

Алгоритмы поиска подграфов типа «черная дыра» в ориентированном графе

Иванов Д.Е., Семенов А.С.

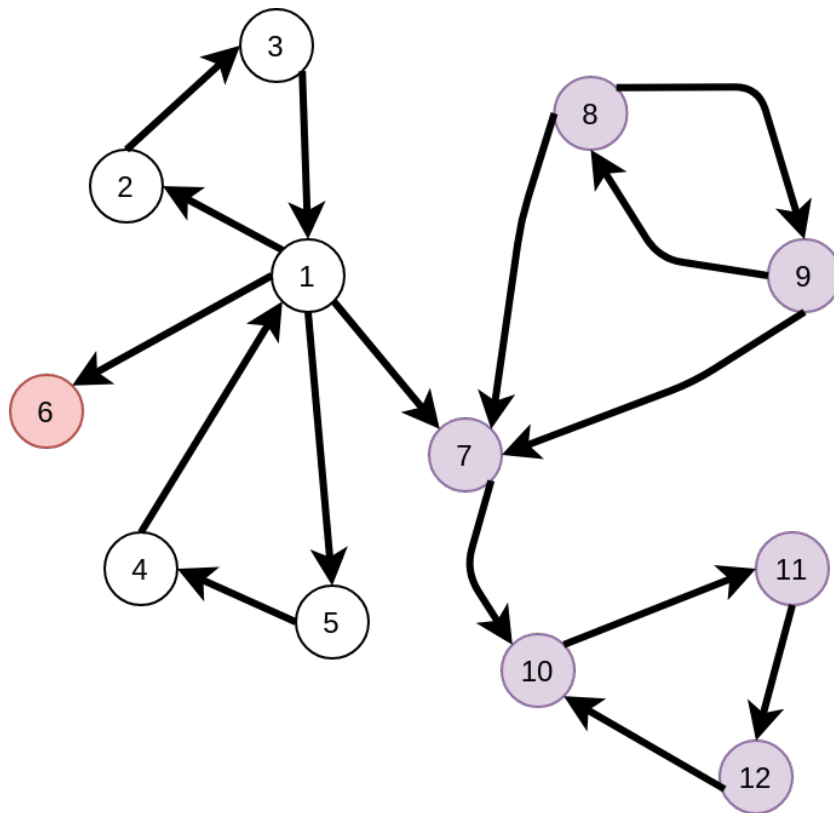
Москва 2020

Определение

Дан ориентированный граф без весов $G(V,E)$

«Черная дыра» – подмножество вершин графа:

- Слабо связанное
- Без исходящих ребер



Приложения

- Поиск мошеннических схем на рынке [Li et al. 2010]
- Предсказание неблагоприятных ситуаций в городской среде [Hong et al. 2015]
- Поиск схем отмывания денег [Semenov et al. 2017]

Постановка задачи

В ориентированном графе за ограниченное время найти как можно больше черных дыр

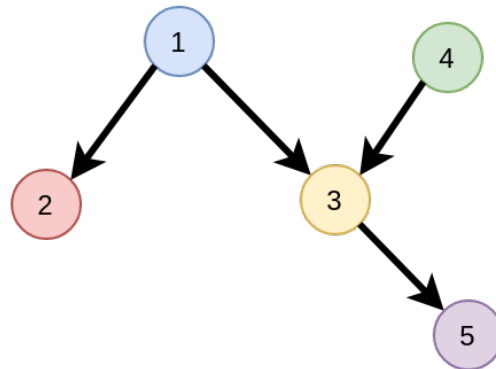
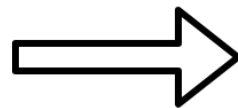
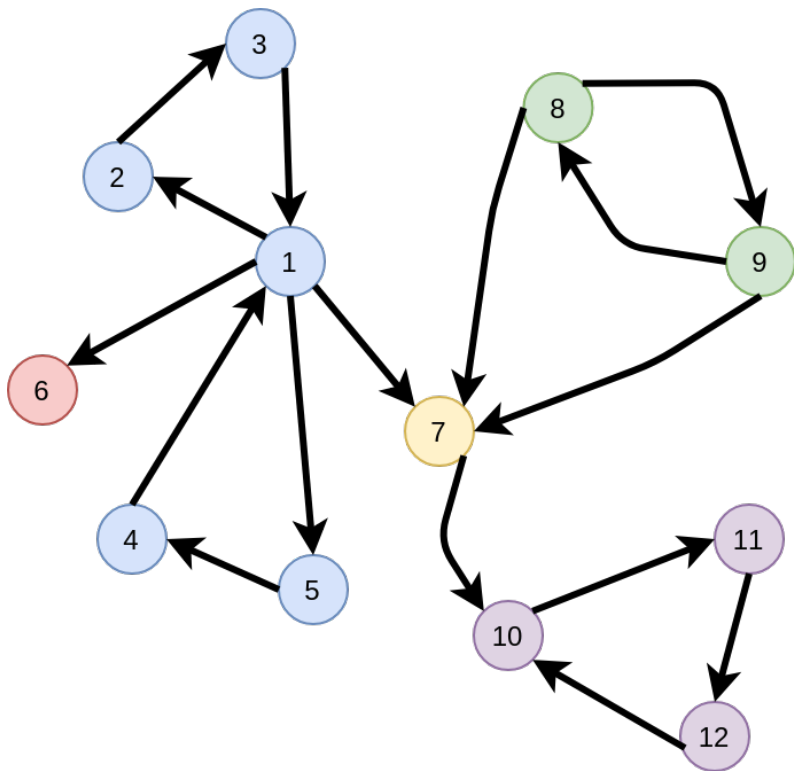
Для этого потребуются:

- Изучить существующие решения
- Разработать новый конкурентоспособный алгоритм для поиска черных дыр
- Разработать параллельные реализации представленных алгоритмов
- Провести сравнительное тестирование разработанных алгоритмов на графах сравнительно больших размеров

Существующие проблемы

- Алгоритм iBlackholeDC [Li et al. 2010]
 - Плохо справляется со SmallWorld графами (например, граф, состоящий из одной большой SCC)
 - В конце каждой итерации происходит перебор вариантов
 - В исследовании рассмотрены лишь графы с количеством вершин не более 1500, а также черные дыры небольшого размера

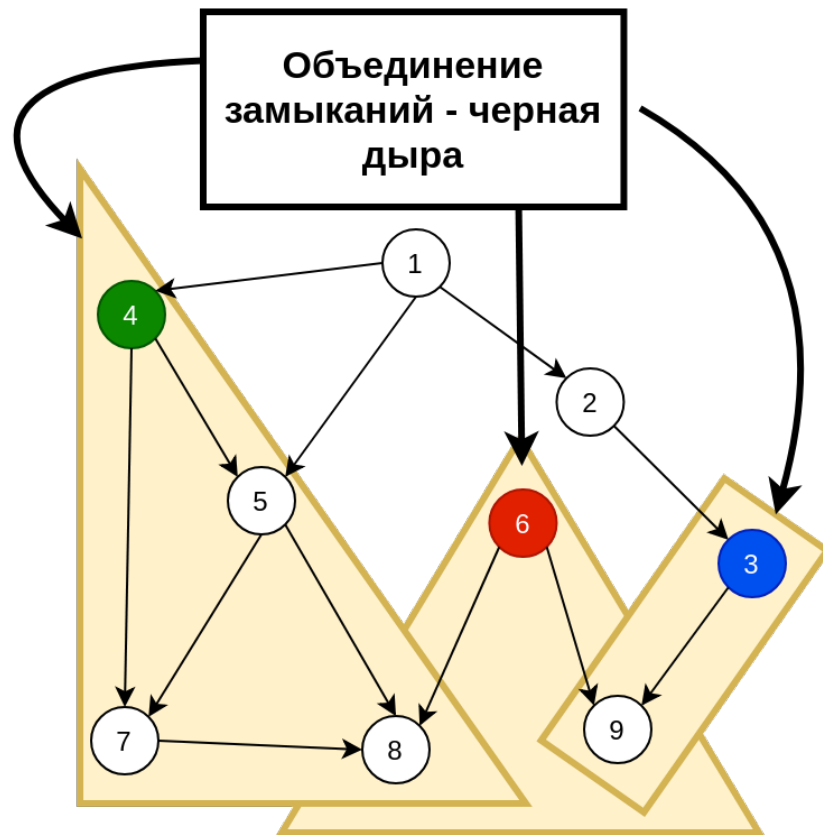
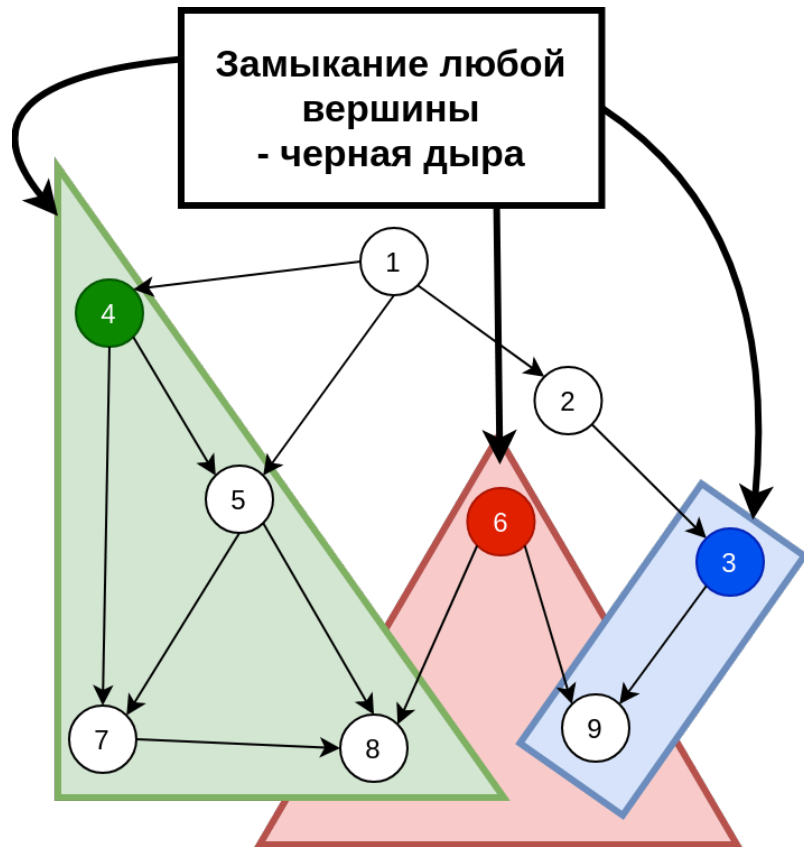
Предложение 1: Конденсация графа



- Ориентированный
- Без циклов
- Меньше вершин

Для поиска компонент сильной связности применяется алгоритм Шарира [Sharir, 1981].

Перебор: Анатомия черной дыры



Предложение 2: Организация полного перебора

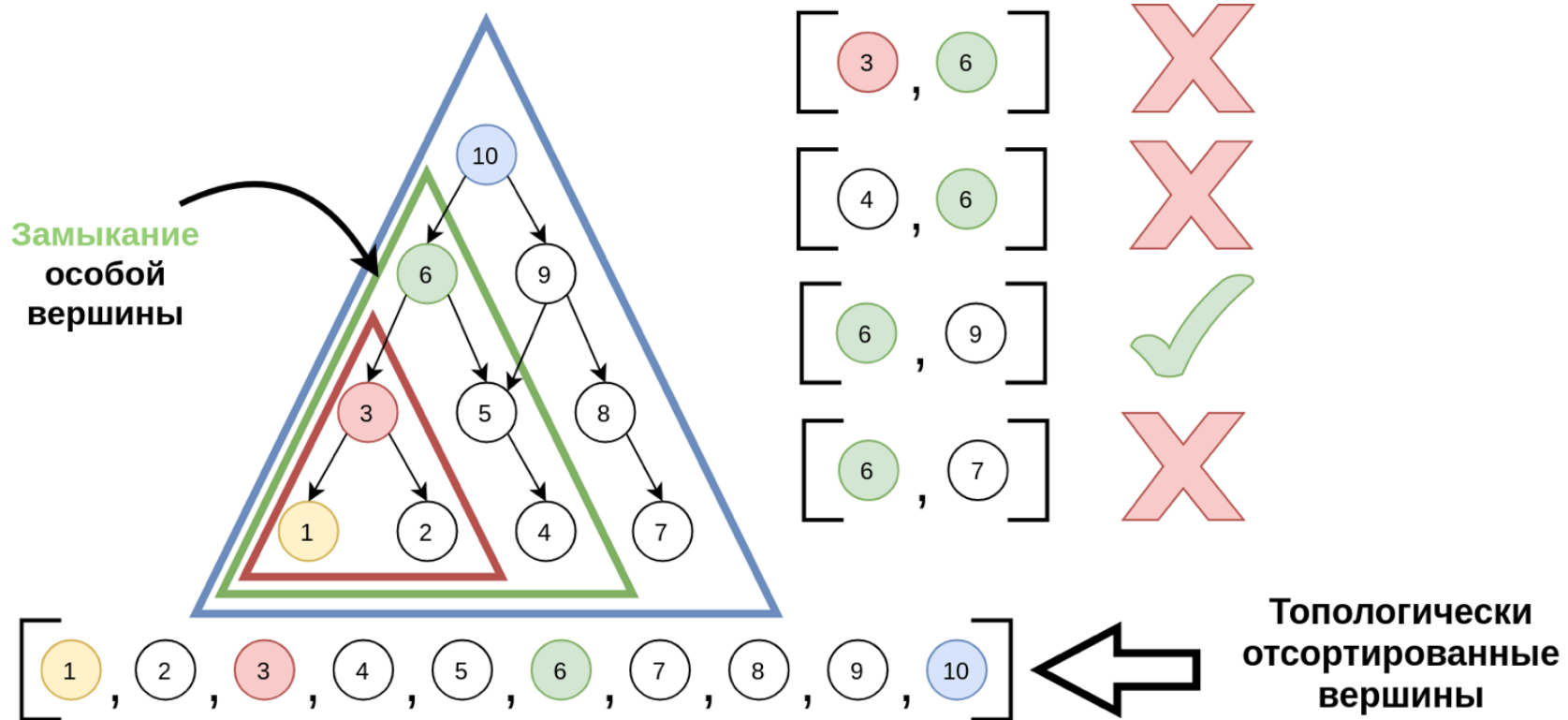
1. Перебираются множества вершин размера 1, затем размера 2 и т.д.
2. Объединение замыканий вершин множества является черной дырой при наличии слабой связности.

Ресурс для оптимизации:

- Много повторов черных дыр
- Дорогая проверка на связность

Предложение 3: Сокращение перебора

Особая вершина – если размер её замыкания равен индексу в топологическом массиве



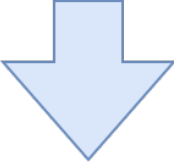
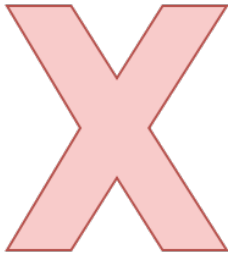
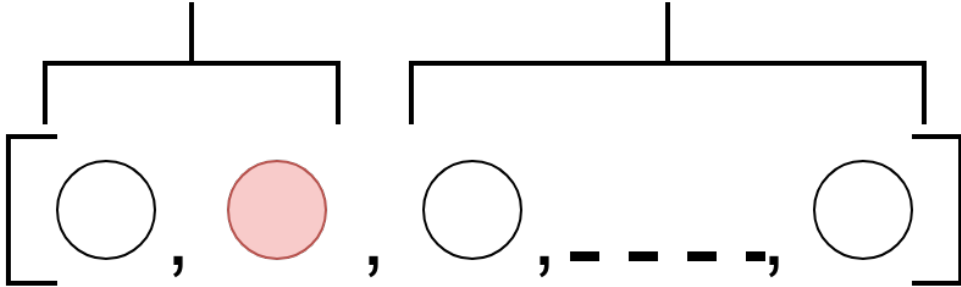
Алгоритм TopSort_BH

1. Конденсация графа
2. Перебор черных дыр
3. Сокращение перебора на основе топологии

Модификация SkipFast

Некорректный префикс

Произвольный суффикс



Ресурс параллелизма

В графе:

- слабо связанные области можно обрабатывать порознь

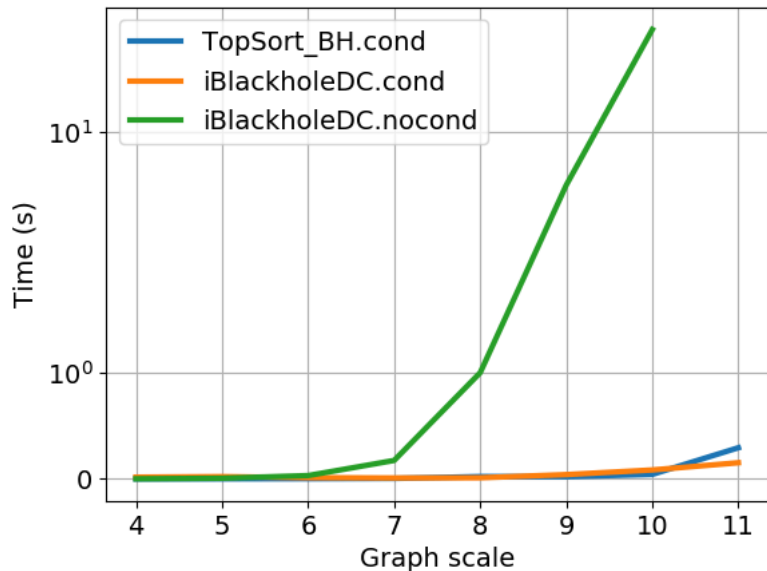
В алгоритме:

- Принимать решение по каждому из кандидатов можно независимо от других

Основные сравнительные результаты

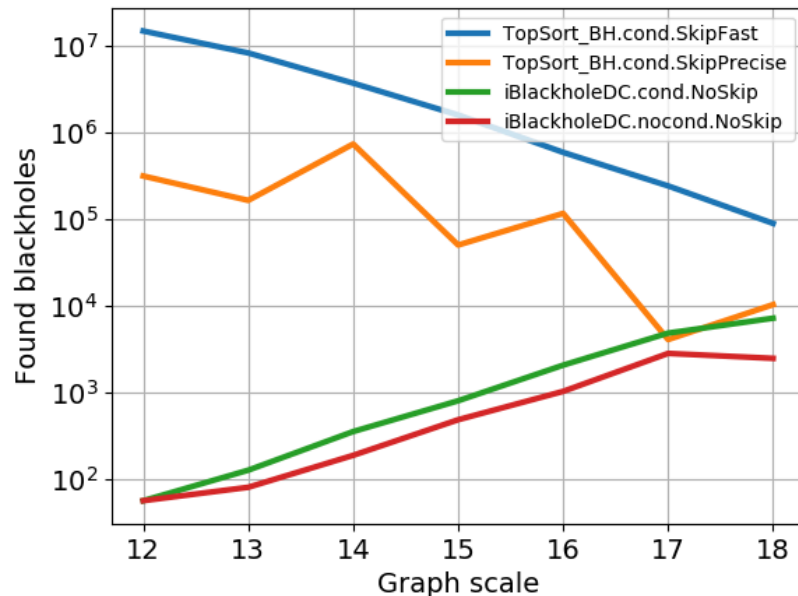
Алгоритм отработал полностью

RMAT



Отработал 30 минут

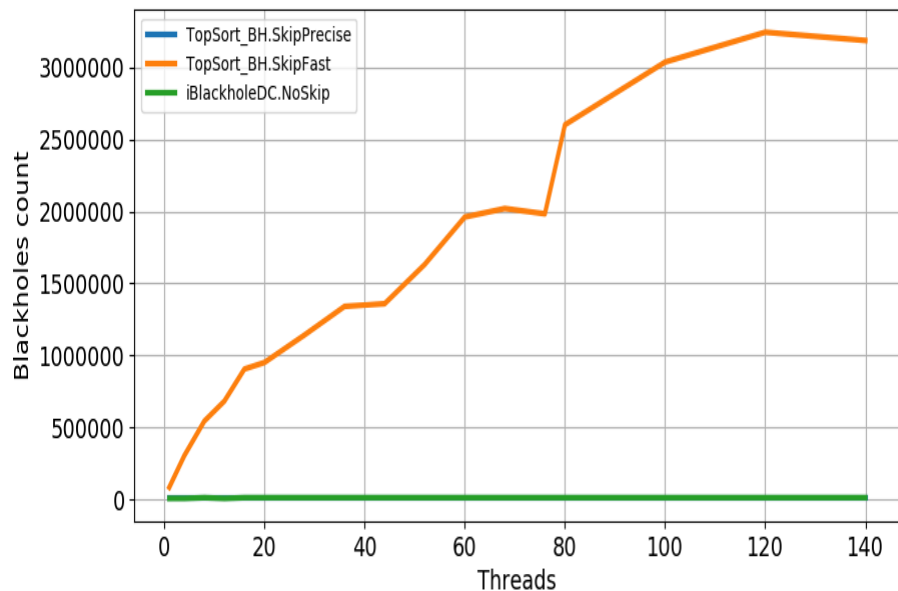
RMAT



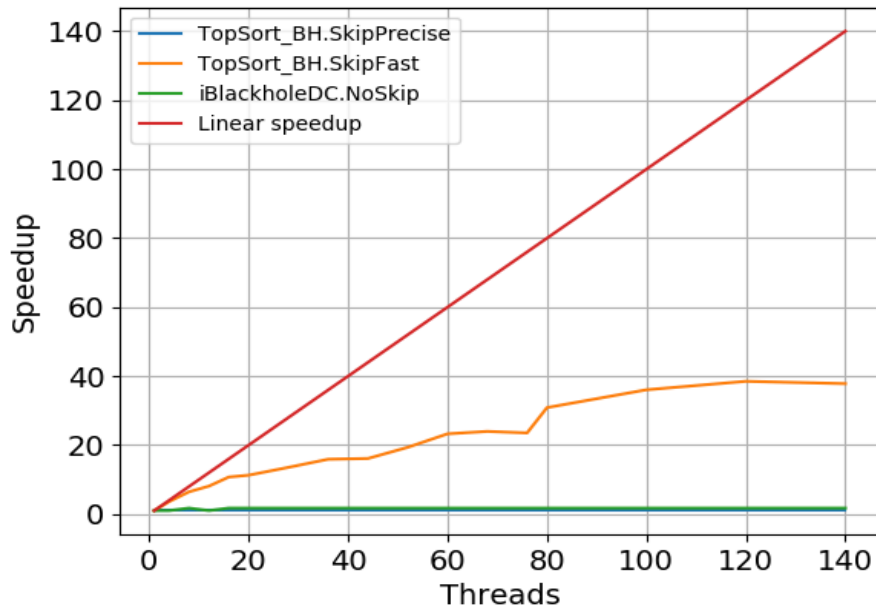
Один узел системы Polus (2x IBM Power8, 40 ядер, 160 потоков)

Основные сравнительные результаты

RMAT



RMAT



Один узел системы Polus (2x IBM Power8, 40 ядер, 160 потоков)

Полученные результаты

- Разработан алгоритм TopSortBH, включающий: конденсацию, новый подход к перебору, правила сокращения перебора
- Предложена эффективная эвристика SkipFast
- Разработаны многопоточные реализации алгоритмов TopSortBH и iBlackholeDC с использованием технологии OpenMP
- Проведено исследование производительности на узле системы Polus
 - Использовано до 140 потоков на одном узле
 - Рассмотрены графы до 2^{18} вершин (по сравнению с 1500 вершин в предыдущих работах)
 - Предложенный алгоритм находит значительно большее количество черных дыр за одинаковое время.

Спасибо за внимание!