Pioneering topological methods for network-based drug-target prediction by exploiting a brain-network self-organization theory. [Accepted in Briefings in Bioinformatics, IF: 8.399]

Claudio Durán[1,†], Simone Daminelli[2,†], Josephine M. Thomas[1,†], V. Joachim Haupt[2], Michael Schroeder[2], and Carlo Vittorio Cannistraci[1]

1. Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Center for Molecular and Cellular Bioengineering (CMCB), Department of Physics, Technische Universität Dresden (TUD), Dresden, Germany.
2. Bioinformatics Group, Biotechnology Center (BIOTEC), Center for Molecular and Cellular Bioengineering (CMCB), Technische Universität Dresden (TUD), Dresden, Germany.
† The first three authors should be regarded as joint first authors.

The bipartite network representation of the drug-target interactions (DTIs) in a biosystem enhances understanding of the drugs multifaceted action modes, suggests therapeutic switching for approved drugs and unveils possible side effects. Since experimental testing of DTIs is costly and time consuming, computational predictors are of great aid. Here, for the first time, state-of-the-art DTIs supervised predictors custom-made in network biology were compared - using standard and innovative validation frameworks - with unsupervised pure topological-based models designed for general-purpose link prediction in bipartite networks [1]. Surprisingly, our results show that the bipartite topology alone, if adequately exploited by means of the recently proposed local-community-paradigm (LCP) theory [2] - initially detected in brain-network topological self-organization and afterward generalized to any complex network - is able to suggest highly reliable predictions, with comparable performance to the state-of-the-art supervised methods that exploit additional (nontopological, for instance biochemical) drug-target interaction knowledge. Furthermore, a detailed analysis of the novel predictions revealed that each class of methods prioritizes distinct true interactions, hence combining methodologies based on diverse principles represents a promising strategy to improve drug-target discovery.
To conclude, this study promotes the power of bioinspired computing, demonstrating that simple unsupervised rules inspired by principles of topological self-organization and adaptiveness arising during learning in living intelligent systems (like the brain), can efficiently equal-perform complicated algorithms based on advanced, supervised and knowledge-based engineering.

[1] Simone Daminelli, Josephine Maria Thomas, Claudio Durán, and Carlo Vittorio Cannistraci. Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks. New Journal of Physics, 17(11):113037, 2015.
[2] Carlo Vittorio Cannistraci, Gregorio Alanis-Lobato, and Timothy Ravasi. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. Sci Rep, 3:1613, 2013.

Evaluation Framework	Unsupervised (5 networks)				Supervised-LCP (4 networks)				
Existing links		LCP	Proj	MF		LCP	BLM	GRMF	wGRMF
	LCP		3.83E-07	5.99E-06	LCP		1	1	1
	Proj	3.83E-07		0.482053	BLM	1		1	1
	MF	5.99E-06	0.482053		GRMF	1	1		1
					wGRMF	1	1	1	
Removal - re-prediction		LCP	Proj	MF		LCP	BLM	GRMF	wGRMF
	LCP		2.41E-07	1.92E-06	LCP		1	0.223789	0.212392
	Proj	2.41E-07		0.944743	BLM	1		0.212392	0.223789
	MF	1.92E-06	0.944743		GRMF	0.223789	0.212392		1
					wGRMF	0.212392	0.223789	1	
Independent validation		LCP	Proj	MF		LCP	BLM	GRMF	wGRMF
	LCP		0.048921	1.42E-06	LCP		0.381089	0.151174	0.224912
	Proj	0.048921		2.41E-03	BLM	0.381089		0.275857	0.422866
	MF	1.42E-06	2.41E-03		GRMF	0.151174	0.275857		0.422866
					wGRMF	0.224912	0.422866	0.422866	

Figure 1. Statistical comparison for the classes of supervised and unsupervised methods. P-values computed by the non-parametric Mann-Whitney test and adjusted by Bonferroni's correction for the classes of supervised and unsupervised methods in 3 evaluation frameworks. Significant differences between classes of methods are highlighted in blue. On the left, comparison between 3 types of unsupervised methods: LCP-based, projection-based, Matrix Factorization-based. On the right, comparison of 3 types of supervised methods: BLM, GRMF and wGRMF with unsupervised LCP-based methods. LCP-methods perform significantly better than the other unsupervised ($p<0.05$), and surprisingly their prediction is comparable to advanced supervised methods ($p>0.05$).

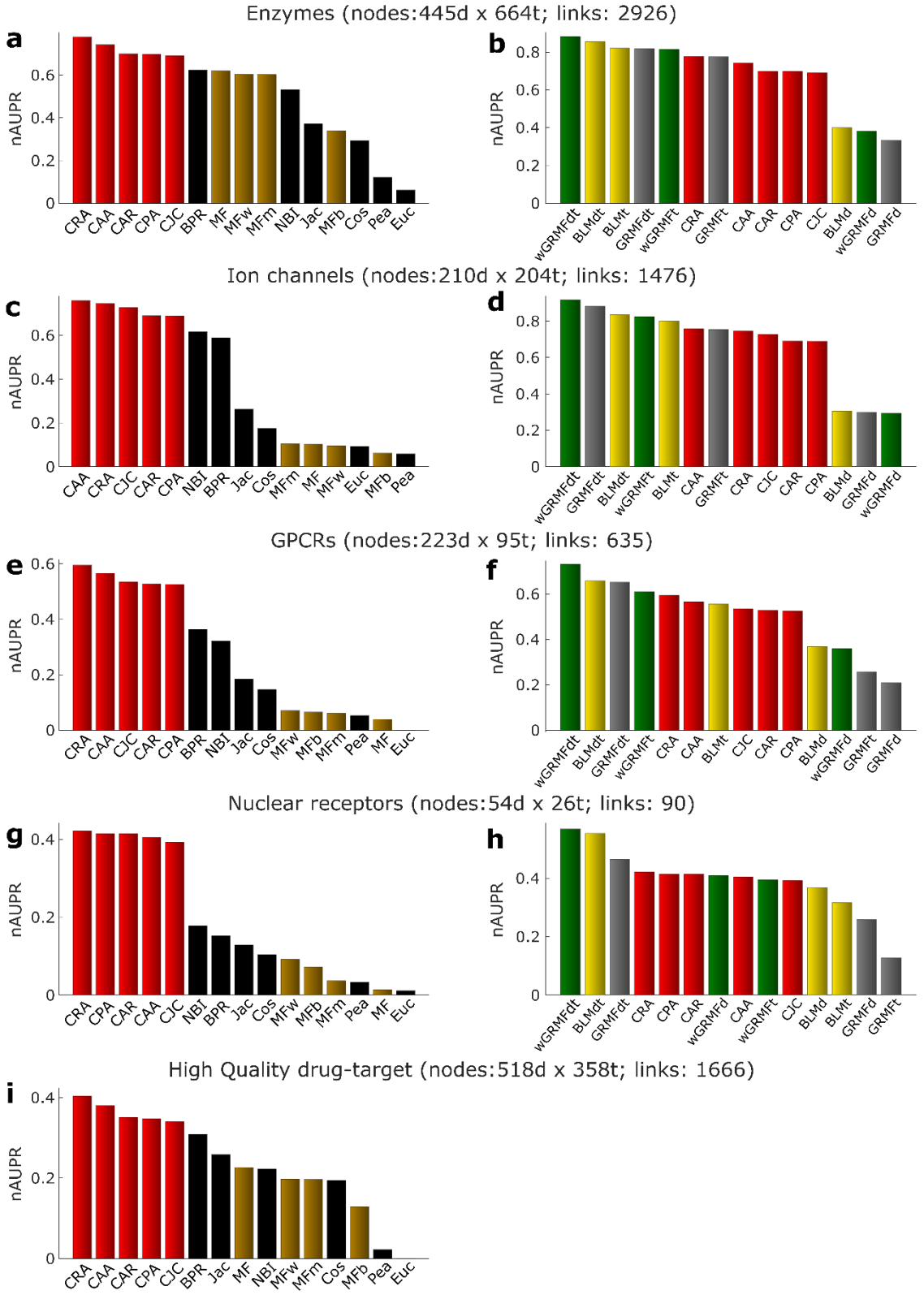

Figure 2. Performance comparison in the existing links evaluation framework.
Normalized AUPR values considering all existing links as True Positive as described in Figure 2a. On the left **(a, c, e, g, i)** comparison of 3 types of unsupervised methods: LCP-based, projection-based, Matrix-Factorization-based. On the right **(b, d, f, h)**, comparison of 3 types of supervised: BLM, GRMF and wGRMF with unsupervised LCP based methods.

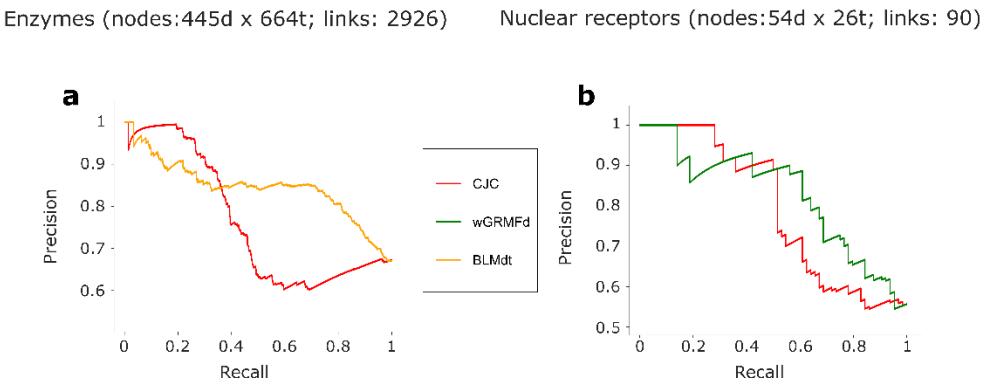

Figure 3. Precision-Recall curves of the Independent validation evaluation framework.
For each class of prediction methods the best method in Enzymes and Nuclear Receptors network were compared with the best LCP-based method: LCP-based (red), BLM (yellow), wGRMF (green). **(a)** Precision Recall curves in Enzyme network, LCP-based (CJC) versus BLM (BLMdt). **(b)** Precision Recall curves in Nuclear Receptors network, LCP-based (CJC) versus wGRMF (wGRMFd).

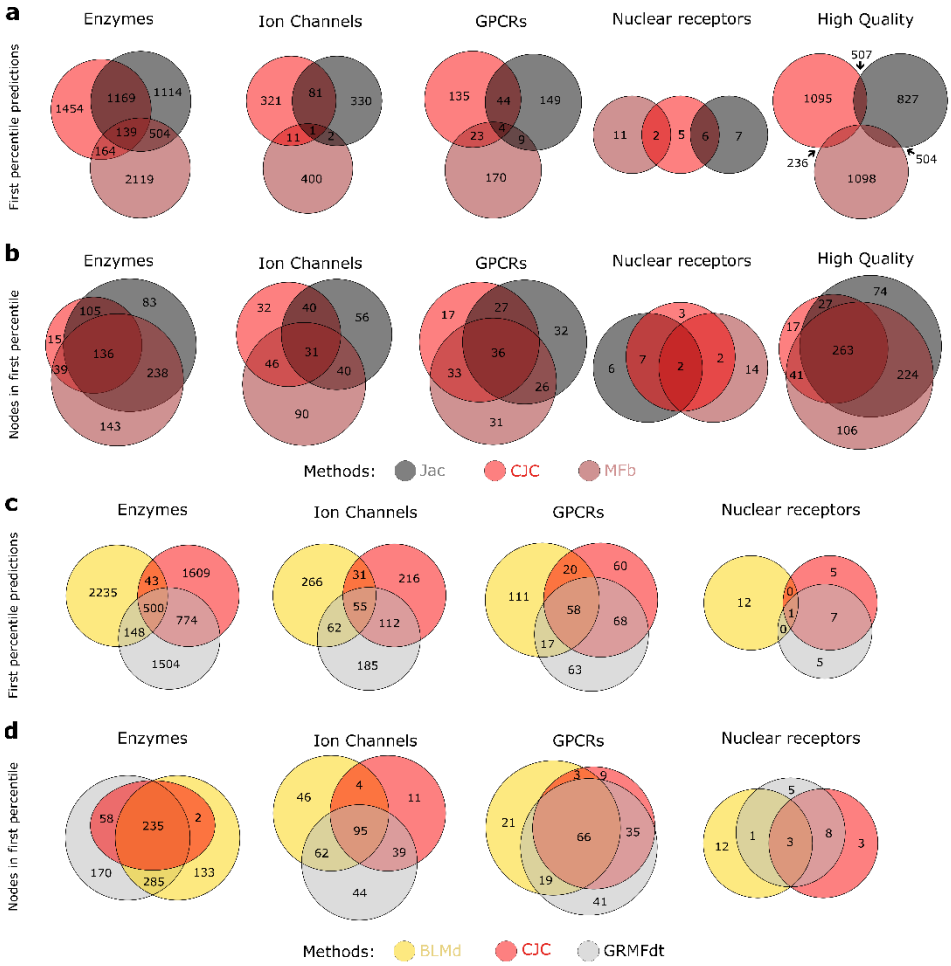

Figure 4. Comparison of novel predicted interactions.
(a) Overlap of the first percentile predictions in 5 networks for unsupervised methods, considering the representative method of each class: MFb (MF based), Jac (projection based) and CJC (LCP-based). **(b)** Overlap of the nodes (drugs and targets) involved in the first percentile predictions for the unsupervised methods. **(c)** Overlap of the first percentile predictions in 4 networks for supervised and LCP based methods, considering the representative method of each class: BLMd (supervised), GRMFdt (supervised), CJC (LCP-based). **(d)** Overlap of the nodes (drugs and targets) involved in the first percentile predictions for the previous methods.