

# СТРУКТУРНЫЙ АНАЛИЗ СИСТЕМ МИНИМАЛЬНЫХ ГРАФОВ СМЕЖНОСТИ

Фильченков А.А., Тулупьев А.Л.

---

УДК 004.8

*Фильченков А.А., Тулупьев А.Л.* Структурный анализ систем минимальных графов смежности.

**Аннотация.** Цель данной работы — анализ структуры минимальных графов смежности и их свойств. Введена система терминов, структурирующая исследуемую область. Исследованы свойства минимальных графов смежности. Доказана структурная теорема о множестве минимальных графов смежности и предложен алгоритм построения такого множества.

**Ключевые слова:** алгебраические байесовские сети, вторичная структура, минимальный граф смежности, автоматическое обучение, структурный синтез.

*Filchenkov A.A., Tulupyev A.L.* Minimal joint graph structure synthesis.

**Annotation.** The goal of this work is the analysis of the structure of minimal joint graph and its properties. A system of terms is given to structure the research space. The minimal joint graph properties are studied. The structure theorem about the minimal joint graph set is proven and an algorithm to build such a set is suggested.

**Keywords:** algebraical Bayesian networks, secondary structure, minimal joint graph, automated learning, machine learning, structure synthesis.

---

**1. Введение и основные обозначения.** Одной из ключевых проблем в развитии различных классов вероятностных графических моделей (в частности, байесовских сетей и марковских сетей) является разработка теоретических основ и алгоритмов автоматического обучения (машинного обучения, machine learning) [9, 11] или, иными словами, структурного и параметрического синтеза таких моделей.

В случае байесовских сетей двумя основными видами автоматического обучения являются

- глобальное обучение (обучение глобальной структуры, синтез глобальной структуры) — поиск оптимальной или «приемлемой» структуры графа-носителя сети [9–11];
- локальное обучение (параметрический синтез, идентификация параметров) — формирование численных оценок вероятностей в тензорах условной или совместной вероятности, приписанных узлам байесовской сети [4–7, 9, 11].

Указанные виды машинного обучения существенно различаются. Подходы к алгоритмизации машинного обучения байесовских сетей доверия (БСД) интенсивно развиваются западными исследователями (см., например, [11]), а соответствующие подходы в теории алгебраических байесовских сетей (АБС) рассматривались в работах [4–7, 9]

Труды СПИИРАН. 2009. Вып. 11. ISBN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)  
SPIIRAS Proceedings. 2009. Issue 11. ISBN 2078-9181 (print), ISSN 2078-9599 (online)

для локального случая и в работах [8, 9] — для глобального случая. Следует отметить заметный вклад в начало исследований алгоритмов автоматического обучения глобальной структуры АБС С.С. Синчука и М.А. Левина, хотя результаты их работы не публиковались.

*Целью* настоящей работы является развитие системы обозначений, понятий и объектов (основа которой заложена в работах [2, 3, 8]), а также изучение и доказательство свойств множества графов смежности с минимальным числом ребер (минимальных графов смежности), построенных над заданным конечным множеством конъюнктов. Полученные результаты должны сформировать теоретическую основу для решения проблемы автоматического построения вторичной структуры АБС в части разработки алгоритмов формирования и перебора минимальных графов смежности.

**Определение 1.1.** *Граф* — пара  $(V, E)$ , где  $V$  — множество вершин графа, а  $E$  — множество ребер, каждое из которых является упорядоченной парой  $(v_i, v_j)$ ,  $i \neq j, v_i, v_j \in V$ .

**Определение 1.2.** Для удобства будем через  $V$  и  $E$  обозначать функции от графа, возвращающие множество его вершин и множество его ребер соответственно:

$$V(G') = V'; E(G') = E', \quad \text{где } G' = (V', E').$$

**Определение 1.3.**  $\subseteq$  — (нестрогое) включение:

$$(A \subseteq B) \Leftrightarrow ((\forall a \in A) a \in B).$$

**Определение 1.4.**  $\subset$  — строгое включение:

$$(A \subset B) \Leftrightarrow ((\forall a \in A, a \in B) \exists b \in B; b \notin A).$$

**Определение 1.5.** *Алфавит* — множество атомарных пропозициональных формул

$$A = \{x_1, \dots, x_n\}.$$

**Определение 1.6.** *Слово*  $V$  — подмножество алфавита.

**Определение 1.7.** *Множество главных конъюнктов максимальных ФЗ, вошедших в байесовскую сеть*, — такое множество слов  $V^* = \{V_i \subseteq A\}_{i=1}^m$ , для которого справедливо следующее:

1) оно не содержит несобственное подмножество алфавита:

$$V_i \neq A, V_i \neq \emptyset;$$

2) никакое слово полностью не содержит никакое другое слово:

$$\forall i \neq j (V_i \not\subset V_j) \text{ и } (V_j \not\subset V_i).$$

**Определение 1.8.** Характеристический вектор слова  $U \in A = \{x_1, \dots, x_n\}$  — бинарная цепочка  $\chi(U) = e_1 \dots e_n$ , где  $e_i \in \{0, 1\}$ . Если

атомарная формула  $x_i$  входит в слово, то в соответствующей ей цепочке  $e_i = 1$ , иначе  $e_i = 0$ .

**Пояснение:** мы будем использовать характеристические векторы для обозначения слов. Так, например, слово  $x_2x_4x_5x_6$  алфавита  $\{x_1, \dots, x_8\}$  будет обозначаться 01011100.

**Определение 1.9.** *Граф максимальных фрагментов знаний* — не-направленный граф, вершины которого соответствуют элементам множества главных конъюнктов максимальных ФЗ, вошедших в алгебраическую байесовскую сеть.

**Пояснение:** с этого момента под словом «граф» мы будем понимать именно граф максимальных фрагментов знаний, если не оговорено обратное. Каждая вершина графа является максимальным фрагментом знаний, ребро между двумя вершинами возможно только тогда, когда пересечение слов, которым соответствуют эти вершины, непусто.

**Определение 1.10.** *Вес  $W(V_i)$  вершины  $V_i$*  — множество атомов алфавита, вошедших в  $V_i$ .

**Замечание 1.1.** Фактически,  $W(V_i)$  в точности совпадает с  $V_i$ , Понятие веса введено для того, чтобы отличать вершину от слова, которому она соответствует. В дальнейшем термины «вес вершины» и «слово» будут абсолютно равноправны и взаимозаменяемы.

**Определение 1.11.** *Вес ребра  $W(\{V_i, V_j\})$ ,  $\{V_i, V_j\} \in E$  графа  $G$*  определяется как пересечение весов тех вершин, которые соединены этим ребром  $W(\{V_i, V_j\}) = W(V_i) \cap W(V_j)$ .

**Определение 1.12.** *Магистральный путь  $M(V_b, V_e)$  от вершины  $V_b$  до вершины  $V_e$* , пересечение весов которых непусто, — это такой путь от вершины  $V_b$  до вершины  $V_e$ , что вес любой принадлежащей ему вершины содержит пересечение весов начальной и конечной вершин (рис. 1):

$$\forall V_i \in M(V_b, V_e) \quad W(V_b) \cap W(V_e) \subset W(V_i).$$

**Определение 1.13.** Граф *магистрально связан*, если между каждой парой несовпадающих вершин  $V_i$  и  $V_j$ , веса которых содержат общие элементы  $W(V_i) \cap W(V_j) \neq \emptyset$ , существует магистральный путь.

**Замечание 1.2.** Если граф магистрально связан, то он связан.

**Определение 1.14.** *Граф смежности* — это магистрально связанный граф МФЗ.

**Определение 1.15.** *Дерево смежности* — это граф смежности, являющийся деревом.



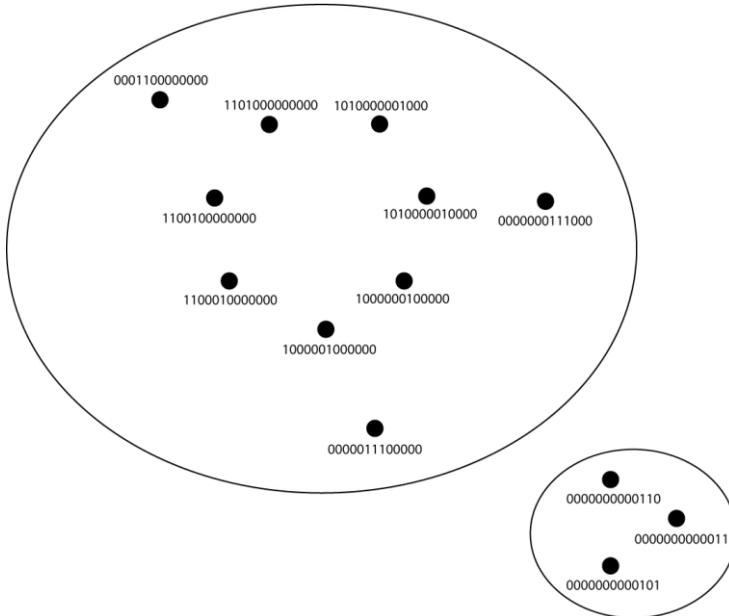


Рис. 2. Набор вершин, ни один граф МФЗ над которыми не будет связан. Так как пересечение весов любых двух вершин из разных выделенных множеств пусто, а, значит, между ними не может быть ребра, то выделенные множества будут несвязны в любом графе, построенном на этих вершинах. В дальнейших примерах будем рассматривать левый набор вершин.

**Утверждение 2.1.** Для заданного множества вершин существует и при этом единственный максимальный граф смежности.

**Доказательство.** Проведем ребра между всеми вершинами, пересечение весов которых непусто (рис. 3). Получившийся граф будет графом смежности, поскольку между любыми двумя вершинами с непустым пересечением весов существует магистральный путь, проходящий по ребру, эти две вершины соединяющему. Так как в графе МФЗ возможны только ребра, соединяющие вершины, пересечение весов которых непусто, и все такие ребра входят в построенный граф, значит, больше ребер добавить нельзя, поэтому построенный граф будет являться максимальным графом смежности —  $G_{\max}$ .

**Определение 2.2.** Сужение  $G \downarrow U$  ненаправленного графа  $G$  на слово  $U$  — это ненаправленный граф, в который входят только те вершины и ребра исходного графа  $G$ , веса которых содержат или равны  $U$  (рис. 4):  $G \downarrow U = \{V_i | V_i \in V^*, U \subseteq W(V_i)\}$ .

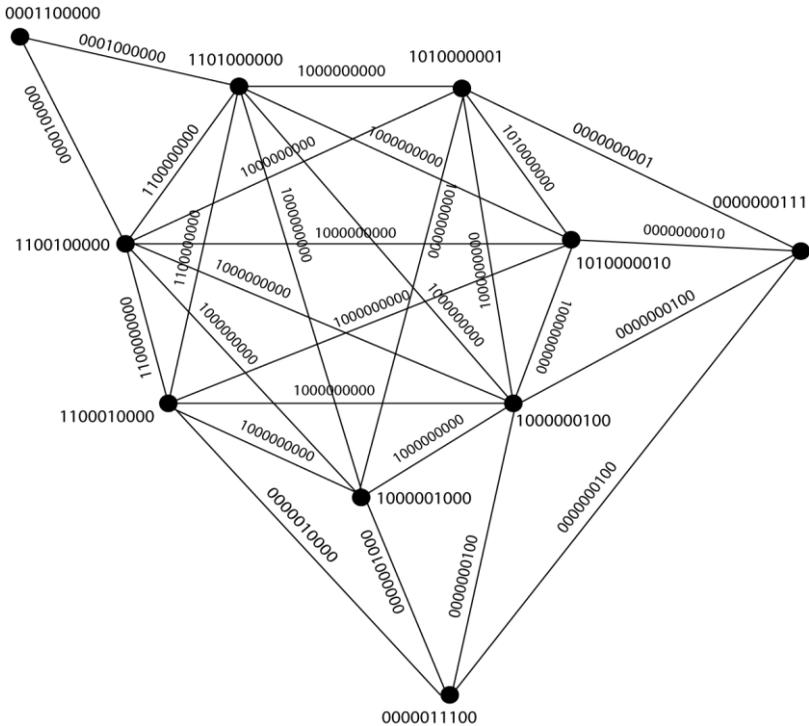


Рис. 3. Максимальный граф смежности.

**Определение 2.3.** *Клика* — полный подграф.

**Утверждение 2.2.** Если сужение  $G_{\max}$  на произвольное слово не пусто, то оно является кликой.

**Доказательство.** Так как все вершины  $G_{\max} \downarrow U$  содержат вес  $U$ , то любые две из них соединены ребром, содержащим вес  $U$ , и это ребро тоже попадет в эту клику.

**Определение 2.4.** *Вес клики* — пересечение весов всех вершин, попавших в эту клику.

**Замечание 2.2.** Вес клики есть наибольшее по включению слово, которое входит в веса всех вершин клики.

**Замечание 2.3.** Вес клики не всегда совпадает со словом, ее породившим.

**Пояснение:** Пускай наше множество вершин состоит из вершин 1101, 1110 и 0011. Вес 1000 порождает клику из вершин 1101 и 1110, но при этом вес данной клики будет 1100.

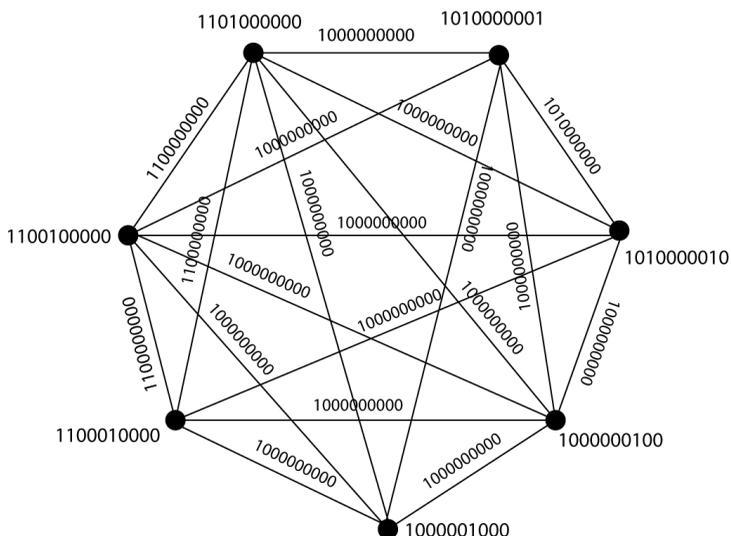


Рис. 4. Сужение.  
Сужение максимального графа на вес 1000000000.

**Определение 2.5.** Clique — множество всех клик графа  $G_{\max}$ , которое:

- не содержит пустых подклик;
- не содержит одноэлементных подклик;
- не содержит одинаковых подклик;
- не содержит подклик с пустым весом;
- содержит только те подклики, вес которых совпадает с весом какого-нибудь ребра графа смежности.

**Определение 2.6.** *Граф клик* — направленный граф, вершинами которого являются клики из множества Clique (рис. 5). Ребро из вершины  $P$  в вершину  $Q$  проведено, если клика из вершины  $P$  содержит клику из вершины  $Q$ .

То же самое, но более формально:

**Определение 2.6'.** Граф клик  $G_{\text{Clique}}$  — пара  $(\text{Clique}, E_{\text{Clique}})$ , где  $\forall P, Q \in \text{Clique}, (P, Q) \in E \Leftrightarrow Q \subset P$ .

**Определение 2.7.** *Направленный цикл* в направленном графе — путь, начинающийся и кончающийся в одной и той же вершине.

То же самое, но более формально:

**Определение 2.7'.** *Направленный цикл* в направленном графе  $G$  — это упорядоченный набор вершин  $C_1, \dots, C_n \in V(G)$ :  $(C_1, C_2) \in E(G), \dots, (C_{n-1}, C_n) \in E(G), (C_n, C_1) \in E(G)$ .

**Утверждение 2.3.** Граф клик не содержит направленных циклов:

$\nexists C_1, \dots, C_n \in G_{\text{Clique}}: (C_1, C_2) \in E_{\text{Clique}}, \dots, (C_{n-1}, C_n) \in E_{\text{Clique}}, (C_n, C_1) \in E_{\text{Clique}}$ .

**Доказательство:** Подобный граф не содержит циклов, поскольку в любом направленном пути этого графа уменьшается число вершин графа смежности в вершине-подклике.

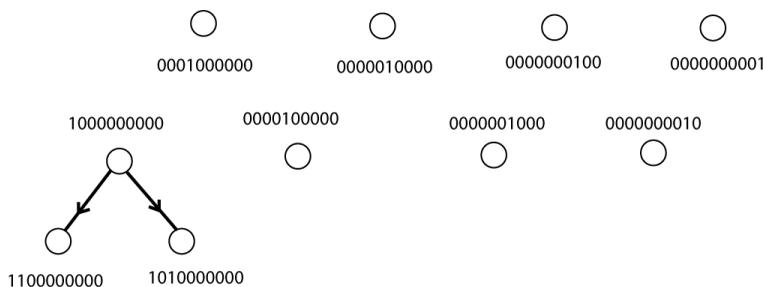


Рис. 5. Граф клик.

**Определение 2.8.** *Упорядоченный граф клик* — граф клик, вершины которого упорядочены по убыванию числа атомов, входящих в их вес (рис. 6).

То же самое, но более формально:

**Определение 2.8'.** *Упорядоченный граф клик*  $G_{\text{Clique}}^{N_o}$  — тройка  $(\text{Clique}, N_o, E_{\text{Clique}})$ , где  $N_o$  — порядок на вершинах  $\text{Clique}$ ,  $N_o: \text{Clique} \rightarrow \mathbb{N}$ , такой, что:

- 1)  $\forall C_1, C_2 \in \text{Clique}, C_1 \neq C_2 \Leftrightarrow N_o(C_1) \neq N_o(C_2)$ ;
  - 2)  $\forall C_1, C_2 \in \text{Clique}, A(W(C_1)) > A(W(C_2)) \Rightarrow N_o(C_1) > N_o(C_2)$ ,
- где  $A(W(C))$  — число атомов, входящих в вес  $W(C)$ .

**Замечание 2.4.** Очевидно, что задать порядок на графе клик можно бóльшим числом способов, а не только лексикографическим. Но так как нас будет интересовать только порядок перебора клик из одного

компонента связности графа клик, потому что только их связи определяют построение множества пучков, то на порядок на графе клик не налагается никаких требований, кроме изложенных в определении 2.8.

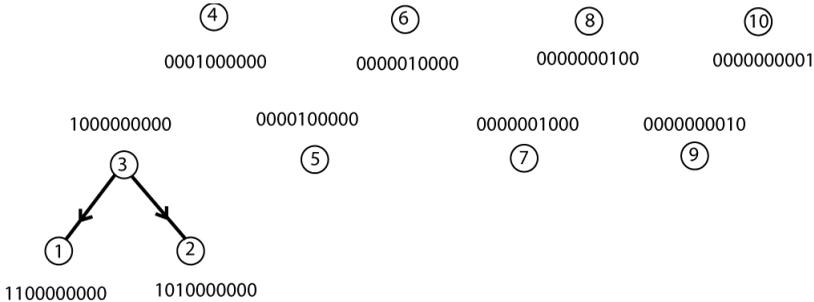


Рис. 6. Упорядоченный граф клик.  
Граф клик из рис. 5, упорядоченный в лексикографическом порядке.

### 3. Множество минимальных графов смежности.

**Определение 3.1.** *Значимое слово* — слово, являющееся весом какой-либо клики из графа клик.

**Утверждение 3.1.** Для любого минимального графа смежности  $M$  его сужение на произвольное значимое слово  $U$ :  $M \downarrow U$  — связно.

**Доказательство.** Возьмем любые две вершины  $V_1, V_2$  из сужения  $M \downarrow U$ . В самом графе они магистрально связны. Рассмотрим соединяющий их магистральный путь. Он содержит пересечение весов взятых вершин  $U' = W(V_1) \cap W(V_2)$ . Так как  $U \subseteq W(V_1)$  и  $U \subseteq W(V_2)$ , то  $U \subseteq U'$ , то есть магистральный путь содержит  $U$ , значит, он попадет в сужение  $M \downarrow U$  и выбранные вершины будут связны в этом сужении посредством рассматриваемого пути. Так как данное утверждение верно для любых двух вершин, то граф связан.

**Следствие 3.1.** Для любого минимального графа смежности  $M$  его произвольное сужение  $M \downarrow U$  магистрально связно.

**Определение 3.2.** *Сын значимого слова  $U$*  — слово, соответствующее весу клики-сына клики веса  $U$  в упорядоченном графе клик.

**Определение 3.3.** *Вассал  $V_i$  значимого слова  $U$*  для минимального графа  $M$  — компонента связности  $M \downarrow S_i$ , где  $S_i$  — сын  $U$ .

**Определение 3.4.** *Домен  $U$*  — множество вершин, входящих в  $M \downarrow U$  и не принадлежащих его вассалам (рис. 7).

**Определение 3.5.** *Значимое сужение  $G \downarrow U$*  — сужение графа  $G$  на значимое слово  $U$ .

**Определение 3.6.** *Владение* — компонента связности значимого сужения  $M \downarrow U$ , из которого удалили все ребра веса  $U$ .

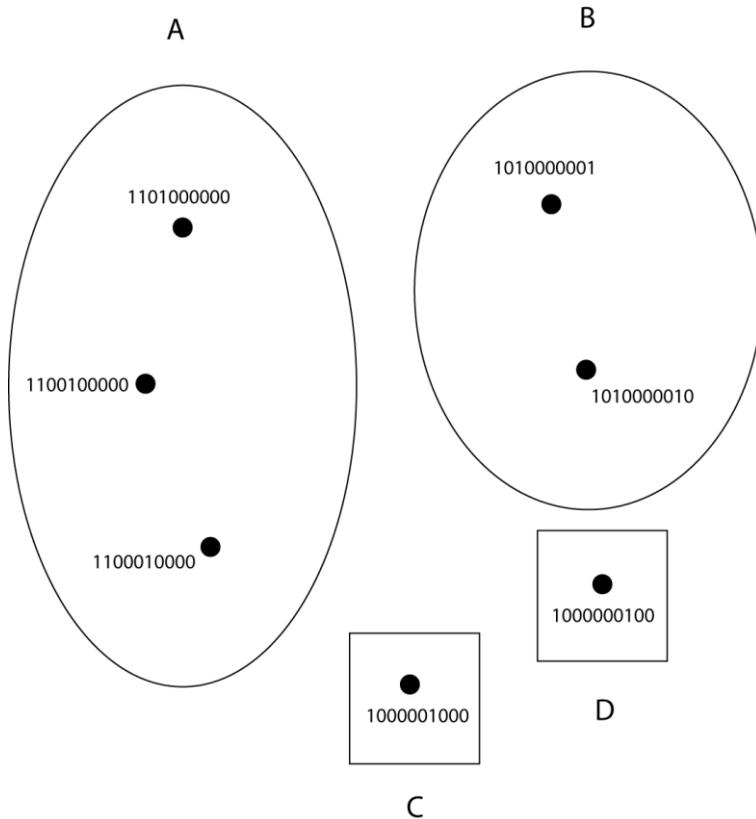


Рис. 7. Домен и вассалы.

В эллипсах расположены вершины, принадлежащие одному вассалу, в квадрат заключена каждая доменная вершина.

**Пояснение:** рассмотрим произвольное значимое сужение  $M \downarrow U$  произвольного минимального графа, разбивающееся на владения (компоненты связности)  $V_1, \dots, V_m$ . Будем так сжимать каждый компонент связности  $V_i$  в одну вершину  $v_i$ , так, что ребра, выходящие из  $V_i$  станут ребрами, выходящими из  $v_i$ . Более формально этот процесс описан в определении 3.7.

**Определение 3.7.** Сжатие подграфа  $V$  в вершину  $v$  — отображение на множестве графов, сопоставляющее графу  $G_1$  с выделенным подграфом  $V$  граф  $G_2$ , в который входят все вершины графа  $G_1$  кроме вершин подграфа  $V$ , которые сжимаются в одну вершину  $v$ , дополнительно входящую в граф  $G_2$ . Также в граф  $G_2$  входят все ребра из графа  $G_1$ , которые соединяют вершины не из  $V$ , а так же ребра, выходящие из вершины  $v$ , которым в графе  $G_1$  соответствовали ребра, соединяющие вершины из  $V$  с вершиной не из  $V$ .

То же самое, но более формально:

**Определение 3.7'.** Сжатие  $p_{V \rightarrow v}: G \rightarrow G, p_{V \rightarrow v}(G_1) = G_2$ , где  $V \subseteq G_1$ , таково, что:

- 1)  $V(G_2) = V(G_1) \setminus V \cup \{v\}$ ;
- 2)  $(v_i, v_j) \in E(G_1); v_i, v_j \notin V \Rightarrow (v_i, v_j) \in E(G_2)$ ;
- 3)  $(v_i, v_j) \in E(G_1); v_i \notin V, v_j \in V \Rightarrow (v_i, v) \in E(G_2)$ ;

**Замечание 3.1.** Сжатие  $p_{M \setminus S_i = V_i \rightarrow v}$  сохраняет связность графа.

**Замечание 3.2.** Доменные вершины при сжатии остаются вершинами.

**Определение 3.8.** Феод значимого слова  $U - v$  из определения 3.7. — вершина, получившаяся сжатием какого-то владения (рис. 8).

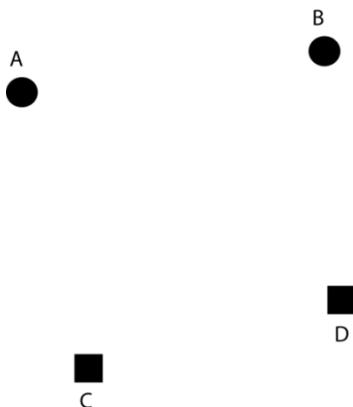


Рис. 8. Феоды.

Феоды, получающиеся путем сжатия владений на рис. 7.

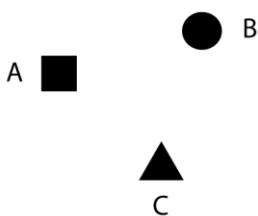
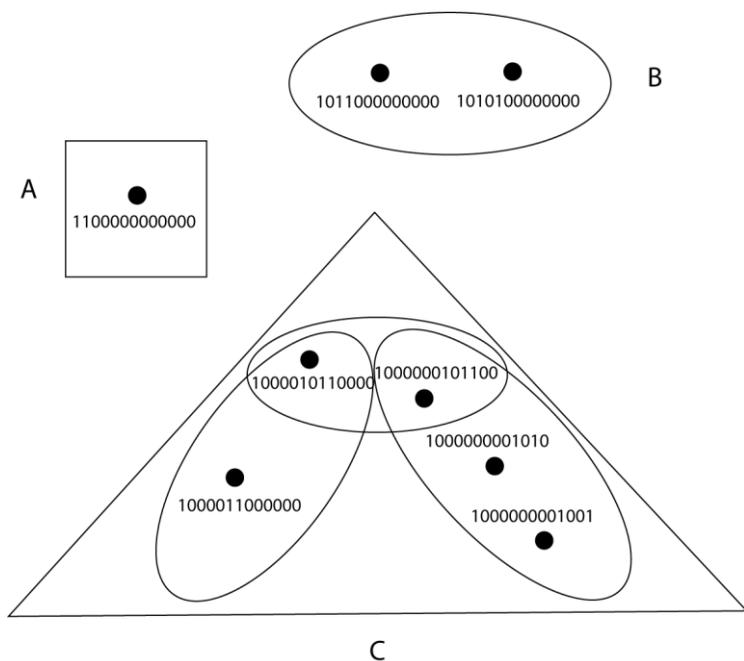


Рис. 9. Владения и феодалы.

В верхней части изображена клика, у которой одна доменная вершина и четыре вассала, которые образуют три компонента связности. В нижней части изображены феодалы, получающиеся путем сжатия владений верхней части картинки.

**Пояснение:** рассмотрим произвольное значимое сужения  $M \downarrow U$  произвольного минимального графа. Если удалить из него все ребра веса  $U$ , то сужение распадется на несколько компонент связности. Ка-

ждая доменная вершина является компонентом связности, потому что не входит ни в один из вассалов, а значит, ребра, соединяющие ее с другими вершинами, имеют вес в точности  $U$ . Если вассал не пересекается ни с каким другим вассалом, то он также образует компонент связности, потому что его внутренние ребра имеют вес, который включает в себя  $U$ , но не равный ему, а все внешние ребра были удалены. Наконец, если какой-то вассал пересекается с каким-то другим вассалом, то он образует с ним компоненту связности, потому что из одного вассала можно добраться в другой через пересечение. Естественно, любой вассал, который пересекается с одним из первых двух, так же входит в эту компоненту связности. Значит, любая компонента связности сужения с удаленными ребрами — либо доменная вершина, либо вассал, либо связный набор пересекающихся вассалов (рис. 9).

**Пояснение:** благодаря операции сжатия для каждого значимого сужения  $M \downarrow U$  существует два множества вершин: соответственно до сжатия (владения) —  $A$ , и после сжатия (феоды) —  $B$ . Рассмотрим какое-нибудь дерево на вершинах второго графа. Ему в соответствие можно поставить множество графов на вершинах множества  $A$ , так, что ребру, соединяющему два феода, может соответствовать любое ребро, соединяющее вершины из владений, сжатых до соответствующих феодов (то есть либо доменную вершину, либо вершину, принадлежащую соответствующему вассалу или объединению вассалов). Подробнее это соответствие раскрыто в определении 3.10.

**Определение 3.9.** Оммаж  $H_i^U$  — дерево, построенное на феодах  $U$  значимого сужения  $M \downarrow U$  (рис. 10).

**Определение 3.10.** Соответствие  $\xi(H_i^U)$  сопоставляет оммажу  $H_i^U$  множество графов, построенных на вершинах значимого сужения  $M \downarrow U$  таким образом, что любой граф  $s \in \xi(H_i^U)$  удовлетворяет следующему требованию: для любого ребра из оммажа  $H_i^U$ , соединяющего два феода, в графе  $s$  существует единственное ребро, соединяющее вершину владения, сжатого до первого феода и вершину владения, сжатого до второго феода.

**Определение 3.11.** Жила  $s_{i,j}^U$  — граф (рис. 11), построенный на вершинах значимого сужения  $M \downarrow U$  и соответствующий оммажу  $H_i^U$ :

$$s_{i,j}^U \in \xi(H_i^U).$$

**Утверждение 3.2.** Ребра любой жилы любого значимого слова  $U$  имеют вес  $U$ .

**Доказательство.** Так как все эти ребра входят в сужение  $M \downarrow U$ , то все они содержат вес  $U$ . Если какое-либо ребро имеет вес  $W \supset U$ , то

оно должно лежать в сужении  $M \downarrow W$ , а, значит, связывает вершины, принадлежащие вассалу  $W$ , следовательно, оно не может входить в жилу.

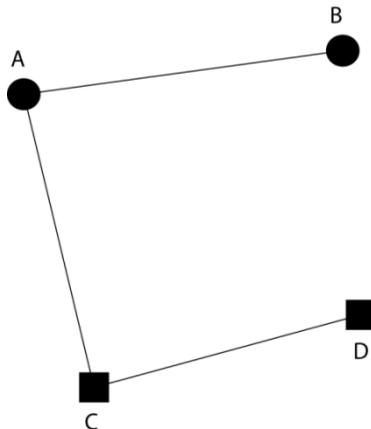


Рис. 10. Оммаж.

**Утверждение 3.3.** Для любого минимального графа смежности  $M$  в произвольном сужении  $M \downarrow U$  множество ребер веса  $U$  является жилой.

**Доказательство.** Из утверждения 3.1. следует, что сужение  $M \downarrow U$  должно быть магистрально связно. Более того, владения  $U$  магистрально связны. Любые два владения не могут быть соединены более чем одним ребром, потому что тогда граф смежности не будет минимальным. Действительно, если вершины  $a_1$  и  $a_2$  владения  $A$  соединены соответственно с вершинами  $b_1$  и  $b_2$  владения  $B$  (какая-то из этих пар вершин может быть на самом деле одной вершиной, эта вершина может быть доменной), то от  $b_1$  до  $a_2$  существует два магистральных пути: через  $b_1$  и через  $a_2$ , что не может сочетаться с минимальностью графа  $M$ . Благодаря такому ограничению связей, мы можем рассматривать сжатие сужения  $M \downarrow U$  — то есть граф на его феодах. В силу того, что  $M \downarrow U$  минимально и связно, данный граф будет деревом, то есть оммажем. А, значит, само сужение  $M \downarrow U$  является какой-то из жил, соответствующих этому оммажу. Так как ребра внутри владений имеют вес, который строго содержит вес  $U$ , то множество ребер веса  $U$  в сужении  $M \downarrow U$  является жилой.

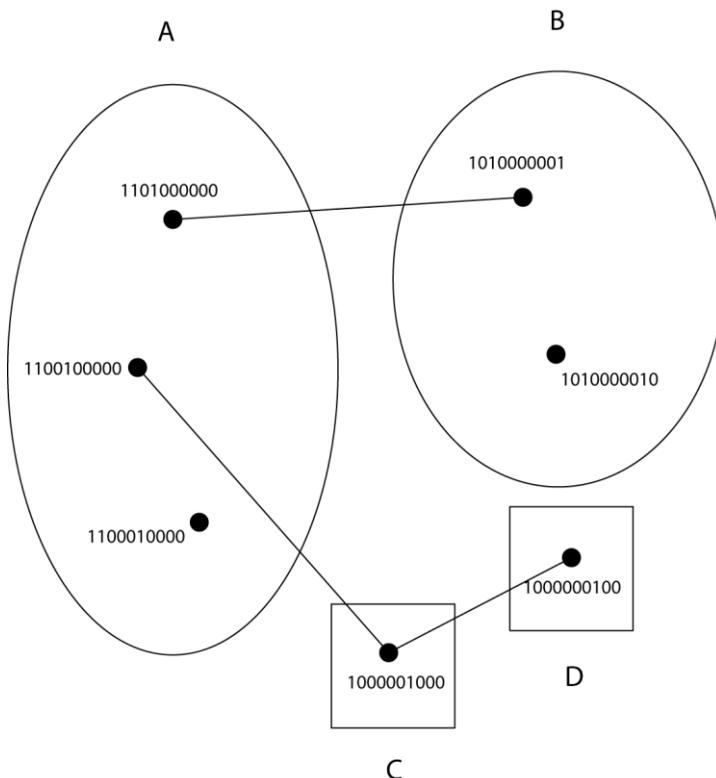


Рис. 11. Жила.

Одна из возможных жил, соответствующая оммажу на рис. 10.

**Определение 3.12.** Объединением множества графов на одном наборе вершин является граф, построенный на этом наборе вершин и содержащий все ребра этих графов.

$$G = \bigcup_i G_i: V(G) = V(G_1), E(G) = \bigcup_i E(G_i), \quad \text{где } V(G_1) = V(G_2) = \dots$$

**Определение 3.13.** Пучок — граф, построенный путем объединения жил, выбранных по одной для каждого значимого слова (рис. 12).

**Теорема 1 (о множестве пучков).** Каждый пучок однозначно задается набором жил по одной для каждой клики.

**Доказательство.** Пойдем от противного и предположим, что существуют два набора жил, по одной для каждой клики, —  $S = \{s^A, s^B, s^C, \dots\}$  и  $T = \{t^A, t^B, t^C, \dots\}$  — таких, что их объединения совпадают и равны  $B$ . Для любого значимого слова  $U$  сужение  $B \downarrow U$  состоит

из ребер с весом  $U$ , составляющих жилу клики  $U$  и ребер с весом, содержащим  $U$ , которые относятся к другим жилам. Таким образом, для любого значимого слова  $U$  жилы  $s^U$  и  $t^U$  из клик  $U$  в наборах жил  $S$  и  $T$  совпадают. Таким образом, сами наборы  $S$  и  $T$  совпадают.

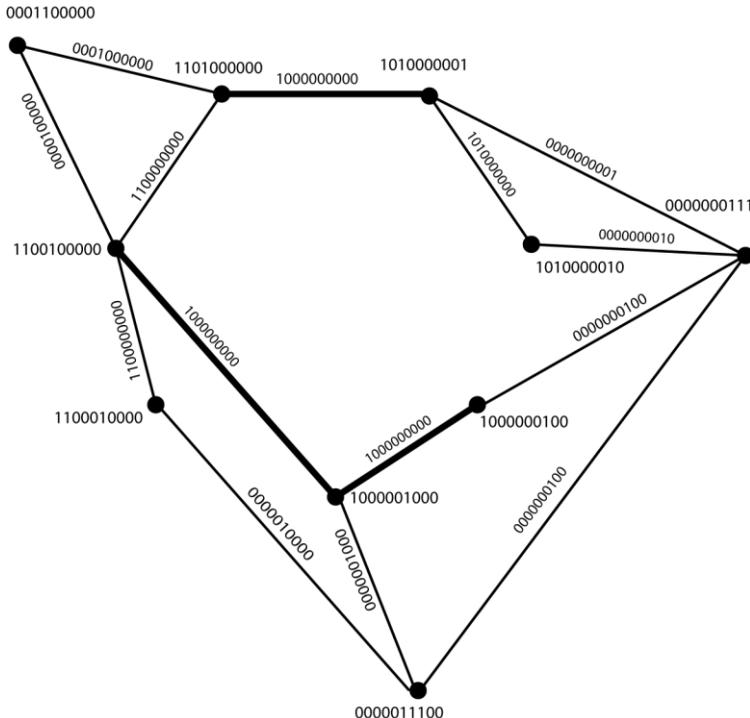


Рис. 12. Пучок.

Пучок, в котором выбрана жила из рис. 11.

**Лемма 3.1.** Для данного набора вершин  $V$  все пучки имеют одинаковое число ребер.

**Доказательство.** Это следует из того, что для любого значимого слова число ребер во всех его жилах равно, так как это деревья на одинаковом наборе вершин; и того, что никакие две жилы разных значимых слов не имеют общих ребер, так как у ребер будут разные веса. Таким образом, число ребер в пучке равно сумме чисел ребер жил для каждого значимого слова, то есть постоянно.

**Лемма 3.2.** Если любое значимое сужение графа магистрально связано, то он является графом смежности.

**Доказательство.** Возьмем любые две вершины  $a$  и  $b$ , у которых пересечение весов не пусто и равно  $U$ . Они обе попадают в сужение графа на  $U$ . В этом сужении между  $a$  и  $b$  существует магистральный путь, а, значит, этот же магистральный путь связывает их в самом графе, следовательно, граф является графом смежности.

**Определение 3.14.** *Бездетным* называется значимое слово, такое, что у порожденной этим словом клики нет сыновей в графе клик.

**Теорема 2 (о множестве минимальных графов смежности).** Множество минимальных графов смежности совпадает с множеством пучков.

**Доказательство.** Докажем, что любой пучок является графом смежности. Будем доказывать, что любое значимое сужение пучка магистрально связано. Проведем некое подобие индукции.

База: сужение пучка на бездетные слова магистрально связано, потому как оно представляет собой жилу только на доменных вершинах, т. е. является деревом. Переход: рассмотрим некое значимое сужение  $P \downarrow U$ , про которое мы знаем, что все сужения  $P \downarrow S_i$  на сыновей  $U$  магистрально связаны. Значит, магистрально связными являются и все владения сужения  $P \downarrow U$ . Рассмотрим жилу пучка, выбранную для  $U$ . Она соединяет все владения  $P \downarrow U$  и все ее ребра имеют вес  $U$ , поэтому  $P \downarrow U$  магистрально связано.

Таким образом, любой пучок оказывается магистрально связным для любого значимого сужения, а значит, любой пучок магистрально связан, то есть он является графом смежности.

Теперь докажем, что любой минимальный граф смежности является пучком. Как мы уже доказали, любое значимое сужение  $M \downarrow U$  является жилой, то есть омажем на феодах. Каждое владение  $V_i$ , соответствующее какому-либо феоде  $F_i$ , является либо доменной вершиной, либо вассалом, либо набором вассалов, каждый из которых в силу индукционного предположения тоже есть жила на владениях  $V_i$ . Кончается подобный «спуск» бездетными весами, их сужение есть жила только на доменных вершинах. Таким образом, мы получили, что  $M$  есть объединение жил каждого значимого слова.

Любой минимальный граф смежности является пучком, а любой пучок является графом смежности и число ребер у всех пучков равно, значит, множества графов смежности и пучков совпадают.

**Следствие 3.2.** Число ребер в минимальных графах смежности одинаково.

**Следствие 3.3.** Множество минимальных графов смежности совпадает с множеством пучков, которое равно декартовому произведению множеств жил каждой клики.

**Следствие 3.4.** Мощность множества графов смежности равна произведению мощностей множеств жил каждой клики.

**Замечание 3.3.** Согласно следствию 3.2, для того, чтобы построить множество графов смежности, достаточно для каждой клики построить множество соответствующих ей жил.

**4. Алгоритм Прюфера.** Данный алгоритм описан в работе [12]. Здесь мы приводим его в нотации, удобной в рассматриваемом контексте.

**Пояснение:** рассмотрим обыкновенный ненаправленный граф из  $n$  вершин, которым произвольным образом присвоены номера с 0 по  $(n - 1)$ .

**Определение 4.1.** *Висячая вершина (листья)* — вершина, из которой выходит только одно ребро.

**Определение 4.2.** *Бирка диаметра  $n$*  —  $(n - 2)$ -местный кортеж из чисел от 0 до  $n - 1$ , встречающихся в нем не более 1 раза.

**Определение 4.3.** *Код Прюфера* для дерева из  $n$  пронумерованных вершин — это бирка диаметра  $n$ , которая строится путем последовательного выполнения  $n - 2$  этапов, каждый из которых состоит из следующих шагов:

- находится висячая вершина с наименьшим номером;
- номер вершины, с которой найденная соединена единственным ребром, записывается в самую левую свободную ячейку бирки;
- найденная вершина удаляется из дерева.

**Определение 4.4.** *Алгоритм Прюфера* строит по коду Прюфера дерево и состоит из  $n - 2$  этапов. Каждый этап состоит из следующих шагов:

- рассматривается крайнее слева число кода Прюфера;
- вершина с данным номером соединяется с вершиной с наименьшим номером, не встречающейся ни в коде Прюфера, ни во множестве уже рассмотренных вершин  $L_T$ .
- упомянутая вершина с наименьшим номером добавляется к множеству  $L_T$ ;
- из кода Прюфера удаляется крайнее слева число.

После завершения этих  $n - 2$  этапов проводится ребро между двумя вершинами, не содержащимися в  $L_T$ .

**Замечание 4.1.** Так как алгоритм Прюфера восстанавливает по коду Прюфера соответствующее коду дерево, а кодом может быть любая бирка, то перебирая каждую бирку и восстанавливая по ней дерево, мы переберем все деревья на данном множестве вершин.

## **5. Схема алгоритма построения минимальных графов смежности.**

**Определение 5.1.** Алгоритм построения минимальных графов смежности строит по заданному набору вершин  $V$ , соответствующих множеству главных конъюнктов максимальных ФЗ, все возможные минимальные деревья смежности графа МФЗ.

Над множеством вершин  $V$  строится максимальный граф смежности  $G_{\max}$ . По этому графу строится граф клик Clique. Каждой клике сопоставляется множество вершин, которые в нее попали. На графе клик задается порядок.

По заданному порядку перебираются все клики из упорядоченного дерева клик. Для каждой клики при помощи алгоритма Прюфера перебираются все оммажи, поскольку оммаж является деревом — и для каждого оммажа перебираются все соответствующие ему жилы. Таким образом, для каждой клики перебираются все жилы, соответствующие ее весу. Все такие жилы записываются в массив жил для данной клики.

Согласно теореме о множестве минимальных графов смежности, нам достаточно построить все возможные пучки, чтобы получить все минимальные графы. Это можно сделать, перебрав всевозможные комбинации жил для каждой клики, объединяя такие жилы в единый граф.

## **6. Граф обязательных ребер и стереоклики.**

**Пояснение:** теорема о множестве минимальных графов смежности утверждает, что для того, чтобы перебрать все возможные минимальные графы смежности, достаточно перебрать все возможные комбинации жил по одной для каждой клики. Заметим, что бывают клики, для которых существует всего одна жила. Было бы удобно не выбирать такие жилы каждый раз для всех комбинаций жил, а объединить их в одном дереве, что уменьшило бы количество элементов, по которым ведется перебор, а, значит, ускорило бы работу программы.

**Утверждение 6.1.** Следующие два утверждения эквивалентны:

- 1) клика  $C$  состоит из двух вершин;
- 2) для клики  $C$  существует всего одна жила.

**Доказательство.**

1)  $\Rightarrow$  2) Очевидно.

2)  $\Rightarrow$  1) Пускай клика  $C$  состоит более чем из двух вершин и существует всего одна жила. Если есть хотя бы три доменные вершины, то жил может быть несколько. Значит, существует хотя бы одно владение  $V$ , состоящий из  $2 \leq n$  вершин. Ему соответствует феодал  $F$ . Каждому ребру оммага, которое выходит из феода  $F$ , может соответствовать  $n$  ребер, выходящих из каждой вершины владения, а, значит, жил не меньше  $n$ , что приводит к противоречию.

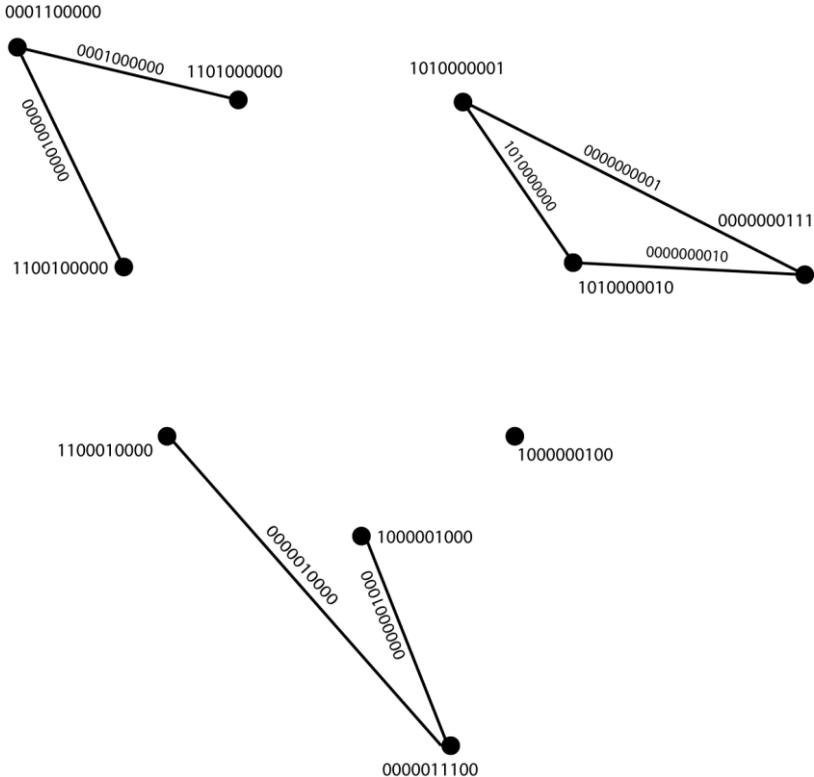


Рис. 13. Граф обязательных ребер.

**Определение 6.1.** *Моноклика* — клика, состоящая из двух вершин.

**Замечание 6.1.** Жила моноклики является ребром.

**Определение 6.2.** *Обязательное ребро* — ребро, являющееся жилой моноклики.

**Определение 6.3.** *Граф обязательных ребер* — граф на вершинах  $V^*$ , состоящий только из обязательных ребер (Рис. 13).

**Определение 6.4.** *Стереоклика* — клика, состоящая более чем из двух вершин.

**Замечание 6.2.** Мощность множества графов смежности равна произведению мощностей множеств жил каждой стереоклики.

**7. Схема улучшенного алгоритма построения минимальных графов смежности.** Благодаря понятиям стереоклики и моноклики мы можем улучшить алгоритм построения минимальных графов смежности. Вместо того чтобы каждый раз выбирать обязательные ребра моноклик, затрачивая на это время, мы можем выделить их в специальный граф, который в дальнейшем добавлять к наборам жил по одной для каждой стереоклики.

**Определение 7.1.** *Алгоритм построения минимальных графов смежности* строит по заданному набору вершин  $V$ , соответствующих множеству главных конъюнктов максимальных ФЗ, все возможные минимальные деревья смежности графа МФЗ.

Над множеством вершин  $V$  строится максимальный граф смежности  $G_{\max}$ . По этому графу строится граф клик Clique. Каждой клике сопоставляется множество вершин, которые в нее попали. На кликах задается порядок.

По порядку перебираются все клики из упорядоченного дерева стереоклик. Если клика является монокликой, то ее ребро добавляется в граф обязательных ребер OEGraph. Если же клика является стереокликой, то при помощи алгоритма Прюфера перебираются все оммажи, поскольку оммаж является деревом — и для каждого оммажа перебираются все соответствующие ему жилы. Таким образом, для каждой стереоклики перебираются все жилы, соответствующие ее весу. Все такие жилы записываются в массив жил для данной стереоклики.

Согласно теореме о множестве минимальных графов смежности, нам достаточно построить все возможные пучки, чтобы получить все минимальные графы. Это можно сделать, перебрав всевозможные комбинации жил для каждой стереоклики, объединяя такие жилы и граф OEGr.

**8. Заключение.** Основным результатом работы представляется структурная теорема о совпадении множества минимальных графов смежности и множества пучков. Она (в совокупности с рядом введенных определений и объектов) обеспечивает инструментарий для строгого описания, а также выполнения исследования и поиска решения задачи, позволяющей автоматически обучить вторичную структуру

алгебраической байесовской сети по ее первичной структуре (в иных терминах — осуществить глобальный структурный синтез логико-вероятностной модели базы фрагментов знаний с неопределенностью). Фактически, установлено, что такого рода синтез будет иметь много возможных результатов — одному и тому же исходному набору главных конъюнктов, как правило, будет соответствовать несколько минимальных графов смежности. Описано сходство и различие таких графов, приведены основные этапы способов их построения.

Заметим, что алгоритм формирования минимального графа смежности (одного из множества возможных) рассматривается в работе [1], а в работе [10] представлена ранняя версия подхода к реализации минимальных графов смежности, послужившая в качестве основы для настоящей работы.

Дальнейшие исследования (развитие полученных результатов) потребует направить на выделение среди минимальных графов смежности оптимальных (или «приемлемых») по ряду критериев: минимальность вторичной структуры (например, по числу связей между фрагментами знаний), минимальность циклов во вторичной структуре (как по «длине», так и по числу), минимальность/максимальность диаметра вторичной структуры (предполагается, что этот показатель влияет на возможность распараллелить вычисления, связанные с логико-вероятностным выводом). Хотя задача описания свойства минимальных вторичных структур в какой-то степени решена, остается актуальным выявление инвариантов таковых структур, причем один из них может сыграть роль «третичной» структуры АБС, которая откроет возможность упростить или оптимизировать реализацию алгоритмов логико-вероятностного вывода в АБС и/или автоматического обучения последних.

### Литература

1. *Опарин В.В., Тулупьев А.Л.* Синтез графа смежности с минимальным числом ребер: формализация алгоритма и анализ его корректности // Труды СПИИРАН. 2009. Вып. 11. СПб.: Наука, 2009. С. 142–157.
2. *Тулупьев А.Л.* Алгебраические байесовские сети: глобальный логико-вероятностный вывод в деревьях смежности: Учеб. пособие. СПб.: СПбГУ; ООО Издательство «Анатолия», 2007. 40 с. (Сер. Элементы мягких вычислений).
3. *Тулупьев А. Л.* Байесовские сети: логико-вероятностный вывод в циклах. СПб.: Изд-во С.-Петербургского ун-та, 2008. 140 с. (Сер. Элементы мягких вычислений.)
4. *Тулупьев А.Л.* Задача локального автоматического обучения в алгебраических байесовских сетях: логико-вероятностный подход // Труды СПИИРАН. 2008. Вып. 7. СПб.: Наука, 2008. С. 11–25.
5. *Тулупьев А.Л.* Автоматическое обучение фрагментов знаний в алгебраических байесовских сетях // Интегрированные модели и мягкие вычисления в искусственном

интеллекте. V-я Международная научно-практическая конференция. Сборник научных трудов. В 2-х т. Т. 1. С. 163–176.

6. *Тулупьев А.Л.* Матрично-векторные уравнения в задачах локального обучения алгебраических байесовских сетей // Региональная информатика-2008 (РИ-2008). XI Санкт-Петербургская международная конференция. Санкт-Петербург, 22–24 октября, 2008 г.: Материалы конференции / СПОИСУ. СПб., 2009. С. 91–99.
7. *Тулупьев А.Л.* Обработка дополнительной нечисловой информации в локальном обучении алгебраических байесовских сетей по выборкам с пропусками // Международная конференция по мягким вычислениям и измерениям. Сборник докладов. 2009. Т. 1. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2009. С. 139–142.
8. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
9. *Тулупьев А.Л., Сироткин А.В., Николенко С.И.* Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во С.-Петербург. ун-та, 2009. 400 с.
10. *Тулупьев А.Л., Столяров Д.М., Ментоков М.В.* Представление локальной и глобальной структуры алгебраической байесовской сети в Java-приложениях // Труды СПИИРАН. 2007. Вып. 5. СПб.: Наука, 2007. С. 71–99.
11. *Korb K.B., Nicholson A.E.* Bayesian Artificial Intelligence. New York: Chapman and Hall/CRC, 2004. 364 p.
12. *Prüfer H.* Neuer Beweis eines Satzes über Permutationen. Arch. Math. Phys. 1918. No. 27. S. 742–744.

**Фильченков Андрей Александрович** — студент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: автоматическое обучение вероятностных графических моделей. Число научных публикаций — 2. [aaafil@mail.ru](mailto:aaafil@mail.ru), СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

**Filchenkov Andrey Alexandrovich** — student of Computer Science Department, SPbGU. Research area: machine learning of probabilistic graphical models. [aaafil@mail.ru](mailto:aaafil@mail.ru), SPIIRAS, 14-th line V.O., 39, St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tulupuyev.

**Тулупьев Александр Львович** — к.ф.-м.н., доцент; и.о. заведующего лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН, доцент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных исследованиях, применение методов биостатистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 160. [ALT@iias.spb.su](mailto:ALT@iias.spb.su), [www.tulupuyev.spb.ru](http://www.tulupuyev.spb.ru); СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

**Tulupuyev Alexander Lvovich** — PhD in Computer Science, Associate Professor; acting Head of Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, Associate Professor of Computer Science Department, SPbSU. Research area: uncertain data and knowledge representation and processing, mathematics and computer science applications in socio-cultural studies, biostatistics, simulation, and mathematical modeling applications in epidemiology,

data intensive software systems development technology. Number of publications — 160.  
ALT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 14-th line V.O., 39, St. Petersburg,  
199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

**Поддержка исследований.** Работа выполнена при финансовой поддержке РФФИ, проект № 09-01-00861-а.

Рекомендовано ЛПИ СПИИРАН, зав. лаб. Юсупов Р.М., член-корреспондент РАН.  
Статья поступила в редакцию 20.12.2009.

## РЕФЕРАТ

### *Фильченков А.А., Тулупьев А.Л.* Структурный анализ систем минимальных графов смежности.

Цель работы — исследование структуры минимальных графов смежности и их свойств.

Введены понятия: магистральная связность, значимое сужение, доменная вершина, вассал, владение, феоде, оммаж, жила, пучок, моноклика, стероклика, граф обязательных ребер. В частности, граф смежности максимальных фрагментов знаний был определен как граф, любые две вершины которого магистрально связны; граф клик был определен как направленный граф сужений максимального графа смежности на веса вершин; жила была определена как граф, построенный на компонентах связности сужения минимального графа смежности на какой-либо вес вершины; пучок был определен как граф, представляющий собой объединение жил.

Предложенная система терминов позволила структурировать и описать исследуемую область и выявить сходства и различия между минимальными графами смежности, построенными на одном и том же множестве максимальных фрагментов знаний.

Доказаны следующие факты: произвольное значимое сужение минимального графа смежности связно; вес ребер любой жилы значимого слова совпадает со словом; в сужении минимального графа смежности множество ребер веса, совпадающего с весом сужения, образуют жилу; все пучки имеют одинаковое число ребер; если любое значимое сужение графа магистрально связно, то граф является графом смежности.

Выведена и доказана структурная теорема о совпадении множества минимальных графов смежности с множеством пучков, которое может быть представлено как декартово произведение множеств подграфов. Это упрощает как представление, так и последовательный синтез всех элементов структуры. На основе указанной теоремы предложена схема алгоритма построения множества минимальных графов смежности и схема улучшенного алгоритма построения минимальных графов смежности. Также на основе теоремы о минимальных графах смежности сделаны выводы о мощности их множества, равной произведению мощностей множеств жил для каждой из клик, и о том, что число ребер в минимальных графах смежности одинаково.

Приведены две схемы алгоритмов построения множества минимальных графов смежности, которые последовательно строят граф клик, множество жил для веса каждой клики, и объединяют эти жилы в множество пучков, таким образом получая множество минимальных графов смежности.

Полученные теоретические результаты создают основу для корректной алгоритмизации глобального машинного обучения (структурного синтеза) алгебраических байесовских сетей. В частности, в работе приведена схема алгоритма, позволяющего для заданного набора максимальных фрагментов знаний построить множество минимальных графов смежности.

## SUMMARY

*Filchenkov A.A., Tulupyyev A.L.* **Minimal joint graph structure synthesis.**

The goal of this work is to research the structure of minimal joint graph and its properties.

The new concepts are given: backbone connectivity, significant narrowing, demesne vertex, vassal, homage, feud, sinew, bunch, monoclique, stereoclique, obligatory edges graph. Particularly, a minimal joint graph of maximal knowledge patterns was defined as the a graph, every two vertexes of which are backbone connected; a clique graph was defined as a directed graph of maximal joint graph narrowings on weights of vertexes; then a sinew was defined as a graph built over connected components of a minimal joint graph narrowing on a weight of a vertex; a bunch was defined as the union of sinews.

The proposed system of terms let to structure and describe the research space and to find the similarities and the difference between minimal joint graphs that were built over different maximal knowledge pattern sets.

The facts such as a significant narrowing of a minimal joint graph is connected; the weight of edges of a sinew of a significant narrowing is equal to the weight of the narrowing; a set of edges with the same weight is a sinew in a minimal joint graph narrowing of the same weight; all the sinews have the same number of edges; if any significant narrowing of a graph is backbone connected than the graph is the joint graph were proven.

As a result the structure theorem about coincidence of the minimal joint graph set and the bunch set that could be represented as a Cartesian product of sets of subgraphs was proven. This simplifies both the representation and consequential synthesis of all structure elements. The scheme of minimal joint graph set building algorithm and the scheme of improved minimal joint graph set building algorithm was given on the base of the said theorem. Also was proven, that the cardinality of the set is equal to a product of cardinalities of set of sinews, chosen one for every clique and the numbers of edges in every minimal joint graph are similar to each other.

In completion as a result two schemes of minimal joint graph building algorithms are given that consequently synthesize the clique graph, sinew sets for each clique and unite these sinews into a bunch set, thus producing a minimal joint graph set.

The theoretical results give a base for a correct global machine learning (structure synthesis) algorithmization of algebraic Bayesian networks. Particularly, the scheme of the algorithm letting to synthesize a minimal joint graph set for given maximal knowledge pattern set is given.