A Study of Medical Drugs and Side Effects: A Network Analysis

Sami Jaffri[1] [44165611] , Ryan Eveson[1] [99775389] , Evan Pasenau[1] [36403509]

[1] The University of British Columbia, Kelowna BC, V1V-1V8, CAN

Abstract. In today's world, it's difficult to find relevant and easy to understand information about medical drugs, what they do, and how they work in order to treat various medical conditions. Patients are often found unaware of how to or what to do with specific medicine which causes the full benefits of the drug to not be realized. According to The National Library of Medicine [1], approximately 50% of all patients do not take their medicine as prescribed, have suboptimal health literacy, and lack involvement in the treatment decision-making process, which all imply poor medication adherence; this does not have to be the case. In our study, we applied network science to understand the flaws in our healthcare system and develop a recommendation tool that provides relevant, and quick information about medical drugs with the aim of bridging the health literacy gap. During our research, we analyzed a dataset of over 3,500 medical drugs to identify their associated side effects. We determined the best medications for a total of 49 medical conditions, discovered the most common side effects linked to drug use, and developed a system that provides analysis to accurately rate medical drugs based on both the quality and quantity of user reviews.

1. Introduction

The newfound intersection between data science and medicine has opened many new paths for analyzing the complex relationships between diseases, treatments, and side effects. Traditional methods of discovering drugs often involved isolated approaches that overlooked critical interdependencies between disease, drugs, and medicine. The use of network science offers a new perspective that addresses these challenges head on. By representing disease, drugs, and side effects as connected nodes and edges in a graph it enables a deeper understanding of the underlying structure of medical data and research.

In this project, our group utilized said network science to construct a graph based representation of medicinal drugs and their connections to various diseases. Each edge in the graph demonstrates a relationship such as a drug used to treat a disease or a side effect associated with that treatment. This approach doesn't only capture direct relationships but also illustrates hidden patterns such as clusters of disease that respond to similar medications sharing overlapping side effects. This insight can guide doctors or even the average individual in tailoring their treatment plans and assist researchers in identifying potential drug repurposing.

Understanding the relationship between side effects and disease is equally as important. Side effects often influence how effective a treatment can be and how willing a patient is to adhere to treatment. By integrating side effect data into our research, we aim to create a tool that can be used to evaluate and streamline the treatment decision-making process. This shows individuals

how to balance efficacy with safety. In other words, our research emphasizes practical applications of network science which attempts to bridge the societal gaps in medical knowledge. By visualizing relationships and providing recommendations we demonstrate how data can contribute to advancing the field.

In the medical field, there are lots of studies showing the effects of medications and how they treat diseases as well as data on how medications work when used together but the task of optimizing the combination of treatments for each patient is left to the healthcare professionals experience. Therefore there is a gap that can be filled with a comprehensive approach backed by data and AI that makes sure there is not another superior treatment option.

Our primary objective is that by analyzing the relationships between diseases, treatments, and side effects on a macro scale, we can create a tool to guide healthcare professionals to which drug is best suited to treat their patient while not conflicting with other medications and limiting side effects that would most negatively affect the patient's life. This helps achieve our goal of allowing patients to choose what side effects they can tolerate while minimizing intolerable side effects in their treatment plan.

2. Related Work

The application of network science in medicine isn't a new topic. In fact, throughout recent years it's been a rapidly evolving field that attempts to showcase the diverse interactions between disease, drugs, and biological pathways. This scientific approach recognizes that diseases aren't usually caused by abnormalities but rather they come from stoppages in cellular and molecular networks. By mapping and illustrating these interactions, individuals, more specifically researchers, can identify diseases, explore relationships, and discover new treatment targets.

A few key studies have been conducted on a similar concept called the diseasome which is a way of representing a network where the nodes are diseases and the edges are shared molecular structures. According to Barabasi, Gulbahce, and Loscalzo in "*Network medicine: a network-based approach to human disease*" diseaseome maps "represent[ing] various molecular relationships between the disease-associated cellular components shows that diseases linked at the molecular level often 'exhibit detectable comorbidity'" [2]. In other words, this framework has assisted in highlighting how diseases with shared genetic bases can show clinical coexistence.

In another related study, researchers used the Jaccard coefficient to measure the similarities between diseases. They discovered that certain drugs could potentially be effective for treating diseases linked by shared biological pathways [3]. This study also revealed that "protein interactions within disease networks could identify new therapeutic targets" [4] for unrelated conditions which supports the idea of repurposing existing drugs for different diseases.

Finally, Galeano, D., Li, S., Gerstein, M., & Paccanaro, A. performed a study that attempted to predict the frequency of side effects with specific drugs and how common they are in specific groups and types of people. They performed this study as "frequencies are experimentally determined in randomised controlled clinical trials" rather than predicting frequencies with data.

This study similar to ours, as it brought insight to side effects and medication, and attempted to inform readers of the health literacy gap and how side effects can dampen the benefits of taking medical drugs

3. Methodology

In this study, we analyzed over 3,500 medical drugs to assess their effectiveness and find the most common side effects related to these drugs. Our goal was to develop a system that accurately rated medications based on past users' reviews and ratings. Before our analysis began, we reviewed other studies to see which network indices would be the most beneficial for our study. Accordingly, we concluded that these following network science approaches: 1- bipartite graphs, 2- Degree centrality index (primarily in degree), 3- Mean and Standard Deviation calculations, 4- correlation calculations. The bipartite graphs are used to separate the medical conditions making visualization and analysis easier as it will create two disjoint sets of data. The degree centrality calculation deals with the total edges leaving and entering different edges creating a weight for each node which signifies how common each side effect node is. The mean and standard deviation calculations are used to compare differences between drugs, and side effects, and are used to calculate how effective each drug is. Correlation calculations identified the correlation between activity and ratings.

For our data collection, we took data from a combination of online datasets from reputable users on platforms like Kaggle. The datasets included the information we needed on medical conditions, drugs, and side effects. We then used data manipulation techniques such as outer, and inner, joins and other transformative methods to merge datasets. This allowed us to organize the most relevant data capturing the relationships suitable for our analysis. Our dataset initially contained a single column listing all side effects associated with each drug. To facilitate analysis, we separated these side effects into individual columns, with one column for each side effect (e.g., hives, swelling, etc.). This was achieved using string manipulation and text searching in R to identify specific side effects. We focused on the 17 most common side effects across all drugs in the dataset. This approach enhances the ease of data manipulation while preserving the dataset's validity and integrity.

$$Weighted\ Effectiveness\ Score\ = Rating\ of\ Treatment^2\ * log(1 + Number\ of\ Review)$$

The weighted effectiveness score is used to give each treatment a general score to give a metric to sort the treatments by which would be preferred for treatment. This formula makes sure that the balance of ratings will be more dominant while still considering review counts. This approach ensures that the highest-rated drugs naturally rank higher, which aligns with your goal of reducing the weight of review volume in the calculation. To get the ratings "users were asked how effective they found the medicine while considering positive/adverse effects and ease of use (1 = not effective, 10 = most effective)."[5] This also gives a rough estimate of the severity of the side effects. Activity represented as a percentage "is based on recent site visitor activity relative to other medications in the list."[5]

4. Results

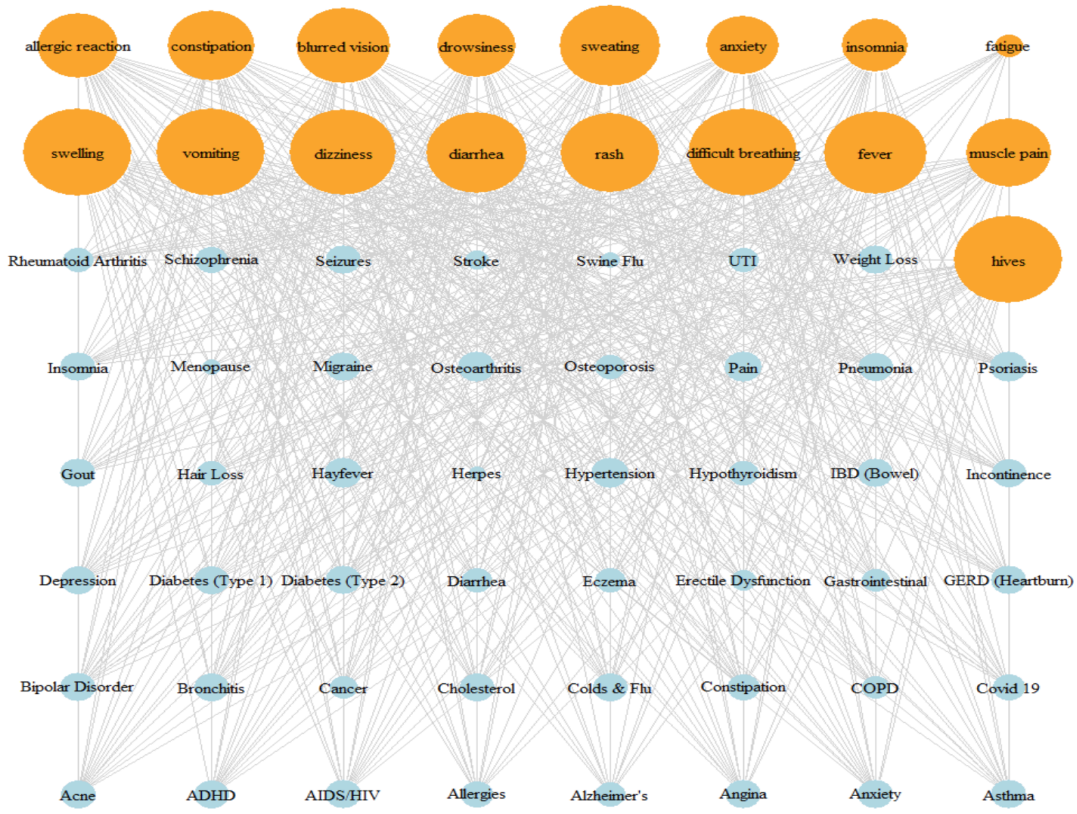

Fig 1. Bipartite graph of medical conditions and their links to side effects they cause where the size of the node represents the degree of the node.

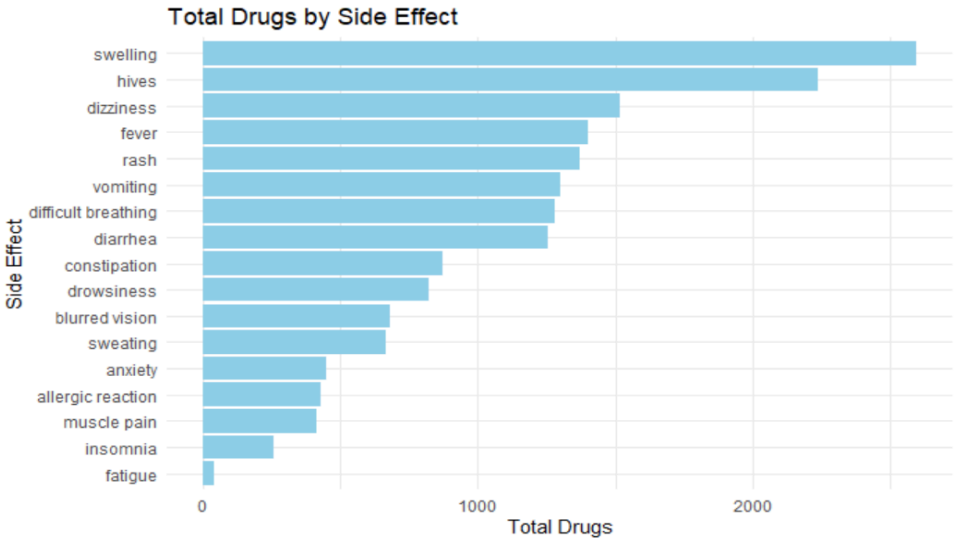

Fig 2. Showing the number of drugs that cause the given symptoms table
Mean = 1034.8, Standard deviation = 687.5

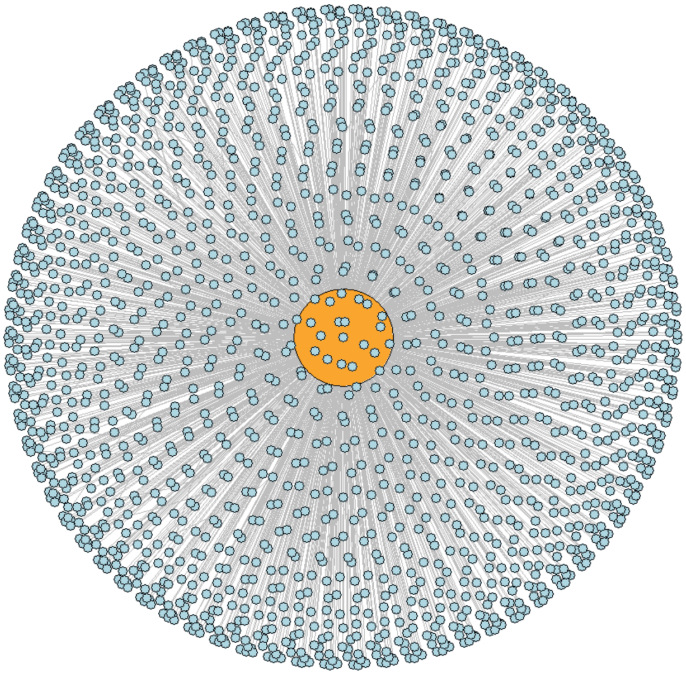

Fig 3.1. Bipartite graph showing the drugs that cause swelling

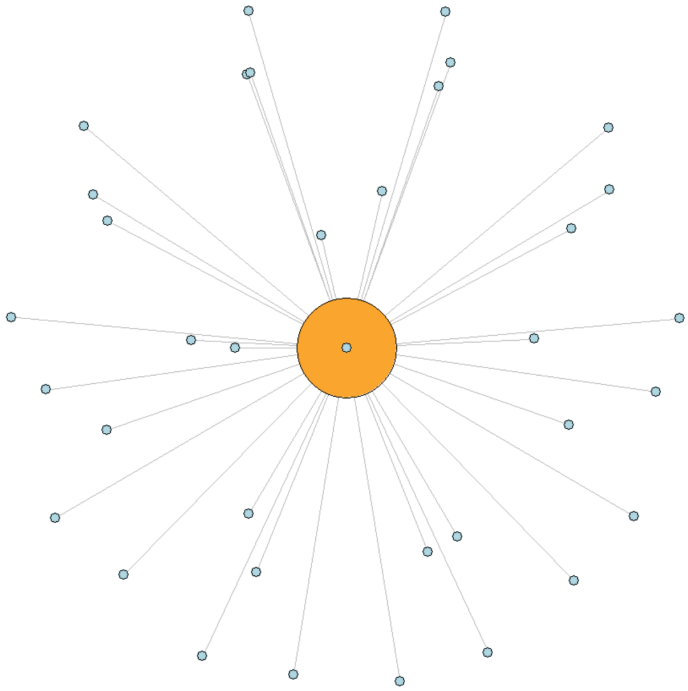

Fig 3.2. Bipartite graph showing the drugs that cause fatigue

medical_condition <chr>	drug_name <chr>	activity <dbl>	rating <dbl>	no_of_reviews <dbl>	weighted_score <dbl>
Weight Loss	phentermine	95	8.7	2934	604.3440
Anxiety	alprazolam	67	8.6	840	498.0904
Constipation	magnesium citrate	18	8.7	473	466.3418
Erectile Dysfunction	Cialis	99	8.6	448	451.6754
Acne	isotretinoin	26	8.0	999	442.0963
Pain	oxycodone	52	8.4	456	432.1577
Bipolar Disorder	Lamictal	96	8.1	600	419.8118
AIDS/HIV	Stribild	60	9.5	89	406.1078
AIDS/HIV	cobicistat / elvitegravir / emtricitabine / tenofovir	33	9.5	89	406.1078
Migraine	rizatriptan	35	8.3	314	396.2947

Fig 4. The top drug for each medical condition by weighted score (only top 10 values ordered by weighted_score descending)

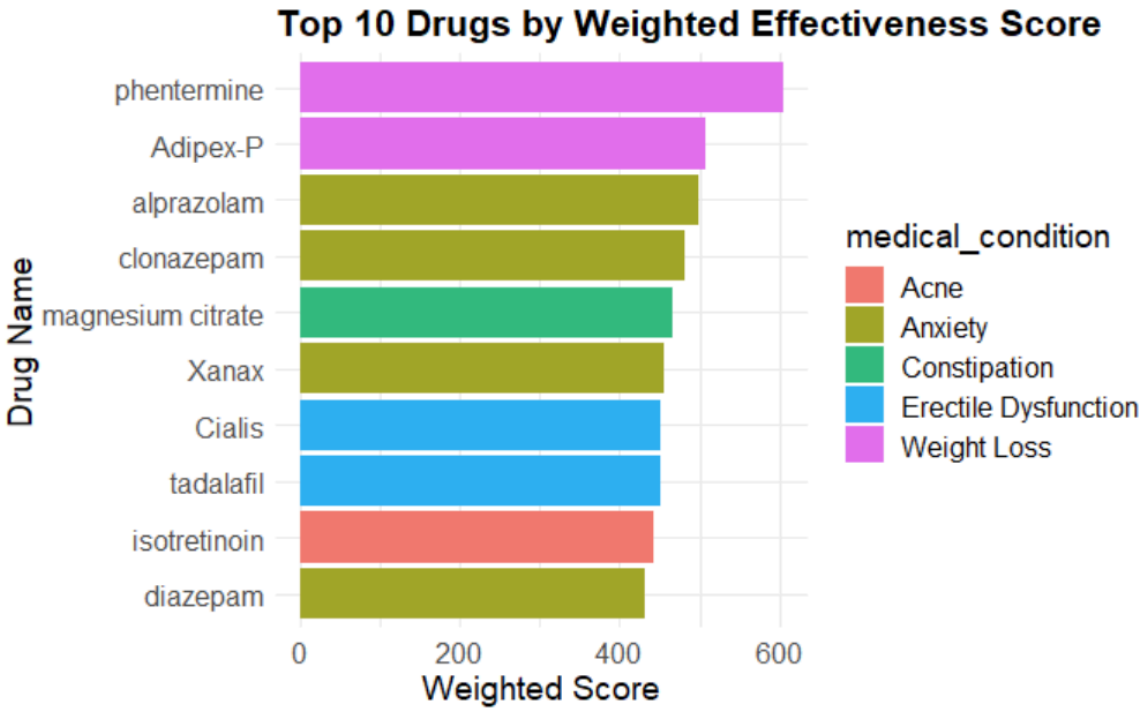

Fig 5. The top drugs by their weighted effectiveness score.
Weighted Effectiveness Score Statistics:
Mean = 280.02, Standard deviation = 111.31, Max = 604.34, Min = 88.99

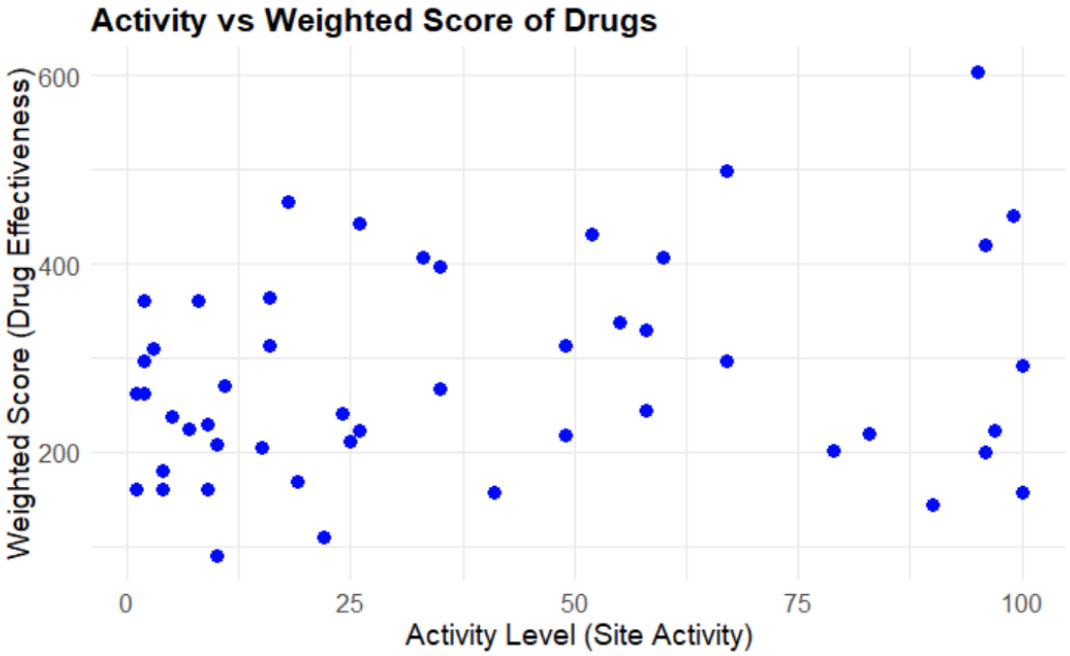

Fig 6. Weighted Effectiveness Score versus Activity Level. The correlation between the weighted effectiveness score and activity is = 0.2373644 = 23.7%

medical_condition <chr>	average_weighted_score <dbl>
Weight Loss	296.36981
Erectile Dysfunction	268.15871
Anxiety	252.03400
Migraine	210.92357
Depression	190.31510
ADHD	186.13696
Gout	179.80542
Bipolar Disorder	178.93701
Acne	152.51258
Pain	152.27199
Cancer	45.24193

Fig 7. Figure showing the average weighted score of all medical conditions (Skipped all drugs below pain to cancer)

5. Discussion

Our research is able to showcase how powerful network science is in medicine and the analysis of optimizing the relationships between drugs, medical conditions, and side effects. We were able to discover the most common and least common side effects associated with each drug and medical condition. The most common being 'swelling' with *2597* total occurrences with a close second being 'hives' with *2239* . Similarly to this, we were able to determine that fatigue is the most uncommon side effect with only *49* of our tested drugs having it [Fig 2]. Next we created a weighted ranking system to determine the quality of each drug and were able to find the best drug for each medical condition, and were able to rank each drug individually. The maximum of the weighted value is *604.34*, the minimum is *88.99*, the mean of these values is *280.02*, and the standard deviation is *111.31* [Fig 5]. From this weighted score, and the activity level we were able to determine that there is little correlation between the best calculated drug and the recent site activity. The correlation coefficient between these two values of '*0.2373644 = 23.7%*' proves that there is a low correlation between the 'best rated drugs' and site activity for that drug. The average weighted score [Fig 7] shows that weight loss has the best average for drugs, with a value of 296.37, and Cancer has the worst average weighted score with 45.24.

Additionally, the research we've done aligns with and attempts to build upon existing studies in networks science with relation to medicine such as the diseaseome framework aforementioned which was proposed by Barabasi et al. The disease framework highlighted the molecular relationships between disease and their genetic bases. By taking inspiration from this approach in our study of drug effectiveness and side effect analysis we contributed to the field by further showcasing how network-based approaches enhance treatment optimization. Furthermore, our research extends the applications of metrics such as but not limited to degree centrality and correlation which helps in understanding drug side effect relationships.

The results of this study underscore the transformative potential of network science in advancing medical treatment. By analyzing over 3,500 drugs and their side effects and symptoms, the research

identified the most effective medications for 49 medical conditions and developed a weighted effectiveness score that balances user reviews and side effect profiles. This unique approach highlights the ability to visualize and optimize the complex relationships between diseases, drugs, and their side effects, providing a framework for informed decision-making. The findings also revealed important new insights, such as the low correlation between drug effectiveness and recent usage activity, underscoring the need for evidence-based evaluations. This proves our hypothesis demonstrating that there is a prevalent health literacy gap in our society. While limitations such as a narrow focus on common side effects and potential review bias exist, this study establishes a strong foundation for leveraging network science to improve medication adherence, optimize treatments, and empower both patients and healthcare professionals in the decision-making process.

6. Conclusion

There is a lack of resources to provide patients with a set of treatments that optimize side effects while maximizing effectiveness. Our research will allow healthcare professionals to minimize the side effects of the patients choosing while still providing an effective treatment.

Our study analyzed well over 3500 different types of drugs, taking special notice of effectiveness, identity, and common side effects using different academic resources such as but not limited to Kaggle, and Oxford Research. We employed network science methods such as bipartite graphs for visualization, degree metrics for centrality, and analyzed statistics such as mean, median, and averages for verification. Additionally, we developed a weighted effectiveness score calculated as:

$$Weighted\ Effectiveness\ Score\ = Rating\ of\ Treatment^2\ * log(1 + Number\ of\ Review)$$

This score helps rank treatments based on effectiveness, ease of use, and side effects using previous studies, and rated user scores on an exponential scale.

The largest limitation is that we used a small number of side effects that were the most common after taking the drug. It would be necessary to expand the number of side effects included in order to get the most beneficial system. Additionally, the review data has the potential to be biased. Due to the fact that our entire project is based on the review data, if we have skewed reviews which lean one way compared to the other and aren't accurate then our recommendation system fails. A way of fixing this could be weighted scores to account for not only low review counts but also account for similarities in reviews. What this means is that if a drug has many reviews saying A as compared to a few reviews saying B then A would get a higher weight making it more relevant.

In saying this, the main takeaway of this research paper is that network science provides an incredibly powerful tool for visualizing, analyzing, and displaying the complex interconnectedness of data. By having a focus on user-centric scores like centrality scores we can put an emphasis on the data that really matters and aggressively bridge the gap between medicinal literacy and informed treatment decision-making.

References

1. Brown, M. T., & Bussell, J. K. Medication adherence: Who cares? *Mayo Clinic Proceedings*, *86*(4), 304–314. (2011) Medication Adherence: WHO Cares? - ScienceDirect
2. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56–68. (2011) Network medicine: a network-based approach to human disease | Nature Reviews Genetics
3. Grosdidier, S., Ferrer, A., Faner, R. et al. Network medicine analysis of COPD multimorbidities. Respir Res 15, 111 (2014). https://doi.org/10.1186/s12931-014-0111-4
4. Nogales, C., Mamdouh, Z. M., List, M., Kiel, C., Casas, A. L., & Schmidt, H. H. (2021). Network pharmacology: Curing causal mechanisms instead of treating symptoms. 1-15. https://www.researchgate.net/publication/356931168_Network_pharmacology_curing_causal_mechanisms_instead_of_treating_symptoms/fulltext/61b3d7f64b318a6970d1ca85/Network-pharmacology-curing-causal-mechanisms-instead-of-treating-symptoms.pdf
5. Varghese, J. (2022, December 10). *Drugs, side effects, and medical condition*. Kaggle. https://www.kaggle.com/datasets/jithinanievarghese/drugs-side-effects-and-medical-condition
6. Galeano, D., Li, S., Gerstein, M., & Paccanaro, A. (2020, September 11). *Predicting the frequencies of drug side effects*. Nature News. https://www.nature.com/articles/s41467-020-18305-y