Uncovering new drug properties in target-based drug-drug similarity networks

Lucreţia Udrescu[1], Paul Bogdan[2], Aimée Chiş[3], Ioan Ovidiu Sîrbu[3,6], Alexandru Topîrceanu[4], Renata-Maria Văruţ[5], Mihai Udrescu[4,6*]

1 Department of Drug Analysis, "Victor Babeş", University of Medicine and Pharmacy Timişoara, Timişoara, Romania
2 Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA
3 Department of Biochemistry, "Victor Babeş" University of Medicine and Pharmacy Timişoara, Timişoara, Romania
4 Department of Computer and Information Technology, Politehnica University of Timişoara, Timişoara, Romania
5 Faculty of Pharmacy, University of Medicine and Pharmacy of Craiova, Craiova, Romania
6 Timişoara Institute of Complex Systems, Timişoara, Romania

* mudrescu@cs.upt.ro

Abstract

Despite recent advances in bioinformatics, systems biology, and machine learning, the accurate prediction of drug properties remains an open problem. Indeed, because the biological environment is a complex system, the traditional approach – based on knowledge about the chemical structures – cannot fully explain the nature of interactions between drugs and biological targets. Consequently, in this paper, we propose an unsupervised machine learning approach that uses the information we know about drug-target interactions to infer drug properties. To this end, we define drug similarity based on drug-target interactions and build a weighted Drug-Drug Similarity Network according to the drug-drug similarity relationships. Using an energy-model network layout, we generate drug communities that are associated with specific, dominant drug properties. However, 13.59% of the drugs in these communities seem not to match the dominant pharmacologic property. Thus, we consider them as drug repurposing hints. The resources required to test all these repurposing hints are considerable. Therefore we introduce a mechanism of prioritization based on the betweenness/degree node centrality. By using betweenness/degree as an indicator of drug repurposing potential, we identify the drug Meprobamate as a possible antifungal. Finally, we use a robust test procedure, based on molecular docking, to further confirm the repurposing of Meprobamate.

Author summary

We address the salient problem of drug repurposing by using an unsupervised machine learning technique (i.e., clustering) on a drug network (i.e., nodes represent drugs and links represent relationships between two drugs). To this end, we build a drug-drug similarity network (DDSN) using information about drug-target interactions, then

perform network clustering and associate the clusters with drug properties. To validate our DDSN clustering approach, we use an old drug database to render the cluster-property pairs and then verify them with the information from the latest database. Indeed, the properties uncovered with our DDSN clustering are confirmed for 86.41% of the drugs; for the remaining 13.59%, we assume that the associated property is not fortuitous, thus representing repositioning hints. As the number of repositioning hints is considerable, we prioritize them with a new approach, based on the node betweenness/degree ratio. The prioritization of repositioning hints renders the repurposing of Meprobamate (known as a hypnotic and sedative drug) as an antifungal drug. We also test and confirm the robustness of our hypothesis using an original, formal procedure based on computer-simulated molecular docking.

Introduction

Conventional drug design has become expensive and cumbersome, as it requires large amounts of resources and faces serious challenges [1,2]. Consequently, although the number of new FDA drug applications (NDAs) has significantly increased during the last decade – due to a spectacular accumulation of multi-omics data and appearance of increasingly complex bioinformatics tools – the number of approved drugs has only marginally grown (see Figure 1) [3,4], calling for more robust alternative strategies [5].

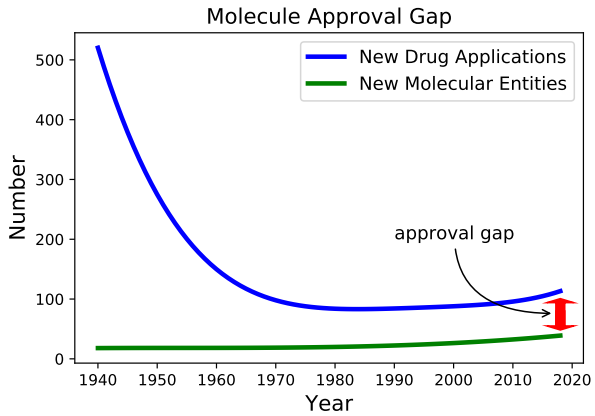


Fig 1. The evolution of New Drug Applications (NDAs) and New Molecular Entities (NME) during the period 1940-2017, according to FDA reports. We used the FDA's annual reports data (https://www.fda.gov) and removed local oscillations by plotting a polynomial data fitting.

One of the most effective alternative strategies is *drug repositioning* (or *drug repurposing*) [6,7], namely finding new pharmaceutical functions for already used drugs. The extensive medical and pharmaceutical experience reveals a surprising propensity towards multiple indications for many drugs [8], and the examples of successful drug repositioning are steadily accumulating. Out of the 90 newly approved drugs in 2016 (a 10% decrease from 2015), 25% are repositionings in terms of formulations, combinations, and indications [4]. Furthermore, drug repositioning reduces the incurred research and development (R&D) time and costs, as well as medication risks, which makes it particularly efficient for developing orphan/rare disease therapies [8,9].

The recent developments confirm computational methods as powerful tools for drug repositioning:

- The trivialization/spread of omics analytical approaches have generated significant

volumes of useful multi-omics data (genomics, proteomics, metabolomics, and others) [10, 11].

- The ubiquity of digitalization in everyday life, including social media, has tremendously expanded the amplitude of the process of gathering data on drug-drug interactions and drug side-effects [12, 13].

- The recent developments in physics, computer science, and computer engineering have created efficient methods and technologies for data exploration and mining, such as complex network analysis, machine learning, or deep learning [11, 14–18].

Complex network analysis has proven to be a useful tool for predicting unaccounted drug-target interactions. Indeed, several state-of-the-art network-based computational drug repurposing approaches use data on confirmed drug-target interactions to predict new such interactions, thus leading to new repositioning hints [19–21]. Some approaches build drug-drug similarity networks, where the similarity is defined based on transcriptional responses [22, 23]. These repositioning approaches analyze the network parameters and the node centrality distributions in either drug-drug or drug-target networks, using statistical analysis [10, 11, 24, 25] and machine learning (i.e., graph convolutional networks) [26–29] to link certain drugs to new pharmacological properties. However, conventional statistics can be misleading when used to predict extreme centrality values, such as degree and betweenness (which particularly indicate nodes/drugs with a high potential for repositioning) [30].

To overcome these challenges, we developed a novel, network-based, computational approach to drug repositioning. To this end, we build a weighted *drug-drug network*, i.e., a complex network where the nodes are drugs, and the weighted links represent relationships between drugs, using information from the accurate DrugBank [31]. In our drug-drug similarity network (DDSN), a link is placed between two drugs if their interaction with at least one target is of the same type (either agonistic or antagonistic). The link weight represents the number of biological targets that interact in the same way with the two drugs. Figure 2 illustrates the building of the DDSN with information on drug-target interactions.

Our methodology for analyzing the drug-drug similarity network (DDSN) consists of the following steps:

1. Generate (using the Force Atlas 2 layout and modularity classes) [32, 33] both topological clusters, and network communities.

2. Relate each cluster and each community to a pharmacological property or pharmacological action (i.e., label communities and clusters according to the dominant property or pharmacological action), using expert analysis.

3. Identify and select (by betweenness divided by degree, $\frac{b}{d}$) within each topological cluster/modularity class community, the top drugs not compliant with the cluster/community label.

4. Validate the hinted repositionings by searching the new versions of DrugBank, the electronic records containing the relevant scientific literature (for merely reconstructed repositionings), and by analyzing molecular docking parameters [34] for previously unaccounted repositionings.

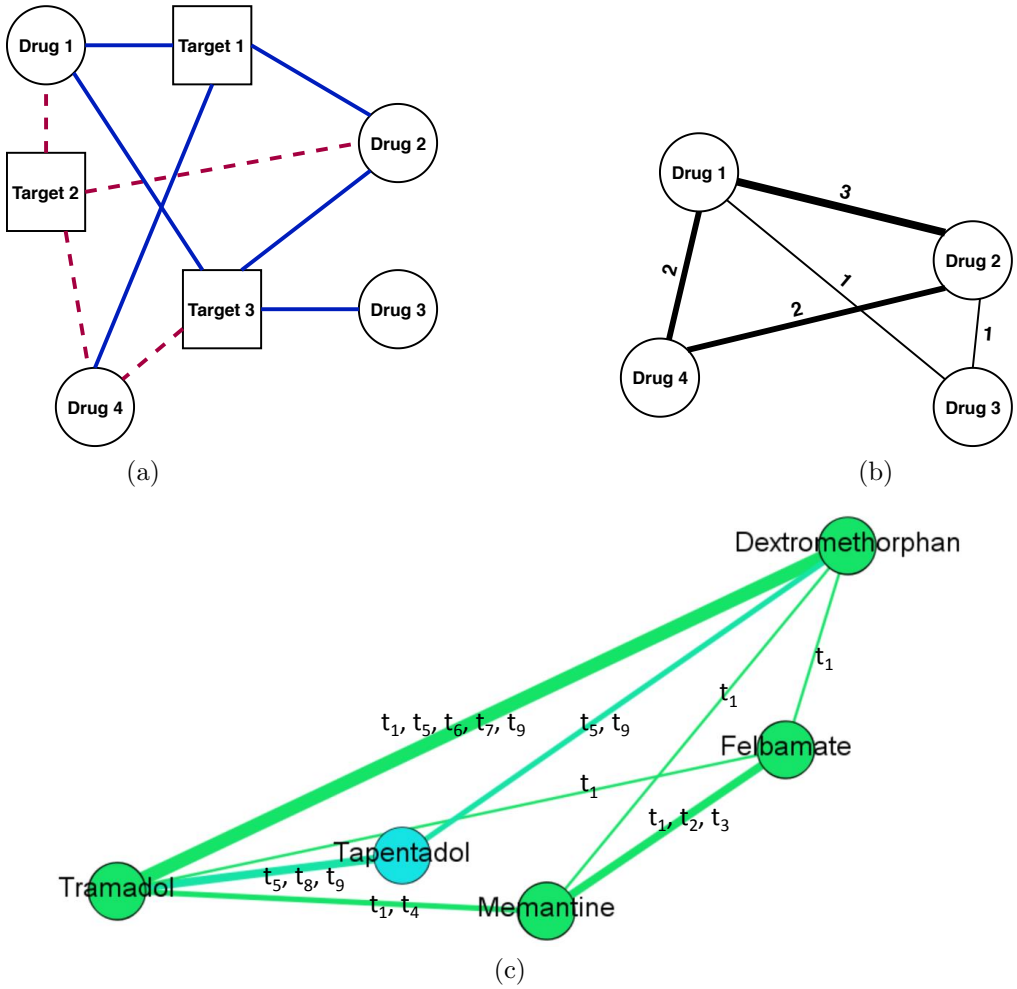


Fig 2. An illustrative example of using drug-target interaction information to build a weighted drug-similarity network. In panel (a), we consider the drug-target interactions between four drugs (i.e., round nodes labeled 1 to 4) and three targets (i.e., square nodes labeled 1 to 3). The dashed red links represent agonist drug-target interactions, whereas the solid blue links represent antagonist drug-target interactions. In panel (b), we show the DDSN corresponding to the interactions in (a). For instance, a link of weight 3 connects the nodes 1 and 2 because Drug 1 and Drug 2 interact in the same way for the three targets, i.e., agonist on Target 2 and antagonist on Targets 1 and 3. Also, a link with weight 2 connects Drug 2 and Drug 4 because they both interact agonistically on Target 2 and antagonistically on Target 1, but they do not interact in the same way with Target 3. In panel (c), we show a DDSN sub-network example, according to drug-target interactions from DrugBank 4.2, containing drugs Dextromethorphan, Felbamate, Tapentadol, Tramadol, and Memantine. We shape the link thickness according to the weight and specify the list of common targets for each link. The weight equals the number of targets in the list, where t_1 = Glutamate receptor ionotropic NMDA 3A, t_2 = Glutamate receptor ionotropic NMDA 2A, t_3 = Glutamate receptor ionotropic NMDA 2B, t_4 = Alpha-7 nicotinic cholinergic receptor subunit, t_5 = Mu-type opioid receptor, t_6 = Kappa-type opioid receptor, t_7 = Delta-type opioid receptor, t_8 = Sodium-dependent noradrenaline transporter, and t_9 = Sodium-dependent serotonin transporter.

Results

Building the DDSN

We built our DDSN as a weighted graph $G = (V, E)$ with $|V| = n$ vertices (nodes) representing drugs and $|E| = n_E$ edges representing drug-drug similarity relationships based on drug-target interactions. The weight associated with an edge $e_{ij} \in E$ represents the degree of *target action similarity* between drugs v_i and v_j, and it is equal with the number of *common biological targets* for v_i and v_j. If $e_{ij} = 0$, then there is no target similarity between v_i and v_j, therefore no edge between these nodes. A *common biological target* is a target $t_k \in T$ (T is the set of targets) on which drugs v_i and v_j act in the same way, either both agonistically or both antagonistically. Within this framework, we build the DDSN graph G using drug-target interaction information from Drug Bank 4.2 [31]. We base our analysis on the largest connected component of the DDSN, consisting of $|V| = 1008$ drugs/nodes and $|E| = 17963$ links resulted from the analysis of the drug-target interactions with $|T| = 516$ targets.

To mine the DDSN topological complexity, we identified the drug clusters (or communities) using both the modularity [33] and the force-directed, energy-based layout Force Atlas 2 [32] algorithms. The two clustering techniques are compatible [35]; however, the energy-based force-directed layout clustering offers more information about the relationship between clusters and acts as an efficient classifier [36]. In the case of DDSN, the clusters correspond to drug communities C_x, $x \in \mathbb{N}^*$, such that $V = \bigcup_{i=1}^{m} C_x$ (We provide the Gephi project of the DDSN in the Supplementary S1 File). Using the constructed DDSN from Drug Bank 4.2 and expert analysis, we label each cluster according to its dominant property (i.e., the property that better describes the majority of drugs in the cluster), which may represent a specific mechanism of pharmacologic action, a specifically targeted disease, or a targeted organ. Figure 3 illustrates the resulted DDSN, where the node colors identify the distinct modularity clusters. We assessed the ability of our method to uncover new repositionings by confronting our results with the latest (version 5.1.4) Drug Bank and with data compiled from interrogating scientific literature databases.

DDSN analysis and repositioning hints

In our characterization of drug-drug similarity networks, a high degree node represents a drug with already documented multiple properties. Also, a high betweenness (i.e., the ability to connect network communities) indicates the drug's propensity for multiple pharmacological functions. By this logic, the high-betweenness, high-degree nodes may have reached their full repositioning potential, whereas the high betweenness, low degree nodes (characterized by high betweenness/degree value $\frac{b}{d}$) may indicate a significant repositioning potential. However, predicting such high-value cases of degree d, weighted degree d_w, betweenness b, and betweenness/degree $\frac{b}{d}$ is difficult because the corresponding distributions are fat-tailed [37]. Although all the estimated DDSN centralities are following a power-law distribution (see Figure 4), the betweenness/degree $\frac{b}{d}$ is the most stable parameter and, hence, the most reliable indicator of multiple drug properties.

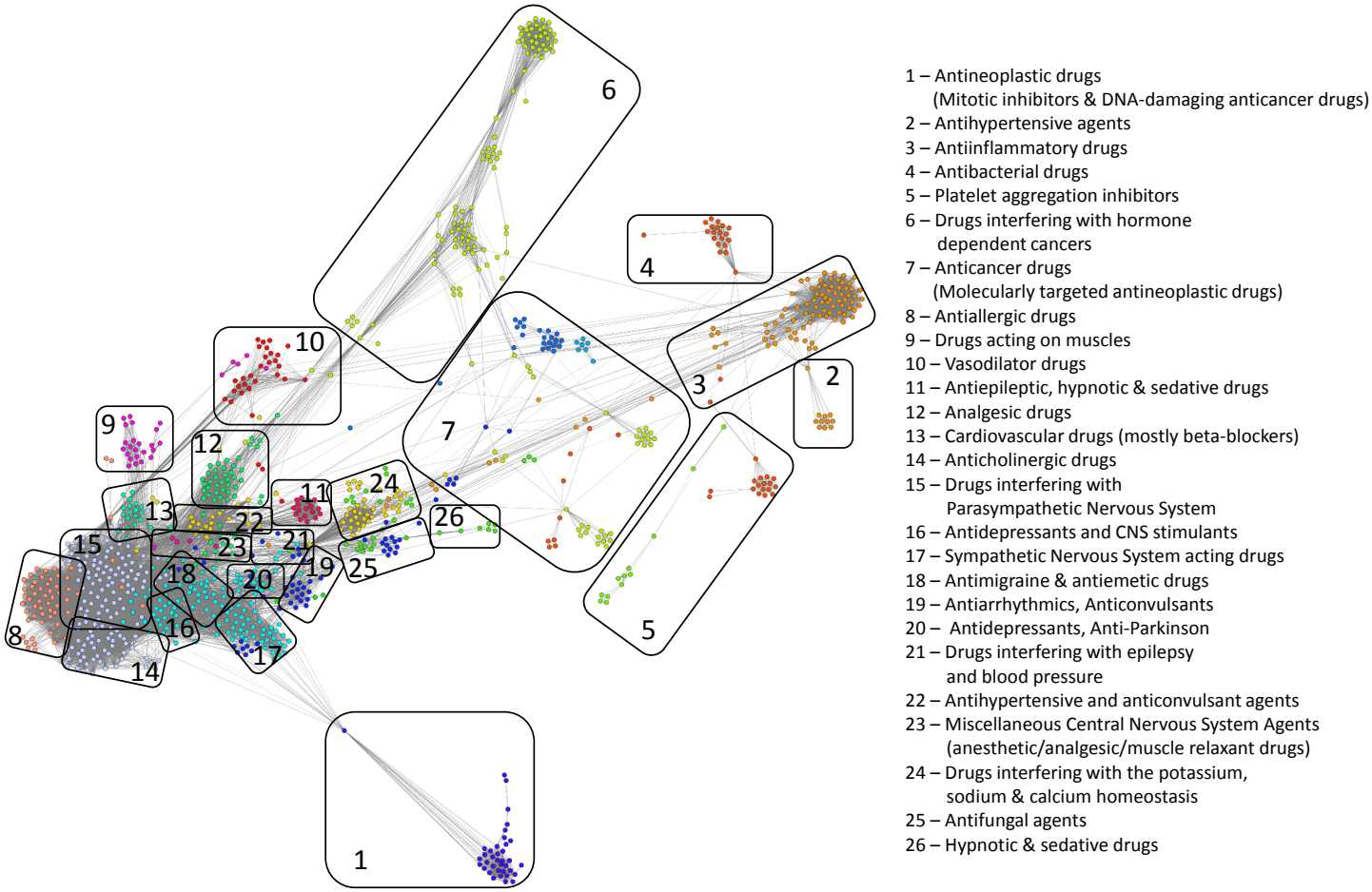


Fig 3. The drug-drug similarity network, where nodes represent drugs and links represent drug-drug similarity relationships based on drug-target interaction behavior. The layout is Force Atlas 2 [32] and the distinct node colors identify the modularity classes that define the drug clusters. We identify the 26 topological clusters with rounded rectangles and provide the functional descriptions for each of them.

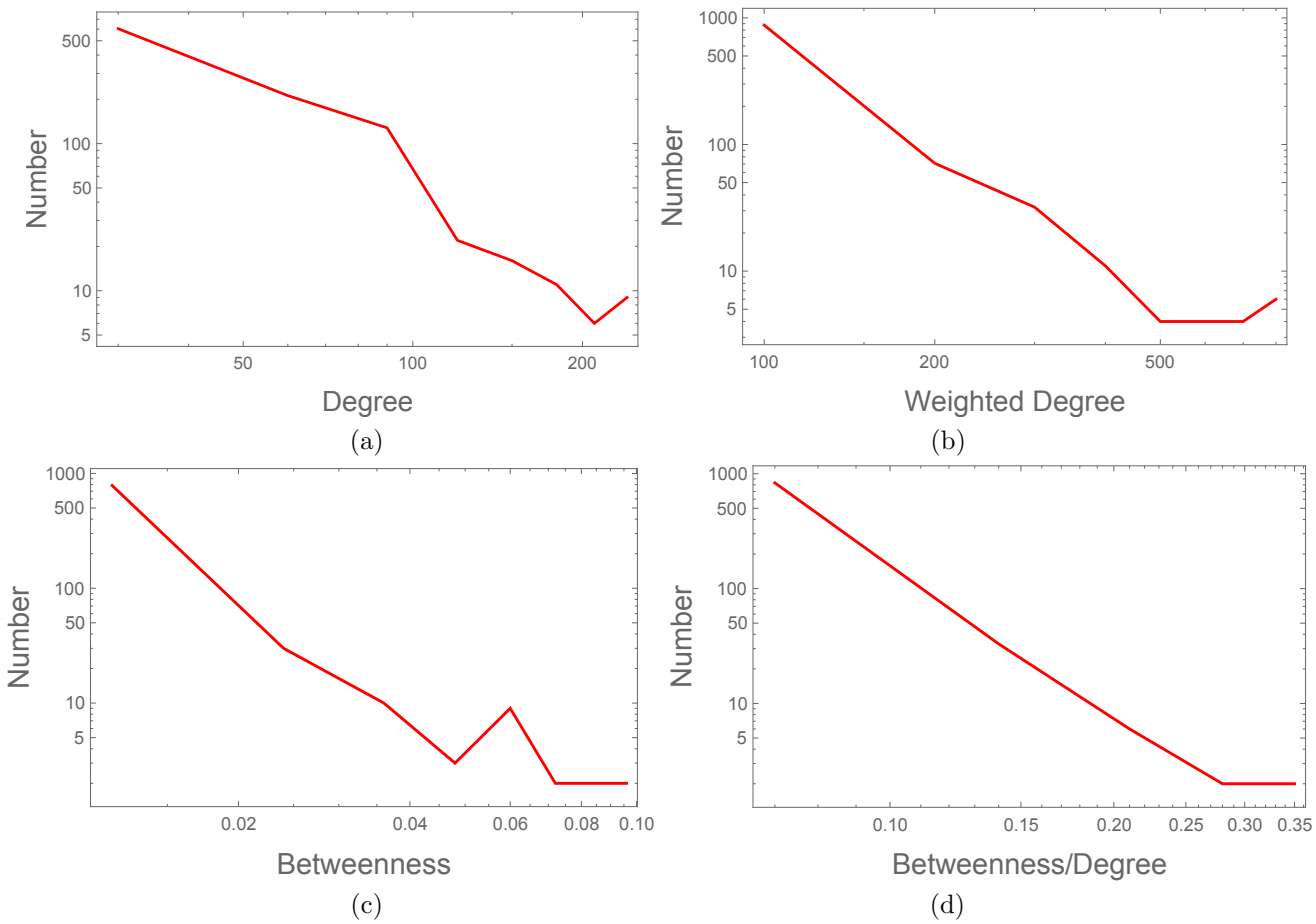


Fig 4. Power-law distributions of centrality parameters in the drug-drug similarity network (DDSN): (a) degree d, (b) weighted degree d_w, (c) betweenness b, and (c) betweenness/degree $\frac{b}{d}$. We represent the distributions according to the guidelines in [38], by using 8 linearly spaced bins for each centrality. The fitting analysis using the *Powerlaw* package in Python [38] indicates the following values for the distribution slope α and cutoff point x_{min}, respectively: 3.436 and 53 for d, 2.598 and 64 for d_w, 2.201 and 0.008 for b, 3.093 and 0.088 for $\frac{b}{d}$. The graphical representations of these centrality distributions show that the betweenness/degree $\frac{b}{d}$ is the most stable parameter; therefore it is the most reliable indicator of multiple drug properties.

To explore the capability of $\frac{b}{d}$ to predict the multiple drug properties, we exploit the community structure of DDSN by following a two-step approach.

1. We uncover the relevant drug properties by generating network communities C_x with $x = \overline{1, m}$ ($m = 26$ in our DDSN). Then, using expert analysis, we assign a dominant property to each community. Figure 3 illustrates the 26 DDSN communities as well as their dominant functionality. The dominant community property can be a pharmacological mechanism, a targeted disease, or a targeted organ. For instance, the community 1 (C_1) consists of antineoplastic drugs which act as mitotic inhibitors and DNA damaging agents; Community 13 (C_{13}) consists of cardiovascular drugs (antihypertensive, anti-arrhythmic, and anti-angina drugs), mostly beta-blockers.

2. In each cluster C_x, we identify the top t drugs according to their $\frac{b}{d}$ values. From these selected drugs, $B_x^t \subset C_x$, some stand out by not sharing the community property or properties, and thus, can be repositioned as such. To this end, for

$x = \overline{1, m}$ eliminated from B_x^t the drugs whose repurposings were already confirmed (i.e., performed by others and found in the recent literature), thus producing $m = 26$ lists of repurposing hints yet to be confirmed by *in silico*, *in vitro*, and *in vivo* experiments, $B_x^h = B_x^t \setminus B_x^c$. Table 1 presents the lists of B_x^t drugs for $t = 5$ and $x = 26$ (i.e., the top 5 $\frac{b}{d}$ drugs in each community). We chose $t = 5$ to provide a reasonable amount of eloquent information in Table 1; we provide the entire B_x sets in the Supplementary S2 File.

Table 1. Top 5 drugs ('B_x^5' with $x = \overline{1, 26}$) according to their $\frac{b}{d}$ values, for each of the 26 DDSN communities/clusters ('C_x'). The properties of drugs written in regular fonts match the properties of their respective communities (according to the Drug Bank). The properties of italicized drugs do not match all the properties of their respective communities, but the latest literature confirms them (drugs in regular fonts and italics pertain to B_x^h). The properties of the drugs written in bold do not match the community properties, and the literature did not confirm them yet; this situation leads to new drug repositioning hints (i.e., the B_x^h drugs). The positions marked with '–' correspond to drugs with $\frac{b}{d} = 0$.

C_x	B_x^5				
	1	2	3	4	5
1	Amsacrine	*Colchicine*	*Podofilox*	*Lucanthone*	**Besifloxacin**
2	–	–	–	–	–
3	*Amiloride*	*Marimastat*	Diclofenac	Thalidomide	Telmisartan
4	Minocycline	Framycetin	Amikacin Tobramycin Netilmicin	Doxycycline Clomocycline Oxytetracycline	–
5	Treprostinil	Iloprost	*Captopril*	*Bimatoprost*	*Candoxatril*
6	*Progesterone*	*Mimosine*	*Fluticasone propionate*	*Danazol*	*Spironolactone*
7	Vandetanib	*Dalteparin*	*Dehydroepiandrosterone*	*Amlexanox*	*Atorvastatin*
8	Olopatadine	Terfenadine	Flunarizine	Astemizole	Epinastine
9	Succinylcholine	Carbachol	Decamethonium	Pilocarpine	Cevimeline
10	*Nicotine*	*Melatonin*	Amrinone	Dipyridamole	*Naloxone*
11	*Quinine*	Phenobarbital Secobarbital Pentobarbital	Barbital Hexobarbital Aprobarbital	–	–
12	*Nimodipine*	Adenosine	Drotaverine	Pentazocine	*Loperamide*
13	*Ketotifen*	Amiodarone	Sotalol	Bevantolol	Penbutolol
14	Disopyramide	Scopolamine	Ethopropazine	Paroxetine	Rocuronium
15	Minaprine	Amitriptyline	*Agomelatine*	Orphenadrine	Imipramine
16	Cocaine	*Chloroprocaine*	*Procaine*	Phenermine	Milnacipran
17	Epinephrine	4-Methoxyamphetamine	Pseudoephedrine	Ephedra	Methamphetamine
18	*Ginkgo biloba*	**Captodiame**	*Cisapride*	*Bromocriptine*	*Carteolol*
19	**Acarbose**	Lidocaine	Mexiletine	*Etomidate*	Flecainide
20	Phenelzine	*Agmatine*	**Quinidine** **Propafenone**	Ephedrine	*Amphetamine*
21	Zonisamide	**Miconazole**	*Ethanol*	**Quinidine barbiturate**	–
22	*Felodipine*	*Bepridil*	*Verapamil*	*Dextromethorphan*	*Amlodipine*
23	Halothane	**Halofantrine**	Tramadol	*Ibutilide*	Tubocurarine
24	*Thiamylal*	*Valproic Acid*	**Progabide**	**Bethanidine**	*Topiramate*
25	**Meprobamate**	**Enflurane**	Tioconazole	Clotrimazole	*Methoxyflurane* **Isoflurane** **Sevoflurane**
26	Flunitrazepam	Eszopiclone	–	–	–

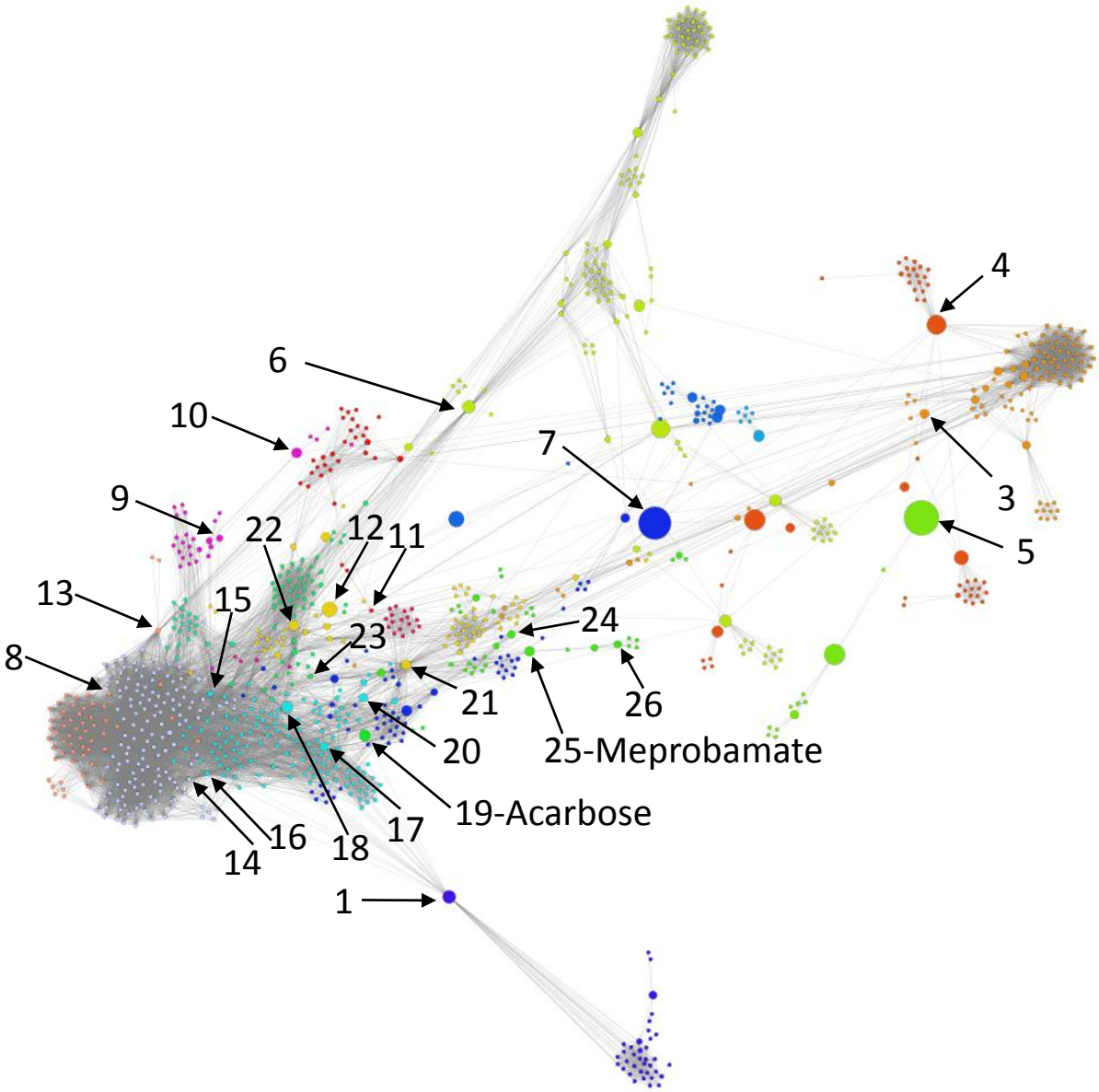


Fig 5. Drug-drug similarity network (DDSN), based on drug-target interactions, where node sizes represent their $\frac{b}{d}$ values. The arrows indicate the top $\frac{b}{d}$ node in each community (for community 2 there is no top node because all drugs have $\frac{b}{d} = 0$). The community id identifies each top $\frac{b}{d}$ node, excepting Meprobamate (top $\frac{b}{d}$ in community 25) and Acarbose (community 19), because these drugs (apparently) do not have their community's property; this indicates Meprobamate as antifungal (i.e., the property of community 25) and Acarbose as antiarrhythmic, anticonvulsant (i.e., the properties of community 19).

To facilitate the visual identification of the repositioning hints, in Figure 5, we shape the size of the nodes of our DDSN representation according to the magnitude of the $\frac{b}{d}$ values. We also identify, by arrows, the top $\frac{b}{d}$ nodes (i.e., drugs) in their respective communities, by indicating their community id. Table 1 shows that $B^1_x = \emptyset$ for all x except 19 and 25 ($B^1_{19} = \{\text{Acarbose}\}$ and $B^1_{25} = \{\text{Meprobamate}\}$), therefore – besides the corresponding community number – we expressly point Acarbose and Meprobamate in Figure 5.

The high percentage of database and literature confirmations of our pharmacological properties predictions are highlighting the robustness of our repurposing method. In the Supplementary Information material (Supplemantary File S2), we show that the confirmation rate $\sum_x B_x^c / \sum_x C_x = 86.41\%$. Table 1, presents a similar situation, with only a few unconfirmed drug properties (these repurposing hints $\in B_x^h$ are in bold).

Our data indicate two top $\frac{b}{d}$ drugs: Meprobamate, in the C_{25} antifungal drugs community, and Acarbose, in the C_{19} (Antiarrhythmics and Anticonvulsants) community. Both repositionings refer to properties currently unaccounted in the DrugBank version 5.1.4 and the scientific literature we have screened (Table 1, Figures 5 and 6); Meprobamate is a hypnotic, sedative, and mild muscle-relaxing drug, while Acarbose is a hypoglycemic drug.

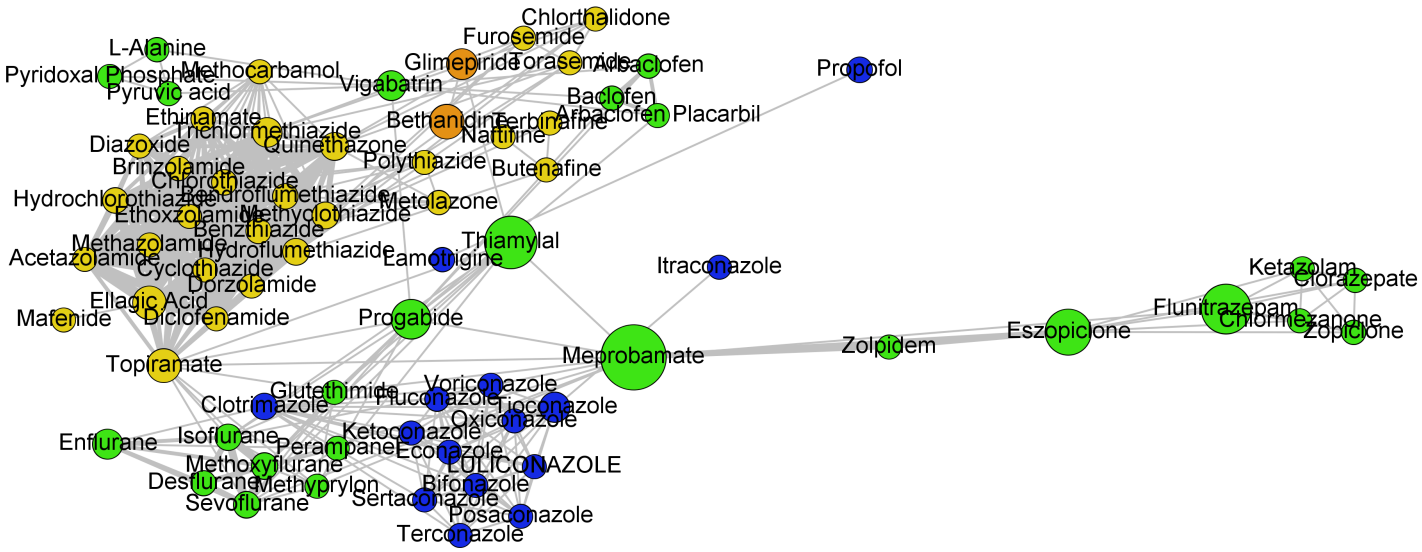


Fig 6. DDSN detail showing Meprobamate as the node with the biggest $\frac{b}{d}$ in its community (the node sizes are proportional to their $\frac{b}{d}$ values). Meprobamate is a member of the (green) community of psychotropic drugs but is also well connected to the (dark blue) community of antifungal drugs. The placement of Meprobamate and the high $\frac{b}{d}$ value suggest that it may also have an antifungal effect.

Repurposing hints testing

Molecular docking uses the target and ligand structures to predict the lead compound or repurpose drugs for different therapeutic purposes. As such, the molecular docking tools predict the binding affinities, the preferred poses, and the interactions of the ligand-receptor complex with minimum free energy. In this paper, we use the AutoDock 4.2.6 software suite [39], which consists of automated docking tools for predicting the binding of small ligands (i.e., drugs) to a macromolecule with an established 3D structure (i.e., target). The AutoDock semi-empirical free energy force field predicts the binding energy by considering complex energetic evaluations of bound and unbound forms of the ligand and the target, as well as an estimate of the conformational entropy lost upon binding.

Testing procedure

To confirm the antifungal property of Meprobamate, we perform molecular docking not only for our hint but also for two drugs that are not known as antifungal agents, thus

facilitating the analysis of their interaction to the target in comparison to that of the reference drugs. As such, we formalize the following testing procedure.

1. We define the drug sets to enter the docking process, consisting of drugs hinted as having the pharmacological property ϕ ($\mathcal{D}_h^\phi$), well-documented drugs with property ϕ (reference drugs $\mathcal{D}_r^\phi$), and drugs with little probability of having property ϕ ($\mathcal{D}_n^\phi$). Our goal is to explore the similarity (in terms of relevant target activity) between the reference drugs $\mathcal{D}_r^\phi$ and the tested drugs $\mathcal{D}_t^\phi = \mathcal{D}_h^\phi \bigcup \mathcal{D}_n^\phi$. In our case, the property ϕ is the antifungal effect and $x = 25$ (corresponding to community C_{25}).

   (a) $\mathcal{D}_h^\phi$ consists of just one hinted drug $\mathcal{D}_h^\phi = \left\{d_h^\phi\right\}$, namely Meprobamate.

   (b) $\mathcal{D}_r^\phi$ consists of two subsets, reference drugs in the DDSN's community C_x ($\mathcal{D}_x^\phi$) and reference drugs not in C_x ($\mathcal{D}_{\overline{x}}^\phi$), with $\mathcal{D}_r^\phi = \mathcal{D}_x^\phi \bigcup \mathcal{D}_{\overline{x}}^\phi$. Therefore, we define $\mathcal{D}_{25}^{\text{antifungal}} = \{\text{Clotrimazole}, \text{Oxiconazole}\}$ and $\mathcal{D}_{\overline{25}}^{\text{antifungal}} = \{\text{Naftifine}, \text{Tolnaftate}, \text{Nystatin}, \text{Natamycine}, \text{Ciclopirox}, \text{Griseofulvin}\}$.

   (c) $\mathcal{D}_n^{\text{antifungal}} = \{\text{Fosinopril}, \text{Furosemide}\}$. Both drugs are very unlikely to have antifungal effects; Fosinopril is an antihypertensive, and Furosemide is a diuretic (the former is topologically far from Meprobamate in the DDSN, while the latter is close).

2. We establish the target sets. Specifically, for pharmacological property ϕ, we take into consideration the targets from DrugBank that interact with the drugs in the hinted drug d_h^ϕ's community C_x having property ϕ ($\mathcal{T}_x^\phi$), and the targets from DrugBank that interact with the drugs with property ϕ not included in DDSN's C_x ($\mathcal{T}_{\overline{x}}^\phi$). As such, for $\phi =$ antifungal and $x = 25$, we have:

   (a) $\mathcal{T}_{25}^{\text{antifungal}} = \{$Lanosterol 14-alpha demethylase, Lanosterol synthase, Intermediate conductance calcium-activated potassium channel protein 4$\}$,

   (b) $\mathcal{T}_{\overline{25}}^{\text{antifungal}} = \{$Squalene monooxygenase, Ergosterol, Sodium/potassium-transporting ATPase subunit alpha, Tubulin$\}$.

3. For the set of tested drugs $\mathcal{D}_t^\phi$, we use molecular docking to check the interactions between all possible drug-target pairs, defined as the Cartesian product of sets $\mathcal{D}_t^\phi$ and $\mathcal{T}^\phi$ (with $\mathcal{T}^\phi = \mathcal{T}_x^\phi \bigcup \mathcal{T}_{\overline{x}}^\phi$),

$$\mathcal{D}_t^\phi \times \mathcal{T}^\phi = \left\{(d_i, t_j) : d_i \in \mathcal{D}_t^\phi, t_j \in \mathcal{T}^\phi, \forall i, j \in \mathbb{N}^*, i \leq |\mathcal{D}_t^\phi|, j \leq |\mathcal{T}^\phi|\right\}. \quad (1)$$

4. For the set of reference drugs, we apply molecular docking on separately designed drug-target pairs for reference drugs in C_x ($\mathcal{D}_x^\phi$), and reference drugs not in C_x ($\mathcal{D}_{\overline{x}}^\phi$) respectively, such that any drug-target pair is well-documented in the literature,

$$\left\{(d_i, t_j) : d_i \in \mathcal{D}_x^\phi, t_j \in \mathcal{T}_x^\phi, \forall i, j \in \mathbb{N}^*, i \leq |\mathcal{D}_x^\phi|, j \leq |\mathcal{T}_x^\phi|, l\,(i, j) = 1\right\} \quad (2)$$

   and

$$\left\{(d_i, t_j) : d_i \in \mathcal{D}_{\overline{x}}^\phi, t_j \in \mathcal{T}_{\overline{x}}^\phi, \forall i, j \in \mathbb{N}^*, i \leq |\mathcal{D}_{\overline{x}}^\phi|, j \leq |\mathcal{T}_{\overline{x}}^\phi|, l\,(i, j) = 1\right\}. \quad (3)$$

   In equations 2 and 3, boolean function l is defined as

$$l\,(i, j) = \begin{cases} 1 \; if \text{ the interaction between drug } d_i \text{ and target } t_j \text{ is listed in DrugBank} \\ 0 \text{ otherwise} \end{cases}. \quad (4)$$

Molecular docking synthetic results

Figure 7 presents the summary of interactions resulted from the molecular docking analysis of the drug-target pairs generated with equations 1, 2, and 3. For reference drugs $\mathcal{D}_r^{\text{antifungal}} = \{$Clotrimazole, Oxiconazole, Naftifine, Tolnaftate, Nystatin, Natamycine, Ciclopirox, Griseofulvin$\}$ the interaction strength is represented as the number of amino acids in the target interacting with the drug molecule (the maximum number in our molecular docking experiments is 24). Because Ergosterol $\in \mathcal{T}_{\overline{25}}^{\text{antifungal}}$ has a steroidal chemical structure, instead of the number of amino acids, we represent interaction strength as the number of hydrophobic alkyl/alkyl interactions. For the tested drugs $\mathcal{D}_t^{\text{antifungal}} = \{$Meprobamate, Fosinopril, Furosemide$\}$, we represent the interaction strength as the number of amino acids from the target (or hydrophobic alkyl/alkyl interactions for Ergosterol) interacting *in the same way* with both the tested drug ($\in \mathcal{D}_t^{\text{antifungal}}$) and at least one reference drug ($\in \mathcal{D}_{25}^{\text{antifungal}}$).

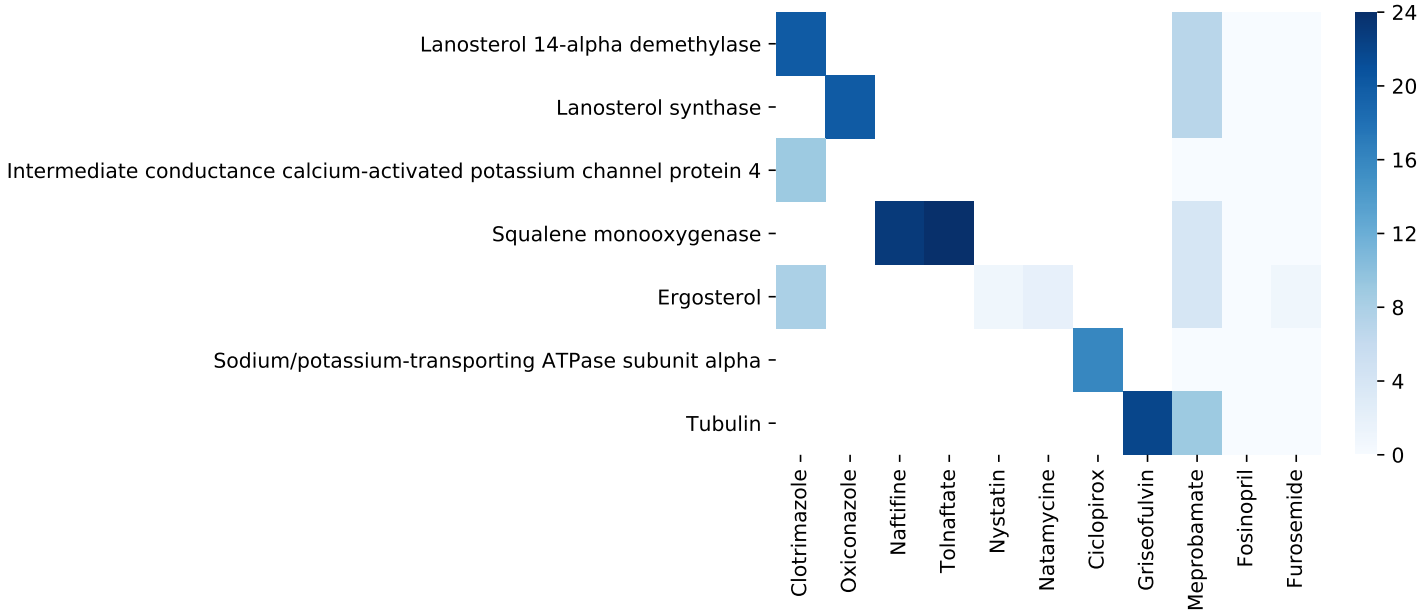


Fig 7. The strength of interactions resulted from running molecular docking on the drug-target pairs. In the left part of the heatmap, we present the interactions between the relevant targets $\mathcal{T}^{\text{antifungal}} = \{$Lanosterol 14-alpha demethylase, Lanosterol synthase, Intermediate conductance calcium-activated potassium channel protein 4, Squalene monooxygenase, Ergosterol, Sodium/potassium-transporting ATPase subunit alpha, Tubulin$\}$ and the reference drugs $\mathcal{D}_r^{\text{antifungal}} = \{$ Clotrimazole, Oxiconazole, Naftifine, Tolnaftate, Nystatin, Natamycine, Ciclopirox, Griseofulvin $\}$. We only test the reference drugs and targets pairs that interact according to DrugBank; all the other pairs are white in our representation because they are not tested. In the right part of the heatmap, we present the interactions between the relevant targets $\mathcal{T}^{\text{antifungal}}$ and the tested drugs $\mathcal{D}_t^{\text{antifungal}} = \{$ Meprobamate, Fosinopril, Furosemide $\}$). For the reference drugs, $\mathcal{D}_r^{\text{antifungal}}$, we measure the strength of interactions with the targets $\mathcal{T}^{\text{antifungal}}$ as the number of amino acids from the target interacting with the drug molecule (from 0 to the maximum number in our experiments, namely 24). In the case of Ergosterol$\in \mathcal{T}_{\overline{25}}^{\text{antifungal}}$, instead of the number of amino acids, we count the number of hydrophobic alkyl/alkyl interactions because this target has a steroidal chemical structure. In the right part of Figure 7 heatmap (corresponding to tested drugs $\mathcal{D}_t^{\text{antifungal}}$, the interaction strength is the number of amino acids from the target (or hydrophobic alkyl/alkyl interactions for Ergosterol) interacting with a $\mathcal{D}_{25}^{\text{antifungal}}$ or $\mathcal{D}_{\overline{25}}^{\text{antifungal}}$ drug and also interacting (in the same way) with the tested drug from $\mathcal{D}_t^{\text{antifungal}}$. The heatmap representation indicates a consistent interaction between $d_h^{\text{antifungal}}$ (i.e., Meprobamate) and almost all the targets from both $\mathcal{T}_{25}^{\text{antifungal}}$ and $\mathcal{T}_{\overline{25}}^{\text{antifungal}}$. For the drugs in $\mathcal{D}_n^{\text{antifungal}}$, there is no relevant interaction with any target from $\mathcal{T}_{25}^{\text{antifungal}} \bigcup \mathcal{T}_{\overline{25}}^{\text{antifungal}}$.

The results confirm a consistent interaction between $d_h^{\text{antifungal}}$ (i.e., Meprobamate) and almost all the targets from both $\mathcal{T}_{25}^{\text{antifungal}}$ and $\mathcal{T}_{\overline{25}}^{\text{antifungal}}$. Conversely, for the drugs in $\mathcal{D}_n^{\text{antifungal}}$, there is no relevant interaction with any target from $\mathcal{T}_{25}^{\text{antifungal}} \bigcup \mathcal{T}_{\overline{25}}^{\text{antifungal}}$.

In-depth analysis

For an in-depth analysis of our docking simulations, we provide the 2D and 3D diagrams for the best docking drug-target conformations in the Supplementary S3 File. For the complexes with Ergosterol, we generate 3D maps, because this target has a non-protein (i.e., steroidal) structure.

The free energy of the complex between Meprobamate and Lanosterol 14-alpha demethylase is -2.77 kcal/mol, whereas that of the complex between Clotrimazole and Lanosterol 14-alpha demethylase is -7.15 kcal/mol, showing a higher stability for the Clotrimazole complex. However, Clotrimazole and Meprobamate similarly bind to the target, as both drugs interact with the same 10 amino acids in the target (7 out of 10 interactions are of the same type). Conversely, Fosinopril and Furosemide interact with none of these 10 amino acids in the target. Figure 8 shows the molecular interaction of (a) Clotrimazole and (b) Meprobamate with Lanosterol 14-alpha demethylase. The results indicate a similarity between Clotrimazole (i.e., a reference drug $\in \mathcal{D}_{25}^{\text{antifungal}}$) and $d_h^{\text{antifungal}}$ = Meprobamate. On the other hand, we notice a clear difference between the $\mathcal{D}_n^{\text{antifungal}}$ = {Fosinopril, Furosemide} and Clotrimazole in terms of interaction at the active binding site of Lanosterol 14-alpha demethylase.

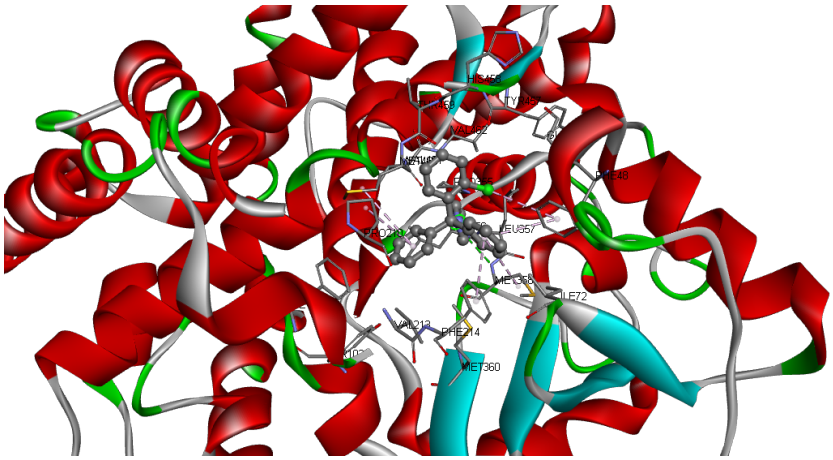
The lowest free energy of the Meprobamate-Lanosterol synthase complex is -3.23 kcal/mol, whereas the Oxiconazole-Lanosterol synthase complex is more stable than the Meprobamate-Lanosterol synthase (-6.24 kcal/mol). We notice that Oxiconazole and Meprobamate interact similarly with Lanosterol synthase through a common set of 11 amino acids, interactions of which 7 are of the same type. Fosinopril and Furosemide interact with none of the 11 amino acids in the active site of the target. Again, we notice a difference between the $\mathcal{D}_n^{\text{antifungal}}$ = {Fosinopril, Furosemide} and Oxiconazole regarding the interaction at the active binding site of Lanosterol synthase.

The molecular docking results reveal that the complex Meprobamate – Intermediate conductance calcium-activated potassium channel protein 4 has the lowest free energy of -1.02 kcal/mol. The complex between the reference antifungal drug Clotrimazole and the same target has the lowest free energy -3.60 kcal/mol. Examining the drug-target interaction, we notice that Meprobamate, Fosinopril, and Furosemide are different from Clotrimazole in the way they interact with the amino acids in the target active site.

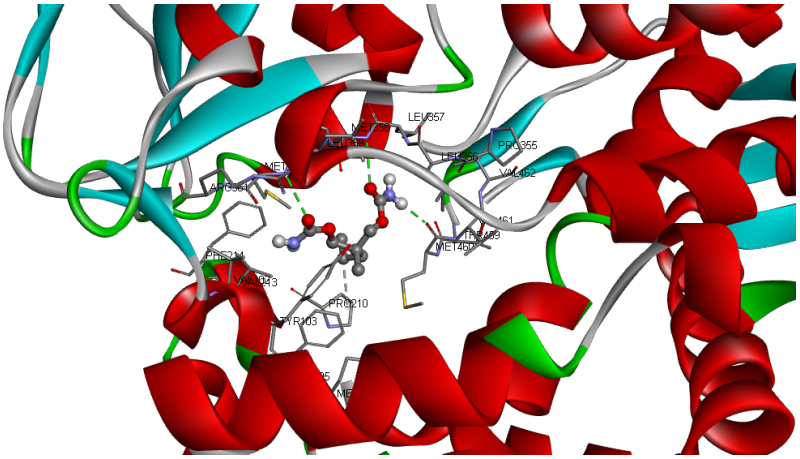
For the Meprobamate – Squalene monooxygenase complex, the lowest free energy is -2.67 kcal/mol, while for the complexes of Tolnaftate and Naftifine with Squalene monooxygenase are -6.75 kcal/mol and 7.47 kcal/mol, respectively; this means that the reference drug-target complexes are more stable than the Meprobamate complex. Meprobamate and Tolnaftate interact with the same 15 amino acids in the target (of which 5 are of the same type). Concurrently, the interaction of both Meprobamate and Naftifine with Squalene monooxygenase shares only 3 amino acids, of which 2 are of the same type. Fosinopril and Furosemide interact with no amino acid in the active site of the target. According to our docking results, Meprobamate's behavior in terms of binding to the target Squalene monooxygenase is more similar to Tolnaftate than to Naftifine.

The interaction with the (non-protein) target Ergosterol is similar for Meprobamate (the repositioning hint) and Clotrimazole ($\in \mathcal{D}_{25}^{\text{antifungal}}$), through weak hydrophobic interactions with the Ergosterol six-atoms cycles. The Clotrimazole-Ergosterol complex (-4.06 kcal/mol) is more stable than the Meprobamate-Ergosterol complex (-3.48

kcal/mol). Of note, the reference antifungal drugs Nystatin and Natamycine ($\in \mathcal{D}_{\overline{25}}^{\text{antifungal}}$) interact differently with Ergosterol, as they form hydrogen bonds with the target's hydroxyl group, while Fosinopril and Furosemide establish hydrophobic interactions with the five-atom cycle of Ergosterol.



(a)



(b)

Fig 8. Molecular interactions analysis of Meprobamate (the repositioning hint) and Clotrimazole (the reference antifungal drug) with target Lanosterol 14-alpha demethylase. (a) The docked complex of Lanosterol 14-alpha demethylase with Clotrimazole, emphasizing the molecular interactions of Meprobamate towards the active site of Lanosterol 14-alpha demethylase; (b) the docked complex of Lanosterol 14-alpha demethylase with Meprobamate, emphasizing the molecular interactions of Meprobamate towards the active site of Lanosterol 14-alpha demethylase. The green dashed lines represent conventional and carbon hydrogen bonds, and the pink dashed lines represent the alkyl and pi-alkyl interactions (a); the green dashed lines represent conventional hydrogen bonds, and the pink dashed lines represent the alkyl interactions (b). The flat ribbon represents the protein target, from which the interacting amino acid residues append. The figures display the drug molecules (figured as ball-and-stick models) in the binding pocket of the target.

The molecular docking of Meprobamate with Sodium/potassium-transporting ATPase subunit alpha, as well as of Ciclopirox with the same target, reveals that the former complex has lower stability (-2.47 kcal/mol) than the latter (-5.38 kcal/mol). Furthermore, Meprobamate and Ciclopirox do not interact with common amino acids in the active site of the target. Fosinopril and Furosemide do not interact with the amino acids in the active site of the target.

The lowest free energy of the Meprobamate – Tubulin complex is -2.99 kcal/mol, and of the Griseofulvin – Tubulin complex is -6.14 kcal/mol, indicating that the Griseofulvin complex has much higher stability than that of Meprobamate. Griseofulvin and Meprobamate similarly bind to the target, as both drugs interact with the same 15 amino acids (9 out of 15 interactions are of the same type). Fosinopril and Furosemide interact with none of the 15 amino acids in the active site of the target.

Discussion

Drug repurposing represents a promising strategy to accelerate drug discovery in sensitive areas of nowadays medicine, such as antibacterial resistance, complex life-threatening diseases (e.g., cancer), or rare diseases. In this paper, we describe a novel weighted drug-drug similarity network whose weights encode the existing known relationships among drugs (i.e., quantifies the number of biological targets shared by two drugs irrespective of the agonist or antagonist effect).

Then, we demonstrate that the ratio between node betweenness and node degree (i.e., a criterion of combined network metrics) can indicate the drug repositioning candidates better than considering simple network metrics (e.g., degree, weighted degree, betweenness). Indeed, the power-law distributions in Figure 4 suggest that our DDSN is a complex system; thus, the conventional statistical analysis of the DDSN can be misleading. Consequently, we introduce a different approach to deciphering the emerging hidden higher-order functional interactions (i.e., interactions that span multiple orders of magnitude and involve multiple nodes) by visualizing and analyzing the community structure in DDSN and determining the culprits (for such unknown functionalities) through combined network metrics criterion. We use the force-directed energy layout Force Atlas 2 to generate network clusters of drugs [32] because it emulates the emerging processes of a complex system. More precisely, the force-directed based network layouts use micro-scale interactions (i.e., adjacent nodes attract and non-adjacent nodes repulse) to generate an emergent behavior at the macro-scale (i.e., topological clusters). Once we identify communities, the combined network metrics criterion selects the drug repositioning most likely candidates. Specifically, our weighted drug-drug network analysis encodes not only information about how pairs of drugs interact with biological targets, but also reveals the unknown functional relationship between drugs, such as the unknown effects on the activation/inhibition of a chemical pathway or cellular behavior. We used a similar methodology – underpinned by force-directed layout clustering – to analyze drug-drug interaction networks (i.e., the DDIN interactome [40, 41]).

As presented in the literature, other computational methods allow for predicting new targets for existing drugs and are currently applied for drug repositioning or explaining off-target effects [42]. Indeed, the computer-assisted methods for predicting alternative targets or similar binding sites are rapidly evolving by making use of the enormous amount of drug data [43]. For instance, Mayr et al. found that the methods based on deep learning are much better than all other in silico approaches, sometimes matching the performance of wet-lab experiments [28].

Molecular docking represents an alternative, in silico simulation approach to drug discovery, which models the physical interaction between a drug molecule and a target

(or a set of targets). With molecular docking, we estimate the free energy values of the molecular interactions to offer a good approximation for the conformation and orientation of the ligand into the protein cavity [44]. Along with many available molecular docking models, DOCK [45] is a dedicated software tool used in drug repurposing. For example, R. L. Des Jarlais et al. used the Dock computer algorithm to find that haloperidol inhibits HIV-1 and HIV-2 proteases [46]. However, molecular docking cannot work efficiently, unless we have some robust repositioning hints; otherwise, the search space for drug repositionings would be exponentially big. To this end, the methodology proposed in this paper provides strong drug-target interaction hints, such that we can build large-scale drug-target interaction profiles [47]. Our approach integrates the molecular docking with complex networks to uncover new pharmacological properties by identifying new sets of biological targets on which the drug acts.

The old paradigm of chemical similarity, which holds that structurally similar drug molecules exert similar biological effects, is challenged by the results reported of Yvonne Martin et al.; they found that only 30% of compounds similar to a particular active compound are themselves active (the compounds are structurally similar if the Tanimoto coefficient is ≥ 0.85 in the Daylight fingerprints) [48]. On the other hand, similar interaction profiles are valuable resources in drug repurposing, as drugs with similar target binding patterns may exhibit a similar pharmacologic activity [47, 49, 50].

The drug we identify as a strong repurposing hint, Meprobamate, is a carbamate derivative. However, the antifungal reference drugs are not structurally similar to Meprobamate: Clotrimazole and Oxiconazole are imidazole derivatives, Ergosterol has a steroidal structure, Terbinafine and Naftifine are allylamine compounds, and Griseofulvin is a 3-coumaranone derivative. Nonetheless, as the chemical similarity is not necessarily a reliable predictor of biological similarity [43, 48], we analyze the binding mode of Meprobamate in comparison with the other known reference drugs.

As shown in our molecular docking analysis, Meprobamate has similar binding modes to that of Clotrimazole with Lanosterol 14 alpha-demethylase, Oxiconazole with Lanosterol synthase, and Griseofulvin with Tubulin. The similarity analysis quantifies the number of distinct amino acids in the target that interact in the same way with the drug molecule. On the other hand, the predicted lowest free energy and the inhibition constant for the Meprobamate-Ergosterol complex (-3.48 kcal/mol; 2.79 mM) are closest to that of Clotrimazole and Nystatin complexes with Ergosterol (-4.06 kcal/mol, 1.05 mM, and -3.98 kcal/mol, 1.21 mM, respectively). Indeed, Clotrimazole is a broad-spectrum antifungal agent with a complex mechanism of action, which inhibits the Lanosterol 14-alpha demethylase target in yeast, as well as Ergosterol target in Candida albicans [51]; therefore, its functional similarity to Meprobamate is probable.

Meprobamate is a known oral drug. However, when considering its potential antifungal activity, we cannot exclude the topical route of administration. To this end, we suggest that further investigations on biopharmaceutical properties (e.g., solubility, lipophilicity, octanol/water partition coefficient) are required.

Materials and methods

Databases

We build our Drug-Drug Similarity Network (DDSN) using drug-target interaction information from the older Drug Bank version 4.2 [31], such that we can use the latest Drug Bank 5.1.4 [51] for testing the accuracy of our drug property prediction.

Network analysis

Our Drug-Drug Similarity Network (DDSN) is an undirected weighted graph $G = (V, E)$, where V is the vertex (or node) set, and E is the edge (or link) set. As such, we have $|V|$ vertices $v_i \in V$ and $|E|$ edges $e_{j,k} \in E$, with $i, j, k \in \{1, 2, \ldots |V|\}$ and $j \neq k$. Each edge $e_{j,k}$ is characterized by a weight $w(e_{j,k}) \neq 0$ (in our DDSN, $w(e_{j,k}) \in \mathbb{N}^*$, $\forall e_{j,k} \in E$). In an unweighted network, $w(e_{j,k}) = 1$, $\forall e_{j,k} \in E$.

Network centralities

Node centralities are complex network parameters, which are characterizing the importance of a vertex/node in a graph [52]. In our analysis, we used the weighted degree, degree, betweenness, and betweenness/degree node centralities.

The weighted degree of a node v_i is the sum of the weights characterizing the links/edges incident to v_i,

$$d(v_i) = \sum_{j \in \{x | e_{i,x} \in E, v_x, v_i \in V\}} w(e_{i,j}). \tag{5}$$

We compute the degree of a node v_i with equation 5, assuming that $w(e_{i,j}) = 1$, $\forall e_{ij} \in E$.

To compute the node betweenness, we must find the shortest paths between all node pairs (v_j, v_k) in graph G, namely $\sigma_{j,k}$. As such, the betweenness of node v_i is the number of minimal paths in graph G that cross node v_i, divided by the total number of minimal paths in G,

$$b(v_i) = \sum_{(j,k) \in \{(x,y) | v_x \neq v_y \neq v_i; v_x, v_y, v_i \in V\}} \frac{\sigma_{j,k}(v_i)}{\sigma_G}, \tag{6}$$

where the total number of shortest paths in G is $\sigma_G = \binom{|V|}{2}$.

The betweenness/degree of node v_i is the ratio

$$b/d(v_i) = \frac{b(v_i)}{d(v_i)}, \tag{7}$$

where equation 5 computes $d(v_i)$ in the unweighted version (i.e., considering $w(e_{i,j}) = 1$, $\forall e_{ij} \in E$).

Community detection

The network layout algorithm we use in this paper places each vertex v_i in a 2D space $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. Therefore, each node $v_i \in V$ has its 2D coordinates $\gamma_i = (x_i, y_i) \in \mathbb{R}^2$, and each edge $e_{i,j} \in E$ has a Euclidian distance $\delta_{i,j} = |\gamma_i - \gamma_j|$.

In an energy-model, force-directed layout, we have a force of attraction between any two adjacent nodes v_i and v_j and a force of repulsion between any two non-adjacent nodes. The expression of these forces is $|\gamma_i - \gamma_j|^f \cdot \overrightarrow{\gamma_i \gamma_j}$, where $f = a$ for attraction and $f = r$ for repulsion. In an energy-model, force-directed layout, we have a force of attraction between any two adjacent nodes v_i and v_j, and a repulsion force between any two non-adjacent nodes. The expression of these forces is $|\gamma_i - \gamma_j|^f \overrightarrow{\gamma_i \gamma_j}$, where $f = a$ for attraction and $f = r$ for repulsion. The attraction force between adjacent nodes (v_i and v_j such that $\exists e_{i,j} \in E$) decreases, whereas the repulsion force between non-adjacent nodes (v_i, v_j such that $\exists! e_{i,j} \in E$) increases with the Euclidian distance. Therefore, we must have $a \geq 0$ and $r \leq 0$.

In this paper, we use the energy-model force-directed layout Force Atlas 2 [32] to assign node positions in the 2D (i.e., $\mathbb{R}^2$) space, based on interactions between attraction and repulsion forces, such that we attain minimal energy in the network layout,

$$\mathcal{E} = \min \left\{ \sum_{(v_i,v_j),i\neq j} \left(\frac{|\gamma_i - \gamma_j|^a}{a+1} - \frac{|\gamma_i - \gamma_j|^r}{r+1} \right) \right\} \tag{8}$$

.

The energy-based layouts generate topological communities because there are specific regions in the network with larger than average link densities. Noack [35] demonstrated that the energy-based topological communities are equivalent to the network clusters based on modularity classes [33], when $a > -1$ and $r > -1$.

The network clustering classifies each node $v_i \in V$ in one of the disjoint sets of nodes (cluster $C_i \subset V$, with $i = \overline{1..m}$, $C_1 \bigcup C_2 \ldots \bigcup C_m = V$). In [33], the authors use modularity to define the node membership to one of the clusters. To this end, the modularity of a clustering $\mathcal{C}_m = \{C_1, C_2, \ldots C_m\}$ is

$$\mathcal{M}_m = \sum_{C_i \in \mathcal{C}_m} \left(\frac{|E_{C_i}|}{|E|} - \frac{\frac{1}{2}{d_{C_i}}^2}{\frac{1}{2}d^2} \right) \tag{9}$$

where $|E|$ is the total number of edges in G, $|E_{C_i}|$ is the total number of edges between nodes in cluster C_i, d is the total degree of nodes in G, and d_{C_i} is the total degree of nodesm in cluster C_i. Thus, $\frac{|E_{C_i}|}{|E|}$ represents the relative edge density of cluster C_i relative to the density of the entire network G, whereas $\frac{\frac{1}{2}{d_{C_i}}^2}{\frac{1}{2}d^2}$ is the expected relative density of C_i.

Molecular docking for repurposing validation

Ligands preparation

We generate the three-dimensional coordinates of all ligands using the Gaussian program suite with the DFT/B3LYP/6-311G optimization procedure.

Targets preparation

We get the X-ray crystal structure of the targets as target.pdb files from the major protein databases Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) and optimize them with the ModRefiner software (https://zhanglab.ccmb.med.umich.edu/ModRefiner/). The targets and their corresponding codes are Lanosterol 14-alpha demethylase (4LXJ code, resolution 1.9 A), Intermediate conductance calcium-activated potassium channel protein 4 (6D42 code, resolution 1.75 A), Lanosterol synthase (1W6K code, resolution 2.1 A), Squalene monooxygenase (6C6N code, resolution 2.3 A), Ergosterol (2AIB code, resolution 1.1 A), Sodium/potassium-transporting ATPase subunit alpha (2ZXE code, resolution 2.4 A), and Tubulin (4U3J code, resolution 2.81). The preparation of targets also requires adding all polar hydrogens, removing the water, and computing the Gasteiger charge.

Docking protocol

We perform the molecular docking analysis using the Autodock 4.2.6 software suite together with the molecular viewer and graphical support AutoDockTools.

In the docking protocol, for the protein targets, we create the grid box using Autogrid 4 with 120 Å × 120 Å × 120 Å in x, y and z directions, and 1 Å spacing from

the target molecule's center. For steroidal target Ergosterol, the grid box is 30 Å × 30 Å × 30 Å in x, y and z directions, with 0.375 Å spacing from the target molecule's center.

For the docking process, we chose the Lamarckian genetic algorithm (Genetic Algorithm combined with a local search), with a population size of 150, a maximum of $2.5 \cdot 10^6$ energy evaluations, a gene mutation rate of 0.02, and 50 runs. We adopted the default settings for the other docking parameters and performed all the calculations in vacuum conditions. Then, we exported all AutoDock results in the PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC), and the Discovery Studio (Biovia) molecular visualization system (Dassault Systèmes BIOVIA, BIOVIA Workbook, Release 2017; BIOVIA Pipeline Pilot, Release 2017, San Diego: Dassault Systèmes, 2019).

Supporting information

S1 File. The DDSN Gephi file. The file *DDSN.gephi* contains our entire Gephi project for describing the drug-drug similarity network, including the betweenness/degree node centrality we use for prioritizing the repositioning hints (as a node ranking attribute).

S2 File. Validation references and statistics. The file *Supplementary-DDSN.xls* contains the data to support the drug cluster/community labeling. The data for each topological cluster is included in the corresponding tab, titled Topo. The syntetic data, corresponding to the entire DDSN, is included in the first tab, titled Topo 1-26.

S3 File. Supplementary molecular docking test results. The file *SupplementaryInformation.pdf* contains tables and figures describing all data resulted from our drug repurposing hints test procedure, based on molecular docking.

References

1. Dickson M, Gagnon JP. The cost of new drug discovery and development. Discovery Medicine. 2009;4(22):172–179.
2. Chen XQ, Antman MD, Gesenberg C, Gudmundsson OS. Discovery pharmaceuticschallenges and opportunities. The AAPS journal. 2006;8(2):E402–E408.
3. Mullard A. 2016 FDA drug approvals. Nature Reviews Drug Discovery. 2017;16(2):73–76.
4. Graul A, Pina P, Cruces E, Stringer M. The year's new drugs & biologics 2016: Part I. Drugs of today (Barcelona, Spain: 1998). 2017;53(1):27.
5. Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. Nature reviews Drug discovery. 2011;10(6):428–438.
6. Csermely P, Korcsmáros T, Kiss HJ, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacology & therapeutics. 2013;138(3):333–408.
7. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nature Reviews Drug Discovery. 2019;18(1):41.

8. Munos B. Lessons from 60 years of pharmaceutical innovation. Nature Reviews Drug Discovery. 2009;8(12):959–968.

9. Shaughnessy AF. Old drugs, new tricks. BMJ. 2011;342:d741.

10. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Briefings in bioinformatics. 2015;17(1):2–12.

11. Lotfi Shahreza M, Ghadiri N, Mousavi SR, Varshosaz J, Green JR. A review of network-based approaches to drug repositioning. Briefings in Bioinformatics. 2017; p. bbx017.

12. Nugent T, Plachouras V, Leidner JL. Computational drug repositioning based on side-effects mined from social media. PeerJ Computer Science. 2016;2:e46.

13. Zhao M, Yang CC. Mining Online Heterogeneous Healthcare Networks for Drug Repositioning. In: Healthcare Informatics (ICHI), 2016 IEEE International Conference on. IEEE; 2016. p. 106–112.

14. Shameer K, Readhead B, T Dudley J. Computational and experimental advances in drug repositioning for accelerated therapeutic stratification. Current topics in medicinal chemistry. 2015;15(1):5–20.

15. Yıldırım MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drugtarget network. Nature biotechnology. 2007;25(10):1119–1126.

16. Mei JP, Kwoh CK, Yang P, Li XL, Zheng J. Drug–target interaction prediction by learning from local information and neighbors. Bioinformatics. 2012;29(2):238–245.

17. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. Bioinformatics. 2014;30(20):2923–2930.

18. Giuliani S, Silva AC, Borba JV, Ramos PI, Paveley RA, Muratov EN, et al. Computationally-guided drug repurposing enables the discovery of kinase targets and inhibitors as new schistosomicidal agents. PLoS computational biology. 2018;14(10):e1006515.

19. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS computational biology. 2012;8(5).

20. Coelho ED, Arrais JP, Oliveira JL. Computational discovery of putative leads for drug repositioning through drug-target interaction prediction. PLoS computational biology. 2016;12(11).

21. Tanoli Z, Alam Z, Ianevski A, Wennerberg K, Vähä-Koskela M, Aittokallio T. Interactive visual analysis of drug–target interaction networks using drug target profiler, with applications to precision medicine and drug repurposing. Briefings in bioinformatics. 2018;.

22. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences. 2010;107(33):14621–14626.

23. Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J. Transcriptional data: a new gateway to drug repositioning? Drug discovery today. 2013;18(7-8):350–357.

24. Cheng F, Kovács IA, Barabási AL. Network-based prediction of drug combinations. Nature communications. 2019;10(1):1197.

25. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nature communications. 2018;9(1):2691.

26. Nguyen T, Le H, Venkatesh S. GraphDTA: prediction of drug–target binding affinity using graph convolutional networks. BioRxiv. 2019; p. 684662.

27. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457–i466.

28. Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chemical science. 2018;9(24):5441–5451.

29. Lo YC, Rensi SE, Torng W, Altman RB. Machine learning in chemoinformatics and drug discovery. Drug discovery today. 2018;23(8):1538–1546.

30. Liu Z, Fang H, Reagan K, Xu X, Mendrick DL, Slikker W, et al. In silico drug repositioning–what we need to know. Drug discovery today. 2013;18(3):110–115.

31. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research. 2007;36(suppl_1):D901–D906.

32. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one. 2014;9(6):e98679.

33. Girvan M, Newman ME. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002;99(12):7821–7826.

34. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015;20(7):13384–13421.

35. Noack A. Modularity clustering is force-directed layout. Physical Review E. 2009;79(2):026102.

36. Grathwohl W, Wang KC, Jacobsen JH, Duvenaud D, Norouzi M, Swersky K. Your Classifier is Secretly an Energy Based Model and You Should Treat it Like One. arXiv preprint arXiv:191203263. 2019;.

37. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661–703.

38. Alstott J, Bullmore E, Plenz D. powerlaw: a Python package for analysis of heavy-tailed distributions. PloS one. 2014;9(1):e85777.

39. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. 2009;30(16):2785–2791.

40. Udrescu L, Sbârcea L, Topîrceanu A, Iovanovici A, Kurunczi L, Bogdan P, et al. Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing. Scientific reports. 2016;6.

41. Udrescu M, Udrescu L. A Drug Repurposing Method Based on Drug-Drug Interaction Networks and Using Energy Model Layouts. In: Computational Methods for Drug Repurposing. Springer; 2019. p. 185–201.

42. Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. Nature chemical biology. 2013;9(4):232.

43. Haupt VJ, Schroeder M. Old friends in new guise: repositioning of known drugs with structural bioinformatics. Briefings in bioinformatics. 2011;12(4):312–326.

44. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. British journal of pharmacology. 2007;152(1):9–20.

45. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. Journal of computer-aided molecular design. 2001;15(5):411–428.

46. DesJarlais R, Seibel G, Kuntz I, Furth P, Alvarez J, De Montellano PO, et al. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. Proceedings of the National Academy of Sciences. 1990;87(17):6644–6648.

47. Cavalla D, Oerton E, Bender A. Drug repurposing review. 2017;.

48. Martin YC, Kofron JL, Traphagen LM. Do structurally similar molecules have similar biological activity? Journal of medicinal chemistry. 2002;45(19):4350–4358.

49. Yang L, Chen J, Shi L, Hudock MP, Wang K, He L. Identifying unexpected therapeutic targets via chemical-protein interactome. PloS one. 2010;5(3).

50. Simon Z, Peragovics Á, Vigh-Smeller M, Csukly G, Tombor L, Yang Z, et al. Drug effect prediction by polypharmacology-based interaction profiling. Journal of chemical information and modeling. 2012;52(1):134–145.

51. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research. 2017;46(D1):D1074–D1082.

52. Barabási AL, et al. Network science. Cambridge university press; 2016.

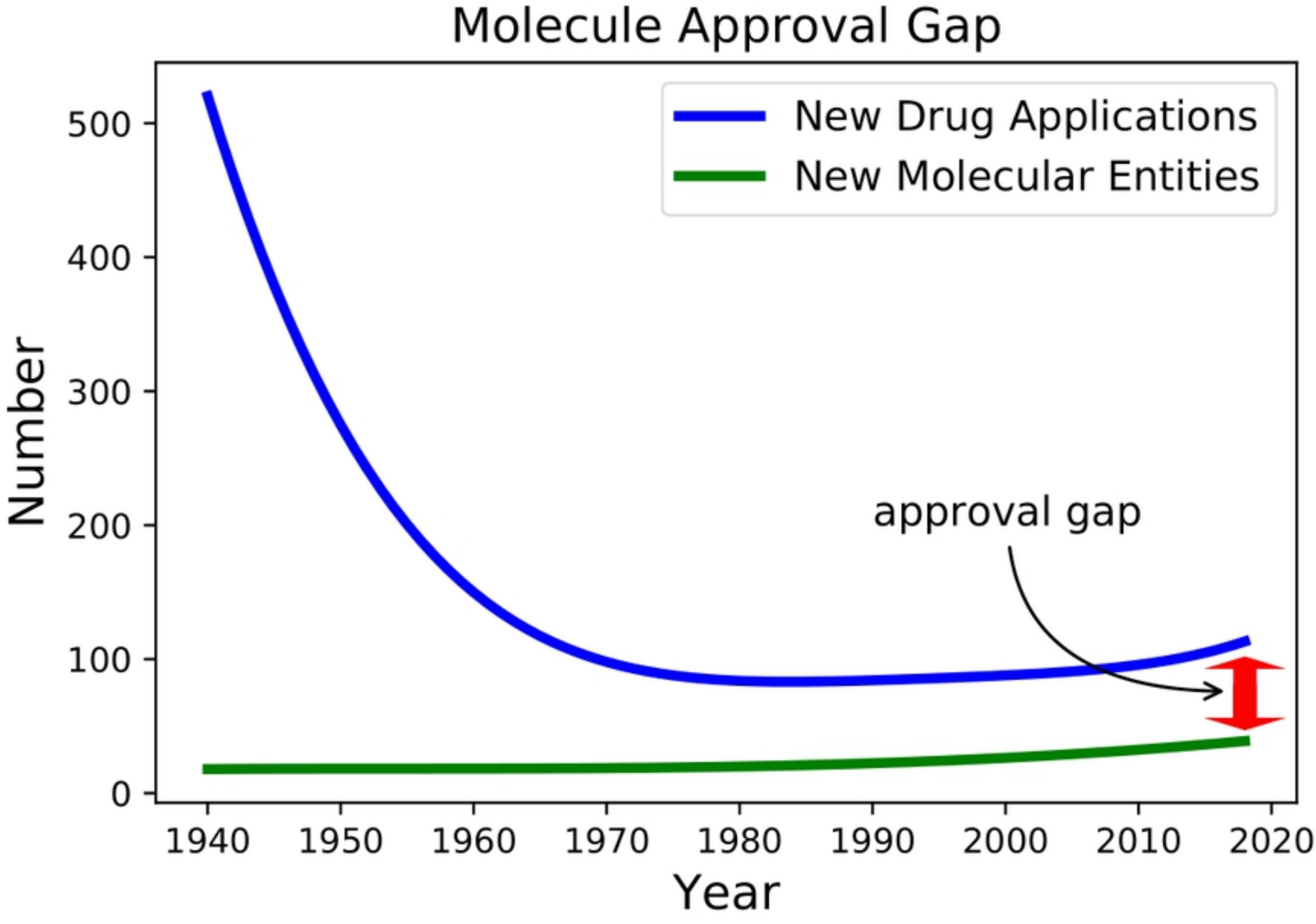
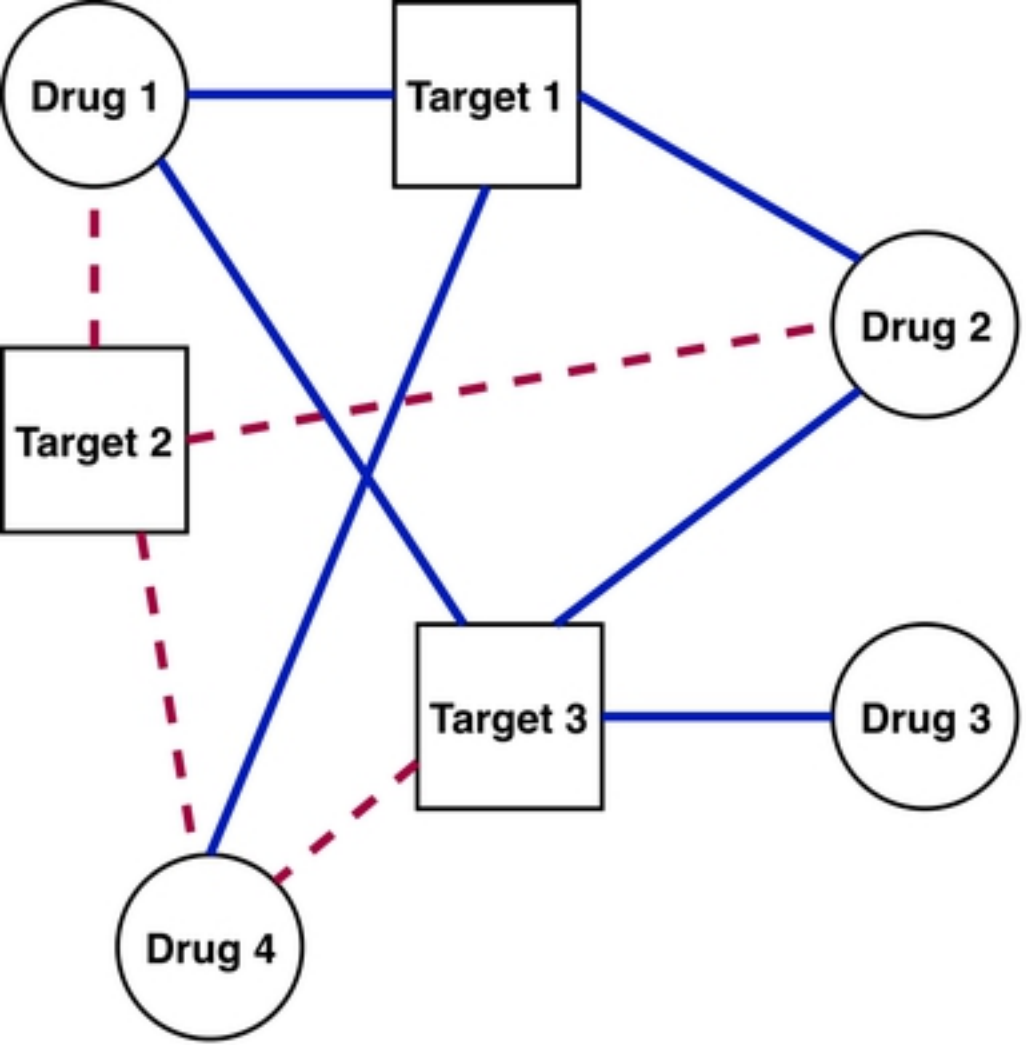

Molecule Approval Gap
New Drug Applications
New Molecular Entities
approval gap
Number
Year
0
100
200
300
400
500
1940
1950
1960
1970
1980
1990
2000
2010
2020

Figure 1


Drug 1
Target 1
Drug 2
Target 2
Target 3
Drug 3
Drug 4

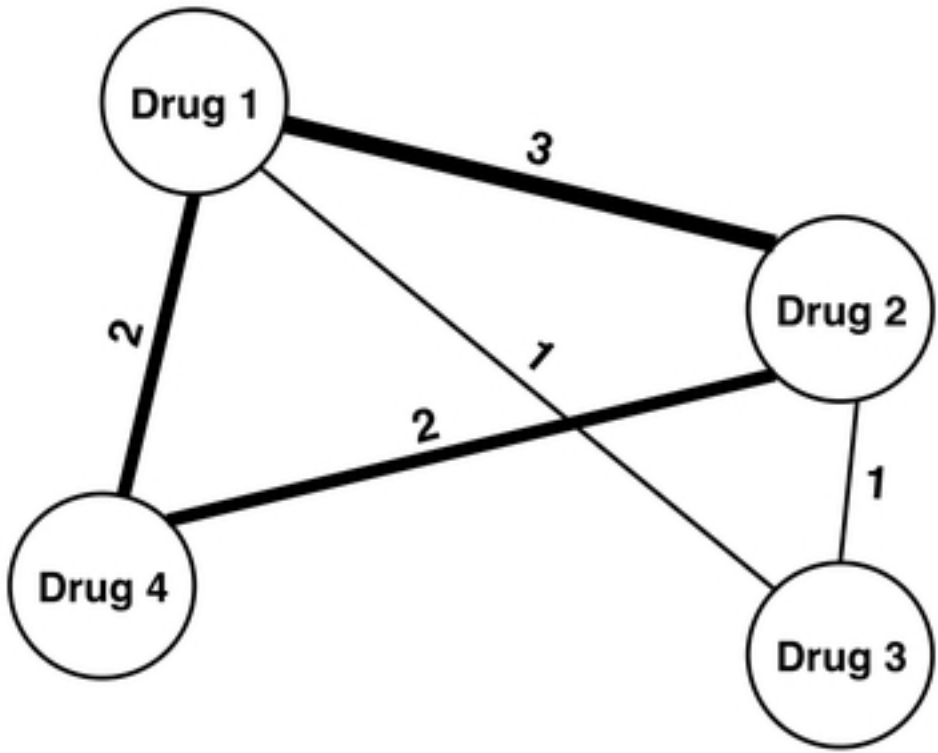

Drug 1
3
Drug 2
2
1
2
1
Drug 4
Drug 3

Figure 2 (a), (b)

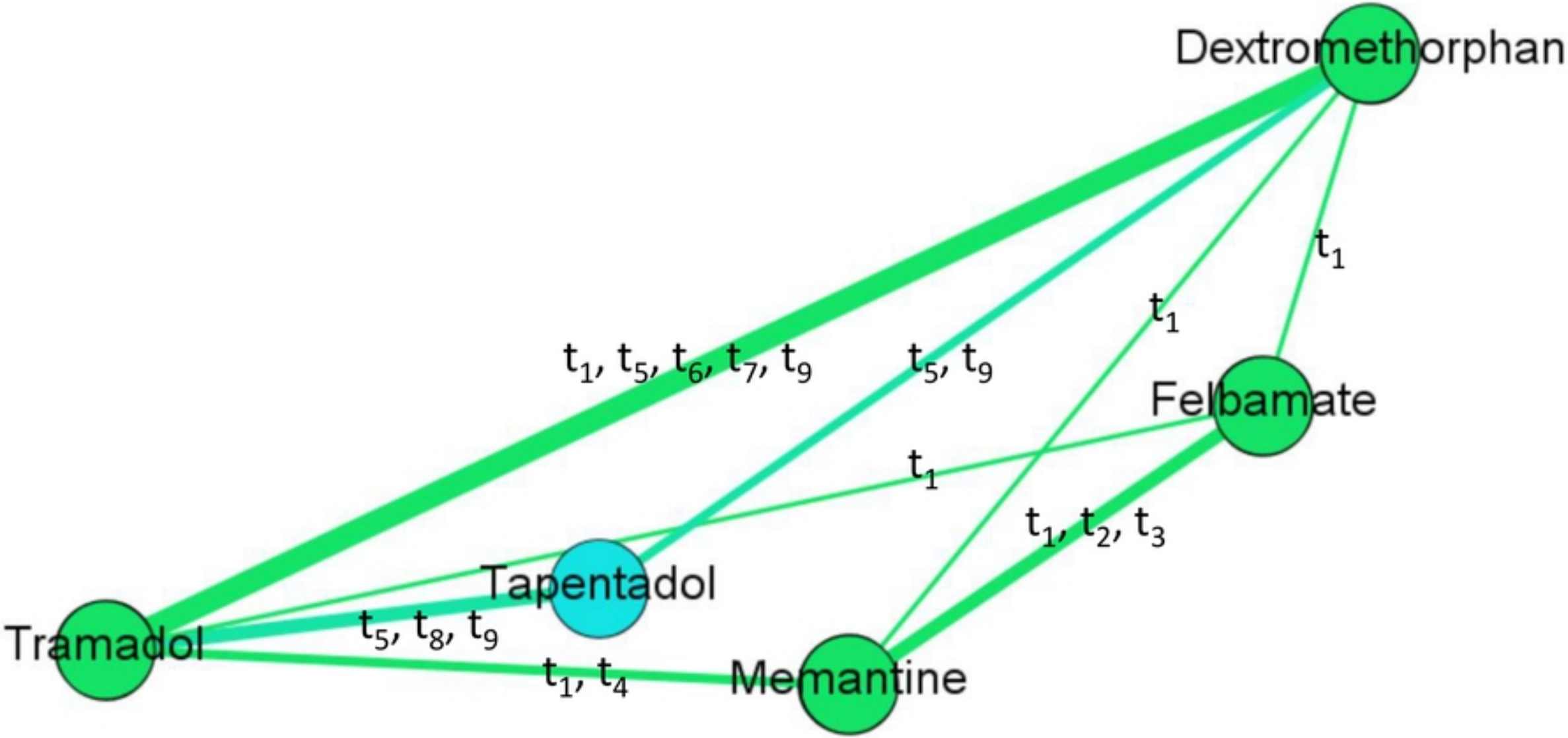
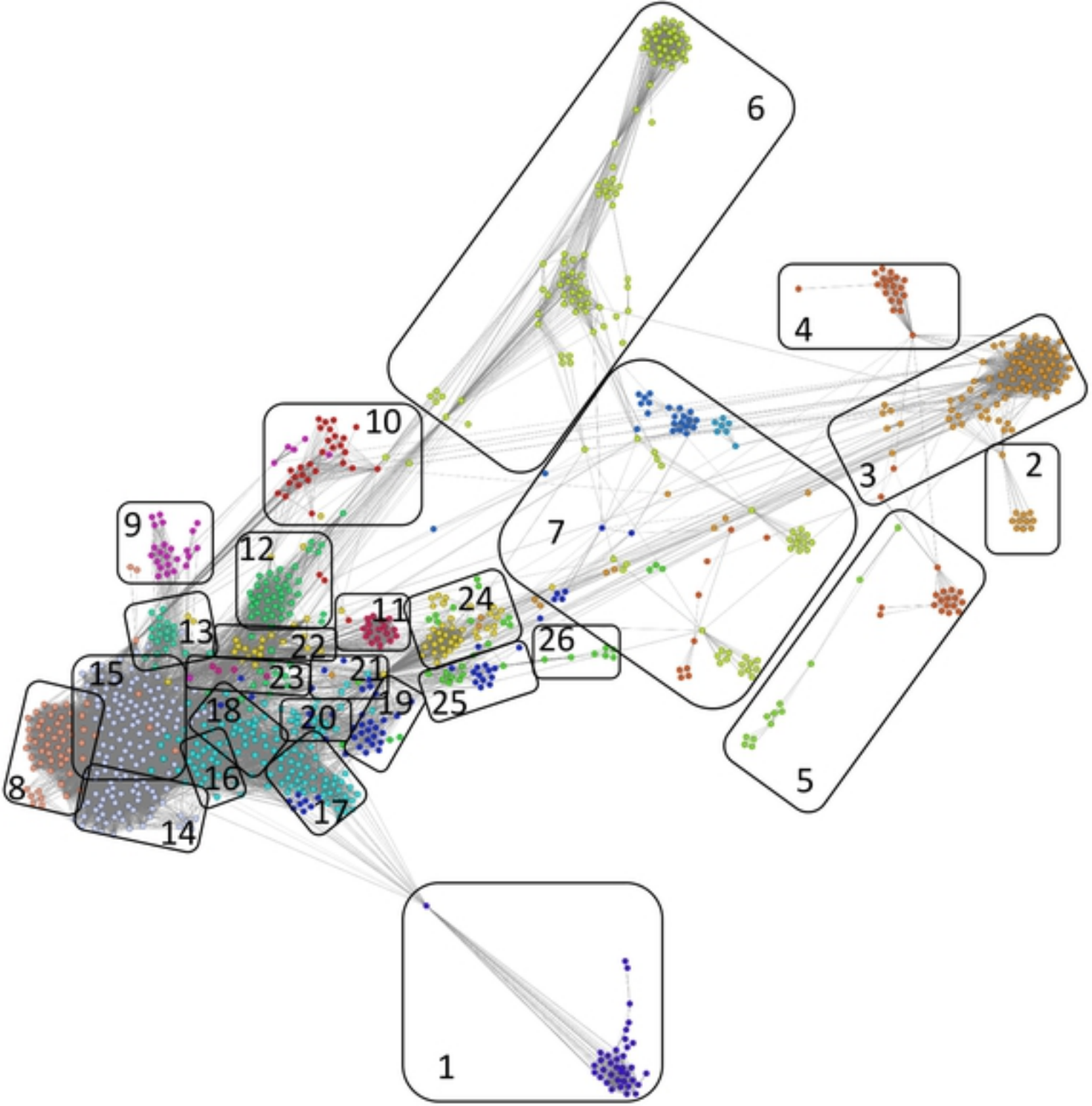

Dextromethorphan
t_1
t_1
t_1, t_5, t_6, t_7, t_9
t_5, t_9
Felbamate
t_1
t_1, t_2, t_3
Tapentadol
Tramadol
t_5, t_8, t_9
t_1, t_4
Memantine

Figure 2 (c)



1 – Antineoplastic drugs
(Mitotic inhibitors & DNA-damaging anticancer drugs)
2 – Antihypertensive agents
3 – Antiinflammatory drugs
4 – Antibacterial drugs
5 – Platelet aggregation inhibitors
6 – Drugs interfering with hormone dependent cancers
7 – Anticancer drugs
(Molecularly targeted antineoplastic drugs)
8 – Antiallergic drugs
9 – Drugs acting on muscles
10 – Vasodilator drugs
11 – Antiepileptic, hypnotic & sedative drugs
12 – Analgesic drugs
13 – Cardiovascular drugs (mostly beta-blockers)
14 – Anticholinergic drugs
15 – Drugs interfering with Parasympathetic Nervous System
16 – Antidepressants and CNS stimulants
17 – Sympathetic Nervous System acting drugs
18 – Antimigraine & antiemetic drugs
19 – Antiarrhythmics, Anticonvulsants
20 – Antidepressants, Anti-Parkinson
21 – Drugs interfering with epilepsy and blood pressure
22 – Antihypertensive and anticonvulsant agents
23 – Miscellaneous Central Nervous System Agents
(anesthetic/analgesic/muscle relaxant drugs)
24 – Drugs interfering with the potassium, sodium & calcium homeostasis
25 – Antifungal agents
26 – Hypnotic & sedative drugs

Figure 3

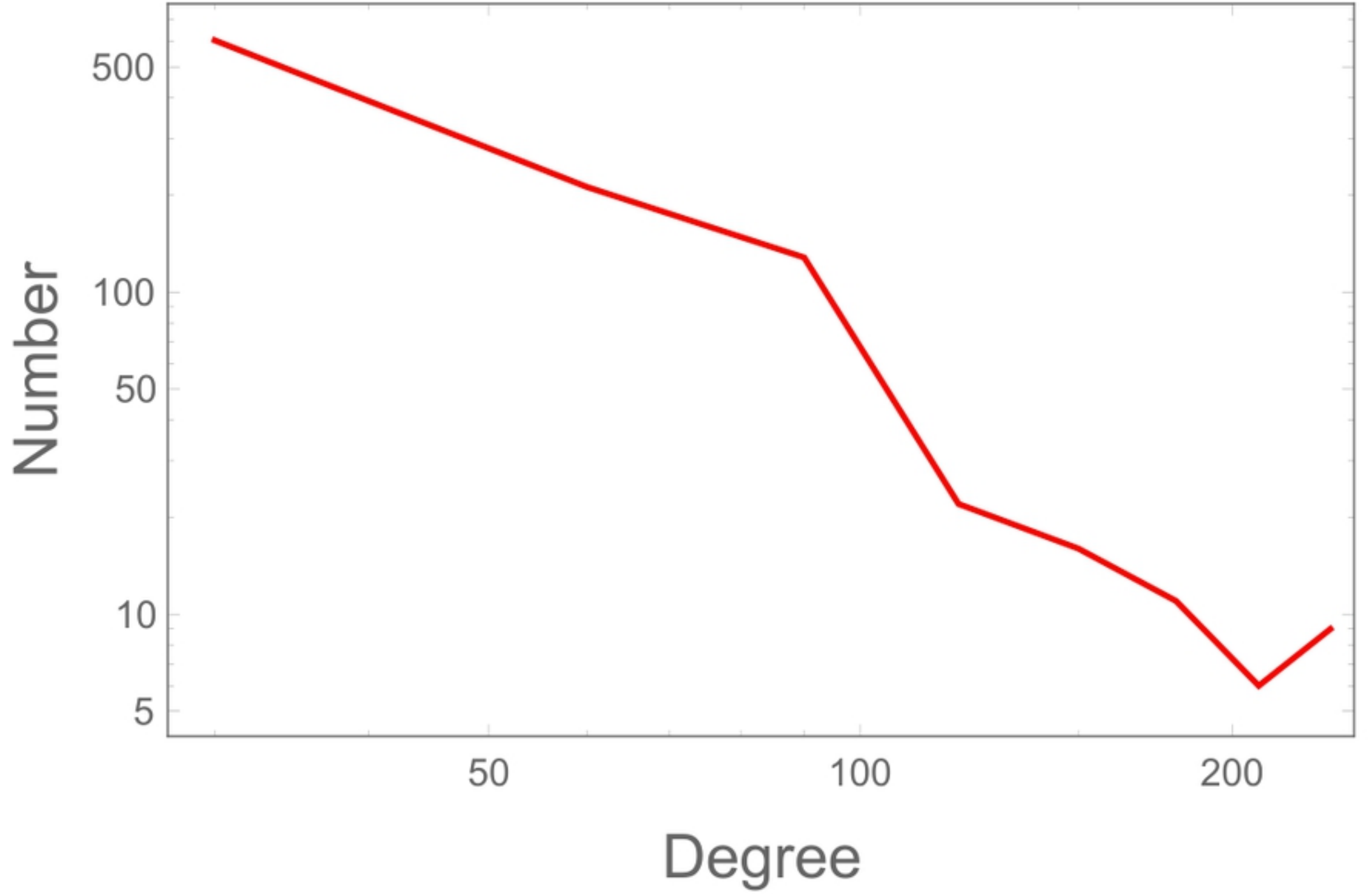


Figure 4 (a)

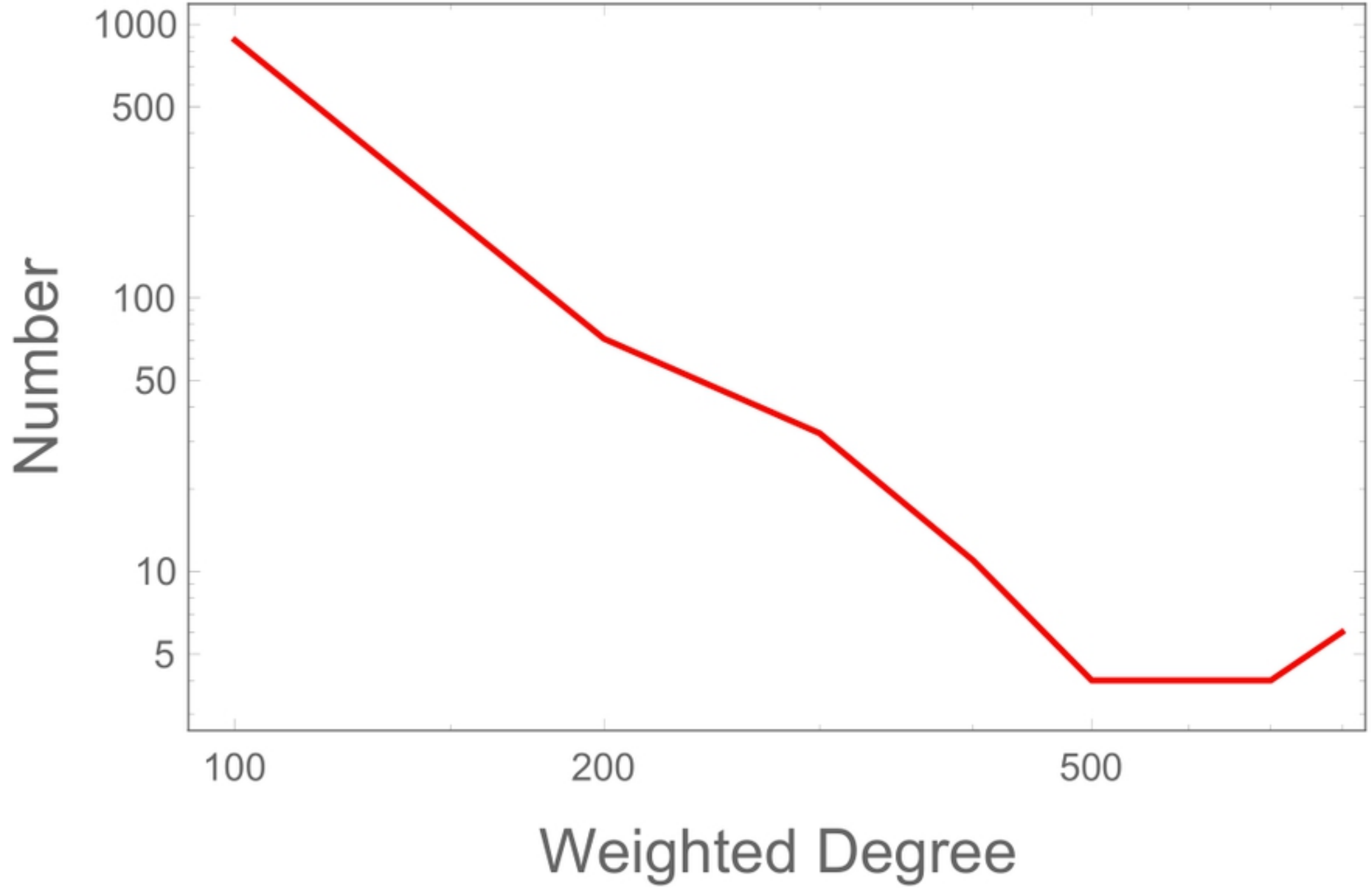


Figure 4 (b)

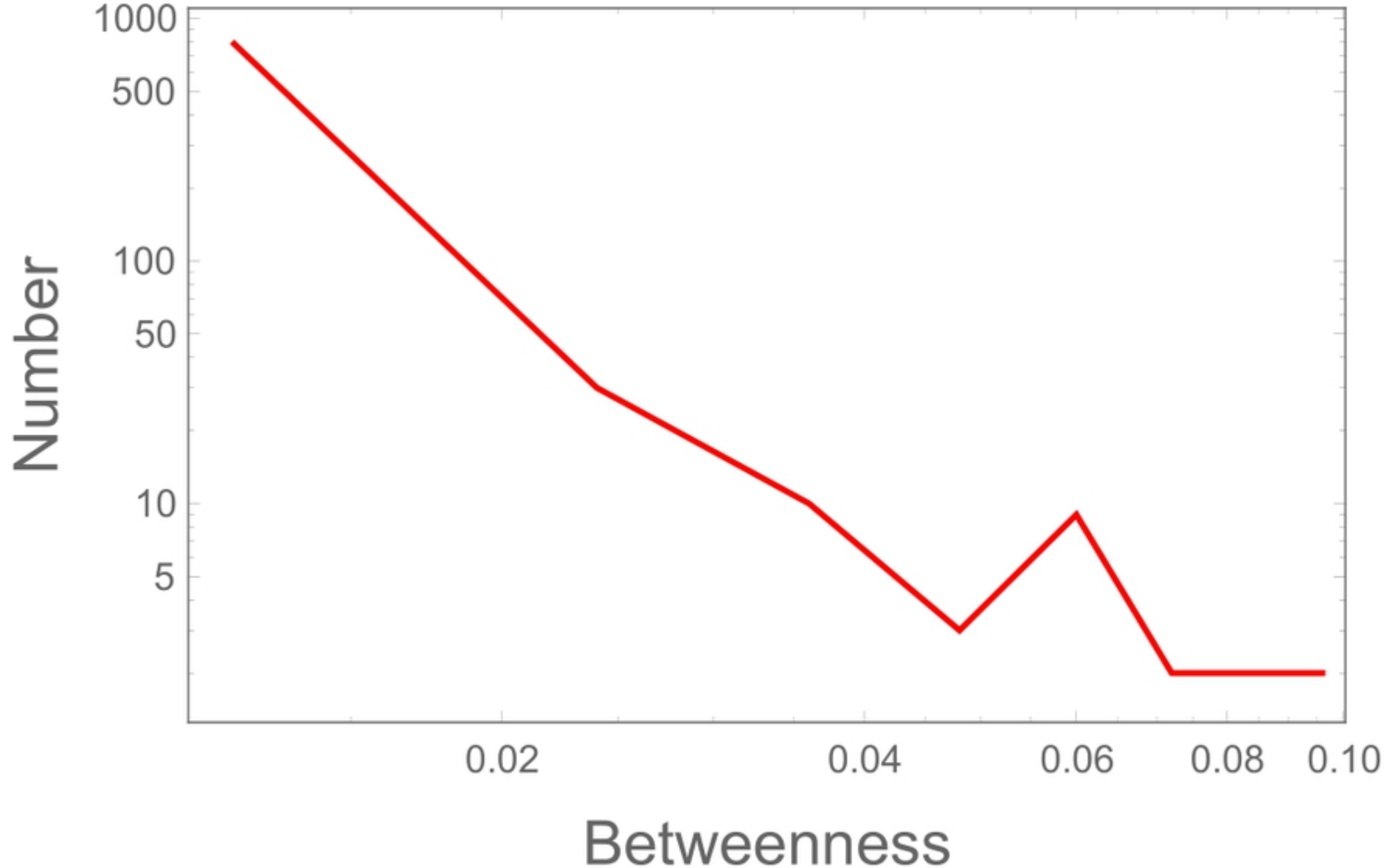


Figure 4 (c)

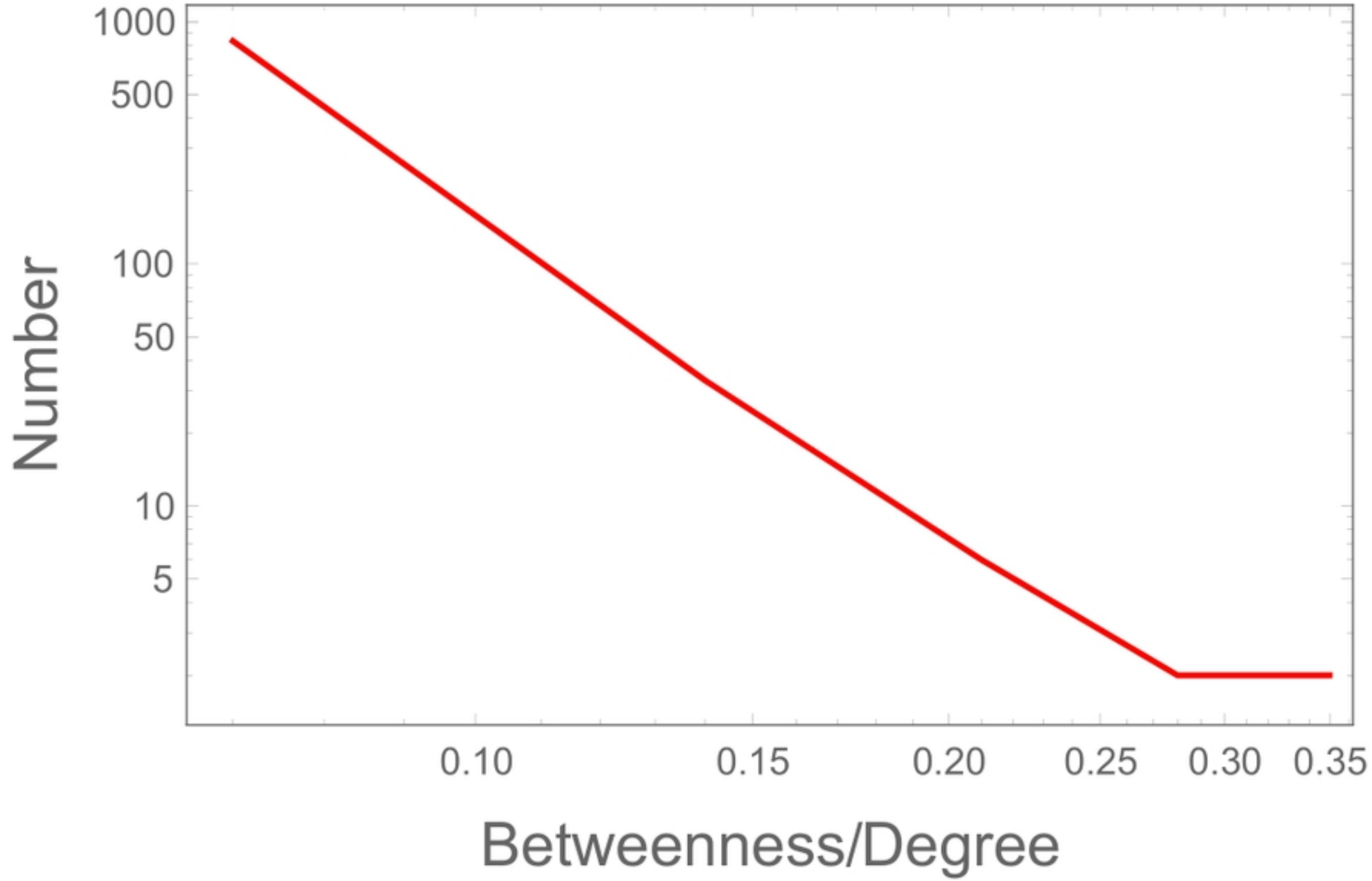

1000
500
100
50
10
5
Number
0.10
0.15
0.20
0.25
0.30
0.35
Betweenness/Degree

Figure 4 (d)

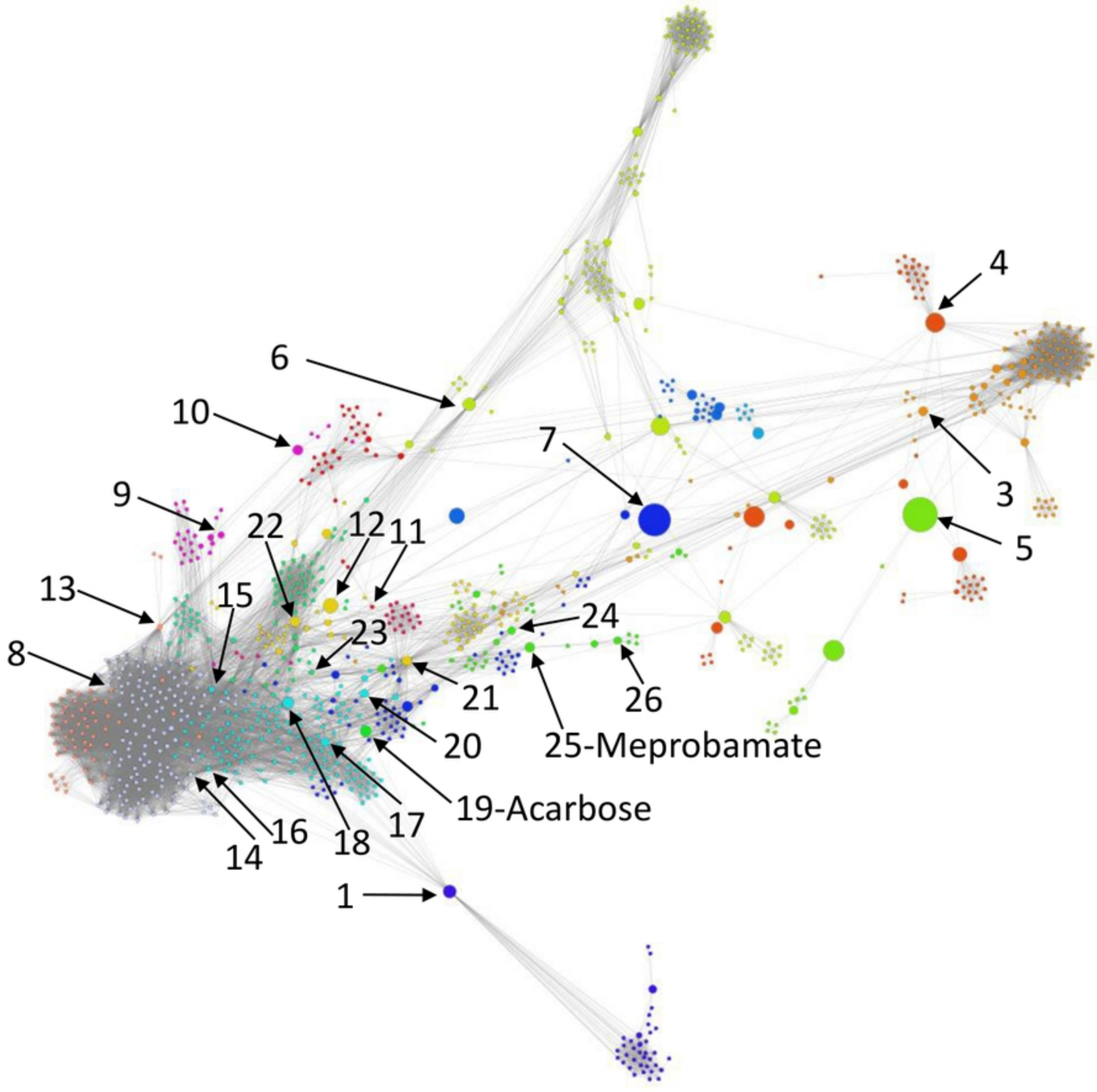

4
6
10
7
3
9
22
12
11
5
13
15
24
23
8
21
26
20
25-Meprobamate
19-Acarbose
16
18
17
14
1

Figure 5

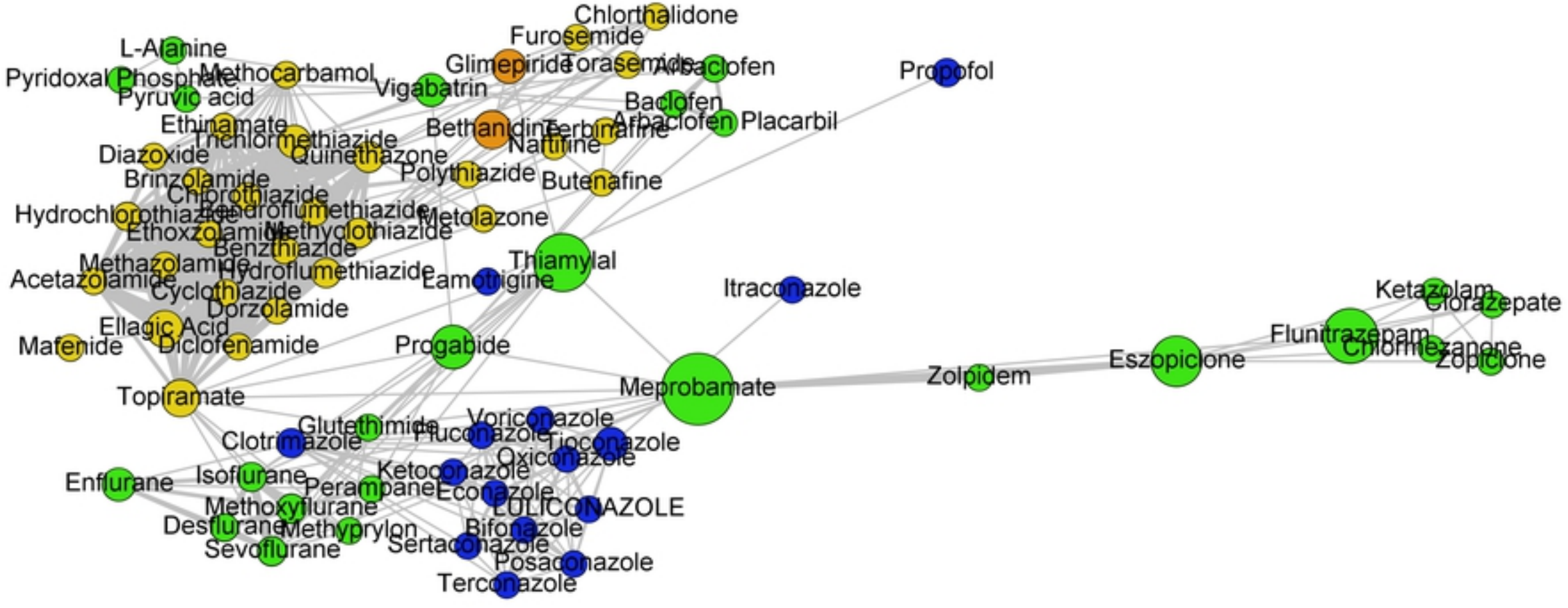

Chlorthalidone
Furosemide
L-Alanine
Pyridoxal Phosphate
Methocarbamol
Glimepiride
Torasemide
Arbaclofen
Propofol
Pyruvic acid
Vigabatrin
Baclofen
Ethinamate
Bethanidine
Terbinafine
Arbaclofen Placarbil
Trichlormethiazide
Naftifine
Diazoxide
Quinethazone
Polythiazide
Brinzolamide
Butenafine
Chlorothiazide
Hydrochlorothiazide
Bendroflumethiazide
Metolazone
Ethoxzolamide
Methyclothiazide
Benzthiazide
Methazolamide
Thiamylal
Acetazolamide
Hydroflumethiazide
Lamotrigine
Itraconazole
Cyclothiazide
Dorzolamide
Ellagic Acid
Mafenide
Diclofenamide
Progabide
Meprobamate
Zolpidem
Eszopiclone
Flunitrazepam
Chlordiazepoxide
Ketazolam
Clorazepate
Zopiclone
Topiramate
Glutethimide
Voriconazole
Clotrimazole
Fluconazole
Tioconazole
Oxiconazole
Ketoconazole
Enflurane
Isoflurane
Perampanel
Econazole
Methoxyflurane
LULICONAZOLE
Desflurane
Methyprylon
Bifonazole
Sertaconazole
Sevoflurane
Posaconazole
Terconazole

Figure 6

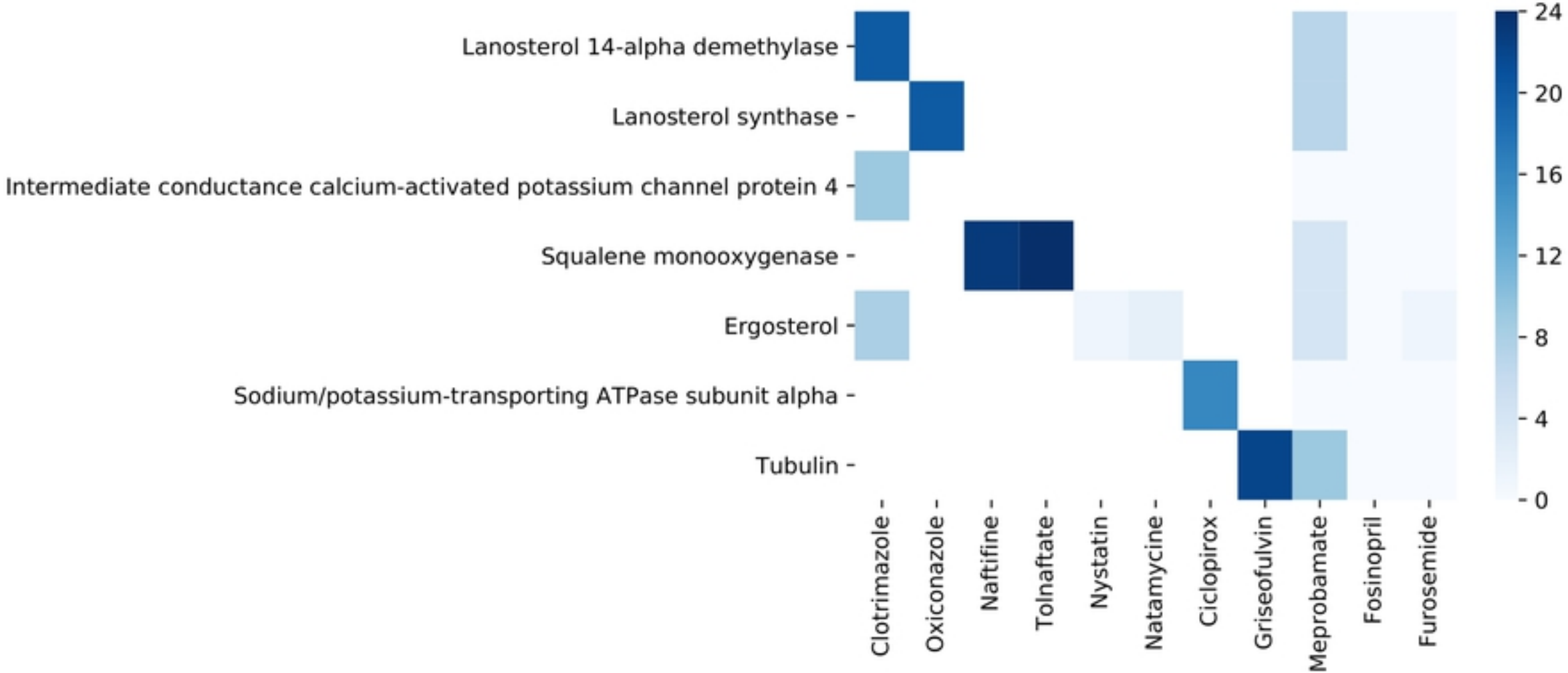

Lanosterol 14-alpha demethylase
Lanosterol synthase
Intermediate conductance calcium-activated potassium channel protein 4
Squalene monooxygenase
Ergosterol
Sodium/potassium-transporting ATPase subunit alpha
Tubulin
Clotrimazole
Oxiconazole
Naftifine
Tolnaftate
Nystatin
Natamycine
Ciclopirox
Griseofulvin
Meprobamate
Fosinopril
Furosemide
24
20
16
12
8
4
0

Figure 7

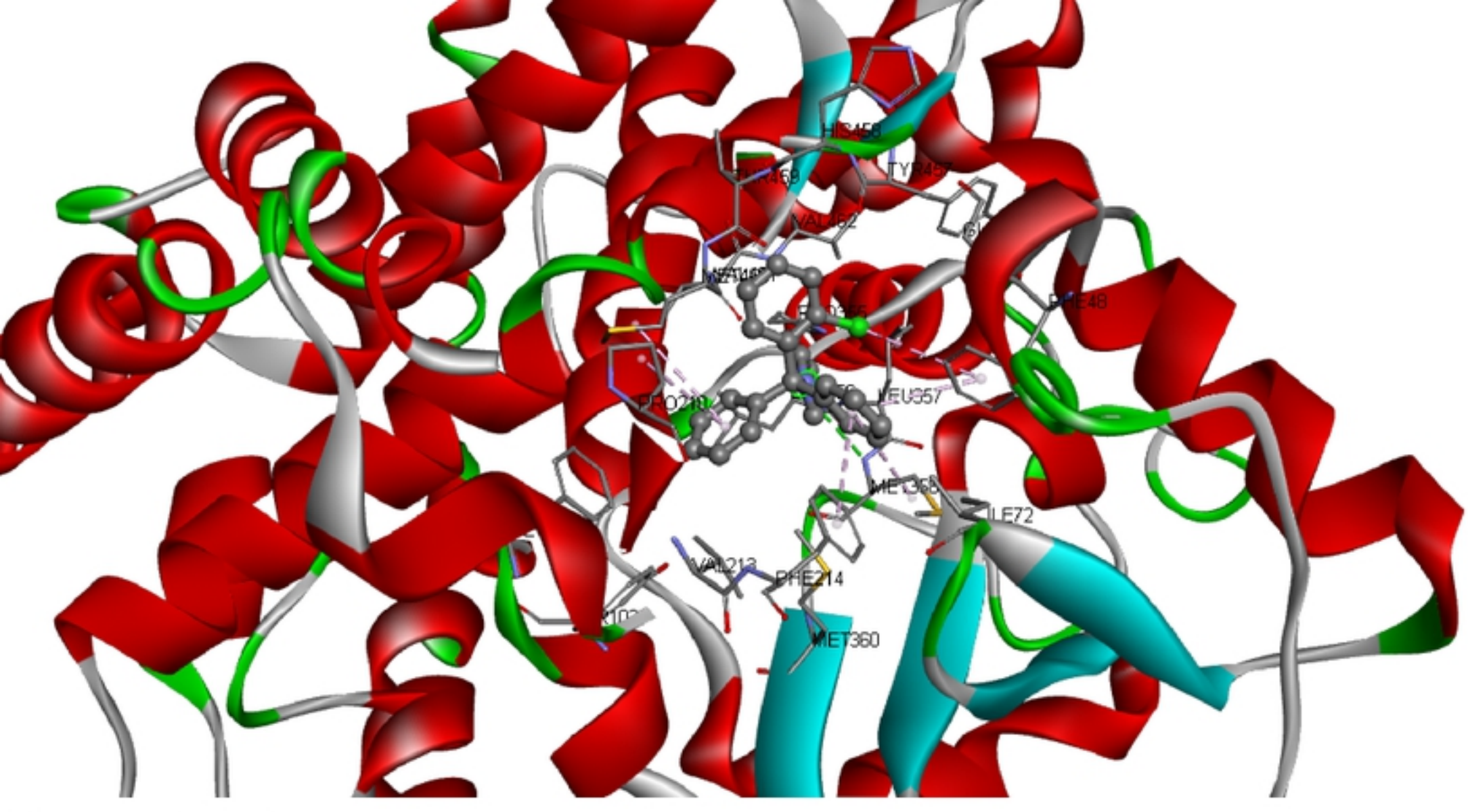


Figure 8 (a)

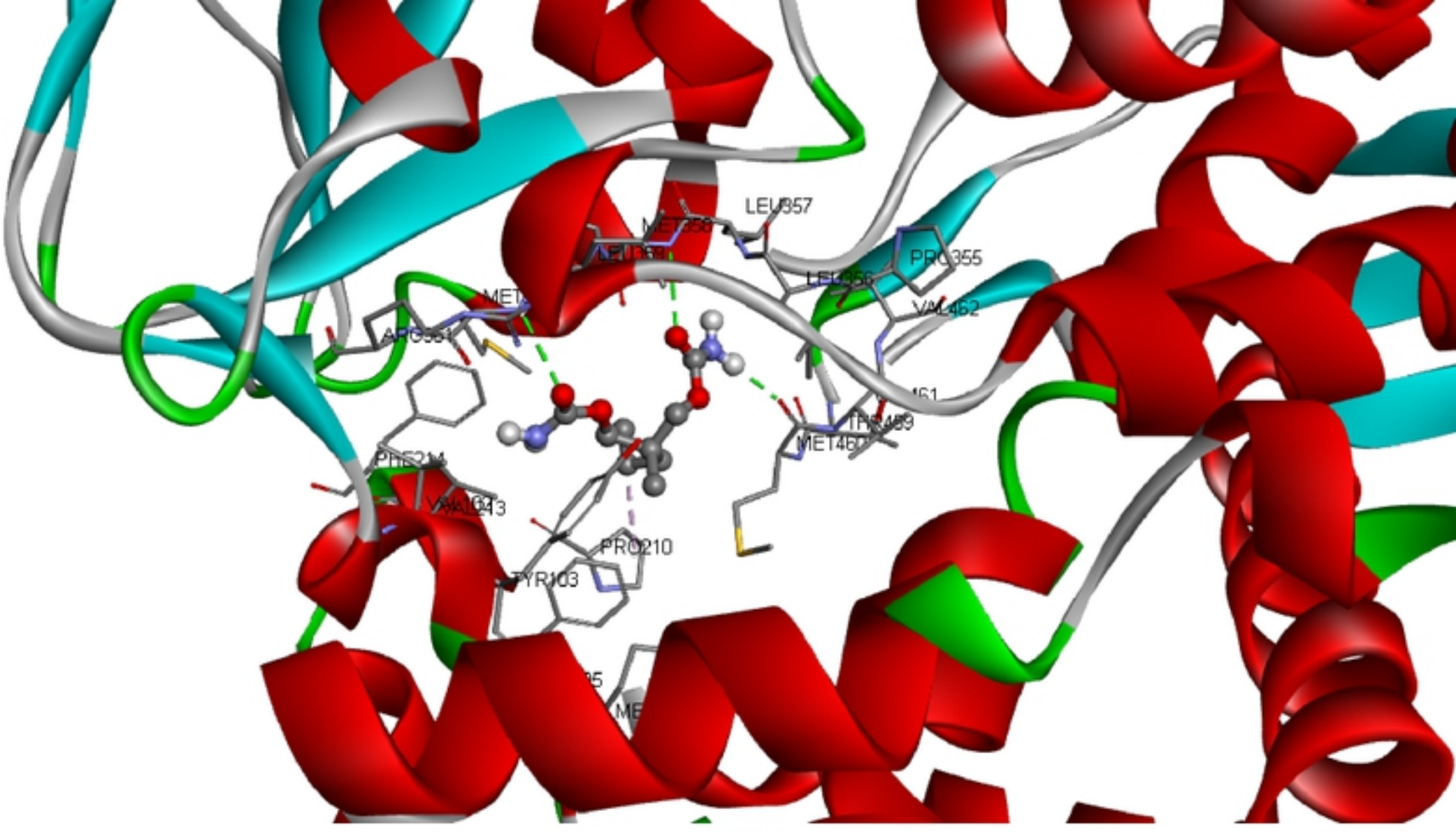


Figure 8 (b)