targeting companies in the tech sector or identifying directors who are well-suited for modern, innovation-driven acquisitions.

4. Complimentary Dataset (Task 4)

I selected the Fortune 100 Board Member Information Dataset from Kaggle due to its concentration on directors from the most influential and resource-abundant companies in the United States, which aligns closely with the venture capital firm's objective of identifying individuals capable of impacting or facilitating a high-value acquisition.

This dataset enhances the current network and compensation information by presenting a carefully selected group of directors who are already situated at the pinnacle of the corporate structure. These individuals are more likely to possess access to executive decision-makers, have established connections with investment banks, and wield influence within the boardrooms of publicly traded companies—rendering them ideal strategic contacts.

Furthermore, the dataset is of a manageable size, well-organized, and compatible with existing director names, facilitating seamless merging and analysis. Here is the link to it:

<https://www.kaggle.com/datasets/thedevastator/fortune-100-board-member-information?resource=download&select=boardmembers.csv>

5. Choosing two refinement options (b & c) (Task 5)

For option (b), In this project, we integrated the Fortune 100 Board Member Information dataset with our existing director network data to identify high-impact individuals who could facilitate a strategic acquisition, as envisioned by the venture capital firm. We commenced by cleaning and standardizing the names in both datasets to allow for precise matching, subsequently filtering out the directors who were present in both sources. These overlapping individuals were then enhanced with essential network metrics—specifically eigenvector, degree, and betweenness centrality—along with compensation data, which acted as a proxy for their corporate influence and seniority. We visualized this data through a scatter plot that illustrated structural network power (eigenvector and betweenness) in relation to compensation, aiding in the identification of directors possessing both social capital and positional authority. From this analysis, we compiled a shortlist of the top five most strategically valuable directors. These individuals are

distinguished as well-positioned to broker or influence acquisition opportunities due to their connections within high-ranking corporate ecosystems. Moving forward, the logical next step is to delve deeper into these directors' backgrounds—focusing on their deal-making history and executive affiliations—to uncover actionable ways the VC firm can utilize their networks in pursuit of a sale. This process guarantees a focused, evidence-based strategy that maximizes the probability of a successful exit.

Answer to option c:

This project brings forth several significant ethical issues, particularly concerning privacy, fairness, and intent. While the data utilized—such as DEF 14A filings and information about Fortune 100 board members—is publicly accessible, repurposing it to assess and rank individuals based on their influence within corporate networks treads a delicate line between strategic analysis and personal profiling. One major concern is that directors are being judged not solely on their professional qualifications, but also on their potential utility in achieving a specific financial objective (i.e., facilitating an acquisition), which could reduce individuals to their network value without their awareness or consent.

Furthermore, there exists an ethical risk of bias in the measurement of influence. Metrics such as compensation and network centrality may unintentionally favor individuals from already privileged backgrounds, thereby reinforcing existing power dynamics and neglecting less connected yet equally competent candidates. Should such models be scaled or institutionalized, they could worsen inequities in corporate leadership.

Additionally, the fundamental aim of leveraging board member influence to orchestrate an acquisition raises concerns regarding transparency. If such strategies are employed to manipulate boardroom dynamics or apply covert pressure, they may erode shareholder trust or circumvent standard governance procedures.

Nevertheless, when applied responsibly and transparently, this form of analysis can assist companies in identifying key stakeholders, nurturing meaningful partnerships, and enhancing decision-making. Ethical practice necessitates recognizing the limitations of the data, refraining from overreaching conclusions, and honoring the autonomy and privacy of the individuals involved.