

# Имитационное моделирование емкостной сложности алгоритма поиска изоморфного подграфа в семантической сети

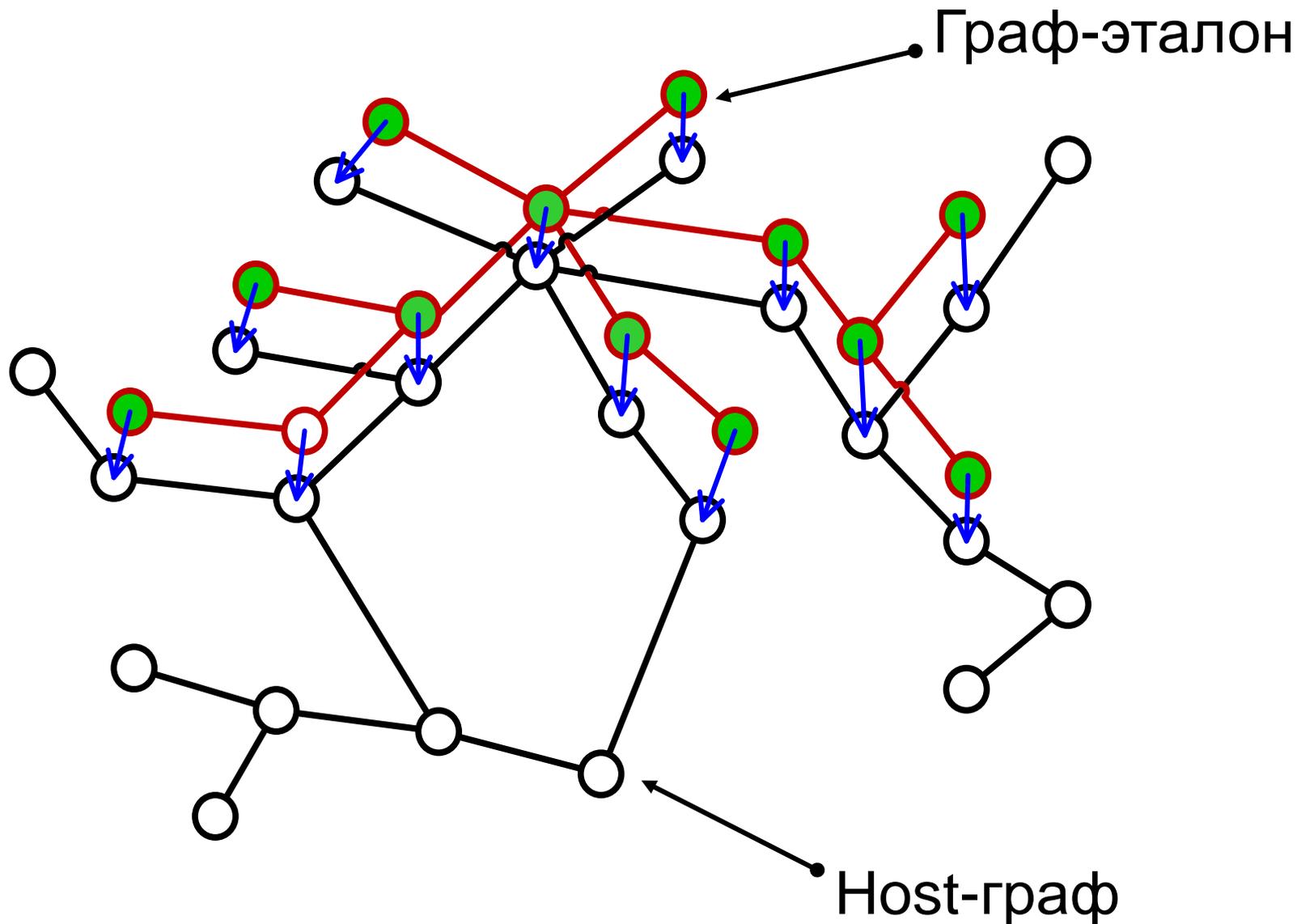
**Салибебян С.М., Жуков А.А., Спирин Г.Г.**

Национального исследовательского университета  
«Высшая школа экономики»,  
Московский институт электроники и математики

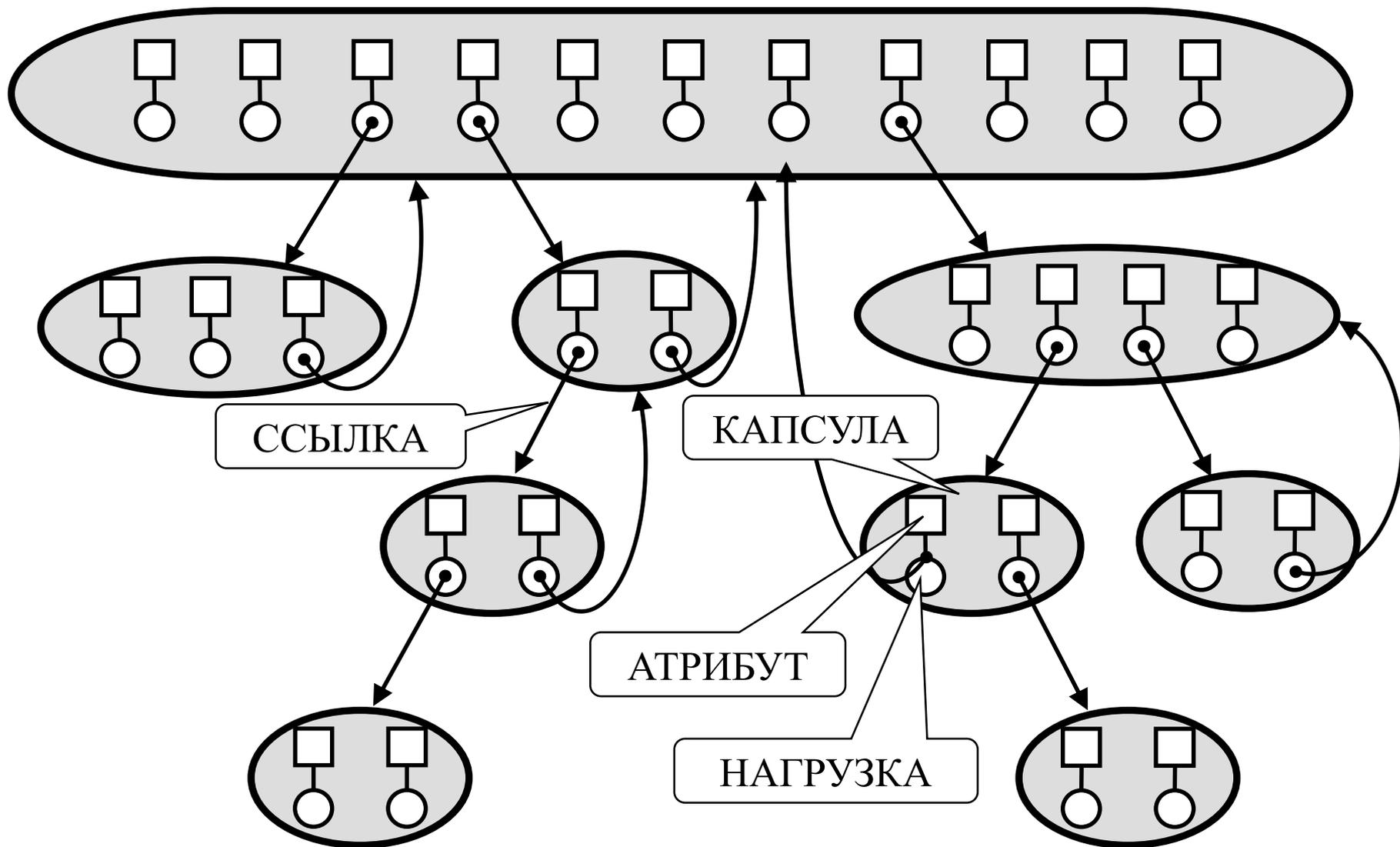


# Проблема поиска изоморфного подграфа (subgraph matching)

2



# Семантическая сеть на базе объектно-атрибутного графа



# Поиска изоморфного подграфа 4 в помеченном графе

Пусть имеется граф  $G=(V_G, E_G)$  (host-граф) и граф-шаблон  $H=(V_H, E_H)$ . Требуется найти подграф  $G'=(V', E')$ , где  $V' \subseteq V_H$  и  $E' \subseteq E_H$  и функцию  $f: V' \rightarrow V_H$ , такие, что  $(f(v_1), f(v_2)) \in E_H$ , где  $v_1, v_2 \in V'$ . (1)

В настоящем исследовании семантическая сеть аппроксимируется помеченным неориентированным графом, поэтому необходимо ввести множество меток  $M$ , а также функцию  $m$ , которая сопоставляет узлы графа с метками  $m: V \rightarrow M$ , где  $V$  – множество узлов графа, а также ввести дополнительное условие в (1) изоморфизма подграфа:  
 $m(f(v_1)) = m(v_1)$ ,  $m(f(v_2)) = m(v_2)$ .

# Области применения поиска изоморфного подграфа

1. Семантический информационный поиск в тексте
2. Математическая химия
3. Оптимизация компьютерных программ
4. Сетевые (графовые) базы данных
5. Системы искусственного интеллекта (базы знаний)

# Постановка задачи

**Объектом исследования** является ёмкостная сложность алгоритма поиска изоморфного подграфа в семантической сети.

**Метод исследования:** аналитическое и имитационное моделирование

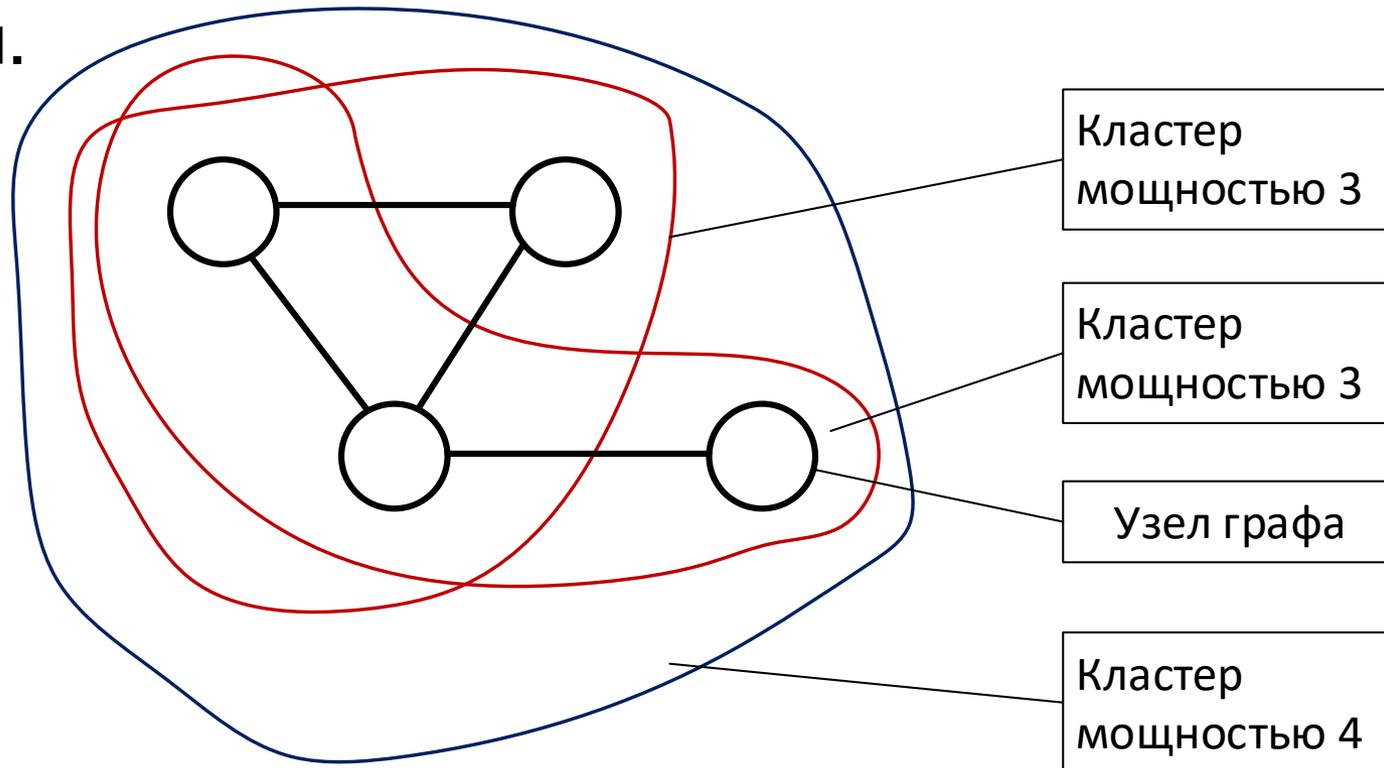
**Основной подход к исследованию:** семантическая сеть аппроксимируется неориентированным помеченным графом, и затем производится моделирование поиска подграфа

**Задачи исследования:**

- Аналитическое исследование ёмкостной сложности алгоритма
- Создание имитационной модели для анализа ёмкостной сложности алгоритма
- Исследование результатов запуска имитационной модели.

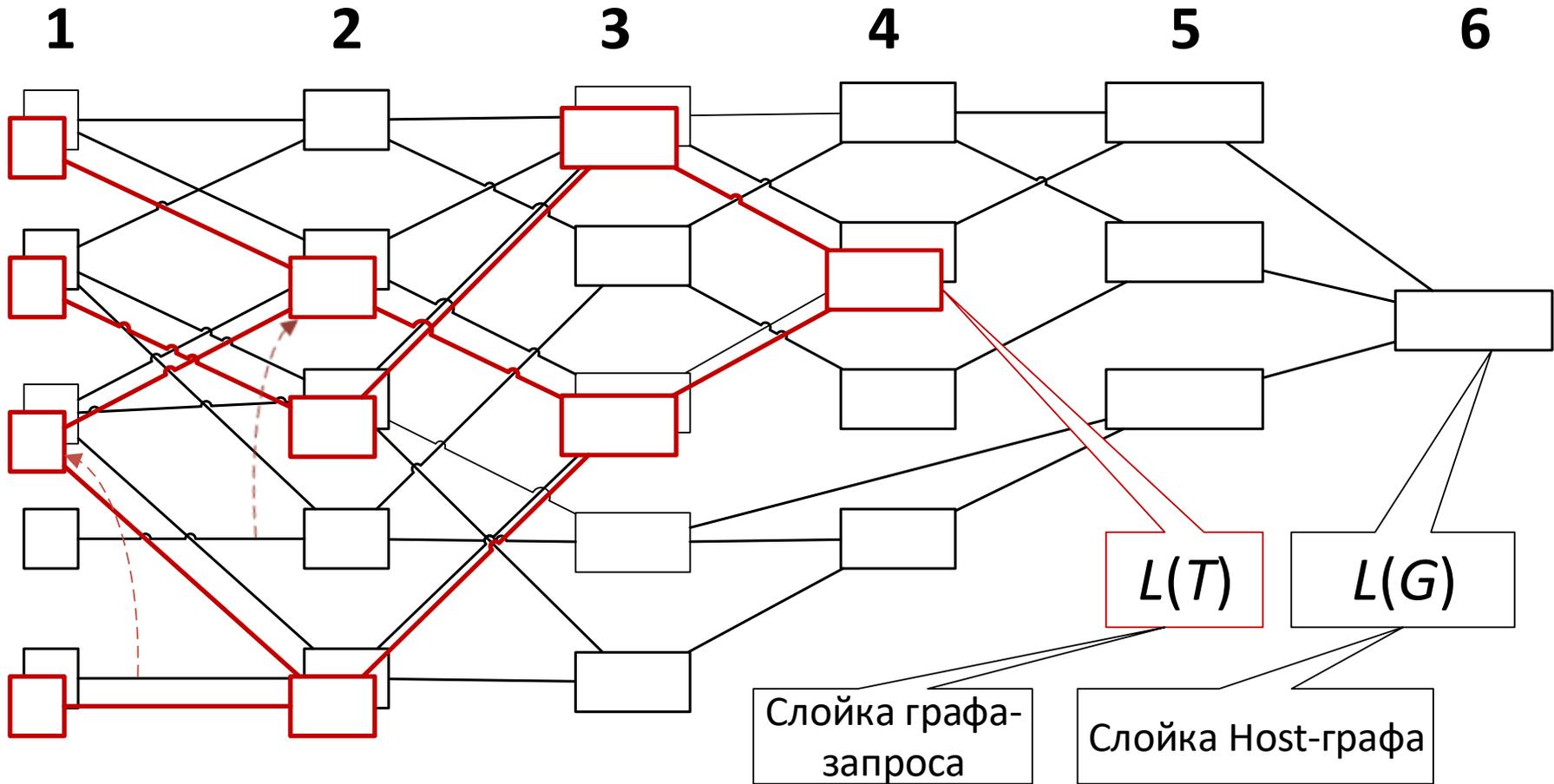
# Алгоритм поиска изоморфного подграфа с помощью графовой слойки

Основная задача алгоритма: повысить скорость поиска изоморфного подграфа за счет предварительного создания информационных конструкций, облегчающих поиск помеченных узлов неориентированного графа за счет механизма хеширования.

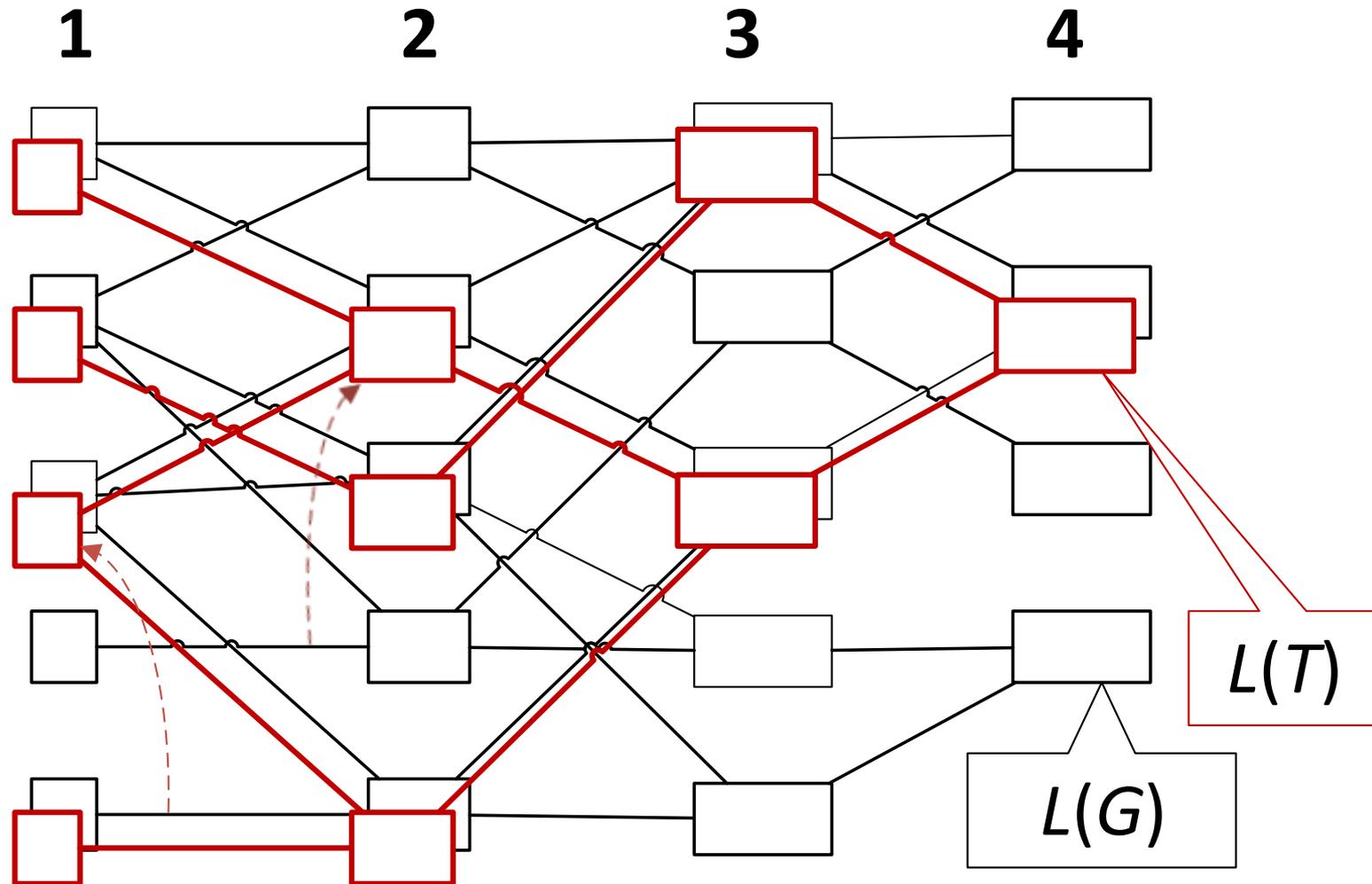




# Алгоритм поиска изоморфного подграфа в помеченном графе



# Алгоритм поиска изоморфного подграфа в помеченном графе



# Модель алгоритма поиска изоморфного подграфа

**Входные параметры:** Количество узлов графа  $N$ , вероятность образования дуги  $p$ , глубина построения слойки  $k$ .

**Выходные данные:** количество узлов и ребер построенной слойки  $L(G)$

# Результаты аналитического моделирования

Количество узлов в слое в зависимости от  $N$  (количество узлов графа) и  $p$  – вероятность образования узла,  $k$  – глубина слоя  $L(G)$ :

$$LV = \frac{1}{p(1-p)} \sum_{i=1}^k \frac{(Np)^i}{i!} \sim N^k$$

Количество ребер в слое  $L(G)$

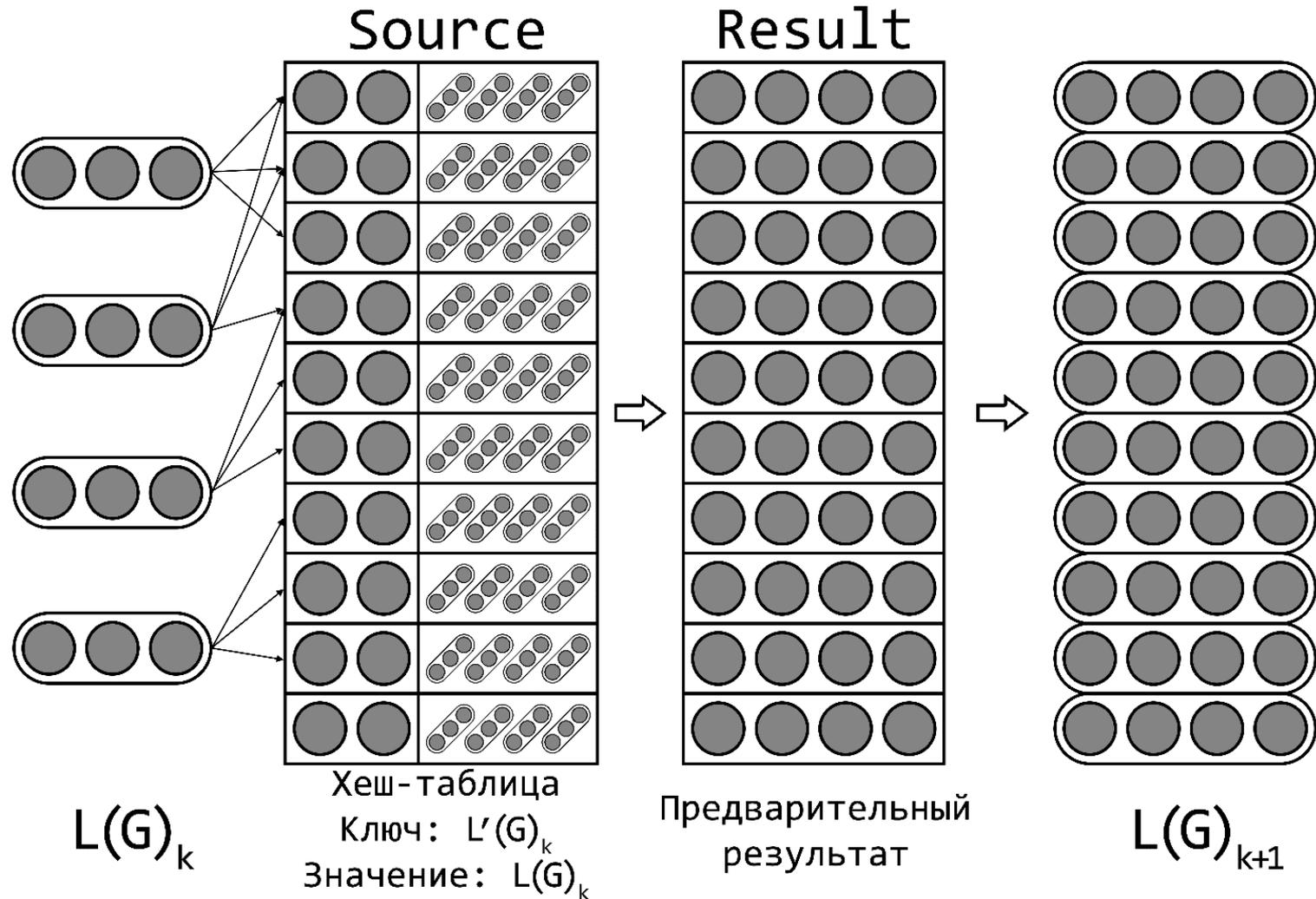
$$LV = \frac{1}{p^2(1-p^2)} \sum_{i=1}^k \frac{(Np^2)^i}{i!} \sim N^k$$

# Программная реализация имитационной модели

## Состав имитационной модели:

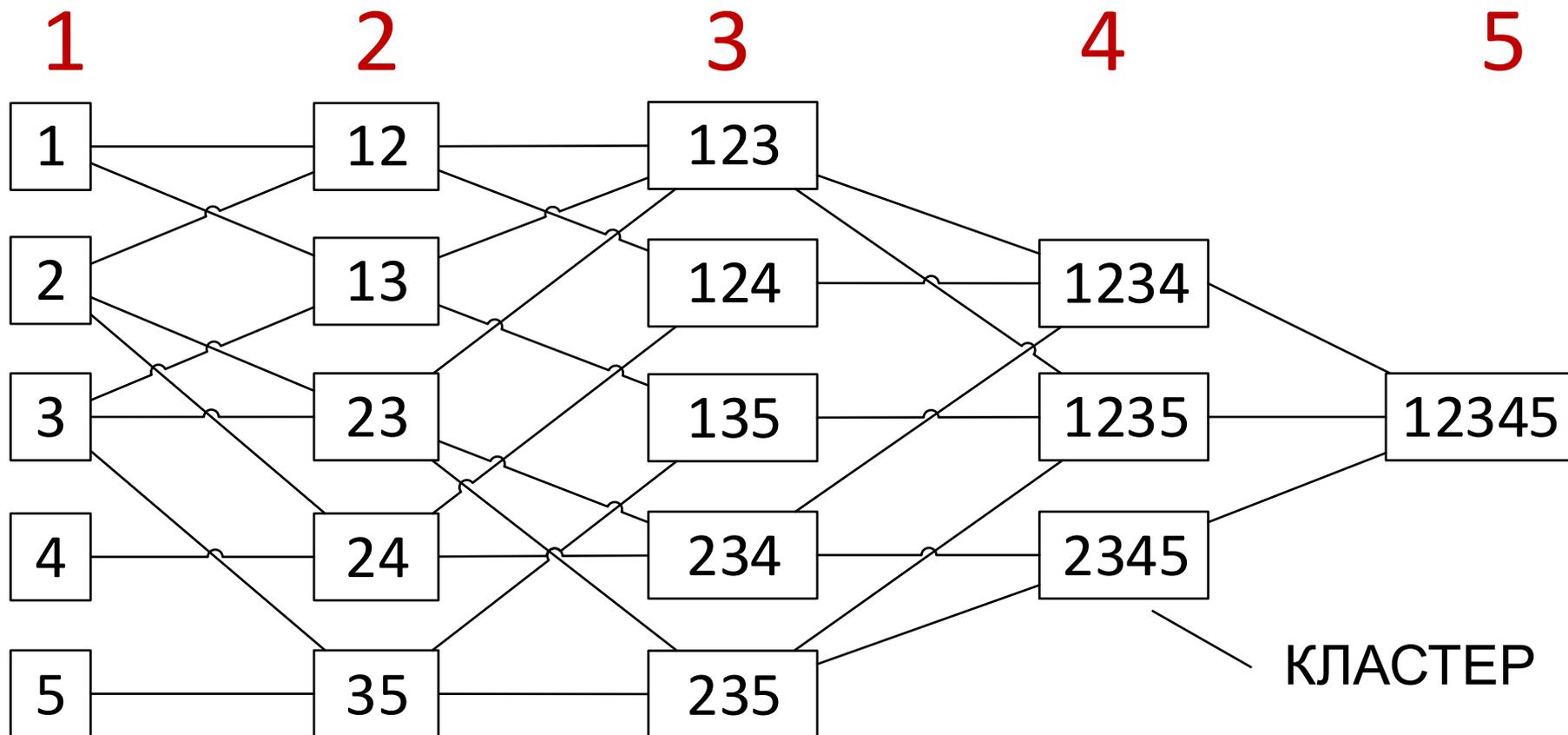
- Блок генерации случайного графа
- Блок построения слойки
- Блок анализа результатов моделирования

# Механизм хеширования для ускорения построения слойки



# Механизм хеширования для ускорения построения слойки

15



# Результаты имитационного моделирования

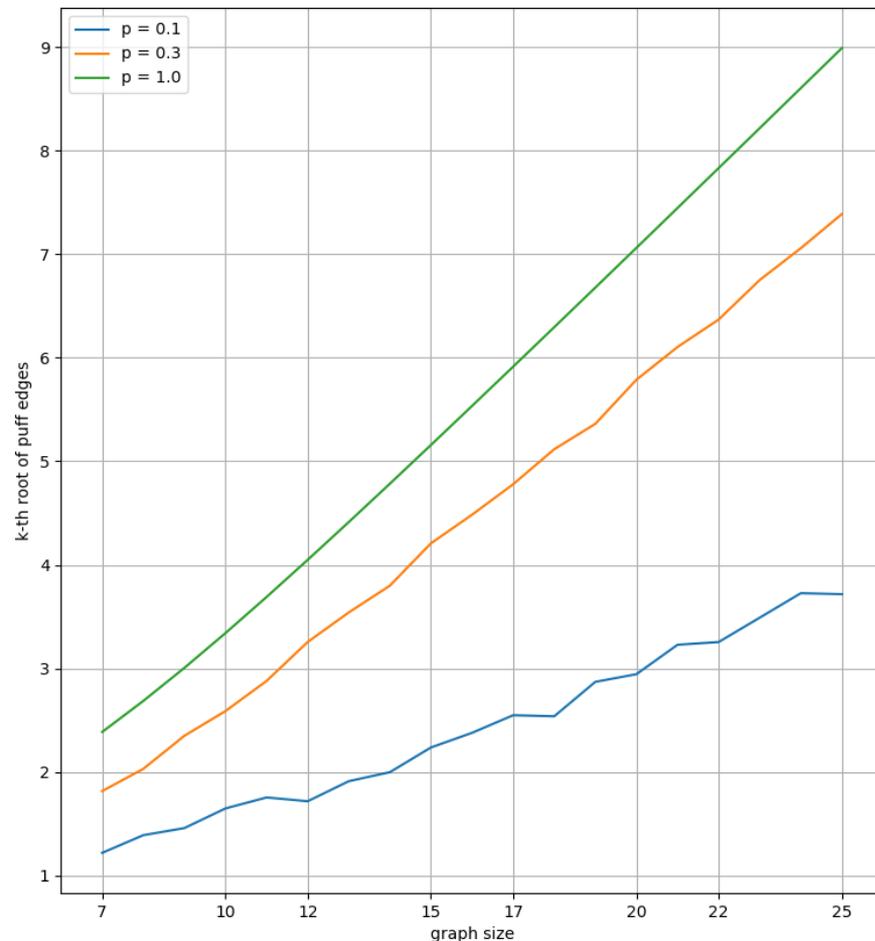
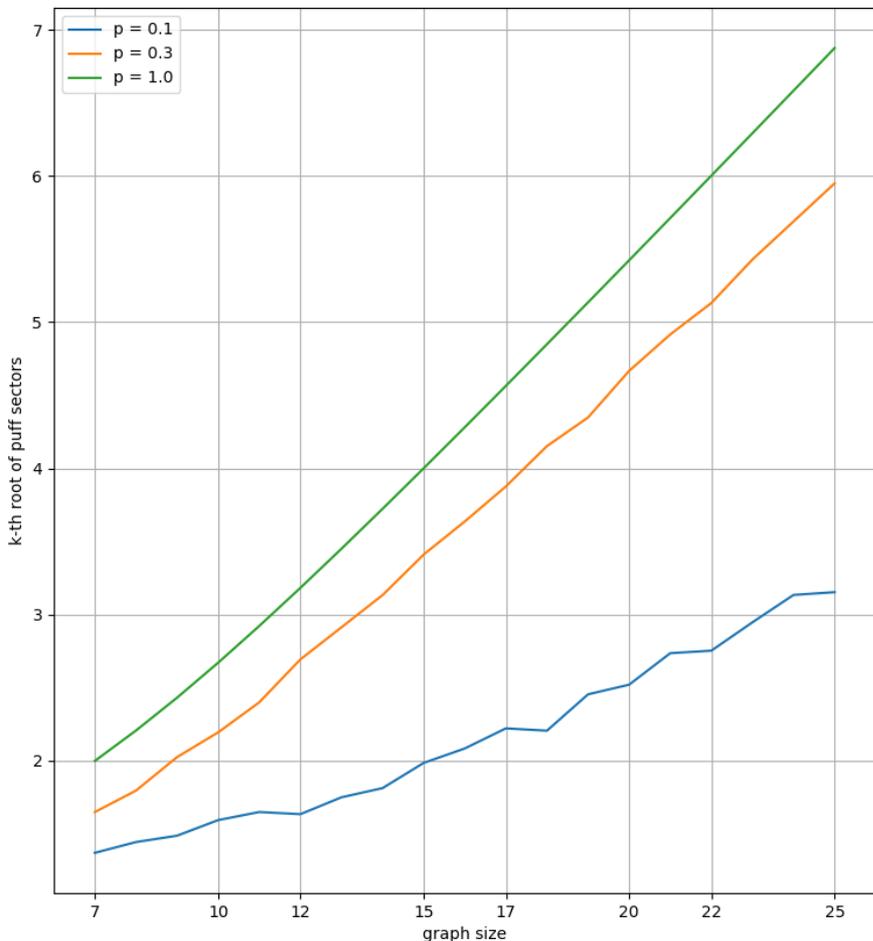


График узловой и рёберной мощностей слойки от  $N$  (по оси  $Y$  отложена величина  $nS$ ) при  $k=7$ ,  $p=0,1; 0,3; 1$

# Результаты имитационного моделирования

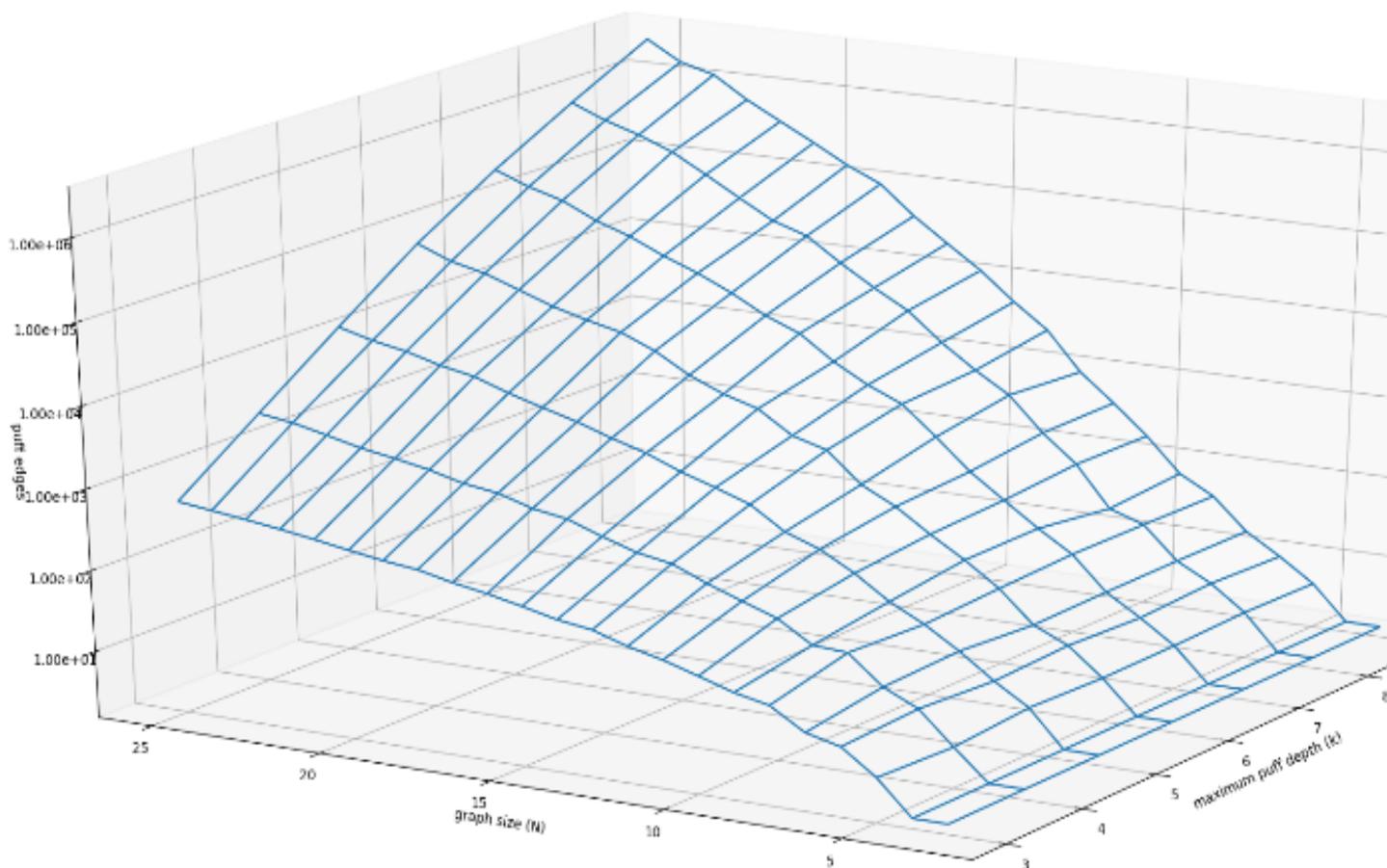


График узловой мощности слойки при  $p=0.2$  в экспоненциальном масштабе

# Результаты имитационного моделирования

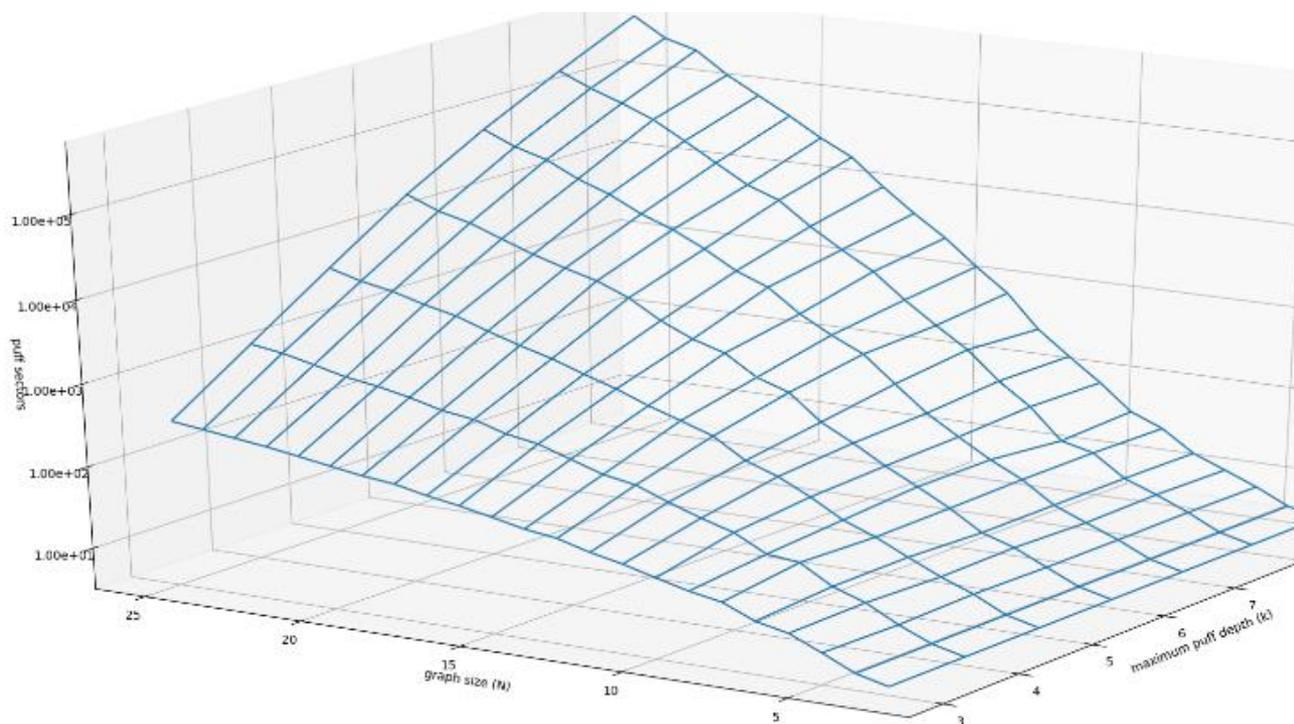


График рёберной мощности слойки при  $p=0.2$  в экспоненциальном масштабе

1. С помощью аналитического и имитационного моделирования было показано, что при  $k \ll N$ , где  $n$ - количество узлов графа-шаблона,  $N$  – количество узлов host-графа, емкостная сложность алгоритма пропорциональна  $N^k$ . Алгоритм такой сложности реализуем на ЭВМ.
2. В ходе исследования были отработаны механизм хеширования для ускорения поиска узлов во время синтеза графовой слойки.
3. Результаты исследования найдут применение в областях: семантический анализ текста, семантический поиск в тексте, сетевые базы данных.

**Спасибо за  
внимание!**