

Международная научная конференция  
Параллельные вычислительные технологии (ПаВТ'2021)  
Волгоград, 30 марта – 1 апреля 2021 г.

---

**Обнаружение аномалий  
в больших временных рядах  
на основе совместного использования  
нейронных сетей  
и параллельных алгоритмов**

Я.А. Краева, М.Л. Цымблер

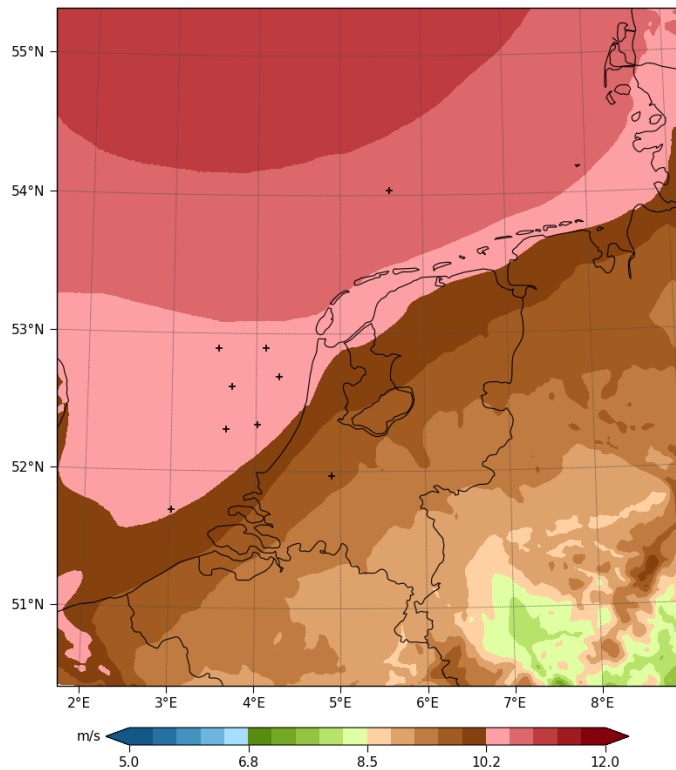
Южно-Уральский государственный университет (Челябинск)

---

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-07-00140) и Министерства образования и науки РФ (гос. задание FENU-2020-0022)

# Поиск аномалий в режиме офлайн

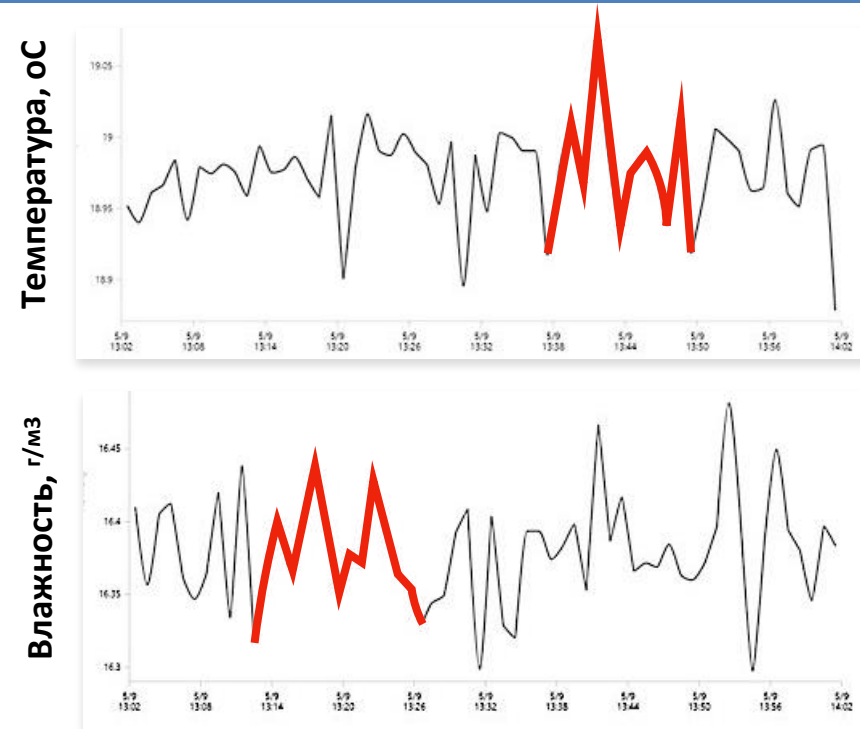
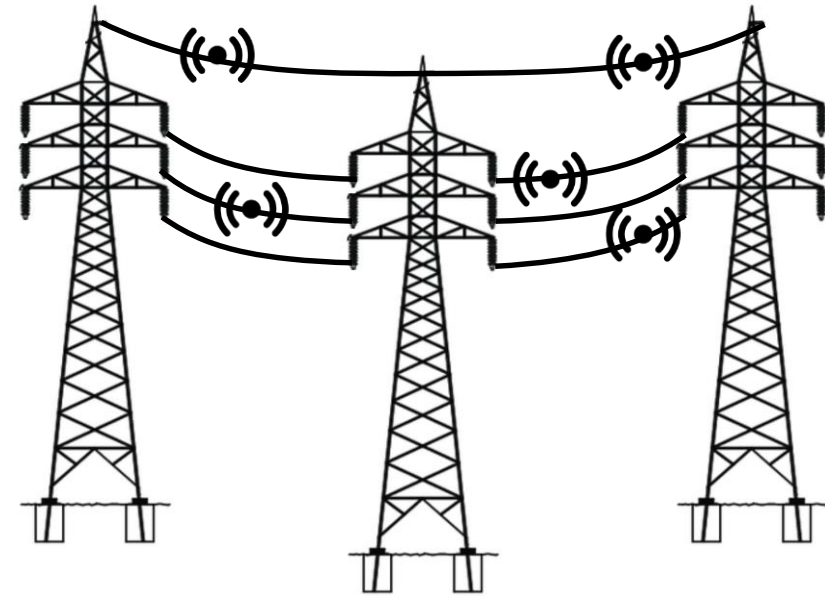
Wind speed  
500m, 2008-2017 mean  
Dutch Offshore Wind Atlas (DOWA)



- Поиск аномалий в данных атласа морских ветров за **10 лет**<sup>1)</sup> позволяет найти локации для установки прибрежных ветряных электростанций, дающие наибольшую производительность
- Длина ряда: **15 млрд.**
- Размер данных: **2.8 Тб**

<sup>1)</sup> De Valk C., Wijnant I.L. Uncertainty analysis of climatological parameters of the Dutch Offshore Wind Atlas (DOWA). 2019. Technical report TR-379. Royal Netherlands Meteorological Institute




# Поиск аномалий в режиме онлайн



- Поиск аномалий в показаниях **более  $10^4$  датчиков** позволяет оперативно обнаружить повреждения в ЛЭП<sup>1)</sup>
- Дискретность показаний датчика: **240 раз в сек.**

<sup>1)</sup> Leon R. A., Vittal V., Manimaran G. Application of Sensor Network for Secure Electric Energy Infrastructure. IEEE Transactions on Power Delivery. 2007. Vol. 22, No. 2. P. 1021–1028.

# Поиск аномалий временного ряда: нейросети **vs.** параллельные алгоритмы

Метод	Длина ряда	Обучение	Использование
Нейронная сеть LSTM <sup>1)</sup>	$10^4$	 $10^3$ сек. GPU (NVIDIA Tesla K80, 4.1 TFLOPS)	 $\leq 1$ сек. CPU (Intel Xeon, 0.4 TFLOPS)
Параллельный алгоритм PDADD <sup>2)</sup>	$64 \cdot 10^6$	 0 сек.	 $10^3$ сек. Кластер 64 узла (Intel Xeon Phi SE10X, 1 TFLOPS)

<sup>1)</sup> Braei M., Wagner S. Anomaly detection in univariate time-series: A survey on the state-of-the-art. 2020.

<sup>2)</sup> Zymbler M., Grents A., Kraeva Ya., Kumar S. A parallel approach to discords discovery in massive time series data. Computers, Materials & Continua. 2021. Vol. 66, No. 2. P. 1867–1876.

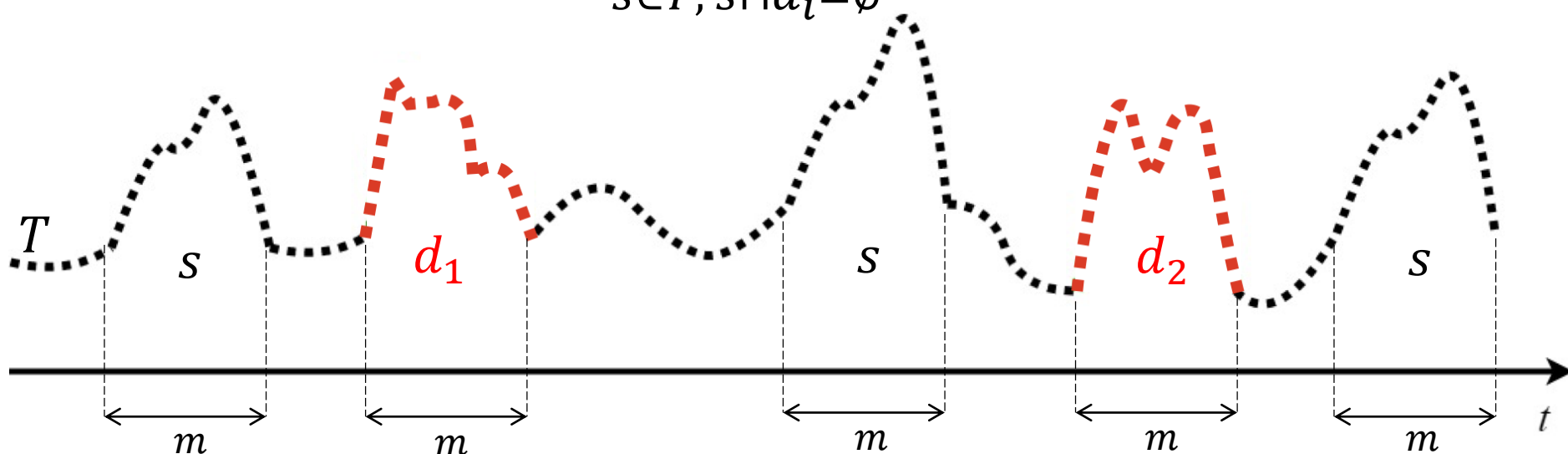
# Поиск аномалий временного ряда: нейросети + параллельные алгоритмы

- Параллельные алгоритмы
  - Найти аномалии
  - Подготовить обучающую выборку для нейросети
- Нейросети
  - Обучиться на подготовленной выборке
  - Выявлять аномалии в режиме онлайн

# Диссонансы

- **Диссонанс**<sup>1)</sup> – подпоследовательность ряда, расстояние от которой до ее ближайшего соседа не ниже заданного порога  $r$

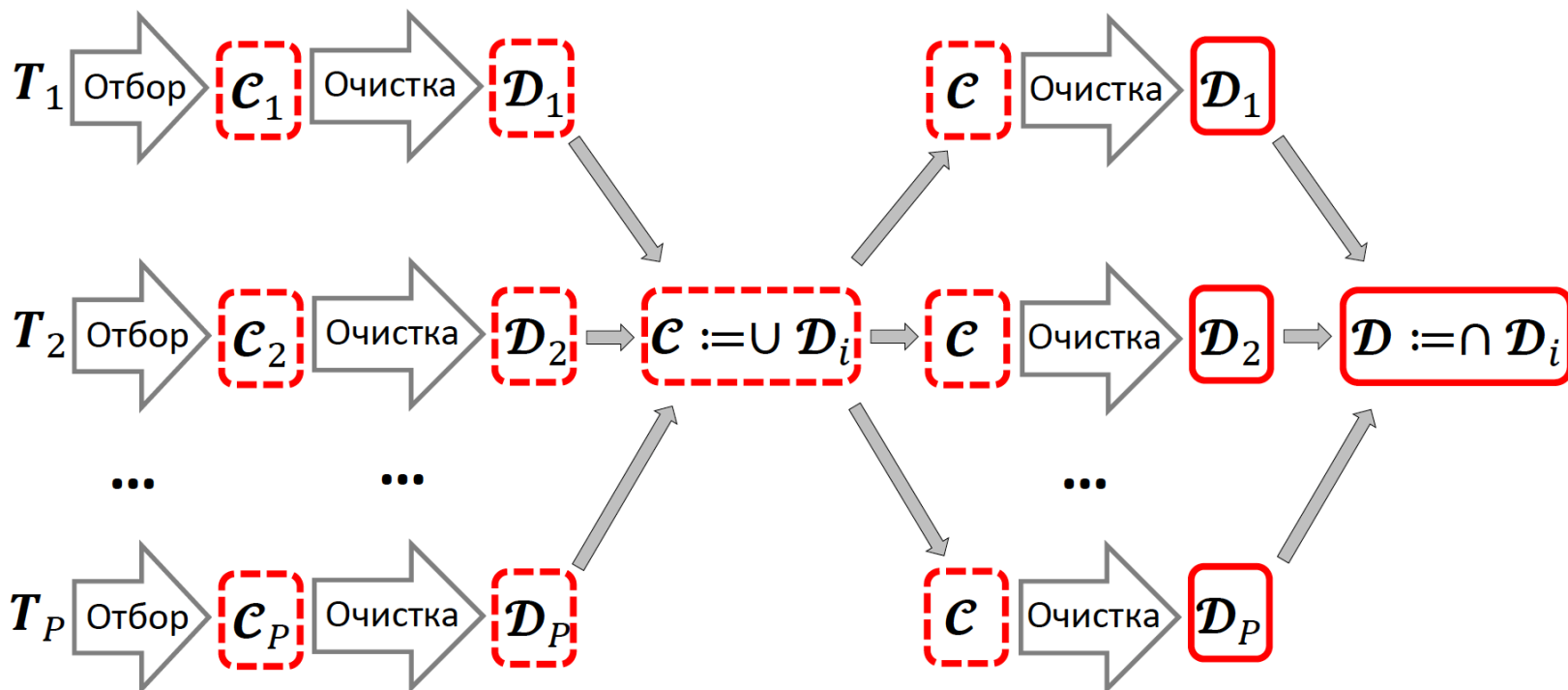
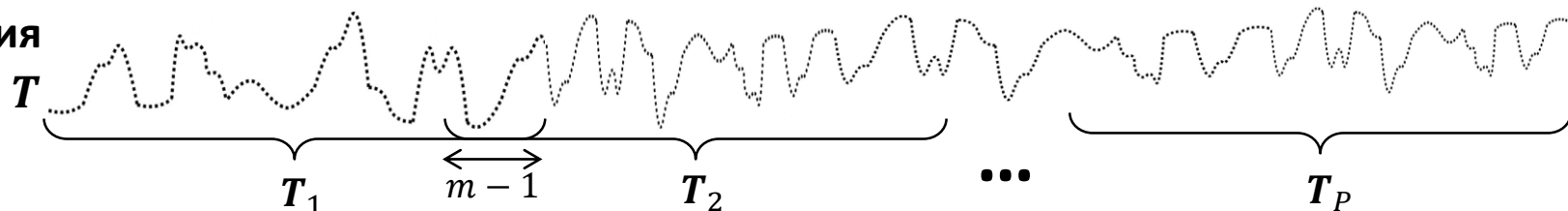
- **Множество диссонансов**  $\mathcal{D} = \{d_i \mid 1 \leq i \leq k\}$ ,  
 $d_i \in \mathcal{D} \Leftrightarrow \min_{s \in T, s \cap d_i = \emptyset} ED(d_i, s) \geq r$



<sup>1)</sup>Yankov D., Keogh E.J., Rebbapragada U. Disk aware discord discovery: finding unusual time series in terabyte sized datasets. Knowl. Inf. Syst. 2008. Vol. 17, No. 2. P. 241–262.

# Параллельный поиск диссонансов<sup>1,2)</sup>

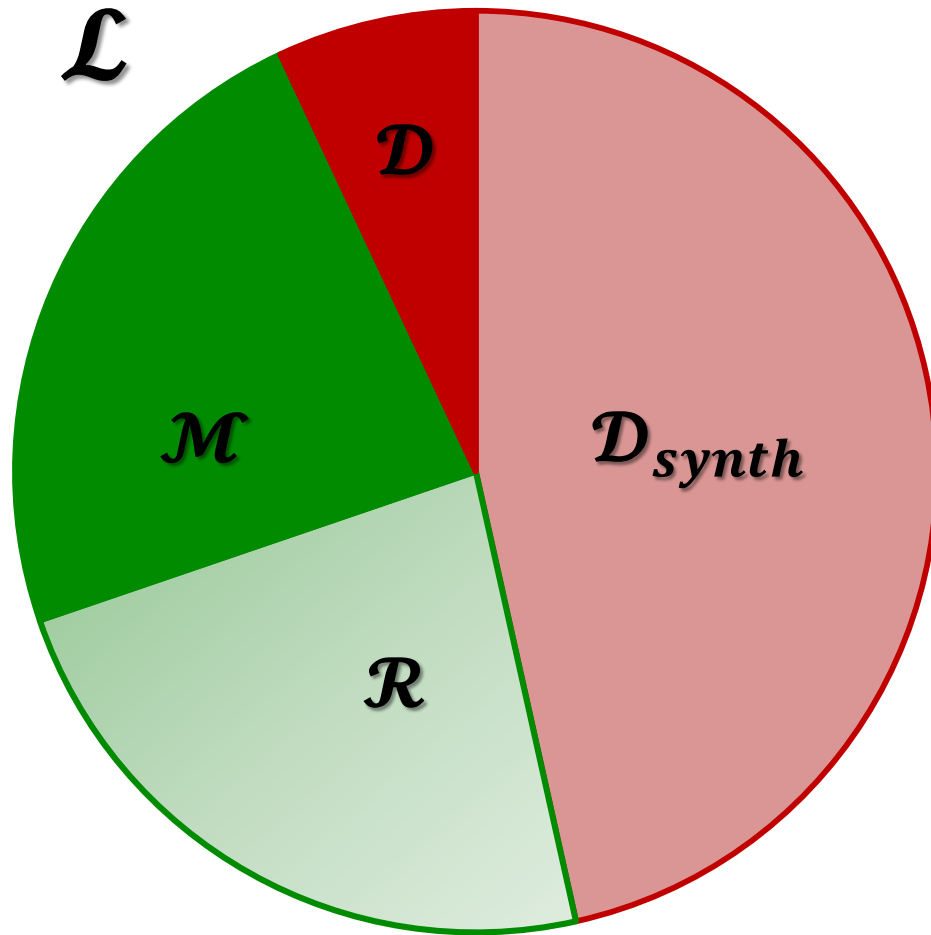
Фрагментация



<sup>1)</sup> Zymbler M., Grents A., Kraeva Ya., Kumar S. A parallel approach to discords discovery in massive time series data. Computers, Materials & Continua. 2021. Vol. 66, No. 2. P. 1867–1876.

<sup>2)</sup> Zymbler M., Polyakov A., Kipnis M. Time series discord discovery on Intel many-core systems // 13th Int. Conf., PCT 2019, Kaliningrad, Russia, April 2–4, 2019. CCIS. 2019. Vol. 1063. P. 168–182.

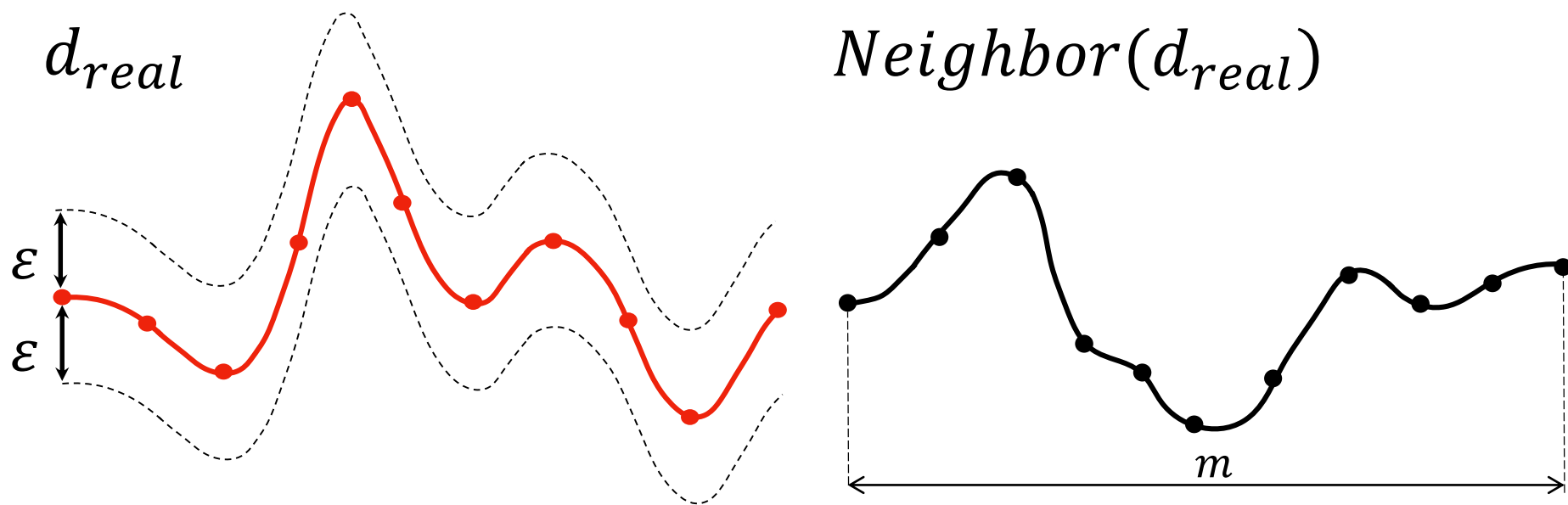
# Формирование обучающей выборки



- $\mathcal{D}$   
реальные диссонансы
- $\mathcal{D}_{synth}$   
синтетические диссонансы
- $\mathcal{M}$   
лейтмотивы (шаблонные подпоследовательности)
- $\mathcal{R}$   
случайные подпоследовательности



# Синтетические диссонансы



Порог схожести

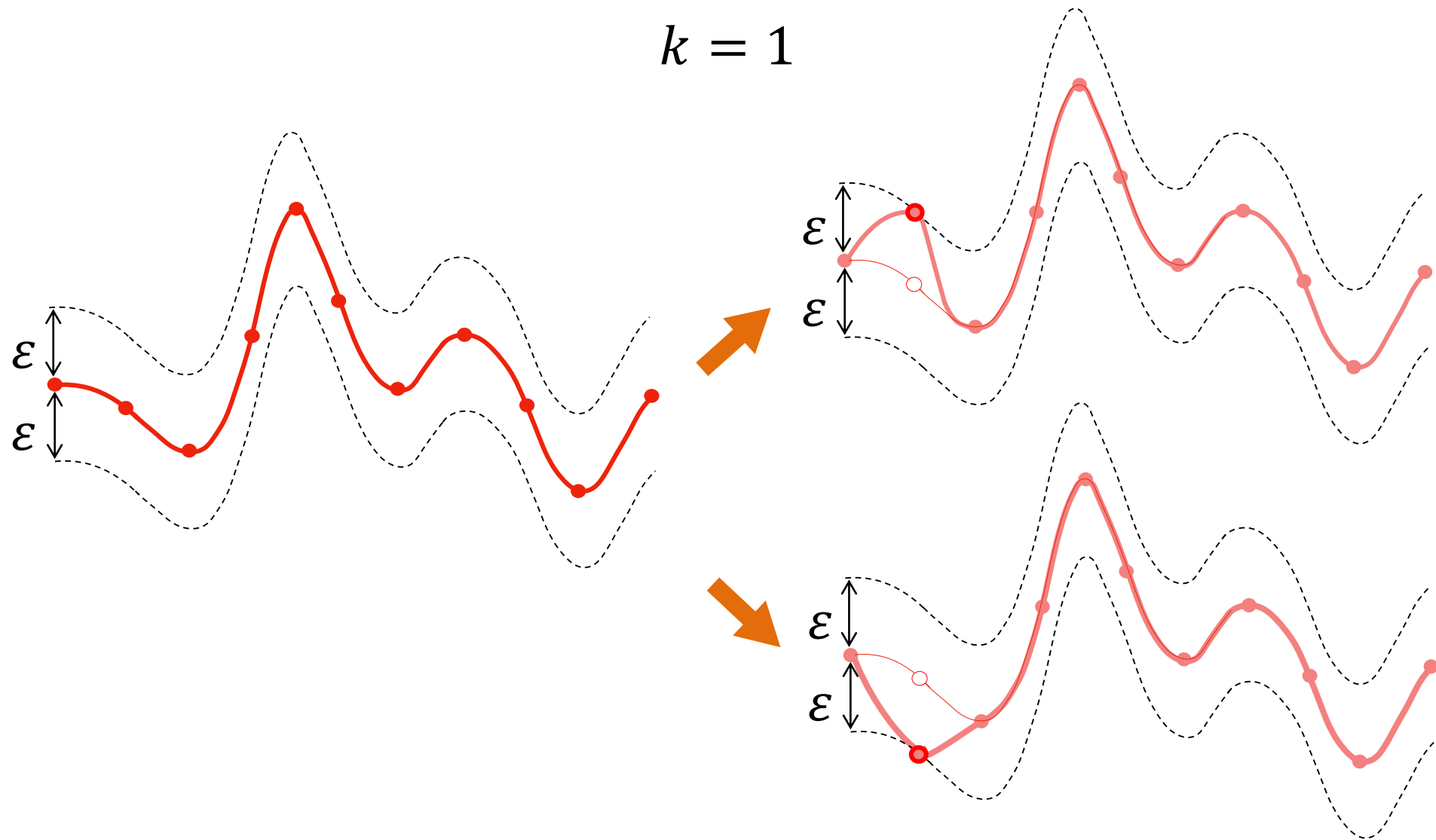
$$\epsilon = ED(d_{real}, Neighbor(d_{real}))$$

Количество разложений  $\epsilon$  на  $m$  слагаемых (в т.ч.  $k$  ненулевых)

$$C_m^{m+k-1} = \frac{(m+k-1)!}{m!(k-1)!}$$

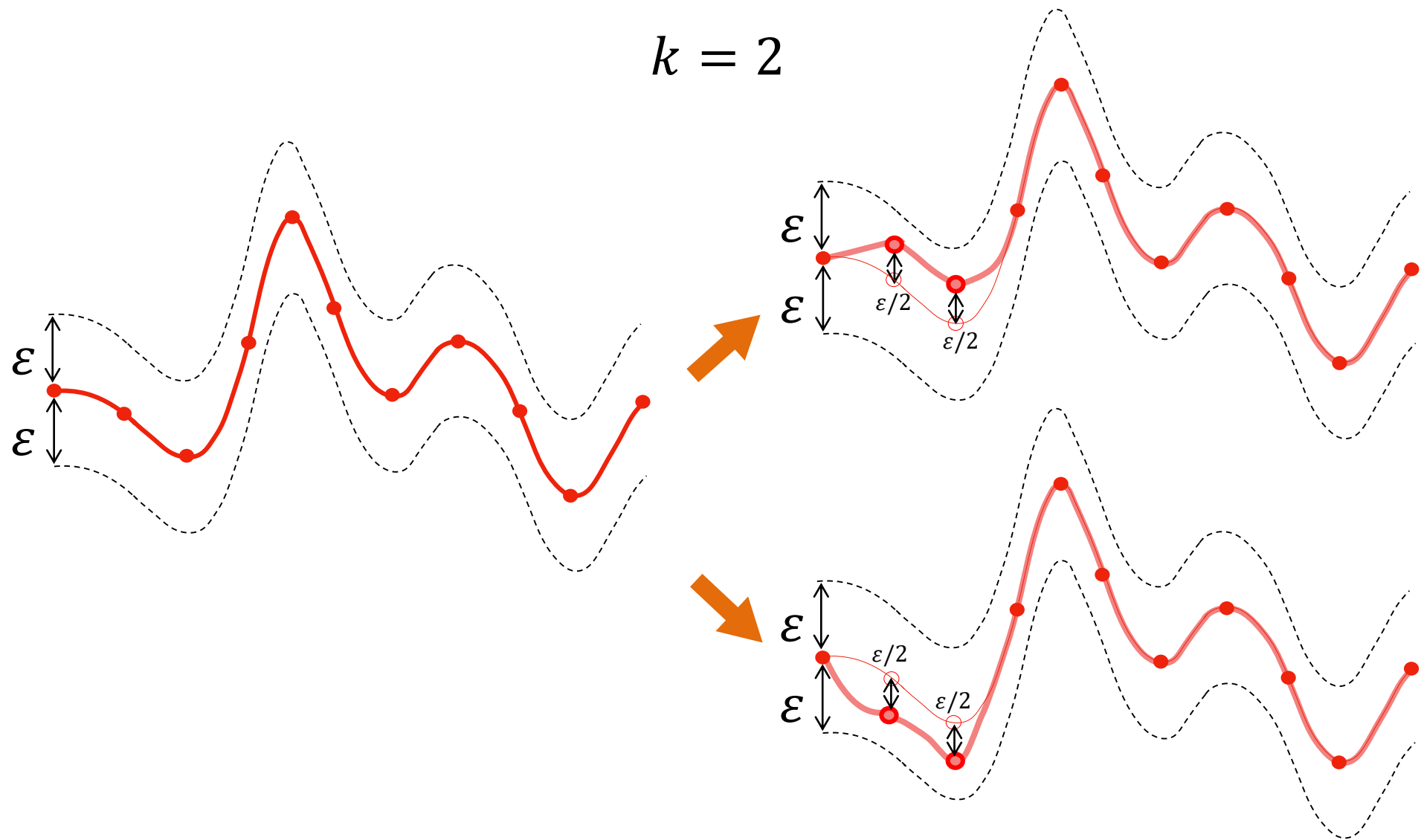
# Синтетические диссонансы

$k = 1$



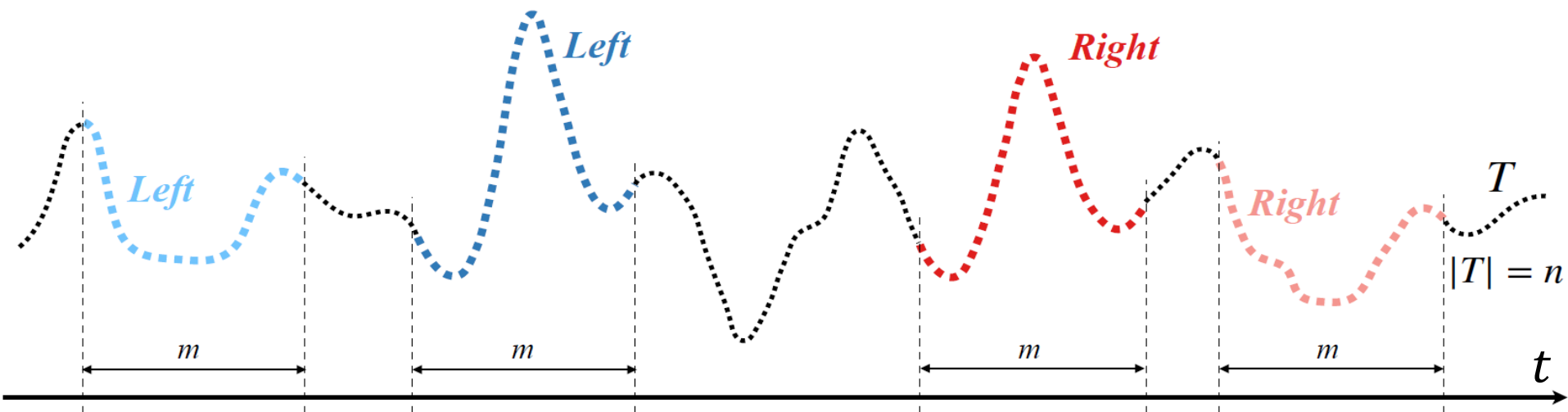
# Синтетические диссонансы

$k = 2$



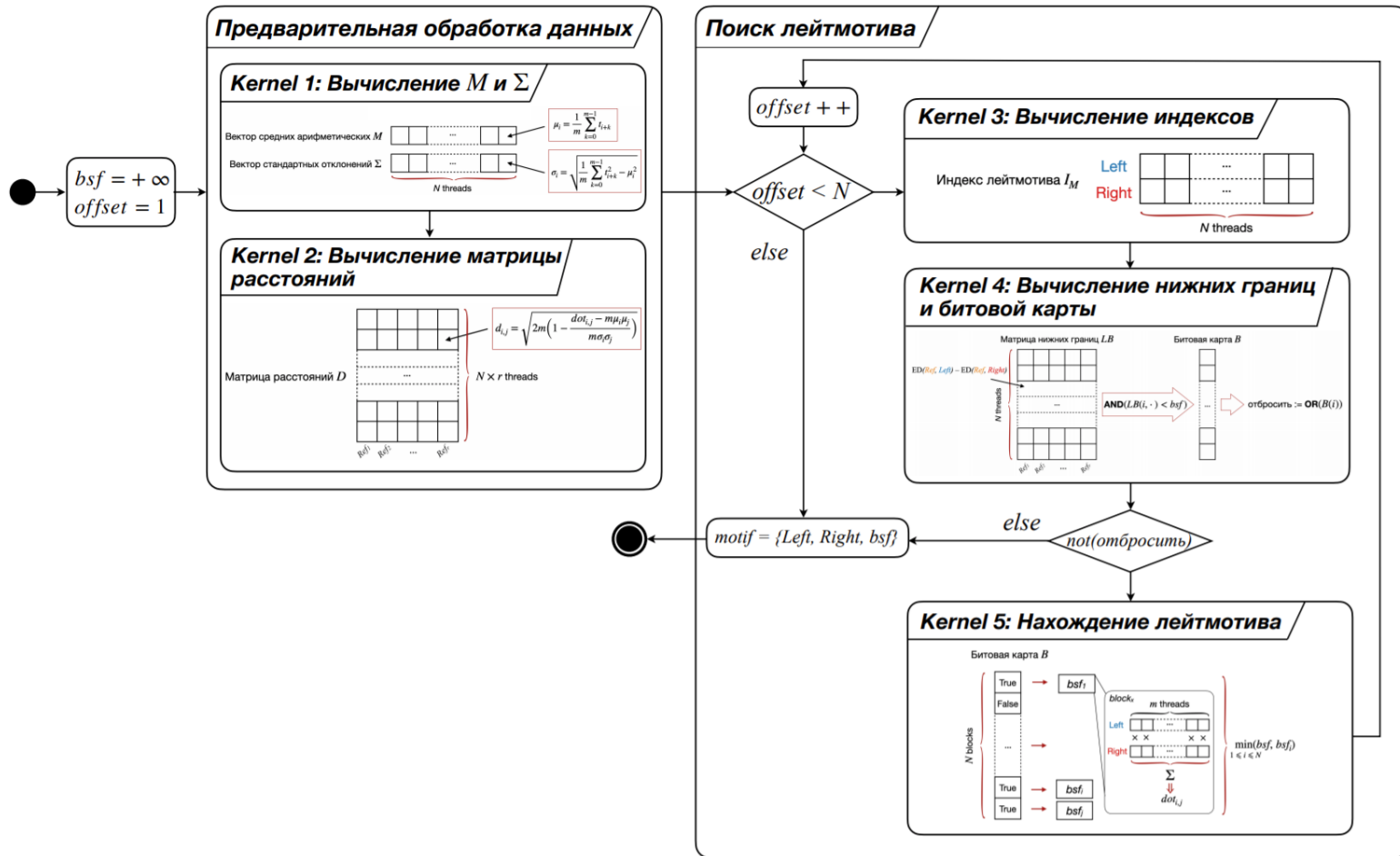
# Лейтмотивы

- **Лейтмотив**<sup>1)</sup> – пара подпоследовательностей ряда, наиболее похожих друг на друга
- **Множество лейтмотивов**  $\mathcal{M} = \{(\ell_i, r_i) \mid 1 \leq i \leq k\}$ ,  
 $(\ell_i, r_i) \in \mathcal{M} \Leftrightarrow \forall \ell \cap r = \emptyset \text{ ED}(\ell_i, r_i) < \text{ED}(\ell, r)$



<sup>1)</sup> Mueen A. , Keogh E. J., Q. Zhu, *et al.* Exact discovery of time series motifs. Proc. of the SIAM Int. Conf. on Data Mining, SDM 2009, April 30 – May 2, 2009, Sparks, Nevada, USA. P. 473–484.

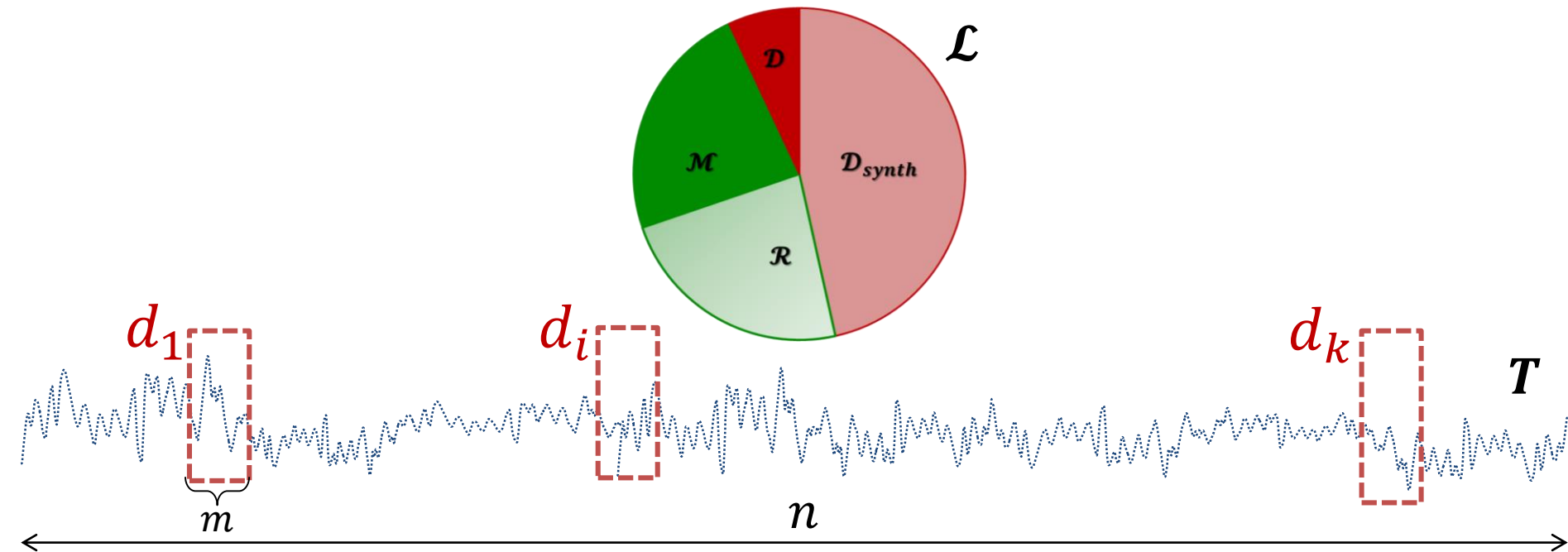
# Параллельный поиск лейтмотивов<sup>1,2)</sup>



1) Zymbler M., Kraeva Ya. Discovery of time series motifs on Intel many-core systems. Lobachevskii Journal of Mathematics. 2019. Vol. 40, No. 12. P. 2124–2132.

2) Цымблер М.Л., Краева Я.А. Параллельный алгоритм поиска лейтмотивов временного ряда для графического процессора. Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2020. Т. 9, № 3. С. 17–34.

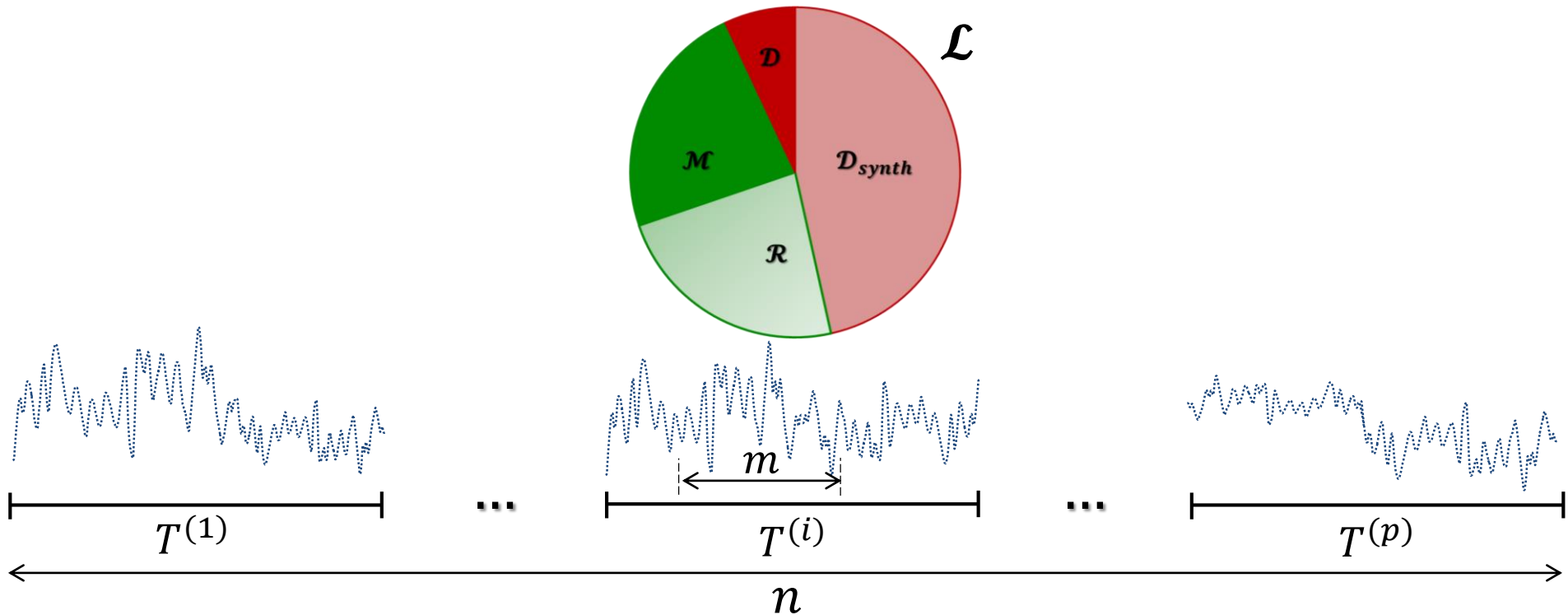
# Параметры выборки: диссонансы



Доля диссонансов  
среди всех подпоследовательностей ряда

$$\alpha = \frac{|D|}{n - m + 1}$$

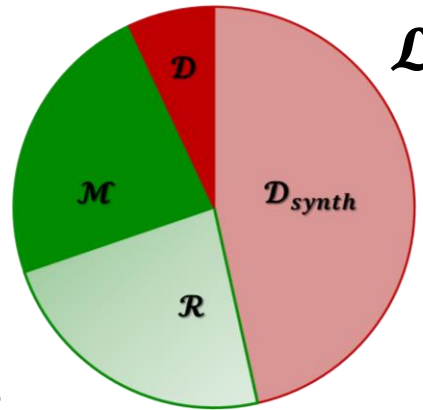
# Параметры выборки: случайные подпоследовательности



Доля случайных подпоследовательностей фрагмента

$$\beta = \frac{p}{n - m + 1} \cdot |\mathcal{R}^{(i)}|$$

# Параметры выборки: лейтмотивы, синтетические диссонансы



- Доля диссонансов

$$r_{discord} = \frac{|\mathcal{D}| + |\mathcal{D}_{synth}|}{|\mathcal{L}|}$$

- Доля лейтмотивов среди нормальных подпоследовательностей

$$r_{motif} = \frac{|\mathcal{M}|}{|\mathcal{M} \cup \mathcal{R}^{(i)}|}$$



# Вычисление размеров подмножеств

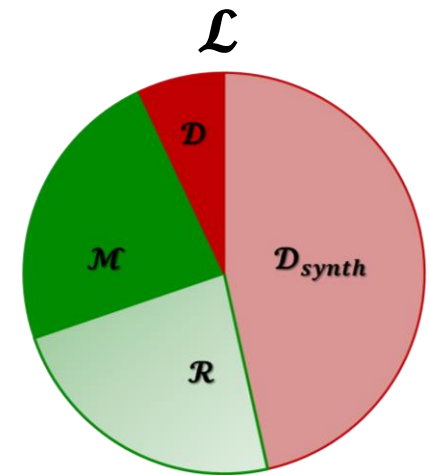
$$N ::= n - m + 1$$

- **Диссонансы**

$$|\mathcal{D}| = \lceil \alpha \cdot N \rceil$$

- **Случайные**

$$|\mathcal{R}^{(i)}| = \lceil \beta \cdot \frac{N}{p} \rceil$$



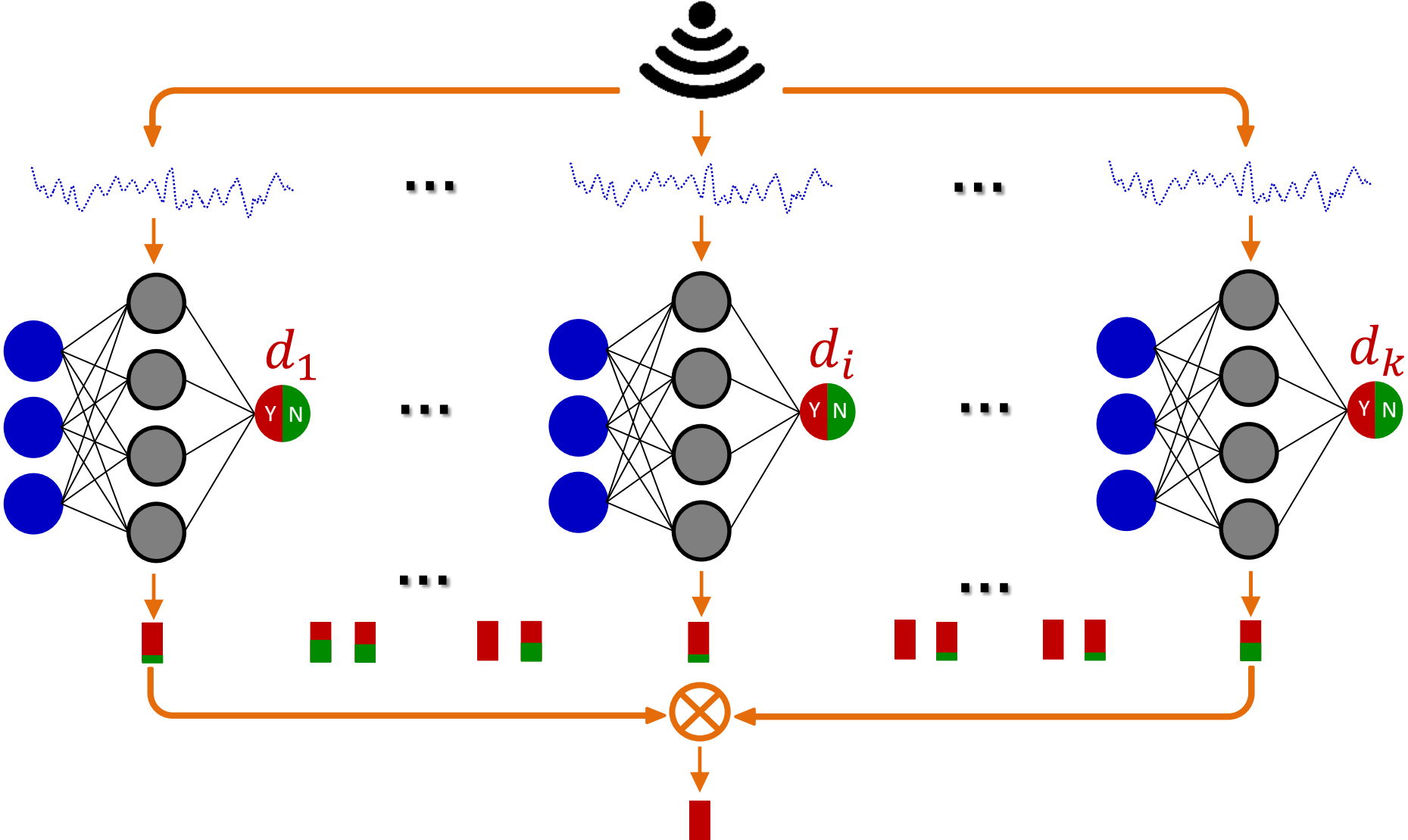
- **Лейтмотивы**

$$|\mathcal{M}^{(i)}| = \lceil N \cdot \frac{r_{motif}}{1 - r_{motif}} \cdot \frac{\beta}{p^2} \rceil$$

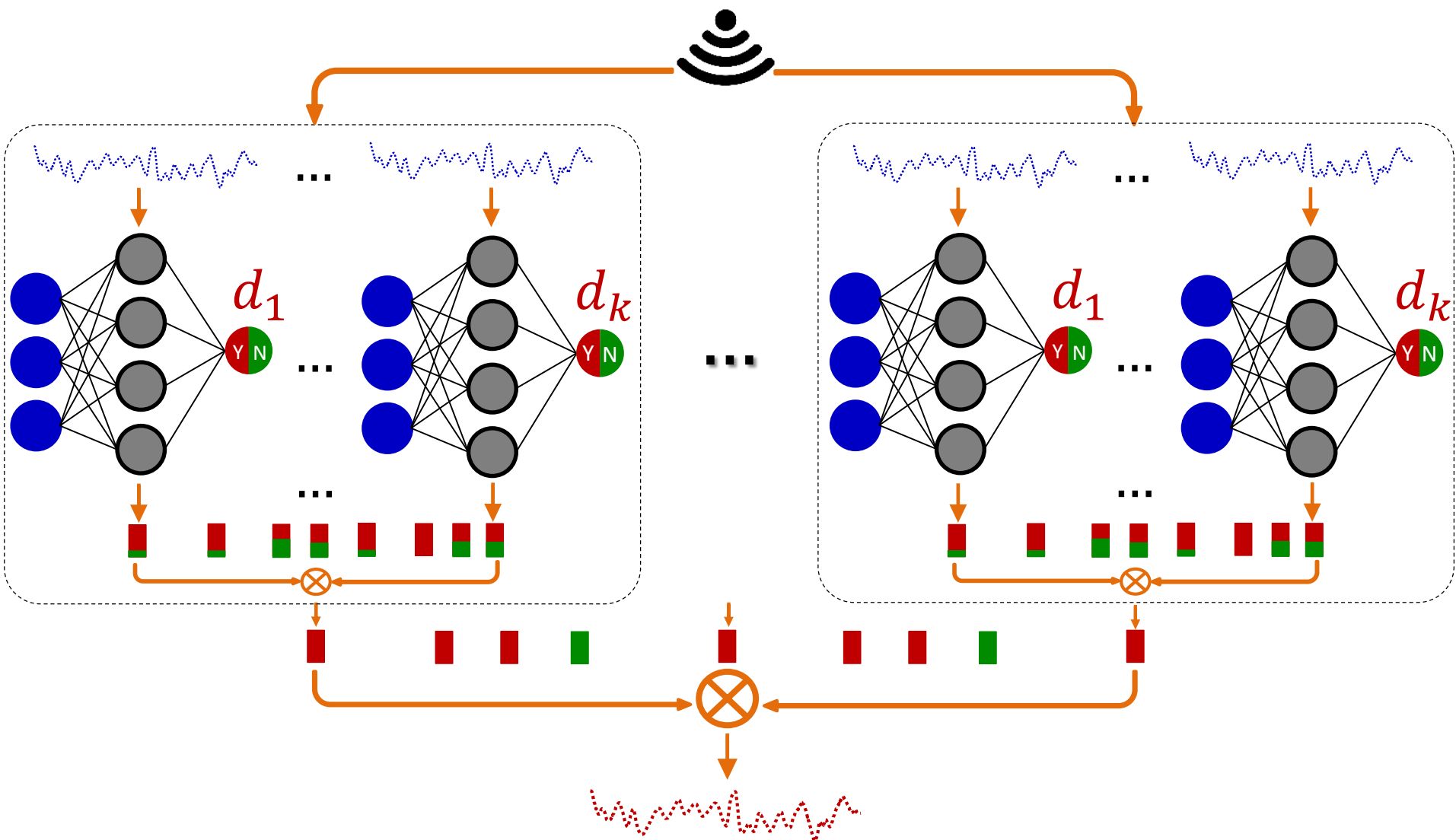
- **Синтетические диссонансы**

$$|\mathcal{D}_{synth}| = \lceil N \cdot \left( \frac{r_{discord}}{(1 - r_{discord})(1 - r_{motif})} \cdot \frac{\beta}{p} - \alpha \right) \rceil$$

# Ансамбль нейронных сетей (один узел кластера)



# Взаимодействие ансамблей нейронных сетей (кластер)



# Начальные результаты и планы

- Разработан подход к обнаружению аномалий в больших временных рядах в режиме онлайн на основе совместного использования нейронных сетей и параллельных алгоритмов
- Планы исследований
  - Эксперименты на больших временных рядах ( $n \geq 10^6$ )
  - Различные архитектуры нейросети-участника ансамбля

Спасибо за внимание!

Вопросы?