



NETWORK DATA MINING (Graph mining)

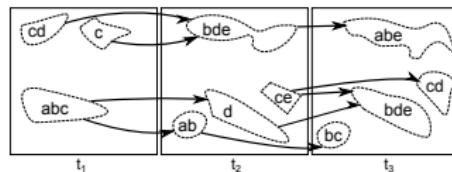
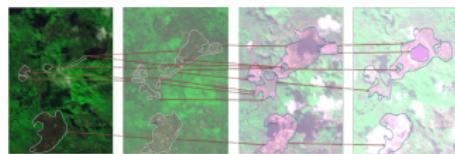
Claude Pasquier

I3S Laboratory

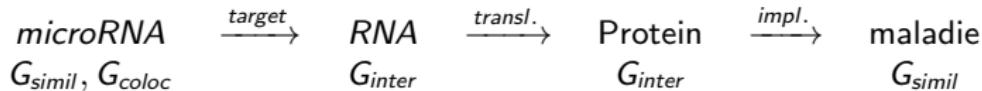
25 février 2019

Data structures

- Elements (nodes/vertices) connected by links
- links (edges) can be directed or undirected
- edges and vertices can be associated with one or more continuous or discrete values (attributes)
- nodes and edges can be associated with different types
- example : spatio-temporal data



- example : biological data



Summary of the presentation

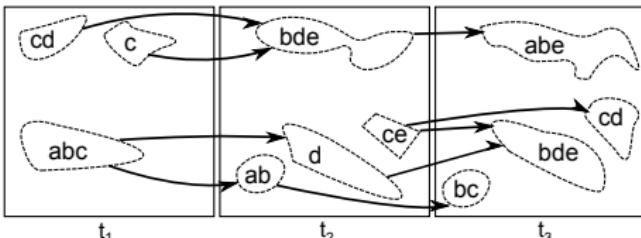
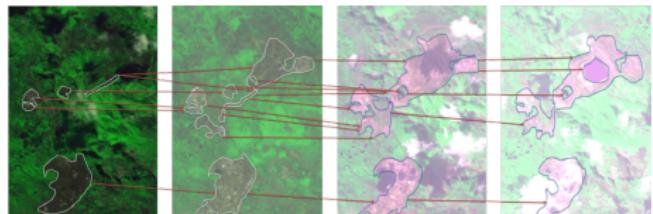
Two different topics:

- Identification of frequent patterns in attributed graphs
- Prediction of associations between microRNAs and diseases

A few words about a work in progress

- Identification of active biological modules

TOPIC 1 : Identification of frequent patterns in attributed graphs

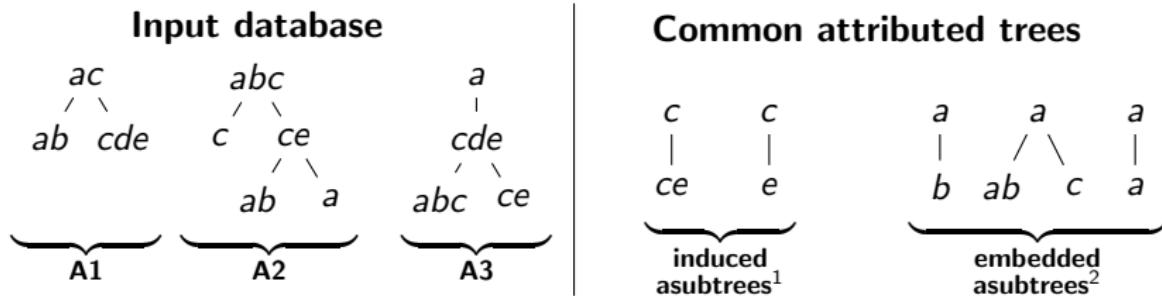


- Input data: a database \mathcal{D} of attributed graphs (**a**graphs)
 - ▶ an agraph $T = (V, E, \lambda)$ with:
 - ★ (V, E) , the underlying graph (V =vertices, E =edges)
 - ★ $\lambda : V \rightarrow D$, association of itemsets to vertices
- Problem type: enumerating all frequent patterns (**a**subgraphs)
- This work combines itemset mining and structural mining

Specificities of our problem

- the itemsets associated to nodes are not necessarily identical
- patterns can be embedded (implicit)

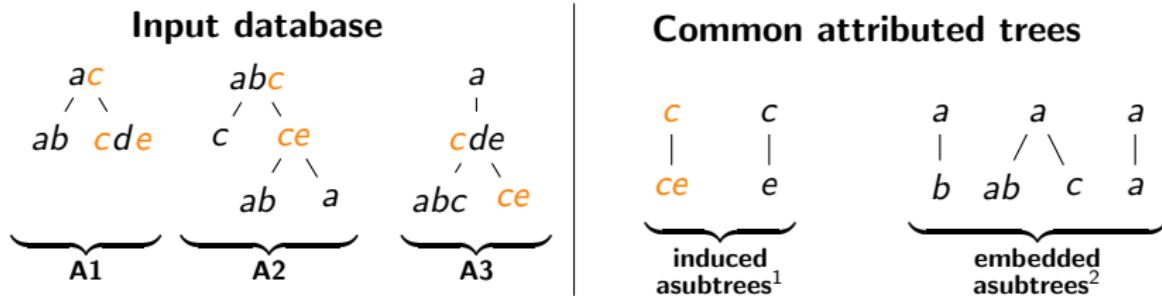
Example on attributed trees



¹isomorphic asubtree preserving the parent-child relationship

²isomorphic asubtree preserving the ancestor-descendant relationship

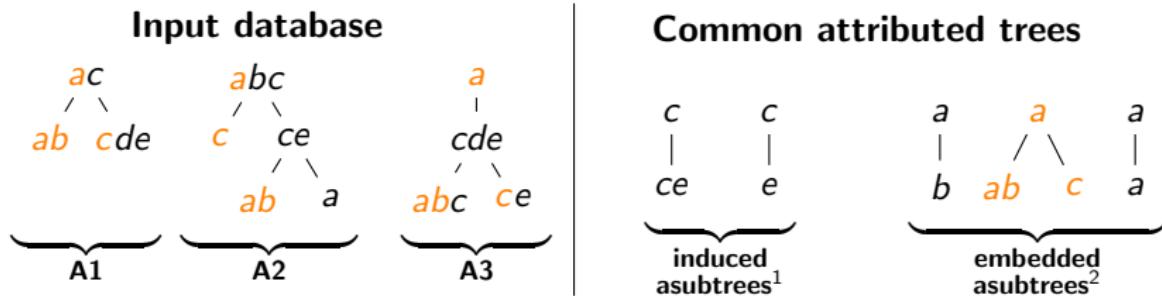
Induced asubtrees



¹isomorphic asubtree preserving the parent-child relationship

²isomorphic asubtree preserving the ancestor-descendant relationship

Embedded asubtrees



¹isomorphic asubtree preserving the parent-child relationship

²isomorphic asubtree preserving the ancestor-descendant relationship

- itemsets mining (Apriori algorithm) [[Agrawal 1994](#)]
- sequential pattern mining [[Agrawal 1995](#)]
- labeled graph mining [[Inokuchi 2000](#)]
- mining of labeled graphs with quantitative attributes [[Miyoshi, 2009](#)]
- mining of subgraphs whose nodes share a set of attributes [[Moser, 2008; Fukuzaki, 2010; Mougel, 2012](#)]
- Identification of correlations between attributes and the graph structure [[Silva, 2012](#)]

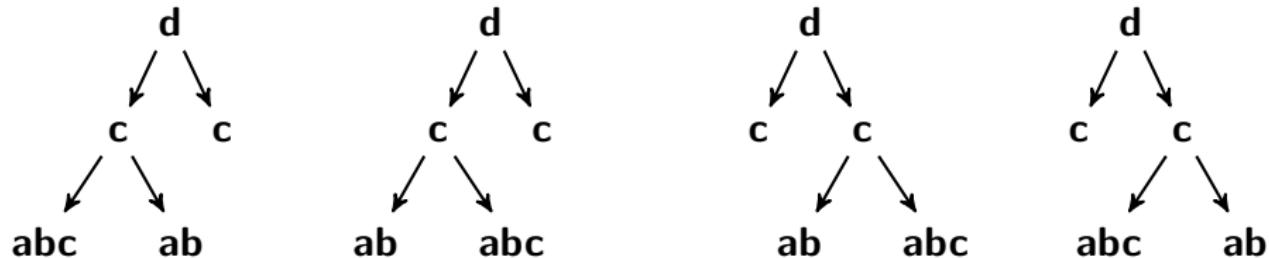
Difficult problem that combines graph mining and itemset mining

⇒ we started by dealing with a simpler problem: attributed tree mining

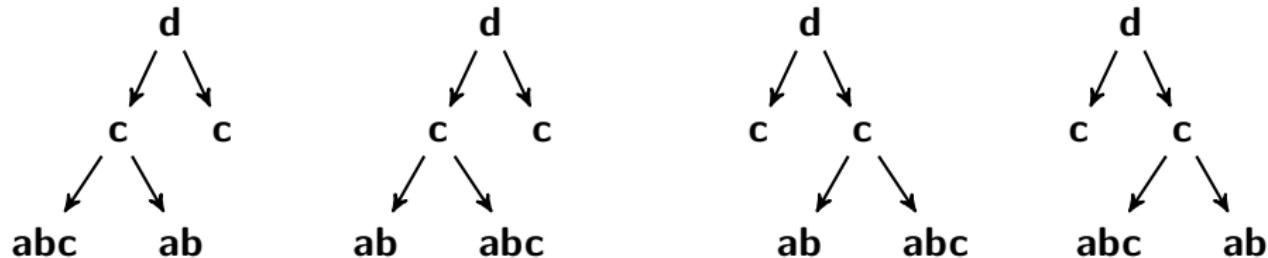
Requirements

- Definition of a systematic, complete and non-redundant enumeration method
- Development of a pruning strategy to avoid exploring unpromising solutions

Definition of an order on itemsets, nodes and attributed trees

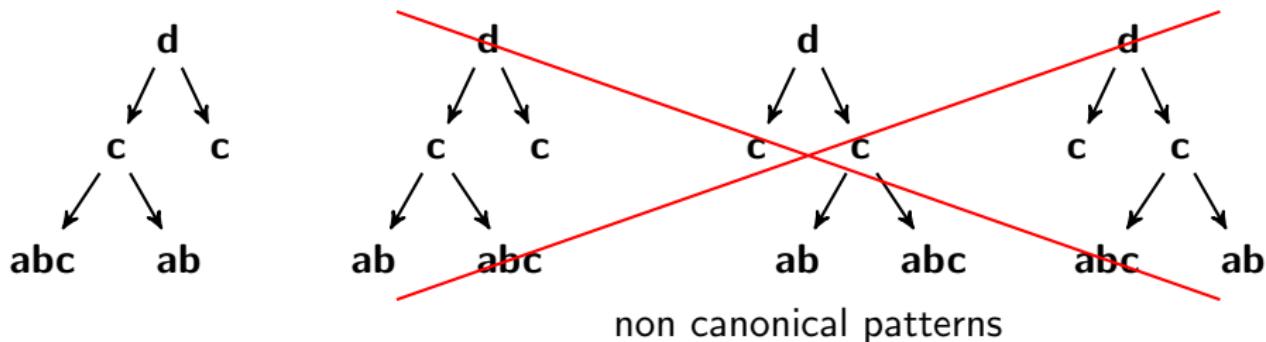


Definition of an order on itemsets, nodes and attributed trees



- Ordering of nodes based on node's associated itemsets
 - Given two itemsets P and Q ($P \neq Q$), $P < Q$ iff:
 - $\forall i \in [1, \min(|P|, |Q|)] : P_i \leq Q_i$
 - if $\forall i \in [1, \min(|P|, |Q|)] : P_i = Q_i$, then $|P| > |Q|$
 - Example: $abc < ab < ac < a < b$
- From nodes' ordering, an ordering, \prec , among atrees is defined
- Canonical form of isomorphic attributed trees is deduced from atrees' ordering

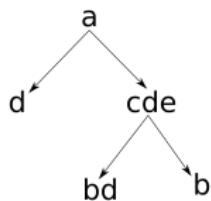
Definition of an order on itemsets, nodes and attributed trees



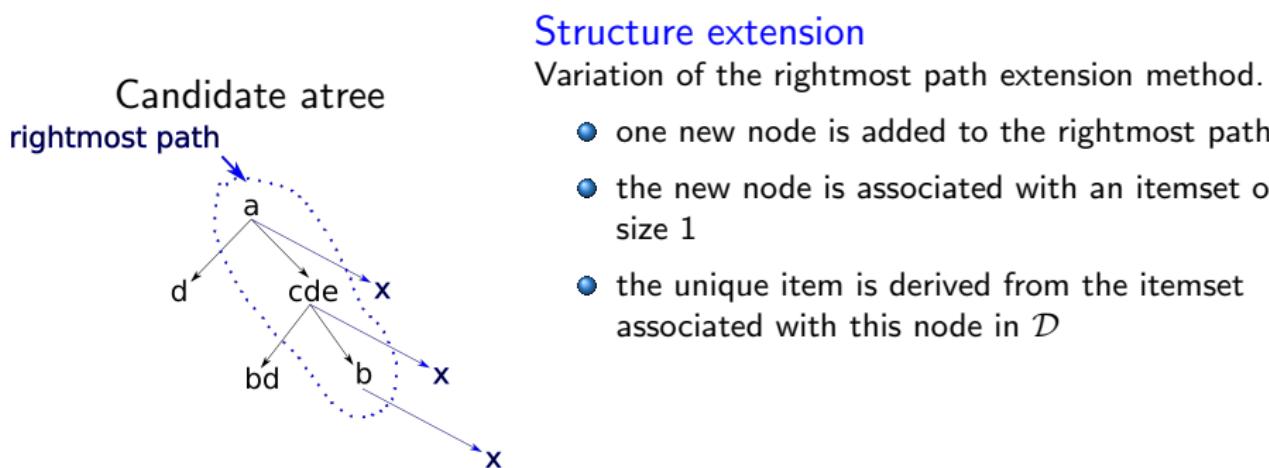
- Ordering of nodes based on node's associated itemsets
 - Given two itemsets P and Q ($P \neq Q$), $P < Q$ iff:
 - $\forall i \in [1, \min(|P|, |Q|)] : P_i \leq Q_i$
 - if $\forall i \in [1, \min(|P|, |Q|)] : P_i = Q_i$, then $|P| > |Q|$
 - Example: $abc < ab < ac < a < b$
- From nodes' ordering, an ordering, \prec , among atrees is defined
- Canonical form of isomorphic attributed trees is deduced from atrees' ordering

- ↲ is used to explore the complete search space
- New candidates are generated either:
 - ▶ by structure extension
 - ▶ by itemset extension

Candidate atree



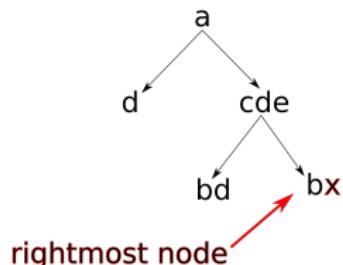
- \nwarrow is used to explore the complete search space
- New candidates are generated either:
 - ▶ by structure extension
 - ▶ by itemset extension



- \prec is used to explore the complete search space
- New candidates are generated either:
 - ▶ by structure extension
 - ▶ by itemset extension

Itemset extension

Candidate atree



For each extension:

- one new items is added to the rightmost node
- this new item is derived from the itemset associated with this node in \mathcal{D}

- Depth-first traversal of the search space
- Enumeration is complete and non-redundant
- Canonical checks required for unordered mining
- Stop condition of the exploration of a branch:
 - ▶ When candidate's per-tree-frequency < minimum support
 - ▶ When candidate is not in canonical form (unordered mining)

IMIT algorithm

IMIT(\mathcal{D} , $minSup$)

```
1:  $\mathcal{C} \leftarrow \{all\ asubtrees\ of\ size\ 1\ in\ \mathcal{D}\}$  //  $\mathcal{C}$  is the set of all candidates
2:  $\mathcal{S} \leftarrow \emptyset$  //  $\mathcal{S}$  is the set of solutions
3: while  $\mathcal{C} \neq \emptyset$  do
4:    $T \leftarrow getFirst(\mathcal{C})$  //  $T$  is the smallest candidate according to  $\prec$ 
5:    $\mathcal{C} \leftarrow \mathcal{C} \setminus \{T\}$  // removes  $T$  from candidates
6:   if  $isCanonical(T)$  and  $frequency(T) \geq minSup$  then
7:      $\mathcal{S} \leftarrow \mathcal{S} \cup \{T\}$  // add  $T$  to solutions
8:      $\mathcal{X} \leftarrow extend(T)$  // compute the extensions of  $T$ 
9:      $\mathcal{C} \leftarrow \mathcal{C} \cup \mathcal{X}$  // add extensions to candidates
10:  end if
11: end while
12: print  $\mathcal{S}$ 
```

IMIT algorithm

IMIT enumerates all solutions

Problem:

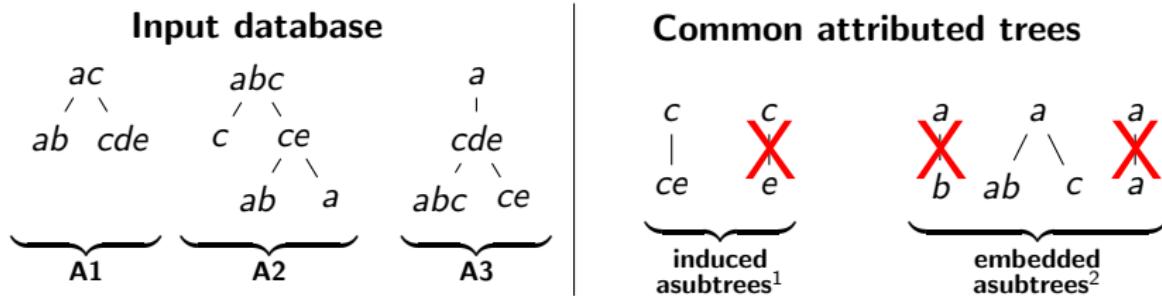
- the search space is huge
- a significant proportion of the solutions are redundant

Example:

- synthetic dataset
- 10,000 atrees
- itemsets of size 3
- minimum support=1%

12 millions frequent patterns in 15 hours !

Concise representation of patterns



¹isomorphic asubtree preserving the parent-child relationship

²isomorphic asubtree preserving the ancestor-descendant relationship

An atree T is a **closed attributed tree** if none of its attributed supertrees has the same support as T

By enumerating only closed atrees, the search space can be significantly reduced

IMIT_CLOSED algorithm

IMIT_CLOSED enumerates all closed solutions

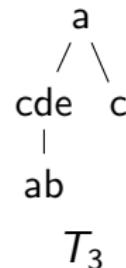
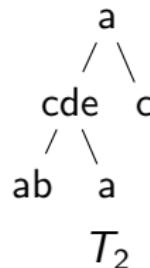
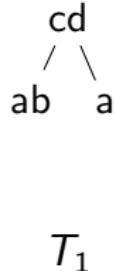
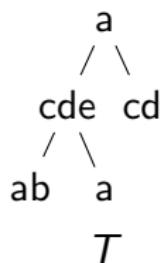
Example: 10,000 atrees, itemsets of size 3, minimum support=1%
103 frequent patterns (vs. 12 millions) in 3 seconds (vs. 15 hours)

Problem:

- The algorithm stores all frequent patterns found
- High cost of numerous subtree isomorphism checks

Example: 10,000 atrees, itemsets of size 3, min support=0.05%
3,219 frequent closed patterns in 33 hours!

Idea: distinguish structural inclusion from itemset inclusion



$T_1 \sqsubset T$
isomorphic
asubtree

$T_2 \sqsubset_I T$
isomorphic
subtree
defined by
itemset
inclusion

$T_3 \sqsubset_S T$
isomorphic
subtree
defined by
structural
inclusion

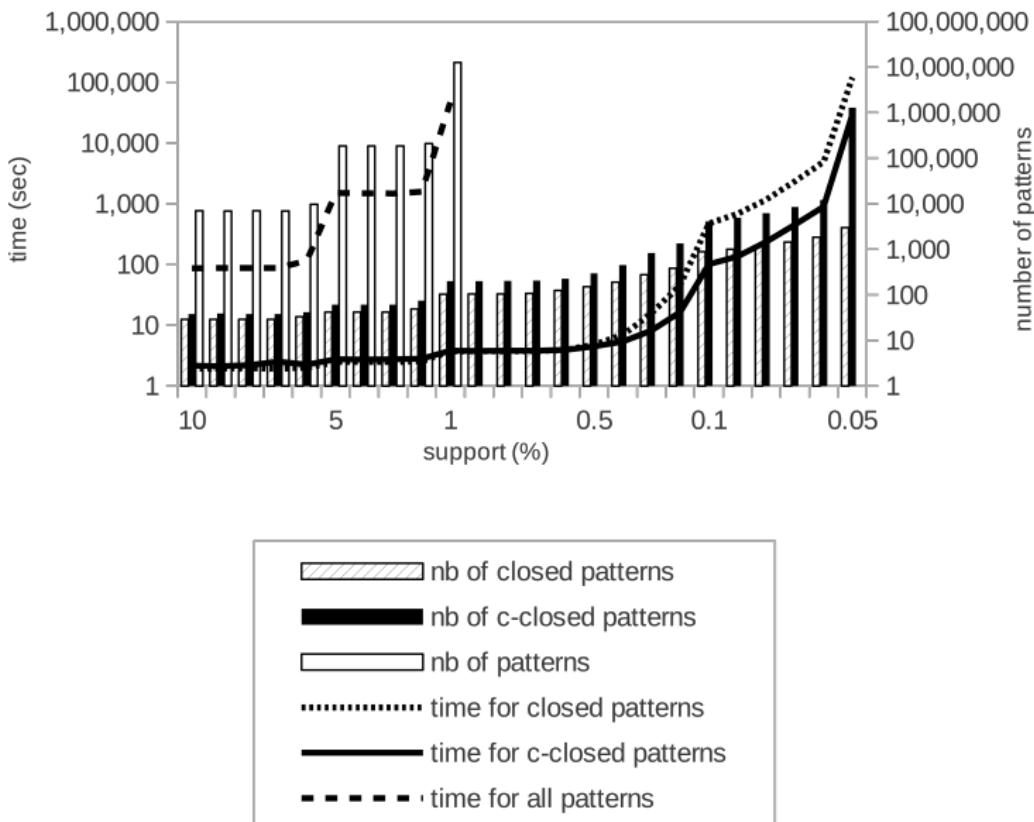
A new condensed representation is defined with respect to the itemset inclusion only (\sqsubseteq_I)

an atree T is a c-closed attributed tree (content closed) if it is not contained in (\sqsubseteq_I) another atree with the same support as T

Advantages:

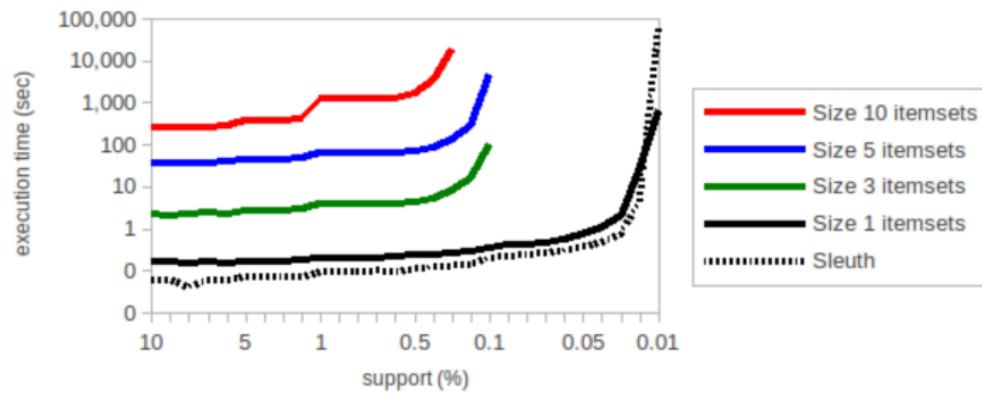
- No need to perform isomorphism checks; only itemset inclusion is tested
- The c-closure test is performed only on candidates generated by itemset extension
- No need to store all frequent patterns found

Induced unordered mining of 10,000 atrees with itemsets of size 3



Performances of IMIT_CONTENT_CLOSED

Induced unordered mining with itemsets of different sizes



Performance is similar to SLEUTH (Zaki 2004) for itemsets of size 1 (labeled trees)

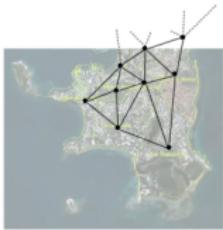
Size of itemsets has a big impact on performances

Application to the mining of dengue fever outbreak

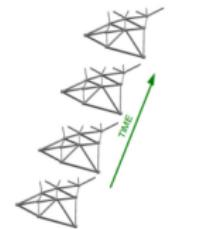
[Pasquier et al. 2015, Knowledge and Information Systems]



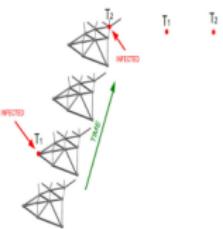
The city is divided in districts characterized by 15 epidemic and environmental attributes.



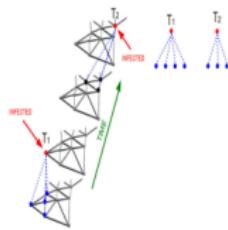
The situation at each time point is represented by an attributed graph.



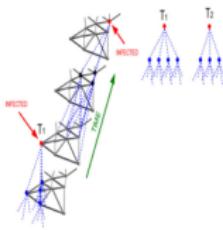
The whole dataset is represented by a succession of attributed graphs.



The roots of attributed trees correspond to infected districts.



Nodes that are just below a root correspond to neighbor zones at time $t - 1$.



Nodes at n^{th} level below a root represent neighbor districts at time $t - n$.

each node is associated to 15 attributes

181 atrees representing the spatio-temporal evolution of attributes before a spike of dengue fever

use of a support of 70%

117 induced patterns mined in 19mn

9265 embedded patterns found in 19h

8 discriminative patterns confirming combinations of factors already known

Extension of the method to the mining of attributed graphs

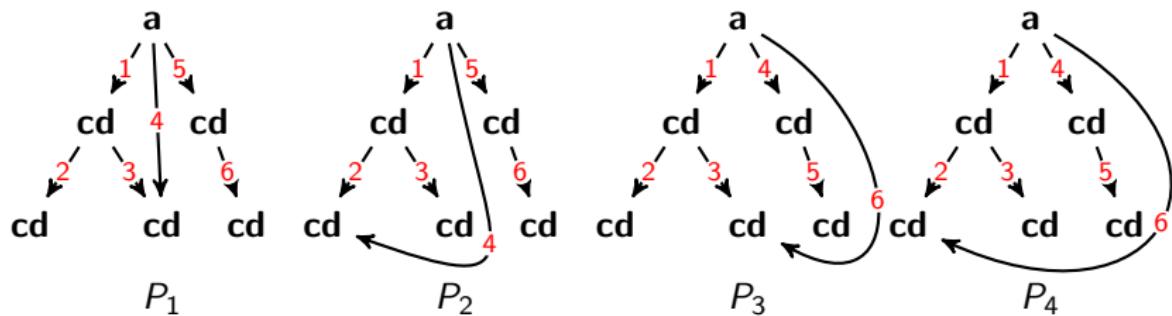
[Pasquier et al. 2017, Knowledge and Information Systems]

- proposal of a subgraph mining strategy based on the enumeration of spanning trees

Extension of the method to the mining of attributed graphs

[Pasquier et al. 2017, Knowledge and Information Systems]

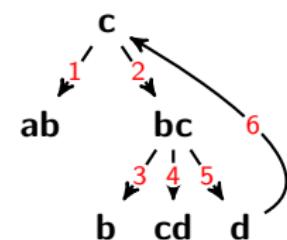
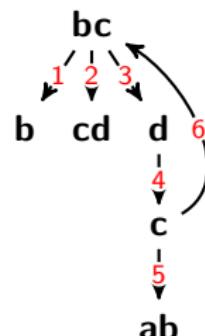
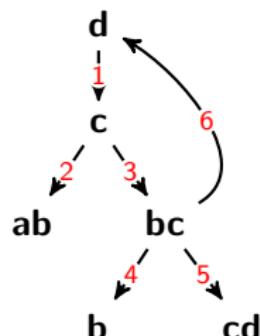
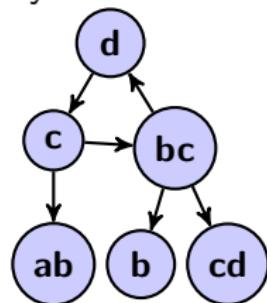
- proposal of a subgraph mining strategy based on the enumeration of spanning trees
- definition of a new canonical form handling reentrant links



Extension of the method to the mining of attributed graphs

[Pasquier et al. 2017, Knowledge and Information Systems]

- proposal of a subgraph mining strategy based on the enumeration of spanning trees
- definition of a new canonical form handling reentrant links
- New enumeration method taking into account reentrant links and cycles



Graph G

pattern P_1

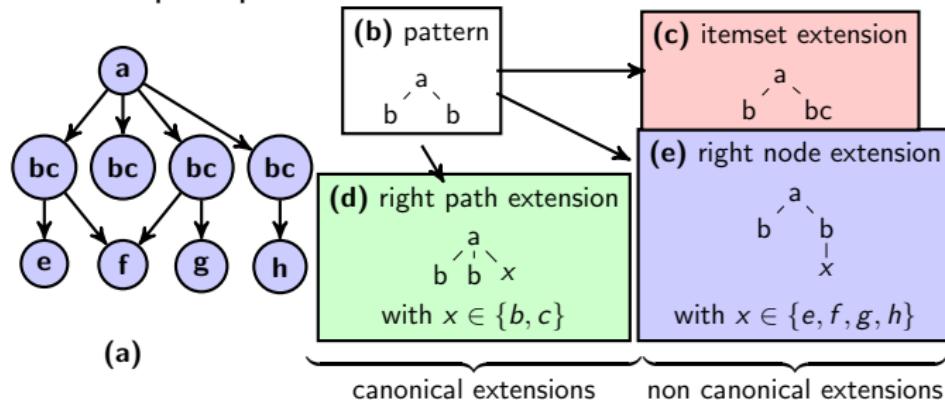
Pattern P_2

Pattern P_3

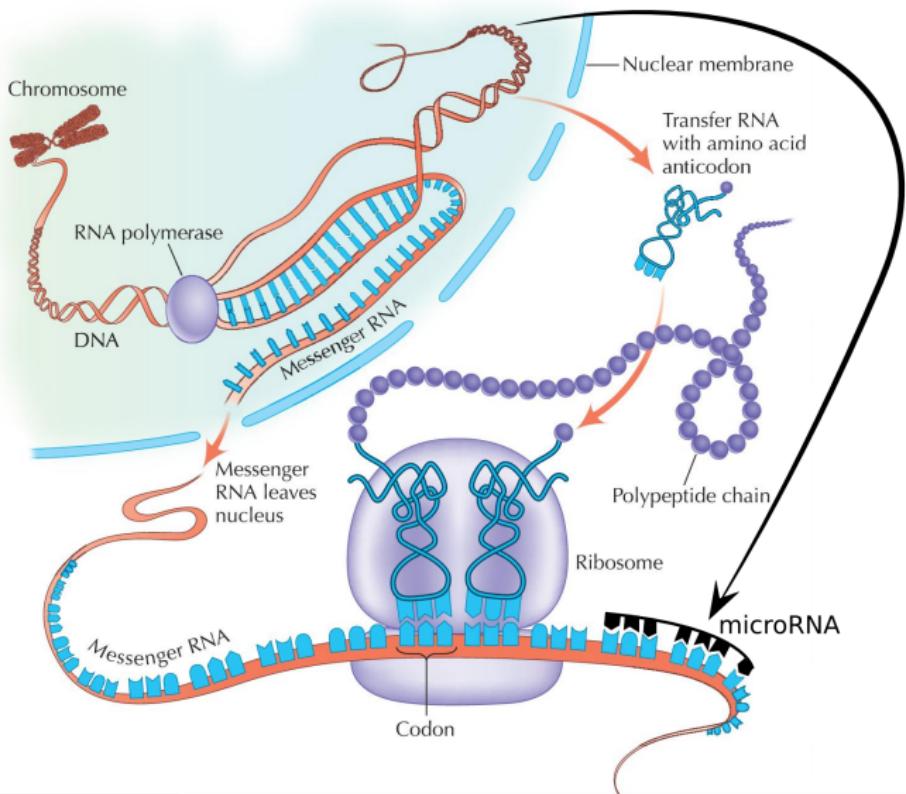
Extension of the method to the mining of attributed graphs

[Pasquier et al. 2017, Knowledge and Information Systems]

- proposal of a subgraph mining strategy based on the enumeration of spanning trees
- definition of a new canonical form handling reentrant links
- New enumeration method taking into account reentrant links and cycles
- pruning method to limit combinatorial explosion in the presence of automorphic patterns



Topic 2: prediction of microRNA diseases associations



- MicroRNAs play critical roles in many physiological processes.
- Their dysregulations are also closely related to the development and progression of various human diseases, including cancer
- Therefore, identifying new microRNAs associated with diseases contributes to a better understanding of pathogenicity mechanisms.

Methods based on disease similarities (2008)

limited performances

Methods based on target similarities (2010-2013)

pb: as the number of verified miRNA-mRNA interactions is currently low, predicted interactions (with high false-positive and false-negative rates) are also used

Methods using supervised learning (2013)

pb: the methods need negative miRNA-disease associations

Methods based on miRNA functional similarities (2012-2015)

pb: miRNA functional similarities remain largely unknown.
All the methods use predicted scores

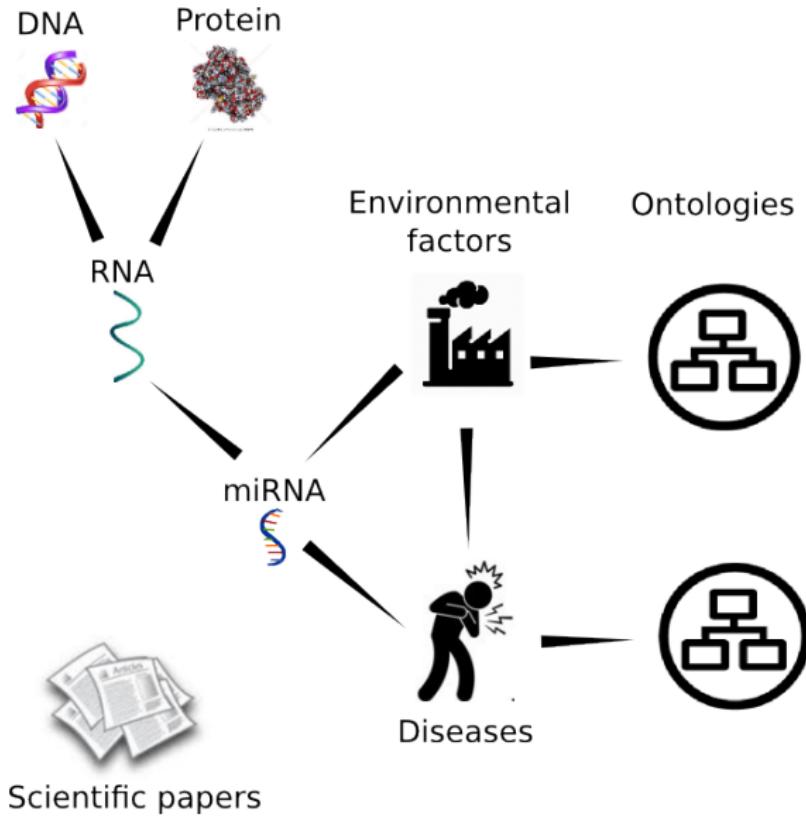
methods using several sources of data (2012-2015)

Best performances obtained with these methods

Observations

- ① best performance is reported with methods that use different sources of data
- ② some knowledge available in the domain is not used in existing programs
 - ▶ miRNA families suggest a common structure configuration in sets of genes that hint to shared functions,
 - ▶ significant correlation between pairs of expressed miRNAs occurs at a distance < 50 kb, suggesting common regulatory pathway,
 - ▶ important information about miRNAs is contained in scientific papers.
- ③ each miRNA is associated with a set of heterogeneous data that makes difficult the use of traditional machine learning algorithms.

Data availability

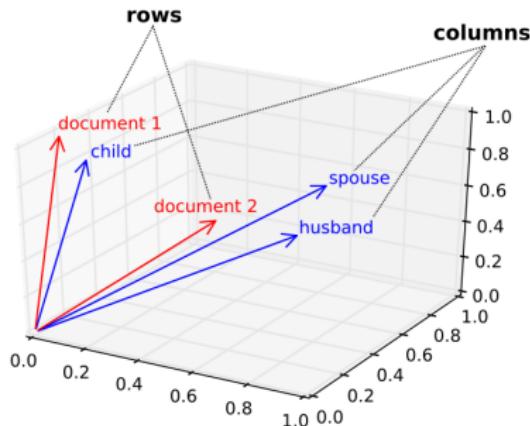


Idea: using distributional semantics

- Distributional semantics
 - ▶ based on an old hypothesis in linguistics (Zellig Harris, 1954) that states that the meaning of words can be deduced from their distributions (context in which the words are used)
- Computational systems proposed from late 80'
 - ▶ spatial models (vector space model)
 - ▶ e.g. Latent Semantic Analysis (LSA), US PATENT in 1988
 - ▶ Latent = identification of hidden relationships
- Used successfully in various applications
 - ▶ finding semantic similarity between words
 - ▶ word sense disambiguation
 - ▶ information retrieval (search engines)

Vector space modeling

- Collecting distributional information in high-dimensional vectors
 - ▶ each "item" is represented by thousands or millions of values
- Performing dimensionality reduction of the original space
 - ▶ use of Singular Value Decomposition
(similar to Principal Component Analysis)
 - ▶ this keeps most important semantic information
 - ▶ this reduces noise and other undesirable artifacts
- Defining semantic similarity in terms of vector similarity



Our hypothesis

If we consider that

- "items" represent miRNAs
- data associated with miRNAs are stored in vectors of values

Then

new knowledge about miRNAs might be highlighted
by using distributional semantics

Problem

Vector Space Model is designed to process textual data

For textual data, the 'classic' following steps can be performed

- ① collecting scientific papers associated with miRNA
- ② constructing a miRNA-term document matrix
- ③ weighting miRNA-term relationships (e.g. use of TF/IDF)
- ④ mapping data to a lower-dimensional space
- ⑤ exploring the reduced space to find miRNA-term associations

LSA example: input text

Let's take the following texts:

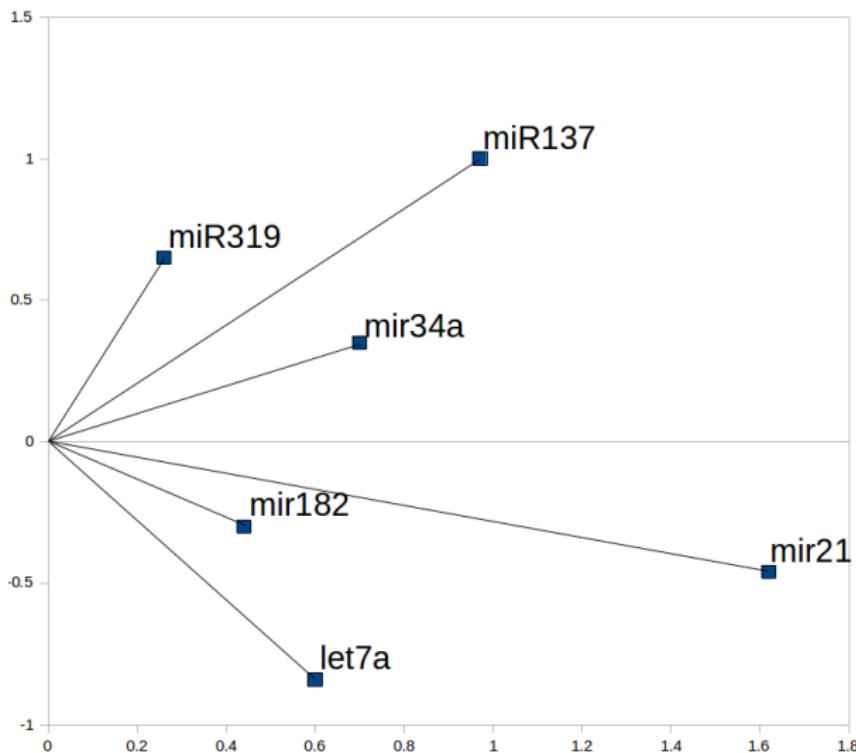
- MiR-21 is found to be highly expressed in numerous cancers like breast cancer, glioblastoma and lung cancer
- let-7a expression correlates with poor survival of lung cancer patients
- Suppression of MIM by mir-182 activates RhoA and promotes breast cancer metastasis
- miR-137 is frequently down-regulated in glioblastoma
- mir-34a targets notch1 and inhibits cell proliferation in glioblastoma multiforme.
- Repression of cell proliferation by miR319-regulated TCP4

LSA example: matrix representation

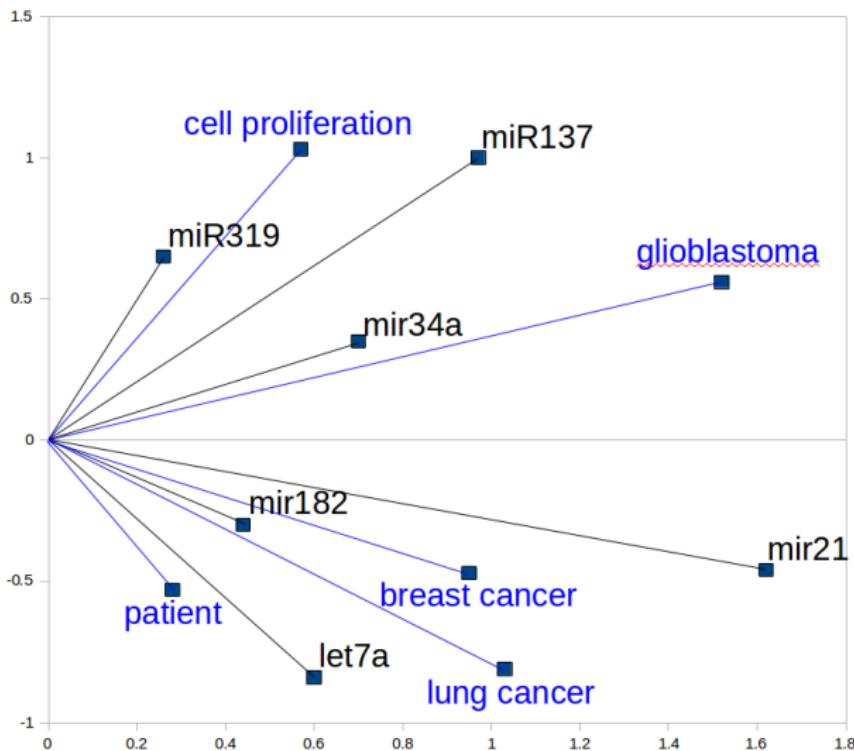
We construct a matrix of 6 rows (miRNAs) by 5 columns (terms)

$$M = \begin{pmatrix} & \textit{breast} & \textit{patient} & \textit{lung} & \textit{gliobl.} & \textit{cell} \\ & \textit{cancer} & & \textit{cancer} & & \textit{prolif.} \\ \textit{mir21} | & 1 & 0 & 1 & 1 & 0 \\ \textit{let7a} | & 0 & 1 & 1 & 0 & 0 \\ \textit{mir182} | & 1 & 0 & 0 & 0 & 0 \\ \textit{miR137} | & 0 & 0 & 0 & 1 & 1 \\ \textit{mir34a} | & 0 & 0 & 0 & 1 & 0 \\ \textit{miR319} | & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

LSA example: mapping miRNAs in reduced space



LSA example: mapping terms in the same reduced space



- MiR-21 is found to be highly expressed in numerous cancers like **breast cancer**, **glioblastoma** and **lung cancer**
- let-7a expression correlates with poor survival of **lung cancer** patients
- Suppression of MIM by **mir-182** activates RhoA and promotes **breast cancer** metastasis
- miR-137 is frequently down-regulated in **glioblastoma**
- mir-34a targets notch1 and inhibits **cell proliferation** in **glioblastoma multiforme**.
- Repression of **cell proliferation** by miR319-regulated TCP4

Non textual data represent biological facts

- Links between miRNAs and diseases, targets or families consist of binary information - There is no weight - This is True or False
 - ▶ Some author weight each association by counting the number of time the link appears in databases. This is not a good idea (selection bias)
- Neighboring between miRNAs are expressed as integers representing a distance

We need to define several weighting schemes

Weighting miRNA-disease associations

Weight of an association between a miRNA m and a disease d is obtained with:

$$x_{md} = \max(simil(d, k)) \text{ for } k \text{ in } \{w | x_{m,w} > 0\}$$

where $simil$ is a similarity measure taking into account the distance between two diseases in the MeSH ontology.

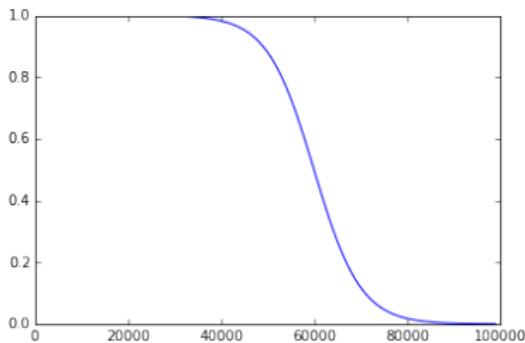
Weight between miRNA m and disease d is the maximum similarity between d and every disease associated with m

Weighting miRNA-neighbor associations

Genomic distance is transformed to a weight, expressing the correlation between two miRNA, with a sigmoid function of the form:

$$w(\text{dist}) = 1 - \frac{1}{1 + e^{-k(\text{dist} - \text{dist}_0)}}$$

where dist is the genomic distance between two miRNAs, dist_0 , the sigmoid midpoint, is set to 6×10^5 (a distance of 60kb is associated with a correlation of 0.5) and k , the steepness of the curve, is set to 2×10^{-4} .



Weighting miRNA-target associations

Associations can be represented by a bipartite network connecting items from a set of miRNAs to items from a set of targets

On this network, we use a recommender system that was primarily designed for online shops to predict user rating or preferences.

Network Based Inference (Zhou et al. 2007)

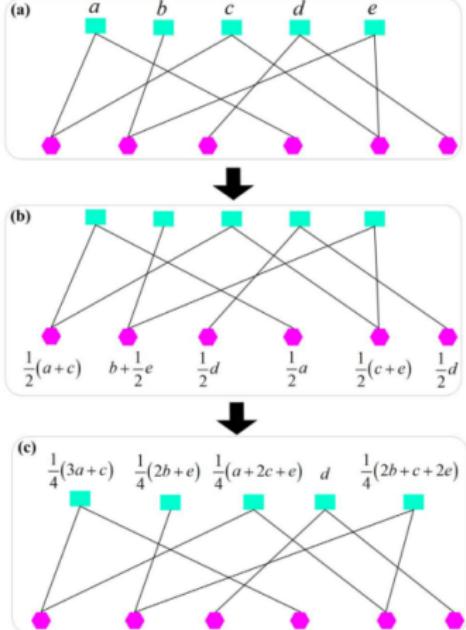


Figure 3. Principle of the resource-allocation process in a bipartite network. The green rectangles represent X nodes and red hexagons represent Y nodes. The whole process consists of two steps: First, the resource flows from X to Y ($a \rightarrow b$), and then returns to X ($b \rightarrow c$).

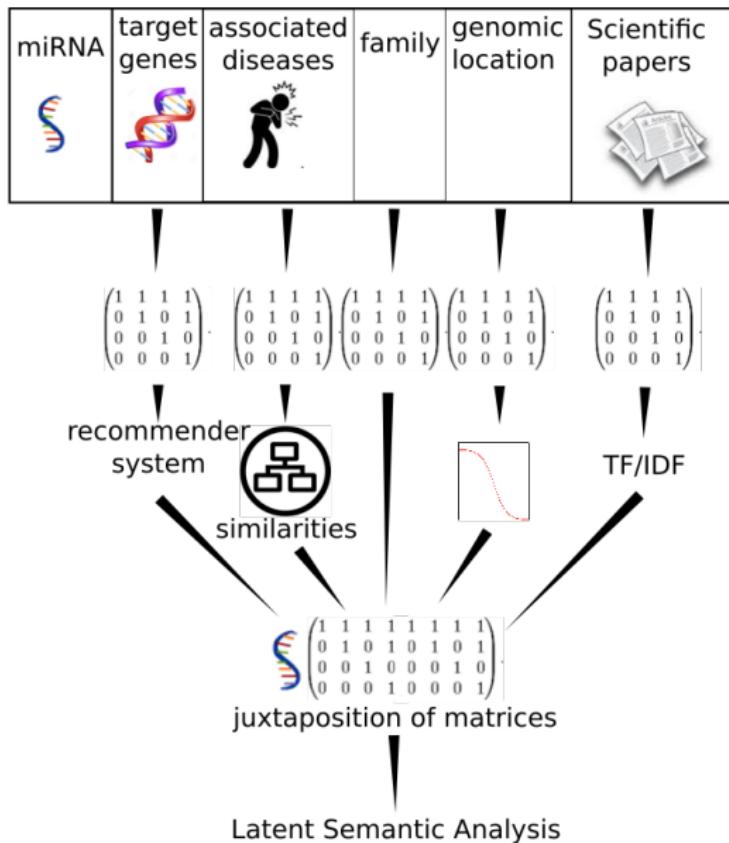
doi:10.1371/journal.pone.0087797.g003

$$M = \begin{pmatrix} a & 1 & 0 & 0 & 1 & 0 & 0 \\ b & 0 & 1 & 0 & 0 & 0 & 0 \\ c & 1 & 0 & 0 & 0 & 1 & 0 \\ d & 0 & 0 & 1 & 0 & 0 & 1 \\ e & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

- (c) is represented by a 5×5 matrix (W) representing the information transfer between blue nodes
- The new weighted matrix is obtained by $M \times W$

$$\begin{pmatrix} a & 1 & 0 & 0 & 3/4 & 1/4 & 0 \\ b & 0 & 3/4 & 0 & 0 & 1/4 & 0 \\ c & 3/4 & 1/4 & 0 & 1/4 & 3/4 & 0 \\ d & 0 & 0 & 1 & 0 & 0 & 1 \\ e & 1/4 & 1 & 0 & 0 & 3/4 & 0 \end{pmatrix}$$

The method (MIRAI)

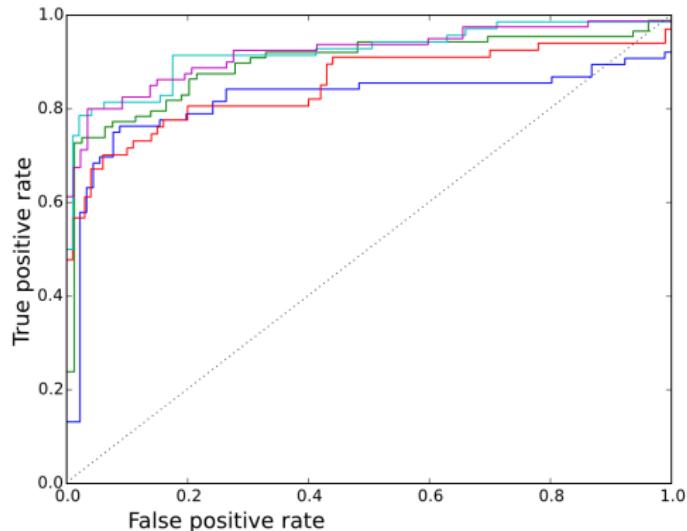


Evaluation process

- Based on the experimentally verified miRNA-disease associations stored in the human miRNA-disease database (HMDD) v2.0 of June, 14, 2014
- 5 fold cross validation
 - 5 partitions: 4/5 of data used for training, 1/5 used for testing
- every partition is used one time for testing
 - ▶ with associations to tested disease removed
 - ▶ with data collected from litterature also removed
- the predictions are performed for 83 diseases (associated with at least 20 miRNAs)
- evaluation consists in measuring whether the program is able to recover the deleted associations

Measuring the performance of the classifier

- plotting Receiver Operating Characteristic (ROC) curve
- an unique measure is obtained with AUC (Area Under the Curve)

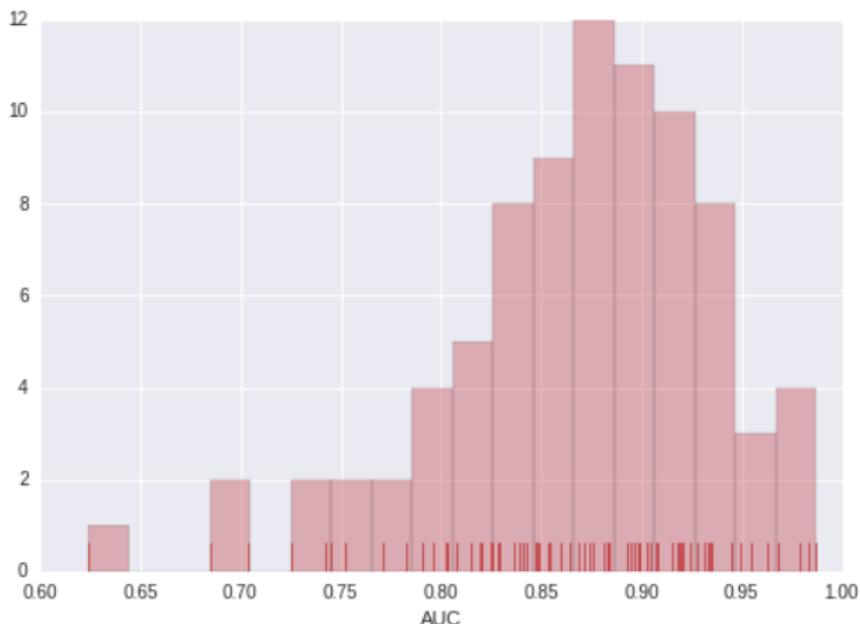


Interpretation of AUC:

- 0.90-1 = excellent (A)
- 0.80-0.90 = good (B)
- 0.70-0.80 = fair (C)
- 0.60-0.70 = poor (D)
- 0.50-0.60 = fail (F)

Curves obtained for the 5 partitions
of Breast cancer data

Univariate distributions of AUC scores



- Average AUC value: 0.867 (min: 0.624, max: 0.987)
- Most of the predictions are above 0.8

Comparison with other methods

Disease name	MIDP 2015	RWR 2012	HDMP 2013	RLS 2014	Chen 2013	MiRAI 2015
Acute myeloid leukemia	0.91	0.84	0.86	0.85	0.72	0.90
Breast neoplasms	0.84	0.79	0.80	0.83	0.65	0.86
Colorectal neoplasms	0.85	0.79	0.80	0.83	0.66	0.86
Glioblastoma	0.80	0.68	0.70	0.71	0.61	0.90
Heart failure	0.82	0.72	0.77	0.74	0.76	0.80
Hepatocellular carcinoma	0.81	0.75	0.76	0.79	0.61	0.81
Lung neoplasms	0.88	0.83	0.84	0.86	0.61	0.90
Melanoma	0.84	0.78	0.79	0.81	0.64	0.85
Ovarian neoplasms	0.92	0.88	0.88	0.91	0.64	0.87
Pancreatic neoplasms	0.95	0.87	0.90	0.89	0.68	0.93
Prostatic neoplasms	0.88	0.82	0.85	0.84	0.63	0.87
Renal cell carcinoma	0.86	0.82	0.83	0.84	0.63	0.87
Squamous cell carcinoma	0.87	0.82	0.82	0.85	0.68	0.88
Stomach neoplasms	0.82	0.78	0.79	0.80	0.63	0.82
Urinary bladder neoplasms	0.90	0.82	0.85	0.85	0.63	0.88
AVERAGE	0.862	0.799	0.816	0.827	0.652	0.867

Prediction results for less studied diseases

Disease name	nb assoc	MiRAI
Arthritis, Rheumatoid	21	0.912
Atrial Fibrillation	21	0.903
Azoospermia	21	0.892
Barrett Esophagus	21	0.966
Coronary Artery Disease	21	0.797
Crohn Disease	21	0.933
Diabetes Mellitus, Type 2	20	0.825
Infertility, Male	20	0.863
Liver Neoplasms	20	0.814
Sarcoma, Kaposi	20	0.920

- Only the following 2 diseases obtain a score lower than 0.70.
 - ▶ **Lupus Vulgaris**, associated with 62 miRNAs obtained from an unique paper analyzing the microRNA expression patterns in renal biopsies of lupus nephritis patients.
 - ▶ **Adenoviridae Infections** is associated with 74 miRNAs extracted from two tables of an unique paper presenting the up and down-regulated miRNAs in adenovirus type 3 (AD3) infected human laryngeal epithelial.
 - ★ Only 14 associations are really confirmed by Q-PCR.
 - ★ By keeping only verified associations, the AUC increases to 0.796

⇒ The database used to develop the method
contains false associations

What next?

Our method achieves state of the art performance on the classification of known association.

⇒ This is good ! But not enough !

The real contribution

Is the method able to predict new (unknown) associations?

Is the method able to detect false associations?

"new" miRNAs associated with breast cancer

miRNA name	described in (PMID)	publication date
hsa-mir-106a	19706389	Sept. 2009
hsa-mir-130a	23528537	Jun. 2013
hsa-mir-138-1	23300839	Dec. 2012
hsa-mir-138-2	23300839	Dec. 2012
hsa-mir-142	25406066	Nov. 2014
hsa-mir-144	25465851	Dec. 2014
hsa-mir-150	24312495	Dec. 2013
hsa-mir-15b	22908280	Sept. 2012
hsa-mir-181c	23524334	Jul. 2013
hsa-mir-19b-2	21059650	Jan. 2011
hsa-mir-208a		
hsa-mir-30e	19432961	May 2009
hsa-mir-378a	20889127	Oct. 2010
hsa-mir-99a	21575166	May 2011

All are confirmed by recent litterature excepted mir-208a that was proposed as a novel miRNA associated to breast cancer

Sun et al.

MiR-208a stimulates the cocktail of SOX2 and β -catenin to inhibit the let-7 induction of self-renewal repression of breast cancer stem cells and formed miR208a/let-7 feedback loop via LIN28 and DICER1.

Oncotarget. 2015 Oct 20;6(32):32944-54

Dubious associations with breast cancer

hsa-mir-1323	hsa-mir-1469	hsa-mir-215	hsa-mir-2355
hsa-mir-3130-1	hsa-mir-3130-2	hsa-mir-3186	hsa-mir-411
hsa-mir-4257	hsa-mir-4306	hsa-mir-718	

- 9 miRNAs come from a table of an unique publication that lists the overexpressed circulating miRNAs. None of them were confirmed
- the 2 remaining miRNAs (mir-215 and mir-411) are specified in an unique paper focusing on circulating miRNAs used as biomarkers

9 out of the 11 association are not verified

According to recent papers, the conclusions based on circulating miRNAs should be taken with caution

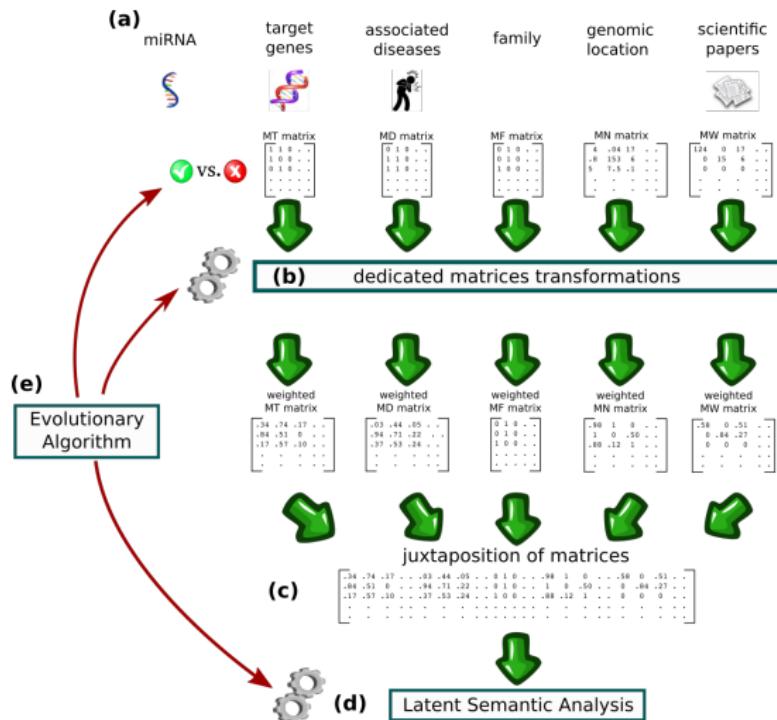
Results

list of 811 potential miRNAs associated to 93 diseases

list of 144 potential miRNAs falsely associated with 42 diseases

Optimisation of MiRAI with an evolutionary algorithm

[Pallez et al. 2017, Scientific Reports]



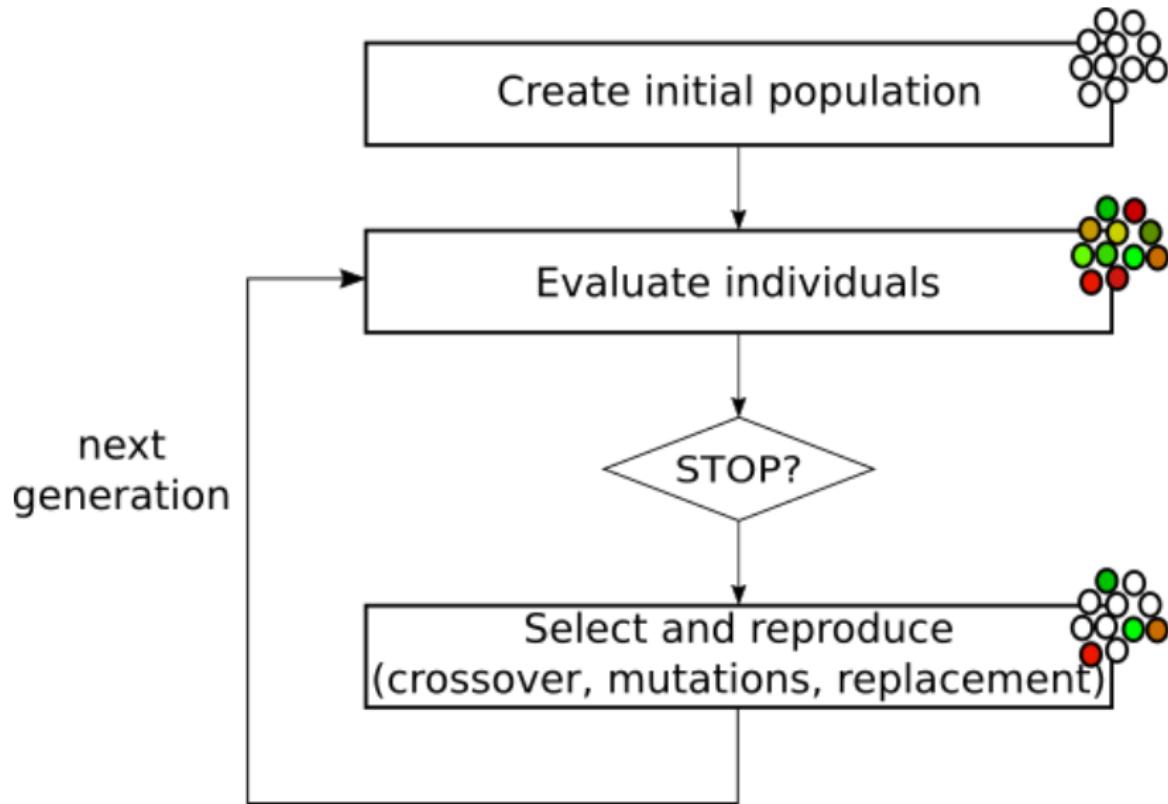
35 binary parameters
2³⁵ possibilities

expensive evaluation of each configuration

⇒ Binary Differential Evolution

⇒ use of a surrogate model that approximates the expensive real fitness function

Evolutionnary algorithm



Improvements brought by automatic parameter setting

Disease name	RWR 2012	HDMP 2013	RLS 2014	MIDP 2015	MiRAI 2015	MiRAI+EA 2017
Acute myeloid leukemia	0.84	0.86	0.85	0.91	0.90	0.91
Breast neoplasms	0.79	0.80	0.83	0.84	0.86	0.86
Colorectal neoplasms	0.79	0.80	0.83	0.85	0.86	0.87
Glioblastoma	0.68	0.70	0.71	0.80	0.90	0.87
Heart failure	0.72	0.77	0.74	0.82	0.80	0.85
Hepatocellular carcinoma	0.75	0.76	0.79	0.81	0.81	0.82
Lung neoplasms	0.83	0.84	0.86	0.88	0.90	0.93
Melanoma	0.78	0.79	0.81	0.84	0.85	0.87
Ovarian neoplasms	0.88	0.88	0.91	0.92	0.87	0.91
Pancreatic neoplasms	0.87	0.90	0.89	0.95	0.93	0.92
Prostatic neoplasms	0.82	0.85	0.84	0.88	0.87	0.87
Renal cell carcinoma	0.82	0.83	0.84	0.86	0.87	0.88
Squamous cell carcinoma	0.82	0.82	0.85	0.87	0.88	0.89
Stomach neoplasms	0.78	0.79	0.80	0.82	0.82	0.85
Urinary bladder neoplasms	0.82	0.85	0.85	0.90	0.88	0.90
AVERAGE	0.799	0.816	0.827	0.862	0.867	0.880

The expectations of transcriptomic analyses have evolved

- Biologists carry out more precise experiments
- They expect much more targeted results

Identification of active modules

[PhD of Leandro Correa, started in january 2018]

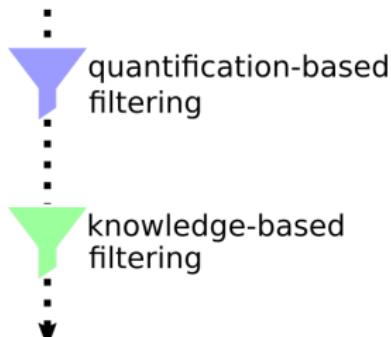
Quantification

- transcriptomique
- protéomique

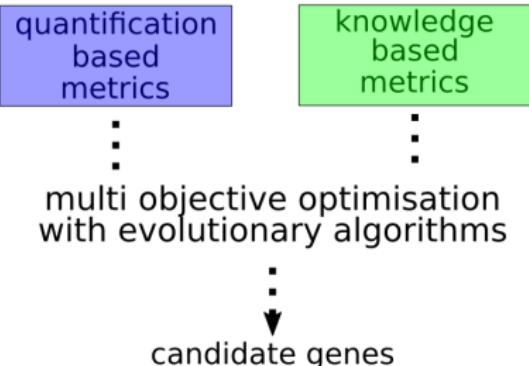
Biological knowledge

- interaction networks
- molecular pathways
- ontologies

Traditional pipeline



Our proposal



Current collaborations and fundings

- Transcriptomic analysis based on multi-criteria optimization
 - ▶ collaboration with iBV
 - ▶ funding from the Fondation ARC
 - ▶ funding from INCa
 - ▶ PhD funded by Université Côte d'Azur
- Prediction of microRNA-diseases associations
 - ▶ collaboration with Biomanda
- Study of ARN ⇒ ARN and ARN ⇒ ADN interactions
 - ▶ collaboration with institut Sophia Agrobiotech

