Редукция данных с гарантиями эффективности

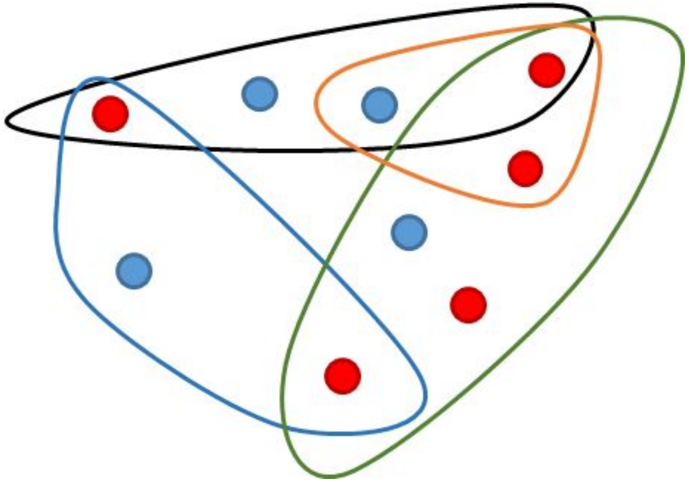
Первый воркшоп МЦА

Афанасьев В.А. ван Беверн Р.А. Кирилин А.М. Начын А.О.
Скачков Д.А. Смирнов П.В. Цидулко О.Ю. Эмдин Г.Д.

f-кратное вершинное покрытие гиперграфа

Дано: гиперграф $H=(V,E)$ с **вершинами V** и **рёбрами $E \subseteq 2^V$**,
и **спросы** $f:E \to \{1,...,\alpha\}$.

Найти: наименьшее подмножество вершин $S \subseteq V$ такое, что $\forall e \in E$: $|e \cap S| \geq f(e)$.



Задача **NP-трудна!**

Надо **сокращать** объём **данных**.

Результат 1: Новые правила редукции данных

(НР) Над**р**ёбра: если $\exists\, e, g \in E$, т. ч. $g \subseteq e$ и $f(g) \geq f(e)$, то удалить e.

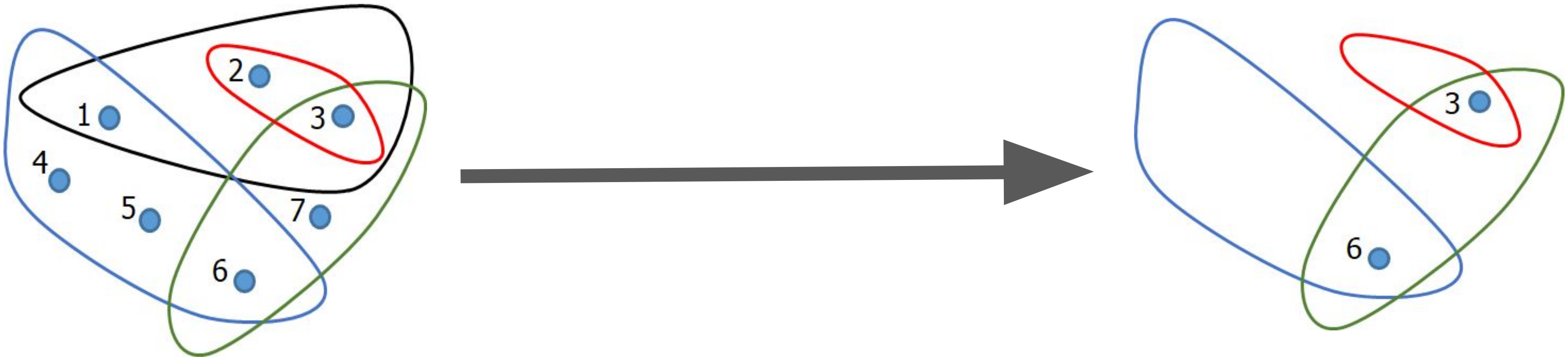
(ПР) Полные **р**ёбра: если $\exists\, e \in E$, т.ч. $f(e)=|e|$, то для всех $g \in E$ уменьшаем $f(g)$ на $|g \cap e|$ и все вершины $v \in e$ добавляем в решение.

(Д) Доминирование**:** если $f \equiv 1$ и $\exists\, u,v \in V$, т. ч. $\forall\, e \in E : u \in e \rightarrow v \in e$, то удалить v.

Новые правила обобщают старые:

(ПС) Проталкивание **с**проса: если $\exists\, e,g \in E$, т. ч. $f(g) - |g/e| \geq f(e)$, то удалить e.

(КД) Кратное **д**оминирование: удалить $v \in V$, если $|\{u \in V \mid \forall e \in E : v \in e \rightarrow u \in e)| > \max_{e \in E}\{f(e) \mid v \in e\}$.



Результат 2: Снижение трудоёмкости (в том числе известных правил)

Исчерпывающее применение новых правил (КД и ПС)*

на одном процессоре: за время $O(|H|\ (|V|+|E|))$,

на $\Theta((|V|+|E|)^3)$ процессорах: за время $O(M \log |H|)$, где $M \le \min\{|V|, |E|\}$ – размер наибольшего паросочетания в графе инцидентности.

Для распараллеливания алгоритм представлен как последовательность операций над большими матрицами.

---

*Новые правила **КД** и **ПС** включают в себя старые **Д** и **НР** соответственно.

Результат 3: Оценка результативности

После применения алгоритма $|V|+|E| \leq \alpha \cdot \nabla(G)$,

где $\nabla(G)$ = число Дилуорса графа инцидентности гиперграфа.

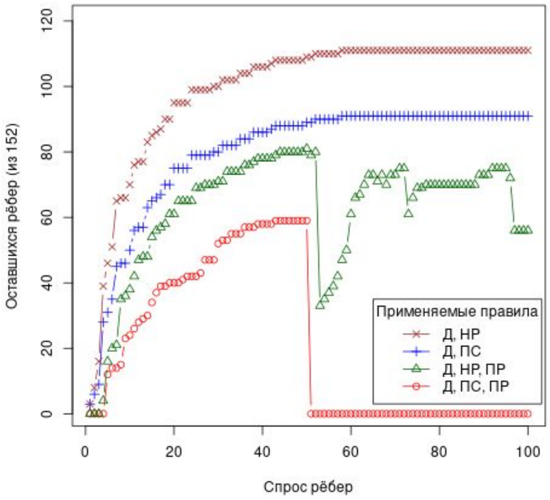
Параметр графов $\nabla(G)$ был введён ещё в 1978г.

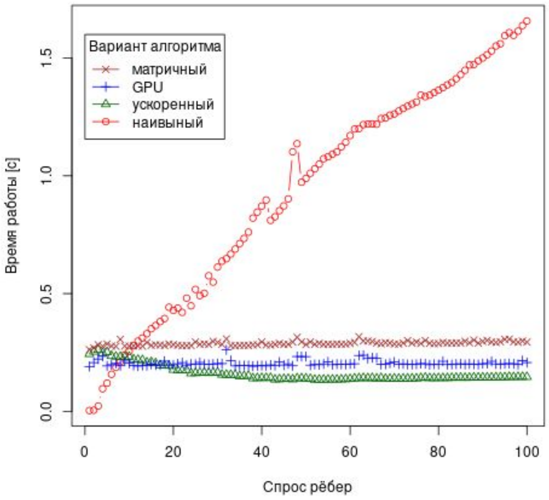
Он **лучше, чем известный** параметр **neighborhood diversity**,

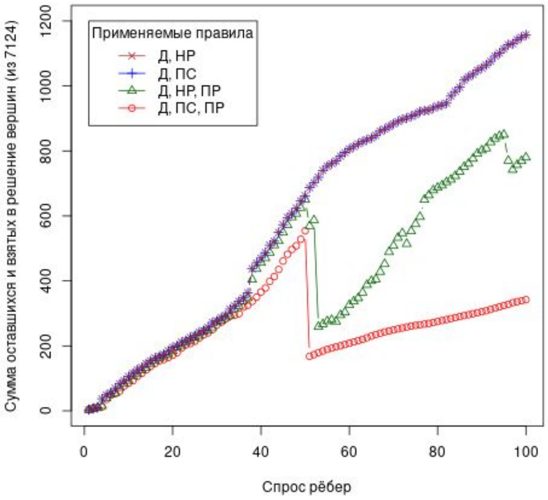
но в области параметризованной теории сложности ещё **не исследовался**.

Результат 5: Вычислительный эксперимент

На примере отбора веществ, противодействующих росту штаммов раковых клеток







160 штаммов раковых клеток, около 60 000 веществ.

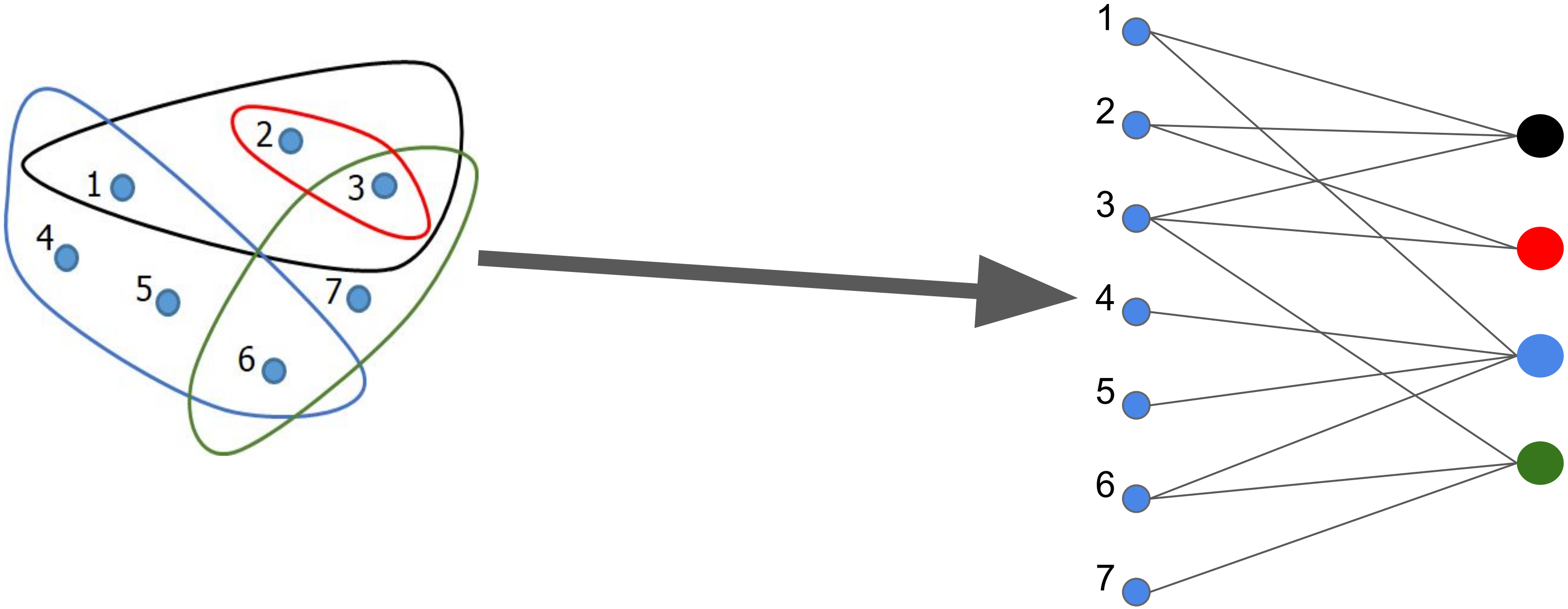
Работа на будущее

Правило ПД улучшает старый аналог только тогда, когда у рёбер разные спросы: нужно разработать правила, повышающие спрос рёбер.

Вариант алгоритма, который работает с внешней памятью.

Закончить статью.

Граф инцидентности гиперграфа



Идея доказательства оценки результативности

Теорема Дилуорса: $\nabla(G)$ = наименьшее число цепей (линейно упорядоченных множеств), покрывающих все вершины.

После применения алгоритма посмотрим граф инцидентности G:

1. Во множестве V останется не более $\nabla(G)$ минимальных элементов, каждый минимальный элемент имеет не более α доминаторов.
2. Множество E можно покрыть $\nabla(G)$ цепями, по каждой цепи f(e) строго возрастает, поэтому длина каждой цепи ограничена сверху α.

→ Число вершин и рёбер в итоговом гиперграфе не превышает $\alpha \cdot \nabla(G)$.