

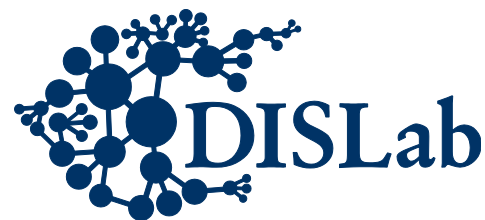
Спецкурс

«Параллельная обработка больших графов»

Лекция 5

А.С. Семенов

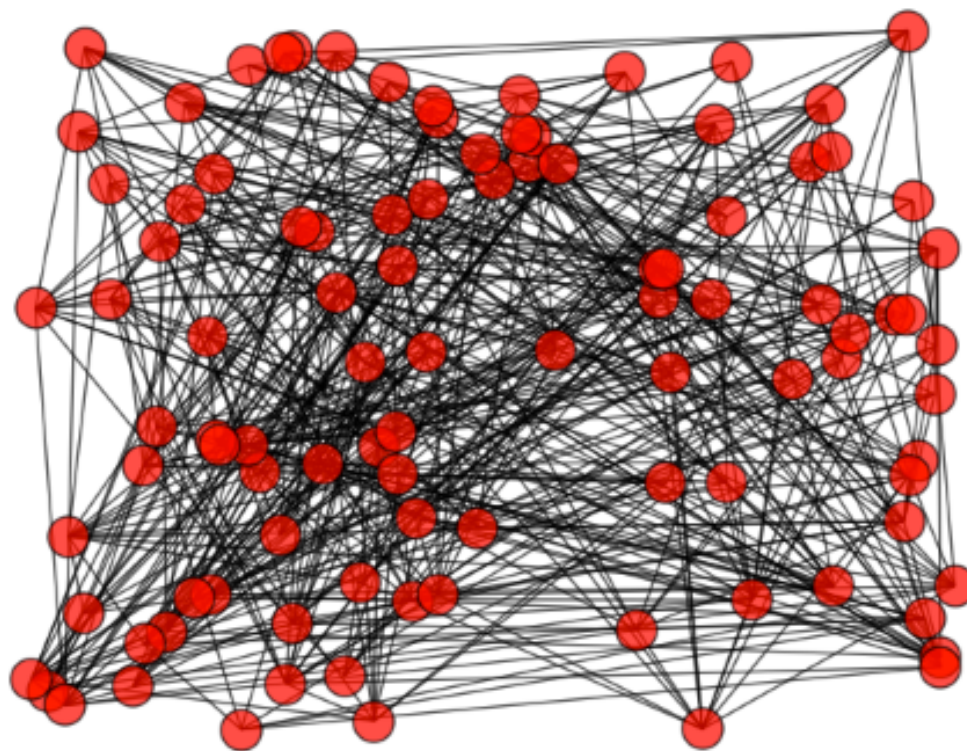
dislab.org



Виды графов

Виды графов. Случайные графы

- Random, Random Uniform, Erdos Renyi
- N вершин, M ребер, k – средняя связность вершины



Виды графов. Степенной закон

- WWW, Социальные сети, Биоинформатика

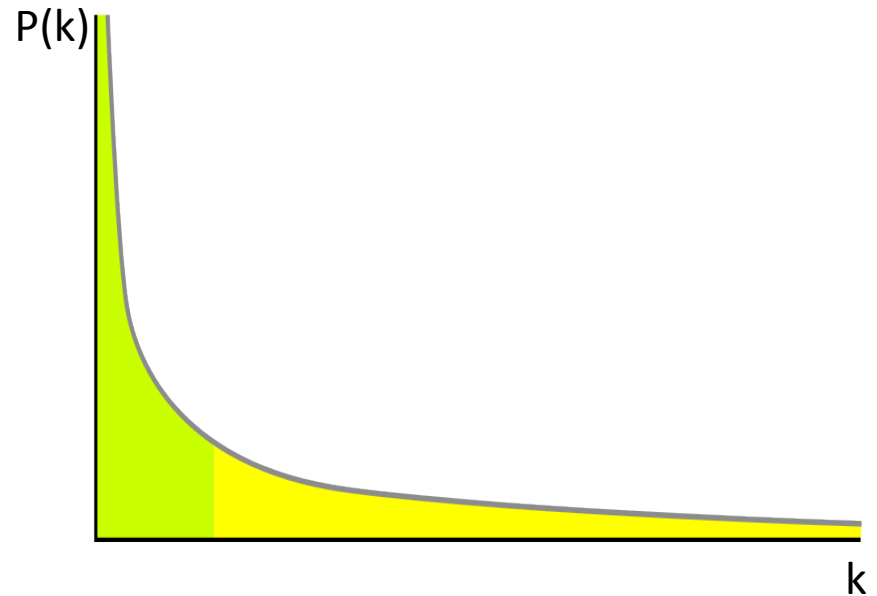
- Графы small-world

$L \sim \log N$

- scale-free – графы,
доля $P(k) \sim k^{-\tau}$, $2 < \tau < 3$

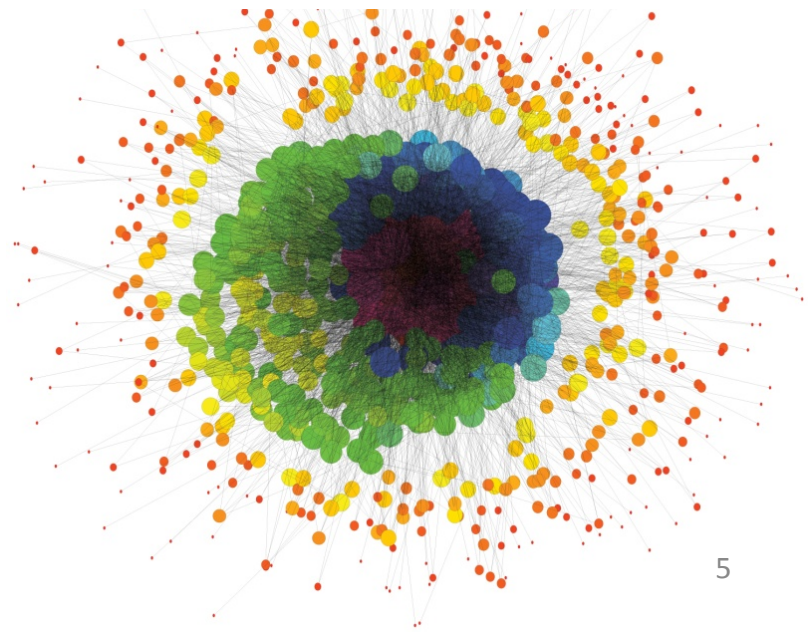
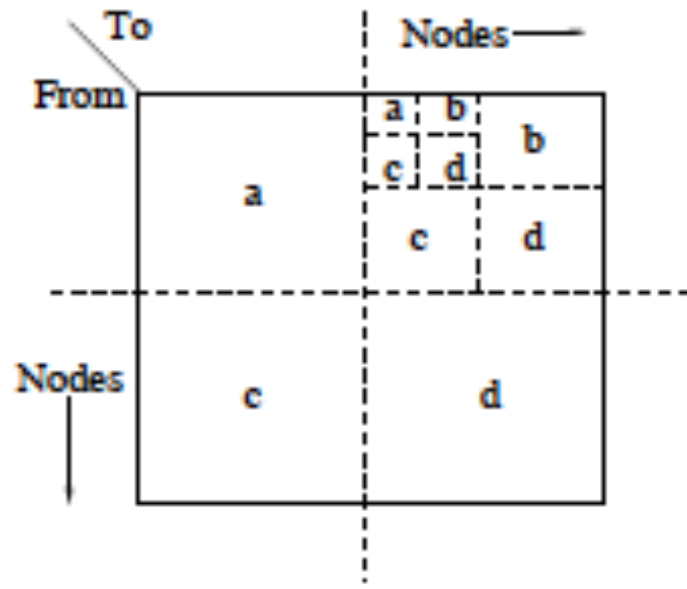
k – связность вершины

$L \sim \log \log N$



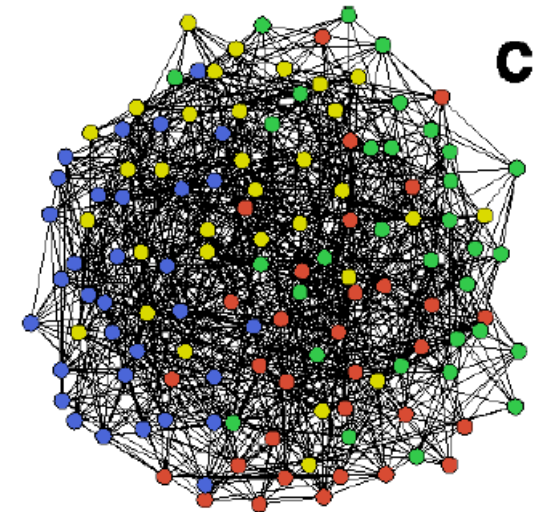
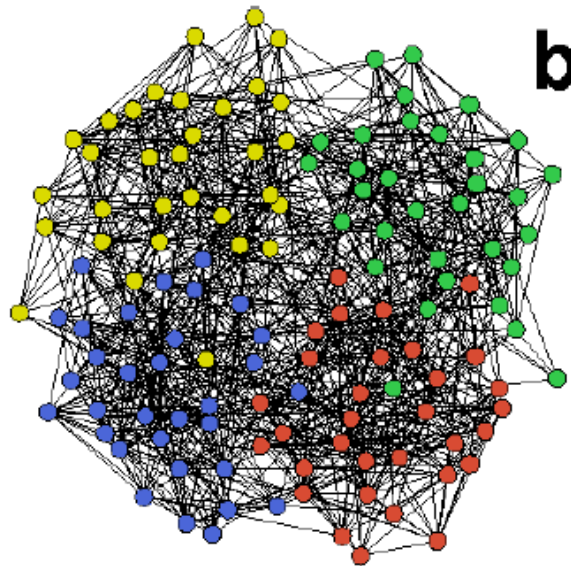
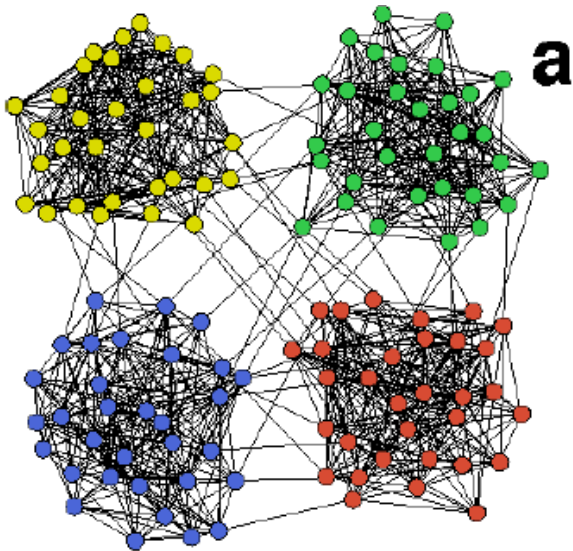
Виды графов. RМAТ-граф

- $a+b+c+d = 1$
- Сообщества:
 - a и d – сообщества
 - b и c – связи между ними
 - наличие «подсообществ»
- может быть scale-free при $a \geq d$
- случайная перестановка вершин



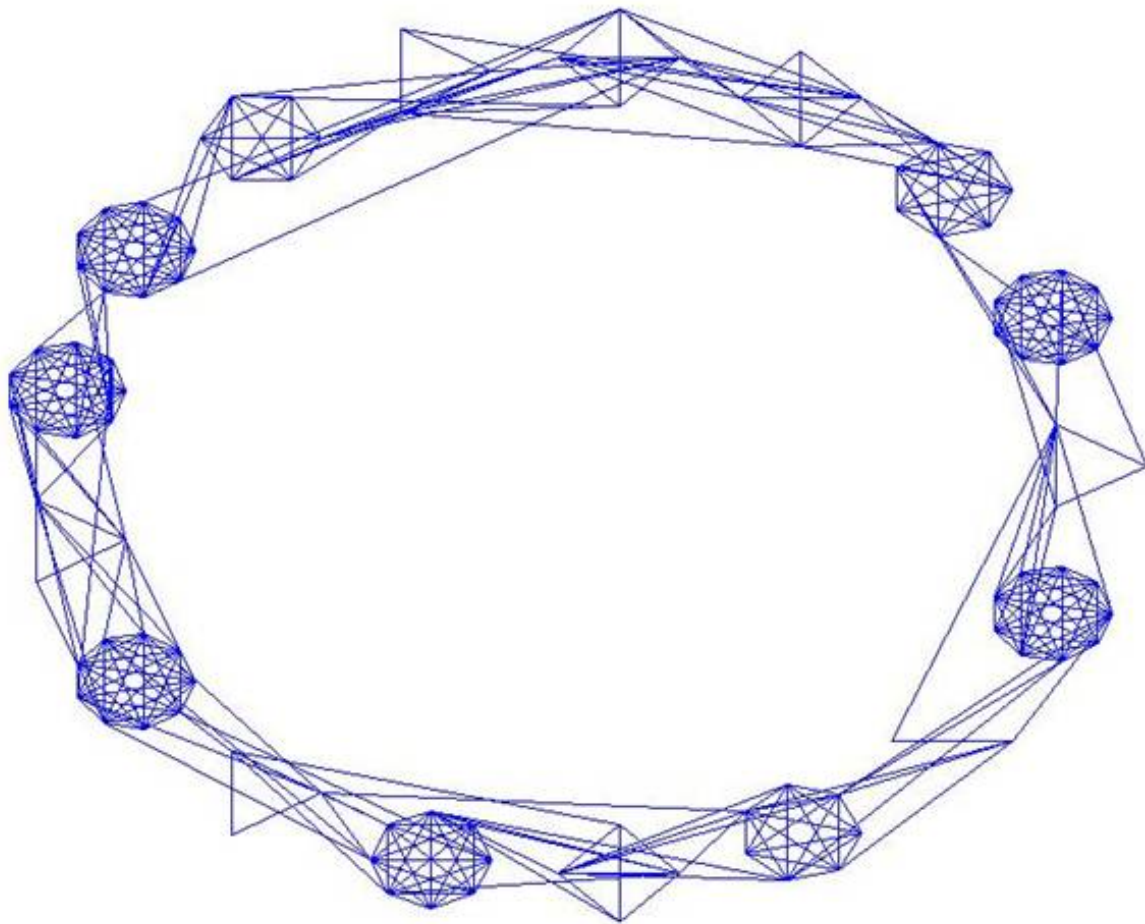
Виды графов. LFR*-граф

- **Параметры:**
 - $m_i \in [0;1]$, показывает количество связей вне сообщества
 - com_tau – показатель степени в законе распределения размеров сообществ
 - deg_tau – показатель степени в законе распределения степеней вершин



Виды графов. SSCA2-граф

- Равномерное распределение случайных параметров
- случайная перестановка вершин

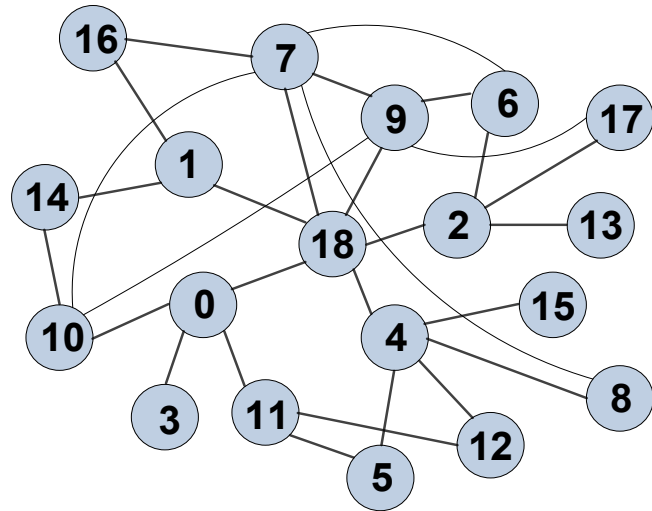


Проблемы и подходы к решению задач обработки графов в рамках одного вычислительного узла

Проблемы анализа больших графов

- **Data-driven computations.** Зависимость вычислений от данных (топологии графа). Невозможность применения методов статического распараллеливания вычислений.
- **Unstructured problems.** Работа с нерегулярными, неструктурированными данными, трудность распараллеливания.
- **Poor locality.** Низкая пространственно-временная локализация обращений к памяти.
- **High data access to computation ratio.** Преобладание команд доступа к памяти над командами выполнения арифметических операций.

Представление графа



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|----|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
| 0 | | | | 1 | | | | | | | 1 | 1 | | | | | | | 1 |
| 1 | | | | | | | | | | | | | | | 1 | | 1 | | 1 |
| 2 | | | | | | | 1 | | | | | | 1 | | | | | 1 | 1 |
| 3 | 1 | | | | | | | | | | | | | | | | | | |
| 4 | | | | | | 1 | | 1 | | | | 1 | | | 1 | | | | 1 |
| 5 | | | | | 1 | | | | | | | 1 | | | | | | | |
| 6 | | | 1 | | | | | 1 | | 1 | | | | | | | | | |
| 7 | | | | | | | 1 | | 1 | 1 | 1 | | | | | | | 1 | 1 |
| 8 | | | | | 1 | | | 1 | | | | | | | | | | | |
| 9 | | | | | | | 1 | 1 | | | 1 | | | | | | | | 1 |
| 10 | 1 | | | | | | | 1 | | 1 | | | | | 1 | | | | |
| 11 | 1 | | | | | 1 | | | | | | | 1 | | | | | | |
| 12 | | | | | 1 | | | | | | | 1 | | | | | | | |
| 13 | | | 1 | | | | | | | | | | | | | | | | |
| 14 | | 1 | | | | | | | | | 1 | | | | | | | | |
| 15 | | | | | 1 | | | | | | | | | | | | | | |
| 16 | | 1 | | | | | | 1 | | | | | | | | | | | |
| 17 | | | 1 | | | | | | | 1 | | | | | | | | | |
| 18 | 1 | 1 | 1 | | 1 | | | 1 | | 1 | | | | | | | | | |

Форматы представления разреженных матриц

- **Доля ненулевых элементов мала**

Можно хранить только позиции и значения ненулевых элементов

- **Compressed Row Storage (CRS)**
- **Coordinate list (COO)**
- **DIA**
- **ELLPACK**
- **SELLPACK**
- **Оптимизированный под задачу**

Внутреннее представление Compressed Row Storage (CRS)

Sparse Matrix

| | | | | |
|----|---|---|---|----|
| 10 | 0 | 0 | 0 | -2 |
| 3 | 9 | 0 | 0 | 0 |
| 0 | 7 | 8 | 7 | 0 |
| 3 | 0 | 8 | 7 | 5 |
| 0 | 8 | 0 | 9 | 13 |

Row pointer array

| | | | | | |
|---|---|---|---|----|----|
| 0 | 2 | 4 | 7 | 11 | 14 |
|---|---|---|---|----|----|

rowsIndices

Column indices array

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 4 | 0 | 1 | 1 | 2 | 3 | 0 | 2 | 3 | 4 | 1 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

endV

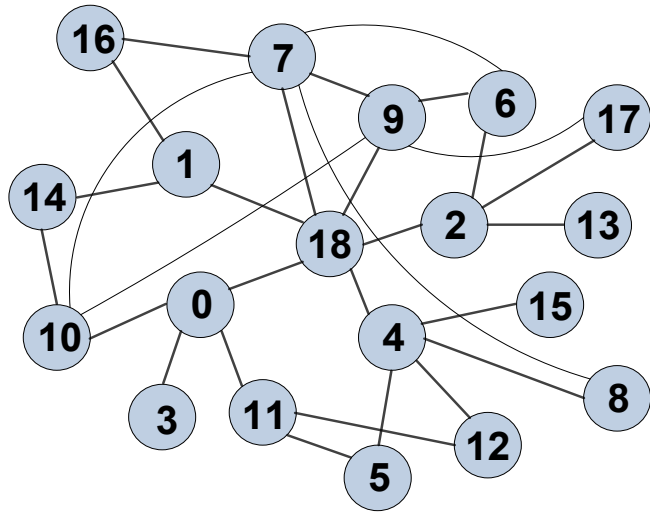
Values array

| | | | | | | | | | | | | | |
|----|----|---|---|---|---|---|---|---|---|---|---|---|----|
| 10 | -2 | 3 | 9 | 7 | 8 | 7 | 3 | 8 | 7 | 5 | 8 | 9 | 13 |
|----|----|---|---|---|---|---|---|---|---|---|---|---|----|

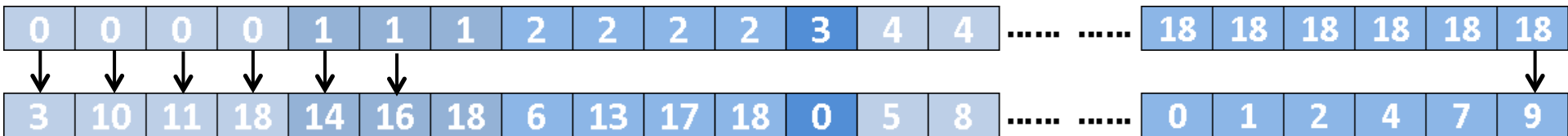
weights

```
for (int u = 0; u < G->n; u++) {  
    for (int j = G->rowsIndices[u]; j < rowsIndices[u+1];  
        j++) {  
        const int v = G->endV[j];  
        const int w = G->weights[j];  
        // обработка ребра u->v  
    }  
}
```

Coordinate list (COO)



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|----|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
| 0 | | | | 1 | | | | | | | 1 | 1 | | | | | | | 1 |
| 1 | | | | | | | | | | | | | | | 1 | | 1 | | 1 |
| 2 | | | | | | | 1 | | | | | | | 1 | | | | 1 | 1 |
| 3 | 1 | | | | | | | | | | | | | | | | | | |
| 4 | | | | | | 1 | | 1 | | | | | 1 | | 1 | | | | 1 |
| 5 | | | | | 1 | | | | | | | 1 | | | | | | | |
| 6 | | | 1 | | | | | 1 | | 1 | | | | | | | | | |
| 7 | | | | | | | 1 | | 1 | 1 | 1 | | | | | | 1 | | 1 |
| 8 | | | | | 1 | | | 1 | | | | | | | | | | | |
| 9 | | | | | | | 1 | 1 | | | 1 | | | | | | | 1 | 1 |
| 10 | 1 | | | | | | | 1 | | 1 | | | | | 1 | | | | |
| 11 | 1 | | | | | 1 | | | | | | | | 1 | | | | | |
| 12 | | | | | 1 | | | | | | | 1 | | | | | | | |
| 13 | | | 1 | | | | | | | | | | | | | | | | |
| 14 | 1 | | | | | | | | | | 1 | | | | | | | | |
| 15 | | | | | 1 | | | | | | | | | | | | | | |
| 16 | | 1 | | | | | | 1 | | | | | | | | | | | |
| 17 | | | 1 | | | | | | | 1 | | | | | | | | | |
| 18 | 1 | 1 | 1 | | 1 | | | 1 | | 1 | | | | | | | | | |



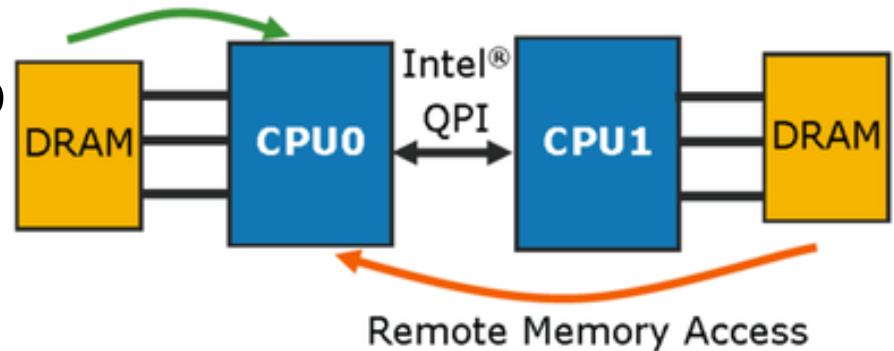
Общие методы программной оптимизации для SMP x86 систем

Архитектура вычислительного узла

- 2 сокета с Intel Xeon E5-2683 v3 2.00 GHz
- Включен режим Turbo до 3 GHz
- кэш 35 МБ
- NUMA
- 14 ядер на сокет
- Hyper-Threading (в сумме 56 потоков)
- QPI 9.6 GT/s
- 64 GB
- Intel Xeon Phi 5110P

SMP узел:

Local Memory Access



Задержка доступа к памяти, тактов

| | NODE |
|--------|------|
| Кэш L1 | 4 |
| Кэш L2 | 15 |
| Кэш L3 | 40 |
| DRAM | 170 |
| NUMA | 400 |