Overlapping Community Detection on a Graph of Chemicals, Diseases and Genes for Drug Repositioning and Adverse Reactions Prediction

María Elena García-Ochagavía
mariae@matcom.uh.cu
Universidad de La Habana

Yudivián Almeida-Cruz
yudy@matcom.uh.cu
Universidad de La Habana

Suilán Estévez-Velarde
sestevez@matcom.uh.cu
Universidad de La Habana

Aimée Alonso-Reina
a.alonso@matcom.uh.cu
Universidad de La Habana

María Elena Ochagavía-Roque
ochagavia@nauta.cu
Centro de Ingeniería Genética y Biotecnología

ABSTRACT

Developing a drug from scratch is a very long and expensive process that has a small probability of success. For this reason, pharmaceutical companies are devoting their efforts to find drugs that could be repositioned. When using a drug to treat a disease is necessary to consider what adverse reactions it may cause, this is why the prediction of adverse reactions is highly related to drug repositioning. We propose the detection of overlapping communities over a biological network of chemicals, diseases and genes in order to find drug-disease pairs that could be used as basis for later drug repositioning and adverse reactions prediction analysis. Of the evaluated overlapping community detection algorithms, OSLOM got the best results, producing 724 communities from which was possible to extract 215944 drug-disease pairs not present in the analyzed graph. We illustrate the usefulness of this set through examples of associations between pairs found in the scientific literature.

KEYWORDS: Drug repositioning, adverse reactions, overlapping community detection, biological network

INTRODUCTION

Developing a drug from scratch is an expensive process that takes several years and has a very small probability of success. For this reason, several pharmaceutical companies are using the strategy of reusing licensed drugs for new medical indications, this practice is call drug repositioning.

Drug repositioning has several advantages over the traditional process of developing a drug. The primary advantage is that the drug to be reused have already passed a high number of toxicity tests and many of its side effects are known. Hence, a repositioned drug has a higher probability of success and the cost of developing it and introducing it in the market with a new functionality is lower. When administering a new drug to a patient, the doctor must consider the possible side effects it might have, for this reason the prediction of adverse reactions to drugs is a subject that is deeply related to drug repositioning.

Multiple strategies have been used to reposition drugs, such as the search of drugs with chemical similarity, the modeling of physical interaction and virtual screening of compound libraries (Eckert and Bajorath 2007). With these strategies the problem is reduced to find chemicals capable of reestablish the affected pathway. However, the complex biological network in which they interact is not taken into account.

The accumulation of biological data in recent years have made possible to start using biological networks as basis for the development of drug repositioning methods. The search of common molecular mechanisms between drugs and diseases and the use of supervised inference methods (Chen, Zhang et al. 2015) are some of the explored strategies.

Another strategy to drug repositioning using biological networks has been the detection of communities. In a previous work (Wu, Gudivada et al. 2013), a network was built from gene-disease and gene-drug relationships extracted from KEGG(Kanehisa and Goto 2000). In the built network, drugs and diseases are represented by nodes while an edge exists between two nodes if they share relationships with genes, biological processes, pathways or combinations of these characteristics. This network was used to detect non-overlapping communities of drugs and genes. Finally, from each community all drug-disease pairs were generated and proposed as candidates for a latter drug repositioning analysis.

When detecting communities on biological networks, its complex nature makes it likely that a node may belong to more than one community. Taking this into consideration, we propose a similar approach using overlapping communities instead of non-overlapping communities.

In the present work an analysis of overlapping community detection was performed over a graph of chemicals, diseases and genes generated from a database that integrates extensive biological information from multiple online publicly available sources: BisoPharma (Ochagavia, Martin et al. 2003).

OVERLAPING COMMUNITY DETECTION ALGORITHMS

A number of overlapping community detection algorithms have been developed using different strategies such as: clique percolation, line graph and link partitioning, local expansion and optimization, fuzzy detection and agent-based and dynamic algorithms. In the current section some of them will be reviewed.

The Clique Percolation Method (CPM)(Palla, Derenyi et al. 2005) assumes that a community is a set of overlapped fully connected subgraphs. CPM identifies all cliques of size *k* and creates a new graph where nodes represent the k-cliques in the original graph and an edge exists between two nodes if the k-cliques that they represent have *k-1* nodes in common. The connected components of the new graph determine the overlapping communities by considering a community as the union of the k-cliques represented by the nodes in a connected component. Various adaptations of CPM have been developed, such as CPMw, an adaptation of CPM for weighted graphs (Illés, Dániel et al. 2007).

Line Graph and Link Partitioning algorithms are based in the idea of partitioning edges instead of nodes. This way, if a node has two edges assigned to different clusters, this node belongs to both communities causing the communities to overlap. This approach has the advantage that allows hierarchical clustering in a simple manner by using a hierarchical disjoint community detection algorithm over the links (Ahn, Bagrow et al. 2010).

Local Expansion and Optimization algorithms are based on creating a natural community or a partial community by optimizing a local function that measures the quality of communities. Baumes (Baumes, Goldberg et al. 2005) proposes a combination of two algorithms to follow this approach: Ranck Removal(RaRe) and Iterative Scan(IS). RaRe finds small connected components that serve as seed communities to expand by IS by adding or removing nodes to optimize a local density function. LMF (Lancichinetti and Fortunato 2009) expands communities from randomly selected nodes. MONC (Frank, Michael et al. 2011) uses a modification of LFM's fitness function that allows a node to be a community by itself. OSLOM (Lancichinetti, Radicchi et al. 2011) uses the statistical significance as a metric to evaluate communities and calculates the statistical significance of a community as the probability to find it in the null model: the configuration model (Molloy and Reed 1995). EAGLE (Andrea, Santo et al. 2009) is an agglomerative algorithm that starts by the set of maximal cliques of size higher that a parameter *k*. It creates a dendrogram by merging the two more similar communities in each step. Finally, it cuts the dendrogram in the level that yields the highest value of a modified modularity measure to obtain the cover.

Fuzzy community detection algorithms model the association of a node to a community as a real number between zero and one that represents the strength of the association. They calculate a soft membership vector for each node. A disadvantage of these algorithms is the need to know *a priori* the length of the vector (the number of communities). Nepusz (Nepusz, Petroczi et al. 2008) models the fuzzy community detection problem as a nonlinear constrained optimization

problem that can be solved using simulated annealing methods. Zhanga (Zhanga, Wangb et al. 2006) proposes an approximate mapping of the graph into a Euclidian space and then uses a fuzzy c-means clustering to find the communities. Another approach to fuzzy community detection is the use of mixture models due to their probabilistic nature, this is the case of SPAEM (Derenyi, Palla et al. 2005).

Label Propagation Algorithm (LPA) (Raghavan, Albert et al. 2007) is a non-overlapping community detection algorithm where communities are detected by label propagation with an almost linear running time. It begins by assigning a label to each node and during each iteration each node is assigned the most frequent label among its neighbors. COPRA (Gregory 2010) is an LPA generalization for overlapping community detection that receives as a parameter the maximum numbers of communities that a node can belong to. Another adaptation of LPA is SLPA (Xie, Szymanski et al. 2011) which is based on a speaker-listener information propagation process.

QUALITY METRICS FOR OVERLAPPING COMMUNITIES

The selection of the best algorithm to find a cover for a specific network requires the use of metrics to evaluate the quality of the cover. Also, metrics that evaluate the quality of communities are necessary to find the most interesting communities for drug repositioning purposes.

Community quality metrics

The use of metrics to evaluate the quality of a cover is necessary in order to select the best algorithm. There is no exact definition of what constitutes a community, however there is a widely accepted notion that a community is a subset of nodes of the graph that are more densely connected among them that with the rest of the graph. The metrics shown in Table 1 are based on the previous notion.

Table 1 Community quality metrics

Metric	Formula		Description
Conductance (Leskovec, Lang et al. 2010)	$f(S) = \frac{c_S}{2m_S + c_S}$	*(1)*	Measures the fraction of total edge that points outside the cluster
Expansion (Leskovec, Lang et al. 2010)	$f(S) = \frac{c_S}{n_S}$	*(2)*	Measures the number of edges per node that point outside the cluster
Internal density (Leskovec, Lang et al. 2010)	$f(S) = 1 - \frac{m_S}{n_S(n_S - 1)/2}$	*(3)*	Measures the internal edge density of the cluster
Cut Ratio (Leskovec, Lang et al. 2010)	$f(S) = \frac{c_S}{n_S(n - n_S)}$	*(4)*	Measures the fraction of all possible edges leaving the cluster
Baumes Density (Baumes, Goldberg et al. 2005)	$f_{db}(S) = \frac{m_S}{m_S + c_S}$	*(5)*	Measures the fraction of the total of edges with both nodes inside the community with respect to the total number of edges with at least one node in the community
Difference between intra-cluster and extra-cluster density (Fortunato 2010)	$f_{dif}(S) = \frac{m_S}{n_S(n_S - 1)/2} - \frac{c_S}{n_S(n - n_S)}$	*(6)*	Is tradeoff be-tween internal density and cut Ratio

Legend: S is the community being evaluated, n_S is the number of nodes in S, m_S is the number of edges between nodes of S and c_S is the number of edges with exactly one node in S

When using the metrics *(3,(5(*6 the higher the value the better the community and in the resting metrics occurs the opposite. The metrics of equations *(1*, *(5* and *(6* were considered of interest in this work because all of them take into account both: that the nodes of the graph are well connected internally and poorly connected to the rest of the graph. However, the metrics *(1* and *5* are highly correlated, thus it would be redundant to use both.

Cover quality metrics

In the scientific literature there are several generally accepted metrics that allow to evaluate the quality of partitions, but that is not the case for evaluating the quality of covers. For this reason, several authors have used adaptations of metrics that where originally created to evaluate the quality of partitions.

The most popular quality function to evaluate partitions is Newman and Girvan Modularity (Girvan and Newman 2002). Several previous works have developed cover detection algorithms based in modularity optimization using adaptations of Newman and Girvan Modularity to allow community overlapping. However, none of these adaptations have become a generally used metric to evaluate the quality of covers.

The use of additive functions is another commonly used approach to evaluate partitions (Fortunato 2010). In this approach the quality of a partition is determined by the summation of the qualities of each community of the partition. When evaluating a partition, the number of summands, which is equal to the number of communities, is limited by the number of nodes in the graph, but this is not the case for covers, where the number of summands (communities) is only limited by the amount of possible subsets of nodes of the graph. This makes the use of additive functions not appropriate to evaluate covers without normalizing.

OVERLAPPING COMMUNITY FOR DRUG REPOSITIONING AND ADVERSE REACTIONS PREDICTIONS

This research proposes an overlapping community detection analysis on a graph of chemicals, disease and genes in order to find drug-disease pairs to be considered as candidates for drug repositioning or that could cause adverse reactions.

Graph of chemicals, diseases and genes (CDG)

The analysis of overlapping community detection in biological network proposed in this research was performed on a graph which nodes represent biological entities like chemicals, diseases and genes and its edges represent gene-chemical, gene-disease and chemical-disease associations. The information used to generate this graph was extracted form BisoPharma (Ochagavia, Martin et al. 2003) database.

BisoPharma is a system for biological network generation and visualization, composed by tree main components: a relational database, a web service and a Cytoscape's plugin (Shannon, Markiel et al. 2003) that plays the role of a client application. The BisoPharma database integrates data related to chemical, diseases and genes extracted from multiple online publicly available sources. This sources are: CTD (Davis, Grondin et al. 2015), OMIM (Amberger, Bocchini et al. 2015), GAD (Becker, Barnes et al. 2004), NHGRI GWAS Catalog (Welter, MacArthur et al. 2014), TTD (Yang, Qin et al. 2016), DRUGBANK (Wishart, Knox et al. 2006), SysBiomics (Martin, Ochagavia et al. 2010), DO (Kibbe, Arze et al. 2015) and MeSH[1].

[1] https://www.nlm.nih.gov/mesh

The graph obtained from BisoPharma has 79400 nodes and 722715 edges. It contains 31066 connected components, of which all except one have 6 nodes or less. The largest connected component has 47255 nodes and 721634 edges. The study of the small connected components of the graph has no value for an analysis of community detection, for this reason in this research only the largest component was considered. The proportions of each node type and edge type in the final graph are shown in the figures Figure 1 and Figure 2, respectively.

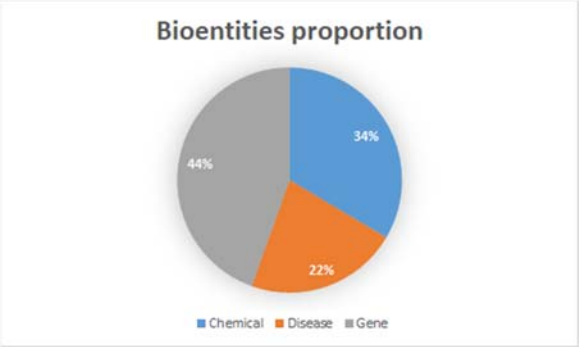

Figure 1 Bioentities proportion

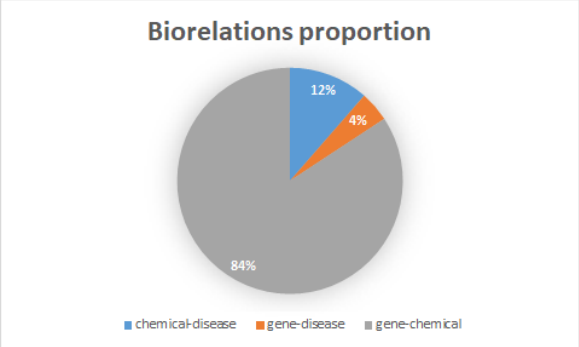

Figure 2 Biorelations proportion

The Cytoscape's plugin NetworkAnalyzer (Assenov, Ramírez et al. 2008) version 3.2.1 was used for a more thorough analysis of the graph. Figures Figure 3, Figure 4 and Figure 5 show the degree distribution, the minimum cost path distribution and the number of common neighbors' distribution respectively.

Cover detection in CDG

The selection of criteria to compare the selected algorithms was necessary to obtain the best cover. These criteria include the selection of metrics to evaluate the quality of communities and covers and the development of a strategy to compare the overall performance of the algorithms. The used methodologies are described in the following sections.

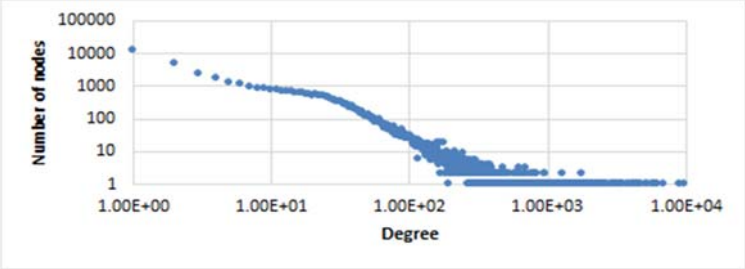

Figure 3 Node degree distribution

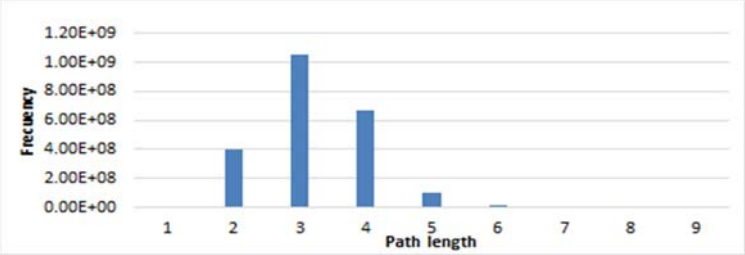

Figure 4 Minimum cost path distribution

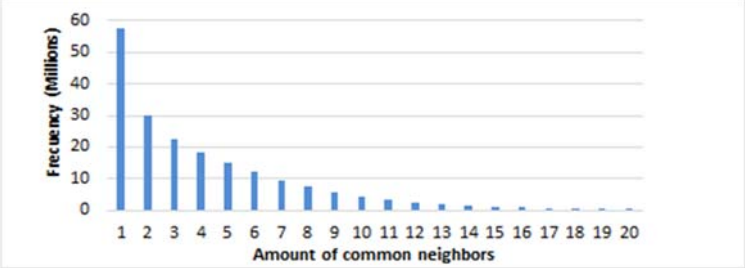

Figure 5 Common neighbors' distribution

Selection of community quality metric

The metrics Baumes density and difference between intra-cluster and extra-cluster density were used to evaluate the quality of the communities obtained from CDG. These metrics were selected because they consider a balance between a high amount of internal edges and a small amount of intra cluster edges, which fits the idea that a community is a set of nodes more connected internally than to the rest of the graph. However, for both metrics the resulting value is affected by the size of the community as shown in figure Figure 6 and Figure 7. For this reason, equal resulting values of the metrics, yielded from communities of different size, do not have the same meaning, thus it is not convenient to use them to compare communities of different sizes.

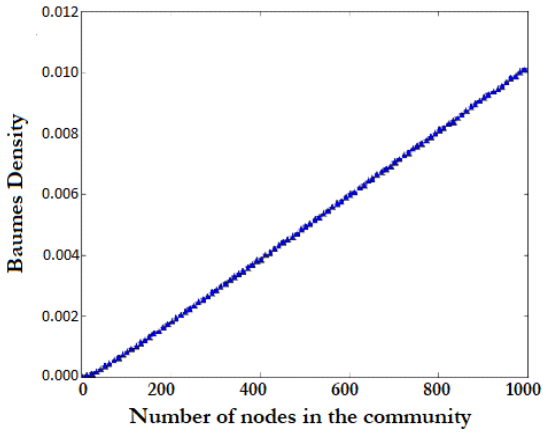

Figure 6 Relation between Baumes density and the size of a community

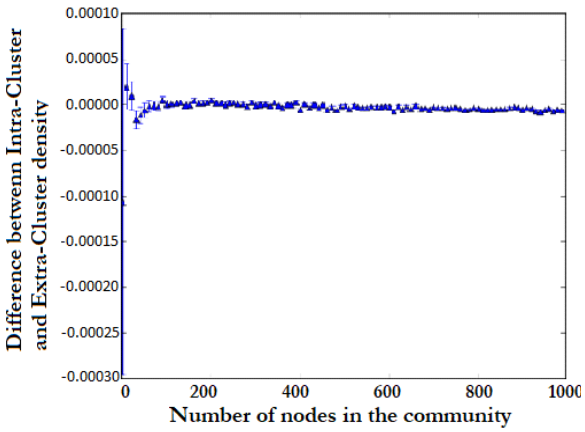

Figure 7 Relation between the Difference between intra-cluster and extra-cluster density and the size of a community

As a solution to this problem, a z-score for each community size was created for both metrics. A z-score measures how far is a sample value from the population's mean in units of the standard deviation. To build a z-score form the selected metrics, the mean and the standard deviation for subgraphs of size k were estimated generating random sets of nodes of size k for each possible community size k. The resulting metrics were:

$$f'_{db}(S) = \frac{f_{db}(S) - X^S_{db}}{\sigma^S_{db}} \quad (7)$$

$$f'_{dif}(S) = \frac{f_{dif}(S) - X^S_{dif}}{\sigma^S_{dif}} \quad (8)$$

Where $X^S_{db}(X^S_{dif})$ and $\sigma^S_{db}(\sigma^S_{dif})$ represent the estimated mean and standard deviation of $f_{db}(f_{dif})$ evaluated on graph of the same size as S.

Definition of cover quality criteria

Additive normalized metrics based on equations (7 and (8 where used to evaluate the quality of covers. The resulting formulas are:

$$Q_{db}(C) = \frac{\sum_{S \in C} f'_{db}(S)}{c} \quad (9)$$

$$Q_{dif}(C) = \frac{\sum_{S \in C} f'_{dif}(S)}{c} \quad (10)$$

For this work, it was decided to discard communities larger than 1000 because it is computationally expensive to estimate the mean and the standard deviation for large community sizes. Also, the objective of this work is to use the resulting communities for a later analysis of drug repositioning, for which loo large communities are not convenient.

When discarding communities of size higher than 1000 is likely that some nodes are not assigned to a community. This makes necessary to take into account the amount of nodes not assigned to any community when comparing covers.

Another criterion that is necessary to consider is the amount of communities of a single node (singleton).

Summing up, the criteria used to evaluate covers are:

1- The average of the z-score of Baumes density of each community
2- The average of the z-score of the difference between intra-cluster and extra-cluster density of each community
3- Amount of nodes not assigned to any community (assigned only to communities with more than 1000 nodes).
4- Amount of singletons

Experiments design

The algorithms OSLOM, COPRA and SLPA were selected to be used on the graph CDG because of their good performance over sparse networks (Xie, Kelley et al. 2013). Several parameter values were evaluated for each method in order to obtain their optimal behavior in the community detection goal.

In OSLOM the *p-value* was set to 0.05, 0.10 and 0.15. This parameter establishes the significance threshold for communities. The cover parameter, which is used by the algorithm to decide to keep some modules or their union, was evaluated for values 0.25, 0.5 and 0.75.

In COPRA only the parameter *v* was explored, this paremeter determines the maximum number of communities to which a node can belong. I was set to values ranging from 1 to 10 with a step of 1.

In SLPA parameters *r* and *v* were investigated. In SPLA each node receives a value between 0 and 1 which represents its degree of belonging to each community. The parameter *r* establishes a belonging threshold that determines the binary assignment of a node to a community. The *v* parameter establishes the selected interaction rule and can take values 1, 2 or 3 each of which represents a specific rule. In this work SLPA was ran using all possible interaction rules and *r* values of 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45 and 0.5.

To obtain the best parameter combination for each method the criteria discussed in the previows section were combined as follows. First, each combination was evaluated using each criterion separately to create a ranking per criterion and, afterward, a general ranking was created summing the ranks of each combination regarding each criterion. Finally, to select the best method, the covers yielded by running each algorithms with the best parameter combination, were compared using the same methodology.

RESULTS AND DISCUSSIONS

The results of evaluating the selected parameter combinations using the previously discused criteria yielded that the best combinations were $v = 3$ and $r = 0.45$ for SLPA, *p-value* = 0.10 and cover = 0.25 for OSLOM and $v = 14$ for COPRA.

Each algorithm was excecuted 30 times using the best parameter combinations to obtain statisticaly significant results, which are shown in Table 2. The best results were obtained from the algorithm OSLOM, because it did not produced singletons, only a small number of nodes were not assigned to communities and it yielded the best result of the difference between intra-cluster and extra-cluster density. The best results of Baumes density were obtained by COPRA, however, it produced a high number of singletons and obtained the worst results of the difference between intra-cluster and extra-cluster density.

Table 2 Algorithms Results

Algorithm	NAN	NANR	S	SR	B	BR	DIE	DIER	Sum
OSLOM	595.43	1	0	1	22837.92	3	9983.84	1	6
SLPA	15982.67	3	0	1	34786.39	2	6740.79	2	8
COPRA	7598.57	2	39250	3	182057.34	1	5753.40	3	9

Legend: NAN is the number of nodes not assigned to any community, NANR is the NAN ranking, S is the number of singletons, SR is the S ranking, B is the Baumes Density, BR is the B Ranking, DIE is the Difference between intra-cluster and extra-cluster density and DIER is DIE Ranquing.

The cover proposed to perform drug reposition analysis was generated by OSLOM because it obtained the best results of the metrics. The cover contains 724 communities, most of which have an appropriated size for visual inspection (between 50 and 150 nodes as can be seen in Figure 8 Community size distributionFigure 8). Figure 9 shows the Baumes density distribution while Figure 10 shows the difference between intra-cluster and extra-cluster density distribution.

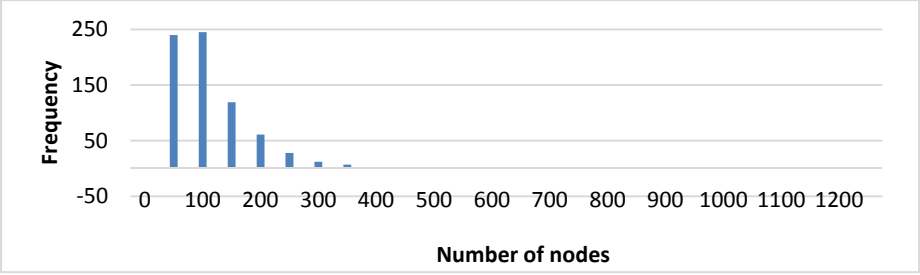

Figure 8 Community size distribution

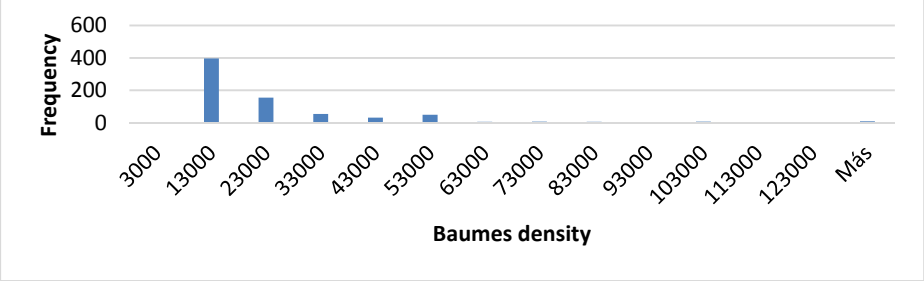

Figure 9 Baumes density distribution

In this work it is proposed the list of all drug-disease pairs assigned to the same community as candidates to perform drug repositioning analysis. This is based in the idea that drugs and diseases in the same community are likely to share molecular mechanisms.

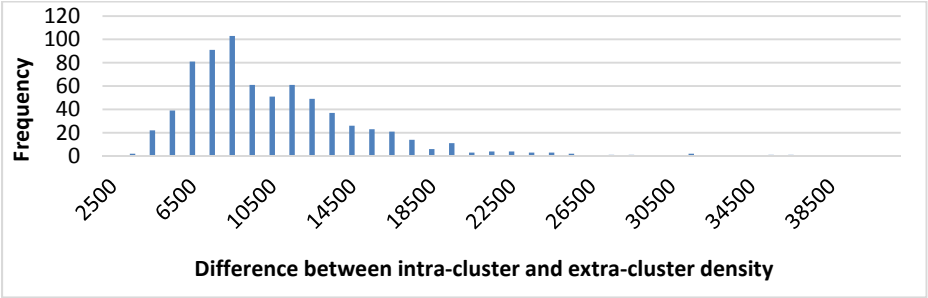

Figure 10 Difference between intra-cluster and extra-cluster density distribution

The list contains 215944 drug-disease pairs of which only 15406 (7.13%) exist in BisoPharma, leaving a large amount of pairs to be explored. In order to illustrate the usefulness of this cover, some communities were analyzed more thoroughly, and some associations of the proposed pairs were found in the literature. **¡Error! No se encuentra el origen de la referencia.**3 shows some examples of these associations.

Table 3 Drug-disease associations fount in the literature

Drug	Disease	Association Type	Reference
Atenolol	Schizophrenia	Repositioning	[2]
Atenolol	Akathisia, drug induced	Repositioning	[4]
Teicoplanin	Cellulitis	Repositioning	(Turpin, Taylor et al. 1988)
Etanercept	Parakeratosis	Adverse Reaction	(Echeverri, Vidal et al. 2015)
Levodopa	Carcinoma and squamous cell	Adverse Reaction	[3]
Efedrina	Parkinson Disease	Adverse Reaction	(Sikk, Haldre et al. 2011)
Leucovorin	Dariel Disease	Repositioning	(Holcmann and Sibilia 2015)

In the following paragraphs some of these associations will be explained with more detail. The community shown in **¡Error! No se encuentra el origen de la referencia.** contains Atenolol

[2] http://www.schizophrenia.com/szresearch/archives/001174.html
[3]http://www.labome.org/research/Basal-cell-carcinoma-and-squamous-cellcarcinoma-in-a-patient-with-Parkinson-disease.html

and Schizophrenia with no edge between them, nevertheless the search in the literature of this pair showed the following: when treating schizophrenic patients is common the use of antipsychotics which may have as a side effect akathisia. Akathisia is a state of severe restlessness that could lead to suicide. One of the treatments for akathisia is the oral administration of beta-blockers like Atenolol. This is an example of the use of Atenolol with positive effect on patients with Schizophrenia.

The community shown in 2 contains Leucovorin and Darier's disease without a direct link between them. However, the research of the literature yielded a patent (Holcmann and Sibilia 2015) that protects the treatment with epidermal growth factor receptor (EGFR) inhibitors, like Leucovorin, of genetic skin disorders that exhibit a high percentage of penetrance, or complete penetrance, such as Darier's disease. This is an example of successful repositioning of the drug Leucovorin to treat Darier's disease.

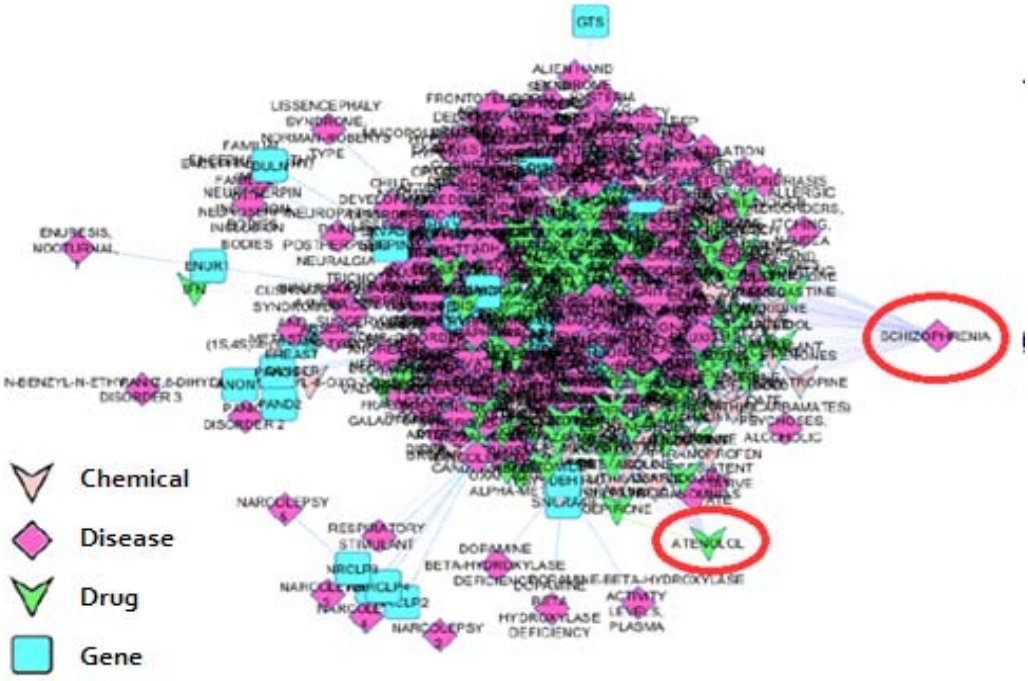

Figure 11 Association between Atenolol and Schizophrenia

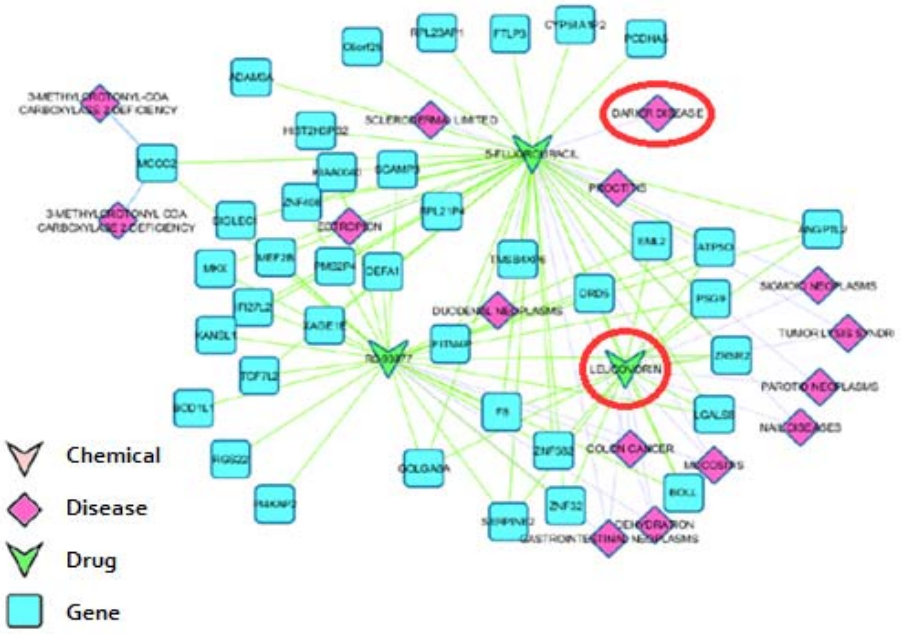

Figure 12 Association between Leucovorin and Darier's disease

CONCLUSIONS

Recently, several pharmaceutical companies are devoting their efforts to find drugs that could be repositioned. This is because the traditional process of developing a drug is very long, expensive and has a small probability of success. When considering a drug to be repurposed to treat another disease, it is necessary to take into account the adverse reactions that it may cause when targeting this disease. For this reason, the prediction of adverse reactions is highly related to the search of drugs candidates to be repositioned.

We approached the subject of finding drugs candidates to be repositioned and adverse reaction prediction through an overlapping community detection analysis of a complex biological network built from BisoPharma database.

The algorithms considered to detect communities over the graph were OSLOM, SLPA and COPRA. These algorithms were compared using additive normalized metrics based on Baumes density and the difference between intra-cluster and extra-cluster density. The evaluation of the covers resulting from these algorithms indicated that OSLOM generated the best cover.

Finally, we provide the list of drug-disease pairs contained in the same community of the obtained cover to be used as base for drug repositioning analysis and adverse reaction prediction. The usefulness of this list was illustrated through examples of associations between pairs found in the scientific literature.

REFERENCES

Ahn, Y. Y., J. P. Bagrow and S. Lehmann (2010). "Link communities reveal multiscale complexity in networks." Nature 466(7307): 761-764.

Amberger, J. S., C. A. Bocchini, F. Schiettecatte, A. F. Scott and A. Hamosh (2015). "OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders." Nucleic acids research 43(Database issue): D789-D798.

Andrea, L., F. Santo and K. János (2009). "Detecting the overlapping and hierarchical community structure in complex networks." New Journal of Physics 11(3): 033015.

Assenov, Y., F. Ramírez, S.-E. Schelhorn, T. Lengauer and M. Albrecht (2008). "Computing topological parameters of biological networks." Bioinformatics 24 2: 282-284.

Baumes, J., M. K. Goldberg, M. S. Krishnamoorthy, M. Magdon-Ismail and N. Preston (2005). Finding communities by clustering a graph into overlapping subgraphs. IADIS AC. N. Guimarães and P. T. Isaías, IADIS: 97-104.

Becker, K. G., K. C. Barnes, T. J. Bright and S. A. Wang (2004). "The Genetic Association Database." Nature Genetics 36: 431.

Chen, H., H. Zhang, Z. Zhang, Y. Cao and W. Tang (2015). "Network-Based Inference Methods for Drug Repositioning." Computational and Mathematical Methods in Medicine 2015: 7.

Davis, A. P., C. J. Grondin, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B. L. King, T. C. Wiegers and C. J. Mattingly (2015). "The Comparative Toxicogenomics Database's 10th year anniversary: update 2015." Nucleic acids research 43(Database issue): D914-D920.

Derenyi, I., G. Palla and T. Vicsek (2005). "Clique percolation in random networks." Phys Rev Lett 94(16): 160202.

Echeverri, A. F., A. Vidal, C. A. Canas, A. Agualimpia, G. J. Tobon and F. Bonilla-Abadia (2015). "Etanercept-induced pityriasis lichenoides chronica in a patient with rheumatoid arthritis." Case Rep Dermatol Med 2015: 168063.

Eckert, H. and J. Bajorath (2007). "Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches." Drug discovery today 12 5-6: 225-233.

Fortunato, S. (2010). "Community detection in graphs." Physics Reports 486(3): 75-174.

Frank, H., H. Michael, S. Alexander and G. Jochen (2011). "Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels." Journal of Statistical Mechanics: Theory and Experiment 2011(01): P01023.

Girvan, M. and M. E. Newman (2002). "Community structure in social and biological networks." Proc Natl Acad Sci U S A 99(12): 7821-7826.

Gregory, S. (2010). "Finding overlapping communities in networks by label propagation." New Journal of Physics 12(10): 103018.

Holcmann, M. and M. Sibilia (2015). "Mechanisms underlying skin disorders induced by EGFR inhibitors." Molecular & cellular oncology 2(4): e1004969-e1004969.

Illés, F., Á. Dániel, P. Gergely and V. Tamás (2007). "Weighted network modules." New Journal of Physics 9(6): 180.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res 28(1): 27-30.

Kibbe, W. A., C. Arze, V. Felix, E. Mitraka, E. Bolton, G. Fu, C. J. Mungall, J. X. Binder, J. Malone, D. Vasant, H. E. Parkinson and L. M. Schriml (2015). Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic Acids Research.

Lancichinetti, A. and S. Fortunato (2009). "Community detection algorithms: a comparative analysis." Phys Rev E Stat Nonlin Soft Matter Phys 80(5 Pt 2): 056117.

Lancichinetti, A., F. Radicchi, J. J. Ramasco and S. Fortunato (2011). "Finding statistically significant communities in networks." PloS one 6(4): e18961-e18961.

Leskovec, J., K. J. Lang and M. Mahoney (2010). Empirical comparison of algorithms for network community detection. Proceedings of the 19th international conference on World wide web. Raleigh, North Carolina, USA, ACM: 631-640.

Martin, A., M. E. Ochagavia, L. C. Rabasa, J. Miranda, J. Fernandez-de-Cossio and R. Bringas (2010). "BisoGenet: a new tool for gene network building, visualization and analysis." BMC bioinformatics 11: 91-91.

Molloy, M. and B. Reed (1995). "A critical point for random graphs with a given degree sequence." Random Struct. Algorithms 6(2-3): 161-180.

Nepusz, T., A. Petroczi, L. Negyessy and F. Bazso (2008). "Fuzzy communities and the concept of bridgeness in complex networks." Phys Rev E Stat Nonlin Soft Matter Phys 77(1 Pt 2): 016107.

Ochagavia, M. E., A. Martin, F. Torres, J. Miranda, J. Fernández-de-Cossio, C. Suárez, J. M. Fernández and R. Bringas; (2003). BisoPharma: An integrated system for gene-disease-drug network building.

Palla, G., I. Derenyi, I. Farkas and T. Vicsek (2005). "Uncovering the overlapping community structure of complex networks in nature and society." Nature 435(7043): 814-818.

Raghavan, U. N., R. Albert and S. Kumara (2007). "Near linear time algorithm to detect community structures in large-scale networks." Physical review. E, Statistical, nonlinear, and soft matter physics 76 3 Pt 2: 036106.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res 13(11): 2498-2504.

Sikk, K., S. Haldre, S. M. Aquilonius and P. Taba (2011). "Manganese-Induced Parkinsonism due to Ephedrone Abuse." Parkinsons Dis 2011: 865319.

Turpin, P. J., G. P. Taylor, M. N. Logan and M. J. Wood (1988). "Teicoplanin in the treatment of skin and soft tissue infections." Journal of Antimicrobial Chemotherapy 21(suppl_A): 117-122.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T.

Manolio, L. Hindorff and H. Parkinson (2014). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations." Nucleic Acids Res 42(Database issue): D1001-1006.

Wishart, D. S., C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang and J. Woolsey (2006). "DrugBank: a comprehensive resource for in silico drug discovery and exploration." Nucleic Acids Res 34(Database issue): D668-672.

Wu, C., R. C. Gudivada, B. J. Aronow and A. G. Jegga (2013). "Computational drug repositioning through heterogeneous network clustering." BMC Systems Biology 7(5): S6.

Xie, J., S. Kelley and B. K. Szymanski (2013). "Overlapping community detection in networks: The state-of-the-art and comparative study." ACM Comput. Surv. 45(4): 1-35.

Xie, J., B. K. Szymanski and X. Liu (2011). "SLPA: Uncovering Overlapping Communities in Social Networks via a Speaker-Listener Interaction Dynamic Process." 2011 IEEE 11th International Conference on Data Mining Workshops: 344-349.

Yang, H., C. Qin, Y. H. Li, L. Tao, J. Zhou, C. Y. Yu, F. Xu, Z. Chen, F. Zhu and Y. Z. Chen (2016). "Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information." Nucleic Acids Res 44(D1): D1069-1074.

Zhanga, S., R.-S. Wangb and X.-S. Zhanga (2006). Identification of overlapping community structure in complex networks using fuzzy c-means clustering.

Editor in Chief: Prof. Dr. Luis Camilo Ortigueira-Sánchez