

# Поиск изоморфных подграфов

Слесарев Илья, 371 группа

## Применение графов в биологии:

- молекулярные сети
- белковые сети

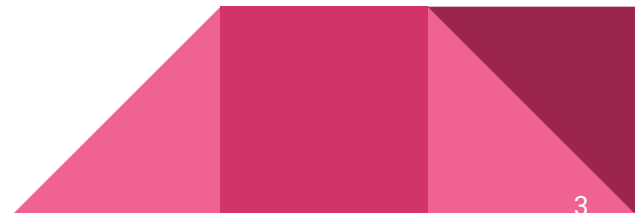
Цель: избавиться от неудачных “матчей” максимально недорого

# Основные понятия

Изоморфизм

Мономорфизм

Густой(плотный) и редкий графы

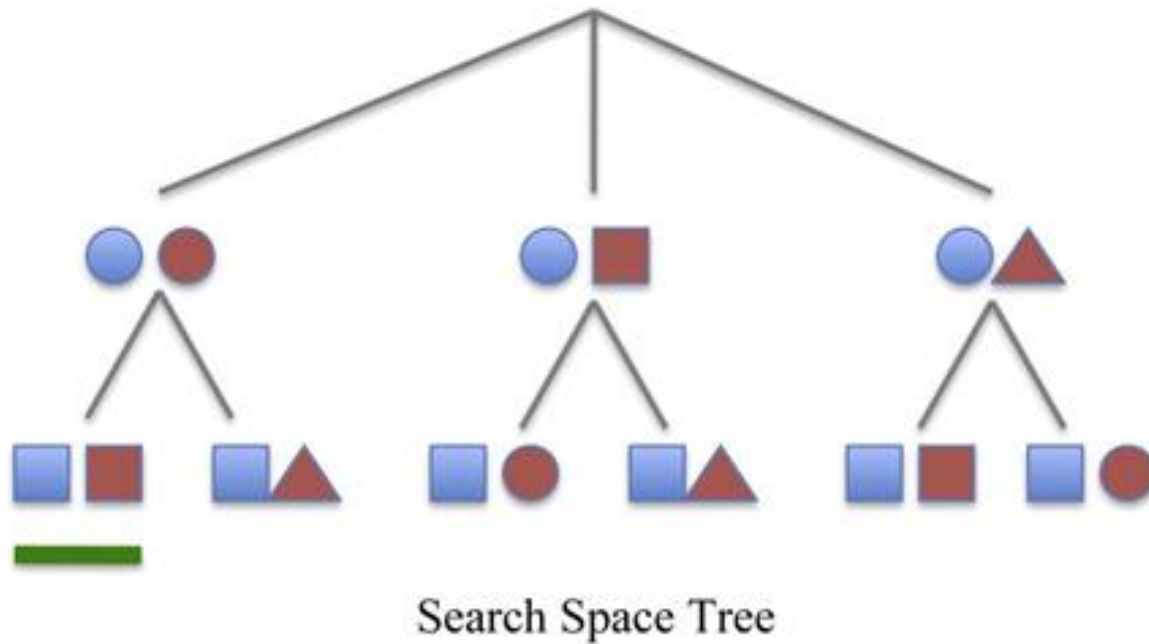
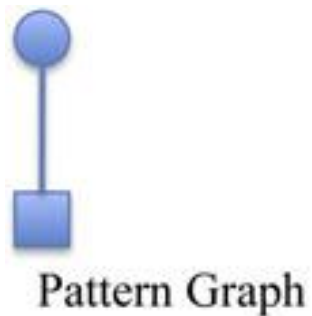


# Принцип работы алгоритмов

1. Построение дерева области поиска (Search space tree)
2. Уменьшение этого дерева, путем отсекаания целых веток

	<b>Search Strategy</b>	<b>Reduce Search Space</b>	<b>Preprocessing Data</b>	<b>x Data Structure</b>
FocusSearch [29]	Static Semi-target dependent	Local domain reduction	Yes	List
Lad [30]	Dynamic Target dependent	Domain reduction until convergence	Yes	Matrix
VFlib [3]	Dynamic Target dependent	Two-Look-Head pruning rules	No	List
RI	Static Target independent	Fast and light pruning rules	No	List

# Дерево области поиска



# VFLib

Учитывая частичное решение, берет несматченные вершины шаблона, имеющие общие ребра с уже сматченными, аналогично берет несматченные вершины целевого графа, имеющие общие ребра с уже сматченными

# VFLib. “Заглядывания”

Две вершины ( $u$  и  $u'$ ) можно считать сматченными, если:

1. Они обе соседи уже сматченных
2. Число несматченных, смежных с  $u$ , меньше или равно числу несматченных, смежных с  $u'$
3. Число оставшихся несматченных, смежных с  $u$ , меньше или равно числу несматченных, смежных с  $u'$



# LAD

Идея:

Две вершины шаблона не могут быть сматченны с одной и той же вершиной в целевом графе в частичном решении с условием сохранения соответствия ребра между ними. Это ограничение применяется до сходимости (отсюда определение Domain reduction until convergence)

# FocusSearch

Идея в подготовке:

1. The first one, called prematch, fills domains by filtering them using vertex invariants based on labels and topology.
2. Аналогично LAD. Две вершины шаблона не могут быть сматченны с одной и той же вершиной в целевом графе в частичном решении с условием сохранения соответствия ребра между ними.

Далее вершины располагают в таком порядке, что каждая вершина из шаблона во главе последовательности, а далее располагаем вершины в порядке уменьшения общих веток с частичным решением.

Существуют также и другие подходы, но они основаны на вероятностных функциях

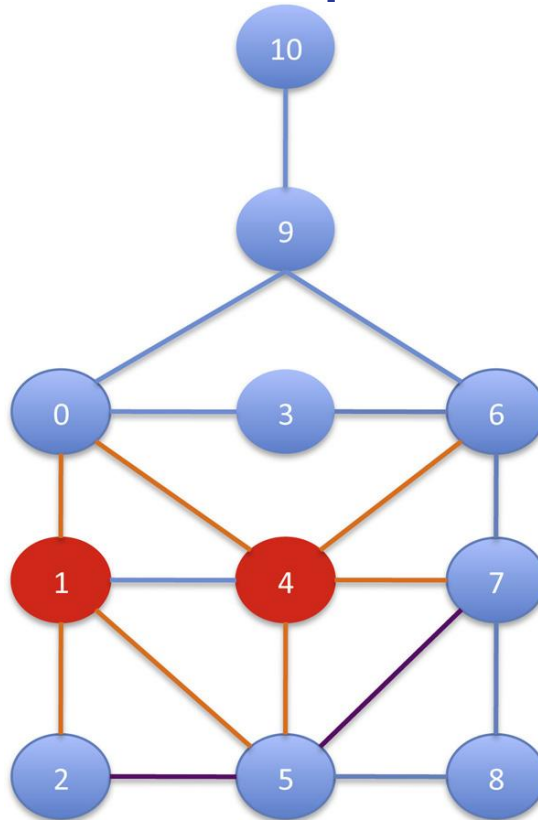
# RI

	<b>Search Strategy</b>	<b>Reduce Search Space</b>	<b>Preprocessing Data</b>	<b>x Data Structure</b>
FocusSearch [29]	Static Semi-target dependent	Local domain reduction	Yes	List
Lad [30]	Dynamic Target dependent	Domain reduction until convergence	Yes	Matrix
VFlib [3]	Dynamic Target dependent	Two-Look-Head pruning rules	No	List
RI	Static Target independent	Fast and light pruning rules	No	List

# RI алгоритм

1. Упорядочивание вершин - GreatestConstraintFirst
2. Непосредственно match - ReduceSearch

# RI - упорядочивание вершин



# RI - reduce search

$$M(u_i) = u_i'$$

1.  $u_i$  и  $M(u_i)$  еще не сматченны
2.  $u_i$  и  $M(u_i)$  соответствуют друг другу ( $\text{lab}(u_i) == \text{lab}(M(u_i))$ )
3. Число ребер в целевом графе, смежных с  $M(u_i) \geq$  аналогичному числу ребер в шаблоне
4. Все ребра из данного частичного пути, смежные с  $u_i$ , также соответствуют друг другу.

# Версии алгоритма RI

RI-Ds: добавляет проверку на ребрах в областях для вершин

Ri-DsPm: добавляет идею из FocusSearch(Called prematch, fills domains by filtering them using vertex invariants based on labels and topology.) Т.е. фильтрует области при помощи инвариантов вершины, основанных на соседних метках (label) и топологии.



Table 1

Statistics of biochemical datasets.

	Min Vertices	Min Edges	Max Vertices	Max Edges	Avg (SD) Vertices	Avg (SD) Edges	Avg (SD) Degree	Total Labels	Avg (SD) Labels
<i>AIDS</i> Small Sparse	4	8	245	500	44.98 (21.68)	93.91 (45.05)	4.17 (2.28)	62	4.36 (0.86)
<i>PDBSv1</i> Large Sparse	240	480	33067	61546	5663.6 (6954.82)	86661.27 (12365.7)	3.21 (2.52)	14	5.9 (1.04)
<i>PDBSv2</i> Medium Sparse	1683	3414	7979	16302	3614.1 (1772.06)	7386.2 (3814.08)	4.08 (17.47)	13	4.63 (0.76)
<i>PDBSv3</i> Small Dense	7	16	883	18832	376.86 186.66	8679.48 3814.08	44.78 (17.47)	21	18.86 (3.48)
<i>Graemlin</i> Medium Dense	1081	12961	6726	230468	3167.6 (1568.66)	87759.6 (75939.2)	48.14 (63.61)	31676	3167.6 (1568.66)
<i>PPI</i> Large Dense	5720	51464	12575	332458	7827.1 (2120.15)	107135 (82730.9)	28.66 (47.44)	78271	7827.1 (2120.15)