

ТЕОРЕТИКО-МНОЖЕСТВЕННОЕ МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СОСТОЯНИЯ СОЦИАЛЬНОЙ СРЕДЫ

Аннотация.

Актуальность и цели. Актуальность работы заключается в решении задач информационной поддержки управления социально-экономическими системами с использованием больших объемов неструктурированных данных сети Интернет. Целью работы является математическая формализация взаимодействующих объектов социальной среды для последующей формализации методов решения задачи социального профилирования в рамках создания системы на основе технологий Больших данных.

Материалы и методы. Материалами исследования выступают данные социальной среды и ее составных элементов. Социальная среда представлена совокупностью взаимодействующих персональных и групповых социальных профилей (СП). Персональный СП детализирован до статической и динамической частей. Модели социальных профилей и среды используют методологические основы теории множеств, семантических сетей и метаграфов.

Результаты. Формализованы понятия социального явления, персонального социального профиля человека и группы людей. Разработаны математические модели для описания социальных профилей личности и группы людей, а также основанной на них социальной среды. Они могут найти практическое применение в задачах социального и экономического управления.

Выводы. Полученные результаты в совокупности позволяют систематизировать процессы сбора и анализа данных о социальной среде, обеспечивая возможность объединения СП людей в группы и проведения над ними прикладных научных исследований.

Ключевые слова: социальный профиль, социальная среда, социальные явления, теория множеств, теория графов, неструктурированные данные, Big Data.

А. Ю. Timonin

MATHEMATICAL MODELING OF THE SOCIAL ENVIRONMENT STATE BASED ON SET THEORY

Abstract.

Background. The work relevance is to solve the information support problems for the socio-economic system management. That uses large amounts of unstructured WWW data. The work purpose is to formalize interacting social environment objects. It is necessary to the subsequent formalization of the social profiling methods based on a Big Data system.

Materials and methods. The research materials are the information of the social environment and its constituent elements. Social environment is represented by a set

of interacting personal and group social profiles (SP). Personal SP is detailed to static and dynamic parts. Social profiles and environments models use the methodological foundations of set theory, semantic networks and metagraphs.

Results. The concepts of social fact (SF), social profile of a person (PSP or SP) and a group of people (GSP) are formalized. Mathematical models have been developed to describe the social profiles of an individual and a group of people, as well as the social environment based on them. They can find practical application in the tasks of social and economic management.

Conclusions. The aggregated results make it possible to systematize the processes of collecting and analyzing the social environment data. It helps to uniting the social profiles into groups and conducting applied research on them.

Keywords: social profile, social environment, social phenomena, set theory, graph theory, unstructured data, Big Data.

Введение

«Социальная среда представляет собой совокупность материальных, экономических, социальных, политических и духовных условий существования, формирования и деятельности индивидов и социальных групп» [1]. Традиционно выделяют социальную макросреду и микросреду. Макросреда – общечеловеческая, включает в себя общественные институты, культуру и экономику. Ее исследование представляет интерес при анализе глобальных социальных закономерностей и может использоваться в процессах социально-экономического управления на уровнях региона, страны или всего мира. Микросреда описывает непосредственное окружение человека – семью, работу, учебу и т.д. Задачи, связанные с анализом микросреды, в настоящее время становятся все более популярными ввиду доступности подробной статистической и открытой биографической информации об отдельных людях в сети Интернет и возможности более глубокого изучения процессов социального взаимодействия.

Исследование социальной среды вызывает ряд трудностей у аналитиков. В первую очередь это связано со сбором и обработкой большого числа слабоструктурированных текстовых, геопространственных, мультимедийных и статистических данных, а также сложностью всеохватывающей интерпретации ввиду наличия запутанных связей между отдельными элементами социальной среды. Другой проблемой является потеря актуальности исходных данных с течением времени, что вызвано высокой динамикой социальных коммуникаций и требует периодического обновления картины социальной среды. Такая среда называется социальным континуумом. Нельзя исключать влияние внешних воздействий и человеческого фактора. Восприятие человека субъективно и ограничено возможностями его мозга. Поэтому при исследовании такой сложной системы, как социальная среда, возможны такие проблемы, как противоречие получаемых результатов, ошибочное истолкование отдельных высказываний, неучет неявных взаимосвязей, принятие решений не на основе объективных выводов, а на личном субъективном опыте и т.д.

Для упрощения взаимодействия лиц, принимающих решения, с информацией, описывающей социальную среду, в данной работе предлагается ее разделить на логически связанные составные части. Перефразируем определение, данное выше: социальная среда представляет собой динамическую совокупность отдельных персон, социальных групп и явлений. При таком раз-

делении становится возможным проведение исследований отдельных составляющих среды и определение явно и неявно связанных социальных структур, в частности социальных профилей людей.

Персональный социальный профиль (СП) [2] – это множество, состоящее из информации, которая способна тем или иным образом охарактеризовать социальные свойства человека, причем она должна быть наглядно структурирована как для удобства человеческого восприятия, так и для последующей автоматизированной обработки. Из этого определения следует, что СП состоит из структурированных социализированных (определяющих конкретного человека) данных, представленных в виде неоднородной семантической сети [3]. Помимо модели социальной среды, на математической модели СП основываются методы решения задачи построения СП [4]. СП личности – простейший случай, в котором имеется единственный центральный элемент (уникальный идентификатор рассматриваемой персоны). Как следствие этого – у личного профиля наименьшее число неявных взаимоотношений между его составными частями и более понятная структура. И тем меньшее влияние оказывает человеческий фактор на результаты анализа.

Социальный профиль может быть групповым (ГСП), учитывающим более общие социальные тенденции, такие как общие интересы, межличностные взаимодействия, поведенческие закономерности. Он представляется в виде совокупности персональных профилей, объединенных одним или несколькими критериями, имеет произвольное количество общих свойств. Социальные группы в данном случае по своей природе могут являться конкретными, явно обозначенными и абстрактными, неявными, которые определяются только по ряду соответствующих им признаков. Таким образом, можно сделать вывод, что ГСП представляют большой интерес в задачах управления.

При составлении социальных профилей возникает ситуация, когда требуется описать какой-либо факт, важный для целостного представления личности, группы или социальной среды. Такие факты в рамках исследования состояний социальной среды могут оказаться центральными объектами. Будем называть их социальными явлениями (СЯ). Они представляют собой явную или косвенную информацию о событиях реального мира, идеях, результатах взаимодействия личностей и их групп между собой в социальной среде. Например, в задачах управления СЯ могут быть представлены общественным мнением, новаторской деятельностью, законодательными инициативами, мероприятиями и т.д. При этом они обладают отличительными свойствами, такими как время и место возникновения, множество вовлеченных участников, множество сопутствующих объектов реального мира, список связанных понятий, дающих описательную характеристику данного явления, список источников, подтверждающих его существование, связи с другими СЯ. Их можно группировать по совпадающим свойствам и делать сравнительные выборки.

Для обеспечения возможности автоматизированной компьютерной обработки представление социальной среды должно быть формализовано, учитывать все ключевые свойства исследуемых людей с возможностью использования больших неструктурированных данных в процессах управления. Требуется иметь возможность обращения к любой связанной информации,

доступной в ходе сбора исходного материала СП, разделять их составные части по природе и ролям с выделением списка уникальных атрибутов в анализе персоны.

1. Обобщенное моделирование социального профиля человека

СП личности состоит из множества взаимосвязанных статических и динамических характеристик деятельности человека в обществе и других объектов социальной среды, неким образом связанных с рассматриваемой персонной. Они представляют собой атомарные объекты – актанты (понятия, события, свойства, процессы) и предикаты. Связь определяется либо фактически, либо более детально с учетом интонационных окрасок, весовых коэффициентов и т.д. Степень важности каждой социальной характеристики зависит как от расстояния до центрального объекта, являющегося отображением рассматриваемого человека, так и от внешних факторов, определяющих вес данной характеристики. На основе анализа тональности отдельных элементов социального профиля можно сделать вывод о настроении исследуемой персоны, является ли оно позитивным или негативным.

Социальный профиль личности в общем случае можно описать следующей формулой:

$$PSP = S(X, v) \cup R(Y, u) \cup Q(m+n), \quad (1)$$

где $S = \{S_1, \dots, S_m\}$ – множество социальных объектов рассматриваемой персоны (темы, события, персоны и т.д.); $R = \{R_1, \dots, R_n\}$ – множество социальных связей между социальными объектами рассматриваемой персоны; m, n – количество социальных объектов и связей соответственно, причем $m-1 \leq n$; $Q(m+n) = \{Q_1, \dots, Q_{m+n}\}$ – множество «интонаций» социальных характеристик, на основе которых строится комплексная оценка настроения рассматриваемой персоны, зависящая от количества социальных характеристик, представляет собой нечеткую величину, причем $Q \in [-1; 1]$; v, u – весовые коэффициенты социальных объектов и связей соответственно, представляют собой нечеткие величины; X, Y – множества свойств социальных объектов и связей соответственно, причем

$$\begin{cases} X = AO \times SO, \\ Y = AE \times SE, \end{cases} \quad (2)$$

где $AO = \{AO_1, \dots, AO_f\}$ – множество атрибутов (свойств), связанных с социальным объектом; $AO_i = \{O_1, \dots, O_{fx}\}$ – множество значений свойства для единичного определения социального объекта; SO – множество синонимов и повторений определения социального объекта в исходной информации; $AE = \{AE_1, \dots, AE_g\}$ – множество атрибутов (свойств), связанных с социальным отношением; $AE_i = \{E_1, \dots, E_{gy}\}$ – множество значений свойства для единичного определения социального отношения; SE – множество синонимов и повторений определения социального отношения в исходной информации; f, g – количество свойств у социальных объектов и связей рассматриваемой

персоны соответственно; fx, gy – количество вхождений свойства (в виде синонимов в том числе) для социальных объектов и связей соответственно.

Социальный профиль группы состоит из совокупности личных профилей, взятых по ряду совпадающих характеристик, таких как статус, интересы, связи с общими объектами и т.д. Примерами таких групп являются: пенсионеры по старости, рыболовы-любители, множество людей, живших когда-либо по определенному адресу. Количество внутренних связей в коллективном СП больше суммарного числа связей личных СП, так как к рассмотрению добавляются взаимосвязи одной личности с другой. Как следствие этого, социальный профиль группы будет содержать большее число социальных объектов, а значит и иметь собственные уникальные свойства, которых нет ни у одного включенного в него профиля личности.

Соответственно СП группы лиц в общем случае будет представляться в следующем виде:

$$GSP = GSP' \cup PSP \cup GS(X', gv) \cup GR(Y', gu) \cup Q(GSP', PSP, m' + n'), \quad (3)$$

где $PSP = \{PSP_1, \dots, PSP_t\}$ – множество социальных профилей членов группы; t – количество личных СП в группе; $GSP' = \{GSP'_1, \dots, GSP'_{tt}\}$ – множество подгрупп, входящих в группу; tt – количество подгрупп в группе; $GS = \{GS_1, \dots, GS_m\}$ – множество социальных объектов, связанных непосредственно с группой; $GR = \{GR_1, \dots, GR_n\}$ – множество непосредственных социальных связей группы; m', n' – количество социальных объектов и связей группы соответственно, причем $m' - 1 \leq n'$; $Q = Q(m' + n') \cup Q(PSP) \cup Q(GSP')$ – множество «интонаций» социальных характеристик группы и входящих в нее подгрупп и людей, на основе чего строится комплексная оценка настроения рассматриваемой группы (4):

$$Q = \frac{\sum_{l=1}^{t+tt} Q_l}{t + tt}, \quad (4)$$

представляющая собой нечеткую величину, причем $Q \in [-1; 1]$; gv, gu – весовые коэффициенты социальных объектов и связей группы соответственно, представляют собой нечеткие величины; X', Y' – матрицы свойств социальных объектов и связей группы соответственно;

$$\begin{cases} X' = AO' \times SO', \\ Y' = AE' \times SE', \end{cases} \quad (5)$$

где $AO' = \{AO'_1, \dots, AO'_{f'}\}$ – множество атрибутов (свойств), связанных с социальным объектом, принадлежащим рассматриваемой группе; $AO'_i = \{O'_1, \dots, O'_{f'_{x'x'}}$ – множество значений свойства для единичного определения социального объекта, принадлежащего рассматриваемой группе; SO' – множество синонимов и повторений определения социального объекта, принадлежащего рассматриваемой группе, в исходной информации; $AE' = \{AE'_1, \dots, AE'_{g'g'}$ – множество атрибутов (свойств), связанных

с социальным отношением, принадлежащим рассматриваемой группе; $AE'_i = \{E'_{1'}, \dots, E'_{g'_{y'}}\}$ – множество значений свойства для единичного определения социального отношения, принадлежащего рассматриваемой группе; SE' – множество синонимов и повторений определения социального отношения, принадлежащего рассматриваемой группе, в исходной информации; f' , g' – количество свойств у социальных объектов и связей группы соответственно; fx' , gy' – количество вхождений свойств (в виде синонимов в том числе) для социальных объектов и связей группы соответственно.

В коллективных СП можно найти скрытые зависимости. Если у одного человека из группы не определено какое-либо свойство, а у остальных оно приблизительно одинаково, то с высокой вероятностью можно сделать вывод, что у первого – то же свойство. Например: возраст студентов из одного потока, факт наличия научных работ и общая область интересов у участников конференции, наличие транспортного средства у членов гаражного кооператива и т.д.

2. Разработка концептуальной модели СП для аналитической системы на основе Big Data

В рамках разработки методов функционирования системы социального профилирования, основанной на принципах Big Data, к проектированию СП предъявляются особые требования. Необходимо обеспечить использование преимуществ подхода Больших данных по параллельной обработке неструктурированных данных, при этом ядро социального профиля должно оставаться структурированным. Требуется привести разгруппирование социальных характеристик по смысловым категориям и степени значимости в процессе идентификации личности. Будем учитывать, что вся исходная информация сводится к текстовому виду как наиболее полному и понятному с точки зрения представления семантики. Тогда комплексная модель СП примет вид

$$SP = \{PP, PG, FACT, PDYN\}, \quad (6)$$

где PP – статическая часть СП личности; PG – статическая часть СП группы; FT – описание связанных СЯ; $PDYN$ – динамическая часть. Покажем ее на рис. 1.

Для объективной интерпретации больших объемов исходной информации, учета равнозначимых выражений и уменьшения влияния человеческого фактора на конечные результаты построения СП необходимы четко обозначенные определения и категории, применяемые к составным социальным характеристикам. Их множества могут быть представлены в виде как динамических дополняемых справочников, так и готовых тезаурусов (толковые, отраслевые, многоязычные словари).

Таким образом, перейдем к подробному рассмотрению социального профиля человека с учетом разделения его данных на статическую и динамическую части:

$$PP = \{ppid, PNAME, PDATE, PLOC, PLIN, PGLINK, PTLINK, PDLINK, PSOURCE, TESLINK\}, \quad (7)$$

где $ppid$ – уникальный числовой идентификатор СП, необходимый для их различения; $PNAME = \{FNAME, LNAME, PATR, ALIAS\}$ – множество имен рассматриваемой персоны (ФИО, псевдонимы, никнеймы); $PDATE$ – множество ключевых дат жизни персоны; $PLOC = \{pcoord, PNAMELOC, PLOCDATE, PLOCNOTE \mid PLOCDATE \subset PDATE\}$ – множество связанных геолокаций, включающее координаты местоположений, их названия, даты и комментарии; $PLINK = \{ppid \mid \exists x, y \in PP : x \cap y \neq \emptyset, ppid \langle x \rangle \neq ppid \langle y \rangle\}$ – множество ссылок на другие СП, с которыми установлены связи; $PGLINK = \{ppid, pgid, role, matter \mid \exists x \in PP, \exists y \in PG : ppid \in x, pgid \in y, x \cap y \neq \emptyset\}$ – ссылки на множество связанных социальных групп, учитывающие социальные роли персоны и сущность отношений с группой; $PTLINK = \{ppid, idtopic \mid \exists x \in PP, \exists y \in FACT : ppid \in x, idtopic \in y, x \cap y \neq \emptyset\}$ – множество ссылок на наиболее значимые для персоны социальные явления и характеристики; $PDLINK = \{ppid, dynid \mid \exists x \in PP, \exists y \in PDYN : ppid \in x, dynid \in y, x \cap y \neq \emptyset\}$ – множество ссылок на объекты динамической части с наиболее устойчивыми связями; $PSOURCE = \{ppid, sourcelink, sourcename, sourceget, x \mid x \in PP \cap SOURCE\}$ – сведения об источнике исходной информации с указанием его ссылки, названия и даты получения; $TESLINK$ – ссылки на словари.

Динамический контент непосредственно является частью СП, но в информационную карту не входит, а рассматривается как отдельное множество. Учет динамических характеристик СП позволяет проектировать адаптивные программные инструменты [5] для ЛПР:

$$PDYN = \{dynid, DATATYPE, DSERVICE, DNOTE, DTAG, PDLINK, DGLINK, DTLINK, DSOURCE, TESLINK\}, \quad (8)$$

где $dynid$ – уникальный числовой идентификатор динамического социального объекта; $DATATYPE = \{\text{текстовый, аудио, видео, графика, геоданные, бинарный}\}$ – тип данных; $DSERVICE = \{servicevalue, servicetype\}$ – сервисная информация (размер исходных данных, html теги и т.д.); $DNOTE$ – комментарии, описание динамического объекта; $DTAG = \{\text{Существо, Организация, Адрес, Телефон, Электронная почта, Деятельность, Достижение, Специализация, ...}\}$ – множество тематических тегов; $PDLINK$ – множество ссылок на информационную карту персонального СП; $DGLINK = \{dynid, pgid \mid \exists x \in PDYN, \exists y \in PG : dynid \in x, pgid \in y, x \cap y \neq \emptyset\}$ – множество ссылок на групповые СП; $DTLINK = \{dynid, idtopic \mid \exists x \in PDYN, \exists y \in FACT : dynid \in x, idtopic \in y, x \cap y \neq \emptyset\}$ – множество ссылок на социальные явления; $DSOURCE = \{dynid, sourcelink, sourcename, sourceget, x \mid x \in PDYN \cap SOURCE\}$ – сведения об источнике исходной информации; $TESLINK$ – ссылки на словари.

Факт (тема или социальное явление) является крупной социальной характеристикой, имеющей собственный набор уникальных атрибутов. При анализе состояния социальной среды возможна ситуация, когда СЯ оказывается центральным элементом среды по некоторым критериям.

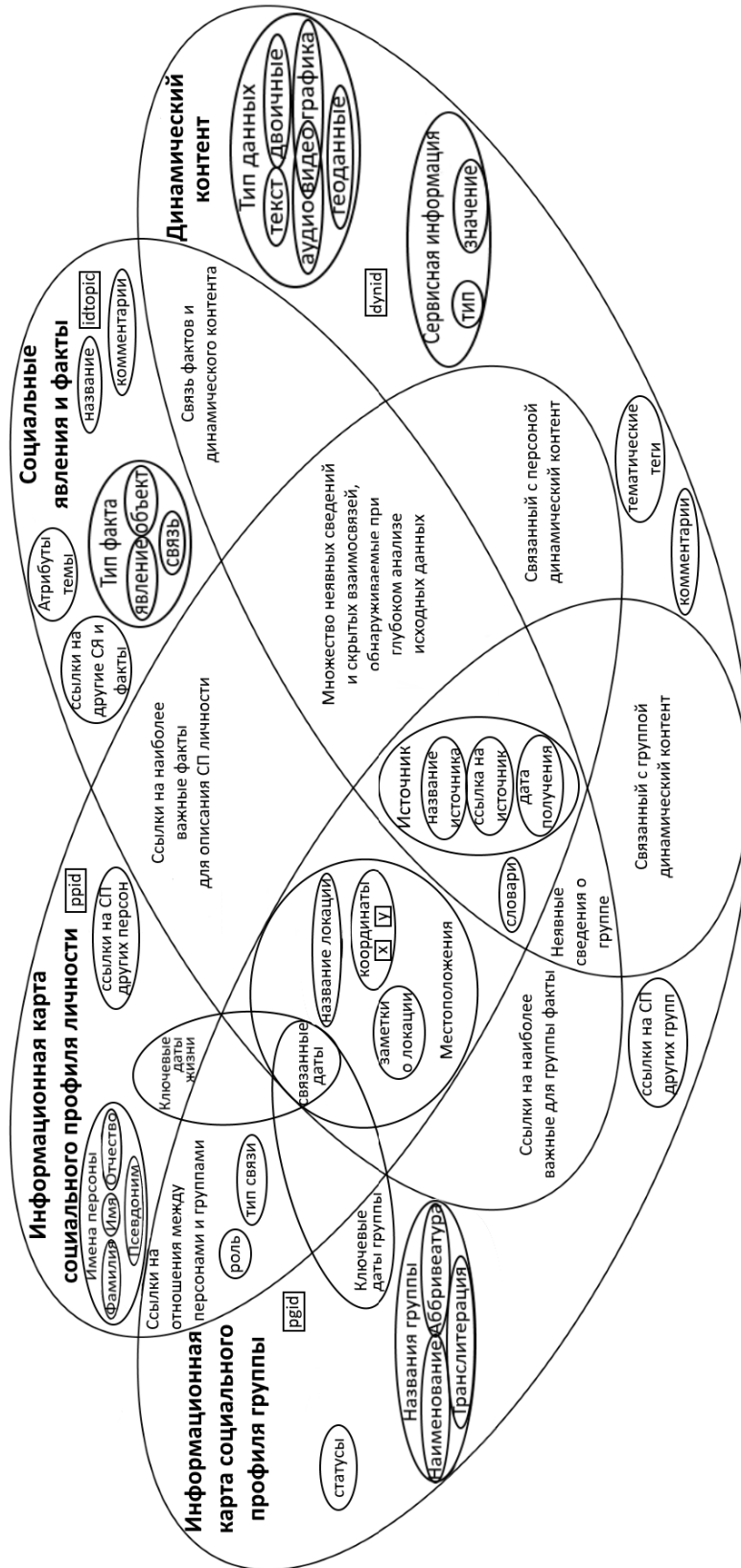


Рис. 1. Представление составных множеств социальных профилей на диаграмме Венна – Эйлера

Опишем его представление на теоретико-множественном уровне:

$$FACT = \{idtopic, TNAME, TFEATURE, TTYPE, TNOTE, TLINK, PTLINK, GTLINK, DTLINK, TSOURCE, TESLINK\}, \quad (9)$$

где *idtopic* – уникальный идентификатор факта; *TNAME* – множество наименований (ключевых слов) факта; *TFEATURE* – множество атрибутов, свойств факта; *TTYPE* = {социальное явление, социальный объект, социальная связь} – тип факта; *TNOTE* – комментарии к описанию факта; $TLINK = \{idtopic \mid \exists x, y \in FACT : x \cap y \neq 0, idtopic \langle x \rangle \neq idtopic \langle y \rangle\}$ – множество ссылок на другие факты; *PTLINK* – множество ссылок на связанные СП личностей; $GTLINK = \{pgid, idtopic \mid \exists x \in PG, \exists y \in FACT : pgid \in x, idtopic \in y, x \cap y \neq 0\}$ – множество ссылок на связанные групповые СП; *DTLINK* – множество ссылок на динамический контент; $TSOURCE = \{idtopic, sourcelink, sourcename, sourceget, x \mid x \in FACT \cap SOURCE\}$ – сведения об источнике исходной информации факта; *TESLINK* – ссылки на словари.

Групповой СП состоит из профилей его членов и собственных характеристик. Представление социальных групп абстрактно, они могут быть не только явно определенными, но и чисто формальными в рамках прикладных исследований. Их параметры могут кардинально меняться со временем, поэтому персональные СП не входят в них целиком, а имеют отдельный связывающий элемент. Модель социального профиля группы примет следующий вид:

$$PG = \{pgid, PGNAME, GTAG, PGDATE, PGLOC, GLINK, PGLINK, DGLINK, GTLINK, GSOURCE, TESLINK\}, \quad (10)$$

где *pgid* – уникальный идентификатор группы; *PGNAME* = {*RLNAME*, *ABBR*, *TRNAME*} – множество названий группы (официальное, аббревиатура, перевод) при наличии; *GTAG* – множество тегов для описания социального статуса группы; *PGDATE* – множество ключевых дат группы; $PGLOC = \{pgcoord, PGNAMELOC, PGLOCDATE, PGNOTE \mid PGLOCDATE \subset \subset PDATE \cup PGDATE\}$ – множество связанных с СП группы геолокаций, включающее название локаций или их координаты, даты и комментарии; $GLINK = \{pgid \mid \exists x, y \in PG : x \cap y \neq 0, pgid \langle x \rangle \neq pgid \langle y \rangle\}$ – множество ссылок на другие профили групп, с которыми установлены связи; *PGLINK* – множество ссылок на участников группы; *DGLINK* – множество ссылок на атрибуты группы из динамического контента; *GTLINK* – ссылки на связанные СЯ; $GSOURCE = \{pgid, sourcelink, sourcename, sourceget, x \mid x \in PG \cap SOURCE\}$ – сведения об источнике исходных данных группы; *TESLINK* – ссылки на словари.

В модели социального профиля для информации, получаемой из различных словарей и справочников, выделено отдельное множество:

$$TESAURUS = \{definition, SEMANTICPROPERTIES, VALUE\}, \quad (11)$$

где *definition* – понятие, определение или термин из словаря; *SEMANTICPROPERTIES* – его семантические отношения, *VALUE* – значение (смысл) понятия, определения или термина.

3. Теоретико-множественное (графовое) представление модели социальной среды

При представлении модели СП в терминах теории графов [3, 6] узлы будут подразделяться на два вида:

– актанты – социальные объекты, состоящие из акторов и артефактов (темы, события, сообщества, местоположения, параметры, медиа-контент и т.д.);

– предикаты – являющиеся социальными отношениями между ними.

Таким образом, простой СП изображается в виде концептуального графа. Для удобства выражения семантики высказываний методология концептуальных графов имеет операции их преобразования обобщением и специализацией:

- Копирование исходного графа.
- Замена исходного графа на граф, в котором вместо меток всех вершин понятий графа используются метки специализации этих понятий.
- Объединение вершин и ребер двух графов.
- Упрощение графа посредством удаления дубликатов вершин и инцидентных им ребер.

При наличии достаточного числа подробных связанных СП личностей возможен переход к задачам исследования социальной микросреды и, в дальнейшем, макросреды. Социальная среда представляет собой безмасштабную (масштабно-инвариантную) сеть [7], способную к постоянному расширению и включающую в свой состав множества социальных профилей и явлений. Она, в отличие от СП, допускает любой актант в качестве центрального объекта и множественные взаимосвязи (косвенные и скрытые в том числе) для них. Понятие социальной среды тесно связано с понятием социальной сети, и в целом для ее исследования могут быть использованы те же методы [8].

Социальная среда обладает свойством эмерджентности, т.е. у нее могут присутствовать свойства, которые проявляются только у среды в целом, а не у отдельных составных частей. То же самое применимо и к ее элементам, представляющим собой множества из более мелких социальных объектов. Например, социальные группы внутри среды должны рассматриваться как монолитные объекты, обладающие уникальными параметрами, но в то же время они включают в себя описания людей, которые рассматриваются не только как атрибут группы, но и как отдельные социальные характеристики. Опять же люди могут быть членами нескольких социальных групп и перемещаться между ними, что ведет к изменению семантики этих групп. Поэтому для описания социальной среды предлагается использовать аппарат метаграфов [9]:

$$SE = \{EV, SP, ER, SR\}, \quad (12)$$

где EV – множество вершин, представленное атрибутами и свойствами социальной среды, взятое из динамического контента; SP – множество метавершин, представленное СП личностей, групп и явлений; ER – множество ребер, показывающее взаимосвязи внутри социальной среды; SR – множество мета-ребер-отношений между включенными социальными профилями. Пример метаграфа показан на рис. 2.

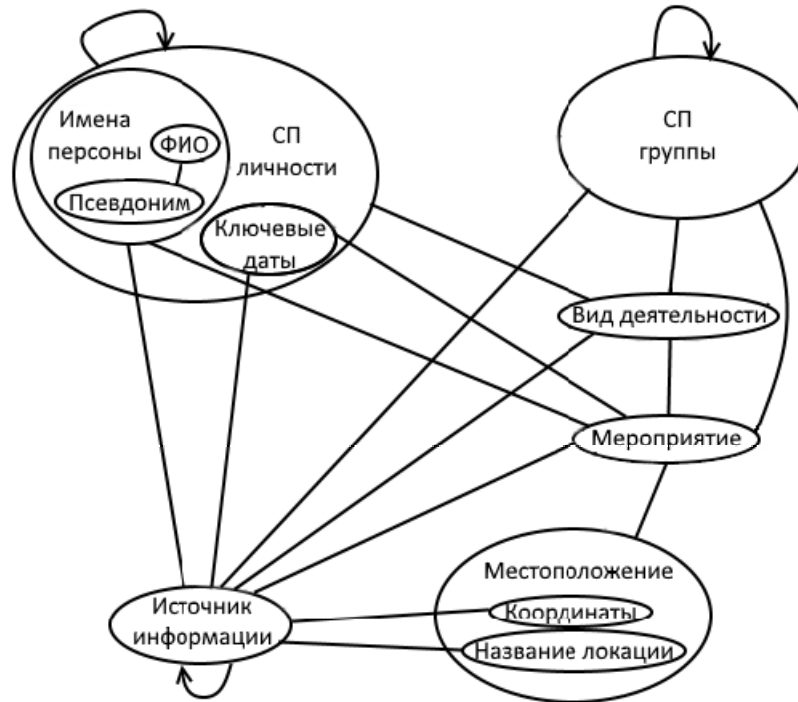


Рис. 2. Изображение части структуры социальной среды в виде метаграфа

В рамках анализа социальной среды в виде метаграфа выделим ряд ключевых расчетных метрик [8, 10]:

1. Число ребер N_{ER} , метаребер N_{SR} , вершин N_{EV} , метавершин N_{SP} в метаграфе.
2. Входная (13) и выходная (14) степени – число ребер и метаребер, инцидентных вершинам:

$$deg_{in}(x_i) = \sum_j y_{ij}, \quad (13)$$

$$deg_{out}(x_i) = \sum_i y_{ij}, \quad (14)$$

где $x_i \in EV \cup SP$; $y_{ij} = [0; 1]$ – отсутствие/наличие ребра между двумя вершинами. Распределение степеней вершин графа социальной среды k подчиняется степенному закону [7]:

$$P(k) \sim k^{-\gamma}, \quad (15)$$

где $\gamma \in [2; 3]$.

3. Плотность графа – отношение числа существующих ребер и метаребер к их возможному количеству, может принимать значение больше единицы. В прикладных задачах исследования социума показывает, насколько сильно связаны между собой социальные элементы, и позволяет вычленять из социальной среды отдельные части для их более детального анализа. Другим применением этой метрики является сравнение состояния социальной среды

в разные моменты времени. В то же время, чем выше плотность, тем сложнее провести анализ, учитывающий все взаимосвязи:

$$D(SE) = \frac{N_{ER} + N_{SR}}{(N_{EV} + N_{SP})(N_{EV} + N_{SP} - 1)}. \quad (16)$$

4. Эксцентриситет – наибольшее геодезическое расстояние между двумя вершинами. Частными случаями являются диаметр графа – максимальный эксцентриситет между любыми двумя вершинами; радиус графа – минимальный эксцентриситет между любыми двумя вершинами:

$$d(SE) = \max_i \left(\max_j \left(L(x_i, x_j) \right) \right), \quad (17)$$

$$r(SE) = \min_i \left(\max_j \left(L(x_i, x_j) \right) \right), \quad (18)$$

5. Промежуточность – число путей L , включающих заданную вершину $x_i : L(x_j, x_k, x_i)$. Позволяет определить суммарное количество связей личности или группы при выделении рассматриваемого СП из описания социальной среды.

6. Центральность графа – характеристика, показывающая наиболее значимые вершины графа. Нахождение центральности в задачах социального управления позволяет определить лидеров и наиболее мобильных представителей социальных групп, значимые явления и объекты социальной среды, авторитетные источники информации и т.д.

Рассмотрим три типа центральности:

1) по близости: обратна сумме длин кратчайших путей от рассматриваемой вершины до всех прочих:

$$C_{prox}(x_i) = \frac{1}{\sum_{j=1}^{N_{EV} + N_{SP}} L_{min}(x_i, x_j)}; \quad (19)$$

2) по степени вершины: центральность прямо пропорциональна числу связей рассматриваемой вершины, является локальной характеристикой ввиду учета только непосредственных соседей и ближайшей окрестности вершины:

$$C_{deg}(x_i) = \frac{deg(x_i)}{deg_{max}}; \quad (20)$$

3) по посредничеству: определяется как доля минимальных расстояний, соединяющих все пары вершин и проходящих через рассматриваемую вершину, является глобальной характеристикой вершины:

$$C_{betw}(x_i) = \sum_{j < k} \frac{L_{min}(x_j, x_k, x_i)}{L_{min}(x_j, x_k)}, \quad (21)$$

7. Средняя длина пути (близость) – среднее арифметическое расстояний между центральной и остальными вершинами в графе. В задачах исследова-

ния социума может показывать степень различия взглядов, уровня жизни, культурных особенностей, удаленность «центра» и «периферии»:

$$Prox(x_i) = \frac{\sum_{j=1}^{N_{EV}+N_{SP}} L(x_i, x_j)}{|L(x_i, x_j)|}. \quad (22)$$

8. Длина цикла $cl(SE)$. Обхват графа – наименьший цикл в графе. Окружность графа – наибольший цикл в графе. На практике возможно применение для исследования механизмов регуляции и обратной связи внутри социума, а также нахождения замкнутых субсред и элит внутри сообществ.

9. Связность – минимальное число ребер N_r , при удалении которых граф разбивается на несколько частей. Позволяет найти ключевые связи и отношения в социуме.

10. Мост-вершина $x \in EV \cup SP$, связывающая отдельные части графа социальной среды. При ее удалении структура сети разрывается, что приводит к росту отдельных узлов. В прикладных задачах управления используется вместе со связностью.

11. Сбалансированность – отсутствие ситуаций «положительное взаимодействие между А и Б, а также А и В, но негативное взаимодействие между Б и В». Считается, что сбалансированные сети более комфортабельны [8] для акторов, что может быть полезным при анализе устойчивости социальных групп и среды в целом.

12. Транзитивность – выполнение условия «если есть взаимодействие между А и Б, а также между Б и В, то имеет место взаимодействие между А и В». Используется для выявления неявных закономерностей и связей внутри социальной среды.

13. Коэффициент кластеризации – позволяет находить ближайших соседей и степень взаимодействия (сплоченность) между ними:

$$CC_i = \frac{2(N_{ER}(x_i) + N_{SR}(x_i))}{deg(x_i)(deg(x_i) - 1)}. \quad (23)$$

Предложенная модель социальной среды позволяет автоматизировать и алгоритмизировать решение ряда задач социального и экономического управления [11]. Например: мониторинг распространения мнений, реакций и идей в обществе в ответ на социально значимые воздействия; предотвращение угроз правонарушений и оценка напряженности общества в территориальном и тематическом аспектах, выявление причин и определение возможных действий по ее уменьшению; управление показателями производства с целью максимизации прибыли и удовлетворения требований потребителей; коллаборативная фильтрация с целью продвижения коммерческих товаров и услуг; медицинские и социальные исследования, направленные на повышение качества жизни населения; управление социальной средой через системы Интернета вещей и «Умный город», где обезличенные статистические данные, взятые из групповых социальных профилей, служат для тонкой подстройки информационно-технического окружения в соответствии с желаниями и потребностями людей.

Заключение

Итоговыми результатами представленной работы можно считать математические модели СП человека, группы и социального явления, разделенных на статическую и динамическую части. Области их практического применения составляют задачи управления социальными процессами с использованием больших неструктурированных данных сети Интернет и учета человеческого фактора. Например, прогнозирование последствий внедрения определенных инноваций и законопроектов или корректировка медицинских рекомендаций на основании сведений о текущей деятельности пациента. Была выявлена необходимость в наличии справочников и словарей для полноценного анализа социальных данных и последующей разработки СП. Был рассмотрен вопрос о представлении социальных профилей отдельных личностей в задаче анализа социума в целом или отдельных социальных групп. Направлением дальнейшей работы является разработка новых и адаптация имеющихся методов к задачам социального профилирования и составления карты состояний социальной среды. Также возникает потребность в использовании нечетких критериев для определения истинности и ценности высказываний, содержащихся в исходных данных СП. Кроме того, для наиболее точной оценки рассматриваемых персон при принятии управленческих решений необходим анализ тональности их контента.

Библиографический список

1. **Осипов, Г. В.** Социологический энциклопедический словарь / Г. В. Осипов. – Москва : М-Норма, 1998. – 480 с.
2. **Бождай, А. С.** Требования к типу и содержанию данных социального профиля человека / А. С. Бождай, А. Ю. Тимонин // Актуальные вопросы современной науки: теория и практика научных исследований : сб. науч. тр. Всерос. науч.-практ. конф. – Пенза : Изд-во ПензГТУ, 2017. – С. 257–260.
3. **Sowa, J. F.** Conceptual Structures: Information Processing in Mind and Machine / J. F. Sowa // Computational Linguistics. – 1986. – Vol. 12, № 3. – С. 218–219.
4. **Тимонин, А. Ю.** Разработка социального профиля на основе имеющихся инструментальных средств / А. Ю. Тимонин // Новые информационные технологии и системы : сб. науч. ст. XII Междунар. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2015. – С. 221–224.
5. **Бождай, А. С.** Метод рефлексивной самоадаптации программных систем / А. С. Бождай, Ю. И. Евсеева // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – № 2 (46). – С. 74–86. – DOI 10.21685/2072-3059-2018-2-7.
6. **Целых, А. А.** Теоретико-графовые подходы к моделированию актор-сетей в исследованиях науки и технологий / А. А. Целых, М. А. Дедюлина // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6, № 4. – URL: https://moit.vivt.ru/wp-content/uploads/2018/10/TselykhDedyulina_4_18_1.pdf (дата обращения: 25.11.2018).
7. **Barabási, A. L.** Emergence of scaling in random networks / A. L. Barabási, R. Albert // Science. – 1999. – Т. 286, № 5439. – С. 509–512.
8. **Чураков, А. Н.** Анализ социальных сетей / А. Н. Чураков // Социологические исследования. – 2001. – № 1. – С. 109–121.
9. **Basu, A.** Metagraphs and Their Applications / A. Basu, R. W. Blanning. – Springer US, 2007. – 172 с. – DOI 10.1007/978-0-387-37234-1

10. **Бабкин, Э. А.** Принципы и алгоритмы искусственного интеллекта : монография / Э. А. Бабкин, О. Р. Козырев, И. В. Куркина. – Нижний Новгород : Нижегородский государственный технологический университет, 2006. – 132 с.
11. **Тимонин, А. Ю.** Social profiles – Methods of solving actual socio-economic problems using digital technologies and Big Data / А. Ю. Тимонин, А. С. Бождай, А. М. Бершадский, О. С. Кошевой // Digital Transformation and Global Society. DTGS 2018. Communications in Computer and Information Science (Россия, г. Санкт-Петербург, 30 мая – 2 июня 2018 г.). – Springer International Publishing, Switzerland, – 2018. – Т. 858. – С. 436–445. – DOI 10.1007/978-3-030-02843-5.

References

1. Osipov G. V. *Sotsiologicheskiiy entsiklopedicheskiy slovar'* [Sociological encyclopedic dictionary]. Moscow: M-Norma, 1998, 480 p. [In Russian]
2. Bozhday A. S., Timonin A. Yu. *Aktual'nye voprosy sovremennoy nauki: teoriya i praktika nauchnykh issledovaniy: sb. nauch. tr. Vseros. nauch.-prakt. konf.* [Actual issues of modern science: theory and practice of research: proceedings of All-Russian scientific and practical conference]. Penza: Izd-vo PenzGTU, 2017, pp. 257–260. [In Russian]
3. Sowa J. F. *Computational Linguistics*. 1986, vol. 12, no. 3, pp. 218–219.
4. Timonin A. Yu. *Novye informatsionnye tekhnologii i sistemy: sb. nauch. st. XII Mezhdunar. nauch.-tekhn. konf.* [New information technologies and systems: proceedings of XII scientific and engineering conference]. Penza: Izd-vo PGU, 2015, pp. 221–224. [In Russian]
5. Bozhday A. S., Evseeva Yu. I. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki* [University proceedings. Volga region. Engineering sciences]. 2018, no. 2 (46), pp. 74–86. DOI 10.21685/2072-3059-2018-2-7. [In Russian]
6. Tselykh A. A., Dedyulina M. A. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii* [Modeling, optimization and information technology]. 2018, vol. 6, no. 4. Available at: https://moit.vivt.ru/wp-content/uploads/2018/10/TselykhDedyulina_4_18_1.pdf (accessed Nov. 25, 2018). [In Russian]
7. Barabási A. L., Albert R. *Science*. 1999, vol. 286, no. 5439, pp. 509–512.
8. Churakov A. N. *Sotsiologicheskie issledovaniya* [Sociological studies]. 2001, no. 1, pp. 109–121. [In Russian]
9. Basu A., Blanning R. W. *Metagraphs and Their Applications*. Springer US, 2007, 172 p. DOI 10.1007/978-0-387-37234-1
10. Babkin E. A., Kozyrev O. R., Kurkina I. V. *Printsipy i algoritmy iskusstvennogo intellekta: monografiya* [Principles and algorithms of artificial intelligence: a monograph]. Nizhniy Novgorod: Nizhegorodskiy gosudarstvennyy tekhnologicheskiiy universitet, 2006, 132 p. [In Russian]
11. Timonin A. Yu., Bozhday A. S., Bershadskiy A. M., Koshevoy O. S. *Digital Transformation and Global Society. DTGS 2018. Communications in Computer and Information Science (Rossiya, g. Sankt-Peterburg, 30 maya – 2 iyunya 2018 g.)*. Springer International Publishing, Switzerland, 2018, vol. 858, pp. 436–445. DOI 10.1007/978-3-030-02843-5.

Тимонин Алексей Юрьевич

аспирант, Пензенский государственный университет (Россия, г. Пенза, ул. Красная, 40)

E-mail: c013s017b301f018@mail.ru

Timonin Aleksey Yurievich

Postgraduate student, Penza State University (40 Krasnay street, Penza, Russia)

Образец цитирования:

Тимонин, А. Ю. Теоретико-множественное математическое моделирование состояния социальной среды / А. Ю. Тимонин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2019. – № 1 (49). – С. 18–33. – DOI 10.21685/2072-3059-2019-1-2.