Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления» (ИУ)

КАФЕДРА «Системы обработки информации и управления» (ИУ5)

Домашнее задание по дисциплине «Методы машинного обучения» на тему: «Исследование и разработка метаграфового хранилища»

Студент группы ИУ5-21М ________________ Р.Р. Сафин

Преподаватель ________________ Ю.Е. Гапанюк

2022 г.

Содержание

Введение

В настоящее время графовые модели на основе сложных сетей все чаще используются для решения широкого класса задач, этому, в значительной степени, способствует тот факт, что в них отказываются от плоского расположения вершин и ребер, вследствие, чего такие модели могут оказать крайне положительное влияние при описании сложных моделей данных. Одной из наиболее перспективных моделей подобного типа, можно назвать, метаграфовую модель.

Стоит отметить тот факт, что метаграфовая модель была предложена несколько раз, и в настоящее время существует несколько ее разновидностей. Будучи изначально предложенной А. Базу и Р. Блэннингом в монографии 2007 года [12], модель в дальнейшем получила ряд расширений, независимо развиваемых различными группами исследователей.

Одним из важных вопросов работы с метаграфами является вопрос хранения метаграфовой модели данных и разработки программных модулей для доступа к хранилищу. В связи с этим, было решено подробнее изучить и спроектировать полноценную метаграфовую СУБД, для решения описанных ранее проблем.

Постановка задачи

Данная работа является исследованием в области развития графовых моделей. В ходе работы ставятся следующие задачи: исследование теоретических графовых моделей, обзор различных моделей графового представления: сложные графы, сложные сети, гиперграфовая модель, гиперсетевая модель, модель многоуровневой сети и метаграфовая модель. Для решения задачи исследования используются различные учебно-методические пособия и книги, представленные в конце данной.

Теоретическая часть

Сложные сети и сложные графы

В настоящее время термины «сложная сеть» или «комплексная сеть», которые являются различными переводами англоязычного термина «complex network») и термин «сложный граф» (англ. «complex graph») часто употребляются как синонимы. В работе [1] отмечается, что термин «сложная сеть», как правило, употребляется для обозначения реальной исследуемой системы, в то время как термин «сложный граф» обычно используют для обозначения математической модели такой системы.

Наибольшие разночтения вызывает термин «сложный» применительно к графовым моделям. Как правило, термин «сложный» трактуется в двух вариантах:

1. Вариант 1 сложного графа. Это плоские графы (сети) очень большой размерности. Такие сети могут включать миллионы и более вершин. Ребра, соединяющие вершины, могут быть ненаправленными или направленными. Иногда используется модель мультиграфа, в этом случае две вершины могут соединяются не одним, а несколькими ребрами.

   Именно такую модель в литературе чаще всего называют «сложной сетью». Исследования данной модели проводятся в основном специалистами в области математики. Исследователи рассматривают такие параметры как распределение количества связей между вершинами, выделение сильно связанных подграфов. Часто для связей вводится количественная метрика, которая обычно трактуется как расстояние между вершинами. Активно исследуются динамические модели, в которых к существующей сложной сети случайным образом добавляются вершины и ребра. Такие модели представляют интерес при изучении социальных сетей, глобальных компьютерных сетей, различных социологических и биологических моделей. Но они не очень хорошо помогают при описании сложных моделей данных и знаний.

2. Вариант 2 сложного графа. Сложные графы, в которых используется сложное (комплексное) описание вершин, ребер и/или их распоряжения. Часто в таких моделях отказываются от плоского представления вершин и ребер. Именно подобные модели могут быть наиболее полезны при описании сложных моделей данных. На сегодняшний день известны четыре подобных модели: гиперграф, гиперсеть, метаграф и многоуровневая сеть (которая является упрощенным вариантом гиперсети).

   В настоящее время в литературе ещё не появился единый «собирательный термин» для моделей такого класса. Авторы моделей, как правило, используют собственные названия для каждой модели, не всегда указывая на родство предлагаемой модели со сложными графами (сетями).

Для подобного класса моделей предложим такое «собирательный термин» как «ансамбли сложных сетей (графов)» или «ансамблевые модели сложных сетей (графов)». Для гиперсетевой и метаграфовой моделей может быть использован термин «сложные сети (графы) с эмерджентностью», так как данные модели реализуют принципе эмерджентности, используемый в общей теории систем.

Гиперграфовая модель

Рассмотрим формализованную модель гиперграфа в соответствии с [2]:

$$HG = \langle V,HE\rangle, hej = \{vi\}, vi \in V, hej \in HE, \quad (1.1)$$

где HG – гиперграф; V – множество вершин гиперграфа; HE – множество непустых подмножеств V , называемых гиперребрами; vi – вершина гиперграфа; hej – гиперребро гиперграфа.

Гиперграф может быть направленным или ненаправленным. Гиперребра ненаправленного гиперграфа показывают включение вершин, в то время как гиперребра направленного гиперграфа задают порядок обхода вершин. Таким образом, гиперребро рассматривается как множество (коллекция) вершин, которое в случае направленного гиперграфа может быть упорядоченным.

Пример описания гиперграфа показан на рисунке 1. Данный ненаправленный гиперграф включает три гиперребра: he1, he2, he3. Гиперребро he1 включает вершины v1, v2, v4, v5. Гиперребро he2 включает вершины v2 и v3. Гиперребро he3 включает вершины v4 и v5. Гиперребра he1 и he2 имеют общую вершину v2. Все вершины гиперребра he3 также являются вершинами гиперребра he1.

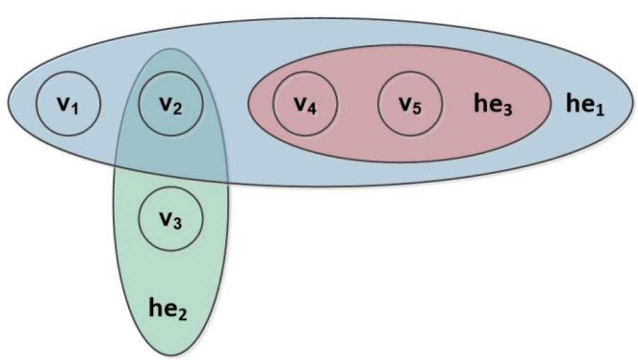

Рисунок 1 — Пример описания ненаправленного гиперграфа.

В соответствии с определением гиперграфа, операция иерархической вложенности для гиперребер не определена явным образом, можно говорить только о пересечении множеств вершин, вложенных в гиперребра.

Поэтому, хотя классический гиперграф и содержит гиперребра, но не позволяет моделировать сложные иерархические зависимости и не является полноценной «сложной сетью с эмерджентностью».

Необходимо отметить, что некоторые современные СУБД (в частности, HypergraphDB) реализуют концепцию иерархического гиперграфа, который можно определить следующим образом:

$$HHG = \langle V, HE \rangle, hej = \langle \{vi\}, HE' \rangle, vi \in V, \tag{1.2}$$

при этом выполняется условие:

$$HE' \subset (HE \setminus (HE'' \cup hej)), \tag{1.3}$$

где HHG – иерархический гиперграф; V – множество вершин гиперграфа; HE – множество гиперребер гиперграфа; vi – вершина гиперграфа; hej – гиперребро гиперграфа; HE′ – множество гиперребер гиперграфа, вложенных в гиперребро hej; HE″ – множество гиперребер гиперграфа, вложенных по отношению к гиперребрам, входящим в множество HE′.

Отличие иерархического гиперграфа от обычного состоит в том, что в иерархическом гиперграфе гиперребро может включать не только вершины, но и другие гиперребра. Условие (1.3) говорит о том, что гиперербро не может включать само себя и гиперребра, находящиеся ниже по иерархии вложенности, что позволяет избежать зацикливания при определении вложенности.

Вершина в иерархическом гиперграфе может рассматриваться как частный случай «пустого» гиперребра.

Операция вложенности для гиперребер становится иерархической, что позволяет моделировать сложные иерархические зависимости и делает иерархический гиперграф полноценной «сложной сетью с эмерджентностью».

Гиперсетевая модель

В данном разделе рассмотрим более «богатую» гиперсетевую модель, которая строится на основе гиперграфовой модели.

Удивительным является факт, что гиперсетевая модель (с одинаковым названием) была открыта дважды. При этом само изложение материала авторами моделей достаточно убедительно свидетельствует об отсутствии возможных заимствований. Сложно даже точно сказать какая модель была предложена ранее, потому что обе модели развивались авторами в течении длительного времени.

Гиперсетевая модель (по В.К. Попкову) предложена профессором В.К. Попковым. Первые публикации можно отнести к 1980-м годам. Гиперсетвая модель наиболее подробно рассматривается в трехтомной монографии [3; 4; 5] а также в статьях [6; 7]. В качестве математического аппарата используются теория графов и теория категорий. Модель применяется в различных прикладных задачах, особенно много применений в области организации транспортных перевозок.

Гиперсетевая модель (по Дж. Джонсону) предложена профессором Джеффри Джонсоном. Несмотря на то, что основная монография профессора Дж. Джонсона [8] вышла в 2013 году (значительно позже работ профессора В.К. Попкова), необходимо отметить, что профессор Дж. Джонсон в этой монографии ссылается на свои работы 1970-х годов, в том числе, опираясь на еще более ранние статьи своего учителя профессора Рональда Аткина. В качестве математического аппарата используются в основном методы комбинаторики, теория решеток, методы топологии(симплициальные комплексы,топологические пространства), методQанализа (предложенный профессором Рональдом Аткиным). Модель применяется в основном в области социологии.

Рассмотрим определение абстрактной гиперсети (по В.К. Попкову) в соответствии с [4].

Пусть даны гиперграфы PS ≡ WS0,WS1,WS2,...WSK . Гиперграф PS или WS0 называется первичной сетью. Гиперграф WSi называется вторичной сетью i–го порядка.

Пусть также задана последовательность отображений между сетями различных уровней Φ_i: $\{WS_K \xrightarrow{\Phi_K} WS_{K-1} \xrightarrow{\Phi_{K-1}} \dots WS_1 \xrightarrow{\Phi_1} PS\}$.

Тогда иерархическую абстрактную гиперсеть порядка K можно определить как $AS^K = \langle PS, WS_1, \dots, WS_K; \Phi_1, \dots, \Phi_K \rangle$.

Эмерджентность в гиперсети (по В.К. Попкову) возникает при переходе между уровнями за счет использования отображений Фi между «слоями» гиперребер.

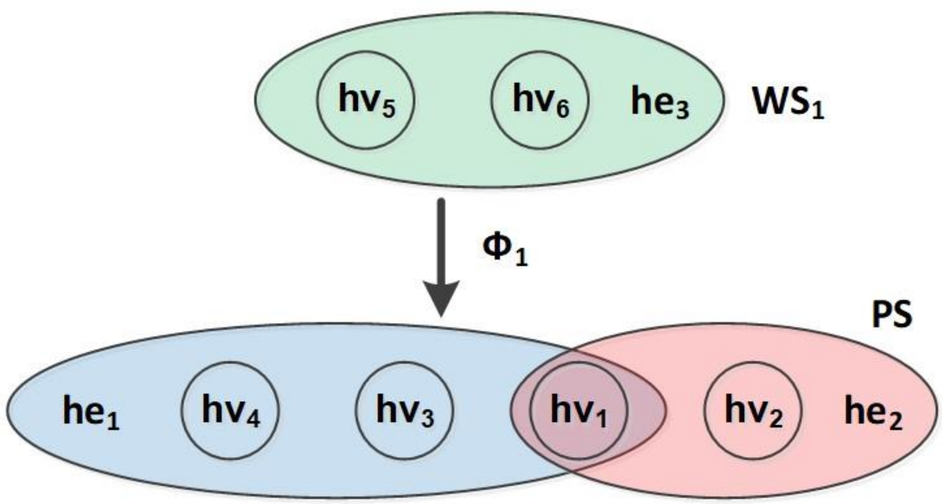

Рисунок 2 — Пример гиперсети по В.К. Попкову.

Пример гиперсети по В.К. Попкову представлен на рисунке 2. Первичная сеть PS образуется вершинами гиперребер he1 и he2. Первый уровень вторичной сети WS1 образуется вершинами гиперребра he3. Связь между уровнями задается с использованием отображения Ф1.

В монографии [8] профессор Дж. Джонсон дает формальное определение гиперсимплекса через достаточно громоздкую цепочку определений, приводить которую целиком автор считает избыточностью. При этом необходимо отметить, что в статье [9] профессор К.В. Анохин приводит удачное неформальное определение гиперсимплекса по Дж. Джонсону: «основание гиперсимплекса содержит множество элементов одного уровня, а его вершина образуется описанием их отношений и приобретает интегральные свойства, делающие ее элементом сети более высокого уровня». В своей модели когнитома [9], профессор К.В. Анохин использует модель гиперсети именно по Дж. Джонсону. Эмерджентность в гиперсети (по Дж. Джонсону) возникает при переходе между уровнями за счет возникновения гиперсимплексов. Основание гиперсимплекса содержит множество элементов одного уровня, а его вершина образуется описанием их отношений и приобретает интегральные свойства, делающие ее элементом сети более высокого уровня.

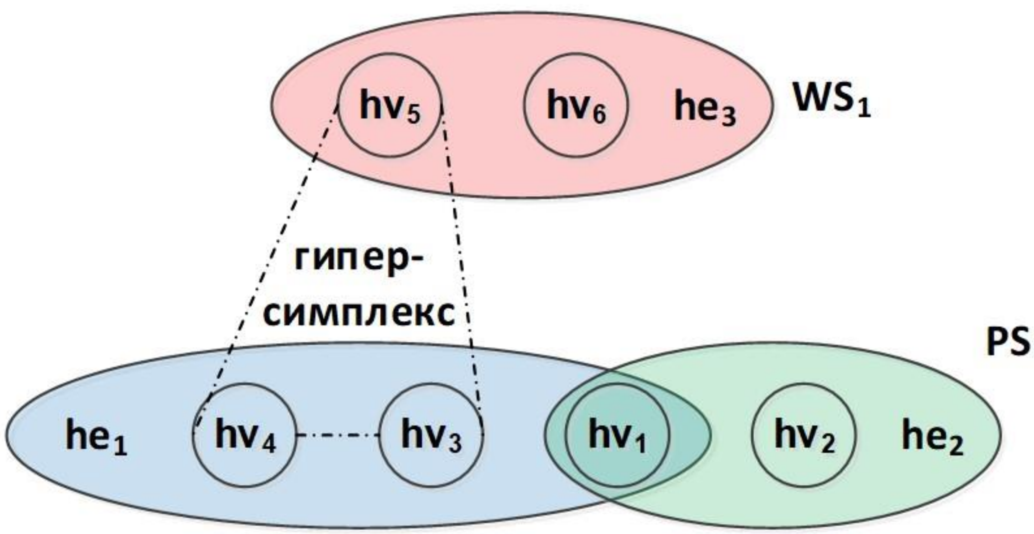

Рисунок 3 — Пример гиперсети по Дж. Джонсону.

Пример гиперсети по Дж. Джонсону представлен на рисунке 3. Как и в случае гиперсети по В.К. Попкову, первичная сеть PS образуется вершинами гиперребер he1 и he2, а первый уровень вторичной сети WS1 образуется вершинами гиперребра he3. Но для связи между уровнями используется не отображение Фi, а гиперсимплекс hv3 −hv4 −hv5. Элементы hv3 и hv4 образуют основание гиперсимплекса, а элемент hv5 является вершиной гиперсимплекса.

В обеих моделях гиперсеть является конструкцией, содержащей слои гиперграфов, при этом только соседние слои могут быть «сцеплены». Разница между моделями состоит в том, как «сцепляются» слои. В модели В.К. Попкова «сцепление» производится за счет отображений Фi. В модели Дж. Джонсона «сцепление» производится за счет использования гиперсимплексов.

Отметим, что в отличие от гиперграфовой модели, гиперсетевая модель является полноценной «сложной сетью с эмерджентностью» и позволяет моделировать сложные иерархические зависимости между элементами. Но при этом гиперсеть состоит из разнородных элементов (гиперграфов и отображений по В.К. Попкову, гиперграфов и гиперсимплексов по Дж. Джонсону). Использование разнородных элементов для описания слоев и связей между слоями усложняет описание систем на основе гиперсетевой модели.

Модель многоуровневой сети

Модель многоуровневой сети представляет собой попытку исследователей перейти от сложных сетей в варианте 1 к сложным сетям в варианте 2. Детальный обзор модели приведен в работах [1; 10; 11].

В работе [1] отмечается, что в последнее время исследователи все чаще обращают внимание на многоуровневый характер реальных сложных систем. Многоуровневая сеть, предлагается в качестве графового формализма для описания сложных систем.

В работе [10] дается следующее формализованное описание многоуровневой сети MN:

$$MN = \langle \{G_1,...,G_L\},\{E_{ij} \subset X_i \times X_j, i,j \in \{1,...,L\}, i \neq j\}\rangle,$$
$$G_K = \langle X_K, E_K \rangle \quad (1.4)$$

В приведенной формуле {G1,...,GL} – семейство плоских графов GK = ⟨XK,EK⟩, где XK – множество вершин графа GK, EK – множество ребер графа GK. Множество {G1,...,GL} называется уровнями в многоуровневой сети MN.

Граф GK может быть направленным или ненаправленным, взвешенным или невзвешенным, а также мультиграфом.

Каждый элемент Eij ⊂ Xi×Xj,i,j ∈ {1,...,L},i ≠ j является множеством связей между вершинами графов уровней i и j. Условие i ≠ j говорит о запрете циклических связей на одном уровне, связи могут быть только между элементами соседних уровней.

Пример многоуровневой сети представлен на рисунке 4. Сеть содержит три уровня G1, G2, G3. В данном примере каждый уровень является плоским ненаправленным графом. Ребра графов EK показаны сплошными линиями. Примеры связей Eij между уровнями показаны пунктирными линиями.

При условии L = 1 многоуровневая сеть превращается в обычную одноуровневую сеть.

В работе [1] для многоуровневой сети MN вводится операция проекции на одноуровневую сеть:

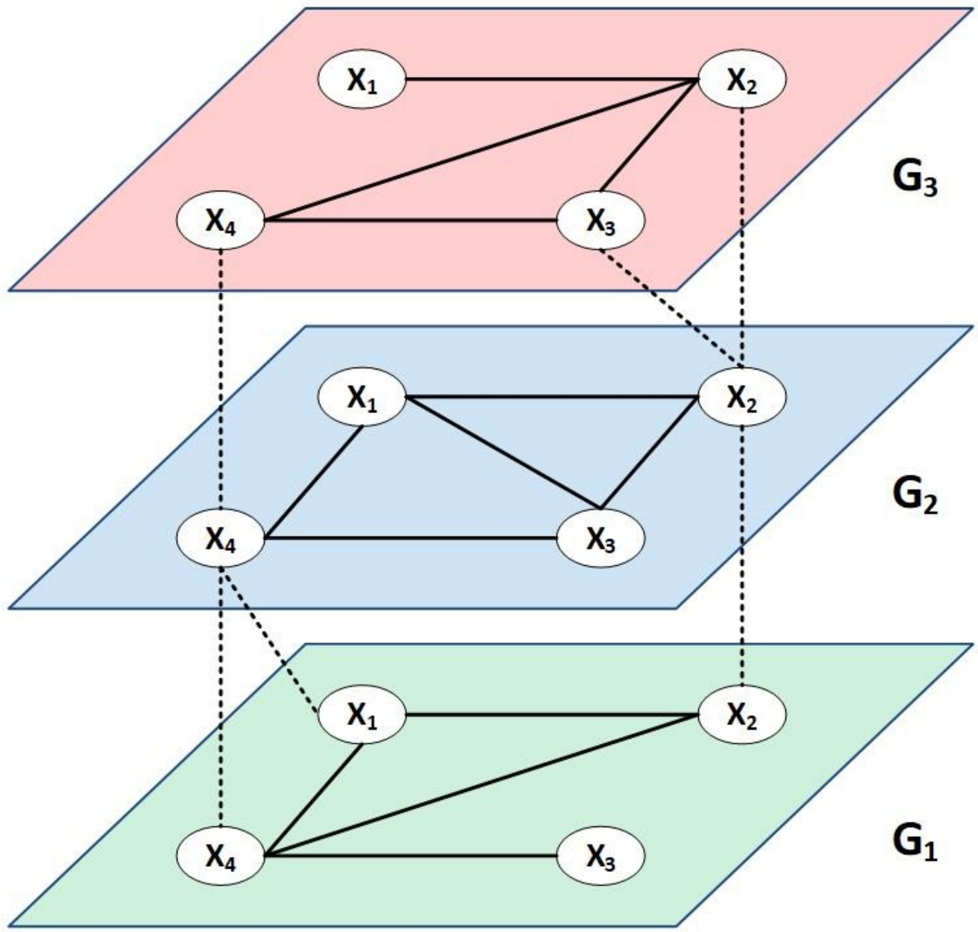

Рисунок 4 — Пример многоуровневой сети.

$$
\begin{gathered}
proj(MN) = \left\langle X^{proj}, E^{proj} \right\rangle, \\
X^{proj} = \bigcup_{k=1}^{L} X_k, X_k \subset G_k, \\
E^{proj} = \left(\bigcup_{k=1}^{L} E_k \right) \cup \left(\bigcup_{1 \leqslant i \neq j \leqslant L} E_{ij} \right), E_k \subset G_k
\end{gathered} \tag{1.5}
$$

Проекция proj(MN) многоуровневой сети MN является одноуровневой сетью (плоским графом). Множеством вершин графа Xproj является объединение множеств вершин графов Gk, соответствующих уровням сети.

Множеством ребер графа Eproj является объединение множеств ребер между вершинами на соответствующих уровнях сети EK и связей между уровнями Eij.

В работе [1] также отмечается, что частным случаем многоуровневых сетей являются мультиплексные сети (англ. multiplex networks). В этом случае связи между слоями разрешены только для одноименных вершин графов (вершин с одинаковыми индексами). Формально это ограничение может быть записано следующим образом:

$$Eij = \{\langle xik,xjk\rangle,xk \in X\},i,j \in \{1,...,L\},i \neq j. \quad (1.6)$$

Это означает что связь в мультиплексной сети может быть установлена только между одноименными элементами xk (у обоих элементов одинаковый индекс k) расположенными на соседних уровнях i и j.

В качестве примера рассмотрим рисунок 4. Например, связь X2 − X2 между уровнями G1 − G2 не нарушает условия мультиплексности сети, так как соединены одноименные элементы (с одинаковыми индексами). Но связь X1−X4 между уровнями G1 − G2 нарушает условие мультиплексности сети, так как соединены разноименные элементы (с различными индексами).

Сравним модель многоуровневой сети и гиперсетевую модель. Как и гиперсетевая модель, модель многоуровневой сети является послойной. Если в гиперсетевой модели на каждом уровне применяются гиперграфы, то в многоуровневой сети уровнем является более простая модель – обычный плоский граф. Связи между уровнями Eij можно рассматривать как частный случай отображения Фi в гиперсети на основе модели В.К. Попкова. Таким образом, модель многоуровневой сети можно считать частным упрощенным случаем гиперсетевой модели в интерпретации В.К. Попкова.

Необходимо также отметить, что модель многоуровневой сети фактически не содержит элементов эмерджентности, так как состоит из плоских графов и связей между вершинами, расположенными в различных слоях. Единственной эмерджентной составляющей в данной модели можно считать слои. Таким образом, модель многоуровневой сети относится к ансамблевым моделям сложных сетей, но в силу отсутствия эмерджентных составляющих не может быть признана полноценной «сложной сетью с эмерджентностью».

Практическая часть

Метаграфовая модель

Метаграфовая модель, несомненно, является наиболее интересной и перспективной моделью сложной сети. Необходимо отметить, что метаграфовая модель отчасти повторила историю гиперсетевой модели – она была предложена несколько раз, и в настоящее время существует несколько ее разновидностей. Будучи изначально предложенной А. Базу и Р. Блэннингом в монографии 2007 года [12], модель в дальнейшем получила ряд расширений, независимо развиваемых различными группами исследователей.

Все элементы метаграфовой модели могут быть снабжены атрибутами:

$$ATR = \{atr_k\}, atr_k = \langle atr_k^n, atr_k^t, atr_k^v \rangle, \quad (1.7)$$

где AT R – множество атрибутов, принадлежащих заданному элементу метаграфовой модели; atrk – атрибут; atrn k – наименование атрибута; atrt k – тип атрибута; atrv k – значение атрибута.

Атрибут является атомарным, то есть значение может содержать только один элемент. При этом, значение атрибута соответствует типу атрибута.

Вершина метаграфа характеризуется множеством атрибутов:

$$v_i = \langle ATR \rangle, v_i \in V, \quad (1.8)$$

где vi – вершина метаграфа; AT R – множество атрибутов, принадлежащих вершине; V – множество вершин метаграфа.

Метавершина метаграфа характеризуется множеством атрибутов и вложенным фрагментом метаграфа:

$$mv_i = \langle ATR, MGF \rangle, mv_i \in MV, \quad (1.9)$$

где mvi – метавершина метаграфа; AT R – множество атрибутов, принадлежащих вершине; MGF – фрагмент метаграфа, вложенный в метавершину; MV – множество метавершин метаграфа.

Определения метавершины метаграфа и фрагмента метаграфа являются взаимно-рекурсивными, далее фрагмент метаграфа будет также определен через метавершину.

Необходимо отметить, что в соответствии с определениями (1.8) и (1.9), вершина является частным случаем метавершины, для которой вложенный фрагмент метаграфа является пустым:

$$v_i \equiv mv_i | MGF = \varnothing. \quad (1.10)$$

Назовем данное свойство «свойством соответствия между вершиной и метавершиной аннотированного метаграфа».

Характеризуется множеством атрибутов, исходной и конечной вершиной и признаком направленности:

$$e_i = \langle v_S, v_E, eo, ATR \rangle, e_i \in E, eo = true|false, \quad (1.11)$$

где ei – ребро метаграфа; vS – исходная вершина (метавершина) ребра; vE – конечная вершина (метавершина) ребра; eo – признак направленности ребра (eo = true – направленное ребро, eo = f alse – ненаправленное ребро); AT R – множество атрибутов, принадлежащих ребру; E – множество ребер метаграфа.

Фрагмент метаграфа в общем виде может содержать произвольные вершины, метавершины и ребра метаграфа:

$$MGF = \{ev_j\}, ev_j \in (V \cup MV \cup E), \tag{1.12}$$

где MGF – фрагмент метаграфа; evj – элемент, принадлежащий объединению множеств вершин V, метавершин MV и ребер E метаграфа.

Необходимо отметить, что фрагмент метаграфа представляет собой множество элементов метаграфовой модели (точнее, объединение множеств вершин, метавершин и ребер модели). Поэтому, по отношению к фрагменту метаграфа допустимо применение понятий пустого множества, подмножества, а также операций над множествами.

Фрагмент метаграфа не включает атрибуты, так как фрагмент метаграфа является служебной структурой, предназначенной для определения вложенности элементов. Свойства появляются у сложных элементов метаграфовой модели, содержащих вложенные элементы в виде фрагмента метаграфа.

Выводы

В данной работе было проведено исследование в области развития графовых моделей, а именно: были проведены исследования моделей сетей, графов, многоуровневых сетей, гиперсетей, гиперграфов и метаграфов.

Теоретическая часть работы содержит основную информацию исследованных разновидностей моделей сложных графов. Практическая часть работы содержит информацию о метаграфовой модели, а также описание возможного метода реализации данной модели, в частности для возможного создания метаграфового хранилища данных.

Список использованных источников

1. Intentional risk management through complex networks analysis [Text] / V. Chapela [et al.]. — 1st ed. — Springer, 2015. — 126 p.
2. Voloshin, V. Introduction to Graph and Hypergraph Theory [Text] / V. Voloshin. — Nova Science Publishers, 2009. — 287 p.
3. Попков, В. К. Математические модели связности. Том 1: Графы и сети [Текст] / В. К. Попков. — Новосибирск : ИВМиМГ (ВЦ) СО РАН, 2000. — 174 с.
4. Попков, В. К. Математические модели связности. Том 2: Гиперграфы и гиперсети [Текст] / В. К. Попков. — Новосибирск : ИВМиМГ (ВЦ) СО РАН, 2001. — 180 с.
5. Попков, В. К. Математические модели связности. Том 3: Представления графов [Текст] / В. К. Попков. — Новосибирск : ИВМиМГ (ВЦ) СО РАН, 2002. — 169 с.
6. Попков, В. К. Применение теории S-гиперсетей для моделирования систем сетевой структуры [Текст] / В. К. Попков // Проблемы информатики. — 2010. — № 4. — С. 17—40.
7. Попков, В. К. О моделировании городских транспортных систем гиперсетями [Текст] / В. К. Попков // Автоматика и телемеханика. — 2011. — № 6. — С. 179—189.
8. Johnson, J. Hypernetworks in the Science of Complex Systems [Text] / J. Johnson. — Imperial College Press, 2013. — 348 p. — (Series on Complexity Science).
9. Анохин, К. В. Когнитом: гиперсетевая модель мозга [Текст] / К. В. Анохин // XVII Всероссийская научно-техническая конференция «Нейроинформатика2015»: Сборник научных трудов. В 3 частях. Ч.1. — Москва : НИЯУ МИФИ, 2015. — С. 14—15.

10. The structure and dynamics of multilayer networks [Text] / S. Boccaletti [et al.] // Physics Reports. — 2014. — Nov. — Vol. 544, no. 1. — P. 1—122.
11. Aleta, A. Multilayer networks in a nutshell [Text] / A. Aleta, Y. Moreno // Annual Review of Condensed Matter Physics. — 2019. — Mar. — Vol. 10, no. 1. — P. 45—62.
12. Basu, A. Metagraphs and their applications [Text] / A. Basu, R. W. Blanning. – – Springer, 2007. — 172 p.