

Синадский Н. И., Сушков П. В.

МОДИФИКАЦИЯ МЕТОДОВ АНАЛИЗА СОЦИАЛЬНЫХ ГРАФОВ НА ОСНОВЕ ПРИМЕНЕНИЯ АТТРИБУТИВНЫХ КОМПОНЕНТОВ УЧЕТНЫХ ЗАПИСЕЙ ДЛЯ ИДЕНТИФИКАЦИИ СООБЩЕСТВ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ

В статье рассматривается метод идентификации сообществ, организованных относительно требуемого пользователя социальных сетей. Предложенный метод позволяет определять схожие структуры в социальных графах на основе существующих методов анализа графов: метода дифференциации вершин и метода выделения частичного изоморфизма. Учитывая значительные размеры социальных графов в статье предложена модификация указанных методов для уменьшения вычислительной сложности разрабатываемого алгоритма идентификации сообществ с учетом особенностей социальных сетей: применение начальной подстановки, исключающей переборные операции при формировании подстановки сходства, а также использование текстовых атрибутов социальных вершин для более точной и быстрой дифференциации.

В статье также рассматривается архитектура разработанного программного средства, реализующего алгоритм идентификации сообществ в социальных сетях на основе модифицированных методов анализа графов.

Ключевые слова: идентификация сообществ, определение сходства структуры графов, анализ социальных графов.

MODIFICATION OF METHODS FOR ANALYZING SOCIAL GRAPHS BASED ON THE APPLICATION OF ATTRIBUTIVE COMPONENTS OF ACCOUNTS FOR THE IDENTIFICATION OF COMMUNITIES OF SOCIAL NETWORKS USERS

The article deals with the method of identification of communities organized with respect to the required user of social networks. The proposed method makes it possible to determine similar structures in social graphs on the basis of existing graph analysis methods: the method of vertex differentiation and the method of extracting partial isomorphism. Considering the considerable size of social graphs, the article proposes a modification of these methods to reduce the computational complexity of the developed communities identification algorithm, taking into account the features of social networks: the use of initial substitution, eliminating unnumbered operations in the formation of similarity substitution, and the use of text attributes of social vertices for more accurate and rapid differentiation.

The article also discusses the architecture of the developed software that implements the algorithm of communities identification in social networks based on modified graph analysis methods.

Keywords: communities identification, determination of graphs structure similarity, analyses of social graphs.

Социальной сетью (далее – СС) будем называть отображение социальных взаимоотношений, построенное и управляющееся с помощью информационно-технологических средств. В последние годы сайты СС обрели огромную популярность. Многие пользователи являются членами одного или нескольких таких сайтов и склонны к предоставлению большого количества персональной информации в сети Интернет¹.

Решение проблемы идентификации сообществ пользователей СС, также рассматриваемой как кластерная или групповая идентификация, является современной и востребованной задачей^{2,3,4,5}. Существующие методы идентификации сообществ социальных сетей^{3,4,5} накладывают множество ограниче-

ний, связанных с наличием начальных знаний о взаимосвязях пользователей. В тоже время в своих работах^{7,8} авторы предлагают на основе метода дифференциации вершин и метода выделения частичного изоморфизма графов оценивать сходство структур объектов, заданных графовыми моделями. Подход, использующий указанные методы анализа графов, может быть применен при решении задачи идентификации сообществ, поскольку СС может быть представлена в виде социального графа.

Однако, для решения задачи определения сходства графов в рассмотренных работах^{7,8} требуется перебрать все подстановки сходства и выбрать среди них такую, которая соответствует максимальной общей части

сравниваемых графов. Таким образом, число перебираемых подстановок сходства при равном количестве вершин $N1=N2$ сравниваемых графов $G1$ и $G2$ составляет $N1!$, а при $N1 < N2$ увеличивается на число сочетаний из $N2$ по $N1$. Несмотря на то, что для сокращения переборных операций авторы предлагают оценивать перспективность каждой выбранной подстановки за счет количества удаленных ребер в процессе дифференциации и при необходимости прерывать данный процесс, выполнить все переборы в социальном графе, насчитывающем миллионы вершин, невозможно.

При этом модификация рассмотренных методов с учетом особенностей социальных сетей позволит уменьшить вычислительную сложность разрабатываемого алгоритма идентификации сообществ. С этой целью предлагается произвести дополнение указанных методов, которое заключается в применении начальной подстановки, исключающей переборные операции при формировании подстановки сходства, а также в присвоении текстовых атрибутов социальным вершинам и в использовании их для более точной и быстрой дифференциации.

Для решения задачи идентификации сообществ предлагается сформулировать ее, оперируя понятиями теории графов при описании СС.

Социальный граф — граф, узлы которого представляют социальные объекты (пользователей), а ребра — социальные связи между ними¹. В общем случае структура социального графа представляется в виде $G=(U, E)$, где U — множество вершин графа; E — множество ребер графа; $|U|=N$ — количество вершин в графе, $|E|=L$ — число связей (ребер). Социальный граф G задается матрицей смежности A размера $N \times N$, где N — число участников сети, $e_{ij} \in \{0, 1\}$ — элемент матрицы, показывающий наличие или отсутствие взаимосвязи между участниками u_i и u_j .

Учетная запись (далее — УЗ) представляет собой совокупность параметров пользователя, необходимых для его опознавания (аутентификации) и предоставления доступа к его личным данным и настройкам. УЗ обладает рядом **признаков**, позволяющих описывать различные характеристики пользователя в рамках СС: имя и фамилия, дата рождения, город проживания, социальные связи и т.д. УЗ пользователей обычно содержат очень подробную информацию об участнике СС.

Для описания вершины графа целесообразно использовать находящиеся в открытом доступе атрибутивные компоненты УЗ (далее — атрибуты), которые можно разделить на две категории: текстовые и структурные. Набор текстовых атрибутов, публикуемых пользователем для описания своей УЗ в социальной сети, описывается как вектор признаков $Atr(u_x^G) = \{id(u_x^G), name(u_x^G), sern(u_x^G), maid(u_x^G)\}$, где $u_x^G \in U$ обозначает x -го пользователя социального графа G , $x=1, 2, \dots, N$, $id(u_x^G)$ — уникальный идентификатор пользователя СС, $name(u_x^G)$ — имя пользователя, $sern(u_x^G)$ — фамилия пользователя, $maid(u_x^G)$ — вторая фамилия пользователя, если такая существует. Структурные атрибуты, представляющие собой систему взаимоотношений между пользователями СС, описываются как окрестность социальной вершины. **Окрестность** вершины $u_x \in U$ $O(G, u_x^G) = (O(G, u_x^G), EO(G, u_x^G))$ — подграф, порожденный этой вершиной и всеми смежными с ней в графе G^6 .

Рассмотренные параметры позволяют достаточно полно описать пользователя социальной сети:

$$u_x^G = \{O(G, u_x^G), Atr(u_x^G)\},$$

Через $G1=(U1, E1)$, $G2=(U2, E2)$, $Gm=(Um,$

$Em)$ определим различные социальные графы. Количество таких графов m определяется существующими в настоящее время СС. Вершины $u_a^{G1} \in U1$, $u_b^{G2} \in U2$ и $u_z^{Gm} \in Um$ обозначают пользователей в различных социальных сетях, при этом $a=1, 2, \dots, N1$; $b=1, 2, \dots, N2$; $z=1, 2, \dots, Nm$; $N1, N2, Nm$ — количество вершин графов $G1, G2$ и Gm соответственно.

Любой пользователь может одновременно являться участником нескольких СС. Следовательно, существует возможность формирования сообществ, состоящих из одних и тех же пользователей, но в различных СС. Под **сообществом** будем понимать группу пользователей, объединенную относительно некоторого исходного пользователя u_x^G , который обладает социальной взаимосвязью с каждым участником данной группы.

Таким образом, в процессе идентификации любого сообщества необходимо решить задачу поиска во всех существующих социальных графах подграфов, структурно схожих с окрестностью $O(G, u_x^G)$ исходного пользователя u_x^G :

$$\gamma: (O(G, u_x^G); \{G1, G2, \dots, Gm\}) \rightarrow \{O(G, u_x^G), S1, S2, \dots, Sm\},$$

где $S1, S2, \dots, Sm$ обозначают определенные в социальных графах $G1, G2, \dots, Gm$ сооб-

щества, обладающие максимальным структурным сходством с сообществом в исходном социальном графе.

В качестве решения предлагается модифицировать ранее предложенный метод^{7,8}, позволяющий оценивать сходство между окрестностью $O(G, u_x^G)$ и всеми вершинами социальными графами $\{G_1, G_2, \dots, G_m\}$ на основе метода дифференциации вершин (далее – ДВ) и метода выделения частичного изоморфизма (далее – ЧИ) социальных графов.

Задача определения сходства структур рассматривается как выделение в сравниваемых графах $O(G, u_x^G)$ и G_m наибольшей общей части — графа $GQ = (UQ, EQ)$. Для решения данной задачи в работах^{7,8} предлагается перебрать все возможные подстановки вершин исследуемых графов и в каждой из них определить число совместившихся ребер, образующих общую часть. Подстановка, образующая граф GQ с наибольшим числом ребер $|EQ|$, именуется **подстановкой сходства**, а сформированная на основе нее общая часть является наибольшей.

В качестве меры при оценке сходства структур графов используется параметр $\mu(u_x^G)$, который обозначает отношение количества ребер $|EQ|$ подграфа GQ к количеству ребер $|EO(G, ux^G)|$ окрестности $O(G, u_x^G)$:

$$\mu(u_x^G) = |EQ| / |EO(G, ux^G)|$$

Поскольку в работах^{7,8} описывается процесс выделения изоморфизма для самого общего случая, предлагается модифицировать метод путем введения нескольких положений, позволяющих конкретизировать задачу определения сходства с учетом особенностей СС.

Положение 1. Для исключения перебора всех возможных подстановок вершин при формировании наибольшей общей части вводится **начальная подстановка**, которая определяет всю дальнейшую подстановку сходства за счет применения метода ДВ.

Под **начальной подстановкой** (далее – НП) будем понимать две пары смежных вершин, для которых принимается условие взаимно однозначного соответствия (биективного отображения) с сохранением отношений смежности этих вершин, т.е. для пары вершин $u_1^G, u_2^G \in U$ и условно соответствующих им вершин $u_1^{G_m}, u_2^{G_m} \in U_m$ ($u_1^G \leftrightarrow u_1^{G_m}, u_2^G \leftrightarrow u_2^{G_m}$) имеет место $u_1^G, u_2^G \in E \leftrightarrow u_1^{G_m}, u_2^{G_m} \in E_m$. Символом « \sim » обозначается неупорядоченная пара вершин.

Положение 2. Для определения НП не-

обходимо использовать набор текстовых атрибутов $Atr(u_1^G)$, публикуемых пользователями для описания своих УЗ в СС, при формировании поискового запроса на список претендентов на соответствие.

Существующие поисковые сервисы СС позволяют сформировать запрос на поиск интересующего пользователя с использованием его имени и фамилии. Полученный в результате список наиболее релевантных претендентов на соответствие используется для формирования НП. Учитывая, что в качестве НП выступает пара смежных вершин $\{u_1^G, u_2^G\}$, то необходимость использования всех предложенных претендентов отсутствует. Для НП будет выбрана лишь та пара пользователей из двух списков, которая состоит в «дружеских» отношениях. Список претендентов в СС G_m на соответствие пользователю $u_1^G \in U$ будем обозначать как множество $PR^{G_m}(u_1^G)$.

Положение 3. Последовательность определения вершин для НП формируется с учетом количества претендентов $|PR^{G_m}(u_1^G)|$ и $|PR^{G_m}(u_2^G)|$ на соответствие пользователям u_1^G, u_2^G , что позволяет сократить количество проверяемых на смежность вершин.

В связи с тем, что среди предложенных претендентов может не оказаться отображения пользователя (пользователь не зарегистрирован в СС, в УЗ указаны ложные имя или фамилия) или может отсутствовать социальная связь в одной СС у пользователей, связанных «дружескими» отношениями в другой СС, возникает необходимость в последовательной проверке вершин для НП на выполнение условия биективности с сохранением смежности: $u_1^G, u_2^G \in E \leftrightarrow u_1^{G_m}, u_2^{G_m} \in E_m$. Для сокращения вычислительных затрат наиболее приоритетной является пара пользователей с минимальным количеством претендентов на соответствие в другой СС.

Положение 4. Для более быстрого и точного достижения полной дифференциации необходимо использовать не только структурные параметры вершин, но и назначенные текстовые атрибуты $\{\text{name}(u_1^G), \text{sern}(u_1^G), \text{maid}(u_1^G)\}$ и $\{\text{name}(u_1^{G_m}), \text{sern}(u_1^{G_m}), \text{maid}(u_1^{G_m})\}$.

Поиск интересующего пользователя по его имени и фамилии во всей социальной сети предоставляет несколько претендентов, что исключает возможность установления взаимно однозначного соответствия. Однако, если такой поиск происходит в достаточно небольшой области социального графа, локализованной относительно НП, то поиск

отображения для пользователя становится вполне достижимым. Поэтому, в случае установления уникального соответствия для текстовых атрибутов $\{name(u_i^G), sern(u_i^G), maid(u_i^G)\} \leftrightarrow \{name(u_i^{Gm}), sern(u_i^{Gm}), maid(u_i^{Gm})\}$ можно принимать условие соответствия самих вершин $u_i^G \leftrightarrow u_i^{Gm}$, что позволяет их дифференцировать относительно других.

Кроме того, рассматриваемая задача идентификации сообществ допускает разбиение на подзадачи, что позволяет применять технологии параллельных вычислений. Такой подход дает возможность увеличивать производительность системы, реализующей алгоритм идентификации.

Предложенные положения направлены на исключение переборных операций при определении подстановки сходства за счет формирования НП с учетом текстовых атрибутов в порядке, основанном на количестве претендентов на соответствие. Сформированная таким образом НП составляет минимальную общую часть в сравниваемых графах. Дальнейшие действия по определению наибольшей общей части происходят на основе ДВ как по структурным, так и по атрибутивным различиям. Сформированная максимальная общая часть позволяет оценить сходство структур сравниваемых графов в виде показателя ЧИ μ .

Таким образом, предложенная модификация методов ДВ и выделения ЧИ позволяет реализовать алгоритм, который для любого сообщества, образованного относительно определенного пользователя, проводит поиск его отображения в другой СС с расчетом показателя сходства μ . Учитывая значительные размеры графов, образованных существующими СС, введенные положения позволят значительно сократить вычислительные затраты и повысить производительность алгоритма идентификации.

В качестве реализации модифицированных методов ДВ и выделения ЧИ разработан прототип программного средства (далее – ПС) идентификации сообществ, архитектура которого представлена на рис. 1.

В качестве исходных данных для работы ПС выступает идентификатор исходного пользователя, например, в социальной сети «Одноклассники». Этот параметр поступает в модуль управления процессом формирования наибольшей общей части. Данный модуль является центральным и отвечает за последовательное выполнение процедур по

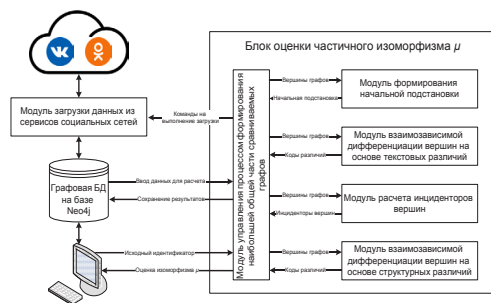


Рис. 1. Архитектура ПС идентификации сообществ

вычислению показателя ЧИ. Для актуализации процесса идентификации загрузка данных о состоянии сервисов социальных сетей происходит в момент запуска ПС, начальные сведения о пользователях и структуре их социальных связей отсутствуют. За данный процесс отвечает модуль загрузки данных из сервисов социальных сетей. На текущий момент взаимодействие организовано посредством API-функций с социальными сетями «Одноклассники» и «ВКонтакте». Для хранения собранных данных наилучшим образом подходит структура графовых баз данных, основными элементами которых являются узлы и связи между ними (ребра). К каждому такому узлу или ребру существует возможность добавления признаков, определяемых атрибутами УЗ пользователя. По мнению ряда авторов⁹, для задач с естественной графовой структурой данных графовые СУБД могут существенно превосходить реляционные по производительности, а так же иметь преимущества в наглядности представления и внесении изменений в схему БД. Для реализации задачи хранения выбрана графовая БД Neo4j¹⁰.

Процедура расчета показателя ЧИ начинается с формирования окрестности $O(G, u_1^{GOK})$ исходного пользователя u_1^{GOK} в социальном графе сети «Одноклассники» GOK. Одновременно с загрузкой данных о пользователях и их социальных связях происходит оценка количества претендентов на соответствие $|PR^{GVK}(u_1^{GOK})|$. Для более точной дифференциации на основе структурных различий производится формирование окрестности всех вершин, смежных с искомым пользователем. Последовательность формирования НП начинается с пары смежных вершин, имеющих минимальное количество претендентов на соответствие. Данные вершины, а точнее их текстовые атрибуты «Имя», «Фамилия»

и «Вторая фамилия», определяют поисковый запрос к сервису «ВКонтакте». Предложенные в ответе пользователи проверяются на наличие между ними социальной связи. В случае успеха, найденное ребро и инцидентные с ним вершины формируют НП, в противном случае выбирается следующая по количеству претендентов пара смежных вершин. Для выбранных вершин производится обновление информации: загружаются данные о всех «друзьях» пользователя с учетом текстовых атрибутов и структуры социальных связей. Дальнейшая работа по построению наибольшей общей части оцениваемых графов $O(G, u_i^{GOK})$ и графа социальной сети «ВКонтакте» GVK строится на основе взаимозависимой дифференциации вершин по текстовым и структурным различиям.

В качестве демонстрации на рис. 2, 3 и 4 представлены начальные этапы работы ПС идентификации сообщества, образованного пользователем «Николай Николаев» u_1^{GOK} , в СС «ВКонтакте» GVK . Так на рис. 2 сформирована окрестность для исходного пользователя в графе GOK СС «Одноклассники». Поскольку еще ни одна вершина не дифференцирована относительно других, коды различия у них равны 1.

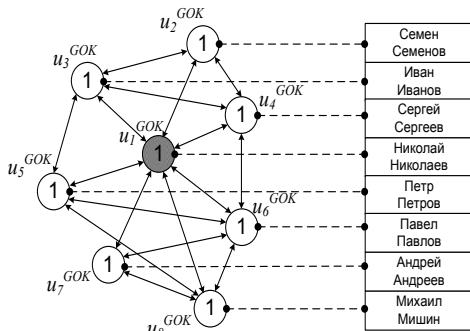


Рис. 2. Формирование окрестности исходного пользователя u_1^{GOK}

Допустим, что в представленном примере пара пользователей «Сергей Сергеев» u_4^{GOK} и «Павел Павлов» u_6^{GOK} имеет минимальное количество претендентов на соответствие в СС «ВКонтакте» (рис. 3). Следовательно, указанные пользователи в первую очередь используются для формирования НП. На основании имени и фамилии данных пользователей формируются поисковые запросы в СС «ВКонтакте». В результате поисковый сервис предложил четырех претендентов на соответствие для пользователя «Сергей Сергеев»

и трех для пользователя «Павел Павлов». Среди всех предложенных претендентов имеется пара пользователей u_4^{GVK} и u_6^{GVK} , состоящих в «дружеских» отношениях. Именно эти вершины и образуют НП. Для данных пользователей производится обновление информации о социальных связях, а также происходит присвоение вершинам u_4^{GOK} и u_4^{GVK} кода различия 2, а вершинам u_6^{GOK} и u_6^{GVK} – 3 (рис. 4). Далее сформированный подграф GVK участвует в процедуре взаимозависимой дифференциации вершин.

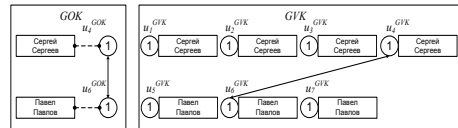


Рис. 3. Формирование НП

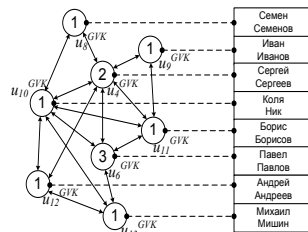


Рис. 4. Обновление информации о вершинах, составляющих НП

Процесс дифференциации представляется в виде динамической системы, функционирование которой можно представить последовательностью смены состояний $D^0, D^1, \dots, D^k, \dots, D^n = D$, от начального D^0 до конечного $D^n = D$, соответствующего полной дифференциации вершин.

Процесс взаимозависимой дифференциации начинается с анализа текстовых атрибутов вершин, входящих в подграф социальной сети «ВКонтакте», сформированный на основе НП. Исходное состояние системы представляется в виде векторов $D^k(GOK) = \{d_j^k(GOK)\}$ и $D^k(GVK) = \{d_j^k(GVK)\}$, $i = 1, 2, \dots, NOK, j = 1, 2, \dots, NVK$, NOK и NVK – количество вершин графов GOK и GVK соответственно, $d_j^k(GOK)$ и $d_j^k(GVK)$ обозначают коды различия (цифровое обозначение) вершин графов GOK и GVK . Для перехода системы из $D^k(GOK)$ и $D^k(GVK)$ в $D^{k+1}(GOK)$ и $D^{k+1}(GVK)$ производится анализ атрибутов «Имя», «Фамилия» или «Вторая фамилия»: в случае их уникального соответствия у двух вершин из различных социальных графов GOK и GVK , они дифференцируются, им присваиваются коды различия $d_j^k(GOK)$ и $d_j^k(GVK)$ (рис. 5), после чего происходит обновление

информации о них. Добавленные в результате обновления вершины также проверяются на соответствие по текстовым атрибутам. Цикл добавления и проверки продолжается до тех пор, пока возможна дифференциация вершин.

Как видно из примера на рис. 5, пользователей, у которых выполняется соответствие имен и фамилий, можно дифференцировать относительно других и присвоить им коды различий в графах GOK и GVK : «Семен Семенов» – код 4, «Иван Иванов» – код 5, «Андрей Андреев» – код 6, «Михаил Мишин» – код 7. Для пользователей «Николай Николаев», «Петр Петров», «Коля Ник» и «Борис Борисов» однозначное соответствие по текстовым атрибутам не установлено.

При дальнейшей невозможности дифференцировать вершины по текстовым атрибутам выполняется дифференциация на основе структурных различий.

системы в дискретном времени позволяет улавливать разнообразие отношений между кодами различия и на этой основе дифференцировать вершины. Для перехода системы из $D^k(GOK)$ и $D^k(GVK)$ в $D^{k+1}(GOK)$ и $D^{k+1}(GVK)$ автоматы под управлением блока дифференциации выполняют единый набор действий:

- обмен состояниями $d_i^k(GOK)$ и $d_j^k(GVK)$ между всеми автоматами вершин $u_i GOK$ и $u_j GVK$;
- формирование инцидентов всех вершин $F(d_i^k(GOK))$ и $F(d_j^k(GVK))$ с состоянием $d_i^k(GOK)$ и $d_j^k(GVK)$;
- передача в блок дифференциации инцидентов $F(d_i^k(GOK))$, $F(d_j^k(GVK))$ и состояний $d_i^k(GOK)$, $d_j^k(GVK)$ в форме $d_i^k:F(d_i^k(GOK))$, $d_j^k:F(d_j^k(GVK))$.

Инцидентор⁸ $F(u_i)$ — совокупность вершин $u_j \in U, i \neq j$, которые в графе G связаны ребрами $e_{ij} = (u_i, u_j)$, инцидентными вершине u_i . В блоке дифференциации на основе анализа

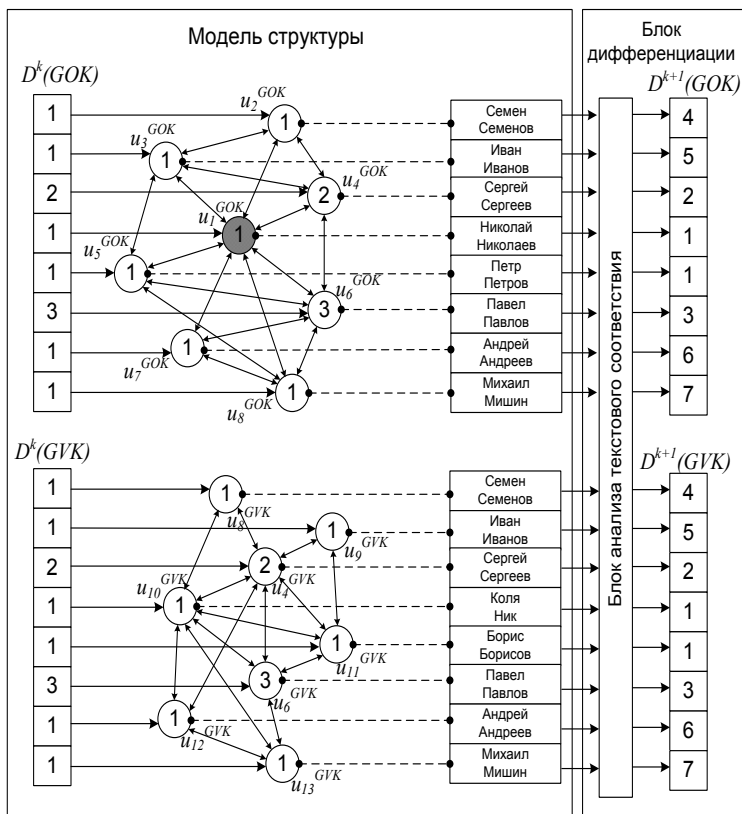


Рис. 5. Дифференциация вершин на основе атрибутивных различий

Метод взаимозависимой дифференциации вершин на основе структурных различий описывается с помощью модели структуры графа, представленной сетью автоматов (рис. 6). Функционирование такой динамической

векторов $\{d_i^k:F(d_i^k(GOK))\}$ и $\{d_j^k:F(d_j^k(GVK))\}$ происходит расчет показателей структурного соответствия вершин, после чего определяется новое состояние системы $D^{k+1}(GOK)$ и $D^{k+1}(GVK)$: пара вершин, имеющая наиболь-

ший показатель, дифференцируется с присвоением кодов различия. Формирование инцидентов, а также расчет показателей структурного соответствия производится только для недифференцированных вершин с единичными значениями кодов различия. Кроме того, при анализе инцидентов $\{d_i^k:F(d_i^k(GOK))\}$ и $\{d_j^k:F(d_j^k(GVK))\}$ учитываются только не единичные коды. После дифференциации вершин происходит обновление информации о них.

Процесс построения наибольшей общей части продолжается до полной дифференциации вершин, смежных с искомым пользователем.

Результатом работы разработанного ПС является формирование наибольшей общей части для окрестности исходного пользователя u_x^{GOK} в СС «Одноклассники» $O(GOK, u_x^{GOK})$ и СС «ВКонтакте» GVK , а также расчет показателя ЧИ $\mu(u_x^{GOK})$ указанных социальных графов.

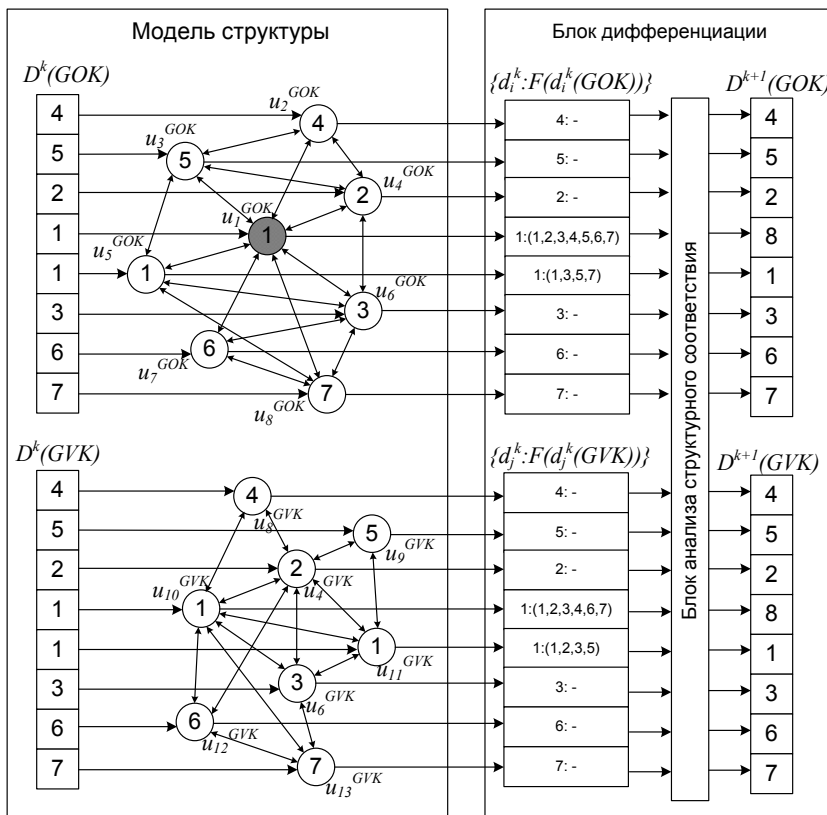


Рис. 6. Дифференциация вершин на основе структурных различий

Так, в примере на рис. 6 пользователи «Николай Николаев» и «Коля Ник» описывают одного и того же человека, но описание его УЗ в различных СС отличается. Однако за счет анализа векторов $\{d_i^k:F(d_i^k(GOK))\}$ и $\{d_j^k:F(d_j^k(GVK))\}$ можно сделать следующий вывод о показателях структурного соответствия: для пользователей u_1^{GOK} и u_{10}^{GVK} показатель составляет 5, для u_1^{GOK} и u_{11}^{GVK} – 3, для u_5^{GOK} и u_{10}^{GVK} – 2, для u_5^{GOK} и u_{11}^{GVK} – 2. Наибольшее значение у пользователей u_1^{GOK} и u_{10}^{GVK} . Таким образом, пользователи «Николай Николаев» и «Коля Ник» дифференцируются относительно других вершин и им присваивается код различия 8 в графах GOK и GVK .

Таким образом, в работе предложен вариант модификации метода дифференциации вершин и метода выделения частичного изоморфизма за счет применения начальной подстановки и использования текстовых атрибутов социальных вершин для уменьшения вычислительной сложности алгоритма идентификации сообществ в социальных сетях. Такой подход создает возможность определения сообществ, организованных одними и теми же пользователями в различных социальных сетях, за счет анализа находящейся в открытом доступе информации о пользователях социальных сетей и их социальных связях.

Примечания

1. Безуглая М.В., Патрушева О.М., Синадский Н.И., Сушков П.В. Расчет показателя сходства учетных записей пользователей социальных сетей на основе анализа атрибутов и структуры социальных связей // Безопасность информационного пространства: сборник трудов XIV Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых / сост. А.А. Захаров. — Тюмень: Издательство Тюменского государственного университета, 2016. — С. 19-23.
2. Форман Дж. Много цифр: Анализ больших данных при помощи Excel / Джон Форман; Пер. с англ. А. Соколовой. — М.: Альпина Паблишер, 2016. — 461 с.
3. Veldman I. Matching Profiles from Social Network Sites. Master's thesis, University of Twente, 2009.
4. Бартунов С.О., Коршунов А.В. Идентификация пользователей социальных сетей в Интернет на основе социальных связей // Доклады Всероссийской научной конференции "Анализ изображений, сетей и текстов". — АИСТ'2012. Екатеринбург, 16-18 марта 2012 г.
5. Vosecky J., Hong D., Shen V.Y. User identification across multiple social networks. In Proc. of First International Conference on Networked Digital Technologies, 2009
6. Зыков А.А. Основы теории графов. — М.: Изд-во «КомКнига», 2004. — 644 с.
7. Погребной Ан.В., Погребной В.К. Метод дифференциации вершин графа и решение проблемы изоморфизма // Известия Томского политехнического университета. — 2015. — Т. 326. — № 6. — С. 34-45.
8. Погребной А.В. Метод определения сходства структур графов на основе выделения частично-го изоморфизма в задачах геоинформатики // Известия Томского политехнического университета, 2015. — Т. 326. — № 11. — С. 56-66.
9. Robinson I., Webber J., Eifrem E. Graph Databases. — O'Reilly Media, Incorporated, 2013. — 178 p.
10. Neo4j, the world's leading graph database // Neo4j Graph Database. URL: <https://neo4j.com> (дата обращения 05.07.2017)

Павел Владимирович СУШКОВ, аспирант кафедры алгебры и фундаментальной информатики Института естественных наук и математики УрФУ им. первого Президента России Б.Н. Ельцина; 620002, г. Екатеринбург, ул. Мира, 19, pavelsu@e1.ru, (343) 375-95-40.

Николай Игоревич СИНАДСКИЙ, к.т.н., доцент, доцент кафедры алгебры и фундаментальной информатики Института естественных наук и математики УрФУ им. первого Президента России Б.Н. Ельцина; 620002, г. Екатеринбург, ул. Мира, 19, nickis@e1.ru, (343) 375-95-40.

Pavel SUSHKOV, Postgraduate at the Department of Algebra and Fundamental Informatics, Institute of Natural Sciences and Mathematics, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Russian Federation, Yekaterinburg, Mira str., 19 pavelsu@e1.ru, (343) 375-95-40.

Nikolay SINADSKY, Candidate of Engineering Sciences, Docent, Associate Professor at the Department of Algebra and Fundamental Informatics, Institute of of Natural Sciences and Mathematics, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Russian Federation, Yekaterinburg, Mira str., 19, nickis@e1.ru, (343) 375-95-40.