**Duplicate Therapy Risk Identification in Order Sets Using Network Analysis**

**Jeff Watson**

**Department of Data Science, University of Wisconsin – Eau Claire**

**DS785: Capstone**

**May 5, 2025**

**Abstract**

This project investigated patterns of duplicate medication therapy within standardized inpatient order sets using social network analysis. Initiated by the Patient Safety and Quality team of Advocate Health’s Division of Pharmacy, the study aimed to identify clinically relevant duplication risks not evident through traditional one-by-one order set reviews. Electronic Medical Record (EMR) data from October 2024 were used to construct a medication co-occurrence network derived from over 815,000 active orders across multiple healthcare markets. In the network, nodes represented medications and edges denoted co-occurrence within the same order set. NetworkX in Python was employed to calculate centrality metrics, detect community structures, and visualize network topology. Key findings included the identification of highly central and frequently duplicated medications, particularly among perioperative pain management and diagnostic support regimens. Opioid analgesics, specific antibiotic classes, and insulin analogs formed dense clusters, highlighting areas of therapeutic redundancy risk. During exploratory analysis, pharmaceutical subclass was selected as the primary categorization, improving clinical specificity compared to broader therapeutic classes. The insights generated provide practical value for refining order set design, strengthening clinical decision support tools, and mitigating duplication-related safety risks. This project demonstrated the utility of network analysis in uncovering systemic patterns in medication use and established a foundation for future work involving modularity scoring, community refinement, and expanded duplicate therapy mapping.

**Keywords:** Order sets, duplicate therapy, duplicate therapy detection, network analysis, pharmaceutical subclass, NetworkX, electronic medical records, healthcare analytics, centrality measures, patient safety, clinical decision support, network visualization

**Table of Contents**

**List of Figures 6**

**Chapter 1: Introduction 7**

**Introduction 7**

**Defining Duplicate Therapies 8**

**Background 8**

**Statement of Business Case or Problem 9**

**Project Objectives 10**

**Significance of the Project 11**

**Definition of Terms 12**

**Assumptions, Limitations, and Delimitations 13**

**Conclusion 14**

**Chapter 2: Review of the Literature 16**

**Introduction 16**

**Prevalence of Duplicate Medication and Therapies 17**

**Risks to Patient Care and Systemic Failures 19**

**Network Analysis as a Solution 21**

**Conclusion 23**

**Chapter 3: 26**

**Introduction 26**

**Exploratory Data Analysis 27**

**Medication Co-occurance Network 33**

**Duplicate Therapy Network 34**

**Community Detection 39**

**Community Structure and Visualization 40**

**Subclass-Homogeneous Community Spotlights 41**

**Duplicate Therapy Concentrtion Across Order Sets and Subclasses 43**

**Conclusion 45**

**Chapter 4: 46**

**Introduction 46**

**Results and Interpretation 46**

**Limitations and Stakeholder Considerations 49**

**Systemic Implications and Scalability 49**

**Conclusion 50**

**Chapter 5: 52**

**Introduction 52**

**Summary of Project 52**

**Implications of Project Results 53**

**Limitations and Next Steps 54**

**Conclusion 55**

**References 57**

**Appendix A: Data Dictionary 60**

**Appendix B: Code 64**

**List of Figures**

**Figure 1 *Top 20 Pharmaceutical Subclasses – Medication Co-occurrance Network* 28**

**Figure 2 *Degree Distribution – Medication Co-occurrence Network* 29**

**Figure 3 *Degree Distribution – Duplicate Therapy Network* 30**

**Figure 4 *Top 20 Pharmaceutical Subclasses – Duplicate Therapy Network* 31**

**Figure 5 *Visualization of Largest Connected Component* 32**

**Figure 6 *Top 20 Medications by Degree Centrality* 36**

**Figure 7 *Top 20 Medications by Betweenness* 38**

**Figure 8 *Community Detection of Medication Co-occurance Network* 41**

**Figure 9 *Top 5 Subclass-Homogeneous Communities* 43**

**Figure 10 *Heatmap of Duplicate Therapy Across Top 25 Order Sets* 44**

**Chapter 1: Introduction**

**Introduction**

Healthcare operates under the dual imperative of providing safe, effective patient care while improving efficiency and reducing costs. One widely adopted tool to support these goals is the use of order sets. Order sets are standardized collections of medical orders designed for specific clinical scenarios. These sets enhance treatment efficiency and adherence to evidence-based guidelines. Due to their typically narrow scope, most patients require multiple order sets concurrently to address all their medical conditions. For example, a patient with hypertension and diabetes will have one or more order sets for each of these conditions, as well as additional ones for other hospital-related treatments like pain management. This opportunity for overlap in order sets can lead to the unintended duplication of therapies, introducing potential medication errors, and heightening the risk of adverse drug events. Managing these redundancies is an administrative concern and a vital patient safety issue.

With the rise of big data and advanced analytical tools, data science has provided new opportunities to tackle complex healthcare challenges. Network analysis offers a framework to map and analyze relationships within the hospital order set ecosystems. This project applied network analysis to identify patterns of therapeutic duplication that may not be evident through traditional data review processes. Network analysis was employed to uncover interconnections between medications across different order sets using a dataset derived from electronic medical records (EMRs) over a defined period. The insights generated have informed clinical committees to streamline order sets, which enhances patient safety and cost efficiency.

**Defining Duplicate Therapies**

Order sets are developed and refined through rigorous processes to ensure clinical effectiveness. It is important to note that not all instances of the same medication/therapeutic class appearing within an order set were considered duplication. Some redundancies are intentional, reflecting the structured nature of medical treatment. Pain management protocols often incorporate medications from the same therapeutic class at different intensity levels. Low pain scores may be treated with over-the-counter (OTC) analgesics like acetaminophen or ibuprofen. As pain severity increases, mild opioids such as codeine may be introduced, often in combination with an OTC agent. For severe pain, stronger opioids like fentanyl or hydromorphone are used. These intra-order set redundancies are clinically justified and essential for flexible patient care. To avoid inflating the network with misleading duplication signals, the definition of duplicate therapies was refined to focus on medications from concurrently ordered, distinct order sets rather than intra-set redundancies.

**Background**

Medical order sets serve as pre-configured templates to standardize, streamline, and enhance treatment efficiency in healthcare settings. They are designed to reduce variability in practice and align care with evidence-based guidelines. Hospitals rely on these sets for a range of scenarios, from routine care to emergency interventions. Managing all these order sets presents significant challenges for clinicians and clinical specialists. With hundreds or thousands of order sets maintained across different departments and service lines, unintended redundancies and therapeutic conflicts are common yet difficult to detect due to the complexity and volume of data. Regularly updating and reviewing order sets is resource-intensive and prone to human error, making it challenging to keep pace with evolving clinical guidelines.

Network analysis offers a powerful tool to systematically evaluate these relationships. By modeling order sets as an interconnected network, this project aimed to reveal hidden patterns of duplication, providing actionable insights to improve patient safety and operational efficiency. This initiative aligns with broader industry efforts to incorporate technology into healthcare management, ultimately lowering costs and enhancing patient outcomes.

**Statement of the Business Problem or Challenge**

This capstone project addressed the issue of duplicate therapies within medical order sets across multiple healthcare facilities. While order sets are designed to enhance efficiency and standardization, unintended redundancies can lead to unnecessary medication exposure, increased risk of drug interactions, and inflated costs. These duplications also complicate inventory management and place additional strain on healthcare resources. It is also worth noting that regulatory bodies that certify healthcare facilities for reimbursements from the Centers for Medicare and Medicaid (CMS) look for duplicate therapy and can and will cite these as findings during facility reviews.

Identifying duplicate therapies was challenging due to the sheer volume of data, the diversity of clinical contexts, and the complexity of medical terminology. Additionally, the dynamic nature of healthcare, with frequent updates to treatment guidelines, made maintaining optimized order sets an ongoing challenge.

Duplicate therapies introduced multiple risks:

* Increased likelihood of medication side effects and adverse drug interactions.
* Unnecessary resource allocation, leading to higher operational costs.
* Added burden on clinical staff for oversight and correction.
* Sanctions from regulatory bodies like Centers for Medicare and Medicaid

Effectively addressing this issue required advanced analytical techniques capable of detecting patterns in complex, interconnected data. Network analysis offered a structured approach to identifying and visualizing these duplications, providing a foundation for targeted improvements in order set management. By leveraging data science methodologies, this project developed strategies for optimizing order sets, improving both patient safety and healthcare efficiency.

**Project Objectives**

The primary objective of this project was applying network analysis to identify and analyze duplications within medical order sets across multiple healthcare facilities. The project sought to achieve several specific goals. First, mapping interdependencies by constructing a comprehensive network that visualized relationships between **medications appearing within order sets**, highlighting areas of overlap and redundancy. Second, identifying duplication patterns by using network metrics to quantify duplication trends and assess their impact. Third, informing clinical committees by providing data-driven insights to inform clinical committees, which may consider streamlining order sets to eliminate unnecessary redundancies.

Beyond these primary objectives, the project also aimed to accomplish several secondary objectives. Enhancing data-driven decision-making by equipping healthcare administrators with quantitative insights for more informed decision-making was a key goal. Additionally, promoting patient safety by minimizing the risk of adverse drug interactions and improving overall care quality was a central aim. Furthermore, improving resource utilization by reducing waste and improving efficiency in healthcare delivery rounded out the project's objectives.

The methodological approach to achieve these objectives involved several key steps. Data collection aggregated electronic medical record (EMR) data related to order sets used over a specified period. Network construction built a network graph where nodes represented individual order sets and edges represented shared therapeutic components. Analytical techniques applied centrality analysis to identify influential order sets and clustering methods to detect patterns of duplication. Evaluation of findings assessed the impact of duplications on resource use and patient safety, correlating findings with clinical outcomes where possible.

The project delivered several key outputs. A detailed report analyzing the order set network, including visualizations and duplication trends, provided comprehensive insights. Additionally, data-driven insights provided to clinical committees for potential action offered practical application of the project's findings.

**Significance of the Project**

This project provided a systematic approach to identifying and addressing duplicate therapies within order sets. The project contributed to both patient safety and healthcare efficiency. Streamlining these sets helped minimize the risk of medication errors and reduced unnecessary healthcare spending. Beyond immediate clinical benefits, this project underscored the broader role of data science in transforming healthcare management. By demonstrating the utility of network analysis in identifying inefficiencies, it encouraged further adoption of analytical techniques in hospital administration. The methodologies developed here are scalable and adaptable, offering a framework that can be applied across different healthcare systems. By establishing a replicable model for order set optimization, this project supports the broader adoption of evidence-based, data-driven decision-making in healthcare administration.

**Definition of Terms**

To ensure clarity and consistent understanding throughout this document, the following terms are defined:

Order Sets: Standardized collections of medical orders designed for specific clinical scenarios or conditions that provide clinicians with evidence-based treatment protocols.

Electronic Medical Records (EMR): Digital versions of patients' medical histories including diagnoses, medications, treatment plans, and test results maintained by healthcare providers.

Therapeutic Duplication: The prescribing of multiple medications from the same therapeutic class or with similar mechanisms of action, potentially leading to additive toxicity without increased therapeutic benefit.

Network Analysis: A method of data analysis that focuses on the relationships (edges) between entities (nodes) to identify patterns, influences, and structures within complex systems.

Centrality Analysis: A set of methods within network analysis that identify the most important or influential nodes in a network based on their connections and positions.

Clinical Decision Support (CDS): Electronic systems designed to assist healthcare providers in clinical decision-making by providing relevant information and knowledge.

Adverse Drug Events (ADEs): Injuries resulting from medical intervention related to a drug, including medication errors, adverse drug reactions, and drug-drug interactions.

Computerized Provider Order Entry (CPOE): An electronic process that allows healthcare providers to enter medical orders directly into a computer system for processing.

**Assumptions, Limitations, and Delimitations**

Several assumptions, limitations, and delimitations shaped the scope and execution of this project. The primary assumption was that the electronic medical record data accurately reflected actual clinical practice and medication administration. It was also assumed that the order sets examined were actively used by clinicians rather than existing as dormant templates within the system. Furthermore, the project operated under the assumption that reducing therapeutic duplication would result in improved patient outcomes and reduced costs.

Limitations of this project included the retrospective nature of the data analysis, which prevented real-time assessment of decision-making processes leading to order set selection. Data quality issues, such as incomplete or incorrectly documented order set usage, potentially affected the comprehensive mapping of medication relationships. Additionally, the project was limited by the inability to directly measure patient outcomes resulting from identified duplications, relying instead on established literature connecting therapeutic duplication to adverse events.

This project was deliberately delimited to focus on medication duplications across order sets rather than examining all aspects of order set optimization. The analysis concentrated on identifying common patterns of duplication rather than evaluating individual clinical decisions. Geographically, the study was confined to healthcare facilities within a specific healthcare system, potentially limiting generalizability to facilities with different patient demographics, clinical specialties, or electronic health record systems.

**Conclusion**

This chapter has established the foundation for understanding the challenge of therapeutic duplication within medical order sets and the application of network analysis as a data science approach to address this issue. The dual imperative of enhancing patient safety while improving operational efficiency underscores the significance of this project. By mapping the complex relationships between order sets and identifying patterns of medication duplication, this project aimed to provide healthcare administrators and clinical committees with actionable insights to optimize clinical decision support tools. The methodology combined data science techniques with healthcare domain knowledge to create a systematic approach for analyzing and addressing a pervasive challenge in modern healthcare delivery. The findings from this project contribute to the growing intersection of data science and healthcare management, demonstrating how advanced analytical methods can be applied to improve clinical care processes.

**Project Organization**

The remainder of this document is organized into four chapters. Chapter 2 presents a comprehensive review of the literature, exploring existing research on therapeutic duplication, order set management, and applications of network analysis in healthcare. Chapter 3 details the data science methodology employed in this project, including data collection procedures, preprocessing steps, network construction techniques, and analytical approaches. Chapter 4 discusses the analysis results, presenting the identified patterns of duplication and their implications for clinical practice. Finally, Chapter 5 provides a summary of the project's findings, recommendations for healthcare administrators and clinical committees, limitations of the study, and suggestions for future research in this area. Throughout these chapters, the document maintains a focus on how data science methodologies can enhance healthcare delivery systems through the identification and resolution of therapeutic duplications.

**Chapter 2: Review of the Literature**

**Introduction**

The safe and effective use of medications is a foundational goal of modern healthcare, yet challenges persist even in technologically advanced systems. Clinicians have continued to prescribe duplicate medications and therapies, instances where patients receive multiple drugs with overlapping therapeutic effects, despite efforts to reduce such errors. These duplicates may arise unintentionally through fragmented workflows, poorly designed electronic ordering interfaces, or ineffective clinical decision support (CDS) mechanisms. The consequences, ranging from adverse drug events to diminished therapeutic efficacy and increased healthcare costs, underscore the need for systematic identification and mitigation of duplicate therapy risks. This project seeks to examine the problem of duplicate therapy within standardized order sets, using methods drawn from network analysis to uncover prescribing patterns that contribute to clinical risk. The purpose of this chapter is to provide a comprehensive review of the literature that supports this approach.

The review opens by examining how frequently duplicate medication and therapy orders occur, especially in settings where Computerized Provider Order Entry (CPOE) and CDS systems are in place. Following this, the review examines clinical and systemic risks associated with such errors, including the role of CDS design, alert fatigue, and communication breakdowns in perpetuating the problem. The chapter then turns to the growing body of research applying social network analysis (SNA) and graph theory in healthcare, particularly in medication safety and prescribing behavior studies. Through this review, the chapter provides a clear rationale for the application of network analysis in this project, demonstrating how it can uncover latent structures and inform targeted interventions in the ordering process.

The review incorporated literature identified through a targeted search of academic databases, emphasizing peer-reviewed articles from the past 15 years related to medication safety, order set design, clinical decision support systems, and network analysis in healthcare. Key sources include both empirical studies and systematic reviews, encompassing quantitative analyses of prescription patterns, case studies of CDS implementation, and methodological explorations of social network metrics applied to healthcare data.

Together, these studies provide both evidence of the problem and support for the analytical framework chosen for this project. The following sections review this literature in detail, organized into three thematic areas: (1) the prevalence of duplicate therapies, (2) the associated clinical and systemic risks, and (3) the application of network analysis to medication safety and ordering practices.

**Prevalence of Duplicate Medications and Therapies**

Despite widespread adoption of electronic health records (EHRs) and CPOE systems, the issue of duplicate medication and therapy orders remains a significant and well-documented concern in clinical practice. These duplicates can involve either the same drug ordered more than once, or multiple drugs from the same therapeutic class prescribed simultaneously, often without clinical justification. This issue is particularly problematic in environments that rely on standardized order sets, as overlapping defaults and inadequate safeguards frequently result in unintended duplicate entries.

Studies have quantified the persistence of duplicate ordering errors, even in systems with embedded CDS. In a landmark study, Wetterneck et al. (2011) found that the implementation of CPOE and CDS in a 400-bed teaching hospital was associated with a more than threefold increase in duplicate medication errors, rising from 1.16 to 4.16 errors per 100 patient-days post-implementation. Many of these errors involved exact duplicate orders or the same medication prescribed in different forms or dosages, often placed by multiple providers during team rounds or handoffs. These findings illustrate both the technical limitations of CDS systems in detecting clinically relevant duplicates and the contribution of workflow complexity to these errors.

Similarly, Bocknek et al. (2022) reviewed 377 incident reports related to duplicate medications and found that over 80% represented true medication errors. Many of these were attributed to flaws in system design, such as defaulted orders in pre-built order sets and inadequate alerting mechanisms. These findings reinforce the notion that duplication errors are not only common but often systemic in origin.

Recent reviews of broader medication safety concerns echo these findings. Saatchi et al. (2023), in a systematic scoping review of social network interventions in healthcare, noted that coordination and communication breakdowns—particularly in high-complexity environments like inpatient medicine—frequently result in redundant or conflicting orders. Inadequate feedback loops and poor integration across multidisciplinary teams exacerbate these issues.

Although CDS systems aim to prevent such errors, studies show that clinicians frequently bypass, misunderstand, or misconfigure them. Duplicate-detection alerts often trigger too easily for clinically appropriate co-therapies or miss meaningful therapeutic overlaps when providers order drugs via different routes or frequencies. This is consistent with observations by Wetterneck et al. (2011), who noted that even "true positive" alerts were commonly overridden, particularly when buried among high volumes of irrelevant or redundant warnings.

Taken together, these findings demonstrate that duplicate medication and therapy orders remain both prevalent and under-detected in modern clinical environments. The literature makes it clear that while CPOE and CDS systems have advanced medication safety in many respects, they have not resolved the problem of duplication. In fact, in some settings, they have introduced new risks by shifting errors from handwriting and transcription to order set design, interface usability, and communication practices.

**Risks to Patient Care and Systemic Failures**

Duplicate medication and therapy orders are more than just administrative or procedural nuisances; they represent serious risks to patient safety and healthcare quality. At best, they may result in redundant therapy and increased cost. At worst, they can contribute to adverse drug events (ADEs), medication overdoses, or interactions that significantly compromise patient outcomes. These risks intensify in acute care settings, where high patient volumes, rapid decisions, and team-based care models add complexity to the ordering process.

Research has consistently shown that duplicate therapy errors often bypass system safeguards and reach the patient. Wetterneck et al. (2011) found that CPOE and CDS systems failed to prevent duplicate orders, particularly when the same drug was ordered via different routes or when one order had already been administered. Because the underlying medication database treated oral and intravenous forms as distinct products, duplicate checking algorithms often failed to trigger alerts. Even when presented by alerts, providers frequently overrode them, sometimes without reviewing the content, due to alert fatigue and poor interface design.

Bocknek et al. (2022) also observed that duplicate orders were commonly associated with transitions of care, shift changes, and high-alert medications. Despite the presence of CDS, clinicians still made errors during routine tasks such as electrolyte replacement, ordering analgesics, or initiating overlapping anticoagulant therapies. These observations align with the broader theme that CDS systems, in their current form, are often insufficient to fully mitigate duplicate therapy risk, especially in fast-paced, cognitively demanding clinical environments.

Systemic failures extend beyond individual alerts. Francis et al. (2024) emphasized that communication breakdowns, lack of clear role delineation, and insufficient understanding of team-based workflows are major contributors to medication-related errors. Their review of SNA applications in healthcare noted that many quality and safety issues arise from unexamined interpersonal and interdepartmental relationships, which influence how information is shared and how decisions are made.

Moreover, Smit et al. (2020) found that few healthcare interventions address the complexity of these dynamics during implementation planning. Their scoping review revealed that while SNA is well-positioned to uncover structural weaknesses in communication networks, most intervention studies fail to apply it during the development phase—missing opportunities to anticipate and correct the very conditions that allow duplicate ordering errors to persist.

Locke et al. (2019) contributed an important perspective by evaluating a pharmacy-led intervention program in a Texas hospital. This initiative used four-hourly report generation to identify duplicate pain medication orders based on shared pain scale, route, and drug class. Pharmacists identified 172 duplications over three weeks where more than 90% involved opioids. Most cases resulted in pharmacist- or physician-led interventions. The most frequent solution involved adjusting the pain scale associated with one or more of the medications. The program demonstrated that pharmacy oversight combined with structured review processes can significantly reduce duplicate therapy events, especially within high-risk medication classes. It also highlights the value of tailored approaches, where clinicians assess not only the existence of duplicates but their clinical context and risk severity.

These studies converge on a critical point: the persistence of duplicate therapy errors is not simply a technological problem—it is a systemic one. Factors such as EHR configuration, alert design, staffing patterns, communication practices, and organizational culture all play a role. Addressing these challenges requires tools capable of illuminating these interdependent relationships, which often remain hidden in traditional workflow analyses. This sets the stage for network analysis as a promising and underutilized approach to this longstanding clinical problem.

**Network Analysis as a Solution**

Given the complex, interconnected nature of clinical workflows and the persistent limitations of existing decision support tools, network analysis has emerged as a powerful method for uncovering hidden structures and inefficiencies in healthcare systems. In the context of medication safety, researchers are increasingly using SNA and related graph-based techniques to examine co-prescribing behavior, clinical collaboration patterns, and systemic risk factors related to medication use.

Askar et al. (2021) provide a comprehensive introduction to network analysis methodologies applied to medication use. They describe the construction of drug prescription networks, in which medications are represented as nodes and their co-prescription relationships as edges. This framework enables researchers and clinicians to visualize and quantify relationships that would otherwise remain obscured in transactional data. Importantly, the study highlights the utility of common graph metrics—such as centrality, modularity, and density—for identifying patterns of polypharmacy and clusters of potentially problematic prescribing behaviors.

Cavallo et al. (2013) applied similar principles in their analysis of drug prescription data in a regional health system. Their findings revealed that prescription networks often exhibit scale-free properties, meaning only a few medications serve as hubs with disproportionately high co-prescription rates. This network structure has direct implications for medication safety interventions, as it suggests that targeting a small set of highly connected nodes may have an outsized impact on reducing duplication risk.

Bazzoni et al. (2015) extended this work by applying network analysis to co-prescription data in elderly populations. They found that the resulting networks were dense, highly clustered, and modular, reflecting the clinical complexity and interdependence of therapies in older adults. Notably, they emphasized the value of modularity analysis in identifying therapeutic clusters—groups of medications frequently prescribed together—which could signal opportunities for standardization or, conversely, red flags for potential duplication.

These network-based approaches provide a system-level view of medication prescribing, moving beyond the binary “duplicate-or-not” logic typically used in CDS alerts. This is especially relevant for order sets, where predefined combinations of medications may reinforce co-prescription habits that are clinically questionable or redundant. By examining how these combinations occur across populations and specialties, network analysis can reveal which order sets—and which medications within them—are disproportionately associated with duplicate therapy risk.

Researchers have also applied network analysis to broader clinical service patterns beyond medications. Niyirora and Aragones (2019) used both social network analysis and principal component analysis to detect clusters of clinical services in emergency department workflows. Their methods for visualizing and analyzing service networks offer a promising framework for translating similar techniques to medication ordering patterns, particularly when investigating how specific drugs or therapeutic classes cluster within order sets.

Brunson and Laubenbacher (2018) further support the validity of this approach through their systematic review of network analysis applied to routinely collected healthcare data. They catalog a wide array of use cases, from physician collaboration networks to clinical co-occurrence and institutional exchange patterns. Their review reinforces the point that network analysis is not only methodologically sound but increasingly feasible and scalable in real-world healthcare settings, thanks to the availability of rich transactional datasets and open-source graph analytics tools.

Together, these studies present a compelling case for the use of network analysis in addressing duplicate medication orders. By shifting focus from isolated transactions to the broader relational context of prescribing, network-based approaches can uncover risk structures that would remain invisible to traditional CDS systems. In doing so, they offer a path forward not just for identifying duplicate therapy, but for redesigning the systems and workflows that enable it.

**Conclusion**

The literature reviewed in this chapter underscores the persistent and systemic nature of duplicate medication and therapy orders in modern clinical environments. Despite the widespread implementation of CPOE and CDS systems intended to reduce medication errors, duplicate orders continue to occur with surprising frequency—often due to workflow complexity, alert fatigue, inadequate CDS logic, and the limitations of order set design. These errors not only increase the risk of adverse drug events but also undermine clinician trust in safety systems and contribute to inefficiencies in patient care.

The studies reviewed demonstrate that the issue is not merely a matter of individual behavior or user interface design but reflects deeper organizational and structural dynamics. Communication failures, inconsistent role delineation during care transitions, and fragmented decision-making processes all contribute to the perpetuation of duplicate therapy risks. These findings emphasize the need for solutions that move beyond isolated alerts and focus instead on understanding the systemic patterns and relationships underlying medication ordering behaviors.

Network analysis emerges from this review as a particularly well-suited methodology to address the complexity of duplicate therapy detection. By representing medications, prescribers, or order sets as nodes in a graph, and analyzing the connections among them, researchers can uncover high-risk clusters, central hubs, and patterns of co-prescription that are not readily apparent through traditional reporting or rule-based alerts. Metrics such as centrality, modularity, and density provide a richer vocabulary for evaluating prescribing behavior and identifying leverage points for intervention.

This chapter has laid the foundation for the analytical approach that follows in this project. Chapter 3 will detail the methodology used to construct and analyze order set and medication networks using institutional data. Guided by the insights from this literature, the goal is to identify specific patterns and structures that contribute to duplicate therapy risk and to inform strategies for improving medication safety at both the system and clinical decision-making levels.

**Chapter 3: Data Science Application**

**Introduction**

This project applied graph theory to identify duplicate therapy risks embedded in clinical order sets. Medications were modeled as nodes, and edges represented co-occurrence relationships derived from real-world ordering behavior. Graph theory was chosen over conventional tabular modeling due to its strength in capturing interdependencies and revealing latent community structures. Traditional classification and clustering methods, while initially considered, were set aside because they require flattening of the data and consequently obscure the very relationships that graph analysis is designed to illuminate. Early exploratory work revealed that therapeutic class, as a grouping construct, was too coarse and inconsistent to support meaningful duplication analysis. Pharmaceutical subclass was ultimately selected for its tighter clinical specificity and superior alignment with clinical decision-making.

The analysis relied on data extracted from an internal data warehouse. SQL queries were used to retrieve relevant data tables, which were exported as CSVs and processed externally using Python in a Jupyter Notebook. Python’s data science ecosystem enabled a robust and scalable approach. Pandas was used for data wrangling, NetworkX powered graph construction and analysis, and the community-louvain package facilitated modularity-based clustering. Visualization tools included Matplotlib and Seaborn for static plots and Plotly for interactive community visualizations.

The foundation for analysis was built on the integration of multiple curated datasets:

* Patient\_Orderset\_c\_Start\_End.csv: This file provided patient-level linkage to order sets, allowing for a temporal window of prescribing behavior. Only orders placed via standardized order sets during the selected time frame were retained.
* Medication-Orderset pairs.csv: This dataset outlined all observed medication pairings within order sets. It defined co-occurrence frequency and enabled pairwise edge construction within the graph.
* Medications\_Detail.csv: This file enriched the analysis with medication-level metadata, including pharmaceutical class, subclass, and route.
* Ordersets.csv: This file translated cryptic order set IDs into readable clinical terms to support visual interpretation.

Cleaning involved removal of ambiguous entries, duplicates, and rows containing non-medication entities or null subclass information. The final dataset represented approximately 800,000 valid co-occurrences.

The chapter now includes the original Chapter 4 material—analysis, findings, and interpretive visuals—organized into the data science lifecycle to meet course expectations. New references and deeper dataset context have also been added.

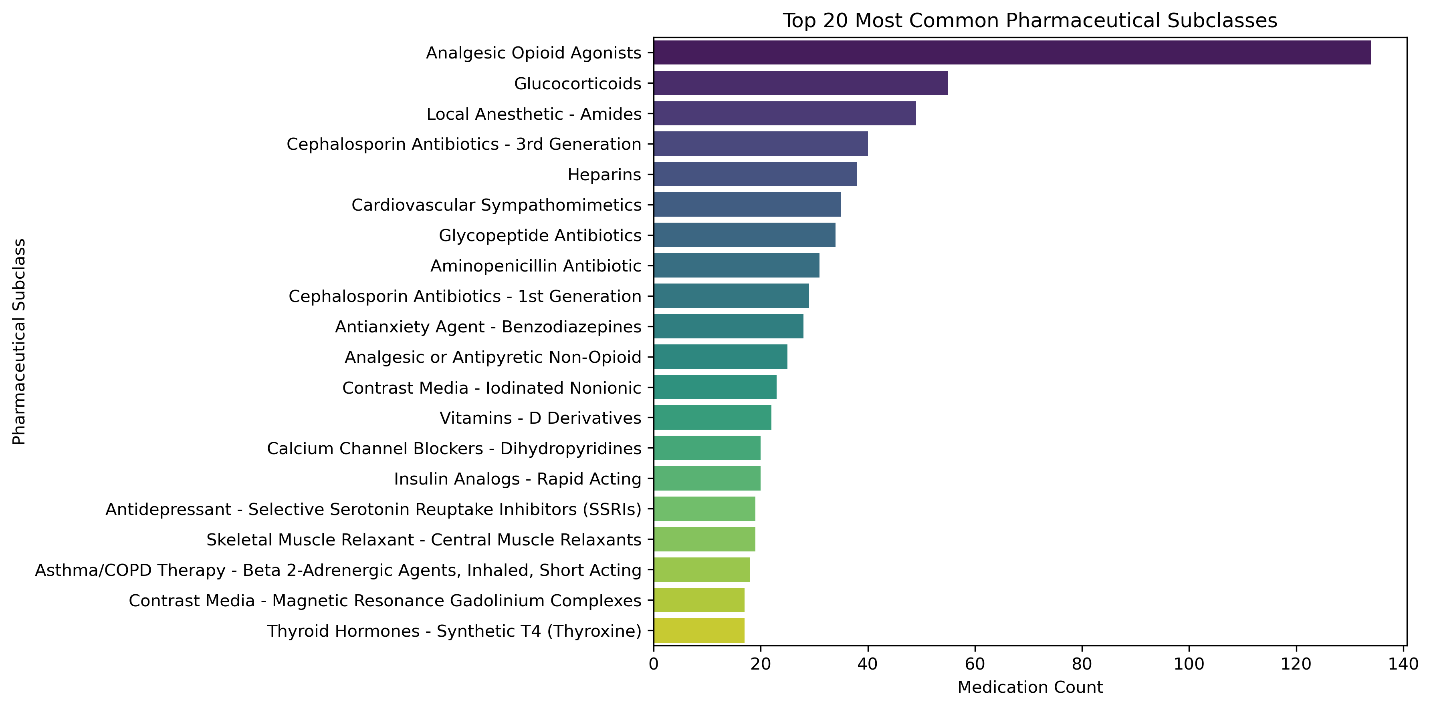
**Exploratory Data Analysis**

The analysis began with a foundational review of medication and order set data to understand the scope, scale, and distribution of entries prior to network modeling. After filtering for valid medication-order set pairs and removing outliers or null values, the dataset consisted of 30,409 unique medication-order set pairs spanning 2,909 distinct medications across 1,109 clinical order sets.

A key focus of the exploratory process was understanding the breadth and frequency of pharmaceutical subclasses within the data. Figure 1 below shows the top 20 most common subclasses based on their overall presence across order sets. Analgesic opioid agonists were by far the most frequent subclass, followed by glucocorticoids (steroids), local anesthetics, cephalosporin antibiotics, etc. This subclass-level distribution reflects typical hospital prescribing patterns, but it also foreshadows some of the primary areas of concern later identified in the duplicate therapy analysis.

**Figure 1**

*Top 20 Most Common Pharmaceutical Subclasses Across all Order Sets*

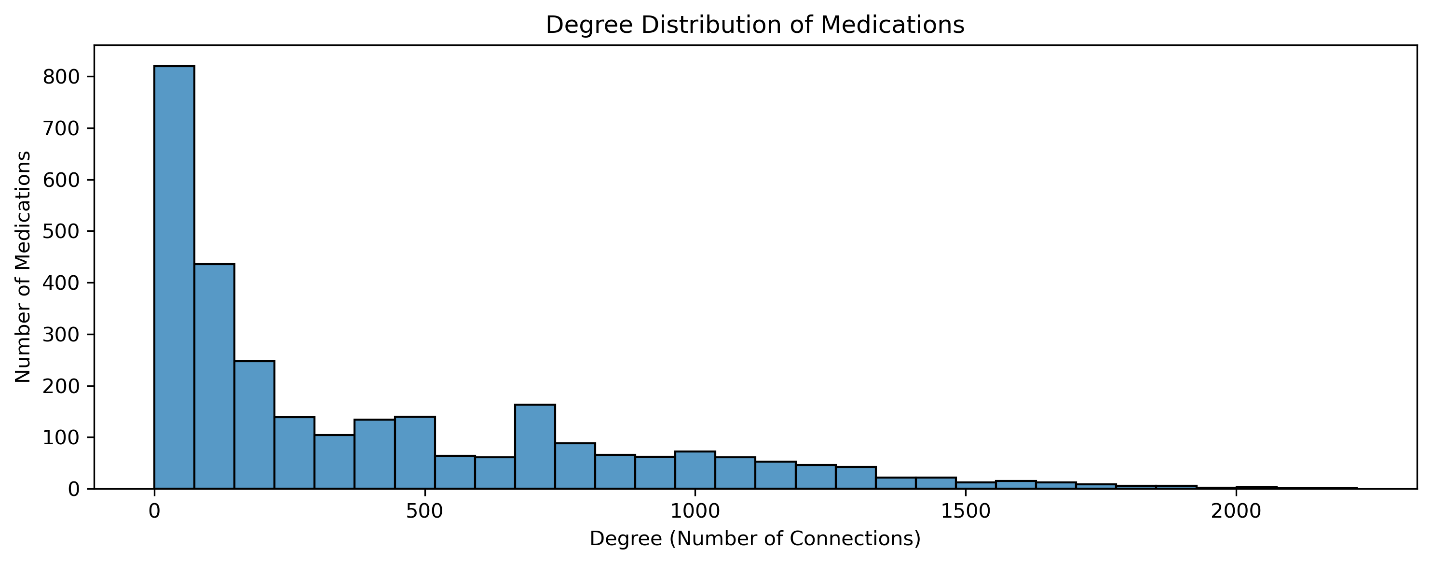


The exploratory review also included a comparison between the full medication co-occurrence network and the filtered duplicate therapy subnetwork. Both were constructed using the same data foundation but diverged in scope and complexity. The full network captures all medications that co-occur in the same order set, whereas the duplicate network isolates only those co-occurrences identified as potential clinical redundancies.

Figure 2 presents the degree distribution of the full co-occurrence network. As shown, the distribution is right-skewed, with most medications connecting to relatively few others and a small set of high-degree medications acting as central hubs. These hub medications often appear across many clinical contexts and serve as backbone nodes in the network.

**Figure 2**

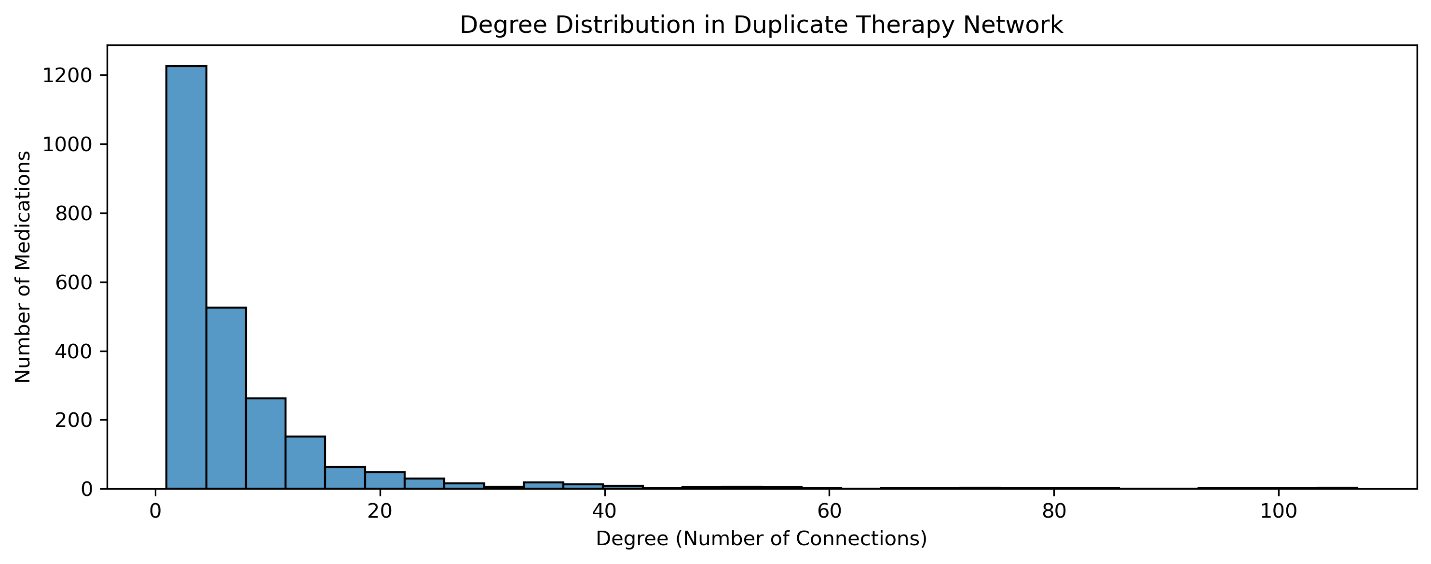
*Degree Distribution for the Full Medication Co-occurrence Network*



In contrast, Figure 3 shows the degree distribution for the duplicate therapy network. This network is far more selective, with fewer total edges and a much flatter degree distribution. The result is a sparser, more targeted network that pinpoints potential risk zones rather than general co-use. This structural difference highlights the analytical benefit of filtering by duplicate criteria thus reducing noise and clarifying actionable insights.

**Figure 3**

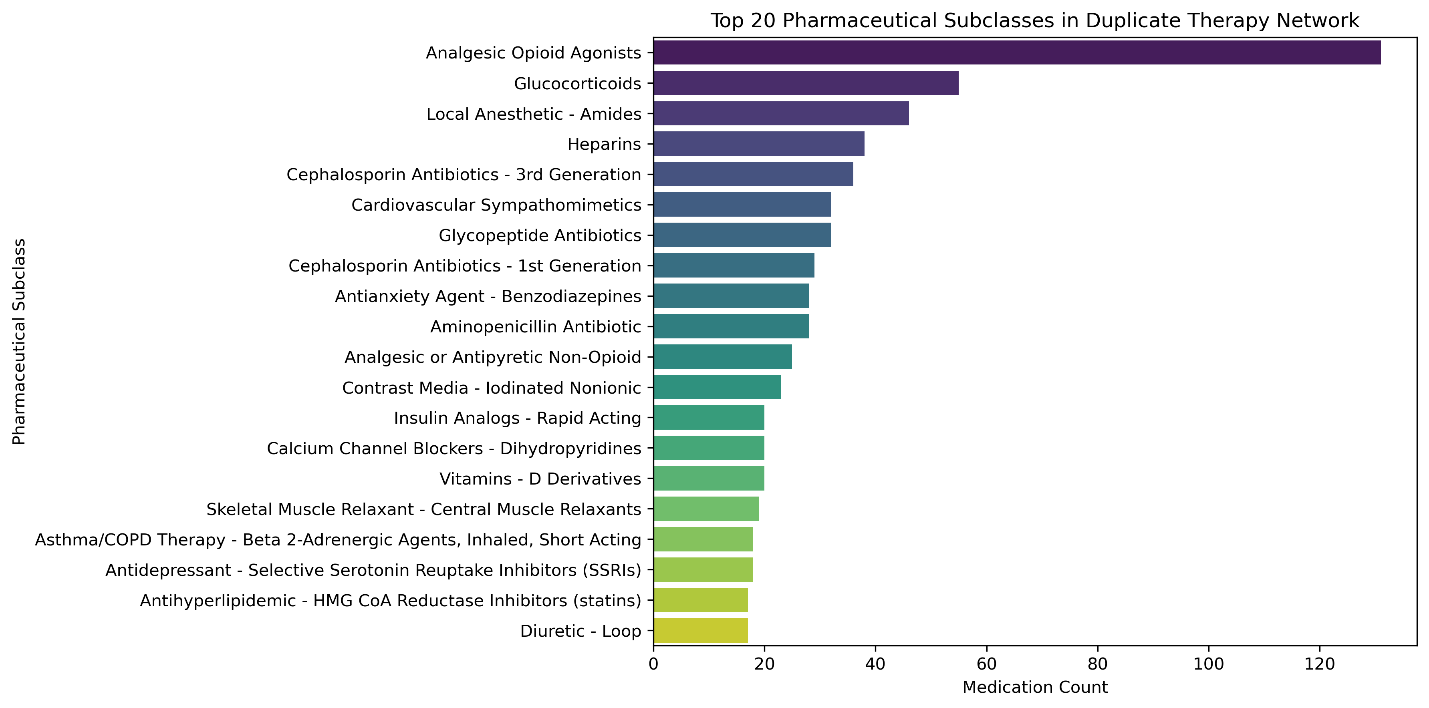
*Degree Distribution for the Duplicate Therapy Network*



These differences are also evident in the subclass distributions of each network. As shown in Figure 1, the full co-occurrence network includes a broad array of therapeutic areas reflecting general prescribing frequency. In contrast, the duplicate therapy network (Figure 4) emphasizes not just commonly used medications, but those with a higher propensity for therapeutic redundancy. While many top subclasses appear in both networks, the duplicate therapy version shows a stronger presence of drug classes prone to multi-formulation duplication, including statins, loop diuretics, and SSRIs, highlighting subtle but important shifts in risk concentration.

**Figure 4**

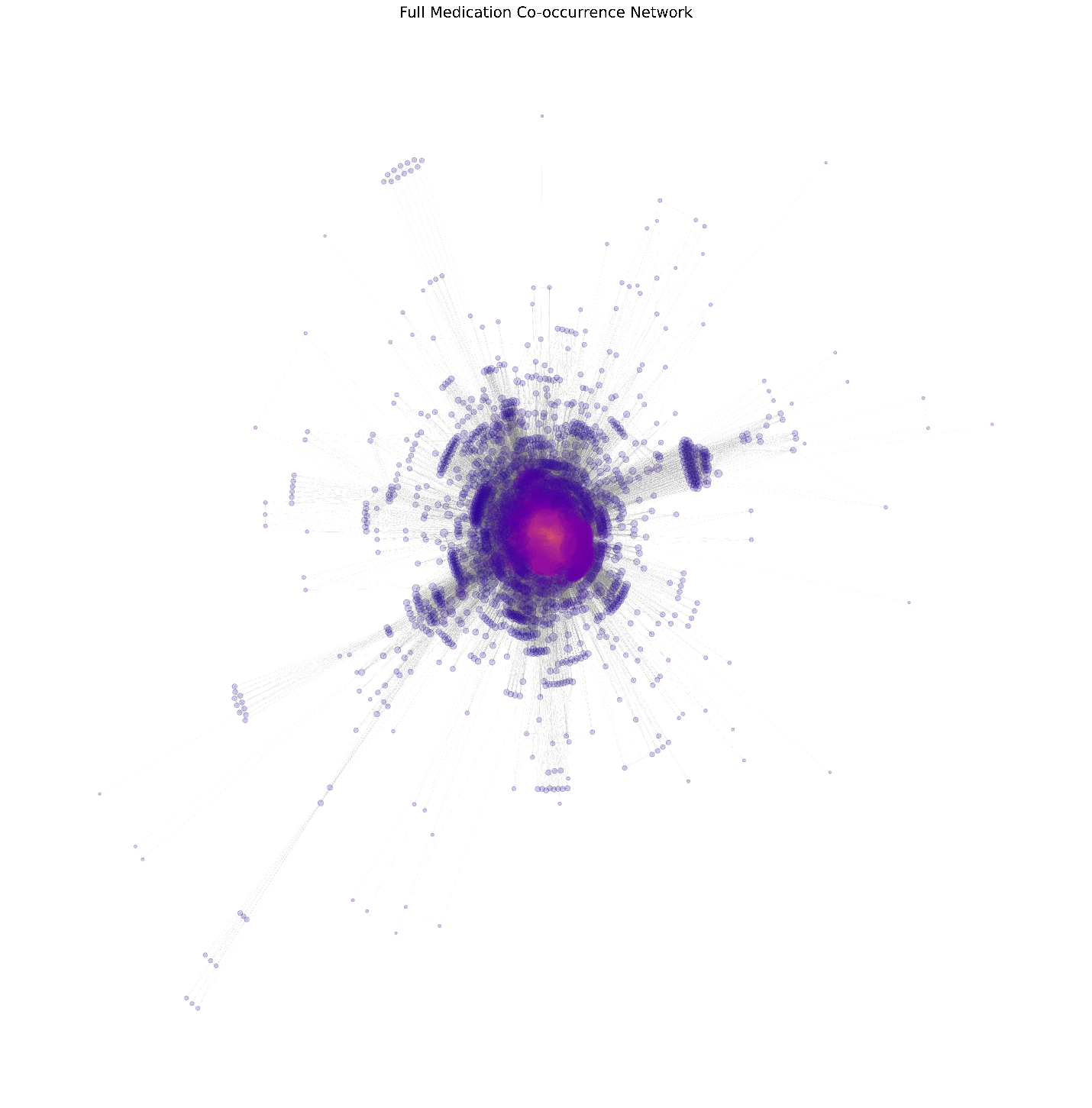
*Top 20 Pharmaceutical Subclasses in the Duplicate Therapy Network*



To visualize the extent of interconnection even before filtering, a network graph of the largest connected component of the full co-occurrence network was generated. This structure, shown in Figure 5, visually reinforces the analytical complexity of medication data at scale.

**Figure 5**

*Visualization of the Largest Connected Component of the Full Medication Co-occurrence Network*



Even after isolating the largest cluster of medication connections, the resulting network forms a dense, entangled structure that defies intuitive understanding. This visualization underscores the analytical challenge posed by raw co-occurrence data and reinforces the need for structured analysis techniques such as centrality scoring and community detection. The dense central core represents highly connected, frequently co-ordered medications, while peripheral nodes and loosely linked branches suggest specialized or context-specific drug clusters.

**Medication Co-occurrence Network**

To move beyond descriptive analysis and explore underlying structure, a medication co-occurrence network was constructed using all medication-order set pairings. In this undirected graph, each node represents a unique medication, and edges are formed between medications that appear together in the same order set. The resulting network consisted of 2,909 nodes and 584,419 edges, indicating a dense landscape of co-prescription relationships.

Initial inspection of the network structure revealed a high degree of connectivity, consistent with the role of certain medications as frequent components across multiple care pathways. The degree distribution (Figure 2) confirms a right-skewed pattern, with a minority of medications connecting to a disproportionately large number of others. These high-degree nodes typically represent widely used drugs—such as analgesics, fluids, or common antibiotics—that serve as anchors in clinical workflows.

The network also includes a small number of disconnected components, but most medications are part of a single giant connected component. This was visualized in the previous section (Figure 5), showing the largest cluster of interconnected medications as a dense and tangled structure. The complexity of this unfiltered network emphasizes the need for downstream techniques, such as filtering for duplicate therapy edges and applying community detection, to surface more interpretable and actionable patterns.

To better understand the dominant therapeutic relationships within this network, subclass frequency was also visualized (Figure 1). Analgesic opioid agonists again emerged as a dominant force, reinforcing their prevalence in general prescribing patterns and hinting at their central role in subsequent duplication risk.

**Duplicate Therapy Network**

While the full co-occurrence network provides a broad view of medication relationships across order sets, it includes all co-prescribed medications regardless of therapeutic relevance or clinical risk. To refine this view, a duplicate therapy network was constructed by filtering the original data for medication pairs identified as potential duplicates. This network emphasizes co-occurrences within the same therapeutic subclass or class that may represent therapeutic redundancy, dosing overlap, or unnecessary combination.

The resulting duplicate therapy network contains 2,402 nodes and 8,960 edges, reflecting a significant reduction in both scale and complexity compared to the full network. However, this smaller network is more targeted, capturing the subset of medications most likely to be implicated in clinical duplication scenarios.

The degree distribution for this network, shown in Figure 4.3, is much flatter than that of the full network. While a small number of medications still exhibit relatively high connectivity, most medications participate in only a few duplicate pairs. This structural profile reflects the focused scope of the network: rather than capturing every co-prescription, it isolates likely sources of redundant therapy.

Therapeutic subclass analysis reinforces this shift in focus. Figure 4.4 shows the most common pharmaceutical subclasses represented within the duplicate network. While many of the top subclasses mirror those in the full network, such as analgesic opioid agonists, glucocorticoids, and antibiotics, the duplicate network exhibits greater emphasis on drug classes prone to multi-formulation prescribing, including loop diuretics and statins. These patterns suggest areas where therapeutic duplication is most likely to arise due to parallel therapies, dosage overlaps, or unclear clinical intent.

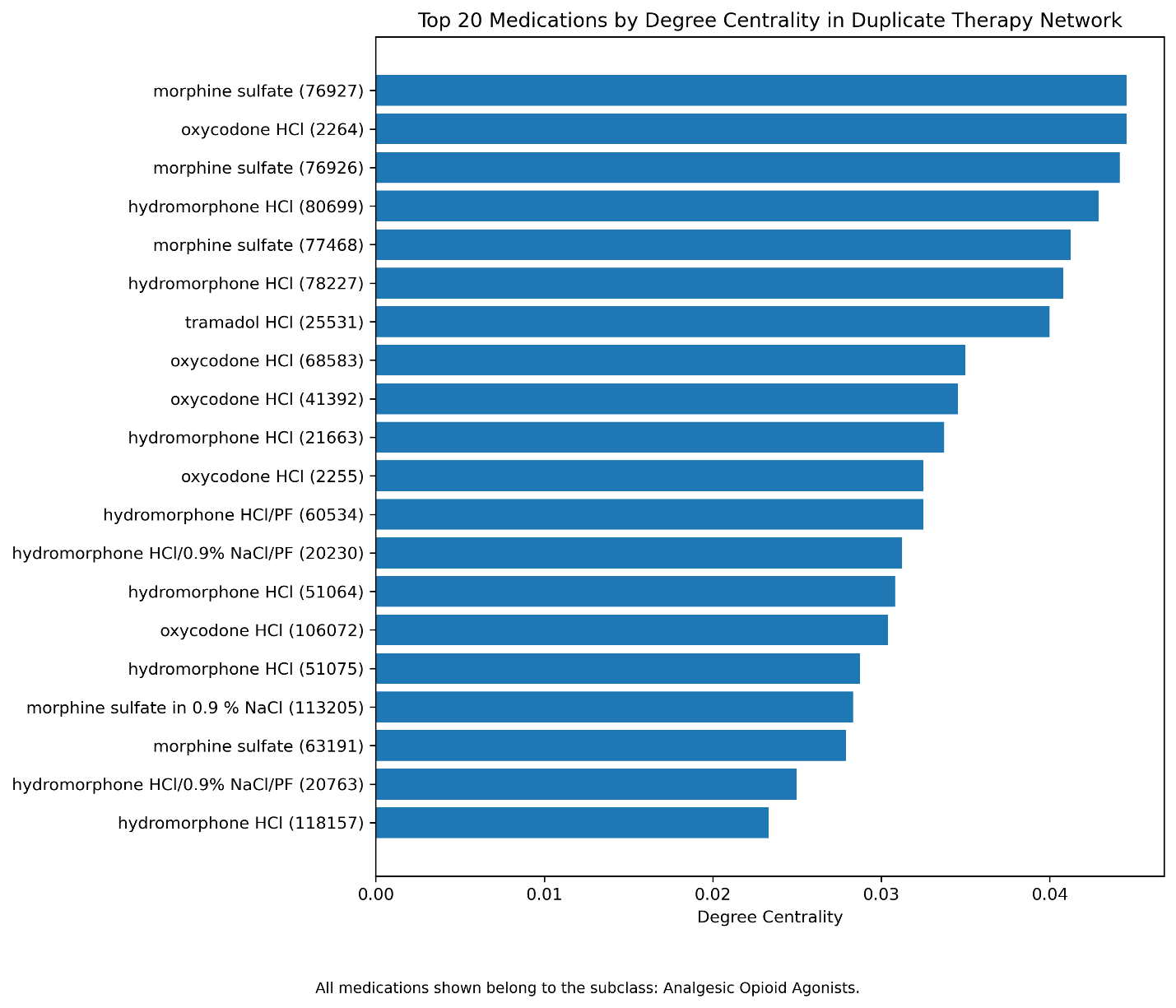
Network-level centrality metrics were also examined to identify medications with disproportionately high involvement in duplicate therapy. Degree centrality, which measures how many connections a node has, highlights medications that frequently appear in duplicate relationships (Evans & Chen, 2022). Betweenness centrality, which quantifies how often a node lies on the shortest path between other nodes, reflects a medication's importance as a potential bridge between clusters of duplicates (Xiang et al., 2023). Medications such as morphine, oxycodone, and hydromorphone appear repeatedly among the top-ranked nodes by both metrics, highlighting their frequent presence in potentially redundant combinations. These drugs are all part of the opioid analgesic class, reinforcing prior observations about their centrality in both prescribing frequency and duplication risk.

Although the duplicate therapy network reduces the overall complexity of the medication landscape, its internal structure still exhibits variation and clustering. This observation sets the stage for more advanced analysis using community detection methods, which are explored in the following section.

To complement the table-based metrics, Figure 6 presents a horizontal bar chart of the top 20 medications by degree centrality within the duplicate therapy network. Each medication is labeled by name and MedicationKey, an internal identifier, to distinguish formulations that share a common name. The visualization highlights the central role of opioid analgesics in duplicate relationships and emphasizes how certain medications, such as morphine and hydromorphone in various formulations, consistently rank as structural hubs within the network.

**Figure 6**

*Top 20 Medications by Degree Centrality in the Duplicate Therapy Network*

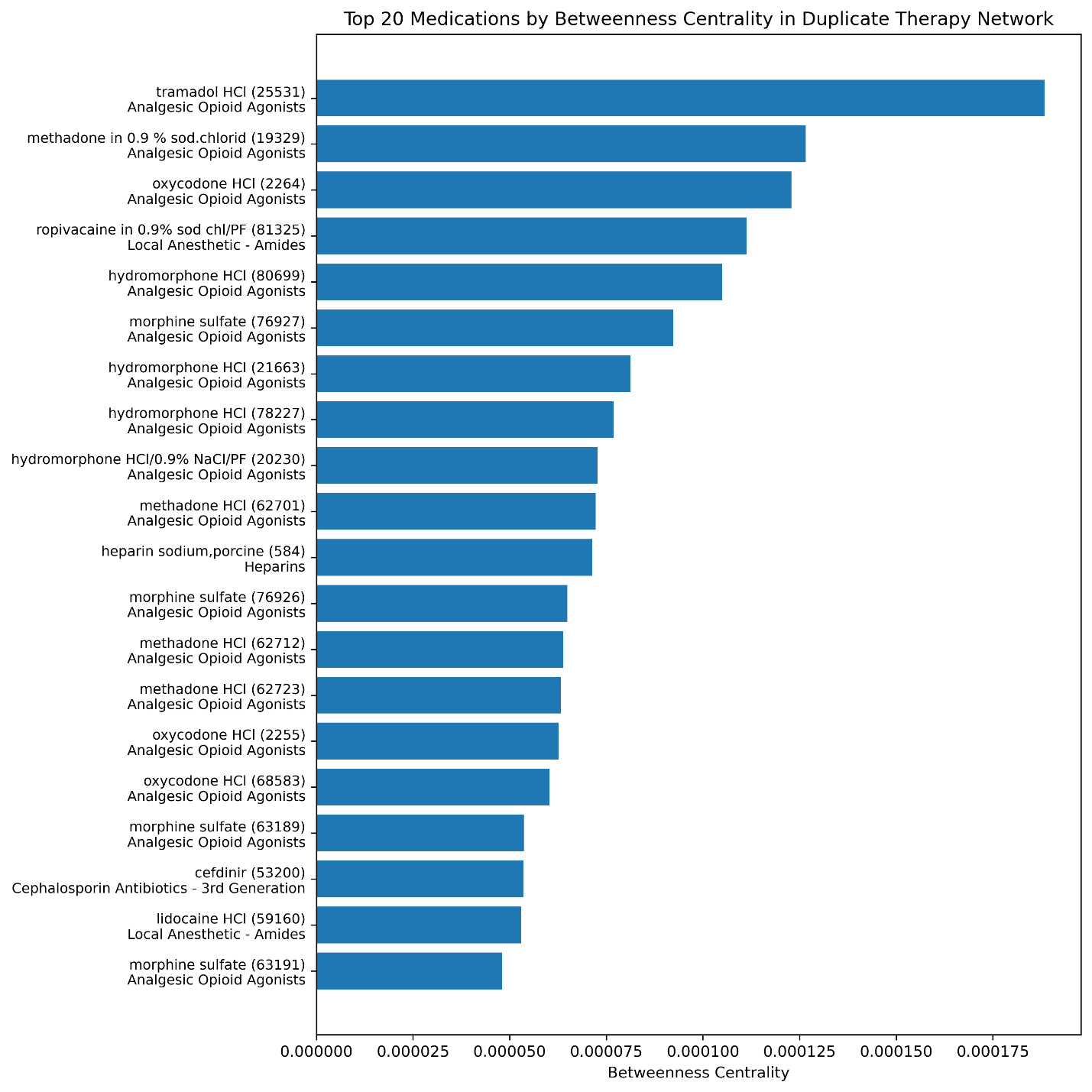


*Note.* All medications shown belong to the subclass: Analgesic Opioid Agonists.

To complement degree centrality, Figure 7 shows the top 20 medications by betweenness centrality, a measure of how frequently a medication sits on the shortest path between other medications in the duplicate therapy network. While many high-degree opioids still rank near the top, this view also surfaces bridge medications that may not be the most duplicated but play a pivotal role in connecting disparate parts of the network. Notably, several non-opioid medications appear on this list, including ropivacaine and lidocaine (local anesthetics), heparin sodium (an anticoagulant) and cefdinir (a third-generation cephalosporin). These results underscore how betweenness centrality can reveal clinically important medications that would otherwise be overlooked when focusing solely on raw frequency.

**Figure 7**

*Top 20 Medications by Betweenness Centrality in the Duplicate Therapy Network.*

*Note.* This measure highlights medications that act as bridges between duplicate-prone clusters, including a subset of non-opioid agents.

**Community Detection**

While centrality measures help identify key medications within the duplicate therapy network, they do not fully capture the meso-level structure, that is, how groups of medications tend to cluster together based on shared relationships. To explore this internal structure, community detection was applied to the network to identify clusters of medications that are more densely connected to each other than to the rest of the network. These clusters, or “communities,” can represent functionally similar medications, clinical practice patterns, or areas of therapeutic overlap that may not be immediately obvious.

The primary algorithm used was the Louvain method, a modularity-based community detection algorithm that is well-suited for large networks. Louvain works by optimizing modularity, a measure of how well the network can be partitioned into communities with dense internal connections and sparse connections between groups (Dugué & Perez, 2022). For validation and comparison, two alternative algorithms—greedy modularity and label propagation—were also tested.

Across all three methods, the number of communities detected ranged between 403 and 427. The Louvain algorithm achieved the highest modularity score of 0.9035, reflecting an exceptionally well-partitioned network. In general, modularity scores above 0.3 are considered meaningful in large networks; a score above 0.9 suggests that the duplicate therapy network contains highly distinct and internally cohesive clusters (Dugué & Perez, 2022).

These communities provide a powerful lens for interpreting duplicate risk—not just in terms of individual medications, but in terms of group-level behavior. In the following sections, both interactive and static visualizations are used to showcase the most prominent and internally consistent communities, offering insight into the therapeutic domains most prone to duplication.

**Community Structure and Visualization**

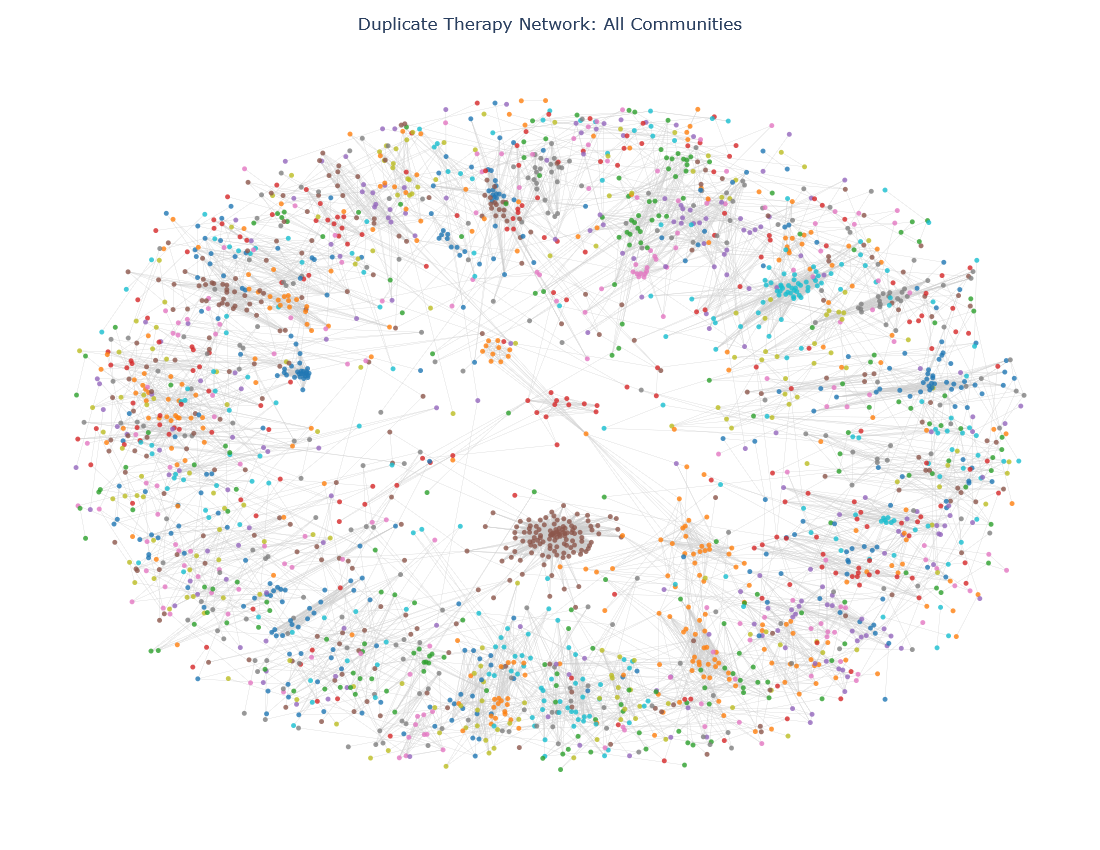
After applying the Louvain algorithm to the duplicate therapy network, a total of 403 distinct communities were detected. These communities vary in size, density, and therapeutic composition. To explore the full scope of this structure, an interactive visualization was generated to display the complete network layout with communities assigned by color. Each node represents a medication, and connections represent duplicate relationships. Node size is scaled by degree centrality, making highly connected medications visually prominent.

Figure 8 provides a static snapshot of this interactive layout. The full version is available in HTML format and allows for zooming, panning, and tooltip-based inspection of individual nodes. This interactive approach is particularly useful for navigating the network's complexity, which includes tightly packed clusters of opioid medications, antibiotics, anticoagulants, and other duplication-prone groups.

The spatial separation of communities reflects meaningful structural segmentation within the network. Medications in the same therapeutic subclass often appear clustered together, though some communities include multiple subclasses that frequently co-occur as duplicates. This layout serves as the foundation for the targeted visualizations in the following section, which highlight selected communities with high subclass consistency and duplication density.

**Figure 8**

*Community Layout of the Duplicate Therapy Network Using the Louvain Algorithm*



*Note.* Nodes are colored by community membership. An interactive version is available for detailed exploration.

**Subclass-Homogeneous Community Spotlights**

While the full network provides a structural overview of duplication risk, individual communities offer more focused insight into specific therapeutic areas where duplication is both common and clinically relevant. The communities highlighted here were selected for their high degree of subclass homogeneity meaning most medications in the group share the same pharmaceutical subclass. This structural consistency suggests frequent redundancy in practice, often within the same mechanism of action or formulation family.

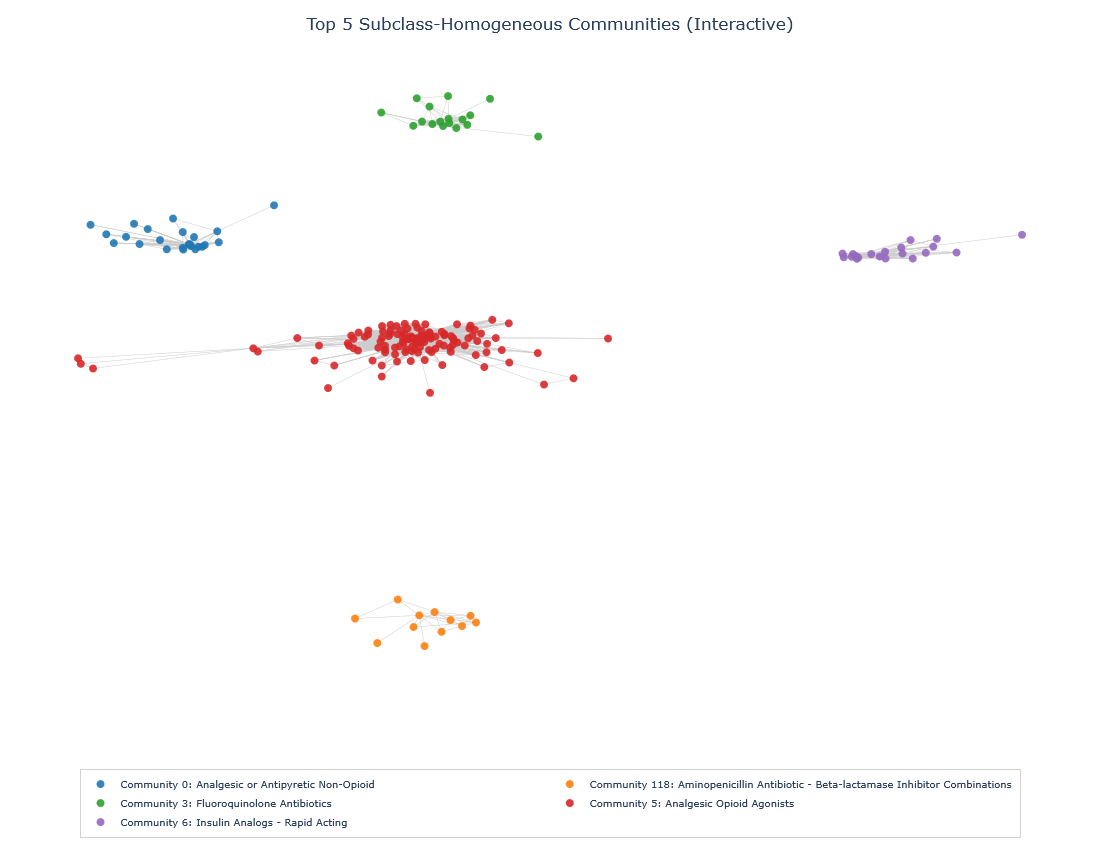
Figure 9 displays five such communities, each representing a distinct profile of duplicate therapy risk. Nodes represent medications, edges represent duplicate co-occurrences, and communities are color-coded and annotated with their dominant pharmaceutical subclass. The most densely connected cluster (red) is composed of Analgesic Opioid Agonists, a subclass that includes morphine, hydromorphone, oxycodone, and tramadol in multiple formulations. This group's density suggests widespread redundancy, likely driven by overlapping PRN pain options embedded in clinical workflows. Other notable communities include Fluoroquinolone Antibiotics (blue), Insulin Analogs – Rapid Acting (purple), Aminopenicillin Antibiotics – Beta-lactamase Inhibitor Combinations (orange), and Analgesic or Antipyretic Non-Opioids (green).

These visualizations serve to illustrate different community profiles, from tightly knit opioid clusters to redundant combinations of antibiotics or non-opioid analgesics, where duplication may arise due to provider habit, legacy defaults, or therapeutic overlap. While Figure 9 provides a static overview of these high-risk groups, a fully interactive version will be linked in the appendix. The HTML version includes hover-based annotations such as therapeutic class, subclass, community ID, and associated order sets, allowing for deeper exploration of network structure and clinical context.

These communities demonstrate how network clustering can inform medication safety reviews. Interventions targeting entire subclasses or clusters may offer more scalable and effective solutions than reviewing duplicate medications one by one.

**Figure 9**

*Top Five Subclass-homogeneous Communities Identified in the Duplicate Therapy Network*



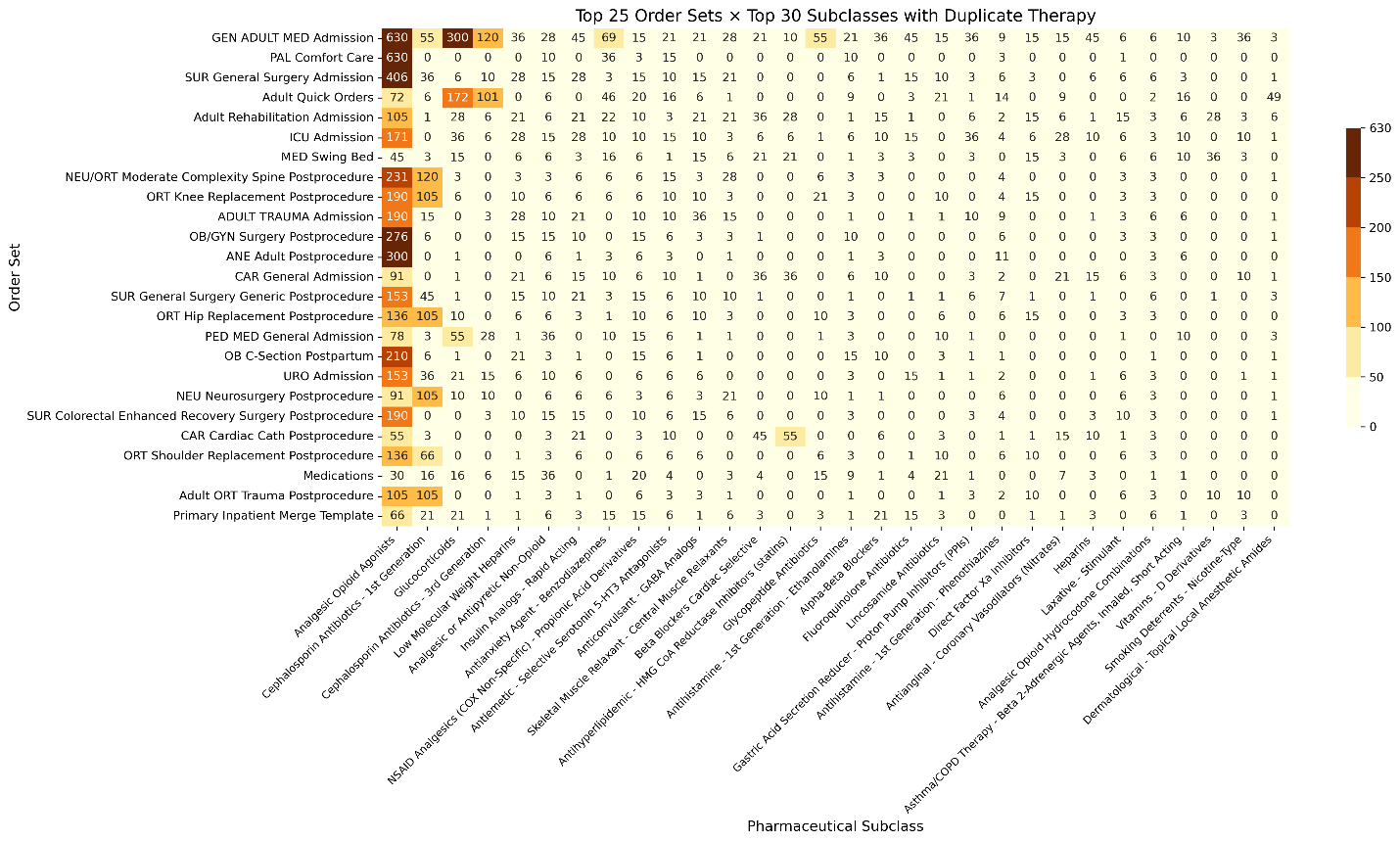
*Note.* Each cluster reflects a group of medications with high internal duplication density and shared pharmacological subclass.

**Duplicate Therapy Concentration Across Order Sets and Subclasses**

To synthesize the findings from centrality analysis, community structure, and subclass duplication patterns, Figure 10 presents a heatmap of the top 25 order sets against the top 30 pharmaceutical subclasses involved in duplicate therapies. Each cell represents the number of duplicate pairs within that order set and subclass combination, with warmer colors indicating higher concentrations.

**Figure 10**

*Heatmap of Duplicate Therapy Concentrations Across the top 25 Order Sets and Top 30 Pharmaceutical Subclasses*



The most prominent hotspot is the GEN ADULT MED Admission order set, which contains over 600 duplicate opioid pairings and substantial overlap in other high-risk categories such as antibiotics and steroids. Similarly, PAL Comfort Care and various postprocedure surgical sets show concentrated duplication in pain management subclasses, reinforcing the clinical relevance of earlier community findings.

By presenting the data in this format, the visualization enables both broad and granular comparisons highlighting not just which order sets are problematic, but also which subclasses drive the risk within each. This dual view supports prioritization efforts, whether targeting specific therapeutic categories (such as opioid agonists or rapid-acting insulin) or particular high-risk order sets.

**Conclusion**

As a closing note to this chapter, the heatmap illustrates how structural network patterns manifest in real clinical workflows. These insights will form the basis of Chapter 4, which will explore strategic opportunities for reducing duplication risk, optimizing order sets, and improving medication safety at scale.

**Chapter 4: Discussion of Analysis**

**Introduction**

Having applied network science methods to identify patterns of duplicate therapy risk within order sets, this chapter interprets those findings in the context of the original business problem: improving medication safety and reducing redundancy in standardized clinical workflows. The analysis presented in Chapter 3 yielded a rich set of visual and quantitative outputs that not only demonstrated widespread co-occurrence of potentially duplicative medications but also surfaced the specific therapeutic areas and order sets where such risks are most concentrated.

This chapter bridges the gap between analytic results and operational implications by examining how the structure and dynamics of the duplicate therapy network align with institutional priorities for safe prescribing, pharmacy review, and order set governance. Key findings from centrality analysis, subclass-specific patterns, and community clustering are revisited through the lens of clinical interpretability and quality improvement. In doing so, the discussion surfaces actionable insights for health system leadership, identifies limitations, and offers considerations for scaling the methodology more broadly.

**Results and Interpretation**

The degree centrality analysis identified medications such as morphine, oxycodone, and hydromorphone as highly connected nodes within the duplicate therapy network. These agents frequently, often embedded in multiple distinct order sets. Their centrality within the network reflects both the clinical ubiquity of opioids and the structural risk posed by redundant prescribing pathways. In many cases, these agents were ordered as part of multimodal pain regimens that did not explicitly distinguish between acute and chronic use or between standing and as-needed orders. The high connectivity of these nodes suggests that interventions to improve opioid stewardship should not be isolated to formulary restriction but should also target order set logic and decision support pathways that currently allow for unintentional duplication. The findings reinforce what was already expected within the organization: opioids present an outsized risk for duplication due to their near-ubiquitous presence across order sets.

Beyond opioids, the analysis highlighted the role of medications that, while less duplicated in isolation, serve as critical bridges across therapeutic domains. Betweenness centrality calculations revealed agents such as ropivacaine, lidocaine, and heparin as frequent intermediaries in the shortest paths connecting other medications in the network. Their positioning suggests that these agents participate in multiple care contexts, linking procedural, inpatient, and surgical workflows. Although they may not present direct duplication risk in every instance, their integrative role amplifies the downstream effect of their misuse. These results highlight the value of reviewing how these agents are incorporated across disparate order sets and whether their usage is governed by consistent clinical logic.

Community detection further deepened these insights. The Louvain clustering algorithm revealed a network composed of tightly interconnected communities, with a modularity score of 0.9035 indicative of an exceptionally well-partitioned system. These communities mapped closely to known therapeutic groupings, such as analgesics, antimicrobials, and antidiabetic medications. In practical terms, the emergence of structurally distinct clusters reinforces the idea that duplication risk is not isolated but embedded within how order sets are constructed and how clinical teams organize care. For instance, the presence of multiple opioid agonists within a single community implies that clinicians may be inadvertently co-prescribing multiple agents from the same class without clear differentiation of purpose, indication, or route.

Of particular interest was the community composed primarily of fluoroquinolone antibiotics. This class is frequently targeted by antimicrobial stewardship programs due to its association with resistance, C. difficile infection, and other adverse effects. The identification of this class as a coherent, densely connected community within the duplicate therapy network suggests that stewardship efforts could benefit from focusing not only on individual prescriptions but on the order set configurations that normalize fluoroquinolone co-prescription. Similar observations were made for insulin analogs and beta-lactamase inhibitor combinations, classes that require precise titration and indication-specific dosing, making potential duplication particularly dangerous.

The heatmap illustrating co-occurrence between top pharmaceutical subclasses and high-frequency order sets provided an operational lens through which to interpret the network structure. The concentration of duplication activity in a small number of order sets, such as GEN ADULT MED Admission and PAL Comfort Care, was striking. These sets function as foundational components of hospital workflows and are often initiated at the point of care transition. As such, they are prone to include a wide range of commonly used medications many of which overlap with other therapeutic orders initiated later in the same episode of care. The result is a structural vulnerability to duplication. This observation highlights a compelling opportunity for targeted intervention: if even a handful of foundational order sets can be revised to reduce duplication in key classes, the overall system burden may decline substantially.

These patterns emphasize that duplication is not a sporadic or random occurrence. Rather, it is a predictable byproduct of system design and an emergent property of how medications are grouped, presented, and selected within the electronic health record. Subclass-level analysis proved especially useful in surfacing these patterns, offering a level of granularity that neither medication-level nor therapeutic class-level analyses could achieve. Where therapeutic classes were too broad to meaningfully interpret, pharmaceutical subclasses revealed clinically coherent clusters of risk.

**Limitations and Stakeholder Considerations**

While the analysis was limited to a single health system, that system spans approximately 30 hospitals across three states, encompassing a broad range of prescribers and clinical workflows. The inclusion of diverse geographic and institutional contexts improves generalizability, though the study remains constrained by its focus on one system's formulary, governance structures, and electronic health record build. Furthermore, this analysis was scoped to order set–based prescribing only. Orders placed independently or via other documentation workflows were not included, as order sets are the area where the division of pharmacy can most directly influence change.

Insights from this analysis are intended to inform downstream review processes. Preliminary results may be presented to the medication safety team for prioritization, after which proposals for specific order set changes would route through clinical governance committees. Those committees, composed of physicians, pharmacists, and operational leaders, would ultimately decide whether and how to adjust prescribing tools or workflows. While the data provides a compelling case for action, the actual implementation of changes falls outside the scope of this project.

**Systemic Implications and Scalability**

This methodology could be replicated in any health system that has access to prescribing data at the medication and order set level. Since the implementation of the Affordable Care Act and the resulting acceleration of electronic medical record adoption, most health systems now possess the infrastructure to conduct a similar analysis. The only prerequisite is the ability to query co-occurrence patterns between medications within structured workflows. Because the data model and graph-based logic are portable, the approach demonstrated here provides a framework for scalable risk identification across institutions.

**Conclusion**

The network analysis successfully transformed a complex prescribing dataset into a set of interpretable, clinically relevant insights. Centrality measures identified medications of concern, not merely because of their frequency of use, but due to their structural importance in facilitating duplication. Community detection revealed patterns of co-prescription that are deeply embedded in the logic of clinical care delivery. The subclass-order set heatmap provided a crosswalk between abstract graph theory and actionable institutional priorities.

Taken together, these results offer a foundation for redesigning order sets in a way that systematically reduces duplication. Rather than relying solely on reactive pharmacist intervention or provider education, health systems can use network-derived insights to strategically restructure the decision-making environment itself. This aligns with broader goals of clinical decision support, patient safety, and standardization of care.

Chapter 5 will serve as the final component of this report, consolidating the outcomes of the analysis and mapping them to the original project objectives. It will also outline specific recommendations for integrating these findings into pharmacy workflows, order set governance processes, and ongoing quality improvement initiatives. By translating insights into tangible next steps, Chapter 5 will help establish a framework for scalable, system-wide enhancements to medication safety.

**Chapter 5: Summary and Recommendations**

**Introduction**

This final chapter provides a comprehensive summary of the capstone project, discusses the broader implications of the findings for clinical operations and patient safety, addresses the limitations encountered during the project, and outlines potential directions for future work. The goal is to synthesize the application of social network analysis techniques within the context of healthcare operations and highlight the significance of the insights generated through a refined investigation of duplicate therapy risks.

**Summary of Project**

The primary objective of this project was to identify opportunities for reducing duplicate therapy risks embedded within standardized inpatient order sets through the application of social network analysis. The project began by constructing a broad medication co-occurrence network using EMR data from October 2024. In this initial network, each node represented a unique medication, and edges indicated that two medications appeared together within the same prebuilt order set. This design allowed for a systemic visualization of medication pairing patterns across the institution’s order set architecture, yielding a dense network consisting of 2,909 nodes and 584,419 edges.

While the co-occurrence network provided a valuable overview of general medication relationships, it did not specifically distinguish between appropriate clinical combinations and those that might represent therapeutic redundancy. To address this, a second, more refined network, the Duplicate Therapy Network, was constructed. This network was filtered to retain only medication pairs that shared the same pharmaceutical subclass, focusing the analysis on co-occurrences that posed a meaningful risk of duplicate therapy. The resulting duplicate therapy network consisted of 2,402 nodes and 8,960 edges, reflecting a substantial reduction in scale while increasing the clinical relevance of the connections under study.

Analyses were conducted primarily on this refined network. Centrality measures, including degree and betweenness centrality, were applied to identify medications most frequently implicated in duplicate therapy risks and those acting as bridges between therapeutic areas. Community detection using the Louvain algorithm produced a modularity score of 0.9035, indicating that the network contained highly distinct, internally cohesive clusters. These quantitative findings reinforced the structural validity of the network model and provided a strong foundation for generating clinically actionable insights. High-centrality medications, densely clustered therapeutic groups, and concentrated duplication patterns across key order sets were all successfully identified, meeting the project’s primary objectives and supporting data-driven opportunities for clinical improvement.

**Implications of Project Results**

The insights derived from this analysis have significant implications for patient safety, operational efficiency, and clinical governance. By moving beyond isolated medication reviews to analyze systemic patterns of therapeutic duplication, the project enables the Patient Safety and Quality team to prioritize their efforts based on objective network-derived risk signals. High-centrality medications within the duplicate therapy network represent logical starting points for review, as interventions targeting these hubs have the potential to influence a wide array of order sets and clinical pathways.

The identification of densely clustered groups of medications with overlapping indications highlights specific therapeutic areas where duplicate therapy risks are particularly pronounced. Addressing these patterns could help to reduce unnecessary redundancies in treatment, minimize adverse event risks, and enhance clinical decision support system performance by reducing alert fatigue. The use of a network-based approach also offers scalability for future initiatives, as the methodology is compatible with routine EMR data structures and can be adapted to evolving clinical content without major reengineering.

The project's emphasis on pharmaceutical subclass as the primary categorization method enhanced the clinical interpretability of the results. This level of granularity allowed for a focus on duplication risks that would likely be clinically relevant, rather than being obscured by overly broad therapeutic groupings. The approach provides a framework for integrating data science techniques into routine clinical quality assurance activities, ultimately supporting safer, more consistent care delivery.

**Limitations and Next Steps**

Several important limitations were identified during this project. First, the data were drawn exclusively from a single large healthcare system spanning approximately 30 hospitals across three states. While this provided a diverse dataset, results may not generalize to other systems with different clinical practices, formulary structures, or order set designs.

Second, the project focused on structured order sets and did not account for medications prescribed through free-text or ad hoc ordering pathways. Duplication risks occurring outside the standardized order set environment were not captured in this analysis. Also, the filtering process that prioritized pharmaceutical subclass pairings, while enhancing specificity, inherently excluded certain duplications that might occur between related but differently classified medications.

Another limitation stems from the static nature of the data used. The project evaluated medication co-occurrence based on order set design as of October 2024, and did not assess dynamic patient-level usage or real-time clinical decision-making. As a result, the findings reflect potential duplication risks embedded in the system's design rather than realized patient exposures.

Building upon this foundation, several opportunities exist for future work. Expanding the analysis to incorporate patient-level medication administration data would enable a more direct evaluation of realized duplicate therapy events. Integrating medication indication metadata, where available, could allow the differentiation of clinically justified therapeutic overlaps from inappropriate redundancies. Additionally, extending the network model into a bipartite or multiplex structure, incorporating patients, medications, and order sets simultaneously, could provide a richer understanding of the interactions driving duplicate therapy risks. Operationalizing the duplicate therapy network as part of a continuous patient safety monitoring strategy could allow health systems to track emerging risks over time and proactively intervene as clinical practice evolves.

**Conclusion**

This project demonstrated the effective application of social network analysis to the identification of duplicate therapy risks within inpatient medical order sets. By constructing and refining a medication co-occurrence network into a clinically focused duplicate therapy network, the analysis revealed systemic patterns of therapeutic redundancy that may otherwise have remained hidden within the complexity of standardized order sets.

The emphasis on pharmaceutical subclass as the primary analytical framework significantly enhanced the specificity and clinical relevance of the findings. High-centrality medications and densely clustered therapeutic areas identified through network analysis provide practical targets for review and optimization efforts. The exceptionally high modularity score achieved through community detection validated the network structure and confirmed the presence of distinct clinical groupings susceptible to duplication risk.

While the project had inherent limitations regarding scope and generalizability, it successfully established a scalable, data-driven methodology that can be adapted and expanded for future quality improvement initiatives. The synthesis of technical data science approaches with clinically actionable outputs reflects the interdisciplinary nature of modern healthcare analytics. This work contributes to a growing body of evidence supporting the use of advanced analytical techniques in the pursuit of safer, more efficient, and more effective healthcare delivery.

**References**

Askar, M., Cañadas, R. N., & Svendsen, K. (2021). An introduction to network analysis for studies of medication use. *Research in Social and Administrative Pharmacy, 17*(12), 2054–2061. <https://doi.org/10.1016/j.sapharm.2021.06.021>

Bazzoni, G., Marengoni, A., Tettamanti, M., et al. (2015). The drug prescription network: A system-level view of drug co-prescription in community-dwelling elderly people. *Rejuvenation Research, 18*(2), 153–161. <https://doi.org/10.1089/rej.2014.1628>

Bocknek, L. S., et al. (2022). Duplicate medication order errors: Safety gaps and recommendations for improvement. *Patient Safety Network (PSNet)*. https://psnet.ahrq.gov/issue/duplicate-medication-order-errors-safety-gaps-and-recommendations-improvement

Brunson, J. C., & Laubenbacher, R. C. (2018). Applications of network analysis to routinely collected health care data: A systematic review. *Journal of the American Medical Informatics Association, 25*(3), 210–221. https://doi.org/10.1093/jamia/ocx052

Cavallo, P., Pagano, S., Boccia, G., et al. (2013). Network analysis of drug prescriptions. *Pharmacoepidemiology and Drug Safety, 22*(2), 130–137. <https://doi.org/10.1002/pds.3384>

Dugué, N., & Perez, A. (2022). Direction matters in complex networks: A theoretical and applied study for greedy modularity optimization. *Physica A: Statistical Mechanics and Its Applications*, *603*, 127798. https://doi.org/10.1016/j.physa.2022.127798

Evans, T. S., & Chen, B. (2022). Linking the network centrality measures closeness and degree. *Communications Physics*, *5*(1). https://doi.org/10.1038/s42005-022-00949-5

Francis, T., et al. (2024). Exploring the use of social network analysis methods in process improvement within healthcare organizations: A scoping review. *BMC Health Services Research, 24*, Article 175. <https://doi.org/10.1186/s12913-024-11475-1>

Locke, S. C., Le, A. D., Dunn, R. L., & Cho, J. C. (2019). Preventing pain medication duplications: A pharmacy-led intervention program. *Pain Medicine, 20*(6), 1255–1256. https://doi.org/10.1093/pm/pny321

Niyirora, J., & Aragones, O. (2019). Network analysis of medical care services. *Health Informatics Journal, 25*(4), 1538–1550. <https://doi.org/10.1177/1460458219887047>

Purdue Online Writing Lab. (n.d.). *General Writing FAQs.* Purdue Online Writing Lab. <https://owl.purdue.edu/owl/general_writing/general_writing_faqs.html>

Saatchi, A. G., et al. (2023). Network approaches and interventions in healthcare settings: A systematic scoping review. *PLOS ONE, 18*(2), e0282050. <https://doi.org/10.1371/journal.pone.0282050>

Smit, L. C., et al. (2020). Value of social network analysis for developing and evaluating complex healthcare interventions: A scoping review. *BMJ Open, 10*(11), e039681. <https://doi.org/10.1136/bmjopen-2020-039681>

Wetterneck, T. B., et al. (2011). Factors contributing to an increase in duplicate medication order errors after CPOE implementation. *Journal of the American Medical Informatics Association, 18*(6), 774–782. <https://doi.org/10.1136/amiajnl-2011-000255>

Xiang, N., Wang, Q., & You, M. (2023). Estimation and update of betweenness centrality with progressive algorithm and shortest paths approximation. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-44392-0

Appendix A: Data Dictionary

This appendix contains a breakdown of the files used in the project including column name, data type, definition, and example data.

**Patient\_Orderset\_c\_Start\_End.csv**

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Definition | Example Value |
| MaskedPatientID | String | De-identified patient identifier linking patient encounters to order sets. | P123456789 |
| StartDateKey | Integer | Start date of order set initiation, formatted as YYYYMMDD. | 20241015 |
| StartTimeOfDayKey | Integer | Start time of order set initiation, in HHMMSS 24-hour format. | 084500 |
| EndDateKey | Integer | End date when the order set ended or was discontinued, formatted as YYYYMMDD. | 20241015 |
| EndTimeOfDayKey | Integer | End time when the order set ended or was discontinued, in HHMMSS 24-hour format. | 123000 |
| OrdersetKey | Integer | Unique identifier for a specific order set. | 100245 |

**Medication-Orderset pairs.csv**

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Definition | Example Value |
| MedicationKey | Integer | Unique identifier for a medication. | 50123 |
| OrdersetKey | Integer | Unique identifier for the order set that includes the medication. | 100245 |

**Medications\_Detail.csv**

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Definition | Example Value |
| MedicationKey | Integer | Unique identifier for a medication. | 50123 |
| SimpleGenericName | String | Simplified generic name of the medication. | Acetaminophen 500 mg Tablet |
| TherapeuticClass | String | Broad therapeutic class describing the medication's primary clinical use. | Analgesics |
| PharmaceuticalSubclass | String | Specific subclass within the therapeutic class for more detailed classification. | Non-Opioid Analgesics |

**Ordersets.csv**

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Definition | Example Value |
| OrdersetKey | Integer | Unique identifier for a standardized medical order set. | 100245 |
| OrderSetName | String | Name of the order set used in clinical documentation and workflows. | Adult Pain Management Order Set |

Appendix B: Code

This project was completed using Jupyter Notebook. The code and interactive html files are available in a GitHub repository.

<https://github.com/jka99/Capstone/blob/main/Capstone.ipynb>

<https://github.com/jka99/Capstone/blob/main/interactive_all_communities.html>

<https://github.com/jka99/Capstone/blob/main/interactive_top_communities.html>