Exploiting Semantics to Predict Potential Novel Links from Dense Subgraphs

Alejandro Flores, Maria-Esther Vidal, Guillermo Palma

Universidad Simón Bolívar, Caracas, Venezuela
{aflores, mvidal, gpalma}@ldc.usb.ve

Abstract. Knowledge graphs encode semantic knowledge that can be exploited to enhance different data management tasks, e.g., query answering, ranking, or data mining. We tackle the problem of predicting interactions between drugs and targets, and propose esDSG, an unsupervised approach able to predict links from subgraphs that are not only highly dense, but that comprise both similar drugs and targets. The esDSG approach extends a state-of-the-art approximate densest subgraph algorithm with knowledge about the semantic similarity of the nodes in the original graph, and then predicts potential novel interactions from the computed dense subgraph. We have conducted an initial experimental study on a benchmark of drug-target interactions. Our observed results suggest that esDSG is able to identify interactions in graphs where existing approaches cannot perform equality well. Further, a large number of esDSG predictions can be validated using external databases as STITCH[1] and Kegg[2]. These results, although initial, reveal how semantics in conjunction with topological information of the knowledge graph may have a great impact on pattern discovery tasks.

1 Introduction

Knowledge graphs are networks of semantically related concepts, described in terms of attributes, types, and relationships. For example, biomedical knowledge graphs represent relationships among drugs, targets, side effects, and so on. Exploring these semantically rich networks can lead to novel discoveries [3], e.g., interactions between drugs and targets, or drug side effects. We focus on the problem of suggesting novel associations between concepts from knowledge graphs where concepts of the same type are characterized by similarity measures. We apply our approach to predict drug-target interactions; however, the approach is general enough to be used in other domains.

Different approaches have been proposed to predict interactions in networks [1, 4, 5, 10, 14, 15]; these approaches have been applied to drug-target graphs, where edges between drugs and targets represent drug-target interactions, and drug-drug and target-target similarity functions are defined. The *prediction hypothesis* is that similar drugs interact with the same targets, and similar targets interact with the same drugs [4]. The main challenge addressed by existing approaches is the development of strategies that allow for detecting portions of the network where novel predictions can be suggested. To illustrate the problem of suggesting drug-target interactions, consider Figures 1(a),

---

[1] http://stitch.embl.de/

[2] http://www.genome.jp/kegg/

(a) Drug-target interactions

(b) Dense Subgraph

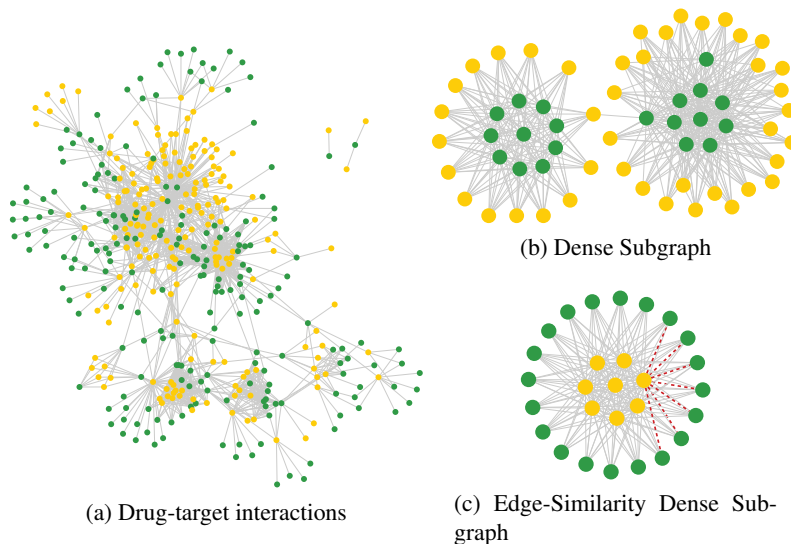(c) Edge-Similarity Dense Sub-graph

Fig. 1: Ion Channel dataset, representing known interactions between targets (yellow) and drugs (green). Predicted interactions are shown in red-dashed lines.

(b), and (c). Figure 1(a) visualizes the network of drug-target interactions of Ion Channels in the dataset studied by Ding et al. [3]; drug-drug and target-target relationships are not drawn. To avoid the pair-wise comparison of all the drugs and targets in the graph, existing approaches limit the search to dense portions of the graph that comprise both similar drugs and similar targets. For example, the semEP [10] approach represents the network of drug-target interactions as a bipartite graph, and partitions the edges (interactions) of the graph according to a density function that measures how similar are the drugs and targets that are placed in one partition. Further, the number of interactions between drugs and targets in each partition are taken into account by this measure. semEP is able to explore portions of the search space and suggest predictions that other state-of-the-art approaches do not discover. However, in graphs such as the one reported in Figure 1(a), semEP and state-of-the-art approaches do not exhibit as good behavior as in other networks. The Ion Channel graph is composed of 414 nodes and 1,476 edges. It also has three connected components; two of these components are comprised of five nodes and three edges. The rest of the nodes are part on the main connected component, from which the dense subgraph of the original network can be computed; this dense subgraph is presented in Figure 1(b). We hypothesize that in graphs with this topology, limiting the search to highly dense components could lead to better predictions. However, dissimilar drugs or targets may comprise these highly connected components. We address the problem of identifying subgraphs that are not only dense, but are composed of highly similar drugs and targets. We name these graphs *edge-similarity dense subgraphs*; Figure 1(c) shows an *edge-similarity dense subgraph* for the Ion Channel dataset. Evaluating the *prediction hypothesis* in this subgraph allows for suggesting seven predictions shown in red-dashed lines.

We propose a novel technique named esDSG, that is able to produce edge-similarity dense subgraphs and to suggest novel associations from these subgraphs. We have conducted an experimental study on the benchmark of drug-target interactions studied by Ding et al. [3]. Interactions suggested by semEP and five state-of-the-art machine learning based techniques are compared to the interactions produced by esDSG. The observed results suggest that the search spaces explored by all these approaches are different. Further, esDSG identifies interactions in graphs where existing approaches cannot perform equality well. More importantly, a large number of the esDSG predictions can be validated using the external databases STITCH and Kegg. These results suggest that considering the topology in conjunction with semantics encoded in the graph, e.g., similarity values, can have a positive impact on the effectiveness of discovery tasks.

The rest of the paper is as follows: Section 2 describes the esDSG approach and Section 3 summarizes related approaches. Experimental results are presented in Section 4. Finally, we conclude in Section 5 with an outlook to future work.

2 Edge-Similarity Densest Subgraph Problem

Our goal is to discover novel drug-target interactions by exploiting relationships among existing interactions and values of similarity between both drugs and targets. However, to effectively suggest such interactions, the *prediction hypothesis* should be evaluated in graphs that maximize: *i*) graph density, and *ii*) drug-drug and target-target similarities. We cast the problem of generating these graphs into a single-objective optimization problem, called the edge-similarity Densest Subgraph (esDSG) problem.

Let $D = \{d_1, d_2, \ldots, d_n\}$ be a set of drugs, and let $s_d(d_i, d_j) \in [0 \ldots 1]$ be the similarity score function defined between every two drugs in D. Similarly, let $T = \{t_1, t_2, \ldots, t_m\}$ be a set of targets, and let $s_t(t_i, t_j) \in [0 \ldots 1]$ be the similarity score function defined between every two targets in T. Given the sets T and D, we call E the set of interactions between targets and drugs, where $E \subseteq T \times D$. These interactions can be naturally represented as a bipartite graph $G = (T, D, E)$.

For each node v in the subgraph, the goal is to maximize not only the number of interactions (i.e., the degree), but also the average of the pair-wise similarity values of v with respect to the rest of the nodes in the subgraph. To do this, we define the similarity measure of a single node v as follows:

Definition 1 (Single-node similarity). *Consider a bipartite graph $G = (T, D, E)$ of drug-target interactions. The single-node similarity of $v \in T$ (resp., $v \in D$) corresponds to the arithmetic mean of the similarity measure values of v with every node $v' \in T$ (resp., $v' \in D$). This is denoted as $s_n(v)$.*

$$s_n(v) = \begin{cases} \frac{1}{|T|} \sum_{t \in T} s_t(v, t) & \text{if } v \in T \\ \frac{1}{|D|} \sum_{d \in D} s_d(v, d) & \text{if } v \in D \end{cases} \tag{1}$$

Figure 2(a) presents a subgraph of the Ion Channel dataset. Similarity values allow for computing the *single-node similarity* of each node. For example, the drug D00332

is dissimilar to the rest of the drugs in the subgraph. This is represented with low values of the *single-node similarity*, i.e., $s_n(\mathsf{D00332}) = \frac{1}{3}(1 + 0.16 + 0.17)$. On the other hand, the drug $\mathsf{D05077}$ is highly similar to the rest of the drugs in this subgraph; thus, a higher value of *single-node similarity* is assigned to $\mathsf{D05077}$, i.e., $s_n(\mathsf{D05077}) = \frac{1}{3}(1 + 0.97 + 0.17)$. Note that *single-node similarity* values depend on the context in which they are computed, i.e., only the nodes in a particular subgraph are considered.
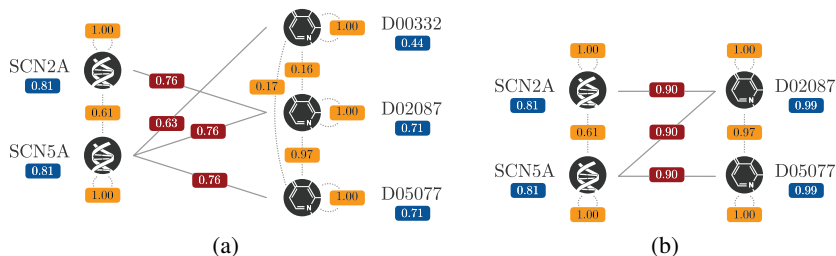

Fig. 2: Subgraphs of drug-target interactions of the Ion Channel dataset, where (b) is subgraph of (a). Drug-Drug and Target-Target similarities in yellow, single-node similarities in blue, and edge similarities in red.

We can characterize edges in a graph based on the *single-node similarity* values of the nodes that connect each edge. For example, because $\mathsf{D00332}$ is less similar than $\mathsf{D05077}$ to the rest of the drugs in the subgraph of Figure 2(a), the interaction that relates $\mathsf{D00332}$ with the target $\mathsf{SCN5A}$ is not as important in terms of similarity as the interaction between $\mathsf{D05077}$ the same target $\mathsf{SCN5A}$. We model the importance of an interaction in terms of similarity as a function named *edge similarity*, that averages the single node similarities of the target and the drug that are related with this interaction.

Definition 2 (Edge similarity). *Consider a bipartite graph $G = (T, D, E)$ of drug-target interactions, and the single-node similarity value of every drug and target in the graph. The edge similarity score of an interaction $(t, d) \in E$, denoted as $s_e(t, d)$, corresponds to the arithmetic mean of the single-node similarity values of t and d.*

$$s_e(t, d) = \frac{1}{2}\big(s_n(t) + s_n(d)\big) \tag{2}$$

Based on this definition, the *edge similarity* of the interaction between $\mathsf{D00332}$ and $\mathsf{SCN5A}$ is 0.63, while the *edge similarity* of the interaction between $\mathsf{D05077}$ and $\mathsf{SCN5A}$ is 0.76. These values of the *edge similarity* function indicate the importance of these two interactions in terms of the similarity of the drugs and targets that they relate, with respect to the rest of the drugs and targets in the subgraph.

Normally, the definition of density for a bipartite graph, denoted as $D(G)$, is given by $D(G) = \frac{|E|}{\sqrt{|T||D|}}$. Using this definition, each edge of the graph is equally relevant in terms of the density of the graph. Our approach relies on the importance of edges in terms of similarity, i.e., the *edge similarity*, to compute a similarity-based density.

Definition 3 (Edge-Similarity Density). *Consider a bipartite graph $G = (T, D, E)$, where T represents the set of targets, D the set of drugs, and E the set of drug-target interactions. Let $s_e(t, d)$, for $(t, d) \in E$, be the edge similarity value for every edge in the graph. The Edge-Similarity Density of G can be defined as:*

$$D_{es}(G) = \frac{\sum\limits_{(t,d) \in E} s_e(t, d)}{\sqrt{|T||D|}} \tag{3}$$

Using the *edge similarity* values computed for every edge in the subgraph in Figure 2(a), the *edge-similarity density* is equal to 1.19, while the graph density for this subgraph is 1.63. Removing the drug D00332 from this subgraph results in the subgraph G' in Figure 2(b) with a decreased density value of 1.50. However, the *edge-similarity density* in this new graph is higher, i.e., $D_{es}(G') = 1.35$. Applying the *prediction hypothesis* to this subgraph would result in the prediction of an interaction between the target SCN2A and the drug D05077; this interaction can be validated in Kegg.

Based on the definition of the *edge-similarity density*, we define the Edge-Similarity Densest Subgraph problem (esDSG) as follows:

Definition 4 (Edge-Similarity Densest Subgraph problem (esDSG)). *Given a bipartite graph $G = (T, D, E)$ as described above, representing the set of interactions between targets and drugs, esDSG identifies the subgraph $G^* \subseteq G$ such that the edge-similarity density of G^*, $D_{es}(G^*)$, is maximized.*

2.1 A greedy approximate algorithm for esDSG

We propose a greedy approximate algorithm to the esDSG problem. The resulting subgraphs are both highly dense and similar, and they will be used to predict novel interactions between targets and drugs. The *edge-similarity density* is used to guide the greedy algorithm into spaces of dense and similar subgraphs.

The greedy decision of removing the nodes that contribute the least to the *edge-similarity density* may result in higher values of the measure. For example, as shown in Figure 2, removing the drug D00332 with both minimum *single-node similarity* and degree values, as well as removing the interaction between this drug and the target SCN5A, leads to an increase of the *edge-similarity density* from 1.19 to 1.35.

After applying the definition of *edge similarity* and some algebraic manipulation, the *edge-similarity density* can be reformulated as a function of the degree values and the *single-node similarity* of each node in a graph G, as follows:

$$D_{es}(G) = \frac{\sum\limits_{v \in T \cup D} degree(v) \cdot s_n(v)}{2\sqrt{|T||D|}} \tag{4}$$

We propose a modification of the densest subgraph 2-approximation algorithm proposed by Khuller and Barna [7]. In our algorithm, the greedy decision will be made not only in terms of the degree, but also the *single-node similarity* of the nodes of the graph. Along with the use of the *edge-similarity density*, this constitutes the most important difference with Khuller & Barna's algorithm, where the greedy strategy indicates

the removal of nodes, according only to the degree.

Algorithm 1: Edge-Similarity Dense Subgraph

Input: A bipartite graph $G = (T, D, E)$
Output: A subgraph $H^* \subseteq G$

1 $i \leftarrow |T| + |D|, H_i \leftarrow G$
2 **while** $H_i \neq \varnothing$ **do**
3   Let v be a vertex with minimum value $degree(v) \cdot s_n(v)$ in H_i
4   Delete all incident edges in v
5   Delete all nodes with no incident edges
6   Update single-node similarity values of all nodes
7   Call the new graph H_{i-1}, $i \leftarrow i - 1$
8 **return** H^*, *a graph with maximum edge-similarity density, $D_{es}(H^*)$, from all the H_i's*

Algorithm 1 performs an ordered removal of the nodes of a bipartite graph, removing a node v with minimum product of the *single-node similarity* and degree values, i.e., a node such that $degree(v) \cdot s_n(v)$ is the minimum value. Further, the incident edges on v are removed from the graph. The algorithm iteratively performs this greedy decision until no more edges can be removed. In each iteration i of the algorithm (**while loop** in lines *2* to *7*), a subgraph H_i is created; the algorithm outputs the subgraph H^* with the maximum value of $D_{es}(H^*)$. Predictions correspond to the missing edges in H^*. For example, consider the subgraph in Figure 1(c), this is a dense graph produced by Algorithm 1 for the Ion Channel dataset; seven red-dashed lines represent missing edges in this subgraph and correspond to the interactions predicted by our approach. It is important to highlight that four out of the seven predicted interactions could be validated in STITCH with prediction score greater than 0.16; only two of these interactions were among the top-10 predictions of semEP.

Finally, for a bipartite graph $G = (T, D, E)$, Khuller's algorithm requires $\mathcal{O}(|E| + |T \cup D|)$ time, while the time complexity of Algorithm 1 is $\mathcal{O}(|T|^2 + |D|^2)$. The increased complexity is a consequence of the need to compute the *single-node similarity* values. i.e., after each removal the sets of drugs and targets changes, thus, the *single-node similarity* values of the nodes in the graph need to be updated.

3 Related Work

We briefly describe existing approaches for graph data mining and link prediction. Graph mining approaches focus on the problem of detecting patterns in graphs by conducting clustering and ranking methods on knowledge graphs [6, 9, 12]. RankClus [12], GNetMine [6], and JointCluster [9] apply clustering techniques to place multi-typed concepts in the same cluster. RankClus exploits link analysis-based ranking with clustering to assign highly ranked entities to highly ranked clusters. GNetMine applies learning based methods on different graph properties, e.g., graph topology, type attributes, or edge labels, to classify and partition nodes in a graph. Finally, JointCluster [9] is a *simultaneous clustering* such that nodes within each cluster in the partition are highly connected, and the number of inter-cluster edges is minimized. These clustering

methods have shown to be effective, but they cluster nodes instead of edges. Further, they mainly focus on the topology of the graph and do not exploit any semantics, constraints, or similarity information between the nodes to enhance the quality of the clustering. Dissimilar nodes may be placed in a cluster, reducing thus the chances to predict edges whenever the *prediction hypothesis* is evaluated.

Several approaches have been proposed for link prediction, and particularly, to suggest annotations for genes [11, 13] and interactions between drugs and targets [1, 4, 5, 10, 14, 15]. Techniques to identify dense subgraphs [11] and graph summarization [13] have been extended to predict Gene Ontology annotations of genes that comprise a clique or from super nodes in a summary of a graph, respectively. These approaches have shown to successfully analyze patterns in graphs. However, because they do not consider semantics encoded in the graph or similarity between the graph concepts, they may be grouping dissimilar concepts in a dense graph or a summary of a graph. Thus, applying the *prediction hypothesis* graphs produced by these approaches may lead to a small dataset of predictions or just to produce no predictions at all.

Machine learning methods [1, 4, 5, 14, 15] have been exploited to identify the portions of the knowledge graph where the application of the *prediction hypothesis* can lead to the discovery of drug-target interactions; a detailed evaluation of all these approaches can be found at [3]. Additionally, semEP [10] implements an unsupervised method able to partition edges that represent drug-target interactions; the *prediction hypothesis* is applied to each component of the partition to predict new interactions. Components of a partition maximize the graph density and pair-wise similarity of the drugs and the targets. Although all these approaches are able to suggest drug-target interactions that can be validated, they may not exhibit the same good performance in graphs as the one illustrated in Figure 1(a). We hypothesize that because esDSG reduces the prediction search to the space of edge-similarity dense graphs, the overlap between the esDSG predictions and the ones suggested by the other approaches will be small; further, esDSG may be able to exhibit better performance from this type of graphs.

4 Evaluation of esDSG and State-of-the-Art Methods

Dataset: We use the dataset created by Bleakley et al. [1] to evaluate esDSG quality. This dataset is used in the comparison of existing interaction prediction methods [3, 10]. The dataset comprises 900 drugs, 1,000 targets, and 5,000 interactions from KEGG BRITE[3], BRENDA[4], SuperTarget[5], and DrugBank[6]. The dataset provides a drug-drug chemical similarity score based on the hashed fingerprints from the SMILES resource, and a target-target similarity score based on the normalized Smith-Waterman sequence similarity score. Data is divided into four groups: Nuclear receptors (NR), Gprotein-coupled receptors (GPCR), Ion channels (IC) and Enzymes (E); details in Table 1.
State-of-the-art Methods: We use semEP [10] and the following machine learning methods reported in [3] to evaluate the quality of the esDSG predictions: *i*) BLM:

---

[3] http://www.genome.jp/kegg/brite.html

[4] http://www.brenda-enzymes.de/

[5] http://bioinf-apache.charite.de/supertarget_v2/

[6] http://www.drugbank.ca/

Table 1: Statistics for the Drug-Target Interaction Dataset [1].

Statistics	NR	GPCR	IC	E
Number of drugs	54	223	210	445
Number of targets	26	95	204	664
Interactions	90	635	1,476	2,926
Average interaction count per target	3.46	6.68	7.23	4.4
Average interaction count per drug	1.66	2.84	7.02	6.57
Density	2.40	4.36	7.13	5.38

Bipartite Local Method [2]; *ii*) LapRLS: Laplacian Regularized Least Squares [15]; *iii*) GIP: Gaussian Interaction Profile [14]; *iv*) KBMF2K: Kernelized Bayesian Matrix Factorization with twin Kernels [5]; and *v*) NBI: Network-Based Inference [2]. Additionally, Khuller & Barna's algorithm is used as baseline.

4.1 Dense Subgraphs and Edge-Similarity Dense Subgraphs

We compare the dense subgraphs, DSG and esDGS, produced by Khuller & Barna's algorithm and our approach, respectively. Table 2 describes these subgraphs in terms of number of drugs, targets, interactions, density, and average similarities. We can observe that DSGs are larger than esDSGs and also have greater values of the *standard* density. For datasets NR, IC, and E, the esDGS graphs are denser in terms of *edge-similarity density*, and result on less interactions but all predicted from similar drugs or targets. However, although the DSG approximation algorithm does not maximize

Table 2: Statistics of the Dense Subgraphs and Edge-Similarity Dense Subgraphs computed for each graph in the Drug-Target Interaction Dataset [1].

Statistics	NR		GPCR		IC		E	
	DSG	esDSG	DSG	esDSG	DSG	esDSG	DSG	esDSG
Targets	6	5	9	9	74	8	61	13
Drugs	3	3	16	13	53	20	37	13
Interactions	18	15	107	87	815	153	994	168
Predictions	0	0	37	30	3107	7	1263	1
Density	**4.24**	3.87	**8.91**	8.04	**13.01**	12.09	**20.92**	12.9
Edge-Similarity Density	1.94	**2.01**	**2.66**	2.65	1.79	**6.88**	3.47	**5.53**
Average Target Similarity	0.40	**0.52**	0.33	0.33	0.06	**0.69**	0.10	**0.54**
Average Drug Similarity	0.50	0.50	0.25	**0.32**	0.21	**0.43**	0.23	**0.31**

edge-similarity density, DSG has higher values of this measure than esDSG on GPCR. Further, DSG and esDSG have 24 common predicted interactions. This suggests that GPCR regions that are dense or similar are the same, i.e., the search spaces of the approximation algorithms for DSG and esDSG overlap.

4.2 Validation using STITCH and Kegg

We evaluate the quality of esDSG predictions in terms of the number of interactions that can be validated. We use the latest online version of STITCH [8] and Kegg to

validate each drug-target interaction predicted by esDSG. Table 3(a) reports on both the number of esDSG validated drug-target interactions, and the total number of predicted interactions. We can observe that esDSG predicted 12 out of 30 predictions that could be validated in STITCH and Kegg; while the 57% of the esDSG predicted interactions in the Ion Channel dataset can be also validated. As observed in Table 2, the IC esDSG is the one with the highest values of *edge-similarity density*, and average target and drug similarities. Furthermore, Table 3(b) reports on the number of validated interactions out of the top-5 novel predicted interactions of all methods; novel interactions are not in the original datasets and have the highest prediction score; these scores are computed by each prediction method. It is clear that esDSG outperforms existing approaches in the Ion Channel dataset. This supports our hypothesis that novel interactions can be predicted from highly dense graphs composed of similar drugs and targets.

Table 3: (a) Number of esDSG predicted interactions and the number of those interactions that can be manually validated with STITCH and Kegg. (b) Top-5 novel interactions manually validated with STITCH; entries highlighted in bold correspond to the largest number of novel validations

(a) esDSG Validated Predictions

Dataset	Number of predictions	Validation		
		STITCH	Kegg	Total
NR	0	-	-	-
GPCR	30	11	3	12
IC	7	4	0	4
E	1	0	0	0

(b) Validated Predictions of State-of-the Prediction Methods

Method	NR	GPCR	IC	E
semEP	4	**5**	1	**4**
BLM	2	1	0	0
NBI	1	1	1	2
GIP	3	3	1	1
LapRLS	**5**	3	**2**	2
KBMF2K	3	4	**2**	2

4.3 Effectiveness of esDSG Predictions

Effectiveness of esDSG is also measured as the overlap between the esDSG predicted interactions and the top-10 interactions predicted by semEP [10], and the machine learning methods: BLM, NBI, GIP, LapRLS, and KBMF2K [3]. Top-10 predictions correspond to the interactions with the highest prediction score. Table 4 compares esDSG and the corresponding prediction method in terms of both: *i*) the number of interactions that both methods predicted, labelled as *Equal*, and *ii*) the number of interactions predicted by esDSG and that are not part of the top-10 predictions of the corresponding method, labelled as *Diff*. We can observe that esDSG predicts interactions that are not predicted by any of the other methods. These results suggest that the edge-similarity subgraphs from where esDSG generates the predicted interactions, correspond to portions of the search space that none of the other approaches is able to explore.

5 Conclusions

We defined the *edge-similarity densest subgraph* problem to predict drug-target interactions. We extended Kuller & Barna's algorithm with the greedy decision of removing the nodes that contribute less to the *edge-similarity density* values. Experimental results suggest our approach is able to predict novel drug-target interactions from highly dense

Table 4: Overlap of esDSG predictions and the top-10 predictions generated by existing methods [3, 10]. Entries in bold indicate that all esDSG predictions are different

Method	NR		GPCR		IC		E	
	Equal	Diff.	Equal	Diff.	Equal	Diff.	Equal	Diff.
semEP	0	0	0	**30**	3	4	0	**1**
BLM	0	0	0	**30**	0	**7**	0	**1**
NBI	0	0	4	26	0	**7**	1	0
GIP	0	0	2	28	4	3	1	0
LapRLS	0	0	0	**30**	3	4	1	0
KBMF2K	0	0	0	**30**	0	**7**	0	**1**

subgraphs composed of both similar drugs and targets. We plan to extend esDSG to traverse the different connected components and identify esDSGs from each one.

References

1. K. Bleakley and Y. Yamanishi. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics*, 25(18):2397–2403, 2009.
2. F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, and Y. Tang. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS computational biology*, 8(5):e1002503, 2012.
3. H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu. Similarity-based machine learning methods for predicting drug-target interactions: A brief review. *Briefings in Bioinformatics*, 2013.
4. S. Fakhraei, B. Huang, L. Raschid, and L. Getoor. Network-based drug-target interaction prediction with probabilistic soft logic. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014.
5. M. Gönen. Predicting drug–target interactions from chemical and genomic kernels using bayesian matrix factorization. *Bioinformatics*, 28(18):2304–2310, 2012.
6. M. Ji, Y. Sun, M. Danilevsky, J. Han, and J. Gao. Graph regularized transductive classification on heterogeneous information networks. In *Machine Learning and Knowledge Discovery in Databases*, pages 570–586. Springer, 2010.
7. S. Khuller and B. Saha. On finding dense subgraphs. In *Automata, Languages and Programming*, pages 597–608. Springer, 2009.
8. M. Kuhn, D. Szklarczyk, A. Franceschini, C. von Mering, L. J. Jensen, and P. Bork. Stitch 3: zooming in on protein–chemical interactions. *Nucleic acids research*, 40(D1), 2012.
9. M. Narayanan, A. Vetta, E. E. Schadt, and J. Zhu. Simultaneous clustering of multiple gene expression and physical interaction datasets. *PLoS Computational Biology*, 6(4), 2010.
10. G. Palma, M. Vidal, and L. Raschid. Drug-target interaction prediction using semantic similarity and edge partitioning. In *ISWC*, 2014.
11. B. Saha, A. Hoch, S. Khuller, L. Raschid, and X. Zhang. Dense subgraphs with restrictions and applications to gene annotation graphs. In *RECOMB*, 2010.
12. Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu. Rankclus: integrating clustering with ranking for heterogeneous information network analysis. In *EDBT*, 2009.
13. A. Thor, P. Anderson, L. Raschid, S. Navlakha, B. Saha, S. Khuller, and X. Zhang. Link prediction for annotation graphs using graph summarization. In *ISWC*, 2011.
14. T. van Laarhoven, S. B. Nabuurs, and E. Marchiori. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics*, 27(21), 2011.
15. Z. Xia, L.-Y. Wu, X. Zhou, and S. T. Wong. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC systems biology*, 4(Suppl 2):S6, 2010.