A Network Approach to Polypharmacy: An Exploratory Analysis of Drug-to-Drug Side Effects Using Patient-Centered Data

Ludovica Benini, Digital Humanities, 0001120870
Niccolò Molinati, Digital Humanities, 0001140894
Pietro Tisci, Digital Humanities, 0001142978

04/04/2025

1 Introduction

The application of network analysis methods has brought significant advancements in a multitude of fields. Among the domains that have increasingly adopted network-based approaches are bioinformatics and pharmacology, where they are used, for example, to analyze drug interactions, identify adverse effects, and explore therapeutic relationships[1]. These techniques provide a systematic way to explore large-scale biomedical data, uncovering hidden patterns that might not be apparent through traditional analytical methods.

Our report is constructed around a dataset produced during research in this specific domain. The goal of the project was to identify interactions between two different drugs (or proteins) using graph convolutional networks[2]. This approach enabled an extensive mapping of polypharmacy side effects, creating a valuable resource for the treatment of complex medical situations that typically involve multiple drugs. Understanding these interactions is critical in precision medicine, where minimizing adverse effects while maximizing therapeutic efficacy is essential.

Building upon this dataset, our work focuses on analyzing drug combinations side effects through traditional network analysis and statistical methods. Additionally, we enriched our analysis by integrating the original dataset with external sources, allowing both for a broader exploration of drug interactions and their associated risks and for a deeper understanding of the practical implications of the combination of very common drugs.

[1] Marianna Milano and Mario Cannataro, "Network models in bioinformatics: modeling and analysis for complex diseases", *Briefings in bioinformatics* 24, no. 2 (Jan. 2023), https://doi.org/https://doi.org/10.1093/bib/bbad016.

[2] Marinka Zitnik, Monica Agrawal, and Jure Leskovec, "Modeling polypharmacy side effects with graph convolutional networks", *Bioinformatics* 34, no. 13 (July 2018): i457–i466, https://doi.org/https://doi.org/10.1093/bioinformatics/bty294, https://academic.oup.com/bioinformatics/article/34/13/i457/5045770.

2 Problem and Motivation

Dealing with a graph-based representation of drug interaction and related side effects, our main focus was exploring information about each specific drug that can be extracted from its relation to others. For this, we will make use of two concepts, reactivity and harmfulness of drugs, which are strictly related to the nature and quantity of connections. We considered a number of assumptions inferred from the meaning of said connections, starting from the idea that measures relative to positions of specific nodes in the common paths connecting other nodes (like betweenness centrality) hold little to no significance for our study. Ideally, whether a drug finds itself in common pathways between different nodes tells nothing about its properties, since edges signify relations (the side effect derived from substance interaction) limited to the nodes they connect - therefore, the concept of path itself loses meaning in our case study.

On the other hand, large scale measures will help us visualize network properties and characteristics much more we could do thanks to a simple graph visualization, due to the extension of our data. Our focus, however, shifted more on what we can know about nodes themselves. The general assumption is that nodes with a high rate of connections will be "more reactive" than the average one, meaning that the drug is more likely to generate side effects whenever taken alongside another substance. The likelihood to generate a side effect given another random drug, however, is a misleading measure, as doesn't really allow us to infer the "harmfulness" of that drug, since we don't have a clear insight on the quality of the side effect. For that reason, we compare that reactivity to the type of side effects, by weighting each connection and obtaining a general idea of the "harmfulness" related to that drug.

Once we have developed an intuitive understanding of our measures, we can begin formulating research questions. In section 5, we will focus on semantically interpreting mathematical measures of node properties within the network, dividing the process into three phases. First, we will conduct a general exploration of the graph (5.1). Next, we will concentrate on specific nodes identified through various metrics (5.2). Finally, we will apply these measures to the most prescribed drugs in the US, aiming to uncover relevant insights into the properties of these medications (5.3). Our study aspires to provide a semantic interpretation of both quantitative and qualitative data — sourced from academic research as well as crowd-sourced — that will help medical professionals better relate drug side effects to patient perspectives. At the same time, we would like to test with mathematical measures the harmfulness and reactivity of the most prescribed (according to a 2022 study) medicinal drugs in the US. Our hypothesis is that both of these values will be significantly lower than the median. This assumption stems from the idea that, in order to be widely available and prescribed, these drugs likely represent the least "dangerous" and reactive option among similar treatments targeted at specific pathologies.

3 Datasets

The final dataset[3] that we decided to explore was pulled by the authors of the original paper from the SIDER (Side Effect Resource)[4], the OFFSIDES[5], and TWOSIDES[6] databases, which contain polypharmacy side effect information. The resulting unified network contained 645 drug and 19,085 protein nodes, connected by 715,612 protein-protein, 4,651,131 drug-drug, and 18,596 drug-protein edges. From this massive multimodal network, we selected the graph encompassing only the nodes representing drugs and their edges. It is an undirected homogeneous multigraph (since a pair of drugs can lead to multiple side effects).

The multigraph, stored as a CSV file, was then enriched by us with the addition of data gathered from another study aimed at ranking adverse drug reactions through crowdsourcing. This step proved to be quite convoluted and requires a detailed explanation. Side effects in our network only partially matched those in the study assigning a rank to them (around 100 perfect matches). This was mainly due to the different naming conventions used in the two datasets. We therefore proceeded with four cycles of matching based on cosine similarities of the embeddings, using the all-MiniLM-L6-v2 model from the transformers[7] library. In each cycle, the threshold for accepting a cosine similarity score as a match was gradually lowered (0.85, 0.65, 0.55) and manually checked. After the third cycle, the threshold had to be reduced so much that it became unreliable. Fortunately, only 61 of the 1,319 side effects cited in the network remained unmatched, so we manually produced those matches.
This process allowed us to obtain a multigraph in which each edge is weighted based on the ranking, which we interpreted as a perceived fear score.
In addition, we collected the list of the top 300 prescribed drugs in the U.S., retrieved from the ClinCalc DrugStats Dataset website[8]. While this list was not used to directly build the graph, it served as an external reference for part of our analysis.

4 Validity and Reliability

We also wanted to assess whether our final graph could be considered representative of the day-to-day usage of medicines. To achieve this, we compared the drugs present as nodes in our network — unequivocally identified using the PubChem Compound Identifier (CID)[9] —to a list of the top 300 prescribed drugs in the United States in 2022[10], as said in the previous

[3]Marinka Zitnik, Monica Agrawal, and Jure Leskovec, *BioSNAP: Network datasets: Poypharmacy side-effect association network*, 2018, `https://snap.stanford.edu/biodata/datasets/10017/10017-ChChSe-Decagon.html`.

[4]Michael Kuhn et al., "The SIDER database of drugs and side effects", *Nucleic Acids Research* 44, no. D1 (Oct. 2015): D1075–D1079, `https://doi.org/https://doi.org/10.1093/nar/gkv1075`, `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702794/`.

[5]Tatonetti Lab, *databases for drug side effects and drug interactions*, 2023, `https://nsides.io/`.

[6]Tatonetti Lab, *databases for drug side effects and drug interactions*, 2023, `https://nsides.io/`.

[7]Thomas Wolf et al., "Transformers: State-of-the-Art Natural Language Processing", in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (Online: Association for Computational Linguistics, Oct. 2020), 38–45, `https://www.aclweb.org/anthology/2020.emnlp-demos.6`.

[8]Sean P Kane, *The Top 300 Drugs of 2022*, 2022, `https://clincalc.com/DrugStats/Top300Drugs.aspx`.

[9]2025, `https://pubchem.ncbi.nlm.nih.gov/`.

[10]Sean P Kane, *The Top 300 Drugs of 2022*, 2022, `https://clincalc.com/DrugStats/Top300Drugs.aspx`.

paragraph. We implemented this comparison using PubChem's API to convert the commercial names of drugs into their corresponding CIDs and the BeautifulSoup Python library to extract the 300 drugs into a CSV file. We then proceeded to match the drugs in our network with those in the ranked list, obtaining a total of 105 matches. Despite the relatively low overall match, 45 of these were among the top 100 ranked drugs, indicating that they represent a reasonable sample. Despite its size, our network represents only a limited number of drugs. As of 2022, the entire PubChem dataset included just over 290 million described substances, whereas our network contains only 645. This limitation stems from the lack of a comprehensive resource on polypharmacy side effects, as highlighted in the original paper[11].

However, as mentioned in the previous paragraph, our comparison with the top 300 prescribed drugs shows that a significant number of commonly used drugs are included in our dataset. This makes the network — and the resulting analysis — a representative sample of the most frequent drug combinations.

Our data manipulation process is fairly reproducible, as the tools used are publicly available and reasonably easy to use. The only potentially ambiguous step involves matching the names of side effects from the original paper with those in the perceived fear ranking. For this reason, we thoroughly detailed our approach in the previous paragraph.

5 Measures and Results

5.1 Overview of the network

Due to the significant dimension of the dataset, we think it would be appropriate to start the study by presenting the general properties of the network. Since visualization tools proved to be of little help in clearly showing the data, we relied on wide-ranging measures to perform the exploration.

Degree distribution

The degree of a node represents the total number of edges connected to it, including all parallel edges (1). The degree distribution— which captures the degree of all nodes in the graph—serves as a key indicator of the network's overall connectivity, revealing both the prevalence of highly connected nodes and their proportion relative to the total. Since we account for all edges connecting two drugs, it is unsurprising that the number of connections far exceeds the number for most nodes: with only 645 drugs, we observe extreme cases where up to 70k–80k edges interconnect them. As shown in Fig. 1, the distribution aligns more closely with a lognormal function rather than a power law. This is further confirmed by the estimated power-law exponent, $\alpha = 6.93$, which is significantly higher than typical values observed in scale-free networks, and by a log-likelihood ratio of -1.77 for the comparison between the power law and the lognormal model. The log-log plot of the same distribution (Fig. 2) further reinforces that the network does not conform to a power-law fit and, consequently, cannot be classified as scale-free.

$$deg(v) = \sum_{e \in E(v)} 1 \tag{1}$$

[11] Marinka Zitnik, Monica Agrawal, and Jure Leskovec, "Modeling polypharmacy side effects with graph convolutional networks", *Bioinformatics* 34, no. 13 (July 2018): i457–i466, `https://doi.org/https://doi.org/10.1093/bioinformatics/bty294`, `https://academic.oup.com/bioinformatics/article/34/13/i457/5045770`.

where $E(v)$ represents the set of edges connected to node v, including parallel edges.

K-core and layers of the network

K-core decomposition is an iterative algorithm that helps us dissect the network in order to find high-density areas in our data and understand their level of connectivity. A k-core is defined as a maximal subgraph in which every node has at least k connections (degree minor or equal to k) to other nodes within that subgraph. We found that the highest density point is a single k-core with a k-value of 184 and composed of 234 nodes. Gradually decreasing the k-value reveals that the aforementioned cluster is the core of the entire network, as the number of clusters stays the same, but the number of nodes in it increases (Table 1).

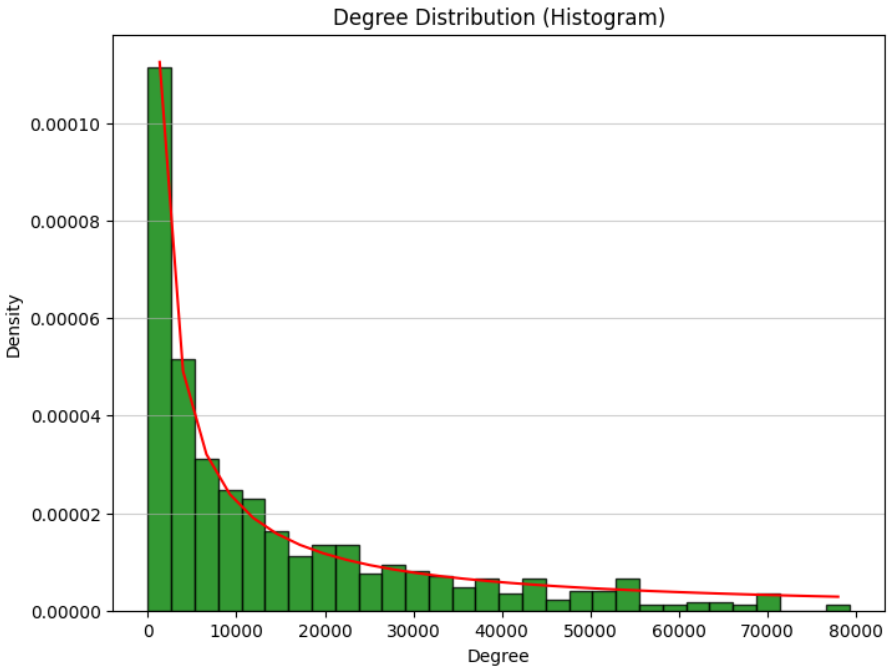

Figure 1: Degree distribution histogram (green) and lognormal function (red)

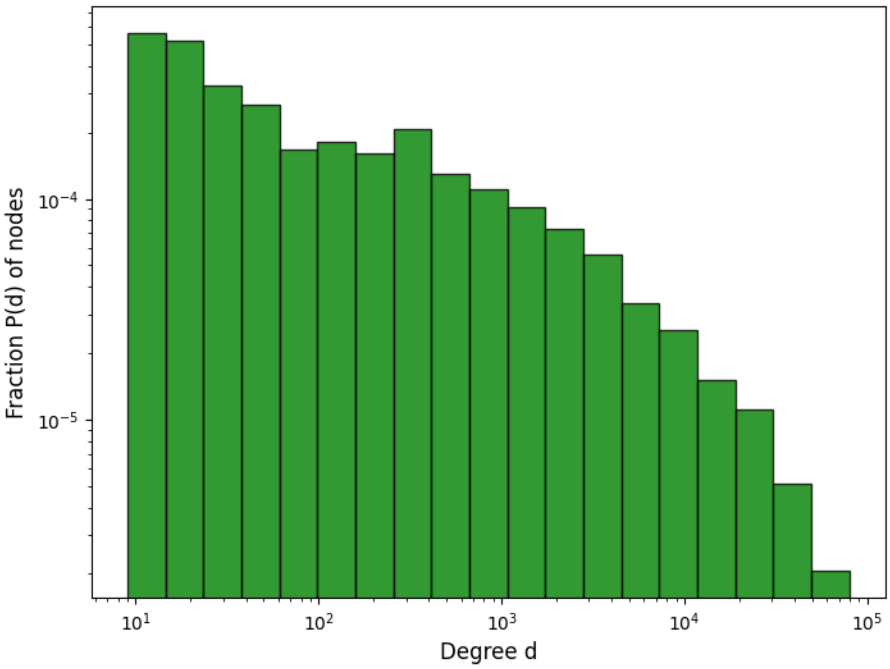

Figure 2: Degree distribution in log-log scale

K-value	Number of cores	Number of nodes
1	1	645
31	1	569
61	1	502
91	1	437
121	1	380
151	1	330
181	1	245
184	1	237

Table 1: Network dissection through k-core analysis

Local clustering coefficient distribution

The local clustering coefficient (LCC) measures how close a node's neighbors are to forming a complete clique. In our specific case, distribution of LCC values (Fig. 3) reveals interesting insights about the behavior of neighboring nodes relative to their degree, and therefore to the density of the network in various regions. While the study confirms that the network generally exhibits a decline in LCC as the degree increases, the rate of decay is lower than initially anticipated. Notably, a substantial number of low-degree nodes form complete cliques - with an LCC of 1.0 - whereas higher-degree nodes display LCC values around 0.4. This indicates that despite the overall network having a high connectivity, central nodes trade off local connection density for their role as "hubs".

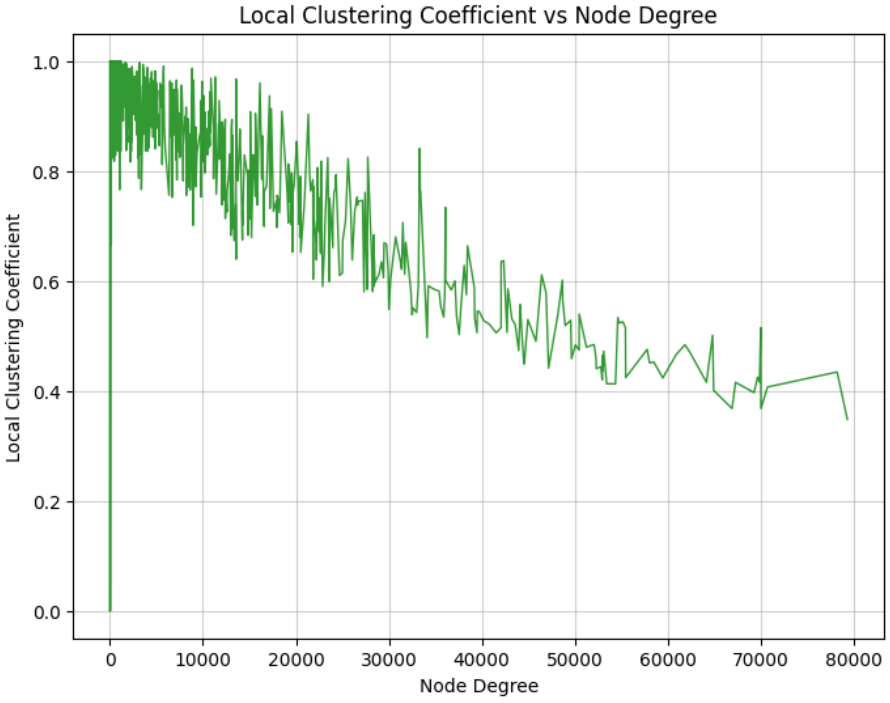

Figure 3: Local clustering coefficient distribution

5.2 Node Metrics

The following metrics are computed at the node level, as opposed to the entire network.

Simple Degree Centrality (Unique Reactivity)
We converted the multigraph into a simple graph by removing multiple edges between the same pair of nodes. This transformation allows us to compute degree centrality in its standard form, where each node's degree is divided by the total number of nodes in the graph minus one (2). In other words, each connection is counted only once, regardless of how many interactions exist between two drugs.
By applying this method, we can identify which drugs have the highest number of **unique** interactions with other drugs.

$$C_D(v) = \frac{\deg(v)}{N-1} \tag{2}$$

where $deg(v)$ represents the degree of the node v, excluding parallel edges and N is the total number of nodes in the graph.

CID	Drug Name	Simple Degree Centrality
CID000004594	Omeprazole	0.93
CID000001983	Acetaminophen	0.90
CID000000853	DL-Thyroxine	0.89
CID000003883	Lansoprazole	0.86
CID000005039	Ranitidine	0.85
CID000003958	Lorazepam	0.84
CID000054454	Simvastatin	0.83
CID000004679	Pantoprazole	0.83
CID000005732	Zolpidem	0.83
CID000002771	Citalopram	0.83

Table 2: Top 10 Drugs by Simple Degree Centrality

The results of the measure can be seen in Table 2. Omeprazole (0.93) has the highest simple degree centrality, indicating that it is directly connected to 93% of all other drugs in the network. Acetaminophen (0.90) and DL-Thyroxine (0.89) follow closely, suggesting that these drugs are also highly connected.
In terms of practical implications, these drugs may require special attention from healthcare specialists, particularly in polypharmacy scenarios (patients taking multiple drugs simultaneously), due to their broad interaction profiles.

Degree and Multi-edge Degree Centrality (Cumulative Reactivity)
We defined each drug's reactivity by measuring each node's degree (1), already defined and used above. This measure expresses each drug's reactivity because it indicates the number of drugs it interacts with and how many side effects each drug produces. Compared to the previous measure, it highlights both the number and frequency of drug interactions.
To obtain a more interpretable measure of reactivity, we applied degree centrality. However, since our network includes parallel edges, the standard degree centrality formula must be adjusted. Instead of normalizing by the total number of nodes minus one, we use the total sum of all edges in the network, excluding those connected to the current node, as the denominator. The denominator then is divided by two to avoid double counting the same edges. The nominator is the sum of the current node's edges (including the parallel ones) (3). This adjustment

ensures that the resulting relative score ranges from 0 to 1, making comparisons between drugs more meaningful and the results clearer.

$$C_D^{\text{multi}}(v) = \frac{\sum e_{vu}}{\frac{1}{2}\sum_{u \in V, u \neq v} \sum e_{ui}} \tag{3}$$

where e_{vu} represents an edge between the current node v and the node u, $V, u \neq v$ is the set of all the nodes excluding v and e_{ui} is the edge between the node u and the node i.

CID	Drug Name	Degree	Multi-edge Degree Centrality
CID000004594	Omeprazole	79335	0.0171
CID000004900	1,4-Pregnadiene-17alpha,21-diol-3,11,20-trione	78225	0.0168
CID000003958	Lorazepam	70740	0.0152
CID000000853	DL-Thyroxine	70030	0.0151
CID000005090	Rofecoxib	70015	0.0151
CID000005732	Zolpidem	69876	0.0150
CID000002083	Albuterol	69666	0.0150
CID000003883	Lansoprazole	69290	0.0149
CID000002771	Citalopram	67299	0.0145
CID000001983	Acetaminophen	66924	0.0144

Table 3: Top 10 Drugs by Degree and Multi-edge Degree Centrality

It is useful to compare the results in Table 2 with the results in Table 3. Most drugs appear in both tables (except for three), but their rankings differ completely — apart from the first one. This highlights the need to treat these measures separately (Table 4). The change in a drug's placement between the two tables highlights the difference between unique reactivity and cumulative reactivity

For instance, Acetaminophen ranks 2nd in Table 2 but drops to 10th in Table 3, meaning that while it interacts with many different drugs, the number of side effects per interaction is relatively low compared to higher-scoring drugs. Similarly, Lansoprazole ranks 4th in Table 2 but drops to 8th in Table 3.

Conversely, some drugs appear in Table 3 but are absent in Table 2, such as 1,4-Pregnadiene-17alpha,21-diol-3,11,20-trione (0.0168) and Rofecoxib (0.0151). This indicates that these drugs may not interact with as many unique drugs but are involved in frequent interactions with a smaller subset of drugs.

On the other hand, Lorazepam and Zolpidem retain high positions in both rankings, indicating that they not only interact with many drugs but also generate multiple side effects per interaction, making them particularly important in polypharmacy risk assessments.

Unique Reactivity	Cumulative Reactivity
Counts only unique connections	Counts all edges, including parallel ones
Emphasizes the number of distinct drugs a drug interacts with	Highlights both the number and frequency of drug interactions

Table 4: Reactivity measures comparison

Weighted Degree (Cumulative Reactivity Combined with Harmfulness)

Since each edge has a weight representing the perceived harmfulness of the associated side effect, we applied the weighted degree. This measure functions similarly to the standard degree, but instead of counting each connection as one, it multiplies each connection by its weight and then sums the results (5.2). The value of this measure lies in its ability to account for the severity of each side effect linked to a node, rather than just the total number of side effects. Thus, a drug with a higher weighted degree indicates both a greater number of side effects and a higher intensity of those effects. This measure accounts for both the number and the severity of side effects associated with a drug.

$$WD(v) = \sum_{u \in N(v)} w_{vu} \tag{4}$$

where $N(v)$ is the set of nodes connected to v and w_{vu} is the weight of the edge between node v and node u.

CID	Drug Name	Weighted Degree
CID000004594	Omeprazole	32807.67
CID000004900	1,4-Pregnadiene-17alpha,21-diol-3,11,20-trione	32716.54
CID000003958	Lorazepam	29996.60
CID000005732	Zolpidem	29128.42
CID000001983	Acetaminophen	29060.58
CID000000853	DL-Thyroxine	28999.38
CID000002083	Albuterol	28841.81
CID000005090	Rofecoxib	28798.69
CID000003883	Lansoprazole	28628.52
CID000002771	Citalopram	27862.70

Table 5: Top 10 Drugs by Weighted Degree (Rounded to Two Decimal Places)

Averaged Weighted Degree (Harmfulness)

This is a modified version of the weighted degree, where the total weighted degree is divided by the node's degree (5). This adjustment isolates the severity of side effects from the number of connections, as in the weighted degree, the number of connections tends to have a greater mathematical influence than severity. Thus, this measure helps identify the nodes associated with the most perilous side effects.

$$WD(v) = \sum_{u \in N(v)} w_{vu} \tag{5}$$

$WD(v)$ is the weighted degree of node v and $deg(v)$ is the degree (number of connections) of node v.

CID	Drug Name	Weighted Degree
CID000145068	Nitric Oxide	0.52
CID000004675	[17-acetyloxy-10,13-dimethyl-2,16-bis(1-methylpiperidin-1-ium-1-yl)-2,3,4,5,6,7,8,9,11,12,14,15,16,17-tetradecahydro-1H-cyclopenta[a]phenanthren-3-yl] acetate	0.51
CID000036339	Ethyl 1-(1-phenylethyl)-1H-imidazole-5-carboxylate	0.51
CID000036811	Dobutamine	0.49
CID000002232	1-[2-[(3-Methyl-1,2,3,4-tetrahydroquinoline-8-yl)sulfonylamino]-5-guanidinopentanoyl]-4-methylpiperidine-2-carboxylic acid	0.48
CID005282044	No name found	0.48
CID000005486	2-(Butylsulfonylamino)-3-[4-(4-piperidin-4-ylbutoxy)phenyl]propanoic acid	0.48
CID000041693	Sufentanil	0.48
CID000004993	Pyrimethamine	0.48
CID000003562	Halothane	0.48

Table 6: Top 10 Drugs by Averaged Weighted Degree (Rounded to Two Decimal Places)

When comparing Table 5 and Table 6, it is interesting to note that they do not share any drugs. This is because Table 6 removes the influence of the frequency of the interaction, focusing only on the (perceived) severity of side effects per interaction. As a result, the drugs listed in Table 6 are perceived as the most harmful but are not particularly reactive.
Another notable observation is the higher presence of compounds with unregistered names in Table 6, indicating that many of these drugs are experimental compounds or less commonly used medications.
Even if these drugs are not frequently involved in interactions, they still warrant careful monitoring due to their potential for severe adverse effects.

Boxplot of nodes' averaged weighted degree
We constructed a boxplot of the averaged weighted degrees to contextualize future results. The plot brought to our attention several outliers, both on the left and on the right side of the plot, which demonstrate a skewed distribution of values (Fig. 4).

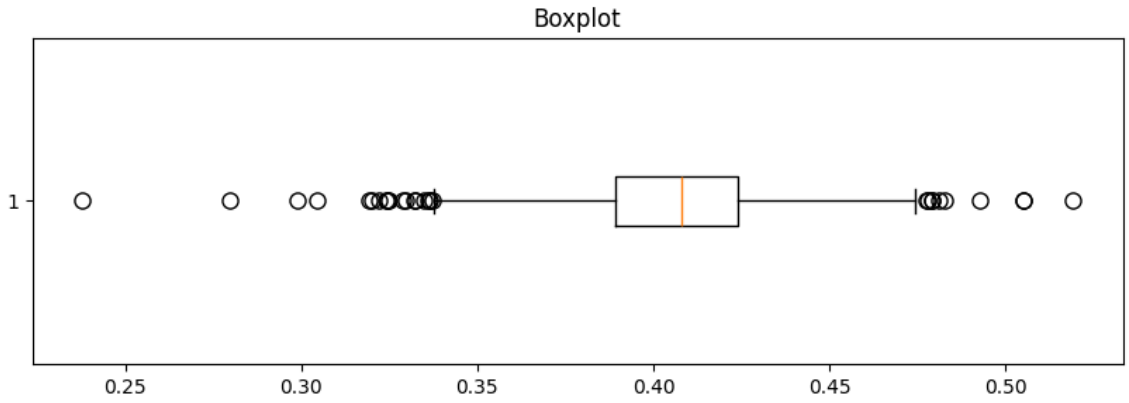

Figure 4: Boxplot on averaged weighted degree distribution with circles corresponding to outliers

Side effects perceived fear distribution in top 10 ranked drugs per averaged weighted degree

Since when we consider the averaged weighted degree of a specific node, we consider the average of perceived fear deriving from the side effects connecting to it, we are still left with unclear information about how specifically the side effects affect the general score. For example, an AWD of 0.30 may imply different scenarios: that the drug considered only causes a couple of side effects around that side effect perceived fear value; that it actually causes many side effects, all averaging on 0.30; or that the drug has a number of tame side effects and a smaller, yet significant, amount of very hazardous side effects, bringing the average way up in respect of, for example, the mode or median value. On the other hand, another question raised was the impact the number of edges of each node had on the AWD: Our concern was whether a low-degree node had a hire percentage of dangerous connections than a high-degree node, both having the same AWD; in parallel, we wondered wether top positions in the AWD were only reserved for high-degree nodes. We studied the perceived fear distribution of the side effects of the top 10 drugs ranked by AWD in the hope of highlighting unusual findings (Figs. 5-14).

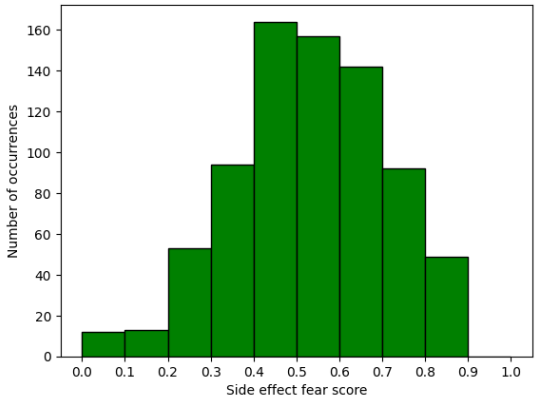

Figure 5: CID000145068

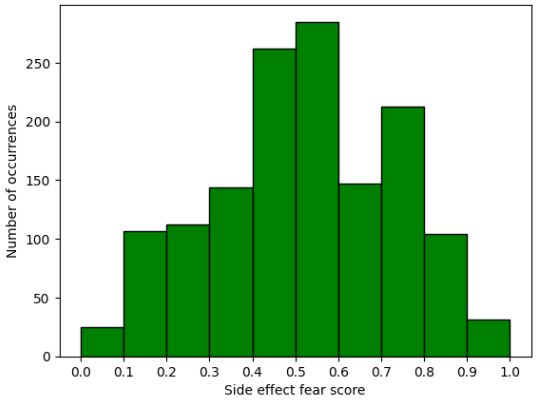

Figure 6: CID000004675

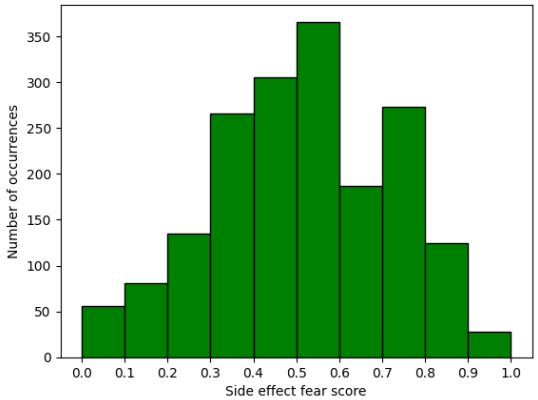

Figure 7: CID000036339

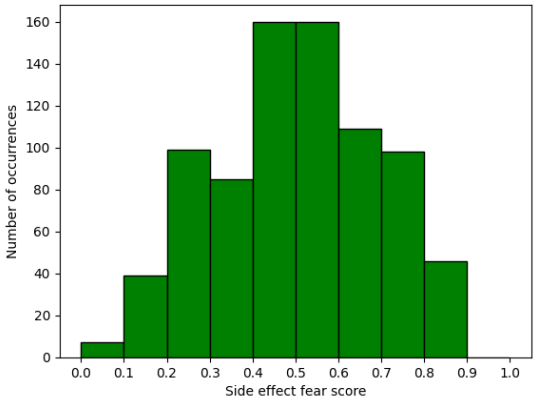

Figure 8: CID000036811

We can now confirm that the most common situation is a normal-like distribution around the average - and therefore, median, average and mode tend to be very close to each other. Similarly, the degree doesn't seem to influence neither the behavior nor the the AWD, as we can find both values higher than 1000 (Fig. 6) and as small as 13 (Fig. 14)

Edge Multiplicity (Drug-to-drug reactivity)
We applied edge multiplicity to assess the reactivity between each pair of drugs. Edge multiplicity measures the number of edges connecting a given pair of nodes, including parallel edges (6). In this context, it represents the number of distinct side effects resulting from the interaction between the two drugs.

$$EM(v,u) = \sum_{e \in E(v,u)} 1 \tag{6}$$

where $E(v,u)$ represents the set of edges connecting nodes v and u, including parallel edges.

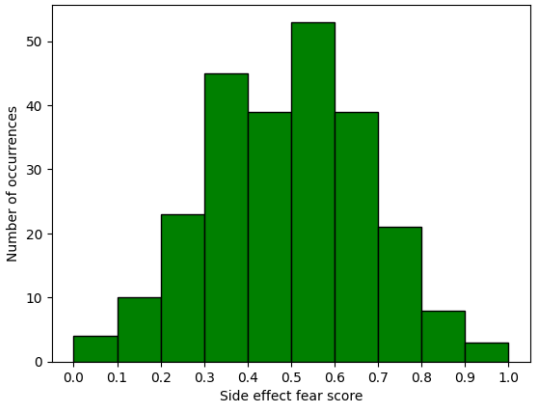

Figure 9: CID000002232

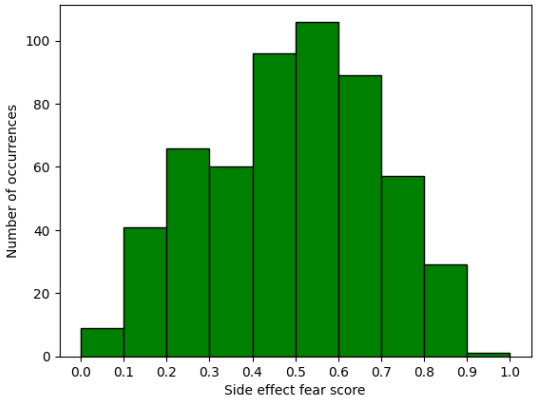

Figure 10: CID005282044

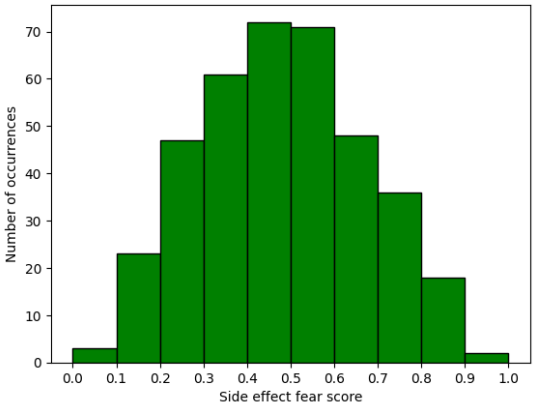

Figure 11: CID000005486

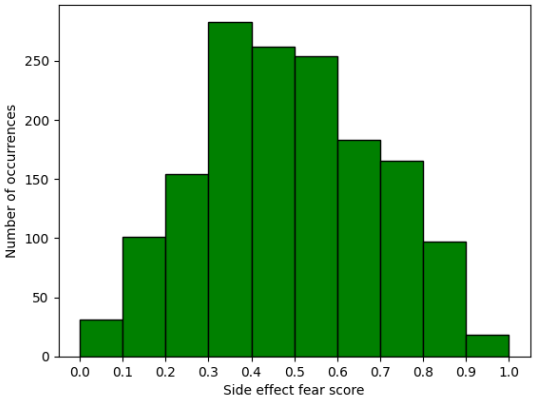

Figure 12: CID000041693

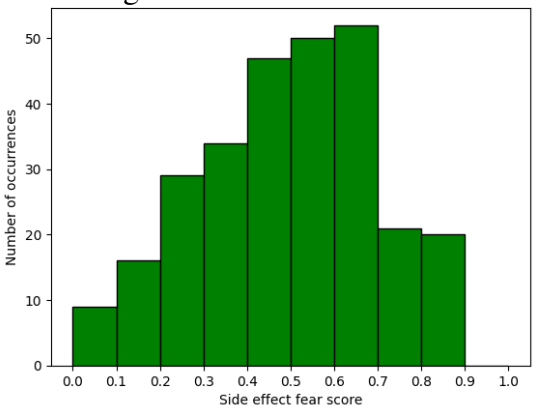

Figure 13: CID000004993

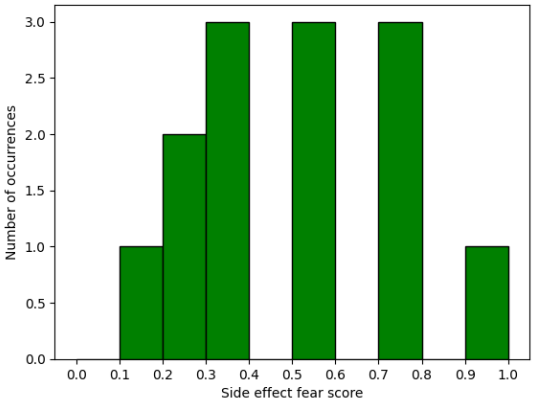

Figure 14: CID000003562

CID 1	CID 2	Drug Name 1	Drug Name 2	Score
CID000003883	CID000004594	Lansoprazole	Omeprazole	540
CID000003883	CID000002662	Lansoprazole	Celecoxib	524
CID000005090	CID000004594	Rofecoxib	Omeprazole	494
CID000005090	CID000005732	Rofecoxib	Zolpidem	478
CID000003446	CID000002662	Gabapentin	Celecoxib	476
CID000002678	CID000004900	Cetirizine	1,4-Pregnadiene-17alpha,21-diol-3,11,20-trione	468
CID000005090	CID000003958	Rofecoxib	Lorazepam	466
CID000005090	CID000004900	Rofecoxib	1,4-Pregnadiene-17alpha,21-diol-3,11,20-trione	465
CID000002662	CID000004900	Celecoxib	1,4-Pregnadiene-17alpha,21-diol-3,11,20-trione	462
CID000004635	CID000004594	Morphinan-6-one, 4,5-epoxy-14-hydroxy-3-methoxy-17-methyl-, (5alpha)-	Omeprazole	461

Table 7: Top 10 Drugs-Pair by Edge Multiplicity

Closeness centrality (Predictability of reactivity)

Analyzing closeness centrality in our case could serve as a significant measure of the predictability of a drug's reactivity. Suppose node A has a closeness centrality value X_A; if X_A is lower than the average, we can hypothesize that A is closer to the "center" of the network and more connected to highly connected nodes. Given that a highly connected node can be interpreted as a drug with a higher level of reactivity, we can also consider this high reactivity as "predictable" — meaning that, given another random drug, a connection between them is likely already present.

Thus, the closeness centrality value of A can be interpreted as an indicator of the predictability of its reactivity, as it directly correlates with the number of highly connected (and therefore more reactive) drugs it is linked to. To reformulate, the more a drug is reactive and connected to highly reactive drugs, the more its reactivity is predictable. We extracted the ten lowest closeness centrality values in the graph, as shown in Table 8, and compared them with the respective averaged weighted degree of each drug. While the two values do not exhibit a direct correlation, we found it interesting to explore the possibility of some form of relationship between them. However, our initial assumption was that no such correlation would emerge, and the data supports this hypothesis.

CID	Drug name	Closeness cent.	Av. weig. degree
CID000001134	1-[4-Hydroxy-5-(hydrox...	0.393162	0.401264
CID000004011	Maprotiline	0.425083	0.324730
CID000002177	3-[(4-Amino-benzenesulf...	0.452247	0.409967
CID000003562	Halothane	0.454481	0.477384
CID000051634	Miglustat	0.456738	0.400150
CID000005052	Isoreserpin	0.459344	0.406939
CID000003291	Ethosuximide	0.461318	0.449093
CID000039860	1-hydroxy-6,6-dimethyl-3-...	0.463309	0.393288
CID003086258	3-[18-(2-Carboxyethyl)-8...	0.465991	0.409241
CID000002170	Amoxapine	0.470760	0.408273

Table 8: Lowest ten closeness centralities values

5.3 Measures of the top 300 drugs prescribed in the US (2022)

As explained in Section 3, we used a dataset of the most prescribed drugs in the U.S. in 2022 to explore the results of our measures in a real-world scenario. We applied simple degree, multi-degree, weighted degree, and average weighted degree to the top 300 drugs that matched those in our network. Additionally, we checked for the distribution of perceived fear score distributions for of each node, confirming the results shown in the same examination applied to the top 10 ranked drugs per averaged weighted degree in 5.2 (Fig 15).
To make the analysis more readable and insightful, we did not report the raw values of the measures directly, but rather their ranks in the ordered list of all node results. We then converted each rank into a percentage expressing the normalized rank (where a rank of 645 equals 0% and a rank of 1 equals 100%) to simplify the comparison. For example, a percentage of 93% for the simple degree applied to Metformin means that, according to this measure, the drug ranks higher than 93% of the 645 items.

Looking at the ten most prescribed drugs in our network Table 9, it is striking to observe the consistently high percentages they exhibit across most measures. Although the average weighted degree shows slightly lower values, it still remains above the halfway mark in all cases. Given how commonly these drugs are prescribed, it is important to consider the potential risk of interaction — not only from a medical perspective but also from the standpoint of consumer safety.
It is interesting to note that Omeprazole, which ranks first in almost all global measures — simple degree centrality, multi-edge degree centrality, weighted degree, and average weighted degree — is also the ninth most prescribed drug.
Taking into consideration all the matching drugs in the top 300 (the CSV file with the measures is available in our GitHub repository[12]), there is slightly more variability in the results; however, higher percentages still prevail, with values below 50% being rare exceptions.

[12]Tisci Benini Molinati, *top 300 drugs with measures*, `https://github.com/side-effects-network/analysis-/blob/main/data/top_300_measures.csv`, 2025.

Rank	Drug Name	CID	Simple Degree	Multi Degree	Weighted Degree	Avg Weighted
2	Metformin	CID000004091	93%	93%	93%	70%
5	Amlodipine	CID000002162	91%	87%	88%	75%
6	Metoprolol	CID000004171	98%	97%	97%	68%
7	Albuterol	CID000002083	97%	99%	99%	61%
8	Losartan	CID000003961	92%	91%	91%	60%
9	Omeprazole	CID000004594	100%	100%	100%	59%
10	Gabapentin	CID000003446	93%	96%	96%	46%
12	Hydrochlorothiazide	CID000003639	94%	94%	94%	51%
16	Pantoprazole	CID000004679	99%	98%	98%	68%
18	Trazodone	CID000005533	67%	77%	76%	36%

Table 9: Matching drugs in the top 300 prescribed drugs in US (2022) and their ranking in the different measures expressed in percentages

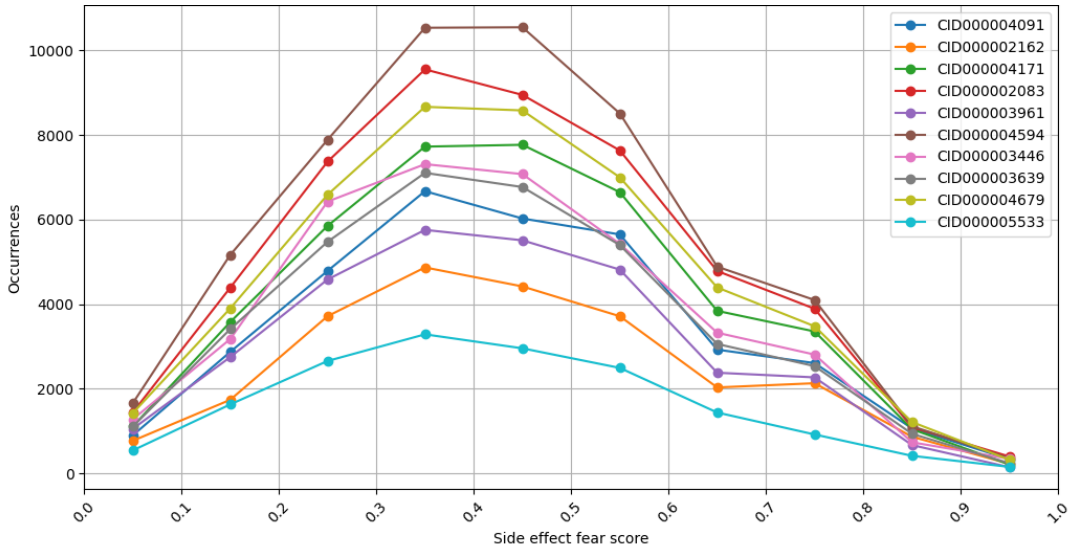

Figure 15: Perceived fear score distribution for drugs in top 300

6 Conclusion

From the general exploration of the network we were able to reconstruct a detailed image on shape and connectivity, alongside significant properties that enables us to contextualize later findings. The most prominent feature is that despite being a very large dataset, it is highly connected and cliques are very common. Not only that, but each node shows numerous multiple connections with each of its neighbors, resulting in a high-density network difficult to clearly navigate. We then discovered that the network contains a tight-knit k-core composed of more than one third of the nodes, around which we still find a well connected periphery. As previously said, this exploration has the goal to contextualize information about nodes and to understand the significance of the metrics we adopted, along with the results we obtained, in making sense of data characterized by high levels of complexity.

Regarding the study we conducted on specific samples and the metrics we applied, we'll re-

view the most prominent points briefly. At the same time, we would like to highlight possible concrete case scenarios in which these values might be relevant. identified two distinct types of drug reactivity: unique reactivity and cumulative reactivity. The first emphasizes the number of interactions with different drugs, while the second accounts for the frequency of interactions between the same pairs of drugs. In real-world terms, knowing that a drug has high unique reactivity — and that a patient is already taking other medications — can lead a healthcare practitioner to prescribe an alternative drug with a lower reactivity score. However, the practitioner should also consider cumulative reactivity to select a drug associated with fewer side effects. It may be even more useful to directly examine the Weighted Degree of the potential prescription, which combines cumulative reactivity with the perceived harmfulness of interactions. This helps avoid drugs that are both highly reactive and likely to cause dangerous side effects.
The Averaged Weighted Degree can be used to identify drugs associated with more severe side effects, regardless of their overall reactivity. This is particularly valuable because — even if the likelihood of a side effect occurring is lower due to limited reactivity — its severity may still warrant careful consideration in the practitioner's decision-making process.
In addition, Edge Multiplicity can be a crucial tool in cases of polypharmacy. It allows the practitioner to assess how intensely a potential prescription is likely to interact with the drug(s) the patient is already taking. If the score is too high, the practitioner can consider safer alternatives to minimize the risk of compounded interactions.
The analysis of the top 300 prescribed drugs in the U.S. (2022) provided the foundation for testing our hypothesis — that the values of our reactivity measures for these drugs would fall below the median. Interestingly, this hypothesis was strongly disproven: the scores for most metrics were consistently above the 90th percentile and, overall, significantly higher than the median. This finding suggests that drug interactions may be underestimated in current medical practice and warrant greater attention. We hope our analysis demonstrates the potential of the network-based approach in addressing the complex issue of drug interactions.

7 Critique

In our study we were presented with two main challenges. The first was that graph visualization was not a useful resource due to the graph's densely connected structure and large size, which limited the amount of meaningful information it could convey. We addressed this by using general measures to assess the structure and connectivity of the graph. The second, and more important, concern was identifying and applying measures that carry semantic meaning in the context of polypharmacy — and justifying their use. We believe this was achieved by thoroughly explaining what each measure implies for drug utilization, both theoretically and through concrete examples.

Another important aspect of our study was the application of our semantically meaningful measures to the most commonly prescribed drugs in recent years, in order to test our hypothesis. The available data was sufficient to disprove our initial assumption — thus leading to a meaningful finding — but the results would have been more robust with a higher number of matches between our network and the external dataset.

Overall, the main problems arising from our study could be articulated in two points: one stemming from the nature of the network, the other from the limitation of data available. The former can be identified in the insignificance of distance measures in the network, severely limiting our possibilities when conducting a network-centered study; the latter simply means

we could have gathered a larger amount of more diverse data, but data on a larger sample were unavailable, affecting our possibility to extract further knowledge. In conclusion, we believe to have tackled the research question, which was set to be an exploration in the first place, and consequently a study on drug interaction and consumption behaviors. However, the field in question poses a series of limits that could only be resolved by taking into consideration either upscaling into a quantitative analysis or, contrarily, downscaling with a narrower segment of the network and a more specific research question.

References

Gottlieb, Assaf, et al. "Ranking Adverse Drug Reactions With Crowdsourcing". *Journal of Medical Internet Research* 17, no. 3 (Mar. 2015). https://doi.org/https://doi.org/10.2196/jmir.3962.

Kuhn, Michael, et al. "The SIDER database of drugs and side effects". *Nucleic Acids Research* 44, no. D1 (Oct. 2015): D1075–D1079. https://doi.org/https://doi.org/10.1093/nar/gkv1075. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702794/.

Zitnik, Marinka, Monica Agrawal, and Jure Leskovec. *BioSNAP: Network datasets: Poypharmacy side-effect association network*, 2018. https://snap.stanford.edu/biodata/datasets/10017/10017-ChChSe-Decagon.html.

– . "Modeling polypharmacy side effects with graph convolutional networks". *Bioinformatics* 34, no. 13 (July 2018): i457–i466. https://doi.org/https://doi.org/10.1093/bioinformatics/bty294. https://academic.oup.com/bioinformatics/article/34/13/i457/5045770.

Wolf, Thomas, et al. "Transformers: State-of-the-Art Natural Language Processing". In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. Online: Association for Computational Linguistics, Oct. 2020. https://www.aclweb.org/anthology/2020.emnlp-demos.6.

Kane, Sean P. *The Top 300 Drugs of 2022*, 2022. https://clincalc.com/DrugStats/Top300Drugs.aspx.

Lab, Tatonetti. *databases for drug side effects and drug interactions*, 2023. https://nsides.io/.

Milano, Marianna, and Mario Cannataro. "Network models in bioinformatics: modeling and analysis for complex diseases". *Briefings in bioinformatics* 24, no. 2 (Jan. 2023). https://doi.org/https://doi.org/10.1093/bib/bbad016.

2025. https://pubchem.ncbi.nlm.nih.gov/.

Benini, Tisci, Molinati. *top 300 drugs with measures*. https://github.com/side-effects-network/analysis-/blob/main/data/top_300_measures.csv, 2025.