

Бикластеризация в масс-спектрометрии

- Научно-исследовательская работа
- Научный руководитель:
- Сергей Николенко

- Магистрант:
- Сергей Князев

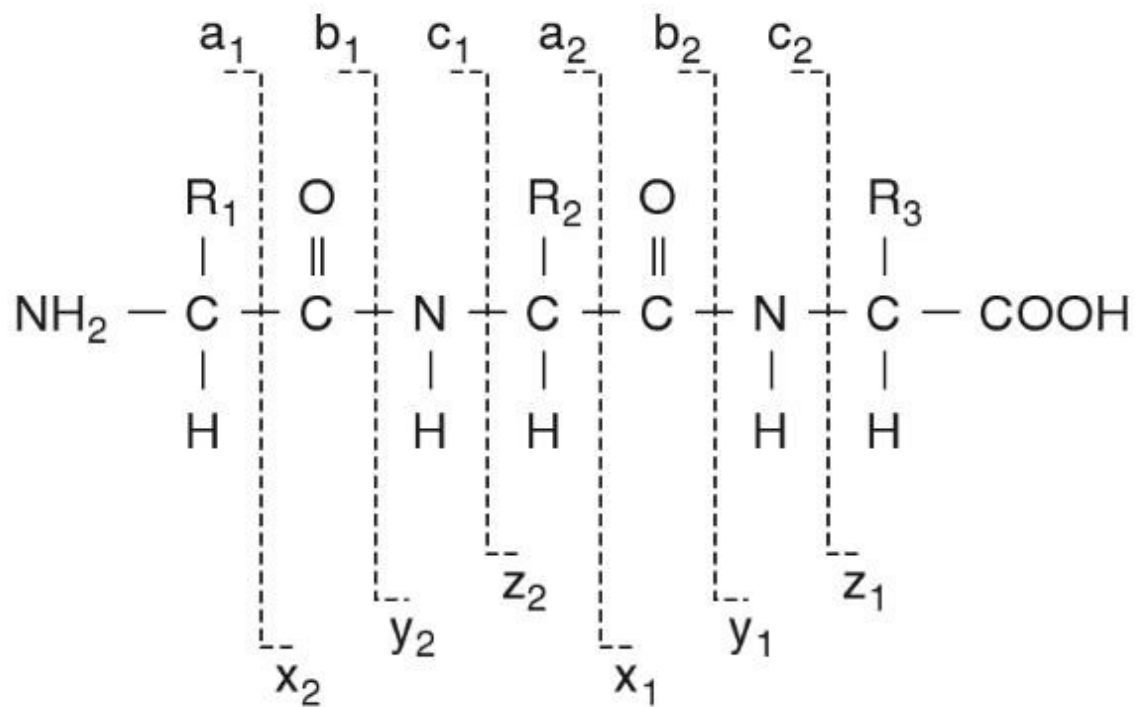
Что такое масс-спектрометрия

- Масс-спектрометрия - метод исследования вещества путём определения отношения массы к заряду.
- Как масспектрометрия работает:
 - Ионизация смеси пептидов.
 - Разделение пептидов на части.
 - Измерение массы/заряда частей.

Масс-спектрометрия

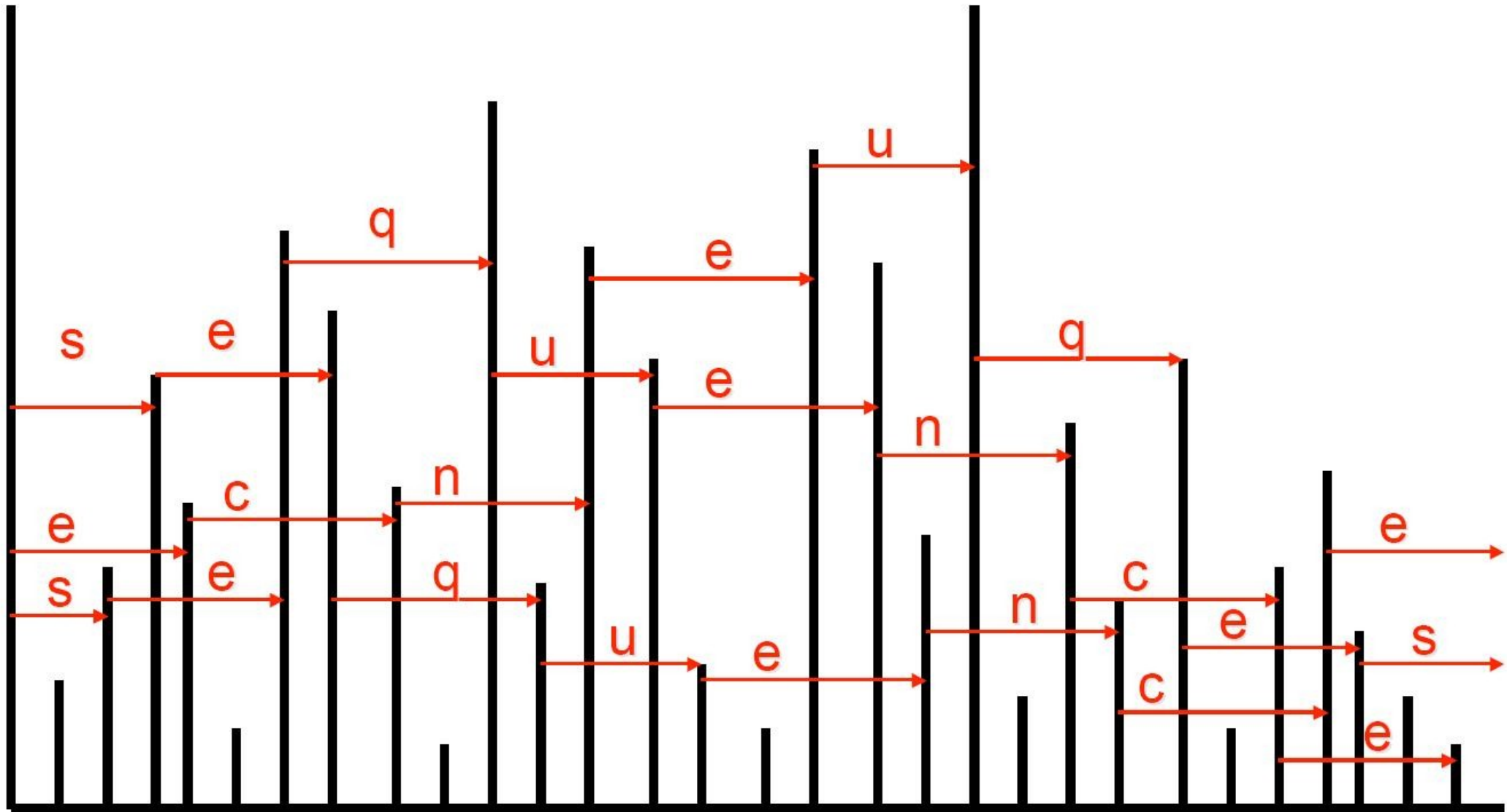
- Места разделения пептидов:

N-Terminal ions

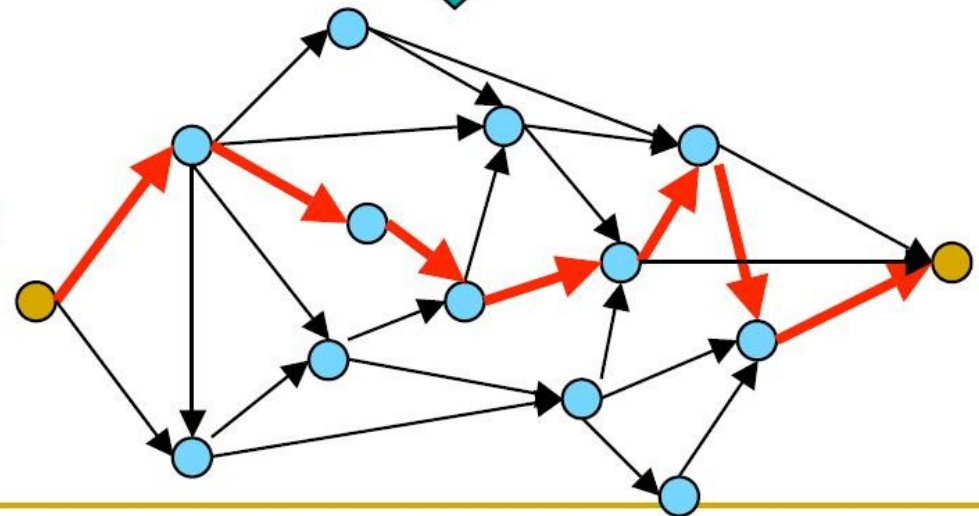
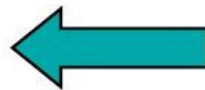
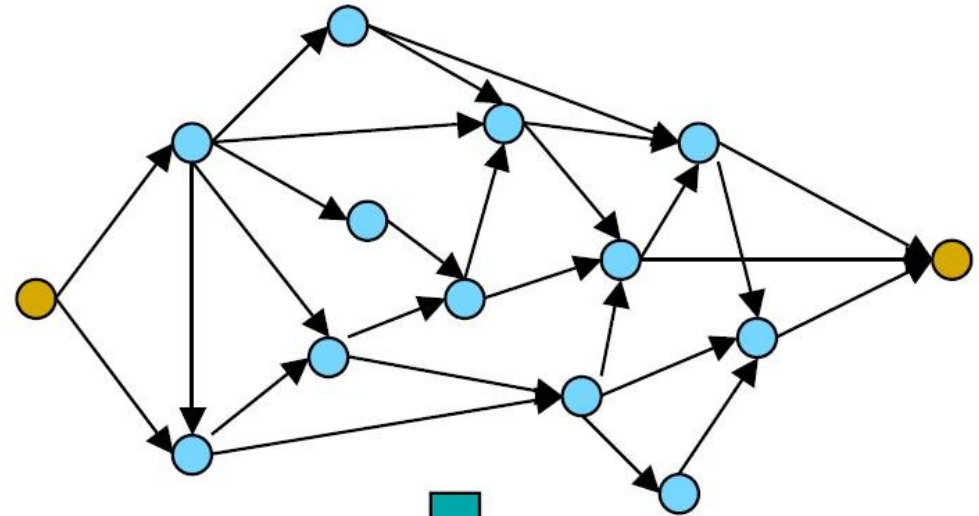
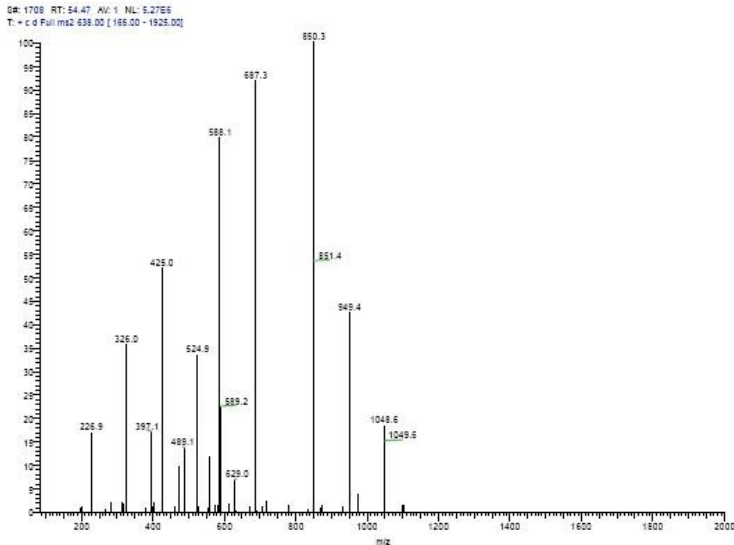


C-Terminal ions

Масс-спектр



Определение пептида по масс-спектру



Sequence

Постановка задачи

- Дана матрица

- Строки X – масса частицы

- Столбцы Y – номер исхода эксперимента

- $X = \{x_1, \dots, x_n\}$ $Y = \{y_1, \dots, y_n\}$

- Ячейки – интенсивность спектра a_{ij}

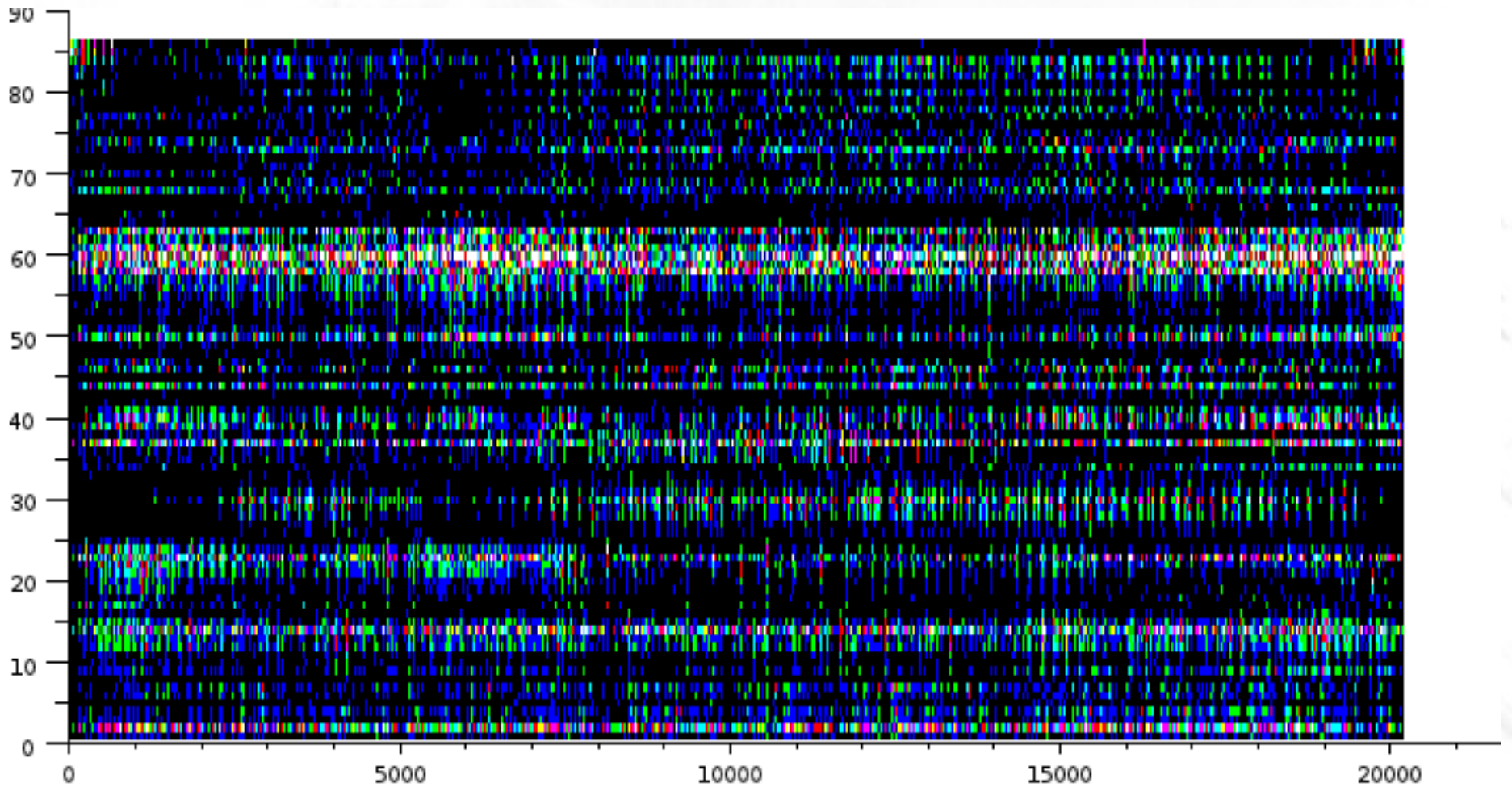
- Задача

Найти субматрицу, в которой ячейки будут похожими - бикластер

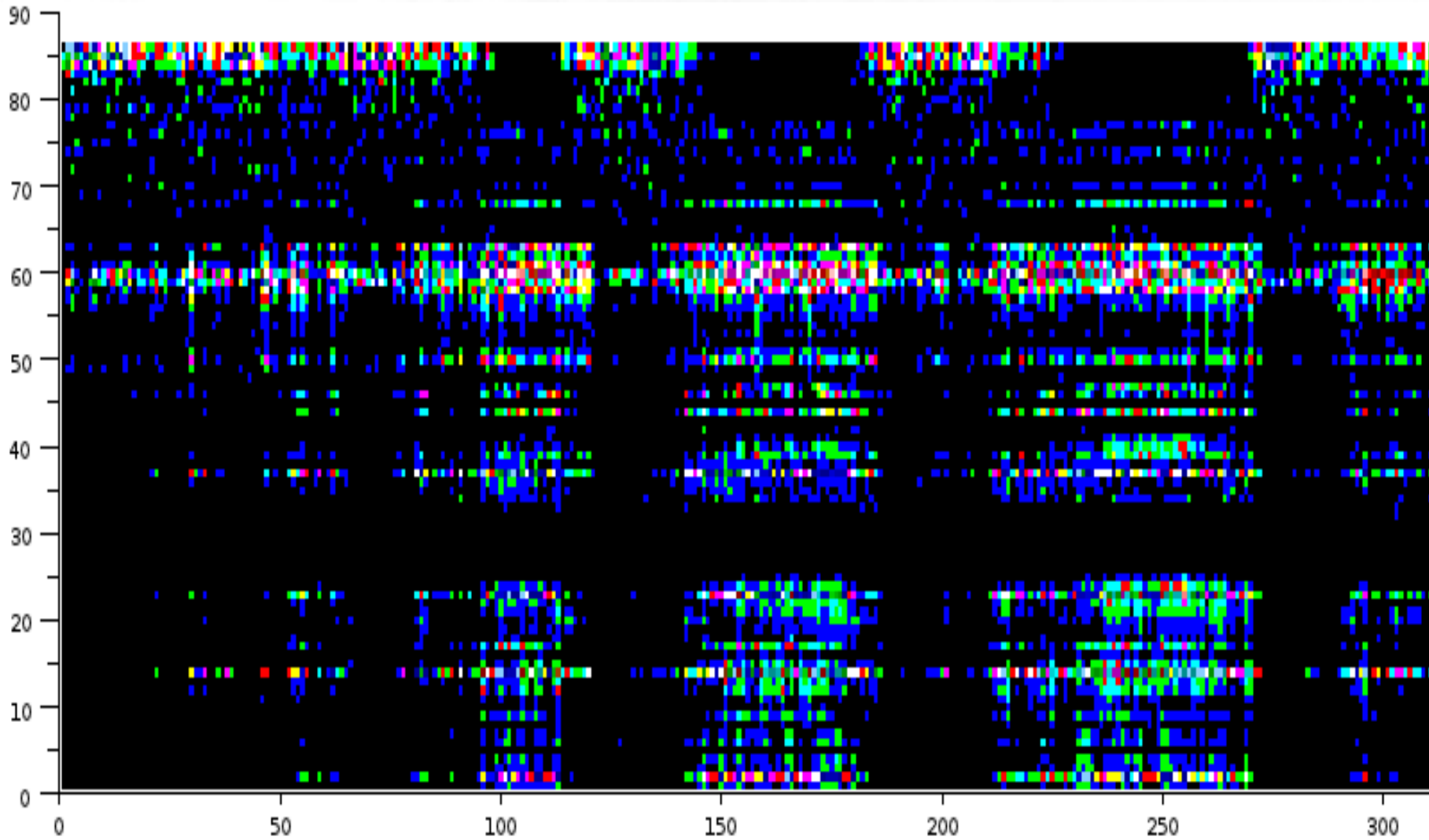
- $I \subseteq X, J \subseteq Y, A_{IJ} = (I, J) =$ субматрица

- $I = \{i_1, \dots, i_k\}$ $J = \{j_1, \dots, j_s\}$

Исходные данные



Исходные данные



Виды бикластеров

- Константные значения во всем бикластере
- Константные значения в строках или столбцах
- Когерентные значения

1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0

1.0	1.0	1.0	1.0
2.0	2.0	2.0	2.0
3.0	3.0	3.0	3.0
4.0	4.0	4.0	4.0

1.0	2.0	3.0	4.0
1.0	2.0	3.0	4.0
1.0	2.0	3.0	4.0
1.0	2.0	3.0	4.0

1.0	2.0	5.0	0.0
2.0	3.0	6.0	1.0
4.0	5.0	8.0	3.0
5.0	6.0	9.0	4.0

1.0	2.0	0.5	1.5
2.0	4.0	1.0	3.0
4.0	8.0	2.0	6.0
3.0	6.0	1.5	4.5

S1	S1	S1	S1
S1	S1	S1	S1
S1	S1	S1	S1
S1	S1	S1	S1

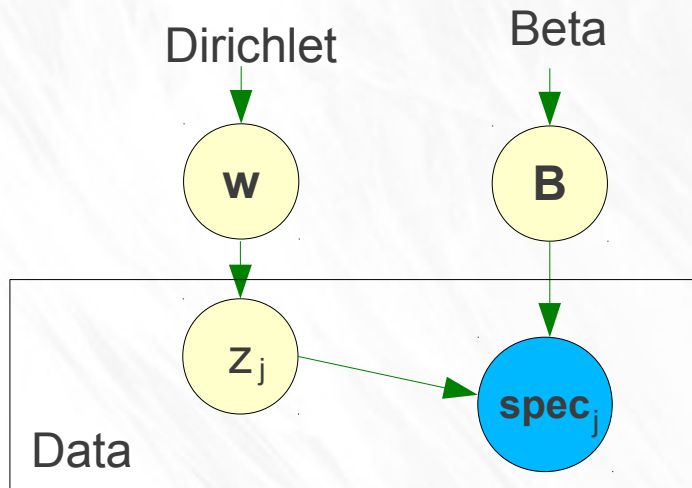
S1	S1	S1	S1
S2	S2	S2	S2
S3	S3	S3	S3
S4	S4	S4	S4

S1	S2	S3	S4
S1	S2	S3	S4
S1	S2	S3	S4
S1	S2	S3	S4

70	13	19	10
49	40	49	35
40	20	27	15
90	15	20	12

↖	↗	↘	↙
↘	↙	↗	↖
↗	↖	↘	↙
↙	↘	↖	↗

Clustering



$$p(w|a_1, a_2, \dots, a_n) = \text{Dirichlet}(a_1, a_2, \dots, a_n)$$

$$z = \{0, 1, 2, \dots, n\}$$

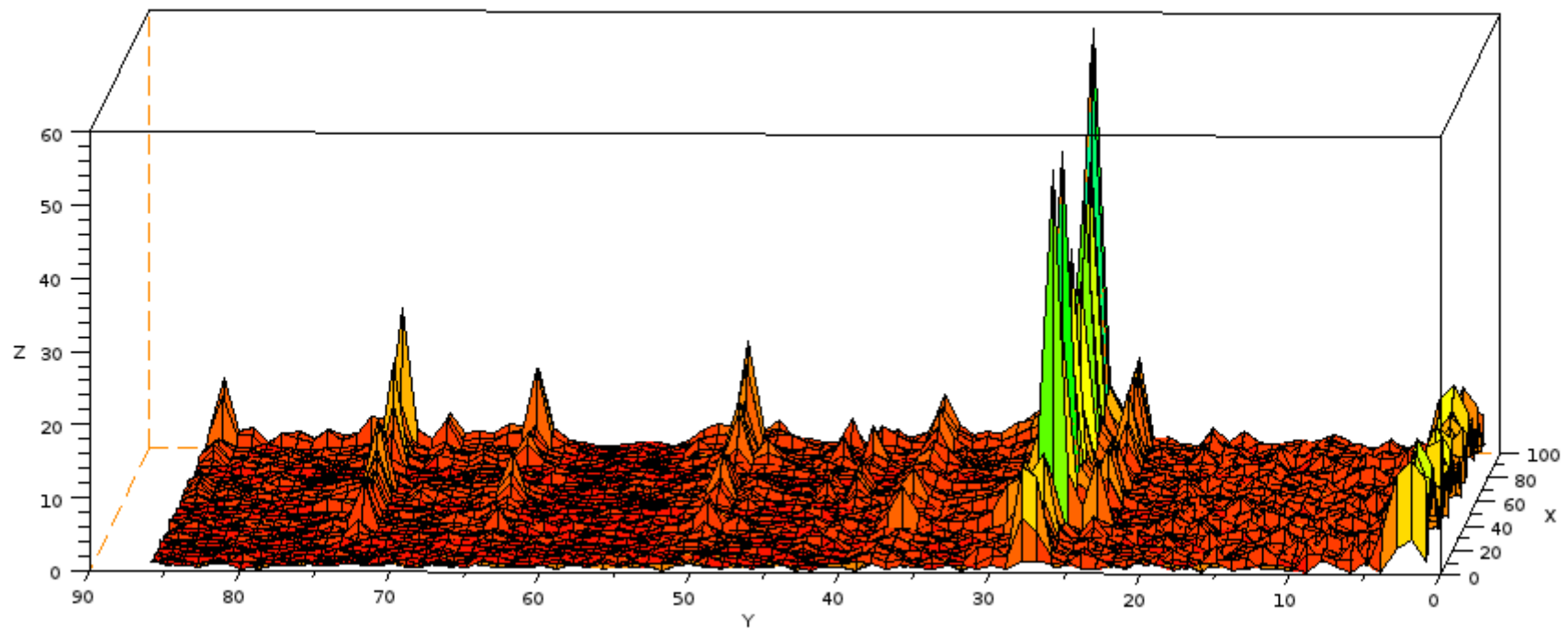
$$p(z=i|w) = w[i]$$

$$p(\text{spec}_j[\text{peak}] = 1 | z_j = i) = B[i, \text{peak}]$$

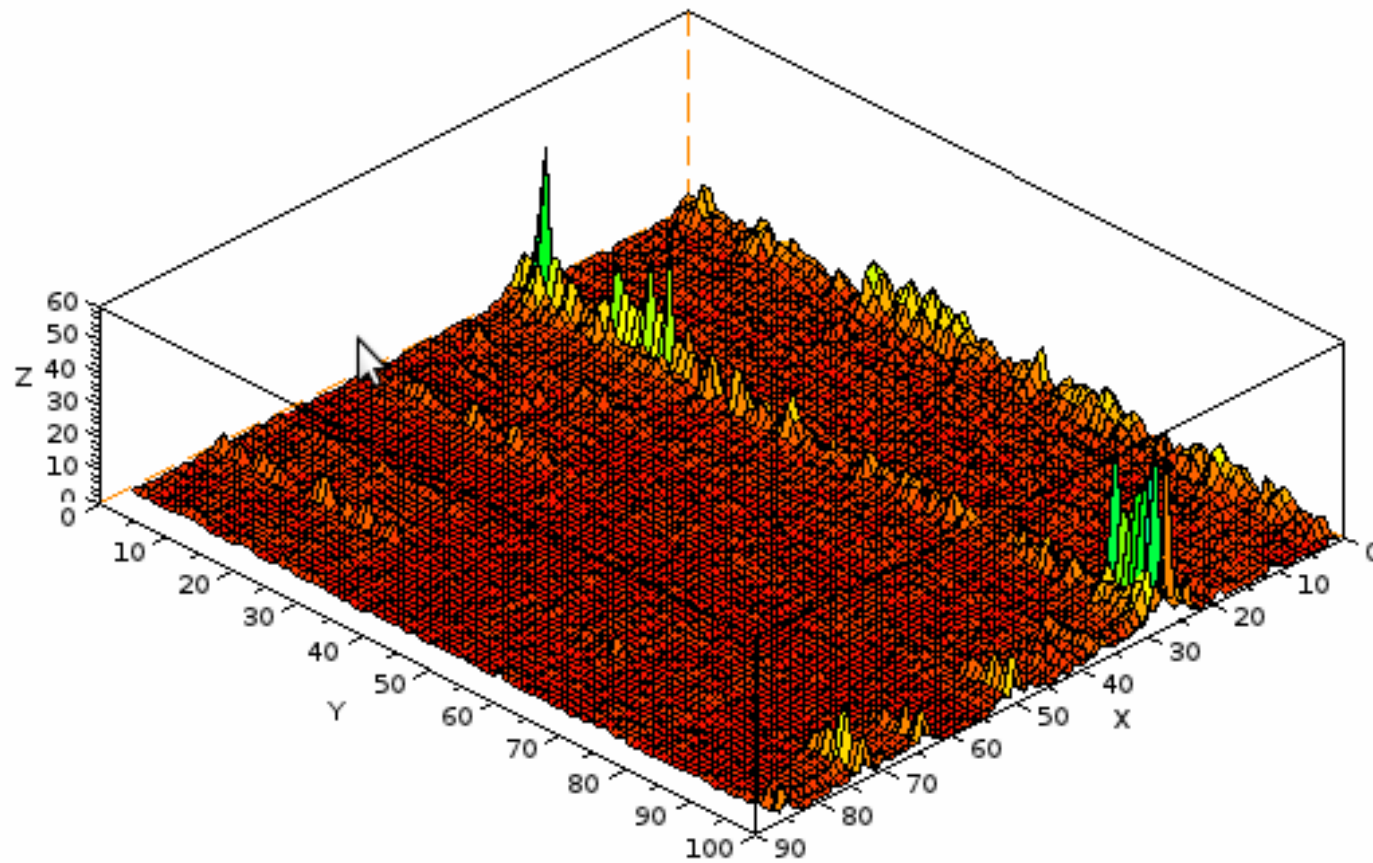
$$p(B[i, \text{peak}]) = \text{Beta}(\alpha[i, \text{peak}], \beta[i, \text{peak}])$$

Исходные данные

- Масс-спектр



Clustering



Clustering

- Output Data:
- Dirichlet(796.4 653.1 1255 961.6 1195 1460 1611 668.2 1462 1185 1473 751.5 725.9 498 1078 988.5 1068 572.8 575.5 1226)
-
- [0] Discrete(3.746e-44 1.061e-23 3.244e-27 2.786e-15 2.994e-43 4.585e-11 3.698e-30 1.433e-48 2.849e-11 1.12e-35 1 8.582e-30 7.565e-12 1.519e-38 5.514e-54 3.062e-56 1.877e-33 2.42e-25 1.213e-36 8.911e-31)
- [1] Discrete(2.033e-37 1.463e-16 3.829e-20 5.942e-12 3.136e-35 4.04e-08 1.427e-27 2.265e-41 1.579e-11 3.158e-29 0.9995 7.256e-25 0.0005335 9.56e-31 3.709e-48 1.053e-49 5.432e-26 1.257e-20 1.316e-28 1.49e-26)
- [2] Discrete(6.683e-36 7.982e-11 2.299e-17 8.78e-11 2.445e-29 8.829e-10 1.516e-23 9.281e-44 1.345e-09 3.66e-24 0.1761 9.159e-23 0.8239 1.423e-26 3.634e-51 1.843e-50 2.639e-25 2.218e-18 1.69e-24 5.045e-30)
-

Clustering

-
- [0,0] Beta(1.691,795.7)[mean=0.002121]
- [0,1] Beta(12.03,642.1)[mean=0.01839]
- [0,2] Beta(23.05,1233)[mean=0.01836]
- [0,3] Beta(6.068,956.6)[mean=0.006303]
- [0,4] Beta(22,1174)[mean=0.01839]
- [0,5] Beta(28.16,1433)[mean=0.01928]
-
- [23,2] Beta(1193,62.48)[mean=0.9502]
- [23,3] Beta(507.8,454.8)[mean=0.5275]
- [23,4] Beta(1185,10.58)[mean=0.9912]
-

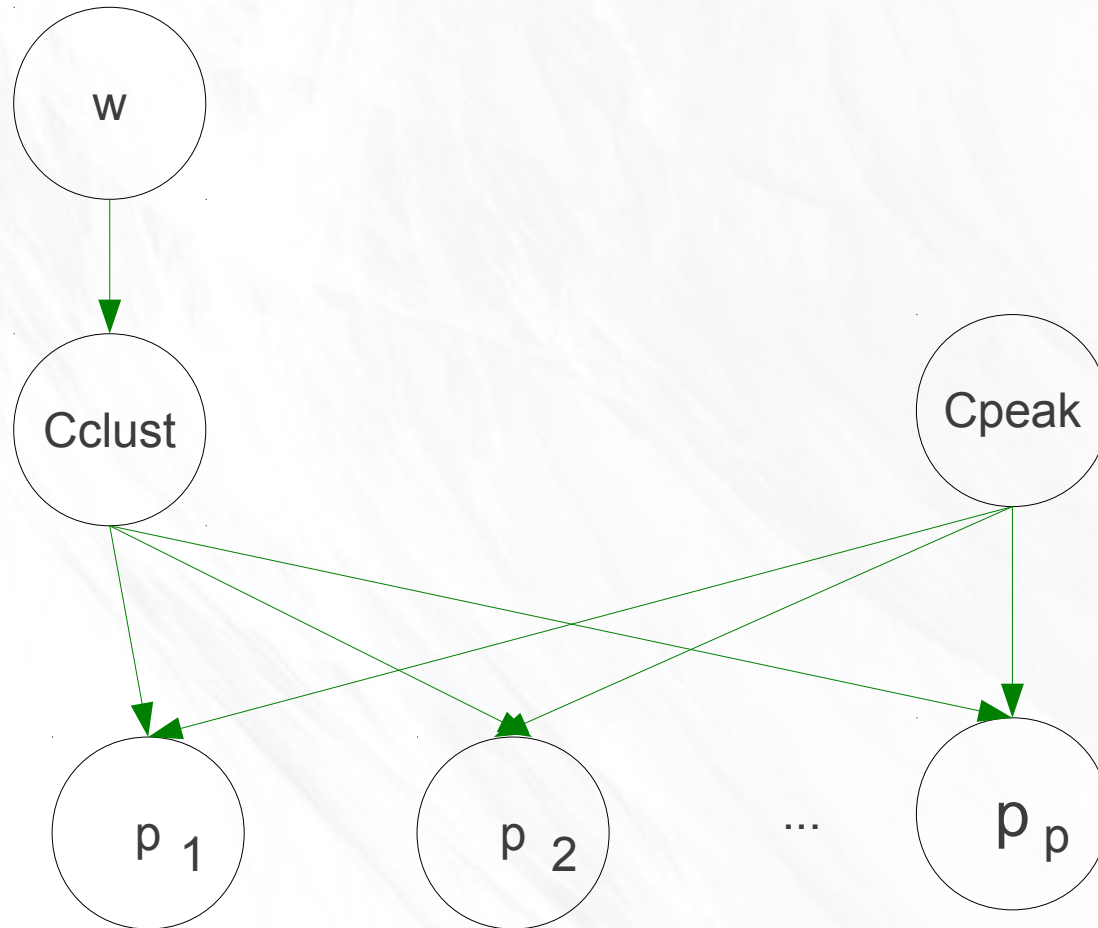
Biclustering

-
- [24,1] Beta(312.7,341.4)[mean=0.478]
- [24,2] Beta(1031,224.2)[mean=0.8214]
- [24,3] Beta(21.55,941.1)[mean=0.02239]
- [24,4] Beta(1154,41.48)[mean=0.9653]
- [24,5] Beta(120.8,1340)[mean=0.08272]
- [24,6] Beta(3.621,1608)[mean=0.002246]
- [24,7] Beta(637.7,31.44)[mean=0.953]
- [24,8] Beta(1.16,1462)[mean=0.0007926]
-

Clustering

- Preprocessed data:
- 95 1 0.5937
- 98 1 0.8154
- 101 1 0.6695
- 103 1 0.5244
- 104 1 0.8831
- 107 1 0.9963
- 145 1 0.9444
- 150 1 0.4687
- 159 1 0.6575
-

Biclustering



Спасибо!