Network-based Drug Ranking and Repositioning with respect to DrugBank Therapeutic Categories: Supplemental Material

Matteo Re, and Giorgio Valentini

The supplemental material reports additional results, tables and figures relative to the main paper and a section relative to the results of drug ranking methods with Therapeutic Categories characterized by low cardinality.

List of Tables

List of Figures

• *Matteo Re, and Giorgio Valentini are with DI, Dipartimento di Informatica, Università degli Studi di Milano, Via Comelico 39, Milano, Italy, e-mail: {re,valentini}@di.unimi.it*

DRUG RANKING OF THERAPEUTIC CATEGORIES CHARACTERIZED BY LOW CARDINALITY

In this section we evaluate the performance of ψ*NetPro* and kernelized score functions using relatively small DrugBank TCs. More precisely, from the 131 therapeutic classes having from 5 to 15 drugs we randomly selected 10 TCs for each of the 5 to 8, 9 to 12 and 13 to 15 subgroups. The resulting TCs are listed in Tab. 3.

As expected, average AUC and P40R results across TCs (Tab. 4) register a certain decrement with respect to the larger TCs analyzed in Section 4.2 (Tab. 1) of the main paper. This decrement is more significant in terms of P40R, while the average AUC shows a less marked decrease. Tab. 4 also shows that with relatively small TCs the ψ*NetPro* projection and integration works nicely, leading to a significant increment of both AUC and P40R independently of the drug ranking method used. A visual clue of this fact is offered also by Fig. 3 and 4 that show respectively the per-class AUC and P40R results achieved by the ranking methods with $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ and $\boldsymbol{W}_3$ pharmacological spaces. Note however that P40R values are more scattered, revealing a significant decay in performance with respect to results relative to the TCs with more than 15 drugs (Fig. 2 and 4).

With the "small" TCs S_{AV} and S_{kNN} confirm their best average results, but also *RWR* and *RW* 1 step obtain competitive results (Tab. 4). Interestingly enough, S_{AV} and S_{kNN} with 10 steps random walk kernel register the best AUC results, while this is not true for the classic *RW* 10 steps, as just observed with the "large" TCs analyzed in Section 4.3 of the main paper.

But the more significant fact with "small" TCs is that the integration introduces a more consistent advancement in both AUC and P40R: all the methods on the average improve the AUC of about 20 percent points and approximately double the precision passing from $\boldsymbol{W}_1$ to $\boldsymbol{W}_3$ (Tab. 4), while with "large" TCs the improvement is about 10 percent in terms of AUC and the P40R is augmented of at most one half by passing from $\boldsymbol{W}_1$ to $\boldsymbol{W}_3$ (Tab. 1 of the main paper). This fact more evident analyzing the per-class results (Fig. 3 and 4): some TCs such as "Angiotensin_II_Receptor_Antagonists" approximately double their average AUC by integrating more sources of data through ψ*NetPro*, independently of the ranking method used, and others, such as "Anticholesteremic_Agents" or "Antitubercular_Agents" improve P40R results from values close to 0 with $\boldsymbol{W}_1$ to values close to $0.9/1.0$ with $\boldsymbol{W}_3$, with all the ranking methods or at least with the most performing S_{AV}, S_{kNN} or *RWR*.

Despite the fact that the development of new drugs is historically based on the comparison of chemical structures of putatively active compounds and on the availability of expensive to produce information about chemicals-targets interactions, a closer look to Fig. 4 is enough to see that some of the largest observed improvements in terms of P40R are due to the integration of the chemicals-chemicals interaction stored in the STITCH database. This effect is particularly evident for the "Muscle_relaxants" and "Antitubercular_Agents" TCs. A manual inspection of the evolution of the reciprocal wiring patterns of the "Antitubercular_Agents" moving from $\boldsymbol{W}_1$ to $\boldsymbol{W}_3$ revealed that out of 7 drugs (DB00233, DB00330, DB00339, DB00609, DB00951, DB00978 and DB01208) only two (DB00978 and DB01208) are linked in $\boldsymbol{W}_1$, two pairs (DB01208-DB00978 and DB00951-DB00609) are connected in $\boldsymbol{W}_2$ and nearly all are connected in $\boldsymbol{W}_3$. Given that STITCH provides, for each integrated chemical-chemical interaction score, the sub scores used during its computation it was possible to verify that these interactions are mainly due to biomedical literature text mining, confirming the potential of ψ*NetPro* for the effective integration of different information about drugs, especially in difficult learning task such as in the repositioning of drugs belonging to small classes.

TABLE 1
DrugBank Therapeutic Categories (TC) with more than 15 drugs considered in the experiments. The first column reports the abbreviated name, the second the full DrugBank name and the third the cardinality of the TC.

Therapeutic categories with more than 15 drugs		
Abbreviated name	**Full DrugBank name name**	**Card.**
Adren.A.	Adrenergic_Agents	26
Adren.In.	Adrenergic_Uptake_Inhibitors	20
Adren.a.	Strategic_alpha.Agonists	23
Adren.b.	Adrenergic_beta.Antagonists	25
Analges.	Analgesics	40
Analg.Op.	Analgesics._Opioid	24
Anti.Aller.	Anti.Allergic_Agents	35
Anti.Arrh.	Anti.Arrhythmia_Agents	42
Anti.Bact.	Anti.Bacterial_Agents	103
Anti.HIV	Anti.HIV_Agents	22
Anti.Inf.A.	Anti.Infective_Agents	29
Anti.Inf.	Anti.Infectives	19
Anti.Ulcer	Anti.Ulcer_Agents	19
Anti.anx.	Anti.anxiety_Agents	35
Anti.infl.	Anti.inflammatory_Agents	49
Antiarr.A.	Antiarrhythmic_Agents	29
Anticonv.	Anticonvulsants	46
Antidysk.	Antidyskinetics	23
Antiemetics	Antiemetics	34
Antifungal	Antifungal_Agents	22
Antihist.	Antihistamines	24
Antihypert.	Antihypertensive_Agents	105
Antimetab.	Antimetabolites	26
Antineopl.	Antineoplastic_Agents	86
Antineopl.H.	Antineoplastic_Agents._Hormonal	18
Antipark.	Antiparkinson_Agents	27
Antipsyc.A.	Antipsychotic_Agents	39
Antipsyc.	Antipsychotics	27
Antiviral	Antiviral_Agents	25
Bronchodil.	Bronchodilator_Agents	33
Ca.Ch.Block.	Calcium_Channel_Blockers	22
Cephalosp.	Cephalosporins	30
Cycloox.Inh.	Cyclooxygenase_Inhibitors	24
Dietary.sup.	Dietary_supplement	47
Diuretics	Diuretics	23
Dopam..Ant.	Dopamine_Antagonists	29
Enzyme.Inh.	Enzyme_Inhibitors	64
GABA.Mod.	GABA_Modulators	26
Glucocort.	Glucocorticoids	21
Hist.H1.Ant.	Histamine_H1_Antagonists	28
Hypnot.Sed.	Hypnotics_and_Sedatives	41
Hypoglyc.	Hypoglycemic_Agents	22
Immunosup.	Immunosuppressive_Agents	20
Micronutr.	Micronutrient	45
Musc.Ant.	Muscarinic_Antagonists	23
Narcotics	Narcotics	22
Penicillins	Penicillins	20
Sympathol.	Sympatholytics	24
Sympathomim.	Sympathomimetics	32
Vasoconstr.	Vasoconstrictor_Agents	25
Vasodilator	Vasodilator_Agents	55

TABLE 2
The 10 best and the 10 worst AUC values of DrugBank Therapeutic Categories achieved by the S_{AV} score with 3 steps random walk kernel. The last three columns report the precision at 10, 20 and 40 % recall.

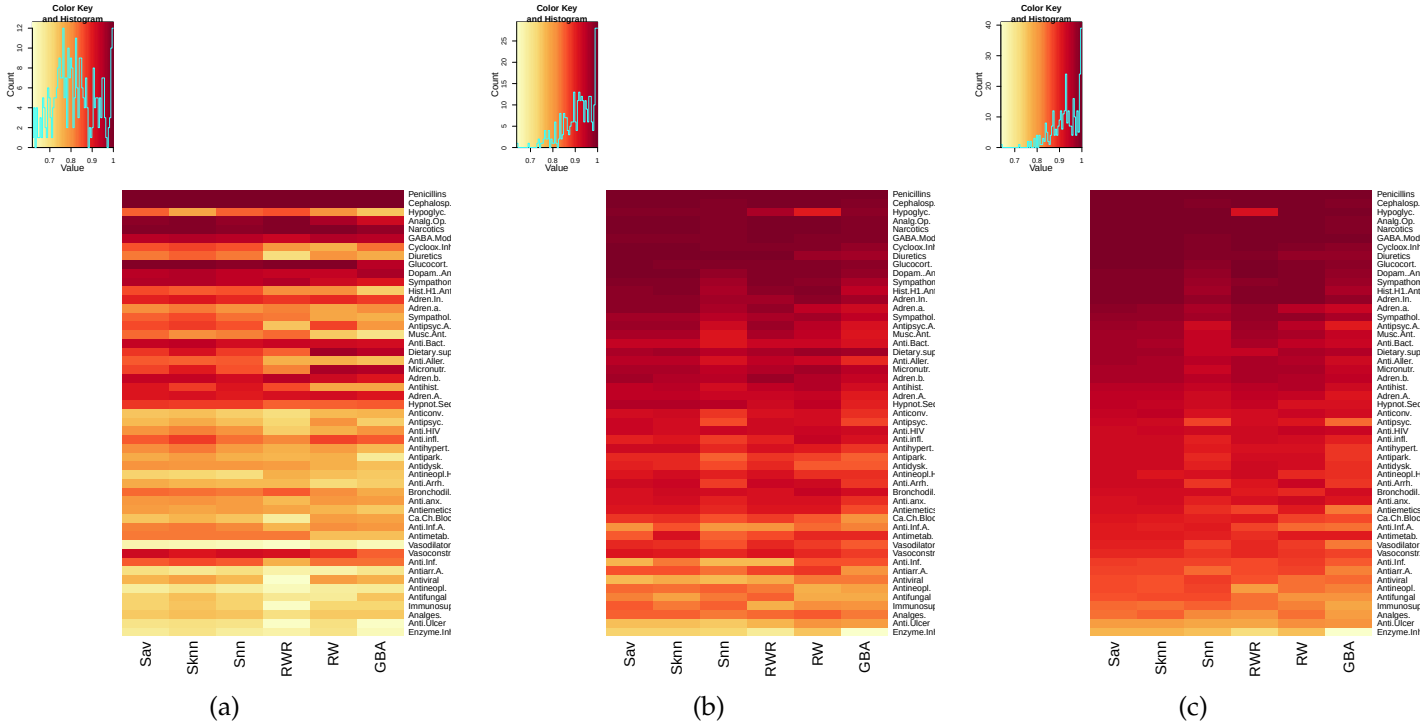
The 10 best ranked TCs				
TC	AUC	$P10R$	$P20R$	$P40R$
Penicillins	0.9999	1.0000	1.0000	1.0000
Cephalosporins	0.9995	1.0000	1.0000	1.0000
Hypoglycemic_Agents	0.9990	1.0000	1.0000	1.0000
Analgesics._Opioid	0.9979	1.0000	1.0000	0.8333
Narcotics	0.9977	1.0000	1.0000	1.0000
GABA_Modulators	0.9966	1.0000	1.0000	0.9166
Cyclooxygenase_Inhibitors	0.9950	0.7500	0.8333	0.9090
Diuretics	0.9948	0.6000	0.6250	0.6923
Glucocorticoids	0.9948	1.0000	1.0000	0.7692
Dopamine_Antagonists	0.9940	1.0000	1.0000	0.7500
The 10 worst ranked TCs				
TC	AUC	$P10R$	$P20R$	$P40R$
Vasoconstrictor_Agents	0.8975	0.7500	0.8333	0.6666
Anti.Infectives	0.8776	0.4000	0.5000	0.3478
Antiarrhythmic_Agents	0.8719	0.1363	0.0895	0.1428
Antiviral_Agents	0.8681	1.0000	1.0000	0.5555
Antineoplastic_Agents	0.8657	0.6923	0.6428	0.4605
Antifungal_Agents	0.8616	0.7500	0.8333	0.9000
Immunosuppressive_Agents	0.8428	0.6666	0.5714	0.4705
Analgesics	0.8402	0.1600	0.2285	0.2758
Anti.Ulcer_Agents	0.8001	1.0000	1.0000	1.0000
Enzyme_Inhibitors	0.7701	0.6363	0.5909	0.4000

TABLE 3
DrugBank Therapeutic Categories (TC) with less than 15 drugs considered in the experiments. The first column reports the abbreviated name, the second the full DrugBank name and the third the cardinality of the TC.

Therapeutic categories with less than 15 drugs		
Abbr. name	**Full DrugBank name**	**Card.**
Antipyr.	Antipyretics	7
Antigl.A.	Antiglaucomic_Agents	5
Antirh.A.	Antirheumatic_Agents	7
Antitub.A.	Antitubercular_Agents	7
Ang.Rec.Ant.	Angiotensin_II_Receptor_Antagonists	5
Antithr.A.	Antithrombotic_Agents	5
Osteopor.Pr.	Osteoporosis_Prophylactic	5
Antidotes	Antidotes	5
Nasal.Dec.	Nasal_Decongestants	7
Corticost.	Corticosteroids	5
Antichol.A.	Anticholesteremic_Agents	11
Dermat.A.	Dermatologic_Agents	11
Gastroint.A.	Gastrointestinal_Agents	10
Sulfonamides	Sulfonamides	10
Antid.Tric.	Antidepressive_Agents._Tricyclic	10
Muscle.Rela.	Muscle_Relaxants._Central	11
Anti.Asth.A.	Anti.Asthmatic_Agents	9
Fibrin.A.	Fibrinolytic_Agents	11
Nootropic.A.	Nootropic_Agents	9
Phenothiaz.	Phenothiazines	11
Antibiotics	Antibiotics	14
Anticoag.	Anticoagulants	15
Antid.IIGen.	Antidepressive_Agents._Second.Generation	14
Antihypoc.A.	Antihypocalcemic_Agents	13
AntiInf.Ur.	Anti.Infective_Agents._Urinary	14
CNS.Stim.	Central_Nervous_System_Stimulants	14
Neuroprot.A.	Neuroprotective_Agents	13
NSAIAs	Nonsteroidal_Anti.inflammatory_Agents_.NSAIAs.	14
Seroton.Inh.	Serotonin_Uptake_Inhibitors	13
Ang.Enz.In.	Angiotensin.converting_Enzyme_Inhibitors	13

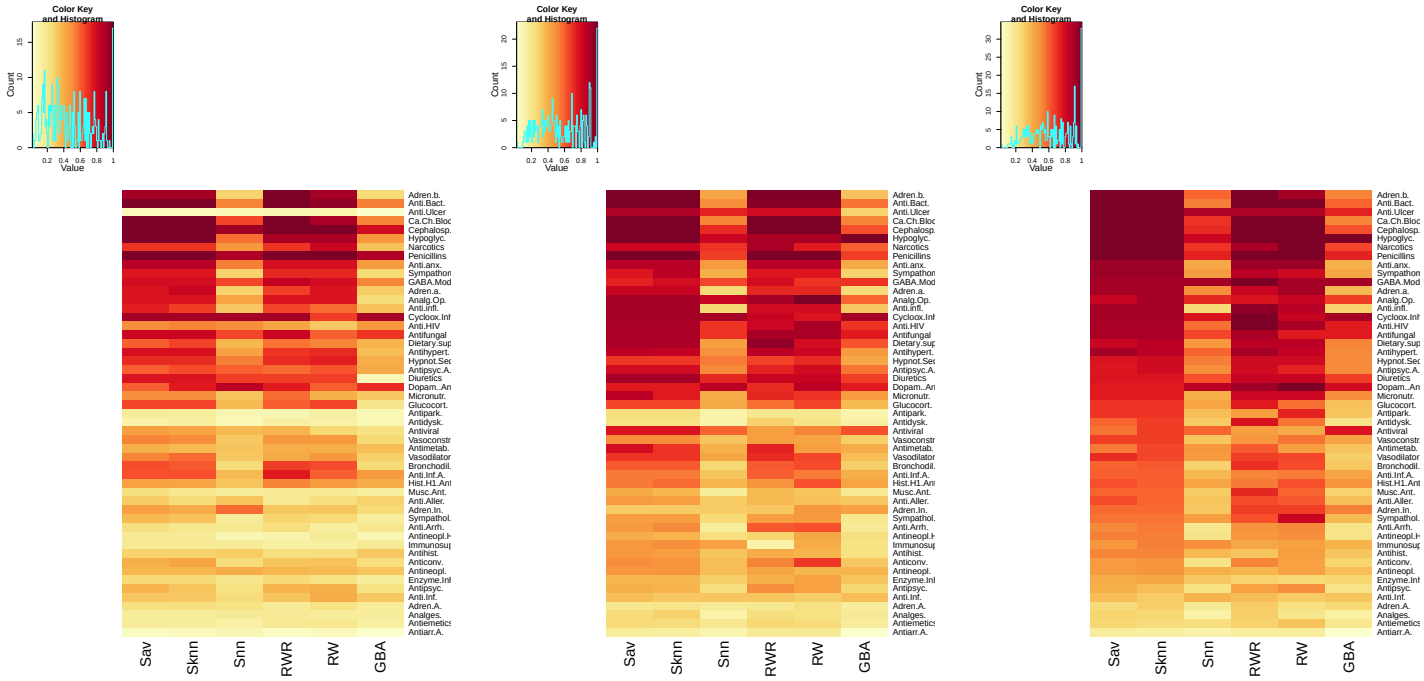
TABLE 4
Average AUC and precision at 40% recall across the DrugBank categories with less than 15 drugs.

Methods	**AUC**			**P40R**		
	W_1	W_2	W_3	W_1	W_2	W_3
S_{AV} 1 step	**0.6924**	0.8635	0.8984	0.2882	**0.4217**	**0.5082**
S_{AV} 10 steps	0.6455	**0.8650**	**0.9153**	0.2710	0.4048	0.4952
S_{kNN} 1 step k=9	**0.6924**	0.8635	0.8983	0.2878	0.4204	**0.5082**
S_{kNN} 10 steps k=9	0.6447	0.8640	0.9115	**0.2920**	0.4143	0.4958
S_{NN} 1 step	0.6916	0.8614	0.8959	0.2436	0.3399	0.4079
S_{NN} 10 steps	0.6447	0.8606	0.9116	0.2522	0.3741	0.4319
RW 1 step	0.6840	0.8620	0.9007	0.2606	0.3707	0.4818
RW 10 steps	0.6130	0.8206	0.8128	0.1605	0.2440	0.2840
$RWR\ \theta = 0.3$	0.6394	0.8601	0.9110	0.2360	0.3869	0.4915
GBA	0.6853	0.8598	0.8909	0.2146	0.3208	0.4105



(a) (b) (c)

Fig. 1. TCs with more than 15 drugs: per class AUC scores compared across methods. Yellow corresponds to the lowest AUC values, while red to the highest AUC values. (a) $\boldsymbol{W}_1$, (b) $\boldsymbol{W}_2$ and (c) $\boldsymbol{W}_3$ pharmacological networks.



(a) (b) (c)

Fig. 2. TCs with more than 15 drugs: per class precision at 40% recall scores compared across methods. Yellow corresponds to the lowest precision values, while red to the highest values. (a) $\boldsymbol{W}_1$, (b) $\boldsymbol{W}_2$ and (c) $\boldsymbol{W}_3$ pharmacological networks.

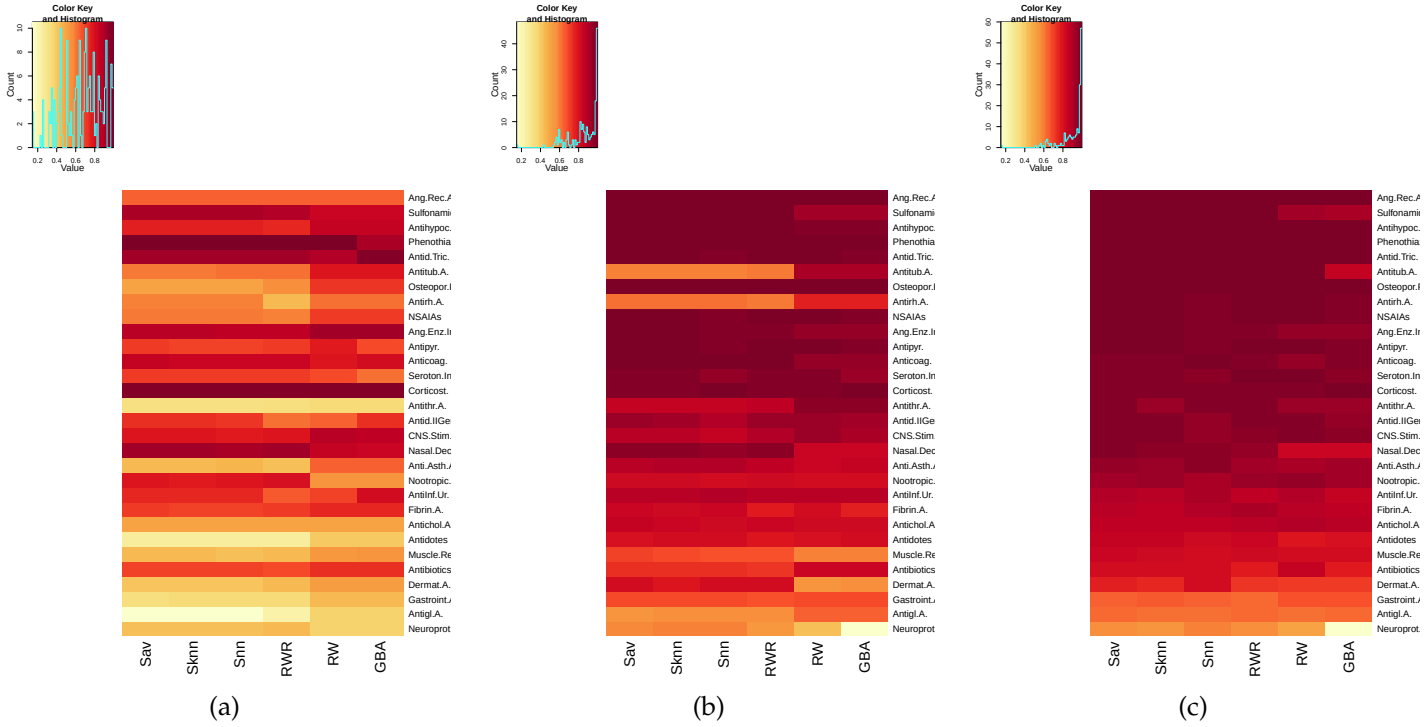


Fig. 3. TCs with less than 15 drugs: per class AUC scores compared across methods. Yellow corresponds to the lowest AUC values, while red to the highest AUC values. (a) $\boldsymbol{W}_1$, (b) $\boldsymbol{W}_2$ and (c) $\boldsymbol{W}_3$ pharmacological networks.

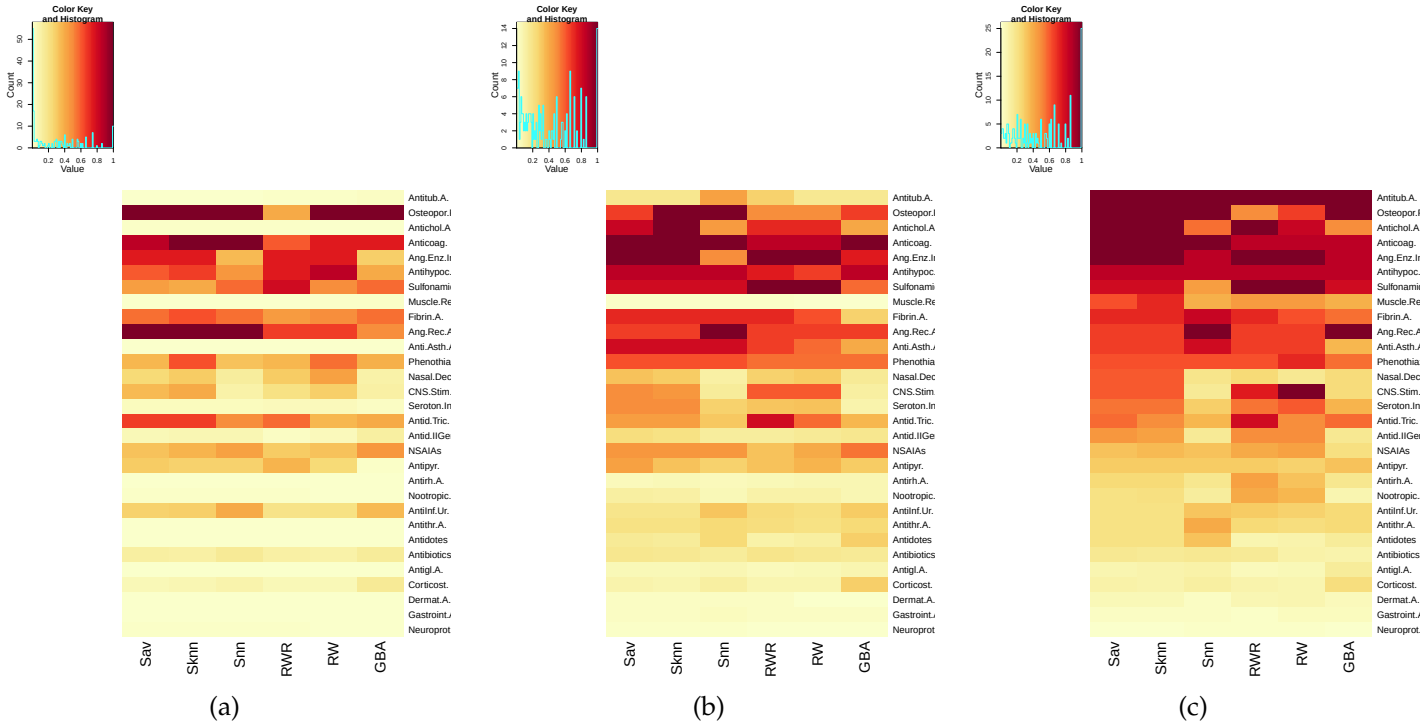


Fig. 4. TC with less than 15 drugs: per class precision at 40% recall scores compared across methods. Yellow corresponds to the lowest precision values, while red to the highest values. (a) $\boldsymbol{W}_1$, (b) $\boldsymbol{W}_2$ and (c) $\boldsymbol{W}_3$ pharmacological networks.

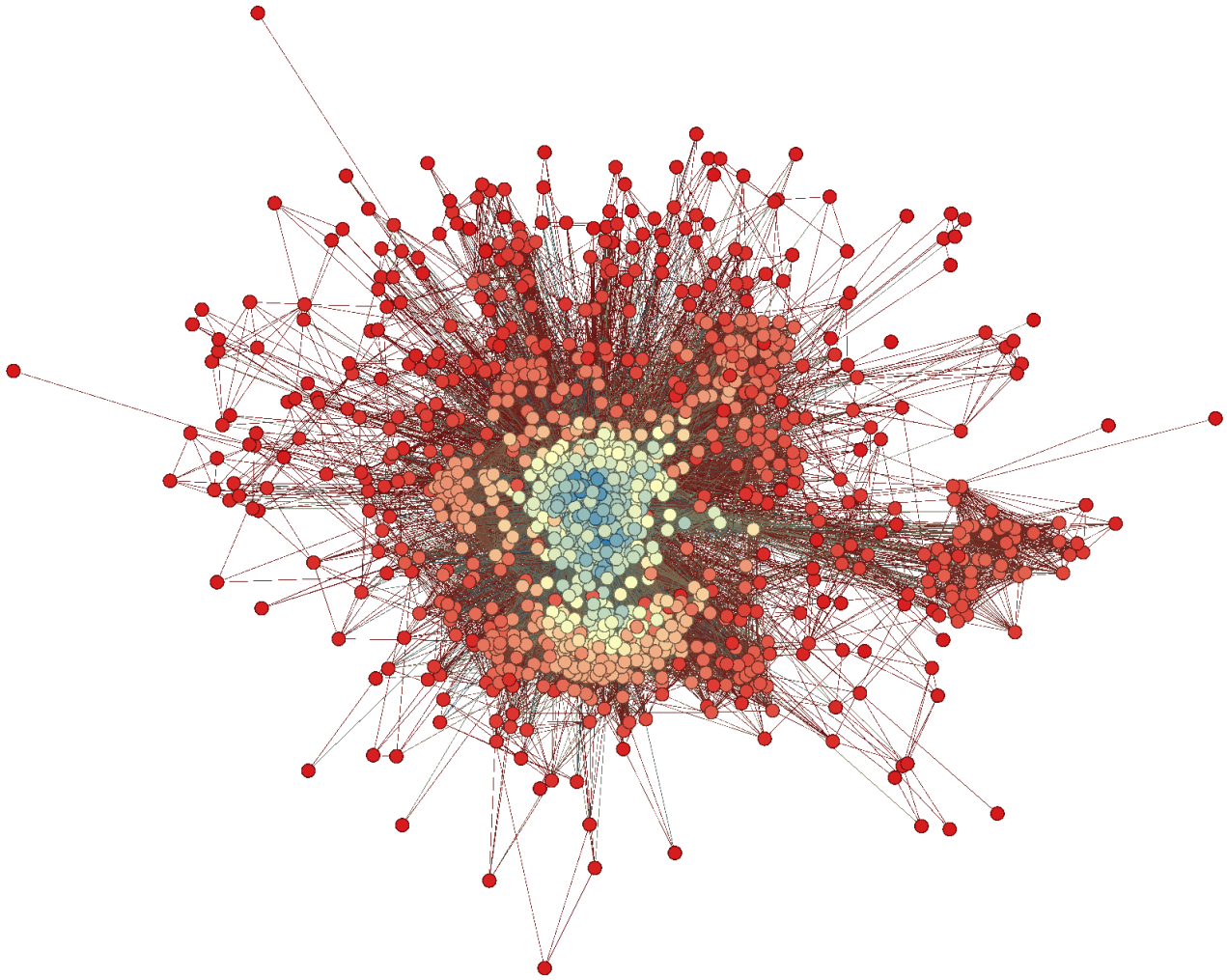


Fig. 5. Graph of the integrated $\boldsymbol{W}_3$ pharmacological network (1253 nodes and 96711 edges). Lighter nodes represent drugs with a higher number of connections (edges) with other drugs in the integrated pharmacological space.